Intermediate Tableau Workshop

Digital Media Commons

Fondren Library Basement B42

dmc-info@rice.edu

(713) 348-3635

http://dmc.rice.edu

Intermediate Tableau Workshop Handout

Jane Zhao

janezhao@rice.edu

This handout was developed with Tableau 2021.3,

last updated on Tuesday, November 16, 2021.

Note: This handout assembles some materials from Tableau's free training videos
https://www.tableau.com/learn/training, no claims on its originality for that part.

Objectives

Participants will be able to

1) use Data Interpreter to clean data
2) do Join and Union, the two basic ways to merge data
3) create dual axis chart and scatter plot
4) use measure names and measure values to build views with multiple measures
5) create calculated fields
6) use basic quick table calculations
7) plot geographic data on a map
8) create and use parameters in filters and calculated fields
9) add analytics objects to the view

Outline

1. Workshop objectives
2. Intermediate Tableau
     I. Data preparation for Excel files
     II. Merge data from different data sheets/datasets
     III. Dual axis chart and scatter plot chart
     IV. Compare multiple measures on the same chart and build a text table with multiple measures
     V. Calculated fields
     VI. Table calculations and quick table calculations
     VII. Mapping your geographic data
     VIII. Using parameters
     IX. Adding analytics objects to the view

I. Data Preparation for Excel Files

Ideally, the Excel file should be prepared as the rules specified below.

      a. The first row of data only includes column headers.

      b. The sheet doesn't contain any comments/notes and column/row totals.

c. Each column of data contains the same type of data throughout the column and each row of data is one data item.
d. The sheet/table is row oriented. That means variables are stored in the row values. In contrast, in column orientated sheet/table, the variables are stored as column headers.

When data is formatted in column-oriented sheet/table, **the variables are stored as column headers. It is easy for us to read. Tableau cannot aggregate and compare data in this format, because each column is a unique variable/measure, the values for different columns will not aggregate as a whole for us to work with later.**

Column Orientated Table A. e.g. The number of workshops each employee offered in the past 3 years

id	x	y	z
1	a	b	c
2	d	e	f

id	2017	2018	2019
1	30	35	39
2	26	38	41

Tableau Desktop is optimized for row-oriented data. In a row-oriented table, the variables are stored in the row values.

Row Orientated Table B. e.g. The number of workshops each employee offered in the past 3 years. (The year is a single column/variable/dimension and the number of workshops is a single column/variable/measure.)

id	n	v
1	x	a
1	y	b
1	z	c
2	x	d
2	y	e
2	z	f

id	n	v
1	2017	30
1	2018	35
1	2019	39
2	2017	26
2	2018	38
2	2019	41

If your data is not in the optimal Tableau format, you can connect it to Tableau Desktop and use Data Interpreter or other Tableau's built-in configurations to clean it or reformat it. In other cases, you might need to edit the Excel data directly before connecting it to Tableau Desktop. See Tips for Working with Data for more details.

Activity 1.1 Use Data Interpreter to clean data, use Pivot function to convert the data to a row-oriented table, use Split function to split a field.

1) Open the Excel file *data_prep_-_flights.xlsx* with Excel.
   a. The *Resolved Incidents* sheet has extra information other than raw data.
   b. The *Resolved Incidents* sheet shows the number of resolved incidents per Employee per month. The table is wider. This is a typical column orientated table, which is not in the optimal Tableau format.

c. The *Ideal* sheet shows a "Date" column and each row contains the number of resolved incidents for each unique combination of date and employee. This data is taller, with more rows, rather than wider, with more columns. This is a typical row orientated table, which is the format Tableau prefers for optimal analysis.

2) In Tableau, connect to the Excel file data_prep_-_flights.xlsx, drag out the *Resolved Incidents* sheet to the canvas area.

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Flights Data Summary	null	null	null	null	null	null	null	null	
This report was gene...	null	null	null	null	null	null	null	null	
Employee	1/1/2015	2/1/2015	3/1/2015	4/1/2015	5/1/2015	6/1/2015	7/1/2015	8/1/2015	9/1,
B-002	4	1	5	2	3	0	3	1	2
E-055	1	2	1	3	4	1	4	0	2
E-075	14	17	16	15	18	16	14	17	12
B-066	4	4	5	2	5	0	0	2	0
C-025	17	13	17	18	17	17	12	15	17
E-030	2	2	1	1	0	3	5	5	0
C-001	14	14	14	14	13	18	17	14	13

3) We see wrong field names, and null values because there are comments in the first few rows in the original Excel file and the column headers have been mis-interpreted as values in the data.

4) In the side panel, click to turn on the Data Interpreter. Instantly, the data sheet has been cleaned up. All the original column headers have been correctly identified.

Resolved Incidents.Employee	2/1/2015	3/1/2015	4/1/2015	5/1/2015	6/1/2015	7/1/2015
B-002	4	1	5	2	3	0
E-055	1	2	1	3	4	1
E-075	14	17	16	15	18	16
B-066	4	4	5	2	5	0
C-025	17	13	17	18	17	17
E-030	2	2	1	1	0	3
C-001	14	14	14	14	13	18
E-038	4	1	0	4	0	2
C-054	2	5	4	4	2	3
A-081	3	2	4	5	2	2

5) Click on *Review the results* link on the left pane. This will open an Excel file describing what Tableau has done to clean the data. Click on *Resolved Incident* sheet, we see which

fields are being used as headers, in red, and which are considered data, in green.

	A	1/1/15	2/1/15	3/1/15	4/1/15	5/1/15	6/1/15	7/1/15	8/1/15	9/1/15	10/1/15	11/1/15	12/1/15	Header
1	Flights Data Summary													
2	This report was generated on 1-1-16													
3														
4	Employee	1/1/15	2/1/15	3/1/15	4/1/15	5/1/15	6/1/15	7/1/15	8/1/15	9/1/15	10/1/15	11/1/15	12/1/15	Header
5	B-002	4	1	5	2	3	0	3	1	2	0	2	5	Data
6	E-055	1	2	1	3	4	1	4	0	2	1	4	0	Data
7	E-075	14	17	16	15	18	16	14	17	12	13	14	12	Data
8	B-066	4	4	5	2	5	0	0	2	0	1	0	3	Data
9	C-025	17	13	17	18	17	17	12	15	17	17	14	15	Data
10	E-030	2	2	1	1	0	3	5	5	0	2	4	1	Data
11	C-001	14	14	14	14	13	18	17	14	13	18	15	14	Data
12	E-038	4	1	0	4	0	2	5	0	2	2	2	2	Data
13	C-054	2	5	4	4	2	3	0	5	5	5	3	5	Data
14	A-081	3	2	4	5	2	2	2	4	1	4	2	0	Data
15	B-031	14	14	14	14	15	13	15	14	12	16	12	18	Data
16	D-019	2	3	0	0	4	4	1	2	5	0	5	5	Data
17	E-096	2	0	4	4	5	3	3	0	5	4	2	0	Data
18	D-026	0	2	0	2	5	3	1	0	0	2	5	4	Data
19	E-022	3	3	4	3	4	2	0	3	2	3	3	1	Data
20	C-015	1	5	3	5	2	1	3	3	1	1	5	2	Data
21	B-062	14	12	16	16	16	18	12	12	18	16	12	17	Data
22	E-029	5	1	2	4	0	3	5	4	5	3	4	5	Data
23	A-037	2	2	0	2	3	4	2	0	2	1	2	2	Data
24	E-087	14	17	13	17	18	13	13	12	13	16	13	16	Data
25	C-040	5	0	4	5	3	5	2	1	1	4	2	1	Data
26	A-077	3	5	3	5	2	3	5	3	4	5	4	4	Data
27	C-041	18	18	15	15	15	17	17	12	18	17	16	16	Data
28	D-005	4	0	5	3	2	3	5	1	0	2	3	3	Data
29	E-046	12	13	14	17	16	14	14	18	13	16	15	17	Data
30	C-053	2	5	5	0	2	1	0	4	5	1	1	3	Data
31														
32														
33	This report is generated once a month													
34	For questions, contact flights@support.com													
35	Employee field consists of Location-Employee ID													
36														

6) Back to Tableau, we want to change the table from column orientated to row orientated.
7) Click on the first column header 1/1/2015 to select the first column, scroll to the right, shift-click the last column header 12/1/2015 to select all the dates column, open the menu, select Pivot.

8) This pivot feature essentially merges the information from the original columns and rows into two new columns – Pivot field names, and Pivot field values. We can see that *Pivot field names* is actually our Date, double click to rename to Date, and *Pivot field values* can be renamed *Resolved Incidents*.

Pivot is a powerful method to aggregate the values that we are ultimately after. It moves the dates into a single dimension column Date and the number of resolved incidents into a single measure column. We will then have a row for each date and the number of resolved incidents.
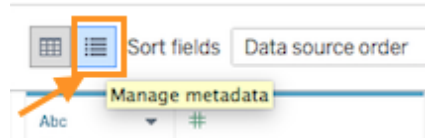
9) Click on Metadata Grid icon to view the original/remote field names.

10) Click on the menu on the upper right corner of *Employee* field, choose *Split*. The field Employee is automatically split into two fields. Double click to rename them to *Employee-Location* and *Employee-ID*.

Pivot Date	Pivot Resolved Incidents	Resolved Incidents Employee	Calculation Employee - Location	Calculation Employee - ID
1/1/2015	4	B-002	B	002
10/1/2015	0	B-002	B	002
11/1/2015	2	B-002	B	002
12/1/2015	5	B-002	B	002
2/1/2015	1	B-002	B	002
3/1/2015	5	B-002	B	002
4/1/2015	2	B-002	B	002
5/1/2015	3	B-002	B	002
6/1/2015	0	B-002	B	002
7/1/2015	3	B-002	B	002

11) DIY. Click on Sheet 1, build a view with the following requirements
   a. Make sure your dimensions and measures have the correct data type
      i. Change the data type at Data Source interface
      ii. Or right click on the field name in the data pane, -> Change Data Type
   b. Build a hierarchy with Employee-Location and Employee-ID
   c. Show the total number of resolved incidents per Employee-ID per year
   d. Create a filter with Employee-Location
   e. Your finished chart should like the screenshot below
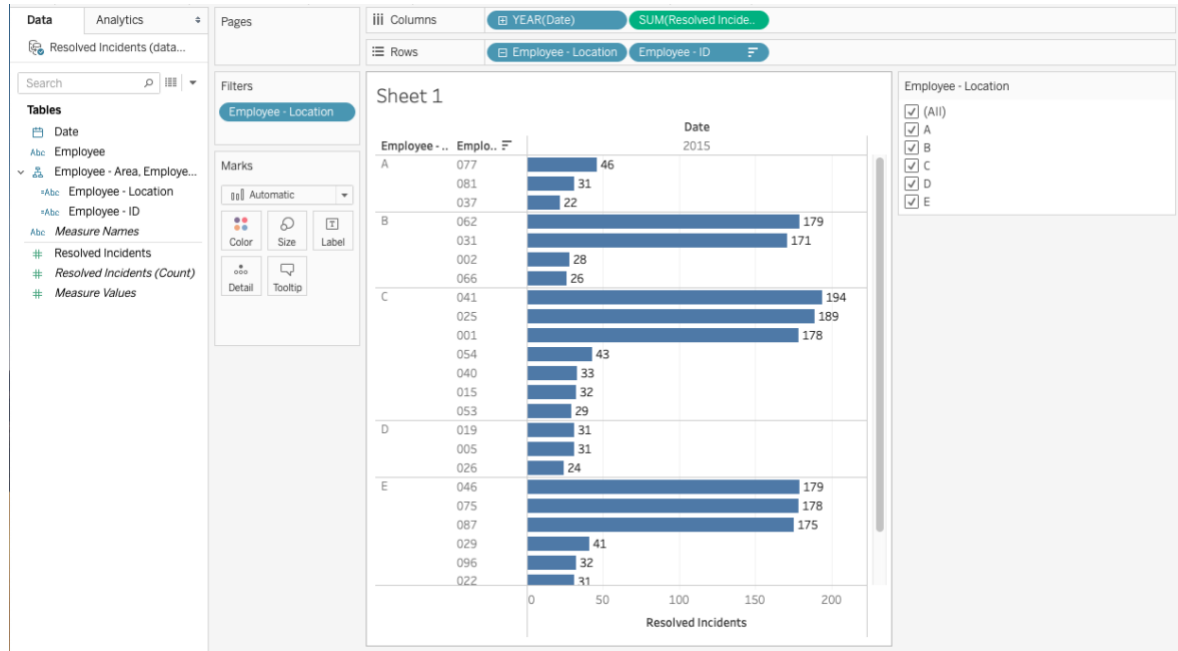
II.   Merge data from different data sheets/datasets

It's not practical to store all data in a single sheet/table/dataset. There are two basic ways to merge datasets. They are Join and Union.

Join will combine columns of two datasets. It merges different columns for matching rows, aligning them next to each other. Tableau tries to automatically decides which field to use to match up the rows. To join, the tables may come from different connections.

- Inner Joins – preserves only the rows that have the same key field.
- Left Joins – brings in all the information for the rows from the table on the left and any information from table on the right for rows with the same key field.
- Right Joins - brings in all the information for the rows from the table on the right and any information from table on the left for rows with the same key field.
- Outer Joins – brings in all names listed in all tables, and fills in nulls wherever there isn't information for a given column for that row.
- For more, go to Join Your Dat*a at* https://help.tableau.com/v2020.2/public/desktop/en-us/joining_tables.htm.

Union will merge rows of two or more datasets. It merges different rows for matching columns, stacking more datasets on top of each other. This is usually used when more data comes in during the project. Tableau tries to match up columns with identical names automatically. To union data, the tables must come from the same connection.

The Data Source Page Canvas

The data source page canvas has two layers: a logical layer and a physical layer. The default view is the logical layer that helps create a data model of data sources visually. When a table is first added to the canvas, it is a logical table at a logical layer. Double-click a logical table to go to the physical layer. The Join and Union have to be done at the physical layer. In the physical

layer, the table is a physical table. You can't join and union data at the logical layer. You can combine data using Relationships at the logical layer. For more details, go to The Tableau Data model at https://help.tableau.com/v2021.3/public/desktop/en-us/datasource_datamodel.htm.

1) We have two different file-based data sources – one Excel, one CSV. The data is structured such that the sales data has a product ID but no other information about the product sold. The product data in the CSV file is a simple table of product ID, category, sub-category, and the product name. If we perform a left join Product data to the Sales data on Product ID, we'll add that product information to each relevant sales transaction.
2) In Tableau, connect to *Sales 2016.xlsx* Excel file. The Sales table will be added to canvas area automatically.
3) Click the Add button by the Connections area, Choose Text File, and open *Products 2016.csv* file.
4) Double click Sales in the canvas area to open the logical table, going from the logical layer to the physical layer. Drag Products 2016.csv to the canvas area to create a Join operation. Name the data source more descriptively *Sales and Products 2016*.
5) Click on the join icon, and we'll select a left join. We can see that our data has been joined at the row level on Product ID.
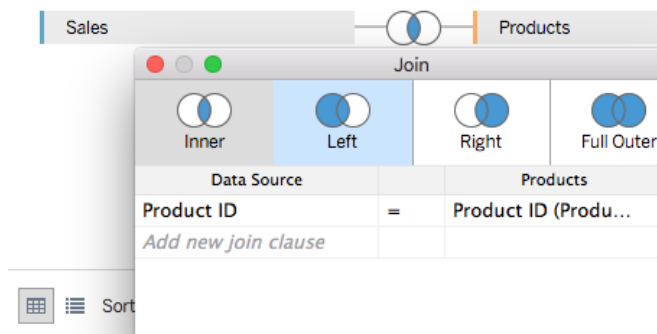
6) In the grid, we can see the product information (in yellow) has been added to each row from the Sales data (in blue).
7) Test out our newly joined single data source

a. We see we have a single data source in the data pane, but our fields are broken out by table.
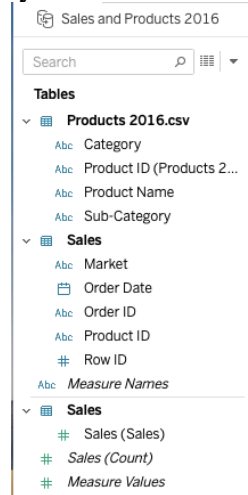
b. Right click inside the data pane, choose *Group by Folder* instead of Group by Data Source Table.
c. Right click over *Product ID(Products 2016)*, Hide it. Because it's just there for the join clause.
d. DIY. Click on Sheet 1, build a view with the following requirements
      i. Build a hierarchy with Category, Sub-Category, and Product Name
     ii. Show the sales by year and product Sub-Category
    iii. Create a filter with Category
     iv. Your finished chart should like the screenshot below. An integrated data source functions like any other single-source data source.
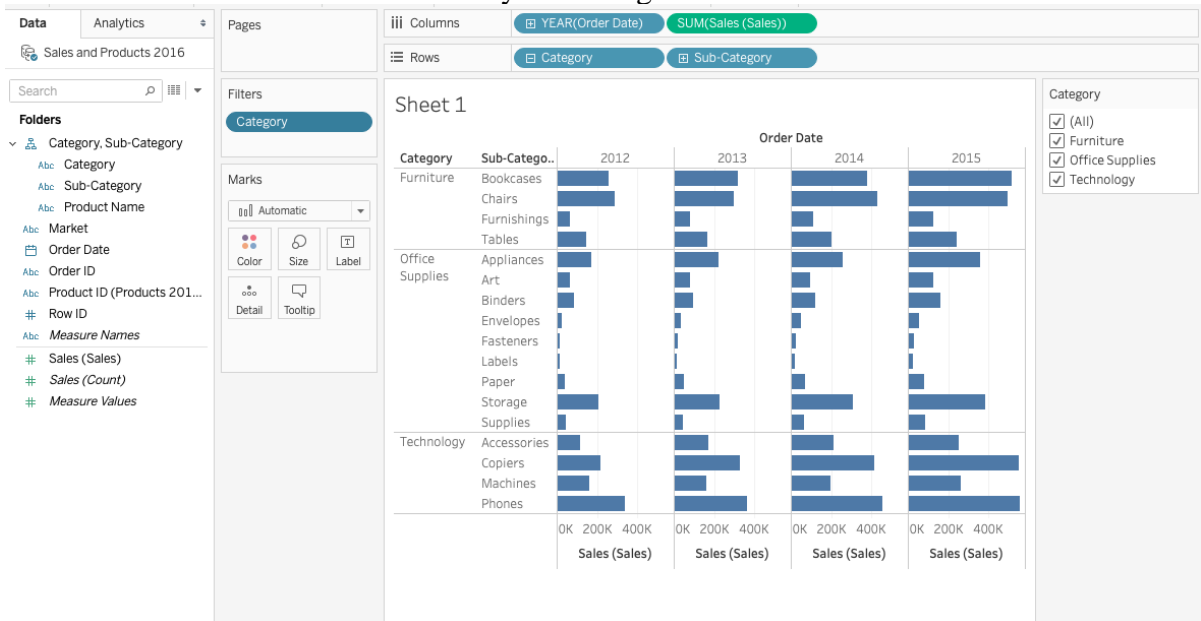
Activity 2.2 Union two tables from the same Excel file

We need to add the rows of data in the sheet *DMC-charges-2018-072019, DMC-charges-2017, DMC-charges-2014-2016* all together.

1) Open Tableau and connect to *DMC-charges-data-three-sheets.xlsx.*
2) The Excel file contains 3 sheets: DMC-charges-2014-2016, DMC-charges-2017, and DMC-charges-2018-072019.
3) Click New Union in the left pane of Data Source interface. Drag the three sheets to the popped-up window.
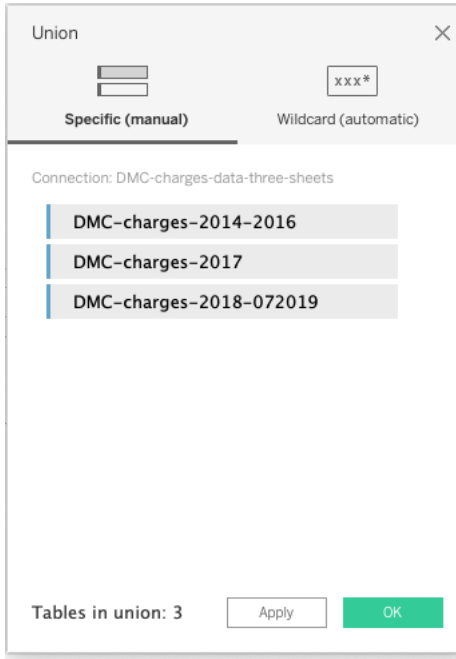
4) Click OK to close the window.

5) Change **Time Charge** from *String Abc* to *Date & Time*
6) Scroll the table all the way to the right, you will see Tableau has added two new columns *Sheet* and *Table Name* to show where the rows originally came from. This may become very helpful later.
7) Note: Since you haven't created any worksheets, dashboards yet, Tableau Public's save functions are greyed out. That means you can't save your work to anywhere at the moment. So if you clean the canvas area, and start it over, remember to correct the data type. Make sure all the fields have the right data types.
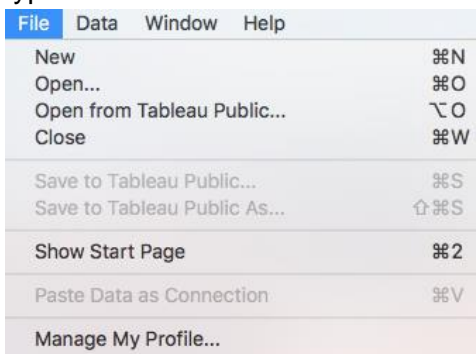
8) Test out our newly unioned single data source by creating a view to see how equipment has been checked out over the years 2014 through July 2019.

9) Since the data for 2019 is not complete, exclude the data point for 2019.
10) Line chart
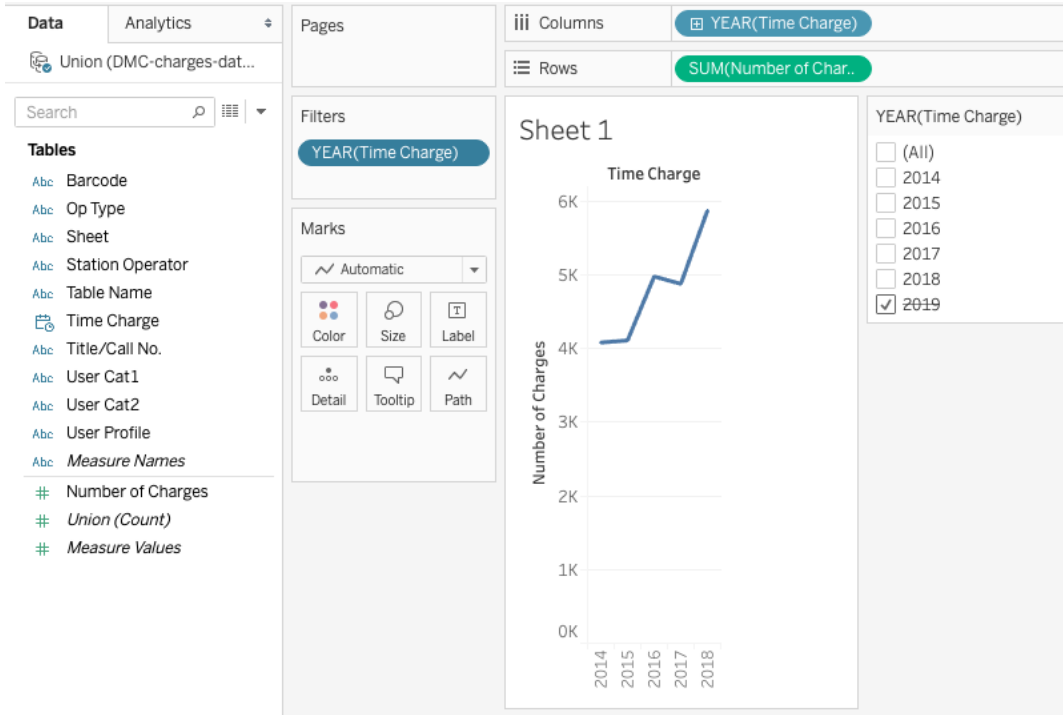
11) Bar chart
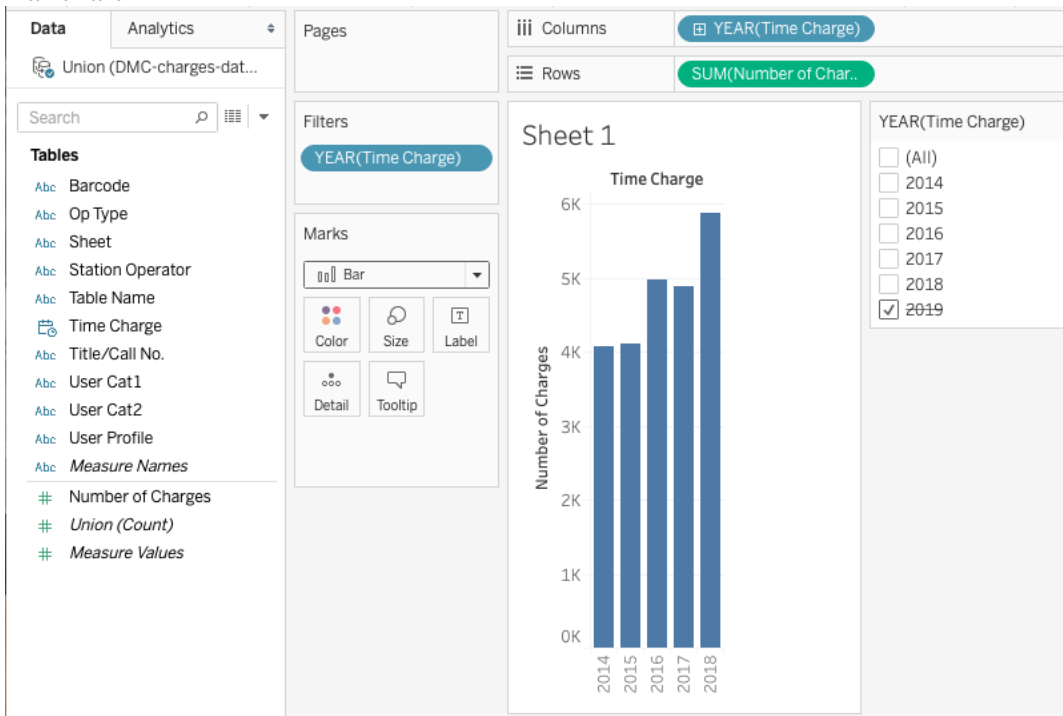
- Connect to Sample_Superstore.xls

- Create an Inner join using order ID
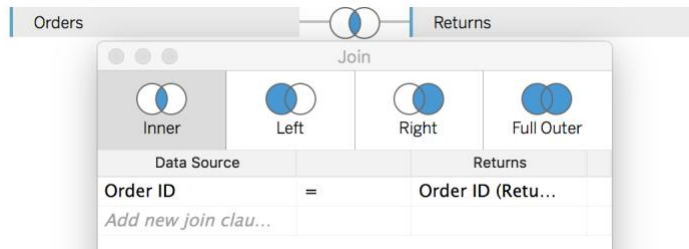
- Create a view that uses joined table to show the number of returned orders in each Sub-Category.
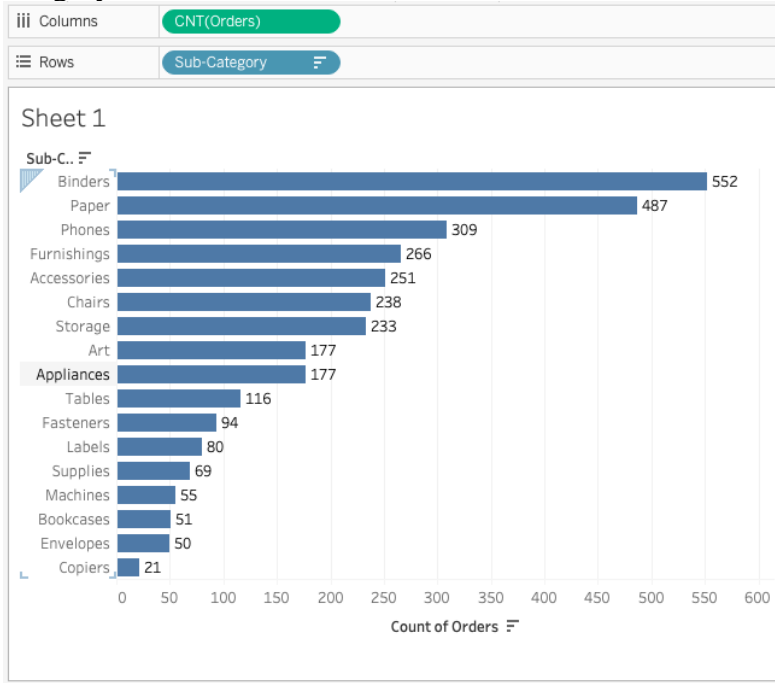
III. Dual axis chart and scatter plot

It is important to any analysis to compare measures with each other. Tableau has dual axis chart and combo chart to compare and contrast two measures. **A dual axis chart** uses two separate measures and a single dimension, comparing two measures in one view. A dual axis chart with multiple mark types is also called **a combo chart**.

<mark>Activity 3.1 Build a Dual Axis chart</mark>
1) Connect to *Sample-Superstore* excel file.
2) Drag Orders table to canvas.
3) Drag *Sales* measure to *Rows*, drag *Order Date* dimension to *Columns.* And select continuous month for Order Date.

4) Drag Profit measure to Rows. From the Profit pill drop down menu, choose Dual Axis.
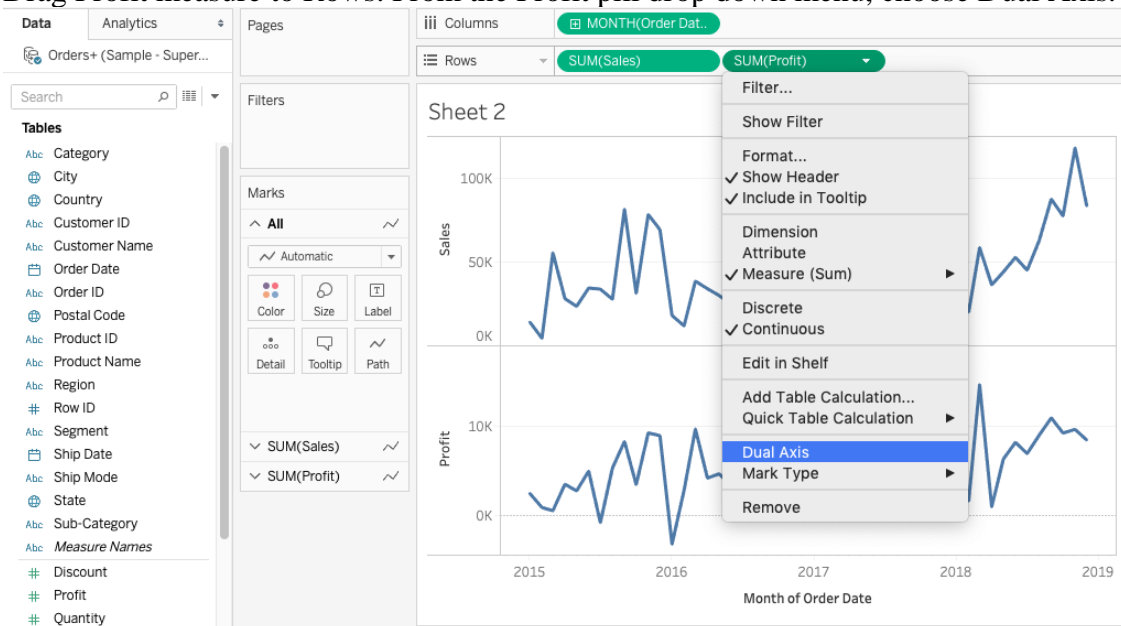
5) Right click over the right Axis, choose *Synchronize Axis* and de-select *Show Header*.
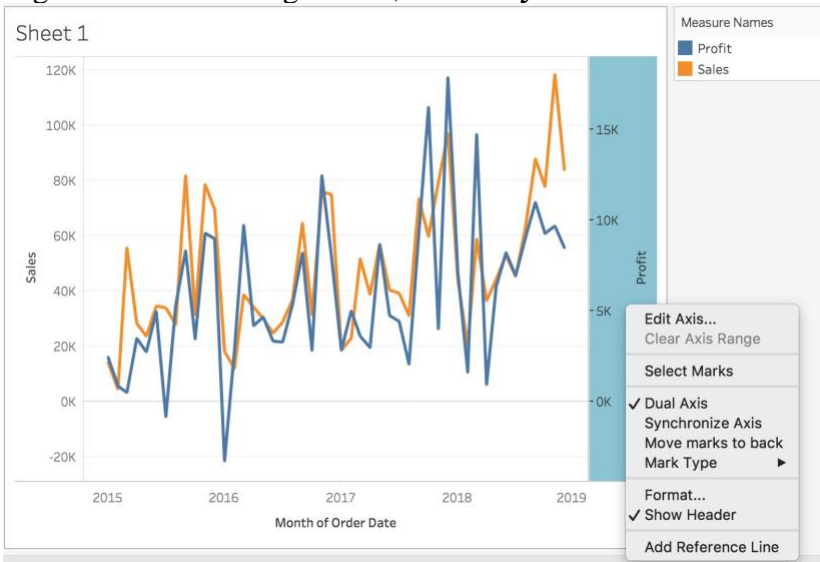
6) From the Marks card, you can change mark type, color, and size.

Building scatter plots

Scatter plot is a useful chart to visualize relationships between numerical values and to analyze the correlation between the two measures. It uses position as visual cue. Marks often represented as hollow circles are used to compare multiple measures based on where they are located in relation to other marks. One advantage of using scatter plot is that you can plot a lot of data at once and be able to spot trend, clusters, and outliers.

Activity 3.2 Build a Scatterplot

13

1) Connect to *Sample-Superstore* excel file.
2) Drag *Sales* measure to *Columns*, drag *Profit* measure to *Rows*.
   Tableau aggregates the measures as sums and creates a horizontal axis and a vertical axis.
3) Now the Scatterplot has one mark, showing the sum of sales and profit values across every row in the data source.
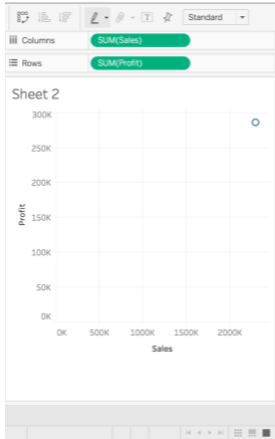
   a. Disaggregate measures to the row level
      i. Analysis -> Aggregate Measure, and de-select to disaggregate the data. Now, the Scatterplot has 9994 marks – one for each row in the original data source.
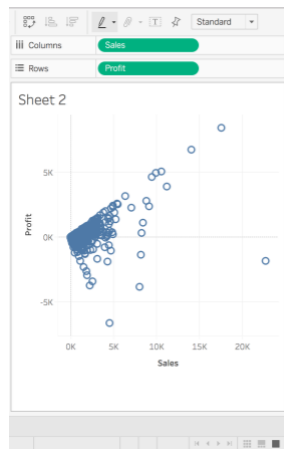
ii.   We can easily spot the outliers. But it is hard to tell any insights. Drag Region to Color. Not much discernable.
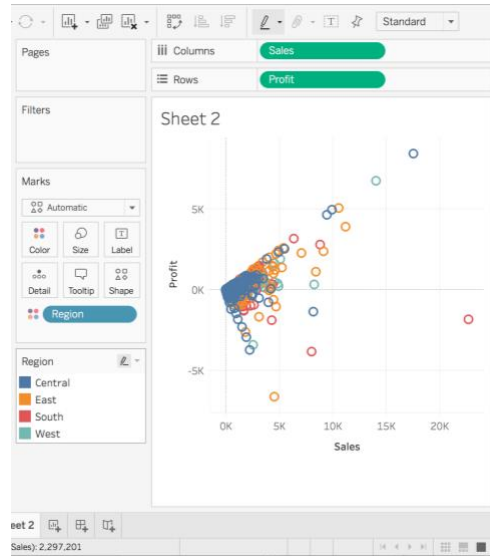
iii.   Remove Region from Color. Drag Category to Color. It looks like Furniture sales doesn't profit very much.
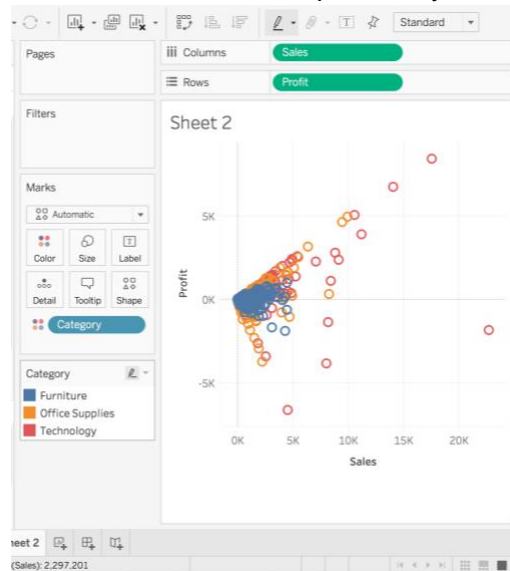
iv.   Trend lines are used to predict the continuation of a certain trend. Click on Analytics tab next to Data tab.

v. Drag Trend Line label and drop it on to Linear model.
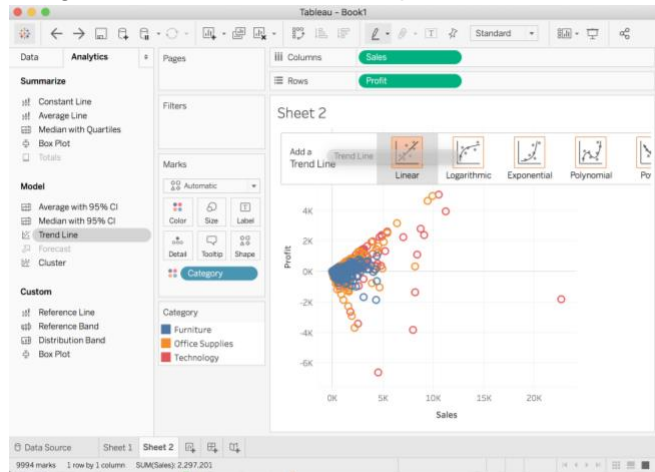
vi. The blue trend line for Furniture does show low profitability for furniture sales.
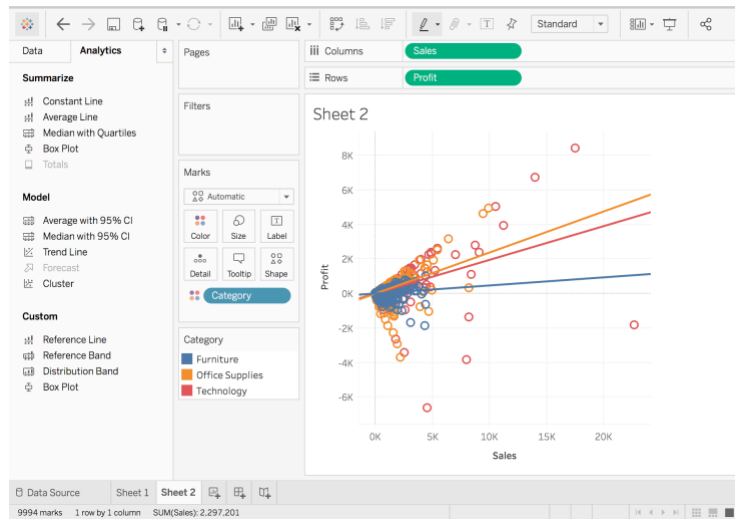
vii. Edit colors for the color legend like below

Sales and Profit Trend by Category

Profit = 0.0469134*Sales + -7.71263
R-Squared: 0.0301056
P-value: < 0.0001
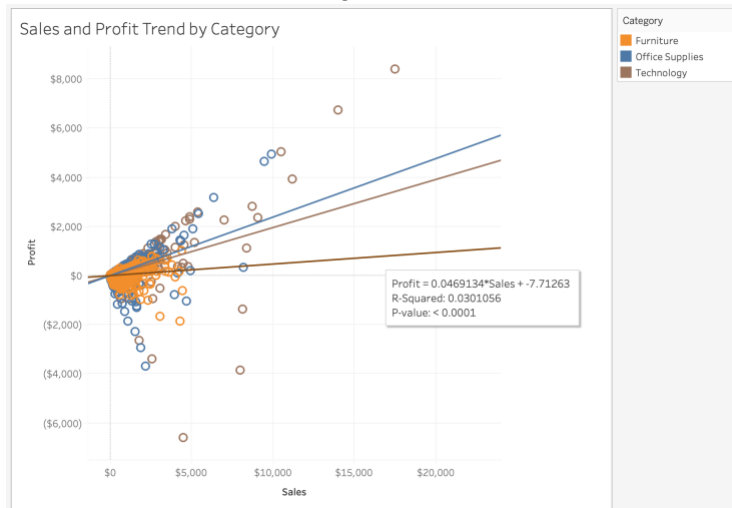
viii. Note the tooltip of the trend line shows the equation used to calculate the trendline, and R-Squared and P-Value. Those are statistical measures that help tell how well your data fits the data model.

ix. Also note that at the row level, you can't select data marks to keep only or exclude to filter data because you cannot add a data source filter this way.

b. Disaggregate measures by adding dimensions
    i. Drag Customer Name to Detail mark card
    ii. The view changes from 1 mark to 785 marks.
    iii. Outliers can be easily selected and excluded at this level.
    iv. Name the sheet "Sales and Profit by Customer"

IV. Compare Multiple Measures on the Same Chart and Build a Text Table with Multiple Measures

With Measure Names and Measure values, we can get three measures in the same line chart. The Data pane always contains a number of fields that do not come from your original data, two of which are **Measure Names** and **Measure Values**. Tableau automatically creates these two fields so that you can build certain types of views that involve multiple measures.

- The **Measure Names** field contains the names of all measures in your data, collected into a single field with discrete values.
- The **Measure Values** field contains all the measures in your data, collected into a single field with continuous values. Drag individual measure fields out of the Measure Values card to remove them from the view.

Activity 4.1 Compare Sales, Profit, and Discount Amount on the same line chart
1) Drag Order Date to Columns, choose continuous month
2) Drag Measure Values to Rows
3) Drag Measure Names to the filter shelf, select Sales, Profit, and Discount Amount since they have same measure unit dollar.
4) Drag Measure Names to Color.
Now we have three measures on the same axis!
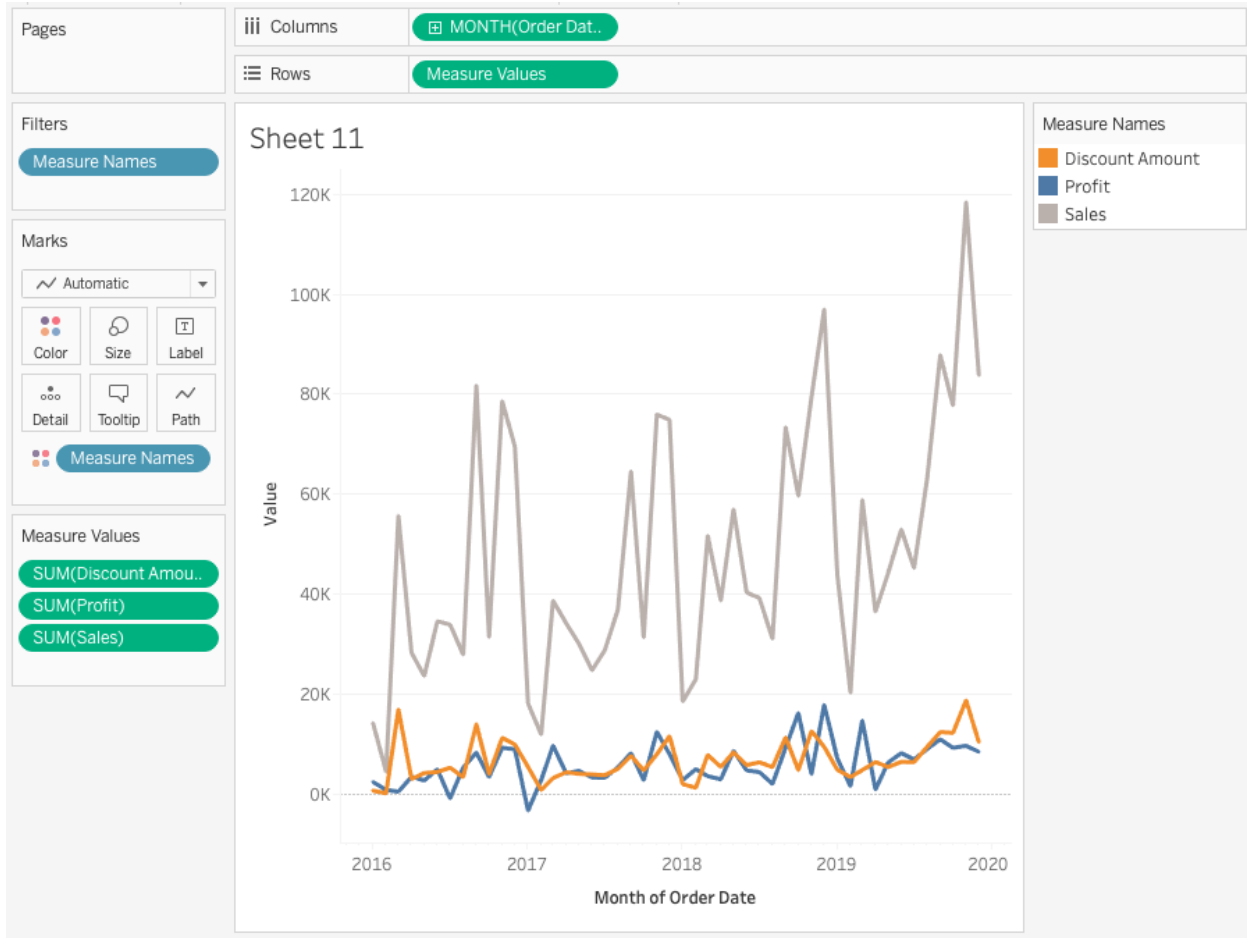
- Double click Measure Names
- Change aggregation for Discount to Average
- Change the number format for Discount Amount to currency
    - Right click over Discount Amount -> Default Properties->Number Format… Currency Standard

V. Calculated Fields

If your underlying data doesn't include all the fields you need to answer your questions or to explore your data, you can create new fields from existing data fields. These fields are called calculated fields.

When you create a calculated field, you are essentially creating a new field (or column) in your data source, the values or members of which are determined by a calculation that you control. This new calculated field is saved to your data source in Tableau and can be used to create more robust visualizations. Tableau doesn't change the data source, but can create an extract where the calculations will be visible.

In other words, the calculation is done at the data source level. For instance, Sales minus Profit as in the Sample Super Store data, is passed as part of the query that Tableau asks of the data source and the computation is handled by the data source itself with only the result set being returned to Tableau.

Note that there's a little equal sign in front of the calculated field – this indicates it's a calculated field, not natively from the data source.

There are four basic components to calculations in Tableau:

- **Functions** - Statements used to transform the values or members in a field, such as IF, THEN, ELSEIF, ELSE, and END.
- **Fields** - Dimensions or measures (columns) from your data source.
- **Operators** - Symbols that denote an operation, such as > and <=.
- **Literal expressions** - Constant values that are represented "as is", such as "Profitable" and "Unprofitable".

Functions in Tableau fall into one of several main categories: Number, String, Date, Logical, Type, etc. See Functions in Tableau for more.

The right-hand side of the calculation editor provides information about the functions. It can be collapsed to save space or expanded. Functions are colored blue in Tableau calculations, with the exception of logical functions, which are colored black. Field names are colored orange in the formula.

<mark>Activity 5.1 Build an Executive Overview worksheet to show some overview figures at a glance. Create a calculated field "Total Orders" to count the total number of orders</mark>
1) Click on the drop-down menu in the data pane, choose "Create Calculated Field…"
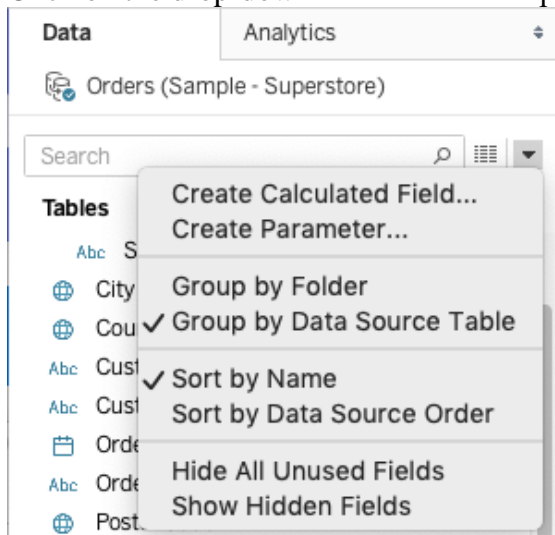
2) On the calculation editor, name the field "Total Orders", input COUNTD([Order ID]). Click OK to close the editor window.
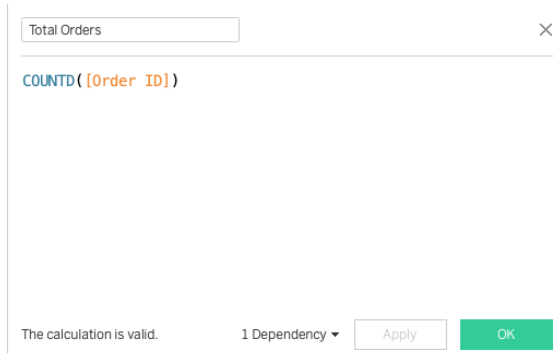
3) Drag Total Orders to Measure Values card

			Discount				
	Avg. Discount	Amount	Item Quantity	Profit	Sales	Total Orders
	15.62%	$322,582.13	37,873	$286,397.02	$2,297,200.86	5,009

Activity 5.2 Create a calculated field "Profit Ratio" to calculate profit ratio
1) Click on the drop-down menu in the data pane, choose "Create Calculated Field…"
2) On the calculation editor, name the field "Profit Ratio",
   input SUM([Profit])/SUM([Sales]). Click OK to close the editor window.
3) Drag Profit Ratio to the Measure Values card.
4) Right click over Profit Ratio -> Default Properties->Number Format… Percentage

DIY Activity: Create a calculated field "Profit per Order" to calculate profit per order
1) On the calculation editor, name the field "Profit per Order",
   input SUM([Profit])/COUNTD([Order ID]).
2) Change the number format to Currency standard

DIY Activity: Create a calculated field "Profit per Customer" to calculate profit per customer
1) On the calculation editor, name the field "Profit per Customer",
   input SUM([Profit])/COUNTD([Customer ID])
2) Change the number format to Currency standard

Format -> Format Borders-> Rows-> Row Divider, Pane -> None

The final Executive Overview worksheet will look similar like below.

	Avg. Discount	Discount Amount	Item Quantity	Profit	Sales	Total Orders	Profit Ratio	Profit per Order	Profit per Customer
	15.62%	$322,582.13	37,873	$286,397.02	$2,297,200.86	5,009	12.47%	$57.18	$361.16

VI.   Table Calculations and Quick Table Calculations

A table calculation is a secondary calculation that performed on top of the returned result set. This computation is done in the view and stay there, locally in our worksheet. They aren't stored in the data source. An example is as Running Total of Sales. A table calculation is indicated by this delta symbol on the pill.

Table Calculations **are based only on the information in the view.** In a given view, the underlying data table (hence "table calculations") contains the information that makes up the marks. If you're working with a crosstab, this data is exactly what you see. For any other view, the underlying data is represented differently, but the values are the same as the crosstab. **Adding or removing fields changes the view, which in turn alters the Table Calculation.**

Table Calculations can either be written like any other calculation, using the Table Calc functions in the calculation editor or **there is 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, etc.

Activity 6.1 Calculate Percent of Total for each product category per year.

1) Connect to Sample-Superstore Excel file.
2) Build the view
   a. Place Profit on the Columns shelf.
   b. Drag Order Date and Category to the Rows shelf.
   c. Drag Profit to the Columns again.
3) Create a Quick Table Calculation
   a. Right click on the second SUM(Profit) field.
   b. On the drop-down menu, choose Quick Table Calculation -> Percent of Total

c. Notice the delta symbol on the pill? That indicates it is a table calculation. Do your math and check the results. Do these numbers seem correct?

d. Click on the quick table calculation field, from the drop-down menu, choose Format -> Pane -> Numbers -> Percentage, make sure 2 decimal places is selected.

4) We see a percentage across year 2016-2019.

5) Add Subtotals for each year
   a. In the data pane, click on the Analytics tab.
   b. Click on Totals and drag it into the view, select Subtotals when a menu opens up.
   c. The totals for each year are added to the view. But the percentage is still the % of Total Number of Records along Table (Down) (i.e. the total number of records across all the years.)

6) Change how the table calculation is computing
   a. Right click on the second field in the Columns shelf
   b. From the drop-down menu, choose Compute Using -> Pane (down).
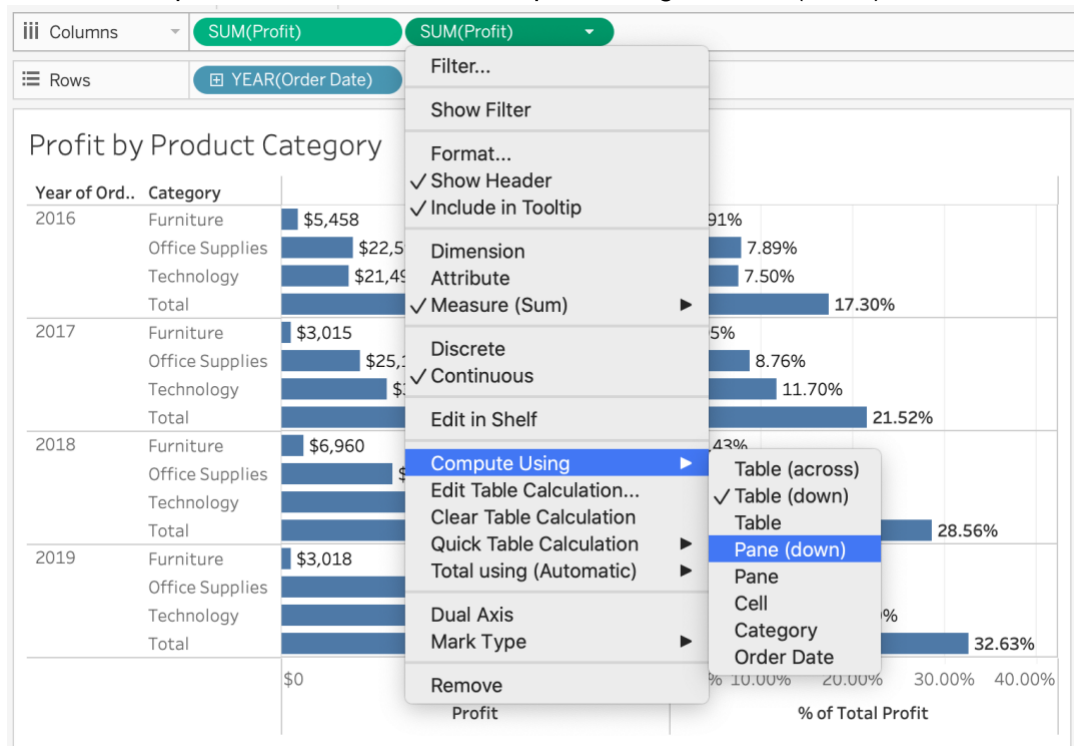
c. The percentage changes to % of Total Number of Records along User Profile(group) per year as shown below.

d. Tableau has defaulted to have the calculation run along Table (Down), meaning we see Percent of Total across all four years. If we want to see Percent of Total for each product category per year, we need to change the scope from Table to Pane, computing the Percent of Total for each Pane, or Year.

7) Back to Data pane, drag Category to Filters, and Show Filter.

8) Uncheck Furniture on the Category filter, you will see percent of total changed. Percent of Total is a Table Calculation. Keep in mind that table calculations are based on the data in the view, adding or removing data to the view will change the calculations.
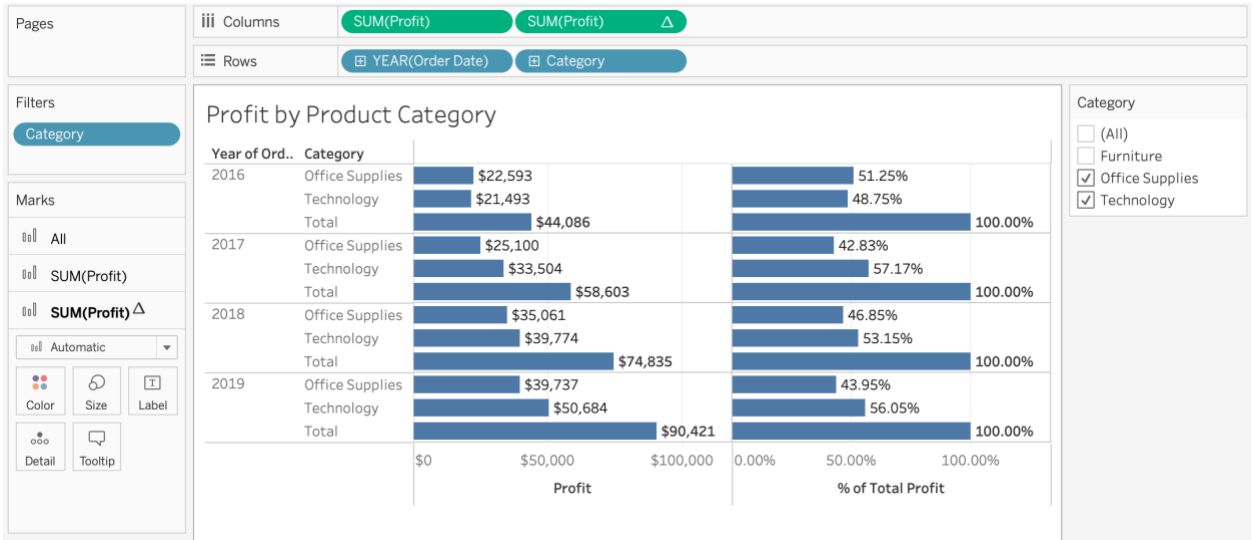

9) Drag Category to Color mark card
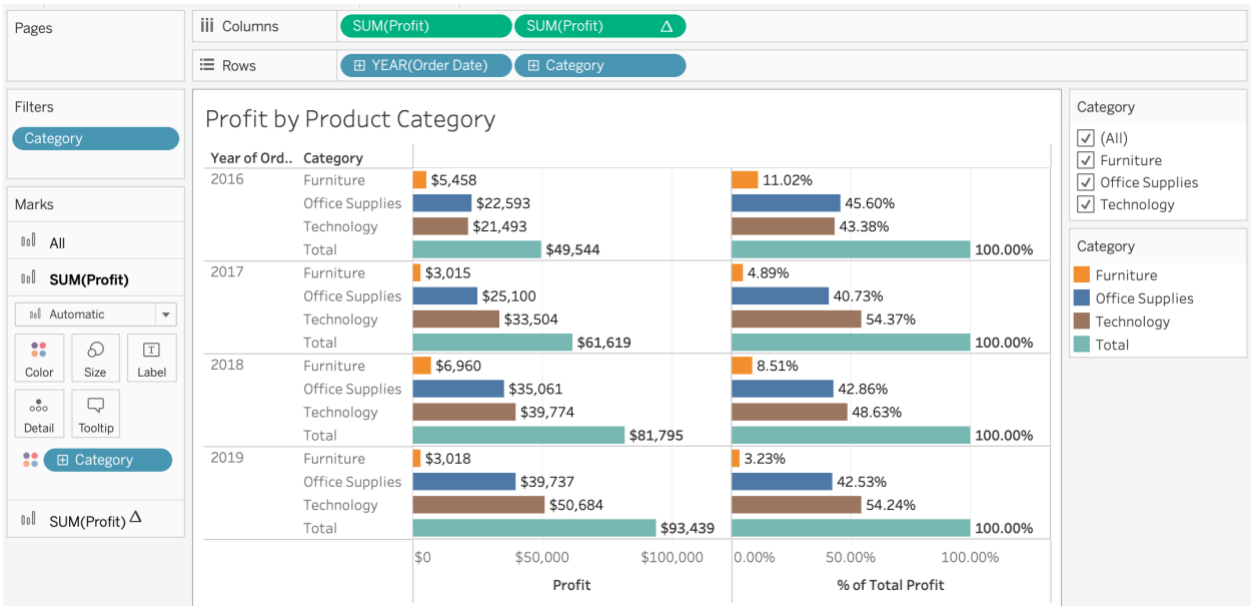
DIY Activity: Show the total sales by year and a running total over years. Create a dual axis chart.
1) Connect to Sample-Superstore Excel file.
2) Build the view.
   a. Drag Sales to the Rows shelf, Order Date to the Columns shelf, Drag Sales to the Rows shelf again.
   b. Change the first Sales on the Rows shelf to a bar chart.

3) Create a running total table calculation.
   Right click on the second field on the Rows shelf, choose Quick Table Calculation ->
   Running Total.
4) Create a Dual Axis view.
   Right click on the second field on the Rows shelf, choose Dual Axis.
   a. Right click on the left or right Y axis, choose Synchronize Axis.
   b. Turn on the Mark Label. Here is the final view.
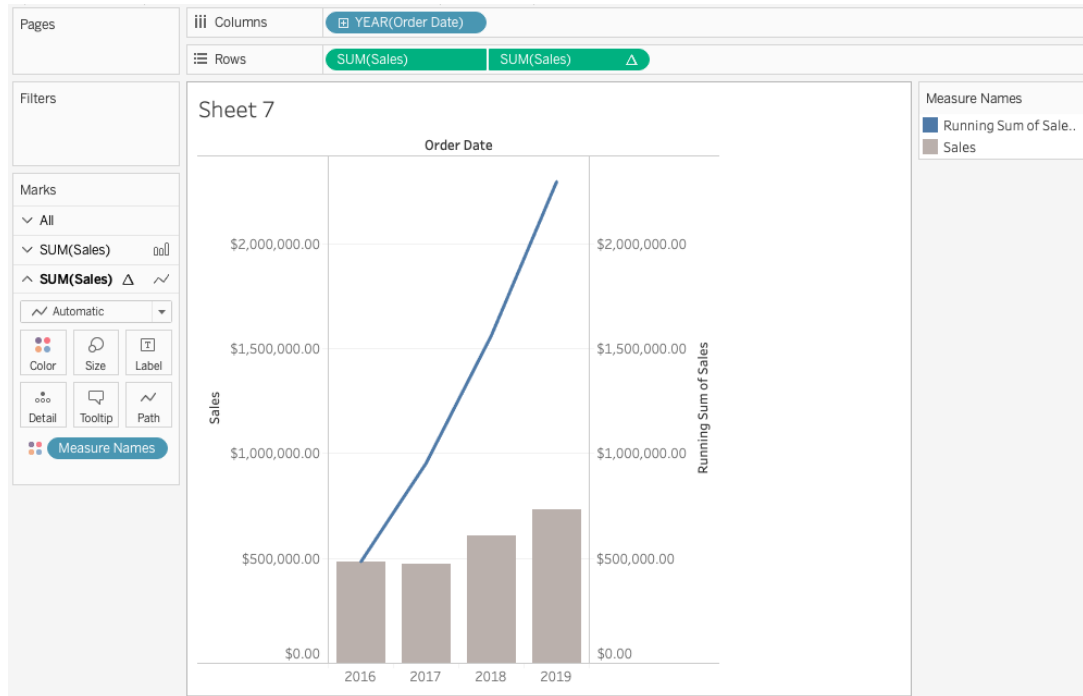
VII. Mapping Your Geographic Data

If your data contains geographic data such as state and zip code, you can plot it on the map to visualize it.

After you set up your data source, you might need to prepare your geographic data for use in Tableau. Depending on the type of map you want to create, you must assign certain data types, data roles, and geographic roles to your fields (or columns). For example, in most cases, your latitude and longitude fields should have a *data type* of **number (decimal)**, a *data role* of **measure**, and be assigned the **Latitude** and **Longitude** *geographic roles*, state field should have a *data type* of **string**, a *data role* of **dimension**, and be assigned the **state** *geographic roles*.

In Tableau, a *geographic role* associates each value in a field with a latitude and longitude value. When you assign the correct geographic role to a field, Tableau assigns latitude and longitude values to each location in that field by finding a match that is already built into the installed geocoding database. This is how Tableau knows where to plot your locations on the map.

Activity 7.1 Build a view of Profit by State

25

1) Connect to Sample_Superstore Excel file.
2) Double click State field.
3) Drag Profit to Color mark card
4) Edit colors to your choice
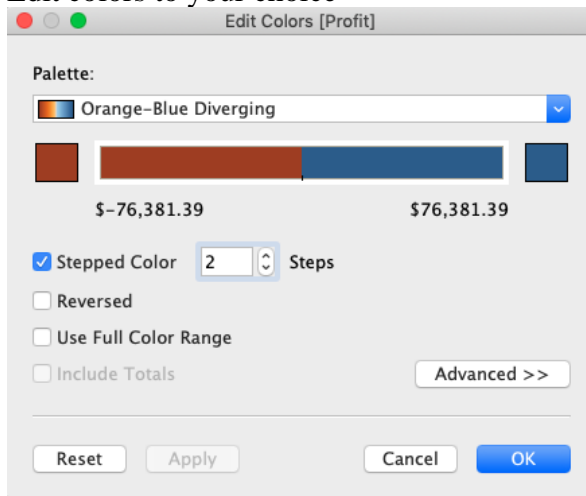
5) Add Region to Filters shelf. Your finished worksheet will look something like below.
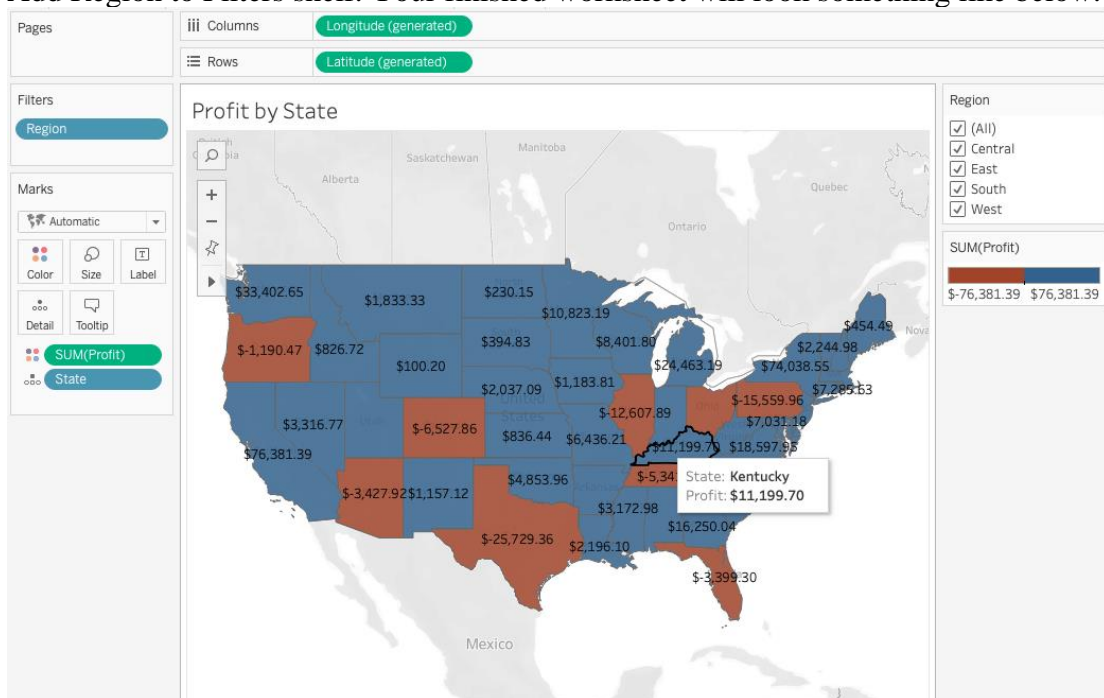

DIY Activity: Build a view to show the average time to ship by state on a map.
Clues: use Date calculations – DATEDIFF(date_part, start_date, end_date), which returns the difference between start_date and end_date expressed in units of date part. Use it to find out how long something takes.

1) Connect to Sample_Superstore Excel file. Drag Orders table to Canvas.

2) We want to know the average time it takes for an order to ship for each state.
3) Right click in the data pane and select Create Calculated Field. Name it "Time to Ship".
4) In the Calculation editor, type in DATEDIFF('day',[Order Date],[Ship Date]), click OK.
5) Double click State dimension to create a basic view.
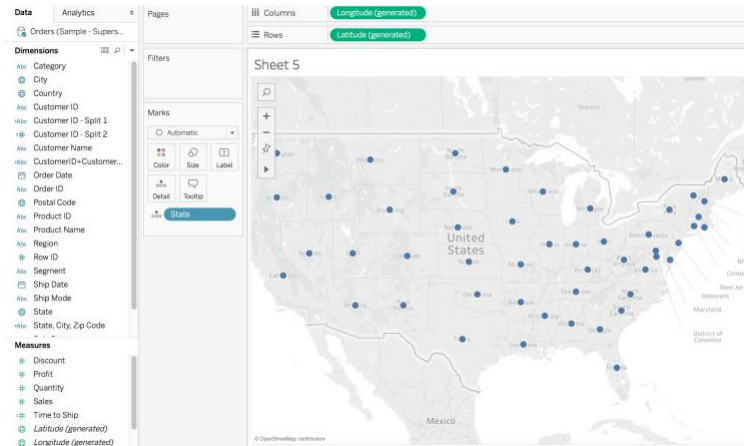
6) Let's color our states by using Time to Ship by dragging Time to Ship to the Color mark. Right click over SUM(Time to Ship), change the aggregation to Average.
7) Click on the Color mark, change the color to a red-black diverging palette and we'll reverse it because more time is bad.
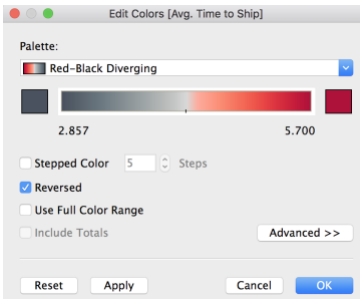
8) Now we can quickly see the states of the longest shipping times and we can focus on those areas.
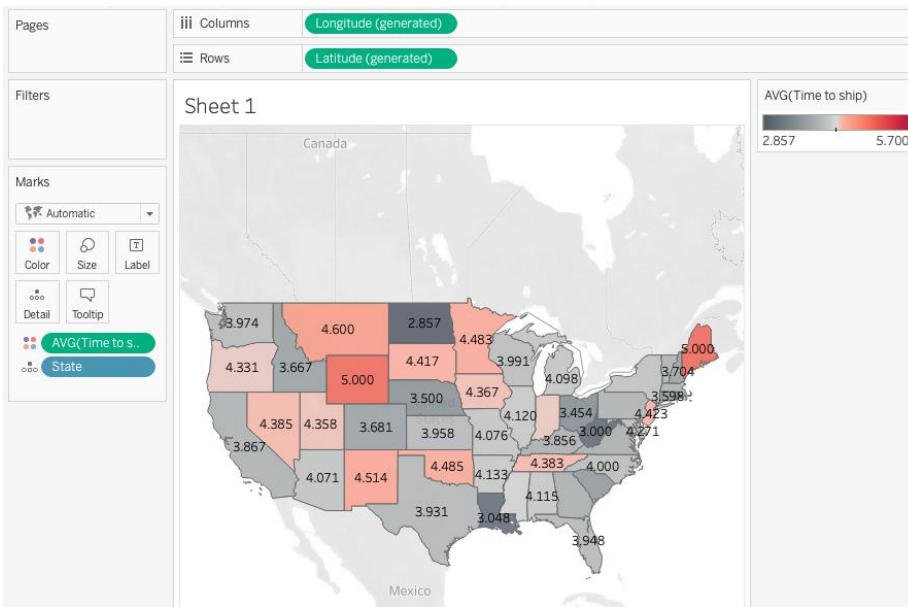
9) Change the Time to ship number format. Click on the dropdown menu of Time to ship, -> Default properties -> Number Format, Number(custom), decimal places – 1 and have stepped color, 5 steps checked.
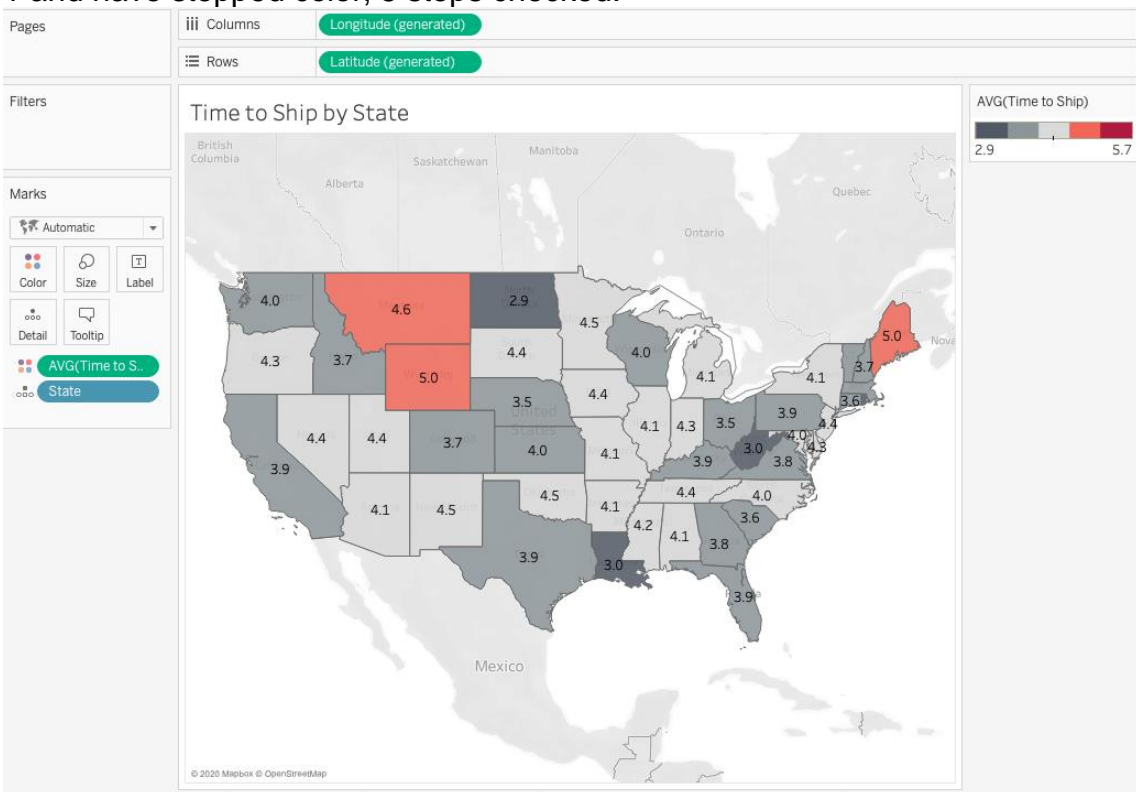
10) Optional steps
   a. Drag State to Filters, then Show Filter

Time to Ship by State
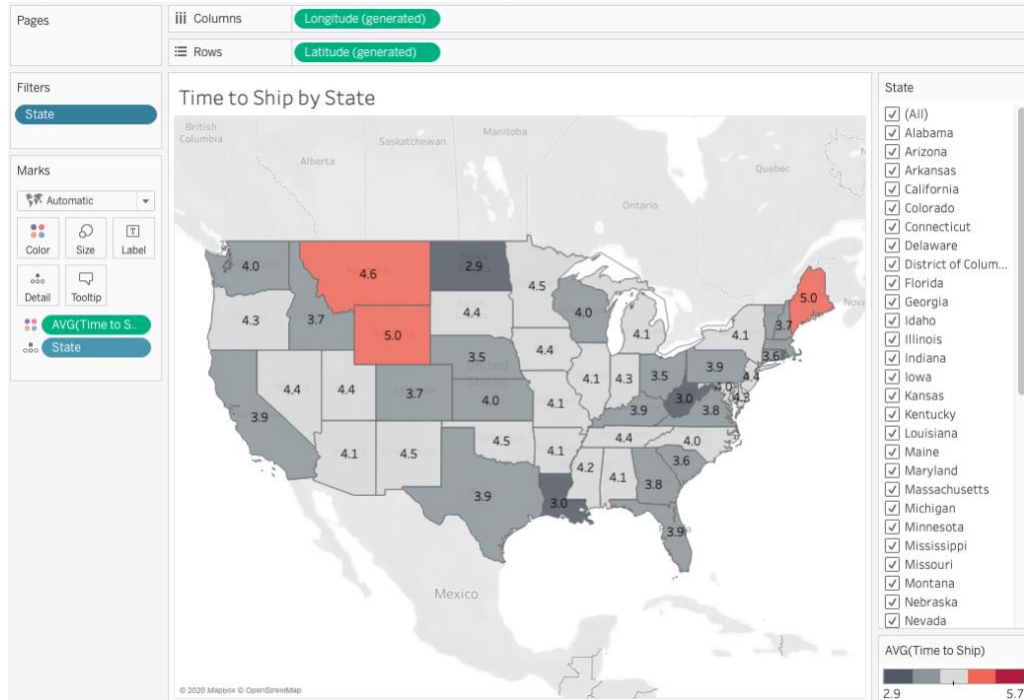
b. Uncheck any state from the State filter, the remaining Time to ship values are not affected. Keep in mind that regular calculations are done in the data source, which is not affected by the data in the view.

VIII. Using Parameters

Parameters are dynamic values that can replace constant values in calculations, filters, and reference lines. Parameters allow you to give your users control over the visualization.

Activity 8.1 Create a parameter to show Top N customers by Profit.
1) Connect to **Sample_Superstore** Excel file.
2) Drag Orders sheet to the Canvas area.
3) Create a parameter named *Top N*.
   a. Click on the drop-down menu on Data pane. Choose create Parameter…

b. Change the settings as below.

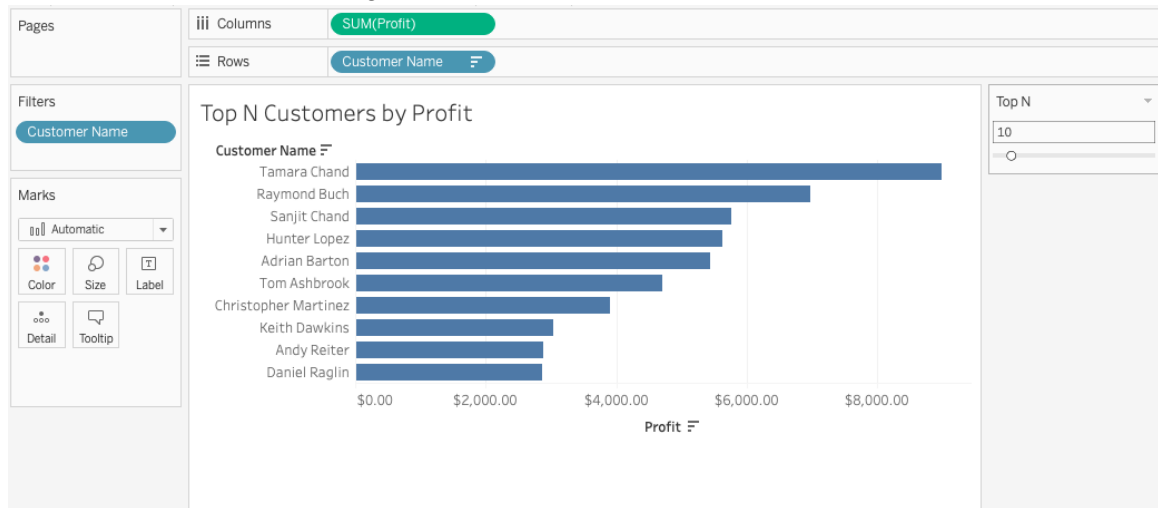c. Right click on the *Top N* parameter, choose *Show Parameter Control*.
4) Create a new worksheet and build a view.
   a. Drag Customer Name to Rows, Profit to Columns
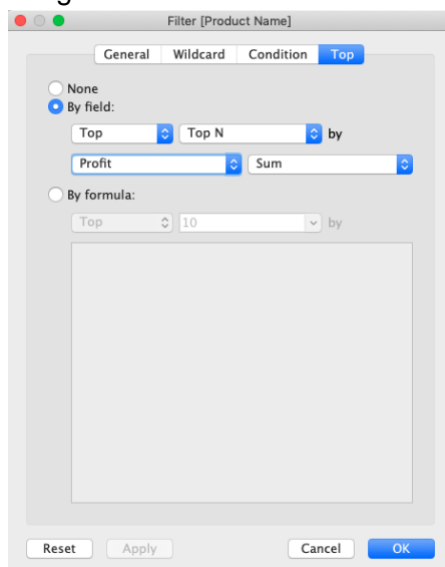   b. Sort it be descending order
5) Drag Customer Name to Filters shelf, set it up like below.

6) The view will look something like below.

1) Build the view by dragging Profit to Rows, and Product Name to Columns.
2) Sort it by descending order.
3) Drag Profit to the Rows shelf again
4) Create a running total table calculation.
   Right click on the second Profit pill on the Rows shelf, choose Quick Table Calculation -> Running Total.
5) Create a Dual Axis view.
   Right click on the second Rows pill on the Rows shelf, choose Dual Axis.
6) Right click on the left or right Y axis, choose Synchronize Axis. Select the right Axis, choose Line for Marks type.
7) Drag Product Name to Filters shelf and set up the filter

8) The final view look like below.
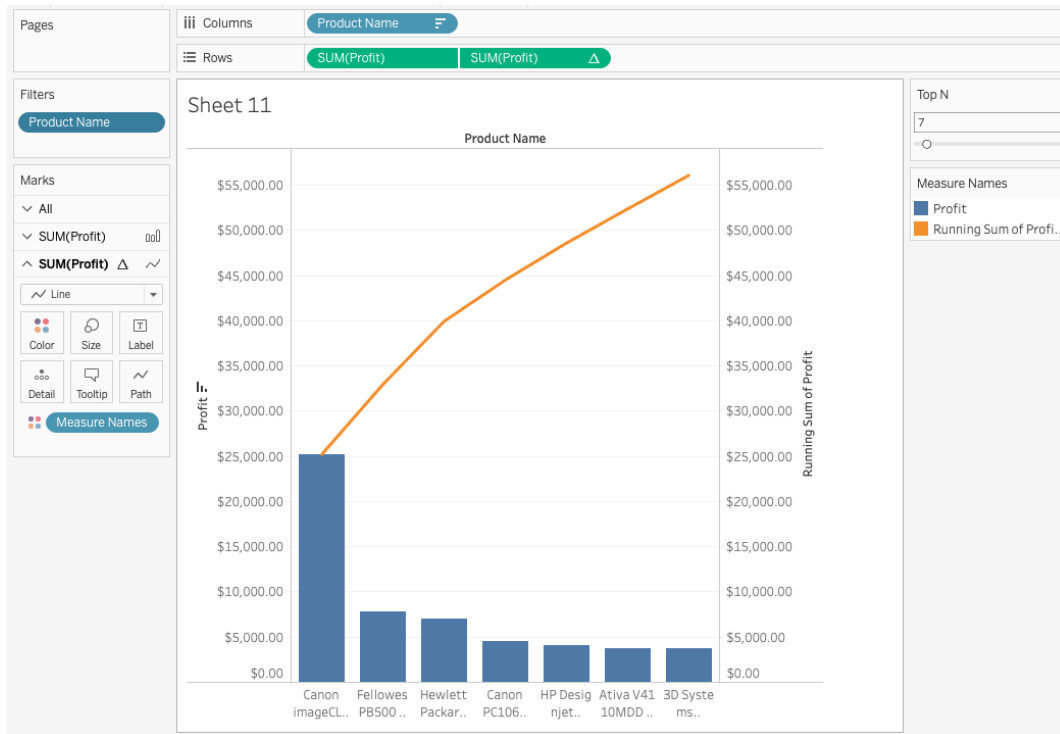
1) Build the view
   a. Drag Quantity to Rows, Product Name to Columns
   b. Sort by descending order
2) Create a parameter
   a. From the data pane, click on the drop-down menu, choose Create Parameter…
   b. Name it Pick a Number and set it as below.

c. Right click over the parameter *Pick a Number*, choose *Show Parameter Control*

3) Create a Calculated Field
a. From the data pane, click on the drop-down menu, choose Create Calculated Field…
b. Name it *less than Pick a Number* and set it as below. This is a logical calculation that returns true if the condition meets, false otherwise.

4) Split the view
a. Drag *Less Than Pick a Number* from **Measures** to **Color** mark.
b. Slide the Pick a Number slider to test it out.
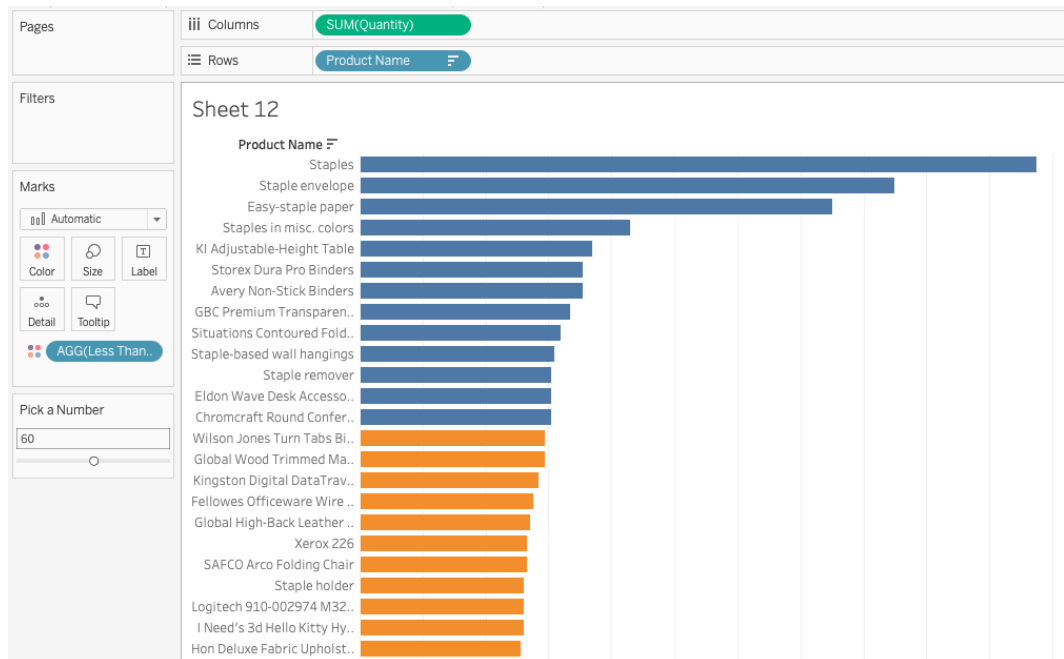
IX. Adding analytics objects to the view

In Tableau Desktop, options for adding Analytics objects to the view are available in the **Analytics** pane or menu, or in context in the view.

To add an item from the **Analytics** pane, drag it into the view. When you drag an item from the **Analytics** pane, Tableau shows the possible destinations for that item. The range of choices varies depending on the type of item and the current view.

Reference lines do exactly what it sounds like they'd do – they add a point of reference to a view.

Activity 9.1 Reference line options

1) Connect to Sample_Superstore Excel file.
2) Create a basic view by adding Profit to Rows, Category and Segment to Columns.
3) Click Analytics pane, choose Custom-> Reference Line, drop to scope Table, click OK. We will see a reference line has been added across the bar chart.
4) Right click over the line, choose Remove.
5) Let's redo it and use Pane this time.
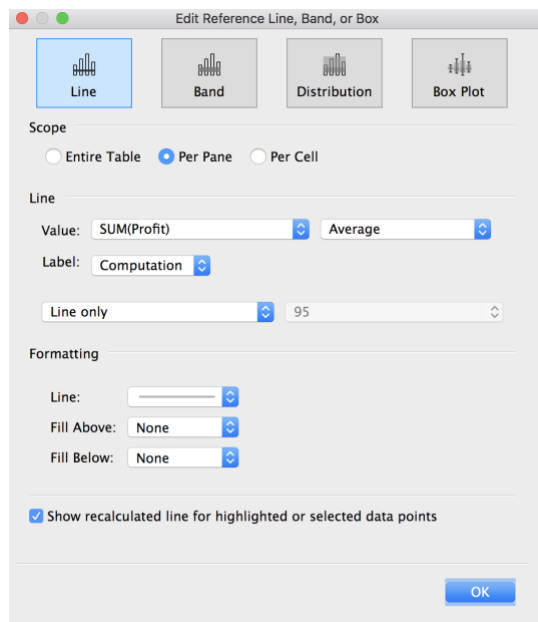
6) Scope. Scope is set to per pane, but can also be across the entire table or per cell. If we click Entire Table, we can see that the average line switched from three different lines, per Category, to a single average across the entire table.
7) Line. We'll make this an Average Profit line.
   a. Value lets us select which measure we want the line to be in reference to. Here, we only have Profit in the view.
   b. The default aggregation for the line is average, but we can change that.
   c. Next is Label. Here we can set what shows as the label on the line.
      i. None is fairly self-explanatory.

34

ii. Value displays whatever the value aggregation of the measure is, so the actual numerical average of profit.

iii. Computation displays what the aggregation of the measure represents, so here it would be the words "Average".

iv. Custom gives us the option to type in whatever we want, and the arrow to the side offers fields that can be inserted. We'll use <Value>:<Computation> of <Field Name>.
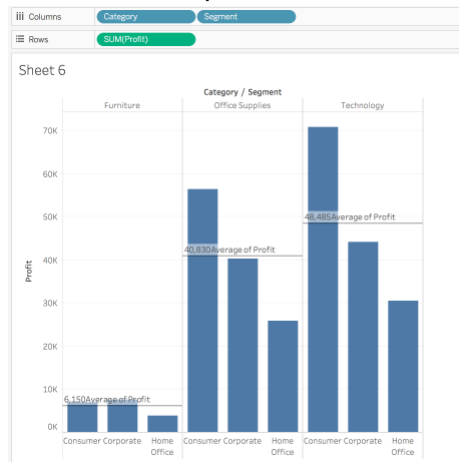
X.   Using parameters to build interactive Dashboards

1) Create a dashboard in a fixed size of 1024x768
2) Check "Show dashboard title", from the title drop-down menu, choose Format Title…, Shading, use the color picker to pick up Rice blue at https://brand.rice.edu/design-elements, title text color, choose Rice Gray.
3) Drag Horizontal object and drop it below the dashboard title. The object container can help distribute objects evenly.
4) Drag Top N Customers by Profit worksheet and drop it to the Horizontal object
5) Make the Top N parameter floating
6) Drag Sales and Profit by Customer and drop it next to the Top N Customers by Profit worksheet on the Horizontal object
7) Back to the Sales and Profit by Customer worksheet, drag Customer Name to the filter shelf, and set up the filter "Top N by Profit".
8) Test "Top N" parameter, see if it controls both sheets on the dashboard
9) Drag a Text object and drop it to the bottom of the dashboard. From Layout tab, choose Background, use the color picker to pick up Rice blue.
10) Drag a Blank object to drop it below the dashboard title, which will hold space for the Top N parameter.

11) The finished dashboard looks similar like below

Resources
1) Tableau online help
   https://www.tableau.com/support/help
2) Tableau free training videos
   https://www.tableau.com/learn/training