Tableau Desktop tutorial

Professor Roy A. Ruddle, School of Computing, University of Leeds.

September 2019

Contents

1 Introduction

This tutorial will help you to learn how to create visualizations using Tableau Desktop, which is an industry-leading information visualization application that grew out of Stanford University (see http://careers.tableausoftware.com/ourstory).

In the tutorial, Datasets and other files are written in "quotation marks", variables are underlined, and Tableau terminology and interface components are *italicised*.

Section 2 provides an overview of Tableau's layout, terminology and icons.

Section 3 shows you how to get started, and create worksheets, dashboards and stories.

Section 4 introduces you to a variety of chart types, types of calculation you can perform to create new variables on-the-fly, and some useful interactive functionality.

Once you have completed all of this tutorial you are ready to work out for yourself how to visualize data by tackling some *Tableau Challenges*, which are described in a separate document!

2 An overview of Tableau Desktop

A Tableau worksheet is divided into two main areas – the *data pane* and an area for *shelves and cards* (see Figure 1). The data pane is divided into sections for *dimensions*, *measures*, *sets* and *parameters*.

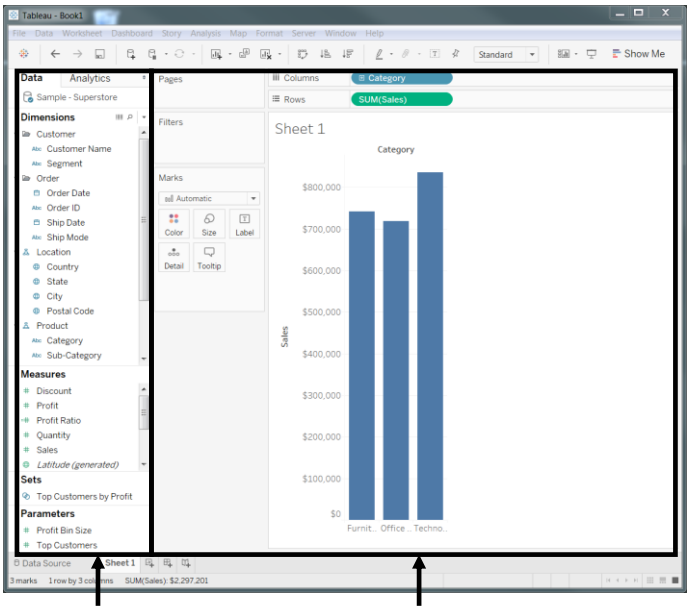


Figure 1: The data pane, shelves and cards on a Tableau worksheet.

Measures are data fields that contain numerical values, with which you can perform mathematical calculations. Dimensions are all other data fields (categorical, ordinal, date, etc.). The icons that Tableau uses for data types are shown in Table 1.

When you drag a field to the columns or rows shelf the field is shown as a *blue* or *green pill*. (see Figure 1). Blue pills are discrete (a finite number of values) whereas green pills are continuous (in principle they contain an infinite number of values). Most dimensions are discrete, but they can be continuous (dates can be discrete or continuous). Most measures are continuous, but they can be discrete (e.g., the number of items that a customer purchased). The types of your data dictate the types of chart that you can create.

Icon	Data type
Abc	Text
	Date
	Date/time
#	Numerical
T|F	Boolean
	Geographic

Table 1: Tableau's data types (see http://onlinehelp.tableau.com/current/pro/desktop/en-us/datafields_typesandroles_datatypes.html).

3 Getting started with Tableau Desktop

If you didn't do this before the workshop, work through Steps 1 – 5 of "Getting Started with Tableau Desktop" (http://onlinehelp.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.html).

3.1 Connect to your data

In this step you will connect to an example dataset ("Sample – Superstore") about retail sales.

3.2 Drag and drop to take a first look

In this step you will create a line chart showing year vs. sales (see Figure 2a) and then sub-divide the sales in each year by the product category (see Figure 2b). The chart types changes automatically because product category is discrete. However, if you swap Year and Category then then line charts are shown again, because year is continuous (see Figure 2c). Tableau's *Show Me* functionality shows all of the types of chart that may be created (see Figure 2d).

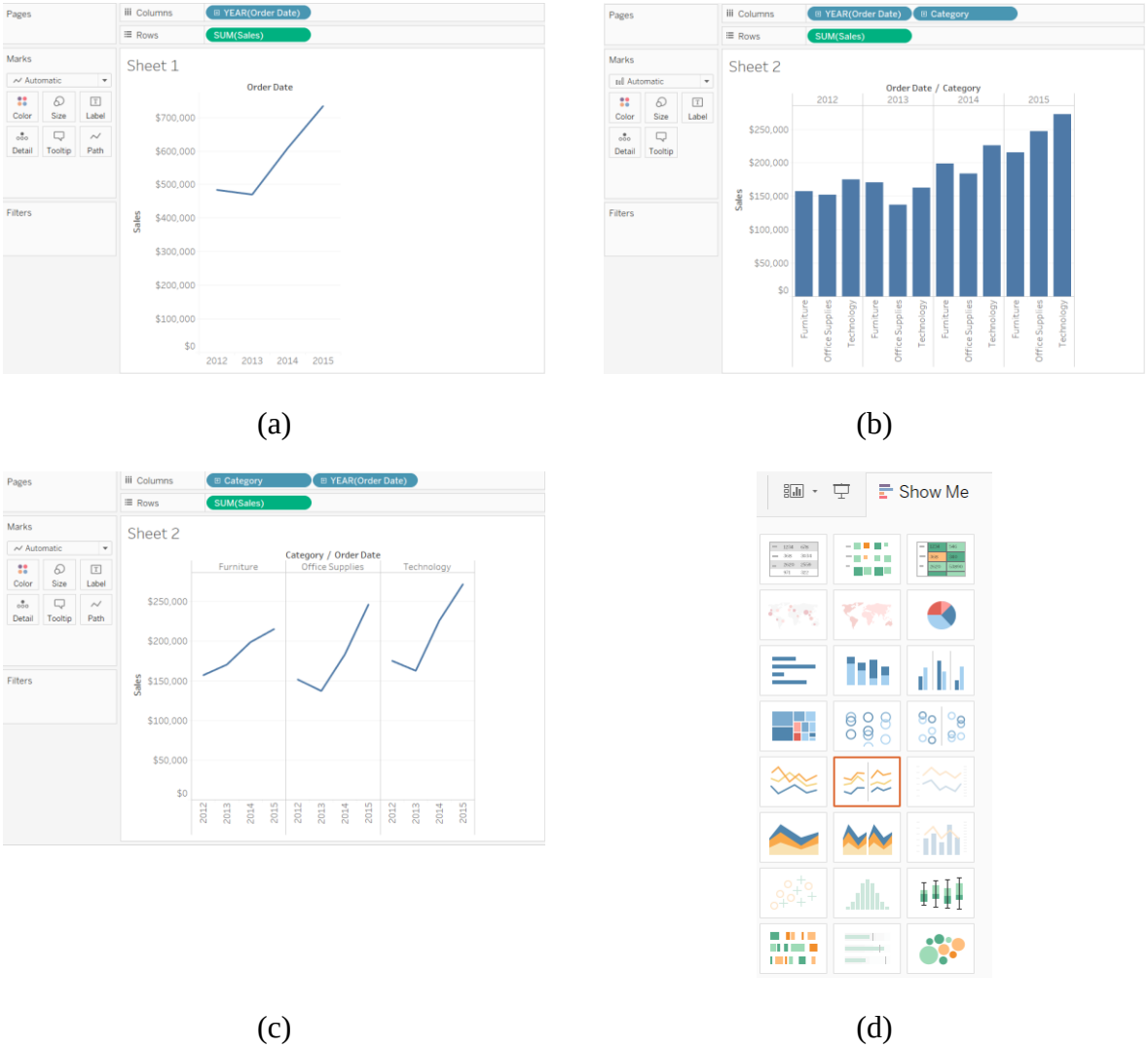
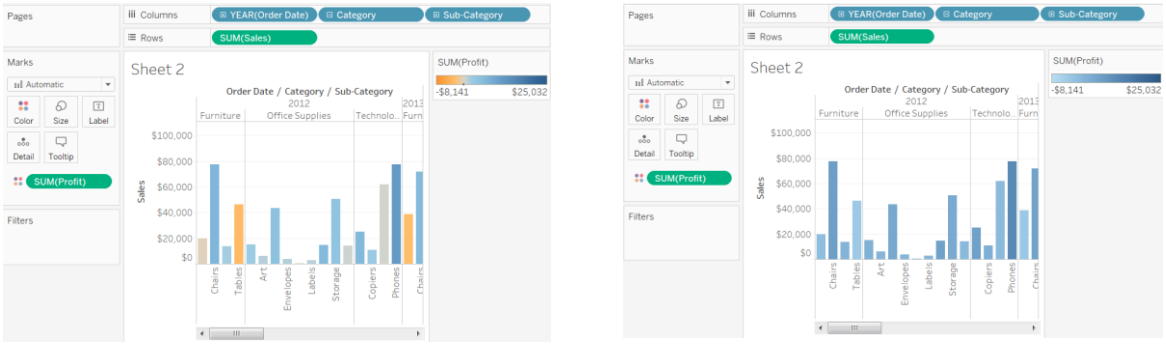


Figure 2: Sales shown as: (a) a line chart for year vs. sales, (b) a bar chart for year/category vs. sales, (c) a line chart for category/year vs. sales, and (d) the Show Me panel for choosing chart types.

3.3 Focus your results

In this step you will add interactive filters and use colour to distinguish profitable products from those that are unprofitable.

The colour map affects the insights that you can gain. Tableau automatically assigned a diverging colour map because the Profit data contained both negative and positive values. However, unprofitable products are much harder to spot if you use an ordinary (single-hue) colour map (see Figure 3), which you can display by selecting *Edit Colors* from the *SUM(Profit)* legend.



(a) (b)

Figure 3: The same chart with: (a) a diverging colour map, and (b) a single-hue colour map.

3.4 Explore your data geographically

In this step you will learn about symbol and filled maps, and Tableau's *Undo* and *Redo* buttons which should encourage you to explore your data and try out alternative designs of visualization.

3.5 Drill down into the details

In this step you will learn about some of the powerful functionality that allow you to interactively investigate detailed, and sometimes hidden, patterns in your data.

You will also learn that you need to keep your wits about you, because Tableau will do what you told it to do and that is sometimes different to what you intended! The reason is that Tableau operates like a database system, issuing a sequence of queries to generate the data that is shown in each chart. Sometimes that query logic is different to your mental model, to the extent that sometimes the result is simply 'wrong'. A good example is where you filtered the data to show the five cities that have the lowest profit, but six cities were displayed (see Figure 4).

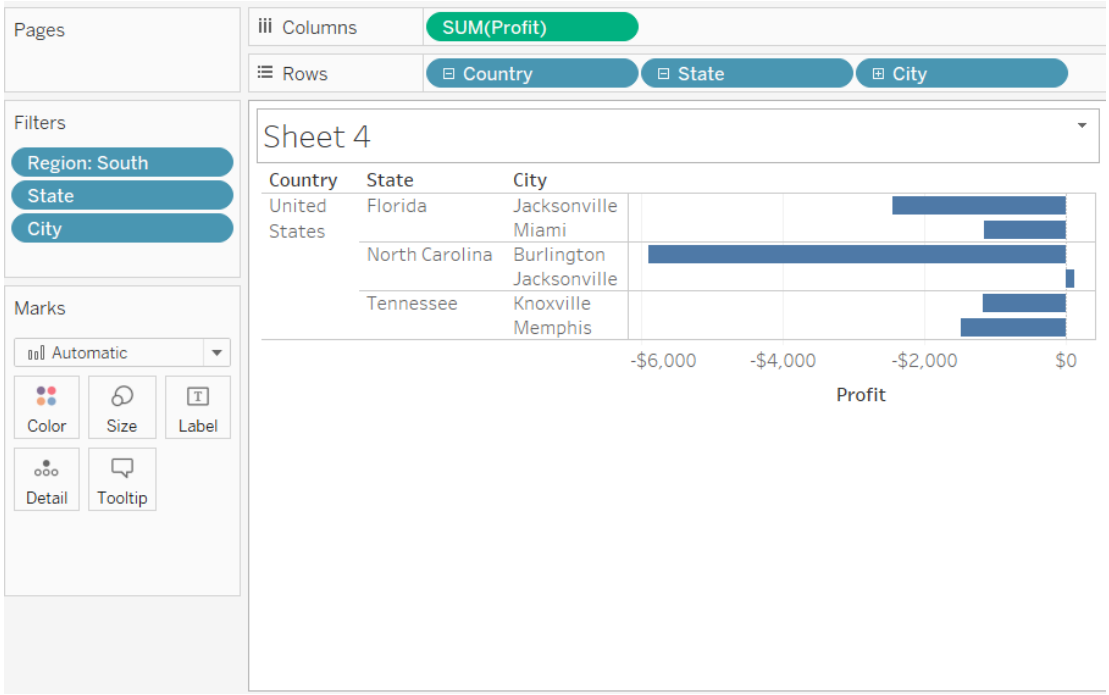


Figure 4: Why does filtering to show the bottom 5 performers actually display six cities?

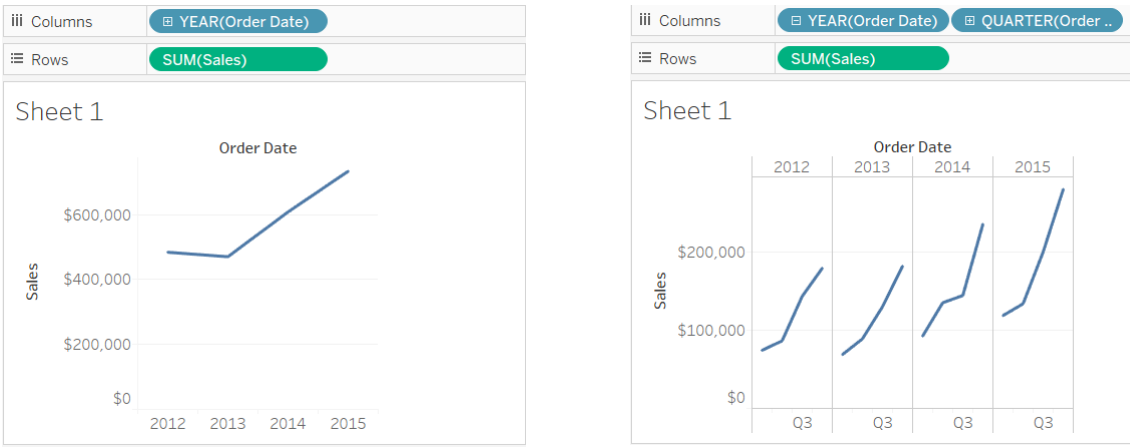
4 A tour of some additional functionality

Now that you have got started with Tableau Desktop, it is time to learn about some of the additional functionality. Start by closing your current workbook (*File -> Close*).

4.1 Date and time data

The granularity of *Date* and *Date/time* data, and whether or not it is treated as discrete (a blue pill) or continuous (green pill) can be changed interactively to help you explore your data.

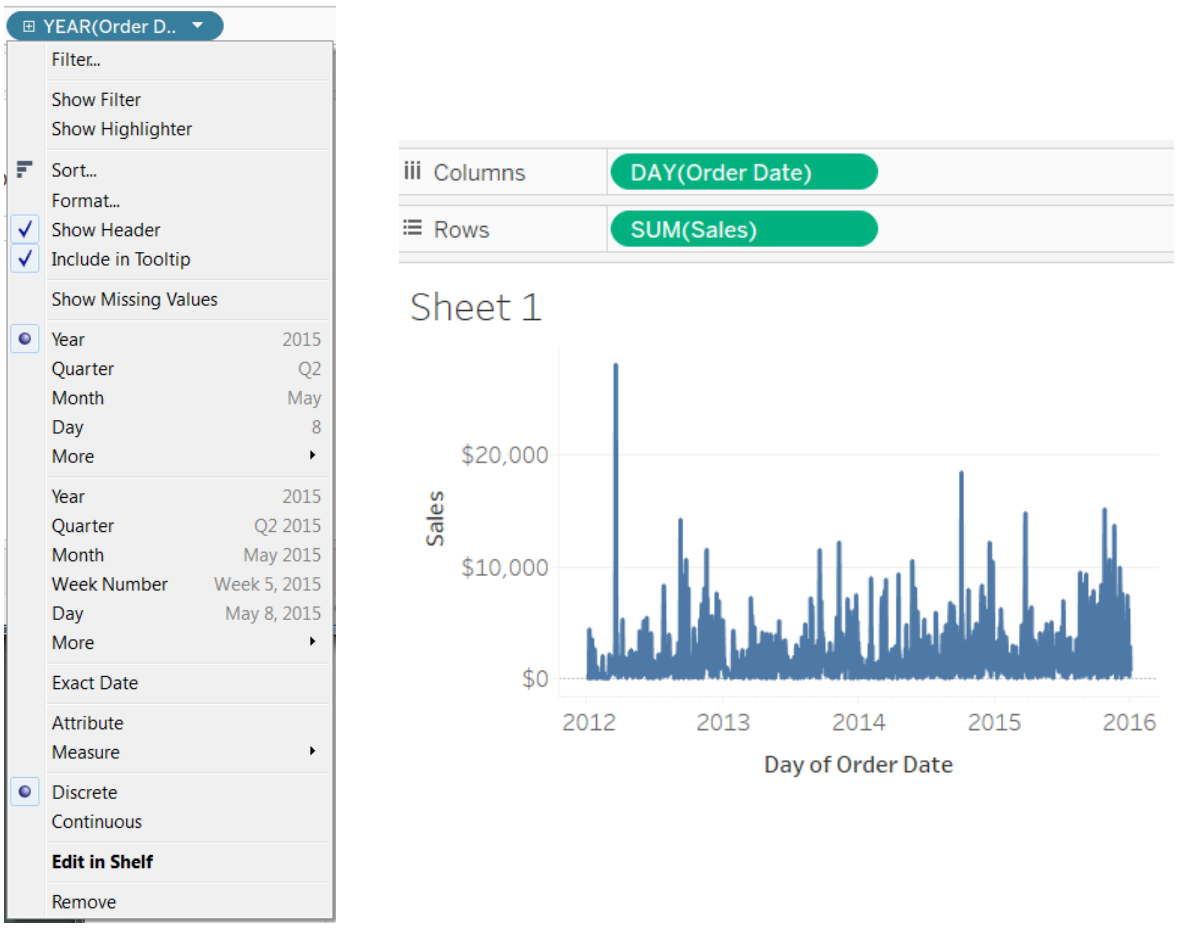
To illustrate this, close your workbook, open the "Sample – Superstore" dataset, and create a chart that shows *Order Date* vs. *Sales*. The date defaults to discrete at year granularity, but by clicking on the '+' of the Order Date pill you can make the chart finer-grained (see Figure 5).



(a) (b)

Figure 5: Sales data shown for discrete dates at the granularity of: (a) a year, and (b) a quarter.

Alternatively, you can select the drop down menu of the Order Date pill and both change the granularity and display the data as continuous (see Figure 6).



(a) (b)

Figure 6: (a) Options for displaying date data, and (b) sales data shown for continuous dates at the granularity of a day.

4.2 Chart types

Now close your workbook and open the "World Indicators" dataset.

4.2.1 Bar chart

If you select Region vs. Internet Usage then Tableau creates a bar chart. *Show Me* uses an orange rectangle to indicate that a bar chart is the recommended chart type, but several other chart types are possible (see Figure 7).

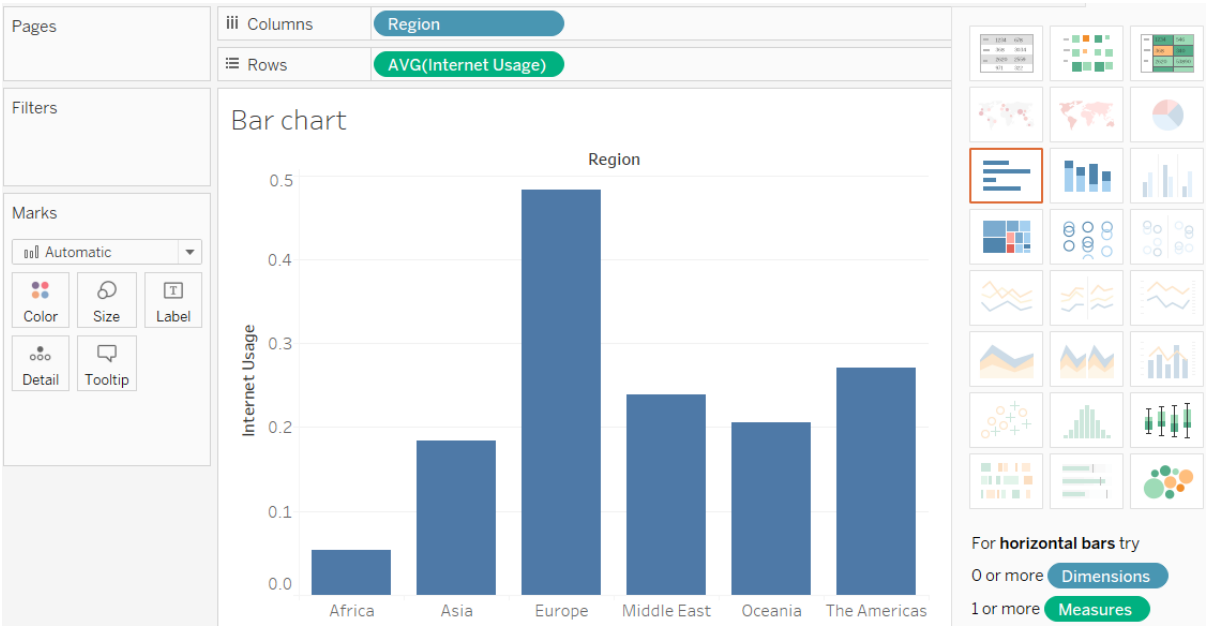


Figure 7: A bar chart, which in this case is the recommended Show Me chart type.

4.2.2 Text table

To see the values that are show in the bar chart, select *text tables* from *Show Me*.

4.2.3 Highlight table

This colours a text table according to the magnitude of the value in each cell. In other words, a highlight table is like combining a text table with a heat map.

4.2.4 Line chart (independent axes; dual-axis)

Create a line chart of Year vs. Region and average CO2 Emisssions. Tableau creates a stack of six line charts (one for each Region) and, by default, all of the axes have the same range. That helps you to make absolute comparisons between the regions, but Africa's emissions are so low you cannot see whether or not they are changing.

Double-click within the area of any of the CO2 Emissions axes to open the *Edit Axis* dialog window, and choose *Independent axis ranges for each row or column* and then *Apply*. Now you can see that Africa's emissions are rising, like those of Asia and the Middle East (see Figure 8).

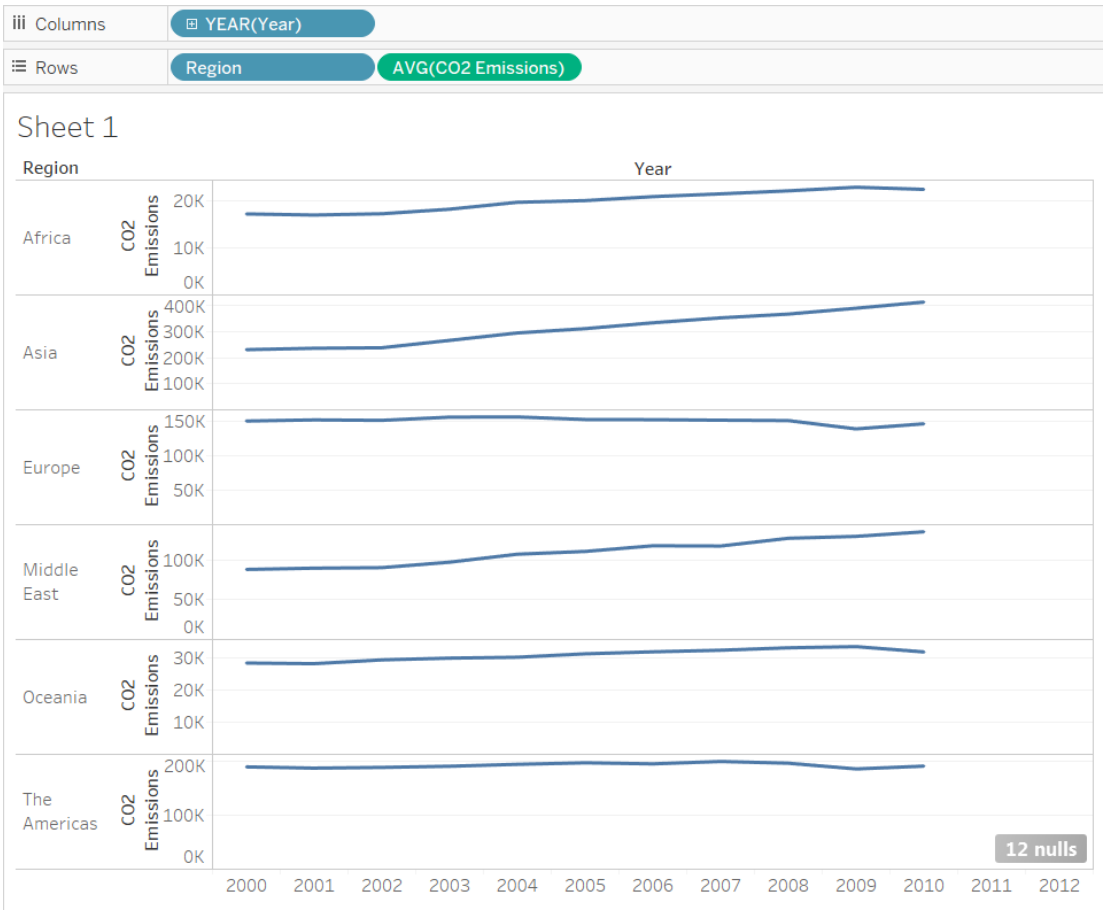
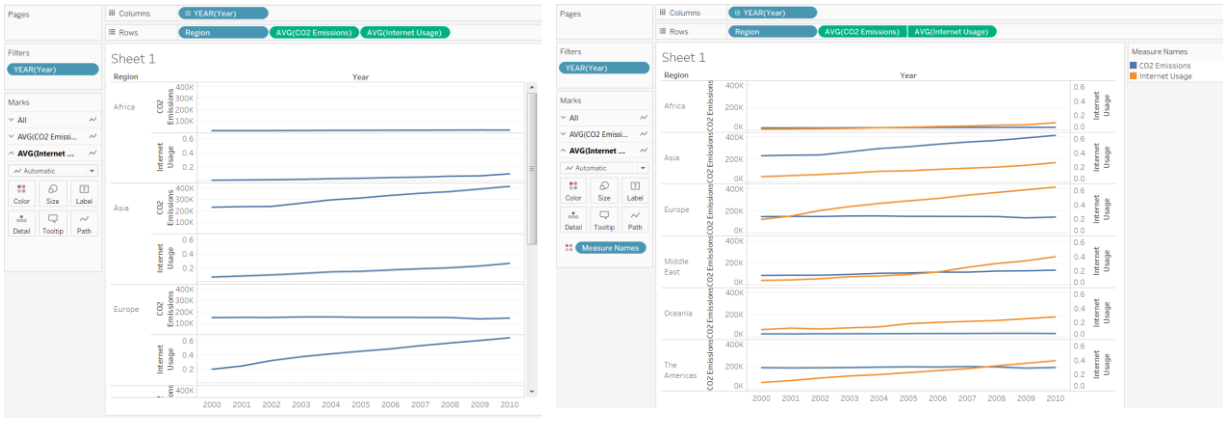


Figure 8: Using an independent axis range to show changes in each region's CO2 emissions.

Go back to a uniform axis range (press *Undo*) and add average Internet Usage to the Rows shelf. This creates a stack of two line charts for each region, which clearly displays each measure but requires a lot of display real estate so you will probably have to scroll the worksheet to see all of the charts. An alternative is to display both measures in a dual-axis chart by selecting either *Dual Axis* from the Internet Usage pill or the *dual lines* chart type from *Show Me* (see Figure 18).

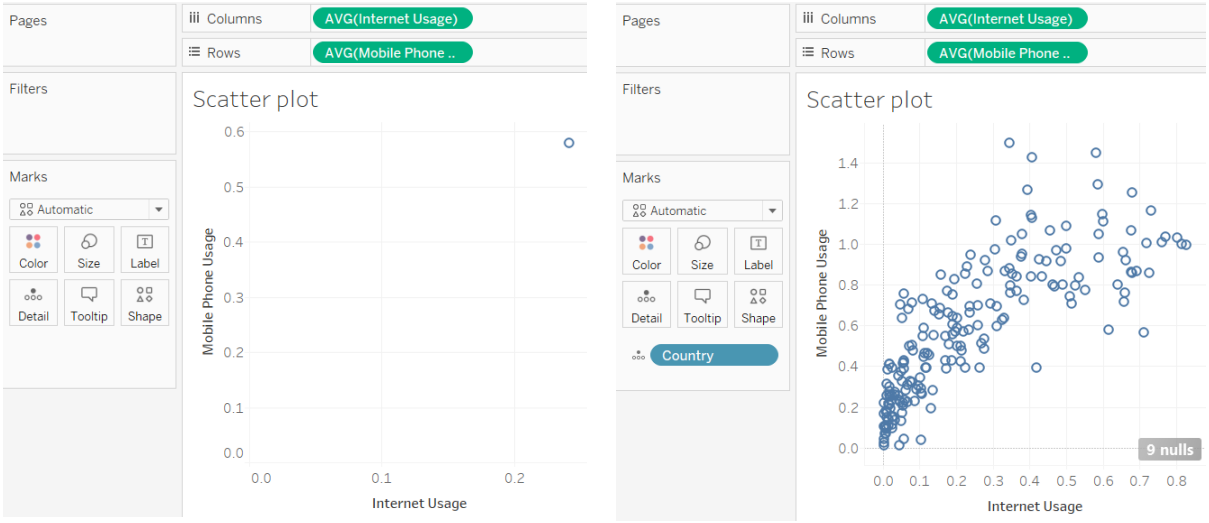


(a) (b)

Figure 9: CO2 Emissions and Internet Usage in each Region visualized using: (a) stacked charts, and (b) dual-axis charts.

4.2.5 Scatter plot

How do internet and mobile phone usage relate? One way of investigating that is to create a scatter plot, but by default Tableau only shows one data point – the average of all the values in the dataset. You know that the dataset actually contains a value for each combination of country and year, so to create a scatter plot that shows the average usage for each country drag *Country* into the *Detail* box of the *Marks* card (see Figure 10).



(a) (b)

Figure 10: A scatter plot showing: (a) the default level of detail, and (b) one data point per country.

4.2.6 Box-and-whisker plot

The scatter plot indicates that the internet and mobile phone usage data are both skewed. A box plot lets us see the distribution of the data for the countries in each region (see Figure 11).

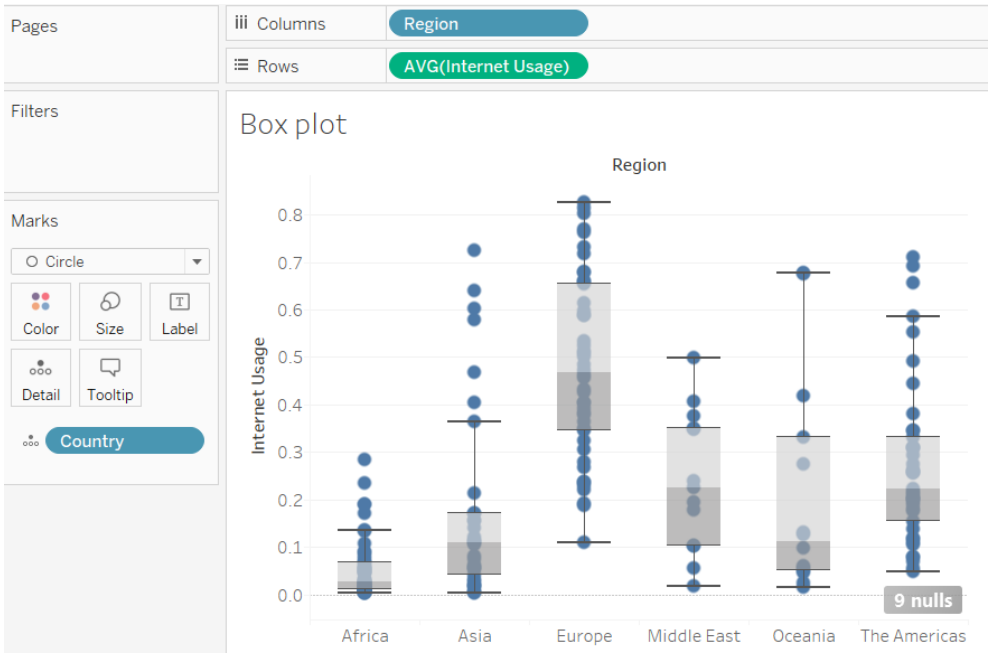


Figure 11: A box plot showing the Internet Usage for the countries in each region.

4.2.7 Histogram (with and without the Pages Shelf)

Another way of looking at data distributions is to create a histogram. To do that, drag *Internet Usage* onto the *Rows Shelf* and select *Histogram* from *Show Me* (see Figure 12).

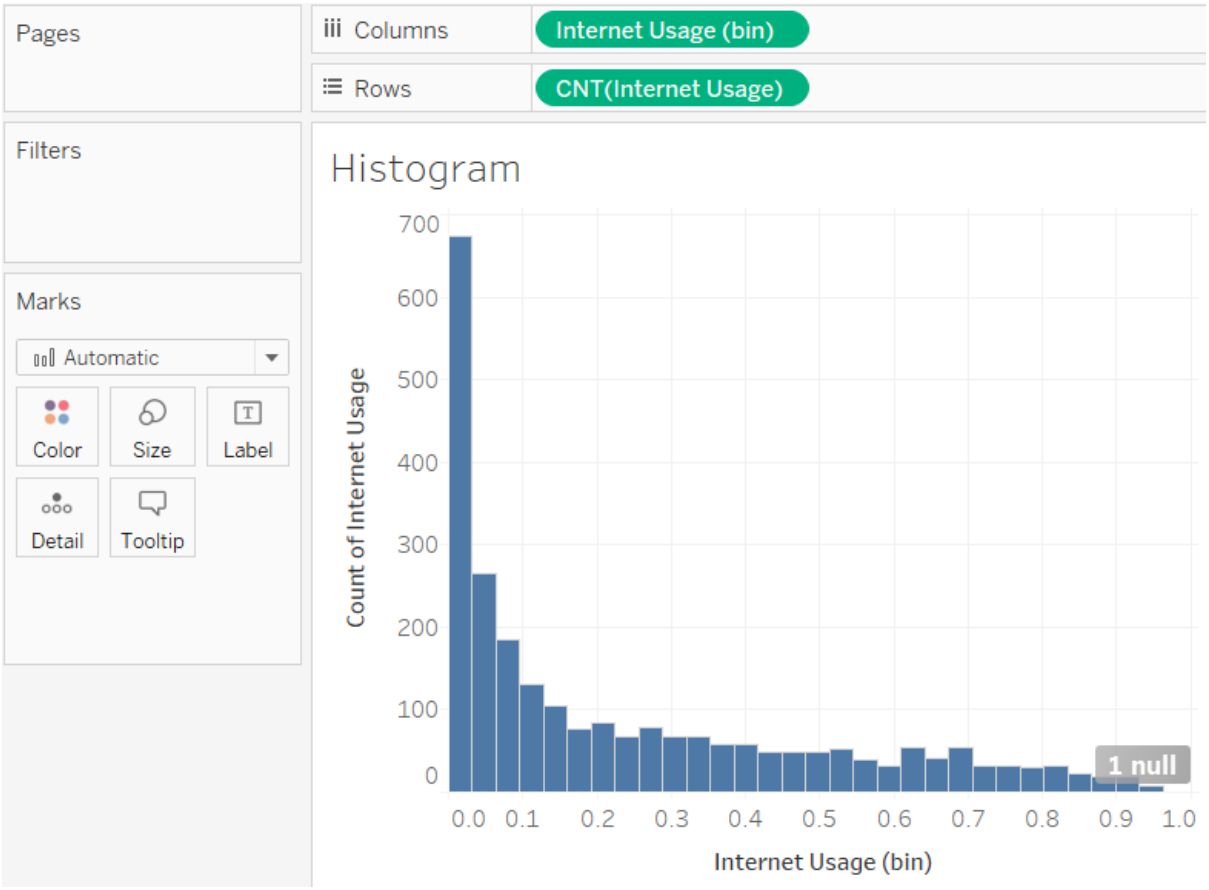


Figure 12: A histogram of Internet Usage.

That histogram shows the data for every combination of country and year. There are many ways that you can subdivide the data, and one is to show a separate histogram for each Region (drag that dimension onto the *Rows Shelf*) and animate the Year (drag that dimension onto the *Pages Shelf*). See Figure 13.

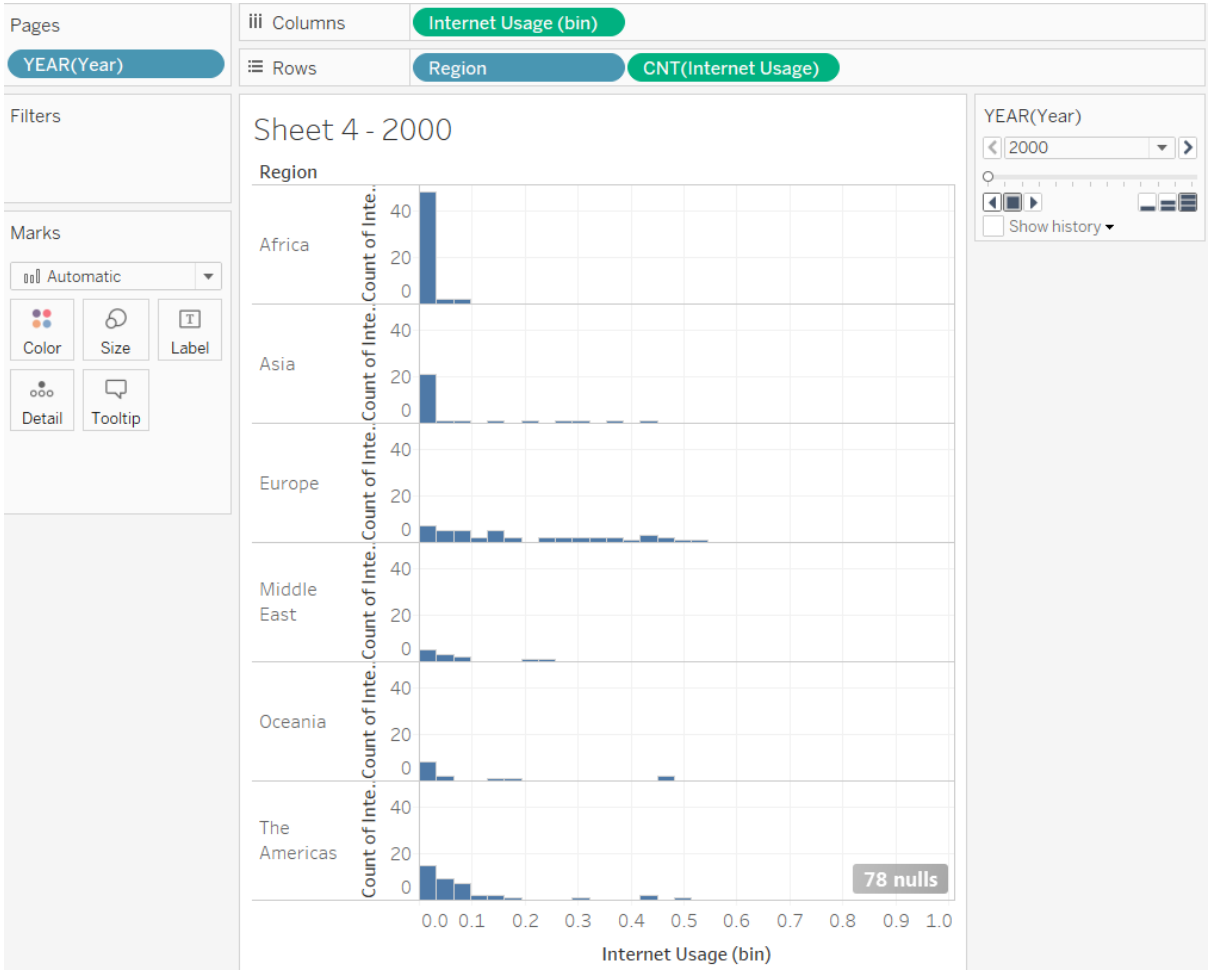


Figure 13: Using the Pages Shelf to display separate histograms for each year.

4.2.8 Heat map

Heat maps are helpful to gain an overview of how a measure (e.g., Internet Usage) varies with two dimensions (e.g., Country and Year). By choosing an appropriate colour map you can emphasise certain differences (e.g., low vs. high usage; see Figure 14).

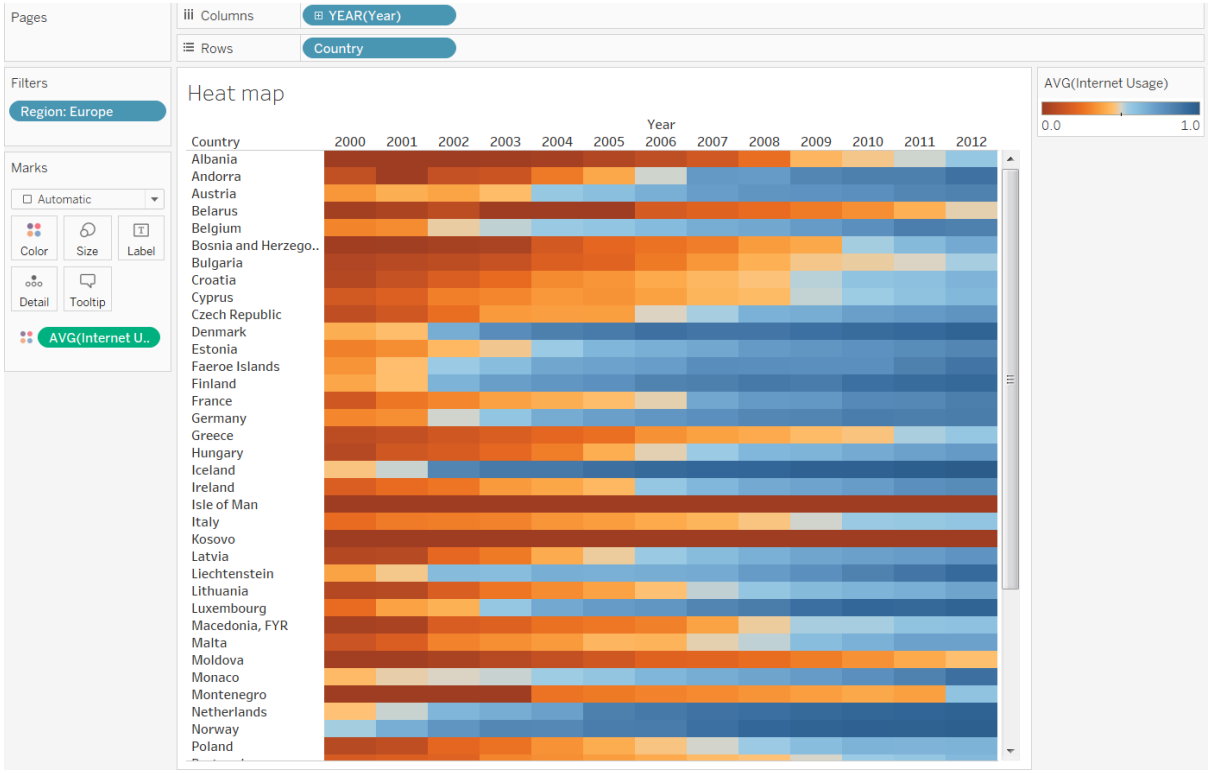
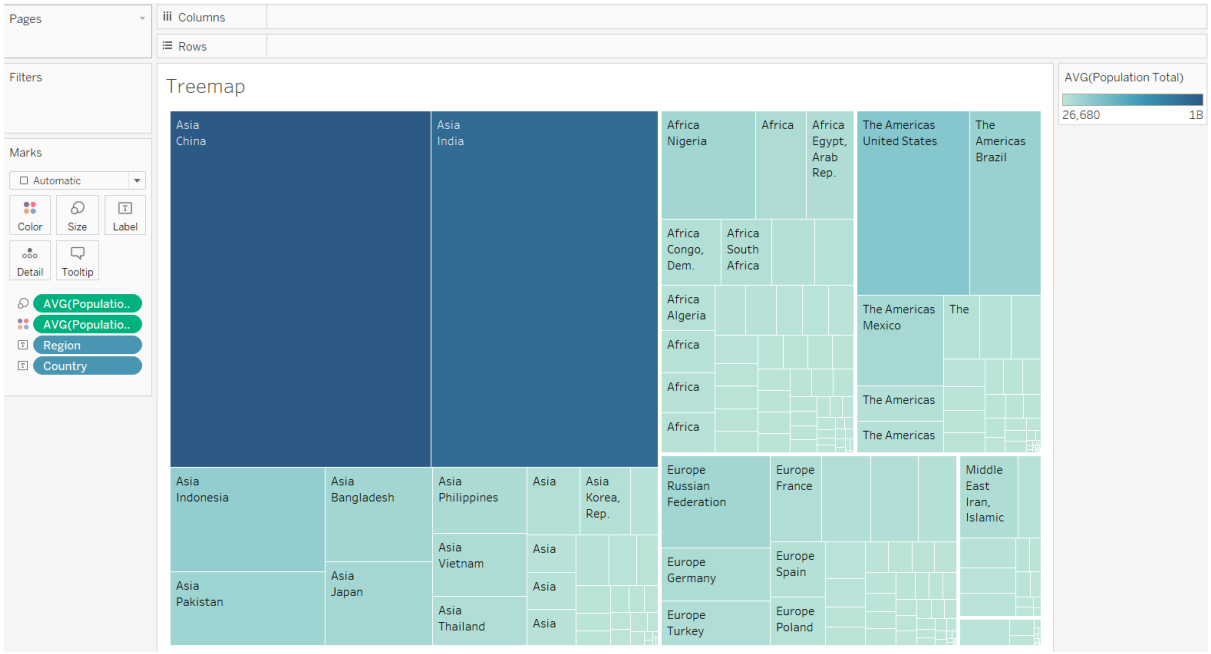


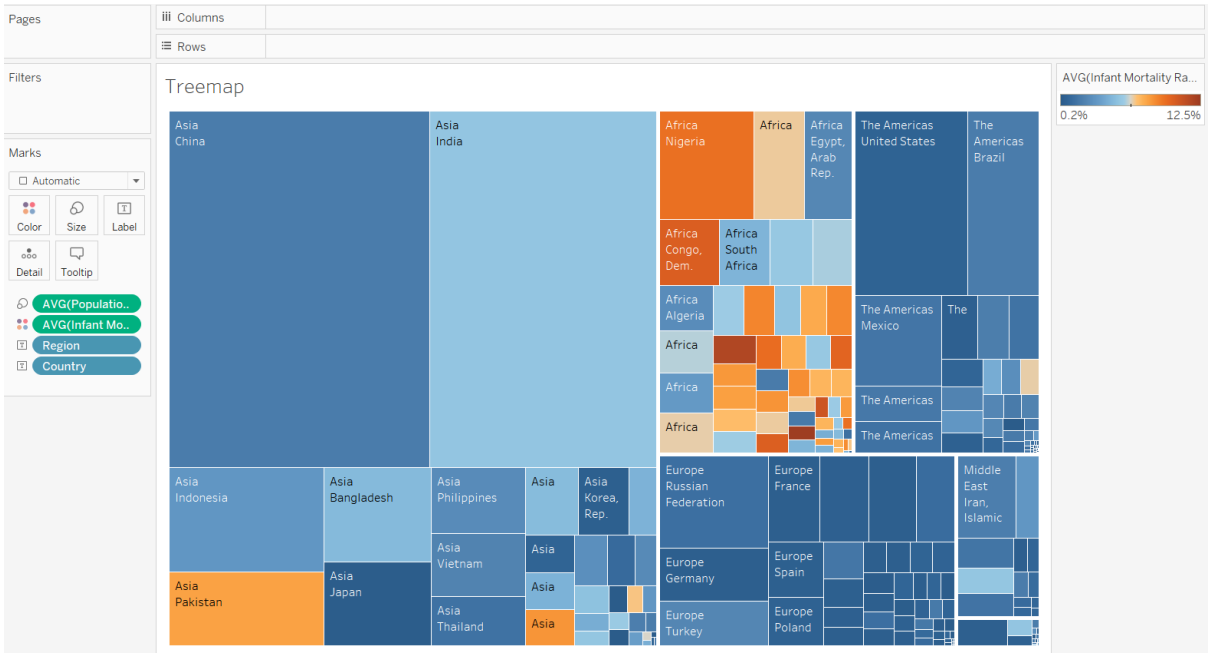
Figure 14: A heat map showing Internet Usage for European countries from 2000 – 2012.

4.2.9 Treemap

Treemaps are useful for hierarchical data, for example the population of countries in regions. By default Tableau uses both area and colour to encode one measure, but if you drag another measure onto *Color* in the *Marks Card* then you can see contrasts between the measures (see Figure 15).



(a)



(b)

Figure 15: A treemap showing: (a) the Population of each region and country, and (b) the Population and Infant Mortality Rate (with a diverging, reversed colour map high-mortality countries are emphasized in shades of red for 'danger').

4.3 Maps

Tableau can create choropleth (filled) maps, and a variety of visualizations that superimpose other types of chart and networks onto a map background (see

http://onlinehelp.tableau.com/current/pro/desktop/en-us/maps_build.html). All of these map visualizations require geocoded data, which can be problematic.

To specify that a dimension or measure is a *Geographic* data type (see Table 1), click on the dimension/measure and select the appropriate *Geographic Role*. However, **be warned** because for UK geography:

- Metropolitan counties are missing (e.g., West Yorkshire) because Tableau uses the Metropolitan Boroughs instead (Bradford, Calderdale, Kirklees, Leeds & Wakefield).
- Tableau only recognises the first segment of a postcode (the *outcode*, e.g., LS2) not the whole postcode (e.g., LS2 9JT).

To experience these problems first-hand:

- *Connect* to the "GP surgeries in Yorkshire.xlsx" spreadsheet.
- Choose the "County and Postcode" sheet
- In a worksheet, double-click on *County* and notice that
    - Only North Yorkshire is displayed
    - There are two unknown geographic locations (West Yorkshire, and South Yorkshire).
- Double-click on *Postcode – full* and notice that nothing is filled on the map ("> 189K unknown").
- Double-click on *Postcode – Outcode* and, hey presto, you see a circle in each postcode location!

Fortunately, Tableau can create maps from spatial files (https://onlinehelp.tableau.com/current/pro/desktop/en-us/maps_shapefiles.html), and here is an example:

- Close your current workbook (*File -> Close*).
- *Connect* to the "English Ceremonial Counties.shp" spatial file (it is provided with the tutorial data).
- In a worksheet, double-click on the *Geometry* measure to display the counties in a *Filled map*.
- Now colour the map according to *Area* and notice that all of the counties are the same colour. What is wrong?
    - The problem is that, by default, *Geometry* has been aggregated into a single mark using the COLLECT aggregation.
    - Fix that by dragging *Name* onto *Detail* in the *Marks* card (see Figure 16).

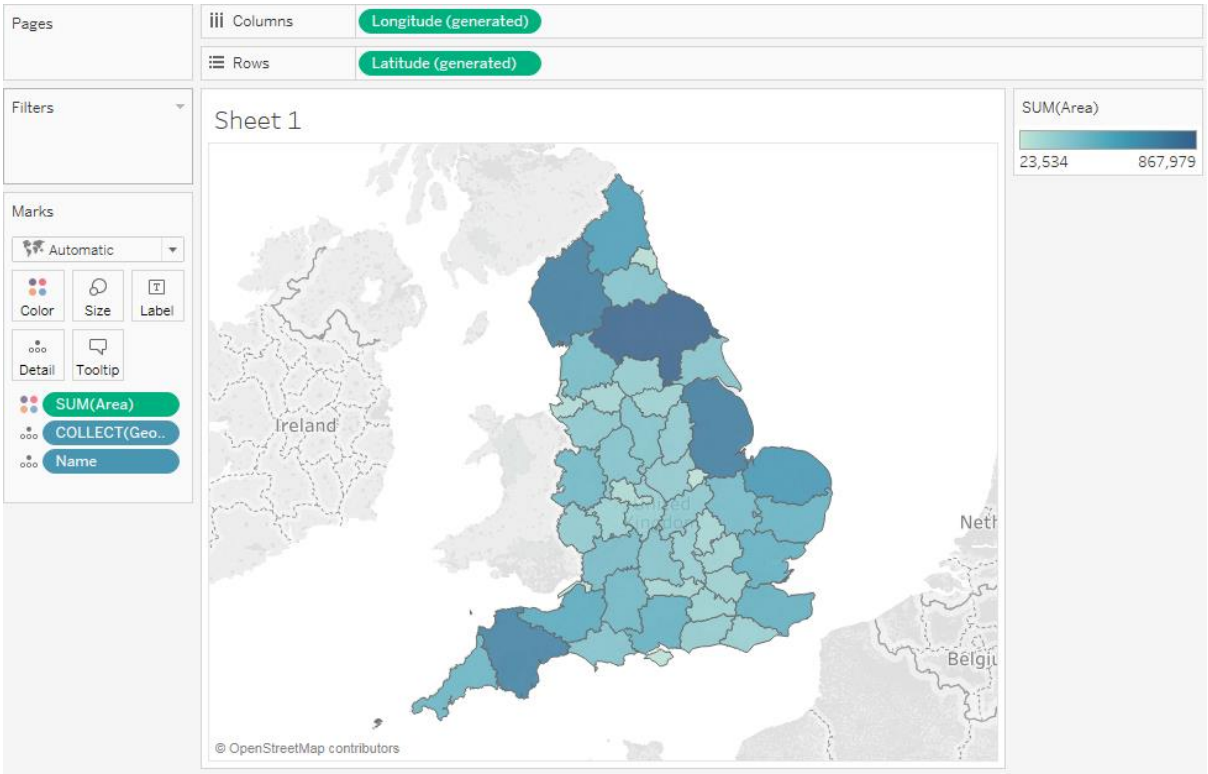


Figure 16: A filled map showing the area of each county, created from a spatial file.

- Now go back to the *Data Source* and add a connection to the "GP surgeries in Yorkshire.xlsx" spreadsheet.
- Join "English Ceremonial Counties" to the "County and Postcode" sheet (use the *Name* and *County* columns, respectively).
- Join the ""County and Postcode" sheet to the "GP surgeries and Postcodes" sheet (use the *Postcode - full* and *Postcode* columns, respectively).
- Colour the map according to the *Number of Patients* (see Figure 17).

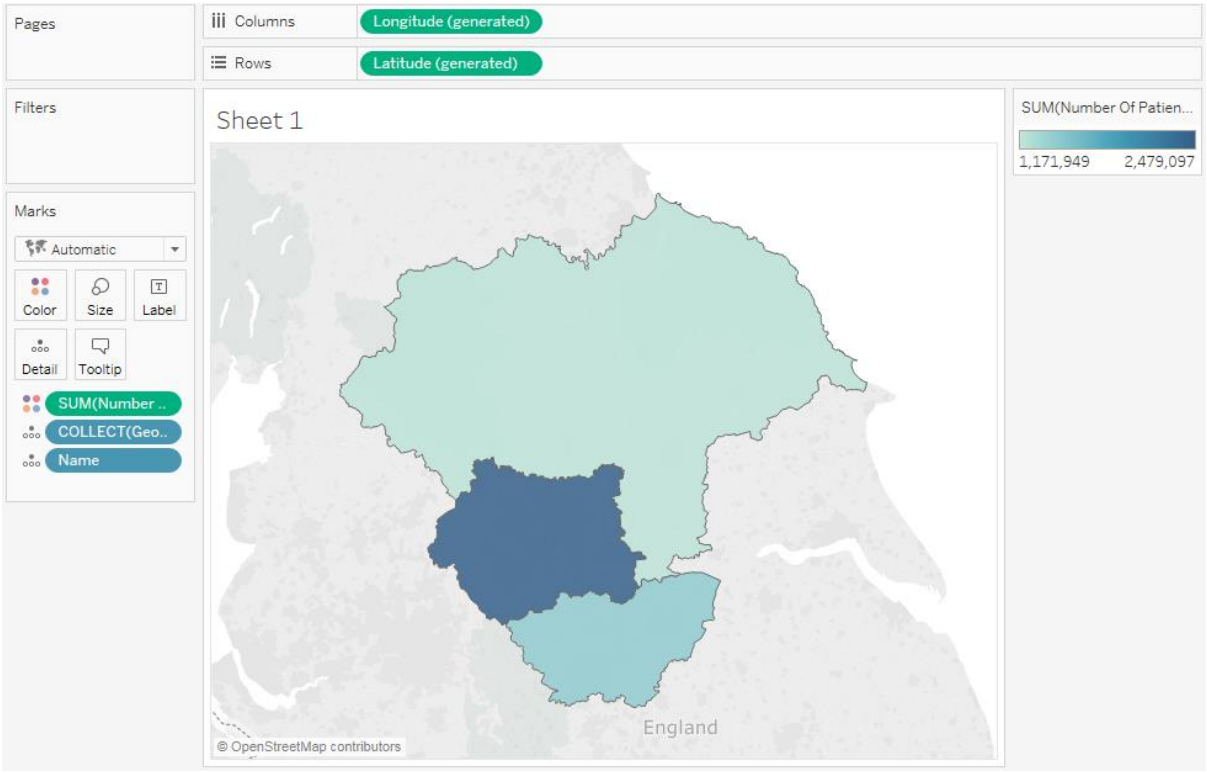


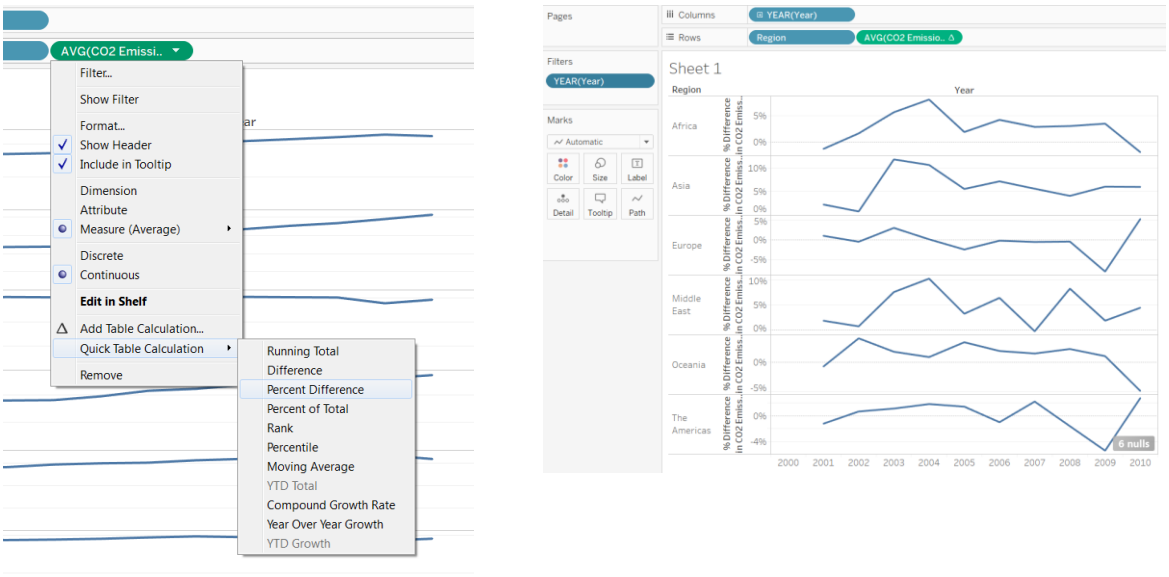
Figure 17: A filled map showing the number of patients in North, South and West Yorkshire.

4.4 Calculations

So far you have only visualized aggregations of values that are stored in a dataset (e.g., average Internet Usage for a given Region), but often you want to visualize data that needs to be calculated from the content of your dataset. Tableau provides several ways of doing that, and mastering calculations is essential if you are to unlock the full power of Tableau. However, be warned – it is often non-trivial to determine how a calculation should be formulated to produce the data and visualizations that you want. There's no substitute for systematic thinking, plenty of practice and experience!

4.4.1 Quick Table Calculations (e.g., percentage change)

Go back to the line chart you created to show Year vs. Region and average CO2 Emissions (see §4.2.4). Instead of visualizing the quantity of CO2 Emissions you might want to visualize the change in those emissions. You can show the percentage change by choosing *Quick Table Calculation -> Percent Difference* from the *CO2 Emissions* pill (see Figure 18a). If you filter the years to remove the artefact that caused by null data in 2011 and 2012 you can see details of the year-on-year change (see Figure 18b). You can also edit the calculation, for example, to calculate emissions as a percentage of the first year.



(a) (b)

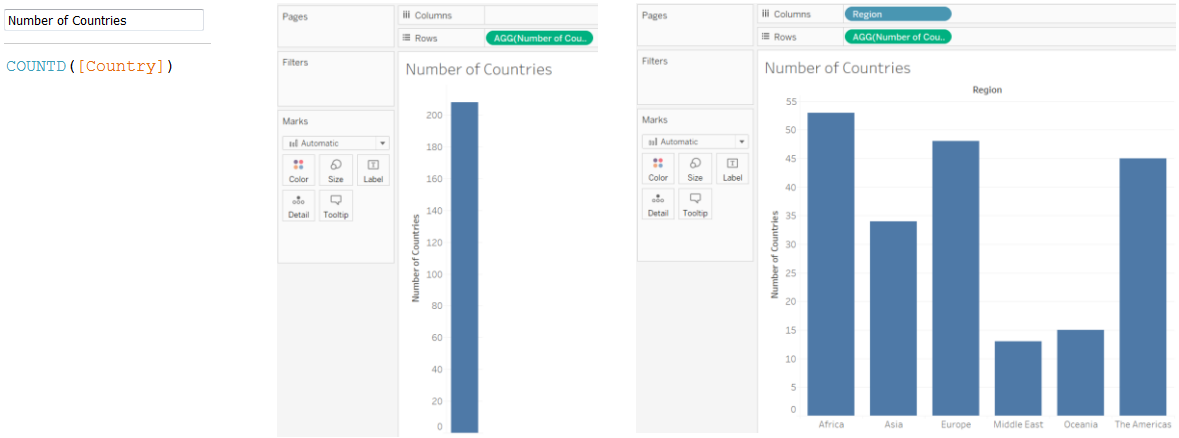
Figure 18: CO2 emissions: (a) Calculating the annual percentage difference, and (b) the result.

4.4.2 Calculating a new field

Calculated fields may contain elements for fields, functions (SUM, COUNT, COUNTD, etc.), operators (+, -, etc.), parameters and comments (any text that follows '//' characters). A simple example is this. The dataset contains Tourism Inbound and Tourism Outbound fields. To create a field for the surplus, select *Tourism Inbound -> Create -> Calculated Field* and enter the formula that is shown in Figure 19.

Figure 19: A calculated field for Tourism Surplus.

A data aggregation example is as follows. The number of countries may be calculated by counting the distinct (COUNTD) Country names. This effectively defines a database query that Tableau executes when the *Number of Countries* field is added to a worksheet, so the result depends on other fields (see Figure 20).



(a) (b) (c)

Figure 20: Calculating the number of countries: (a) the formula, (b) a bar chart showing the number for the whole world, and (c) a bar chart showing the number for each region.

4.5 Interaction

Start this section of the tutorial by recreating the bar chart described in §4.1.

4.5.1 Mark labels

To add display data values on a chart click on *Label* in the *Marks* card and select *Show mark labels* (see Figure 21).

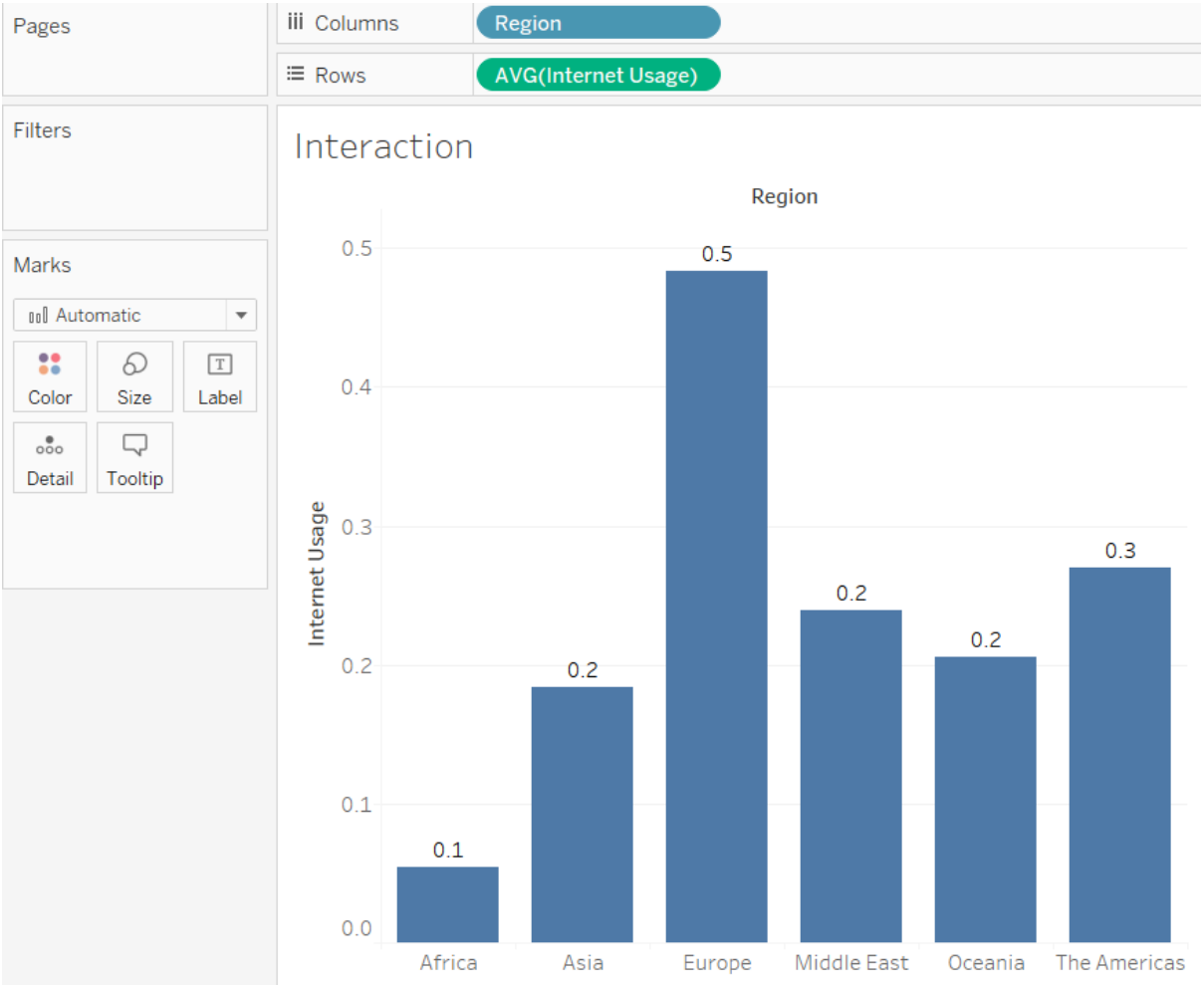


Figure 21: A bar chart with each data value labelled.

4.5.2 Sorting

To change the order of the bars, click on the Region pill and select *Sort* (see Figure 22).

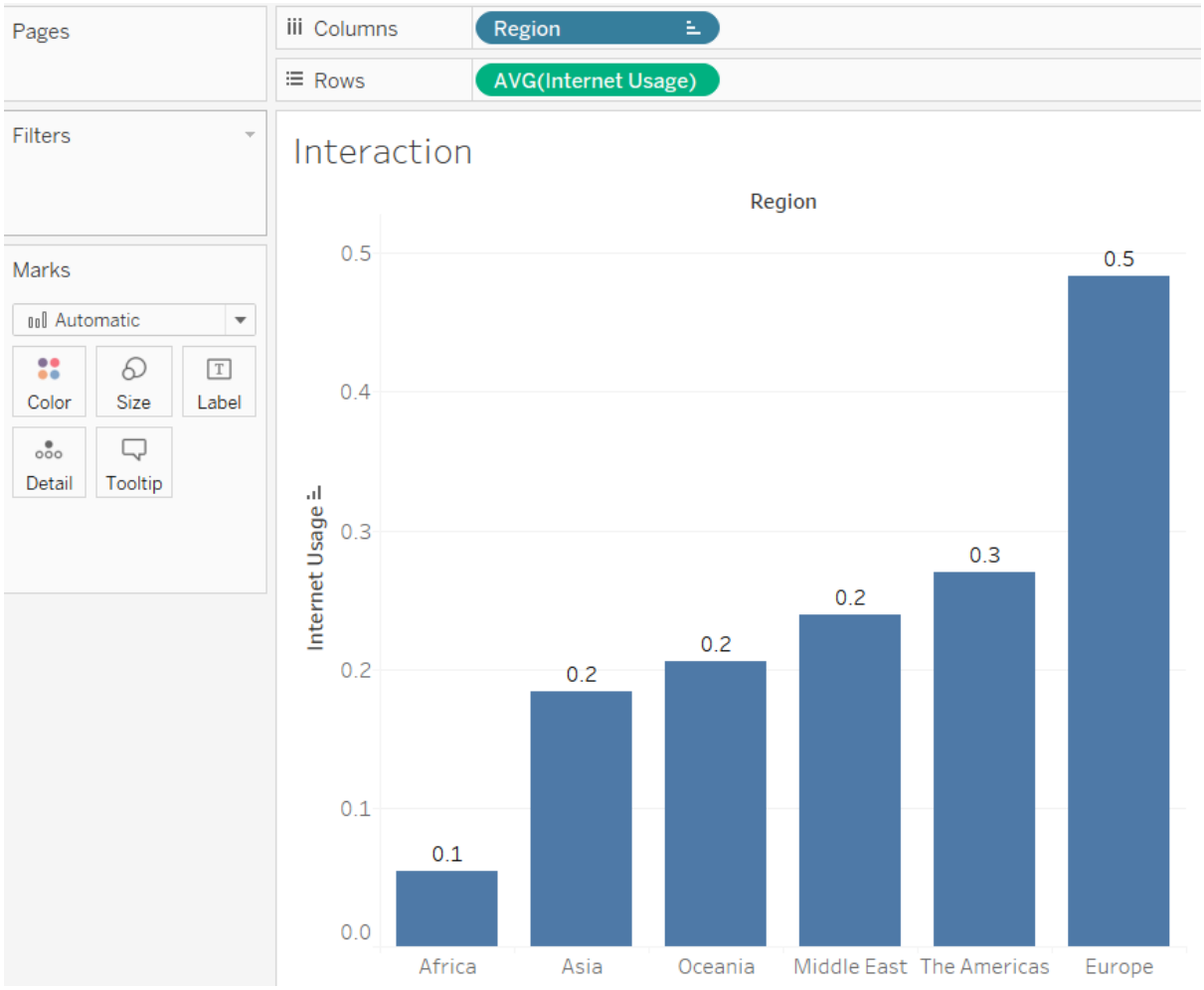


Figure 22: Reorder a dimension by sorting.

4.5.3 Grouping

Sometimes a variable contains many values, making a chart difficult to read because of the amount it needs to be scrolled. You could filter the data to (say) show the top 10 values. Alternatively, this is a situation where it is useful to group together some of the values.

The grouping process is rather convoluted, but you may perform it as follows (for other approaches, see Tableau's online help):

- Create a bar chart of Internet Usage for each country in the Middle East, and sort the chart into descending order.
- Select the countries that you want to group by clicking on their bars or highlighting them by dragging the cursor to create a box. This creates a new dimension called *Country (group)*.
- Click on that dimension in the data pane, and select *Edit Group* (see Figure 23).
    - Right click on the group you created (Iran, Islamic Rep., Iraq, Jordan and 5 more) and rename it *Other countries*.
    - Select the *Other* group and press the *Ungroup* button, to separately list each of the countries that you did not explicitly group together.
    - Remove *Country (group)* from the Marks card

- In the rows shelf, replace *Country* with *Country (group)*

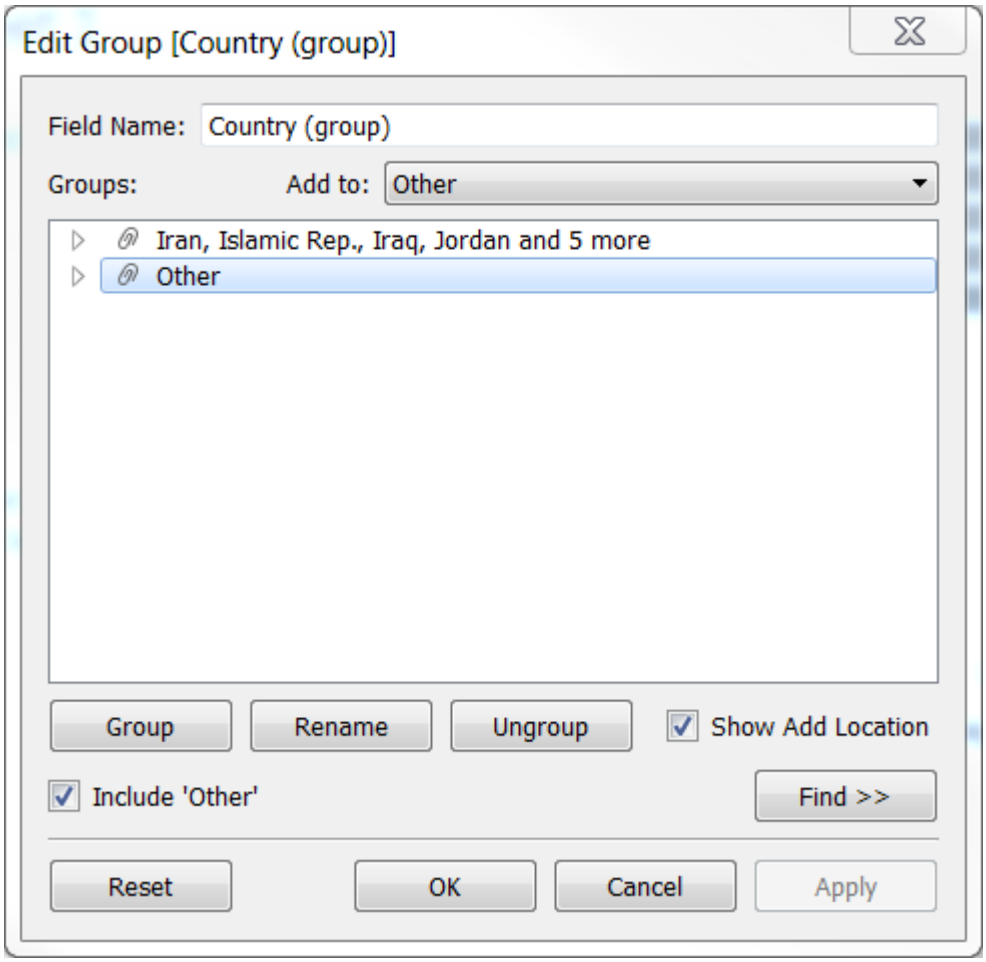


Figure 23: Editing a group of countries.

4.5.4 Parameters

You can replace constant values in calculations, etc. by interactive parameters. For example:

- Create a new worksheet and double-click on Country to create a map.
- Click on Life Expectancy Male, select *Create -> Parameter* and then click *OK*.
- Click on *Life Expectancy Male Parameter* and select *Show Parameter Control*.
- Create a dimension called *Male Life Expectancy Category* by creating a new calculated field with the formula shown in Figure 24.
- Display low and high life expectancy as coloured circles on the map and adjust the parameter to explore how the pattern varies with different parameter values (see Figure 25).

Male Life Expectancy Category

```
IF [Life Expectancy Male] < [Life Expectancy Male Parameter] THEN "Low" ELSE "High" END
```

Figure 24: A calculated field for Male Life Expectancy Category.

Figure 25: Using a parameter to set a life expectancy threshold.

4.6 Saving workbooks and exporting data

4.6.1 How to save your workbook

You can save a workbook in two formats:

- Tableau Workbook (*.twb): This is an XML file that defines how your visualizations are created. If you want to see what the XML looks like then open it in a text editor.
- Tableau Packaged Workbook (*.twbx): This contains your data source(s) and the *.twb file. **WARNING** if you send someone your *.twbx file then you have also sent them your data.

Workbooks may also be saved in a format that is compatible with an older version of Tableau (File -> Export As Version).

4.6.2 How to export the data you used

See https://help.tableau.com/current/pro/desktop/en-gb/save_export_data.htm.

4.7 Integration with R, Python and Matlab

See https://www.tableau.com/developer/data-science-integration.