Data Visualization Course Handout

August 2024 Version

Contents

Shortcuts for managing workbooks, sheets, and files

https://onlinehelp.tableau.com/current/pro/desktop/en-us/shortcut.htm

Description	Windows shortcut	Mac shortcut
New workbook	Ctrl+N	Command+N
New worksheet	Ctrl+M	Command+T
Describe sheet	Ctrl+E	Command+E
Cycle forward through open worksheets	Ctrl+Tab, Ctrl+F6	Shift+Command+Right Bracket
Cycle backward through open worksheets	Ctrl+Shift+Tab, Ctrl+Shift+F6	Shift+Command+Left Bracket
Switch in and out of Presentation Mode	F7 , Ctrl+H	Option+Return
Switch in and out of Full Screen mode		Control+Command+F
Open file	Ctrl+O	Command+O
Save file	Ctrl+S	Command+S
Save file as	Ctrl+Shift+S	
Revert workbook to last saved state	F12	Option+Command+E
Close the current workbook	Alt+F4	Command+W
Print	Ctrl+P	Command+P
Open Help	F1	Control+Command+Question Mark

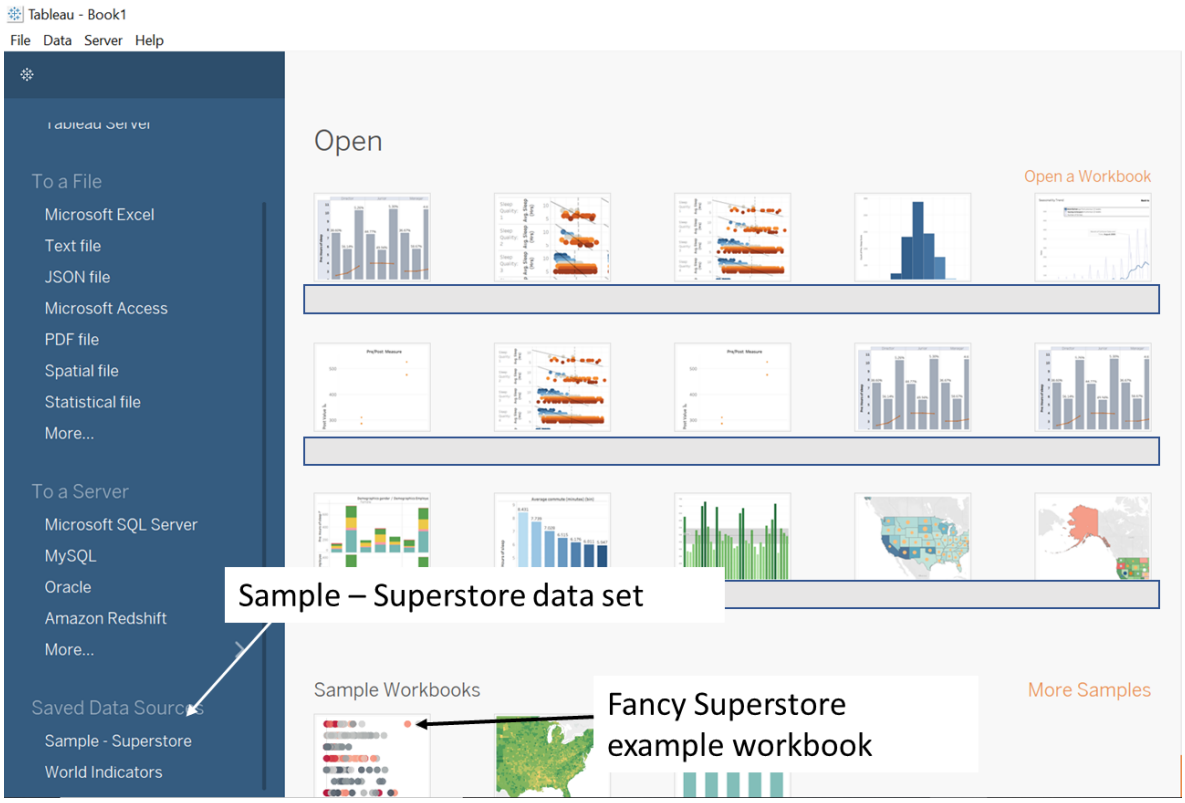
Topic 1: The Tableau User Interface

Downloading Tableau

Download a free version of Tableau for students at https://www.tableau.com/academic/students.

- Use your Georgetown email address when applying for a free 1-year license.
- You may be prompted to upload a picture of your Georgetown Student ID with a valid expiration date.
- Expect the turnaround time for the application to be less than 3 business days. Once approved, you will receive an email with your license key and further instructions.
- If you have not downloaded Tableau prior to the start of class, you can download the free 14-day trial and update the software with the free 1-year license key once you receive it in an email.

When Tableau is opened, the screen looks like this:



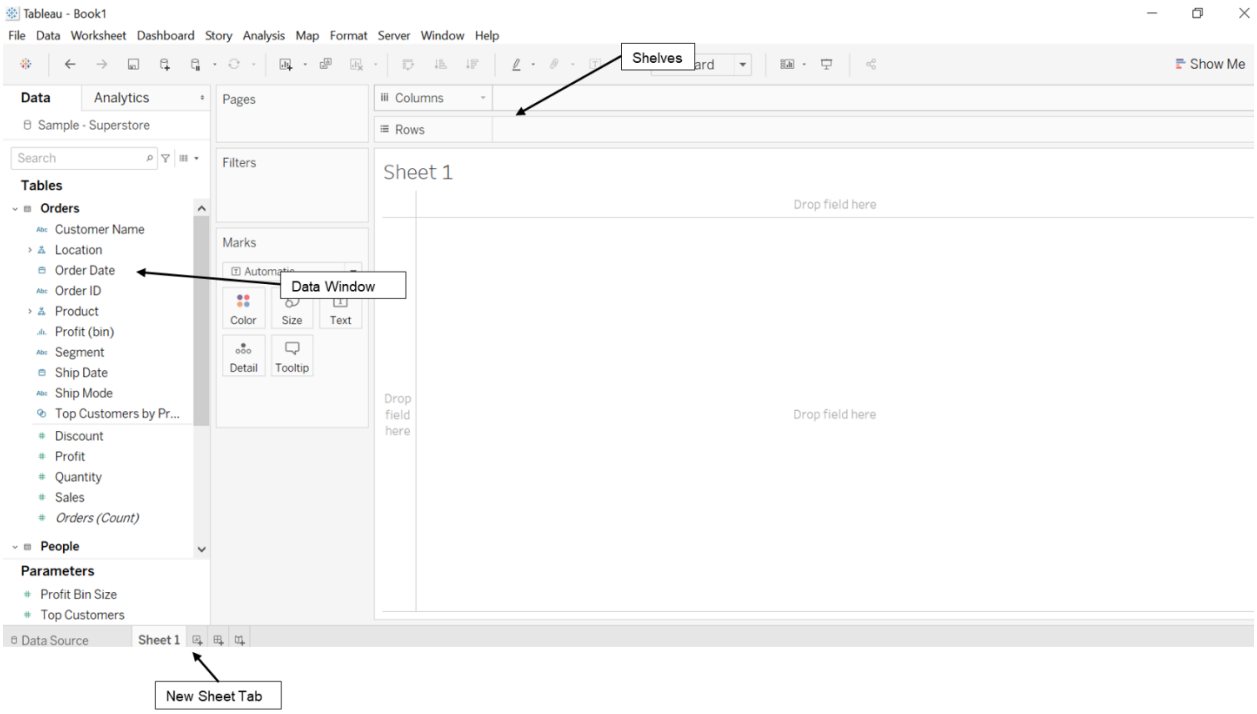
This screen allows you to navigate to where your data are. For many of the examples, you will use the Superstore Sample data provided by Tableau. You can access this data by clicking on it in the lower left of the opening screen under Saved Data Sources. There is also an Excel version of the file available for download from Canvas. NOTE: Tableau makes small changes to the Superstore dataset annually (usually just changing the dates). Depending on the example,

your visualization may not match all the examples in the handout exactly because they were created with an older version of Superstore data.

In the built-in saved superstore data set (rather than the Excel file), some hierarchies have been made (Location is a Hierarchy of Country/Region, Region, State/Province, City, Postal Code; Product is a Hierarchy of Category, Subcategory, Manufacturer, Product Name), and some fields have been hidden (Customer ID, Product ID). You can unhide hidden fields by clicking the small drop-down arrow located at the top-right corner of the Data Pane, choosing "Show Hidden Fields." You still need to right-click on the Customer ID and unhide it before you can use it in a visualization. There are a handful of products which do not have unique product names (i.e., more than one Product ID associated with the same product name) and there is one Customer Name (Harry Olson) who corresponds to multiple Customer ID's so Customer Name is nearly a unique identifier. You can assume it is for the purposes of this class.

Menus & Toolbar

Use the Superstore data to explore various areas of the screen.



New sheet tabs are found at the bottom. You can create sheets, dashboards, and stories with these tabs. You can also do things like rename the sheets, drag to rearrange them, duplicate sheets, copy formatting, and many other things.

Creating a new sheet provides a new blank sheet to build visualizations.

Menus are across the top (The layout may look slightly different on a Mac than on a PC.)

Below is the toolbar, with buttons like undo (there is no limit to how much you can undo).

In the toolbar you also have save. There is no automatic save in Tableau, so make sure to save your work periodically.

The logo button to the left of save brings you back to the start experience, where you can access saved data sources or recently opened workbooks.

Data Window

On the left of the screen is the Data window. When active, the Data tab lists all open data connections, and depending on which one is selected, the fields from that data source are listed, broken out into dimensions and measures with a light gray line dividing the two field types (dimensions are above the line and measures are below the line). The data window will also show any sets or parameters you may have created.

- If the Map options or the Format pane are opened, these temporarily cover the data window. To get back to the data window, simply close the other pane by clicking the X in the upper right corner.

- At the bottom of the screen is the status bar. This can provide helpful information like the number of marks in the view as well as other summary information.

If you switch from the Data tab to the Analytics tab, you are brought to an entirely new pane. The Analytics tab will be discussed later, but this is where you can use additional analysis elements (things like trend lines).

Shelves and Cards

Tableau uses its own unique terminology and if you are communicating with other Tableau users it is important that you know the terms. Views are built by dragging and dropping fields from the data window into the canvas directly, or onto the shelves.

- Columns Shelf: Fields placed here will create columns on the view

- Rows Shelf: Fields placed here will create rows on the view

- Pages Shelf: The Pages Shelf allows you to flip through a page for each dimension member and/or add an animation to a view.

- Filters Shelf: Any dimension or measure that you filter a view by will be displayed here.

- Marks Card (Marks Shelf): Each square in this area is called a Marks Card, which are called that because they influence the marks on the view. Each Marks Card resides on the Marks Shelf. Note that depending on the chart type you are creating, additional cards will show up such as for Shape or Path. For example, select the **Shape card** to change

the shape of icons from the default shapes to people, bugs, weather, etc. The **Path card** will be used in some examples when it is important to establish a desired sequence.

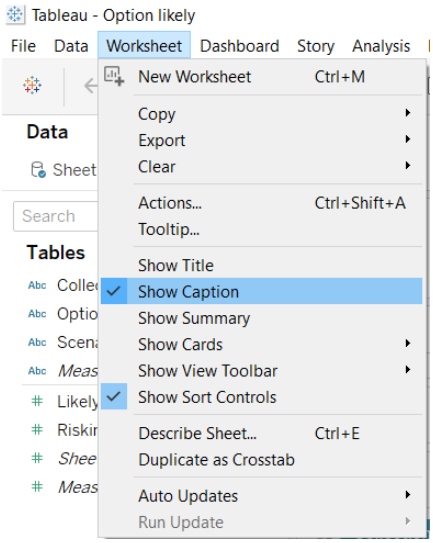
- “Pill”: the slang term for fields being used on a worksheet. This term is used due to the oblong shape dimensions and measures inherit once they are placed on a shelf

Legends

Legends, such as for color, size, and shape, will automatically be created when a field is placed on the color, size or shape card. You can also change the order of fields in a visualization by dragging them around in the legend. Legends can be removed by clicking on the menu and then selecting hide. To bring a legend back, use Legends under the Analysis menu.

Captions

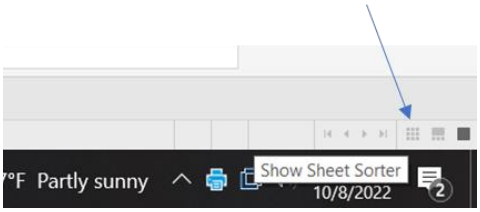
On each worksheet, you can add a Caption below the visualization. You do this by selecting Show Caption from the Worksheet menu.

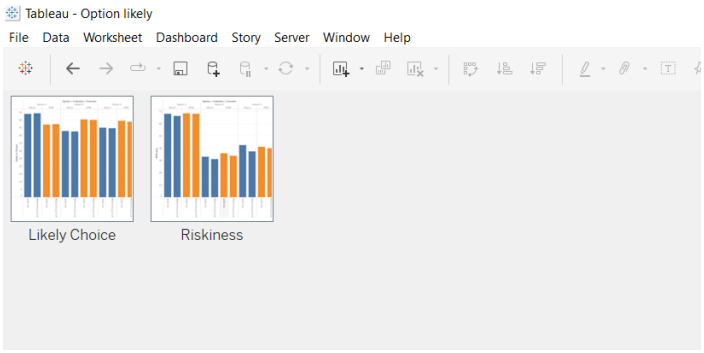


This could be a useful place for you to take notes during the class examples if there are things you want to remember about particular visualizations. This is also helpful for providing additional text to explain your visualizations.

Navigating your Tableau workbook

If you only have a small number of sheets in your workbook, you can easily navigate through them with the tabs at the bottom of the workbook. However, when you have a large number of sheets, it may be easier to use the sheet sorter view. You can also reorder sheets, create new sheets, and duplicate or delete existing sheets in the sheet sorter view. To return to the normal view, click the last icon “Show Tabs”.





Exporting Images

You can quickly copy an individual view as an image and paste it into another application, such as Microsoft Word or Excel.

Select Worksheet > Copy > Image.

To create an image file that you can reuse, export the view rather than copy it.

Select Worksheet > Export> Image.

In the Export Image dialog box, select the elements you want to include in the image. If the view contains a legend, under Image Options, select the legend layout.

Click Save.

In the Save Image dialog box, specify a file location, name, and format. Then click Save.

Exporting Data

You might want to share or reuse the data in its new form.

Select Worksheet > Export > Data.

Select a location and type a name for your file.

Click Save.

Because the .csv format is one of the simplest structured formats for data, it's supported by a wide range of tools, databases, and programming languages. Exporting your data in the Tableau data source using this format creates an independent data set and can be a convenient and flexible way to share your data with others.

You can export directly to Excel the data used to generate the view formatted as a crosstab. When you export your view as a crosstab, Tableau automatically opens the Excel application and pastes a crosstab version of the current view into a new Excel workbook.

Although this option provides a direct method for exporting your data to another application, performance of the export can be affected because it is simultaneously copying and formatting the data. If the view you are exporting contains a lot of data, a dialog box opens asking whether you want to export the formatting. In this case, if you choose to exclude the formatting from the export, performance of the export might improve. In Tableau Desktop:

Select Worksheet > Export > Crosstab to Excel.

If you're using a Mac, this option opens a dialog box where you can save the file. You must then manually open the file in Excel.

Combining/Merging Multiple Workbooks

With the existing workbook open, select File > Import Workbook.

Select the workbook that contains the sheets you want to import from another workbook, and click Open. When the imported workbook contains a duplicate sheet name, Tableau adds a number after the name of the imported sheet.

Alternatively, you can copy sheets in tableau from one instance to another. First, open both workbooks. if you have a lot to copy it's often a good idea to do this from the sheet sorter. Multi-select sheets using shift+click, or drag over the sheets you want to copy, right click and click copy sheet and then move to the other workbook and right click on the sheet area and click paste sheet.

When you copy, save, or export selected sheets, the data source or sources that are used on that sheet are also copied. This includes any calculations, parameters, sets, etc. Custom shapes and colors are also included. If the sheet being copied is a dashboard or story, all of the sheets that are used on that dashboard are also copied, whether they are hidden or not.

When you paste or import sheets from a different workbook, some items might already exist in the destination workbook, or some items might have the same name in both places. If Tableau encounters an exact duplicate item in the Data pane, such as a calculation, it does not paste or import that item into the destination workbook. However, if an item in the Data pane has the same name but is defined differently, Tableau imports and renames it.

Tableau also pastes or imports sheets and data sources with duplicate names—whether only the name is identical, or their names and contents are identical—and it renames the newer copy.

If you want to extract a subset of information from a larger workbook to maintain as a standalone file, you can export or save selected sheets to a new workbook. You can then import that workbook into an existing one to incorporate its sheets and other objects into the existing workbook.

If in combining workbooks, you also get multiple copies of the data source, you may need to Replace the Data Source generated by the merge with the original/primary one. Any fields that the user creates in the original data source (for example, calculated fields, groups, sets, etc.), that are included in the view are successfully copied over to the new data source. Fields that are not included in the view that don't exist in the new data source are not automatically copied and will need to be manually copied and pasted to the new data source before replacing the old one.

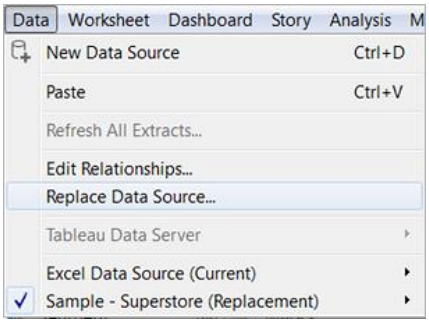
The two data sources do not have to be identical, however, any differences between the data sources will affect the sheets in the workbook and the fields in the view. More specifically, any fields, groups, sets, and calculated fields that do not exist in the new data source (or have a different name) are removed from the Data pane.

1. Open a workbook that connects to the *original* data source.

2. Select Data > New Data Source and then connect to the *new* data source.

3. On the Data Source page, drag a table to the canvas to set up the data source (if this is not automatically done for you).

4. Go to the sheet tab and select Data > Replace Data Source.



Submitting Files

Whenever you save your work in Tableau, the default file extension is .twb, or Tableau Workbook. However, you have the option to save your work under a different extension, .twbx.

A .twbx file is a Tableau Packaged Workbook, meaning it is the original .twb file grouped together with the datasource(s) in one package. .twbx files can be considered analogous to specialized zip files, in which these “zip” files contain all the information necessary to work in Tableau. The primary advantage to using .twbx files is that analysis can be performed without network/internet connections to your data because your data is already present in this packaged file.

For your assignments to work properly to be graded, you will need to submit them as a Packaged Workbook, i.e., *.twbx file. This will work for all projects in this class. If later in life, you are connecting to data in a database or server, you need to create a tableau data extract (.tde) before you can take advantage of the packaged workbooks.

Benefits of Tableau vs. Excel

Tableau Benefits:
- Doesn’t change or touch the original data so no one will mess up the data or accidently delete it.
- Makes creating maps easy.
- Great for pulling together diverse data stored in disparate worksheet files.
- Easy to create interactive and visually appealing dashboards that allow for dynamic data exploration.
- Simplifies the process of creating visualizations and dashboards with an intuitive drag-and-drop interface.

Excel Benefits:
- Great for storing data.
- Generally more affordable than Tableau.

Topic 2: Connecting and Organizing Data

The first step to access your data occurs on the opening Tableau screen.

Connect to Data Screen

In the Connect pane there are all sorts of data sources. This is a real strength of Tableau, but here, we will focus on a very common data format – an Excel file.

Rather than selecting the Superstore data from the Saved Data Sources, download the latest Superstore Excel file provided on the course website. Select Microsoft Excel (from the To a File option), navigate to where you saved the file, and click open.

Data Connections

Tableau supports two different data connections Live and Extract.

Live connections connect Tableau directly to the data source. Any changes in the data source will be directly reflected in the Tableau view. Your connection relies on the speed of your database.

Live connections are great when you need just-in-time (near real-time) data and/or you want to leverage your database performance/investments.

Extracts create saved subsets of data that can be used to improve performance and allow you to take advantage of specific Tableau functionality.

Advantages of extracts include:

- Supporting large data sets
- Improved performance
- Additional functionality
- Offline access to data
- Can be refreshed on a scheduled/periodic basis
- Incremental Refresh to limit the amount of data refreshed

Connecting to Tables

Now Tableau brings you to the data connection window.

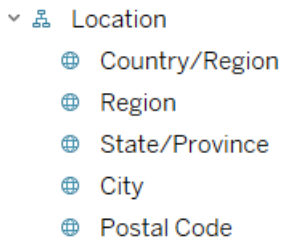
- At the top you can see the name of the file – and you can click to rename the connection if desired.
- In the Sheets area, you can see all the sheets in the Excel file.
    - Sheets in Excel are treated the same as tables in databases, and you can choose to connect to a single table or multiple tables.
- Drag a sheet that you want to include into the data connection canvas.
- You can preview the data in sheets by clicking the “View Data” icon to the right of the sheet name.

Using the Superstore Excel file, only drag the “Orders” sheet in the file to the data connection canvas (you will add the other sheets in section: Connecting to Multiple Tables later). Then Select the Sheet 1 to further organize the data.

Organizing Data via Hierarchies

When you first opened up the Superstore data, there was a dimension called Products that included the Category, Sub-Category, Manufacturer, and Product Name for each of our products. However, this isn’t something that Tableau does automatically for us when importing new data. To create this Hierarchy, drag Sub-Category on top of Category until you see a small plus sign. A “Create Hierarchy” pop-up box will appear asking you to name the Hierarchy – name it Products. Now drag Product Name and Manufacturer into the hierarchy. Put them in order from largest to smallest (Category, Sub-Category, Manufacturer, Product Name). Drag the Category Dimension to the Rows Shelf. You see the plus next to the Category pill and can drill into these other dimensions. (If you are using the Excel file, you will not have the Manufacturer).

Repeat the steps to organize the Location fields.



Poorly Formatted Data

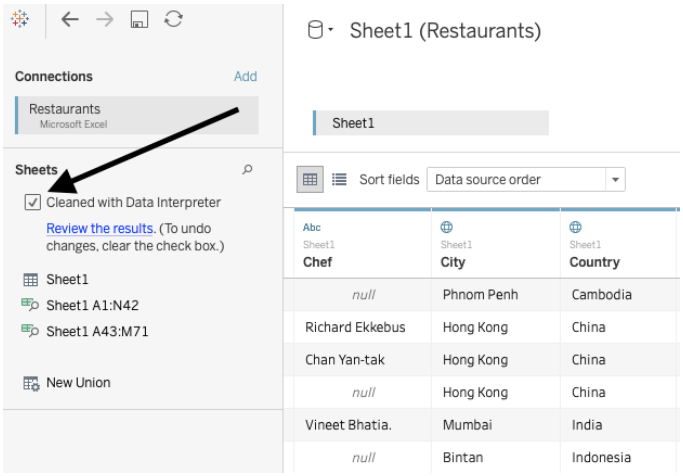
The Superstore data in Excel was already formatted nicely – you just brought it straight into Tableau. In reality, data files are not always so well-formatted. The next section explains how to prepare poorly formatted data when you run into it.

Review the Restaurant data to get a sense of why it might need some additional preparation.

Data Interpreter

In Tableau's Data Source tab, you can see a preview of the data which will give you an idea of how clean it is. For example, the data may be missing column names or headers. Tableau should recognize this and suggest the Data Interpreter (Tableau's built-in tool for preparing your data for analysis). Note that if the data source is already in a format that Tableau can interpret and Tableau Desktop doesn't need extra help from the Data Interpreter to handle unique formatting or extraneous information, the Data Interpreter will not be available.

- Turn on the Data Interpreter by checking the box shown below.



- Data Interpreter is useful for striping out nulls/headers and identifying columns properly.
- If you want more specifics on what the Data Interpreter did, you can click "Review Results" on the right. This will open an Excel file describing the changes.
- If you click to the tab you used, you see which fields are being used as headers, in orange, and which are considered data, in green.

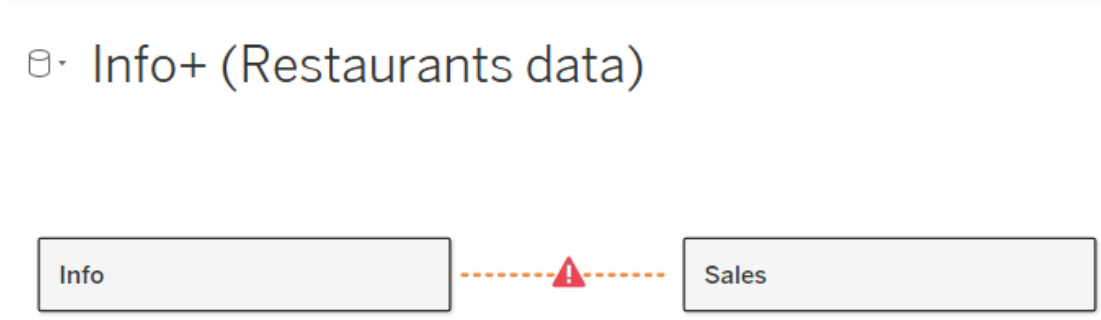
Connecting to Multiple Tables

To connect data from multiple tables in the *same data source* (i.e. other worksheets in an Excel file), you need to connect the tables.

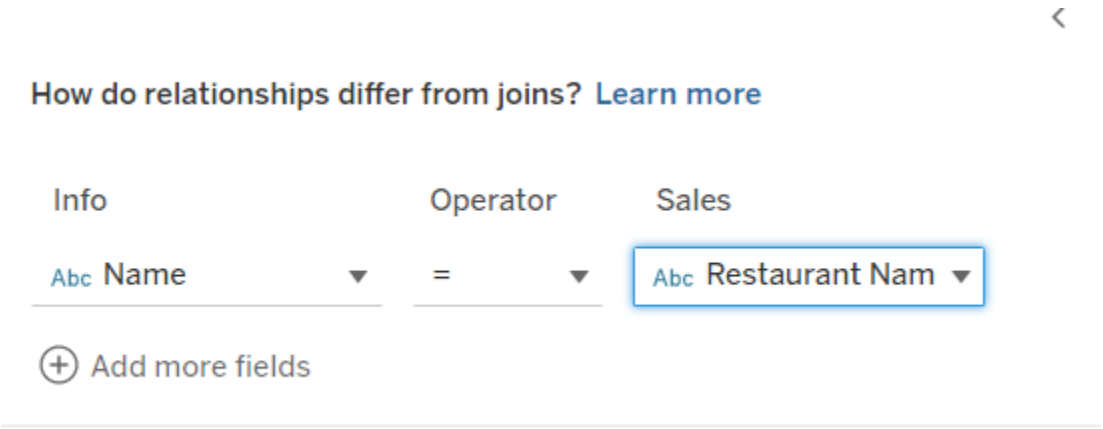
You should still be on the Data Source tab (lower left of screen), and we want to link the fields in information table with those in the sales table.

Double click or drag out the Info table and the Sales table.

- The icon here indicates Tableau has automatically linked these tables (established a ‘relationship’.



But it doesn’t know how these are related. Because of the way we ‘cleaned’ our data the Restaurant name in the Info table is “Name” and in the Sales table is “Restaurant Name”. We need to tell Tableau these are the same.



Notice how you can click back and forth between the two tables to view the data.

NOTE: Prior to Tableau 2020.2, you actually joined data between tables. When users did this, it was easy to miss relevant data. Now Tableau only establishes relationships between tables. During analysis, Tableau adjusts join types intelligently and preserves the native level of detail in your data. You can see aggregations at the level of detail of the fields in your visualizations rather than having to think about the underlying joins. Joins would have filtered data, but relationships always preserve all measures. And unlike joins, relationships won’t double your trouble by duplicating data stored at different levels of detail. Tableau is claiming that relationships are better because unlike joins or unions, relationships form a data source without flattening multiple tables into a single table. That means each field keeps its context or level of detail. But if you need to join data for some reason see this website for more information:

https://help.tableau.com/current/pro/desktop/en-us/datasource_relationships_learnmorepage.htm

Pivot

Tableau has a preferred data arrangement: taller (often called long data), with more rows, rather than wider (called wide data), with more columns. How do we make this adjustment in Tableau? For example, it is preferred to have a row for each date rather than multiple dates in a row that show a common field.

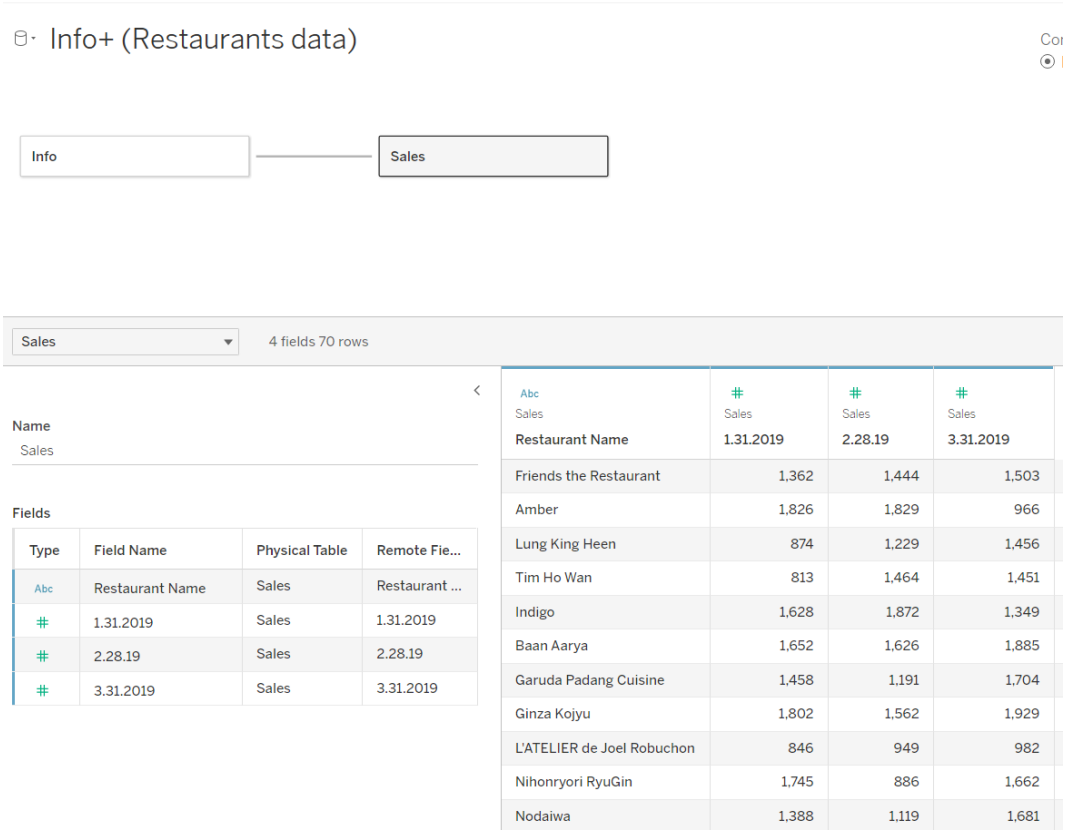
Preferred:

Sales	Date
11,203	11/30/19
11,104	12/31/19
12,346	1/31/20

Not Preferred:

Date	11/30/19	12/31/19	1/31/20
Sales	11,203	11,104	12,346

In Tableau, you may want to change the format from that column-per-month layout into a single date column and a single column. Looking at the Sales sheet into the data connection canvas as shown below, notice how each line has multiple monthly sales figures. This is not a good format for Tableau so we need to use the pivot tool.



To do this easily, you'll simply select all the date columns at once. Then right click on one of the highlighted columns and select "Pivot".

Abc Sales Restaurant Name	Abc Pivot Pivot Field Names	# Pivot Pivot Field Values
Friends the Restaurant	1.31.2019	1,362
Friends the Restaurant	2.28.19	1,444
Friends the Restaurant	3.31.2019	1,503
Amber	1.31.2019	1,826
Amber	2.28.19	1,829
Amber	3.31.2019	966
Lung King Heen	1.31.2019	874
Lung King Heen	2.28.19	1,229
Lung King Heen	3.31.2019	1,456
Tim Ho Wan	1.31.2019	813
Tim Ho Wan	2.28.19	1,464

- This pivot feature merges the information from the original columns and rows into two new columns – Pivot field names, and Pivot field values. See how there are now three rows for the Amber restaurant, one for each month of sales.
- You can see that “Pivot field names” is actually the Date, so you can click to open the menu and select rename.
- Similarly, “Pivot field values” can be renamed Restaurant Sales, so rename that field also.

Abc Pivot Date	# Pivot Restaurant Sales	Abc Sales Restaurant ...
1.31.2019	1,362	Friends the R...
2.28.19	1,444	Friends the R...
3.31.2019	1,503	Friends the R...
1.31.2019	1,826	Amber
2.28.19	1,829	Amber
3.31.2019	966	Amber
1.31.2019	874	Lung King Heen
2.28.19	1,229	Lung King Heen

Correct Date Format

Now the date and restaurant sales are oriented correctly, the formatting of the date is still problematic if we want it recognized as a date. This is just the first of many examples where we will use a calculated field to solve a problem that we encounter with our data.

- Right-click on the field of concern (in this case Date) and select create calculated field.
- You will need to give this new field a name (consider Date Corrected).
- Type DATE(DATEPARSE(“M.d.yy”,[Date]))
- The DATEPARSE function lets you clearly define which parts of your field are which parts of a date. In essence, you are translating the string into a date field.

Note: once corrected, you could choose to only retain the corrected Date field and rename it Date, but we will keep both since this is an example. You could also choose to hide the original field. If you hide fields and need to unhide them later, select the small arrow by view data to find hidden fields, i.e., “Show Hidden Fields”.

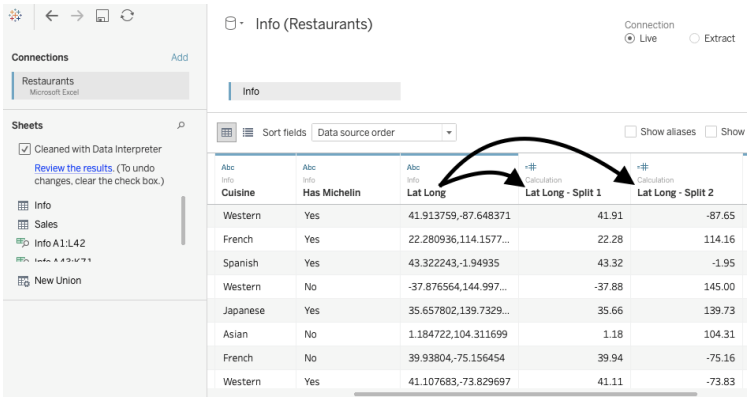
Also, don’t think you can use any combination of M/m, D/d, Y/y, because each has a specific meaning. For example, capital “M” means months, and small “m” means minutes. See this

website for more details on DATEPARSE notation:
https://help.tableau.com/current/pro/desktop/en-us/data_dateparse.htm

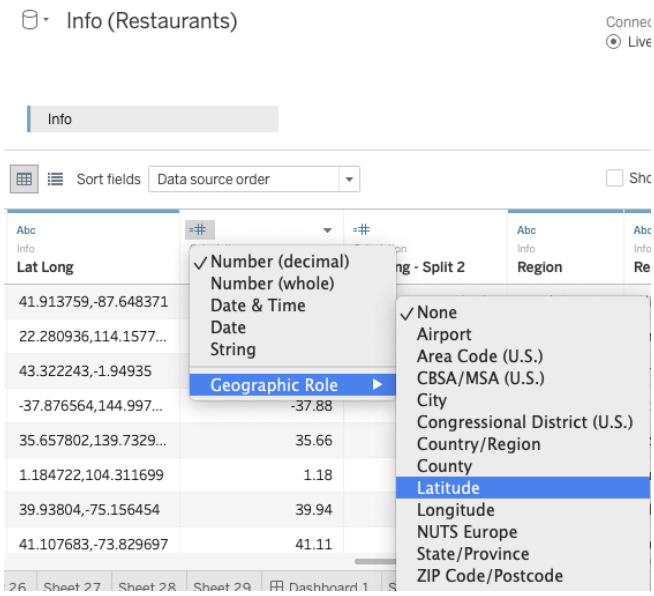
Split

There's another common data preparation tool known as a split. Switch the data source to the Info tab instead of the Sales tab. In the Info tab of Restaurant.xlsx, see how the longitude and latitude values are combined in one field called 'Lat Long'. To separate these two pieces of info, we will use the split tool.

- Right click on the Lat Long column and select Split.



- There are now new fields – Split 1 and Split 2.
- Right click on the split column headers to update the names to latitude for Split 1 and longitude for Split 2.
- Select the # icon above the new Latitude column and update the field type to a latitude so Tableau knows how to treat it.



Do the same for the new Longitude column, but this time mark it as a longitude.

Data Validation - Nulls to Zeros

A common issue with real-world datasets is the presence of null values, or blanks, in the data. This can make data analysis challenging and frustrating (as you may or may not know, some statistical techniques will not work at all if there are nulls in the data). Here is how to remove Nulls from a data file:

- Right-click on the field of concern (in this case Michelin Stars) and select create calculated field.
- You will need to give this new field a name (consider Michelin Stars Revised).
- Type ZN(in the box and you will start to see the function auto-fill. Formula should be ZN([Michelin Stars])
- The ZN function returns a zero wherever there is a null. You can think of this function as asking “Is there a value for this record? If not, give it a zero.” You can then use this field instead of the one that has null values in it.

Pro Tip: in this case the Null values had a logical interpretation as zeros. But if it is truly missing data, and you want to exclude the rows with Nulls from our analysis, you can use a Data Source Filter:

1. Right click on the data source and select “Edit Data Source Filters”
2. Click Add and select the field that contains the Nulls

3. Click on Special and select Non-Null Values

Make sure you save the Restaurant dataset with the pivot, split, corrected date field, and zeros for null values for future practice problems.

Topic 3: Working with Data

Dimensions vs. Measures

Tableau classifies each field as a dimension or a measure. Tableau will group the fields by their dimension or measure classification on the left-hand side of the workspace in the Data Window. Prior to the 2020.2 release, this distinction was more prominent. Now dimensions are above a light gray line and measures are before the line. While less prominently featured, the distinction is still important.

Measures are numeric values that can be used in math operations such as average or sum. Tableau will assume that any field containing numeric information is a measure. For example, Sales is quantitative and is treated as a measure.

In Tableau, measures come into the canvas as aggregates. The level of granularity is set by the dimensions in the view. Think of them as the data elements on which you want to perform calculations. The value of a measure therefore depends on the context of the dimensions. For example, the result for the sum of **Sales** is different if there are no dimensions in the view (just a single overall sum) versus when **State** is set as a dimension. The later computes a sum for each state.

In contrast, Tableau treats any field containing qualitative, categorical information as a dimension. For example, sales may be broken down by Region. You can think of a dimension as a field that can group a measure. For example, you could use region to group sum of sales.

Generally, the measure is the number; the dimension is what you 'slice' the number by.

There are exceptions to this rule. Order numbers can be numeric so Tableau would try to classify this as a measure. But we would want to 'slice' by order numbers? A good rule of thumb is if it does not make sense to sum the number, it is probably a dimension. (What value is the sum of order numbers?) If a field is misclassified, reclassify it by right-clicking the field name in the Data pane and choosing "convert to dimension" or "convert to measure" as appropriate. Or simply drag and drop the field into the correct area of the Data pane.

Remember:

- Dimensions show in the view as themselves
- Measures show in the view aggregated

The second way Tableau classifies a field is as discrete or continuous. This classification has an impact on what types of visualizations you can create, so it is important. It is easy to know if a field is discrete or continuous based on its color.

When you bring a field into the view from the Data Window, Tableau creates a pill. The pill color indicates whether the pill is continuous or discrete. Discrete pills are BLUE, Continuous pills are

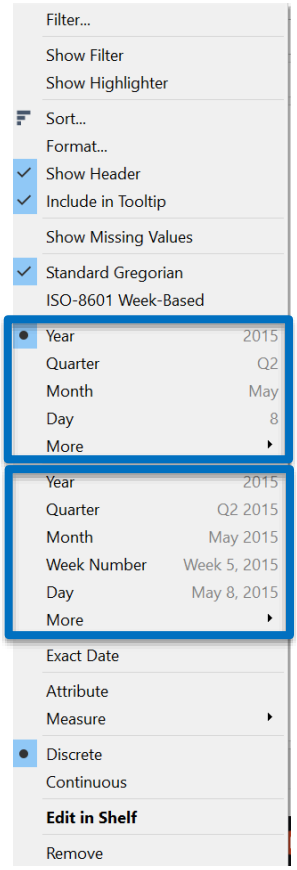
GREEN. Text and categories are inherently discrete. Numbers can also be discrete if they can only take one of a limited set of distinct, separate values. On the other hand, numbers are continuous if they can take on any value in a range.

Most of the time, dimensions are discrete and measures are continuous. However, it is possible to change this. If you want to change a continuous measure to a discrete measure, you can right click and select “convert to discrete.” Remember that measures can actually be used as discrete fields or continuous fields, and the same is true for some dimensions such as dates.

Continuous versus Discrete Dates

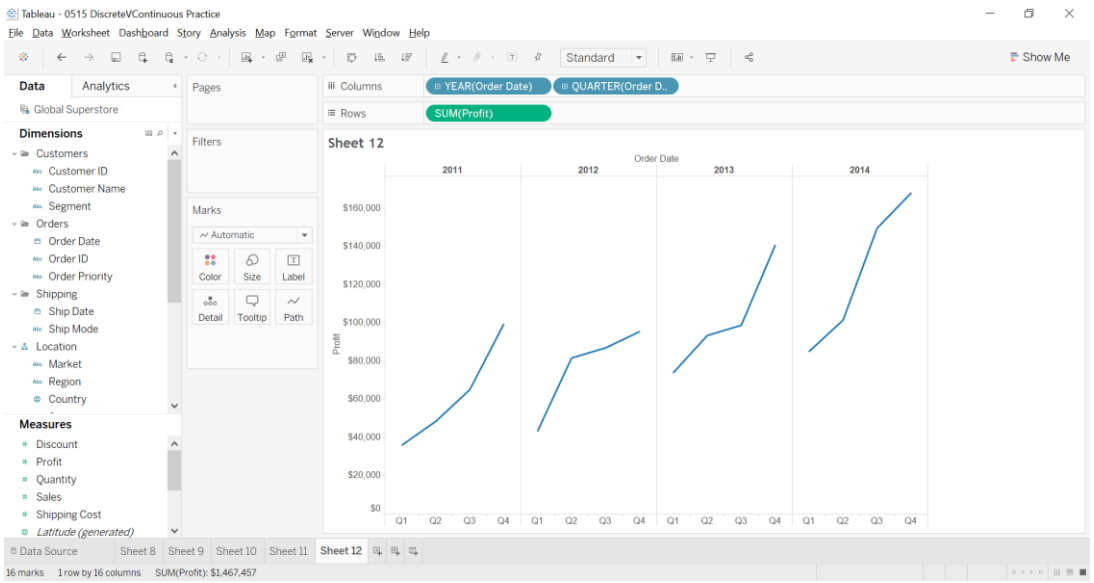
The continuous versus discrete nature of date data is very important to understand. This is because dates can be brought into the view either way. If the date icon is blue in the data window, then the default when we bring out that pill will be discrete, otherwise it will be brought out as continuous. Users can Right-Click-Drag (Option-Click-Drag on a Mac), to set the date type.

Once used in a visualization, date data can be changed from continuous to discrete or vice versa from the pill’s menu. Note that the options on top are for discrete dates, and the options in the lower section are for continuous dates.

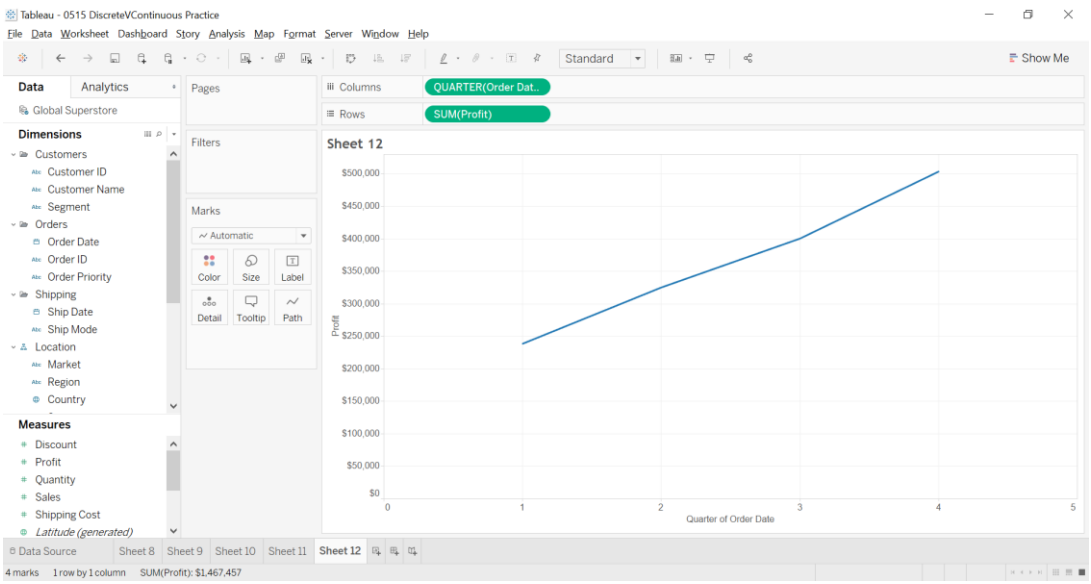


Quick Continuous versus Discrete Dates Example

Using the Superstore dataset, start by double-clicking **Profit** (or dragging the field to rows), then bring **Order Date** to columns. The blue pill indicates that this is a discrete date. This means that each part of the date is treated like a category, and clicking the **+** will allow for drilling down to from years to smaller time date increments. There will be a separate section of the visualization for each year, rather than one continuous line. Go to the **Analytics** tab and drag out a trendline. Now there is a separate line for each year. Discrete dates can be useful because they allow each date part (year, month, etc.) to be treated as a separate variable. For example, to filter sales by month (e.g. “show me our sales in July”), a discrete date works best. The visualization should look like this, with Year and Quarter as separate pills:



Next drag the **Quarte**r pill out of the view (or click the minus sign next to Year) so that only **Year** remains, and remove the trendline (drag it off). Right-click the blue **Year** pill and select the lower Year (continuous). The line should now become continuous, and when the **+** is clicked it simply adds resolution rather than adding new pills since each part of the date is no longer considered a category. Now drag a trendline out, and a single line for the entire (continuous) range should appear. The visualization should look like this, with Quarter overwriting Year as we drill down:



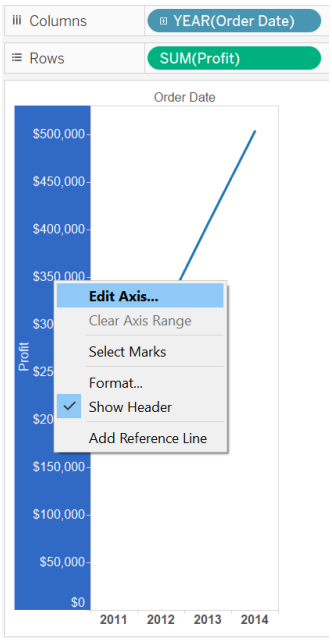
To learn more about discrete versus continuous dates check out this website:

http://sqlbelle.com/2015/01/29/do-you-know-how-dates-work-in-tableau-understanding-continuous-vs-discrete-dates/#more-2628

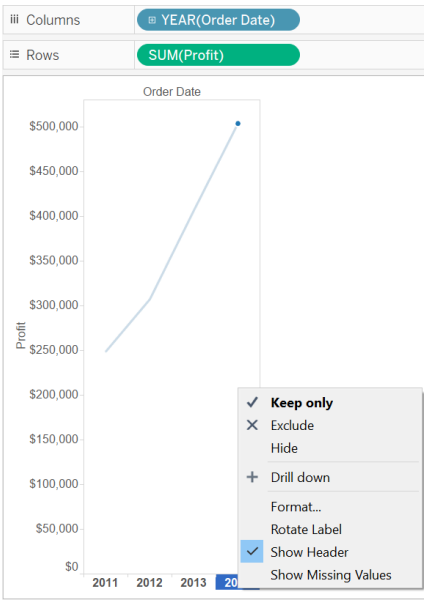
Axis versus Label

The use of a **label** versus an **axis** for a visualization is an important distinction in Tableau. On the surface they may appear very similar. However, the distinction impacts how data are visualized. Between labels, additional values cannot exist. Thus, discrete data fields (i.e., blue pills) will be presented as labels. Typically, but not always, labels correspond to dimensions. As a label example, an order was either placed in the Fall, Spring, Summer, Winter. A label would be used to describe when the order was placed. Another example of labels would be specific years 2015, 2016, and 2017.

Alternatively, continuous data fields (i.e., green pills) will be presented as data and may be used as axes in visualizations. Values can exist on a scale at any point along an axis. An example of an axis would be the time spanning from 2015 to 2017. An axis would be used to describe exactly when the order was place between 2015 and 2017.



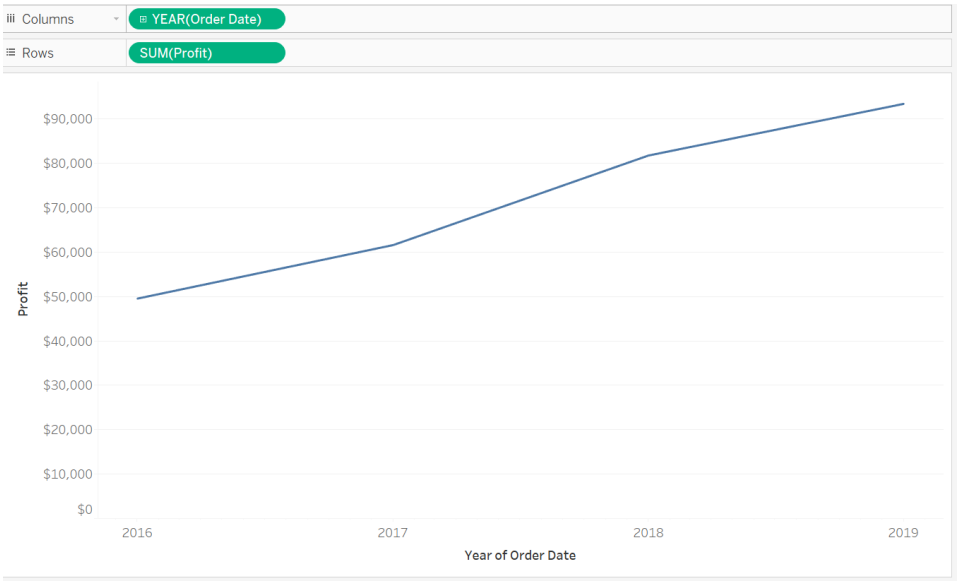
Axis (Vertical)



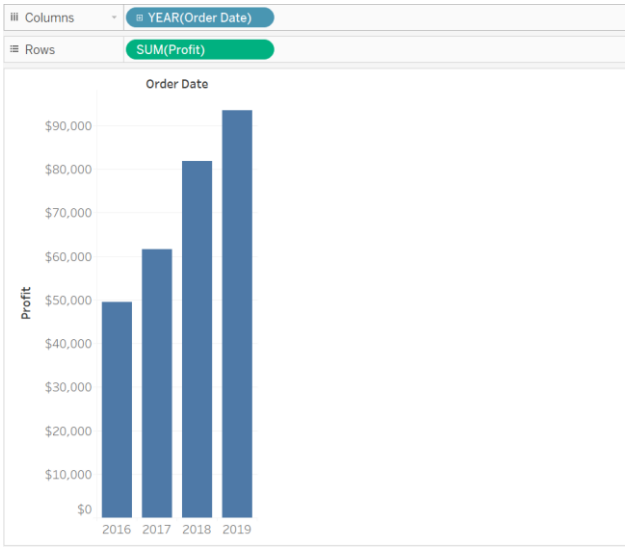
Labels (Horizontal)

Quick Axis versus Label Example

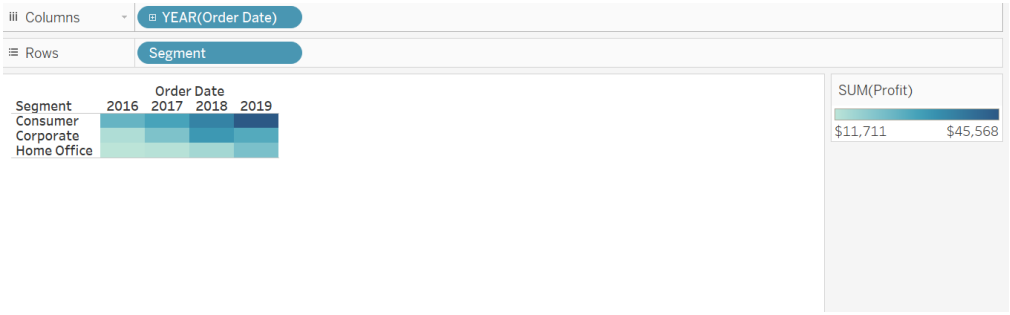
Using the Superstore dataset, create a new sheet. Drag **Order Date** to Columns. If Order Date is a discrete date, change it to a continuous date. Notice how the date labels have changed to an axis. Drag **Profit** to Rows. The chart shows profit over time.



Now change **Order Date** to a discrete date and the chart to bar chart (use Marks card to change Chart type). Now labels show the profit measure for each year.



With another new sheet, drag **Order Date** to columns again. Then drag something discrete to rows such as **Segment**. A table is created instead of an axis. Drag **Sales** to the Color mark to create a heat map (more on those later). If the visualization does not look like the heat map below, change the mark **Shape** to automatic or square.



Filtering

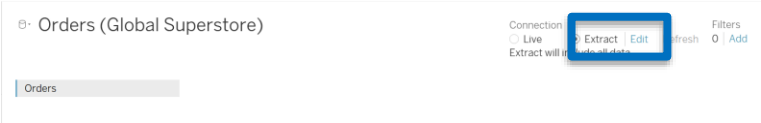
In Tableau, a filter uses a dimension or measure to limit the scope of the visualization. For example, a region filter can be used to show sales of a specified region as shown below. This allows data visualizations to be more interactive and flexible for different types of users.

In Tableau there are different types of filters, and it is important to understand the order in which filters are executed, known as the order of operations, as this will impact the filter results.

Filters are executed in the following order:

1. Extract filters

Extract filters are used to filter the extracted data from the data source. This is set in the Data Source tab.



2. Data source filters

   Data source filters are used at the source data level. They restrict the records present in the data set. This filter is similar to an extract filter except that it can operation on either live or extract connections.

3. Context filters

   Context filters create a separate dataset from the source data and compute the selections made at the worksheet level. All other filters used in the worksheet build on the Context filter. This means that after a context filter is set it is an independent filter (i.e., data are assessed without considering other filters). Any other filters will then be considered dependent filters because they will process only the data that has already passed through the context filter. If you have two dimensions (or two measure) filters in a single view, and you want them to execute in a particular order instead of simultaneously, you can convert the one to execute first to a Context filter.

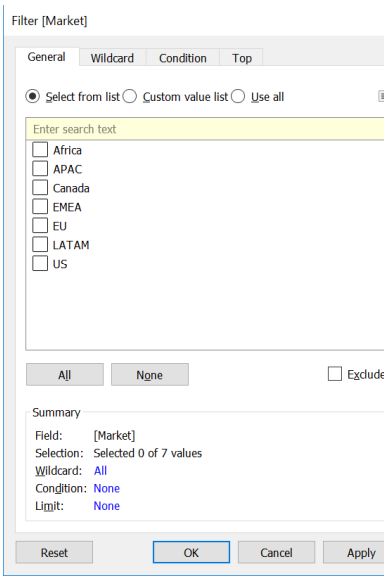
4. Filters on dimensions (whether on the Filters shelf or in filter cards in the view)

   When a dimension is used to filter the data in a worksheet, it is called as Dimension filter. The members present in a dimension can be included or excluded from the list using this filter. This type of filter can be shown on a worksheet or dashboard.

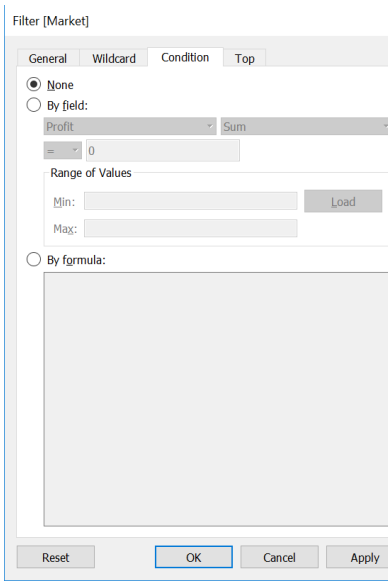
5. Filters on measures (whether on the Filters shelf or in filter cards in the view)

   A measure filter can filter the data based on the values present in a measure. The aggregated measure values can be used in a measure filter to modify the data that are presented.

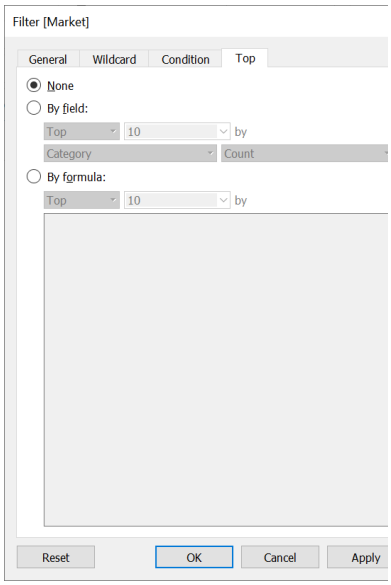
When filtering discrete fields, the filter can be set to include (or exclude) data by selecting individual categories. For continuous fields, the filter can be set by setting a range of allowable values. Conditions can also be set to return a subset of data records that meet a particular condition. When a discrete field is brought to the **Filter** shelf, the dialog box below is displayed:



General Filter

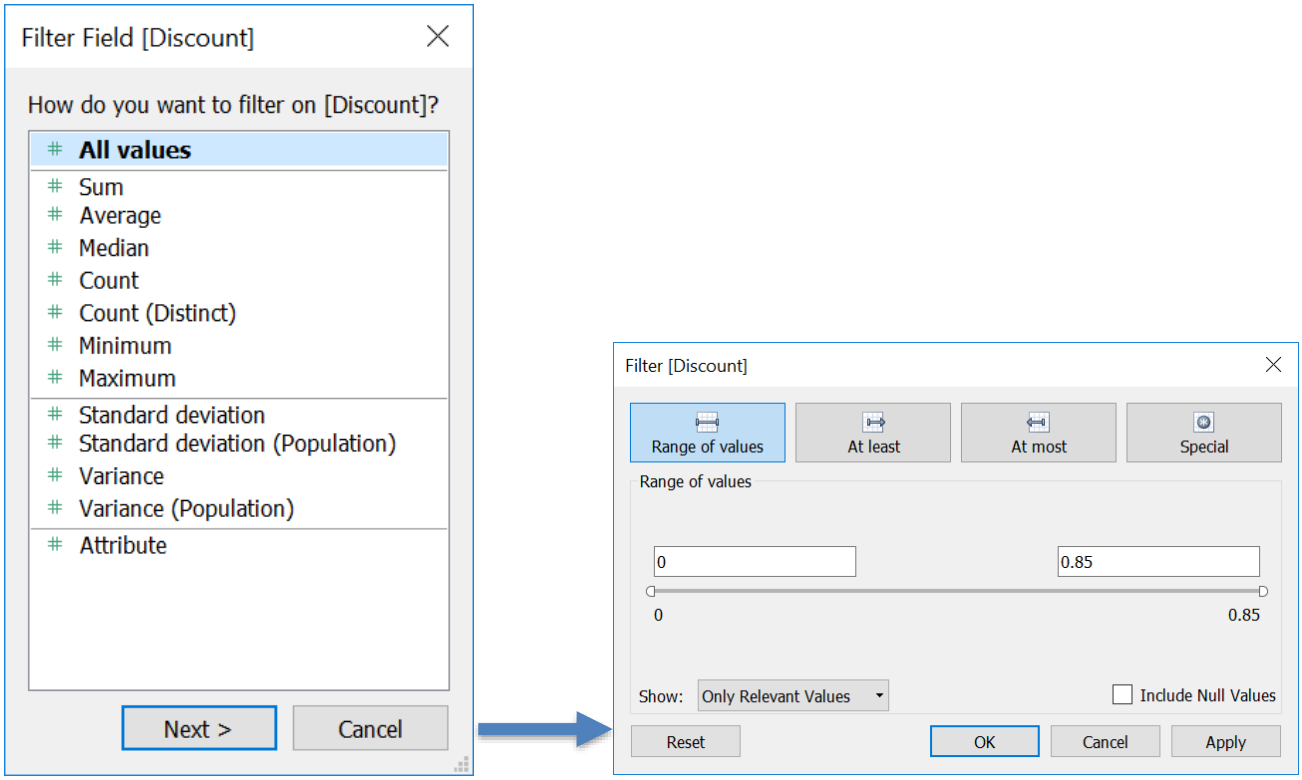


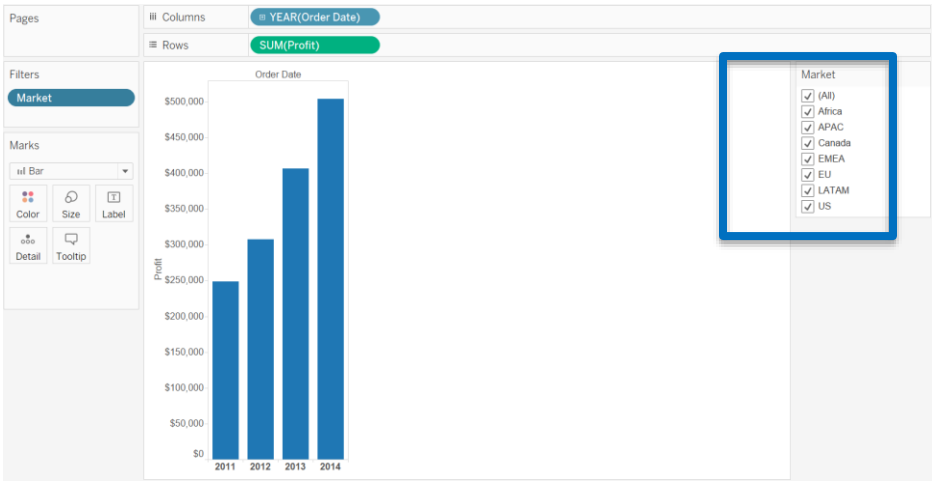
Condition Filter



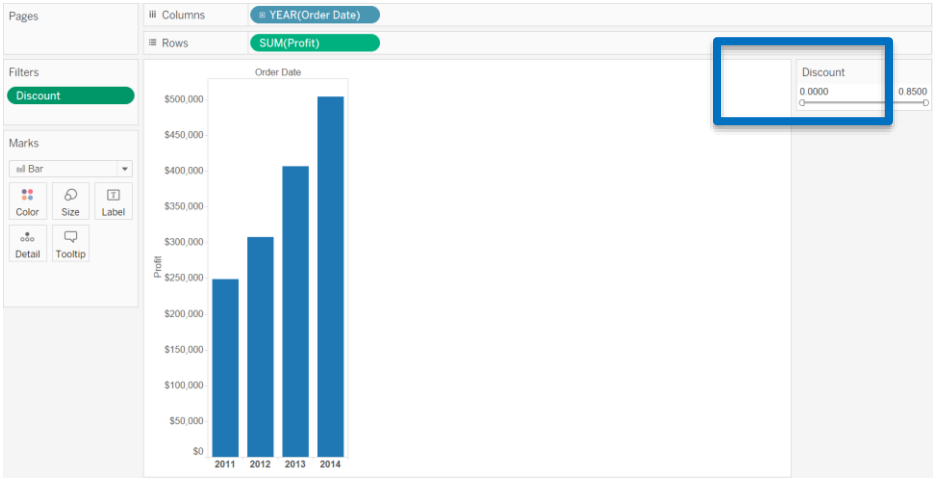
Top N Filter

Alternatively, when a continuous field is brought to the **Filter** shelf, the dialog boxes below are shown:





Discrete Filter

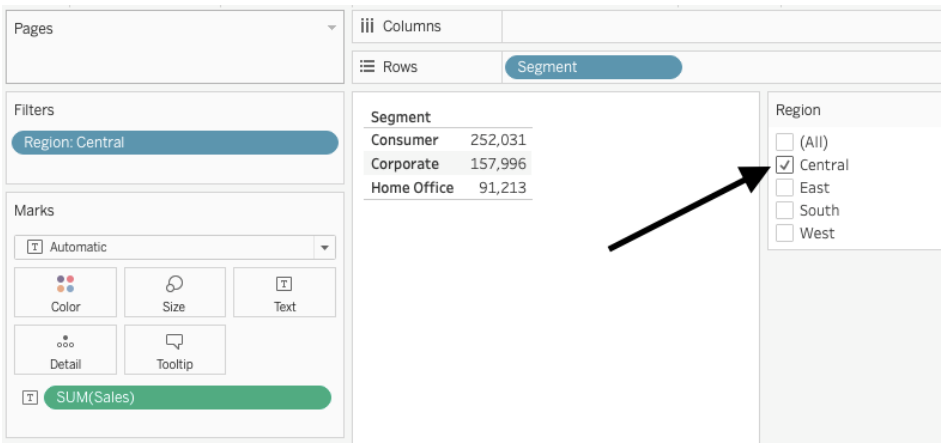


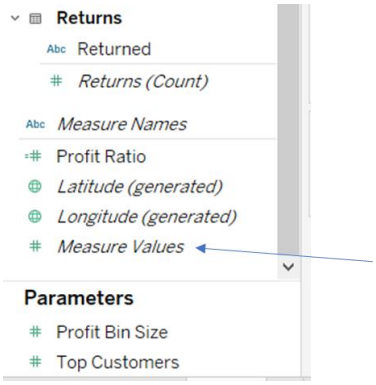
Continuous Filter

For more information on filters, review the websites: https://www.guru99.com/filter-data-tableau.html#6 or https://www.guru99.com/filter-data-tableau.html.

Quick Filtering Example

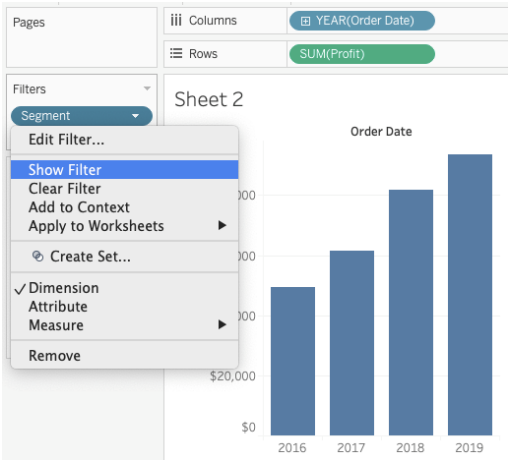
The regional manager from the South and East could use the same dashboard and select their respective filters. NOTE: the default is to not show the title for the Sales column. If you double-click “Measure Values” in the Data Pane.



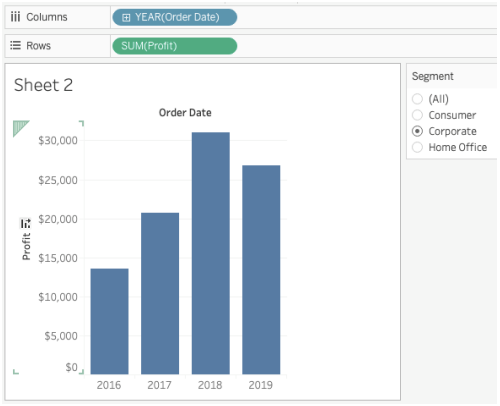


Continuing with the Superstore data and a new sheet, begin with bar chart of profit by year of order date.

Filtering on a discrete pill brings up options related to the specific list of values for that pill. Trying filtering by Segment, i.e., drag Segment to Filters shelf. Then Show Filter. The first time it may pop-up on its own. After that you need to click on drop-down arrow by Segment and select show filter.



Here is an example of the Sales data filtered on Corporate segment.



Context Filtering Example

By default, all filters that you set in Tableau are computed independently. That is, each filter accesses all rows in your data source without regard to other filters. However, you can set one or more categorical filters as context filters for the view. You can think of a context filter as being an independent filter. Any other filters that you set are defined as dependent filters because they process only the data that passes through the context filter.

You may create a context filter to:

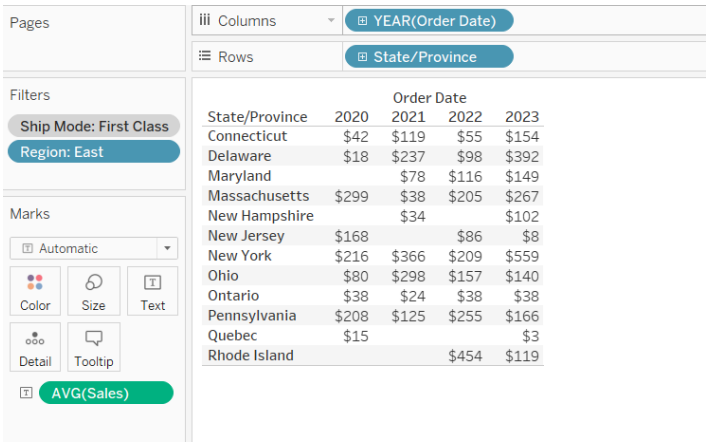
- Force a filter to be carried out first.
- Create a dependent numerical or top N filter. You can set a context filter to include only the data of interest, and then set a numerical or a top N filter (discussed in Topic 6).

To create a context filter, select Add to Context from the context menu of an existing categorical filter. The context is computed once to generate the view. All other filters are then computed relative to the context. Context filters:

- Appear at the top of the Filters shelf.
- Are identified by a gray color on the Filters shelf.

- Cannot be rearranged on the shelf.

As shown below, the Ship Mode dimension is set to be the context for a view. The Region filter is computed using only the data that passes through Ship Mode.



	Order Date			
State/Province	2020	2021	2022	2023
Connecticut	$42	$119	$55	$154
Delaware	$18	$237	$98	$392
Maryland		$78	$116	$149
Massachusetts	$299	$38	$205	$267
New Hampshire		$34		$102
New Jersey	$168		$86	$8
New York	$216	$366	$209	$559
Ohio	$80	$298	$157	$140
Ontario	$38	$24	$38	$38
Pennsylvania	$208	$125	$255	$166
Quebec	$15			$3
Rhode Island			$454	$119

You can modify a context filter by:

- Removing the field from the Filters shelf – If other context filters remain on the shelf, a new context is computed.
- Editing the filter – A new context is computed each time you edit a context filter.
- Selecting Remove from Context – The filter remains on the shelf as a standard filter. If other context filters remain on the shelf, a new context is computed.

If you are having a problem with your filters, remember to consider the context.

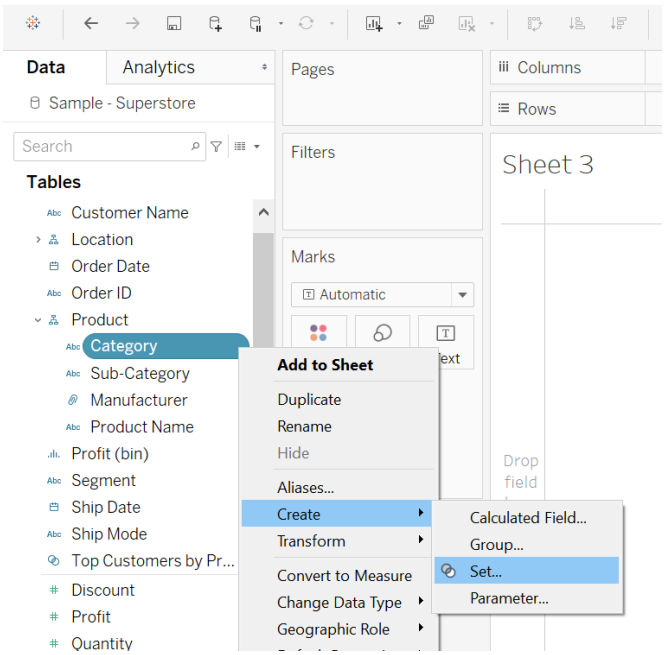
Sets

Sets allow you to create new fields based on existing fields so that the data is easier to work with or tells a better story. You will use a set when you want to create a custom field that defines some binary subset of data based on some conditions. For example, this could include:

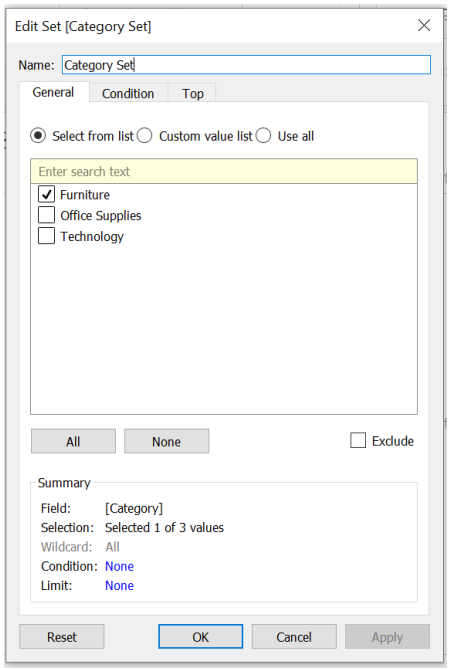
- Splitting data into two buckets (binary choice) based on a custom condition
- Filtering data in a more meaningful way
- Using specific data in a calculation
- Summarizing scattered data for visualizations.

Set: Furniture vs Other Products

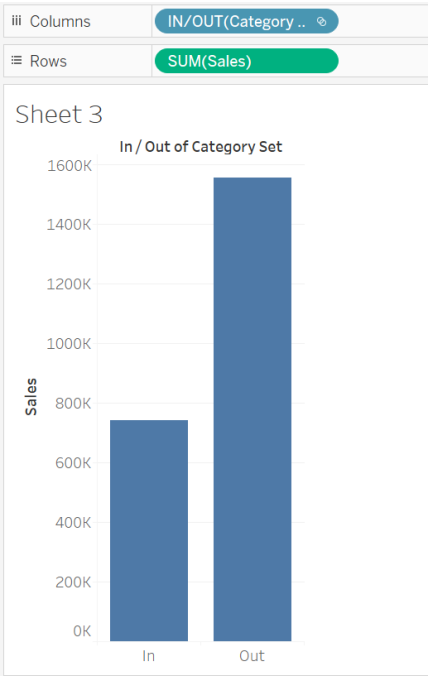
Assume you want to better understand your Furniture product versus the other products (Office supplies and Technology). Right-click the Category field and choose Create -> Set.



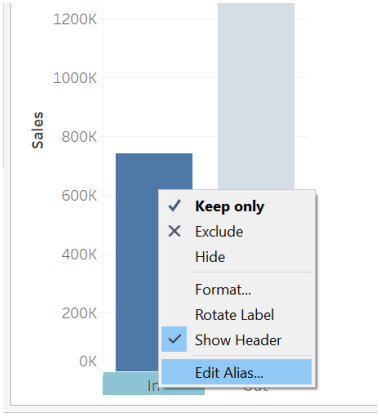
Name this set, Category Set, and select Furniture to be in the set.



Make a bar graph that shows the sales for Furniture versus Not Furniture (i.e., Office Supplies and Technology).



It may be hard to remember what is in the set or not, so to make your labels more descriptive, right-click on the ‘In’ label and choose ‘Edit Alias’ and give In and Out more descriptive, e.g., Furniture and Office Supplies & Technology.

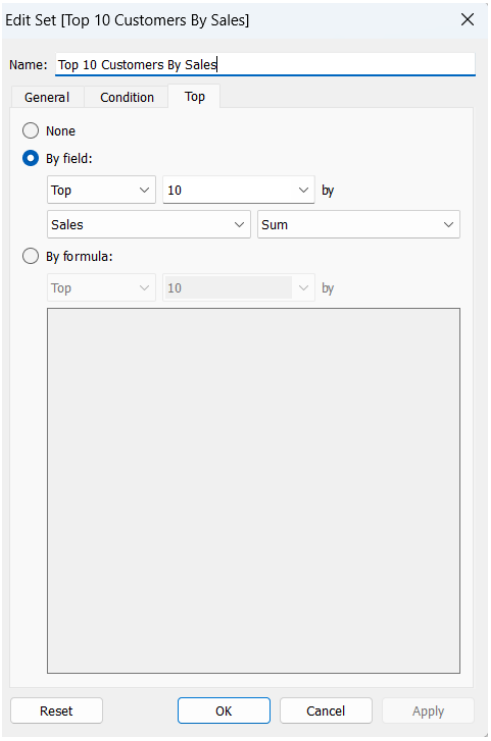


You may want to use Sets as filters but remember Sets are binary so when you add them as a filter, you will only have the option to choose whether the marks on the view are in or out of the set.

You can also create sets by selecting dimensions on your visualization and then right-clicking and choosing Create Set. Whatever items were highlighted will be in the set, and the other items will be out of the set.

Set: Top 10 Customers versus the Rest

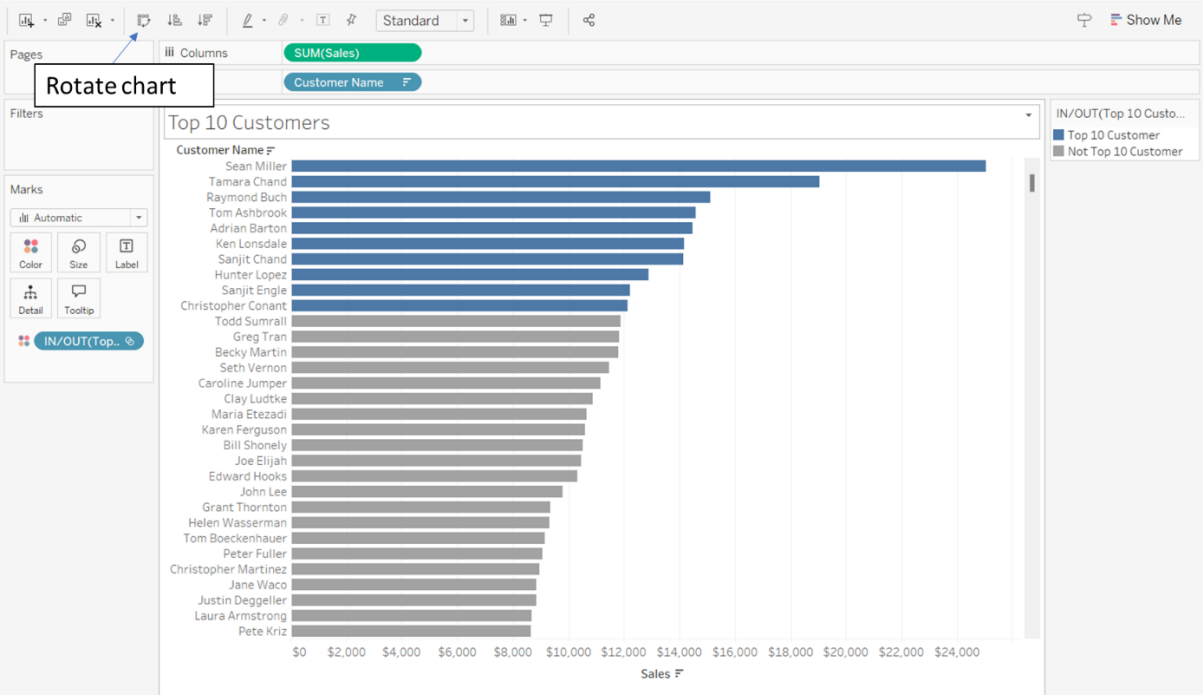
Assume you want to identify the top 10 customers by total sales in the Superstore dataset. From the "Data" pane, right-click on the "Customer Name" field, and select "Create" > "Set." In the "Create Set" dialog box, name the set "Top 10 Customers by Sales." Switch to the "Top" tab, choose "By Field," and set it to the top 10 by "Sum of Sales." Click "OK."



Create a new worksheet to compare the total sales of these top 10 customers against all other customers. You can do this by dragging "Sales" to the Rows shelf and "Customer Name" to the Columns shelf.

Drag the "Top 10 Customers by Sales" set onto the color mark to differentiate the top customers from the rest.

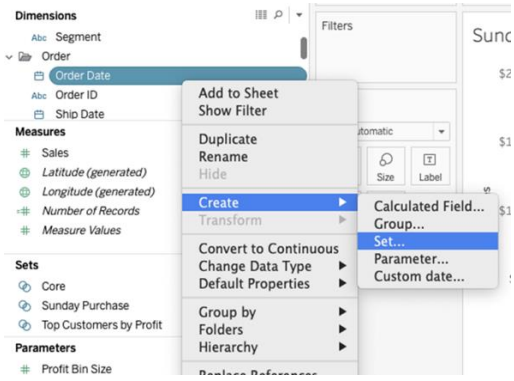
Sort the data so that the top 10 customers are first and then rotate the chart so that the names are easy to read. Edit Aliases to be more descriptive.



Set: Sunday versus non-Sunday

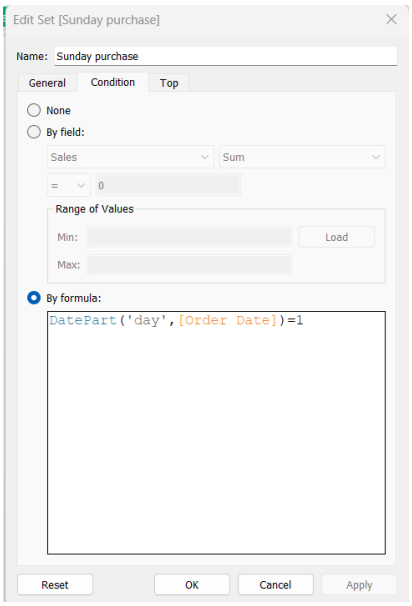
Here is a more complicated example of a Set. Assume for Superstore data, there is a belief that more shopping happens on Sundays. Instead of working with dd/mm/yyyy dates, we can use a set to isolate Sunday vs. non-Sunday order dates.

Right click the ‘Order Date’ dimension and select Create and then Set. A box will appear and you can click on the Condition tab.

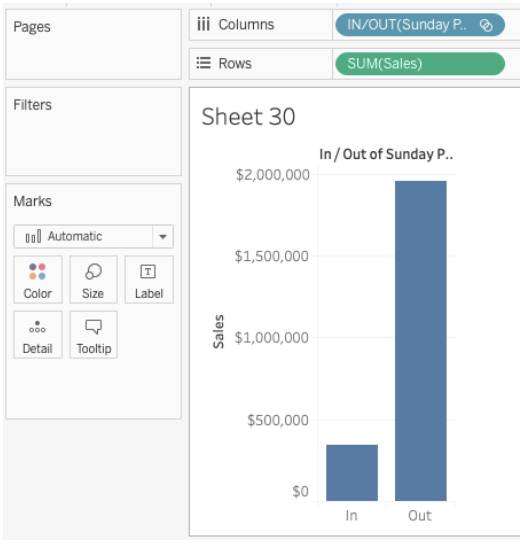


Fill in a name for the Set such as Sunday Purchase. Select ‘By Formula’ option. The DATEPART function in this example takes a date and assigns it a number value from 1 (Sunday) to 7 (Saturday). Enter the Formula:

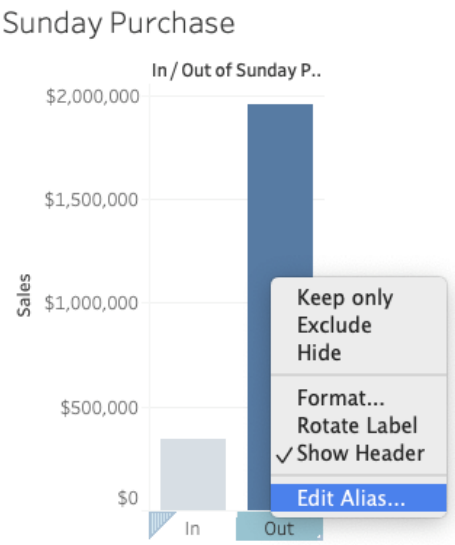
DATEPART(“day”,[Order Date]) = 1



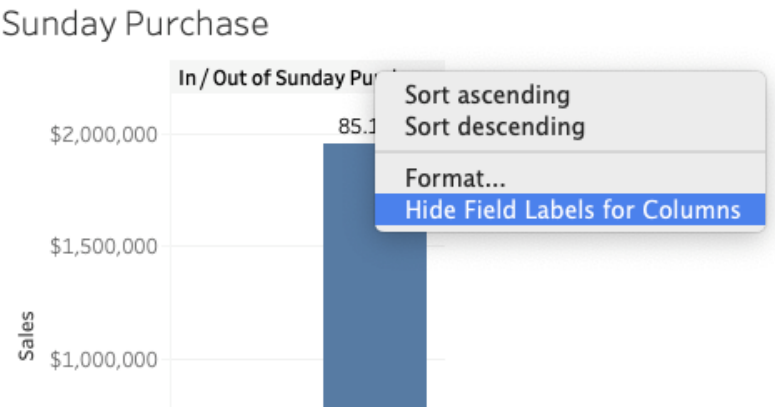
The Set can be used in the visualization to better understand shopping behaviors on Sunday vs. other days of the week. Double click sales and then the new ‘Sunday Purchase’ set you just created.



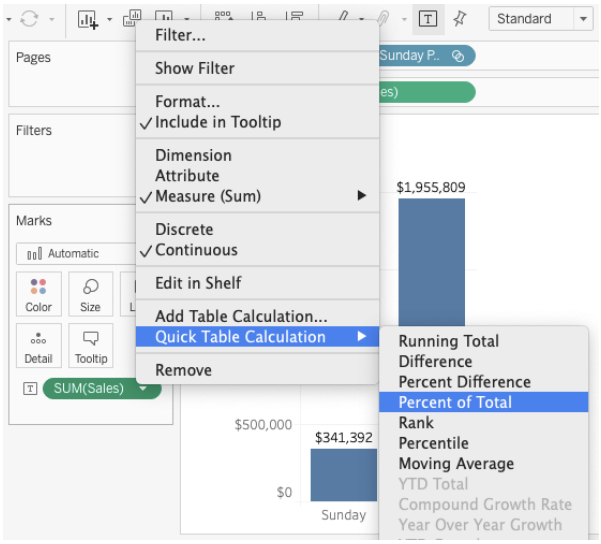
Right click on the title and edit it so that it shows ‘Sunday Purchases’. Instead of using the default ‘Out’ label, right click it and edit alias. Change it to ‘Non-Sunday’. Do the same for the ‘In’ label, but name it ‘Sunday’ this time.



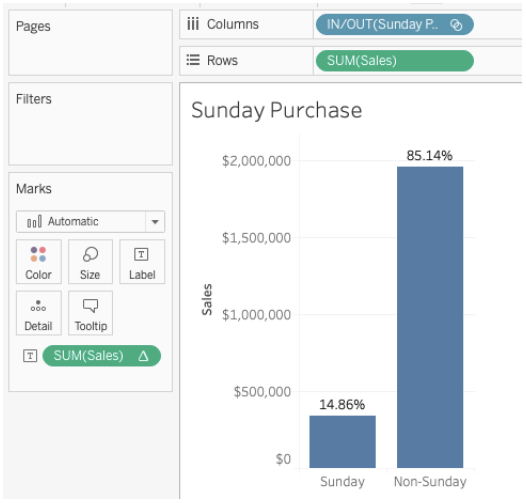
Right click on the field label above the chart and select ‘Hide Field Labels for Columns’.



Drag a sales pill onto the label mark card. Right click on the pill and change it to a ‘percent of total’ quick table calculation as shown below.



Here is the final visualization below. Remember, yours will be slightly different because as Tableau updates the years, which days are Sundays also changes.



Topic 4: Calculated Fields and Table Calculations

This section introduces Tableau calculations. Similar to formulas in Excel, calculations allow you to manipulate data in any number of ways, and in many cases you will need to create new data from existing data through calculated fields.

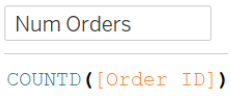
If you want to put notes or comments in a data field calculation, simply put two forward slashes (//) in front of the text or code you want to be a comment. Tableau will render this piece 'light-gray', and the text or code will be ignored in the calculation.

Simple Calculations in Tableau

Example 1:

Assume you want to find the average order value (AOV) in the Superstore dataset. You will learn an easier way to do create this calculation in Topic 5 (Level of Detail Calculations), but for now let's look at the steps to find it with simple calculated fields.

1. Create a calculated field that counts the distinct number or orders.



Num Orders will be the denominator of the average order value calculation.

2. Create a calculated field called AOV which is the aggregated Sum of Sales divided by [Num Orders]. Num Orders is already aggregated by COUNTD in the previous step.

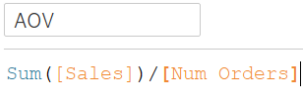
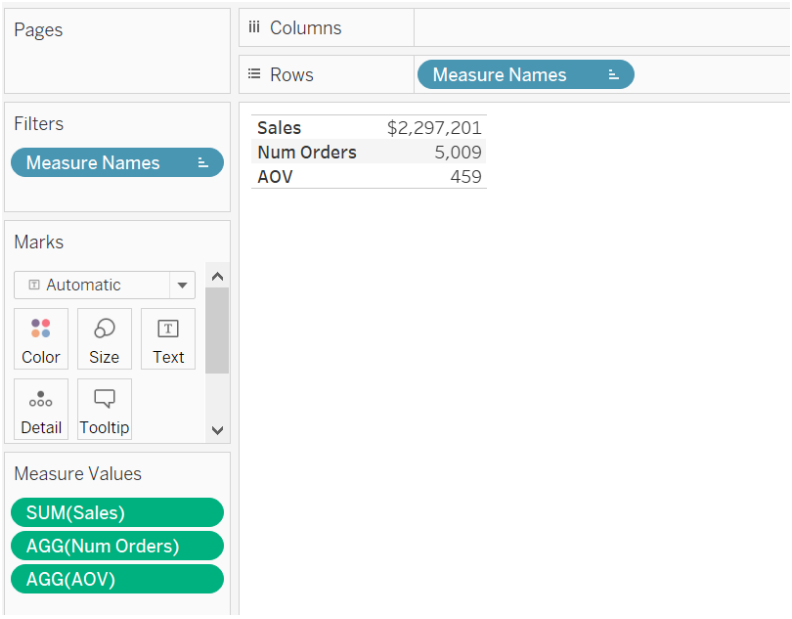


Tableau said that the calculation is valid but how can you tell it was correct?

Can you make a table to show the sales, number of orders, and the newly created AOV measure. Is it correct?

(HINT: Double click on each of the three variables.)



Example 2:

Start with Superstore sample data. Add **Profit** to *Columns*, **Segment** and **Sub-category** to *Rows*, Filter **Category** to Furniture.

What if you want, for emphasis, to have one distinct color for positive profit and one for negative profit?

Calculated fields are created by defining a formula:

1. Select the Analysis menu (or click the down triangle for a variable and select Create -> Calculated Field) and then Create Calculated Field.

2. At the top name it **Sign of Profit**

3. Put the following into the calculation box:

```
IF SUM([Profit]) > 0
  THEN "Positive"
  ELSE "Negative"
END
```

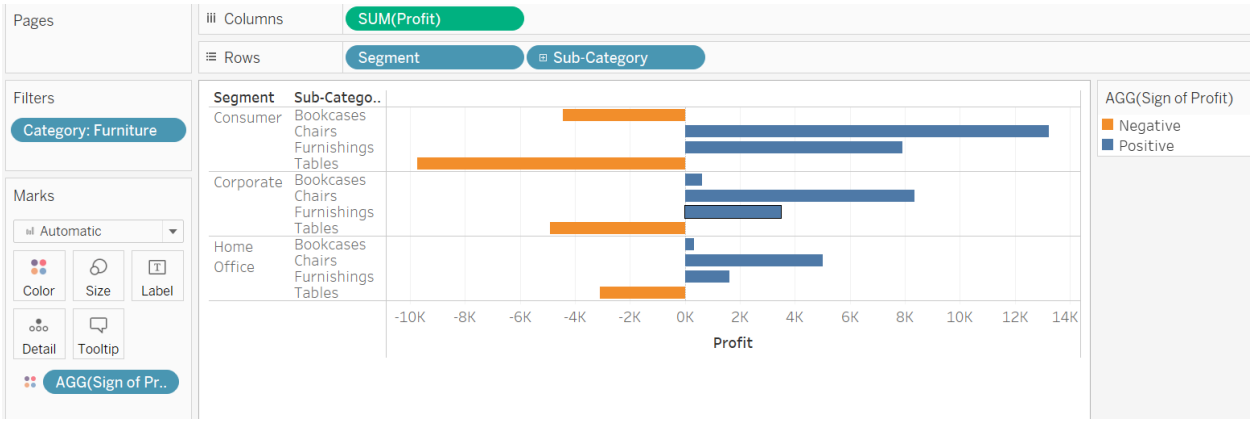
4. The SUM here will add the profit for all observations in the aggregation – remember that Measures need an aggregation level.

5. Note that if Profits are actually zero you are calling it positive. This is just for simplicity.

6. I indent because I think it makes your calculated field script easier to read

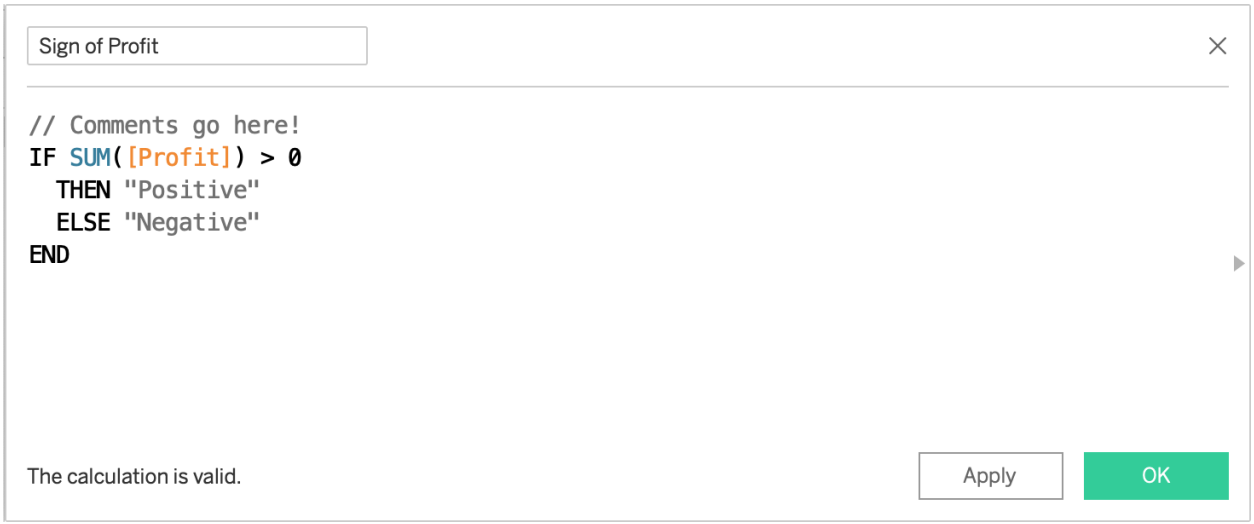
7. Note the little equal sign in front of the field that indicates it's a calculated field, which is not from the data source

8. Bring the calculated field to *Color.* If you don't like the default colors, click on *Color* and you can change the default colors.

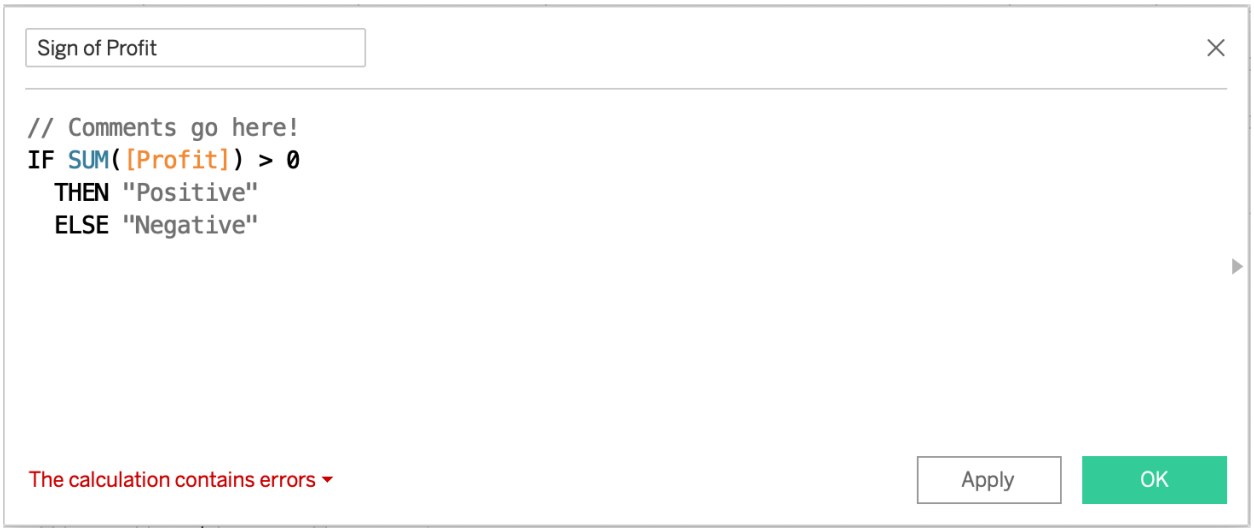


How do I know the calculation will work?

Tableau will tell you in the bottom left corner of the calculation (notice the small gray text) if you are creating a valid calculated field. This one is correct:



But this one is not:



If you click on the red text you will get a (slightly informative) guess about why the formula doesn't work. In this case the final 'END' is missing on the last line so Tableau doesn't know when to stop looking for it:

But be careful because a valid calculated field is not always valid math.

Example 3:

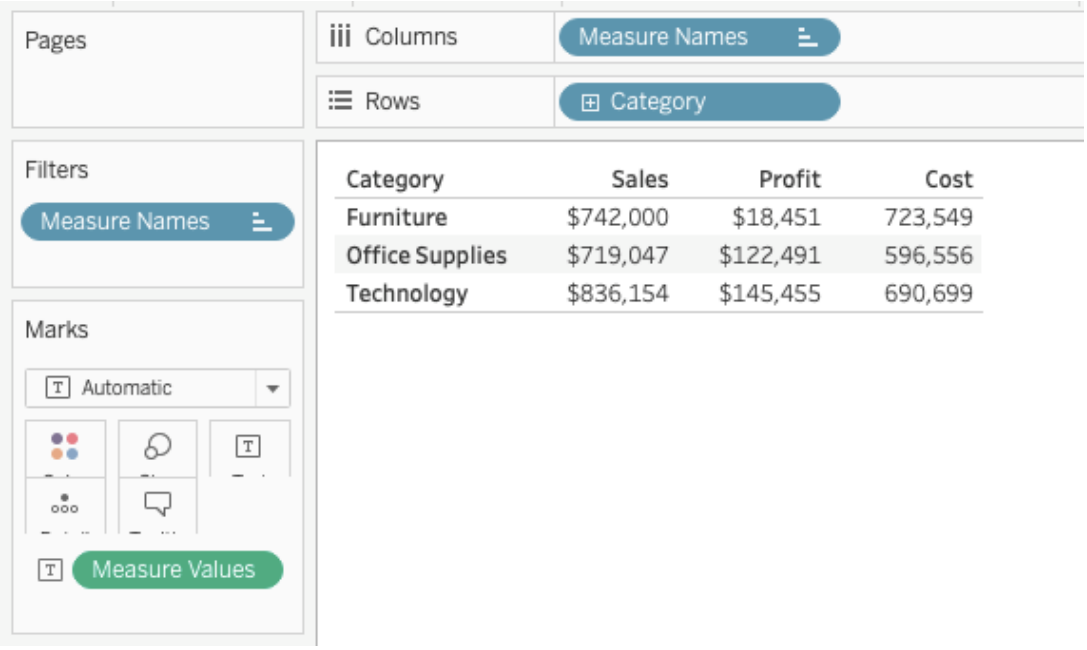
Let's visualize Cost, defined as Sales minus Profit.

Add a Calculated Field named **Cost**:

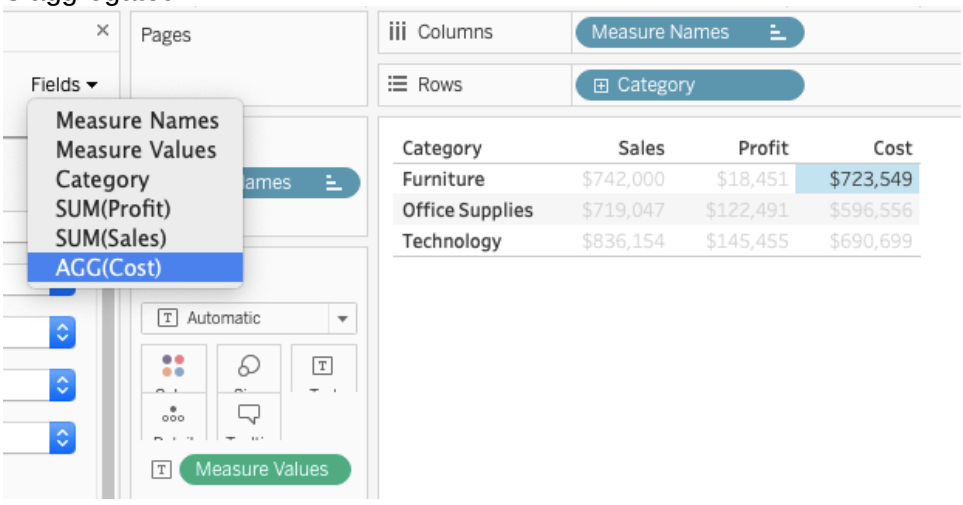
```
SUM([Sales]) - SUM([Profit])
```

In this case you could do [Sales] – [Profit] and get the same result but be careful because there are situations when calculations with aggregated fields and those with disaggregated fields will not yield the same result. We will see some of these examples later.

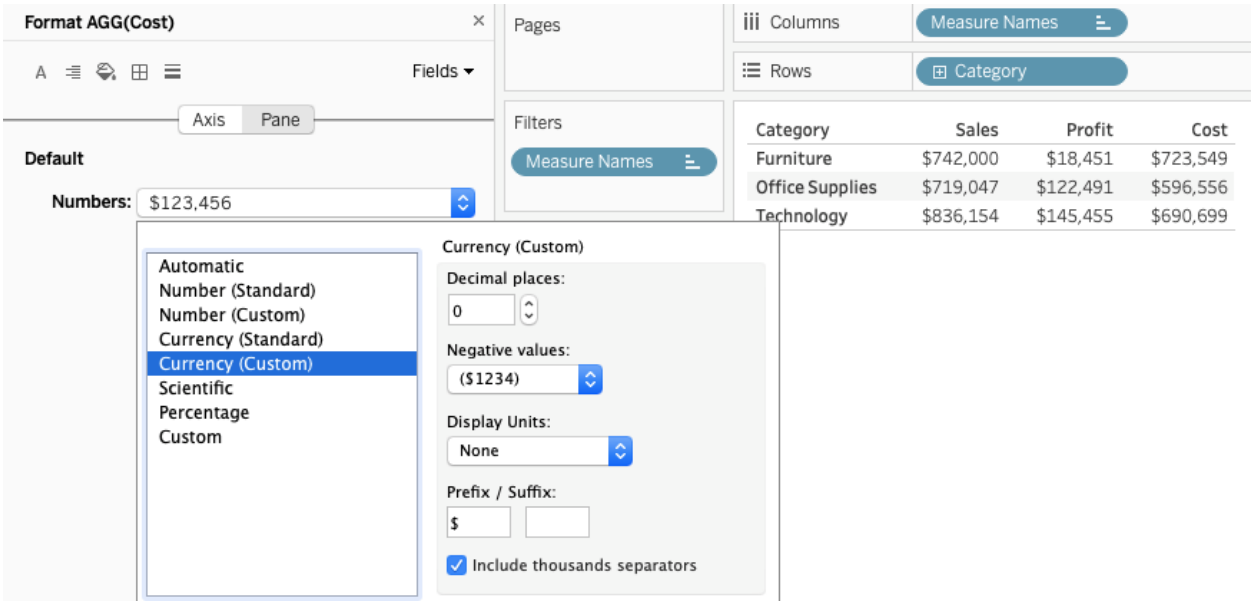
You can see how it is aggregated by category below:



In order to format the Cost column to show dollars, right click on of the values in the column and select ‘Format’. On the left side of the screen, a formatting pane will open and you should select the new Cost field by using the dropdown arrow as shown below. It will show as AGG(Cost) because it is aggregated.



Under the Numbers dropdown, select Currency (Custom) and reduce the number of decimal places down to zero so that it matches the other fields.



Row Level vs. View (Aggregation) Level Calculations

Aggregation is an important concept to consider when creating calculated fields. How you write a calculation can have different results. The following two calculations will calculate a profit ratio:

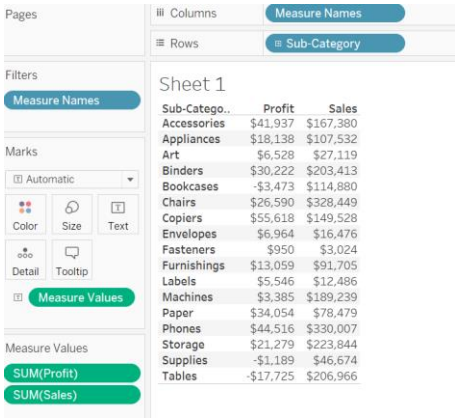
```
[Profit]/[Sales]
```

And

```
SUM([Profit])/SUM([Sales])
```

These are two very different calculations. The first is not aggregated and performs the division for each row. The second performs the sum and then performs the division. Try it to see the difference:

- Drag Sales to Text, Sub-Category to Rows, and Profit to the view directly over the Sales numbers.



Sub-Catego..	Profit	Sales
Accessories	$41,937	$167,380
Appliances	$18,138	$107,532
Art	$6,528	$27,119
Binders	$30,222	$203,413
Bookcases	-$3,473	$114,880
Chairs	$26,590	$328,449
Copiers	$55,618	$149,528
Envelopes	$6,964	$16,476
Fasteners	$950	$3,024
Furnishings	$13,059	$91,705
Labels	$5,546	$12,486
Machines	$3,385	$189,239
Paper	$34,054	$78,479
Phones	$44,516	$330,007
Storage	$21,279	$223,844
Supplies	-$1,189	$46,674
Tables	-$17,725	$206,966

- Open the Calculated Field editor to create a calculated field “Row Level Profit Ratio” and write the calculation: [Profit]/[Sales]
- Drag the field to the view to see another column of numbers.
- Format these numbers as percentages and see VERY LARGE profit ratios. Something is wrong. The problem is that the calculation is a row level calculation. You are calculating a row level profit ratio for thousands of rows of data. You are then aggregating these row level results into an insanely large decimal number.
- To fix this, open the calculated field editor and create a field called Aggregate Profit Ratio that is SUM([Profit])/SUM([Sales]).
- Drag the field to the view. You will now see another column of numbers. Format these numbers as a percentage. Your profit ratios should look more reasonable now.

Sub-Category	Profit	Sales	Row Level Profit Ratio	Aggregate Profit Ratio
Accessories	$41,937	$167,380	16911.25%	25.05%
Appliances	$18,138	$107,532	-7310.11%	16.87%
Art	$6,528	$27,119	20031.00%	24.07%
Binders	$30,222	$203,413	-30398.33%	14.86%
Bookcases	-$3,473	$114,880	-2887.39%	-3.02%
Chairs	$26,590	$328,449	2708.61%	8.10%
Copiers	$55,618	$149,528	2156.92%	37.20%
Envelopes	$6,964	$16,476	10747.75%	42.27%
Fasteners	$950	$3,024	6492.00%	31.40%
Furnishings	$13,059	$91,705	13117.25%	14.24%
Labels	$5,546	$12,486	15639.75%	44.42%
Machines	$3,385	$189,239	-828.30%	1.79%
Paper	$34,054	$78,479	58307.25%	43.39%
Phones	$44,516	$330,007	10598.83%	13.49%
Storage	$21,279	$223,844	7539.00%	9.51%
Supplies	-$1,189	$46,674	2128.75%	-2.55%
Tables	-$17,725	$206,966	-4712.48%	-8.56%

Functions Available to a Calculated Field

There are many different functions available to you when you create a calculated field. To see a list start creating a Calculated Field as before. If you start typing you will see some functions pop up. However, how do you know what is even available? There is a small arrow pointing to the right just to the right of the dialog where you define the function. If you click that arrow you

will see many different functions. You can search for them in the box or filter with the dropdown. You can find out more here: https://onlinehelp.tableau.com/current/pro/desktop/en-us/functions_functions_tablecalculation.htm

Some of the most common are:

Number Functions:

ABS: This function returns the absolute value of the specified number.
Example: ABS(-10) = 10
CEILING: It rounds the given number to the nearest integer of equal or greater value.
Example: CEILING(5.1265) = 6
FLOOR: It rounds the given number to the nearest integer of equal or lesser value.
Example: FLOOR(5.1265) = 5
MIN: Returns the minimum value of an expression across all records
Example: MIN([Profit]) – This expression will return the minimum value of the profit across all the profit values
MAX: Returns the maximum value of an expression across all records
Example: MAX([Profit]) – This expression will return the maximum value of the profit across all the profit values
ZN: "ZN" stands for "Zero if Null". It means it returns the given expression if it is not null, otherwise returns zero.

String Functions:

CONTAINS: This function returns "True" if the user's mentioned sub-string is present in the given string.
Example: CONTAINS("University", "versi") = True
ENDSWITH: It returns "True" if the given string ends with the user's mentioned sub-string
Example: ENDSWITH("University", "sity") = True
FIND: Similar to excel "Find" functions, in Tableau also FIND functions returns the position of the mentioned sub-string within the given string. If the sub-string isn't found then it returns 0(zero)
Example: FIND("University", "versi") = 4
ISDATE: This function is used to check whether the given string is a valid date. If the string is valid date then it returns "True"
Example: ISDATE("2017-05-12") = True
LEFT: It returns the specified number of characters from the start of the string.
Example: LEFT("Calculation", 4) = Calc
LEN: This function returns the number of characters in the given string.
Example: LEN("Calculation") = 11
LOWER: Convert a text string to all lowercase letters
Example: LOWER("CalCulation") = calculation
MID: It returns the characters from the middle of a text string given a starting position and

length.
Example: MID(“Tableau”,2,4) = able
RIGHT: It returns the specified number of characters from the end of the given string.
Example: RIGHT(“Calculation”, 4) = tion
TRIM: Returns the string with both trailing and leading spaces removed
Example: TRIM(“ Budget “) = Budget
UPPER: Convert a text string to all uppercase letters
Example: LOWER(“Budget”) = BUDGET

Date Functions:

DATEADD: Add an increment to the specified date and returns the new date.
Example: DATEADD(‘month’, 3, #2017-07-12#) = 2017-10-12 12:00:00 AM
DATEDIFF: Returns the difference of the two dates where start_date is subtracted from end_date
Example: DATEDIFF(‘month’, #2017-04-12#, #2017-07-15#) = 3
DATENAME: Returns the part of the given date as string.
Example: DATENAME(‘month’, #2017-05-14#) = May
MONTH: Returns the month of the given date as an integer
Example: MONTH(#2018-02-19#) = 2
TODAY: Returns the current date
YEAR: Returns the year of the given date as an integer
Example: YEAR(#2018-02-19#) = 2018

Logical Functions:

AND: This function performs the logical conjunction of two or more expression. “AND” returns “True” when all the given expressions are true.
CASE: This function finds the first that matches the given and returns the corresponding
Example:
CASE [Name] WHEN “John” THEN 90
WHEN “Emma” THEN 95
END
IF-ELSE: It tests a series of expressions and returns only that value corresponding to the first expression is true
Example:
IF [Profit] >0 THEN “Profitable”
ELSEIF [Profit] = 0 THEN “Breakeven”
ELSE “Loss”
END
OR: This function performs the logical disjunction on two or more expression. “OR” returns “True” when any of the given expressions are true.

Aggregation Functions:

AVG: It returns the average value of the given expression or array of values
COUNT: It returns the count of items in the group. NULL values are not counted
COUNTD: It returns the unique count of items in the group.
MAX: Returns the maximum value of an expression across all records
Example: MAX([Profit]) – This expression will return the maximum value of the profit across all the profit values
MEDIAN: Returns the median value of an expression across all records
Example: MEDIAN([Profit]) – This expression will return the median value of the profit across all the profit values
MIN: Returns the minimum value of an expression across all records
Example: MIN([Profit]) – This expression will return the minimum value of the profit across all the profit values
PERCENTILE: an aggregate calculation that returns the percentile value from the given expression, corresponding to the specified percentile value
Example: PERCENTILE([Sales],0.9) = This expression will return the 90th percentile value from all the Sales value
SUM: It returns the sum of the given expression or array of values

Window Functions:

WINDOW_SUM([Field]): Calculates the sum of a specified field over a specified window, which can be defined by dimensions or table calculations.
WINDOW_AVG([Field]): Calculates the average of a specified field over a specified window.
WINDOW_MIN([Field]): Computes the minimum value of a specified field within a defined window.
WINDOW_MAX([Field]): Computes the maximum value of a specified field within a defined window.
WINDOW_COUNT([Field]): Counts the number of records or rows within a specified window. It can be useful for creating calculated fields that involve counting.
WINDOW_PERCENTILE([Field], [Percentile]): Calculates the specified percentile of a field within a defined window. You can specify the desired percentile as a parameter.
WINDOW_RANK(): Assigns a rank to each row within a partition defined by dimensions. Rows with the same values in the partition receive the same rank.
WINDOW_PERCENTILE_RANK([Field]): Calculates the percentile rank for each record in the data based on a specified field.
WINDOW_FIRST_VALUE([Field]): Returns the first value of the specified field within a specified window. This is useful for finding the first occurrence of a value within a partition.
WINDOW_LAST_VALUE([Field]): Returns the last value of the specified field within a specified window. It's helpful for finding the last occurrence of a value within a partition.
WINDOW_SUM([Field], FIRST:LAST): Calculates the sum of a specified field over a range of rows within a window. You can specify the range using the FIRST and LAST keywords.
WINDOW_STDDEV([Field]): Computes the standard deviation of a specified field within a specified window.
WINDOW_VARIANCE([Field]): Calculates the variance of a specified field within a specified window.

WINDOW_MEDIAN([Field]): Calculates the median of a specified field within a specified window.
WINDOW_NTH_VALUE([Field], [N]): Returns the value of the specified field in the Nth row within a defined window.

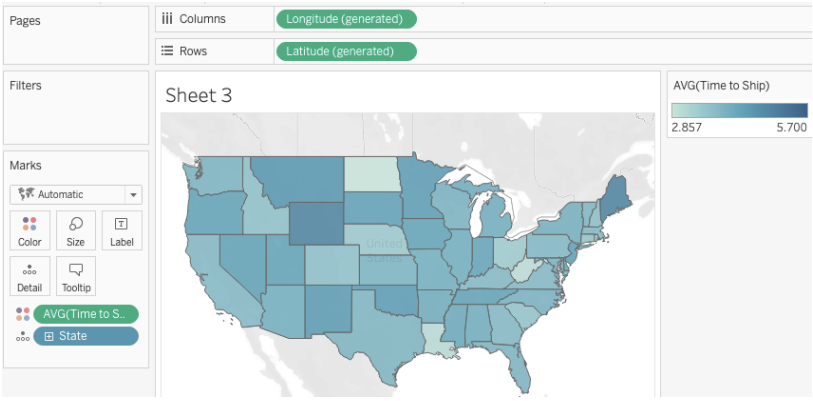
Example: Time between Dates

A common question could be something like “How much time did something take?” when you have a start and an end date in your data. A date calculation can handle that.

- Let’s look at the average time it takes for an order to ship
    - Click the menu in the data window and select Create Calculated Field
- Name it “Time to Ship”
- You’ll use a function called DATEDIFF to get the amount of time between Order Date and Shipping Date.
- Search for DATEDIFF; the first argument needed is the date part. Type ‘day’ and a comma
- Next, drag out Order Date, another comma, and Ship Date:

```
DATEDIFF('day', [Order Date], [Ship Date])
```

- Be sure to check that the calculation is valid, then click OK
- Now you can color our states by using Time to Ship
    - Create a new sheet
    - Select State and Time to Ship (hold down Control key to select), Choose maps from Show Me
    - You’ll change the aggregation to Average by clicking Time to Ship pill in the marks card



Example: Time to Now

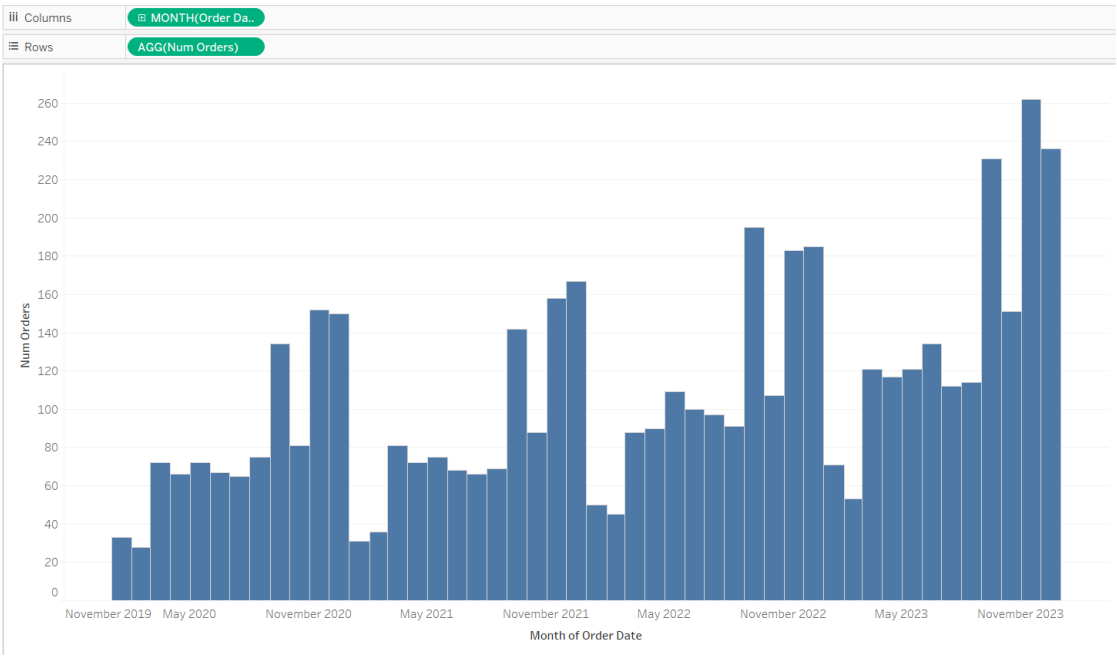
- Click the menu in the data window and select Create Calculated Field
- Name it “Days Since Last Order”
- You’ll use a function called DATEDIFF to get the amount of time between Order Date and Now.

```
DATEDIFF('day', [Order Date], Today())
```

- Be sure to check that the calculation is valid, then click OK

Using a Calculated Field to Highlight a Value

You want to use a Calculated Field to identify the maximum (or minimum) value in a visualization so that you can use color to make it stand out. Using our previously created calculated field Num Orders, create a visualization for Num Orders by Month:



Next, try to make a calculated field that you can use to highlight the maximum bar in this visualization. Try this (note: it isn't going to work):

Highlight Best

```
If [Num Orders] = Max([Num Orders]) Then
1
ELSE
0
END
```

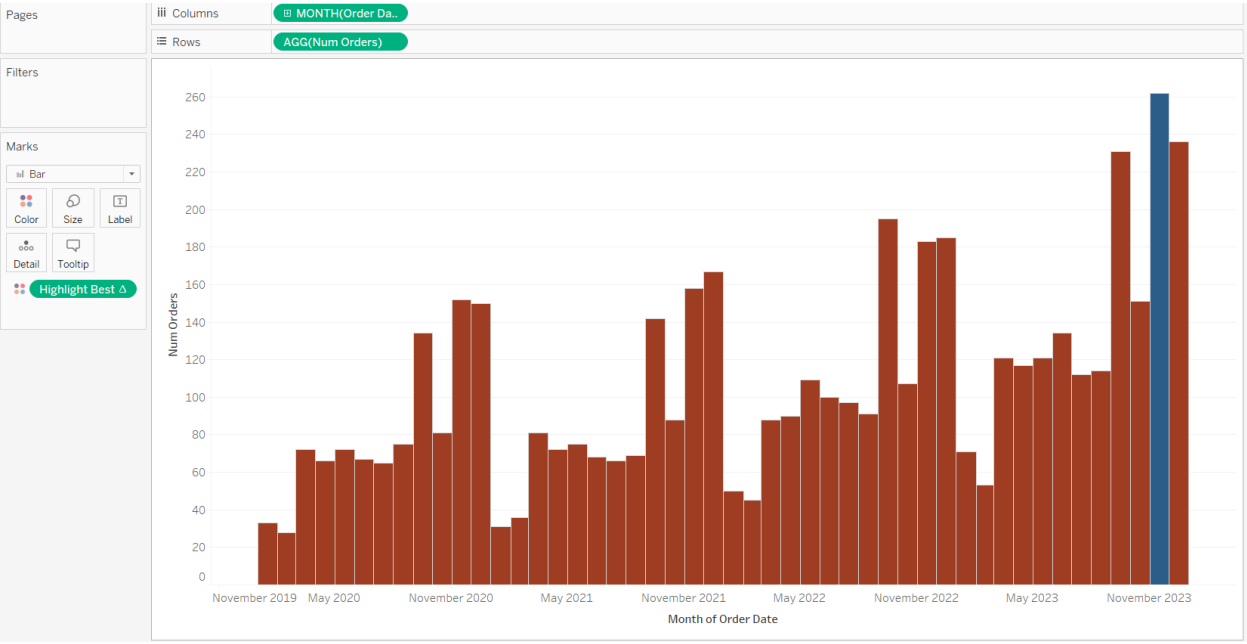
It is giving you an error that says "Cannot mix aggregate and non-aggregate arguments with this function." It is getting confused because of the way Tableau handles the aggregation of data. The error occurs because Tableau is designed to work with aggregate functions at multiple levels of detail in your visualization, and it expects you to be explicit about how you want the aggregation to be applied.

You can fix this by specifically telling it that it should be the Maximum of the Window/Visualization.

Highlight Best

```
Results are computed along Table (across).
If [Num Orders] = WINDOW_MAX([Num Orders]) Then
1
ELSE
0
END
```

This formula compares the number of orders to the maximum number of orders within the window (your dataset). If the number of orders equals the maximum, it will be labeled as "1"; otherwise, it will be labeled as "0". If you then add this calculated field to the color marks card, it will make the maximum value stand out in the visualization.



Note: this still does not always work. If you try to do this with for example [Sales], this won't work, but that is because it does not know what you want to do with sales.

What doesn't work:

Highlight Best Sales

```
If [Sales] = WINDOW_MIN([Sales]) Then
1
ELSE
0
END
```

It doesn't know what about Sales that you are trying to do. You can identify the maximum (or minimum) Average Sales or Sum of Sales:

Highlight Best Sales

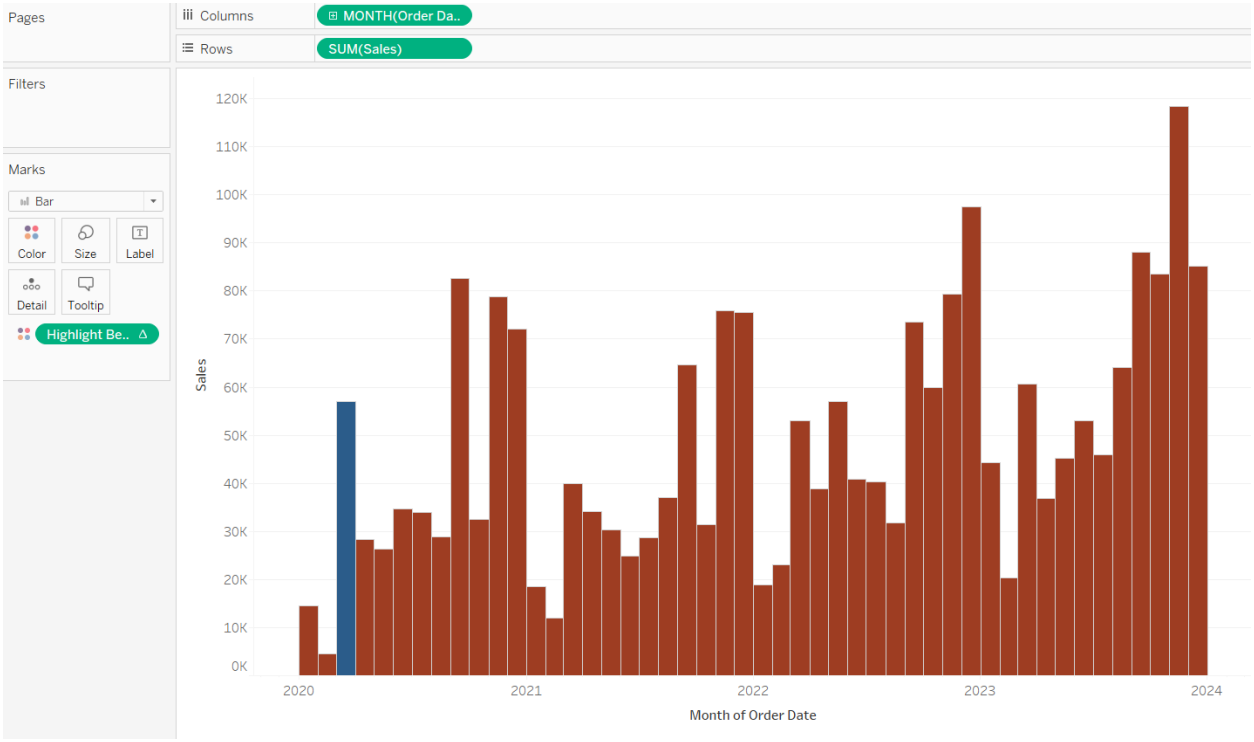
```
Results are computed along Table (across).
If Avg([Sales]) = WINDOW_MAX(Avg([Sales])) Then
1
ELSE
0
END
```

And this will work:

Highlight Best Sales

```
Results are computed along Table (across).
If Sum([Sales]) = WINDOW_MAX(Sum([Sales])) Then
1
ELSE
0
END
```

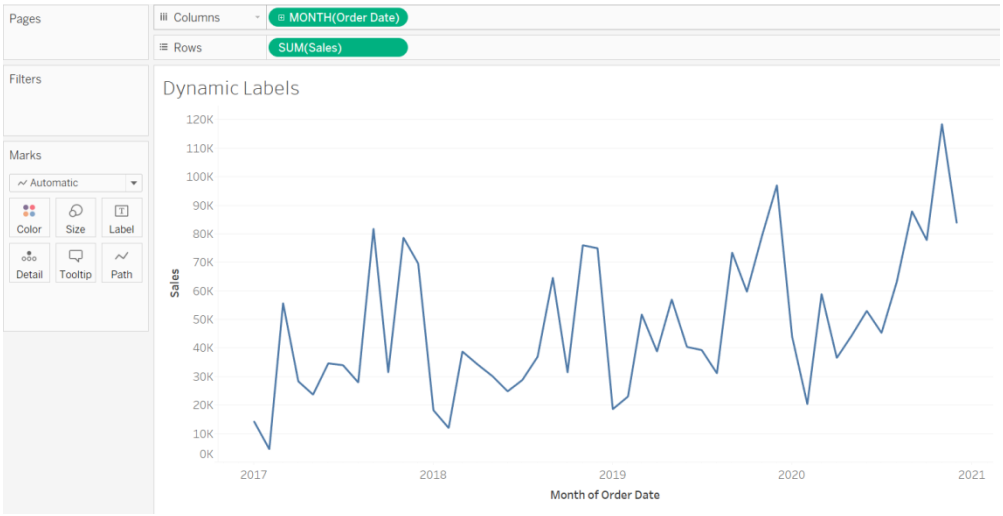
This visualization shows the sum of sales but then highlights the month with the highest average sales:



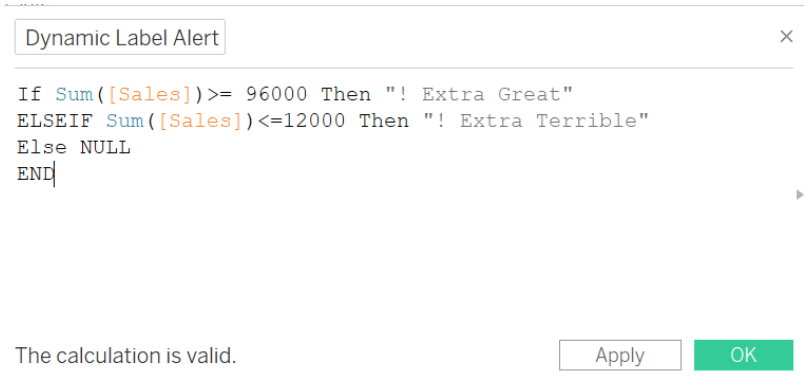
Dynamic Labels

1. What if we want mark labels to only show up when something extra important happens? The trick is to build the logic for what is “extra important” in a calculated field.

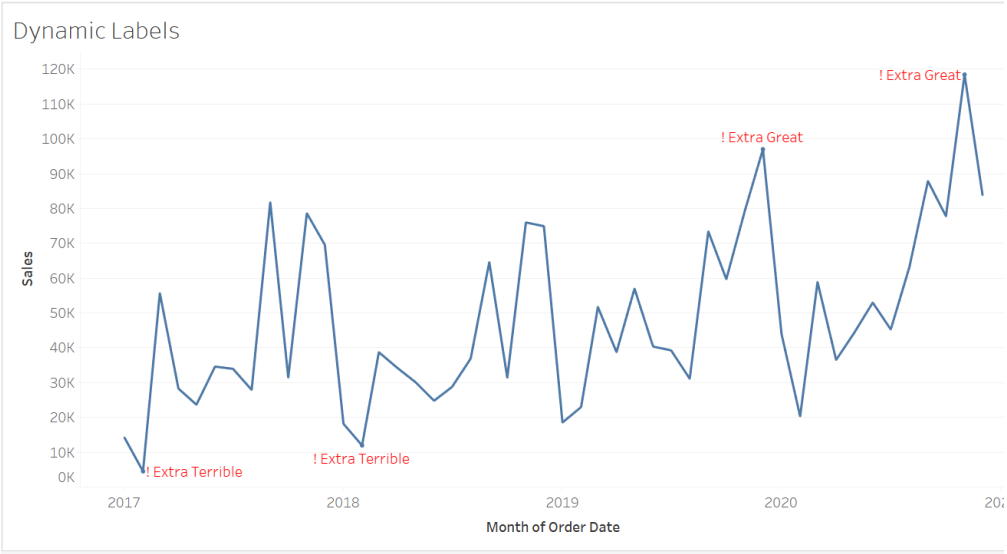
2. Create a Sales over time visualization with Months of the Order Date on the Columns shelf and Sum of Sales on the Rows shelf.



3. Create a calculated field called Dynamic Label Alert which has this logic:



4. Place the calculated field on the Label Marks Card. If you want to change the color of the Alert, edit the Label Font properties.

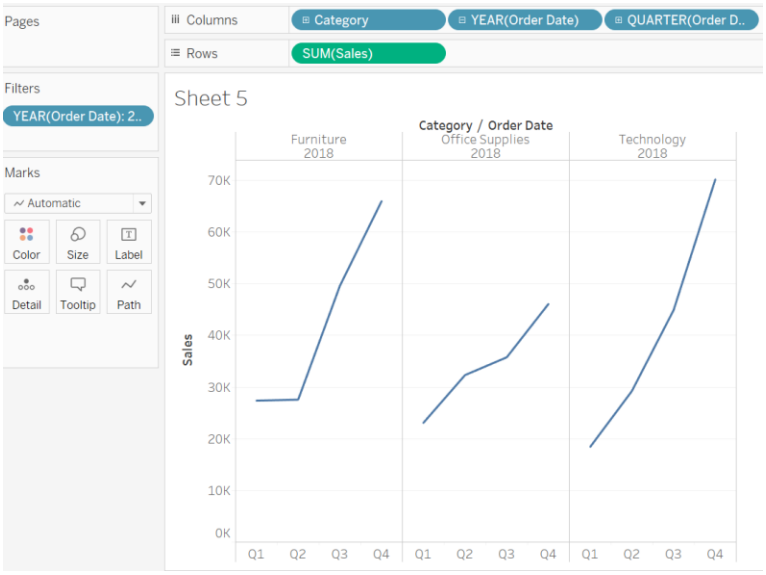


Introduction to Table Calculations

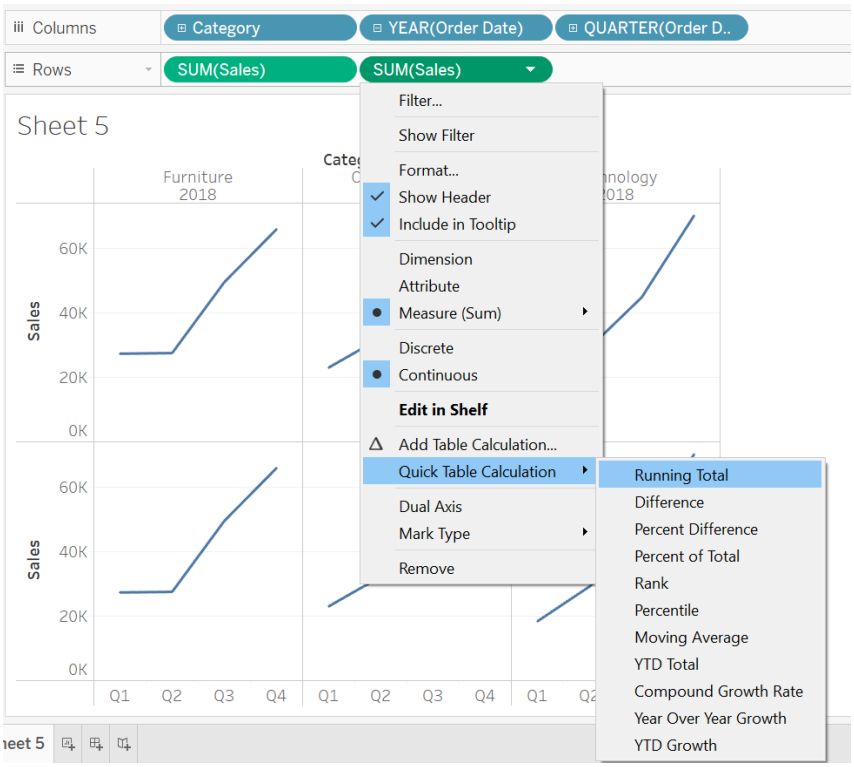
A Table Calculation is a secondary calculation that is performed on top of a result set. Tableau includes a set of pre-defined, commonly used computations called Quick Table Calculations. These include options like Running Total, Percent of Total, and Year over Year Growth.

Start with a New Sheet. Add Category to Columns and Sales to Rows. Order Date also to Columns (right of category), drill in to Quarters by clicking the plus sign. Drag order date to the filter pane and select the year 2021 only.

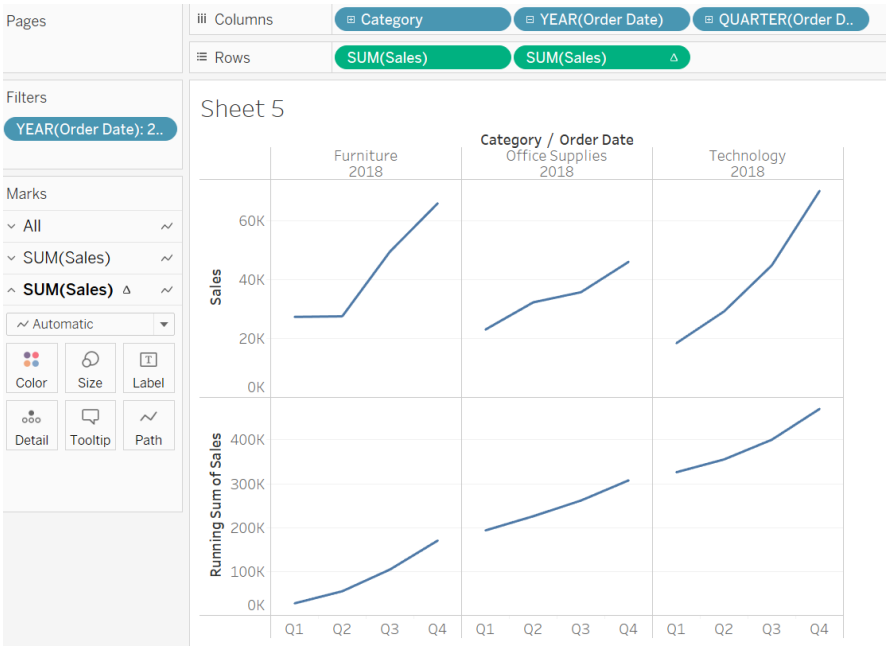
The visualization should look like this so far:



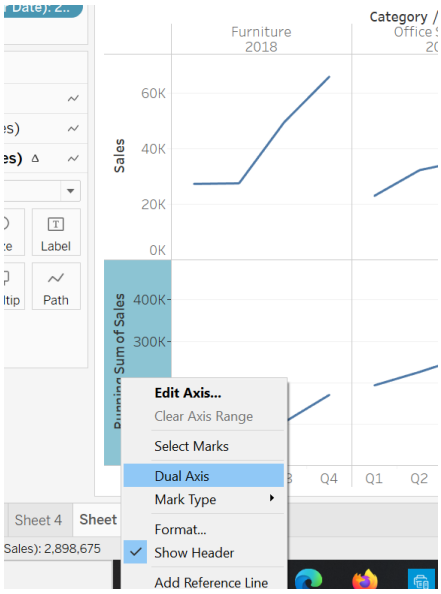
Now bring Sales to Rows again, right click the second pill and change to a Running Total Quick Table Calculation as shown below. You can tell that something is a table calculation by the delta symbol on the pill.



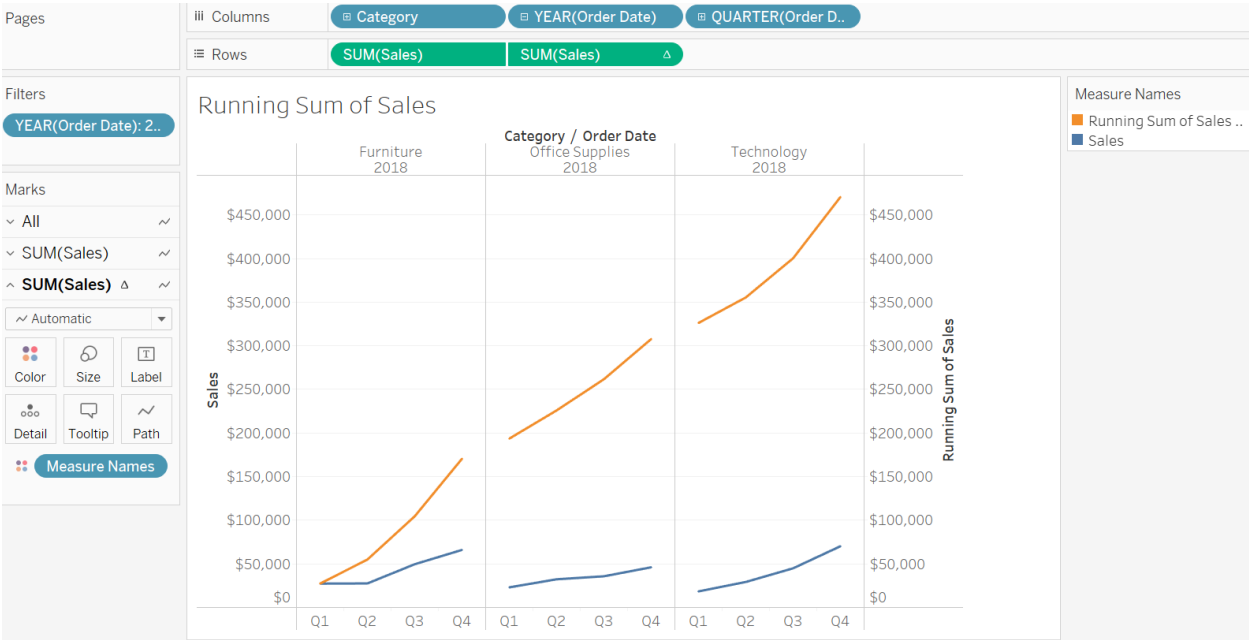
Now you have a set of graphs with quarterly sales as well as a running total.



If you want to view the graphs on top of each other (in one chart), you can right click on the Running total Table Calculation and select Dual Axis.



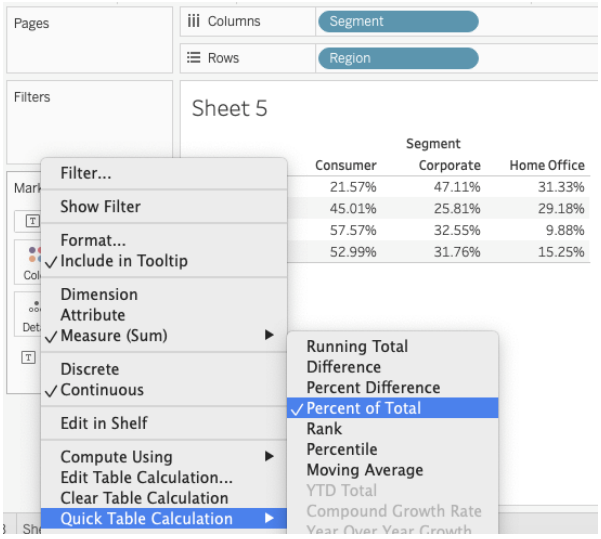
Right click an axis and select synchronize axis so that they each use common tick marks. This will produce the running total across the all the categories. If you want the running total across each pane, right click on the Sum([Sales]) quick table calculation pill and change compute using to “pane” rather than “table.”



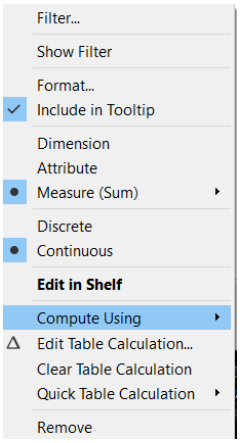
Once synchronized, if you want to remove the right-most axis, right click and unselect “Show Header.” If you remove it and change your mind and want to show it again, right click on the Sum([Sales]) quick table calculation pill, and select Show Header.

Other Table Calculation Options

Table calculations are not limited to the Quick Table calculations. Create a new sheet. Now, put Region in rows and Segment in columns and drag Profit out to get a text table. Now right click on the Profit pill and create a table calculation. As you click on different options you can see how the visualization changes. If you turn this into a visualization (e.g., a bar chart as opposed to a text table) then pay attention to the numbers as they will show you how the visualization calculations are done as you change Compute Using.

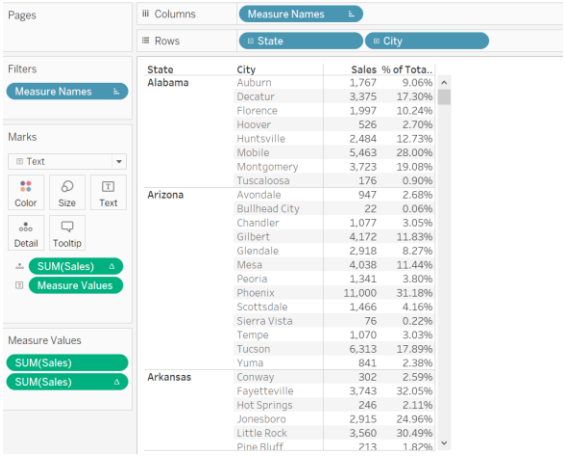


Note: Table Calculations are defined by how they are grouped and computed. In the above example, you could have the percent of the total profit by segment or by region. The default is to calculate the percent of total by rows. If you want percent of total by columns, select Compute Using and change from table(across) to table(down). There are other options too to check if your calculations are not as expected.

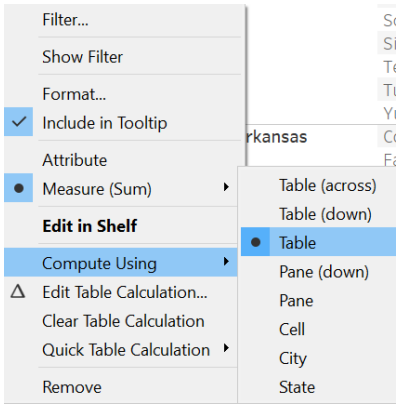


Another Quick Table Calculation Example

Using the Superstore Data, show Sales by State and City. Use the quick table calculation to show the sales percent that each City comprises in the State. To do this, change the compute using to City.

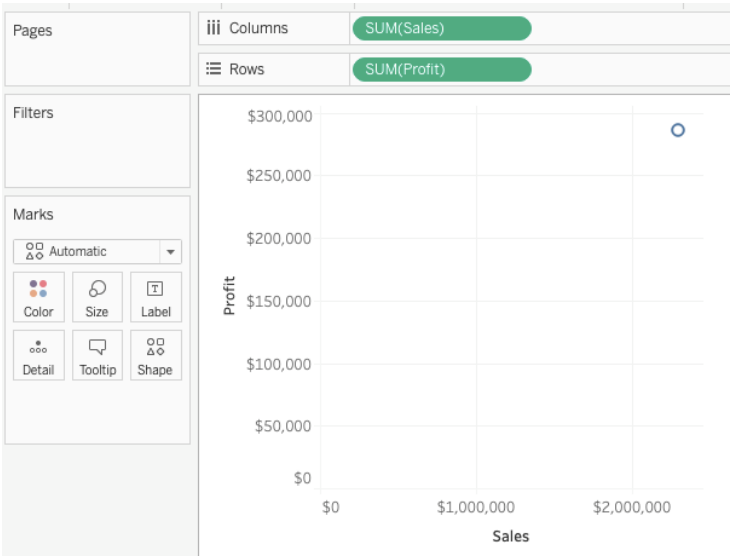


Compare this to the sales percent that each City comprises in the entire data set. To do this, change the compute using to Table.

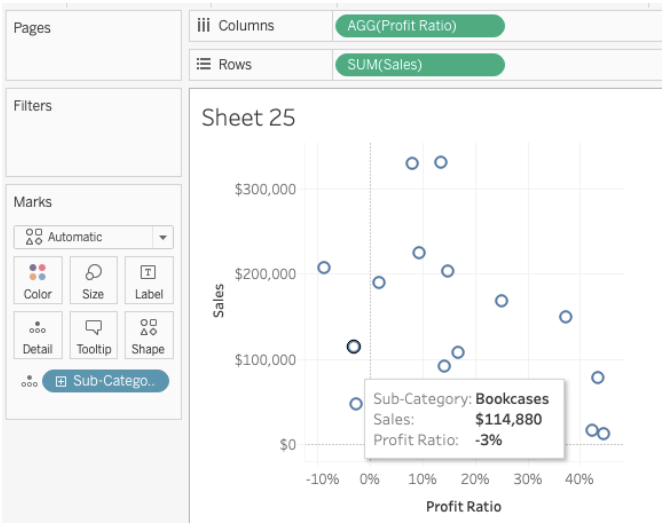


Topic 5: Level of Detail

This topic introduces Level of Detail calculations using Superstore data. In Tableau, you can change the level of detail or aggregation by bringing dimensions into the detail mark card. The example below simply shows profit ratio by sales:



When we bring a new dimension, such as sub-category, to the Detail Marks card, the level of detail for aggregation creates a much different visual.



The level of detail of the view determines the number of marks in your view. When you add a level of detail calculated field to the view, Tableau must reconcile two levels of detail—the one in the view, and the one in your expression.

The behavior of a level of detail calculated field in the view varies depending on whether the expression's level of detail is coarser, finer, or the same as the level of detail in the view.

Introduction to Level of Detail Calculations

Level of detail (LOD) calculations provide a way to compute aggregations that are not at the level of detail of the visualization. In particular, a level of detail calculation can compute an aggregation of a specified dimension. It's a good way to drill down to more detail or broaden your view with less detail. This will be a hard concept to get but very valuable once you do finally figure it out.

There are three different types of LOD calculations: FIXED, INCLUDE, and EXCLUDE, each of which alters the scope of the LOD calculation. FIXED LODs apply before any additional filters are applied. You can think of a FIXED LOD as something that doesn't change no matter what other pills you add to the view or what normal filters you apply. INCLUDE and EXCLUDE are calculated relative to dimensions in the view.

FIXED: fix the measure at a certain level of detail

INCLUDE: include dimensions that are not on the view

EXCLUDE: exclude dimensions that are on the view

There are cases where an INCLUDE/EXCLUDE and FIXED calculation may return a similar result depending on the dimensions in the view.

Usually the question you are trying to answer or the information you are trying to communicate will determine which LOD calculation to use. However, there are multiple ways to do the same things in some cases. In general, FIXED is used when you want something to stay put no matter what you do in the visualization.

Fixed Level of Detail Calculations

Examples for FIXED: Average Order Size by Region

Using the Superstore dataset, suppose you are analyzing the sales performance of each region and would like to know which region has the highest (or lowest) average order size? To look at average order size, we need to group by order id and then average within each order instead of every single record being averaged. To do this, we must create a calculated field that aggregates sales considering the order id.

Again, you need to calculate the size of each order (sum the sales corresponding to each Order ID), and then average those values by region.

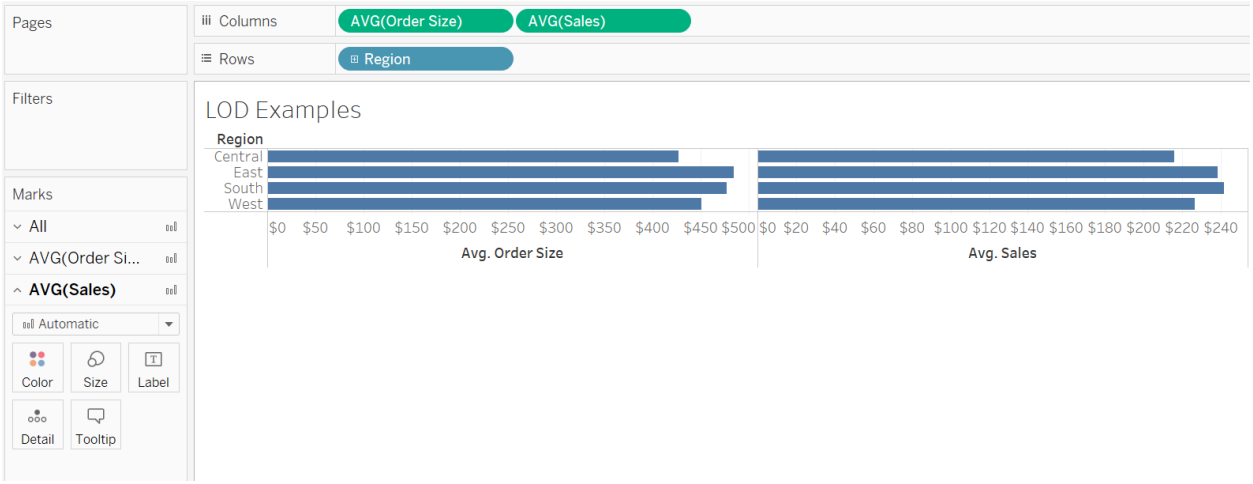
LOD calculations are written in the calculation editor.

Create a calculated field named Order Size:

Order Size

```
{Fixed [Order ID]:sum([Sales])}
```

The first element is the scoping keyword. The second element specifies dimensions references by the scoping keyword when evaluating the aggregate expression. The final part is the calculation performed. The above expression sums the purchases in each Order ID. You can use the Order Size to determine which region has the highest or lowest average order size and compare it to average sales by region. Drag average order size and average sales to the columns shelf and region to the rows shelf. Now select a horizontal bar chart. To format the average order size axis, right click it and select format. Chose the AVG(Order Size) option from the fields dropdown and make sure to select the axis tab instead of pane. Change the numbers format to currency with zero decimal places.



The average sales is the average of all line items (rows) in each Region. The LOD calculation first determines the size of each order (sum of sales for all line items within that order) and then averages the resulting orders, and it makes sense that the average order size would be larger than the average sales of each line item.

INCLUDES/EXCLUDES will be discussed next, but in this case, you can get the same result if you used INCLUDES instead of FIXED, i.e., sum the sales including the Order ID:

Order Size

```
{INCLUDE [Order ID]:SUM ([Sales])}
```

Another FIXED Example: Customer Order Frequency

What if you wanted to know the number of customers who made one order, two orders, three orders, and so forth? To build this view, you must break up the number of customers by the number of orders made. In this data set, we need to use Customer Name and assume this is a

unique ID. A simple LOD expression can turn the number of orders into a dimension that breaks out the number of customers, i.e., we want to count the distinct number of orders by each customer.

Orders per Customer

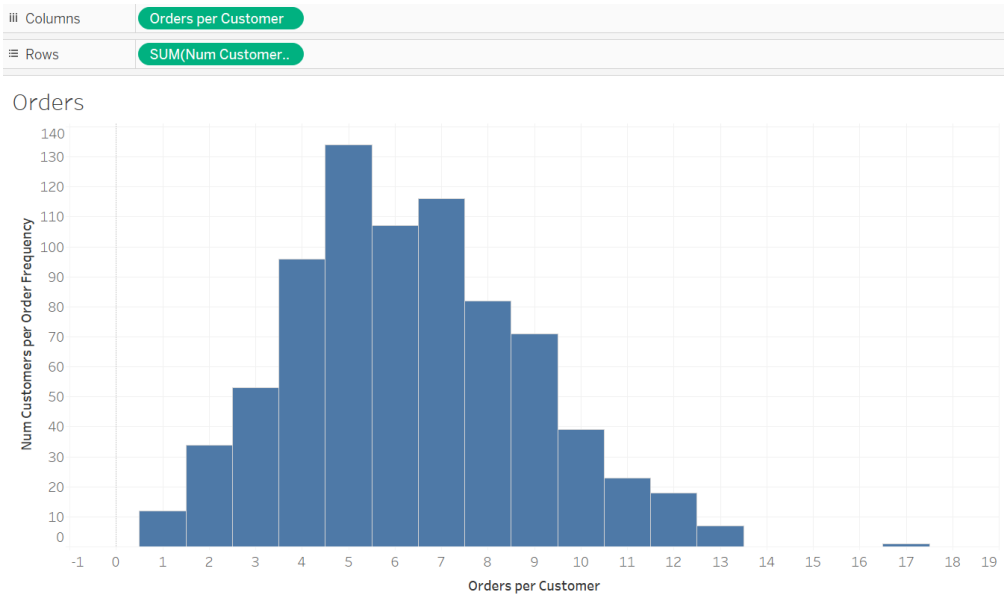
```
{FIXED [Customer Name]:COUNTD([Order ID])}
```

To create your own histogram of the number of customers who made one order, two orders, etc., you will also need a count of number of customers by orders per customer. Call the calculated field ‘’Num Customers per Order Frequency”:

per Order Frequency

```
{ FIXED [Orders per Customer]:COUNTD([Customer Name])}
```

Then create a bar chart. If your bars are too narrow to look like a proper histogram, then on the size mark card, change the option to ‘Fixed’. You may also need to change alignment to ‘Center.’ (Note: If you use the instructions in Topic 7 to make a histogram of orders per customer, you do not need to calculate the second variable because it will do it for you.)

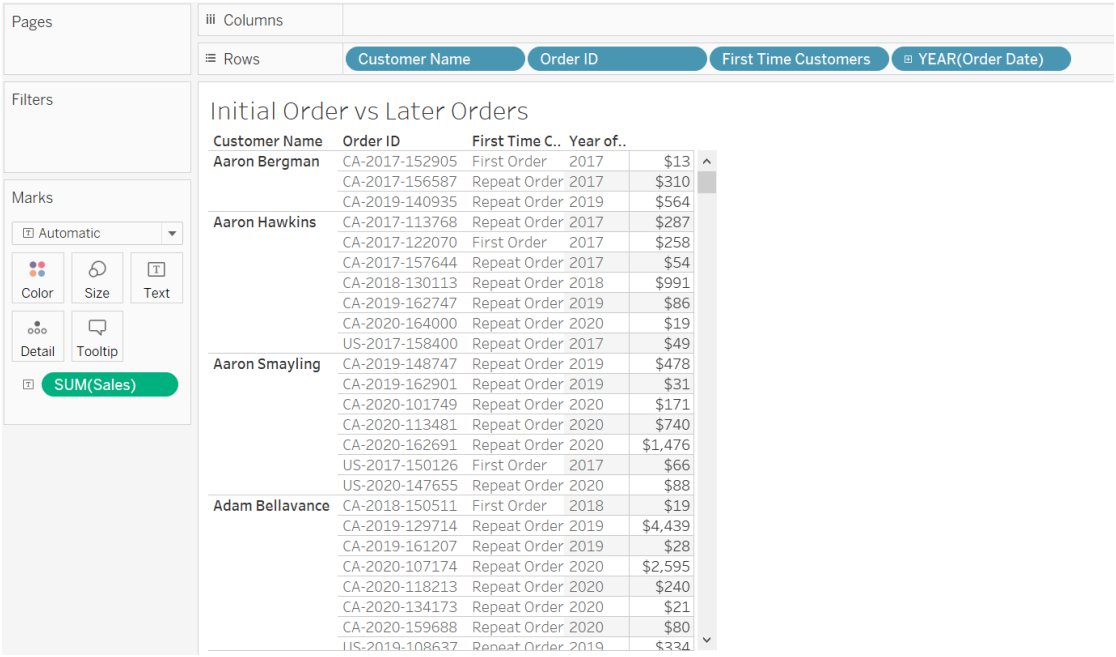


Another FIXED Example: Identify Each Customer’s Initial Order

Do new customers purchase more or less in their first order versus later orders? To determine this, you would need to identify the first order of a customer which would occur when the current order date is the minimum order date for a customer.

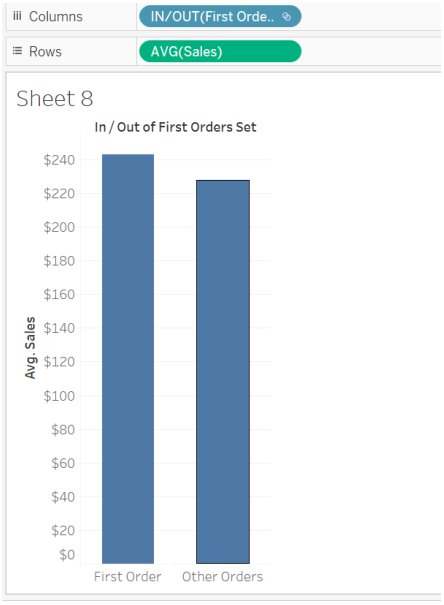
First Time Customer

```
If [Order Date] = {Fixed [Customer Name] : Min([Order Date])}
THEN "First Order"
ELSE "Repeat Order"
END
```



Customer Name	Order ID	First Time C..	Year of..	
Aaron Bergman	CA-2017-152905	First Order	2017	$13
	CA-2017-156587	Repeat Order	2017	$310
	CA-2019-140935	Repeat Order	2019	$564
Aaron Hawkins	CA-2017-113768	Repeat Order	2017	$287
	CA-2017-122070	First Order	2017	$258
	CA-2017-157644	Repeat Order	2017	$54
	CA-2018-130113	Repeat Order	2018	$991
	CA-2019-162747	Repeat Order	2019	$86
	CA-2020-164000	Repeat Order	2020	$19
	US-2017-158400	Repeat Order	2017	$49
Aaron Smayling	CA-2019-148747	Repeat Order	2019	$478
	CA-2019-162901	Repeat Order	2019	$31
	CA-2020-101749	Repeat Order	2020	$171
	CA-2020-113481	Repeat Order	2020	$740
	CA-2020-162691	Repeat Order	2020	$1,476
	US-2017-150126	First Order	2017	$66
	US-2020-147655	Repeat Order	2020	$88
Adam Bellavance	CA-2018-150511	First Order	2018	$19
	CA-2019-129714	Repeat Order	2019	$4,439
	CA-2019-161207	Repeat Order	2019	$28
	CA-2020-107174	Repeat Order	2020	$2,595
	CA-2020-118213	Repeat Order	2020	$240
	CA-2020-134173	Repeat Order	2020	$21
	CA-2020-159688	Repeat Order	2020	$80
	US-2019-108637	Repeat Order	2019	$334

If you then create a set of First Orders vs. Later Orders and check the average of sales, customer do on average order more in their first order.



INCLUDES vs EXCLUDES

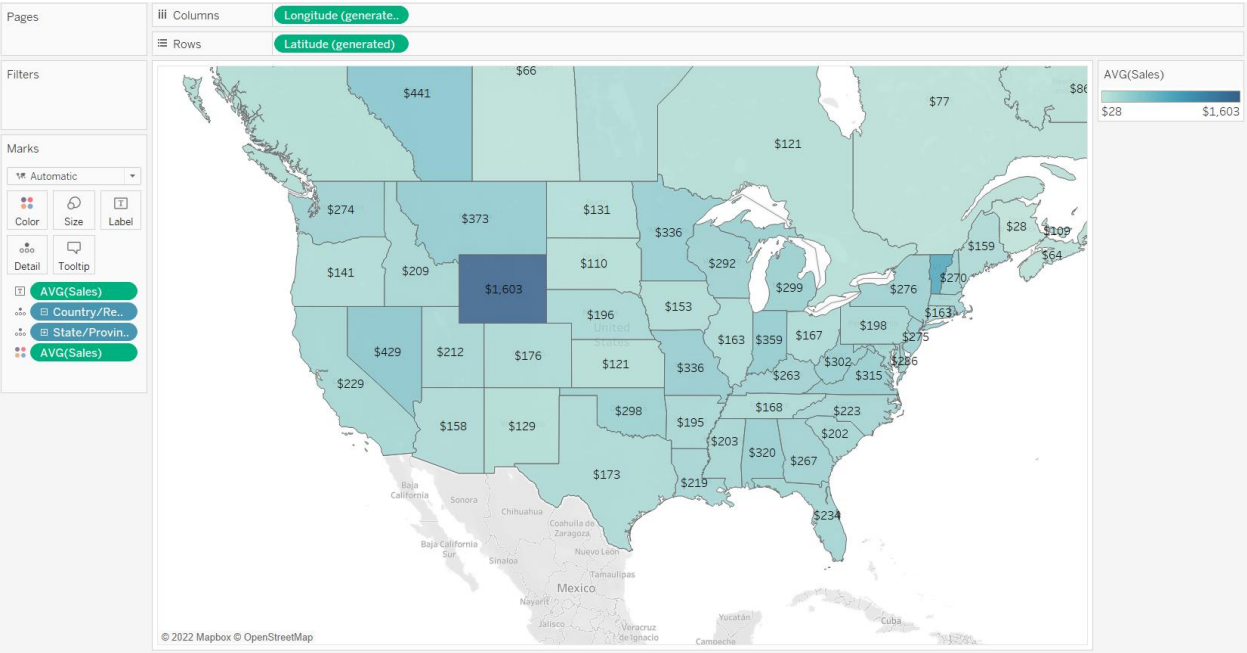
Again, INCLUDE and EXCLUDE are calculated relative to dimensions in the view, and sometimes one or the other will behave the same as FIXED and sometimes they won't. In general, FIXED should be used when you are doing something independent of your view, e.g., the average order size per customer. You would use INCLUDES or EXCLUDES when you are doing something specific to your view in your visualization.

The INCLUDE level of detail expression calculates the specified Dimensions in your syntax in addition to the dimensions already in your view. In other words, the INCLUDE expression is for calculating a higher level of detail without changing your view.

The EXCLUDE function allows you to calculate a coarser granularity than the dimensions in your view. To calculate your desired expression, Tableau will first remove the specified excluded dimensions from the visualization LOD and then perform the calculation as though the dimension is not present. All dimensions in the view will be aggregated except the excluded dimension.

Include Example: Average Sales by State and by Customer

First create a visualization of Average Sales by State.

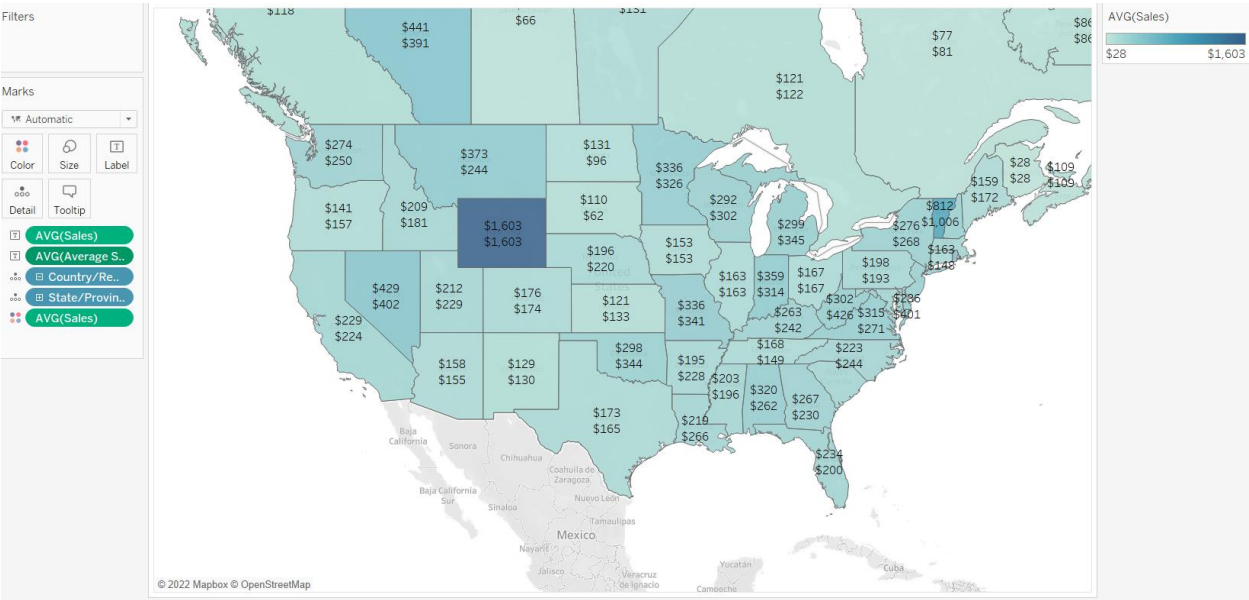


What if we also wanted to understand the sales averaged by customers in a state? For this to happen we must create a new field that shows average sales including the Customer Name which is not currently included in this level of visualization.

You will need to include Customer Name for the calculation of Average Sales. This means that Total Sales will still be influenced by state but will also consider Customers in that state.

Avg. Sales by Customer

```
{INCLUDE [Customer Name]: AVG([Sales])}
```



Note: if you are checking the Average Calculations make sure you are averaging at the item level/product name for the Average Sales values. Average Sales by Customer is not the same as taking the Sum of Sales and dividing by the number of customers. The value in the above visualization is the average of the average of Sales each customer in a state.

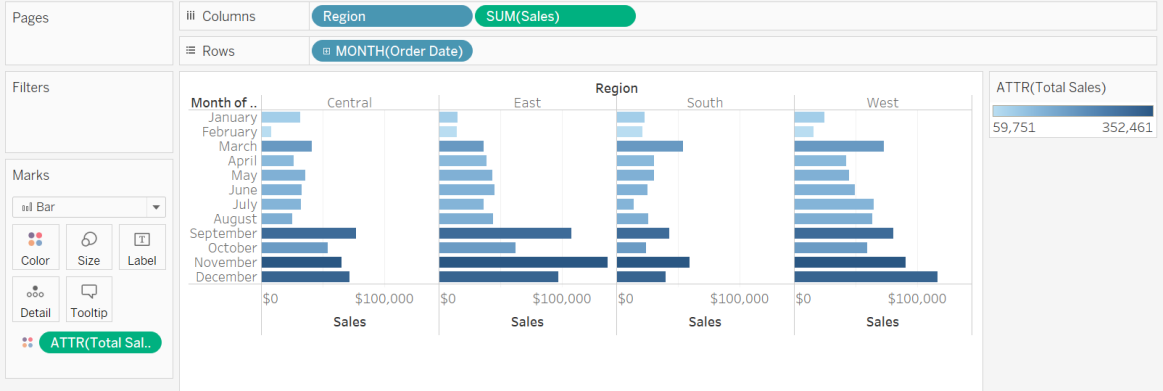
Exclude Example: Total Monthly Sales independent of region

What if the question is what are the total monthly sales as well as the total sales by Region? For this to happen we must create a new field that shows total sales while excluding the grouping by region.

You will need to exclude Region from the calculation of monthly Total Sales. This means that the aggregation of Total Sales will not be influenced by a region grouping. In the chart below, notice how Total Sales is used only on the color mark card. Each month's bar is the same color across regions because it is excluding region, unlike the sales measure which includes it.

Total Sales

```
{EXCLUDE [Region]:SUM ([Sales])}
```



The color now represents the average monthly Sales and is computed using the EXCLUDE LOD calculation. The sum of Sales by Region (the length of the bar) is not based on an LOD expression.

Level of Detail Calculations Comparison

Determining when to use a particular LOD calculation can be tricky. Especially because there are cases where an INCLUDE/EXCLUDE and FIXED calculation may return a similar result depending on the dimensions in the view.

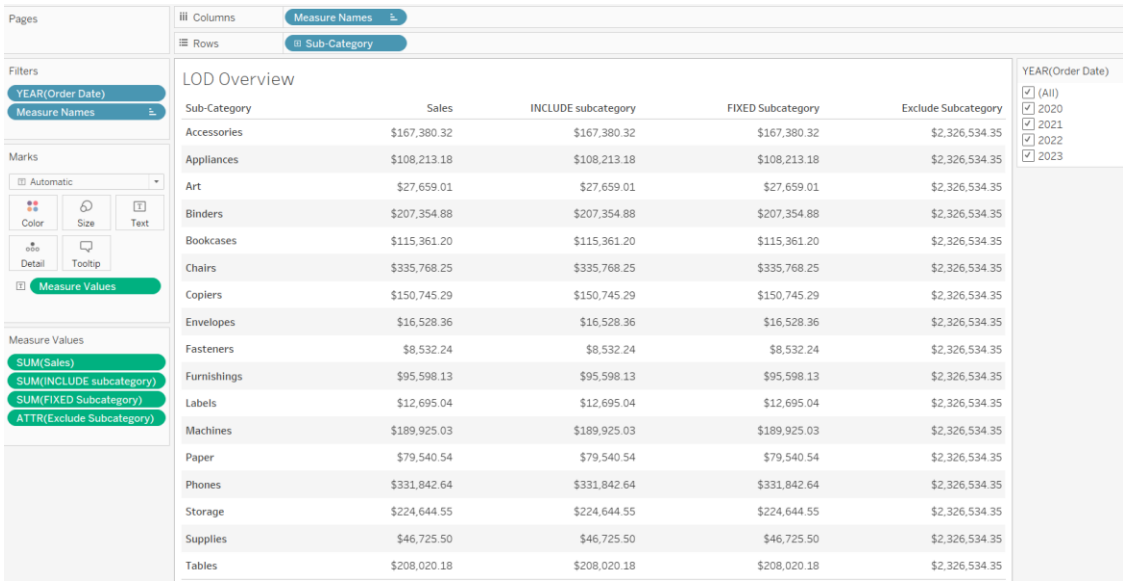
Usually the question to be answered, or the information to be communicated, will determine which LOD calculation to use. However, in some cases there are multiple ways to do the same thing, which can make LOD calculations confusing – especially FIXED versus INCLUDE. In general, use FIXED when something should remain the same no matter how the visualization is adjusted, and use INCLUDE when a calculation should adjust along with the visualization.

Comparative LOD Example

To explore the difference between types of LOD calculations, begin by creating a text table of **Sales** by **Sub-category**. Add a filter on **Order Date** aggregated by **Year** and select ALL. Make sure Order Date is discrete. Then create three new LOD calculations and add them as columns to the text table:

> Fixed Sub-Category Sales: `{FIXED [Sub-Category]: SUM([Sales])}`
>
> Exclude Sub-Category Sales: `{EXCLUDE [Sub-Category]: SUM([Sales])}`
>
> Include Customer ID Sales: `{INCLUDE [Customer ID]: SUM([Sales])}`

The view should look like this:

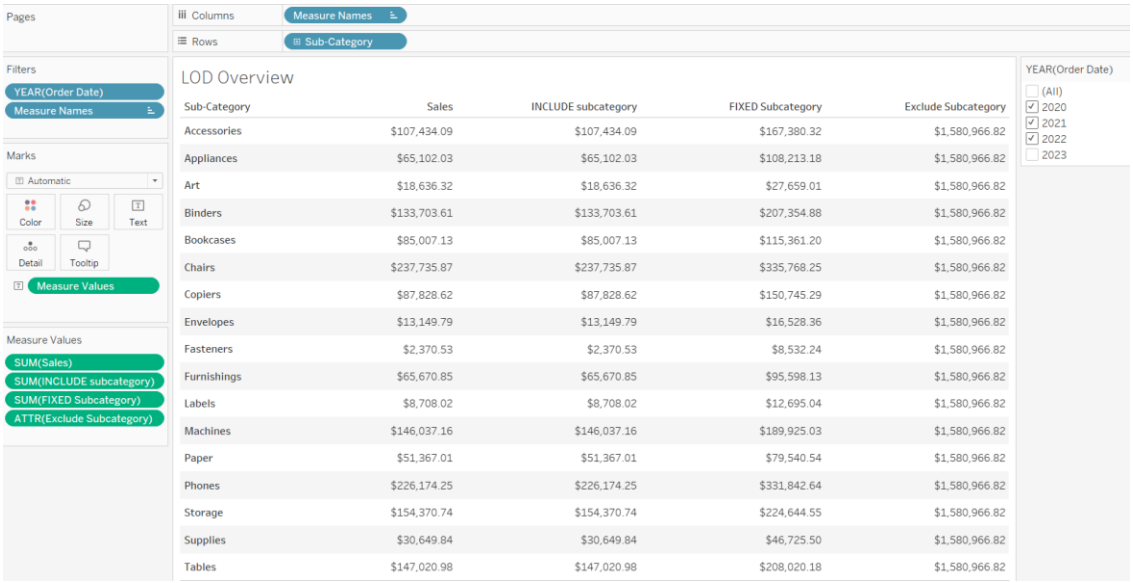


Sub-Category	Sales	INCLUDE subcategory	FIXED Subcategory	Exclude Subcategory
Accessories	$167,380.32	$167,380.32	$167,380.32	$2,326,534.35
Appliances	$108,213.18	$108,213.18	$108,213.18	$2,326,534.35
Art	$27,659.01	$27,659.01	$27,659.01	$2,326,534.35
Binders	$207,354.88	$207,354.88	$207,354.88	$2,326,534.35
Bookcases	$115,361.20	$115,361.20	$115,361.20	$2,326,534.35
Chairs	$335,768.25	$335,768.25	$335,768.25	$2,326,534.35
Copiers	$150,745.29	$150,745.29	$150,745.29	$2,326,534.35
Envelopes	$16,528.36	$16,528.36	$16,528.36	$2,326,534.35
Fasteners	$8,532.24	$8,532.24	$8,532.24	$2,326,534.35
Furnishings	$95,598.13	$95,598.13	$95,598.13	$2,326,534.35
Labels	$12,695.04	$12,695.04	$12,695.04	$2,326,534.35
Machines	$189,925.03	$189,925.03	$189,925.03	$2,326,534.35
Paper	$79,540.54	$79,540.54	$79,540.54	$2,326,534.35
Phones	$331,842.64	$331,842.64	$331,842.64	$2,326,534.35
Storage	$224,644.55	$224,644.55	$224,644.55	$2,326,534.35
Supplies	$46,725.50	$46,725.50	$46,725.50	$2,326,534.35
Tables	$208,020.18	$208,020.18	$208,020.18	$2,326,534.35

In this example, the FIXED LOD is at the level of Sub-Category, and that is the finest level of granularity in the visualization. Since Customer ID is not in the visualization, Sales are summed to the finest level of granularity (i.e., Sub-Category), and the FIXED and INCLUDE LOD expressions return the same result.

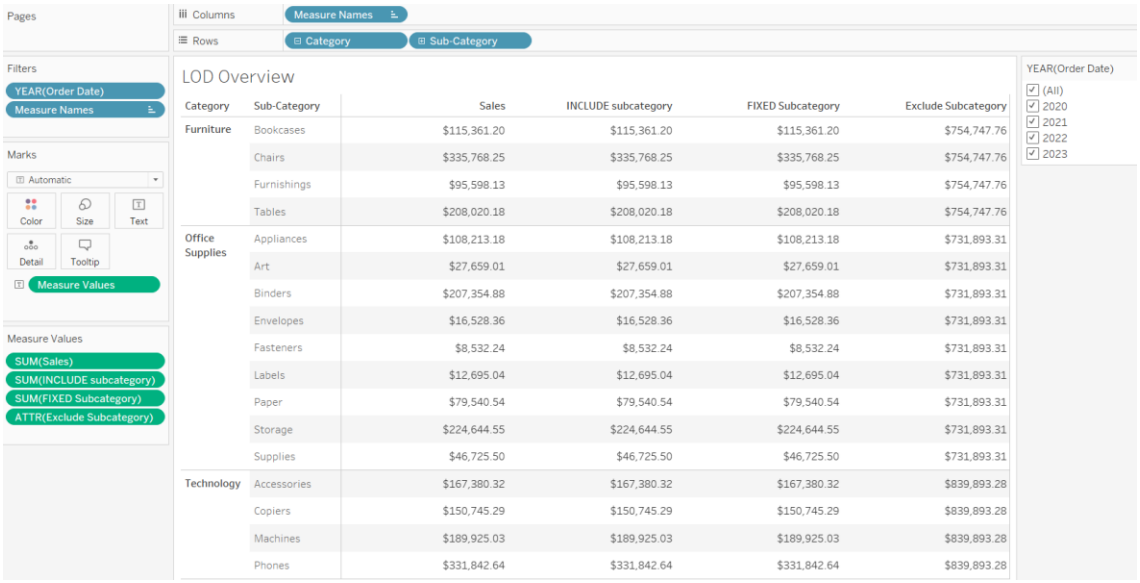
Note also that since the EXCLUDE calculation excludes the highest level of aggregation, Sub-category, it instead calculates the value for the entire data set (i.e., total sales). Insert a new sheet and bring Sales to the view to confirm this.

When filtered to show only 2020, 2021, and 2022, the INCLUDE, EXCLUDE, and normal Sales calculations change, but the FIXED calculation remains the same:



Sub-Category	Sales	INCLUDE subcategory	FIXED Subcategory	Exclude Subcategory
Accessories	$107,434.09	$107,434.09	$167,380.32	$1,580,966.82
Appliances	$65,102.03	$65,102.03	$108,213.18	$1,580,966.82
Art	$18,636.32	$18,636.32	$27,659.01	$1,580,966.82
Binders	$133,703.61	$133,703.61	$207,354.88	$1,580,966.82
Bookcases	$85,007.13	$85,007.13	$115,361.20	$1,580,966.82
Chairs	$237,735.87	$237,735.87	$335,768.25	$1,580,966.82
Copiers	$87,828.62	$87,828.62	$150,745.29	$1,580,966.82
Envelopes	$13,149.79	$13,149.79	$16,528.36	$1,580,966.82
Fasteners	$2,370.53	$2,370.53	$8,532.24	$1,580,966.82
Furnishings	$65,670.85	$65,670.85	$95,598.13	$1,580,966.82
Labels	$8,708.02	$8,708.02	$12,695.04	$1,580,966.82
Machines	$146,037.16	$146,037.16	$189,925.03	$1,580,966.82
Paper	$51,367.01	$51,367.01	$79,540.54	$1,580,966.82
Phones	$226,174.25	$226,174.25	$331,842.64	$1,580,966.82
Storage	$154,370.74	$154,370.74	$224,644.55	$1,580,966.82
Supplies	$30,649.84	$30,649.84	$46,725.50	$1,580,966.82
Tables	$147,020.98	$147,020.98	$208,020.18	$1,580,966.82

To explore the EXCLUDE LOD, add Category to the Rows shelf and note what happens to the EXCLUDE LOD. It now computes total sales for the Category since Sub-Category is excluded. Essentially, it rolled-up the computation.



Category	Sub-Category	Sales	INCLUDE subcategory	FIXED Subcategory	Exclude Subcategory
Furniture	Bookcases	$115,361.20	$115,361.20	$115,361.20	$754,747.76
	Chairs	$335,768.25	$335,768.25	$335,768.25	$754,747.76
	Furnishings	$95,598.13	$95,598.13	$95,598.13	$754,747.76
	Tables	$208,020.18	$208,020.18	$208,020.18	$754,747.76
Office Supplies	Appliances	$108,213.18	$108,213.18	$108,213.18	$731,893.31
	Art	$27,659.01	$27,659.01	$27,659.01	$731,893.31
	Binders	$207,354.88	$207,354.88	$207,354.88	$731,893.31
	Envelopes	$16,528.36	$16,528.36	$16,528.36	$731,893.31
	Fasteners	$8,532.24	$8,532.24	$8,532.24	$731,893.31
	Labels	$12,695.04	$12,695.04	$12,695.04	$731,893.31
	Paper	$79,540.54	$79,540.54	$79,540.54	$731,893.31
	Storage	$224,644.55	$224,644.55	$224,644.55	$731,893.31
	Supplies	$46,725.50	$46,725.50	$46,725.50	$731,893.31
Technology	Accessories	$167,380.32	$167,380.32	$167,380.32	$839,893.28
	Copiers	$150,745.29	$150,745.29	$150,745.29	$839,893.28
	Machines	$189,925.03	$189,925.03	$189,925.03	$839,893.28
	Phones	$331,842.64	$331,842.64	$331,842.64	$839,893.28

Referencing two dimensions

If you want to create a LOD expression with multiple dimensions, the syntax is:

{FIXED [Segment], [Category] : SUM([Sales])}

In conclusion, some people think it is helpful to remember:

Fixed: absolute dimensions (calculation regardless of dimensions in view)

Include/Exclude: relative dimensions (calculation relative to dimensions in view)

Learning More about LOD Calculations

LOD calculations can get intense and are very useful. Check out this blog for an introduction to some common uses: https://www.tableau.com/about/blog/LOD-expressions

Topic 6: Parameters and Interactive Filters

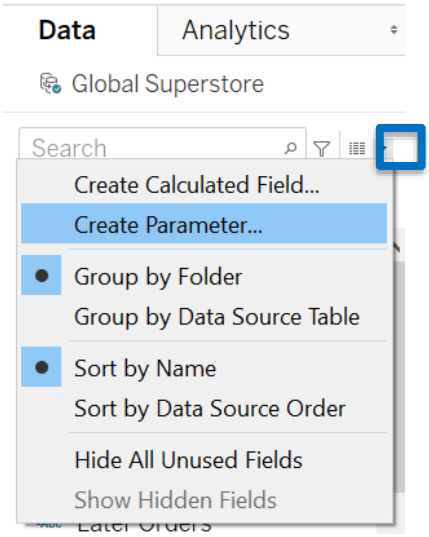
Introduction to Parameters

Parameters facilitate user engagement by providing users the ability interact with worksheets and dashboards. They serve as placeholders for fixed values to enable variable values that can change, thus impacting information included in visualizations.

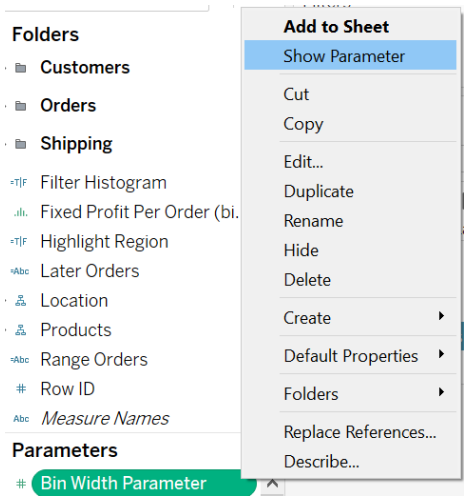
While filters use measures or dimensions to interact with worksheets and dashboards by restricting the data records included in a visualization, parameters function like variables (and are thus used in calculated fields) to interact with worksheets and dashboards. Because of this, parameters are more powerful than filters. Parameters are commonly used for such things as What if? analysis, user controlled thresholds, dynamically changing visualization source data, or filtering a variable number of data records.

In order to use a parameter, four things must be completed:

- The parameter must be created. To do this, click on the dropdown arrow in the top corner of the **Data** pane, and choose **Create Parameter**.



- The parameter must be used in a calculated field.
- The parameter control must be shown to the user. Click the dropdown arrow of the parameter pill and select **Show Parameter**.

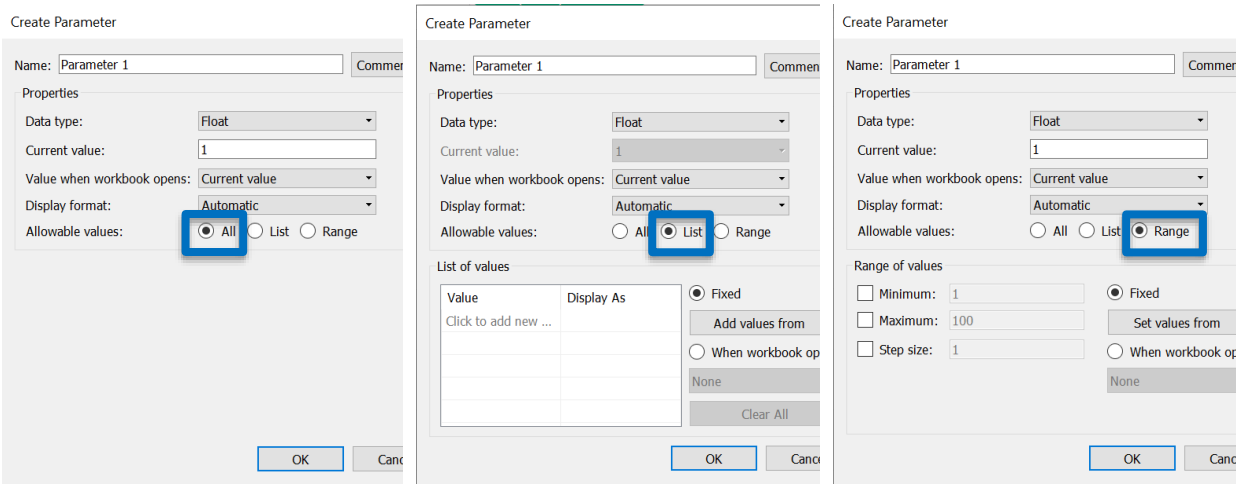


- The parameters must be connected to the visualization, usually by including the calculated field containing the parameter in the visualization.

When creating a parameter, a data type must be selected. The available choices are:

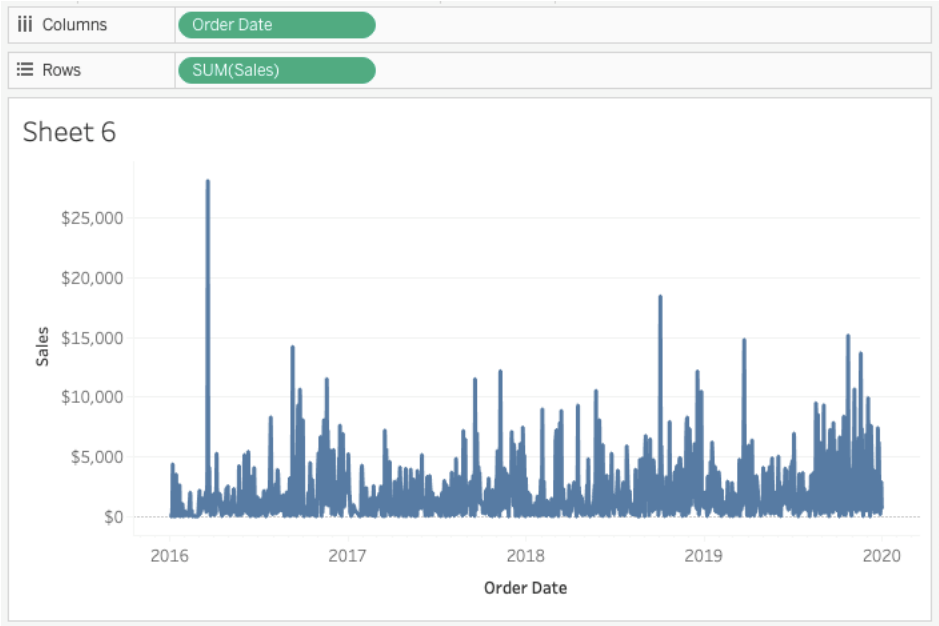
- Float: is a number with decimals
- Integer: a whole number without decimals
- String: text
- Boolean: True or False
- Date: this is the date without a timestamp
- Date & Time: this is the date with the timestamp

Data type, along with additional properties, are set in the Create Parameter dialog box. As the Allowable values option buttons are changed, the bottom of the dialog box updates:



User Controlled Threshold Example: Color After a Certain Date

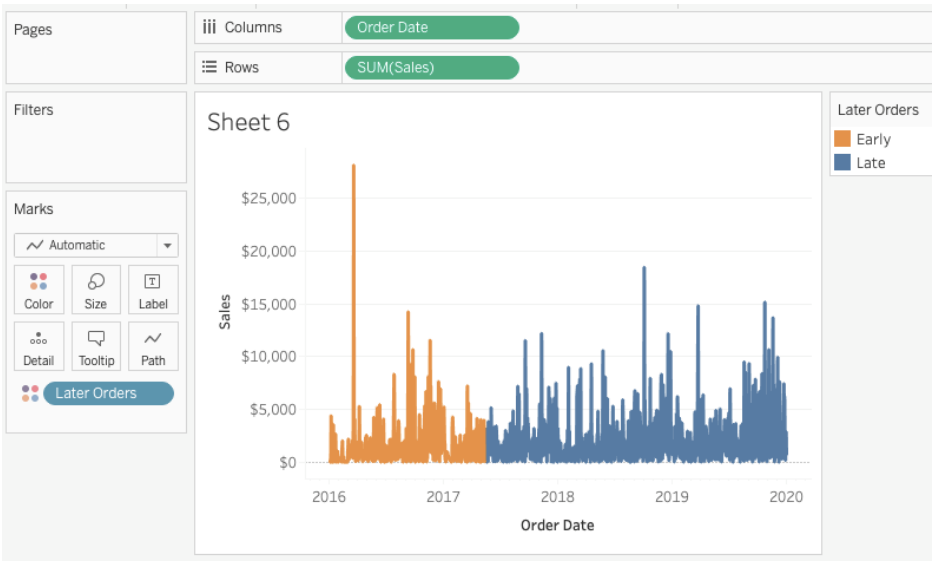
Create a line plot of SUM(Sales) vs. Order Date (continuous). Expand the **Order Date** to Day and select Exact Day by right clicking Order Date.



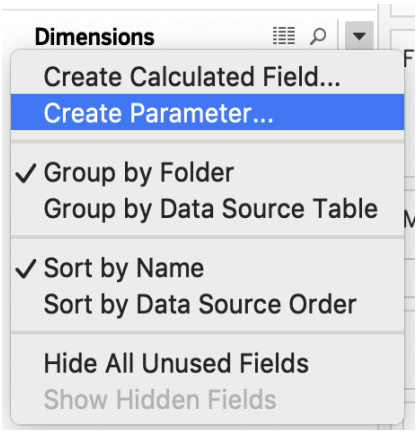
Now, let's create a calculated field Later Orders:

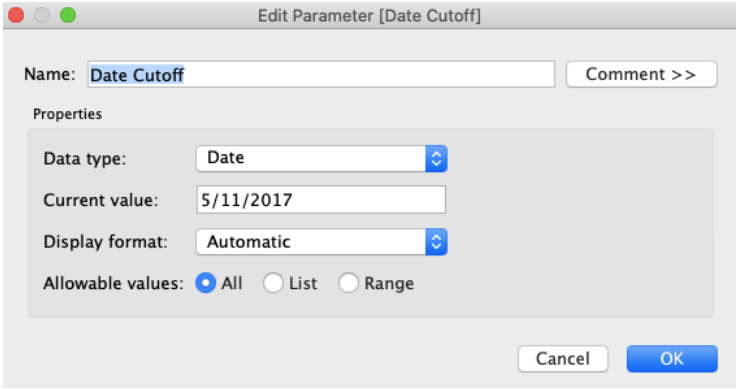
```
IF [Order Date] >= DATE("2017-05-11")
  THEN 'Late'
  ELSE 'Early'
END
```

Drag this new field to color. You see everything after (and including) May 11, 2017 as a different color now. What you want to do is let a user change this date as you go. You can do this with a parameter.

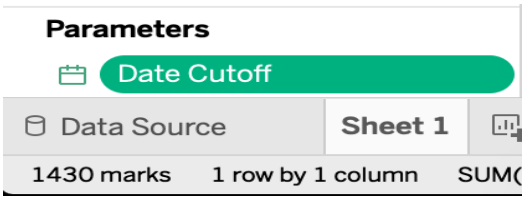


Now, create a parameter: click on the dropdown arrow icon near the magnifying lens in the Dimensions window and select "Create Parameter..." Call the parameter Date Cutoff and change the Data type to Date. Press OK.

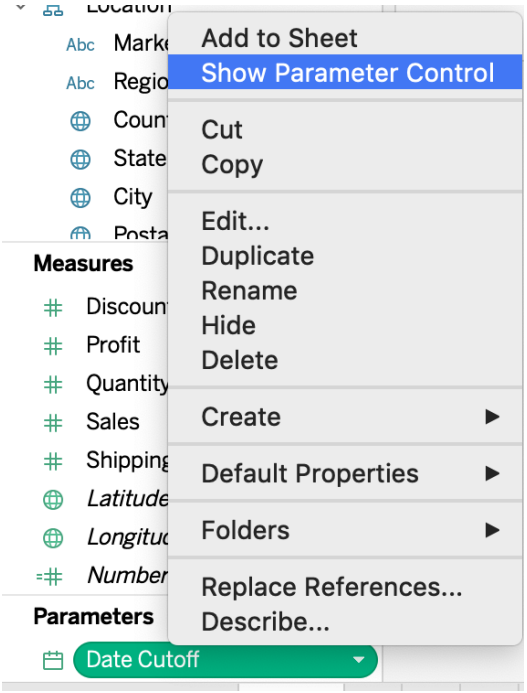




You now have a parameter, but nothing has changed. You can see the parameter in the Parameters section (below Measures):



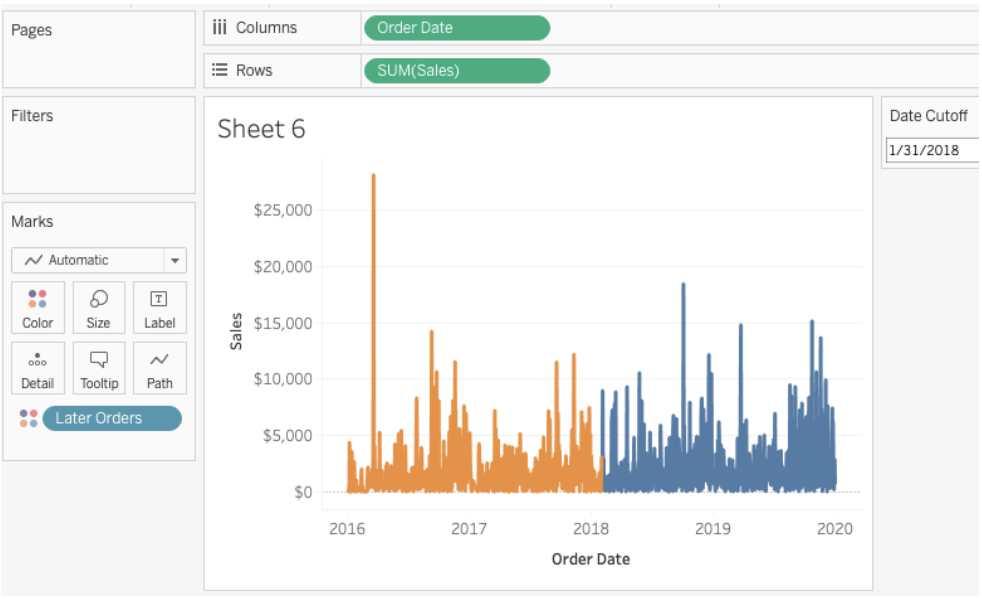
If you want your users to control the parameter value, you will need to show the parameter control. To do this, right click on the parameter, and select "Show Parameter Control" (or go to Analysis, Parameters, and select one to show).



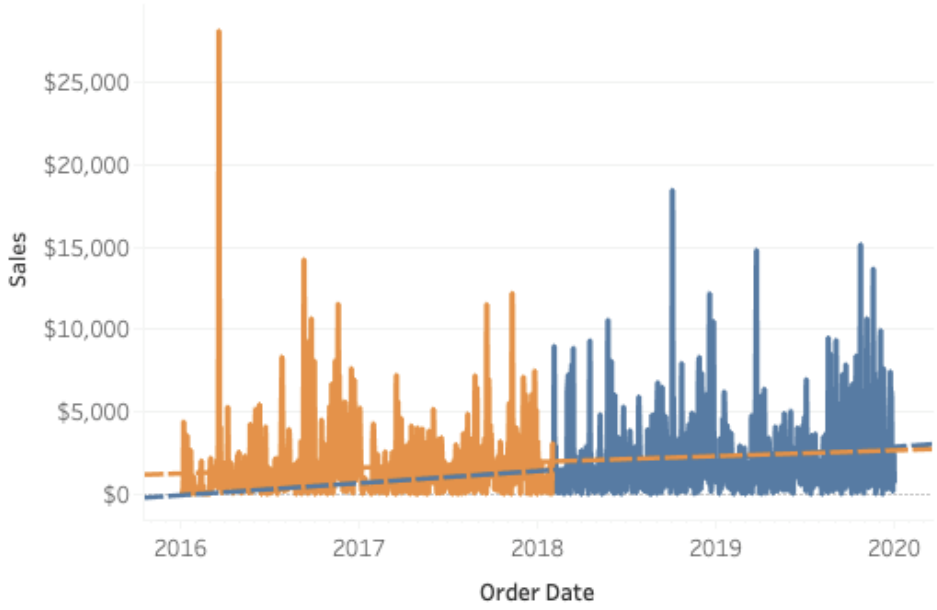
Now, let's edit our calculated field Later Orders:

```
IF [Order Date] >= [Date Cutoff]
  THEN 'Late'
  ELSE 'Early'
END
```

Change the cutoff to 1/31/2018 (or something within the time range) and you will see the visualization dynamically adjust.



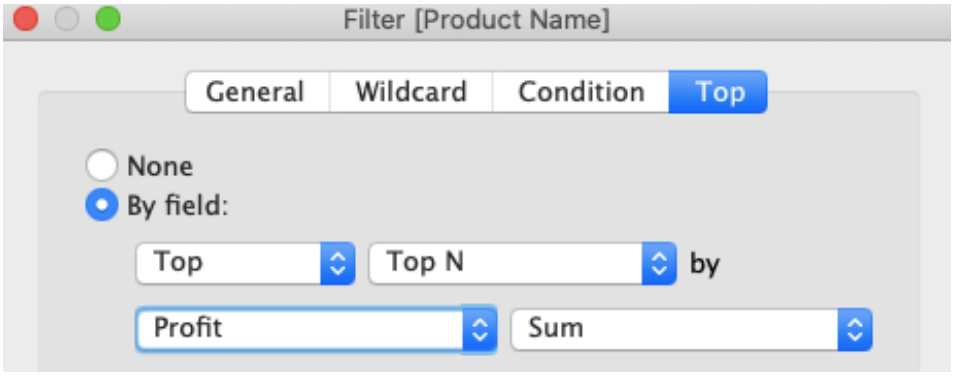
Now, go to Analytics and add a linear trend line. You should see two trend lines. Change the date again and see what happens to your visualization.



Parameter Filter to the Top N

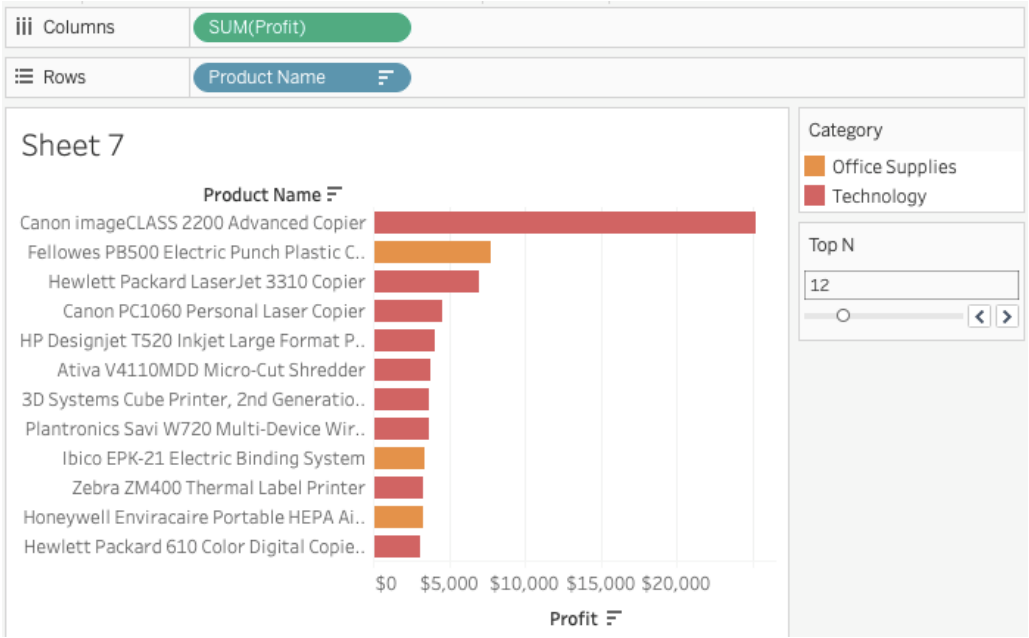
Create a new Sheet. Make a horizontal bar chart with **Product Name** (rows) and **Profit** (columns). Sort by **Profit**. Now bring **Category** to color. This is too long to really know what is going on. You may want to restrict the view to the top N items and let the user decide what N will be. Again, you need to use a parameter.

Create Parameter and name it “Top N” with data type Integer. For **Allowable** values select **Range** then Min = 1, Max = 50, and step = 1. Click Ok. Right click on the **Product Name** pill in Rows and select **Filter.** Click on Top. In By field, set the Top by Top N and make sure you having sum of Profit below.



Now you are set up but the parameter is defaulting to 1 so you only see one product. Right click on **"Top N"** and select **Show Parameter Control.** Now you use the slider to select the Top N products by profit. Magic.

If you don’t like the slider, you can change the slider to a Type In (click on the down triangle) and set the number to something higher.



More Advanced Topic: Parameter Filter to the Top N with a Context Filter or with the INDEX function

What if you want to show the Top N products in the Office Supply Category? If you don't do anything special you will get this visualization not the top 12 items in the Office Supply category.

Why?

Filters by default act as AND statements so all criteria have to be met, i.e., items in the Office Supply category AND the Top 12. The view shows the intersection of the results.

What if we use a Context Filter?

To find out what the top 20 furniture products are we need to make the Category filter a context filter. Right-click the field on the Filters shelf and select Add to Context. Now we can see the top 20 furniture, office supplies, or technology.

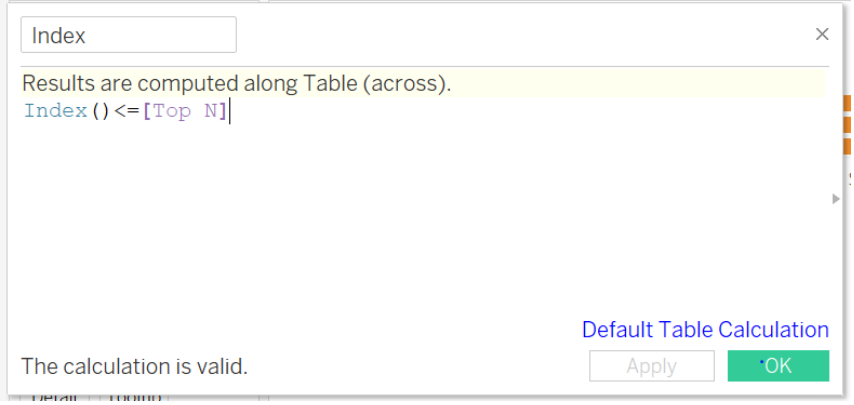
If we include more than one category in the filter, then we again get 20 items total (if N = 20).

How else can we do it?

If what we want is the Top N for each category, then we need to do something more complicated and use the INDEX() function.

The INDEX() function returns the index of the current row in the partition without any sorting with regard to value. The first-row index starts at 1. When INDEX() is computed, the index of each row is 1, 2, 3, 4,… etc.

Create a new calculated field called Index like this (Note: it shows the Table(across) comment when you reopen):

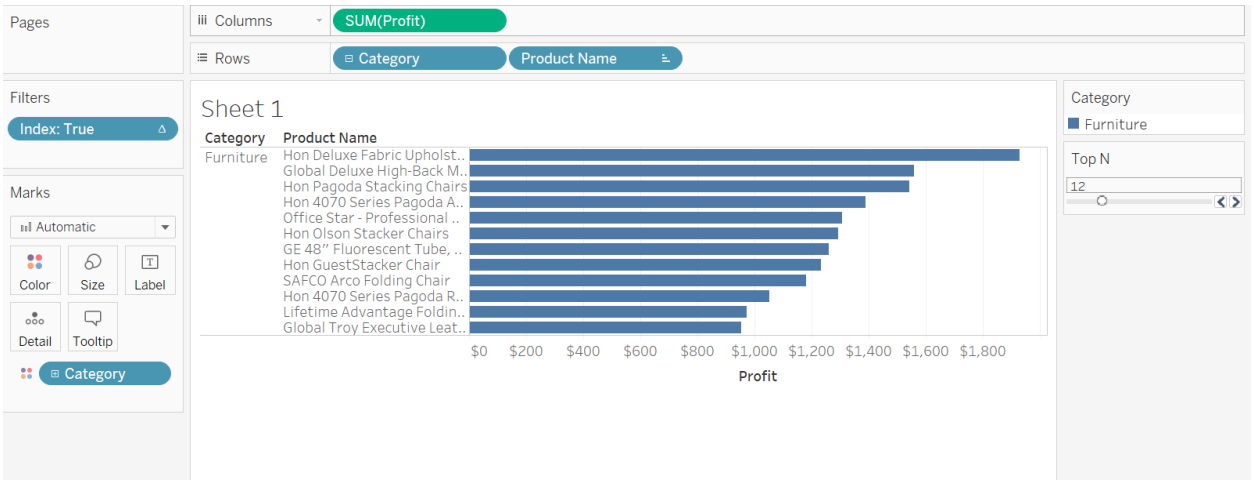


Remove the Product Name Filter and then drop the Index calculated field into the filter shelf and set to true.

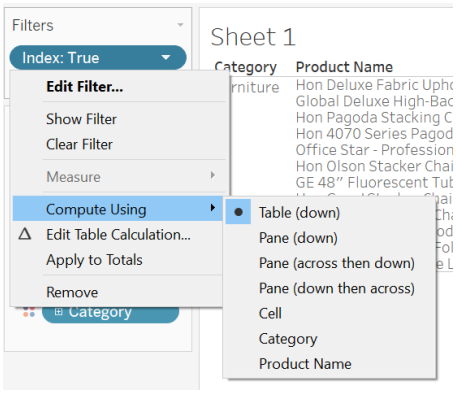
This will correct the visualization above to show all Top N fields for the chosen category.

But what if you would like to see the Top N products based on profit for each of the 3 categories in the Superstore?

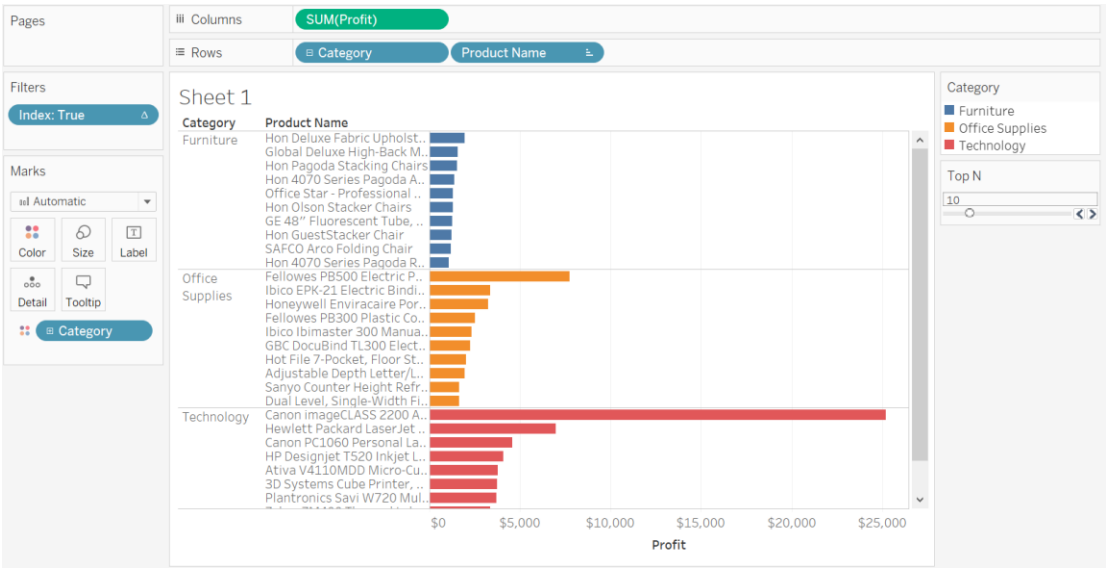
Remove Category as a Filter and include it in the Row shelf. But then you get the following picture which still isn't quite right. It is only showing the first 12 rows for the first category (Furniture).



When you are using Table Calculations, you may need to check that Tableau is correctly interpreting how you want the calculations since the default is to compute across the table. You can change the Table Calculation default with the “Compute Using” option (right-click on the Index filter).



Because we want to see the Top N product names for each category, we want to compute the Index filter using the Product Name. Select Product Name from the Compute Using option, and you will get the following visualization – the Top N products for each of our 3 categories.



Use a Parameter to Control Histograms

Create a new sheet.

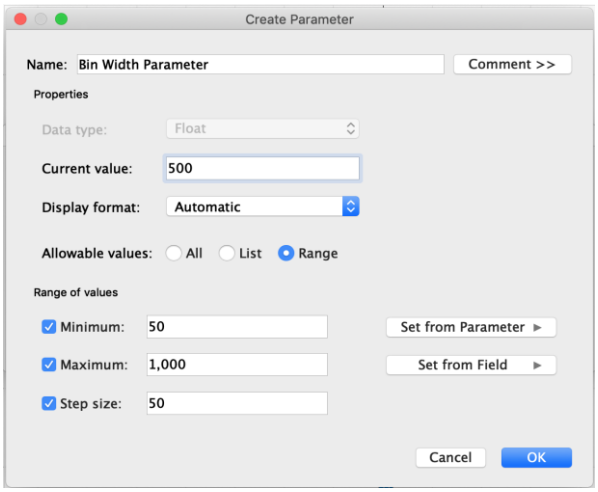
Create a new calculated field Fixed Profit Per Order (refer to the level of detail topic for more information on the fixed calculation):

```
{FIXED [Order ID]: SUM([Profit])}
```

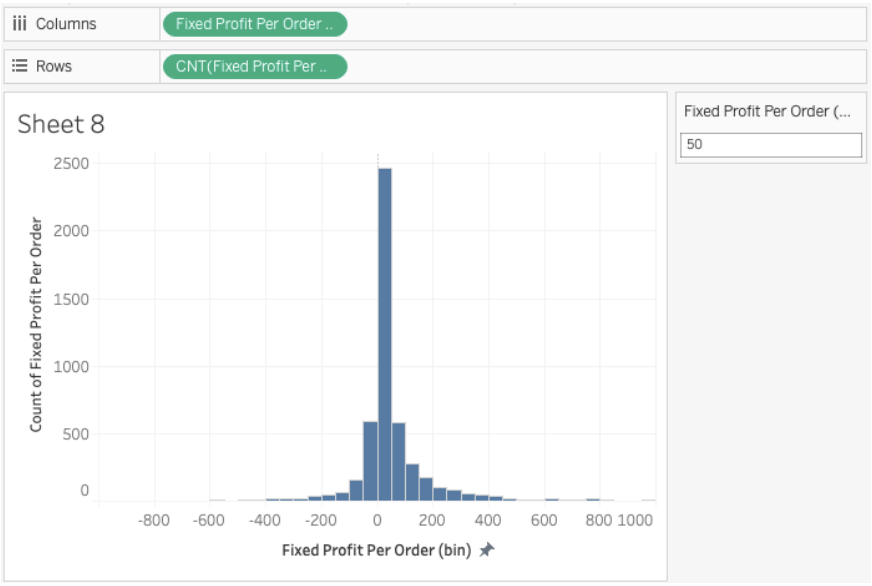
Select **Fixed Profit Per Order** and Show Me to select Histogram.

Edit the **Fixed Profit Per Order (bin)** dimension that Tableau created for you. Click on size of bins and select Create a New Parameter.

Name the new parameter Bin Width Parameter with a current value of 500, min of 50, max of 1000, and step size of 50. NOTE: If you create the parameter first then you need to go to Fixed order profit per order (bin) that it created for you.



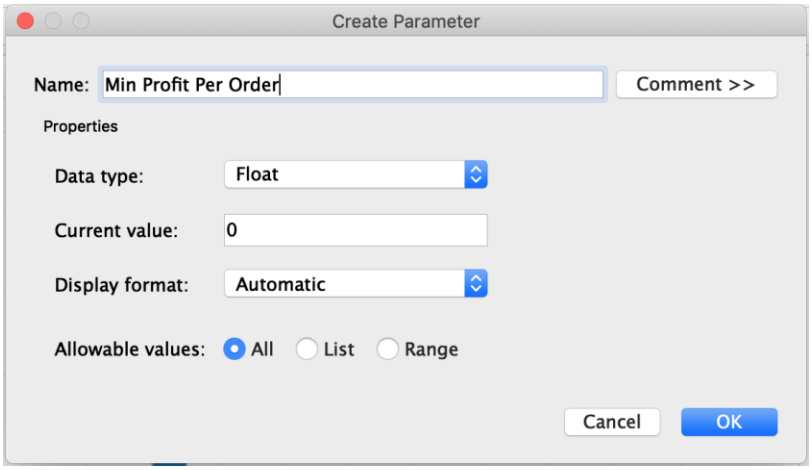
Now when you change the bin size on the right, you directly edit the visualization. You may need to adjust the min and max values on the axes to improve the appearance. Note: If nothing is showing, make sure the max value is greater than the min value.



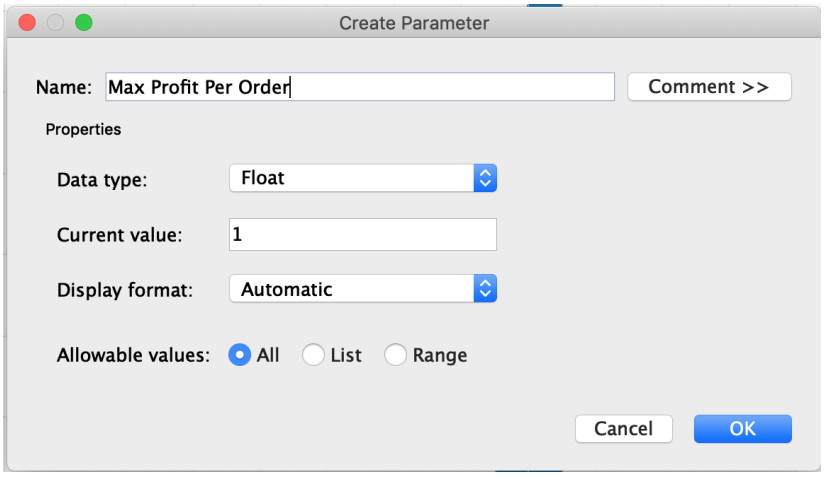
Use a Parameter to Filter Data

Look at your histogram. There are a lot of really large and really small values. You may want to zoom in only the middle of the histogram. You can do that by filtering out rows of data where Profit Per Order is too big or too small. Might as well do that with a parameter.

Create a parameter called Min Profit Per Order that is a Float with default 0. (Float is just the type of number when it can have decimals in comparison to integers that are only whole numbers.)



And another called Max Profit Per Order that is a float with default 1

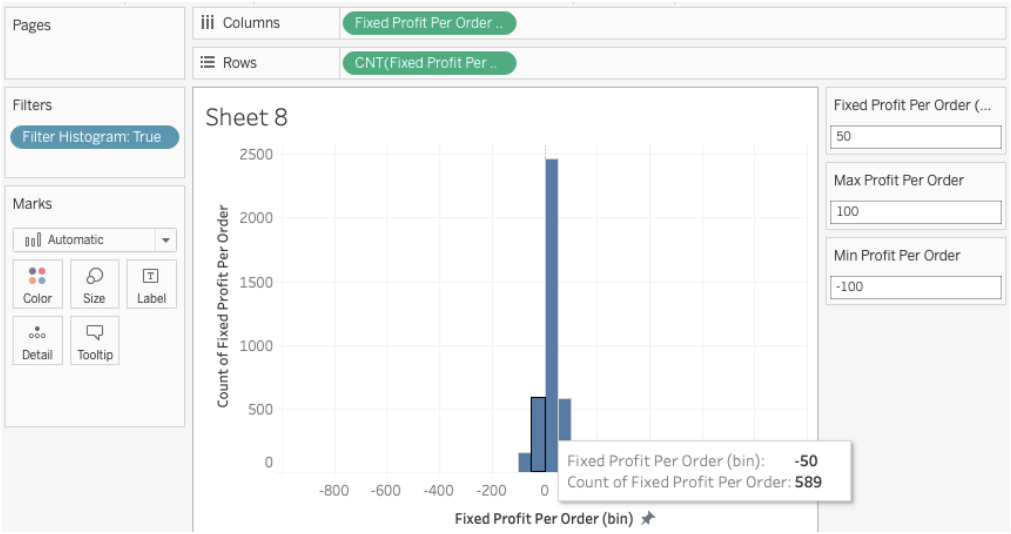


Show both of the parameters by right clicking on them and selecting Show Parameter

Then create a calculated field called Filter Histogram:

```
[Fixed Profit Per Order] > [Min Profit Per Order] AND [Fixed Profit Per Order]< [Max Profit Per Order]
```

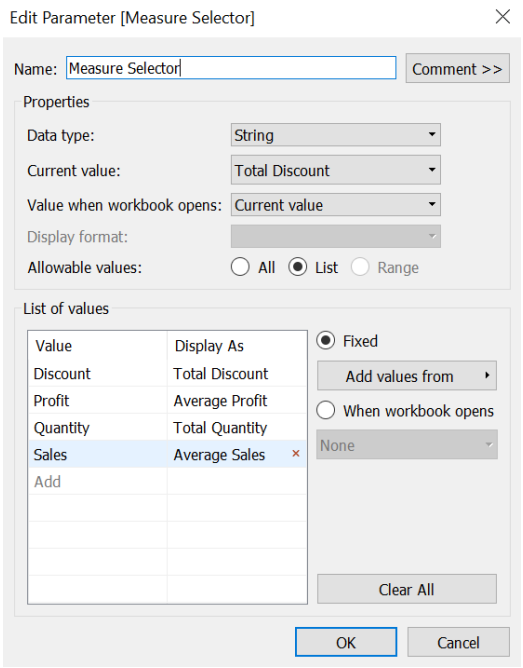
Bring Filter Histogram to filters and select True. Now change the parameters and watch everything change.



Swap Measures with Parameters

Create a new Sheet and create a new parameter called Measure Selector. In the Create Parameter dialog box, do the following:

1. Name the parameter Measure Selector.
2. Set the Data type to String
3. Skip down to the Allowable values field, and choose List.
4. Type individual measure names in the List of values area: Discount, Profit, Quantity, and Sales.
5. Change the display names to be Total Discount, Average Profit, Total Quantity, Average Sales.



Show this parameter.

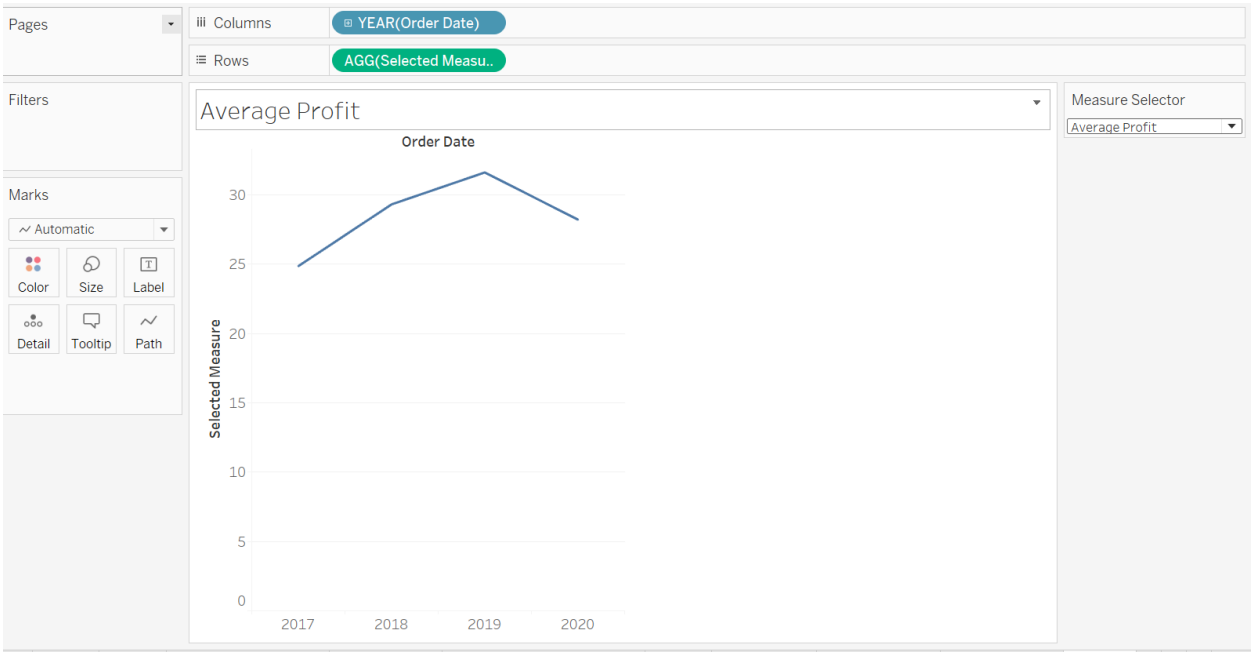
Now create a new calculated field called Selected Measure:

Selected Measure

```
Case [Measure Selector]
 WHEN "Discount" THEN Sum([Discount])
 WHEN "Profit" THEN AVG([Profit])
 WHEN "Quantity" THEN SUM([Quantity])
 WHEN "Sales" THEN AVG([Sales])
END
```

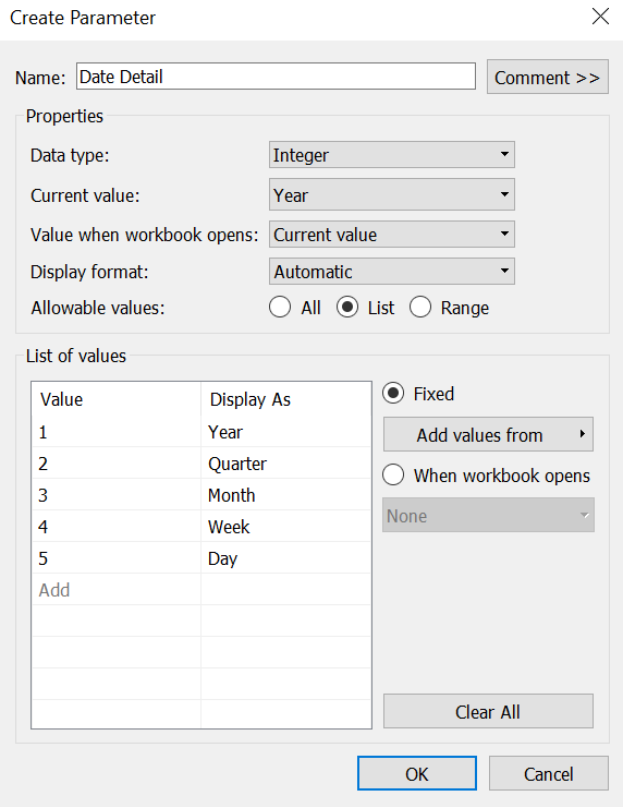
Now create a line chart with your new calculated field and **Order Date** (year - continuous). Change the measure selector and your visualization will change automatically.

Right click on the title for your plot and click Edit Title. Now you can change the title to say <Parameters.Measure Selector> which will change the title to match the measure selected by the parameter.

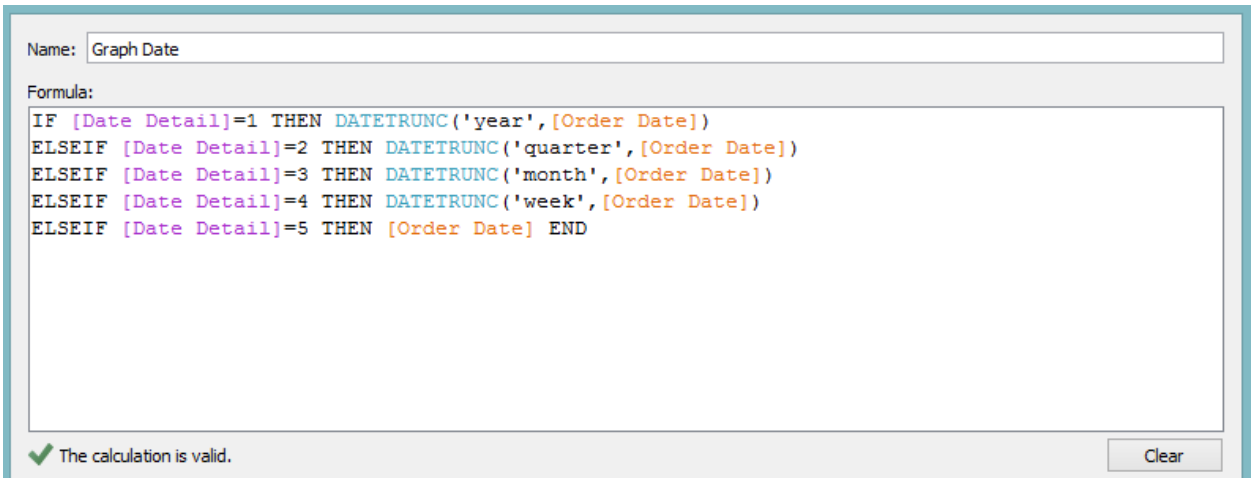


Adjust the level of drill down for dates with a parameter

First, create a parameter with variables to represent each of the date parts you want users to be able to select from.



Next, create a calculated field that truncates the date variable for the line chart based on the parameter selected.



Use the new Graph Date variable in place of the original Order Date in the line chart and set the displayed date part to day. With this, you can now use the parameter to drill down and up inside of a date range.

Pros:

- Even if users click on the plus sign at the end of the date axis, the chart stays the same.

Cons:

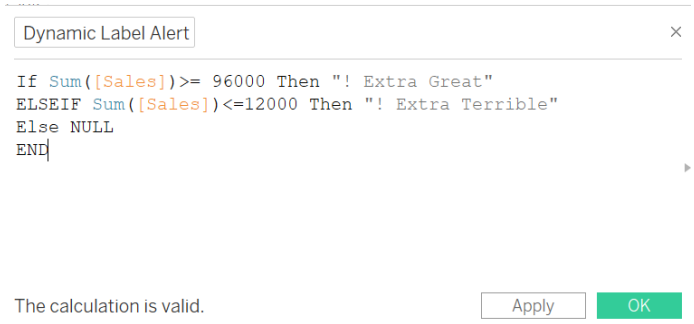
- A single date format must be selected to be used across all date parts.
- You have to find a place for an additional drop-down on the dashboard.

Revisiting Dynamic Labels with a Parameter

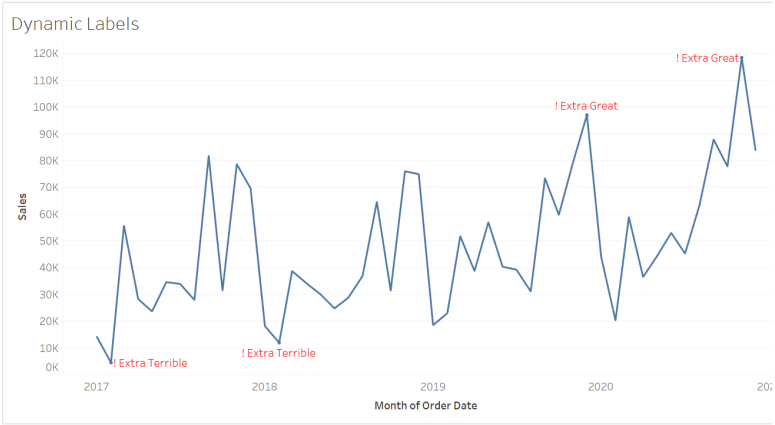
Now, we want mark labels to only show up when something extra important happens, and we want to give the user control over what denotes extra important with a Parameter.

Previously, we created a Sales over time visualization with Months of the Order Date on the Columns shelf and Sum of Sales on the Rows shelf.

Then we created a calculated field called Dynamic Label Alert which has this logic:



Then we placed the calculated field on the Label Marks Card and changed the color of the Alert with the Label Font properties.



Now replace the hard coded values with parameters:

More on parameters

Parameters are very powerful. Check out some other uses here:
http://www.vizwiz.com/2016/05/tableau-tip-tuesday-12-use-cases-for.html

Topic 7: Charts

This section provides an overview of several commonly used charts available in Tableau. The example uses Superstore data**.**

Bar Charts

Bar charts are one of the most common data visualizations. They are useful for highlighting differences between categories, identifying trends and outliers, and presenting historical highs and lows. They are a valid visualization for data that can be divided into different groups (e.g., product categories, project completion times).

Make Note!

Bar charts are of the most common, simple, and effective charts to presenting discrete data.
Use these when data can be categorized.
Bars use length to convey data values.
Cluster or stack bar charts to add depth to analysis.
IF time is variable, the chart should be presented vertically, and the time periods should be consistent.
IF time is constant, the chart should be presented horizontally.
IF time is not a relevant part of the data visualization, the chart can be presented either vertically or horizontally.

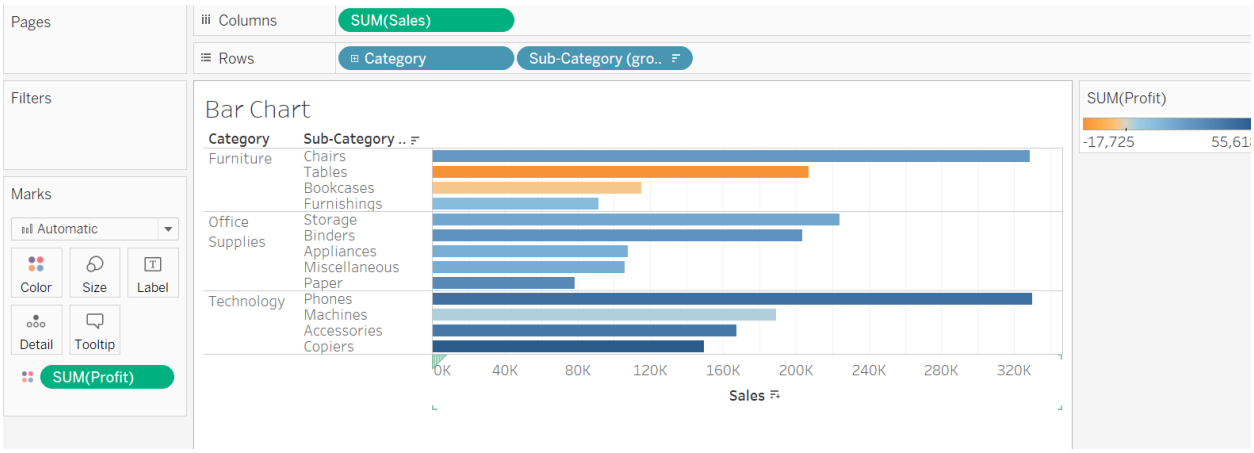
Let’s Analyze Sales (double click) by Category (double click) and drill down to Sub-Category. To make this easier to read let’s rotate it (click Swap in the toolbar) and sort it (click Sort icon).

This gives a very nice look at Sales. For example, Furniture (especially Chairs) seems to be selling well. But what about Profits?

Drag Profit to the Color shelf. It is clear that Tables are not profitable – and they are selling lots of them! Not good. Depending on what Tableau picked as a color scheme, you may need to adjust it since people with color blindness have problems with red/green. You may want to switch to orange-blue diverging.

Reminder: the Show Me pane provides other possibilities for visualizing the same data – you can click a couple to see other options, but then use the Back button to return to the bar chart you were creating.

Notice that Office Supplies has a number of smaller categories that are cluttering the analysis. This is a common issue in a dataset, but you do not want to remove these small categories (fasteners, labels, envelopes, art & supplies) because they may be significant in aggregate. Instead, select the smaller subcategories (hold down CTRL and select the words or CMD on Mac) and clicking the paper clip to Group them together. Right-click and Edit Alias to rename the group to Miscellaneous.

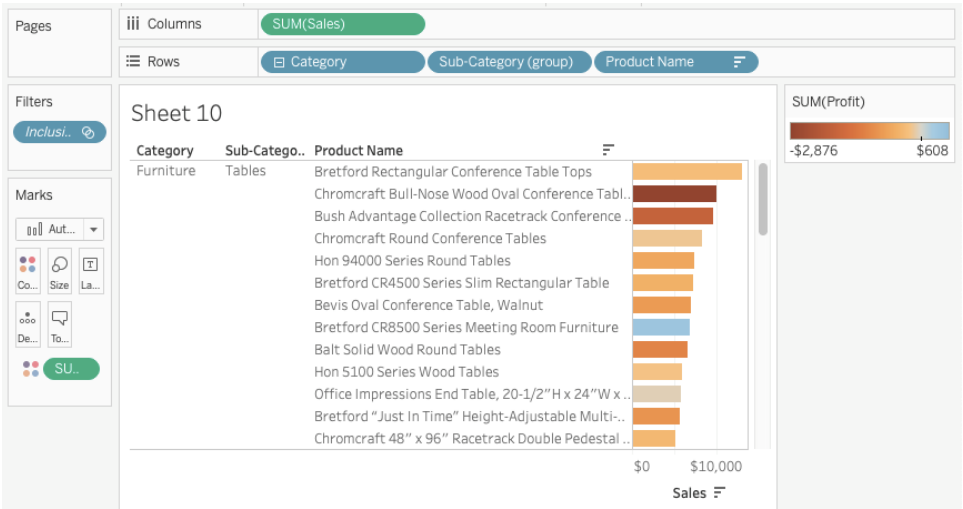


It is important to understand that this does not change the underlying data – the original sub-categories are still there. What this has done is simply created a Group in the visualization.

Now you see that there is a problem in Tables. Right-click Tables and select Keep Only.

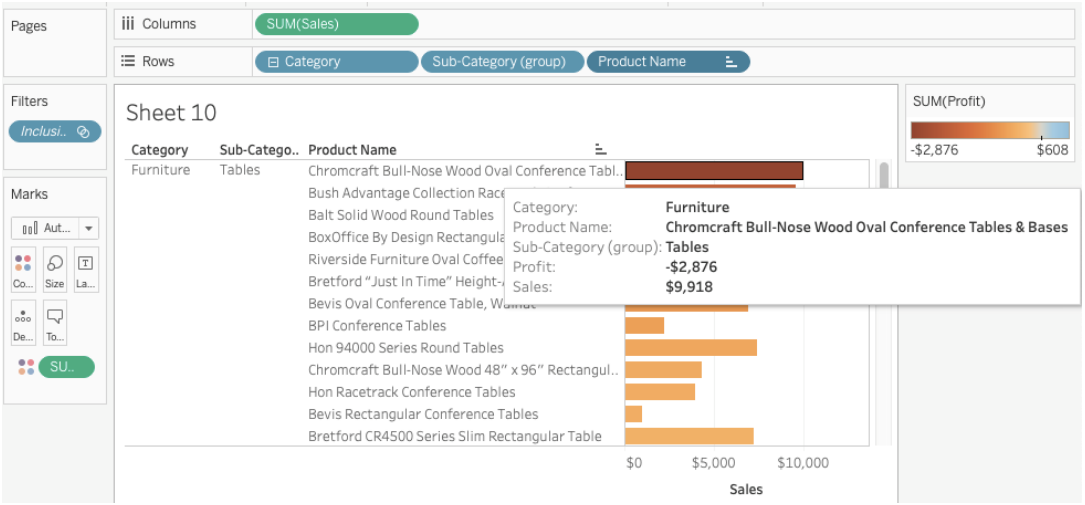
Now that you are just looking at Tables, drill all the way down to the Product level. Sort descending by Sales.

Finally let us look at the actual underlying data from the top selling products: Click the first bar… View Data, Full Data. And now you have a nice list of transactions that you could want to investigate.



Sorting by something not already in the visualization

What if you wanted to see the product that is losing the most money (as opposed to the top selling product)? You can see this by sorting the bar chart by profit. To do this right click the Product Name pill and select Sort. Set the Sort Order to Ascending and Sort By Field with Profit selected. Press OK. Now the top of the visualization shows the product that is losing us the most money, although the bar still shows sales.



Scatterplots

Scatterplots are a very easy and powerful way to visualize the relationships between numeric variables (Measures).

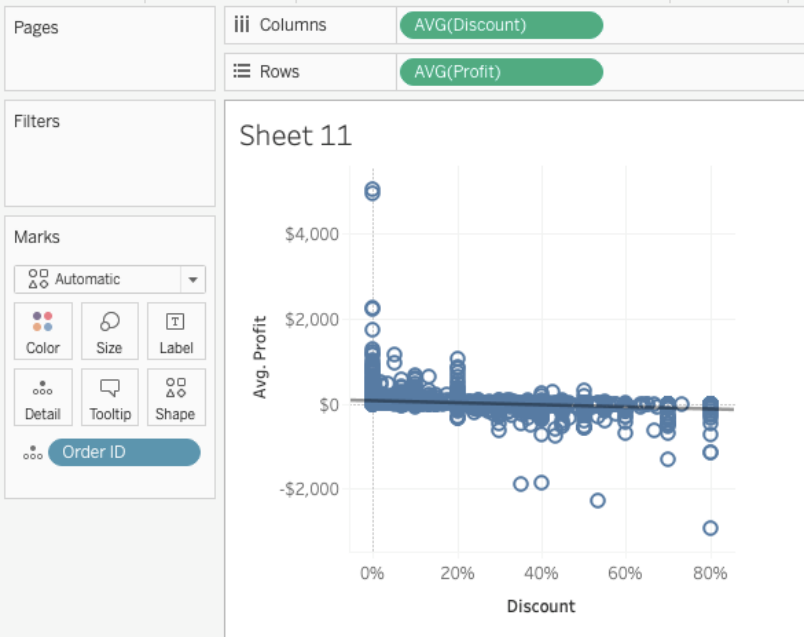
Make Note!

Use scatterplots to show relationships between numerical measures.
Use scatterplots to identify patterns, trends, clusters, or outliers.
Scatterplots lend themselves to statistical analysis, so it is often helpful to include statistical information such as trendlines.

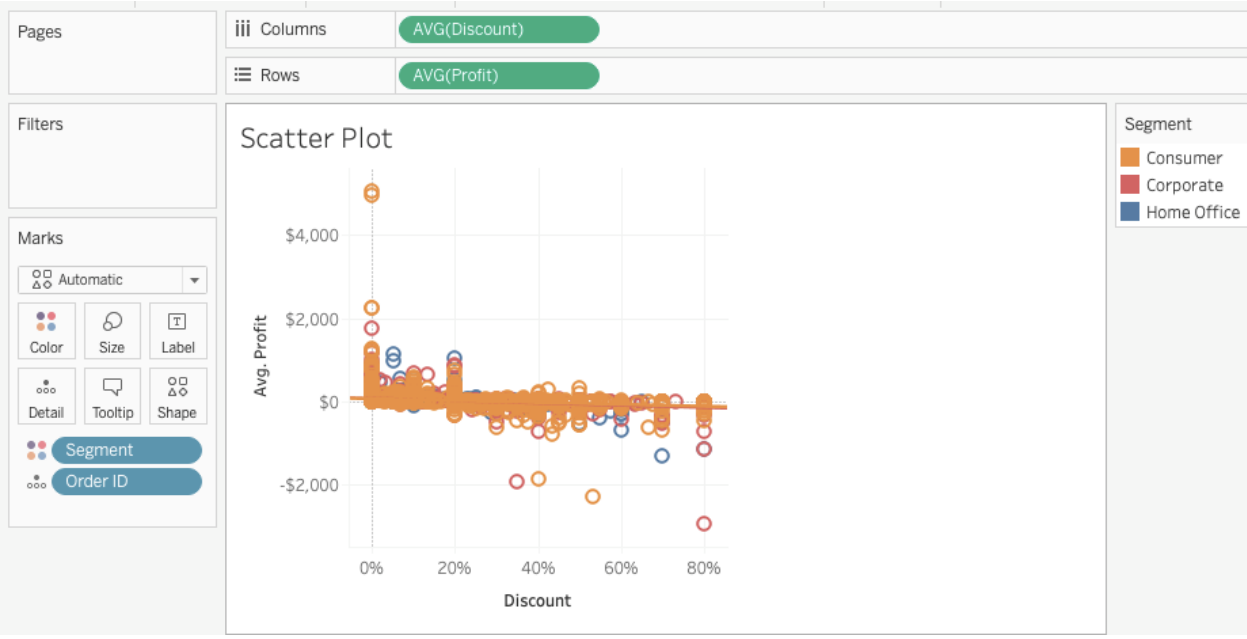
For example, is there a relationship between Discount and Profit (seems like a reasonable possibility)? Let's take a look by selecting those two fields and clicking Scatterplot on the Show Me menu. Go to Analysis, uncheck Aggregate Measures. This shows me all of the data. Uncheck this to go back to the data. Now drag Order ID to Detail.

Convert the discount to average by right clicking on the pill and setting aggregated to average.

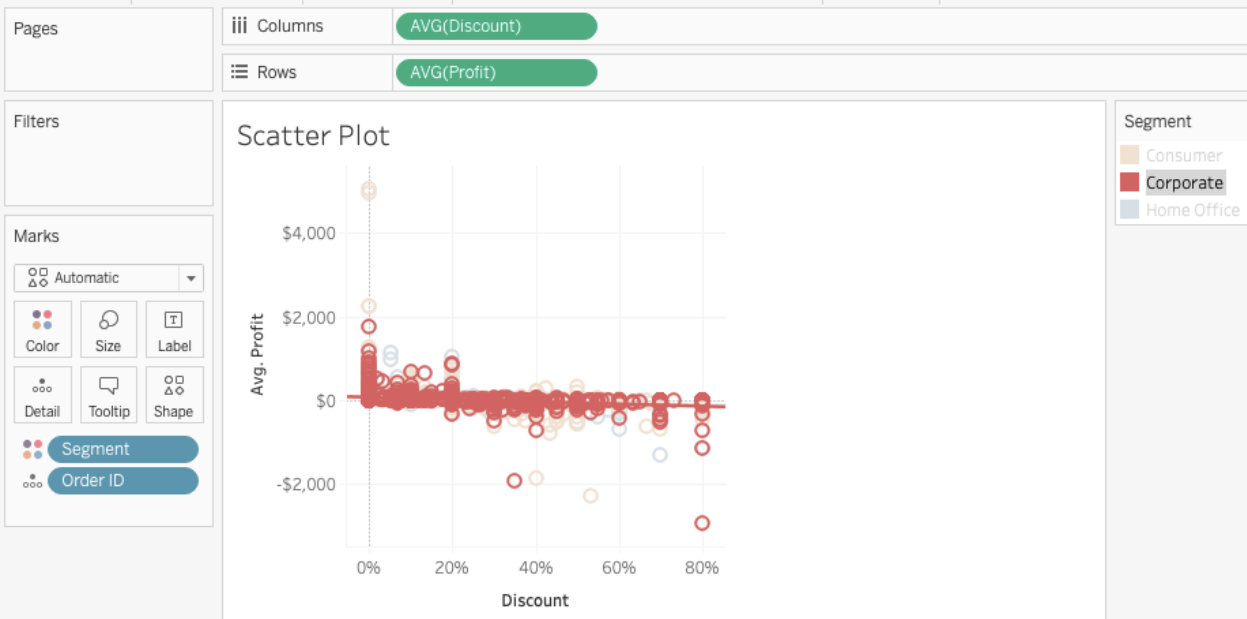
Swap if needed to get Discount on the horizontal axis – looks like Profit tends to decrease as a larger Discount is given (confirms intuition).



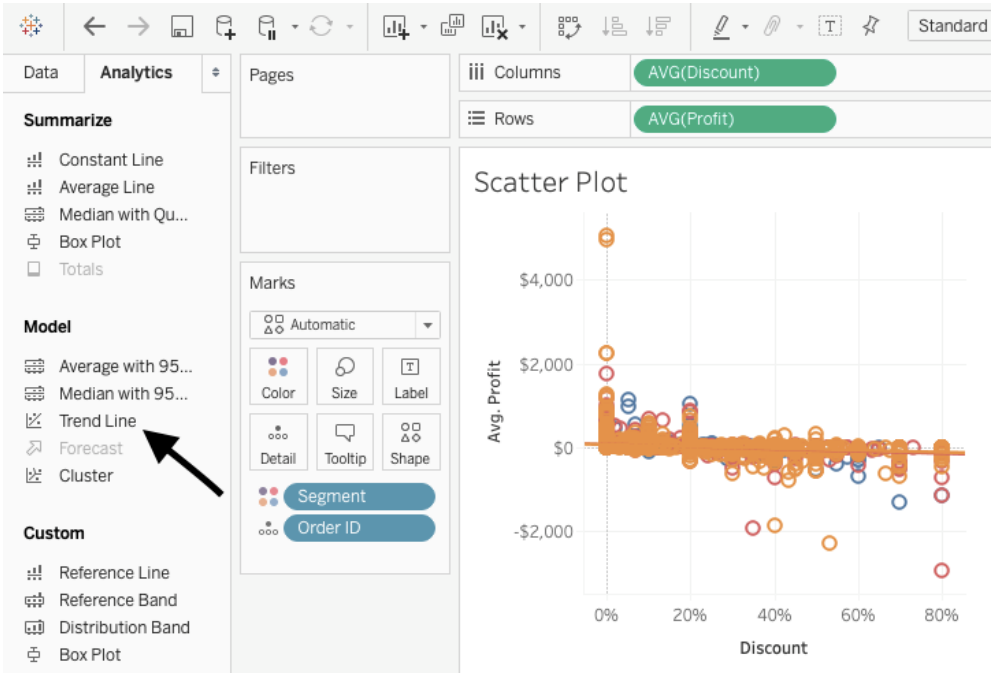
Remember that with Tableau, unlike many other tools, you don't need to stop here. It is very easy to keep adding data. For example, drag Segment to Color. This makes it easy to see what segment certain outliers belong to.



If you prefer to focus on just one segment (while leaving all the data displayed), you can choose an option in the Legend and select, for example, to focus our visualization on the Corporate segment by highlighting it.



Go back to viewing all the data. Go to the Analytics pane and drag out a trend line (choose linear).



This trend line does not look very impressive but take a look at the y-axis scale. You need to investigate more before dismissing this as a negligible relationship. Hover over the trend line. The p-value is very small, so in fact the relationship between these variables is statistically significant. And you can see by the negative coefficient that the relationship is negative. However, the R-Squared value is somewhat low, so the simple trend line has not explained very much of the relationship. Drag the trendline out of the visualization to remove it.

You can also add a Segment filter to the plot so you can look at this for each segment individually.

Line Graphs

Like Scatterplots, line graphs are good for looking at the relationship between 2 variables (although a scatterplot is often better for this). Line graphs are often used for looking at how something changes over time. This is called time series analysis.

Make Note!

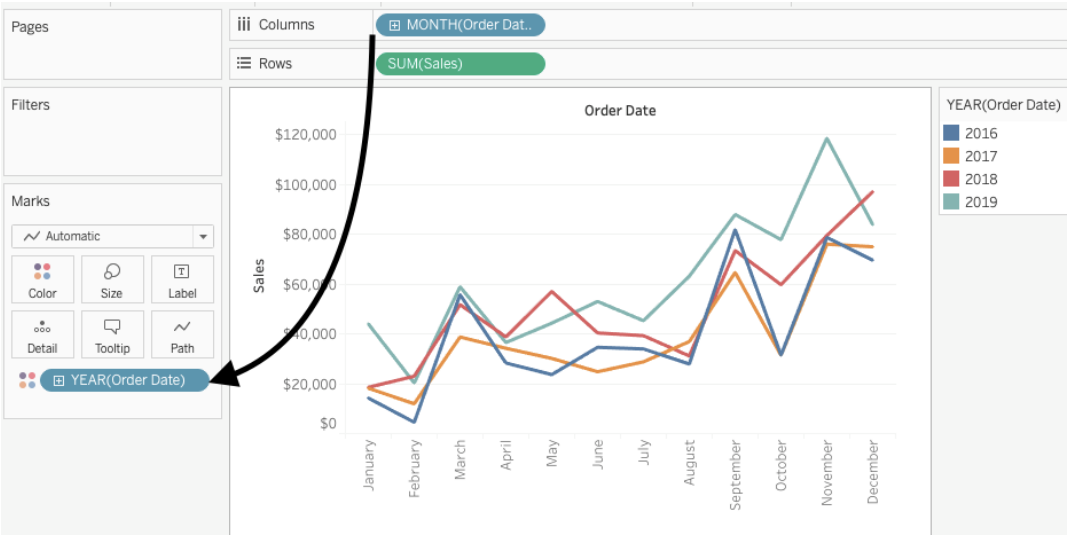
Use line graphs to show trends over time.
Use line graphs when interpolation of data makes sense.
Lines imply continuation so may not be appropriate for discrete data.
Each line should be clearly distinguishable from the others.
More than 4-6 lines in a single chart tends to be too much because the chart becomes confusing.

Let’s start by looking at sales over time. Double-click Sales, double-click Order Date. Tableau understands date fields and automatically breaks them down by Year, Quarter, etc.

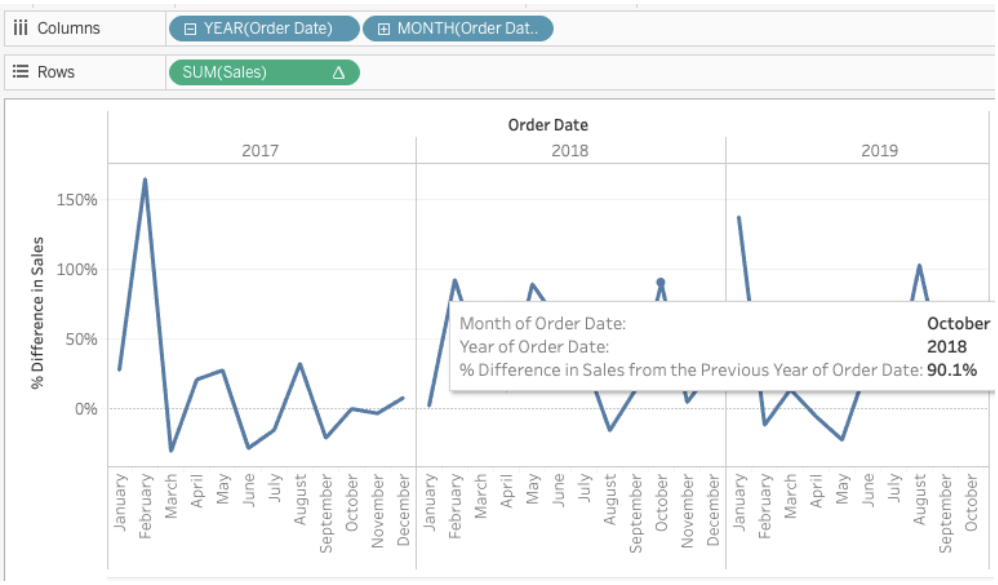
If you need more detail than yearly sales results, expand to Quarter and then Month. For this, we really don’t care about Quarter so drag it out of the visualization. This will leave month and year in the column shelf.

Note that dual axis is very easy in Tableau. To demonstrate, drag Profit to the sales vertical axis. Then Undo to get back to Sales only for now.

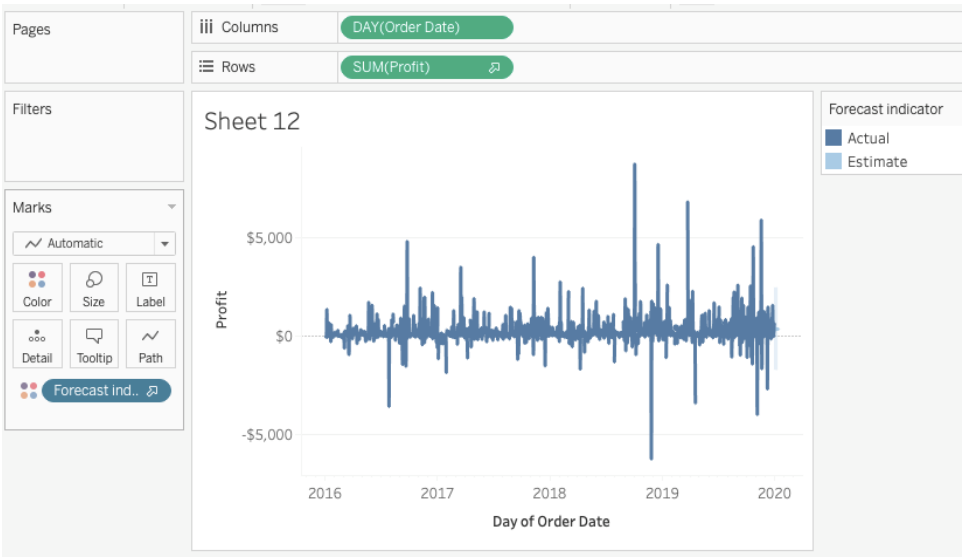
Remember that you can now drag things around to change the visualization. You have discrete dates right now (you have already learned how, with discrete dates, each part of the date is treated separately). So for example you can drag Year from the column shelf to the Color mark card and see direct comparisons between the years.



Let us get back to a separate line chart for each year (drag the year pill back to the column shelf). If you want to look at a common calculation for time series data, such as year over year growth: Right-click on the green Sum(Sales) pill. Select Quick Table Calculation… Year over Year Growth. Now you should see month(s) in which sales went down year over year (where the line dips below 0%).



With Tableau it is also easy to add a Forecast. Change the date field to continuous days and change the sales pill to a profit pill (as shown below). Go to the Analytics pane and drag out Forecast (interested students can learn about the specific forecasting algorithm in the Tableau documentation)



Side Note: Forecast by default will not include the last mark. If you want to include that mark in your forecast you should change the forecast options.

Histograms

Histograms display the shape of a distribution and can tell you if the data is skewed. This provides a good overview of the dataset. For histograms in Tableau, you must utilize a ”bin” in addition to a measure. Bins are used as slots for measures to be positioned in the visualization.

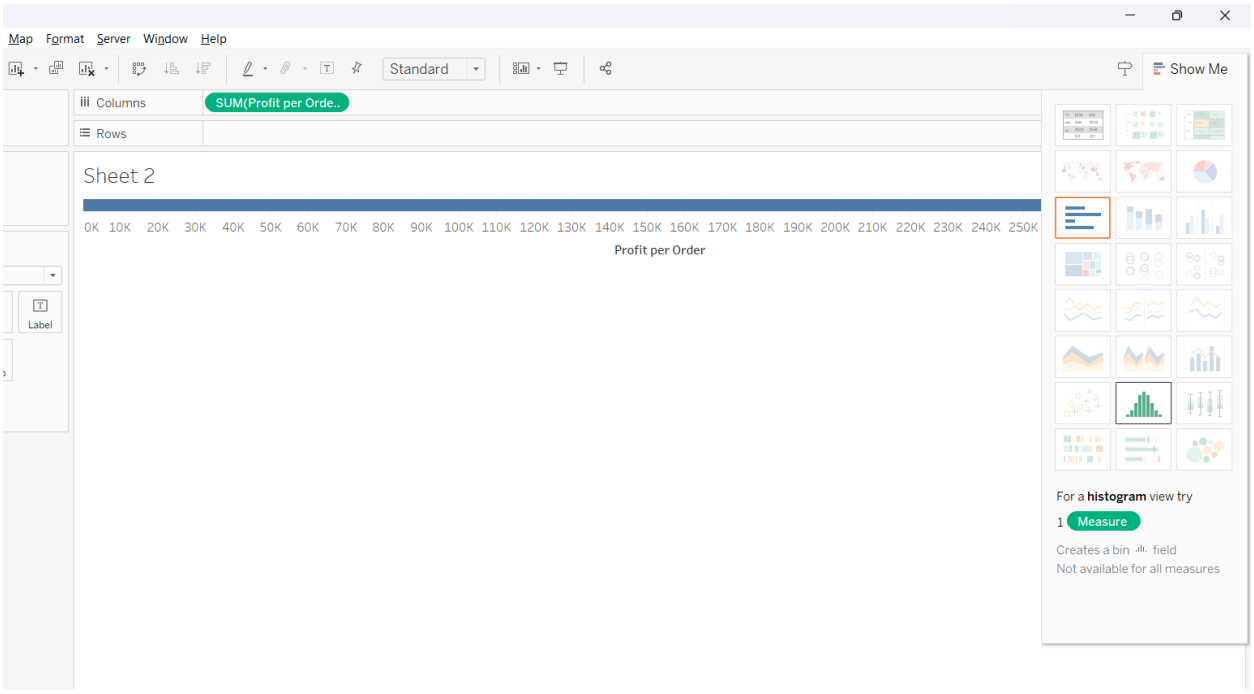
Make Note!

Use histograms to show distributions within a group or to show proportions of data that fit within particular groupings.
Bars use length to convey data frequency.
Data should be presented with data groupings, each of equal size, on the horizontal axis.

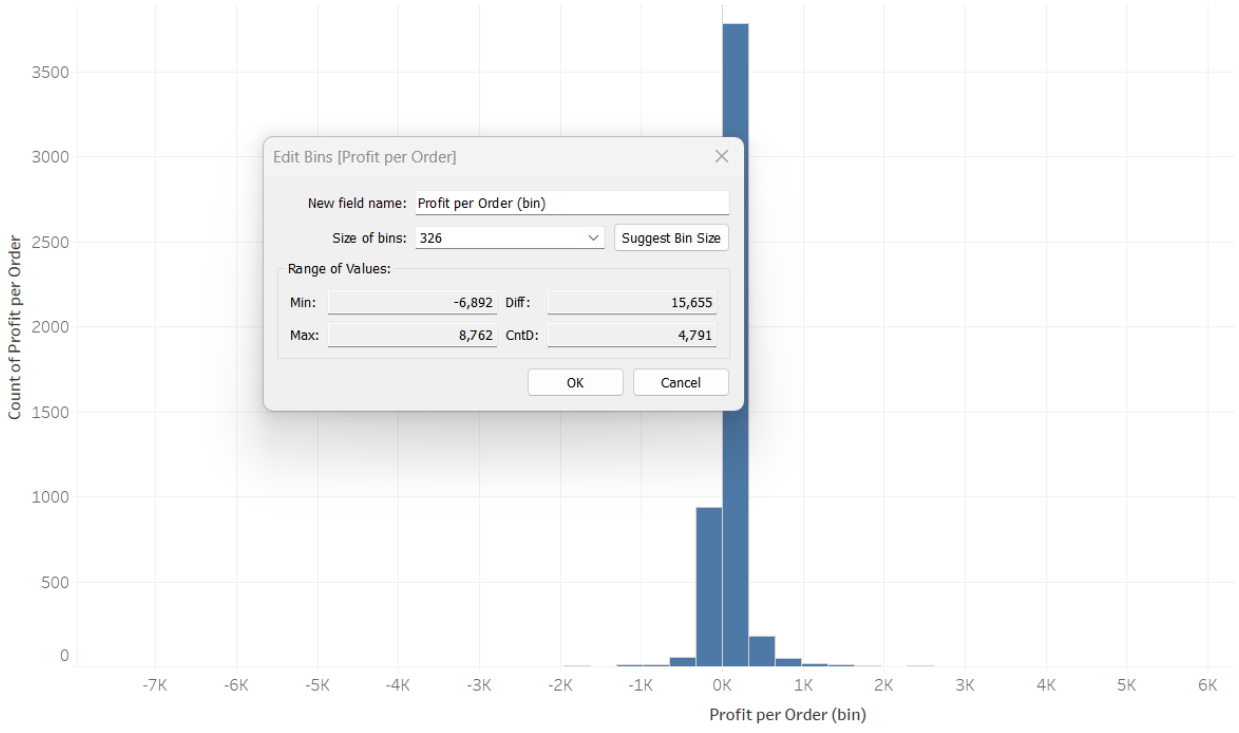
This exercise will display Profit Distribution across all sales.

1. Create a new calculated field called Profit per Order. You will need a calculated field: {Fixed [Order ID]: Sum ([Profit])}.

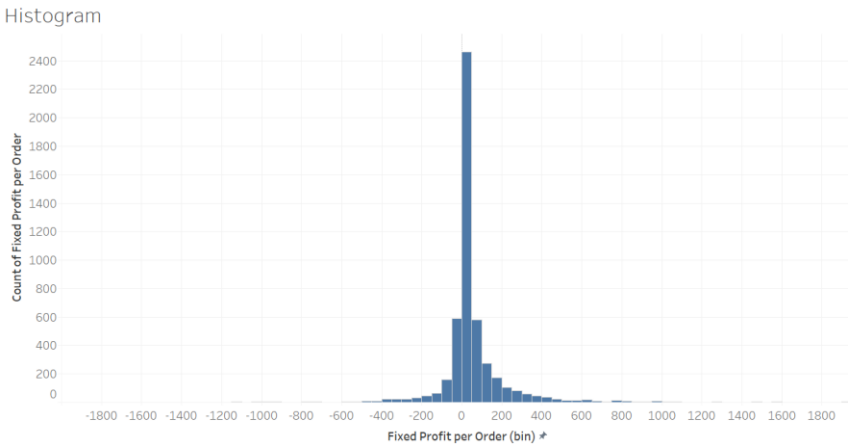
2. Drag Fixed Profit per Order to the columns shelf and then select the histogram under the ‘Show Me’ chart selection.



3. When you select histogram, Tableau automatically creates a ‘bin’ for each Profit per Order to fit into for the histogram. Adjust the size of the bin to 50.



4. You will need to edit the Axis (Min -2000, Max 2000) to match this picture



If you want to make the size of the bins a dynamic parameter, see Topic 6 for that example.

Adding a Cumulative Distribution to a Histogram

This section provides a detailed walk-through of how to add a cumulative distribution to a histogram in Tableau using the previous histogram.

From the previous example, you had a calculated field: {Fixed [Order ID]: Sum ([Profit])}. You selected Fixed Profit per Order and chose the histogram from Show Me. You edited the Axis (Min -2000, Max 2000) and the Fixed Profit Per Order (bin) to 50 to match the picture above.

To then add the cumulative distribution to the histogram:

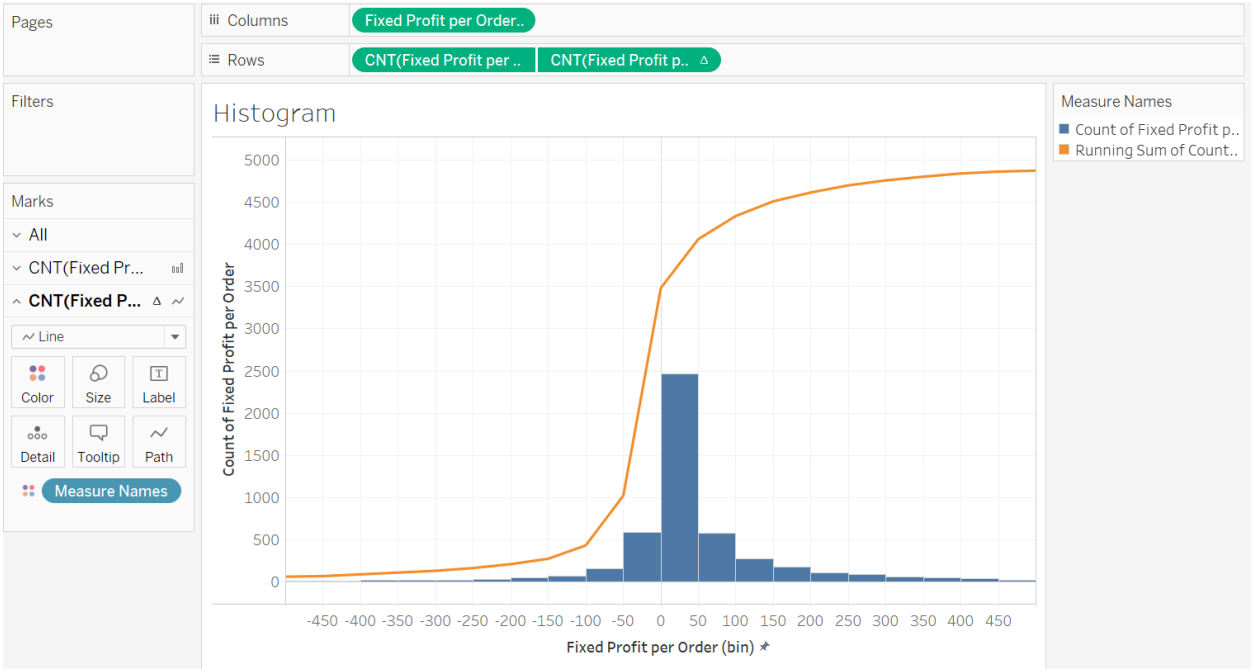
1. Duplicate the CNT(Profit) pill on Rows by CTRL-drag (CMD-drag on Mac) the pill next to itself on the Rows shelf. You should see two histograms stacked on top of each other now.

2. This gives you a new marks card – now you can create 2 different types of mark for a single data field. On the new marks card “CNT(Profit)(2)”, change the mark type to a line. Change the color as well, if desired.

3. Apply a Quick Table Calculation of Running Total to the duplicated field

4. Right click on the duplicated field on the Rows shelf and make it a dual axis. Synchronize the axes to get an accurate perspective.

You may want to further trim the outliers from the Axis (-500, 500)



Heat Maps

Heat maps use color to display patterns in data. They are particularly useful for large data sets as the use of color enables very fast identification of patterns.

Make Note!

Heat maps use color to show patterns.
Like treemaps (see below), heat maps are for finding patterns in large data sets.
Unlike treemaps (see below), these are not for organizing hierarchical data.

Let us use heat maps to visualize the margin that you are making on your products.

If at this point you do not have a calculated field for Profit Ratio, make one:

Create Calculated Field called **Profit Ratio**:

```
SUM([Profit])/SUM([Sales])
```

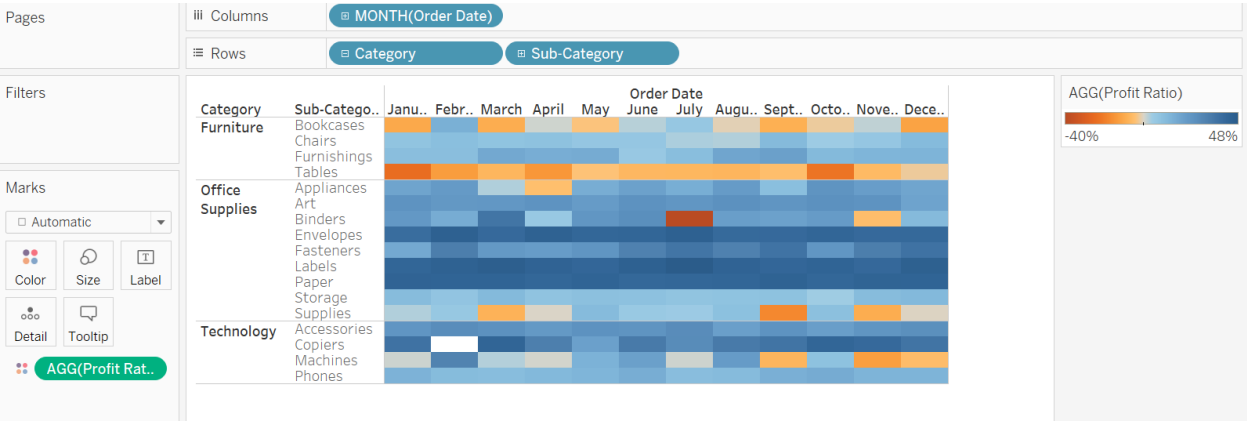
Drag **Category** and **Subcategory** to **Row**

Drag **Order Date** to **Columns**, and then change to **Month (discrete)**

Drag your **Profit Ratio** field to the visualization (drop in the center of the table). You have a table of numbers. Not the most user-friendly…

Drag **Profit Ratio** to the Color Marks card (and remove from the Text Marks card). Change color to Orange-Blue Diverging (if needed) This is much better than the table of numbers! You

immediately see some insights – e.g. maybe you are selling tables as a loss leader and making it up on chairs?



Packed Bubble Charts

Packed bubble charts are used to display data in a cluster of circles. Dimensions define the individual bubbles, and measures define the size and color of the individual circles.

Make Note!

Packed Bubble Charts display categorical data with quantitative values where the size of the bubbles represents the magnitude of the data.
The data should have a clear quantitative aspect that can be effectively represented by the size of the bubbles and that there are distinct categories or groups to compare.
This chart conveys meaningful insights through visual comparison of bubble sizes.

The basic building blocks for a packed bubble chart are as follows:

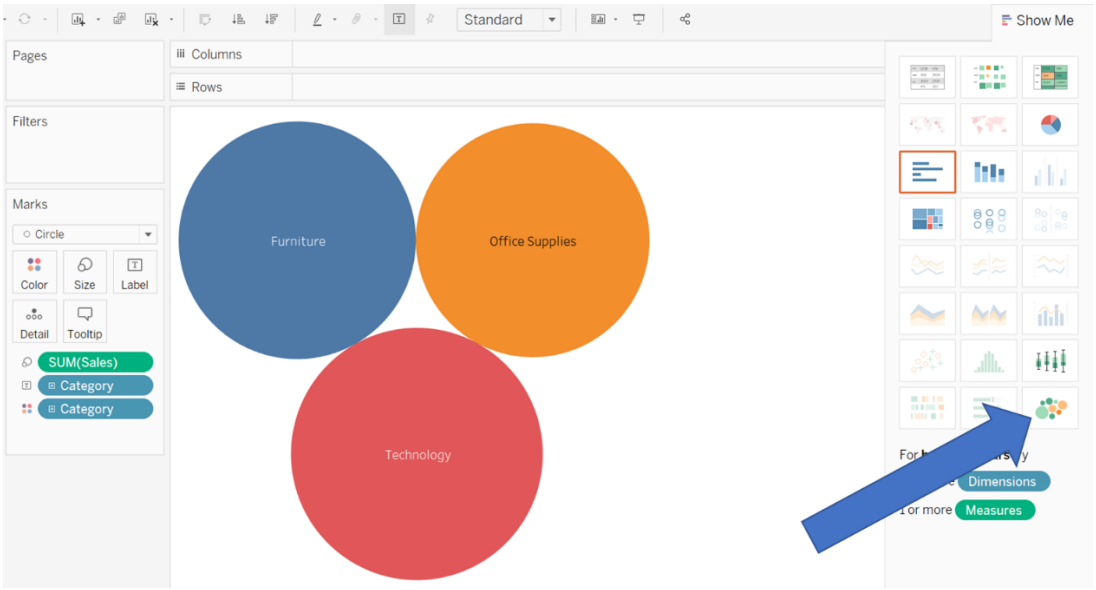
Mark type: Circle
Detail: Dimension
Size: Measure
Color: Dimension or Measure
Label: Dimension or Measure

To create a basic packed bubble chart that shows sales and profit information for different product categories, follow these steps using the Superstore data:

1. Drag **Category** to Columns.

2. Drag the **Sales** to Rows.

3. Click Show Me on the toolbar, then select the packed bubbles chart type.

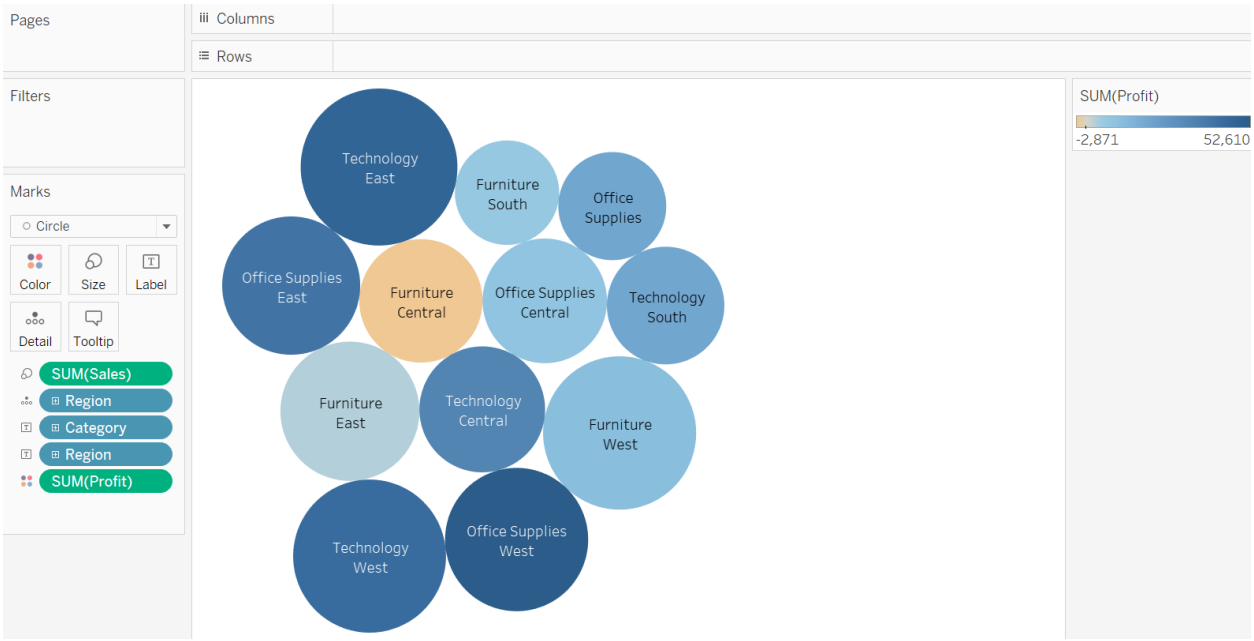


4. Drag **Region** to Detail on the Marks card to include more bubbles in the view.

5. Drag **Profit** to Color on the Marks card.

6. Drag **Region** to Label on the Marks card to clarify what each bubble represents.

The size of the bubbles shows the sales for different combinations of region and category. The color of the bubbles shows the profit (the darker the blue, the greater the profit).



Pie Charts

Pie Charts are often criticized because in general they are poor at communicating data. They take up more space and are harder to read than some other alternatives. Also, the brain's not very good at comparing the size of angles and because there's no scale, reading accurate values is difficult.

Make Note!

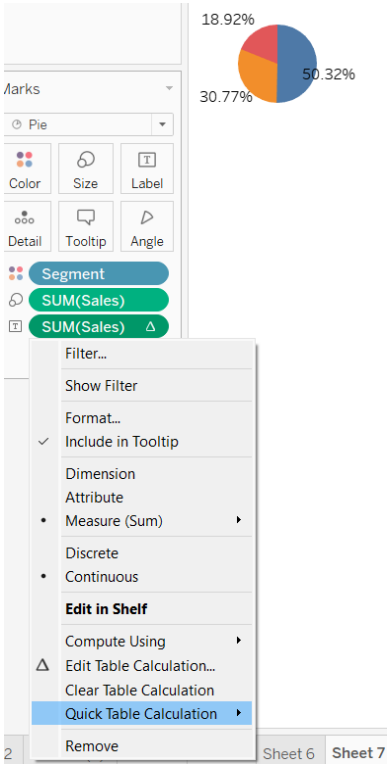
Pie charts can be difficult to interpret because of how humans understand angles.
Use these to compare parts of a whole or percentages.
Use these when they contribute to a story, but not necessarily for analysis.

But regardless, they are commonly used in visualizations, so you may need to know how to make one. Pie charts require a dimension and a count of the dimension (measure).

1. Select **Segment** and **Sales**

2. Show me pie

You can add labels by dragging Sales into the label mark card. Now right click the Sales pill and use a quick table calculation to display % of total on the pie chart.



In the center of the toolbar, switch the View to “Entire View” to make the pie chart larger and easier to see.

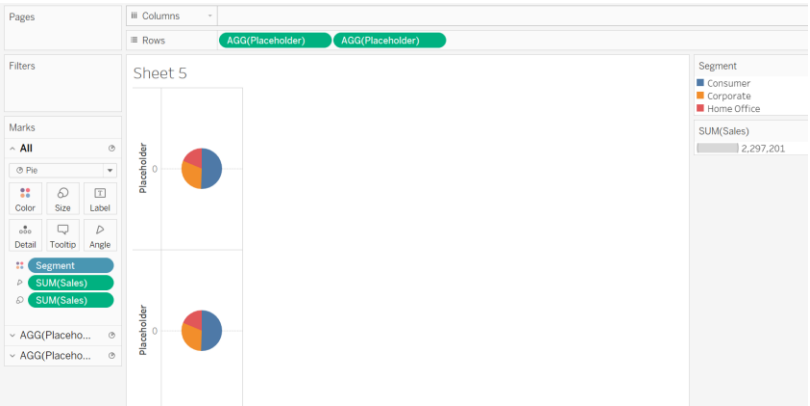
Donuts

Pie charts can look like a donut. It is not a great way to actually communicate information but it does look cool so let us try it out.

Make Note!

Like pie charts, donut charts can be difficult to interpret because of how humans understand angles.
Donut charts can be made more understandable by including a takeaway in the otherwise empty center.

1. Select **Segment** and **Sales**
2. Show me pie
3. Create a Calculated Field called “Placeholder: that has this formula MIN(0). There are multiple formulas that will work here AVG(0) does also.
4. Bring **Placeholder** to rows TWICE. This is just making two copies of the pie chart centered at 0.



5. Select Dual Axis for the lower pie chart, and you’ll see one pie chart. It’s like you’ve done nothing but what you really have is two pie charts on top of each other.
6. Go to the lower marks options (this controls the top pie chart)

a. Remove Segment from color

b. Remove Sales from Size

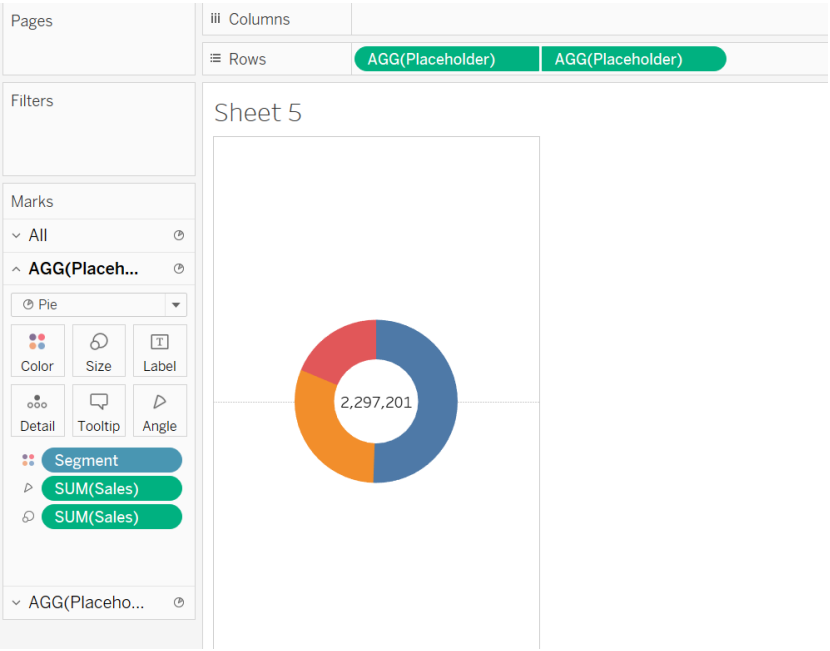
c. Color to White

d. Size a little smaller

e. Sales to Label

7. Remove axes by right clicking and selecting “Show Header”

8. You can rescale by dragging the right side of the view if you are zoomed in too much. If you have problems with them sizing, make sure you removed the measures and dimensions from the Marks shelf for mark you want to be on top of the pie chart.



To get rid of the center line, choose Format, then lines, then no zero line.

Tree Maps

A tree map is a visualization that nests rectangles in hierarchies so you can compare different dimension combinations across one or two measures (one for size; one for color) and quickly interpret their respective contributions to the whole. When used well, some portion of the tree map will be composed of large rectangles where additional context can be added as labels. Also, similar to a scatter plot, tree maps allow you to reasonably communicate and consume hundreds of marks on a single view, making it easier to spot patterns and relationships that you would otherwise not be able to see.

Make Note!

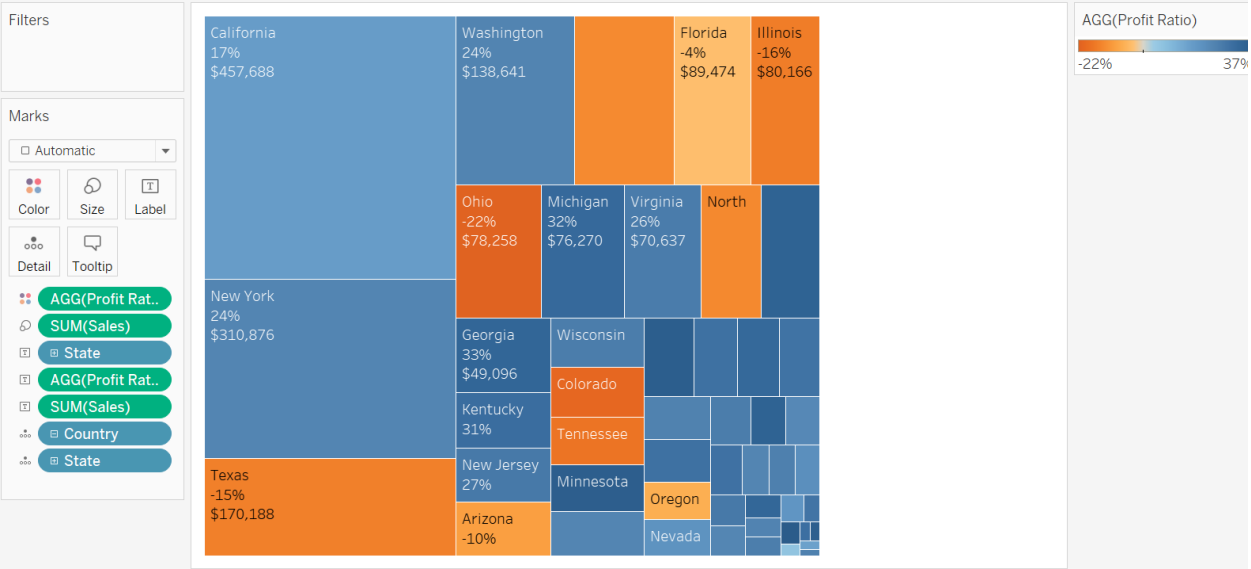
Use treemaps to present large, complex data that is hierarchical.
Use these with discrete data.
Use these when data can be categorized.
Sunburst charts are generally better for showing hierarchies but can be confusing in other ways.
When used for data that are not hierarchical, treemaps show relative contributions to the whole but are essentially rectangular pie charts.

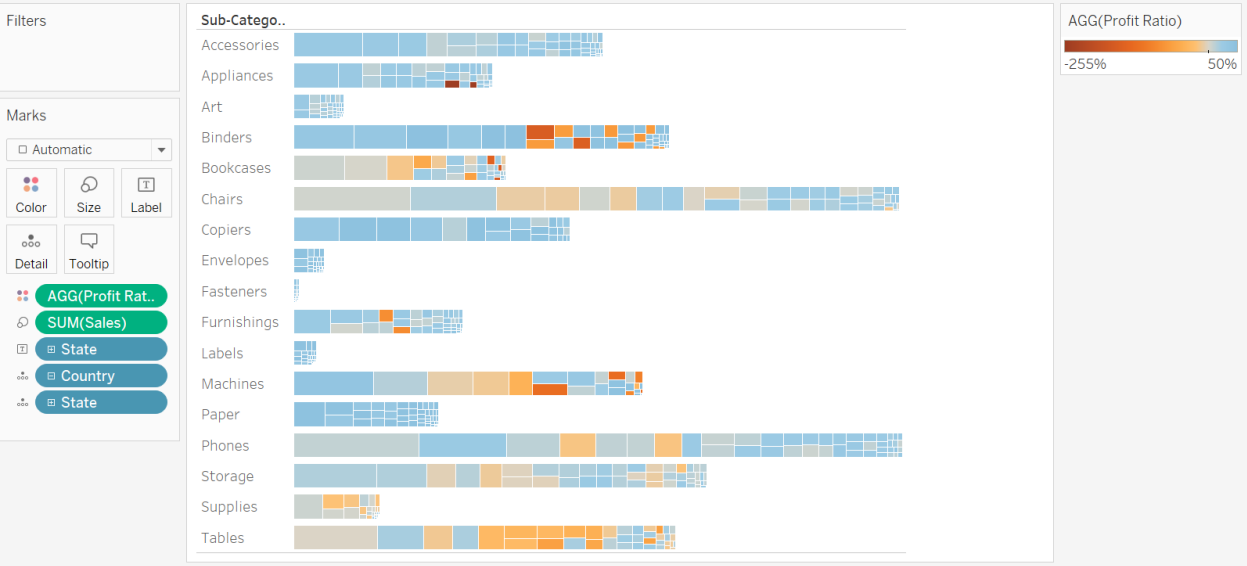
To make a tree map, begin by changing the mark type for a view from Automatic to Square. Then drag the primary measure that you want to evaluate to the Size Marks Card and the secondary measure to the Color Marks Card. Thus, the primary measure will control the size of the squares and the secondary measure the color of the squares.

For example, drag Sales to Size and Profit Ratio to Color. Very little has happened because there is no detail.

Drag State to the Detail Marks Card. Then also drag State to the Label Marks Card. Also drag Sales and Profit Ratio to the Label Marks Card.

If you want to see the same analysis at the Category level, drag Category to the Rows Shelf.





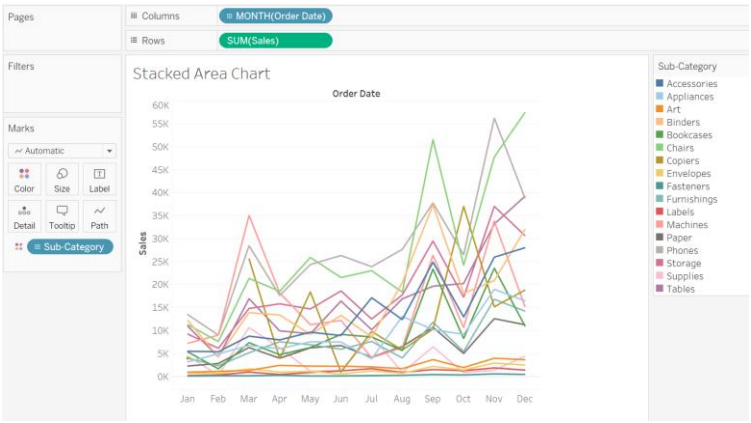
Stacked Area Charts

A stacked area chart ‘stacks’ trends on top of each other to illustrate how a part-to-whole distribution changes over time. Combined with a table calculation that computes the percent of total for each dimension member, stacked area charts are an effective way to evaluate distributions.

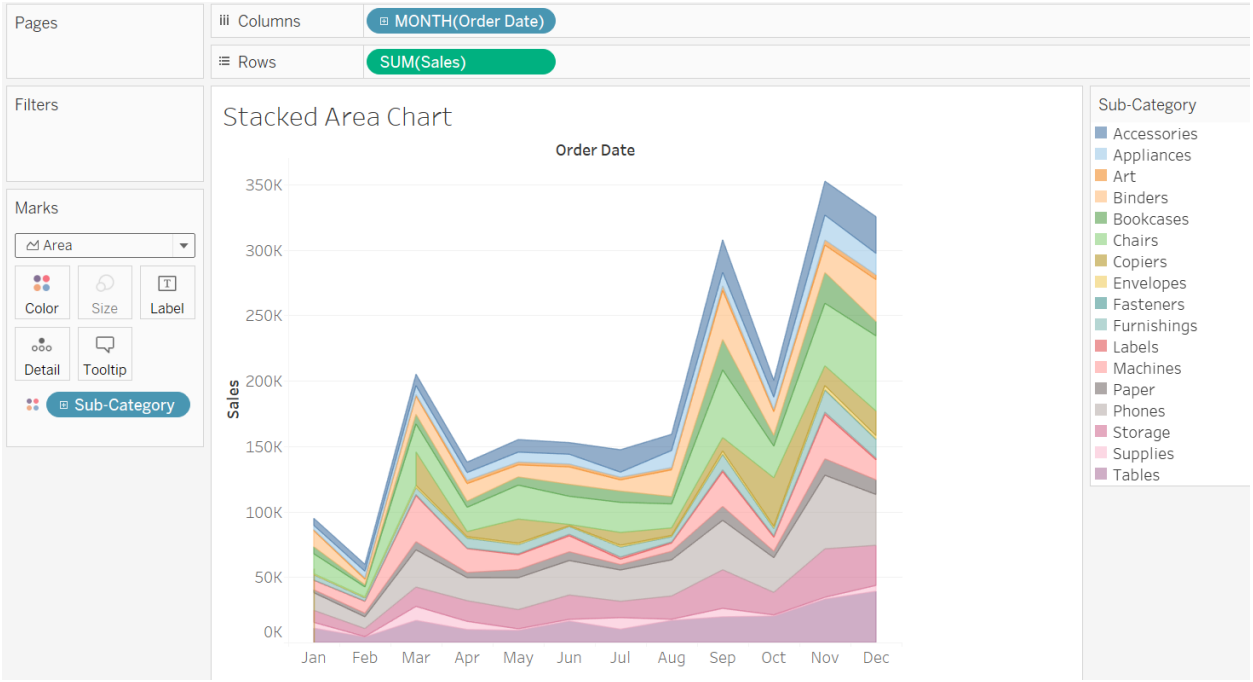
Make Note!

Use stacked area charts to compare a data subset to the whole.
Ideally, stacked area charts show percentages instead of absolute values.
Stacked charts are best to show how distributions are changing over time.
The order of data series, i.e., layering order, will impact the visualization and potentially its interpretation.
Be careful with stacked area charts – the change of individual data subsets over time should not be evaluated because data subsets inherit the trend(s) of the data subsets lower in the chart.

Start with a line graph showing Sales by Sub-category by discrete Month of Order Date from the Superstore dataset.



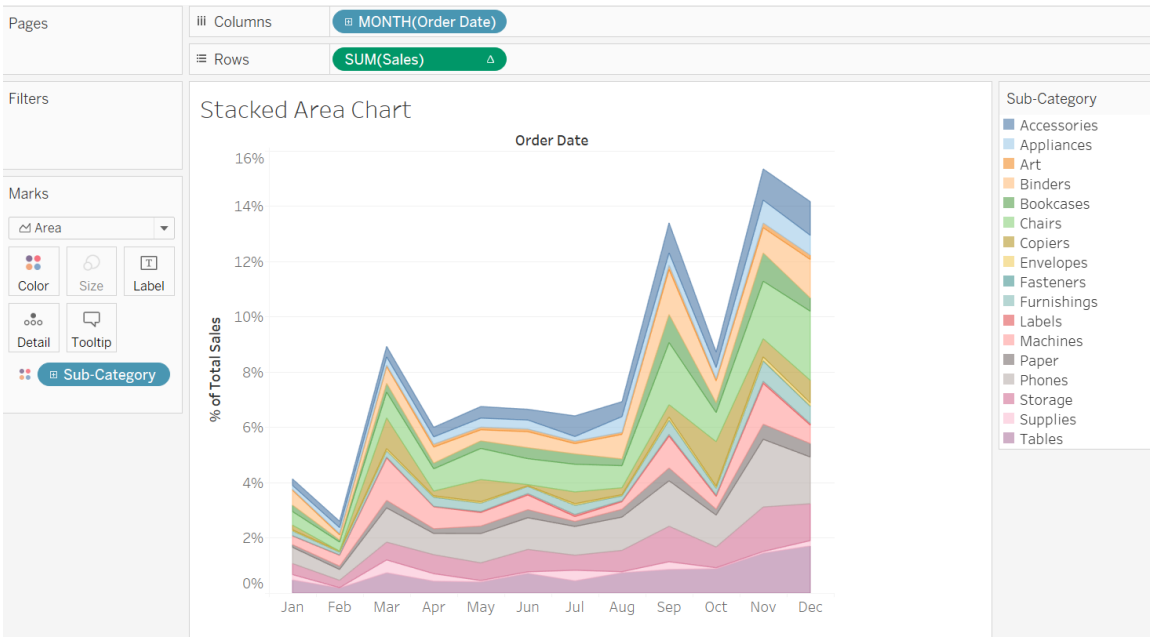
To create a stacked area chart, change the mark type from Automatic (or Line) to Area:



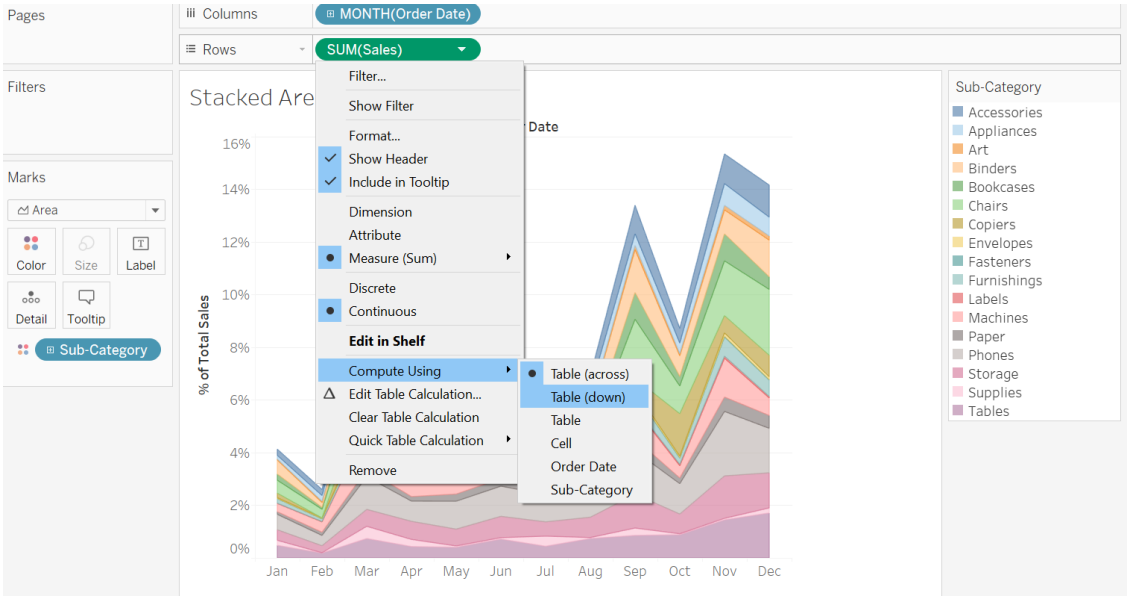
When the mark type was changed from Line to Area, the values for each sub-category were stacked on top of each other. This helps us understand the monthly total across all 17 sub-categories, but it is very hard to see trends (since after the first trend on the bottom of the stack, each subsequent trend inherits the trend below it).

This is a good visualization when the total axis equals 100% and each individual dimension is displayed as a percentage of the total.

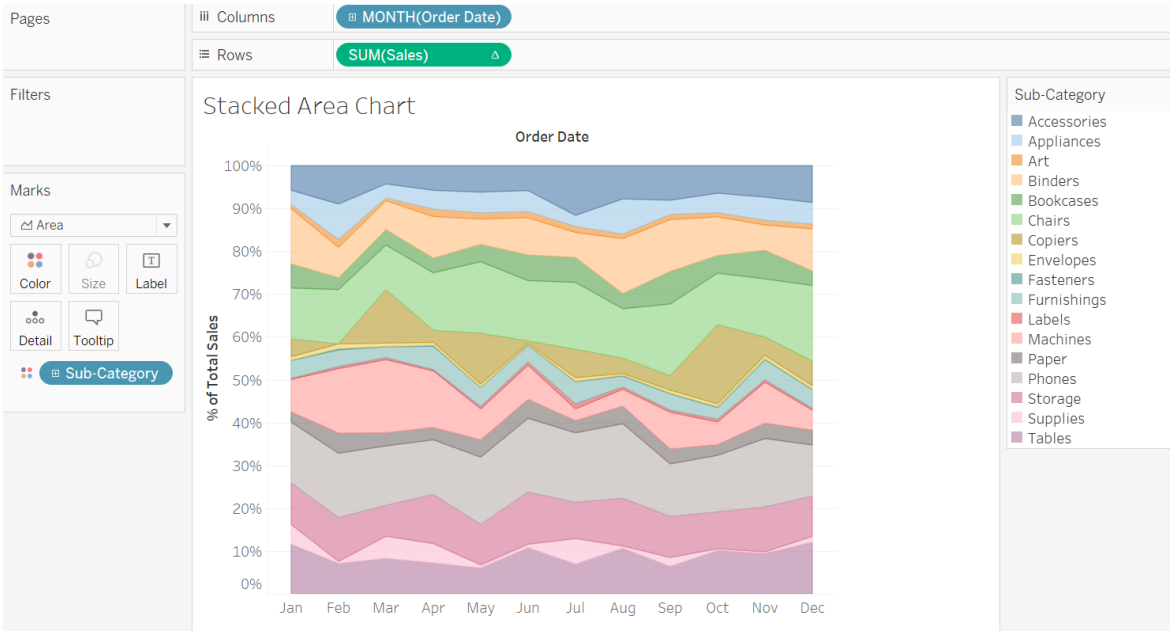
Right-click the Sum(Sales) and choose Quick Table Calculation, Percent of Total which changes the above visualization to the below visualization.



To change the view so that the axis totals 100% and the areas represent each contribution to the month's total, the table calculation needs to be changed to compute using Table (down).



Now we can see each sub-categories' monthly contribution:



Stacked charts are best to show how distributions are changing over time.

Box Plots

Box plots, also known as box-and-whisker plots, show the distribution of values along an axis. Boxes indicate the middle 50 percent of the data (that is, the middle two quartiles of the data's distribution).

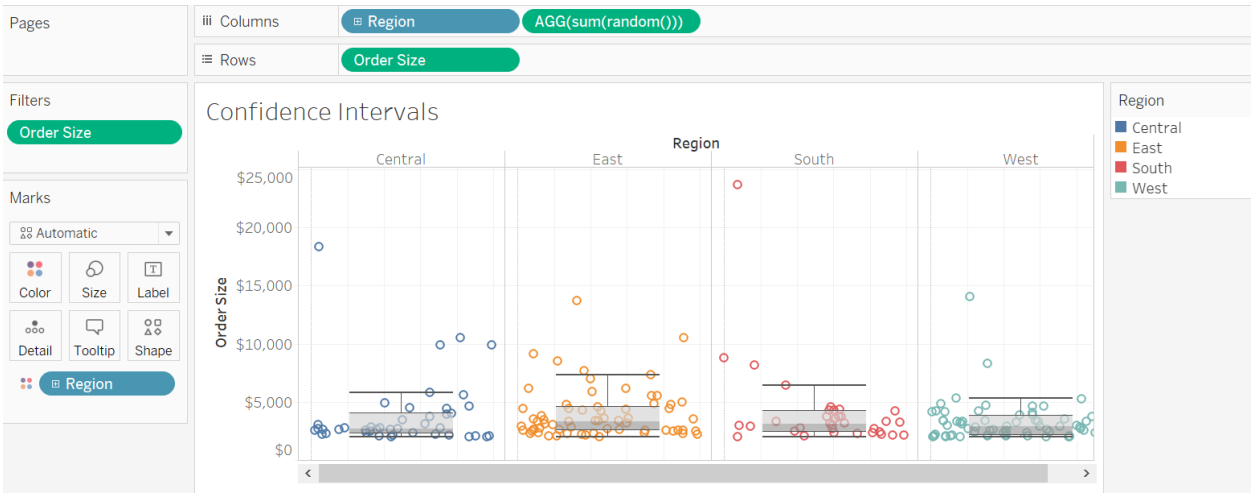
.

Make Note!

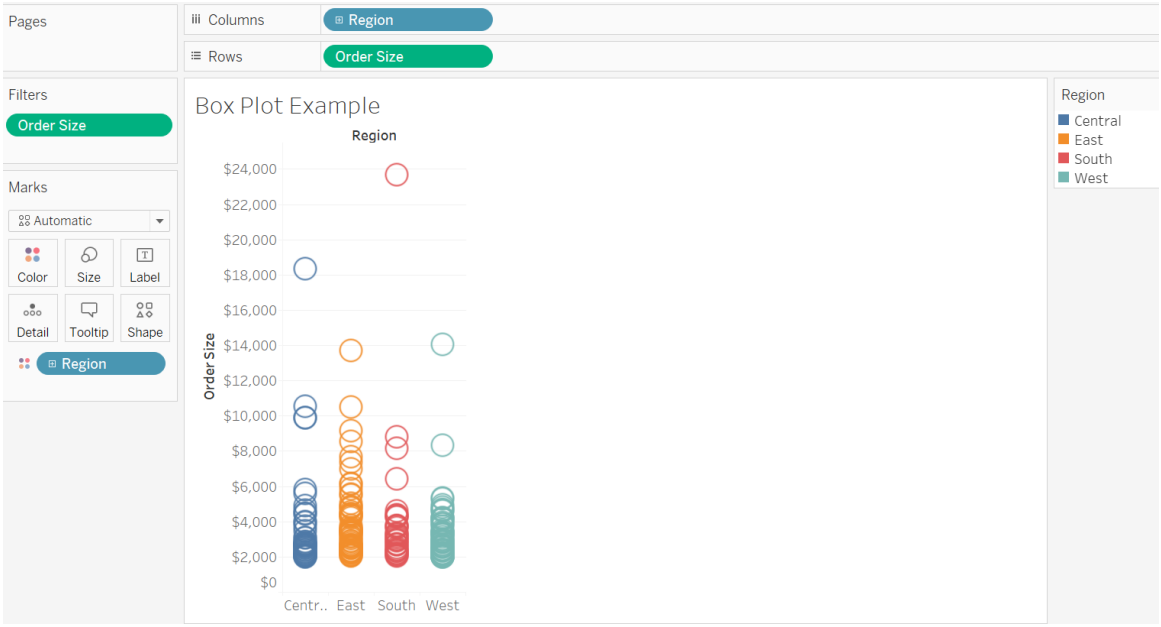
Box and whisker plots (or box plots) are best suited for displaying the distribution and variability of a dataset.
They effectively summarize data using key descriptive statistics.
When using box and whisker plots, ensure that the data is continuous and has enough variability to make the plot informative.

You can configure lines, called *whiskers*, to display all points within 1.5 times the interquartile range (in other words, all points within 1.5 times the width of the adjoining box), or all points at the maximum extent of the data.

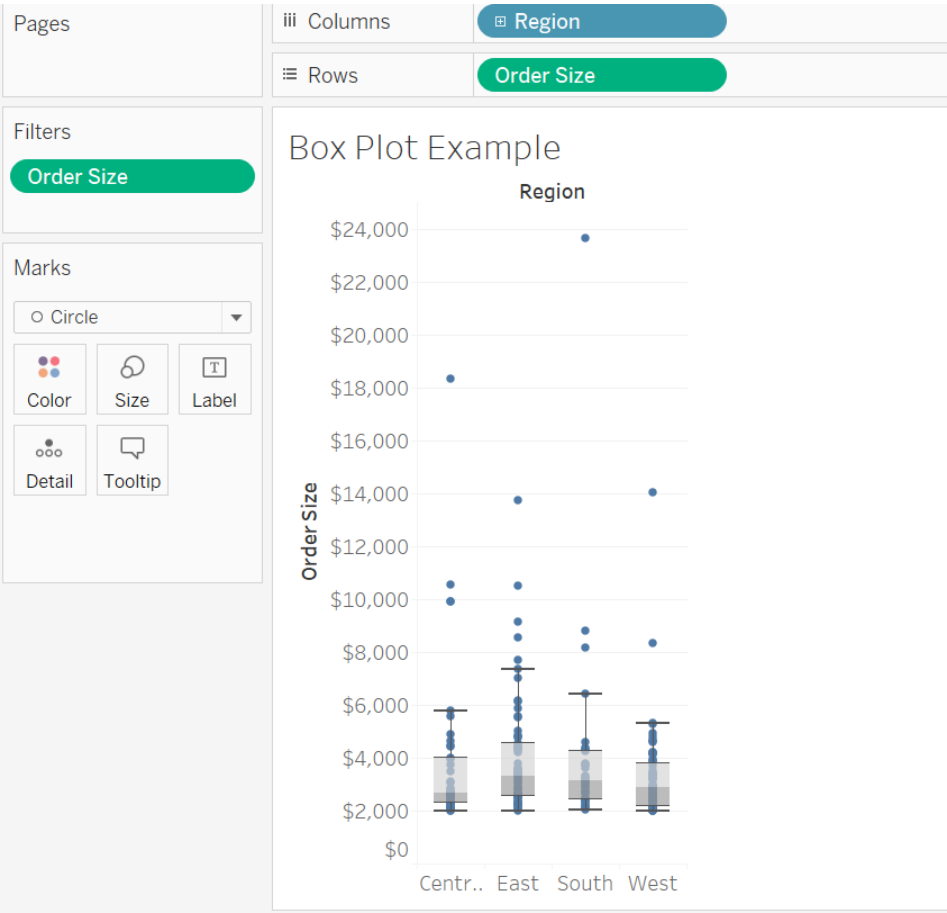
The Analytics Tab has a Box Plot that you can drag out. This is what the previous two examples look like with the Box Plot from the Analytics Tab:



But a better alternative may be to make your own box plot. On a new sheet, drag Region to Columns, and Order Size to Rows. To be consistent with the previous example, again filter Order Size to only values greater than $2,000 and deselect Aggregate Measures.



Then from the Show Me Tab, choose Box and Whisker plot:



Given the quantity of orders that we have in our data set, this is a good visualization to show the median and the outliers.

Another Box Plot Example

To create a box plot that shows discounts by region and customer segment, follow these steps:

1. Using the Sample - Superstore data source, drag the Segment dimension to Columns.

2. Drag the Discount measure to Rows. Make it Average Discount rather than Sum.

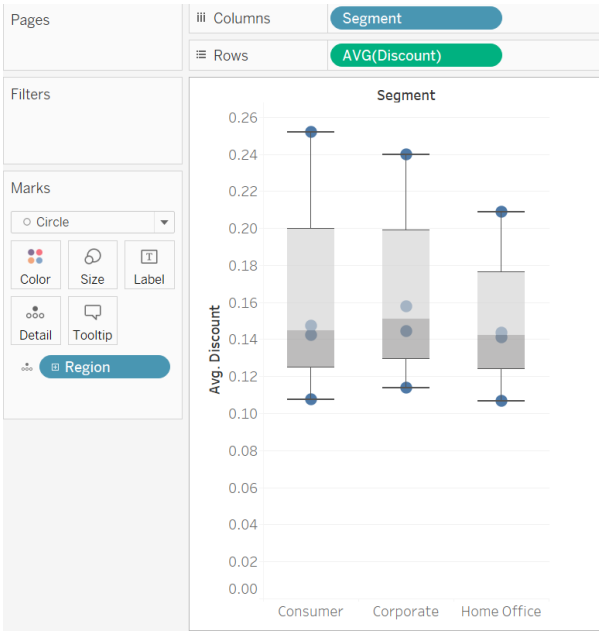
   Tableau creates a vertical axis and displays a bar chart—the default chart type when there is a dimension on the Columns shelf and a measure on the Rows shelf.

3. Drag the Region dimension to Columns, and drop it to the right of Segment.

   Now you have a two-level hierarchy of dimensions from left to right in the view, with regions (listed along the bottom) nested within segments (listed across the top).

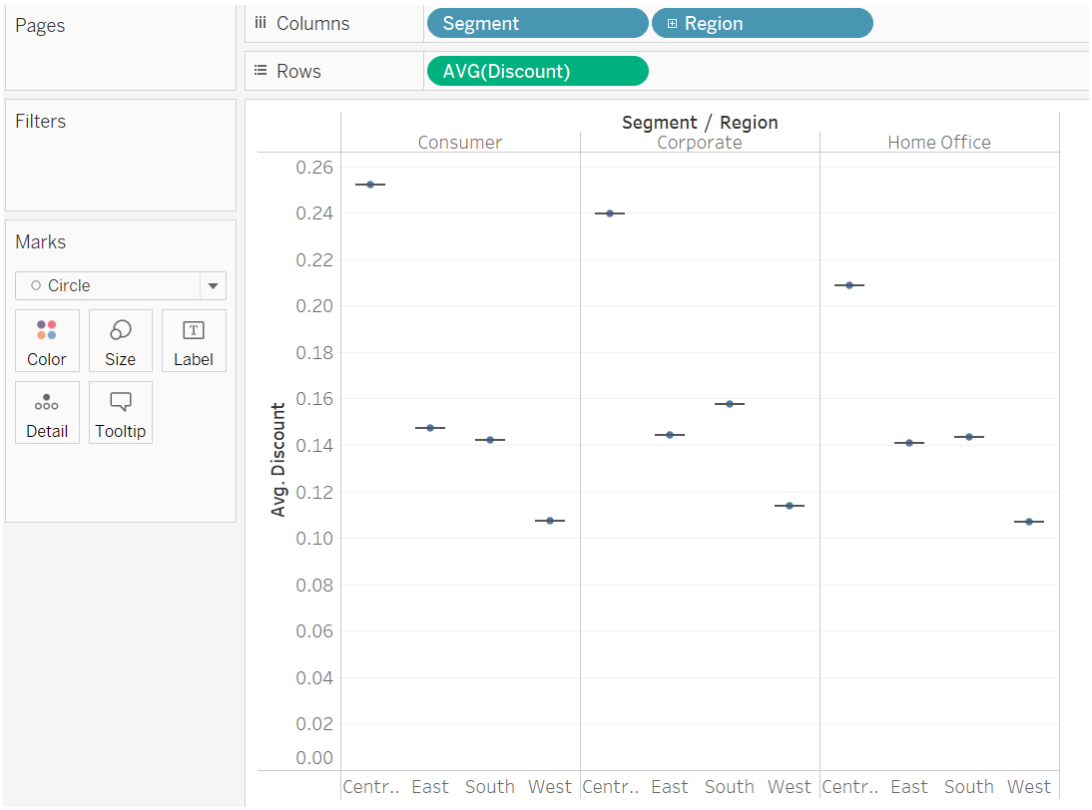
4. Click Show Me in the toolbar, then select the box-and-whisker plot chart type.

Tableau displays a box plot:



Notice that there are only a few marks in each box plot. Also, Tableau reassigned Region from the Columns shelf to the **Marks** card. When you changed the chart type to a box plot, Tableau determined what the individual marks in the plot should represent. It determined that the marks should represent regions. We'll change that.

5. Drag Region from the **Marks** card back to Columns, to the right of Segment.

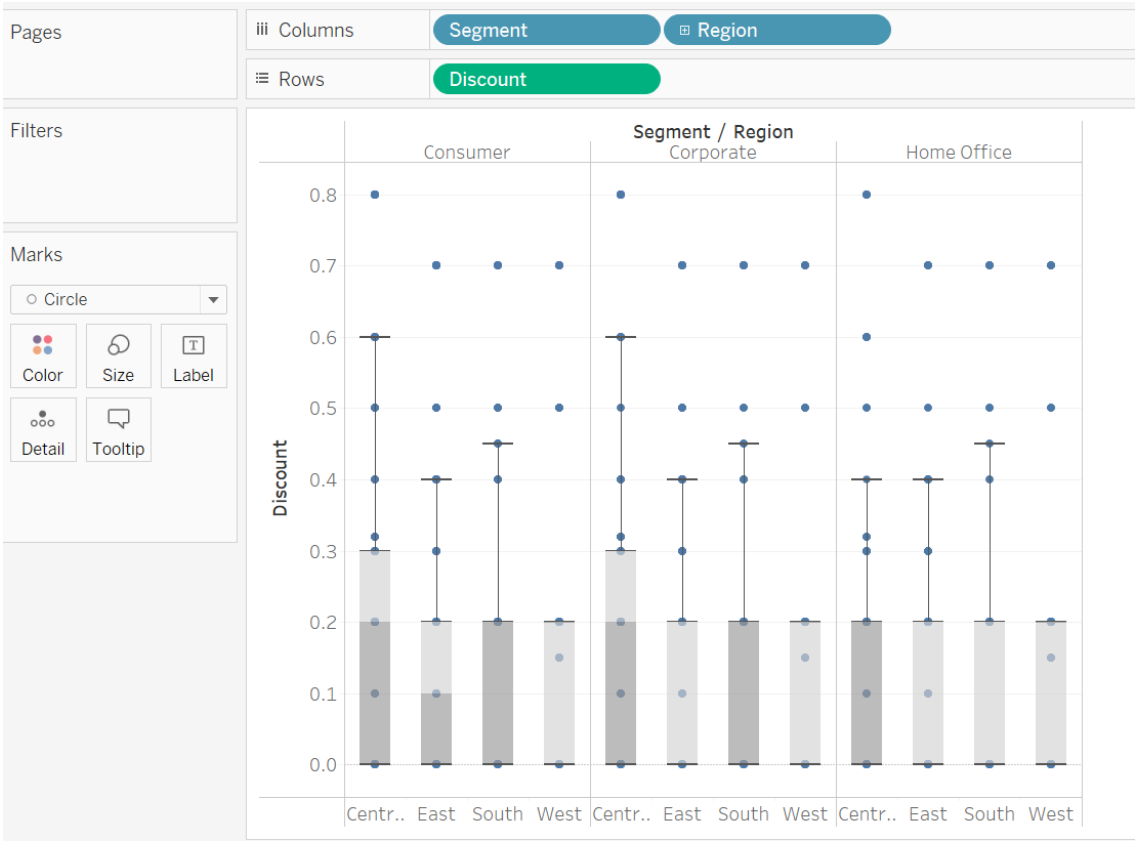


The horizontal lines are flattened box plots, which is what happens when box plots are based on a single mark.

Box plots are intended to show a distribution of data, and that can be difficult when data is aggregated, as in the current view.

6. To disaggregate data, select Analysis > Aggregate Measures.

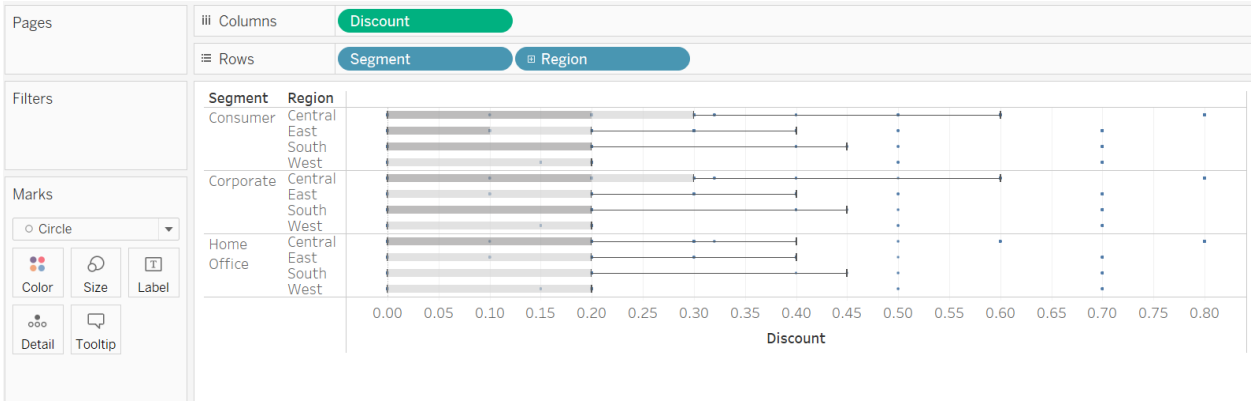
Now, instead of a single mark for each column in the view, you see a range of marks, one for each row in your data source.



The view now shows the information we want to see. The remaining steps make the view more readable and appealing.

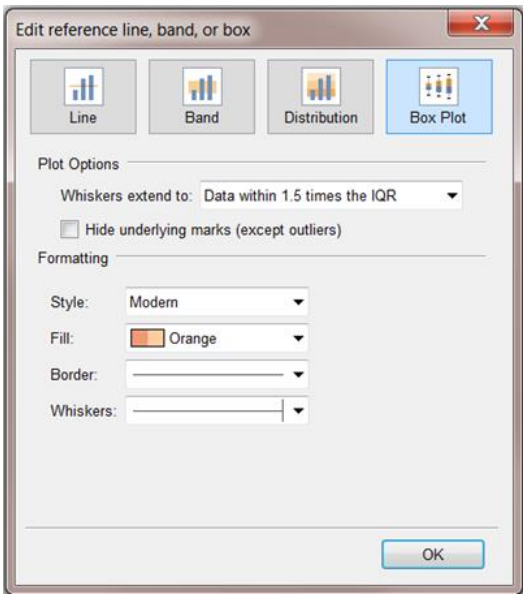
7. Click the Swap button to swap the axes:

The box plots now flow from left-to-right:

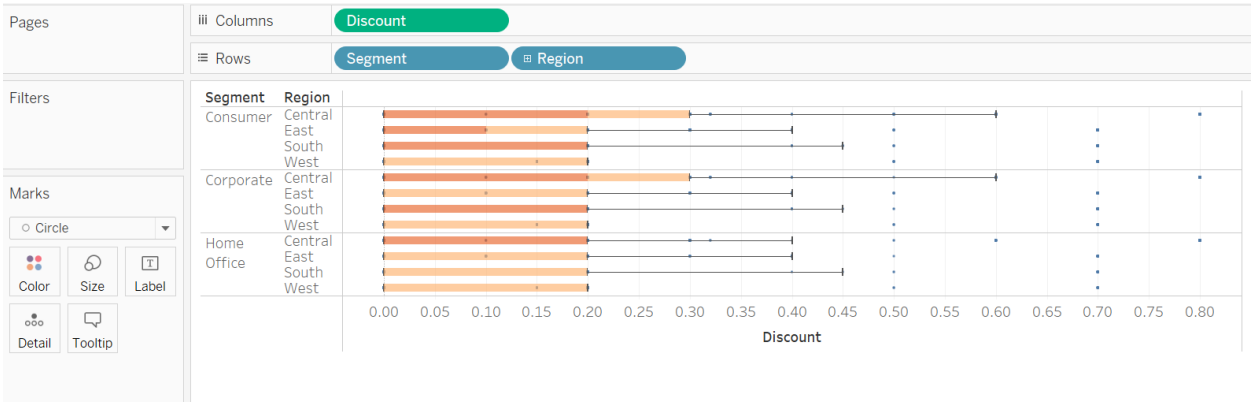


8. Right-click the bottom axis and select Edit Reference Line.

9. In Edit Reference Line, Band, or Box dialog box, in the Fill drop-down list, select an interesting color scheme.



Now your view is complete:



You can see that the discount was the same for all segments in the West. You can also see that the interquartile range (from the 25th percentile to the 75th percentile) for discount was greatest in the Central region for the consumer and corporate segments.

Topic 8: Specialized Charts

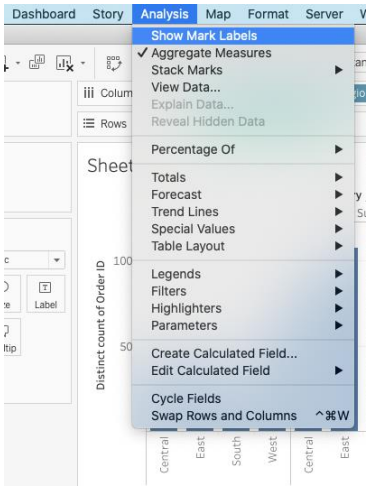
Grouped Charts

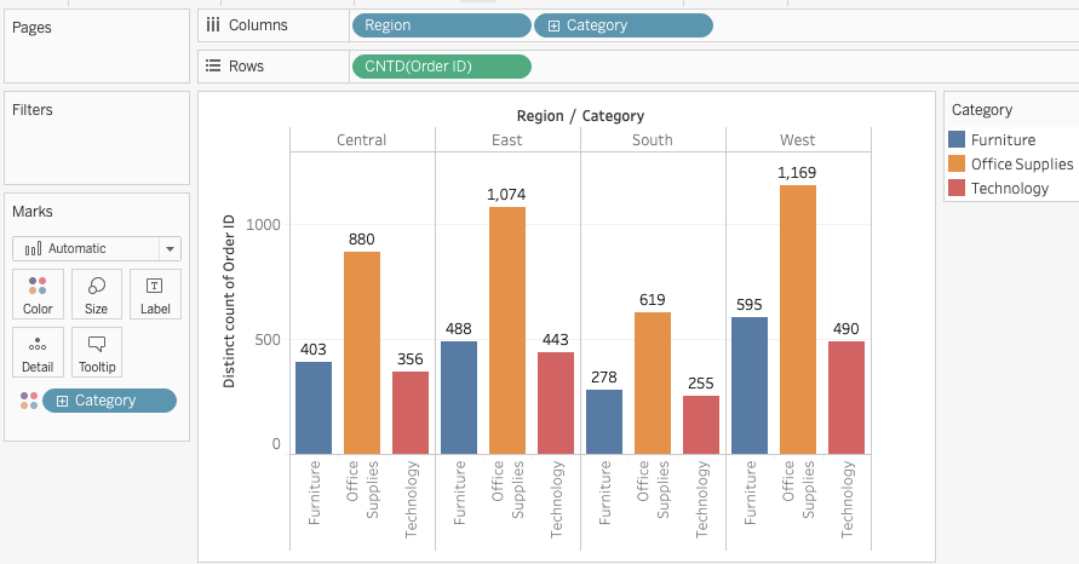
It is helpful to group bar charts (or other types of charts) when you want to compare how a measure differs within a dimension breakout. This requires a dimension to split out the data.

For example, you could use a grouped bar chart to show how the percentage of international students differs by cohort. The following example uses superstore data to determine how the number of orders per category differs between regions.

Drag the Region and Category dimensions to columns shelf. Drag the Order ID dimension to the rows shelf and convert to a count measure. Notice how the bar chart is grouped into panes so that you can now compare counts for each region side by side. Drag the category to the color mark card so that it is easier to compare across regions.

From the Analysis menu, select show mark labels to label the count totals for each bar.





Bullet Graphs

Bullet graphs are a variation of a bar chart invented by Stephen Few. (Optional video by Mr. Few, note: his demo uses an old version of Tableau: http://www.tableau.com/videos/bullet-graphs)

Bullet graphs improve bar charts by providing additional points for comparison. For example, a bullet graph could show last year's sales or a target sales amount. Bullet graphs are a Show Me option in Tableau but the results may not always be exactly what you are expecting. Bullet graphs are just a combination of bars and reference lines.

Since the Superstore dataset does not contain goal information, assume that the comparison will be last year's performance. Assume that 2019 is last year and 2020 is this year. (NOTE: Tableau keeps changing the dates in the Superstore dataset, so use the latest year as this year, and the year previous to the latest as last year.)

Create a Calculated Field for last year: If Year([Order Date]) = 2019 THEN [Sales] END

Last Year's Sales

```
IF  Year([Order Date])=2019
  Then [Sales]
END
```

Repeat for This Year's Sales.

This Year's Sales

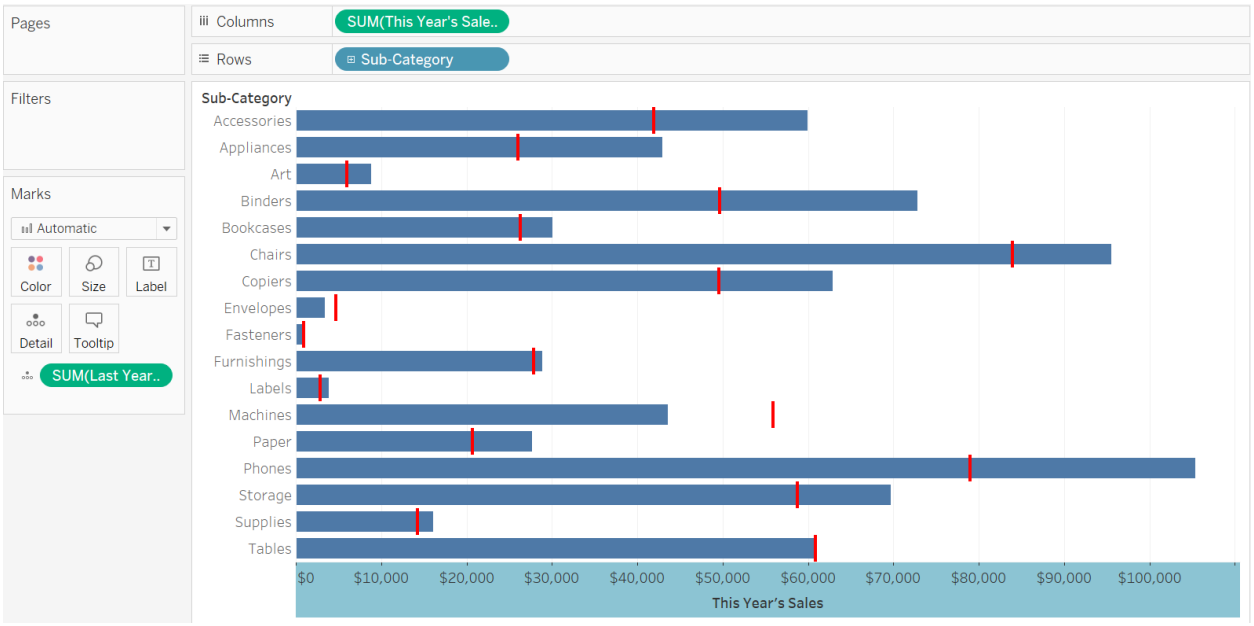
```
IF  Year([Order Date])=2020
  Then [Sales]
END
```

To save yourself some typing, note that you can duplicate a calculated field to create a copy and then modify.

Make a bar chart with This Year's Sales on the Columns Shelf and Sub-Category on the Rows Shelf.

To add last year's sales, drag the Last Year's Sales field to the Details Marks Card. This does not change the view but now the data are available to use as a reference line.

Then, right-click the x-axis and select Add Reference Line. Change Value to Sum(Last Year's Sales), Label to None, and Tooltip to Custom -> Field Name. Also (VERY IMPORTANT), change the Scope radio button from Per Pane to Per Cell. This will give you a reference line for each distinct sub-category. You may also choose to make the line a bolder, heavier weight from the same dialog box.



Small Multiples

Small multiples is a chart visualization tool popularized by Edward Tufte to compare multiple trends side by side on a common scale. They are often used to show trends for different measurements across time and grouped by different traits such as segment or brand.

Well-designed small multiples are:

- Inevitably comparative
- Narrative in content
- Multivariate
- Shrunken, high-density graphics
- Based on a large data matrix
- Efficient in interpretation

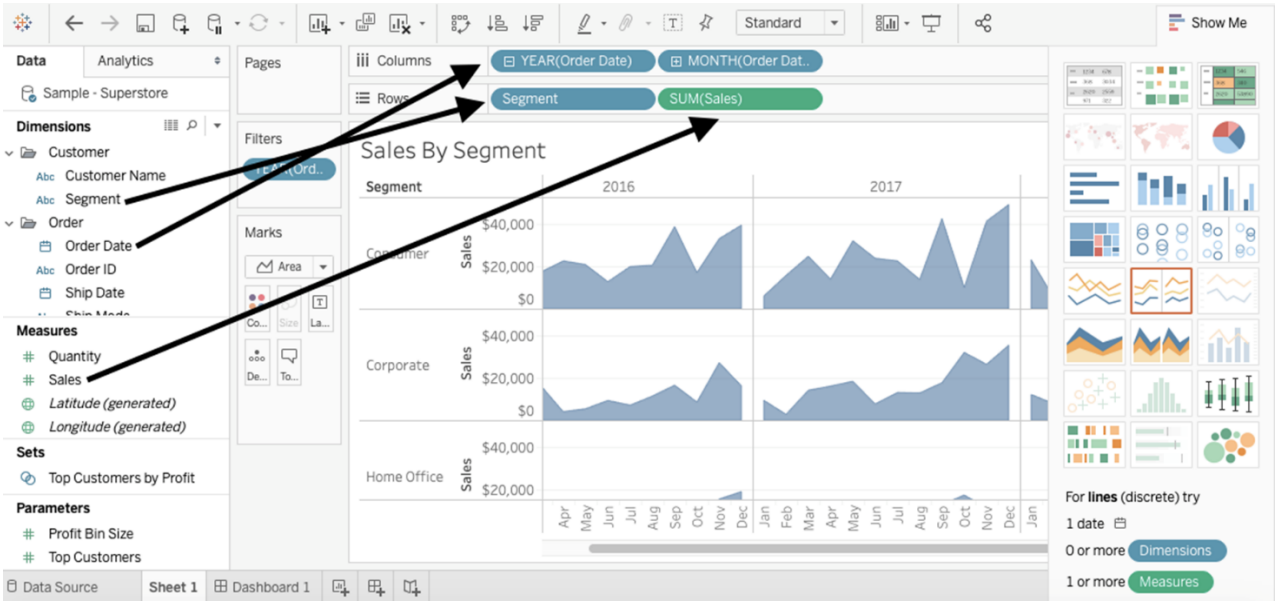
For example, a retail company may want a comparison of how sales for different products are changing throughout time to know which product to discontinue or focus advertising efforts on.

The following example uses the superstore data.

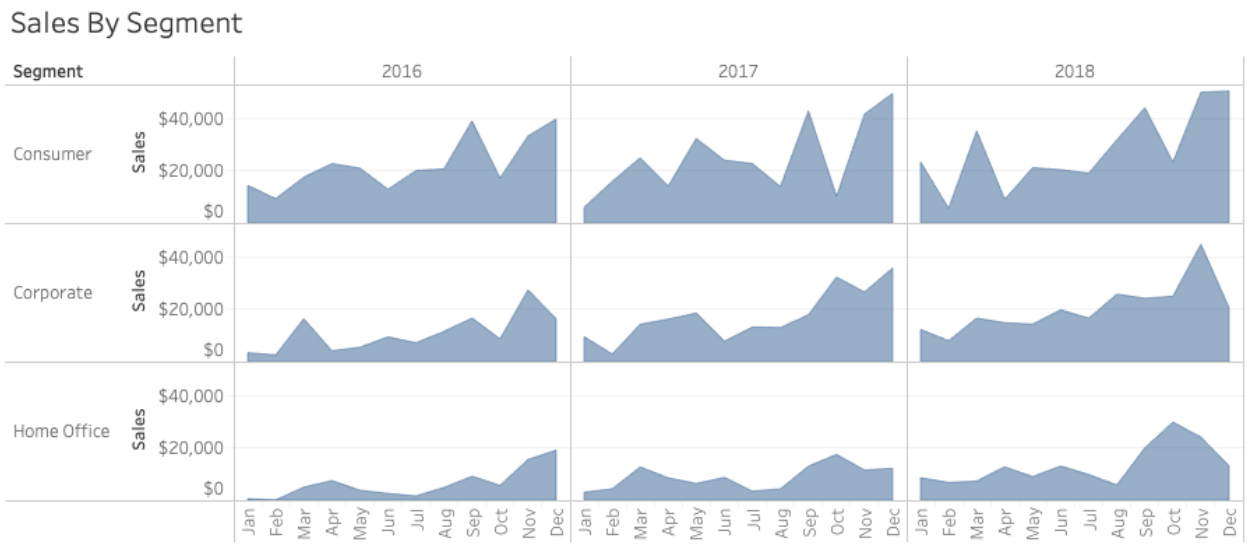
In the following example, we are looking at how sales have changed throughout time for three different business segments of the grocery business.

Steps:

- The ‘Sales’ field will be used as a sum measure in the Rows shelf.
- The ‘Segment’ field will be used a dimension in the Rows shelf to group (or separate) the data.
- The ‘Order Date’ field will be used as a time dimension in the Columns shelf.
- Expand ’Order Date’ to months by clicking on the plus sign next to ‘Year’ in the columns shelf.



The results should look like this (you may need to change the Mark Type to Area to look exactly like this):

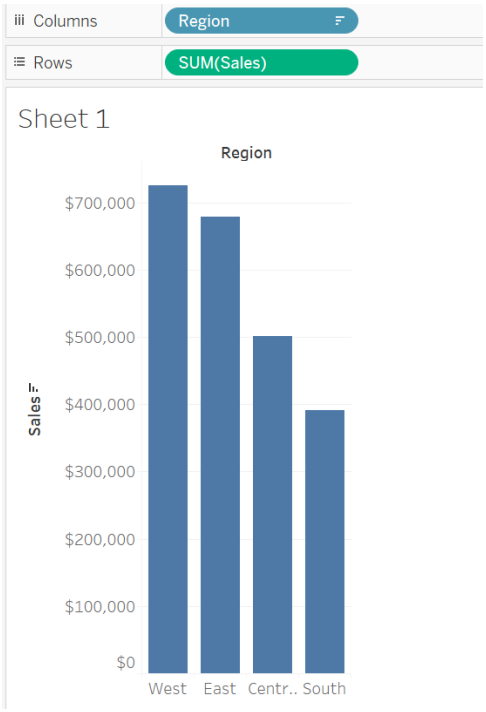


Pareto Charts

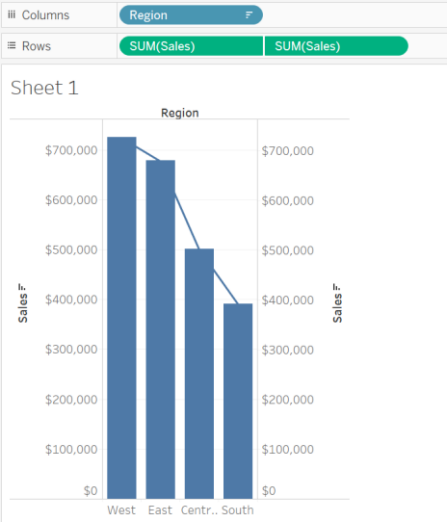
This class demonstration provides a detailed walk-through of how to create a Pareto Chart in Tableau. For a Pareto Chart, on the primary axis, bars are used to show the quantities of a dimension sorted in descending order. The secondary axis shows a line graph with the cumulative total in a percent format. It is traditionally used to identify the biggest contributors to a cumulative total.

Making the Line Portion of the Pareto

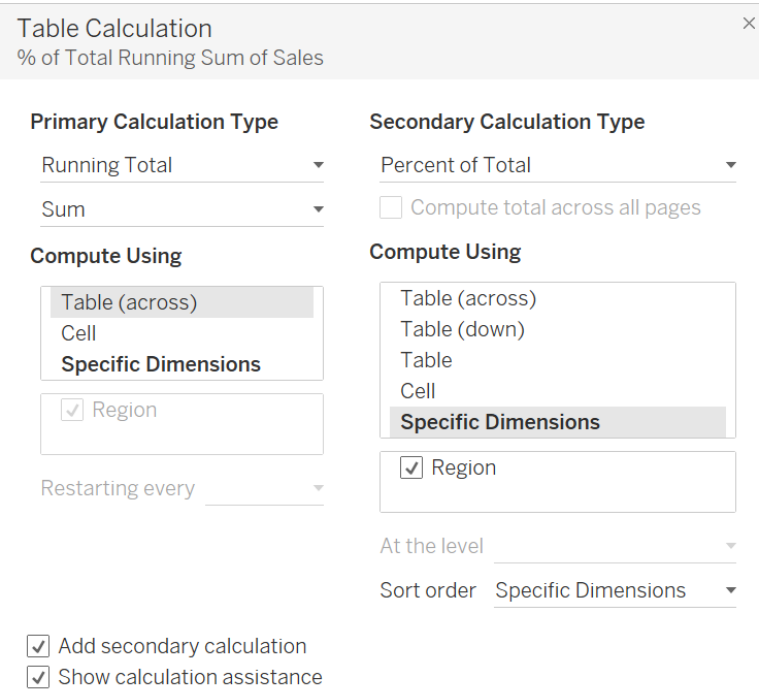
First, you need to create a bar chart that shows Sales (rows) by Region (columns) and sort in descending order:



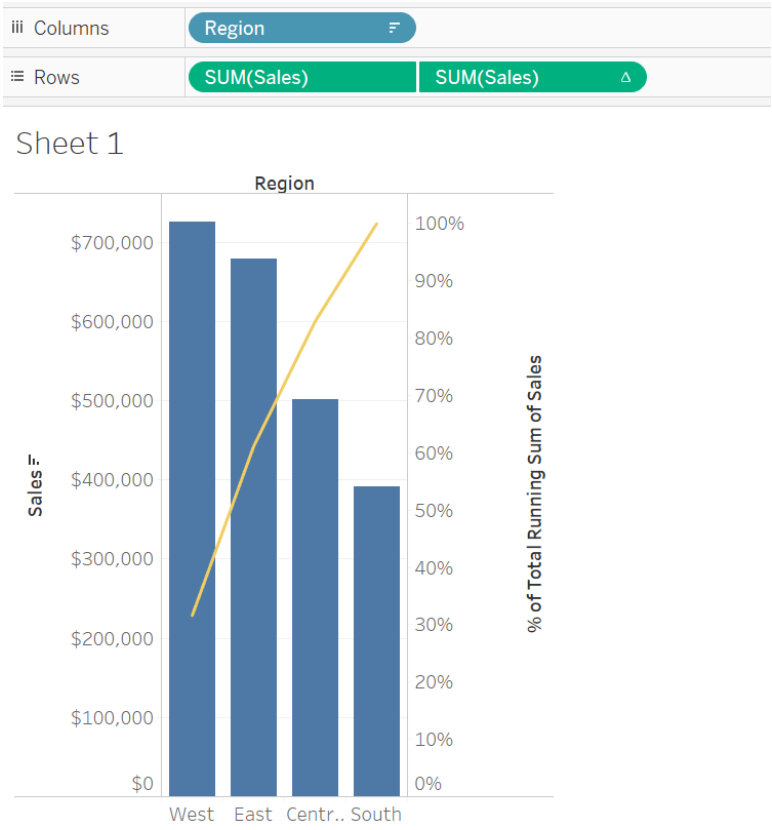
Next, create a dual-axis combination chart by dragging Sales to the right axis, changing the primary axis back to a bar chart, and selecting a line chart for the secondary axis.



Then, you need a table calculation. Click the second sales pill, select quick table calculation, and choose running total. This will be the raw running total for each region. Click the second sales pill again and select edit quick table calculation. Add a secondary calculation for percent total and make it specific to the region. It will appear as cumulative.



You might want to change the colors at this point to help the line stand out.



Bump Charts

What if you wanted to create a line plot, but where the vertical position of the line is determined by the ranking of different sales across markets (rather than the actual sales values). This is called a bump chart. To say it another way, it is used to show changes in rank of a value over some dimension often a time dimension.

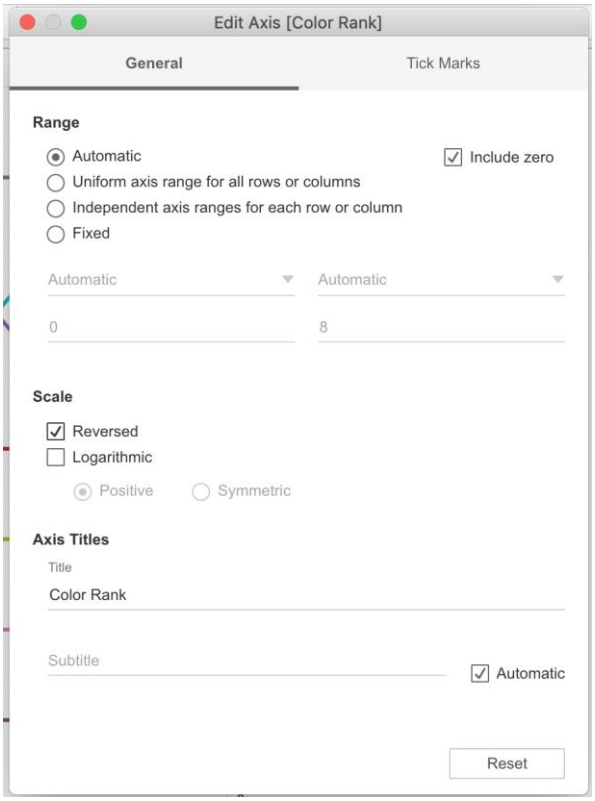
1. Create a new calculated field called Color Rank:

```
RANK_UNIQUE(SUM([SALES]))
```

2. Bring **Region** to Color
3. **Order Date** to Columns
4. **Color Rank** to Rows
5. Change to compute using Region (on **Color Rank** pill)
6. **Color Rank** to Rows again.
7. Change to compute using Region (on Color Rank pill) again.

8. Use the drop-down combo box on the lower Marks Card to change Marks to Circles
9. Size as you wish
10. Right click on Color Rank and choose Dual Axis.

At this point you have a bump chart but it is in reverse order. You can see that if you hover over some of the points. Fix that by reversing BOTH axes. Double-click the axis and select Reversed.



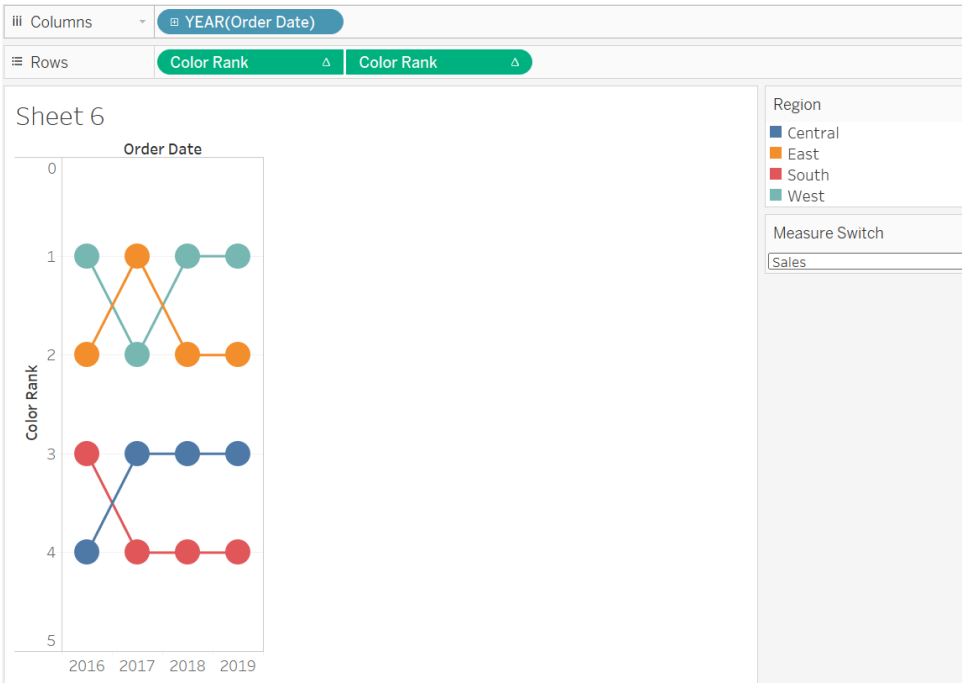
To further customize, bring Sales to size of the circles. Bring Profit to size of the lines. Right click on the right axis and select “Show Header” to get rid of it.

If you want to let the user decide which measure they want to use (as opposed to Sales). You can do that by adding a parameter (similar to an example in Topic 6):

1. New Parameter called “Measure Switch”
2. List Parameter of type String
3. Add Sales, Profit to the list
4. Create a calculated field with a case statement called Selected Measure

```
CASE [Measure Switch]
  WHEN "Sales" THEN [Sales]
  WHEN "Profit" THEN [Profit]
END
```

5. Now you change the **Color Rank** calculation to sum over **Selected Measure**
6. Show Parameter Control and switch between **Profit** and **Sales**



Waterfall Charts

This section provides a detailed walk-through of how to create a Waterfall Chart in Tableau with superstore data.

Creating a Waterfall Chart using the Gantt mark type

A waterfall chart shows the running total of a particular measure (usually with positive and negative values) over a dimension. It shows the relationship of each dimension member against the running total – how each member is contributing to the overall total, either positively or negatively. These can be useful when you want to track something that goes up and down a lot.

1. Build a view with Month(Order Date) on columns and Profit on rows
2. Add a Quick Table Calculation to Profit of Running Total
3. Filter by Year – choose 2017 only
4. Change the Mark to Gantt Bar
5. Create a calculated field called “Negative Profit”: -[Profit]
6. Bring Negative Profit to Size
7. Bring Profit to Color, change to Orange-Blue Diverging

Explanation: The Gantt mark type draws a line at the value of each month’s profit. Putting negative profit on size essentially stretches each Gantt bar’s height to line up with the start of

the previous bar – note that negative bars need to grow upwards, and positive bars need to grow downwards, hence the need to size by *negative* profit to flip that.

Market Basket Analysis

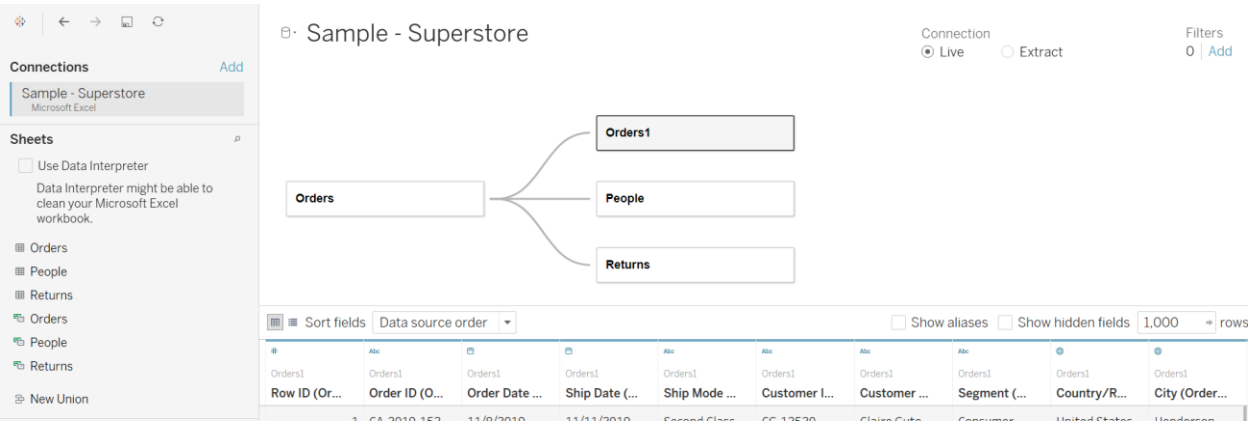
This section provides a detailed walk-through of how to create a market basket analysis in Tableau. This topic uses superstore data.

Market basket analysis looks to see if there are combinations of products frequently co-purchased to drive recommendation engines, marketing campaigns, or placement of items.

To accomplish this in Tableau you need to link orders to a copy of itself.

1. Open the superstore data and go to the Data Source.

2. In the data connection window, create a relationship between Orders and itself by dragging out the Orders table twice. You can remove People & Returns or leave it (either is fine).



3. You will be prompted to edit the relationship and set: Customer Name = Customer Name

    - This is the level of detail for the analysis—Customer. If you want to know which individual *orders* had which combinations of sub-categories then use Order ID instead of Customer Name

4. In the worksheet, bring Sub-Category from the first table to Rows and Sub-Category from the second table to Columns.

5. Add Count (Distinct) of Customer Name to Color, click Label and check Show Mark Labels.

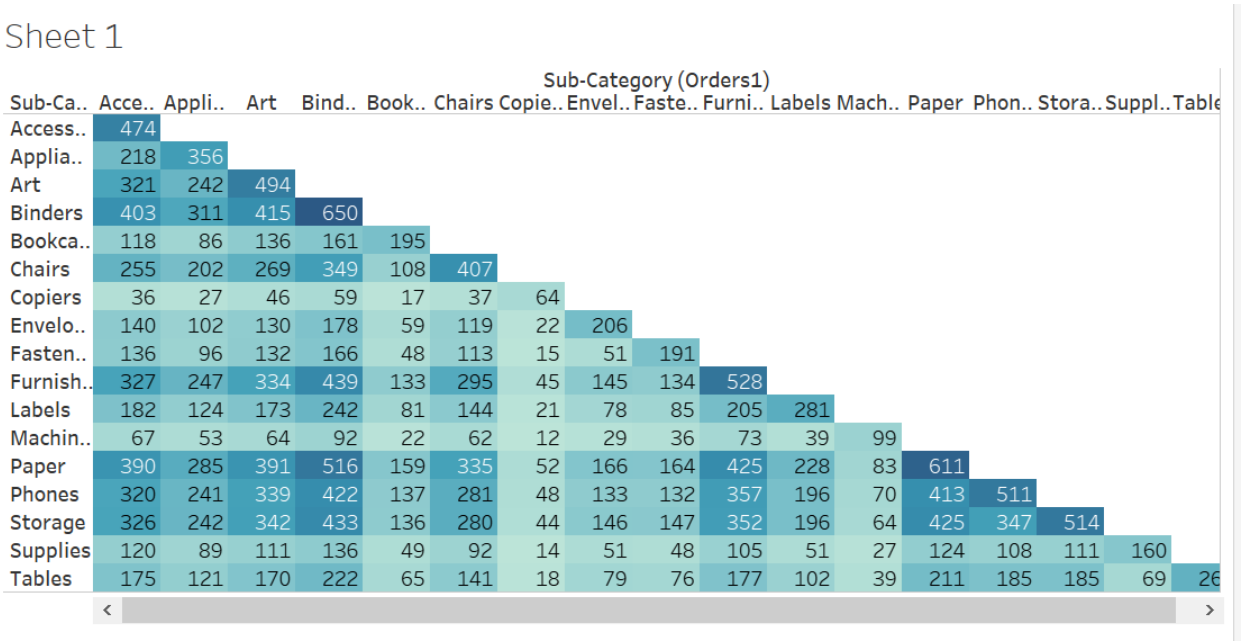
    - Again, Customer is the level of analysis here. If you wanted to know the number of orders that included both sub-categories you would count the unique number of Order IDs.

This gives us a full matrix. To remove the mirrored upper half of the matrix, create the Display Sub-Category giving us only the desired bottom triangle. To do this, create a calculated field: Display Sub-Category

isplay Sub-Category

```
IF [Sub-Category]< [Sub-Category (Orders1)]
 THEN 1
 ELSE 0
END
```

Add this to a filter. You can now format the filter as you like.



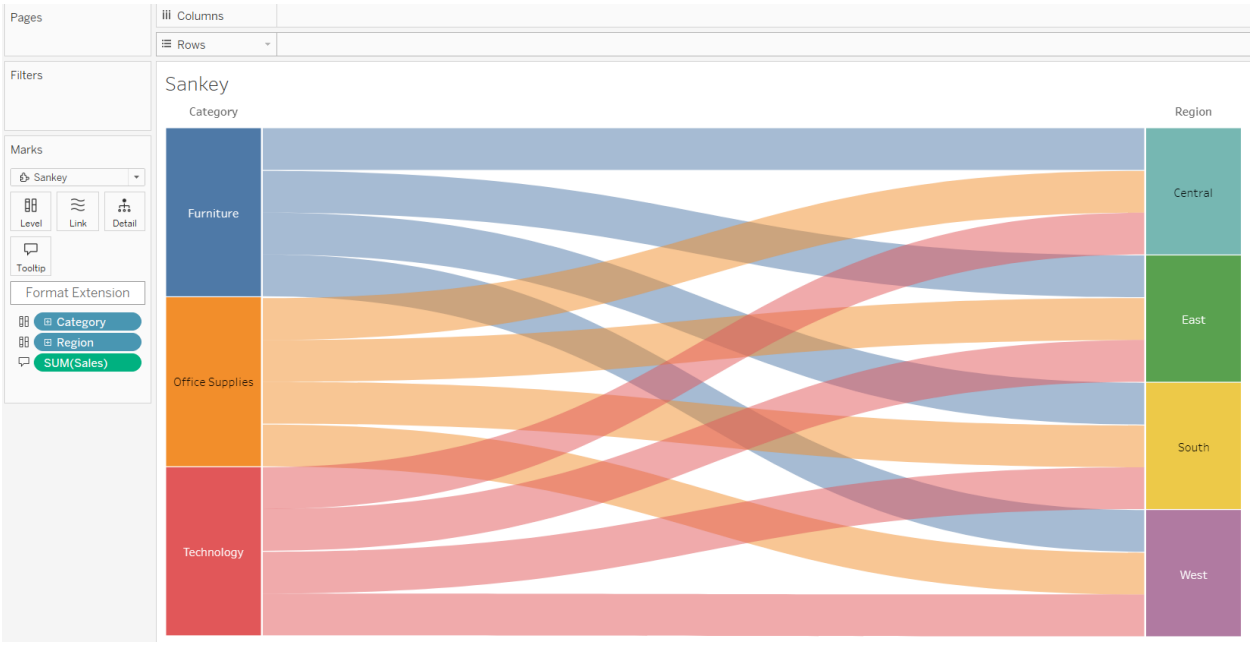
Sub-Ca..	Acce..	Appli..	Art	Bind..	Book..	Chairs	Copie..	Envel..	Faste..	Furni..	Labels	Mach..	Paper	Phon..	Stora..	Suppl..	Table
Access..	474																
Applia..	218	356															
Art	321	242	494														
Binders	403	311	415	650													
Bookca..	118	86	136	161	195												
Chairs	255	202	269	349	108	407											
Copiers	36	27	46	59	17	37	64										
Envelo..	140	102	130	178	59	119	22	206									
Fasten..	136	96	132	166	48	113	15	51	191								
Furnish..	327	247	334	439	133	295	45	145	134	528							
Labels	182	124	173	242	81	144	21	78	85	205	281						
Machin..	67	53	64	92	22	62	12	29	36	73	39	99					
Paper	390	285	391	516	159	335	52	166	164	425	228	83	611				
Phones	320	241	339	422	137	281	48	133	132	357	196	70	413	511			
Storage	326	242	342	433	136	280	44	146	147	352	196	64	425	347	514		
Supplies	120	89	111	136	49	92	14	51	48	105	51	27	124	108	111	160	
Tables	175	121	170	222	65	141	18	79	76	177	102	39	211	185	185	69	26

Sankey Chart

Tableau added a Sankey Chart as part of their Viz extensions now available in version 2024.2. This example requires this latest version.

In the Marks card, chose Sankey. If Sankey is not available, the last item says “Add Extension.” Then choose the Tableau developed Sankey extension. You can review other Extensions available (Radial charts, Network charts) but realize that some may cost money to use.

To build the visualization with this extension, assign two categorical dimensions to the Level encoding box and assign one measure to the Link encoding box.

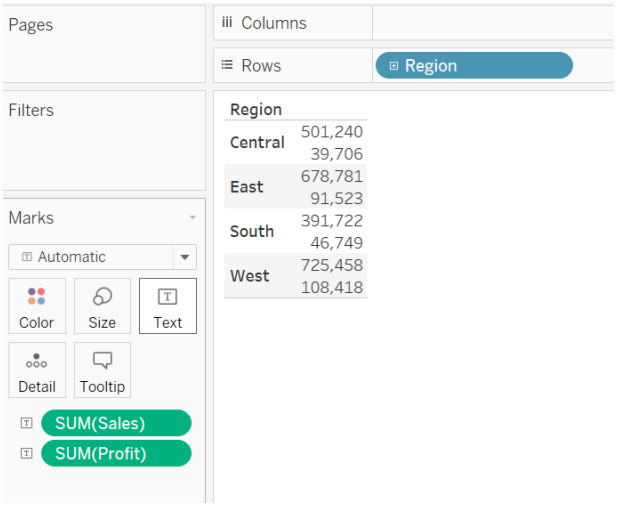

Pages
Columns
Rows
Filters
Marks
Sankey
Level
Link
Detail
Tooltip
Format Extension
Category
Region
SUM(Sales)
Sankey
Category
Region
Furniture
Office Supplies
Technology
Central
East
South
West

Topic 9: Other Analytic Tools

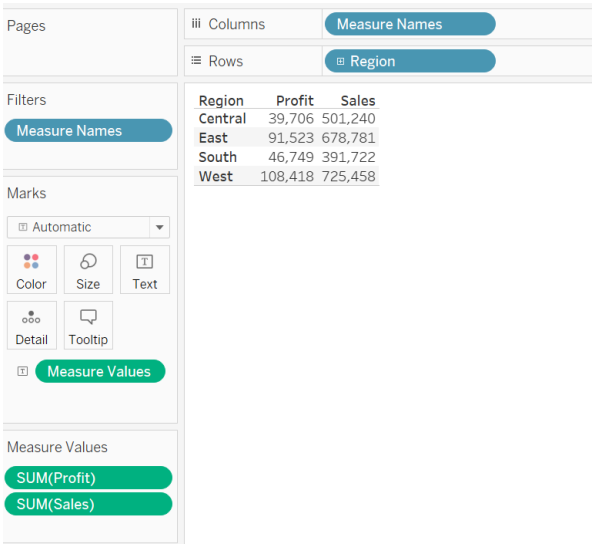
This section provides more details on text in tables, an overview of additional items in the Analytics pane, and examples of other useful analytic calculations. This topic uses superstore data.

Text in Tables (Measure Names and Measure Values)

Drag Region to Rows and Profit and Sales to the Text Marks Card. Your text table could look like this which is not very helpful since it is hard to compare and not properly labeled.



If instead of dragging Profit and Sales to the Text Marks Card, you double-clicked on each field, you would create a different text table:



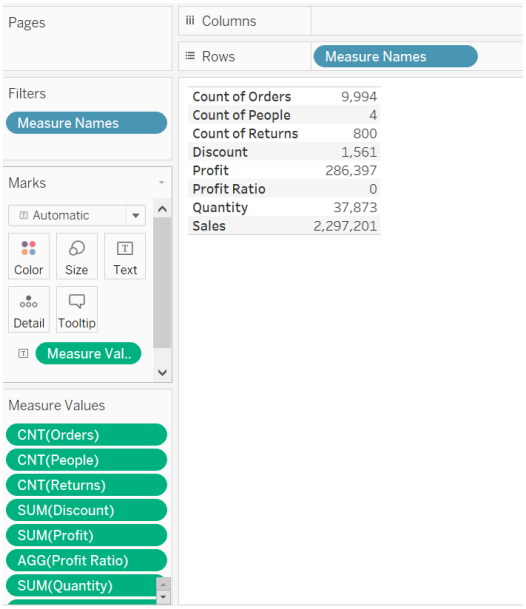
This text table is easier to understand and is well labeled. Also note that the Column pill shows “Measure Names” and the Text Marks Card shows “Measure Values.”

Tableau automatically creates these fields so that you can build certain types of views that involve multiple measures.

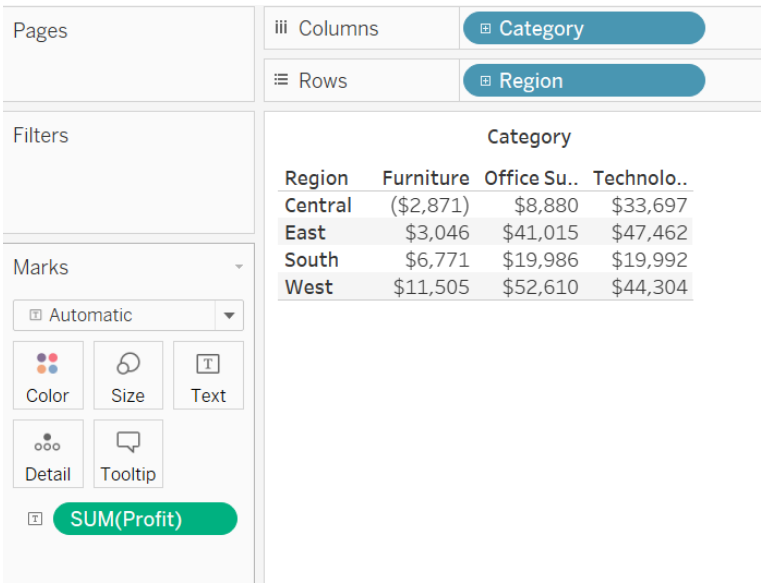
- The Measure Values field contains all the measures in your data, collected into a single field with continuous values.

- The Measure Names field contains the names of all measures in your data, collected into a single field with discrete values.

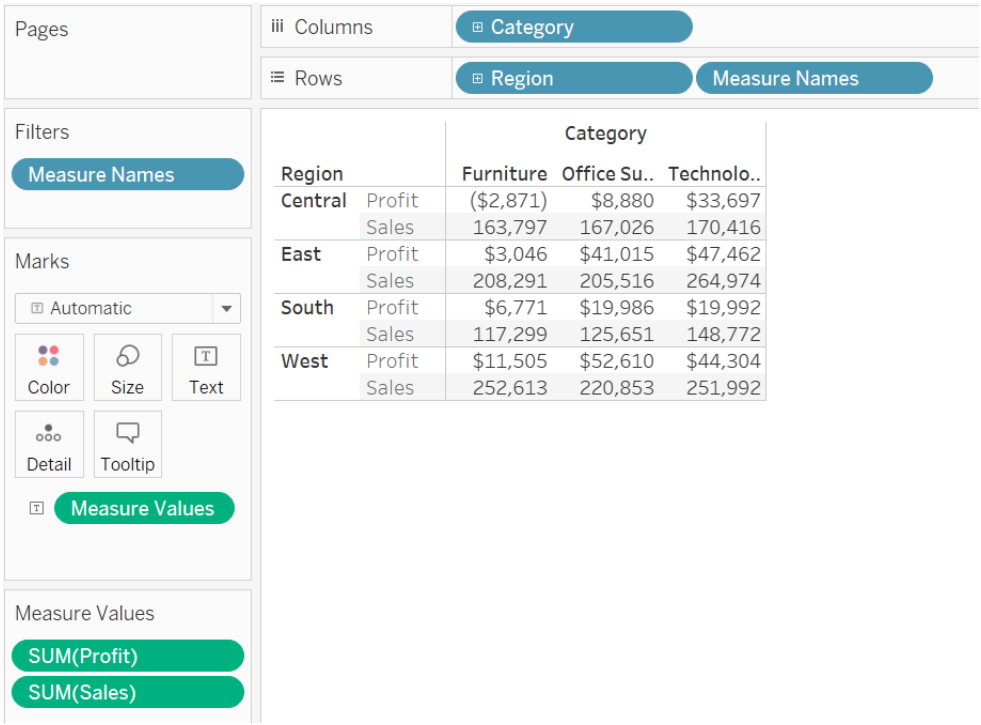
Drag Measure Names to the Row shelf and Measure Values to the Text Marks Card. When Measure Values is in the view, Tableau creates a Measure Values card that lists the measures in the data source with their default aggregations. You can remove individual measures from the view by dragging them out of the Measure Values card.



When you want to show multiple measures in a view, you can use the Measure Values and the Measure Names fields. When you add Measure Names to a view, all of the measure names appear as row or column headers in the view. The headers include each measure name. This feature becomes useful when you are working with a text table that shows multiple measures. For example, suppose you have a text table containing the aggregated profit of each product category by region.



Now suppose you want to show both the Profit and the Sales for each category and region. When you add Sales to the text table (by dragging it and dropping it in the view), the measures are combined and the Measure Values field is added to Text. The Measure Names field is automatically added to the Rows shelf.

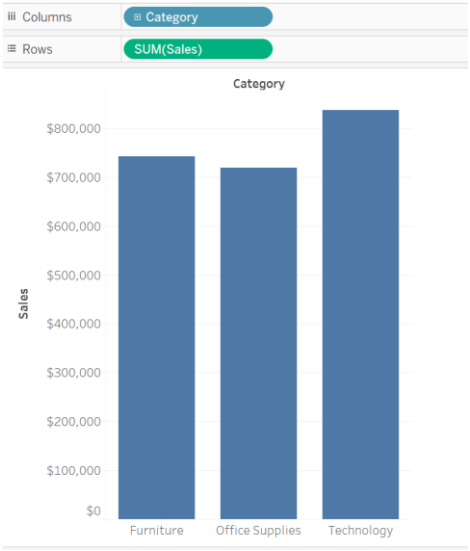


You may want to include the aggregations or call it "Total Profit" and "Total Sales." To change the measure names, right-click the Measure Names field on the Rows shelf and select Edit Aliases. Make the changes and click OK.

Region		Category Furniture	Office Supplies	Technology
Central	Total Profit	($2,871)	$8,880	$33,697
	Total Sales	163,797	167,026	170,416
East	Total Profit	$3,046	$41,015	$47,462
	Total Sales	208,291	205,516	264,974
South	Total Profit	$6,771	$19,986	$19,992
	Total Sales	117,299	125,651	148,772
West	Total Profit	$11,505	$52,610	$44,304
	Total Sales	252,613	220,853	251,992

Create a visualization using Measures Names and Measure Values

By now, everyone knows how to make a chart of Sales by product category.



But assume for some reason, the Sales by product category where individual calculations that you made from your data.

Furniture Sales

```
If [Category] = "Furniture" Then
 [Sales]
END
```

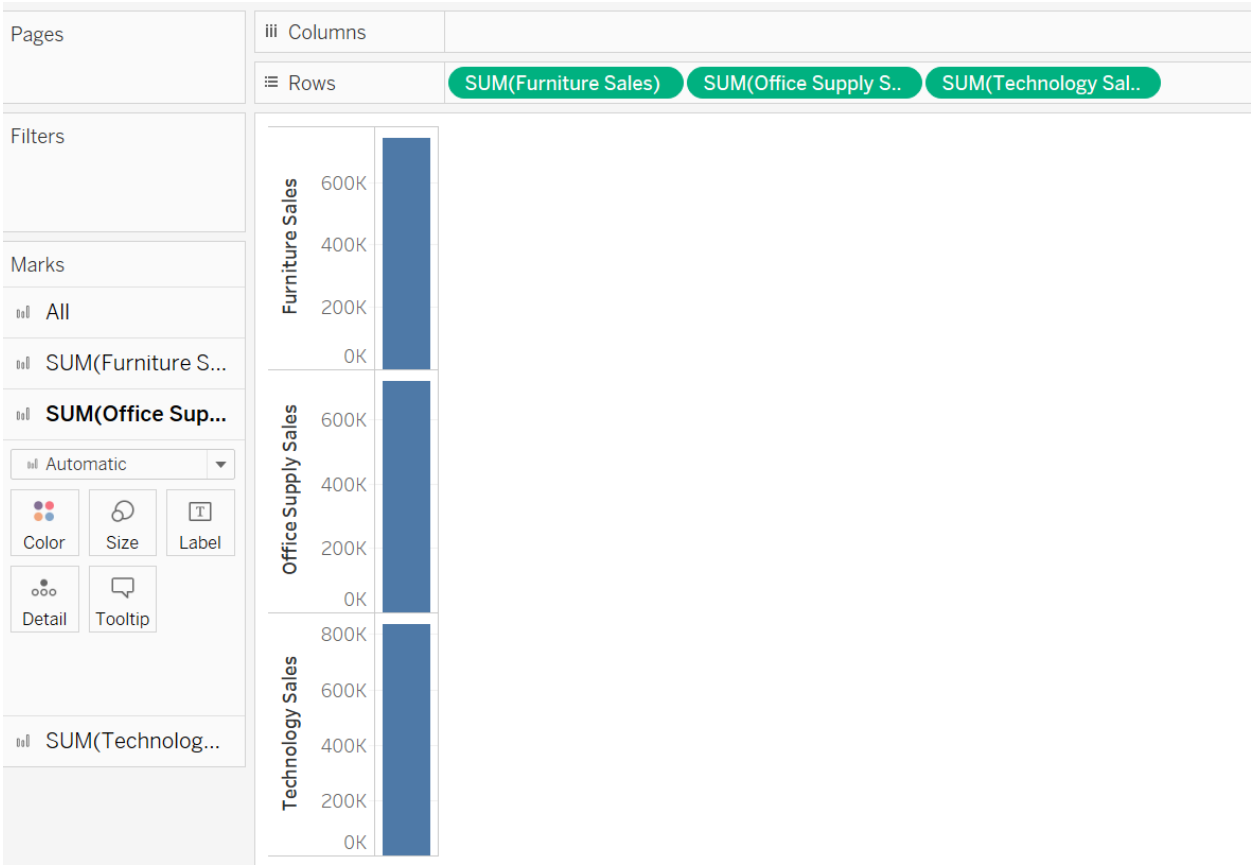
Office Supply Sales

```
If [Category] = "Office Supplies" Then
 [Sales]
END
```

Technology Sales

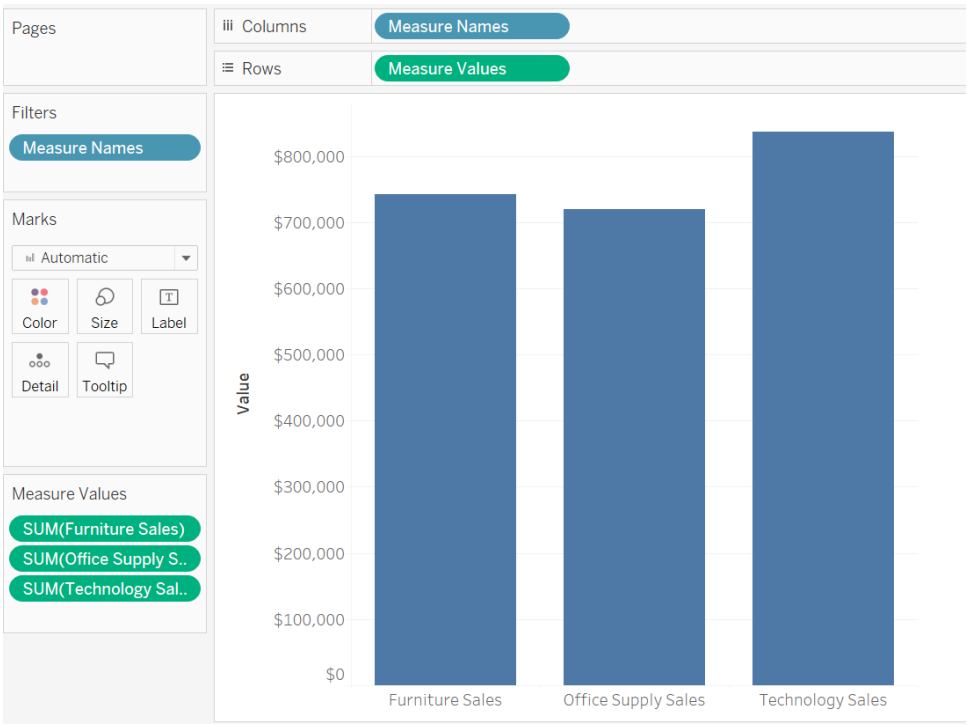
```
If [Category]="Technology" THEN
 [Sales]
END
```

If you try to make a visualization of these three calculated fields, the visualization will make the vertical and separate rather than side-by-side.



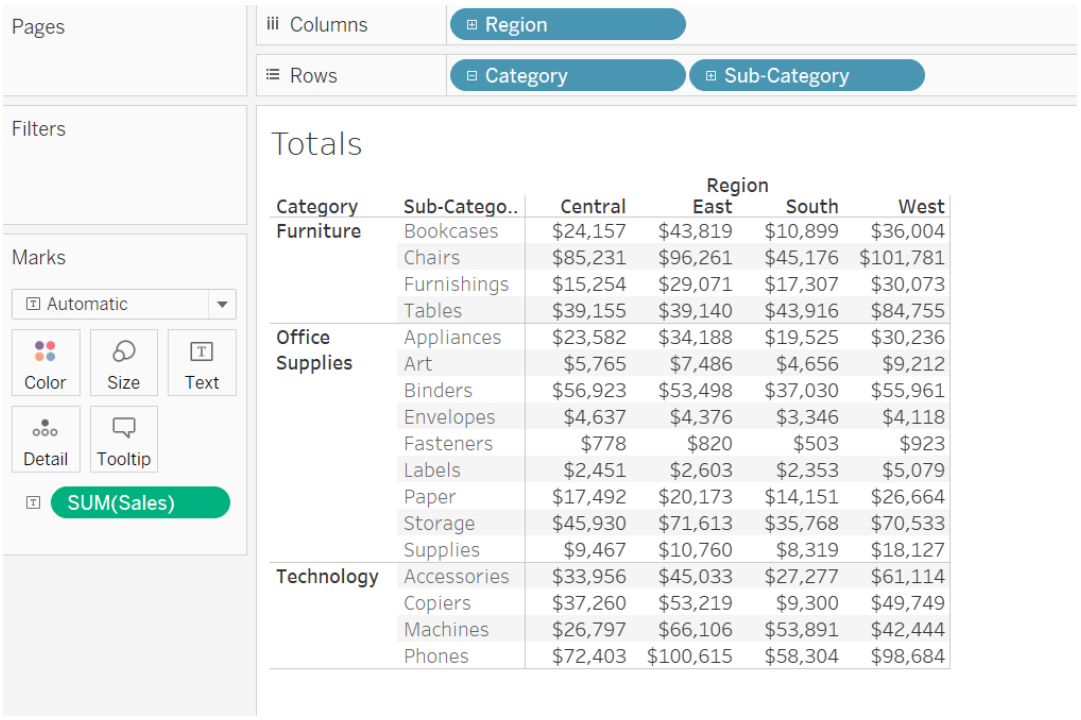
You could try to make them a dual axis but that will only work for 2 not 3 fields and it has its own problems with spacing.

Instead add the Measure Names to the Columns shelf and the Measure Values to the Rows shelf and only keep the three Measure Values in the Measure Values card that you want to show. While this example made a simple chart more difficult, the point was to demonstrate when you have multiple calculated fields of your choosing from the data.



From the Analytics Tab: Totals (Grand Totals, Subtotals)

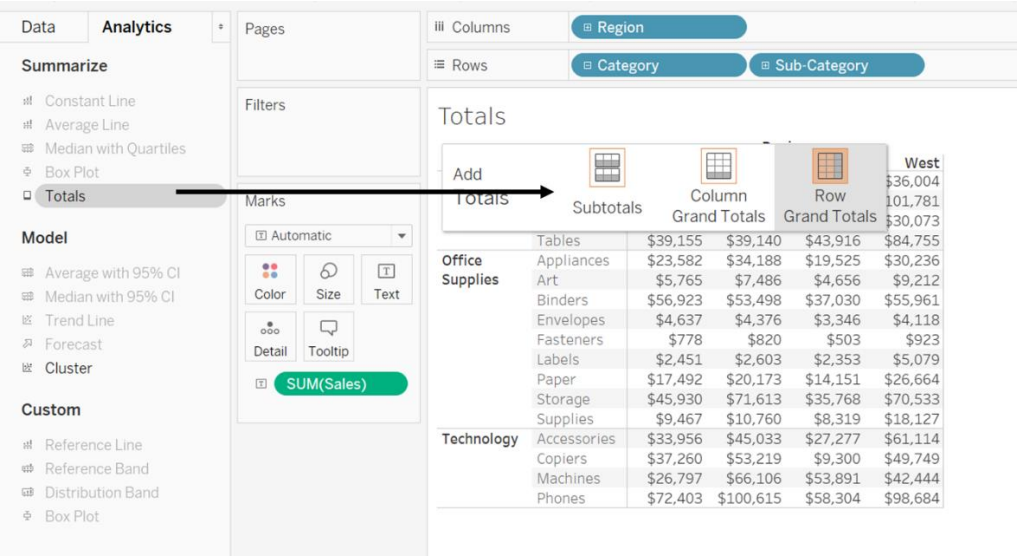
Set up a table with Region in Columns, Category and Sub-Category in Rows, and the Sum of Sales as Text for the Table.



Totals

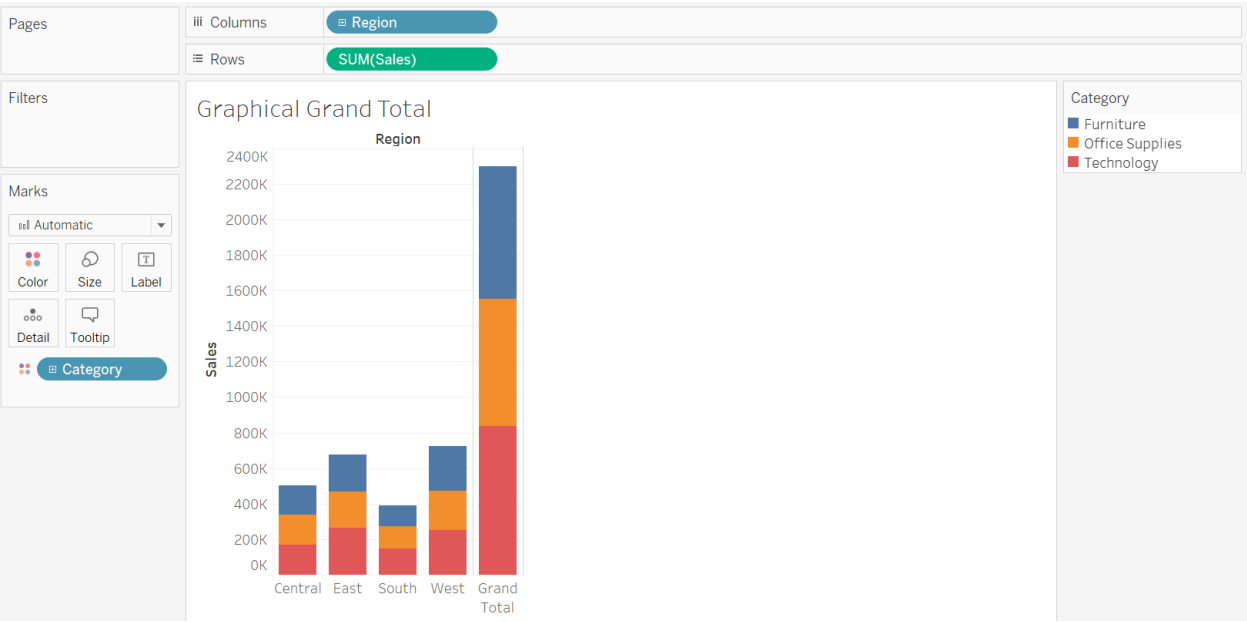
Category	Sub-Catego..	Region Central	East	South	West
Furniture	Bookcases	$24,157	$43,819	$10,899	$36,004
	Chairs	$85,231	$96,261	$45,176	$101,781
	Furnishings	$15,254	$29,071	$17,307	$30,073
	Tables	$39,155	$39,140	$43,916	$84,755
Office Supplies	Appliances	$23,582	$34,188	$19,525	$30,236
	Art	$5,765	$7,486	$4,656	$9,212
	Binders	$56,923	$53,498	$37,030	$55,961
	Envelopes	$4,637	$4,376	$3,346	$4,118
	Fasteners	$778	$820	$503	$923
	Labels	$2,451	$2,603	$2,353	$5,079
	Paper	$17,492	$20,173	$14,151	$26,664
	Storage	$45,930	$71,613	$35,768	$70,533
	Supplies	$9,467	$10,760	$8,319	$18,127
Technology	Accessories	$33,956	$45,033	$27,277	$61,114
	Copiers	$37,260	$53,219	$9,300	$49,749
	Machines	$26,797	$66,106	$53,891	$42,444
	Phones	$72,403	$100,615	$58,304	$98,684

In the Analytics pane, under Summarize, drag Totals into the Add Totals dialog, and drop it over either the Row Grand Totals or Column Grand Totals option.

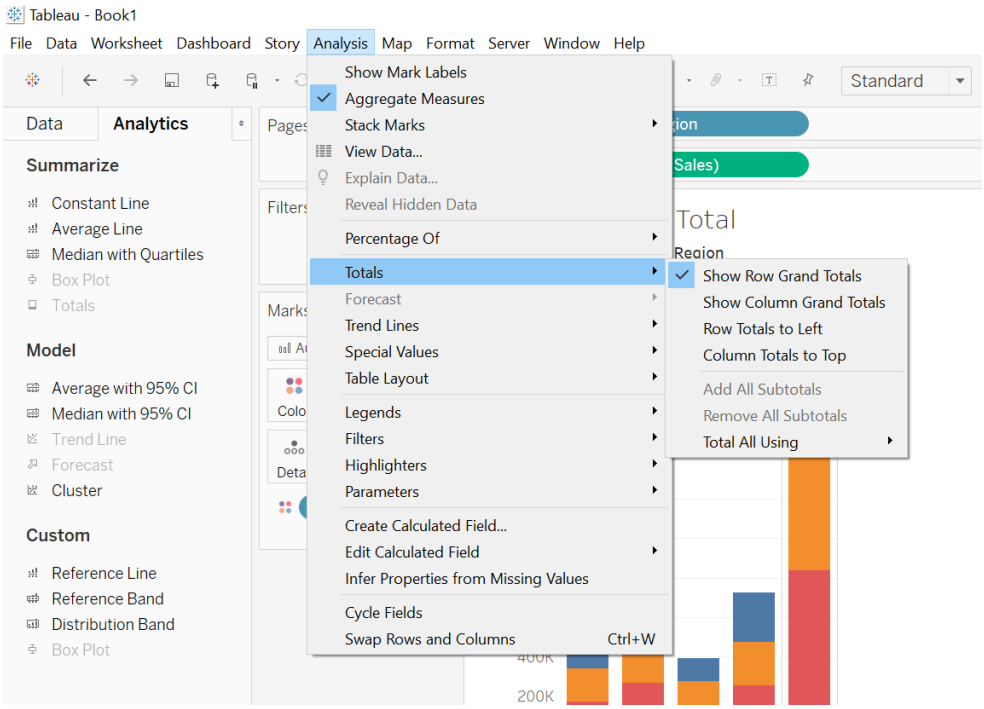


Row grand totals appear automatically on the right-side of the visualization. Column grand totals appear automatically at the bottom of the visualization.

Use can use the same process to add totals to some graphical views too.



There are numerous options associated with Totals that can be adjusted by selecting Totals from the Analysis menu.



Bar Charts with Error Bars

You have created many bar charts so far, but what if you want to show error bars (to support analysis that sales in two regions are statistically different). You should recall from statistics that the standard error of the mean is a measure of the dispersion of sample means around the population mean. It is the standard deviation of the 'scores' per condition divided by the square root of the number of participants in that condition.

If we want to look at the average size of orders by region, we first need the order size (previously found in Topic 5).

Order Size

```
{Fixed [Order ID]: Sum([Sales])}
```

Then we want the standard error of Order Size:

Standard Error

```
STDEV([Order Size])/sqrt(COUNTD([Order ID]))
```

Once you have the standard error, you can create new fields for the upper and lower standard error limits with:

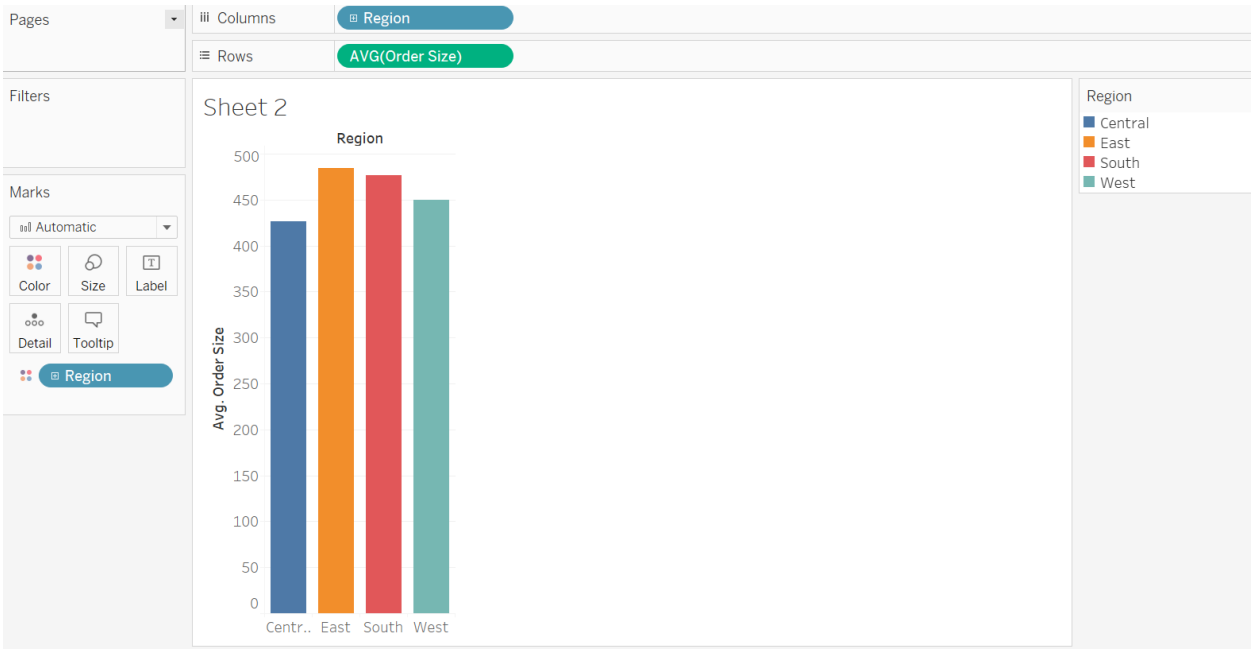
Upper Std Error

```
AVG([Order Size])+[Standard Error]
```

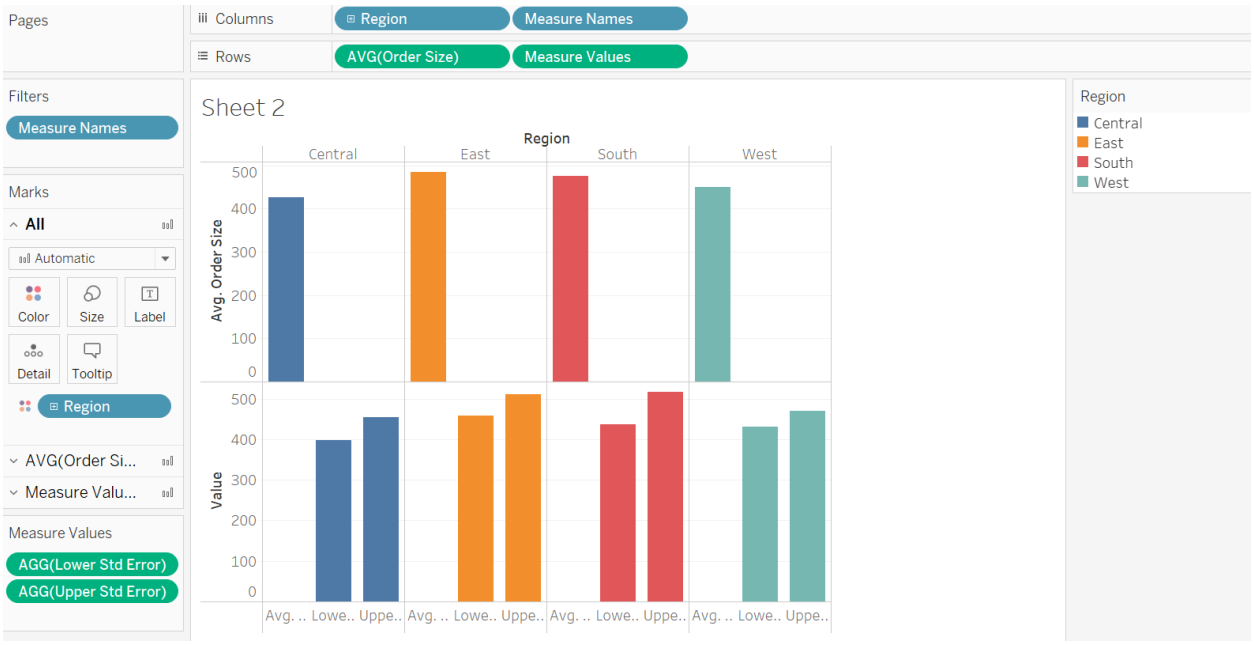
Lower Std Error

```
AVG([Order Size])-[Standard Error]
```

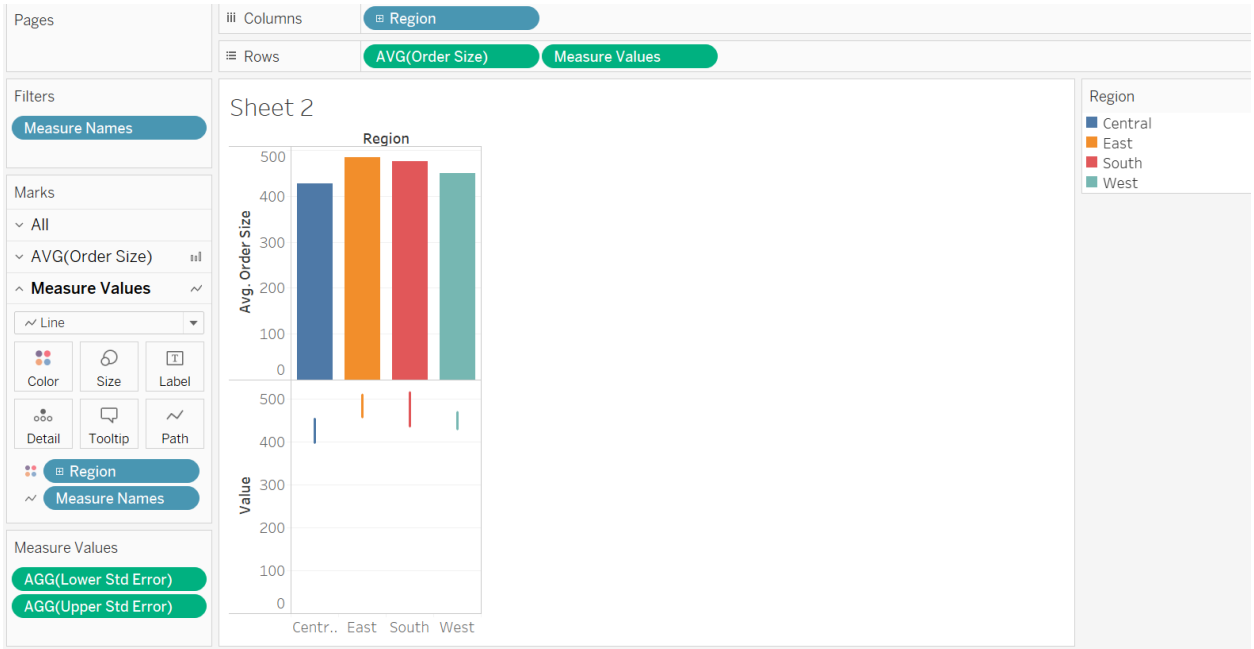
To create a visualization with standard error bars, first create your bar chart for Order Size by Region.



Then drag the lower standard error field onto rows to create a separate graph. Drag the upper standard error field onto the same axis of the new graph to set up a measures names/measures values situation:



Then for the ‘Measures Values’ marks (i.e., the lower set of marks), switch the measures value mark type to line and drag measures names from columns to the path card.

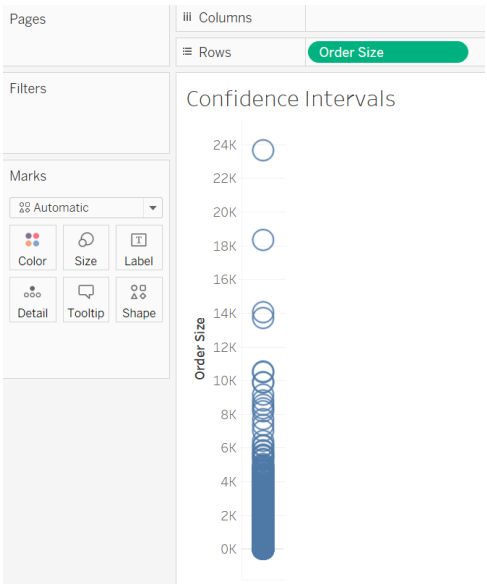


Then create a dual axis graph, synchronize the axes and remove region from color on the standard error lines.

Now you have a bar graph with standard error bars. The performance in the Central Region versus the East Region are significantly different based on one standard error.

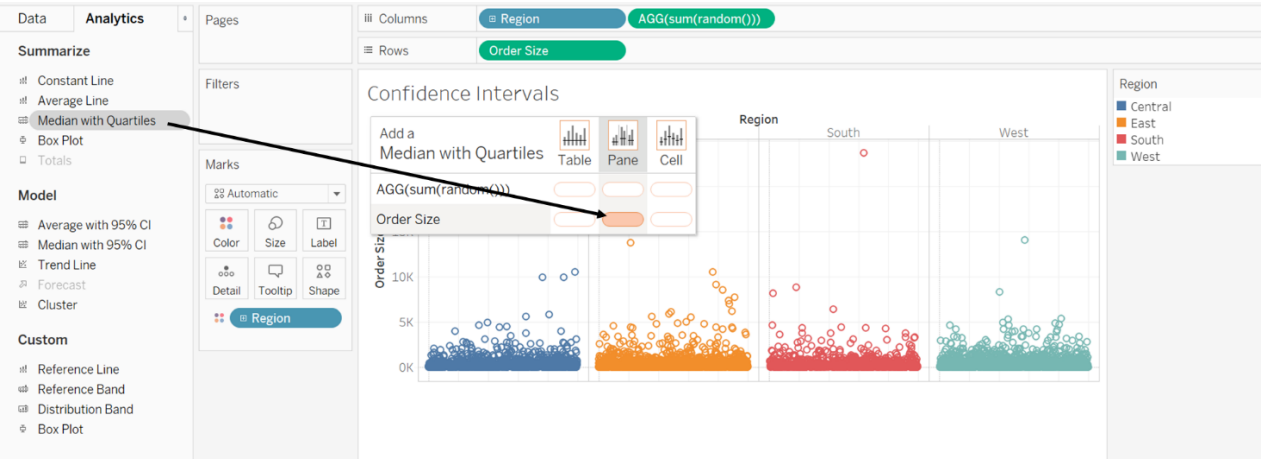
Visualization with Confidence Intervals

Let us next look at confidence intervals. Begin by dragging [Order Size] to the Rows shelf and verifying that the 'Aggregate Measures' selection is *unchecked* under the Analysis tab. The individual order sizes should appear as individual shapes oriented in a vertical line as shown below.

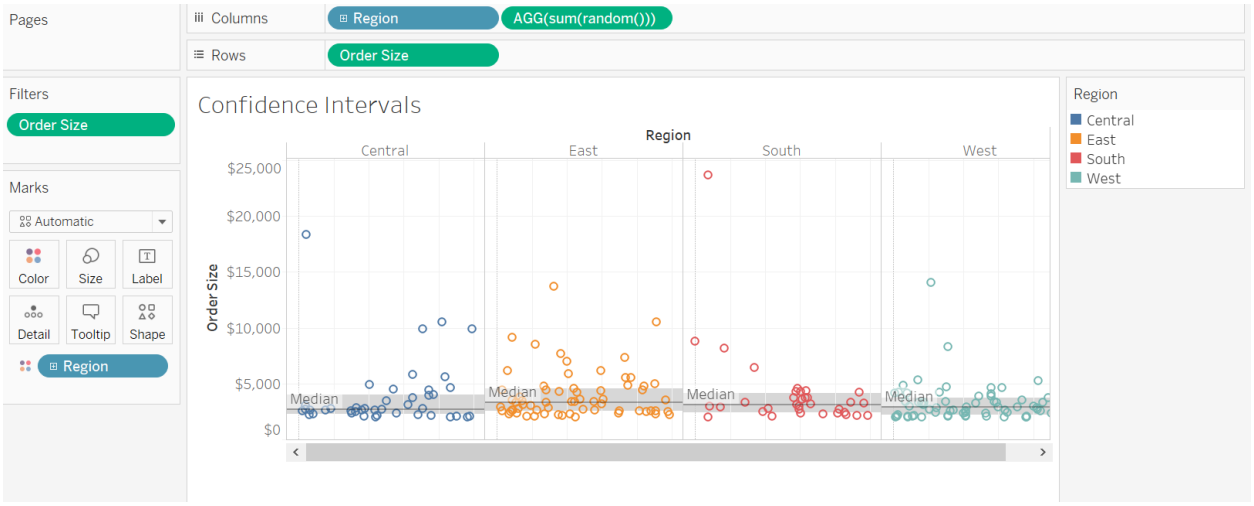


To decongest the data, type “sum(random())” into the Columns shelf. This is a trick that randomly spreads the data points across the horizontal axis. Note that a point’s horizontal position suggests nothing about it. The trick is only to improve the visualization of the data points. Add [Region] to the Columns shelf so that each section’s data is shown side by side. You can right click on the horizontal axis and unselect ‘Show Header’ to hide it from the view. Drop [Region] on the color card to further distinguish the different sets of data. If preferred, you can change the size and shape of each data point under the Marks toolset.

To visualize how concentrated each region’s orders are, drag the ‘Median with Quartiles’ pill from the Analytics tab into the visualization. Make sure that the pill is dropped into the ‘Pane’ and ‘Order Size’ intersection as shown below. This will create separate medians with quartiles in each region’s pane.

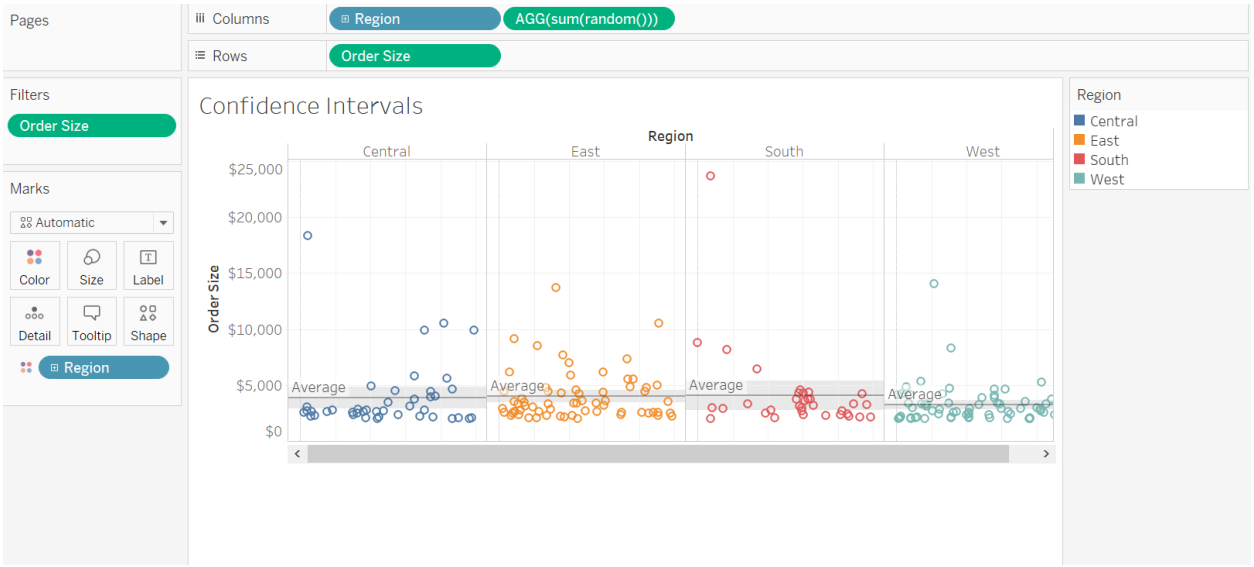


We can conclude that all the regions tend to have many small orders. To improve this example, let us assume that we are focusing on larger orders so filter out orders less than $2,000. Now we can see that the East Region has a higher Median than the Central Region for orders greater than $2,000.



You can now clear the ‘Median with Quartiles’ feature by right clicking the band and selecting ‘Remove’. (It may take a few attempts to remove.) Let us instead look at the confidence intervals for each region relative to the Average rather than the Median. Drag the ‘Average with 95% CI’ pill over to the same intersection we used before (Pane & Order Size).

The Confidence Interval bands use Standard Error to predict what the population’s mean would be, with 95% confidence.

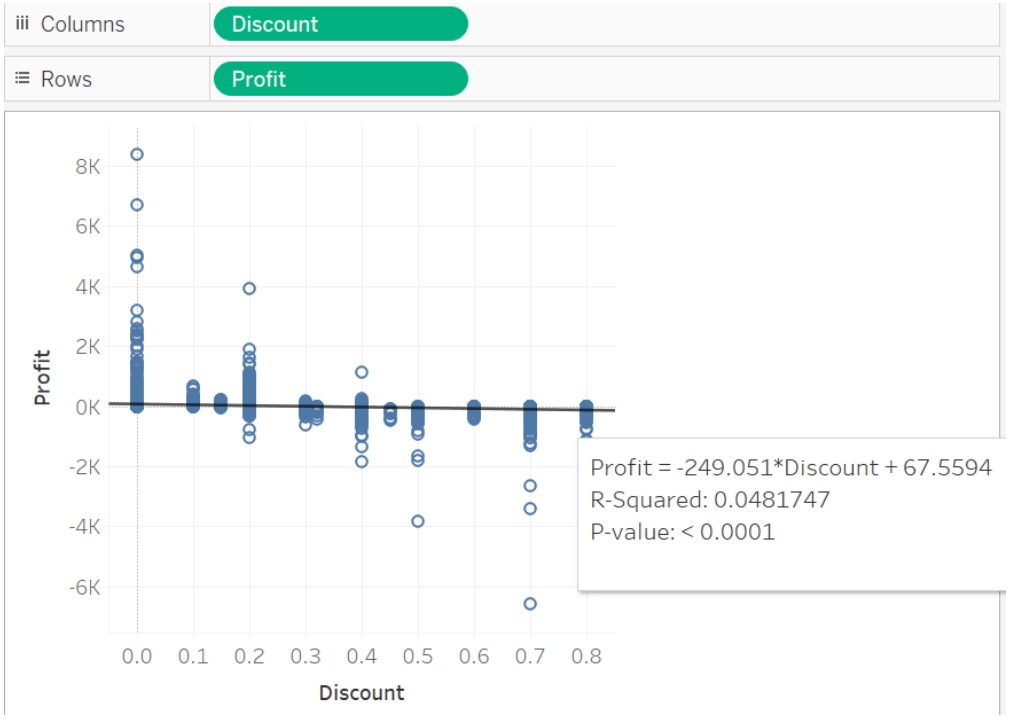


When we are only looking at Order Sizes greater than $2,000, the South has the highest Average and the widest confidence intervals.

Residuals from a Trendline

You would use a residual plot to help determine the fit of your linear regression model. Since residuals show the remaining error after the line of best fit is calculated, plotting residuals gives you an overall picture of how well the model fits the data and, ultimately, its ability to predict.

Consider the scatterplot of Discount vs. Profit and the associated trendline. The trendline has a very small p-value (showing highly significant), but with a small R-square. Perhaps we could learn more from seeing the residuals.



In order to calculate the residuals that we wish to plot, we can create calculated fields to find the slope and y-intercept using the equations:

Slope = [correlation] * ([std deviation of y] / [std deviation of x])

Y-intercept = Avg[y variable] – [slope] * Avg[x variable]

or in our example:

Slope

```
Corr([Discount],[Profit])*((STDEV([Profit]))/STDEV([Discount]))
```

Y intercept

```
Avg([Profit])-[Slope]*Avg([Discount])
```

Then we need to calculate the predicted dependent variable:

predicted y-variable = {[slope]} * [observed x-value] + {[y-intercept]}

or in our example:

Predicted y

```
{[Slope]}*[Discount] + {[Y intercept]}
```

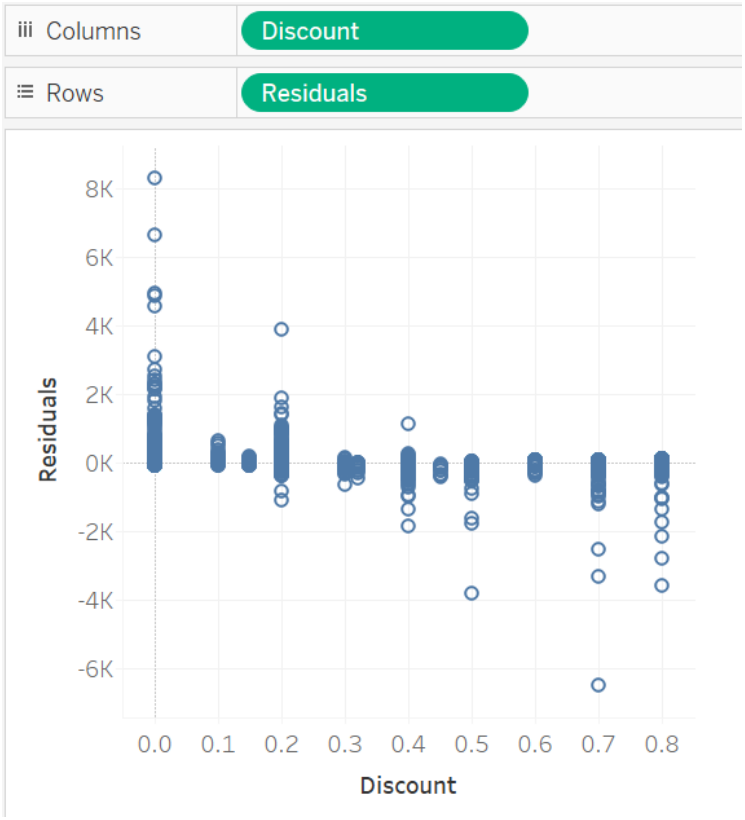
Since the slope and y-intercept will not change value for each Discount, but we need a new predicted output (y) for each value (x), we use a level of detail calculation. The curly brackets tell Tableau to hold the slope and y-intercept values at their constant level for each discount. This is equivalent to a FIXED LOD with no scoping keyword which means define at the table level - {Fixed : [Slope]}

The Residuals are then observed – predicted:

Residuals

```
[Profit]-[Predicted y]
```

.



Inspect your residual plot.

Don't forget to inspect your residual plot for clear patterns, large residuals (possible outliers) and obvious increases or decreases to variation around the center horizontal line. Decide if the model should be used for prediction purposes.

- The residual plot makes it easier to see the amount of error in your model by "zooming in" on the liner model and the scatter of the points around/on it.
- Any obvious pattern observed in the residual plot indicates the linear model is not the best model for the data.

In the plot above, the residuals increase for the more extreme discount values (either 0 or 0.08). This means the error in predicting the discount gets larger as the discount gets very small or larger. Perhaps this model is not effective in predicting profit for extreme discount values.

Topic 10: Maps

Some examples of relevant business questions that maps can answer include:

- In what state do we sell the most products?
- What regions of the U.S. have the highest obesity rates?
- What countries have the lowest per capita income?
- Which metro station is the busiest for each metro line?
- Where are customers checking out and returning rental cars most frequently?

To build a simple map, your data source must contain location data (for example, location names or latitude and longitude coordinates).

This section provides an overview of how to use Maps in Tableau. This topic uses superstore data. There are two basic kinds of maps – symbol maps use marks as symbols and maps are just filled maps. Filled maps cannot change size because size is determined by geography. Sometimes this distinction is blurred if you have symbols on filled maps.

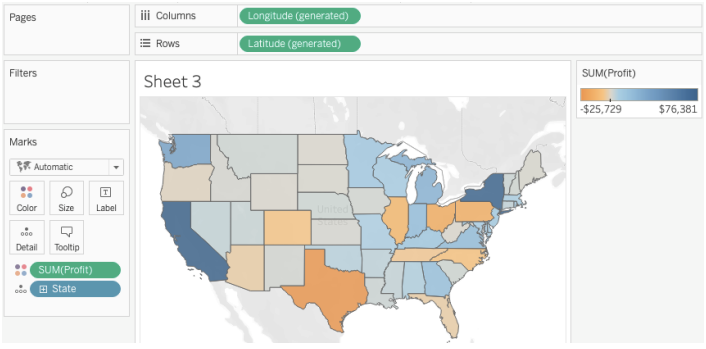
Color and Maps

Placing a measure on the color marks defaults to a filled map. A dimension on color defaults to a symbol map. If the pill is continuous, Tableau will show a color gradient. If it is discrete Tableau will show a palette of distinct colors. This sounds more complex than it is – let's do a couple maps and see it in action.

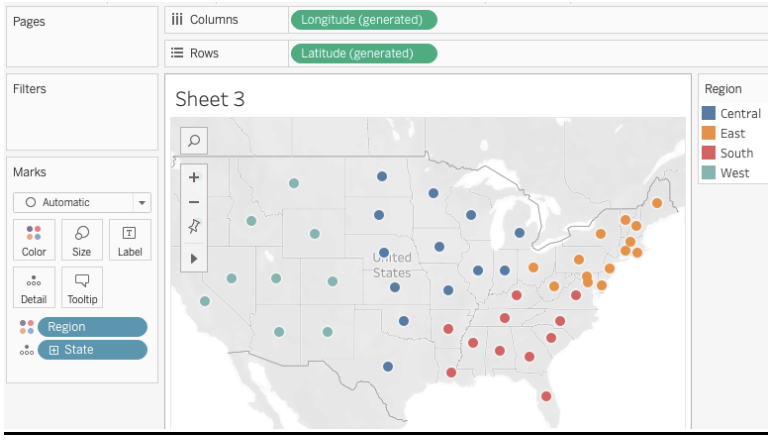
With the Superstore Data, start a new sheet. Double click State to get a map.

Drag a continuous measure (Profit) to color – You get a filled map with gradient shading as shown below. Now undo, and drag a discrete dimension (Region) to color – You get a map with a unique color for each region. You could add more data to this map, for example, by dragging a measure (Discount) to size.

Gradient

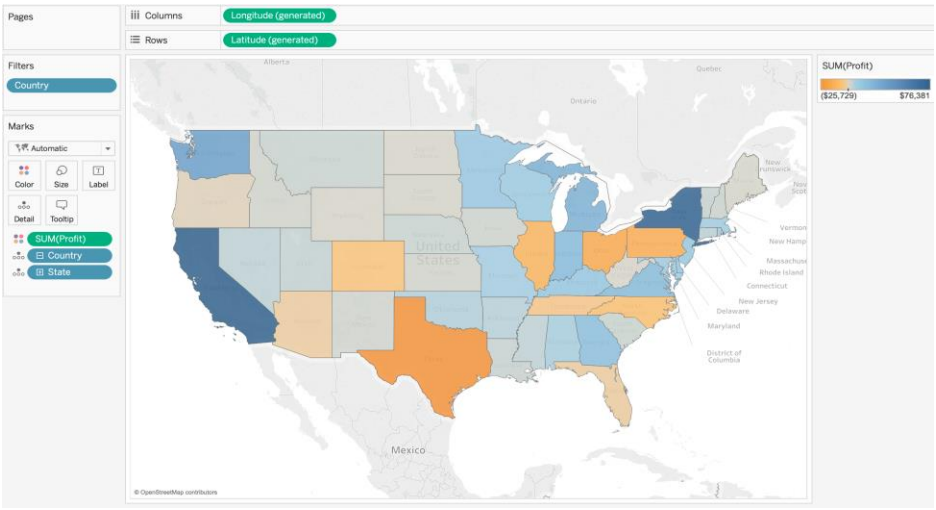


Discrete Colors



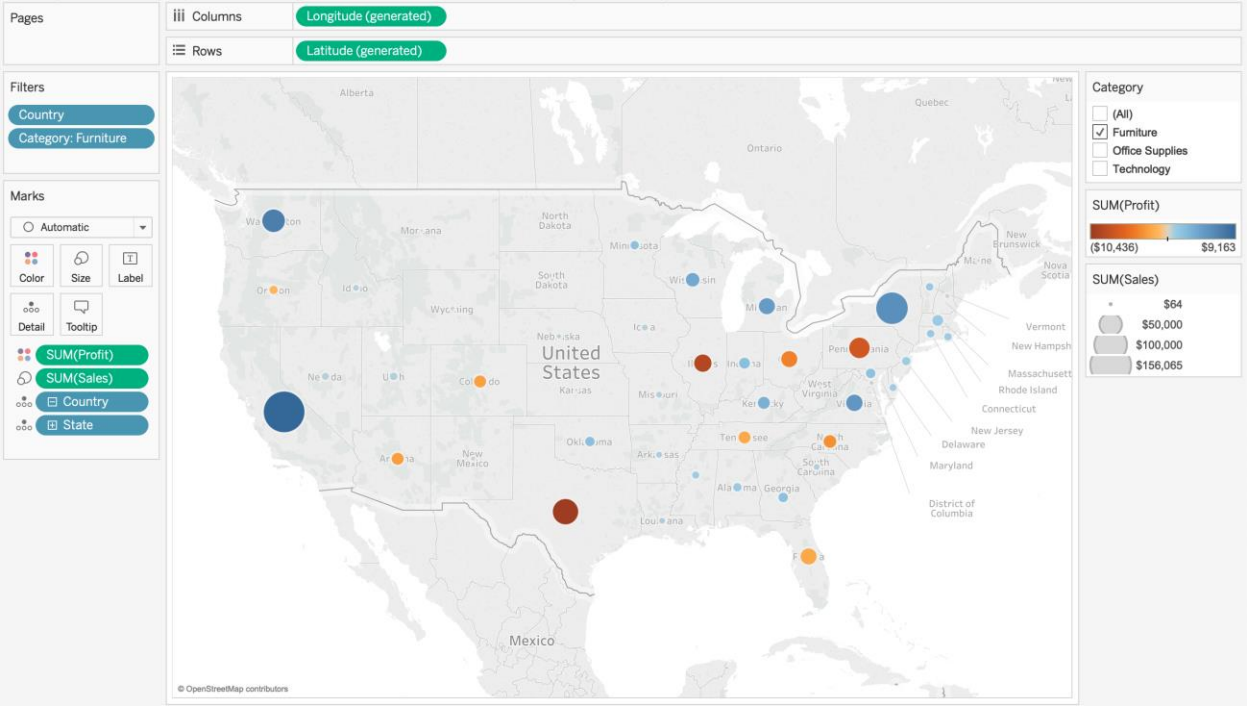
Working with Maps

Continuing with the gradient chart, overall, Texas looks like a challenging market for us. Now let's say you want to drill down and just look at the furniture category:

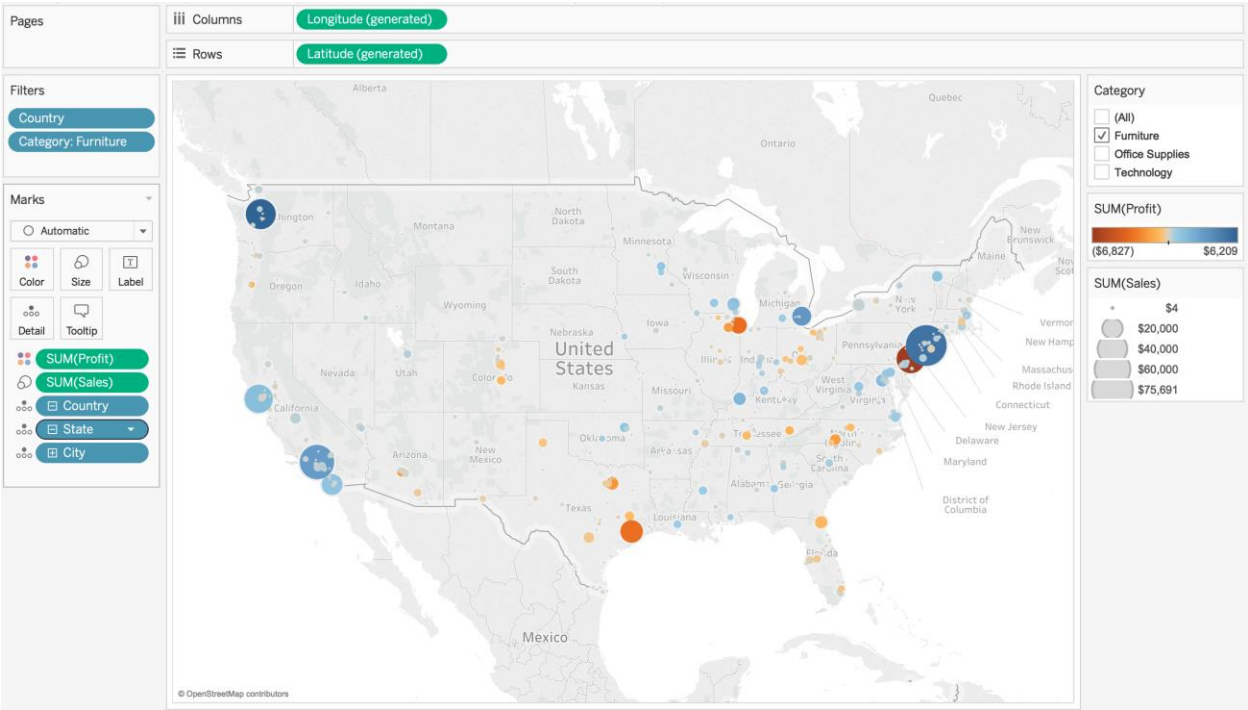


Right-Click on **Category** and select Show Filter. Select just the Furniture category.

Filled maps like this are great when you just want to see one measure on your map. But I would like to see both profits and sales. Only one attribute can be on color, so let us drag **Sales** to Size on the Marks card. Now you see them both together (click Size to make the bubbles larger).



Finally, it is easy to explore different levels of geography: click the plus by **State** on the Marks card to switch from State to City bubbles, then Undo.



Side note: support for geographic roles varies from country to country. If you are interested in a specific non-US region, research geographic support for that region. Tableau is getting better at this with every new release.

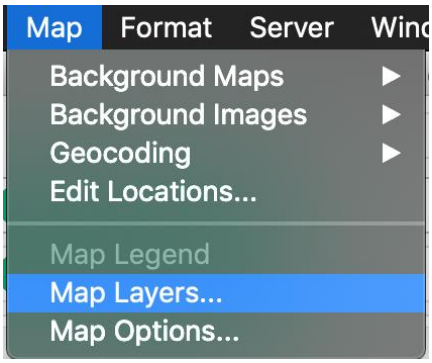
What is Latitude (generated) and Longitude (generated)?

When you add a geographic dimension to your data, Tableau tries to match it to geographies it knows. So when you add something called Country, Tableau looks at all the country names in that column and generates a latitude and longitude for that country (at somewhere around the geographic center of the country). There is a distinction between latitude/longitude that Tableau generates and values you might find in a data source.

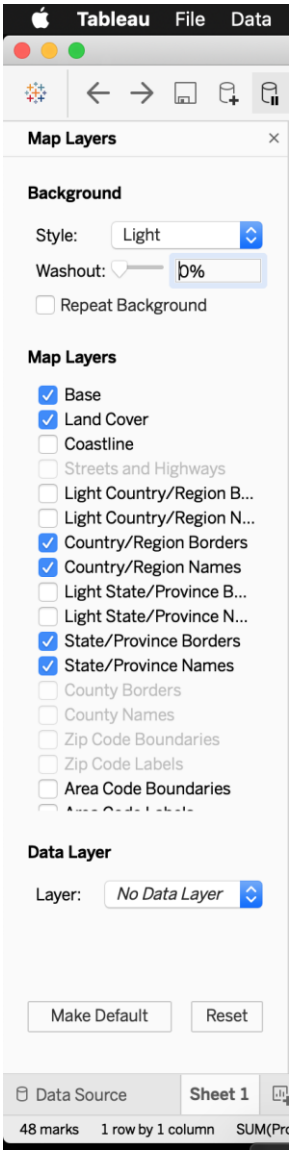
Map Options

When working with maps, there are map-specific formatting options available.

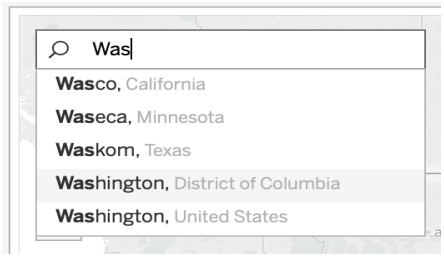
- Open the Map menu, then select Map Layers.

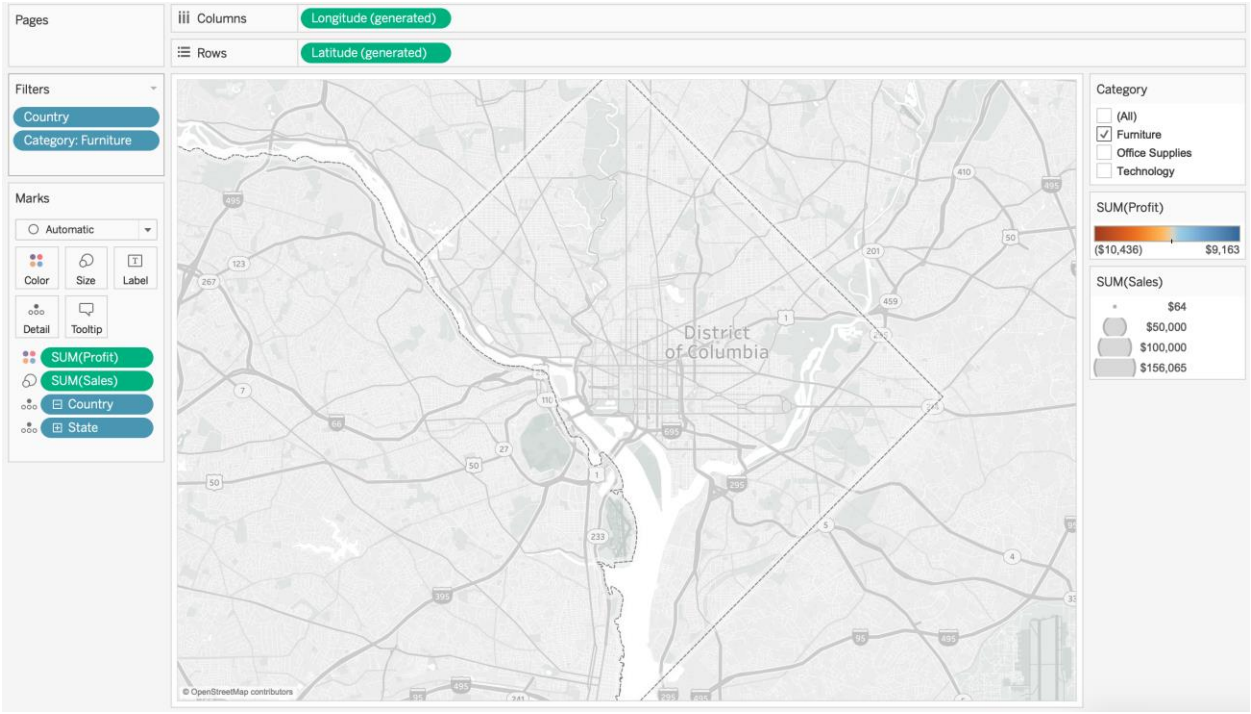


- A pane opens over the data window just like the formatting pane.



- Here, you can change our map style – Dark, Light, or Normal
- Control the washout (transparency of the map).
- And control map layers, such as borders and names.
    - Some options, such as streets and highways, can be turned on only when the map is sufficiently zoomed in (currently greyed out).
- Hover over the map and click the magnifying glass. Let's search for DC, and now you can turn on Streets and Highways.



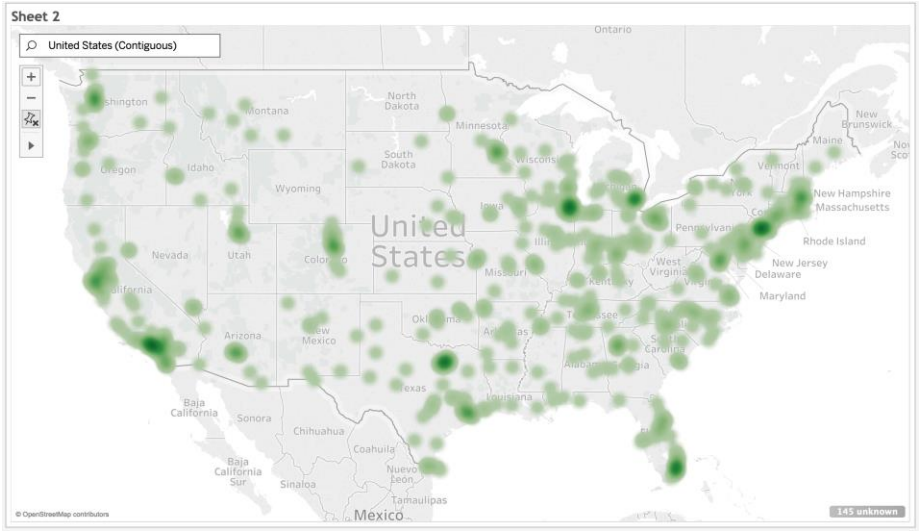


Depending on your map's location, other data layers can also be added, such as census data. I won't cover that here, but you can play with the data layer dropdown to see what is available.

Heat Maps (Density Marks)

Sometimes you want to see how sales or traffic or something is dispersed geographically. This is the perfect time for a heat map (on a geographic map). It's a very simple process:

New sheet. Select **Sales** and **Postal Code** then Show Me a Symbol Map. Change the mark to density. Zoom in to see only the contiguous US.



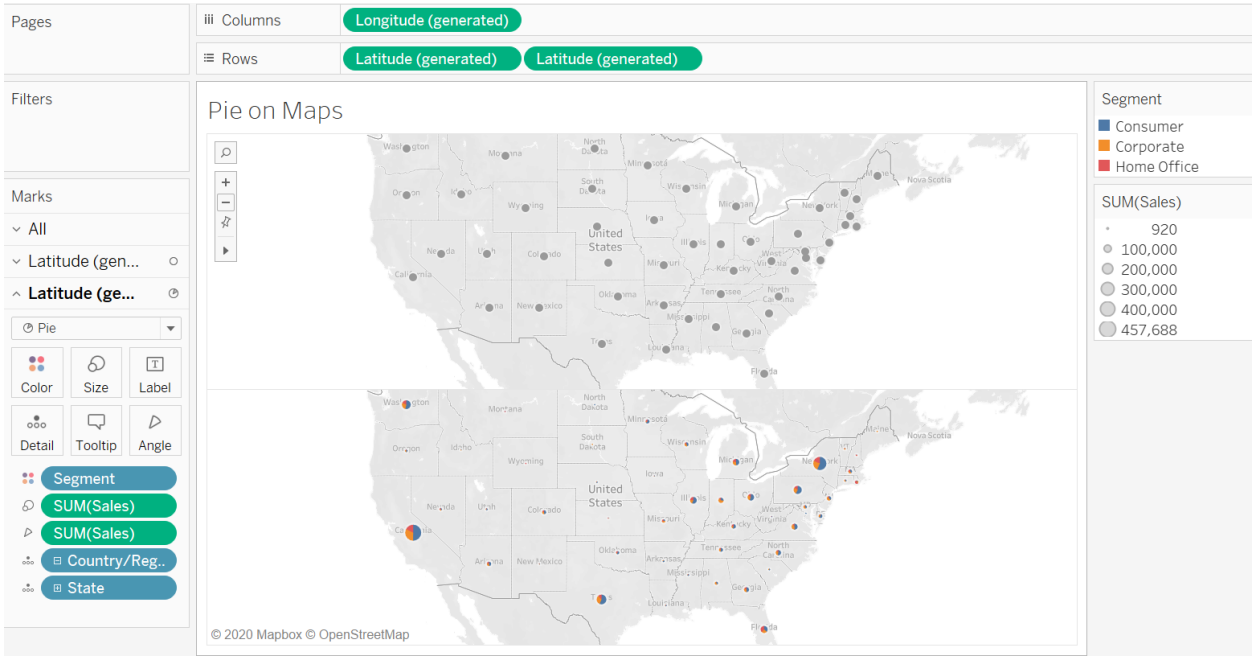
Pies on Maps

No one is a fan of pie charts. However, they can be useful when placed on top of a map.

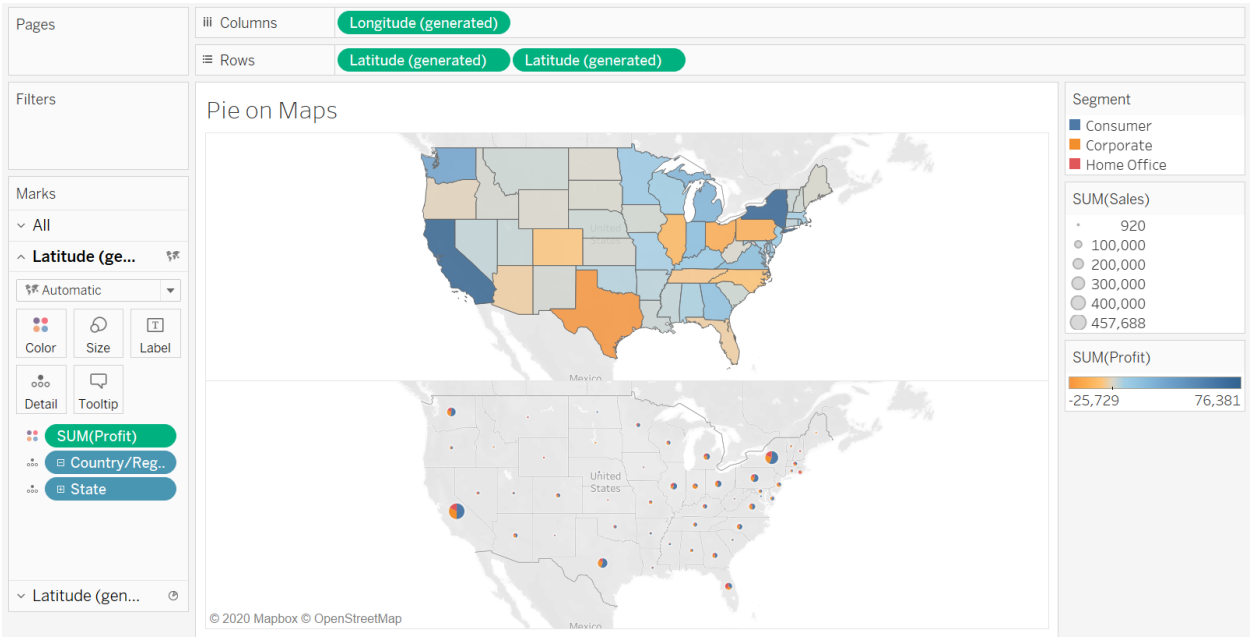
1. Double click on **State**
2. Bring **Latitude (generated)** to Rows again.
3. Customize the bottom chart:
    a. Segment to color

b. Set the marks to pie. At this point you have a bunch of pie charts where the size and angle really just show me which segments I've sold in each state.

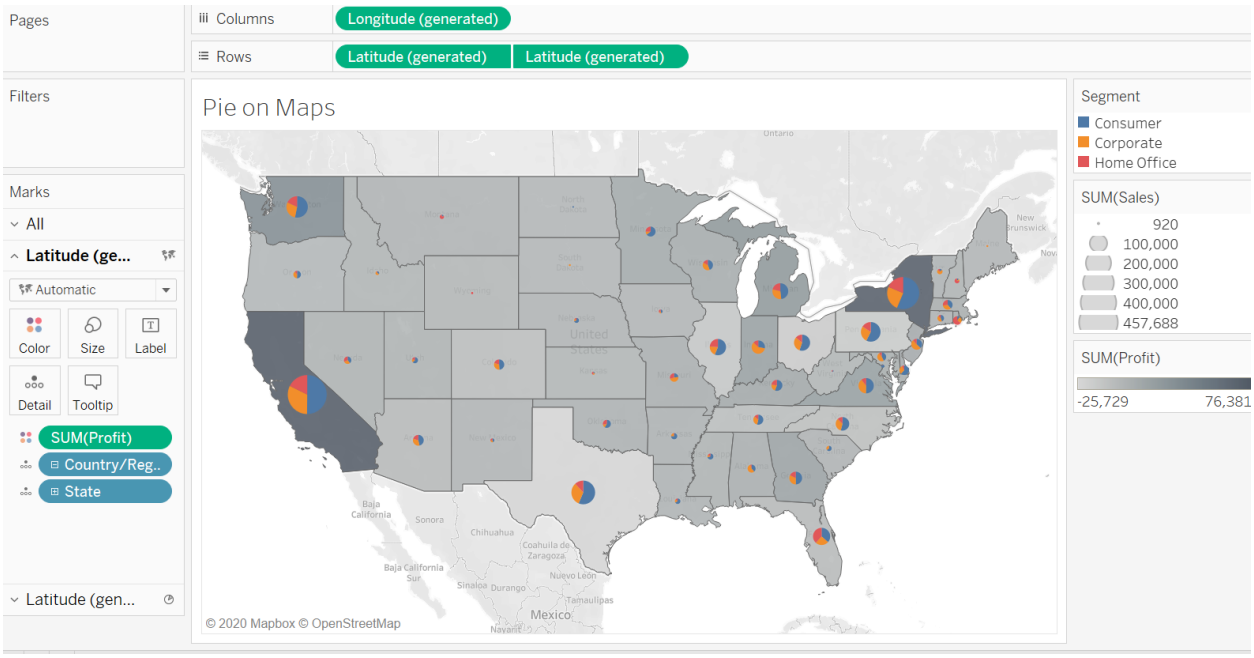
c. Bring Sales to angle and size



4. Customize the top chart: Drag Profit to color to create a filled map



5. Right click on **Latitude (generated)** and create a dual axis. To improve readability, change the color of the filled map to gray scale.

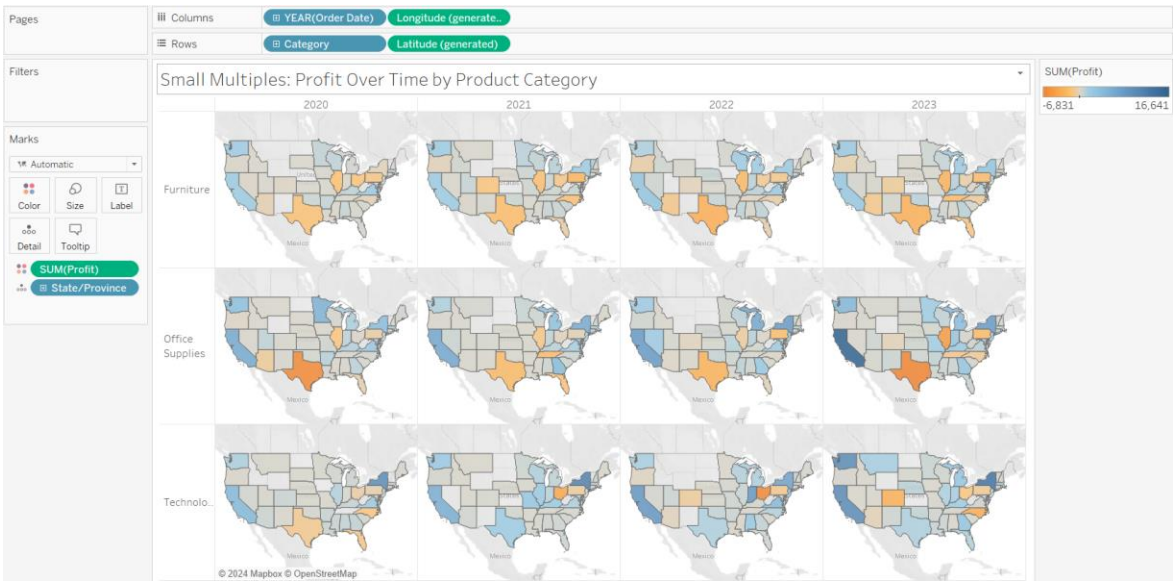


Small Multiple Maps

Using small multiples with maps shows trends for different measurements across time and grouped by different traits such as segment or brand. By adding the map element; however, small multiples add the additional dimension of geography so that trends can be viewed across both time and space.

In the Superstore data set, we can view how profits vary over time and space by product category.

To create a visualization with small multiples of maps, begin by creating the initial map. First bring **State** to the canvas, if necessary filter on US in the **Country** field. Bring **Profit** to the Color mark and change the color to Orange-Blue diverging. The map is now the original choropleth map showing total profit per state across the U.S. To shift to small multiples, a dimension must be added to both the Rows and Columns shelves. Here, add **YEAR(Order Date)** to the Column shelf and **Category** to the Rows shelf.



Additional information on creating small multiple maps is provided at https://dataremixed.com/2014/09/how-to-make-small-multiple-maps-in-tableau/.

Topic 11: Summary of Formatting Worksheet Tips

This topic uses the superstore data. Much of it is review at this point, but there are a few new ideas.

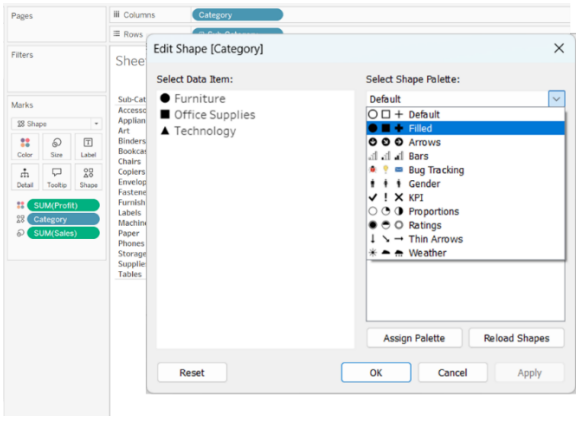
Changing Shapes in Visualization

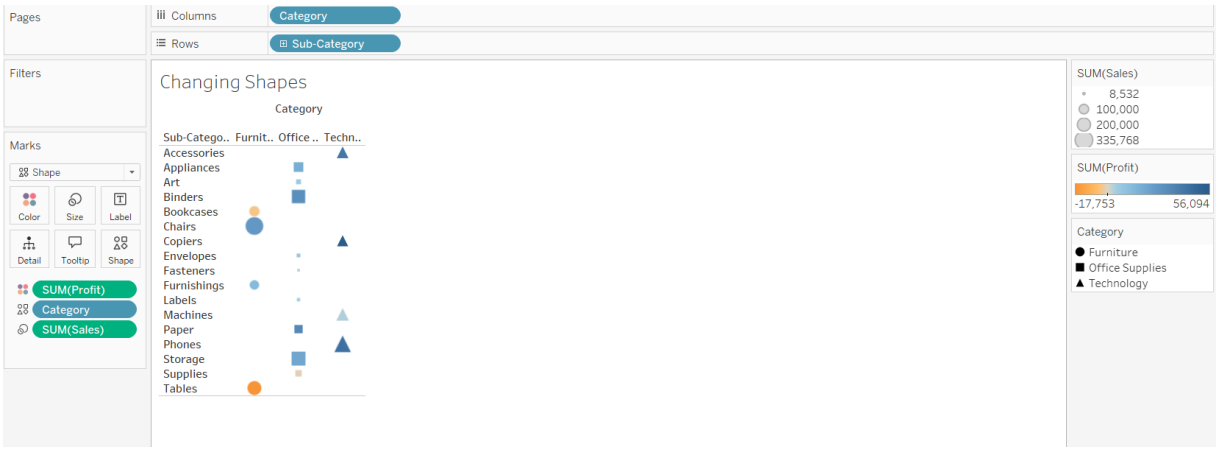
Using the superstore data, drag "Category" to the Columns shelf. Drag "Sub-Category" to the Rows shelf. Drag "Sales" to the Size on the Marks card. Drag "Profit" to the Color on the Marks card.

On the Marks card, click on the drop-down menu and select "Shape." Then drag "Category" to the Shape on the Marks card. Then click on the Shape icon, to open the Edit Shape dialog box. Select different shapes for each Category. For example:

- Furniture: Circle (Filled)
- Office Supplies: Square (Filled)
- Technology: Triangle (Filled)

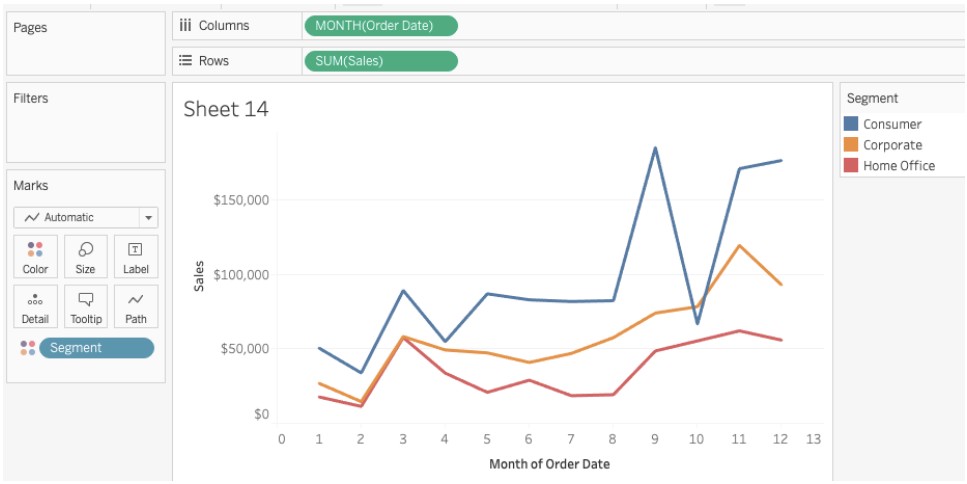
Click on the Color legend to choose appropriate colors for positive and negative profits. Adjust the size slider on the Size legend to make sure the shapes are clearly visible.



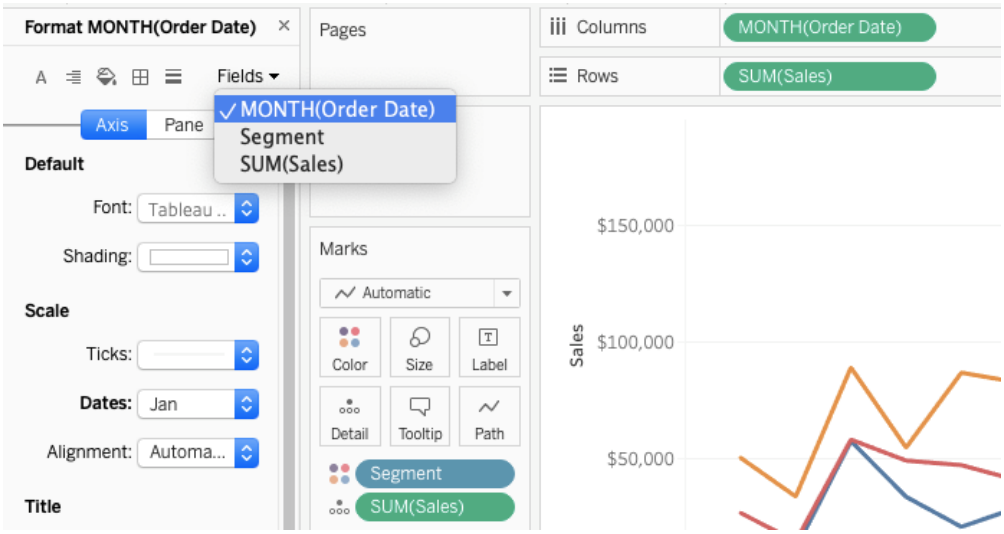


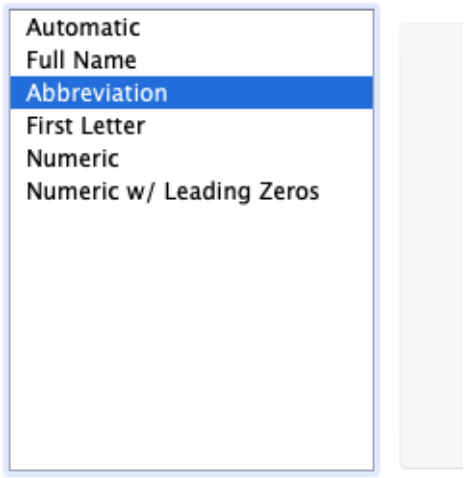
Labels

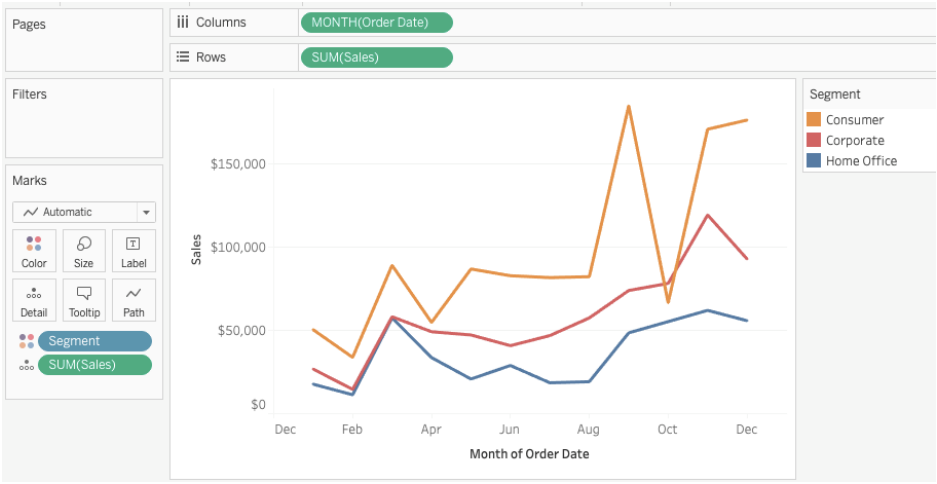
Start by making a line chart – Sales by month (Discrete) of Order Date, Segment onto color.



In order to change the axis labels to months instead of numbers, right click the month axis and select format. Under the fields dropdown, select MONTH(Order Date) and then change the Dates dropdown to Abbreviation.



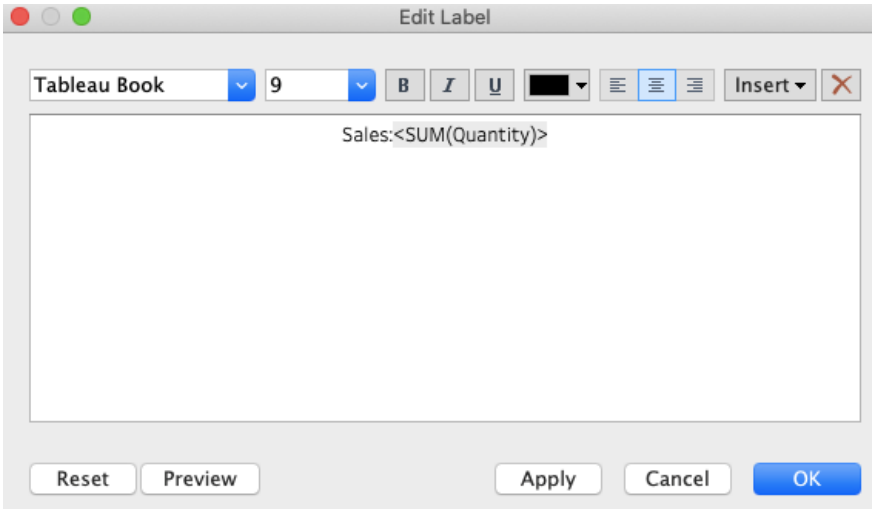




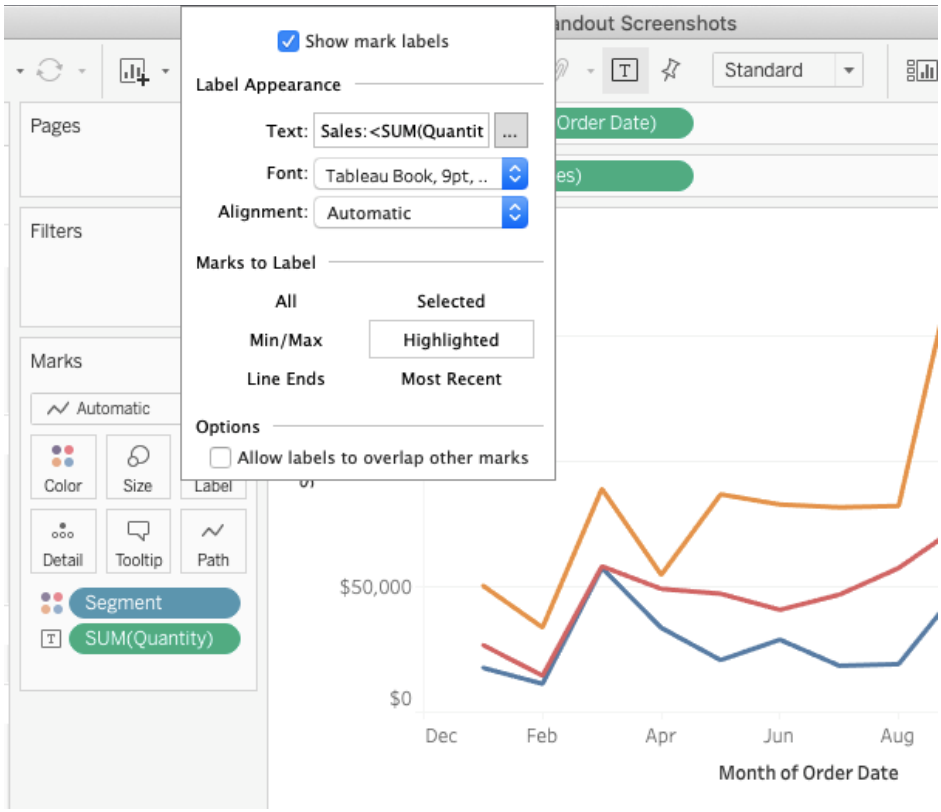
Click the Label Marks card to see the options:

Checking the ‘show mark labels’ box will turn labels on in the view, the same as clicking the [T] label button in the toolbar.

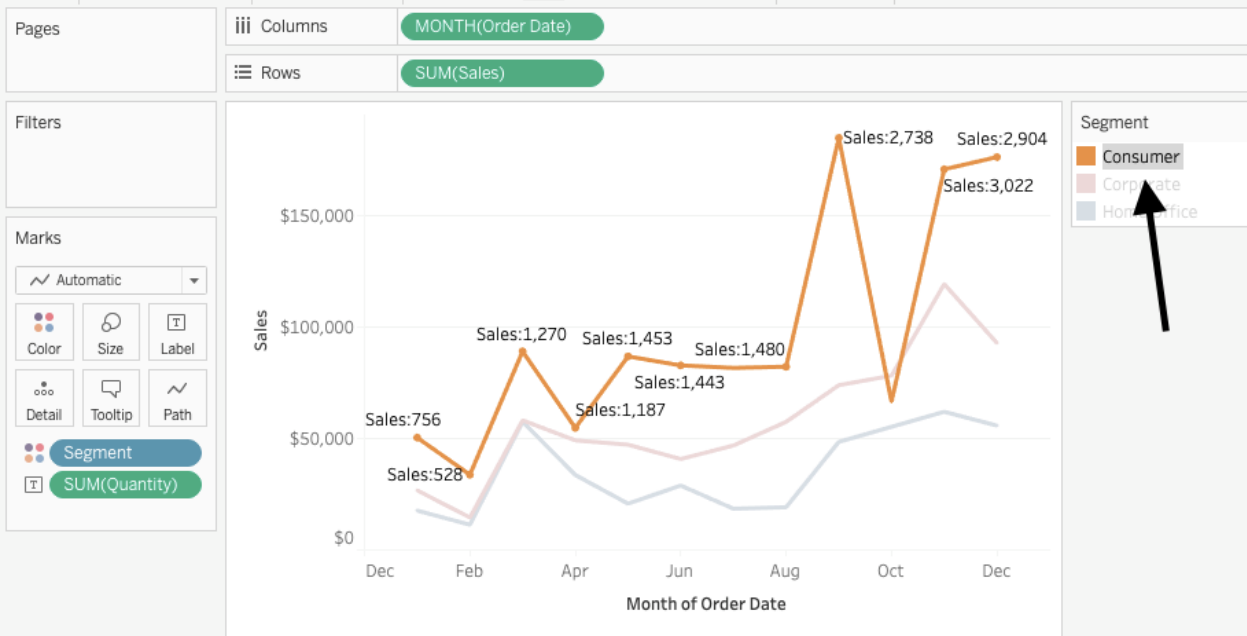
Drag Quantity to the [T] Label Marks card. This adds it to the label (you can temporarily turn it off with the [T] in the toolbar). Click the Label Marks card and select the ellipses […] to bring up the full text editor to customize the label so that viewers know that the number label represents sales. Type in Sales: before the existing text as shown below.



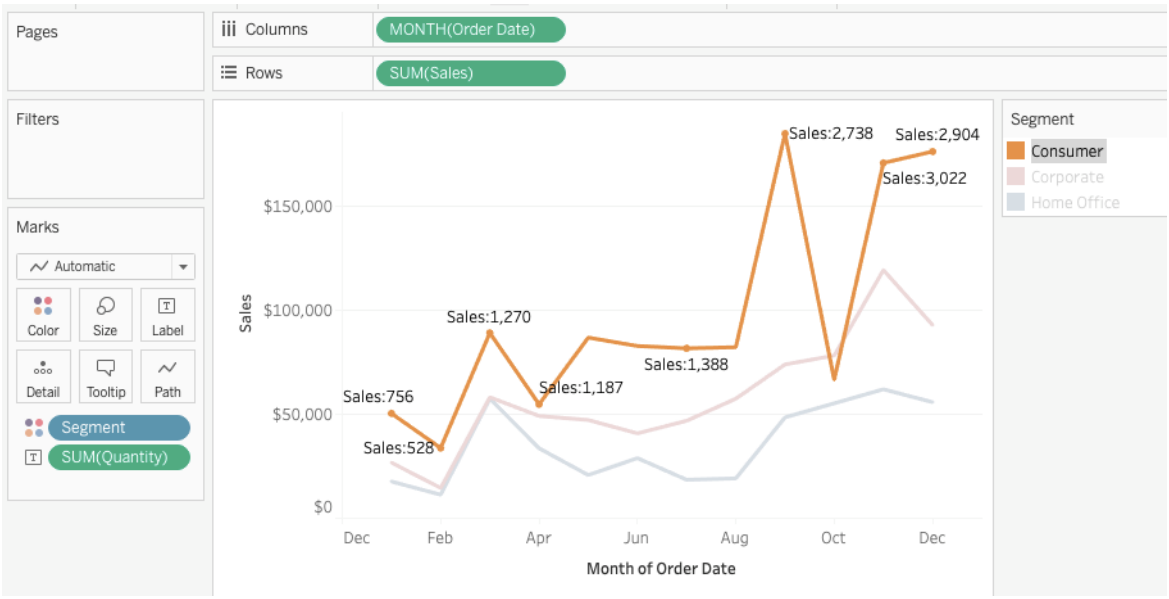
Select the label shelf again and check the show marks label box. Click on ‘highlighted’ for marks to label and make sure the last box is not checked so that labels do not overlap.



We have set up quantity labels so that they only show when a line is highlighted. If you select a segment on the legend, the quantity labels should appear as shown below.



Notice how some parts of the chart are too busy with all of the labels. If you want to never (or always) see a specific label, just right click on the point, go to Mark Label, and choose the desired behavior, such as Never Show. Try removing labels for May, June and August.

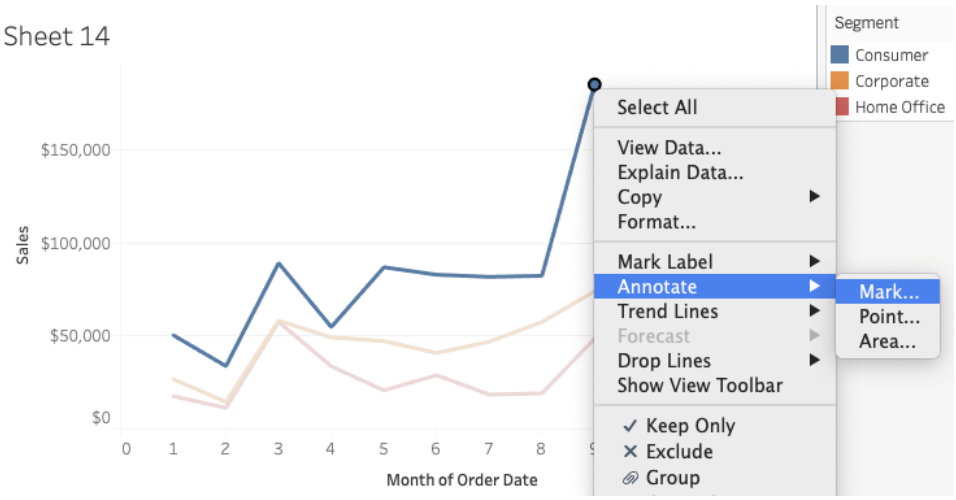


But you often do not have much space in the view for labels. Often, the best way to add additional information is through the use of annotations or Tooltips.

Annotations

Annotations are text boxes used to call out something specific in the view.

- A mark annotation stays with the mark itself, regardless of whatever changes are made on the view.
- A point annotation stays in the same relative position on the view, independently of the mark, as does an area annotation.
- Let us start with a Mark annotation – right-click a data point in a visualization and select Annotate…Mark



- These are not simple text boxes; like labels, annotations can have fields added to them as dynamic inputs.
- The appearance of annotations is easy to customize just as you would change text within Word.
- A point annotation is similar, but it will stay in a particular place in the view rather than be tied to a mark, which can move as you interact with the view.

Tooltips

To show Tooltips, create a new worksheet:

Select **Sales** and **Profit**, choose Scatterplot from Show Me

Segment to Color and **Region** to Detail.

Change the Mark to Circle and adjust the size if you want.

Hover – the standard Tooltip here has a lot of information. But with just a little work you can make it more user-friendly and engaging:

Click Tooltip on the Marks card.

Change it to read something like (and you could change the size and color of some words):

Total of <SUM(Sales)> in sales with <SUM(Profit)> in overall profit.

Remember that if you want to include a field in the tooltip that is not in your visualization, just drag it to the Tooltip shelf. Also, when you insert a field, use the “Insert” button in upper right of dialog box to bring up the list of fields available to the tooltip.

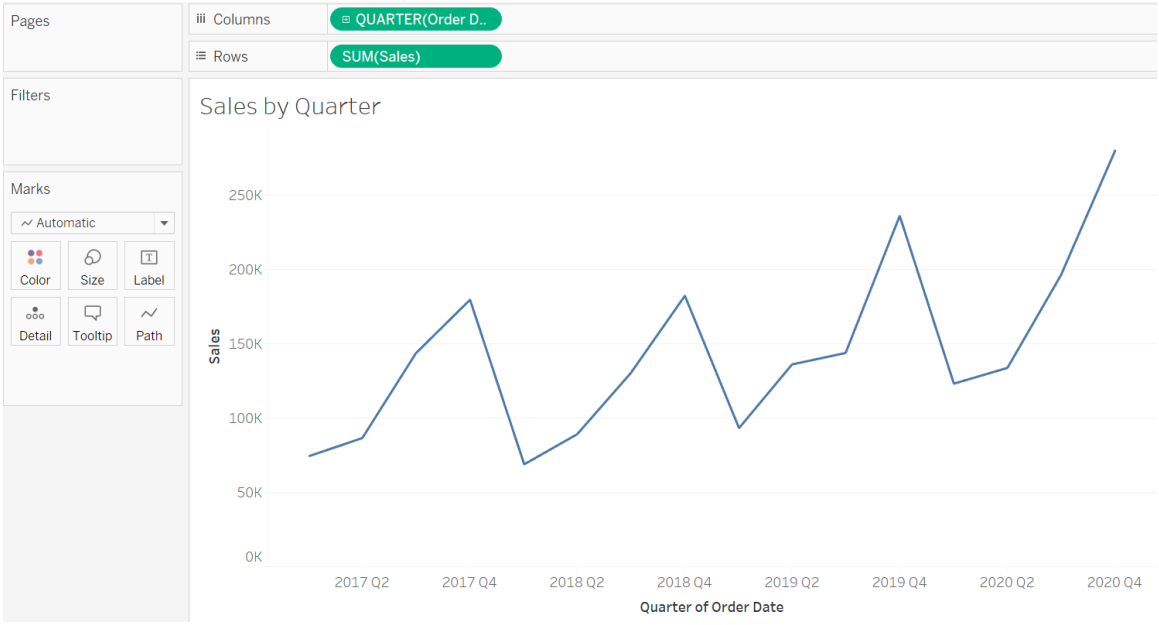
In addition, if you prefer not to see the command buttons (keep only, group, view underlying data, etc), you can turn them off with a checkbox in the Tooltip window.

With every visualization you make, do not forget to pay attention to the Tooltips! They are often an important part of the user interactive experience.

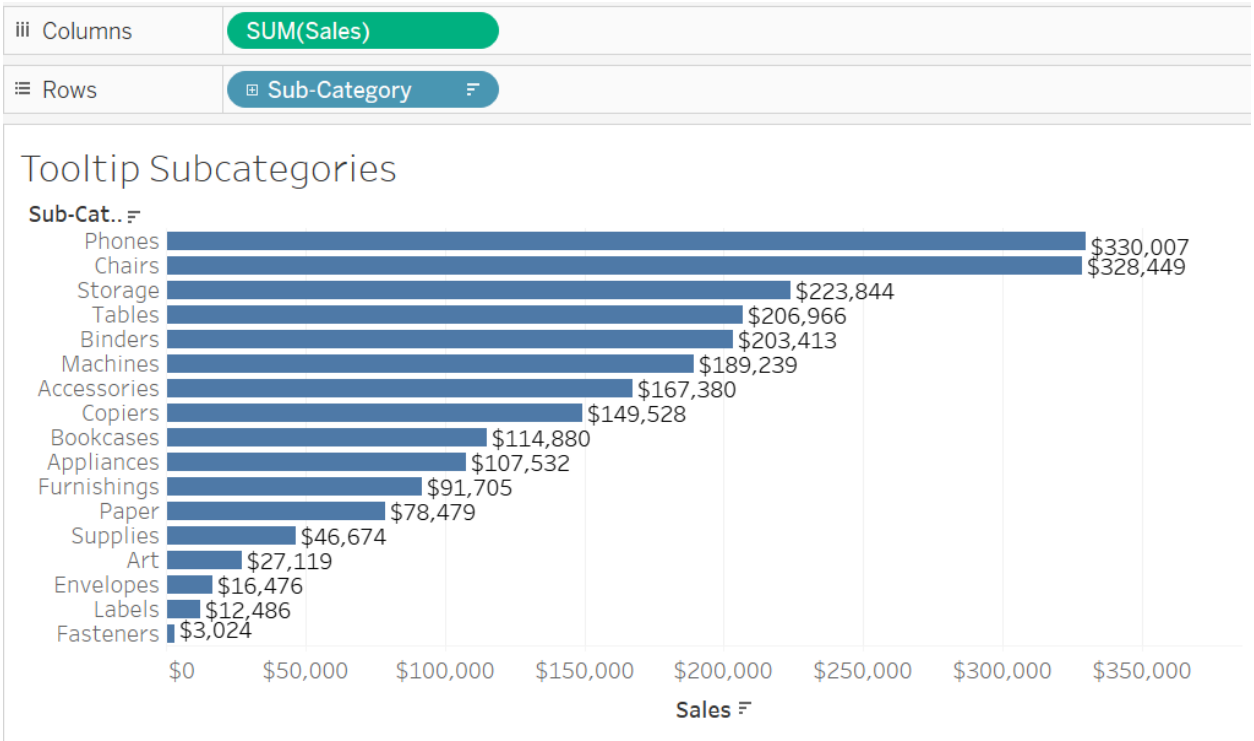
Including a Visualization in a Tool tip

You can show related visualizations in tooltips to help your audience engage with the data at a different or deeper level, while keeping them in the current context and maximizing the space available for the current view.

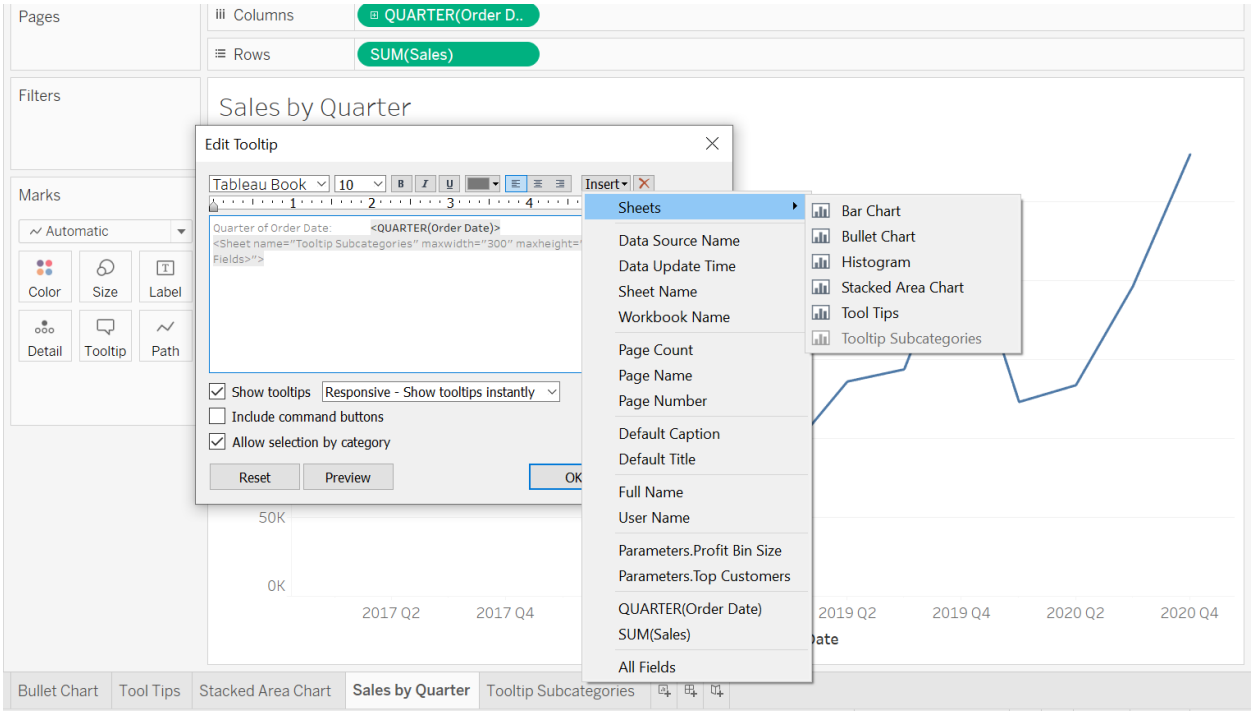
1. Create a visualization in the *source* worksheet. In this example, we want to show sales by Quarter for the Order Date.



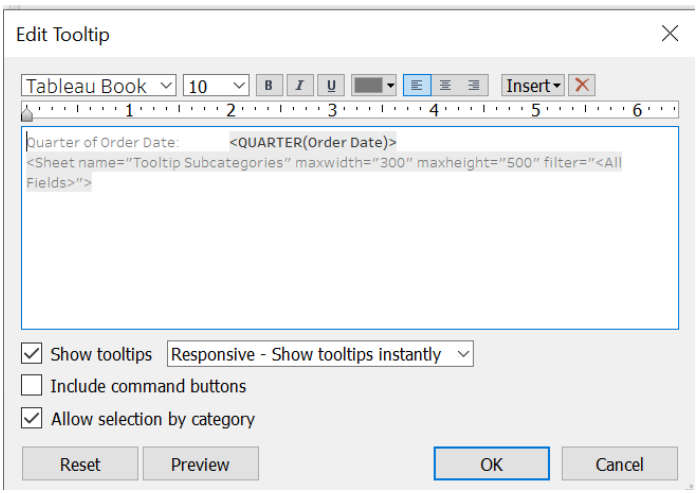
2. Create a visualization in a *target* worksheet view to serve as the visualization in the Tooltip. In this case, we want to show the sub-category sales.



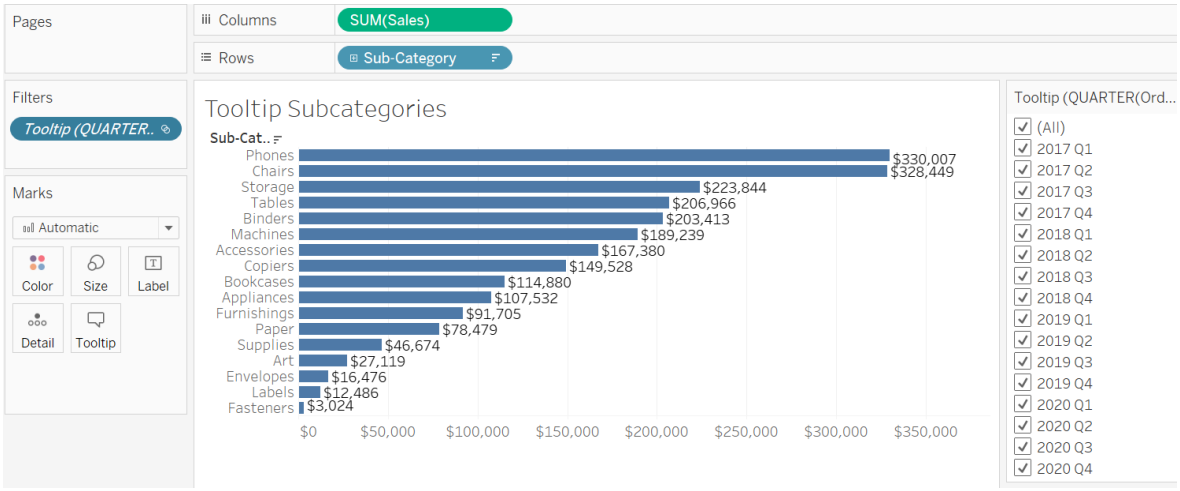
3. In the source worksheet, click Tooltip in the Marks card. In the Tooltip Editor, insert a reference to the target worksheet.



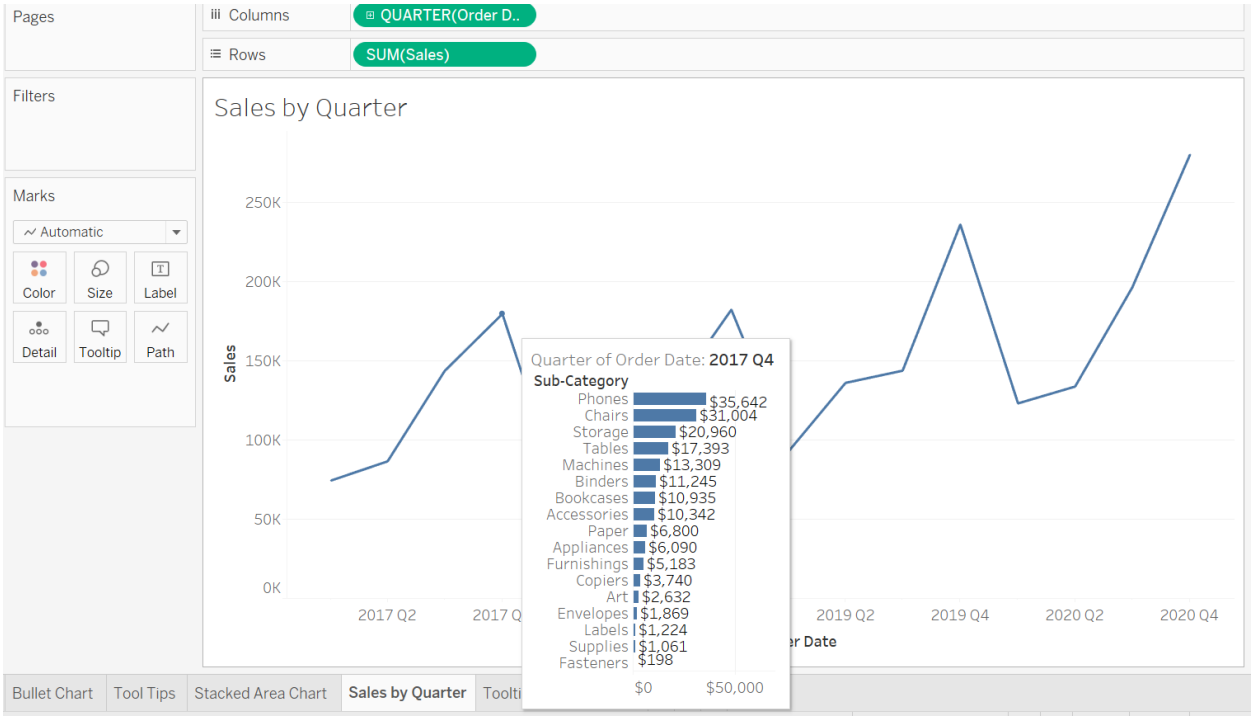
4. Adjust the 'title' in the Tool tip. You may also need to adjust the size (here, maxheight was changed to 500).



5. Test to see if this is working. Check the target worksheet. You will see that Tableau has added a Filter to the target worksheet that relates to the Tooltip.



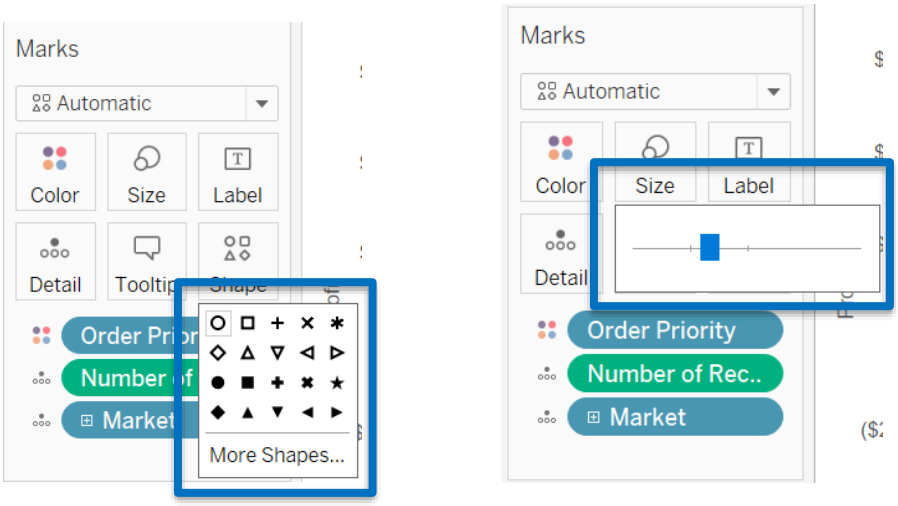
6. If you are satisfied that the Tooltip visualization is correct, you can hide the target sheet by right-clicking the worksheet tab and choosing ‘Hide’. If you need to unhide, right-click any remaining worksheet tab and choose ‘Unhide All Sheets.’



Formatting Specific Parts of the View

Many parts of the view can be independently formatted. To see if a field, number, legend, or other part of the view can be formatted, right-click or bring up the menu to see if the format option is available.

Marks can be edited by first activating the data field of interest and then changing the settings.



Editing and Formatting Axes

To adjust anything having to do with the layout of the axis, right click and select Edit Axis.

- This dialog controls the range, tick marks, and the title.
- You can edit the range, to a fixed range that you prefer. You can change the tick marks, axis title, etc. also
- Additionally, this pin icon has appeared – this indicates the axis is no longer automatic.
- To hide an axis label entirely, right click on the axis and uncheck Show Header.

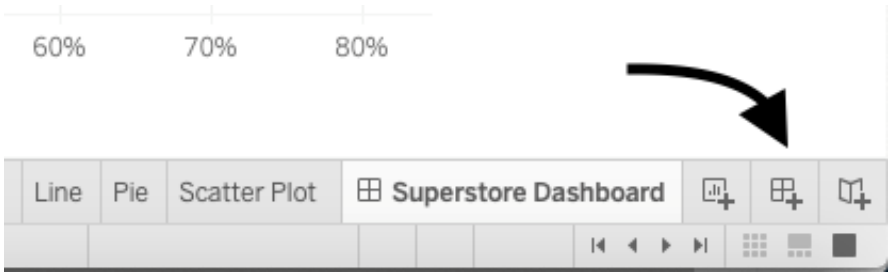
Topic 12: Dashboards & Stories

This section covers how to create dashboards and stories in Tableau. This topic builds on the workbooks created using superstore data previously. Individual worksheets can provide powerful insights but often it makes sense to combine these worksheets into a single dashboard. Dashboards make it easier to analyze different aspects of the data in context of each other.

Open the Tableau file that you created during in the Basic Charts section (Topic 7). The worksheets do not have to be exactly the same but try to name them so that you can differentiate them as shown below. You can name a worksheet by right clicking on its tab.

Creating a Basic Dashboard

Click the icon to create a new dashboard.



Name the dashboard. Go to the Dashboard drop-down and select Show Title.

Notice the Dashboard options. You should always consider these options before starting your Dashboard. You should probably leave the size setting to Desktop, but remember that Laptop can be a good choice too if you are choosing a Fixed size. “Automatic” is OK sometimes, but not if you want to have floating legends (which you will), since these will move around as the screen resizes under the Automatic setting.

Sheets – these are the individual worksheets in your workbook that can be added to the dashboard.

Objects – Horizontal: Adds a horizontal layout container that additional objects can be added to.

Vertical: Adds a vertical layout container that additional object can be added to.

Text: Allows you to add and format text like titles.

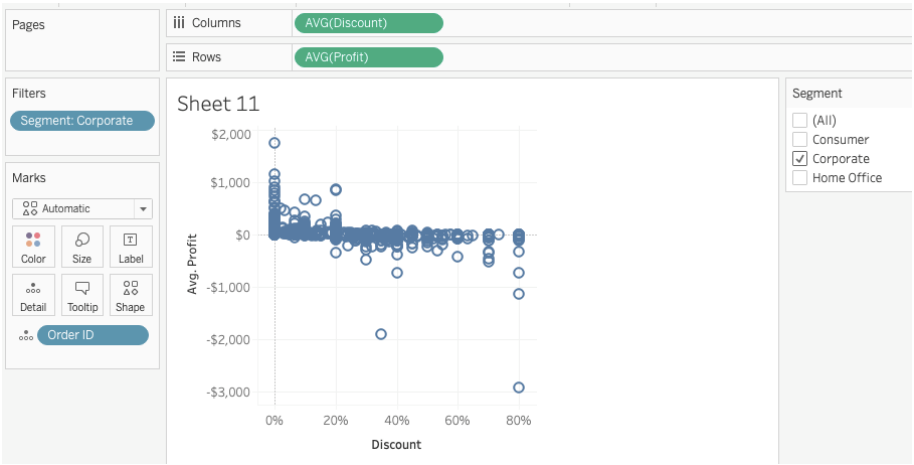
Image: Adds an image to the dashboard.

Webpage: Embeds a webpage in the dashboard.

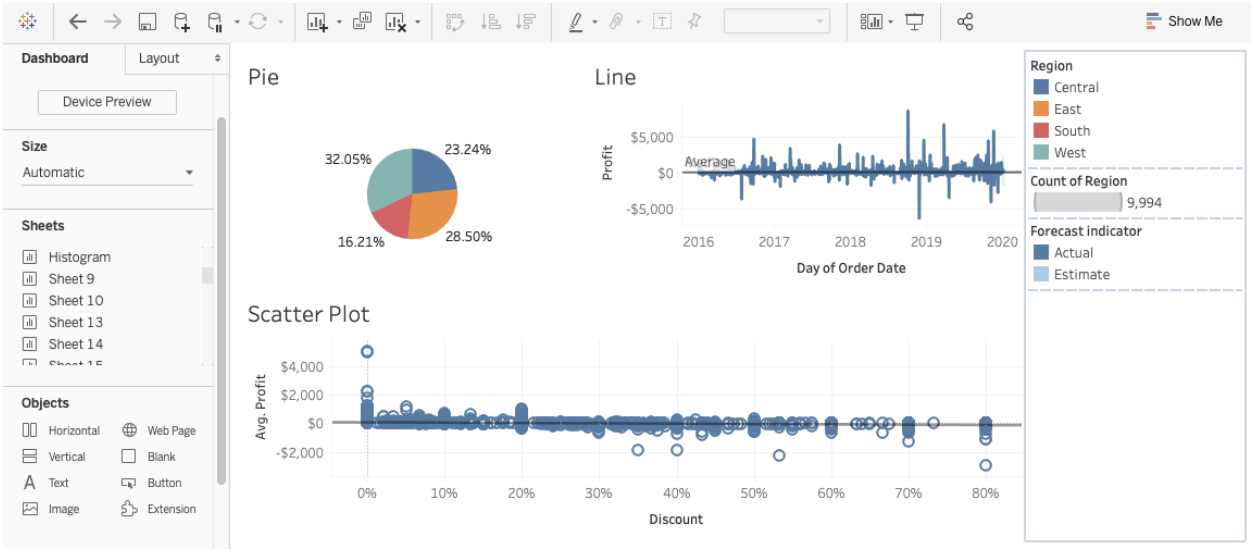
Blank: Adds blank space to the dashboard which can be helpful when elements are too close to each other.

Tiled or Floating: Tiled fills all available space in tiles. With floating, you can control the exact size and location on the dashboard.

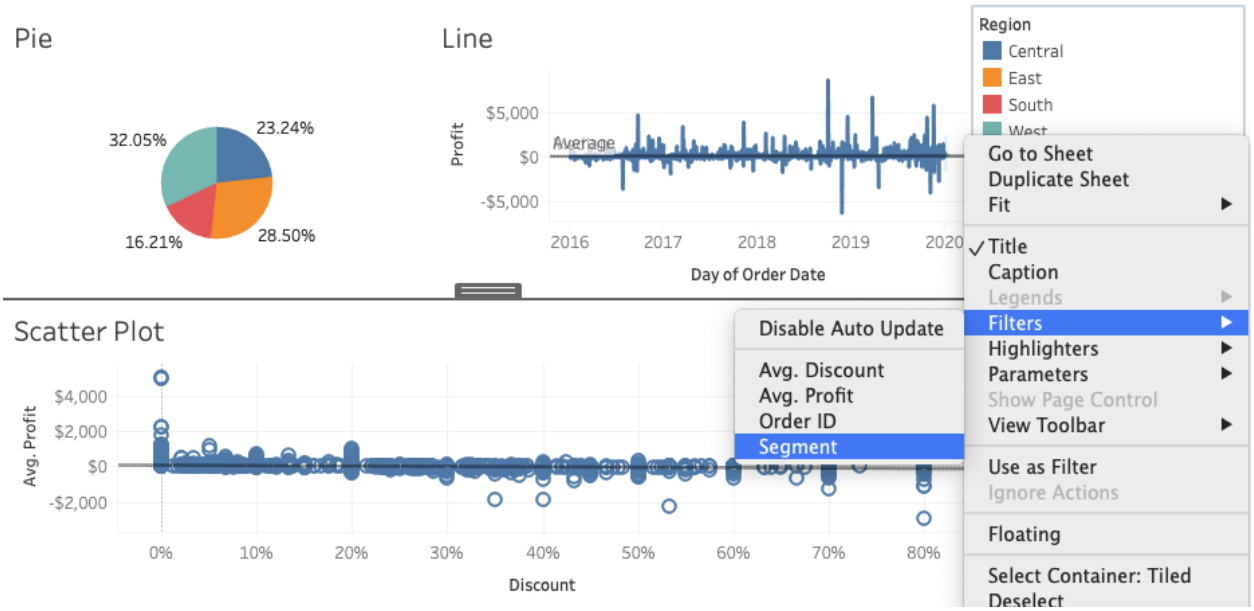
Before populating this new dashboard, add a Segment filter to the Scatterplot worksheet as shown below, and select Show Filter by right clicking the filter. This step requires moving to the scatterplot tab. Now you are ready to populate the dashboard.



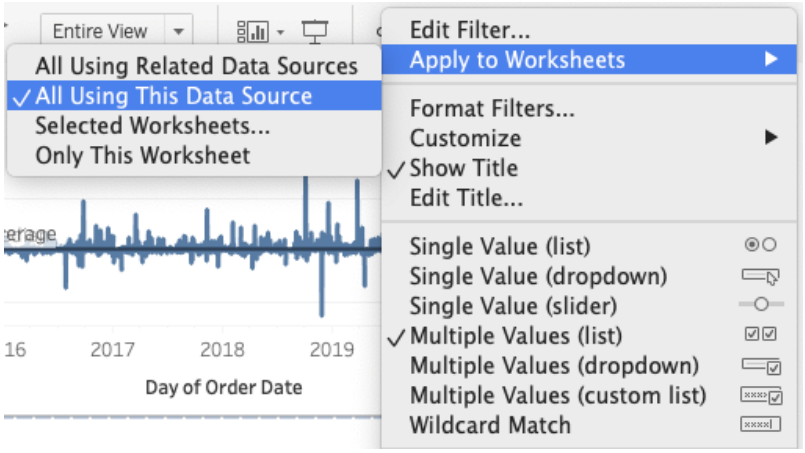
Go back to the tab for the new dashboard. Under the Sheets section, drag in the Pie chart to the dashboard canvas, drag the scatterplot below it, and add the line chart on the top right. The sheets should fit into default spaces on the dashboard. Recall that these are charts you created earlier in Topic 7. The main purpose of a dashboard is to bring together worksheets in an interactive, connected way.



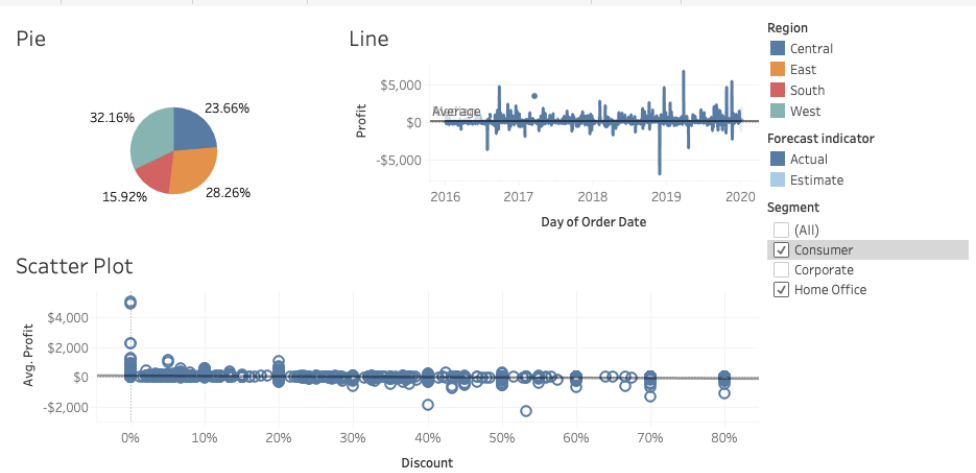
Your dashboard should look something like the example above. Notice how the legends follow the sheets into the dashboard. You can select the dropdown on the count of region legend to exclude it from the dashboard. You can also add the segment filter by selecting the dropdown on the scatter plot (and choose apply to all worksheets using this data source).



Select the new filter's dropdown arrow to make sure the segment filter applies to all sheets.



Make a change to the filter to make sure each sheet in the dashboard responds:



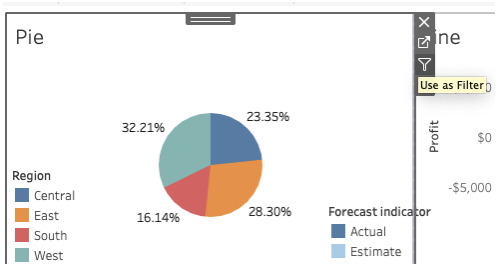
You can make a filter or legend “float” by clicking the legend/filter’s drop-down arrow and selecting Floating. This allows you to arrange them anywhere on the dashboard as opposed to specific slots.

Float the Region color legend over the pie chart by clicking on the down arrow on the legend.

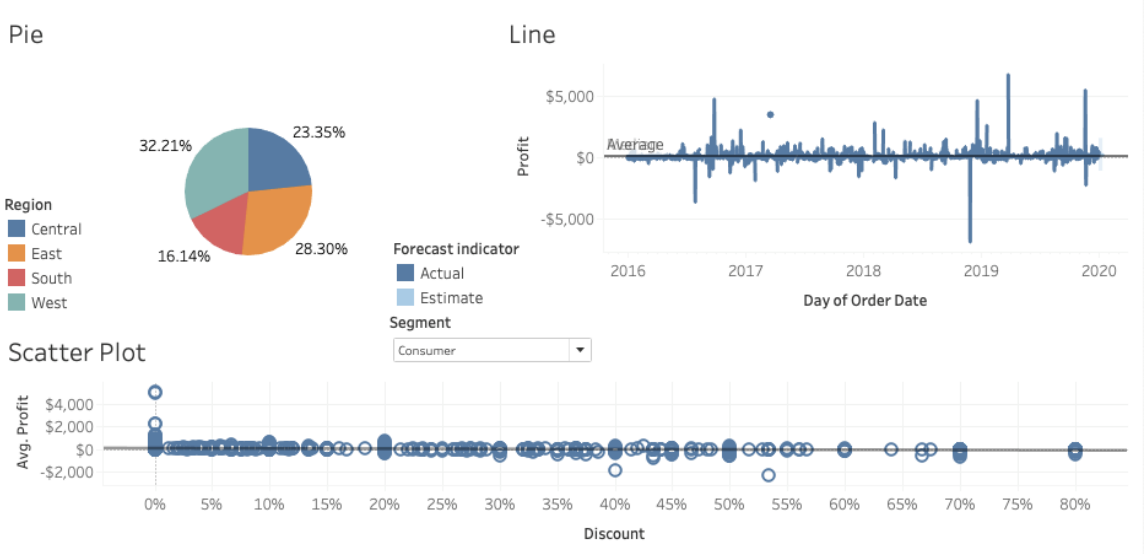
Set the Segment filter to a single-value drop down to save space.

In addition to filters, you can also use the worksheets themselves as filters. For example, I would like to focus on the different sections of the pie chart:

Select the ‘use as a filter’ icon on the pie chart (next to the drop-down arrow).



Click a region in the pie chart and see how the other worksheets change.



Other Formatting Options for Dashboards

Dynamic Titles

The user-friendliness of a dashboard can often be improved by inserting a dynamic value into the view's title.

- On a worksheet in the dashboard, Right Click, Edit Title.
- If you click on Insert, you will get a list of things that can be put in the title. This depends on what fields are in use in the view
- For example, you can set the title to dynamically change as you change a filter.

Format Menu

Select Format…Dashboard from the top menus to open the format pane. You can customize a lot of things here.

- One thing to note is that if you change the shading, say to yellow, only dashboard items are shaded.
- Views have their own formatting, so if you want everything yellow you would need to edit each view individually.

Images, Blank Objects, & Web Pages

- If you have an image (say a logo) that you want to add, you can drag out Image from the left and just navigate to where your image is stored.

    - You also have the option through the menu to make the image be a hyperlink to a URL (link to URL in this case).

- Blank objects are useful if you want to add space to the visualization, push an item to one side, etc.

- If you want to add a web page, you simply bring out the web page part

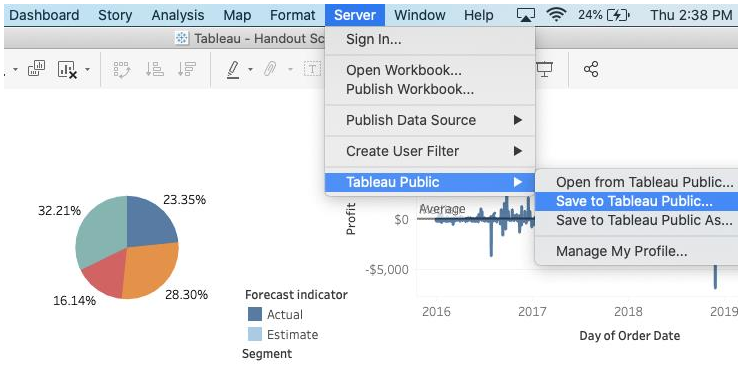
    - For the URL, put msb.georgetown.edu

    - This is treated just like any other part of the dashboard (drag out a worksheet, resize things).

    - And it is still a fully interactive webpage, just like in a browser.

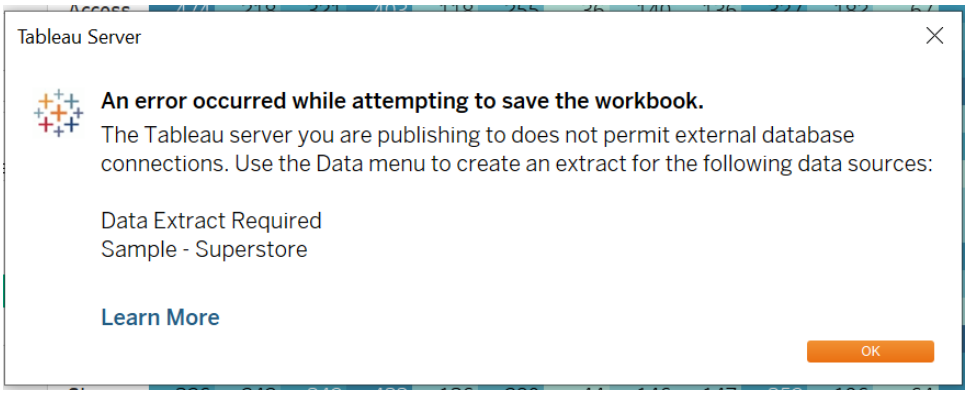
Publishing a Dashboard to Tableau Public Server

In order to publish this dashboard on Tableau Public Server, click on the server menu and navigate to ‘Save to Tableau Public..’ as shown below. You will then be prompted to sign in or create a new account (if you do not have one yet).

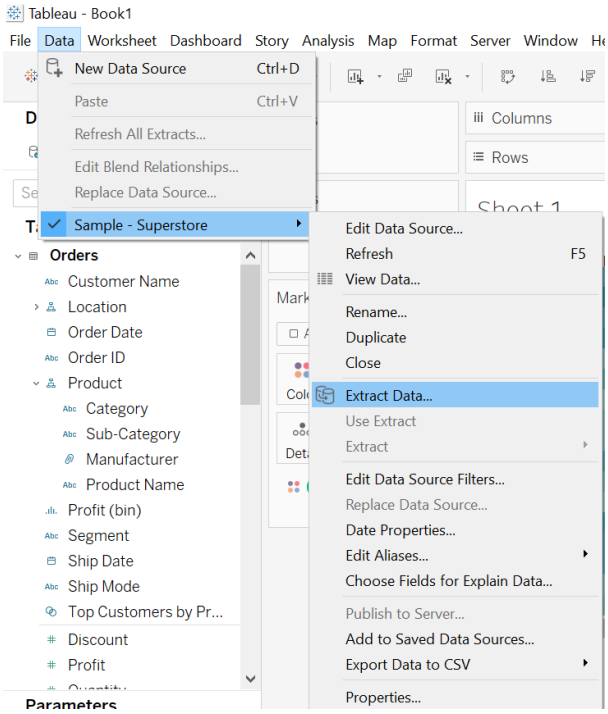


The dashboard is now available online under your individual Tableau profile and its link can be shared with others to view! If the dashboard looks out of sorts in the browser, make sure you are in desktop mode. This can be done by clicking on the laptop icon at the bottom of the Tableau Public site.

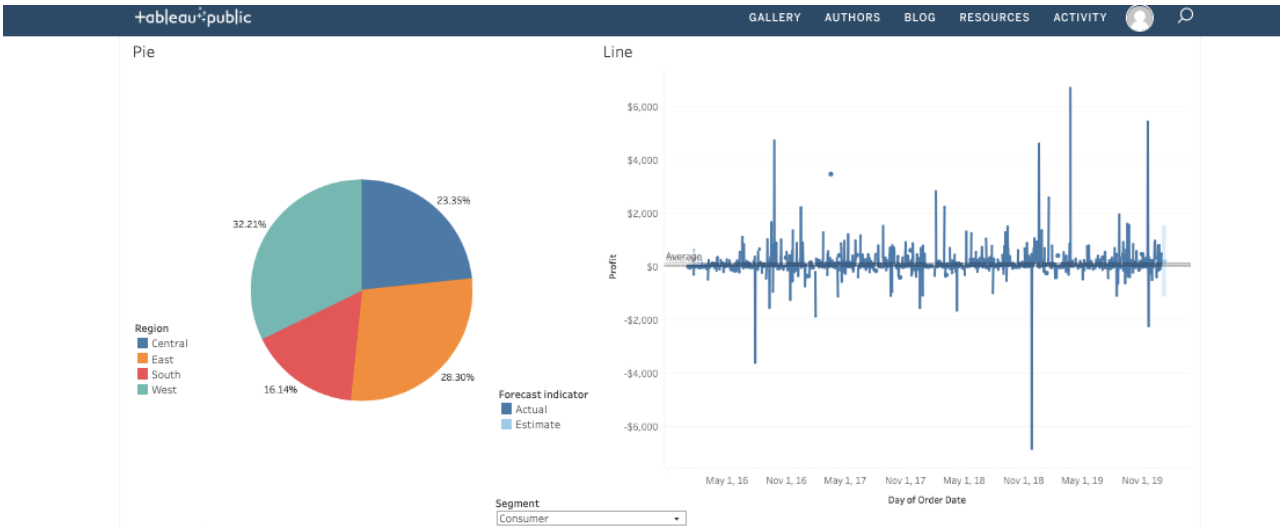
You may get an error at this step when you try to publish your dashboard – it looks like the following:



If you get this error, you need to go to Data and select Extract and check defaults to make sure you extract the relevant data for your dashboard.



Here is an example of what you should see on Tableau Public Server.



Create a Story

In Tableau, a story is a sequence of visualizations that work together to convey information. You can create stories to tell a data narrative, provide context, demonstrate how decisions relate to outcomes, or to simply make a compelling case.

Stories connect a series of facts (thought of as story points) to tell a data narrative, provide context, demonstrate the rationale upon which decisions are based, or simply present a compelling case. Each story point can be based on a different view or dashboard, or the entire story can be based on the same visualization seen at different stages, with different filters and annotations. The methods and tools for designing and formatting sheets and dashboards also apply to stories.

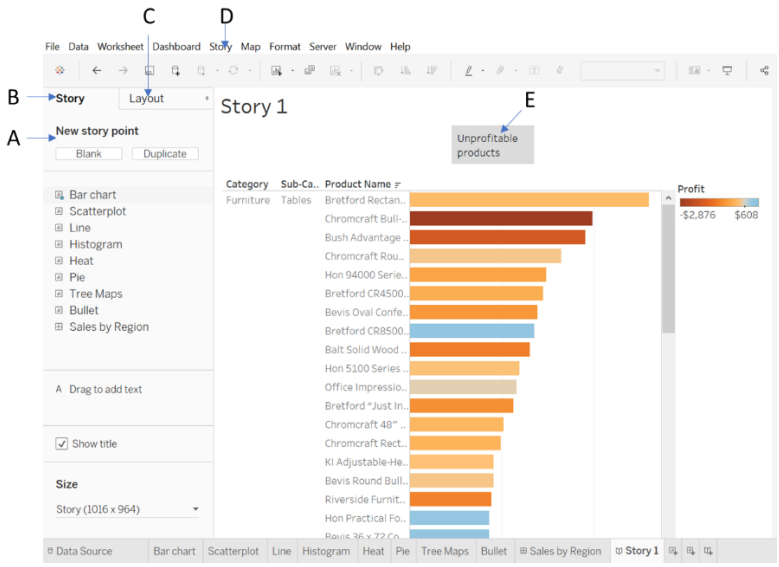
Different types of stories can be told, but they all include characters (the data), a plot (the insights) and a narrative (the visualizations that convey the insights). Story types include the following:

Trend	This type of story uses a chronology to illustrate a trend over time. This type of story highlights *Why did trend happen or keep happening? What can we do prevent or encourage the trend?*
Drill Down	This type of story establishes context so you're your audience understands what is going on in a particular category and emphasizes differences in categories. This type of story highlights *Why is this category (person, place or thing) different? How does the performance of this category*

	compare to others? Why is it different? How can we get to perform like another category?
Zoom Out	This type of story shows how an issue relates to the bigger picture. This type of story highlights *How does the issue compare to the bigger picture? What effect does one area have on the bigger picture?*
Intersection	This type of story identifies important shifts when one category overtakes another. This type of story highlights *What caused a shift? Are the shifts good or bad? How do shifts affect other aspects of business?*
Factors	This type of story explains a subject by dividing it into types or categories. This type of story highlights *Is there a particular category that requires additional attention?, How is(are) the category(ies) affecting important metrics?*
Outliers	This type of story shows anomalies or points out where things are exceptionally different. This type of story explains *Why are there outliers? Are outliers good? Bad?*

Side Note: Stories can be a powerful way to support a decision, but they should not be used to avoid designing a dashboard in which all key insights are presented in a dashboard.

As you work on a story, you can use the following controls, elements, and features.



A. Options for adding a new story point: Choose Blank to add a new point or Duplicate to use the current story point as the starting place for your next point.

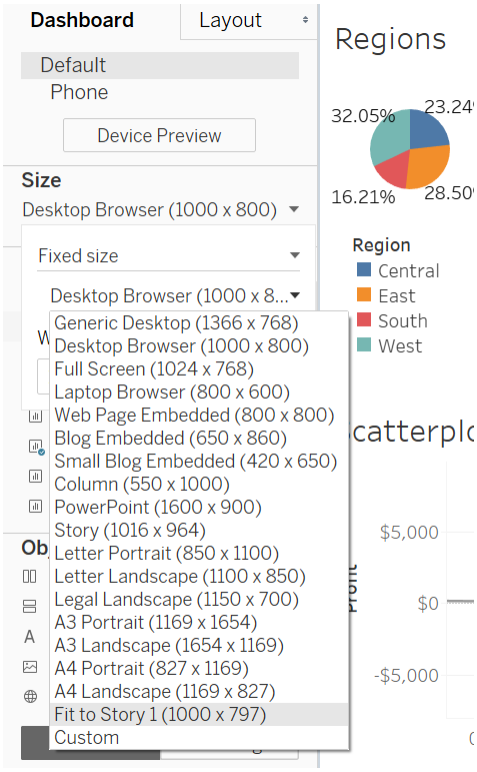
B. The Story pane: Use this pane to drag dashboards, sheets, and text descriptions to your story sheet. This is also where you set the size of your story and display or hide the title.

C. The Layout pane: This is where you choose your navigator style and display or hide the forward and back arrows.

D. The Story menu: Use this menu in Tableau Desktop to format the story or copy or export the current story point as an image. You can also clear the entire story here.

E. The navigator: The navigator allows you to edit and organize your story points. It's also how your audience will step through your story. To change the style of the navigator, use the Layout pane.

Dashboards are a common ingredient in Tableau stories. For dashboards that you plan to include in your story, you can use the Fit to option under Size on the Dashboard pane. It will resize your dashboard so that it is the right size for the story you are creating.



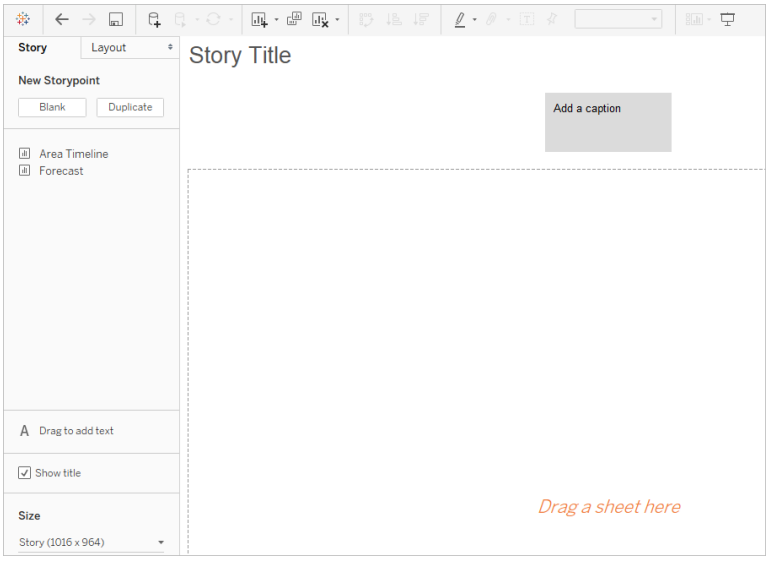
Each story point can be based on a different view or dashboard, or the entire story can be based on the same visualization seen at different stages, with different filters and annotations.

Creating a story point

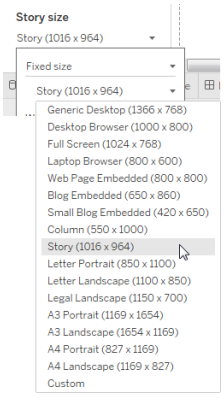
1. Click the New Story tab.

   Tableau opens a new story as your starting point:



2. In the lower-left corner of the screen, choose a size for your story. Choose from one of the predefined sizes, or set a custom size, in pixels:

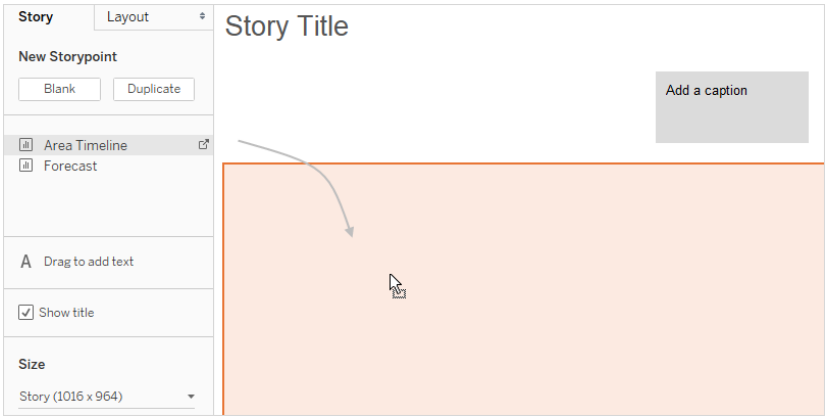


Note: Choose the size your story will be viewed at, not the size you're authoring in.

3. By default, your story gets its title from the sheet name. To edit it, right-click the sheet tab, and choose Rename Sheet.

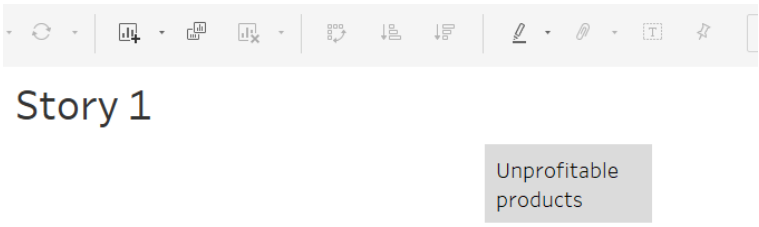
4. To start building your story, double-click a sheet on the left to add it to a story point.

   In Tableau Desktop, you can also drag sheets into your story point.

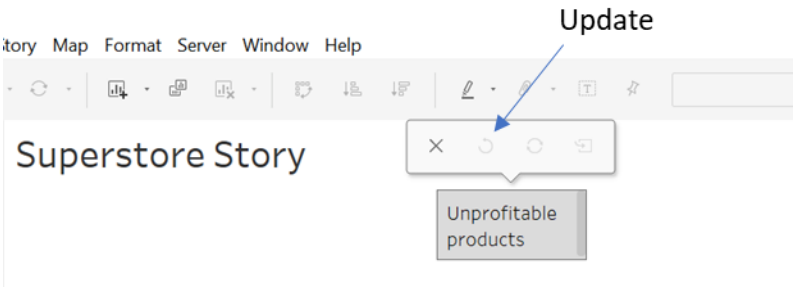


When you add a sheet to a story point, that sheet remains connected to the original sheet. If you modify the original sheet, your changes will automatically be reflected on the story points that use it.

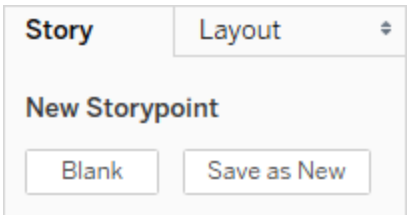
5. Click Add a caption to summarize the story point.



6. To further highlight the main idea of this story point, you can change a filter or sort on a field in the view. Then save your changes by clicking Update on the story toolbar above the navigator box:



7. Add another story point by clicking Blank to use a fresh sheet for the next story point.



Explore layout options

You can refine the look of your story using the options on the Layout tab.

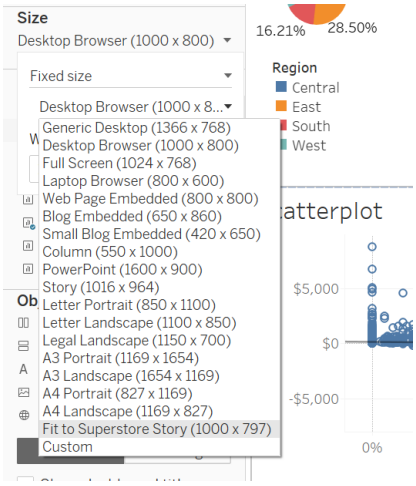
1. Click the Layout tab.

2. Choose a navigator style that best suits your story, and show or hide the next and previous arrows.

Fit a dashboard to a story

You can fit a dashboard to the exact size of a story. For example, if your story is exactly 800 by 600 pixels, you can shrink or expand a dashboard to fit inside that space.

From the dashboard, click the **Size** drop-down menu and select the story you want the dashboard to fit inside.



Delete a story point

Click the X in the toolbar above the point's caption.

Present your story

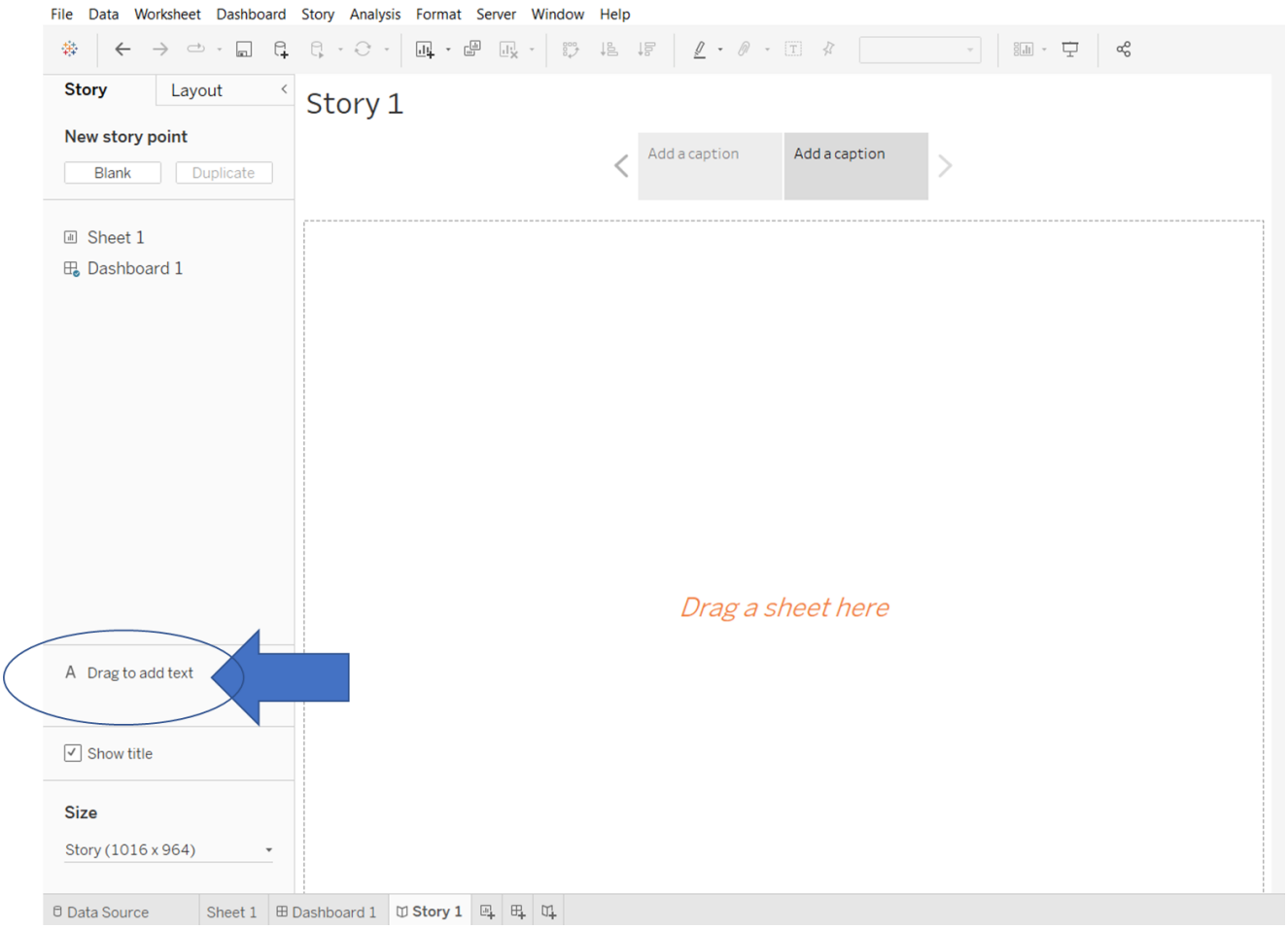
1. In Tableau Desktop, click the Presentation Mode button on the toolbar. Or publish the story to Tableau Online, and click the Full Screen button in the upper-right corner of the browser.

2. To step through your story, click the arrow to the right of the story points. Or, in Tableau Desktop, use the arrow keys on your keyboard.

3. To exit Presentation or Full Screen mode, press **Esc**.

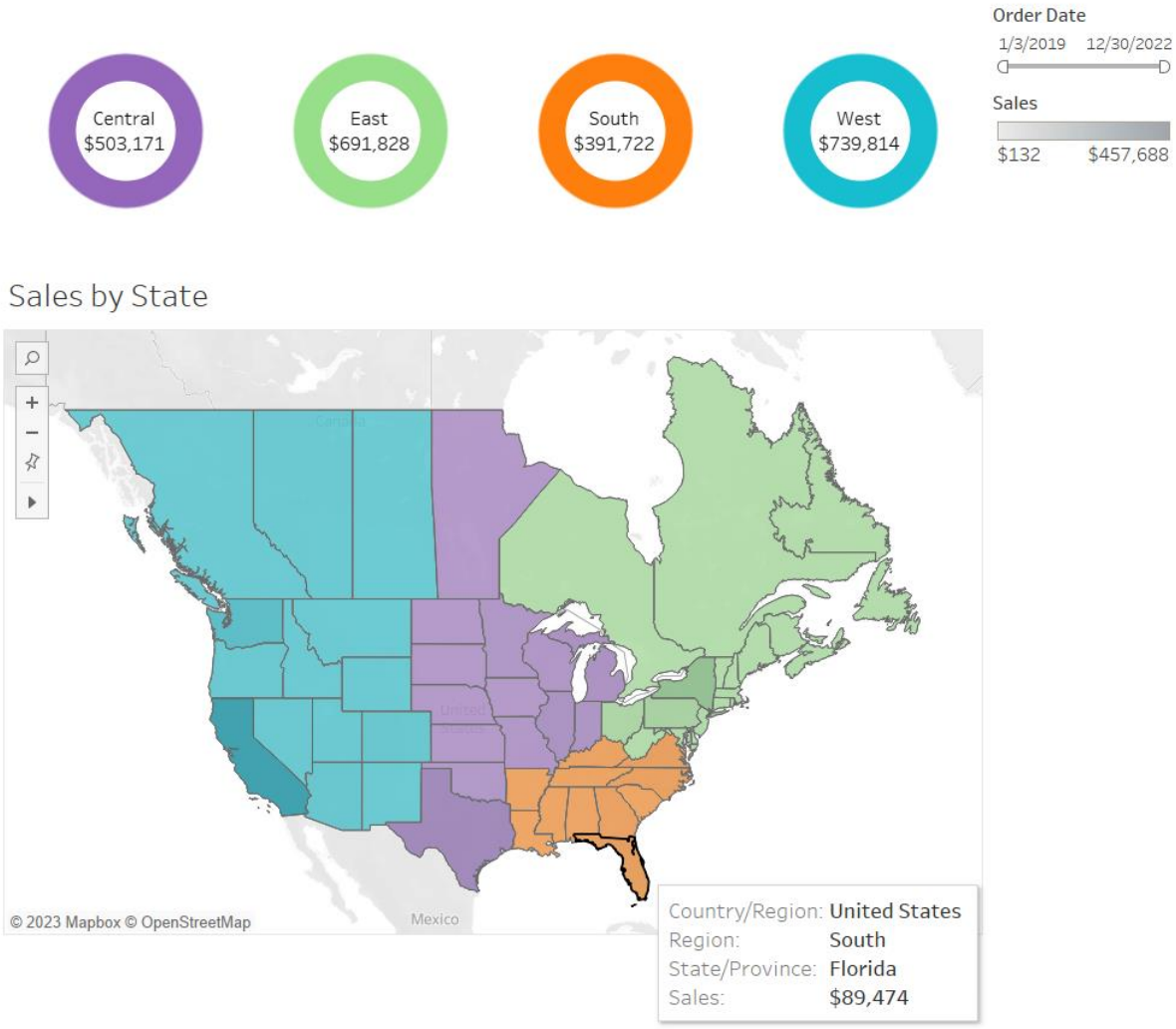
Tips for using a Story to Present

Story points can be used as a presentation. You can add slides with text to provide information besides your visualizations.



More Practice Making a Dashboard

Your goal is to use the skills you have learned to create the following dashboard with the Superstore data:



Hints

Step 1: Make the Sales by State worksheet. This should be straight forward. Make sure you are happy with the colors for the regions.

Step 2: Make the Donuts worksheet.

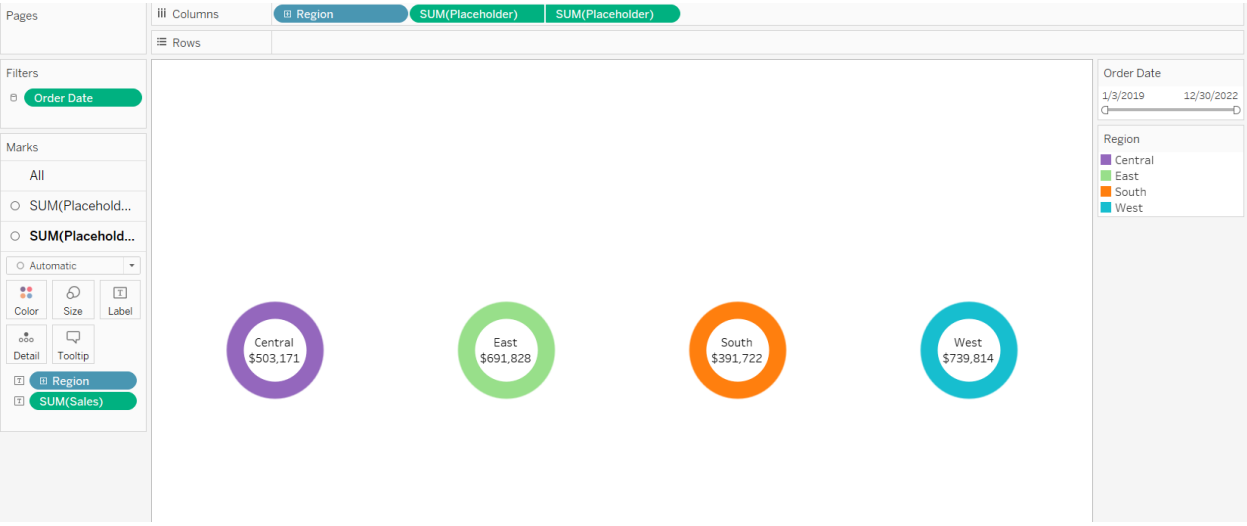
Step 3: You will need a placeholder field (Placeholder: 0).

Step 4: On your columns shelf include Region, Placeholder, and Placeholder

Step 5: Adjust marks cards. First marks card – include Region for color, Sum of Sales for Detail, Make sure marks are Circles, Adjust the size bigger, Edit colors to match Sales by State worksheet. Second marks card – make color white, marks are circles, Sum of Sales and Region as Labels, Adjust size smaller than First card.

Step 6: Right-click 2nd placeholder pill, and choose dual axis.

Step 7: Other formatting: you may need to change from Standard View to Entire View, you may need to adjust format alignment pane to horizontal center, you may need to adjust format borders to make all borders none, you may need to adjust format lines to make all lines none, and format sales as currency. Also, need to unselect Show Headers.



Step 8: Use order date as a filter and show filter.

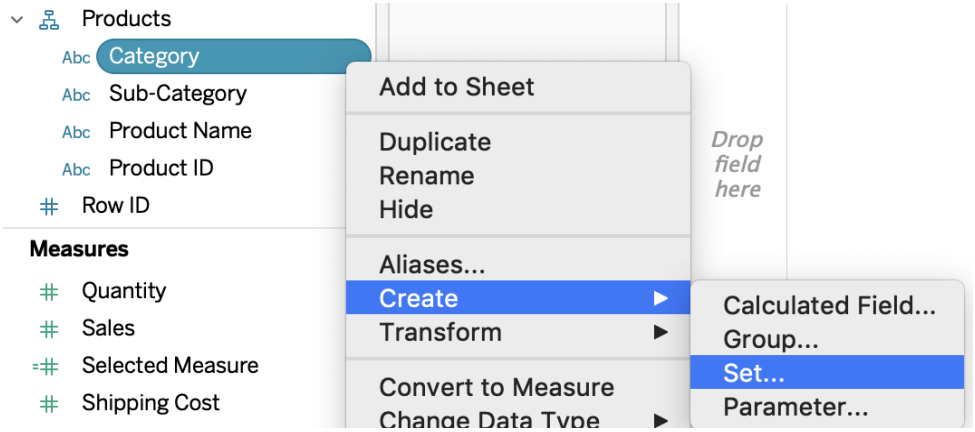
Topic 13: Advanced Tableau Examples

This topic provides some examples of advanced Tableau functionality. This topic uses Superstore data.

Asymmetric Drill Down Using Set Actions

Drilling down usually expands everything from one level to another. But that is not always what you want. Set Actions are a great way to drill down only in one part of the visualization.

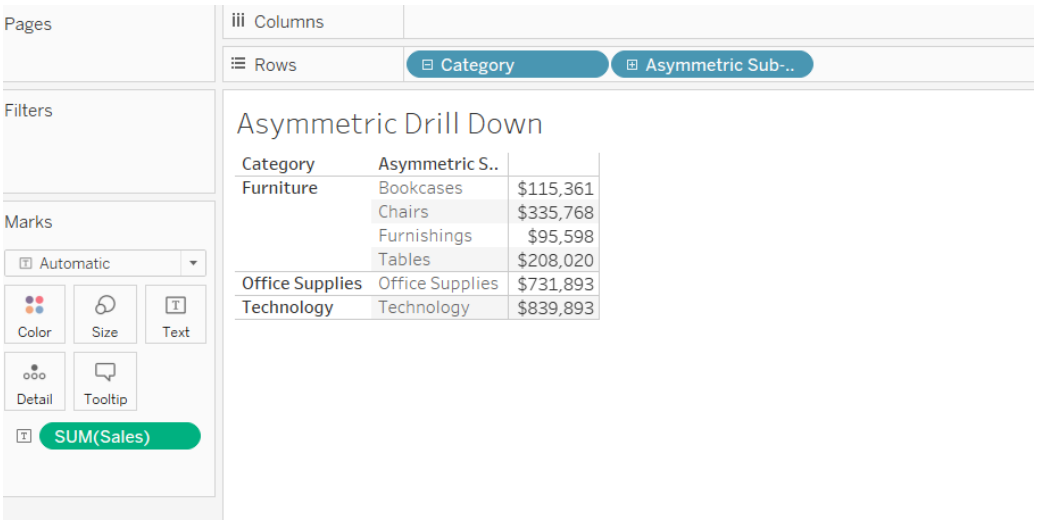
1. Create a set for the Category field named Category set. Right click on Category > Create > Set



2. Choose Furniture to be a member of that set. The membership is temporary and will be overwritten by the set action.
3. Create a calculated field named Asymmetric Sub-Category.

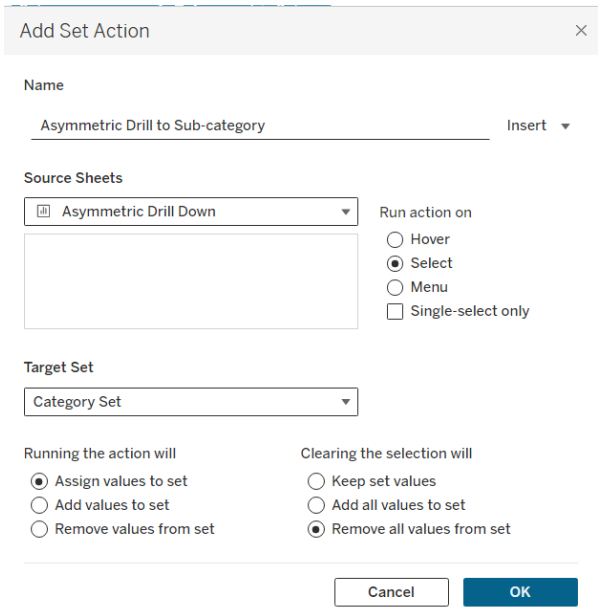
```
IF [Category Set]
    THEN [Sub-Category]
    ELSE [Category]
END
```

4. If the category of the mark or marks in the view is in the Category Set, the calculation will return the Sub-categories for that Category. Otherwise, it will return the Category again. The set isn't in use in the view yet and has no action associated with it, so it is static for the moment.
5. Bring **Asymmetric Sub-Category** below the Sub-Category pill in the data panel
6. Build the view by dragging the **Category** and **Asymmetric Sub-Category** to Rows (in that order). Drag **Sales** to Text in the Marks card.

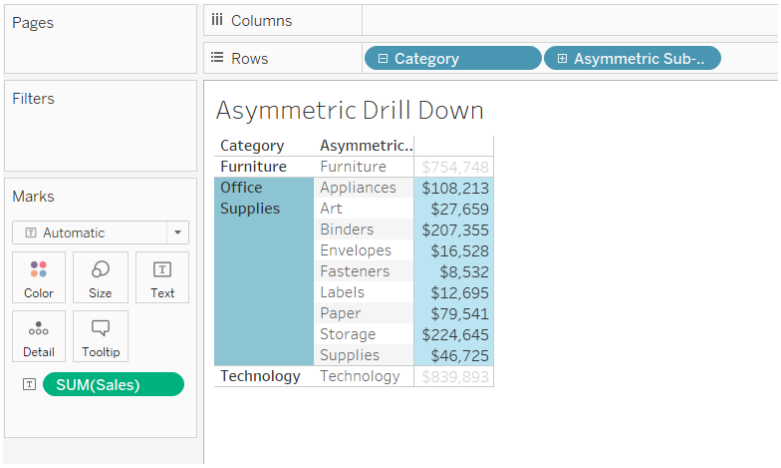


7. Create a set action.
    a. On the Worksheet menu, select Actions. Click Add Action, and then select Change Set Values.

    b. Name the action Asymmetric Drill to Sub-Category and select the following options:
        i. Source Sheets: Make sure the only the current sheet is selected.
        ii. Run action on: Choose Select. This will make the action apply when the user selects a mark or marks in the view.
        iii. Target Set: Select the current data source from the first drop-down list and the Category set from the second drop-down list.
        iv. Clearing the selection: Select Remove all values from set.
        v. Click OK to save your changes.



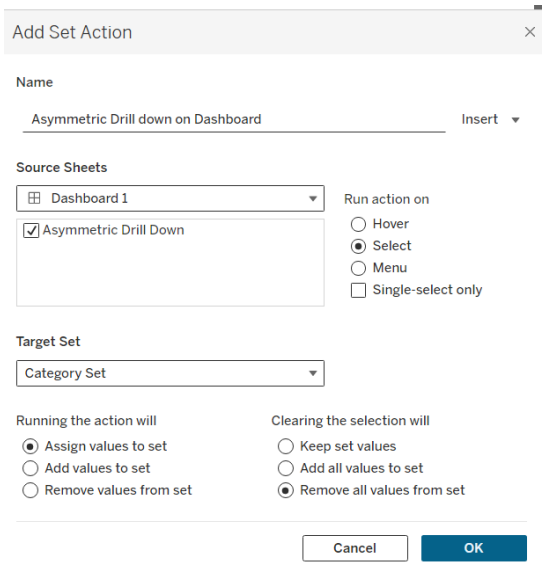
When you click on a header or mark in the view, it will expand out to the next level. Clicking a category bar will expand its sub-categories.



The example showed how to make a Worksheet Dynamic but if you use that worksheet in a Dashboard, the dynamic actions do not (always) automatically convey. You may need to set up the actions on the Dashboard for the dynamic behavior to appear there.

9. Create a Dashboard with this Worksheet on it. Select Dashboard from the Menus and select **Actions**. Select **Add Action.**

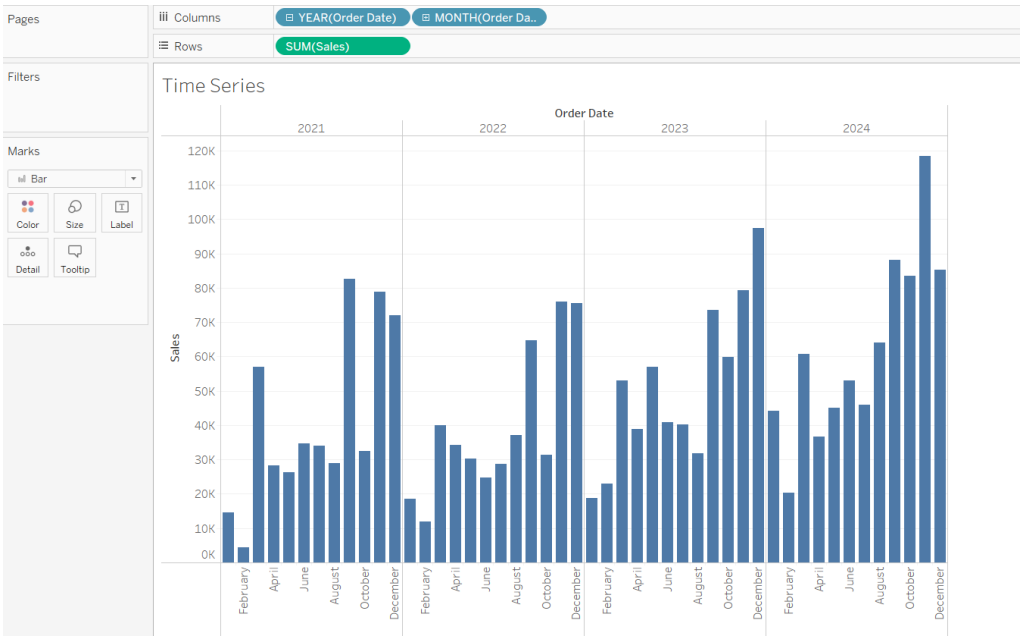
10. The settings are similar to those used to set up the action on the worksheet except that the Dashboard is the Source Sheet with the linked Worksheet (here named Dynamic Drilldown) selected.



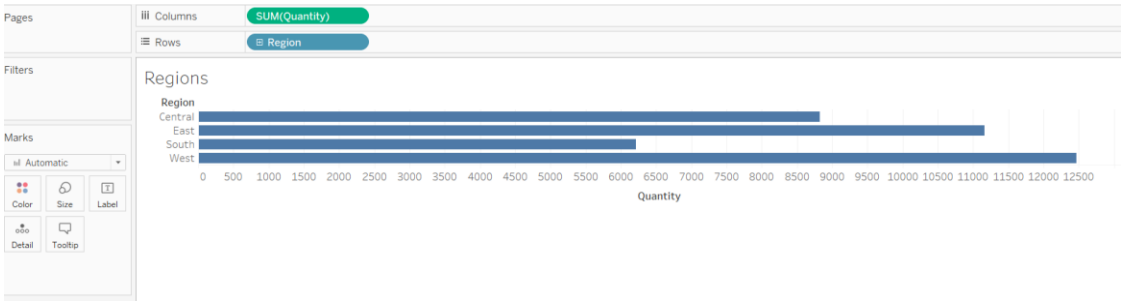
11. If the Dynamic Drilldown on the Dashboard did not work before, it should now.

Another Dynamic Dashboard Example Using Set Actions

1. Create the first visualization: a time series of sales by month over time. Drag order date to the columns shelf and expand to months. Drag Sales to the row shelf. Change visualization to a bar graph.

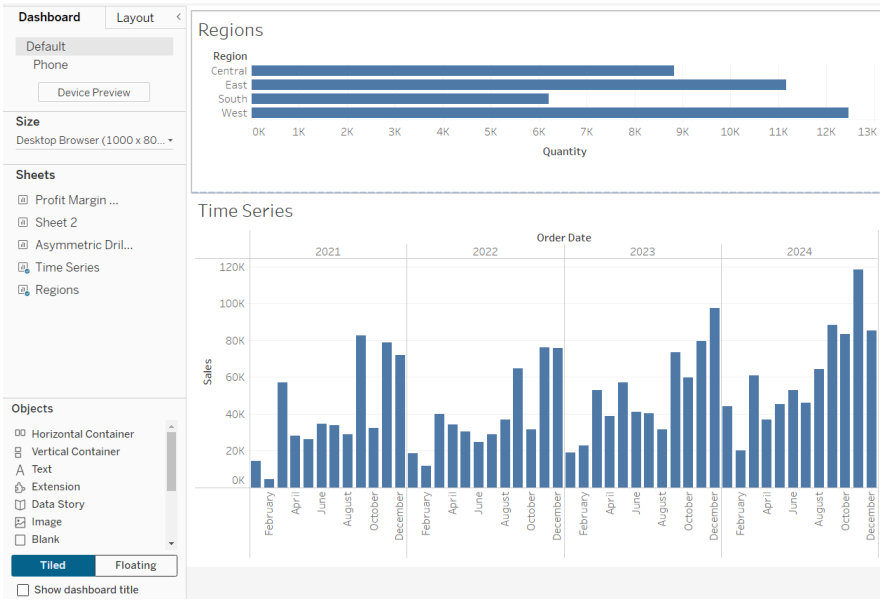


2. Create the second visualization that shows quantity by region. Add region to rows and quantity to columns.



3. Use Region to create a set and call it **Region Set**. You can leave the set empty for the moment.

4. Create a Dashboard with the Time Series visualization on the bottom and the Quantity by Region visualization at the top.

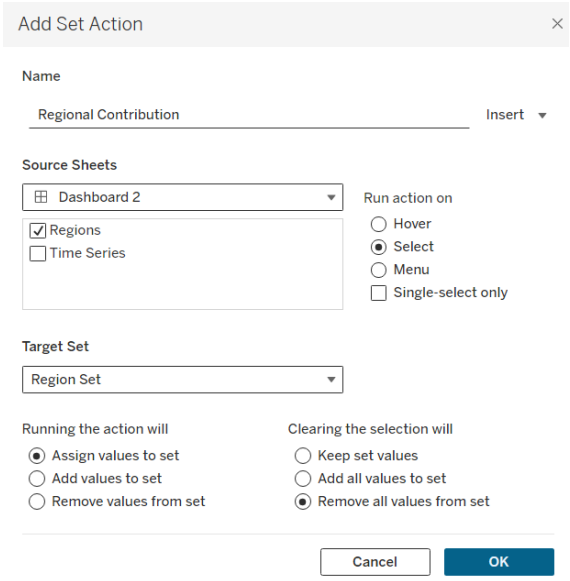


5. We will use the Region to change what is in the Time Series by showing which region contributed how much to sales in a particular month.

6. On the Time series worksheet and drag **Region Set** to color. If nothing is in the set, everything will look gray.

7. On the Dashboard, from the Dashboard menu, selection **Action**, and **Add an Action** to create a new action for **Change set values**.

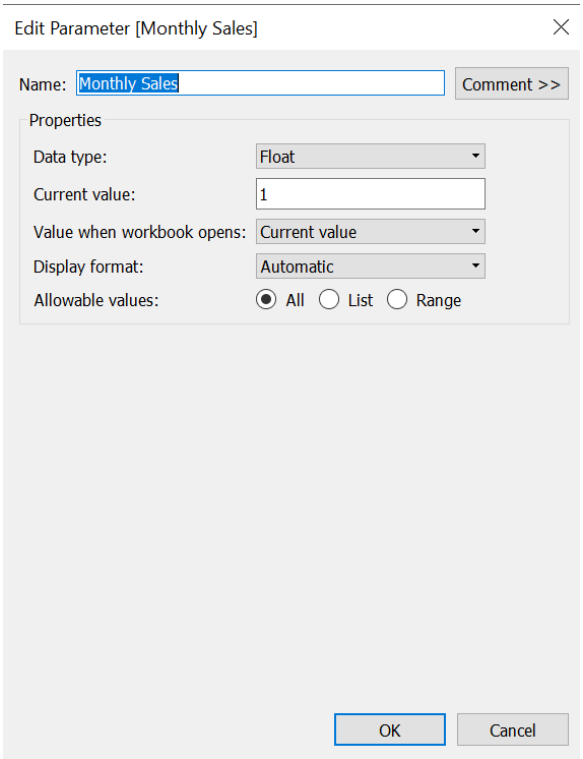
8. The Source Sheets selects which sheet will ‘drive’ the dynamic action. Select **Regions**. Make sure the **Target Set** Set is the Region set. The other options are similar to the previous example.



9. When you select one or more regions from the top visualization, the coloring on the Time Series should change to reflect the contribution of the region. My feeling is the coloring is reversed. Return to the Time Series worksheet, right click on the Region Set in the Color Marks, and select **Sort**. Change to Descending and then re-click Ascending. Now the coloring should be at the bottom of the bars rather than the top.

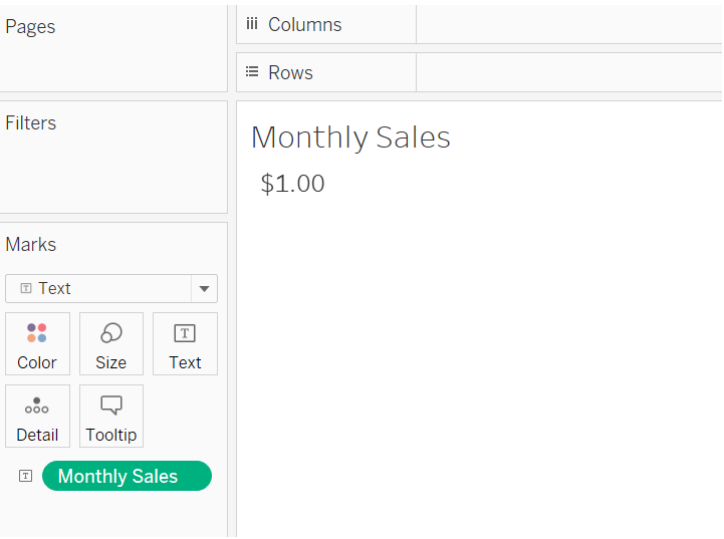
Parameter Dashboard Actions

To demonstrate how parameter actions work, first create a parameter with a data type Float and allowable vales of All.



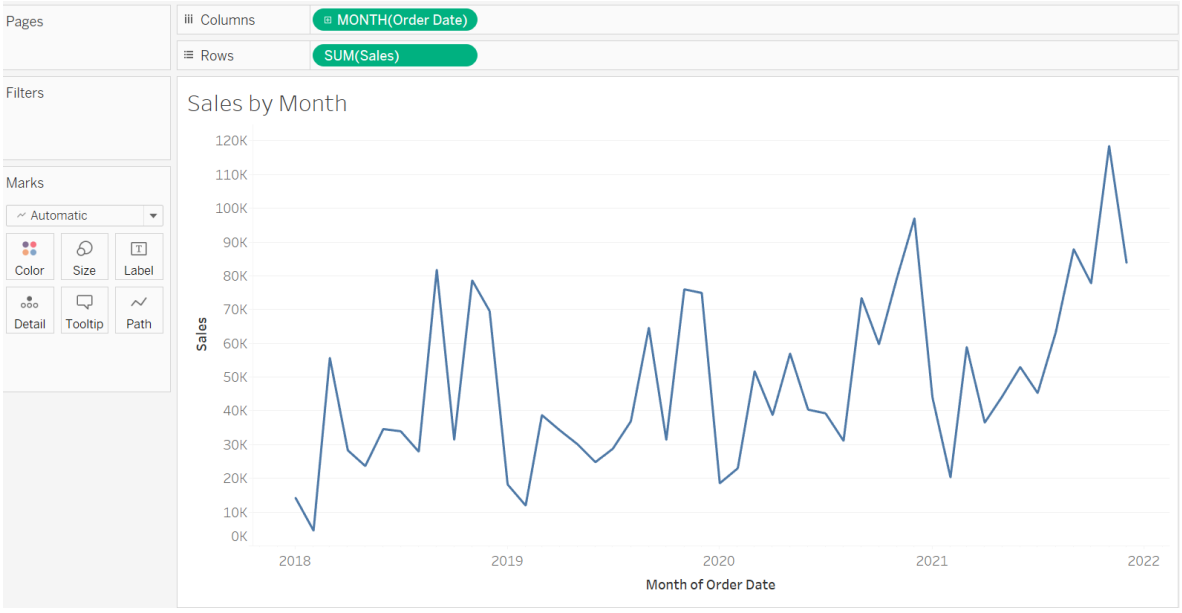
These settings mean that this parameter can be populated with *any* number including those with decimals. Then show the value of the new parameter as Text.

(NOTE: I could only show the Parameter to the Text Marks Card, if I first added another field such as Profit. Once I added the Monthly Sales parameter, I could remove Profit and it would stay. If anyone can figure out a better way to do this or explain why this is happening, that would be great. Thanks.)

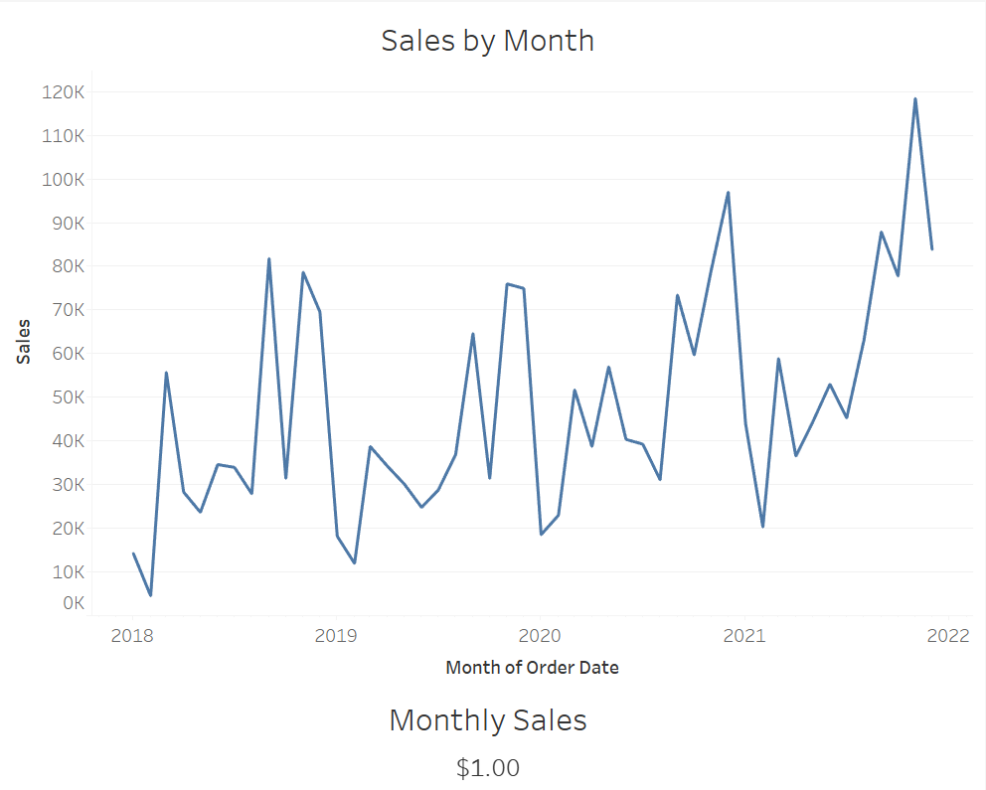


I also changed the title to Monthly Sales and changed the formatting of the parameter to currency with 2 decimal places.

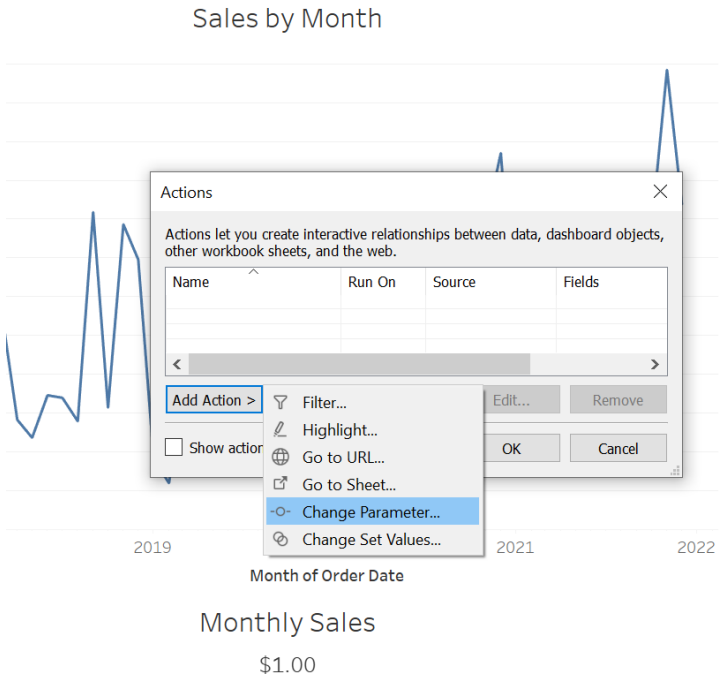
Next, set up a line graph that looks at the measure of Sales in the sample data by continuous month of the Order Date.



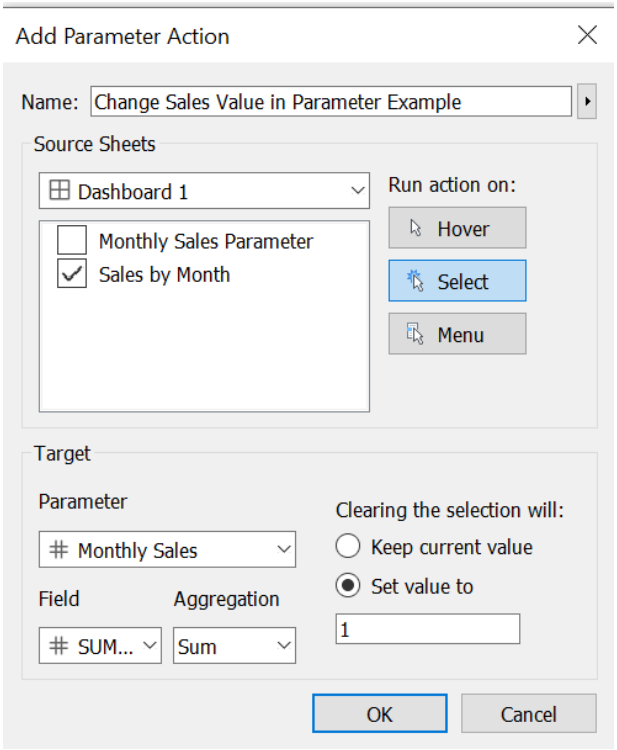
And then place both sheets on a dashboard together. If the Monthly Sales Parameter is not displaying well, select the More options for the pane and choose Fit Width. You may also need to adjust the Formatting of the Monthly Sales Parameter including the Alignment.



To add a parameter action, click “Dashboard” from the navigation menu and select Actions. Then click, the Add Action button, and choose Change Parameter.



Then, adjust the Parameter Action Example parameters:



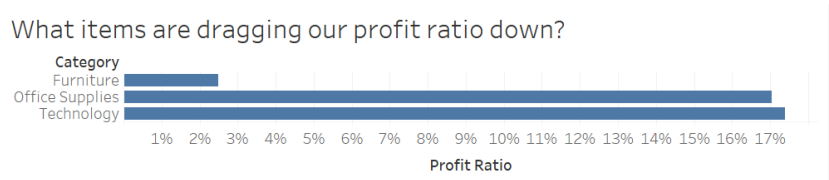
Check if it works. Clicking on a value within the line graph should overwrite the current value of the parameter which is being displayed in the text sheet. Since the aggregation of our parameter action is SUM, clicking more than one data point will add them together.

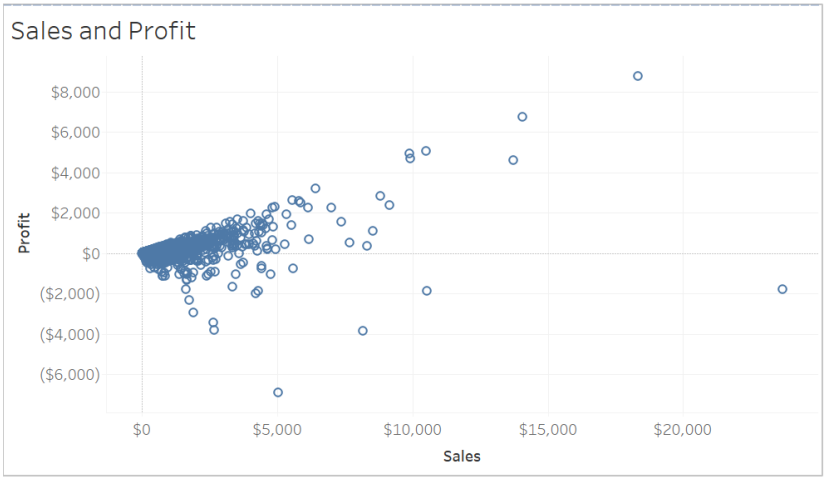
Highlight Actions

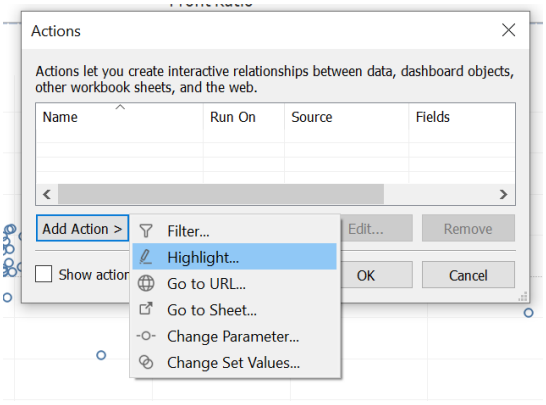
Create the worksheets and then the simple dashboard that will show Profit Ratio by Category and Profit by Sales.

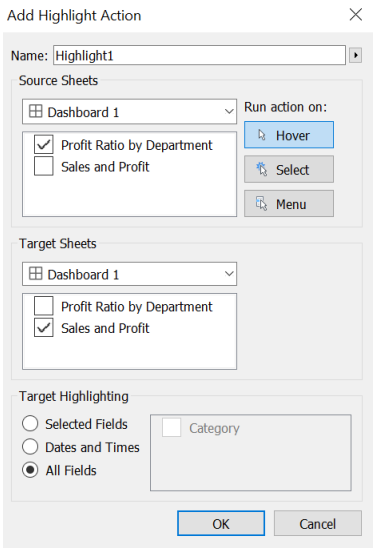
Then select Actions from the Dashboard menu, and choose Add Action, and select Highlight.

You want to be able to hover over a category in the bar chart it and highlight the corresponding items in the scatter plot.


What items are dragging our profit ratio down?
Category
Furniture
Office Supplies
Technology
1% 2% 3% 4% 5% 6% 7% 8% 9% 10% 11% 12% 13% 14% 15% 16% 17%
Profit Ratio


Sales and Profit
$8,000
$6,000
$4,000
$2,000
$0
($2,000)
($4,000)
($6,000)
Profit
$0
$5,000
$10,000
$15,000
$20,000
Sales


Actions
Actions let you create interactive relationships between data, dashboard objects, other workbook sheets, and the web.
Name
Run On
Source
Fields
Add Action >
Filter...
Highlight...
Go to URL...
Go to Sheet...
Change Parameter...
Change Set Values...
Edit...
Remove
Show actior
OK
Cancel


Add Highlight Action
Name: Highlight1
Source Sheets
Dashboard 1
Profit Ratio by Department
Sales and Profit
Run action on:
Hover
Select
Menu
Target Sheets
Dashboard 1
Profit Ratio by Department
Sales and Profit
Target Highlighting
Selected Fields
Dates and Times
All Fields
Category
OK
Cancel

If you try it, it doesn’t work 😟, instead everything is grayed out. The problem is that we are trying to highlight items in the scatterplot by what category they belong to, but we didn’t use the Category field anywhere in that worksheet. If you add Category for example to the Tooltip, this will highlight the corresponding data in the scatterplot when the user hovers over the Category in the bar chart.

Topic 14: Tableau and R (Expert Level)

R is a popular statistical language used to perform sophisticated statistical analysis and predictive analytics, such as linear and nonlinear modeling, statistical tests, time-series analysis, classification, clustering, etc.

Tableau Desktop can connect to R through calculated fields and take advantage of R functions, libraries, packages and even saved models. These calculations dynamically invoke the R engine and pass values to R via the Rserve package and are returned back to Tableau. This feature is primarily targeted for users who are already proficient at R. It is NOT meant for beginners with R. Anyone who wishes to use the new functions must first learn how to use R in order to leverage its capabilities in Tableau.

Basic Steps

1. Download and Install R. (If you don't already have R and are familiar with it, this section is not for you.)

2. Install the Rserve package in R for Tableau to connect to.

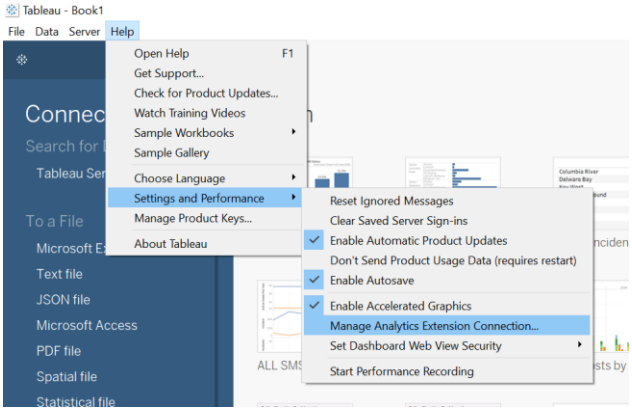
In the R console, enter the following commands:

```
install.packages("Rserve")
library(Rserve)
Rserve()
```

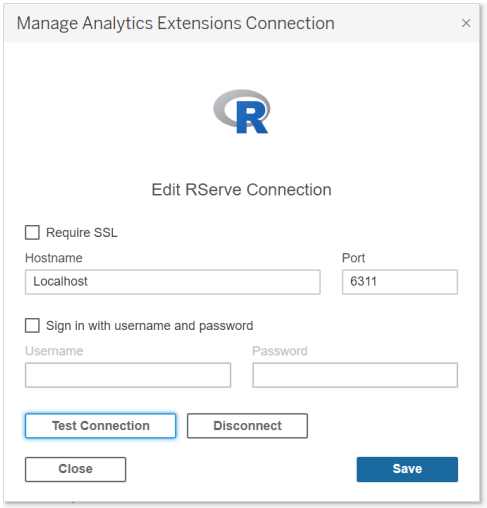
Note: after the first time you install Rserve, you will only need to use lines 2 and 3 to open Rserve.

3. Connect Tableau to the R Server. Once Rserve is installed, open Tableau Desktop and follow the steps below.

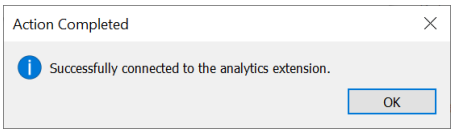
a. Go to the Help menu, "Settings and Performance" and select "Manage External Service Connection".



b. Select Rserve and then enter a server name of “Localhost” and a port of “6311”.



c. Click on the “Test Connection” button to make sure everything runs smoothly.

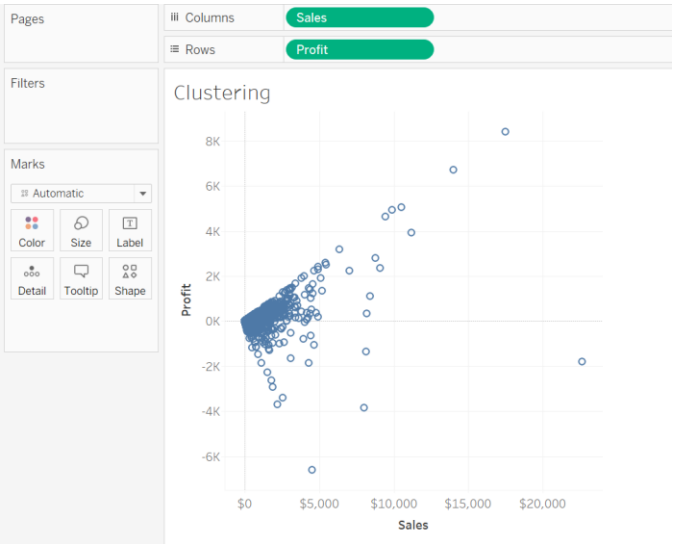


d. You should see a successful message. Click OK to close.

4. Start using the R scripts in Tableau. Now you will be able to create new calculated fields in Tableau Desktop that utilize the SCRIPT_* functions to make R functional calls.

Example: Clustering (K means)

Clustering means combining (clustering) data into similar groups based on available variables. We will cluster products from the Superstore dataset that have similar sales and profits.



If you create the above chart and see only one point, remember to unselect Aggregate Measures from the Analysis menu.

There are four new built-in functions that are used to call specific R models and functions. The functions are:

- `SCRIPT_REAL`
- `SCRIPT_STR`
- `SCRIPT_INT`
- `SCRIPT_BOOL`

These functions are distinct only in the type of result they return: a real number, a string, an integer, or a Boolean.

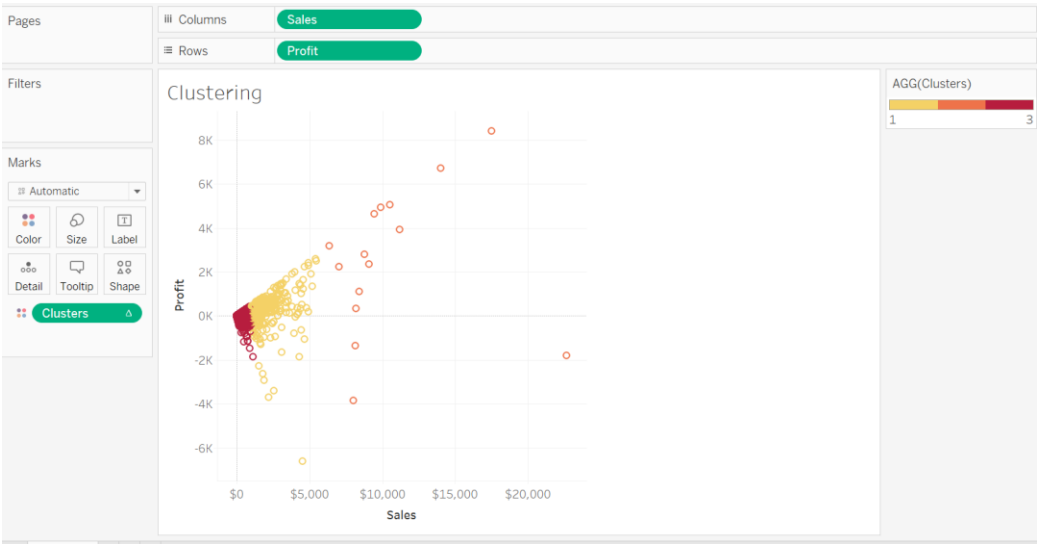
The arguments you pass into each of these functions include R-language scripts and function calls. You can pass 1 or more arguments to R, which are then passed dynamically via Tableau. For instance, you can pass the sales of each customer on a visualization.

For example, the SCRIPT that will use the R clustering function to cluster data points on sales and profit is:

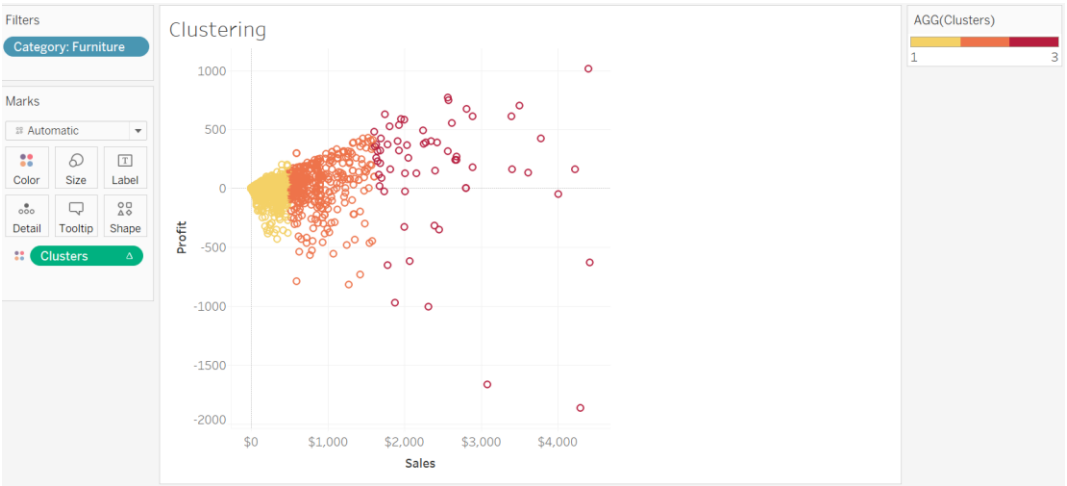
```
SCRIPT_INT('result <- kmeans(x = data.frame(.arg1,.arg2), 3, iter.max
= 10)
  result$cluster',
  SUM([Sales]),SUM([Profit])
)
```

Create a Calculated Field called Clusters and use the above script. This will create 3 clusters iterating 10 rows at a time to prevent the program from crashing.

Then drag Clusters to Color to color the 3 clusters. You may want to adjust the color scheme of the clusters.

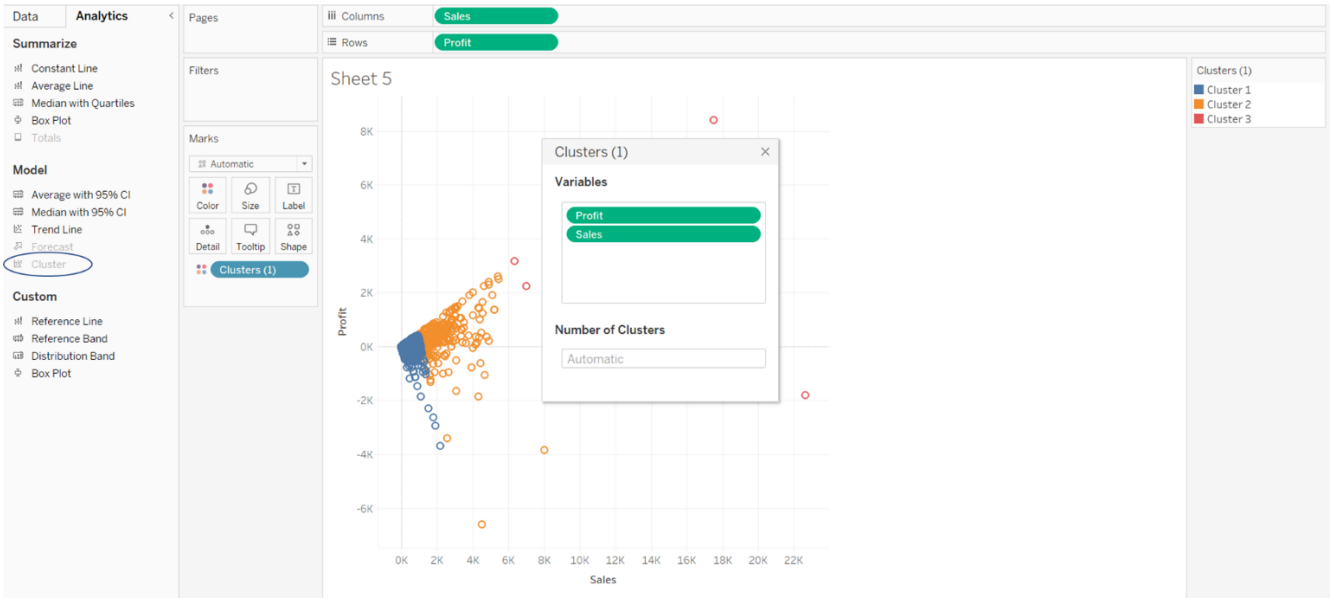


Try filtering the data, for example, only the Furniture Category. The clusters automatically recalculate.



This is just an example to show you what is possible and not a comprehensive tutorial on the multiple ways to incorporate R calculations in Tableau. That is for you to explore on your own.

You can compare this example to the Clustering available with the built-in Analytics tool.



Course Conclusion

After successfully completing the content in this document, students should be well versed in data visualization, comfortable with all of the core functionality in Tableau, and prepared to successfully become Tableau certified https://www.tableau.com/learn/certification. Note: while all the content covers all the concepts on the exam, this is not a Tableau certification prep course.