Tableau: Creating Interactive Data Visualizations

Illustrate your data in a more interactive way by implementing data visualization principles and creating visual stories using Tableau

Packt

LEARNING PATH

Table of Contents

Tableau: Creating Interactive Data Visualizations

Tableau: Creating Interactive Data Visualizations

Illustrate your data in a more interactive way by implementing data visualization principles and creating visual stories using Tableau

A course in three modules

BIRMINGHAM - MUMBAI

Tableau: Creating Interactive Data Visualizations

Published on: August 2016

Published by Packt Publishing Ltd.

Livery Place

35 Livery Street

Birmingham B3 2PB, UK.

ISBN: 978-1-78712-419-6

www.packtpub.com

Credits

Authors

Jen Stirrup

Ashutosh Nandeshwar

Ashley Ohmann

Matt Floyd

Reviewers

Shweta Savale

Darwin Witt

Victor Blær

Mohanganeesh Dorairaj

Joshua N. Milligan

Content Development Editor

Abhishek Jadhav

Graphics

Kirk D'Phena

Production Coordinator

Shantanu N. Zagade

Preface

With increasing interest for data visualization across the media, businesses are looking to design constructive dashboards that captivate the interest as well as liaise data. Tableau makes data available to everyone, and is a great way of dispensing enterprise dashboards across the business. Tableau is an extensive toolkit that lets you create high quality data visualizations effectively.

What this learning path covers

Module 1, *Tableau Dashboard*, introduces you to the theory and practice of delivering dashboards using Tableau through a step-by-step process of creating the building blocks of a dashboard. We then proceed towards the designing principles of putting the dashboard items together. You will learn how to summarize data as a way of conveying key messages on your dashboard along with the introduction to calculations. This module will facilitate structured investigation of data using guided analysis in Tableau. We will also focus on the graphical presentation of data using sparklines, KPIs, maps, and so on. Towards the end of the module, we will look at theming and adding more details to the dashboard by providing examples of more advanced features of Tableau.

Module 2, *Data Visualization with Tableau*, acquaints you with Tableau's user interface and creates perspicacious visualizations. In this module we start off by connecting various data sources, including text, Excel, as well as data sources on a Server. We move on further to create univariate, bivariate, and multivariate charts. This module will also help you create maps by setting geographic variables, placing markers, and overlaying demographic data. We will create new fields using predefined functions, calculate percentages, apply the if-then logic, discretize and aggregate data, manipulate text, and so on. You will be able to modify visualizations by adding information, changing the default marker size and shape settings. Finally, we not only learn to export images and data from the workbook and share them on the Web, but we also explore some of the advanced features of Tableau, such as customizing marker shapes, adding various selectors, and creating animated visualizations.

Module 3, *Creating data stories with Tableau Public*, provides guidelines on how to pursue an enthralling, rich story with data that will enlighten others. By the end of this module, we will create an ideal example of a dashboard that focuses on an issue that impacts everyone. We begin with an overview of the functions of Tableau Public along with its installation. Furthermore, you will be familiarized with various features in Tableau Public, such as cards, shelves, and ShowMe. This module will teach you how to format source data and explain some basic data modeling, such as Dimensions, Measures, and Joins. Topics such as Visualization, Calculation and Dashboard designing, which are studied in the previous modules, will be covered in detail. Finally, the module will explain how to build filters with their use in dashboards and familiarize you with the various methods to embed data visualization in blog posts, websites, and offline documents

What you need for this learning path

You need the following in order to work with Tableau:

- Tableau Version 8.2
- Windows Live login ID and password
- Microsoft Excel
- Internet access

Users only need to download the Tableau Public client. The technical specifications for Tableau Public mirror those of Tableau Desktop Personal and are listed on the Tableau website at http://www.tableau.com/products/desktop. According to Tableau system requirements, PC users require the following minimum specifications:

- Microsoft Windows Vista SP2 or newer (32-bit and 64-bit)
- Microsoft Server 2008 R2 or newer (32-bit and 64-bit)
- Intel Pentium 4 or AMD Opteron processor or newer (SSE2 or newer required)
- 2 GB memory
- 750 megabytes minimum free disk space

Who this learning path is for

Data scientists who have just started using Tableau and want to build on the skills using practical examples. Familiarity with previous versions of Tableau will be helpful, but not necessary.

Reader feedback

Feedback from our readers is always welcome. Let us know what you think about this book—what you liked or disliked. Reader feedback is important for us as it helps us develop titles that you will really get the most out of.

To send us general feedback, simply e-mail <feedback@packtpub.com>, and mention the book's title in the subject of your message.

If there is a topic that you have expertise in and you are interested in either writing or contributing to a book, see our author guide at www.packtpub.com/authors.

Customer support

Now that you are the proud owner of a Packt book, we have a number of things to help you to get the most from your purchase.

Downloading the example code

You can download the example code files for this book from your account at http://www.packtpub.com. If you purchased this book elsewhere, you can visit http://www.packtpub.com/support and register to have the files emailed directly to you.

You can download the code files by following these steps:

1. Log in or register to our website using your e-mail address and password.
2. Hover the mouse pointer on the **SUPPORT** tab at the top.
3. Click on **Code Downloads & Errata**.
4. Enter the name of the book in the **Search** box.
5. Select the book for which you're looking to download the code files.
6. Choose from the drop-down menu where you purchased this book from.
7. Click on **Code Download**.

You can also download the **code files** by clicking on the Code Files button on the book's webpage at the Packt Publishing website. This page can be accessed by entering the book's name in the **Search** box. Please note that you need to be logged in to your Packt account.

Once the file is downloaded, please make sure that you unzip or extract the folder using the latest version of:

- WinRAR / 7-Zip for Windows
- Zipeg / iZip / UnRarX for Mac
- 7-Zip / PeaZip for Linux

The code bundle for the book is also hosted on GitHub at https://github.com/PacktPublishing/Tableau-Creating-Interactive-Data-Visualizations. We also have other code bundles from our rich catalog of books and videos available at https://github.com/PacktPublishing/. Check them out!

Errata

Although we have taken every care to ensure the accuracy of our content, mistakes do happen. If you find a mistake in one of our books—maybe a mistake in the text or the code—we would be grateful if you could report this to us. By doing so, you can save other readers from frustration and help us improve subsequent versions of this book. If you find any errata, please report them by visiting http://www.packtpub.com/submit-errata, selecting your book, clicking on the **Errata Submission Form** link, and entering the details of your errata. Once your errata are verified, your submission will be accepted and the errata will be uploaded to our website or added to any list of existing errata under the Errata section of that title.

To view the previously submitted errata, go to https://www.packtpub.com/books/content/support and enter the name of the book in the search field. The required information will appear under the **Errata** section.

Piracy

Piracy of copyrighted material on the Internet is an ongoing problem across all media. At Packt, we take the protection of our copyright and licenses very seriously. If you come across any illegal copies of our works in any form on the Internet, please provide us with the location address or website name immediately so that we can pursue a remedy.

Please contact us at <`copyright@packtpub.com`> with a link to the suspected pirated material.

We appreciate your help in protecting our authors and our ability to bring you valuable content.

Questions

If you have a problem with any aspect of this book, you can contact us at <`questions@packtpub.com`>, and we will do our best to address the problem.

Part 1. Module 1

Tableau Dashboard

Over 40 recipes for designing professional dashboards by implementing data visualization techniques

Chapter 1. A Short Dash to Dashboarding!

In this chapter, we will cover:

- Preparing for your first dashboard
- Showing the power of data visualization
- Connecting to data sources
- Introducing the Tableau interface
- Interacting with your first data visualization
- Sharing your visualization with the world

Introduction

This chapter starts with you being a Tableau beginner, then quickly moves you forward to creating your own visualizations and explains how to interact with the Tableau sample dashboards—how to find, open, and interact with them.

We can create visualizations by using Tableau in order to produce meaningful dashboards that communicate clearly. The six recipes in this chapter will explain how we can get up to speed with Tableau very quickly in order to produce dashboards that facilitate and expedite the decision-making process for strategic decision makers and operational team members within your organization.

Tip

For this book, we will be using version 8.1 to work with Tableau.

Preparing for your first dashboard

The following definition has been taken from the *Intelligent Enterprise* magazine's March 2004 issue:

> *A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance.*
>
> --*Stephen Few*

For an enterprise, a dashboard is a visual tool to help team members throughout the ranks of the organization to track, monitor, and analyze the information about the organization in order to make decisions to support its current and future prosperity. In this recipe, we will interact with Tableau's sample dashboards, which are constructed from worksheets. People often learn by example, and this is a straightforward way of inspiring you with dashboard samples while also learning about Tableau.

What do dashboards help you to do?

- **Evaluate**: Dashboards answer questions such as "Have the goals and objectives been met? Are we on track?"
- **Reveal**: Dashboards help you view and digest information very quickly, which means you have more time for strategic planning.
- **Communicate**: Using a visual tool can help to get the message across in a common format and create impact.
- **Certainty**: Dashboards help you to have confidence in your insights.

Dashboards help key team members to gain insights and discern the health of the organization very quickly. Tracking, monitoring, and analyzing the organization's data is an essential part of making accurate decisions.

Tableau provides a number of example dashboards, both online and as part of the Tableau Desktop installation. We will find, open, and interact with sample Tableau dashboards.

We can also use the example dashboards as a basis to make our own dashboards. They can form a source of inspiration to make your own compelling visualizations. For the purpose of this recipe, we will focus on the sample Sales workbook.

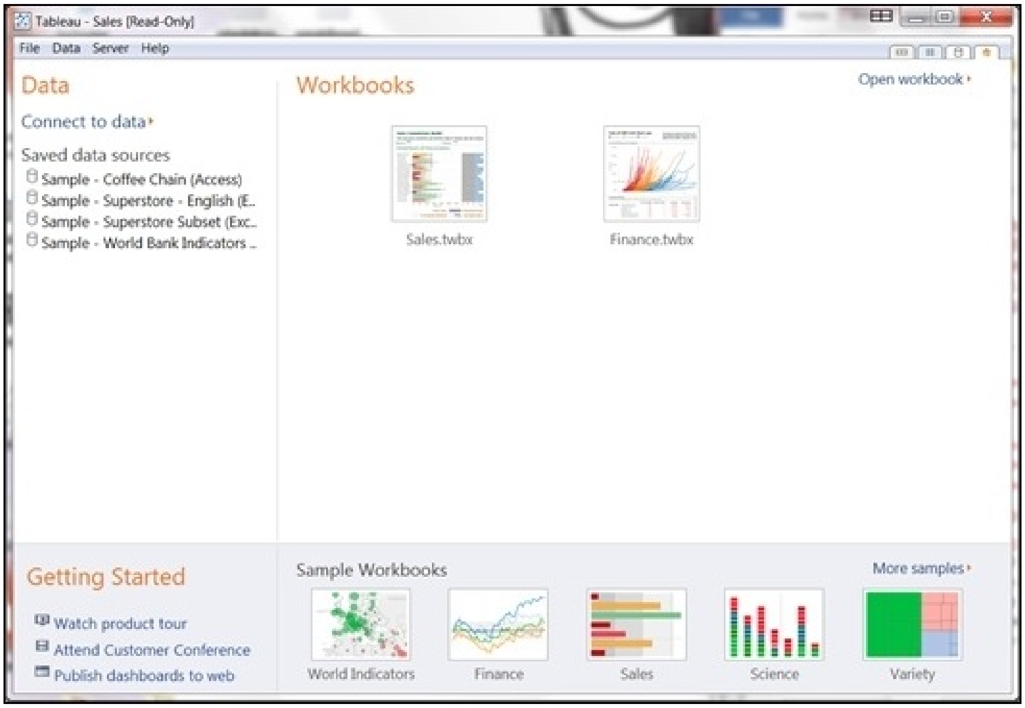
A key feature of dashboards is that they are interactive. There are different types of dashboards, and some references are included at the end of this recipe. A key feature of dashboards is their interactivity. Dashboards are not simply a set of reports on a page; they should tell a story about the business when they are put together. They should answer a clear business question. In order to facilitate the decision-making process, interactivity is an important part of assisting the decision-maker to get to the heart of the analysis as quickly as possible.

Fortunately, it is straightforward to interact with a dashboard that has been implemented in Tableau.

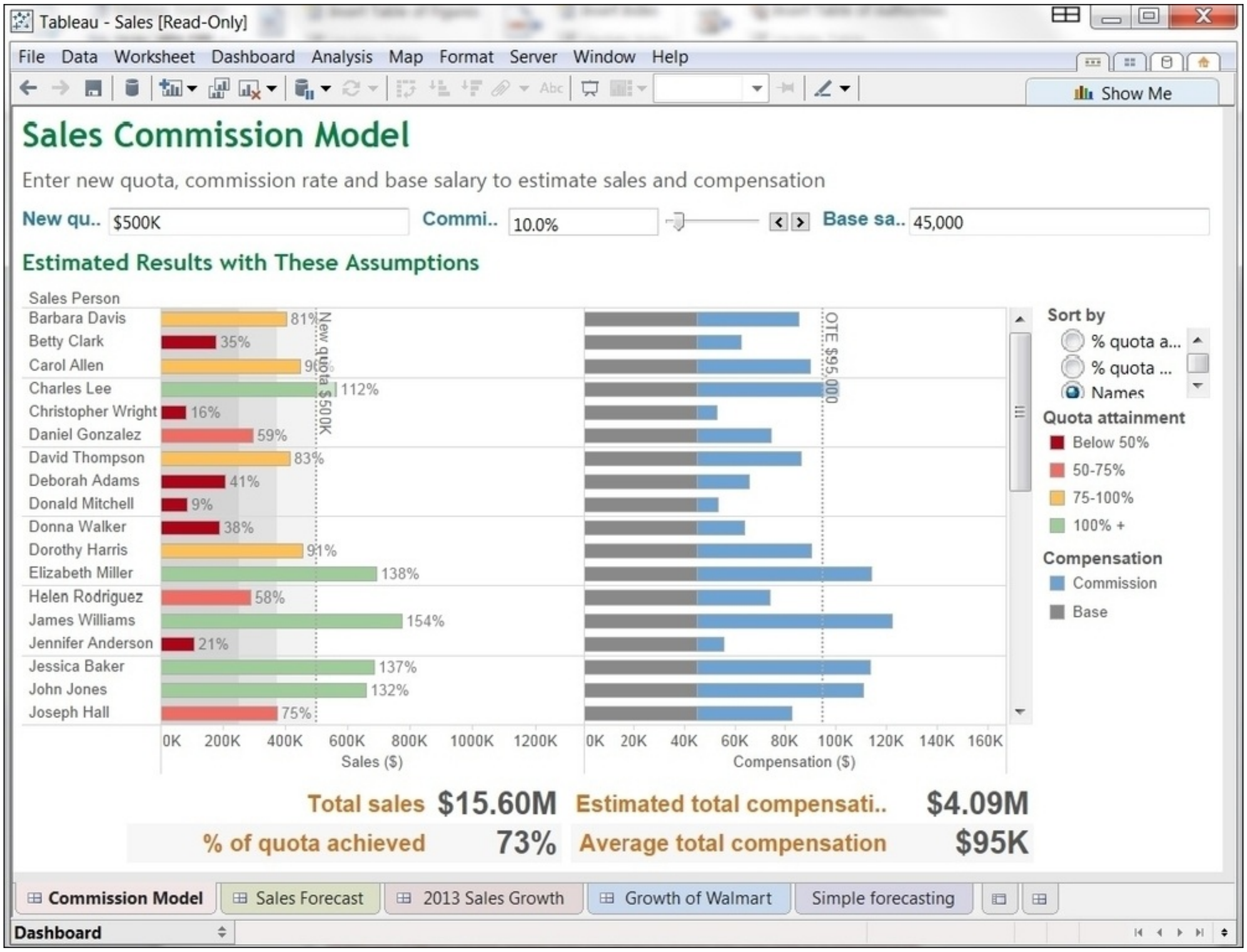
How to do it...

We will perform the following steps to see how we can interact with a dashboard:

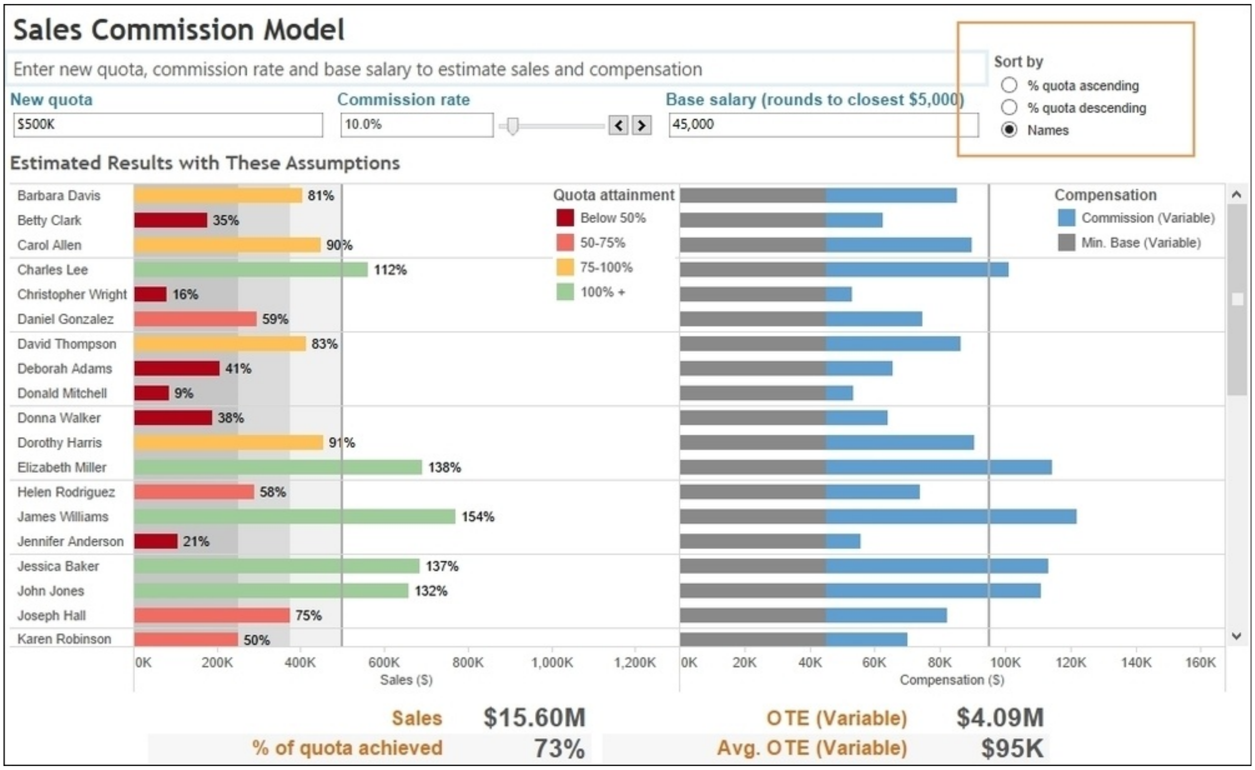
1. Open up the Tableau Desktop and you can see the **Getting Started** page. The following screenshot is an example:



- At the bottom of the entry page, you can see a section called **Sample Workbooks** that contains some examples.
- Let's take a look at the **Sales** dashboard. If you double-click on the **Sales** example, it will open and you will see the sample **Sales** dashboard, as shown in the following screenshot:



- A worksheet is like a tab in Excel; it is a data visualization on its own. A workbook, on the other hand, is a collection of worksheets. In Tableau, a dashboard allows you to combine and manipulate the worksheets together. Let's interact with this dashboard straightaway using the **Sales** dashboard sample that has been provided by Tableau. On the right-hand side of the dashboard, you can see a box called **Sort by**. You can see an example of this in the following screenshot, where the relevant section has been highlighted with a box:

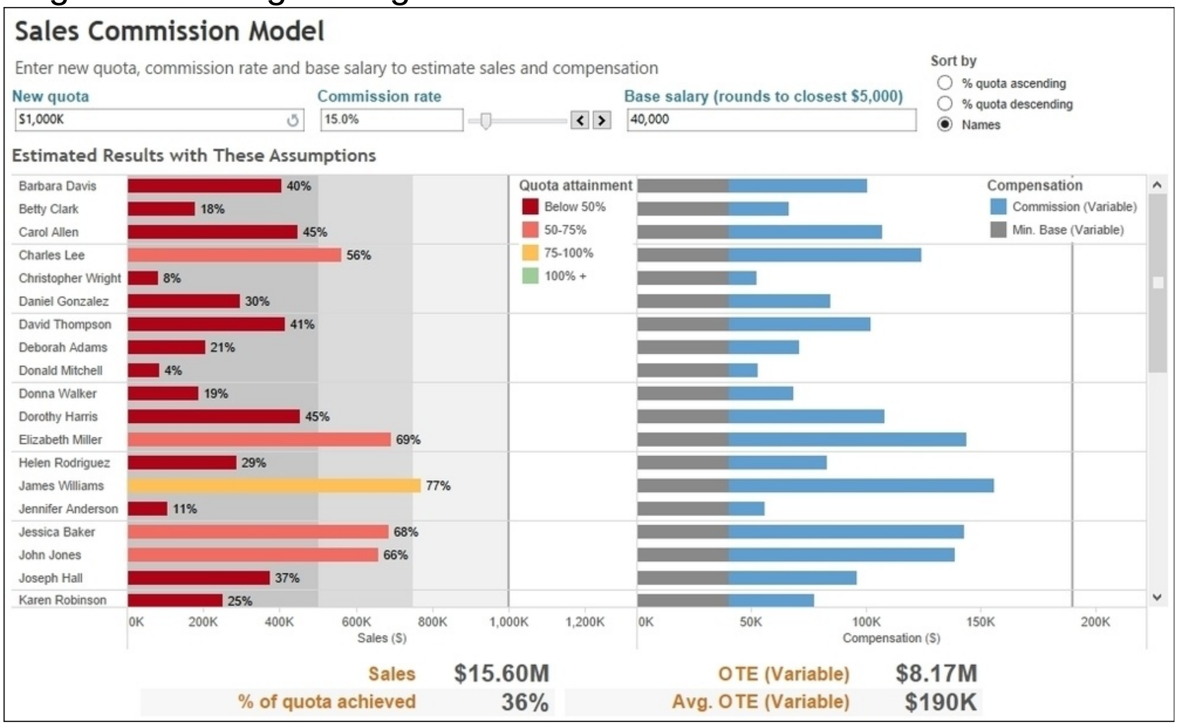


When you click on the middle item, denoted as **% quota descending**, you can see that the horizontal bar charts in the main area of the dashboard change very quickly in response to the user interaction. The dashboard now looks quite different from the previous Tableau example, where the bars were sorted by **Names**. The rapidity of the change means that decision makers can "think as they click" in order to focus on their analysis.

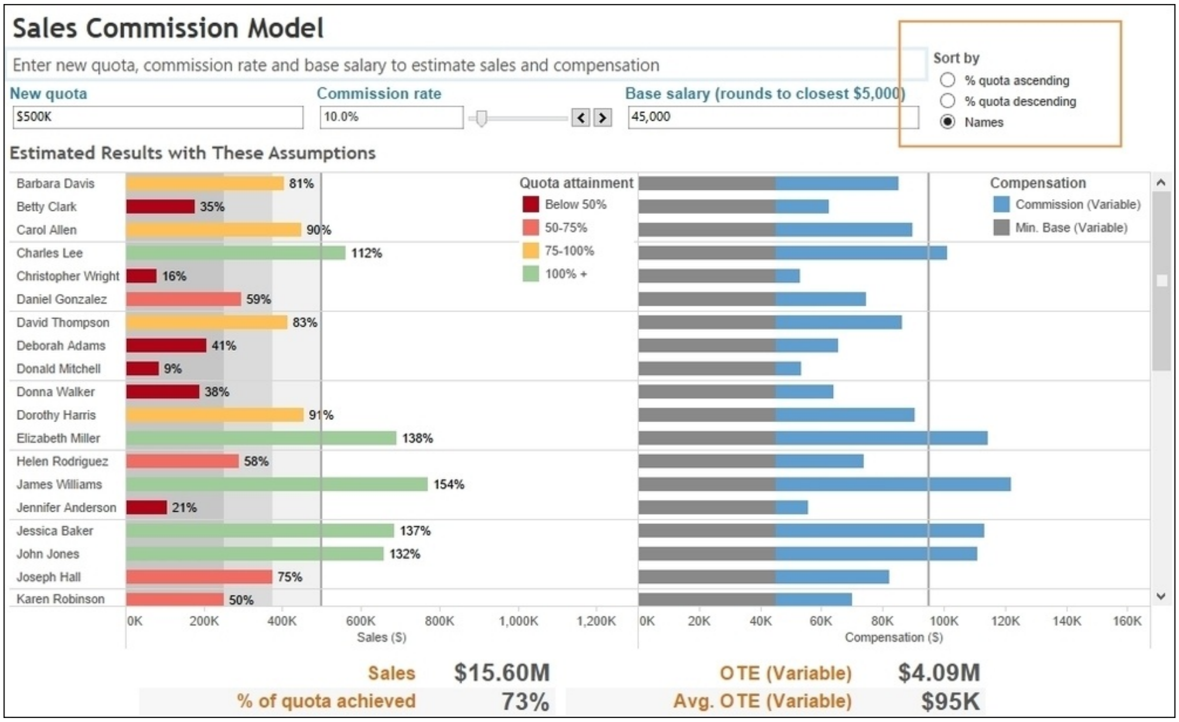
There are a number of different ways in which Tableau can offer useful interactivity for dashboards. For example, we can include sliders, filtering by color, moving from dashboard to dashboard, radio buttons, drop-down lists, and timelines. For example, another interesting feature is that users can enter values into parameters in order to see the impact of their activity. A parameter is a dynamic value that responds to user input. In this example, we use it to filter the data by replacing constant values in calculations. We use the following steps to view the interactivity:

1. Let's see the impact of interactivity on the performance information given by the dashboard. In the **Sales** dashboard, increase the **New quota** level to `$1,000,000`.

2. Next, increase the value in the **Commission rate** textbox to **15.0%** by moving the slider to the right.
3. Decrease the base salary to `$40,000` by inserting this value in the **Base salary** textbox. Note that the estimated results are now quite different. You can see from the following screenshot that the number of people making the sales target decreases, and the chart now shows a significant increase in the number of people nearing their target or missing it altogether:



In the previous screenshot, note that the colors of the **Estimated Results with These Assumptions** bars have changed so that all of them now show red or yellow. The green bars have disappeared. This gives a visual cue that the estimated results have changed considerably for the worse after we made changes to the filter. We can also see this due to the presence of the target line, which shows whether the individual met his/her target or not. The following screenshot depicts this:



How it works…

Tableau gives you a series of sample dashboards as part of the installation. You can also see more samples online. Some samples are provided by Tableau team members, and you can also visit the Tableau website for samples submitted by keen data visualization fans from around the world. These samples can help to inspire your own work.

In this topic, we compared the changes on a dashboard in order to see how Tableau responded to changes. We noted that the color has changed along with the values. The dashboard provides quick feedback that the values do not change favorably for the new quotes, commissions, and base salary. When decision makers are interacting with dashboards, they are expecting quick-as-a-flash responsiveness from the dashboard, and the sample Tableau dashboards meet this expectation well.

See also

Tableau offers a number of sample dashboards on its website, and it is worthwhile to check the site for ideas and brainstorming for your own dashboards. Please take a look at www.tableausoftware.com for examples. If you are interested in dashboard theory in general, then you can look at the following references:

- *Dashboard Confusion* by *Stephen Few*, *Intelligent Enterprise*, *2004*
- *5 Best Practices for Creating Effective Dashboards* by *Tableau Software* (http://www.tableausoftware.com/learn/whitepapers/5-best-practices-for-effective-dashboards)

Showing the power of data visualization

Dashboards rely on the power of visualization in order to let people see the message of the data to make effective decisions. How can you show the power of a dashboard when compared to a crosstab table?

In this recipe, we will see how a data visualization can have more impact than a straightforward crosstab. We will make a crosstab table in Tableau and then turn it into a data visualization to see the impact in action!

Understanding your data is an essential part of data visualization, regardless of the technology you are using. Tableau can help you to understand your data by automatically distinguishing between measures and dimensions. How do you know which are which? Look at the title of a report or dashboard. For example, if a dashboard is called `Sales by Country`, then anything that comes after the `by` word is a dimension and the item being counted is a measure. Dimensions and measures are explained as follows:

- **Dimensions**: Dimensions describe data
- **Measures**: Measures are usually numbers

In this recipe, we will look at the difference between a plain table and a graphical representation of the data. While tables are data visualizations in themselves, Tableau's power lies in its ability to visualize data graphically and quickly. This recipe will demonstrate the ease of going from a table to a picture of the data. We will create a map, and the color intensity of the map coloring reflects the value.

Getting ready

Let's start by opening up Tableau to get ready for your first visualization.

We will need to get some data. To obtain some sample, download the

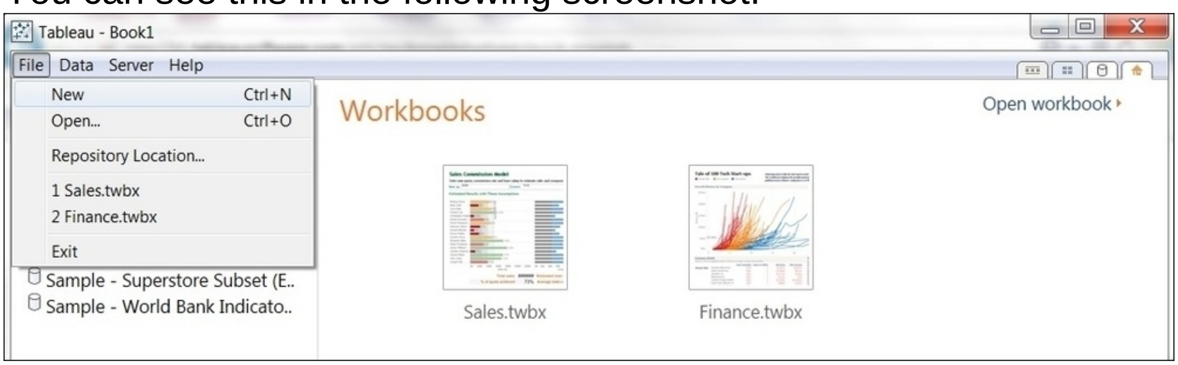
Unicef Report Card spreadsheet from the following link:
http://bit.ly/TableauDashboardChapter11Unicef

It will have the following columns:

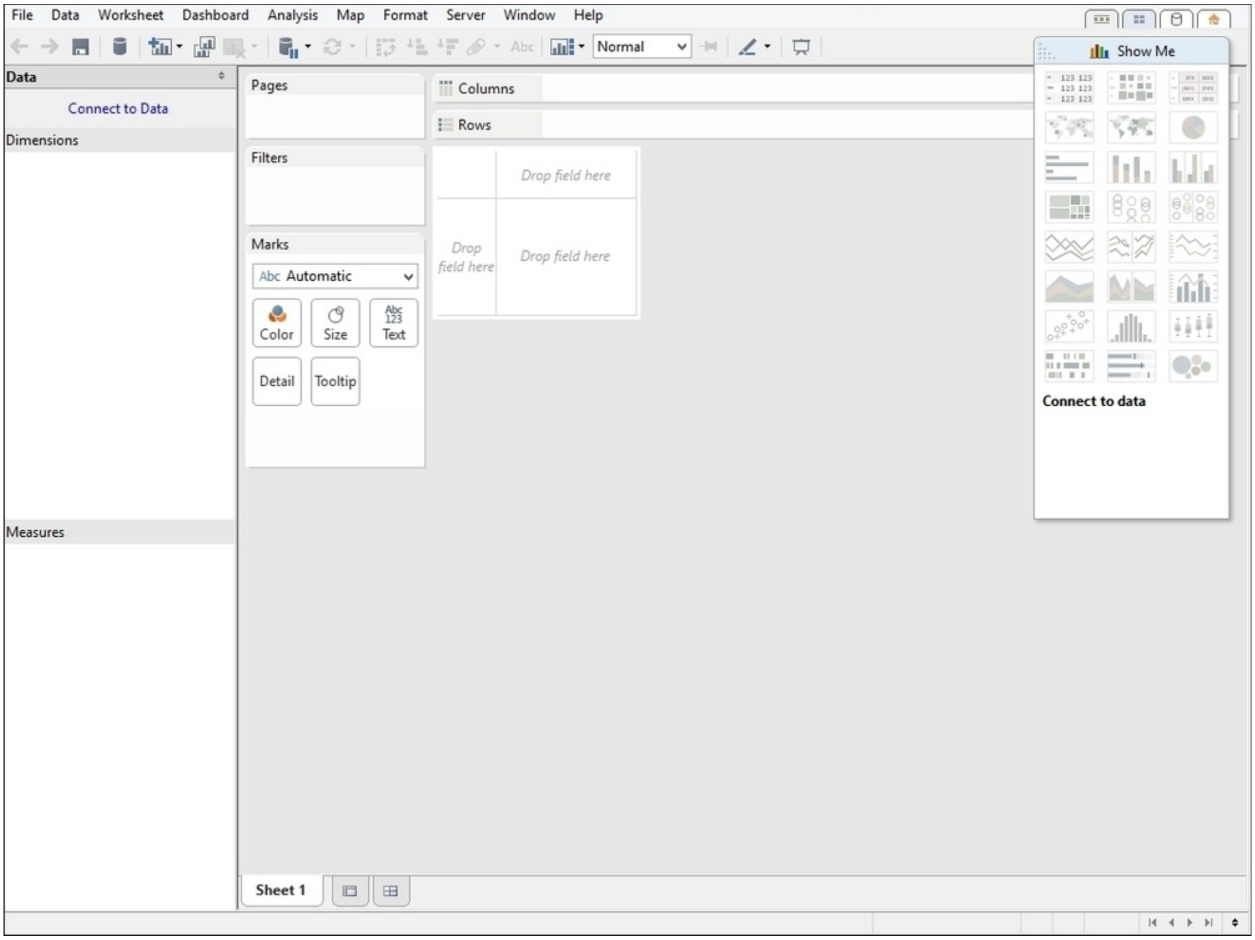
- **Country**
- **Average ranking position (for all 6 dimensions)**
- **Material well-being**
- **Health and Safety**
- **Educational well-being**
- **Family and peer relationships**
- **Behaviors and risks**
- **Subjective well-being**

How to do it…

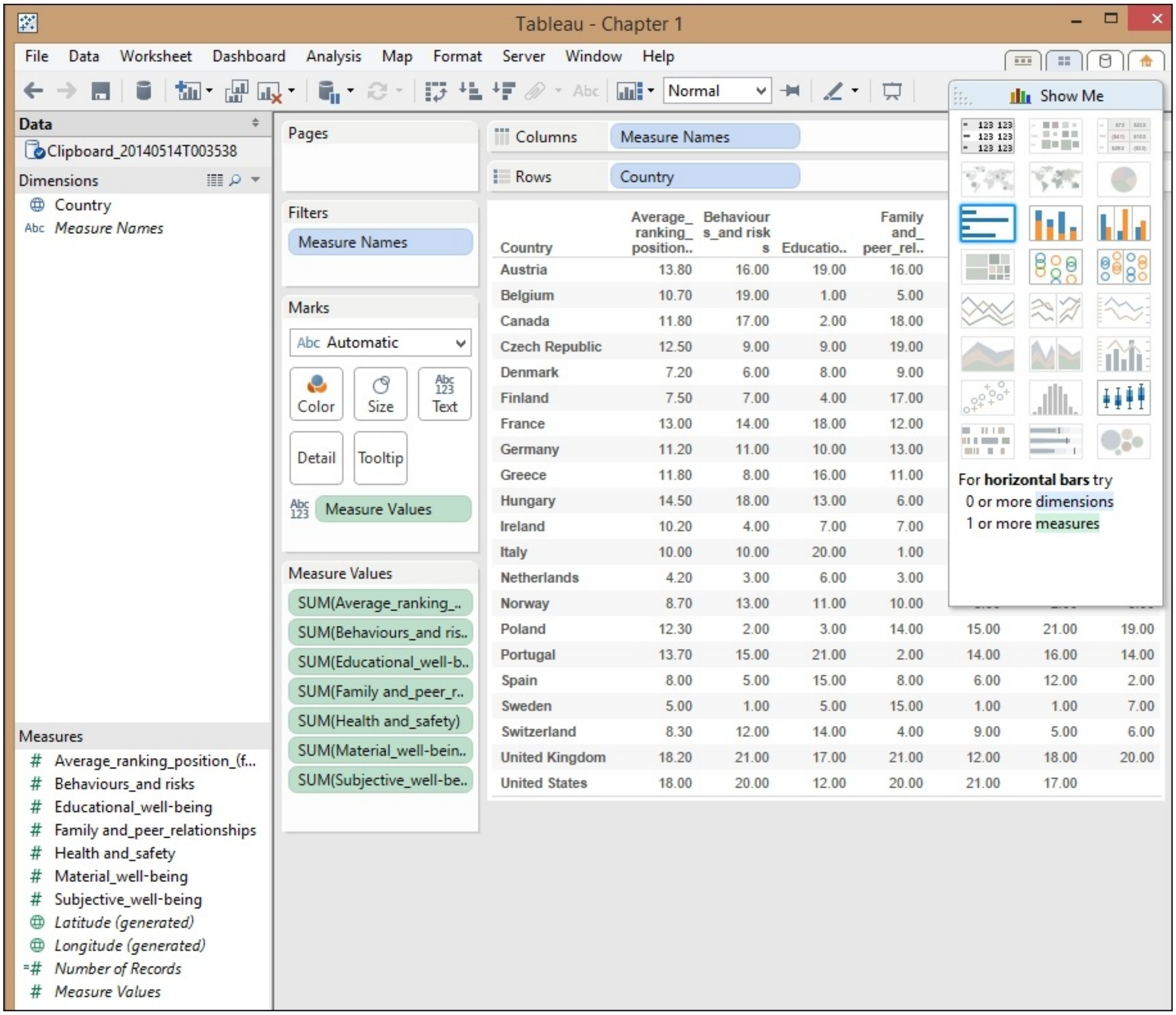
1. In Tableau, click on **File** in the top left-hand corner and click on **New**. You can see this in the following screenshot:



- When you've clicked on **New**, you will get a blank Tableau workbook.
This is shown in the following screenshot:



- Let's insert our downloaded data. To do this, go to the Excel spreadsheet and select all of the data by pressing *CTRL + A*.
- Once you have done this, go to Tableau and press *CTRL + V* to paste it. Here is an example of the workbook:



The following points describe the different panels in Tableau:

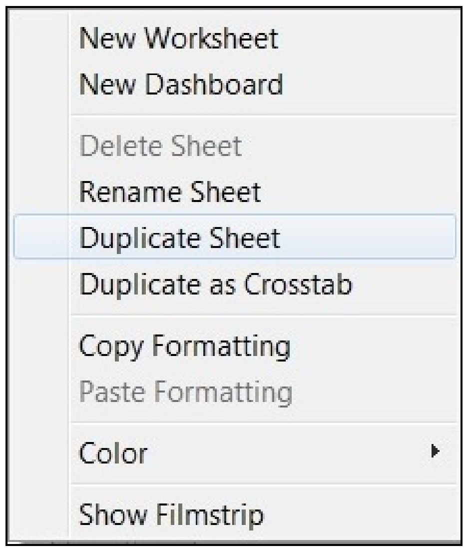
- **Data window:** This holds the measures, dimensions, and calculations in the data
- **The Tableau view:** This shows the items held in the **Rows**, **Columns**, **Marks**, **Pages**, or **Filters** shelf
- **Shelf:** This is a place where you drag a measure or a dimension

The following steps can be performed to create a quick visualization:

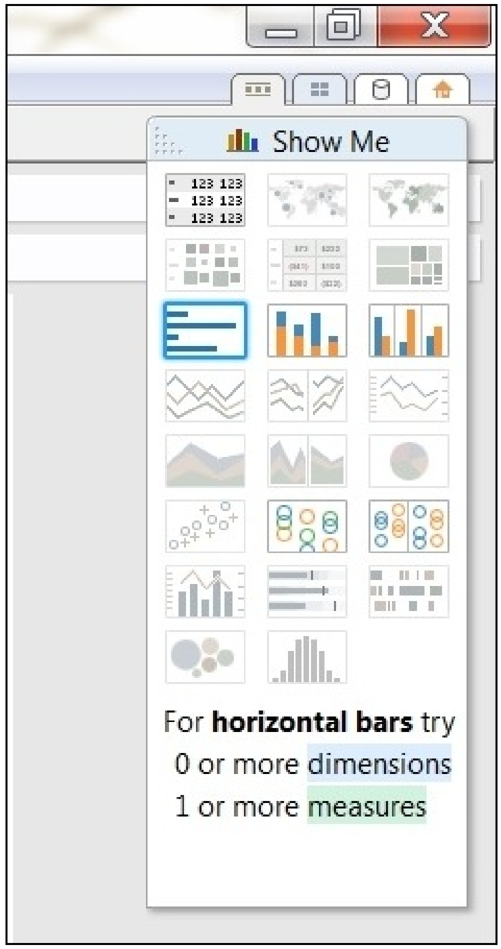
1. When you paste the data, it appears as a crosstab. We can see the data, but it is quite difficult to see any patterns in the data.
2. Using the preceding list as a basis, it is very simple to create a quick

visualization.

3. Let's take a copy of our work so that we can compare before and after. To do this, click on the **Sheet 1** tab at the bottom of the worksheet. Right-click on the worksheet tab at the bottom of the Tableau interface and a pop-up menu appears.
4. Select the option **Rename Sheet** and rename the worksheet as `Before`.
5. Then, choose the option **Duplicate Sheet**, as shown in the following screenshot, to take a copy of the worksheet, and rename the new copy as `After`.



6. In the `After` worksheet, look for Tableau's **Show Me** feature. This is a key feature of Tableau, and you can see the **Show Me** toolkit at the right-hand side of the Tableau interface, as shown in the

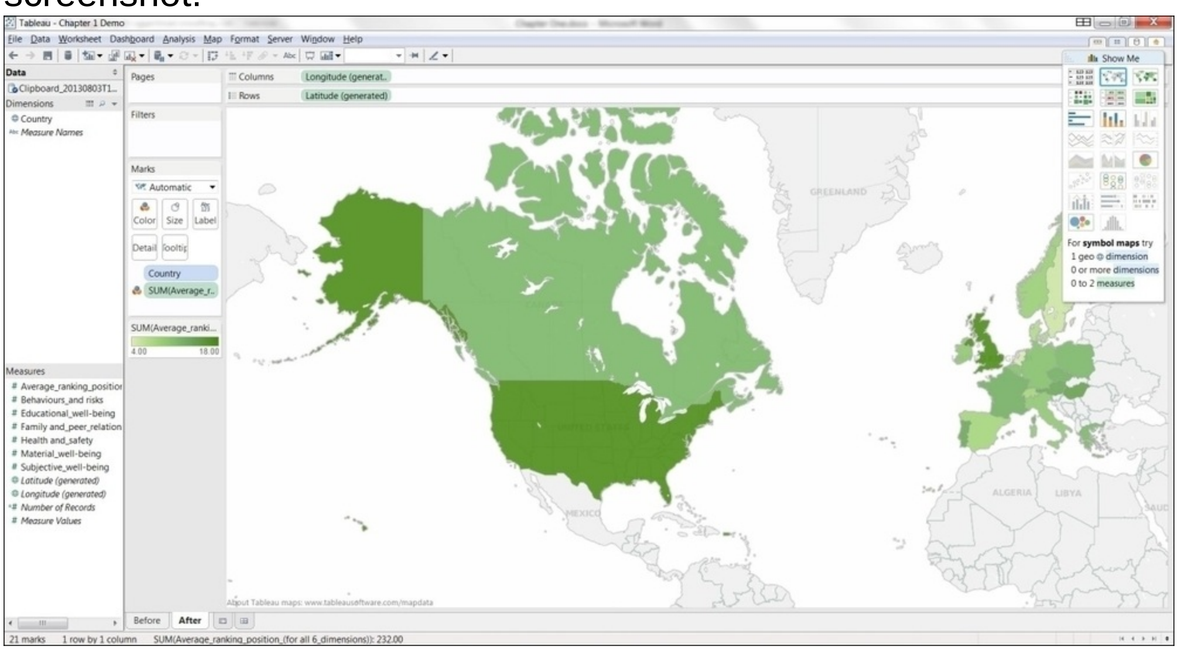
following screenshot: 

For the purposes of this recipe, we will choose a map visualization.

7. Using the `After` worksheet, click on the first **Measures** column called **Average ranking position_(for all 6 dimensions)** to select it. Right-click on the column and choose **Keep Only**. This excludes the rest of our measures, retaining only this column. The result can be seen in the following screenshot:

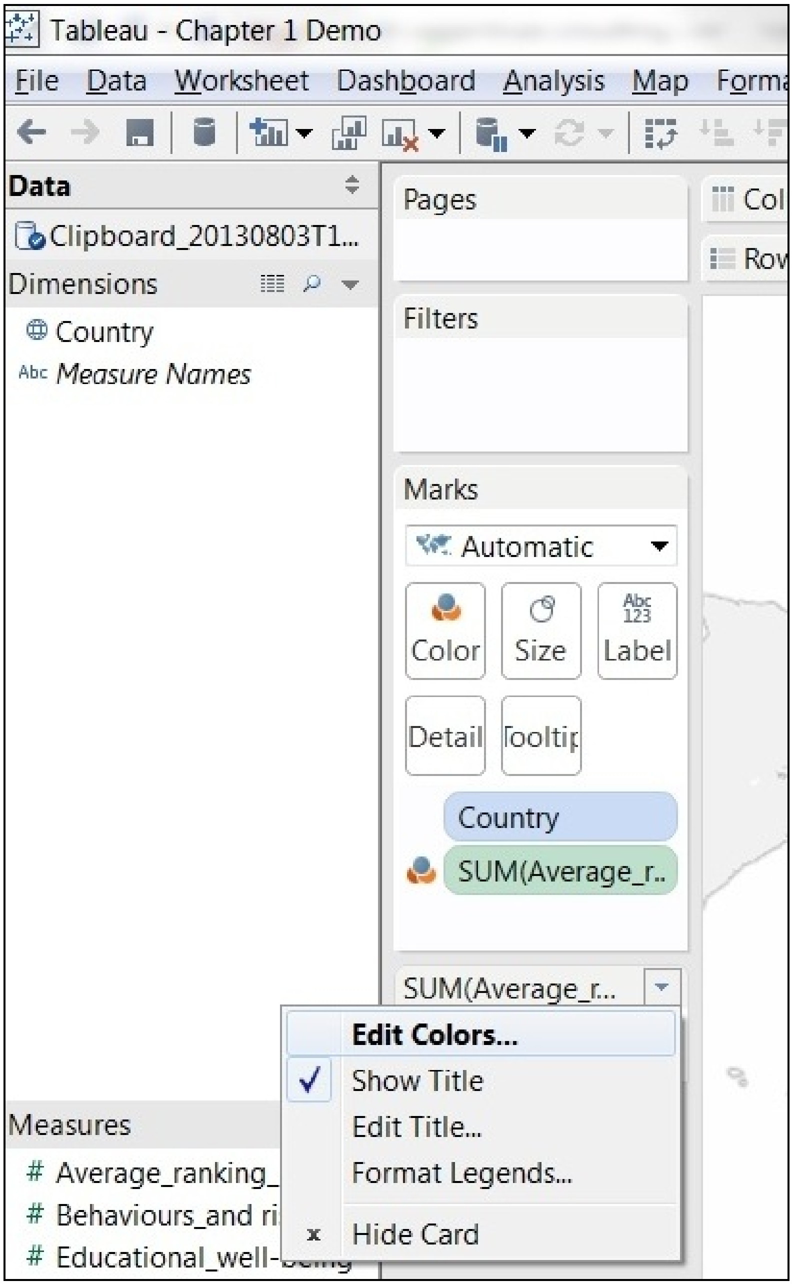
8. When we exclude the other options, the **Show Me** toolkit changes in response to the amendments that have been made in the data table. Now, the map options are available to us. The **Show Me** toolkit changes can be seen in the following screenshot:

When we select the **filled maps** option, which is bordered with a heavy line at the top right-hand side row, our screen now changes to look like a filled map, in which each color corresponds to the average rank of each country. An example is shown in the following screenshot:

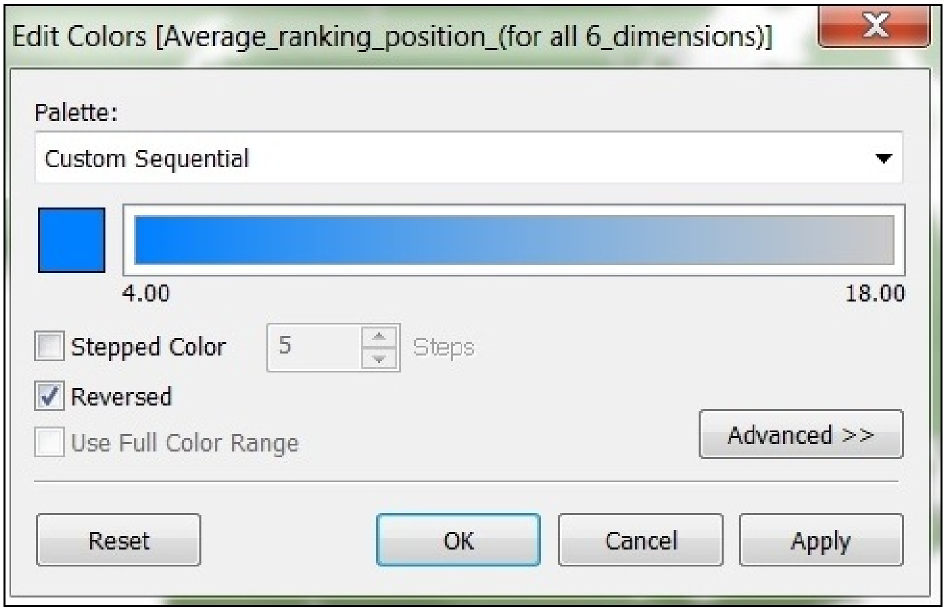


> **Note**
>
> We have Denmark ranked as 7 and the United Kingdom is ranked at 18. Denmark is considered as having a higher ranking, even though it has a lower number.

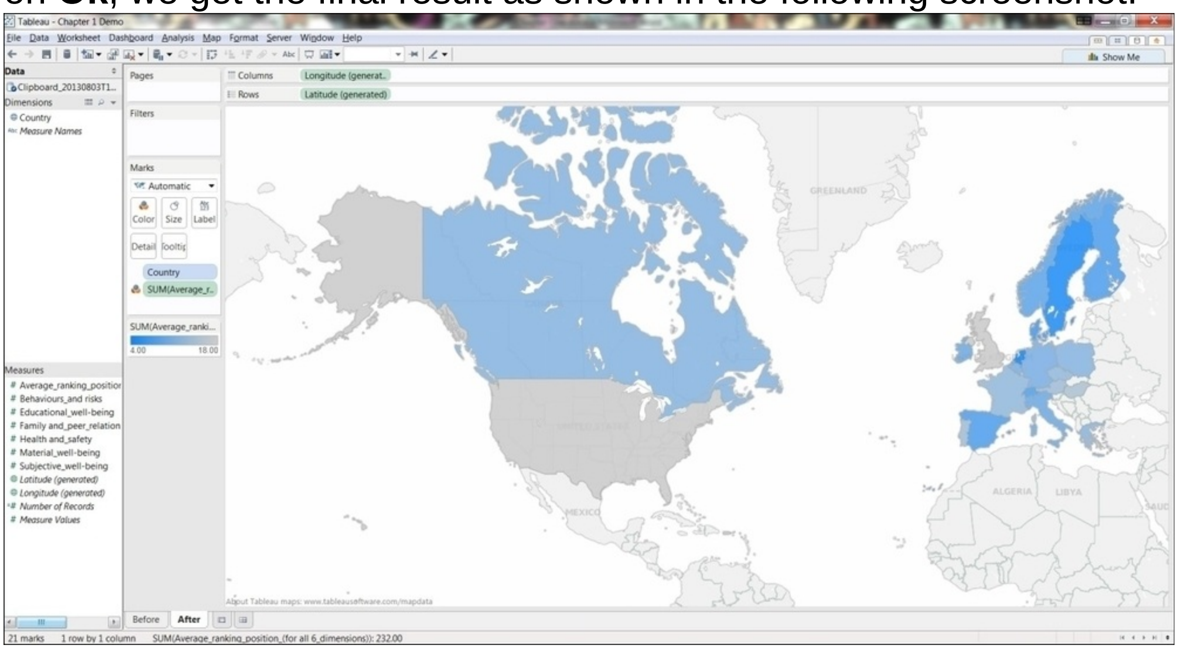
9. To change the color settings, we right-click on the colors item that is located on the left side of the screen, centered vertically. We can see an example in the next screenshot:



The **Edit Colors** dialog box appears. An example can be found in the next screenshot:



10. Using the square box, you can change the color. Here, it has been changed to blue. The important item to note here is the **Reversed** option. This option allows us to reverse the color so that lower numeric values are represented by higher intensities. When we click on **Ok**, we get the final result as shown in the following screenshot:



How it works…

The **Show Me** toolkit takes the guesswork out of what data visualization tool to choose by offering you a selection of visualizations that are based on your data types.

The **Show Me** button helps you to choose which data visualization is most suited to your data. It does this using an in-built, intelligent, knowledge-based system that is part of Tableau. This helps to take the guesswork out of selecting a data visualization, which can often be a contentious issue among data consumers and business intelligence professionals alike.

Data visualization is telling a story; the value is depicted by a corresponding color intensity. This example topic involved ranking data. Therefore, the higher the number, the lower the value actually is. Here, the value refers to the country rank.

How can we make the message clearer to the users? When we visualize the data in a map, we can still use color in order to convey the message. Generally speaking, we assume that the brighter or more intense a color is, then the higher the value. In this case, we need to adapt the visualization so that the color is brighter in accordance with the rank, not the perceived integer.

There's more…

Color theory is a topic in itself, and you will see practical applications as we proceed throughout this book. For further references, please see the *See also* section.

See also

- Data visualizations can also be known as dataviz for short. On Twitter, `#dataviz` is a well-used hashtag

Connecting to data sources

In the previous recipe, we inserted data into the Tableau workbook by simply copying and pasting. In the real world, however, we need to be able to connect to different data sources that may contain large amounts of data.

We will now look at connecting to multiple data sources at a time. This is a useful way of enriching our data. We have access to multiple data sources. We can open up Tableau and connect numerous data sources.

First, we will see how we can connect to the Windows Azure Datamarket cloud data source, and then continue to connect to the local Excel file. Windows Azure Marketplace is an online market to buy and sell finished **Software as a Service** (**SaaS**) applications and premium data. Some data on Windows Azure Datamarket is free. We will be using one of the free data samples, which will give us a lot of information about individual countries, such as the country code, population, size, and so on. In data warehousing terminology, this data can be considered as a dimension, which is another way of describing data. In this definition, it is a field that can be considered an independent variable, regardless of the datatype. Tableau has a more specific definition of a dimension. Tableau treats any field containing qualitative, categorical information as a dimension, such as a date or a text field.

To connect the online data and local data, we will connect to Windows Azure Datamarket using OData, which is a standardized protocol to provide **Create, Read, Update, Delete** (**CRUD**) access to a data source via a website. It is the data API for Microsoft Azure, but other organizations use it as well, such as eBay, SAP, and IBM.

Getting ready

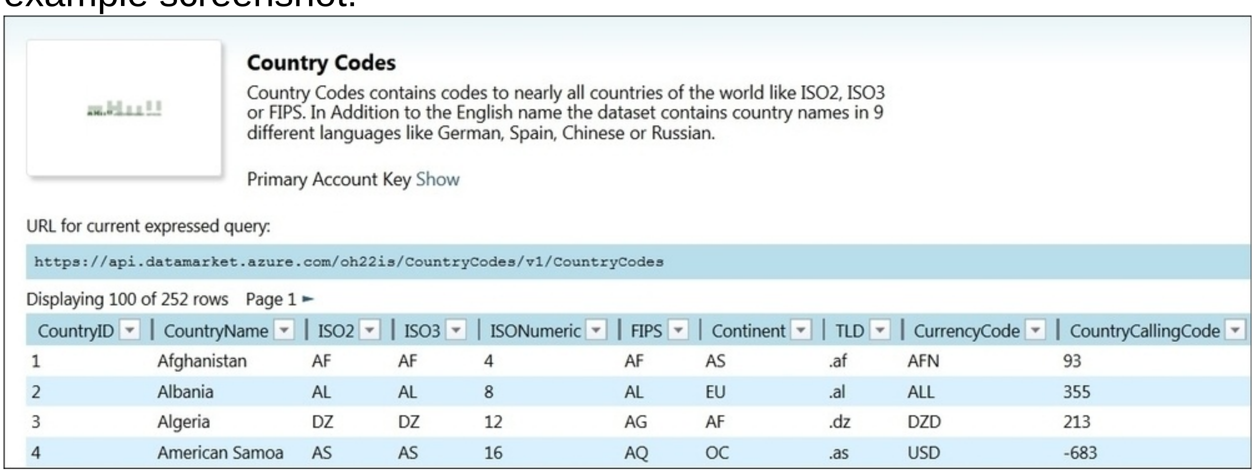
Before you start, you need to create a folder where you can download data to run through the examples. You should pick a folder name that is meaningful for you. Also, be sure to select a location that has plenty of space. In this example, we will use the following location to store data: `D:\Data\TableauCookbook`. For the example in this chapter, we will create a folder called `Chapter 1`.

How to do it…

1. To connect to Windows Azure Datamarket, please sign up for a free account using a Windows Live ID. To do this, please visit https://datamarket.azure.com/ and follow the instructions. This may involve activating your account via a link, so please follow the instructions carefully.
2. Sign in to Windows Azure Datamarket and navigate to the following URL:
   https://datamarket.azure.com/dataset/oh22is/countrycodes#schema

3. About half way down the page, look for the **Sign Up** button and click on it.
4. This will take you to a terms and conditions page. After you've read the terms and conditions, and, if you agree with them, tick the box to specify that you agree and click on **Sign Up**.
5. This will take you to a **Thank You** page. Look for the **EXPLORE THIS DATASET** link on this page and click on it, as shown in the following screenshot:

- When you click on **EXPLORE THIS DATASET**, you will be able to see the data appear in the browser, which you can slice and dice. Here is an example screenshot:

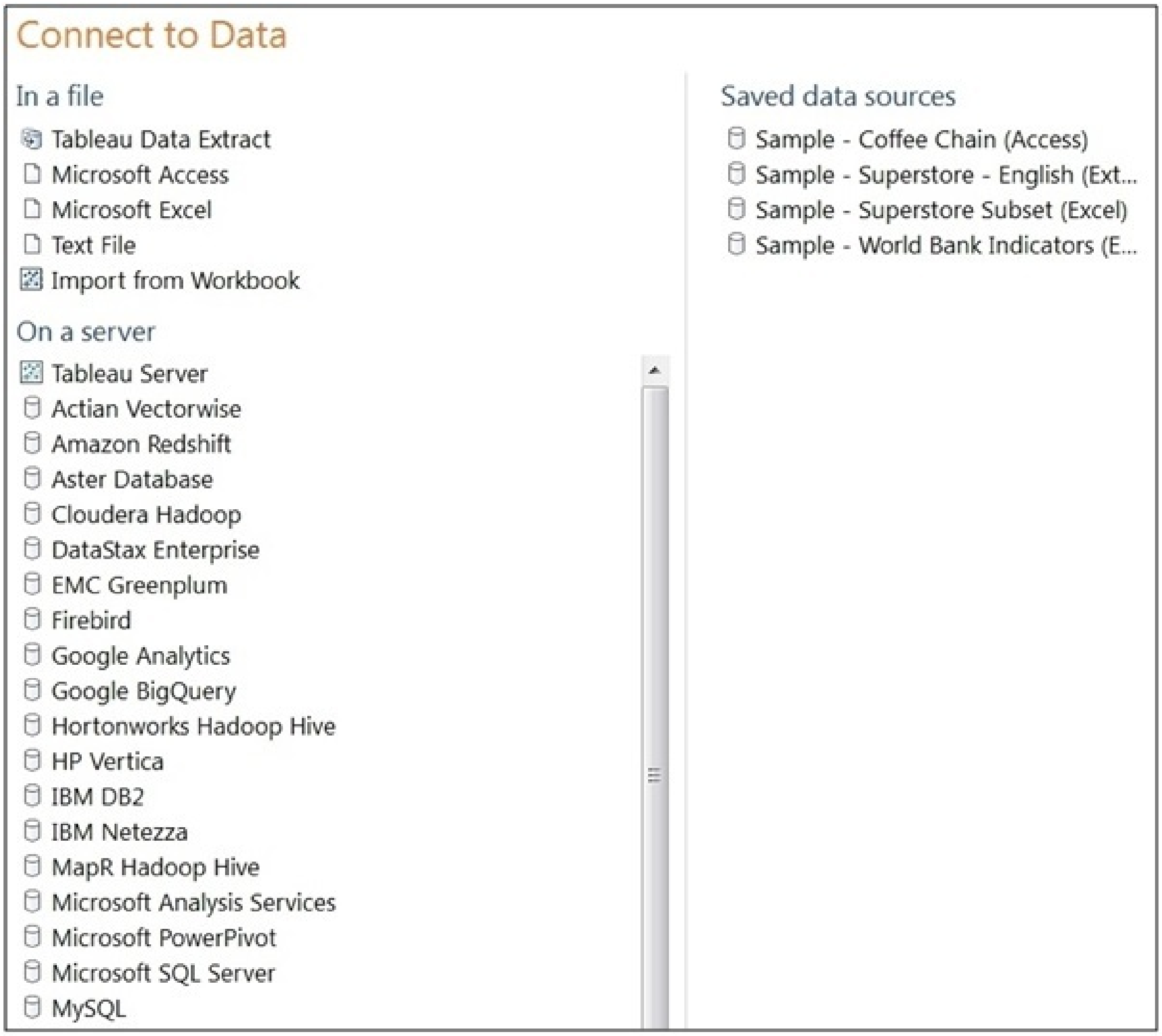


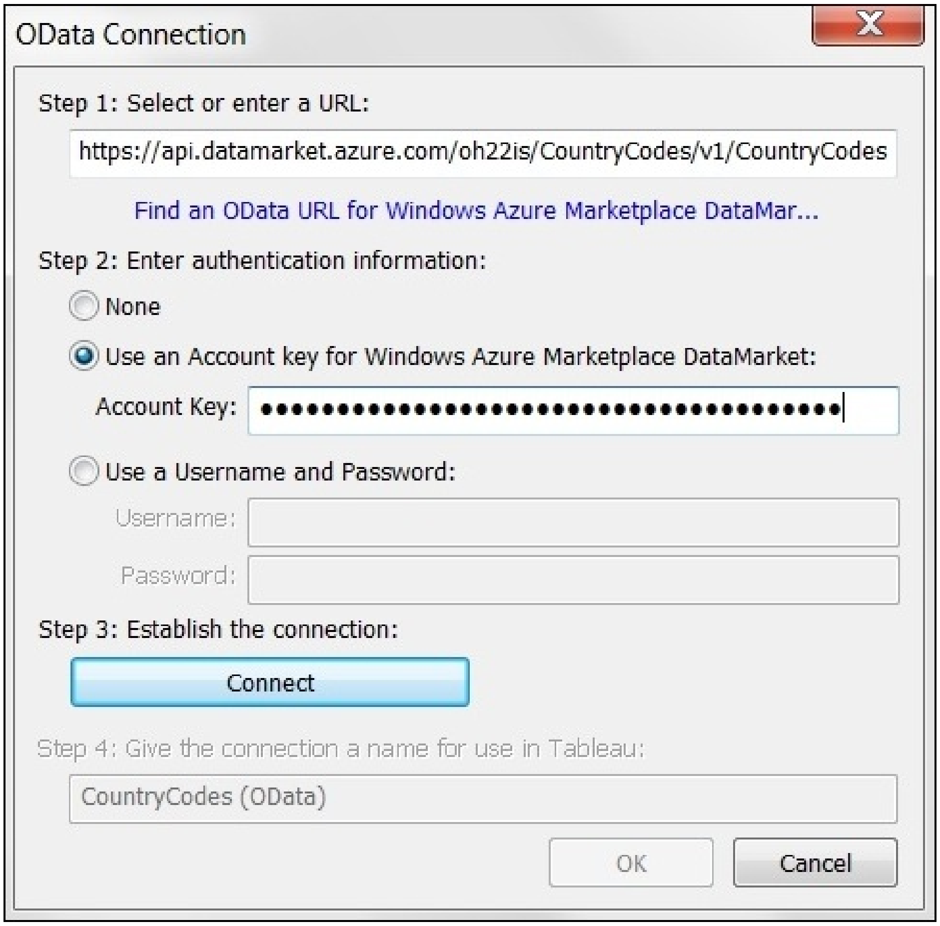
- In this example, we will load the data in Tableau rather than in the Data Explorer URL. To do this, we need the primary account key. In Windows Azure Datamarket, this is easy to obtain. From the previous example, we can see a feature called **Primary Account Key**. If you click on the **Show** link next to **Primary Account Key**, then your primary account key will appear.

- Copy the primary account key to your clipboard by selecting it and pressing the *CTRL + C* keys. You will need the primary account key to access the data using Tableau.
- You will also need to get the OData feed for the `Country Codes` data of the Windows Azure Datamarket `Country Codes` store. To get the OData feed, you can see it under the sentence **URL for current expressed query**, and you should copy this information.
- Before you proceed, you should note the OData URL and the primary account key. Select them and press the *CTRL + C* keys simultaneously. The following table shows an example of how your data might look:

OData URL	https://api.datamarket.azure.com/oh22is/CountryCodes/v1/Coun
Primary account key	**Aaa0aaAa0aAa00AAaAAA0aaA0AaaOa0aAaeAaA1AAA**

- To connect to Windows Azure Datamarket, let's open up Tableau and open the `Chapter 1 Demo` workbook that we started in the *Getting ready* section of the *Showing the power of data visualization* recipe.
- Go to the **Data** menu item and choose **Connect to Data…**.
- This action takes you to the **Connect to Data** window, and you can see that there are a variety of data sources for you to choose from! A sample of the list can be seen in the next screenshot:



- In this example, we are interested in connecting to Windows Azure Datamarket. Here, we will use the information that we saved earlier in this section. You will need the **OData** connection link. The connection panel only needs a few items in order to connect to the `Country Codes` data in Windows Azure, and an example can be seen in the next screenshot: 

- Insert the OData URL into the textbox labeled **Step 1: Select or enter a URL**.
- Next, take a look at the step labeled **Step 2: Enter authentication information**, select the radio button next to the option **Use an Account key for Windows Azure Marketplace DataMarket**, and insert the account key into the textbox. Then, click on the **Connect** button.
- All being well, the data connection will be successful and we can save the Tableau workbook before proceeding to connect to the Excel data source.
- We will download the GNI data from the World Bank. The URL is as follows: http://data.worldbank.org/indicator/NY.GNP.PCAP.CD?page=1
- To do this, open an Internet browser and navigate to the URL. You can see the web page in the following screenshot:

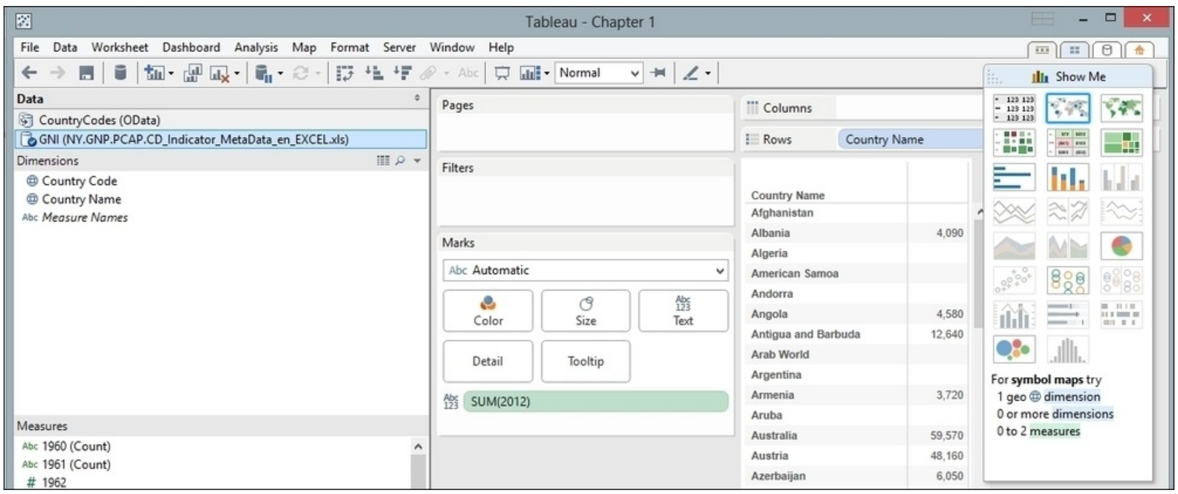
- You will see a button called **DOWNLOAD DATA**, which is on the right-hand side.
- Click on this button and you will be presented with two options: **EXCEL** and **XML**. We will download all of the data in Excel format.
- Before accessing the data source, let's save the file into the directory that you created earlier.
- Once the file is saved, open it in Excel and take a look. If you don't see any data, don't be alarmed.

You will see that there are three sheets and the workbook may open on Sheet 2. This will only provide metadata about the data held in the worksheet, and we need to look at Sheet 1. Then, we'll perform the following steps:

1. Let's rename Sheet 1 to something more meaningful. Right-click on the sheet tab name and rename it as `GNI`.
2. Remove the first two rows of the file. They will only add noise to the import.
3. Once you've done this, save the workbook. Now, you can exit Excel. We will go back to Tableau to connect to the data.
4. To connect to the Excel file, go to the **Data** menu item. Select **Connect to Data…** and a browser will appear.
5. Navigate to the location where the files are stored.
6. Select the worksheet to analyze and double-check whether you are looking at the correct file.
7. We are given the option of selecting a single file, multiple files, or a custom SQL.
8. At step 3, labelled **Does the data include field names in the first**

row?, select the option **No**.

9. At step 4, labelled **Give the connection a name for use in Tableau**, type in `GNI` and click on **OK**.
10. We can now decide whether to import the data or to connect *live*. We will connect *live* to the Excel workbook.
11. Now, we can see the Tableau workbook in the following screenshot. In the **Data** view at the top, we can see two connections: our Windows Azure Datamarket connection and our Excel file connection.



12. If we want to flip between each data source, we can click on each connection and see that the dimensions and measures change in response.

How it works…

Tableau connects to each data source and talks to it using drivers that are specific to each datatype. For example, Tableau has some connectors to popular programs, such as R, Google Analytics, and Salesforce.

You can find more information about drivers on the Tableau website at the following link: http://www.tableausoftware.com/support/drivers

There's more…

Tableau will connect to each data source independently. Even though they are different types of data sources, they appear to look the same in Tableau. From the user perspective, this is very useful since they should not be distracted by the differences in the underlying data source technologies. This means that the user can focus on the data rather than trying to put the data into one data source. Further more, it means that the sources of data can be refreshed easily because the Tableau visualization designer is able to connect directly to the source, which means that the data visualization will always be up to date.

See also

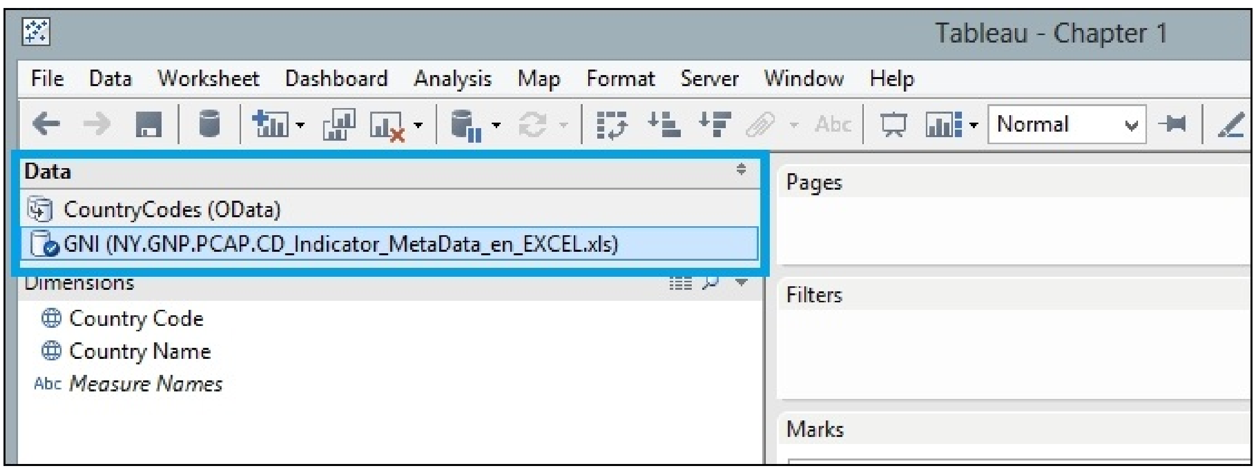
- Tableau can import data into its own in-memory engine. We will look at this in a later section.

Introducing the Tableau interface

In this recipe, we will look at the components of the Tableau interface and use these features in order to create a simple Tableau visualization. In the previous recipe, we connected to data in Windows Azure Datamarket and a local Excel spreadsheet. We will use these data sources in our example here in order to produce a quick and easy data visualization.

Getting ready

Make sure that you have a copy of the `Chapter 1` Tableau data visualization open. You should be able to access both data sources. To do this, click on the Tableau **Data** connection that you will see in the top left-hand corner of the Tableau interface, as shown in the following screenshot:

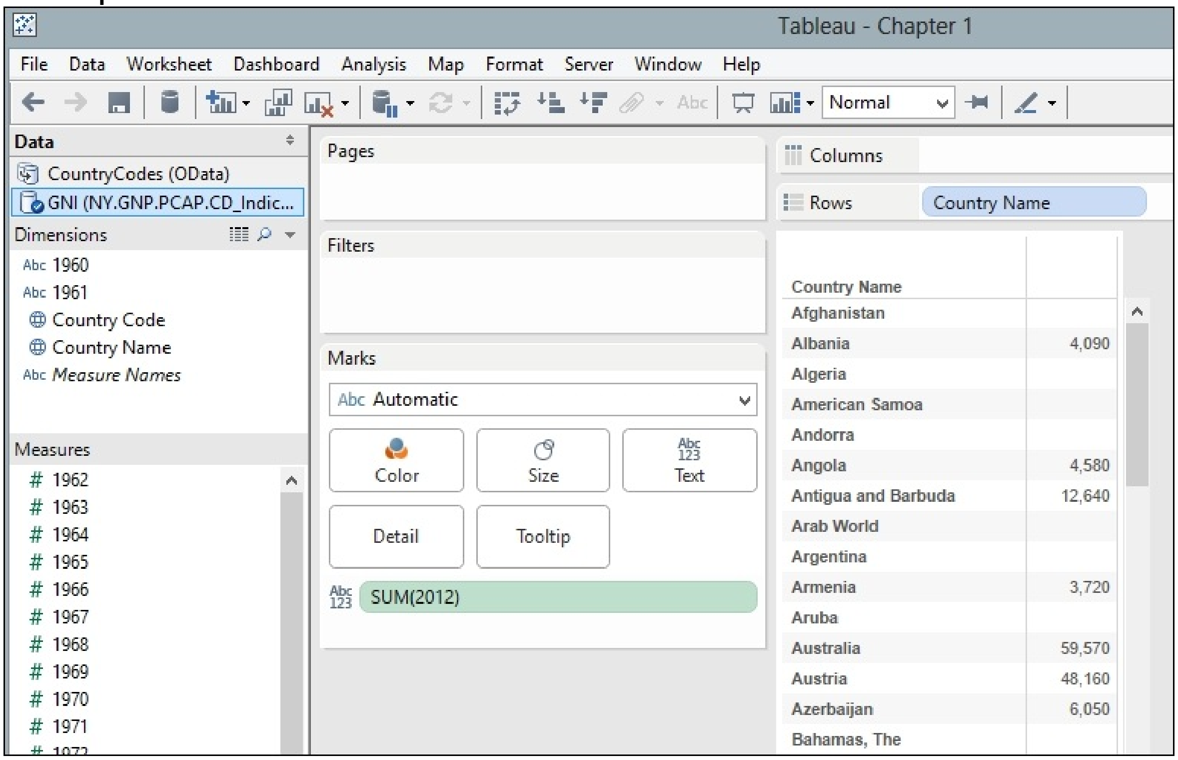


You should be able to click on the **CountryCodes** and the **GNI** connections alternately, and see the differences in the dimensions and metrics contained in the two data sources.

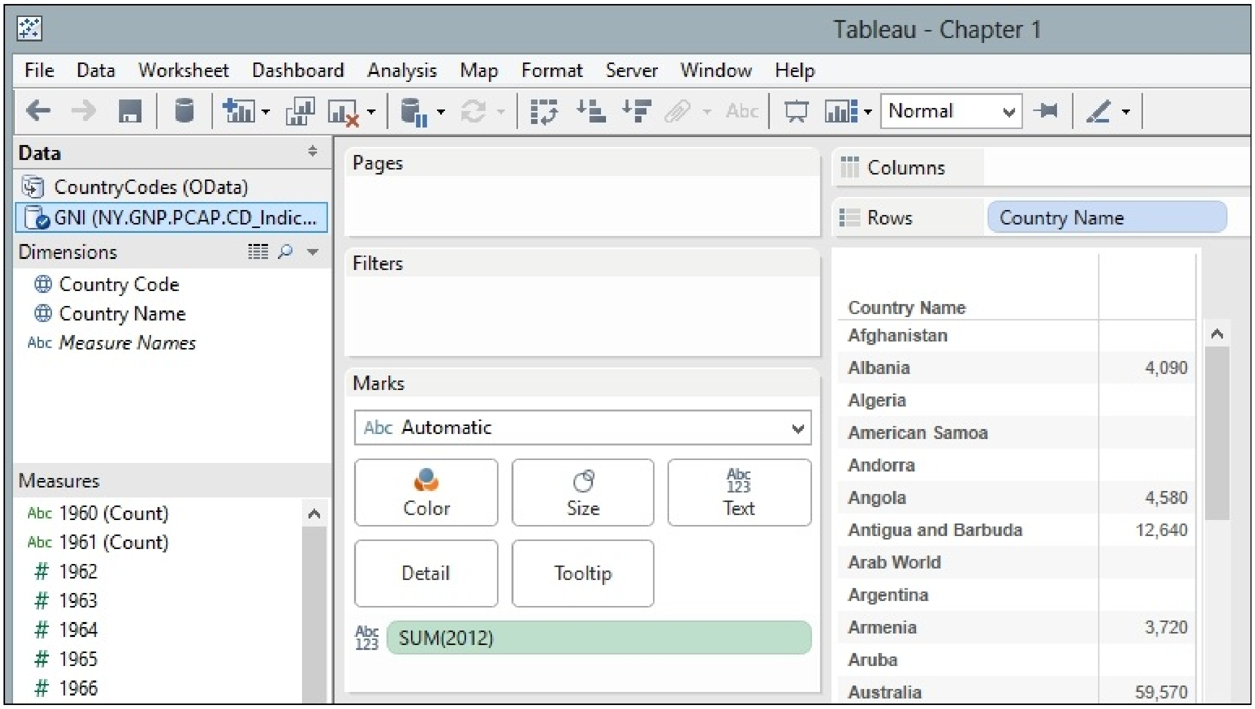
How to do it…

1. In the `Chapter 1` Tableau data visualization, click on the GNI data source. This will change the dimensions and measures, which you can see in the left-hand side column of the Tableau interface. An

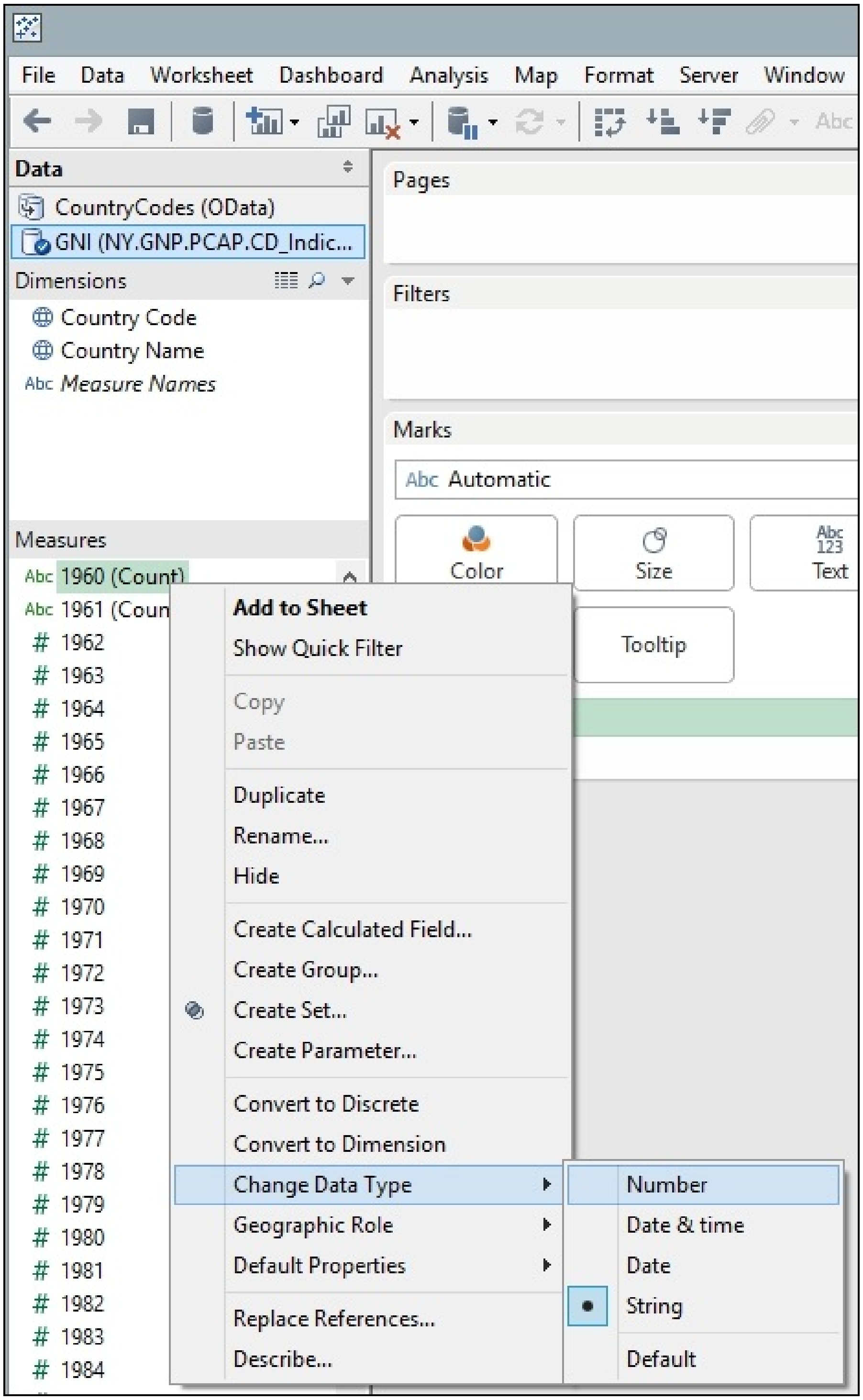
example is shown in the next screenshot:



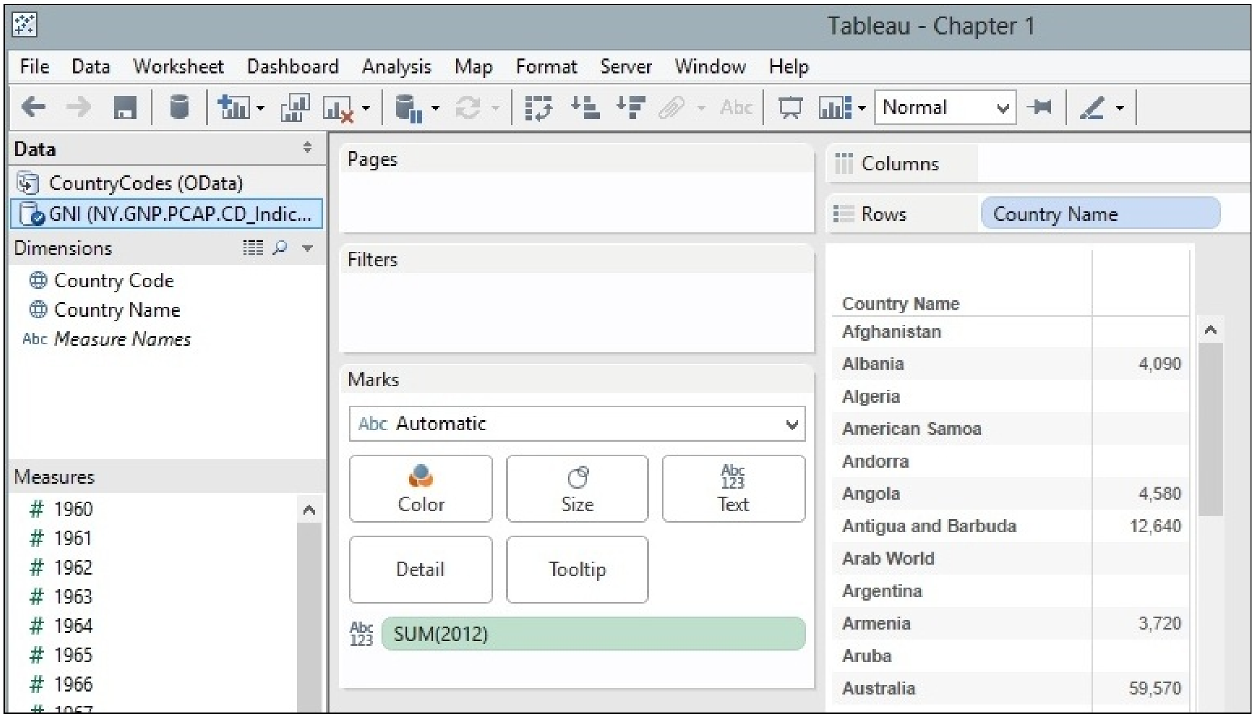
- You might notice that some of the dimensions are called `1960` and `1961`, but the rest are considered to be metrics.
- Fortunately, this is very easy to change. You can simply drag the `1960` and `1961` dimensions down to the **Measures** area. The Tableau interface now looks like the following screenshot:



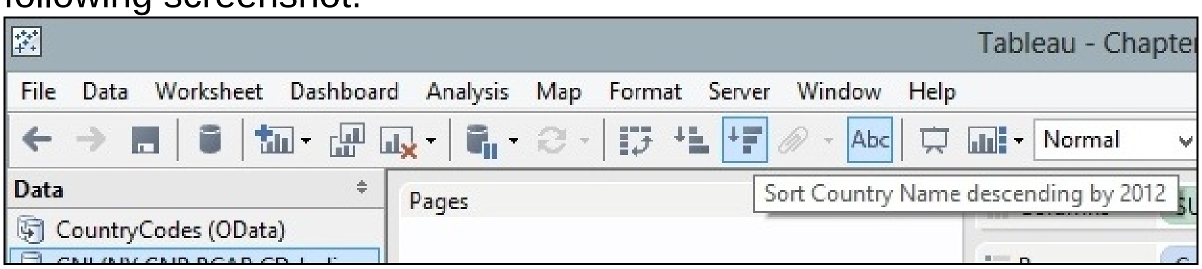
- Now that we see the measures, you can see that they are still specified as a string datatype and they are specified as a **Count**.
- Fortunately, this is also very easy to change. If you right-click on the measure **1960**, a pop-up menu will appear. You can see an example of the pop-up menu in the next screenshot:



- If you do this for both `1960` and `1961`, you can change both the datatypes to number. The result can be seen in the next screenshot:



- Now that the data has been prepared, let's move to visualizing the data.
- Earlier, we were introduced to the **Show Me** panel. Before we use the **Show Me** panel, however, we need to put some data on the shelves. This is a location where we drag-and-drop the dimensions and metrics in order to make them part of the data visualization.
- Pick the dimension **Country Name** and drag it onto the **Rows** shelf.
- Pick the metric `2012` and place it on the **Columns** shelf.
- You can now see that the data visualization has changed from a table to a horizontal bar chart. We can make it look better by sorting the bars in descending order. This allows us to quickly identify the highest GNI amounts for the top *n* countries.
- To sort in descending order, look for the button that shows a downward arrow next to a horizontal bar chart. When you wave the mouse over it, you will see that it sorts by the metric. An example is shown in the following screenshot:



Once you've sorted the data, it will look neater and easier to understand. We can see this in the following screenshot:

How it works…

One of Tableau's features is that it works out automatically whether the data is a dimension or a measure. Tableau does this by looking at the datatype in the columns. So, for example, in this case, it has identified text and geographical types as dimensions and integers as measures.

You may be wondering why we have data that has a year for each column rather than a column **Year**. This is a good question to ask, and we will look at different ways of shaping the data and how that affects the resulting visualization throughout the course of this book.

Tableau has an internal knowledge base that it uses in order to determine the most appropriate visualization for the data that it sees. Initially, in this case, it has suggested a horizontal bar chart in blue. Why is this the case?

We have a horizontal bar chart rather than vertical because we can read more easily along rather than up and down. For people in the West, we tend to read left to right, so we see the country name on the left followed by the bar and the value on the right.

By having horizontal bars, it is easy to see how the bars compare within the chart itself. We have the visual information from the bar itself as well as the metrics labelled at the end of the bar.

See also

- A book list will be provided at the end of the book for people who are interested in research on data visualization

Interacting with your first data visualization

In this recipe, we will learn about interacting with your first visualization and look at different visualizations that are available to you in Tableau. The **Show Me** panel provides you with a range of options to create data visualizations. Some of these can be adapted so that they pack a lot of information into a very small space, which is ideal for dashboarding. In this recipe, we will look at creating a bullet chart, which has been designed to retain a balance between packing the maximum amount of information into the minimum amount of space while also retaining clarity.

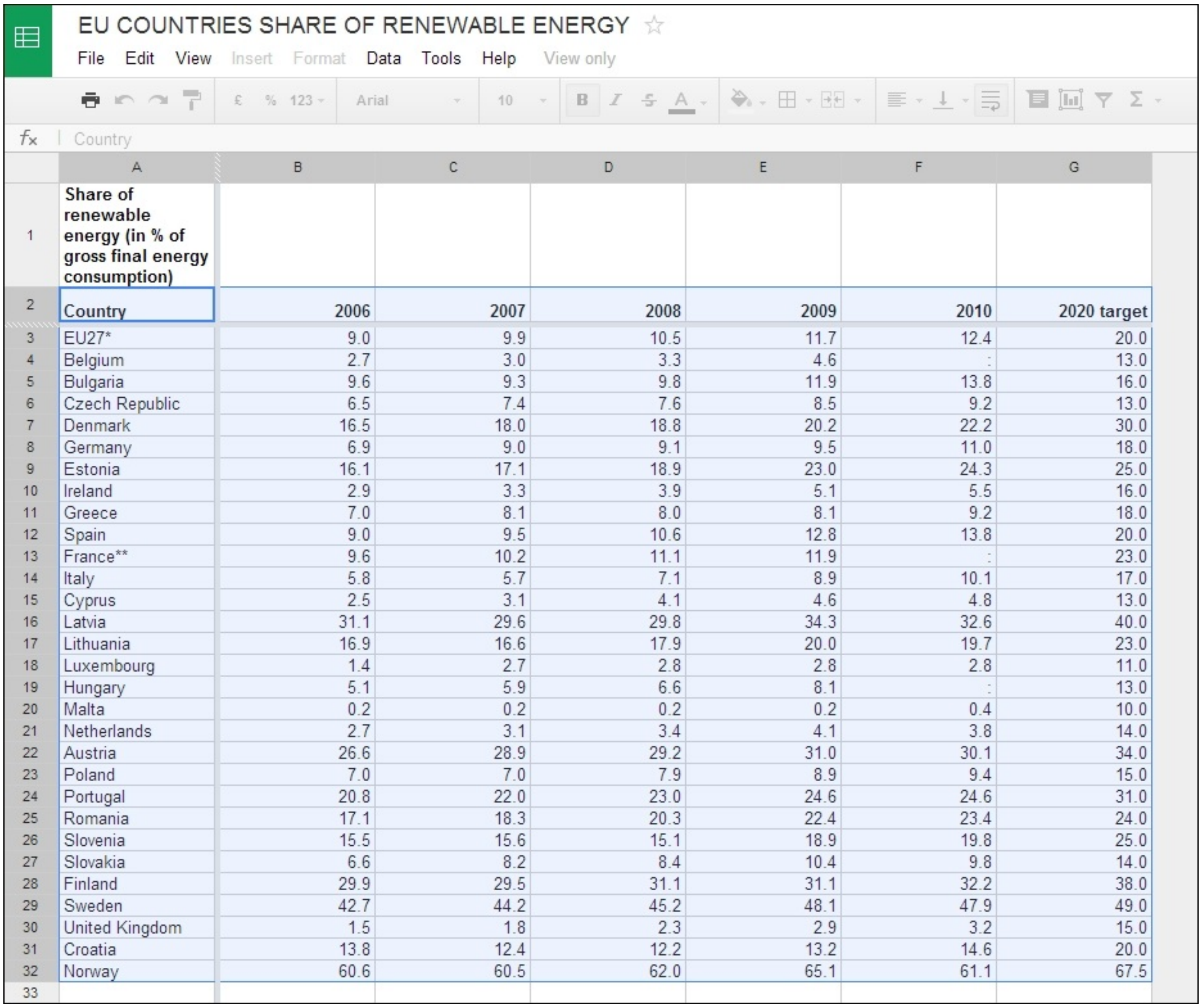
The bullet chart was devised by a data visualization expert and thought leader, Stephen Few. It is designed to replace charts and graphs that show a lot of ink or take up a lot of space on the page but do not show a lot of data. The bullet graph is effective because it takes up little space and allows the viewer to see whether the actual data is comparable to the target by reading from left to right along the bar. Playing with the colors on the bullet chart is a useful way to understand this useful chart better.

We are using a very simple dataset as a starting point, and we will move towards more complexity in terms of data and visualizations for dashboarding as we proceed throughout the book.

Getting ready

Before we open Tableau, let's download the data from a Google Docs spreadsheet provided by the Guardian Datastore, which is provided by *The Guardian* newspaper that is published in the UK. You can visit the following link: http://bit.ly/TableauCh1TargetData.

You may need a Google account to open the spreadsheet. Once you have opened the spreadsheet, you copy the data that you see highlighted in the following screenshot:



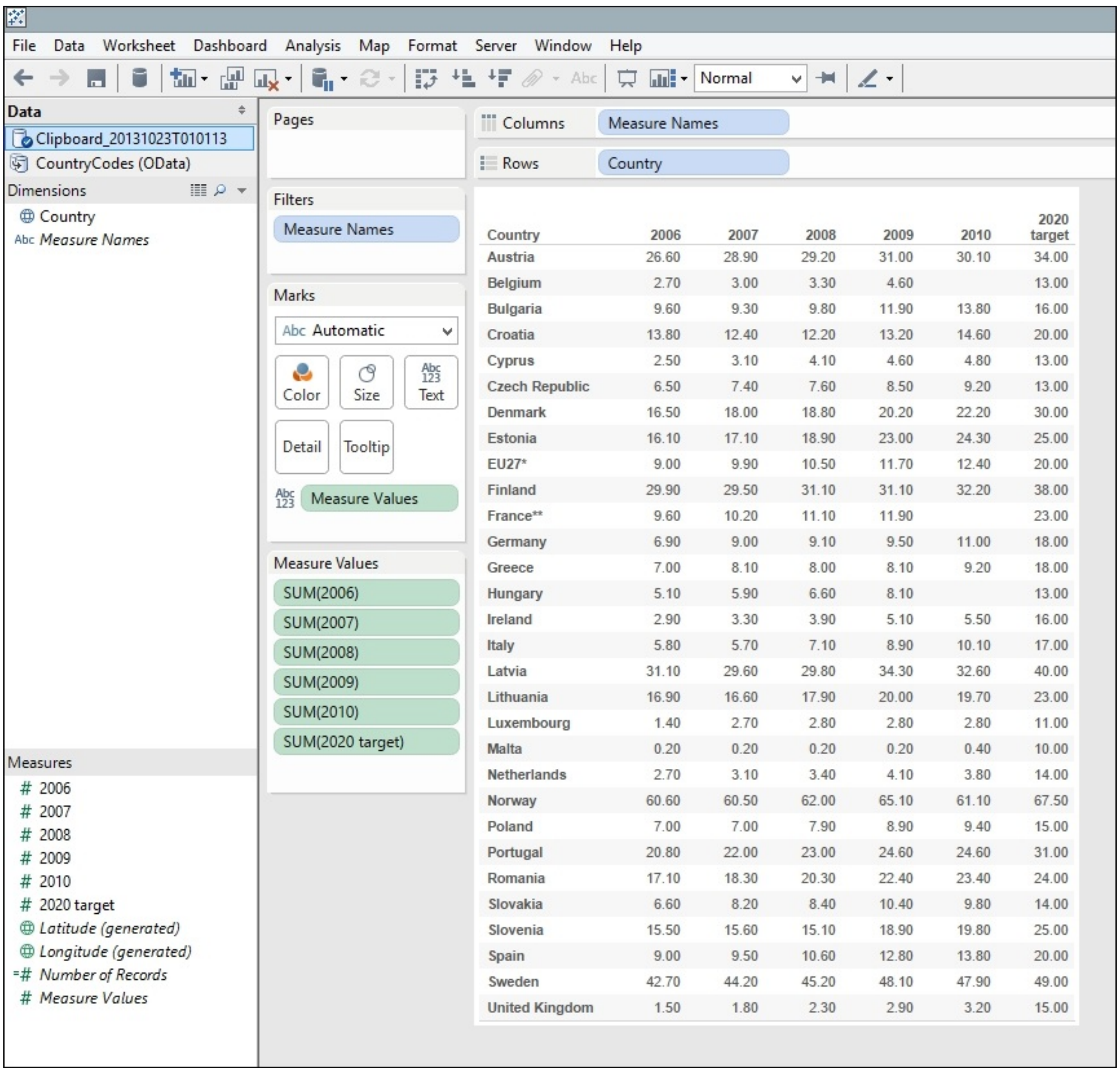
Share of renewable energy (in % of gross final energy consumption)						
Country	2006	2007	2008	2009	2010	2020 target
EU27*	9.0	9.9	10.5	11.7	12.4	20.0
Belgium	2.7	3.0	3.3	4.6	:	13.0
Bulgaria	9.6	9.3	9.8	11.9	13.8	16.0
Czech Republic	6.5	7.4	7.6	8.5	9.2	13.0
Denmark	16.5	18.0	18.8	20.2	22.2	30.0
Germany	6.9	9.0	9.1	9.5	11.0	18.0
Estonia	16.1	17.1	18.9	23.0	24.3	25.0
Ireland	2.9	3.3	3.9	5.1	5.5	16.0
Greece	7.0	8.1	8.0	8.1	9.2	18.0
Spain	9.0	9.5	10.6	12.8	13.8	20.0
France**	9.6	10.2	11.1	11.9	:	23.0
Italy	5.8	5.7	7.1	8.9	10.1	17.0
Cyprus	2.5	3.1	4.1	4.6	4.8	13.0
Latvia	31.1	29.6	29.8	34.3	32.6	40.0
Lithuania	16.9	16.6	17.9	20.0	19.7	23.0
Luxembourg	1.4	2.7	2.8	2.8	2.8	11.0
Hungary	5.1	5.9	6.6	8.1	:	13.0
Malta	0.2	0.2	0.2	0.2	0.4	10.0
Netherlands	2.7	3.1	3.4	4.1	3.8	14.0
Austria	26.6	28.9	29.2	31.0	30.1	34.0
Poland	7.0	7.0	7.9	8.9	9.4	15.0
Portugal	20.8	22.0	23.0	24.6	24.6	31.0
Romania	17.1	18.3	20.3	22.4	23.4	24.0
Slovenia	15.5	15.6	15.1	18.9	19.8	25.0
Slovakia	6.6	8.2	8.4	10.4	9.8	14.0
Finland	29.9	29.5	31.1	31.1	32.2	38.0
Sweden	42.7	44.2	45.2	48.1	47.9	49.0
United Kingdom	1.5	1.8	2.3	2.9	3.2	15.0
Croatia	13.8	12.4	12.2	13.2	14.6	20.0
Norway	60.6	60.5	62.0	65.1	61.1	67.5

Select the table of data as in the preceding screenshot, copy it using *Ctrl + C*, and then paste it into Tableau. This will import the copied data into the model contained in the Tableau worksheet. Alternatively, you could download the Google spreadsheet as an Excel spreadsheet by navigating to **File | Download as | Microsoft Excel (.xlsx)**. Since we will be changing the original visualization in the `Chapter 1` workbook, it is good practice to take a copy of your current visualization and work on the copy. When you work in Tableau, it is very easy to keep clicking around and changing visualizations. However, if you want to roll back to an earlier point, you might find that you've easily clicked away quite far from your preferred point.

In this example, we will work on a copy of the `Chapter 1` workbook so we can compare our progress from start to finish quite easily. We will use data from the Guardian Datastore which shows whether countries are on target to meet their environmental targets according to the Kyoto agreement. This is a good preliminary example of dashboard data, because we are displaying the actual versus target data, and this is a common dashboarding scenario.

How to do it...

1. Once the data is copied into Tableau, the workbook will appear as follows:



- If the years appear as dimensions, then drag them to the **Measures** pane on the left-hand side.

- Our starting point is a table. In our duplicate sheet, go to the **Show Me** panel at the right-hand side. Select the **horizontal bars** option. You can see a sample of the **Show Me** panel in the next screenshot:

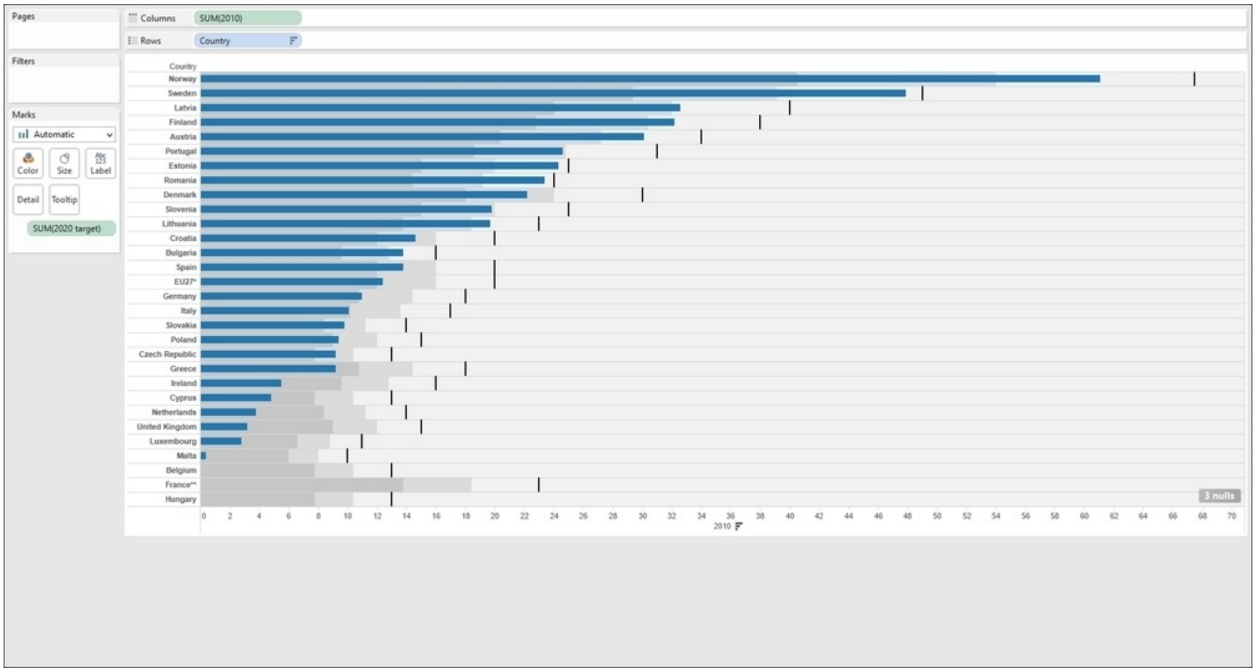
- Once you have selected the **horizontal bars** option, your screen will look like the following screenshot:

- We are interested in the target data. To show the scenario of comparing actual data with target data, remove all of the green pills from the **Columns** shelf, except **SUM(2010)** and **SUM(2020 Target)**.

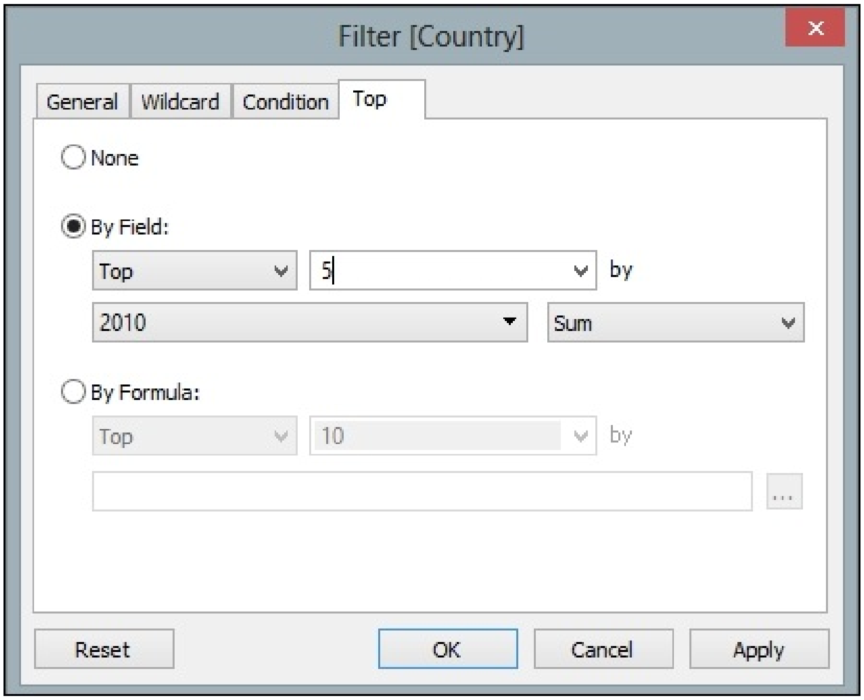
- Once these columns have been removed, the **Show Me** panel will show more options. We will choose the **bullet graphs** option, which is highlighted with a blue box in the following screenshot:

- Once the **bullet graphs** option has been clicked on, look for the small icon that looks like a horizontal bar chart on the taskbar. You will find it below the menu items. When you wave the mouse over it, you will see that it is a tooltip that says **Sort Country Descending by 2012**. It is circled in the following screenshot: 

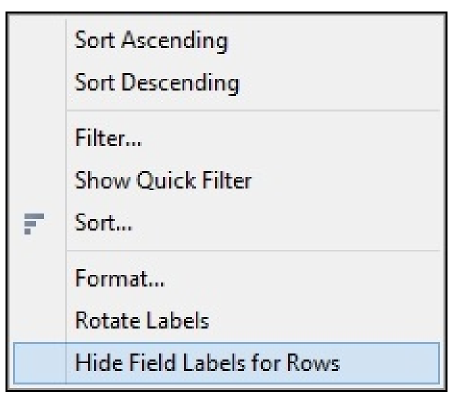
- The result can be seen in the next screenshot, which shows rows of bullet charts:

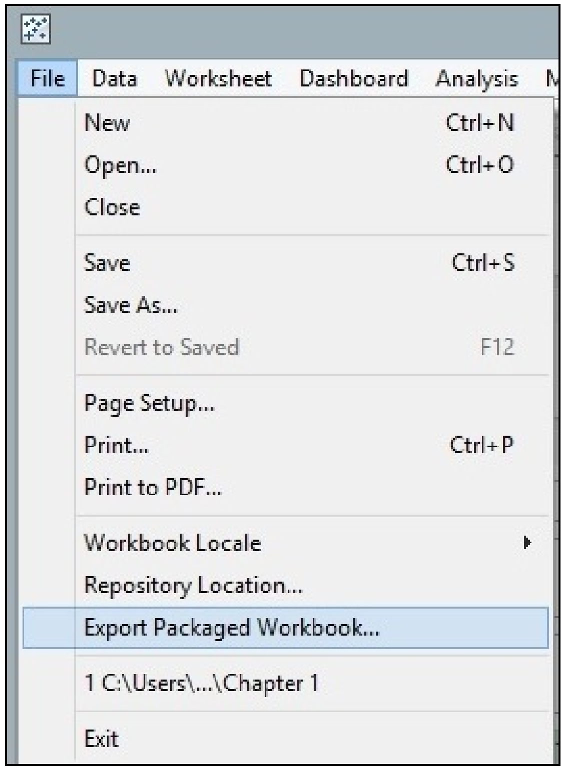


- This is still a lot of data to show on a dashboard and still be sure that the data consumer is able to remember and understand it quickly. The idea is that the thick horizontal line displays the actual data and the vertical line on each row displays the target. We can resize it so that the rows are smaller in height. To do this, you can resize by grabbing the bottom of the white canvas and pulling it upwards. This will make the data visualization smaller.
- We could filter this further in order to show the top five countries who have the greatest share of renewable energy sources in **2010**. To do this, drag the **Country** dimension from the left-hand side of the Tableau workbook to the **Filter** panel located just above the **Marks** panel. The following wizard will appear:



- Select the **Top** tab and select the **By Field** radio button.
- Then, put the number 5 into the textbox and select the **2010** column from the drop-down list.
- Click on **OK** to clear the **Filter** wizard.
- Then, right-click on **Country** in the visualization and select the option **Hide Field Labels for Rows**, as shown in the following screenshot. This will remove unnecessary ink from the screen, which means that there are fewer unnecessary items to distract the viewer.



- Once this is done, resize the visualization so that it is only a few inches in length. To do this, go to the right-hand side of the visualization and drag the end along to the desired size. The data visualization now looks like the following screenshot: 

How it works…

Copying and pasting the data into Tableau is a great way of importing data quickly. Note, however, that this data is static and will not change with any changes in the data source.

There's more…

Removing unnecessary ink from the screen is a useful way of cutting down the items displayed on the dashboard. In this example, the label was redundant and its removal made the graphic neater.

If you require more information on the bullet chart, please visit the following link: http://bit.ly/BulletGraphbyStephenFew

Sharing your visualization with the world

In the first recipe, we specified communication as one of the key features of a dashboard. We need to be able to share the information to the right audience at the right time, to the right people in the right format.

Tableau offers a number of different ways to share the dashboard in order to help team members throughout the organization to track, monitor, and analyze the metrics about their organization, and we will look at these in the current section.

Given that Tableau offers a number of ways to share a dashboard, what is the best way to do this? The best way to decide which method to use to share your information fundamentally rests on the user requirements. These are listed in the following table:

Objective	**Method**
For other Tableau users who don't have access to the data	Exporting a Tableau packaged workbook
To view data online and share the data	Sharing your workbook with Tableau Public
For Tableau users who do have access to the data	Sharing your workbook with Tableau Server

In this recipe, we will look at the first two methods of sharing data: exporting a Tableau packaged workbook and sharing your workbook with exporting a Tableau packaged workbook and sharing your workbook with Tableau Public. When we export a workbook as a packaged workbook, it wraps up the data as part of the Tableau workbook. Why would you want to do this? The following are some reasons:

- You may want to send the workbook to someone who does not have access to the data source
- You may be prototyping a workbook with some sample data
- You may find it quicker to develop offline

When we save a file as a packaged workbook, the workbook points at its own internal copy of the data via the data source connection. If it is a packaged workbook with a data extract, then it no longer references the data from the original data source. Instead, all of the references point to the workbook's internal version of the data via the data source connection, not the original source. Logo images, for example, that are part of the dashboard, are stored as part of the packaged workbook rather than externally referenced.

The workbook is now insulated from changes in the data source, and it won't be impacted by changes in the data source. Individuals who do not have access to the original data source can still see the workbook and manipulate the data, but cannot impact the data source in any way.

If you want to save a workbook to Tableau Public, then you must use a workbook that has a packaged data source. There are a number of criteria which must be met in order to publish the dashboard to Tableau Public. The data extract may not include more than 1 million rows. Only workbooks with a data extract will be published to Tableau Public. Finally, if the workbook has multiple data connections, then you will need an extract for each data connection.

We will look at this issue first, and then we will look at uploading this workbook to Tableau Public.

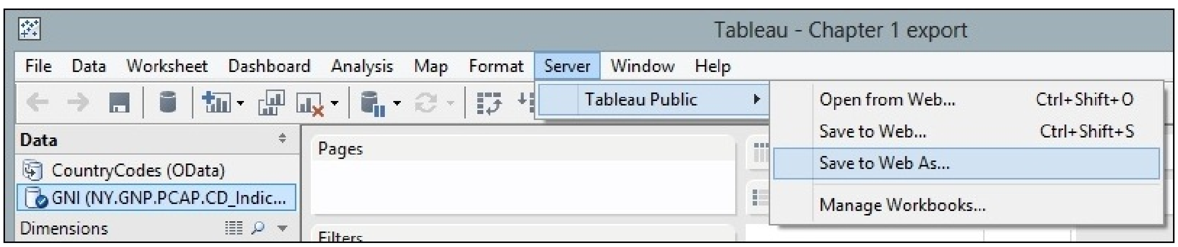
Getting ready

Check that your workbook has less than 1 million rows. In this example, it does. So, we can proceed. However, for your own work, you may find that this is not always the case.

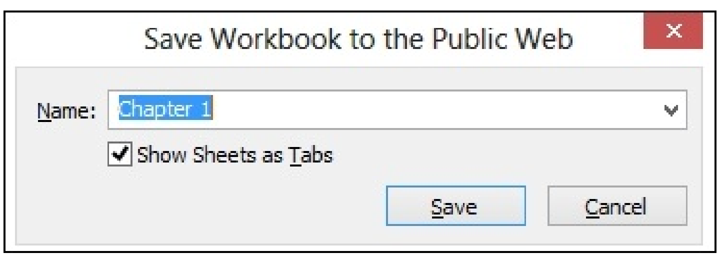
Check that you have a login for Tableau Public. If not, please visit the Tableau website in order to set up a login and a password (www.tableausoftware.com).

How to do it…

1. To save a workbook in order to upload it to Tableau Public, you need to save it as a packaged workbook. To do this, go to the **File** menu item and choose the option **Save As…**.
2. Enter the filename in the **File Name** textbox.
3. Go to the **Save As** option from the drop-down list, choose **Tableau Packaged Workbook**, and then click on **Save**.
4. Now, go to the **Server** menu item and you will see one option called **Tableau Public**. From here, you can get to a small menu which is called **Save to Web As…**. You can see an example of this in the following screenshot:



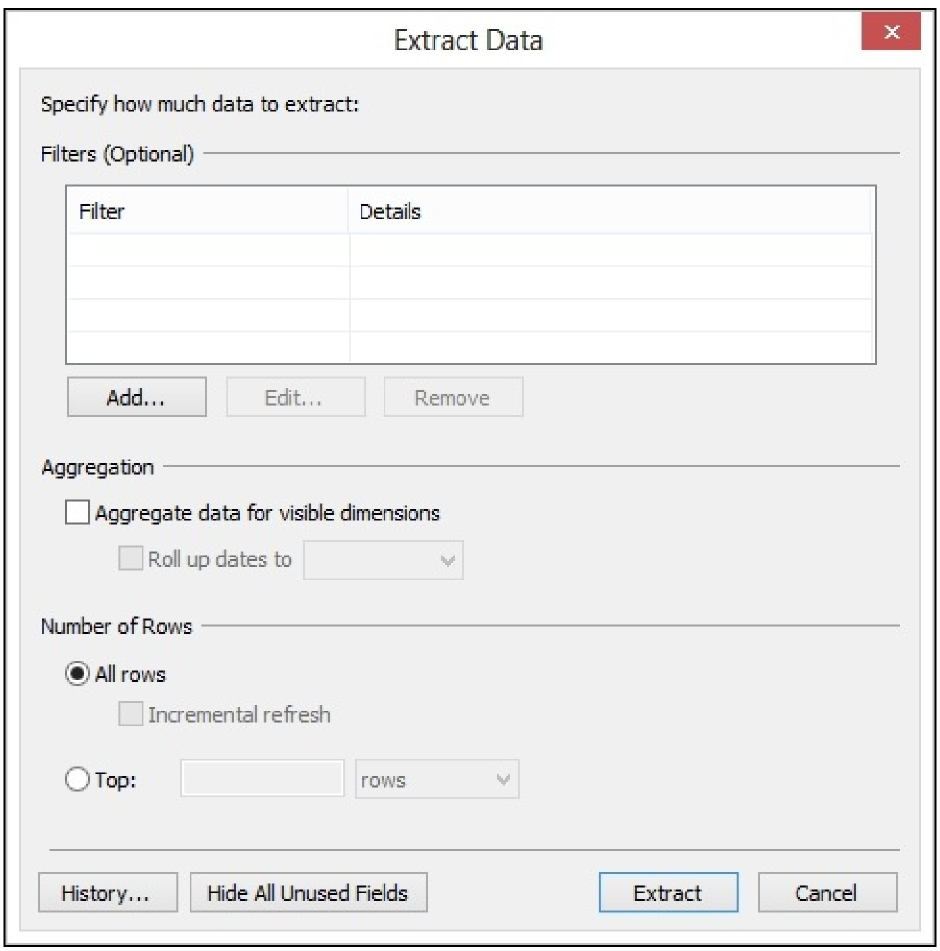
5. You will get the following dialog box:



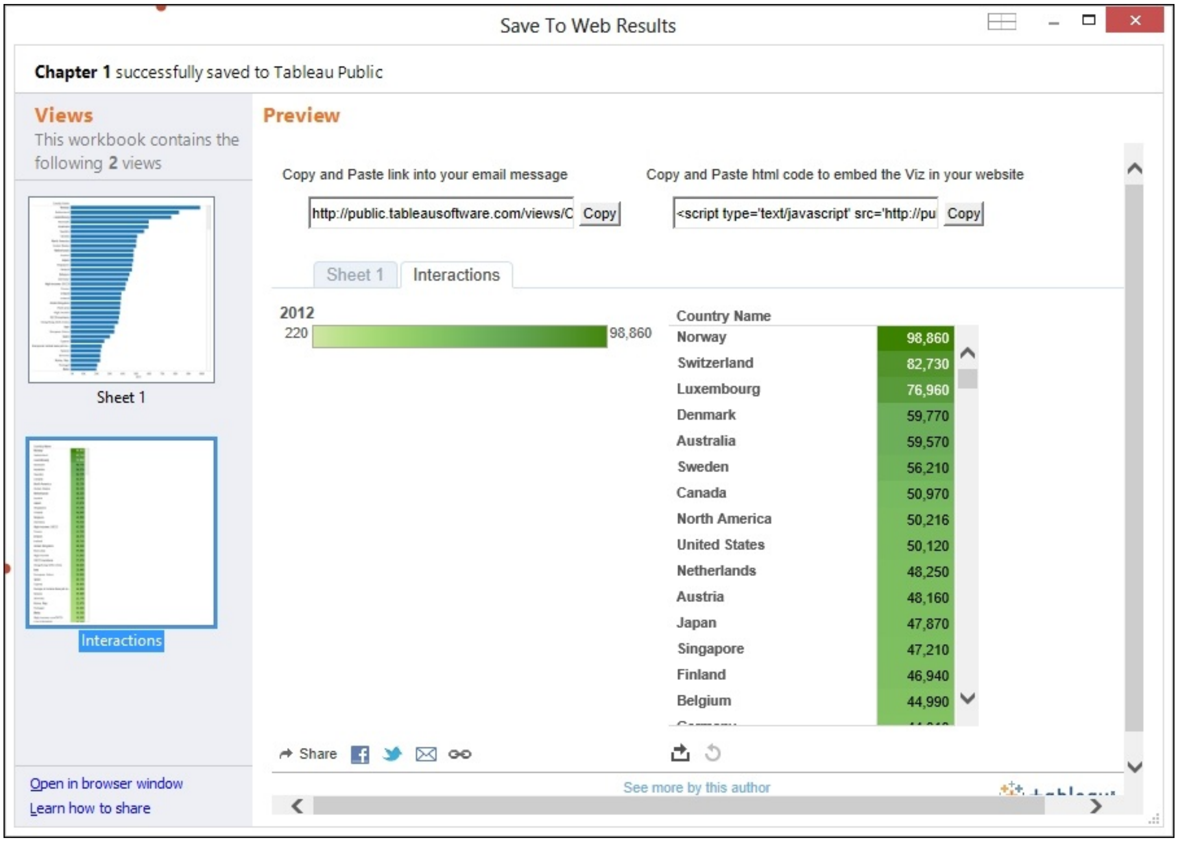
6. You will get a message asking you to log in to Tableau Public with your login ID and password. When you have entered these details, click on **OK**.

7. Next, you will get the message shown in the following screenshot:

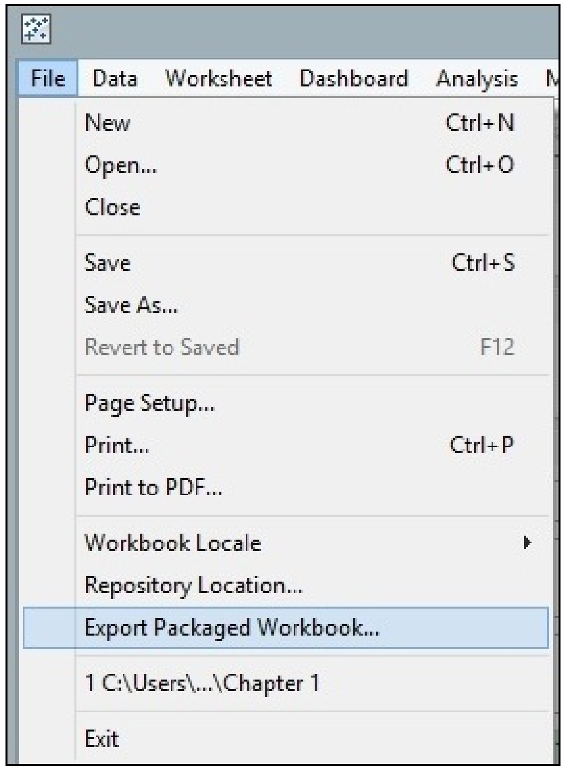
8. Select the link **Create Data Extract**.
9. You will now get a filter box. We wish to extract all of the data, so click on **Extract**. You can see an example in the next screenshot:



10. Now, you will see your results in an Internet browser, as shown in the following screenshot:



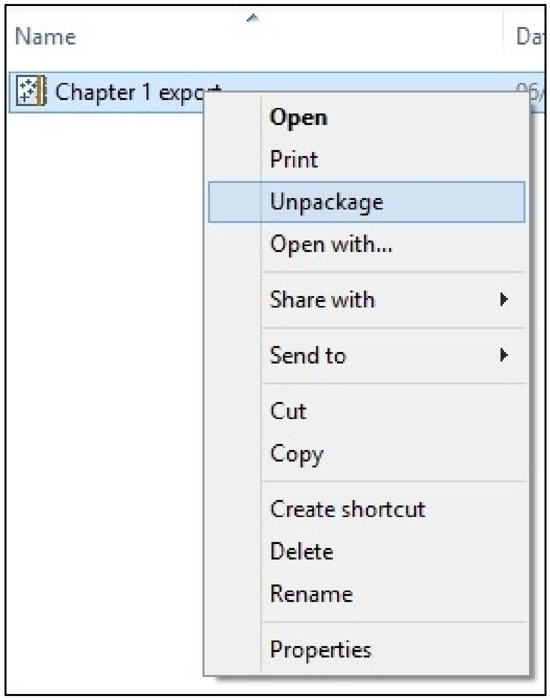
11. If you want to share your visualization, you can use the links to share your work, as shown in the next screenshot:



Tip

Tableau-packaged workbooks have the file extension `*.twbx`. Tableau workbooks have the extension `*.twb`.

12. If you are using Microsoft Windows and Tableau Desktop, it is possible to unpack the file. This is not possible in Tableau Reader or in the Mac version of Tableau. Once the file is saved, it is possible to unpack the original data source by unpacking the original workbook. To do this, navigate to the file in Windows Explorer and right-click on it. In the resulting menu, you will see an option to **Unpackage** the workbook. You can see an example of this in the following screenshot:



Then, you will be asked where you would like to unpackage the workbook. Select a location on your computer to unpackage the workbook; for example, you could use the location that we created at the beginning of this chapter: `D:\Data\TableauCookbook\Chapter 1`. We will keep the filename as it is.

How it works…

You can publish your workbook to the whole world using Tableau Public. The data is saved to Tableau's data centers and you can access the workbook from anywhere in the world via the Internet.

Tableau allows you to publish easily from your desktop. However, there are a few restrictions on using Tableau Public. Also, be careful about sharing your work; once the Tableau workbook is published to Tableau Public, anybody can download the data.

There's more…

At the end of your first chapter, you've moved from learning about the Tableau interface to creating your own dashboards and even publishing them on the Internet. The next chapters of this book will help you to create dashboards that fulfill the requirement to answer the business questions and increase the productivity and interactivity of the dashboards. We will look at ways to enhance understanding and also tips and tricks on how to make the dashboards look compelling in the next chapter.

Chapter 2. Summarizing Your Data for Dashboards

In this chapter, we will cover:

- Arithmetic – the queen of mathematics!
- Dashboards and dates
- Grouping your data with calculations
- Correlations with calculations
- Using cross-tabs flexibly
- Simplifying your business rules with custom calculations

Introduction

It isn't enough just to make data look beautiful; we know that we can do that with Tableau. The data has to be accurate as well in terms of the business rules of an organization.

Calculations make things easier for the business user. The idea of "intelligent laziness" is often ascribed to Napoleon Bonaparte. The core idea is that people put effort where it is most needed rather than wasting time and resources. Calculations can help you in many ways, such as removing the implementation of repeatable calculations through automation. They also allow you to implement your business calculations so that they are consistently used in your dashboards.

Arithmetic – the queen of mathematics!

This recipe explains how to use simple descriptive statistics and arithmetic as the first step toward analyzing your data. We will also look at ways to import data into Tableau. When we import data, Tableau will create a **Tableau Data Extract** file behind the scenes. This is also known as a TDE file for short.

Descriptive statistics are a great starting point when analyzing data. They are very helpful in delivering an initial overview of the data to help you interpret it. We can glean information about the spread of the distribution of the data, measures of variability around the mean, and measures of deviation from the normal curve. Descriptive statistics have a variety of uses, for example, to help you identify outliers, which are unusual cases in the data that may warrant further investigation. Missing data points are important as well because missing data can mislead our analysis, and data visualization can help us to profile the data to check for potential instances of missing data.

How do we calculate descriptive statistics? Once the `FactInternetSales` data has been imported, we will calculate its mean, median, and mode. These are measures of central tendency that allow us to see the shape of the data. Many business questions are quite simple: what are my average sales? What are my total costs? How well can a dataset be summarized by one number?

We can look at the data in order to see how well it can be described by a single number; this is called the **measure of central tendency**. Often, when we talk about business questions, people listening to us would want to know the average of something. However, when we look at the average, things become more complex. The average may be skewed, for example, by an outlier. We need to know whether our average is a representative of a dataset.

The average, median, and mode tell us about the symmetry of our data in terms of its distribution. They give us an initial picture of the data, a simple summary. Further, knowing about the symmetry of the data can help us look at important factors such as probability, which may form part of your analysis.

As you might expect, there are many ways to perform descriptive statistics in Tableau. This recipe will show you how to perform simple and quick descriptive statistics that will help you begin analyzing your data; this will be useful in understanding whether the average is effective in describing the data overall.

In this recipe, we will import some data and look at using some descriptive statistics to describe our data. Firstly, we will calculate the average, which is the most well-known measure of central tendency. Then, we will look at the median and mode.

Getting ready

For this recipe and future recipe, you will need to download a mix of Excel and CSV files. Perform the following steps:

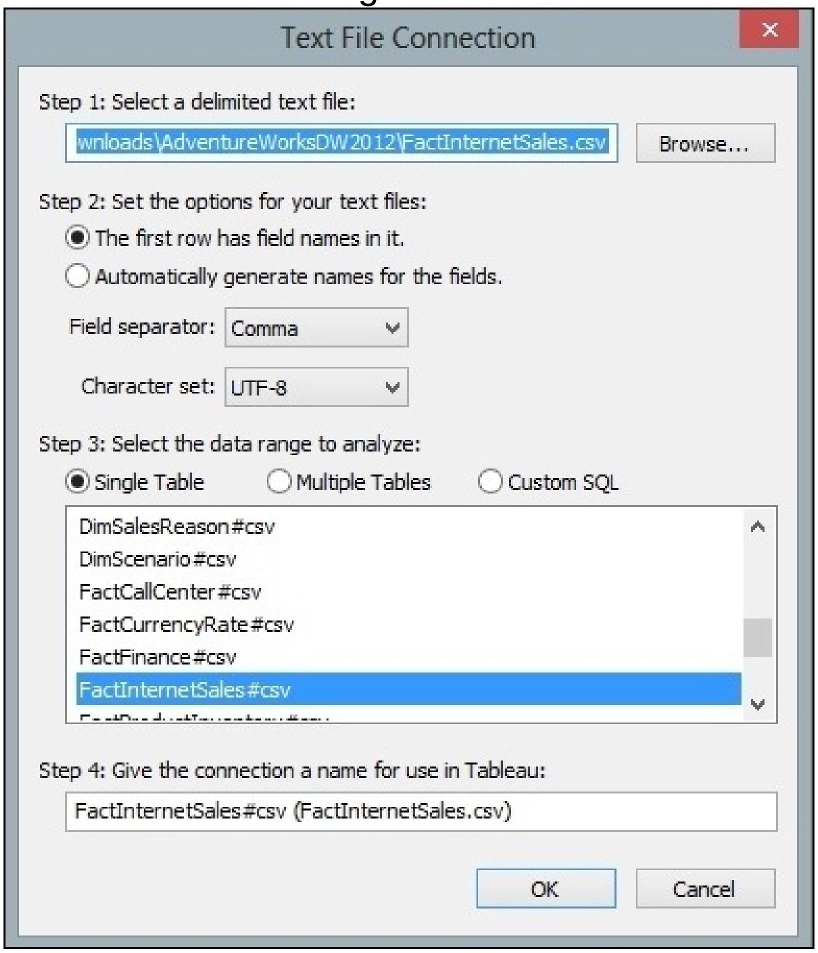
1. Set up a folder on your computer where you can store the files. As an example, you could call it `TableauCookbookData` and locate it on `D:`. The path for the folder would be `D:\TableauCookbookData`.
2. Go to http://bit.ly/TableauDashboardCookbookSampleData and download the ZIP file.
3. Right-click on the ZIP file and select **Extract To**.
4. Extract the files to your folder. So, in our example, you would extract the files to `D:\TableauCookbookData`.

How to do it...

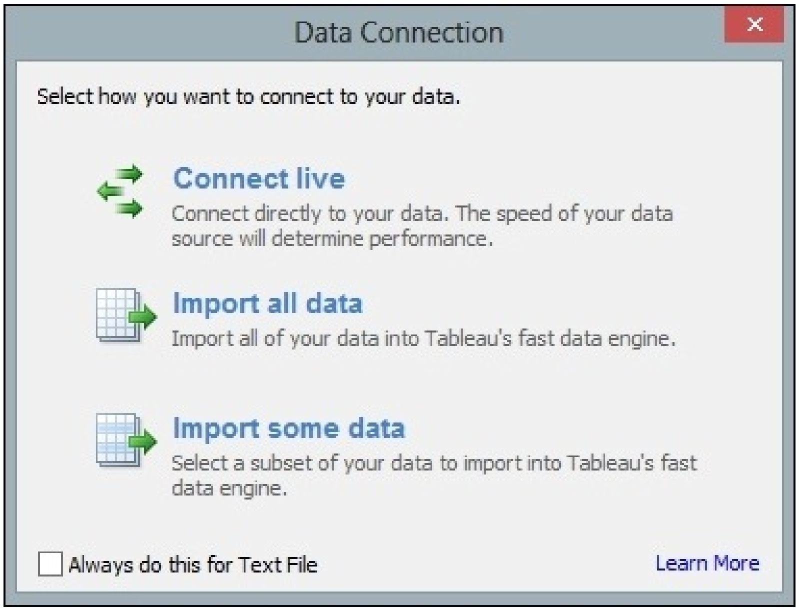
1. Open up Tableau and navigate to **File | New**.
2. Save the file as `Chapter 2`.
3. We will connect to the data and import it into Tableau's internal data store mechanism. To do this, click on **Connect to Data**.
4. Then, select the link **Text File** and a file browser will appear.

Navigate to the folder where you stored the CSV files.

5. Navigate to the FactInternetSales.csv file and select it to open it.
6. Tableau will ask you to save the connection and give it a name. You can see this in the dialog box named **Text File Connection**, as shown in the following screenshot:



- If you look at the name in **Step 4** of the **Text File Connection** dialog box, you will see that it is not very user-friendly. It reads **FactInternetSales#csv (FactInternetSales.csv)**. Let's rename the connection to Connection_FactInternetSales and then click on **OK**.
- Tableau will then ask you how you would like to connect to the data, or whether you'd like to import some or all of the data. We will import all of the data as per the **Data Connection** dialog box that is shown in the following screenshot. Select the **Import All Data** option.

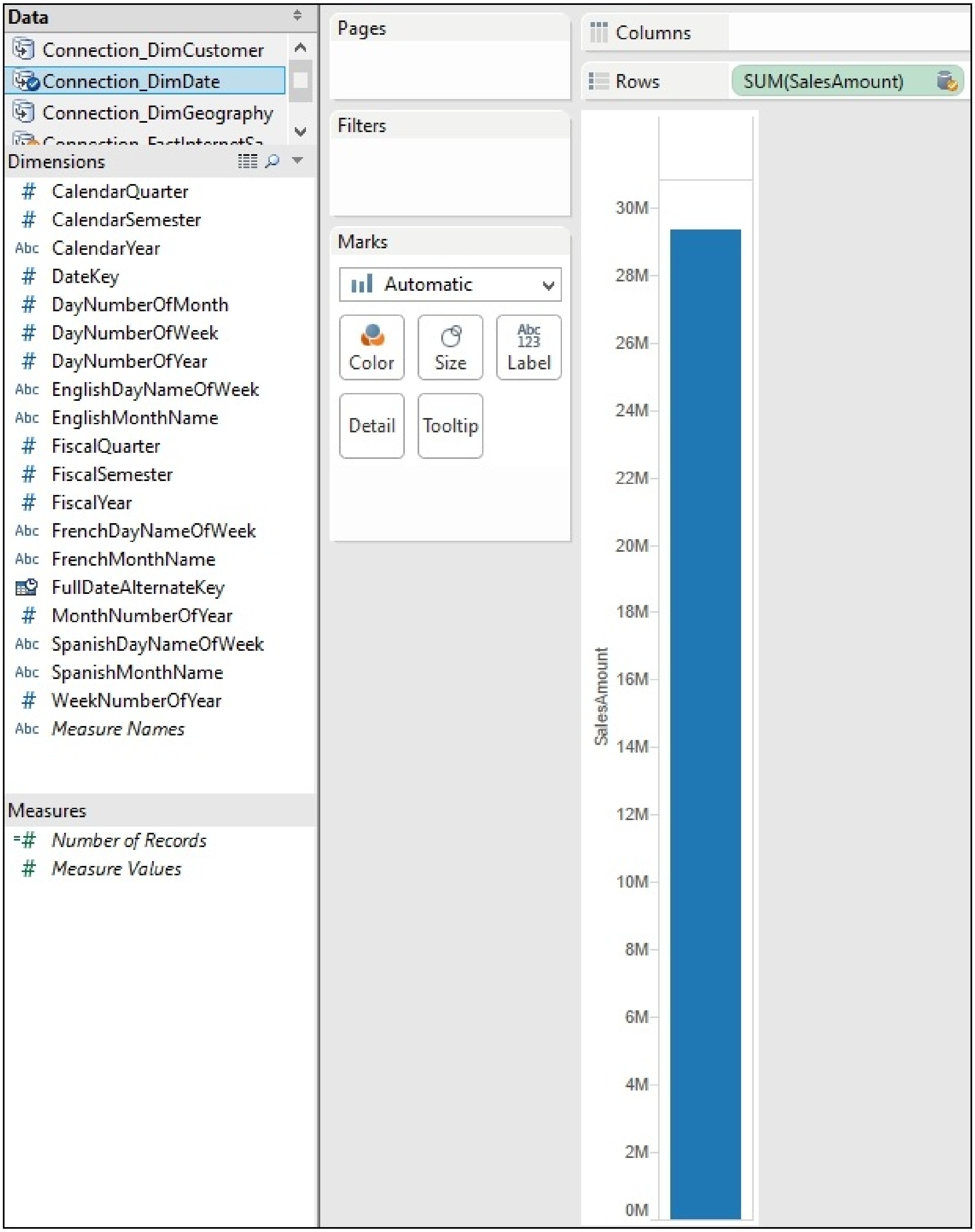


- Once the data is imported, Tableau will ask where you would like to store the Tableau Data Extract file.

Tip

If you store the TDE file in a location that is synchronized with SkyDrive, you will need to rename the file to remove the # file since SkyDrive will not save the file due to the presence of this character.

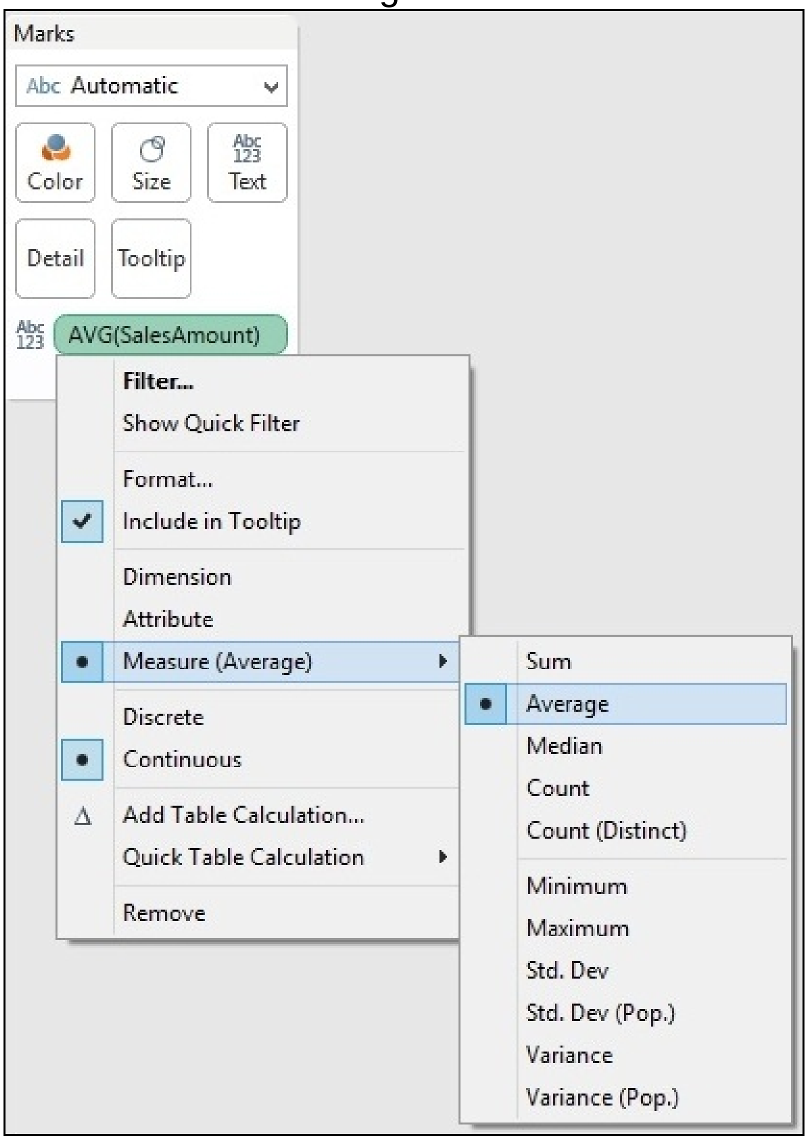
- Firstly, we will calculate the average of the sales amount. Take **SalesAmount** and drag it to the **Rows** shelf. You will see that Tableau immediately visualizes the data and turns it into a bar chart with one vertical bar. You can see this in the following screenshot:



- We would like to see the actual figure, so we will turn the data into a table. To do this, go to the **Show Me** panel and select the table. The **SalesAmount** pill will disappear from the **Rows** shelf and will reappear in the textbox in the **Marks** shelf, since Tableau is now showing a number.

The first visualization that Tableau selects is a sum of **SalesAmount**. However, since we are interested in the measures of central tendency, we are interested in the average, median, and mode. We will calculate these values to perform a quick summary of the data and also to understand how much we can rely on one number, the mean, to summarize the data.

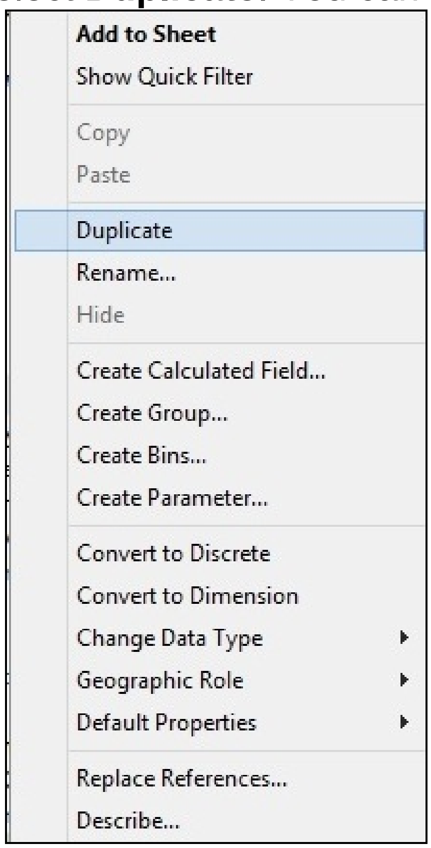
- To calculate the average, simply click on the **SalesAmount** lozenge in the **Rows** shelf and click on the down arrow to show the menu. You can see this in the following screenshot:



- The average sales amount is $486.09.
- The average is calculated and visualized in a table. However, we would like to see the median as well. To do this, make sure that the

SalesAmount pill is on the **Rows** shelf.

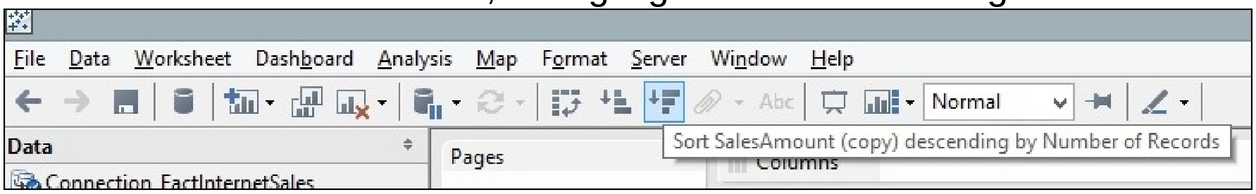
- Right-click on the pill and select **Median**. The median **SalesAmount** is given as 29.99.
- Next, remove **SalesAmount** so that we have a clean canvas again.
- Then, we need to calculate the mode, which can be defined as the number that occurs most frequently. To calculate the mode, we need to make a copy of **SalesAmount** and make it a dimension. This is a workaround to help the Tableau user to work out how many times each price occurred. To do this, right-click on the **SalesAmount** column in the **Measures** window and select **Duplicate**. You can see the menu item in the following screenshot:



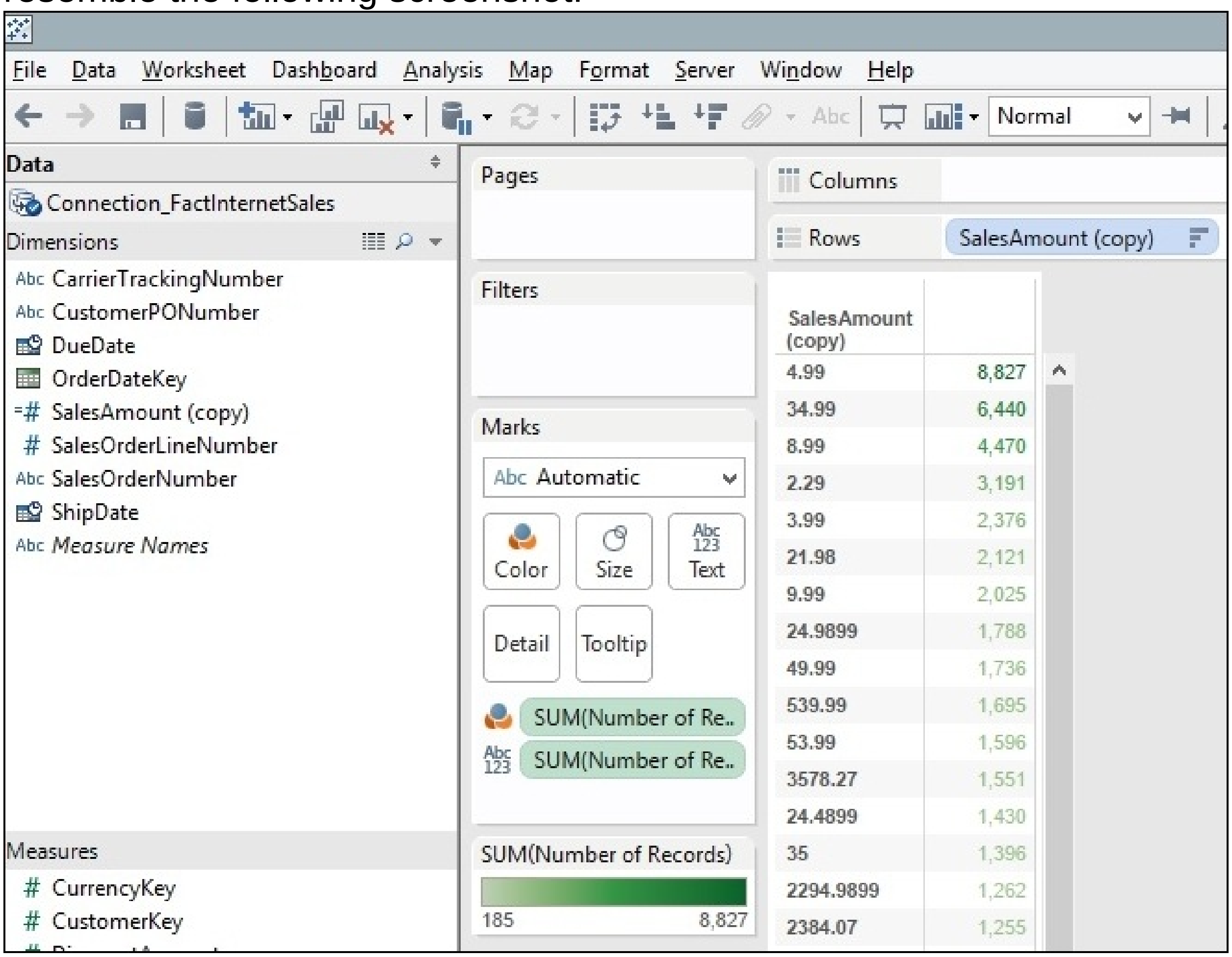
- Once you have duplicated the **SalesAmount** figure, you will see that the duplicate is called **SalesAmount (copy)**.
- Next, drag **SalesAmount (copy)** to the **Dimensions** shelf in the **Data** sidebar so that the table works properly.
- Drag the **SalesAmount (copy)** item to the **Rows** shelf.
- Next, you can go back to the **Measures** pane and move the **Number of Records** pill to the canvas area, next to the **SalesAmount (copy)** column.
- This will give us the number of records for each sales amount.

However, it is quite difficult to work out the most commonly-occurring sales amount from the table simply by looking at it. What we can do instead is sort the data so that the maximum is at the top, and the top item provides us with the mean.

- To sort the data, go to the top of the Tableau interface and look for the horizontal bar chart symbol with the downward arrow. This will sort the data by the number of records in descending order. You can see this under the **Server** menu item, as highlighted in the following screenshot:



- Once you've done this, drag the **Number of Records** item on to the **Color** shelf of the **Marks** shelf. Once this is done, your screen should resemble the following screenshot:



- The mode is actually the topmost number; the highest frequently occurring **SalesAmount** is $4.99.

How it works…

To summarize, in this section, we learned the following:

- Basic calculations
- Changing the color
- Basic sorting

We can see that the average and median values are quite different. The mode is $4.99, but the median is $29.99. Since the average is much higher than the median and the mode is different from the other two numbers, the data is not symmetrical. Often, when business analysts look at data, they try to find out whether the data is close to a normal curve or not. The average, median, and mode help us to determine whether the data is close to a normal distribution or whether the data is shaped. Therefore, we can't just simply use the average to summarize the data; we need to know about the other items too. This helps us understand the skewedness of the data, or how far it is from the central measures.

There's more…

If you are interested in learning more about analyzing data and the normal curve, then you can take a look at http://en.wikipedia.org/wiki/Normal_curve.

If statistics interest you, why not look at doing a Khan Academy course? This is a free facility for learning statistics yourself online; refer to https://www.khanacademy.org/.

Dashboards and dates

In Business Intelligence, dates are an essential part of analysis, and they are an important part of Business Intelligence projects. Data warehouses, for example, have a Date dimension as a way of helping business users to describe their data by date. People's business questions often include a **when** element. Additionally, dashboards will often reference dates.

Comparison is fundamental to analysis. Time is a fundamental part of comparison. Dashboards will often display comparisons between periods of time, so time is an essential part of the dashboard display. It is easy to envisage the following business questions that involve time:

- What are my numbers compared to last year?
- How did my sales region perform this month as compared to the last month?
- When will my department reach its target?

This recipe explains how to use dates in order to analyze your data using the Dates functionality in Tableau. In the last recipe, we imported the `FactInternetSales` table. Once the `FactInternetSales` data has been imported, we will also do the same for the `DimDate` table. To analyze our data, we will perform the following actions:

- Join data together
- Activate links in relationships
- Date analysis in Tableau

Getting ready

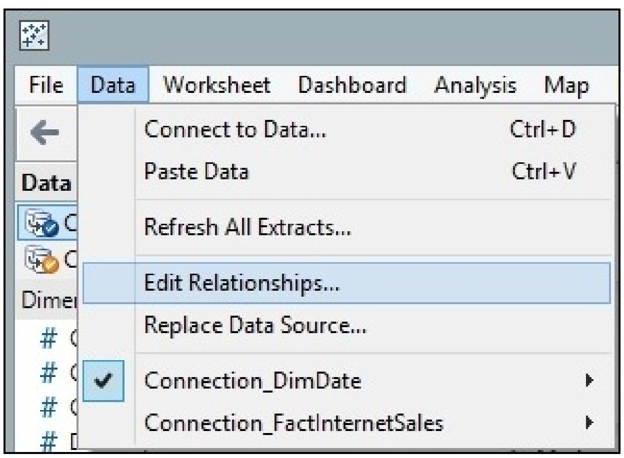
We will continue to use the same Tableau workbook we set up in the first recipe. It was called `Chapter 2`.

How to do it...

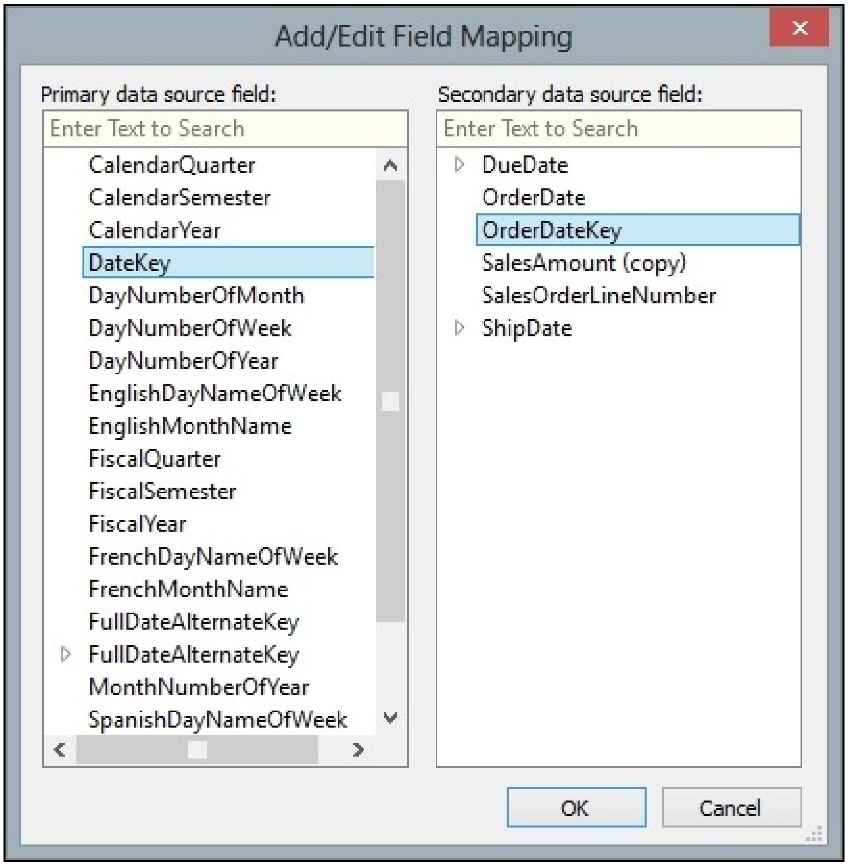
1. We will connect to an Excel file called `DimDate.xls` and import it into

Tableau's internal data store mechanism. To do this, navigate to **Data | Connect to Data**.

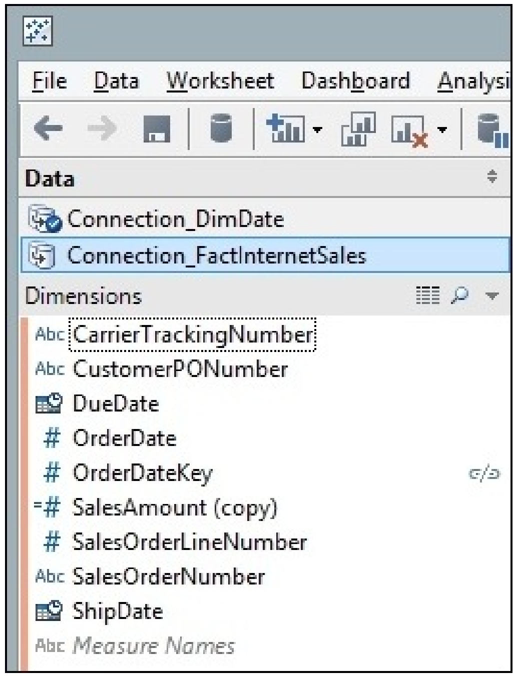
2. Then, select the link **Microsoft Excel** and a file browser will appear. Navigate to the folder where you stored the downloaded Excel and CSV files.
3. Navigate to the `DimDate.xls` file and select it.
4. In the **Excel Workbook Connection** dialog box, change the **Step 4** name so that it reads `Connection_DimDate`.
5. We have two data sources which will need to be joined together so that the data can be analyzed by date. We can join the `DateKey` and `OrderDateKey` columns together for the `DimDate` and `FactInternetSales` tables respectively. Make sure that the `DateKey` and `OrderDateKey` columns are both dimensions, not measures. So, ensure that both fields are contained in the **Dimensions** field; if not, drag them from the **Measures** pane to the **Dimensions** pane. If they are measures, then we cannot join them.
6. To do this, we edit the relationship between the two tables. Ensure that you have selected the `Connection_DimDate` data source in the **Data** pane in the sidebar.
7. Next, go to the **Data** menu item and select **Edit Relationships**. You can see this illustrated in the following screenshot:

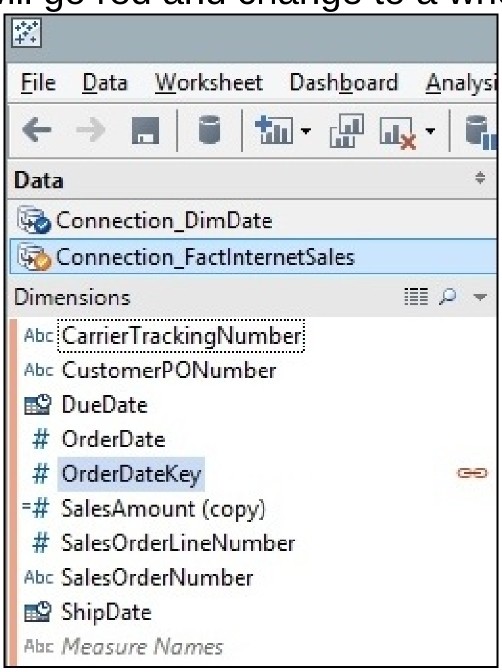


- You can now see the **Relationships** dialog box. Click on **Custom** and then click on **Add**.
- Next, you will see two columns: one for the `DimDate` columns and the other for the `FactInternetSales` columns. Each column represents a table, and the items in the list are source fields.

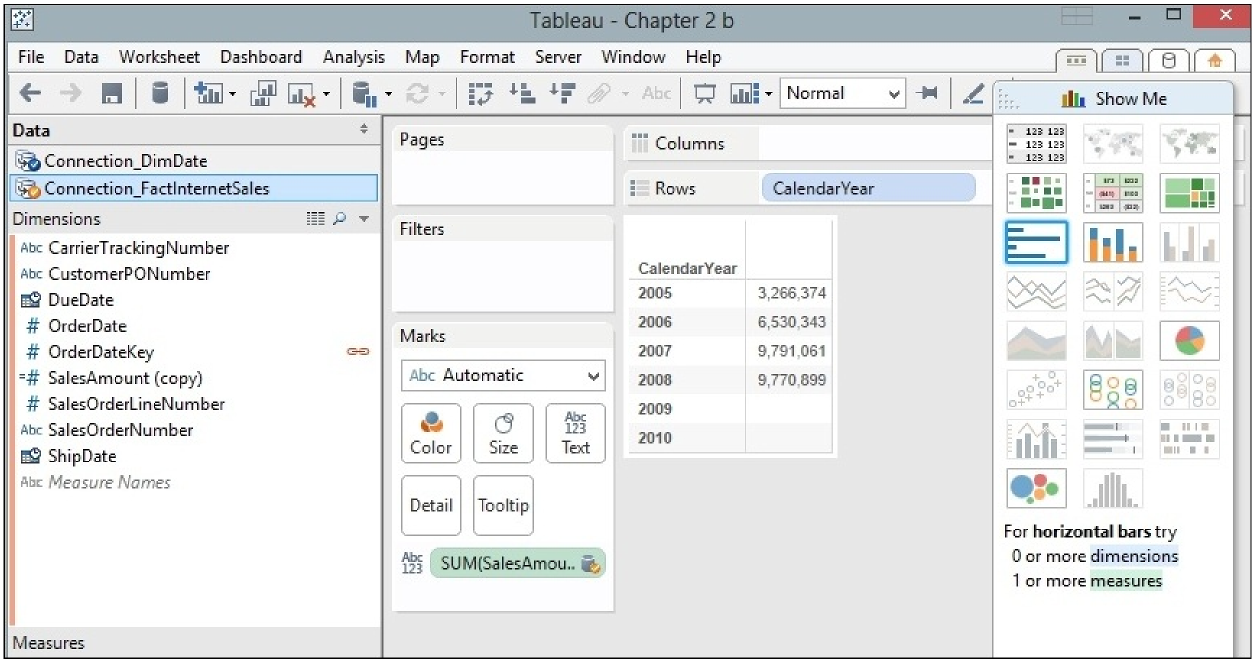
- Select **DateKey** on the left-hand side, **OrderDateKey** on the right-hand side, and then click on the **OK** button. You can see this in the following screenshot: 

- Let's create a new sheet by going to the tab at the foot of the worksheet with the worksheet name on it. Right-click to select **Duplicate Sheet**. Let's make the canvas clean again by removing all of the pills from the **Columns**, **Rows**, and **Marks** shelves.
- Let's use the new worksheet to proceed.
- In the **Data** pane, select the **Connection_DimDate** dimension. Select **Calendar Year** and drag it into the **Rows** shelf. Make sure it is set to the String data type by right-clicking on **Calendar Year** on the **Data** pane, then navigating to **Change Data Type**, and finally selecting **String**.
- Next, click on the **Connection_FactInternetSales** connection and you will see a broken link symbol next to **OrderDateKey**. You can see this in

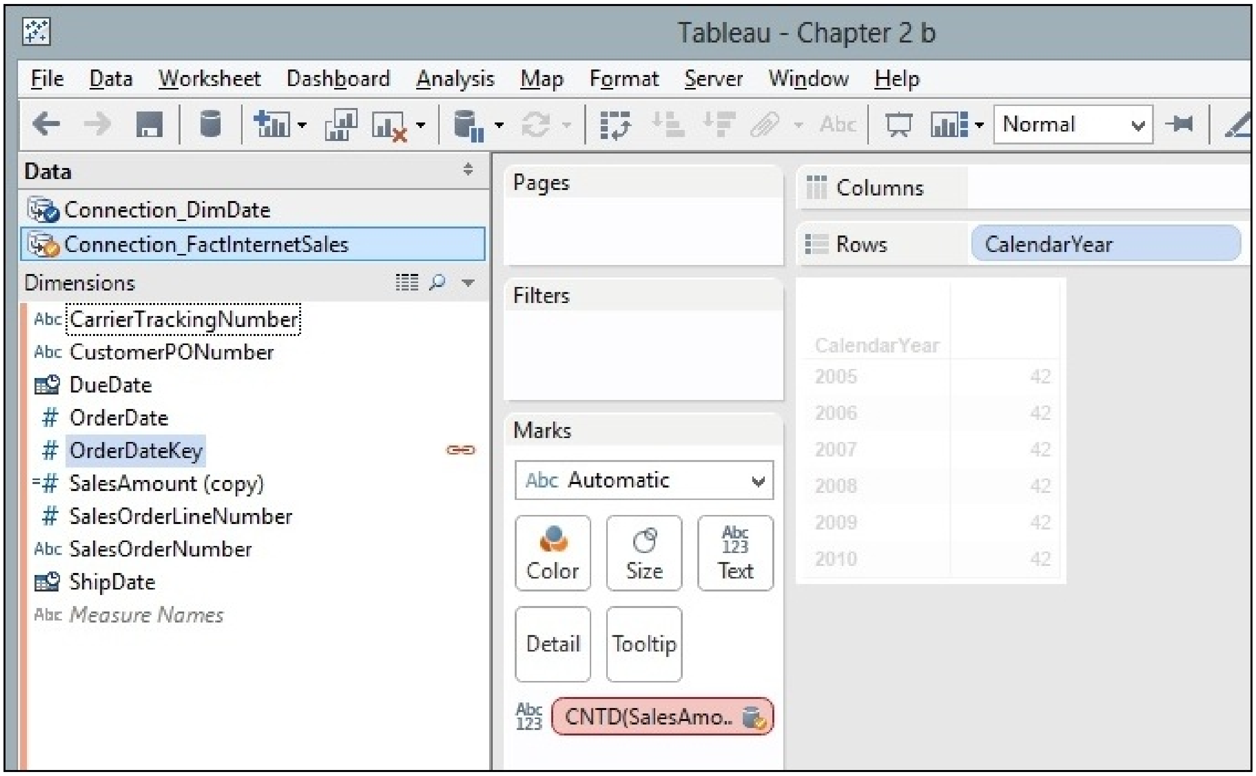
the following screenshot: 

- We need to make the relationship between the data sources active. If we don't, then the data sources will not be related to one another and our analysis will not make sense. To activate the link, click on the broken chain; it will go red and change to a whole link. The following is an example: 

- Staying in the `FactInternetSales` table, drag the **SalesAmount** column from the **Measures** pane to the **Canvas** shelf. You can see this in the following screenshot:



- If you break the link by clicking on it again, then you will notice that the figures change. Each calendar year is now set to 29,358,677, rather than the figures shown in the previous screenshot. This is because the link between the tables is now broken, so Tableau cannot sum the data according to the years; instead, it issues a query to both data sources and simply returns the product of the total.
- Let's look at what happens if we use a different aggregation, **count distinct**. This returns a count of the distinct **SalesAmount** values. When we link the `OrderDateKey` column and select **Count Distinct** for the **SalesAmount** figure, Tableau turns the **SalesAmount** pill to red and grays out the whole screen. You can see an example of this in the following screenshot:



- In Tableau 8, the aggregation is considered not valid, so Tableau saves the user from themselves by graying out the screen until it is fixed. You have the facility to undo the last step. To do this, press *Ctrl* + *Z*.

How it works…

Tableau allows business users to enrich their data through the addition of calculations that get stored as part of the workbook. This is useful for data analysis, since we can look and see how simple steps can quickly affect the data.

Why is this the case? Tableau issues separate queries to each data source and joins the two data sources together. Then, it conducts the aggregation on the joined data sources at the lowest level of detail in the view of the data from the Tableau interface.

Unfortunately, if the level of detail in the underlying query is different from the level of detail in the view, then the calculation will not be correct.

The moral of the story is to keep the relevant dimensions in the Tableau view, that is, put their features in the **Columns** and **Rows** shelves or in the **Marks** shelf. By putting more elements in the view, these shelves will move the query towards serving up data that can be used for matching the data sources. Then, the user can try to incorporate the query by including as many dimensions as possible, which in turn will produce as much detail in the query as possible in order to facilitate matching between the tables.

There's more…

Dashboards use a lot of calculations to summarize data. Research by specialists such as Ben Shneiderman shows that people tend to want to see the summary first, followed by zooming and filtering the data, and then finally see the details on demand. This is a very natural way of engaging with data. Shneiderman calls this the "Visual Information Seeking Mantra".

If you are interested in the psychology of how individuals interact with data, then Ben Shneiderman's paper *The Eyes Have It* (1996), which you can find at http://dl.acm.org/citation.cfm?id=834354, will be of help.

Grouping your data with calculations

In the first recipe, we specified communication as one of the key features of a dashboard. We need to be able to share the right information with the right audience, at the right time, to the right people, and in the right format.

Sometimes, data needs to be translated so that it matches the business rules and the business understanding of the organization. Tableau offers a number of different ways that help you translate data into something that the business decision makers will understand.

Grouping is one way of making data meaningful to the business. In this recipe, we will look at grouping some dimension members into a single member. Rolling up some of the members in one dimension is a good way of summarizing data for dashboards.

In this recipe, we will look at grouping dimension members; then, we will look at more complex grouping of calculations. The business question is an investigation into the characteristics of customers, for example, those who have children, and those who do not. We will group the **NumberChildrenAtHome** dimension members into the group of customers who have children and those who do not.

Then, we will look at a more advanced example of grouping the data by measure rather than dimension. To do this, we can create a calculation that will distinguish the values that are below the average sales amount and above the average. Results that are classified as above average are labelled **Above or Equal to Average**, and below average sales are labelled **Below Average**. We can then use this calculation to convey a visual message to the business user; for example, we could color the above average sales in one color and the below average sales in another in order to make the distinction easily identifiable.

Getting ready

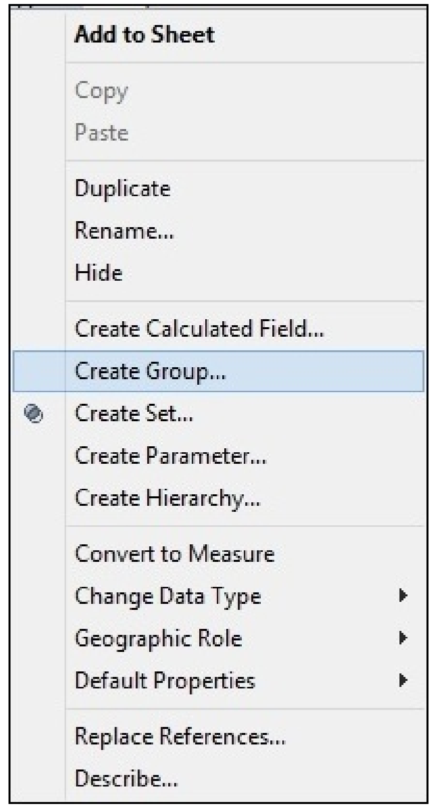
We will need to add in a new data source for this recipe.

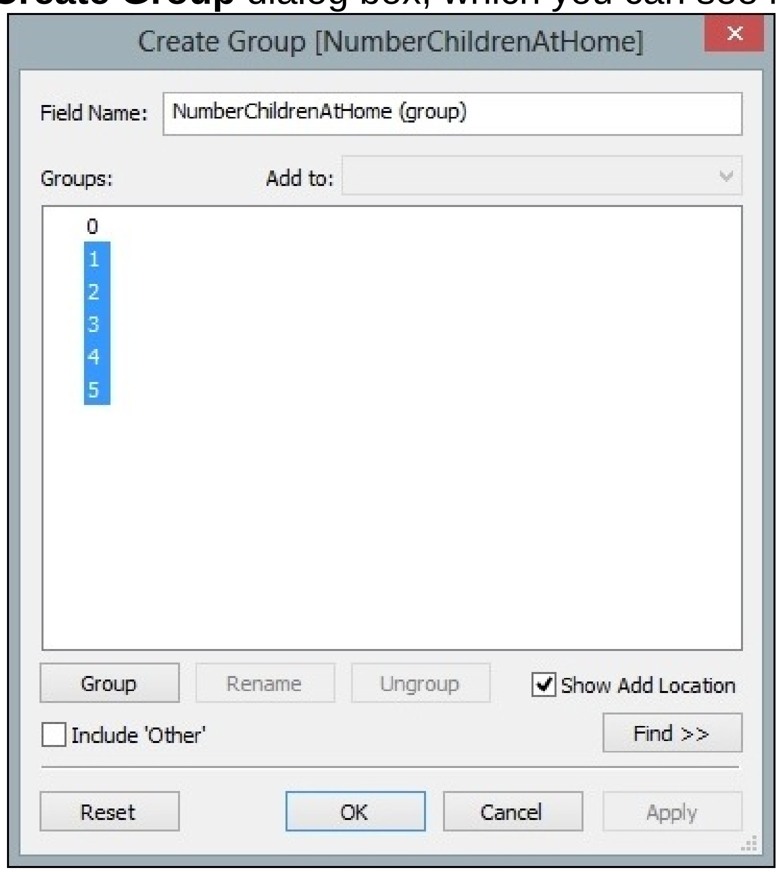
How to do it...

1. We will connect to an Excel data file called `DimCustomer.xls` and import it into Tableau's internal data store mechanism. To do this, navigate to **Data | Connect to Data**.
2. Then, select the link **Text File** and a file browser will appear. Navigate to the folder where you stored the downloaded Excel and CSV files.
3. Navigate to the `DimCustomer.csv` file and select it.
4. In the **Excel Workbook Connection** dialog box, change the **Step 4** name so that it reads `Connection_DimDate`.
5. You can now see the new data source connection on the **Data** pane.
6. Click on *Ctrl* + *M* to get a new worksheet in Tableau. Alternatively, go to the **Worksheet** menu item and select **New Worksheet**.
7. Click on the **Connection_DimDate** source and drag **Number of Records** from the **Measures** pane to the **Rows** shelf.
8. Drag the **NumberChildrenAtHome** dimension attribute to the Tableau canvas, which is to the left of the `NumberOfRecords` column. You can see an example in the next screenshot.
9. When we look at the **NumberChildrenAtHome** dimension, we see the following members and the number of customer records associated with each member:

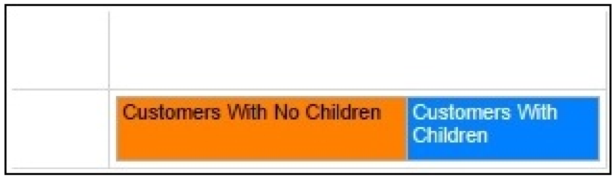
NumberChildrenAtHome	
0	11,116
1	2,460
2	1,648
3	1,204
4	1,089
5	967

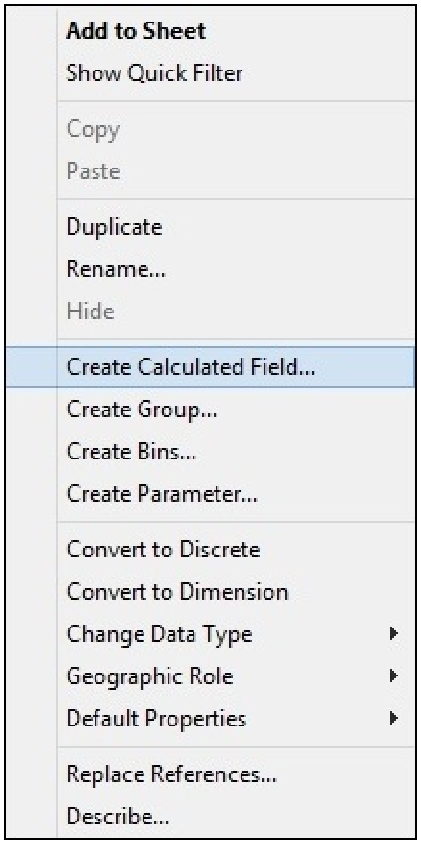
- We will group the dimension members so it is easier to see which customers have children and which do not.
- Go to the dimension called **NumberChildrenAtHome**, right-click on it,

and select the **Create Group** option as shown: 

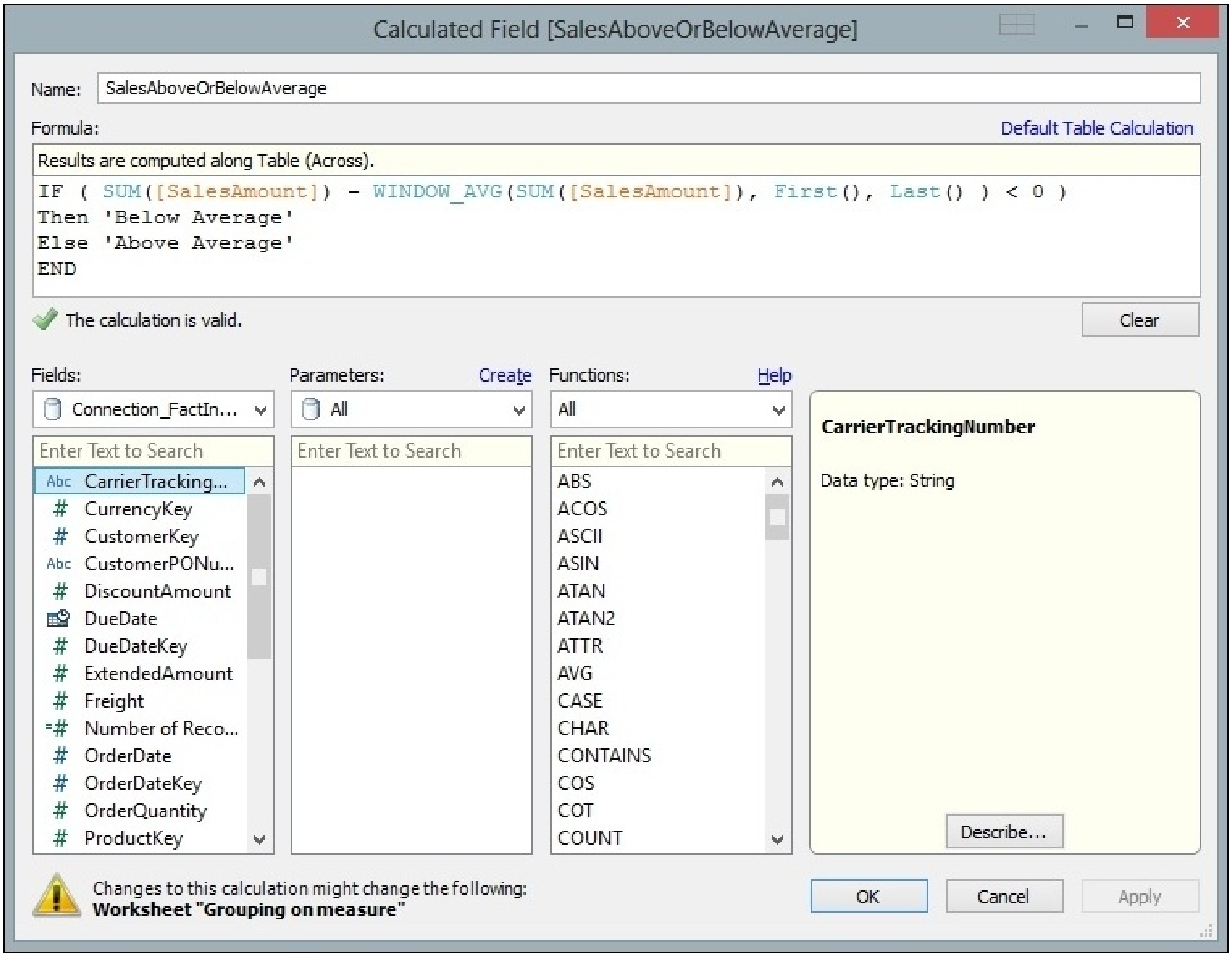
- This produces the **Create Group** dialog box, which you can see in the following screenshot: 

- In the **Field Name** field, enter the name of the field.
- Multiselect the numbers 1 through to 5 by holding the *Shift* key and clicking to select more than one number at a time.
- Click on **Group**, rename the group to `CustomersWithChildren`, and click on **OK**.
- You can then see the new group on the left-hand side on the **Dimensions** pane. Drag your new grouping to the **Rows** shelf.
- Then, choose **Number of Records** from the **Measures** pane and put it into the **Columns** shelf. You can see that the table now only has two rows in it: one that consists of a zero and another that has **CustomersWithChildren**.
- In order to make things clear, click on the zero and select **Edit Alias**.
- Rename the zero to `Customers with No Children`.
- You can also rename the **Group** alias to `Customers With Children` so that it matches the format used elsewhere. To do this, select the numbers 1 to 4 in the table and right-click on them.
- In the pop-up menu, select **Group**.
- Right-click on the new group and select **Edit Alias**.
- In the textbox, enter `Customers With Children` and click on **OK**.
- It is now clear from the table that over 11,000 customers have no children, whereas just over 7,000 customers do have children. We can visualize this information in a better way, and we will do this for the rest of the exercise.
- To do this, change the visualization to a heat map using the **Show Me** panel.
- We can then change the colors by dragging **Number of Records** to the **Color** panel on the **Marks** shelf.
- In order to show the difference between the customers who do and do not have children, the custom diverging color palette has been selected here.
- Two diverging colors of orange and blue are selected. Orange is selected for **Customers With No Children** because there is a greater number of customers who do not have children; brighter and more intense colors are often used in order to denote higher values, as shown

in the following screenshot: 

- To do this, stay in the `Chapter 2` workbook and continue to work in the existing worksheet.
- Right-click on the **SalesAmount** measure.
- Select the measure **Create Calculated Field** that is illustrated in the following screenshot: 

- When you select this option, you will get the following dialog box:



- In the **Name** field, enter the name of the calculated field: `SalesAboveOrBelowAverage`.
- In the **Formula** field, we will put in a formula that will calculate whether or not the sales amount is above or below the average amount. The formula is as follows:

```
IF ( SUM([SalesAmount]) -
WINDOW_AVG(SUM([SalesAmount]), First(), Last() ) < 0 )
Then 'Below Average'
Else 'Above Average or Equal To Average'
END
```

Tip

Downloading the example code

You can download the example code files for all Packt books you have purchased from your account at http://www.packtpub.com. If you purchased this book elsewhere, you can visit http://www.packtpub.com/support and register to have the files e-mailed directly to you.

- Once you've placed the formula into the calculation editor and clicked on **OK**, you will be returned to the main Tableau interface. You will see your new calculation on the left-hand side in the **Measures** pane.
- Drag the **SUM(SalesAmount)** measure to the **Columns** shelf.
- Drag **NumberCarsOwned** from the **Dimension** pane to the **Rows** shelf.
- Drag your new calculation **SalesAboveOrBelowAverage** to the **Color** button on the **Marks** shelf.
- Your screen should now look like the following:

To summarize, we have created a calculation that is meaningful to a business user. It provides the color display of the measure, which helps the business user understand things more efficiently. To summarize, it is simple and effective to conduct a grouping of dimension members into a binary grouping. This is useful for dashboards in order to provide an "at a glance" metric visualization that shows the organization has more customers who do not have children than those who do.

Essentially, this formula uses the `WINDOW_AVG` function to work out the average of the values that are in the Tableau view of the data. Basically, this average works out the value of the data that is viewable in the Tableau canvas and does not include data that has been filtered.

It uses `First()` and `Last()` to work out the average of all the data from the first row right until the last row. The calculation takes the current **SalesAmount** value and compares it with the average **SalesAmount** value.

How it works...

Tableau allows you to group data together by simply arranging fields of your data source on a Tableau worksheet.

When you group fields in a worksheet, Tableau queries the data using standard drivers and query languages (such as SQL and MDX). It then groups data together wherever necessary. Finally, it presents a visual analysis of the data.

Correlation with calculations

Data visualization is all about communicating a message using data. In the first chapter, we specified communication as one of the key features of a dashboard.

The problem with data visualization is that people don't always like what visualization tells them about their business. It can defy commonly held assumptions about a business that goes against the grain of what people believe. This can be particularly uncomfortable if people have been with an organization for a long time and perhaps have not changed their perspective as the business moved ahead.

In this recipe, we will look at how we can use correlation to test a hypothesis and then display this information so that the message of the data is easy to understand.

Getting ready

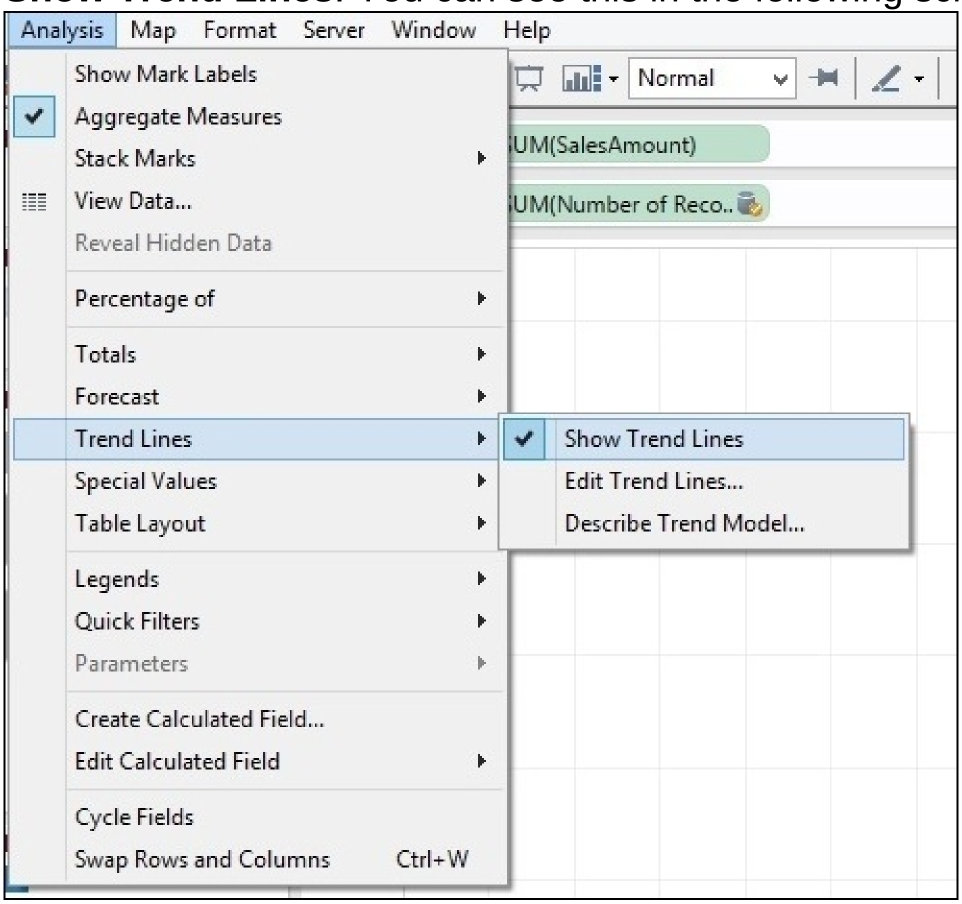
We will continue to use the same worksheet as we have used in the previous recipes of this chapter.

How to do it...

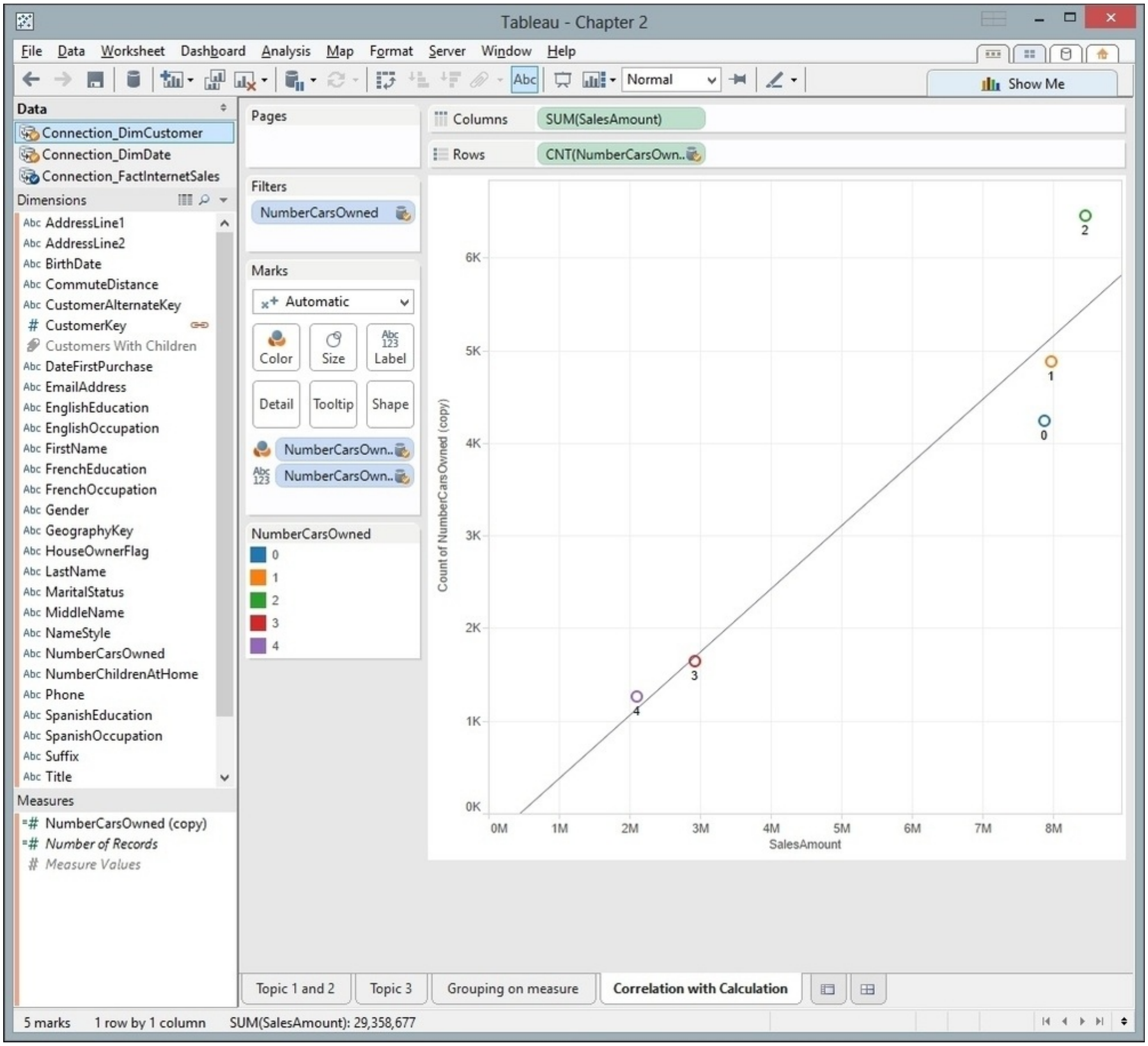
In this recipe, we will use a real-life example where a business analyst wants to know if people with more cars spend less on bikes. We will test whether there is a correlation between the number of cars owned by a customer and how much money is spent on bikes. Since the AdventureWorks store is selling bikes, there is an assumption that people who own fewer cars will spend more on bikes. However, this hypothesis would need to be tested, and data visualization can help us "sense-check" the data and see patterns easily. Perform the following steps to create correlations in Tableau:

1. Open up the `Chapter 2` workbook and create a new worksheet by pressing *Ctrl* + *M*, or by going to the **Worksheet** menu item and then

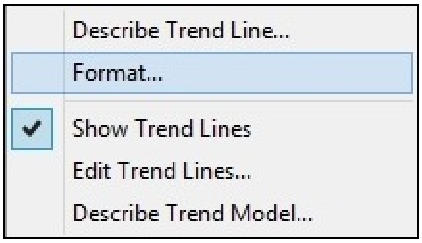
selecting **New Worksheet**.
2. Rename the tab `Correlation with Calculations`.
3. Take the dimension named **NumberCarsOwned** from the `DimCustomer` table and drag it on to the **Rows** shelf. Change the measure to **Count**.
4. Take the **SalesAmount** measure from the **FactInternetSales** table and place it on the **Columns** shelf.
5. Take the **NumberCarsOwned** member from the **DimCustomer** dimension and place it on the **Color** attribute in the **Marks** shelf.
6. Next, we will add a trend line order to convey the relationship between the sales amount and the number of sales. To do this, go to the **Analysis** menu item and the **Trend Lines** menu item and select **Show Trend Lines**. You can see this in the following screenshot:



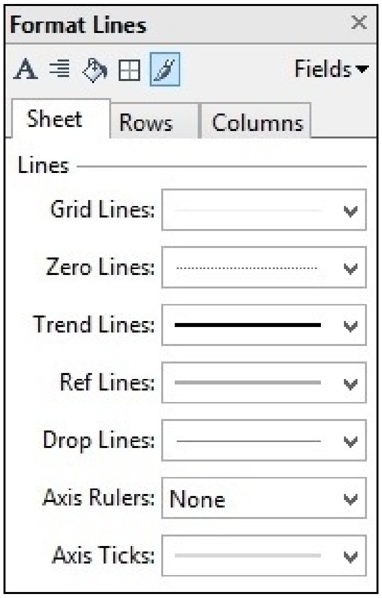
- The visualization now appears as follows:



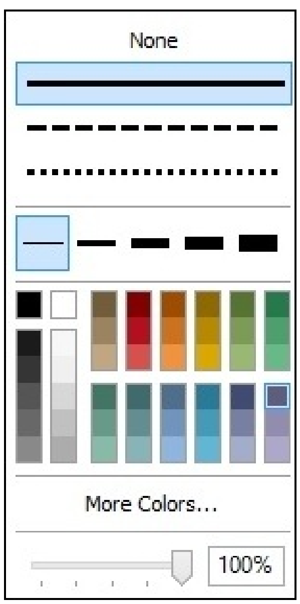
- The image needs some more detail in order to clarify the message of the data.
- We can add labels to the data points that will help us identify the data points more clearly. To do this, we can drag the **NumberCarsOwned** dimension to the **Label** option in the **Marks** shelf.
- The black trend line looks quite harsh, and we can soften it by turning it into a light gray color. This would still get the message of the data across, but without attracting too much attention away from other pieces of the image.
- To change the image, right-click on the trend line and you will get the

following menu: 

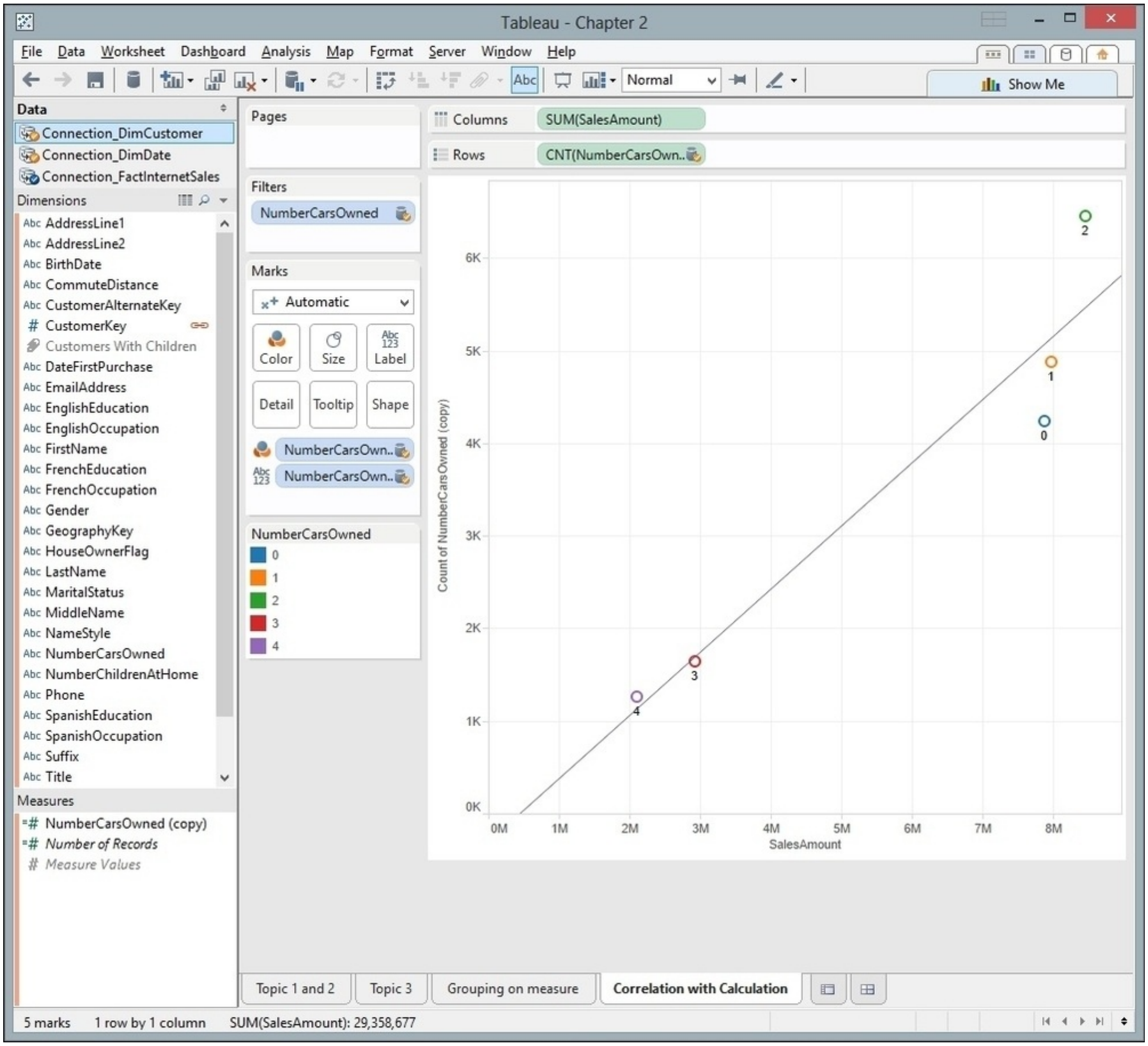
- Select **Format** and the following menu item will appear:



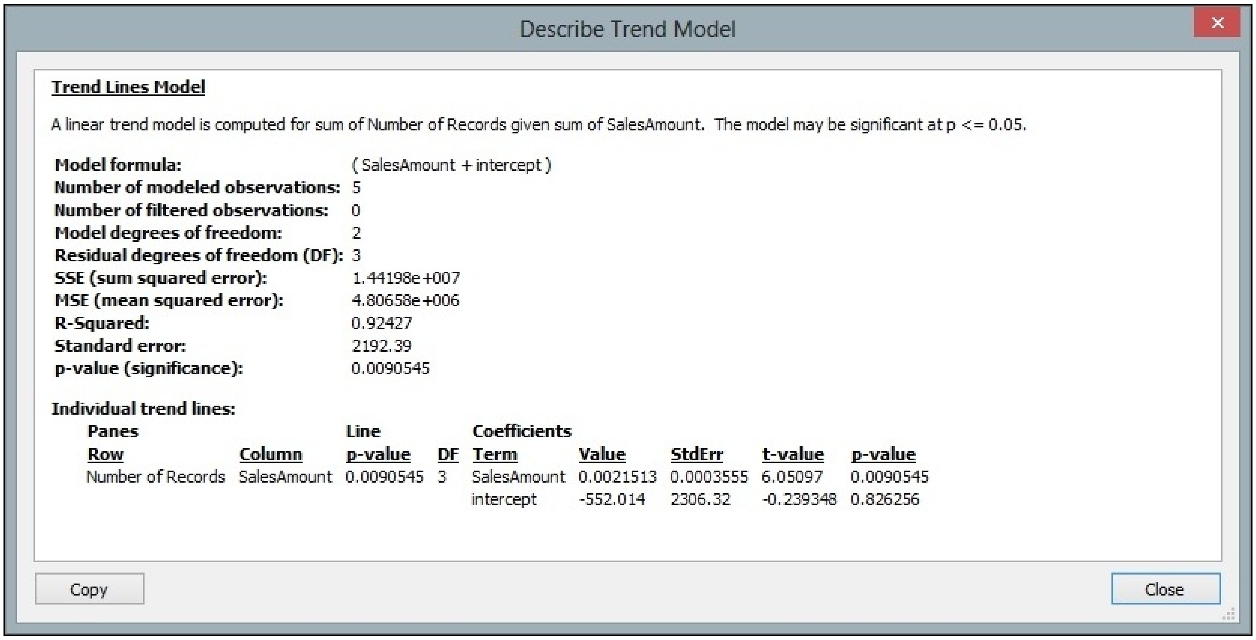
- We can soften the trend line by changing the width. To do this, click on the arrow and the following panel appears. This allows us to soften the trend line.



- The resulting visualization appears as you can see in the next screenshot:



- We can learn more about the trend line by selecting **Describe Trend Model** from the **Trend Lines** menu. This provides us with the following information:



- We can see the results of the analysis in the resulting window.

How it works...

Tableau provides a way to test hypotheses quickly and help business users to identify the relationships between variables. Tableau does this by simplifying the process of creating visualizations quickly and on-the-fly.

The visualization shows us that a linear correlation is produced, which shows how the sum of transactions or **Number of Records** is related to the sales amount.

We can see that the R-squared value is 0.92, and when we find the square root of this number, we will find that *R = 0.95*. This shows there is a very strong relationship between the two variables. We could then proceed to do some more analysis, but this would provide us with a great starting point.

There's more...

From a dashboard perspective, Tableau also helps business processes by allowing analysts to test assumptions that are perhaps long-held and subtle. When a visualization like this is placed with other visualizations that tell a similar story, then the dashboard becomes a valuable tool for promoting real business change.

Using cross-tabs flexibly

When we create dashboards, we are conveying a unified message of data that can be made up of moving parts. The context of the data can help to ensure that data pieces *move together* and are consistent with one another. This can mean that showing an aggregated number is not the most meaningful form of data. Instead, percentages or differences can be more meaningful.

In Tableau, we can create very simple calculations using table calculations. What is a table calculation? **Table calculations** are easy calculations that are provided by default as part of the Tableau interface. Table calculations are quick to create and are a powerful tool that can help to enhance your understanding of the data. From a dashboard perspective, we need to maximize the amount of information in a small space, and table calculations will help you to compress the message of the data while helping the data to be meaningful.

In this recipe, we will create a table calculation and see how it adds to the comprehension of the data in a small space. In this recipe, we will create a chart that shows the percentage of sales attributed to each region, rather than the number.

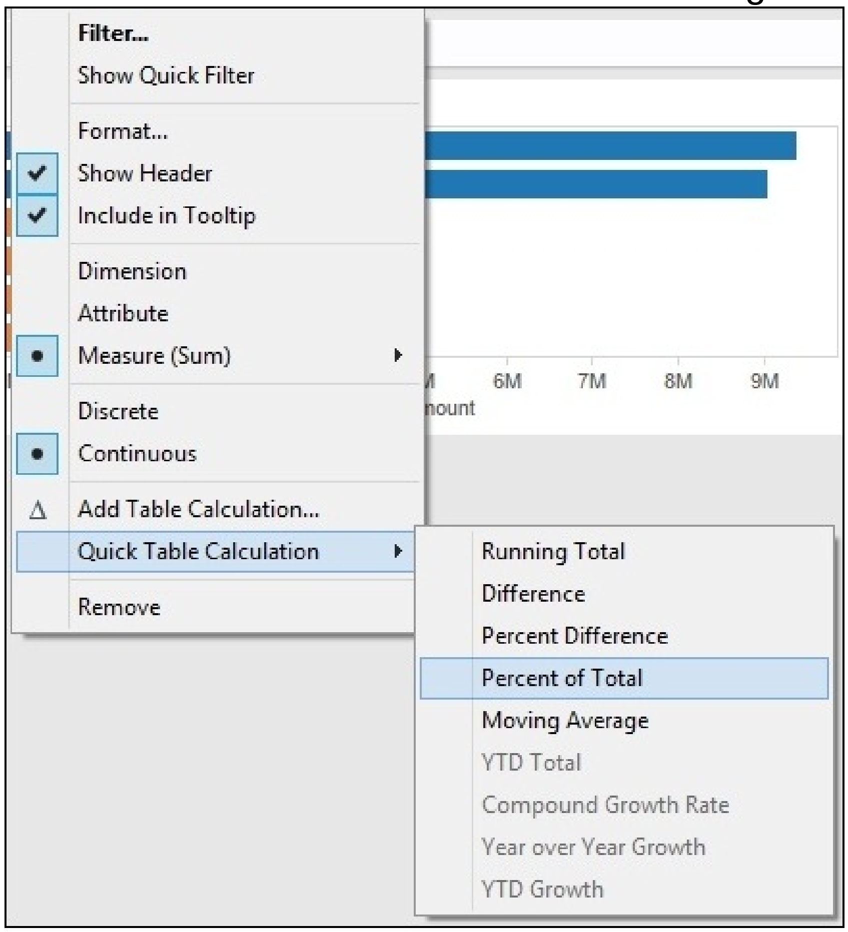
Getting ready

Open up the Tableau `Chapter 2` workbook and start a new sheet by selecting *Ctrl* + *M*.

How to do it...

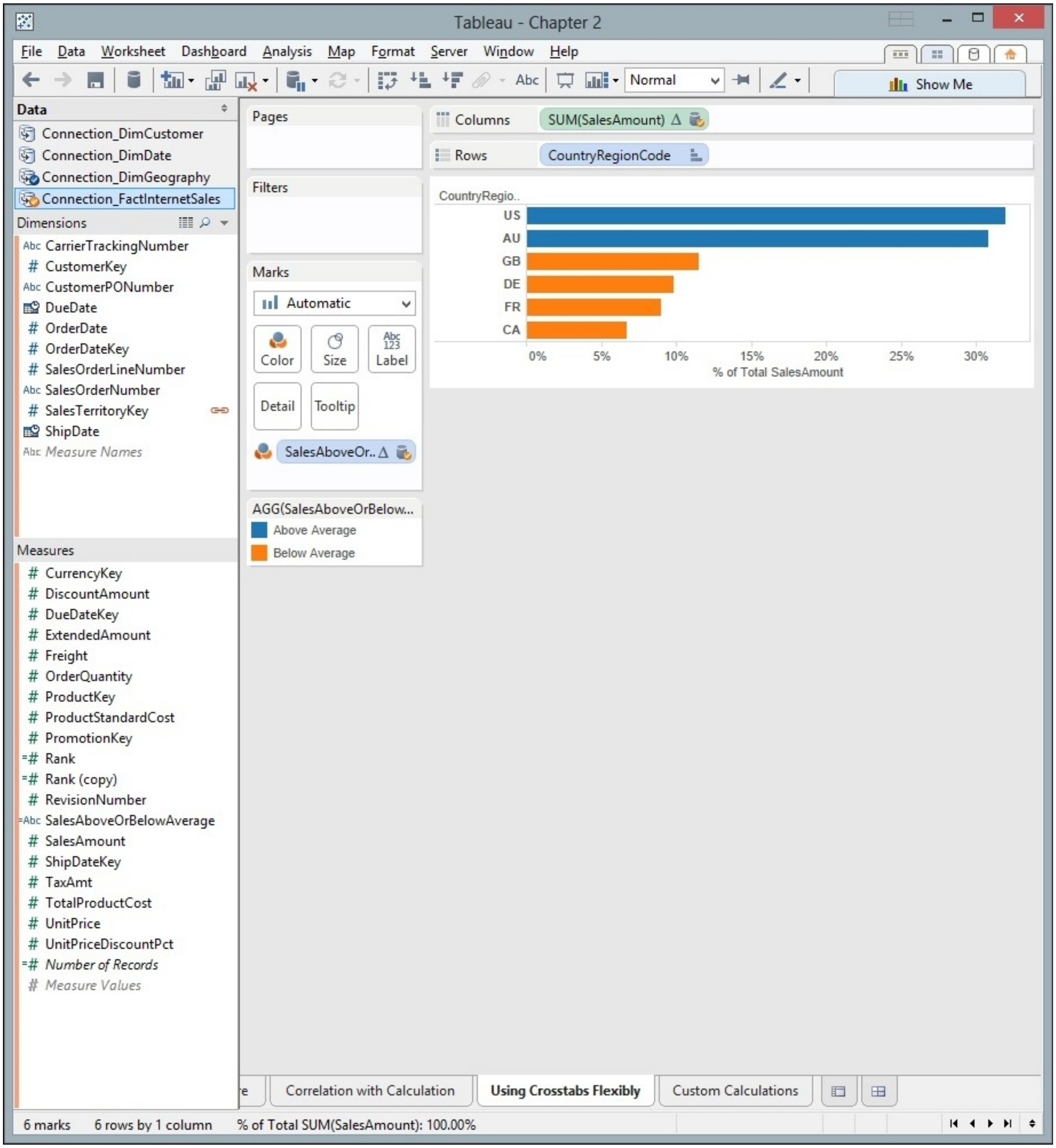
1. To do this, go to the **DimGeography** dimension in the **Dimensions** pane, select the **CountryRegionCode** member, and place it on the **Rows** shelf.
2. Then, go to **FactInternetSales** in the **Dimensions** pane, select **SalesAmount**, and drag it to the **Columns** shelf.

3. Right-click on the **SalesAmount** blue pill on the **Columns** shelf and navigate to **Quick Table Calculation**. You can see the following menu item appear:



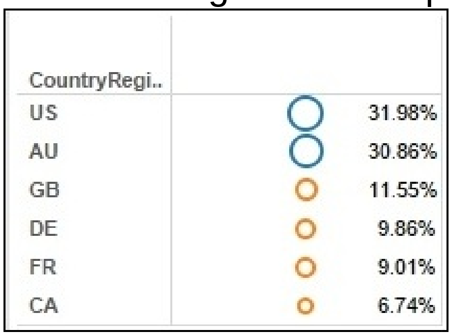
- For our example, we will choose **Percent of Total**. The **SalesAmount** figure will change, and you will see a small triangle appear on the right-hand side. This feature is illustrated as follows: 

- Take the **SalesAboveOrBelow Average** calculation we created earlier and place it in the **Color** section.
- For the visualization, choose a horizontal bar chart from the **Show Me** panel. Your screen should be the same as the following screenshot:



- Now, it is important to note that the *x* axis labeling has changed to show **% of Total SalesAmount**. This is more accurate, but how could we make it more obvious to the data consumer that they are in fact looking at percentages without cluttering the "real estate" on the dashboard? We will look at these steps next.
- We can set up a variation on a heat map. From the **Show Me** shelf, select the **heatmap** option.

- We will use a label to convey the percentage. To do this, take **SalesAmount** from the **Measures** pane and drag it to the **Label** option. Make sure that you change the calculation so that it uses the **Percentage of Total** option, as we did previously in this recipe. The label will now read the percentage of the total rather than the actual value.
- Then, go back to the **Marks** shelf and select the drop-down list so that it shows **Circle**; after this, space the heat map so that it looks more even. The following is an example of the resulting chart:



To summarize, the data is shown in a very compressed way, which provides details as well as a visual message about the data in question. It's clear that the US and Australia regions have above average sales; if more detail is required, the percentages are also given.

How it works...

Tableau allows us to tell a story simply in a small space! The use of table calculations is a very simple way to enhance the data, both visually and also in terms of details. Sometimes, however, using the default table calculations are not enough to meet the business needs and custom calculations are required. This is the subject of our next recipe.

There's more...

The author of the *Le Petit Prince* books, Antoine de Saint-Exupéry, was once quoted as saying the following:

> *"A designer knows he has achieved perfection not when there is nothing left to add, but when there is nothing left to take away."*

The simplicity of a design is often an asset in data visualization, and is better than adding more detail that deviates from the purpose of data visualization.

Edward Tufte coined a term called **chartjunk**, which he defined as follows:

> *"Chartjunk refers to all visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information."*

This is an important aspect of data visualization. When we are creating dashboards, we have to be careful about adding nonessential items. This is particularly important in situations where we have a small screen space, such as a dashboard or a mobile device.

Table calculations are useful because they help us maximize the space on the screen by providing more enhanced information and encoding business rules effectively. They help us get across essential business information, which is key to the visualization. If you are interested in reading more about chartjunk, it is recommended that you read *The Visual Display of Quantitative Information*, *Edward Tufte* (1983).

Simplifying your business rules with customer calculations

Dashboarding concentrates on representing data in a small space while still getting across key concepts in the data. It is important to use the space effectively. There should be a balance between representing too much information and representing not enough information.

Some people circumvent the issue altogether by avoiding data visualization or perhaps simply not requiring it for their roles. For example, you might run up against individuals who are not interested in the "pretty pictures" and want to see the numbers. How do you combine this requirement with other people's requirements to have a visual representation of the data?

Fortunately, there are a number of ways in which we can use data visualization techniques to convey the message of the data in a table. In itself, a table is a valid data visualization technique because tables are very good at representing detail. In particular environments, such as finance, it is the minute details within the table that make all the difference. In data visualization, we also have to cater to the people who just want the numbers as well as those who have a more visual requirement.

In this recipe, we will create a crosstab that has a custom calculation in it, which allows us to enhance the representation of the data for those who prefer a more visual approach; in addition to this, we'll also provision numbers for those who simply want a table.

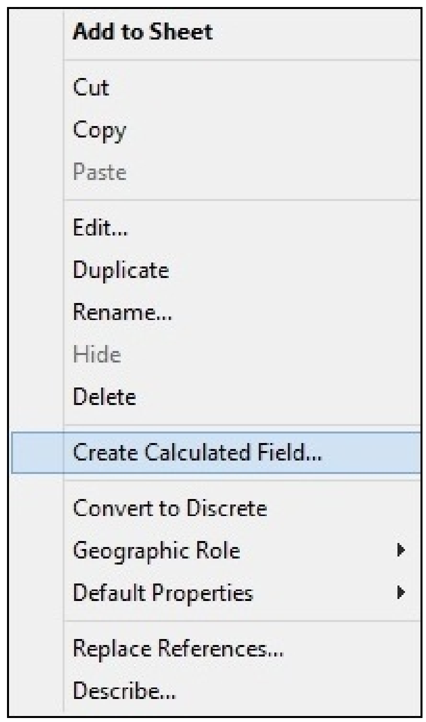
In this example, we will use the calculation `Index()`. This ranks members depending on their value. We can use the ranking calculation in order to restrict our data visualization so that it only shows the top three best performers. This is a concept that is key to many dashboards, and we will look at implementing this element in this recipe.

Getting ready

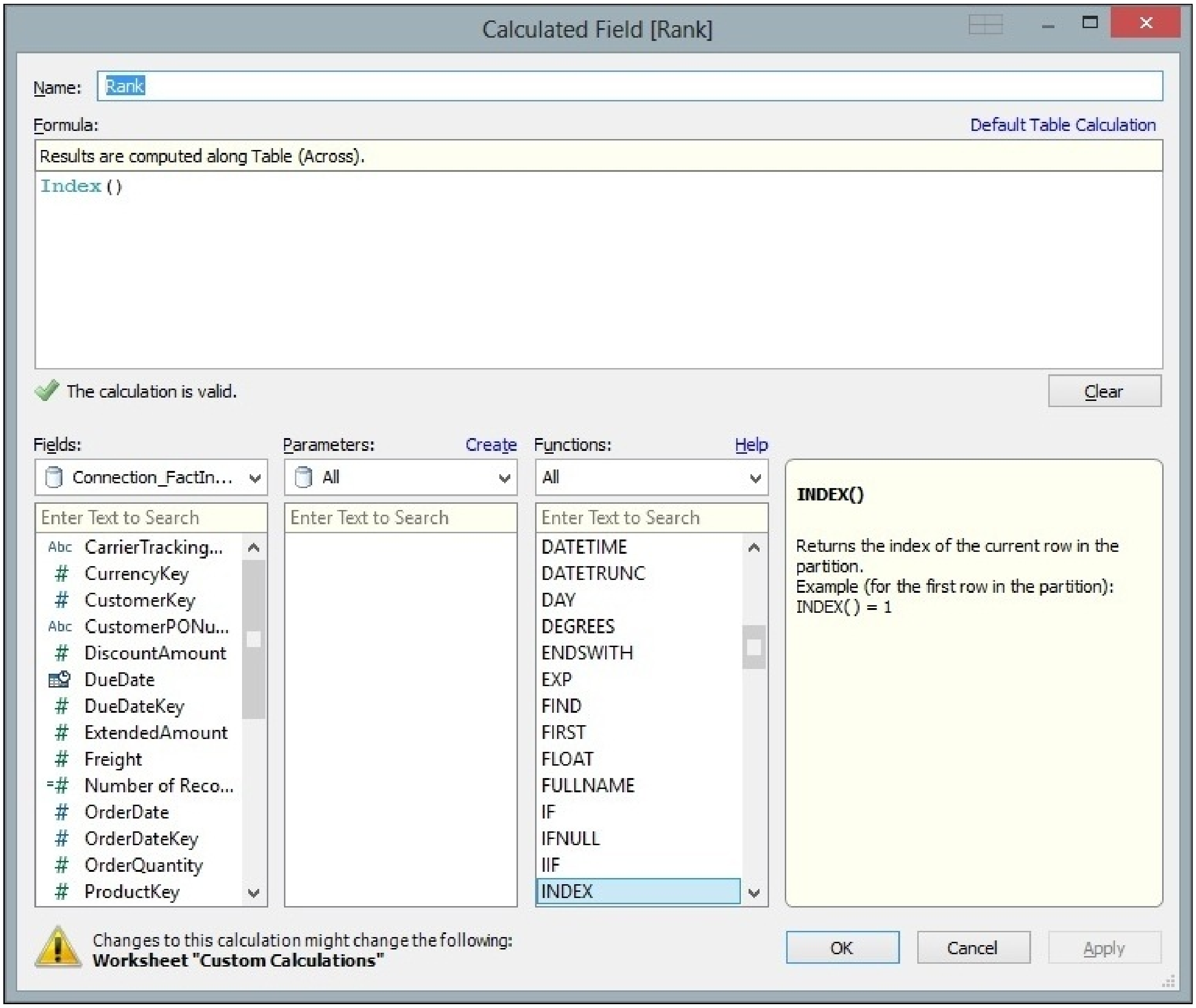
We will continue to work on the Tableau `Chapter 2` workbook and start a new sheet by selecting *Ctrl* + *M*.

How to do it...

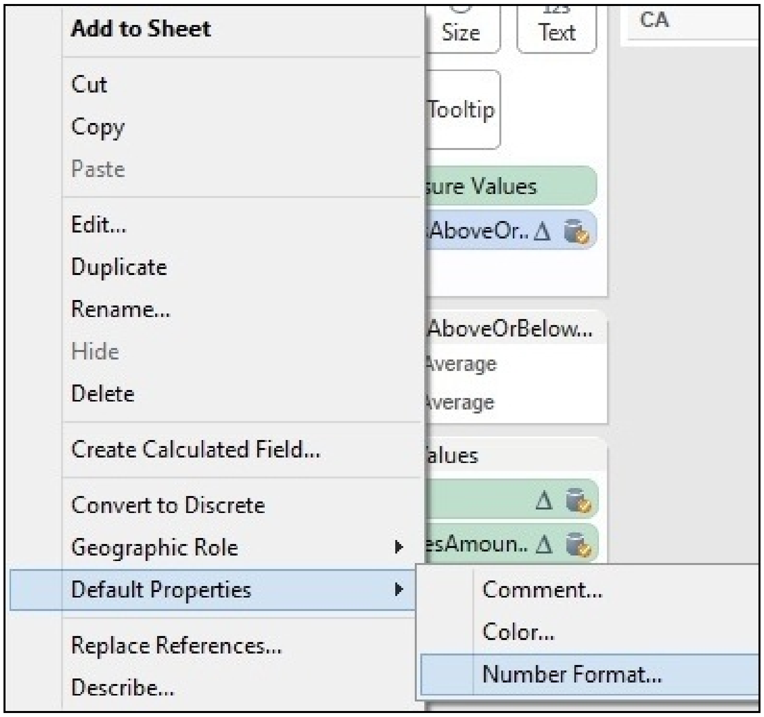
1. Open up the Tableau `Chapter 2` workbook and duplicate the worksheet called **Using Crosstabs Flexibly**. Rename it to `Custom Calculations`.
2. Click on the **Connection_FactInternetSales** data view and duplicate the **SalesAmount** measure by right-clicking on it and selecting **Duplicate**.
3. Rename it to `Rank` and then right-click on it to see a pop-up menu. Select the option called **Create Calculated Field**. You can see an example in the following screenshot:



- Once you have renamed the **Calculated Field** to `Rank`, click on it and choose **Edit**.
- You will see the **Calculated Field** dialog box; an example of this dialog box is shown in the following screenshot:



- To use this calculation as a rank, simply type `Index()` in the **Formula** field.
- Fortunately, Tableau provides a helpful guide to show you whether the formula is correct. Here, we can see that the calculation is valid. Once you have typed `Index()` in, click on **OK** and you are returned to the main Tableau interface.
- Next, change the visualization to a table using the **Show Me** panel.
- Now drag the **Rank** calculated field to the **Columns** shelf. You will see that **Rank** appears as a number with two decimal points. To change the format, simply go back to the **Measures** pane and right-click on **Rank**. Under the heading **Default Properties**, you will see an option specified as **Number Format**. You can see this in the following screenshot:

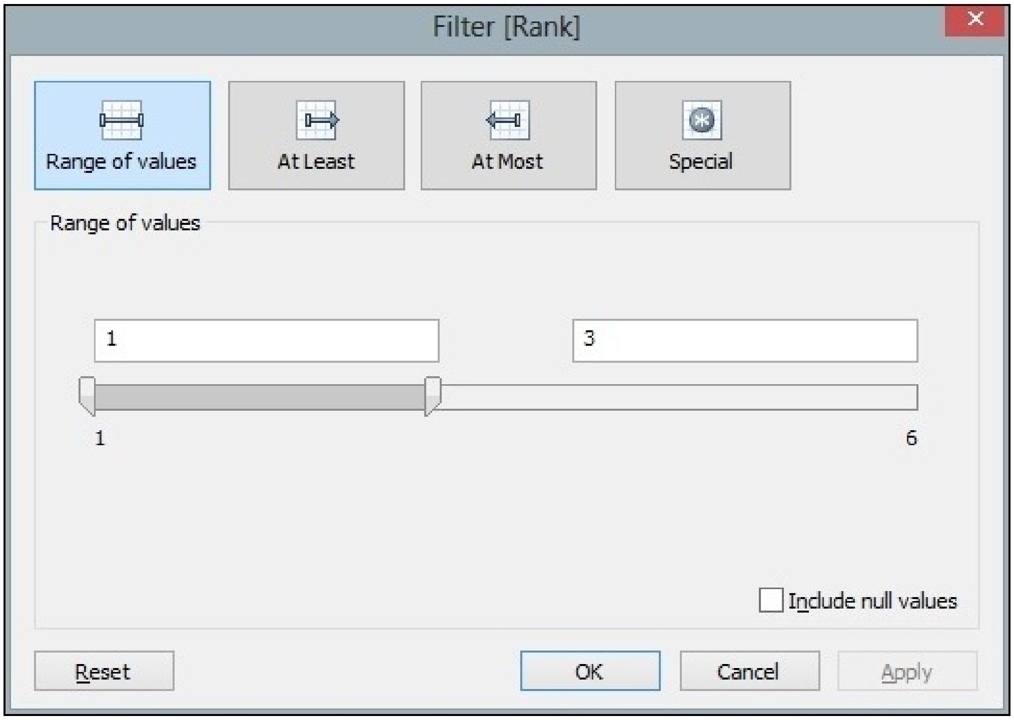


- We will change the number format to an integer. When we select **Number Format**, we get the **Default Number Format** dialog box.
- Select the option **Number (Custom)**, reduce the decimal places down to zero, and then click on **OK**.
- The **Rank** number will appear as an integer, and our data visualization appears as illustrated in the following screenshot:

CountryRegi..	Rank along CountryRegionCode	% of Total SalesAmount along CountryRegionCode
US	1	31.98%
AU	2	30.86%
GB	3	11.55%
DE	4	9.86%
FR	5	9.01%
CA	6	6.74%

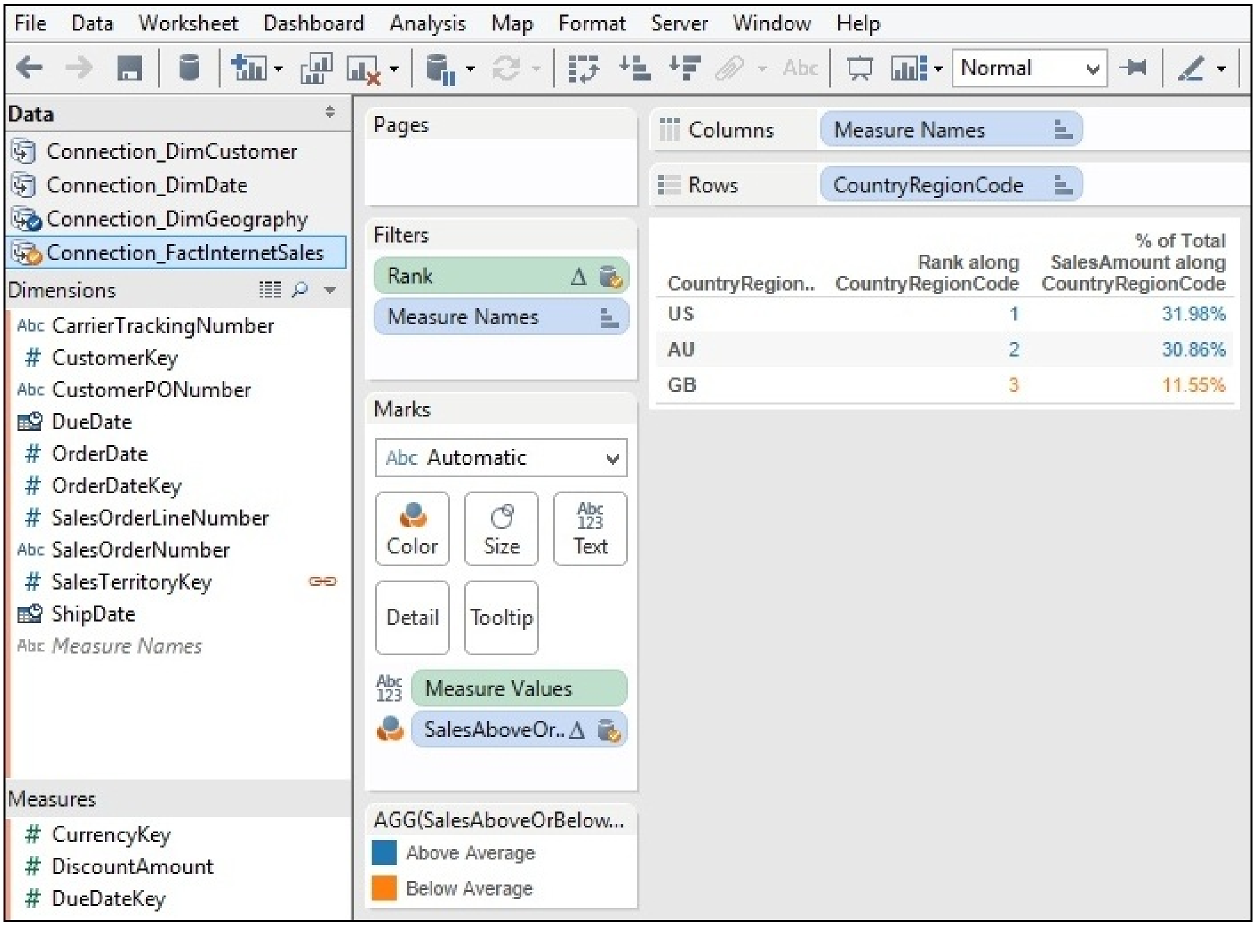
Now, let's take the scenario whereby we want to display only the top three performers by region.

- Simply drag **Rank** from the **Measures** pane through to the **Filters** pane. When you do this, you will get the following dialog box:



Make sure you enter the number 3 on the right-hand side rather than 6. Although there are six options, this filter means that Tableau will only display the top three best-selling regions.

You can see the final data visualization in the following screenshot:



How it works…

Tableau offers us many interesting ways to compact the data down to its minimum design while still helping us to show the message of the data. Although the main visualization is very small, it helps us to use this element in a later dashboard because we have packaged a lot of information into a very small space.

The visualization shows the following:

- Top three performers
- The third performer is actually below average
- The overall percentage of sales as a number in order to provide numerical detail as well as some ways of visualizing the data so it provides "at a glance" information too

To summarize, making data tell a story is a challenge, particularly when there is not much space to play with. However, making data tell a story is fun with Tableau.

There's more…

Often, the column names in the data source are not meaningful to business users. They may need to be translated into something that is useful. Fortunately, Tableau allows you to use aliases to translate dimension names into something else.

Chapter 3. Interacting with Data for Dashboards

In this chapter, we will cover the following topics:

- Fun with filters – grouping your data with clarity
- Hierarchies for revealing the dashboard message
- Classifying your data for dashboards
- Actions and interactions
- Drilling into the details
- Working with input controls

Introduction

A key aspect of dashboarding is that a dashboard should convey its message clearly and simply in order to help team members draw the right conclusions. Dashboards became more interesting to businesses when Kaplan and Norton introduced their *Balanced Scorecard* methodology in the 1990s. This introduced the dashboard as a way of measuring business performance with a particular focus on **Key Performance Indicators** (**KPIs**), which helped to measure the success and direction of the organization. With the Enron scandal in 2001, businesses realized that it was perfectly possible to drown in data and not really understand what is going on at the executive level. Therefore, the dashboard concept gained renewed interest, which continues to date.

Creating dashboards is both a top-down and bottom-up process. It is top-down because we need to be able to summarize and put all of the pieces together. It is also a bottom-up process because the dashboard is made up of its constituent parts. In this chapter, we will look at making the most of the constituent parts, so this chapter uses the bottom-up approach. Later on in this book, we will look at top-down processes while creating the dashboard.

In order to achieve the objective of conveying the message of the data effectively, users should be able to interact with data to get the information that they need. The subject of this chapter is to help you create dashboards that will facilitate team members to get the most out of their dashboards by setting up interactivity, navigation, and an awareness of the underlying data. According to the guidelines, such as *Eight Golden Rules of Interface Design* by Professor Ben Shneiderman, it is vital to allow users to interact with the data by offering the filtering, categorizing, and zooming in functionalities to access the details. Business users gain trust in the data by having a look at the details, which allows them to validate the truth of the data.

Interacting with dashboards is a vital way of allowing business users to understand the data better. It also allows components to tell stories individually as well as provide a coherent story of the data as a whole.

From the practical perspective, it allows us to make the most of the space. Instead of having lots of reports with different dimensions and filters, we can help the user move towards **Self-Service Business Intelligence** (**SSBI**). We do this by furnishing the user with the data that they require in a dashboard format while allowing them to focus on the dimensions and attributes that are most important to them.

Dashboards are different from reports in that users expect to be able to view the data and understand it at a glance. In other words, very little interaction is required as the necessary data should be presented; that said, it is expected that the data is highly integrated and that the various elements of the dashboard are highly coupled together.

This section will help you to see the different ways of facilitating interactions with the data on the dashboard while getting the message of the data across as quickly and effectively as possible. Currently, we are in a bottom-up part of dashboard creation, and not top-down. In Tableau, we create worksheets that then go into the dashboards. This is why we will initially focus on worksheets.

Fun with filters – grouping your data with clarity

Filters are a useful way of helping users focus on particular aspects of the data that they are most interested in. This helps them to investigate and compare data and perhaps look for outliers and exceptions in the data.

Filtering data is an essential part of a dashboard. Users like to interact with data in order to understand it better, and it is natural to filter data so that users can pinpoint the data that particularly interests them.

Tableau allows users to filter measures, calculations, and dimensions, which is extremely useful in a dashboard. For example, take the case where you need to see the sales figures that are less than the given amount over a period of time.

In this recipe, we will import some more data and look at taking some descriptive statistics. We will also look at filters. Users have a lot of flexibility when it comes to combining filters, which means that you can have a lot of creativity in your analyses.

Getting ready

For this recipe, we will need to have Tableau open and ready to create a new workbook. For the exercises in this chapter, we will import multiple tables into our new workbook as the basis for creating calculations.

Let's create a new folder to store the Tableau workbook. For example, in this book, we will use the folder `Chapter Three` under `D:\Data\TableauCookbook\`. Therefore, you will need to have the folder open where you will download the data files. In Chapter 1, *A Short Dash to Dashboarding!*, we specified `D:\Data\TableauCookbook` as an example. We will be importing the following files:

- DimProductCategory
- DimProductSubCategory
- DimProduct
- FactInternetSales
- DimDate

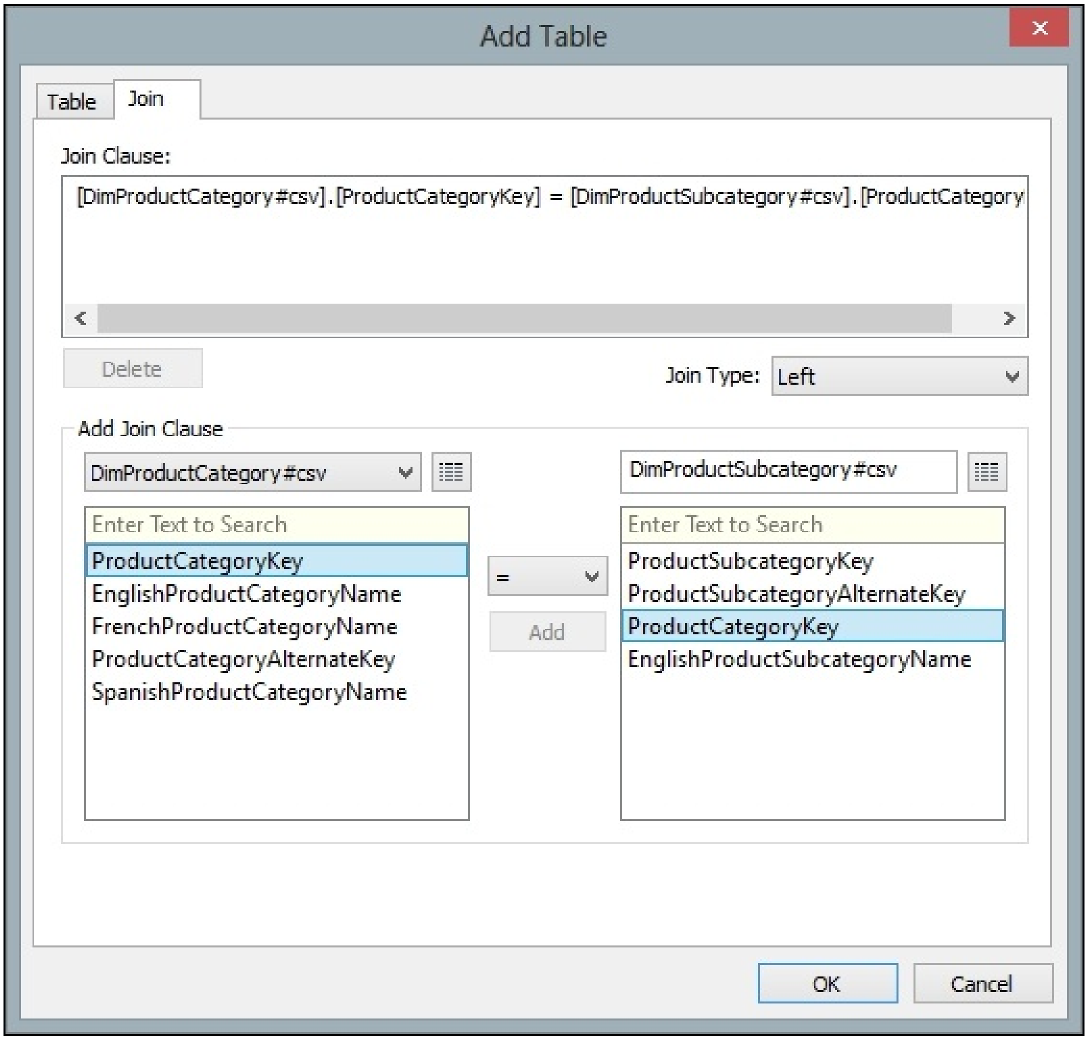
How to do it...

1. Open up Tableau and navigate to **File | New**. This will start a new workbook in Tableau.
2. Go to **File** and then click on **Save As**.
3. Save the file as Chapter Three.
4. We will connect to the data and import it into Tableau's data store. To do this, select **Connect to Data**, which you will see under **Data**.
5. Navigate to the location where you stored the CSV files. The following table shows the links between the tables:

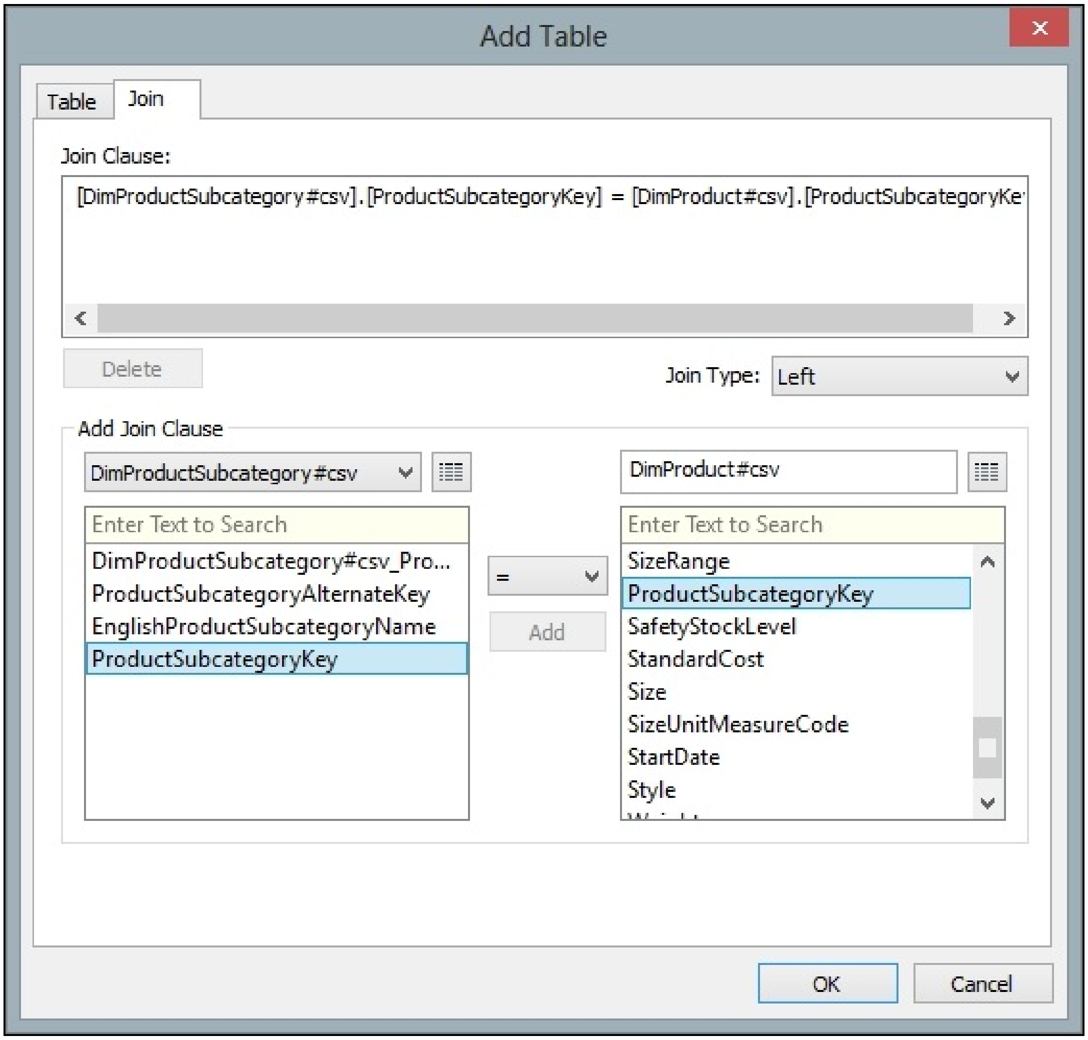
Table name	Linked To	Join
DimProductCategory	DimProductSubCategory	ProductCategoryKey
DimProductSubCategory	DimProduct	Product
DimProduct		
FactInternetSales		
DimDate		

- Next, select **Text File** from the options of **Connect to Data**. You will see it in the section labeled **In a file**.

- The first file we will import is called `DimProductCategory`, so select it and click on **Open**.
- The **Text File Connection** dialog will appear. In **Step 1**, you will see the file path and the name of the selected `DimProductCategory` file.
- In **Step 2**, select the option **The first row has field names in it**.
- In **Step 3**, select the **Multiple Tables** option. You will see that the **Text File Connection** dialog will change, and you will see the name of the `DimProductCategory` file appear in the **Table Alias** part of **Step 3**.
- We will be joining multiple tables, so click on the **Add Table…** button. The **Add Table** dialog box will appear.
- In the **Table** tab, select `DimProductSubCategory`. You will see the fields for this table appear in a box under the **Table Alias** heading.
- In the **Join** tab, you will see that the `DimProductCategory` table and the `DimProductSubCategory` table are joined together via `ProductCategoryKey`.
- The value of **Join Type** should be set to **Left**.
- In the following screenshot, you can see the join between the `DimProductCategory` and `DimProductSubCategory` tables:

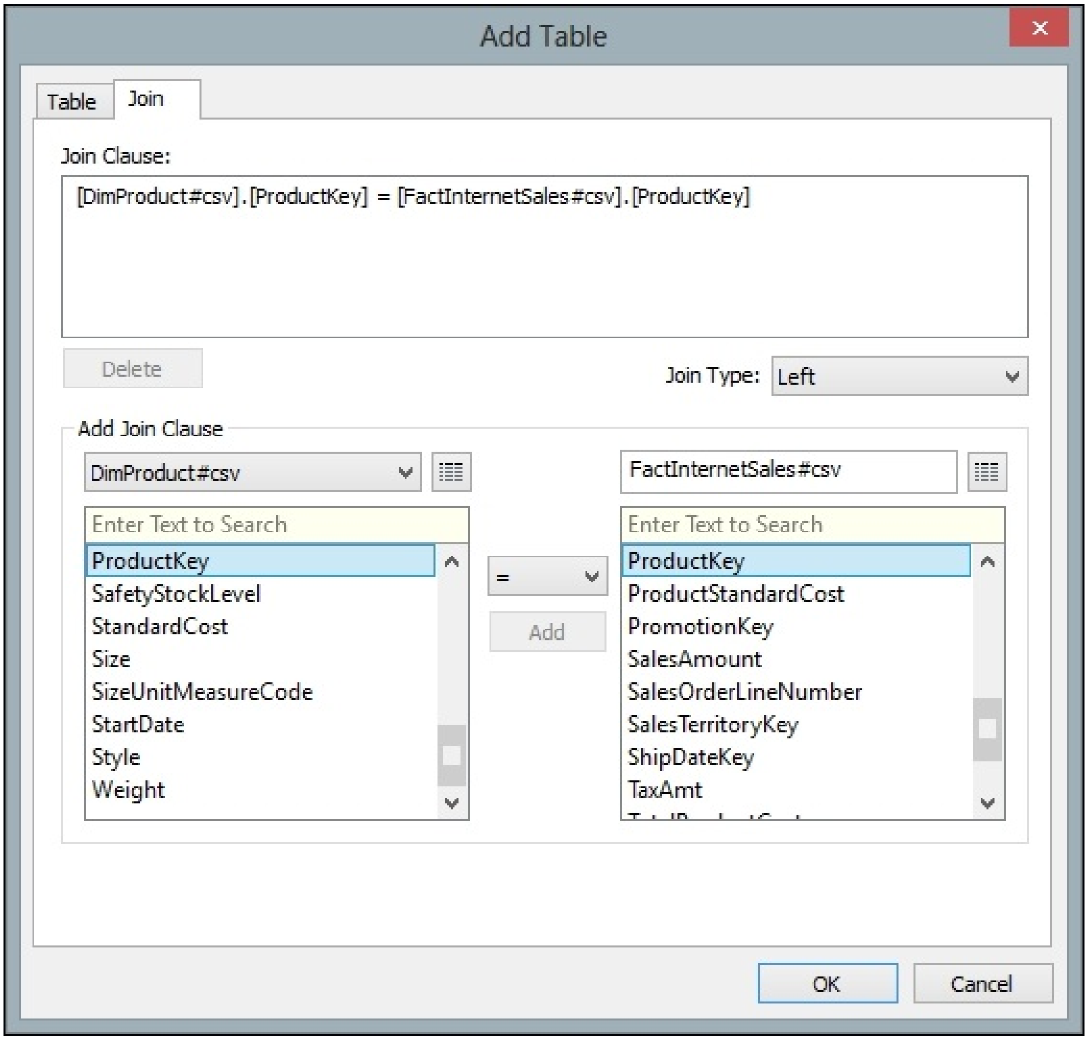


- Click on **OK**. This will take you back to the **Text File Connection** dialog box.
- The next step is to join the `DimProductSubCategory` and `DimProduct` tables together. To do this, click on the **Add Table…** button again.
- In the **Table** tab, click on `DimProduct#csv`.
- In the **Join** tab, you will see that the `DimProductSubCategory` and `DimProduct` tables are joined together by `ProductSubcategoryKey`. The join type should be **Left**.
- In the following screenshot, you can see the join between the `DimProductSubCategory` and `DimProduct` tables:



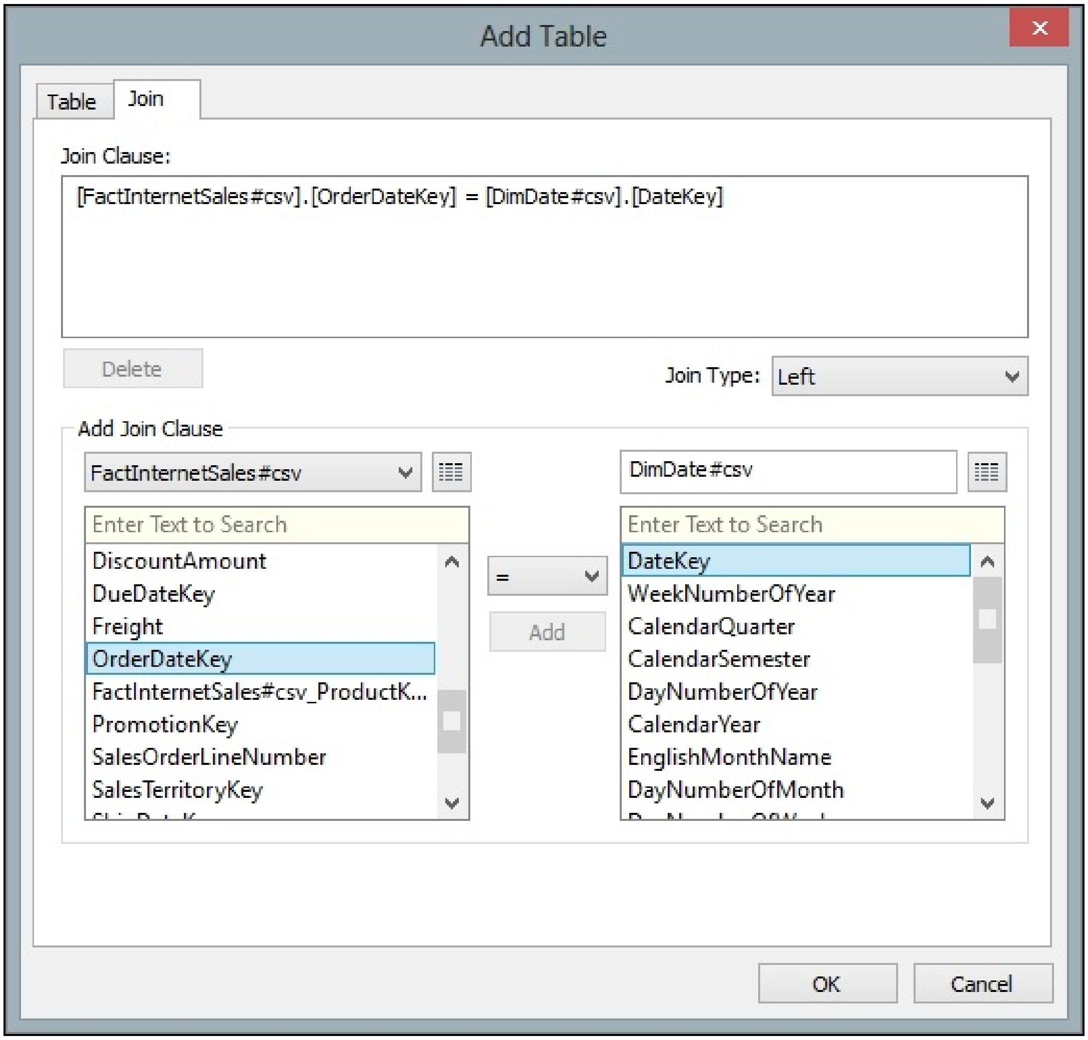
- Click on **OK**. You will be taken back to the **Text File Connection** dialog box.
- Let's now join the `DimProduct` and `FactInternetSales` tables together. In the **Text File Connection** dialog, click on the **Add Table…** button again, which will bring up the **Add Table** dialog box.
- In the **Table** tab, look for the file named `FactInternetSales#csv`, which contains the data for the `FactInternetSales` table, and select it. You will see the fields for this table appear in a box under the **Table Alias** heading.
- In the **Join** tab, you will see that the `DimProductCategory` table and the `FactInternetSales` table are joined together via `ProductKey`.
- The join type should be set to **Left**. Click on **OK**, and you will be taken to the **Text File Connection** dialog box.
- In the following screenshot, you can see the join between the

`DimProduct` and `FactInternetSales` tables:



- Finally, let's join the `DimDate` and `FactInternetSales` tables together. In the **Text File Connection** dialog box, click on the **Add Table…** button again, which will bring up the **Add Table** dialog box.
- In the **Table** tab, look for the file named `DimDate#csv`, which contains the data for the `DimDate` table, and select it.
- You will see the fields for this table appear in a box under the **Table Alias** heading in the **Add Table** dialog box.
- In the **Join** tab, you will see that the `DimDate` and `FactInternetSales` tables are joined together via `DateKey`.
- The join type should be set to **Left**. Click on the **Add** button in the **Add Join Clause** section, which you will see in the middle of the screen.
- Click on **OK**, and you will be taken to the **Text File Connection** dialog box.

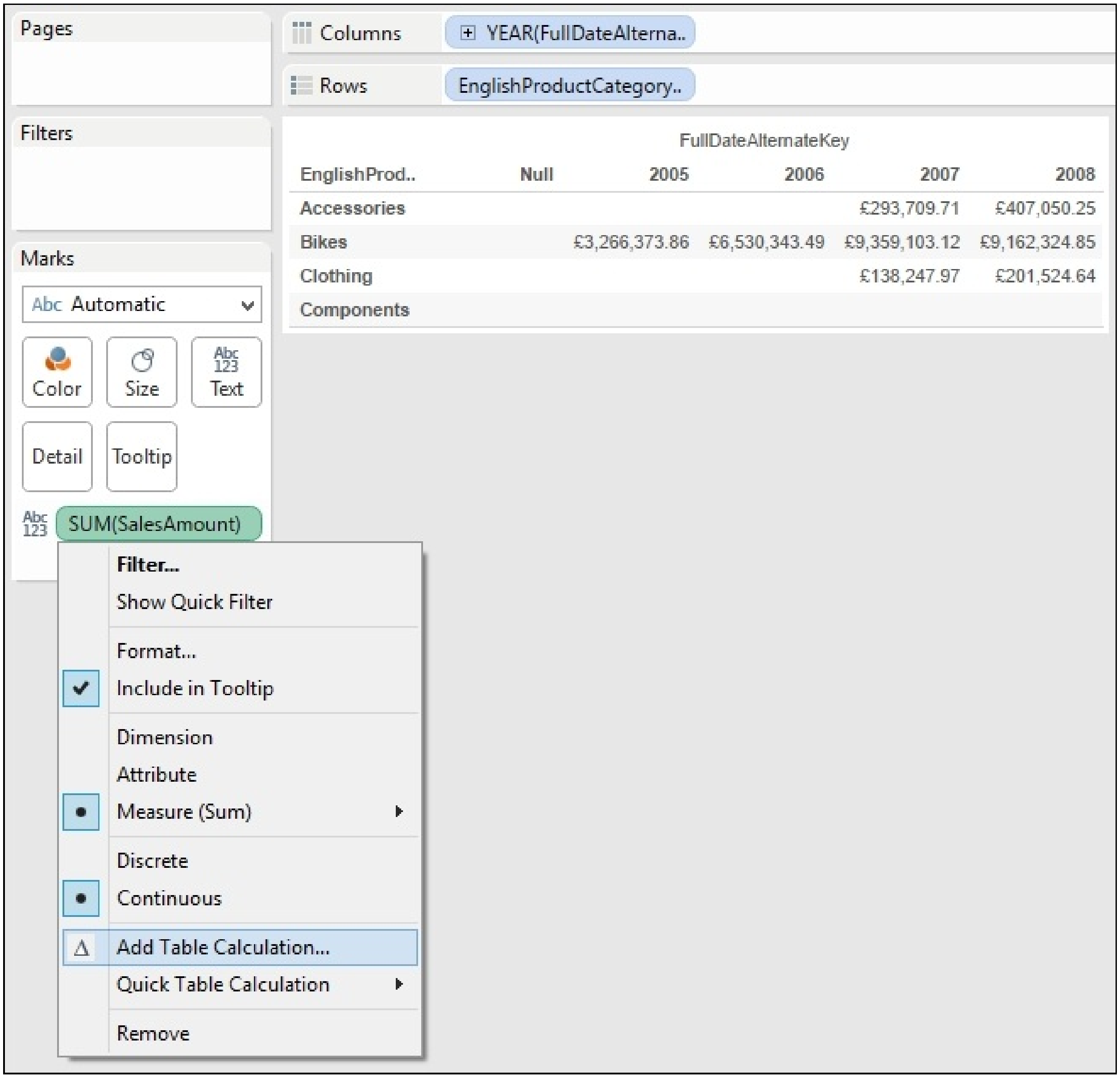
- In the next screenshot, you can see the join between the `DimProduct` and `FactInternetSales` tables.



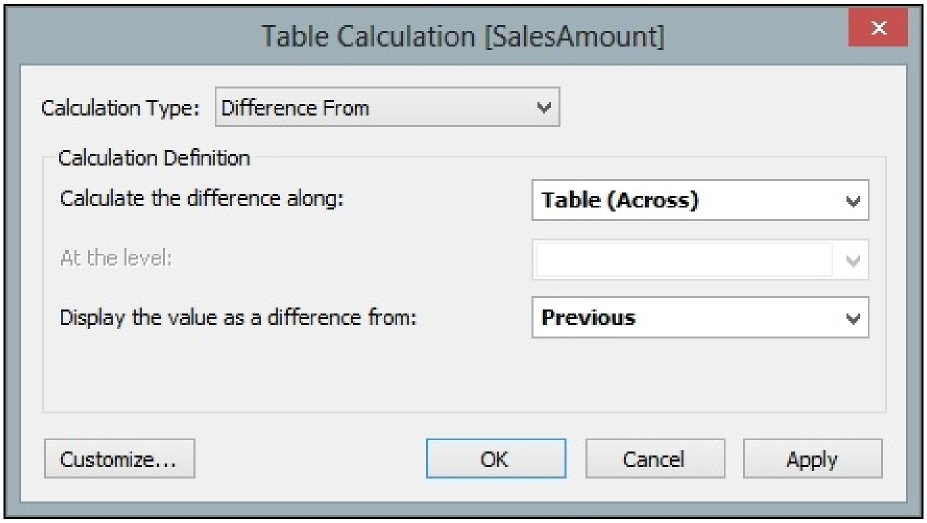
- Click on **OK**. When prompted, select the **Import all data** option.
- Once all of the tables have been imported, you will be taken to the Tableau canvas.
- On the **Data** pane in the Tableau side bar, let's rename the data source to something more meaningful. Right-click on the data source under the **Data** pane. By default, Tableau will have given it a name that combines the first table and then the file type. Here, it will be called `DimProductCategory #csv`. Right-click on it and select **Rename**.
- Enter `CombinedProductsWithFacts` so we know that this data source is a combination of facts and products.
- We can look at putting in dimensions and metrics in order to make a

start and be productive straightaway with Tableau.

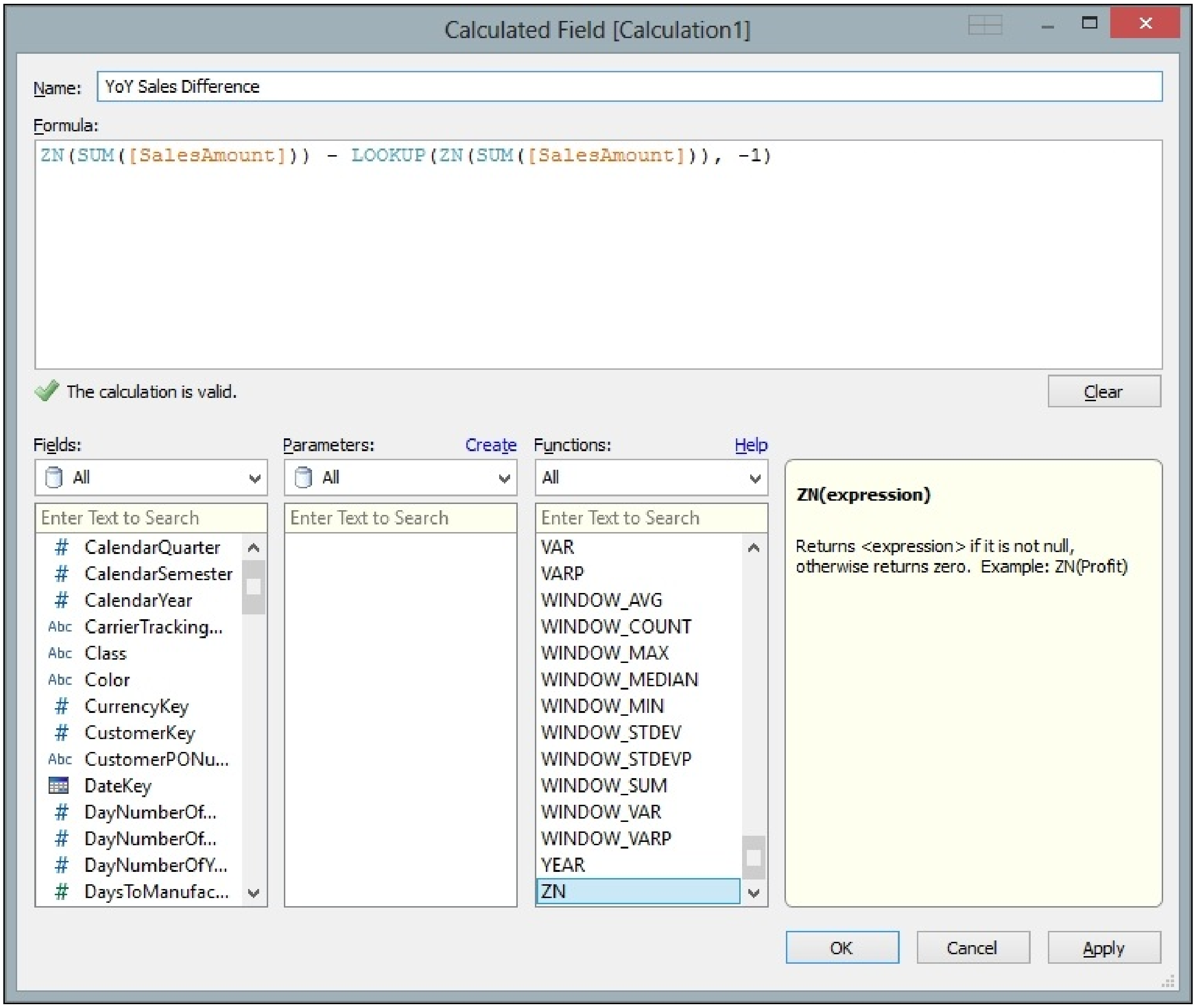
- Tableau places the **DateKey** field into **Measures**. The next step is to drag **DateKey** from **Measures** to **Dimensions** in order to be able to set up some of the relationships.
- Next, we will visualize a table as the starting point. Take the **FullDateAlternateKey** field from the **DimDate** table and drag it onto the **Columns** shelf. Tableau will automatically recognize that this is a date, and it will aggregate the data according to the year level. Therefore, it will appear as **Year(FullDateAlternateKey)**.
- Next, take the **EnglishProductCategory** attribute from the **DimProductCategory** table and place it on the **Rows** shelf.
- Drag **SalesAmount** to the canvas.
- Then, we will add a few table calculations as an exercise to explore this concept more while also adding to our filters in this exercise.
- On the **Marks** shelf, right-click on the **Sum(Sales Amount)** metric and select the option **Add Table Calculation**, as shown in the following screenshot:



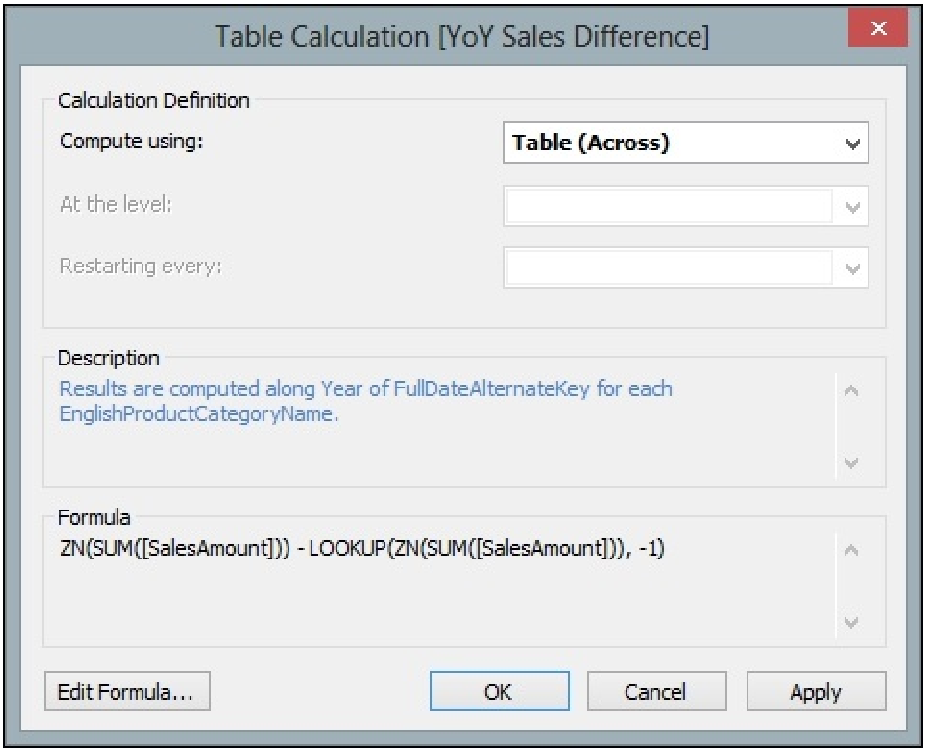
- When we right-click on the **Sum(Sales Amount)** metric, we get the **Table Calculation** window, as shown in the following screenshot. For our purposes, we want to choose **Difference From** as the value for **Calculation Type**.



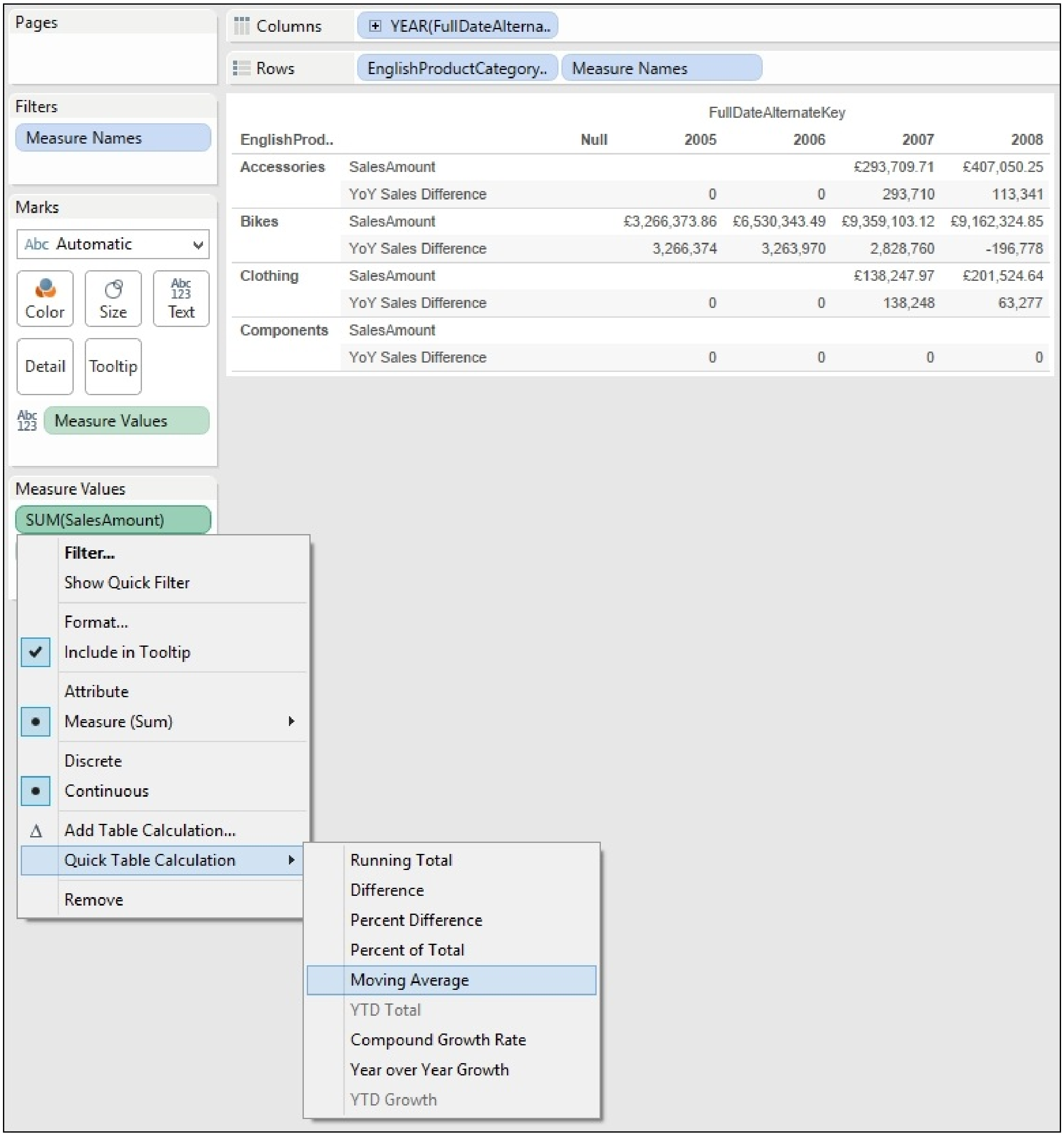
- We will calculate the difference along the table, so we will choose to calculate the difference along with the **Table(Across)** option.
- In the **Calculation Definition** panel, we will choose the **Previous** option under the **Display the value as a difference from:** dropdown.
- Once these options have been selected, we can customize the table calculation further by renaming the calculation to something meaningful. To do this, click on the **Customize** button, which can be found at the bottom-left corner of the **Table Calculation** box.
- After we have clicked on the **Customize** button, we will get the **Calculated Field** dialog, which you can see in the next screenshot. The text button at the top is labeled **Name:** and we can insert a different name in this textbox.
- Here, we will rename the table calculation to `YoY Sales Difference`. The formula itself works out the current sales amount and compares it to the previous sales amount. If a null value is found, for example, where there is no previous sales amount available because we are looking at the data for the first year, then a zero is returned; this is the job of the **ZN** expression. Once you have renamed the table calculation, click on **OK**.



- You are then returned to the previous window, and you will see a description of the formula in the **Description** window. You will also see the formula in the **Formula** box. When you reach this point, click on **OK**, as shown in the following screenshot:



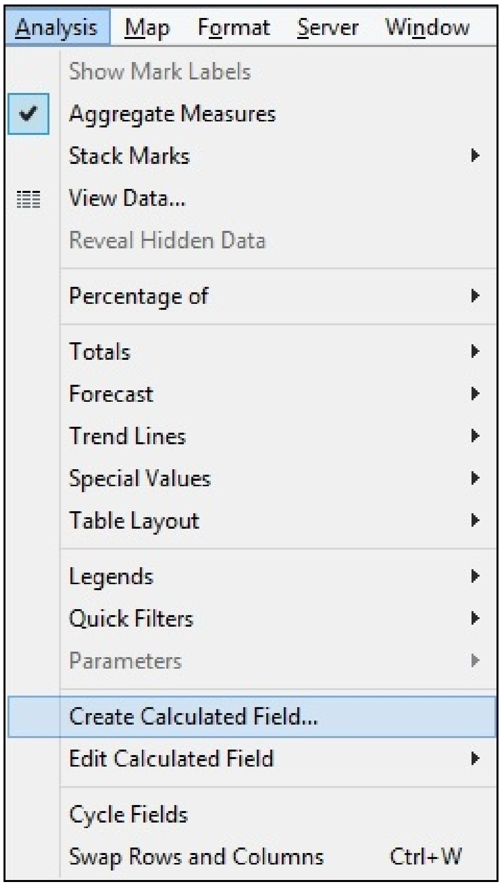
- Drag **Sum(Sales Amount)** from the **Measures** part of the side bar over to the **Marks** shelf.
- Drag **Sum(Sales Amount)** from the **Measures** part of the sidebar over to the canvas so that it appears in the table along with **YoY Sales Difference**.
- Right-click on the **Sum(Sales Amount)** measure on the **Marks** shelf and choose the option **Quick Table Calculation** from the menu list. Then, select the **Moving Average** option. You can see this in the following screenshot:



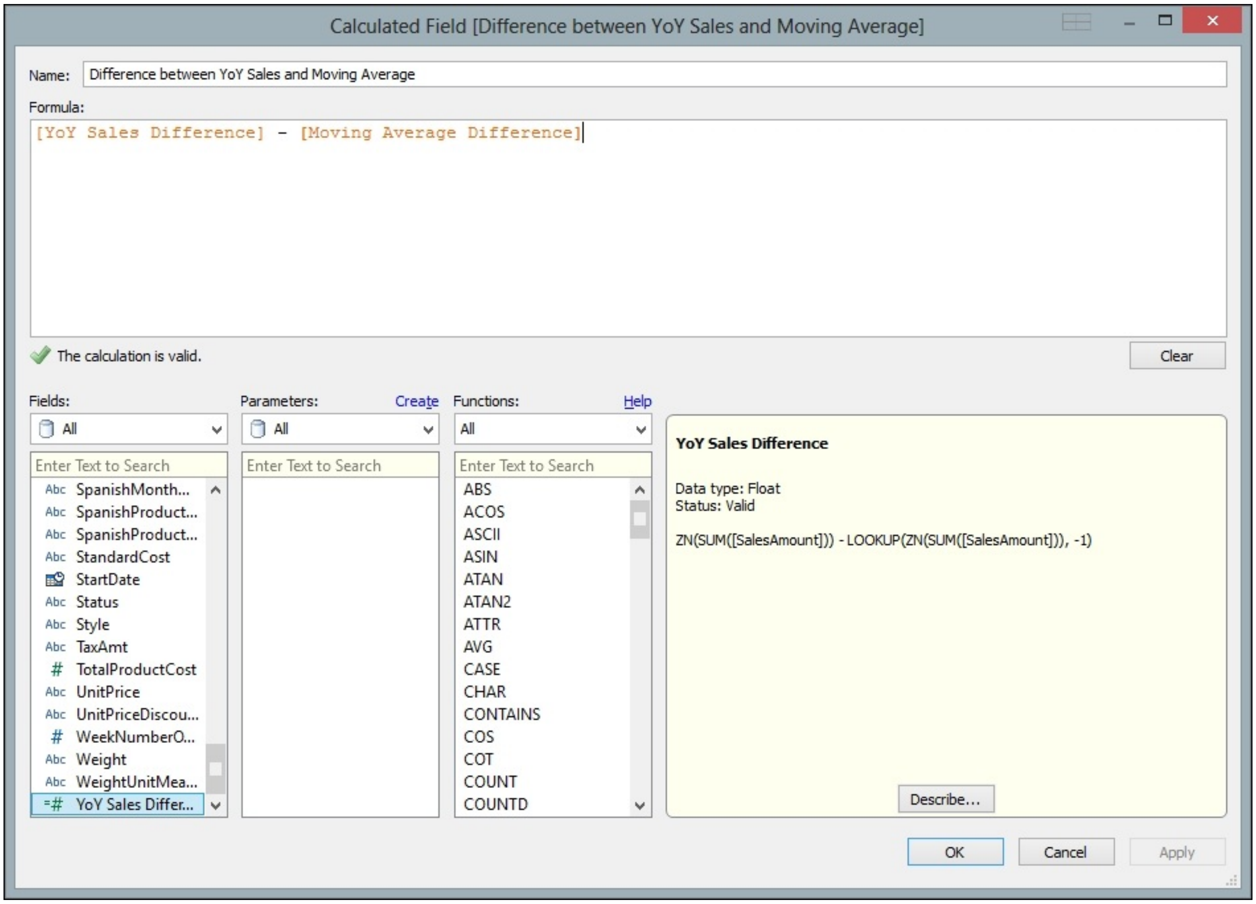
- This will create a new measure that shows the moving average. To rename the new calculated measure, right-click on the **SUM(SalesAmount)** measure in the **Measure Values** shelf and choose the **Edit Table Calculation** option.
- In the **Table Calculation [Moving Average of SalesAmount]** dialog box, select the **Customize…** option.
- In the **Name:** box, rename it to `Moving Average` and click on **OK**.
- To summarize, we will now have two measures: one for **Year on Year**

changes and another for the **Moving Average Difference** over time.

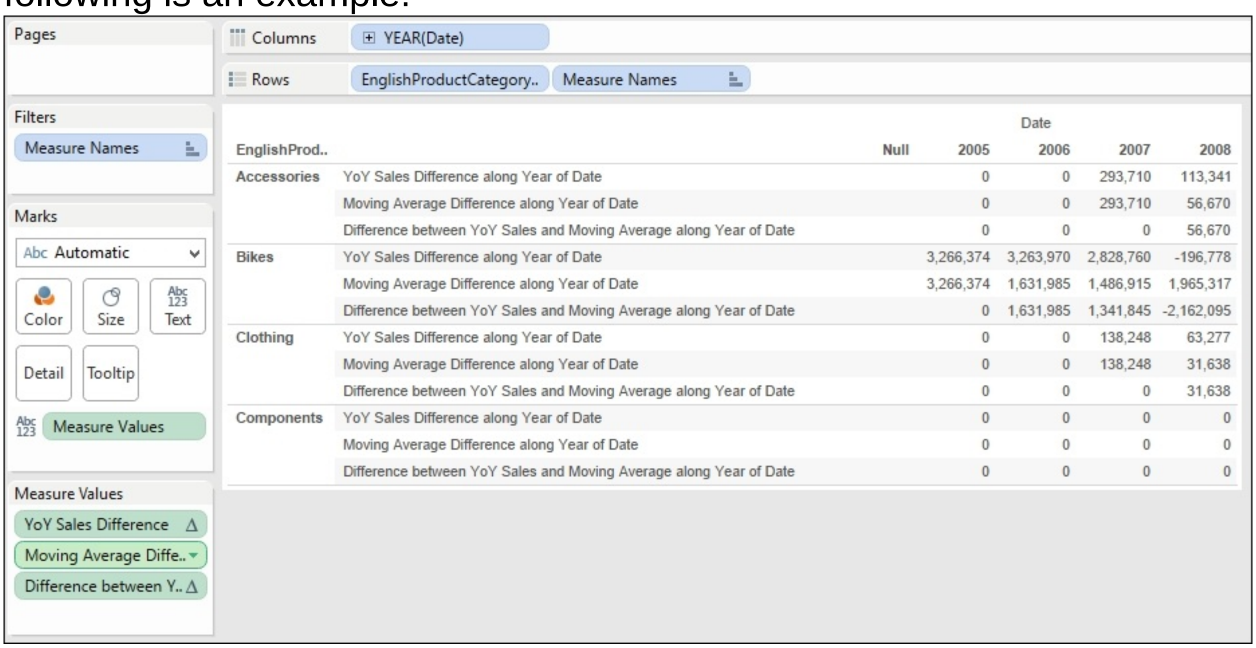
- Our next step is to work out the difference between the two calculations that we have just made. In other words, what is the difference between the year-on-year change and the moving average for the sales amount?
- Our first step in this process is to create a new calculated field that will work out the difference between the year-on-year change and the moving average. To do this, firstly we will need to go to the **Analysis** tab at the top menu item and select the **Create Calculated Field** option. You can see this in the menu in the following screenshot:



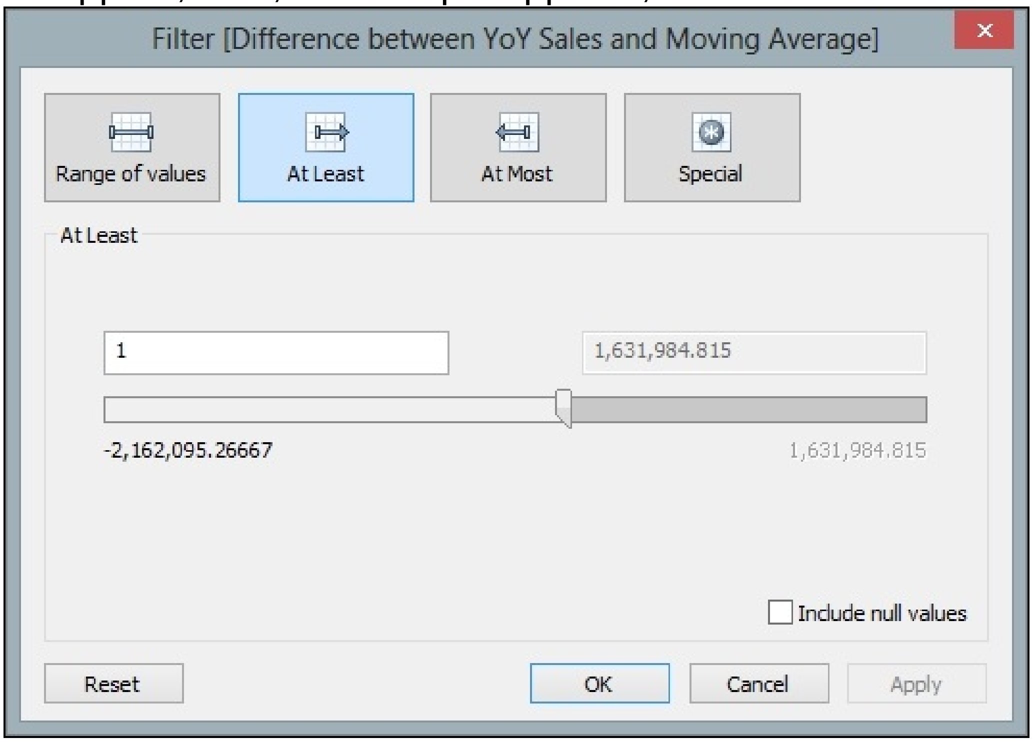
- We will now get the **Calculated Field** editor box, and we need to subtract **Moving Average** from the **Year on Year** average. We can see this in the following illustration:



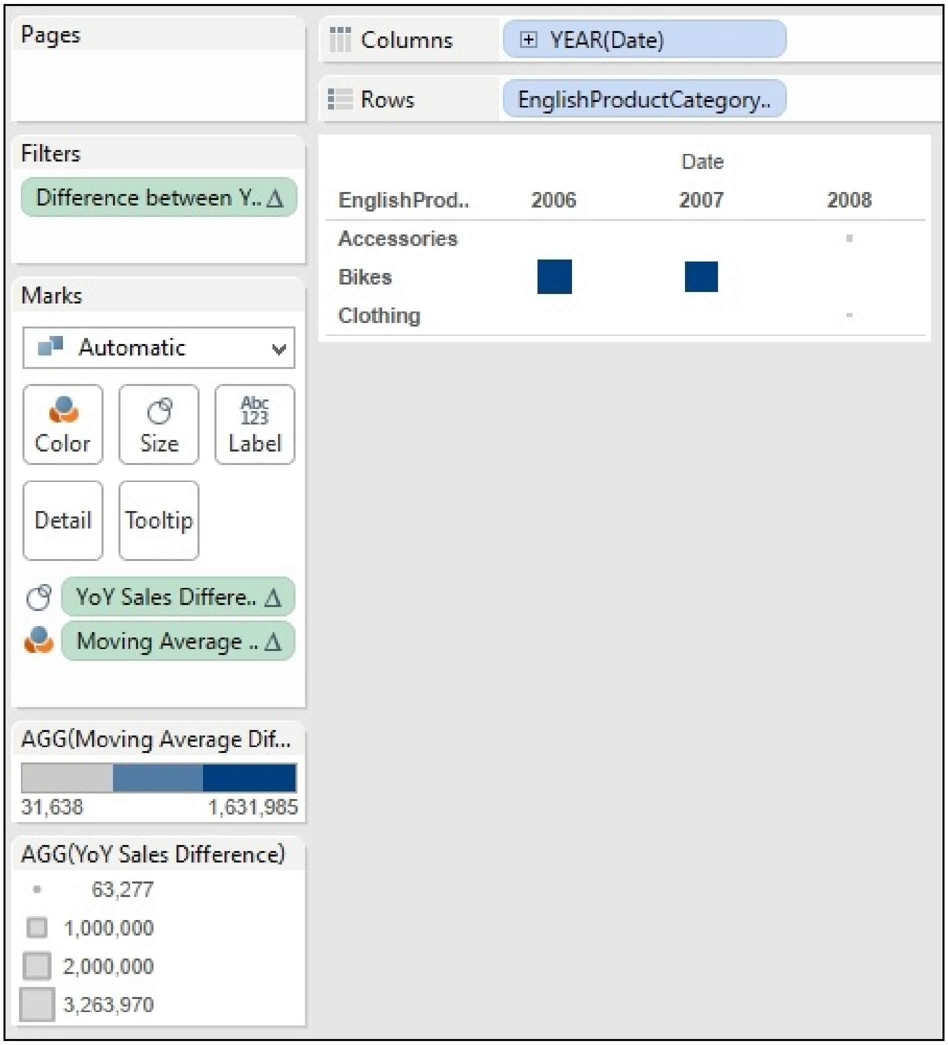
- When we place all the three calculations on the table, it looks a little confusing with a lot of numbers, and it is hard to differentiate the difference between patterns and outliers contained in the data. The following is an example:



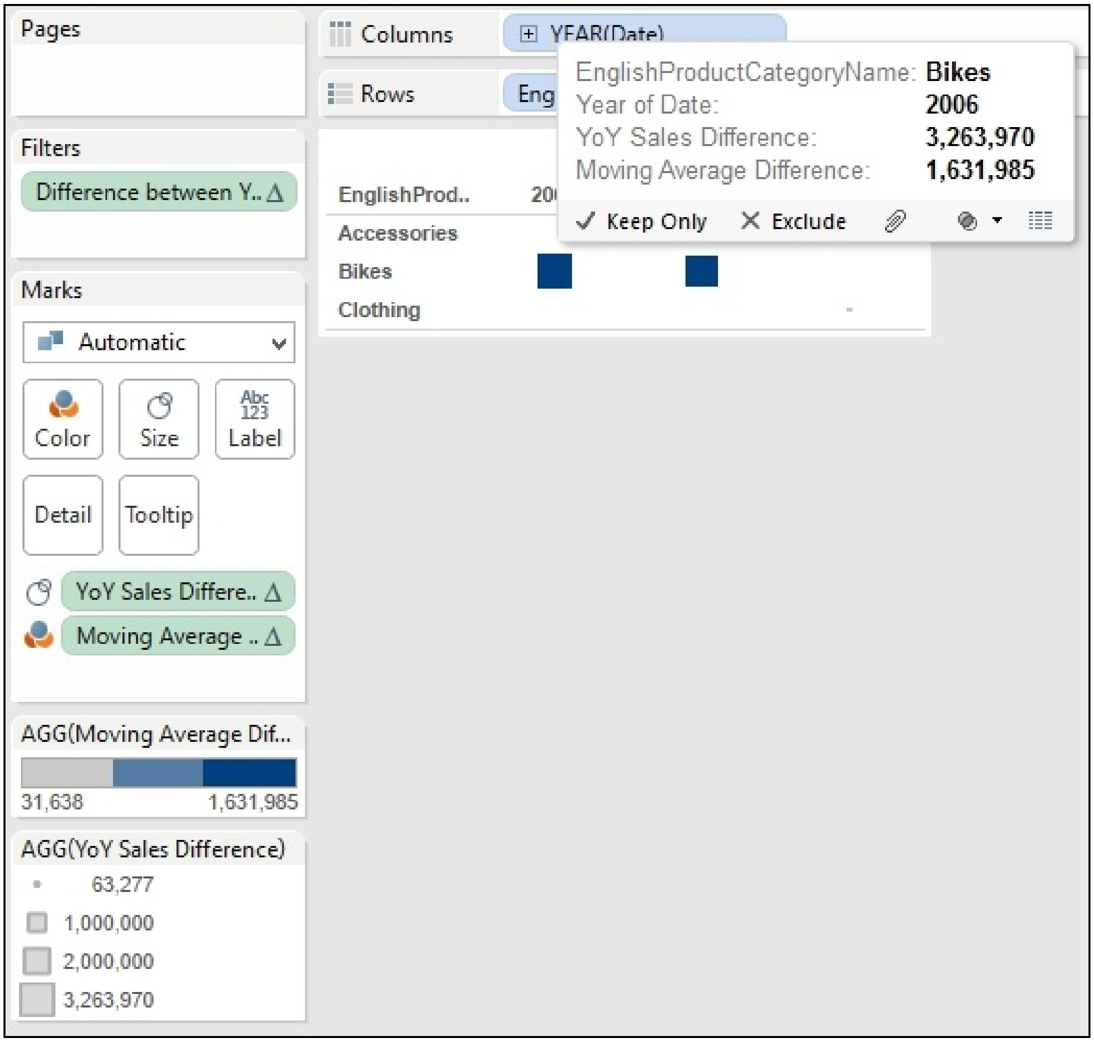
EnglishProd..		Null	2005	2006	2007	2008
Accessories	YoY Sales Difference along Year of Date		0	0	293,710	113,341
	Moving Average Difference along Year of Date		0	0	293,710	56,670
	Difference between YoY Sales and Moving Average along Year of Date		0	0	0	56,670
Bikes	YoY Sales Difference along Year of Date		3,266,374	3,263,970	2,828,760	-196,778
	Moving Average Difference along Year of Date		3,266,374	1,631,985	1,486,915	1,965,317
	Difference between YoY Sales and Moving Average along Year of Date		0	1,631,985	1,341,845	-2,162,095
Clothing	YoY Sales Difference along Year of Date		0	0	138,248	63,277
	Moving Average Difference along Year of Date		0	0	138,248	31,638
	Difference between YoY Sales and Moving Average along Year of Date		0	0	0	31,638
Components	YoY Sales Difference along Year of Date		0	0	0	0
	Moving Average Difference along Year of Date		0	0	0	0
	Difference between YoY Sales and Moving Average along Year of Date		0	0	0	0

- We can use our measures in order to filter the data, and it's very simple to do this. Drag the measure **Difference** between **YoY Sales** and **Moving Average** to the **Filters** shelf on the left-hand side. A **Filter** window will appear; then, an example appears, as shown in the following window: 

- In this example, we have chosen **At Least** as the filter option, so the Tableau worksheet will only show rows where the difference between **YoY Sales** and **Moving Average** is at least 1. This means that negative numbers will not be shown along with rows where there is no difference.
- If we change the visualization of the chart to a heat map, the largest differences are clearly seen in the years 2006 and 2007.



- If we click on the data points, we can obtain more details of specific values. Tooltips are an extremely useful way of providing additional details *on demand* without cluttering the real estate on the screen for people who do not need it.



- To summarize, in this section, we have shown that we can use table calculations and measures in order to filter data to show the information that we would like to see on the dashboard. This helps to provide the "at a glance" purpose of a dashboard.

How it works...

Tableau matches are based on column names initially. It also makes a decision on whether a column is a dimension or a measure, based on the data type. So, in this example, the `OrderDateKey` field was inserted as a measure, and we needed to drag it to the **Dimension** area of the **Data** pane. Tableau does this by default because the `DateKey` field is an integer. If we leave it as a measure, it will not show up as a column available to define the join.

See also

- *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, *Ben Shneiderman*, *Cath Plaisant*, *Maxine Cohen*, *Steven Jacobs*, *Prentice Hall*

Hierarchies for revealing the dashboard message

It can become difficult to manage data, particularly if you have many columns. It can become more difficult if they are similarly named too. As you'd expect, Tableau helps you to organize your data so that it is easier to navigate and keep track of everything.

From the user perspective, hierarchies improve navigation and use by allowing the users to navigate from a headline down to a detailed level. From the Tableau perspective, hierarchies are groups of columns that are arranged in increasing levels of granularity. Each deeper level of the hierarchy refers to more specific details of the data.

Some hierarchies are natural hierarchies, such as date. So, say Tableau works out that a column is a date and automatically adds in a hierarchy in this order: year, quarter, month, week, and date. You have seen this already, for example, when you dragged a date across to the **Columns** shelf, Tableau automatically turned the date into a year.

Some hierarchies are not always immediately visible. These hierarchies would need to be set up, and we will look at setting up a product hierarchy that straddles across different tables. This is a nice feature because it means that the hierarchy can reflect the users' understanding of the data and isn't determined only by the underlying data.

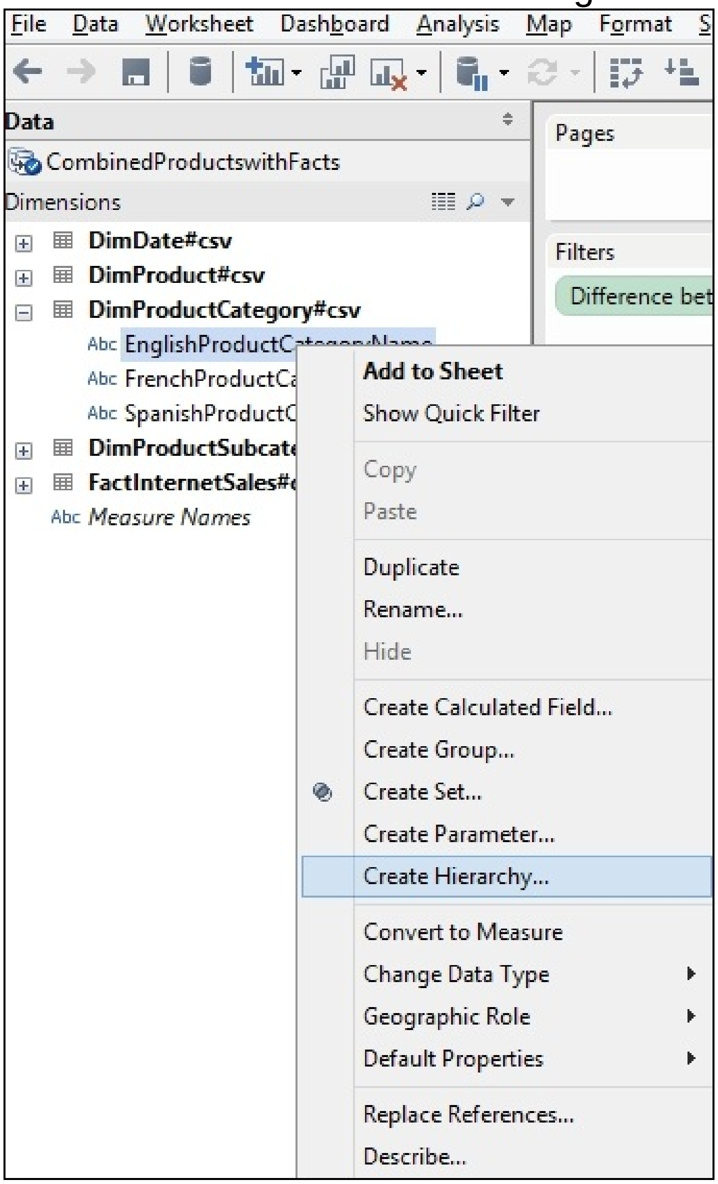
Getting ready

In this recipe, we will use the existing workbook that you created for this chapter.

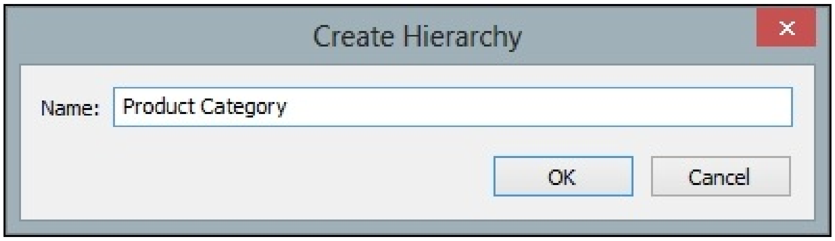
We will use the same data. For this recipe, let's take a copy of the existing worksheet and call it `Hierarchies`. To do this, right-click on the **Worksheet** tab and select the **Duplicate Sheet** option. You can then rename the sheet to `Hierarchies`.

How to do it...

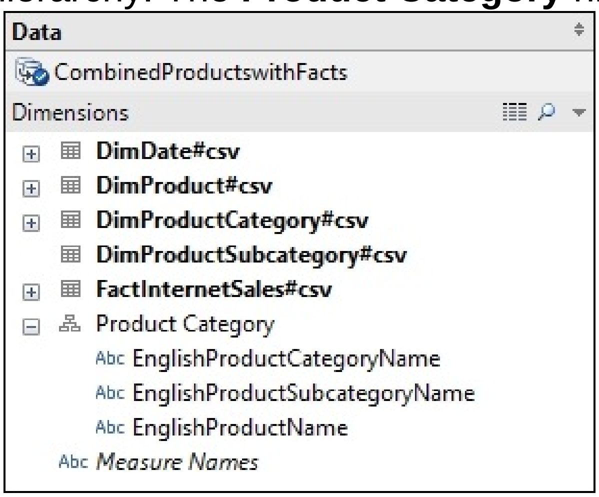
1. Navigate to the `DimProductCategory` dimension and right-click on the **EnglishProductCategoryName** attribute.
2. From the pop-up menu, select the **Create Hierarchy** feature. You can see its location in the following illustration:



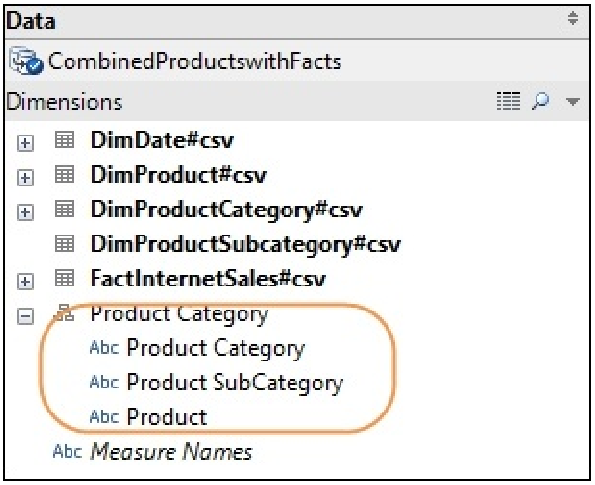
- When you select the option, you will get a textbox entitled **Create Hierarchy**, which will ask you to specify the name of the hierarchy.
- We will call our hierarchy `Product Category`. Once you have entered this into the textbox, click on **OK**.



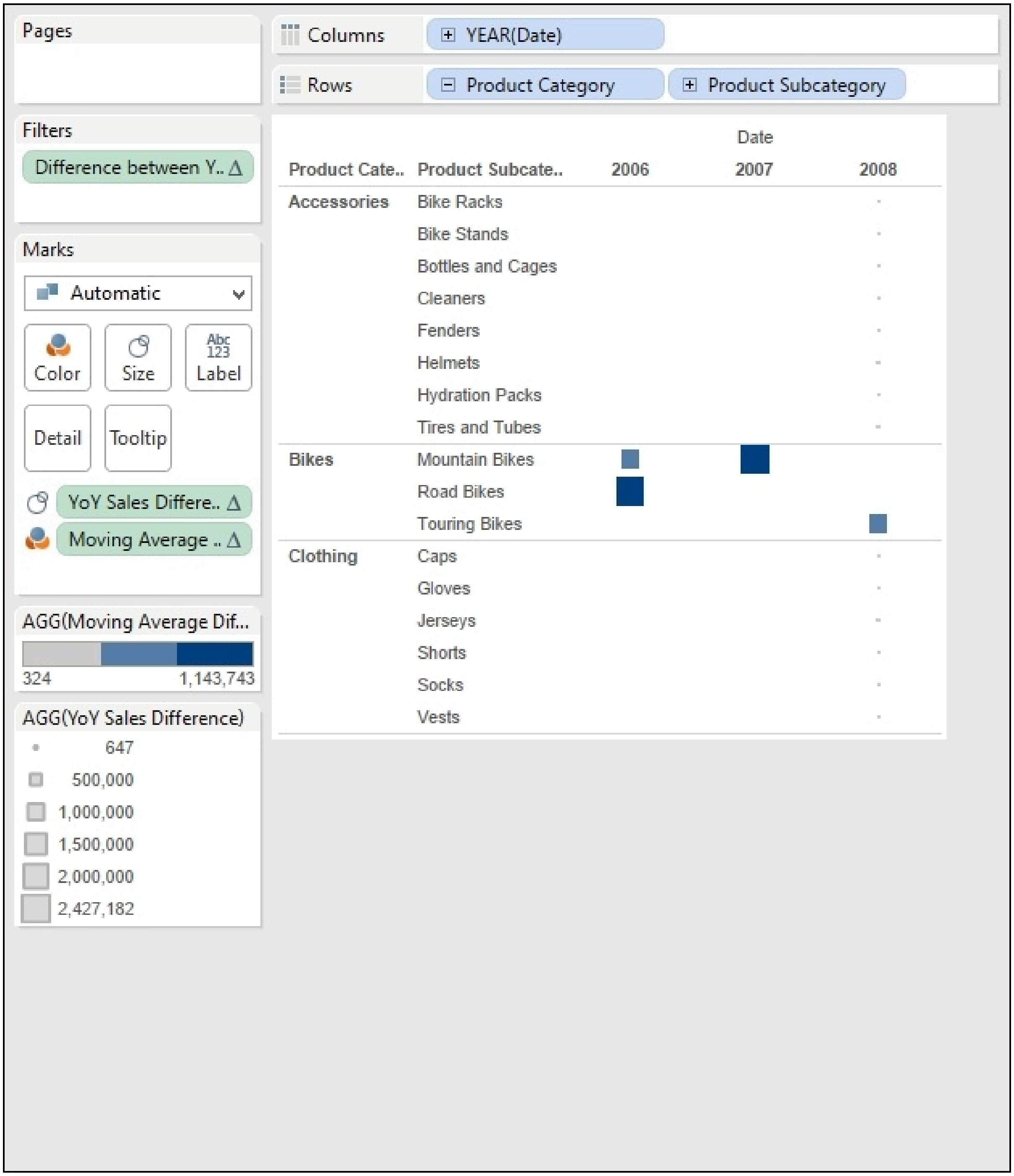
- Your hierarchy will now be created, and it will appear at the bottom of the **Dimensions** list on the left-hand side of Tableau's interface.
- Next, go to the **DimProductSubcategory** dimension and look for the **EnglishProductSubCategoryName** attribute. Drag it to the **Product Category** hierarchy under **EnglishProductCategoryName**, which is already part of the **Product Category** hierarchy.
- Now we will add the **EnglishProductName** attribute, which we will find under the **DimProduct** dimension. Drag-and-drop it under the **EnglishProductSubCategoryName** attribute that is already under the **Product Category** hierarchy. The **Product Category** hierarchy should now look as follows:



- The **Product Category** hierarchy will be easier to understand if we rename the attributes. To do this, right-click on each attribute and choose **Rename**. Change **EnglishProductCategoryName** to `Product Category`.
- Rename **EnglishProductSubcategoryName** to `Product Subcategory` by right-clicking on the attribute and selecting **Rename**.
- Rename **EnglishProductName** to `Product`.
- Once you have done this, the hierarchy should look as follows:



- You can now use your hierarchy to change the details that you wish to see in the data visualization. Now, we will use **Product Category** of our data visualization rather than **Dimension**.
- Remove everything from the **Rows** shelf and drag the **Product Category** hierarchy to the **Rows** shelf. Then, click on the plus sign; it will open the hierarchy, and you will see data for the next level under **Product Category**, which are subcategories.
- An example of the Tableau workbook is given in the following illustration. You can see that the biggest differences occurred in the **Bikes** product category, and they occurred in the years 2006 and 2007 for the **Mountain Bikes** and **Road Bikes** categories.



- To summarize, we have used the **Hierarchy** feature in Tableau to vary the degree of analysis we see in the dashboard.

How it works…

Tableau saves the additional information as part of the Tableau workbook. When you share the workbook, the hierarchies will be preserved.

The Tableau workbook would need revisions if the hierarchy is changed, or if you add in new dimensions and they need to be maintained. Therefore, they may need some additional maintenance. However, they are very useful features and worth the little extra touch they offer in order to help the dashboard user.

There's more...

Dashboarding data usually involves providing "at a glance" information for team members to clearly see the issues in the data and to make actionable decisions. Often, we don't need to provide further information unless we are asked for it, and it is a very useful feature that will help us answer more detailed questions. It saves us space on the page and is a very useful dashboard feature.

Let's take the example of a business meeting where the CEO wants to know more about the biggest differences or "swings" in the sales amount by category, and then wants more details. The Tableau analyst can quickly place a hierarchy in order to answer more detailed questions if required, and this is done quite simply as described here. Hierarchies also allow us to encapsulate business rules into the dashboard. In this recipe, we used product hierarchies. We could also add in hierarchies for different calendars, for example, in order to reflect different reporting periods. This will allow the dashboard to be easily reused in order to reflect different reporting calendars, say, you want to show data according to a fiscal year or a calendar year. You could have two different hierarchies: one for fiscal and the other for the calendar year. The dashboard could contain the same measures but sliced by different calendars according to user requirements.

The hierarchies feature fits nicely with the Golden Mantra of Information Visualization, since it allows us to summarize the data and then drill down into it as the next step.

See also

- http://www.tableausoftware.com/about/blog/2013/4/lets-talk-about-sets-23043

Classifying your data for dashboards

Bins are a simple way of categorizing and bucketing values, depending on the measure value. So, for example, you could "bin" customers depending on their age group or the number of cars that they own. Bins are useful for dashboards because they offer a summary view of the data, which is essential for the "at a glance" function of dashboards.

Tableau can create bins automatically, or we can also set up bins manually using calculated fields. This recipe will show both versions in order to meet the business needs.

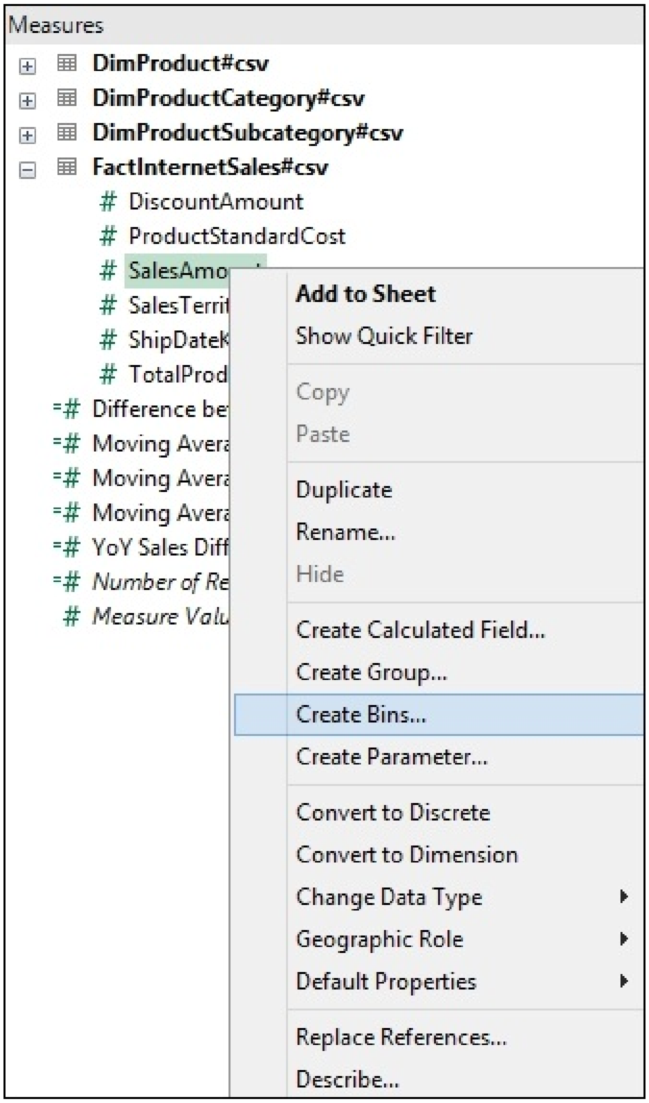
Getting ready

In this recipe, we will use the existing workbook that you created for this chapter.

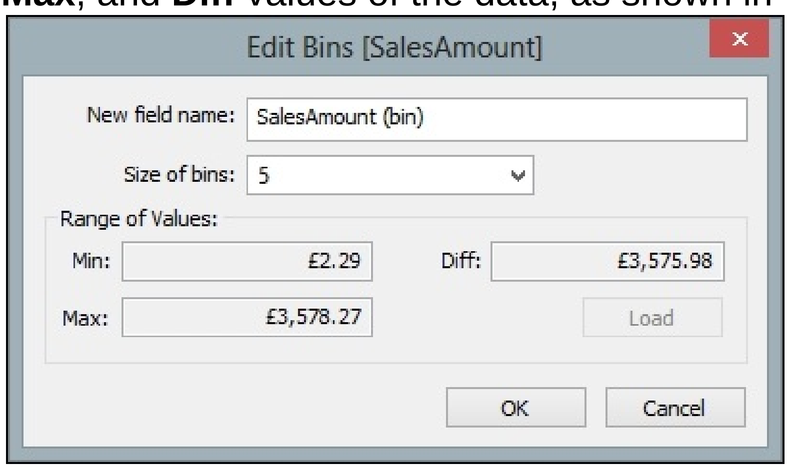
We will use the same data. For this recipe, let's take a copy of the **Hierarchies** worksheet and by right-clicking on the **Worksheet** tab, select the **Duplicate Sheet** option. You can then rename the sheet to `Bins`.

How to do it...

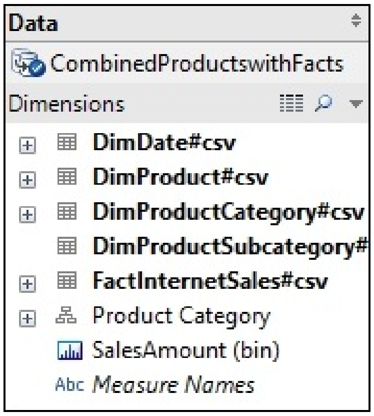
1. Once you have your **Bins** worksheet in place, right-click on the **SalesAmount** measure and select the **Create Bin** option. You can see an example of this in the following screenshot:



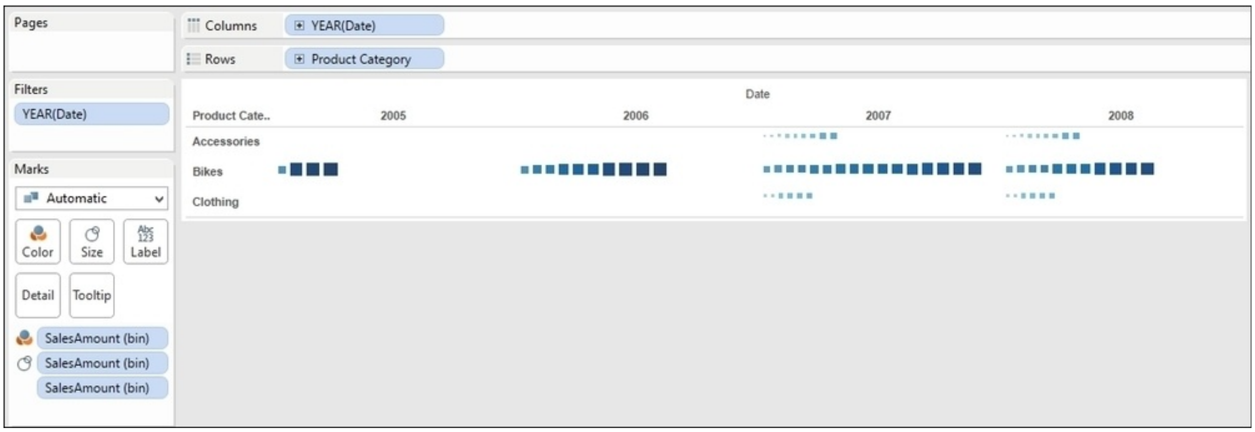
- We will change the value to 5. Once you've done this, press the **Load** button to reveal the **Min**, **Max**, and **Diff** values of the data, as shown in the following screenshot:



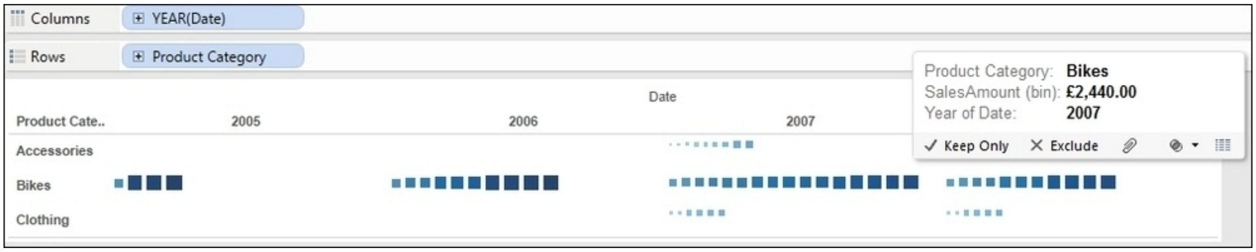
- When you click on the **OK** button, you will see a bin appear under the **Dimensions** area. The following is an example of this:



- Let's test out our bins! To do this, remove everything from the **Rows** shelf, leaving only the **Product Category** hierarchy. Remove any filters from the worksheet and all of the calculations in the **Marks** shelf.
- Next, drag **SalesAmount (bin)** to the **Marks** area under the **Detail** and **Tooltip** buttons. Once again, take **SalesAmount (bin)** and drag it to the **Color** button on the **Marks** shelf. Now, we will change the size of the data points to reflect the size of the elements. To do this, drag **SalesAmount (bin)** to the **Size** button.
- You can vary the overall size of the elements by right-clicking on the **Size** button and moving the slider horizontally so that you can get your preferred size.
- To neaten the image, right-click on the **Date** column heading and select **Hide Field Names for Columns** from the list.
- The Tableau worksheet should now look as follows:



- This allows us to see some patterns in the data. We can also see more details if we click on the data points; you can see an illustration of the details in the data in the following screenshot:



- However, we might find that the automated bins are not very clear to business users. We can see in the previous screenshot that the **SalesAmount(bin)** value is **£2,440.00**. This may not be meaningful to business users.

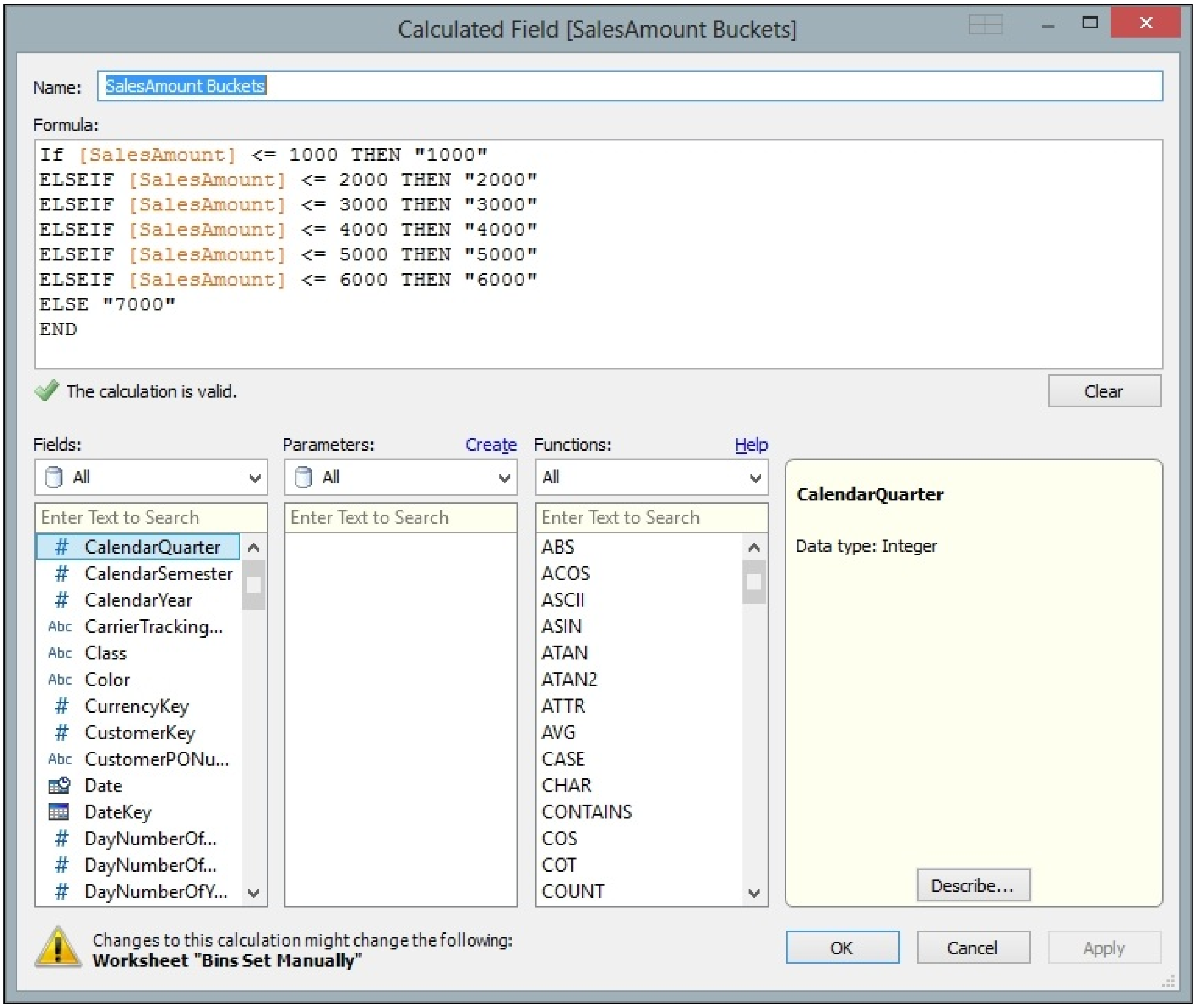
How can we set the bins so that they are meaningful to business users, rather than being automated by Tableau? For example, what if the business team wants to know about the proportion of their sales that fell into well-defined buckets, sliced by years?

Fortunately, we can emulate the same behavior as in bins by simply using a calculated field. We can create a very simple `IF… THEN... ELSEIF` formula that will place the sales amounts into buckets, depending on the value of the sales amount. These buckets are manually defined using a calculated field, and we will see how to do this now.

- Before we begin, take a copy of the existing worksheet called **Bins** and rename it to `Bins Set Manually`.
- To do this, right-click on the **Sales Amount** metric and choose the **Create Calculated Field** option.
- In the calculated field, enter the following formula:

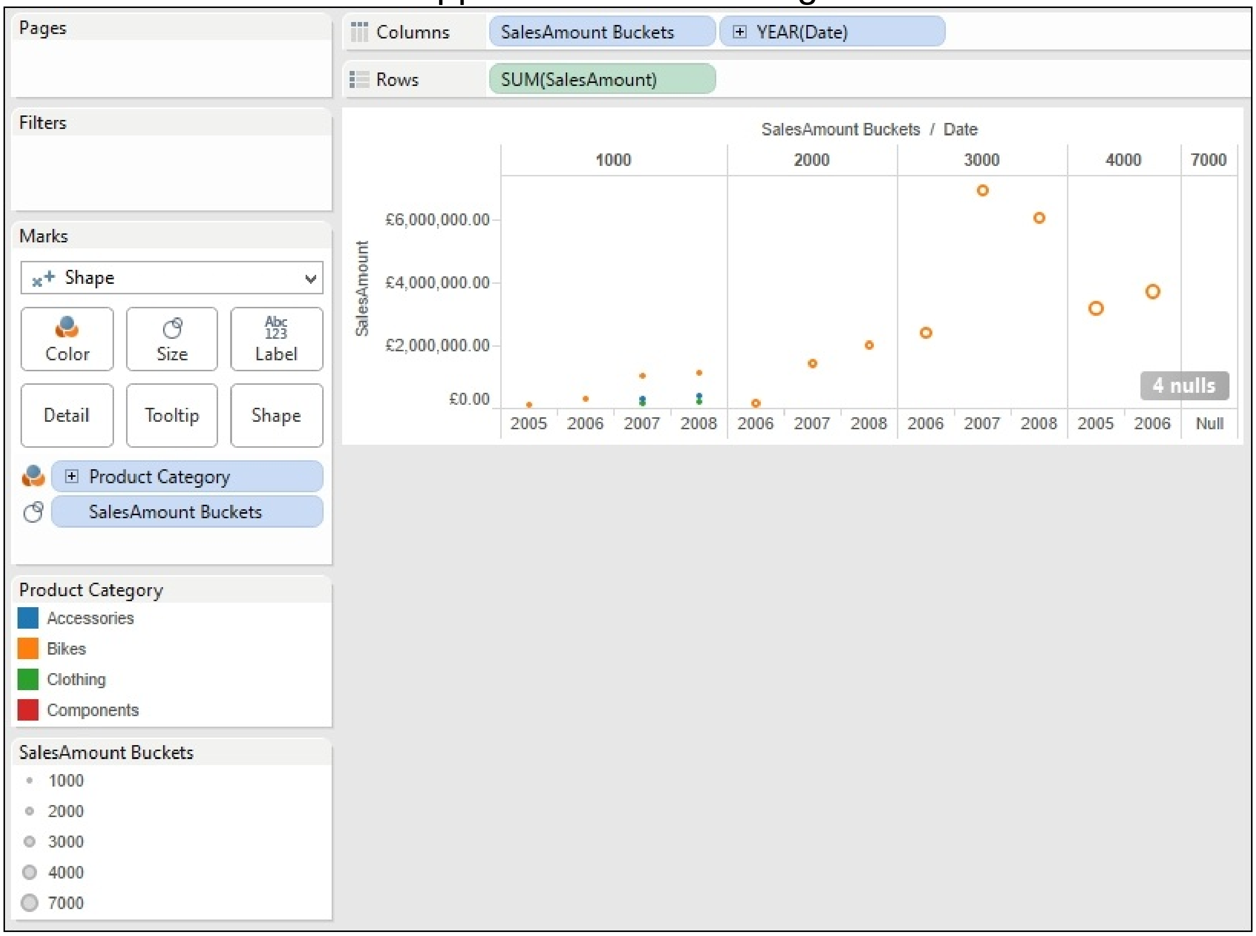
```
If [SalesAmount] <= 1000 THEN "1000"
ELSEIF [SalesAmount] <= 2000 THEN "2000"
ELSEIF [SalesAmount] <= 3000 THEN "3000"
ELSEIF [SalesAmount] <= 4000 THEN "4000"
ELSEIF [SalesAmount] <= 5000 THEN "5000"
ELSEIF [SalesAmount] <= 6000 THEN "6000"
ELSE "7000"
END
```

- When this formula is entered into the **Calculated Field** window, it looks like what the following screenshot shows. Rename the calculated field to **SalesAmount Buckets**.



- Now that we have our calculated field in place, we can use it in our Tableau worksheet to create a dashboard component.
- On the **Columns** shelf, place the **SalesAmount Buckets** calculated field and the **Year(Date)** dimension attribute.
- On the **Rows** shelf, place **Sum(SalesAmount)** from the **Measures** section.
- Place the **Product Category** hierarchy on the **Color** button.
- Drag **SalesAmount Buckets** from the **Dimensions** pane to the **Size** button on the **Marks** shelf.
- Go to the **Show Me** panel and select the **Circle View** option. This will provide a **dot plot** feel to data visualization. You can resize the chart by hovering the mouse over the foot of the *y* axis where the £0.00 value is located.

- Once you're done with this, drag-and-drop the activities. The Tableau worksheet will look as it appears in the following screenshot:



To summarize, we have created bins using Tableau's automatic bin feature. We have also looked at ways of manually creating bins using the **Calculated Field** feature.

How it works...

Bins are constructed using a default **Bins** feature in Tableau, and we can use **Calculated Fields** in order to make them more useful and complex. They are stored in the Tableau workbook, so you will be able to preserve your work if you send it to someone else.

In this recipe, we have also looked at dot plot visualization, which is a very simple way of representing data that does not use a lot of "ink". The data/ink ratio is useful to simplify a data visualization in order to get the message of the data across very clearly. Dot plots might be considered old fashioned, but they are very effective and are perhaps underused. We can see from the screenshot that the **3000** bucket contained the highest number of sales amount. We can also see that this figure peaks in the year 2007 and then falls in 2008. This is a dashboard element that could be used as a start for further analysis. For example, business users will want to know the reason for the fall in sales for the highest occurring "bin".

See also

- *Visual Display of Quantitative Information*, *Edward Tufte*, *Graphics Press USA*

Actions and interactions

We can make the dashboard more effective by highlighting certain aspects of data visualization. Basically, when the user hovers over the data point, it will highlight the column and row where the data point is found. Highlighting the data means that other irrelevant data points are grayed out, thereby emphasizing the relevant data points. This is a useful dashboarding tool because the relevant features are made more prominent, thereby enhancing the speed with which the data is understood.

We can create highlights using the **Actions** feature in Tableau. To create a highlight action, use the following options:

- For workbooks, we can find an **Actions** option under the **Worksheet** menu item.
- When we move towards creating a full dashboard, we can find dashboard actions under the **Dashboard** menu item. For now, we are looking at creating components that will go onto a dashboard, so we will stick with the worksheet feature for now.

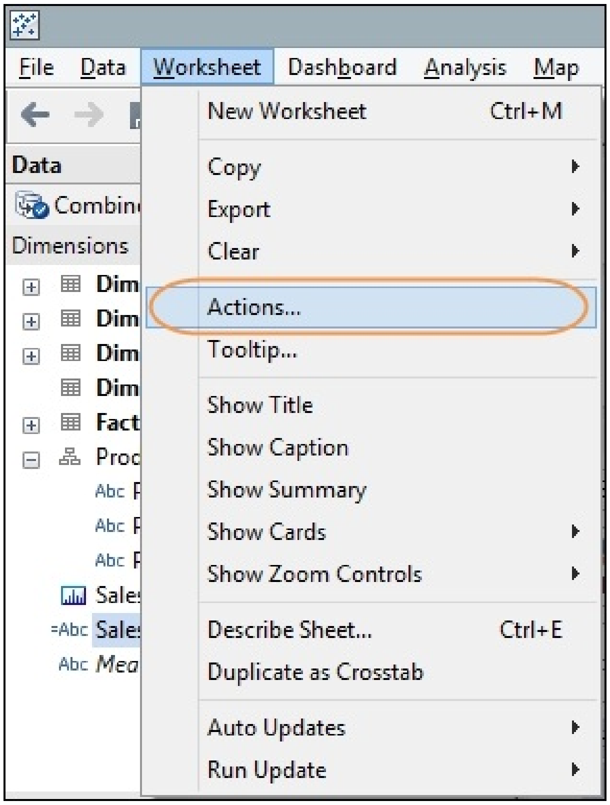
Getting ready

In this recipe, we will use the existing workbook that you created for this chapter.

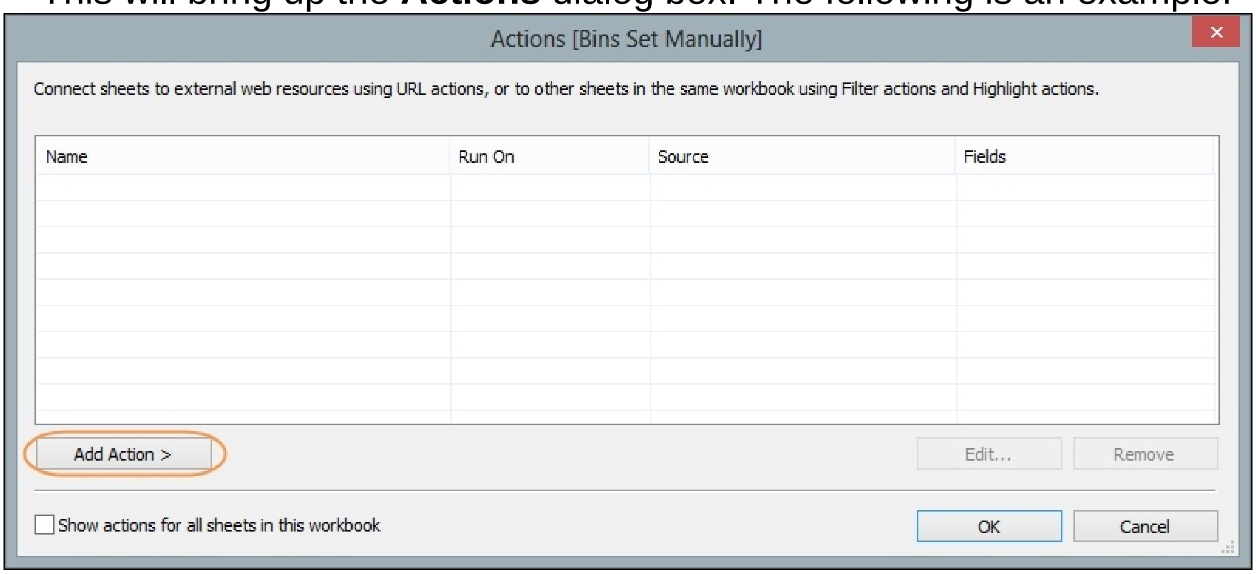
We will use the same data as before. For this recipe, we will take a copy of the **Bins Set Manually** worksheet and select the **Duplicate Sheet** option. You can then rename the sheet to `Actions`.

How to do it...

1. Once you have your worksheet in place, you will need to locate the correct **Actions** item. To do this, go to the **Worksheet** menu and look for **Actions**. You can see an example of this in the following

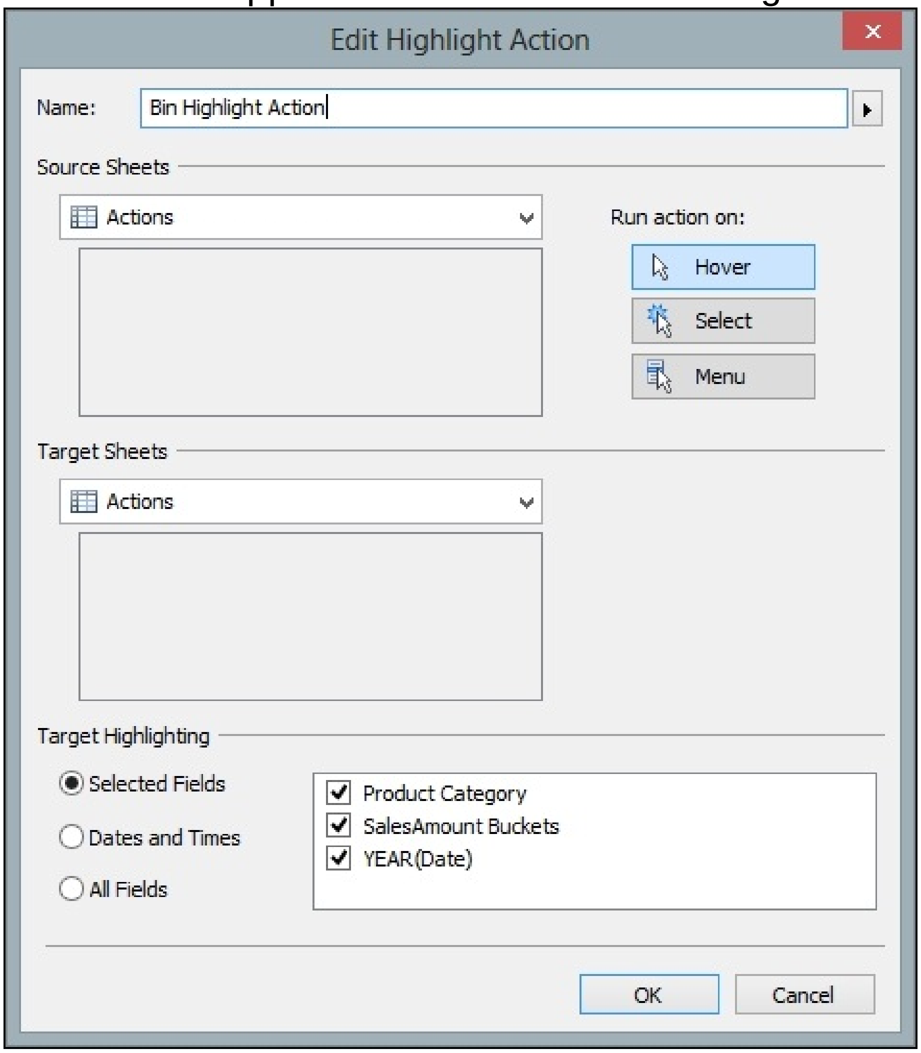
screenshot: 

- This will bring up the **Actions** dialog box. The following is an example:



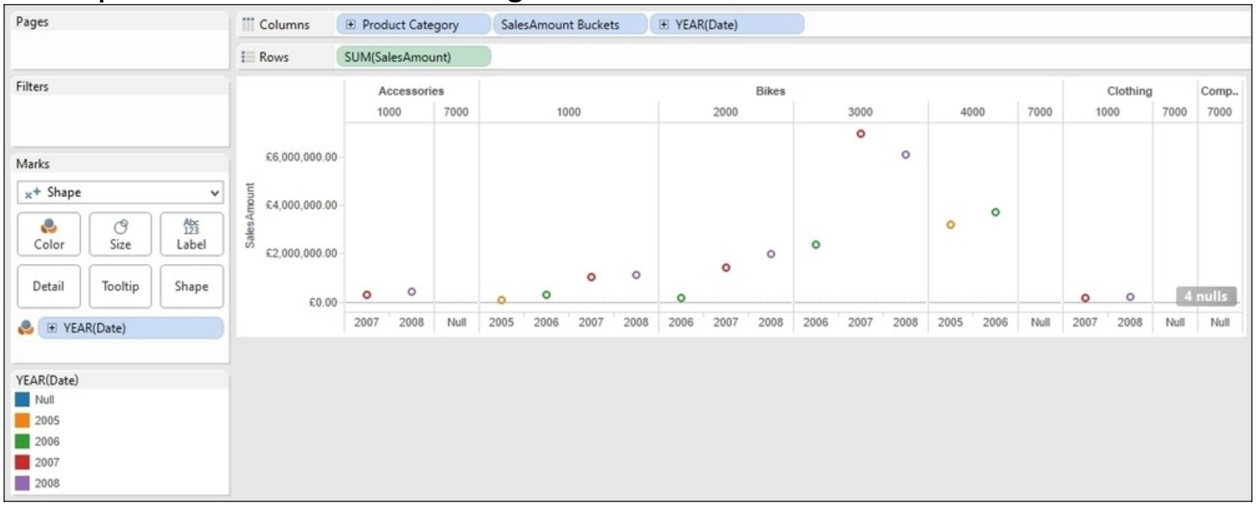
- In the **Actions** dialog box, select the **Add Action** button; this will bring up some options. We will choose the **Highlight** option.
- Once we have selected the **Highlight** option, you will see the **Edit Highlight Action** dialog box appear, which you can see in the next screenshot.

- We will call this `Bin Highlight Action`, and it will be based on the **Actions** worksheet.
- We will then choose the **Hover** option.
- For the **Target Highlighting** option, select all of the fields. The dialog box will then appear as shown in the following screenshot:

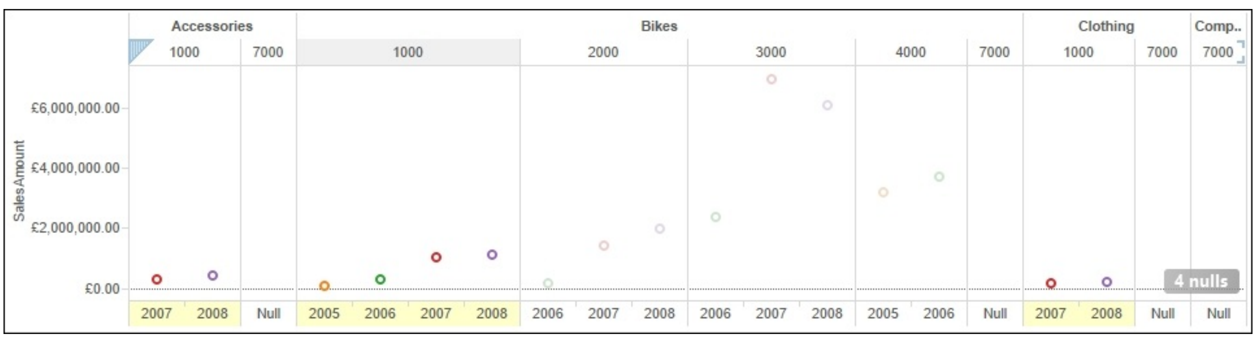


- Now, click on **OK** and then go back to the Tableau worksheet. We will change the Tableau worksheet so that we can see the result of the action.
- On the **Columns** shelf, place the **Product Category** hierarchy, the **SalesAmount Buckets**, and the **Year(Date)** dimension attribute.
- On the **Rows** shelf, select **SUM(SalesAmount)**. We will place **Year(Date)** on the **Color** button.
- Finally, select **Discrete (Lines)** from the **Show Me** panel in order to show the patterns over the years for each bucket type. You can see an

example of this in the following screenshot:



- If you hover the mouse over one of the bucket names, you will see that the relevant data points are highlighted. In the following example, when we hover the mouse over the **1000** bucket, we can see that it lights up the data points for that bucket; plus, the relevant years are highlighted. It's clear that other data points are grayed out.



How it works...

To summarize, we can use **Actions** to highlight data, and this functionality assists with the comparison process. Business users do not have to type in any information to achieve this result; a simple mouse hover will give them the patterns that they are looking for.

See also

- *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, *Stephen Few*, *Analytics Press*

Drilling into the details

Filters are a useful way to help users focus on particular aspects of the data that they are most interested in. This helps them to investigate and compare data and perhaps look for outliers and exceptions in the data.

Filtering data is an essential part of a dashboard. Users like to interact with data in order to understand it better, and it is natural to filter data so that users can pinpoint the data that particularly interests them.

Tableau allows users to filter data based on measures, calculations, and dimensions, which is extremely useful in a dashboard. For example, take the case where you need to see sales that are less than a given amount over a period of time. Users have a lot of flexibility when it comes to combining filters, which means that you can have a lot of creativity in your analyses.

In this recipe, we will look at combining the high level with the detail view. In dashboards, we tend to stick to the summary data only. However, very occasionally, business users may ask for details as well, and this recipe caters to this particular scenario.

We create a dashboard, make the dashboard appear more appealing to business users, and help to make it read better. Further, more we will explore some of the options for delivering detailed data to the business users.

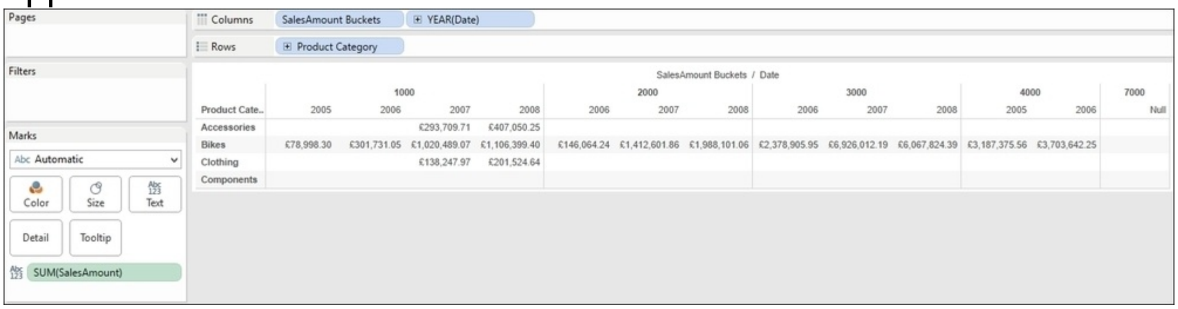
Let's make the current dashboard look better!

Getting ready

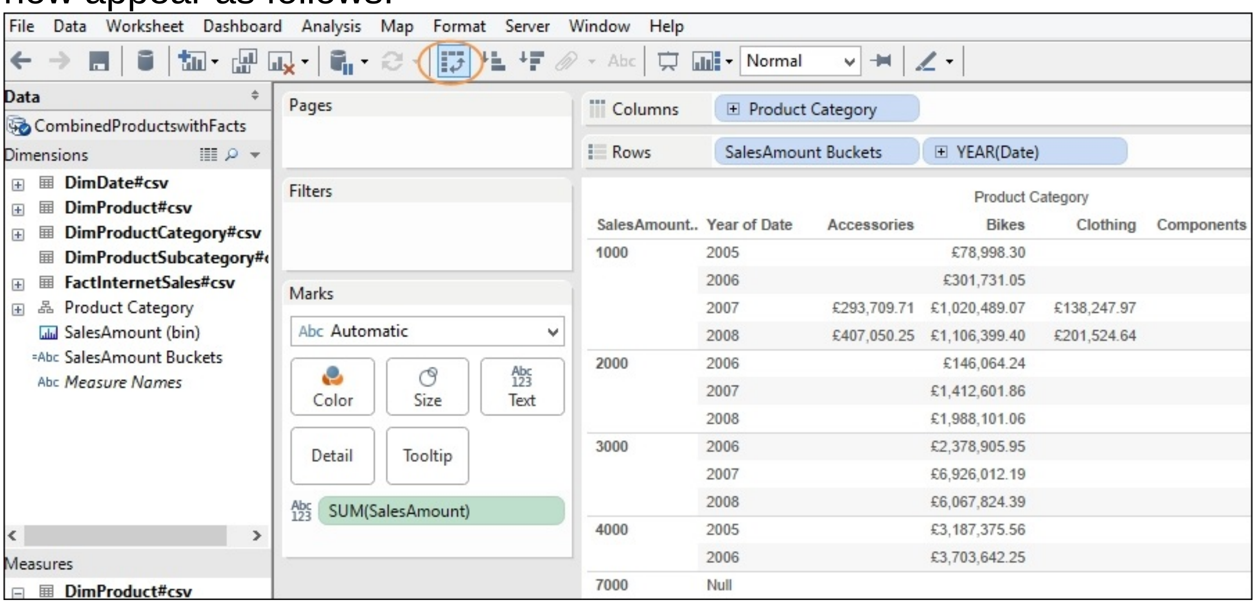
For the exercises in this chapter, we will continue to use the **Chapter Three** workbook. In this recipe, we will duplicate the **Bins Set Manually** worksheet and rename the duplicated sheet to `Bins Set Manually Table`.

How to do it...

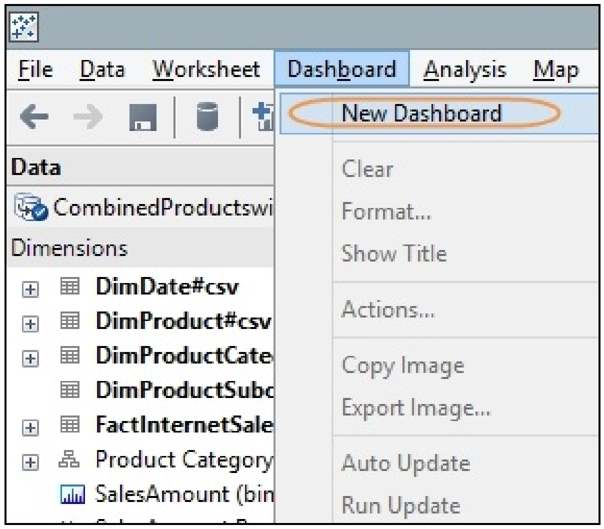
1. Drag **SalesAmount Buckets** to the **Columns** shelf.
2. Drag **Date** to the **Columns** shelf to the right of **SalesAmount Buckets**.
3. Drag **Product Category** to the **Rows** shelf. The visualization should appear as follows:



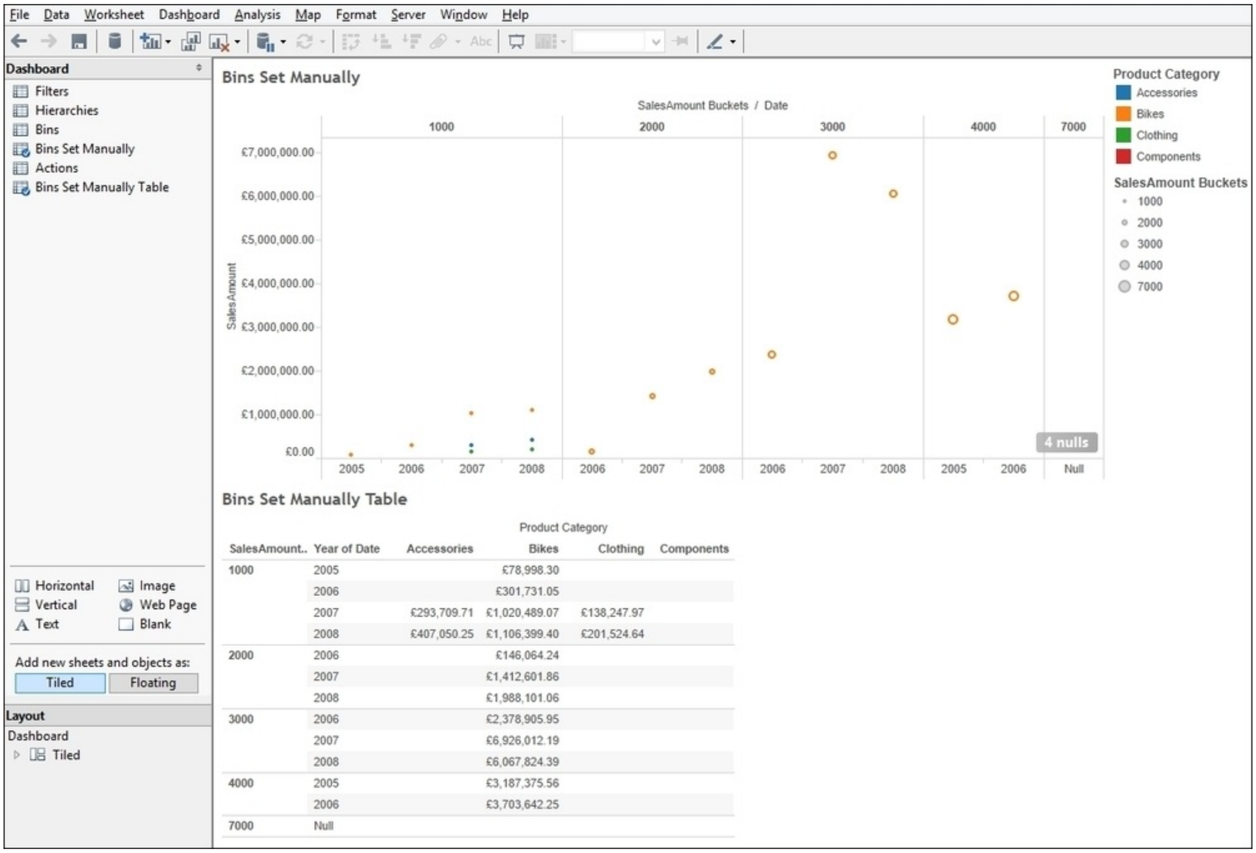
- We will swap the rows and columns so that the table does not use up as much space on the page. To do this, click on the **Swap** button that is located just under the **Format** menu item. The Tableau workbook will now appear as follows:



- Now we need to show the chart and the table next to each another. To do this, we create a dashboard, which is how Tableau combines different charts and tables together.
- We create a dashboard by navigating to the **Dashboard** menu item and selecting the **New Dashboard** option. You can see an example of

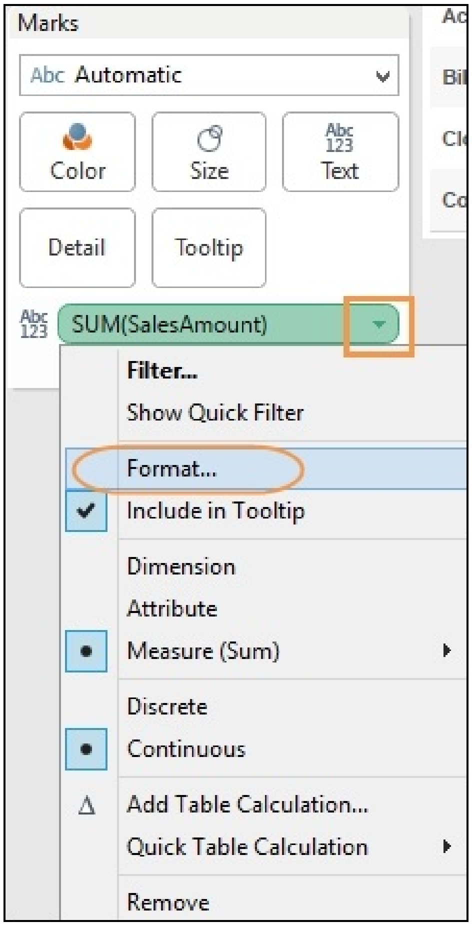
this in the following screenshot: 

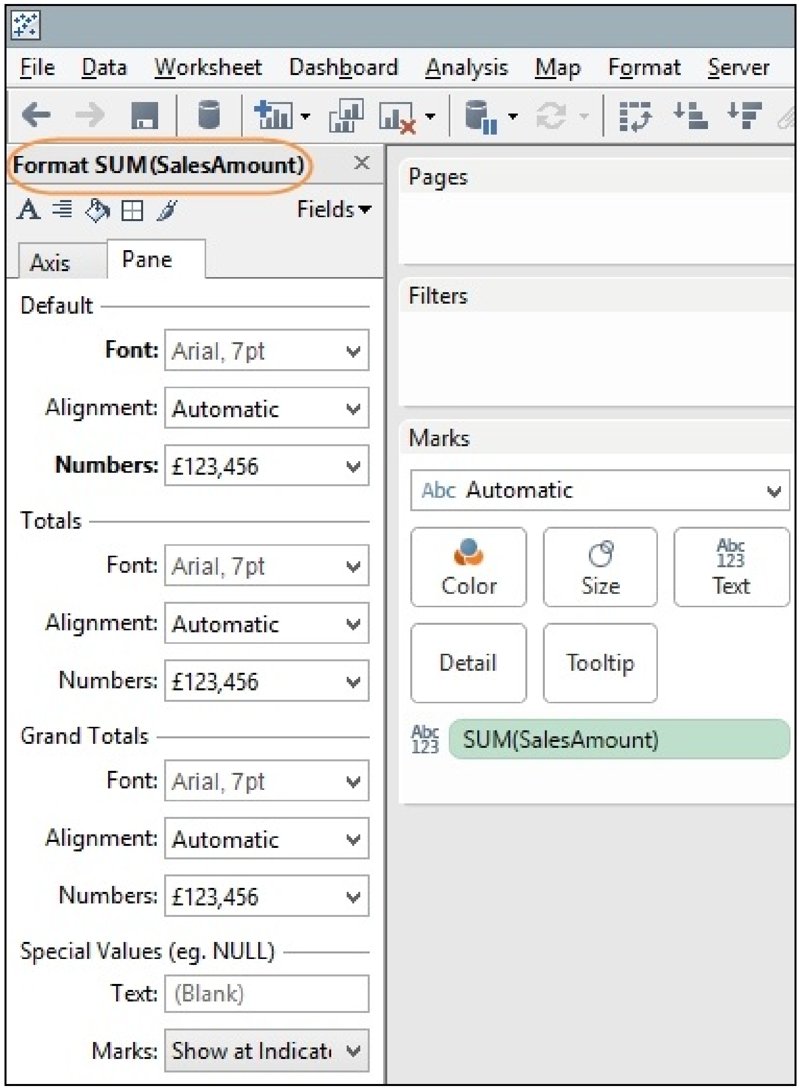
- A new dashboard will appear, and the tab will simply read **Dashboard 1**. To rename it, right-click and choose the **Rename** option. Here, we will call the dashboard, `Bucket Analysis`.
- Currently, our dashboard is empty, so let's populate it with some tables. On the left-hand side, you will see a section named **Dashboard**, which contains the names of other worksheets in the Tableau workbook.
- In our example, let's drag the workbook named **Bins Set Manually** to the dashboard area on the right-hand side. Then, drag the workbook named **Bins Set Manually Table** so that it sits under the **Bins Set Manually** chart. Your dashboard should now look as follows:



SalesAmount..	Year of Date	Accessories	Bikes	Clothing	Components
1000	2005		£78,998.30		
	2006		£301,731.05		
	2007	£293,709.71	£1,020,489.07	£138,247.97	
	2008	£407,050.25	£1,106,399.40	£201,524.64	
2000	2006		£146,064.24		
	2007		£1,412,601.86		
	2008		£1,988,101.06		
3000	2006		£2,378,905.95		
	2007		£6,926,012.19		
	2008		£6,067,824.39		
4000	2005		£3,187,375.56		
	2006		£3,703,642.25		
7000	Null				

- You will notice in the screenshot that we could make the columns align for each chart. So let's swap the bottom chart back to its original layout. Firstly, you will need to resize the bottom table so that it has the same width as the top chart.
- Then, to improve clarity, you will need to resize the columns in the bottom table so that the column lines match the lines in the top table. You may find that resizing has affected the readability of some of the numbers in **Bins Set Manually Table**. Tableau gives us the ability to change the number format plus the appearance of the data in the table. We can make changes in order to make the data more readable.
- We can easily change the format of the data so that it is much more readable, so let's do that first. To change the format properly, go back to the **Bins Set Manually Table** worksheet.
- In the first instance, let's change the format of the data in the table by going to the **Format** menu item. You can find it easily by going to the **Marks** shelf, then to the **Sum(SalesAmount)** lozenge and clicking on the downward-pointing arrow on the right-hand side. You can see this in the

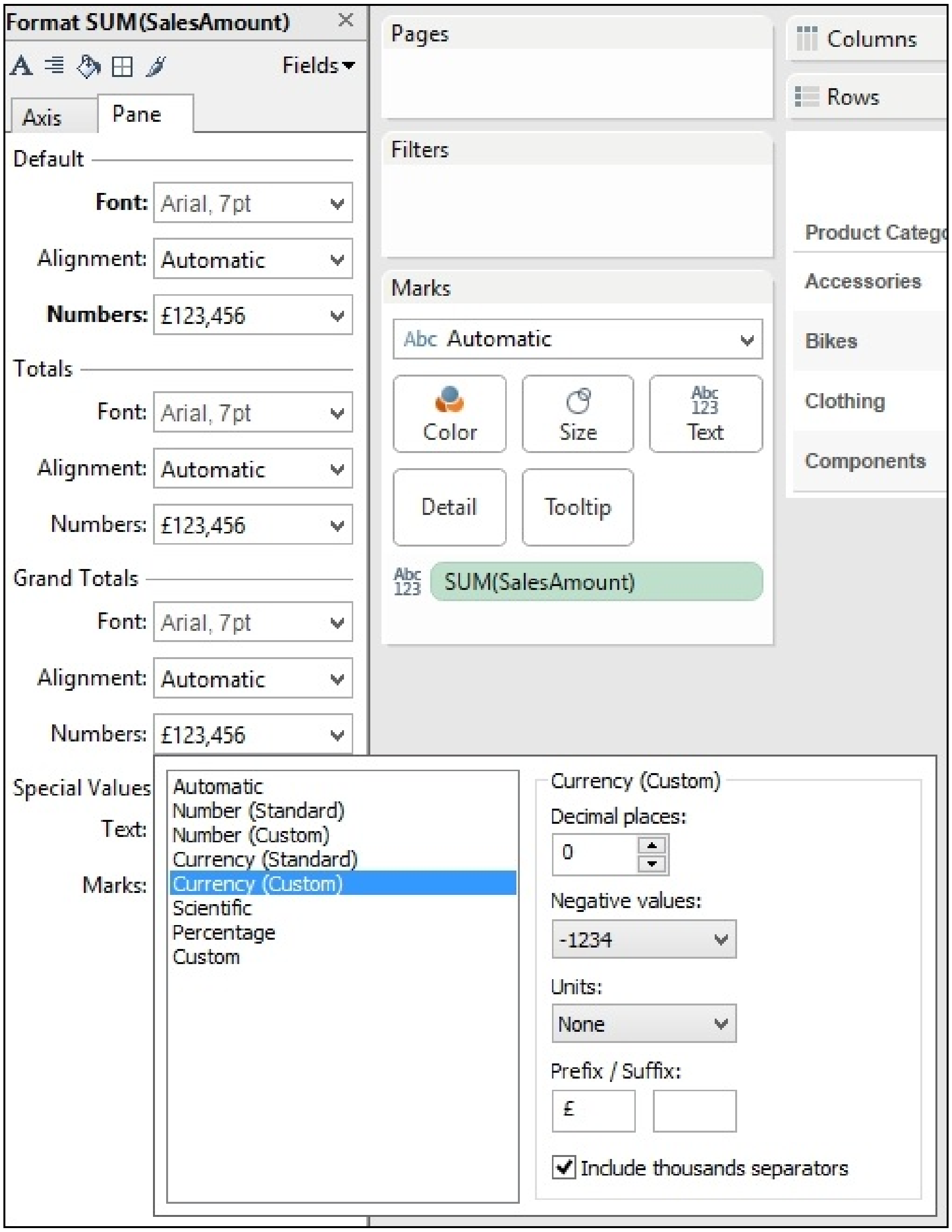
following screenshot: 

- When you select the **Format** option, the left-hand side of the Tableau workbook changes to a specific format-based panel. Normally, the left-hand side panel is dedicated to **Data**. However, it changes flexibly in response to the user needs. You can see the **Format** panel in the following screenshot: 

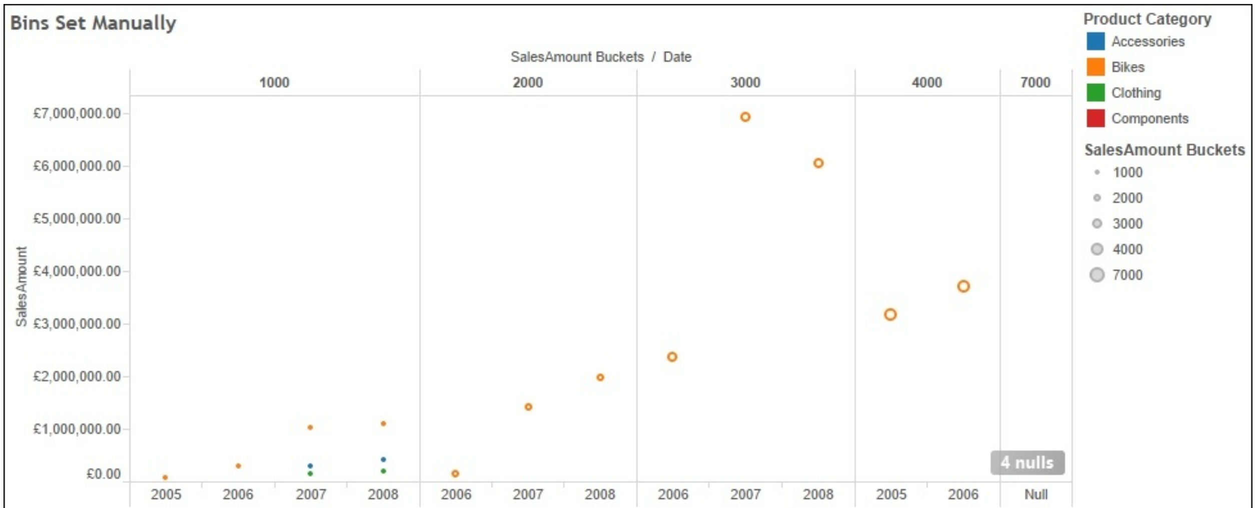
- As the first step, we can remove unnecessary data by changing the number format to remove the pennies. This costs us two extra characters in space, which may not seem very much per cell. However, across the 13 columns, this soon adds up to 26 unnecessary characters. To remove the pence, go to the **Numbers** option in the **Format** panel and select the **Currency (Custom)** option.

Make sure that the decimal places are set to zero; this will remove the pennies. This gives us more free space for more important data while retaining the width of the table to match the top chart.

- You can also reduce the font size by selecting the **Font** option under the **Default** heading, as shown in the following screenshot:



- The worksheet now appears as you can see in the following screenshot:

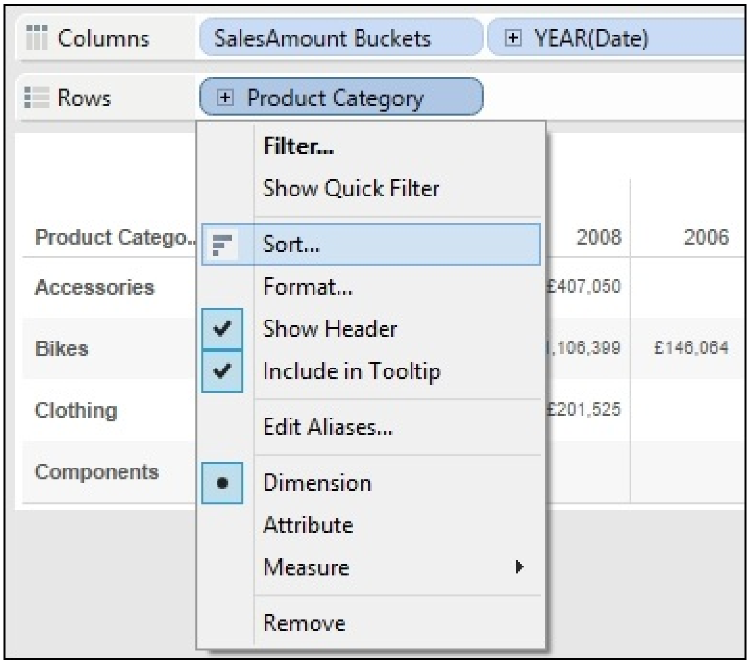


Bins Set Manually Table

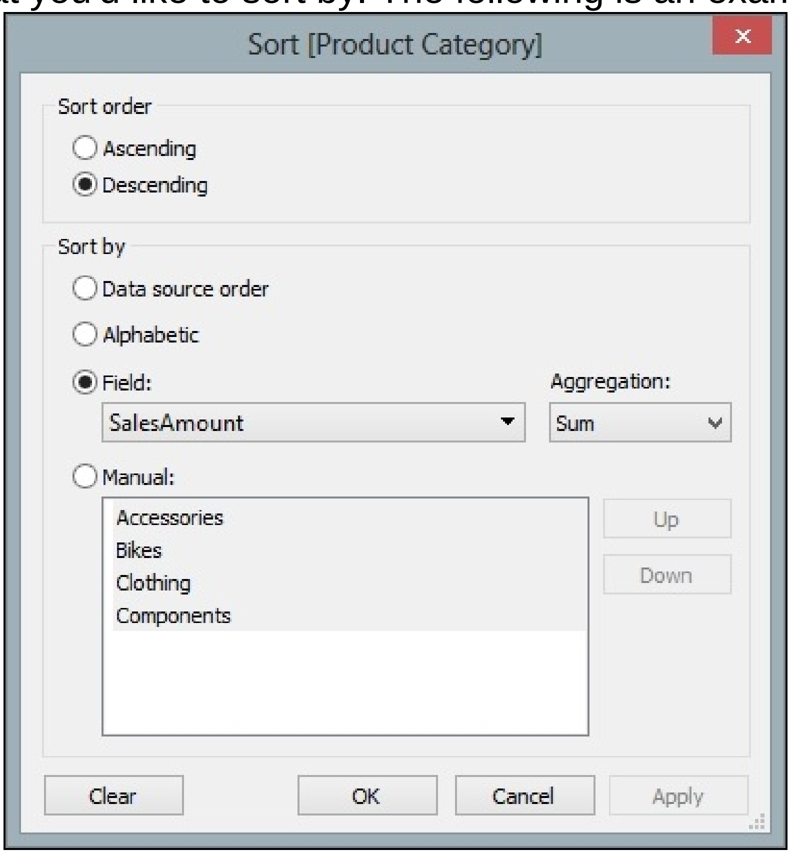
SalesAmount Buckets / Date

	1000				2000			3000			4000		7000
Product Catego..	2005	2006	2007	2008	2006	2007	2008	2006	2007	2008	2005	2006	Null
Accessories			£293,710	£407,050									
Bikes	£78,998	£301,731	£1,020,489	£1,106,399	£146,064	£1,412,602	£1,988,101	£2,378,906	£6,926,012	£6,067,824	£3,187,376	£3,703,642	
Clothing			£138,248	£201,525									
Components													

- To change the sorting of the table so that the rows are sorted by the value of the data and not the name of the product category, we will need to go back to the **Bins Set Manually Table** worksheet.
- Go to the **Rows** shelf, and you will find the **Product Hierarchy** dimension attribute. Right-click on it, and you will get a pop-up menu. Click on **Sort**, as shown in the following screenshot:

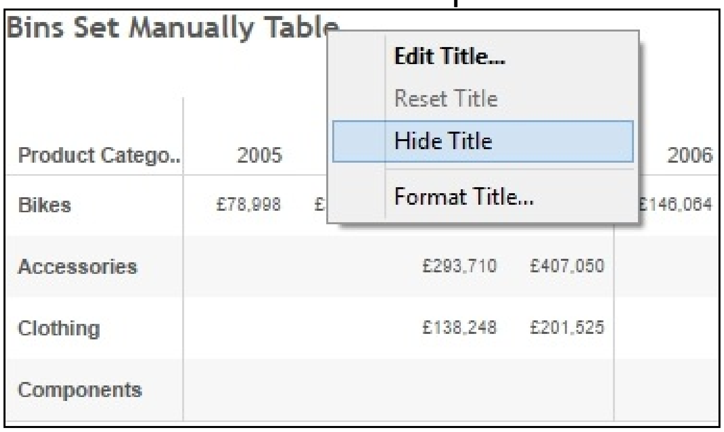


- Once you have clicked on **Sort**, you will get a dialog box where you specify what you'd like to sort by. The following is an example of the

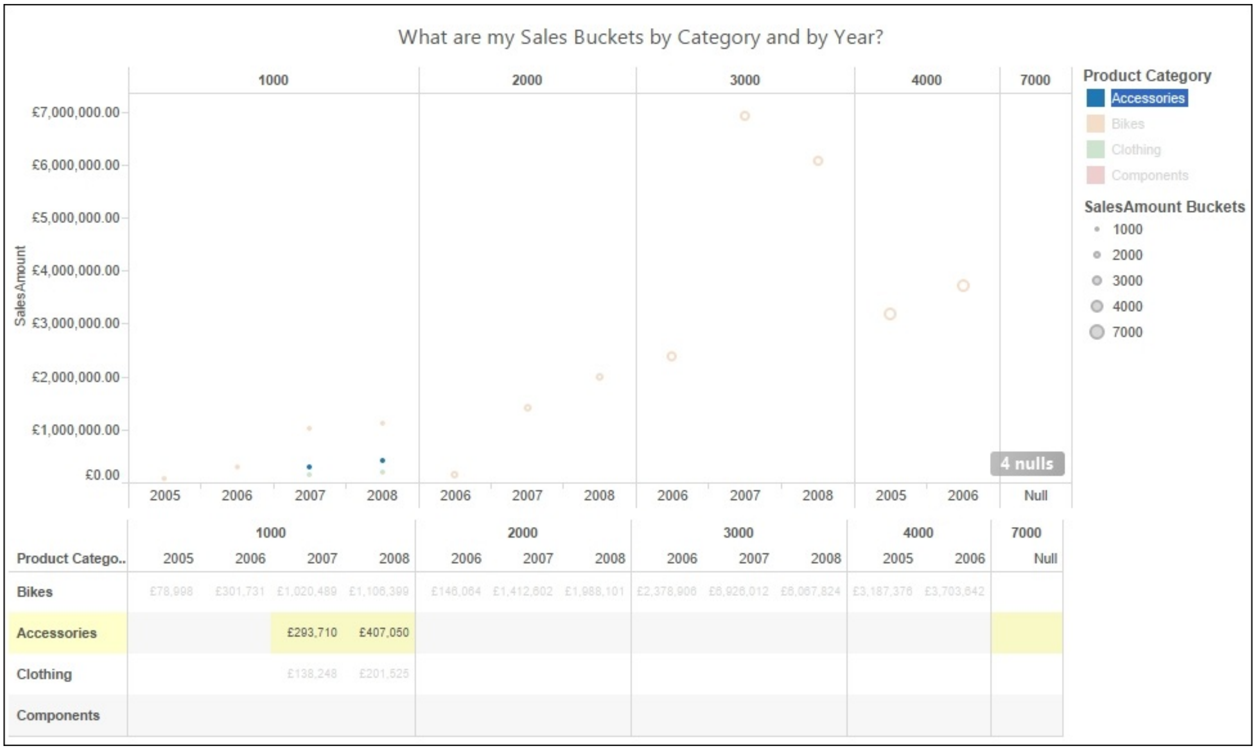


dialog box:

- If we sort the data in descending order, by value, then it will be clearer for the business users to understand it quickly. As shown in the previous screenshot, you will find a series of options for sorting the data. Here, we have chosen **Descending** as the value of **Sort Order**, and we have sorted by the **SalesAmount** field.

- Now that the data is sorted, we can go back to our dashboard, and we see that the table corresponds much better with the chart. To make the dashboard flow better, we can remove some unnecessary items, thereby adhering to our earlier discussion on chartjunk as a distraction in dashboards. For example, if we right-click on the title of the **Bins Set Manually** table, we can hide the title so it does not form an interruption between the chart and the table. The option is shown in the following screenshot: 

- Now that we have removed the titles, the image is simpler. Further, more the chart and table are consistent with one another. We can also add in text to make the dashboard clearer.
- We can see this from the following final screenshot:



How it works...

Tableau helps you to marry the summary and the detailed data by placing them together on the same page. When you click on a value in the legend, it highlights the appropriate values on the page. We have set up the configuration visually using Tableau so that we don't have to handcraft all of the programming language that occurs behind the scenes. This is stored in the `.twb` file, and we don't need to know any programming language to create dashboards.

Brushing the data gives us the opportunity to highlight and filter relevant data, which helps the business user to see emerging patterns in the data.

To summarize, we have used a combination of charts and details to create a dashboard. Normally, in dashboards, we focus on the summary of the most important information to help people make decisions. Occasionally, however, users will ask to be provided with the details and this technique allows people to see the pattern of the data alongside the actual detail. Some of you will notice, however, that there is something wrong with the ordering of the data in the table. The rows are ordered by the **Product Category** hierarchy, and these are ordered in alphabetical order. However, this does not match the data that appears in the chart above it. For example, if you take the **1000** bucket, you will see that the data points are accurately representing the value. So, the **Bikes** sales are represented by the values that appear at the highest point of the *y* axis. However, in the table below, the **Bikes** sales appear in the second row with the **Accessories** value above them.

Working with input controls

Removing some of the data can actually reveal more of the message of the data by narrowing the focus. Using filter controls in worksheets and dashboards is a way to pinpoint the data that you would like to show. Filters are very easy to set up, and their "clickiness" can help maintain "stickiness" in interacting with the dashboard itself.

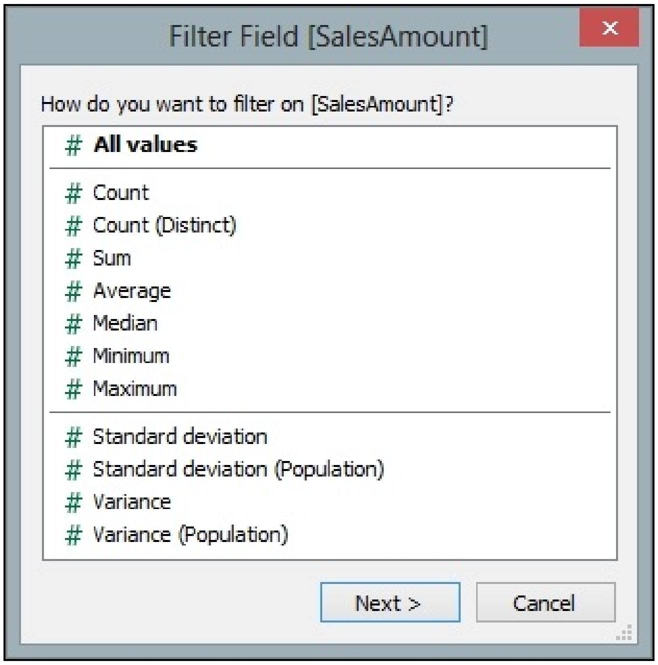
In this recipe, we will create dashboards that provide a summary while adding filters to include an interactive aspect to the dashboard, thereby engaging users further in the data. We will filter by measure and then show how this filter can be used in a dashboard.

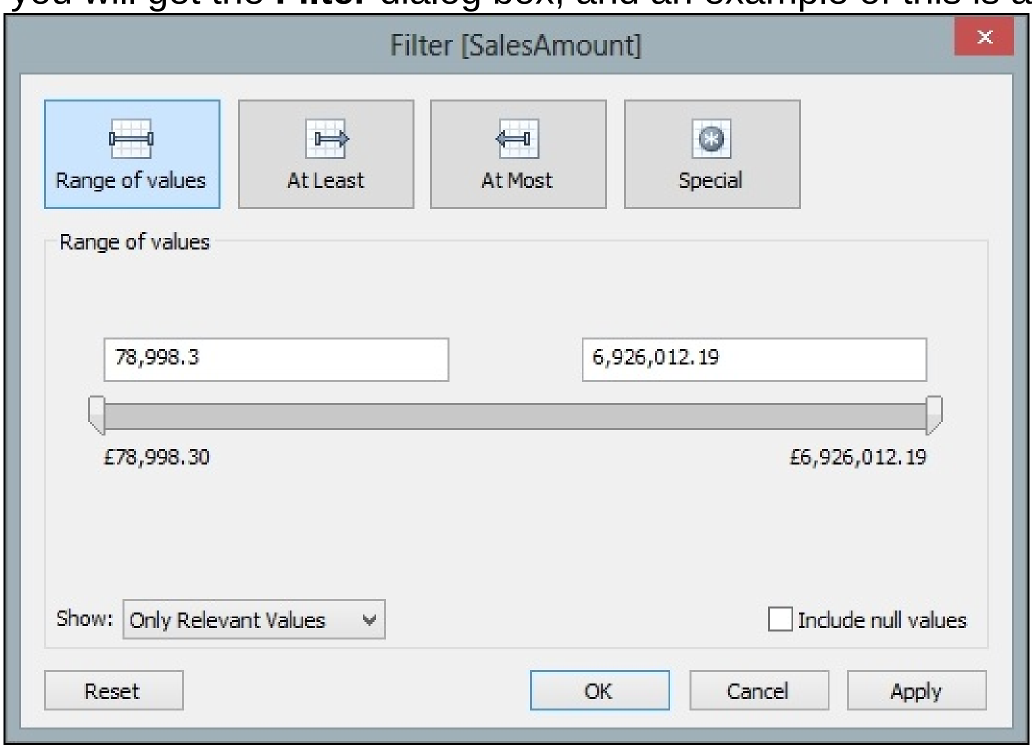
Getting ready

For the exercises in this chapter, we will continue to use the **Chapter Three** workbook. In this recipe, we will focus on the **Bins Set Manually** worksheet and rename the dashboard from **Dashboard 1** to `Sales Dashboard`.

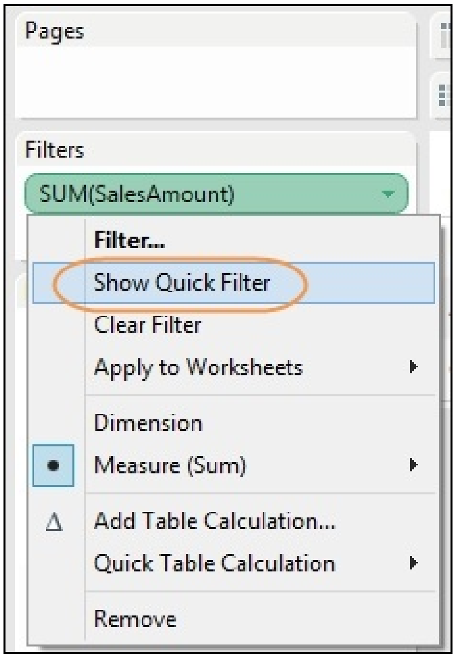
How to do it...

1. Take the **SalesAmount** metric from the **Measures** pane and put it in the **Filters** panel.
2. When you release the measure into the panel, you will see the following **Filter Field** dialog box pop up:

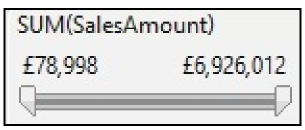


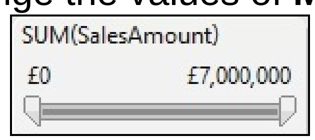
- We will filter on the **Sum** value, so select **Sum** and click on **Next**.
- Then, you will get the **Filter** dialog box, and an example of this is as follows: 

- In the worksheet, you can get a slider filter by simply clicking on the **OK** button. If you want the filter to appear as a separate slider, then right-click on the **SUM(SalesAmount)** filter and select the **Show Quick Filter** option. You can see this option in the following screenshot:

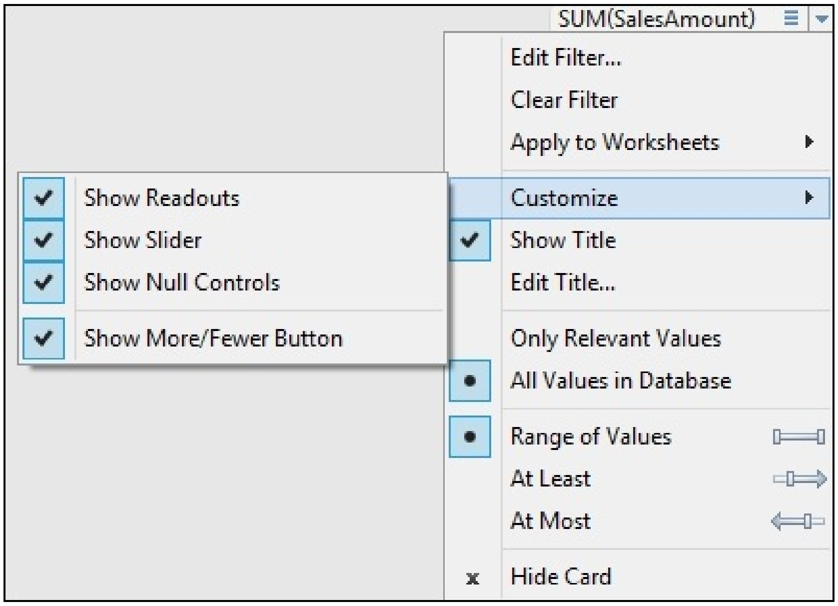


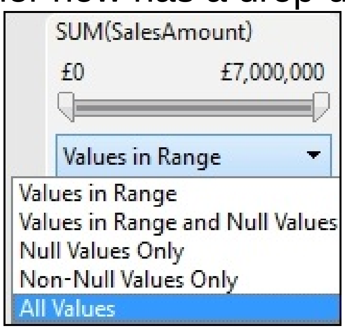
- You can now see that the slider is located on the right-hand side of the dashboard. The following screenshot shows a close-up of the slider:



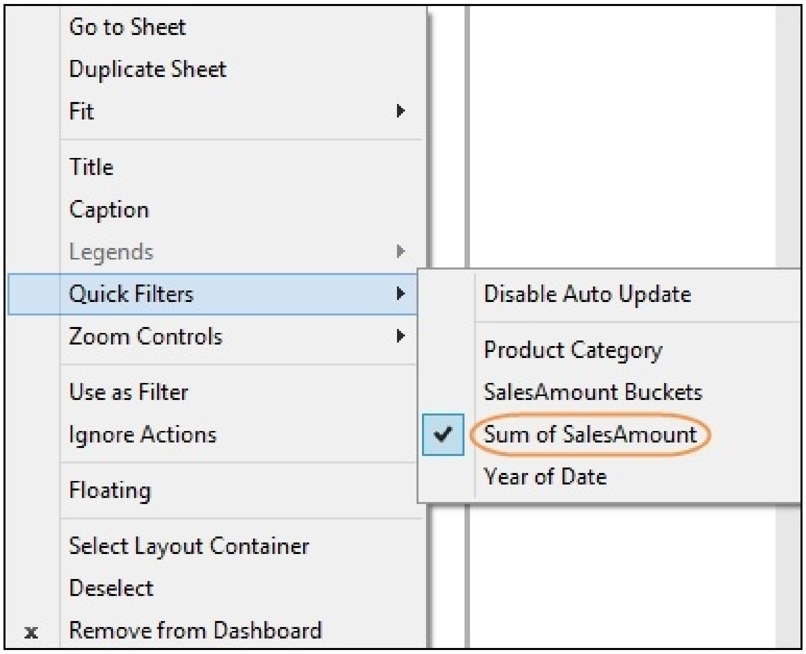
- The numbers look a little odd, don't they? Let's make it look better. If you select the option to edit the filter, you can change the values of **MIN** to `0` and **MAX** to **£7,000,000**. Now it looks cleaner: 

- We have to decide whether or not we should include the `NULL` values. The option to show nulls is labeled **Show Null Controls**, and you can find it the pop-up menu, as the following screenshot shows:

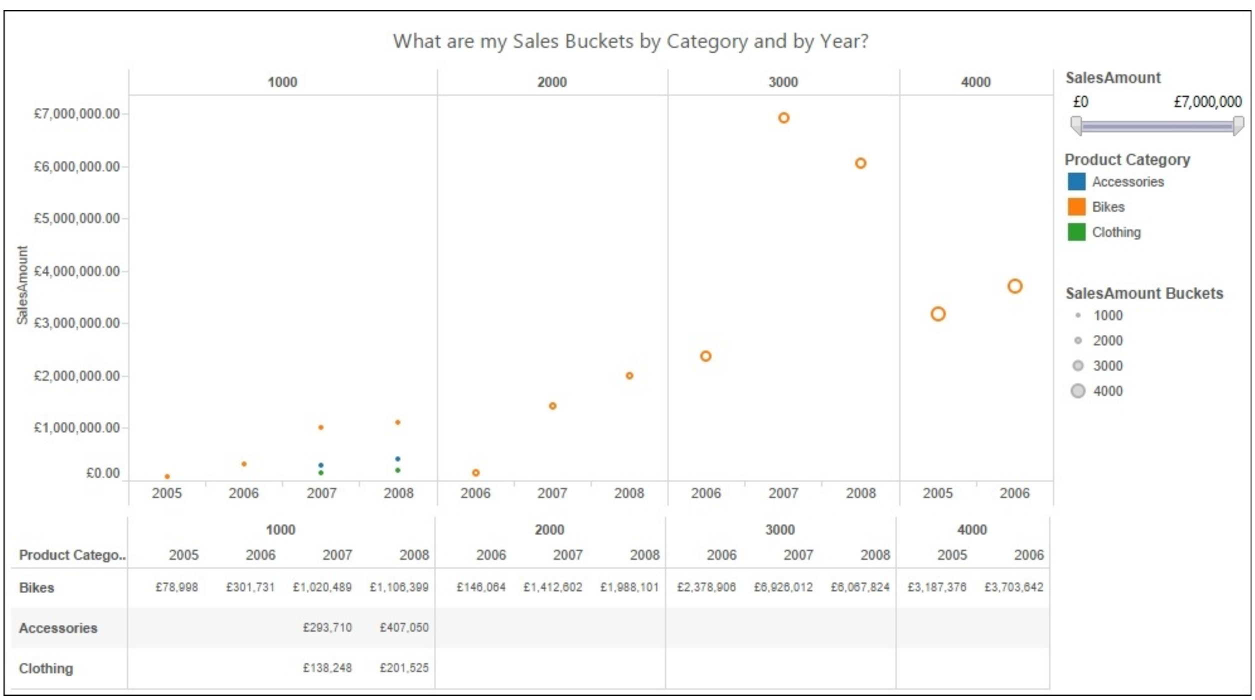


- The slider now has a drop-down list, which you can see in the following example: 

- Now, if you go to the **Dashboard** worksheet, you cannot see the **SalesAmount** filter. How do we make it appear? If you click on the small downward arrow at the end of the dashboard representation of the table, you will get a pop up.
- Navigate to **Quick Filters** and select **Sum of SalesAmount**; you'll see the filter appear in the dashboard, as the following screenshot shows:



- You can see the resulting dashboard in the next screenshot. Note that you can use the legend to filter along with the slider bar and the data points themselves. Users have many ways of interacting with the data, thereby increasing adoption.



How it works...

Filters are one of the most important features that automatically come with Tableau. As before, we set up everything visually, and the programming code is automatically generated behind the scenes. This means that it is easy to create Tableau dashboards as well as use them.

There's more...

Using filters is a key part of dashboards, since it allows people to understand the data better so that they can make decisions based on the data. Filters help business users by allowing them to interact with the data. Further "brushing" the data is a technique whereby we highlight the selected data points in order to see more details and gray out the irrelevant items.

Humans are thought to be able to hold only a limited amount of information in their heads at any one time; the "magic number" is thought to be seven, but it can often depend on the researcher. Highlighting selected data points helps us by focusing our attention on these data points and filtering the irrelevant material.

Chapter 4. Using Dashboards to Get Results

In this chapter, we will cover:

- Enriching data with mashups
- Page trails
- Guided analytics with Tableau
- Sharing your results in a meeting
- Notes and annotations
- Using external data to enrich your dashboard

Introduction

Dashboards are more than visual tools to display data; they are tools that can help to move your business forward. Dashboards are used as decision-making tools to obtain results quickly.

You can help your business users to make decisions fast by producing dashboards that are in line with the current research and thinking about dashboard structure. You can help users get results from their dashboard by:

- Improving the availability of data
- Facilitating the user to explore and understand the data quickly
- Sharing information with team members and beyond
- Providing adaptability in the dashboard
- Allowing flexibility for users to add notes to their dashboard

Enriching data with mashups

Business intelligence is all about people. We need to help people to understand data as quickly as possible so that they make strategic decisions more quickly.

Unfortunately, decision makers can end up being distracted by a need to mash data together in Excel. Worse, they may even expend time in trying to understand the data in the first place, rather than using the time to analyze the data to make an informed decision. This has a pernicious impact on the organization since decision makers are diverted from their role and contribution to the organization.

One proposal is to take a step back and re-evaluate the business questions and how they are answered. Due to the requirement that the data is correct and current, users could be provisioned with data that is insulated from the operational systems and merged together. This could help to answer the business need for a decision to be made using the data on the dashboard.

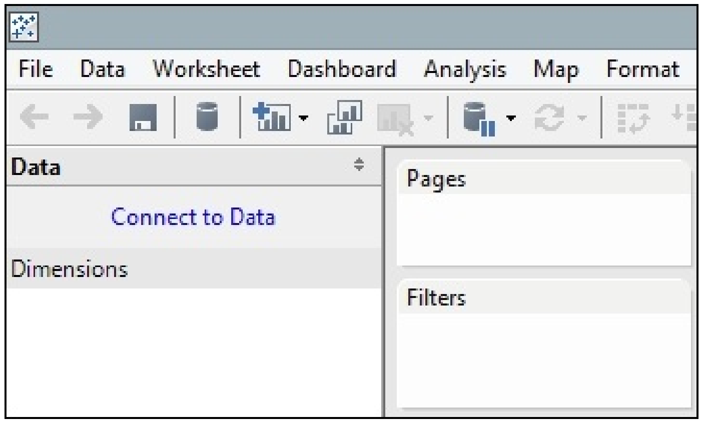
In this recipe, we will load the fact table called `FactInternetSales` first, and then we will load the dimension tables. We will look at joining tables together and putting the tables into Tableau's memory. When we do this, we relieve some of the pressure that is produced when we create an unnecessary proliferation of snapshots of data. In turn, data silos need to be mashed up by business users in order to get the information that they need. We will join tables together so that the business users can get results more quickly using Tableau, rather than trying to merge lots of smaller spreadsheets together.

Getting ready

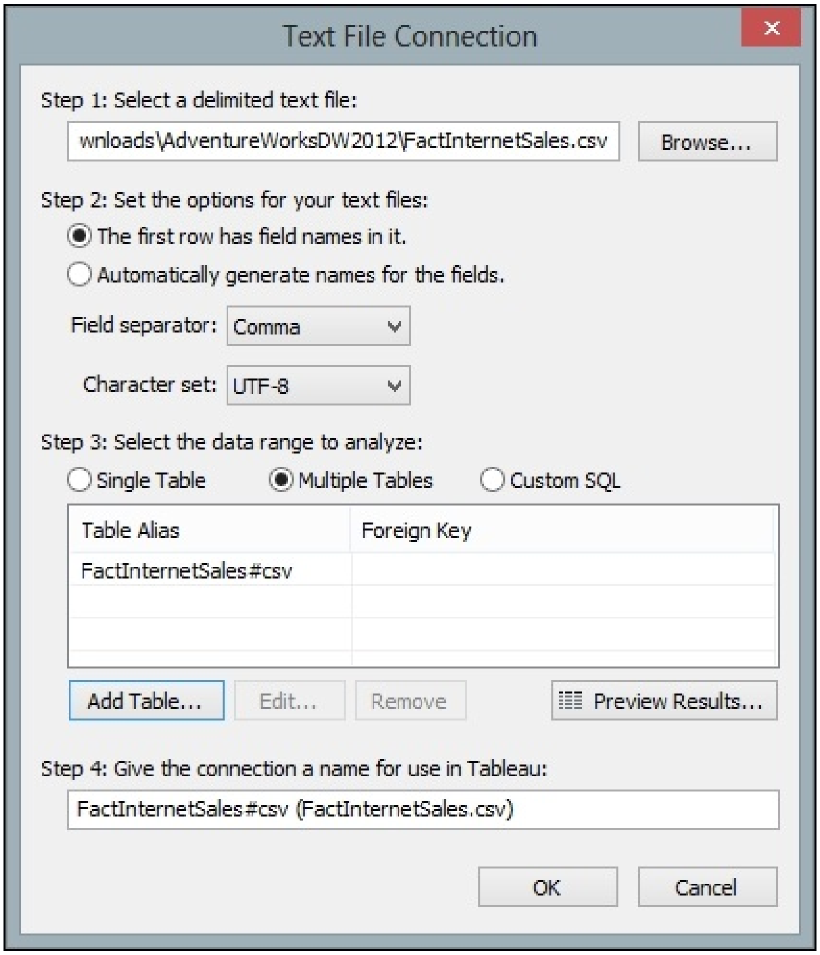
For the exercises in this chapter, open a new Tableau workbook and name it `Chapter Four`.

How to do it…

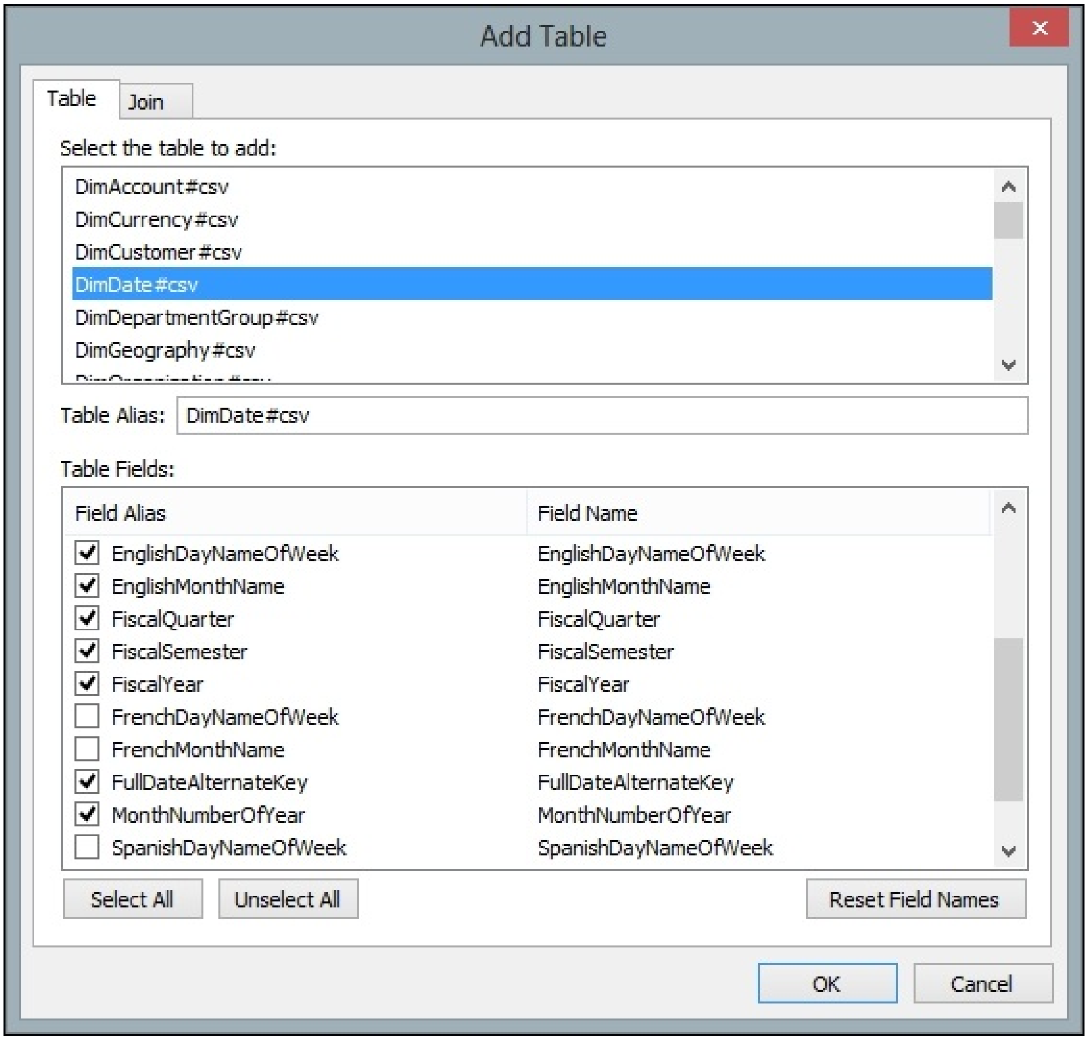
1. On the initial Tableau page, select **Connect to Data** and navigate to the location where you stored the data files, as shown in the following screenshot:



2. Navigate to the `FactInternetSales` file and select it.
3. Next, select the option to load multiple tables. You can see this in the next screenshot:

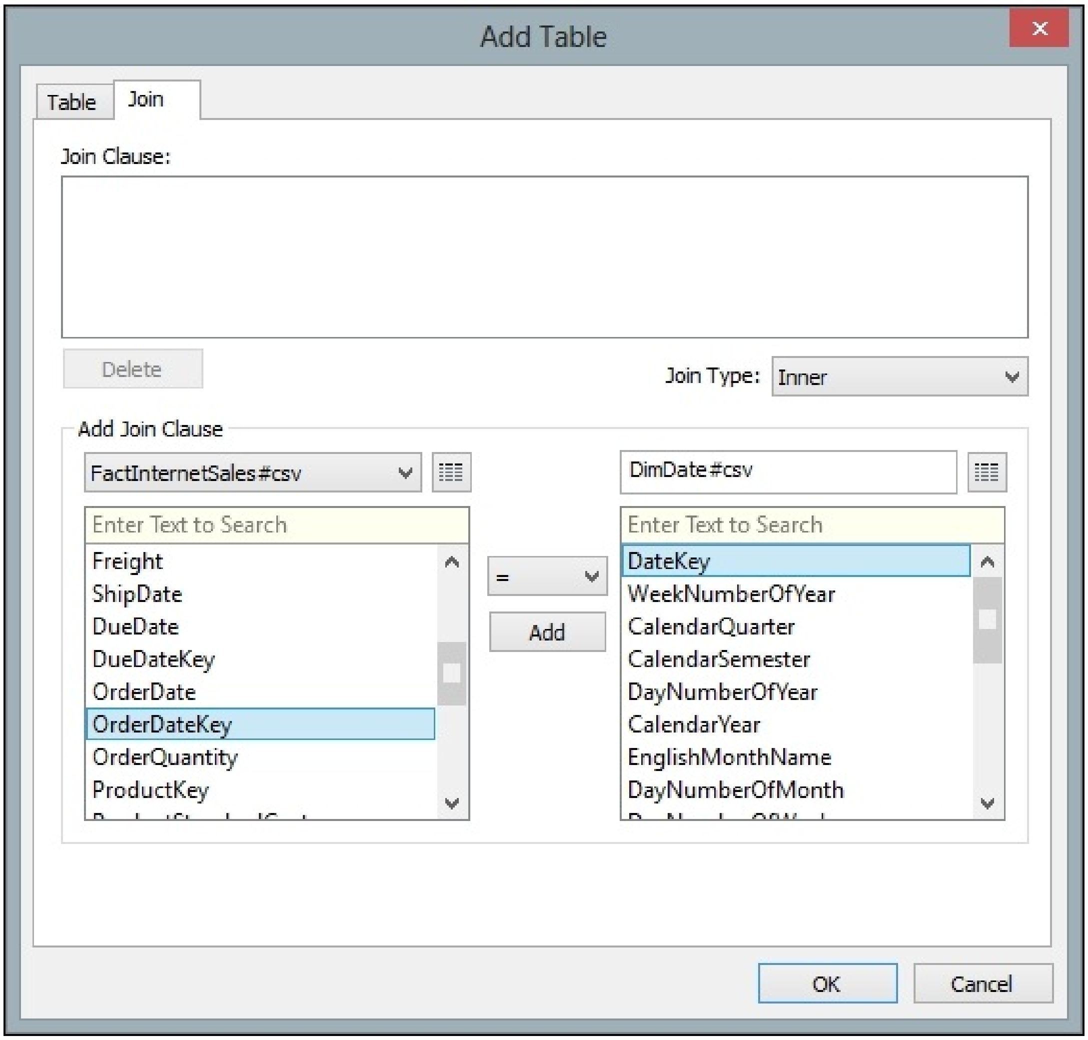


4. When we click on the **Add Table…** button for `FactInternetSales`, the following dialog box will appear:

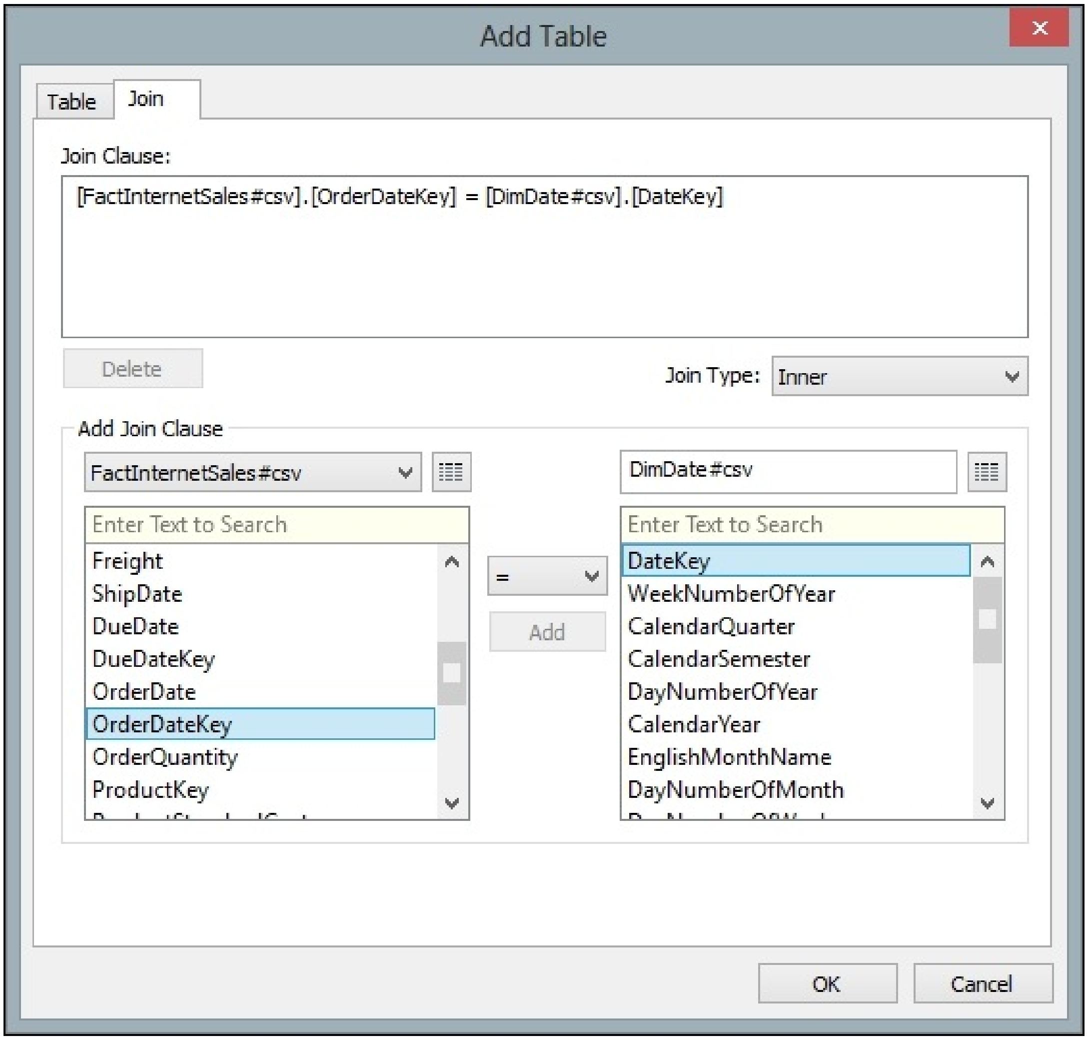


5. You can see that the previous screenshot has two tabs: **Table** and **Join**.
6. We will add the date dimension first. The connection to the table is called **DimDate#csv**, and you can see it in the previous screenshot.
7. You don't need to take all of the columns. In fact, it would be a good idea to take across only the columns that you need, or it will be confusing for the end user to see too many unnecessary columns. Here, we have removed the international language columns for clarity.
8. Now, let's click on the **Join** tab. We will join the tables together by the `OrderDateKey` column in the `FactInternetSales` table and the `DateKey` column in the `DimDate` table. The `FactInternetSales` table has more than one date: order date, ship date, and due date. The `DimDate` table has one key, which is called `DateKey`. A key is a

database column that is used to establish relationships between tables in a database so that it does not have lots of redundant information. You can imagine how confusing it gets if a data store has the same column located all over the place in many different tables. So, we use keys to link tables together so that we can reuse the same information as much as possible. You can see the example in the following screenshot:



9. When you click on the **Add** button in the middle, a dialog box will appear as follows:



10. You can now click on **OK**, and you will be taken back to the **Add Table** dialog box.
11. We will now add in other tables—all in the same way. Tableau attempts to work out the keys for you. In case you need a list, the following table is a summary:

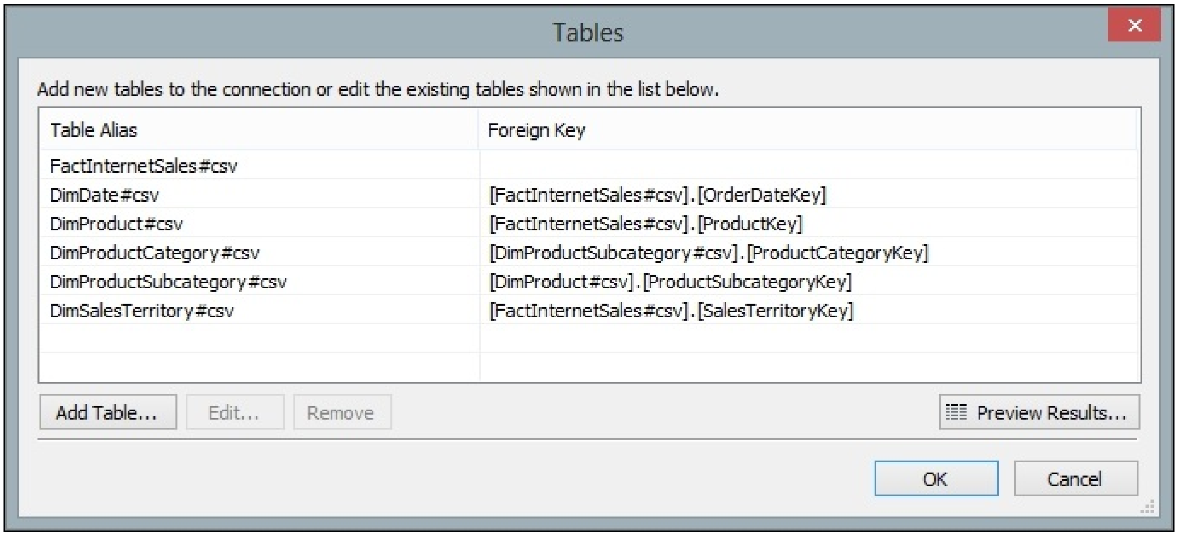
Table name	**Key**
`DimProduct`	`ProductKey`
`DimProductSubCategory`	`ProductSubCategoryKey`

DimProductCategory	ProductCategoryKey
DimSalesTerritory	SalesTerritoryKey

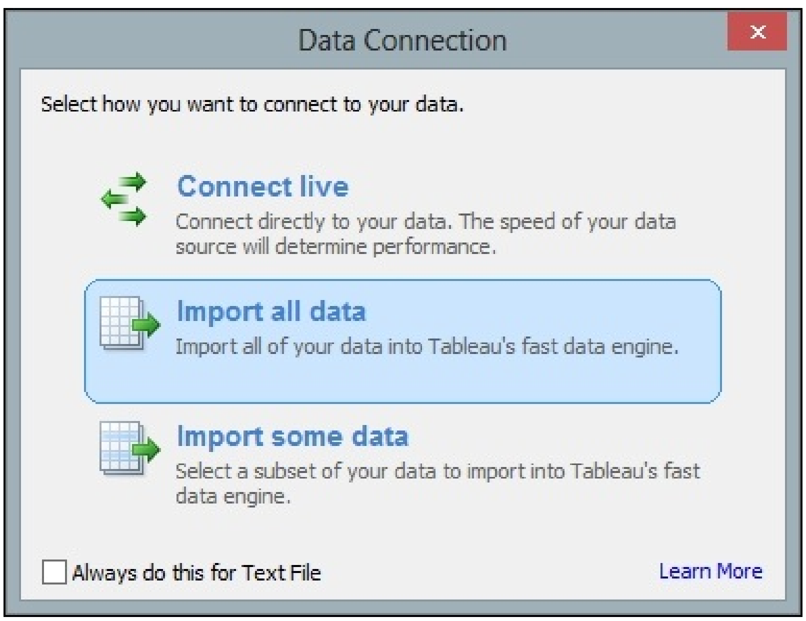
12. For each additional table, we will add an inner join. What are these joins? An inner join, also known as an equi join, selects only the rows from both the tables that have matching values. Rows with values in the joined field that do not appear in both of the database tables will be excluded from the result set. So, for example, if there is a row in the `FactInternetSales` table which does not have a value in the `DimDate` table, then it will not be returned, and vice versa.
13. With a left join, the selected rows will include all of the records in the first database table. In this case, it will return all of the rows in the `FactInternetSales` table whether there is a match in the `DimDate` table or not.

> ## Note
>
> One or more fields can serve as the join fields. In this simple example, we have selected only one field for clarity.

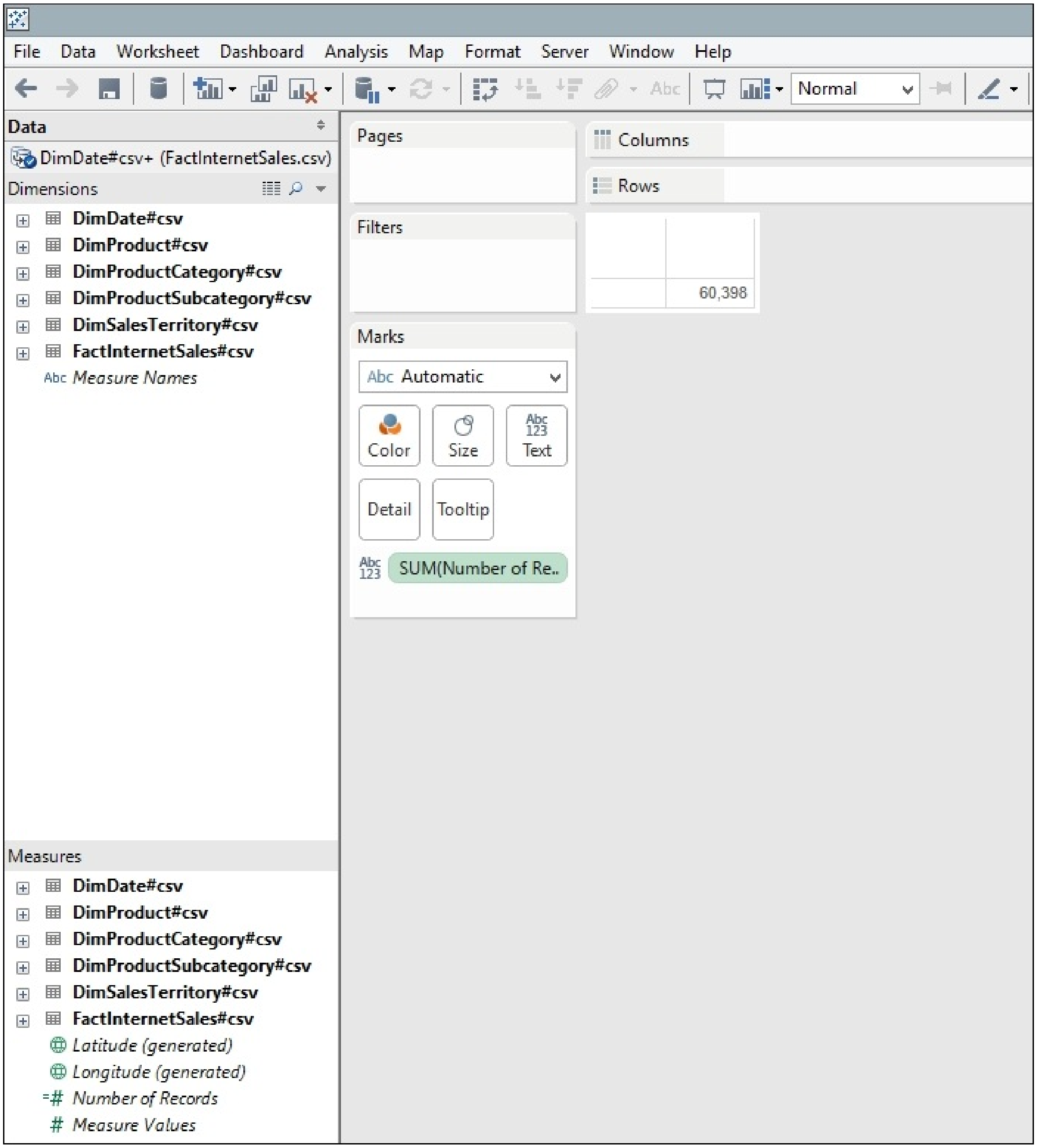
14. Once you have added all of the tables, your dialog box will appear as you can see in the following screenshot:



15. Once you have clicked on **OK**, you can import all the data or just some of it. We will import all of the data into Tableau's data engine. A list of the options is shown in the following screenshot:



16. Once you have imported your data, your Tableau worksheet will show all of the tables on the left-hand side. You can see the next screenshot as an example:



You can see that the tables are located in the **Data** pane on the left-hand side.

How it works…

To summarize, in this section, we have shown different ways of joining tables in order to alleviate the situation where people are copying and pasting data all over the place. By unifying all the data that the users need into one distinct place, it will save them time and energy that they could use to make better decisions.

Tableau has its own data engine, which is an analytics database. It uses compression, which means that it can store a lot of data. It also involves techniques to make data retrieval very fast. It is a flexible data model, and you can work on the data very quickly in the same way as it is represented on the disk.

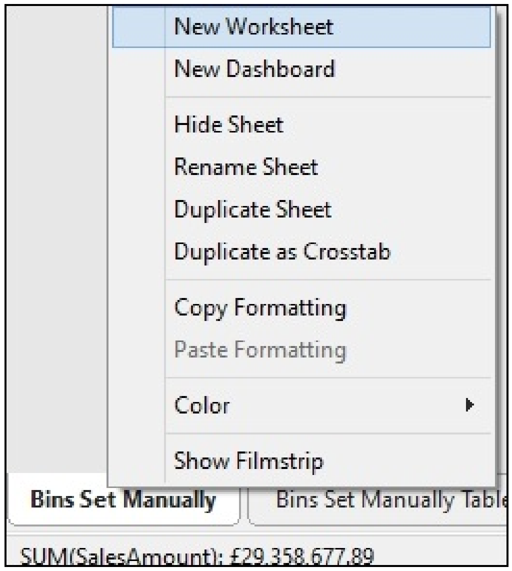
We saved the data extract, which is a separate file from the Tableau workbook. It has the file extension `.tde`, which stands for Tableau Data Extract. This extract file can be reused directly without having the source connection details. This is extremely useful for portability. You can also use it as a way to prototype the dashboard. Further, you can anonymize a dataset and load it into an extract file. Then, you can ask the developer to work from the anonymized extract file. Once the development is complete, you could change the source connection so that Tableau connects to real-world data.

There's more…

Here is a trick: if you want to verify the number of rows loaded into Tableau's data engine, then take the **Number of Records** metric and put it into the **Marks** shelf. This will give you a quick check to see if the number of rows loaded matches what you expect. If the number is much lower or higher, then one of your joins may be wrong. It's best to do this before you start or you will have to redo the work!

What happens if you have made a mistake and want to delete worksheets? You can delete worksheets by right-clicking on the worksheet tab and selecting **Delete Sheet**.

When you attempt to delete a worksheet, you will notice that there is no option to delete the sheet. You can see an illustration of this feature in the following screenshot:



If the worksheet is used in a dashboard, Tableau will not allow you to delete a worksheet if it is reused elsewhere. In order to get around this feature, you need to delete the dashboard first, and then you'd need to delete the original worksheet. Tableau doesn't let you delete all of the worksheets, however, since it needs a worksheet in order to show data!

Page trails

Websites often have page trails that help users to find their way around the site. Similarly, in Tableau, we can add features that will help business users to reduce the number of actions that they need to take in order to navigate through the workbook. These actions can help to make worksheets more findable in a Tableau workbook. Findable refers to the ease with which a website can be found, and it is also relevant to finding pages within a Tableau workbook.

Linking workbooks together is an effective visual tool that helps the user to understand where they are in terms of the user's location within the workbook. It also helps to add context. In this recipe, we will look at user-oriented trails in a Tableau workbook.

Getting ready

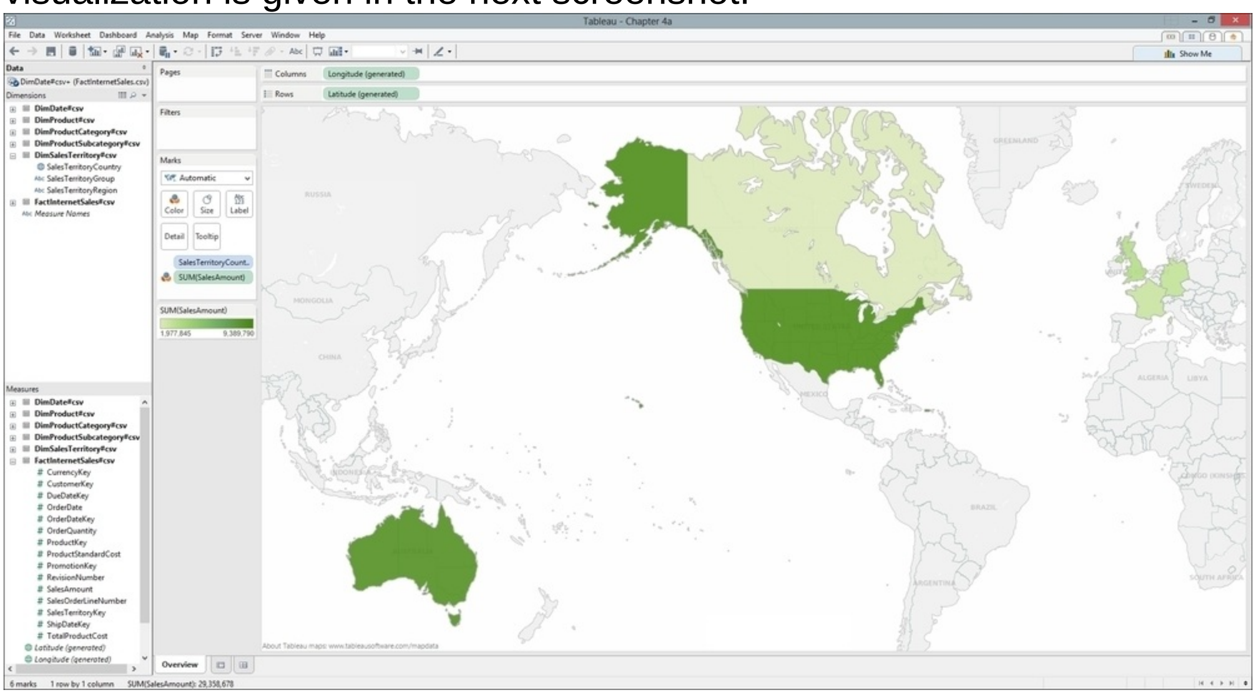
In this recipe, we will use the existing workbook that you created earlier in this chapter. If you have taken the **Number of Records** field and put it onto the white canvas, remove it.

How to do it…

1. First, let's rename the worksheet to `Overview`.
2. Let's take the **SalesAmount** metric from the `FactInternetSales` table and place it onto the white canvas.
3. Then, navigate to the **DimSalesTerritory** dimension, look for the `SalesTerritoryCountry` attribute, and drag it onto the white canvas.
4. Once you have the fields in place, let's select the filled map from the **Show Me** panel. To help you find it, you can see the next

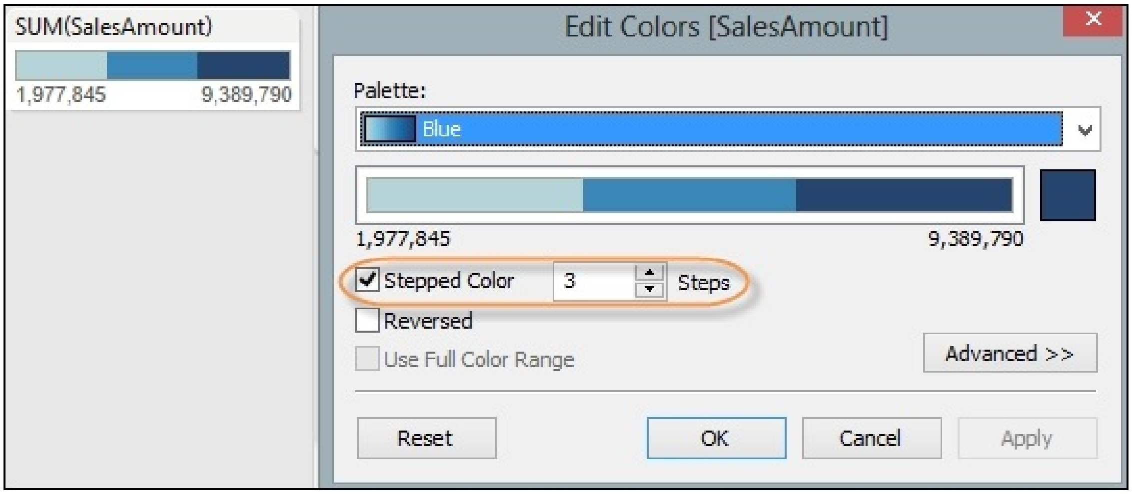
screenshot: 

- Let's work with the visualization of data. An example of data visualization is given in the next screenshot:

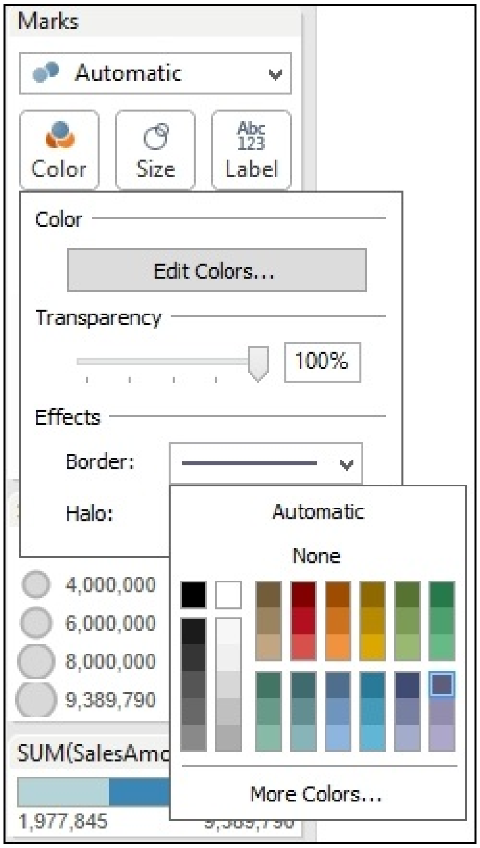


- Now, let's work on the color and the size. For the color, let's drag the **SalesAmount** metric from the `FactInternetSales` table and place it onto the **Color** button on the **Marks** shelf.

- When you see the **SUM(SalesAmount)** metric on the **Marks** shelf, right-click on the arrow on the right-hand side of the dialog box. You can see this in the next screenshot:

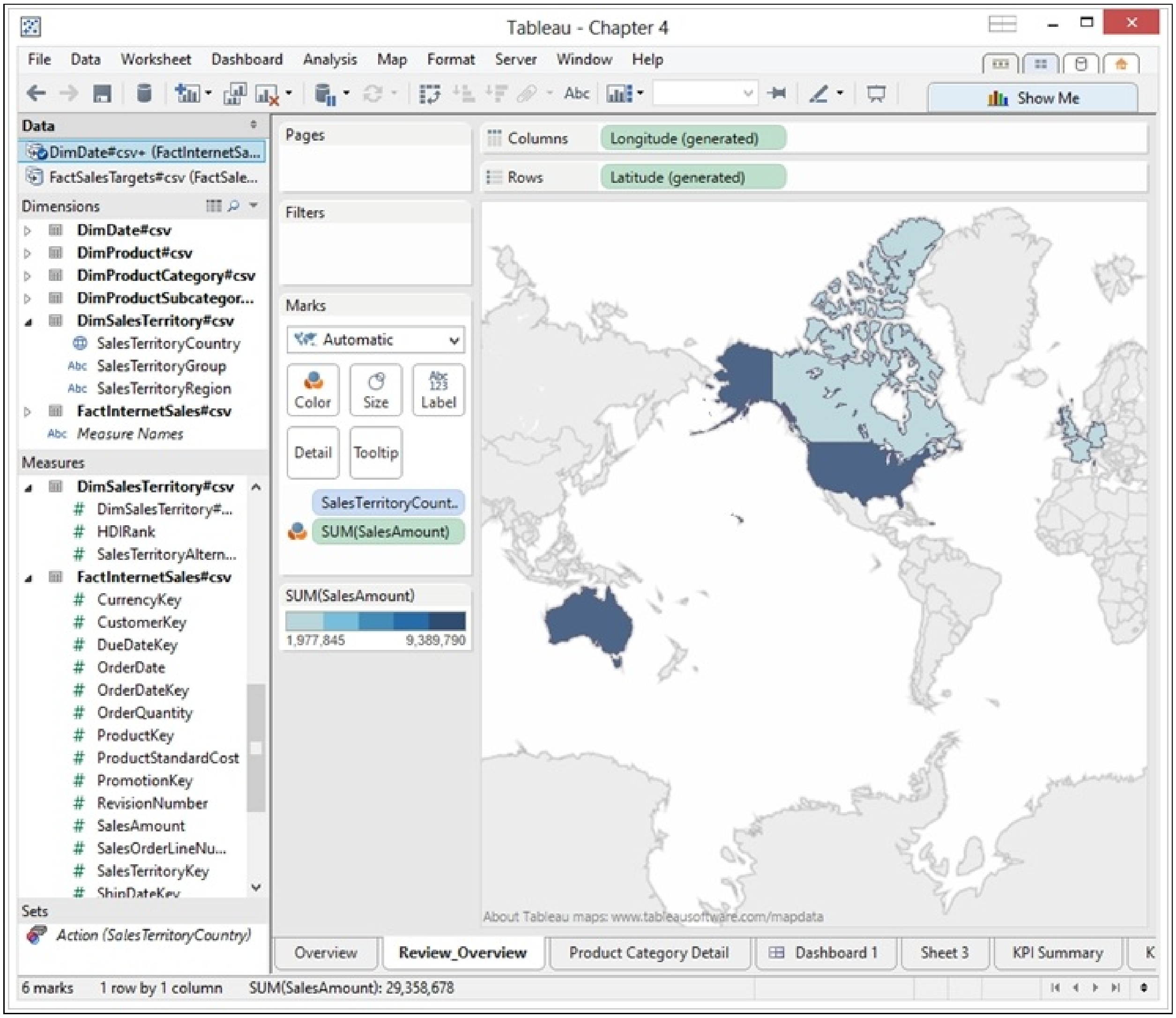


- From the drop-down list, select the **Blue** option for color.
- Then, we will select the **Stepped Color** option.
- Next, select **3** steps and click on **OK**.
- Some of the colors might appear a little pale on the screen, so we will give a very light border to the shapes. To do this, click on the **Color** button on the **Marks** shelf.
- Look for the **Effects** section for the **Border** option. Here, you will get a drop-down list that gives you the option to change the color. In this example, the border has been changed to a mid-purple color since it is softer than black or dark grey, as shown in the following screenshot:

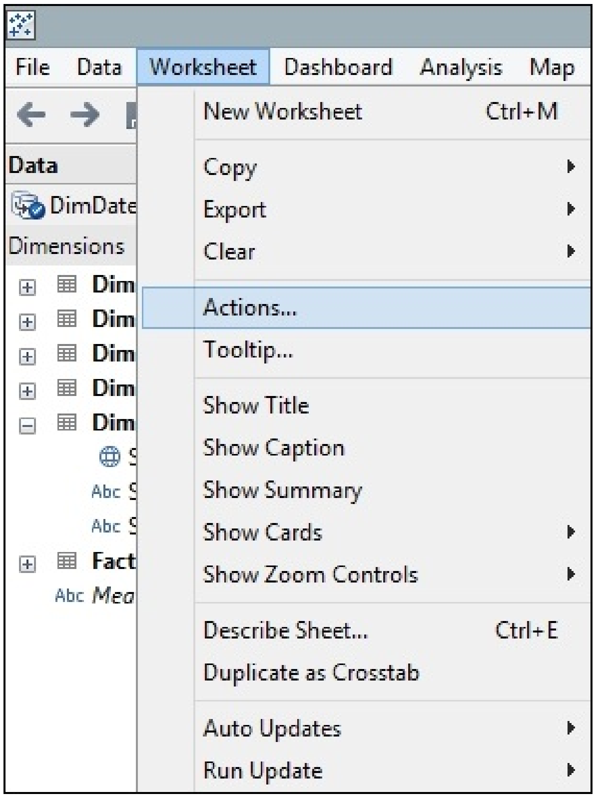


- Finally, drag **SalesAmount** to the **Size** button on the **Marks** shelf.

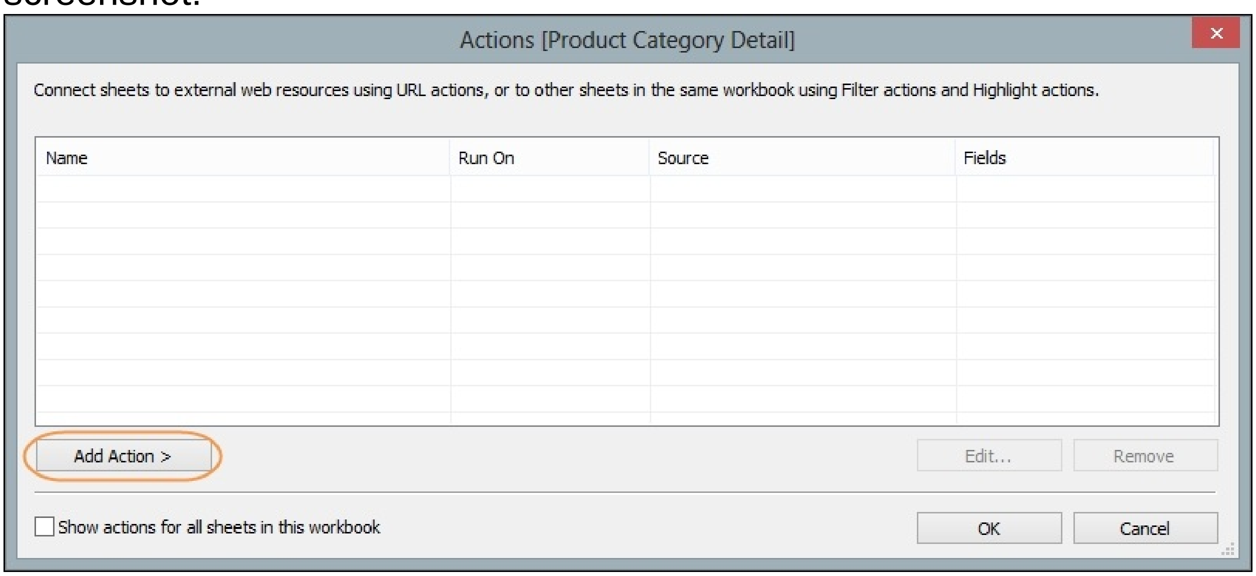
Once you have made these changes, the screen will appear as you can see in the following screenshot:



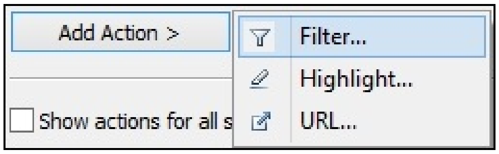
- So, let's proceed towards creating our `Product Category Detail` worksheet. Now that we have an `Overview` worksheet, we will create a `Product Category Detail` worksheet that we would like to navigate to. To do this, simply right-click on the tab and select **Duplicate Sheet**.
- Rename the duplicated worksheet to `Product Category Detail`.
- Navigate to the `DimProductCategory` dimension, look for the attribute `EnglishProductCategoryName`, and drag it to the **Rows** shelf.
- Next, let's add an action that will allow the user to simply right-click on country on the `Overview` worksheet and they are presented with the `Product Category Detail` worksheet. To add an action, go to the **Worksheet** menu item and choose the **Actions…** option. You can see where to find this option in the following screenshot:



- Now, when we click on the **Actions…** menu item, we get a dialog box. Initially it is empty; you can see an example of this in the following screenshot:



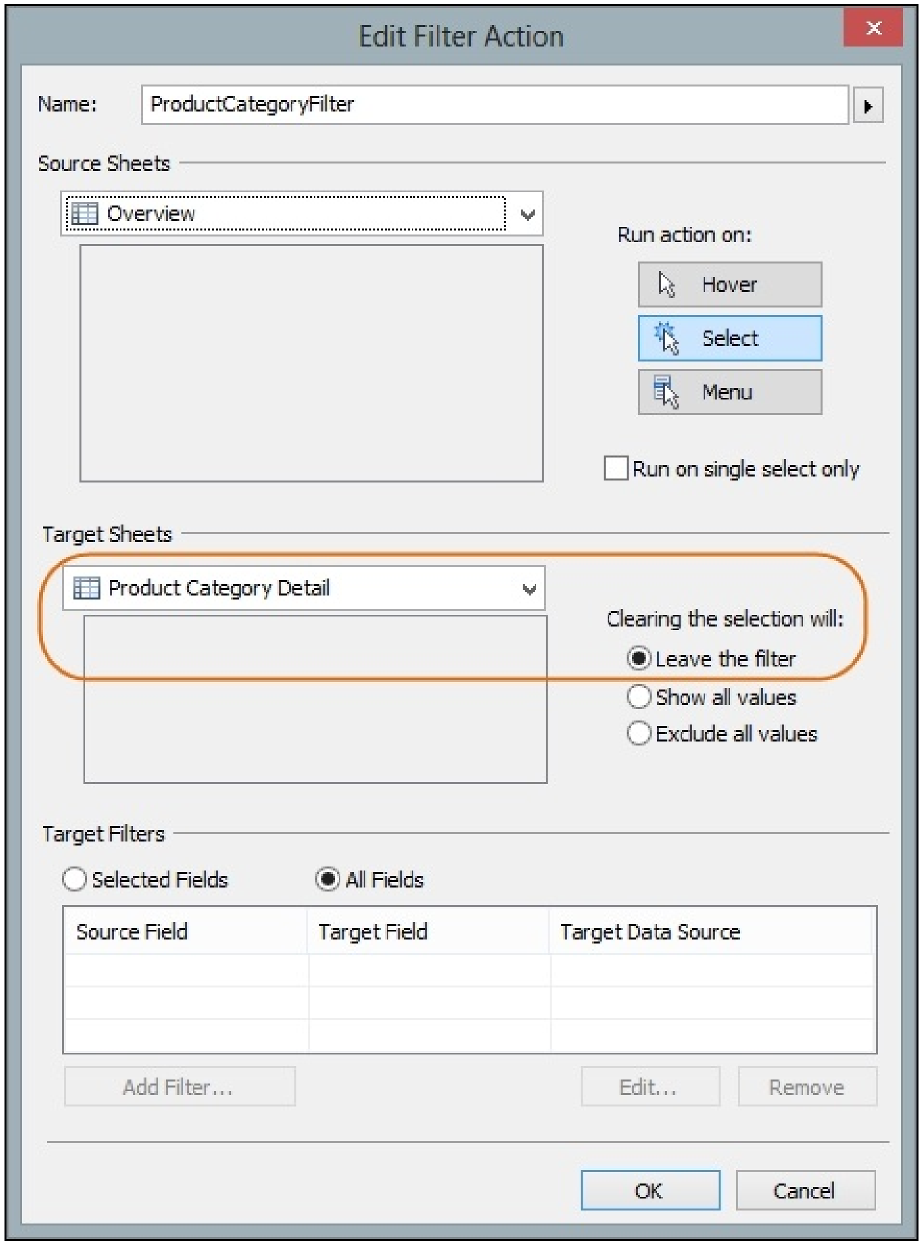
- We should now click on the **Add Action >** button, and this will give us a number of options, as shown in the next screenshot:



We have three options, as follows:

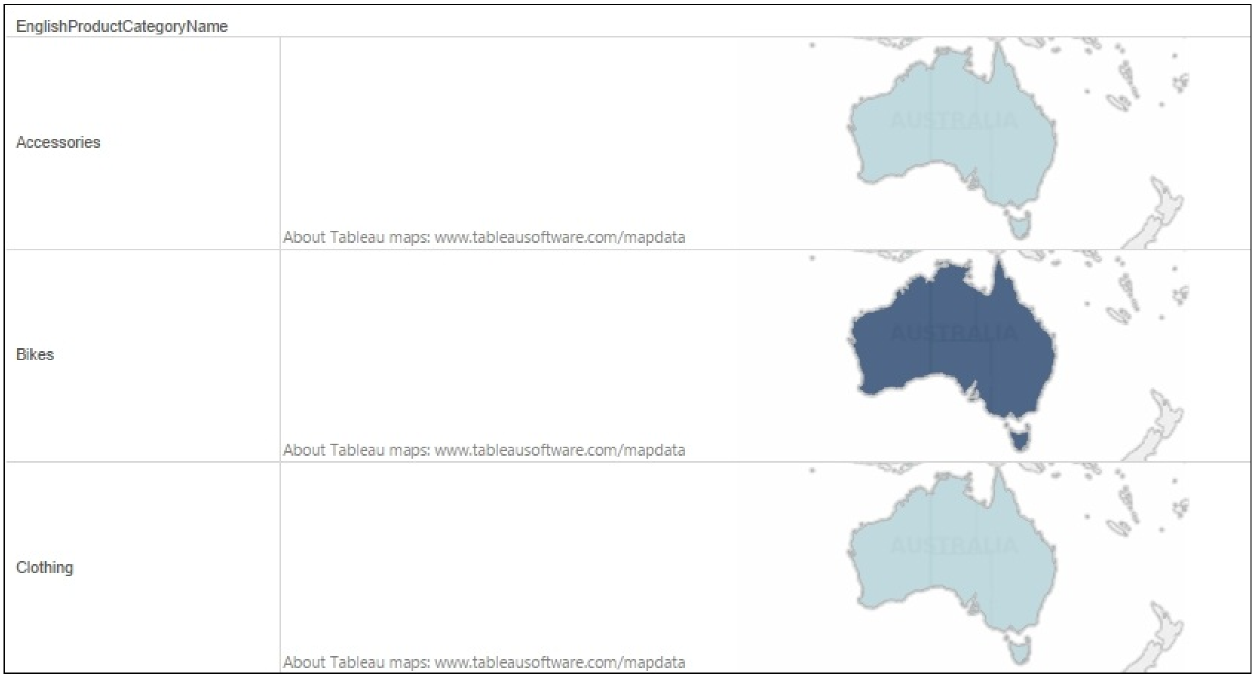
- **Filter**: This action means that you can make a trail between worksheets-from the summary to more specific data— going down to the details
- **Highlight**: This action emphasizes specific data points dependent on the rules you set up
- **URL**: This action allows you to link to external data, such as a website or a SQL Server Reporting Services report

- In this example, we will choose a straightforward **Filter** example. When you select the **Filter** option, you will get the dialog box that appears in the



next screenshot:

- We will set the `Overview` worksheet as the source sheet, and the `Product Category Detail` worksheet as the target sheet.
- When you've selected the correct sheets, click on **OK**, and your action is all set up.
- We can test if the action works simply by going to the `Overview` worksheet and right-clicking on a country, and then Tableau shows you the `Product Category Detail` worksheet. So, for example, if you click on Australia, then the `Product Category Detail` worksheet will appear as you can see in the following screenshot:



- To summarize, we have set up a simple Tableau action that links worksheets together. This improves the user's experience in navigating through the data in that they can get the results from the dashboard quickly.

How it works…

In this recipe, we have used actions so that we can create worksheets that are more detailed than the `Overview` worksheet. Research has shown that people tend to prefer to navigate from the summary data down towards the details, so our page trails will work in the same way.

There's more…

An important item to note is the item marked **Target Sheets**, which was circled in the previous illustration. This feature allows you to preserve the filter or release the filter when the user goes from one worksheet to another. In our example, we preserved the filter. This means that we are facilitating user navigation by going from a summary view to a more detailed, filtered view.

Why did we only choose three colors to represent the **SalesAmount** value? We are not distinguishing the colors at a fine-grained level.

Instead, we are using color to broadly distinguish the value of **SalesAmount** into three categories. The lower values are represented by a light color, and the higher values are represented by a darker color—a more intense blue. Research has shown that people tend to associate lighter colors with smaller values, and more intense, bright, or dark colors with higher values. You can follow this up by looking at *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, *Second Edition*, *Stephen Few*, *Analytics Press*.

Using color in this way does not provide you detail, but it can help you to see patterns in the data very quickly. This is extremely useful for dashboarding.

See also

- *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, *Fifth Edition*, *Shneiderman B.*, *Plaisant C.*, *Cohen M.*, *Jacobs S.*, *Prentice Hall* (2009)
- *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, *Second Edition*, *Few S.*, *Analytics Press*

Guided analytics with Tableau

Industry reports have shown that guided analytics is becoming an increasingly important requirement for mobile business intelligence requirements (Dresner (2013)). What is guided analytics? Guided analytics is defined as cases where knowledge workers can use data models to follow pathways of investigation towards their own results.

Users follow a pathway down the data, starting at a high-level summary of all of the available data, down to the specifics that they are interested in. Users don't have to start with a specific business question, but they can be directed down a pathway towards data that might be interesting and produce results.

Tableau can be used to create dashboards that can be adaptable in response to users' data explorations while also providing them with a pathway that is intuitive and helpful towards the goal of producing results using the dashboard. The topics in this recipe will help you to create dashboards that use guided analysis.

In order to enrich the guided analysis, we will add some new target data, which we can use to compare with the actual data. The comparison between actual and target data is fundamental to dashboards. We will set up actions to highlight fields in one worksheet based on input from another worksheet. We will set up some menu actions, which will pop up some additional menu items in order to help users to flow naturally through the data.

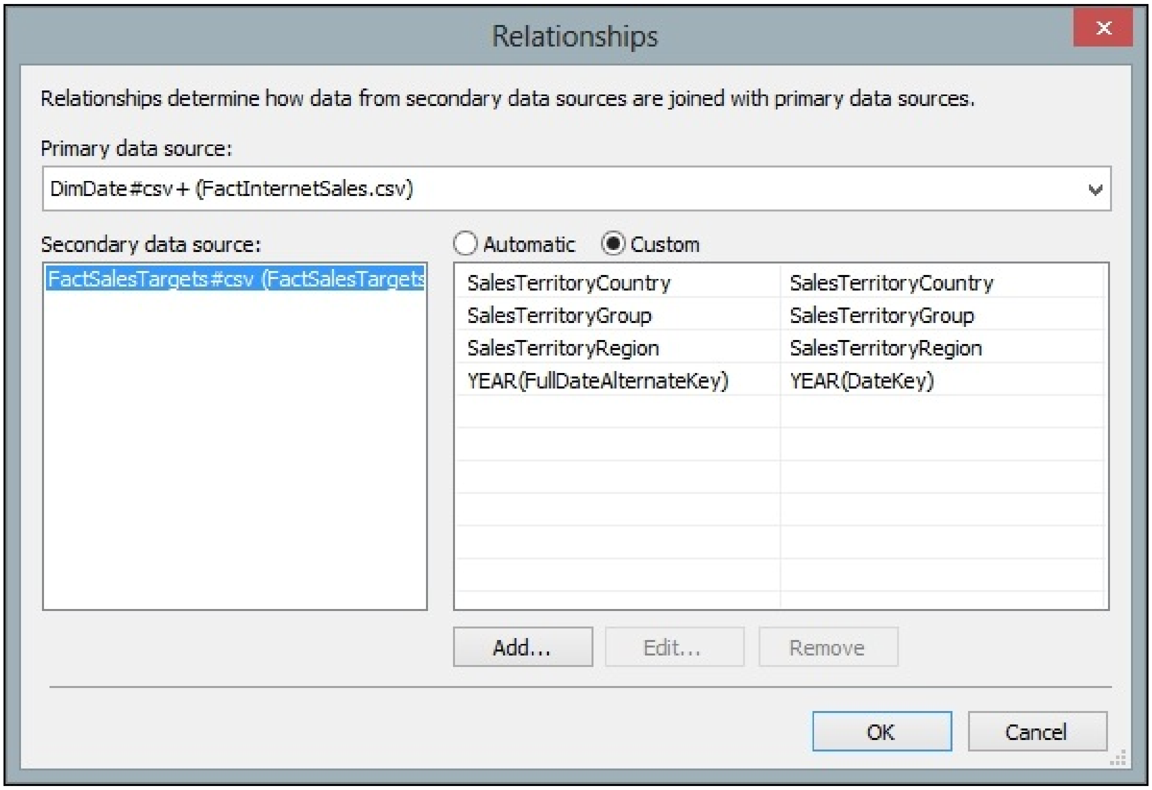
Getting ready…

In this topic, we will use the existing workbook `Chapter Four`.

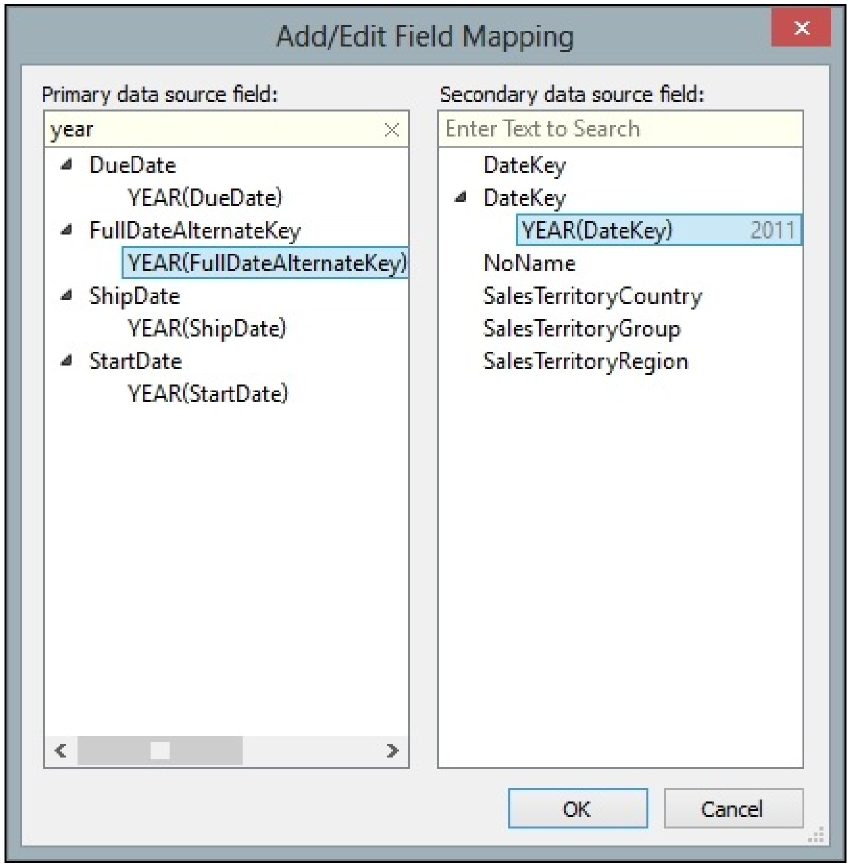
To proceed, we are going to import some more data.

How to do it…

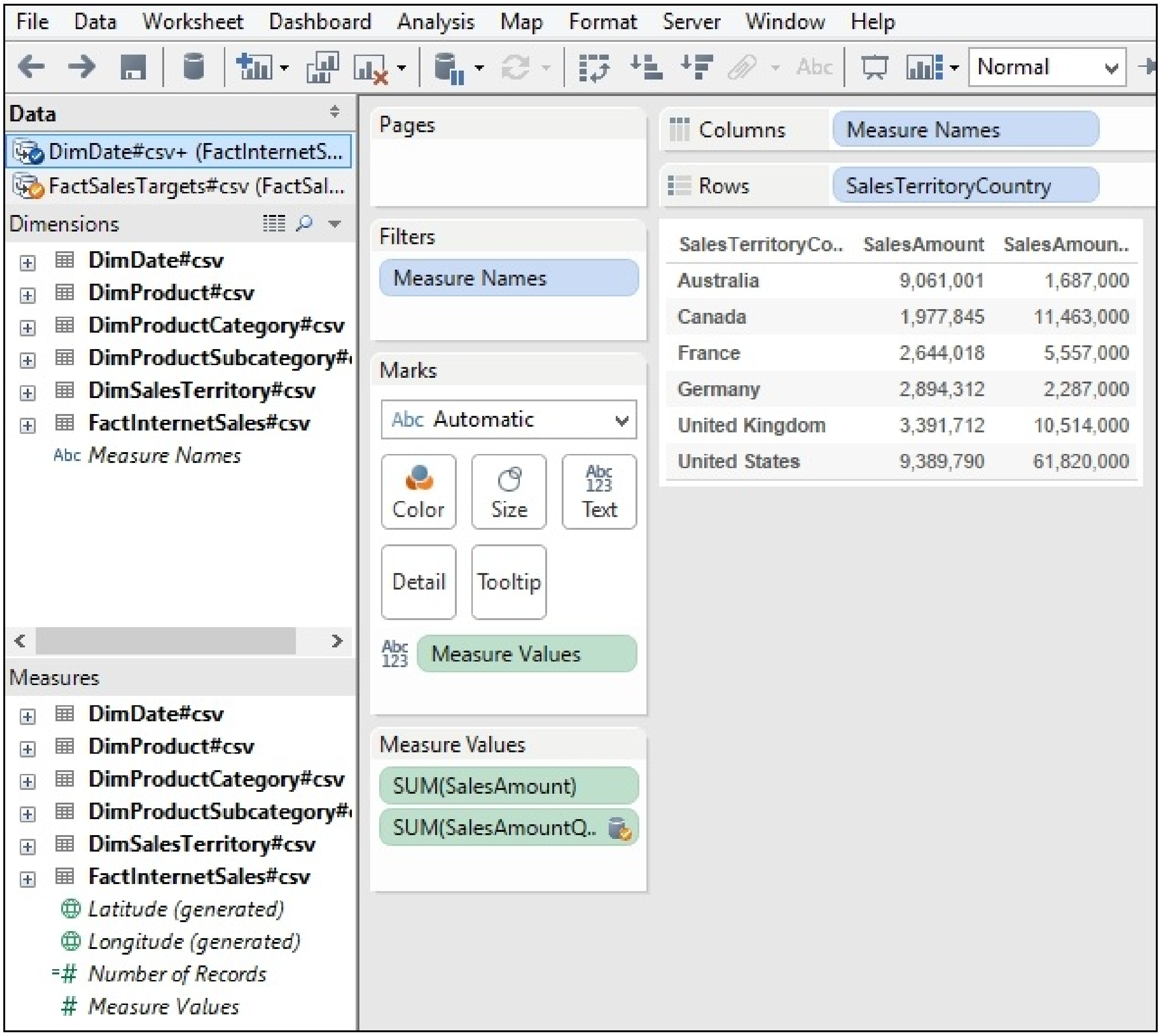
1. Create a new worksheet and call it `KPI Summary` by going to **Worksheet** and then selecting **New Worksheet**.
2. We will import new data that will give us target metrics for the purposes of our visualization. To do this, go to **Data** and then select the option **Connect to Data**.
3. For our purposes, we will select the file called `FactSalesTarget` and open it.
4. Then, we will select the **Import all data** option to import all the data into the worksheet. When we have imported the additional data, the **Data** shelf will hold two connections—one for the original data and another for the target data. We need a way to tie the two connections together. To do this, go to the **Data** menu item and select the **Edit Relationships…** option from the list.
5. The **Relationships** dialog box helps to associate the data sources with each other. The data has the country, region, and group information in common, along with the year. We will need to link the columns together. You may find that the country information, region information, and group information have already been associated with each other by Tableau, and now we simply need to add the year. We are aiming to have the **Relationships** dialog box look like the example in the following screenshot:



- To add the year, click on the **Add…** button. This will bring up the **Add/Edit Field Mapping** dialog box. Then, type `year` in the **Primary data source field** and the **Secondary data source field** dialog boxes. You will find the **YEAR(FullDateAlternateKey)** and the **Year(DateKey)** fields.
- Click on both of these fields and click on **OK**, as shown in in the

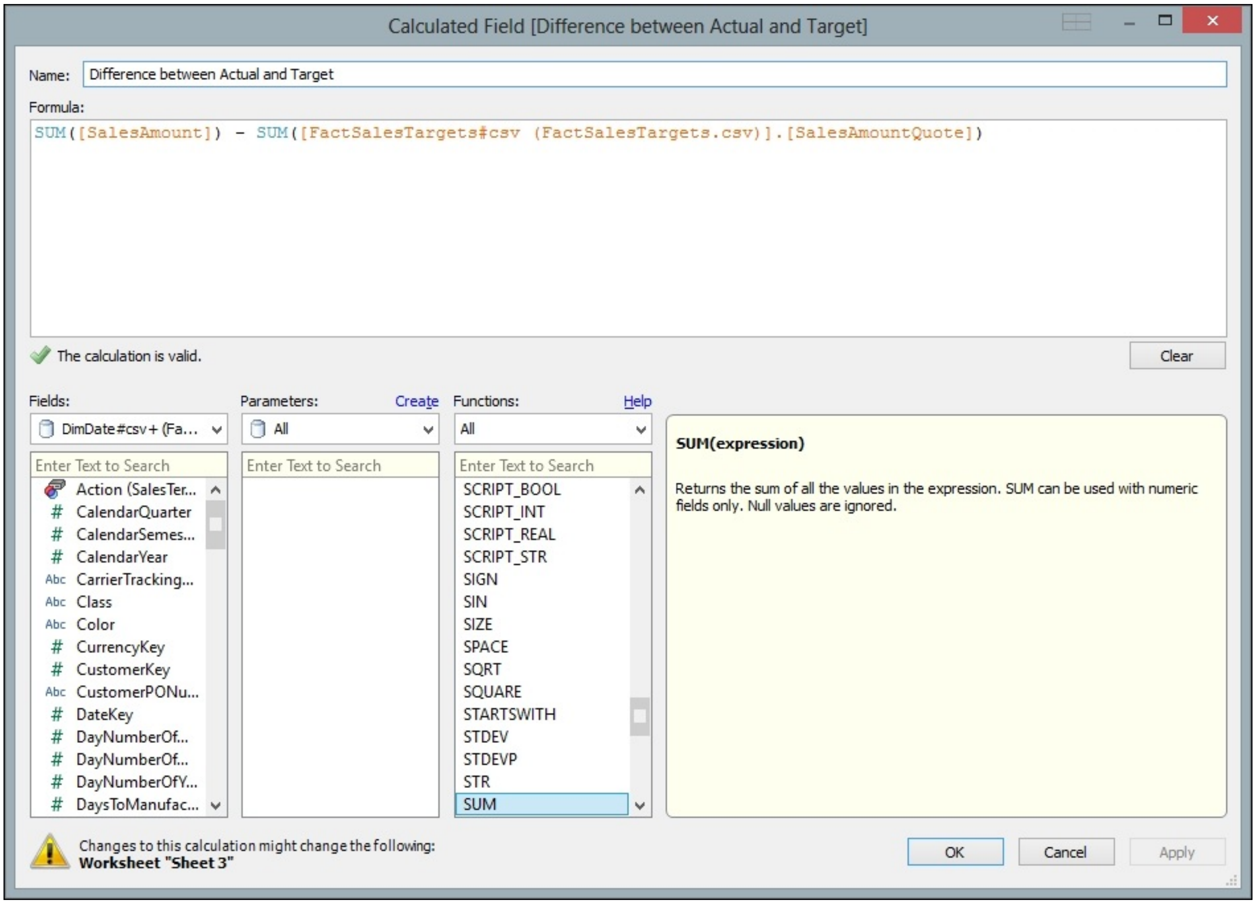
following screenshot: 

- Once we have set up the relationships, we can now add the data to the Tableau canvas. Click on the **DimDate#csv** data source in the **Data** shelf of the sidebar.
- Drag **SalesAmount** from the **Measures** pane and onto the **Column** shelf.
- Click on the **FactSalesTargets#csv** data source in the **Data** shelf of the sidebar.
- Drag **SalesAmountQuote** from the **Measures** pane onto the **Columns** shelf.
- Then, add the **SalesTerritoryCountry** dimension to the **Rows** shelf.
- From the **Show Me** panel, select the table option first of all so that we can see the data.
- You will find that one of the **SalesTerritoryCountry** fields is marked **NA**. Right-click on **NA** and select **Exclude**.
- Your Tableau canvas should appear as shown in the following screenshot:



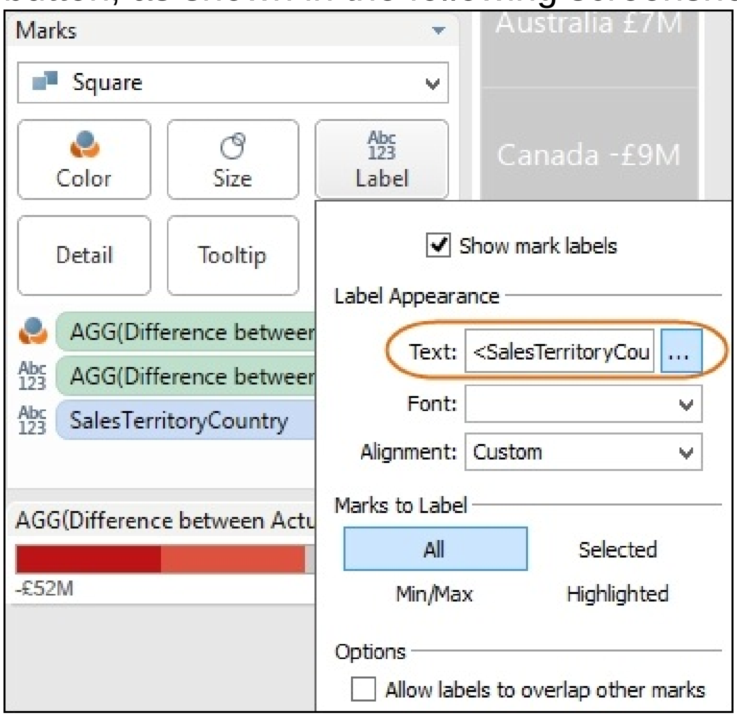
- Now that we have blended our data together, we can create a calculated field in order to show very quickly whether there is a difference between the actual amount and the target amount.
- To set up the calculated field, go to the **Analysis** menu item and select the **Create Calculated Field…** option from the list. This will give us the **Calculated Field** dialog box, which you can see in the next screenshot.
- We will call our **Difference between Actual and Target** calculation, and it is a very simple calculation to do. We will work out the difference between the actual and the target profit. The target profit is called `SalesAmountQuote` in the dataset.
- To do this, we subtract the target sales amount, called `SalesAmountQuote` in the dataset, from the actual sales amount.
- You can see an example in the following screenshot. When you have

set up the calculation, click on **OK**.

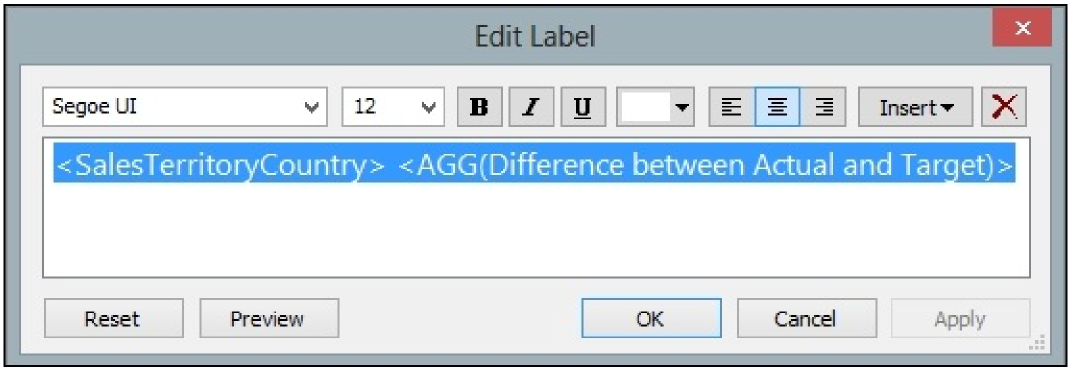


- Now, we have set up our actual metric, our target metric, and the difference between the two. We can now proceed to do some interesting guided analytics in putting the data together.
- We will start by creating KPIs, which will serve as our summary data. From this vantage point, we can drill down into the detailed picture of the data.
- Take our **Difference between Actual and Target** calculation and drag it to the **Color** button.
- Click on the **Color** button and choose the option **Edit Colors…**.
- In the **Edit Colors** dialog box, choose **Red-Blue Diverging** from the **Palette** drop-down list. We will use the red and blue diverging color palette.
- We will use stepped color as before. However, instead of using the default five steps, we will use only three. Enter 3 in the stepped color **Steps** box and then click on **OK**.

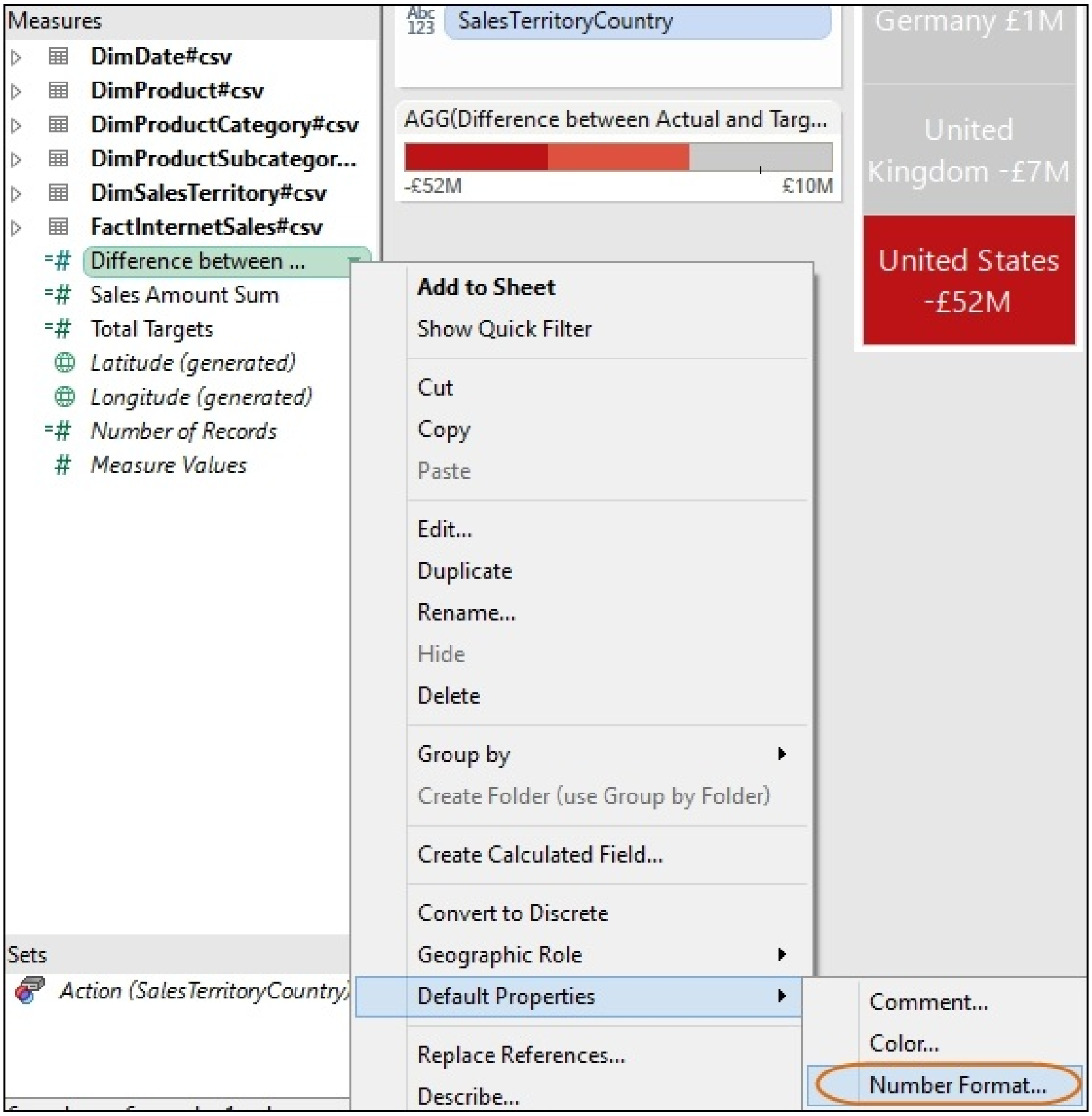
- In the **Marks** shelf, change the mark type from **Automatic** to **Square**.
- Drag the **Difference between Actual and Target** calculation onto the **Label** button. Let's add some labels so that the boxes appear more like KPI tiles.
- Change the color to white and then click on the **Edit Label** dialog box. You will see the change in the table.
- Click on the **Label** button and change the font so that we can use the **Segoe UI** font with size **12**.
- Next, we will make the KPI look more like a tile than a table. To do this, drag the **SalesTerritoryCountry** dimension attribute onto the **Label** button, as shown in the following screenshot:



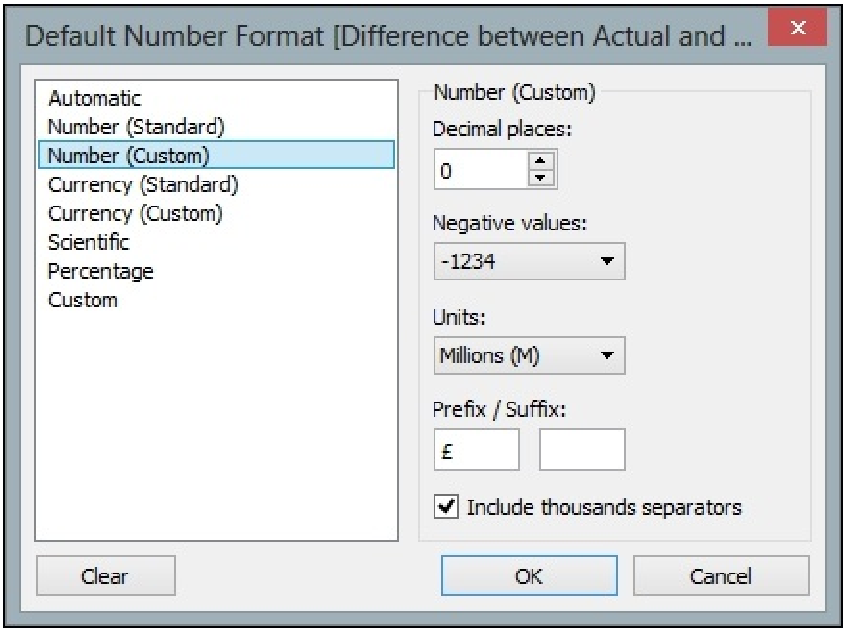
- Now, we will concatenate the **SalesTerritoryCountry** name and the **Difference between Actual and Target** calculation value together so that the label reads nicely. When you click on the blue edit label button, a dialog box appears, which you can see in the next screenshot.
- Next, we will format the text so that we can use the **Segoe UI** font with size **12**.
- When you click on the downward-facing arrow on the **Insert** button, you can select the **SalesTerritoryCountry** name and the **Difference between Actual and Target** calculation in order to show the country and the associated difference.



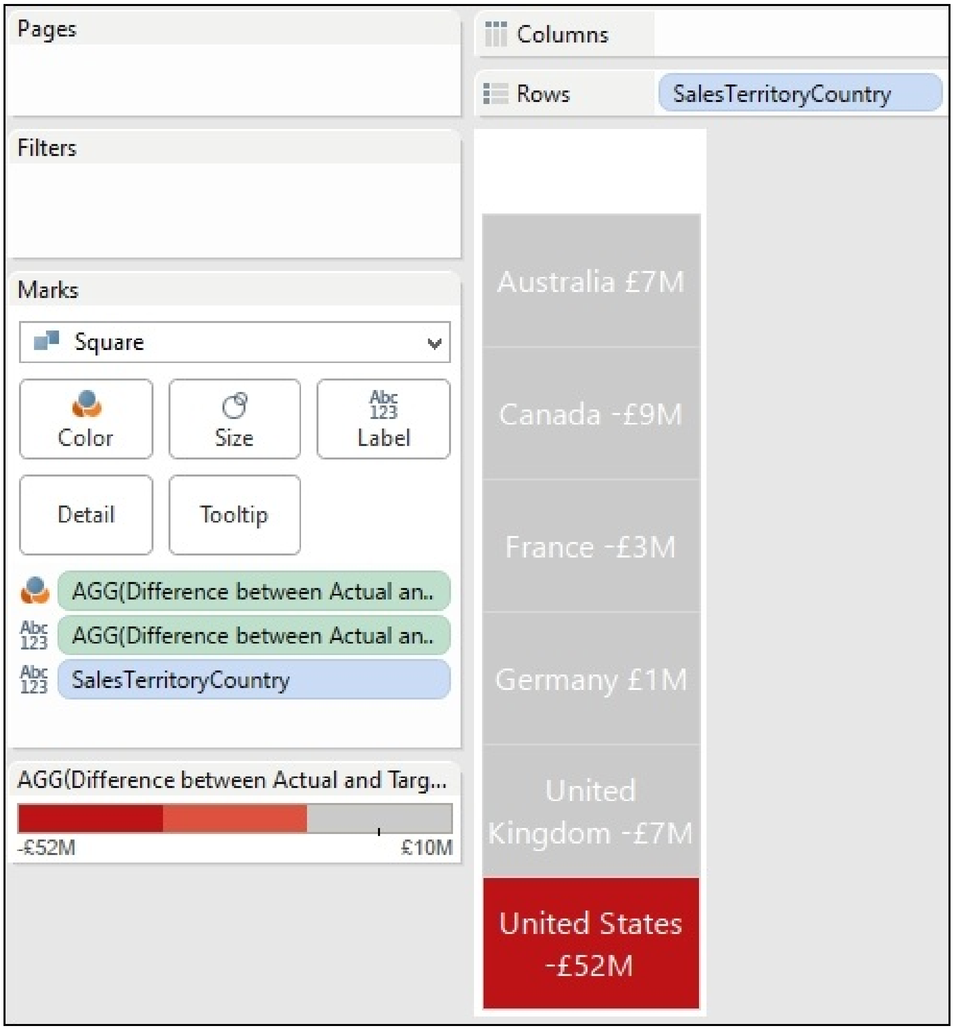
- You will find that the number format for the **Difference between Actual and Target** calculation means that the whole number is shown. This isn't very clear. In order to stay with our principles of eliminating chartjunk, let's amend the format so that it simply shows the million figure. To do this, right-click on the **Difference between Actual and Target** calculation.
- Go to **Default Properties** and click on **Number Format…**. This will bring up the **Default Number Format** box.
- Choose **Number (Custom)**, as shown in the following screenshot:



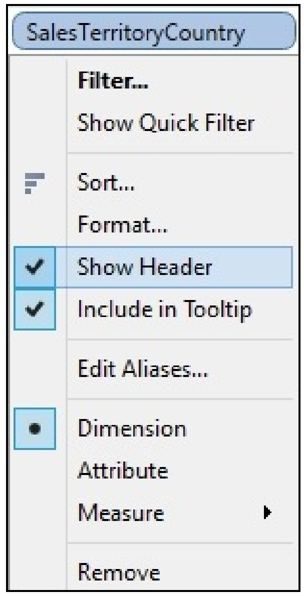
- Firstly, let's get rid of the chartjunk. Let's get rid of the pennies; when we are talking about millions of pounds or dollars, pennies do not matter so much. Let's put the units to **Millions (M)** and prefix the amount with a pound sign or a dollar sign—whatever you prefer! Once you have done this, click on **OK**, as shown in the following screenshot:



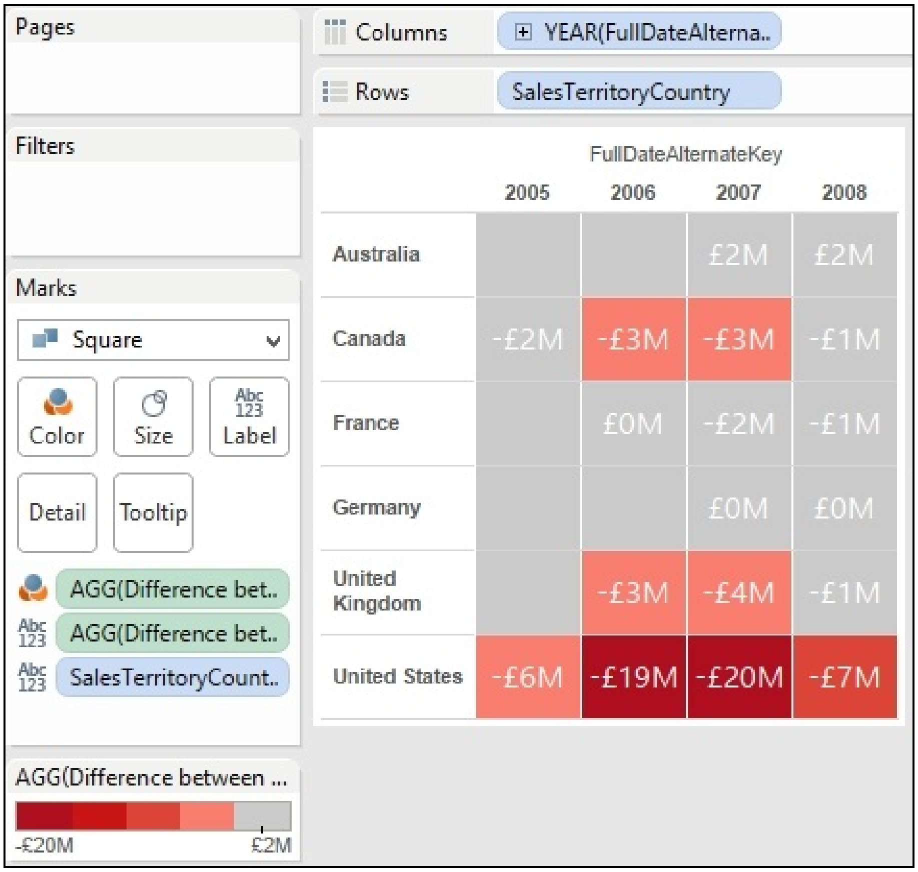
- Now, you will notice that the labels read much more nicely in our KPI tile.
- Remove the row headers from the canvas by right-clicking on the heading and selecting the **Show Header** option.
- Our KPI tiles should now look as shown in the following screenshot:



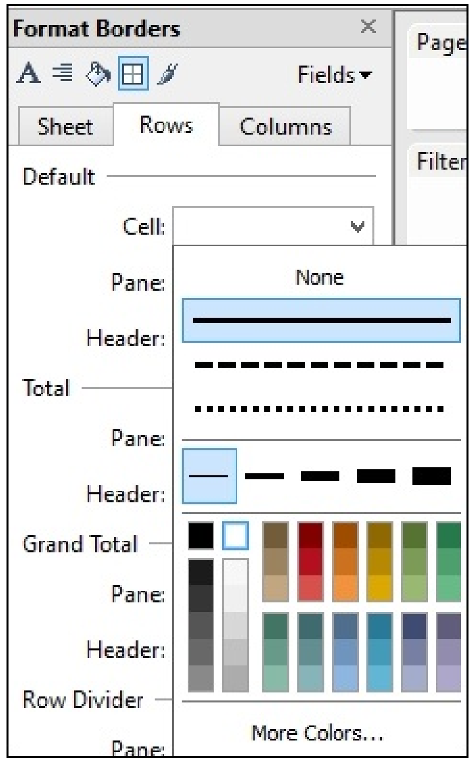
- Let's add some actions so that we can navigate around. Before we move forward, let's duplicate the existing `KPI Summary` sheet and rename it to `KPI by Year`.
- We can now add some actions so that we can navigate around.
- In the new `KPI by Year` worksheet, click on the **DimDate** dimension, look for the **FullDateAlternateKey** dimension attribute, and drag it across onto the **Columns** shelf.



- Let's add the row labels back to the left-hand side so that we can compare more easily. To do this, click on the **SalesTerritoryCountry** dimension member in the **Rows** shelf and select the **Show Header** option.
- Let's change the color settings so that we show five steps rather than three, as per the previous example. This will reveal more fine-grained variations in the data.
- Once you have completed these simple steps, your Tableau visualization will appear as shown in the following screenshot:



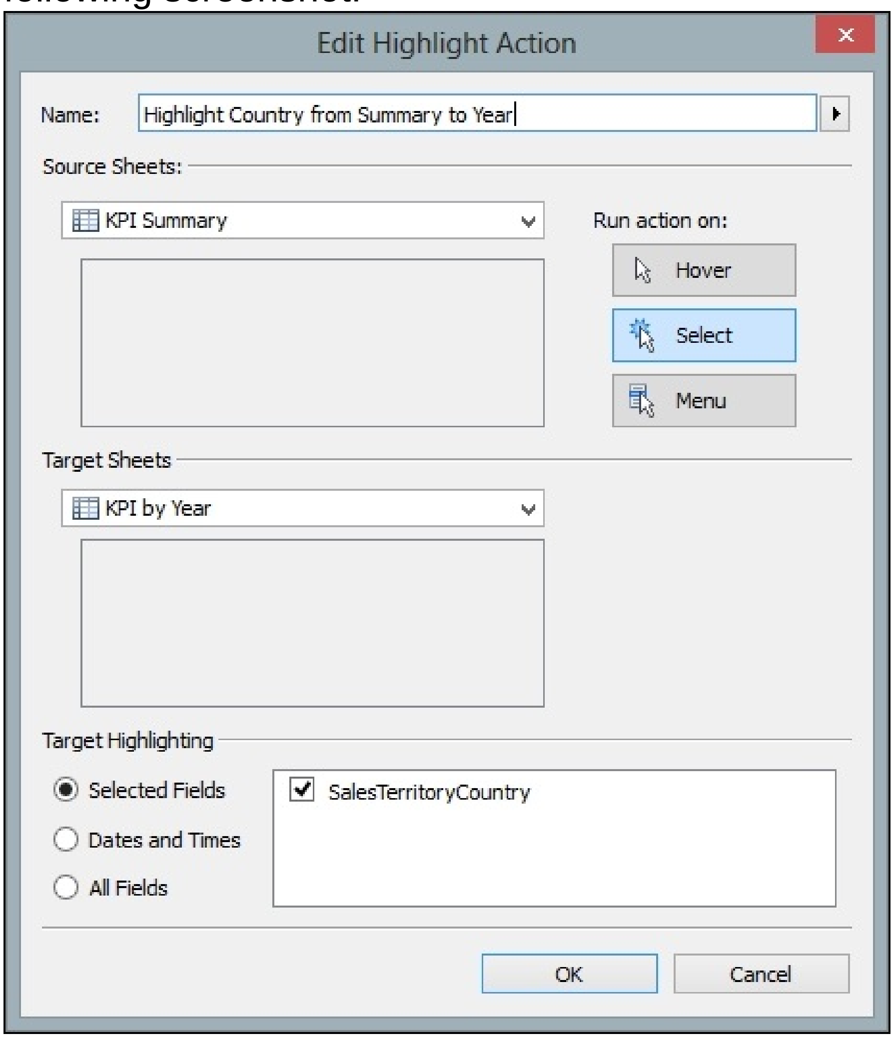
- Let's add the row labels back to the left-hand side so that we can compare more easily. To do this, click on the **SalesTerritoryCountry** dimension and select the **Add Headers** option.
- We can also format the borders of the cells so that they are neater. To do this, click on the **SalesTerritoryCountry** dimension and select the **Format…** option.
- Using this dialog box, you can neaten up the borders by making them all white in order to give a nice finish to the cells. You can see how to do

this in the following screenshot: 

- Let's add another worksheet which will provide more detail. Duplicate the `KPI by Year` worksheet and call it `KPI by Q`. This only requires a few simple changes in order to provide a more detailed chart.
- To start, simply click on the **+** sign on the **Year(FullDateAlternateKey)** dimension so that we see the **Quarter(FullDateAlternateKey)** dimension exposed on the **Columns** shelf. Now, we will see more columns on the visualization.
- To summarize, we now have three sheets: `KPI Summary`, `KPI by Year`, and `KPI by Q`. We will now set up actions in order to link them together. Our Tableau actions will respond to user input. For example, if we click on **Canada** on the `KPI Summary` worksheet, then we can see that the Canada fields are highlighted on the `KPI by Year` worksheet. This is a very simple action which helps to draw attention to particular aspects of the dashboard, thereby helping the user interaction flow.
- In all, we will create five actions that will facilitate all of these activities. Let's start with the actions to highlight fields. To do this, let's first go to the `KPI Summary` worksheet and begin adding some new actions by going to the **Worksheet** option and choosing the **Actions…** option.
- Let's add in some highlight actions in the **Actions** dialog box by clicking on the **Add Action >** button and choosing the **Highlight…**

option. We will create three highlights, one for each of the worksheets: `KPI Summary`, `KPI by Year`, and `KPI by Q`.

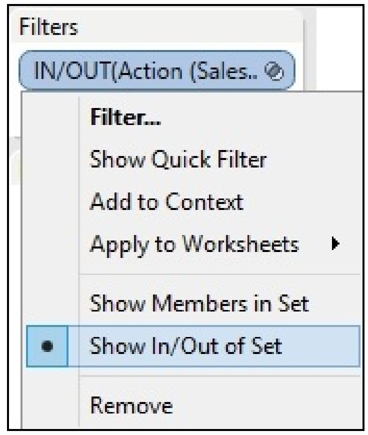
- For the first filter, let's rename it to `Highlight Country from Summary to Year`. We will choose to run the action on the **Select** action. Otherwise, we will have a lot of noisy actions generated if the user simply hovers over a data point. This might mean that the user loses his/her thread in navigating the data, and we do not want that to happen.
- For the **Source Sheets** option, let's stay with `KPI Summary`. For the **Target Sheets** option, let's choose `KPI by Year`. In the **Target Highlighting** section, let's choose to highlight some fields, which will be done by selecting the **Selected Fields** option. Once you've done this, your **Edit Highlight Action** dialog box should appear as shown in the following screenshot:



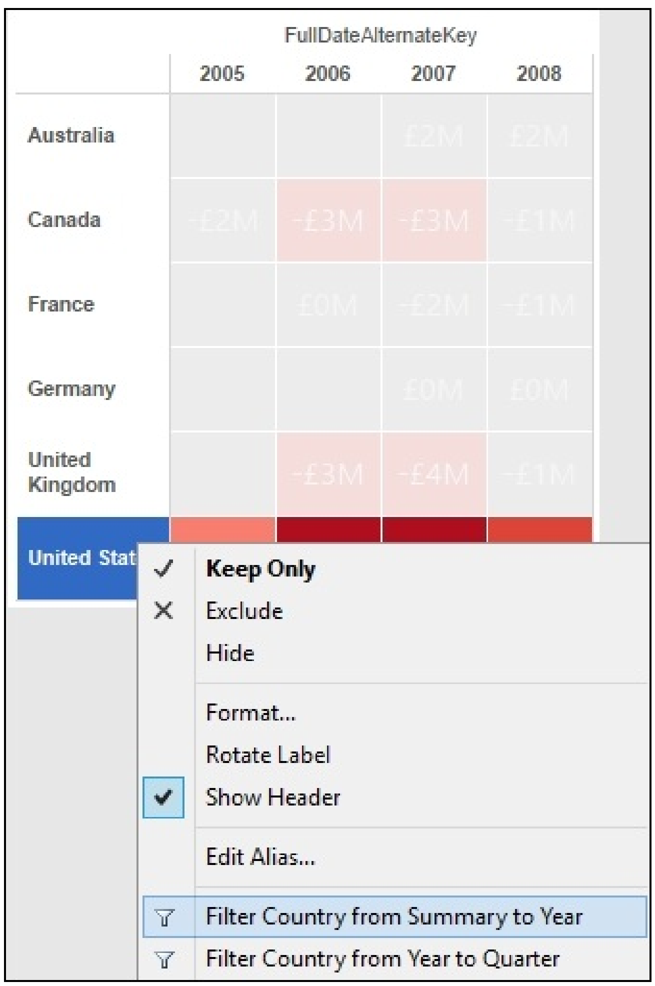
- For the remainder of the actions, let's repeat them using the details given in the following table:

Name	Run on	Source sheet	Target sheet	Fields
`Highlight Country from Summary to Q`	Select	`KPI Summary`	`KPI by Q`	`SalesTerritoryCountry`
`Highlight Country from Year to Q`	Select	`KPI Summary`	`KPI by Q`	`SalesTerritoryCountry.` `YEAR(FullDateAlternateKey`
`Filter Country from Summary to Year`	Menu	Connection name, for example, `DateDate#csv` + (`FactInternetSales`)	`KPI by Year`	`SalesTerritoryCountry`
`Filter Country from Year to Quarter`	Menu	Connection name, for example, `DateDate#CSV` + (`FactInternetSales`)	`KPI by Quarter`	`SalesTerritoryCountry`

- Once you have repeated the creation of the actions with these attributes, we can test out some action scenarios.
- Let's test out the highlight actions first. On the worksheet, let's click on United States on the `KPI Summary` worksheet. When we do this, we can go to the `KPI by Year` worksheet, and you can see that United States is highlighted.
- You can vary which countries you show in the canvas by clicking on the action lozenge in the **Filters** shelf. For example, you could choose the option **Show In/Out of Set** to decide which details you want to show,

as shown in the following screenshot: 

- In setting up these actions, you have configured additional menu options as well. So, for example, if you right-click on the United States row in the `KPI by Year` workbook, you will be presented with a number of options for filtering, as shown in the following screenshot. These menu options were set up automatically for you by Tableau when you set up the menu actions.



- So, if you select the **Filter Country from Year to Quarter** option, then you are taken to the `KPI by Q` worksheet, and you can see the relevant data for the United States. Note also that the years are highlighted; this is as a result of the addition of these fields in the **Selected Fields** option.

How it works…

To summarize, we have set up a lot of different options for navigating around worksheets and dashboards using Tableau actions. You can see that Tableau is extremely flexible and adaptable for customization, which helps users to get results from their dashboard explorations.

In setting up these actions, you have configured additional menu options as well. So, for example, if you right-click on the United States row in the `KPI by Year` worksheet, you will be presented with a number of options for filtering. These menu options were set up automatically for you by Tableau when you set up the menu actions.

When we changed the color, we removed the blue color completely from the visualization, and only red and grey remained. Using a neutral color conveys the message that those data points are OK since they have been brushed out. This means that the strong red color, representing the high negative difference between actual and target totals for the United States here, is emphasized by default.

Depending on your version of Tableau, the **Label** button may be called the **Text** button. However, the rest of the items have stayed the same.

See Also

- *2013 Wisdom of Crowds Business Intelligence Market Study: Buyer's Guide Edition*, *Dresner, Howard*, *CreateSpace Independent Publishing Platform* (2013)

Sharing your results in a meeting

Dashboards are designed to communicate the meaning of the data to the decision maker. However, decisions are not usually made by one person, which means that dashboards need to be shared. This is very easy if you all have Tableau.

What about the case, however, where not everyone has Tableau? What do you do then in order to share your results? In Tableau, it is very easy to share your results in PowerPoint and via PDF files.

In this topic, we will look for different ways in which to embed a PowerPoint presentation into a Tableau workbook. Later on in this book, we will go through ways of embedding a Tableau workbook in a PowerPoint presentation, but in the meantime, we will start with the simplest option. Embedding a PowerPoint presentation in a Tableau workbook can help to share results since it might help to add context.

Sometimes, you may need to consider printing Tableau workbooks, and we will look at tips to make this easy.

Getting ready

In this recipe, we will continue to use the workbook we created for previous recipes in this chapter.

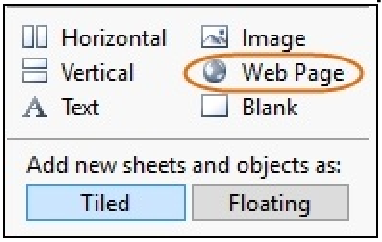
Before you proceed to do anything in Tableau, you should upload a PowerPoint presentation of your choice to SkyDrive. If you do not have a SkyDrive account, then you are welcome to use the following link: http://bit.ly/TableauChapter4

How to do it...

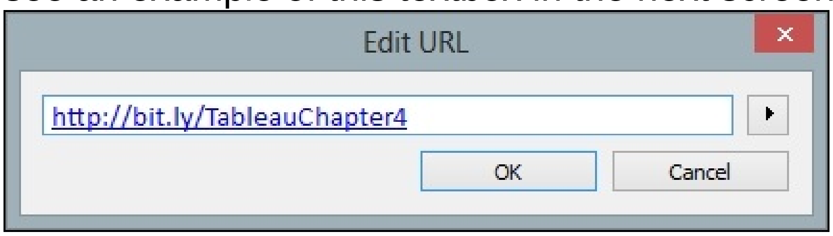
1. In this recipe, create a dashboard in the `Chapter Four` Tableau workbook.
2. First, we will add a web page component to the dashboard canvas

by clicking on the **Web Page** element on the left-hand side of the dashboard canvas.

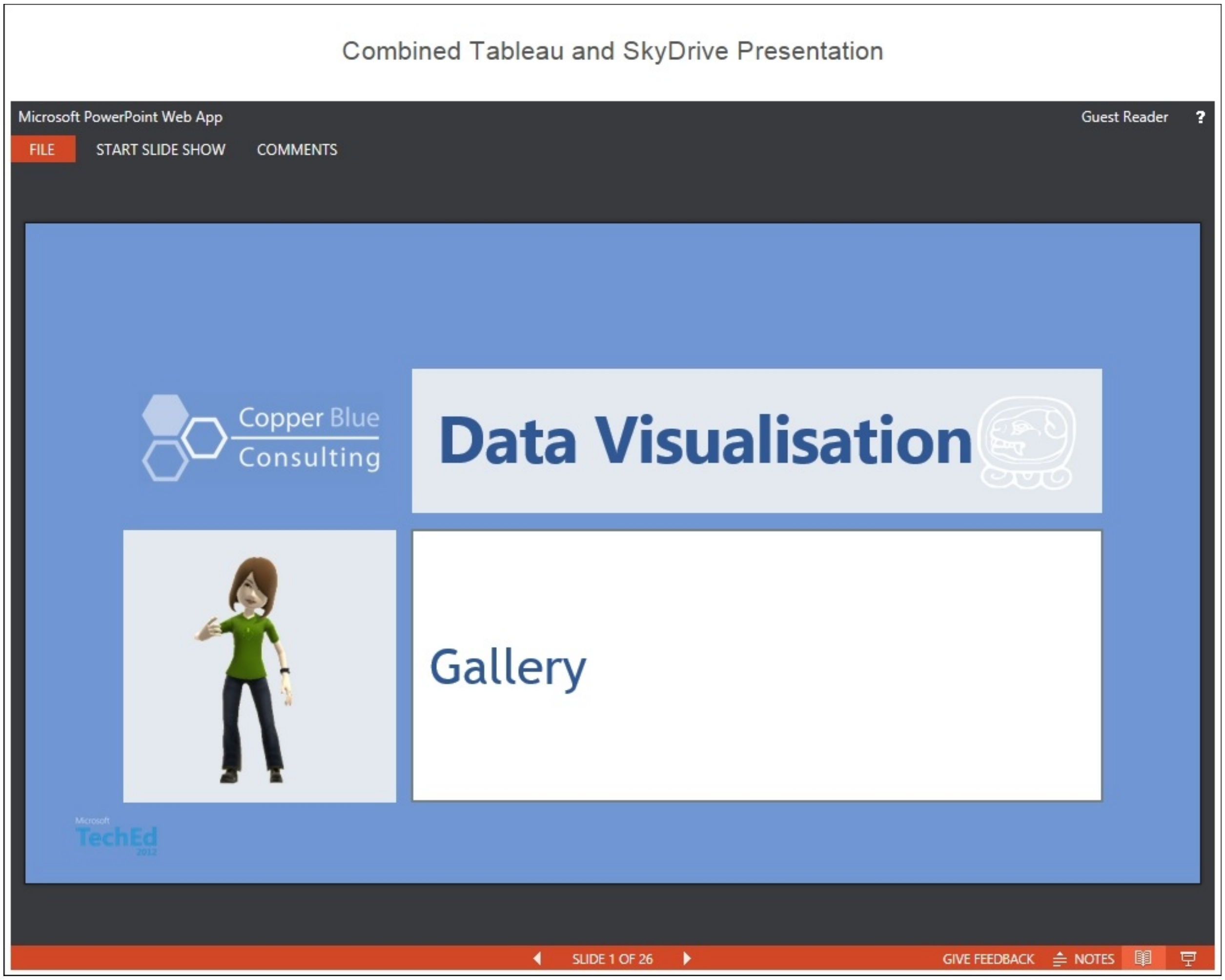
3. You can see an example of this element in the following screenshot:



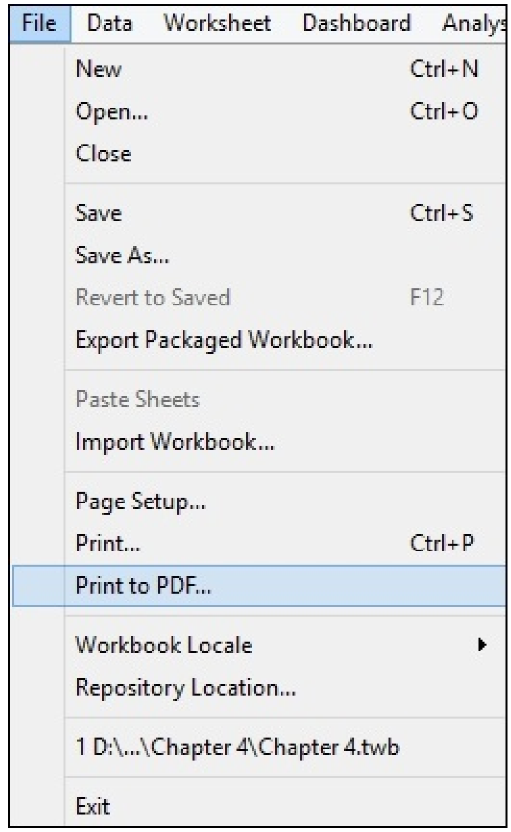
- When you drag the **Web Page** element onto the canvas, Tableau will present you with a textbox where you can put the website URL. You can see an example of this textbox in the next screenshot:

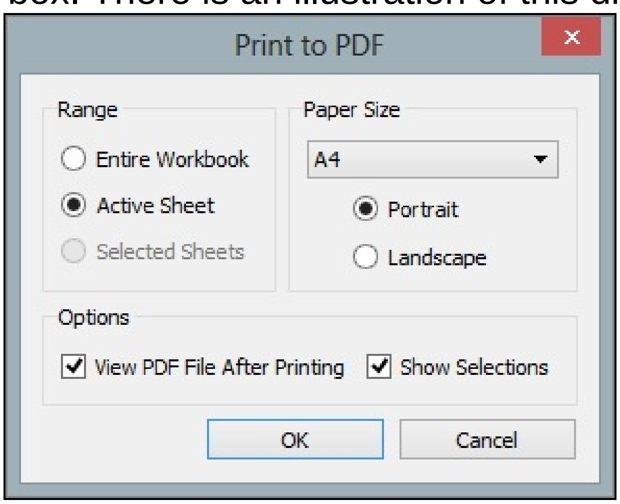


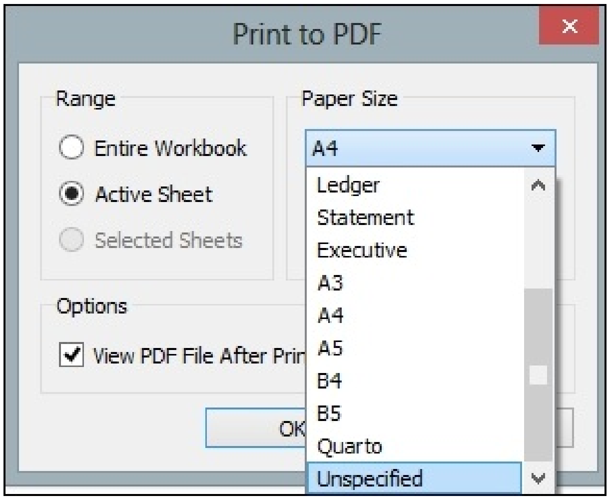
- You can also add a textbox to the sheet and give it a title. When you have added a title, the Tableau dashboard will appear as follows:



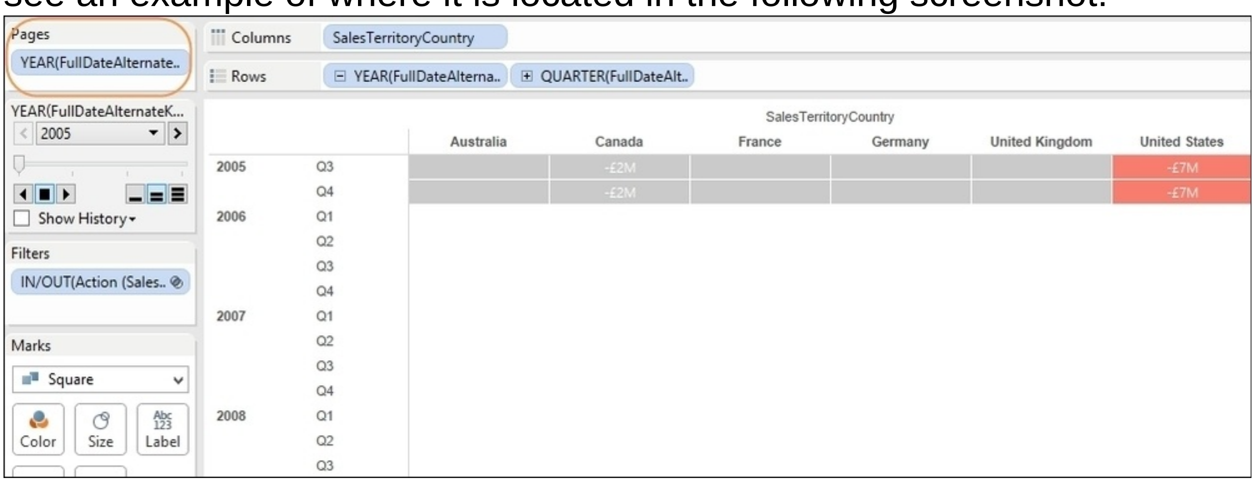
- If you like, you can add some of the worksheets onto the dashboard in order to combine Tableau worksheets with the PowerPoint presentation.
- What happens if you need to print out the Tableau workbook? To do this, go to the **File** menu item and select **Print to PDF…**. You can see an

example of this in the following screenshot: 

- When you select the **Print to PDF…** option, you are presented with a small dialog box. There is an illustration of this dialog box in the next screenshot: 

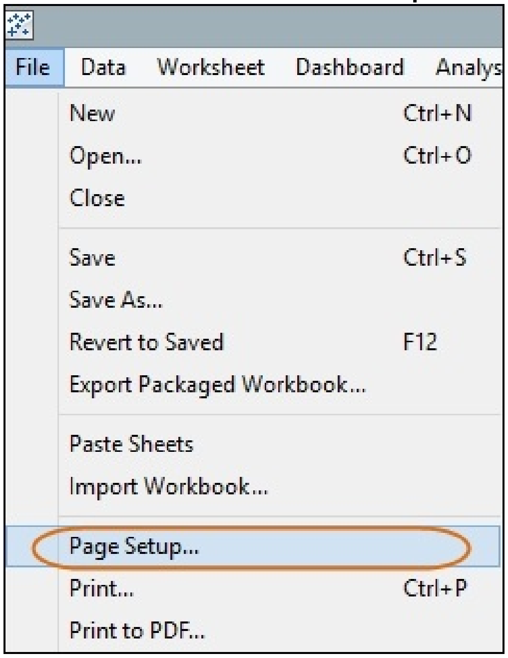
- There are a few tips to note, however, when printing PDF files. For example, in the **Paper Size** drop-down list, there are a number of page sizes available. You can see a sample of the drop-down list in the next screenshot: 

- The highlighted option is **Unspecified**, and it is recommended that you do not select this option, or the paper size of your Tableau workbook will stretch and stretch!
- You can also use the dimensions in order to split the printout into different pages. For example, you could have a new page per year, or per country as in our example here. Let's go through an example of this Tableau feature now.
- Let's duplicate the `KPI by Q` worksheet and swap the dimensions around so that the **SalesTerritoryCountry** dimension is on the **Columns** shelf and the **DimDate** dimension attributes are on the **Rows** shelf.
- Now, let's drag the **Year(FullDateAlternateKey)** dimension attribute onto the **Pages** shelf, which is located above the **Filters** area. You can see an example of where it is located in the following screenshot:

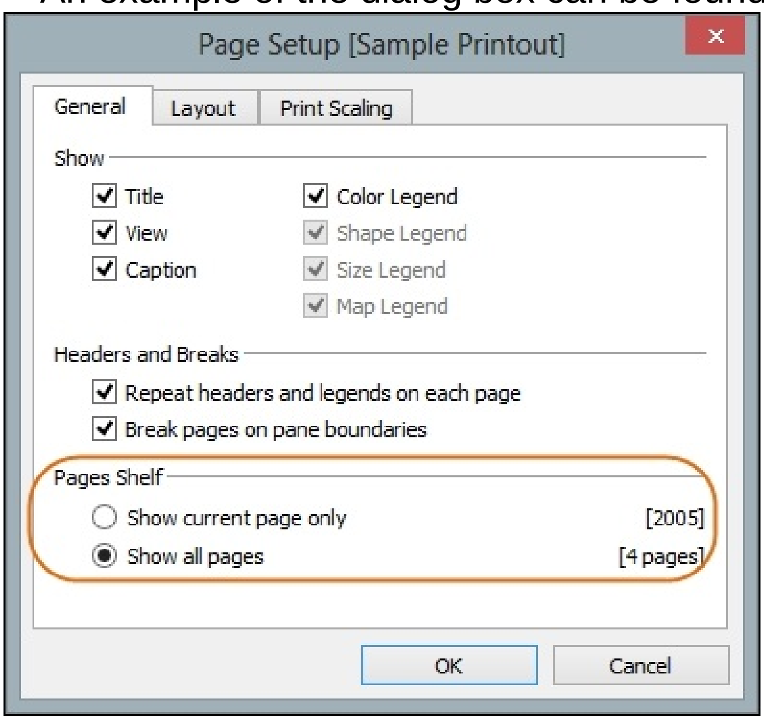


- We need to do an extra bit of work so that the PDF printout will have a new page per year. To do this, go to the **File** menu item and then **Page Setup…**.

- You can see an example of this location in the following screenshot:



- When you select this option, you are presented with a dialog box where you can specify how you would like the pages to be printed out.
- An example of the dialog box can be found in the following screenshot:



- For our purposes, we will select the **Show all pages** option under the **Pages Shelf** configuration since this will print out a new page for each dimension member. In our example, this will print out a new page for each year.

- Once you have clicked on the **OK** button, you should now proceed to choose the **Print to PDF…** option, which you can find under the **File** menu option.
- Tableau will ask you where you would like to store the file, and once you have saved it to the disk, you can review your file.
- To summarize, Tableau has facilities that will help you to use dashboards as a tool for making decisions and sharing those insights with your colleagues. Even in today's world, some people work best with a printout, and it is good to be able to serve these people in the way that they understand data best.

How it works…

The analytical power of the dashboard is amplified if it is shared with the right audience. Tableau is a great data visualization tool, but you also need to get the message out to the world. Adding it as part of PowerPoint means that you can add contextual information to the dashboard.

By embedding the dashboard and the PowerPoint presentation together, you might find that you end up with scroll bars. If this is the case, be careful that you are not scrunching everything together on one page. If so, it is time to rethink what you are showing on the screen. Less is more! You could think about increasing the interactivity so that it is not so confusing and scrunched up. You could also think about using filters or highlights in order to emphasize some data over others.

You can be careful with the dashboard size by emphasizing the most important data through the careful use of color, size, shape, or the inclusion of annotations, which we will discuss in the next recipe.

Notes and annotations

Annotations are little notes that help us to inform and engage the dashboard consumer, and they can make the difference between understanding the dashboard and even rejecting it due to a misunderstanding.

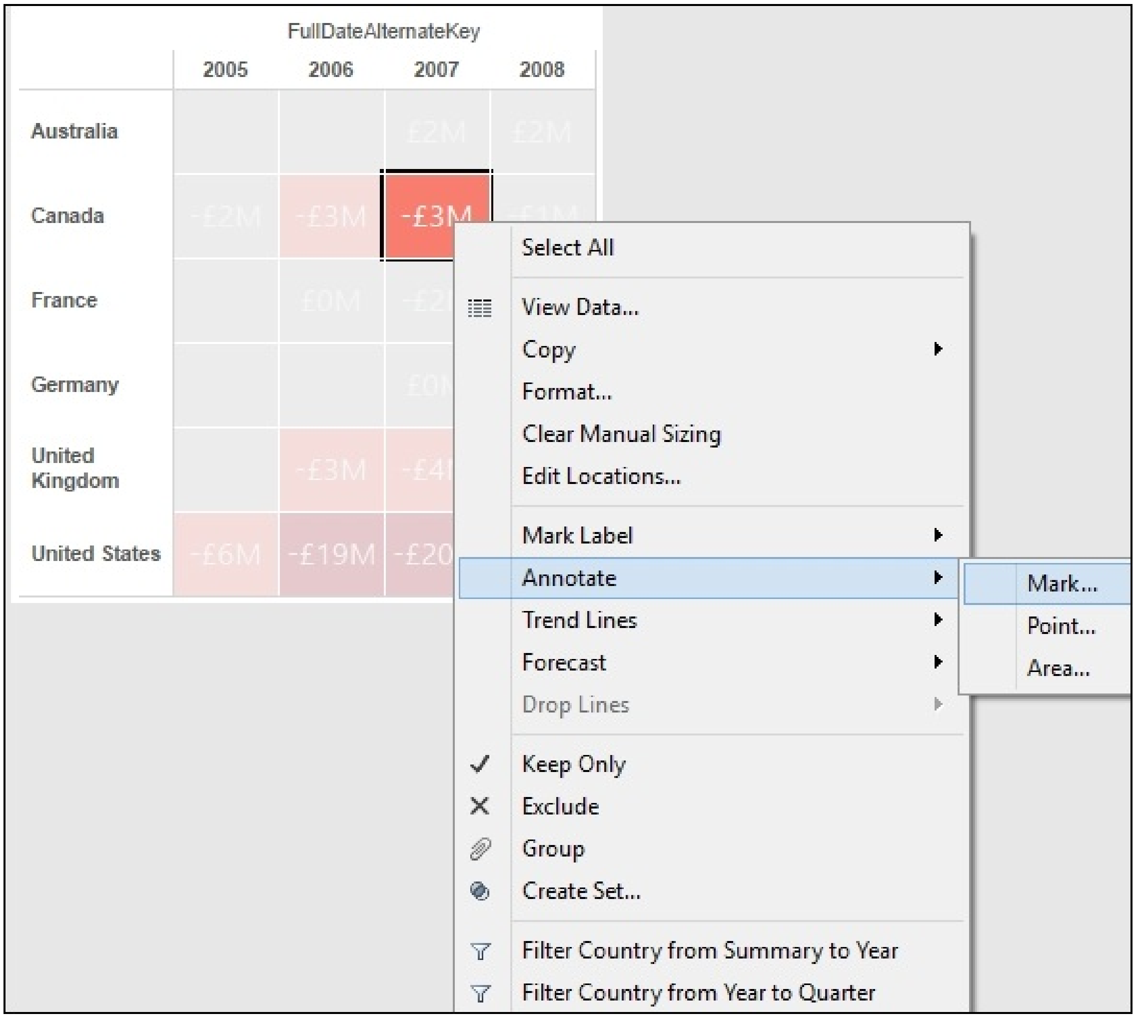
Annotations are useful for all sorts of reasons. For example, you might find them useful for jotting down ideas on the Tableau dashboard as we go along. Annotations can help to add context to your dashboard. For example, if a sales figure is an aberration in some way, then we can add an annotation to explain why this result has been found.

Getting ready

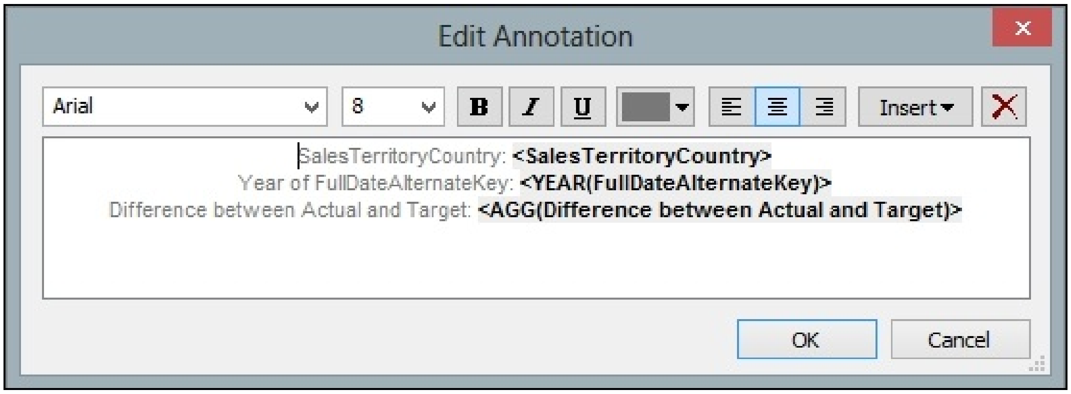
For the exercises in this chapter, we will continue to use the `Chapter Four` workbook. In this recipe, we will look at creating and formatting annotations. We will use the `KPI by Year` worksheet.

How to do it…

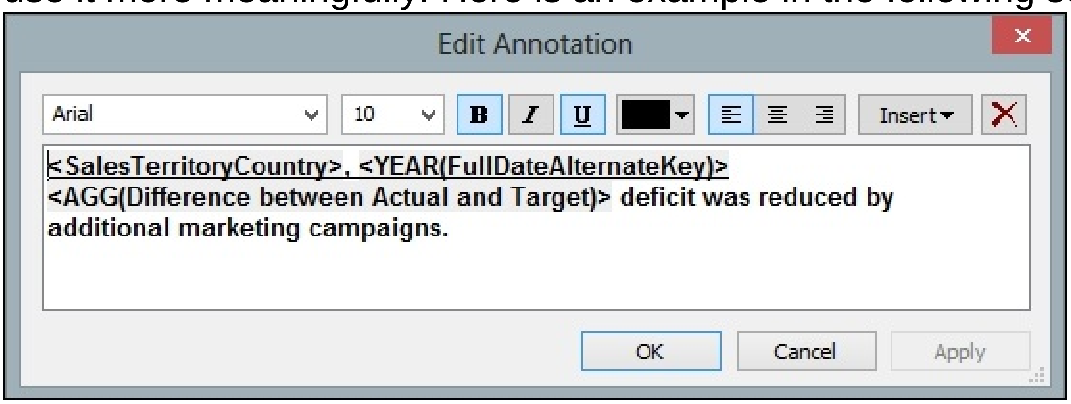
1. On the `KPI by Year` worksheet, right-click on the country **Canada** for the year **2007**.
2. On the pop-up menu, you will get the **Annotate** option, as shown in the following screenshot:



- You have three options for how you would like to annotate: by the mark, the point, or the area. When you annotate using the **Mark…** option, this means that the specific highlighted mark will be annotated. Alternatively, if you would like to annotate a specific point on the canvas that may not be at a specific data mark, select the **Point…** option. Finally, you could annotate an area of the canvas, and for this, you would choose the **Area…** option. For our example, we will annotate the specific mark for **Canada** in **2007**. When you choose the **Mark…** option, you are presented with a text editor with some pre-populated variables. You can see an example of this in the next screenshot:



- We can amend the default annotation, however, so that we can provide some more meaningful information. We can move around some of the features to make it more readable and add some additional text. For example, if we make the text black and add or remove some of the default text, we can take up the same amount of space on the page but use it more meaningfully. Here is an example in the following screenshot:



- When we click on **OK** and return to the main screen, we can see that our annotation has been added. When we click on the mark for **Canada** in **2007**, we can see the mark, as shown in the following screenshot:

- If the annotation is deemed noisy and is distracting, then we can try to reduce the font and move it to a location where there is less noise. Here is an example in the following screenshot:

- To summarize, annotations can be a useful way of adding some context to the data visualization. One barrier for data visualization is that data consumers can become bogged down in the details if they do not understand a finer point. Adding an annotation can help the conversation to move forward and hopefully help in the dashboard consumer getting results from your visualization.

Using external data to enrich your dashboard

We can provide a better perspective on data by adding external data sources. There are many applications where external data is extremely useful to the organization. For example, your marketing organization may be interested to know how your marketing campaigns impact your website, which in turn impacts your sales. Tableau is well suited to this type of analysis, and it provides a range of connectors so that you can mix external data sources together.

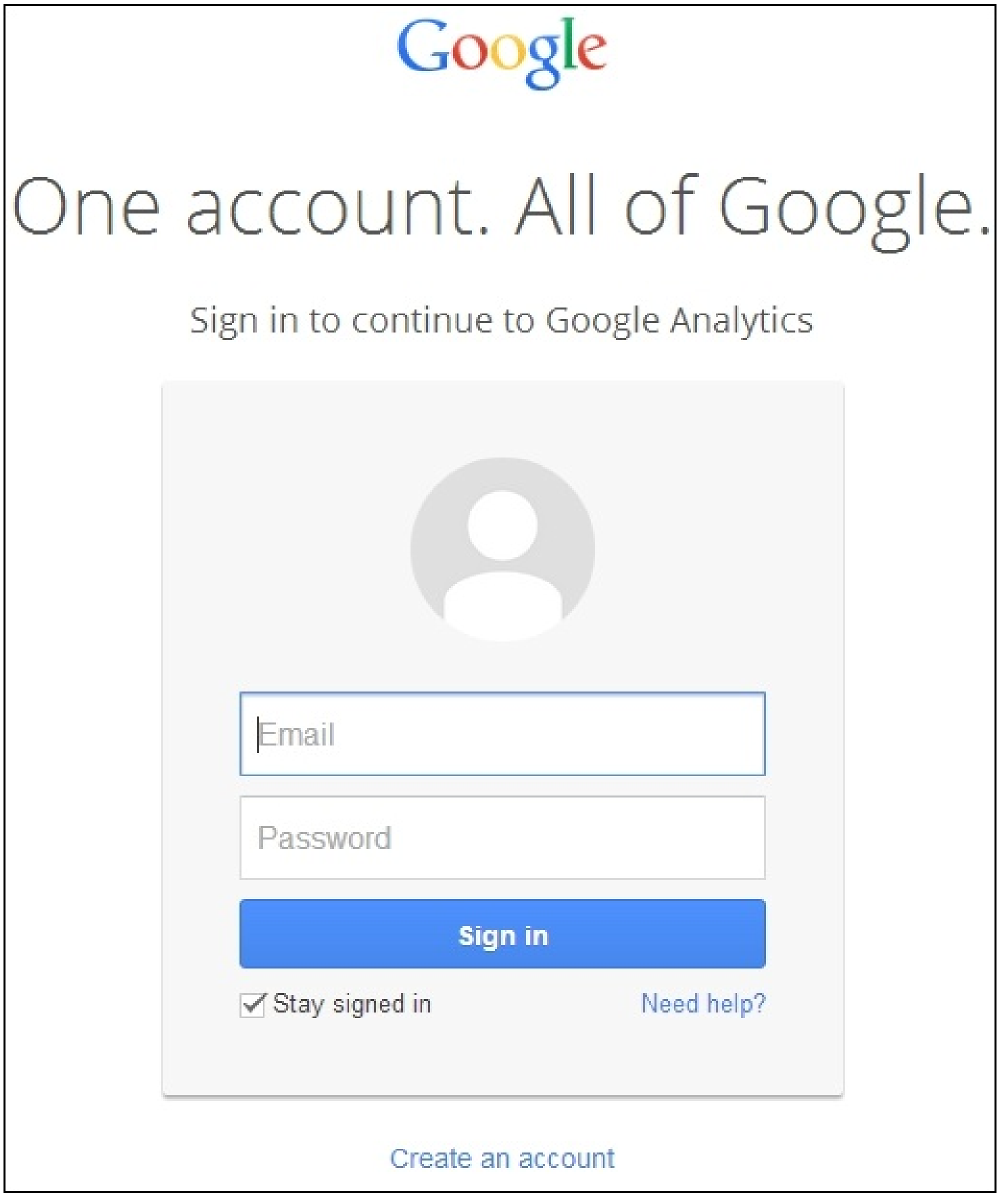
In an earlier chapter, we retrieved the country data from the Microsoft Windows Azure Datamarket. In this recipe, we will take a look at retrieving data from Google Analytics using the Tableau connector. Google Analytics is a tool that helps you to analyze traffic to your website, and many businesses make use of this free facility.

Getting ready

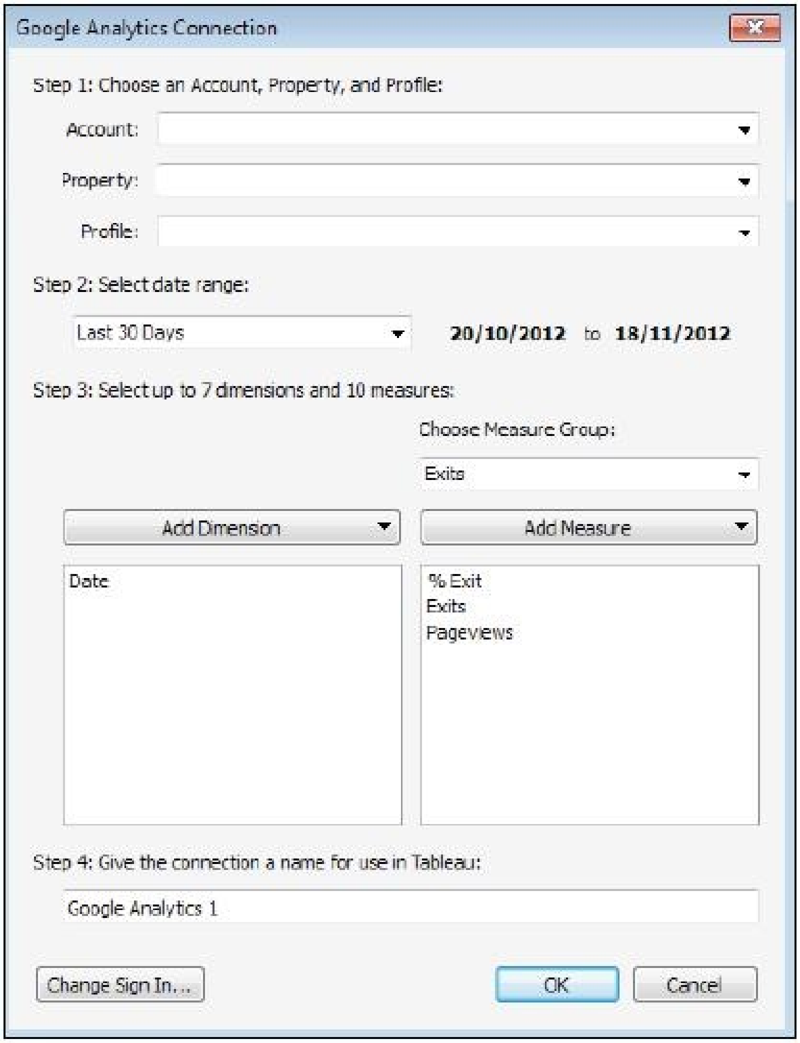
As in the case with the previous recipes in this chapter, we will continue to use the `Chapter Four` workbook.

How to do it…

1. Let's select the **Connect to Data** option on the Tableau welcome page on the Tableau desktop. Select the option for **Google Analytics** and the **Google** web page will ask you for a username and password for your user account, as shown in the following screenshot:



- Once you have entered this information, you will be asked for the Google Analytics connection details. You can see an example of this feature in the next screenshot:



- Once you see the **Google Analytics Connection** dialog box, you can fill in your details, such as **Account**, the **Property**, and the **Profile**.
- When you click on **OK**, you can see the dimensions and measures in the Tableau worksheet.
- Now that you can see your data, you can combine the data by simply grouping it together, which makes it easier to analyze.
- For example, you can combine page groups in order to analyze web pages on a section-by-section basis, or even exclude certain groups from your analyses.

How it works...

To summarize, mashing up your external data with your internal data can mean that you get richer results and analyses from the data. This means that your dashboards are essential tools for helping to drive decisions and insight throughout your organization, and your colleagues will thank you for it!

Chapter 5. Putting the Dash into Dashboards

In this chapter, we will cover the following topics:

- Choosing your visualization
- Using parameters in dashboards
- Using custom geocoding in Tableau
- Profiting from Big Data to rev your visualization
- Filtering your data for focus
- Creating choices in dashboards using conditional logic

Introduction

Dashboards are more than visual tools to display data; they are tools that can help to move business forward. Dashboards are used as decision-making tools to obtain results quickly.

You can help your business users make quick decisions by producing dashboards that are in line with the current research and thinking of dashboard structure. You can also get results from dashboards by performing the following actions:

- Improving the availability of data
- Facilitating the user's understanding of the data quickly
- Sharing information with team members and beyond
- Opening accessibility to users via adaptability
- Allowing flexibility for users to add notes to their dashboard

Choosing your visualization

Visualization is about democratizing the data and making it accessible to the people who need to know.

Today, there are many hot trends in both consumer and enterprise technology that increase accessibility by being highly visual. Think of the popularity of iPads, Surfaces, eReaders, and large screens. Everyone wants their data in the best resolution possible, with crisp graphics and colors.

Executives are engaging with the charm of visualization and putting it firmly onto enterprise business intelligence roadmaps. According to a survey by Howard Dresner, the extensive use of color, size, shape, and motion were more appealing than other buzzwords such as Big Data and the cloud. A study by Dresner Advisory Services found that advanced visualization and dashboards ranked high in terms of importance.

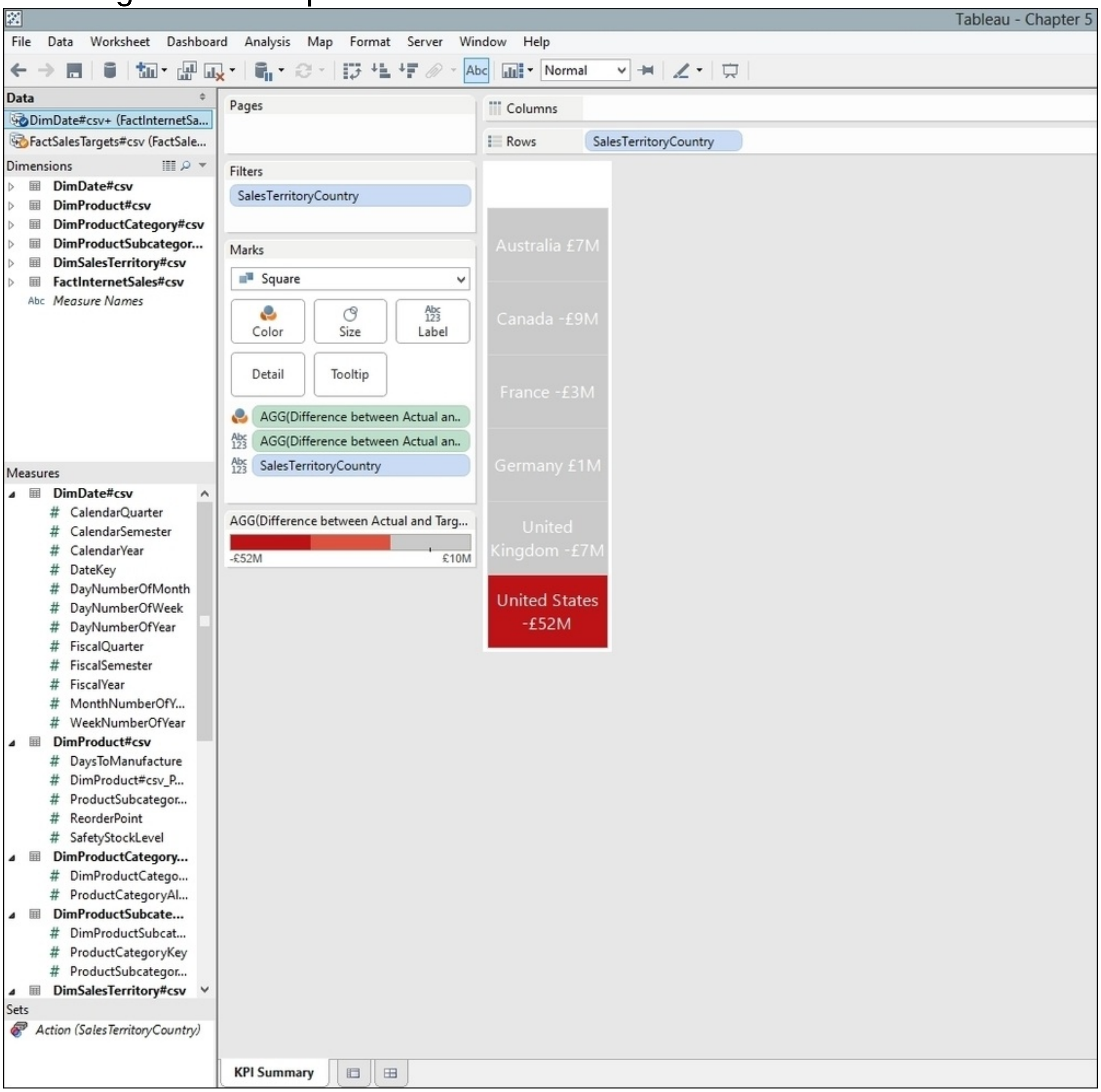
Why is visualization so useful? It's more than pretty pictures. Data visualization helps us to understand meaning in data via a communication medium that we are "geared" towards every day—our vision. Through discrimination and effectiveness of data visualizations, we reach insights and decisions. Technology allows us to create magic with our data, which engages us towards better decision making.

Given its power, how do we choose which is the right visualization? Throughout this book, we will talk about different visualization choices as we proceed. Because we are looking at dashboarding, we will look at dashboarding features such as KPIs.

In this recipe, we will look at different considerations when choosing your visualization. We will look at some of the default settings of Tableau and how they are affected by color blindness. We will also look at **sparklines**, which aim to provide as much context in as small a space as possible. This will be very useful in creating dashboards.

Getting ready

For the exercises in this chapter, take a copy of the `Chapter 4` workbook and name it `Chapter 5`. This workbook already has the data for the exercises in this chapter, so we do not need to make any changes to the data. We will delete all of the sheets except the `KPI by Q` sheet. The following is an example of how the workbook will look:



How to do it...

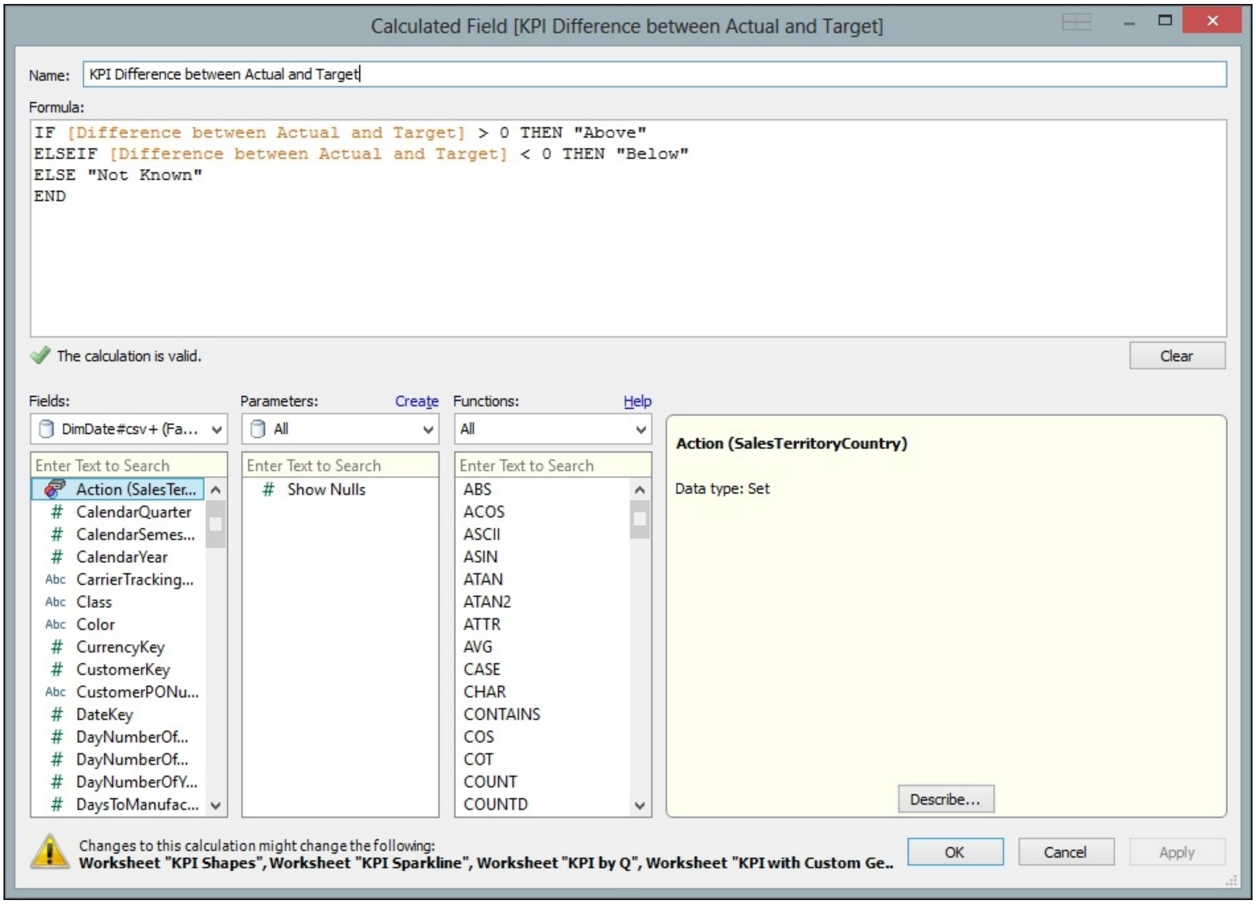
1. To start, let's rename the worksheet to `KPI Summary` by right-clicking on the tab at the bottom of the worksheet and selecting **Rename Sheet**.
2. Drag **SalesTerritoryCountry** to the **Rows** shelf.
3. Drag **Difference between Actual and Target** to the canvas area.
4. The boxes will appear colored, but you won't see any text. Drag **Difference between Actual and Target** to the **Label** button.
5. Drag **SalesTerritoryCountry** to the **Label** button.
6. Click on the **Label** button and then click on the button with ellipses next to **Text**.
7. In the **Edit Label** textbox, use the **Insert** drop-down box to choose the field names that should be displayed. This will read as follows:

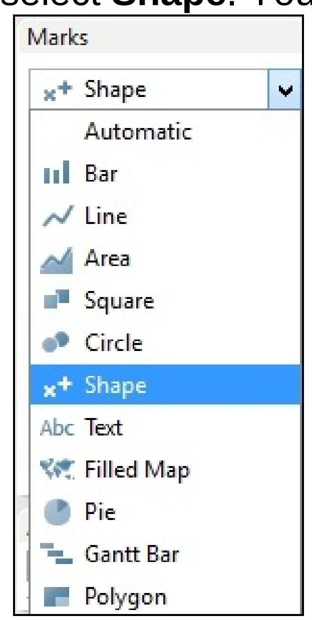
```
<SalesTerritoryCountry> <AGG(Difference between
Actual and Target)>
```

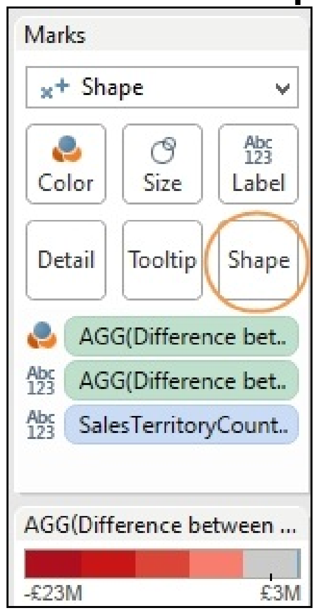
- Click on **OK**.
- Let's duplicate the worksheet by right-clicking on the tab at the bottom and selecting **Rename Sheet**.
- Rename the duplicated worksheet to `KPI by Q`.
- Drag **Year(FullDateAlternateKey)** to the **Columns** worksheet.
- We will start by creating a very simple KPI at first using the `KPI by Q` worksheet. To do this, we will create a new calculated field by going to the **Analysis** menu item and selecting the **Create Calculated Field…** option.
- In the **Calculated Field** textbox, enter the following formula which you can see here:

```
IF [Difference between Actual and Target] > 0 THEN
"Above"
ELSEIF [Difference between Actual and Target] <= 0
THEN "Below"
ELSE "Not Known"
END
```

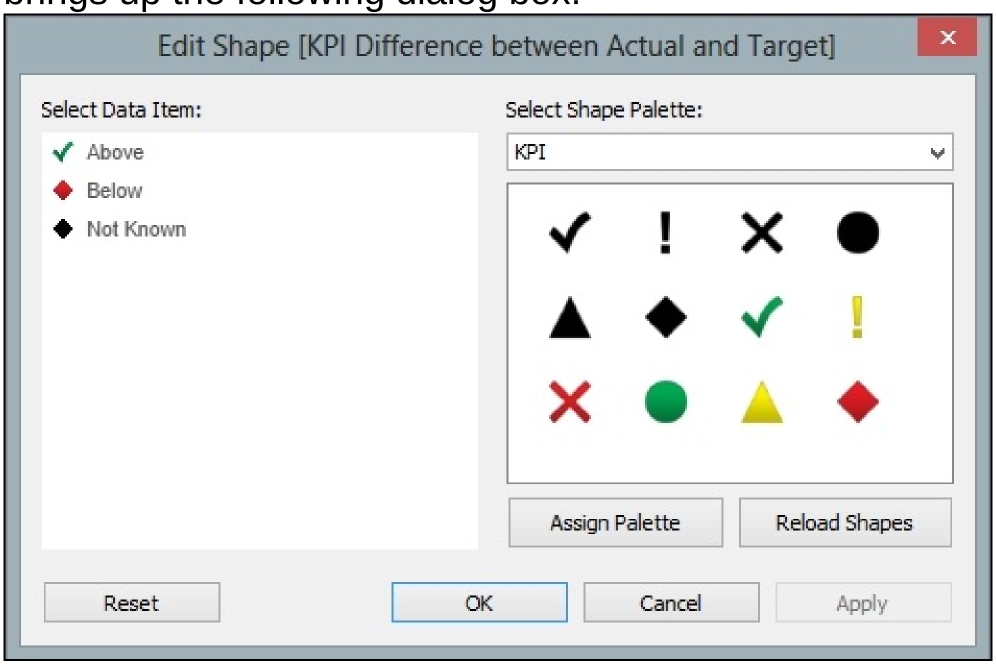
- The following screenshot shows you an example:



- Once you've created your calculation and returned to the main Tableau canvas, you'll see that there are five buttons in the **Marks** shelf. We can add a sixth button, **Shape**. To do that, go to the drop-down list below **Marks** and select **Shape**. You can see an example in the following screenshot: 

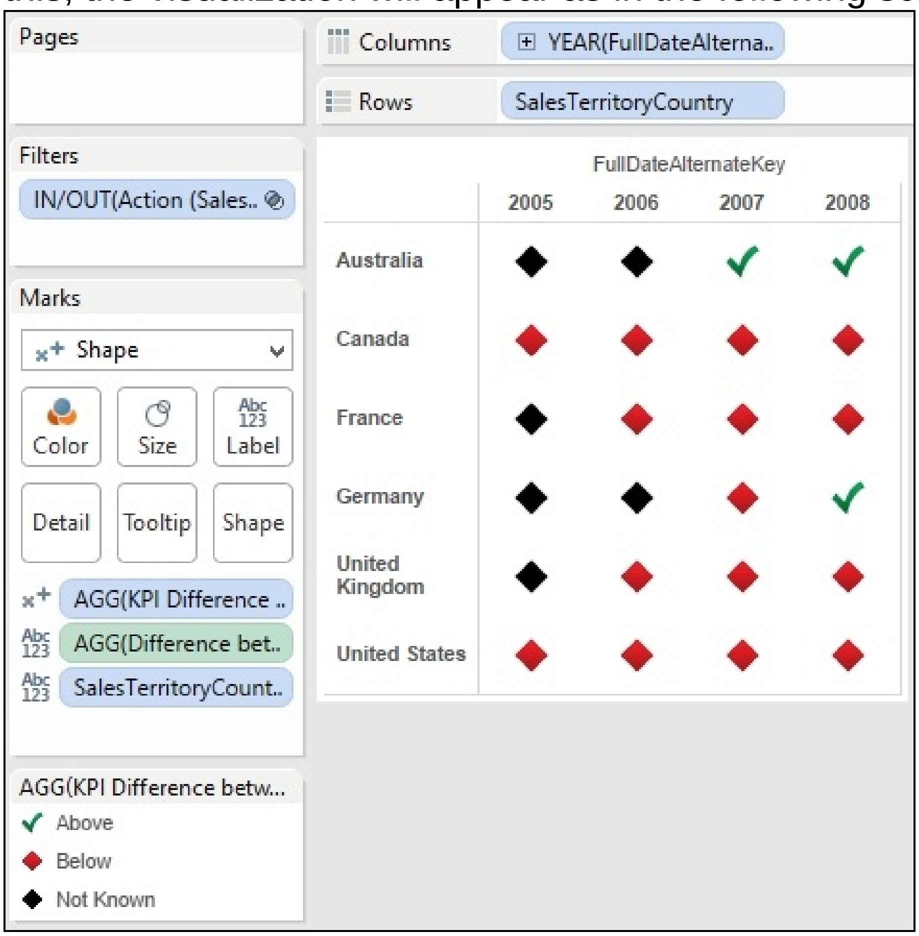
- When you select **Shape**, you can see that we now have a sixth button called **Shape**. Your screen should now show the **Shape** button, as you can see in the following screenshot: 

- You can then click on the **Shape** button to edit the shape's style and appearance. Drag the **Difference between Actual and Target** calculation onto the **Shape** button.
- When you click on the **Shape** button, you get a dialog box that offers you a number of options. Let's choose the **Edit Shape** option, which brings up the following dialog box:



- Here, we have selected the default **KPI** palette. A more traditional KPI selection selects the green tick for the **Above** option, which identifies the metrics that exceed the KPI success criteria. It follows that the **Below** option is indicated by a red cross. Finally, the **Not Known** criteria identifies the areas where there is no value present; in other words, it is a `NULL` value.

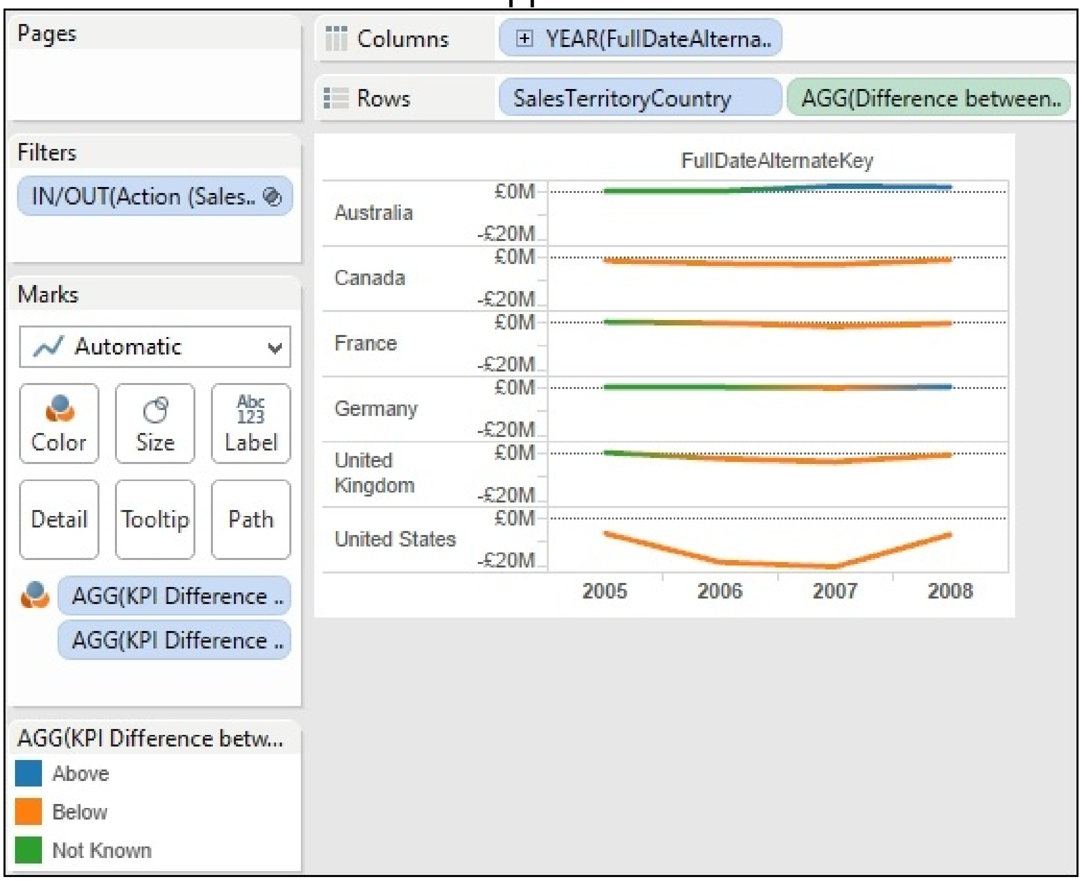
- Remove the color marks from the **Marks** shelf. At this point, you will probably have **AGG(Difference between Actual and Target)** as a color mark, but this will prevent the KPI shape from working properly.
- Remove the **Label** marks from the **Marks** shelf.
- Drag **Action(SalesTerritoryCountry)** from the **Sets** pane in the **Data** sidebar to the **Filters** shelf. On the downward arrow at the right-hand side of the **Action(SalesTerritoryCountry)** pill, select **Show In/Out of Set**.
- On the dialog box, select **In** and then select **OK**.
- Once you have selected this option, click on **OK**. Once you have done this, the visualization will appear as in the following screenshot:



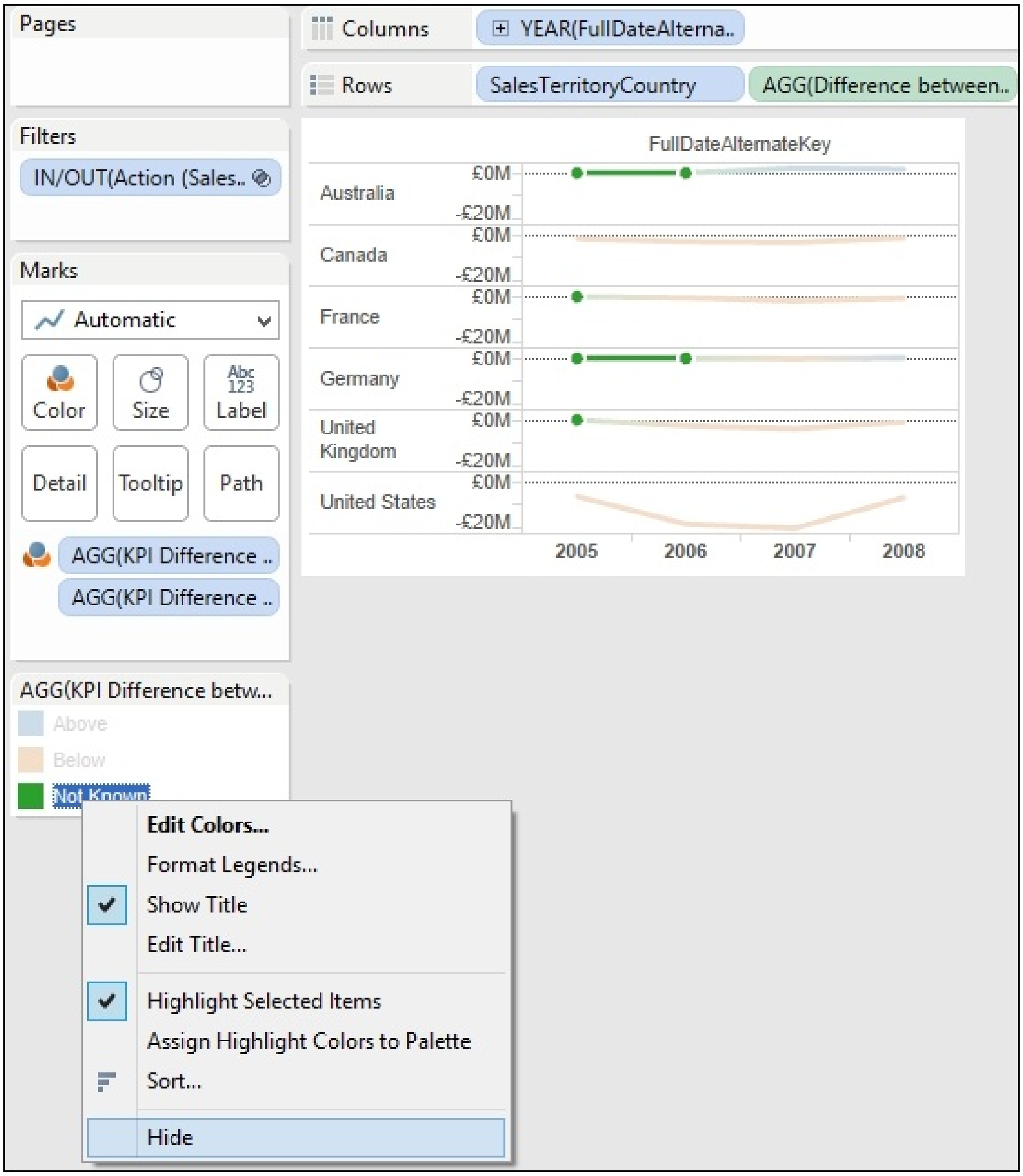
- Although we have created a KPI image, we would like to change it in order to cater to people who have color blindness. How can we show a

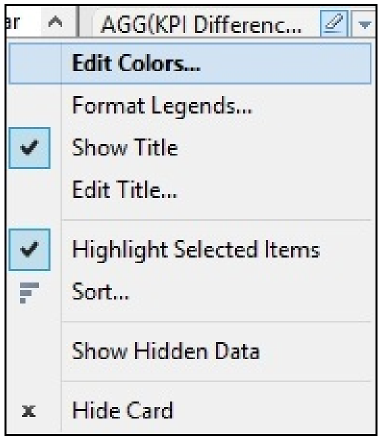
KPI so that color blindness is taken into account?

- Let's take a look and see what Tableau does when we try to create a sparkline. To do this, duplicate the existing visualization so that we have a "point in time" of our work to date.
- Make sure that the **SalesTerritoryCountry** attribute is in the **Rows** shelf. Then, drag the **Difference Between Actual and Target** calculation to the **Rows** shelf.
- Select a line graph visualization from the **Show Me** panel. Make sure that the **SalesTerritoryCountry** attribute is in the **Rows** shelf.
- Drag **the Difference Between Actual and Target** calculation to the **Color** button.
- The Tableau visualization appears as follows:

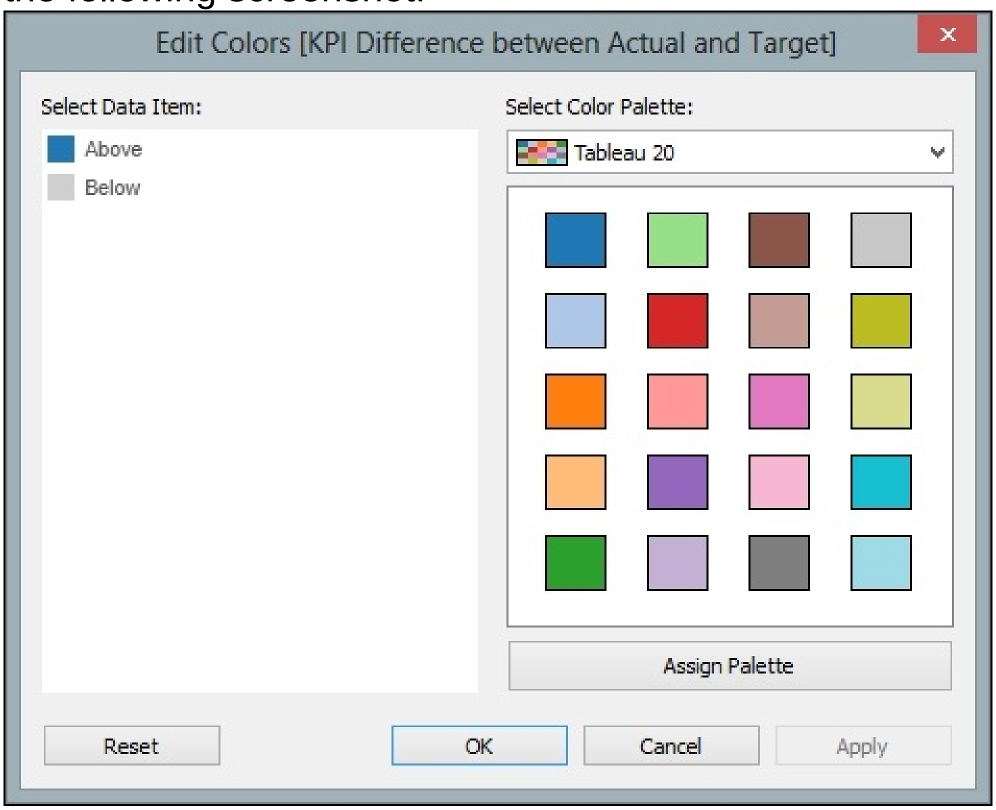


- This still leaves us with the problem of showing data that isn't there—in other words, the `NULL` data. While we are building up to a sparkline, it can look misleading because the data looks as if it starts from 0 and curves upwards; in fact, there isn't any data there for the highlighted marks, which you can see in the following screenshot. We can opt to **Hide** the **Not Known** data.

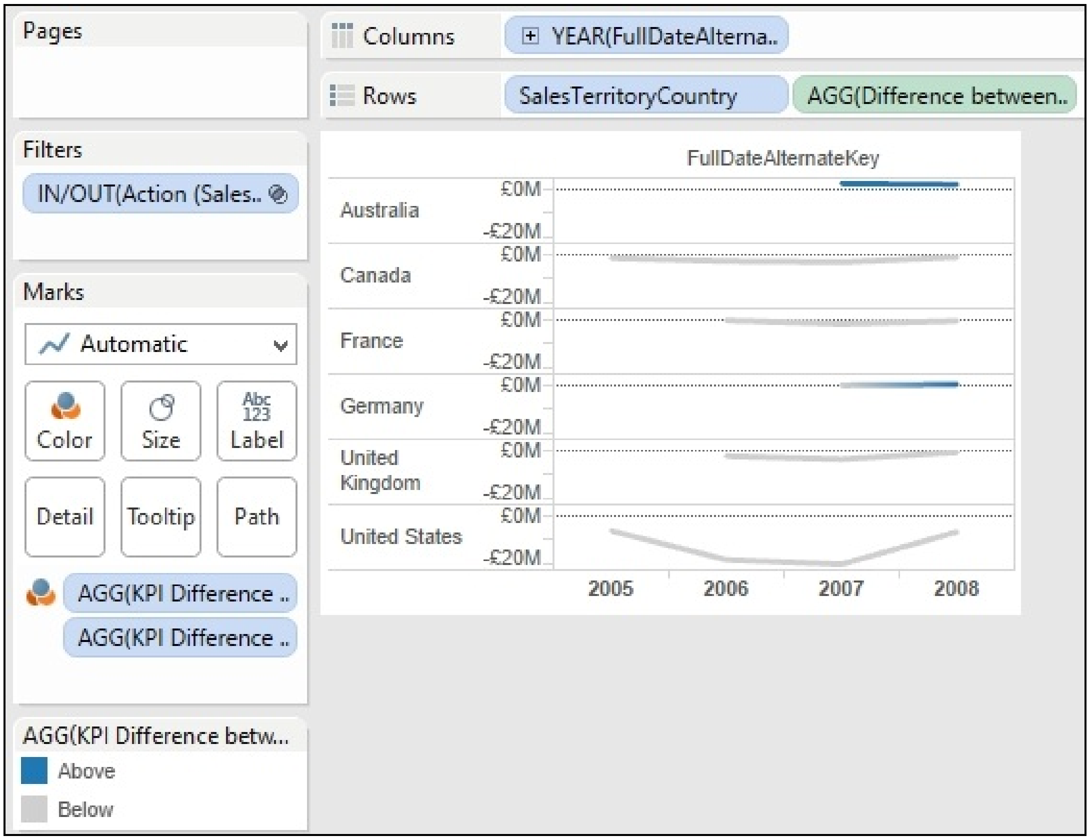


- First things first, however, let's fix the colors so that they are more appealing. To do this, click on the **AGG (Difference between Actual and Target)** metric and select the **Edit Colors…** option, as shown in the following screenshot: 

- Since we have hidden the **Not Known** data, we only have the **Above** and **Below** data to worry about. For now, let's choose blue for the **Above** value and gray for the **Below** value. You can see a sample selection in the following screenshot:



- Now let's see how the completed visualization looks once we have resized it, as shown in the following screenshot:



To summarize, in this recipe, we have started to look at some dashboard visualizations. We will progress to look at others throughout the rest of the chapter.

How it works...

In this recipe, we have looked at some of the default settings in Tableau for creating KPIs along with some options for configuring them. It is clear that we can drastically change the appearance of the visualization by simply making a few changes to the default settings.

When creating data visualizations, it is vital to remember the audience. This may seem a simple, obvious statement. However, how many charts have you seen that mix red and green together? A lot, probably! In this recipe, we looked at alternatives to using red and green to convey a message. Here, we have made a color choice based on the business question. We went from a straight red-and-green visualization to one which used blue to represent the data which was the focus of the business question. There are a few issues with our initial KPI though. First, a `NULL` value is represented by an icon when there is no data. It would be better to simply not show anything at all; this would reduce unnecessary "ink" on the page and would require less effort to assimilate.

Another issue is that the image shows red and green on the same visualization, which isn't good for color-blind viewers. Although additional information is provided by the shape of the indicator, it's probably best to simply avoid it if possible. At that point, we could have created only a simple, red KPI to denote the data points where the target was not met and then moved on to a different topic.

Instead, we then changed from using a shape to represent the data to a sparkline. Sparklines were devised by Edward Tufte in order to express a "small, intense, simple, word-sized graphic" to represent data. Sparklines are supposed to be able to be embedded into sentences. They are used to express data in a very compressed, at-a-glance way, which means that they are perfect for dashboards.

There's more...

Red and green KPIs are popular, but they are not always the best solution. Sparklines can give us more context by providing more information over time, but this can be misleading in certain cases. In the next recipe, we will look at ways in which we can improve the sparkline to manage this scenario.

Using parameters in dashboards

Parameters are dynamic "placeholders" that can help to control the dashboard appearance by storing information to help drive the flexibility of the dashboard.

We can use parameters to make our visualizations more accurate by using them to control which data is to be displayed and which is to be hidden in response to user input. In this recipe, we will look at using parameters in order to drive the display and enhancing the accuracy of the presentation by controlling the inclusion of `NULL` values in the dashboard.

If `NULL` values are present in the data, then this might cause a misleading representation of the data. On the other hand, `NULL` values might be useful because they could tell us about the quality of the data itself. Sometimes, the data visualization project turns into a data quality project. Often, when data is visualized, it is the first time that business users have seen their data. This means that they can sometimes get a nasty surprise! Data can be missing, incomplete, or perhaps plain wrong. Therefore, it is important that dashboard consumers understand about data quality issues rather than being distracted by the shininess of technology.

Ultimately, the effectiveness of the data visualization rests on the accuracy of the data. In other words, even if the dashboard is perfectly formed, it will be no good if it shows inaccurate data.

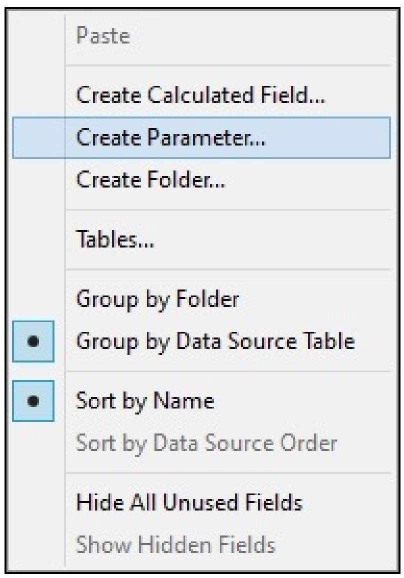
In this recipe, we will use a parameter to give the user the opportunity to reveal or hide some data, dependent on the user's choice. This recipe is quite involved, because it covers parameters, dual axes, calculated fields, and some aspects of data visualization. So let's get started!

Getting ready

For the exercises in this recipe, let's continue with our `Chapter 5` workbook so that we can see the progression from the initial KPIs to the end of the chapter. So, let's take the `KPI by Q` worksheet and make a copy of it; we will rename it `KPI Shapes`.

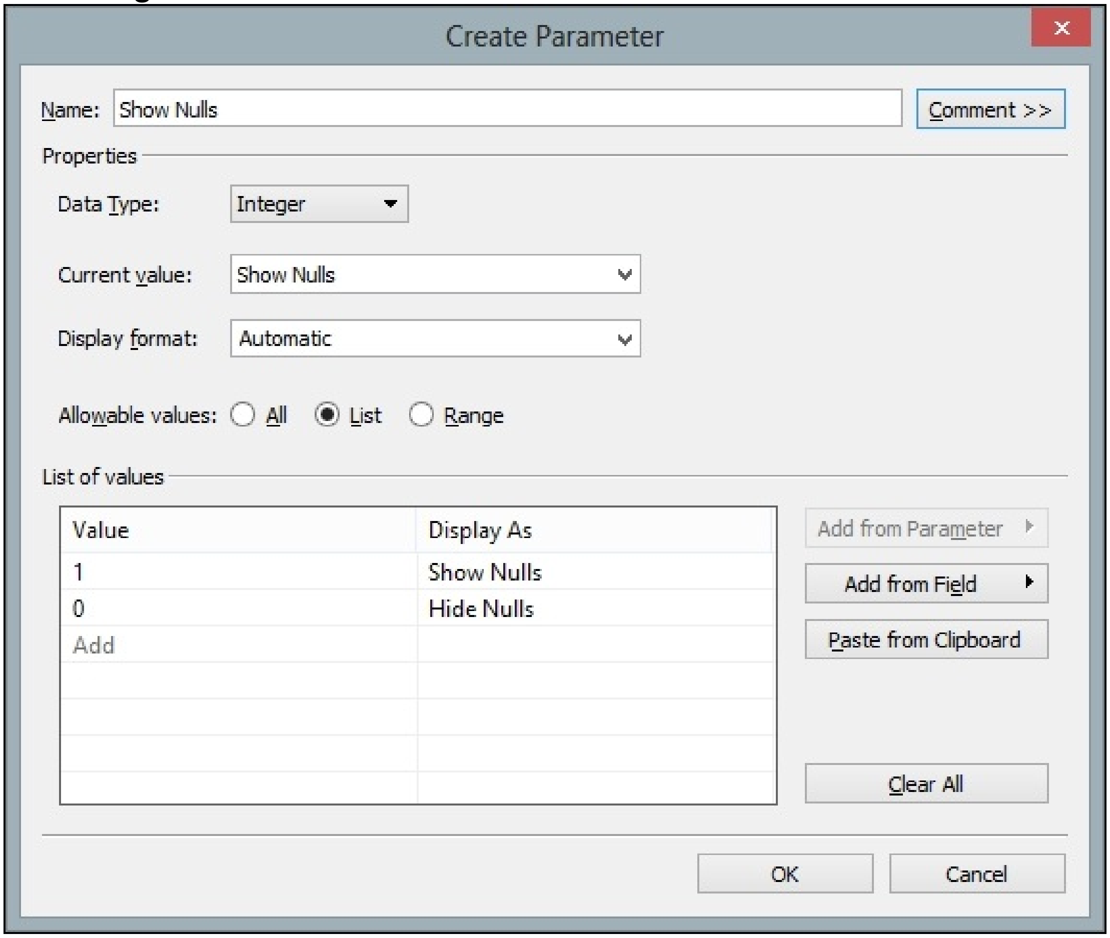
How to do it...

1. Once you've made a copy of the `KPI by Q` worksheet and renamed it `KPI Shapes`, let's create our first parameter. This is very easy to do; simply go to the **Measures** box at the sidebar of the Tableau workbook. Right-click anywhere inside this box and you will get a pop-up menu. You can see an example of this pop-up menu next. When it appears, simply click on **Create Parameter...**, which is highlighted in the following screenshot.



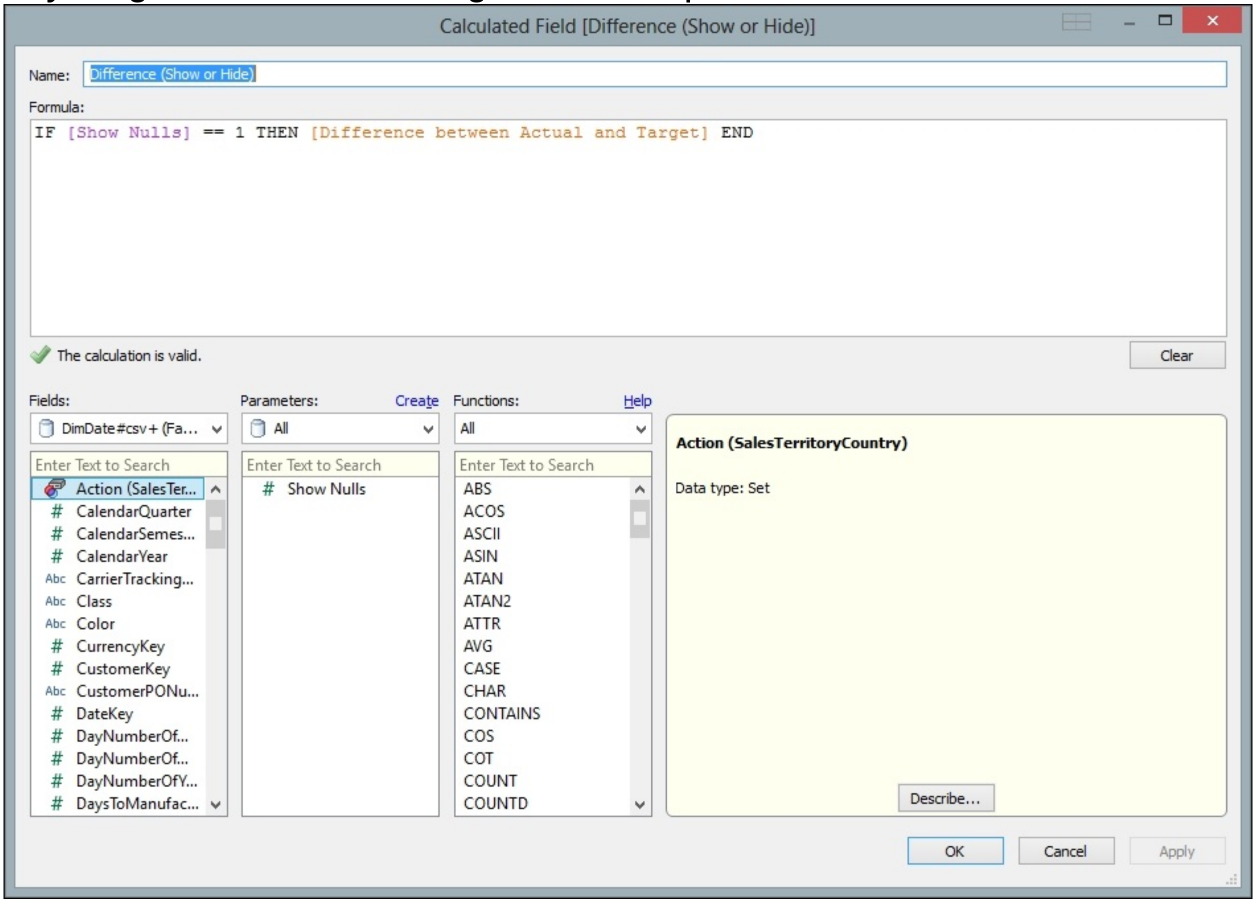
2. Once you have clicked on the **Create Parameter** option, you will get a dialog box named **Create Parameter**. This allows you to configure the parameter. We will set up a parameter that will allow the users to choose whether or not they want to display Null values.
3. First, let's give our parameter a name so that its purpose is explained precisely. We will call it `Show Nulls`. This parameter is very simple. It is set to `1` if `NULL` values are to be shown and set to `0` if the `NULL` values are to be hidden. Since we are using integers as a setting, we should keep the **Data Type** setting as **Integer**. The **Current Value** option gives the parameter a value as a starting

point. Once you have completed these fields, your **Create Parameter** dialog box should appear exactly as shown in the following screenshot:

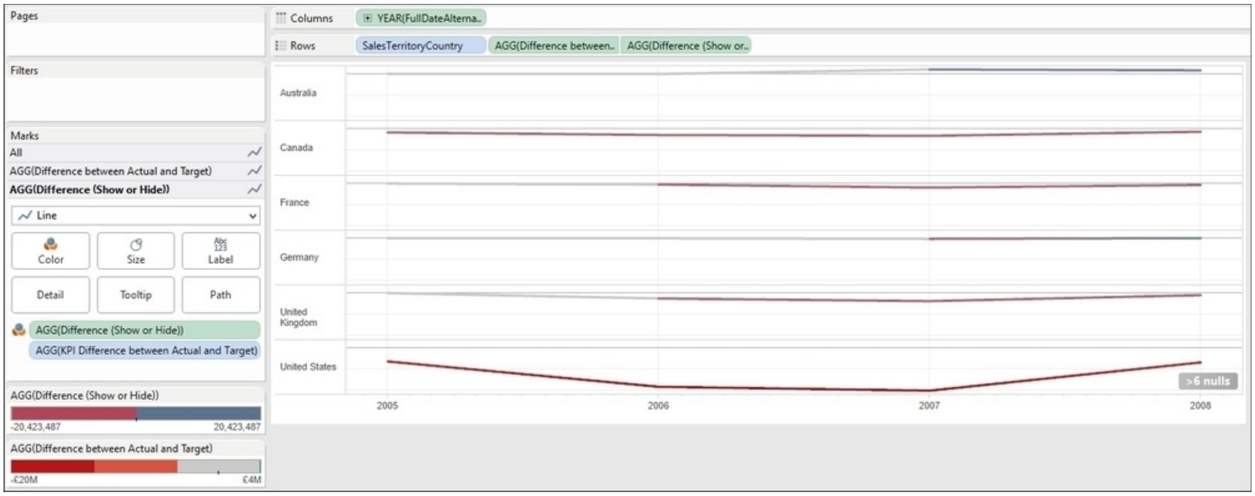


- We need to set up a calculation to control the parameter. To do this, create a new calculated field by right-clicking in the **Parameter** box again and choosing the **Create Calculated Field** option. We will use the calculation to make a rule that will drive the parameter. Our rule will specify that if the parameter setting is to show the `NULL` values, the parameter value is set to `1`. This will display the second copy of the difference between the **Actual** and **Target** measure. If it is set to `0`, only then will the line graph show on its own, which will show the `NULL` values as well as the actual data.
- You can see our calculated field in the next screenshot. The calculated field is called **Difference (Show or Hide)**. The calculated rule incorporates a rule that says that if the parameter **Show Nulls** is equal to `1`, then show the **Difference Between Actual and Target** metric. If

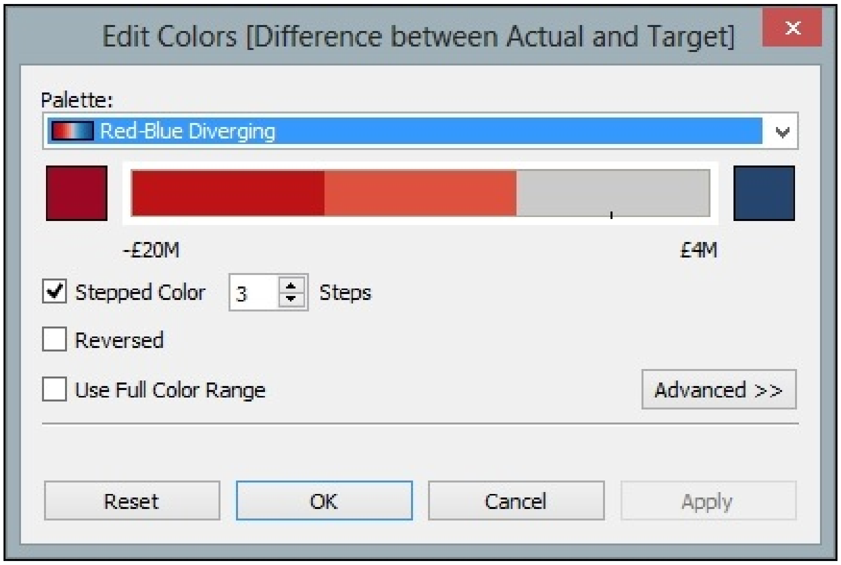
Show Nulls is not equal to `1`, then the rule fails, and it does not show anything at all. The following is an example of the calculated field:



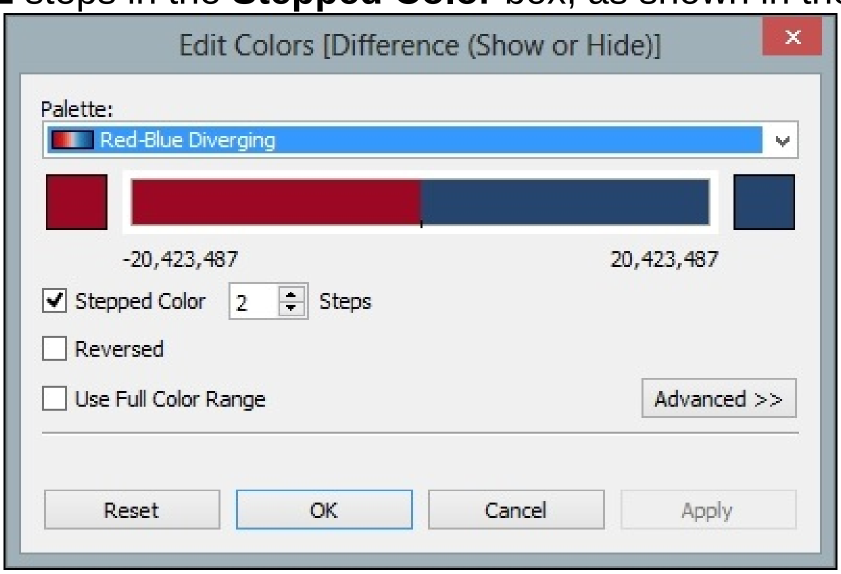
- Once you have created the calculated field, click on **OK**, and this will take you back to the Tableau workbook. Then, drag the **Difference between Actual and Target** measure and **Difference (Show or Hide)** to the **Rows** shelf, making sure that **SalesTerritoryCountry** is also on the rows. Then, choose the **Dual Line Axis** option from the **Show Me** panel. Your Tableau worksheet will look similar to the following screenshot:

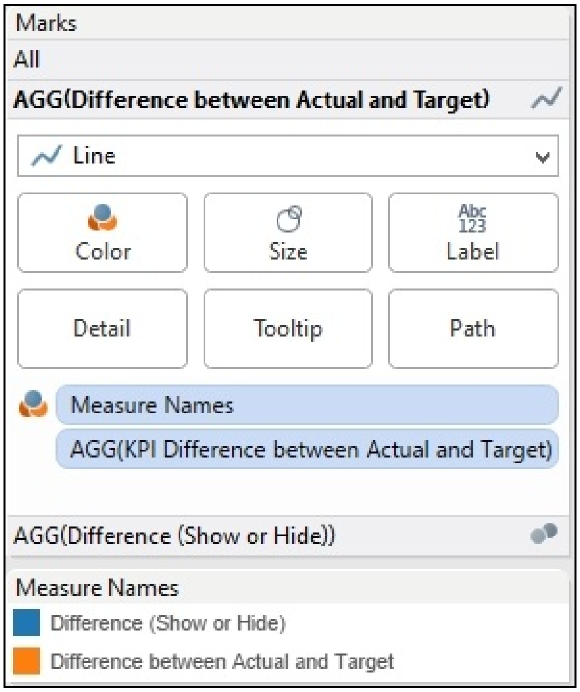


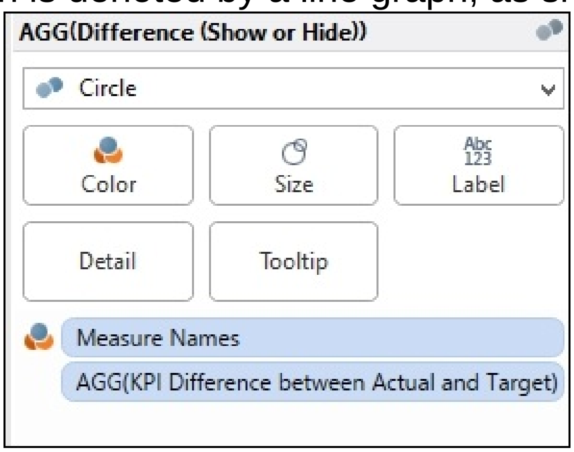
How can we clearly differentiate where the actual data points are? The problem with the dual axis, as it stands, is that it will start at 0 if there is no data. This is because the axis is aligned to the year and the country, and if there is no data, it will simply map the data as having a value of 0. This can be misleading, however. For example, the data for Germany shows that there is a line showing data for 2005 and 2006, which ends up at a value of £1 million for the year 2008. However, this is a bit misleading; in fact, there was no data for the years 2005 or 2006; there was only data for 2007 and 2008. It would be better if this was clearer to the user.

Let's make the story of the data clearer to the business user by setting the colors and the line chart. This KPI panel is illustrating the data to illustrate an answer to a business question: which countries failed to meet their targets and when? This means that we are interested in emphasizing the losses made. We can do this by coloring these data points in red, which is a color normally used to denote a warning or a loss. Since we are not so interested in data where the countries made their targets, we will use the color gray so that this data takes a "background". Let's do this first for the **Difference between Actual and Target** data by dragging this measure onto the **Color** button. This will give us the following **Edit Colors** dialog box. Although we choose the option for **Red-Blue Diverging**, if we select the **Stepped Color** option and set it to **3** steps, then, we can get two different shades of red and one gray color. You can see an example of this setting in the following screenshot: 

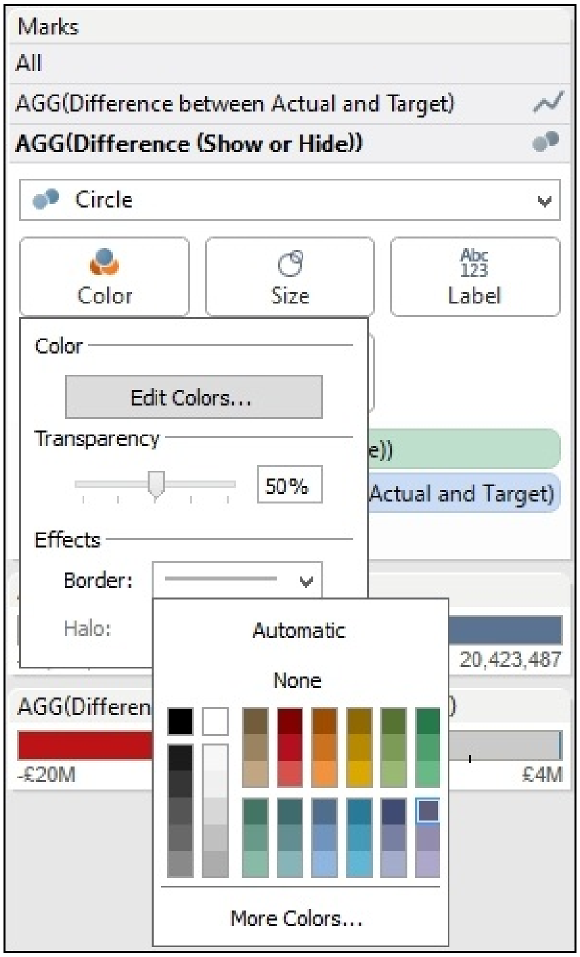
- Once you have configured the color for the **Difference between Actual and Target** metric, let's move forward to set the color for the **Difference (Show or Hide)** metric. Drag the **Difference (Show or Hide)** metric onto the **Color** button and you will get the **Edit Colors** dialog box. This time, we will select the same **Red-Blue Diverging** option, but we will choose **2** steps in the **Stepped Color** box, as shown in the following

screenshot: 

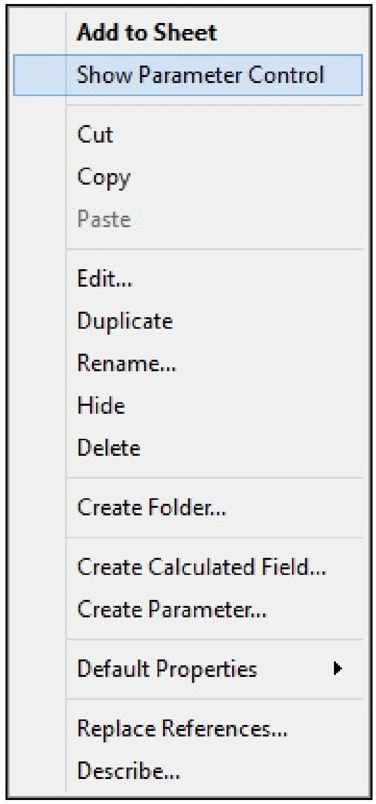
- Now let's make sure that the **Difference between Actual and Target** metric is set to a line. We can see this because the mark for the **Difference between Actual and Target** metric has a small line next to it, denoting that it is set to a line chart, as shown in the following screenshot: 

- Now, click on the **AGG(Difference(Show or Hide))** metric under the buttons, and this will reveal the buttons for editing this metric. In the drop-down list, choose **Circle** for the **AGG(Difference(Show or Hide))** metric. This will distinguish it from the **Difference between Actual and Target** metric, which is denoted by a line graph, as shown in the following screenshot: 

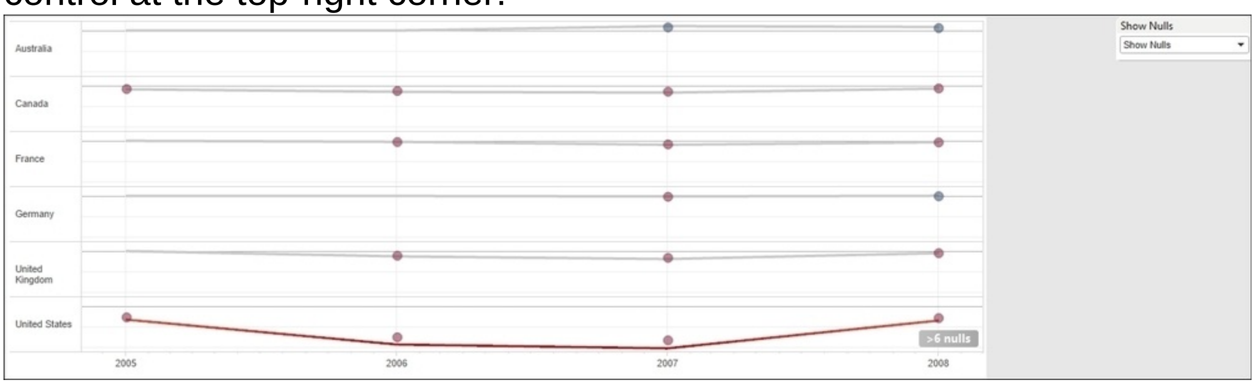
- We can set a border around the circles so that they are defined. At the same time, we can make the color transparent so that we get a layered effect. To do this, click on the **Color** button and you will get a pop-up menu. Set the transparency to 50 percent and choose a light purple for the border. You can see the options in the following screenshot:



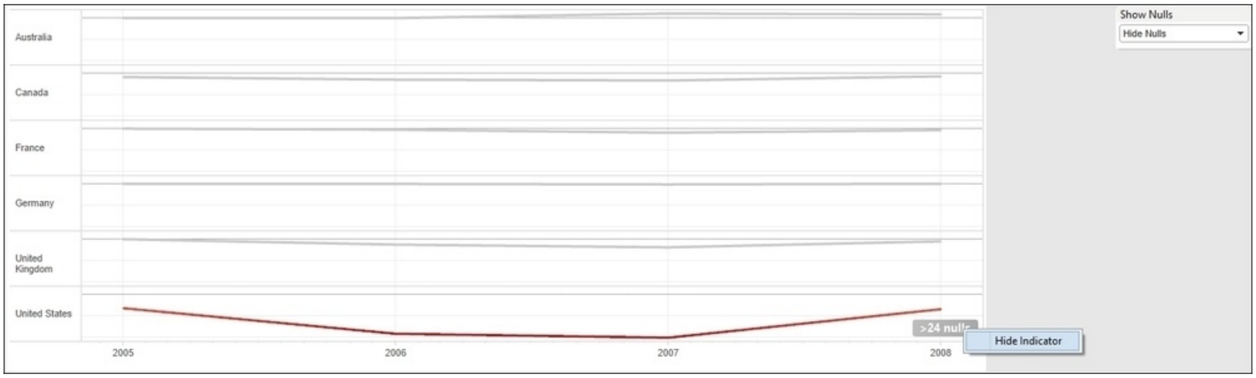
- The last thing we need to do is to show the parameter control, which will give the user the option of explicitly showing the data points that are not `NULL` or leaving the chart as is. To show the parameter control, right-click on the **Parameters** section and select the **Show Parameter Control** option from the pop-up list.



- The data visualization has now been completed, so let's test it out to see how it looks. In the following screenshot, you can see the parameter control at the top-right corner:



- If you choose the **Show Nulls** option, then you will get the data points appearing. This allows you to see which points are actual data and which points are `NULL`.
- If we choose the **Hide Nulls** option, then we can see that the United States has engendered a loss which was unacceptable, but the other countries have not. However, this shows the `NULL` values, which assumes that all of the countries have commenced at the same starting point, as shown in the following screenshot:



- To summarize, using parameters to drive the data visualization, we can make our dashboards interactive and more sensitive to data quality.

How it works...

To sum up, in this recipe, we have looked at data quality, calculations, parameters, and data visualizations. These are all interesting topics in their own right, and the objective of this recipe was to show that we can put them together in interesting ways in order to produce a dashboard. Tableau allows us to be very creative with our data in order to satisfy user requirements.

How did we use parameters in Tableau? To set up this visualization, we set up a dual line axis which has two measures on it: one is **Difference between Actual and Target** and the other is a calculated field that has a rule in it, which shows or hides a copy of the **Difference between Actual and Target** measure. Yes, in other words, we show this measure twice on the dual axis, or show the measure only once, depending on the choice of the user. The difference is in the way in which we represent each copy of the measure. One copy of the measure is a line graph, which is always shown, and the second copy is a dot plot, which only shows the data that is present. The parameter shows, or hides, the second version of the measure in order to show which data points actually exist.

Using custom geocoding in Tableau

Organizations often have their own definitions of geographic data. Although country names stay relatively static, their classification can change as the organization emerges from one level of maturity to another. Sometimes, for example, an organization can start with a very simple division: North America and EMEA. However, as the organization grows, it might split off into North America, Europe, Asia Pacific, and Rest of the World, for example. This can mean that the geography has a business context and meaning as well as describes a physical location.

Since some geography is fairly standard, Tableau offers a default interpretation of certain geographic data to help you automatically create maps from your data. The default interpretation includes countries, states, and area codes, for example. However, Tableau's default interpretation can be tailored to align with the business interpretation of geographic data.

In this recipe, we will add in some customized geographic data by importing a custom file and then using the customization to create a data visualization. The data is taken from the Human Development Index research, which is part of the United Nations Development Programme, which in turn is an organization that has the goal of "advocating for change and connecting countries to knowledge, experience, and resources to help people build a better life." The **Human Development Index** (**HDI**) is a new way of measuring development by using metrics such as life expectancy, educational success, and income and combining them into one measure. You can find more information about the HDI metric at http://hdr.undp.org/en/statistics/hdi/.

In this recipe, we will look at importing custom geocoding. One interesting feature of this exercise is how we go about using color to indicate rank.

Getting ready

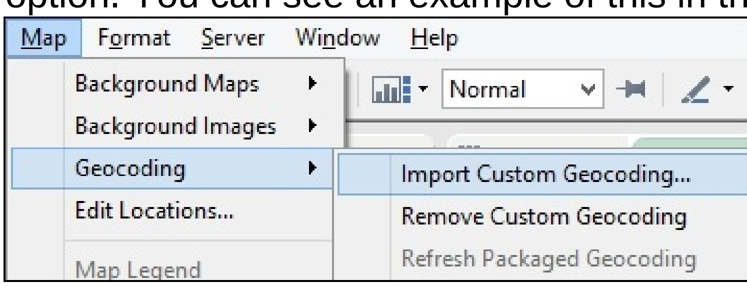
Let's continue to use the `Chapter 5` workbook. We have an amended `DimSalesTerritory` to reimport, which contains the HDI rank of each country in the `AdventureWorks` database. To do this, replace the existing `DimSalesTerritory.csv` file with the `DimSalesTerritory.csv` file of the `Chapter 5` workbook. If you open the new file, you will see that it contains an additional column: `HDIRank`. To refresh the data, simply go to the **Data** menu option and select **Refresh All Extracts**. You should see a new column called `HDIRank` in the `DImSalesTerritory` dimension.

For the purposes of creating data for customizing geocoding in Tableau, you need data that follows a number of rules:

- The filename must be called the same as the key of the data
- The file must be in CSV format

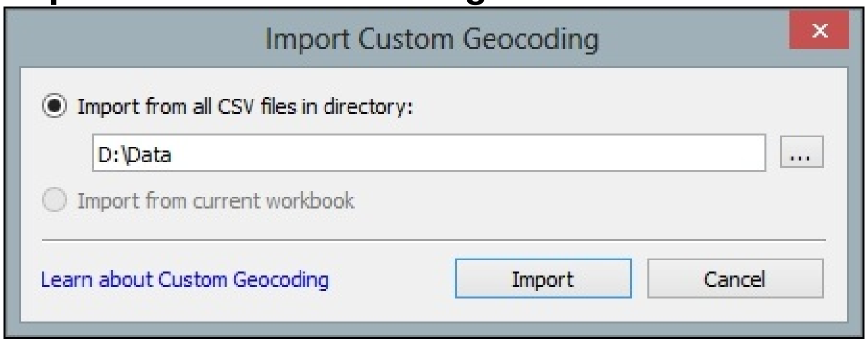
For the purposes of this example, there is a small file that you can download at http://bit.ly/TableauBookCh5HDIRank. The data file contains three columns: a nominal latitude and longitude of the countries contained in the `AdventureWorks` database along with their HDI ranks according to the HDI 2013 report.

How to do it…

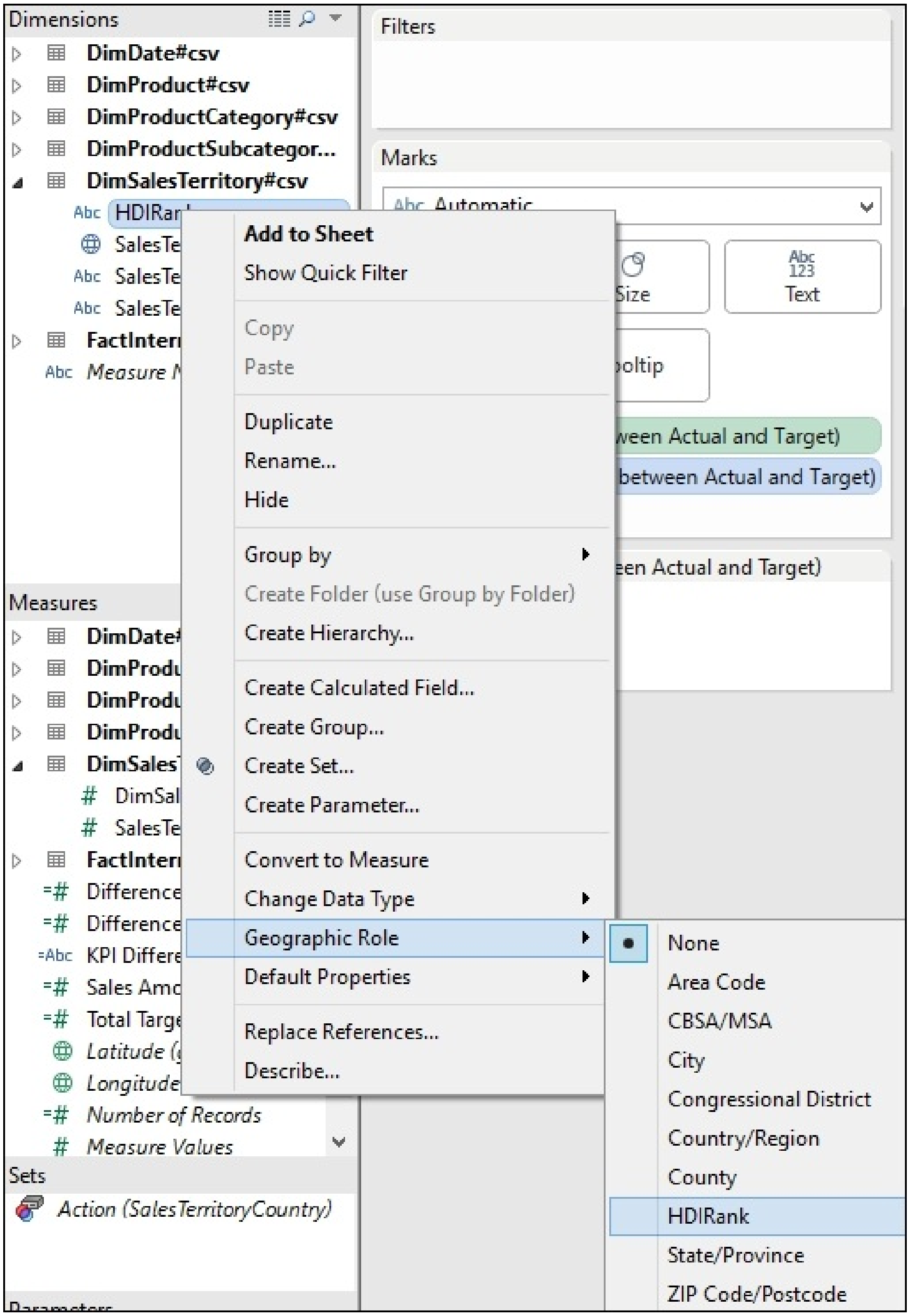
1. Once you have made a copy of the data file, let's open our `Chapter 5` workbook in Tableau and proceed to import the custom data file. This is very simple. Simply go to the **Map** file menu item, then go to the **Geocoding** file menu item, and then select the **Import Custom Geocoding** option. You can see an example of this in the following screenshot: 

- Once you have selected this option, you will get a small dialog box that asks you for the location of the file. The following is an example of the

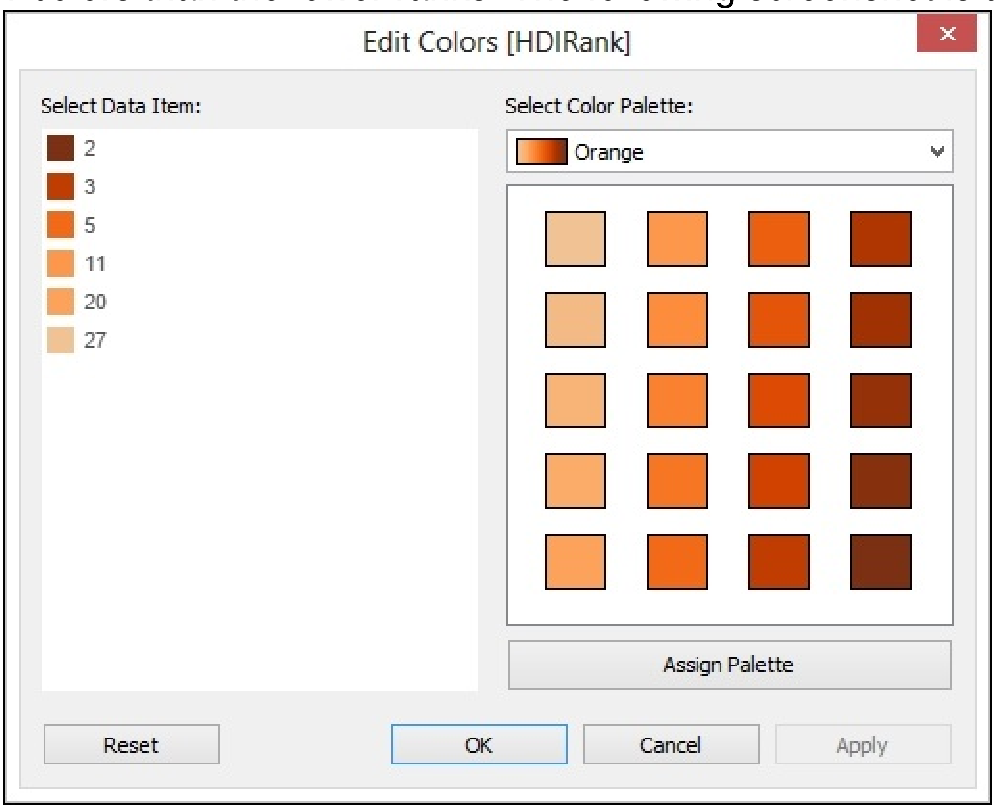
Import Custom Geocoding textbox:



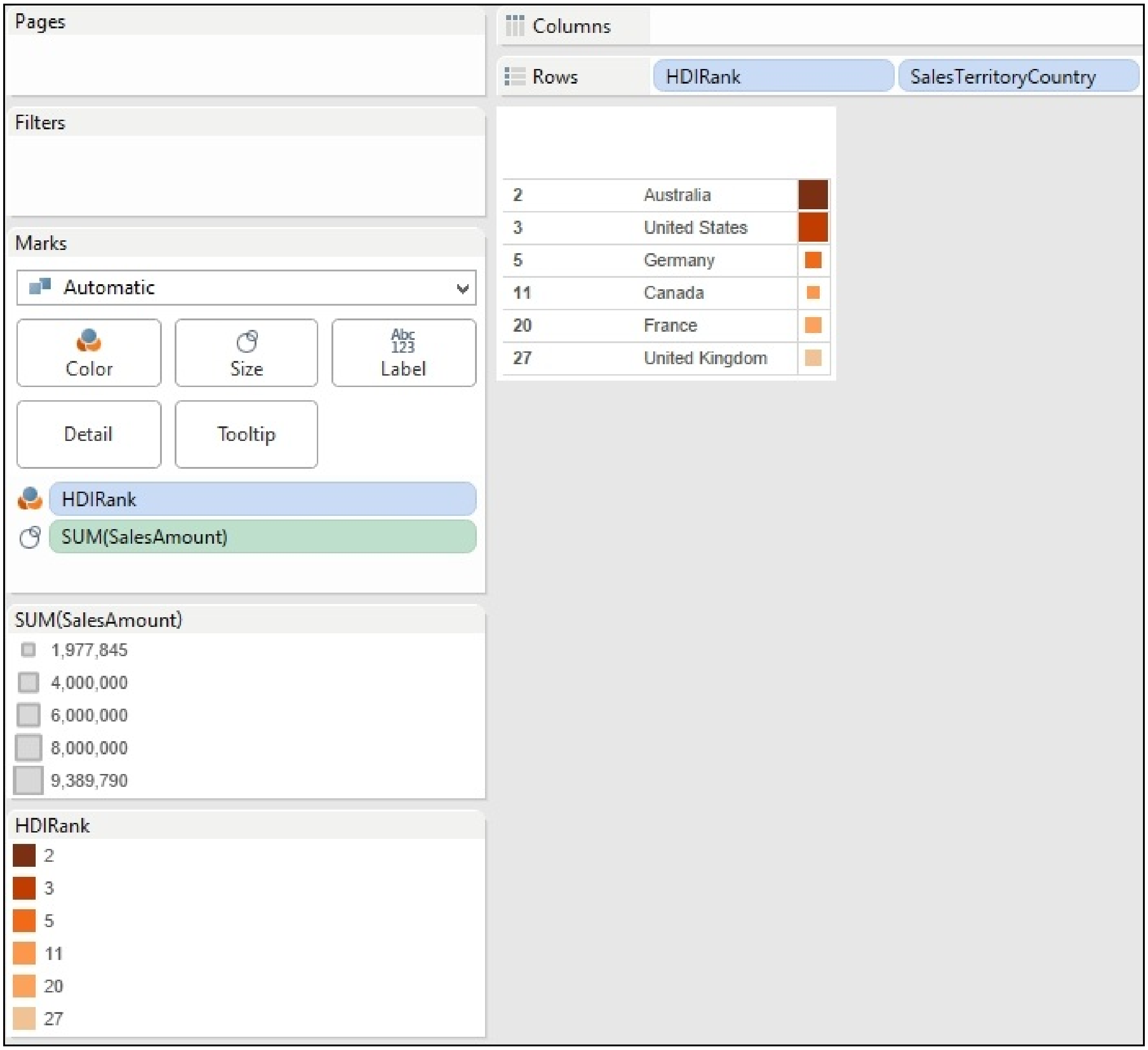
- When you have navigated to the file, select the **Import** button and your import is complete. Once you have imported the customized geography, you should be able to see it as part of the geographical role options in the Tableau drop-down list. You can see this in the next screenshot.
- If you go to the **DimSalesTerritory** measures pane in the sidebar, you will see the new **HDIRank** column. Drag it to the **Dimensions** pane to make it a dimension.
- Now, if you right-click on it and look under **Geographic Role**, you will see we have a new option called **HDIRank**, as shown in the following

screenshot: 

- The next step is to assign the **HDIRank** geographical role to the **HDIRank** dimension attribute. To do this, right-click on **HDIRank**, navigate to **Geographical Role**, and then select **HDIRank** under **Geographical Role**. Once you have done this, we will be able to use our **HDIRank** in our data visualization. You can see that the **HDIRank** symbol changes to show that this has a custom geographical data role. The following is an example of the symbol: 

- For example, we can use **HDIRank** to drive the color that denotes each country. To do this, drag it to the **Color** button. Since we are using a rank to distinguish the countries, we can use a sequential palette to show that the data is on a continuum rather than in separate categories.
- In this example, we will select the orange color.
- We can change the colors quite easily using the **Edit Color** dialog box. You will need to change each color so that the lower values have brighter and darker colors than the lower ranks. The following screenshot is an example: 

- Once you have edited the colors, let's use **Sales Amount** to denote the size. Then, choose the **Heatmap** option from the **Show Me** panel and our visualization now looks like the following screenshot:



To summarize, we have used our custom geography to help identify the rank of each country. Since the new role appears in the drop-down list as part of the Tableau interface, it is very easy for report developers to use it as part of their dashboards.

How it works...

Once again, we see that color plays a vital role in conveying the message of the data. In line with research on how to visualize data, Tableau will assume that the lower values should be assigned a less intense color, and higher values should be assigned a darker, brighter, or more intense color. If our data was rational or interval in nature, this would be correct. However, we are looking at ranking data, so the situation is reversed. In other words, the lower the number, the higher the rank.

See also

- If you'd like to see the full HDI 2013 report, you can find it at http://hdr.undp.org/en/media/HDR2013_EN_Summary.pdf

Profiting from Big Data to rev your visualization

We live in a world where everything is Big Data. Many organizations are burdened with too much data, and it is a common problem. The problem is made worse by the fact that many people aren't sure what to do with the data due to its size and complexity. In today's enterprises, data is often in disparate locations as well as growing in size. This situation is reflected in this recipe since it requires a lot of "moving parts" to be put together, such as the downloaded data, the Hortonworks Sandbox, and Tableau.

For the purposes of simplicity and clarity, we will simply use a small amount of data rather than a Big Data source. This will help you to manipulate the data more easily since it is in an accessible format. Often, the key factor in the importance of data is how often it is used and how many business processes depend on the data, rather than its size. So, don't ignore the little data!

If you don't have access to Big Data technologies, don't feel excluded from the party. There is no need for you to skip over this chapter. We will base our example on the Hortonworks Sandbox, which is freely available over the Internet for you to use. Further more, it is already preconfigured for you, so it is the easiest way possible to ramp up towards Big Data for free.

Getting ready

To use Windows Azure DataMarket, you will need a Microsoft account, such as a live account, Hotmail, an MSN account, or others of the kind. This is free to set up if you don't already have one. To do this, visit https://login.live.com/ and look for the **Sign Up Now** link to follow the wizard through the process.

To learn more about Big Data solutions, a great place to start is the

Hadoop Sandbox, generously provided by Hortonworks for free and preconfigured for you to get started straightaway. To get started, you need to download the Hortonworks Sandbox from the Hortonworks website at www.hortonworks.com. Sandbox is a virtual machine, and Hortonworks offer it using Hyper-V or VMware. You can download it in your preferred VM mechanism. If you are not sure about using Hyper-V or VMware, you can download the free VMware Player, which is very easy to use and will work with the Hortonworks Sandbox. You can find it at https://my.vmware.com/web/vmware/free#desktop_end_user_computing/v
.

Once you have configured the Sandbox, you need to download some sample code to put onto the Sandbox. To enrich the data, we will use a country code set of data, which will give us a lot of information about individual countries. For this example, we will reuse the country file that we downloaded from Azure in Chapter 1, *A Short Dash to Dashboarding!*. Instead of connecting directly to the file, we will download it to a CSV file in Excel. To do this, connect to Windows Azure DataMarket using a Windows Live ID. To do this, please visit https://datamarket.azure.com/.

The data can be found at https://datamarket.azure.com/dataset/oh22is/countrycodes#schemaAbout and, halfway down the page, look for the link to **Explore this Dataset**. On the right-hand side, you will see the option **Download Options**. Select the option to download as CSV.

You will see that there are a lot of columns, and we won't need them for this example. In order to make the example simple, let's keep only the following columns:

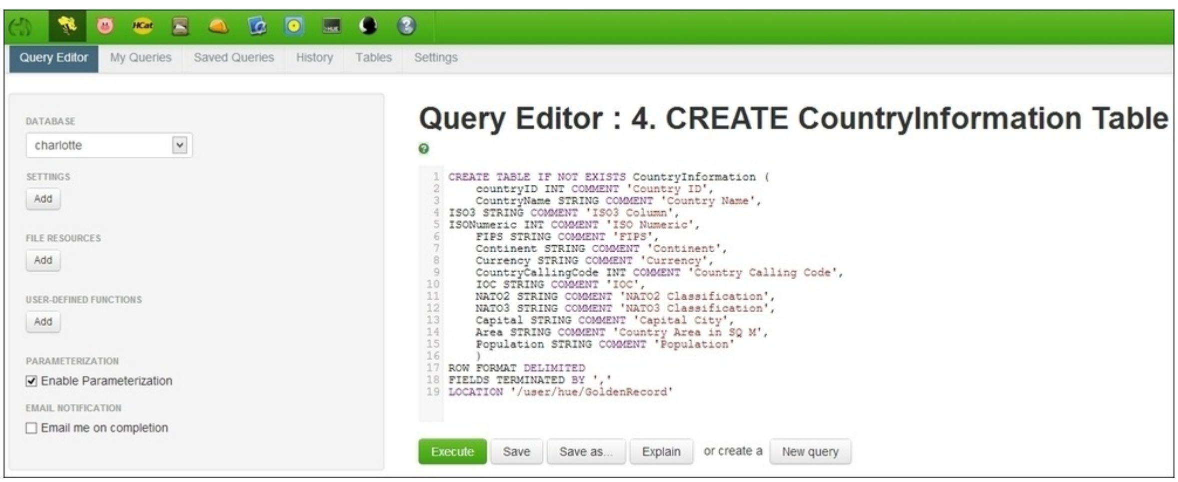
- `Area`
- `Capital`
- `Continent`
- `Countrycallingcode`
- `Countryid`
- `Countryname`
- `Currency`

- `Fips`
- `IOC`
- `ISO3`
- `Isonumeric`
- `Nato2`
- `Nato3`
- `Population`

Now we are ready to use the Big Data technology in order to enrich our data in Tableau.

How to do it…

1. Let's upload our CSV file into the Hortonworks Sandbox. This is straightforward. Once you have the Hortonworks Sandbox open in your browser, create a directory called `GoldenRecord`. To proceed, go to the **File Explorer** option and select **Upload File** to the `GoldenRecord` directory.
2. Once you have uploaded the file, let's create the table. To do this, we will run a query in the **Query Editor** interface. You can see an example of this interface in the following screenshot:

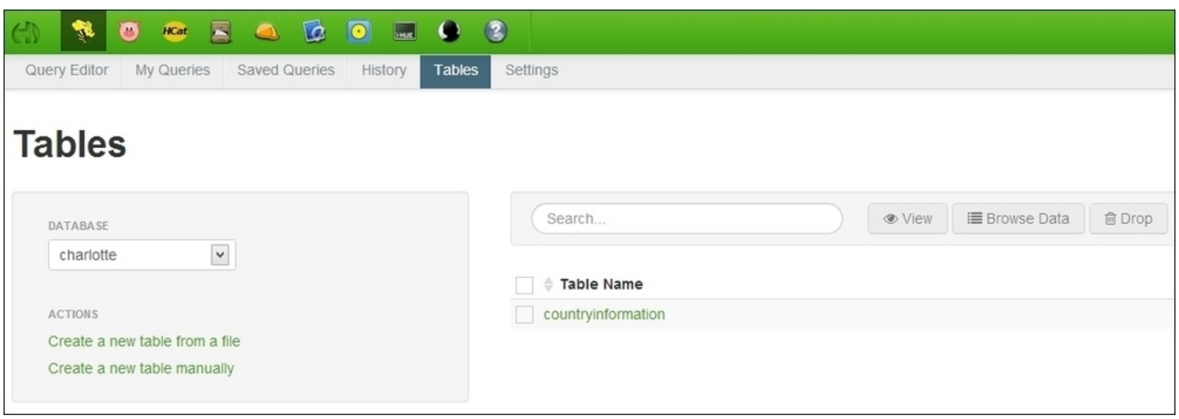


3. Copy the following script and paste it into the query editor as shown in the preceding screenshot:

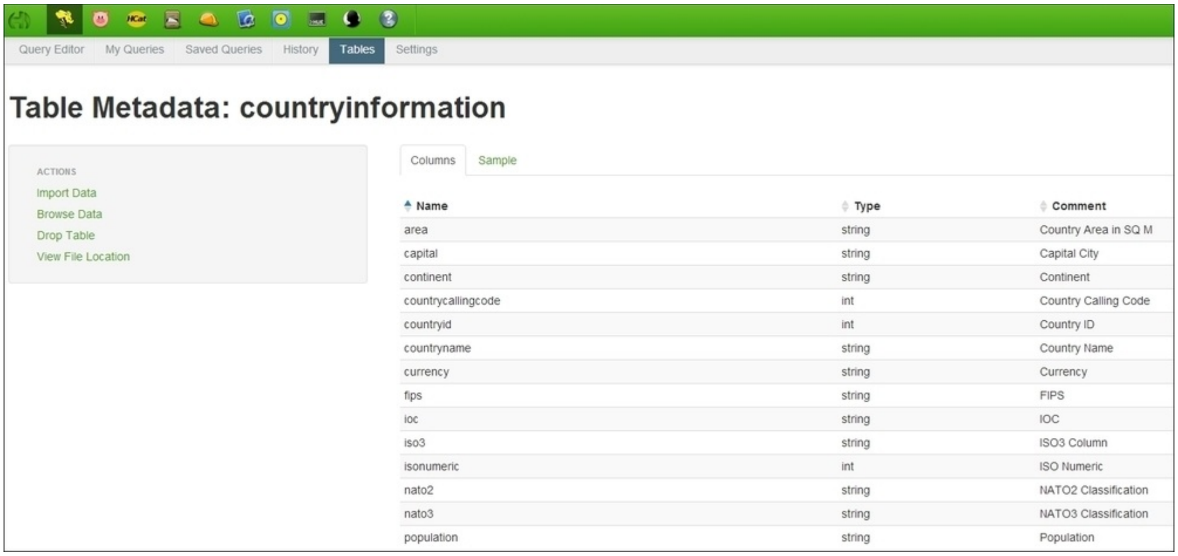
```
CREATE TABLE IF NOT EXISTS CountryInformation (
    countryID INT COMMENT 'Country ID',
```

```
    CountryName STRING COMMENT 'Country Name',
ISO3 STRING COMMENT 'ISO3 Column',
ISONumeric INT COMMENT 'ISO Numeric',
    FIPS STRING COMMENT 'FIPS',
    Continent STRING COMMENT 'Continent',
    Currency STRING COMMENT 'Currency',
    CountryCallingCode INT COMMENT 'Country
Calling Code',
    IOC STRING COMMENT 'IOC',
    NATO2 STRING COMMENT 'NATO2 Classification',
    NATO3 STRING COMMENT 'NATO3 Classification',
    Capital STRING COMMENT 'Capital City',
    Area STRING COMMENT 'Country Area in SQ M',
    Population STRING COMMENT 'Population'
    )
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LOCATION 'userhue/GoldenRecord'
```

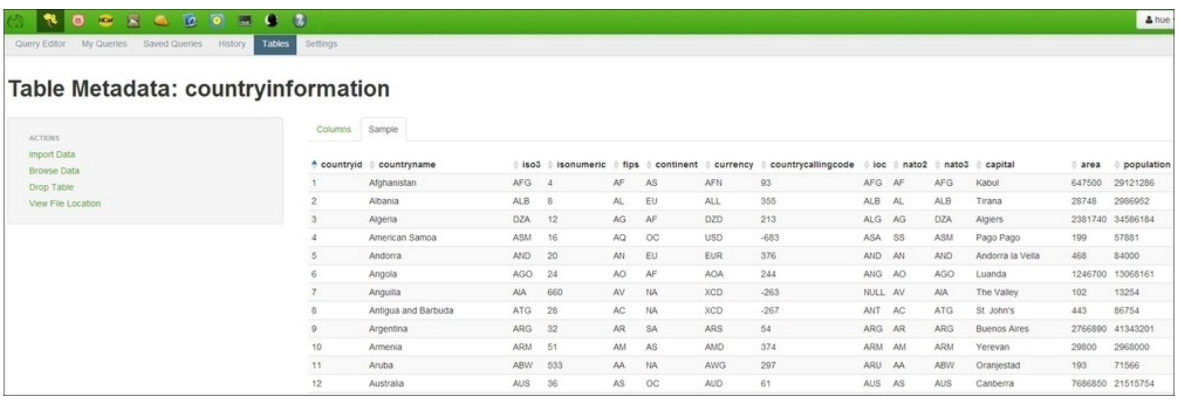
4. Once you have done this, click on the **Execute** button and the script will create a table called `CountryInformation`. When you click on the **Tables** option, you can see the `CountryInformation` table. You can see this in the following screenshot:



5. If you click on the **CountryInformation** link, you will see the columns that you created. The following is an example of the columns as seen in the Hue browser:



6. You can also see a sample of the data by clicking on the **Sample** link. The following is an example of the data:

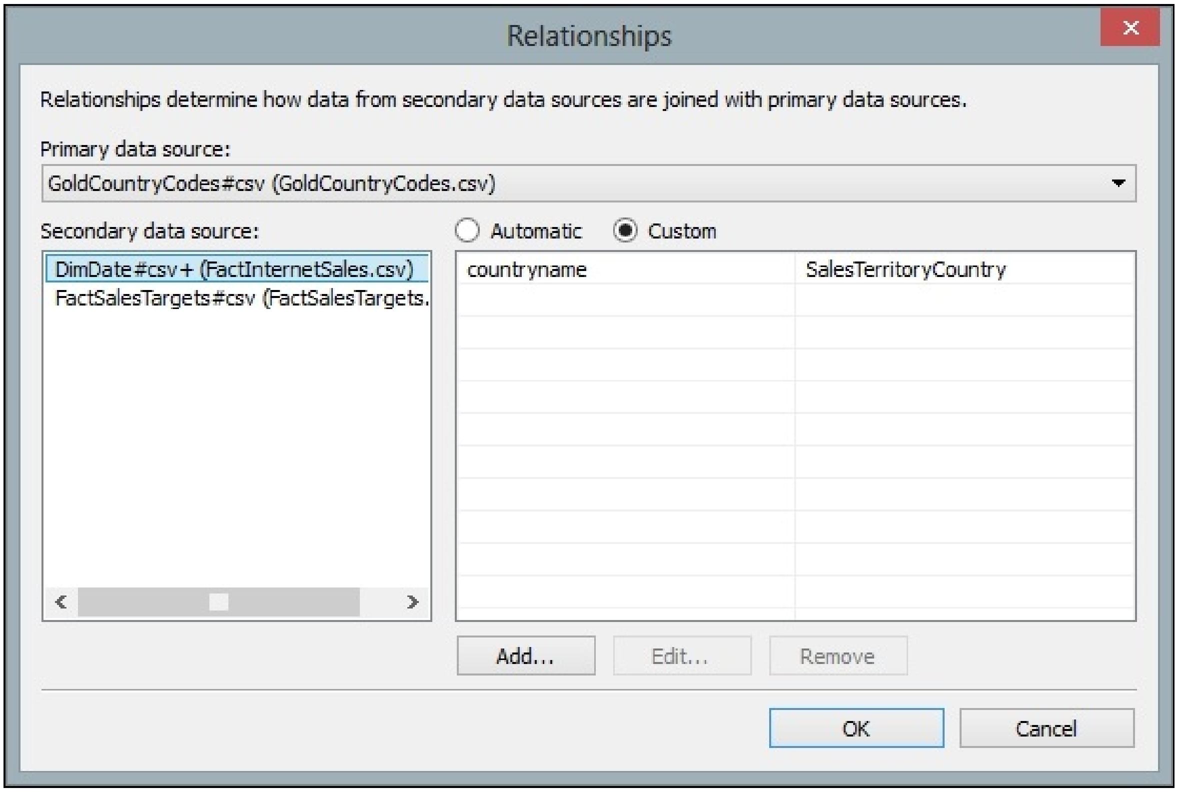


We are now finished working with the Hortonworks Sandbox. Next, we need to use Tableau to connect to the Hortonworks store. There are two ways to do this. If you have Tableau Professional edition, then you have enabled connectivity to Hortonworks. You can see this if you go to **Connect to Data** on the Tableau workbook. If you have Tableau Desktop edition, you will need to use Excel to connect to the Hortonworks Sandbox and download the data from there. Once the data is downloaded to Excel, you can store it and connect to it easily. If this was a real-life scenario, this wouldn't be satisfactory because the data might go out of date very quickly.

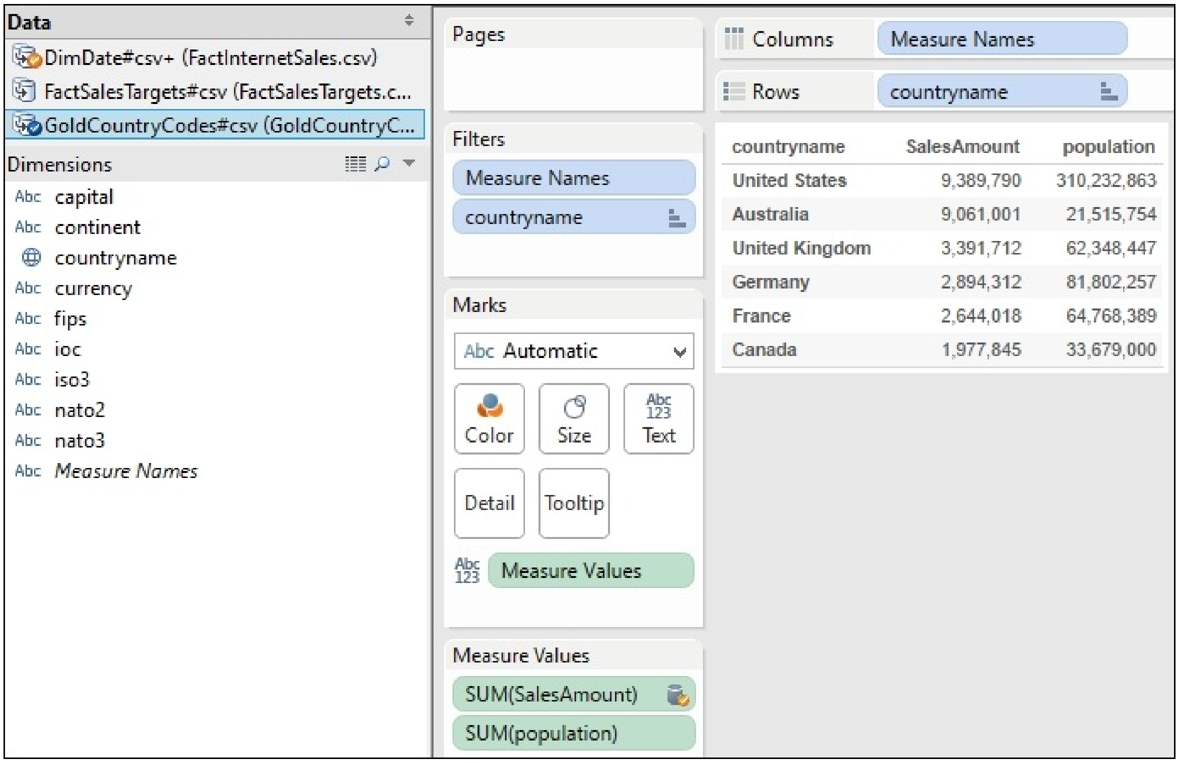
7. However, for the purposes of our simple example, the data could be

loaded from Excel into Tableau Desktop edition. However, we'll connect to the data source, and let's call it `GoldCountryCodes`.

8. Once the data is in the Tableau workbook, we can join it to the other tables using the country name. You can see an example of the joining in the following screenshot:



9. Once the data is joined together, we can visualize it in Tableau. We will have a mashup of the CSV files and Big Data technology—all in the same Tableau workbook.
10. For example, you could use the **SalesAmount** measure and put it next to the **Population** data from the external file. As a starting point, you could make a table and then see where the data takes you! The following screenshot shows your starting point:



Why not try some of the visualizations in Tableau based on this data to see how it looks?

How it works...

Note that this recipe only uses small data files as an example, and it is not intended to be a real-world Big Data exercise where we are transferring petabytes of data. Excel is used as an accessible example of a data source for training purposes.

The beauty of mixing Big Data sources with Little Data is that the user is insulated from the size of the data. Instead, they can visualize their data from different stores and different formats.

Fortunately, Tableau offers us a royal road to understanding the data by helping us to visualize it quickly and easily. It also allows us the ability to explore the data so that it starts to make sense regardless of whether it is Big Data or the important Little Data that makes up the data currency of the enterprise. We can enrich our existing data stores by using Big Data technologies, and this is the theme that we explored in this recipe.

In this recipe, we also made changes to the data by enriching it with a Big Data source. Big Data solutions are becoming more prevalent, but there is still a need for simplicity in accessing data regardless of its size. In this example, Tableau used a simple ODBC connector to access the data held in the Hortonworks Sandbox. A common experience among data analysts is not being able to get access to the data that they want. Therefore, the simplicity of accessing the data is vital, and ODBC is a common way of accessing data that is familiar to IT professionals.

Once we have access to the data via ODBC, there is no stopping us! Tableau then sees the data as another data source, in the same way as it sees data from Excel or OData, for example. In other words, this mechanism is a great "leveler" of data access since the data is accessible regardless of its size.

Tableau obviously cannot suck in petabytes of data (yet!), and this is one scenario where Big Data will need to stay outside of Tableau as an external data source. On the other hand, as we saw earlier, Business Intelligence requirements often involve summarizing data for averages, counts, and so on. It can be useful to crunch the data down into manageable summaries, and Tableau could access the summary data rather than the full Big Data itself. These issues are architectural questions, but a summary is a good place to start before moving forward to bigger questions.

There's more...

If you are interested in learning more about Big Data, you will find that the Hortonworks Sandbox already has a number of preconfigured tutorials. This is a great resource to get you started looking at Hadoop.

Filtering your data for focus

Dashboards are very compact ways of communicating data because they are constrained by space. One way in which we can make more of the "real estate" on the dashboard is to use filters.

Tableau has three ways to filter dashboards. Global filters apply to every part of the workbook that uses the same data source. This might be a problem, however. What happens if you want the filter to apply in some cases, but not others?

Local filters are specific to only one region of the dashboard. However, this may make them too restrictive. Tableau 8.1 now has a new filter feature, which allows you to stipulate a selection of worksheets for the filter rather than being specific to a data source. We can apply the filter to all worksheets that use the data source, and to do this, you can choose the **All Using This Data Source** option. It is also possible to let the filter apply to only the current worksheet, and you would select the **Only This Worksheet** option for this.

In this recipe, we will look at the new filter advancements of Tableau 8.1. We will work towards changing a chart into a filled map to show the sales amount, filtering by year. We will add in some new dashboard elements and get them to "talk" to one another by the use of filters. We will apply our filter to selected worksheets rather than the previous editions of Tableau, where it was more "all or nothing" in terms of filtering the data visualization.

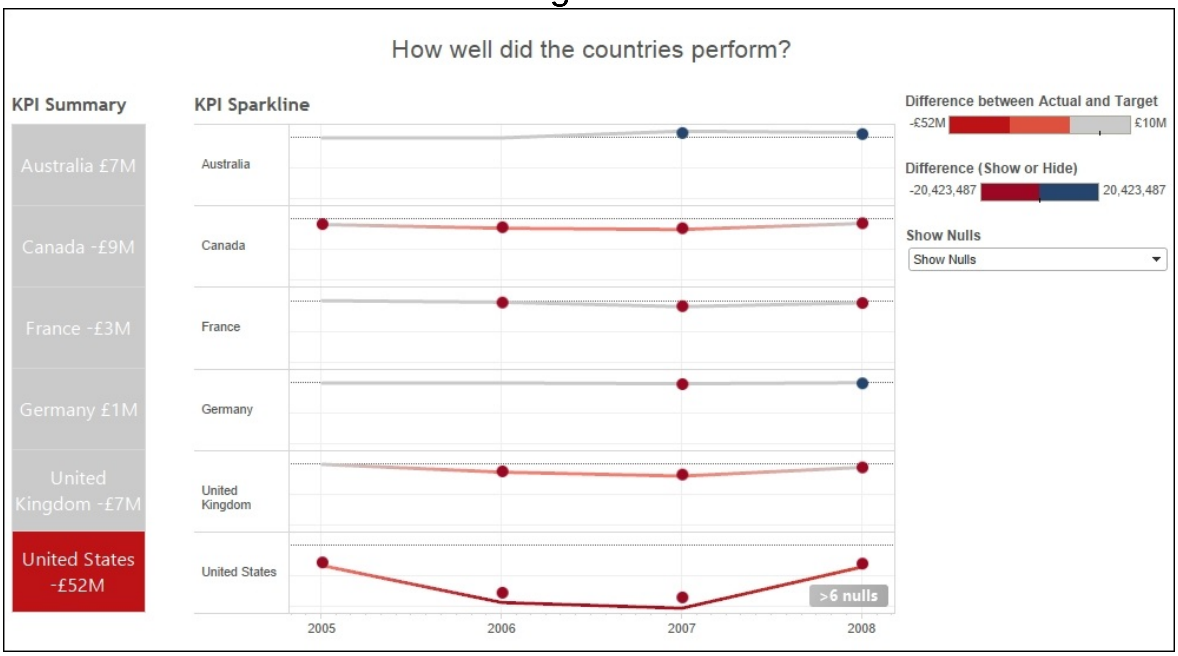
For the purposes of this recipe, we will want to select only some of the worksheets, so we'll select the **Selected Worksheets** option.

Getting ready

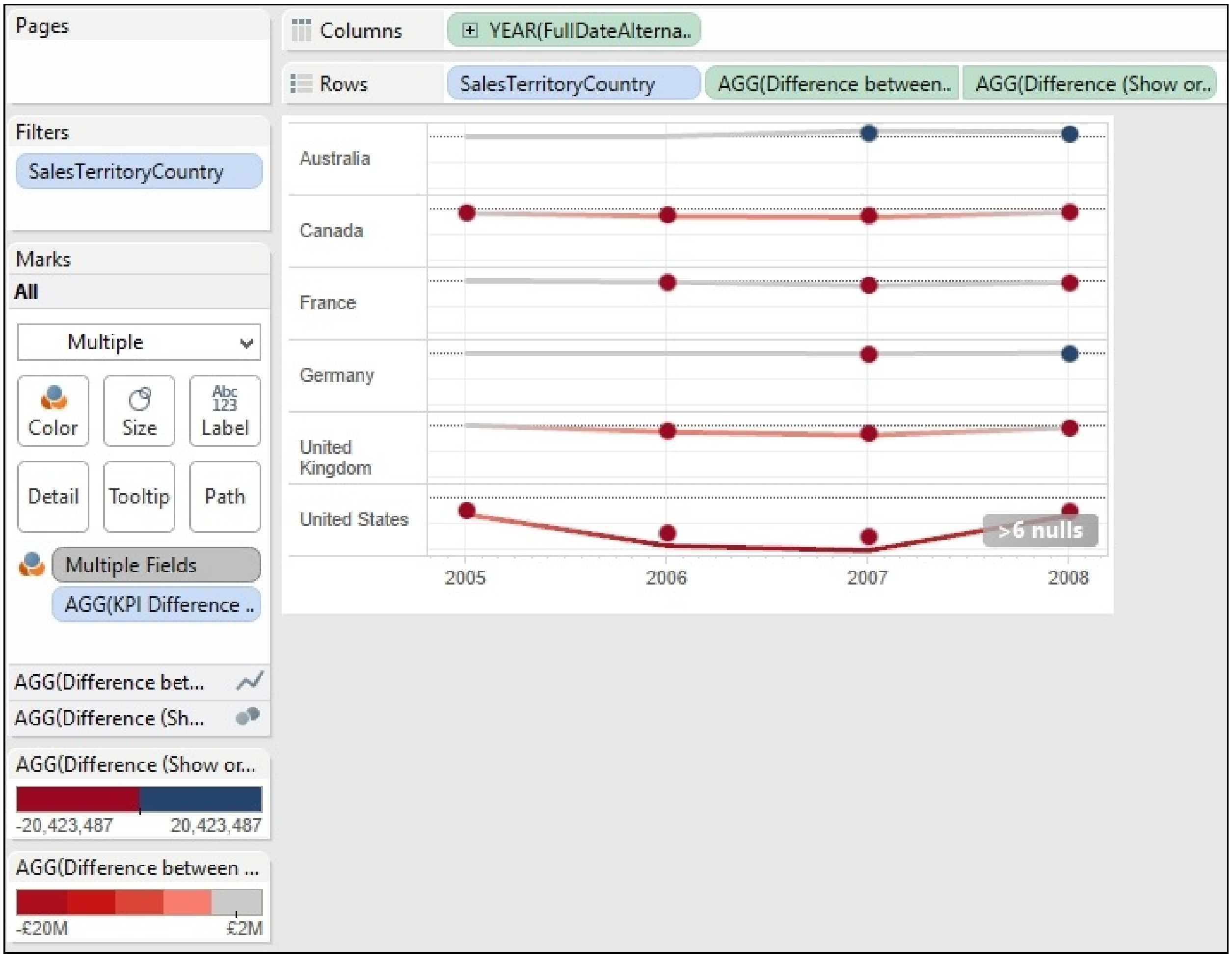
In this recipe, we will continue to use the workbook we created for previous recipes of this chapter. There is no need to add more data sources.

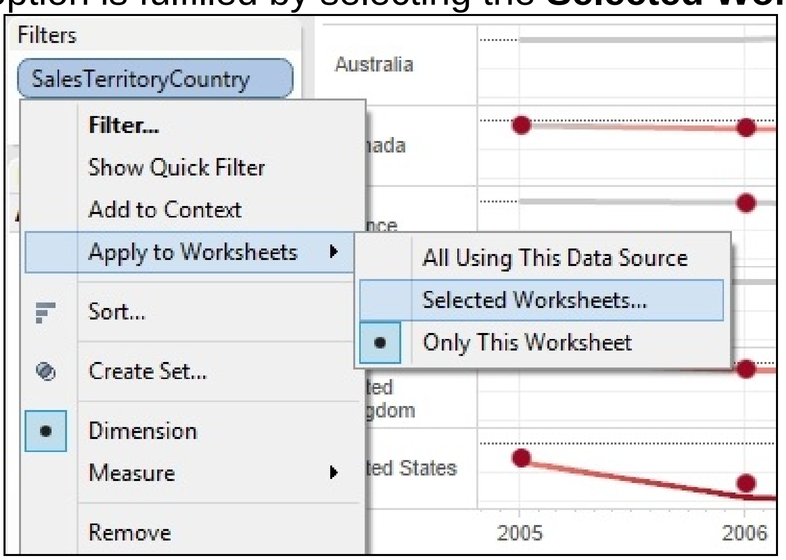
How to do it…

1. Create a new dashboard called **KPI Analysis**.
2. Take the **KPI Summary** and **KPI Sparkline** worksheets and put them in the dashboard by dragging-and-dropping them into place.
3. Next, add a title at the top of the dashboard, asking the question `How well did the countries perform?`
4. Place a **Blank** object underneath the two worksheet objects so that the countries are aligned and read left to right. Your dashboard should now look like the following screenshot:

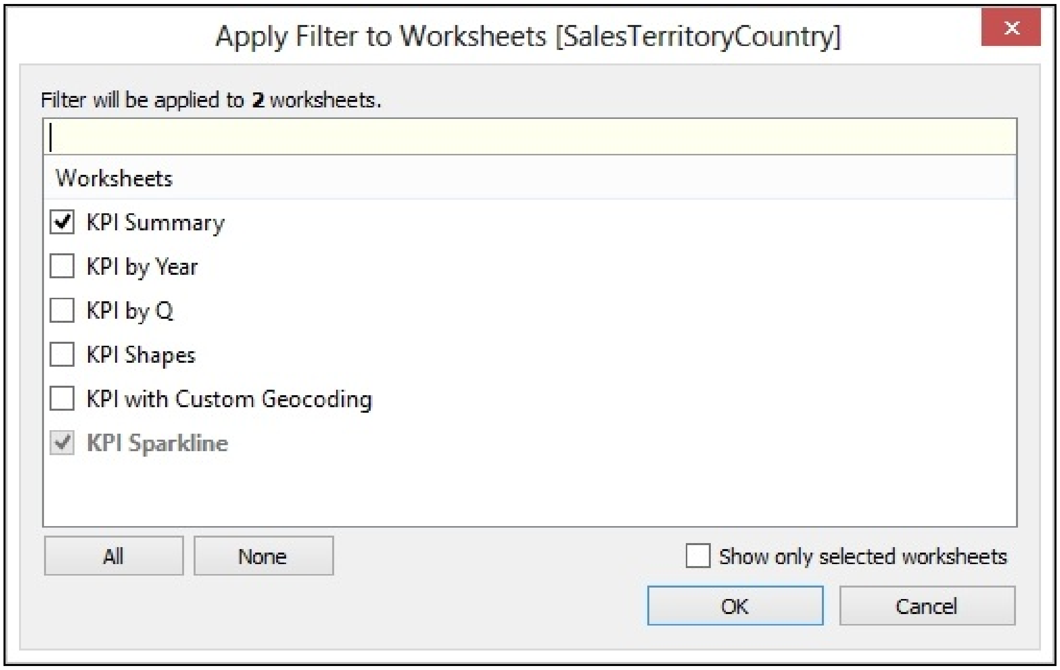


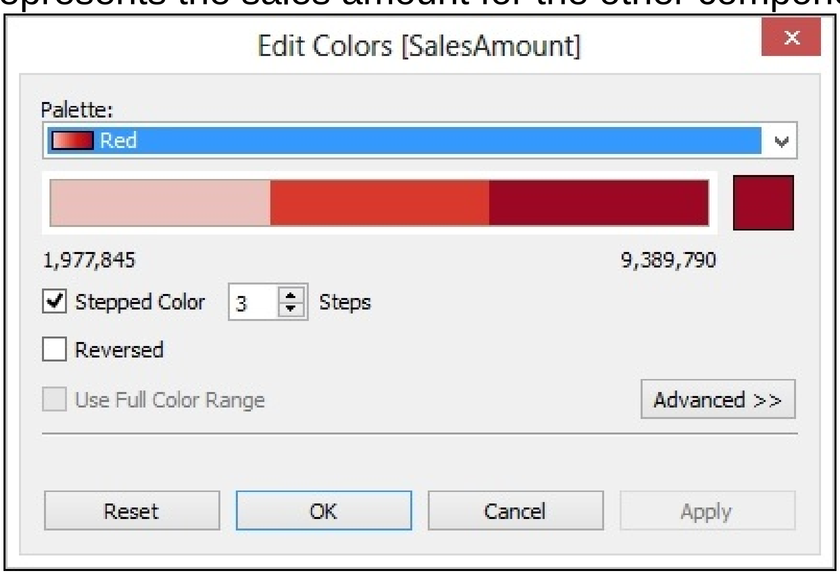
- When we go to the original **KPI Sparkline** sheet, we will add the **SalesTerritoryCountry** filter so that we can filter by country. We do this by dragging the **SalesTerritoryCountry** attribute to the **Filters** shelf. The following screenshot is an example:



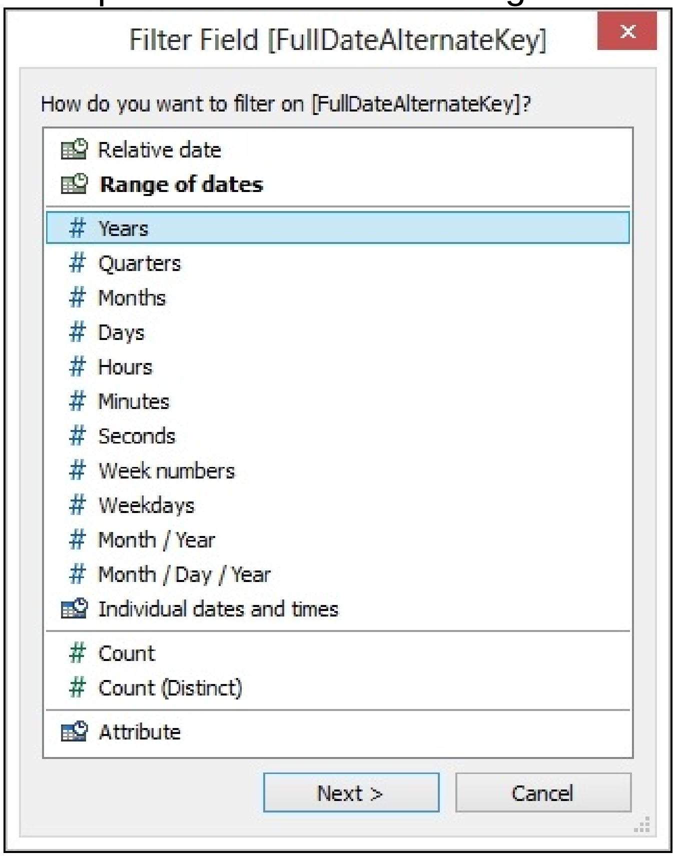
- This option is fulfilled by selecting the **Selected Worksheets** option as follows: 

- For the purposes of this recipe, we will select the **KPI Sparkline** worksheets. We do this by simply checking the boxes next to these names. The following screenshot shows an example of the dialog box:

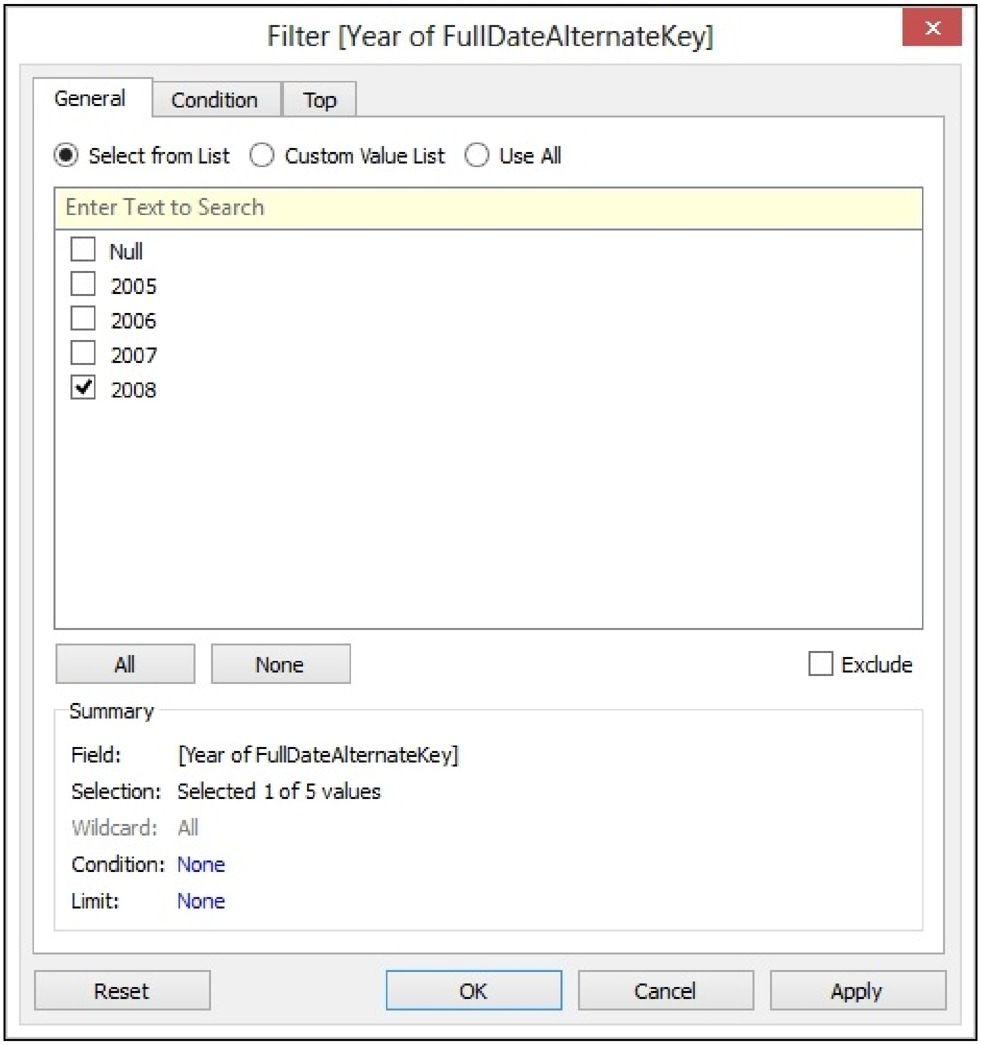


- Once you have clicked on the checkboxes next to the worksheets, click on **OK**. Once the filters are created, you will need to click on them to display the filters. To do this, click on the filter and choose the **Show Filter** option.
- Now, go to our **Golden Record** worksheet.
- To change the **Golden Record** visualization to a map, simply select the **Filled Map** option from the **Show Me** pane.
- Remove **Population** if it is present, and use **SalesAmount** to illustrate the sales amount value. In the following example, red has been used because it represents the sales amount for the other components of the dashboard: 

- Let's filter the filled map to only show the latest year's data by simply taking **FullDateAlternateKey** and dragging it to the **Filters** shelf. When you do this, a **Filter Field** dialog box will be initiated. We will select the **Years** filter and will restrict the data so that only the data from the latest year is shown, which in this case is the year 2008. You can see an example of this in the following screenshot:

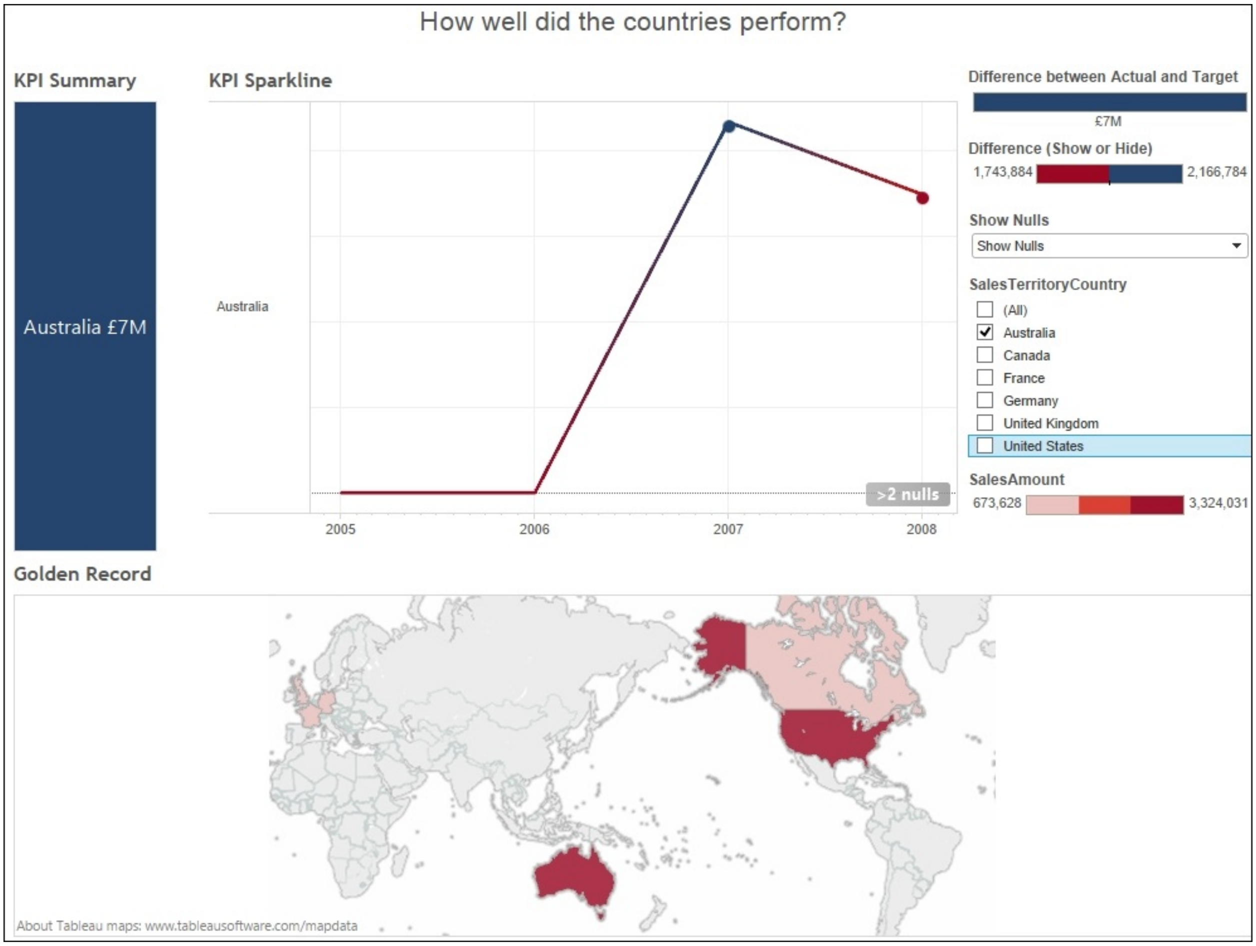


- The next step, as shown in the next screenshot, is to select the year **2008** from the **Filter** dialog box; this will filter the map. However, we are not applying this filter to the rest of the dashboard. This means that the map will stay static even though we have filters shared across the other components of the dashboard, **KPI Summary** and **KPI Sparkline**.



- Once you have selected the year **2008**, click on **OK**, and we are returned to the map.
- Finally, let's put the **Golden Record** worksheet with the filled map into the **KPI Analysis** dashboard. It is static data, and we will place it down the bottom so that it acts as an anchor for the rest of the more detailed data.

The Tableau dashboard will now look as follows:



How it works...

Filtering data is a key part of the *Visual Information-Seeking Mantra* article by Professor Ben Shneiderman, and users expect to be able to filter and interact with their data. Although the principles will not hold if people only want very detailed row-level data, the mantra is good to keep in your head when you are thinking about designing your dashboard. Dashboards are about actionable overviews rather than the detail about one row. Hence, the summary is an essential part of providing the overview.

Filtering is a good way to promote engagement with your dashboards. In marketing, stickiness refers to anything that encourages readers to stay on your website. In dashboard creation and reporting, stickiness can refer to features that increase the likelihood that users will stay on your dashboard and use it.

We can use filters in order to make our dashboards more flexible in response to user input, which may help to keep the dashboard engaging and interesting for data consumers. A key aspect to dashboarding is that we need to make the most of the space while engaging the user in the key facts of the data. Filters can help us to do that easily in Tableau.

Once again, color is key to conveying the message of the data. In this example, red is used in both the **KPI Summary** and **KPI Sparkline** worksheets, and the color is split into three steps in order to simplify the classification of the sales amount. People don't always distinguish fine-grained nuances of color, and using the **Stepped Color** feature of the **Edit Color** panel makes the data simpler to understand.

There's more...

For the purposes of this recipe, we have only selected some of the worksheets, and the **Selected Worksheets** option serves our purpose.

See also

- http://www.ifp.illinois.edu/nabhcs/abstracts/shneiderman.html

Creating choices in dashboards using conditional logic

Logical calculations can make your analyses richer. They can also make things easier for a dashboard consumer. For example, logical calculations can help you to funnel the analysis to specific-dimension members, combine members to follow a business rule, or even remove values that are irrelevant to your investigation.

Normally, when we use filters, we select the attributes within a dimension. In this recipe, we will implement logical calculations so that users can choose different dimensions to describe the data. We will place a small control on the dashboard so that users can simply click to choose the dimension that they would like to see, which describe the sales amount data. Users can simply click on which dimension they would like to see, for example, by color, country, or product line.

In this recipe, we will need to make a calculated field using a logical calculation and a number of parameters, amend the colors, and so on. Our sequence is to set up some parameters, a calculated field, and then some filters. So, let's get started!

Getting ready

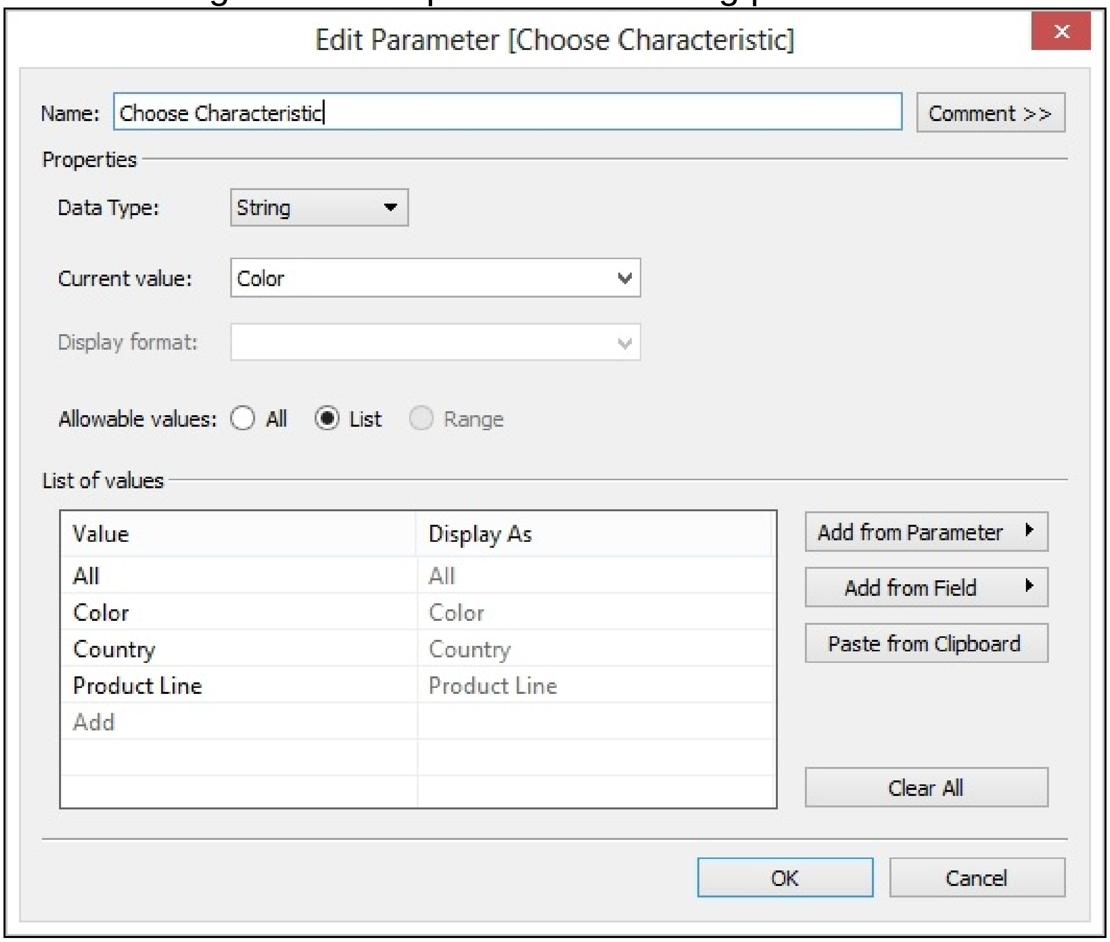
For the exercises in this recipe, we will continue to use the `Chapter 5` workbook. There is no need to add any more data. Let's make a copy of the **KPI Sparkline** worksheet and call it `KPI Dimensions`. We will also make a copy of the dashboard and call it `KPI Dimension Analysis`.

How to do it...

1. Firstly, let's set up a parameter so that the user can choose a metric. To set up a parameter, right-click on the **Measure** box and select the **Create Parameter** option. We will call our parameter `Choose`

`Characteristic` and will set up a list for each metric. Make sure that you set up **List** as an allowable value and type each metric name in the list of values.

2. The following is an example of the resulting parameter:



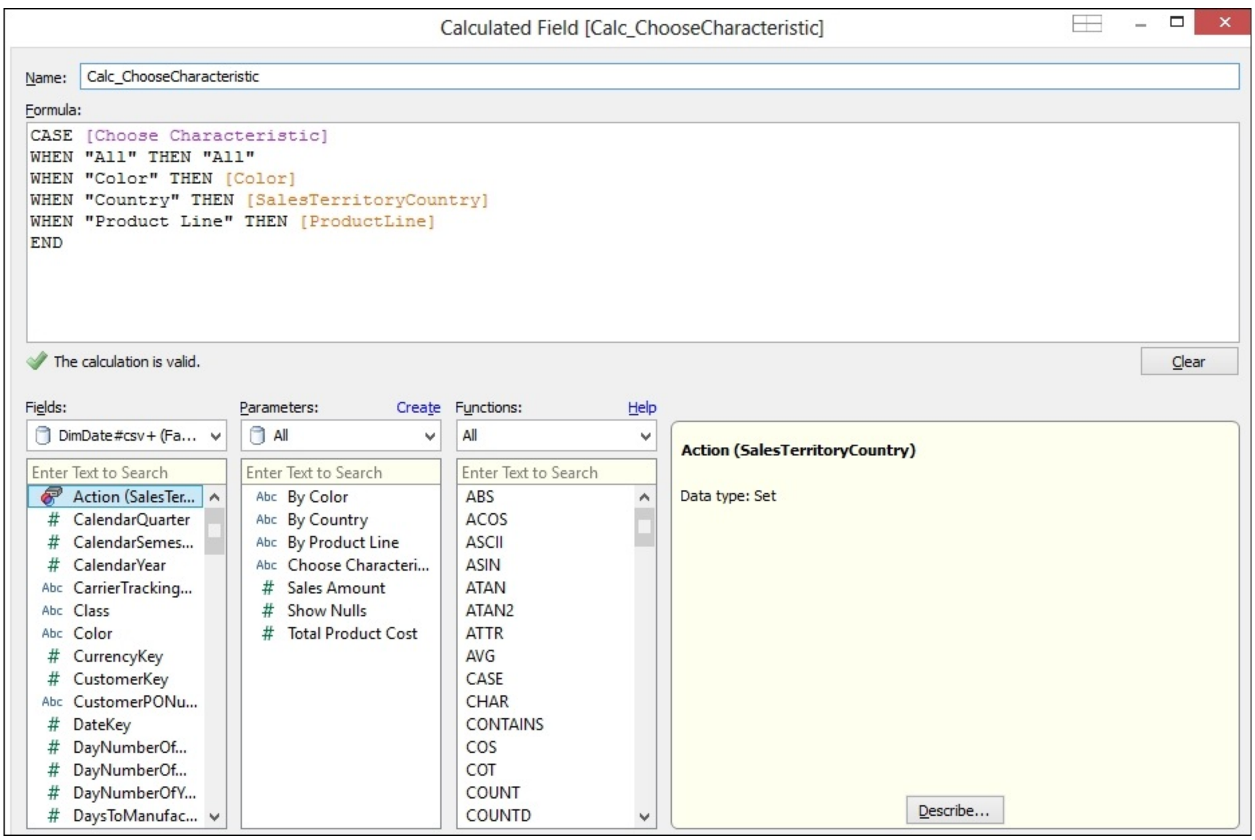
- Next, we need to add a calculation that will help us choose different dimensions. We will use a `CASE` statement, which is simply like lots of `IF… THEN…ELSE` statements strung together. The calculation will execute the first statement that it finds to be true.
- We will set up a `CASE` statement that chooses between dimensions, dependent on the user selection. We will offer the following choice of different measures so the dashboard consumer can select the measure they would like to see on the dashboard:

  - Sales amount
  - Sales amount quote

- To do this, right-click on the **Dimension** part of the Tableau workbook and select the **Create Calculated Field** option. Our logical calculation is written as follows:

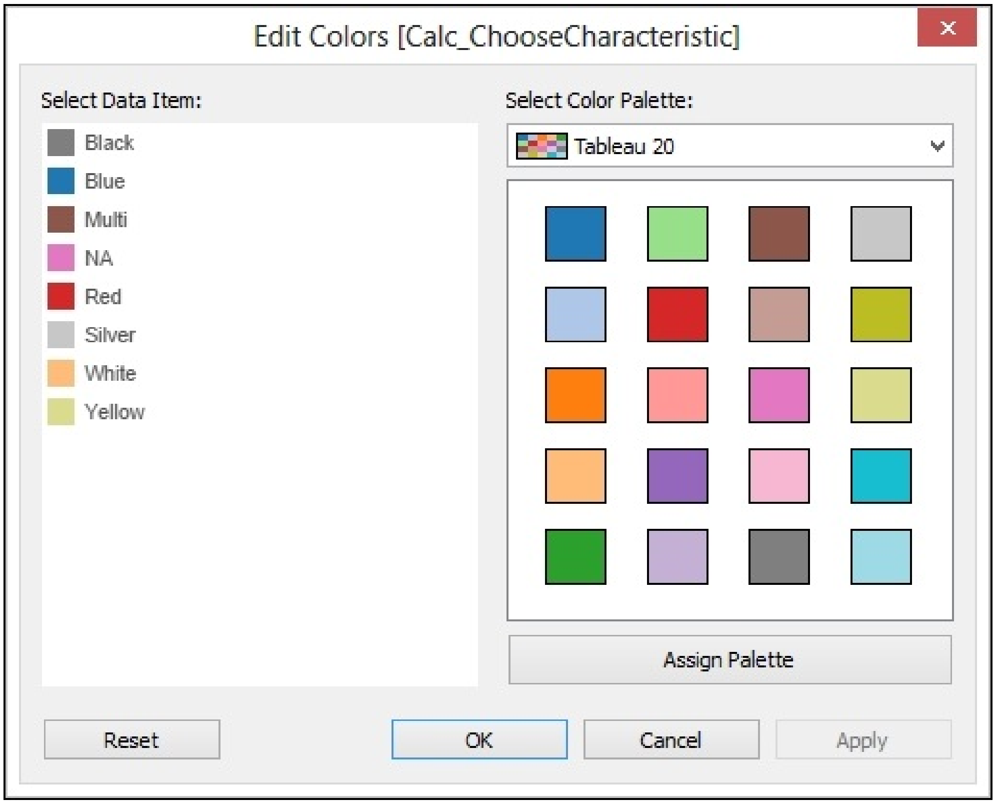
```
CASE [Choose Characteristic]
WHEN "All" THEN "All"
WHEN "Color" THEN [Color]
WHEN "Country" THEN [SalesTerritoryCountry]
WHEN "Product Line" THEN [ProductLine]
END
```

- The `CASE` calculation allows us to simply show all of the data not described by any dimension, or show by color, country, or product line.
- The logical calculation can be seen in the calculation editor, as shown in the following screenshot. Simply copy and paste the mentioned calculation and put it into the textbox. We will call the calculation `Calc_ChooseCharacteristic`.

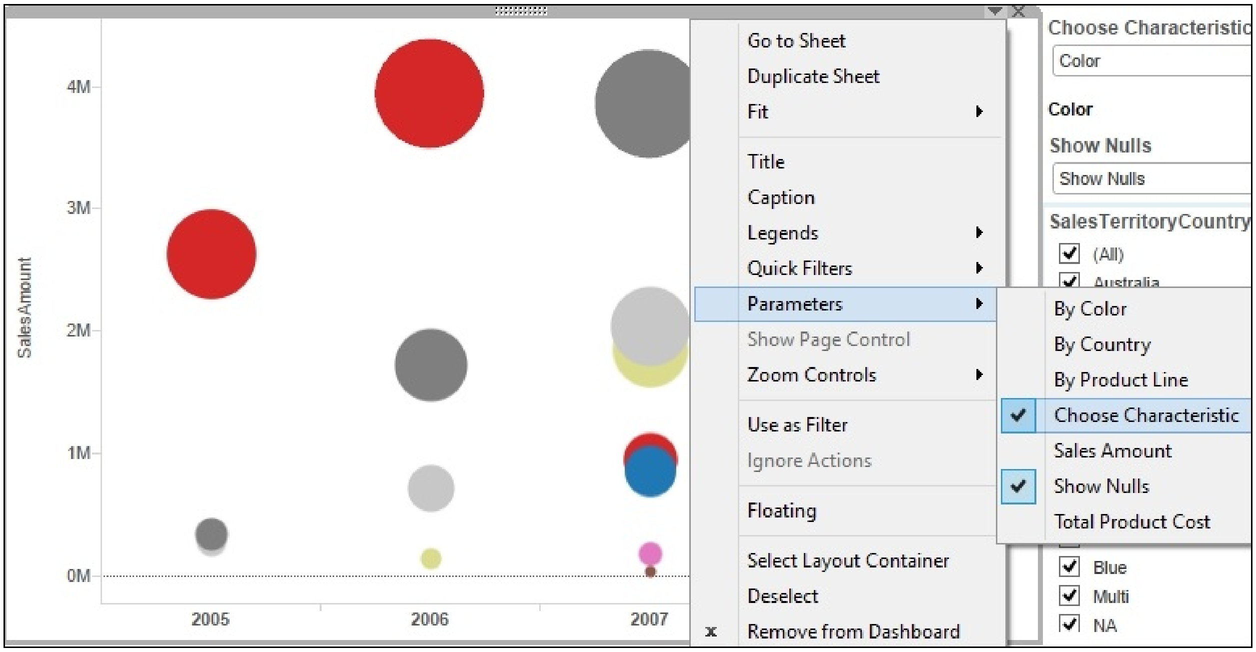


- Once you have clicked on **OK**, you will see it in the **Dimensions** pane in the sidebar in the Tableau workbook.

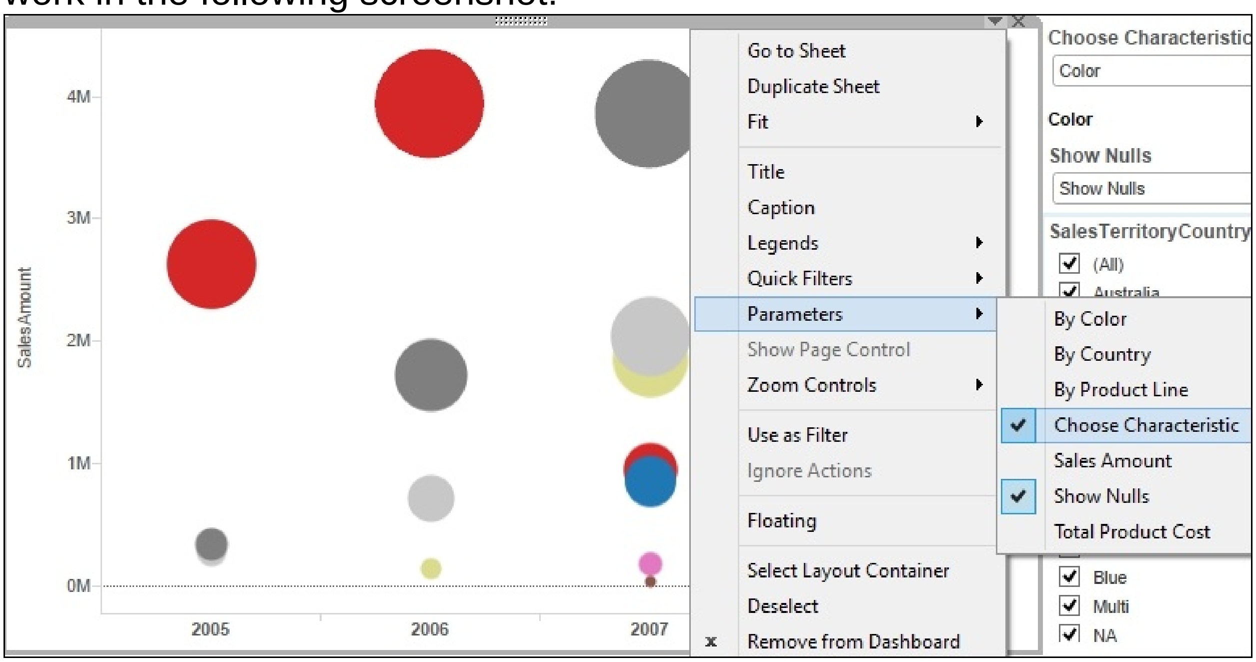
- Now, we need to set up some pills so that our workbook visualization is filtered according to the user-selected dimension. Now, take the **Calc_ChooseCharacteristic** filter and drag it to the **Marks** shelf so that the detail is retained.
- Next, we need to drag **Calc_ChooseCharacteristic** and put it to the **Color** button. This is a key part because it tells Tableau that it needs to change the display dependent on the selected dimension. The calculated fields implement the logic to denote which dimension should be displayed.
- This selected **Dimension** value is held in the **Choose Characteristic** parameter. The color is dependent on the result of the `CASE` statement evaluation in the **Calc_ChooseCharacteristic** calculation, and this is how Tableau differentiates in the display. We will need to make sure that users can select their preferred dimension, and to do this, they will need to see the parameter control. And you will need to see it in order to test it out! Simply go to the parameter called **Choose Characteristic**, right-click on it, and select the **Show Parameter Control** option.
- Since we are using **Color** as a potential dimension, we can make the palette-defined colors match the actual colors of the merchandise. To do this, click on the **Color** button and select the **Edit Colors** option. We can then set the color attribute **Red** to be red, the color attribute **Blue** to be set to the color blue, and so on. The following screenshot is an example:



- To finalize the visualization, you need to make sure that **Year(FullDateAlternateKey)** is on the **Columns** shelf and that **Sum(SalesAmount)** is on the **Rows** shelf.
- Let's set **Marks** to be **Circle** rather than **Automatic**.
- We can also add **SUM(SalesAmount)** so that the diameter of each circle becomes a representative of the **SUM(SalesAmount)** value.
- Let's go to our dashboard called **KPI Dimension Analysis**, which we copied earlier. Let's remove the **KPI Sparkline** workbook and insert the new **KPI Dimensions** workbook instead. We will need to show our parameter control; to do this, click on the **KPI Dimensions** area of the dimension, select **Parameters**, and select the **Choose Characteristic** option. We should also add in the **Calc_ChooseCharacteristic** filter, which is under **Quick Filters** in the same menu. You can see an example in the following screenshot:



- You can also remove the title and any headings so that you are making the most of the space. Overall in the dashboard, you will need to change the business question that has been posed. Here, we have changed it to `How well did sales perform, by different characteristics?`.
- You can see our final dashboard and the culmination of this chapter's work in the following screenshot:



- Our dashboard is now complete, and you can have fun changing the parameters and switching dimensions. This will create interesting patterns in the data for your users.

How it works...

In this recipe, we implemented logical calculations that allow users to change the dimension that appears on the dashboard. We are starting to make our dashboards look more interesting and more interactive.

Showing different dimensions helps the dashboard designer to make the most of the space, while maximizing the choices available to the dashboard user. Setting up all of the moving parts is not as quick as other visualizations, but it is worth the effort to help the users.

For this recipe, we will use the **KPI Dimension** worksheet and see the result in our dashboard.

There's more…

In this chapter, we have looked at many different ways to help users to engage with dashboards, such as choosing visualizations and making use of items to help with analysis such as parameters, Big Data technologies, and conditional logic. These tools will allow us to create dashboards that are powerful and rich, and who knows, maybe even fun!

Chapter 6. Making Dashboards Relevant

In this chapter, we will cover the following topics:

- Adding an infographic to your Tableau dashboard
- String manipulation using dashboards
- Correcting data exports from Tableau to Excel
- Blending data
- Optimizing tips for efficient, fast visualization

Introduction

Performance dashboards are used by management to gauge performance and how the business is making progress towards business goals. They can be hard to define since they apply to a wide spectrum of objectives, such as evaluating a business strategy globally or looking specifically at one department or team.

Tableau is a very easy software to use. By now, you are probably running through lots of scenarios where Tableau is useful to your business. Perhaps some of your colleagues are starting to eye your work so far and are looking at ways in which Tableau could be applicable to their teams.

Although Tableau is very easy to use, dashboards often fail their objectives because they are not aligned with the business goals. Perhaps the objectives themselves are poorly defined; in such cases, dashboards will simply reflect the poorly thought out objectives. It is possible that the dashboard will simply show a mediocre strategy, and accordingly, the business will only become good at executing a mediocre strategy. However, this situation must be better than executing no strategy at all and basing business decisions on a month-by-month reporting calendar. A dashboard is a picture that communicates the business vision clearly.

How can we be sure that a dashboard meets the expectations of the business audience? There are a number of important factors, which are as follows:

- Appropriate characterization of the target audience
- Who are the consumers of the dashboard, and what are their objectives and responsibilities?
- How well do they respond to change?

Often, the dynamics of an organization would be such that you cannot introduce a change too quickly. A dashboard can help you identify the key drivers that departments use when evaluating performance; then, you can start to align them with the drivers of other departments so that end user departments can start to work together. Communication is often a key failure point in many organizations, and this alignment is a step towards getting everyone headed in the right direction. The process of creating a dashboard can help in defining the clarity that is needed across the organization, since this process begins with communication.

It is essential to have a fitting definition of the metrics. What is going to be measured? If you are creating dashboards across the organization, then flexibility is going to be a key factor in order to facilitate the alignment. The metrics should be meaningful to the consumers with a logical structure and repeatable results. The dashboard should help generate and translate the data into actions that are aimed towards the organizational goals.

Further, more we must also have manageability of the metrics. Who is going to manage the dashboard? Is there a data steward within the organization? Dashboards are no good if nobody looks at them. Within many organizations, a lot of work has gone into reports and dashboards that have then been ignored and gone into obscurity. Sometimes, this happens because people do not like what they say! That being said, placing unrealistic metrics on a dashboard is a certain route to dashboard failure, since it will result in a lack of support, ultimately rendering the dashboard irrelevant.

It can be hard to meet these success criteria for dashboards. It can often be a challenge to define metrics that target strategic objectives such as return on investment or governance. The most useful dashboards are the return on investment of governance. The most useful dashboards are the ones that are implemented with a project sponsor who is senior in the organization and able to push the organization through a change. Is there a change champion within the organization? Are they onboard?

Dashboarding and the scorecard approach will become more prevalent with the emphasis on Big Data. Implementing Big Data approaches is only part of the Big Data story, and we will always need ways to learn from our data. This chapter will help us make our dashboards more meaningful to our organization using Tableau.

Adding an infographic to your Tableau dashboard

Sometimes people want an infographic rather than a data visualization. An infographic is a picture or a poster that illustrates a part of the story that is contained in the data so that the viewer understands it very quickly. Data visualization, on the other hand, allows the viewers to make up their own minds about the data, which can often be a longer process.

In this recipe, we will look at adding an infographic or a picture to the Tableau dashboard and putting it together with data visualizations that tell a story of the comparison of certain results between the United Kingdom and Australia. This recipe consists of a number of steps. Generally, we will use a background image and then configure the properties of the image using Excel. We will then go back to using Tableau. We will use the image and the Excel file to configure the appearance of the Tableau worksheet.

Getting ready

For the exercises in this recipe, take a copy of the `Chapter 5` workbook and name it `Chapter 6`. We don't need to add in any more data for now.

Configure the dimensions of your image in an Excel workbook and save it separately to connect to it later. We can record the length and width of the image as `X` and `Y` values.

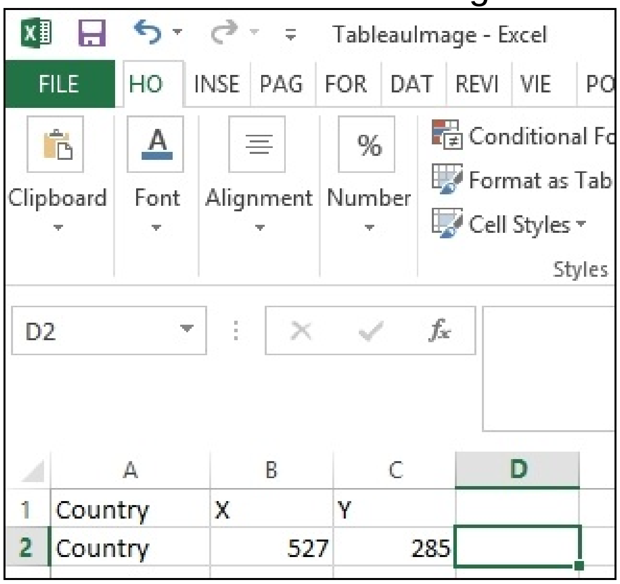
How to do it…

Firstly, we will choose your background image and use it as a base for your infographic. You can choose your own background image as well. For the purposes of this recipe, we have provided you with a sample image that you can download from http://bit.ly/TableauImage. Then, perform the following steps:

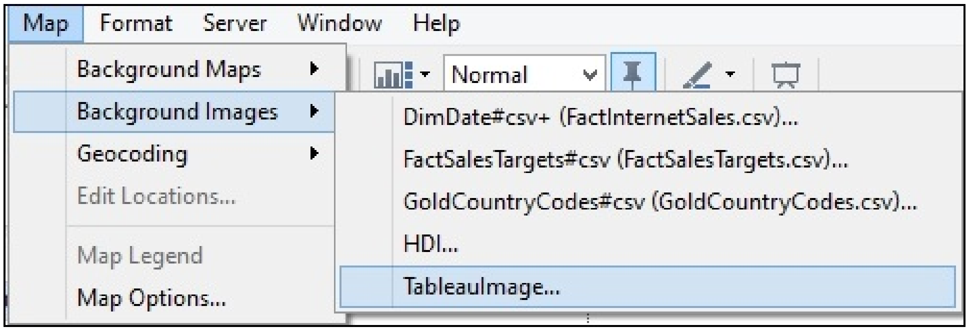
1. Create a new Excel workbook and call it `TableauImage.xlsx`.
2. In the Excel workbook, create a headings row by typing the following three items into the following cells:

    - Cell A1: `Country`
    - Cell B1: `X`
    - Cell C1: `Y`

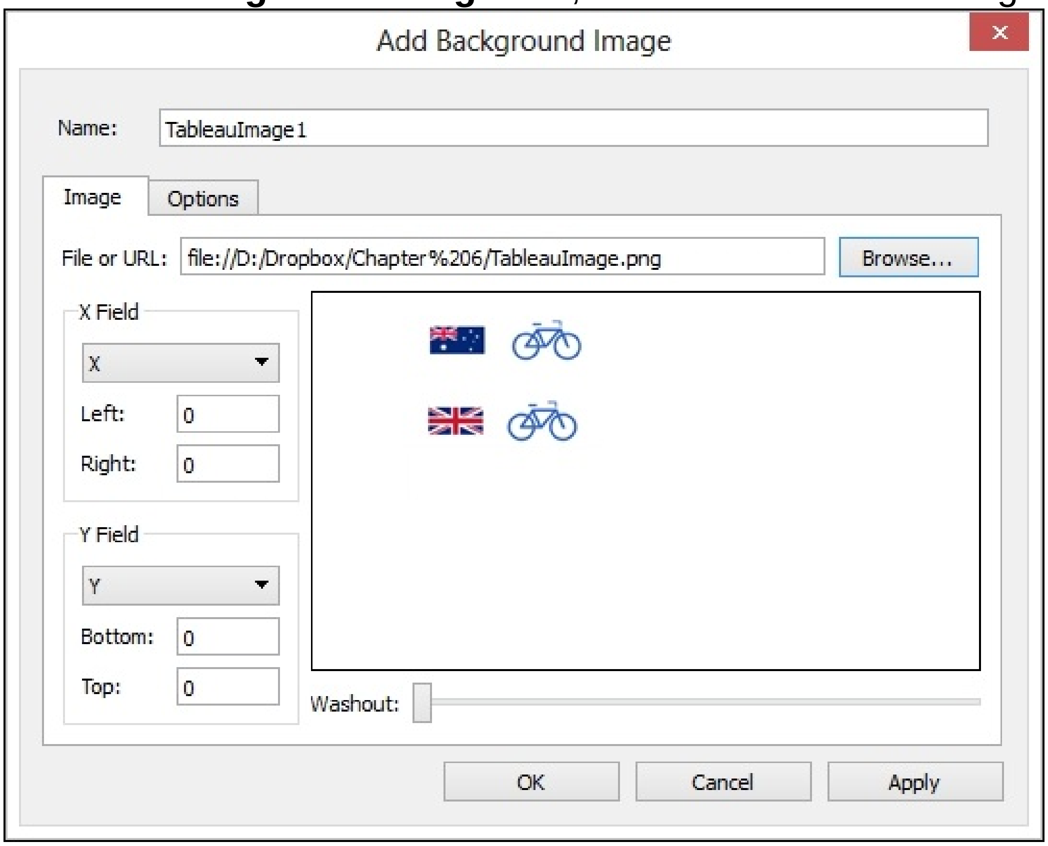
- Simply enter the word `Country` into cell A2.
- Next, we will record the width of the image in pixels in cell B2, which is the `X` value. For our example, the width is `527`. Your Excel workbook should look like the following screenshot:



- Now, we will record the height of the image in pixels in cell C2, which is the `Y` value. For our example, the height is `285`.
- Save your Excel workbook as `TableauImage.xlsx` and close it.
- Go back to Tableau and create a new worksheet called `KPI Poster`.
- Next, connect to the Excel workbook that you just created. To do this, go to **Data**, then click on **Connect to Data** and select the option **Microsoft Excel** on the left-hand side.
- Navigate to the `TableauImage.xlsx` file and choose the option **Connect live**.
- Next, let's connect to the image by setting it as a map. To do this, navigate to **Map | Background Images** and then select the Excel spreadsheet that contains the image dimensions, as shown in the following screenshot:

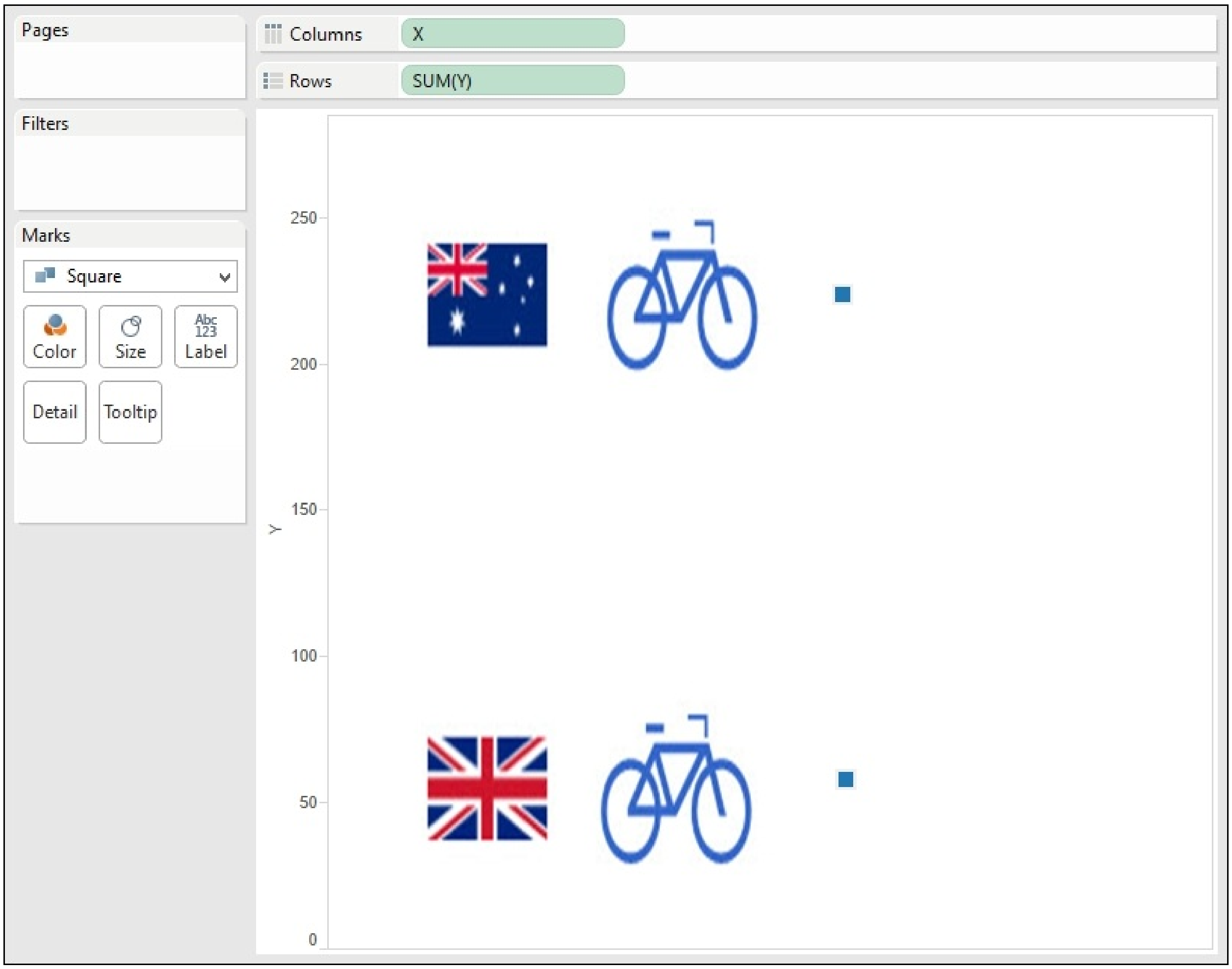


- Choose **Add Image** and then browse for the location of the image file in your `Chapter 6` folder. Now, select the file and you will see it appear in the **Add Background Image** box, as shown in the following screenshot:



- Next, we will need to enter the dimensions of the file in the **X** and **Y** axis fields. For the **X** field, leave **Left** as `0` and enter the width of the image in the **Right** field. In our example, the width is `527`, so we should enter this value in the **Right** field.
- For the **Y** field, leave **Bottom** as `0` and enter the height of the image in the **Top** field. In our example, the height is `285`, so enter this figure in the **Top** field.
- Click on **OK** and you will return to the Tableau worksheet. Rename the worksheet to `KPI Poster`.

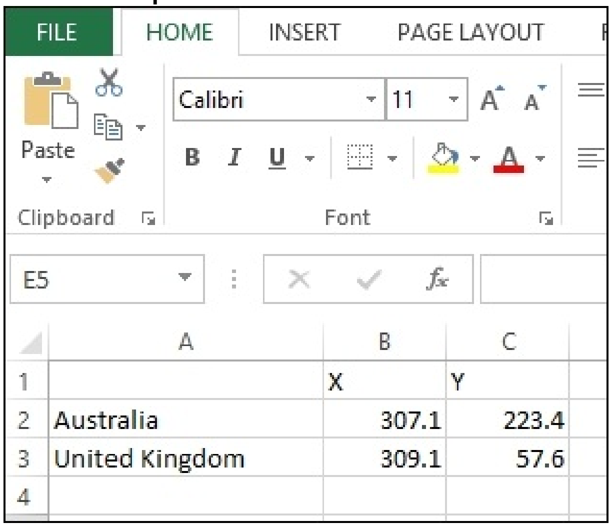
- Now, we need to plot our **X** and **Y** fields.
- Drag **X** on to the **Columns** shelf and right-click on it to select **Dimension**.
- Drag **Y** on to the **Rows** shelf. Your image will now appear.
- Right-click on the **Y** green pill on the **Rows** shelf and select **Dimension** from the pop-up menu.
- You may notice that you have the axis headers appearing on the *x* and *y* axes. To remove each axis, right-click on the axis and uncheck the option **Show Header**. A sample of this is illustrated in the following screenshot:



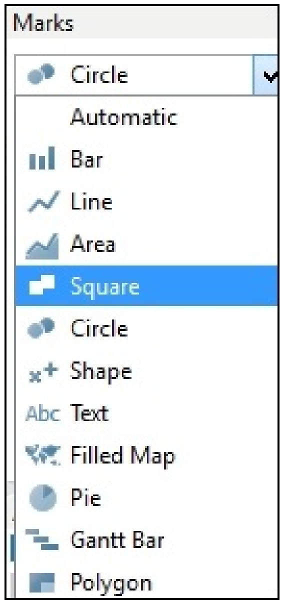
- Now, we can select the point at which we would like the data to appear. To do this, pick a spot to the right of the Australian flag, right-click on it, and select **Annotate** and then **Point**. When you click on **OK**, you can see the *x* and *y* coordinates for any point.
- Similarly, select a point for data related to the United Kingdom flag as well. To do this, pick a spot to the right of the United Kingdom flag, right-

click on it, and select **Annotate** and then **Point**. When you click on **OK**, you can see the *x* and *y* coordinates for any point.

- When you have the points that you want, open the Excel workbook called `TableauImage.xlsx`.
- Enter `Australia` in cell A2.
- Enter `United Kingdom` in cell A3.
- Your spreadsheet should look like the following screenshot:

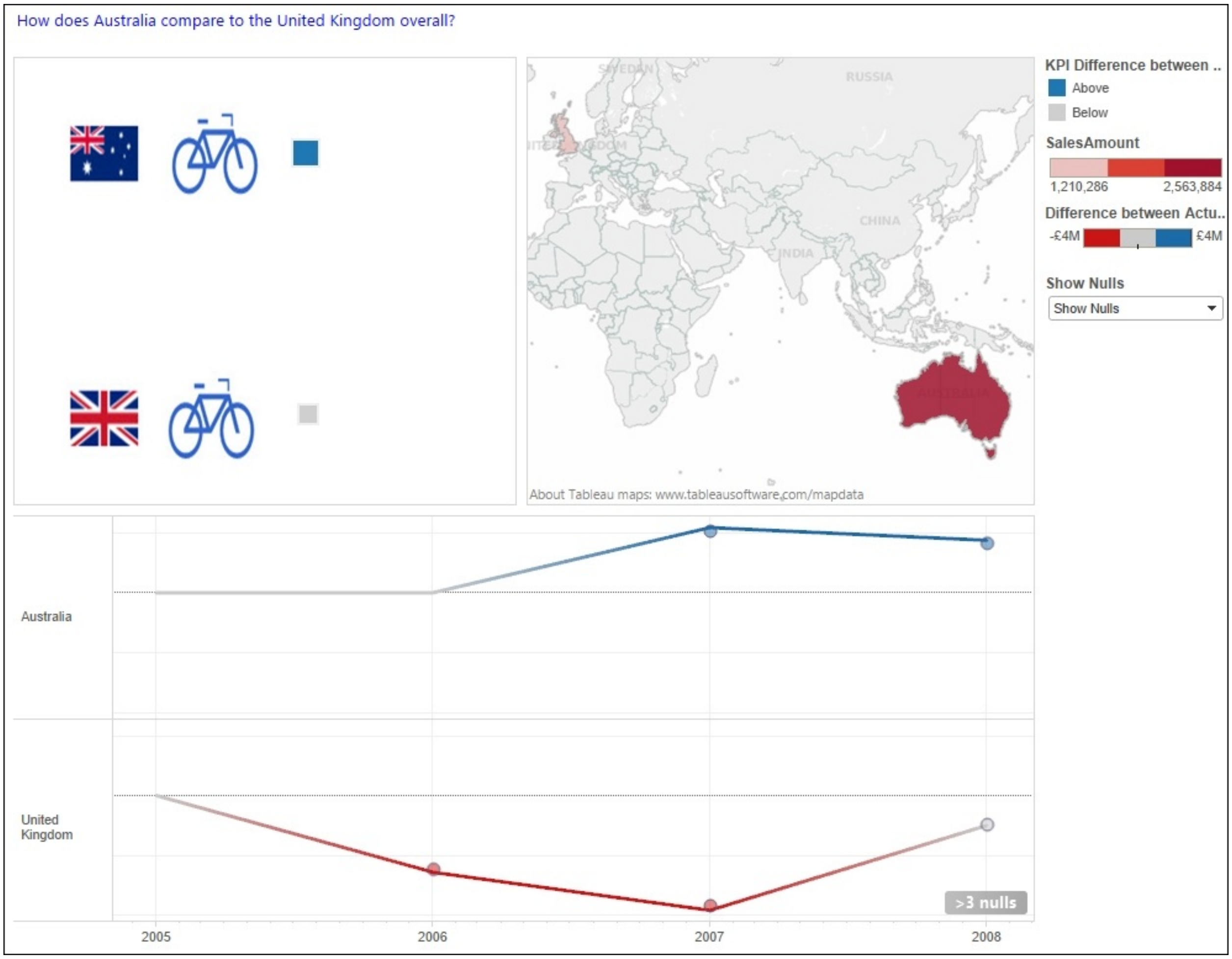


- Save and close the Excel file.
- Go back to your Excel worksheet.
- Right-click on the data source to refresh it. To do this, go to the **Data** field at the top left-hand side and look for the data source `TableauImage`. Right-click on it and select **Refresh**.
- Now, place the **Country** field on the **Level Detail** shelf of the **Marks** shelf.
- Drag **KPI Difference betweeen Actual and Target** on to the **Color** button on the **Marks** shelf.
- Now, **KPI Difference betweeen Actual and Target** is driving the color of the KPI. So, for the above KPI data, the color should be blue, and for the below KPI data, the color should be gray.
- You can choose to select the **Square** option to change the mark types. To do this, go to the **Marks** shelf and then to the drop-down list right under the **Marks** label. Look for **Square** and select it, as shown in the

following screenshot: 

- Let's turn this into a dashboard. Don't worry if the images look strange; the dashboard will be fixed when we put it together.
- Create a new dashboard and call it `Infographic Dashboard`.
- On the dashboard, drag our new image on to the top left-hand side.
- Drag the sheet **Golden Record** on to the top right-hand side.
- Drag **KPI Shapes** on to the bottom of the dashboard.
- For the **Golden Record** chart, simply right-click on France, Canada, Germany, and the United States and choose the **Exclude** option. Only the data for Australia and the United Kingdom should be displayed on the dashboard.
- To filter the **Golden Record** sheet, right-click on the arrow that is located at the top of the chart and select **Go To Sheet**. From there, drag the **countryname** attribute on to the **Filter** shelf and choose the **Filter** option. In the list of countries, select **Australia** and **United Kingdom** and then click on **OK**.
- Remove the title from each component of the dashboard. To do this, click on each of the components and look for the arrow at the top right-hand side. In the pop-up menu, click on **Title** to uncheck it.
- Finally, let's add a title so that the intent of the dashboard is clear. Drag the **Text** button from the left-hand side of the dashboard over to the top of the dashboard. For the text itself, enter `How does Australia compare to the United Kingdom overall?` into the textbox, changing the font to royal blue.
- Now, you've completed the steps. The dashboard will appear as

follows:



How it works…

In this section, we used a custom image and overlaid it with data points in order to produce an infographic that displays up-to-date data. This infographic can be used as part of a dashboard to complement and highlight the main message of the data.

To do this, we need a base image and an Excel workbook as a small data store. We use the Excel workbook in order to hold the image size of the picture, and we add to it throughout. When Tableau starts up, it loads the data from this Excel workbook, such as the file properties, so that the image can be loaded properly. It provides some information that helps us locate the data points at particular points on the image.

There's more…

If you want to know more about infographics versus data visualization, refer to http://www.jenstirrup.com/2010/12/data-visualisation-and-infographics.html.

String manipulation in dashboards

Annotations are useful for adding additional context to a data visualization. Data can be difficult to interpret without any context, and additional commentary can help to save the business user from having to find out additional information.

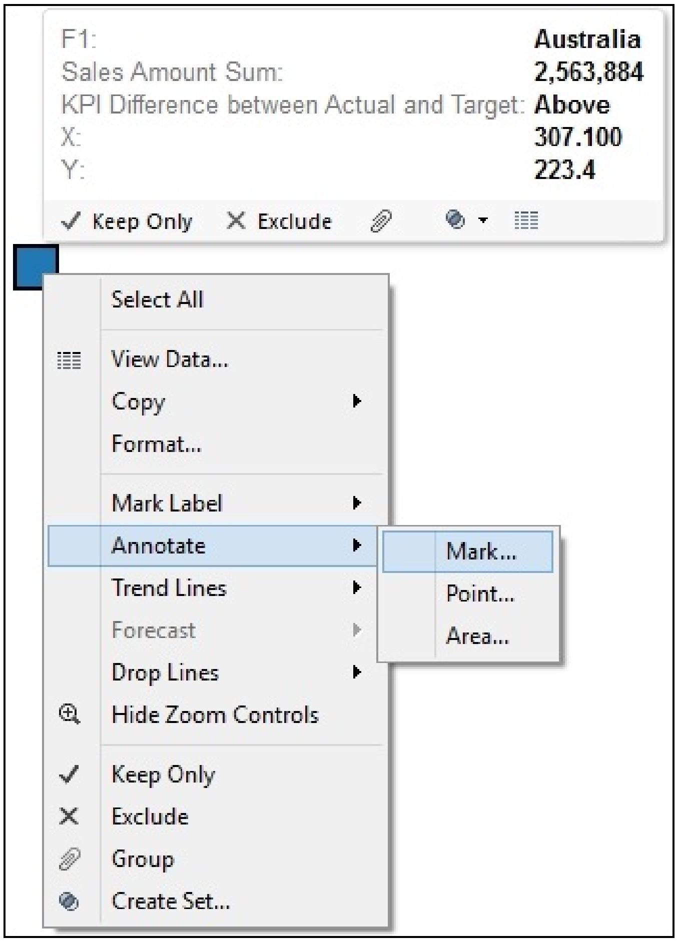
In this recipe, we will look at using calculated fields to amplify the message of the data and to automatically add new information to the annotations.

Getting ready

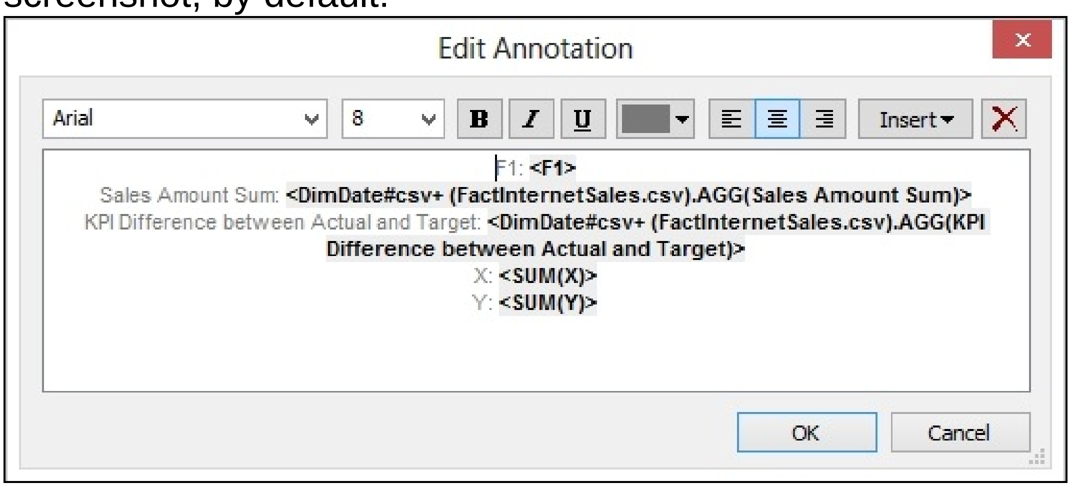
For the exercises in this recipe, we will build on the existing `Chapter 6` dashboard. We don't need to add in any more data for now.

How to do it…

1. Let's start by adding an annotation to the dashboard. Right-click on the square that we created near the Australia flag in the last recipe. In the pop-up menu, select **Annotation** and then **Mark**. You can see an example of this activity in the following screenshot:

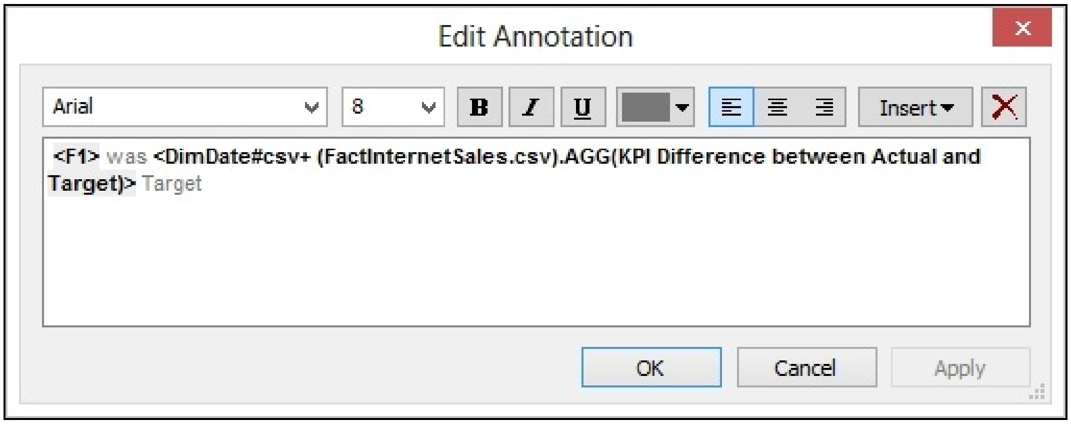


- Now, you will get a pop-up window that contains a default specification of the annotation. However, we will amend the annotation so that it is relevant to the data points. The pop-up window looks like the following screenshot, by default:

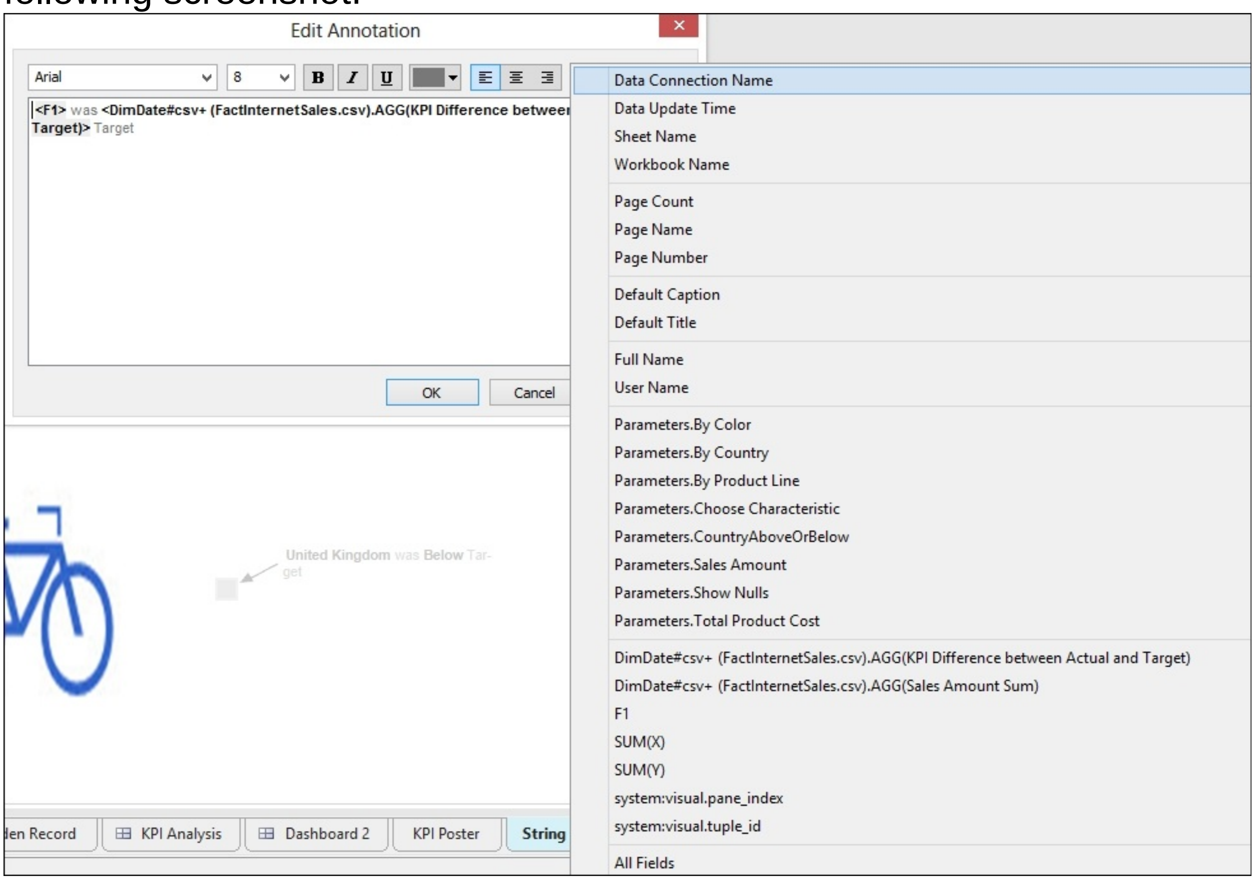


- Next, let's make the annotation more relevant to the visualization.

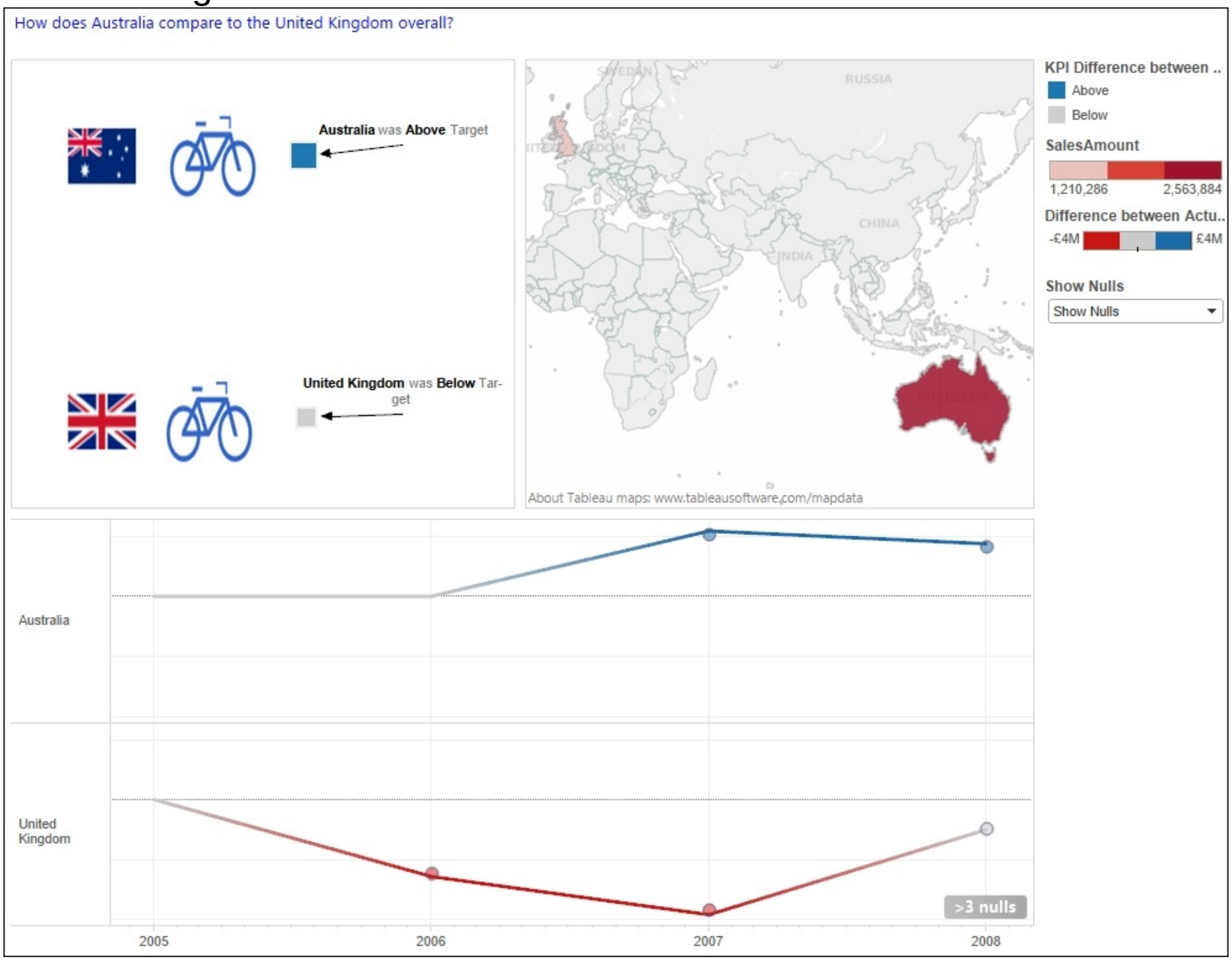
Remove the text in the annotation and enter <F1> was <DimDate#csv+ (FactInternetSales.csv).AGG(KPI Difference between Actual and Target)> Target. The pop-up window now looks like the following screenshot:



It is easy to add the content of our calculated fields using the **Insert** button on the right-hand side. When you click on the **Insert** button, it gives you a list of all the fields. You can see an example of this in the following screenshot:



- Go back to the **Edit Annotation** window by clicking on it and then click on **OK** in order to create the annotation.
- The annotation uses a mix of calculated fields and field names to construct an appropriate annotation. Our resulting annotation looks like the following screenshot:



How it works…

Strings can be set up using calculated fields, parameters, field names, or a combination of these. It is a nice feature to make your visualization punchy.

Correcting data exports from Tableau to Excel

IT agility can be increased by improving the quality of the data, for instance, by creating enterprise data standards. However, how can you increase data quality if you can't see the problems in the data? Data quality isn't just about wrong numbers or missing data. It can also refer to surplus data stores that your IT team end up looking after unnecessarily.

This is where Tableau steps in. By making data accessible and visible, the issues are made visible for all to see. It's at this point that the business needs to decide whether to tackle the problems head on or go along as they did before.

Sometimes people don't like numbers. Numbers can deliver hard news, for example, something like job losses might be on the cards. People need to be completely sure of the data so they have confidence that they are taking the right decisions. People need to prove to themselves and their managers that the numbers are correct.

The information-seeking mantra defines drilling to detail as a key concept of people interacting with their information. This is particularly important if the numbers show a message that the business consumers do not like.

When looking at the details, it is important to work out what details we want to see. Sometimes, people will want to see the detailed data behind the whole workbook. On other occasions, they will simply want to see the data that lies behind the particular dashboard that is currently on view. In this recipe, we will look at both of these scenarios.

In Tableau, it isn't possible to export all of the data at the dashboard level. Instead, you need to export data at the worksheet level only. Therefore, if you go to a dashboard and try to export data, you would see that the option is grayed out. This makes sense because the dashboard-level data may appear confusing if it is placed in a crosstab format, and it may be hard to relate the columns to their appearance on the dashboard.

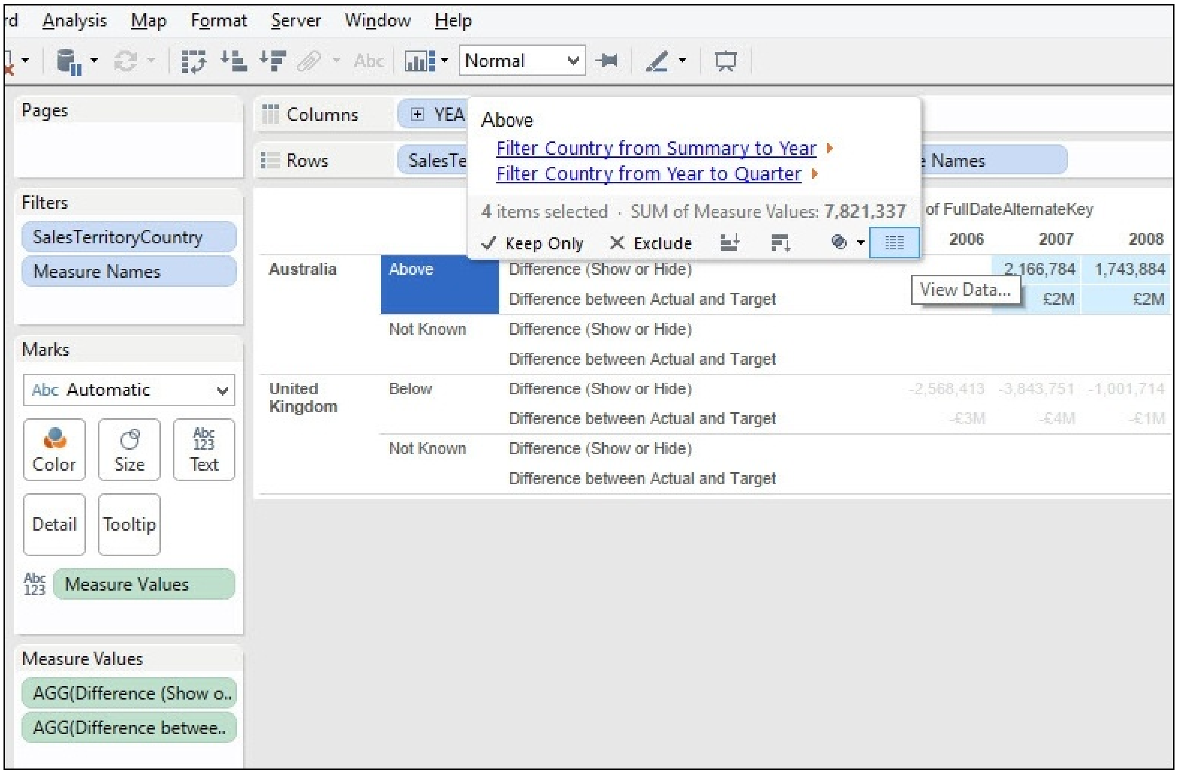
You can select the option **Export the data to CSV**. However, this will only export the data for the specific data points that you see. If you want to export all of the underlying data, then you need to select the **Underlying** tab rather than the **Summary** tab. It is easy to miss this step, and we will call it out in this recipe.

Getting ready

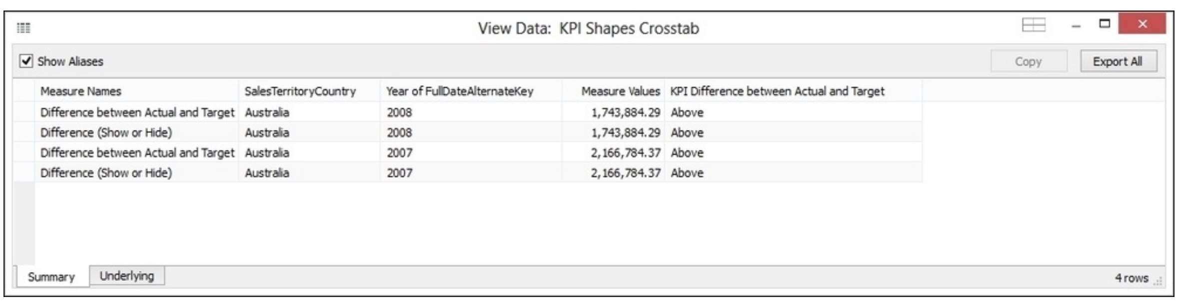
For the exercises in this recipe, continue to work on the `Chapter 6` workbook.

How to do it…

1. To export all of the data, go to the `KPI Shapes` worksheet.
2. Go to the **Worksheet** menu item and then choose the option **Duplicate as Crosstab**. You will then generate a new worksheet that has a neat crosstab which displays all of the data.
3. Rename your new worksheet as `KPI Shapes Crosstab`.
4. Now, we will export just a part of the data rather than the whole set. If you want to simply export the data of one particular data point, right-click on the data point.
5. Click on the **Above** label for Australia and you will see a pop up appear.
6. Click on the **View Data…** button at the bottom right-hand side of the pop-up window that will appear. You can see an example of this in the following screenshot:



7. Now, you will get the following **View Data** dialog box which opens on the **Summary** tab, as shown in the following screenshot:



8. Click on the **Underlying** tab to reveal all of the data. You can see this at the bottom-left of the **View Data** dialog box.

How it works…

In this recipe, we took a dashboard and conducted various exports on the data. We exported all of the data and then looked at taking filtered exports based on the data displayed on the screen. We also looked at obtaining the underlying data which supports the summary that we see on the screen.

When you create a new crosstab worksheet, you could use this crosstab as the basis for further visualization or to export data to other packages such as Excel.

It is also good for checking the data. People get comfort from knowing about the data from the cradle to the grave, particularly if the data is contentious.

Blending data

If a workbook uses data from more than one data source, you can blend data. Blending data is different from joining tables. Blending data means that you combine data from different sources. Tableau makes it very simple to perform this activity.

If you have more than one source of data, you can blend data together in Tableau. Alternatively, you might want to blend the data in a data warehouse or a data store in a single place, outside of the Tableau software.

Sometimes, when we put data together, we get an error message saying **Fields cannot be used from the data source…**. This recipe will explain how to get around this issue. Tableau could not associate the new imported file with the existing file because it could not match any of the column names.

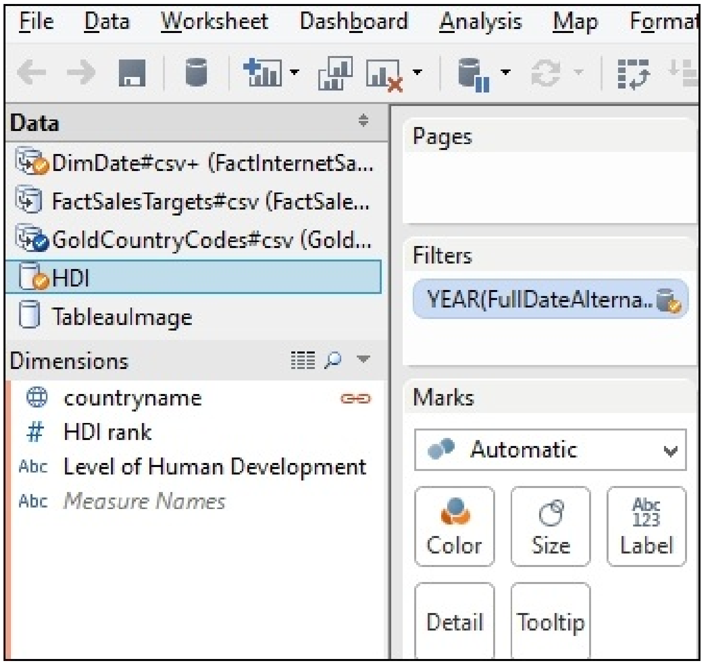
Getting ready

For the exercises in this recipe, let's continue working on the existing `Chapter 6` workbook.

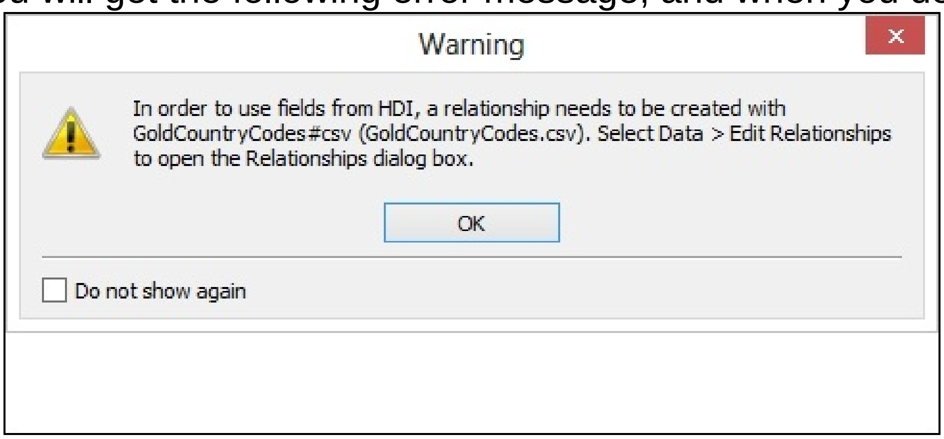
How to do it…

1. Duplicate the `Golden Record` worksheet and rename the copy to `Blending Example`.
2. Next, we will import the `HDI.xlsx` file by going to the **Data** menu item and then choosing the option **Connect to Data…**.
3. In the browser that appears, connect to the file `HDI.csv` and click on **Open**.
4. In the **Connection** dialog box, rename the data source connection to `HDI`.
5. In the **Import** dialog box, let's import all of the data.
6. When all of the data has been imported, you will see the new HDI

connection on the **Data** shelf. Here is an example of this:

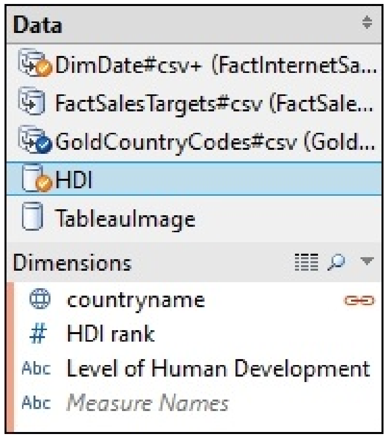


- Drag the **HDI Rank** field over to the **Color** button on the **Marks** shelf.
- You will get the following error message, and when you do, click on



OK:

- We will now work to sort the error message. Click on the **HDI** data source in the **Data** shelf that you set up earlier in this recipe.
- Right-click on the **Country** field and click on **Rename**.
- Rename **Country** to `countryname` and click on **OK**. Note that the case must match exactly as is or Tableau will not be able to match the fields.
- On the right-hand side of `countryname`, you should now see a data blend mark appear; you can see an example of this in the following

screenshot: 

- The field **HDI Rank** will appear as a measure, which means that Tableau will try to aggregate it. Drag the field up to the **Dimensions** pane so that it is used as a way of describing the data, rather than an ordinal piece of data that can be added.
- Now, repeat the step of dragging the **HDI Rank** field over to the **Color** button on the **Marks** shelf.

How it works…

In this recipe, we looked at blending two disparate data sources together, and we saw how easy it is to relate data sources together based on the column name.

Blending data is a key part of creating guided analytics for the user. This feature is useful in solving end-to-end problems since the user can get up to speed quickly, without a detailed understanding of the underlying data sources. Furthermore, it allows users to quickly connect to the data without having to associate the data together. With this, you can assume that the data is clean, of course. It also helps that non-matching column names can be associated with one another simply by renaming one of the columns so that the columns match and are recognized as the relationship between the tables.

One word of warning however: it is easy to become blithe about setting up relationships in the data based on the column name if you are unfamiliar with the sources. It can be possible to have identical column names in different data sources, but the data can mean totally different things and refer to different business processes. It is easy to set up the relationships, but it's also good to sanity check the relationships too.

relationships, but it's also good to sanity check the relationships too. What's technically correct may not be correct from the business perspective.

Optimizing tips for efficient, fast visualization

By having access to a lot of data that you've never seen before, it is easy to get excited and engrossed in loading data and pushing it around. Tableau is a great tech toy, and people enjoy playing with data.

If you are seeing lots of new data for the first time, how can you make sure that Tableau is interacting as quickly as possible with the data? People want to interact with data, and Tableau helps people to explore their data quickly. Nothing, however, puts business users off a new system more than poor response times. People want their data, and they want it now.

Evaluating a system often involves questions such as response time, data-load time, and utilization. However, it can often be difficult to work out the quickest way to access data. The best way is to test, but this recipe will offer some different ways in which you can get a head start in optimizing data.

This recipe isn't intended to suggest that Tableau is slow to access data; in fact, it is extremely fast. In today's world of ever-increasing data sources, it can be hard to work out the best way to access data quickly. It is important to make the right decision at the very start, so this topic is aimed at setting you on the right path.

If you have multiple data connections that are large and take a long time to query, using a join can increase query time dramatically. In this case, it is possible to consider joining the data earlier on in the process before Tableau sees it.

Tableau can help you work out the relationships within the data. For example, by visualizing the data, it can help you to see whether you need to bring data into a single platform or whether the data can live in its existing sources.

This recipe helps you to see a process for making your data as fast as

This recipe helps you to see a process for making your data as fast as possible. We prepare the data at the source. We import the aggregated data into the Tableau data source for speed.

In this example, we preaggregate the data by turning it into an average data. This means that the calculation is already done when it goes into Tableau, and Tableau does not have to spend time making the calculation; it just lifts the data as it sees it.

Importing the whole data into Tableau's own engine is often a useful strategy since Tableau does not have to connect to large data sources, query them, and then bring the data back across the network.

Ultimately, the best way is to provide the best performance for your environment and your data. The method suggested here is one way to give you a head start.

Getting ready

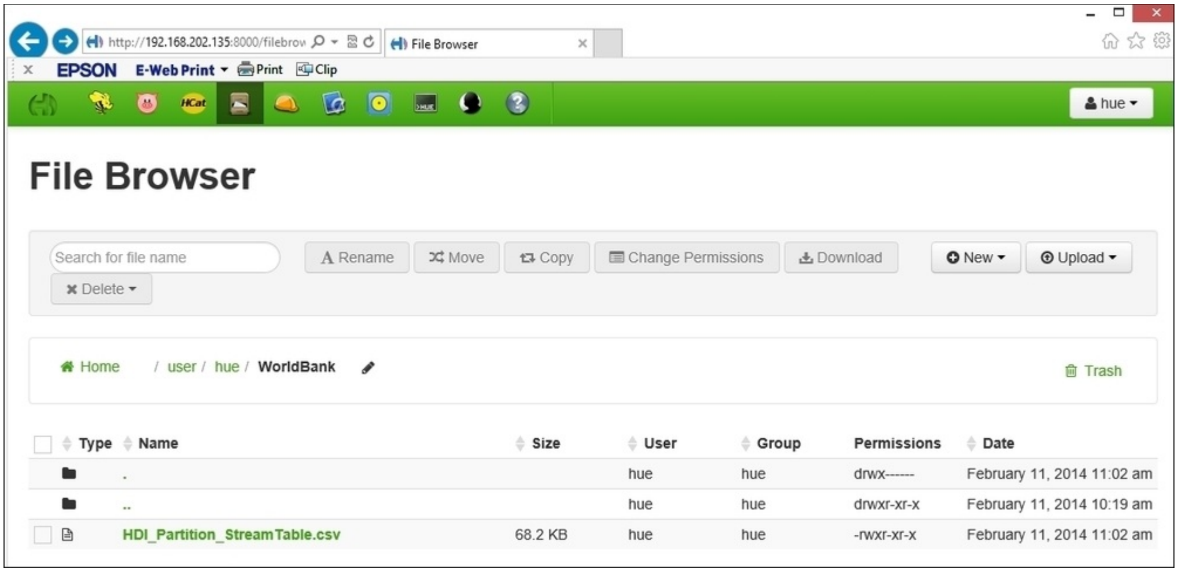
Let's continue with the Tableau workbook `Chapter 6` as is.

As in the previous chapter, we will use the Hortonworks Sandbox to connect to some data and use it as a data source. Alternatively, if you don't have access to this source, you will find details of a location where you can download the sample files so that your work is not impeded.

You can download the data source file and the resulting CSV files from http://sdrv.ms/1aHDtib.

How to do it…

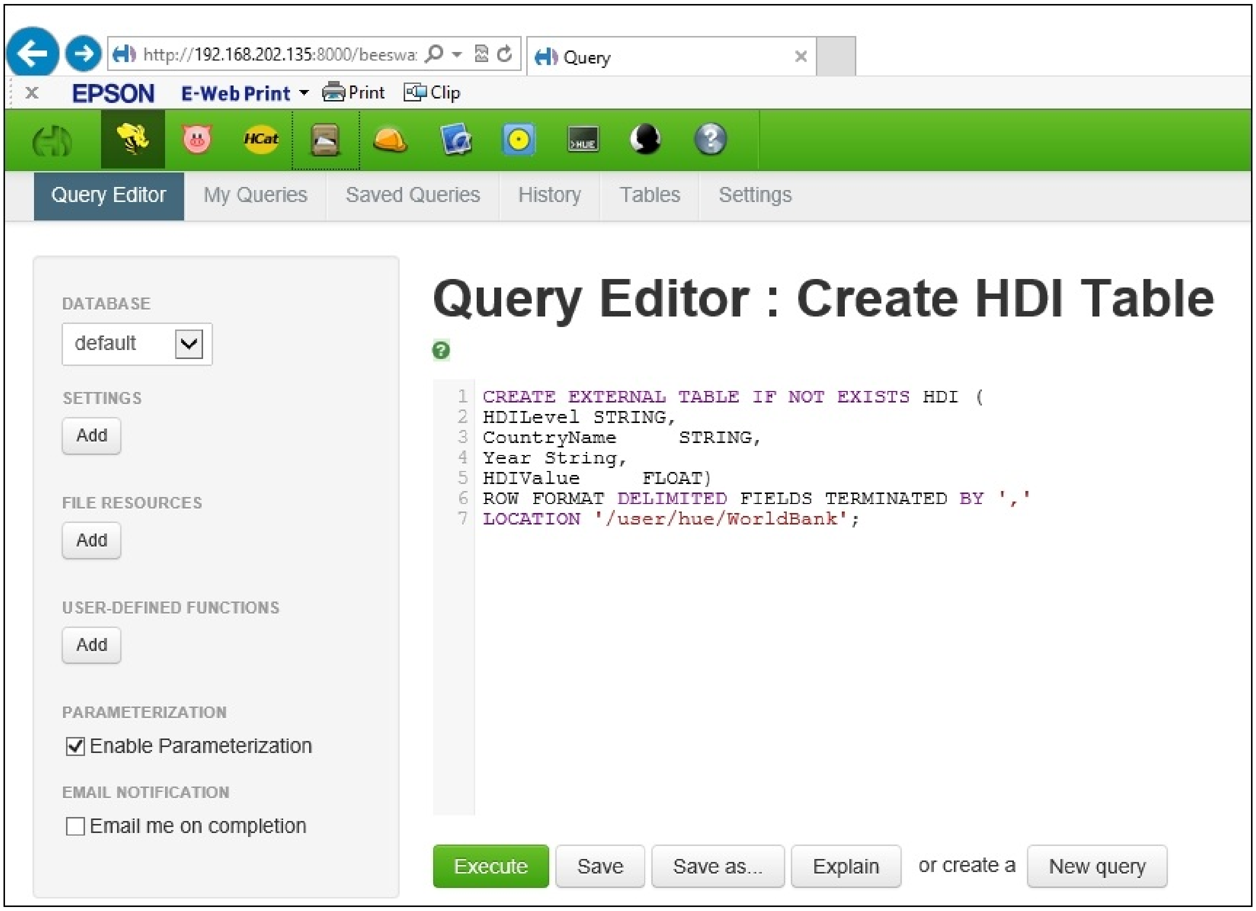
1. Upload the source file to the Hortonworks Sandbox using the File Explorer to a directory called *hue*`WorldBank`. You can see an example of this in the following screenshot:



- Create a table in Hive using the following command:

```
CREATE EXTERNAL TABLE IF NOT EXISTS HDI (
HDILevel      STRING,
CountryName     STRING,
Year              STRING,
HDIValue      FLOAT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION '/userhueWorldBank';
```

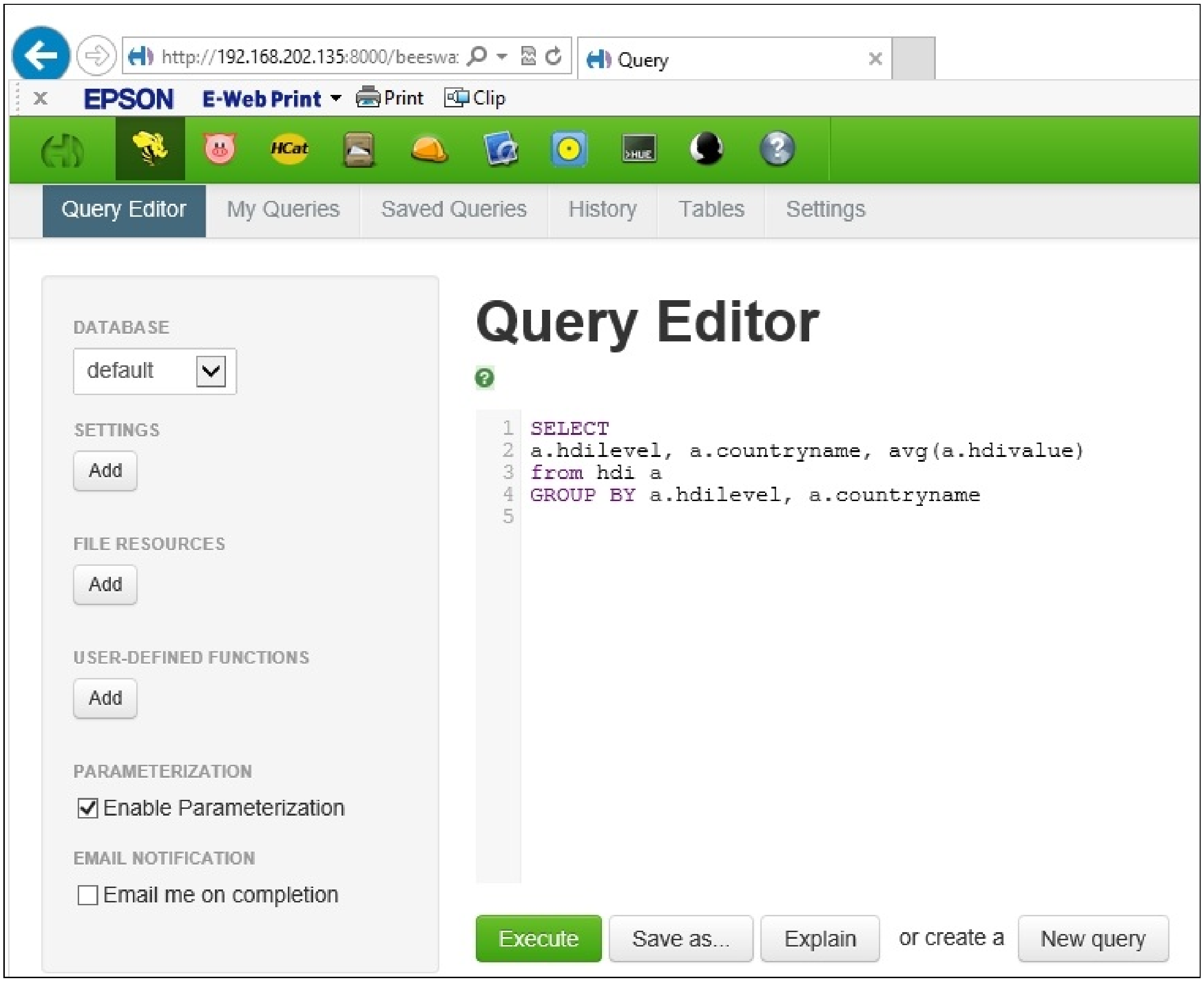
The following screenshot is an example of how the query will look in the Hue web interface:



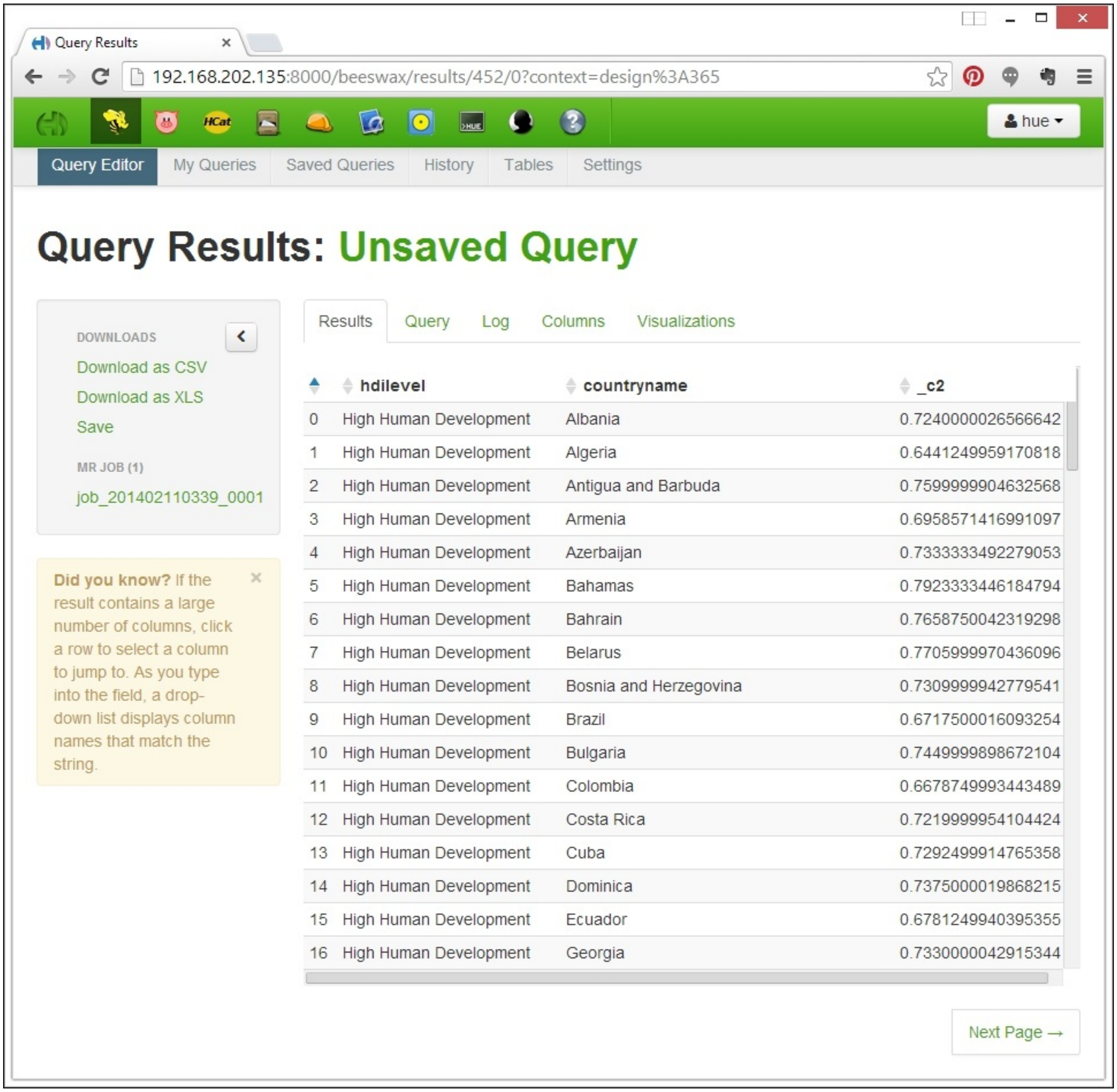
- Run the following query in the Hortonworks Hive query table. It is designed to obtain the average HDI value, which is based on the HDI level and the country.

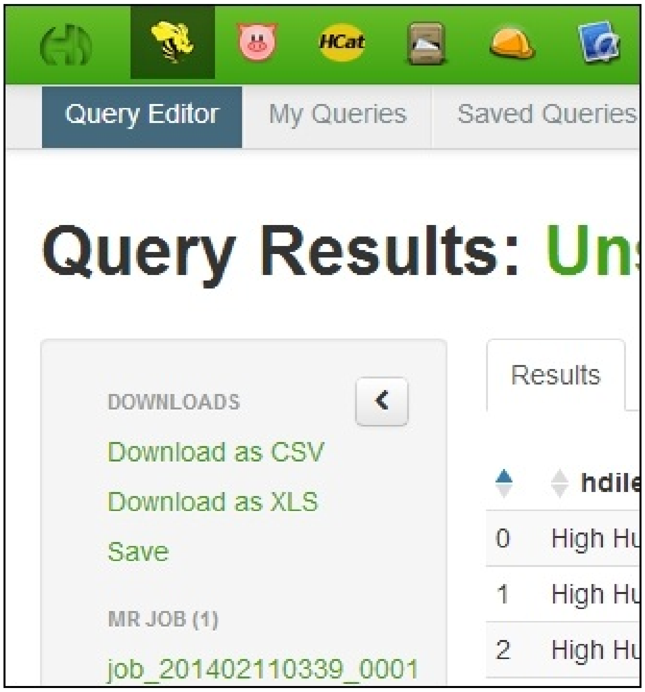
```
SELECT
a.hdilevel, a.countryname, avg(a.hdivalue)
from hdi a
GROUP BY a.hdilevel, a.countryname
```

The Hue screen will appear as follows:

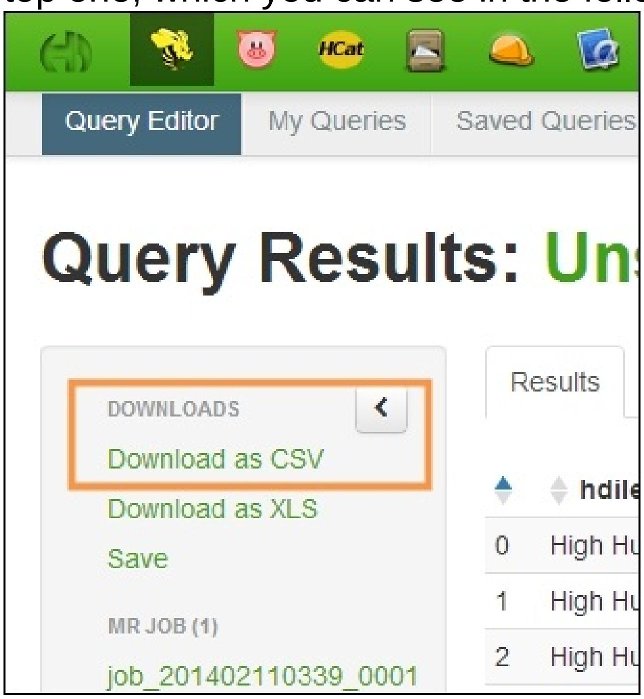


- Once you have created the table, make sure that the data has been loaded by going to the **Results** view of the query, as shown in the following screenshot:

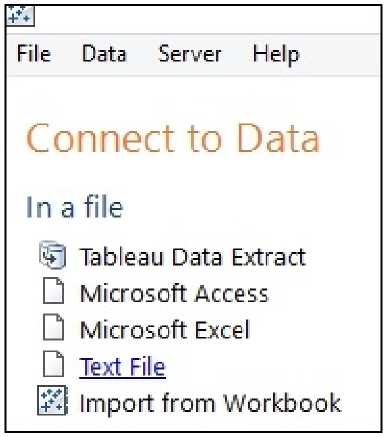


- Once you have done this, you can save the file by going to **Download as CSV** on the right-hand side. Here is an example of this in the following screenshot: 

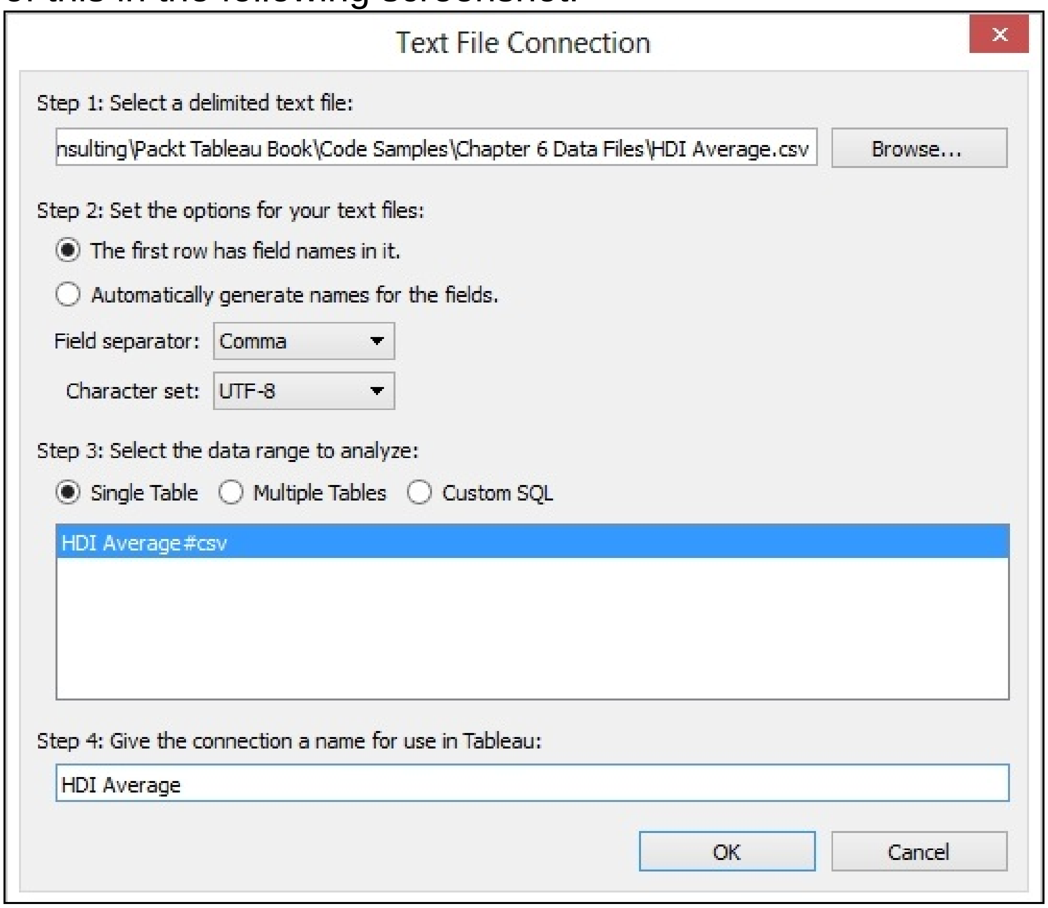
- In case it isn't clear, there are a number of options, and you need the top one, which you can see in the following screenshot:



- Alternatively, download the file from http://sdrv.ms/1aHDtib.
- The CSV file will be downloaded with the name `query_results.csv`. Rename it to `HDI Average`.
- We will connect to the text file in Tableau. Open Tableau and create a new worksheet using the option **Connect to Data**, as shown in the

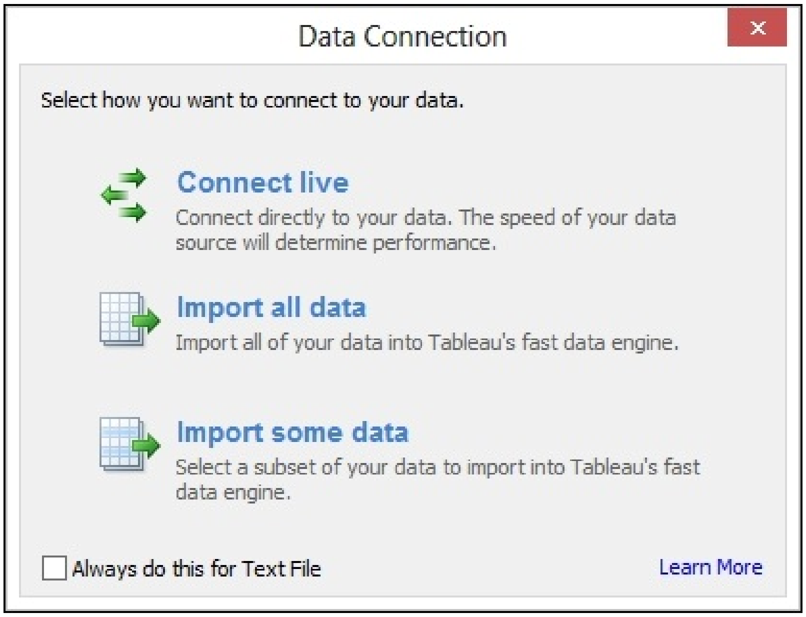
following screenshot: 

- Next, rename the connection to `HDI Average`. You can see an example of this in the following screenshot:

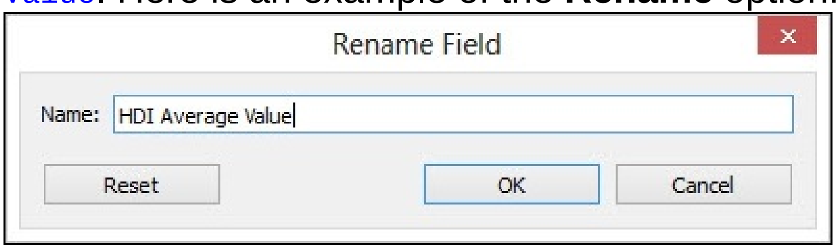


- Click on **OK** and return to the Tableau worksheet.
- Let's import all of the data into Tableau's internal data store. So, in the **Data Connection** tab, select the option **Import all data**, as shown in the

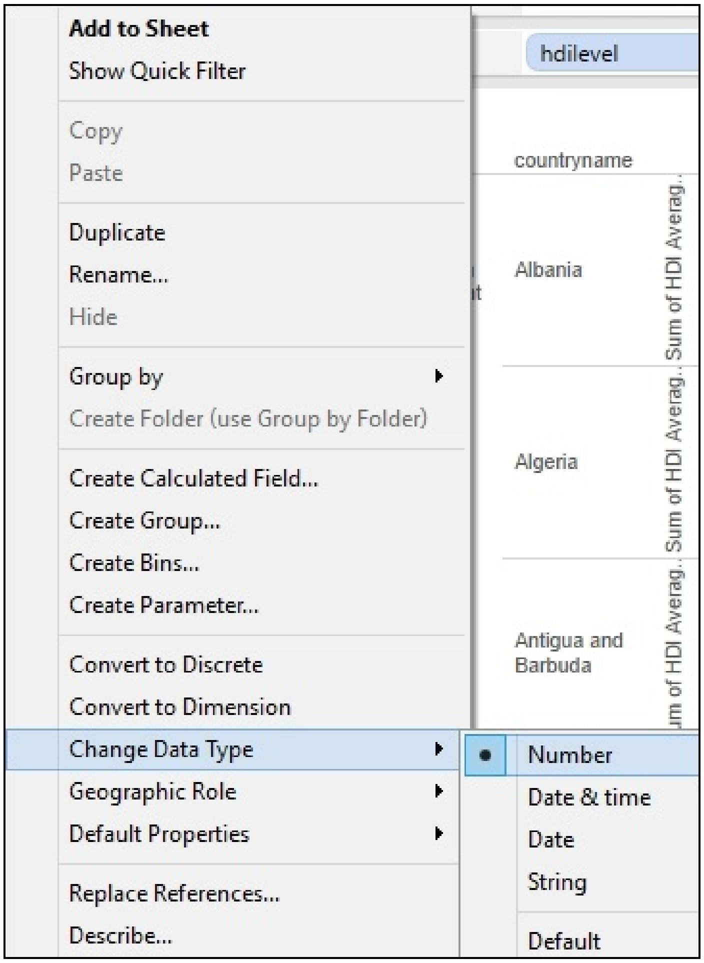
following screenshot:

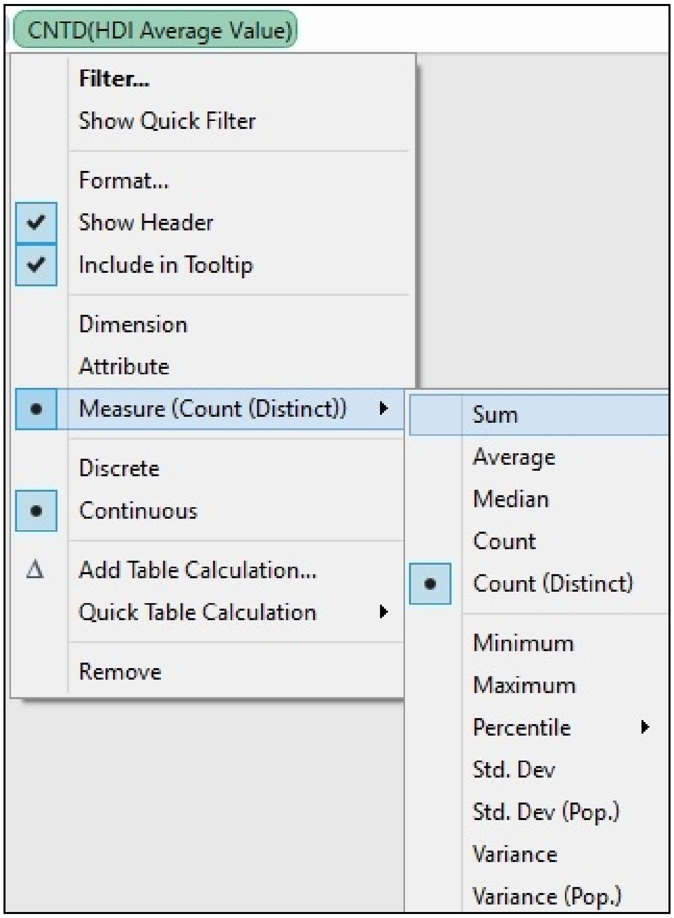


- Tableau will then ask you where to store the data abstract as a TDE file; select a file location on your laptop that suits you best.
- When the data is imported, you will see the fields on the left-hand side. The value field may be transported with a default column name. For example, it may read **c2**, which is simply Tableau's placeholder name for the column during the transport. It is shorthand for column 2. In this case, right-click on it and select the **Rename** option. Rename it to `HDI Average Value`. Here is an example of the **Rename** option:

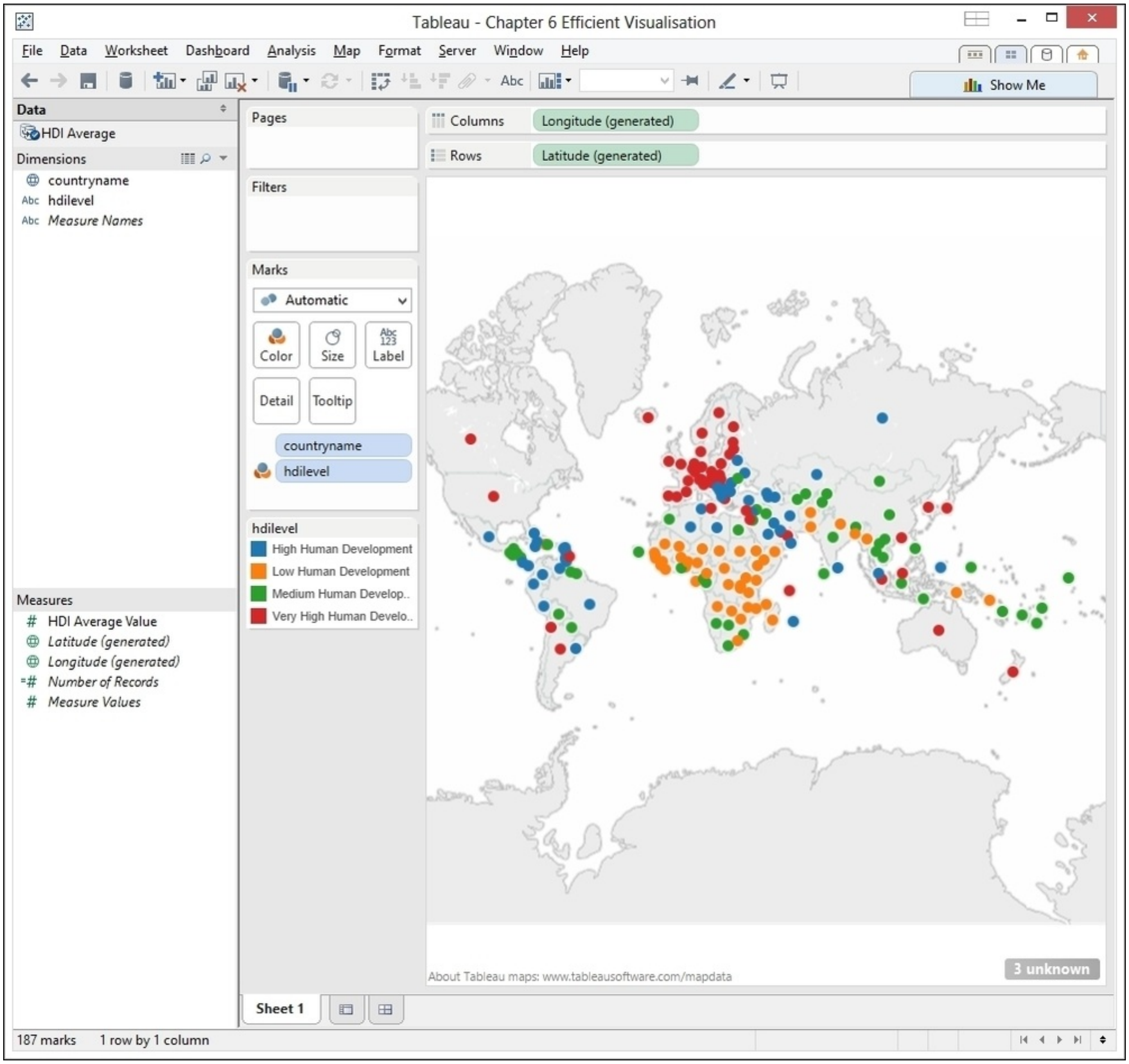


- `HDI Average Value` will appear as a string, so we will need to make it a measure. To do this, drag it from the **Dimensions** shelf down to the **Measures** shelf.
- The field will still be in the string format, so we will need to change it to a decimal format. To do this, right-click on the field and select **Change Data Type**, and change the type to a number. Here is an example:



- Now, when you drag **HDI Average Value** over to the **Rows** column, you will notice that it is a count measure rather than an average measure. Tableau does this because the measure was originally a dimension that could be counted rather than summed. To change the metric from **COUNT** to **AVG**, right-click on it in the **Rows** shelf and select **Measure(Count(Distinct))**, and then select **AVG**. Here is an example of this setting: 

- Drag **countryname** on to the **Rows** shelf so it is on the left-hand side of **HDI Rank**.
- Select a symbol map from the **Show Me** panel. You should have **Longitude** in the **Columns** shelf and **Latitude** in the **Rows** shelf.
- Drag the **hdilevel** column on to the **Color** button. This will use **Color** to categorize the countries in terms of their development.
- Your dashboard should now look like the following screenshot:



- Our data is now visualized very simply using a mix of data that was aggregated at the source, and then importing the data into Tableau's memory engine.

How it works…

Tableau can connect to live data or data that is held in memory, or both. The Tableau data engine uses different levels of memory at different times, so people can explore their data more quickly. It also means that the business users are not touching the underlying data source, which means that pressure can also be relieved from the system.

A nice thing about the Tableau data engine is that we can combine data

A nice thing about the Tableau data engine is that we can combine data that has been held in different formats, such as SQL Server, Excel, or Teradata, and combine it together into one source. Regardless of the source of the software solution, users can connect to the data, consolidate it, and then analyze and visualize the data.

It isn't always easy to work out when to use in memory and when to use the source. Live connections are better if you have fast-moving data and a fast database since the data is changing all the time and you need a fast connection.

In memory is better where the underlying data source is slow or has a lot of user and operational pressure on it. Also, taking the data into Tableau means that you can access the data offline; for example, if you are working on a train or somewhere where you don't have good Internet access to connect to your company's data sources.

To summarize, it is good that Tableau gives us several options; we have looked at one method of connecting to data here, which combines the features of the source system as well as the fast features of Tableau.

Chapter 7. Visual Best Practices

In this chapter, we will cover:

- Coloring your numbers
- Dueling with dual axes
- Where is the three dimensional data?
- Eating humble pie – pie charts or not?
- Sizing to make a data story

Introduction

Why is data visualization so important, and how can we do it well? Data visualization is often the initial pain point in a project. People don't have their reports and data visualizations, and they simply want more of them. In order to build a business strategy, leaders and decision makers need to understand what they want to achieve, and they also need to understand the existing terrain of the organization.

Businesses require that operational reporting solutions deliver results that can be predicted, operationally efficient, and robust, while delivering corporate accountability and transparency. This makes operational reporting more important. If the stakeholder needs are not fulfilled, then they will simply resort to more home-grown solutions rather than insightful long-term decision-making tools.

Business intelligence can enhance and extend an enterprise by supporting its decision-making ability. It can have a direct impact on the overall performance of the organization by promoting a cycle of continuous innovation, along with better decision making. This is more important in today's fast-paced and demanding environment, particularly given the amount of data that we produce every day. By understanding the data better using visual best practices, we are giving ourselves the opportunity to make better decisions. This is particularly important in today's Big Data world.

This chapter will help you to see some of the theories and best practices that underpin visual design and display in a dashboard. Why is this important? You will want to share your dashboards with team members or perhaps with senior management in your organization. Even though every visualization is different, there are common themes that will help you create your dashboards in such a way that you are more likely to get your message across to the right people in the right way.

John Stuart Mill, the English philosopher, once held the utilitarian principle that the right course of action is the one that maximizes utility. This usually translates as the one that maximizes benefit or the one that makes most people happy. In data visualization, ultimately the goal is to build a visualization that suits your audience rather than building something that is best for you, the dashboard creator.

This is only a brief overview, and it's recommended that you follow up with the references provided in each section.

Coloring your numbers

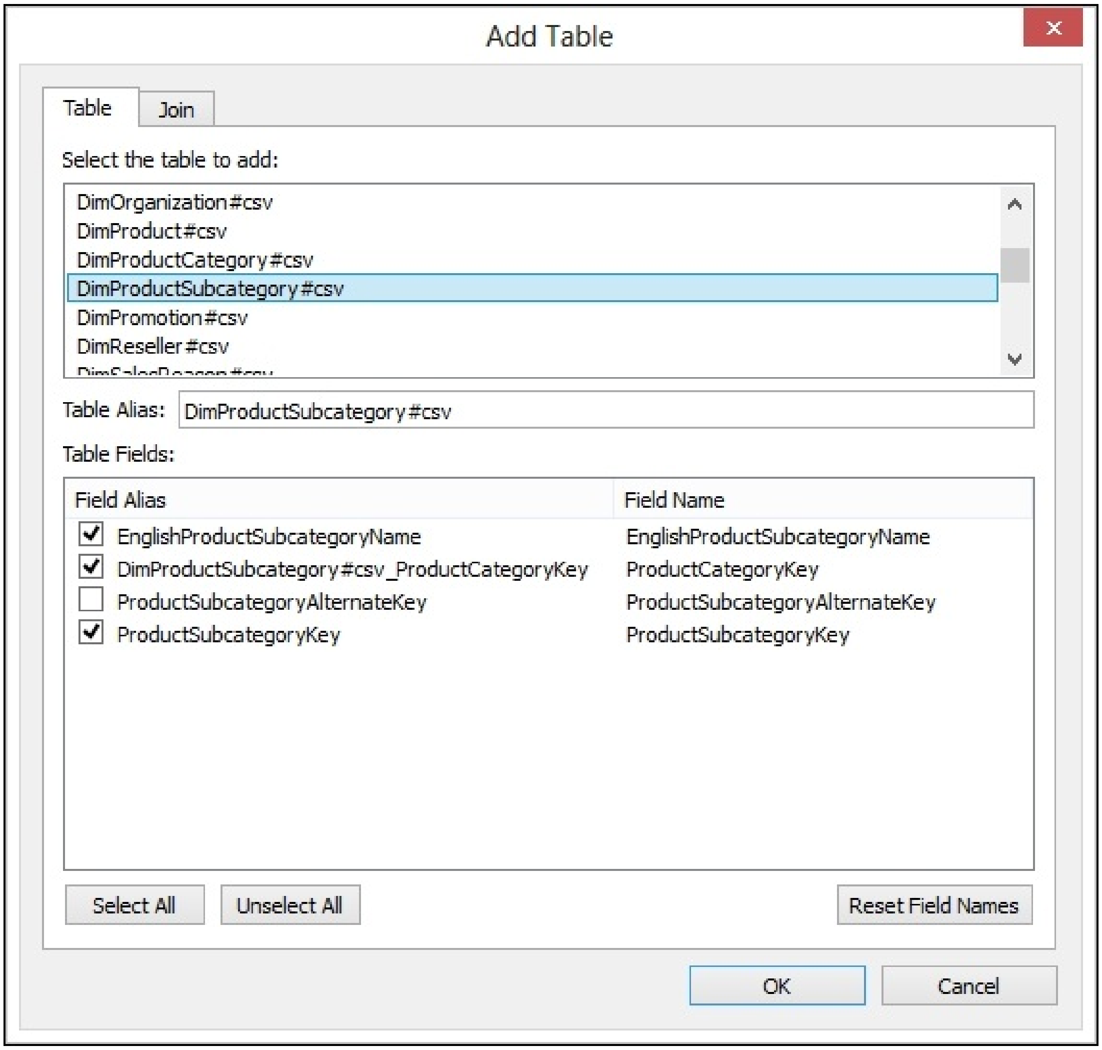
In this recipe, we will look at the use of color to convey a message. Since we are looking at dashboarding, we need to know how to use color effectively to make the most of a small space. Here, we will use a box and whisker plot to convey a lot of information about the data in a small space, along with additional information on the figures themselves using color.

Getting ready

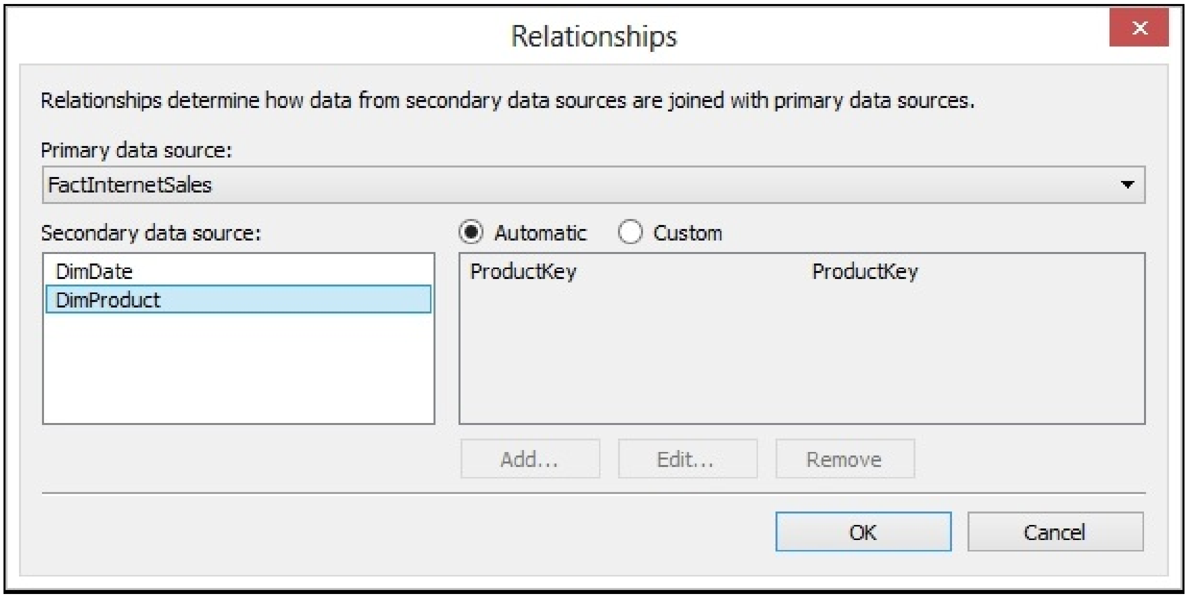
For the exercises in this recipe, let's start with a fresh Tableau workbook. There are no other requirements for this recipe.

How to do it…

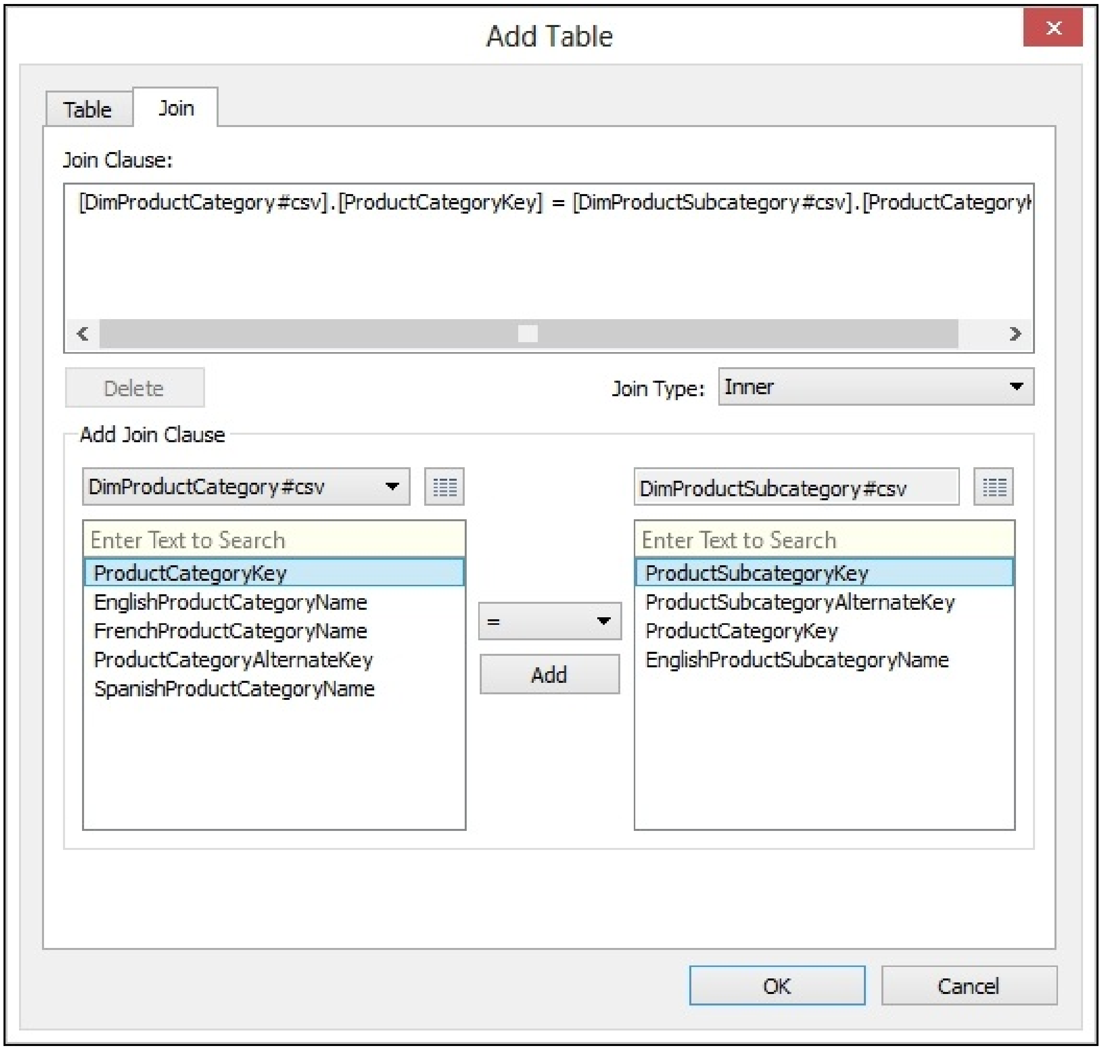
1. Go to Tableau and navigate to **File | New**.
2. Select the **Connect to Data** link at the top-left corner of the screen.
3. Navigate to the `DimProductCategory.csv` file, which is located in the folder where you downloaded the code samples, and click on the **Open** button to import it into the Tableau workbook.
4. When the **Text File Connection** dialog box appears, select the option **Multiple Tables**.
5. Next, click on **Add Table** to import the file `DimSubProductCategory.csv`. The dialog box shown in the following screenshot appears:



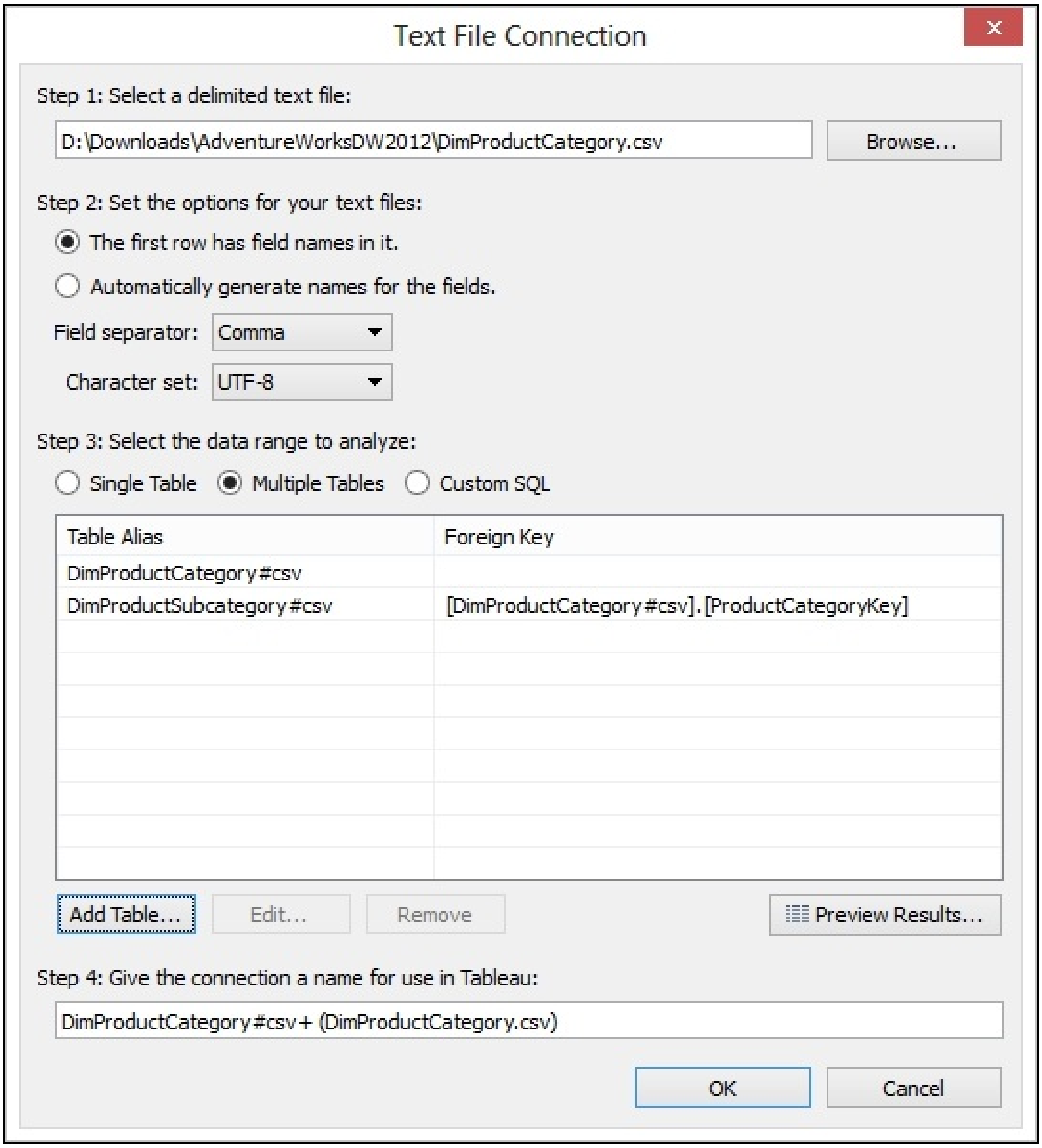
- Uncheck the column **DimProductSubcategory #csv_ProductCategoryKey**, as you can see in the preceding screenshot. Click on **OK**.
- When the **Data Connection** dialog box appears, select the option **Import all data**.



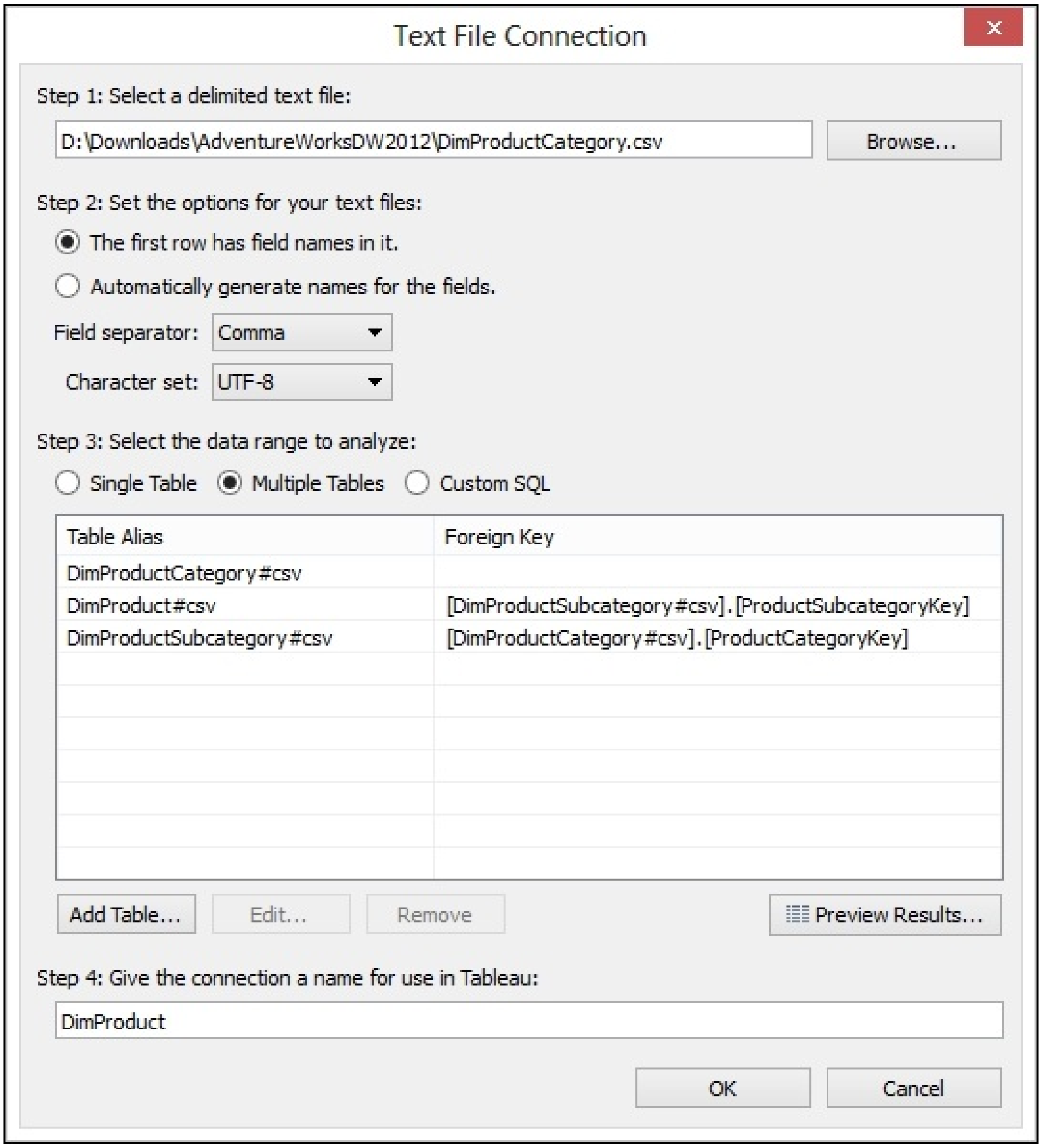
- Now we will join the data together. Click on the **Join** button and you will see that the column **ProductCategoryKey** is the join for each table. Click on **OK**. Here is an example of the join clause in the following screenshot:



- The **Text File Connection** dialog box will now look like the following screenshot:

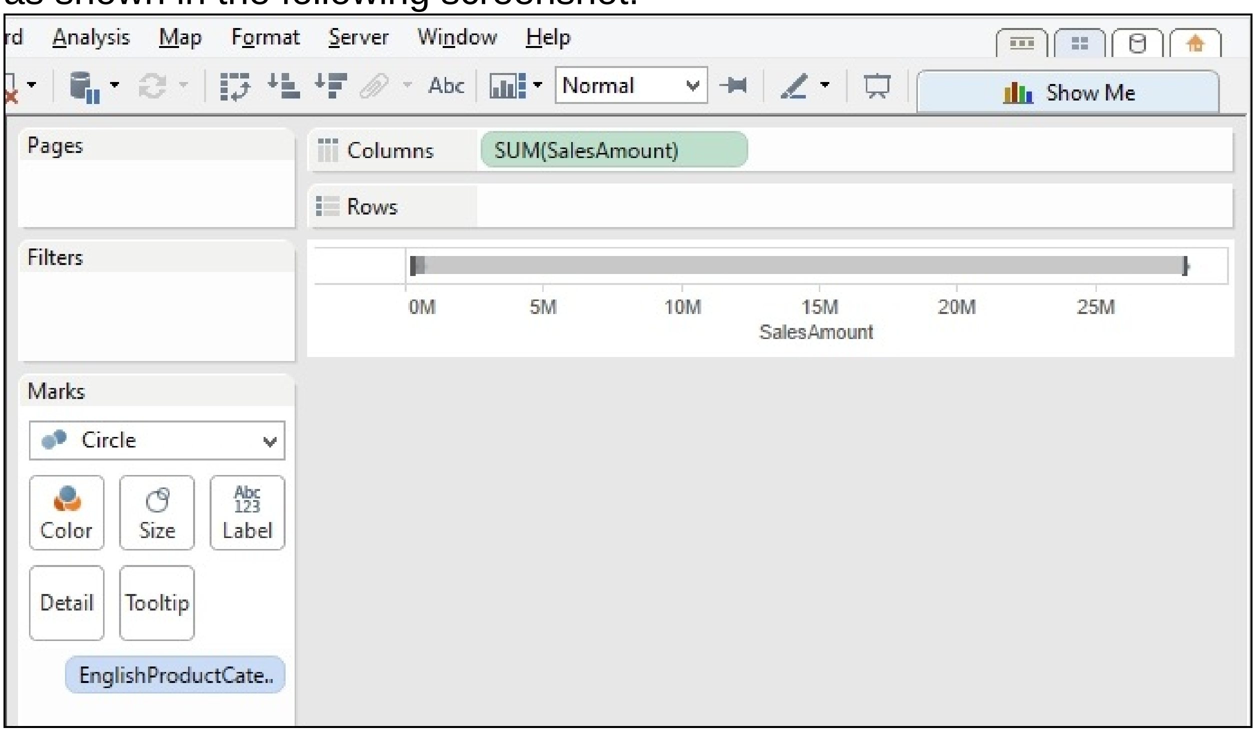


- Now, we will add the `DimProduct` file by clicking on the **Add Table…** button again and click on it once more for the `DimProduct#csv` table.
- Click on the **Unselect All** button. Tick only the following columns: **EnglishProductName**, **ProductKey**, and **Color**, and then click on **Join**. The join clause will show that there is a join between the **ProductSubcategoryKey** column for each table, and this is correct. Click on **OK**.
- The **Text File Connection** dialog box will appear as shown in the following screenshot. Rename the connection to `DimProduct` and click on **Add Table…**.

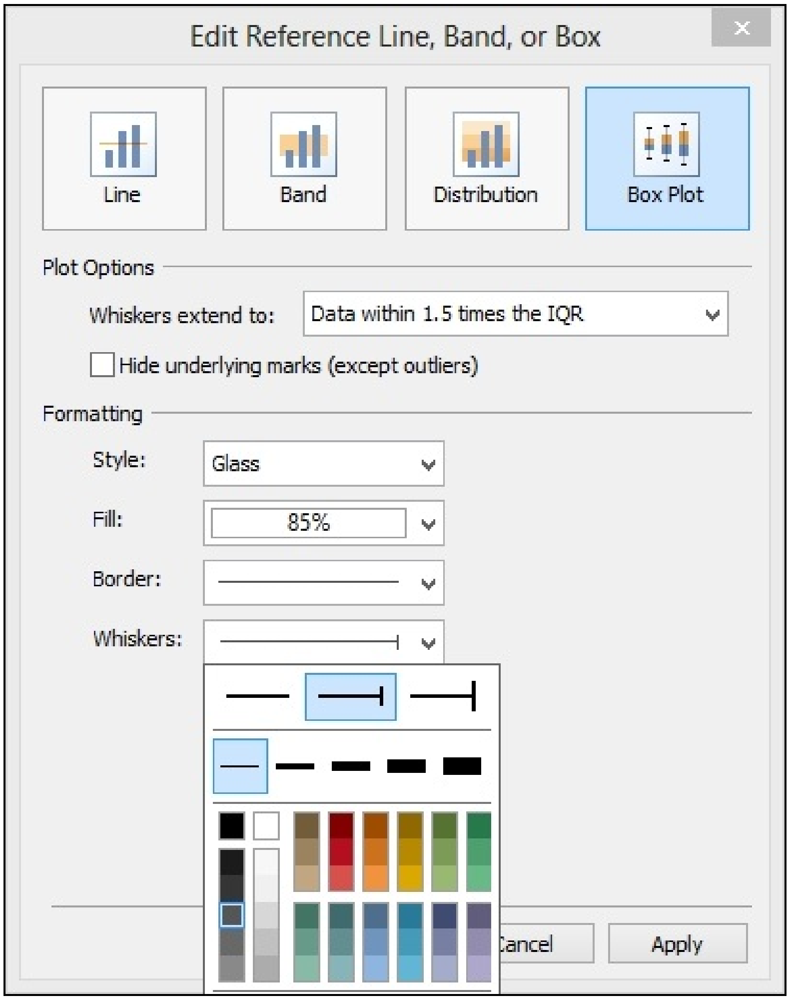


- Now, we will add in the **FactInternetSales** table by selecting it from the list of available tables in the **Text File Connection** dialog box.
- Choose the option **Unselect All** and select only the following fields: **Sales Amount**, **Order Quantity**, **ProductKey**, and **OrderDate**, and then click on **Join**.
- In the join clause on the left-hand side, make sure that **DimProductCategory#csv** is selected from the drop-down list.
- Select the **ProductKey** column for the `DimProduct` table from the left-hand side of the join clause.
- Make sure that **FactInternetSales#csv** is on the right-hand side of the join clause in the drop-down list.

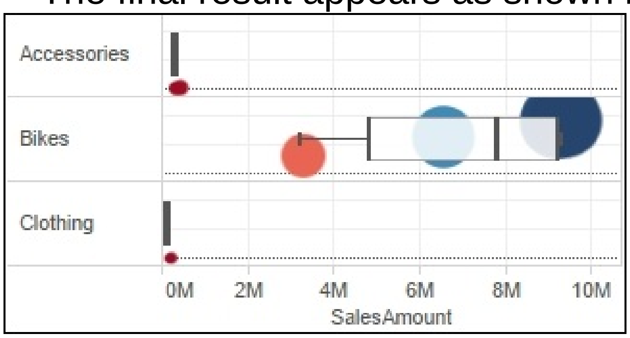
- On the right-hand side, select **ProductKey** from the **FactInternetSales** table and set the join type to **Inner**. Click on **OK**.
- Next, let's add the table **DimDate#CSV** by selecting it from the list of tables in the **Text File Connection** dialog box.
- In the join clause, select **FullDateAlternateKey** for the `DimDate` table and **OrderDate** for the `FactInternetSales` table.
- In the **Text File Connection** dialog box, click on the **DimDate#CSV** table and then click on **Unselect All**.
- Select only the following fields: **Max year**, **FullDateAlternateKey**, **FiscalYear**, **FiscalSemester**, and **FiscalQuarter**.
- Click on the **Join** button to connect `FullAlternateDateKey` from `DimDate#CSV` and the `FactInternetSales#csv` column `OrderDate`.
- In the **Data Connection** dialog box, click on **Import all data**. You should now see the Tableau worksheet with your data source on the left-hand side. Rename the data source to `Chapter Seven`.
- Drag **SUM(SalesAmount)** onto the **Columns** shelf.
- Drag **EnglishProductCategory** onto the **Rows** shelf.
- Select the **box-and-whisker** plot from the **Show Me** panel.
- You'll see that the selected dimensions and measures change. Click on the **Swap** button after navigating to **Map | Format**. The screen will look as shown in the following screenshot:



- Drag **EnglishProductCategory** back onto **Rows**.
- Drag **Year(FullDateAlternateKey)** onto the **Marks** shelf.
- Filter **Year(FullDateAlternateKey)** so that only the year **2008** is selected. Drag **Year(FullDateAlternateKey)** onto the **Filters** shelf. In the pop-up dialog box that appears, select **Years**.
- In the **Filter[Year of FullDateAlternateKey]** editor box, deselect **2005**, **2006**, and **2007**. Leave only **2008** checked and click on **OK**.
- Drag **SUM(SalesAmount)** onto the **Rows** shelf.
- Right-click on the **SalesAmount** axis and deselect **Show Header**.
- Select **Circle** from the drop-down list on the **Marks** shelf.
- Drag **SUM(SalesAmount)** onto the **Size** button.
- Click on the **Size** button and move the slider so that it is half way between the start and the end of the slider. This will increase the size of the circles for **SalesAmount**.
- Drag **SUM(SalesAmount)** onto the **Color** button.
- Click on the **Color** button and select **Edit Colors…**.
- From the drop-down list, select **Red-Blue Diverging** and click on **Apply**, and then click on **OK**.
- Reduce the size of the rows by clicking down and pressing on one of the **Rows** lines and dragging it upwards.
- Right-click on the **box-and-whisker** plot and select **Edit…**.
- In the **Edit Reference Line, Band or Box** dialog box, go to the **Formatting** section and choose **Glass** for the **Style** setting.
- Set **Fill** to be white from the drop-down list.
- For the border, select the thinnest border from the available selection.
- For **Whiskers**, select the middle option. You can see an example of the settings in the following screenshot:



- The final result appears as shown in the following screenshot:



How it works…

Ever played with a Rubik's cube? Color is a vital way of understanding and categorizing what we see. We can't order colors in terms of low to high value, for example, red plus yellow gives blue, since people experience colors differently. However, we can use color to tell a story about the data. We can use color to categorize, order, and display quantity.

In this recipe, we chose color to highlight some elements over others, and we also used it to convey a message. Red was used to denote smaller values, and blue was used to denote higher values. Red is often seen as a *warning* color in the West. We reduced the color intensity in the **box-and-whisker** plot so that the circles could be seen through them. This allows us to add visualizations on top of one another but not occlude one another. The users can click on the **box-and-whisker** plot to get more detail about the data.

Data visualization is about displaying high-dimensional data onto a low-dimensional canvas. Color can help us to distinguish between the dimensions that you want to display. Bright colors pop at us, and light colors recede into the background. We can use color to focus attention on the most relevant parts of the data visualization. This is very important when we are dealing with Big Data sources. We tend to spot things that stand out.

In Tableau, we can see that there are a number of ways in which we can choose colors. Further, more we know that a percentage of the population is color blind, so their color perception is reduced. We can choose colors that feel natural, thereby bringing the dashboard closer to the viewer, and they can understand it better. Fortunately, Tableau often helps you to choose the right type of color for the data.

Color choice depends on the numbers that you are trying to represent. If you are looking at ordering data, you can choose a sequential palette. This is where you choose one color to reflect the metric, but the intensity, brightness, or darkness of the color increases as the value increases. You may want to use a sequential palette to represent age, for example, where lighter values represent younger age groups and darker colors represent older age groups.

Alternatively, if you are looking at distinguishing metrics, you could use a diverging palette. For example, the diverging palette could diverge from red right through the spectrum to white and then on to blue. This palette could be used to represent profit and loss, for example, white could represent zero or thereabouts, red could indicate a loss, and blue could indicate profit.

If you are looking at categorizing data, you could use different colors to represent different dimensional attributes. For example, you could use a different color to represent a different country or a different product group.

Picking color isn't easy. We can't say precisely that this color of blue is twice as blue than another shade of blue. However, Tableau does give you a helping hand.

See also

If you want to know more about color choice and theory, there are plenty of resources. Here are some good places to start:

- *Show Me the Numbers*, *Stephen Few*, *Analytics Press* (2012)
- *Now You See It*, *Stephen Few*, *Analytics Press* (2009)

Dueling with dual axes

Charts with dual axes can be a mixed blessing. Adding an additional axis can help with comparison purposes. Comparison is one of the essential tools of analyzing data. You can often hear it expressed in user questions, such as how does that figure compare to last year's? Or where are we in relation to our target?

On the other hand, dual axes are best used where the viewer really understands the data. They can be very misleading. For example, if we have units on one axis and currency on another, the chart can be hard to understand. Further, more if the axes are contracted whereby they don't start at zero, or only show a band of the data, then the naïve user may find it misleading. Normally, due to these issues, dual axes charts are best avoided where people don't understand the data very well. This is particularly the case for a dashboard, where people are expecting to pick up information very quickly.

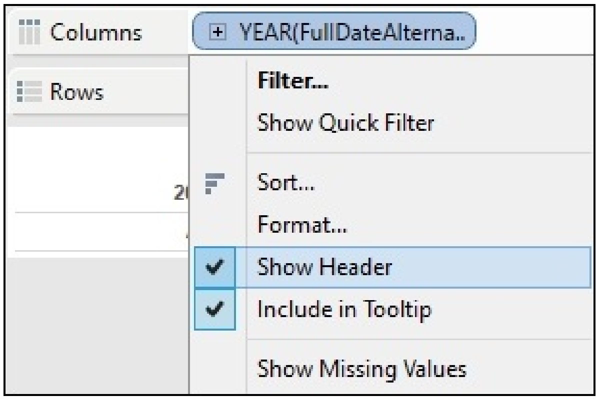
In Tableau, however, the use of dual axes can be very useful to display the same data in different ways in order to enhance the message of the data. In this recipe, we will look at using dual axes charts as another neat trick for visualizing data.

Getting ready

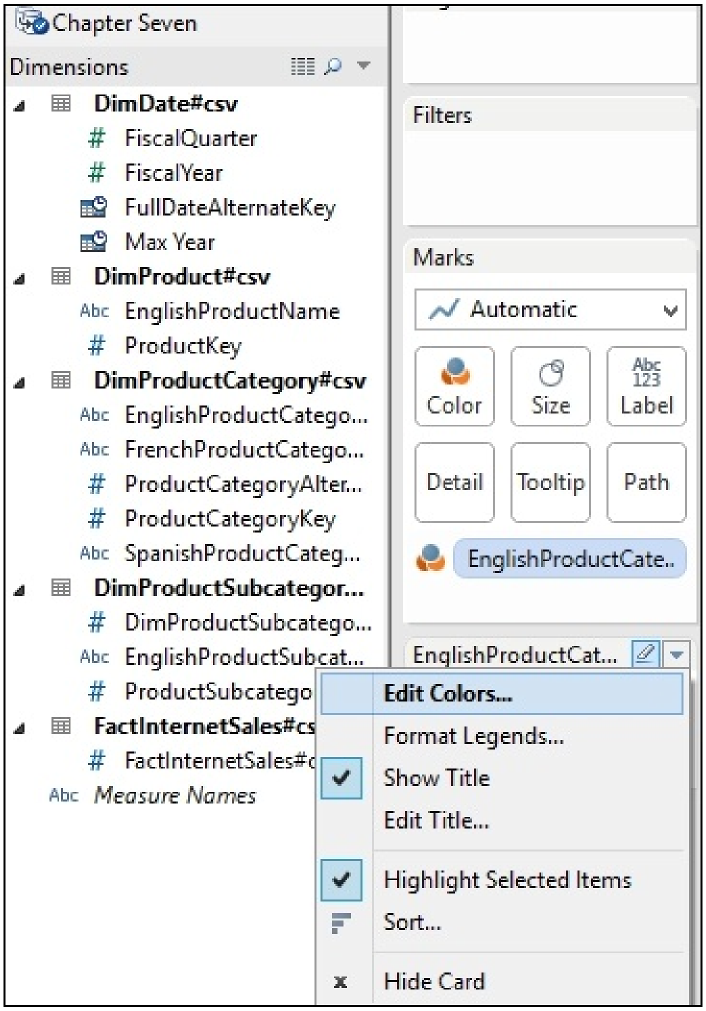
For the exercises in this recipe, we will build on the existing `Chapter 6` dashboard. We don't need to add in any more data for now.

How to do it…

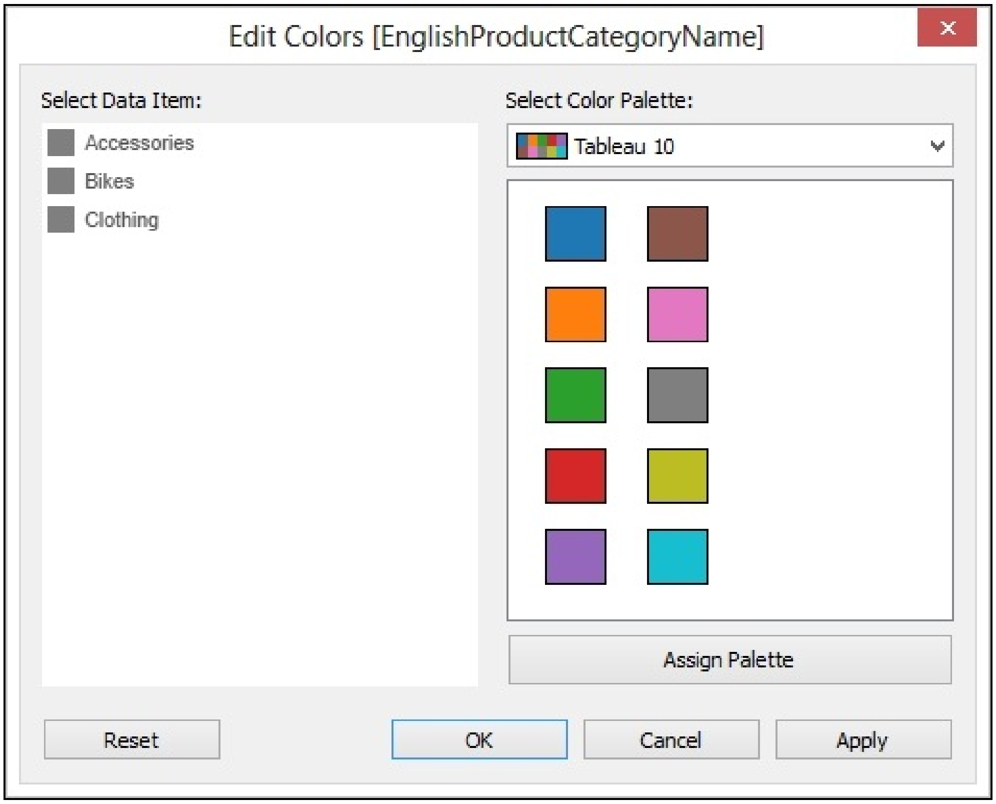
1. Drag **FullAlternateDateKey** onto the **Columns** shelf.
2. Remove the header by right-clicking on the blue pill and deselecting **Show Header**. You can see an example in the following screenshot:



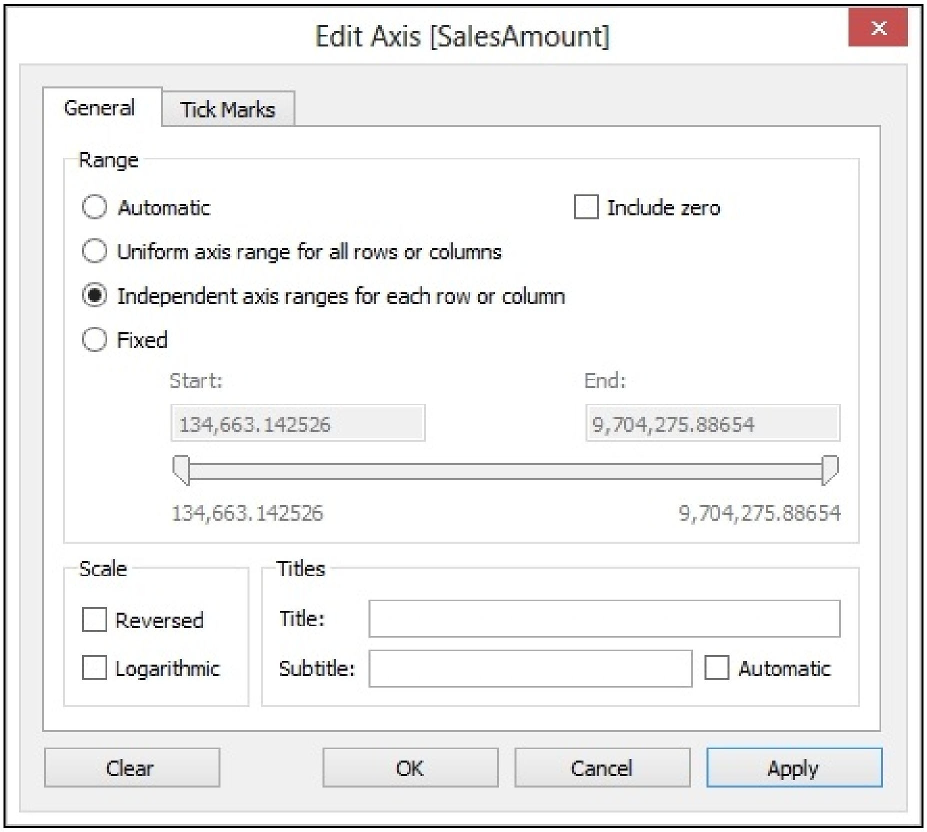
- Drag **EnglishProductCategoryName** onto the **Rows** shelf.
- Drag **SalesAmount** onto the right-hand side of **EnglishProductCategoryName** on the **Rows** shelf. You should get line charts now.
- To change the lines to make them gray for all the product categories, drag **EnglishProductCategoryName** onto the **Color** button and click on the right-hand side downward-facing arrow and select the **Edit Colors…** option. Here is an example in the following screenshot:



- Click on each category in turn and select the color to be gray. You can see an example in the following screenshot:



- You can see that **Bikes** has a much higher sales amount than **Accessories** or **Clothing**. The sales amount value for **Bikes** is a behemoth next to the other categories, which unfortunately means that we cannot see the patterns in the data for these categories. To solve this problem, we need to change the axes so that they are synchronized and we can see the patterns. Right-click on the **SalesAmount** *y* axis and choose the option **Edit Axis**.
- In the **Edit Axis** dialog box that appears, deselect the **Include Zero** checkbox and select the option **Independent axis ranges for each row or column**. Then, remove **SalesAmount** from the **Title** textbox and deselect **Automatic**. Then, we click on **Apply** and **OK**, as shown in the following screenshot:

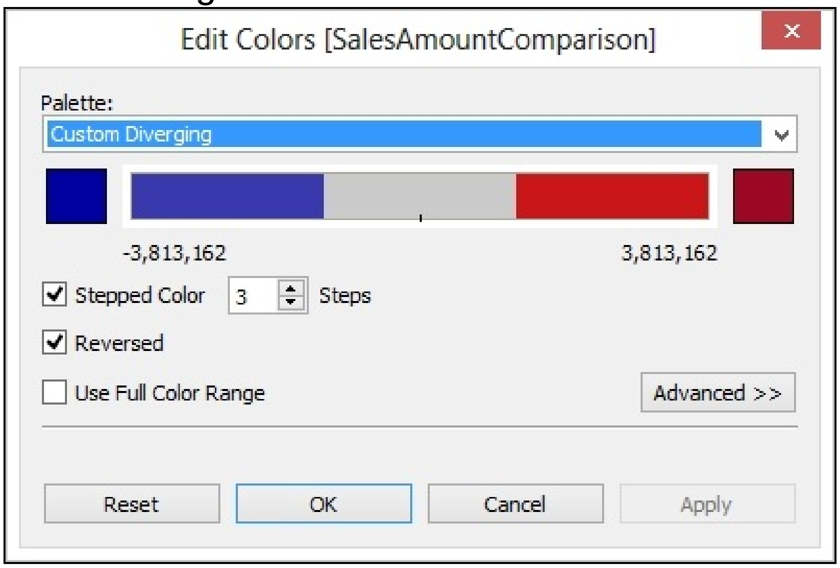


- Next, click on the **SalesAmount** green pill and uncheck the option **Show Header**.
- Finally, go to the **Columns** shelf, click on **Year(FullSalesAmount)**, and select **Discrete**.
- Right-click on **EnglishProductCategoryName** and deselect **Show Header**.
- Right-click on **SalesAmount** and select **Create Calculated Field**.
- We will create a calculated field that calculates whether the latest sales amount is greater than the average amount of sales for each row. In the **Name** field of the **Calculated Field** editor, type `SalesAmountComparison`.
- In the **Text** field of the **Calculated Field** editor, type the following formula and click on **OK**: *ZN(SUM([SalesAmount])) - Window_AVG(SUM([SalesAmount]))*

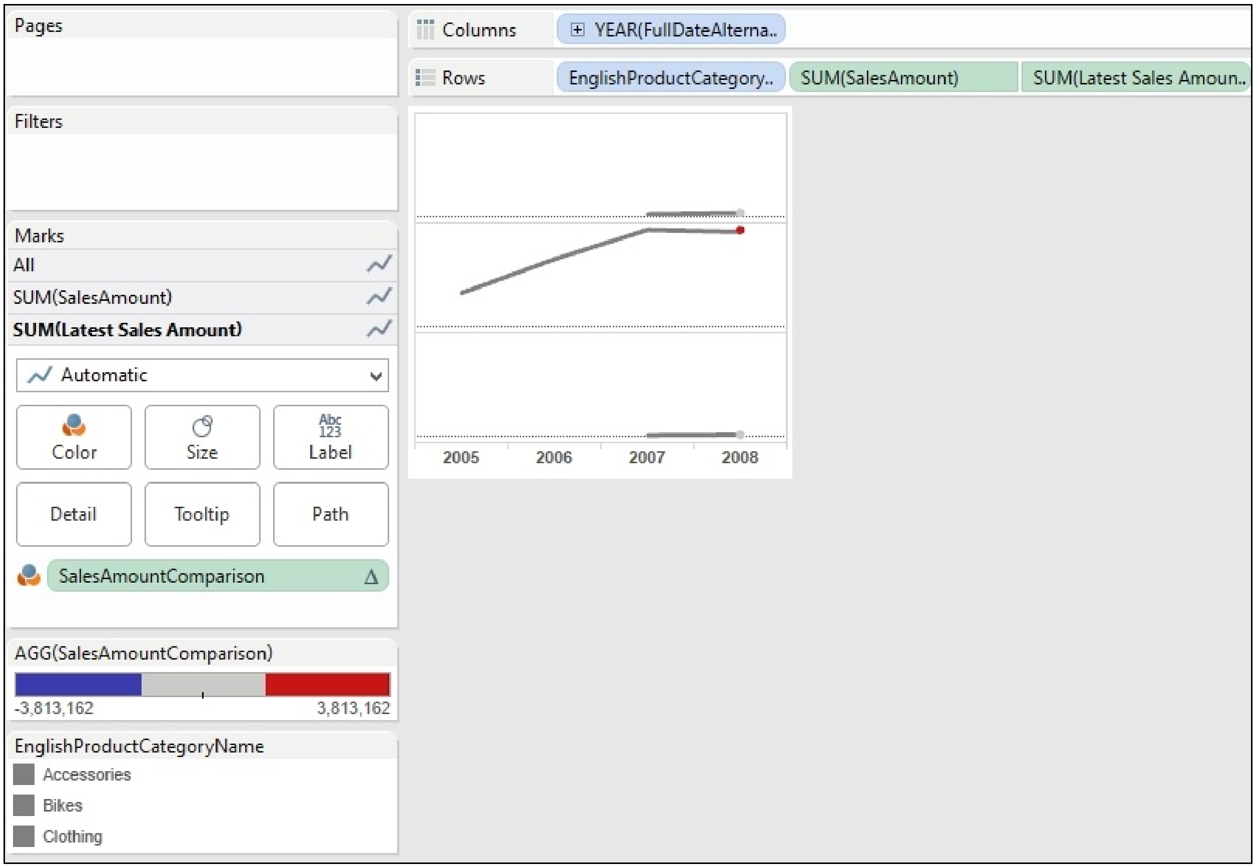
- On the **Marks** shelf, click on **SUM(FullSalesAmount)** and drag **SalesAmountComparison** onto the **Color** button.
- Let's use color to signify the result. We will categorize the color into three types: red for below average, gray for close to the average sales amount, and blue for greater than the average sales amount. On the

Color mark for **SalesAmountComparison**, select **Edit Colors…** from the right-hand side downward-facing arrow button.

- Select **Stepped Color** and type the number 3 to represent three steps. Instead of the green color, click on the green box at the left-hand side of the green bar and a color dialog box will appear. Select blue. Then, select the **Reversed** option and click on **OK**. You can see the result in the following screenshot:

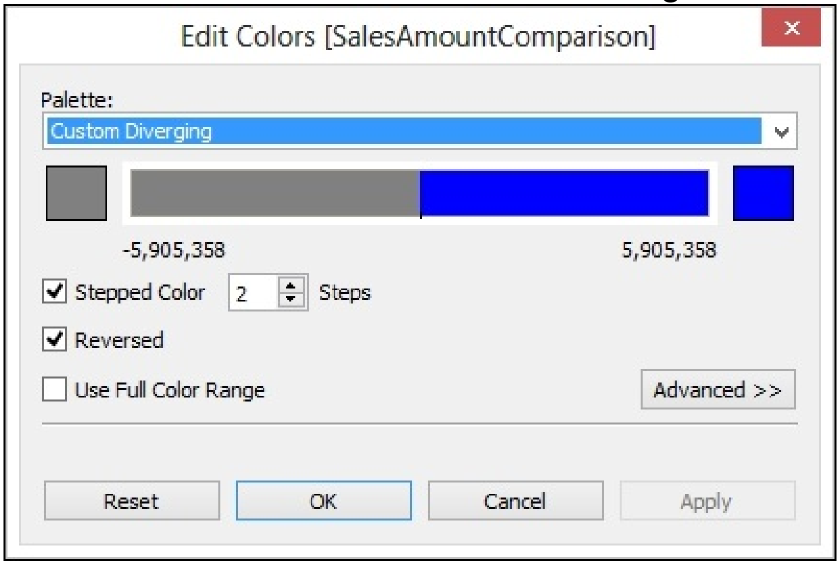


- Finally, hide the field name for the columns by selecting **FullDateAlternateKey** in the visualization and reselecting **Show Header**.
- The result so far should appear as shown in the following screenshot:

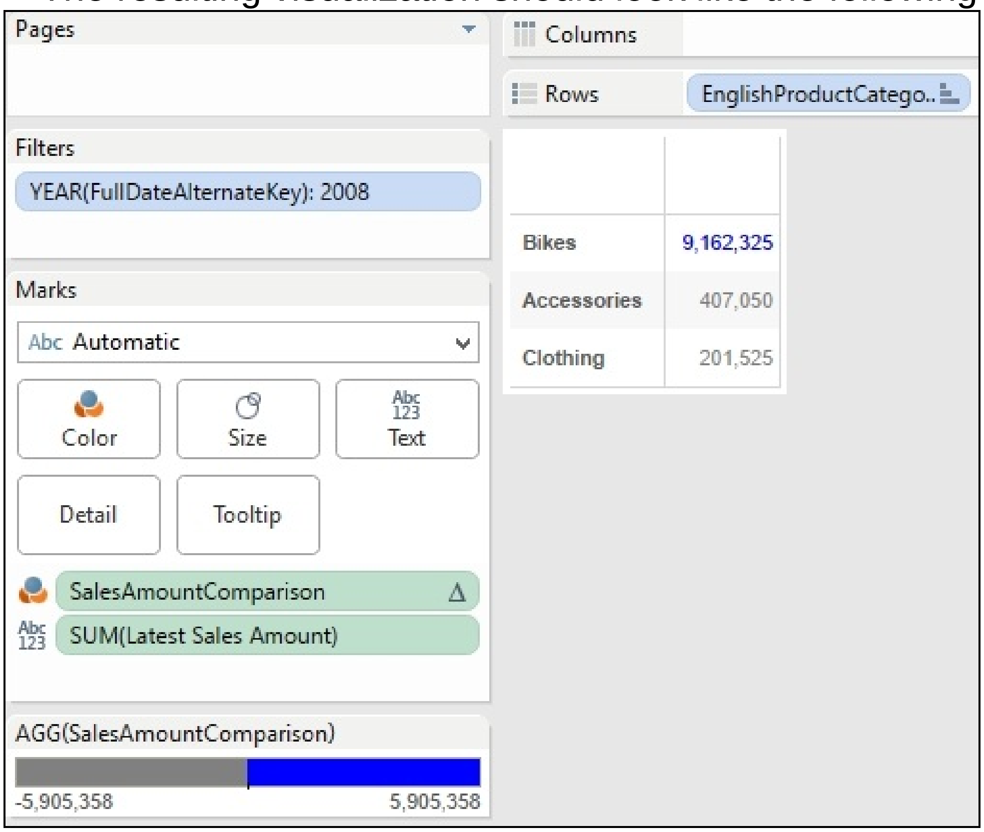


- Rename the sheet `Topic 1 Color Sparkline`.
- Duplicate the sheet by going to the tab name, right-clicking on it, and selecting **Duplicate Sheet**.
- Rename the sheet `Topic 1 Color Sparkline`.
- Go to the **Show Me** tab and select the **table** visualization.
- Drag **EnglishProductCategory** onto the **Rows** shelf.
- Drag **Year(FullDateAlternateKey)** onto the **Filters** shelf and filter by **Years** so that only data for the year **2008** is showing.
- Drag **Latest Sales Amount** onto the canvas area to show the numbers.
- On the **Marks** shelf, drag **SalesAmountComparison** onto the **Color** button.
- Click on the **Color** button and choose the option **Edit Colors…**.
- On the **Edit Colors** dialog box, choose the **Stepped Colors** option and enter the number `2`.
- Choose the **Reversed** option.
- Select the left-hand side square box, and in the color dialog box, select gray and click on **OK**.

- Select the right-hand side square box in the color dialog box, select royal blue, and then click on **OK**. You can see the final settings in the following screenshot:

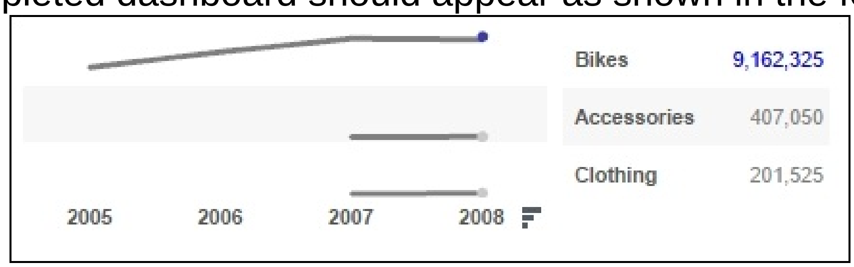


- Click on **SUM(SalesAmount)** and sort in descending order.
- The resulting visualization should look like the following screenshot:



- Now, let's put them together in a dashboard. To create a new dashboard, right-click on the name tab at the bottom and select **New Dashboard**.

- On the dashboard sheet, select the worksheets **Topic 1 Color Sparkline** and **Topic 1 Color Average** and put them next to each other on the canvas.
- On the dashboard, hide the title for the **Topic 1 Color Sparkline** worksheet by clicking on the downward-facing arrow on the right-hand side and deselecting **Title**.
- Now, hide the title for the **Topic 1 Color Average** worksheet by clicking on the downward-facing arrow on the right-hand side and deselecting **Title**.
- Resize the rows on each worksheet so that they match nicely.
- Navigate to **Format | Shading**.
- To add banding, go to the sheet tab on the format shading series of options that appear on the left-hand side of the screen. Select **Row Banding** and move the **Band Size** slider until it is halfway along the slider.
- Navigate to **Format | Lines**. Set each of the lines to **None**.
- Now click on the right-hand side visualization in the dashboard and then select the **Lines** option from the **Format** menu. In each of the **Rows** options, select **None**.
- Now go back to the **Format** menu item and select the **Borders** option. On the **Sheet** menu item, select **None** for the **Row Divider** option.
- Your completed dashboard should appear as shown in the following screenshot:



How it works…

Dual axes can be difficult to interpret, particularly if each axis is showing different measurements. However, here, dual axes can help us to create a visualization. Using a dual-axes chart here allows us to set the size and color of the circle highlights and the line chart independently.

In this recipe, we used a *ZN* formula. The *ZN* formula is used when you want to replace a zero with null values. We saw the impact of null values in an earlier chapter, and this is one option for us.

There's more…

Tableau has some great functionality which means that you can have fun with the appearance of sparklines as well as provide more information. For example, you could use the **Line End** option for the label and use advanced editing on the text label to format in order to provide more detail for the dashboard users.

Where is the three dimensional data?

The objective of data visualization is to present data so that it's easier for people to consume, spot trends, and understand the story of the data. There is a debate over the use of three dimensional in charts, and people looking at Tableau may wonder how they can make visualizations that are three dimensional. Three dimensional is not available in Tableau. Three dimensional requires the viewer to spend more time trying to understand the data being presented to them than necessary.

People often consider three dimensional as a way to bring more information into the view. We don't need to use three dimensional to make visualizations beautiful and informative. Beauty can come in the form of simplicity and understanding the data story as easily as possible. Flashy isn't always better, particularly if it misleads the viewer.

If you can't use three dimensional but still want to show different metrics on the visualization, then what are your options? If you want to display a number of variables, then we can use a scatterplot matrix, also known as a splom. This is a grid of scatter charts. What is a scatter chart? It is also known as a scatter graph, scatter chart, or simply a scattergram, and it is simply a dataset plotted as points on a graph. The *x* axis represents one variable, and the *y* axis represents another. We can arrange them in a grid so that the viewer can easily compare along the variable, and also vertically between the graphs. We can also enhance the scattergram by adding in trend lines and using color to convey a story.

In this recipe, we will look at creating a scattergram matrix using Tableau and adding in a few reference lines. This is a good option when you are asked to create something that shows the relationships between variables. In this recipe, we will create a small dashboard segment that looks at whether the sales differ on a quarterly basis.

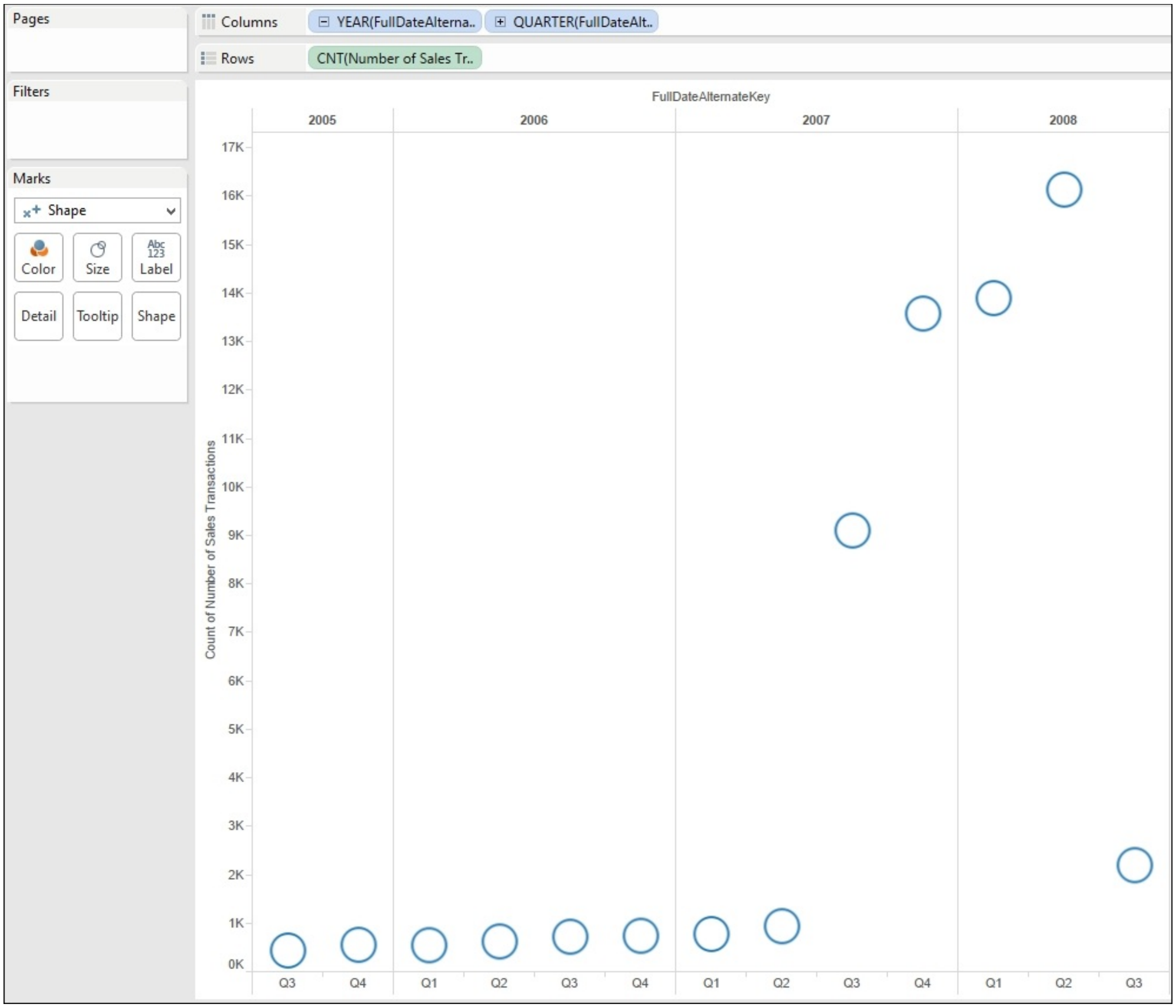
Getting ready

For the exercises in this recipe, continue to work on the `Chapter 7` workbook.

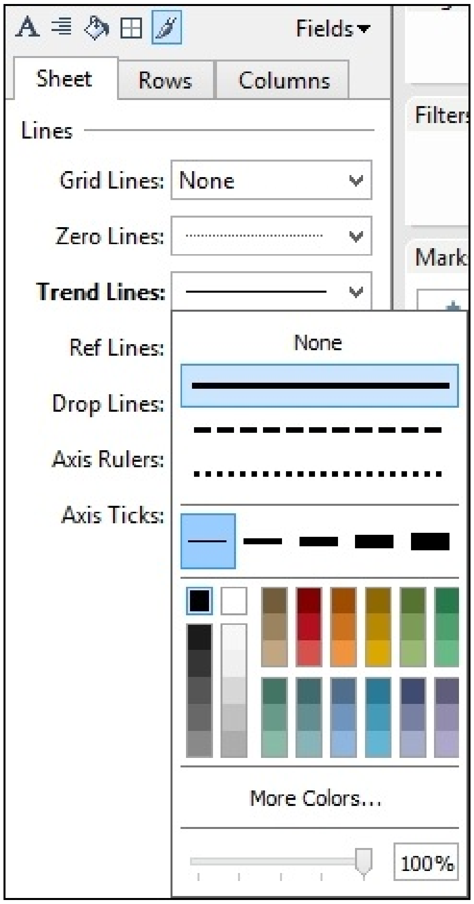
How to do it…

1. Create a new worksheet by clicking on **File** and then **New Worksheet**.
2. Drag **Year(FullYearAlternateDate)** onto the **Columns** shelf.
3. On the pill **Year(FullYearAlternateDate)**, click on the small plus sign on the left-hand side so that you now see the **Quarter(FullYearAlternateDate)** pill.
4. On the **Measures** pane at the left-hand side, rename **Number of Records** to `Number of Sales Transactions`.
5. Drag **Number of Sales Transactions** onto the **Rows** shelf.
6. Change the measure calculation to **Count** by right-clicking on **Number of Sales Transactions**, navigating to **Measure (SUM)**, and selecting the **Count** option.
7. Let's create a new calculation called **Diff from All Year's Average**. Firstly, let's right-click on the **Measures** part of the **Data** sidebar and select the option **Create Calculated Field**.
8. In the textbox, type the following formula and then click on **OK** to return to the main Tableau workbook canvas: *ZN(SUM([SalesAmount]))-Window_AVG(SUM([SalesAmount]))*

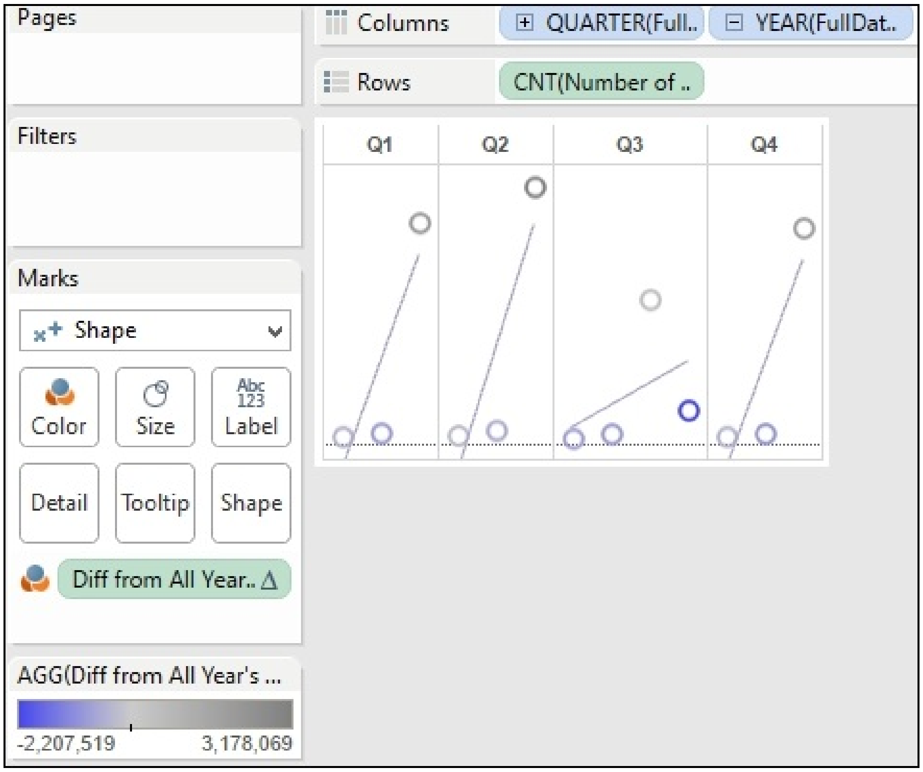
9. Let's change the visualization to a scattergram by going to the **Marks** shelf and selecting **Shape** from the drop-down list. The chart will now change to a scattergram. Here is an example in the following screenshot:



- Take the calculation **Diff from All Year's Average** and drag it onto the **Color** button. The color scheme will change to red and green.
- Click on the **Color** button and select **Edit Color…**.
- Change the red color to royal blue and click on **OK**.
- Change the green color to gray and click on **OK**.
- Next, let's add in a reference line to show the trend lines. To do this, go to the **Analysis** menu item and choose the option **Show Trend Lines**.
- The trend lines will appear for each year. We will swap it around to show a quarter by dragging **Quarter(FullDateAlternateYear)** onto the **Columns** shelf to the left-hand side of **Year(FullDateAlternateYear)**.
- Click on one of the reference lines and select the **Format** option. The **Data** pane will now change to the **Format Lines** option pane.
- For the **Trend Lines** option, click on the drop-down list and select the thinnest line. You can see an example in the following screenshot:



- Let's soften the color of the trend line by selecting the light purple color at the bottom, which is located on the bottom row at the far-right side. You will see the changes made on the trend line.
- Next, let's remove the headers in order to make the visualization as compact as possible. Remove the header **FullDateAlternateKey** by right-clicking on it and selecting **Hide Field Names for Columns**.
- On the *y* axis, right-click and deselect the option **Show Header**.
- On the *x* axis, right-click and deselect the option **Show Header**.
- The completed visualization now looks like the following screenshot:


Pages
Columns
QUARTER(Full..
YEAR(FullDat..
Rows
CNT(Number of ..
Filters
Marks
Shape
Color
Size
Abc
123
Label
Detail
Tooltip
Shape
Diff from All Year.. Δ
AGG(Diff from All Year's ...
-2,207,519
3,178,069
Q1
Q2
Q3
Q4

Eating humble pie – pie charts or not?

Pie charts are probably the most ubiquitous data visualization form. However, there is a debate regarding their effectiveness. To summarize, humans are not very good at distinguishing and comparing area or angles. Pie charts use both of these mechanisms to convey a message about the data.

There is a temptation to use lots of pie charts when visualizing data. Why? Because people like them. However, they take up a lot of space on one page to give one message on the data. It isn't possible, for example, to show timelines very effectively, and they can't be used to convey multi-dimensional data properly.

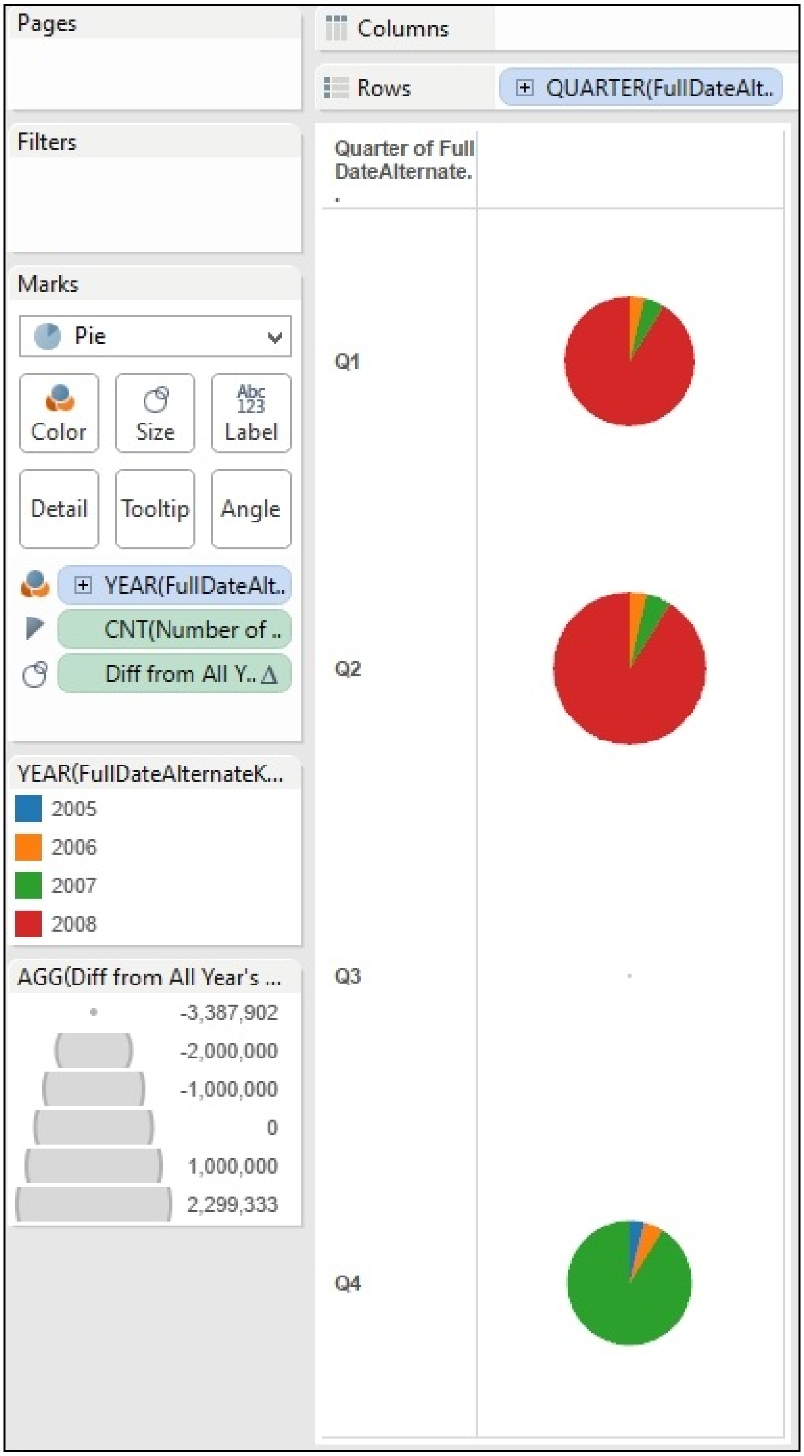
In this recipe, we will look at creating a pie chart in Tableau, and then we will take a look at a better way of visualizing the same data. This will help you to see more of the debate surrounding the humble pie chart and to see when they are most effectively used.

Getting ready

For the exercises in this recipe, we will build on the existing `Chapter 7` dashboard.

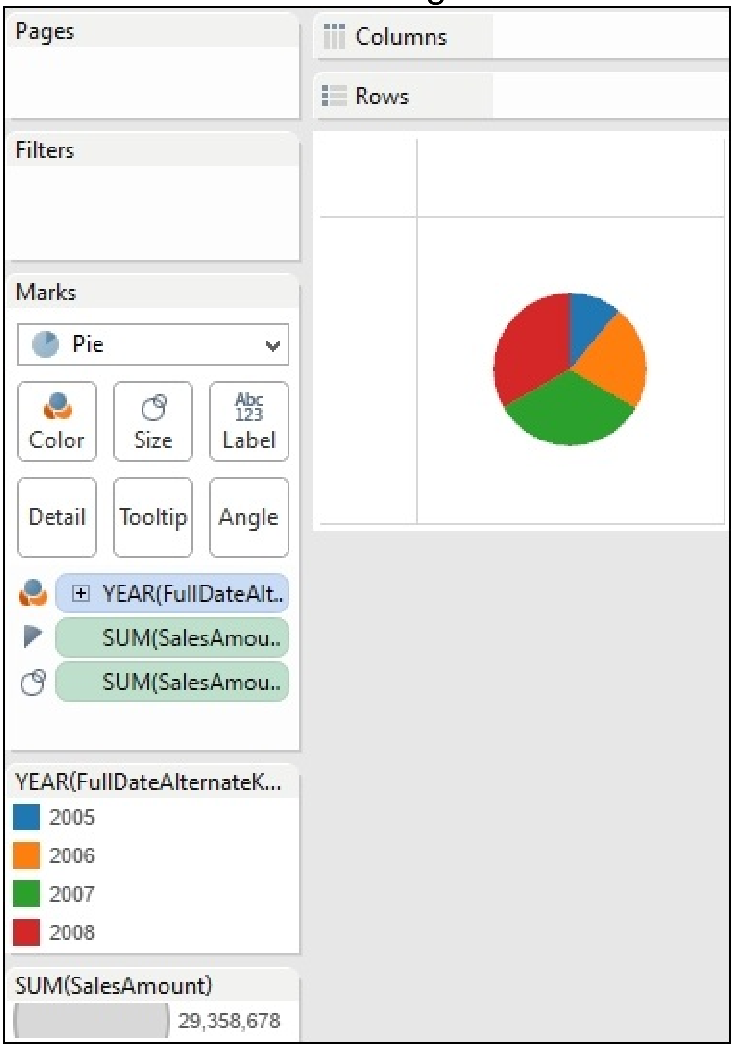
How to do it…

1. Duplicate the `Topic 3 Splom` worksheet by right-clicking on the tab and selecting **Duplicate Worksheet**.
2. Rename the duplicated worksheet to `Topic 4 Humble Pie`.
3. On the **Show Me** tab, select the **pie chart** option.
4. Your screen will now change to display pie charts. You can see an example in the following screenshot:

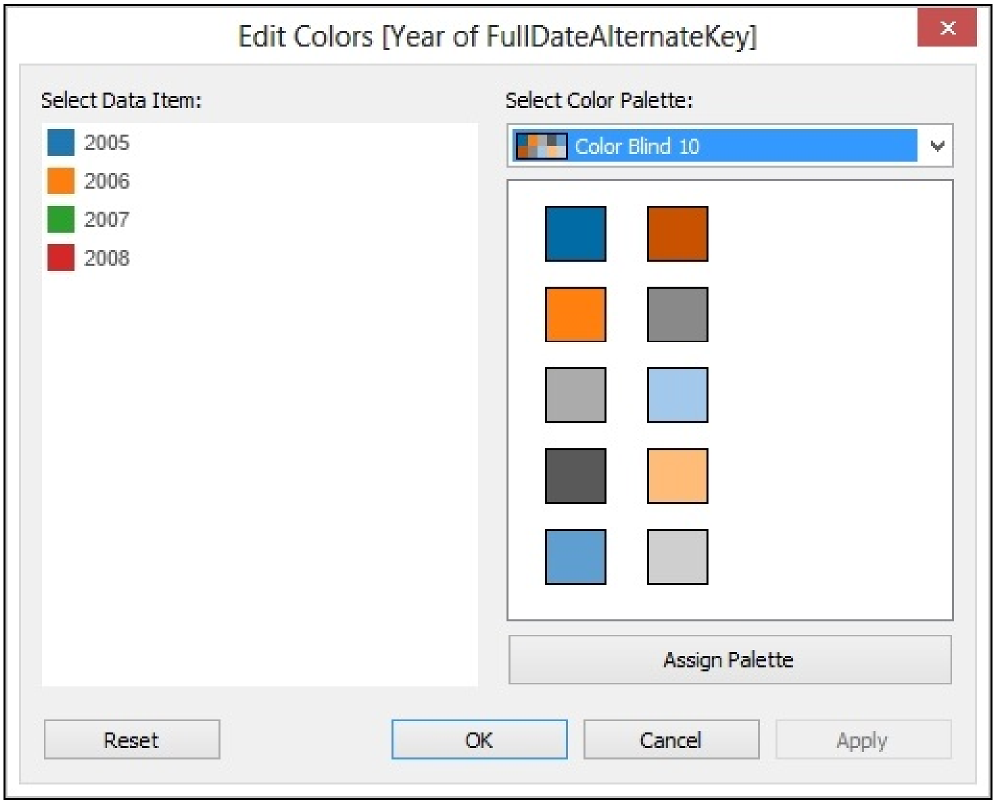


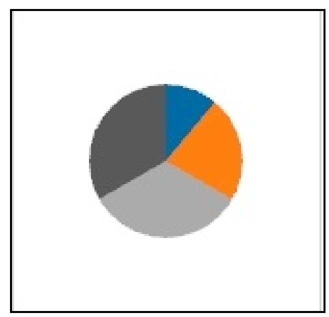
- You can see that the pie chart doesn't make any sense. Tableau has tried to display different metrics on the same pie chart, and it does not show any meaningful information.
- Let's change the pie chart so that it shows one metric only. Remove everything from the columns and rows in order to start.

- Drag **Year(FullDateAlternateKey)** onto the **Color** button and remove **CNT(Number of Sales Transactions)** from the **Marks** shelf.
- Drag **SalesAmount** onto the **Size** button. The data visualization will now look like the following screenshot:



- For some reason, Tableau has put red and green next to one another. Let's change the palette to one that is sensitive to the needs of color-blind people. Go to the **Color** button and click on it, and select **Edit Color…**.
- In the **Edit Colors** dialog box, choose **Color Blind 10** from the drop-down list, click on **Assign Palette**, and then click on **OK**. Here is an example in the following screenshot:



You can see the final pie in the following screenshot: 

How it works…

Generally speaking, pie charts are to be avoided. Humans are not accurate when they estimate area.

A key part of data analysis is comparison. Pie charts don't allow easy comparison because we find it harder to compare areas, and it is also harder for us to compare slices which are not next to one another.

So when should we use pie charts? If there are not too many slices, and the pie chart slices represent a percentage of a single measure, then this helps them to be more easily and quickly understood. They can be used when the area of a pie chart represents a measure. In the previous examples, we used Sales Amount as a single metric and showed it sliced by different categories. If we used it for lots of categories, then this would be harder to read. As a rule of thumb, no more than five slices is a good guideline.

Pies are useful when we want to put data on a map. Even then, the guidelines on slices should remain the same.

Pies can be nice as a way of filtering dashboards. So, they could filter dashboards by allowing the user to click on a particular slice to filter the whole dashboard. Pies are useful if you are displaying data on a map, as long as there are not too many slices, or you will lose the value of the data visualization.

See also

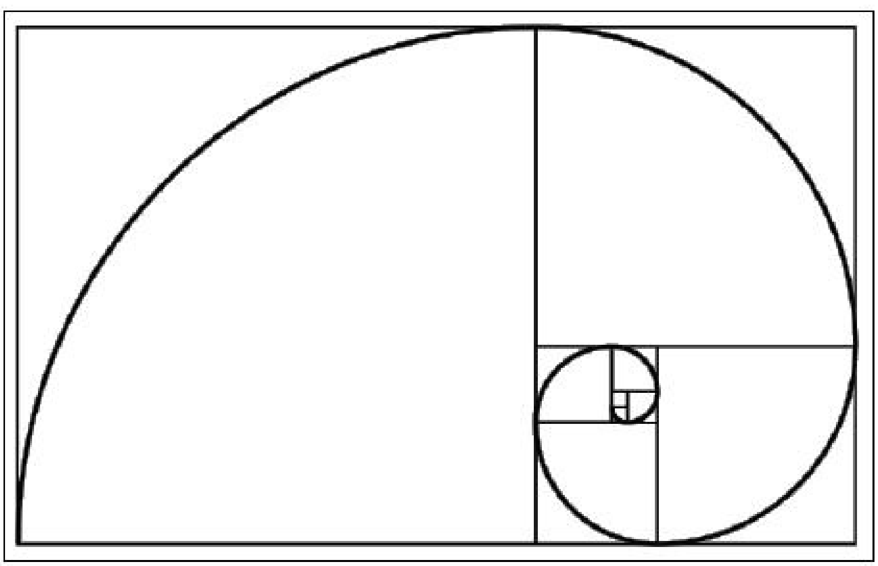
- If you want to know more about the use and abuse of pie charts, there are plenty of resources. Stephen Few's blog, *Perceptual Edge*, is a great place to start. You can view it at the following link: www.perceptualedge.com.

Sizing to make a data story

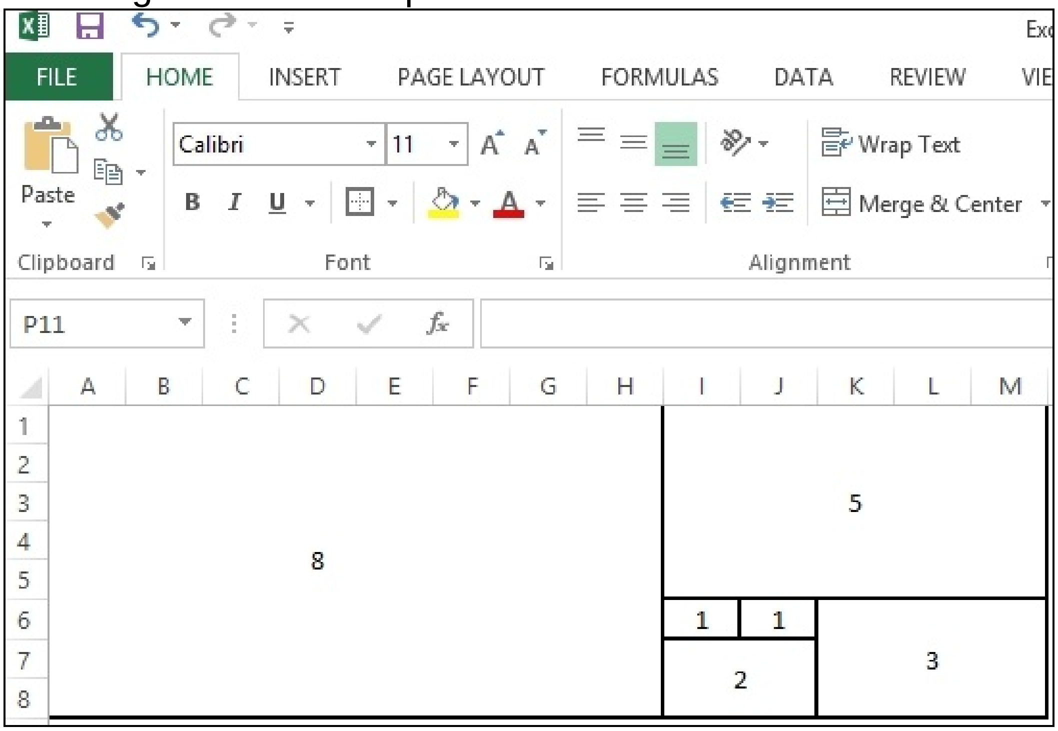
It is easy to balance columns in a report, but it is not very easy to lay out a dashboard. The size and location of dashboard elements is critical in helping users to understand the data better. How can you know how to put the dashboard elements together in order to facilitate understanding?

One key concept in design is called the golden ratio. It is based on the Fibonacci sequence, which is found in nature all around us, and it is familiar. You find the Fibonacci sequence in a number of flower petals, for example. You can also find it in art. For example, artists such as Salvador Dali and Le Corbusier have set out their works to approximate the golden ratio in the placement of the items within the painting. The golden ratio itself is 1:1.61, and is represented by the Greek character phi.

By using the golden ratio to design your dashboard, it means that you are following a natural order that is harmonious and familiar. You can use the following diagram as a way of helping you in the placement of the dashboard elements.



How can this help you to design your dashboard? The golden ratio can be used as a heuristic to help you to tell a story with the data. It can help you to decide what goes where. Here is the preceding design translated into an Excel template, which has been set into blocks. To try things out, you can swap your dashboard components around and use rectangle blocks to fill out blanks until you have finished with this aspect. The following screenshot depicts this:



We can use size as a way of determining how the dashboard components are laid out. How can we determine size and positioning? Here are some helpful tips:

- Prioritize your components in terms of their importance, and then assign the size in the order of their importance.
- Don't include dashboard components that aren't relevant. We are in the business of making the dashboard clear, not busy!
- Get your users involved. It is always possible that your prioritization and sizing is not the same as their prioritization. You could generate a series of prototypes of the dashboard. For example, you could paste pictures of your components onto the preceding template, make a picture, and then show them the different options. If you want your users to use your dashboard, why not involve them earlier in the process?

In this recipe, we will create a dashboard from the visualizations that we

In this recipe, we will create a dashboard from the visualizations that we created earlier in this chapter. We will look at the different ways in which we can place components on the dashboard and the tricks to resize things.

Getting ready

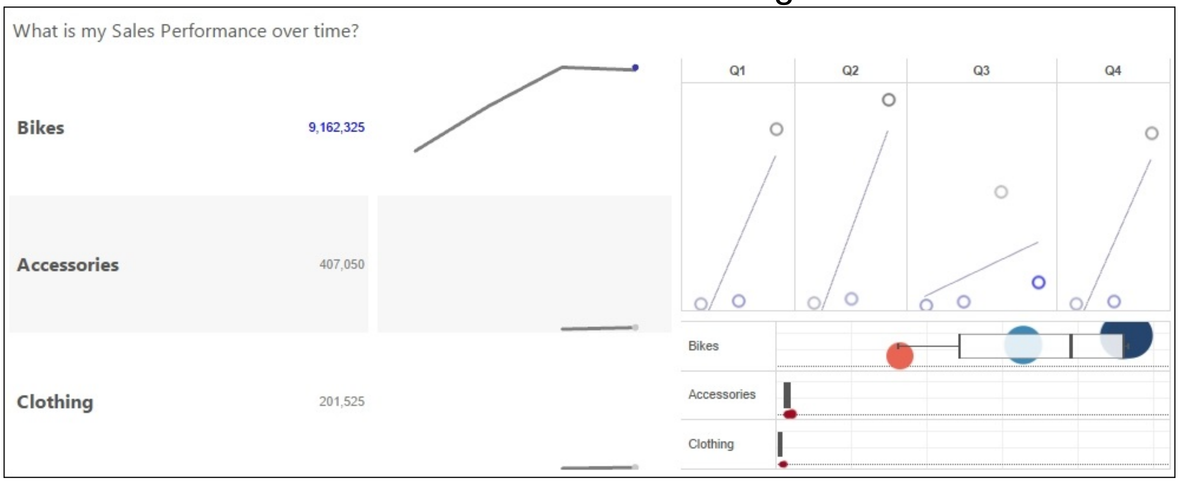
For the exercises in this recipe, we will build on the existing `Chapter 7` dashboard. We don't need to add in any more data for now.

How to do it…

1. On one of the tabs at the bottom, right-click and select the option **Create Dashboard**.
2. Name the new dashboard `Dashboard Sizing`.
3. On the new dashboard, we can treat the visualizations from the first topic as one topic. Therefore, drag **Topic 1 Color Average** onto the left-hand side of the dashboard.
4. Now, drag **Topic 1 Color Sparkline** to the right-hand side of the element.
5. Drag **Topic 3 Sales Transactions** onto the right-hand side of the element.
6. Drag **Topic 2 Dual Axes Box Whiskers** underneath **Topic 3 Sales Transactions**.
7. Remove the title from **Topic 1 Color Average** by going to the right-hand drop-down list and deselecting **Title**.
8. Remove the title and headers from **Topic 1 Color Sparkline** by going to the right-hand drop-down list and deselecting **Title**.
9. Remove the year from the *x* axis in **Topic 1 Color Sparkline** by right-clicking on one of the years and deselecting **Show Header**.
10. Insert a title at the top by dragging the text to the top of the dashboard.
11. In the textbox that appears, type `What is my Sales Performance Over Time?`.
12. Change the font to **Segoe UI**.
13. Change the font size to **12**.
14. Let's resize the dashboard by inserting a blank tile underneath the

other visualizations. Making it smaller will make it neater! To do this, drag **Blank** to the right-hand side.

15. Your dashboard will now look like the following screenshot:



How it works…

In this recipe, we have experimented with layout. If we had more product categories, the sparkline would be smaller, we would have more rows, and it would appear much smaller. However, the purpose was to illustrate the importance of sizing. We can see that using the golden ratio as a template is a good starting point, and you can bring the users along with you.

See also

If you want to know more about dashboard design, Stephen Few's book is an excellent read and a follow-on from this topic. The details are as follows:

- *Information Dashboard Design: The Effective Visual Communication of Data*, *Stephen Few*, *O'Reilly Media*

Part 2. Module 2

Data Visualization with Tableau

Over 70 recipes for creating visual stories with your data using Tableau

Chapter 1. Connecting to Data Sources

We will cover the following topics in this chapter:

- Connecting to text files
- Connecting to Excel files
- Connecting to Access databases
- Connecting to a SQL Server
- Pasting from a clipboard
- Connecting to other databases
- Connecting to Windows Azure Marketplace
- Understanding dimensions and measures
- Changing data types
- Applying filters
- Merging multiple data sources

Introduction

This chapter will cover the basics to get Tableau connected with various data sources, such as text files, Excel/Access files, SQL Server, ODBC sources, and the clipboard. We will cover simplistic versions of data files, where data is clean and ready-to-use. This chapter also covers how to apply filters to reduce the available data for analysis as well as merging two different data sources.

Connecting to text files

When you open Tableau for the first time, you should see a screen similar to the one shown in the following screenshot. This image shows the various data sources available for analysis. Tableau provides you with two sample data sources, **Sample - Coffee Chain (Access)** and **Sample - Superstore Sales (Excel)**, as shown in the following screenshot:

Getting ready

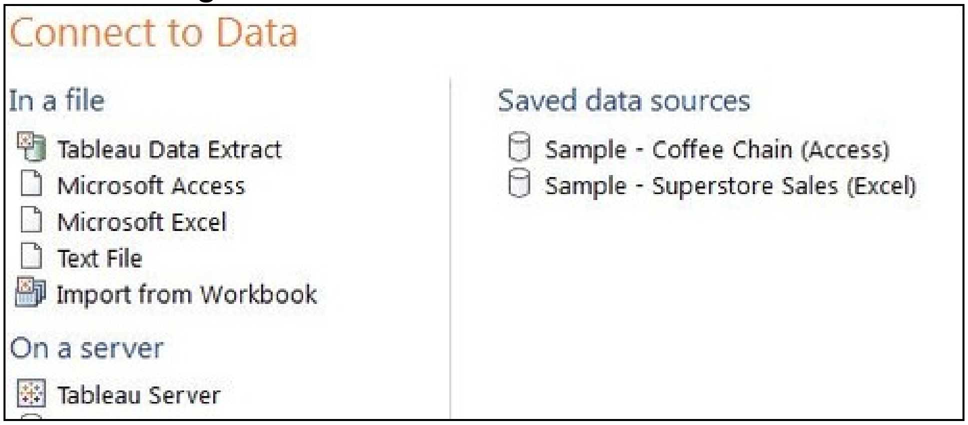
To prepare for the recipe, download and save `titanic.txt` from http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic.txt on your local hard drive. Remember this location, as we will use this file for this recipe. This file lists all the passengers (and their details) that boarded Titanic on its disastrous voyage.

> **Tip**
>
> **Downloading the example code**
>
> You can download the example code files for all the Packt books you have purchased from your account at http://www.packtpub.com. If you have purchased this book elsewhere, you can visit http://www.packtpub.com/support and register to have the files e-mailed directly to you.

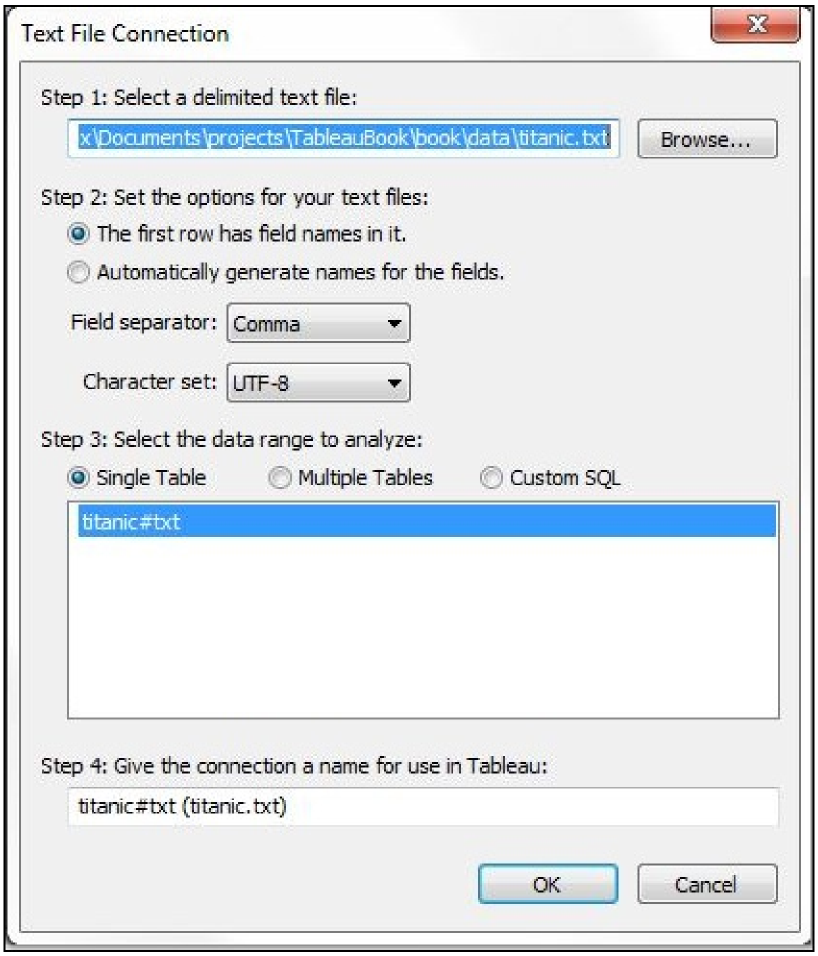
How to do it...

Once you have downloaded the text file, perform the following steps to get the data in Tableau:

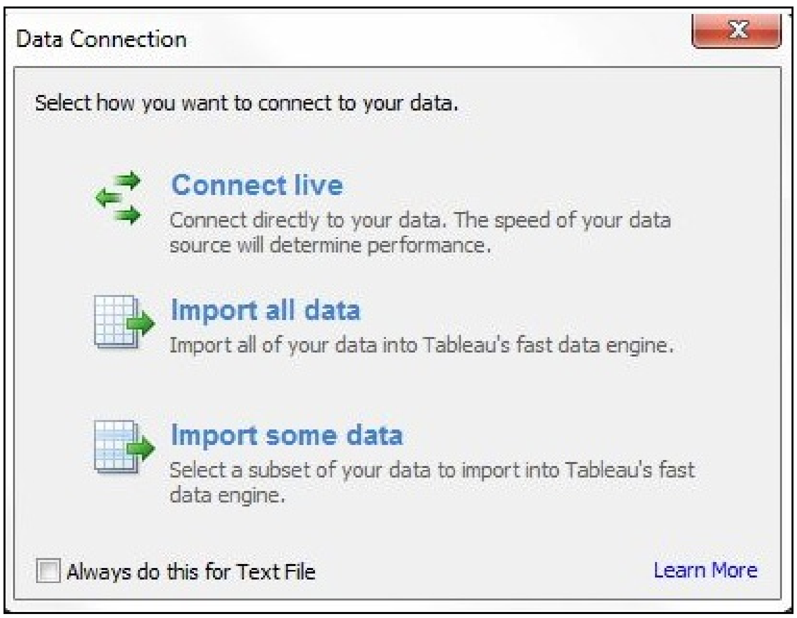
1. Click on the **Connect to data** link to expand that area as shown in the following screenshot:



- Click on **Text File** under the **In a file** section to launch the following screen:



- Find and select `titanic.txt` in the given **Open** dialog box.
- As Tableau loads the data, it will prefill some of the options. For example, the field separator or delimiter and the header row. In this case, the field separator is a comma and the first row does have field names in it. So, hit **OK** on the dialog box as shown in the previous screenshot.
- Tableau provides three options to allow you to interact with the data, which is a text file in this recipe. These three options are shown in the following screenshot. By using the **Connect live** option, we can use the file connection as it is, and by using the **Import all data** or **Import some data** option, we can speed up the analysis by importing the data in Tableau's own format. In this case, let's just use the **Connect Live** option to load all 1,313 rows in the `titanic.txt` file.



- As you can see, Tableau determined the data types and put some fields from the text file in the **Dimensions** section and others in **Measures**. Tableau determines data types of various fields using the Microsoft Jet Database Engine driver. Due to the driver's limitations, however, some fields are at times misinterpreted as measures when they should be detected as dimensions and vice versa. Since the field **survived** shows up as a **Measures** section, but contains a binary value of zero and one (no and yes), it would make sense to convert that field to a **Dimensions** section. To do so, simply drag the field over to the **Dimensions** section or right-click on the field and click on **Convert to**

Dimension.

How it works…

We used a text file as a data source and connected to it using Tableau's data source connection options. Although most of the time Tableau can determine data types accurately, sometimes you need to pay attention to changing the data types to reflect the actual data type. In this case, we converted a binary field (containing zero and one) from the **Measures** field to the **Dimensions** field.

There's more...

In its online Knowledge Base, Tableau discusses how to handle situations where data types are misinterpreted because of Microsoft's Jet Database Engine's limitations. You can find that article at http://kb.tableausoftware.com/articles/knowledgebase/jet-incorrect-data-type-issues.

Connecting to Excel files

Since Microsoft Excel is a very commonly used tool for analyzing data, Tableau makes it easy for the users to connect to Excel files.

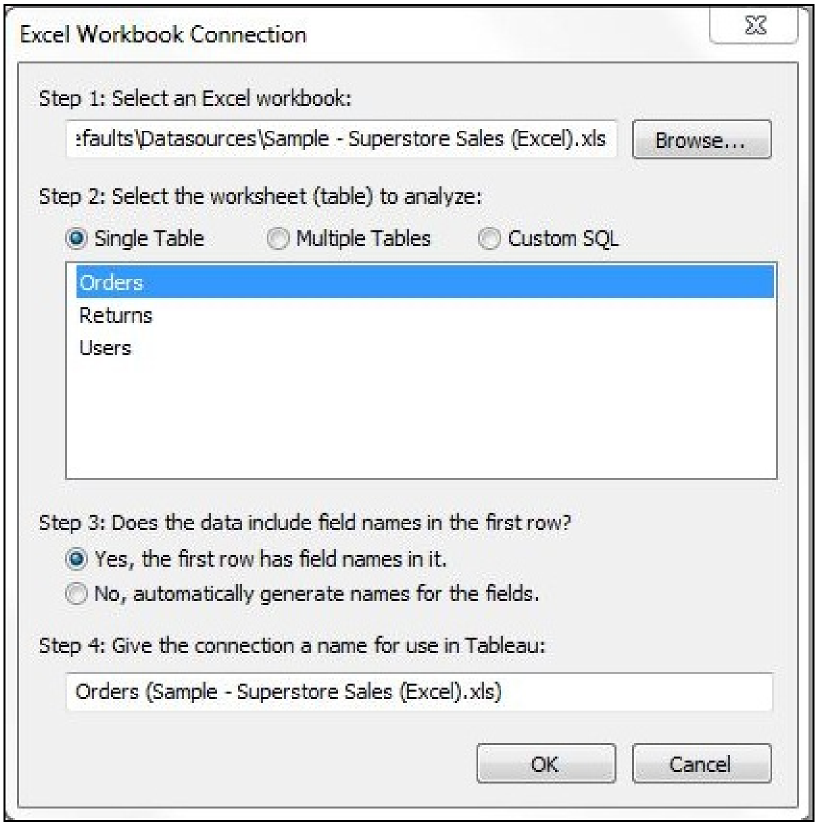
Getting ready

To use an Excel file as a data source, let's use the sample file that comes with the Tableau installation. Unless you have customized your Tableau installation, you should find the `Sample – Superstore Sales (Excel).xls` file when you navigate to **My Documents | My Tableau Repository | Datasources**.

How to do it...

Once you have identified the presence of the sample Excel file, perform the following steps to connect to the Excel file:

1. From Tableau's main screen, click on **Connect to Data** as shown in the following screenshot.
2. Under the **In a file** option, select **Microsoft Excel** as the connection option.
3. Browse and select the file `Sample - Superstore Sales (Excel).xls`.
4. Tableau will determine the number of sheets in the file and provide an option to import a single worksheet (also called tabs or sheets) or multiple worksheets.



5. Select the **Orders** sheet and hit **OK**.
6. Use the **Connect live** option to get the data loaded as is from the Excel file.
7. You'll see that Tableau determines the field types based on the available data.

How it works...

When Tableau connects to Excel files, it provides options for connecting to a single worksheet or multiple worksheets. Custom SQL commands can also be written to access data from multiple worksheets of an Excel file. In this recipe, we connected to a well-formatted worksheet from the sample Excel file. As was the case with plain text files, Tableau determines the data types of the fields using Microsoft's Jet Engine Driver.

There's more...

Although we used a well-formatted Excel file for this recipe, we know that analysts spend a lot time cleaning and manipulating data before any analysis. Before connecting to Tableau, we have to make sure the Excel files are formatted according to what Tableau is expecting as a data source. The *Preparing Excel Files for Analysis* article in the **Knowledge Base** section at the following link provides more information on how to prepare Excel files to be used in Tableau for analysis: http://kb.tableausoftware.com/articles/knowledgebase/preparing-excel-files-analysis

Connecting to Access databases

Microsoft Access is a good tool to store smaller datasets in a relational database format without purchasing and installing complete data storage solutions, such as Microsoft SQL Server, Oracle, or MySQL. Tableau provides an option to connect to Access databases.

Getting ready

Let's use the Access database (`Sample - Coffee Chain.mdb`) that came with the standard installation. As with the Excel file used in the previous recipe, unless you made any customization during installation, the database file should be found by navigating to **Documents | My Tableau Repository | Datasources**.

How to do it...

Once you have located the Access database file, perform the following steps to connect to the sample Access database file:

1. From Tableau's main screen, click on **Connect to Data**.
2. Under the **In a file** option, select **Microsoft Access** as the connection option.
3. Browse and select the file `Sample - Coffee Chain.mdb`.
4. Tableau will determine and list tables found in the database and will ask whether to analyze a single table or a query. Select **CoffeeChain Query** and hit **OK**.
5. Use the **Connect live** option to connect to the data from the Access database.
6. You'll see that Tableau loaded the query from the Access database and also determined the data types using Microsoft's Jet Engine Driver.

How it works...

Using Microsoft's Jet Engine Driver, Tableau connects to Microsoft Access and determines the data types of the fields of a table. Just like the connection to Excel files, Tableau allows the user to select a single table, multiple tables, or write custom SQL commands.

Connecting to a SQL Server

Although it is pretty easy to connect to Access, Excel, and other flat files, data is frequently stored on some sort of relational database on a server, such as on the SQL Server or Oracle. Tableau offers connections to various data stores too. Here, we'll focus on Microsoft SQL Server.

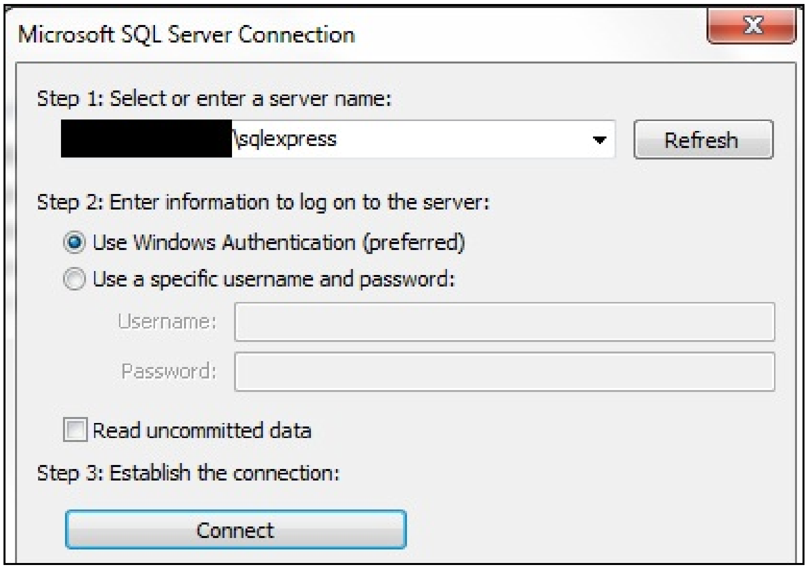
Getting ready

Security roles, server connections, authentication issues, port and firewall details, and other factors can create problems while trying to access data from a server. The solutions to these problems are out of the scope of this book but you should make sure that you are able to access the server database from the same computer on which Tableau is installed.

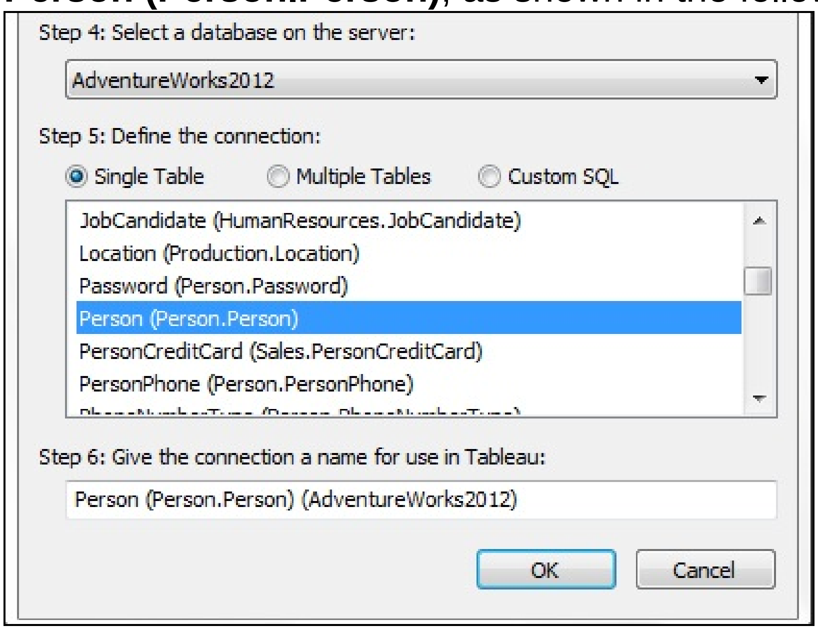
How to do it...

Once you have made sure you have access to the database server and the database, perform the following steps to connect to a Microsoft SQL Server table:

1. From Tableau's main screen, click on **Connect to Data**.
2. Under the **On a server** option, click on **Microsoft SQL Server**.
3. In the **Microsoft SQL Server Connection** dialog box, enter the server name as shown in the following screenshot.
4. Click on the **Connect** button under the **Establish the connection** option.



5. Select **AdventureWorks2012** as a database on the server.
6. Under **Define the connection**, select **Single Table** and then **Person (Person.Person)**, as shown in the following screenshot:



- Hit **OK**.
- Select the **Connect Live** option to connect to the SQL Server database directly.

How it works...

Using the the **Connect Live** option in Tableau, we connected to a SQL

Server database directly. This option allows users to create visualizations that will be refreshed as the underlying data changes. If connected live, Tableau will create results based on the SQL Server's settings, which are usually set to maximize performance.

Pasting from a clipboard

Sometimes it is easier to just paste data from the clipboard than pasting it to Excel or CSV files and then importing them again in Tableau. Tableau does provide a quick import method from the clipboard.

Getting ready

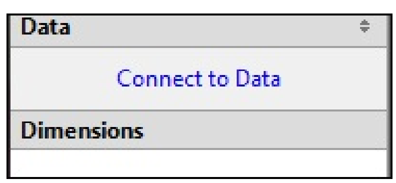
Let's use the Titanic dataset that we used in the *Connecting to the text files* recipe. The file is at http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic.txt.

How to do it...

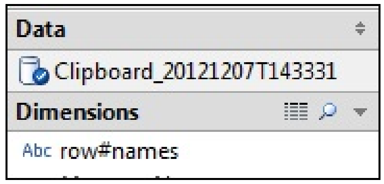
Once you have determined that the Titanic dataset file is on your local drive, perform the following steps to copy the data to Tableau using a clipboard:

1. Open the file in Notepad.
2. Select all the data and copy everything from the file to the clipboard (*Ctrl* + *A* and then *Ctrl* + *C* on Windows), and perform the following steps:

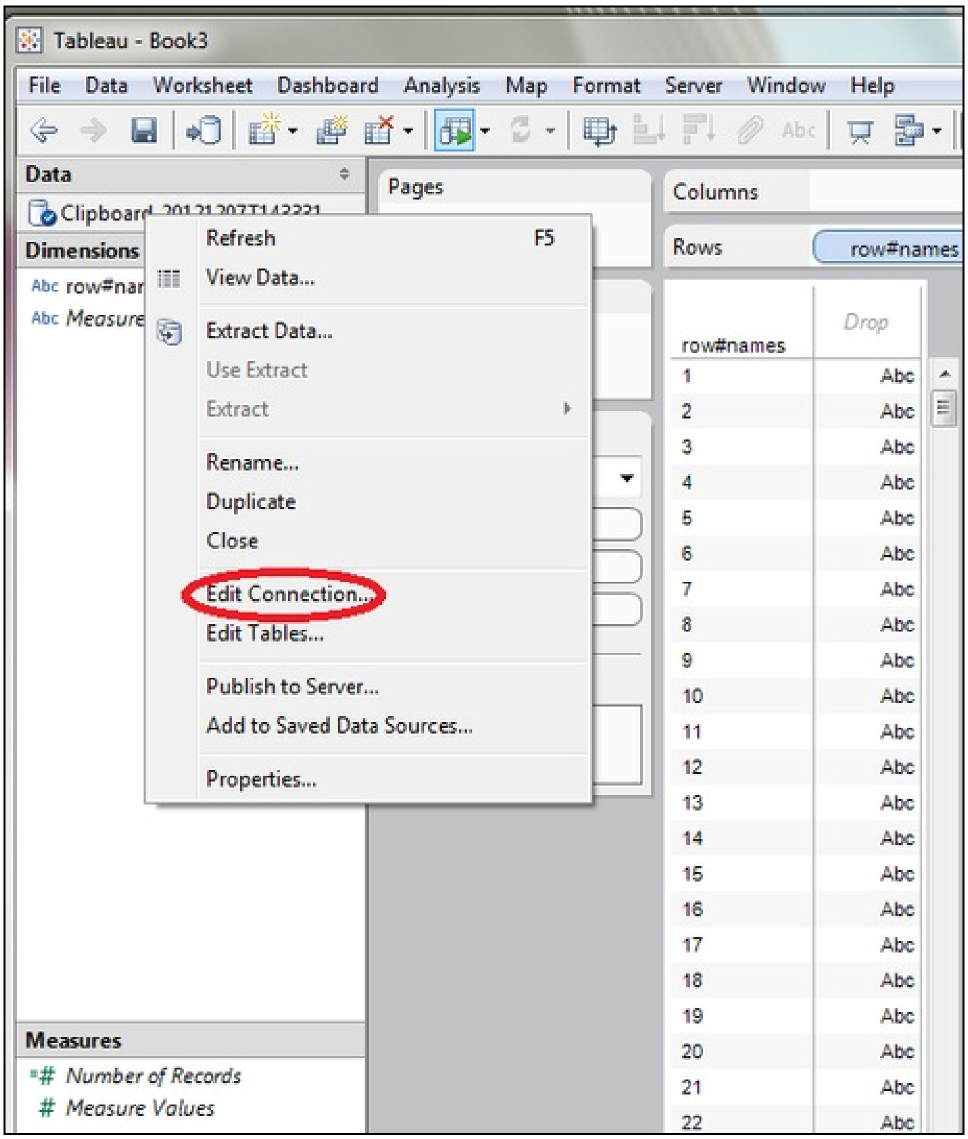
    1. In Tableau, navigate to **File** | **New** to open a new blank workbook.

    2. Click on any open area on the workbook and paste the copied data by going to **Data | Paste Data** (*Ctrl* + *V* on Windows). Note the top-left area of the workbook, where the data connections are shown. It should show text such as **Clipboard_timestamp**, where **timestamp** is the time and date when the paste occurred.



- As you can see from the workbook, the data was improperly imported. To fix this problem, right-click on the data connection named as **Clipboard_timestamp** under **Data** and click on **Edit Connection**.



- From the **Field** separator drop-down menu, select **Comma**; we do this because the original file was separated by commas and Tableau used tab as the separator.

- After making your selection, hit **OK** and you'll see that all the fields from the Titanic data text file are shown in **Dimensions** and **Measures**.

How it works...

While copying data from the clipboard, Tableau uses tab as the default separator of data. This causes improper import of data when the data is separated by other delimiters such as a comma. You can easily correct this problem by editing the connection to the clipboard file saved in Windows' temporary folder.

There's more...

If you created visualizations based on the data from the clipboard and you need to regularly update this visualization, you will find the Tableau online article, *Editing Pasted Data*, in the **Knowledge Base** section at the following link:
http://kb.tableausoftware.com/articles/knowledgebase/editing-pasted-data

This article explains how to save the data source of the clipboard and modify the data source.

Connecting to other databases

Connecting to most of the databases on a server is straightforward. Providing the server name and authentication details is usually sufficient. There are at times, however, when either Tableau does not provide a direct connection to that database server or you want to use an ODBC connection.

Getting ready

The easiest way to make a connection to a database server is using an ODBC connection. To use this type of connection, we must first set up a **Data Source Name** (**DSN**). For this recipe, we will set up a DSN to connect to a database server.

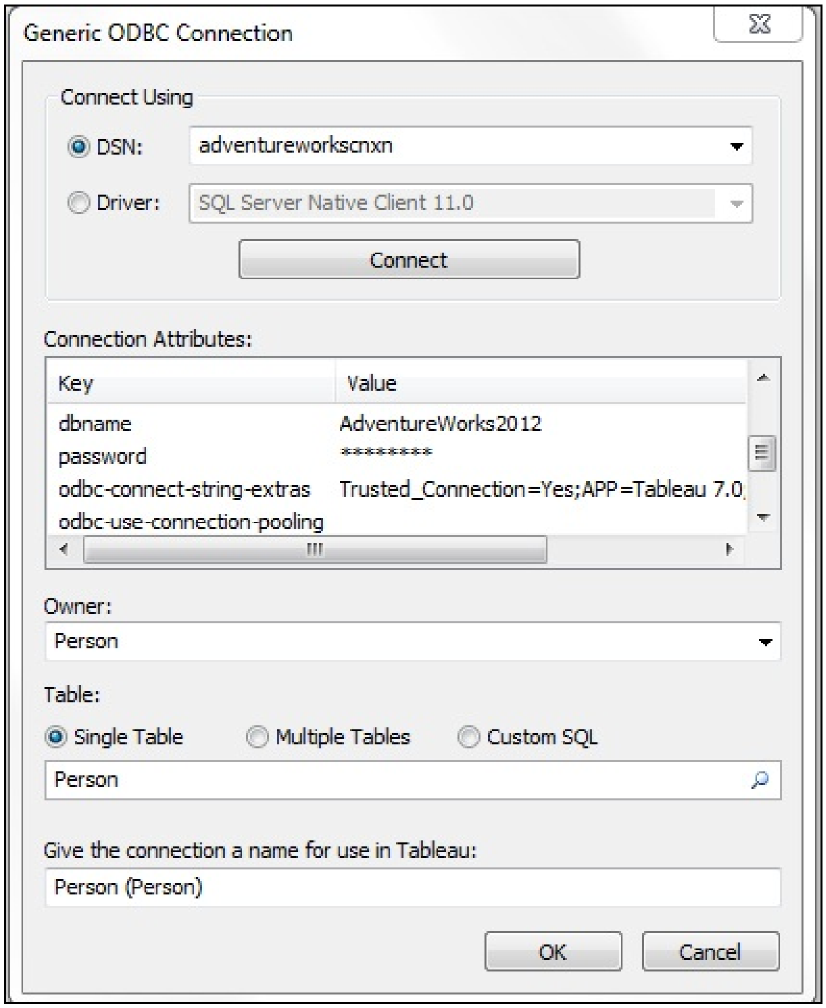
In **Control Panel**, click on **System and Security** and then on **Administrative Tools**. Double-click on **Data Source (ODBC)**. You'll see a **Data Source Administrator** window. Under **User DSN**, click on **Add** and follow the steps to create a DSN for your database. If you have a SQL Server instance installed and the `Adventure Works` database populated, select **SQL Server Native Client** and hit **Finish**. In the **Name** field, enter a name that you'll remember easily; remember that it cannot contain spaces. I chose `adventureworkscnxn`. Under **Server**, either enter the database server name or select **Local**. Continue with the default selections until you see a **Finish** button.

How to do it...

Once the DSN is set up, open a new worksheet in Tableau and perform the following steps to connect to a database using ODBC:

1. Click on **Connect to data** and select **Other Databases (ODBC)**.
2. In the **DSN** dropdown, select the DSN that you created earlier.
3. Click on **Connect** to test the data connection.
4. Under **Owner**, select **Person**.

5. Among the **Table** selection radio buttons, select **Single Table**; search for a table name by clicking on the magnifying glass icon.
6. Under the **Give the connection a name …** textbox, enter a name and hit **OK**.



7. If you can see the **Connect live**, **Import all data**, and **Import some data** options in the **Data Connection** page, you were able to successfully connect to the SQL Server using ODBC and DSN.

How it works...

With ODBC, Tableau provides an option to connect to the data sources that otherwise do not have native support in Tableau. This option provides flexibility to connect any data source that has an ODBC drive.

Although SQL Server is supported directly by Tableau, in this recipe, we saw how easy it is to create a data connection using an ODBC driver and DSN.

There's more...

Microsoft's online support provides an excellent overview of ODBC connectivity at http://msdn.microsoft.com/en-us/library/windows/desktop/ms710252%28v=vs.85%29.aspx.

Connecting to Windows Azure Marketplace

Microsoft created an online platform called Windows Azure Marketplace for trading **Software as a Service** applications and data. Users can choose to buy and sell various datasets, and that makes it a great place to use datasets hosted on the cloud with Tableau.

Getting ready

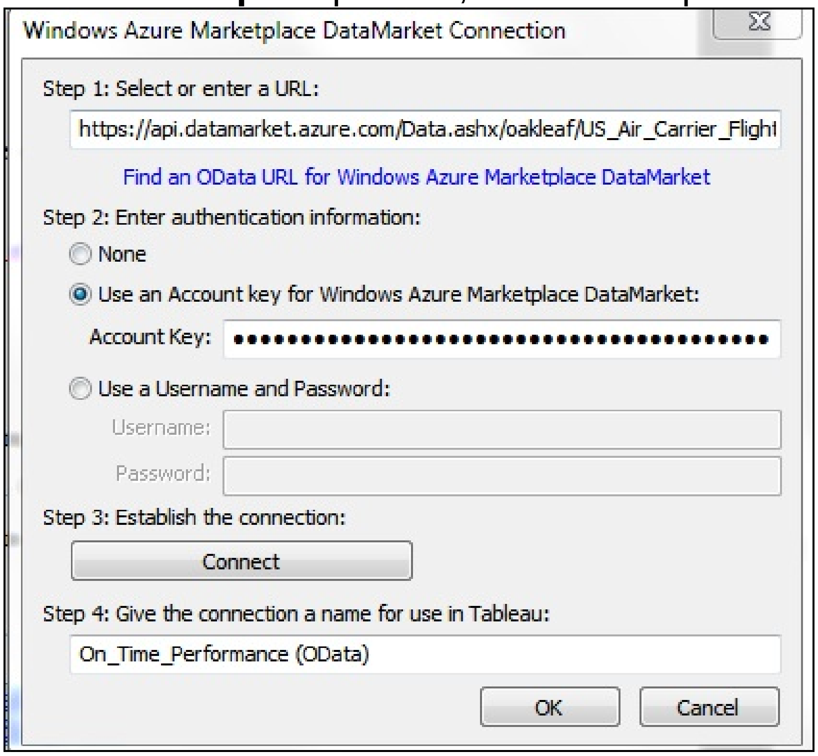
To access the Azure Marketplace datasets, you'll need to create or already have an account with Microsoft (which was earlier called Windows Live ID). Once you are logged in with your Live ID, you'll also need to complete the registration process.

How to do it...

Once you have completed the registration process for the Azure Marketplace, perform the following steps to get the data from the Marketplace to Tableau:

1. Log in to Windows Azure Marketplace (datamarket.azure.com) with your account credentials.
2. Click on **Data** on the top navigation menu.
3. Click on the **US Air Carrier Flight Delays** dataset.
4. Click on **Sign-up** to subscribe to the dataset, and perform all the required steps.
5. Once the subscription process is complete, click on the **Explore this Dataset** link.
6. In the **Build Query** area, select **On_Time_Performance** from the **Query** field and hit **Run Query**.
7. You should see a URL for the currently expressed query, which would look something like https://api.datamarket.azure.com/Data.ashx/oakleaf/US_Air_Carrier_

$top=100.
8. Copy this URL.
9. In Tableau, go to the **Connect to Data** options screen.
10. Click on **Windows Azure Marketplace DataMarket**.
11. In the **OData Connection** pop-up box, shown in the following image, under the **Step 1** input box, enter the copied URL:



- In **Step 2** of this process, either select the **Account key** or **Username** option and enter the credentials (you'll find the **Account key** value below the OData URL information; you need to click on **Show** to make the key visible).
- In **Step 3**, click on the **Connect** button. If you do not see any message, you will be able to connect to the data file.
- In **Step 4**, enter a name for the connection and hit **OK**.

How it works....

Microsoft Azure Marketplace offers a data market for users to explore datasets, which are usually scattered everywhere on the Web. This marketplace creates a central repository of datasets, and with Tableau's integration of this marketplace, it is very easy to analyze various datasets.

There's more...

Microsoft's online support provides detailed information for users, who could be consumers or publishers of the data, as well as developers of Azure applications and services. You can read it at http://msdn.microsoft.com/en-us/library/windowsazure/gg315539.aspx.

Understanding dimensions and measures

Tableau divides the data in two main types: dimensions and measures. Dimensions are usually those fields that cannot be aggregated; measures, as its name suggests, are those fields that can be *measured*, aggregated, or used for mathematical operations. Dimension fields are usually used for row or column headings; measures are usually used for plotting or giving values to the sizes of markers.

When you import the data for the first time, Tableau determines whether to consider a field as a dimension or a measure. This determination involves considering fields with all text (nominal or other text) values and fields with numeric values. Depending on the data source, Tableau also uses Microsoft's Jet Engine Driver to classify fields into dimensions and measures.

Tableau visualizations are heavily dependent on the structure of dimensions and measures. Thus, organizing data properly into dimensions and measures is important, and if Tableau's determinations are wrong about the field data types, it is easy to convert these fields to the other category. Simply dragging the field to the pane works just like right-clicking on the field and clicking on **Convert to Dimension** or **Convert to Measure**.

Changing data types

Depending on the data source and connection, Tableau tries to determine the field data type. Most often, the field data types are identified correctly; sometimes, however, changing data types becomes necessary.

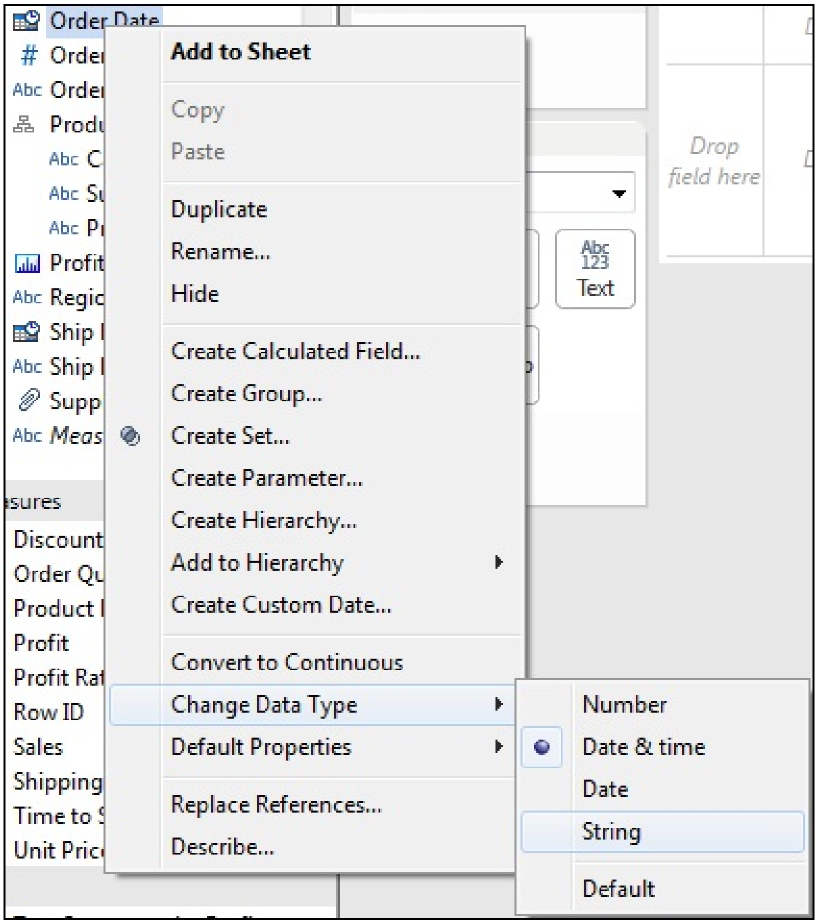
Getting ready

We will use the sample superstore sales saved data source for this exercise. Open a new worksheet and connect to the **Sample – Superstore Sales (Excel)** data source.

How to do it....

Once the sample file is loaded on the worksheet, perform the following steps to convert data types:

1. In the **Dimensions** pane, right-click on **Order Date**.



2. You'll notice some data types in Tableau: **Number**, **String**, and **Date**. **Date & time** is also a type, which is suited for data with a timestamp.
3. Select **String** as the data type for this field. Next to this field name, you'll notice a symbol with letters (**Abc**); this symbol indicates that this field contains data of type String.
4. Drag the **Order Date** field from the **Dimensions** to the **Measures** pane.
5. You'll notice in the **Measures** pane that the field **Order Date** has an aggregation of **Count**.
6. Right-click on **Order Date (Count)** in the **Measures** pane and select **Change Data Type**. Select **Date** as the new data type.
7. You'll notice that the **Order Date** field is back in the **Dimensions** pane.

How it works...

Since the data type and role of a field (dimension or measure) determines how the data will be used in the visualizations, it is critical to have the right data type for fields in the data. You will notice that, if you convert a field in the **Measures** pane to a **Date** type, that field will be moved to **Dimensions**. If a field from the **Dimensions** pane is converted to **Number**, it will stay in the **Dimensions** pane. If a field from the **Measures** pane is converted to **String**, the default aggregation changes to **Count**.

Applying filters

If you want to reduce the amount of data available for visualizations or restrict the data for a particular field value, applying filters is a very good solution. This recipe will provide a basic overview of filters, and later in the book you'll see some other uses of filters.

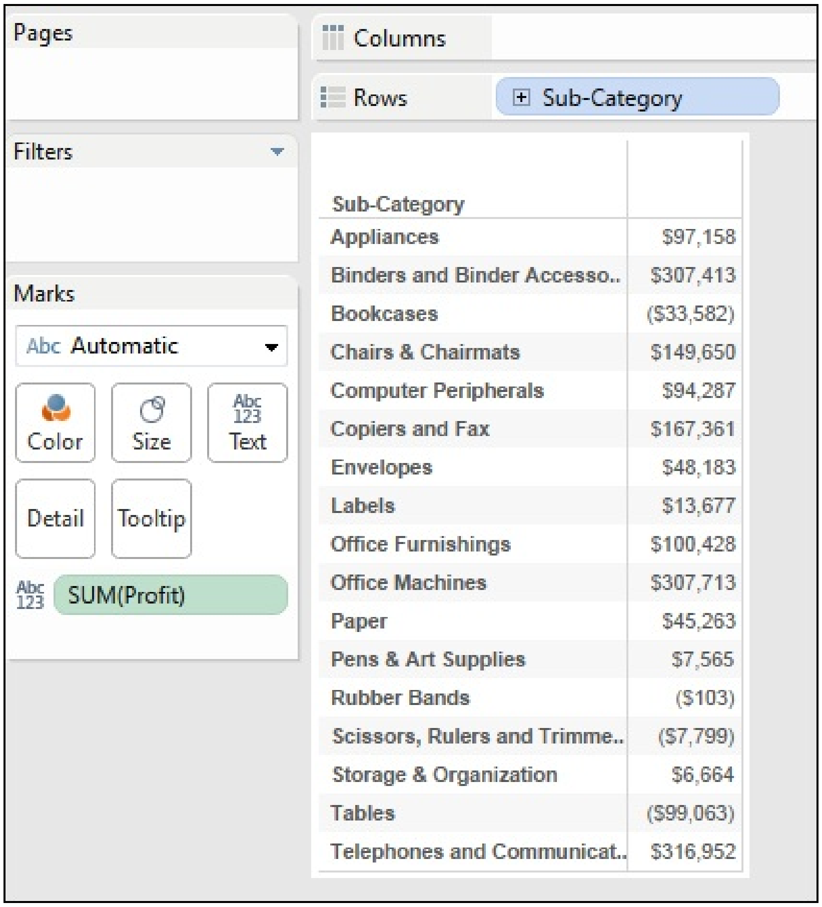
Getting ready

We will use the sample superstore sales saved data source for this exercise. Open a new worksheet and connect to the **Sample – Superstore Sales (Excel**) data source.

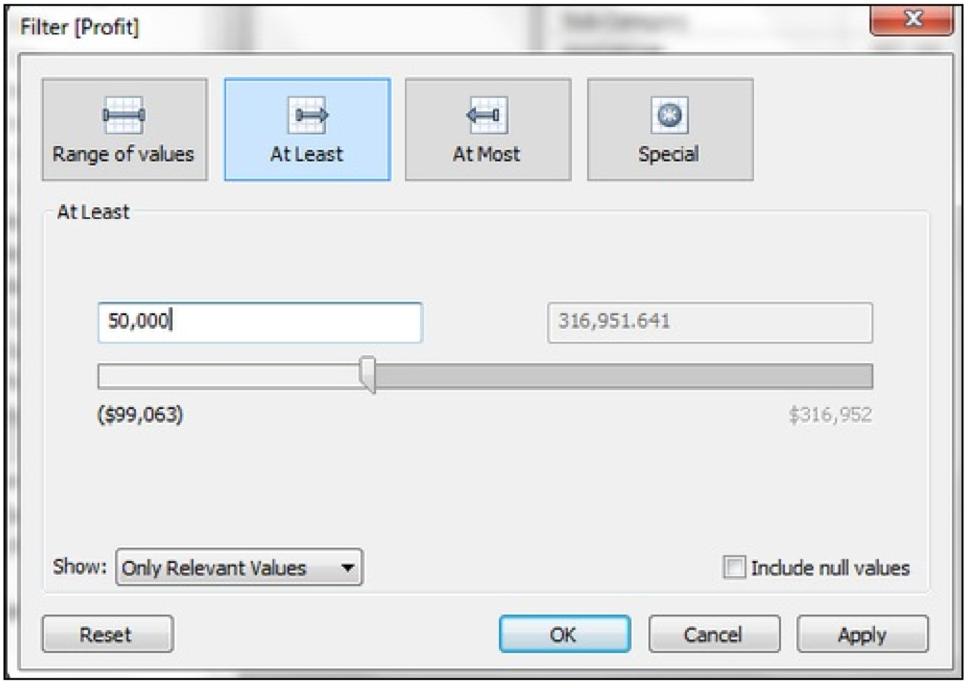
How to do it...

Once the sample file is loaded on the worksheet, perform the following steps to explore the **Filters** feature in Tableau:

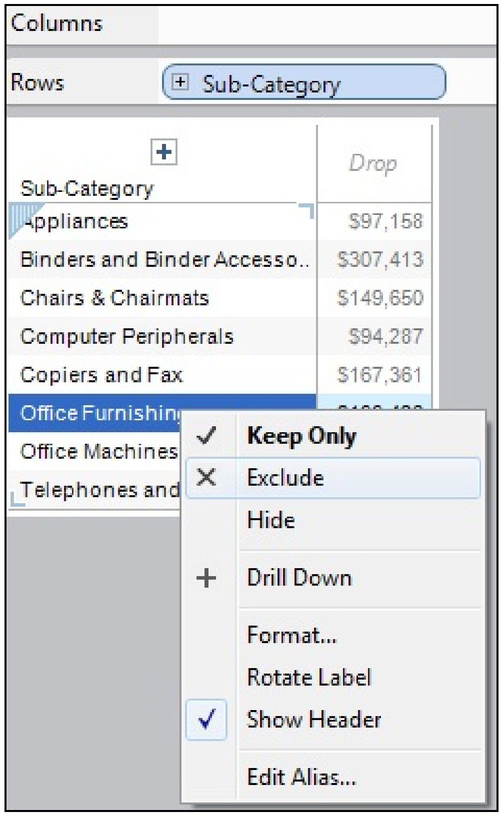
1. Drag-and-drop **Subcategory** from **Dimensions** into the **Rows** shelf.
2. Then drag-and-drop **Profit** from **Measures** into the **Text Marks** box. You can also right-click on **Profit** and click on **Add to Sheet**. Your worksheet should look like the following screenshot:



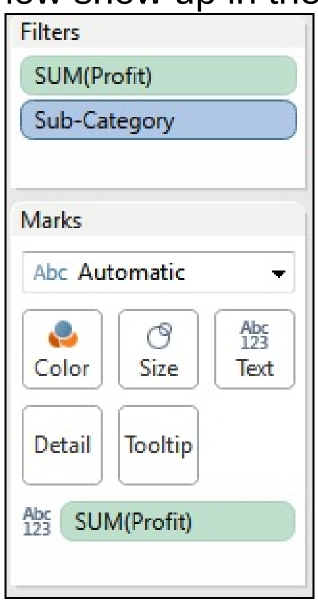
- If we want to see the subcategories that generated profit of more than $50,000, right-click on the **Text** box from the **Marks** pane box where it says **SUM(Profit)**, and click on **Filter**
- In the **Filter** dialog box, click on the **At Least** option and either drag the slide to **50,000** or type `50,000` in the input box and hit **OK**.



- Once the filter is applied, you'll see eight subcategories that generated profit of more than $50,000.
- We can continue to filter this information further. For example, to remove **Office Furnishings** as a subcategory, right-click on **Office Furnishings** and click on **Exclude**. Now only seven rows of subcategories are visible.



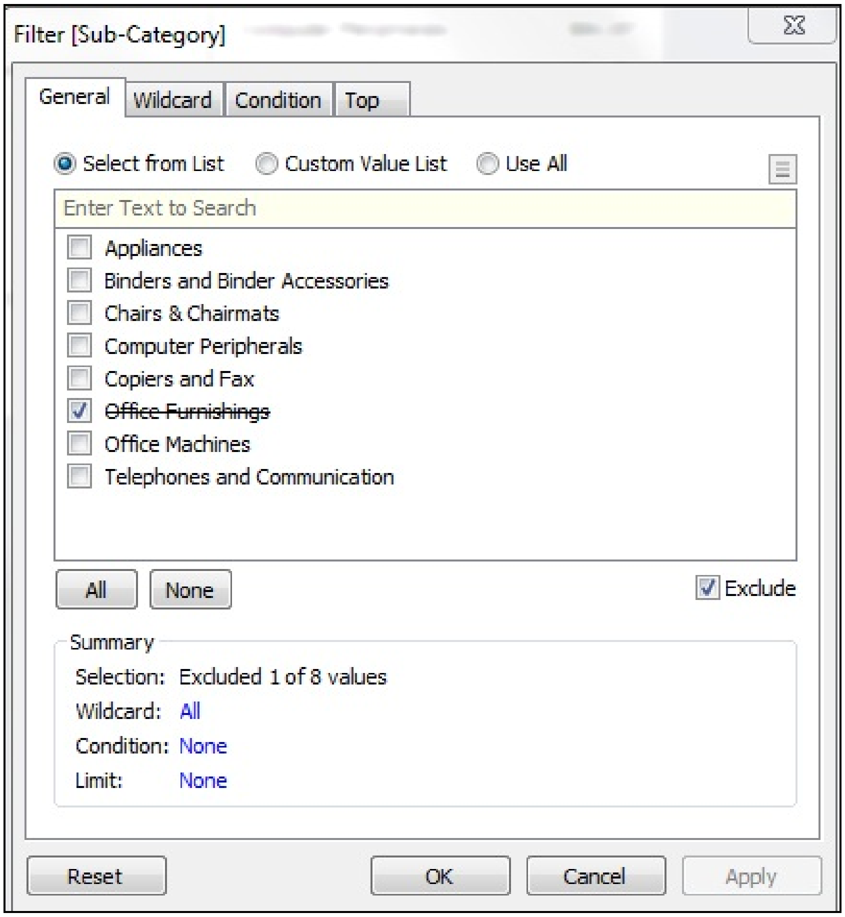
- Both the filters will now show up in the **Filters** pane, as shown in the



following screenshot:

- To remove the filters, right-click on the **Filters** pane and select **Clear Shelf**.
- To change a filter, right-click on the **SubCategory** filter from the **Filters** pane and select **Filter**.
- In the **Filter** window, you'll see tabs such as **General**, **Wildcard**,

Condition, and **Top**.



- In the **General** tab, you can type or select a value from the field. In the **Wildcard** tab, you can enter approximate string values to match certain patterns. In the **Condition** tab, you can enter conditions by the **Fields** or **Formula** values. In the **Top** tab, you can select the top *n* or bottom *n* items by a field or using a formula.

How it works...

Filters are a great way to manipulate the data on a worksheet. Depending on the field data type, various types of filters can be applied to a field. These filters can be numeric conditions to limit a numeric field or text patterns to limit a string field.

Merging multiple data sources

Often, our data is stored in different formats or different files. In relational databases, if two different tables have a common field, we can join these two tables with this field and pull the data in one single query. Tableau supports joins within a single data source connection; however, to merge multiple data source connections, Tableau uses a concept called **data blending**. In this recipe, we will look at how to blend two different data sources.

Getting ready

Download the following Google Spreadsheet, which contains the U.S. population by states, after signing in: http://bit.ly/12rUIh3

Download it as a CSV on your local hard drive and name it `USStatesPopulation.csv`.

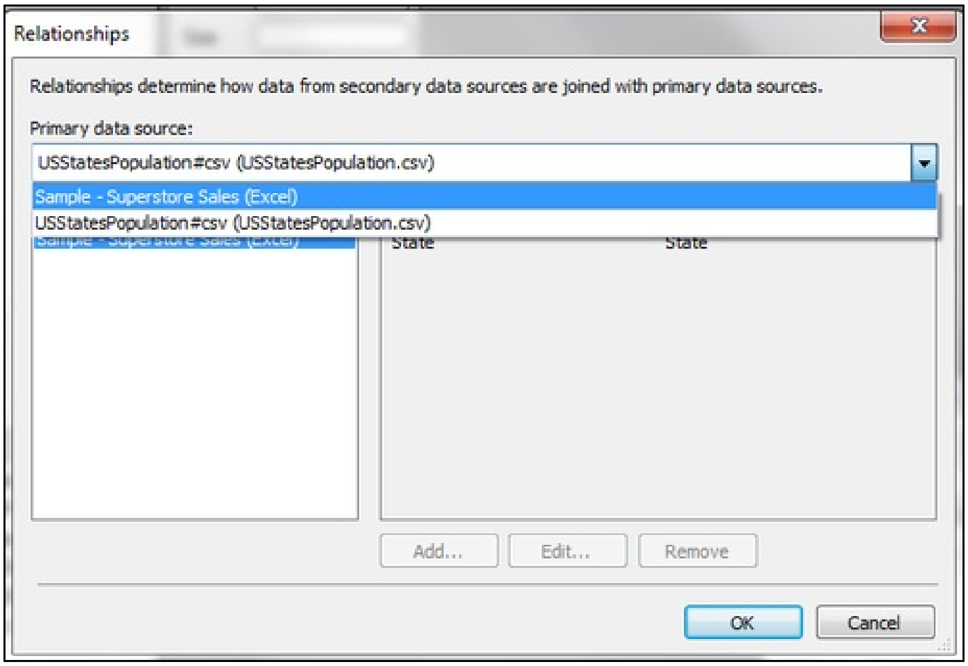
How to do it...

Once you have downloaded the CSV file, create a new worksheet in Tableau and perform the following steps to merge the CSV file and an Excel file:

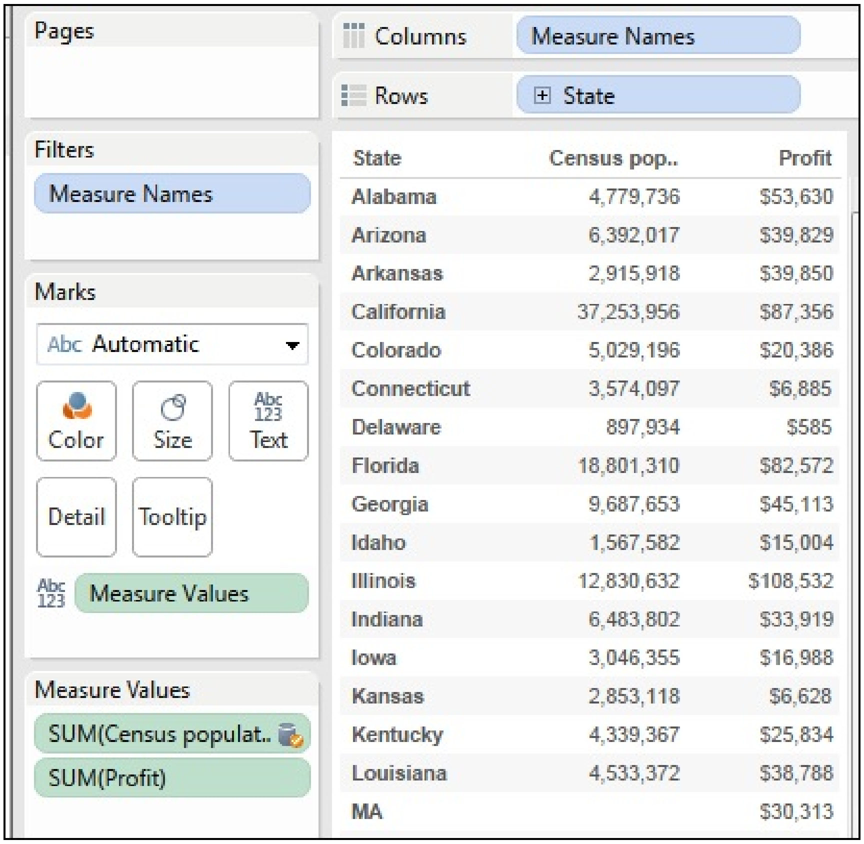
1. In a new workbook, connect to the **Sample – Superstore Sales (Excel)** data source.
2. Once the data is loaded and you can see **Dimensions** and **Measures** populated, click on **Connect to Data** in **Data** and select the text file `USStatesPopulation.csv`.
3. Accept all the defaults in the **Text File Connection** dialog box and hit **OK**.
4. Choose the **Connect Live** option in the next dialog box.
5. Tableau will match field names, and if it finds the same field names in both the data sources, it will create relationships between those common fields. To manually create relationships, click on **Data** and

select **Edit Relationships**.

6. In the **Relationships** dialog box, select **Sample – Superstore Sales (Excel)** as the **Primary** data source. Tableau will make `USStatesPopulation.csv` a secondary data source file.



7. Click on the **Custom** radio button and select **State** from the left-hand side column and **State** from the right-hand side column and hit **OK**.
8. To see profit by state, drag-and-drop the **State** value from the **Sample – Superstore Sales (Excel)** data source into the **Rows** shelf and the **Profit** measure into the **Text Marks** box.
9. Click on the **USStatesPopulation#csv** data source in the **Data** pane, and right-click on **Census population_April 1, 2010** from the **Measures** pane and select **Add to Sheet**.
10. As shown in the following screenshot, you should see three measure values in the **Measure Values** pane, **Measure Names** in the **Columns** shelf and **Census population_April 1, 2010** and **Profit** in

the datasheet: 

How it works...

Tableau can merge two or more different data sources in the same worksheet by creating relationships among common fields of these data sources. You can customize the blending operation by specifying the common fields in the data sources in the relationships. You should also note that this blending is different from joining two tables, because when we join tables, we create row-level joins and we can add fields from both the tables. Whereas, in blending, we merely show different fields from different data sources in a single visualization.

There's more...

Since the blending or merging of multiple data sources can prove challenging, it might be easier to understand this concept better by watching somebody actually do it. A YouTube user named *James Wright* uploaded a video of blending data at http://youtu.be/-G0lIz7y6y0.

Chapter 2. Creating Univariate Charts

The recipes covered in this chapter are:

- Creating tables
- Creating bar graphs
- Creating pie charts
- Sorting the graphs
- Creating histograms
- Creating line charts
- Using the Show Me toolbar
- Creating stacked bar graphs
- Creating box plots
- Showing aggregate measures
- Showing the top 10 items

Introduction

An analysis involving one measure is called **univariate analysis**. In this chapter, we will cover various univariate charts. Note that charts are also referred to as graphs or plots, and you will find these terms used interchangeably throughout this book.

Creating tables

Sometimes the best way to present the data is using a table. Tables use very little space and pack a lot of information in a very small area without losing any detail. Please note that, in Tableau, we refer to tables as cross tabs as well.

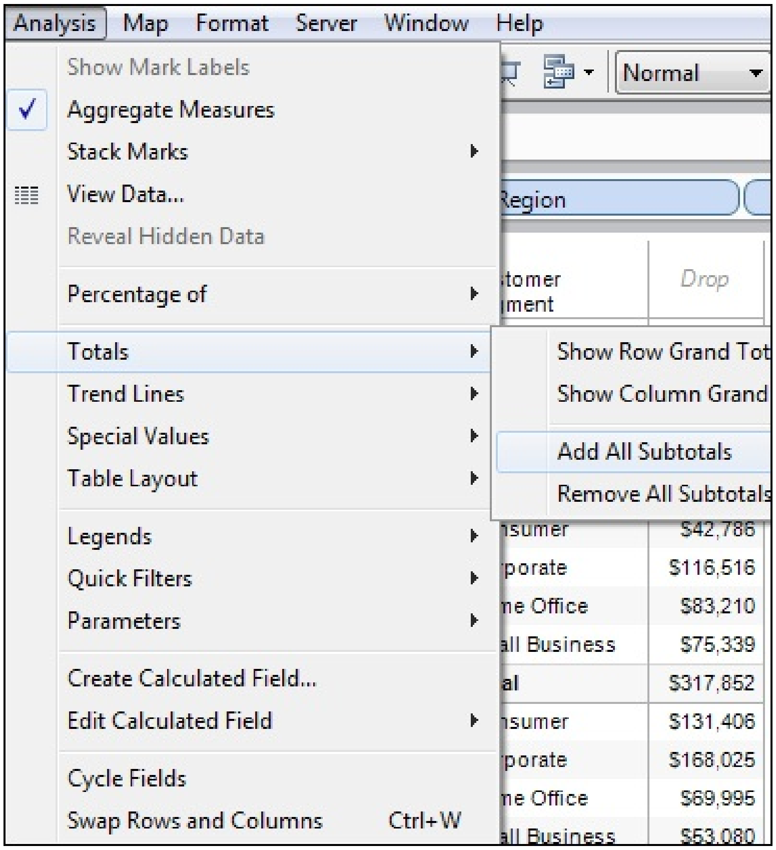
Getting ready

For this recipe, let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Although we can create complicated tables, in this recipe let's create a simple table with two dimensions and one measure.

1. Drag-and-drop **Region** into the **Rows** shelf (you can also right-click and select **Add to Sheet**).
2. Drag-and-drop **Customer Segment** into the **Rows** shelf.
3. Drag-and-drop **Profit** into the **Text** marks box. Now, you should see the total of profits as per **Region** and **Customer Segment**.
4. To add totals, click on **Analysis** from the toolbar and select **Totals** and click on **Add All Subtotals**, as shown in the following

screenshot: 

How it works...

To a graph designer, using tables instead of charts is a strong alternative as tables can provide very minute details of the data to the reader. Tables used with other graphs can create a compelling narrative for the reader, and Tableau makes it easy to create and combine tables and graphs by creating dashboards, which are covered later in the book.

Creating bar graphs

As bar graphs are very easy to understand, they are the most common type of graphs. The graphs that have a horizontal orientation are called **bar graphs** and the graphs that have a vertical orientation are called **column graphs**. The length of the bar represents the quantity of a particular measure. They are best used with categorical information, such as gender, state, regions, countries, business types, and others. One very important thing to note with the column bar charts is that this type of chart's y axis must always start at zero, otherwise it is very difficult and misleading to encode the length of the bar to a measure.

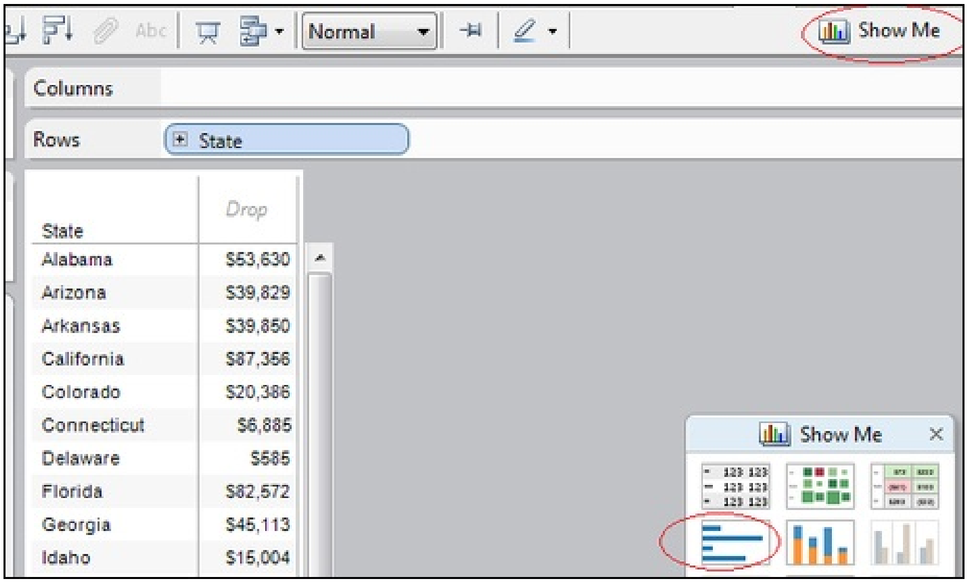
Getting ready

Let's use the sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to create a simple bar graph:

1. Drag-and-drop **State** into the **Rows** shelf.
2. Drag-and-drop **Profit** into the **Columns** shelf.
3. You should see the **Profit** totals by **State** in the **Columns** area in plain numbers.
4. To change the text representation to a bar graph, click on the **Show Me** button on the toolbar (shown in the following screenshot).
5. From the **Show Me** toolbar, select **horizontal bars** as shown in the following screenshot:



How it works...

By changing the default behavior of showing text tables to showing bar charts, we can make it easier for the reader to compare various measures by comparing the lengths of the different bars. In addition, since bar charts are common in various publications, the reader is adept at understanding bar charts.

Creating pie charts

Pie charts and their variations are one of the most controversial types of charts. Many experts in the information graphics and information visualization fields have warned against the use of pie charts but they are still quite common in business presentations as well as reports. There are a few key things to consider while creating such a graphic:

1. Limit the number of *slices* to three to four. In addition, slices must be large enough for easy differentiation.
2. Limit the use of color (if there are only three to four slices to show, it is easy to use one color with different hues). Colors are better used only for differentiating one item from another and not for decorating. If many colors are used, the reader faces the difficulty of distinguishing items by color.
3. Start the largest slice at 12 o'clock and move to the right; next to it on the left must be the next biggest slice. The smallest slice should be close to the bottom. This helps the reader see the bigger slices and make comparisons easily.
4. Do not use any 3D pie charts as they make matters worse. Tableau saves the users from this trouble as 3D pie charts are not supported.
5. Do not allow a sliced pie section to be sliced further, as it distracts the reader. Again, Tableau does not support *exploding* a slice.
6. Make sure that the pie totals up to 100 percent of the measure; any other total will render the pie meaningless. For example, if you want to plot the profit by store location, the pie should total to the total profit by *all* stores; that is, 100 percent of the profit.
7. Avoid excessive labels on the slices. Although labels will remove any guessing by the reader, if you label all the slices, then perhaps it is better if you replace the chart with a table.

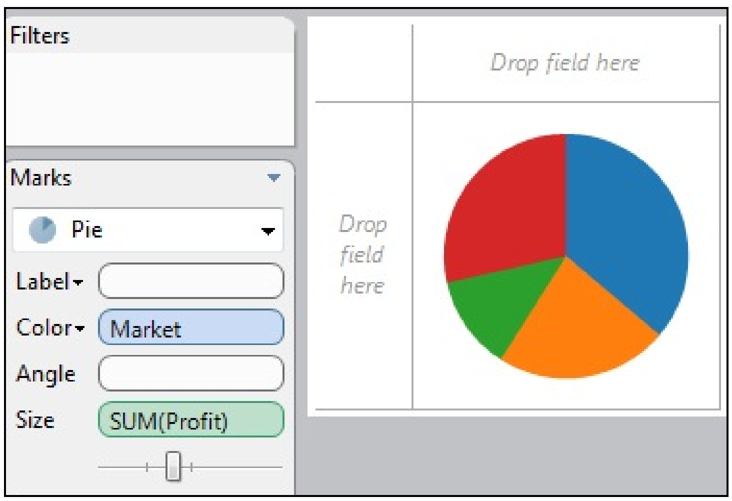
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to create a simple pie chart:

1. Drag-and-drop **Profit** into the **Size** box in the **Marks** pane.
2. Drag-and-drop **Market** into the **Rows** shelf.
3. Click on the **Show Me** button to make the toolbar visible.
4. Click on the pie chart icon to create the pie chart.
5. Right-click on **SUM(Profit)** in the **Angle** box in the **Marks** pane and hit **Remove**.
6. To make the chart bigger, click on the zoom slider that can be seen after clicking the **Size** mark box.



How it works...

Tableau's settings for default colors for pie charts as well as the setting for the placement of the various slices creates good-looking and effective pie charts; however, by tweaking colors and adjusting sizes, we can make the charts even better. The topic of changing color palettes is covered later in this book.

Sorting the graphs

Although Tableau generates default graphs using the best practices in information visualization, often they need modification for meeting the business needs and sometimes for better representation. Tableau provides various ways to adjust and modify various aspects of the graph. Sorting is useful to display the most or least influential number or category at the top or bottom.

Getting ready

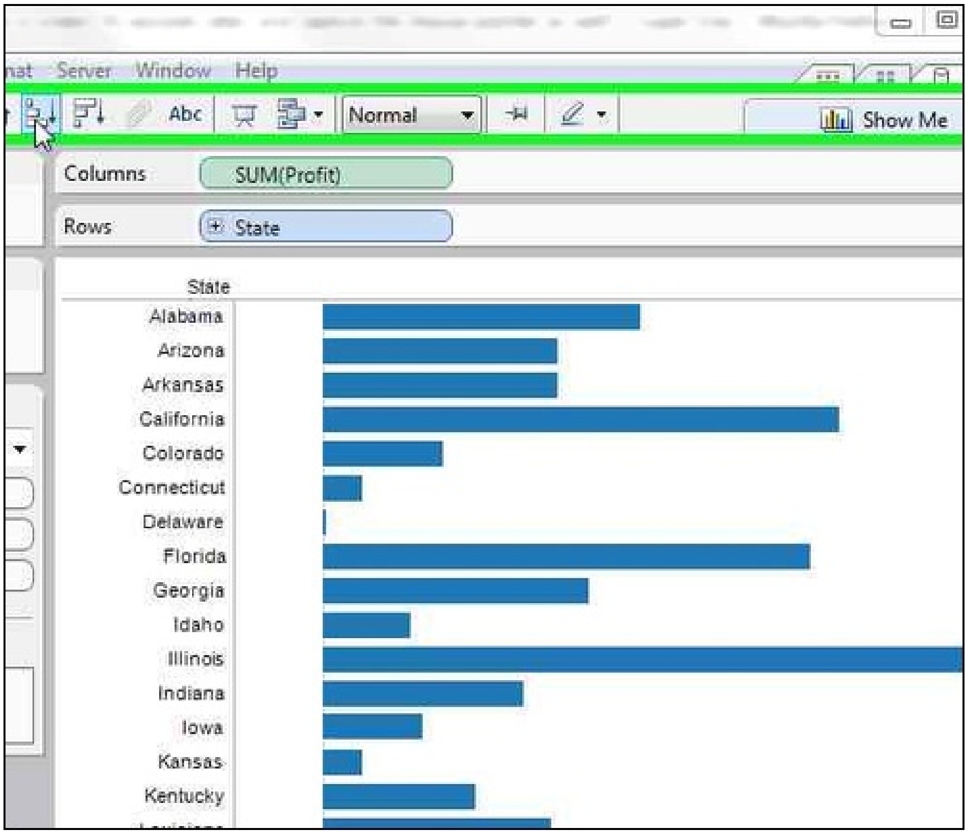
Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to create and sort a simple bar graph:

1. Drag-and-drop **State** into the **Rows** shelf.
2. Drag-and-drop **Profit** into the **Columns** shelf.
3. You should see the **Profit** totals by **State** in the **Columns** area in plain numbers.
4. To change the text representation to a bar graph, click on the **Show Me** button on the toolbar (shown in the following screenshot).
5. From the **Show Me** toolbar, select **horizontal bars** (also shown in the following screenshot).
6. Once the bar chart is created, click on the **Sort** button, which has the shortest bar on the top, on the toolbar (again, shown in the following screenshot) to show the least profitable **State** value up at the top.
7. You'll see that **Montana** is the least profitable state with **$9,127** in profit. To see the most profitable **State** value, click on the **Sort** button (this button has the longest bar at the top) to the right.
8. You'll see that **Illinois** is the most profitable state with **$108,532** in

Profit.



How it works....

Since sorted bar charts do not require the additional tasks of identifying the longest bar and comparing lengths of various bars, they are better and more effective than unsorted bar charts.

Creating histograms

Histograms show counts or density of a measure, which is then discretized (binned) to make counting meaningful. They are best used to observe the distribution of the measure. They are sometimes confused with plain bar charts, which can be modified to show counts but usually encode the measure value as the length of the bar.

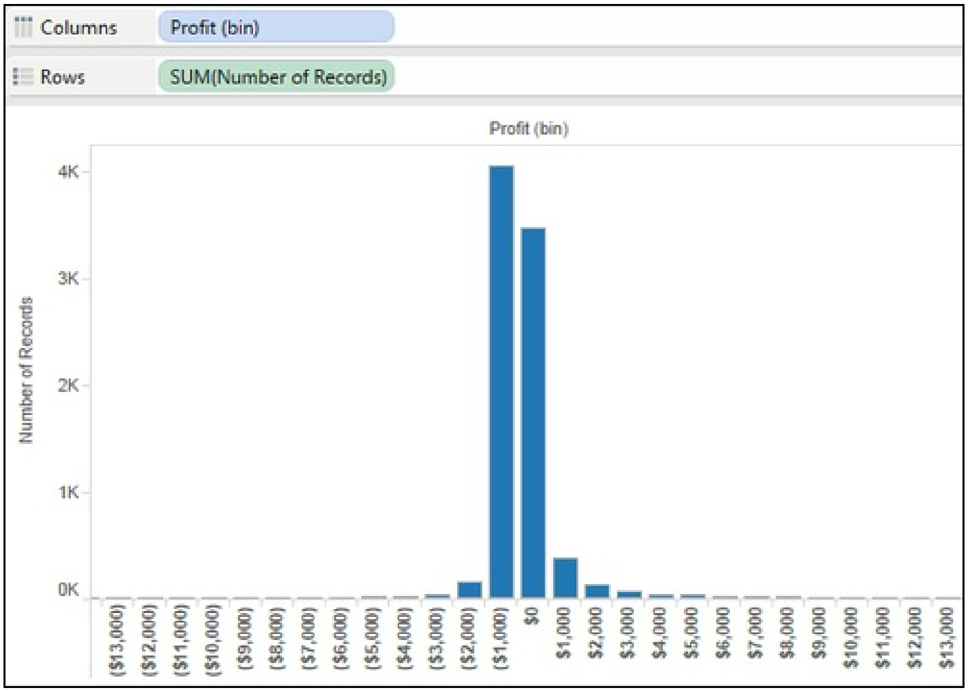
Getting ready

Let's use the sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to create a histogram:

1. Drag-and-drop **Profit (bin)** from the **Dimensions** pane into the **Columns** shelf.
2. Drag-and-drop **Number of Records** from the **Measures** pane into the **Rows** shelf.
3. You should see **SUM(Number of Records)** in the **Rows** shelf now and also a histogram with a very narrow distribution, as shown in the following screenshot:



How it works...

Histograms are very effective charts in observing the distribution of the measure of interest; however, sometimes the distribution is quite skewed or centered at one range or position, and in such cases rebinning (creating different bins is covered later in Chapter 6, *Calculating User-defined Fields*, in the *Discretizing data* recipe) assists the reader in observing the detailed distribution or uncovering some patterns.

Creating line charts

Although line charts are best used for time-series data to observe trends by various time units, such as day, week, month, quarter, and year, they could be used for other types of data as well; however, the ups and downs in the lines themselves are less important in such cases.

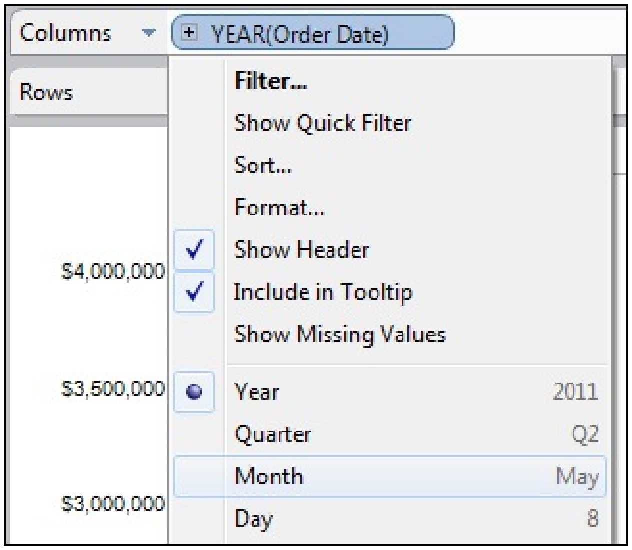
Getting ready

Let's use the sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

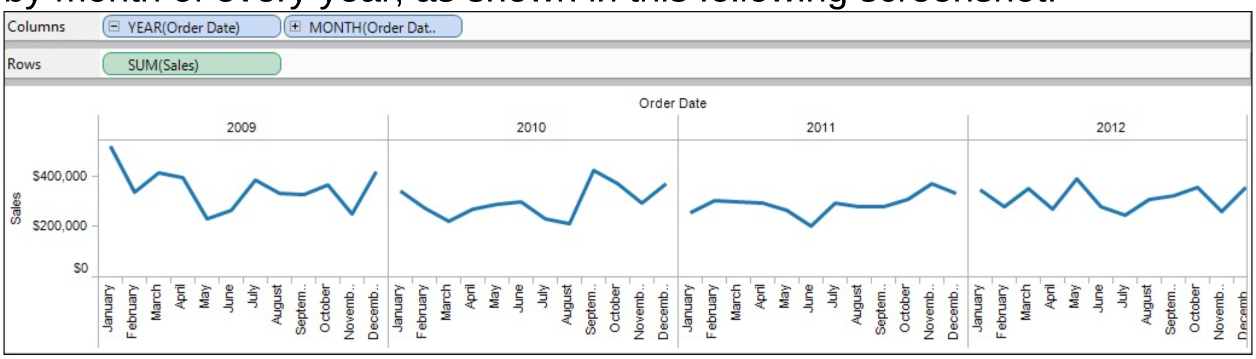
How to do it...

Once the data is loaded, perform the following steps to create a line chart:

1. Drag-and-drop **Order Date** into the **Columns** shelf. Tableau, by default, will show **YEAR(Order Date)**.
2. Drag-and-drop **Sales** into the **Rows** shelf.
3. To observe the trends by month of orders, click on **YEAR(Order Date)**, which is shown in the **Columns** shelf, and click on **Month**, as shown in the following screenshot:



- Drag-and-drop **Order Date** into the **Columns** shelf again, but this time in front of the **MONTH(Order Date)** field. This change will show the trend by month of every year, as shown in this following screenshot:



How it works...

Line charts are quite effective in representing trends over time. These trends, however, could be misrepresented if improper zoom level, axis scale units, or aspect ratios are used. For example, if a reader is observing a line chart that has hour as the unit, the reader may think that the observed measure fluctuates quite often. But if the axis unit is changed to months, the lines will show trends over a longer time period and will not show major fluctuations, as shown in the chart with hour as the axis unit. The chart designer should carefully select the unit of time for the x axis.

Using the Show Me toolbar

The **Show Me** toolbar is one of the most powerful features of Tableau. It removes many steps required to create a graphic and automatically determines the axis location of the variables used.

Getting ready

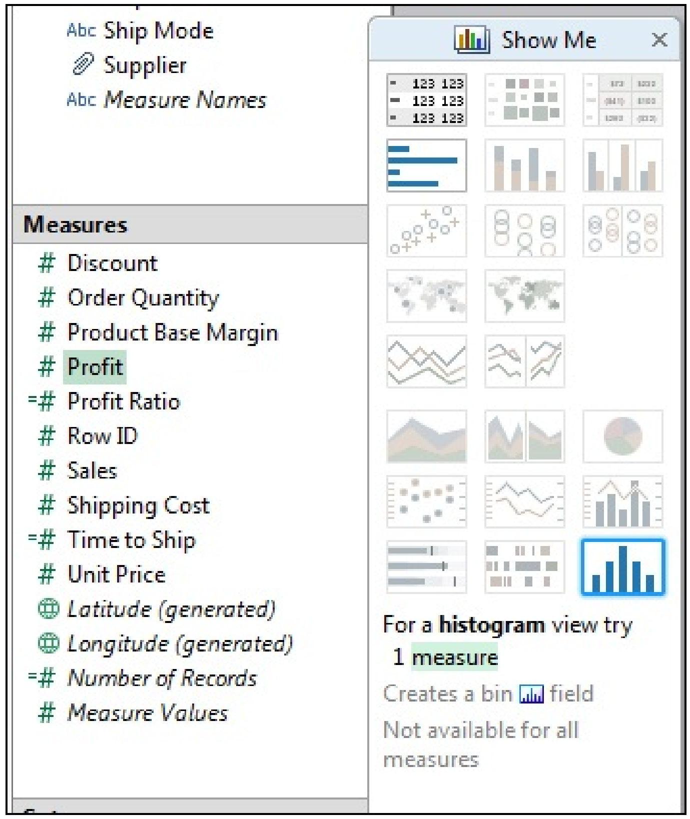
Let's recreate the histogram from an earlier recipe using the **Show Me** toolbar. We will use the same sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to reproduce the histogram from an earlier recipe using the **Show Me** toolbar:

1. If the **Show Me** toolbar is not visible, click on the **Show Me** button to make the toolbar visible.
2. Click on **Profit** from **Measures**.
3. While **Profit** from **Measures** is still highlighted, click on the histogram graph button from the **Show Me** toolbar as shown in the

following screenshot:



How it works....

As you can see by comparing the number of steps listed in this recipe with the number of steps listed in the earlier histogram recipe, the **Show Me** toolbar reduces the effort required to produce the same graphic. Similarly, you can generate various types of graphs in two clicks. The **Show Me** toolbar also lists the number of measures and the number of dimensions required to create a graphic. If a graphic cannot be drawn with the selected measures or dimensions, the graph button is shaded gray (disabled) to indicate unavailability of the graphic.

Creating stacked bar graphs

In stacked bar graphs, various categories of the same field are plotted on top of each other. One of the biggest problems with the stacked bar graphs is that the length of the bars is hard to measure, except for the bottom bar in the stack. Some people argue that it is good at showing the proportion or comparison of two or more categories; however, if comparison is the objective, there are much better alternatives, such as facets and small multiples, which have higher efficacy in comparing data. If you must use a stacked bar graph, limit the number of stacks to two to three categories and avoid very disproportionate stacks, such as 99 percent and 1 percent stacks.

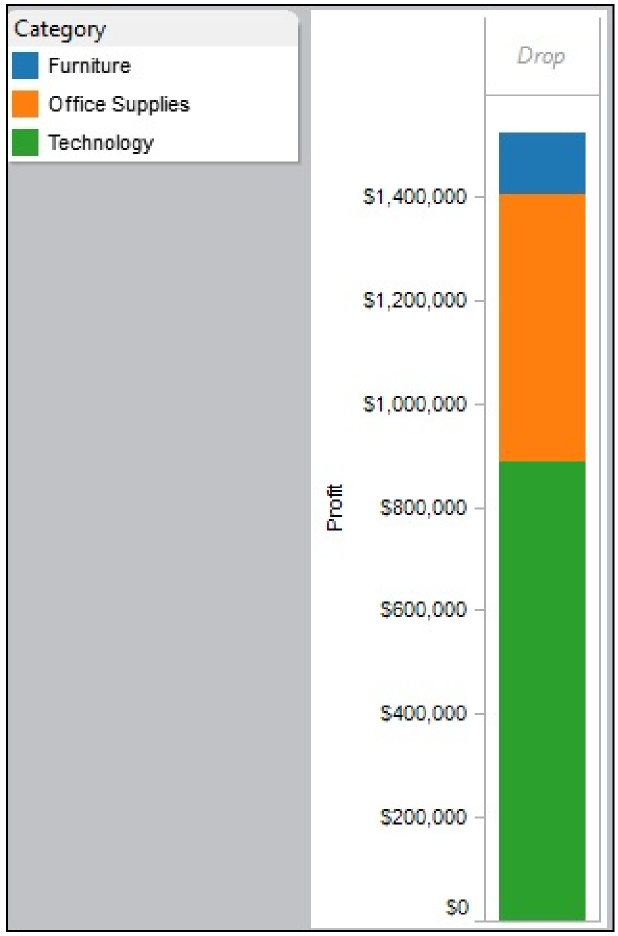
Getting ready

Let's use the sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to create a simple, one-variable stacked bar graph:

1. Drag-and-drop **Profit** from **Measures** into the **Rows** shelf.
2. Drag-and-drop **Category** under **Products** from **Dimensions** into the **Color** box from the **Marks** pane.
3. Change the **Marks** type to **Automatic**. The resulting graph is shown

in the following screenshot: 

How it works...

Since we use **Color** to encode the **Category** variable, Tableau automatically assigns the default colors to various **Category** types and, in effect, produces the stacked bar graph. You could also select **Category** from **Dimensions** and **Profit** from **Measures** and click on the stacked bars graph button on the **Show Me** toolbar to create the same graph. As you can see from the graph, the reader has to look at the legend colors to distinguish various **Category** types, thus increasing the difficulty in understanding the information presented. An alternative to this type of chart is creating multiple charts or facets. To create such facets, in this recipe, we can drag-and-drop **Category** into the **Rows** shelf before **Profit**.

Creating box plots

Box and whisker plots, also known as box plots, show the distribution of the observed measure. This distribution includes the 25th, 50th, and 75th percentile as well as the minimum and the maximum values of the measure. A box surrounds the interquartile values of the 25th, 50th, and 75th percentile, and whiskers represent the minimum and the maximum values. Since Tableau does not support creating a box plot directly, this recipe is a workaround to create a box plot in Tableau.

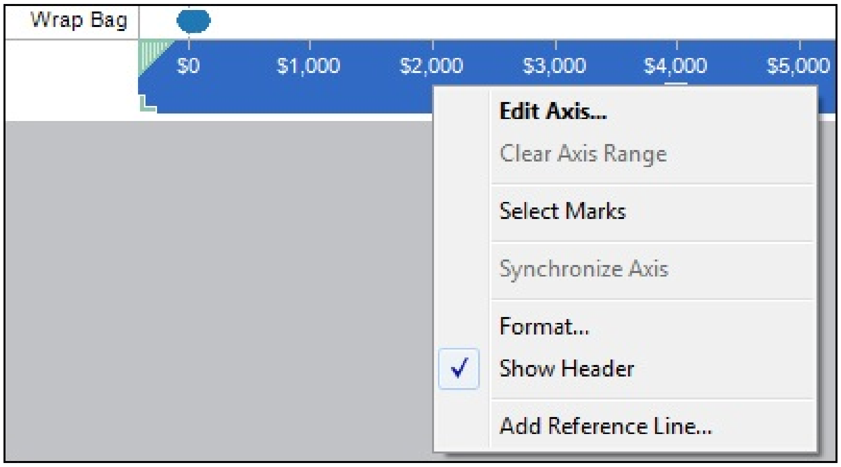
Getting ready

Let's use the sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

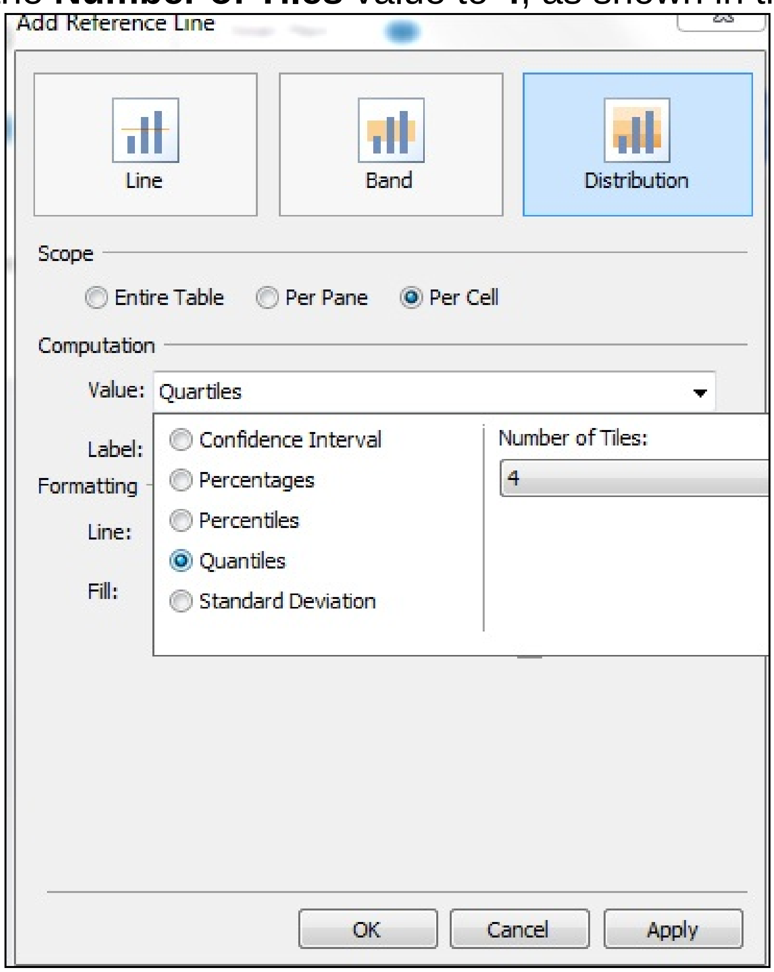
How to do it...

Once the data is loaded, perform the following steps to create a box plot of shipping cost by container types:

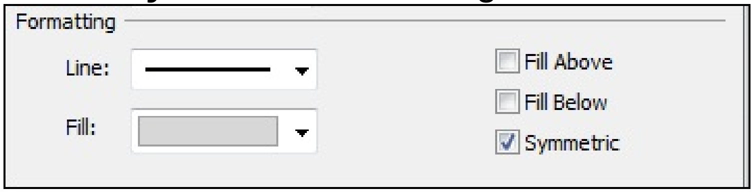
1. Drag-and-drop **Container** from **Dimensions** into the **Rows** shelf.
2. Drag-and-drop **Unit Price** from **Measures** into the **Columns** shelf.
3. Click on the **Analysis** menu option from the top toolbar and uncheck **Aggregate Measures** to remove aggregation.
4. Right-click on the x axis and click on **Add Reference Line**, as shown in the following screenshot:



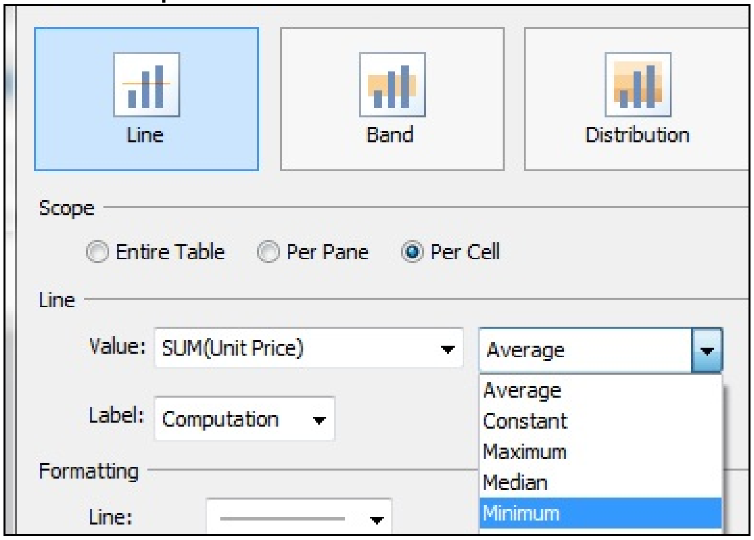
- Click on the **Distribution** pane in the **Add Reference Line** options box.
- Select the **Per Cell** button under **Scope**.
- Under **Computation**, in the **Value** dropdown selection, select **Quantiles** and keep the **Number of Tiles** value to **4**, as shown in the following screenshot:



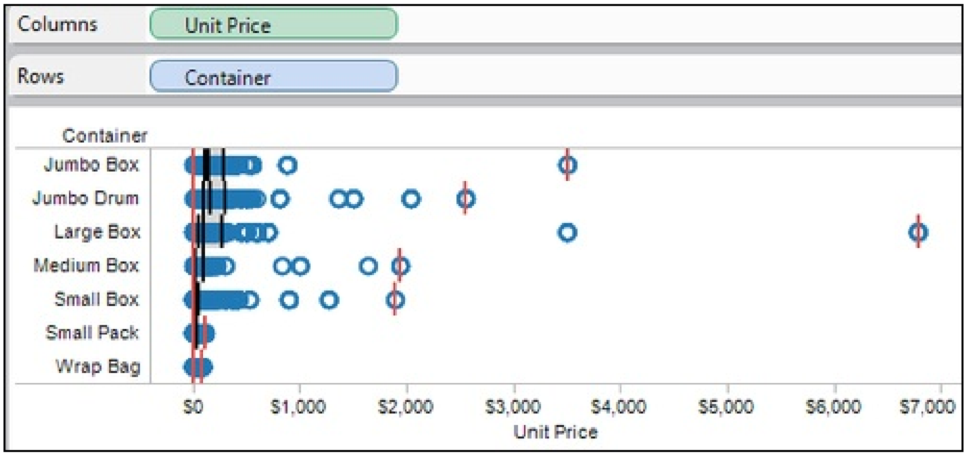
- Under **Computation**, in the **Label** dropdown selection, select **None**.
- Under **Formatting**, in the **Line** dropdown selection, select the first thick and solid line.
- Under **Formatting**, check the **Symmetric** formatting box as shown in the following screenshot:



- Under **Formatting**, in the **Fill** dropdown selection, keep the default gray color, which is in the first column and fourth from the top.
- Hit **OK**.
- To add whiskers for the minimum values, right-click on the x axis and click on **Add Reference Line**.
- Keep the **Line** pane selected.
- Select the **Per Cell** option under **Scope**.
- Under **Line**, in the **Value** dropdowns, select **Minimum** as shown in the following screenshot:



- Under **Line**, in the **Label** dropdown, select **None**.
- Under **Formatting**, in the **Line** dropdown, change the line color to red.
- Hit **OK**.
- Follow steps 13-19 to add the maximum whiskers, and instead of selecting **Minimum**, select **Maximum** in step 16. The final box plot should look like the one in the following screenshot:



How it works...

Although Tableau does not provide a quick way to create box plots, adding reference lines is a very powerful feature that can be used to create box plots. Adding reference lines can be very useful for the reader to observe trends, distributions, and variance. In the case of box plots, we added reference distributions using quantiles (also known as quartiles) and added minimum-maximum lines.

Showing aggregate measures

Tableau, by default, aggregates measure values, and this behavior can be changed to show all individual values of the measures by clicking on the **Analysis** menu option from the top toolbar and unchecking **Aggregate Measures** to remove aggregation. It is also possible to change the aggregation type, such as total, average, variance, and others.

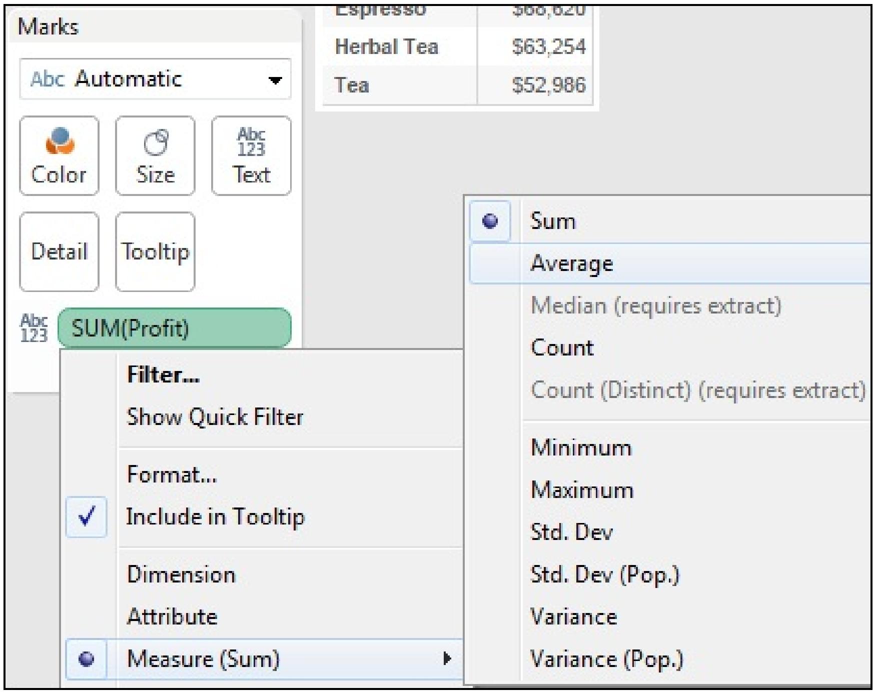
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

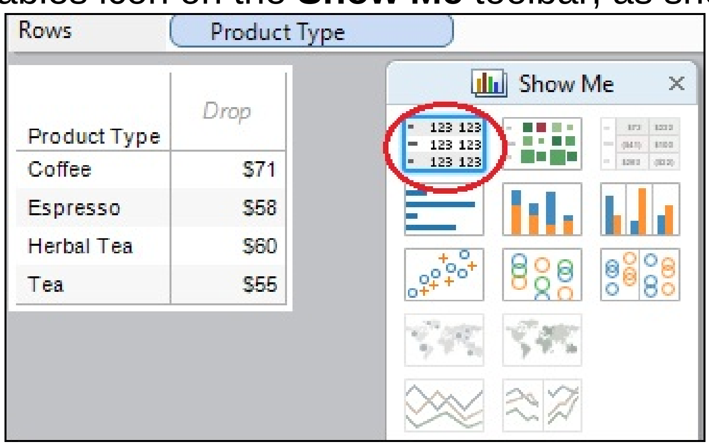
How to do it...

Once the data is loaded, perform the following steps to change and add various aggregate measures:

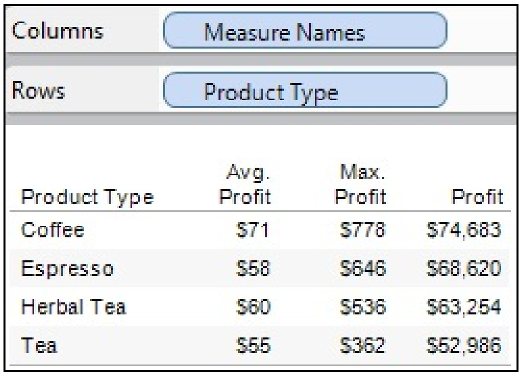
1. Make sure the **Aggregate Measures** option is checked under the **Analysis** menu option on the top toolbar.
2. Drag-and-drop **Product Type** from **Dimensions** into the **Rows** shelf.
3. Drag-and-drop **Profit** from **Measures** into the **Text** input box under **Marks**.
4. To view the average profit by **Product Type**, click on **SUM(Profit)** in the **Text** input box, expand **Measure (Sum)**, and select **Average**, as shown in the following screenshot:



- Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
- To add total profit by **Product Type**, click on **Profit** from **Measures** and click on the text tables icon on the **Show Me** toolbar, as shown in the following screenshot:



- To add the maximum profit value by **Product Type**, click on **SUM(Profit)** under **Measure Values**, expand **Measure (Sum)**, and select **Maximum**.
- To add the total profit by **Product Type** again, drag-and-drop **Profit** from **Measures** into the **Measure Values** pane. Once all the aggregate values are added, the table should look like the one in the following

screenshot: 

Product Type	Avg. Profit	Max. Profit	Profit
Coffee	$71	$778	$74,683
Espresso	$58	$646	$68,620
Herbal Tea	$60	$536	$63,254
Tea	$55	$362	$52,986

How it works...

Adding various aggregate measures of the same measure is somewhat counterintuitive and is hardly straightforward. Since Tableau allows one type of aggregation only once, users must change the aggregation type of the already displayed aggregation and add the measure again. You could also create duplicate copies of the measure and add the new measure with a different aggregation, but you will have to rename the measures.

Showing the top 10 items

At times, it is just easier to view the top 10 items by a certain measure, such as the top 10 most profitable customers or the top 10 least expensive vendors, rather than viewing all the items of a field. This approach, although easier, must be used with caution since anomalous items or patterns could be missed by viewing only the top *n* or bottom *n* items.

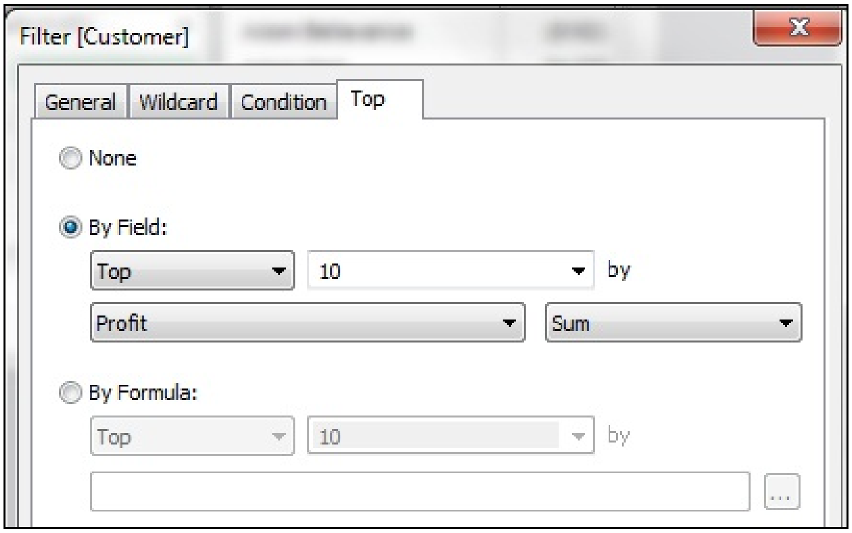
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

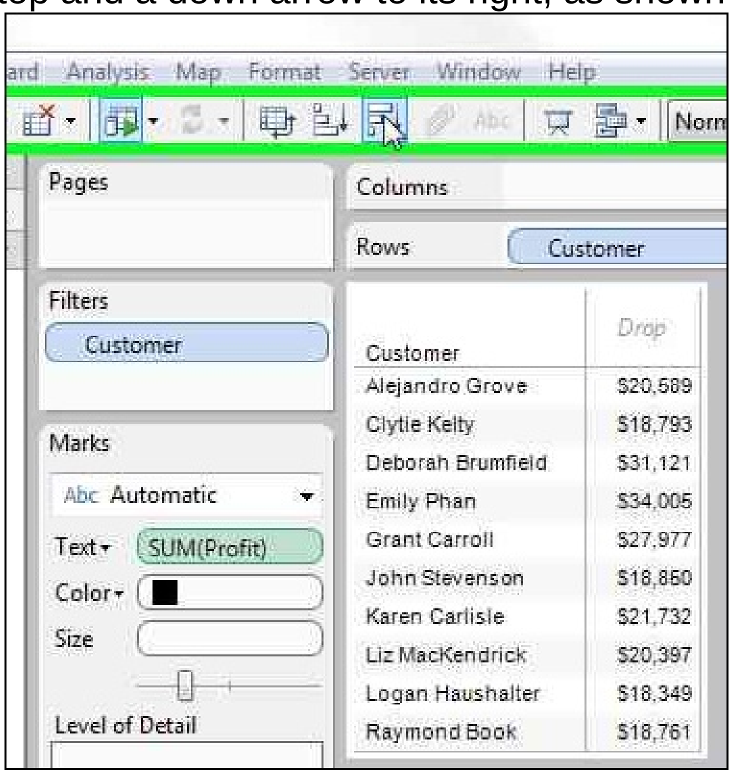
How to do it...

Once the data is loaded, perform the following steps to view a table with the top 10 customers by total profit:

1. Drag-and-drop **Customer** from **Dimensions** into the **Rows** shelf.
2. Drag-and-drop **Profit** from **Measures** into the **Text** input box under **Marks**.
3. Drag-and-drop **Customer** from **Dimensions** into the **Filters** pane.
4. Click on the **Top** tab in the **Filter [Customer]** options box.
5. Check the **By Field** option.
6. Make sure your options (top 10 by **Profit** and aggregation is **Sum**) look like the one in the following screenshot:



- Click on **OK**.
- To sort the customer list in descending order of profit (that is, the most profitable customer up at the top), click on the **Sort** button that has the longest bar up at the top and a down arrow to its right, as shown in the following screenshot:



How it works...

Tableau not only provides filtering by item names, but also provides filtering by aggregate measures such as limiting to top or bottom items by sum, average, and other aggregations.

Chapter 3. Creating Bivariate Charts

An analysis involving two measures is called **bivariate analysis**, and in this chapter we will cover various bivariate charts.

We will be covering the following topics:

- Creating tables
- Creating scatter plots
- Swapping rows and columns
- Adding trend lines
- Selecting color palettes
- Using dates

Introduction

This chapter provides recipes for generating visualizations when using two measures. Such visualizations can help a user with formulating questions that can be answered using data. There are other recipes that manipulate existing data to generate alternative visualizations, such as swapping rows and columns and using color palettes. This chapter also explains how to add trend lines to existing visualizations to extend the effectiveness of a chart.

Creating tables

If you want to present any data with all the details, tables often are a good choice as they retain all the information and reduce the chances of misrepresentation of data. Tables are also effective in presenting data with precision. For example, a reader might get confused by a value of 100.8 in a chart to be 100.5 or 101, but in a table, all values are presented accurately and there is no scope for misinterpretation. Tables are great for smaller number of columns or rows but charts are better suited for complex information.

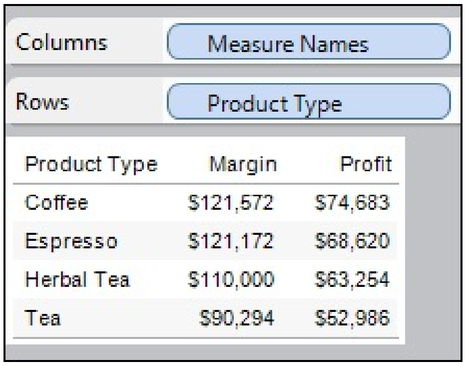
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it....

Once the data is loaded, perform the following steps to create a table with one dimension and two measures:

1. Drag-and-drop **Product Type** from **Dimensions** into the **Rows** shelf.
2. Click on **Margin** and **Profit** from **Measures**.
3. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
4. Click on the text tables icon in the **Show Me** toolbar. Your table should look like the one in the following screenshot:



How it works....

As we click on the two measures and then click on the text tables icon in the **Show Me** toolbar, Tableau will automatically create filters on **Measure Names** to limit the measures to **Margin** and **Profit**, and it will also put the **Measure** names in the **Columns** shelf. By default, Tableau will total the measure value and this can be changed by clicking on the **Measure Values** shelf and changing the aggregate measure to view individual values.

There's more...

You can read up on good arguments when to use tables and when not to use them on the University of Leicester's page: http://www2.le.ac.uk/offices/ld/resources/numeracy/numerical-data

Creating scatter plots

Scatter plots are often used to identify any correlation or observe relationships between two variables. By looking at these plots, the reader can quickly observe any trends, if present. A scatter plot is a very useful tool in any analyst's toolbox.

Getting ready

For this recipe, let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it....

Once the data is loaded, perform the following steps to create a scatter plot of two measures:

1. From the top toolbar, under **Analysis**, uncheck **Aggregate Measures**.
2. Drag-and-drop **Profit** into the **Columns** shelf.
3. Drag-and-drop **Sales** into the **Rows** shelf. The generated scatter plot should look like the one in the following screenshot:

How it works...

By default, Tableau will aggregate measures to show only the aggregated values. In traditional statistics, however, to observe any trends or correlation between two variables, individual data points are plotted across both axes. Therefore, we removed the aggregation for this recipe, but please note that some applications of scatter plots may warrant aggregation.

There's more...

Scatter plots are one of the most common techniques to observe the relationship between two variables. It is important to note, however, that a plot may suggest a correlation between two variables but cannot conclusively prove a causal relationship. You can read more about scatter plots on the **National Institute of Standards and Technology** (**NIST**) exploratory data analysis handbook at http://www.itl.nist.gov/div898/handbook/eda/section3/scatterp.htm.

Swapping rows and columns

Sometimes the data points are located in undesirable locations, which makes looking at the numbers slightly challenging. By swapping rows with columns, we can offer a different point of view to the reader.

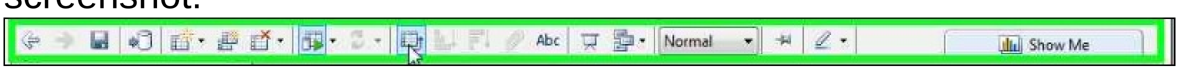
Getting ready

For this recipe, let's use the same sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, we will follow the steps from the previous recipe to reproduce the scatter plot and then swap the rows with columns using the following steps:

1. From the top toolbar, under **Analysis**, uncheck **Aggregate Measures**.
2. Drag-and-drop **Profit** into the **Columns** shelf.
3. Drag-and-drop **Sales** into the **Rows** shelf.
4. Click on the swap button to place **Sales** in the **Columns** shelf and **Profit** in the **Rows** shelf. The swap button is shown in the following screenshot:



How it works...

Swapping of rows with columns and columns with rows works with almost any type of chart and it is a very useful tool when we want to quickly change the orientation or position of the visualization. Tableau makes it very easy to make such a change.

Adding trend lines

Trend lines are very useful in observing the relationship between two variables as well as predicting future values. Trend lines are frequently used in simple linear regression to observe the relationship between two variables. The shape of the trend line explains the type of the relationship between the variables. For example, in the case of simple linear regression, the trend line is a straight line, which is represented by the mathematical equation of a straight line: $y = mx + c$.

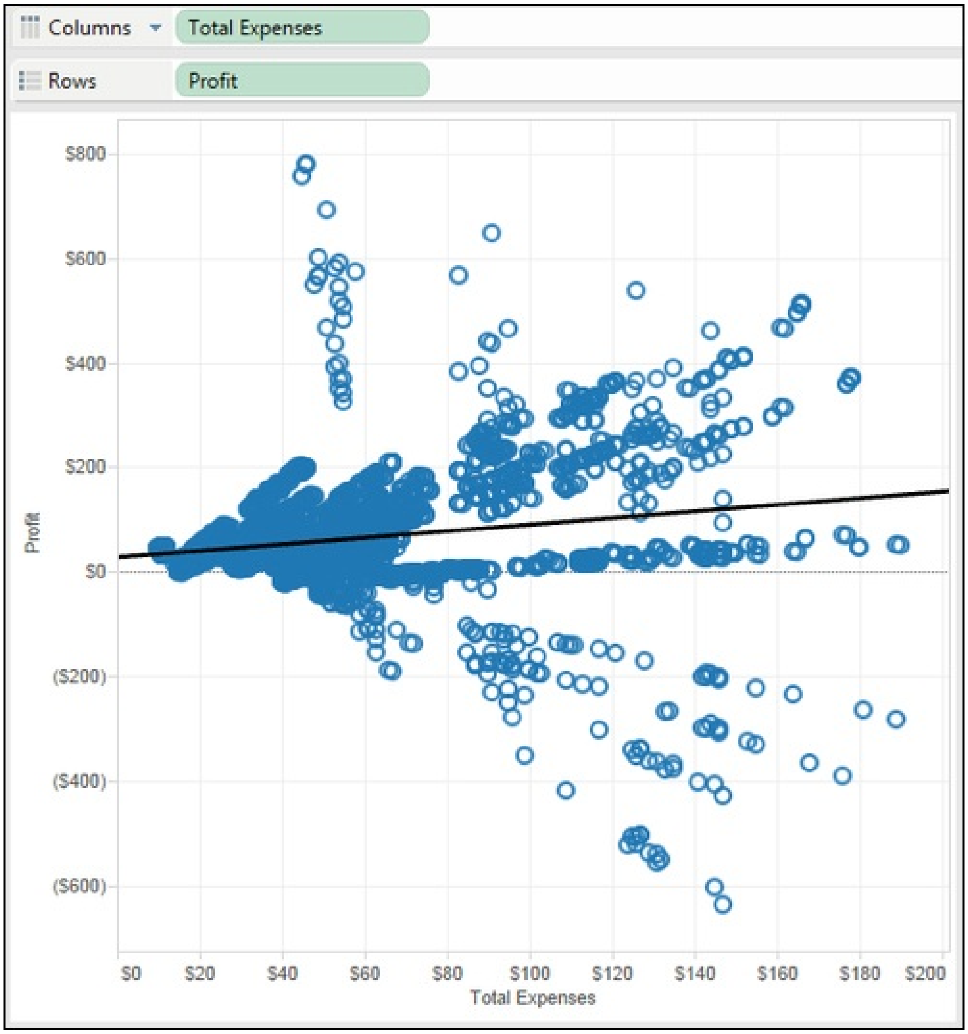
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

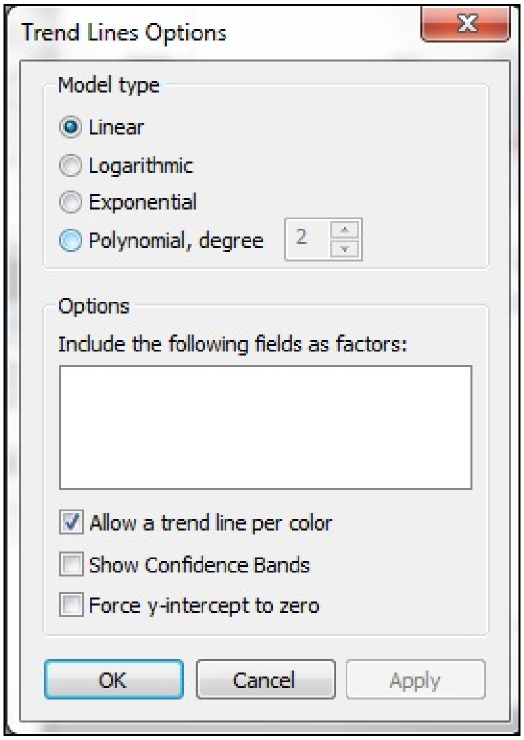
Once the data is loaded, perform the following steps to add a trend line to a plot:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. Select **Profit** and **Total Expenses** from **Measures by pressing the** *Ctrl* **key and clicking on both the fields.**
3. Click on the scatter plots button from the **Show Me** toolbar.
4. Click on the **Analysis** menu option and uncheck the **Aggregate Measures** option.
5. Right-click on any data marker or anywhere in the plot area and click on the **Show Trend Lines** option in **Trend Lines** to see a plot with a linear trend line, as shown in the following screenshot:



How it works....

The trend line that is added by default is a linear trend line, which is the simplest type of trend line explaining the relationship between two variables, and as we would expect in this recipe, the relationship is quite linear; that is, the profits are generally higher for expensive products. This trend line is mathematically represented as *Profit = 0.628675*Total Expenses + 27.1093*. This trend line can be edited to observe complex relationships, such as logarithmic, exponential, and polynomial. To change the trend line, right-click on the trend line and click on **Edit Trend Line**. You'll see a dialog box with various options, as shown in the following screenshot. From this box, select options from **Model type** to observe which trend line fits the data better:



There's more...

When we fit a linear model trend line on the data, we are essentially performing linear regression to fit a straight line to the data. If you are unfamiliar with linear regression, you will find the following resources helpful:

- The Khan academy video at http://youtu.be/OhUkMQtBGmE
- A reference from a Psychology course at Illinois State University: http://psychology.illinoisstate.edu/jccutti/psych340/fall02/oldlecturefile

Selecting color palettes

One of the biggest strengths of Tableau is its color selection for various visualizations. This color selection is based on best practices and the concepts of information visualization. Sometimes, however, we may want to change the default color settings.

Getting ready

For this recipe, let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

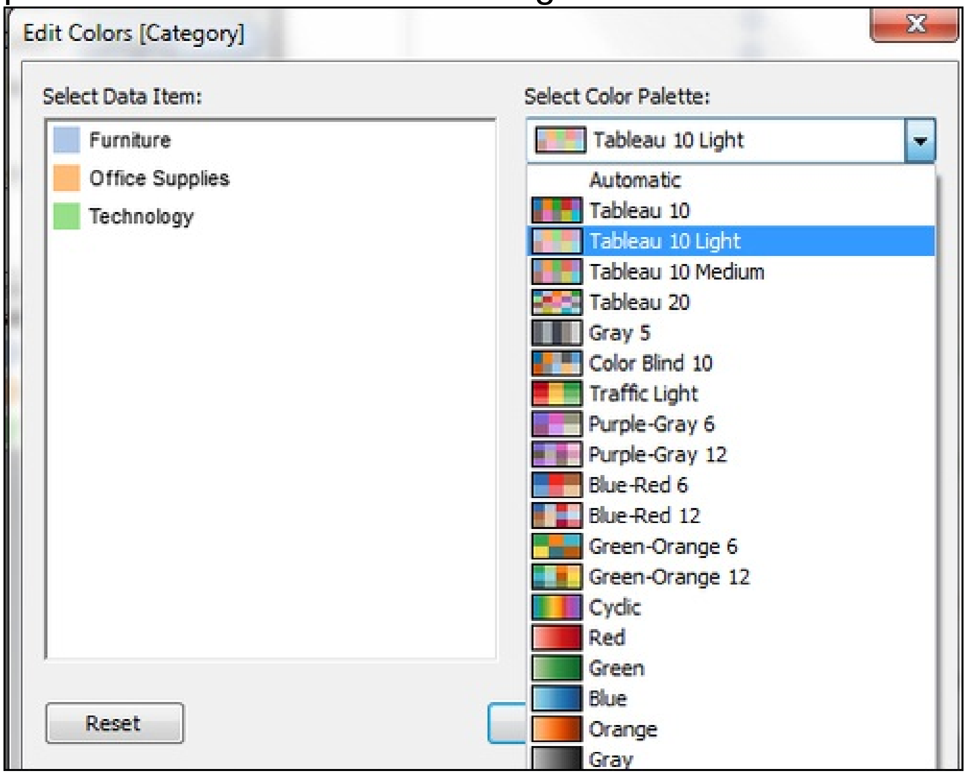
How to do it...

Once the data is loaded, perform the following steps to select different types of color palettes:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. Select **Discount** and **Profit** from **Measures by pressing the** *Ctrl* **key and clicking on both the fields.**
3. Click on the **Scatter Plots** button from the **Show Me** toolbar.
4. Click on the **Analysis** menu option and uncheck the **Aggregate Measures** option.
5. Drag-and-drop **Category** from **Dimensions** into the **Color** box under the **Marks** pane.
6. Click on the small dropdown, which becomes visible after hovering on the **Category** legend pane, as shown in the following screenshot, or double-click on any legend key: 

- Click on the **Edit Colors** option.
- In the **Edit Colors [Category]** properties box, under **Select Color**

Palette, expand the dropdown and select the **Tableau 10 Light** color palette as shown in the following screenshot:



- Click on the **Assign Palette** button to assign the colors from this palette to various categories.
- Click on **OK**.

How it works...

Tableau provides many color palettes to choose from, and these palettes are designed to maximize the effectiveness of colors in the visualizations. It is also possible to change the color of a single legend by selecting a value from the legends and then selecting a color from the palette and clicking on the **Assign Palette** button.

There's more...

If you are interested in learning more about the theory behind using proper colors, you will find the slides on color from the Information Visualization Stat 120 class by *Ross Ihaka* at the University of Auckland very insightful; go to the following link:

https://www.stat.auckland.ac.nz/~ihaka/120/Lectures/lecture14.pdf

Using dates

Tableau provides various options to analyze data fields that are of date type. Some options include grouping by quarters, years, or months.

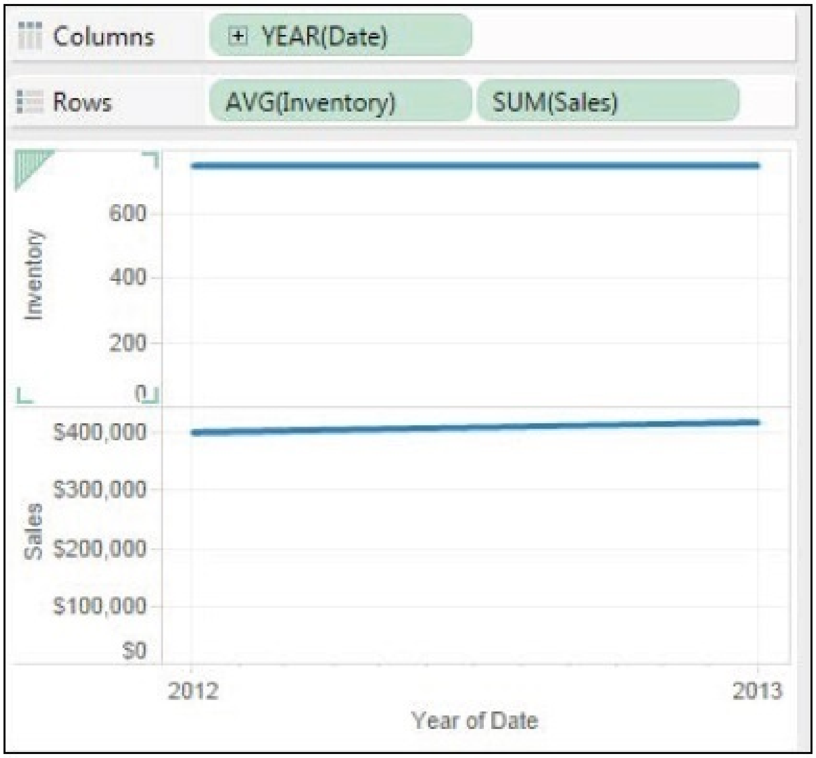
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access`). Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

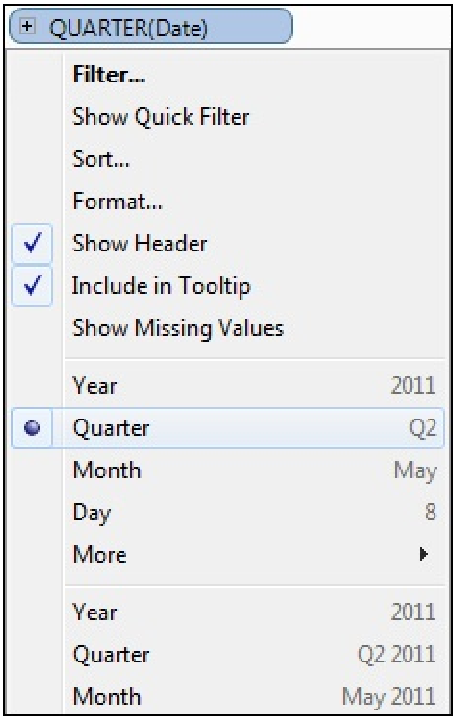
How to do it....

Once the data is loaded, perform the following steps to use dates in your analysis:

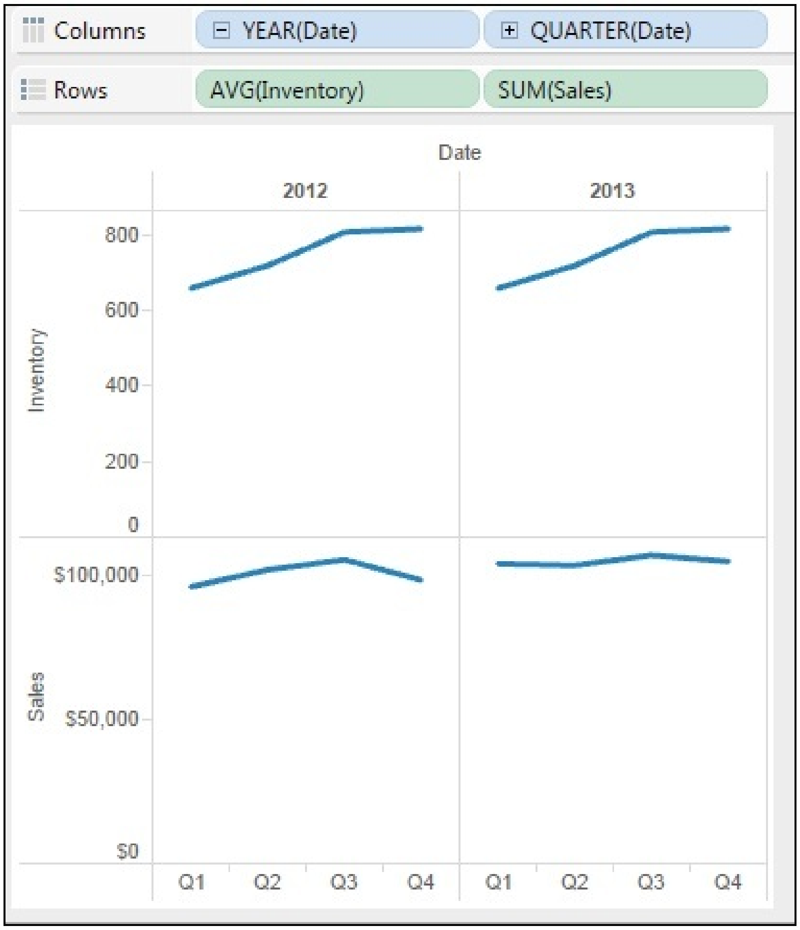
1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. Select **Inventory** and **Sales** from **Measures** and **Date** from **Dimensions** by pressing the *Ctrl* key and clicking on these fields.
3. Select the **lines (continuous)** chart type from the **Show Me** toolbar to generate a visualization as shown in the following screenshot. The following chart shows us the average inventory size and total sales for the years 2012 and 2013:



- To observe the time-series data by the quarters of a year, we need to add the **Date** field again. Drag-and-drop **Date** from **Dimensions** into the **Columns** shelf.
- In the **Columns** shelf, you'll now see two **Date** fields grouped by year. To show the time-series by quarter, click on the second **YEAR(Date)** value.
- From the drop-down menu, select the **Quarter** value that has the format **Q1** or **Q2**, as shown in the following screenshot:



- You can now observe the time-series trend by the quarters of 2012 and 2013, as shown in the following screenshot:



How it works...

Tableau will automatically group a date type of field by year, but it does provide various grouping options, such as quarter, month, day, weekday, and even by hour, minute, and second. To add multiple groupings, we need to add the date field multiple times and change the grouping option. This allows the user to generate data for various types of analyses that generate trends by different date combinations, thus assisting the reader to observe micro or macro trends. We saw the different time-series trends; when the data was grouped only by year, we observed that the time-series lines were almost flat. However, when we added the quarters, we observed that the average inventory went up in the third quarter of 2012 but the total sales went down in the same time period.

Chapter 4. Creating Multivariate Charts

Multivariate analysis involves analyzing multiple measures. In this chapter, we will create graphs that can effectively visualize multiple measures.

We will cover the following topics:

- Creating facets
- Creating area charts
- Creating bullet graphs
- Creating dual axes charts
- Creating Gantt charts
- Creating heat maps

Introduction

With increasing number of variables, any analysis can become challenging and any observations harder; however, Tableau simplifies the process for the designer and uses effective layouts for the reader even in multivariate analysis. Using various combinations of colors and charts, we can create compelling graphics that generate critical insights from our data. Among the charts covered in this chapter, facets and area charts are easier to understand and easier to create compared to bullet graphs and dual axes charts.

Creating facets

Facets are one of the powerful features in Tableau. *Edward Tufte*, a pioneer in the field of information graphics, championed these types of charts, also called **grid** or **panel charts**; he called them small multiples. These charts show the same measure(s) across various values of one or two variables for easier comparison.

Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

Once the data file is loaded on the new worksheet, perform the following steps to create a simple faceted chart:

1. Drag-and-drop **Market** from **Dimensions** into the **Columns** shelf.
2. Drag-and-drop **Product Type** from **Dimensions** into the **Rows** shelf.
3. Drag-and-drop **Profit** from **Measures** into the **Rows** shelf next to **Product Type**.
4. Optionally, you can drag-and-drop **Market** into the **Color Marks** box to give color to the four bars of different **Market** areas. The chart should look like the one in the following screenshot:

How it works...

When there is one dimension on one of the shelves, either **Columns** or **Rows**, and one measure on the other shelf, Tableau creates a univariate bar chart, but when we drop additional dimensions along with the measure, Tableau creates small charts or facets and displays univariate charts broken down by a dimension.

There's more...

A company named Juice Analytics has a great blog article on the topic of small multiples. This article lists the benefits of using small multiples as well as some examples of small multiples in practice. Find this blog at http://www.juiceanalytics.com/writing/better-know-visualization-small-multiples/.

Creating area charts

An area chart is an extension of a line chart. The area chart shows the line of the measure but fills the area below the line to emphasize on the value of the measure. A special case of area chart is a stacked area chart, which shows a line per measure and the area between the lines is filled. Tableau's implementation of area charts uses one date variable and one or more measures.

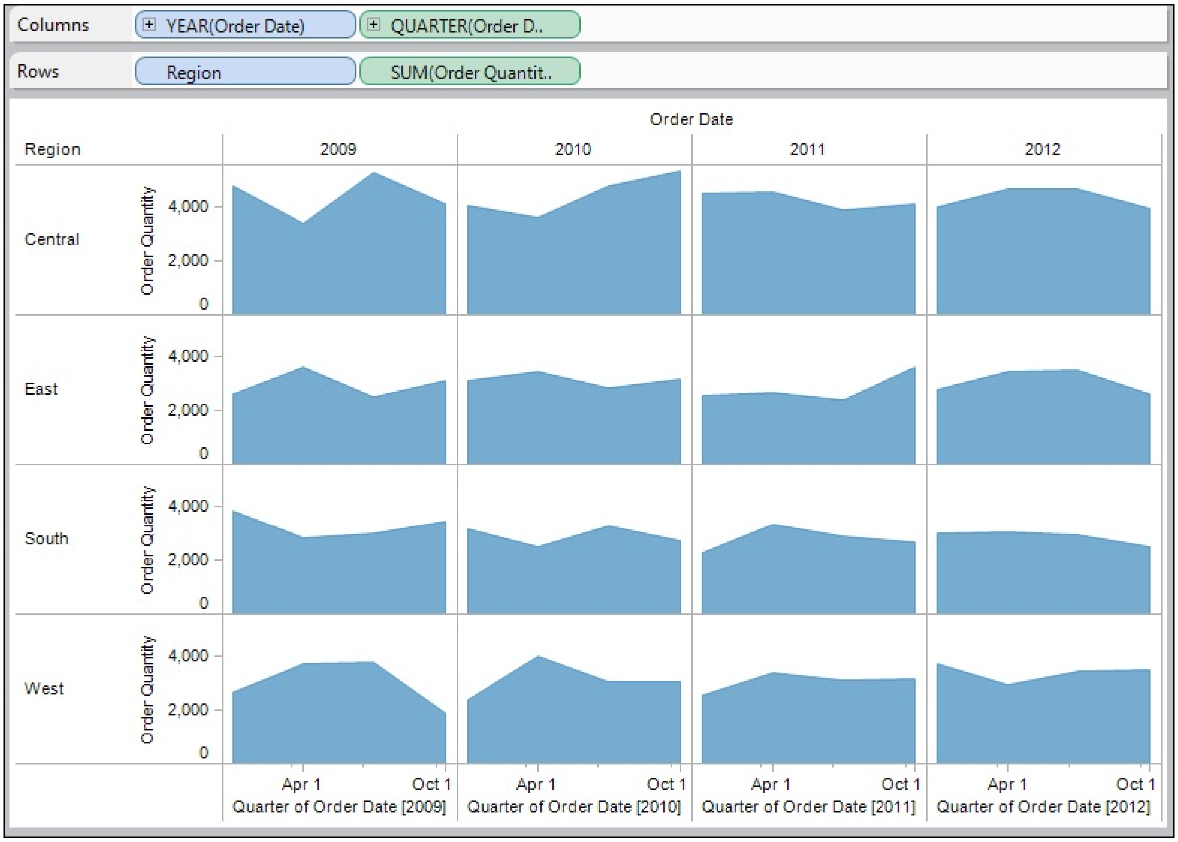
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded on the new worksheet, perform the following steps to create an area chart:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. Select **Order Date** from **Dimensions** and **Order Quantity** from **Measures** by clicking and holding the *Ctrl* key.
3. Click on **Area charts (continuous)** from the **Show Me** toolbar.
4. Drag-and-drop **Order Date** into the **Columns** shelf next to **YEAR(Order Date)**.
5. Expand **YEAR(Order Date)**, seen on the right-hand side, by clicking on the plus sign.
6. Drag-and-drop **Region** from **Dimensions** into the the **Rows** shelf to the left of **SUM(Order Quantity)**. The chart should look like the one in the following screenshot:



How it works…

When we added **Order Date** for the first time, Tableau, by default, aggregated the date field by year; therefore, we added **Order Date** again to create aggregation by quarter of the **Order Date**. We also added **Region** to create facets on the regions that provide trends of order quantity over time.

There's more...

A blog post by `visual.ly`, an information graphics company, discusses the key differences between line charts and area charts. You can find this post at http://blog.visual.ly/line-vs-area-charts/.

Creating bullet graphs

Stephen Few, an information visualization consultant and author, designed this chart to solve some of the problems that the gauges and meters type of charts poses. Gauges, although simple to understand, take a lot of space to show only one measure. Bullet graphs are a combination of the bar graph and thermometer types of charts, and they show a measure of interest in the form of a bar graph (which is the bullet) and target variables.

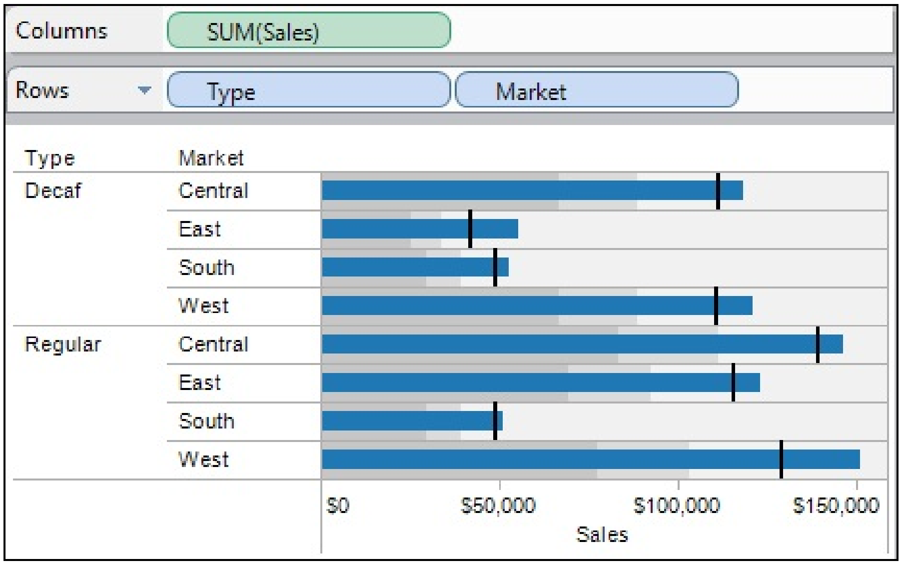
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

Once the data is loaded on the sheet, perform the following steps to create a bullet graph:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Type** and **Market** from **Dimensions** and **Budget Sales** and **Sales** from **Measures**.
3. Click on the bullet graphs icon on the **Show Me** toolbar.
4. Right-click on the x axis (the **Budget Sales** axis) and click on **Swap Reference Line Fields**. The final chart should look like the one in the following screenshot:



How it works...

Although bullet graphs maximize the available space to show relevant information, readers require detailed explanation as to what all the components of the graphic are encoding. In this recipe, since we want to compare the budgeted sales with the actual sales, we had to swap the reference line from **Sales** to **Budget Sales**. The black bar on the graphic shows the budgeted sales and the blue bar shows the actual sales. The dark gray background color shows 60 percent of the actual sales and the lighter gray shows 80 percent of the actual sales. As we can see in this chart, blue bars crossed all the black lines, and that tells us that both the coffee types and all market regions exceeded the budgeted sales.

There's more...

A blog post by Data Pig Technologies discusses some of the problems with the bullet graph. The main problem is intuitive understanding of this chart. You can read about this problem and the reply by *Stephen Few* at http://datapigtechnologies.com/blog/index.php/the-good-and-bad-of-bullet-graphs/.

Creating dual axes charts

Dual axes charts are useful to compare two similar types of measures that may have different types of measurement units, such as pounds and dollars. In this recipe, we will look at the dual axes chart.

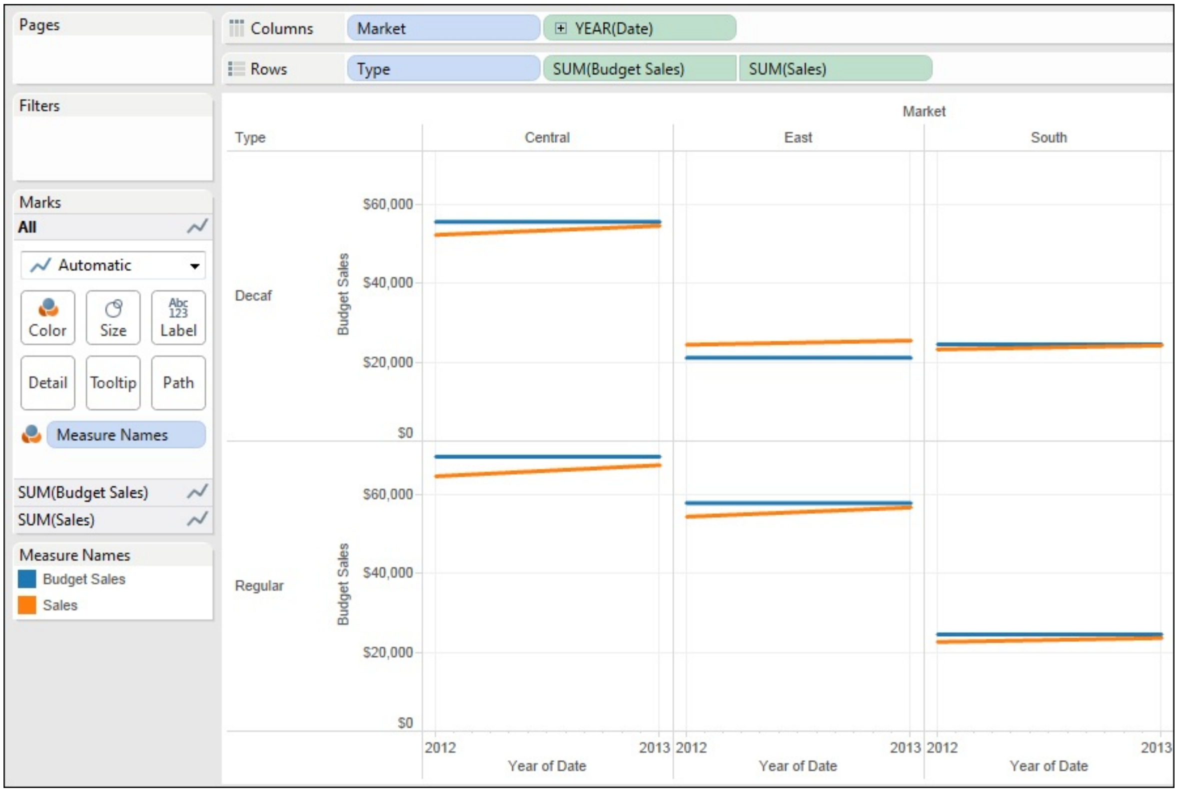
Getting ready

Let's use the same sample file, `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

Once the data is loaded on the sheet, perform the following steps to create a dual axes chart:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Date**, **Type**, and **Market** from **Dimensions** and **Sales** and **Budget Sales** from **Measures**.
3. Click on the dual line graph icon on the **Show Me** toolbar.
4. Click-and-drag **Market** from the **Rows** shelf into the **Columns** shelf.
5. Right-click on the **Sales** vertical axis and click on **Synchronize Axis**. The chart should look like the one shown in the following screenshot:



How it works...

Tableau will create two vertical axes and automatically place **Sales** on one dual axes charts vertical axis and **Budget Sales** on the other. The scales on both the vertical axes are different, however. By synchronizing the axes, we get the same scales on both axes for better comparison and accurate representation of the patterns.

Creating Gantt charts

Gantt charts are most commonly used in project management as these charts show various activities and tasks with the time required to complete those tasks. Gantt charts are even more useful when they show dependencies among various tasks. This type of chart is very helpful when the number of activities is low (around 20-30), otherwise the chart becomes too big to be understood easily.

Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

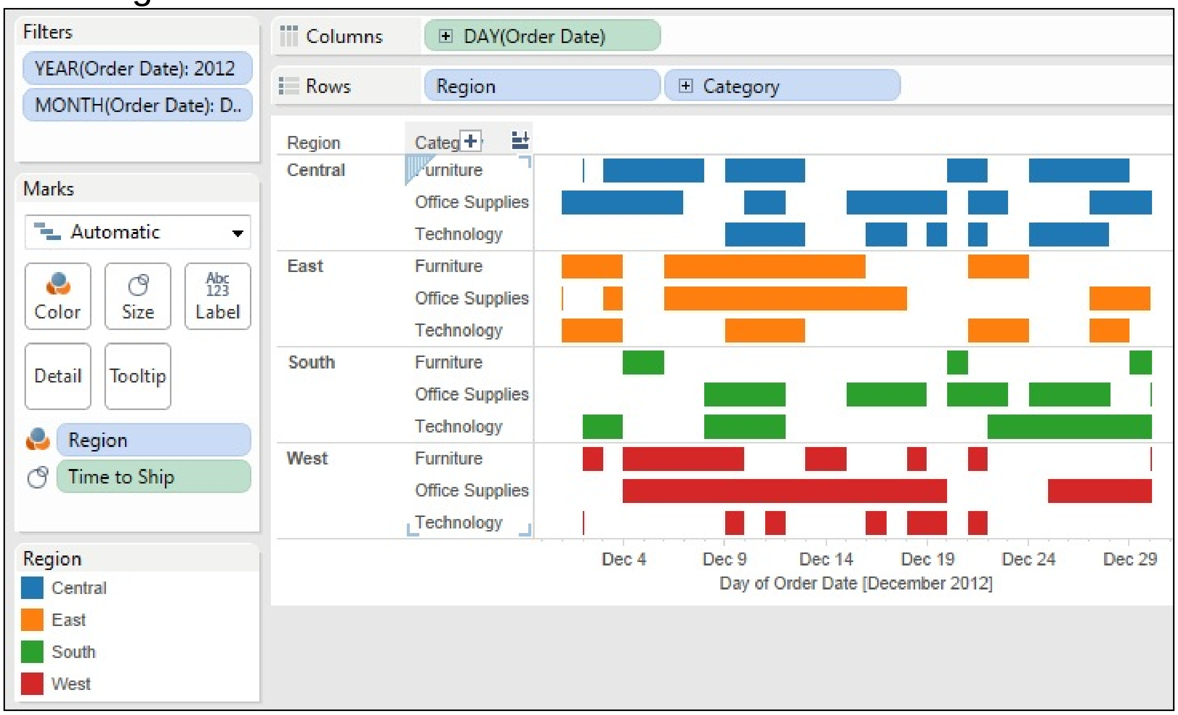
How to do it...

Once the data is loaded, perform the following steps to create a Gantt chart:

1. Click on **Analysis** from the top menu toolbar, and if **Aggregate Measures** is checked, click on it again to uncheck that option.
2. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
3. While holding the *Ctrl* key, click on **Order Date** and **Category** (under **Products**) from **Dimensions** and **Time to Ship** from **Measures**.
4. Click on the Gantt chart icon on the **Show Me** toolbar.
5. Drag-and-drop **Order Date** into the **Filters** pane.
6. Select **Years** from the **Filter Field [Order Date]** options dialog box and hit **Next**.
7. Check **2012** from the list and hit **OK**.
8. Right-click on **YEAR(Order Date)** on the **Columns** shelf and select the **Day May 8, 2011** option.
9. Drag-and-drop **Order Date** into the **Filters** pane.
10. Select **Months** from the **Filter Field [Order Date]** options dialog box

and hit **Next**.

11. Check **December** from the list and hit **OK**.
12. Drag-and-drop **Region** from **Dimensions** into the **Color Marks** input box.
13. Drag-and-drop **Region** from **Dimensions** into the **Rows** shelf before **Category**. The generated Gantt chart should look like the one in the following screenshot:



How it works...

Representing time this way helps the reader to discern which activity took the longest amount of time. We added the **Order Date** field two times in the **Filters** pane to first filter for the year 2012 and then for the month of December. In this recipe, out of all the products shipped in December of 2012, we can easily see the red bars for the **West** region in the **Office Supplies** category is longer, suggesting that these products took the longest amount of time to ship.

There's more...

Andy Kriebel, a Tableau data visualization expert, has a great example of

Gantt charts using US presidential data. The following link shows the lengths of terms in office of Presidents from various parties: http://vizwiz.blogspot.com/2010/09/tableau-tip-creating-waterfall-chart.html

Creating heat maps

A heat map is a visual representation of numbers in a table or a grid such that the bigger numbers are encoded by darker colors or bigger sizes and the smaller numbers by lighter colors or smaller sizes. This type of representation makes the reader's pattern detection from the data easier.

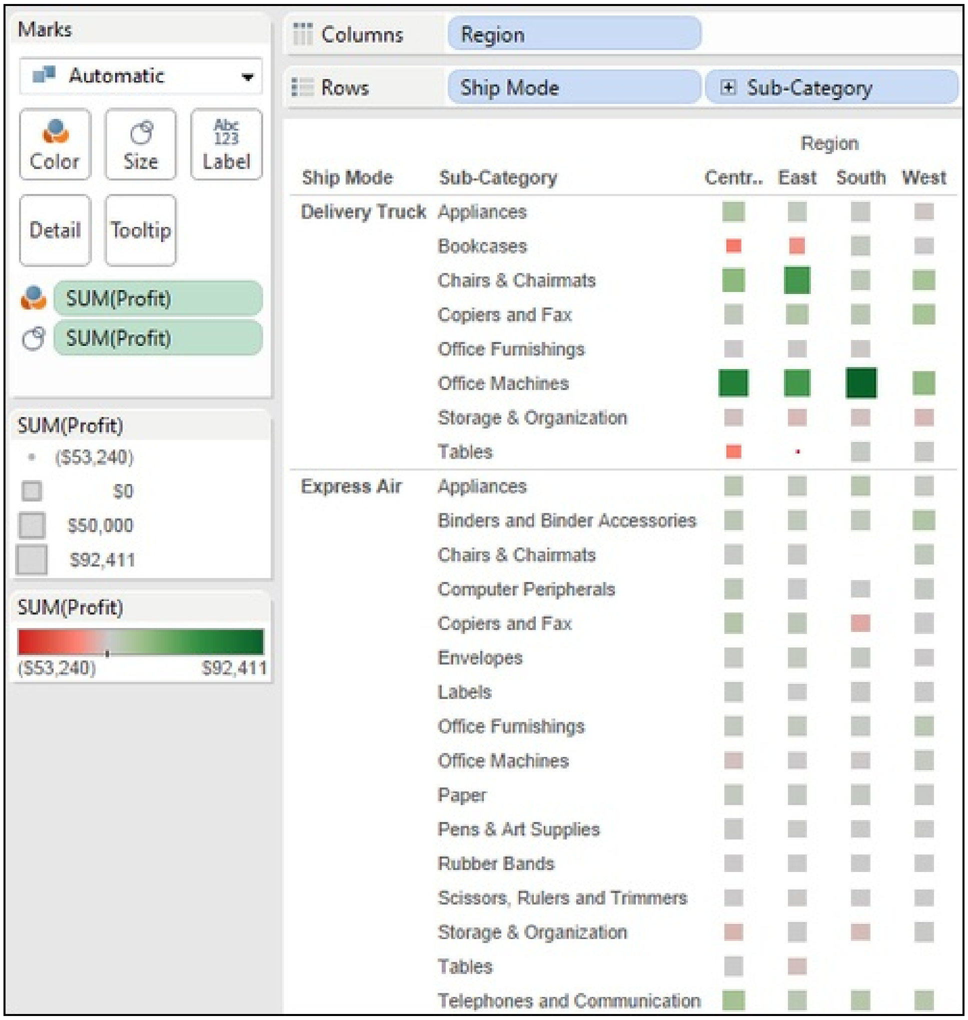
Getting ready

Let's use the same sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to create a heat map chart:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Sub-Category** (under **Products**), **Region**, and **Ship Mode** from **Dimensions** and **Profit** from **Measures**.
3. Click on the heat maps chart icon on the **Show Me** toolbar.
4. Drag-and-drop **Profit** from **Measures** into the **Color Marks** box. The generated chart should look like the one in the following screenshot:



How it works...

When we created the chart for the first time, Tableau assigned various sizes to the square boxes, but when we placed **Profit** as a color mark, red was used for low amounts of profit and green was used for higher amounts of profit. This made spotting of patterns very easy. **Binders and Binder Accessories**, shipped by **Regular Air** in the **Central** region, generated very high amounts of profit and **Tables**, shipped by **Delivery Trucks** in the **East** region, generated very low amounts of profit (it actually created losses for the company).

Chapter 5. Creating Maps

In this chapter, we will cover the following recipes:

- Setting geographic roles
- Placing marks on a map
- Overlaying demographic data
- Creating choropleth maps
- Using polygon shapes
- Customizing maps

Introduction

Overlaying information on top of maps allows the readers to understand and observe data by various regions and geographic boundaries. In some other software, creating such maps would be a time-consuming task; in Tableau, however, it is very straightforward. Although seeing dense data in a map could confuse readers, Tableau provides a couple of options to create insightful maps with the use of colors, shapes, and sizes.

Setting geographic roles

Once the data is loaded, Tableau will determine geographic fields using the field names, such as city, state, and zip code, and will generate latitude/longitude data for those fields. Tableau will denote the geographic fields by placing a globe symbol next to the field name. If Tableau misses the detection of any field, usually due to variations in field names, we can manually set these fields as geographic fields.

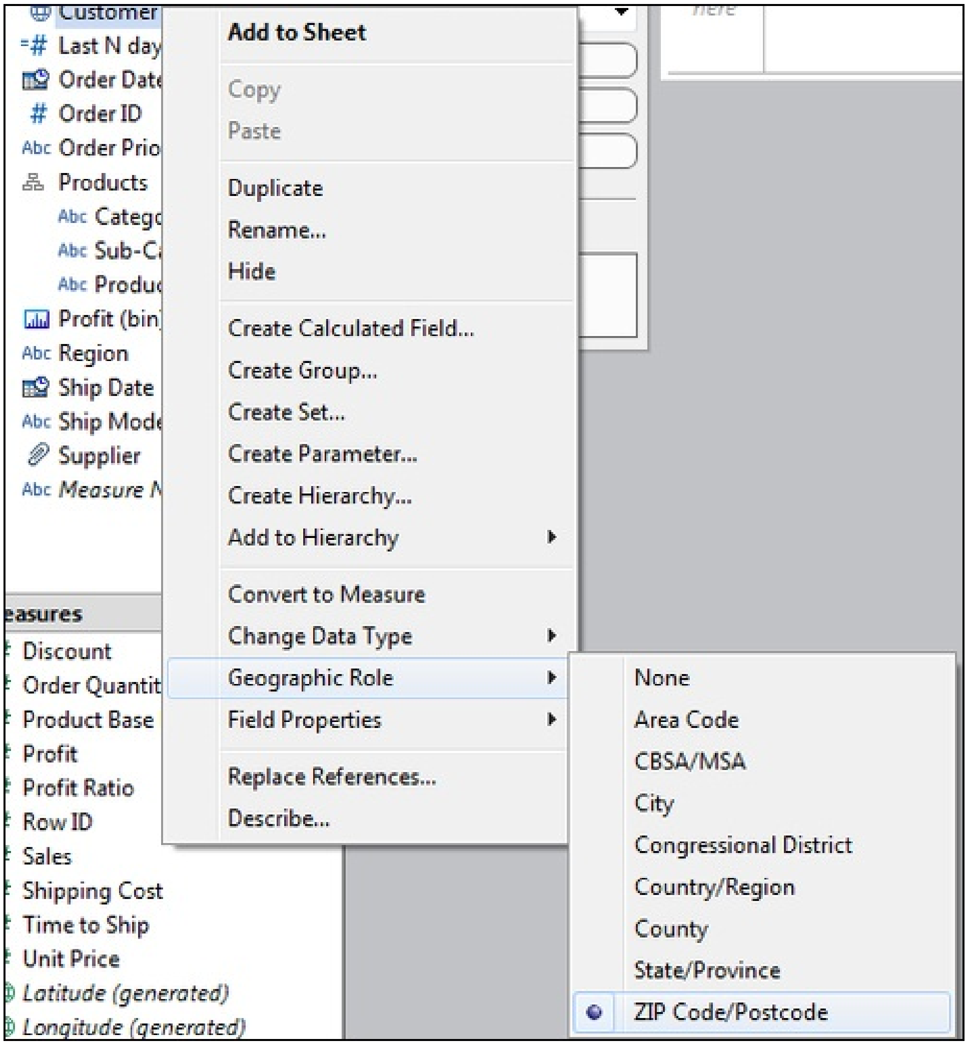
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded in Tableau, perform the following steps to set a few fields as geographic fields:

1. Right-click on **Customer Zip Code** from **Dimensions**.
2. Expand **Geographic Role** from the dropdown.
3. Select **Zip code/Postcode** as shown in the following screenshot.
4. Right-click on **State** under **Customer City** from **Dimensions**.
5. Expand **Geographic Role** from the dropdown.
6. Select **State/Province** from the options.



How it works...

When the field names are different from conventional names, Tableau will not know that these fields are geographic fields. We can manually assign various fields as geographic fields, which generate latitude and longitude to be used in the maps. Sometimes Tableau cannot match a field to its internal data; for example, if there's a state called UH in our data, but Tableau does not have that value in the list of states of the US. In such cases, Tableau will prompt for mapping of the values from the data.

There's more...

In Tableau's documentation on geographic roles, you can find out which fields Tableau can geocode automatically; browse to http://onlinehelp.tableausoftware.com/v8.0/pro/online/en-us/maps_geographicroles.html. Geocoding of fields with information on area code, CBSA/MSA, congressional district, and county are limited to the US only.

Placing marks on a map

One of the ways to encode information on a map is placing a mark for each geographic value and adjusting the size/color of that mark based on some measure. This is the most common type of a map with information used in businesses and media.

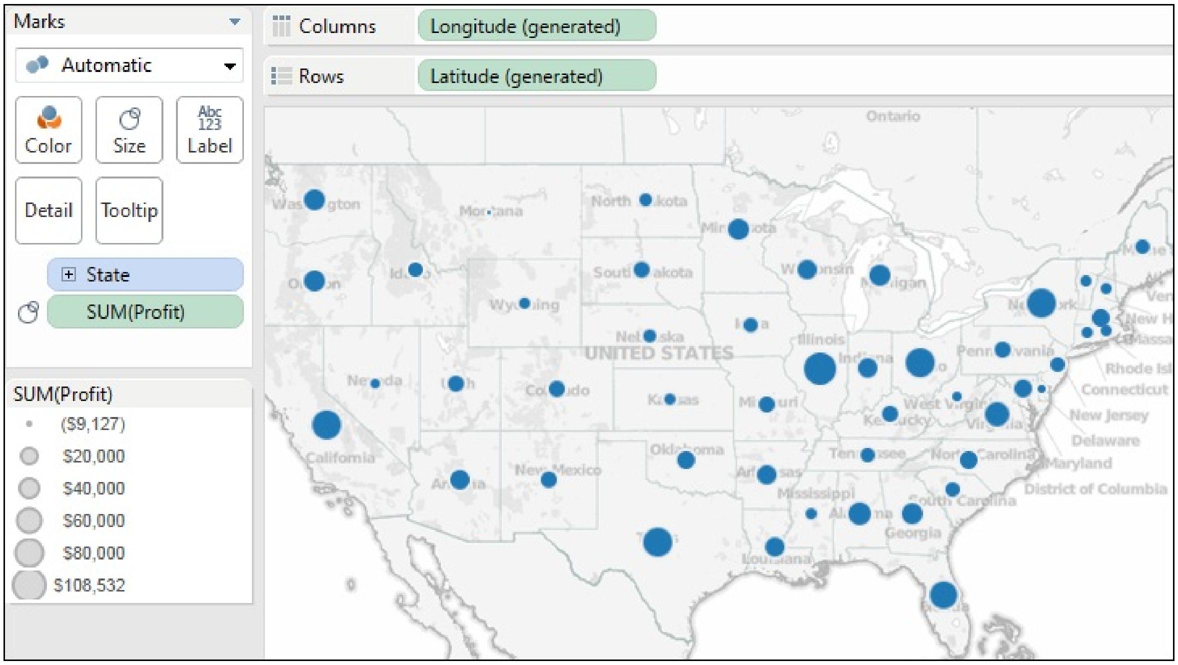
Getting ready

Let's use the sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to create a map with markers encoding information of a measure:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **State** (under **Customer City**) from **Dimensions** and **Profit** from **Measures**.
3. Click on the symbol of maps on the **Show Me** toolbar, and you will see the map as shown in the following screenshot:



How it works...

Once we loaded the data, Tableau assigned geographic roles to **State**, **City**, and **Customer Zip Code** using field names. Tableau also generated **Latitude** and **Longitude** for these geographic roles. When we clicked on the symbol of maps, Tableau automatically added **State** to the **Level of Detail** pane and placed marks for every state that was present in the data. If you add **Customer Zip Code** to the **Detail box**, Tableau will generate markers by zip code and adjust the mark size for the **Profit** value of that zip code.

Overlaying demographic data

Using data from various service providers, Tableau provides a powerful feature of overlaying the US census information, such as median household income, population, race, and others. This allows the reader to compare the measure of interest with some demographic information.

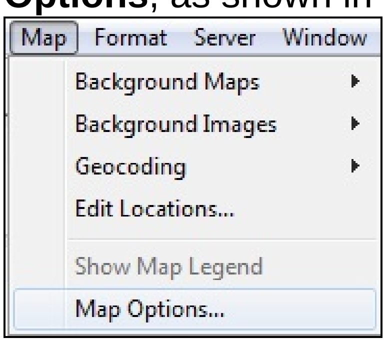
Getting ready

Let's use the sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

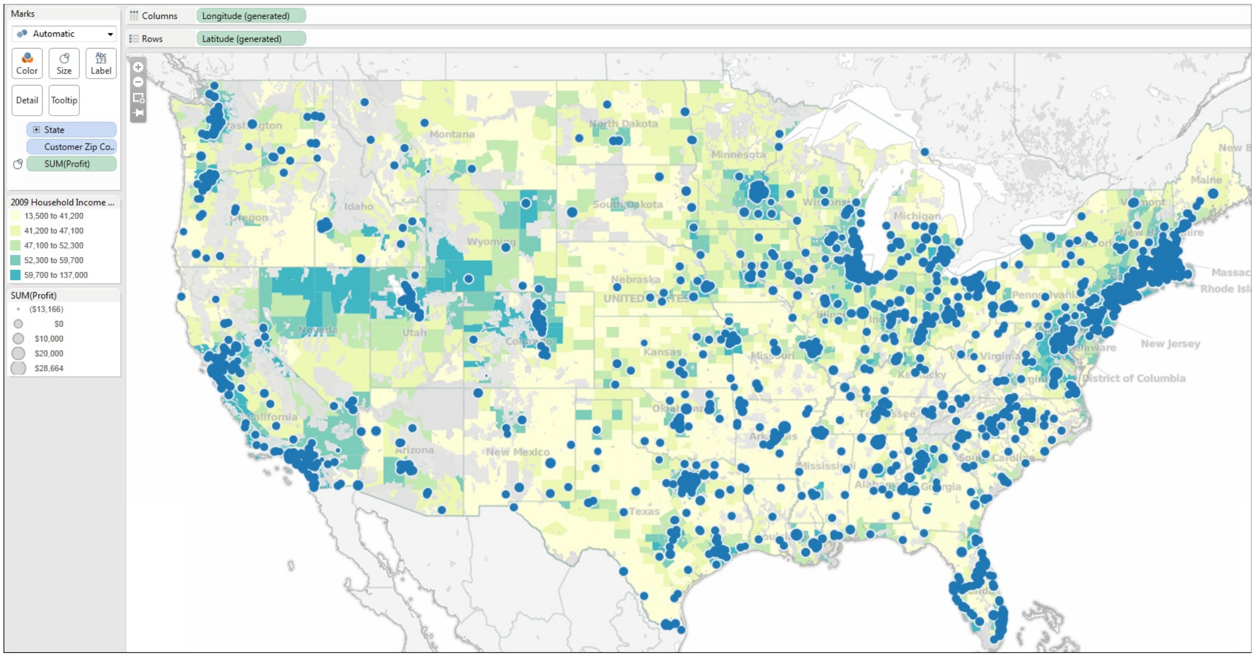
Once the data is loaded on the worksheet, perform the following steps to overlay the demographic data:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **State and Customer Zip Code** from **Dimensions** and **Profit** from **Measures**.
3. Click on the symbol of maps on the **Show Me** toolbar.
4. Drag-and-drop **State** and **Customer Zip Code** from **Dimensions** into the **Level of Detail** box.
5. From the main menu toolbar, click on **Map** and then on **Map Options**, as shown in the following screenshot:



- From the **Map Options** (shown in the following screenshot) box, expand the dropdown **No Data Layer**.

- Select **Household Income (median)** under **US Households**.
- From the **By:** dropdown option, select **County** to generate a map similar to the one shown in the following screenshot:



How it works...

Based on our choice of aggregation level for overlaying demographic information, Tableau fills the level (that is, county, state, zip code, and the block group) with the selected demographic information. In our recipe, we are comparing the median household income of counties with the profits generated by each zip code in the data. We can see that there are many counties in **Nevada** with high median household income, but hardly any profit-generating zip codes. By such comparisons, the reader can identify areas of growth.

There's more...

Tableau provides these powerful features using various data providers listed on Tableau's website: http://www.tableausoftware.com/mapdata. If these maps do not meet your needs, Tableau also has the option of using an open technology called **Web Map Services** (**WMS**) to get the map source. You can read about using a WMS server at http://onlinehelp.tableausoftware.com/current/pro/online/en-us/maps_mapsources_wms.html.

Creating choropleth maps

A choropleth map, known as a **filled map** in Tableau, is a modification of a traditional marks map, in that study areas (regions, states, and counties) are filled with the measure of interest and colors are used with different hues or diverging progression to assist the reader in identifying areas of poor or good performance.

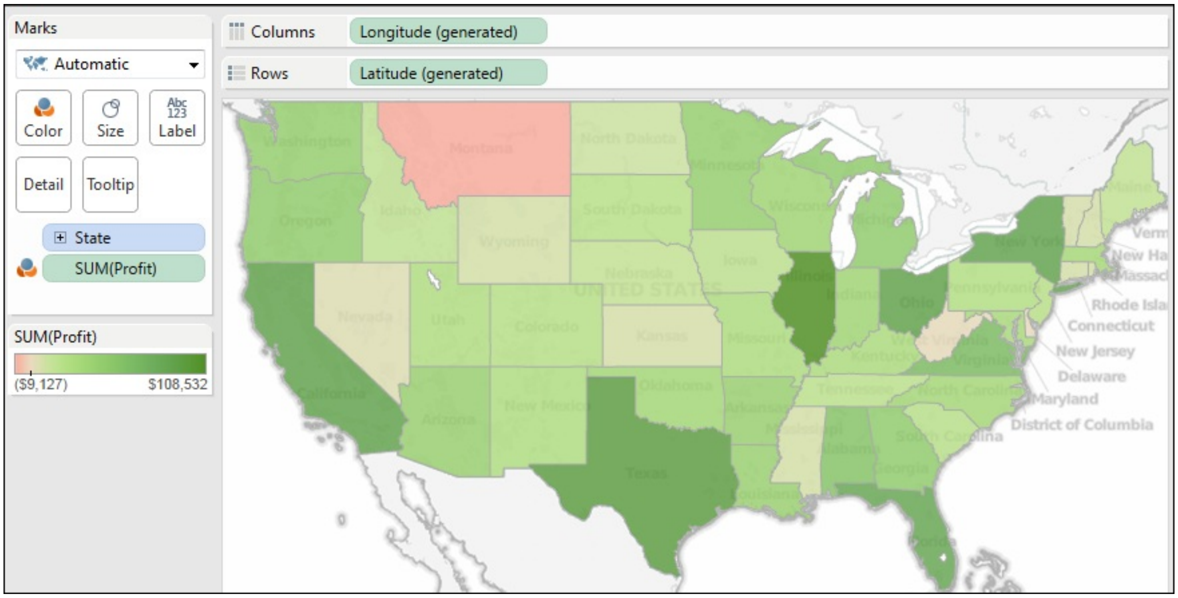
Getting ready

Let's use the sample file, `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to create a choropleth or a filled map:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Profit** from **Measures** and **State** from **Dimensions**.
3. Click on **filled maps** on the **Show Me** toolbar to create a choropleth map as shown in the following screenshot:



How it works...

We selected **State** as the level of detail for this example of choropleth maps (known as filled maps in Tableau), and Tableau created a color range to encode **Profit**; that is, dark green for higher profit and pink for losses. This allows the reader to quickly identify **Montana** as a state with losses and **West Virginia** and **Nevada**, among others, as states with lower profit margins. It is slightly challenging, however, to identify which states grossed the highest profit, as hues of the green color at a higher profit level look very similar, and that is one of the main disadvantages of using this type of map. An alternative to overcome this problem will be creating a sorted bar chart, which, if plotted correctly, will help the reader identify similar states without any guesswork.

Using polygon shapes

Tableau 7 and above have a functionality to create filled maps using fill map marks. These marks are useful when your levels of detail (or the shape of the filled area) are limited to the US counties or county/state combinations. If you want to create custom-shaded maps using geographical boundaries, such as districts of India, you have to use the following steps. You can create such filled maps using a polygon file, which consists of latitude and longitude of various points on the boundaries or shapes of the custom region. An example of such a file is shown in the following screenshot:

```
"long","lat","group","order","region","subregion"
-90.4127349853516,46.5585517883301,1,1,"michigan","north"
-90.3783645629883,46.5642776489258,1,2,"michigan","north"
-90.3153381347656,46.5929260253906,1,3,"michigan","north"
-90.2809600830078,46.6158447265625,1,4,"michigan","north"
-90.2007522583008,46.6330337524414,1,5,"michigan","north"
-90.1377182006836,46.6444969177246,1,6,"michigan","north"
-90.0746994018555,46.6559524536133,1,7,"michigan","north"
-90.0231323242188,46.6788711547852,1,8,"michigan","north"
-89.9543762207031,46.713249206543,1,9,"michigan","north"
-89.9028091430664,46.7418975830078,1,10,"michigan","north"
-89.8741607666016,46.7648162841797,1,11,"michigan","north"
-89.8397827148438,46.7934608459473,1,12,"michigan","north"
```

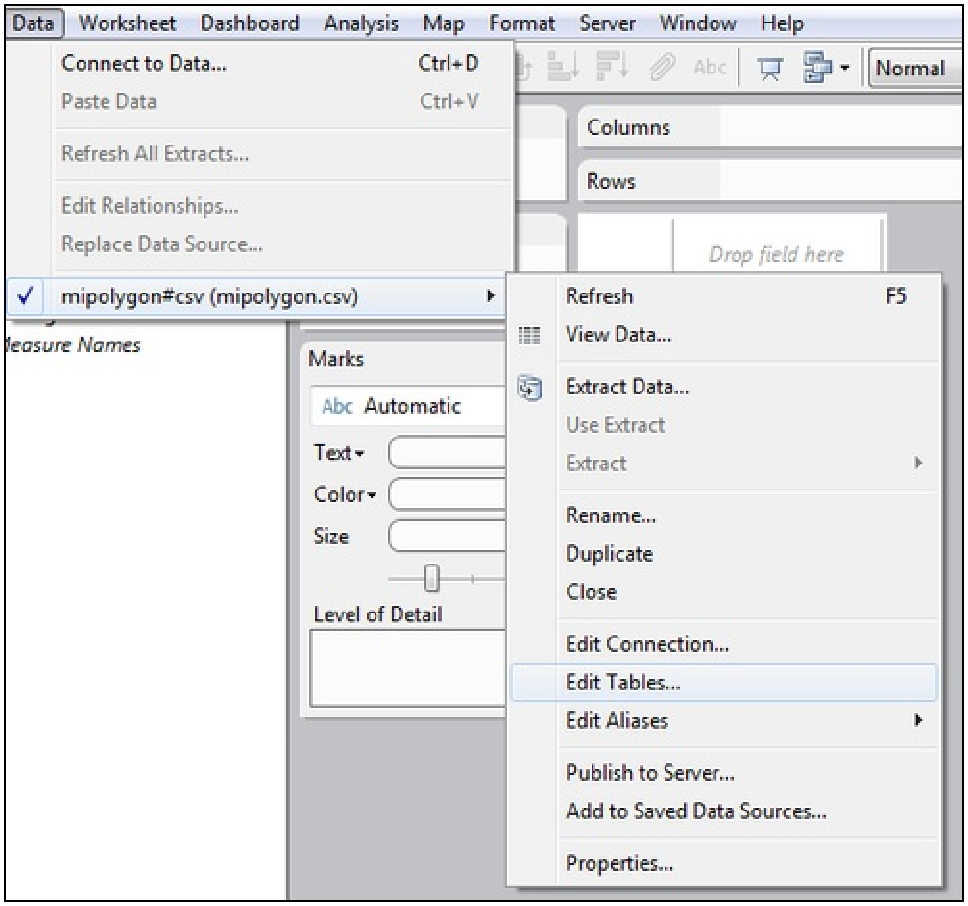
Getting ready

We'll plot the sales figures for the **Upper Peninsula** (**UP**) and **Lower Peninsula** (**LP**) of Michigan. To do so, we'll need two files: a file that contains latitude, longitude, and boundary groupings for Michigan (`mipolygon.csv`) and a file that contains sales numbers for the UP and LP (`misales.csv`). These files are provided with the downloadable code accompanying this book. In the following steps, we will join these two files to get all the required data and use the `mipolygon`.csv file to form the boundaries and the `misales`.csv file to fill the map.

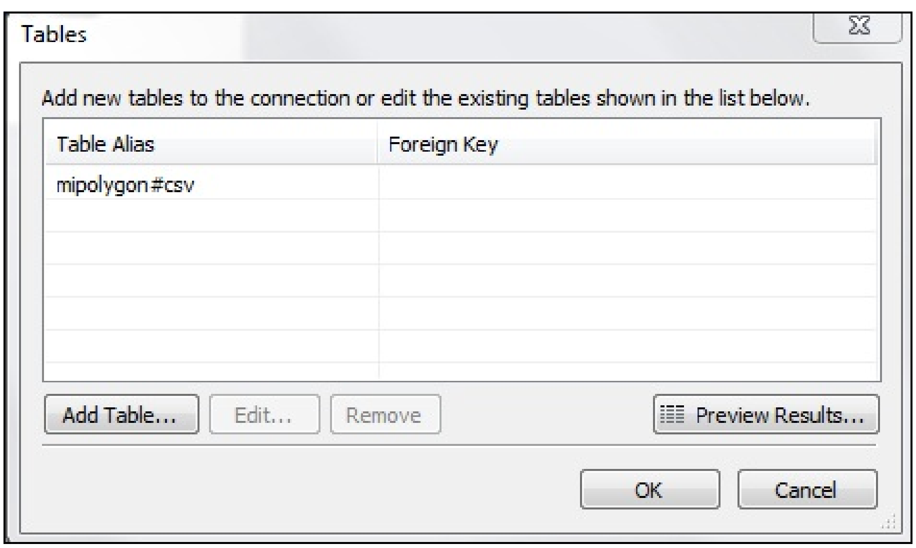
How to do it...

After you download and save `mipolygon.csv` and `misales.csv` locally, perform the following steps to create a map of Michigan with the UP and LP regions filled with the sales figures:

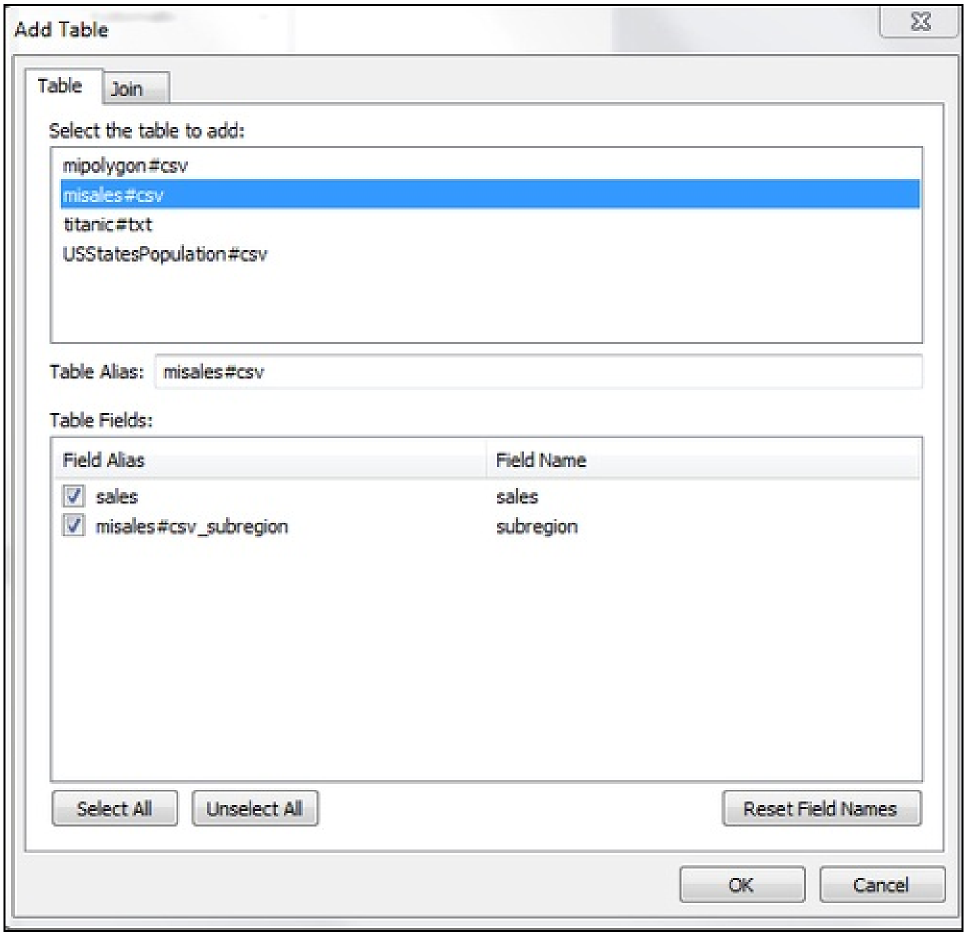
1. Open a new workbook by clicking on **New** under the **File** menu.
2. Click on **Connect to Data** to select the data file.
3. Under the **In a file** options, click on **Text File**.
4. Select `mipolygon.csv` from your downloaded file's location and hit **Open**.
5. Maintain all the default options and hit **OK**.
6. If you see a **Data Connection** option dialog box, select **Connect live**.
7. While holding the *Ctrl* key, click on **Group** and **Order** under **Measures** and then right-click on **Convert to Dimension**. Alternatively, you can drag-and-drop these fields into the **Dimensions** pane.
8. You'll notice that Tableau recognizes the measure **lat** as a geographic field as the field name matches Tableau's internal naming convention for latitude. However, the field **long** is not recognized. To set the geographic role, right-click on **long** under **Measures** and, under **Geographic Role**, select **Longitude**.
9. Click on **Data** from the top menu and expand **mipolygon#csv (mipolygon.csv)** and click on **Edit Tables**, as shown in the following

screenshot: 

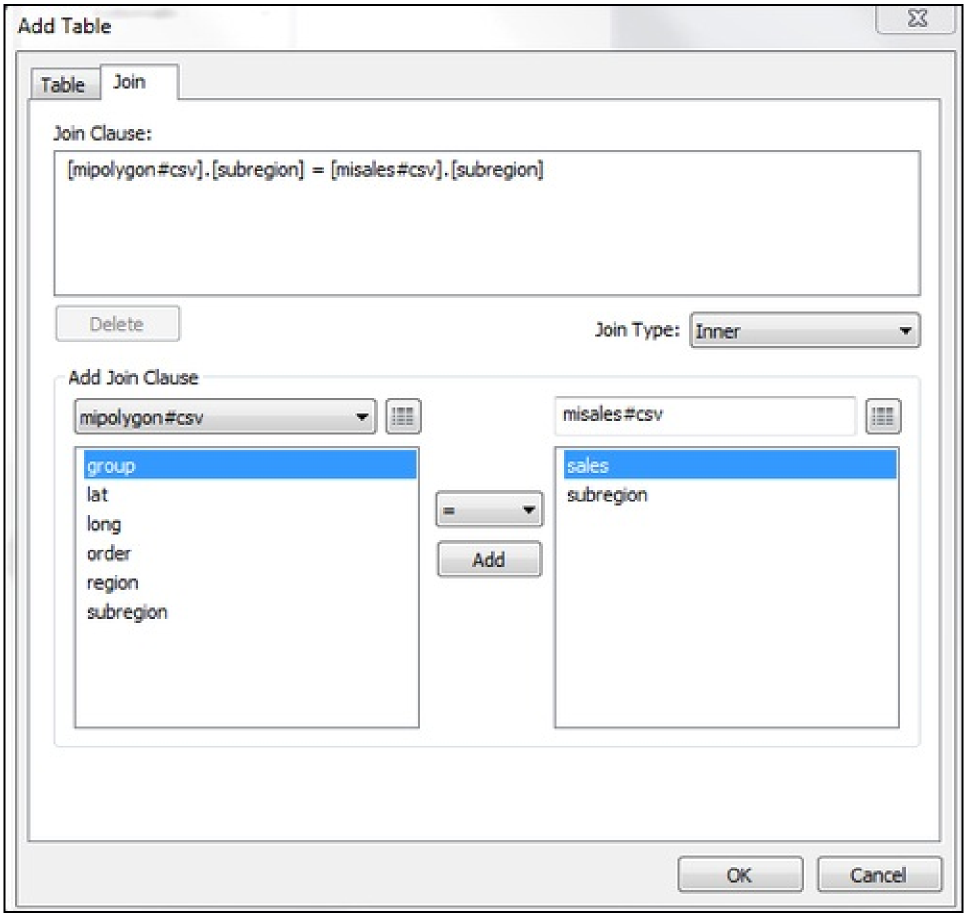
- In the **Tables** dialog box, click on the **Add Table** button.



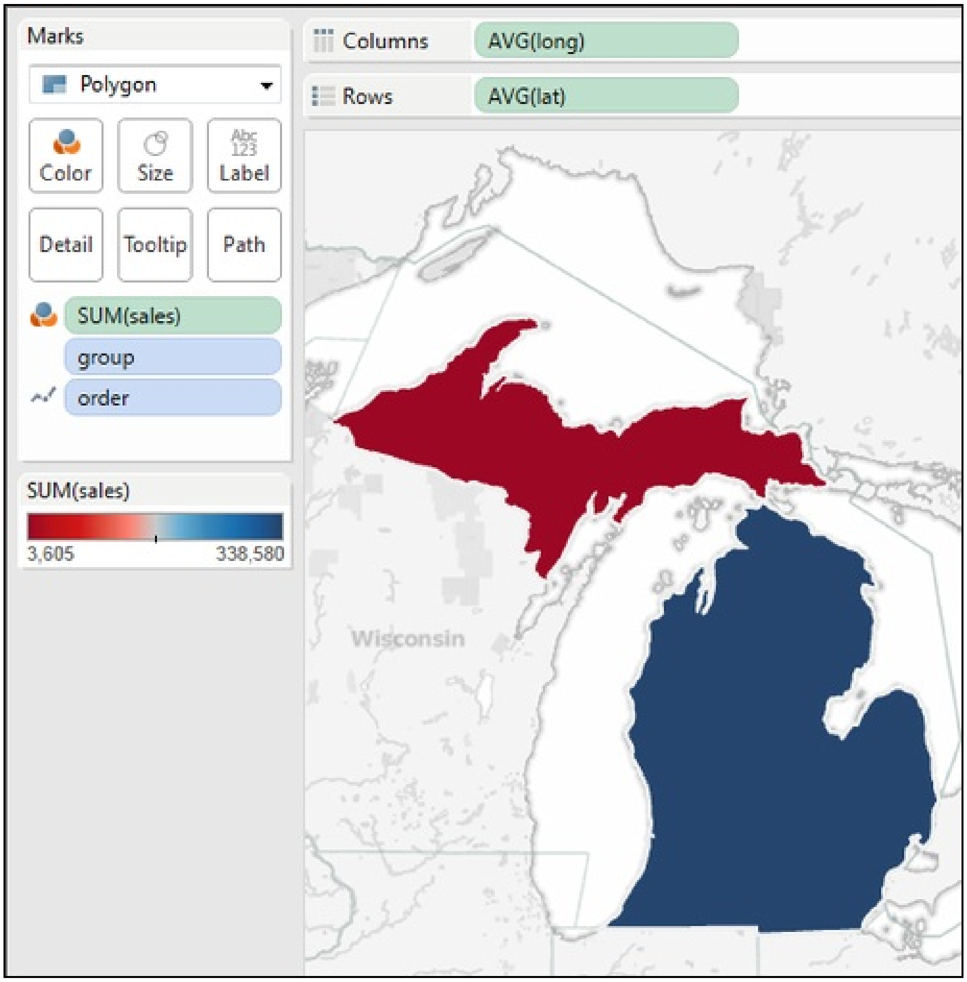
- Select **misales#csv** under the **Table** tab.



- Click on the **Join** tab and make sure that the **Join Clause** pane shows that the **subregion** value from the `mipolygon.csv` file is joined to the **subregion** option of the `misales.csv` file, as shown in the following screenshot. We do so to make sure that all the records from both files where the subregions match are returned. This is called an **inner join**.



- Hit **OK** on the **Edit Tables** options dialog box.
- Hit **OK** on the **Tables** dialog box.
- Drag-and-drop **lat** under **Measures** into the **Rows** shelf.
- Drag-and-drop **long** under **Measures** into the **Columns** shelf.
- Expand the **Marks** dropdown and select **Polygon**.
- Drag-and-drop **order** from **Dimensions** into the **Path** box in the **Marks** pane.
- Drag-and-drop **group** from **Dimensions** into the **Detail** box.
- Drag-and-drop **sales** from **Measures** into the **Color** box.
- Click on the dropdown arrow on the **SUM(sales)** legend and click on **Edit Colors**.
- Select the **Red-Blue Diverging** color from the **Palette** dropdown and hit **OK** to generate a custom polygon-filled map similar to the one shown in this following screenshot:



How it works...

Although we can create filled maps with Tableau 7.0 or higher, we sometimes need custom-filled maps that may not be contained in Tableau's internal data, such as regions within a state. To create such maps, we need a polygon file that has coordinates (latitude and longitude) for the area, an **Order** field to indicate the order of the outline of the polygon, and a grouping variable to indicate the boundaries. In this recipe, we stored two areas (north and south) and also the group field, whose coordinates fell under those two areas. You will need software such as ArcGIS to generate polygon files, which are also called **shape files**. Quantum GIS, an open source geographic information system, can also be used to create such a shape file. The polygon file used in this example was generated using the R language and the `map_data` function of the `ggplot2` package. The code to generate this file is as follows:

```
install.packages('maps')
install.packages('ggplot2')
library(maps)
library(ggplot2)
mimap <- map_data(map = "state", region = "michigan")
write.csv(mimap, file = "mipolygon.csv", row.names =
F)
```

Customizing maps

Tableau provides quite a few options to change the format of a generated map. Some of the options include washout, to make the map transparent, and removing borders. By customizing maps this way, we improve the readability as well as increase the efficacy of the maps.

Getting ready

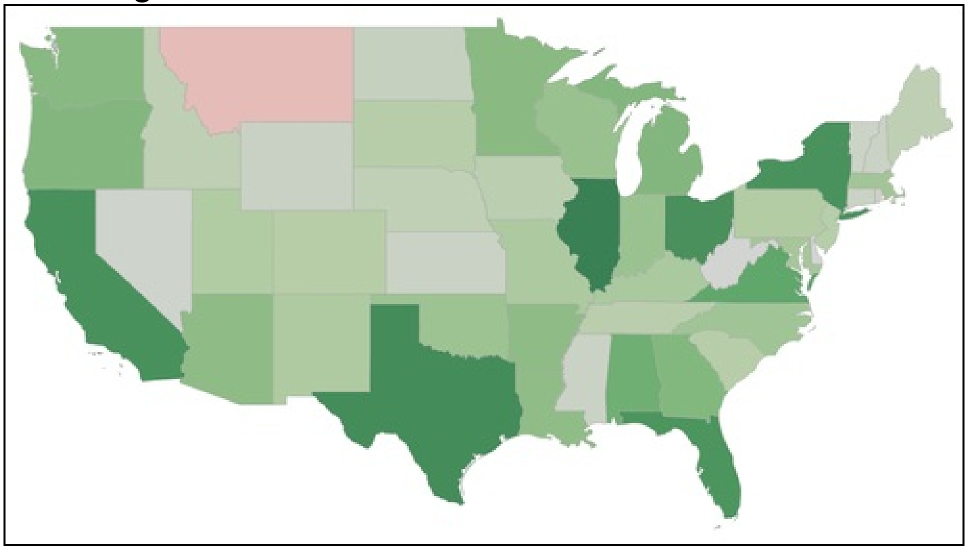
Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

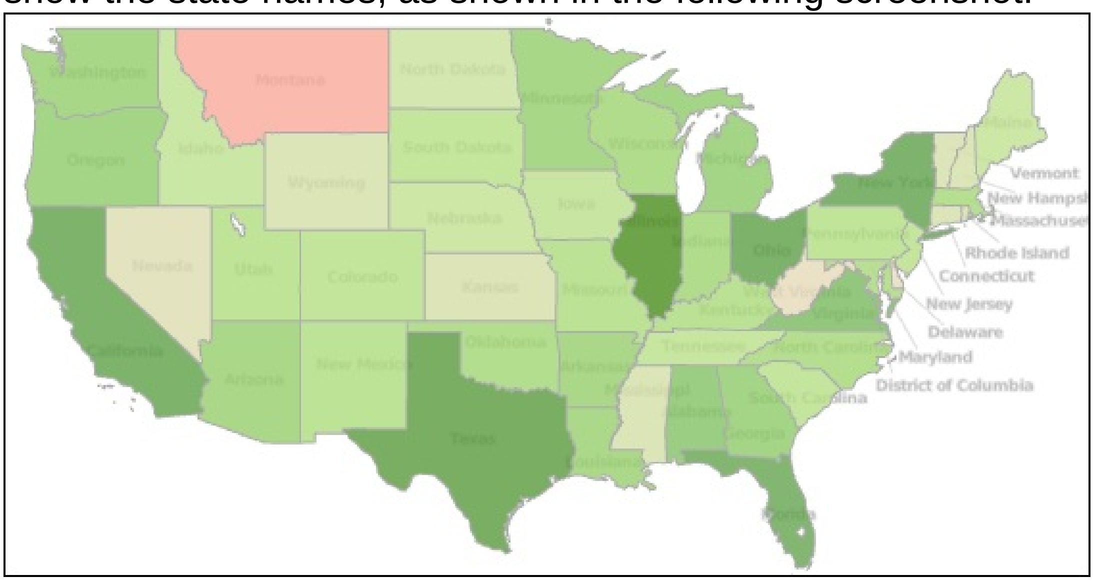
Once the data is loaded, perform the following steps to customize the generated map:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Profit** from **Measures** and **State** from **Dimensions**.
3. Click on **filled maps** on the **Show Me** toolbar to create a choropleth map.
4. From the main menu toolbar, click on **Map** and then on **Map Options**.
5. Note the difference in the background color of the map by changing the **Style** value from **Gray** to **Normal**.
6. Change the **Style** value back to **Gray** again.
7. Uncheck the box in front of **Base** to make the map look more clean and aesthetically pleasing.
8. Uncheck the box in front of **Light State Border & Names** to make the map look even more clean, since the map is already grouping the states.
9. Drag the **Washout** slider to 100 percent if you want to see only the

filled map with no other additional information, as shown in this following screenshot:



- Experiment with the **Washout** slider and check the box **State/Province Names** to create an informative yet good-looking map. For example, by changing the **Washout** slider to 40 percent and checking the box **State/Province Names**, we have minimized the distractions but still show the state names, as shown in the following screenshot:



How it works...

Tableau automatically selects the options that will work in most cases, but we can customize the options even further to make the maps aesthetically pleasing. This is achieved by changing the map layers, which are drawn from an online map provider or Tableau's offline maps. It is very similar to painting a layer of color on top of another on a canvas, with one big difference—our ability to add and remove layers as we please.

Chapter 6. Calculating User-defined Fields

In this chapter, we will cover the following recipes:

- Using predefined functions
- Calculating percentages
- Applying the If-Then logic
- Applying logical functions
- Showing totals
- Showing the percentage of totals
- Discretizing data
- Manipulating text
- Aggregating data

Introduction

Many a times we need to manipulate data in a certain way to generate the desired visualization or text. Tableau provides ways to calculate and create new fields, which could be used to enhance our visualization.

Using predefined functions

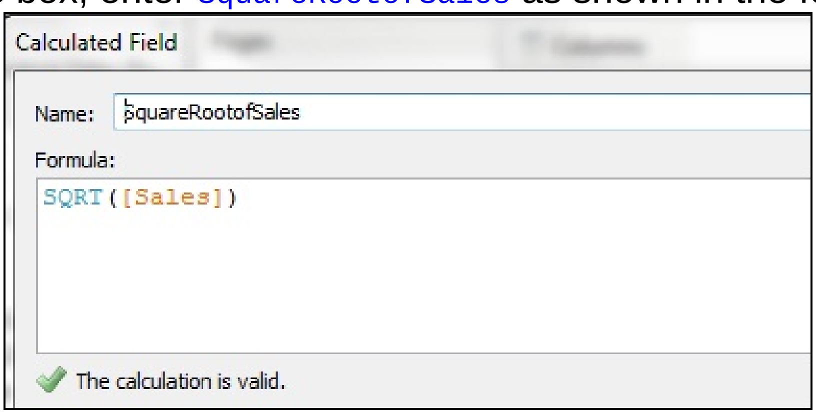
Tableau provides many predefined functions that help us manipulate data in a certain way. These functions are divided by the type of manipulation, such as functions of numeric and string data types, or aggregate operations.

Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to create a new calculated field:

1. To calculate the square root of a number, right-click on the **Measures** pane and select **Create Calculated Field**.
2. In the **Name** box, enter `SquareRootofSales` as shown in the following screshot: 

- In the **Formula** box, enter `SQRT([Sales])` and hit **OK**.
- To compare the square root of sales with the original sales, create a scatter plot of **Unit Price**, **Sales**, and **SquareRootofSales** by **Region**. From the main menu, click on **Analysis** and uncheck **Aggregate Measures**.

- Drag-and-drop **SquareRootofSales** and **Sales** in the **Columns** shelf.
- Drag-and-drop **Region** and **Unit Price** in the **Rows** shelf.
- Change the **Mark** type to **Circle** to create a chart similar to the one in the following screenshot:

How it works...

Using a predefined function, we created a field that houses the square root values of the **Sales** field. This newly created field can be used similarly for other existing fields, as shown in the created chart. Taking the square root of numeric values is a common data-transformation technique used to better observe the distribution of values, including outliers. You can see from the previous chart that the **SquareRootOfSales** values are spread more than the original sales values, which are more clustered around certain areas.

There's more...

You can learn more about data transformation objectives and techniques from the lecture slides of *Regression III: Advanced Methods Workshop* by *William Jacoby* at http://polisci.msu.edu/jacoby/icpsr/regress3/lectures/week1/4.Transformat

Calculating percentages

One of the most common type of measures is the percentage of a value within a population. Tableau provides options for converting values in a row or column to fractions of row or column totals; however, we can also create new fields with some calculated values presented as percentages.

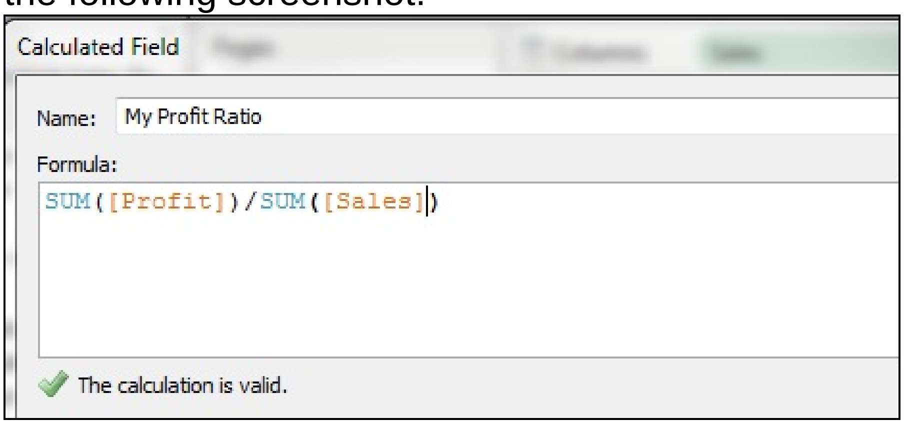
Getting ready

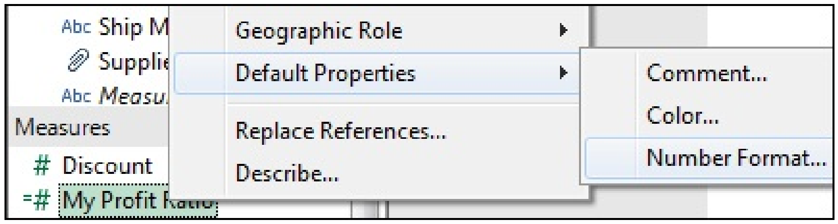
Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to create a new calculated field:

1. To calculate the profit to sales ratio, right-click on the **Measures** pane and select **Create Calculated Field**.
2. In the **Name** box, enter `My Profit Ratio`.
3. In the **Formula** box, enter `sum([Profit])/sum([Sales])` as shown in the following screenshot:



- Right-click on **My Profit Ratio** from the **Measures** pane, expand **Default Properties**, and select **Number Format** as shown in the following screenshot: 

- In the **Number Format** options box, select **Percentage** and hit **OK**.

How it works...

In the calculated field, we aggregated the **Profit** and **Sales** fields before dividing them. By calculating this way, we summed the **Profit** and **Sales** fields individually and then performed the division. This is different from dividing profit and sales first and then summing those values—this type of operation is a row-level operation and is suitable for many occasions. However, we have to remember to aggregate fields to avoid unexpected results.

Applying the If-Then logic

At times, it becomes necessary to report values in a certain way; for example, displaying blank values as dashes or categorizing some values into buckets. Since these modifications are based on logic, they are created using logical functions.

Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to create a new calculated field:

1. To categorize shipping modes into air and ground, right-click anywhere on the **Dimensions** pane and select **Create Calculated Field**.
2. In the **Name** box, enter `Ship Type`.
3. In the **Formula** box, enter `IF [Ship Mode] = 'Delivery Truck' then 'Ground' Else 'Air' End` and hit **OK**.

How it works...

Tableau provides seven logical functions to test logical conditions and return some values depending on the result of the condition. In the previous recipe, we checked whether the `Ship Mode` attribute was `Delivery Truck`, and if it were, we returned the `Ground` value as `Ship Type`. Since we know there are only three types of the `Ship Mode` attribute (`Delivery Truck`, `Express Air`, and `Regular Air`), we don't need to check for other shipping modes as they both are of the `Air` type.

Applying logical functions

Tableau provides various logical functions, such as `CASE`, `IF`, and `IIF`, to create calculated fields based on some conditions. In this recipe, we will create and see the use of some of these logical functions.

Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to create calculated fields based on conditions:

1. Right-click on **Product Type** from **Dimensions** and click on **Create Calculated Field**.
2. In the **Name** box, enter `Coffee or Tea`.
3. In the **Formula** box, enter the `CASE [Product Type] WHEN 'Coffee' THEN 'Coffee' WHEN 'Espresso' THEN 'Coffee' WHEN 'Herbal Tea' THEN 'Tea' ELSE 'Tea' END` formula and hit **OK**.
4. To use `0` instead of missing values of **Sales**, right-click on **Sales** from **Measures** and click on **Create Calculated Field**.
5. In the **Name** box, enter `Non-missing Sales`.
6. In the **Formula** box, enter `ZN([Sales])` and hit **OK**.
7. To see **Total Expenses** in some categories, right-click on **Total Expenses** from **Measures** and click on **Create Calculated Field**.
8. In the **Name** box, enter `Expensive Type`.
9. In the **Formula** box, enter the `IF [Total Expenses] <= 49.99 THEN 'Cheap' ELSEIF [Total Expenses] >= 50 and [Total Expenses] < 100 THEN 'Somewhat Expensive' ELSEIF [Total Expenses] >= 100 and [Total Expenses] < 150 THEN 'Slightly Expensive' ELSE 'Very Expensive' END` formula.

How it works...

The `CASE` and `IF` functions are similar in that they both allow testing of an expression and returning values on various conditions. The `CASE` function is usually easier to read and is usually the preferred way of testing expressions. The `IF` function allows us to test on numeric conditions whereas the `CASE` function doesn't allow that; for example, we cannot write a `CASE [Profit] < 100` condition, but we can write `IF [Profit] < 100`. The `IFNULL` function is very useful when we want to return any value (numbers in case of numeric expressions and a string in case of string expressions) if the expression is null, and the `ZN` function is useful when we want to return `0` if the expression is null.

Showing totals

Although it is useful to show the breakdown of measures by various dimensions, readers value seeing grand totals for rows and columns. If at least one **Columns** or **Rows** value is present, it is very easy to show grand totals.

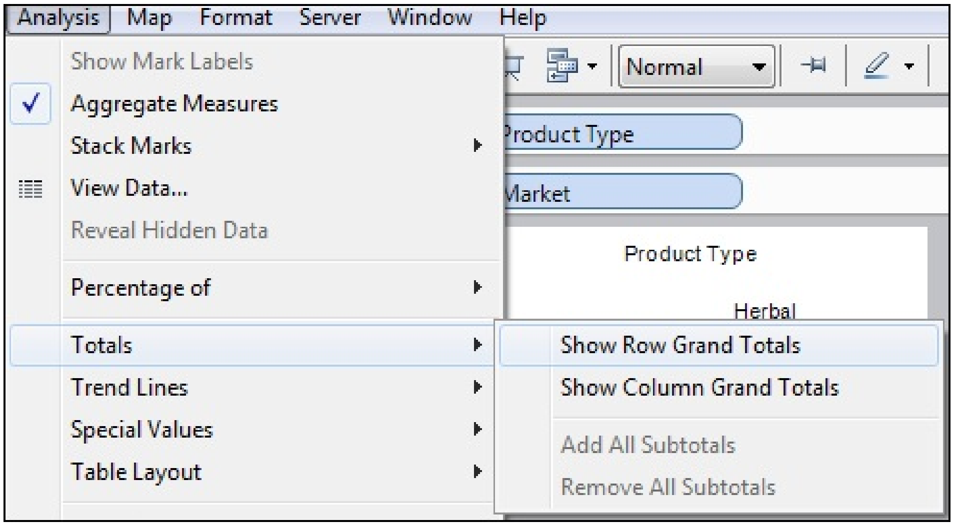
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

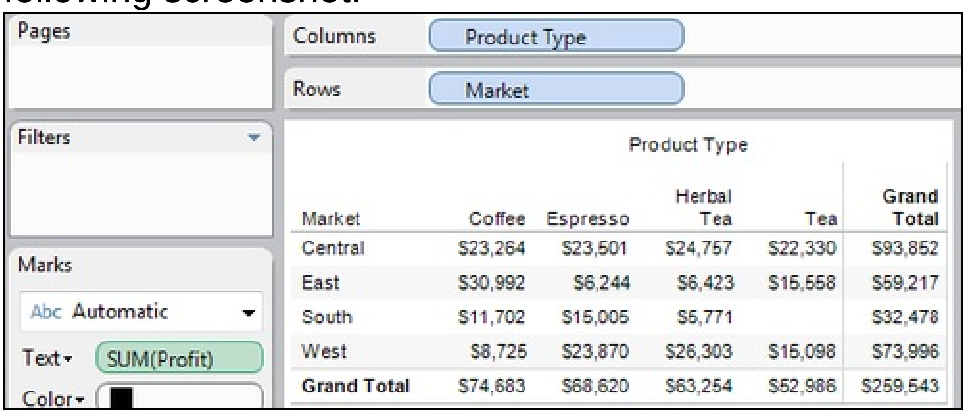
How to do it...

Once the data is loaded on the worksheet, perform the following steps to show grand totals:

1. Drag-and-drop **Product Type** from **Dimensions** in the **Columns** shelf.
2. Drag-and-drop **Market** from **Dimensions** in the **Rows** shelf.
3. Drag-and-drop **Profit** from **Measures** in the **Text** box under the **Marks** pane.
4. From the top menu bar, click on **Analysis**, expand **Totals**, and click on **Show Row Grand Totals** as shown in the following screenshot:



- Again, click on **Analysis**, expand **Totals**, and click on **Show Column Grand Totals** to see both the column and row totals as shown in the following screenshot:



Showing the percentage of totals

Seeing the percentage of each group as compared to the total of all groups is as useful as seeing totals and breakdowns. This gives the reader an idea about the magnitude of every value compared to that of the totals. Tableau offers various options to see values as percentages of the totals. These options include seeing percentages of the row totals, the column totals, or the grand total.

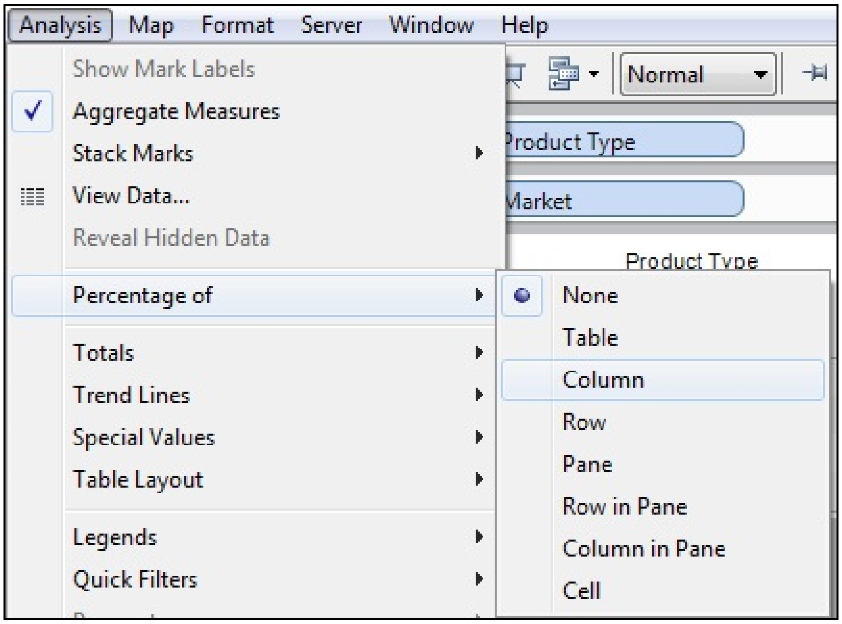
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

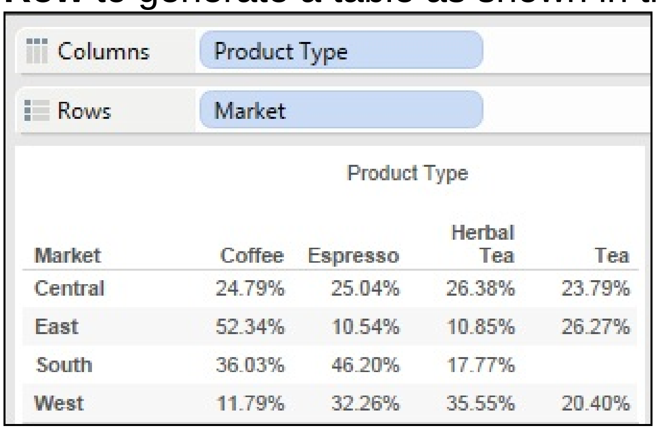
How to do it...

Once the data is loaded on the worksheet, perform the following steps to see percentage values:

1. Drag-and-drop **Product Type** from **Dimensions** in the **Columns** shelf.
2. Drag-and-drop **Market** from **Dimensions** in the **Rows** shelf.
3. Drag-and-drop **Profit** from **Measures** in the **Text** box under the **Marks** pane.
4. To see percentage profit by every **Product Type** in all the **Market** types, expand the **Analysis** menu option from the main menu toolbar. Then, expand the **Percentage of** option and select **Column** as shown in the following screenshot:

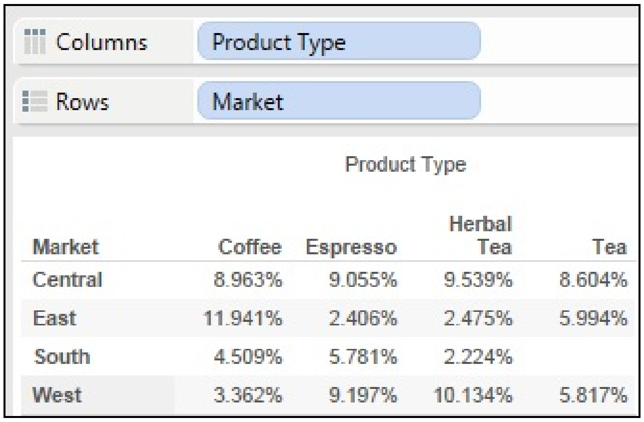


- To see the percentage profit by every **Market** type in all the **Product Type** values, expand the **Analysis** menu option from the main menu toolbar, followed by expanding the **Percentage of** option, and select **Row** to generate a table as shown in the following screenshot:



Market	Coffee	Espresso	Herbal Tea	Tea
Central	24.79%	25.04%	26.38%	23.79%
East	52.34%	10.54%	10.85%	26.27%
South	36.03%	46.20%	17.77%	
West	11.79%	32.26%	35.55%	20.40%

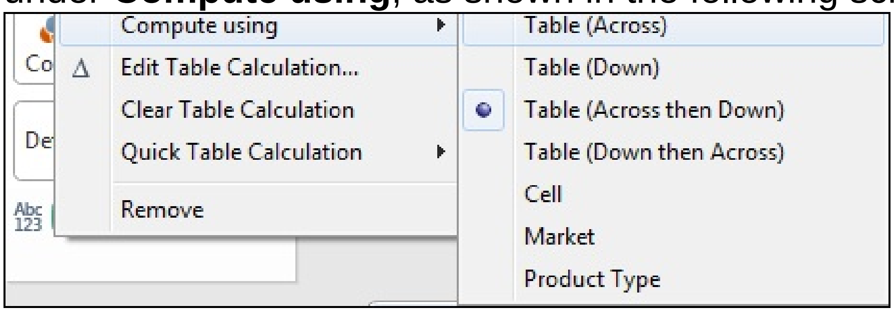
- To see profit by every **Market** and **Product Type** values as a fraction of the total **Profit** value, expand the **Analysis** menu option from the main menu toolbar, followed by the **Percentage of** option, and select **Table** to generate a percentage table as shown in the following screenshot:



Market	Coffee	Espresso	Herbal Tea	Tea
Central	8.963%	9.055%	9.539%	8.604%
East	11.941%	2.406%	2.475%	5.994%
South	4.509%	5.781%	2.224%	
West	3.362%	9.197%	10.134%	5.817%

There's more...

Tableau provides options to change how percentages are calculated, such as across, down, by cell, or by other fields. These options can be changed by clicking on the aggregated measure and expanding options under **Compute using**, as shown in the following screenshot:



Discretizing data

Sometimes we require discretizing (or binning) of numeric data for pretty labeling or meeting some format guidelines; for example, you may need to report the sales amount in thousands, and thus you will need to create a field that will put every sales amount in various bins, for example, 0-1000, 1000-2000, and so on.

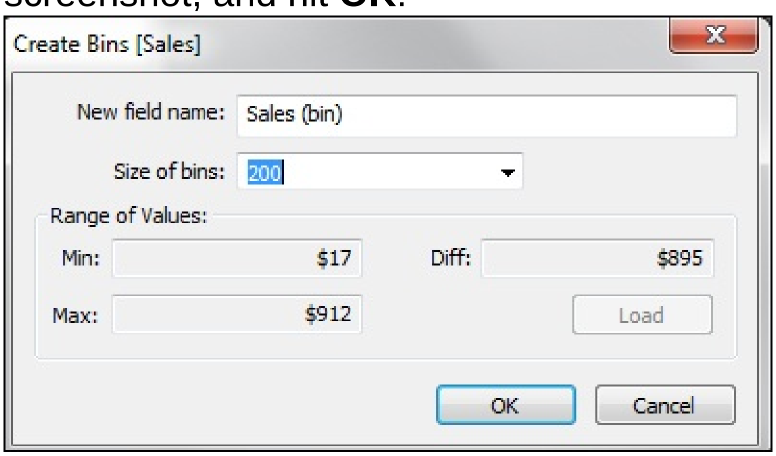
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

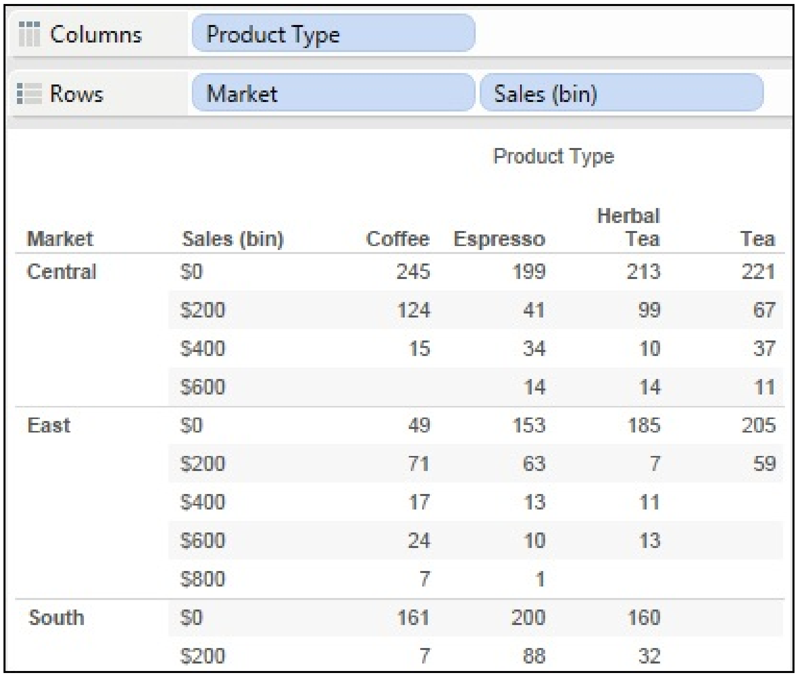
Once the data is loaded on the worksheet, perform the following steps to discretize a numeric value or create bins:

1. Right-click on **Sales** from **Measures** and select the **Create Bins** option.
2. Hit the **Load** button to see the distribution of the **Sales** amount.
3. In the **Size of bins** box, enter `200`, as shown in the following screenshot, and hit **OK**:



- Drag-and-drop **Sales (bin)** from **Dimensions** in the **Rows** shelf.

- Drag-and-drop **Product Type** from **Dimensions** in the **Columns** shelf.
- Drag-and-drop **Market** from **Dimensions** in the **Rows** shelf, but place it before **Sales (bin)**.
- Drag-and-drop **Number of Records** from **Measures** in the **Text** box under the **Marks** pane to create a table similar to the one shown in the following screenshot:



Market	Sales (bin)	Coffee	Espresso	Herbal Tea	Tea
Central	$0	245	199	213	221
	$200	124	41	99	67
	$400	15	34	10	37
	$600		14	14	11
East	$0	49	153	185	205
	$200	71	63	7	59
	$400	17	13	11	
	$600	24	10	13	
	$800	7	1		
South	$0	161	200	160	
	$200	7	88	32	

How it works...

When we hit the **Load** button on the **Create Bins [Sales]** dialog box, Tableau loads the distribution; that is, minimum, maximum, and the difference between the minimum and maximum value of the underlying measure. By looking at those values, we can decide the appropriate number of bins. Once we enter the number of bins, Tableau puts all the individual values of the underlying measure into bins, which start with zero and end with the highest possible value of the range that doesn't exceed the maximum value. For this recipe, we had the maximum value of 895 dollars and the maximum value of the bins was 800 dollars, because the next bin value would be 1,000 dollars and there are no values that are over 1,000 dollars.

Manipulating text

At times, we are required to parse or manipulate text variables to get something meaningful out of those variables; for example, a **Full Name** field may contain both the first name and last name of a sales representative, but our reporting standards may require us to show two different columns for the first and last names. With Tableau's string operators, we can easily manipulate the text to meet our requirements.

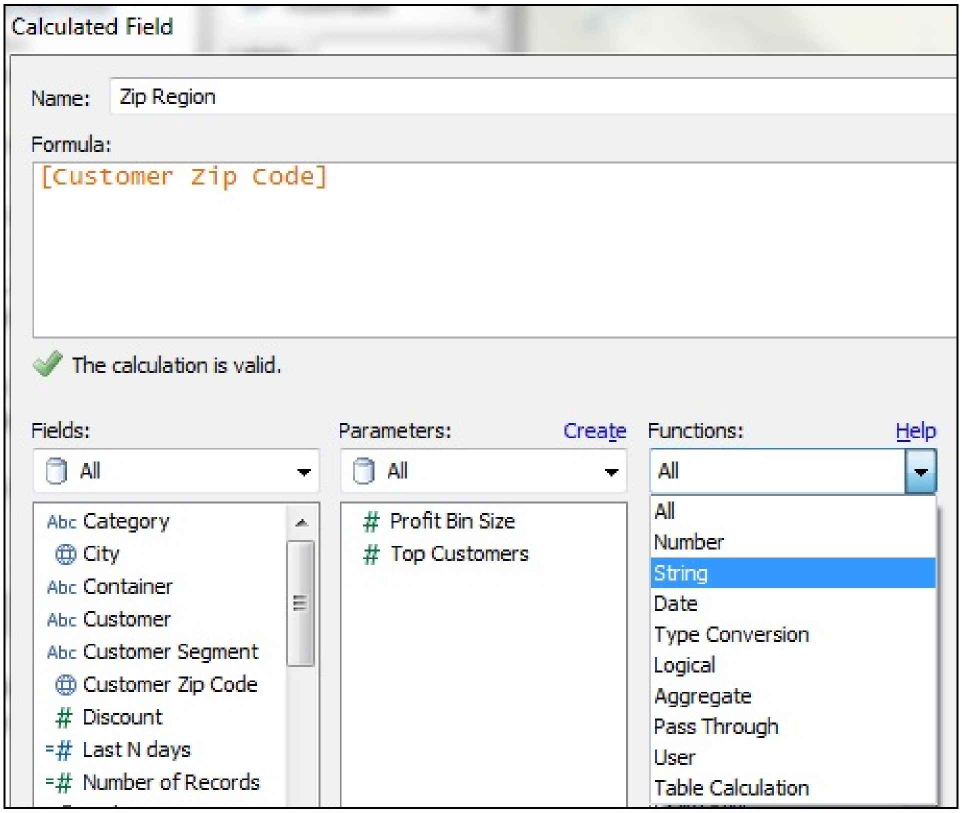
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

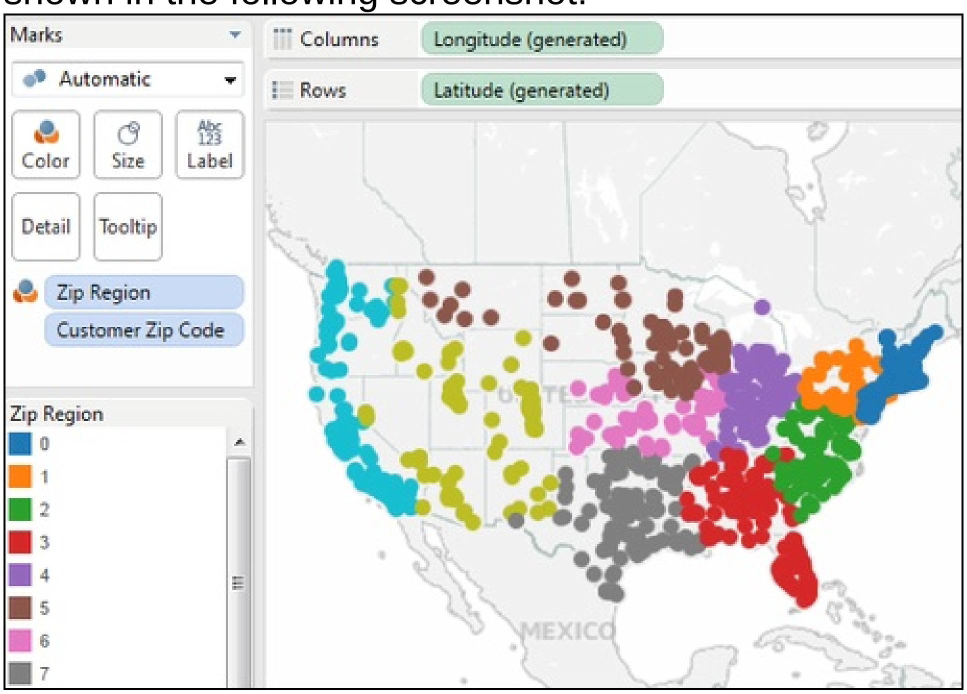
How to do it...

Once the data is loaded on the worksheet, perform the following steps to create new string fields based on existing text variables:

1. Right-click on **Customer Zip Code** from **Dimensions**, and select **Create Calculated Field**.
2. In the **Name** box, type `Zip Region`.
3. From the **Functions** dropdown, select **String**, as shown in the following screenshot:



- Find the `LEFT` function and double-click on it.
- Adjust the formula in the **Formula** box to `LEFT([Customer Zip Code], 1)` and hit **OK**. We can use this newly generated field to create a map as shown in the following screenshot:



- To extract the customer's last name, right-click on **Customer** from **Dimensions** and select **Create Calculated Field**.
- In the **Name** box, enter `Customer Last Name`.
- In the **Formula** box, enter `RIGHT([Customer], LEN([Customer]) - FIND([Customer], " "))` and hit **OK**.

How it works...

The `LEFT` function extracts the specified number of characters from the start of the given string variable. In our recipe, we extracted the first character of the **Customer Zip Code** value. The `RIGHT` function works similarly except that it extracts characters from the end of the given string variable. The `FIND` function returns the position of the searched string within a string variable. To extract the customer's last name, we first found the position of the space between the customer's first and last names in the **Customer** field. Then we computed the number of characters between the space and the end of the string by subtracting the position of the space from the total number (found using the `LEN` function) of characters in the **Customer** field.

Aggregating data

Although the type of aggregation of a measure can be changed from the **Marks** pane, it is sometimes necessary to show different aggregations of the same measure, and we can do this by creating multiple aggregate fields. We can also add the same **Measure** field multiple times to the **Rows** or **Columns** shelf and then change the aggregation type.

Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

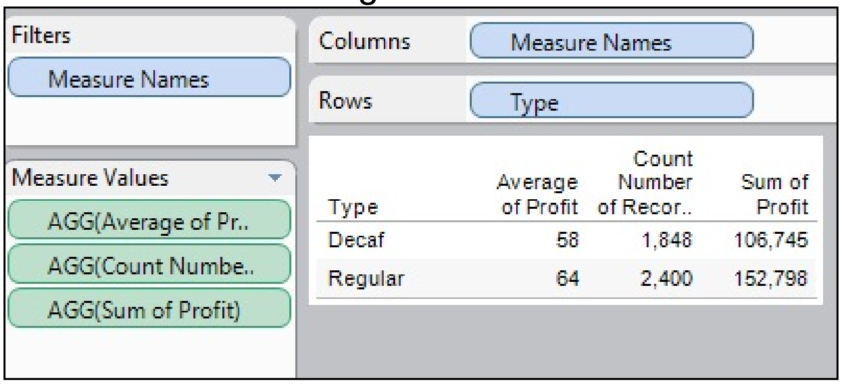
How to do it...

Once the data is loaded on the worksheet, perform the following steps to create calculated fields with different aggregations:

1. Right-click on **Profit** from **Measures**, and click on **Create Calculated Field**.
2. In the **Name** box, enter `Sum of Profit`.
3. In the **Formula** box, enter the `SUM([Profit])` formula and hit **OK**.
4. Right-click on **Profit** from **Measures** and click on **Create Calculated Field**.
5. In the **Name** box, enter `Average of Profit`.
6. In the **Formula** box, enter the `AVG([Profit])` formula and hit **OK**.
7. Right-click on **Number of Records** from **Measures**, and click on **Create Calculated Field**.
8. In the **Name** box, enter `Count Number of Records`.
9. In the **Formula** box, enter the `COUNT([Number of Records])` formula and hit **OK**.
10. Click on the **Show Me** button to display the **Show Me** toolbar on the screen.
11. Select **Sum of Profit**, **Average of Profit**, and **Count Number of**

Records from **Measures** and **Type** from **Dimensions**.

12. Click on **text tables** on the **Show Me** toolbar to create a table as shown in the following screenshot:



Chapter 7. Customizing and Saving

In this chapter, we will cover the following recipes to customize and save files:

- Adding title and caption
- Modifying font sizes and colors
- Applying various marks
- Adding colors
- Adding labels
- Changing marks sizes
- Adding reference lines
- Printing to PDF
- Saving packaged workbooks
- Creating a workbook data extract

Introduction

Once you have generated insightful graphics, you would want to customize it first and share it with others. Tableau offers various customization options including modifying font sizes and colors, applying various marks, adding labels, and others.

Adding title and caption

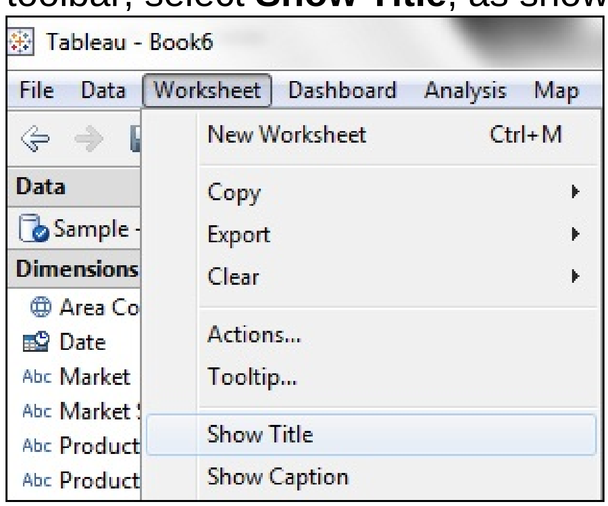
Adding as much information as is possible to the title and caption that describe or summarize some of the important points of the visualization helps readers understand the visualization better. If you have applied filters, Tableau will automatically describe the filters in the caption area, but as a designer, you will have to describe any other important aspects of the visualization.

Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

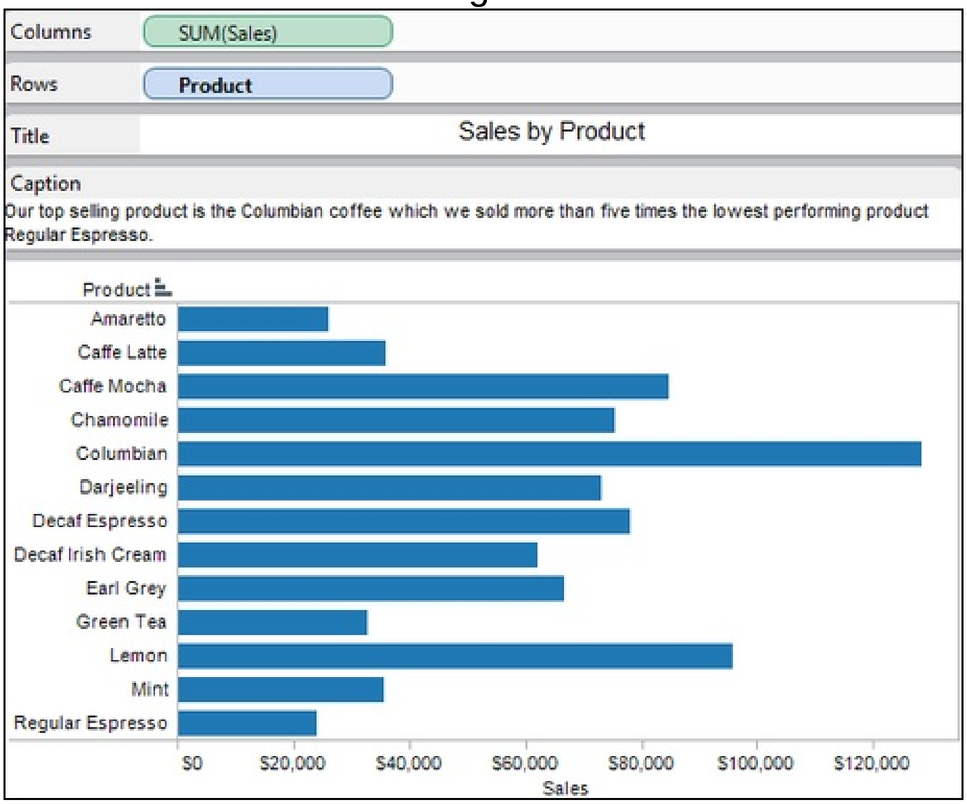
How to do it...

Once the data is loaded on the worksheet, perform the following steps to add a tittle and a caption:

1. Drag-and-drop **Sales** from **Measures** into the **Columns** shelf.
2. Drag-and-drop **Product** from **Dimensions** into the **Rows** shelf.
3. From the main **Menu** toolbar, select **Show Title**, as shown in the following screenshot: 

- From the main menu toolbar, select **Show Caption**.

- Double-click anywhere in the area next to **Title** and in the **Edit Title** box enter `Sales by Product`, and hit **OK**.
- Double-click anywhere in the area next to **Caption** and in the **Edit Caption** box enter `Our top selling product is the Columbian coffee which we sold more than five times the lowest performing product Regular Espresso`, and hit **OK**.
- Move the **Caption** box below the **Title** box and the final visualization should look like the following screenshot:



Modifying font sizes and colors

Tableau provides options to modify font sizes and colors for the whole worksheet or individual components of the worksheet, such as the pane, headers, tooltip, and grand total. Although the default scheme is good enough to be used in production-quality material, there might be instances where you would want to customize these options.

Getting ready

To customize the font and color for this recipe, repeat the *Adding title and caption* recipe.

How to do it...

Once you have recreated the graphic, perform the following steps to customize the font and color:

1. From the main menu toolbar, click on **Format** and then select **Font**.
2. Make sure that the **Format Font** button, which has the letter A in its icon, is highlighted.
3. To modify all the fonts on the worksheet, select a different font size and font from the **Worksheet** dropdown.
4. To modify the font color of the header, select a different font color from the **Header** dropdown.

Applying various marks

Tableau provides various ways to encode data using different marks including **Square**, **Circle**, and **Shape**. The **Square** and **Circle** marks will show data points using a square or circle shape. By using the **Shape** mark, however, we can assign some attribute values to various shapes and help the reader distinguish data points by those shapes.

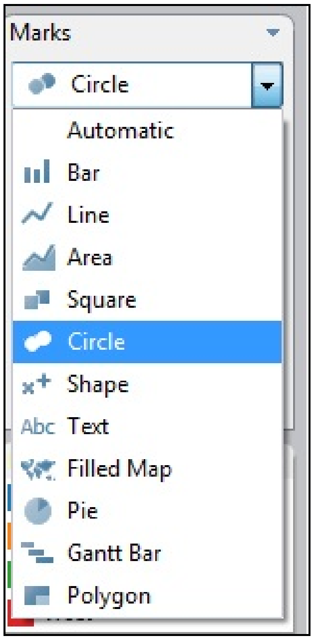
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

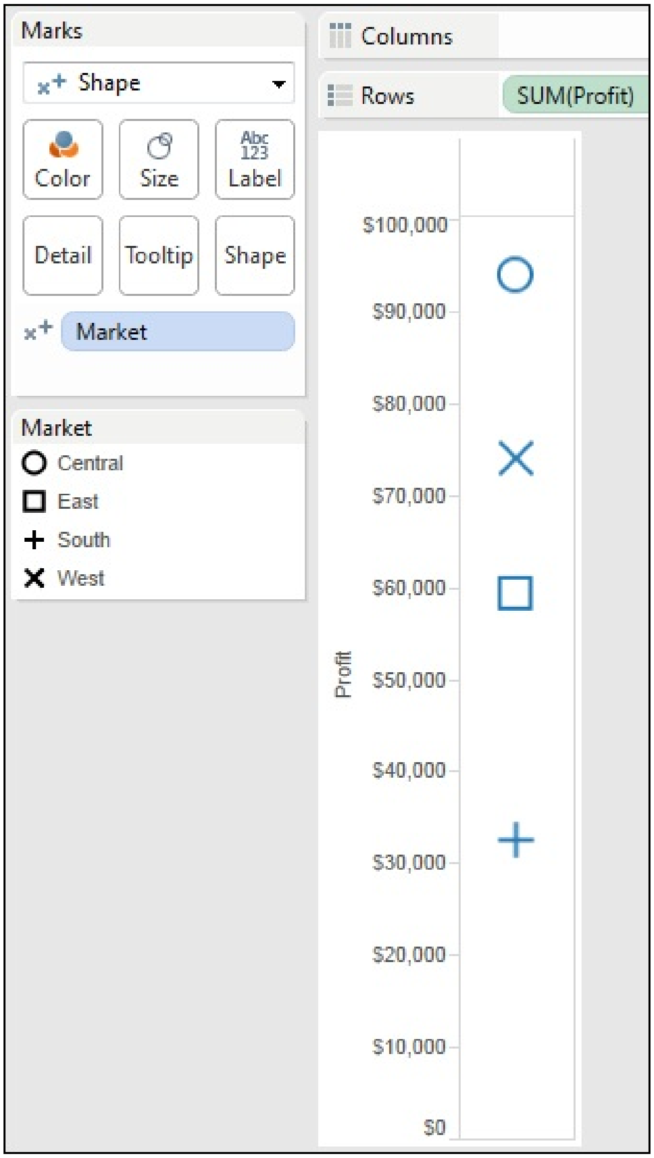
How to do it...

Once the data is loaded on the worksheet, perform the following steps to use various marks to denote the data points:

1. While holding the *Ctrl* key, click on **Market** from **Dimensions** and **Profit** from **Measures**.
2. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
3. Click on **circle views** on the **Show Me** toolbar.
4. To change the shape, from the **Marks** pane, select **Square** or **Circle** from the dropdown as shown in the following screenshot:



- To use a different shape for every **Market** value, select **Shape** from the dropdown in the **Marks** pane.
- Drag-and-drop **Market** from **Color** into the **Shape** box to generate a chart shown in the following screenshot:



How it works...

Although the **Square**, **Circle**, and **Shape** options may appear similar, they provide different ways to view and distinguish various data points. That is especially true in the case of the **Shape** option: as every attribute value is given a shape, it becomes easy to identify and note the different data points. In the case of many data points or many categories, however, identification and distinction of data points even with different shapes is challenging. To assist the readers, you should limit shapes to three to four attribute values.

Adding colors

Depending on the type of chart you created, Tableau may or may not color code any data. Using the **Color** box in the **Marks** pane, however, you could easily add colors to your graphs.

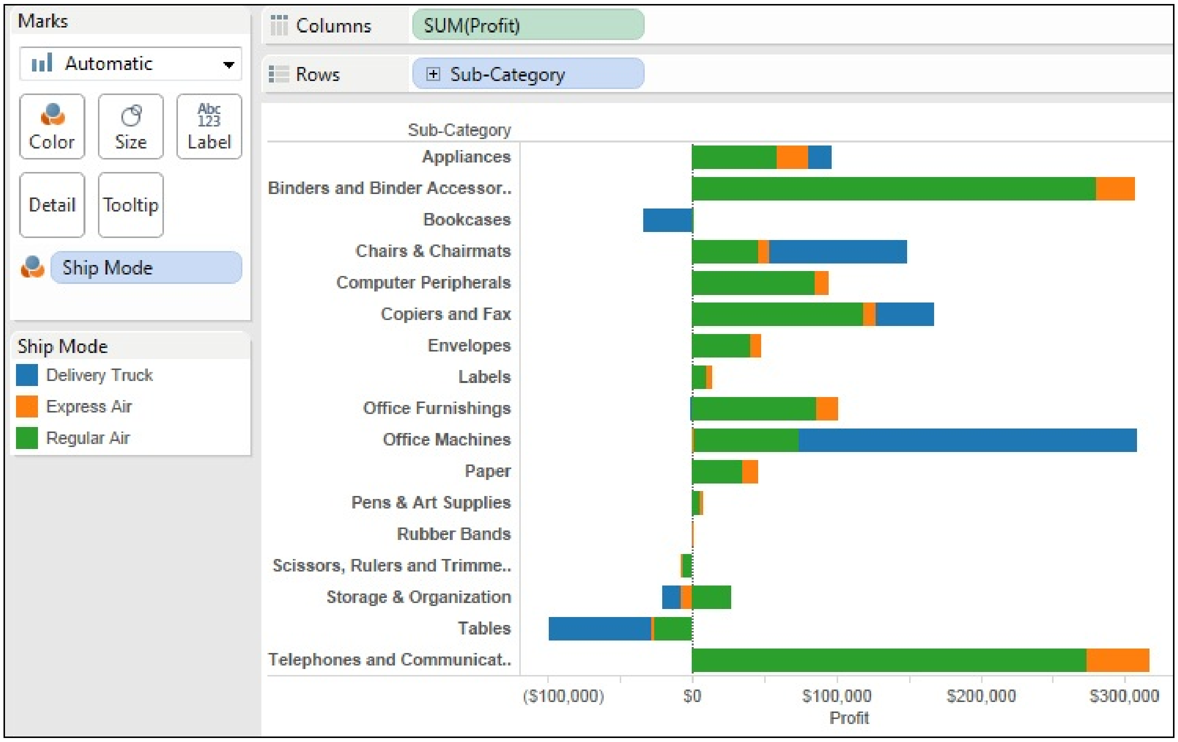
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

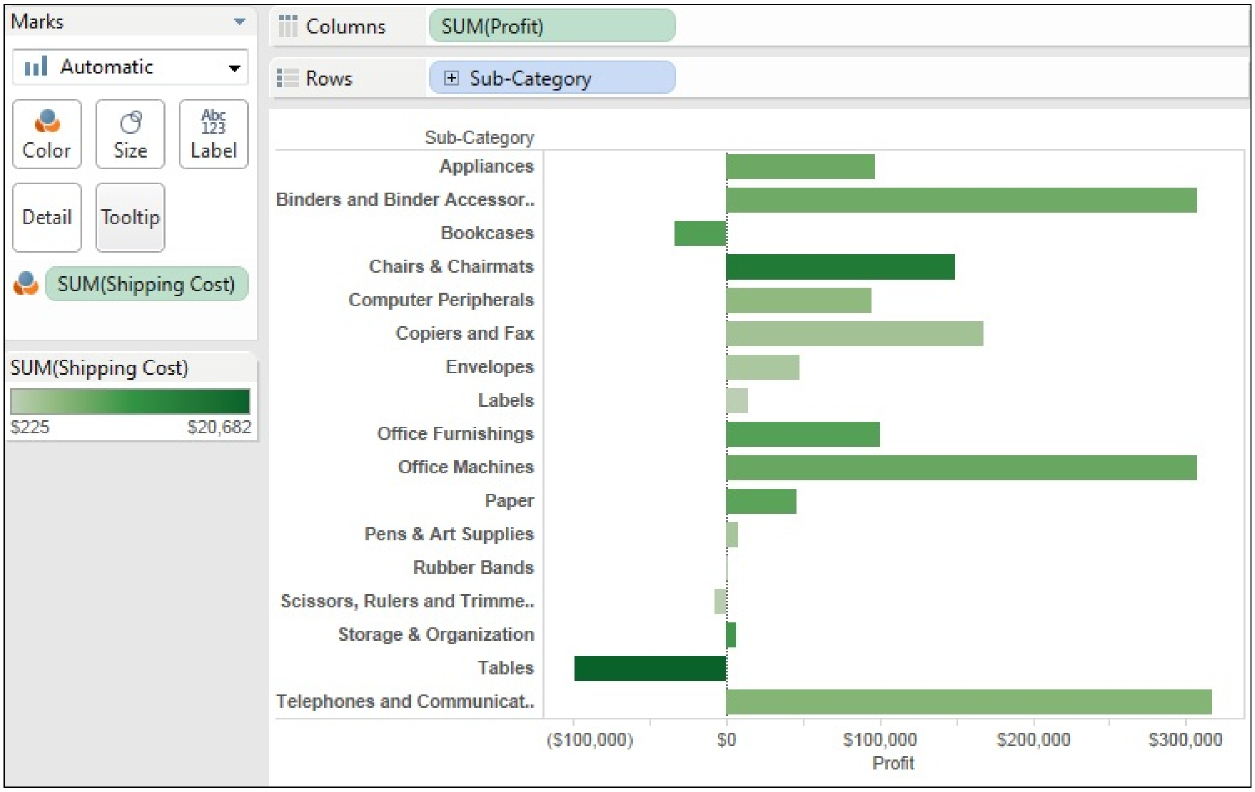
How to do it...

Once the data is loaded on the worksheet, perform the following steps to add colors to encode the data:

1. Drag-and-drop **Profit** from **Measures** into the **Columns** shelf.
2. Drag-and-drop **Sub-Category** from **Dimensions** into the **Rows** shelf.
3. To see the **Profit** values by **Ship Mode**, drag-and-drop **Ship Mode** in the **Color** box in the **Marks** pane to generate a chart shown in the following screenshot:



- To see the **Profit** values by **Customer Segment**, drag-and-drop **Customer Segment** in the **Color** box in the **Marks** pane.
- To compare **Profit** and **Shipping Cost** values, drag-and-drop **Shipping Cost** in the **Color** box in the **Marks** pane to generate a chart shown in the following screenshot. You can see that tables generated a loss and had the highest shipping costs, whereas chairs and chair mats were profitable, though this category incurred the second-highest shipping costs:



How it works...

When you drag a dimension to the **Color** box, Tableau automatically selects a color palette to draw colors from, and these colors are very distinctive and use different hues. When you drag a measure to the **Color** box, however, Tableau selects a gradient of a single color, and bigger data points (larger values) are encoded by a darker gradient, and smaller data points (smaller values) are encoded by a lighter gradient. It is important to note, however, that the underlying value of the dimension or measure also dictates what type of color palette would be chosen: a dimension of a continuous type of data can generate a gradient palette, and a discrete measure can generate a discrete color palette.

Adding labels

Although adding labels to data points is sometimes redundant, Tableau makes it very easy to add labels to your graphs. With the help of data labels, readers of the graph are able to read the exact value of the data point instead of speculating about the values by gauging the heights of bars or sizes of shapes. However, as a designer of a visualization, you must ask this question to yourself: if data labels are important to your graph, can you replace the graph with a simple table to provide all the details?

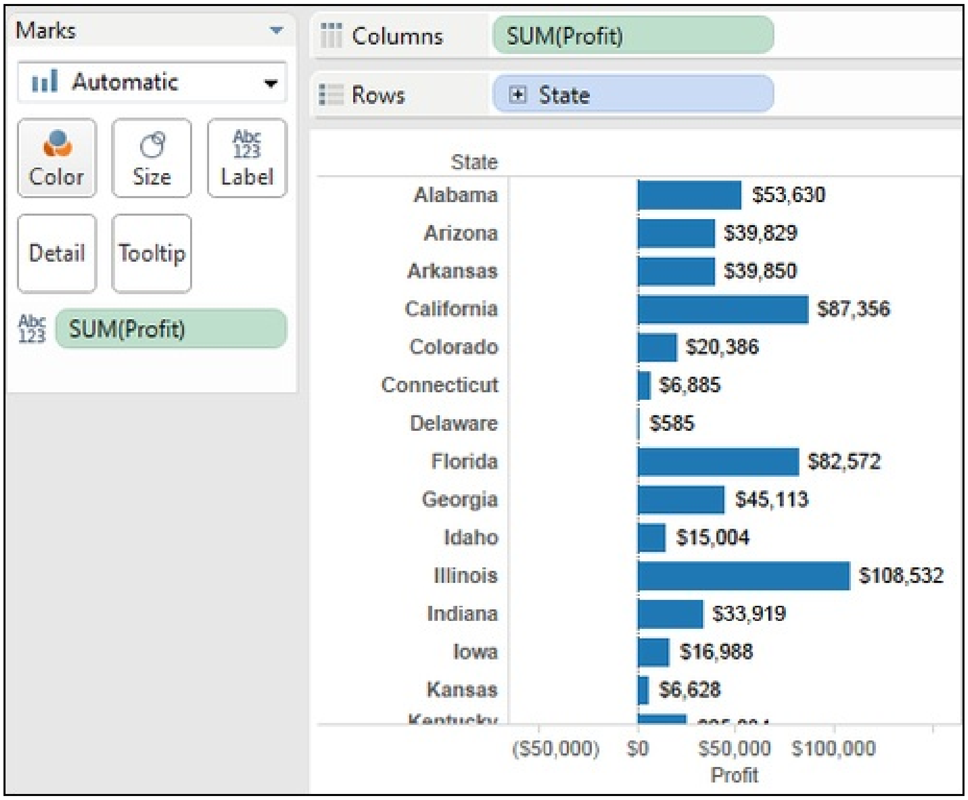
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to add the data labels to your graph:

1. Drag-and-drop **State** from **Dimensions** into the **Rows** shelf.
2. Drag-and-drop **Profit** from **Measures** into the Columns shelf
3. Drag-and-drop **Profit** from **Measures** again into the **Label** box in the **Marks** pane to display data labels next to the bars, as shown in the following screenshot:



How it works...

In the simple cases, as we have seen in this recipe, Tableau places the value next to the marks. In overlapping data points cases, Tableau will hide some of the labels to increase the clarity of the graphic. Instead of repeating the value encoded in a mark, it is possible to show a completely different measure as a label. This could be misleading, however, and could confuse readers.

Changing marks sizes

Since we can tell differences in sizes easily compared to differences in colors, encoding data in various sizes of marks will increase the effectiveness of a graph. If the differences in data points are hard to observe, then a different type of visualization might be needed.

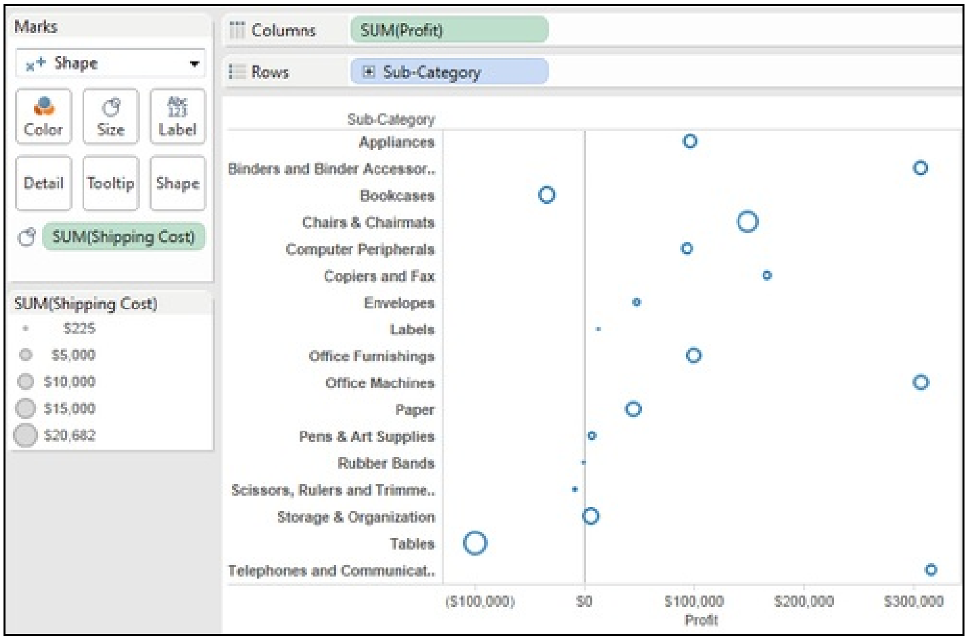
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded, perform the following steps to show marks sizes encoded by the values of an attribute:

1. Drag-and-drop **Sub-Category** from **Dimensions** into the **Rows** shelf.
2. Drag-and-drop **Profit** from **Measures** into the **Columns** shelf.
3. Drag-and-drop **Shipping Cost** from **Measures** into the **Size** box in the **Marks** pane.
4. Change the mark type to **Shape** to see the mark size vary by **Shipping Cost**.
5. Click on the **Size** box to increase the size of shapes by dragging the visible slider. The final chart should look like the one in the following

screenshot: 
Marks
Shape
Color
Size
Label
Detail
Tooltip
Shape
SUM(Shipping Cost)
SUM(Shipping Cost)
$225
$5,000
$10,000
$15,000
$20,682
Columns
SUM(Profit)
Rows
Sub-Category
Sub-Category
Appliances
Binders and Binder Accessor..
Bookcases
Chairs & Chairmats
Computer Peripherals
Copiers and Fax
Envelopes
Labels
Office Furnishings
Office Machines
Paper
Pens & Art Supplies
Rubber Bands
Scissors, Rulers and Trimme..
Storage & Organization
Tables
Telephones and Communicat..
($100,000)
$0
$100,000
$200,000
$300,000
Profit

Adding reference lines

Among many of Tableau's features, adding reference lines to graphs is one of them. By adding reference lines, we can compare data points with either any constant value or any statistical computation such as average of the measure values.

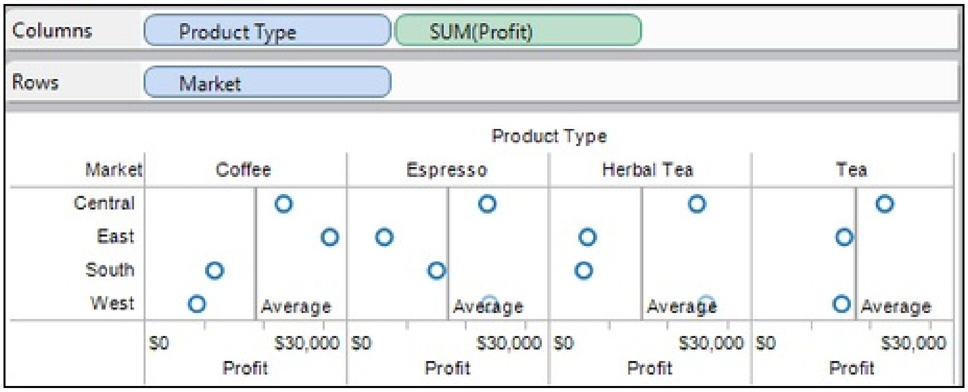
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

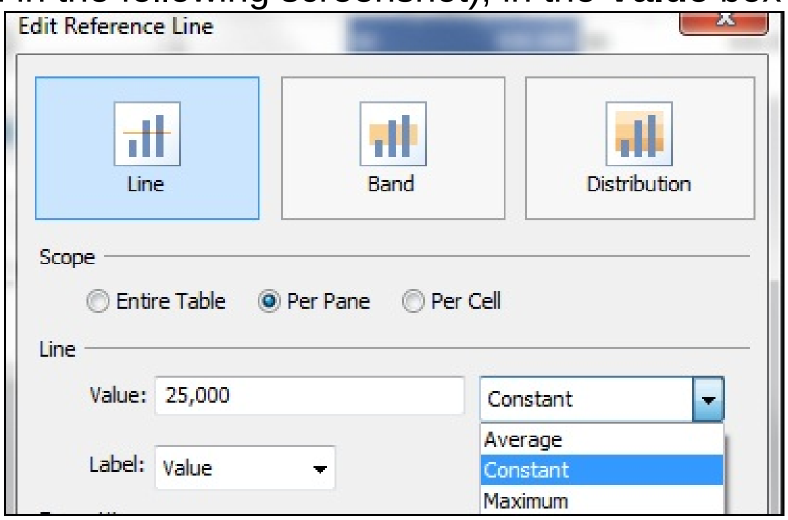
How to do it...

Once the data is loaded on the worksheet, perform the following steps to add various reference lines:

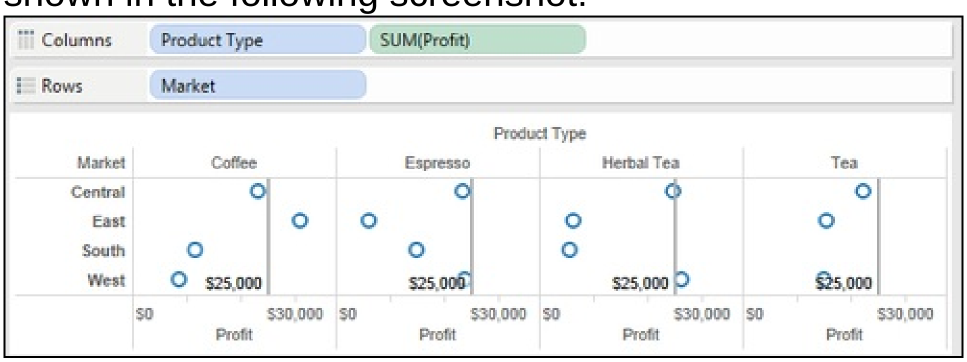
1. Drag-and-drop **Product Type** from **Dimensions** into the **Columns** shelf.
2. Drag-and-drop **Profit** from **Measures** into the **Columns** shelf.
3. Drag-and-drop **Market** from **Measures** into the **Rows** shelf.
4. Change the mark type to **Shape**.
5. To compare **Profit** values for each **Product Type** and **Market** value with the average **Profit** value for all markets and product types, right-click on the **Profit** axis, select **Add Reference Line**, keep the **Scope** option value to **Per Pane**, accept all the default values, and hit **OK**. The graph should look like the one in the following screenshot:



- To compare the **Profit** values to a fixed value, right-click on the **Profit** axis, select **Edit Reference Line**, keep the **Scope** value at **Per Pane**, under the **Line** selections, change the drop-down value to **Constant** from **Average** (as shown in the following screenshot), in the **Value** box enter `25,000`, and hit **OK**:



- The chart with a constant reference line should look similar to the one shown in the following screenshot:



Printing to PDF

One of the easiest ways to share your Tableau graphs is to save them as PDF files. Printing graphs to PDF files is built into Tableau, and you do not require any additional software to print the PDF files.

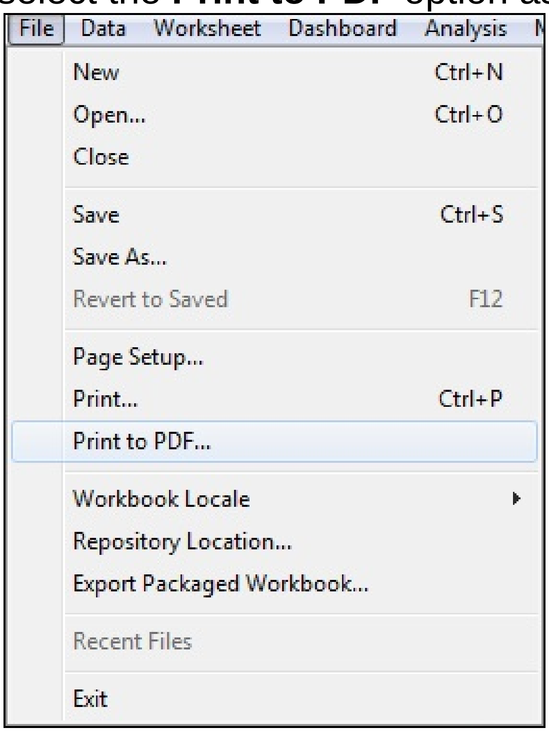
Getting ready

To create a nice looking PDF, repeat the *Adding title and caption* recipe.

How to do it...

Once you have a graph with a title and caption, perform the following steps to print it to PDF:

1. From the main menu toolbar, select **File**.
2. Under **File**, select the **Print to PDF** option as shown in the following screenshot:



- Select the **Active Sheet** button in the **Range** options box.
- Select the **Landscape** button in the **Paper Size** box and hit **OK**.

- Select a folder to save the file in and enter a filename in the **Save As** dialog box, and hit **Save**.

Saving packaged workbooks

When you use local data sources, such as using sample files, Excel, the Access files, or text files, to create a Tableau workbook, sharing could become a challenge when your users do not have access to those data sources. We could overcome such a problem by saving packaged workbooks, which have the workbook as well as the local data.

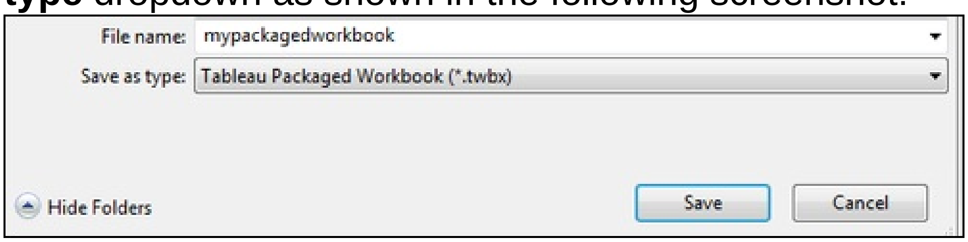
Getting ready

Follow the *Adding labels* recipe to create a workbook using an Excel data source.

How to do it...

Once you have created the workbook with the graph and labels, perform the following steps to save the workbook as a packaged workbook:

1. From the main menu bar, click on **File** and then on **Save As**.
2. Enter a name for the workbook in the **File name** box
3. Select **Tableau Packaged Workbook (*.twbx)** from the **Save as type** dropdown as shown in the following screenshot:



- Hit **Save**.

Creating a workbook data extract

If your data is coming from some connected data sources, extracting such data from a workbook will let you work on that data even if you are disconnected. The extracts are also useful when dealing with large data files as you can apply filters to select only a few rows (based on conditions).

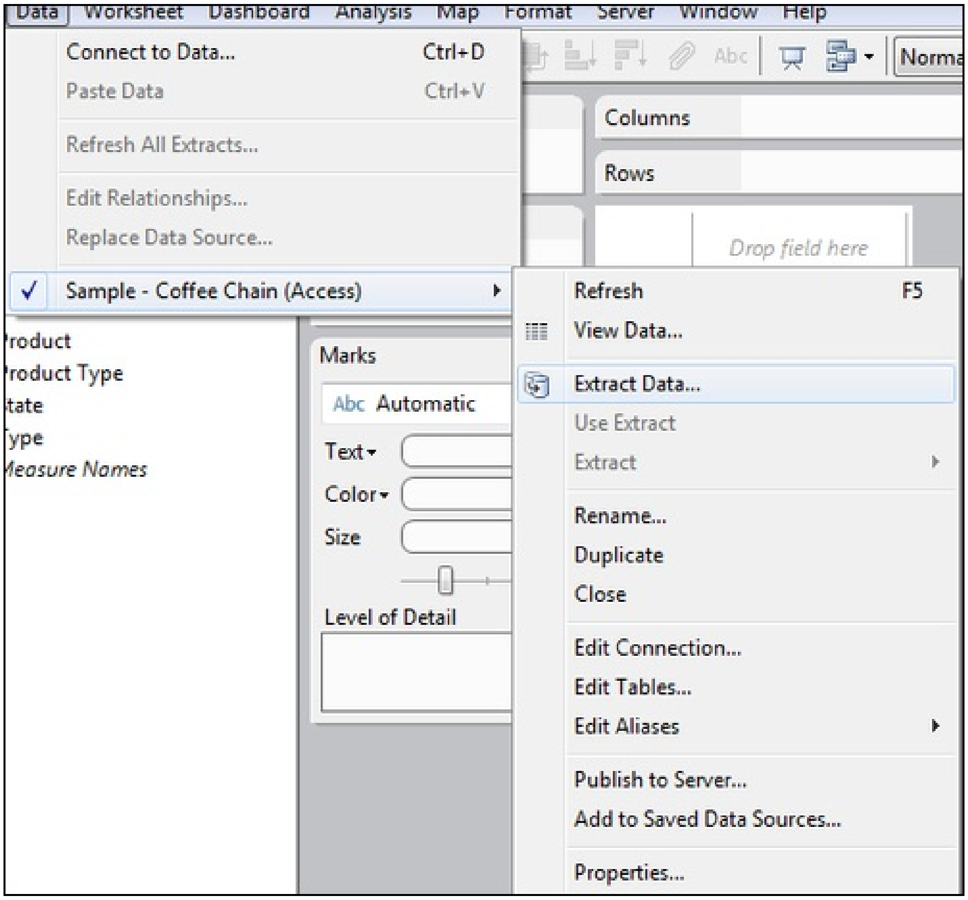
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

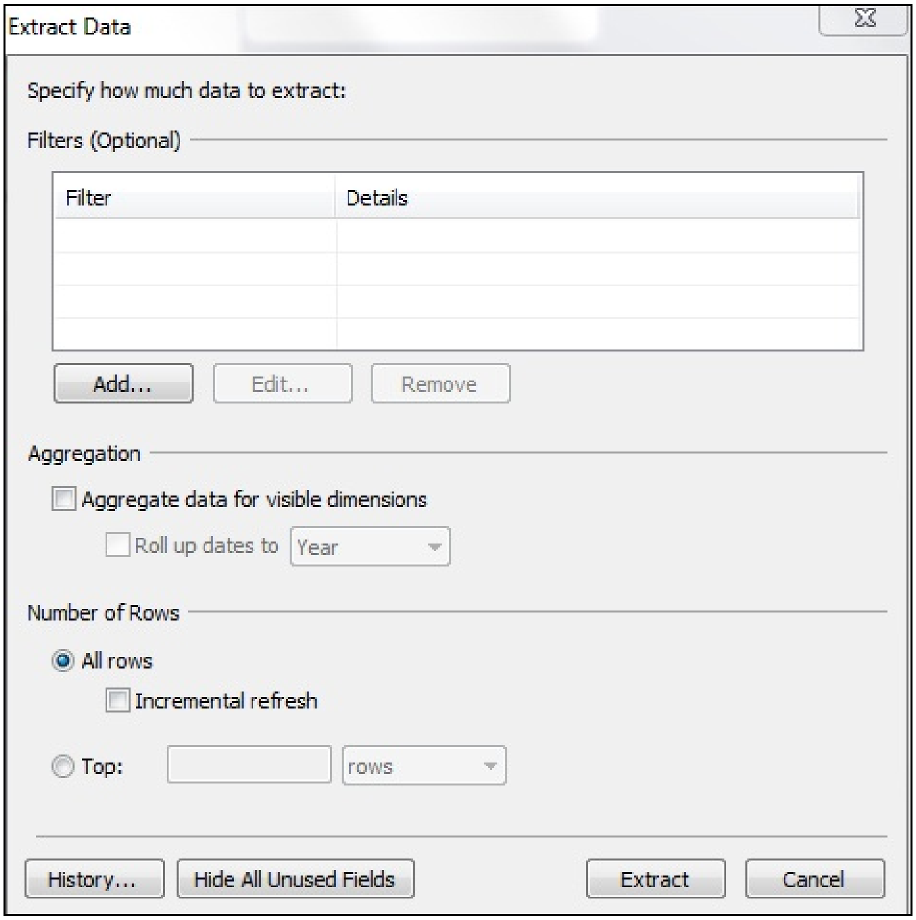
How to do it...

Once the data is loaded on the worksheet, perform the following steps to extract the data:

1. From the main menu toolbar, select **Data**.
2. Under **Data**, expand **Sample - Coffee Chain (Access)**, and select **Extract Data** as shown in the following screenshot:



- To select all rows, keep the **All rows** option selected as shown in the following screenshot:



- Hit **Extract**.
- In the **Save As** dialog box, select a desired location to save the extract.
- Enter the filename of the Tableau data extract file in the **File name** box in the **Save As** dialog box.
- Hit **Save**.

There's more...

Robin Kennedy from the The Information Lab Team, a Tableau consulting firm, has written a great insightful blog article about the reasons for a Tableau data extract. You can find this blog post at http://www.theinformationlab.co.uk/2011/01/20/tableau-extracts-what-why-how-etc/.

Chapter 8. Exporting and Sharing

In this chapter, we will see the following recipes:

- Saving a workbook on a Tableau server
- Sharing a workbook on the Web
- Exporting images
- Exporting data

Introduction

Tableau, apart from a workbook, extracts and prints to PDFs and offers customized options to export and share workbooks and data.

Saving a workbook on a Tableau server

Sharing visualizations on a Tableau server is one of the best ways to ensure that the readers are seeing the latest, and sometimes, live information. With the Tableau 8 server, the readers are also able to interact with the visualization by customizing it to their liking.

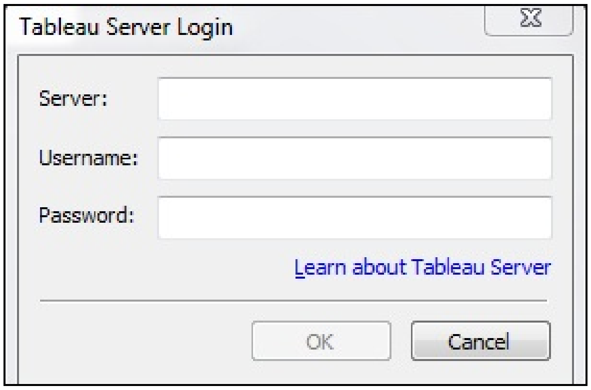
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source. You would also need access to a Tableau server to complete this example.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to save the workbook on a Tableau server:

1. Drag-and-drop **Market** from **Dimensions** into the **Rows** shelf.
2. Drag-and-drop **Sales** from **Measures** into the **Text input** section in the **Marks** pane.
3. From the main menu toolbar, click on **Server** and then click on **Publish Workbook**.
4. In the **Tableau Server Login** window (shown in the following screenshot), enter the **Server** name or the full path and the **Username** and **Password** values, and hit **OK**:



- In the **Publish Workbook** field in the **Tableau Server** dialog box, keep **Sheet1** selected in the **Views to Share** pane.
- Click on **Publish**.

Saving a workbook on the Web

Tableau Public is a great free product offered by Tableau to share any workbook with anyone on the Web. If you do not mind sharing your raw data with the users, then this is a great way of sharing your visualizations as the Tableau server can be very expensive.

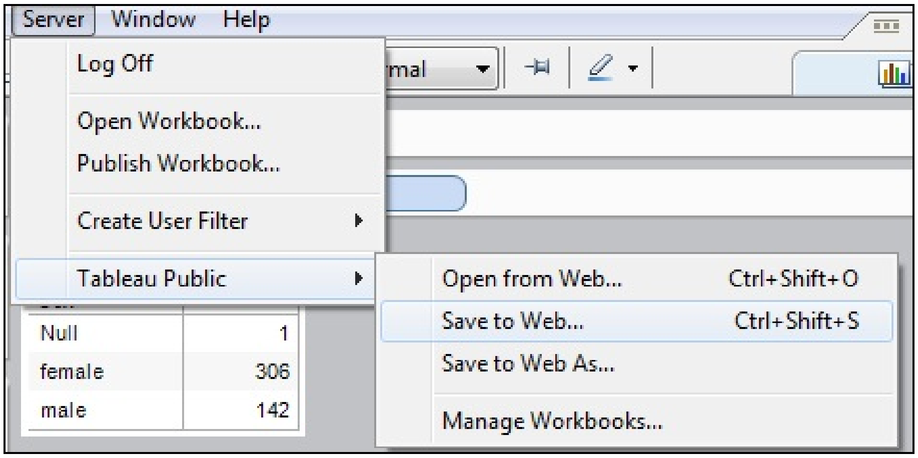
Getting ready

Download and save `titanic.txt` from http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic.txt on your local hard drive. Remember this location, as we will use this file for this recipe. This file lists all the passengers (and their details) that boarded the Titanic on its disastrous voyage. In addition, you'll need to create a Tableau Public free account from https://public.tableausoftware.com/auth/signup.

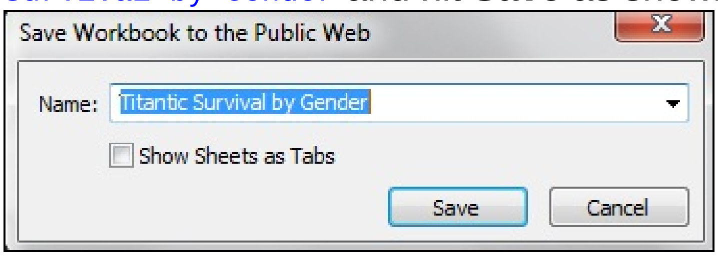
How to do it...

Create a new worksheet and perform the following steps to create a simple graphic and save the workbook on the Web:

1. Click on **Connect to data** to expand that area.
2. Click on **Text File** under the **In a file** section.
3. Find and select `titanic.txt` using the **Browse** button.
4. Keep all the options as is and hit **OK**.
5. In the **Data Connection** dialog box, click on the **Import all data** option.
6. Save the **Tableau Data Extract** file in a familiar location and hit **Save**.
7. Drag-and-drop **sex** from **Dimensions** into the **Rows** shelf.
8. Drag-and-drop **survived** from **Measures** into the **Text** box in the **Marks** pane.
9. From the main menu toolbar, select **Server**, expand **Tableau Public**, and select the **Save to Web** option as shown in the following

screenshot: 

- In the **Tableau Public Login** window, enter your e-mail address and password that you used to create your Tableau Public account.
- In the **Save Workbook to the Public Web** box, enter `Titantic Survival by Gender` and hit **Save** as shown in the following screenshot:



Exporting images

At times, it is quicker or necessary to share visualizations in image formats. Tableau provides an option to export visualizations in JPEG, PNG, BMP, and EMF formats.

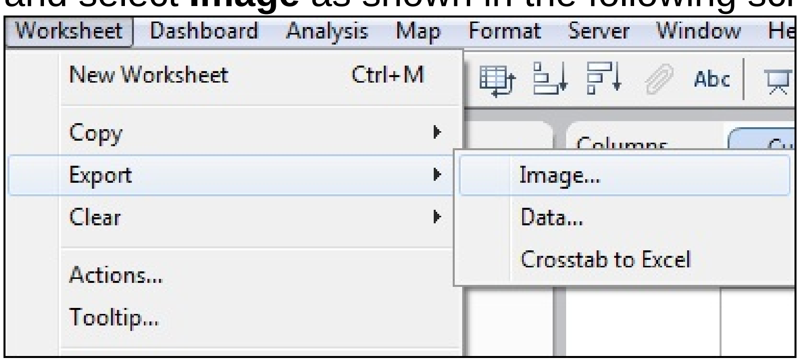
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

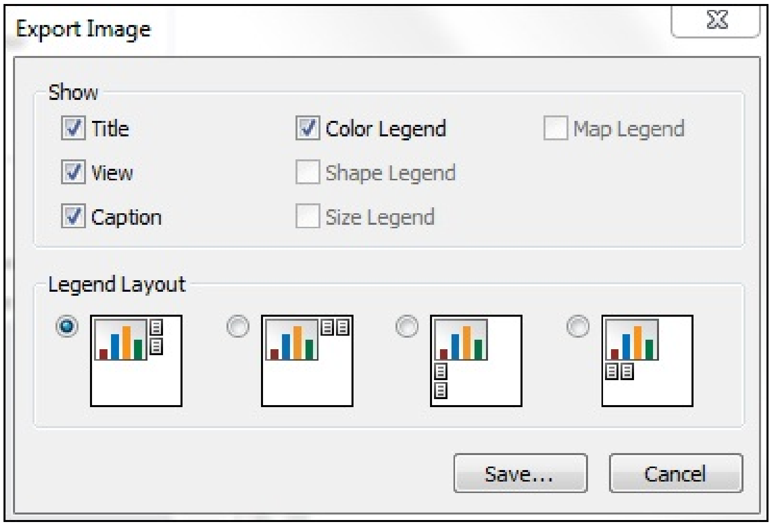
How to do it...

Once the data is loaded on the worksheet, perform the following steps to export a simple visualization in an image format:

1. Click on the **Show Me** button to display the **Show Me** toolbar on the screen.
2. Select **Customer Segment** and **Category** from **Dimensions** as well as **Profit** from **Measures**.
3. Click on the circle views icon on the **Show Me** toolbar.
4. From the main menu toolbar, select **Worksheet**, expand **Export**, and select **Image** as shown in the following screenshot:



- In the **Export Image** dialog box, keep all options selected and hit

Save: 

- In the **Save Image** file box, select a folder where you would like to save the file, give a **File name** value, and hit **Save**.

Exporting data

Raw data is helpful and is needed either for tables or for presentations, and Tableau provides an option of exporting the data behind the visualizations in an MS Access database format. This allows the user to get data from a different source, aggregate/manipulate the data, and export it to a database for further use.

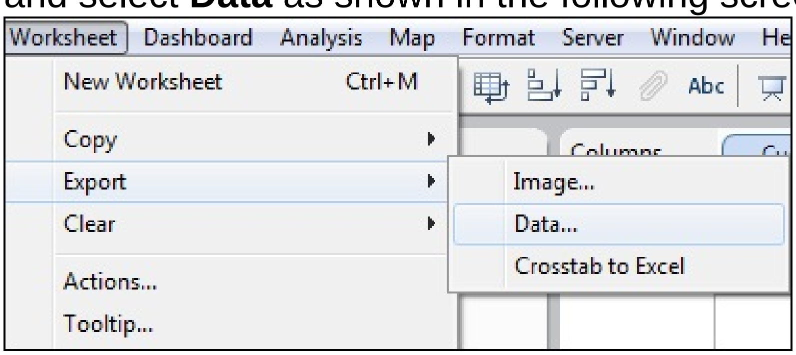
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

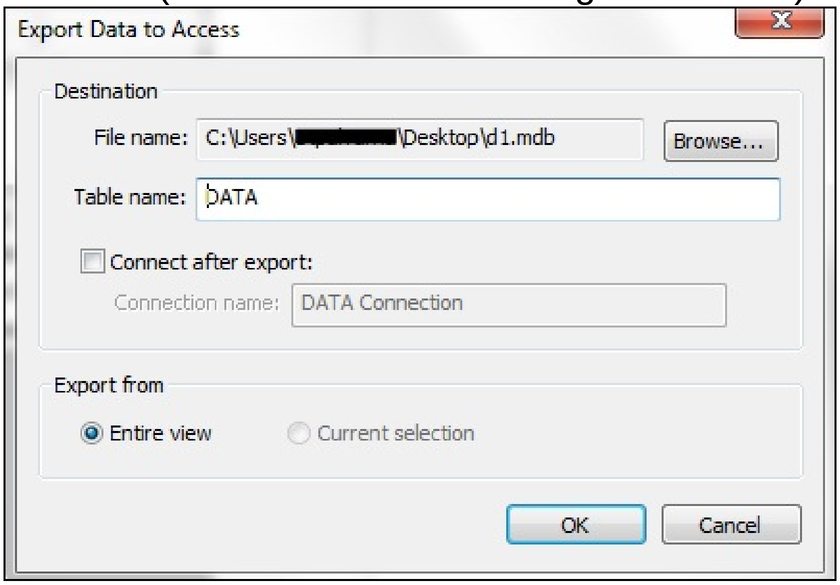
Once the data is loaded on the worksheet, perform the following steps to export a simple visualization in an image format:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Customer Segment** and **Category** from **Dimensions** as well as **Profit** from **Measures**.
3. Click on the circle views icon on the **Show Me** toolbar.
4. From the main menu toolbar, select **Worksheet**, expand **Export**, and select **Data** as shown in the following screenshot:



- In the **Export Data to Access** save file dialog box, select a location to save the Access database and enter a **File name** value, and hit **Save**.

- In the next **Export Data to Access** dialog box, keep the default values selected (as shown in the following screenshot) and hit **OK**:



How it works...

If all the default options are selected, Tableau will export the underlying data supporting the visualization to an MS Access database. If any filters are applied in the visualization, Tableau will export only the data that was not excluded because of filters, that is, the exact same data that is supporting the current visualization. You can expect the exported table `DATA` to look similar to the one in the following screenshot:

Category	Customer Segment	Profit (SUM)
Technology	Consumer	156699.3999
Office Supplies	Consumer	88532.3490000001
Furniture	Consumer	42728.234879
Technology	Corporate	374700.5542
Office Supplies	Corporate	203037.286
Furniture	Corporate	22008.083958
Technology	Home Office	173229.137
Office Supplies	Home Office	121145.782
Furniture	Home Office	23979.176491
Technology	Small Business	181684.4242
Office Supplies	Small Business	105306.041
Furniture	Small Business	28717.490953

Chapter 9. Exploring Advanced Features

The recipes in this chapter are as follows:

- Viewing data
- Changing the mark size
- Using the presentation mode
- Adding annotations
- Excluding data on the fly
- Customizing mark shapes
- Adding drop-down selectors
- Adding search box selectors
- Adding slider selectors
- Creating dashboards
- Creating animated visualizations
- Creating parameters

Introduction

This chapter will cover some advanced features and capabilities of Tableau.

Viewing data

Tableau, by default, aggregates the data and shows the aggregation on the visualization, but it is useful to see the underlying data that is used for that visualization. Using this feature, the user can view data behind the visualization.

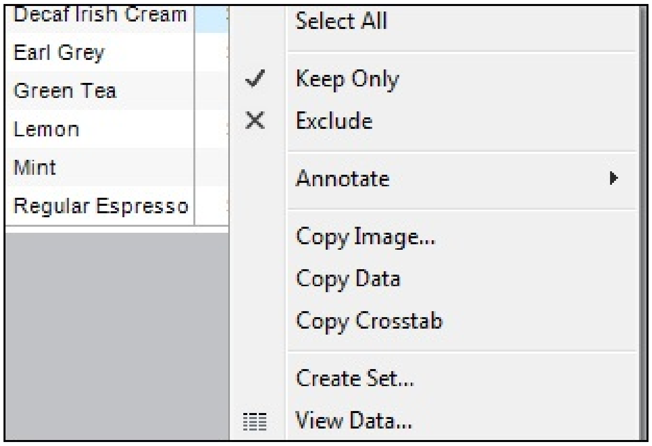
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

Once the data is loaded on the worksheet, follow these steps to create a simple table, and then view the data populating that table.

1. Drag-and-drop **Product** from **Dimensions** into the **Rows** shelf.
2. Drag-and-drop **Profit** from **Measures** into the **Text** box under the **Marks** pane.
3. Hit *Ctrl* + *A* on the keyboard to select all the data or the totaled **Profit** value for a single **Product** value.
4. Right-click anywhere on the selected area in the data table, and click on **View Data** as shown in the following screenshot:



- You can see the summarized data as shown in the following screenshot: 

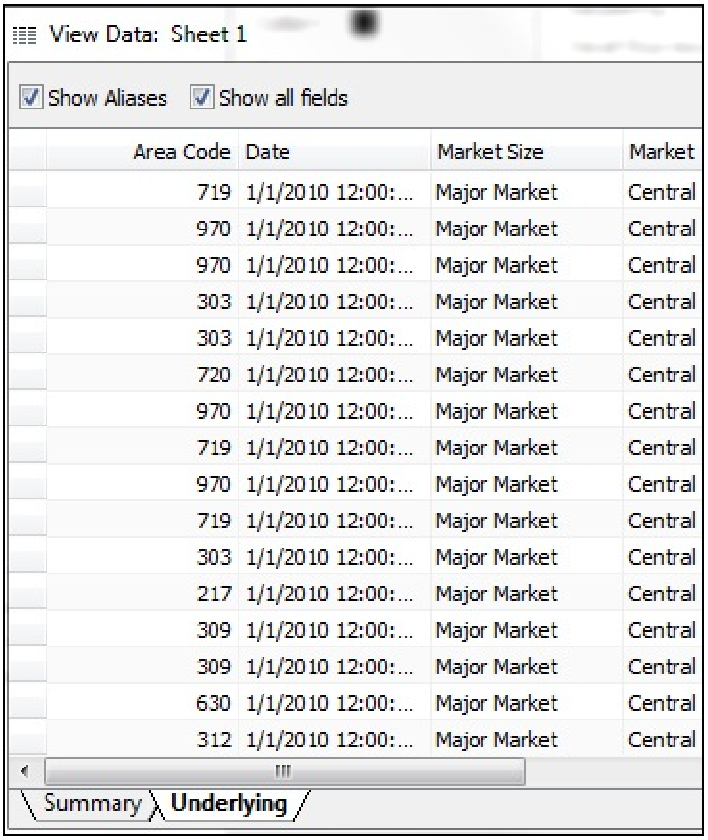
View Data: Sheet 1

Show Aliases

Product	Profit
Regular Espresso	$10,065
Mint	$6,154
Lemon	$29,869
Darjeeling	$29,053
Green Tea	($231)
Earl Grey	$24,164
Decaf Espresso	$29,502
Decaf Irish Cream	$13,989
Columbian	$55,804
Chamomile	$27,231
Caffe Mocha	$17,678
Caffe Latte	$11,375
Amaretto	$4,890

Summary | Underlying

- Click on the **Underlying** tab to see the complete and raw data supporting the summary view as shown in the following screenshot:



Area Code	Date	Market Size	Market
719	1/1/2010 12:00:...	Major Market	Central
970	1/1/2010 12:00:...	Major Market	Central
970	1/1/2010 12:00:...	Major Market	Central
303	1/1/2010 12:00:...	Major Market	Central
303	1/1/2010 12:00:...	Major Market	Central
720	1/1/2010 12:00:...	Major Market	Central
970	1/1/2010 12:00:...	Major Market	Central
719	1/1/2010 12:00:...	Major Market	Central
970	1/1/2010 12:00:...	Major Market	Central
719	1/1/2010 12:00:...	Major Market	Central
303	1/1/2010 12:00:...	Major Market	Central
217	1/1/2010 12:00:...	Major Market	Central
309	1/1/2010 12:00:...	Major Market	Central
309	1/1/2010 12:00:...	Major Market	Central
630	1/1/2010 12:00:...	Major Market	Central
312	1/1/2010 12:00:...	Major Market	Central

- To view fields shown on the worksheet, check the **Show all fields** option.

Changing the mark size

Although Tableau's default mark sizes are good, the option of changing (increasing or decreasing) those mark sizes is also useful. This can help the reader see the data better and reduce the confusion caused by overlapping or small marks.

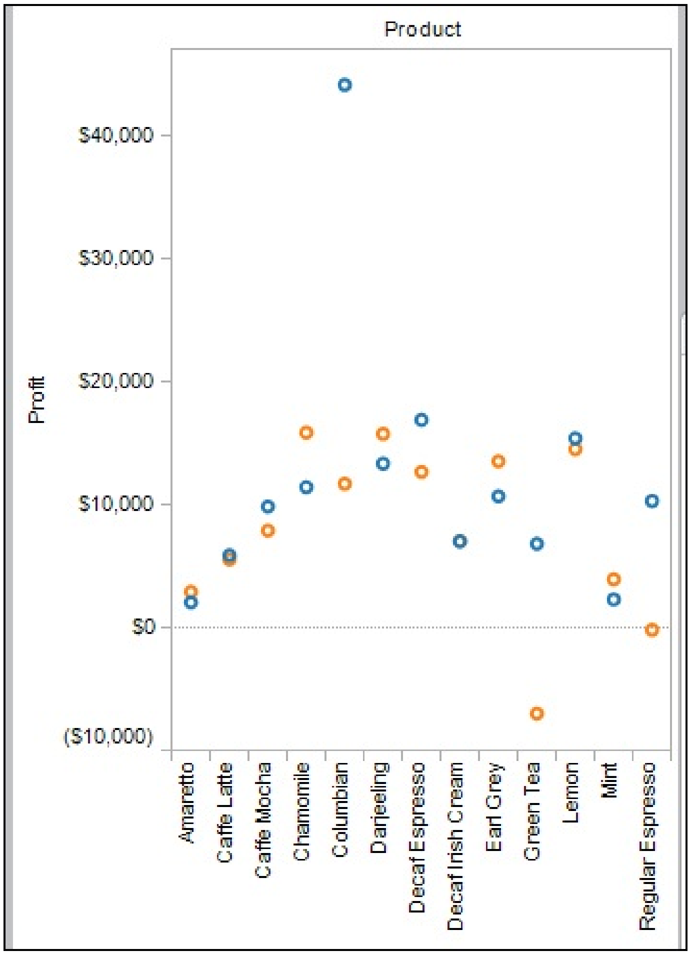
Getting ready

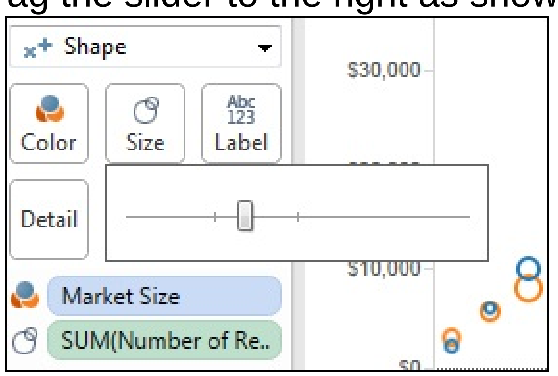
Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to increase or decrease the mark size:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Product** and **Market Size** from **Dimensions** and **Profit** from **Measures**.
3. Click on the circle view icon on the **Show Me** toolbar to generate a visualization as shown in the following screenshot:



- To vary the mark size by the number of records in each **Product** and **Market Size** values, drag-and-drop **Number of Records** from **Measures** into the **Size** box under the **Marks** pane.
- To increase the size of the marks, click on the **Size** box to show the slider and drag the slider to the right as shown in the following screenshot: 

- To decrease the size of the marks, drag the slider to the left.

Using the presentation mode

While editing a Tableau workbook, a user sees many options and a lot of screen real estate is used towards showing such information. Tableau offers a useful feature of displaying the visualization in a presentation mode, removing any clutter or unnecessary options for exploring the visualization.

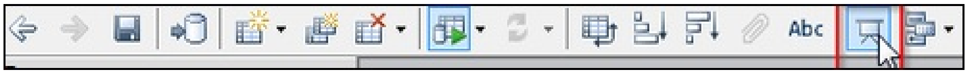
Getting ready

Follow the *Changing the mark size* recipe to create a simple visualization. We will view this visualization in the presentation mode.

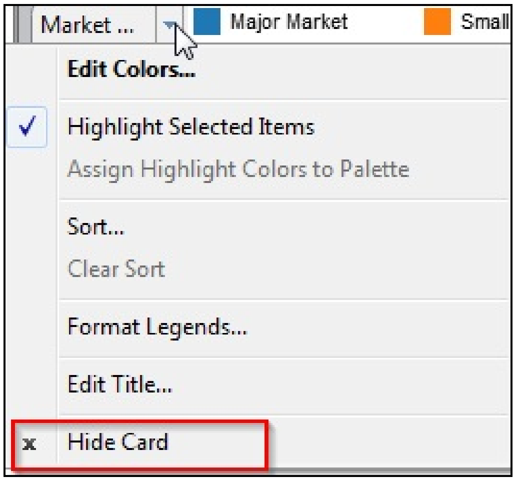
How to do it...

Once you have created the simple visualization, perform the following steps to view it in the presentation mode:

1. Since you cannot change the positions of legends in the presentation mode, move the **Market Size** legend and the **SUM(Number of Records)** legend below the chart.
2. Adjust the height and width of the legend boxes as shown in the following screenshot:

- Once you are satisfied with the positioning and the dimensions of the legends, click on the **Presentation Mode** button shown in the following screenshot: 

- To hide a legend card during the presentation mode, click on the small dropdown, seen by hovering the mouse over the legend title, and click on **Hide Card** as shown in the following screenshot:



Adding annotations

Although many graphs are self explanatory, annotations on visualizations help the reader understand the graphic better and note any important trends or characteristics of the data. Tableau provides three main types of annotations: **Point**, **Area**, and **Mark**. The **Point** annotation creates annotations for a data point in the visualization. The **Mark** annotation creates annotations for a specific, selected mark whereas the **Area** pane covers an area consisting of many other points of interest.

Getting ready

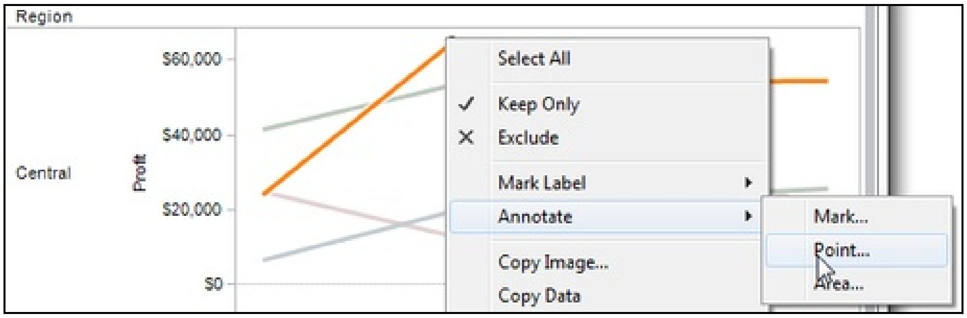
Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

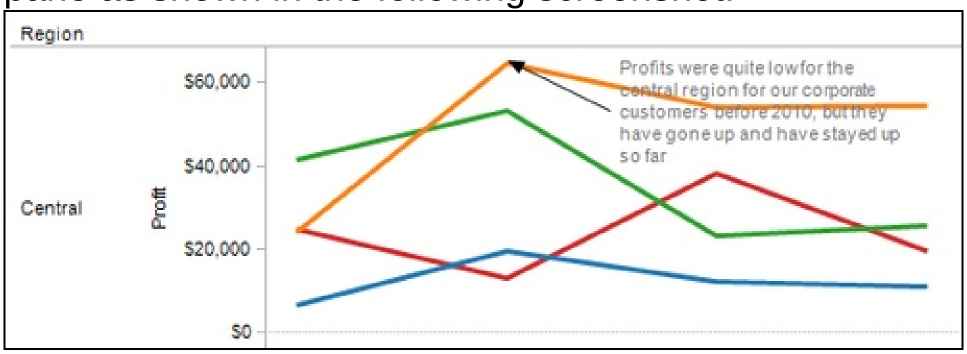
Once the data is loaded on the worksheet, perform the following steps to add annotations to a visualization:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Customer Segment**, **Order Date**, and **Region** from **Dimensions** and **Profit** from **Measures**.
3. Click on **lines (continuous)** on the **Show Me** toolbar to create a visualization as shown in the following screenshot:

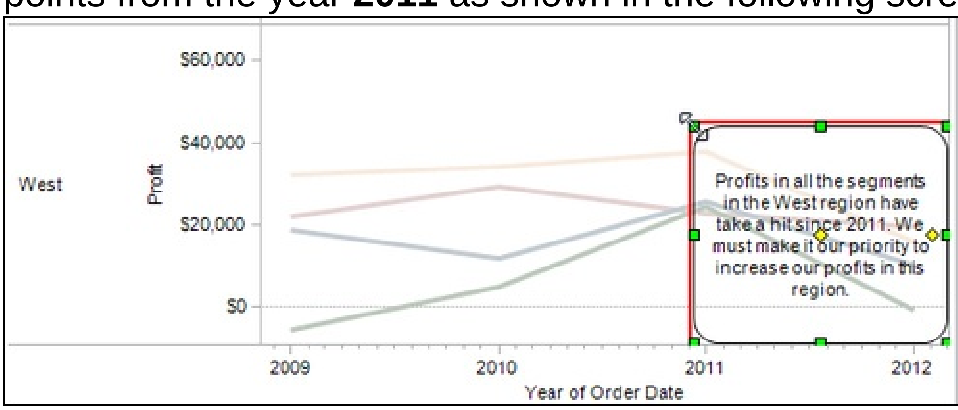
- Right-click on the **Central** region's **Profit** value for the year **2010**, expand the menu for **Annotate**, and select **Point...** as shown in the following screenshot:



- In the **Edit Annotation** box, enter `Profits were quite low for the central region for our corporate customers before 2010, but they have gone up and have stayed up so far` and select **OK**.
- You can see the annotation for the year **2010** in the **Central** region pane as shown in the following screenshot:



- Right-click on the **Home Office Customer Segment** point for the year **2012** in the **West** region, expand the menu for **Annotate**, and select **Area**.
- In the **Edit Annotation** box, enter `Profits in all the segments in the West region have taken a hit since 2011. We must make it our priority to increase our profits in this region.` and hit **OK**.
- Click on the **Annotation** box and drag the top-left corner to cover points from the year **2011** as shown in the following screenshot:



Excluding data on the fly

We can easily set multiple filters on a visualization; however, while exploring the data, the user has a very common need to exclude certain data without explicitly creating a filter beforehand. Tableau offers this feature, and exclusion happens on the selected marks and a filter is created.

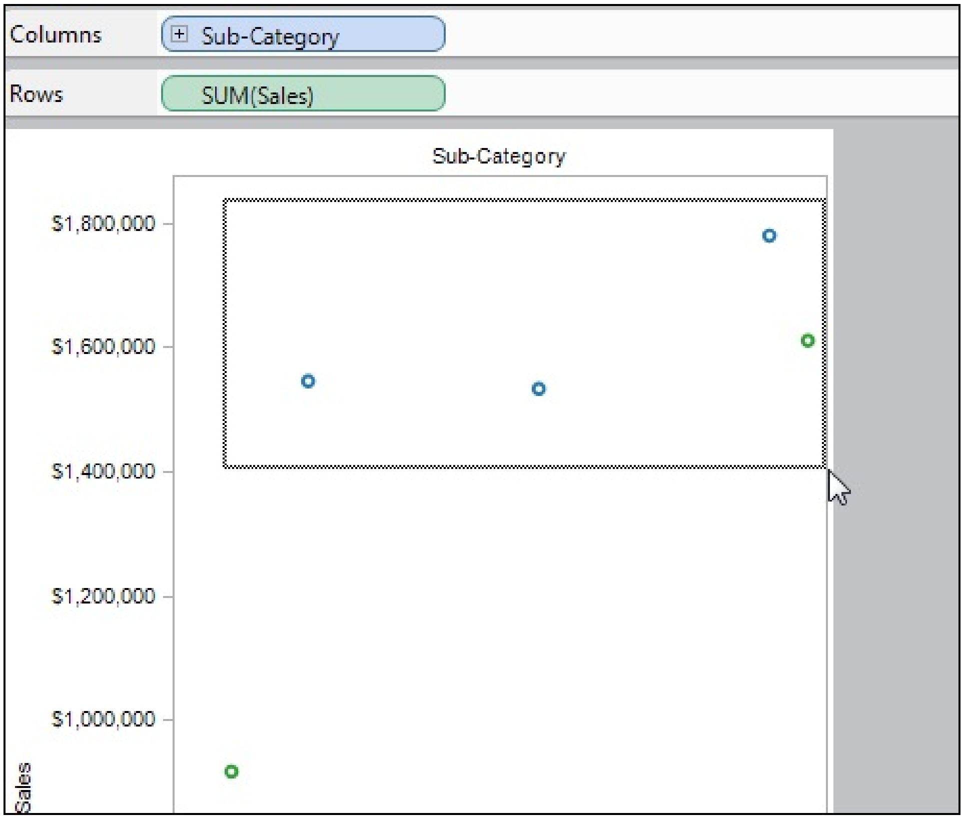
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

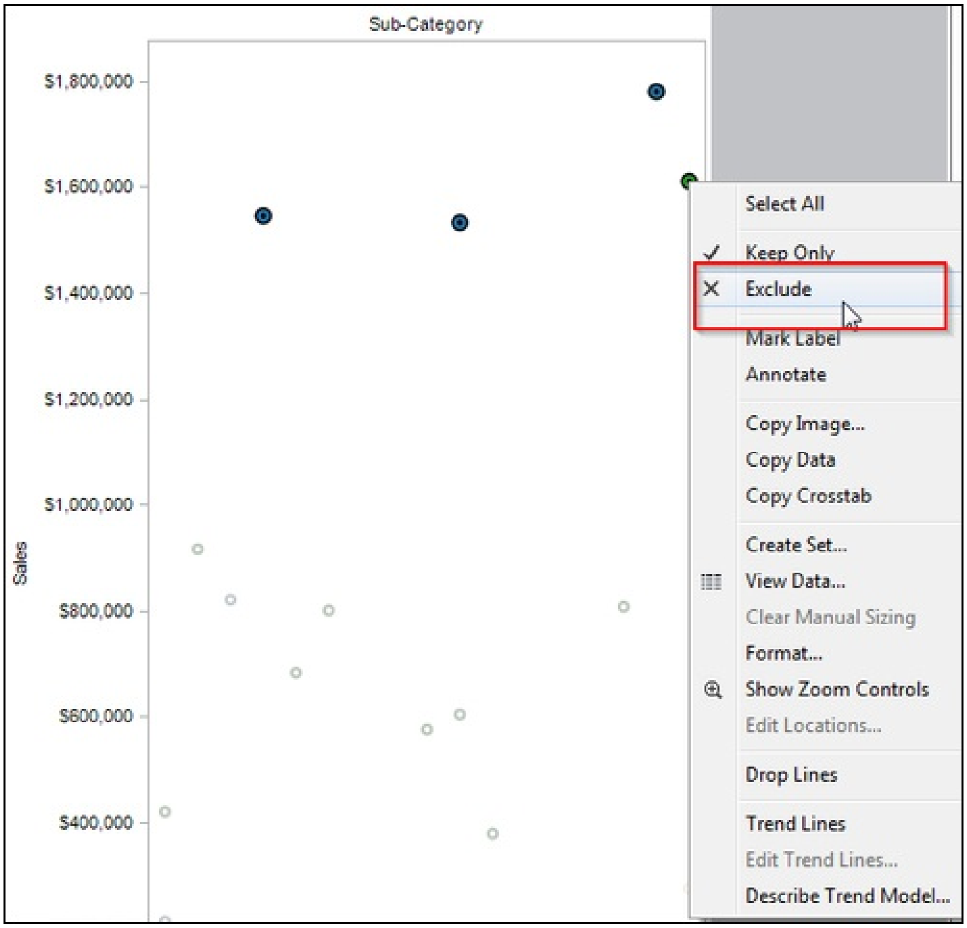
How to do it...

Once the data is loaded on the worksheet, perform the following steps to create a visualization and exclude or filter some data from the visualization:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Ship Mode** and **Sub-Category** from **Dimensions** and **Sales** from **Measures**.
3. Click on the circle view icon on the **Show Me** toolbar.
4. To exclude every product with a **Sales** value of more than 1,000,000 dollars, select all the four marks above **$1,000,000** as shown in the following screenshot:



- The selected marks will become darker in color, and all the other marks will dim down. Right-click on any selected mark, and click-on **Exclude** as shown in the following screenshot:



Sub-Category
$1,800,000
$1,600,000
$1,400,000
$1,200,000
$1,000,000
$800,000
$600,000
$400,000
Sales
Select All
Keep Only
Exclude
Mark Label
Annotate
Copy Image...
Copy Data
Copy Crosstab
Create Set...
View Data...
Clear Manual Sizing
Format...
Show Zoom Controls
Edit Locations...
Drop Lines
Trend Lines
Edit Trend Lines...
Describe Trend Model...

Customizing mark shapes

From branding your company to distinguishing the data, custom mark shapes can help in customizing the visualization. Once you have the desired shape of files, you can easily use them instead of using Tableau's standard mark shapes.

Getting ready

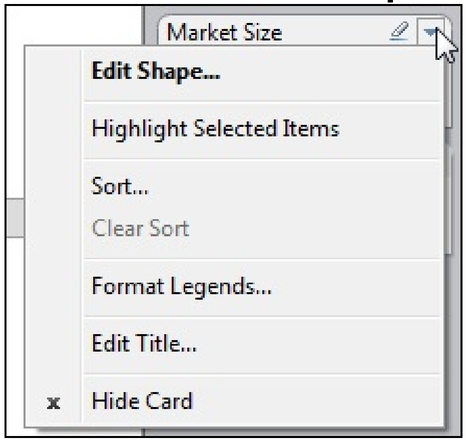
This recipe requires modifying the custom shape image files, which could be of the JPG, BMP, and GIF formats. Specifically, it makes the background transparent while keeps the original size of 32 by 32 pixels. The tutorial given at http://www.interworks.com/blogs/iwbiteam/2012/01/27/using-custom-shapes-tableau explains in detail how to modify the shape properties. For this recipe, you can choose to create and modify your own shapes or download the ones used in this recipe.

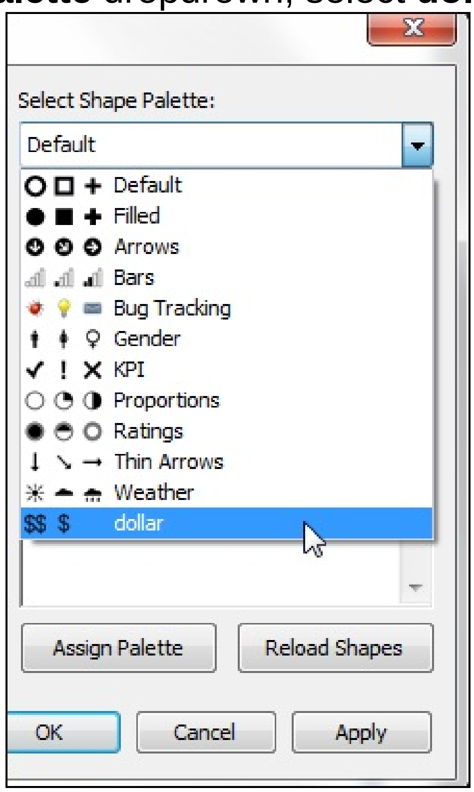
Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

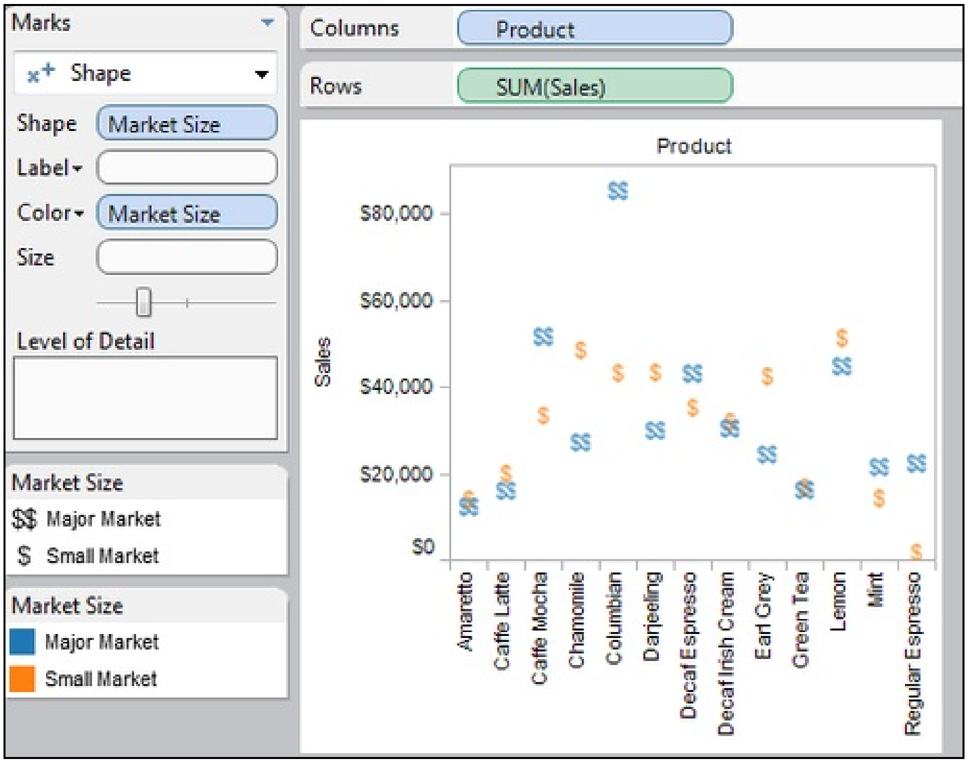
Once the data is loaded on the worksheet, perform the following steps to use the custom shape marks:

1. Create a folder called `dollar` in the `Shapes` folder in `My Tableau Repository`.
2. Copy the `single.png` and `double.png` images in the `dollar` folder.
3. Drag-and-drop **Product** from **Dimensions** into the **Columns** shelf.
4. Drag-and-drop **Sales** from **Measures** into the **Rows** shelf.
5. From the **Marks** dropdown, select **Shape**.
6. Drag-and-drop **Market Size** from **Dimensions** into the **Shape** box.
7. Drag-and-drop **Market Size** from **Dimensions** into the **Color** box.

8. Hover your mouse over the **Market Size** legend shelf, click on the small dropdrown arrow, and select **Edit Shape...** as shown in the following screenshot: 

- In the **Edit Shape [Market Size]** dialog box, hit the **Reload Shapes** button.
- In the **Select Shape Palette** dropdrown, select **dollar** as shown as in the following screenshot: 

- Select **Major Market** under **Select Data Item**, and select the double dollar sign.
- Select **Small Market** under **Select Data Item**, and select the single dollar sign.
- Select **OK** and the final visualization should look like the following

screenshot: 
Marks
Shape
Shape
Market Size
Label
Color
Market Size
Size
Level of Detail
Market Size
$$ Major Market
$ Small Market
Market Size
Major Market
Small Market
Columns
Product
Rows
SUM(Sales)
Product
Sales
$80,000
$60,000
$40,000
$20,000
$0
Amaretto
Caffe Latte
Caffe Mocha
Chamomile
Columbian
Darjeeling
Decaf Espresso
Decaf Irish Cream
Earl Grey
Green Tea
Lemon
Mint
Regular Espresso

Adding drop-down selectors

It is useful to give the readers some control over the visualization, and using quick filters is a great way of doing so. Drop-drown selectors are good if you want the reader to select one value without the risk of misspelling.

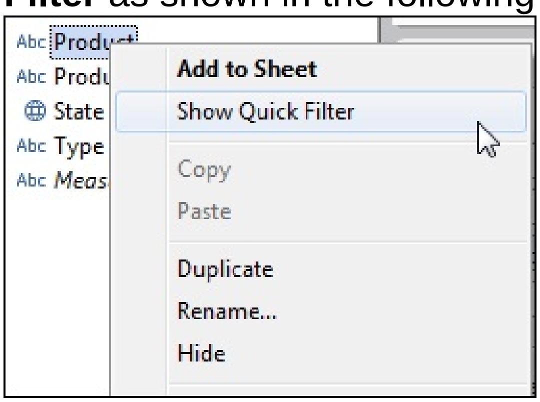
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

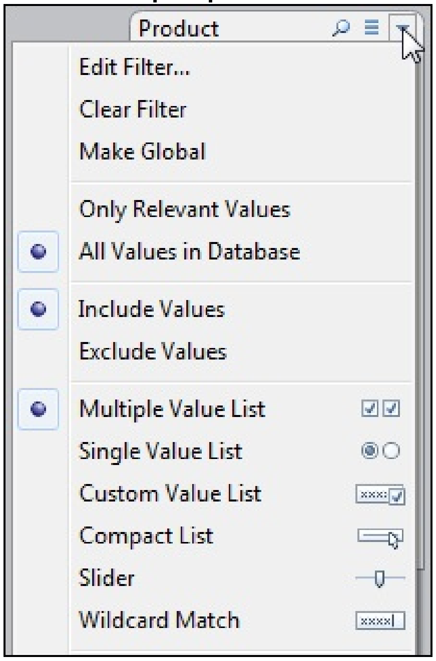
How to do it...

Once the data is loaded on the worksheet, perform the following steps to create a drop-down selector on your visualization:

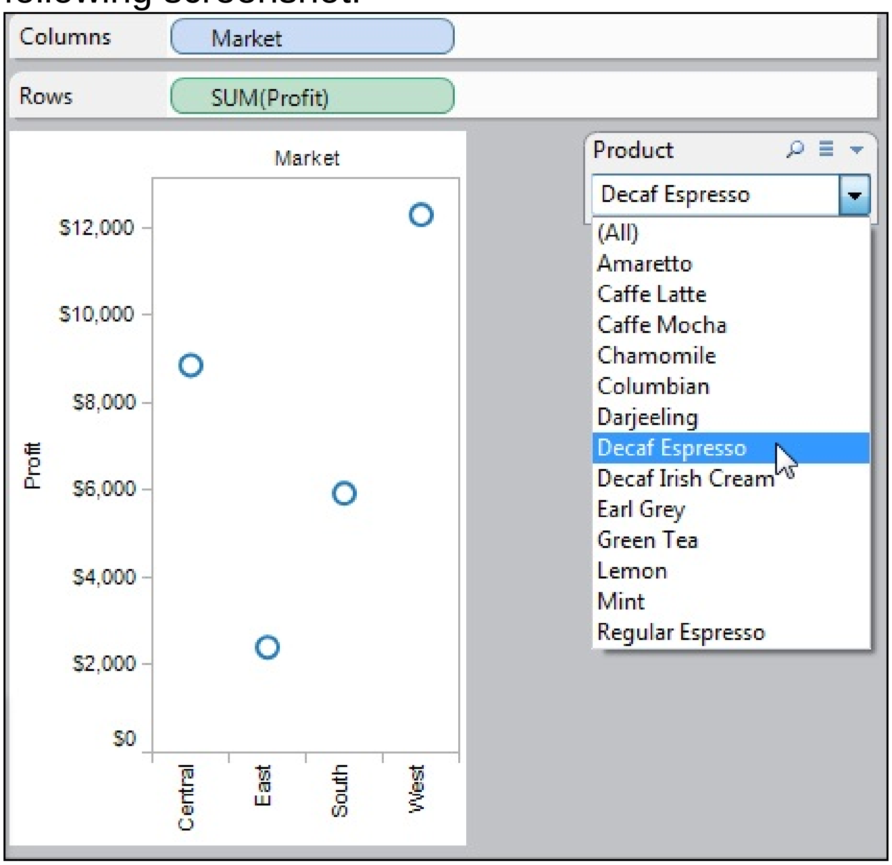
1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Market** and **Type** from **Dimensions** and **Profit** from **Measures**.
3. Click on the circle views icon on the **Show Me** toolbar.
4. Right-click on **Product** from **Dimensions**, and select **Show Quick Filter** as shown in the following screenshot:



- On the **Product** filter box, click on the small drop-down arrow to open the filter properties menu as shown in the following screenshot:



- From this menu, select **Compact List** to create a drop-down selector box. The final visualization will look similar to the one shown in the following screenshot:



Adding search box selectors

Search box selectors are useful when you want the readers to type a part of the value and yet be able to filter the data. This does assume that the readers know the underlying values of the filtered fields.

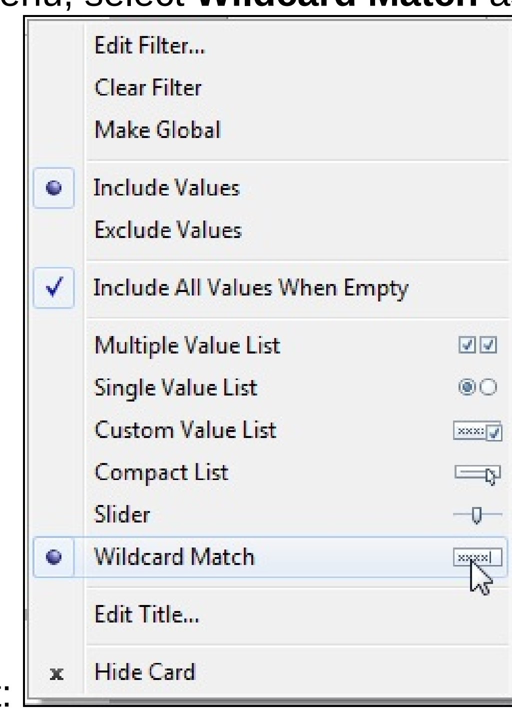
Getting ready

Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source. Follow all the steps, except for the last step, given in the recipe *Adding drop-down selectors*.

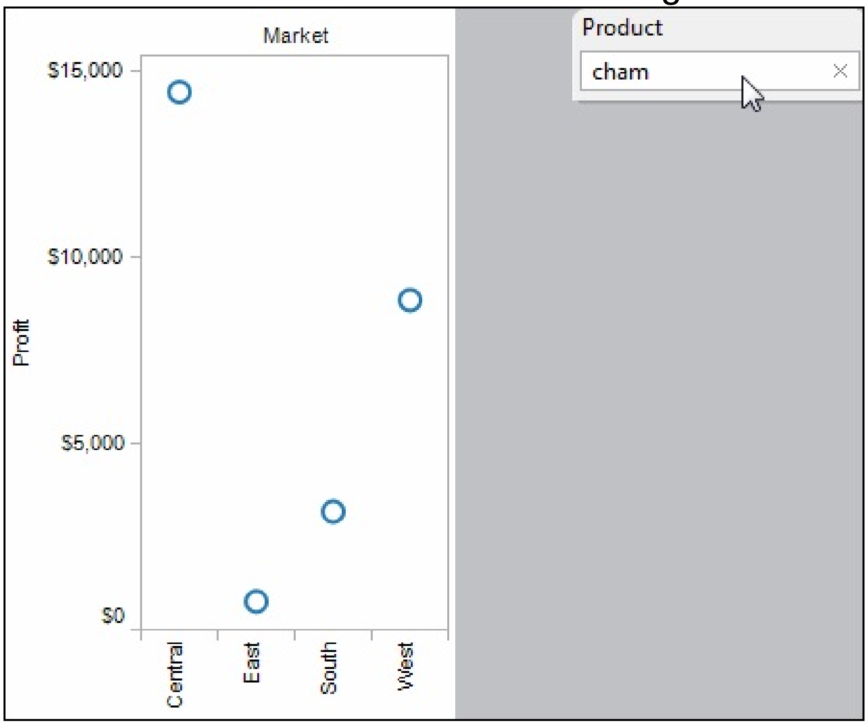
How to do it...

Once the data is loaded on the worksheet and you have placed a quick filter, perform the following steps to add a search box selector to your visualization:

1. From the quick filter drop-down menu, select **Wildcard Match** as shown in the following screenshot:



- To see this search box in action, enter `cham` in the search box and hit *Enter*. You'll see a visualization for the **Product** values that contain `cham` in their names as shown in the following screenshot:



Adding slider selectors

Slider selectors add more flexibility for filtering the numeric or date type of data, though this selector can be used on any type of data.

Getting ready

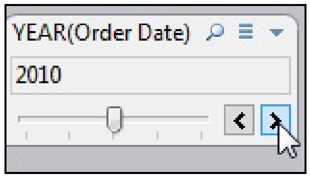
Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to add slider selector to your visualization:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Customer Segment** and **Category** from **Dimensions** and **Sales** from **Measures**.
3. Click on the circle views icon on the **Show Me** toolbar.
4. Right-click on **Order Date** from **Dimensions** and select **Show Quick Filter** to show a multiple checkbox selector as shown in the following screenshot: 

- On the **YEAR(Order Date)** quick filter box, click on the small drop-down arrow to modify the filter properties.
- Select **Slider** from this box to add a slider filter as shown in the

following screenshot: 

Creating dashboards

Dashboards in Tableau are very powerful as they are a compilation of individual visualizations on different sheets. This provides the reader with a lot of information on one single view with all the filters, parameters, and legends of individual visualizations. Complex types of dashboards can be created, such as those allowing drill-through aggregate information and viewing the details.

Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to create a simple dashboard:

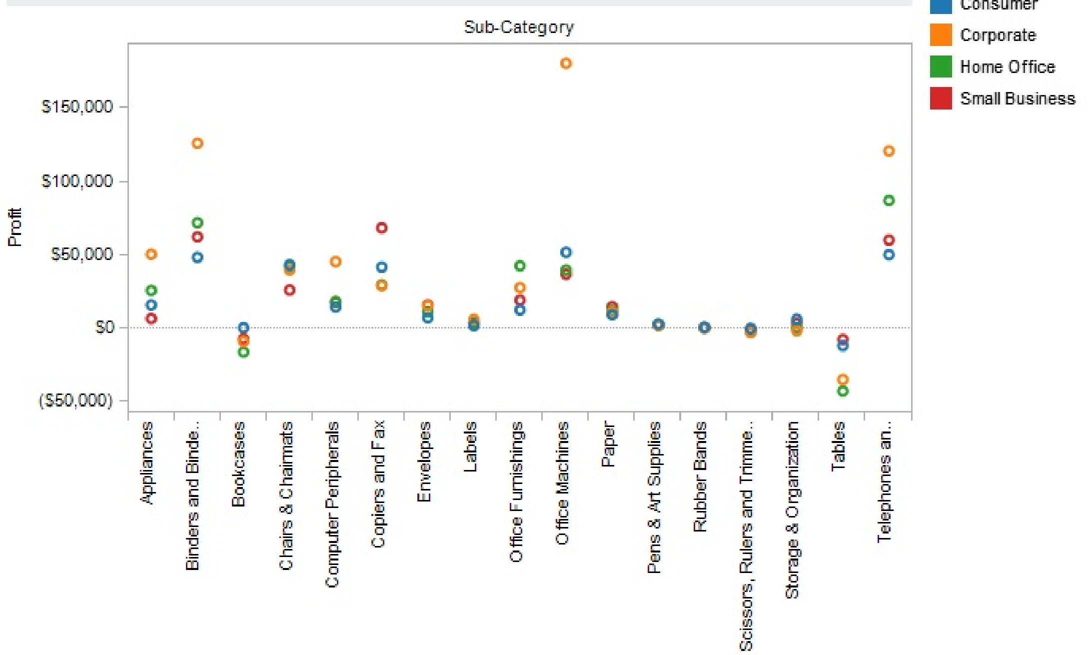
1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Customer Segment** and **Sub-Category** from **Dimensions** and **Profit** from **Measures**.
3. Click on the circle views icon on the **Show Me** toolbar.
4. From the main menu toolbar, select **Show Title** under **Worksheet**.
5. Double-click on the **Title** shelf and change the title to `Profit by Customer Segment and Product Sub-Category`.
6. From the main menu toolbar, select **New Worksheet** under **Worksheet**.
7. Drag-and-drop **Supplier** into the **Rows** shelf.
8. Drag-and- drop **Profit** into the **Text** shelf under the **Marks** pane.
9. Drag-and-drop **Supplier** into the **Filters** shelf.
10. In the **Filters [Supplier]** dialog box, select the **Top** tab and select the **By Field** radio button.

11. Click on **OK**.
12. From the main menu toolbar, select **Show Title** under **Worksheet**.
13. Double-click on the **Title** shelf and change the title to `Top 10 Suppliers by Profit`.
14. From the main menu toolbar, select **New Dashboard** under **Dashboard**.
15. In the **Dashboard** view, drag-and-drop **Sheet 1**.
16. Drag-and-drop **Sheet 2** on top of **Sheet 1** as shown in the following screenshot:

Top 10 Suppliers by Profit

Supplier	
Other	$1,241,628
Canon	$117,476
Fellowes	$93,642
Xerox	$22,082
Hon	$21,155
Belkin	$15,638
Epson	$14,898
Avery	$13,473
3M	$8,111
Boston	$2,095

Profit by Customer Segment and Product Sub-Category



Sheet 1 Sheet 2 Dashboard 1

Creating animated visualizations

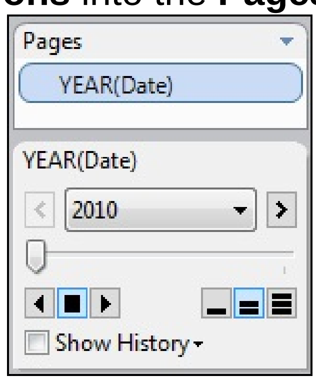
Animated visualizations are useful for spotting a measure in seasonal trends or simply observing measures over a period of time.

Getting ready

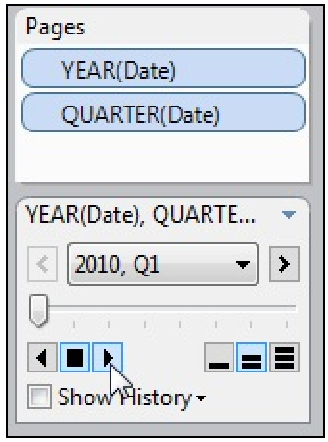
Let's use the sample file `Sample – Coffee Chain (Access)`. Open a new worksheet and select **Sample – Coffee Chain (Access)** as the data source.

How to do it...

Once the data is loaded on the worksheet, perform the following steps to create an animated visualization:

1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Product** from **Dimensions** and **Sales** from **Measures**.
3. Click on the horizontal bars icon on the **Show Me** toolbar.
4. Drag-and-drop **Date** from **Dimensions** into the **Pages** shelf as shown in the following screenshot: 

- To show the graphs by quarters, drag-and-drop **Date** from **Dimensions** into the **Pages** shelf again.
- To play the animation, click on the play button from the **Pages** shelf as

shown in the following screenshot: 

Creating parameters

Parameters allow more interaction with the reader by allowing him/her to change certain values and see how this impacts other measures. By creating parameters, the reader can be put in charge of evaluating various what-if scenarios or given options to choose the number of items to view.

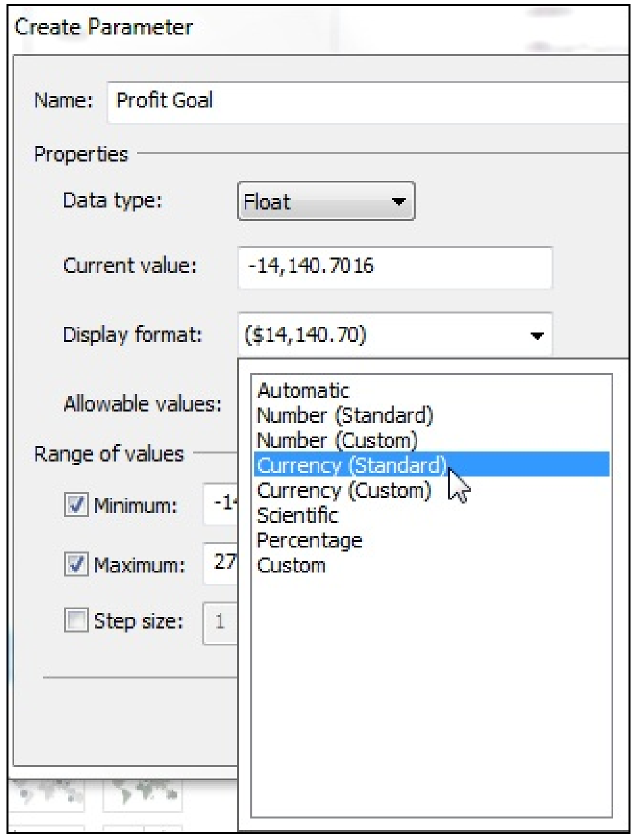
Getting ready

Let's use the sample file `Sample – Superstore Sales (Excel)`. Open a new worksheet and select **Sample – Superstore Sales (Excel)** as the data source.

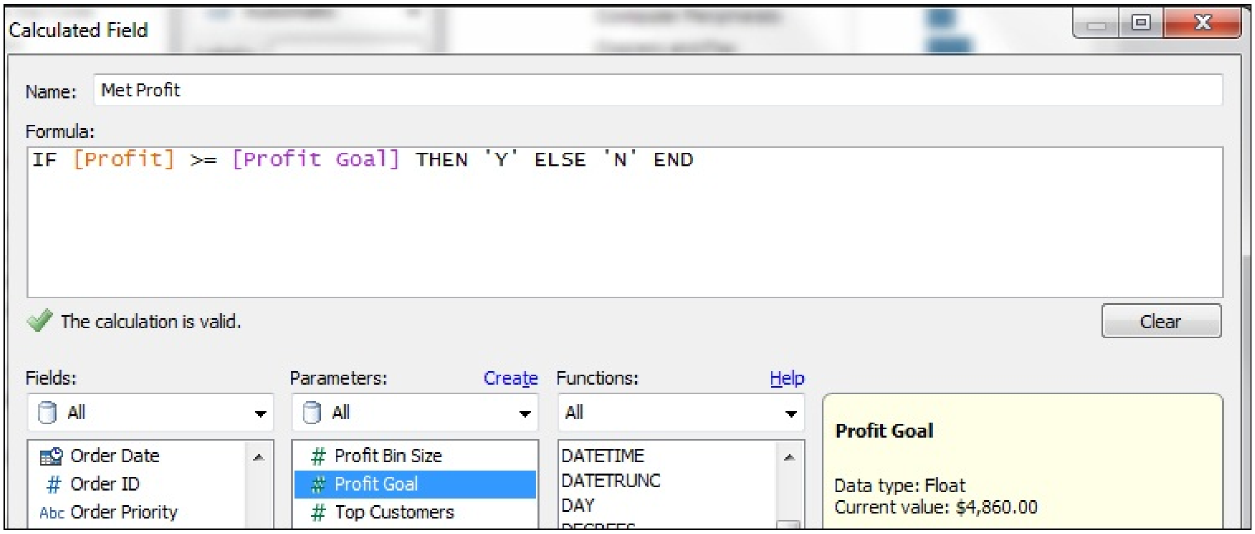
How to do it...

Once the data is loaded on the worksheet, perform the following steps to create a parameter, and use it to dynamically adjust the visualization:

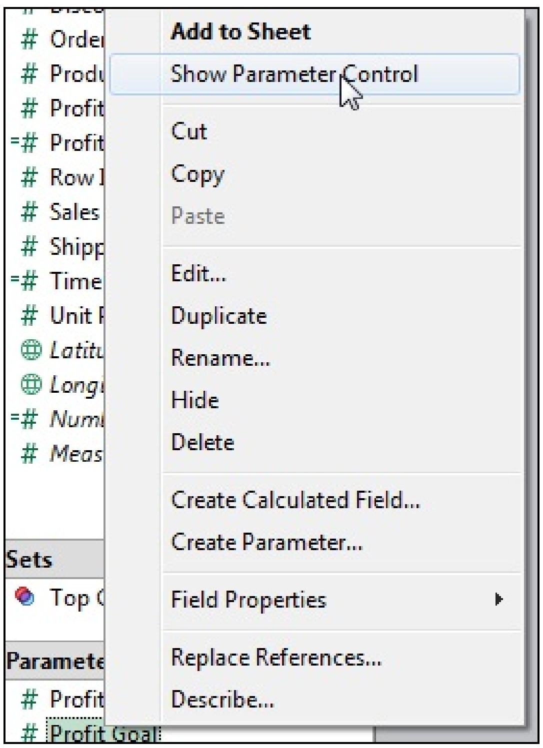
1. Click on the **Show Me** button to bring the **Show Me** toolbar on the screen.
2. While holding the *Ctrl* key, click on **Sub-Category** and **Region** from **Dimensions**, and **Profit** from **Measures**.
3. Click on the horizontal bars icon on the **Show Me** toolbar.
4. Right-click on **Profit** from **Measures** and select **Create Parameter**.
5. In the **Create Parameter** dialog box, enter `Profit Goal` in the **Name** textbox.
6. Change the **Display** format to **Currency (Standard)** as shown in the

following screenshot: 

- Check the **Step size** checkbox and enter `1000`.
- Select **OK**.
- Right-click on the newly created parameter **Profit Goal**, and select **Create Calculated Field**.
- In the **Calculated Field** dialog box, in the **Name** textbox, enter `Met Profit`.
- In the **Formula** box, enter the `IF [Profit] >= [Profit Goal] THEN 'Y' ELSE 'N' END` formula as shown in the following screenshot:



- Select **OK**.
- Drag-and-drop the newly created calculated field **Met Profit Goal** in the **Color** box into the **Marks** pane.
- Right-click on the **Profit Goal** parameter and select **Show Parameter Control** as shown in the following screenshot:



- Change the values of **Profit Goal** using the slider or left and right arrow buttons as shown in the following screenshot:


Region
Sub-Category
Central
Appliances
Binders and Binder Accesso..
Bookcases
Chairs & Chairmats
Computer Peripherals
Profit Goal
$7,860.00

Part 3. Module 3

Creating Data Stories with Tableau Public

Illustrate your data in a more interactive and interesting way using Tableau Public

Chapter 1. Getting Started with Tableau Public

Making sense of data is a valued service in today's world. It may be a cliché, but it's true that we are drowning in data and yet, we are thirsting for knowledge. The ability to make sense of data and the skill of using data to tell a compelling story is becoming one of the most valued capabilities in almost every field, which includes business, journalism, retail, manufacturing, medicine, and public service. Tableau Public (for more information, visit www.tableaupublic.com), which is Tableau Software's free Cloud-based data visualization client, is a powerfully transformative tool that can be used to create rich, interactive, and compelling data stories. It's a great platform if you wish to explore data through visualization. It enables your consumers to ask and answer questions that are interesting to them.

This book is written for people who are new to Tableau Public and who would like to learn how to create rich, interactive data visualizations from publicly available data sources that they then can easily share with others. Once you publish visualizations and data to Tableau Public, they are accessible to, and can be viewed and downloaded by, everyone. A typical Tableau Public data visualization contains public data sets such as sports, politics, public works, crime, census, socioeconomic metrics, and social media sentiment data. You can also create and use your own data. Many of these data sets are either readily available on the Internet, or can be accessed via a public records request or search (if they are harder to find, they can be scraped from the Internet). You can now control who can download your visualizations and data sets, which is a feature that was previously available only to paid subscribers. Tableau Public currently has a maximum data set size of 10 million rows and/or 10 GB of data.

In this chapter, we will walk through an introduction to Tableau, which includes the following topics:

- A discussion on how you can use Tableau Public to tell your data story
- Examples of organizations that use Tableau Public
- Downloading and installing the Tableau Public software
- Logging in to Tableau Public
- Creating your own Tableau Public profile
- Discovering the Tableau Public features and resources
- Having a look at the author profiles and galleries section of the Tableau website so that we can browse other authors' data visualizations (this is a great way to learn and gather ideas on how to best present data)

A Tableau Public overview

Tableau Public allows you to tell your data story and create compelling and interactive data visualizations that invite discovery and education. Tableau Public is sold at a great price—free. It allows you as a data storyteller to create and publish data visualizations without learning how to code or having special knowledge about web publishing. In fact, you can publish data sets of up to 10 million rows or 10 GB to Tableau Public in a single workbook. Tableau Public is a data discovery tool. It should not be confused with enterprise-grade business intelligence tools, such as Tableau Desktop and Server, QlikView, and Cognos Insight. Those tools integrate with corporate networks and security protocols as well as server-based data warehouses. Data visualization software is not a new thing. Businesses have used software to generate dashboards and reports for decades. The new twist comes with data discovery tools, such as Tableau Public. Journalists and bloggers who would like to augment their reporting of static text and graphics can use these data discovery tools, such as Tableau Public, to create compelling, rich data visualizations, which may consist of one or more charts, graphs, tables, and other objects that can be controlled by readers to allow for discovery.

The people who are active members of the Tableau Public community have a few primary traits in common— they are curious, generous with their knowledge and time, and enjoy conversations that relate data to the world around us. Tableau Public maintains a list of blogs of data visualization experts who use Tableau Software.

In the following screenshot, Tableau Zen Masters *Anya A'hearn* of Databrick and *Allan Walker* used data on San Francisco bike sharing to show the financial benefits of **Bay Area Bike Share**, a city-sponsored 30-minute bike sharing program, as well as a map of both the proposed expansion of the program and how far a person can actually ride a bike in half an hour.

This dashboard is featured in the Tableau Public gallery because it relates data to users clearly and concisely. It presents a great public interest story (commuting more efficiently in a notoriously congested city) and then grabs the viewer's attention with maps of current and future offerings. The second dashboard within the analysis is significant, as well. The authors described the Geographic Information Systems (**GIS**), the tools that they used to create innovative maps, as well as the methodology that went into the final product so that the users who are new to the tool can learn how to create a similar functionality for their own purposes:

30 MINUTES

In 2014 over 85% of riders using the **Bay Area Bike Share** within the city of San Francisco were yearly subscribers. For $88 a year, (compared to $792 for a year of MUNI) subscribers are allowed unlimited rides at no additional charge as long as each ride duration is less than 30 minutes. Current bike station locations are concentrated downtown, but new stations are being proposed, making the service a viable commuting option, either on its own, or in conjunction with MUNI or BART service.

Neighborhoods
- ○ Current
- ○ Proposed
- ◉ Both

■ Current
■ Proposed

WHAT ARE THE VIABLE COMMUTE ROUTES?

What train stations could a rider reach in 30 minutes? This network analysis shows the shortest route path between current and proposed stations, and the train stations that could be reached within 30 minutes. Proposed stations are based on the central point of the neighborhoods. Use the filters below to view viable commuting routes and train connections.

From Bike Station
All

To Train Station
All

Estimated Travel Time
0 to 30

- 0
- 1-5
- 6-10
- 11-15
- 16-20
- 21-25
- 26-30

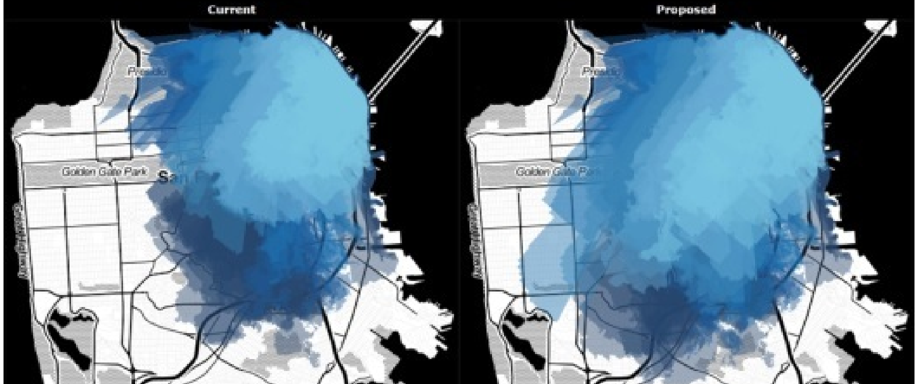
HOW FAR COULD YOU GO?

How far could a rider travel in 30 minutes? Could all of San Francisco be reached? This map shows distance traveled by time areas colored in 5 minute increments from current and proposed locations, in order to show where even more stations could be added to support commuters.

Estimated Travel Time
0 to 30

■ 5 ■ 10 ■ 15 ■ 20 ■ 25 ■ 30

Current **Proposed**



Map Tiles by Stamen Design, under CC BY 3.0 Data by OpenStreet Map, under CC BY SA.

DATAVIZ BY @AllanWalker.. @DataBlick

The preceding image is republished under the terms of fair use. It was created by *Anya A'hearn* and *Allan Walker*. (Source: https://public.tableausoftware.com/views/30Minutes___BayAreaBikeShare

As humans, we relate our experiences to each other in stories, and data points are an important component of stories. They quantify phenomena and, when combined with human actions and emotions, can make them more memorable. When authors create public interest story elements with Tableau Public, readers can interact with the analysis, which creates a highly personal experience and translates into increased participation and decreased abandonment. It's not difficult to embed Tableau Public visualizations into websites and blogs. It is as easy as copying and pasting the JavaScript that Tableau Public automatically renders for you.

Using Tableau Public increases accessibility to stories too. You can view data stories on any mobile device with a web browser and then share it with friends via social media sites such as Twitter or Facebook using Tableau Public's sharing functionality. Stories can be told with text as well as popular and tried-and-true visualization types such as maps, bar charts, lists, heat maps, line charts, and scatterplots. Maps are particularly easier to build in Tableau Public than most other data visualization offerings because Tableau has integrated geocoding (down to the city and postal code) directly into the application. Tableau Public has a built-in date hierarchy that makes it easy for users to drill through time dimensions just by clicking on a button. One of Tableau Software's taglines, *Data to the People*, is a reflection not only of the ability to distribute analyses sets to thousands of people at once, but also of the enhanced abilities of nontechnical users to explore their own data easily and derive relevant insights for their own community without having to learn a slew of technical skills.

Telling your story with Tableau Public

The Tableau Software was originally imagined in the Stanford University Computer Science department, where a research project sponsored by the U.S. Department of Defense was launched to study how people can rapidly analyze data. This project merged two branches of computer science—the understanding of data relationships and computer graphics. This mash-up was discovered to be the best way for people to understand and sometimes digest complex data relationships rapidly and, in effect, help readers consume data. This project eventually moved from the Stanford campus to the corporate world, and Tableau Software was born. The usage and adoption of Tableau has since skyrocketed. At the time of writing this book, Tableau is the fastest growing software company in the world and now, Tableau competes directly with older software manufacturers for data visualization and discovery, such as Microsoft, IBM, SAS, Qlik, and Tibco, to name a few.

Most of these are compared by Gartner in its annual Magic Quadrant. For more information, visit http://www.gartner.com/technology/home.jsp.

Tableau Software's flagship program, Tableau Desktop, is a commercial software used by many organizations and corporations throughout the world. Tableau Public is the *free* version of Tableau's offering, and it is typically used with nonconfidential data either from the public domain or the one that we collected ourselves. This free public offering of Tableau Public is truly unique in the business intelligence and data discovery industry. There is no other software like it—powerful, free, and open to data story authors.

Tip

There are a few terms in this book that might be new to you. You, as an author, will load your data into a workbook, which will be saved into the Tableau Public Cloud.

A visualization is a single graph. It, typically present on a worksheet. One or more visualizations can be on a dashboard, which is where your users will interact with your data.

One of the wonderful features about Tableau Public is that you can load data and visualize it on your own. Traditionally, this has been an activity that was undertaken with the help of programmers at work. With Tableau Public and new blogging platforms, nonprogrammers can develop data visualization, publish to the Tableau Public website, and embed the data visualization on their own website. The basic steps to create a Tableau Public visualization are as follows:

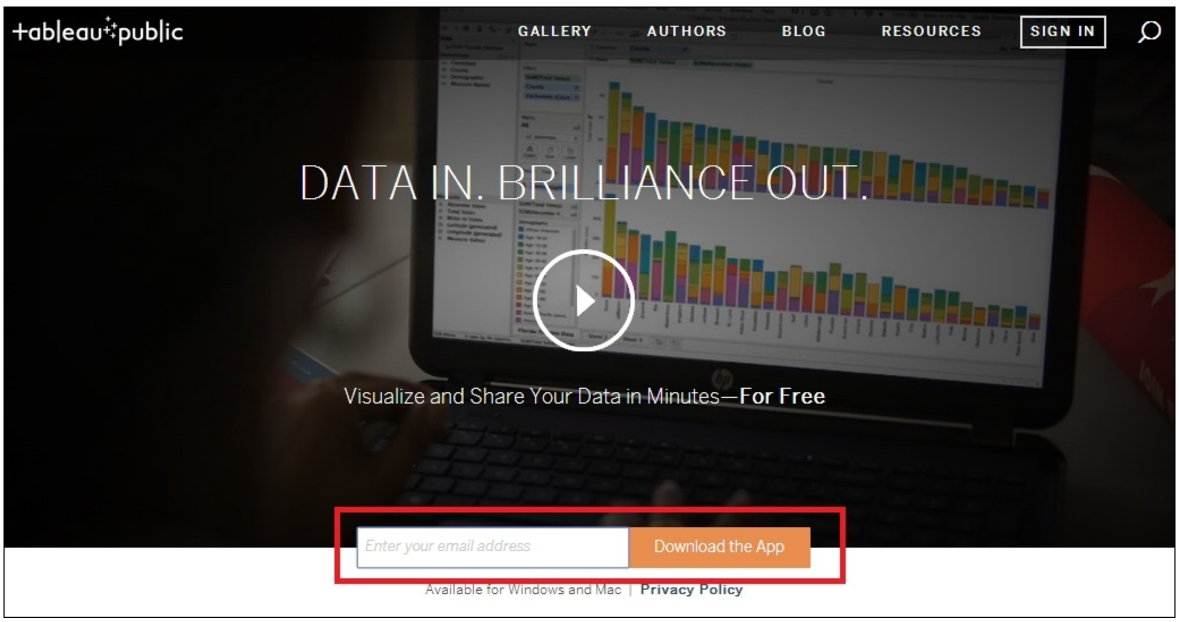
- Gather your data sources, usually in a spreadsheet or a `.csv` file
- Prepare and format your data to make it usable in Tableau Public
- Connect to the data and start building data visualizations (charts, graphs, and other objects)
- Save and publish the data visualization to the Tableau Public website
- Embed your data visualization in your web page by using the code that Tableau Public provides

Tableau Public is used by some of the leading news organizations across the world, including *The New York Times*, *The Guardian* (UK), *National Geographic* (US), the *Washington Post* (US), the *Boston Globe* (US), *La Informacion (Spain)*, and *Época* (Brazil). In the following sections, we will discuss installing Tableau Public. Then, we will take a look at how we can find some of these visualizations out there in the wild so that we can learn from others and create our own original visualizations.

Installing Tableau Public

Now, let's look at the installation steps for Tableau Public:

1. To download Tableau Public, visit the Tableau Software website at http://public.tableau.com/s/.
2. Enter your email address and click on the **Download the App** button located in the middle of the screen, as shown in following screenshot:



Tip

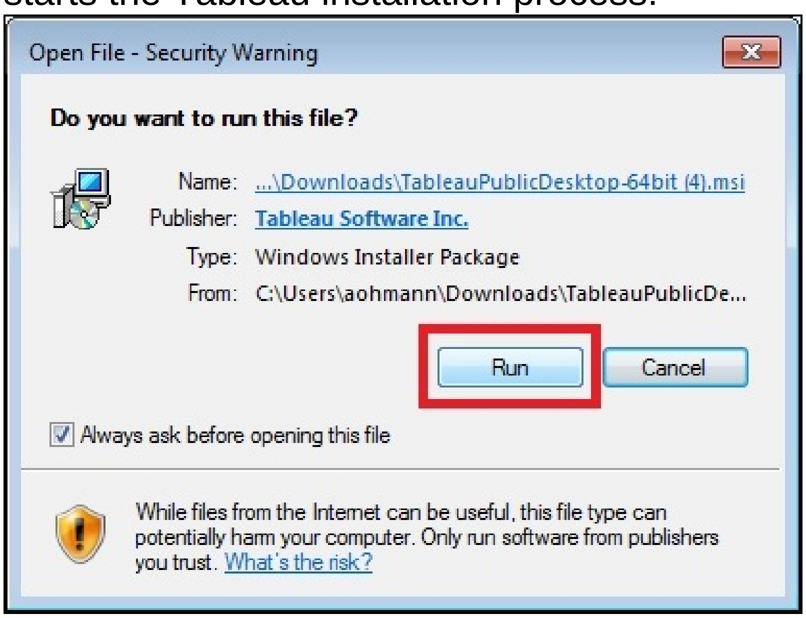
The downloaded version of Tableau Public is free and not a limited release or demo version. It is a fully functional version of Tableau Public.

- Once the download begins, a **Thank You** screen gives you the option of retrying the download in case it does not automatically begin or it is downloading a different version. The version of Tableau Public that downloads automatically is the 64-bit version for Windows. Users of Mac should download the appropriate version for their computers, and users with 32-bit Windows machines should download the 32-bit version.

Tip

Check your Windows computer system type (32-or 64-bit) by navigating to **Start** | **Computer** and right-clicking on the **Computer** option. Select **Properties**, and view the **System** properties. 64-bit systems will be noted as such. 32-bit systems will either state that they are 32-bit systems, or not have any indication of being a 32-or 64-bit system.

- While the Tableau Public executable file downloads, you can scroll to the lower part of the **Thank You** page to learn more about the new features in Tableau Public 9.0. The speed with which Tableau Public downloads depends on the download speed of your network, and the 109 MB file usually takes a few minutes to download.
- The `TableauPublicDesktop-xbit.msi` (where `x` has a value of either 32 or 64 depending on the version that you selected) file is downloaded. Navigate to that `.msi` file in Windows Explorer or the browser window and click on **Open**. Click on **Run** in the **Open File - Security Warning** dialog box that appears in the following screenshot. The Windows installer starts the Tableau installation process:

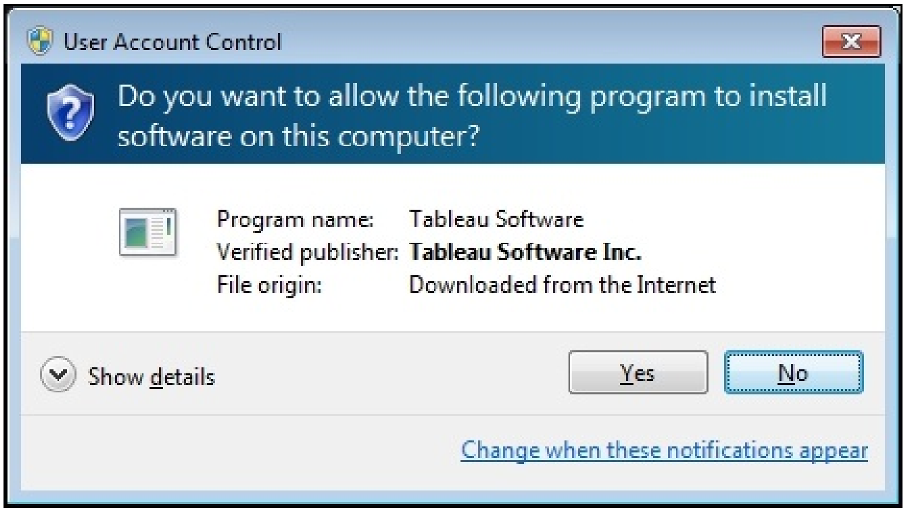


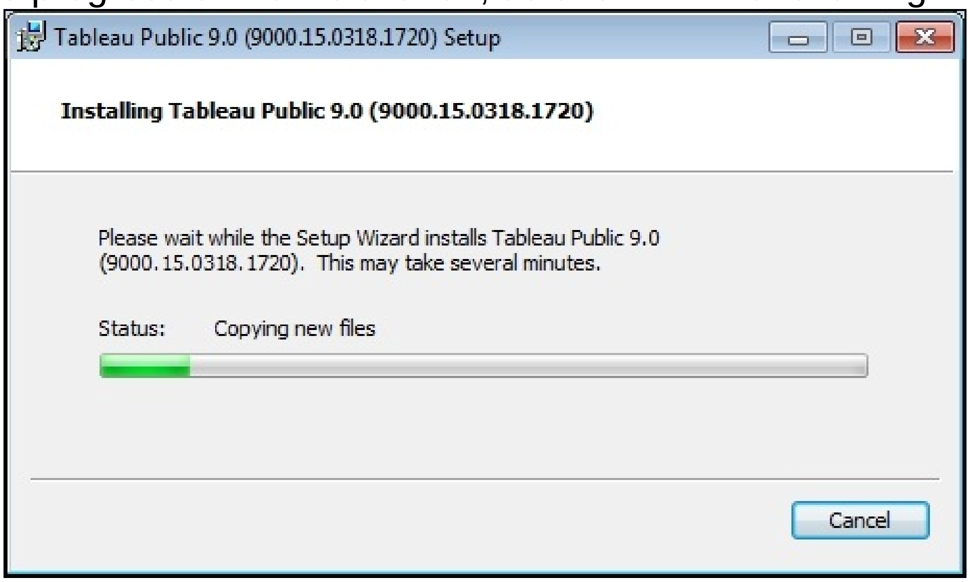
- Once you have opted to **Run** the application, the next screen prompts you to view the License Agreement and accept its terms:

- If you wish to read the terms of the license agreement, click on the **View License Agreement…** button.

(You can customize the installation if you want to. Options include the directory in which the files are installed as well as the creation of a desktop icon and a Start Menu shortcut (for Windows machines). If you do not customize the installation, Tableau Public will be installed in the default directory on your computer, and the desktop icon and the Start Menu shortcut will be created.)

- Select the checkbox named **I have read and accept the terms of this License Agreement** and click on **Install**.
- If a **User Account Control** dialog box appears with the **Do you want to allow the following program to install software on this computer?** prompt, click on **Yes**:



- Tableau Public will be installed on your computer, with the status bar indicating the progress of the installation, as shown in the following screenshot: 

- When Tableau Public has been installed successfully, the **Home** screen opens. The next section discusses its features.

Exploring Tableau Public

The Tableau Public home screen, as shown in the following screenshot, has several features that allow you to **Connect** to data, **Open** workbooks, and **Discover** the features of Tableau Public:

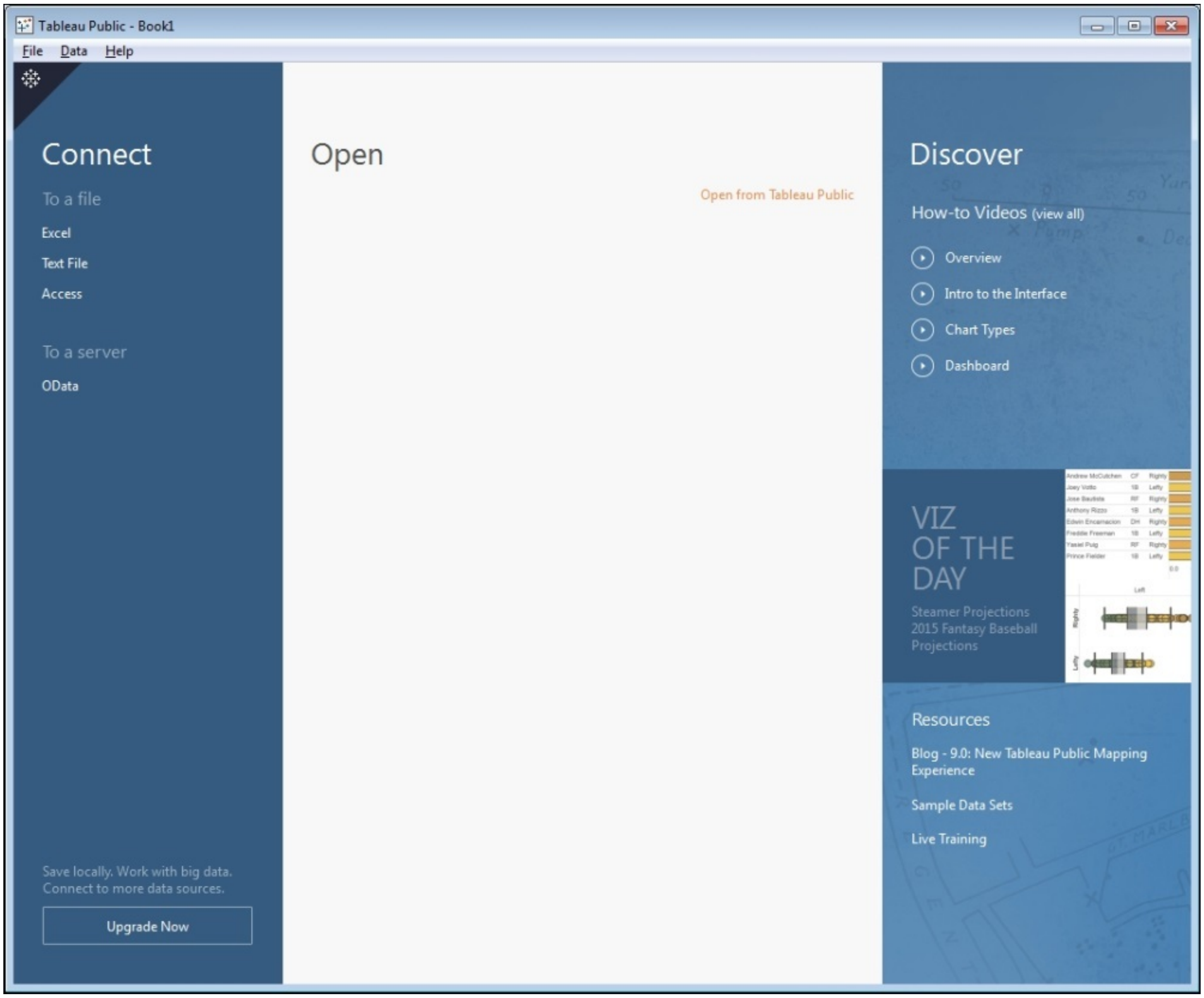
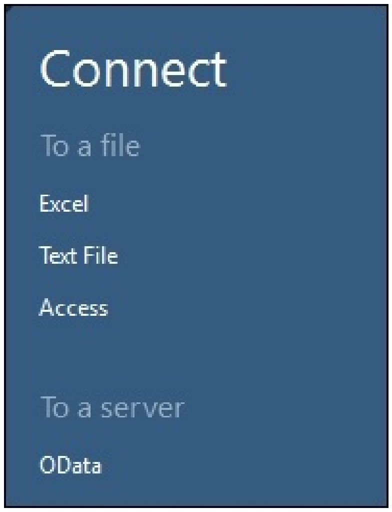


Tableau encourages new users to watch the video on this welcome page. To do so, click on the button named **Watch the Getting Started Video**. You can start building your first Tableau Public workbook any time.

Connecting to data

You can connect to four different data source types in Tableau Public, as shown in the next screenshot, by clicking on the appropriate format name:

- Microsoft **Excel** files
- **Text** Files with a variety of delimiters
- Microsoft **Access** files

- **OData** files 

Chapter 2, *Tableau Public Interface Features*, focuses on connecting to data sources and explains this in detail.

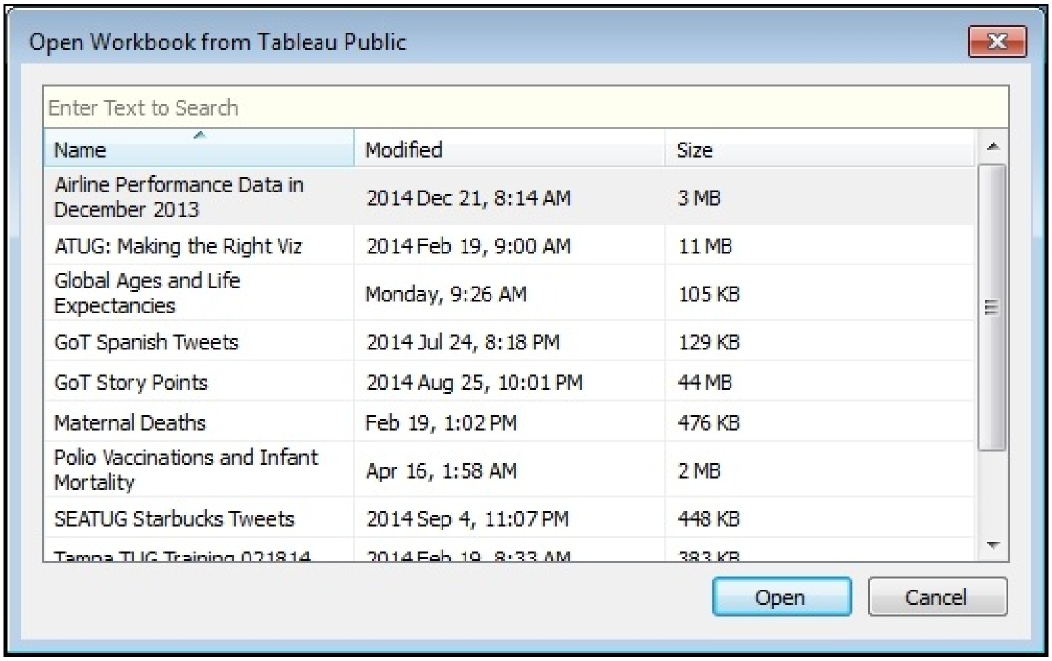
Opening files and creating your profile

You can open the files that you create in Tableau Public by clicking on the **Open from Tableau Public** link. When you click on the link, Tableau Public will prompt you to log in with the e-mail address that you have used to create your account, as shown in the following screenshot:

When you enter your e-mail and password, Tableau Public will verify it. Then, you will be able to select the file that you would like to open.

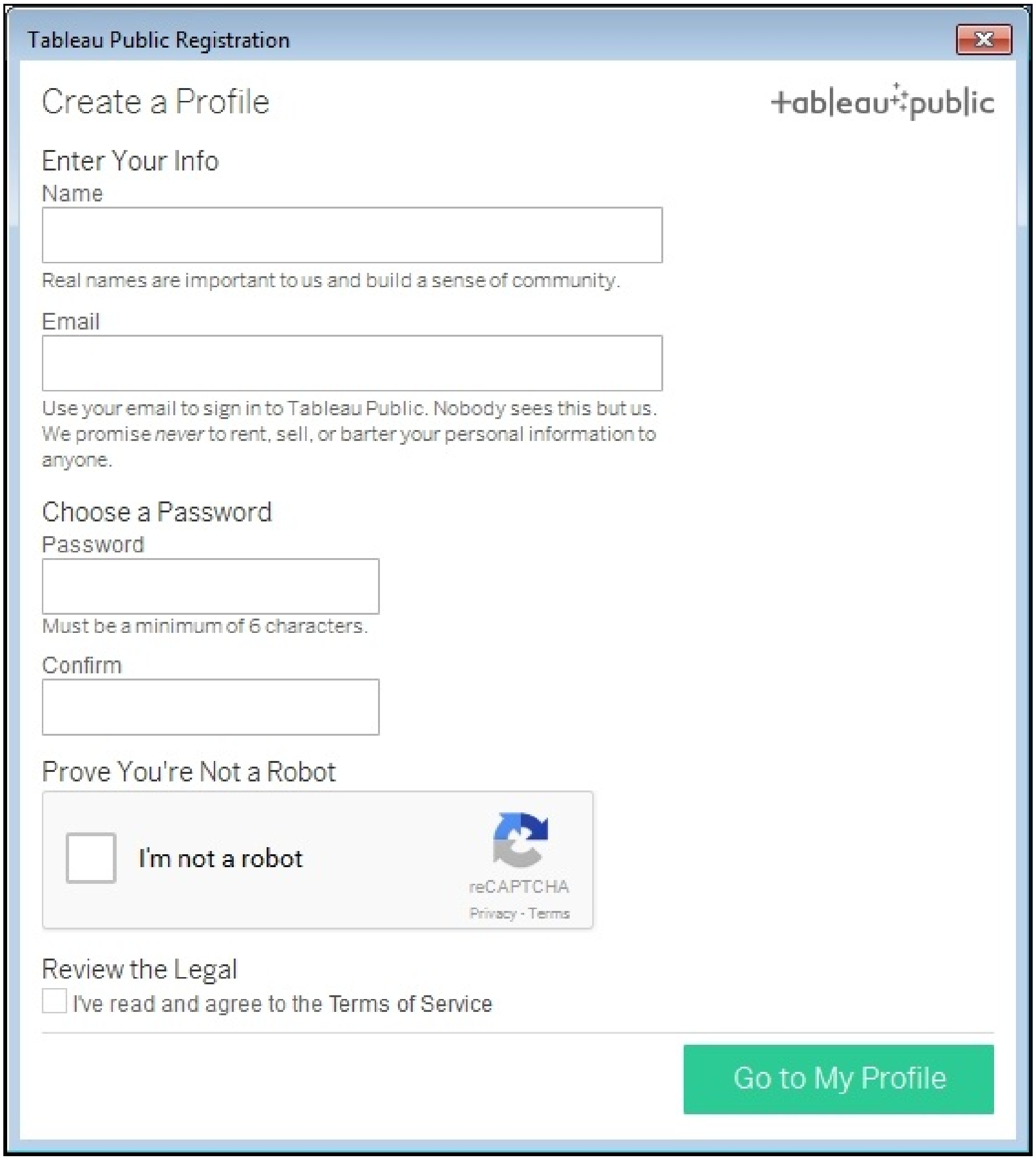
The list of files includes the names, modification dates, and size of the workbooks that you have saved in Tableau Public. It also includes the ability for you to search by entering a string of text.

When you find the workbook that you would like to open, click on **Open**, and then the most recently saved version will open in Tableau Public on your computer, as shown in the following screenshot:



If you have not created an account, click on the link at the bottom of the screen that says **Create one now for free**.

The **Create a Profile** screen requires you to enter information in several fields, as shown in the following screenshot:



Now, let's look at each of these fields:

- **Name**: Your **Name** will be displayed in your profile. You can edit this later if you want.
- Your **Email Address** is the identifier that you will use to log in to Tableau Public.
- **Choose a Password**, which must consist of at least six characters.
- **Confirm** your password.
- **Prove You're Not a Robot**. A CAPTCHA is generated for you to verify that you're not a robot when you click on it.

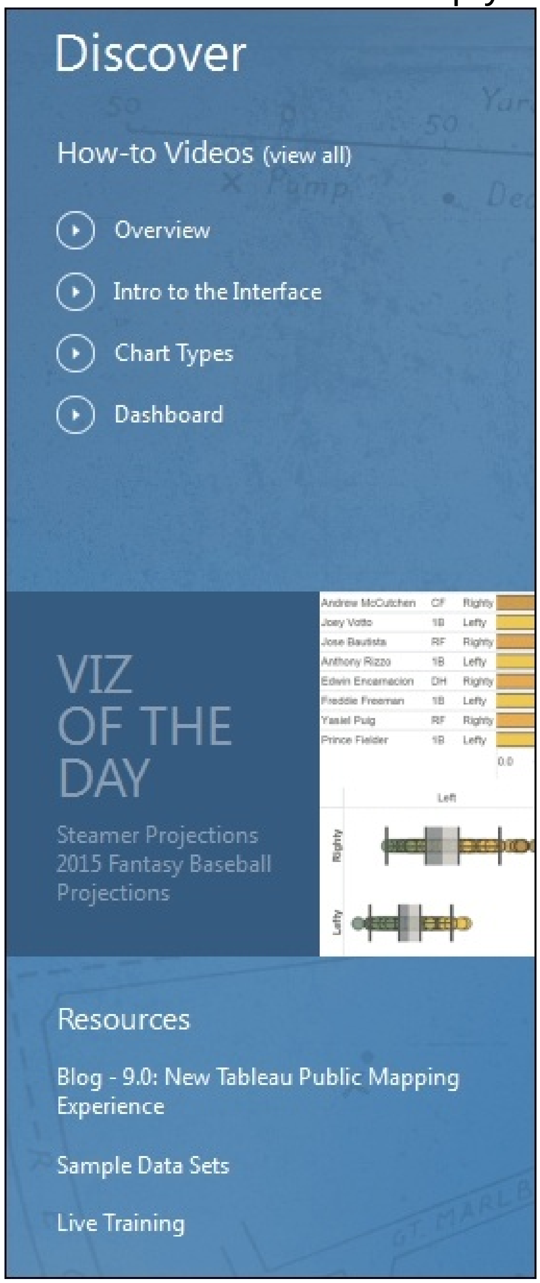
- **Review the Legal** requirements and agree to the terms of service.
- Click on **Go to My Profile** to complete the creation of your profile.

When you click on to **Go to My Profile**, your web browser will open your new profile page on Tableau Public. This is a page that displays information that you enter about yourself and your interests as well as a photograph of your choosing and links to other websites with which you're affiliated.

Your profile page also displays and allows you to manage your Tableau Public workbooks. We will discuss the profile in greater detail in Chapter 9, *Publishing Your Work*.

Discover

The right pane of the Tableau Public 9.0 home screen, as shown in the following screenshot, has several features that help you learn how to use the Tableau Public 9.0:



Let's take a look at each of these features:

- **How-to Videos**: Tableau has a wealth of online videos. You can view them by clicking on the video names in the pane.

- If you would like to explore other videos, click on the **view all** link next to the header. This will open Tableau's training video section of their corporate website in your browser. If the page doesn't open, you can access it by visiting https://public.tableau.com/s/resources.
- **VIZ OF THE DAY**: Tableau Public's staff selects a **VIZ OF THE DAY** from the recent publications on Tableau Public. These are the visualizations that are relevant to current events, explore important questions, and/or innovatively use the functionality of Tableau Public. You can subscribe to the **VIZ OF THE DAY** and view other selections by visiting https://public.tableau.com/s/gallery.

Resources that you can open include the Tableau Public blog, sample Data Sets, and links to live training. You can view all of these on Tableau Public's resources page in your Internet browser by visiting https://public.tableau.com/s/resources.

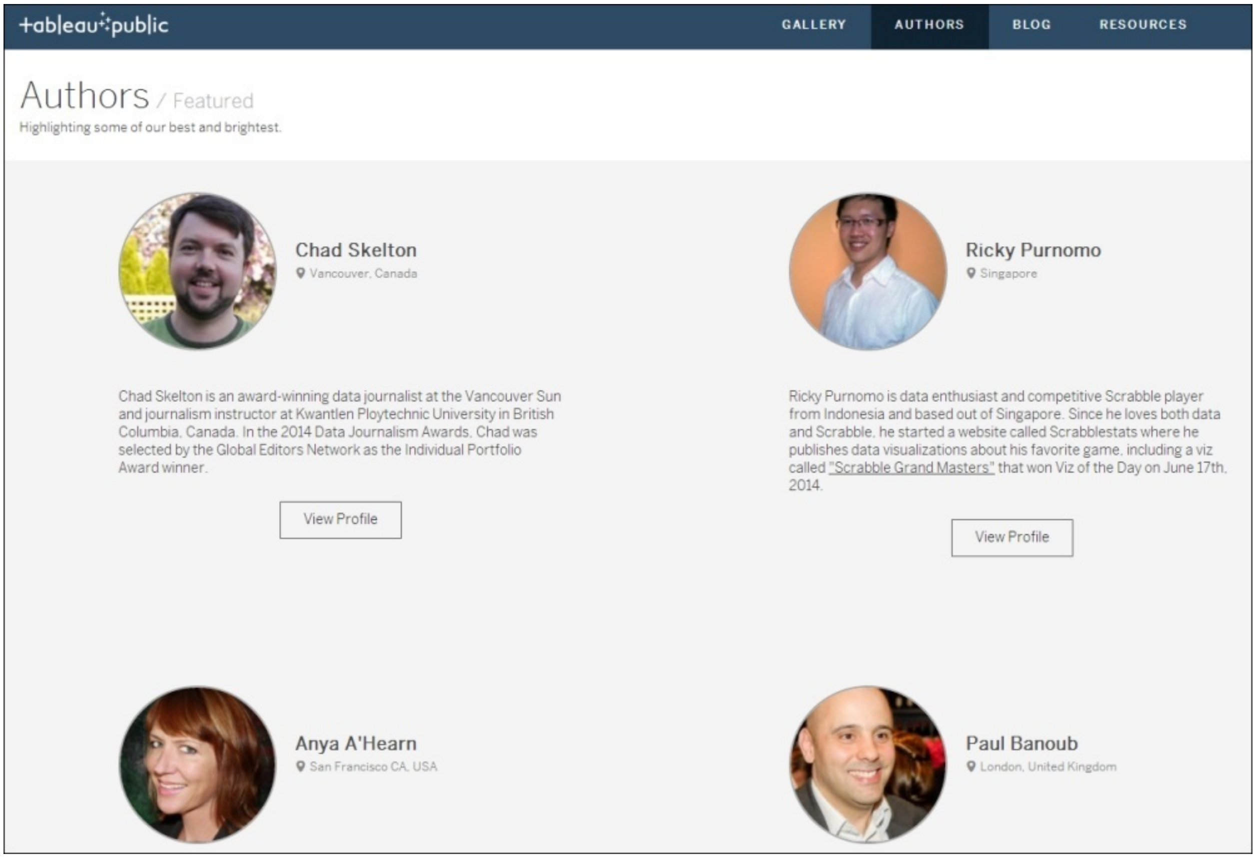
Exploring the visualizations of other authors

We often learn by viewing other people's work. So, let's take a look at a few data visualizations created by other authors. Note that most Tableau Public data visualizations allow you to download the entire workbook. If data is not readily downloadable on the workbook page, you can export the underlying data to Excel while inside the workbook by using the desktop client of Tableau Public. There are several great places to find the best Tableau Public data visualizations, including Tableau Public and the **VIZ OF THE DAY** galleries (for more information, visit https://public.tableau.com/s/gallery) and the Tableau Public blog (to have a look at the blog, visit https://public.tableau.com/s/blog).

To make use of a recommended authors and profile finder, visit https://public.tableau.com/s/authors.

The Tableau Public gallery is an excellent place to look at examples of works of others, and the Tableau Public team has curated a set of popular visualizations by topic and number of views.

The recommended authors page (https://public.tableau.com/s/authors) is a fun place to look at both well-known and established Tableau Public authors (bloggers, journalists, and the Tableau staff) as well as lesser known authors to explore their work. You can also access an author's profile page and see their work by clicking on the **View Profile** button under their name, as shown in the following screenshot:



+ableau public
GALLERY
AUTHORS
BLOG
RESOURCES
Authors / Featured
Highlighting some of our best and brightest.
Chad Skelton
Vancouver, Canada
Chad Skelton is an award-winning data journalist at the Vancouver Sun and journalism instructor at Kwantlen Ploytechnic University in British Columbia, Canada. In the 2014 Data Journalism Awards, Chad was selected by the Global Editors Network as the Individual Portfolio Award winner.
View Profile
Ricky Purnomo
Singapore
Ricky Purnomo is data enthusiast and competitive Scrabble player from Indonesia and based out of Singapore. Since he loves both data and Scrabble, he started a website called Scrabblestats where he publishes data visualizations about his favorite game, including a viz called "Scrabble Grand Masters" that won Viz of the Day on June 17th, 2014.
View Profile
Anya A'Hearn
San Francisco CA, USA
Paul Banoub
London, United Kingdom

Summary

In this chapter, we had a look at how Tableau Public is commonly used. We also discussed how to download and install Tableau Public, explore Tableau Public's features and learn about Tableau Public, and find other authors' data visualizations using the Tableau Galleries and Recommended Authors/Profile Finder function on the Tableau website. In the next chapter, we will explore data connections and manipulations in Tableau Public.

Chapter 2. Tableau Public Interface Features

The user interface for Tableau Public was created to be simple and intuitive. It comes with three primary features (as discussed in Chapter 1, *Getting Started with Tableau Public*), namely connecting to data, opening your work, and discovering Tableau Public. Since Tableau Public is a tool for data discovery as well as data visualization, the interface is designed to encourage discovery through the drag-and-drop features for data. The user interface for Tableau Public is segmented into separate areas, namely data elements, cards, shelves, and the canvas. The data is also divided into two general categories—dimensions and measures. By understanding how data interacts with the user interface, you can design, configure, and polish chart objects that will be built into worksheets. These worksheets can then be assembled into one or more dashboards.

In this chapter, we will cover the following topics:

- The Tableau Public user interface
- The side bar, including the **Data** window and the **Analytics** pane
- The toolbars and menus
- The **Columns**, **Rows**, and **Filters** shelves
- The **Marks** card
- The **Filters** and **Pages** shelves
- The **ShowMe** card

Touring the Tableau Public user interface

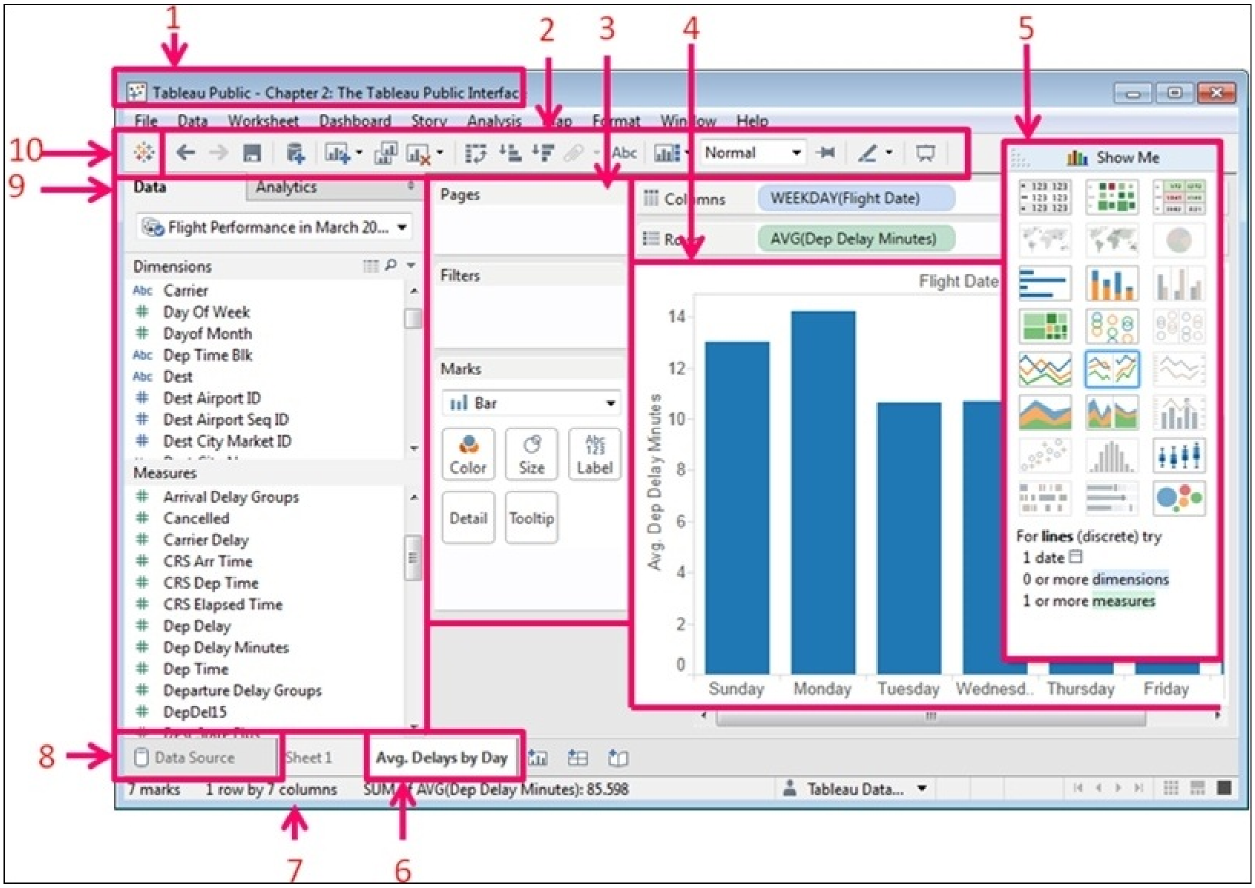
- In the previous chapter, we discussed how to download and install Tableau Public. We also saw what the start screen looks like and how you can use it to connect to data, explore your own work or that of others, or discover how to use the tool. On opening either a data file or an existing workbook with Tableau Public, you will see the worksheet view.

Note

There can be one or more worksheets in a workbook. Tableau Public extends this further by allowing you to place one or more worksheets in a dashboard, with all of this contained within the workbook.

The starting point that appears when opening a new Tableau workbook is the worksheet view. This is the working area where you can build your dashboard. Let's take a quick look at it.

The visualization shown in the following screenshot uses data from the **Federal Aviation Administration** (**FAA**) on every commercial flight at the domestic airports of the United States in March 2015 to average the departure delay (in minutes) for every weekday. You can download this data from http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time. Also, you can download the companion Tableau Public workbook from the profile of this book at https://public.tableau.com/profile/tableau.data.stories#!/. We will use this workbook to explore parts of the Tableau Public interface, as shown in the following screenshot:



The following are the parts of the user interface that are shown in the preceding screenshot:

- **Workbook (1)**: This is the workbook title, which is the name given to the workbook when you save it
- **Toolbar (2)**: This is where you can save your work, among other functions
- **Cards** and **shelves (3)**: These are the areas where you can add fields or filters to the visualization
- The **View**, or the **Visualization (4)**: This is the graph itself
- The **ShowMe** card **(5)**: This prompts you to create different visualization types based on the data selected
- **Sheet tabs (6)**: This allows you to create, rename, or duplicate sheets and dashboards
- The **Status** bar **(7)**: This shows the aggregated totals of the marks on your visualization
- **Data Source (8)**: Links back to data sources

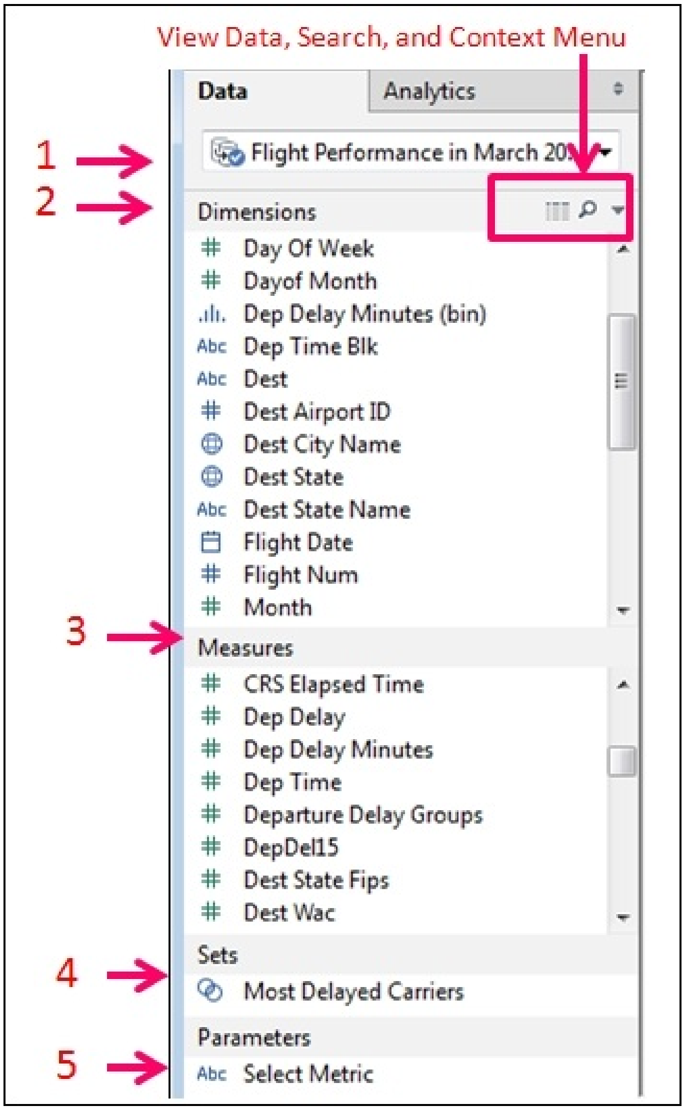
- The **Sidebar (9)**: This contains both the Data window and the Analytics pane
- The **Start** button **(10)**: This takes you back to the home screen

The side bar

Our discussion about the parts of the user interface starts with the side bar because it contains both the **Data** window and the **Analytics** pane. First, we will talk about the **Data** pane. After that, we will discuss the **Analytics** pane.

The Data pane

The Data pane is where your data sources load. In addition to listing the fields in your data source either alphabetically or by folder, the **Data** pane includes visual cues that tell you the type of each field. The following screenshot shows the visual cues of the **Data** pane:



Tip

Tableau Public scans the contents of your data source, groups fields into **Dimensions**, and **Measures** according to their field type. Before you start working with your data, you should look through the fields in the **Dimensions** and **Measures** panes to make sure that each field is in the proper place.

The **Data** pane has the following five different sections:

1. The **data source name (1)**: When you load data, you should provide

the data source with a name that identifies the contents, because that is what you and your consumers will see when they download your workbook from Tableau Public. Once you have added several data sources, you can condense their window in order to save space and then select different sources from the drop-down menu.
2. The **Dimensions** pane **(2)**: This includes categorical fields with qualitative data. The **Dimensions** pane typically consists of a string field, a date field, and a field that has geographical attributes, as well as unique identifiers, such as `ID` fields.
3. The **Measures** pane **(3)**: This usually includes quantitative fields with numerical data that can be aggregated. Tableau Public will automatically group numerical fields, except the ones with the `ID` string in the name as measures.
4. The **Sets** pane **(4)**: This includes user-defined, custom fields that interact just like dimensions and measures do. **Sets** pane can also create subsets of data that you can use just like dimensions.
5. The **Parameters** pane **(5)**: This includes dynamic placeholders that can replace constant values in calculated fields and filters. Parameters are unique to a workbook and not a data source. You'll see the parameters available in your workbook no matter which data source you are viewing.

From the **Data** pane, you can can create fields for the **Data** window, as follows:

1. Right-click on the data source to edit it.
2. Click on the **View Data** icon to see a sample of your data set.
3. Click on the search icon (the little magnifying glass) to search for fields.
4. Click on the arrow that points downwards, which is the **Context** menu, to create calculated fields, parameters, and change the sort/view options for the **Data** window.

Visual cues

Within the **Data** pane, each field name is displayed, but there is also a visual cue that tells you what type of field it is. Field types determine the function and capability of joining data sources. In addition to showing the field type, Tableau Public allows you to change a field's type by right-clicking on it and selecting the **Change Data Type** option. The following list shows the fields and their descriptions:

- The **Abc** field: This indicates that the field is a string field, which means that the contents of the field may include letters, special characters, or numbers
- The calendar icon: This indicates a date, datetime, or time field
- The **#** sign: This indicates a numeric field that can have any type of native numeric format, from `bigint` to `decimal`
- The globe icon: This icon indicates that the field has geographic attributes
- The paper clip icon: This indicates that the field is a group that you have created in Tableau Public
- The Venn diagram icon: This indicates a set

> **Tip**
>
> The parameters have their own data types, which can be established when you create them.

The Analytics pane

The **Analytics** pane is next to the **Data** pane. You can access it by clicking on its header. It provides you with the ability to add summaries, average lines, constant lines, distribution bands, medians, boxplots, forecasts, and other visual analytics to your visualization. You can then customize and edit them by using the reference line and formatting user interfaces.

The **Analytics** that you can add are dependent on the data elements in your visualization, which include the following:

- **Summarized Analytics**: This includes the following elements:

- A constant line: This is an integer on an axis that you can input
- An average line: This displays the mean of the measure that you have selected either across the table, pane, or cell
- Median line with 25% and 75% quartiles: This creates a median reference line and a quartile distribution band that you can edit
- A box plot
- Totals and subtotals

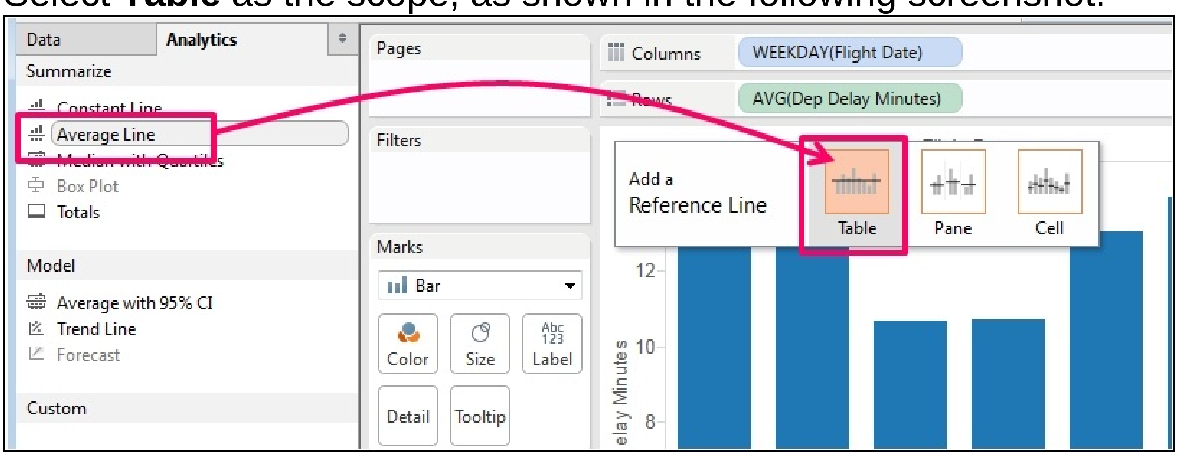
- **Modeling**: This includes the following elements:

  - **Average** (or mean) with confidence intervals
  - Trend lines with the most commonly used models (such as linear, exponential, and so on)
  - Forecast

- **Custom**: With the help of this element, you can add custom reference lines and distribution bands as well as box plots, which can also be used by right-clicking on an axis

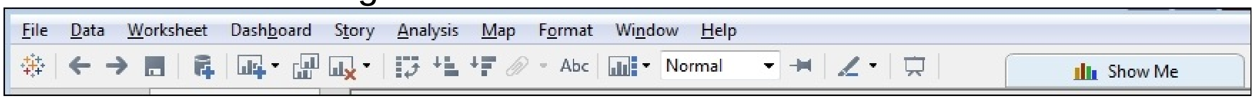
In the following screenshot, we added the average delay in minutes per day to the graph as a reference line by performing the following operations:

1. Click on **Analytics** to see the **Analytics** pane.
2. Drag **Average Line** to the *y* axis, which is the vertical axis.
3. Select **Table** as the scope, as shown in the following screenshot:



Menus and toolbars

The menus in Tableau Public are arranged and named in a way that is similar to those in other modern applications. Their primary uses are shown in the following screenshot:



Let's look at each of them in the following list:

- **File**: You can open and save your work to Tableau Public via this menu. Remember that Tableau Public does not auto-save.
- **Data**: From here, you can add new data sources and modify the existing ones.
- **Worksheet**: From here, you can create new worksheets, copy the visualization or data on your worksheet, and modify the title, caption, tooltip, and other features.
- **Dashboard**: This is used to create and format dashboards as well as to add actions.
- **Story**: This is used to add **Story Points**, which enhance the narrative capabilities of a data visualization. **Story Points** refers to a specific function in Tableau Public, and it's beyond the scope of this book.
- **Analysis**: This can be used to aggregate and disaggregate measures, create forecasts, totals, and trend lines, and create and edit calculated fields.
- **Map**: By changing the **Map** options, you can modify background maps and images, and add custom geocoding and WMS.
- **Format**: This is used to modify the appearance of visualizations.
- **Window**: This can be used to switch between the presentation and development mode as well as other views in the workbook
- **Help**: You can use this to get help and manage performance.

The buttons on the toolbar are graphically descriptive of their function. When you roll over each with your mouse, the instructions for use pop up as well.

The buttons of the toolbar include the following:

- The Start button: This takes you back to the Start screen

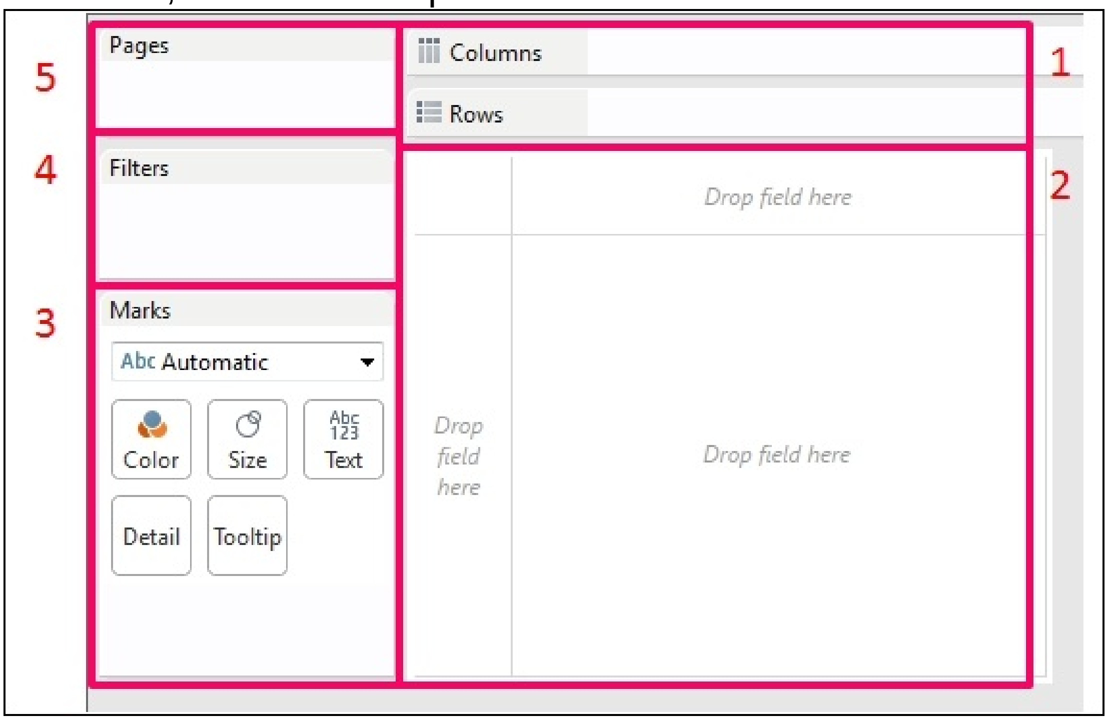
- The Undo button: This reverses the previous action that you took and can go as far as back to the state of the workbook when you opened it
- The Redo button: This repeats actions that you have reversed
- The Save button: This is critical, as Tableau Public does not auto-save
- Add New Data Source: This adds a new data source
- Add New Worksheet: This adds a new worksheet
- The Duplicate button: This duplicates the current worksheet
- The Clear button: This clears the current worksheet
- The Swap button: This swaps the fields on the **Rows** and **Columns** shelves
- Ascending: Sorts in ascending order
- Descending: Sorts in descending order
- Group: This is inactive in the view, as shown in the previous screenshot
- Show Mark Labels
- The Reset cards: These allows you to show legends and cards that you might have hidden or removed
- Fit: This allows you to change the fit of the visualization within a window
- Fit Axes: This allows you to set axis ranges
- Highlight: This allows users to click on dimension members and highlight the related records in other visualizations on the dashboard
- The Presentation mode: This hides the menus, the **Data** pane, or the **Dashboard** pane

Canvas and Column/Row shelves

Tableau Software has a user interface that is very different from that of the older reporting or data analysis tools that you may have used at work or in school. It uses a methodology of dragging and dropping objects for most of the functions that you need to perform to build a visualization. The areas of the workspace where you place objects are called **Shelves** and **Cards**. Many of the tasks in this book instruct you to drag a field to the **Columns** or **Rows** shelf.

The far right side of the screen is called the canvas area. It is where sheet objects, such as a chart, are built. The chart area itself starts out blank, and you must drag the fields that you want to analyze to an axis, header, or the **Columns** or **Rows** shelf, which also determines on which axis or header the field appears.

In the following screenshot, we have highlighted the **Columns** and **Rows** shelves, the canvas, the **Marks** card, the **Filters** shelf, and the **Pages** shelf. This screenshot shows how the workspace looks before we add fields to it, and the description is as follows:



- The **Columns** Shelf **(1)** is where you put fields that you want on the horizontal header or the *x* axis, that is, the axis that goes from left to right. The **Rows** shelf is where you put fields that you want on the *y* axis.
- The **Canvas (2)** has three places where you can drop fields, namely the *x* axis/header, the *y* axis/header, and the visualization space itself.
- The **Marks** card **(3)** includes the controllers that are used to mark color, size, text label, and shape. In addition to this, the level of the **Detail** and **Tooltip** controllers allows you to control the appearance of the visualization. You can change the mark type for each

individual axis as well as control the size, color, label, and shape. (Since the **Marks** card is critical to the function of the visualization, we will describe it in depth in the next section).

- The **Filters** shelf **(4)** is where you put fields that you want to use to limit the values included in your visualization. After dragging a field onto the **Filters** shelf, you can select values in the dialog box of that data element's Filters shelf.
- The **Pages** shelf **(5)** allows users to progress through your visualization based on the fields that you put on it. For instance, you can use a date field on the **Pages** shelf, and your users can click on changes over time without having to manually select the subsequent values.

Using the Columns and Rows shelves

As discussed earlier, the fields that you place on the **Columns** and **Rows** shelves will appear either on the *x* and *y* axis, or the row or column headers of the visualization respectively. The following are a few new concepts related to the uses of these shelves:

- Once a field is on the shelf, it is referred to as a **pill**, or an **active field**
- A pill on a shelf has a context menu

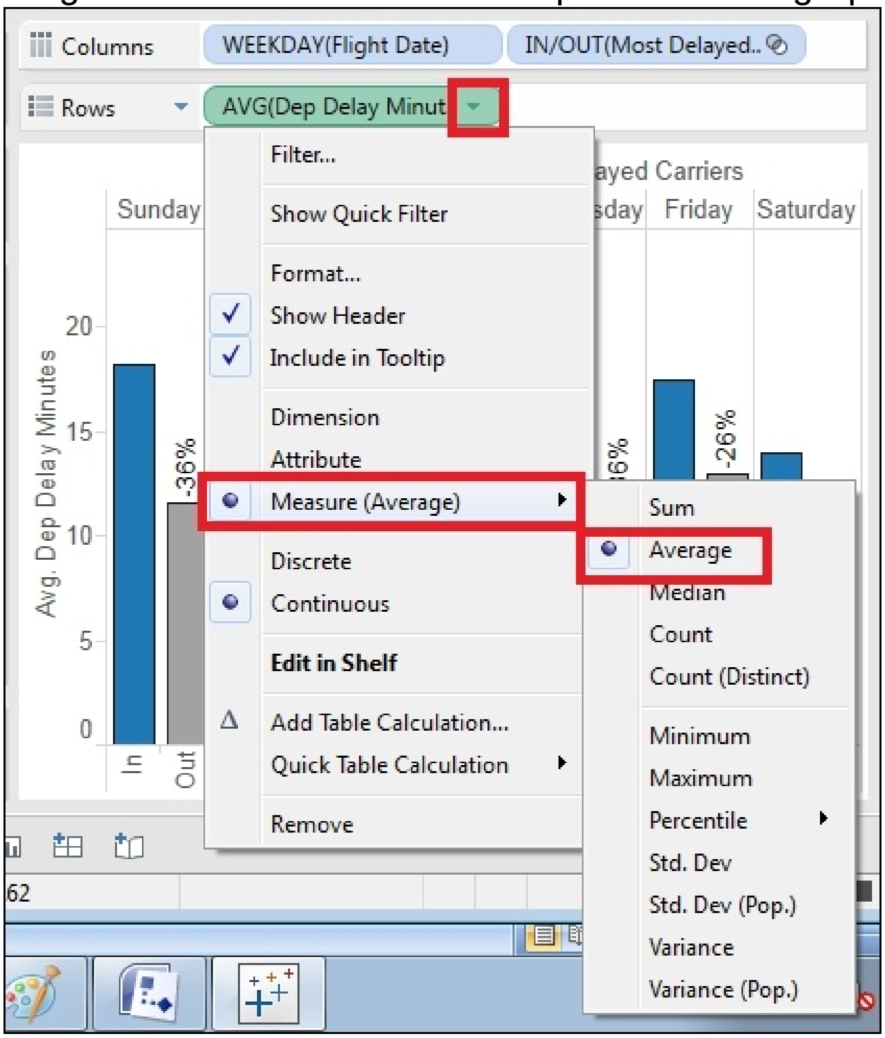
Note

The context menu can be opened by clicking on the small arrow that point downwards.

- Instances of fields that are discrete will appear in blue, and continuous fields will appear in green. The following is a description of both these fields:

  - Discrete fields have specific values, and the range of values is finite
  - Continuous fields, on the other hand, have an infinite range of values with infinite possibilities

- The Information Lab blog does a great job explaining the differences in Tableau Public; you can have a look at the explanation by visiting http://www.theinformationlab.co.uk/2011/09/23/blue-things-and-green-things/
- In case you have multiple pills on a shelf, the discrete pills appear first because they are used to group fields, and the continuous pills appear second because they are used to measure fields

The following screenshot shows an example of a basic graph that we created: 

In order to create the graph shown in the preceding screenshot, we will perform the following steps:

1. Drag **Flight Date** to the **Columns** shelf from the **Dimensions** pane.
2. Choose **WEEKDAY** as the part of the data that we want to show.

3. Add a set that we created that groups airlines by the criteria that we established.
4. When we drag the **Flight Date** field and the set to the **Columns** shelf, they both appear in the column headers. Drag the **Delay** minutes from the **Measures** pane to the **Rows** shelf.
5. When you do this, it defaults to aggregating as a **SUM**. Click on the **Context** menu of the pill and change the aggregation to **Average**.

Next, we will discuss how to use the **Marks** card to add color and labels.

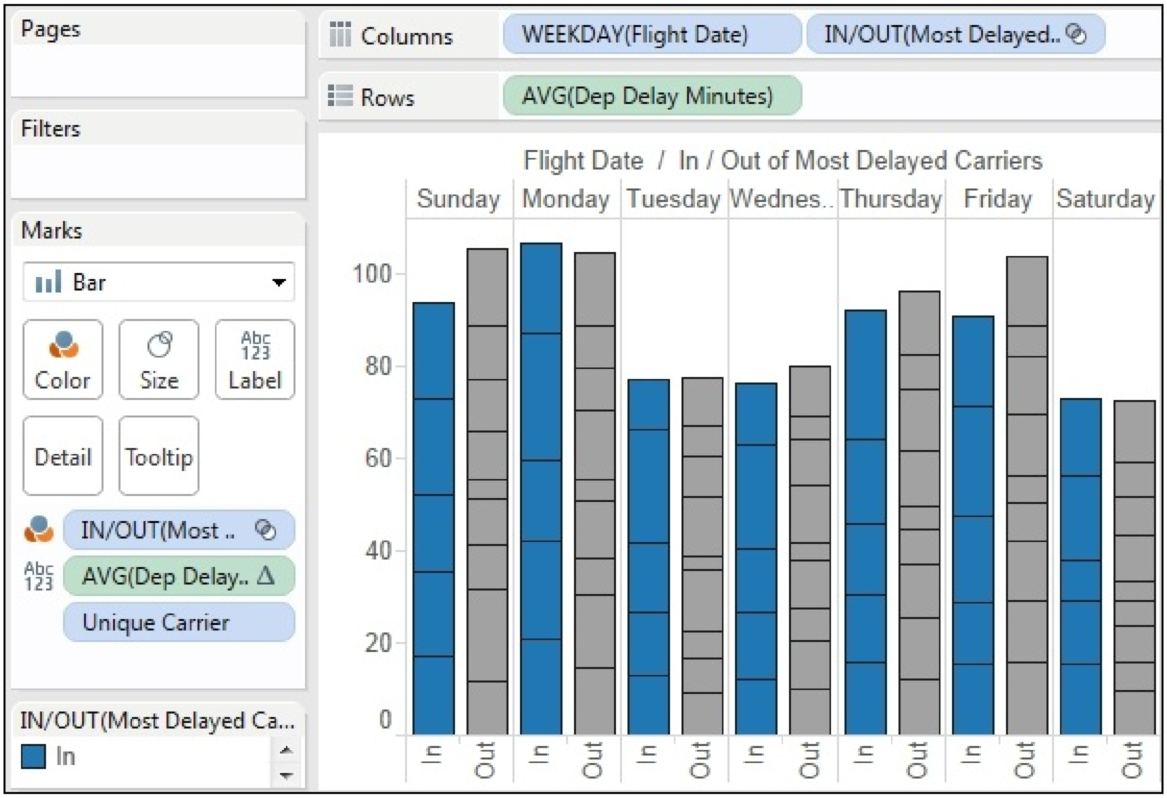
Using the Marks card

Another Tableau invention is the use of cards. Cards are containers for various controllers: which are dialogs in the Tableau workspace that allow various data elements and components to be configured. The most important card is the **Marks** card, which is in the most current version of Tableau Public. It has combined various controllers into one.

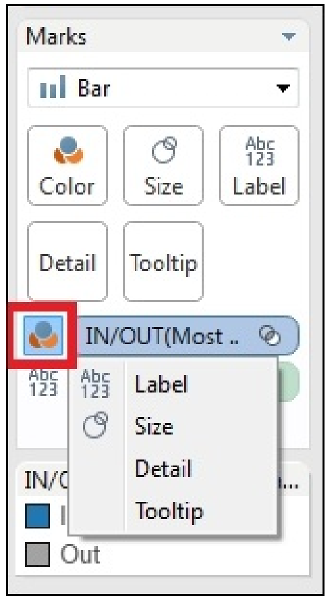
The **Marks** card is a compact yet highly functional area of the worksheet view that contains different controllers for data element chart properties (these data points on a chart are called marks). To use the **Marks** card, drag and drop data elements onto a corresponding shelf (such as **Colors**, **Label**, and **Size**). This will change the chart visualization by changing the chart mark properties.

The different controllers on the **Marks** card, which are commonly referred to as shelves, include the following:

- **Colors**: This changes the colors of marks in the chart (such as bar or line colors). In our example, the **Set** that shows whether an airline is in the top five worst airlines according to their average delay time is on the Color shelf.
- **Size**: This configures the sizes of data points or elements in charts.
- **Label/Text**: This adds a label to the chart for data points, bars, groupings, or lines. In our example, the percentage difference in average delay time between the top five worst airlines and all the others is on the **Label** shelf.
- **Details**: This adds details to the chart or data points, allowing you to keep the main structure of the chart. However, it further categorizes it in detail and with more granularity. For example, if we add **Unique Carrier** to the **Detail** shelf, our visualization will aggregate the average departure time by day for each **Unique** carrier, which means that there is now a bar segment for each carrier, and the story that our graph tells will change dramatically, as shown in the following screenshot:



- **Tool Tip**: This allows you to add context and calls to action, which are critical for telling stories because they instruct users how to progress to the next step and they better illustrate how a data point relates to a user's interests.
- **Shape**: This sets the data point shape in the chart visualization.
- **Path**: This is typically used for routes on a map. This controller allows a path to be sequentially built and placed on a map visualization. This is commonly seen in tornado and hurricane tracking maps visualizations.

Once a field is on a shelf on the **Marks** card, it shows up under the shelves with a small icon for the shelf type followed by the field name, as shown in the following screenshot: 

You can change the shelf on which a field is placed by clicking on the small icon and then selecting the shelf to which you'd like to move the field, as shown in the previous screenshot.

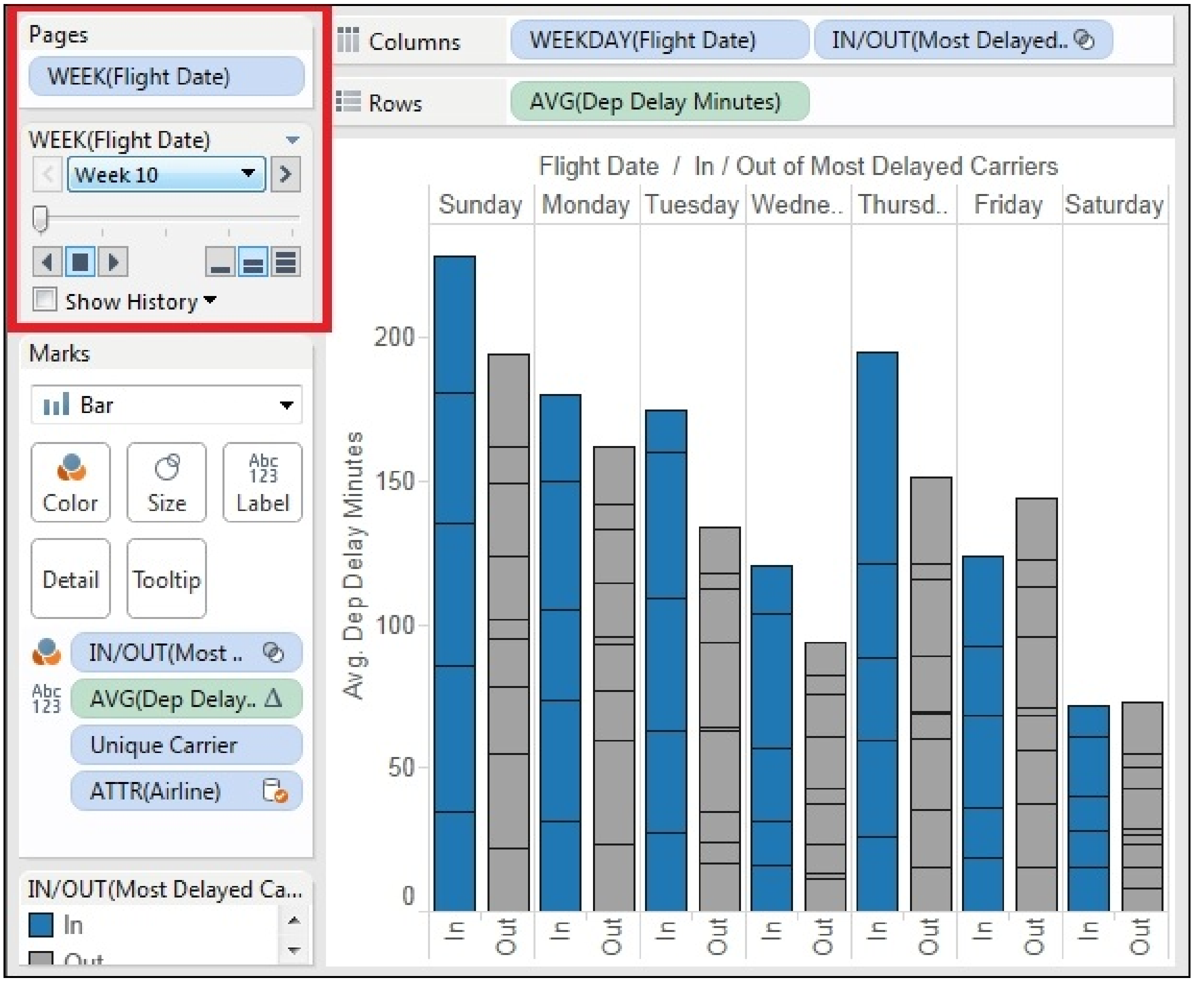
The Filters and Pages shelves

The **Filters** shelf is where you drag fields to limit the data points that your users see in your visualization. When you drag it to the **Filters** shelf, Tableau Public will prompt you to filter a field according to its type.

You can also filter a visualization based on the inclusion in a **Data** set. If you want to show the filters to users, you can right-click on the fields in the data window to **Show Quick Filters** as well. We will describe filtering in depth in future chapters.

In the following screenshot, we have added **WEEK(Flight Date)** to the **Pages** shelf. The controller appears here. We can set the speed at which the visualization changes. In the Tableau Public desktop client, the visualization will progress sequentially once you click on the **Play** button. However, online, it will not progress automatically. Your user will need to click on the **Play** button to proceed.

Since the functionality is limited online and that is where your users will interact with your work, we will condense the discussion of the **Page** shelf:

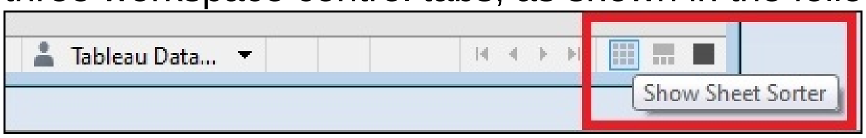


The workspace control tabs of Tableau Public

The tabs at the bottom of the Tableau Public screen are either sheet or dashboard tabs. The tabs with a series of rows and a column on them are **sheet tabs**, where you can design objects that fit on one sheet. The tabs with a rectangle divided into four quadrants are **dashboard tabs**, where you can add one or more sheets in order to form a unified dashboard. When you click on one of the new (blank) worksheet or dashboard tabs, that tab opens and names the sheet or dashboard the next sequential name, such as **Sheet 3** or **Dashboard 2**. The sheet or dashboard is blank when you first open a new tab.

> ## Tip
>
> A dashboard is really just a group of worksheets on the same page. You can assemble a dashboard after creating the various component worksheets that it will contain. Dashboards are assembled and configured by using the **Dashboard** tabs in Tableau Public, which are available in the lower row of the tabs in the Tableau Public interface.

The lower-right hand side of the Tableau Public interface also contains three workspace control tabs, as shown in the following screenshot:



Let's take a look at each of these tabs:

- **Show Sheet Sorter**: This shows the sheets in your workbook and allows you to reorder them easily
- **Show Filmstrip**: This shows the visualization with the thumbnails of other worksheets, which is shown in the previous screenshot
- **Show Tabs**: This shows the default view of the visualization and the named tabs of other worksheets, as shown in the previous screenshot

The Show Me tool

A very useful tool in Tableau Public is the **Show Me** tool, which is available as a floating window when you click on the **Show Me** button on the upper-right hand side of the Tableau Public interface.

This tool, as shown in the following screenshot, is a helpful aid when you wish to select various chart types that can be used with your data:

> ## Tip
>
> The **Show Me** tool is an optional tool. Charts can be created without any help from this tool.

The **Show Me** tool also contains a list of tips to add data to shelves and cards, such as the number of dimensions and measures necessary to create the desired chart type.

The **Show Me** tool also suggests graph types based on the view that the user has created and/or the fields that the user has selected from the **Data** window. Based on the selections and the chart type selected in **Show Me**, the tool also instructs the user what field types are needed to create a certain graph type.

For example, selecting a continuous date field will make the **Show Me** tool prompt the suggestion of a line graph. Selecting a geocoded or latitude/longitude field will make it prompt the selection of a geographic map.

In Chapter 4, *Visualization - Tips and Types*, we will discuss the **Show Me** tool in detail and have an overview of the chart types that are available in Tableau Public. Many of these chart types are covered by the **Show Me** tool suggestions.

Tip

The **Show Me** tool dialog does not auto-hide once the user makes a selection from it. Click on the **Show Me** button again to hide the **Show Me** tool dialog.

Summary

In this chapter, we learned the Tableau Public user interface, from the welcome screen to the worksheet and dashboard tab views. We discussed the concepts of shelves and cards, walked through an example of how the Marks card affects a data visualization, and had a look at how to create a dashboard from the various worksheets created in a Tableau Public workbook. Lastly, we discussed the Show Me tool and how it can aid you in choosing appropriate data visualization types for your data.

In the next chapter, we will discuss the various chart types that you can create in Tableau Public and what some of the best practices and uses are for the chart types.

Chapter 3. Connecting to Data

Visualizations depend on the data in them, and however aesthetically pleasing your visualization might be, it may be misleading or even wrong, unless the data has been formatted, aggregated, and properly represented.

This chapter discusses the major elements of finding, cleansing, understanding, formatting, and aggregating data that you will need to understand in order to produce accurate visualizations that tell compelling stories, including the following elements:

- Where you can get publicly available data and how to use it
- What tables and databases are
- The data formats that Tableau Public connects to
- Databases, tables, dimensions, facts, and field formats and conventions
- Preparing data to load it into Tableau
- Connecting to the data from Tableau Public
- Using the data interpreter
- Pivoting fields

Public data

The data sets that are publicly available or the ones that you have compiled on your own, are ideal for Tableau Public. Since all users will be able to download this data and create their own visualizations once you have published your workbook, your data set should not contain information that may be considered sensitive, which can be anything that can be used to identify a private individual or reveal confidential corporate information or intellectual property.

Public data is readily available online. Tableau Public maintains a catalog of publicly available data. Much of this data is produced by various governments, economic groups, and sports fans, along with a link to, and a rating for each source. This catalog is updated monthly, and it is a great introduction to using publicly available data. You can find it at http://public.tableau.com/s/resources.

The **Google Public Data Explorer** has a large collection of public data, including economic forecasts and global public health data. This tool is unique because it allows users to make simple visualizations from all the original data sources without having to investigate the source data, though most of it is available by linking available resources.

There are several tools available for the scraping of data from public sites too, such as ScraperWiki, import.io, and IFTTT, among others. These industries and such tools pertaining to the industry evolve rapidly. Therefore, we will not discuss any specific tool. Social media applications, such as Twitter, have made it possible for individuals and companies to build application programming interfaces (APIs) to connect to their data streams. This is useful for nonprogrammers because it's often free of cost if you wish to scrape data about specific topics, hashtags, or users with a minimal amount of coding.

Not all data is public data, and it's very important to determine whether a data set is public before using it. None of us wants to end up being sued, with a ruined reputation, or as a victim. If your source data set has identifying characteristics, first and last name, address, and financial, geolocation, medical, or federally or state protected data, it should be removed or de-identified and then saved separately before saving the visualization to Tableau Public (or not used at all). Each state has guidelines on what is considered protected information, and it's a good idea to check the restrictions in case there's even the slightest chance that your data set has sensitive information in it.

Additionally, data from a corporation should never be used unless the corporation has given the permission to use it. (Did we mention lawsuits? Being fired also isn't any fun.)

Tables and databases

Once you have found the data set that's ideal for your visualization, it's helpful to know how data stores are structured and what the different terms are.

Data is stored in tables. A table is an array of items, and it can be as simple as a single word, letter, or number, or as complicated as millions (or more) of rows of transactions with timestamps, qualitative attributes (such as size or color), and numeric facts, such as the quantity of the purchased goods.

Both a single text file of data and a worksheet in an excel workbook are tables, though this may not be apparent. When grouped together in a method that has been designed to enable a user to retrieve data from them, they constitute a database. Typically, when we think of databases, we think of the **Database Management Systems** (**DBMS**) and languages that we use to make sense of the data in tables, such as Oracle, Teradata, or Microsoft's SQL Server. Currently, the Hadoop and NoSQL platforms are very popular because they are comparatively low-cost and can store very large sets of data, but Tableau Public does not enable a connection to to these platforms. They are considered enterprise tools that should be used with Tableau Desktop Professional. Therefore, our discussion about these tools is limited.

Tableau Public is designed in such a way that it allows users in a single data connection to join tables of data, which may or may not have been previously related to each other, as long as they are in the same format. In other words, multiple CSV files or worksheets can be joined in the same excel workbook. Then, users can specify the conditions under which they need to retrieve data from the tables and how to aggregate it (examples are given in following section). Thus, that data connection becomes a de-facto database.

The most common format of publicly available data is in a text file or a **Character-Separated Values** (**CSV**) file. CSV files are useful because they are simple. The rows of data, which may or may not contain a header row, are separated by line breaks. The fields within each row can be separated by a character. Typically, this character is a comma, pipe, or tab. Commas present difficulties because the content of the fields can contain them, which causes the text to shift into a new column.

Many public data sources do allow data to be downloaded as Excel documents. The World Bank has a comprehensive collection, and we will demonstrate the connective capabilities of Tableau Public using one of its data products. Tables can be joined in Tableau Public by manually identifying the common field among the tables.

The data sources that Tableau Public connects to

Tableau Public connects to four different data sources, namely Access, Excel, text file (CSV or TXT), and OData; the first two data sources are bundled with Microsoft Office (in most cases), and the second two are freely available to everyone, regardless of the operating system that they are using. Text files are the default source of origin for most of the data that we will discuss, and anyone can create and distribute them.

Tableau Public does not connect to enterprise tools such as Teradata, Oracle, or Hadoop, and it does not connect to SQL Server Management Studio, though it does connect to flat files' output from these tools and other **Online Analytical Processing** (**OLAP**) systems. SQL Server Management Studio is free for noncommercial use, and it's common to use it to design basic star schemas and clean noncommercial data. Tableau Public is free because it's assumed that people use it only for their personal endeavors and not projects that generate revenue for their employers. If this is not the case, they should upgrade to the Tableau Desktop Professional edition.

The databases, tables, dimensions, facts, field formats and conventions

Data that is retrieved from different sources will invariably have different structures. Some of these data resources need more formatting than others in order to turn them into clean, usable tables.

As previously mentioned, a table might be as simple as having a single digit in a text file. As long as users know what that digit represents, they can assign a qualitative or quantitative value to it. Imagine a situation where you are collecting rainfall measurements. Entering the amount of rainfall as subsequent rows of text into a new file constitutes a table.

The amount of rainfall is a measure; it is a quantitative fact. A dimension is a field that contains qualitative data. In this case, both the time of the day and the location of the measurement will be dimensions. Dimensions are typically formatted as date, string, or character fields, while measures are formatted as numbers. Text files do not have field formats, which are considered metadata, but Microsoft Excel and Microsoft Access do contain this information.

> ## Note
>
> It's important to make sure that the formats for fields of the same type (the date or the primary/foreign key) are consistent between worksheets in a workbook or tables in an Microsoft Access database, because Tableau Public automatically joins only the fields with the same format and the exact same name (including capitalization). If your field names are not the same but they should be joined using join conditions, you can join them manually.

Another common dimension is a unique identifier, which assigns a non repetitive value to each object in a set. A phone number is a unique identifier as it is related to only one phone at a time. The same is the case with a social security number. Within a data set, a person's name will be considered a unique identifier if it is not repeated; if it were, then a different unique identifier would be used to identify individual people. Thus, it's common to use numerical fields as primary keys for individuals, and these numerical fields can be used across multiple tables and across different dimensions.

Tables need to be structured so that the field names (dimensions and facts) go across and the rows of data (the dimension values and measure facts), go down the table. There are some databases that transpose data because their querying engines are optimized to search across columns rather than down the rows, but most DMBSes are not columnar, and Tableau is not built to search rows.

The following table is a great example of a table that is structured properly. This is the 2012 NFL performance data that is freely available:

Player	Team	Receptions	Yards	Average
Calvin Johnson	Det	122	1964	16.1
Wes Welker	NE	118	1354	11.5
Brandon Marshall	Chi	118	1508	12.8
Andre Johnson	Hou	112	1598	14.3
Jason Witten	Dal	110	1039	9.4
Reggie Wayne	Ind	106	1355	12.8
A.J. Green	Cin	97	1350	13.9
Demaryius Thomas	Den	94	1434	15.3
Tony Gonzalez	Atl	93	930	10
Sharod White	Atl	92	1351	14.7

Each column is a field; the dimensions are **Player** and **Team**, and the measures are **Receptions**, **Yards**, and **Average**. The primary key is `Player`. There is only one row for each player. `Team` is the foreign key, as it may be the primary key in other tables, such as the aggregations by team.

Conversely, knowing what not to do is as instructive as knowing what to do. The following table, which shows the population (in millions) by country, is a good example of what not to do:

Country	1950	1951	1952	1953
Afghanistan	8,151	8,277	8,407	8,543
Africa	229,895	234,594	239,501	244,621
Albania	1,215	1,240	1,269	1,303
Algeria	8,753	8,953	9,141	9,326
American Samoa	19	19	20	20
Andorra	6	7	7	8
Angola	4,148	4,219	4,297	4,377
Anguilla	5	5	5	6
Antigua and Barbuda	46	48	50	51
Argentina	17,150	17,506	17,865	18,224

The problem with this table is that the years, which actually are qualitative descriptions of when each population measurement was made, are used as separate columns even though the year is a dimension and should run down the page. (Dates are dimensions too and not facts, because they describe something qualitative). If we loaded this table in Tableau Public, we would see a separate measure field for each year because Tableau Public recognizes each column as a distinct field. (In one of the following examples, we will use Tableau Public's new data interpreter to structure the data source properly).

The correct structure for this table will have three columns, namely `[Country]`, `[Year]`, and `[Population]`, with a separate row for each combination of country and year.

Connecting to the data in Tableau Public

Tableau Public has a graphical user interface (GUI) that was designed to enable users to load data sources without having to write code. Since the only place to save Tableau Public documents is in Tableau's Cloud, data sources are automatically extracted and packaged with the workbook. (The ability to save extracts as separate documents or open extracts and share them with different users is a feature of Tableau Desktop Professional).

Connecting to data from a local file, that is, an access, excel, or text file saved on your computer, takes several steps that have little variability by data source, which will be illustrated as follows with detailed screenshots:

1. Click on the **Connect to Data Link** option from the **Data** menu.
2. Select the data source type.
3. Select the file or website to which you want to connect.
4. For a Microsoft Access, Microsoft Excel, or a text file, determine whether the connection is to one table or multiple tables or it requires a custom SQL connection:

    - If the connection is to one table, select the table.
    - If the connection is to multiple tables, select the option for the connection to multiple tables and identify the join conditions. We will discuss this in detail in the next section.
    - Alternatively, you can type or paste a custom SQL.

5. When all the selections have been made, click on **Ok**.

Now that you have learned what data sources look like and how they are structured, we will give you a couple of examples of data connections.

In the first exercise, we will connect to the World Bank's environment indicators. You can download this data, which is formatted either for Microsoft Excel or as a text file, at www.worldbank.org.

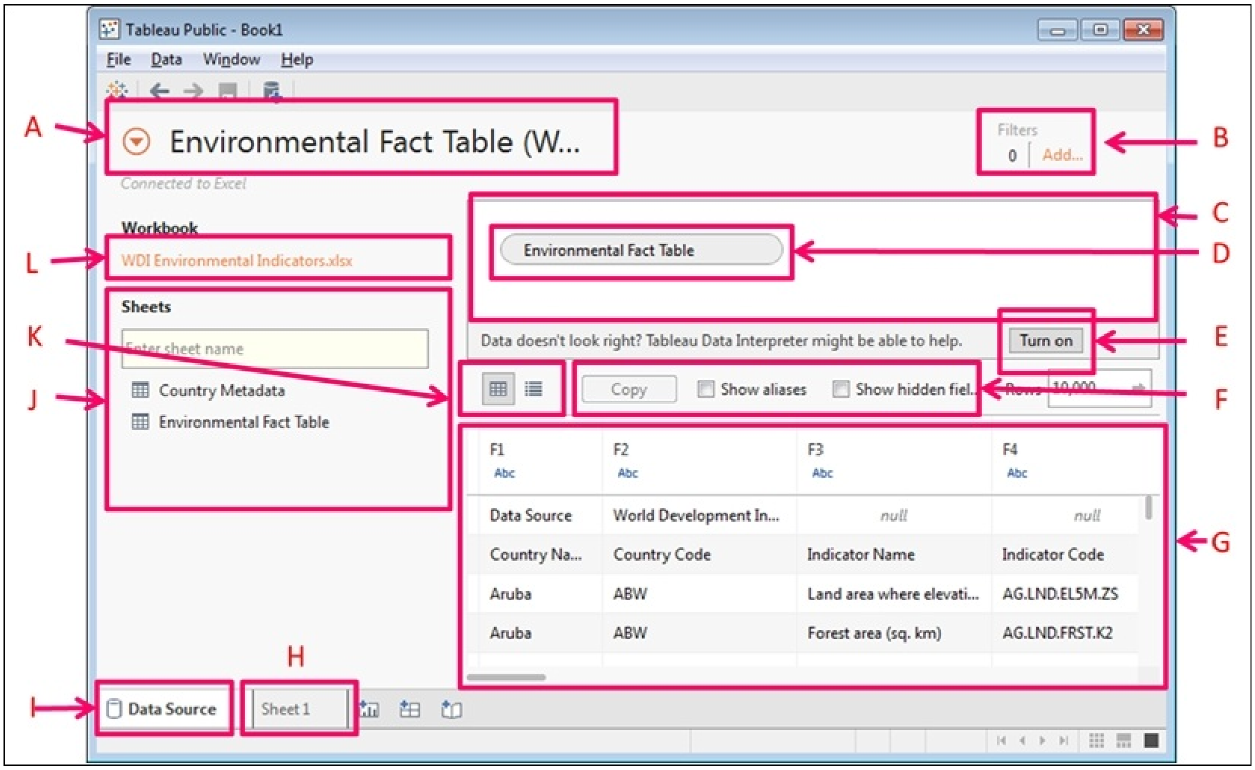
This data source is not formatted properly, which is shown in the following screenshot. It has a spacer row between the top of the worksheet and the headers' rows, and the years in which the measurements were taken were distributed as columns rather than individual dimensional values in a row.

Once we connect to the file, we will use Tableau Public's new data interpreter to clean up and pivot the rows:

	A	B	C	D	AS	AT	AU
1	Data Source	World Development Indicators					
2							
3	Country Name	Country Code	Indicator Name	Indicator Code	2000	2001	2002
4	Aruba	ABW	Land area where elevation is below 5 meters (% of total lan	AG.LND.EL5M.ZS	29.57481		
5	Aruba	ABW	Forest area (sq. km)	AG.LND.FRST.K2	4	4	4
6	Aruba	ABW	Forest area (% of land area)	AG.LND.FRST.ZS	2.222222	2.222222	2.222222
7	Aruba	ABW	Cereal production (metric tons)	AG.PRD.CREL.MT			
8	Aruba	ABW	Access to electricity (% of population)	EG.ELC.ACCS.ZS	84.99329		
9	Aruba	ABW	Electricity production from oil, gas and coal sources (% of to	EG.ELC.FOSL.ZS			
10	Aruba	ABW	Electricity production from renewable sources (kWh)	EG.ELC.RNEW.KH			
11	Aruba	ABW	Electricity production from renewable sources, excluding h	EG.ELC.RNWX.KH			

The data source user interface

Before loading data, it's important to know what the different parts of the data connection user interface are. The Tableau Public 9.x user interface that is used for the connection to data is shown in the following screenshot. Don't forget that the Start button remains in the upper-left corner of the UI. You can click on it from either the data source window or a worksheet to get back to the Start menu:



The parts of the user interface and their descriptions are as follows. The parts have been given in alphabetical references:

References	Description
A	This is the data source name, which will be modified in subsequent exercises
B	These are the data source filters, which can be used to limit the data that you load
C	This is the workspace, where you can add and join tables
D	These are individual tables

E	This is the Data Interpreter, which is available for Microsoft Excel files; we will learn how to turn it on and use it in subsequent exercises
F	Edit data source display by showing/hiding fields
G	This is the data
H	This is a link to sheets; you can click on this to go back to your worksheets
I	This is the Data Source button, which can be clicked on from any worksheet to get back to the data source
J	These are the tables within the data source, which can be dragged to the workspace to join to other workspace
K	This is the pivot or view grid of the data, which will be used in subsequent exercises
L	This is the data source, which can be changed by clicking on the orange link and then browsing to a new file

To load this file into Tableau Public, we will start with a new Tableau Public workbook. You can download the Tableau Public workbook that we used for this chapter by visiting https://public.tableau.com/profile/tableau.data.stories#!/. The following steps will guide you through how to connect a file to Tableau Public:

1. Open a new instance of Tableau Public.
2. From the **Connect** pane, click on the data file type to which you'd like to connect. In this case, we are using an excel file.
3. Browse to the file to which you would like to connect.
4. Drag a table from the list of tables, which is a list of different worksheets in this case, along with the workbook onto the workspace.
5. Note that the values in the data source are now populating the space below the workspace, but at least with this data set, there is no complete set of field headers. We will edit the data source by using the data interpreter in the next exercise.

The name of the data source is showing a concatenation of the name of the workbook and the table name. Click on it (in the previous screenshot, it's **A**) to give the data source a good name. Remember that anything that you publish on Tableau Public is available for other people to download, and since they aren't able to see the actual origin of your data source, it's a good idea to give it an explicit name so that there are no errors of attribution.

Using the data interpreter

Tableau Public 9.x has a new feature that is designed to reduce the amount of transformation that you need to do to your data sources. The data interpreter automatically detects where the first row of headers or data is in a Microsoft Excel data file, and if there are empty or semi-structured rows before the data, it can remove them. (The data interpreter does not work with text files.)

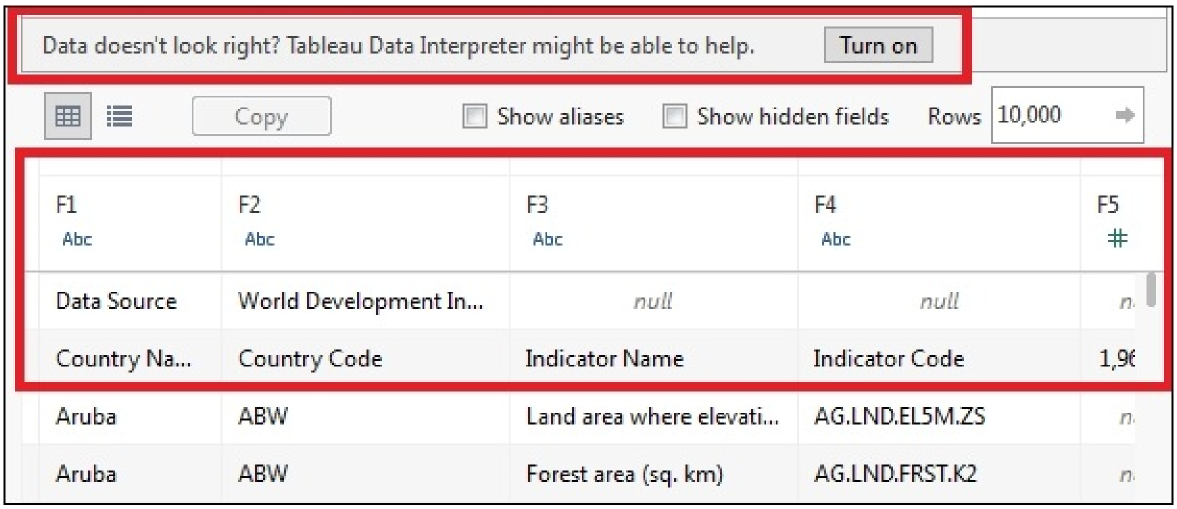
The following data source has the following major errors in it:

- There are two rows of mostly blank, non-data values before the first row of the valid data; we will use the data interpreter to fix this
- The years in which the measurements were taken should be going down a column rather than across the columns; we will pivot the data

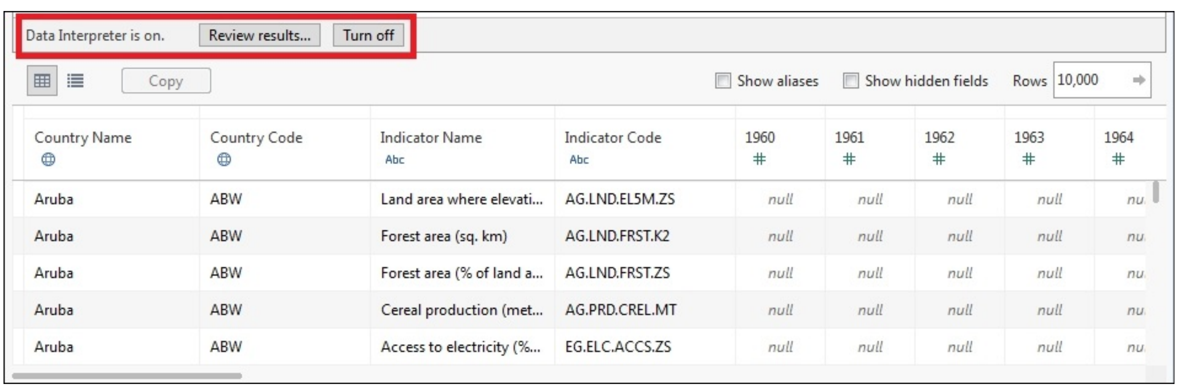
in the next exercise to fix this

We will use the data interpreter to fix the first problem. In the following screenshot, you will see that Tableau Public has suggested that we use the data interpreter. The steps are as follows:

1. Under the workspace, note that Tableau Public has recognized that the data might not be formatted properly and has suggested using the data interpreter. Click on the button that says **Turn on**:



2. Now that the data interpreter is on, you can review the results in the following screenshot and see how the data was transformed:



The two rows of garbage are now gone, and the field headers are populating properly. We still need to resolve the issue of the date dimension going across the columns rather than down a column. We can resolve this by pivoting the data.

Pivoting data

Pivoting data is a capability designed to help you resolve issues within data sources, like in the previous example, where the date dimension is not formatted properly.

By highlighting the headers, you can pivot them from columns into rows by performing the following steps:

1. Highlight the field headers that you need to pivot. In this case, we click on **1960** and scroll all the way to the right, holding down the Shift key as we select columns.
2. Right-click on a selected header and choose **Pivot**.
3. The pivoted fields now have transformed into two new columns—the headers that you selected appear as values in a new column called `pivot field names`, and the measures now appear in a field called `pivot field values`.
4. Right-click on the headers for each of these fields and rename them. We renamed `Pivot field names` to `Year` and `Pivot field values` to `Measure`.
5. Check out the following modified data source. It is now formatted properly, but there is one issue—the numerous rows with null values. We will edit those in the next exercise:

Country Name Environmental Fa...	Country Code Environmental Fa...	Indicator Name Abc Environmental Fact Table	Indicator Code Abc Environmental Fac...	Year Abc Pivot	Measure # Pivot
Aruba	ABW	Land area where elevation is below 5 met...	AG.LND.EL5M.ZS	1960	*null*
Aruba	ABW	Forest area (sq. km)	AG.LND.FRST.K2	1960	*null*
Aruba	ABW	Forest area (% of land area)	AG.LND.FRST.ZS	1960	*null*
Aruba	ABW	Cereal production (metric tons)	AG.PRD.CREL.MT	1960	*null*

Filtering data sources

It's reasonable to expect that you won't need to load all the data in the data source. It is important to load only what you need because the more the data in the data source, the slower it will be. In the current example that we are using, there are many rows with null values. The reason that they have null values is that for the selected measure, no measurement was taken for certain time periods.

Tip

The null values are different from measurements of zero. Values of zero mean that a measurement was taken and the value was zero. Null means that no measurement was taken.

We have no reason to load rows with null values. Therefore, we can filter them as follows:

1. In the upper-right corner under **Filters**, click on **Add**.
2. Click on **Add** again.
3. Select the field that you wish to filter. (We filtered on **Measure**).
4. Since **Measure** is a measure and not a dimension, we see a continuous spectrum of values. But we want to include everything *except* null values.
5. Click on the **Special** button on the upper-right side.
6. Click on **Non-null values**.
7. Click on **OK**.
8. Click on **OK** again.

Note

The data source shows values of zero, but not null. The data source is almost complete. The only item that is remaining before we can start using filter is joining it with another table in the same data source.

Joining tables

In this exercise, we will join the **fact** table with a dimension of the countries so that we can group the countries by region. A join is a logic statement in which you tell Tableau's data engine how two tables are related to each other. There are two parts to it, the join types (the left join, inner join, right join, or outer join) and the join conditions.

- **The left join**: This keeps all the records from the left (or first) table and the corresponding records from the right (or second) table.
- **The inner join**: This keeps only the records from both the tables that match the join condition.
- **The right join**: It is the opposite of the left join; it keeps all the records from the right table and only the corresponding records from the left table. Outer joins keep all the fields from all the tables.

> ## Tip
>
> The availability of join types depends on your data source. For this data source, we can create an inner join or a left join.

A join condition is where you tell Tableau Public's data engine, which is functioning as a database management system in this case, how the two tables are related. In order to join tables, you need to have at least one field whose contents occur in both tables. In the following example, we will join our tables by the country name so that we can see the corresponding region for each country.

> ## Tip
>
> Tableau Public will automatically join your tables on the first fields that occur alphabetically in both the tables and have the exact same field name and field type.

In order to add new tables, you need to drag them from the list of tables on the left into the workspace next to the existing tables:

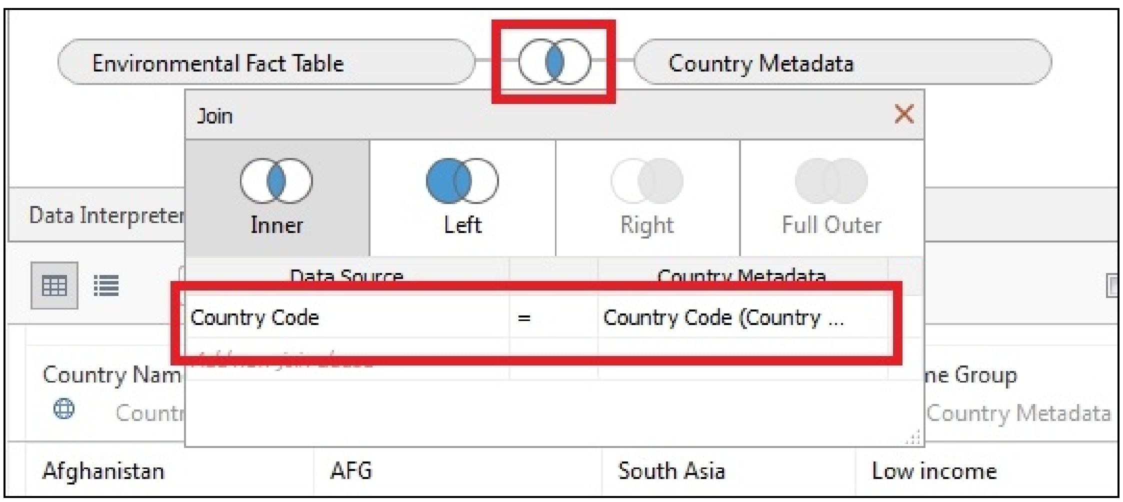
1. Drag the **Country Metadata** table into the workspace and drop it

next to the **Environmental Fact Table**.

2. Tableau Public automatically detects the field that occurs first in both data sources alphabetically and has the following properties:

    - The exact same field name, including capitalization and punctuation
    - The same field type

3. You can view and edit the **Join** details by clicking on the Venn diagram icon between the tables, as shown in the following screenshot. In this case, our data sources are joined by **Country Code**, which is correct. In order to select different fields, click on the name of the joined field and replace it with someone more appropriate:

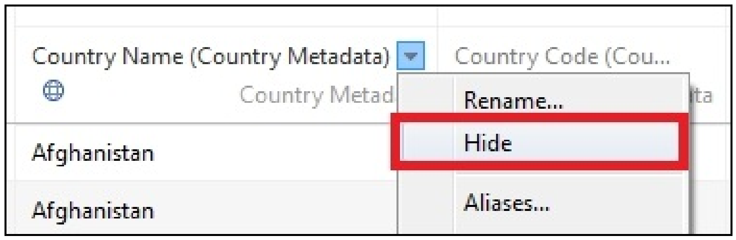


4. When you are satisfied that the join condition is correct, click on the Venn diagram icon again.

> ## Note
>
> The additional fields, with their source table name appended in parentheses, are included in the data set below the workspace.

5. Just because a field is included in a join condition, it does not mean

that you need to load it in the workbook. You also do not need duplicates of existing fields. For that reason, click on the **Country Name (Country Metadata)** and **Country Code (Country Metadata)** fields, which already occur in the **fact** table, and from their context menus, select **Hide**, as shown in the following screenshot:

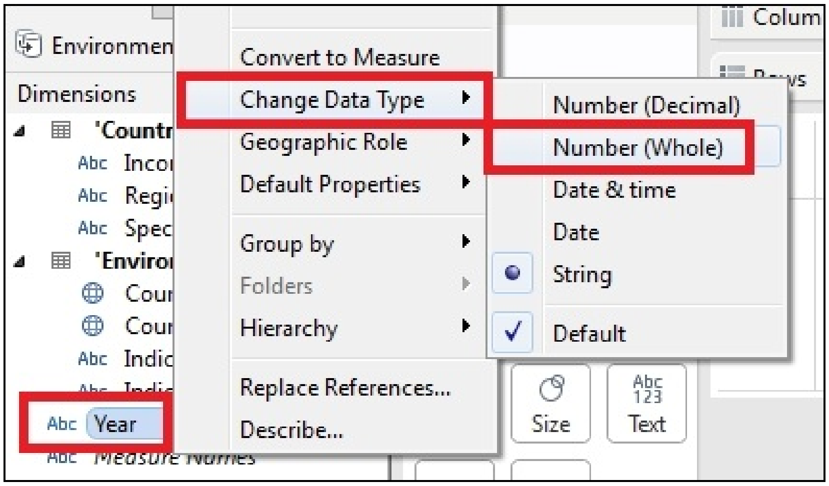


6. If you would like to see the fields that you have hidden, click on the checkbox next to the **Show Hidden Fields** text above the data source.

The data source is now ready to be used in a visualization. To create a visualization, click on a sheet number or name in the ribbon at the bottom. If you'd like to get back to the data source, you can click on the data source icon from any worksheet.

When you load a new data source, the following are some of the several items that you should check before you can use it:

- Confirm that all the dimensions, are in fact in the **Dimensions** pane, rather than in the **Measures** pane
- Confirm that the data source types of all the fields are correct. For instance, in this data source, `Year` is formatted as a string, but it really should be a number
- We can change the data type by right-clicking on the field, selecting **Change Data Type**, and choosing **Number (Whole)**



Connecting to web-based data sources

The steps required to connect to OData are different from the steps required to connect to the previously mentioned sources because they involve web servers and network security. These steps are a subset of the steps in Desktop Professional that are used to connect to a server:

1. Enter the URL of the website.
2. Select the authentication method.
3. Establish the connection.
4. Name the data source.

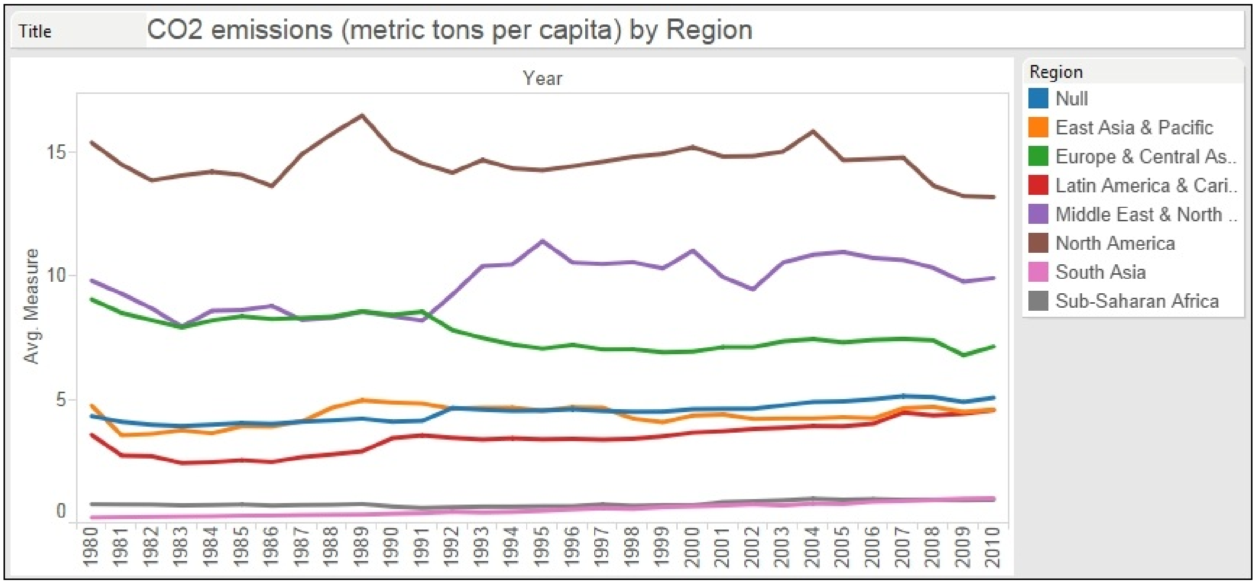
Another big difference between using local sources and online sources is that while the local sources can be refreshed with just a right-click on the data source name, online sources must connect to the website, which needs to be refreshed.

In order to refresh a web-based data source, perform the following steps:

1. Right-click on the data source name in the data pane.
2. Click on **Edit Connection**.
3. In the previous dialog box, which will be populated with the connection parameters, click on the **Connect** button in step 3 of the preceding list. It isn't necessary to repopulate the connection parameters or create a new connection to refresh the data.

Check out the visualization in the following screenshot, where we used

Check out the visualization in the following screenshot, where we used this data source to graph the average CO2 emissions per capita by region since 1980:



Summary

In this chapter, you learned one of the most critical concepts of using Tableau Public and producing accurate, informative data visualizations—connecting to, transforming, and loading data. The precepts that we discussed, such as making sure that your data source is clean and formatted properly, are relevant to working with data.

In the next chapter, we will discuss the different types of visualizations that you can make with your data and how to use each type correctly and in the most compelling manner.

Chapter 4. Visualization – Tips and Types

A journalist or blogger who uses Tableau Public must convert data to visualizations so that readers can quickly understand large quantities of data that have been distilled down to a single graphic. The role of an author is data abstraction, the conversion of real-life data into visual cues (primitives, as some people who work in the field of data science call them), such as colors, shapes, lengths, and positions. The role of your reader is to do the reverse, consume data primitives and convert them back into real-life things, events, and phenomenon. If you have done a poor job choosing the elements of your data visualization, your readers will have a difficult time relating your visualization to the real world, which means that the story that you're telling isn't as compelling as it should be. Your visualization will not have the value that you wanted, because the connection between the data visualization and the underlying data has been broken.

Before you create a visualization for people to utilize, there are several considerations that go into the design. The following are some of these considerations:

- The purpose of the visualization: Which questions should it answer?
- Content: What should be included? (Start small, as less is often more).
- The visualization structure: What do you actually want to measure or show?
- The readers' platform: How have your readers approached your questions before? Which tools have they used to do so?

In this chapter, we will discuss the following topics:

- The lifecycle of visualization
- Iterative design and what it means
- Creating visualizations and visual perceptions
- Types of visualization

- Using discrete and continuous dates and measures
- Filtering, grouping, and sorting

An overview of the development lifecycle

Perform the following steps to develop a high-quality, compelling visualization:

- The first step is to devise a purpose for the visualization—a question that you would like the data to answer—and envision the possible visualization types that can best answer that question.
- Next, define the content and find your relevant data; determine how much data should be shown and the level of details required. A large part of your visualization project should be spent on finding and cleansing data. Do not skimp on this step, because it will make your visualization more clear and useful to readers. You may have to scrape data from the Internet or combine it from multiple sources.
- Wherever you get your data, make sure that you have both validated and attributed it properly.

> ### Tip
>
> We surveyed a few experts in Tableau Public, and anecdotally, they estimate that they spend 70% to 80% of their total development time building data sources.

- The third step (with respect to this book) is to try various view chart types using the **Show Me** feature of Tableau Public. Try various view types to answer the question, experiment with various views, add or subtract data, manipulate components, and explore the question in greater detail or find other questions.
- The next step in building a visualization is considering the platform—

print, desktop, laptop, tablet, and smartphone. The medium that your consumers use to interact with your visualizations and develop insights is critical.

- The last step is publishing your work, or delivering content to your visualization consumers. Your data visualization should start a conversation. So expect feedback.

Ten visualization tips

Data visualization is an art and a science. Much has been researched, studied, and published on the topic of what a great visualization is composed of. In this section, let's attempt to clarify some of these findings via some brief explanations, as shown in following list. As with everything in life, your mileage may vary, and there are always exceptions to rules. However, consider these general rules as condensed, combined, and paraphrased here as compared to the industry standards and academic research as you develop visualizations:

- Readers should be able to understand data visualization quickly. Make it easy for your consumer to read your data visualization. Don't use font sizes that are less than 10 pt if you can manage it, and use common fonts that are designed for ease of use. The key is that visualizations are not too flashy, complex, or artistic just for the sake of it. Do not use 3D in any way, it distracts and can mislead readers by distorting data. Follow the form over function principle, and the amount of time required to understand the data will work itself out.
- Consider the format for the reader. Print, computer screens, and mobile devices use large fonts for titles, and consider the format of consumers' devices. Scrolling is not good in most cases, and you should manage white space carefully.
- Every element should have a reason for being in a virtualization. Choose colors carefully. Use colors for categories instead of quantities; don't use colors randomly. Shapes, colors, legends, and labels should add to the understanding of the visualization.
- Be consistent in the placement of objects, elements, and colors. Your work is your digital brand, it literally is your product. Your consumers may not know you in person. Their perception of you and your skills is determined by how they relate to your visualizations. Do not use variety for variety's sake. The organization of elements and colors is important, as it helps readers navigate through the visualization.
- Use the correct context for data and consistent axes for elements on a chart. Help readers understand the data and what they are looking

at. Do not mislead readers by using mismatched axes for charts that compare the same measure, nonzero axes, and other misleading tactics.

- Simplicity comes from clarity, and this helps readers understand data. Keep it simple! Some data sets are inherently complex. In such cases, try to show a part of the data set, or break up the visualizations to show the various aspects of the data. Be clear in your organization and approach, and simplicity (for readers) will follow.
- Consider calculating differences for readers instead of plotting multiple lines or bars in an effort to show the comparisons and differences. For example, a line chart of the city budget versus the city expenditures may be interesting, but an even better option is to use a bar chart, with a baseline set to the budget and bars above or under the baseline. Alternatively, you can set reference lines in a chart by using mean or target values.
- Choose the best type of view (chart) for the data. Read the following sections and understand the key elements of every major chart type and what the main objectives of each chart type are.
- Understand the different visualization primitives (color, length, area, shapes, position, direction, and angle) and how humans perceive them in terms of data visualization.
- Experiment with the Tableau Public **Show Me** feature to choose a view type and make changes of your own, such as selecting the shelf items that you need to include, and their characteristics. Experiment and refine, and think in terms of a **visualization cycle** of exploration rather than a linear route from data to visualization creation.

The perception of visual clues

Our brains interpret visual signals in specific, predictable ways, and modern user interfaces and consumer products are based on extensive study in this area. The way we interact with visual signals was first formally studied and discussed by *William Cleveland* and *Robert McGill* (at the time, with AT&T Laboratories) in their seminal paper named *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods*, as published in the *Journal of the American Statistical Association* in September 1984. This paper is readily available via a quick Internet search, and it is worth a quick read.

In addition to *Cleveland* and *McGill*, who were statistical scientists (and later professors) focused on researching how we process visual cues, *Stephen Few* has written extensively on this topic in his books and on his blog, which can be viewed by visiting http://www.perceptualedge.com/.

The authorities on this matter have concluded that the human brain correctly perceives numerical values in a way that is different from how it perceives various visual primitives.

In the descending order of accuracy, visual primitives are as follows:

- **Position**: This shows how marks relate to an axis, like in a **scatter plot**.
- **The bar or line length**: This is why line graphs and bar graphs remain the first and the most popular forms of data visualization.
- **Angle**: This should be used sparingly, as the human brain actually is not good at differentiating angles within a circle. **Pie charts** are effective if you wish to communicate measures as long as about five dimensions are included and the slides are labeled with their respective percentages of the whole.
- **Area**: It's actually fairly difficult for us to differentiate numerical values from each other based on the size of a shape. Therefore, you should use bubble graphs and tree maps sparingly and label them properly when you do use them.
- **Saturation and hue**: These are the worst primitives that you can

use to encode numerical values because not only are a significant number of readers color-blind, but also they are used inconsistently and designed for perception rather than judging.

- **Color**: This is best used for categorization into groups rather than to measure quantity. However, you can use it effectively in **heat maps and highlight tables,** as long as you are using it properly and labeling the values.

Your mileage may vary, but consider the hierarchy of visual clue perception carefully. Some conservative data authors will try to stick to scatter plots, line charts, and bar charts (and their derivative chart types) almost exclusively and avoid most other charts such as pie charts, bubble charts, and heat maps.

Leveraging your understanding of how our brains interpret visual clues will help you select the proper chart type. Fortunately, there is a built-in assistant in Tableau Public called **Show Me** that will help you choose a chart type based on the data elements (measures and dimensions) that you select.

Using the Show Me tool to create charts

Tableau Public has a great tool embedded into the software called **Show Me** (or the **Show Me** button). This tool is an expression of Tableau Software's vision of self-service analysis because it allows largely nontechnical users to develop useful graphs and charts based on their data. The **Show Me** button is at the top right-hand side of the Tableau Public interface. Clicking on this button opens the **Show Me** dialog box (with 24 chart type icons in Tableau Public 9.x), as seen in the following screenshot:

The **Show Me** tool suggests graph types based on the view that the user has created and/or the fields that the user has selected from the **Data** window or the shelf areas.

Note

Only the graph icons that fit the data types available or selected are

Only the graph icons that fit the data types available or selected are enabled. For example, none of the geographic maps on icon row number 2 are enabled, because there is no geographic data in the data set used to make this screenshot.

The **Show Me** tool also tells you which additional field types you need to select in order to create a specific visualization, which is a great way to learn.

To use the **Show Me** tool, select one or more dimensions or measures from the **Data** window and then click on the **Show Me** button in the upper right of the Tableau window to see the suggestions in the **Show Me** dialog box. The following are some examples:

- Selecting a continuous date field will prompt the suggestion of a line graph
- Selecting a geographical field will prompt the selection of a map

Tip

You can't auto-hide the **Show Me** dialog box once it is displayed and the user has made a selection from it. Click on the **Show Me** button again to hide the **Show Me** dialog box.

In addition to creating new visualizations using the **Show Me** card, you can modify the existing graph. When you select a new graph type from one of the highlighted buttons, Tableau Public will change your visualization for you, and by doing so, it might remove fields on the visualization that aren't part of the new one.

Show Me contains 24 chart types in Tableau Public 9.x, but many more advanced or customized charts can also be created by adding different data elements and tweaking the existing chart that you have built. Items that help users better understand the data include tool tips, trend lines, color indicators, row separators, and reference lines. We will explore these additions in detail later in the chapter.

Answering questions using Show Me chart types

Of the many chart types available in the **Show Me** tool in Tableau Public 9.x, there are several common chart types that are frequently used—tables, line charts, bar charts, geographic maps, and scatter charts. Pie charts are often used in popular culture, particularly in infographics, but you should use them sparingly and always label the slices with the percent of the total.

Charts and graphs exist to answer questions, and some charts can naturally answer certain types of questions better than other charts.

The following sections discuss in detail some important chart types offered in Tableau Public, but this is not a comprehensive list due to limitations pertaining to space in this book. In order to help you learn how to construct different visualizations, the screenshots in these sections include the placement of different fields on the shelves and cards of the workspace.

About dimensions and measures

In Chapter 3, *Connecting to Data*, we discussed dimensions and measures in detail. Let's bring up a few more points so that we understand the chart creation process a little better. Tableau Public separates data source into dimensions (qualitative fields) and measures (quantitative fields).

When you drag a field onto a worksheet, its new instance is called a **pill**. The dimensions and discrete fields will be blue, while continuous measures will be green pills. Each of these pills has a right-click on the shortcut context menu and can have icons on them that correspond to the Marks card shelves that they belong, to as well as the set and group icons (if they are part of a group or set).

When we use a dimension, Tableau Public creates column or row headers for a chart (view). On the contrary, measures typically create an axis in the chart in case the measure is classified as continuous, as described in the next section of this chapter. Measures can be aggregated (mathematical calculations such as summation, averaging, counting, and so on) based on the selected aggregation function (`SUM`, `AVG`, `COUNT`, and so on) for each item in the dimension used in the chart.

Tableau Public automatically determines whether a data field is a dimension or a measure, but sometimes, the software makes mistakes. For instance, if your data source has numeric unique identifiers, such as account numbers, then they will be grouped as measures by default, unless the field name contains the `ID` string. In this case, you can change a data element to either a dimension or measure by dragging it to the correct pane in the **Data** window (to the **Dimensions** pane or the **Measures** pane). It is more likely that you will have to convert measures into dimensions (because all numeric values are not measures). To convert a measure into a dimension, you can either drag it to the **Dimensions** pane or right-click on the measure and select **Convert to Dimension** from the shortcut context menu. If this conversion is needed for only one chart, you can locally convert the measure on a shelf in the same manner (by selecting the **Convert to Dimension** command from the shortcut context menu).

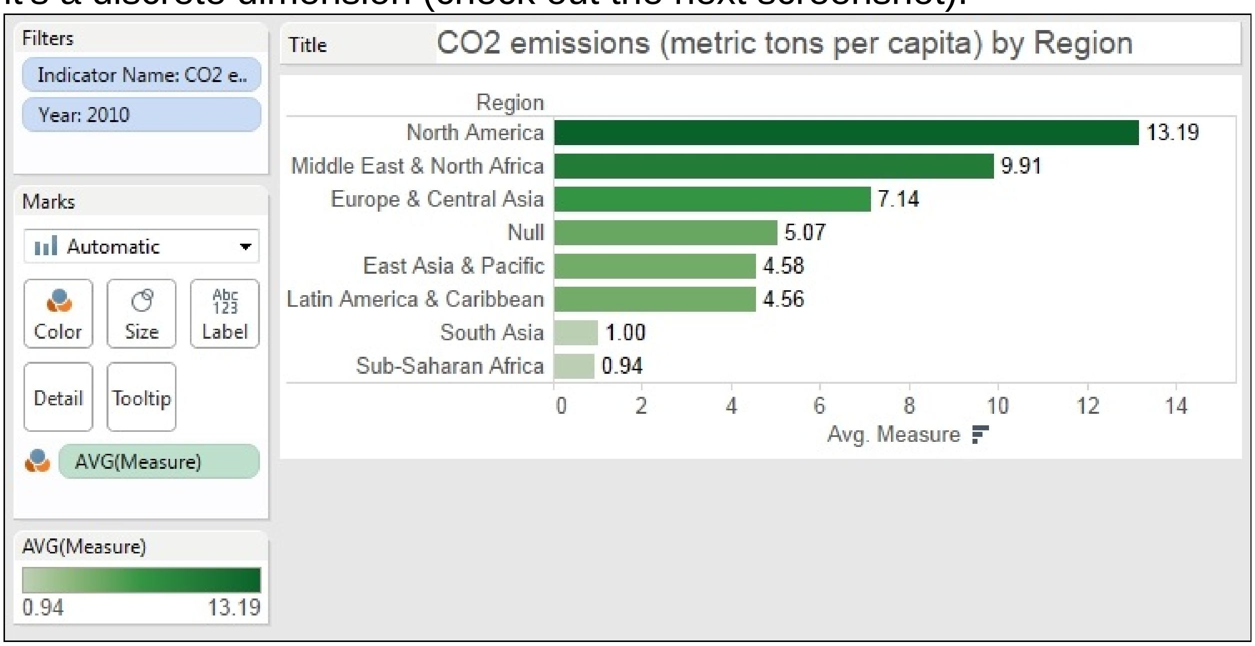
Continuous and discrete dimensions and measures

As learned in the previous section, all data elements are classified in Tableau Public as either a dimension or a measure. These dimensions and measures are further classified into continuous or discrete elements.

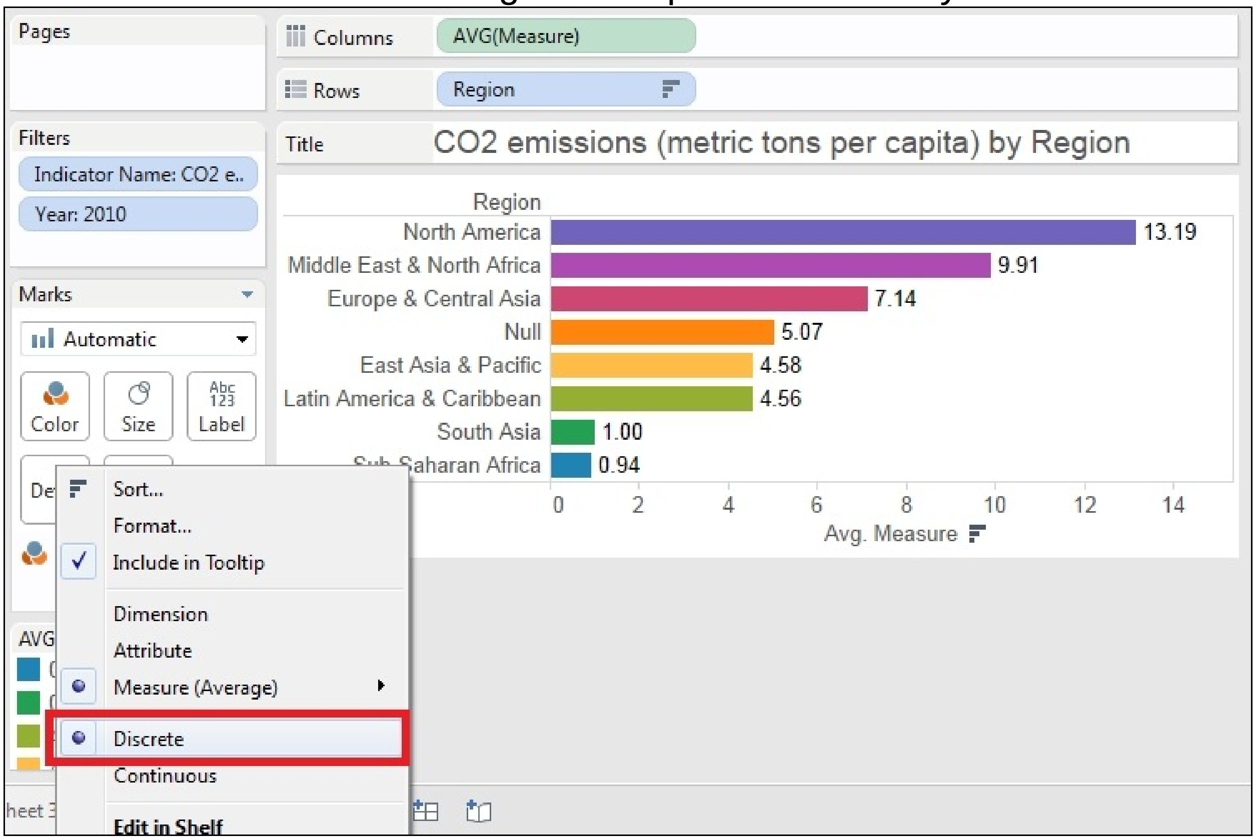
The data elements in the **Data** window and on the shelves in Tableau Public are either light green or light blue in color. The green color indicates that a data element is set to continuous, and the blue data elements are set to discrete elements. A green or blue outline will also appear when selecting the corresponding continuous or discrete data elements. When you are building charts in Tableau Public and adding dimensions and measures to the **Columns** and **Rows** shelves, the continuous data items (usually measures) will create axes, and the discrete elements will create row or column headers.

Discrete and continuous designations for data elements affect how graph elements are built and marks and axes are rendered. For example, a continuous date dimension used in a line chart will create a traditional, unbroken line chart. Using a discrete date dimension will create a line chart with segmented panes, that is, lines that are broken apart with each change in the date part. For instance, for a graph that shows years and months, the lines will break between each year. Using continuous measures on the **Color** shelf will result in a color gradient being used in the chart, with the hue or color in proportion corresponding to the value of the measure. If a discrete measurement is used on the **Color** shelf, the color will not be a gradient but separate colors (this is generally fine for dimensions on the **Color** shelf, but not for measures). Have a look at the following screenshot to see the difference between using a discrete color and a continuous color.

In this example, we're using a continuous measure to color each cell. Note that on the **Marks** card, **AVG(Measure)** is green. This means that it's continuous. **Region**, as compared to **AVG(Measure)**, is blue because it's a discrete dimension (check out the next screenshot):



We converted **AVG(Measure)**, which is the average CO2 emissions per capita in 2010, into a discrete number by clicking on its context menu on the **Marks** card and then selecting **Discrete**, as shown in following screenshot. This version assigns a unique color to every numerical value:



Similarly, if a continuous measure is used on the **Size** shelf, the sizes of the graph marks such as circles, squares, or other shapes, will be in proportion to the value of the measure. Discrete measures used in the **Size** shelf can have a more pronounced size variation, but you can experiment with changing the continuous or discrete settings for measures and dimensions to better understand their effect on the chart. In my opinion, using discrete measures on the **Size** shelf can sometimes yield good results.

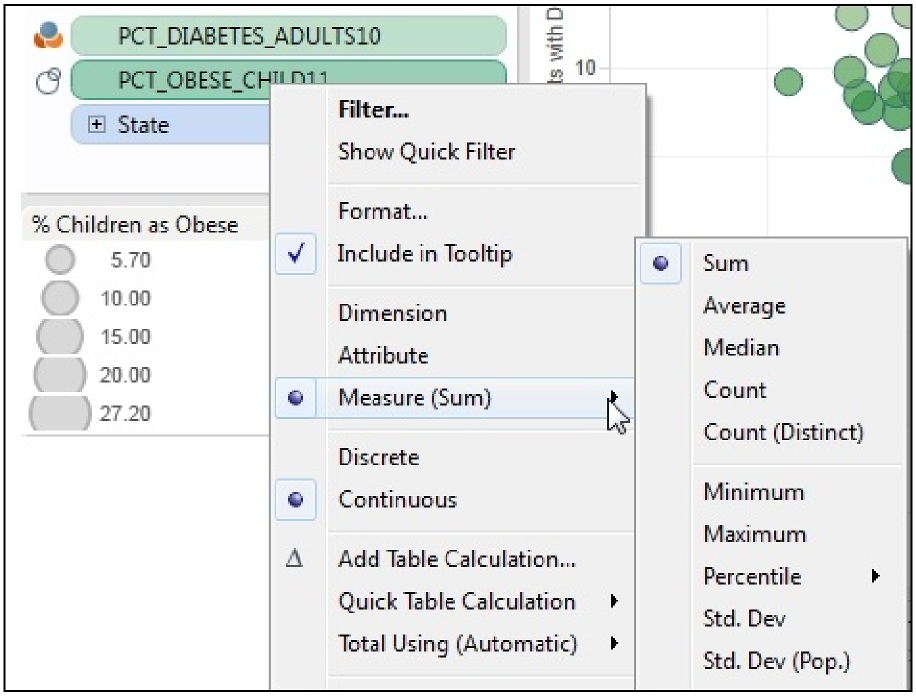
Measures and dimensions can be converted from continuous elements to discrete and vice versa. This allows you to experiment and customize the chart view. Quite often, you will convert continuous elements to discrete elements, as you will also often convert measures to dimensions. You can convert continuous elements to discrete ones by right-clicking on the date element and selecting the **Convert to Discrete** menu option. Also, when you move a data element from the **Measures** pane in the **Data** window to the **Dimensions** pane, it will typically convert the date element from continuous to discrete.

Just like changing data items from measures to dimensions only in one chart, you can also change from continuous to discrete for only the desired chart instead of making the change globally for all the charts and worksheets in the workbook.

Selecting aggregation types for measures

Measures can be aggregated. These numeric values can be added, counted, and averaged, and the median can be chosen. In addition, the aggregation types that are available in Tableau Public also include statistical functions such as maximum and minimum selection, variance, standard deviation, and percentile. The most common aggregation types used in journalism and blogging are sum (addition), average, and count. The count aggregation is useful when you have a data source with rows that have an ID field, such as the `FIPS ID`, which is a numerical identifier for a specific geographic area, such as a county. To get a count of records, use `Count(ID)`, where the ID is the specific ID name in your data source. Extending this example, you can also use **Count(Distinct)** when you only want a count of every unique ID and don't want to count the repeating IDs.

Selecting the aggregation type for measures can be done by right-clicking on a measure from the context menu, choosing the **Measure** command, and clicking on the desired aggregation type. You can also access aggregations from the drop-down menu by clicking on the tiny arrow on the right-hand side of the data element pill. The following is a screenshot of the full aggregation menu that is available from a measure pill's right-click on the shortcut context menu:



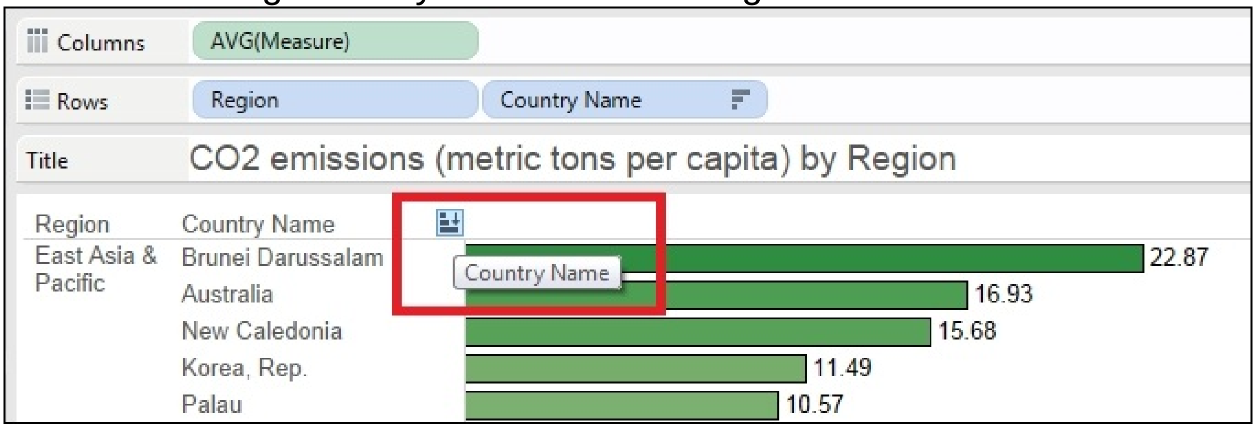
Swapping and sorting

Columns and rows can be swapped (one for another) by using the **Swap** button in the menu bar. Items on shelves can also be moved manually among shelves and to and from the **Data** window. Dimensions and measures can be duplicated by pressing *Ctrl* key and dragging to a new or the same shelf. This is helpful when changing aggregations for a duplicate date element or part of the date (such as changing year to month).

Similarly, you can sort fields in the following three ways:

- Right-click on the field in the **Data** window, click on **Default Properties**, and select **Sort**, which sets the default sort order but does not allow you to sort by a measure
- Click on the context menu of a pill on your visualization and select **Sort**
- Click on the ascending or descending sort buttons in the toolbar, or hover over a header and click on the appropriate icon

For instance, in the following screenshot, the countries in each region are sorted in the descending order by the CO2 emissions per capita. If you click on the quick sort icon that we have highlighted, the countries in each region will be sorted again in an alphabetical order. Conversely, when you hover over the *x* axis and click on the sort button, they will be sorted in the ascending order by the measure being used:



Types of visualization

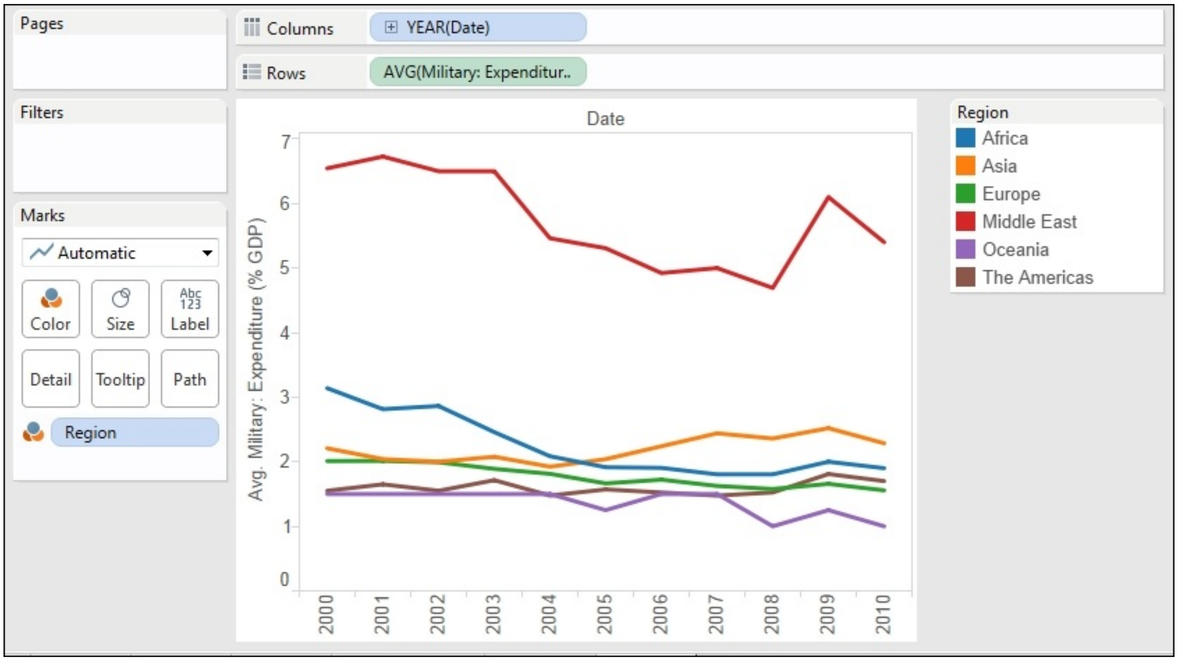
There are many different types of visualization that you can create in Tableau Public, and we will focus on several of the most effective ones in detail. There are some that you should not use without practice, careful construction, and deliberate labels, such as bubble graphs, tree maps, and word clouds. We will not cover those here, since they have very limited uses.

Line graphs

Line graphs show data trends over time. Line charts, along with bar charts, are the most popular chart types used in data visualizations. For this type of time series chart, place a time or date dimension on the **Columns** shelf and a measure on the **Rows** shelf (the horizontal line chart is the most commonly used/useful line chart).

Line charts can then be modified and made more complex by adding a dimension on the **Color** shelf, which adds one or more additional lines to the chart. You can add different measures to the **Color** shelf as well and measures or dimensions on the **Size** shelf. But usually, adding too much complexity to the graph detracts from the understanding of the data. The following steps will help you create a line graph:

1. Create a line graph that shows the average military expenditures as a percentage of the GDP between 2000 and 2010 by loading the World Bank Indicators data.
2. Then, drag **Date** to the **Columns** shelf from the **Dimensions** pane.
3. Drag **Region** from the **Dimensions** pane to the **Color** shelf.
4. Drag **Military Expenditures (% GDP)** from the **Measures** pane to the **Rows** shelf.
5. Click on the context menu on the pill for **Military Expenditures (% GDP)** and select average the measure, as shown in the following screenshot:



From this line graph, we learned that the Middle East has had the most fluctuation in its average military spending over the past 15 years.

Continuous versus discrete date-time elements

This section on line charts is a good place to discuss continuous versus discrete date-time elements. Line charts can be set up either with a continuous time series, or a discrete time series. Most line charts used for visualizations are continuous and unbroken, as the one in this section is. However, time can also be set as discrete, and this will break apart the time series into sections (also called panes or panels), such as by year, quarter, and month. Dashboards are sometimes developed with a discrete time series to allow for a fast comparison of quarters or months. A time-data element can be set as discrete or continuous by right-clicking on the date-time data element and setting it to continuous or discrete. Doing so will radically change the line chart by switching to either a continuous line or a broken line based on the parts of the date selected. Most journalists and bloggers who use Tableau Public will want to use the continuous feature of date-time data to create unbroken line graphs, but experiment with continuous and discrete to check whether it adds value for your readers.

Tables

Tables are usually created to show a fine level of detail for your data. In most business organizations, people have traditionally consumed analytics created in tabular reports, which can be highly inefficient. Tables display counts or measures relative to categorical variables, such as department spending, the number of college graduates in a city, pollution levels in a stream, and so on. They are useful when you wish to look up individual data point values and compare them across one or more levels of dimensional detail. Tables are the most effective when they are used at a high level and they contain data summaries rather than very long and detailed tables. You should also add filters to table charts to narrow down the data displayed to give readers a focus point. Tables are often called crosstabs. Pivot tables are a specific type of table; they are not discussed here.

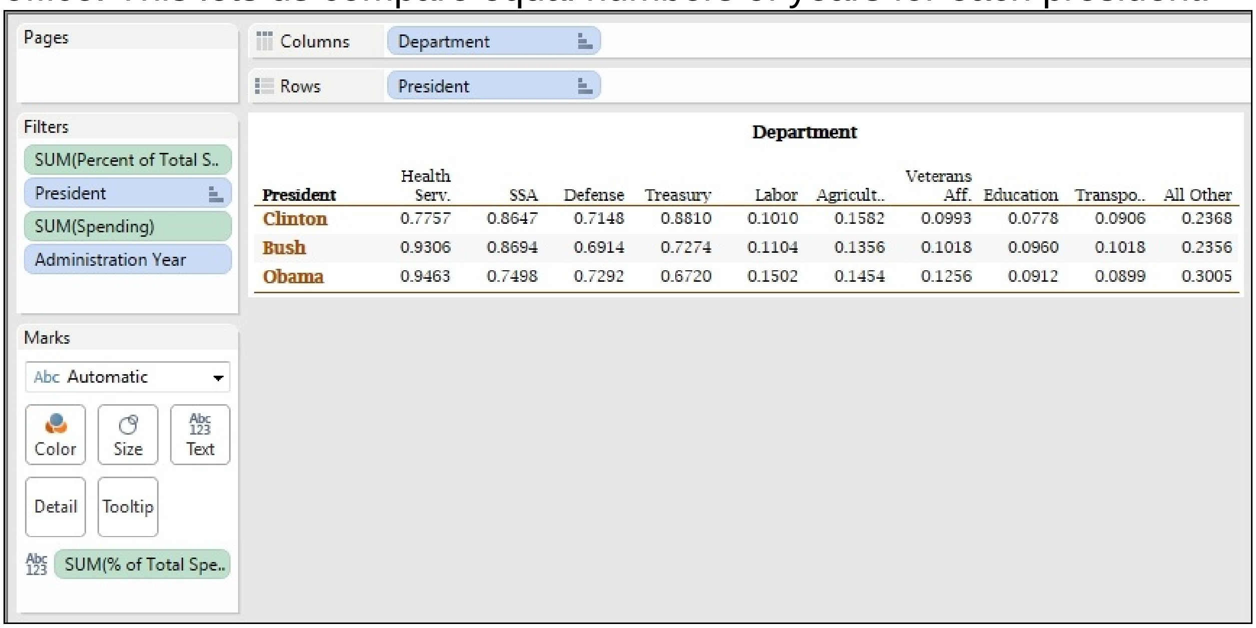
There are three table views in Tableau—**text tables** (sometimes called crosstabs), **highlight tables**, and **heat maps**. Any of these three chart types can be selected and modified easily from one chart to another, or added to and varied as needed. The text table is a good place to start experimenting when building a table.

> ## Tip
>
> The tables on the US government's spending were obtained from Tableau Software. Download the Tableau Public workbook from Tableau by visiting http://tinyurl.com/1962-2012spending and download the source data from Tableau Software by visiting http://tinyurl.com/spending-source-data.

Text tables are common tables, as seen in excel, for instance. This type of chart is available for creation from the top row of the **Show Me** tool. The following screenshot is an example of a text table. This is a simple table that shows the US government's spending as a percentage of the total spending per **Department**, for the US **President**'s administration.

For the table, we set the **SUM(Percent of Total Spending)** to the **Text** shelf, and this exposes the numeric values in the table rows and columns.

Note that the details of the spending (the numeric values) are easily seen, and there is no highlighting or context around the values (in terms of telling the reader whether the numbers are high or low). No shape or color marks are used. Also note that we set up various filters on the **Filter** shelf for this table view, including filters for only the top spending departments, and also limited to the last three presidential administrations (**Clinton**, **Bush**, and **Obama**). Because the data is only current through 2012 (and Obama started his first presidential term in 2009), we also set up a filter for **Administration Year** and filtered to show only years 1 to 4 corresponding to each president's 4-year term of office. This lets us compare equal numbers of years for each president:

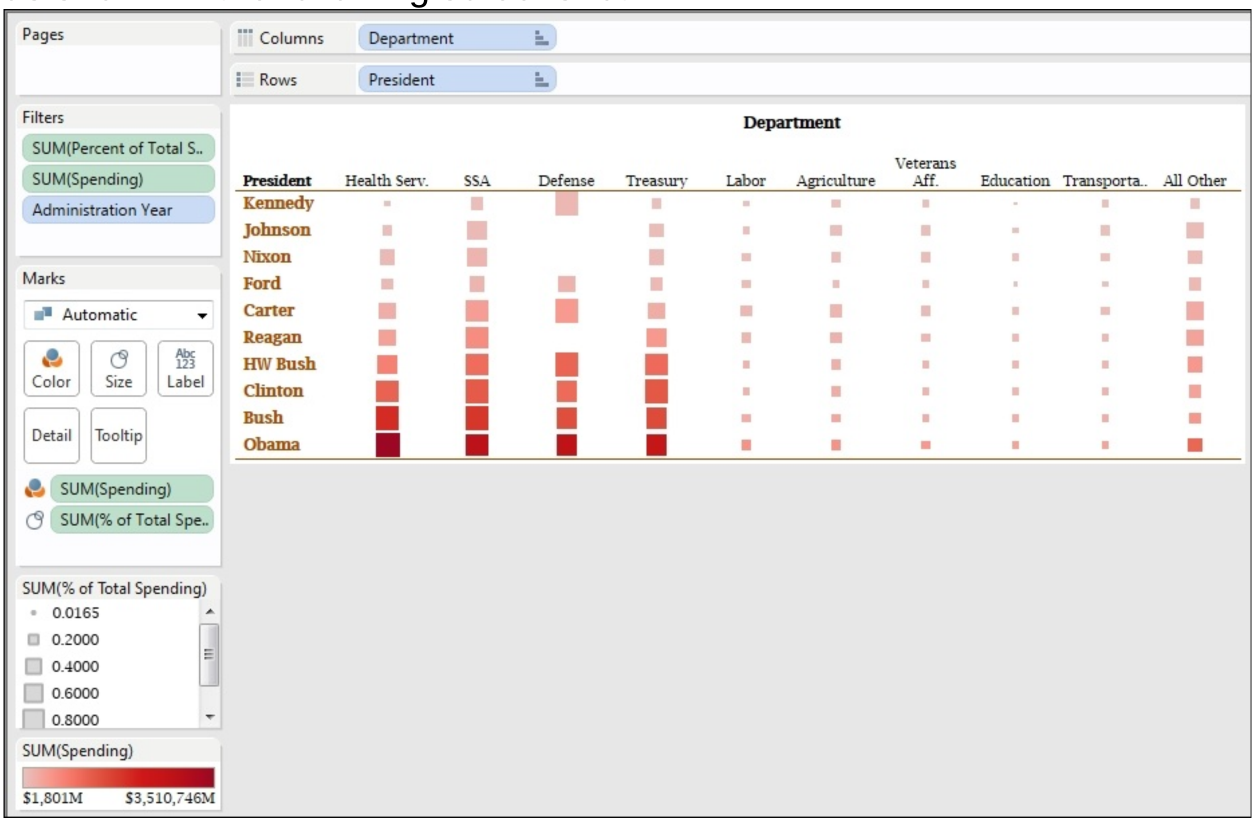


Department

President	Health Serv.	SSA	Defense	Treasury	Labor	Agricult..	Veterans Aff.	Education	Transpo..	All Other
Clinton	0.7757	0.8647	0.7148	0.8810	0.1010	0.1582	0.0993	0.0778	0.0906	0.2368
Bush	0.9306	0.8694	0.6914	0.7274	0.1104	0.1356	0.1018	0.0960	0.1018	0.2356
Obama	0.9463	0.7498	0.7292	0.6720	0.1502	0.1454	0.1256	0.0912	0.0899	0.3005

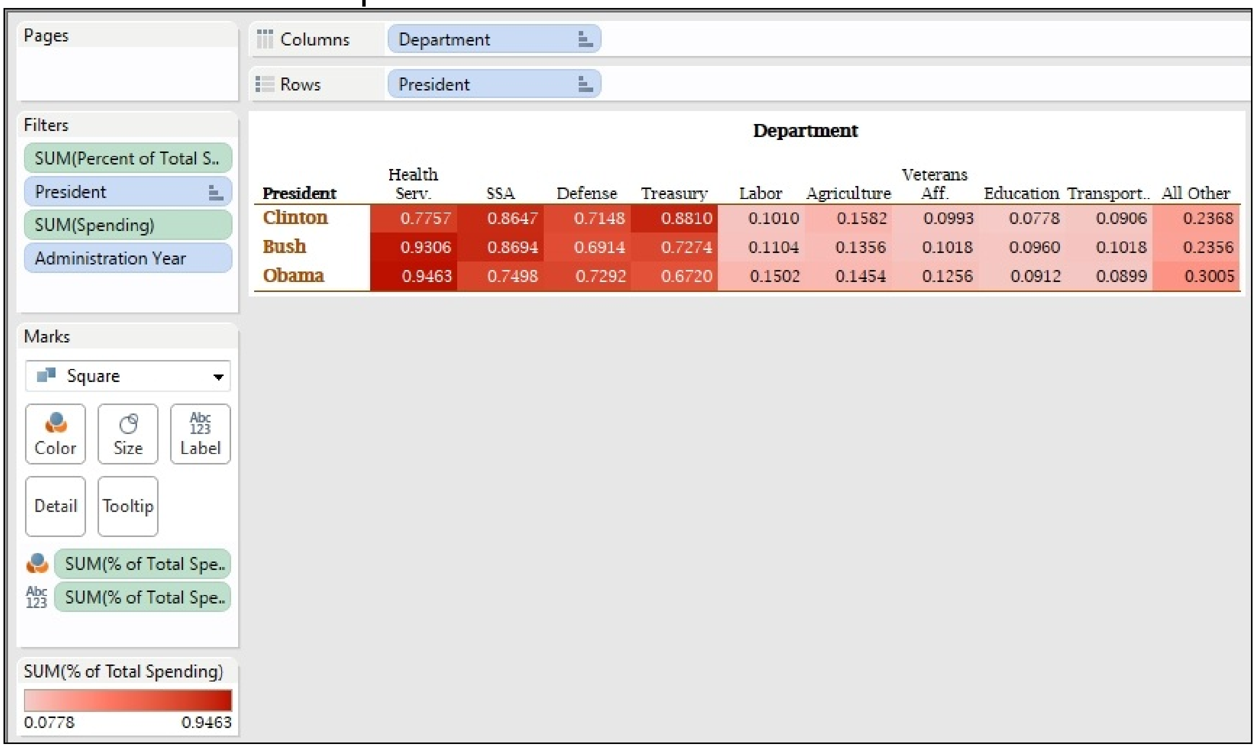
The heat maps are tables that are used to communicate visual cues such as shape and area by referring to up to two quantity measures. The larger the shape (and usually, the deeper the color), the higher the measure value. This type of chart is available for creation from the top row of the **Show Me** tool. The following graphic is an example of a heat map, with the **SUM(Spending**) on the **Color** shelf (in this case, the darker the red color gradient, the more dollars were spent) and the **SUM(% of Total Spending)** is tied to the **Size** shelf (this controls the size of the square). In this case, the larger the square, the more the percentage of the total spending for each government department during successive US presidential administrations. The actual values of spending are obscured here so that the reader is not able to quickly compare the actual values.

The colors used can be easily customized for each graph by double-clicking on the gradient color bar titled **SUM(Spending)**. The shape can be customized by selecting another shape in the dropdown list under the **Marks** title section. Filters that were similar to the **Text** table were set up. But this time, no filtering for presidents was created. All the presidents listed in the source data set are represented, from Kennedy to Obama, as shown in the following screenshot:



Highlight tables are often thought of as heat maps to the industry, as they use color gradient depth, saturation, or hue to highlight a quantity measure (such as highs and lows). The difference between a highlight table and a heat map is the lack of shape and size marks and the ability to see the underlying details in the highlight table. The following screenshot is an example of a highlight table. Note that the darker the color hue, the greater the percentage of spending to the total. The **SUM(% of Total Spending)** is duplicated in the **Marks** card for both the **Color** shelf and the **Text** shelf. This allows the color hue to be affected in proportion to the numeric value and also lets the value itself be shown in the rows and columns.

The colors used, and the other borders and elements, are completely customizable in the tables described in this section. The colors of this chart can be easily customized for each graph by double-clicking on the gradient color bar titled (in this case) **SUM(% of Total Spending).** The filters that have been created are identical to the first table named the **Text** table in this chapter section:



President	Health Serv.	SSA	Defense	Treasury	Labor	Agriculture	Veterans Aff.	Education	Transport..	All Other
Clinton	0.7757	0.8647	0.7148	0.8810	0.1010	0.1582	0.0993	0.0778	0.0906	0.2368
Bush	0.9306	0.8694	0.6914	0.7274	0.1104	0.1356	0.1018	0.0960	0.1018	0.2356
Obama	0.9463	0.7498	0.7292	0.6720	0.1502	0.1454	0.1256	0.0912	0.0899	0.3005

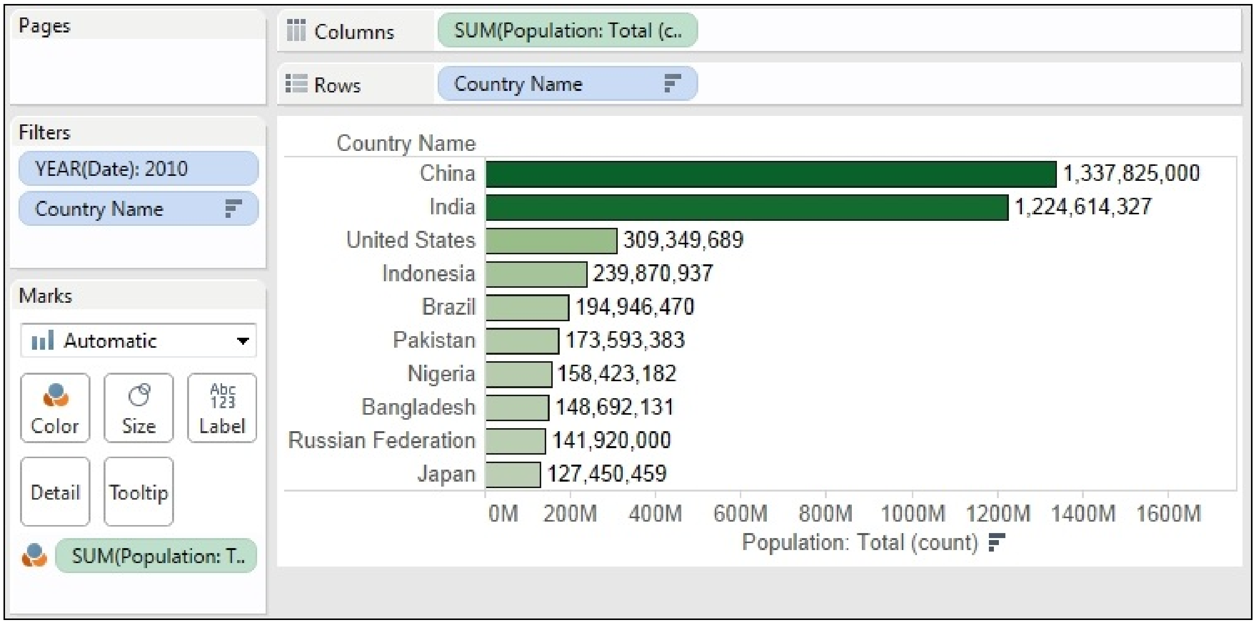
Bar charts

Bar charts are used to compare values and are perhaps the most useful of the chart types. Bar charts allow you to compare values across categorical dimensions (such as age groups, sex, race, cities, states, expense categories, departments, and so on; the list is endless), but they do not display the underlying data details as can be done in a table. Bar charts are used to compare values across dimensions. As we discussed in an earlier section in this chapter, length can be easily and naturally interpreted by readers. Therefore, bar charts are one of the more commonly used and appreciated charts used today.

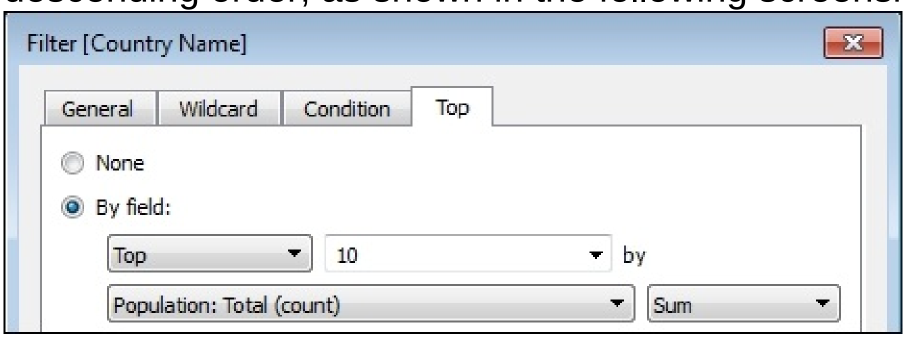
> ## Tip
>
> The global population data used in the following section was downloaded from the World Bank by visiting http://data.worldbank.org/data-catalog/Population-ranking-table.

There are several types of bar charts in Tableau—**horizontal bars**, **stacked bars** (they compare parts to a whole and are the same as pie charts), **side-by-side bars** (they compare two categories), Gantt charts, **histograms** (they measure the frequency of events and plot distribution diagrams), and **bullet graphs** (a relatively new compact chart invented by *Stephen Few*, a visualization expert; you should become familiar with his work). In this section, we will only talk about a couple of these bar charts, but I invite you to experiment on your own and learn about these other chart types if you find them appealing.

The following screenshot shows a horizontal bar chart that depicts the top 10 countries by population. The chart reveals that **China** and **India** are the two largest countries by population (with over 1 billion citizens each), followed by the USA, Indonesia, and other countries with a much lower population:

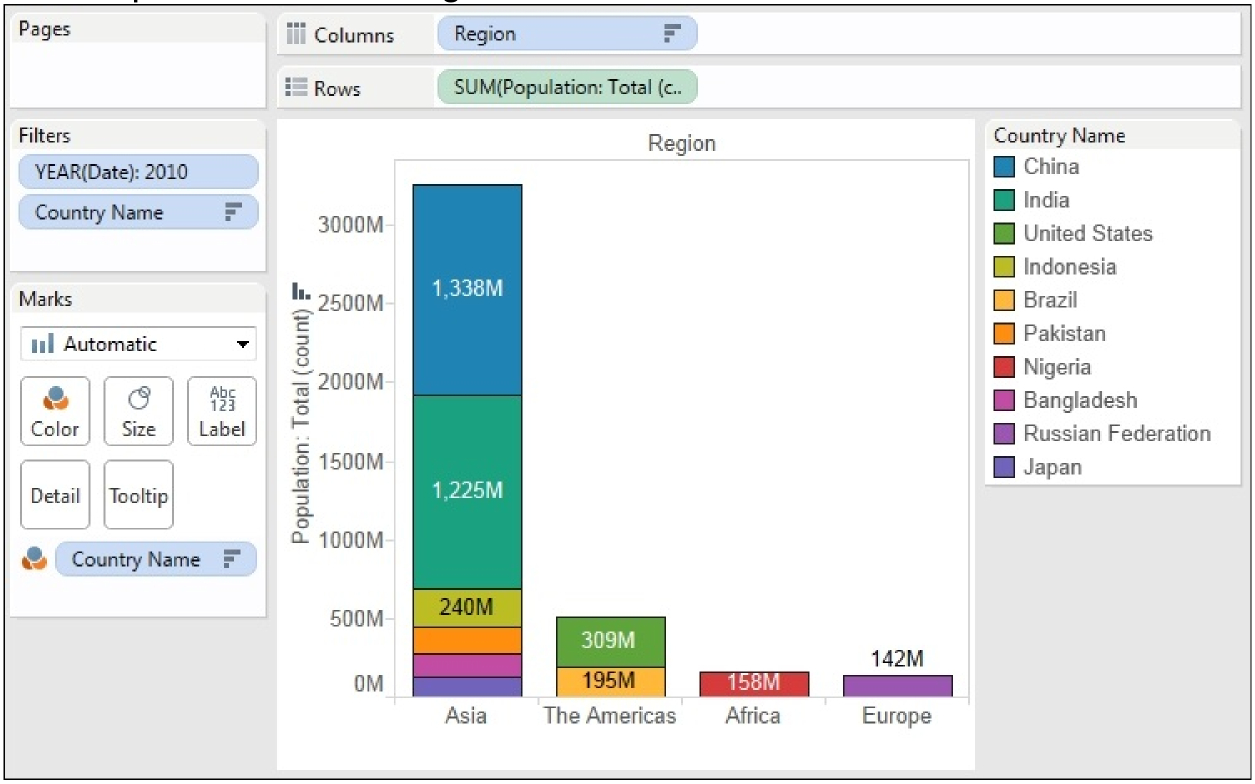


Some key design elements that should be pointed out in the previous horizontal bar chart are we used **Country** as the dimension and **SUM(Population)** as the measure. To limit the graph to the top 10 countries in terms of population, we added the **Country** field to the **Filters** shelf and chose to filter the top 10 countries by population in the descending order, as shown in the following screenshot:



After filtering the data to get information about the top 10 countries in the year 2010, we sorted the bars by clicking on the quick-sort button on the *x* axis to sort the countries in descending order by total population.

The chart shown in the following screenshot is a **stacked bar chart**. In this chart, we have pivoted several fields in the previous two exercises to show, by region, the 10 most populous countries in the world. The stacked bar chart is a **part-to-whole comparison** chart that often works better than a pie chart because it relies on the bar's lengths rather than an interpretation of the angles:



Stacked bar charts are best used when there are just a few different dimensions, because too many dimensions and corresponding colors are difficult to interpret. We created this stacked bar chart by performing the following steps:

1. Drag the **Region** field from the **Dimensions** pane to the **Columns** shelf.
2. Drag **Population: Total(count)** from the **Measures** pane to the **Rows** shelf.
3. Maintain the filters on **Year** for 2010 and the top 10 countries according to the total population.
4. Drag **Country** from the **Dimensions** pane to the **Color** shelf.
5. Click on the **Color** shelf on the **Marks** card and add a black border, which separates the bars more definitively.
6. Clicked on the **ABC** icon in the toolbar to see the mark labels.

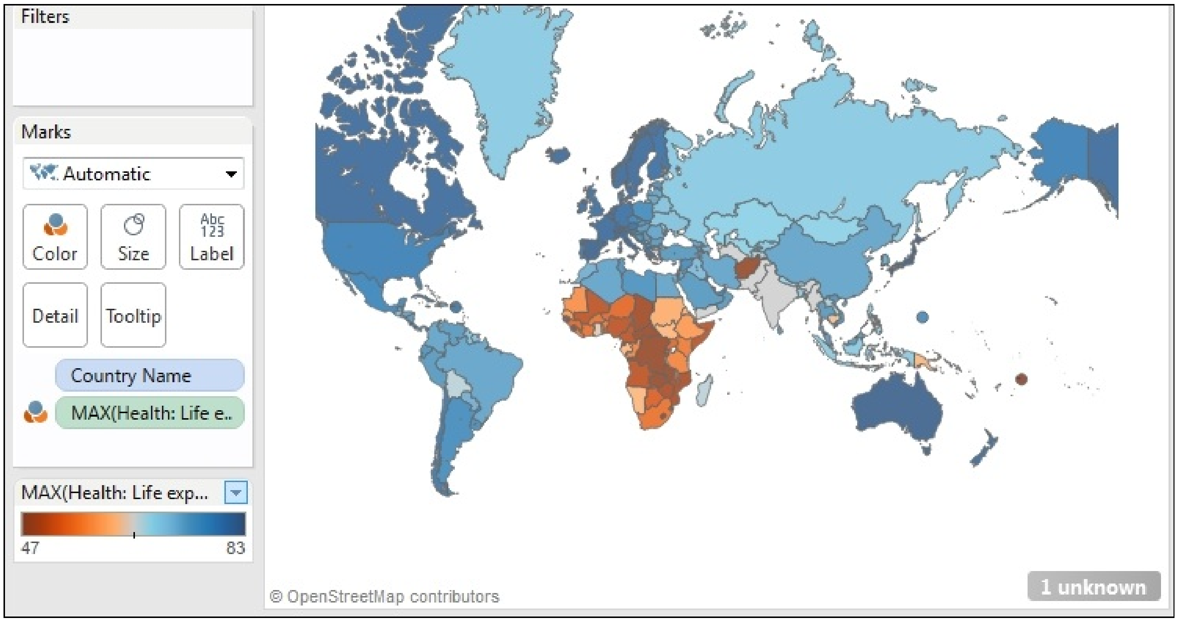
Geographic maps

Geographic maps are a useful chart type that can answer various questions related to locations. Before you select this chart type, consider the fact that having geographic data elements such as longitude, latitude, state, county, city names, or addresses does not mean that you have to use them in every case. Sometimes, a bar chart or another chart type can express a data story as well as a geographic map in, case map data does not actually add value. For example, sometimes it is better to group states, counties, or cities into regions and treat those as categories by using bar charts.

Tableau Public has a powerful **geocoding** function that is built-in and requires no additional programming (though many third-party providers exist for more robust maps with *street* and *topographic* map support). Geocoding is the process of recognizing geographic data, such as addresses, cities, counties, state country names, latitude, longitude, and so on, with the help of software and automatically encoding text strings and numeric values as geographic dimensions.

We built the following map to show the maximum life expectancy by country in the World Bank Indicators data set by performing the following steps:

1. Multi-select **Country** in the **Dimensions** pane and **Life Expectancy at Birth (total)** in the **Measures** pane.
2. Click on the **Show Me** card.
3. Select the filled map.
4. On the **Marks** card, click on the context menu for **Life Expectancy at Birth (Total)** and change it to a maximum, which means that it shows the maximum value for each country within the data set.
5. Click on the context menu for the **Color Legend**, click on **Edit**, and select a color palette that's friendly to people who are color-blind, as shown in following screenshot:



From this map, we learned that the countries of central Africa have relatively worst life expectancies in the world.

When the data is loaded in Tableau Public, the software automatically encodes geographic fields with their average latitude and longitude. When you select one of these newly encoded fields, a small globe icon will be associated with the data element. You can click on one of the map icons on the **Show Me** dialog box, or you can double-click on any of those fields to add it to the visualization pane. The filled map icon is the most commonly used one for data visualizations.

Scatter plots

Scatter plots are a type of chart that show relationships between two measures to establish correlations and comparisons, find trends in data, and expose outliers. Scatter plots are one of the most effective forms of communicating mathematical relationships, and they are one of the seven basic tools of quality control.

You can build a scatter plot by adding a measure to the **Rows** shelf and a measure to the **Columns** shelf, or double-clicking on each of the measures in succession—the first measure will be placed on the **Rows** shelf and the second one on the **Columns** shelf. You can swap the measures (rows swapped with columns) by pressing the *Ctrl* key and clicking on the measures and select the **Swap** menu command icon. Alternatively, you can right-click and click on **Swap** on the shortcut context menu. Then, add one or more dimensions to the **Marks** card, dragging various dimensions to the various shelves, such as **Color**, **Size**, or **Shape**. This will put the scattered measures in context with the dimensions and allow for an analysis of the relationships.

Aggregation of the measures plays an important role here. The default behavior of Tableau Public is to aggregate measures (for example. by using sums or averages). This may be fine for most analyses, but you may want a clearer picture (or you may want to expose outliers) by looking at all the data points. You can override this aggregation behavior by disaggregating the data, which will display all the values in the data source for that measure. To do so, go to the **Start** button and select **Analysis**. Then, click on **Aggregate Measures** and deselect that menu command. Note that disaggregation may remove the information displayed from the tool tip when hovering over a data point. You can add the contextual values displayed in the tool tip by dragging dimensions and measures from the **Data** window to the **Detail** or **Tooltip** shelves in the **Marks** card.

If too many values are displayed, you can add the measures to the **Filters** shelf and set parameters in the filter dialog box to limit the number of data points (marks) that are being shown.

We created the scatter plot shown in the following screenshot, which shows the relationship of GDP and mobile phone users between 2000 and 2010 in Asia by country from the World Bank Indicators data, by performing the following steps:

1. Drag **Finance: GDP (current. USD)** to the **Columns** shelf, which puts it on the x axis and makes it the independent variable.
2. Drag **Business: Mobile Phone Subscribers** from the **Measures** pane to the **Columns** shelf, which makes it the dependent variable.
3. Drag **Date** from the **Dimensions** pane to the **Color** shelf.
4. Click on the context menu for the **Color** legend and assign the **Cyclic** color palette.

5. Drag **Region** to the **Filters** shelf and select Asia.
6. Drag **Country** to the **Label** shelf.
7. Drag **Population: Urban (count)** to the **Size** shelf on the **Marks** card, which further encodes the default circles with the relative urban populations.

From this, we learned that the relationship between the **GDP** and **Mobile phone** subscribers in **China** is linear and that in other countries such as **Japan**, it is relatively consistent. It is also not surprising that the urban population of **China** is the largest, since it has the largest population. In the next chapter, you will learn how to create calculated fields that show you the percentage of the total that is urban rather than just a discrete number.

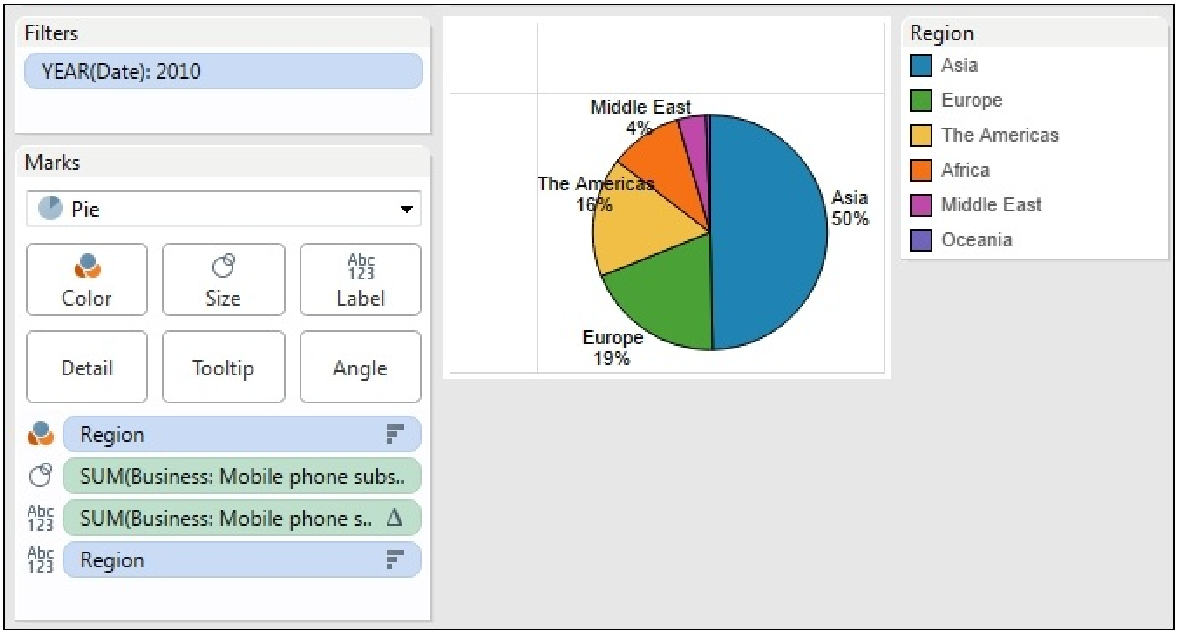
Pie charts

Pie charts allow you to show the composition of parts to a whole, like slices in a pie. Pie charts should be used sparingly because humans cannot interpret angles and area that well. Noted visualization experts, including *Stephen Few* and *Edward Tufte*, are generally against using pie charts, except in limited circumstances. It is often said for data, authors consider using a stacked bar chart instead.

That being said, pie charts can sometimes be used effectively under certain circumstances. Limit the number of slices to 5 (at the most), and make sure that the slices are not too small and are visible. You can use pie charts effectively to give a general sense of how one dimension compares to another, but don't use it to report measures that are close in value. It is too difficult for people to interpret when the measures (the pie slices) are of similar sizes. Tableau Public, unlike other data visualization software packages, does not support 3D pie charts and drill-downs into pie slices. This is a good thing because these characteristics lead to confusing and misleading pie charts.

We made the pie chart shown in the following screenshot, which still uses the World Bank Indicators data to show the relative percentage of mobile phone users in the world by region:

1. On a new worksheet, change the mark type to a pie on the **Marks** card.
2. Drag **Region** to the **Color** shelf.
3. Click on the context menu on the **Color** shelf to change the color palette to **Cyclic**.
4. Drag **Business: Mobile Phone Subscribers** to the **Size** shelf on the **Marks** card.
5. Click on the **Color** controller on the **Marks** card and add a black border.
6. Drag both **Region** and **Business: Mobile Phone Subscribers** to the **Label** shelf on the **Marks** card.
7. Click on the context menu for **Business: Mobile Phone Subscribers** on the **Label** shelf of the **Marks** card, select **Quick Table Calculation**, and choose **Percent of Total**.
8. Filter **Date** by setting it to 2010.
9. Click on the **Context** menu on the **Color** legend, select **Sort**, and sort **Regions** in the descending order by **Business: Mobile Phone Subscribers**:



This pie chart shows that half of the mobile phone users in the world are in Asia. The pie chart is easy to use, and it does not mislead users. It provides them with the context that they need to sort through the information points quickly.

Using groups and sets

It can be useful to group dimension members into groups to report on these groups, as compared to others. For example, in the Life Expectancy Map that we built earlier, it might be helpful to add Canada, USA, and Mexico into a group or set named **North American Countries**.

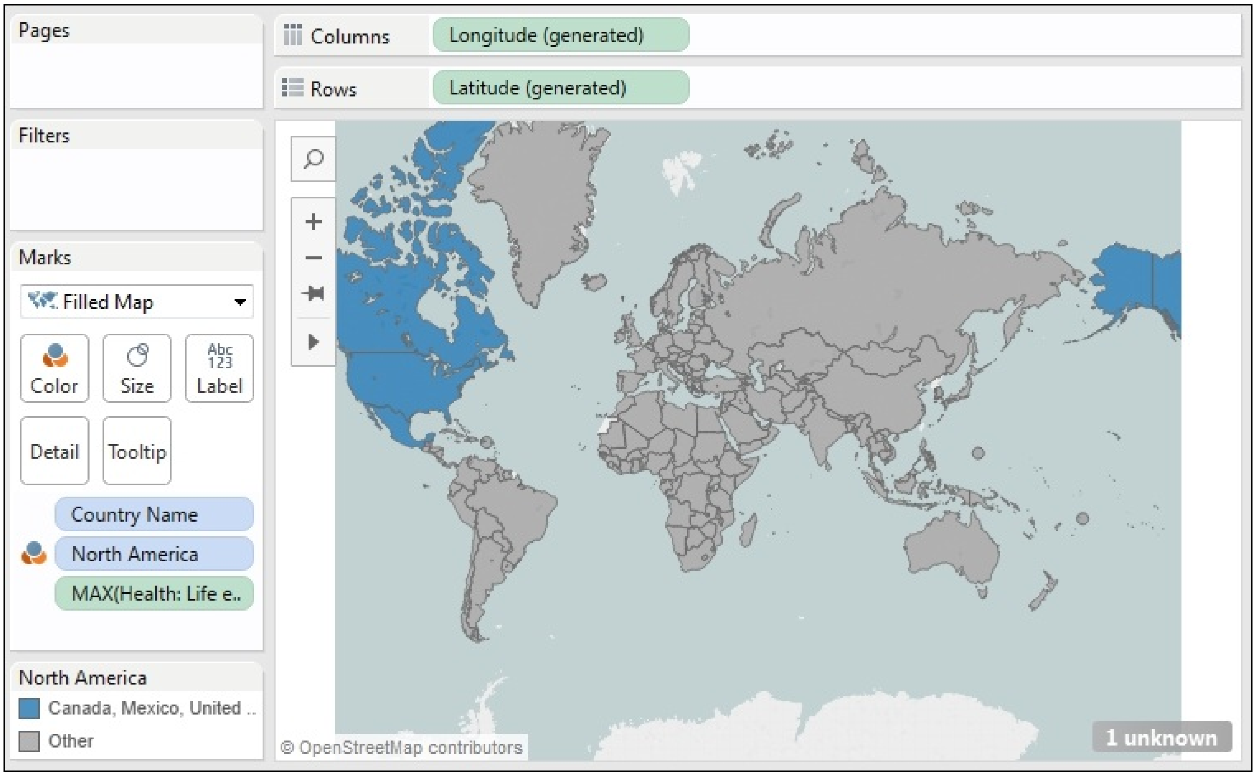
A group is a simple set that is composed of the dimension members that you choose. In the following example, we set up the North American group by performing the following steps:

1. Right-click on the **Country** field in the **Dimensions** pane.
2. Select **Create** and then choose **Group** from the context menu.
3. Give the group a name.
4. Manually select the three countries that we want to put into the group.
5. Click on the **Group** button and then name the group.

6. Click on **OK**.

Note that the group field appears at the bottom of the **Dimensions** pane. It is considered to be metadata. It exists in the Tableau Public workbook, but it does not appear in your original data source.

In the following example, we added the group to the **Color** shelf on the **Marks** card of the map. Countries are colored by their membership in the group. Either they are in, or they are out, as shown in the following screenshot:

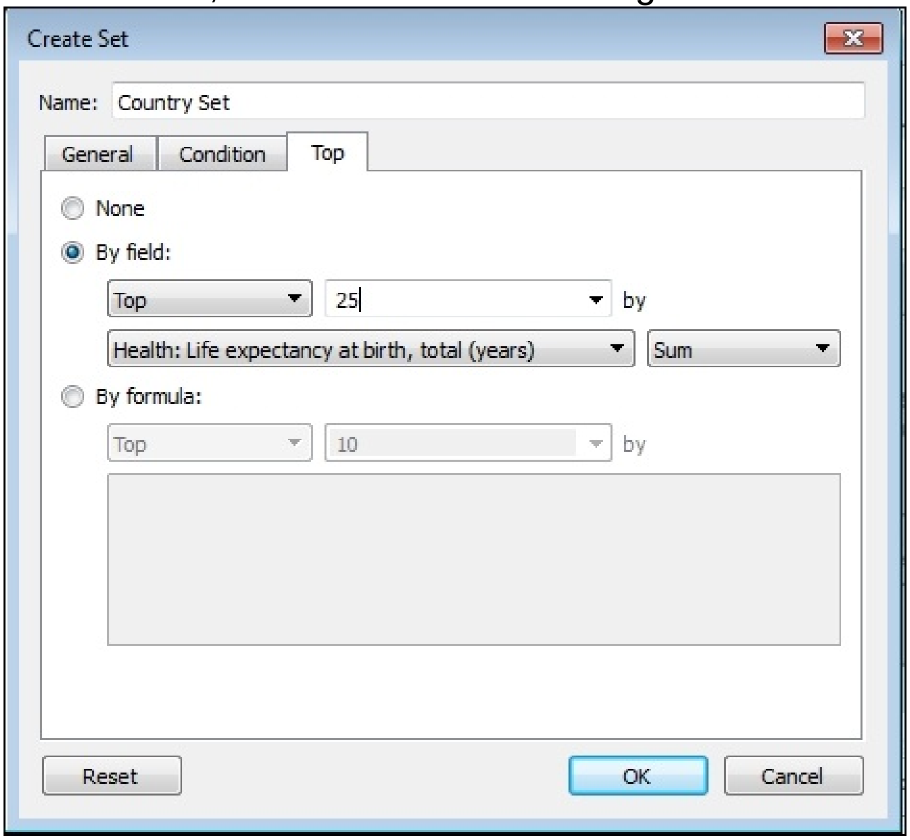


Similarly, we can create a set that groups members of a dimension by their adherence to the criteria that you establish.

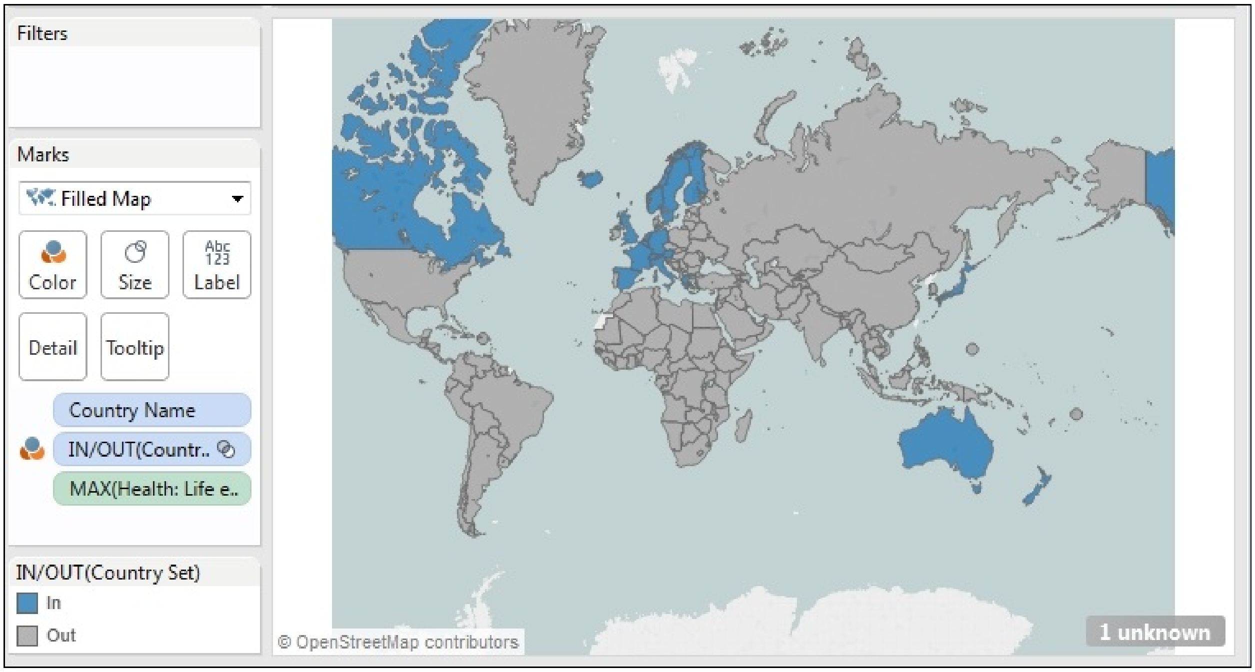
We can create a set that identifies the top 25 countries by life expectancy, which will be used again to modify the map, from the World Bank indicators data source by performing the following steps:

1. Right-click on the **Country** field in the **Dimensions** pane.
2. Click on **Create** and then select **Set**.
3. Rename the **Set** to **Country Set**.

4. Click on the **Condition** tab rather than manually selecting the members of the **Set** so that it will be updated if and when we get new data.
5. Create a condition that includes the top 25 countries by **Health: Life Expectancy at Birth (total years)**.
6. Click on **OK**, as shown in the following screenshot:



We dragged the set from the bottom of the **Dimensions** pane and dropped it on the top of the field on the **Color** shelf on the **Marks** card, thus replacing it. Countries that are in the top 25 countries filtered according to life expectancy are colored blue, and all the others are gray, as shown in the following screenshot:



Filters
Marks
Filled Map
Color
Size
Abc 123
Label
Detail
Tooltip
Country Name
IN/OUT(Countr..
MAX(Health: Life e..
IN/OUT(Country Set)
In
Out
© OpenStreetMap contributors
1 unknown

Summary

In this chapter, we discussed the processes that are the best if you want to design charts, from understanding the requirements, to data discovery, through to an iterative discovery cycle. We discussed the best practices in chart design and data visualization, reviewed the chart types that best answer which types of questions, and explored data dimensions and measures. We also discussed the different impacts that discrete and continuous data elements have on visualizations. Lastly, we talked about the use of reference and trend lines in views.

In the next chapter, we will talk about how to create calculated fields and table calculations.

Chapter 5. Calculations

The data sources that you query to tell stories do not always have all the data points that you need. If they did, someone else would have told the story already. Tableau Public's data engine allows you to create new mathematical calculations of varying complexity, from basic multiplication functions to sophisticated aggregations with specific levels of detail that add color, context, and key insights to your data stories.

In this chapter, we will discuss how to develop new insights for users by understanding and implementing different types of calculated fields. We will use several data sources that we have connected to previously, in order to help you understand the following functions:

- Creating and editing calculated fields
- Types of calculations
- Number functions
- String functions
- Date functions
- Type conversions
- Aggregate functions
- Logic functions
- Blending data sources

Throughout this chapter, we will present and discuss error messages that you are likely to encounter. We cannot emphasize enough that, especially when presenting information to others using a tool that might be new to them, substance is more important than style. Also, it's very difficult to rebuild credibility once it's been compromised.

The next chapter focuses on Level of Detail calculations and table calculations. They both are big topics that build on the capabilities that you will learn in this chapter.

In this chapter, we will use a data set of major global floods since 1985, which was found on Tableau Public's resources page. The original source is at http://floodobservatory.colorado.edu/Archives/index.html.

Creating calculated fields

A calculated field that you create exists only in your data source in Tableau Public; it does not exist in the file that you originally queried, which is the reason behind why you are creating it in the first place. In the following examples, we will be creating calculated fields in a data set of floods. The fields that we create will show up in the **Dimensions** or **Measures** pane in the Data window in Tableau Public depending on their type of field, with the equals sign (=) next to them.

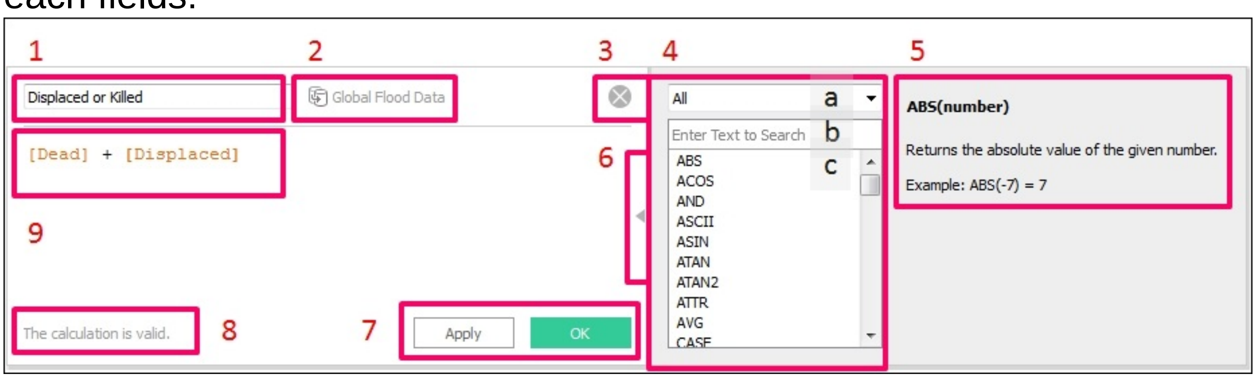
In Tableau Public, there are several ways to create a calculated field. The following steps will guide you:

1. From the **Analysis** menu, select **Create Calculated Field**.
2. Click on the **Context** menu of the data source and then select **Create Calculated Field**.
3. Right-click on a field in the **Dimensions** or **Measure** pane of a data source. Select **Create** and then choose **Calculated Field**.

Tableau Public has a new calculated field dialog box in version 9.x, which is a significant improvement over the previous versions. When it's open, you can perform functions on your visualization, and if you want to add fields to a calculation, you can drag them into the dialog box from the Data window.

You also can resize the dialog box and show or hide the list of functions.

The following screenshot shows the dialog box along with descriptions of each fields:

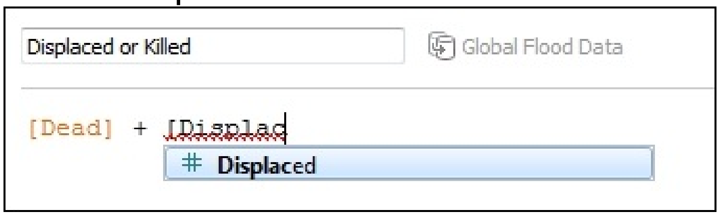


The following are the descriptions of all the functions shown in the preceding screenshot:

- **Calculated field name (1)**: This is where you give the field a good, descriptive name so that others understand what it is.
- **Data source (2)**: The data source name in which you created the calculated field.
- Click on the close icon **(3)** to close the dialog box.
- The list of functions **(4)** includes the following function types and elements:
    - List of function types, which can be used to filter the list of functions. The options, which will be explored in detail later in this chapter, are as follows:
        - Number
        - String
        - Date
        - Type Conversion
        - Logical
        - Aggregate
        - User
        - Table calculation
        - Filter box (here, you can type in the name of the function type that you would like to see)
- **Information box (5)**: Here, you can see the name of the function, what it does, and an example.
- **Button (6)**: This can be used to hide the function list.
- **The Apply and OK buttons (7)**: These buttons apply changes that you have made to the visualization and then close the window respectively.
- **Messages (8)**: This tells you when your calculation is valid, when there are errors, and what the error is.
- **The calculation itself (9)**: Tableau Public 9.x is intelligent. When you start typing in a field name, it will suggest all the fields that include the string that you are typing. You can hit the *Tab* key or click on its suggestion to populate the field. Alternatively, you can drag

fields from the **Dimensions** and **Measures** pane, of both your current data source and others data sources that you have loaded into your workbook, to create new fields.

The parts of the formula include fields, mathematical operators, functions, and often parameters as shown in the following screenshot:



Editing calculated fields

Once you have created a calculated field that is valid, you can edit it by right-clicking on it in the window in which it appears and then selecting **Edit**. You can also rename calculated fields (as well as fields native to your data source) by right-clicking on them and selecting **Rename**.

Types of calculations

There are several types of calculations that you can create, as listed previously. Some of them are common to other data management tools, but some, such as user calculations, are unique to Tableau Public.

When creating calculated fields, the functions that you can apply on a field depends on the type of that field. For instance, you can't perform multiplication functions on a string field (even if it contains only numbers) without first converting it to a numeric field. Similarly, you can't concatenate numeric values without first converting them to strings.

Basic mathematical functions, such as addition (+), subtraction (-), multiplication (*), division (/), and exponents (^), are created by typing the standard operators into the dialog box.

You can add functions by typing their names or filtering the list of functions to the appropriate group, or name, and then double-clicking on the best choice. The different types of functions that you can perform in Tableau Public are as follows:

- **Number**: This includes geometric, trigonometric, and rounding functions, among others
- **Strings**: This includes options to find characters, measure string length, split and parse fields, and match character strings
- **Date**: This includes functions for duration, the addition and subtraction of dates, the truncation of dates, and the identification of date parts
- **Type conversion**: This allows you to convert fields into different types without modifying the metadata of the source field
- **Logical**: This includes the powerful IF, CASE, and ISNULL and allows you to tell Tableau Public how to group fields or relate them to parameters
- **Aggregate calculations**: These are the most commonly used functions; they include sum, count, avg, min, max, and median
- **User functions**: These are more commonly used for Tableau Server rather than Tableau Public; they allow you to create calculated fields

that operate on usernames or groups

- **Table calculation**: This will be detailed in the next chapter; it allows you to compute aggregations based on data points that are visible in the current window

The number functions

Number functions include several functions that you may be familiar with owing to working with applications such as ROUND and ABS. Many of us may not have used several of these functions, such as the trigonometric and exponential functions, since our high school math class. There are some number functions that we will use extensively, such as MIN and MAX, while there are others that have limited uses, such as the geometric and trigonometric functions.

Number functions include the following, in order of decreasing precedence:

- **ABS**: This takes the absolute value of a number. It is commonly found in the denominator of table calculations.
- **CEILING**: This rounds up a decimal to its nearest integer and is the opposite of INT.
- **FLOOR**: This rounds a decimal down to its nearest integer.
- **MAX** and **MIN**: These take the maximum and minimum values in a sequence respectively. They are also considered to be string functions.
- **ROUND**: This allows you to specify how many decimals to round up a float number to.
- **ZN**: This is predominantly found in table calculations. It tells Tableau Public that if a numeric field has a NULL value, you want to replace it with zero.
- **DIV**: This produces the whole number product of a division statement.
- **PI**: This produces the numeric value of pi.
- **SIGN**: This produces a numeric value (1, 0, or 01) based on the input, which may be positive, zero, or negative respectively.
- **Trigonometric and geometric functions**: These are largely beyond the scope of this book. These functions include ACOS, ASIN, ATAN, ATAN2, COS, COT, DEGREES, RADIANS, SIN, and TAN as well as HEXBINX and HEXBINY.
- **Exponential and logarithmic functions**: These functions include

EXP, LN, LOG, POWER, SQRT, and SQUARE.

The date functions

The date functions are useful for a variety of tasks, such as identifying the elapsed time between two events or a part of a date.

Many date functions in Tableau Public operate on dateparts. Dateparts are the small units of measurement of dates, such as a year, quarter, month, week, day, and the units of time, such as an hour, a minute, and a second. Writing date functions in Tableau Public is similar to writing date functions in ANSI SQL or Microsoft Excel with a big difference—dateparts need to be spelled out and enclosed in single quotation marks, which is a relatively simple design and use of punctuation for a programming language.

The most commonly performed analysis is the measurement of change over time, and the prevalence of date functions is high. Therefore, it is worth the time and effort to master them early on in your work with Tableau Public.

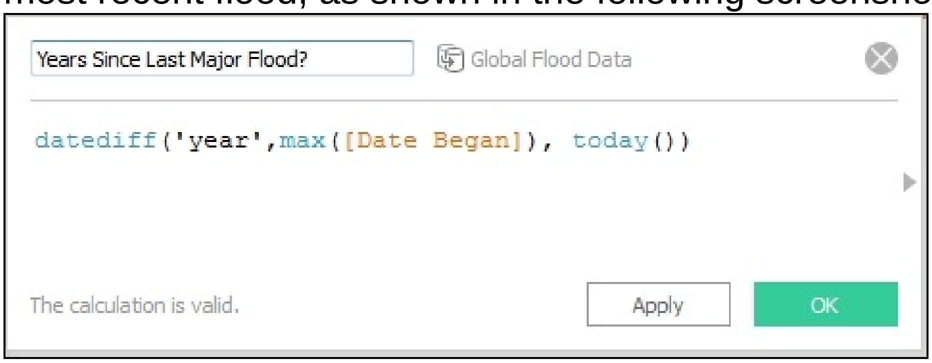
The date functions include the following functions:

- **DATEADD**: This adds numeric values, which can be hard-coded or variable, in a specific number of dateparts to a date.
- **DATEDIFF**: This calculates the elapsed number of specified dateparts between two dates.
- **DATENAME**: This produces the name of the specified datepart of a date. An example that demonstrates the implementation of this function is, `DATENAME('month', #7/21/2015#) = "July"`.
- **DATEPARSE**: This turns a string into a date in the format that you specify.
- **DATEPART**: This is similar to DATENAME. It returns the datepart of a specified date but returns a numeric value. An example that demonstrates the implementation of this function is, `DATEPART('month', #7/21/2015#) = 7`.
- **DATETRUNC**: This rounds up a date to the first date of the datepart that you specify. An example that demonstrates how to implement this function is, `DATETRUNC('month, #7/21/2015#) = '7/1/2015'`.

- **DAY**: This returns the day of the month of a given date. An example that demonstrates how to implement this function is, `DAY(#7/21/2015#) = 21`.
- **ISDATE**: This tests whether a string that you enter is actually a date and produces a Boolean value (`true` or `false`).
- **MONTH**: This produces the numerical month of a date.
- **TODAY** and **NOW**: These produce the `date` and `datetime` of the current moment by using the time settings on your computer.
- **YEAR**: This produces the year of the current date.

In the following example, we will use the floods data to determine the number of years that have elapsed since the last major flood. Most of the countries have had many floods and not just one. Therefore, we used the maximum value of **Date Began**, which is the most recent date in the data set, and compared it with today's date.

We could have omitted the **max** from the **Date Began**, but for each country, Tableau Public would have then aggregated the elapsed time per country, and we really wanted to know how long it has been since the most recent flood, as shown in the following screenshot:



Type conversions

We will begin with type conversions because you cannot create properly functioning calculated fields without having the fields in the right format. For instance, aggregations work on numeric fields, string functions work on strings, and date functions work on dates. Tableau Public automatically identifies the type of field based on its contents, but this may not meet your needs to aggregate or create calculations. Publicly available data sources, such as the ones available at http://public.tableau.com/s/resources, often have issues with either the format or the quality of the data. This will make it necessary for you to convert fields into different types, and it is quite likely that you will need to strip bad data and identify replacement values as well.

If you are familiar with relational database management, you are probably aware that there are many different field types, some of which are used to govern the field length within a database. Tableau Public focuses on the three primary field types, namely; string, which can be anything; numeric fields, which include integers and float fields; and datetime fields; which include the DATETIME, DATE, and TIME field formats.

You can convert a field to another type by creating a calculated field and wrapping it in the appropriate function. In case fields are of the same parent form or are interchangeable, you can right-click on the field name, select **Change Data Type**, and then choose the appropriate type. If you do need to change the format manually, you can use the following functions:

- **DATE**: This converts the strings that you specify (whether you enter them yourself or they are predefined) into date fields, which can then be used to drill through hierarchies.
- **DATETIME**: This converts a field or string into a datetime field, which is a date plus a timestamp.
- **FLOAT**: This converts integers into decimal numbers. This function works only on numeric formats.

- **INT**: This takes the whole number of a numeric expression by rounding up to the nearest whole number.
- **MAKEDATE**: This formats an entry for a year, month, and day into the MM/DD/YYYY format as a proper date. It also accepts fields for the datepart variables.
- **MAKEDATETIME** and **MAKETIME**: These functions accept proper dateparts, just like MAKEDATE, to convert variable or hard-coded sequences into the appropriate field types.
- **STR**: This turns a sequence of characters, numbers, letters, or special characters into strings.

The string functions

The string fields are often the richest in data source as they include other fields with free text, but they can also have the lowest data quality. The string functions in Tableau Public empower you to perform `splice`, `trim`, `find`, `replace`, `match`, `reformat`, and `concatenate` functions.

We will use several of these functions to clean up the floods data source, which was assembled manually and has several data quality issues.

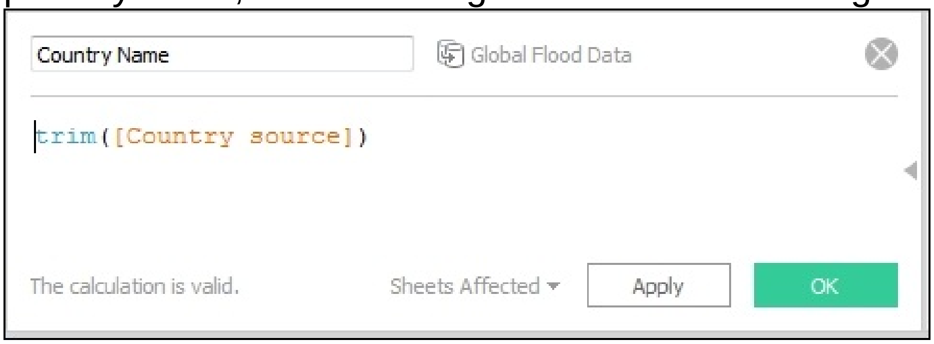
The following are some of the most useful functions that we will use:

- **CONTAINS**: This has a Boolean output. It tests whether a field contains the specified string.
- **FIND**: This finds the place where a string of characters is located within a field.
- **LEFT**, **MID**, and **RIGHT**: These often work in conjunction with FIND and LEN when extracting fixed or variable strings of characters from a field.
- **LEN**: This produces the length of a field.
- **MIN** and **MAX**: These are commonly used on numeric fields. They produce the numerically minimum or maximum values in sequences respectively.
- The **REGEX** expressions: These features were introduced in Tableau Public 9.0. They extract, match, and replace variable strings within fields and are similar to the LIKE function in ANSI SQL.
- **REPLACE**: This replaces a sequence with a specified value.
- **TRIM**, **LTRIM**, and **RTRIM**: These trim leading or lagging spaces from a string.
- **UPPER** and **LOWER**: These are commonly used to normalize the contents of a field.

> ### Note
>
> The Tableau Public does not have the PROPER function that both ANSI SQL and Microsoft Excel use to impose proper

capitalization on inconsistently capitalized fields.

In the next few examples, as shown in the following screenshot, we will use the TRIM and REGEX functions to repair the quality of the most varied misspellings of several countries. It looks like many of the values in the **Country Name** source field have spaces on either side of the primary value, which is not great and it's something that we need to fix:



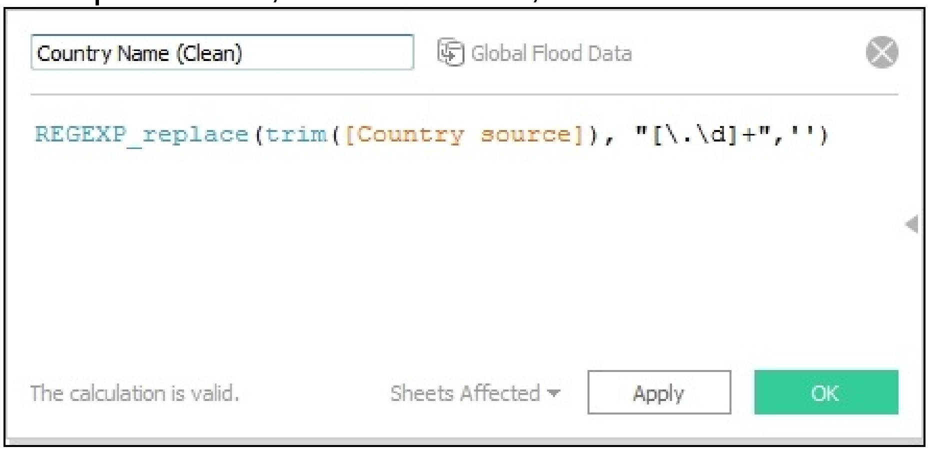
This is helpful. We now have fewer unique entries for country names than before, but bad characters are still present in the data-set.

There are two options for the replacement of characters in a string, namely REPLACE and REGEXP_REPLACE:

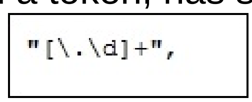
- **REPLACE**: This requires you to enter the string to be searched, the exact pattern to be replaced, and the replacement value. The advantage of using REPLACE is that it is easy to learn and execute. The disadvantage is that Tableau Public can search and replace only one string at a time.
- **REGEXP_REPLACE**: This is a robust function, and while it takes practice to master it, this is widely used in different programming languages. It is a type of regular expression. Among the technical users of Tableau Public, the introduction of regular expressions in 9.x was highly anticipated because the search pattern allows a high degree of variability. You can enter a specific letter, number, or special character as the pattern that you need to search, or you can tell Tableau Public to find letters, numbers, or a variety of special characters. You can also ask Tableau Public to find a combination of these.

Regular expressions are efficient, and they require less hard-coding than the REPLACE function. This means that you will need to modify them less as your data set changes. This is important with data sets that you (or other people) are compiling because humans inevitably introduce errors. Anyone can fat-finger a country name when they're entering data.

The regular expression that we entered includes the string to search (**Country source**), the pattern to search, which is entered in quotation marks, and the replacement, which is blank, as shown in the following screenshot:



The pattern that we entered, which is also called a token, has several parts. This is shown in the following screenshot:



Everything within the quotation marks is a *token*. Working from the outside, the plus sign asks Tableau Public to match any of the expressions inside the brackets. Within the brackets, there are two expressions that need to be found, namely a period and a letter `d`. These expressions are very different from each other; the period is a literal, which means that we want Tableau Public to find all the periods. The `d` is a variable, and it means that we want Tableau Public to find numbers.

The two expressions, namely the period and the `d` variable, are preceded by a backslash. This is an escape. The concept of escapes is beyond the scope of this book. However, it is critical to master them if you want to learn how to structure data because they tell programming applications where to break apart long strings of variable characters, such as URLs. In this case, they ask Tableau Public to look for exactly the expression that we have entered.

If you would like to master the advanced string functions of Tableau Public, which are portable to other programming languages and are a good investment of time in case you're planning on producing advanced analytics or would like to build a career in data science, check out Mark Jackson's (http://www.twitter.com/ugamarkj) blog post at http://ugamarkj.blogspot.com/2015/01/tableau-90-and-regular-expressions.html and Joshua Milligan's (http://www.twitter.com/vizpainter) post at http://vizpainter.com/my-favorite-tableau-9-0-feature/. Mark and Joshua are both Tableau Zen Masters. It's a coincidence that the search pattern in our example matches that of Mark's. In line with full disclosure, Joshua Milligan is one of the reviewers of this book and an author who has worked with Packt Publishing on a book on Tableau as well.

The aggregate functions

Aggregation functions in Tableau Public are performed typically on numeric fields. In this section, we will show you how to use the default aggregations on visualizations as well as how to use them in calculated fields.

The following are the aggregate functions that are available for you to apply on a field are also available in the calculated field dialog box. Tableau Public has arranged them in order from the greatest to the least commonly used functions for the visualization as follows:

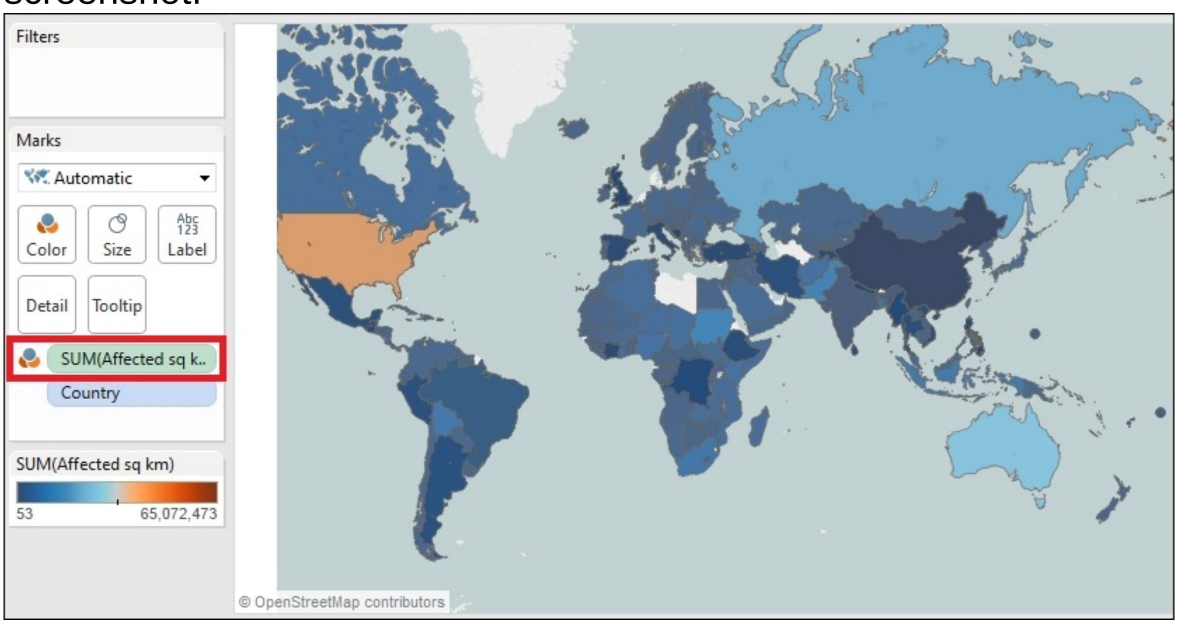
- **SUM**: This adds up the values within a partition
- **Average**: This sums the measure and divides it by the number of dimension members in the partition
- **Median**: This provides the measure value that's halfway between the greatest and least values
- **COUNT** and **COUNTD**: These count the number of dimension members and the number of distinct dimension members respectively; they are typically performed on dimensions rather than measures
- **Minimum** and **Maximum**: These take the least and greatest values in the partition respectively
- **Percentile**: This provides the numeric value at which a percentage of the partition that you specify falls beneath (we'll use this for several examples)
- **Standard Deviation**: This is the square root of the variance for a data set, and it is the unit of measurement of distance from the mean within a data set; it is commonly represented by the Greek letter, sigma
- **Variance**: This is the average distance from the mean value, and it's used to measure the distribution of a data set

Note

The variance and standard deviation are two of the basic measurements in statistics. Advanced statistics in Tableau Public are beyond the scope of this book, but if you understand the concepts, you will find that there are several capabilities within Tableau Public that provide rich statistical functionality.

The most basic aggregation is a sum, and it's the default for numeric fields in Tableau. The second and third most basic aggregations are counts and averages. When you add a measure to a visualization, Tableau Public automatically sums it. In the following example, which uses the data on global floods that we referenced earlier, we do the following three things to create a map:

1. Double-click on **Country** in the **Dimensions** pane to create a symbol map.
2. Double-click on **SUM(Affected sq km)** in the **Measures** pane to change the map into a filled map.
3. Right-click on the **Context** menu for the color legend. Change it to a reversed *orange-blue diverging color spectrum* so that the higher numbers represent greater flood damage, as shown in the following screenshot:



Note that on the **Marks** card, Affected sq km is summed. Therefore, you see the total area affected for each country in the whole data set. There is a major flaw with this method of aggregation, it is not normalized. It would be more useful to have a rate, such as the average area affected per flood, or to add context by creating an aggregation that tells you the percentage of the total area that was affected. For instance, we know that in Pakistan, there were catastrophic floods in 2010, and China traditionally also has had severe floods. However, these countries are in blue, which seemingly indicates that their floods were less severe.

Let's start changing aggregations, and thus add context and reduce the likelihood that someone will misinterpret our visualization, by changing the `Measure` function to an `Average` function. The easiest way to do this is by performing the following steps:

- Click on your aggregation context menu.
- Click on **Measure(SUM)**.
- Click on **Average**.

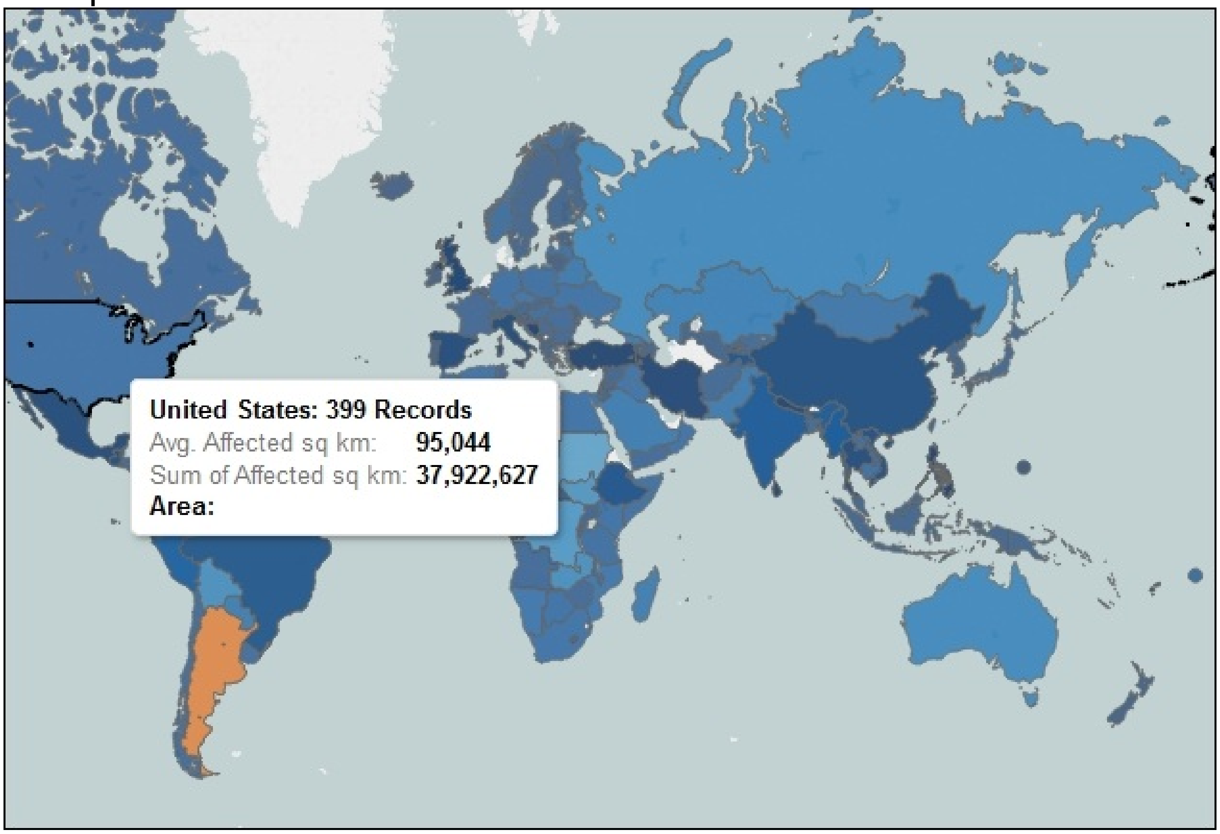
Changing this measure to an **Average** will average the square kilometers of all of the rows related to each country. This means that if there were three floods in the US and they affected 1 million, 2 million, and 3 million square kilometers respectively, then the average would be 2 million.

Tip

If you would like to set the default aggregation for a field, you can right-click on it in the **Measures** pane, click on **Default Properties**, select **Aggregation**, and select the appropriate aggregation for your needs.

The result is a visualization that shows much more appropriate context. From this, we can gauge the average severity in terms of the total area affected by each flood in every country. In the next chapter, we will rank the countries, which adds even more context and relates them to each other for the user.

We added the number of records as well as the sum of the affected square miles to the tooltip so that when you roll over a country, you see the total number of floods, the total area affected, and the average area affected by each flood, as shown in the following screenshot. Using a tooltip is a great way of adding context, which is the art of relating two data points to each other and to the consumer:



The logic functions

Logic functions enable you to tell Tableau Public what to do when certain conditions are met. They are also known as conditional statements. The format is commonly referred to as IF THEN ELSE. You, as a programmer, ask Tableau to test whether a row of data meets a certain condition. If it does meet the condition, then there is an output, which can be another field, a discrete number, or a string. If the condition is not met, then you want a different output.

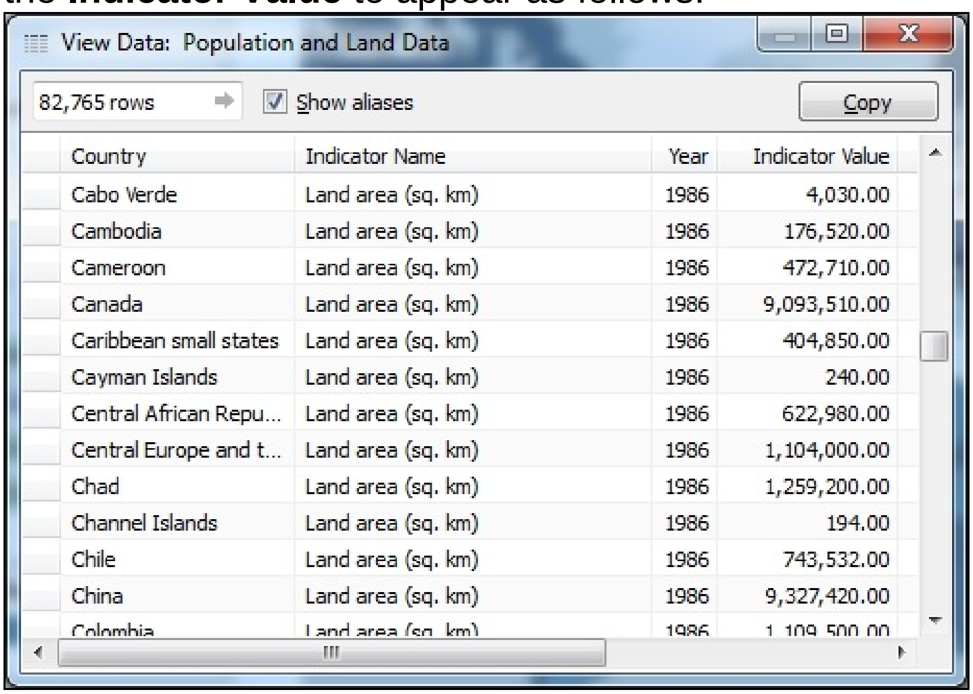
There are several logic functions in Tableau. Some of these functions are sub-functions or parts of others. We will focus on the following major functions:

- **IF**, followed by **THEN**, **ELSE-IF**, or **ELSE**: This tests whether a condition is met and show the result if it is met as well as other conditions that need to be tested and results that need to be produced in case none of the conditions are met.
- The **IIF** statement: This tests for a minimum of one condition and then provides the results when the condition is or is not met. It follows the same format as that of IF in Microsoft Excel.
- The **IFNULL** function: This identifies what Tableau Public should do in case a value is null. This is particularly useful for aggregations of dirty data sets, which are very common.
- The **CASE** statement: This tells Tableau Public what to do when a parameter or a string field has a very specific value. Unlike the CASE statements in ANSI SQL, Tableau Public does not accept aggregations in these.
- **AND**, **OR**, and **NOT**: These link or negate conditions that need to be met.
- **END**: This is critically important, as it terminates the loops of the IF and CASE statements. Tableau Public will tell you in case you need to add it and have not done so.

The IF statements are very useful. You can use them to group members and set thresholds, among many other uses. The condition that needs to be met can be a variety of things. In the first example, we will create a new measure in a transformation of the World Bank indicator data that we have extracted and transformed, and it's called population and land data.

We will create a new measure that gives us the total land area of a country. We want to use this measure to determine the percentage of each country's area that was flooded, but this measure does not exist in the Floods data, but it does exist in the **Population and Land Data**. Later in this chapter, we will blend the data sources to create a calculated field that uses it in the denominator.

The Population and Land Data does not have a metric called total area. The way the data source is structured, the measures are in rows and not columns, with a field representing the corresponding metric value, as shown in the following screenshot. We need to create a calculated field that states that if **Indicator Name** is **Land area (sq. km)**, then we want the **Indicator Value** to appear as follows:

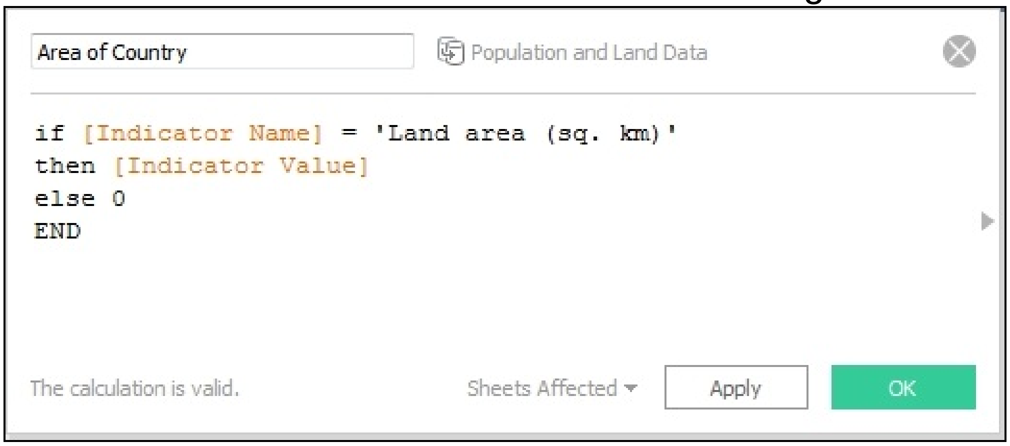


Country	Indicator Name	Year	Indicator Value
Cabo Verde	Land area (sq. km)	1986	4,030.00
Cambodia	Land area (sq. km)	1986	176,520.00
Cameroon	Land area (sq. km)	1986	472,710.00
Canada	Land area (sq. km)	1986	9,093,510.00
Caribbean small states	Land area (sq. km)	1986	404,850.00
Cayman Islands	Land area (sq. km)	1986	240.00
Central African Repu...	Land area (sq. km)	1986	622,980.00
Central Europe and t...	Land area (sq. km)	1986	1,104,000.00
Chad	Land area (sq. km)	1986	1,259,200.00
Channel Islands	Land area (sq. km)	1986	194.00
Chile	Land area (sq. km)	1986	743,532.00
China	Land area (sq. km)	1986	9,327,420.00
Colombia	Land area (sq. km)	1986	1,109,500.00

In order to do this, we will follow these steps:

1. Create a new calculated field in the **Population and Land Data** source by clicking on the **Context** menu in the Data window and selecting **Create Calculated Field**.

2. Name the calculated field *Area of Country*.
3. Start an IF statement that says that if the **Indicator Name** is **Land area (sq. km)**, then Tableau Public should use **Indicator Value**. Otherwise, it should use 0.
4. End the statement.
5. Validate that there are no error messages.
6. Click on **OK**. The result is shown in the following screenshot:



In this statement, if we did not include the ELSE statement, then the products of the field would be null for every value of **Indicator Value** that isn't specified, and we do not want nulls, because performing any kind of mathematical operation on a null value results in a null value even when you're adding or multiplying it by valid numbers. So, we included **else 0**, even though we did not necessarily need to do so, in order to preserve the integrity of the data set.

Another good use of an IF statement is to create groups or establish performance thresholds. In the World Bank Indicators data source, there is a field that identifies the percentage of its GDP that each country spends on public health. The first quartile is at five percent, the median is at seven percent, and the upper quartile is at nine percent. In the following example, we will use mathematical functions to group the countries according to what their average expenditure on public health is. (You can download this data set from www.worldbank.org).

In this example, we have hard-coded the values in the inequalities, which means that the thresholds used for the groups are not dynamic. In the next chapter, we will use window calculations to make them dynamic.

We created a simple box plot that shows the maximum health expenditure as a percentage of the GDP per country in 2010. Since we put **Country** on the **Detail** shelf, Tableau Public graphs one mark per country. In the calculated field, we even accounted for null values, as represented in the ELSE statement. Then, we put this new calculated field on the **Color** shelf of the box plot, as shown in the following screenshot:

Perform the following steps to create this graph:

1. Add **Health**: **Health expenditure, total (% GDP)** to the **Columns** shelf.
2. From the **Analysis** menu, deselect **Aggregate Measures**.
3. Drag **Date** to the **Filters** shelf and select **YEAR**.
4. Filter **YEAR** to 2010.
5. Drag **Military: Expenditures (% GDP)** to the **Size** shelf.
6. Drag **Country** to the **Detail** shelf on the **Marks** card.
7. Create a new calculated field called **Health Expenditure Quartile**, and provide Tableau Public with several conditions to test.

> ## Note
>
> We wrote the conditions in a specific order so that for each country, the accurate value is the output.

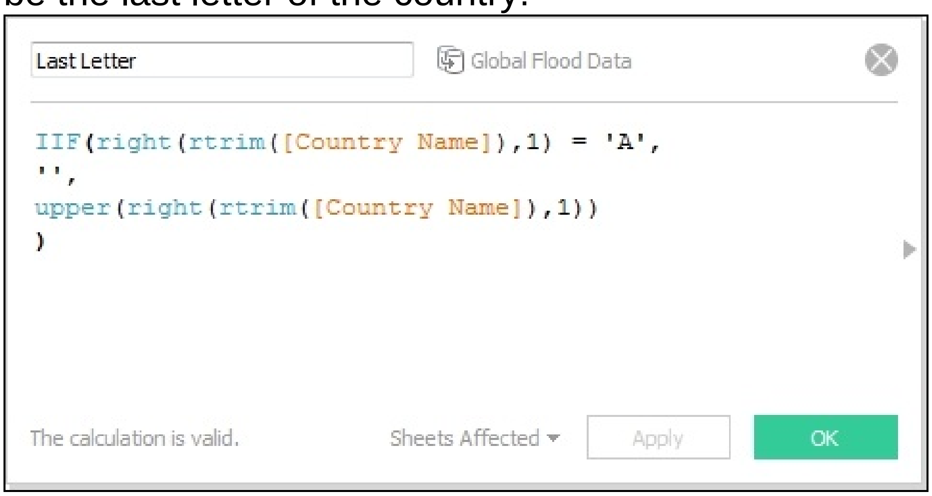
8. Drag the new field named **Health Expenditure Quartile** to the **Color** shelf.
9. Adjust the colors so that the lower quartiles are in red and the upper quartiles are in blue.

This IF statement is similar to the two other types of logic statements in Tableau Public, namely IIF and CASE.

The structure of IIF is the same as that of the IF statements in excel, and it's very similar to the IF statement that we just created in that you are asking Tableau Public to test a condition and then provide the results according to the logical outcome, which can be either true or false. In the following example, we will ask Tableau Public whether the last letter of the name of each country is *A*.

There are two possible outcomes—yes, the condition is met, and no, it is not met. We tell Tableau Public what value to use for each possible outcome. We can use a string, a number, another field or aggregation, or even nothing at all.

In the following example, which is for demonstration purposes, we ask Tableau Public to test whether each country name in the global major floods data, which was trimmed in the section on string functions, ends with the letter A. If the condition is met, the result should be blank (which is different from NULL), but in case it is not met, then the outcome should be the last letter of the country:

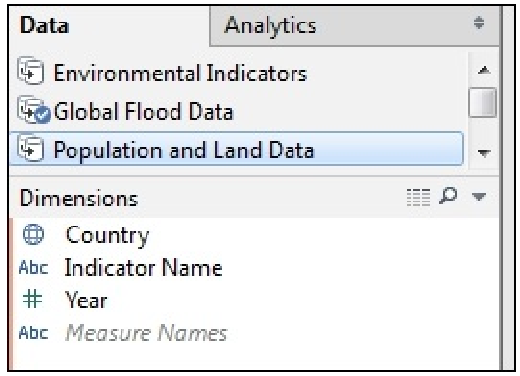


Blending data sources

If you want to create a calculated field that produces the percentage of the total land flooded during an event, we need to use a data source that has this information in it. Most stories need more than one data source. When you are adding in secondary data sources, it's important to know what the level of aggregation is if you have joined data sources with different levels of aggregation and have not accounted for that when blending them because if you fail to do so, you might get the wrong results in your calculated fields.

The flood data is aggregated at the event level. This means that for every major flood, we have summarized data—the start date, end date, location, total land area affected, and so on. This is moderately granular data. If it were even more summarized, it might be summed up as floods per year per country.

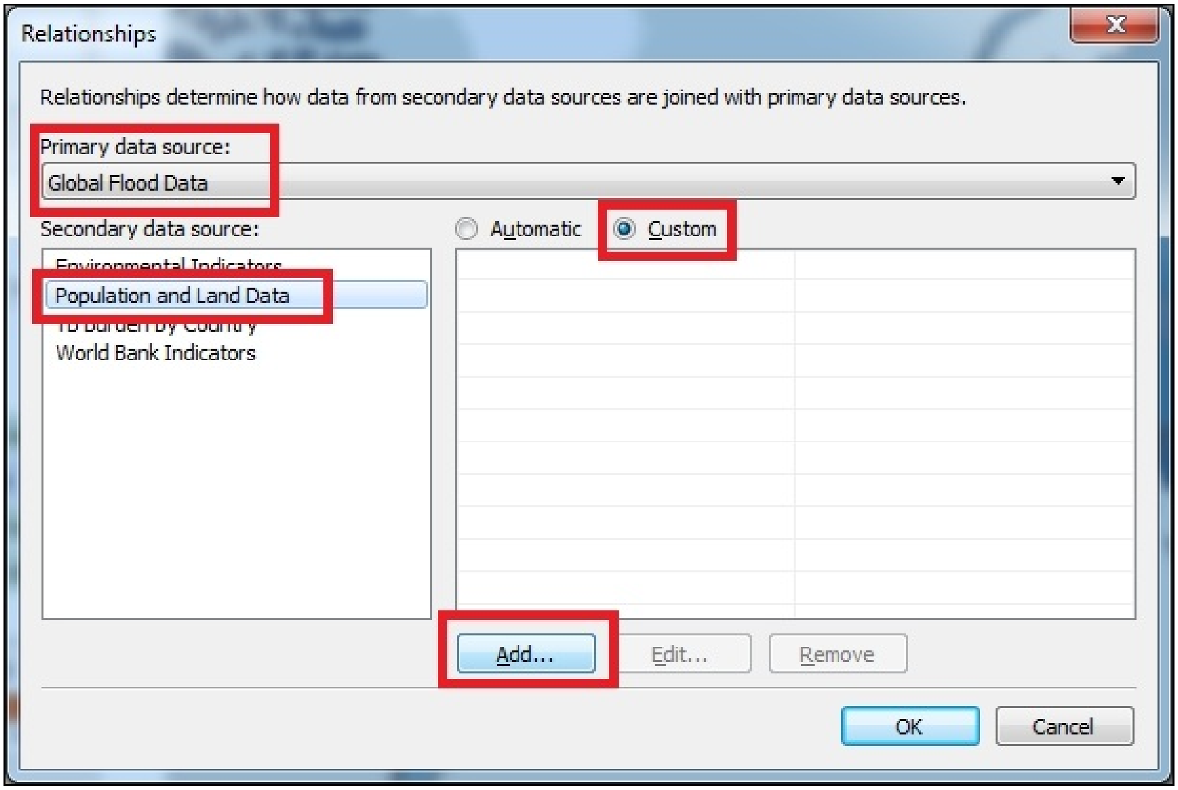
We want to know the total percent of land flooded each year, and for this, we will assume that the area of each country is static throughout time. So, we created a new data source. We filtered the World Bank indicators data (available at http://data.worldbank.org/data-catalog/world-development-indicators/) in this data source to include only the indicators pertaining to the land area or population. The World Bank indicator data is aggregated by country by year, which is not the same level of aggregation as that of the data on floods; it is higher. So, in our calculated field, which will sum the total area flooded and then divide it by the total area of each country, we will need to tell Tableau how to aggregate the total area of each country. Otherwise, we will get the wrong results.

First, we need to join the data sources. In the companion workbook, which can be found at [X], we have a data source called **Population and Land Data**. This is the version of the World Bank indicators that we have filtered. It has only four dimensions, in this case, we have three, **Country**, **Indicator Name**, and **Year**. The data source on floods has Country Name and several other date fields. We want to join on **Country** name, as shown in the following screenshot: 

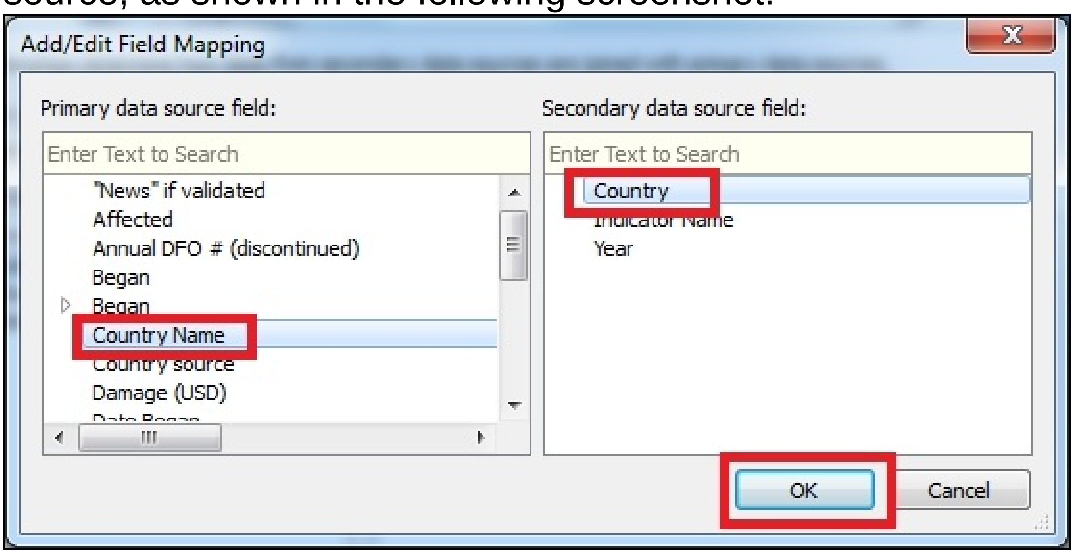
If we look at the **Population and Land Data** source, we do not see any of the gray chain links next to our dimensions that indicate that the fields are linked. There's a reason behind this. The names of the fields are not exactly the same. In order for Tableau Public to identify the join conditions automatically, the field names need to be exactly the same, with both capitalization, punctuation, and content. The field types need to be exactly the same as well.

From the **Data** menu, we can create a join on **Country** name in the Floods, as follows:

1. Click on **Edit Relationships**.
2. In the **Edit Relationships** dialog box, which is shown in the following screenshot, ensure that the appropriate data source is selected as the primary data source, which is the data source from which the first field is added to the visualization.
3. Select the appropriate secondary data source.
4. Click on the **Custom** radio button in case it is not selected by default. In this case, there are no automatic joins. Therefore, Tableau Public assumes that we need custom joins.
5. Click on **Add**:



Next, we need to tell Tableau Public that the Country Name field in the major global floods data, which was created to remediate data quality issues, is the same as the Country field in the major global floods data source, as shown in the following screenshot:



You can create only one custom join at a time. So for instance, if you would like to add more joins, such as a date field, you need to add each one individually by clicking on the **OK** button and then clicking on the

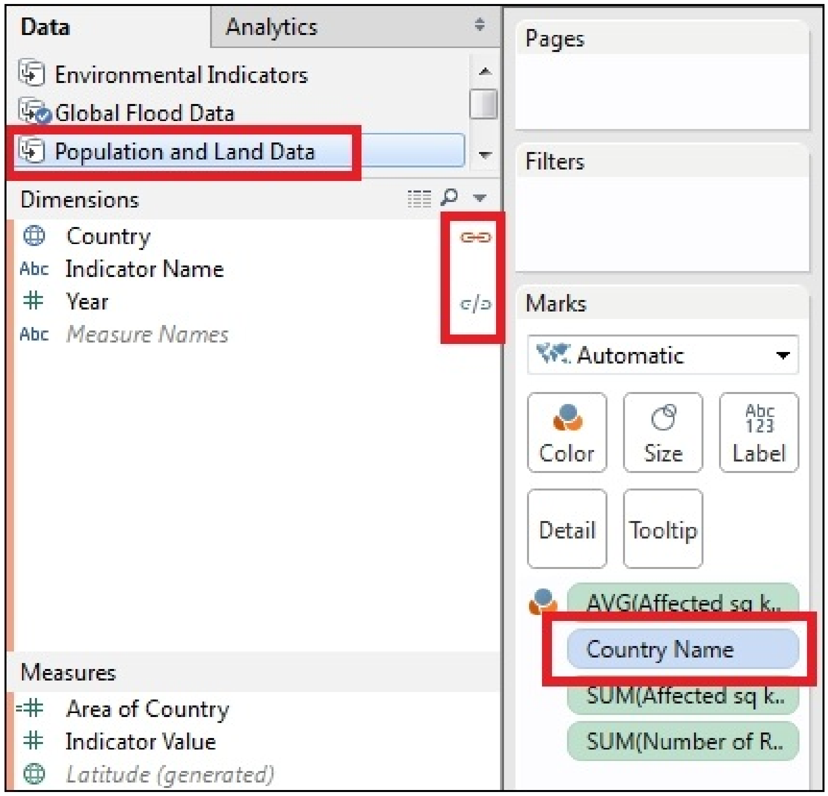
Add button again, which we did in order to add a second join on the year of the event.

Note the small arrow next to the **Began** field, which is above **Country Name**. This field allows you to establish a join condition on a specific date part. Since we also wanted to join the data sources on the year of the flood event, we expanded the list to see the appropriate dateparts, clicked on **Year**, and then selected **Year** in the secondary data source.

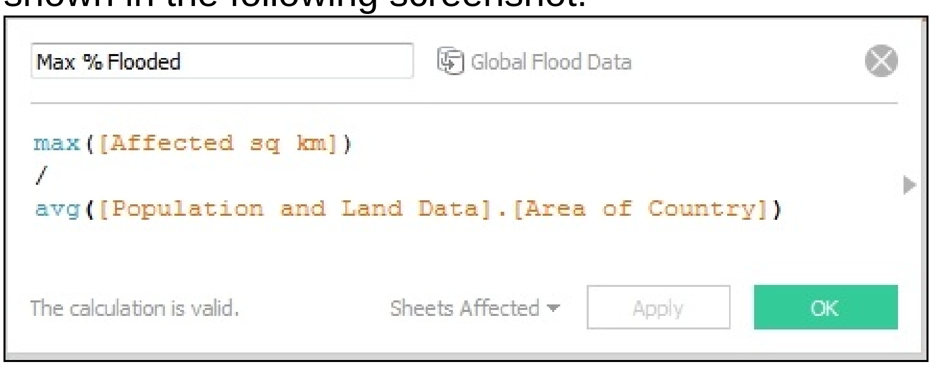
We clicked on **OK** again to return to the visualization, which is still the map that we created that shows the average flooded land.

The view is now different. Looking at the **Data** window for the **Population and Land Data** source, you will see that that there are two chain links, the chain link for **Country**, and the following is orange. Because the linking field is a part of the visualization, you will see that **Country Name** is on the **Detail** shelf on the **Marks** card. However, **Year** has a gray, broken chain link because it is not a part of the visualization. If you want it to be used in the calculated fields, click on the gray chain link to activate the join. Even though the field might not be on the visualization, it still be a part of calculations using fields from its data source, as shown in the

following screenshot:



We can create a calculated field that produces the average amount of land in each country that was flooded. In this case, since some countries might have had major floods that add up to more than 100 percent of the total flooded land, we will create the maximum percent of land flooded, as shown in the following screenshot:



In order to create this calculation, perform the following steps:

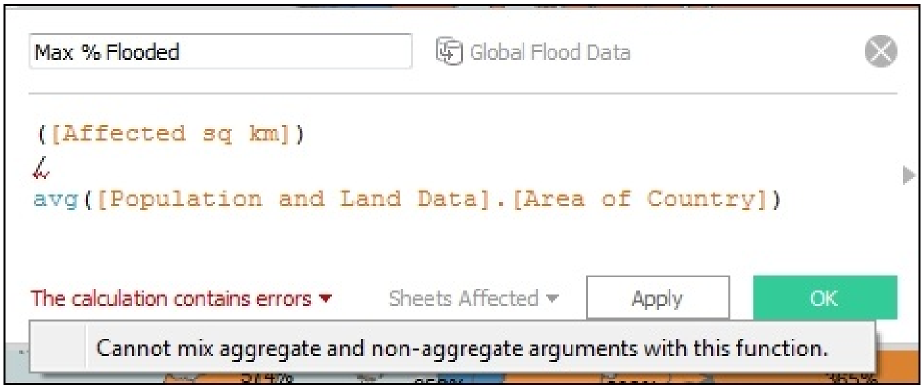
1. Create a new calculated field in the major global floods data source.

> **Note**
>
> It is important that the place where we create the new calculated field determines the data source that is the primary data source.

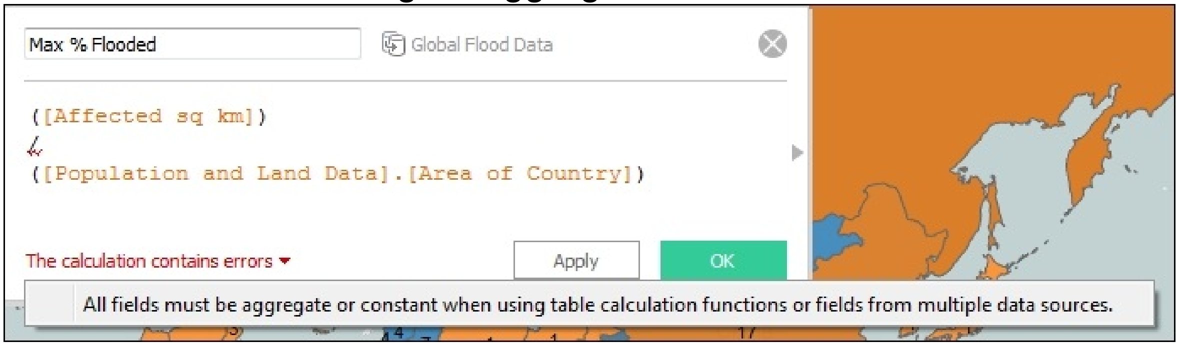
2. Name the field.
3. Add **Affected sq km** to the numerator by dragging it into the dialog box.
4. Wrap it in a **max** function.
5. Add a division sign (***/***) because we want to divide it by the total area of the country.
6. Click on the **Population and Land Data** source and drag **Area of Country** to the denominator.
7. The denominator is automatically summed. You can change the sum to whatever aggregation you would like, but if you do not aggregate the fields that are from the secondary data sources, you will get error messages.
8. Change the aggregation of the field from the secondary data source, which is the field whose name is preceded by its data source name. This is similar to a SQL query that is used to find out an average. The reason behind why we did this is that for every record in the partition, Tableau Public needs to know what to do with the denominator. If we sum it, we will not have the proper calculation.

Check out the correct calculated field, It's important to learn how to handle errors. The next two errors are very common and are as follows:

- The following screenshot shows that we cannot mix aggregate and non-aggregate arguments with this function, with the division function highlighted. This means that in a calculated field, both arguments must be aggregated. We can resolve this by aggregating the numerator or by using the `ATTR` function on it. ATTR tells Tableau to use the discrete value for it:



- All fields must be aggregate or constant when using table calculation functions or fields from multiple data sources. This means that the denominator, which is from a secondary data source, must be aggregated. Tableau Public requires this so that it knows how to aggregate. We resolved this by placing the AVG function around the denominator and adding an aggregation to the denominator:



Summary

In this chapter, we explored the different data types in Tableau Public. You learned how to create various types of calculated fields, which enable you to create new metrics and tell stories that are relevant and impactful to users.

In the next chapter, we will build on these capabilities to create Level of Detail calculations and table calculations.

Chapter 6. Level of Detail and Table Calculations

In the previous chapter, you learned how to use custom calculations to create new fields that add context and insights to the story that you are telling with Tableau Public. In this chapter, we will discuss how you can use table calculations and level of detail calculations to enhance the comparisons that you are making with the data. We will build on several of the calculations that we have created to show you how to make them more dynamic and contextual.

The table calculations are different from the traditional calculations because they are performed locally on the data in the cache, that is, the data that Tableau Public has used to create the visualization. Not all data that is in your data set is in use all the time. The data in the cache is what Tableau Public is using in the memory to render your visualization. Therefore, the data that you have filtered out will not be included in the table calculations.

Level of detail calculations, which are a new feature in Tableau 9.x, allow you to tell Tableau Public at exactly which dimensional level you want the calculations to be aggregated. We are discussing them along with table calculations because some of the concepts are very similar.

Table calculations have the following major functions:

- Calculating change over time as a relative percentage
- Calculating the percentage of a whole attributed to a one-dimensional member
- Calculating change relative to other members of a dimension
- Distinguishing the maximum or minimum values in a partition
- Moving calculations, such as averages and sums
- Running calculations, such as running sums, that are used for waterfall and Pareto graphs

The following are the level of detail calculations that Tableau Public enables you to do:

- Fix the level of aggregation
- Include specific dimensions, that may not be present on the visualization, in calculations
- Exclude specific dimensions, that are present on the visualization, from specific calculations

In this chapter, we will discuss the following table calculations:

- Creating quick table calculations
- Addressing and partitioning table calculations
- Changing over time
- Editing table calculations
- Moving averages – window max and running max
- Ranks and percentiles
- Difference from the average

The level of detail calculation concepts that we will discuss involve fixing, including, and excluding dimensions, as well as nesting calculations and limitations. A brief exercise on editing fields in the shelf is also included in this chapter.

About data sources

We will continue using the World Development Indicators data source, as some of these examples will be included in the dashboard that we will develop later in this book. You can download the data from the World Bank by visiting http://data.worldbank.org/products/wdi.

Creating quick table calculations

Tableau Public has a feature that enables the rapid creation of a table calculation. After dragging a field onto the visualization, typically a measure, you can aggregate and perform table calculations on dimensions, or click on its context menu and select **Quick Table Calculation**. There is one limitation to this—table calculations cannot be created on fields that have forecasting turned on.

The following are some of the many different types of quick table calculations, though not all of them are available at the same time, and the options vary by the type of dimensions and measures that are also on the visualization:

- Running total
- Difference
- Percent difference
- Percent of total
- Rank
- Percentile
- Moving average
- Year to date (YTD) total
- Compound growth rate
- Year over year growth
- Year to date (YTD) growth

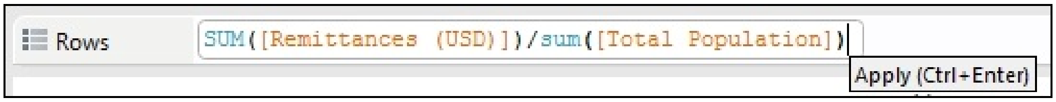
Once you have created a table calculation, you can edit it by either clicking on the **Context** menu on the pill, or right-clicking on it. For each unique table calculation, we will explain how to modify it and what the components of the formula mean.

Changing over time

One of the questions that we commonly ask about data is, how has performance changed over time? This helps us understand how events of decisions affect outcomes. While understanding the relationship between discrete numbers is helpful, it's even more useful to understand the rate of change, or the percentage change over time. The reason why it's important to understand the rate of change is that, while discrete numbers may appear to be increasing, the actual percentage change from year to year may be declining.

In the following example, we will discuss how to graph the amount of remittances over time and we will create a quick table calculation that shows the percent difference:

1. On a new worksheet, using the World Development Indicators data source, drag **Year** from the **Dimensions** pane to the **Columns** shelf.
2. Drag **Remittances (USD**) from the **Measures** pane to the **Rows** shelf.
3. Exclude years before 1980 by using your preferred method. We selected them on the *x* axis, hovered over pill, and selected **Exclude**.
4. We now have a basic line graph that shows change over time, but we're missing context. While we can see that **Remittances** have gone up and down, we know nothing about how they relate to the economy of the region or its population.
5. We can edit the measure to produce a rate per capita. We will do this by editing it in the shelf, which is a new feature in Tableau 9.x, and adding the total population to the denominator (editing in the shelf is a good option for simpler aggregations).
6. On the **Rows** shelf, click on the pill for the measure to see the **Context** menu and select **Edit in Shelf**.
7. Enter a division sign after the closing parenthesis, and from the **Measures** pane, drag **Total Population** to the right of the division sign so that it is in the denominator, as shown in the following screenshot:

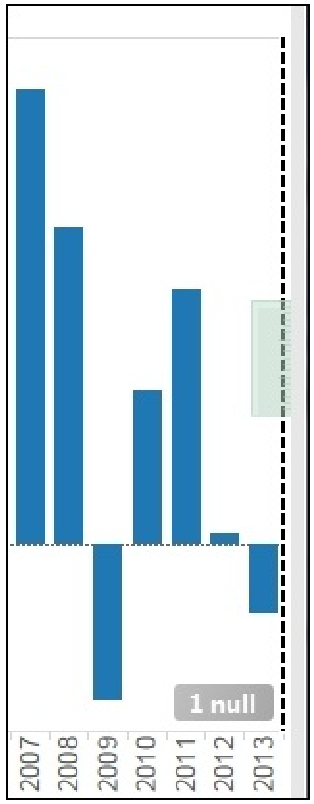


> **Note**
>
> Since the **Remittances (USD)** is summed, we need to sum **Total Population**. In a calculated field, both fields must be aggregated or disaggregated. We could disaggregate, but that would give us the average per country within each region, and when those numbers are added up, they will not represent the data accurately.

- In order to give the new field a name, drag it from the **Rows** shelf to the **Measures** pane, which does not seem intuitive. When prompted, name it **Remittances per Capita**.
- We can create a quick table calculation that shows change over time by right-clicking on the **Remittances per Capita** field on the **Label** shelf, selecting **Quick Table Calculation** from the context menu, and then selecting **Percent Difference**.

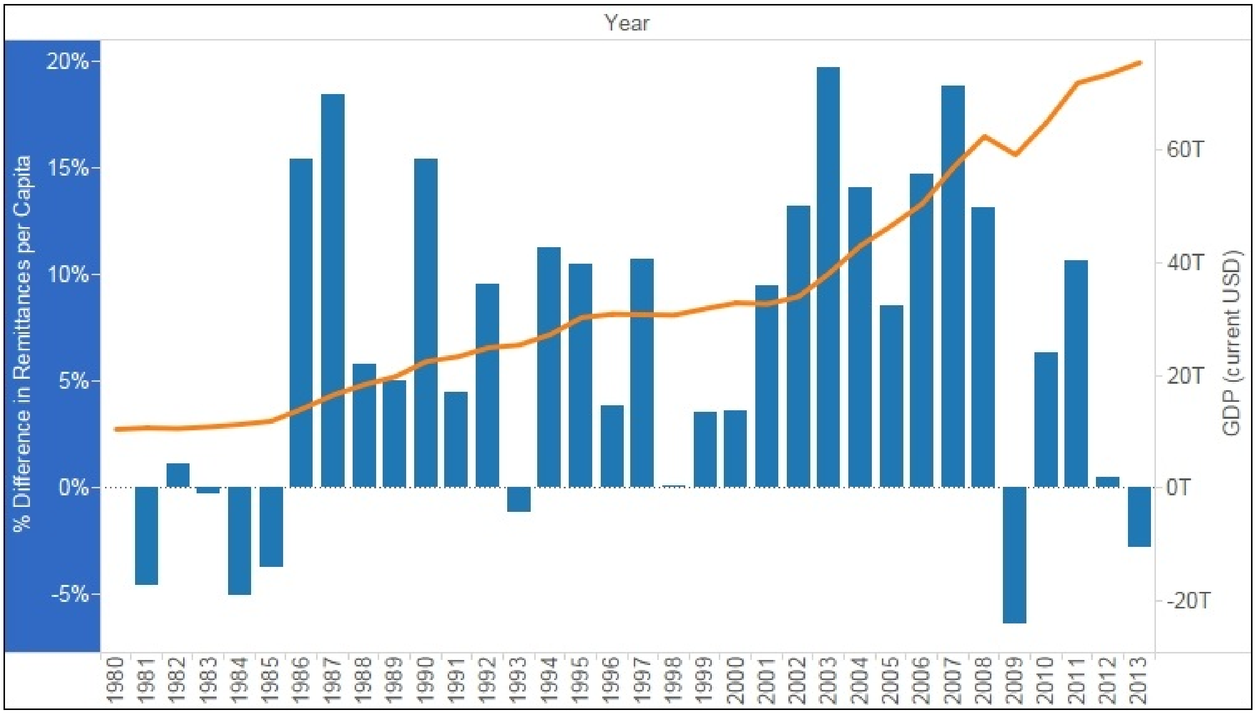
> **Note**
>
> The pill for **Remittances per Capita on the Rows** shelf now has a small delta sign to the right of the name, which signifies that it is a table calculation.

- Change the mark type for **Remittances per Capita** to a bar.
- You can create more context by adding another reference point, namely the GDP. Drag **GDP (current USD)** to the secondary *y* axis, as shown in following screenshot, and then drop it when the secondary *y*



axis has a dashed, horizontal line:

- Right-click on the *y* axis for **GDP (current USD)**, click on the **Mary** type, and select **Line**.
- The following screenshot shows the fluctuation in remittances, as it relates to the global GDP. In 2009, when the global economy was reset, the GDP dropped, and so did the remittances:

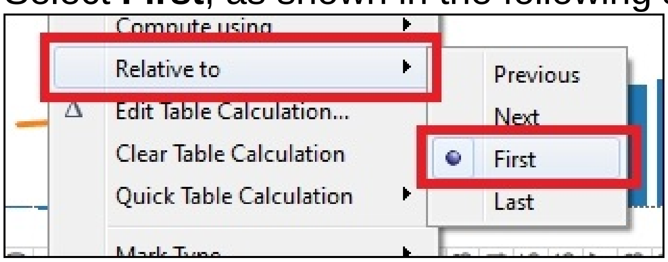

Year
% Difference in Remittances per Capita
GDP (current USD)
20%
15%
10%
5%
0%
-5%
60T
40T
20T
0T
-20T
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013

Compute using

The graph from the previous exercise calculates the percentage difference from year to year, which means that the value for each year is being computed relative to the previous year. We can ask Tableau Public to calculate the percentage difference from different years by clicking on the **Context** menu for a table calculation of this type, and then selecting **Relative to**. We can calculate the change from the first, next, previous (the default one), or last values in a partition.

In the following graph, which is a revision of the previous one, the bar height for each year shows the aggregated percentage difference in **Remittances per Capita** since 1980, which is the first year in the visualization. You can do this by performing the following steps:

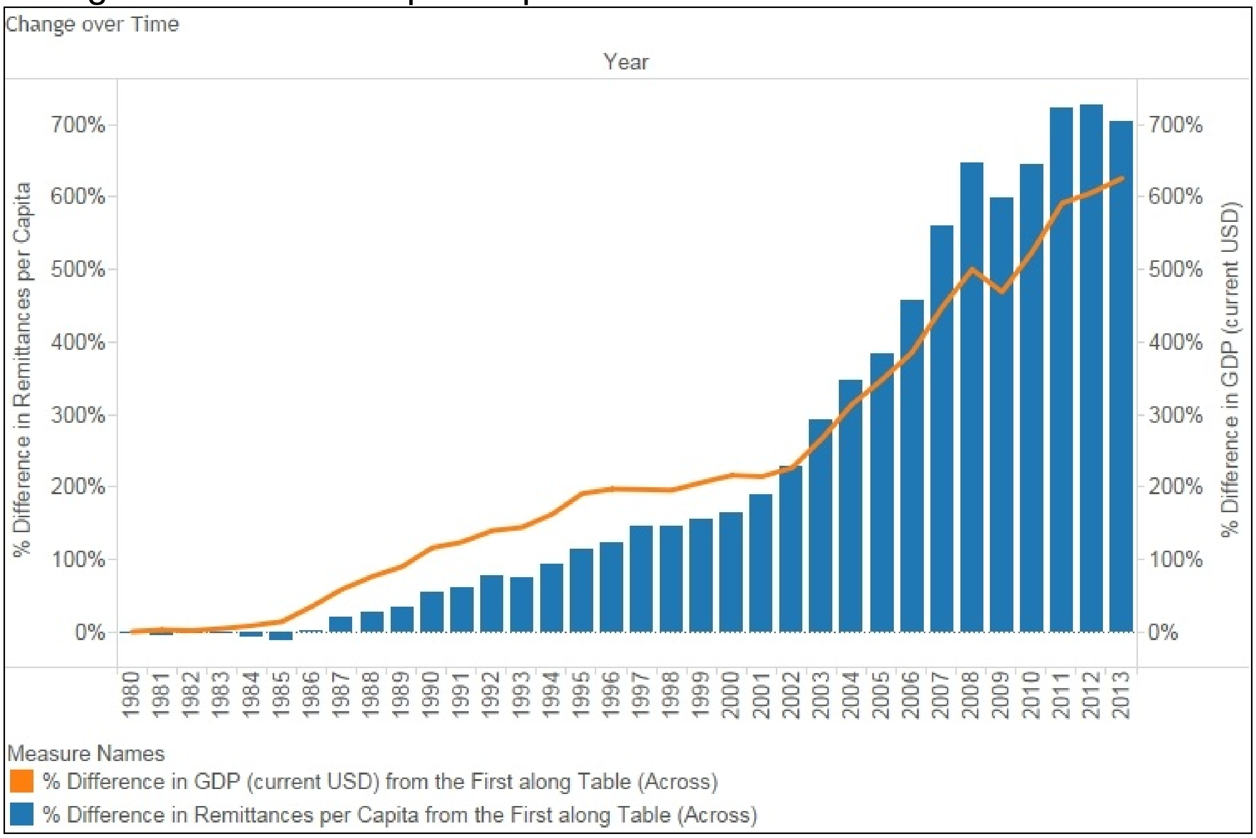
1. Click on the **Context** menu for the **Remittances per Capita** field on the **Rows** shelf.
2. Click on **Relative to**.
3. Select **First**, as shown in the following screenshot:



Then, perform the same calculation on **GDP (current USD)**. We changed its use on the **Rows** shelf into a quick table calculation for the percentage difference and set it to **Relative to** for the **First** value in the partition.

We took one extra step—we right-clicked on the secondary *y* axis, where the **GDP (current USD)** table calculation resides, and selected **Synchronize Axis** so that the axis ranges of both the *y* axes are the same. This is an important step because we want to make sure that the consumers do not perceive a relationship that is different from the one that we intend to communicate.

In the following graph, you will see that just around the turn of the century, the GDP rate of change since 1980 slowed down, and the rate of change of Remittances per Capita overtook it:



Moving average

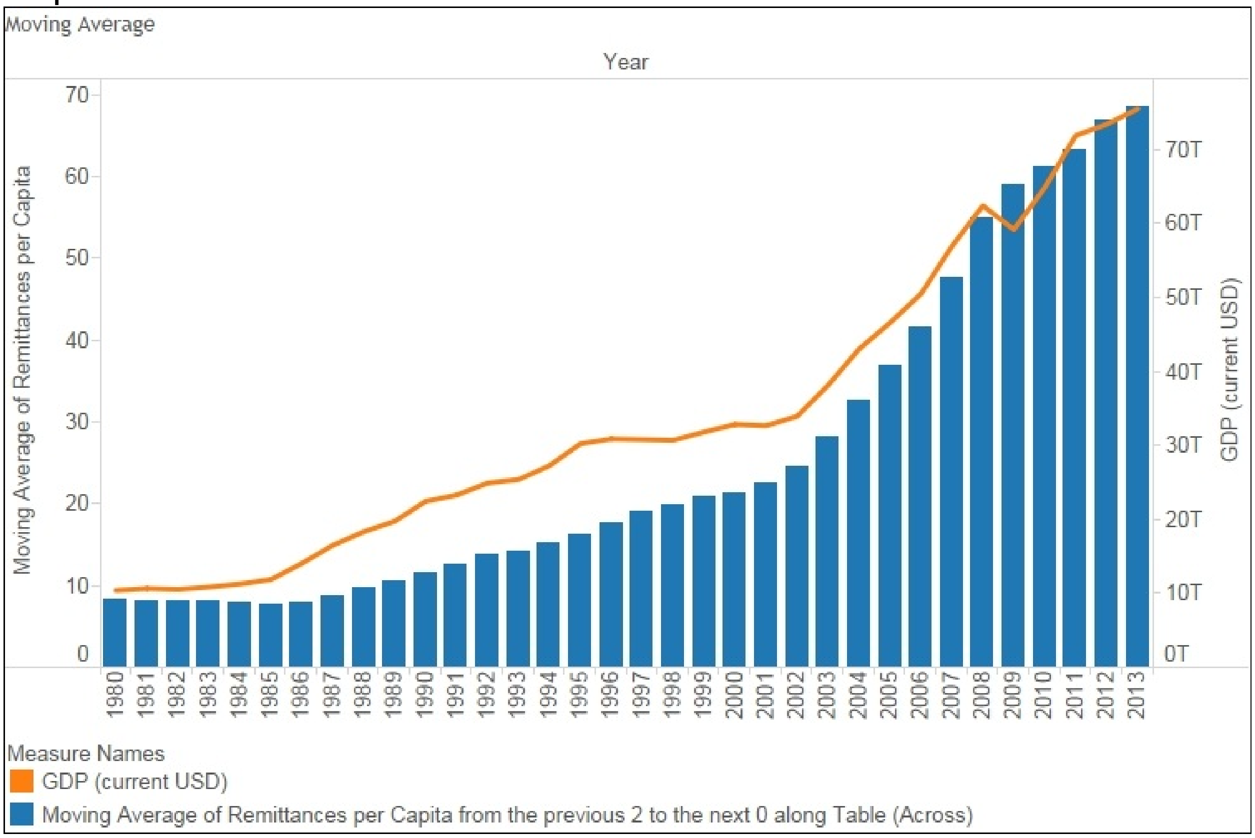
The moving average is a powerful table calculation because it gives you the ability to determine how many previous and future values are included in an average.

For instance, we can change the percent difference table calculation that we performed on **Remittances per Capita** by performing the following steps:

1. Click on the **Context** menu of **Remittances per Capita** on the **Rows** shelf.
2. Click on **Clear Table Calculation**.
3. Click on the **Context** menu of **Remittances per Capita** again, select **Quick Table Calculation**, and then choose **Moving Average**.

We cleared the table calculation for **GDP (current USD)** in the following graph as well.

The following graph shows the moving average of Remittances per Capita since 1980 versus the GDP:

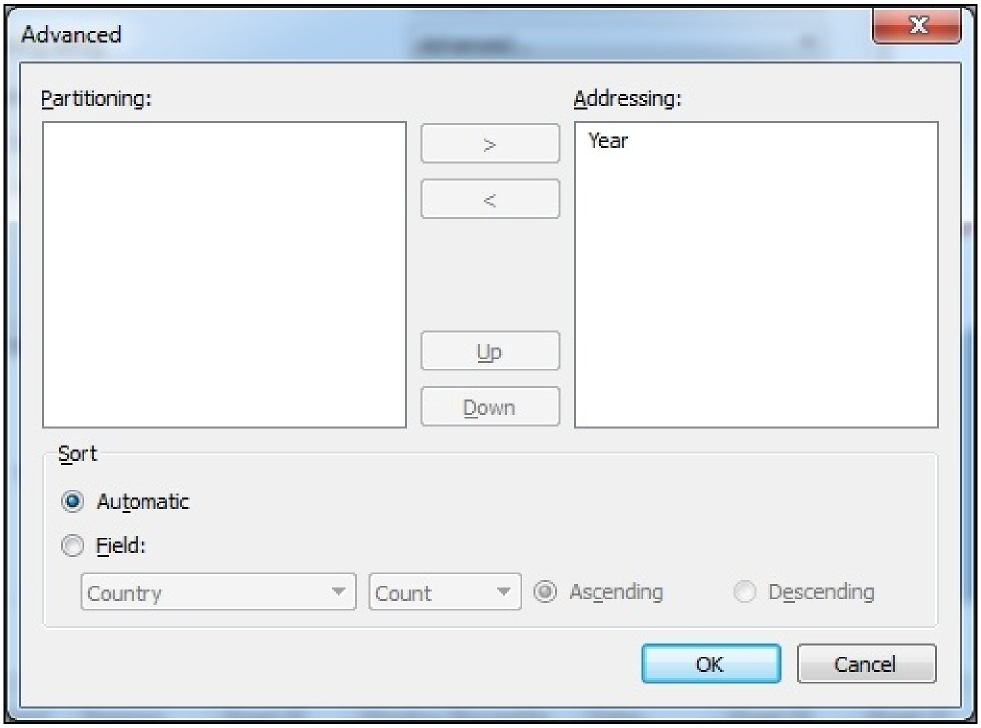


One thing that we don't know is exactly how *smooth* the line is. The term "smooth" means that more than one value is included in the calculation, and if we edit the table calculation, we can control the smoothing. You can also create a parameter that gives users a control over the smoothing. More on this will be discussed later in the book.

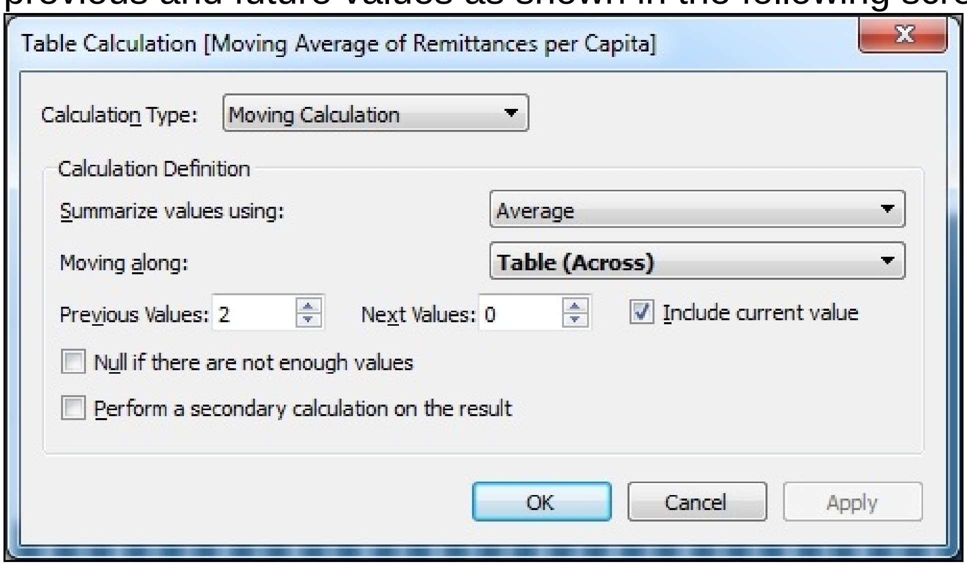
Editing table calculations

So far, we have used only the default settings for the table calculations that we have created. Tableau Public gives us several capabilities. Right-clicking on the pill for the moving average of **Remittances per Capita** shows us the following options that are available for this field:

- **Calculation Type**: We can change the type of calculation via this option. Tableau Public automatically creates a different calculated field depending on our selection, and the options available in this user interface also vary based on the selection.
- **Summarize values using**: This gives us the option to select the aggregation type. Ours is a SUM, but we can use AVG, MIN, and MAX in this type of calculation.
- **Moving along** (or **Compute using**, with other table calculation types): This allows us to address the table calculation. In this example, we have only three options, namely **Table (Across)**, **Cell**, or **Date**. The default is to compute by going across the table, which means that Tableau Public computes the values by going from left to right or top to bottom, depending on the location of the dimensions and measures.
- When you click on the drop-down list, you can customize the way the field is computing by telling Tableau Public which fields to **Compute using** the first field, then the second, and so on. In the following screenshot, we have only one dimension, namely **Year**, in use. In case there were others, they would appear in the **Partitioning** pane, and we could add them to the **Addressing** pane and then sort them again. More on this will be discussed later.
- The dimensions that we do not select for **Compute using** are used for partitioning, which is called grouping:



The next feature allows you to tell Tableau Public how many previous and future values to include, and you can also determine whether the current values should be included. Lines that are smoother include more previous and future values as shown in the following screenshot:



- **Null if there are not enough values**, when selected, does not compute marks for places when there are not enough values. For instance, if you select this for this particular calculation, then the first two years in the visualization will not show a line.

- **Perform a secondary calculation on the result** gives you the power to add additional table calculations. For instance, if this were a running sum that shows the cumulative total, we can use a percentage of the total secondary calculation to show the percentage accumulated for each year instead of the discrete number.

Manually editing table calculations

Learning how to edit table calculations yourself is an advanced capability, but it gives you the opportunity to create rich metrics, such as the percent difference from an average.

In the first example, we will modify the running sum of **Remittances per Capita** so that it includes the preceding and next three values, which can be done from the edit table calculation dialogue by entering those numbers in their respective places.

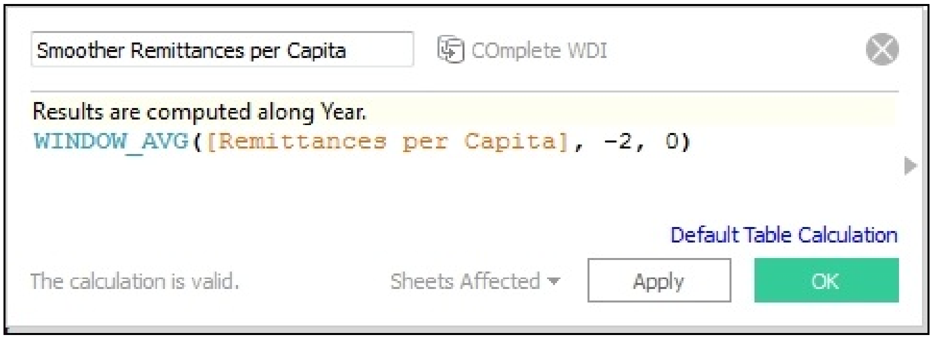
In the second example, we will modify the **Remittances per Capita** percent difference table calculation to show the percent difference from the average.

Let's begin with the first example. We'll modify the running sum in the following way:

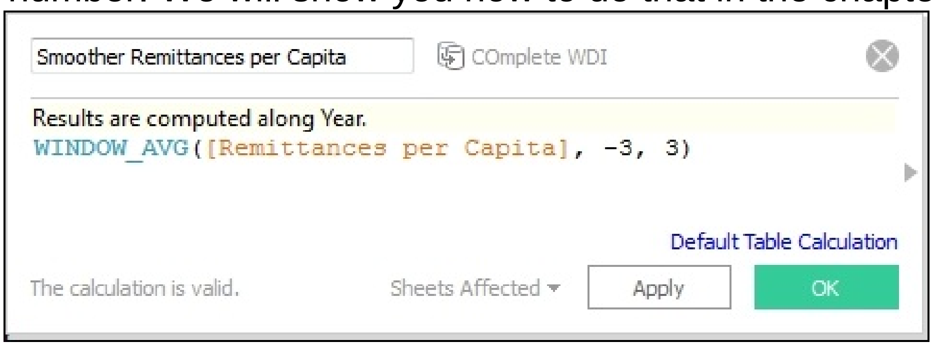
1. In order to edit the table calculation, duplicate the sheet on which we created the original moving average worksheet, which maintains the integrity of the work that we have already done.
2. Then, drag the pill for the moving average of **Remittances per Capita** from the **Rows** shelf to the **Measures** shelf, which prompts us to rename it. Let's call it **Smoother Remittances per Capita**.
3. The original table calculation, as shown in the following screenshot, has one function, one expression, a start offset, and an end offset:

    - The `WINDOW_AVG` function tells Tableau Public that the field can be addressed and we want it to apply the `AVG` function to the values
    - The expression, in this case, is the field on which the calculation is operating
    - The start offset, by default, is two places prior to end offset and hence, it has the negative sign
    - The end offset, by default, is at the current value, as shown in

the following screenshot:



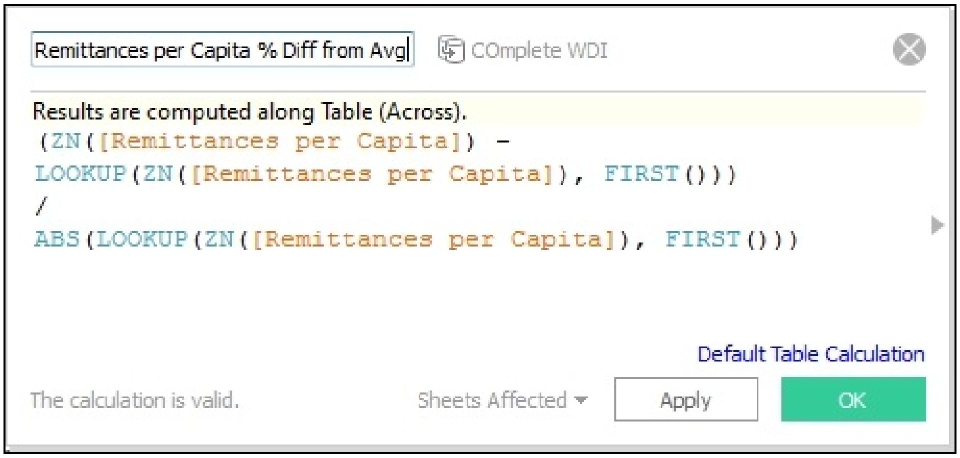
- Modify this calculated field to include three previous and three future values by replacing `-2` and `0` with `3` and `3` respectively, as shown in the following screenshot. You can also create an integer parameter and allow users to set the number. We will show you how to do that in the chapter on parameters:



The result is a calculated field that is much smoother than the original one.

In the second example, we'll calculate the percentage difference from the average **Remittances per Capita**. First, we will show you the original calculated field, and then, we'll show the changes that we made.

We duplicated the sheet on which we were working and then dragged the percent differences of the **Remittances per Capita** field to the **Measures** pane, where we renamed it to **Remittances per Capita % Diff from Avg**, as shown in the following screenshot:



In the formula box, Tableau Public calculates the value for each year, subtracts the value of the first year from it, and divides it with the value for the first year.

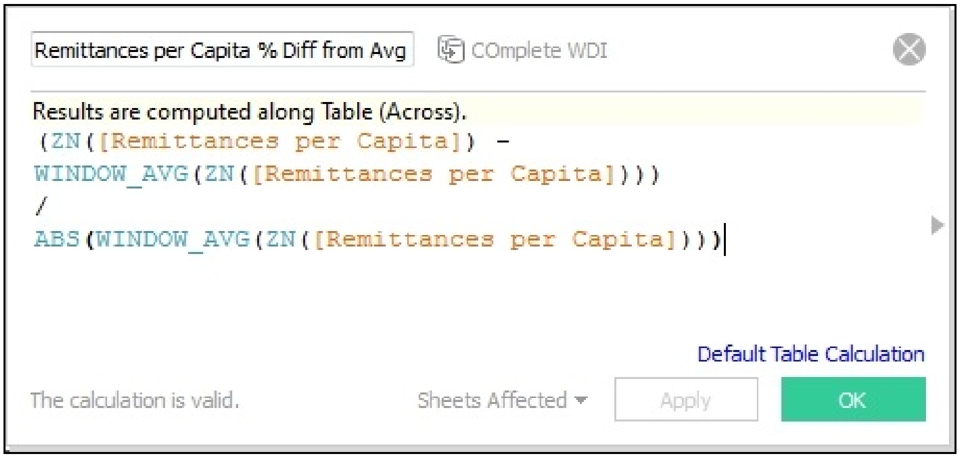
The following are a few new functions in this formula:

- **ZN**: This means that Tableau Public uses 0 if the value is null.
- **LOOKUP**: This finds a value specified by an offset from the current value. In this case it's `-1`, which means the previous value, but it could be anything. The other values are the **FIRST()** and **LAST()** functions.
- **ABS**: This takes the absolute value of the previous value. When writing table calculations, you should always use **ZN** and **ABS**, even if you think you don't need them at the time, because you do not necessarily know what will be in your data set in the future.

In order for this to be the percent difference from the average, change the references to the first value to the references to the **WINDOW_AVG**.

- Replace **LOOKUP** with **WINDOW_AVG**.
- Then, delete both instances of the string and the **FIRST()** function, because in this case, the **WINDOW_AVG** does not need an offset, like we had in the previous exercise. We want to use the average for the whole partition.

The new formula looks like the one shown in the following screenshot:


Remittances per Capita % Diff from Avg
COmplete WDI
Results are computed along Table (Across).
(ZN([Remittances per Capita]) -
WINDOW_AVG(ZN([Remittances per Capita])))
/
ABS(WINDOW_AVG(ZN([Remittances per Capita])))
Default Table Calculation
The calculation is valid.
Sheets Affected
Apply
OK

Ranking

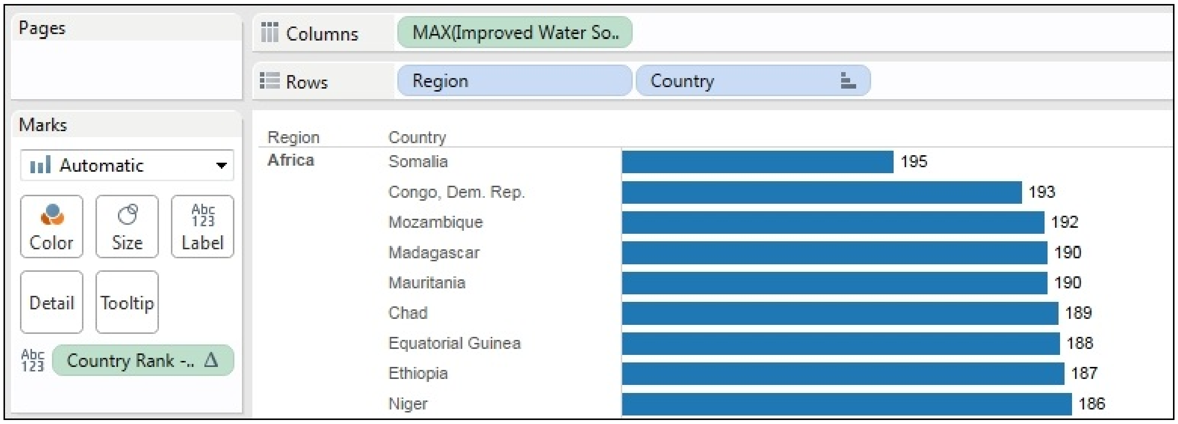
Our examples so far have focused on comparing measures that occur along a time continuum. Table calculations are useful for a comparison between different dimensions as well. Ranking by discrete numbers, as well as percentiles, is powerful. Many people want to know the top or bottom number of members in a dimension. In the next example, the final product shows the bottom 20 percent of the countries in each region according to their availability of improved water sources, and each country will be colored by its percentile of life expectancy.

Start off by creating a basic bar graph from the World Development Indicators data source that shows the maximum percentage of a country's population that has access to improved water sources. Drag **Region** and **Country** to the **Rows** shelf. Then, when we drag **Improved Water Source (%)** to the **Columns** shelf, we aggregate it as a maximum rather than a sum. Select **MAX** based on the assumption that unless there's a major natural disaster (such as floods and earthquakes, which were used as data points in the previous chapter) or war, it's unlikely that the access to clean water will decrease.

Then, sort the countries in the ascending order by clicking on the sort icon on the *x* axis.

The real ranking tasks begin here, with the label. The objective is to show the bottom 20 percent by region. So, we need to do several things along the way that help explain the ranking functionality. Perform the following steps:

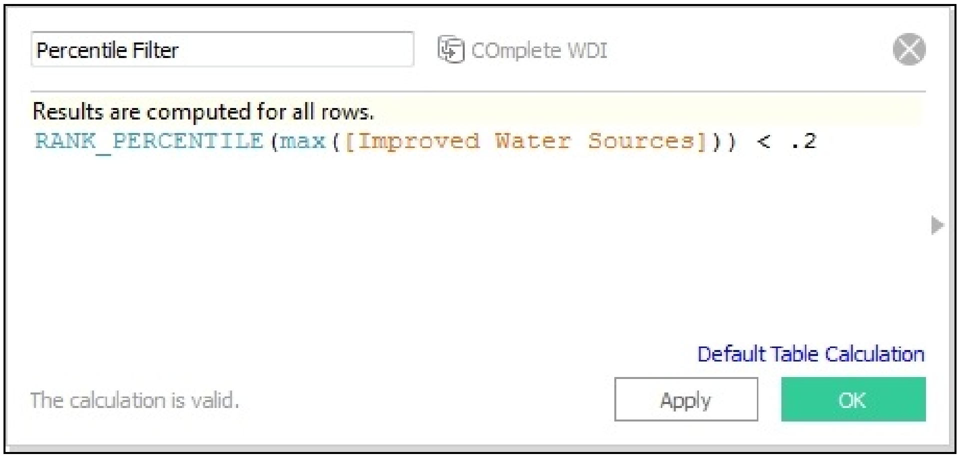
1. Drag **Improved Water Source (%)** to the **Label** shelf.
2. Then, click on it and select **Quick Table Calculation**.
3. The Quick Table Calculation that we selected is **Rank**. Then, drag the field to the **Measures** pane and rename it to **Country Rank – Water**.
4. Check out the snippet of the visualization in the following screenshot; each country is ranked in the descending order by the percentage of citizens who have access to clean drinking water:



- Note that the countries are ranked as a part of the whole table. We would like to rank them by region. Therefore, click on the pill, which is present on the **Label** shelf, select **Compute using**, and choose **Pane** (down).
- To the lower-right of the visualization, there is an indicator that we have several null values. This is because some of the countries have not reported these data points. Click on the indicator to exclude the countries with null values.
- There are other functions, such as `INDEX()`, that perform similar tasks, except that `INDEX()` actually produces the row number in the partition and not the rank. If you change the sort order of the countries shown, the labels stay the same; wherever you see Somalia, it will be ranked 53 in Africa.

The next task is to show the bottom 20 percent of the countries. Showing the top 20 percent would not add much value since the measure cannot exceed 100 percent of the population. In order to do this, we created a new calculated field called **Percentile Filter**, and told Tableau that we want to know whether the percentile of each country is less than 2. This means that it's in the bottom 20 percent of all the countries in its partition. Then, define the partition. The default is that the entire table is the partition, but we want each pane to be the partition.

The filter shown in the following screenshot has three possible results, namely true, false, and null, which makes it a Boolean field that shows up with a **T|F** icon in the data window. Since we excluded the nulls, we will not get null values:



In order to make sure that the total ranking of the countries in each region and showing only the bottom 20 percent, perform the following steps:

1. Click on **Default Table Calculation**.
2. Select **Country**.
3. Click on **OK**.
4. Click on **OK** again.
5. Drag the field to the **Filters** shelf.
6. Select **TRUE**.

You can also allow users to see the percentage of their preference by creating a parameter with a data type of float and a rank from 0 to 1.00 that replaces the hard-coded `.2` in the formula, as shown in previous screenshot.

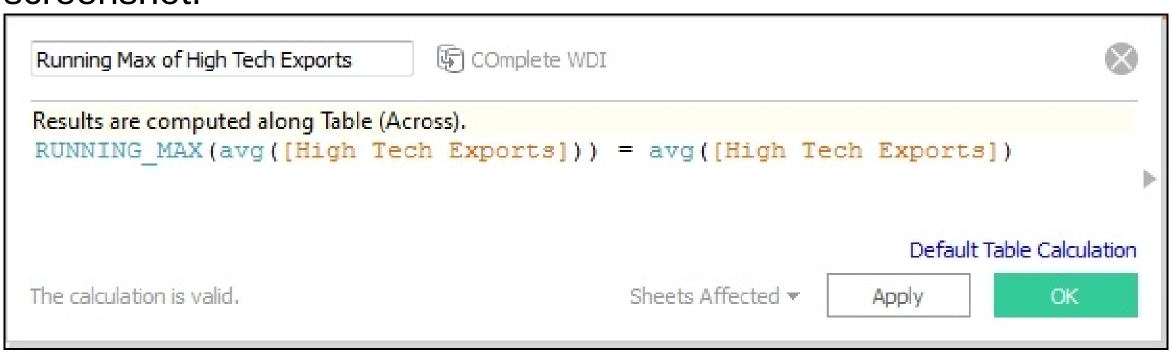
There are several functions within Tableau Public for ranking, and except for `RANK_PERCENTILE`, the other functions vary only in the way they use unique or duplicate ranking values.

Window versus running functions

We have briefly discussed the `WINDOW` functions already. There's also another type of function, called the `RUNNING` function. The `WINDOW` functions have specific partitions, that is, they measure either the whole table, pane, or cell, or from a specific number of previous or future values. Alternatively, the `RUNNING` function compares all the values before the current value in the partition.

In the following example, we graphed the average percentage of the GDP that was from high-tech exports by region and by year in the World Development Indicators data source. This is a simple spark bar graph. We have hidden the header for the *y* axis in this graph.

We created a calculated field that, for each year, tests whether the average **High Tech Exports** as a percentage of the GDP is equal to the running maximum, that is, is the value for any year higher than all the years before it. The results are true and false, as shown in the following screenshot:

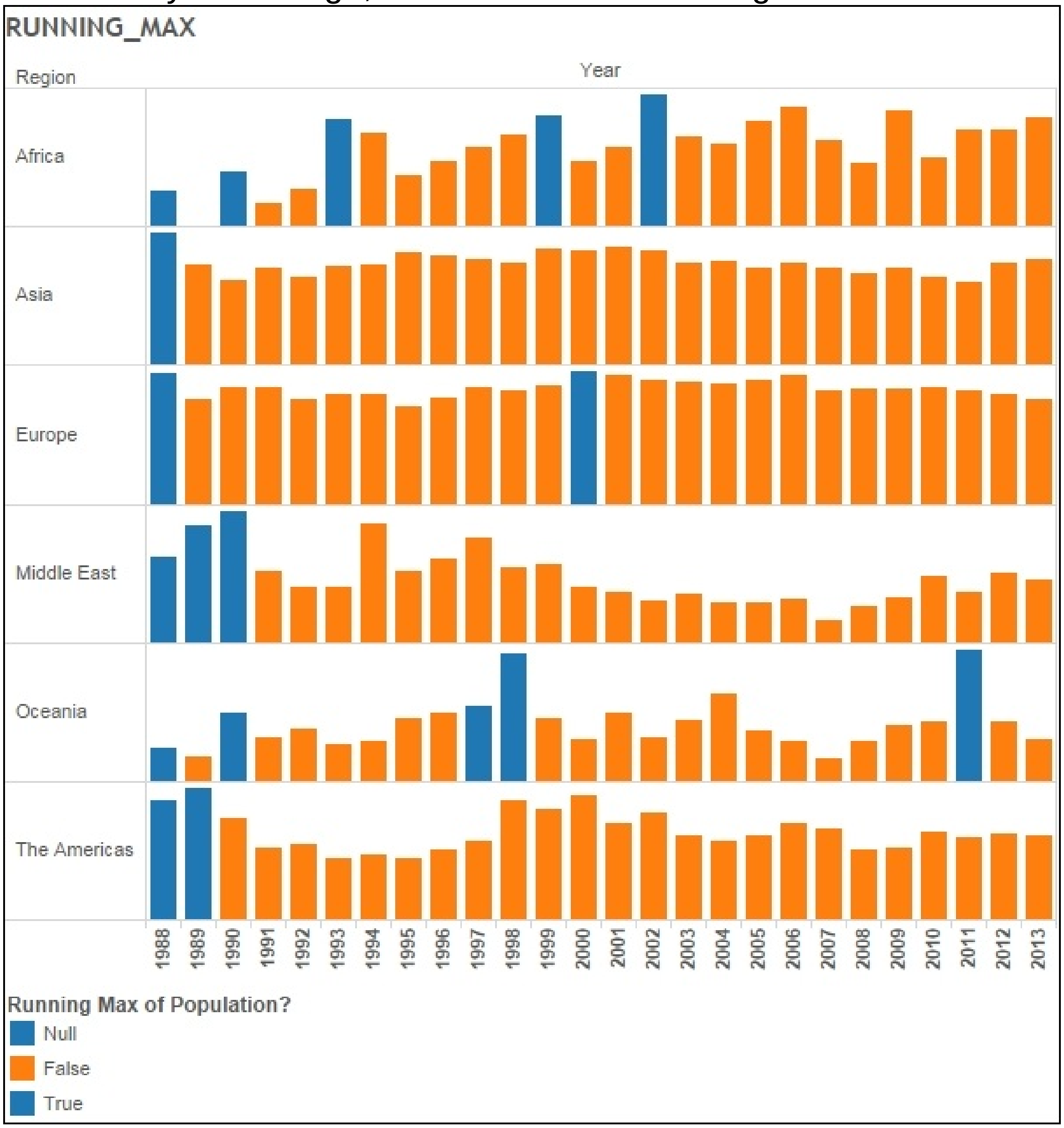


The `RUNNING_MAX` function, which is one of several `RUNNING` functions, does not have the start and end offsets like the `WINDOW` functions. However, you still need to address it.

A note on addressing

The following table calculation computes across the pane. You can also set it to **Compute using Year**. Year is the field along the *x* axis. Addressing the table calculation across the pane tells it to compute across the *x* axis and then to use other fields on the visualization, such as **Region**, as the partition or group. So, this field is addressed across year, and the calculation is performed for each **Region**.

Add the **Region** field to the **Color** shelf in the visualization and then format the true values so that they are blue, and format the false values so that they are orange, as shown in the following screenshot:



The level of detail calculations

The **level of detail calculations** (LOD) enable you to tell Tableau Public at exactly which dimensional level you would like to aggregate without having to place that dimension on the visualization. Additionally, you can include or exclude dimensions in the calculations and create calculations that include all the underlying data.

The following are the three types of LOD calculations:

- **FIXED**: This computes a value for the dimension that you specify
- **INCLUDE**: This computes at a dimensional level and is not included in the visualization itself
- **EXCLUDE**: This excludes a dimension that is a part of the visualization

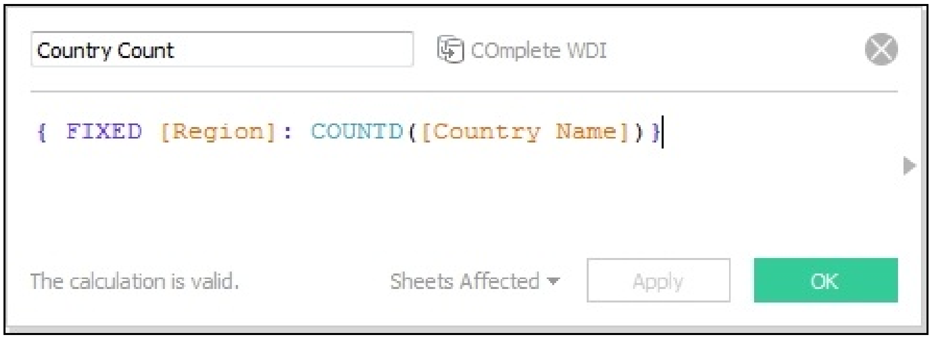
LOD has a big caveat. While you can perform table calculations on them and create aggregations and functions within them, you cannot include table calculations in them.

A FIXED LOD calculation

Each LOD calculation has the following three features that are different from those that we have discussed before:

- The **LOD** expression type: In this case, the LOD expression type is **FIXED**
- The **Dimension** on which the calculation is operating (you can add dimension levels, which should be arranged in the increasing order of granularity by separating them with a comma): In this case, it is **Region**
- The **Aggregation**: In this case, it is **COUNTD([Country Name])**

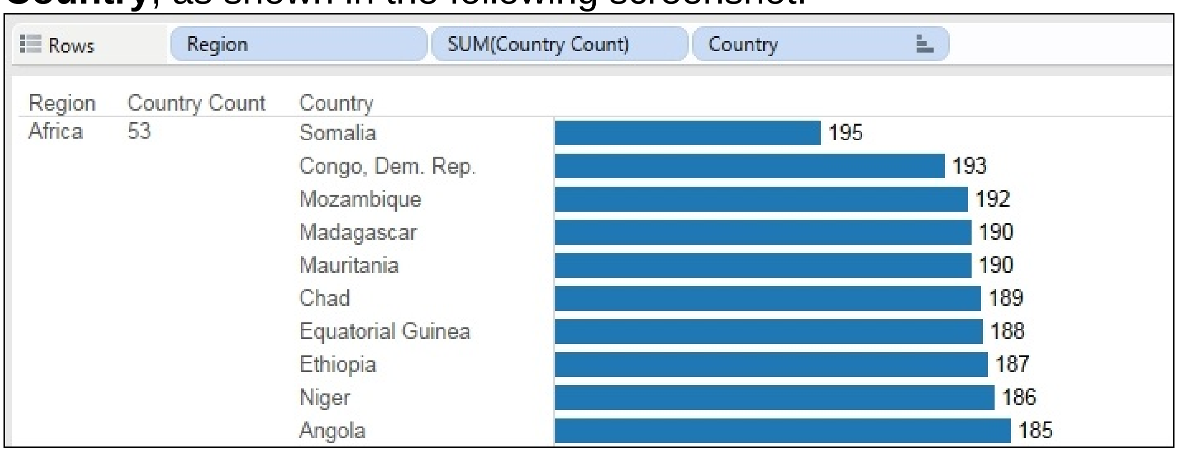
Note that the entire expression is enclosed in curly braces, as shown in the following screenshot. The LOD calculations are the only instances in Tableau Public where you will use curly braces:



This field provides us with the unique number of countries in each region, even though 80 percent of these countries are filtered out. This provides great context.

In the following example, we will discuss:

1. Add this field, which is a measure, to the **Rows** shelf to the right of **Country**.
2. Then, click on its **Context** menu and select **Discrete** instead of **continuous** so that it becomes a dimension.
3. Drag the field so that it's to the right of **Region** and to the left of **Country**, as shown in the following screenshot:



This visualization shows us the bottom 20 percent of the 53 countries in Africa in the data set. It would be even more helpful to know what the overall access to improved water sources is in each region and not just by country.

The following are a few issues that we need to overcome with virtualization:

- The data is not necessarily set up for this.
- The granularity of the data is at the country level, and we only know the percentage in each country that has access to clean water. We don't know the discrete number.
- We do not have the totals by region. We need to determine the overall number of people with access to clean drinking water, and then we need to divide this number by the population of each region.

The INCLUDE and nested LOD calculations

In the following example, we will show you how to solve some problem by using an INCLUDE LOD calculation nested within a FIXED LOD calculation, and each will be aggregated. This is an issue that most people have had to solve by using data sources with different levels of aggregation, which takes a lot of time to develop and render in Tableau Public.

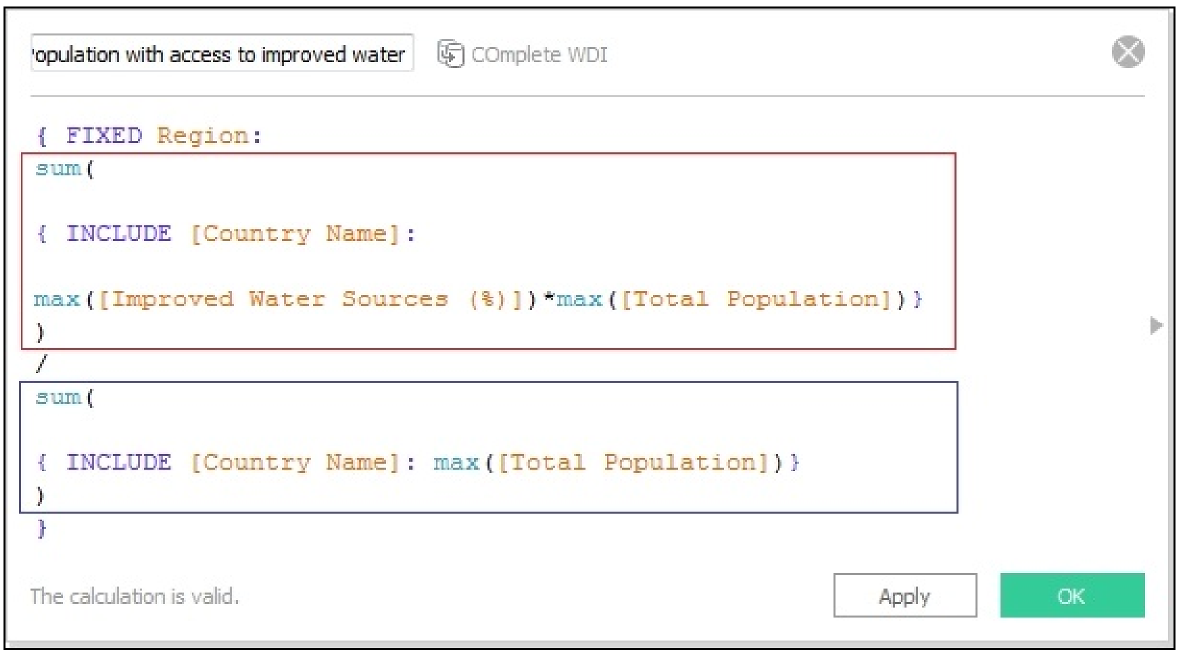
Create a new field within the World Development Indicators data source and name it **Population with Access to Improved Water**. It's kind of a long name, but we want to ensure that no one misunderstands the metric.

So, here's our approach to calculating the overall percentage of each region with access to clean water:

- The primary LOD type is **FIXED** at the **Region** level.
- The numerator (enclosed in the red box shown in the following screenshot) adds up the total population that has access to clean water.
- The secondary LOD type is INCLUDE at the **Country Name** level. We are using INCLUDE because we want to roll up the country-level data to the region.

Rationale: For every country, we need to know the total population that has access to the sources of clean water. We have the percentage of the population. So, we need to multiply the maximum percentage with the total population to get the number of people who have access to clean water.

- The denominator (enclosed in the blue box in the following screenshot) is the total population. We need to know the total population by region. So, we add up the total population of each country.
- We will use the `INCLUDE` function at the **Country Name** level and then use the `MAX` aggregations. The granularity of the data is at the **Year** level. We are not including Year, because that would produce the population at the **Country Name and Year** level. We just want the total population of each country, as shown in the following screenshot:

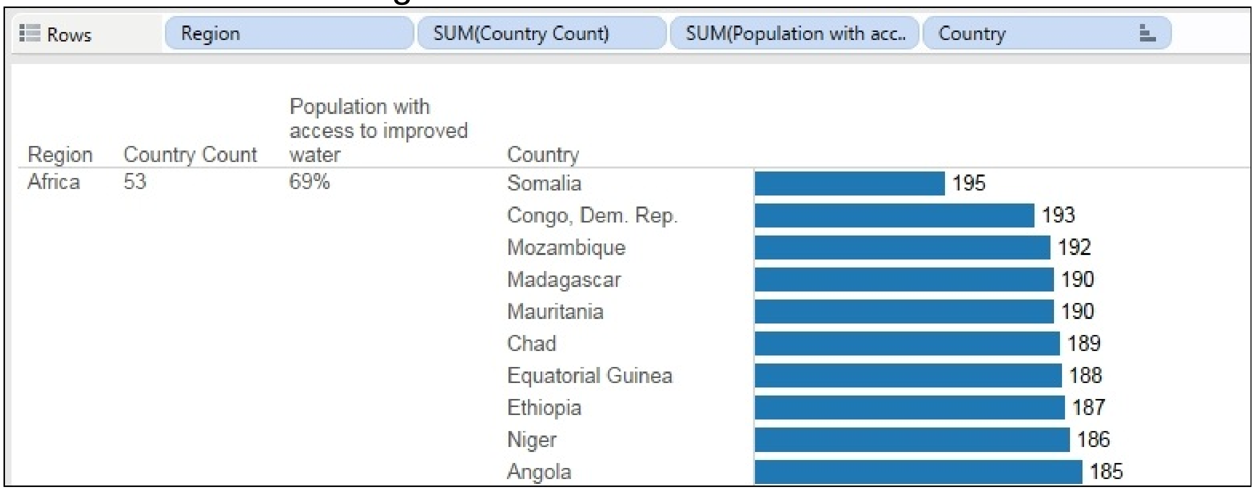


This is a long field, but it accomplishes what has been difficult before, namely aggregating at multiple dimensional levels in one calculation without duplications, and aggregating records that are not a part of the visualization.

Add this field to the **Rows** shelf of the previous bar chart, turn it into a discrete field, and drag it to the left of **Country Name**, just as we did for **Country Count**.

Also, right-click on the new calculated field, click on **Default Properties**, and then select **Format** so that we can set the number format as a percentage.

Here's the finished product. Only 69 percent of the people in Africa have access to clean drinking water:



Summary

In this chapter, you learned how to create quick table calculations to add depth and context to your visualizations. You also learned how to edit table calculations, how the different types of table calculations function, and how to address table calculations and create them manually. One of the important features of data analysis is the comparison of a data point to the others around it so that users understand the context of its performance. You learned how to create a basic, and advanced level of detail calculations to aggregate data points that are not visible, and to aggregate at multiple dimensional levels as well.

In the next chapters, which focus on mapping and parameters, we will discuss using many of the concepts that you learned here to give control to your users and illustrate compelling, rich stories with contextually curated visualizations.

Chapter 7. Dashboard Design and Styling

The purpose of a dashboard is to tell a story. In the scheme of human history, hieroglyphs, the Bible, epic poems, theatre, comic books, and even social media are a few of the many different media through which we recount events to each other. All stories have one of the several primary purposes, to teach values or historical lessons, create culture, motivate people to take action, or entertain, to name a few.

Knowing the purpose of the story that you are telling is imperative to making it memorable. Good stories have several important attributes, which should be kept in mind when you are working. Some of these attributes are as follows:

- They're simple and focused
- They create an emotional response
- Listeners can relate to the actors in them
- They have good, real data

It's a good idea to take a few minutes to think about the story that you want to tell before you even create a new dashboard. Think about the stories that are the most memorable to you and the data points in them. Then, put yourself in the place of your listener. Design with your listener in mind. If you can develop a list of the top three data questions that you think are most important to your listener and then design a dashboard that focuses on answering them simply and evocatively, then your dashboard will likely be impactful.

With Tableau Public, you can tell stories in a variety of ways—with individual visualizations, well-formed dashboards, and story points. You can also integrate your work in Tableau Public into other tools by using the JavaScript API (which is beyond the scope of this book). But just because there are many tools and options at your disposal, it does not mean that you should make your dashboards overly complicated. It's actually just the opposite. Good journalism focuses on keeping stories simple, and you should use the same ethos with your stories. *Just because you can do something doesn't mean that you should do it*.

It truly applies well to designing dashboards; less is more. Knowing your user and their questions is fundamental to a good design, and everything else flows from that.

In Tableau Public, you can make a dashboard from one or more visualizations, and you can add images, free text, filters, legends, parameters, links, and even embedded web pages. All of these capabilities are designed to enable you to create a data story that has the right context for your dashboard, both for the consumer and for its relativity to other events and data elements. While you can publish individual worksheets to Tableau Public, we encourage you to organize your work into dashboards so that you can guide your user through a complete story.

In this chapter, we will discuss some best practices for dashboard creation, the elements of a dashboard, and how to navigate the dashboard worksheet. The following topics will be discussed in this chapter:

- The design process
- Best practices of dashboard design
- The elements of a successful dashboard
- Creating a dashboard
- Adding context with titles, images, links, and tooltips

In the next chapter, we will discuss dashboard filters, actions, and parameters, and after that, we will discuss publishing your work to Tableau Public.

The dashboard design process

The purpose of Tableau Public is to help you tell data stories interactively. Dashboard design, like view design, should tell a concise story that flows from one element to the next. When you are first starting your journey of dashboard design in Tableau Public, you should expect it to be an iterative process in which you design dashboards and revise them over time, often with input from other people. You might also find it helpful to draw out dashboards before building them in Tableau Public. Leave room and time for experimentation when building your dashboard.

You may find that more views, or different views, are needed in order to tell your story. Pulling various views together into a dashboard sometimes exposes flaws in the initial design. In that case, go back and change or add to your views, and then come back to the dashboard design process.

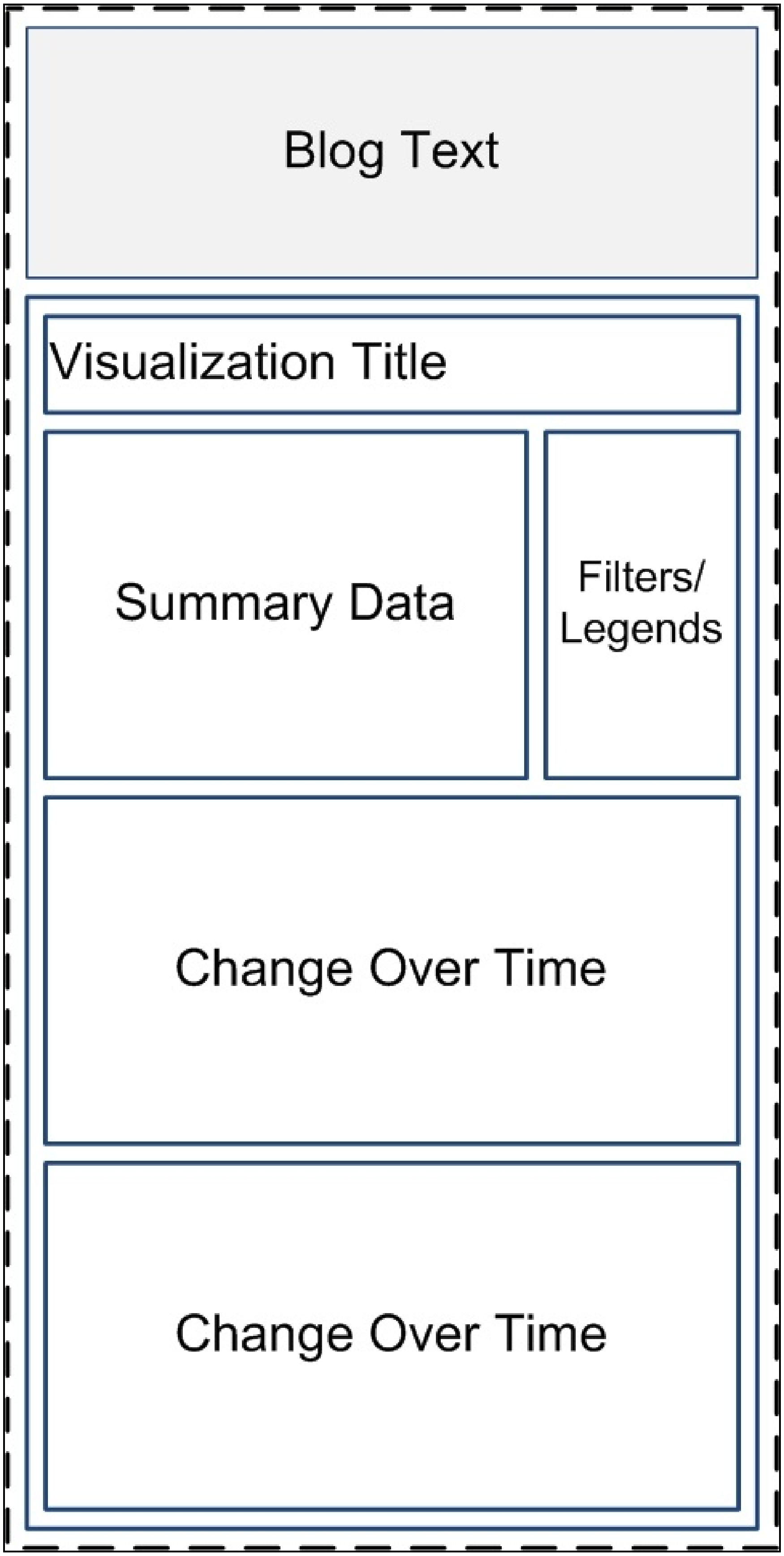
Best practices for dashboard design

Dashboard design should highlight data and decoration should be limited, to add to a reader's understanding. Much has been written about dashboard design for data visualization. The following are some key points that you should consider when designing dashboards:

- **Keep it simple**: Your dashboard needs to answer only one question. If you can construct it so elegantly that it answers several questions, that's great, but according to *Stephen Few*, the maximum number of data points that any of us can remember at once is three.
- **Keep in mind the audience**: Consider what devices will be used to read or consume the data story and content.
- **The method of consumption**: Will users interact with virtualization on their computer, mobile phone, or tablet? Size your dashboard accordingly.
- **Performance**: Like any other web application, the amount of time it takes for a consumer to get what they need is a major factor of performance. Design your visualizations and data sources so that they perform efficiently.
- **Start big and end small**: Always place the most aggregated (macro) data points or summary metrics on the upper left, and guide users to granular, actionable data points.
- Use **colors**, **graphics**, and **fonts** that are appropriate for your subject matter and consumers. Choose a simple, single color or complementary palette colors. Consider red-green color blindness and improve accessibility. Remove all non-data ink except labels, limited titles, and bare-bones instructional text. Additionally, you can create your own custom color palettes in Tableau Public fairly easily. Use labels, titles, and simple instructional text for the dashboard, and format tooltips so that they add context and calls to action for your users.

For the dashboard that we are using in this chapter, we created a rough wireframe in Microsoft Visio. We'll refer to it throughout this chapter so wireframe in Microsoft Visio. We'll refer to it throughout this chapter so that you can see how the functional elements that we are discussing will be included in the final design. It's fairly common for people to draw out the designs that they would like to build in Tableau Public. Also, it's a great way to keep your objectives in perspective when laying out the dashboard in Tableau Public.

Our sample dashboard, which will be included in a blog post at www.dataviz.ninja, follows the **Rule of Thirds**—after adding the title to the dashboard, the three primary visualizations, namely aggregated data, change over time, and granular details, will consume about one-third of the total space, as shown in the following screenshot. We have also provided space for filters and color legends, and we will continue to use this dashboard design concept in the following two chapters, which cover filtering and publishing:


Blog Text
Visualization Title
Summary Data
Filters/
Legends
Change Over Time
Change Over Time

Creating a dashboard

Creating a Tableau Public dashboard is a process of adding the various elements (starting with worksheets), onto the dashboard canvas, adding titles, configuring filters, titles, and layout, and then modifying the behavior of the dashboard.

Now, we will build a dashboard from simple visualizations that were created by using the Climate Change data that was published by the World Bank and which is available at http://data.worldbank.org/data-catalog/climate-change. It's structured in a way that is similar to that of the other data sources that we are using. You can download the sample dashboards from https://public.tableau.com/views/Chaper7-Dashboards/CO2EmissionsDashboard. We will use the data sources that we have discussed in the previous chapters to build sample dashboards so that you can download the source files to practice.

Our workbook includes a summary map of carbon dioxide emissions by country in 2008, as well as a line graph that shows change over time as regards the total emissions and emissions per capita by region. Lastly, it contains a heat map that shows the foreign aid distributed for generic programs by the United States by year.

The objective of this section is to build a dashboard that allows users to navigate from high-level areas of interest, that is, outliers, to more detailed data to identify causation.

To start the dashboard creation process, click on the dashboard tab at the bottom of the Tableau Public interface. The dashboard tab is the tab with an icon of a rectangle divided into four quadrants, which is outlined in the following screenshot.

You can also browse **Dashboard** | **New Dashboard** from the **Menu** bar:

When you select the dashboard tab, the dashboard view is displayed. This is the interface that allows you to build and configure dashboards from the worksheets that you have already created in this workbook. To the left of the window are the available worksheets. From the left side of the window, you can also add to the dashboard various supporting elements such as text boxes, labels, web page windows, and images. To the right is the canvas area of the dashboard, where sheets and objects are dragged to, arranged, and configured.

The dashboard tab interface

The dashboard interface is similar to the worksheet workspace in many ways. In the following screenshot, we have identified its several key areas.

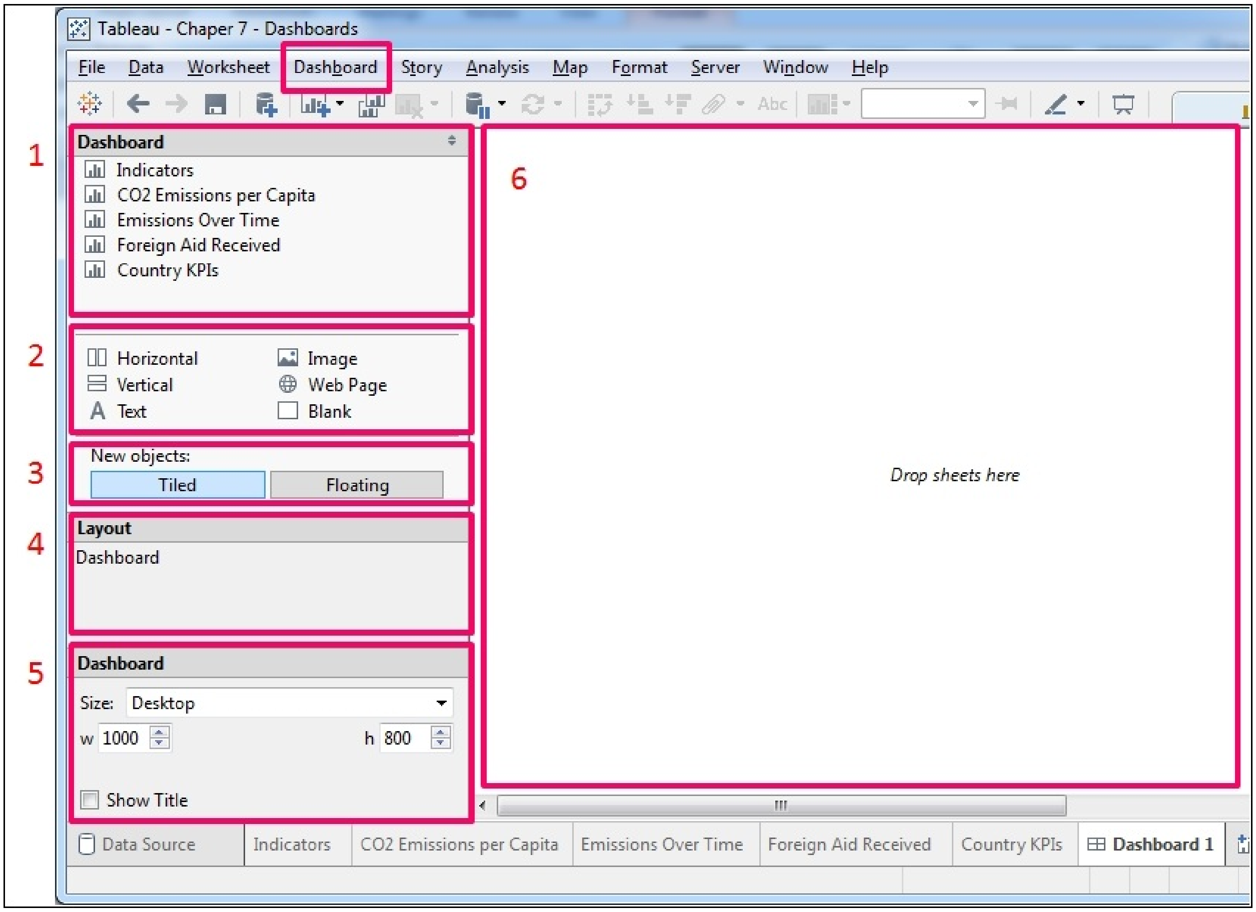
Like the worksheet view, the dashboard has the following elements:

- The menus and toolbar, which are placed horizontally across the top of the page
- The **Dashboard** pane on the left, where you can determine the exact composition of your dashboard
- The canvas, where you will drag worksheets and other visual elements

We have numbered the elements on the dashboard, as shown in the following screenshot. They are as follows:

- A list of worksheets **(1)** that have not been hidden and which are sorted according to the order in which they are arranged in your workbook
- Containers and objects **(2)** that can be dragged onto your dashboard to increase the integrity of the design
- A controller **(3)** for new objects
- A hierarchy of dashboard objects **(4)** that can be used for navigation
- A controller **(5)** for the dashboard size
- The canvas **(6)** where you will compose the dashboard

If you are creating a new dashboard, the canvas will be blank. If you open an existing dashboard, the various existing dashboard elements will be displayed on the canvas. The following screenshot shows the dashboard interfaces:



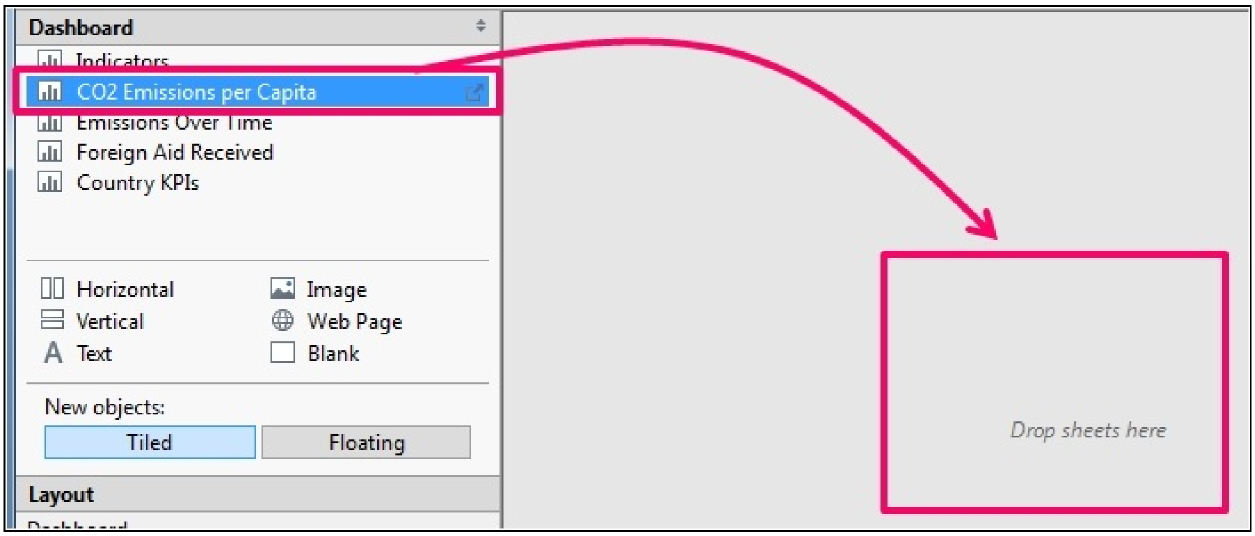
It's important to keep in mind the purpose of your dashboard. We are building a dashboard that tells the story of carbon dioxide emissions over time. The dashboard should allow users to navigate to the areas of interest and then understand the factors of consumption.

The dashboard will have the following key elements, which were previously discussed:

- Title, so that our users know what they are viewing
- At least one visualization with data and proper context is required so that different data points relate to each other and to the user
- A call to action (conclusion) is also needed so that users know how to explore the details within the dashboard or go to another web page

In order to build this dashboard, we will start with a single visualization, which is the visualization with summary data in this case. Drag a sheet onto the canvas by hovering over the visualization that you want to add onto the canvas by hovering over the visualization that you want to add, pressing the left mouse button, and dragging it onto the canvas, as depicted in the following screenshot. You can also double-click on the worksheet name.

If you wish to remove a worksheet from the dashboard, click on the title of the visualization until a grey container bar appears at the top border of the visualization. Then, right-click on the grey bar. Select **Remove from Dashboard** from the shortcut menu:



The result, as shown in the following screenshot, shows the map that we created. This map shows the CO2 emissions per capita in 2008, which was taken from the World Bank official website. Each country is colored by its relative percentile in the data set, and the label shows the CO2 emissions per capita.

However, users don't know all of that yet:

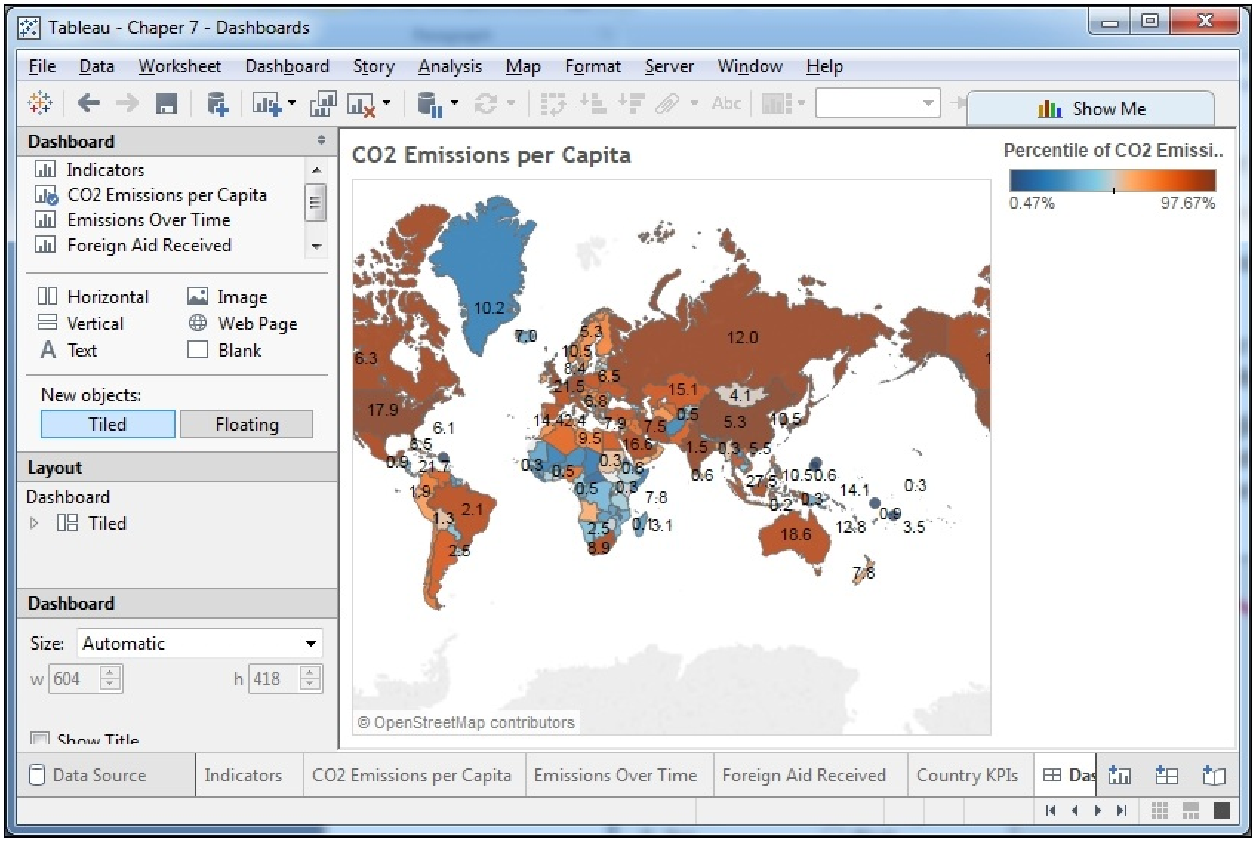


Tableau Public automatically added the following two elements:

- The worksheet name as the title
- The color legend in a vertical layout container to the right

In order to compose an empathetically designed user experience with proper context, we need to use layout containers to rearrange the dashboard and create a flow of both navigation and information, that is, a data story. The next section discusses layout objects, and we will show examples of how to use them.

Layout objects

Tableau Public has several different layout objects that you can add to your dashboard to control the composition.

Objects are versatile, and you can use them to reflect the overall theme of your graphic composition. From the context menu of an object, you can perform the following tasks:

can perform the following tasks:

- Format it with borders and background colors
- Remove it
- Set it as a floating or tiled object
- Deselect it
- Remove it from the dashboard

You can set the height and width attributes for floating objects, just like you can for worksheets that you have added (and for the dashboard as a whole, which is highly encouraged).

The following objects can be added to your dashboard:

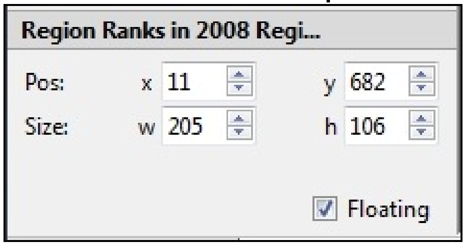
- Horizontal layout containers: You can drag worksheets that you want next to each other into these containers.
- Vertical layout containers: You can drag other objects and worksheets that you want to stack from top to bottom into these containers.
- Text objects: These can be used to add titles and calls to action. Though you can't add field tokens to text objects, you can add parameter tokens, which will be discussed in depth in the next chapter.
- Images: You can use these to browse logos or branding elements that add richness to your dashboard. You can also use images as links to web pages, since each image can have a URL attribute.
- Web page objects: These can be used to add content from the Internet. You can even create dynamic content, which will be covered in the next chapter too.
- Blank objects: These objects can be used to control space. Blank containers are transparent. So, the background colors that you are using will show through them.

Note

In the section of the dashboard pane that says **New objects**, the default object is set to **Tiled**. You should leave it there until you are comfortable with dashboard design. Each new object that you add to your dashboard will have this setting, and though you can easily change the attribute for individual sheets on a dashboard, it's best to keep things simple in the beginning. The **Tiled** automatically aligns objects to the dashboard grid, and floating adds new objects to the dashboard that are detached from the dashboard grid.

When you add objects, like worksheets, into a horizontal or vertical layout container, the widths and heights of the objects will be sized automatically, unless you specify otherwise. In case you're using actions, which will be discussed in the next chapter, it's particularly useful to use containers for sizing automatically.

Setting the size of dashboard elements

Usually, dashboard elements such as containers can be set by manipulating the border handles of each container and resizing them to a suitable size. Tableau Public also offers a **Size** feature, as shown in the following screenshot, that allows for more control in the positioning and sizing of these dashboard elements: 

> ## Note
>
> The dashboard element must be in the floating mode so that it can be sized and positioned. In case the container or element is not already in the floating mode, it can be set to floating by clicking on the dashboard element and checking off the **Floating** checkbox, as shown in the previous screenshot.

To change the positioning of the dashboard element, enter the **x** and **y** coordinates (this will take some trial and error) in the corresponding **Pos:** fields, or scroll up and down with the arrows for fine correction. To change **Size**, enter the width (**w**) and height (**h**) values in the corresponding fields using the up and down arrows for fine adjustment.

In the **Size** section of the Tableau Public controls on the left side of the dashboard, you can also choose to show the title of a particular chart by selecting or deselecting the **Show Title** checkbox. You must first select the dashboard element either in the dashboard, or in the **Layout** section of the Tableau Public controls. To show or hide the title of the dashboard, you can click on the **Dashboard** field in the **Layout** section and select or deselect **Show Title**. This may also be done, as described earlier in this chapter, by using the **Dashboard | Show Title** menu command.

Sizing the dashboard

Tableau Public presents many size options for the entire dashboard. This feature helps you select a size either for a blog or other website, or optimized for a tablet computer. To set the dashboard size, click on **Dashboard** in the **Layout** section of the Tableau Public controls.

There is also an option to resize the entire dashboard. For example, you may find that scrollbars appear to the right, or at the bottom of the dashboard, that you don't want to see. Alternatively, you may prefer matching the dashboard to a standard web page or computer desktop size. Again, in the **Layout** section, select **Dashboard**, or ensure that no worksheet or other elements are selected on the dashboard itself. The **Size** section presents a drop-down menu with a variety of dashboard size options.

Many sizing options are available for selection depending on your needs. **Blog**, **iPad**, **Laptop**, **Desktop**, **Exactly**, and **Automatic** are some of the typical settings. It will take some trial and error to properly size a dashboard for your needs.

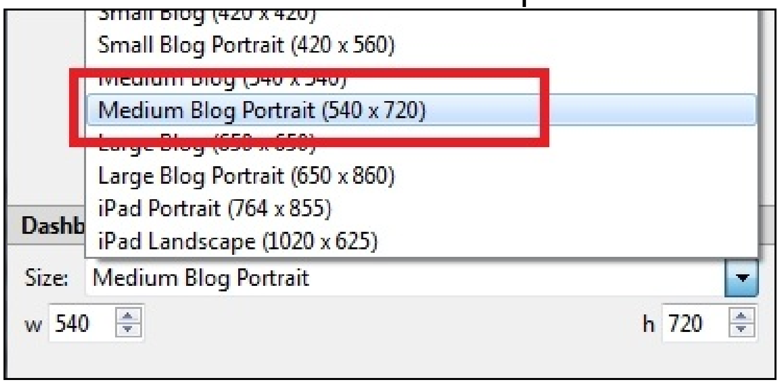
It is important to consider the methods in which you will be publishing your dashboard. If you are publishing your dashboard to your blog, you need to make sure that it properly fits into the layout that you have selected.

Building a dashboard

In order to build a sample dashboard, which has a title, aggregate data, change over time, detailed data, and proper context, as shown at https://public.tableau.com/views/Chaper7-Dashboards/CO2EmissionsDashboard.

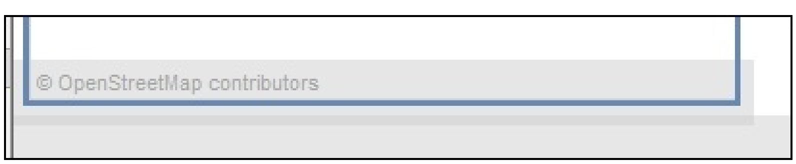
Perform the following steps to build the dashboard:

1. Create a new dashboard by clicking on the dashboard creation icon in the filmstrip.
2. Name it **CO2 Emissions Dashboard**; the name is important because it will be part of the URL for the workbook.
3. Size the dashboard so that it fits a blog, as shown in the following screenshot. This size fits in a typical WordPress blog, and it is also narrower than most mobile devices and laptops. Therefore, the size will work well for users of multiple devices:

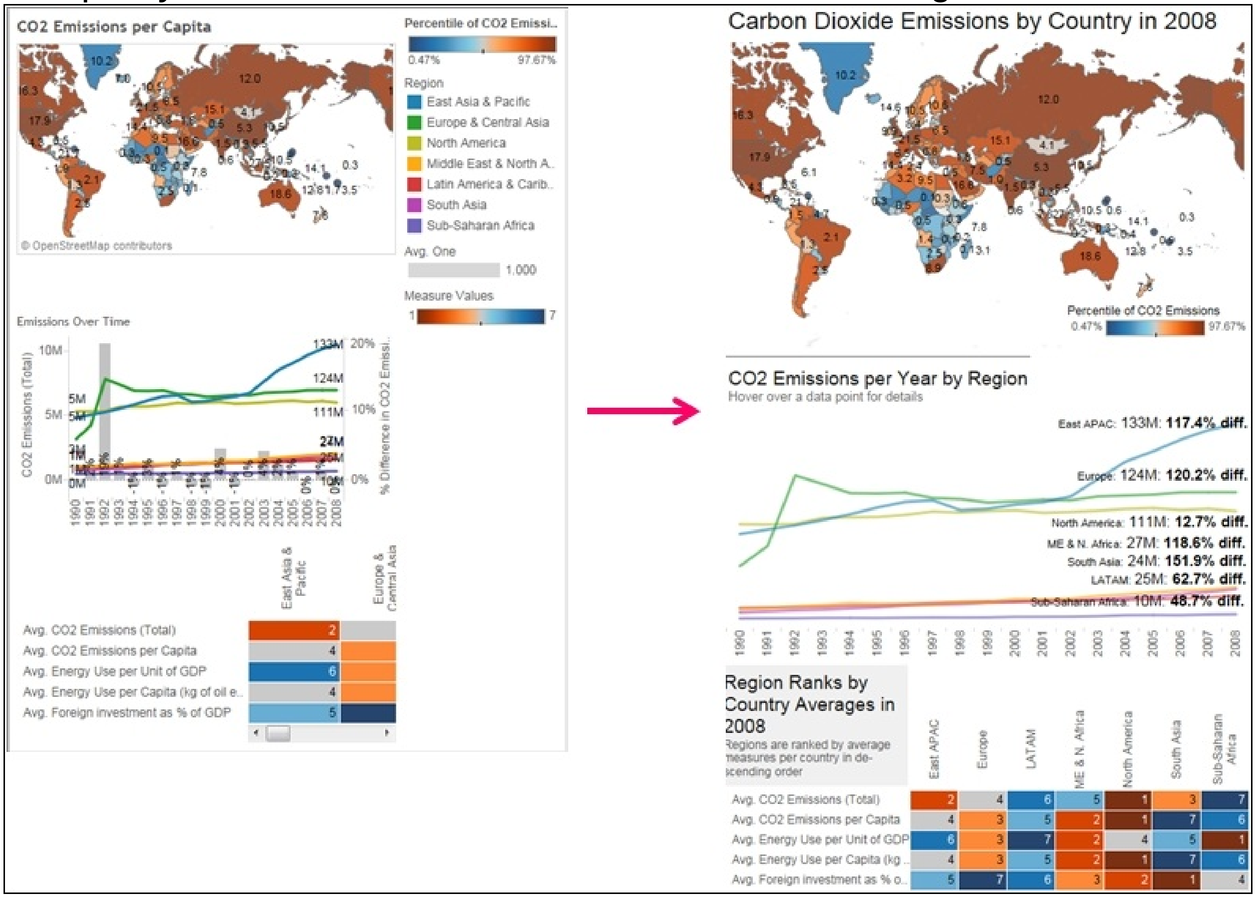


- We added containers knowing that we wanted to have a title, aggregate data, supporting details, and conclusions. We added a vertical layout container so that our visualizations and objects will stack on top of each other. The following steps will guide you to add data-map in the layout container:

1. We dragged the aggregated data-map of **CO2 Emissions per Capita** into the vertical layout container.
2. We then dragged the next worksheet, **Emissions Over Time**, into the very bottom of the vertical layout container so that the worksheet and the map would be automatically sized.
3. We dragged the worksheet from the **Dashboard** pane to the bottom of the container and dropped it when we saw the wide gray border, as shown in the following screenshot:



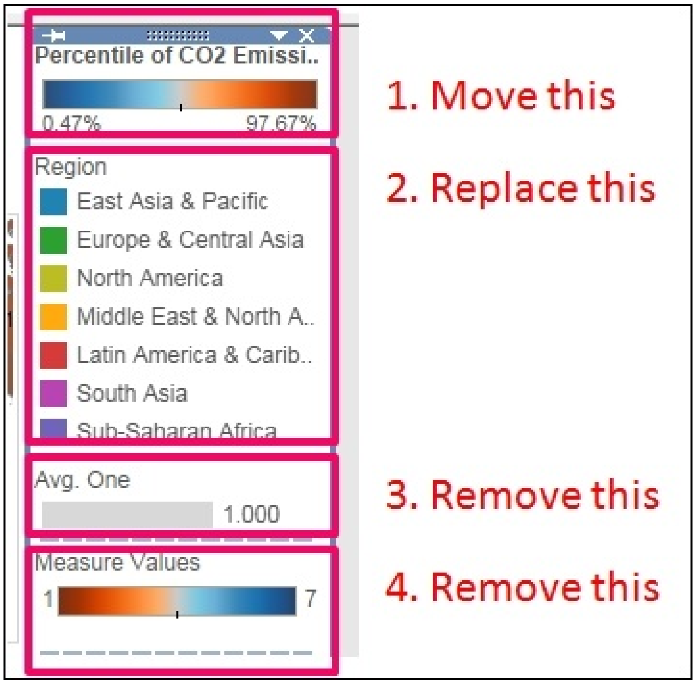
- We assessed that the dashboard shown in the following screenshot isn't pretty. It still needs a title, and it should be rearranged:



- We added a title before rearranging it, by performing the following steps:

1. Drag a `Text` object to the very top of the vertical layout container above the map.
2. Populate the title (**Carbon Dioxide Emissions by Country in 2008**) and then format it in 18 pt black Arial, which is wide enough to fit in the screen.

- The layout needs some work. Color legends need to move closer to their respective visualizations, and we need to change it so it can be fitted in some visualizations, as shown in the following screenshot:



- Move the **Percentile** color legend for the map by performing the following steps:

1. Click on it to select it.
2. Click on its context menu.
3. Change it to floating.

- Move it below the center of the map (we will move it again later).
- We decided to remove the color legend for **Region** altogether.
- Remove the size legend for **Avg.** Once we have actually used a bar in the heat map; you can check it out by clicking on it.
- We removed the last color legend, which is for the heat map. The colors are used to indicate rank, and they follow the same color scheme as that of the map (brown is bad, and blue is good).
- Adjust the fit of the containers as:

    - The heat map needs to be fit differently. Therefore, click on its context menu, select **Fit**, and choose **Fit Width**.

- Next, add titles by performing the following steps:
- Add a text object above the line graph; the title should be highly descriptive. It tells our users exactly what they are looking at, and it also tells them how to get more details. We used 14 pt black Arial font for the

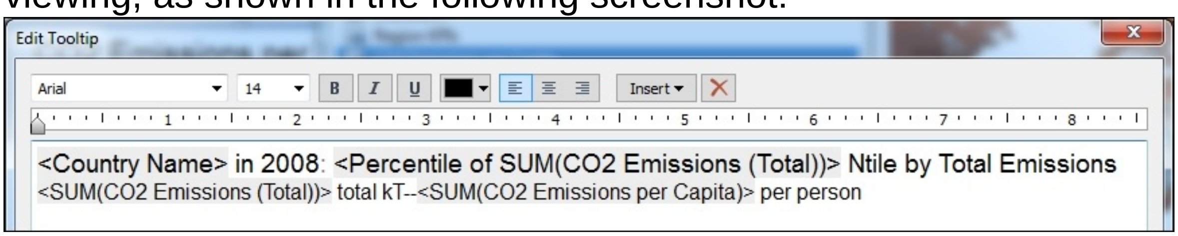
main title, and the call to action (that is, the text that tells the user what to do) is in 10 pt gray Arial, as shown in the following screenshot:

> ## Tip
>
> You can also add an image above the line graph. We added an image of a simple gray line that we created in Microsoft Paint and then saved as a bitmap.

- We also added a floating title for the heat map because we had extra horizontal space but less vertical space. We used the same fonts as that of the previous title and added a light gray shade so that it's visually separated from the graph itself.

- Add more context with tooltips and labels; it's important for users to know exactly what they are looking at, how the data points relate to each other, and how it relates to them. You can modify the tooltip for the map by performing the following steps:

1. Click on **Map Visualization** to select it.
2. Click on the **Worksheet** menu.
3. Select **Tooltip**.
4. Modify the contents and font so that they add obvious context and do not require users to do a lot of work to figure out what they are viewing, as shown in the following screenshot:



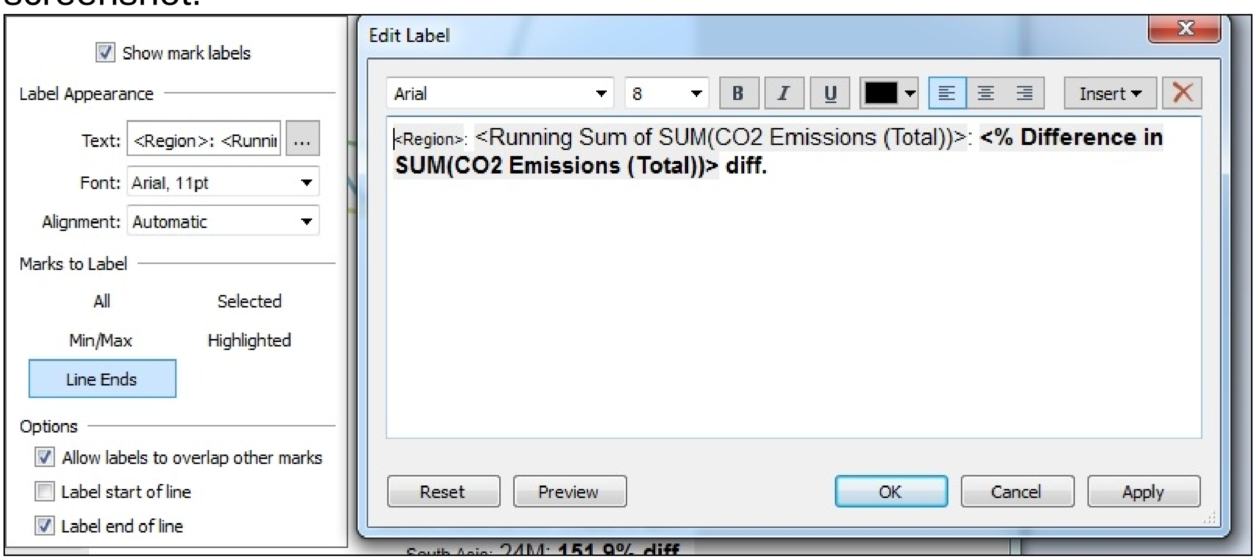
5. Repeat these steps for other visualizations. Make sure that you use the same font face and size in each tooltip. Consistency of design is critical.
6. Modify the data labels, particularly for the line graph.
7. Navigate to the worksheet, and place the appropriate fields on the Label shelf in case they were not there already.

- Format dashboard so that the right emphasis is created.

Change their location relative to each other on the dashboard so that they are all legible by clicking on individual labels and then dragging them elsewhere.

We felt comfortable making custom placements because we do not have a filter on the dashboard. (We will discuss filters in the next chapter.) We are confident that, with a specific size of the dashboard, the labels will always appear where we placed them.

If you want to undo custom placements, you can right-click on the label and select **undo the custom location**, as shown in the following screenshot:



- Adjust the size of the dashboard to 540 x 900 pixels because it appears compressed. Unfortunately, this means that we'll have a vertical scroll bar when not using a studio monitor, but we needed to make sure that the visualizations were legible. It's better to have a vertical scroll bar than a cramped dashboard. If no one can read it, it's a waste of time and effort.

Before publishing the dashboard, right-click on every worksheet in the filmstrip and select **Hide**.

This dashboard, as available at
https://public.tableau.com/views/Chaper7-Dashboards/CO2EmissionsDashboard?:embed=y&:display_count=yes&:s
It tells the story of not only how carbon dioxide emissions have changed geographically over time, but also what compounding factors may be contribute towards the future growth of Tableau Public.

Summary

In this chapter, we discussed best practices of dashboard design as well as how to build an empathetically designed user experience that tells a concise story with data. You also learned that Tableau Public dashboards are composed of one or more sheets (chart visualizations) with added elements such as text, captions, and interactivity such as filters and actions. Dashboards may also contain, or open web pages and can pass dashboard information into the web page in some instances.

Keep in mind that a good dashboard design helps you convey a pertinent message to the reader. Have a goal in mind pertaining to what you wish to get across to the readers. Elements of a good dashboard design start with good data and visualizations, and these visualizations come together in an aesthetically pleasing, intuitive dashboard that helps users tease out conclusions and discoveries of their own. Knowing your audience is critical to dashboard design.

In the next chapter, we will explore filters and actions, which are important tools in adding interactivity to dashboards.

Chapter 8. Filters and Actions

Dashboard filters and actions in Tableau Public enable users to precisely select the information that they would like to explore, which is powerful. Selecting what's interesting makes a data story personal and relevant to users; all of this is done without shifting their attention or their mouse away from the dashboard. Filters and actions create true interactivity. They also give you, as an author, the ability to connect disparate sources of data or web pages to complete user experience.

In this chapter, we will discuss the following topics:

- Adding and using **Quick Filters**
- Filtering the **Data** sources with parameters
- Filtering the **Data** sources with controller worksheets
- Highlighting actions
- Action **Filters**
- URL Actions

Adding and using Filters

Filters are designed to meet several needs of your users; the following are some of these needs:

- Limiting the scope of an analysis
- Allowing users to view only what interests them
- Removing outliers

Using a filter to limit the data that is extracted into your workbook improves the performance of the workbook. However, the following are the two disadvantages of using filters on dashboards:

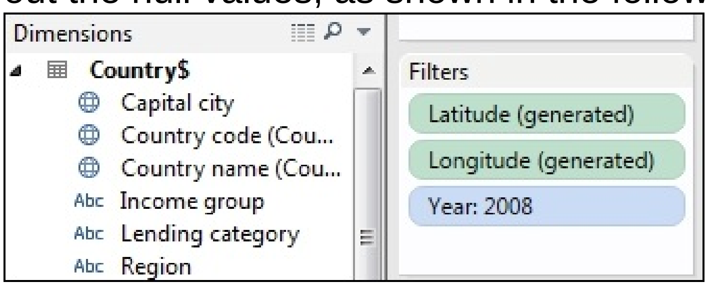
- Using many filter actions with multiple Dimension members can slow down a dashboard's performance.
- Filters can be applied to only one data source at a time. So, in case multiple data sources are being used on one dashboard, you'll need

to use parameters or controller worksheets to filter all of them.

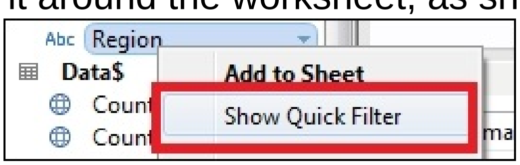
In this chapter, we will use the dashboard that we created in the previous chapter. You can download the finished product by visiting https://public.tableau.com/views/Chaper7-Dashboards/CO2EmissionsDashboard?:embed=y&:display_count=yes&:s
The dashboard uses only one data source at this point. As we progress through this chapter, we'll add another dashboard along with a fourth visualization.

Adding Filters to worksheets

In the dashboard that we are using, we have already filtered the map to show only the data for 2008, which is a blue pill on the **Filters** shelf, because it is a discrete number. You will also see that there are green pills for **Latitude** and **Longitude**. Both of these pills are continuous numbers, and they are included on the **Filters** shelf because we filtered out the null values, as shown in the following screenshot:



If a field is not already on the **Filters** shelf, you can drag it there from the **Data** window. Alternatively, you can create a **Quick Filter** by right-clicking on it in the **Data** window and selecting **Show Quick Filter**, as shown in the following screenshot. Tableau Public will then display the **Quick Filter** on the upper right-hand side of the workspace, but you can easily move it around the worksheet, as shown in the following screenshot:



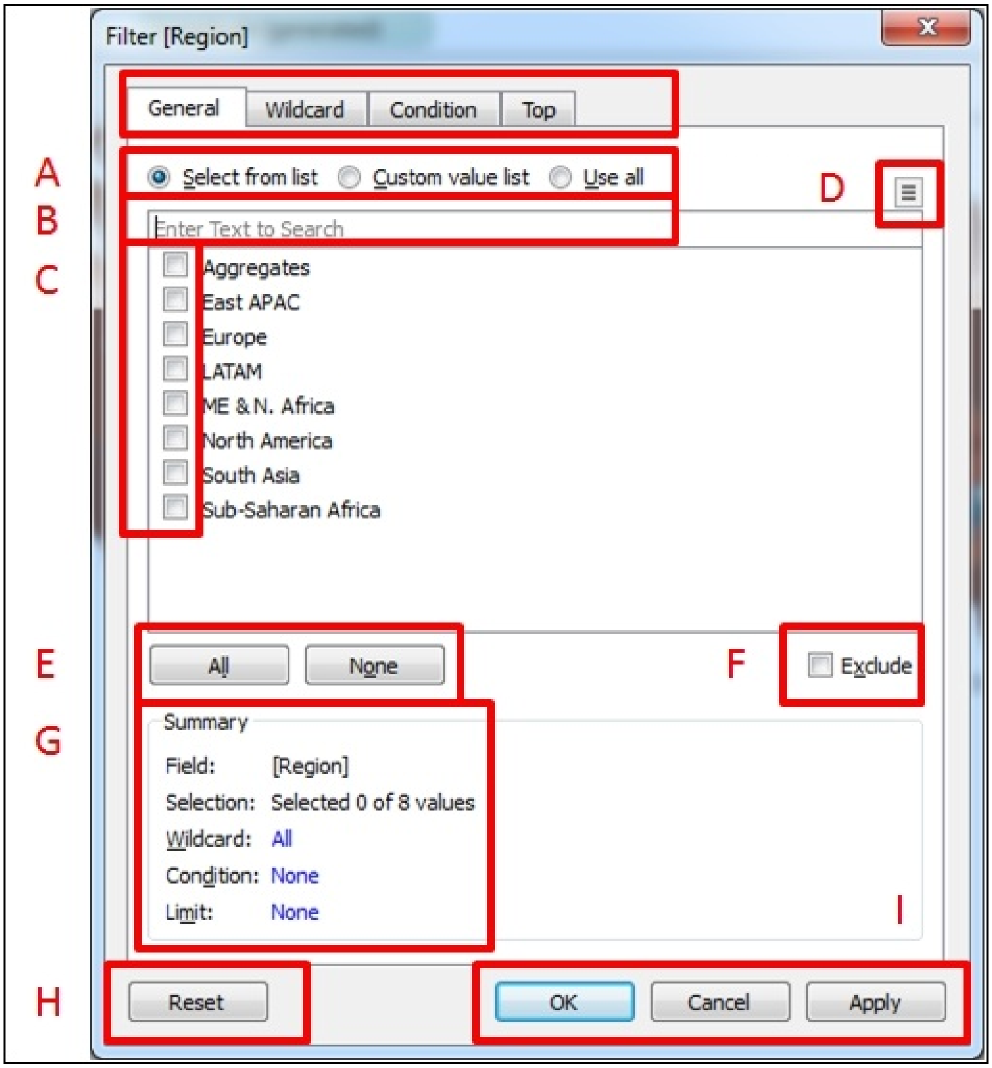
For instance, if a field is a part of a visualization and it's on the **Rows** or **Columns** shelf, the **Marks** card, or anywhere else, you can click on its

Context menu and select **Filter**.

There is a catch to adding filters. If a field is not on the **Filters** shelf of a worksheet, you won't be able to add it as a filter on the dashboard. In the following screenshot, we navigated to the worksheet in the workbook that has the map on it and added the filter there. In the next section of this chapter, we will go back to the dashboard and add the **Quick Filter**.

To add a **Quick Filter** for **Region**, perform the following steps:

1. Right-click on **Region** in the **Data** window.
2. Click on **Show Quick Filter**.
3. Then, it's time to make a selection. With a discrete field such as **Region**, you will see the dimension members that you can choose. The **Filters** dialog box is shown in the following screenshot:



4. In the **General** tab, we can perform the following operations, as shown in the preceding screenshot:

- **Select from list** (**A**): This uses either a custom list of values, or all the values.
- The search box (**B**): We can search for a text string that is present in a dimension member.
- Checkboxes (**C**): Click on the checkboxes next to the dimensions that we want to keep or exclude.
- The filter (**D**): This sets the filter to show fewer values, which means that it shows the values that are automatically limited by the selections of other filters, and it doesn't show all the discrete values in the field.
- **All** or **None** (**E**): You can select **All** or **None**, as required.

- **Exclude** (**F**): This excludes the values that we have selected. If you are excluding values, be careful when exposing the filter to users. They might need some extra instructions about what the exclusion means and how to use it.
- **Summary** (**G**): This shows a summary of your selection.
- **Reset** (**H**): This resets all values.
- **OK**, **Cancel**, and **Apply** (**I**): Click on **Apply** and then on the **OK** button to filter and set worksheet.

5. In our example, we selected all the values and then deselected **Aggregates**.
6. Then, we clicked on **OK**.

The pill for **Region** (which is blue because it's a discrete field) is now in the **Filters** shelf. When you go back to the dashboard, you can add it to the worksheet.

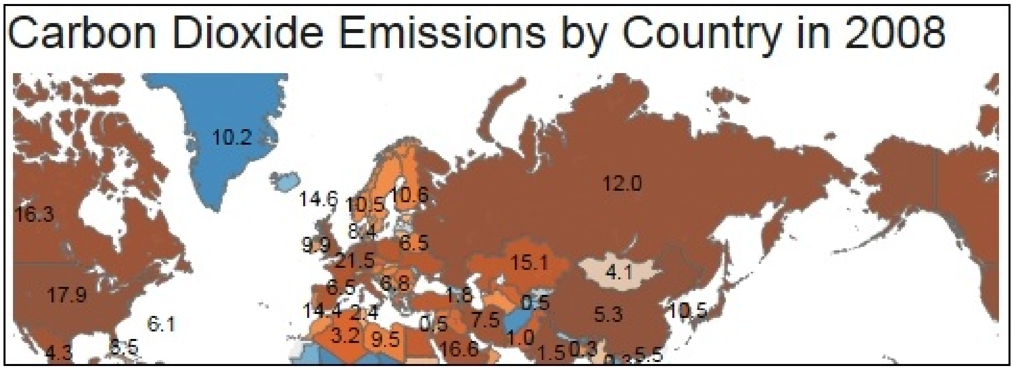
The other **Filter** tabs allow you to further specify the members that you want to retain or exclude in the analysis. You can perform the following operations:

1. Establish the conditions that need to be met
2. Use formulas to determine the members that you want to include
3. Include the top or bottom members by a certain metric

This works particularly well in graphs other than maps because users are often only concerned with the best and worst performers.

Adding Quick Filters to a dashboard

If your worksheet is already on the dashboard, the new filter will not appear there automatically. The dashboard is laid out in a vertical column for the most part, and we do not have filters on it. Its top looks like the following screenshot:



We want to add the **Region Quick Filter** to the dashboard and then apply it to the other worksheets on the dashboard. When we add a filter (or a legend) to a dashboard that we have arranged very specifically, Tableau Public will automatically add a vertical layout container on the far right. Don't be alarmed by this. You can move filters and legends to the locations you want them to be in.

If a field is already a part of a visualization that is on a dashboard because it is one of the fields that is being displayed on the **Detail** or **Tooltip** shelves, you can add it to the dashboard by clicking on the **Context** menu for the visualization and then selecting it from the list of the **Quick Filter** options.

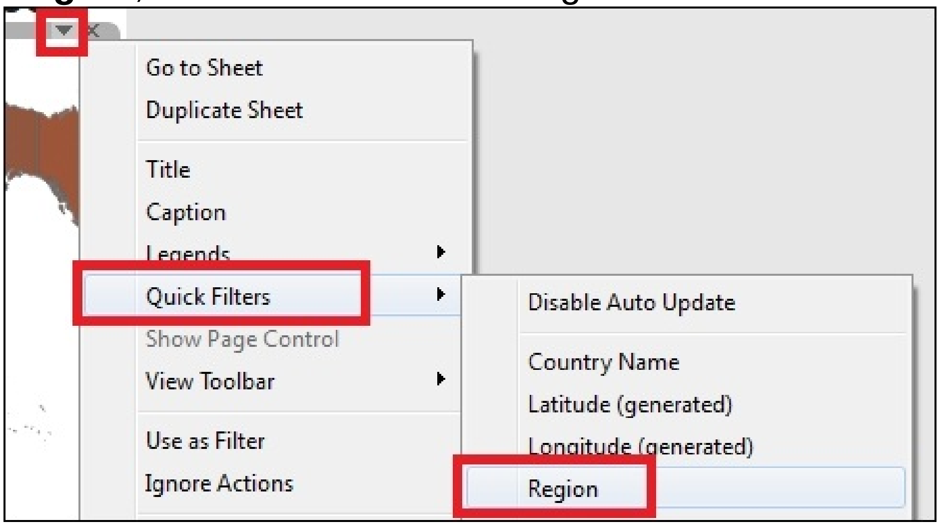
In the next few exercises, we will discuss the following topics:

- Adding the **Region Quick Filter** to the dashboard
- Formatting it to be a drop-down list
- Setting it to show only the relevant values
- Applying it to all the worksheets on the dashboard
- Moving it so that it's in the same container as that of the title of the dashboard

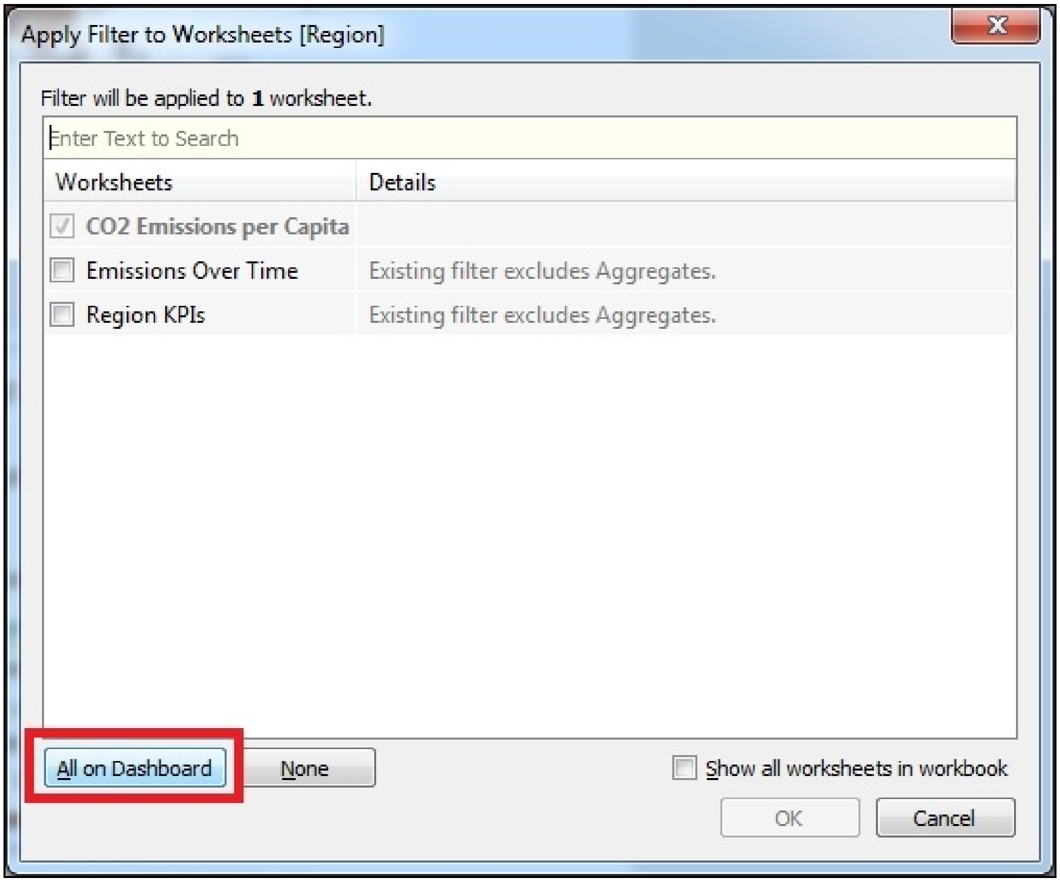
In order to add and modify the **Quick Filter**, perform the following steps:

1. Select the map worksheet on the dashboard by either clicking on the ocean in the map, or selecting it from the list of worksheets in the **Dashboard** menu to the left of the dashboard.
2. Click on the **Context** menu.
3. Hover the pointer of the mouse over the **Quick Filters** and click on

Region, as shown in the following screenshot:



4. The **Region** filter will appear in a **Vertical Layout Container** to the far right of the dashboard. We'll move it later. It's important to be mindful of unused white space.
5. Now, we will change the format of the quick filter. There are many options, and we want the most compact one, which is a multi-select drop-down list. Click on the **Context** menu of the filter and change it to show **Multiple Values (Dropdown)**.
6. There are many other modifications that we can make by clicking on the **Context** menu of the filter. Click on the **Context** menu again.
7. Click on **Only Relevant Values** so that it shows only the **Regions** that meet the conditions of the other filters.
8. Apply it to the other worksheets on the dashboard that are using the same data source. From the **Context** menu, click on **Apply to Worksheets**.
9. In the **Apply Filter to the Worksheets [Region]** dialog box, click on **All on Dashboard**, as shown in the following screenshot, and then click on **OK**. Keep in mind that this only applies **Quick Filters** to worksheets using the same data source:



The other modifications that you can make from the **Context** menu of the filter and which add polish and specificity to your dashboard, are as follows:

- Applying **Quick Filters** to worksheets allows authors to select the sheets to which the filter applies, where all the worksheets are using the same data source
- Formatting **Quick Filters** gives authors control over the font faces and styles used in the title and the list of values
- Customizing **Quick Filters** can remove or show the **All** option
- Showing, hiding, or editing the title can prompt users to take action
- Changing the display of the filter, whether it's radio buttons for short lists or multiple value drop-down lists for longer lists, enables authors to select the best format for both the data and consumers
- Showing relevant values, also known as cascading filters, is a great way of simplifying the filtering experience for consumers so that selections in a filter determine the contents of another

Formatting the filter box so that it has a floating or fixed width is another feature that gives authors control over the appearance of worksheet.

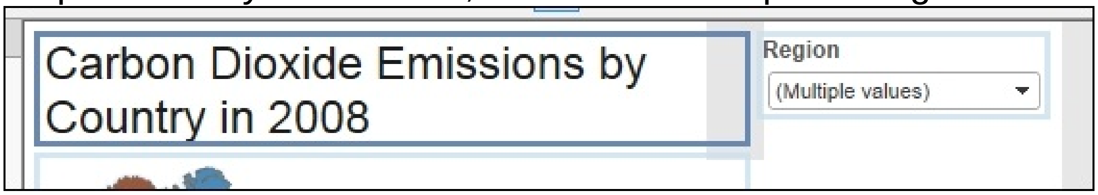
Moving the Quick Filter

The quick filter is still present in the vertical container to the right that Tableau Public automatically created for it. If you have several **Quick Filters** and **Legends**, it's perfectly fine to leave them in the container. Just be conscious of how you are using *white space*.

We plan to add a parameter later, but for now, we want to put this quick filter next to the title for the dashboard. The title is in a horizontal layout container already, which means that we can put other objects next to it.

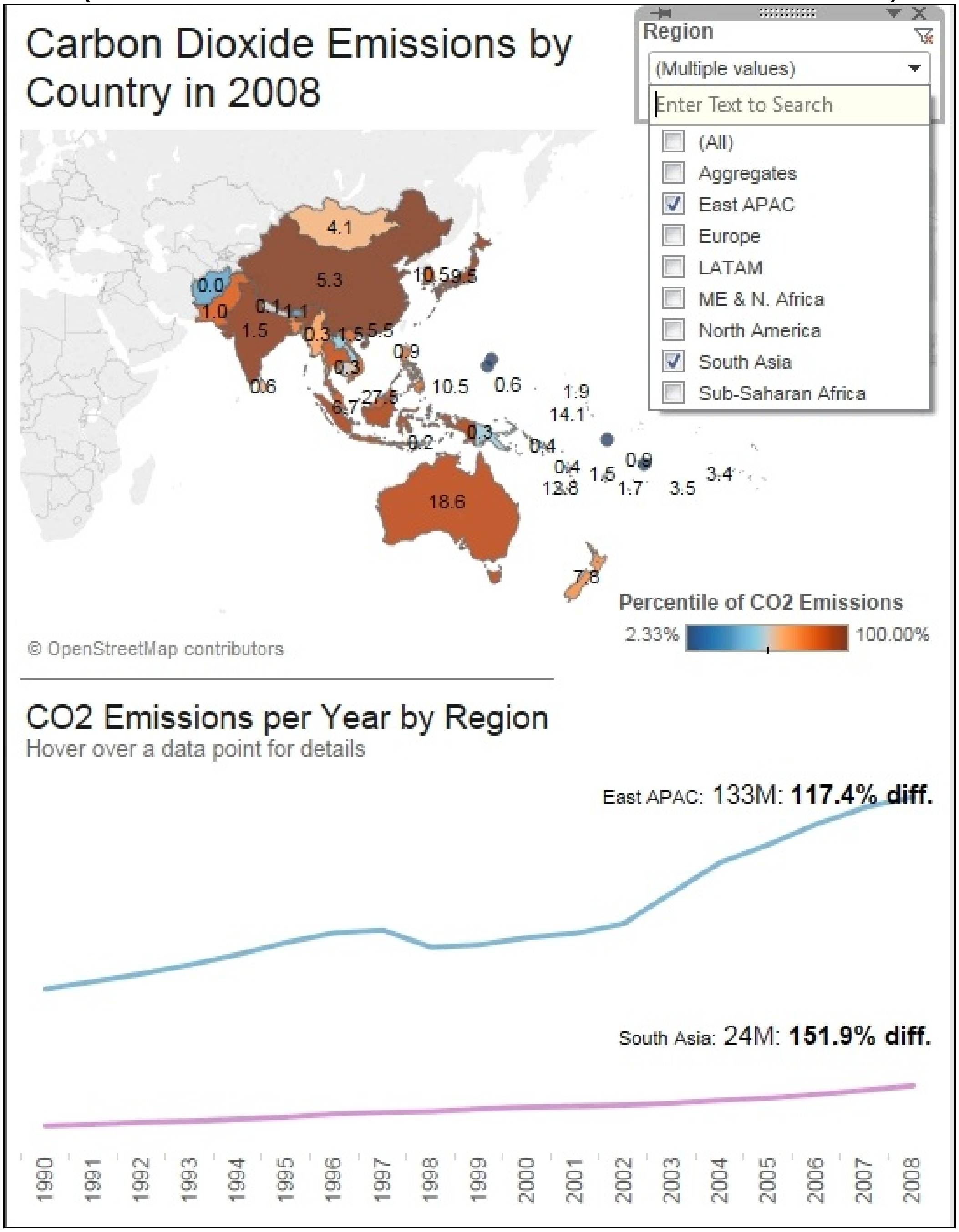
1. In order to drag the **Quick Filter** into the same horizontal layout container with the title, select the **Quick Filter** by clicking on it.
2. Hover the pointer of the mouse over the center of the **Quick Filter**. You can click on the area with the white hashes, which is outlined by a red box, as shown in the following screenshot:

- When you get the white crosshair mouse, click on it and drag the filter into the horizontal layout container shared by the title for the dashboard. You will know that it is in the right place when the right border of the container has a shadow, as shown in following screenshot, and emphasized by the red box, as shown in the preceding screenshot:



- Note that the **Vertical Layout Container** has disappeared on its own.
- Change the filter values and see how the dashboard display changes.

In the following screenshot, the changes that we made to the **Region** filter were applied to the other worksheets to which we had applied the filter (which included all the worksheets on the dashboard):



Filtering across Data sources with parameters

Although the **Quick Filter** can be applied to the worksheets using its data source, it can't be applied to worksheets using other data sources. In order to filter multiple data sources with one list, we need either a reference table or a parameter.

While you can use parameters for many important functions, such as determining which worksheets appear on a dashboard, there are some disadvantages:

- Only one value can be selected at a time
- The list of values for static parameters, and they need to be maintained manually
- Parameters are related to the **Data** sources through the use of calculated fields, which can be complicated to maintain

In the next few exercises, we will accomplish several tasks via the creation of three parameters. We will create parameters that integrate with calculated fields to:

- Show a single year in one graph and a range of years in another graph
- Filter across multiple **Data** sources
- Allow users to select the detail that needs to be explored

There are a couple of key concepts about parameters that you should keep in mind as we go through this chapter:

1. Each parameter is unique to a workbook and not to a **Data** source.

> ## Note
>
> You can copy parameters from one workbook to another without

issues because they are not dependent on data sources.

2. Each parameter has a name, data type, format, and value.

The first task that we need to accomplish by creating a parameter involves using a **Year** filter for the dashboard. Two of the visualizations, namely the map and the line graph, have a **Time** dimension. We want to create a filter from which users can select the year. When a user selects a year, the map will be filtered to show only that year, but the line graph will show five years on either side of dashboard.

In order to build this functionality, we will need to create three new elements. The descriptions of these elements are as follows:

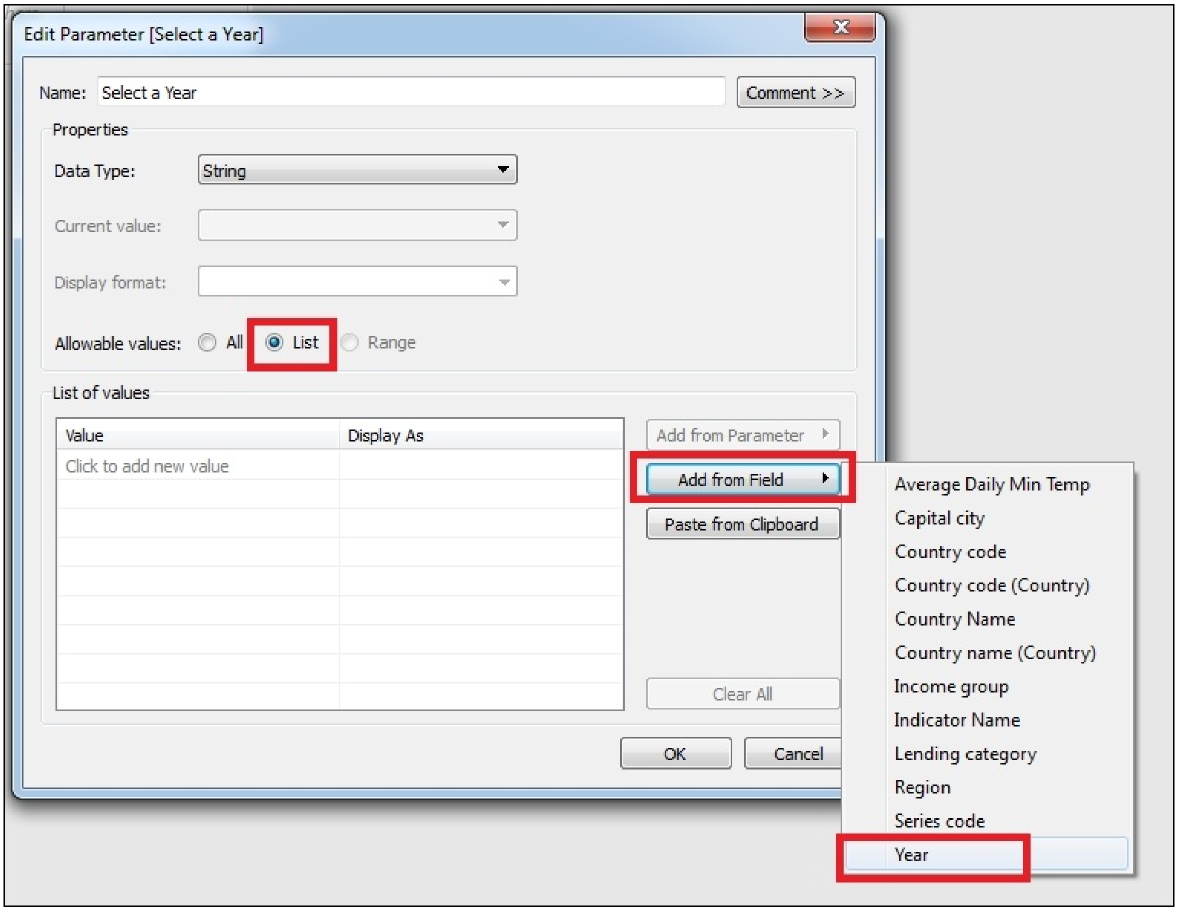
- The `Year` parameter: This has a data type of `INT` and a range of values
- A **Calculated** field: This will be used as a filter on the map so that the year that we are showing in the map matches the selection from the parameter
- A **Second Calculated** field: This will be used as a filter on the line graph that allows five years on either side of the selected year

Let's create a **Year** parameter by performing the following steps:

1. Create a new parameter by clicking on the **Context** menu in the **Data** window or by right-clicking on the white space in the **Data** window and selecting **Create Parameter**.
2. Give the parameter a name that represents what you want users to do. Ensure that the name does not match any of the field names in your data source. Otherwise, this will cause issues.
3. We named our parameter **Select a Year**.
4. Select the data type of the parameter. When you are using a parameter in a **Calculated** field, you can convert it to different types. We want ours to be a string because **Year** is a string field in the **Data** source.
5. If we had selected a different type of field, we would have had to format it accordingly. One of the reasons why we left it as a string is

that we won't be running the risk of displaying improperly formatted values.

6. Select **List** for the **Allowable values**, as shown in the following screenshot:



7. You have the option of populating the list from your own list of values or from one of the fields in either of the two **Data** sources.

Note

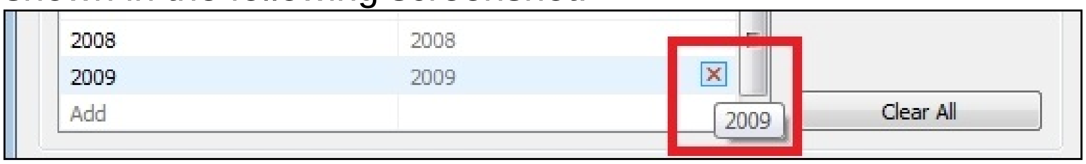
The parameter lists are not updated dynamically from the fields.

8. Click on the **Add from Field** button and then select the **Year** field.

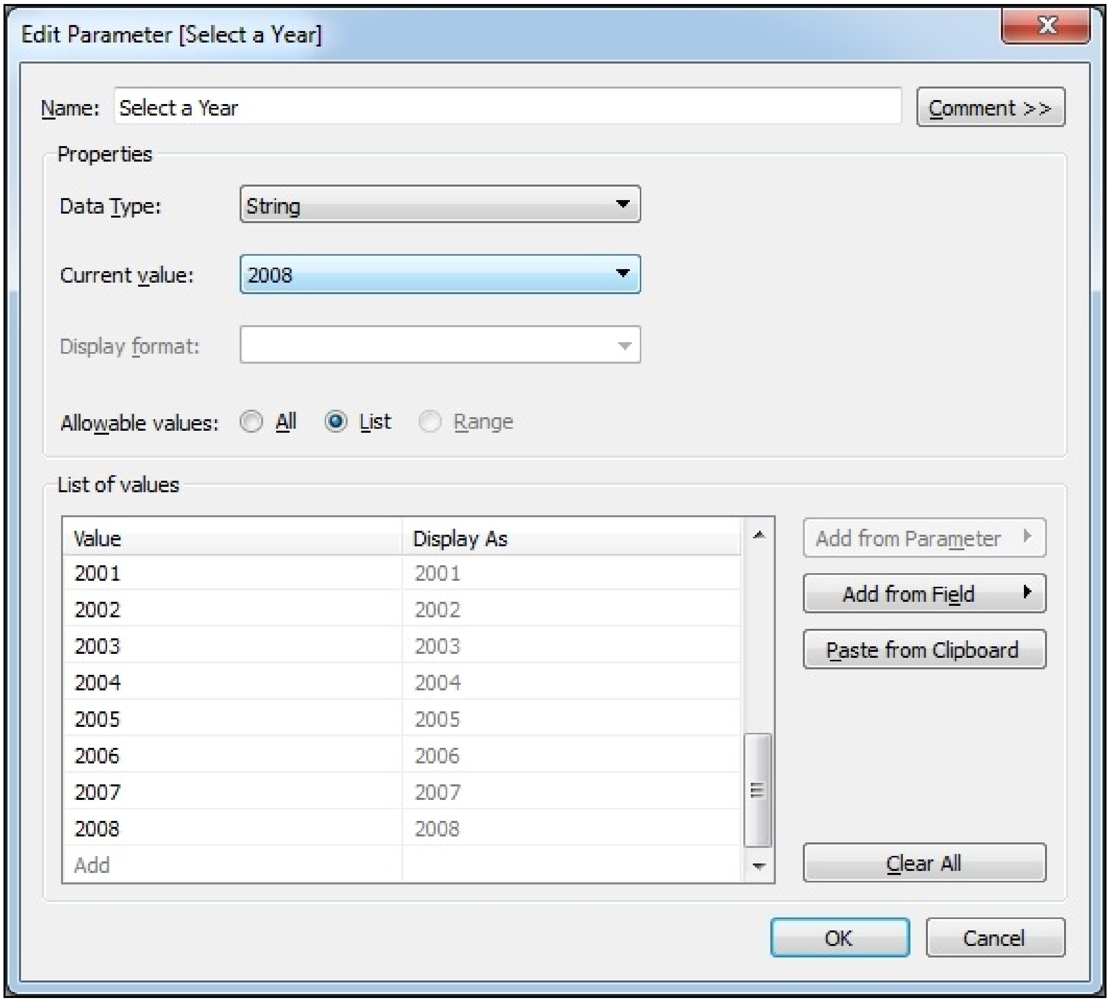
Note

You only see fields of the same data type as a parameter. If there were multiple data sources in the workbook, you would see the others as well.

9. Now, the list of values matches that of the source field from the **Data** source. But from the bottom of the list of values, remove the years after *2008* because the data collected for these years is not complete. You can do this by hovering the pointer of the mouse over the row and clicking on the **X** that appears to the right of a values, as shown in the following screenshot:



10. Also, keep in mind that Tableau is case-sensitive, which means that the contents of the parameter list must match the contents of the field list.
11. The final parameter has been set to a default of **2008**, as shown in the following screenshot:

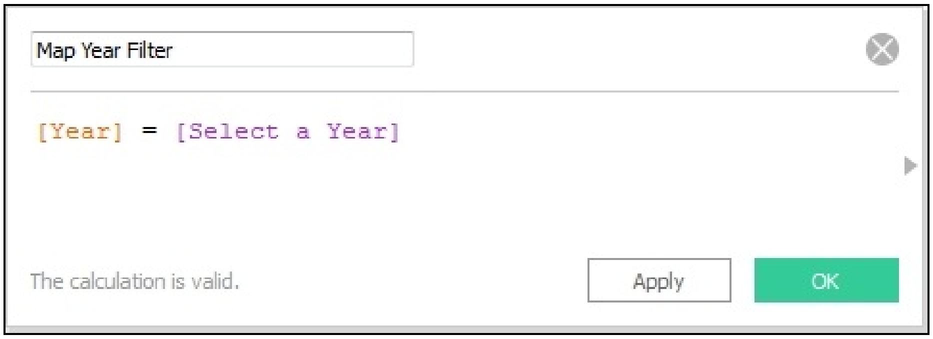


12. Now, click on **OK**.

Using parameters as Filters

You now have a parameter that you can use to filter certain fields. In the following steps, we will create two new calculated fields—the first will filter the map to show only the **Year** selected, and the second will filter the line graph to show five years on either side of the **Year** selected:

1. On the map's worksheet, the first task is to create a new **Calculated** field called **Map Year Filter**.
2. The formula is simple; it is Boolean expression, and it tells Tableau Public that we want only the **Year** values that map the parameter selection, as shown in the following screenshot:

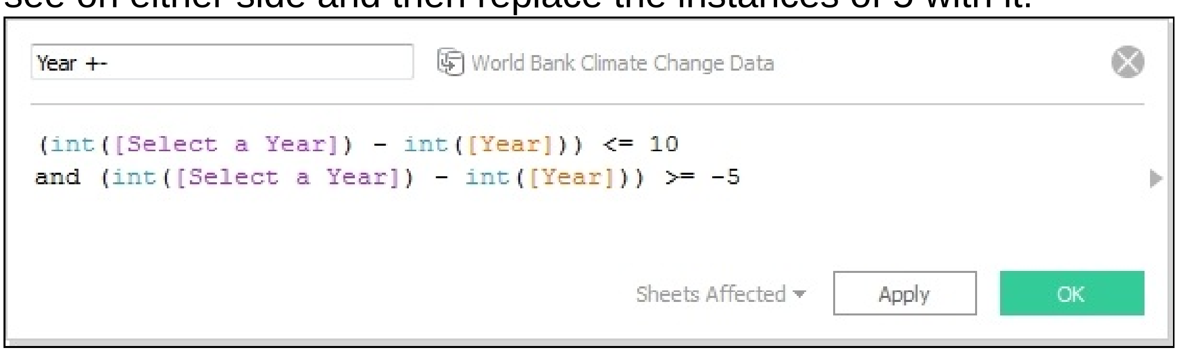


- Remove the filter on the **Year** field, which is a blue pill on the **Filters** shelf.
- Drag this **new** field from the **Dimensions** pane to the **Filters** shelf and select **True**.
- Apply this filter to the **Region KPIs** worksheet.
- Right-click on the **Select a Year** parameter, which is below the **Measures** in the **Data** window, and show the parameter control.

Next, we will create a **Calculated** field for the line graph that filters to show five years on either side of the parameter.

Open the worksheet with the line graph, perform the following steps to create a **Calculated** field:

1. Create a new **Calculated** field and name it **Year +-**.
2. In the following screenshot, we used a formula to convert both the parameter and field to integers so that we can perform mathematical functions on them. If you want to, you can create another parameter to allow users to input the number of years that they would like to see on either side and then replace the instances of 5 with it:



3. Click on **OK**.
4. Add this new field from the **Measures** pane to the **Filters** shelf and

select **True**.

Now, we will return to the dashboard and continue making modifications. The first thing that we want to do is show the parameter.

5. From the **Context** menu for the map worksheets on the dashboard, click on **Parameters** and then click on the **Select a Year** parameter.
6. The parameter automatically appears next to the **Region Quick Filter**. It would look better if the two were stacked. So, add a **Vertical Layout Container** to the **Horizontal Container** with the header and filter, and then move the **Filter** and parameter into it.
7. Add an extra 100 pixels of height to the dashboard.

Modifying titles

We can modify the header of the dashboard to reflect the selection of a year as well. The title is hard-coded to identify CO2 emissions from 2008, but if we select a different year from the parameter, such as 2005, then we need to ensure that the title is updated.

We can add the parameter to the title by performing the following steps:

1. Double-click on the title.
2. Delete the hard-coded reference to **2008**.
3. Click on the **Insert** button.
4. Select the parameter that we want to add, as shown in the following screenshot:

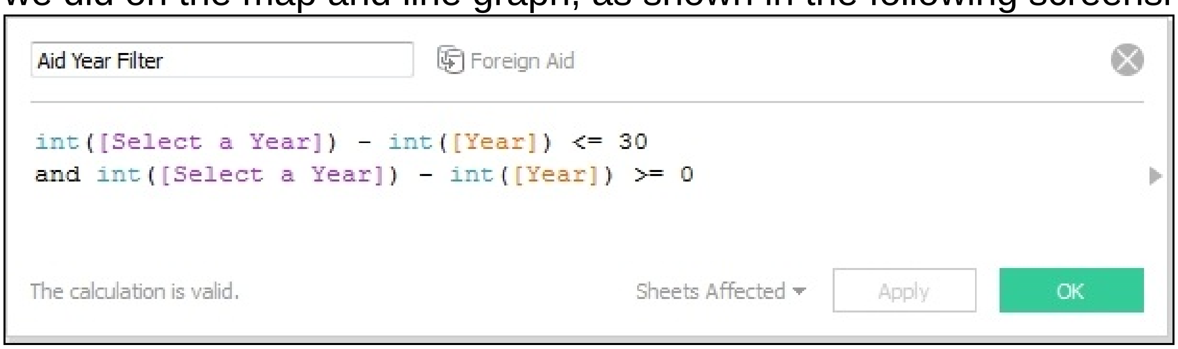
Since this title is actually a text object on the dashboard and not a

Since this title is actually a text object on the dashboard and not a specific worksheet title, we cannot insert field values, but we can insert parameters. It's always a good idea to make it very obvious what your user is looking at so that they don't run the risk of assuming the wrong thing.

Filtering across multiple Data sources with parameters

Parameters are useful when you need to filter across multiple **Data** sources. We loaded a second Data source, which contains data related to the foreign aid given to various countries by the United States since 1948, into our workbook. There are several fields that have the same name, namely **Region**, **Country**, and **Year**. We created a simple bar graph that shows the total aid in dollars over a period of time on a new worksheet called **Aid Graph**.

We created a filtering field for this visualization that shows only the **Year** selected in the parameter and the 30 years before it. In this case, we're using the same **Year** parameter that is being displayed on the dashboard. We also created a **Calculated** field, referencing it in the **Foreign Aid Data** sources, and then added it to the **Filters** shelf, just like we did on the map and line graph, as shown in the following screenshot:



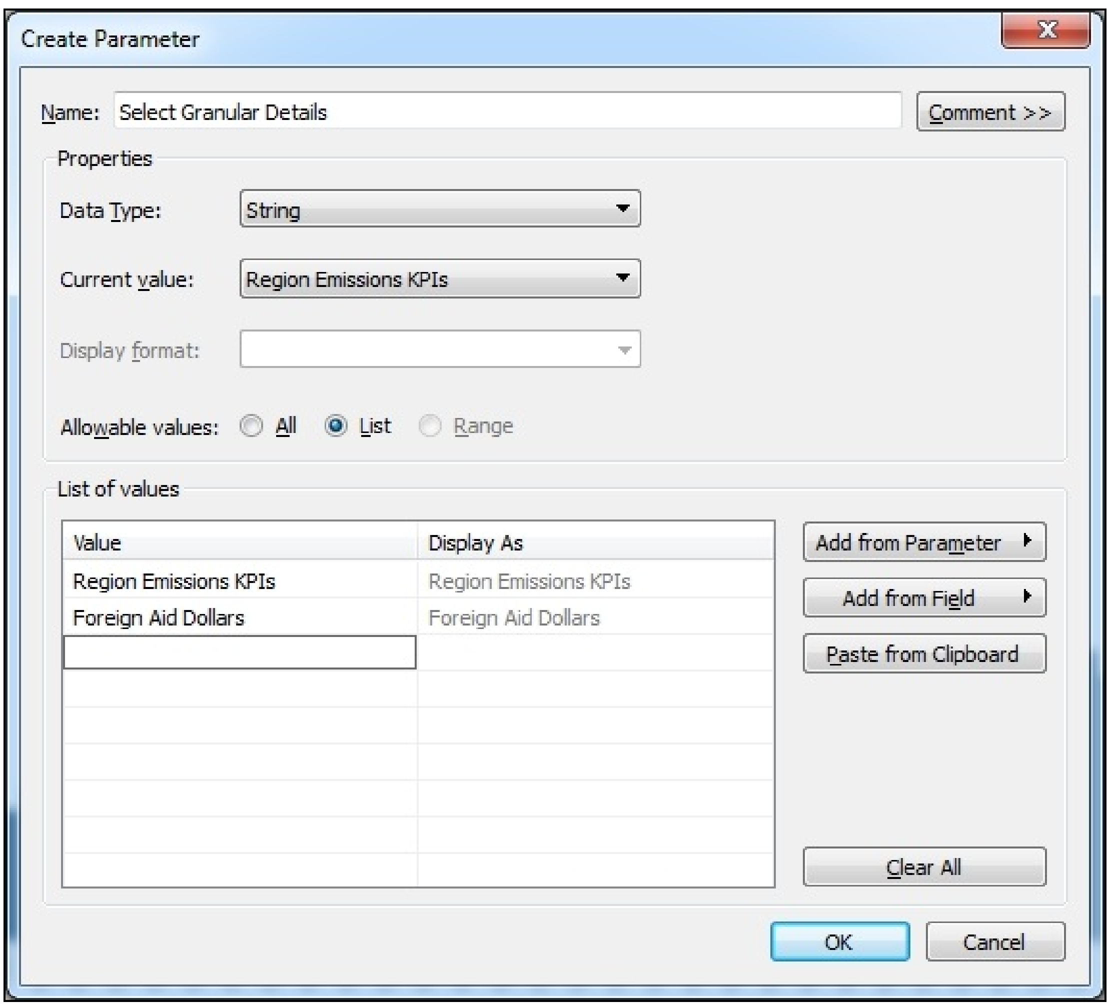
We will create a new parameter that allows users to select the visualization that they wish to view in the third of the last workbook, which is where we show granular details. Currently, the **Region KPIs** worksheet is displayed there, but we would like to provide users with the option of deciding what to see in the workbook.

In order to do this, we will create a new parameter. Then, we will create a **Calculated** field that can be used as a filter in each **Data** sources.

Let's create a parameter by performing the following steps:

1. Create a new parameter and name it **Select Granular Details**, which is a string with a list of two possible values.
2. We have capitalized these two values carefully because Tableau Public is case-sensitive. We will include following values in a **Calculated** field in the next step. The values are shown in the following screenshot:

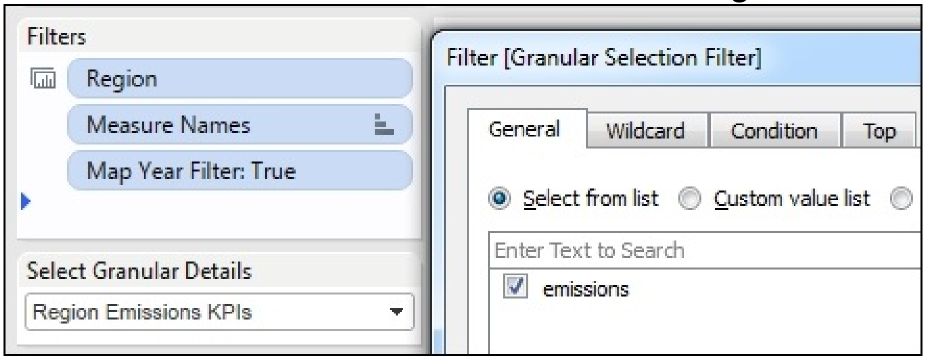
    - **Region Emissions KPIs**
    - **Foreign Aid Dollars**



3. Click on **OK**.
4. In the **Foreign Aid Data** source, create a new **Calculated** field. This field will use a case statement to create a string output based on the parameter's selection.
5. We named the field **Granular Selection Filter**, and we wrote a case statement that states that if **Region Emission KPIs** is selected, then the output is **emissions**; otherwise, it's **Foreign Aid**. If there were more than two values, then we would have another condition, as shown in the following screenshot:

6. Click on **OK**.
7. Add the **Granular Selection Filter** field to the **Filters** shelf of the **Aid Graph**.
8. The string value that corresponds to the current value of the parameter will be displayed in the **Filter** list. Select it and click on **OK**.
9. Select **Show Parameter Control** and select the other value. The graph should now appear.
10. Right-click on the **Granular Selection Filter** field and select **Copy**. We'll paste it exactly as is into the **Data** source that contains the other worksheet on that we want to work.
11. Go to the **Region KPIs** worksheet. In the **Dimensions** pane, right-click on it and select **Paste**. The field that you just created now exists in both the data sources.
12. Right-click on the **Select Granular Details** parameter and select **Show Parameter Control**. You have to do this on every sheet on which you'd like to see it.
13. Ensure that the value selected corresponds to the graph that you're viewing.

14. Drag the **Granular Selection Filter** field to the **Filters** shelf and select **emissions**, as shown in the following screenshot:

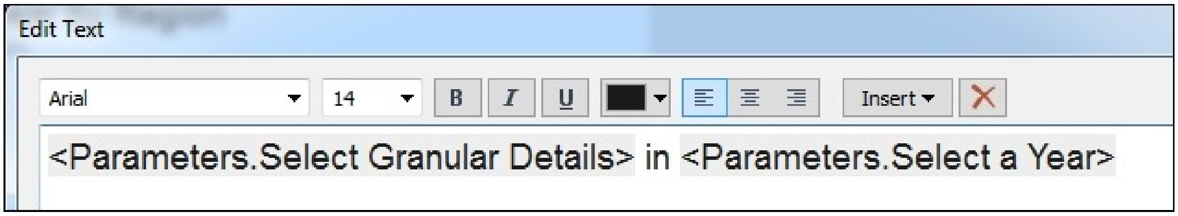


15. Click on **OK**.

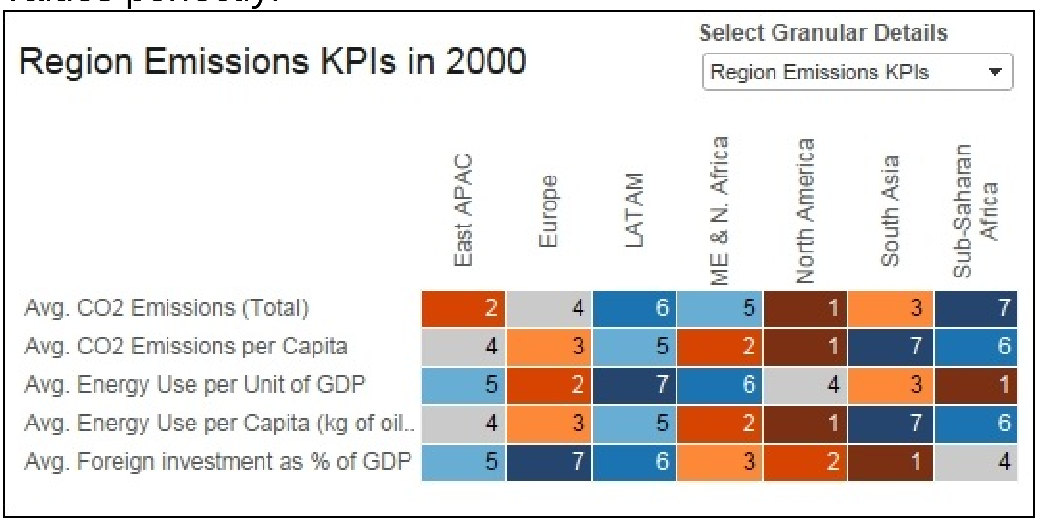
Now, we can go back to the dashboard and add the parameter and the new visualization.

We will stack the visualizations into a **Vertical Layout Container**, and we will hide the title of each so that no space is unnecessarily taken up when it isn't selected. Then, we will modify the space in the title names of account for the parameter selection by performing the following steps:

1. First, add a **Vertical Layout Container** to the bottom of the dashboard and add the **Region KPIs** and **Aid Dollars** worksheets into it.
2. When you do this, the the parameters that appear in these worksheets will be added to the container with the filter and parameter at the top.
3. Add a **Horizontal Layout Container** within the new **Vertical Layout Container**.
4. Change the title for this area so that it is not floating and drag it into the **Horizontal Layout Container**.
5. Drag the parameter from the top into the **Horizontal Layout Container** as well.
6. Modify the text for the title so that it represents exactly what we can see in the following screenshot:



7. Right-click on the title for the **Aid Graph** and hide it.
8. Note that when you change the parameter value, both the title and the visualization change.
9. If you're having issues with the visualizations changing properly, the first thing that you need to check is the capitalization and spelling in the case statement. It's fairly common to make mistakes when writing these statements. Therefore, you should check them to ensure that you have written conditions that match the parameter values perfectly:



Actions

There are three types of actions that you can create on dashboards to add context, relevance, and specificity for your users, namely **Highlight**, **Filter**, and **URL**.

The following list describes these three action types:

- **Highlight actions**: These draw attention to marks related to selections in a visualization without filtering other data points on the visualization.
- **Filter actions**: These show only the selections made in a source visualization. They can be set up in such a way that when the marks in the source sheet are deselected, the target visualization shows all the marks, shows the filtered marks, or excludes all the marks.
- **URL actions**: These allow you to link to generic URLs or dynamically add selected field values into URLs to link Tableau Public dashboards to external resources. As an example, we'll show you how to insert a country's name into a URL to directly link to its page on the World Bank's website.

Highlight actions do not filter visualizations, they only draw attention. URL actions link to external websites. Therefore, they don't have implications in dashboard performance. Filter actions have the following benefits:

- They are comparably faster than **Quick Filters**. Tableau Public does not need to scan the contents of a field before they can function.
- They can filter across multiple data sources without setting up join conditions ahead of time, which will be explored later on in this chapter.
- They allow users to select and view only the granular data that's important to them.

Actions have a disadvantage. For a novice author, maintaining actions can be time-consuming. The actions that we will use in the following exercises are designed to allow users to filter all the visualizations to show only the countries that interest them. Then, we'll add a URL action that links to the World Bank's web page for every individual country.

In the dashboard, we have a map that shows every country for which we have data. We have filters for time and other geographic dimensions. Therefore, we have an opportunity to do the following two things with the countries:

- We can create a filter action that limits the selections in other worksheets to the countries that a user has selected
- We can create a URL action that allows a user to see the World Bank's web page for each country

The first task involves creating the **Filter** action that runs when a user selects a country from the map by performing the following steps:

1. On the dashboard, click on the **Dashboard** menu and then click on **Actions**.
2. Click on **Add Action**, and then click on **Filter**.
3. Name the action **Filter on Country**. It's important to name actions according to their functions so that when you are editing and testing them later, you know where to look first.

   The following attributes need to be selected:

   - The source of the action, that is, the visualization from which it originates
   - The action that triggers it
   - The target sheets
   - The fields on which you need to run an action

4. For the source, select the **CO2 Emissions per Capita** sheet on the dashboard.
5. Run the action on **Select**. Thus, when a user clicks on a field, its correlated marks in the target sheet will be highlighted. The alternatives are running the action when a user hovers over a mark or showing a hyperlink to run the action when a user rolls over **Menu**. We will use **Menu** for the URL action.
6. Target all the sheets on the dashboard except for the **CO2**

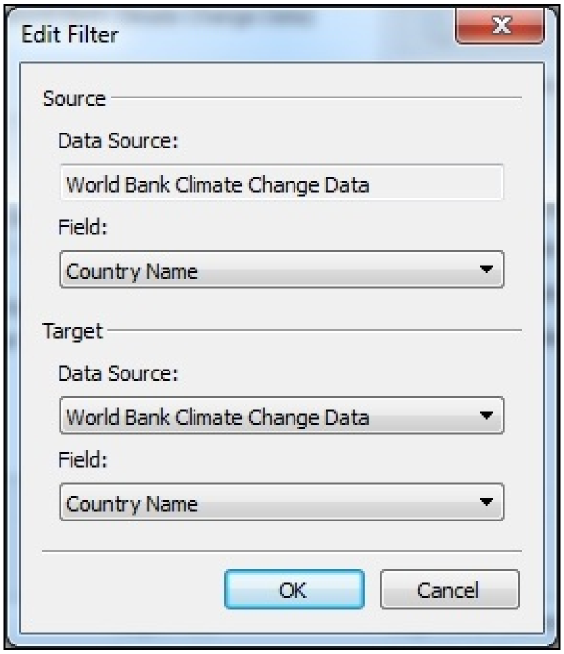
Emissions per Capita sheet.

7. Leave the default values to clear the selection. If we had worksheets that we wanted to hide until someone made a selection, which is a great option, we would have excluded all values.
8. Next, we need to establish the fields on which we want the filter to run. When a user clicks on a country, the value for that country is used as a filter for other sheets.

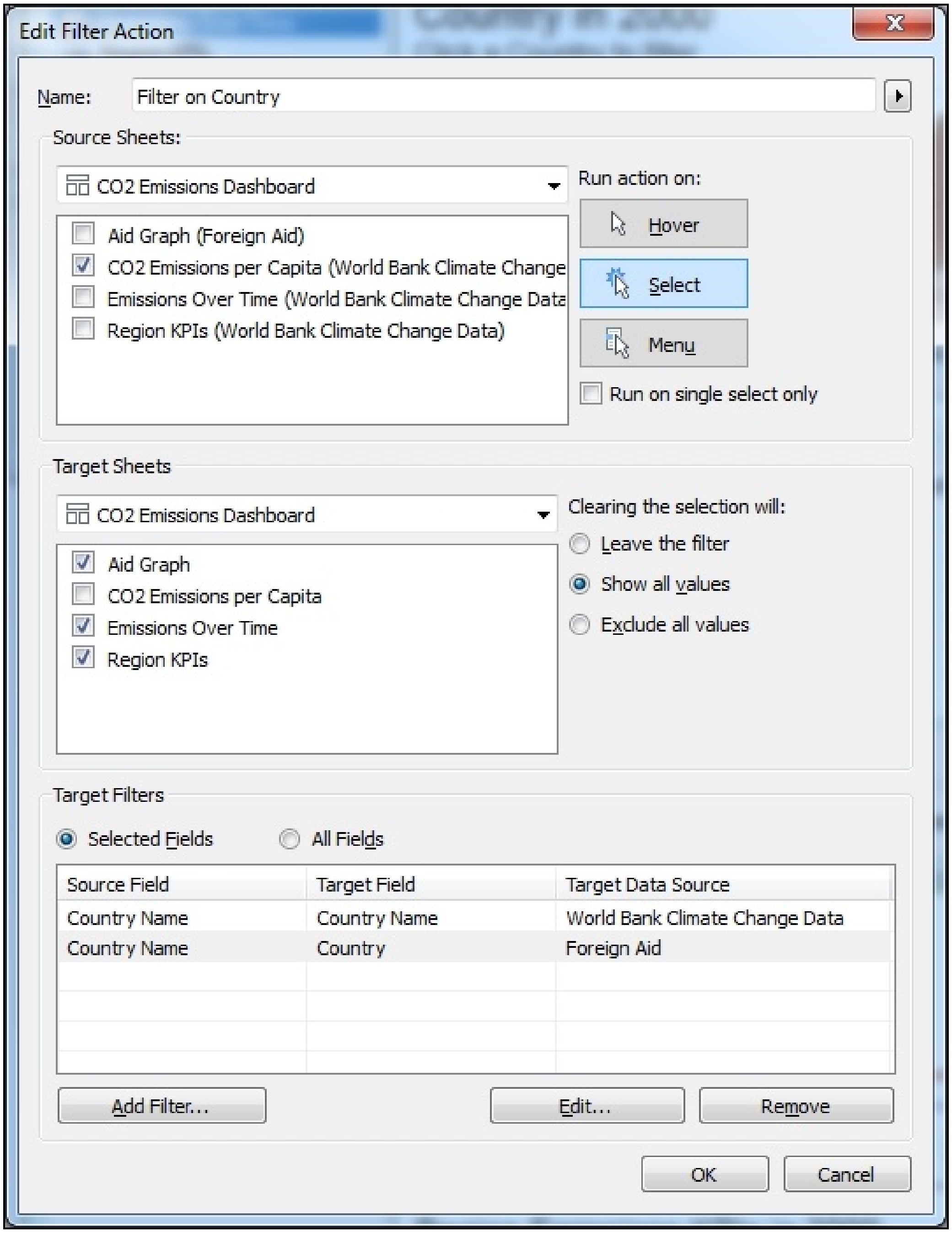
   In case you left the setting on **All Fields**, then all the attributes of the mark will be passed as a filter, but only to fields of the same name. For example, if you were going through the values for region and country, then each sheet with those exact same field names and values would be filtered.

   Since we have a secondary Data source, specifically the **Foreign Aid** data, that does not have the same field name for each country (though the values are the same) and has completely different values for the **Region** field, we need to establish targets.

9. Click on the **Selected Fields** radio button. Though using specific fields can potentially cause the dashboard to run more slowly with very large data sets, we need to establish the fields on which we need to join the **Data** sources.
10. Click on **Add Filter**. Since the action is running from a sheet using the World Bank Climate Change data source, it is the default for the source. Pick the **Country Name** field. Then, for the target, select the same **Data** source and field. This tells Tableau Public that for sheets that use this **Data** source, use the **Country Name** field as both the source and target. Check out the result in the following screenshot:

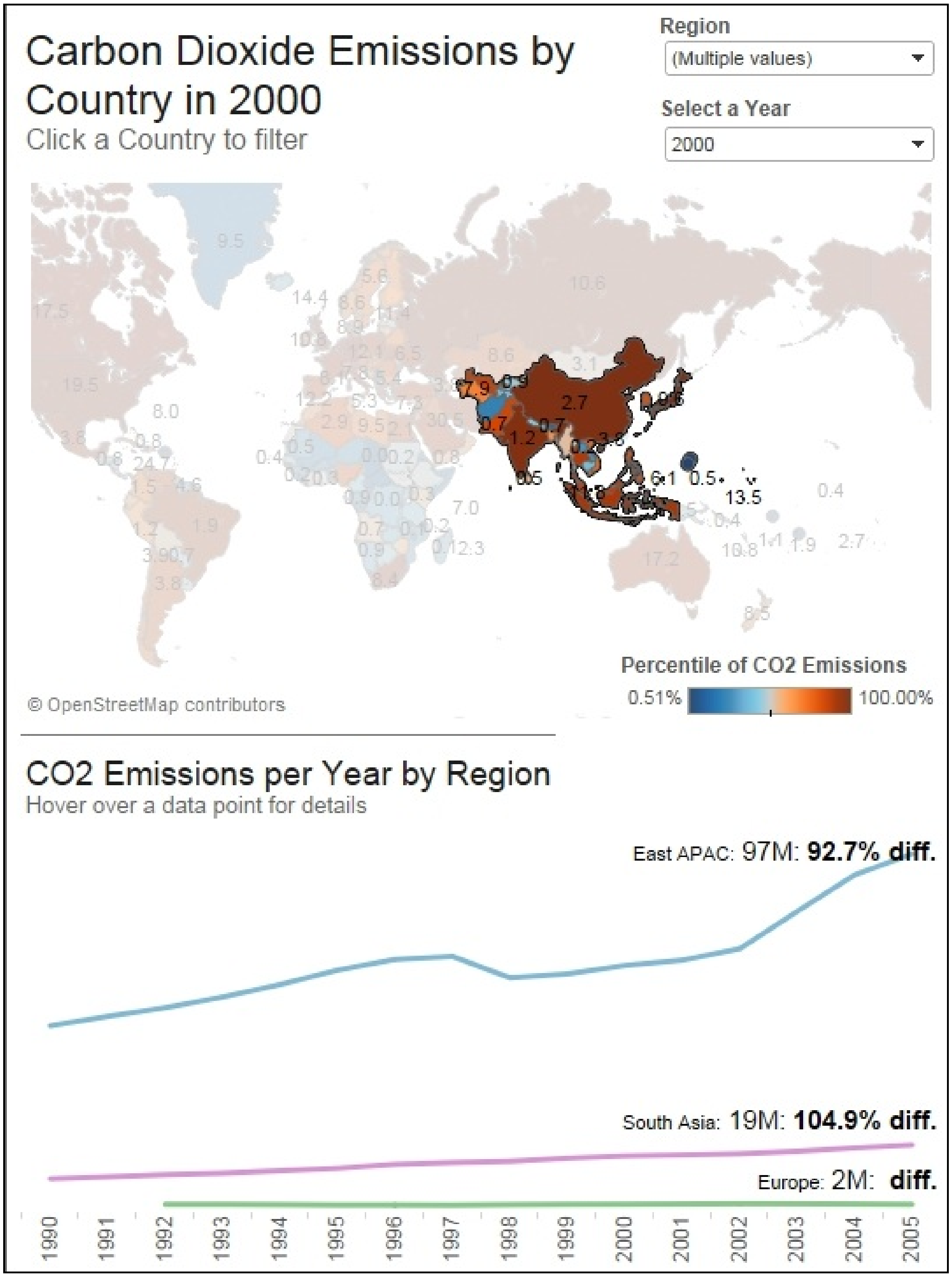


11. Click on **OK**.
12. Click on **Add filter** again. Select the **Country Name** field again as the source, but for the target, select the **Foreign Aid Data** sources and then select the **Country** field as the target.
13. Click on **OK**.
14. Check out the action dialog box, which is now complete, as shown in the following screenshot:



15. You can edit and remove the target filters as required.
16. Click on **OK**.
17. Back on the dashboard, modify the title above the map so that it says **Click on a country to filter** as sub-text in 12 pt Arial, as shown in the following screenshot:

18. Draw a box around an area of interest on the map. Check out the changes made to other visualizations in the dashboard, as shown in the following screenshot:



Note

If we were to add additional visualizations to this dashboard, they automatically would become targets of the actions that we have established. For this reason, it's wise to wait until you have added everything of interest to the dashboard before creating actions.

URL actions

URL actions are useful when you wish to connect users to external and auxiliary data, whether it's from a public interest website or to a related dashboard. In the next example, you will create a URL action that allows users to view the World Bank's website for each country.

The URL action that you will build will append the name of each country as a variable to the root URL for the World Bank's well-organized websites.

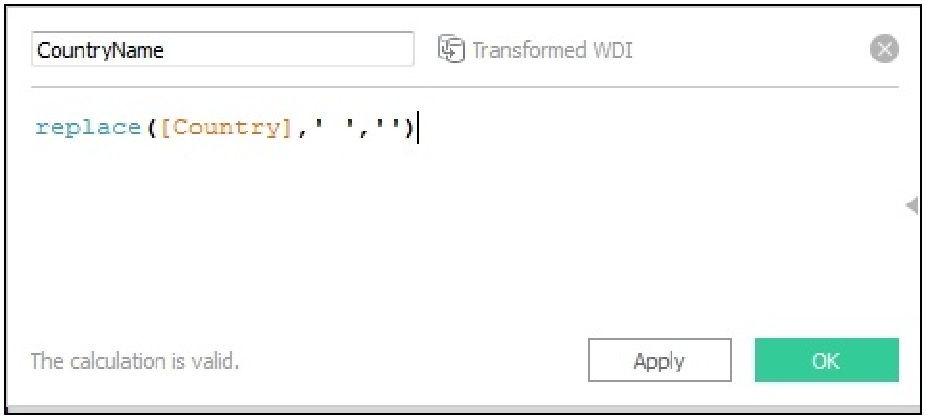
The root URL is http://www.worldbank.org/en/country, and in order to get to a specific country, such as South Africa, you need to add the country name so that it looks like http://www.worldbank.org/en/country/southafrica.

There is one issue—there are no spaces in the country names in the URL and yet, there are spaces in the country names in the data source.

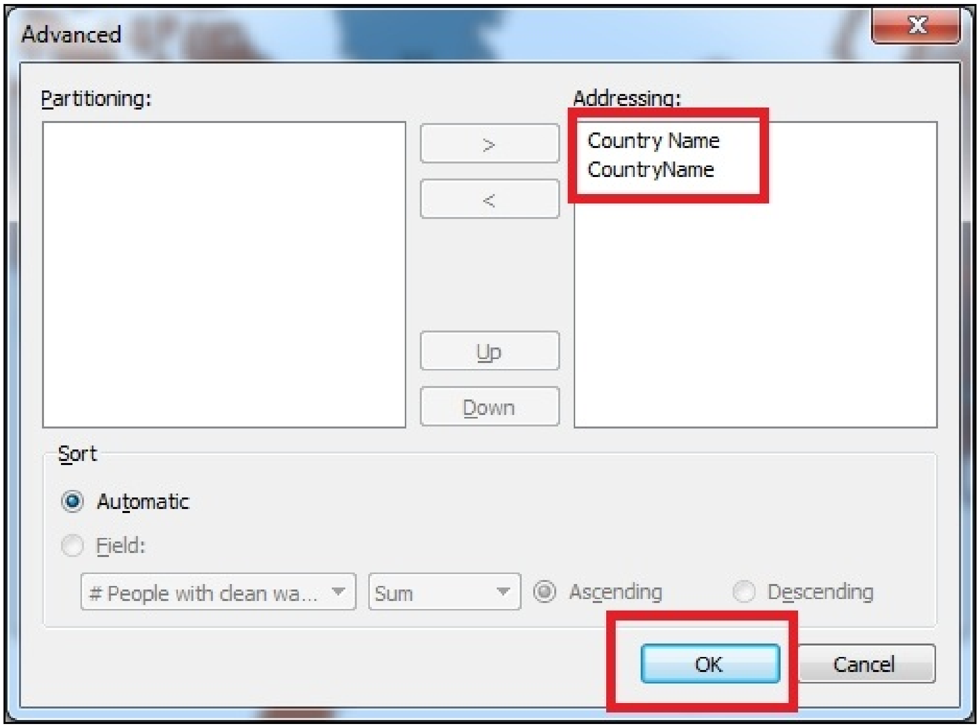
So first, create a **Calculated** field that removes the space in each **Country** name so that we can append it as a token to the URL.

To remove and append the **Country** name with a URL, perform the following steps:

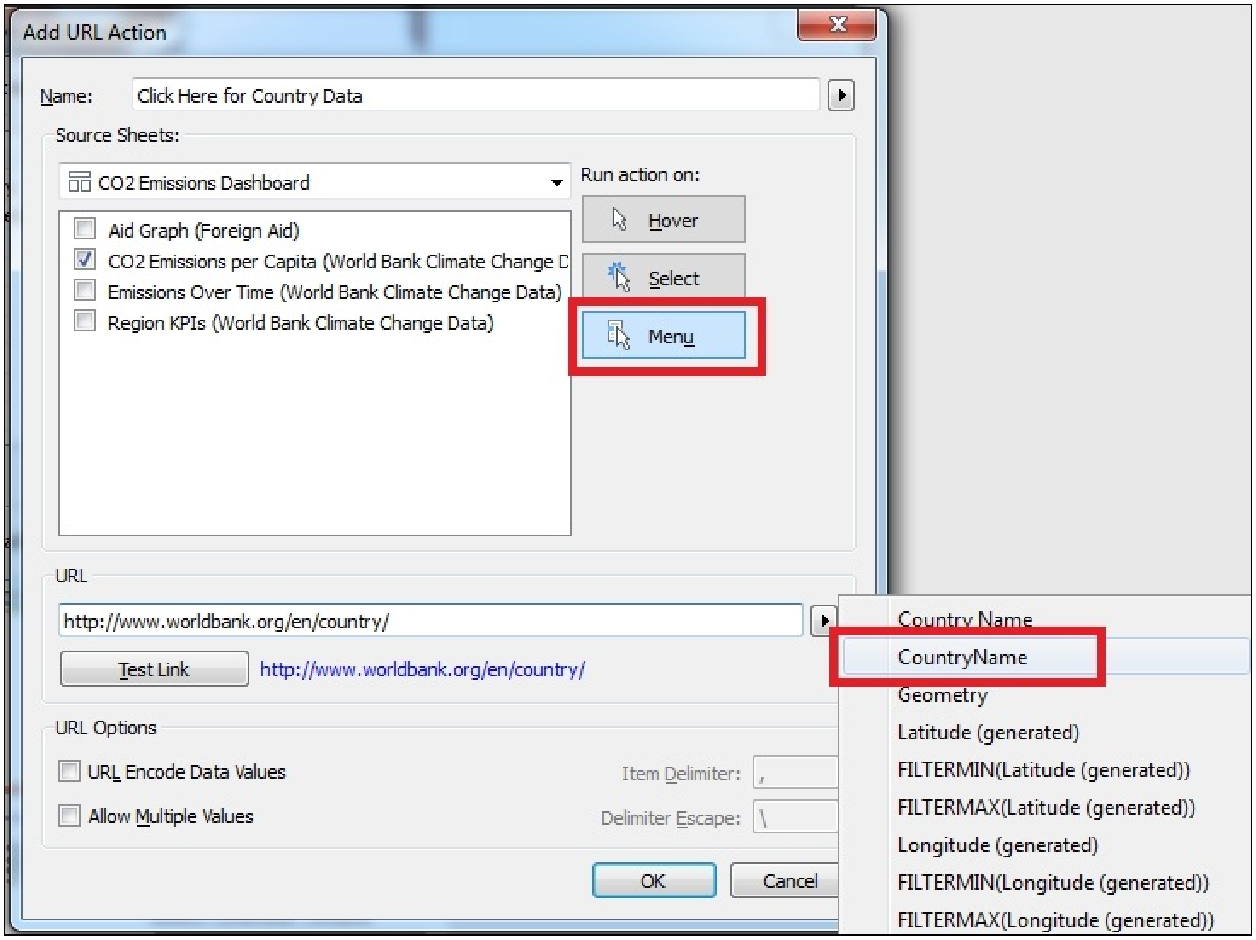
1. Go to the **Map** worksheet and create a **Calculated** field in the **Data** source.
2. Name the **Calculated** field **CountryName**.
3. In the **Calculated** field, write a formula that uses the **Replace** function to remove the spaces in the **Country** field, as shown in the following screenshot. You need to remove the spaces because in the URL, for each country, there are no spaces. Different websites treat spaces differently, depending on their information management strategy and encoding; others might replace spaces with `%20`:



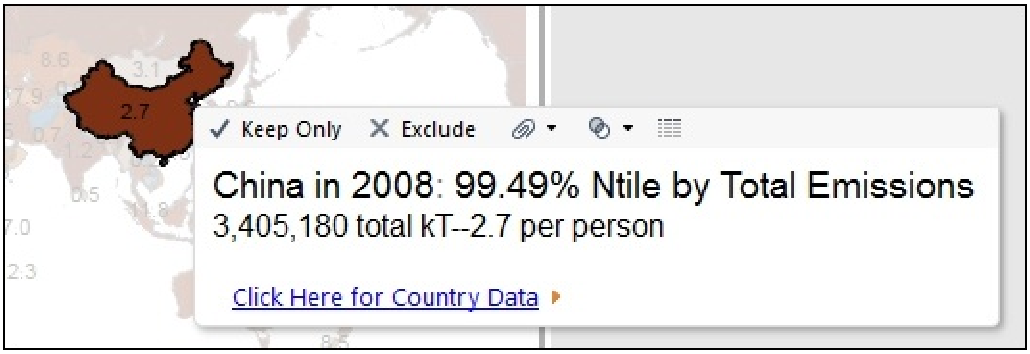
- Add this field to the **Detail** shelf on the sheet.
- Note that the percentile table calculation that generates the color for each country no longer seems to work because it was addressing only by **Country Name** before and it wasn't partitioned (or grouped) by any other values.
- In order to solve this problem, click on the context menu for the field on the **Color** shelf.
- Click on **Edit Table Calculation**.
- In the **Running Along** drop-down list, select **Advanced**.
- Click on the **CountryName** field in the **Partitioning** pane and then click on the **right arrow** to move it over to the **Addressing** pane, as shown in the following screenshot:



- Click on **OK**.
- Go back to the dashboard. From the **Dashboard** menu, select **Actions**.
- Create a new **URL** action.
- Name the **URL** action with the text that will prompt users to perform a specific action, such as **Click on Here for Country Data**. You want the result to be very obvious because this action will be a **Menu** action, which means that the name of the action will appear as a call to an action at the bottom of the tooltip, as shown in the following screenshot:



- Select **CO2 Emissions per Capita** as the source sheet.
- Run the action on **Menu**.
- Paste the root URL for the World Bank into the **URL** box. Then, click on the arrow to the right of the URL box to see a list of fields that you can insert as a variable token.
- Select **CountryName** from the list.
- Test the link by clicking on the **Test Link** button and, if it works, click on **OK**.
- Click on **OK** again.
- Click on a **Country** (**China**, for instance) and check out the call to action at the bottom of the tooltip, which you unfortunately cannot format differently, that guides the user to click on **Click Here For Country Data** to see country data in detail, as shown in the following screenshot:


Keep Only
Exclude
China in 2008: 99.49% Ntile by Total Emissions
3,405,180 total kT--2.7 per person
Click Here for Country Data
2.7
3.1
8.6

Summary

In this chapter, you learned how to use **Quick Filters** and **Dashboard Actions** not only to provide greater interactivity for users, but also to provide an efficient user experience. You also learned how to create parameters to filter single and multiple data sources, as well as to govern the visualizations that appear on a dashboard. You studied the implications in dashboard performance when using different methods of filtering, all of which will give your dashboard users a greater control over the data points that they choose to explore. It will also create a more personal, relevant, and compelling data story.

In the next chapter, we will discuss publishing your work, which is the last step to creating and distributing your work in Tableau Public.

Chapter 9. Publishing Your Work

Publishing your work on the Cloud with Tableau Public is the core of the application's value proposition, and with Tableau Public 9.0, it's easy to save and share your work. There's more to publishing than just locating a file on a web server. Publishing your work and managing your author and workbook profiles is the best way of helping people browse through your visualizations to create more personal connections with your data story. When people are personally connected and relate to you and your work as a human, they are more likely to share it.

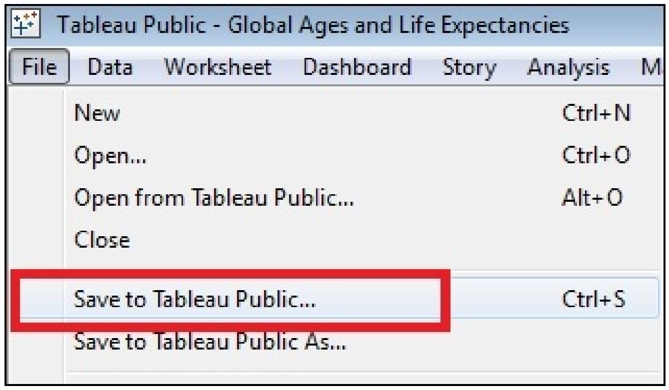
In this section, we'll discuss the following topics:

- Saving your work and logging in to Tableau Public
- Opening your work from the Cloud on your computer
- Managing your profile
- Viewing your work online
- Sharing your work with others
- Managing workbook details

Saving your work and logging in to Tableau Public

Since Tableau Public does not have autosave, we recommend that you save to the Cloud frequently.

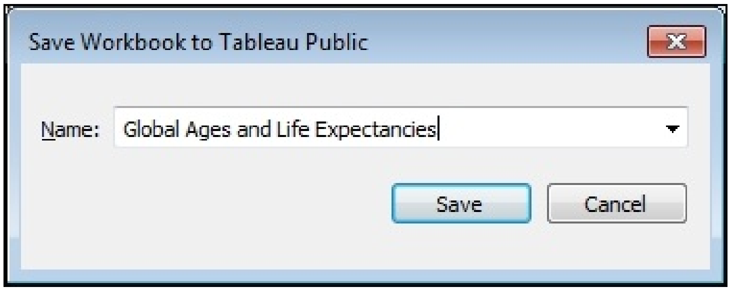
In the **File** menu, you can save your work by clicking on the **Save to Tableau Public...** option, which is present halfway down the list. Alternatively, you can use the same shortcut keys that other applications use (*Ctrl* + *S*):



Once you opt to save your work, you will need to sign in to Tableau Public. Use the same credentials that you used to download Tableau Public initially, as shown in the following screenshot:

Once you have signed in successfully, you can save your work on the Cloud with a name of your choice.

Giving your analysis a good name is critical. First, it's the title that other people will see when they look at your profile on Tableau Public. Also, the name of the workbook and individual dashboards are a part of the URL of the file. It's important not to include special characters, such as ampersands, question marks, and periods, because they can disrupt the format of the URL. It's fine to include numbers though. You can name your workbook as shown in following screenshot:



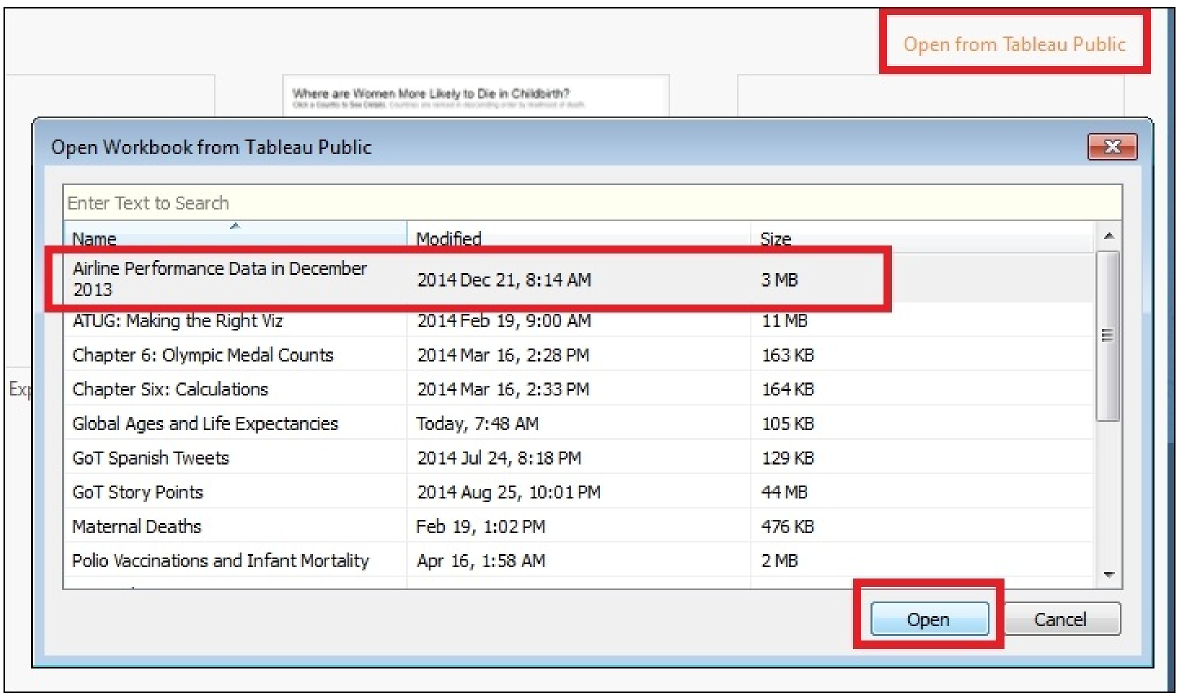
If you want to replace an existing file with the one that you have opened now, you can perform the following steps:

1. Select **Save to Tableau Public As...** from the **File** menu.
2. From the dropdown list of the existing files, select the file that you want to overwrite.
3. Click on **Save**.

Opening work from the Cloud

If you'd like to make changes to your work, you need to perform the following steps:

1. Open Tableau Public on your computer.
2. Navigate to the Tableau Public home screen.
3. Click on the orange **Open from Tableau Public** link on the upper-right corner.
4. Log in with the credentials that you used to create your account, which was covered previously in this chapter.
5. Select the workbook that you want to open, as shown in the following screenshot:



Managing your profile

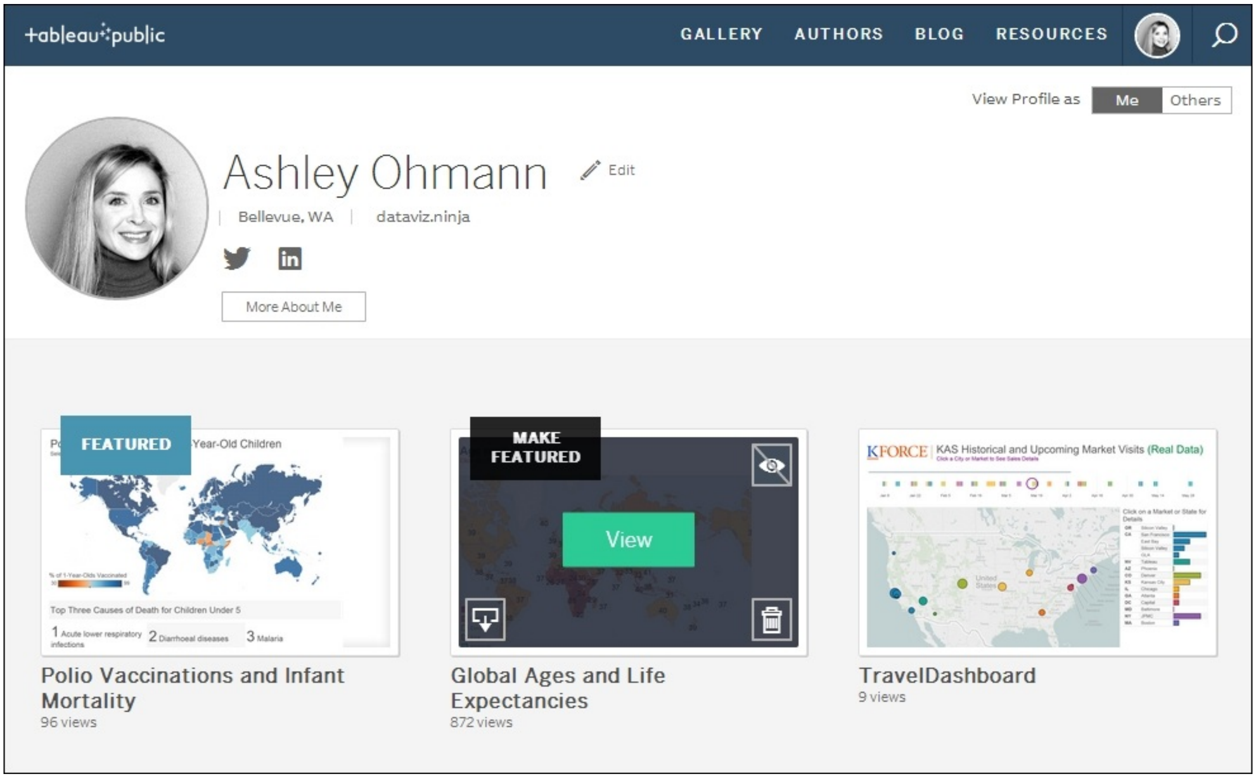
In order to view your work on the Internet, you need to go to Tableau Public's website and log in. From there, you can share your work with others, download it, promote it on social media, and also delete it, should the need arise.

1. On your Internet browser, go to http://public.tableau.com.
2. In the upper-right corner, click on **SIGN IN**, as shown in the following screenshot:

- Enter the credentials that you used to create your account.

The destination page that you reach when you login is your profile page. This is where all the dashboards that you created are organized. It is also a great place for you to add information about yourself. You can add links to your Twitter and LinkedIn accounts as well as the blogs or corporate websites that you would like to promote. (It's important to make sure that you have the permission before presenting URLs of organizations of which you're not an associate).

My profile is shown in the following screenshot:



You can change settings related to specific workbooks by hovering over them. As your catalogue of work grows, you should curate it by being selective about which visualizations are featured.

When you roll over any of the thumbnails for a specific visualization, you can do several things; the following are a few of these things:

1. First, on the upper right-hand side, select **MAKE FEATURED** to make it the featured visualization, which will move it to the upper-most left spot on your profile.
2. On the upper right-hand side, you can change the **visibility** settings.

Note

This does not prevent people from searching for your visualization, and you also cannot control the permissions regarding who can see it.

3. On the lower right-hand side, you can delete workbooks. This is a good idea in case you would like to delete the earlier iterations of a finished product, but keep in mind that you don't have the file saved on your computer. So, be judicious when using the **Delete** button.
4. On the lower right-hand side, you will see the workbook name and the number of views that it has.

> ## Note
>
> A view is each visit to the page by a browser. So, if you go to the workbook five times a day, that's five views. Note that this is not the number of unique visitors.

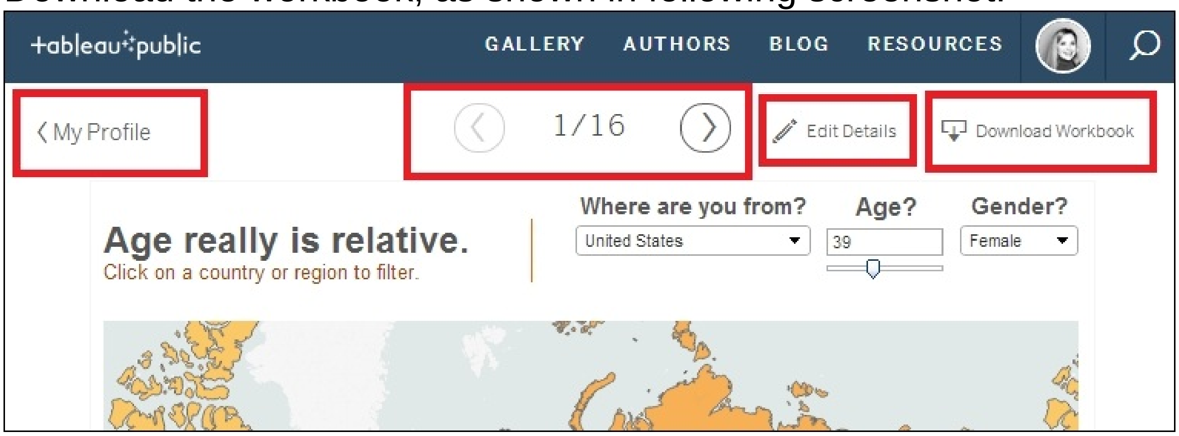
5. Then, above the workbook name, you can download the workbook.
6. At the center, you have the **View** option in case you wish to view the workbook, as shown in the following screenshot:

Viewing your work online

When you view a workbook, you have several options at the top of the page.

From left to right, you can perform the following tasks:

- Go back to your profile
- Keep on clicking and go through other workbooks
- Edit the details of the workbook that you have opened
- Download the workbook, as shown in following screenshot:



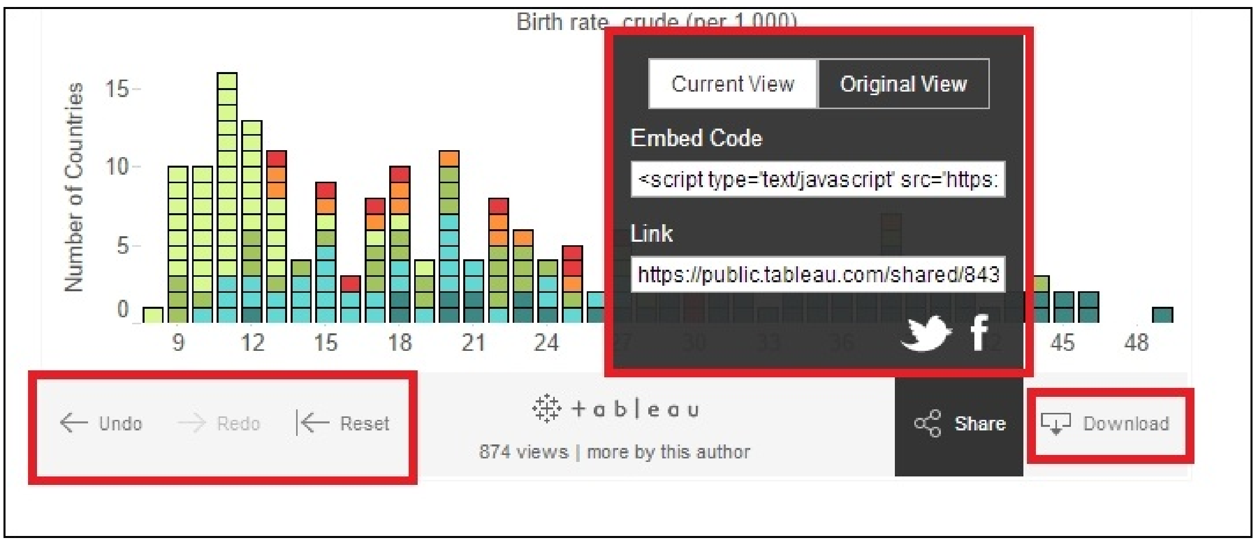
If you scroll to the bottom of the dashboard, you have the following additional controls:

1. **Undo**, **Redo**, and **Reset** any changes that you have made. For instance, if you have filtered your workbook and want to revert to the original state, just click on the **Reset** button.
2. Share your work in the following way:

    - You can share the **Current View**, which includes filter conditions or parameter selections. Alternatively, you can share the **Original View**.
    - **Embed Code** allows you to copy an automatically generated block of HTML that you can use to embed the visualization on another web page. This code identifies the viewing attributes of your workbook, such as whether the toolbars are showing, the

height and width of the workbook, and its title. For instance, if you want to embed the dashboard on your blog, you just need to copy the code. Then, in the management console of your blog, either insert the code directly, or create and then insert a snippet of HTML. If you have the capabilities, you can utilize Tableau Public's JavaScript API to embed the visualization in a webpage and then allow users to interact with it through web objects that are native to the source page. That's beyond the scope of this chapter though.

- To share the workbook with others, you can copy the link as well as simply highlight and copy the text and then paste it into a new e-mail in the tool of your choice.
- You also can share the workbook on Twitter or Facebook by clicking on either of the two buttons that open a pop-up on your browser and ask you to enter your credentials for the medium that you have selected, as shown in the following screenshot:



When you share your work via Twitter or Facebook, the description that you enter in the workbook details will automatically be populated, as shown in following screenshot. You can always change the description and remove the `via @tableau` string, but it's much more likely that people will find your tweet if you leave that tag:


Share a link with your followers
Curious how old you would be if you'd been born in another country?
https://public.tableau.com/shared/843HPJ5BT?:display_count=yes via @tableau
36
Tweet

Managing workbook details

The information that you provide about your work is important; it's the first part of the story that you're telling that people will take in. The workbook name is important, and so is the description and additional links.

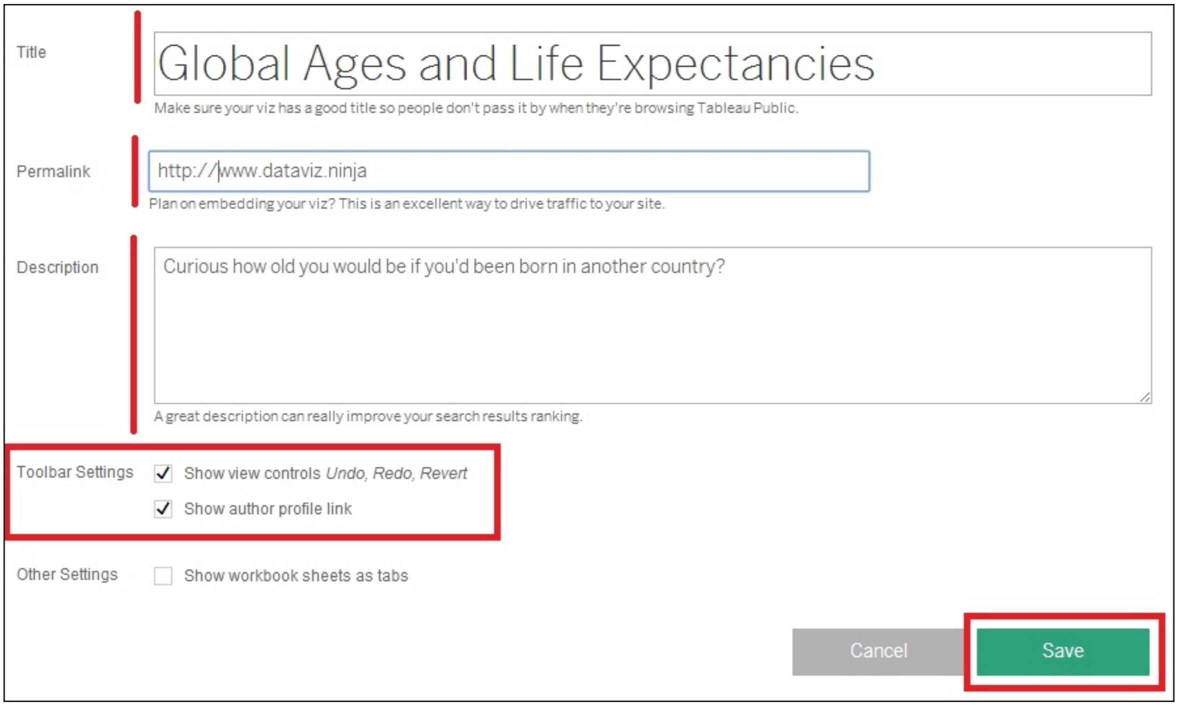
If you scroll down the page past the **Share** button, you will see the workbook title as well as a link that allows you to edit the workbook details, as shown in the following screenshot:

Now, perform the following steps:

1. Make sure that the Title of the workbook says exactly what you want to appear on your profile.
2. Add a **Permalink**. Though this isn't required, it's a good way to promote your blog.
3. Enter a good description in the Description. This is the text that search engines will index. Also, if you want people to tweet your work, then you need to enter a description that's 82 characters or less, because the link that Tableau generates for your workbook will consist of 62 characters.
4. Toolbar Settings typically include control buttons as well as a link to your author profile. If you choose not to include those, then the HTML generated to embed your work will be adjusted.
5. An additional setting that you can adjust is **Show workbook sheets as tabs**. If all the worksheets in your workbook are on your dashboard and you have hidden them, then you don't need to select

this.

6. Save your work by clicking on **Save**, as shown in the following screenshot:



Summary

In this chapter, you learned how to save your work to the Web as well as how to manage your Tableau Public profile and workbook attributes to help you showcase and share your work effectively.

This chapter concludes our discussion of creating data stories in Tableau Public. Throughout the previous eight chapters, we progressed through the basics of the tool, data analysis, and different visualization types, all the way through sophisticated calculations and parameters. You now have the skills to create rich, relevant, and compelling data stories from which others can learn, get inspired, and find the information that they need to improve their own lives and those of others.

Appendix A. Bibliography

This course is a blend of different recipes and texts all packaged up keeping your journey in mind. It includes the content from the following Packt products:

- *Tableau Dashboard Cookbook - Jen Stirrup*
- *Tableau Data Visualization Cookbook - Ashutosh Nandeshwar*
- *Creating Data Stories with tableau Public - Ashley Ohmann, Matt Floyd*

Index

A

B

D

G

M

N

O

P

R

S

T

U

V