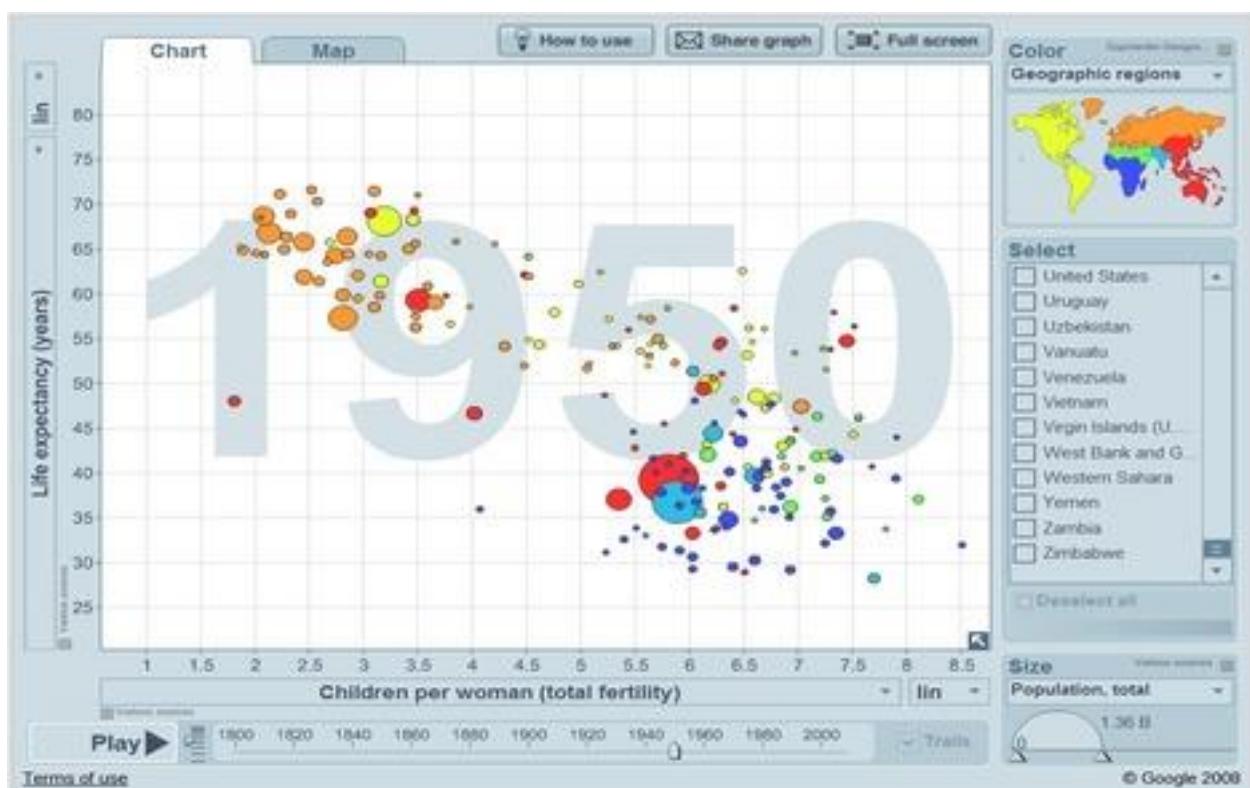


# Visual Analytics (BI Tools -Tableau)

1. Tableau for Beginners – Data Visualisation made easy
2. Intermediate Tableau Guide – For Data Science and Business Intelligence Professionals
3. Advanced Tableau – for Data Science and Business Intelligence Professionals

## Tableau for Beginners – Data Visualisation made easy

The visualisation of tableau, which was created by a famous Swedish statistician, Hans Rosling. He compiled roughly 200 years of World Development Data and presented it in a very simple manner:



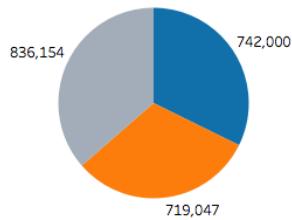
The above image is an excellent example of Data Visualisation, which rather than focussing on what the numbers are, focuses on telling their story. There are multiple Software that are available now at instant access which assist in such easy visualisations and one tool that we are going to cover in this article is **Tableau**.

# Sample Dashboard in Tableau

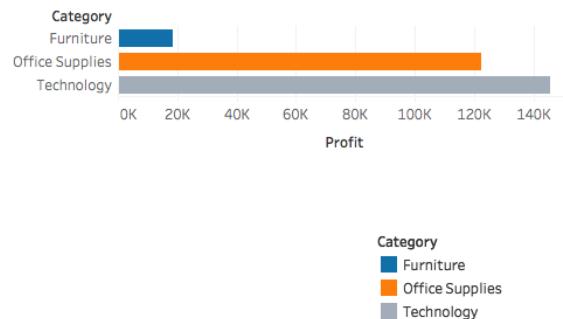
What can you make out from the picture below?

## Category Analysis

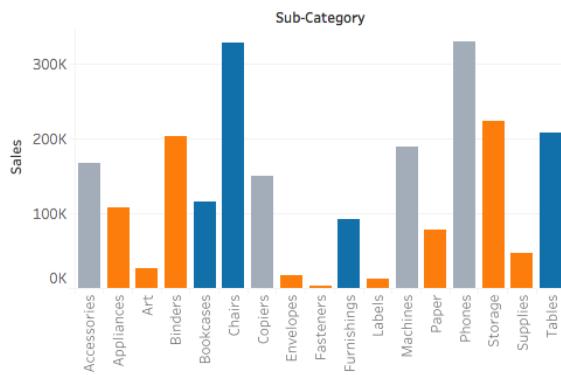
Net Sales Based on Category



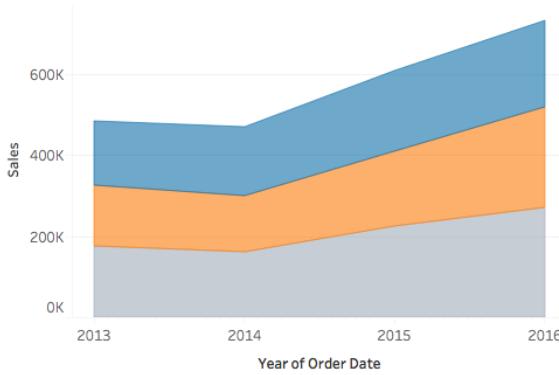
Net Profit



Sales by Sub - Category



Sales Growth over the years



This Dashboard, made on Tableau, represents the Sales and Profit Analysis of a Supermarket.

At a glance, you can see:

1. The Sales distribution of various categories relative to each other
2. Their respective Profit margins.
3. Each Category's Sub – Category Product Sales
4. And finally, the Sales growth of the Categories over the years

So, in this article, we will learn how to make such simple visualisations in Tableau to understand our data well.

# Table of Contents :

1. Overview of Tableau
  1. About
  2. Prerequisites
  3. Installation
2. Getting Started
  1. Connect to Data
  2. Data Visualisations
3. Other Functionalities
  1. Filters
  2. Drill Down and Up
  3. Forecasting
  4. Trend Lines
  5. Clustering
4. Dashboard
5. Story – Bringing it all together
6. End Notes

## 1. Overview of Tableau

### 1.1 What is Tableau?

Tableau is a Data Visualisation tool that is widely used for Business Intelligence but is not limited to it. It helps create interactive graphs and charts in the form of dashboards and worksheets to gain business insights. And all of this is made possible with gestures as simple as drag and drop!

#### What Products does Tableau offer?

	Key Features	Other Features	Operating System	License
Tableau Desktop	Creating Dashboards and Stories locally	Tableau Personal - limited data sources, non connectivity to Tableau Server  Tableau Professional - Full enterprise capabilities	Windows, Mac	Personal - \$999 Professional - \$1999
Tableau Public	A Massive, public, non commercial Tableau Server	All data published in public	-	Free
Tableau Online	Creating Dashboards and Stories on the Cloud	Live Connections	-	\$500 per year per
Tableau Reader	View Dashboards and Sheets locally	Cannot modify workbooks or connect to the server	Windows, Mac	Free
Tableau Server	Connect to Data sources and share Dashboards	Users can directly interact with Dashboards via browser	Windows	Core Licensing

## **1.2 What do you need to know before using Tableau?**

You don't need to know much to use Tableau, but still a basic awareness of all the types of graphs such as bar graph, line charts, histograms etc is preferred.

Along with that it will be beneficial if you possessed some basic understanding of database management (datatypes, joins, drill down, drill up etc ) too. Even if you don't, not a reason to worry since I will be covering all such concepts in this article.

## **1.3 Installation:**

To work on Tableau, you **need** Tableau right?

Out of the five above mentioned products, Tableau Desktop, Public and Online offer Data Visual Creation.

### **Tableau Desktop**

It is available in the following three formats :

1. [Free trial for 14 days](#)
2. If you are a student or a teacher, you get free access to the [Desktop](#) for a full year.
3. [Purchase Tableau](#)

### **Tableau Public**

[Tableau Public](#) is purely free of all costs and does not require any licence. But it comes with a limitation that all of your data and workbooks are made public to all Tableau users.

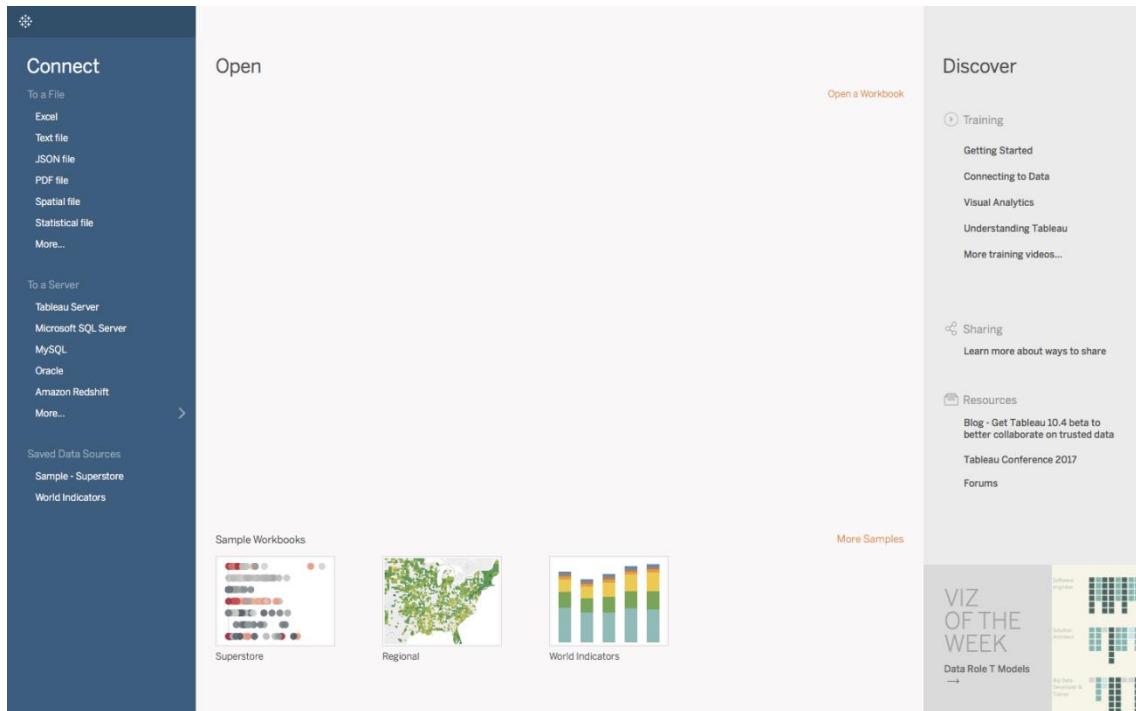
### **Tableau Online**

[Tableau Online](#) is the best option for you, if you wish to make your Workbooks on the Cloud and be able to access them from anywhere.

## **2. Getting Started**

Now that you have the suitable product installed and set up, I am pretty sure that your hands must be tingling with anticipation to finally begin! Well let's not keep you waiting then, go ahead and launch the tool.

## 2.1 Connect to the Data



You should see a screen similar to the one above. This is where you import your data. As is visible, there are multiple formats that your data can be in. It can be in a flat file such as Excel, CSV or you can directly load it from data servers too.

You can see that Tableau itself offers some Sample Workbooks, with pre-drawn charts and graphs. I would suggest going through these later for further exploration.

The best way to learn is to get your hands dirty. Let us start with our Data, which can be found [here](#). The data is that of a United States' Superstore which is deliberating over its expansion. It wishes to know the prospective regions of the country where it could and hence requires your help.

The first thing that you will obviously need to do is import the data onto Tableau. So quickly follow the below steps:

1. Since the data is in an Excel File, click on **Excel** and choose the Sample – Superstore.xls file to get :

The screenshot shows the Power BI desktop application interface. At the top, there's a ribbon with icons for file, home, analyze, and share. Below the ribbon, the title bar says 'Sample - Superstore'. On the left, the 'Connections' pane shows 'Sample - Superstore Excel' is connected. The 'Sheets' pane lists 'Orders', 'People', 'Returns', and 'New Union'. A tooltip 'Drag sheets here' is positioned above the main data grid area. The data grid itself is currently empty. At the bottom, there are standard Excel-like navigation buttons for 'Data Source', 'Sheet 1', and other sheet tabs.

2. You can see three sheets on the screen, but we are only going to be dealing with Orders here, so go ahead and drag the same on *Drag sheets here* :

This screenshot shows the 'Orders' sheet selected in the Power BI desktop interface. The data grid displays the following sample data:

#	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders	#
Data	Superstore Sales	F3	F4	F5	F6	F7	F8	F9	F10	F11	Orders
null	21	null	null	null	null	null	null	null	null	null	null
null	9994	null	null	null	null	null	null	null	null	null	null
null	Order ID	null	null	Ship Mod	Customer ID	Customer Name	Segment	Country	City	State	
7,981	CA-2011-103800	03/01/2013	07/01/2013	Standard Class	DP-13000	Darren Powers	Consumer	United States	Houston	Texas	77,C
740	CA-2011-112326	04/01/2013	08/01/2013	Standard Class	PO-19195	Phillina Ober	Home Office	United States	Naperville	Illinois	60,5
741	CA-2011-112326	04/01/2013	08/01/2013	Standard Class	PO-19195	Phillina Ober	Home Office	United States	Naperville	Illinois	60,5
742	CA-2011-112326	04/01/2013	08/01/2013	Standard Class	PO-19195	Phillina Ober	Home Office	United States	Naperville	Illinois	60,5
1,760	CA-2011-141817	05/01/2013	12/01/2013	Standard Class	MB-18085	Mick Brown	Consumer	United States	Philadelphia	Pennsylvania	19,1
5,328	CA-2011-130813	06/01/2013	08/01/2013	Second Class	LS-17230	Lycoris Saunders	Consumer	United States	Los Angeles	California	90,0
7,181	CA-2011-106054	06/01/2013	07/01/2013	First Class	JO-15145	Jack O'Briant	Corporate	United States	Athens	Georgia	30,6
7,475	CA-2011-167199	06/01/2013	10/01/2013	Standard Class	ME-17320	Maria Etezadi	Home Office	United States	Henderson	Kentucky	42,4
7,476	CA-2011-167199	06/01/2013	10/01/2013	Standard Class	ME-17320	Maria Etezadi	Home Office	United States	Henderson	Kentucky	42,4
7,477	CA-2011-167199	06/01/2013	10/01/2013	Standard Class	ME-17320	Maria Etezadi	Home Office	United States	Henderson	Kentucky	42,4

At the bottom, there are buttons for 'Data Source', 'Sheet 1', and other sheet tabs. A 'Go to Worksheet' button is also present.

Uh oh, the imported data looks a bit different for the first few rows. Don't worry, the solution lies right ahead.

## Data Interpreter

3. You see the option of **Use Data Interpreter**? Click on it to get the following clean view :

The screenshot shows the Tableau Data Interpreter interface. On the left, the 'Connections' pane shows a connection to 'Sample - Superstore Excel'. The 'Sheets' pane lists 'Orders', 'People', 'Returns', 'Orders A1:B3', and 'New Union'. The main workspace displays a clean dataset titled 'Orders (Sample - Superstore)'. The top right shows connection options ('Live' or 'Extract') and filter settings ('0 | Add'). The data preview shows 1,000 rows of clean data, including columns like Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, and City. A note in the workspace states: 'Cleaned with Data Interpreter. Data Interpreter removed some data. [Review the results](#). (To undo changes, clear the check box.)'. A 'Go to Worksheet' button is visible at the bottom left of the preview area.

All that messy data magically disappeared!

If you open the Excel data file, you will see some metadata in it, i.e. information about data :

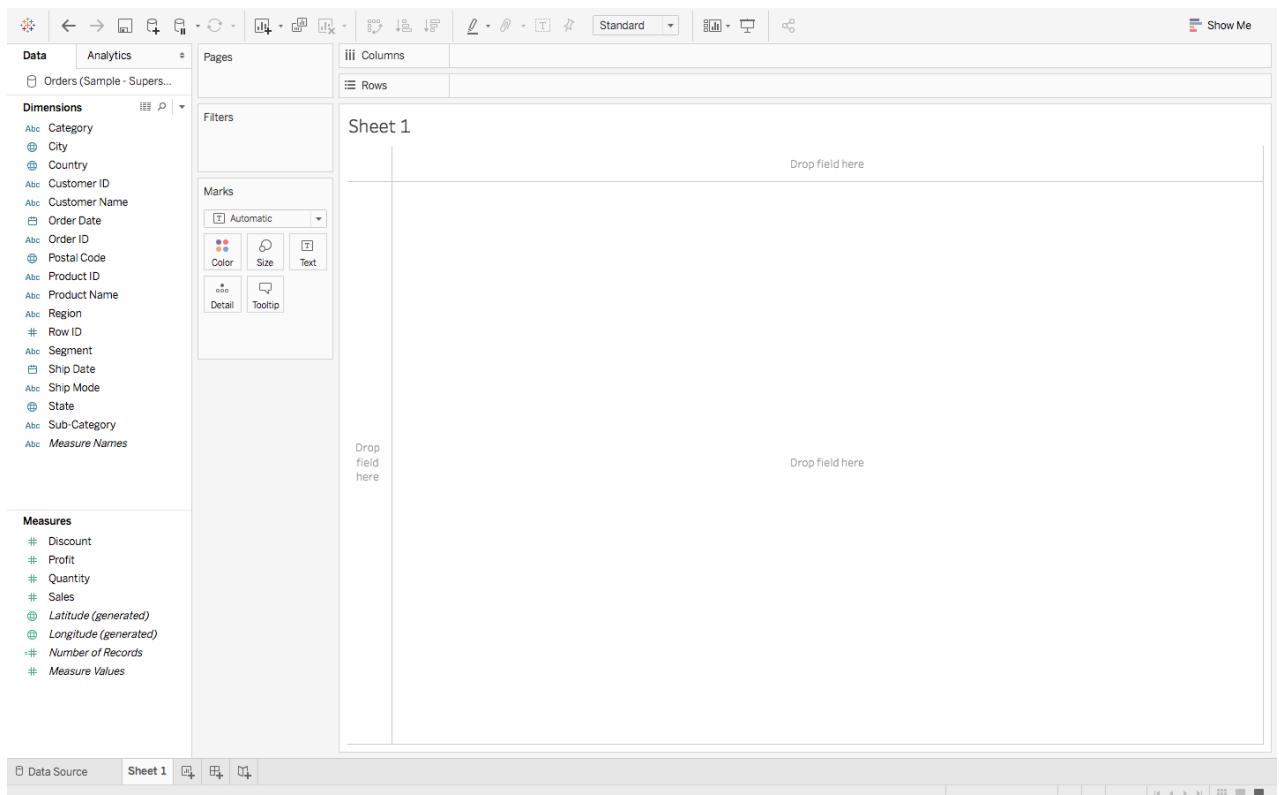
	A	B
1	Data	Superstore Sales
2	Fields	21
3	Rows	9994

Tableau imports the entire data file as is, but anticipating such discrepancies, explicitly provides a solution in the form of a Data Interpreter. If you wish to view the exact changes that it made, click on **Review the results**, and choose the **Orders** tab in the opened Excel sheet.

As it will show, it simply removed the erroneous data.

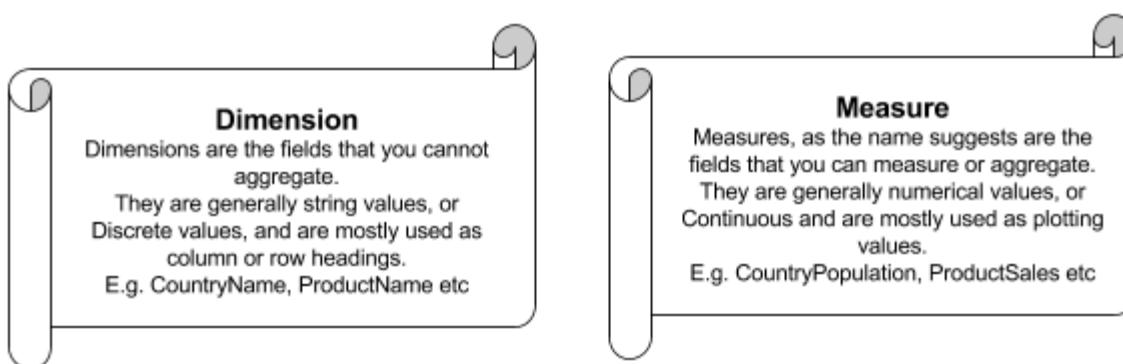
## 2.2 Data Visualisations

As soon as you had imported your dataset, next to the Data Source tab near the bottom of the screen, you immediately must have seen **Go to Worksheet**. A Worksheet is where you make all of your graphs, so click on that tab to reach the following screen :

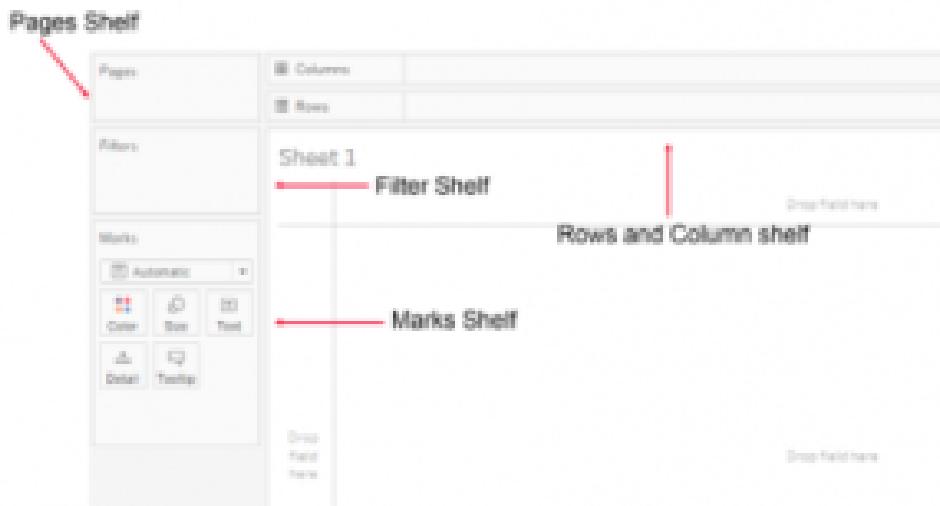


Don't get overwhelmed by the various elements that you see here, we will cover them all one by one.

Let's start with **Dimensions and Measures** :



Moving onto **Shelves** :



Visualisation in Tableau is possible through dragging and dropping Measures and Dimensions onto these different Shelves.

**Rows and Columns** : Represent the x and y – axis of your graphs / charts.

**Filter** : Filters help you view a strained version of your data. For example, instead of seeing the combined Sales of all the Categories, you can look at a specific one, such as just Furniture.

**Pages** : Pages work on the same principle as Filters, with the difference that you can actually see the changes as you shift between the Paged values. Remember that Rosling chart? You can easily make one of your own using Pages.

**Marks** : The Marks property is used to control the mark types of your data. You may choose to represent your data using different shapes, sizes or text.

And finally there is **Show Me**, the brain of Tableau!



When you drag and drop fields onto the visualisation area, Tableau makes default graphs for you, as we shall see soon, but you can change these by referring to the **Show Me** option.

**Note :** Not every graph can be made with any combination of Dimensions or Measures. Each graph has its own conditions for the number and types of fields that can be used, which we shall discuss next.

## 2.3 Various Graphs and Charts

So far we have pretty much covered the requisite theoretical knowledge. Lets finally begin with some visualisations now.

I personally prefer to start from the shallow side of the pool, slowly swimming towards the deeper end. So I would suggest beginning by getting an overview of the Superstore Sales and Profit Statistics. That would include the Net Sales, the Net Profit and the growth of the two measures, to name a few. Here is a gist of what we will be making :

## Overview of the Superstore

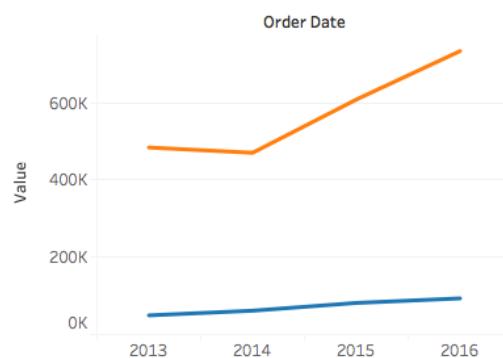
### Net Statistics

Profit	Quantity	Sales
286,397	37,873	2,297,201

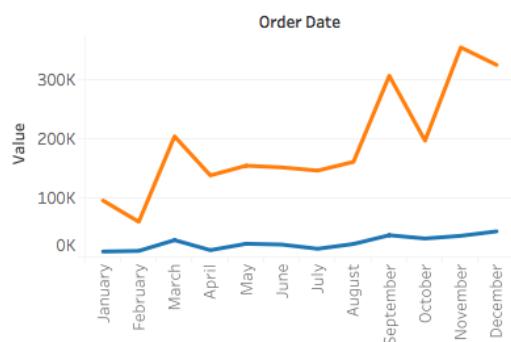
- Year of Order Date
- (All)
  - 2013
  - 2014
  - 2015
  - 2016

- Measure Names
- Profit
  - Sales

### Sales and Profit Growth



### Peak Sales and Profit Months



Year of Order Date	Sales	
2013	484,247.4981	49,544
2014	470,532.509000001	61,619
2015	608,473.83	81,727
2016	733,947.023199999	93,508

Month of Order Date	Sales	
January	95,656.6036	9,203
February	59,640.1474	10,286
March	204,475.8308	28,608
April	138,260.6826	11,567
May	154,732.1607	22,377
June	151,804.5883	21,035
July	146,416.889	13,864
August	161,194.253	22,043
September	307,148.0297	36,889
October	197,098.3697	31,304
November	355,241.428	35,833
December	325,531.8775	43,387

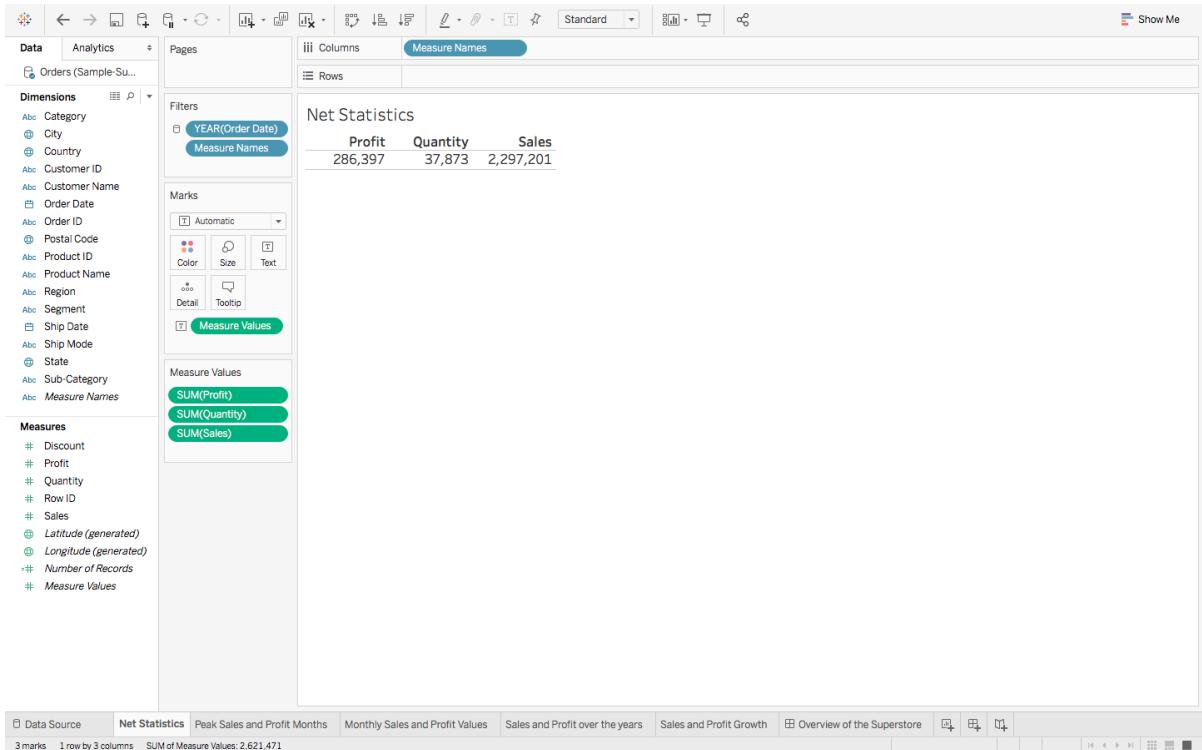
From what can be observed, the net Sales are on the rise, but the Profit is creeping up slowly. We can also quite clearly see the peak Sales Months, which could be attributed to various reasons. We can only know more as we explore more.

Before we start, there is one thing that I would like to recommend and that is you name your Worksheets as being done here. Since I will be referencing them back and forth throughout the article, it will be easier for you to follow.

Let's begin with the simplest visualisation, and that is displaying the **Net Statistics** numbers. Tableau, being as smart as it is, automatically computes such values under **Measure Names** and **Measure Values**. Follow these steps to make what is called a **Text Table** :

1. Drag **Measure Names** from Dimensions onto the central empty area so that you see a Text Table.
2. **Measure Names** will be displayed automatically onto Rows, so drag it from Rows to Columns.

3. Since we don't really need Measures like the Row ID, Discount etc, you can drag them off from below the Marks Pane, to get something like this :

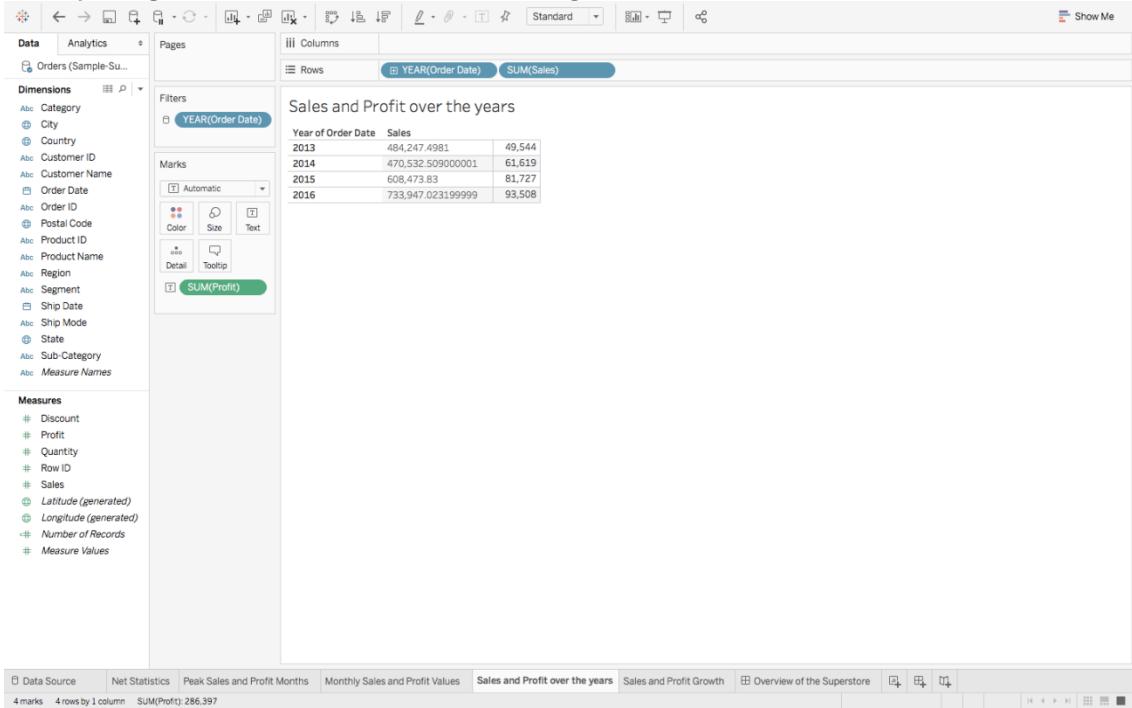


**Note :** Don't get confused by the different colours of the fields that you see. Just remember one small trick : Blue means **Discrete** and Green, **Continuous**.

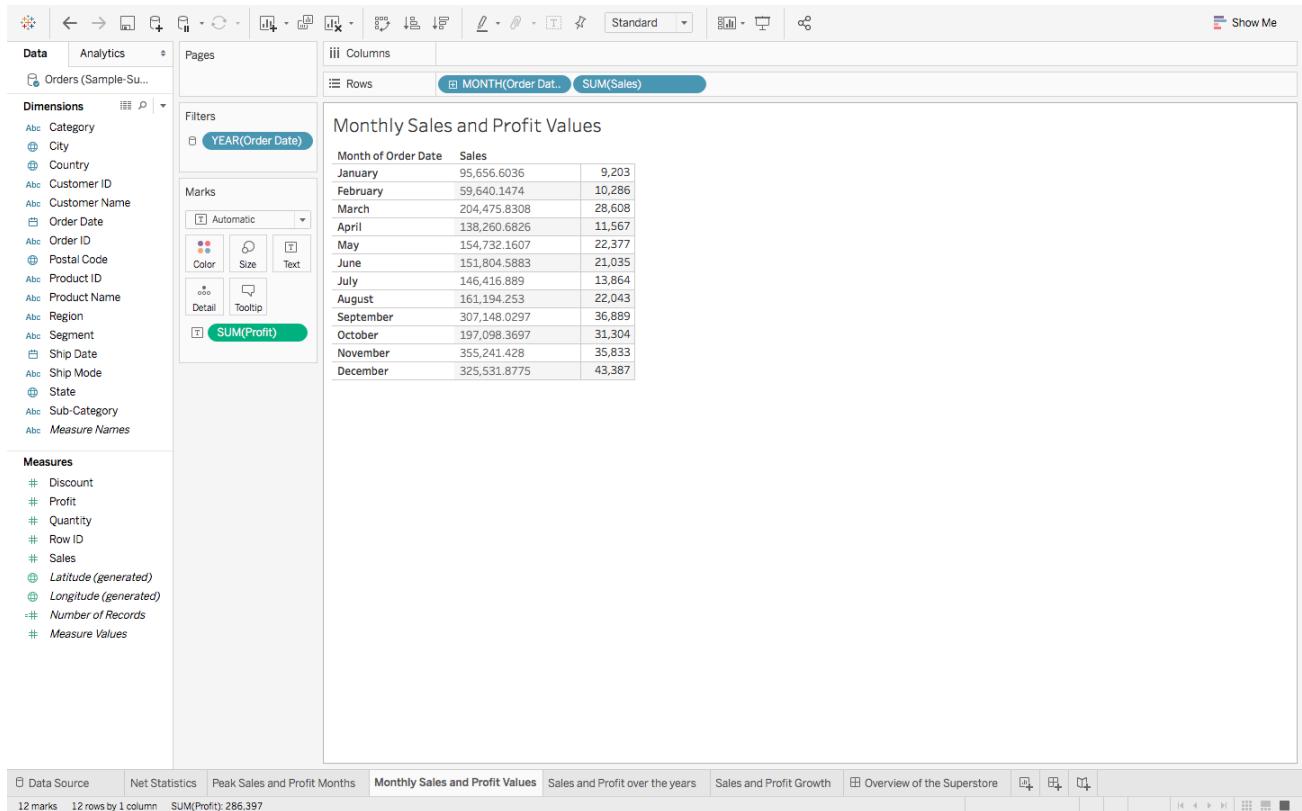
So we have the net Sales and Profit values, let's delve a little deeper by getting the **Sales and Profit Values over the years**. Lets make another, but a more detailed, Text Table :

1. Drag **Order Date** from Dimensions and **Sales** from Measures to Rows.
2. Right click on the green **Sales Pill**, and select **Discrete**, in place of **Continuous**, since we want the explicit values and not the bar graphs.

### 3. Finally drag Profit on the 'abc' column to get :

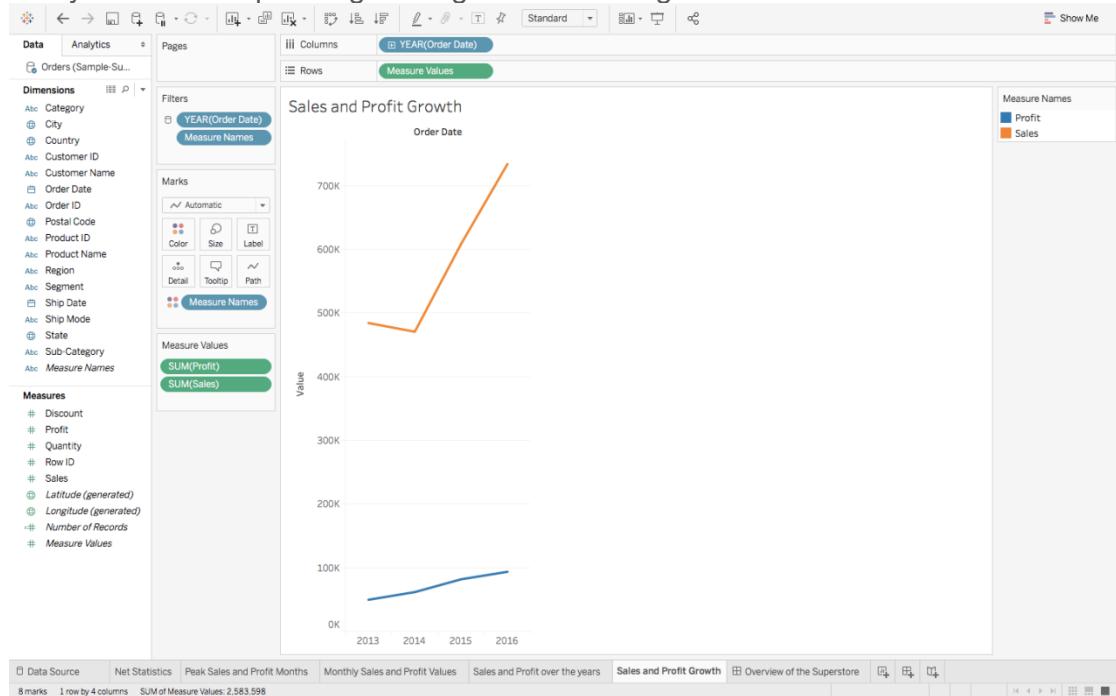


4. Do the same thing for **Monthly Sales and Profit Values**, but this time change the format of **Order Date**, from **Year to Month**, by right clicking on **Order Date** in the **Rows**, and choosing Month, to get something like this :

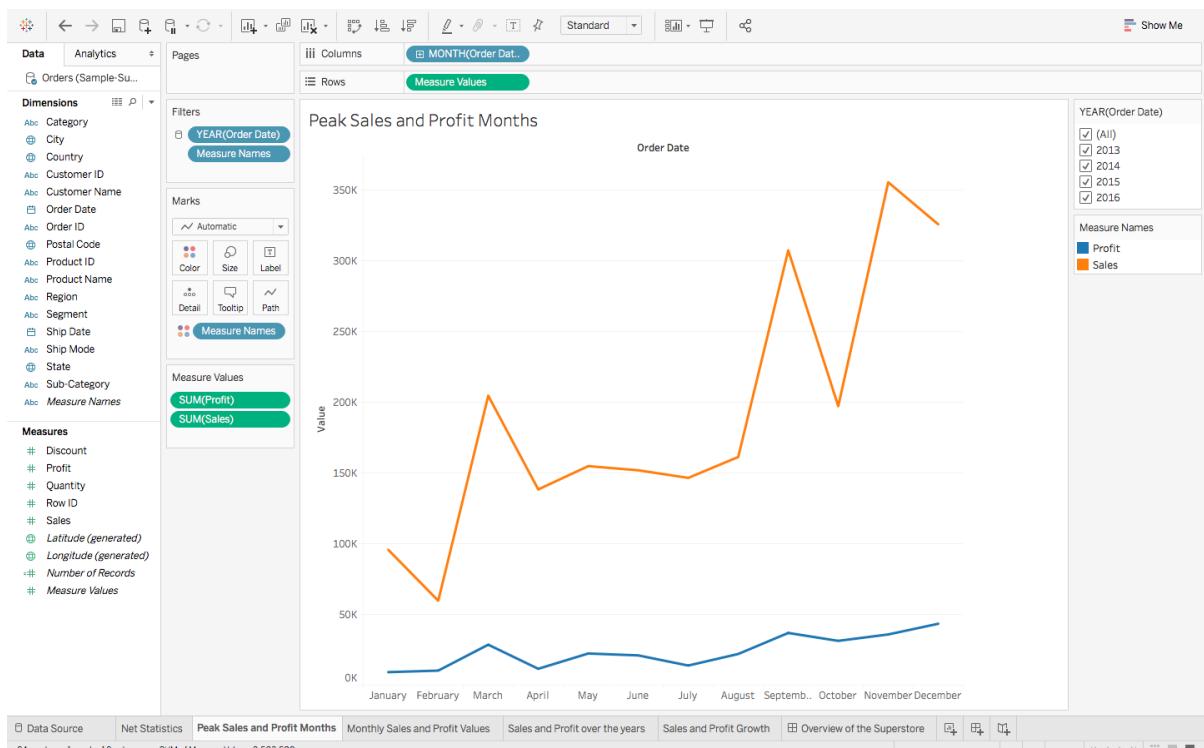


We have just covered the numeric part of the Dashboard, but that is not its selling point. It's the **Line Charts**. Lets quickly learn how to make one :

- To create the chart of **Sales and Profit Growth**, drag **Order Date** over the **Columns**, **Sales** over **Rows** and then **Profit** over the formed **Sales** axis – so that you see an equals sign – to get the following :



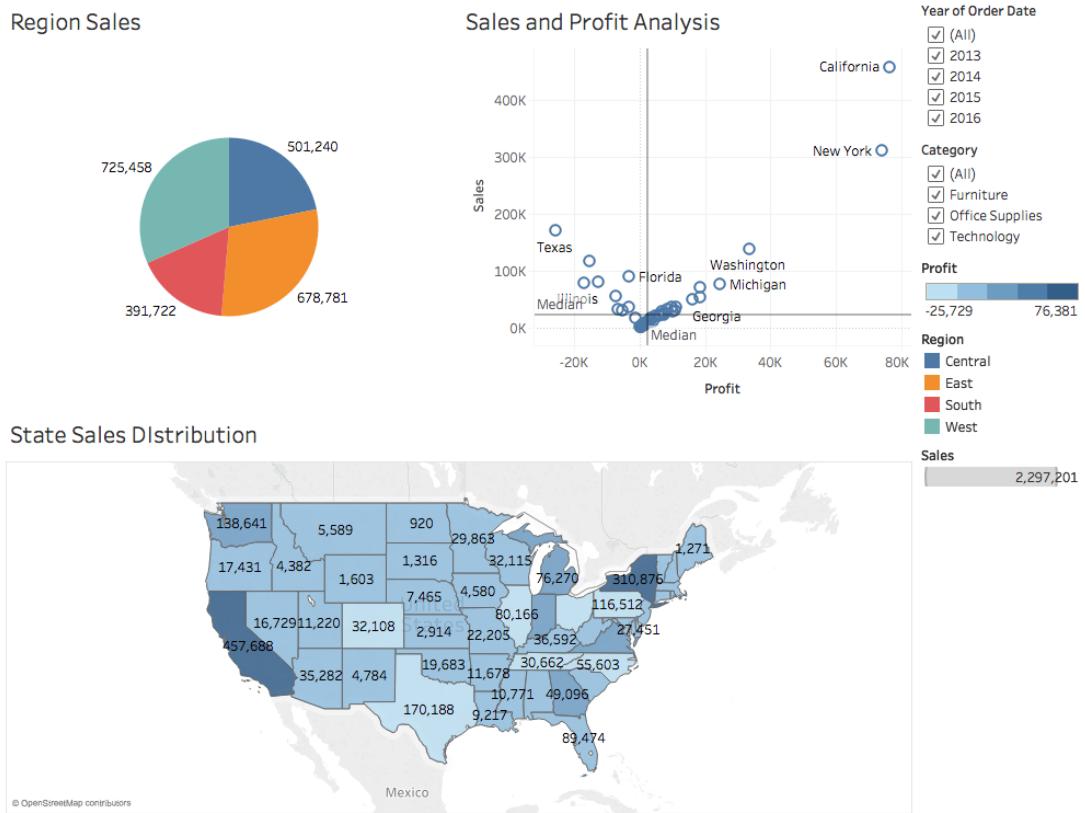
- Repeat the same to find the **Peak Sales and Profit Months**, but again change the format of **Order Date**, from **Year** to **Month**, and get :



If you were to click on **Show Me**, you will see the different types of Line Charts that you can make, and if you were to hover over each of them, you will get to see their Dimension and Measure requirements too. In case you ever feel lost, I recommend referring to Show Me.

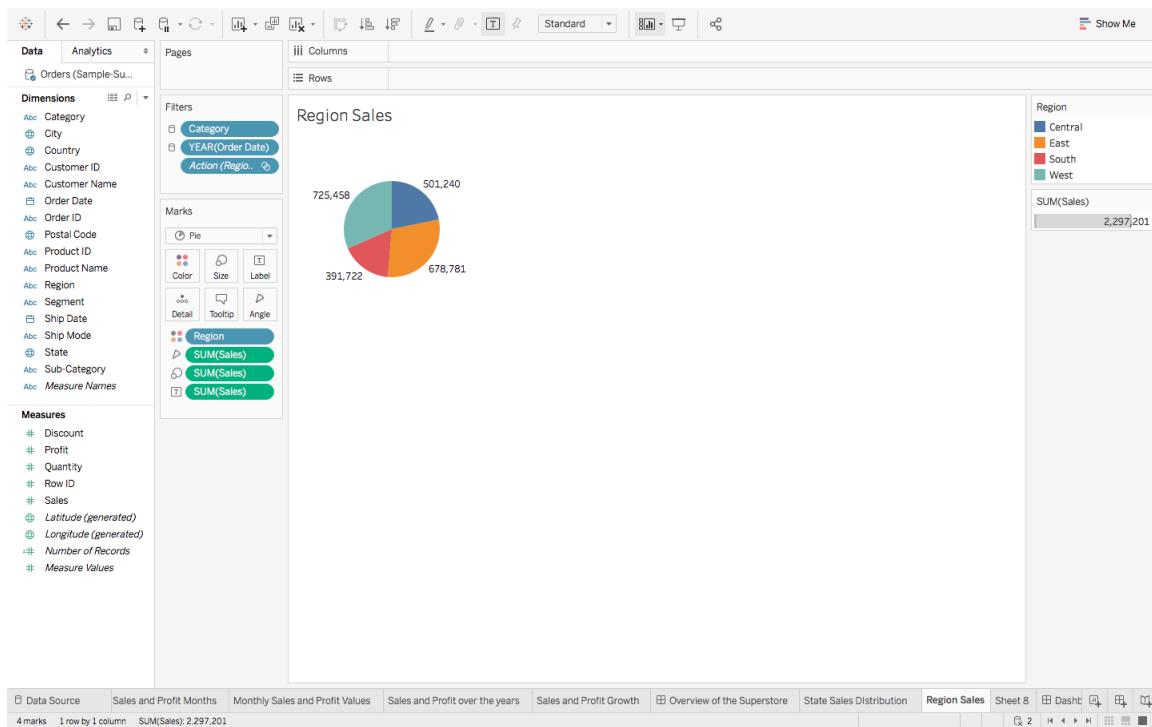
With the previous visualisations, we had gotten a brief overview of the Superstore. Let's dig a little deeper now. The next thing that I can think of exploring is the demographic of the Sales and Profit. What are the States that have the highest Sales Revenue, which ones are generating the maximum Profits:

## Demographic Analysis



Before discussing the inferences, let's first create the **Pie Chart of Region Sales**:

1. Drag **Regions** onto Rows and **Sales** onto Columns.
  2. Go to **Show Me**, and select the **Pie Chart**.
  3. And finally drag **Sales** over the **Label** in the **Marks Pane** to get :

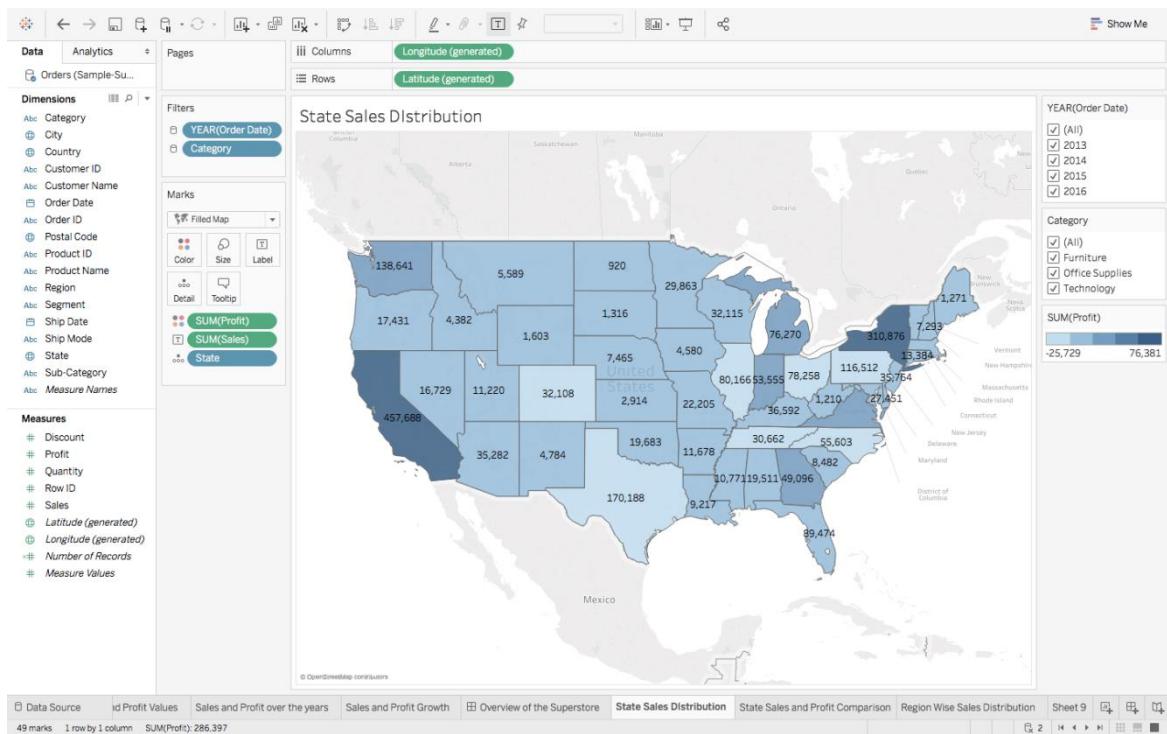


From the visual it's pretty evident that the two opposite ends, East and West are leading in the Sales game. Let's dissect this a bit more.

**Note :** Whenever you have some geographical data, it is always advisable to plot and see it on a Map to gain better insights.

So, we are now going to make the **Map Chart of State Sales Distribution :**

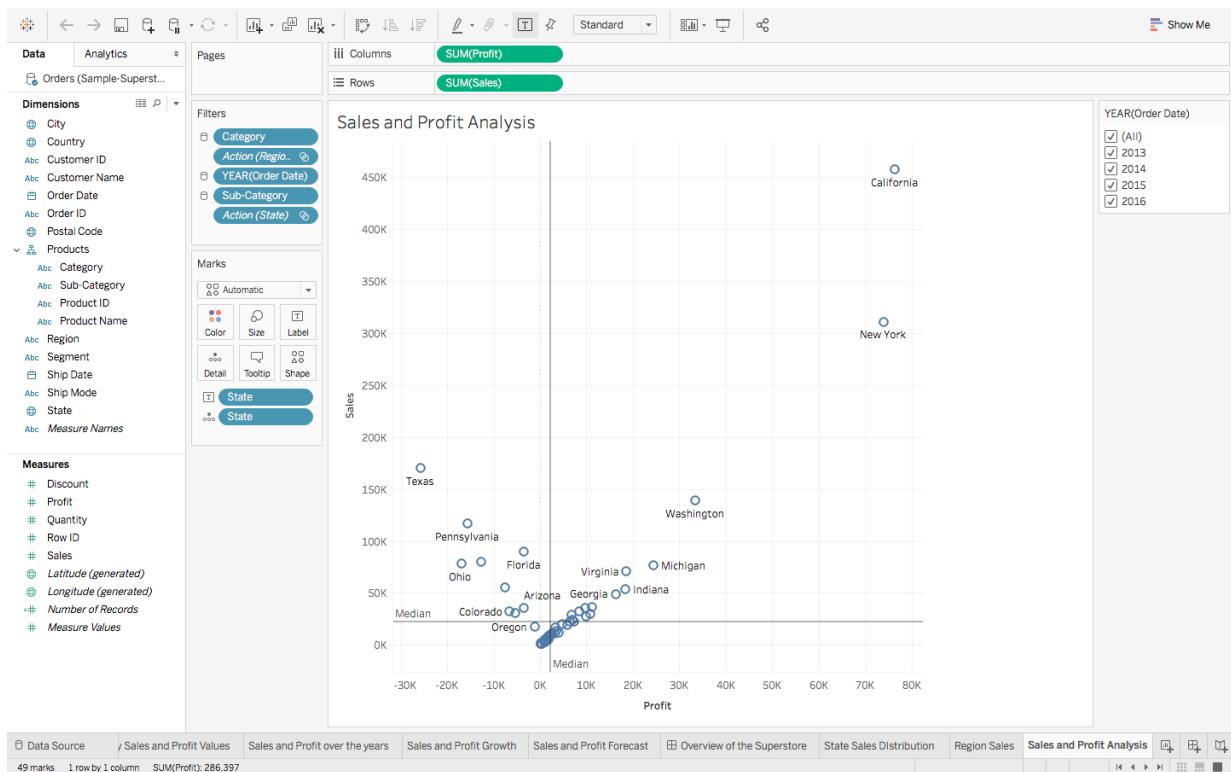
1. Since its the States that we wish to analyse, drag **States** onto the empty area, so that you automatically see a Map, with small Circles. Follow this step by dragging **Profits** next. You will notice the size of these circles changing to represent the varying values of Profits. This is called a **Symbol Map**. But we are going to convert this into a **Filled** one, by going to Show Me, and selecting the **Filled Map**.
2. Drag **Profits** again, but this time onto **Label** in the **Marks Pane**, to view the Profit Values mapped as well, like so :



California and New York are the top most sellers from West and East region, but unfortunately there are other States such as Texas, Colorado which even after having good Sales, have negative Profits! This is certainly not good news for the Superstore. You can perceive a good analysis for the other States as well.

And lastly, here are the steps for making the **Scatter Plot of Sales and Profit Analysis** :

1. Drag **Sales** onto Rows, and **Profit** onto Columns. You will see one tiny circle, which actually represents the Total Sales and Profit Values.
2. To get more information, drag **States** onto the graph created, so that these circles / bubbles scatter to represent the individual States.
3. To better understand the central tendency of the data, we have also added a **Median axis** as Reference Line. This can be easily done by right clicking on the **Sales / Profit Axis –> Adding Reference Line** and choosing Median over the default Average Reference.
4. Finally for some more insight, drag **States** again, but this time onto **Label** in the Marks Pane, and get:



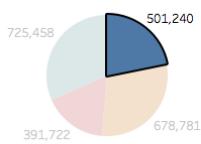
The findings from the **Map chart** become more prominent with the following **Scatter plot** inferences :

- The states in the top right, with high Sales and high Profits mean good business for the organisation.
- States with positive Sales and Profits, but near the two respective axis are the ones where there is some scope of improvement.
- Whereas the states that belong to the 2nd or 3rd quarter are the ones which are not generating much revenue.

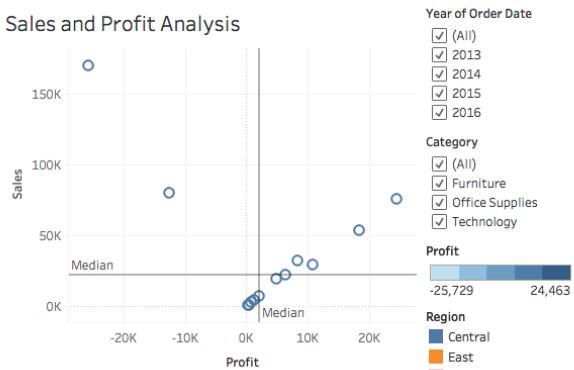
One of the great things about Tableau is that it lets you interact with the visuals. Have a look at an example :

## Demographic Analysis

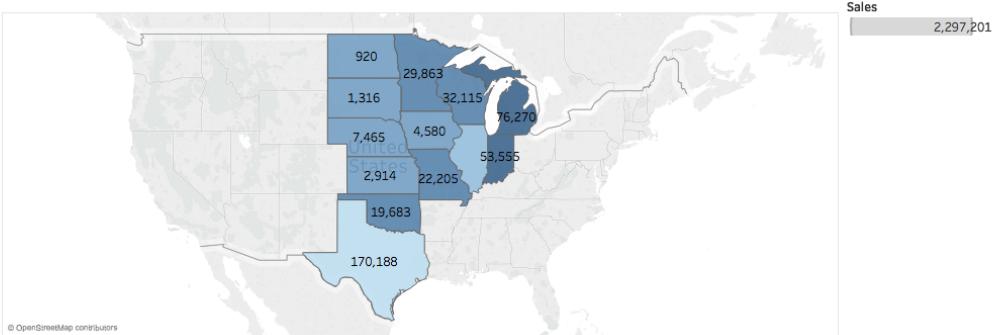
Region Sales



Sales and Profit Analysis



State Sales Distribution



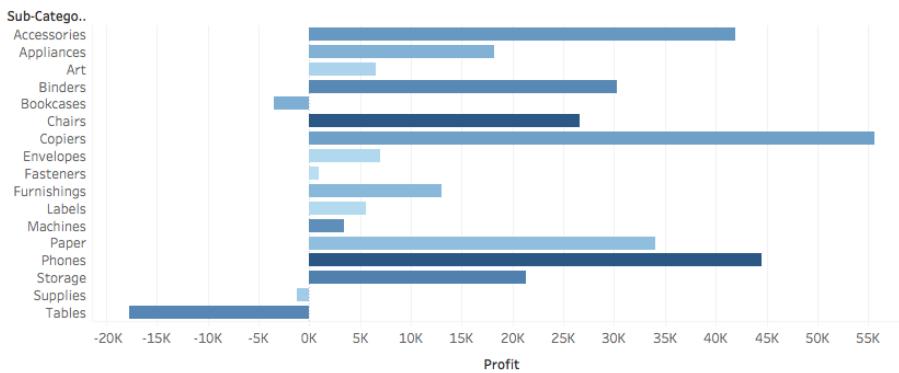
When we clicked on the Central Region, it highlighted and showed the Central States of US, along with their respective Sales and Profit scatter. Here we used the chart as a Filter itself which is a feature of a Dashboard. We shall learn how to make one at a later stage.

There is one pretty important analysis that we have yet to touch, and that is Product Statistics. High Sales could be easily attributed to the high cost of the products being sold. Also, when you are considering expansion, you will want to know the Sales distribution of the Products too:

Product Category Sales

Category	Year of Ord..	Order Date												
		January	February	March	April	May	June	July	August	Septem..	October	Novemb..	Decemb..	
Furniture	2013	6,243	1,840	14,574	7,945	6,913	13,206	10,821	7,320	23,816	12,304	21,565	30,646	
	2014	11,740	3,134	12,500	10,476	9,375	7,714	13,674	9,639	26,273	12,027	30,881	23,086	
	2015	7,623	3,926	12,473	13,406	15,031	12,027	13,199	13,619	26,739	10,131	33,659	37,069	
	2016	5,964	6,866	10,893	9,066	16,958	19,009	11,813	15,442	29,028	21,884	37,057	31,407	
Office Supplies	2013	4,851	1,072	8,606	11,155	7,136	12,953	15,121	11,379	27,423	7,211	26,862	18,006	
	2014	1,809	5,368	15,883	12,559	9,114	10,648	4,720	11,735	19,306	8,673	21,218	16,202	
	2015	5,300	6,683	17,458	10,640	13,007	10,902	12,677	9,219	23,286	14,799	21,428	38,112	
	2016	21,704	7,408	14,550	15,072	13,737	16,912	10,241	30,060	31,896	23,037	31,472	30,437	
Technology	2013	3,143	1,609	32,511	9,195	9,600	8,436	8,004	9,210	30,538	11,938	30,201	20,893	
	2014	4,625	3,449	10,344	11,161	11,643	6,435	10,371	15,525	19,017	10,705	23,874	35,632	
	2015	5,620	12,259	21,255	15,203	28,653	16,502	12,564	10,427	22,883	31,533	27,105	22,057	
	2016	17,035	6,027	33,429	12,383	13,567	17,061	23,210	17,619	26,943	32,856	49,919	21,985	

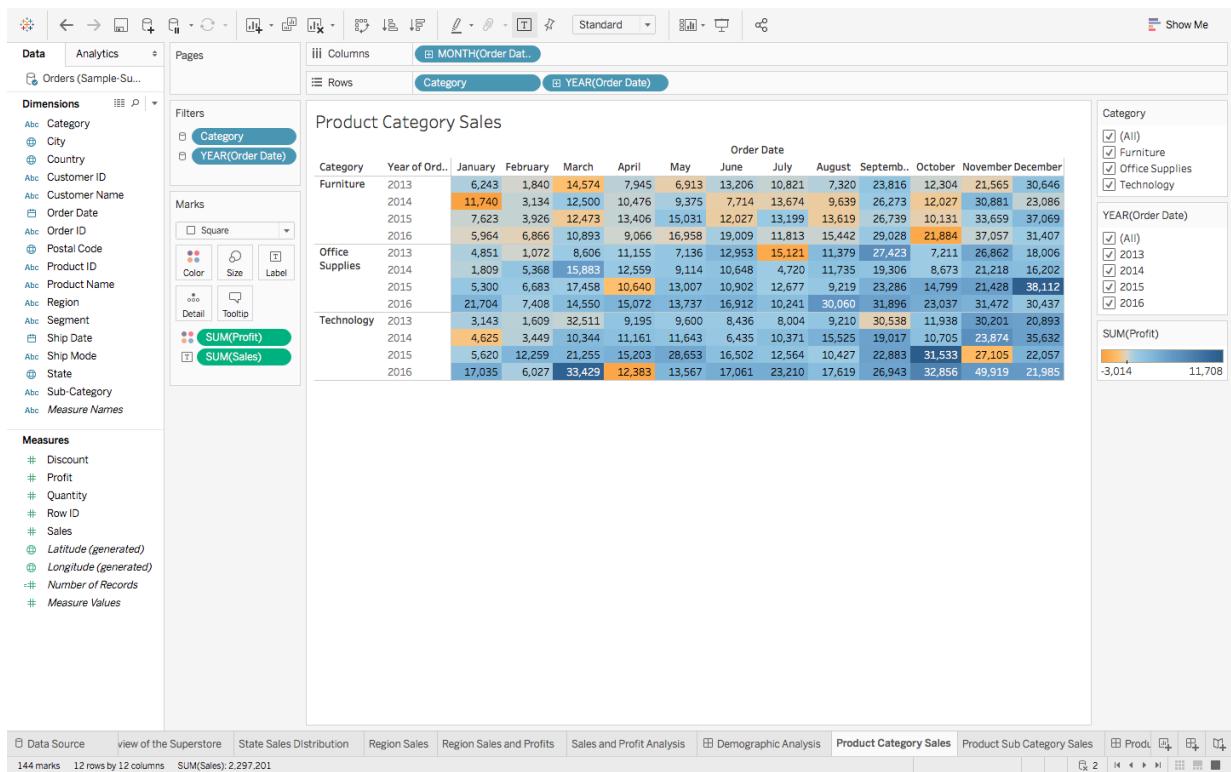
### Product Sub Category Sales



Here we have visualised not just the Sales but also the Profits.

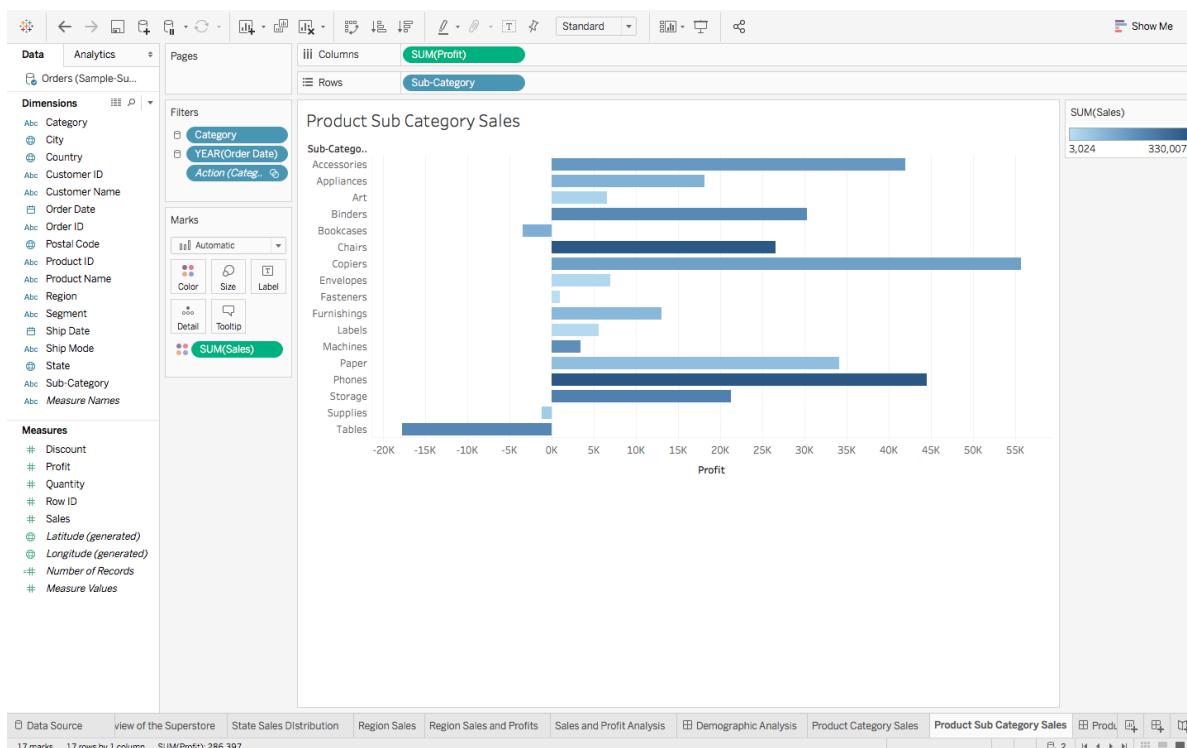
Its quite surprising to see Categories that have high Sales, generating negative profits, like Technology in November 2015, or Furniture in October 2016 and this is inferred from the first chart, which is also called a **Highlight Table**. As the name suggests, it highlights the relative proportion of the Measure Values of our data. So let's learn how to make one :

1. Drag **Category** and **Order Date ( Year )** in Rows.
2. Drag **Order Date (Month)** over Columns, and **Sales** over the empty 'abc' fields
3. Select **Highlight Table** from Show Me, and drag **Order Date ( Year )** back to Rows, in case it got re-shuffled.
4. Finally drag **Profits** over **Colour** in the **Marks Pane**, to get :



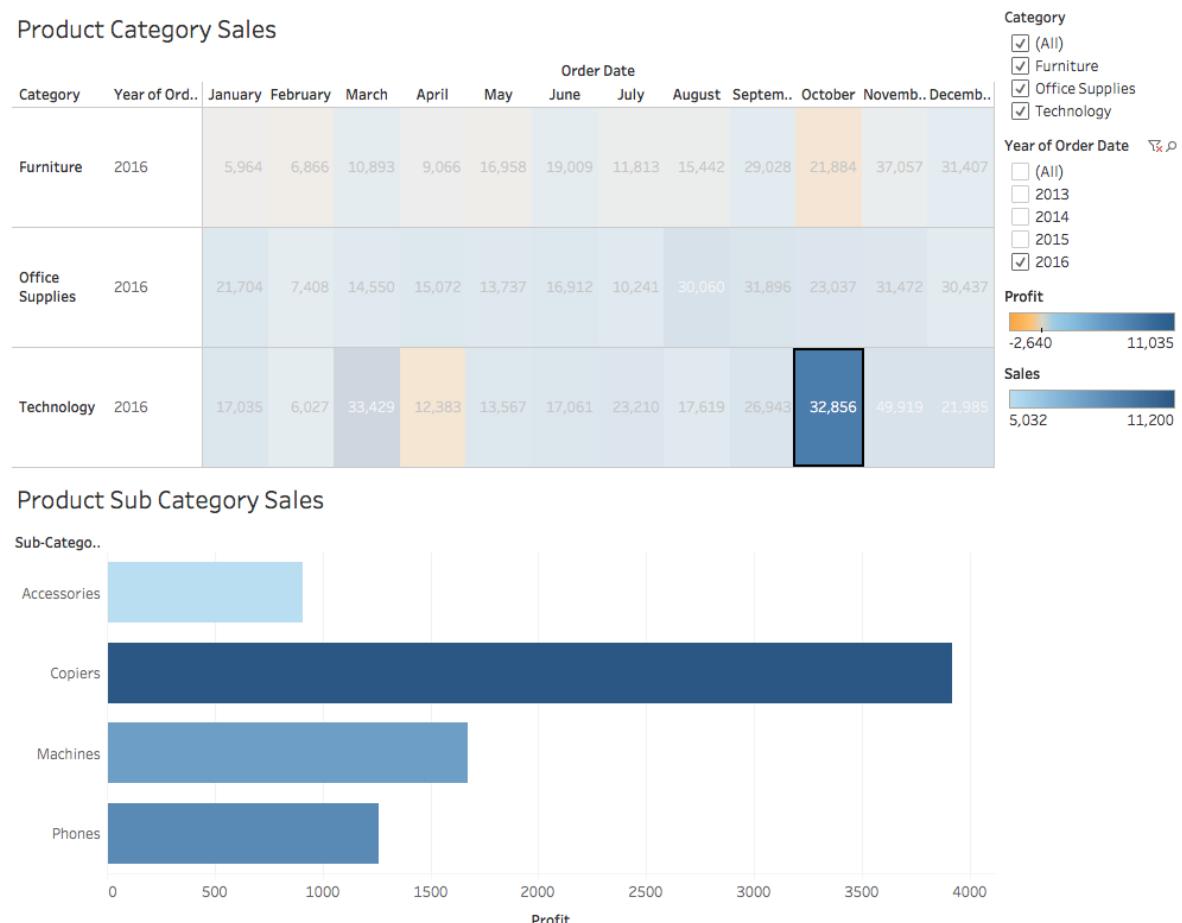
## The Product Sub Category Sales is a Bar Chart, which is also quite easy to make :

1. Just drag **Sub – Category** over to the Rows
2. Drag **Profit** onto the Columns.
3. Go to Show Me and choose the **Horizontal Bars**
4. For some customisation, drag **Sales** over **Colour** in the Marks Pane to attain this final visualisation :



From the the above graph, we are getting a good idea of the Net Sales and Profit margins of the various products. Notice that even though Tables' Sales are quite high on the scale, it's the only product with the least profit.

Now, just like before, consider an interaction with the visualisation :



We are now able to view each Category's Products' Sales and Profits, at a low level granularity of Year and Month!

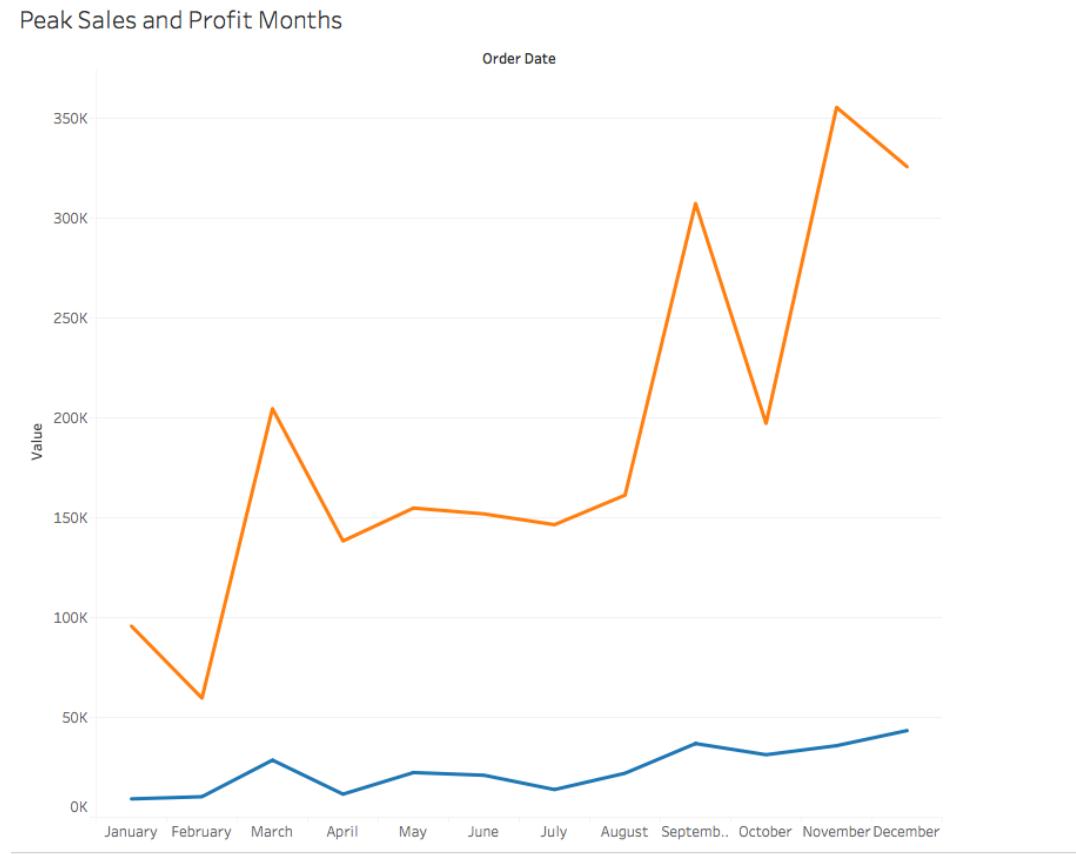
### 3. Other Functionalities

Congratulations! You have now covered one of the important aspects of Tableau! But it's not the end of your learning just yet. Tableau offers some advanced functionalities too, some of which we will cover next :

### 3.1 Filters

Till now we have only made simple charts, that actually provide cumulative data, that is combined data over the lifetime of the Superstore. To look at Sales of a particular Year, a Month, for a certain Product, or to basically view the distinct aspects of the data, **Filters** are the way to go.

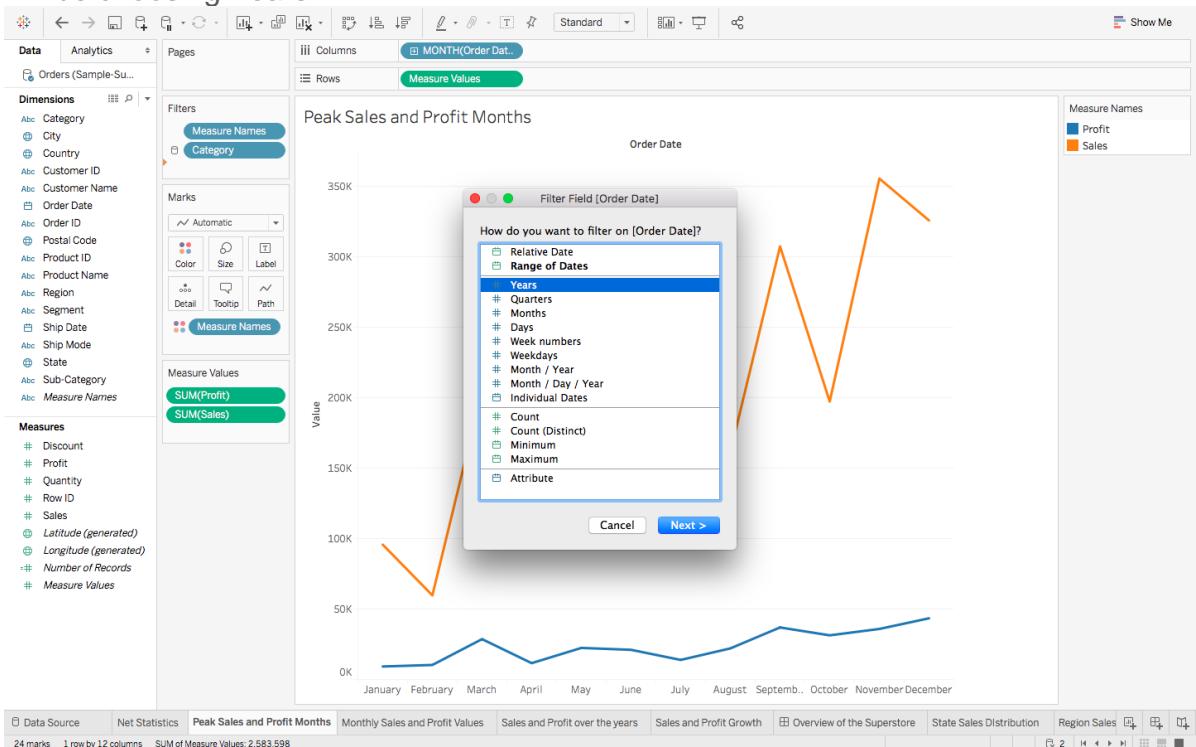
Let's head back to the first ever Chart that we had made, of **Peak Sales and Profit Months** :



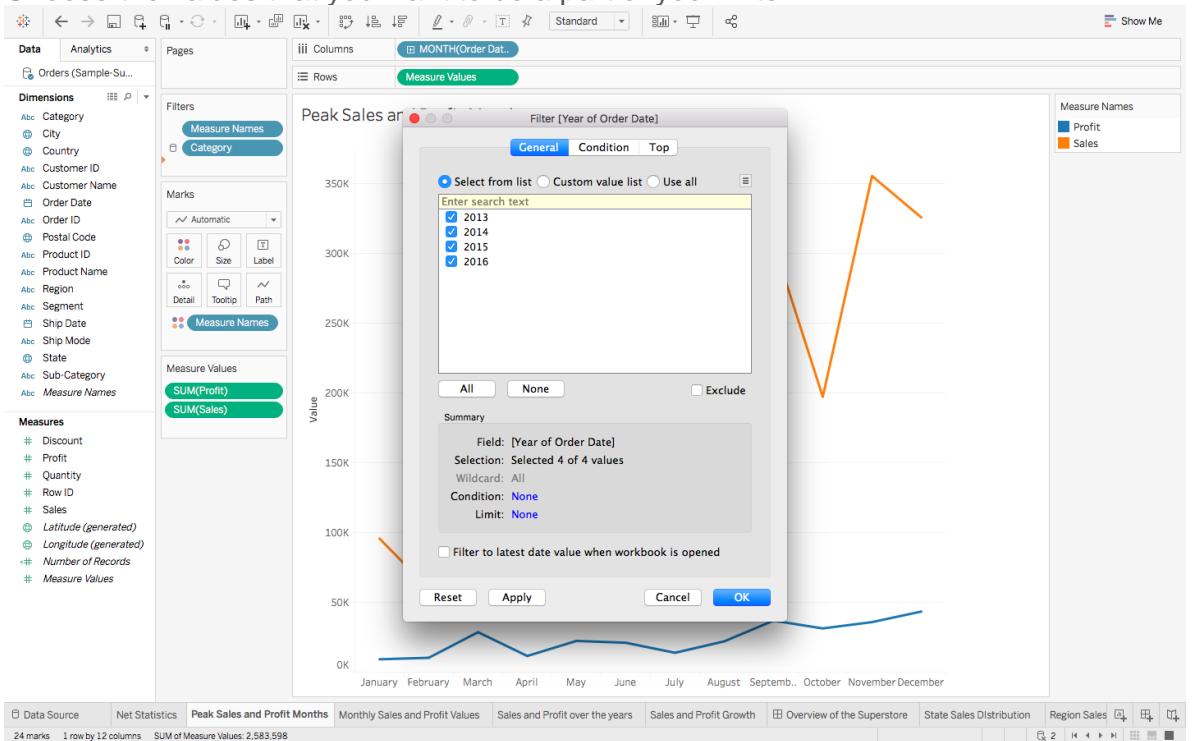
The visual here is an accumulation of all 4 years of data, for all Regions, States, Categories and Sub Categories.

The steps of turning any Dimension into a Filter are the same. Let's first experiment with the Order Date (formatted to Year) :

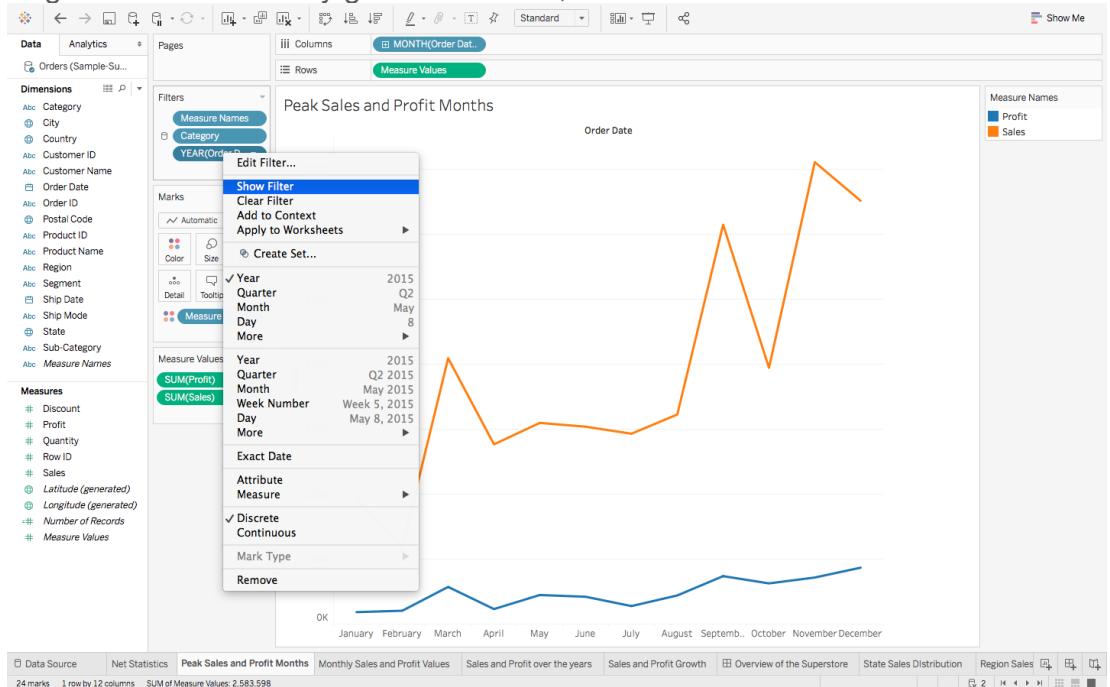
1. Drag the Dimension to the **Filters' Shelf**, to see the following pop up. Here we will be choosing Years :



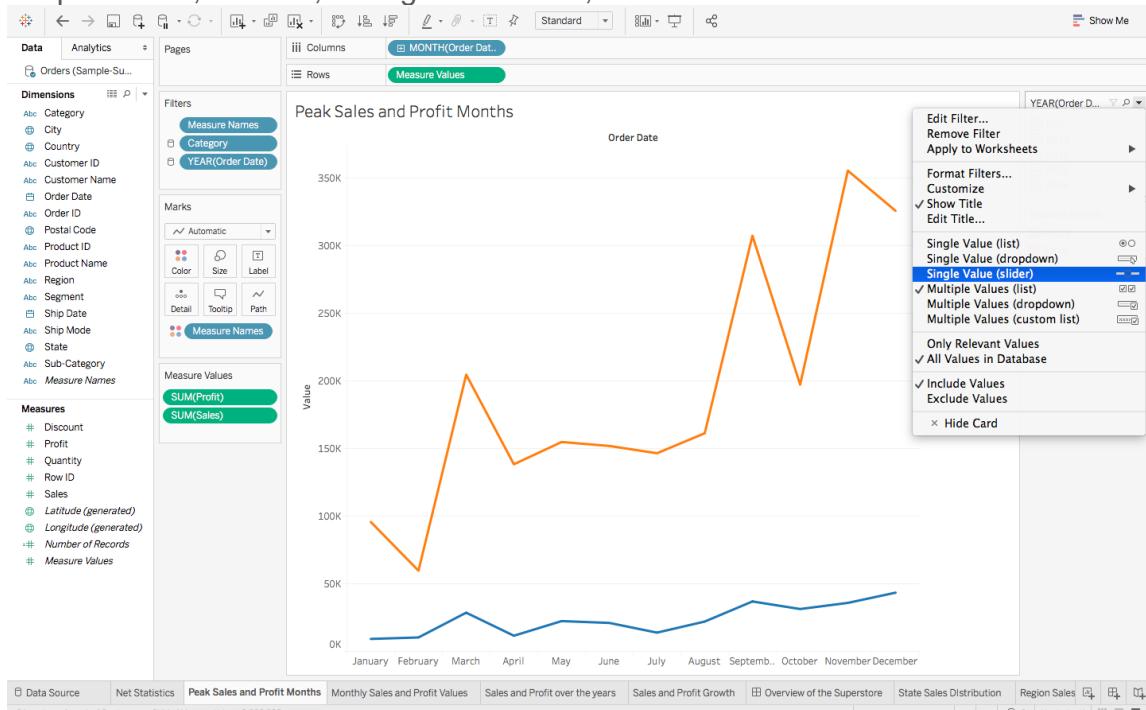
2. Choose the values that you want to be a part of your Filter :



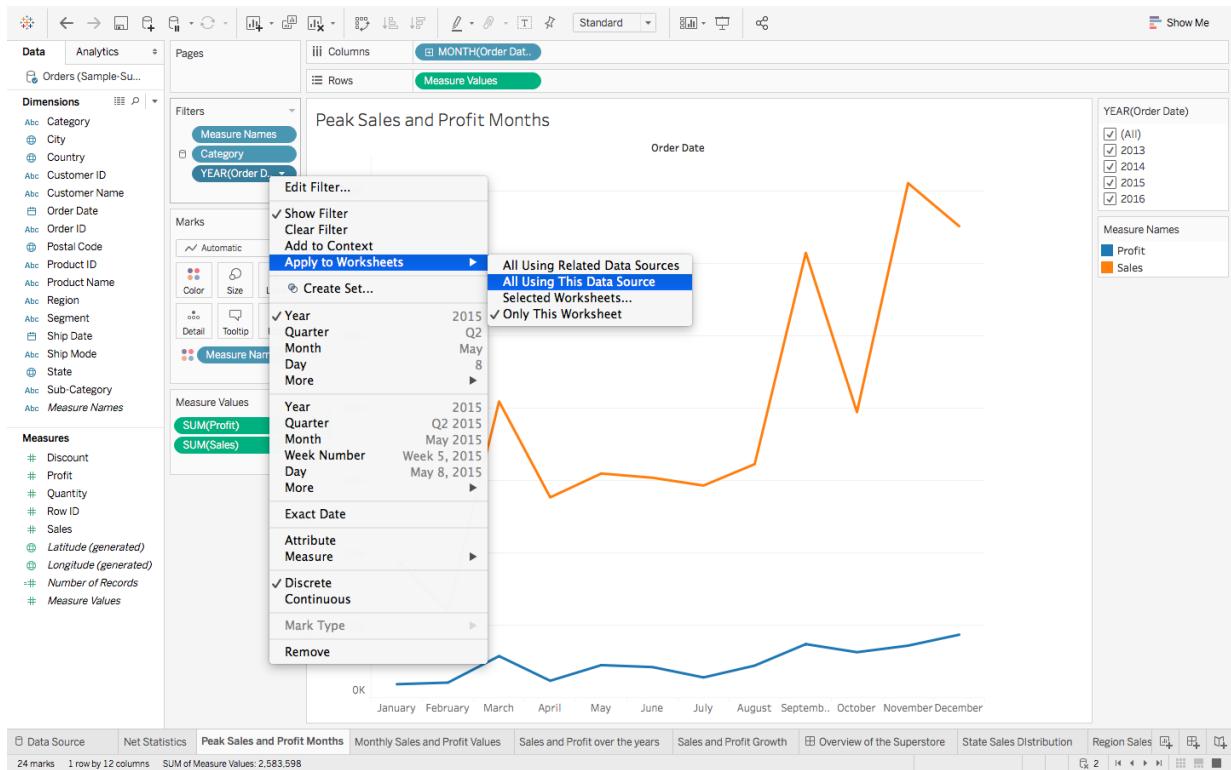
3. Right click on the newly generated Filter, and then choose Show Filter :



4. You can also change the format of your Filter, for example whether you wish for a Dropdown list, a Slider, a Single Value List, etc :



5. If you feel that some of your filters can be applied to other sheets as well, then rather than repeating the steps, you can simply **Apply the Filter** to all other relevant Worksheets :

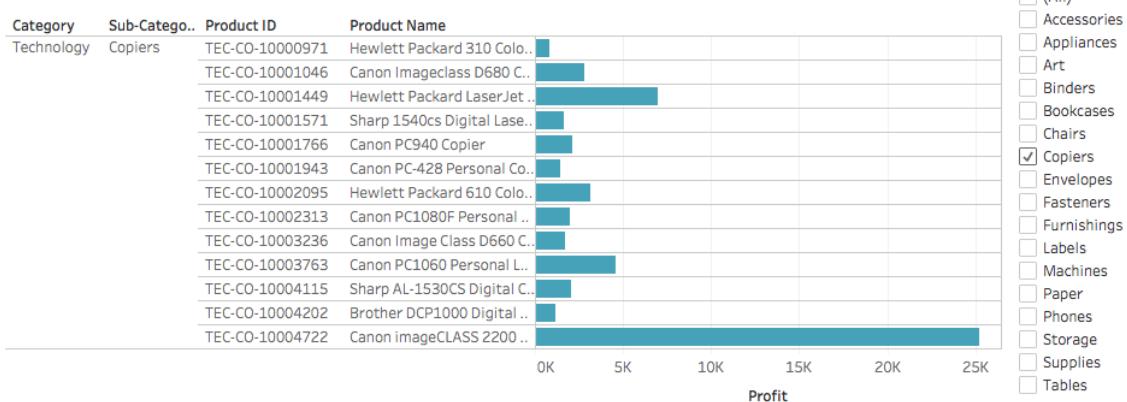


## 3.2 Drill Down and Drill Up

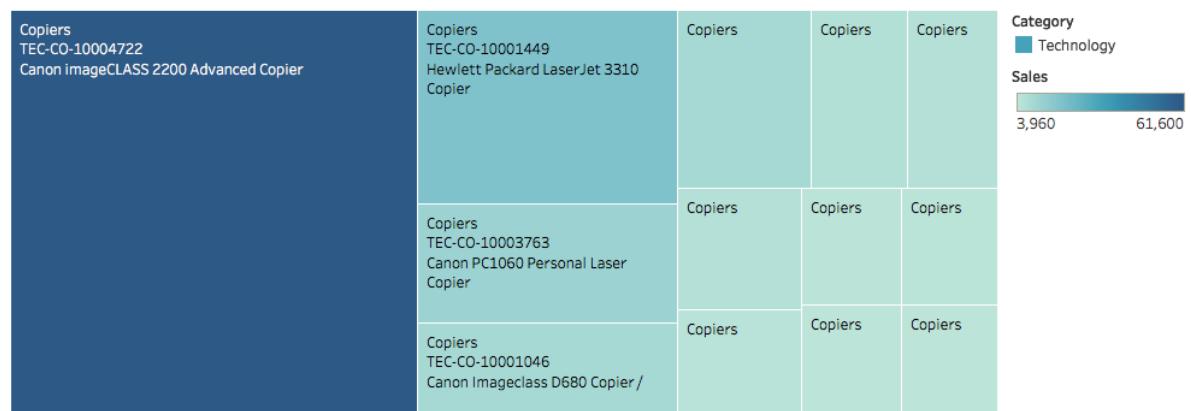
By now you must have gotten some picture of the way our Data is built. We have Category as the main Field, divided into Sub – Category, which is further distinguished into the various Product IDs and their corresponding Product Names.

This concept of breaking down our data to reach the absolute depth is called **Drilling Down** :

### ProductDrillDown



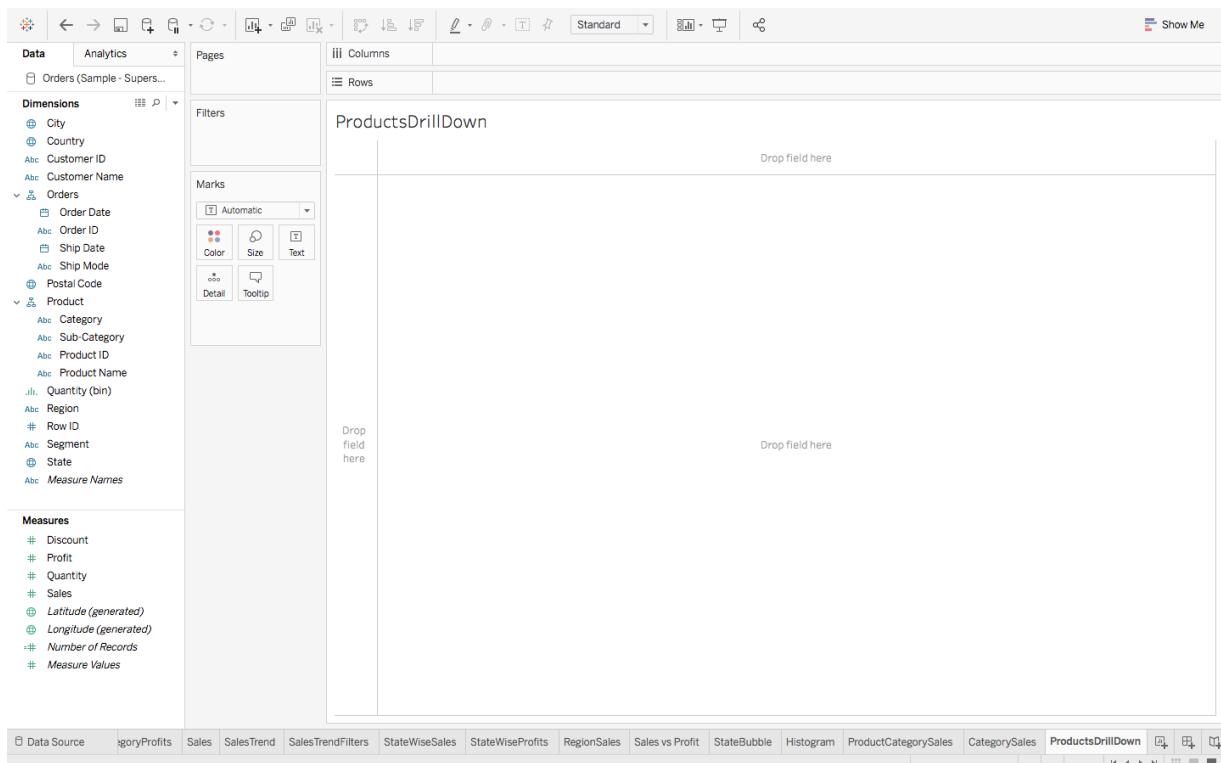
### Tree Analysis of Product Sales



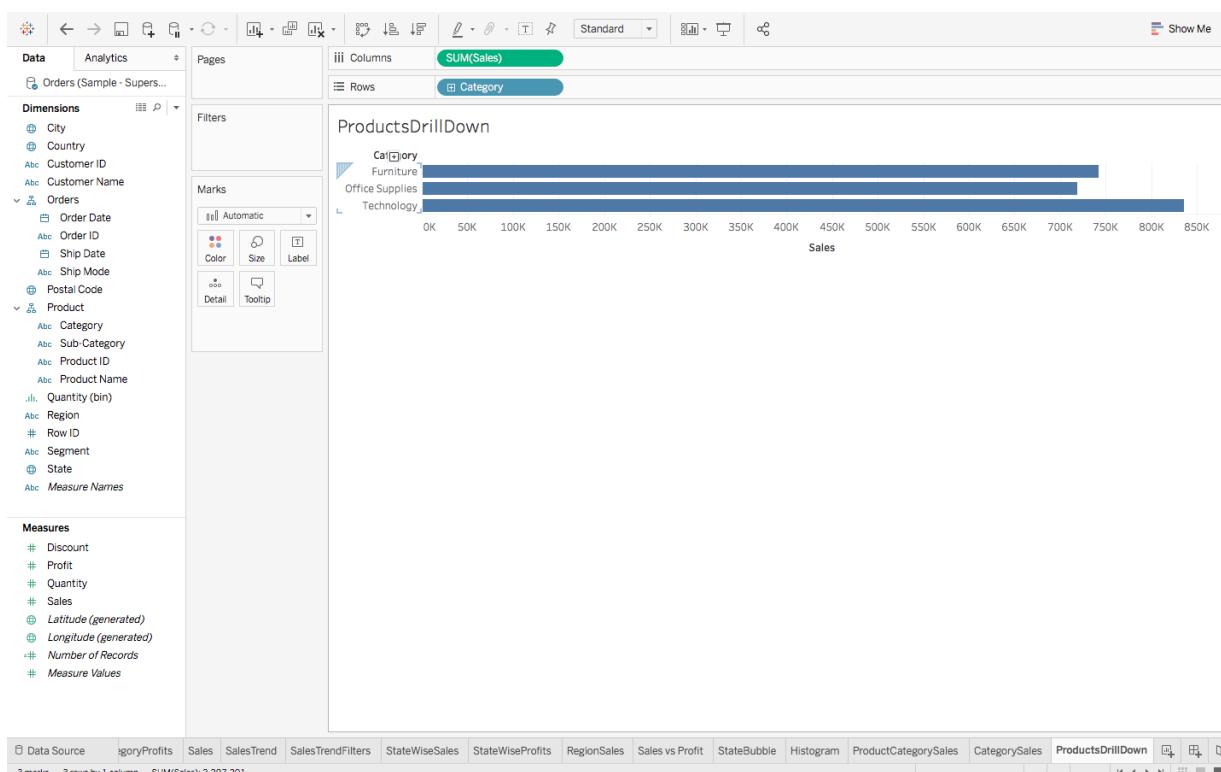
Similarly you can drill down from **Order Date** to Order ID to Ship Date to **Ship Mode**. This is also referred to as making an **Hierarchy** of data.

Let's consider the **ProductDrillDown** first, which is really a **Bar Graph** :

1. First you need to group the Dimensions you want in a single Hierarchy. So, drag **Sub – Category** from Dimensions **on top of Category** in the Dimensions itself, and change the Name of the hierarchy to **Product**.
2. Now drag **Product ID** and **Product Name** over this **Product Hierarchy**
3. Do the same for Order Hierarchy to get :



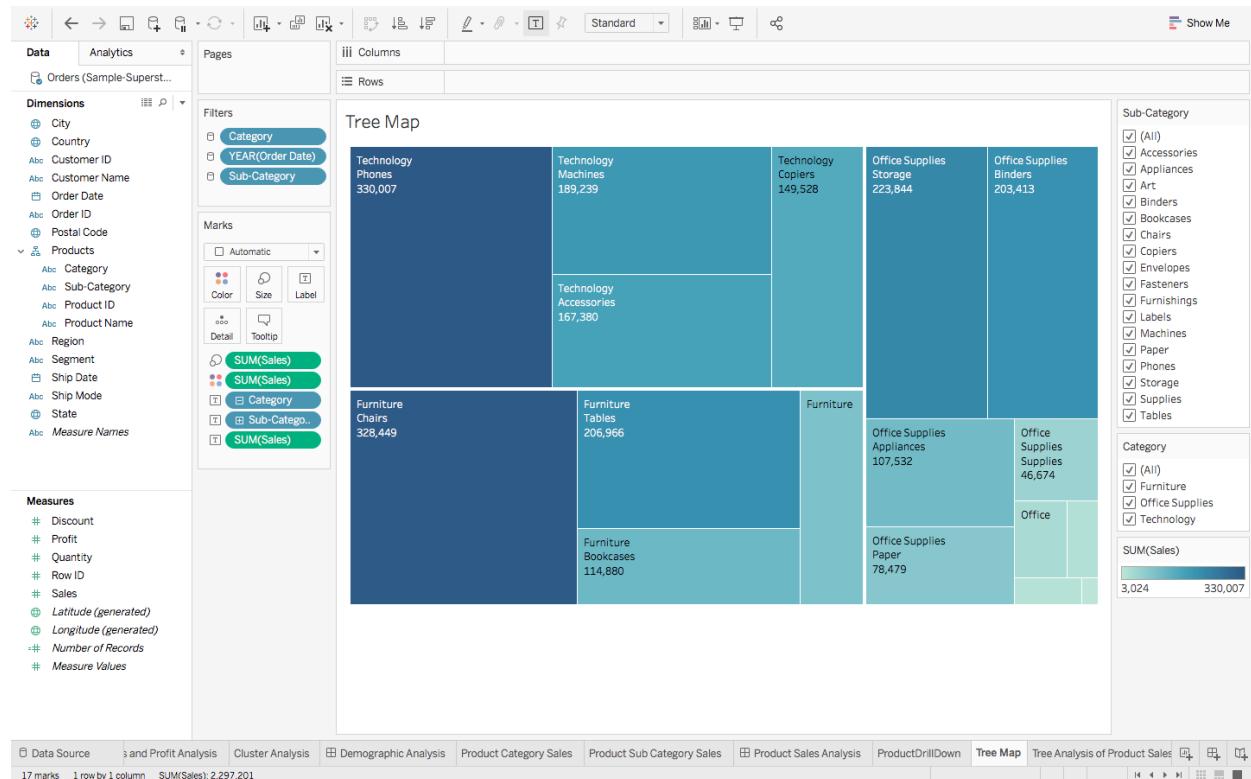
4. To finally plot your data, drag the **Product Hierarchy** onto Rows and **Sales** onto Columns, and get:



This was just a simple Bar Graph, but if you hover over the **Category** axis, you will see a small plus sign. Click on it to get a granulated version of your data. Do the same for the other generated axis as well to get to the absolute depth.

The **Tree Analysis of Product Sales** is a **Tree Map**, which is a great way of representing Drilled Down data, and is quite easy to make :

5. Following the drill down from **Step 4**, simply go to Show Me and select the **Tree Map** chart, to get the following :

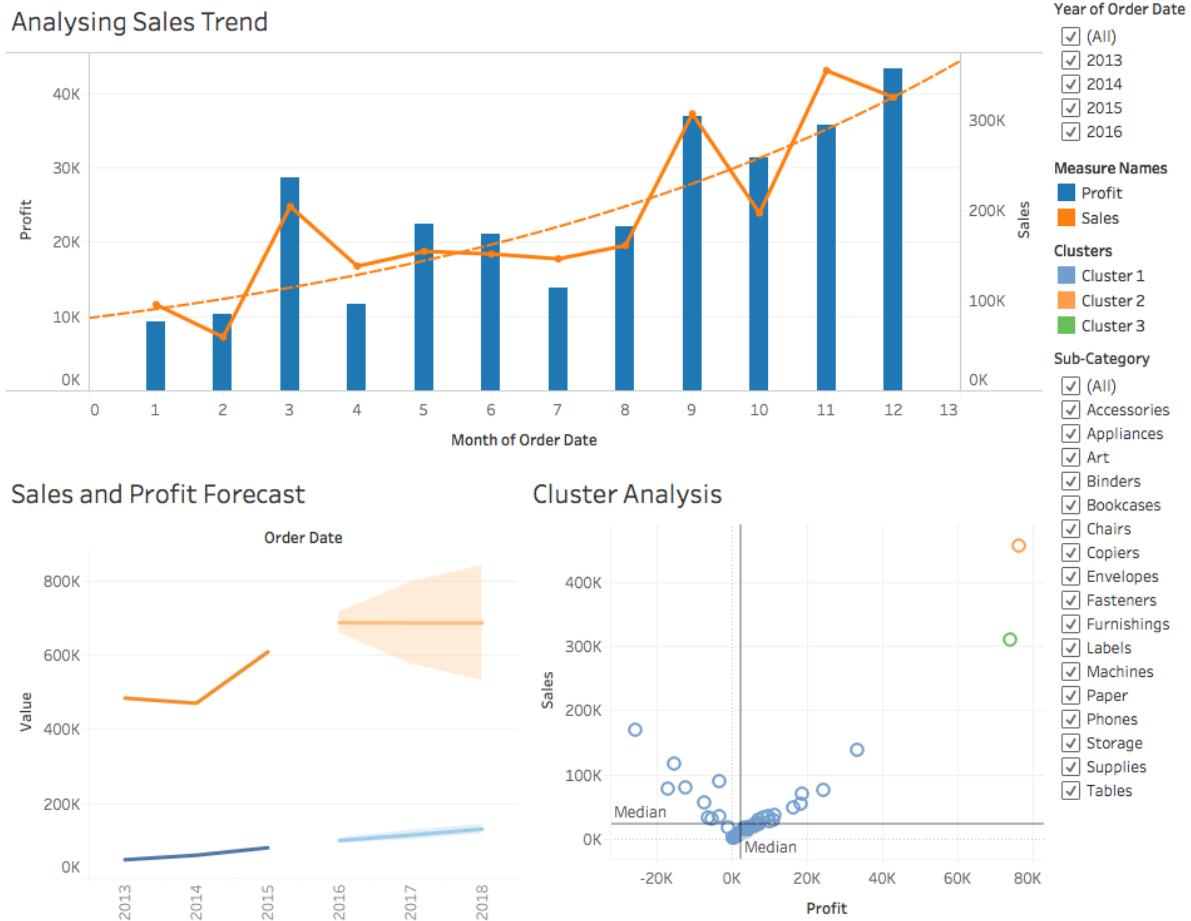


So far you have analysed the present scenario, but for expansion consideration, let's try and analyse the future too.

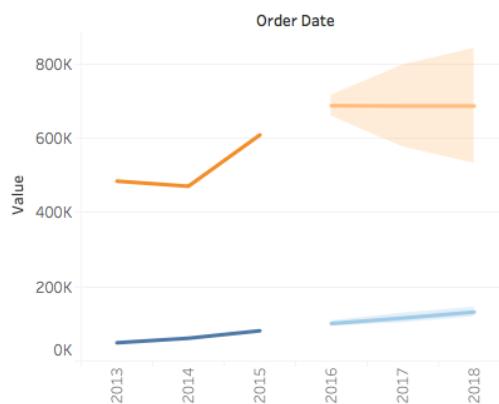
With the following Dashboard, you can not only see the **Trends over the Sales Months**, but also a **Forecast over the Years** too. And both of them tell a different story altogether :

## Trends and Forecasts

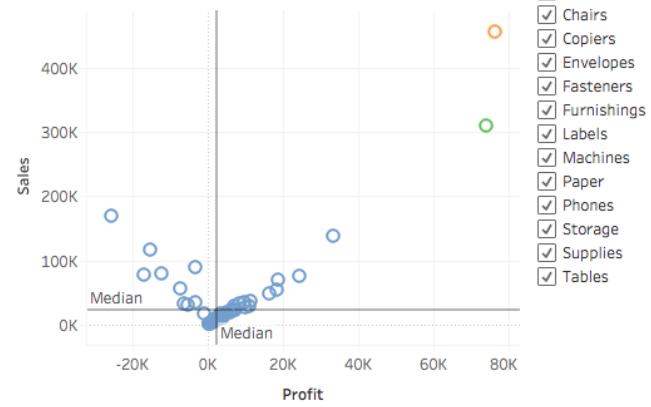
### Analysing Sales Trend



### Sales and Profit Forecast



### Cluster Analysis

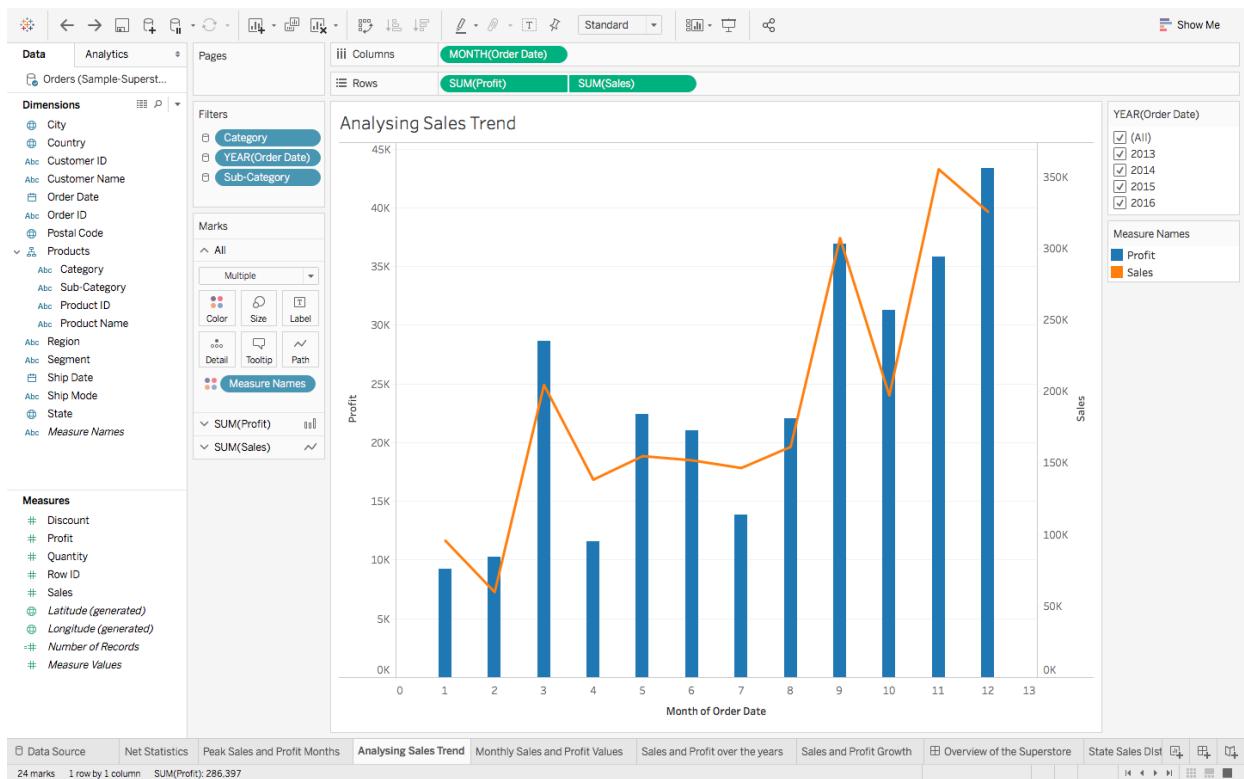


Although the Sales of the Superstore are increasing over the months of a Year, the future in general looks a bit bleak. The sales seem to become constant for the next 3 years, but fortunately for the Superstore, the Profit is increasing steadily. Let's get to making the above now.

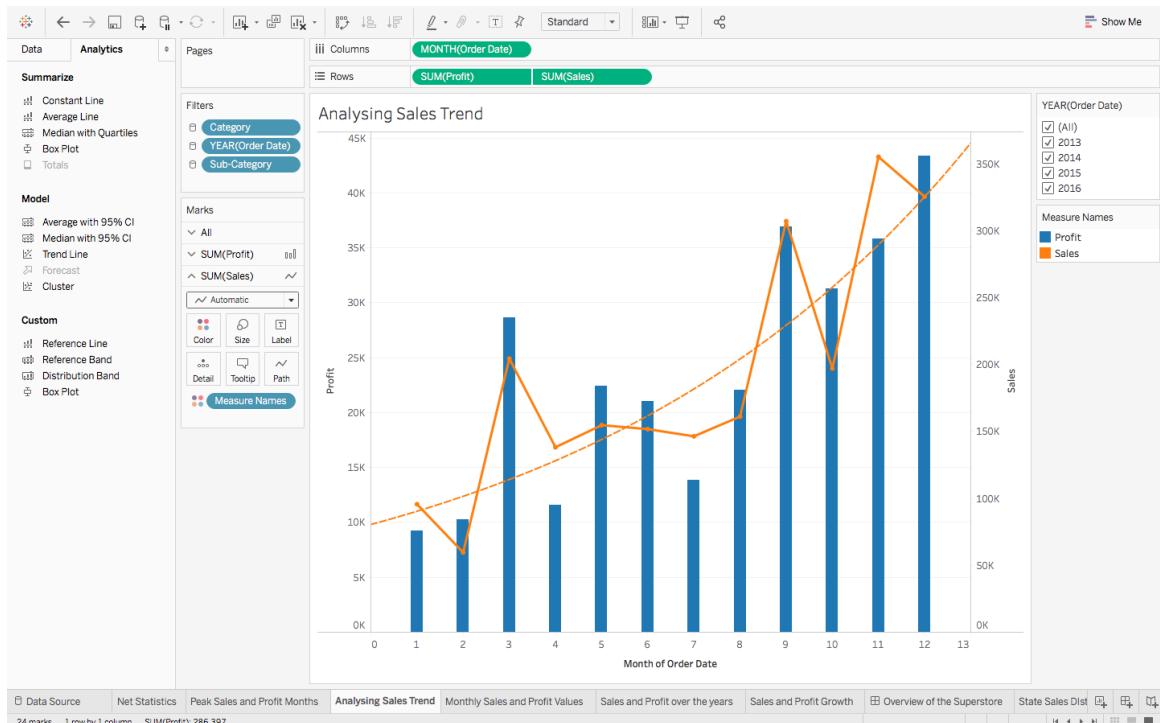
## 3.3 Trend Line

Traverse back to the **Peak Sales and Profit Month Chart** and follow these steps to make a Trend Line of your own :

1. Go to **Show Me** and choose the **Dual Combination** chart, to get this chart :



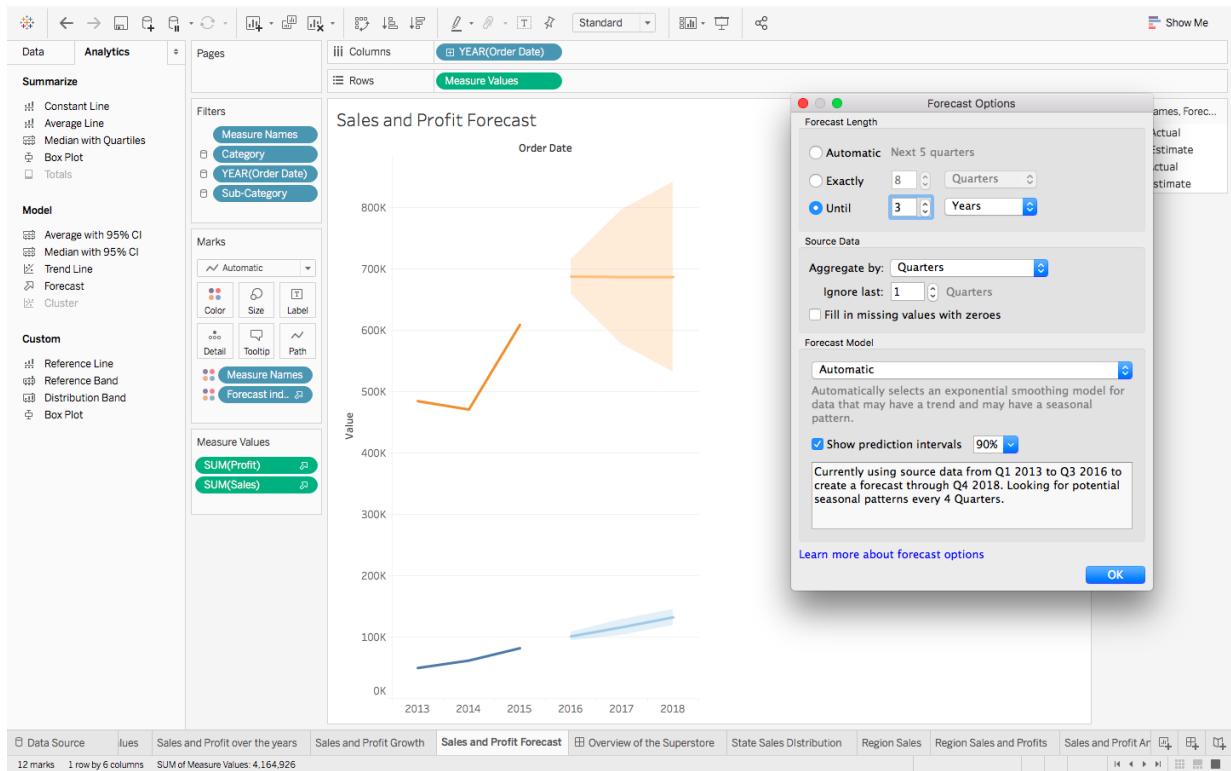
2. To get the Trend Line, go to **Analytics**, and simply drag **Trend Line** over the chart, to get :



## 3.4 Forecasting

For **forecasting**, we are going to deal with the **Sales and Profit Growth** chart. The construction is similar to that of Trend Lines, but with a small change. The steps are :

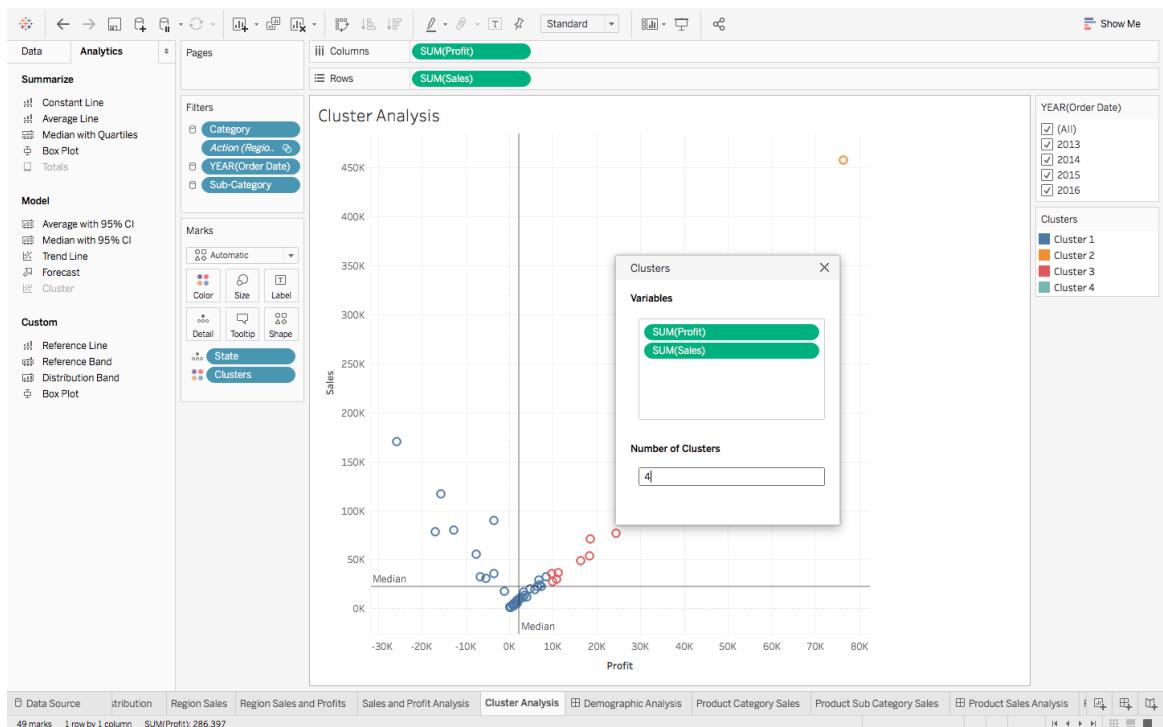
1. Drag **Forecast** over the chart.
2. You can also change the time frame of the Forecast, by right clicking on the Forecast Area and opting for **Forecast Options**, after which you can make your customisations :



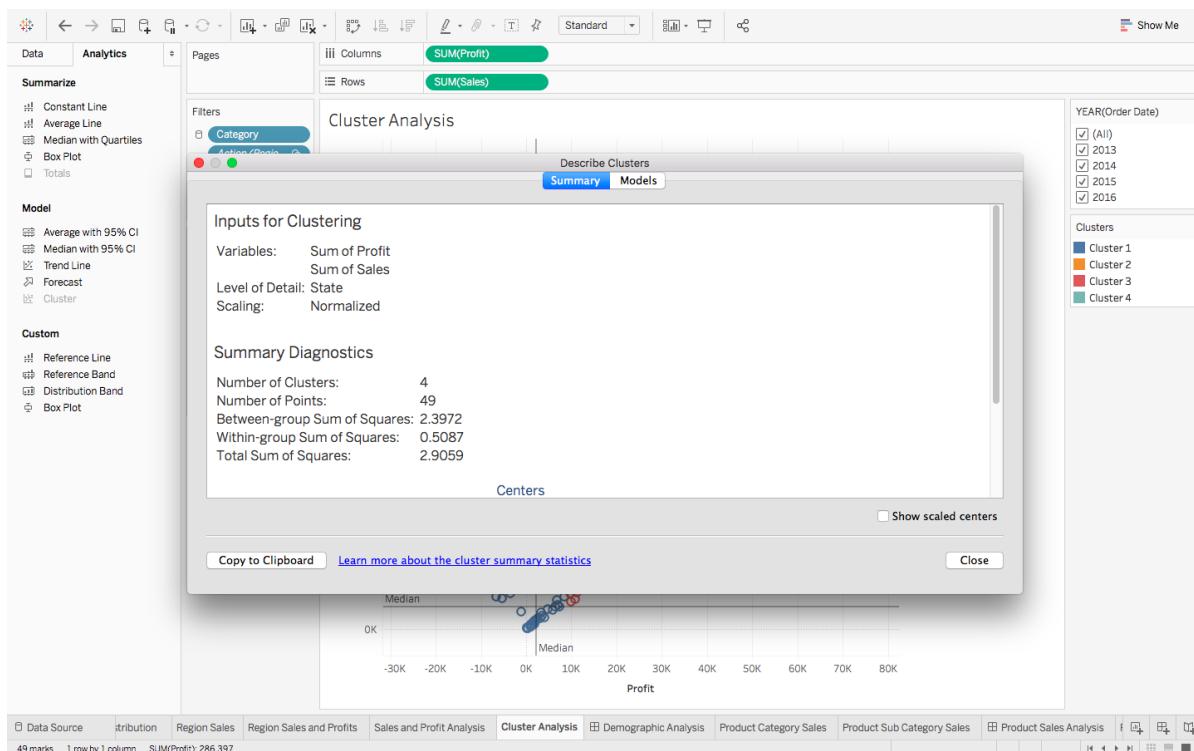
## 3.5 Clusters

Let's head back to the **Sales and Profit Analysis** chart that we had made. Remember the detailed inference that we had generated from it? We are just going to make that a bit more prominent now, using **Clusters**. To make them :

1. Go to **Analytics** and choose **Clusters**.
2. You can format the Cluster formation as per your wishes. Here we are clustering based on the **Sum of Sales and Profit**, choosing the **number of clusters to be 4** :



3. To view the Cluster information, right click on **Clusters** in the Marks Pane, and select **Describe Clusters**, to get this pop up :



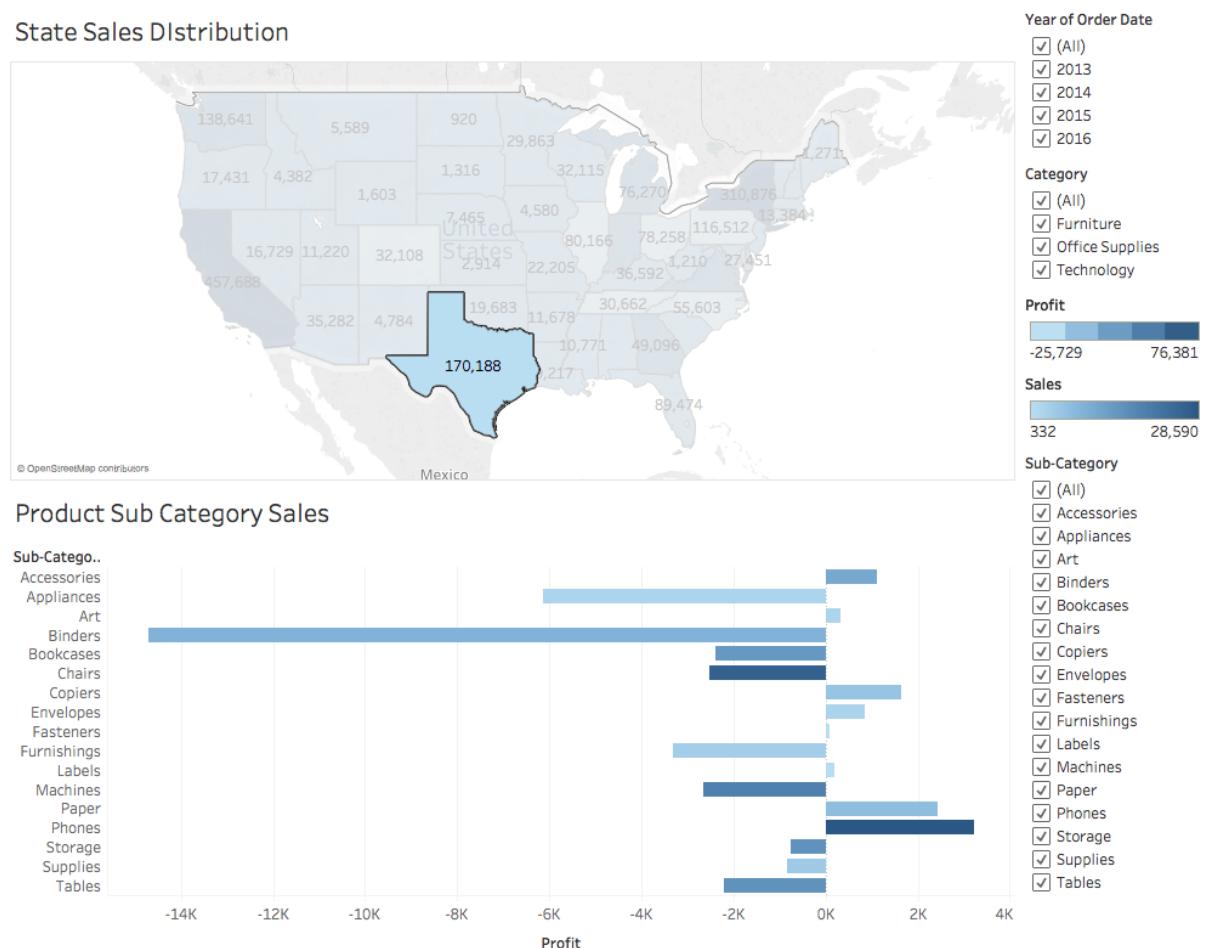
## 4. Dashboard

I am sure by now you must have gotten a pretty good idea of what a Dashboard is, having seen it plenty of times all throughout this article.

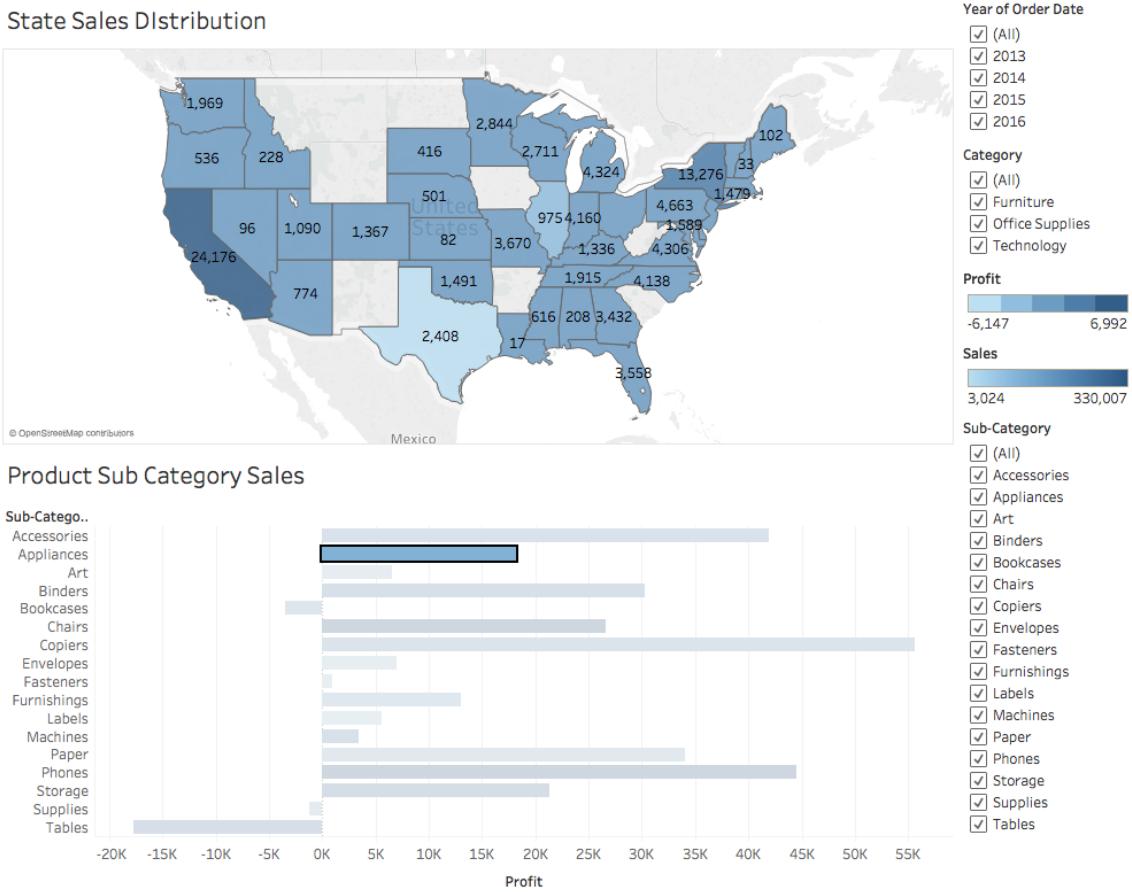
If not, well then a Dashboard is simply a means of combining Worksheets together so that they convey some message. Without much further ado, let's get right to it!

Consider the **State Sales Distribution** Map chart and **Product Sub Categories**.

What if you wanted to know the various Sales margin of each Product within separate States? We had observed that Texas was one of the States with the lowest Profits. By looking at the following Dashboard, you will see that the reason is it's not managing to generate Profits in majority of the Products :



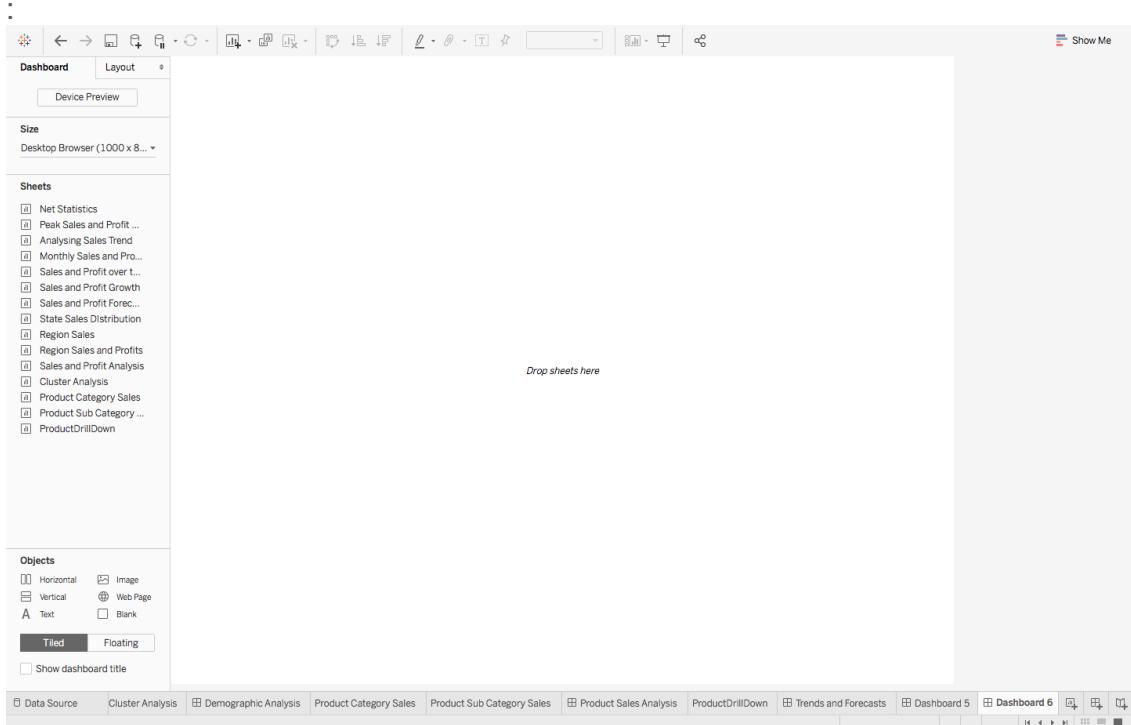
Now consider the state wise Sales distribution of a Sub – Category :



The above beautifully shows the distribution of **Appliances** over the country, where California seems to be the major Profit contributor.

Making such a Dashboard is actually quite easy. Let's see how :

1. This time instead of creating a New Worksheet, we are going to create a new Dashboard. Click on the **window like icon** next to the 'New Worksheet' icon in the bottom pane to get the following

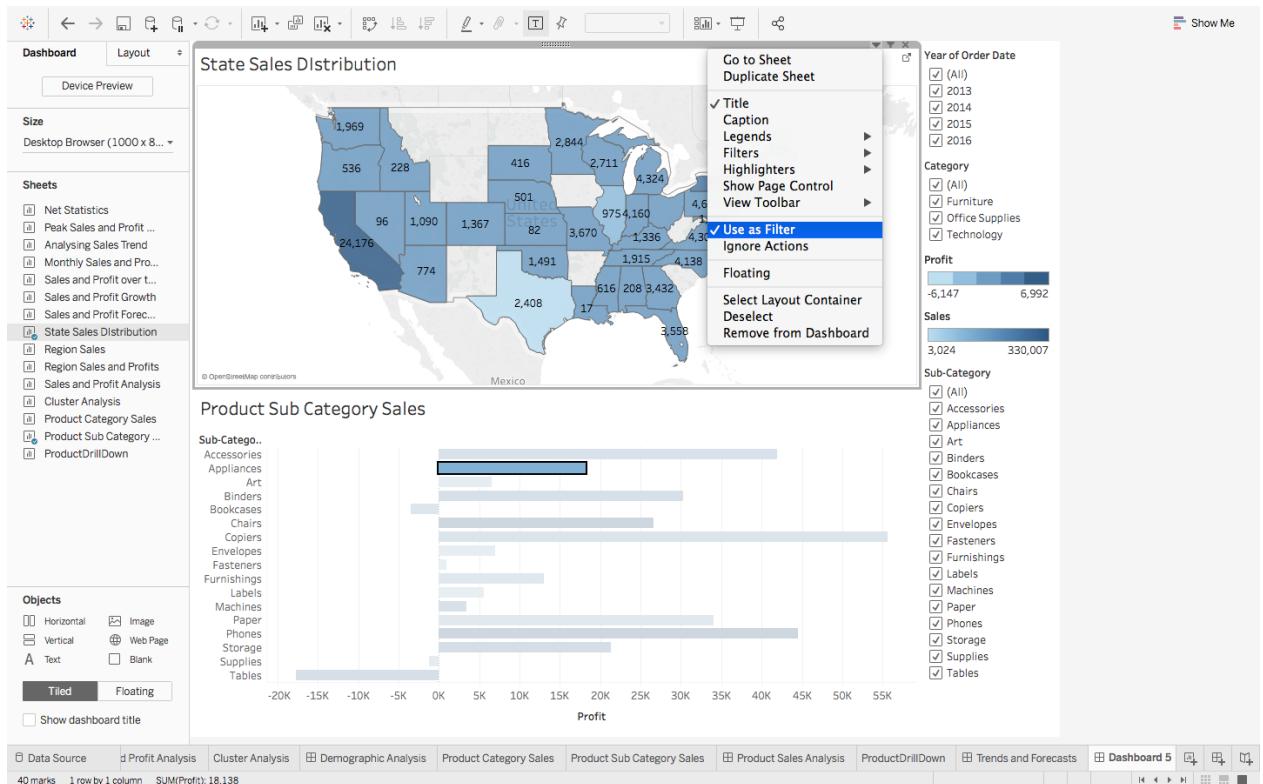


2. See the multiple Worksheets that we had made till now over on the left? All that we have to do to make a Dashboard is drag these sheets from the pane to the empty area '**Drop sheets here**'.
3. So to make the previously displayed Dashboard, simply drag State Sales Distribution and Product Sub Category Sales. The Dashboard will automatically make space available for both of them.

**Note :** Even after the creation of the Dashboards, you can still edit your Worksheets, and the same changes shall be reflected here.

If you were to click on the States or the Products after creating your first ever Dashboard, you won't observe any change. Because for such visuals, we first have to convert the Charts themselves into filters.

4. Simply click on the small Down Arrow on each chart you wish to turn into a Filter, and select **Use as Filter**:

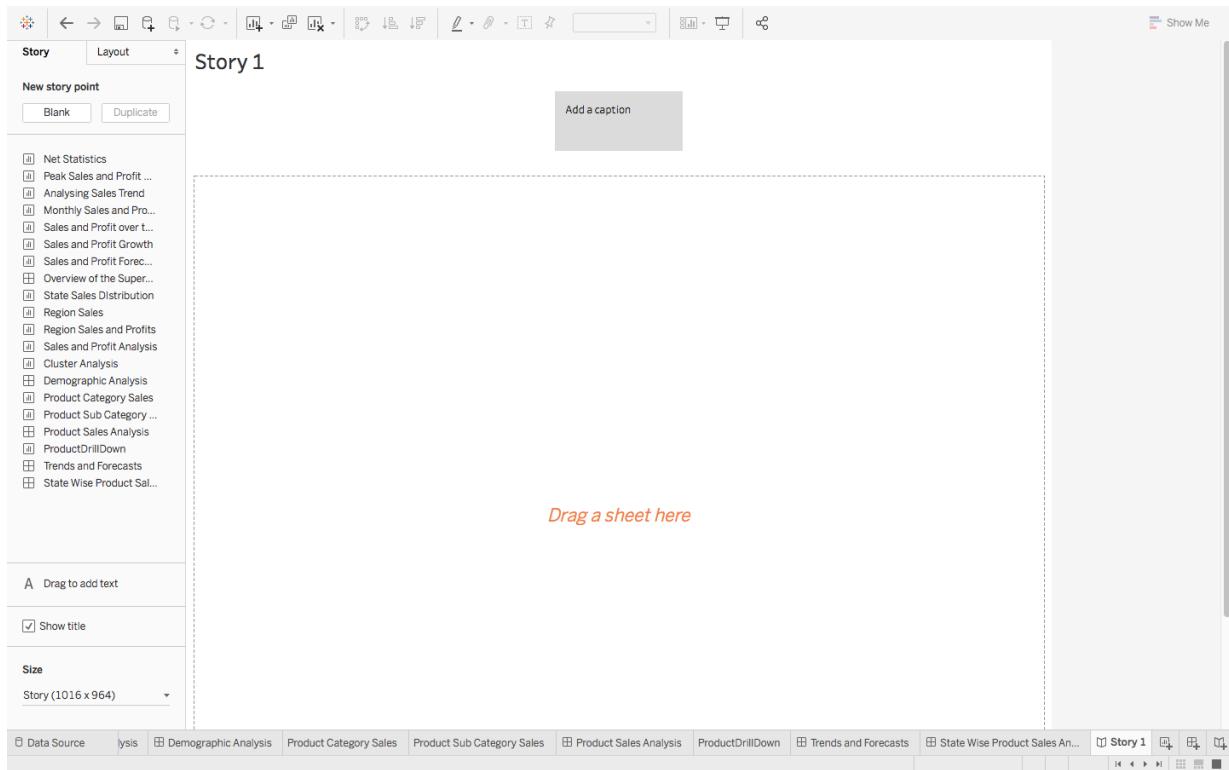


**Note :** While making Dashboards, it is preferred to **use your charts** as Filters, rather than cluttering up the view with custom ones.

## 5. Story – Bringing it all together

Just like Dashboards were a way to combine the Worksheets, a Story is where you combine all the dashboards, and if need be individual Sheets as well, to convey, as the name suggests – a Story.

1. Just like before, you simply drag your Worksheets and Dashboards onto the empty space :



So let's combine all those Dashboards that we had made into what could perhaps make a decent presentation for a beginner. Do ensure to **Add a Caption** to all of your Dashboards, to convey your message clearly :

## Overview of the Superstore

### Net Statistics

Profit	Quantity	Sales
286,397	37,873	2,297,201

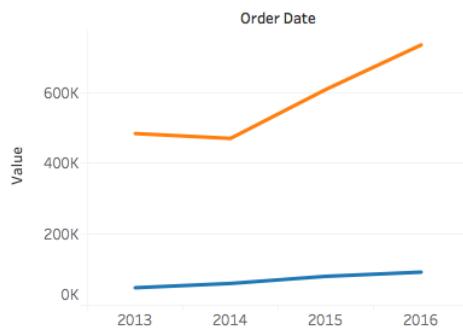
Year of Order Date

- (All)
- 2013
- 2014
- 2015
- 2016

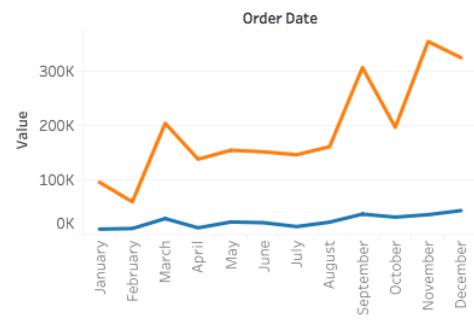
Measure Names

- Profit
- Sales

### Sales and Profit Growth



### Peak Sales and Profit Months

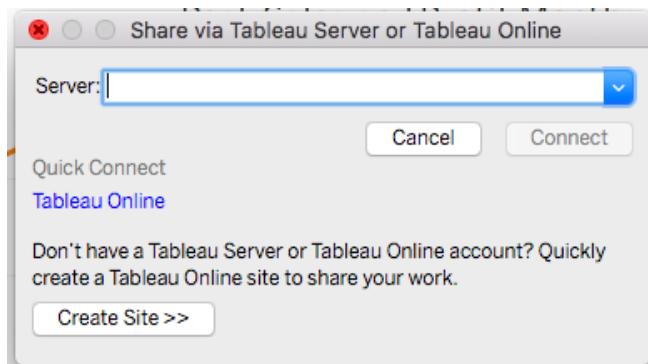


Year of Order Date	Sales
2013	484,247.4981
2014	470,532.509
2015	608,473.83
2016	733,947.0232
	93,508

Month of Order Date	Sales
January	95,656.6036
February	59,640.1474
March	204,475.8308
April	138,260.6826
May	154,732.1607
June	151,804.5883
July	146,416.889
August	161,194.253
September	307,148.0297
October	197,098.3697
November	31,304
December	25,822

If you have ever come across Tableau Stories online, the ones which you could actually interact with, instead of just viewing, that is made possible by publishing your Workbooks onto the [Tableau Server](#).

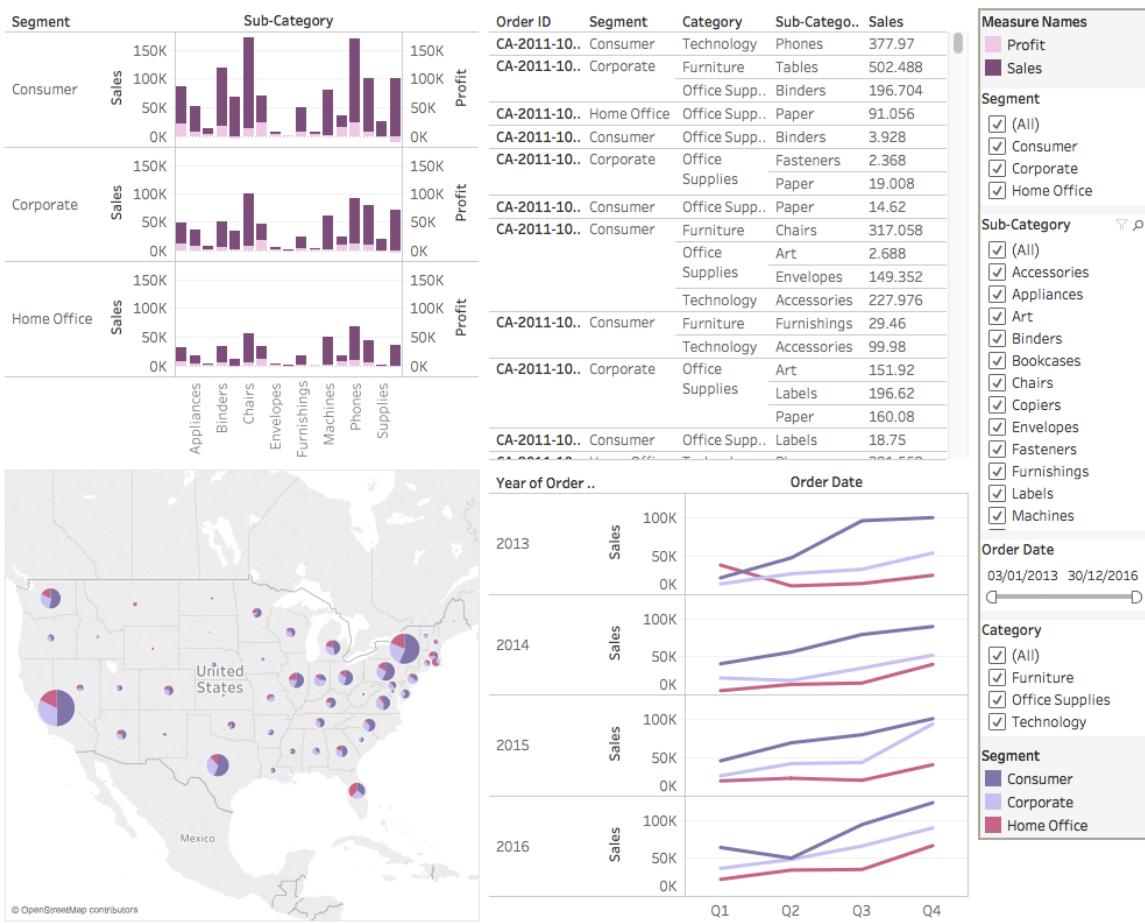
If you have one set up, then all you need to do, after creating your Stories, is go to **Server -> Publish Workbook** and enter the Server Name :



## 6. End Notes of “Tableau for Beginners – Data Visualisation made easy”

What we have covered so far is pretty much the basics of Tableau. As it is said ‘With practice, comes perfection’, it is suggested that you experiment as much as you can with Tableau.

Below is a sample Dashboard that I would encourage everyone of you to try and make. You will not only get to test the skills that you have learned so far, but also hopefully acquire more. The dataset used is the same as the one we had been working with so far



If there are ever any doubts, do leave them as comments

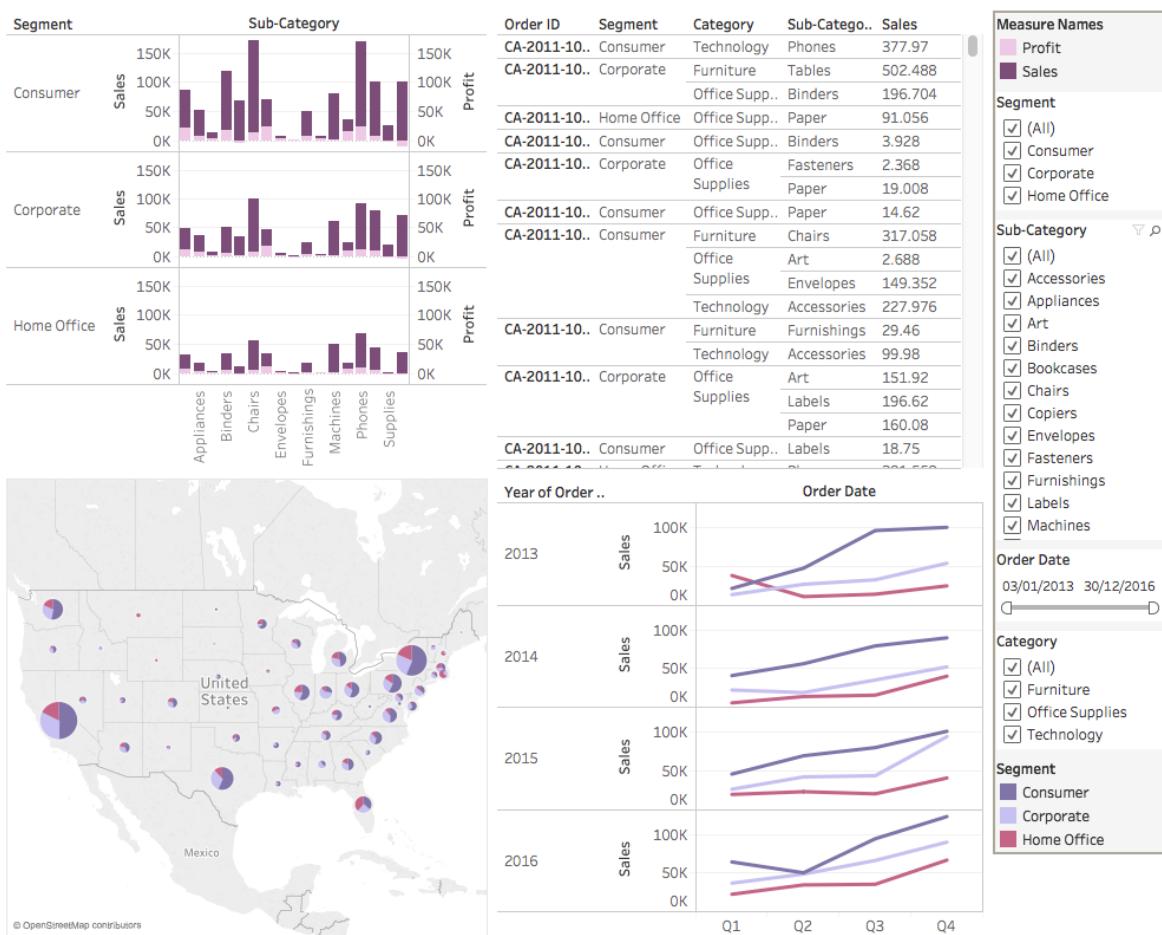
All the best on your journey as a Data Explorer, and stay tuned for my next article on Tableau!

# Intermediate Tableau guide for data science and business intelligence professionals

## Introduction

Let's assume that you have some data with you and you wish to garner some insights from it. Coding is not your forte and you don't know how to get started.

Let me tell you this – you can make something as descriptive / insightful as the image below, with gestures as simple as drag and drop. And it doesn't even require a single ounce of coding. **Now that is the power of Tableau for you!**



For all those reading this who have been acquainted with Tableau, can plot a few basic charts on it, and wish to learn more about its wide horizons, **this** article is meant for you.

As for those who have yet to be introduced to the beauty and simplicity of Tableau. Practice making a few simple visualisations and then rush back here!

In this article we are going to discuss a few core functionalities of Tableau which help in making really dynamic graphs. So let's quickly get started!

# Table of Contents

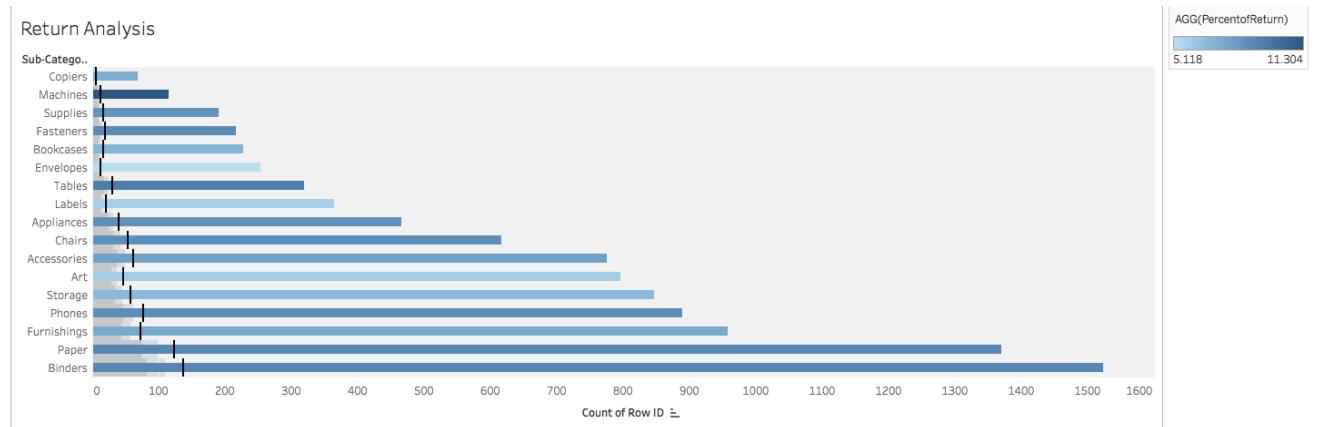
1. Dealing with Different Data Sources
  1. Joins
  2. Data Blending
2. Conditional Combination of Data
  1. Groups
  2. Sets
3. Calculations and Level of Detail
4. Parameter Control
5. End Notes

## 1. Dealing with different data sources

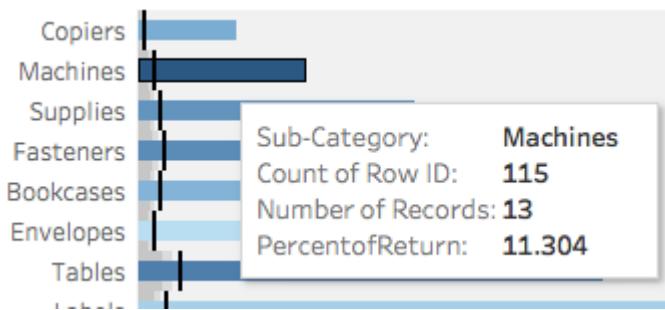
### 1.1 Joins

It's not practical to store all data in a single table. In order to avoid anomalies related to updates, data is almost always distributed in multiple tables that have some relation with each other. Let's understand the same with an example.

Consider the situation where a Superstore, on the verge of expansion, perceives that the number of Returned orders has been increasing by day. To ascertain the analysis and come up with the right action plan, they plotted the following chart to understand the products that were being returned :



As can be observed, **Binders** have the maximum number of items being returned. But judging by the color of the bars, **Machines** and **Tables** have the highest percent of return ( returned / bought ) :



This seems pretty similar to plotting just another chart, but the trick here was that it was created by using the combined data of two tables : **Orders** and **Returns**. This is known as a **Join**.

Let's try making the same to get a better understanding :

1. Go ahead and download the Sample Superstore Sales data from [here](#).
2. Open Tableau and import the same :

The dataset is made up of 3 tables : Orders, People and Returns, and the ones that we are interested in at the moment are **Orders** and **Returns**.

3. Drag **Orders** onto the empty area and use the Data Interpreter to get rid of the erroneous data.
4. Drag **Returns** in the same area to get something like this :

The screenshot shows the Tableau Data Source interface. On the left, under 'Connections', there is one entry: 'Sample-Superstore' (Excel). Under 'Sheets', several tables are listed: 'Orders', 'People', 'Returns', 'Orders A1:B3', and 'New Union'. A note indicates that 'Cleaned with Data Interpreter' has removed some data, with a link to 'Review the results'. Below this, a preview of the 'Orders' table is shown with 18 rows of data. The columns are: Row ID, Order ID, Order Date, Ship Date, Ship Mode, and Custom. The first few rows of data are:

#	Abc Orders	Abc Orders	Abc Orders	Abc Orders	Abc Orders
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Custom
5,737	CA-2011-148614	20/01/2013	25/01/2013	Standard Class	MV-174
5,738	CA-2011-148614	20/01/2013	25/01/2013	Standard Class	MV-174
6,530	CA-2011-103744	23/02/2013	27/02/2013	Standard Class	MG-178
6,531	CA-2011-103744	23/02/2013	27/02/2013	Standard Class	MG-178
7,428	US-2011-164763	17/03/2013	21/03/2013	Standard Class	MH-174
7,429	US-2011-164763	17/03/2013	21/03/2013	Standard Class	MH-174
7,430	US-2011-164763	17/03/2013	21/03/2013	Standard Class	MH-174
7,431	US-2011-164763	17/03/2013	21/03/2013	Standard Class	MH-174
7,432	US-2011-164763	17/03/2013	21/03/2013	Standard Class	MH-174
6,152	CA-2011-111871	18/03/2013	21/03/2013	Second Class	EK-137
6,016	CA-2011-102652	06/04/2013	12/04/2013	Standard Class	AY-105
6,017	CA-2011-102652	06/04/2013	12/04/2013	Standard Class	AY-105
6,018	CA-2011-102652	06/04/2013	12/04/2013	Standard Class	AY-105

At the bottom, there are buttons for 'Data Source' (disabled), 'Sheet 1' (highlighted in orange), and other sheet options.

For two tables to be joined, there has to be the presence of at least one common field. Here, Tableau automatically **Inner Joined** the two tables, based on the commonality of the column **Order ID**. By way of inner join, the combined data only consists of those rows that have the same Order ID in both the tables.

You can change the **Join Type** as well as the **Joining Field** in Tableau, but you need to ensure that it's sensible.

Changing the Join type (Default Inner to Right) :

The screenshot shows the Tableau Data Source Join dialog. At the top, there are four join types: Inner, Left, Right, and Full Outer. The 'Inner' option is selected. Below the join type is a table with two columns: 'Data Source' and 'Returns'. The 'Data Source' column contains 'Order ID' and the 'Returns' column contains 'Order ID (Retu...)' with an equals sign between them. A link text 'Add new join cla...' is also present. Below this table is a preview pane showing the joined data with columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, and Order ID (Retu...). The 'Ship Date' and 'Ship Mode' columns are partially cut off.

Changing the Join field :

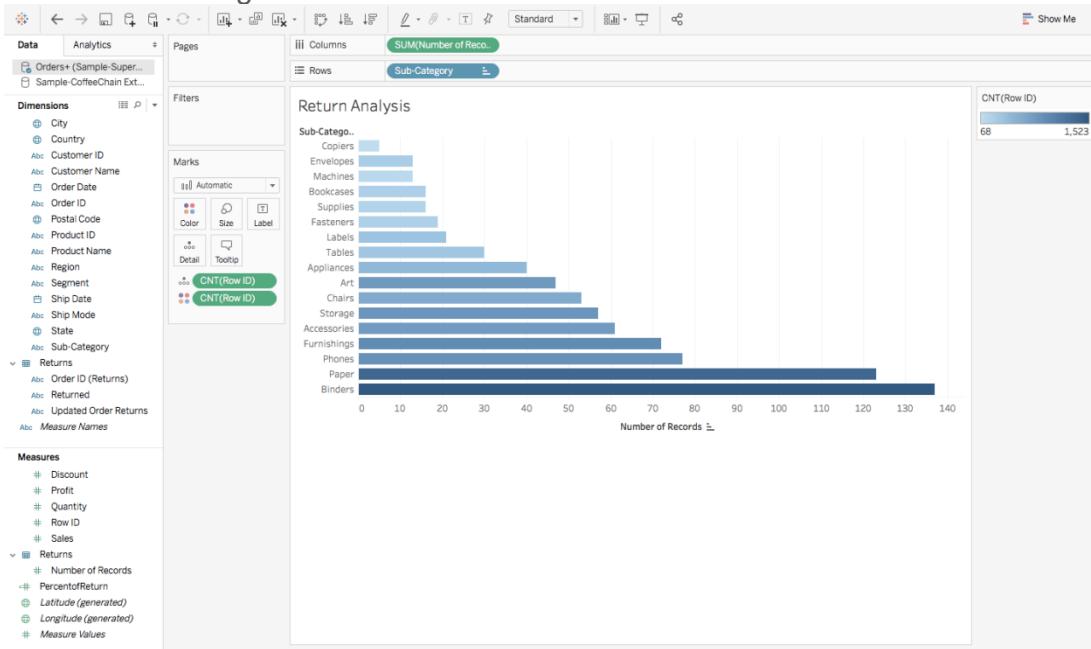
This screenshot is similar to the previous one but shows a 'Left' join instead of 'Inner'. The 'Row ID' field from the 'Orders' table is joined with the 'Order ID' field from the 'Returns' table. The preview pane now includes additional columns: Category, Sub-Category, Product Name, Sales, Quantity, Discount, Profit, and Returned.

See how I tried to join the two tables based on **Row ID of Orders** and **Order ID of Returns**? Since the two are not compatible, we don't see any records, and plus that red mark near the circles shows an error.

Let's get back to working on the chart now :

5. Drag **Sub-Category** of table **Orders** onto **Rows** and **Number of Records** of **Returns** onto **Columns**.

6. Drag **Row ID** onto the Marks Section and convert its measure to **Count** instead of **Sum**. Now drag this onto the chart :



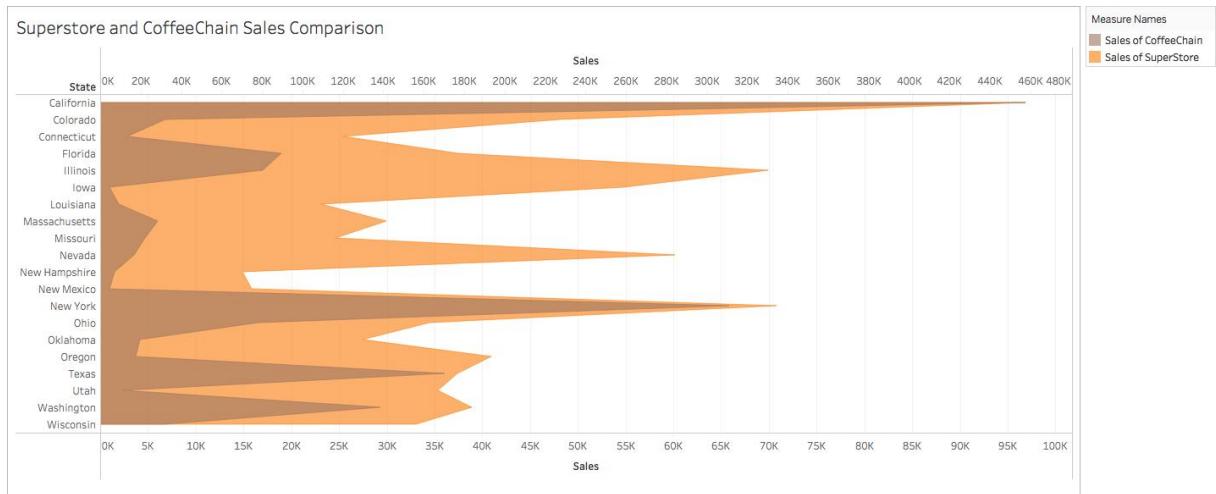
7. Finally go to **Show Me** and choose the **Bullet Graph** to get the desired chart.

Here we have used Inner Join, but you can always choose between Inner, Right, Left and Full Outer based on your requirements.

## 1.2 Data Blending

Data blending is quite similar to Joins, with the difference being that Joining requires the data to be from the same data source. In the above example, we used different tables from the same Excel file. But Data Blending comes into the picture when you are working with different Data Sources. Let's understand the same with an example.

The Superstore has another vertical in the form of a **Coffeechain** which is spread across as many states as the Superstore is. But they are considering shutting down some of the branches after observing the following plot :

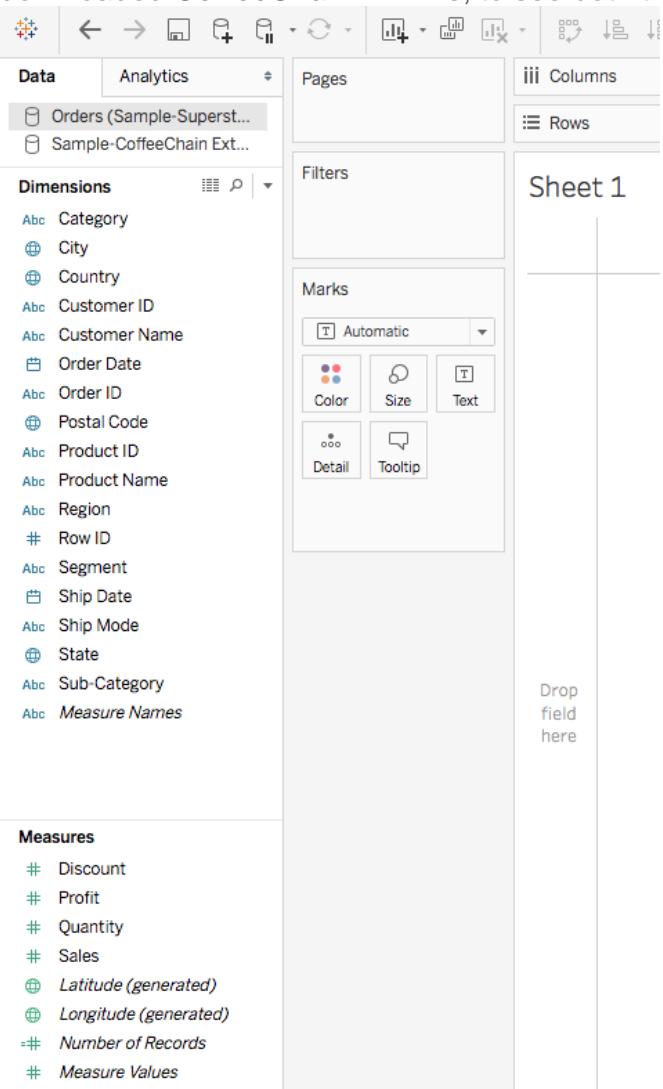


As can be seen there are some branches that are doing equally well as the Superstore such as **California** and **New York**, while many are not, such as **Iowa** and **New Mexico**. Just like in Joins, here the trick is that both the datasets, dealing with the two verticals, were stored in different Data Sources, an Excel file and a TDE database.

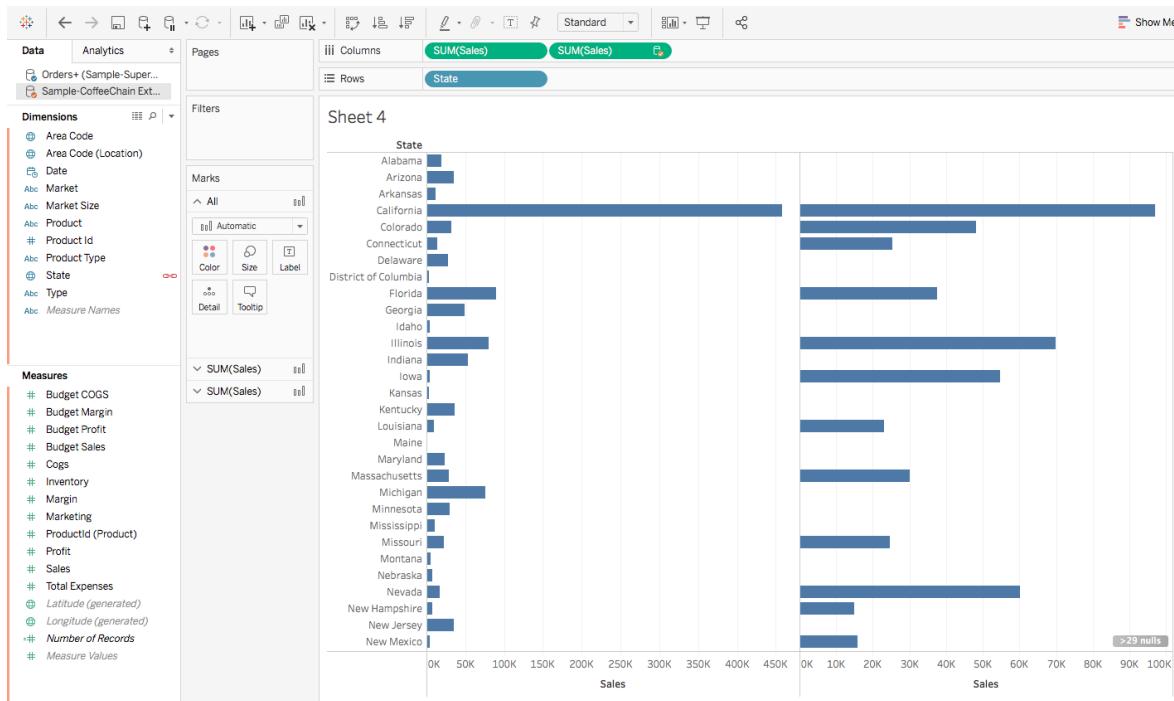
Why not we plot the same as well to get a better understanding? We will begin by blending the Superstore data and the Sample – CoffeeChain database. You can find the data for the latter here as well:

1. First import the Sample-Superstore.xls

2. Go to the Worksheet, click on **Data -> New Data Source** and choose the downloaded **CoffeeChain TDEfile**, to see both the datasets imported :



3. Click on the **States** under the **Superstore** database and drag it onto **Rows** and drag **Sales** onto **Columns**.  
 4. Click on the **CoffeeChain** database and drag **Sales** onto **Columns** as well to get :



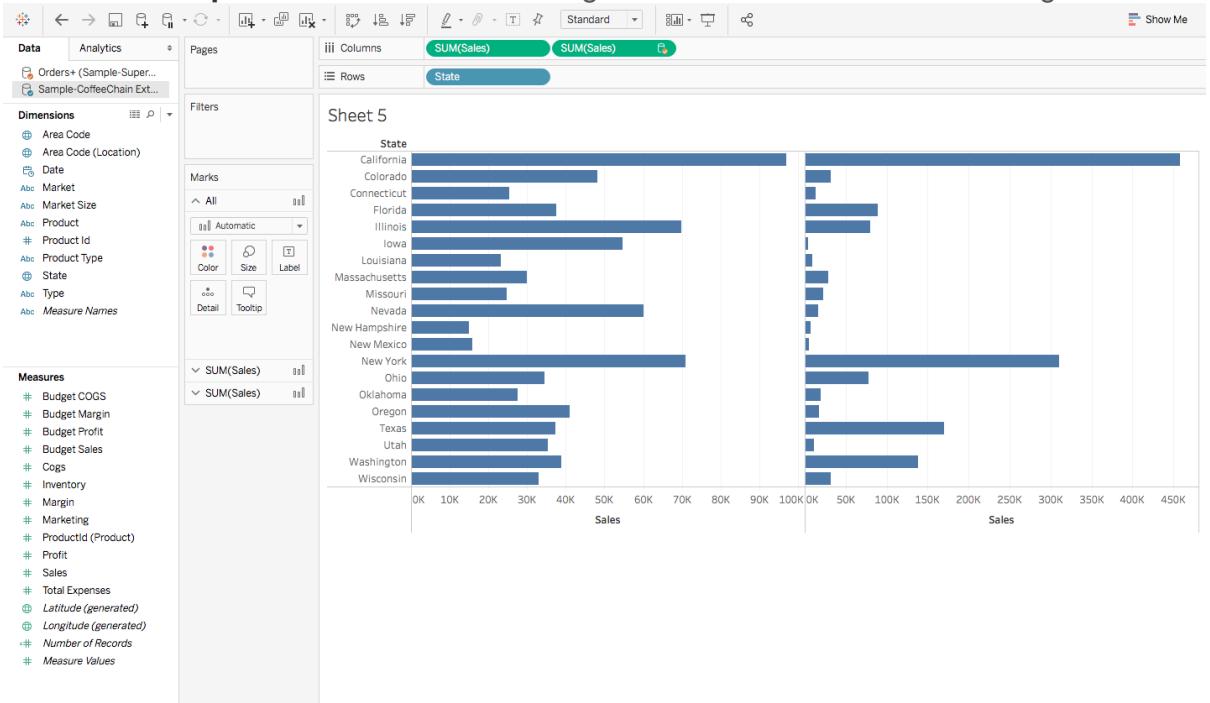
Here you must have observed a few things; let's take them up one by one :

- To the top-left of the screen, the Superstore database now has a **blue** mark on it – this indicates the **Primary Source**.
- The CoffeeChain data has an **orange** mark on it – this indicates the **Secondary Source**.
- There is an **orange chain** link next to **State** under CoffeeChain – This indicates the **Blended Field**, meaning that this field is common in both the datasets.
- There are '**>29 null**' values in the chart. This is because there are not as many states in the CoffeeChain dataset (Secondary Source) as there are in the Superstore dataset (Primary Source).

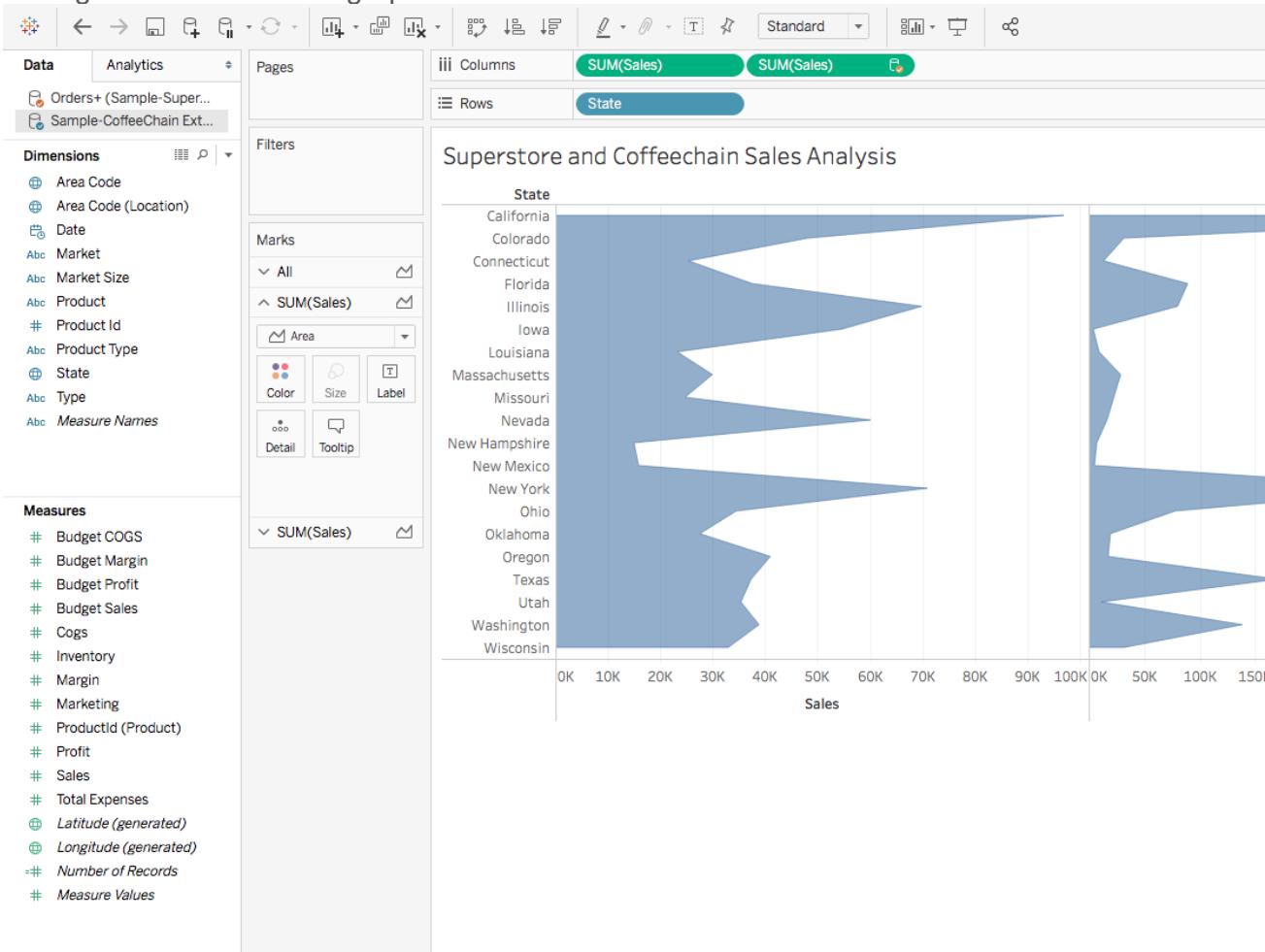
Let's rectify these null values by interchanging **Steps 3** and **4** :

- Click on the **States** under the **CoffeeChain** database and drag it onto **Rows** and drag **Sales** onto **Columns**.

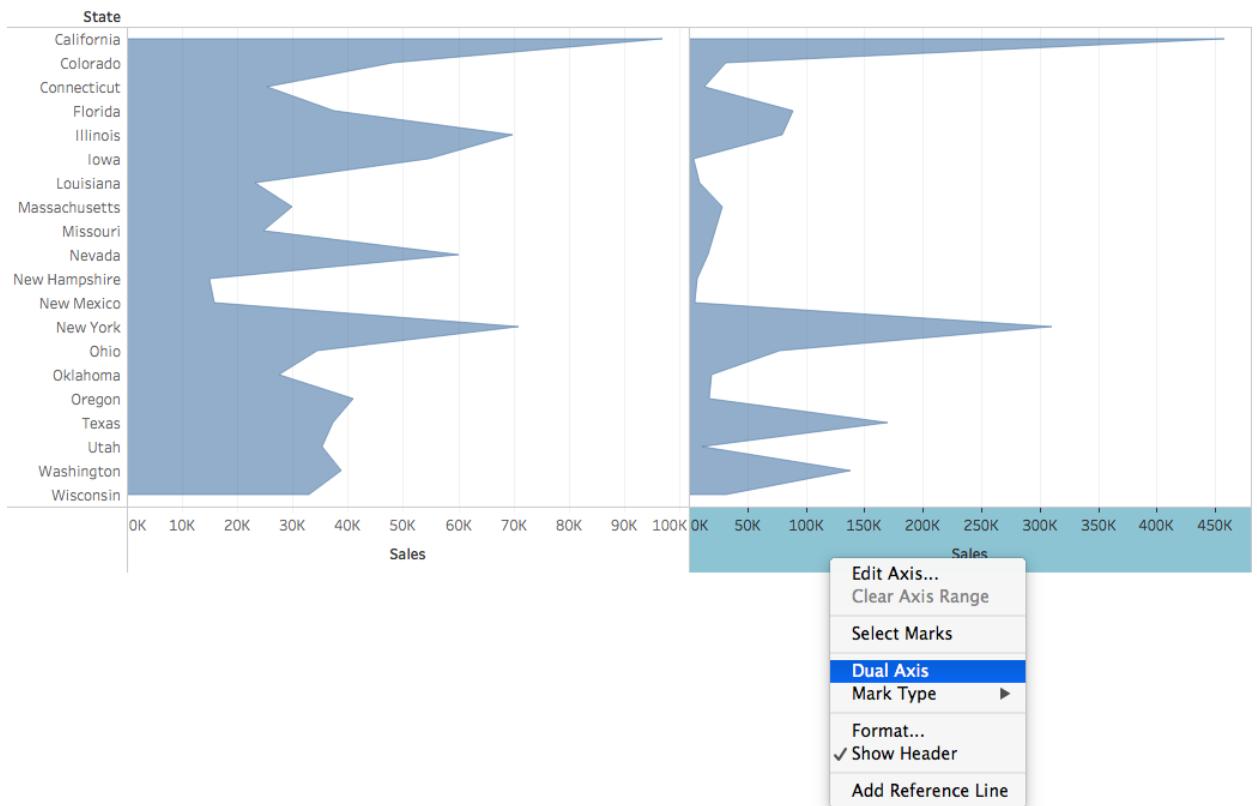
4. Click on the **Superstore** database and drag **Sales** onto **Columns** as well to get :



5. Change the chart of each graph from the **Marks section** to **Area chart** :



6. Right click on the second chart and choose **Dual Axis** to merge the two graphs into one :



7. Now all that is left is changing the color scheme and you just finished blending!

## 2. Conditional Combination of Data

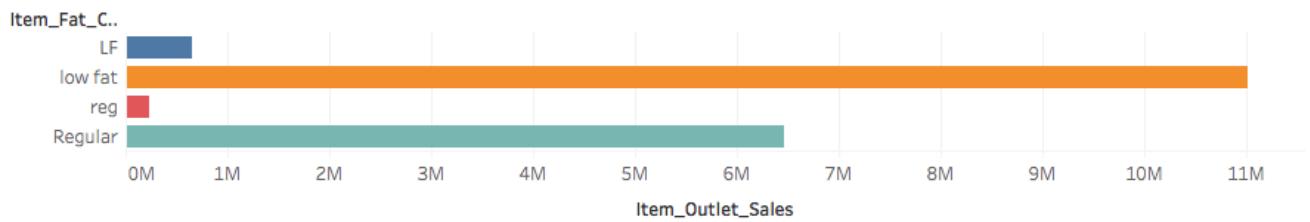
Now that your data is all ready, blended or joined, let's start making some interesting dashboards. From here on out, we will be using only the **Superstore data: Orders + Returns (Left-Joined)**.

### 2.1 Groups

Let's start off by considering the example of a Survey analysis. In a survey of Food Consumption, under the section of **Food Preferences**, instead of 'Low Fat', you may have 'LF', or instead of 'Regular', you may have 'reg'.

In such cases, during data visualisation, you face issues like the following:

## Grouping

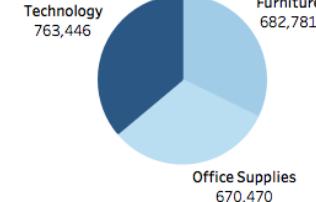
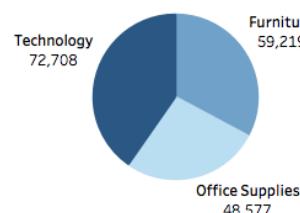
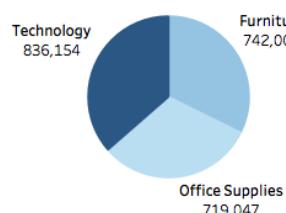
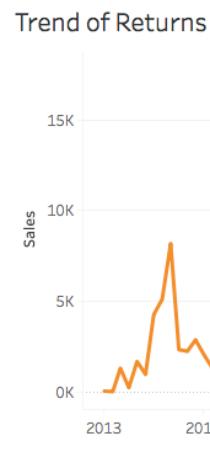
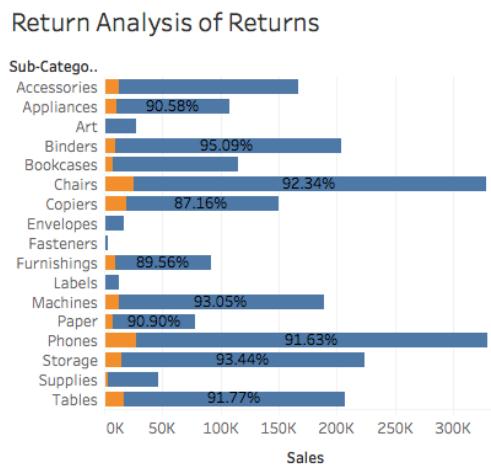


As you can see, due to different nomenclatures, this visualisation is not ideal. So, one possible solution to this is **Grouping** where you can place 'LF' and 'Low Fat' in one group, and 'reg' and 'Regular' in another

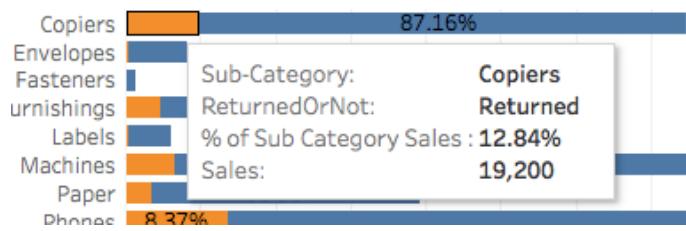
## Grouping



Let's understand this a bit better with the help of the following Dashboard :



The above is the **Returns Analysis** across Categories and their Sub Categories. Although, it cannot be seen, in the bar graph, **Copiers** have the maximum Return percentage, followed by **Furnishings** :

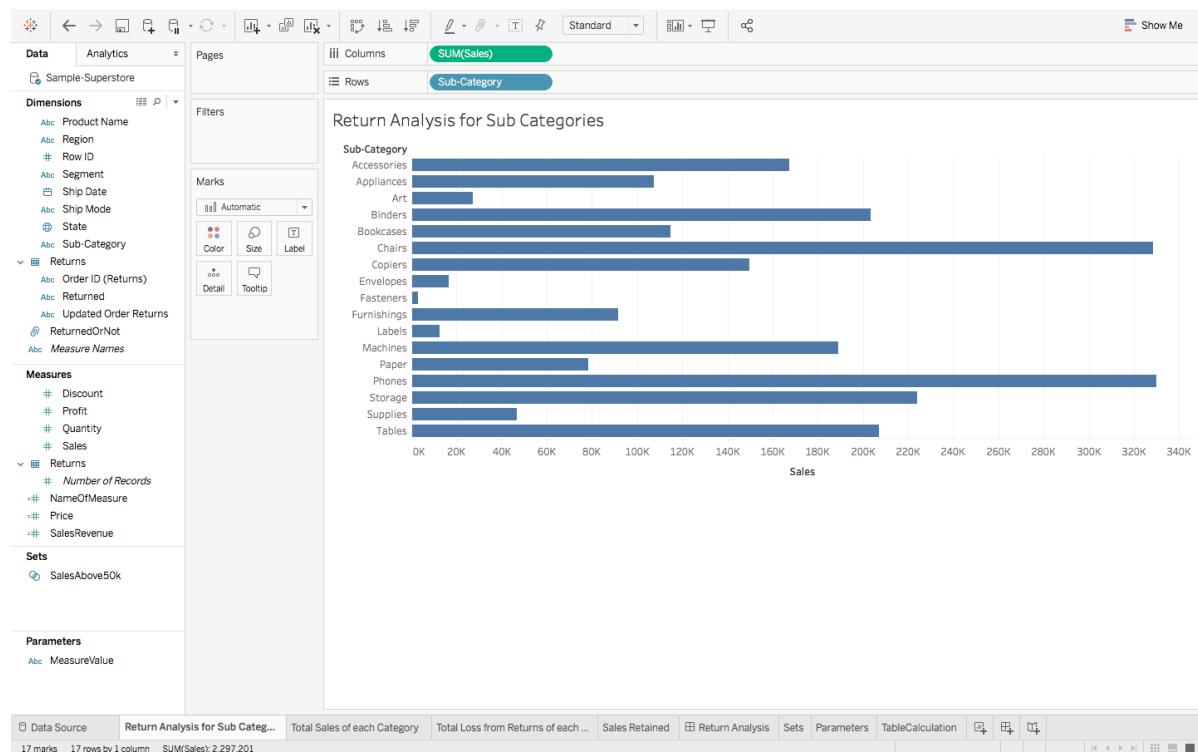


As far as the Line Chart goes, it seems that the Sales Team had been right all along. The Returns were in fact increasing quite rapidly, but luckily from what we can see, the rise is slowly receding.

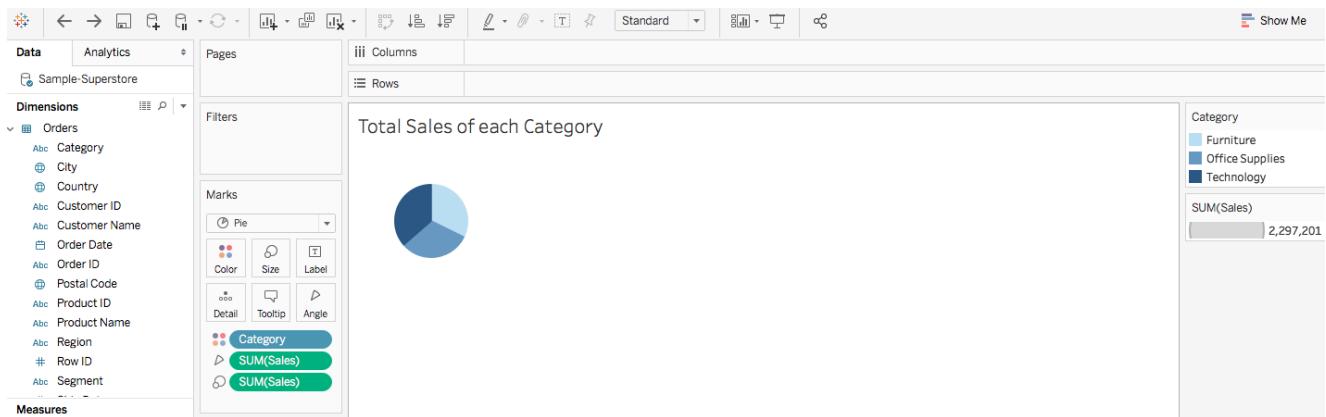
From the pie charts, you can clearly analyse the Returns of each Category. As can be observed, **Technology** suffered the maximum number of Returns.

Another analysis, which I am going to leave for you to make, could be the Return distribution across the various States. Once you finish learning how the above graphs were made, you can easily make this too. So, let's get started :

1. We are going to begin with the following graph, and do take note that the dataset is a **Left Join** of Orders and Returns :



2. In another sheet, make this graph :



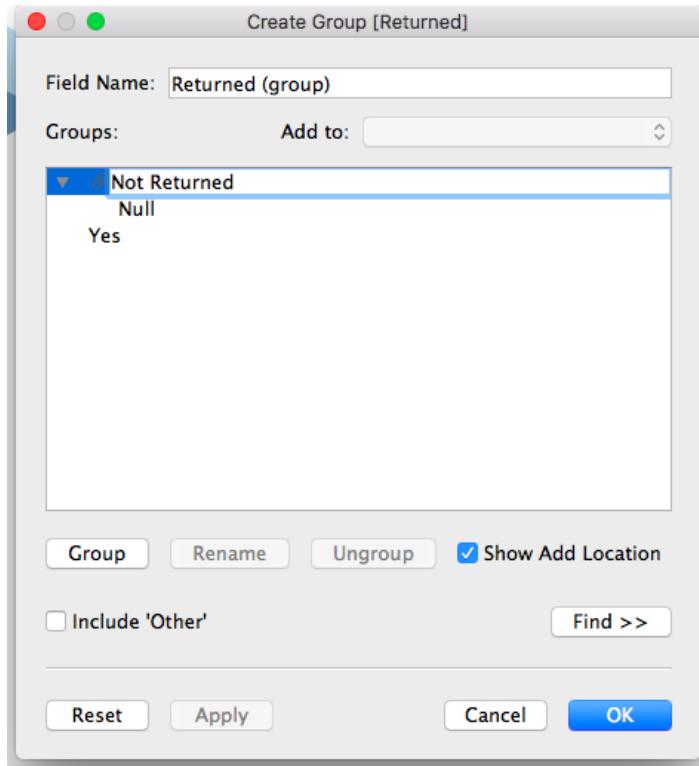
3. You see the Dimension of **Returned** under **Returns**? This is what we are going

**Dimensions**

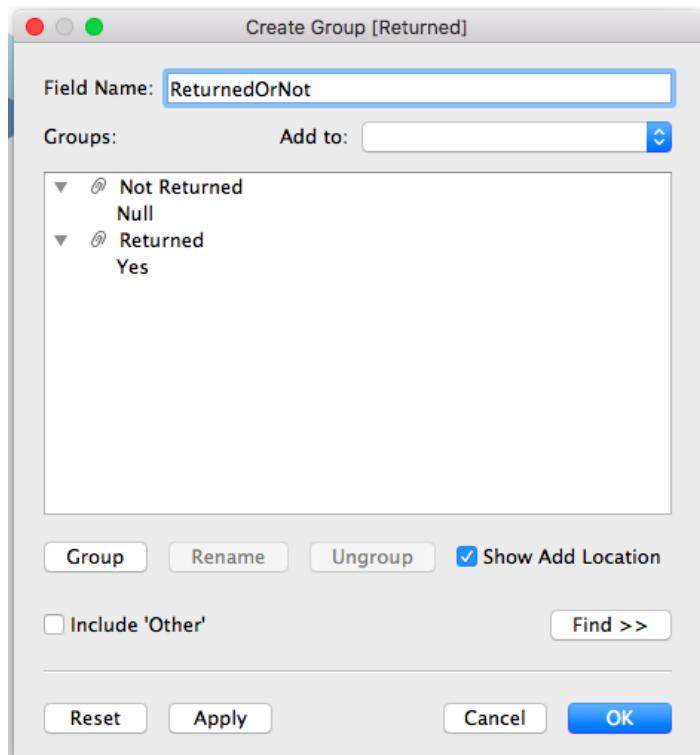
Abc Product Name  
Abc Region  
# Row ID  
Abc Segment  
Abc Sub-Category  
Abc Order ID (Returns)  
Abc Returned

to use to make our groups. Right click on this Dimension :

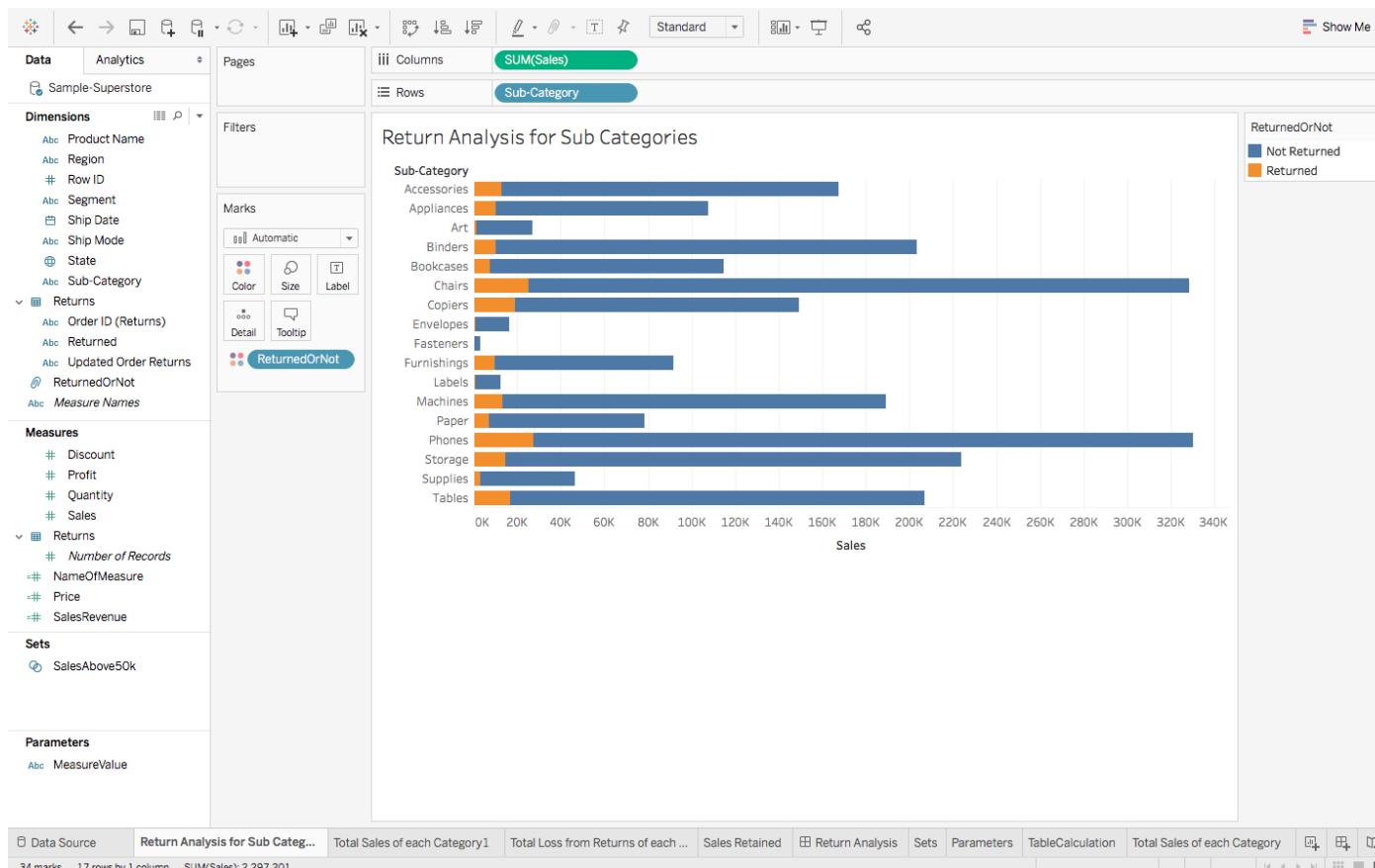
4. Go to **Create –> Group**. Since we are going to make groups of products that were returned or not, click on **Null -> Group**, and change the name :



5. Do the same for **Yes**, and get to the final configuration :



6. Click on **Apply** -> **OK**, after which you will see a group by the name of **ReturnedOrNot** under Dimensions.
7. Drag this new Dimension over the **Color** in the **Marks Pane** to get the following view :



This step automatically segregated the Sales of each Sub Category based on whether the Orders had the **Null or Yes** value under **Returns** or not.

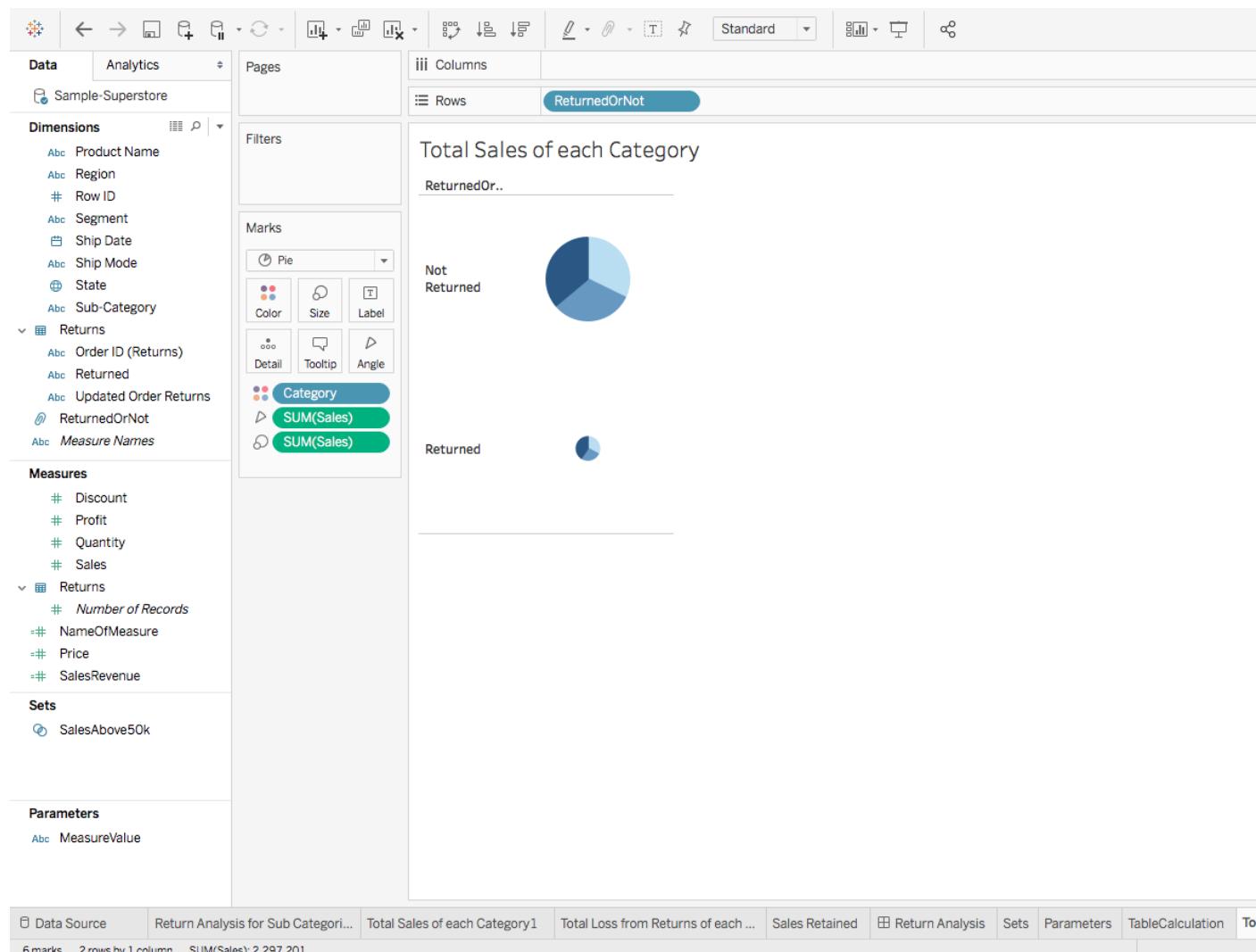
The remaining steps are merely customisation. Let's do those as well :

8. Drag **Sales** onto **Label** in the Marks Pane
9. Right click on the **Sales** pill -> **Quick Table Calculation** -> **Percent of Total**
10. Right click on the **Sales** pill again -> **Compute Using** -> **Table (across)**

Also, in the Legends, you will most likely see 'In/Out' as the aliases. You can change this as per your requirements, by right clicking on the **In/Out blue** pill in the Marks Pane, and choosing **Edit Alias**.

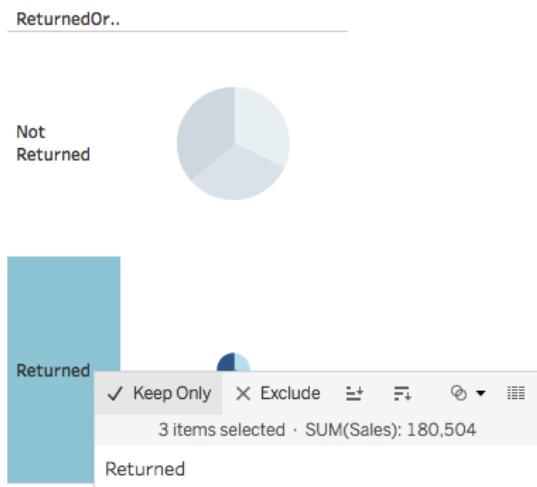
Let's shift to the pie chart that we had made. We are going to apply the same **ReturnedOrNot** group to this as well. First create two duplicates of this sheet, and work on one of them :

1. Drag the group onto the chart, to get something like this :



2. Right click on **Returned**, to get :

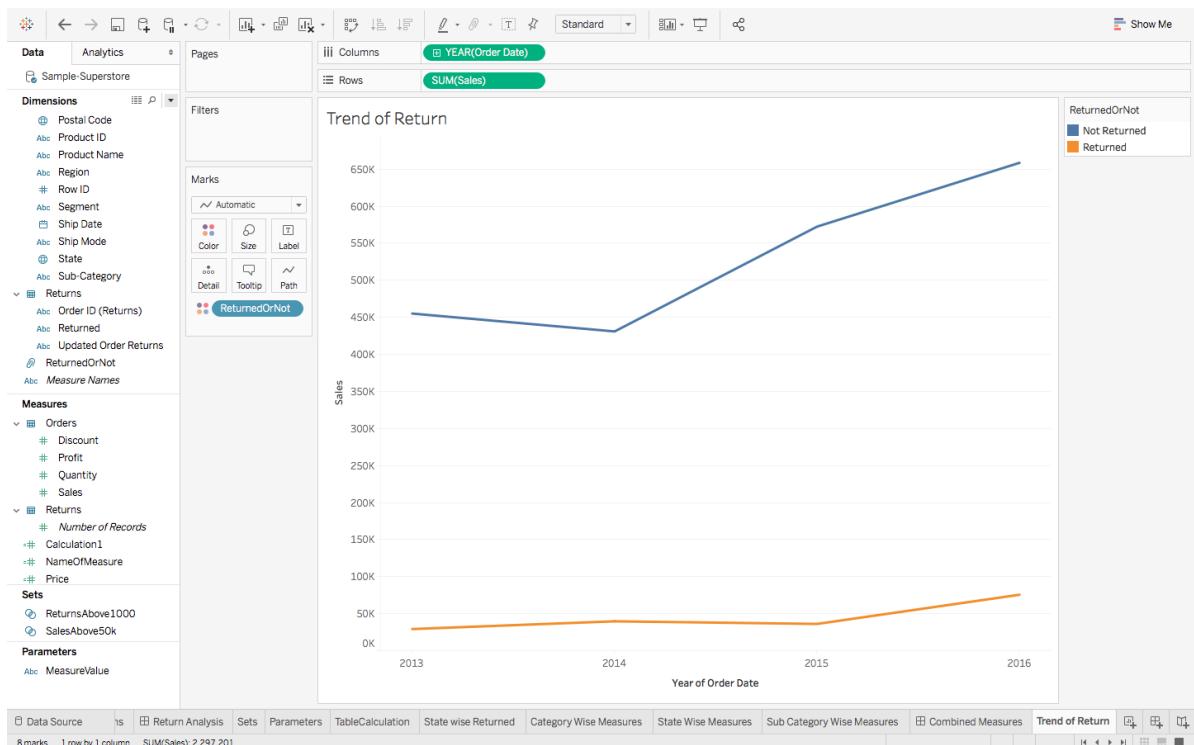
## Total Sales of each Category



3. Choose **Keep Only**, so that you just see the **Returned** graph. This gives you the **Total loss from Returns**. All that is remaining is labelling with the Sales value.
4. Now go to the second duplicate of the sheet, repeat the first 2 steps, and instead of keeping the Returned chart, we are going to keep the **Not Returned** chart. This will give you the sales of the **Remaining / Retained Orders**.

Its equally easy making the Line Chart :

1. Drag **Order Date** to Columns and **Sales** to Rows to get a Line Chart
2. On top of this chart, drag the **ReturnedOrNot** Dimension, but just like you had before, exclude the **Not Returned** Line, right clicking on it and choosing **Exclude**.



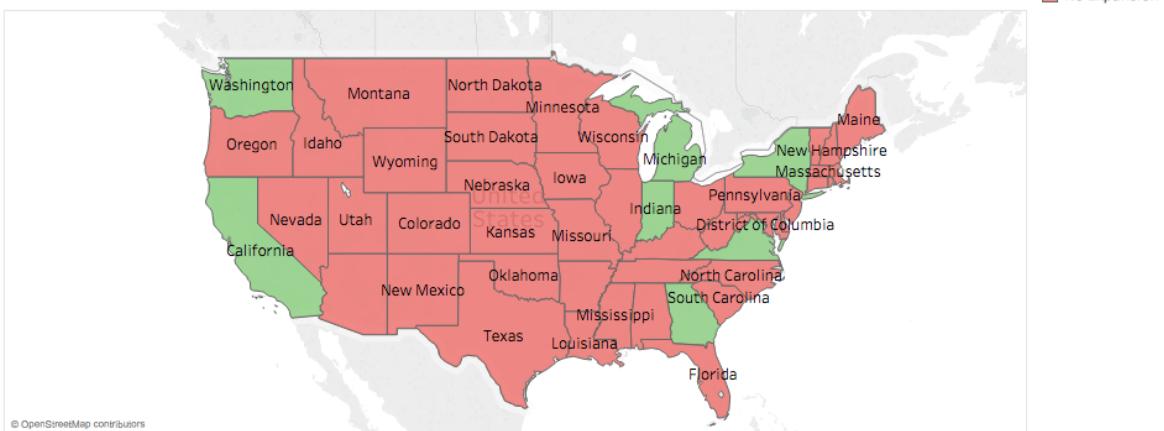
All that is left now, is combining the above Worksheets into one Dashboard. Why don't you try making the chart for State wise Returns distribution as well?

## 2.2 Sets

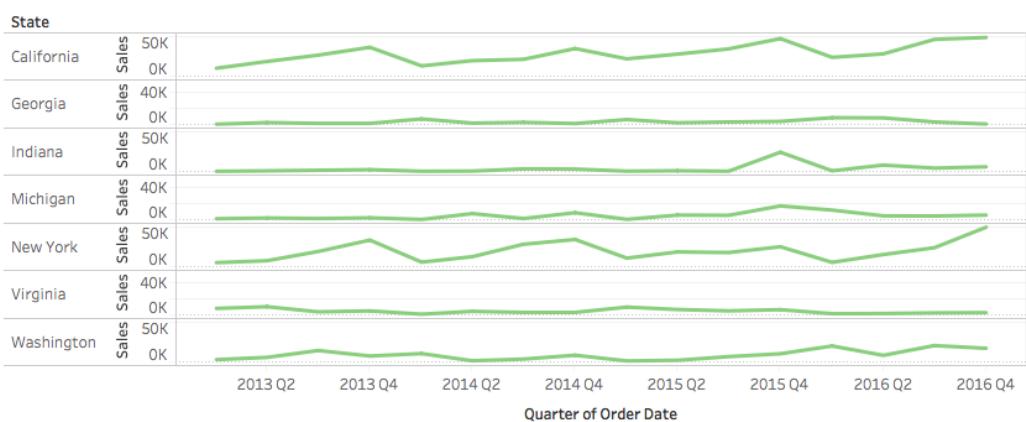
After taking note of the Returns analysis, your organisation decided that the increase in Returns was not that alarming, and that it should not be constituted as a reason for non-expansion.

But the Superstore is only going to expand in those States where the Sales and Profit both have crossed a certain margin, for example, 40,000 and 10,000 respectively :

## States for Expansion



## Sales Growth



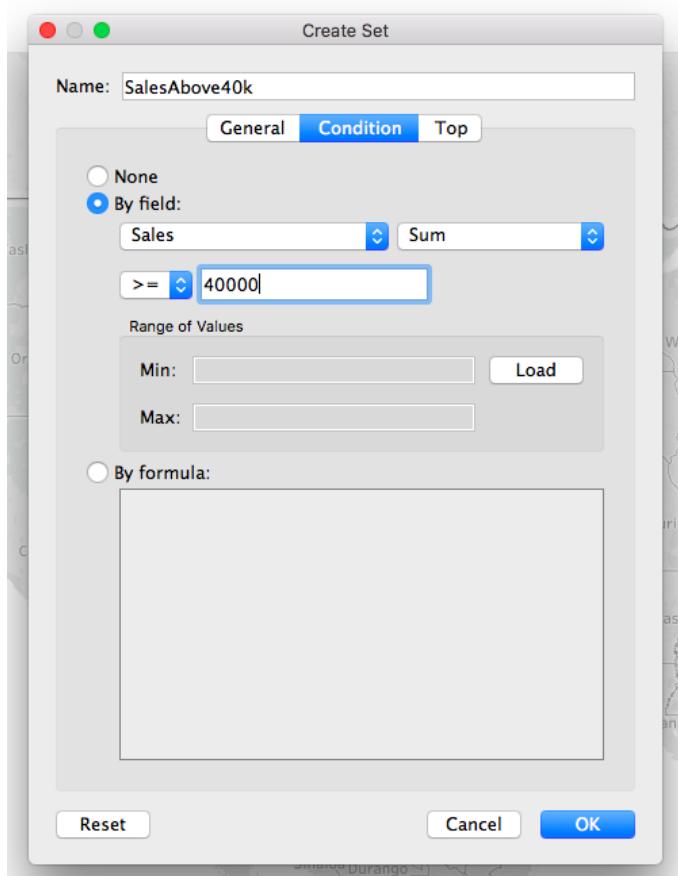
So, Sets, as are created above, are really similar to groups. In Set, you **group** data that fulfills a particular **set** condition. Another interpretation could be: Groups help you attain a higher level hierarchy, as we had seen in the previous example, whereas Sets help you attain a lower level granularity.

Let's understand this better by creating the above Dashboard :

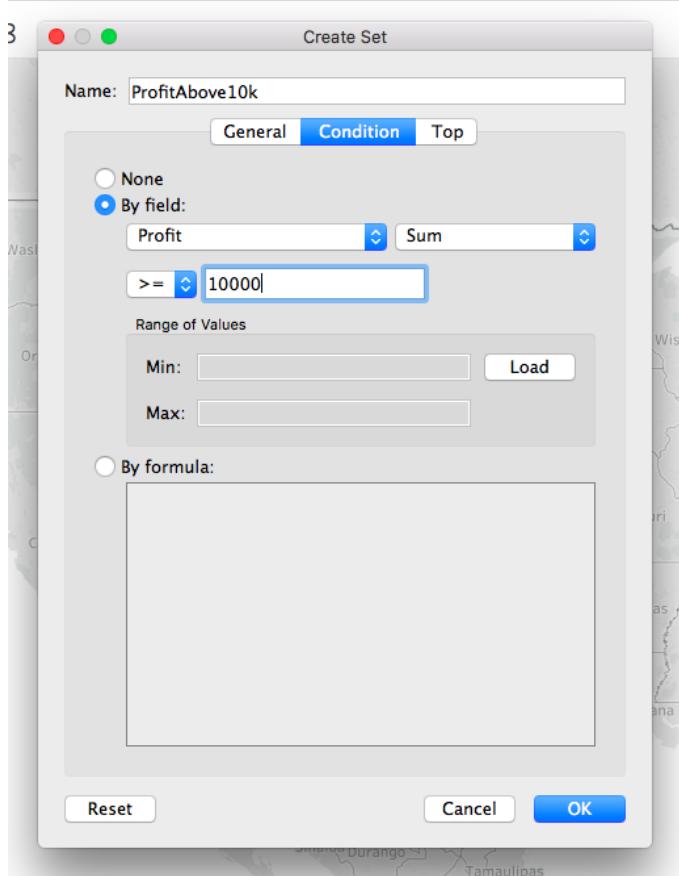
1. We are going to start off with a simple Map Chart. Since we are going to map our Sales and Profits onto this graph, right click on **States** and choose **Create -> Set**.

The screenshot shows the Tableau Data Source pane on the left and a map view on the right. In the Data Source pane, the 'State' field is selected. A context menu is open over the 'State' field, with the 'Create' and 'Set...' options highlighted.

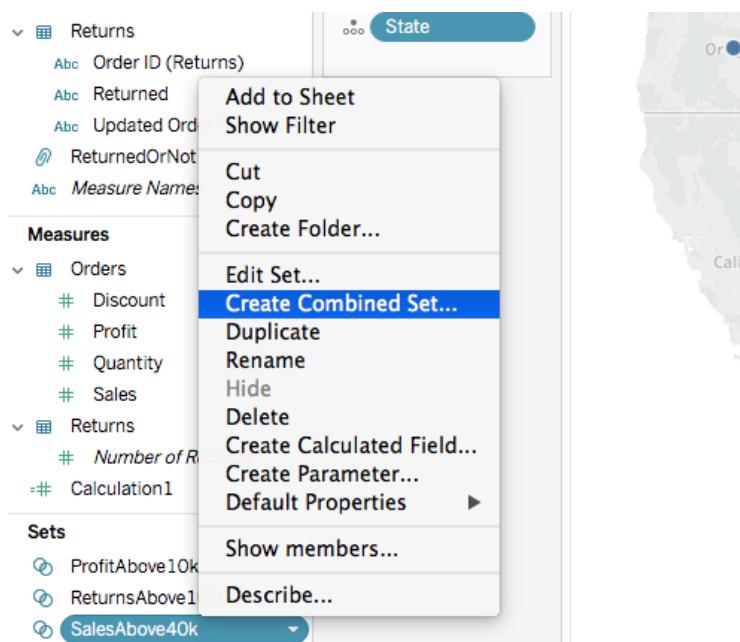
2. Fill in the fields with the following demo information :



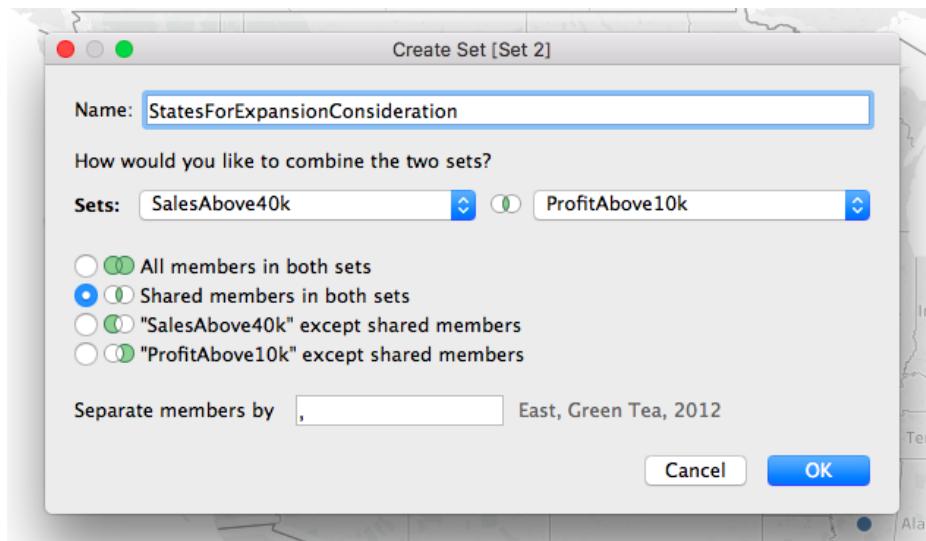
3. Right click on States once again, and this time fill information for the Profit Margin



4. Now we are going to join the two sets to obtain our desired configuration. So right-click on **SalesAbove40kset** -> **Create Combined Set** :

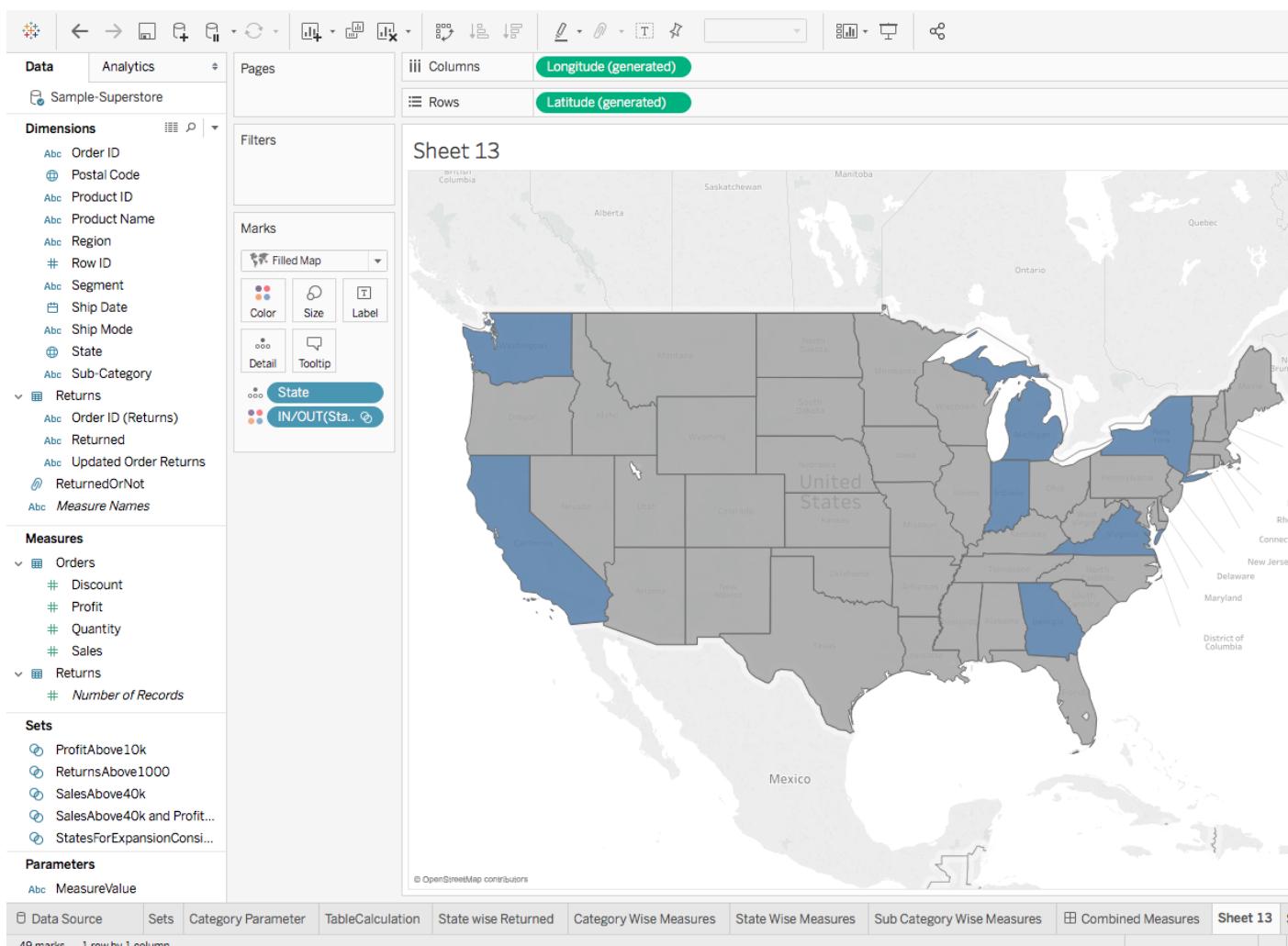


5. Fill in the fields with the following configuration :



This step joins the above two conditions for Sales and Profit, to get the requisite combined computation. To view the results :

6. Drag this new Set on top of the Map Chart, and convert it into a Filled Map instead of the Symbol Map :



You can always customise the above chart by changing the colour, adding labels etc.

The line chart is as easy to make as the one we had made previously for **Trend of Return**. Here we have excluded the States that belong to the **No Expansion** group, like we had excluded the **Not Returned** there.

### 3. Calculated Field

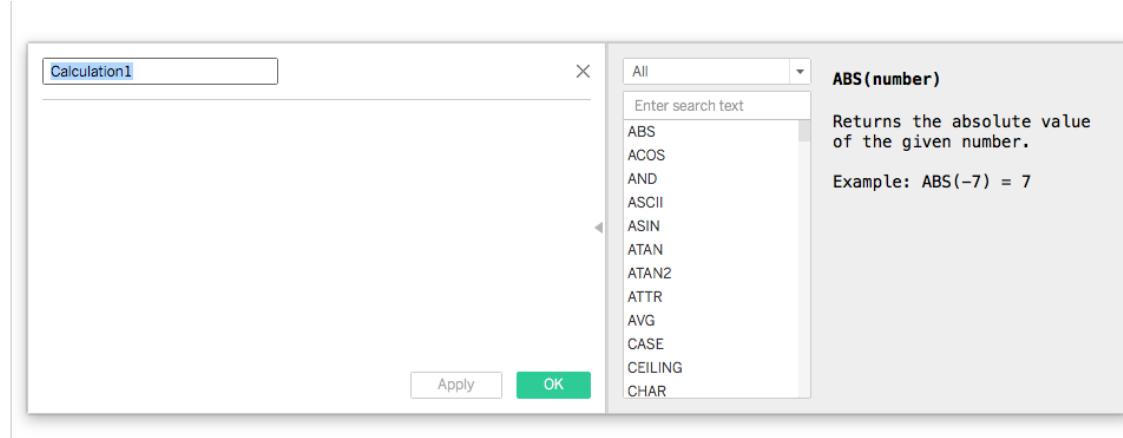
The Superstore dataset is pretty comprehensive. It offers quite a lot of information and field sets. But like all data, there is always the possibility of extracting more features. Calculated Fields help you do exactly that while also allowing you to carry out both simple and complex calculations on the data.

*So what is a Calculated Field?*

To put it in simple words, it's a formula that you apply to your data, where the various Measures act as the variables.

*How to create one?*

Just simply go to **Analysis**, click on **Create Calculated Field** and something like this will pop up :

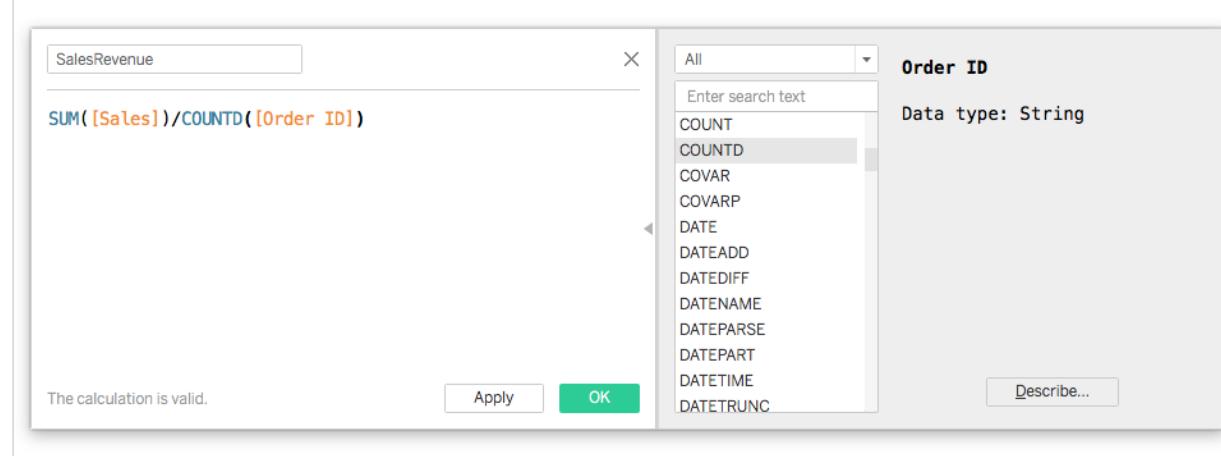


Here is where you write your formulas, and as you can see, Tableau provides you with the various syntaxes too, so that you never feel lost! You can also apply 'If-Else' conditions, 'Case' conditions (as we shall see next) and of course the usual mathematical computations too, which we will explore now.

So what Calculation to start with? Let's begin with something simple, that is, Average Sales associated with the Orders. The most apt formula for the same would be **Total Sales / Total Number of Orders**. To convert this formula in Tableau terms, Total Sales implies the **SUM of Sales**, whereas Total Number of Orders means their **COUNT**.

Let's get to the application then :

1. In the Calculation Box, make the following changes :



2. To see the cumulative **SalesRevenue** (Calculated Field), drag **Measure Names** from Dimensions to the empty area to see the above computed result :

The screenshot shows the Tableau Data Source interface. The left sidebar lists dimensions like City, Country, and various Order-related fields. The "Measures" section contains fields such as Price, SalesRevenue, and Latitude (generated). The "TableCalculation" section is expanded, showing a table of calculations. The table includes:

Discount	1,561
Number of Records	800
Price	608,830
Profit	286,397
Quantity	37,873
Sales	2,297,201
SalesRevenue	459

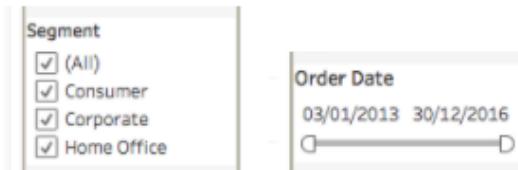
Below the table, there is a list of measure values, including `SUM(Discount)`, `SUM(Number of Rec...)`, `SUM(Price)`, `SUM(Profit)`, `SUM(Quantity)`, `SUM(Sales)`, and `AGG(SalesRevenue)`.

So what you have basically accomplished with the Calculated Field is create a Measure of your own, which you can use just like Sales and Profits.

Obviously this was just a gist of what Calculated Fields can do. They can be used for various complex calculations as well, and the glimpse of one such instance can be seen in the following section.

## 4. Parameter Control

Filters such as the following are a great way of interacting with the visualisations on Tableau :



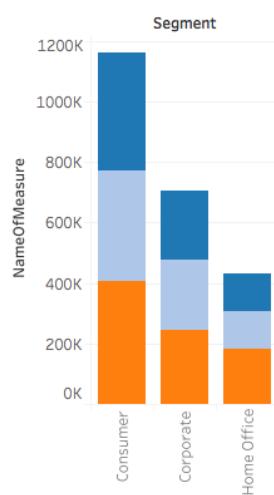
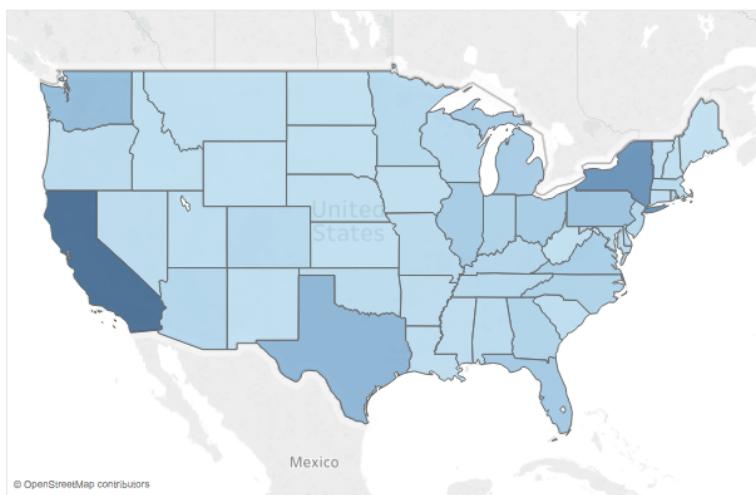
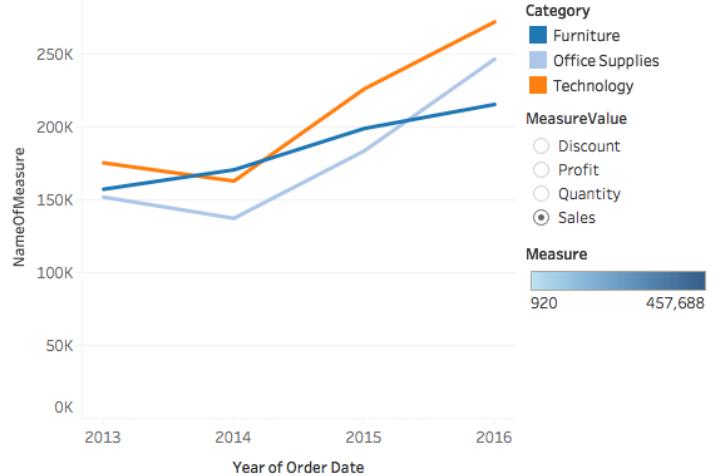
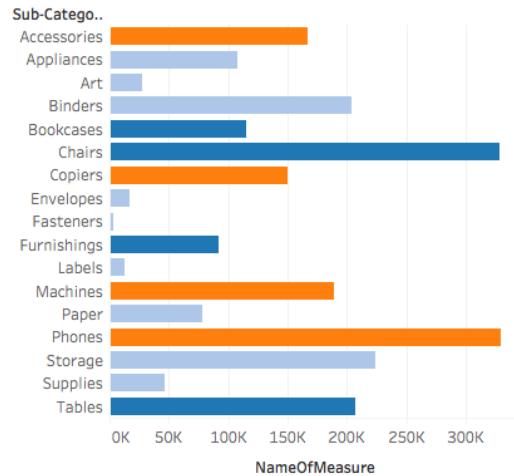
Just like filters are measures through which you can look at various aspects of your data, a Parameter is another great feature. It can be used *in place* of Filters, and can exhibit its own dynamic property too.

So what are Parameters? These act as variables of an equation, which you can change to get different results each time.

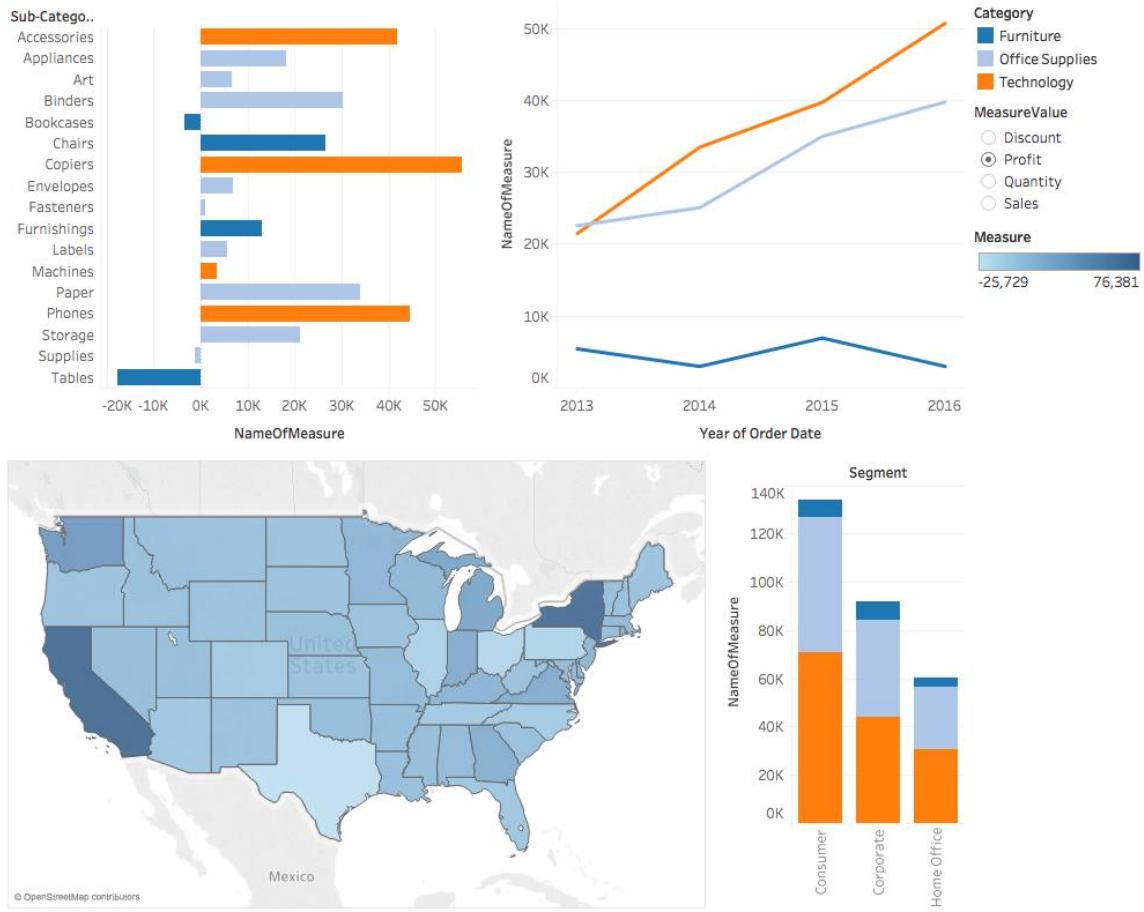
Let's try and understand this with the help of an example. So far we have been making separate graphs for separate Measures. Whenever we had to analyse Sales, Profits, Quantity or Discount of the various Categories of Products, we had to make different charts, everything being the same BUT the Measures.

With the power of Parameter Control, there is actually a pretty easy way of going about this, without the repetitions.

Consider this for Sales :



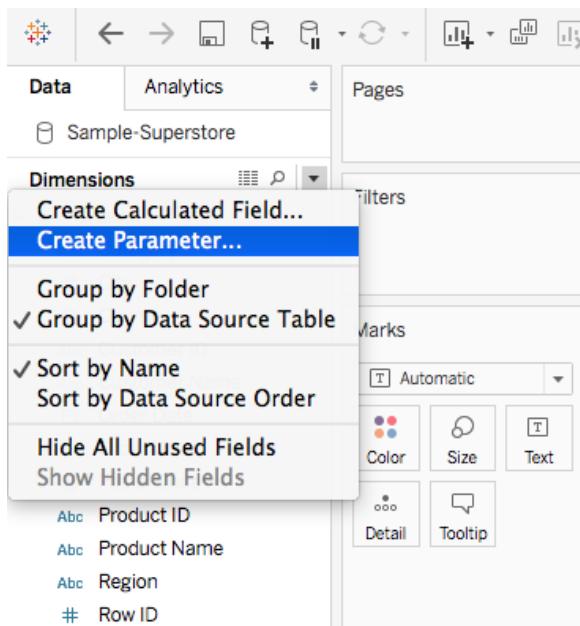
And the following for Profit :



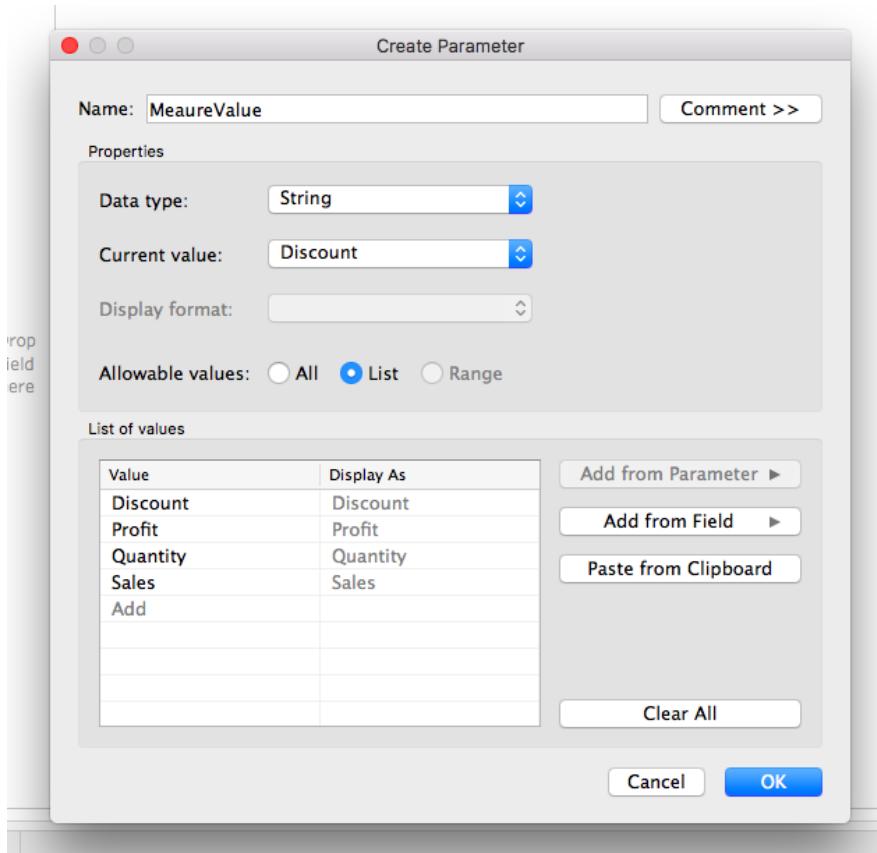
Just with a simple click on the **MeasureValue** list, you are getting cumulative results across various Dimensions. Here, the **MeasureValue** is our Parameter, since we are able to change its value.

Let's get to making one of our own now. Parameters rely heavily on Calculated Fields, so you are going to get a nice chance to practice what you learned above:

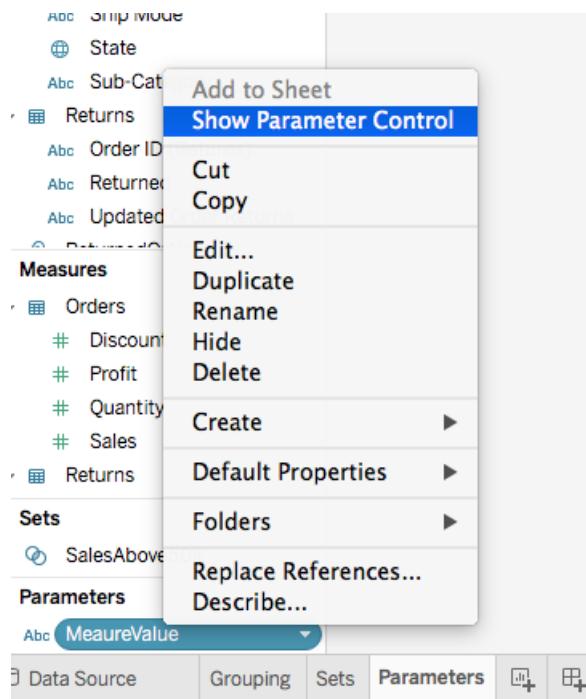
1. We are going to start off with a blank screen this time.
2. Click on the downward arrow near Dimensions, and choose **Create Parameter** :



3. Since we are going to interchange the Measure Values between Sales, Profits, Quantity and Discount, fill in the fields shown below and click OK :



4. Right click on the newly generated Parameter, and choose **Show Parameter Control**:



Do not expect the drop down list to start magically creating graphs on its own. There are still some computations to be done.

So far we have only allocated the names to the Parameter, but not the Values that they are supposed to take up. So for this purpose, we are going to create a Calculated Field.

5. Go to Analysis -> Create Calculated Field, and fill it with the following self explanatory details :

The calculation is valid.

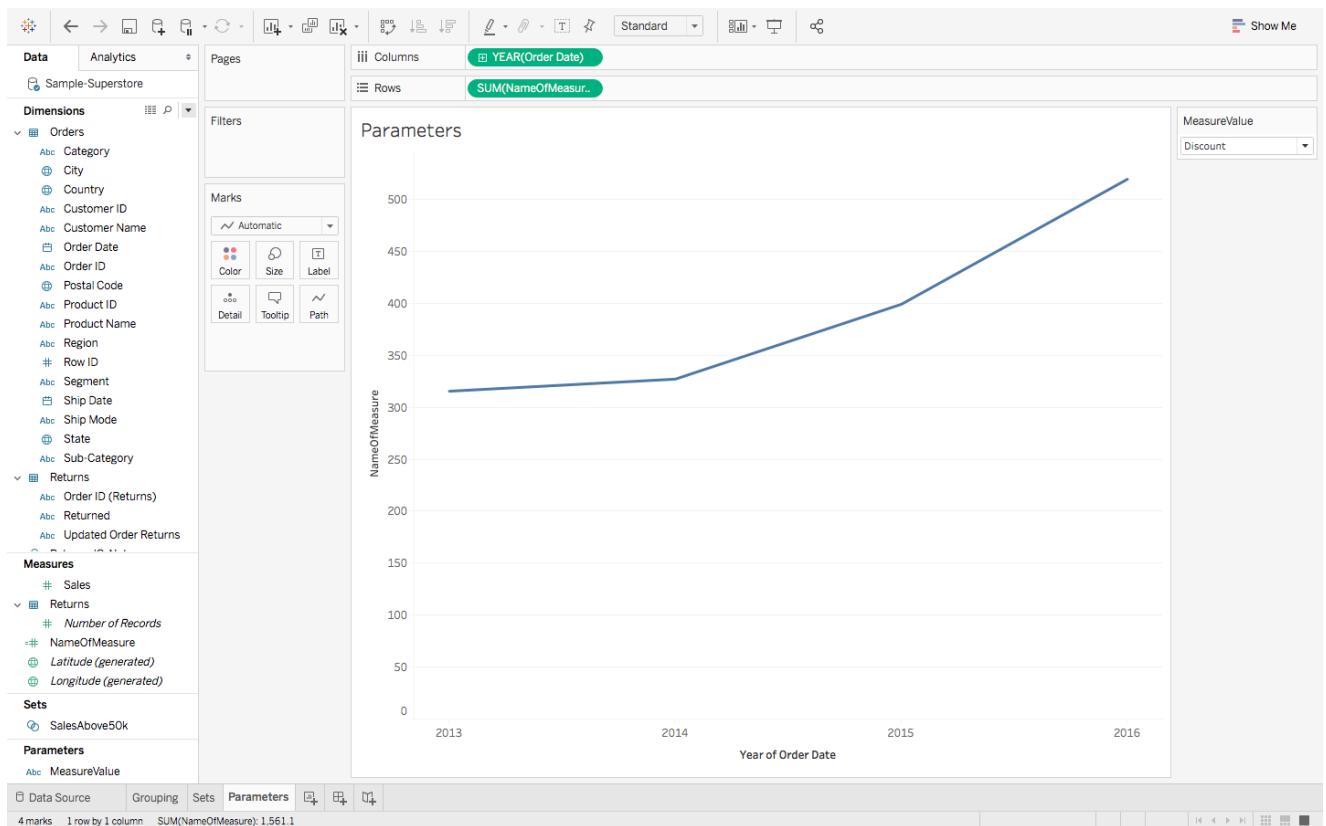
**Apply    OK**

**ABS (number)**

Returns the absolute value of the given number.

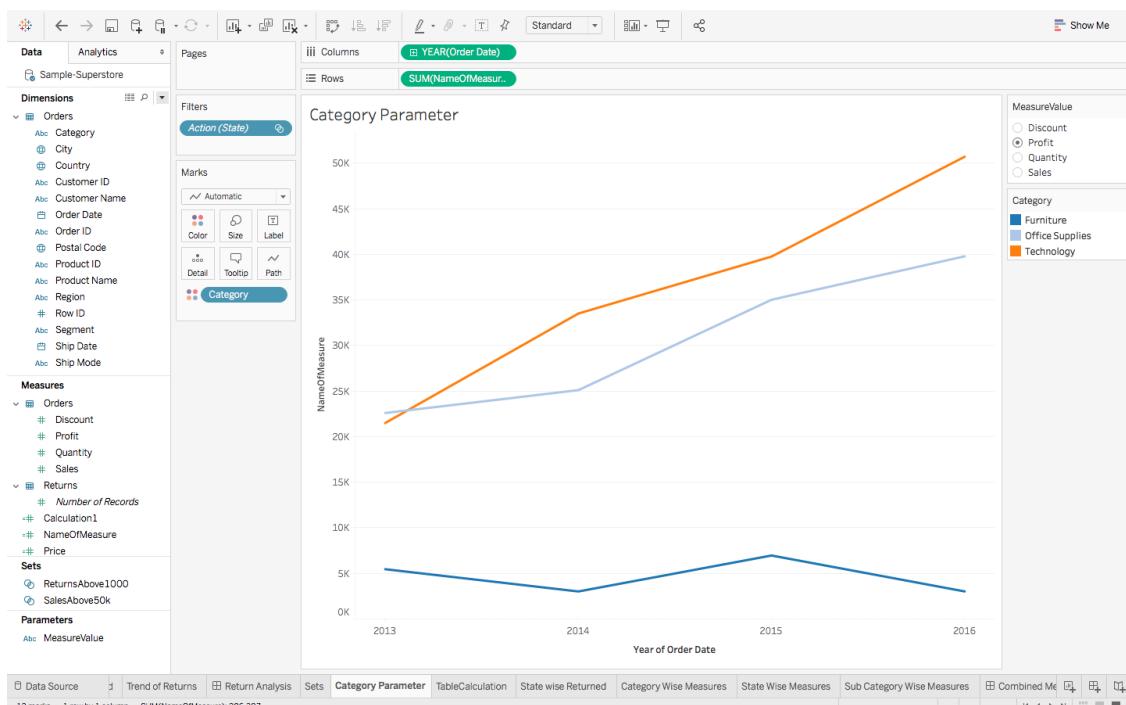
Example:  $\text{ABS}(-7) = 7$

6. You will find the newly created field under Measures. So first drag **Order Date** to Columns and then **NameOfMeasure** to Rows :



There you go! Go ahead and try changing the MeasureValues now.

Now onto the chart we have created in the Dashboard. To see the lines for the individual **Categories**, simply drag that Dimension on top of the chart :



For creating other charts, do as you normally would, but instead of dragging the individual Measures onto the Rows / Columns, drag the **NameOfMeasure** parameter instead.

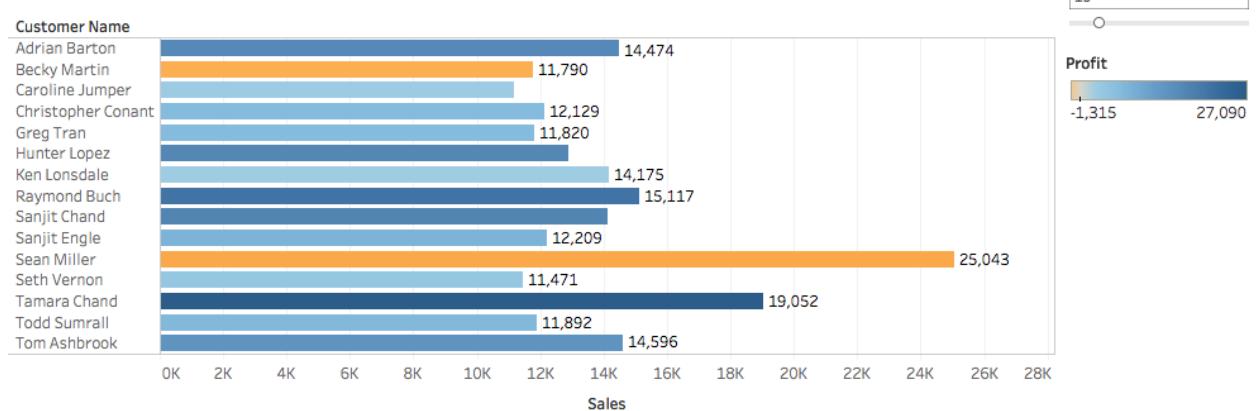
## 5. End Note of Intermediate Tableau guide for data science and business intelligence professionals

That brings us to the end of this article. But don't worry, I will be back again with another article on Tableau!

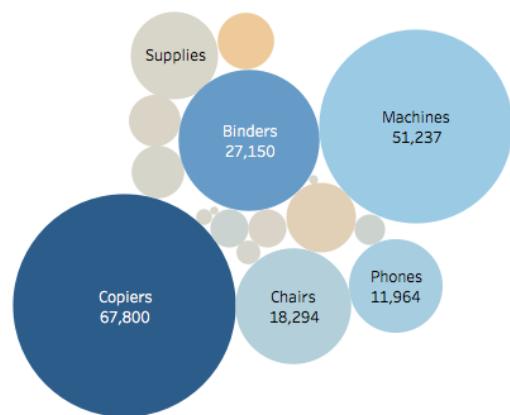
Meanwhile, I think it's customary to give you a Dashboard to make :

### Customer Analysis

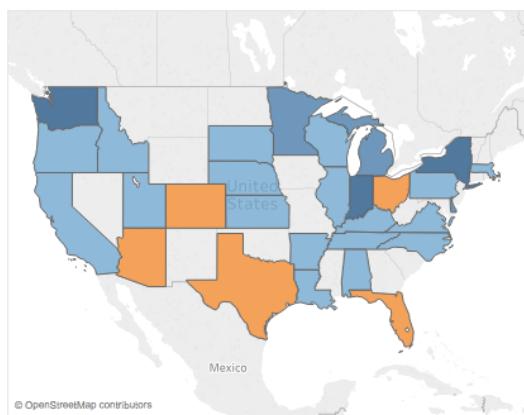
#### Top x Customers



#### Product Purchased by the Top x customers



#### Demographic of the Top x customers



You may face a tad bit of difficulty in the beginning, but if you truly apply yourself, you are bound to get it. And of course if there are ever any doubts, or if you wish for me to cover any concept in the next article, do leave them as comments.

All the best to you Data Explorers!

# A Step-by-Step Guide to learn Advanced Tableau – for Data Science and Business Intelligence Professionals

## Introduction

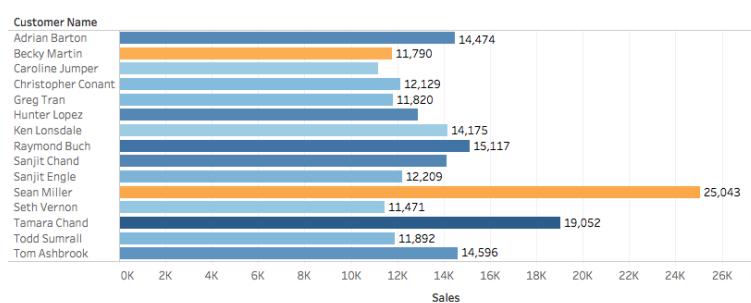
Tableau is one of the most popular Data Visualization tools used by Data Science and Business Intelligence professionals today. It enables you to create insightful and impactful visualizations in an interactive and colorful way.

Its use is not just for creating traditional graphs and charts. You can use it to mine actionable insights thanks to the plethora of features and customizations it offers.

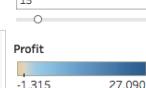
Famous for its ease of use and simple functionalities, making insightful dashboards like the below takes only a few clicks:

Customer Analysis

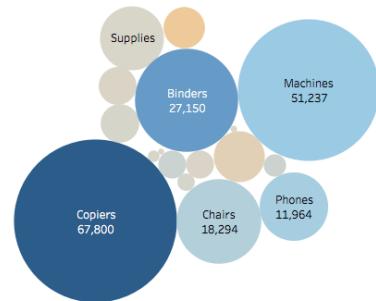
Top x Customers



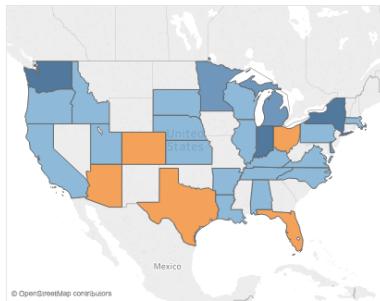
Top x Customers



Product Purchased by the Top x customers



Demographic of the Top x customers



In this article, we will look at a few advanced graphs that go beyond the drag and drop feature. We will create calculations to dive deeper into our data to extract insights. We will also look at how R can be integrated and used with Tableau.

*This article assumes that you possess a fair amount of knowledge about using Tableau, such as basic chart formation, calculations, parameters etc. In case you don't, I would recommend referring to the following articles first and then heading back here:*

1. [Tableau for Beginners – Data Visualisation made easy](#)
2. [Intermediate Tableau Guide – For Data Science and Business Intelligence Professionals](#)

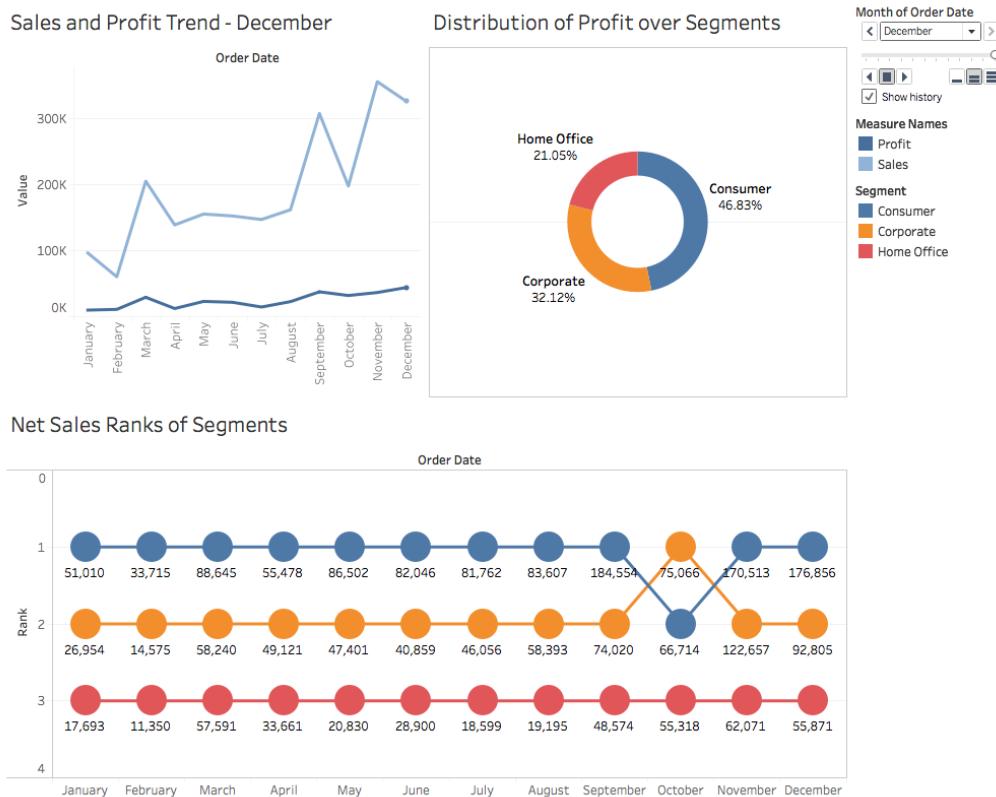
# Table of Contents

1. Advanced Graphs – Visualizing beyond ‘Show Me’
  - o Motion Chart
  - o Bump Chart
  - o Donut Chart
  - o Waterfall Chart
  - o Pareto Chart
2. Introducing R programming in Tableau

## 1. Advanced Graphs – Visualizing beyond ‘Show Me’

Almost all Tableau users are privy to the various elementary graphs, such as those shown in the introductory dashboard. Such charts can be easily made using the ‘Show Me’ feature of Tableau. But since this is an article meant for advanced users, we are going to move beyond ‘Show Me’ and explore graphs that require some extra computations.

First, let’s take a quick look at what we are going to be making in the next few sections. Below is some basic analysis of the Sales and Profit of our Superstore. Simple graphs will serve the same purpose as those in the dashboard, but I think you would agree that there is something exciting and enrapturing about the grandeur of these charts.



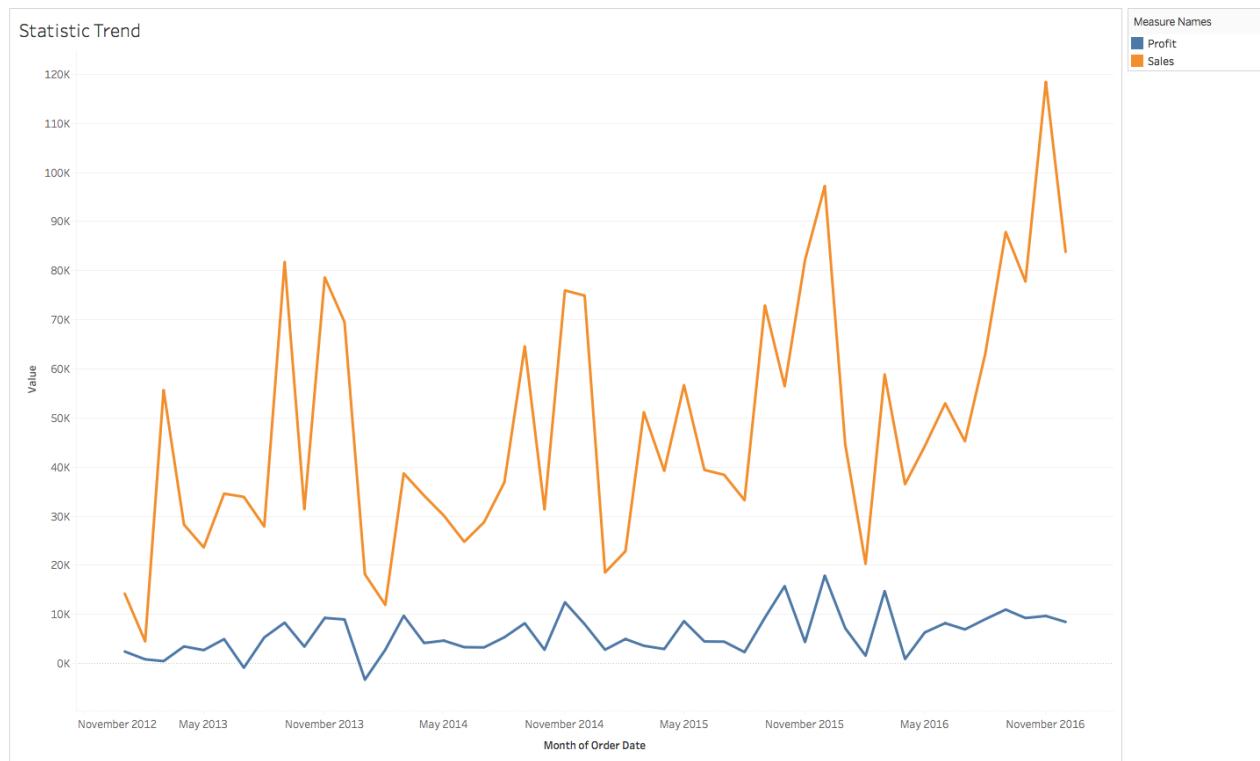
## 1.1 Motion Chart

Before we begin, have a look at [Hans Rosling's World Economics Representation visualization](#). Hit play, and see the magic unfold.

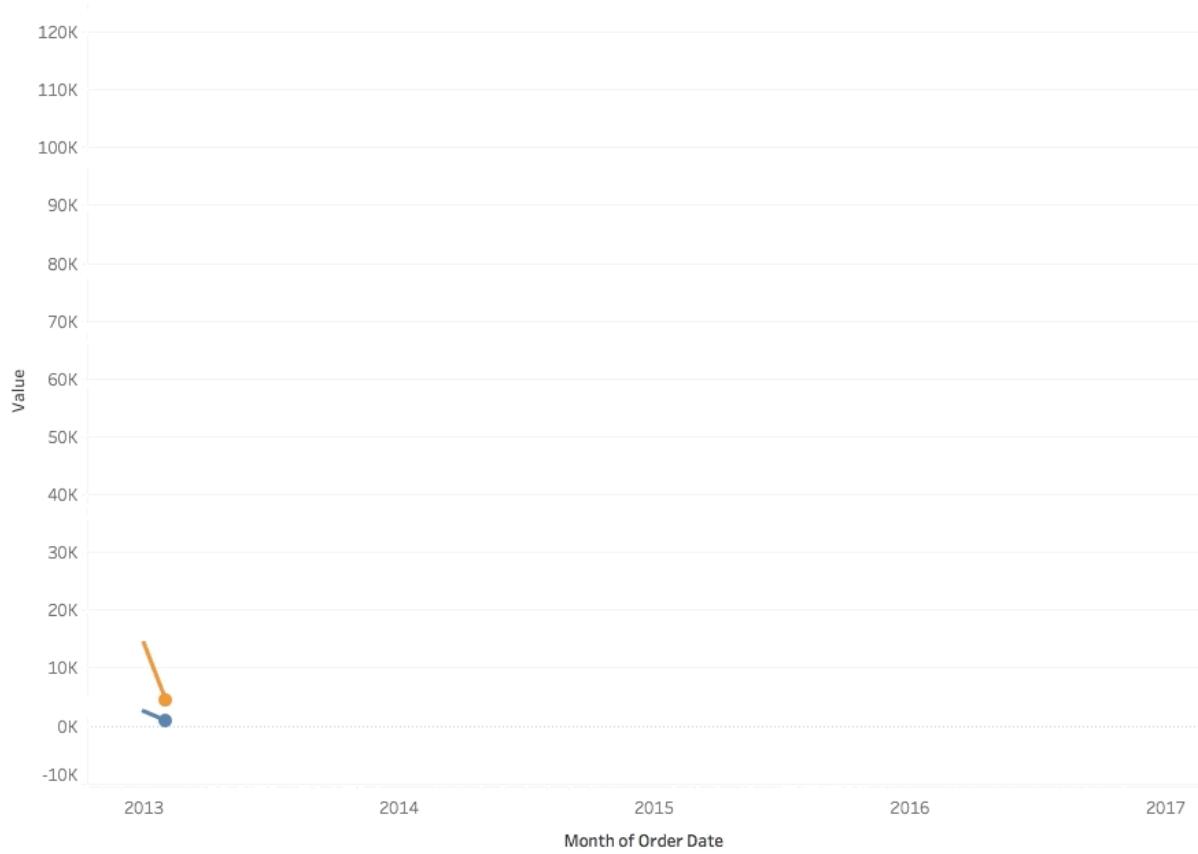
Interested in making one of your own now? If you have already started worrying about animation, don't! What you saw is called a **Motion Chart**. Using this, you can view the changes in your data in real-time.

So let's start by downloading the Superstore dataset which can be found [here](#).

By now making trend lines like the following should be easy for you:

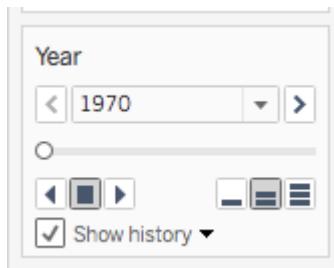


But what we are first going to learn in this section is how to make the below trend lines in motion:



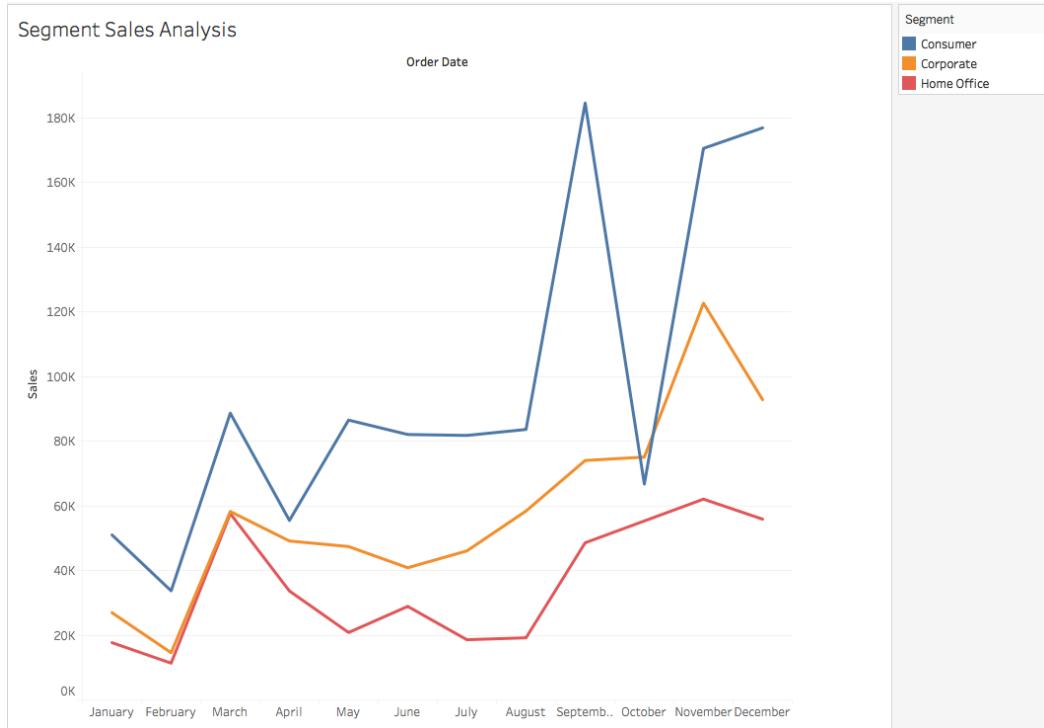
So let's get started!

1. Import your dataset, and create the aforementioned Trend Chart. Our X-axis was the **Order Date** (in the format of Month) and **Sales** and **Profit** are the Measures.
2. All you need to do to make the Motion Chart is drag **Order Date** over to the **Pages** shelf, and change the format again to match with the X-axis.
3. Change the Mark Type from **Automatic** to **Circle**.
4. Go to **Show History**, and select **Trails** to view the trend change. And voilà! Your Motion Chart is ready for launch.
5. Press on the arrow buttons to see the motions, change the Show History customisations, the speed etc.,:



## 1.2 Bump Chart

Suppose you want to explore the Sales of the various segments of the Superstore (for an entire year). One way to do this is the following:



While an alternate option could be the below:

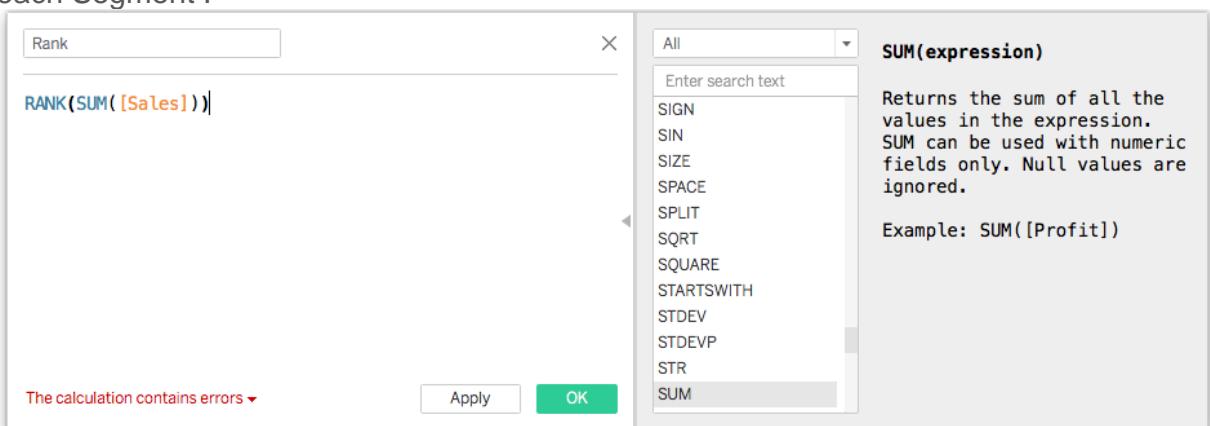


Although the Line chart managed to show the difference of Sales between each Segment, the Bump Chart (in the above image), gave a more clear and concise picture of the same outcome.

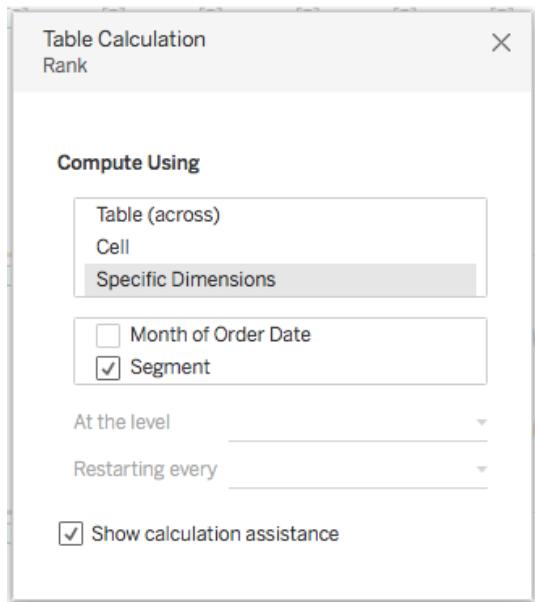
Such charts are mostly used to understand how the popularity of a particular product is changing over the years.

Let's try and make one of our own now:

1. First we need to think of the Measure on the basis of which we wish to rank our Dimensions. Here the **Measure** we have taken is Sales and the **Dimension** is Segment.
2. You need the help of a **Calculated Field** to make Bump Charts. So quickly create a calculation as below below. We are going to rank the **Sum of Sales** for each Segment :

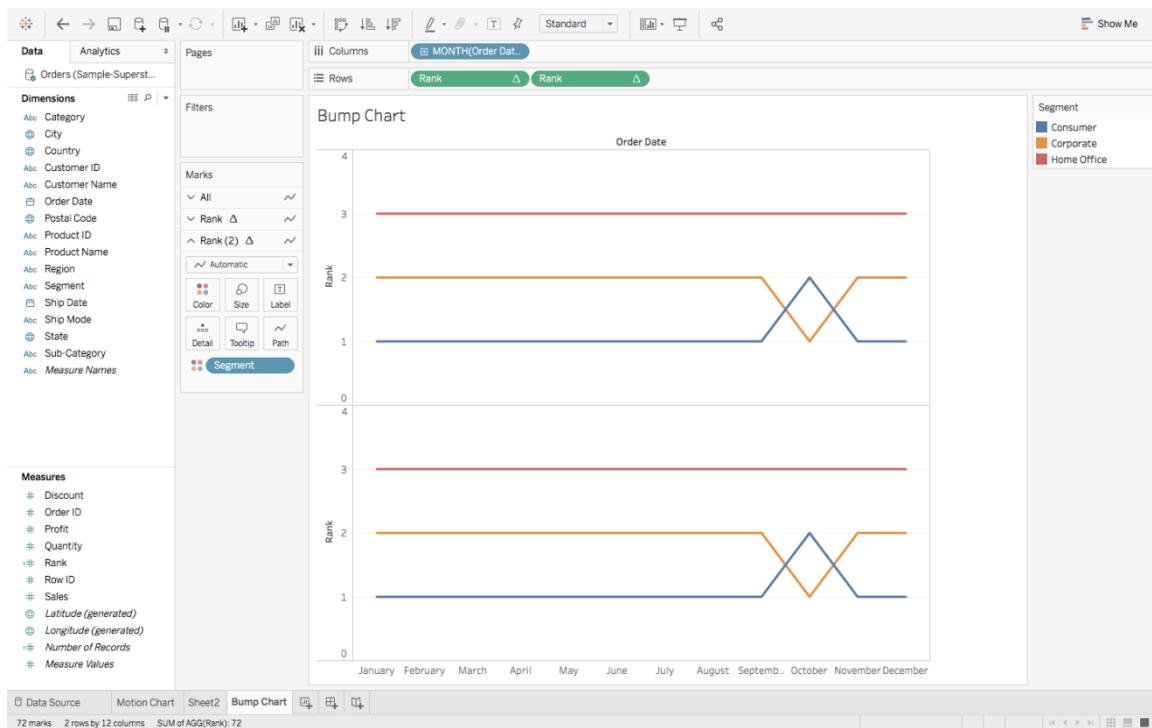


3. Now drag **Order Date** over to the Columns and change the format to Month. Drag **Segment** to Colour in the Marks Pane. And finally drag Rank over to the **Rows**.
4. In the graph that you can see now, the Ranks have been allocated based on the number of months. However, we need them to be on the basis of Segments. So right click on **Rank** in **Rows**, and go to **Edit Table Calculation**.
5. Since we wish to Compute Using **Segment**, change the configuration to:



The chart that you will get won't look like the chart in the dashboard because it lacks the Labels. Let's remedy that quickly, with the help of a **Dual Axis**:

## 6. Drag **Rank** again onto **Rows** and repeat Steps 4 and 5 to get:



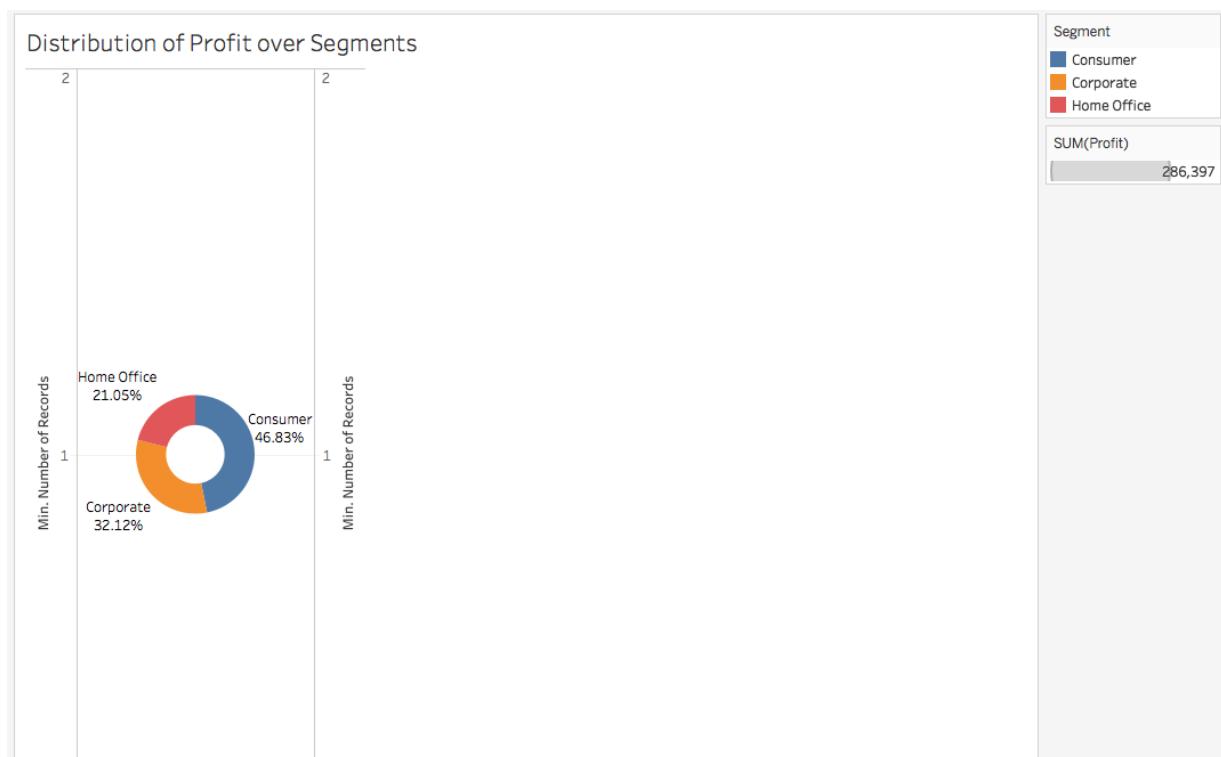
You see Rank and Rank (2) in the Marks Pane? We are going to use these to create those circled Labels.

7. To convert the above into a Dual Axis Chart, right click on the second chart's **Rank** axis and choose **Dual Axis**.
8. In the Marks Pane, choose either Rank or Rank (2), and change the Mark Type to **Circle** instead of **Automatic**.

- Here the Ranks are in descending order. To change it to ascending, right click on the **left** Rank axis –> Edit Axis –> Reversed Scale. Repeat the same for the **right** Rank axis as well.
- Finally, drag Sales onto Labels –> Quick Table Calculation –> Percentage of Total, to get our desired bump chart.

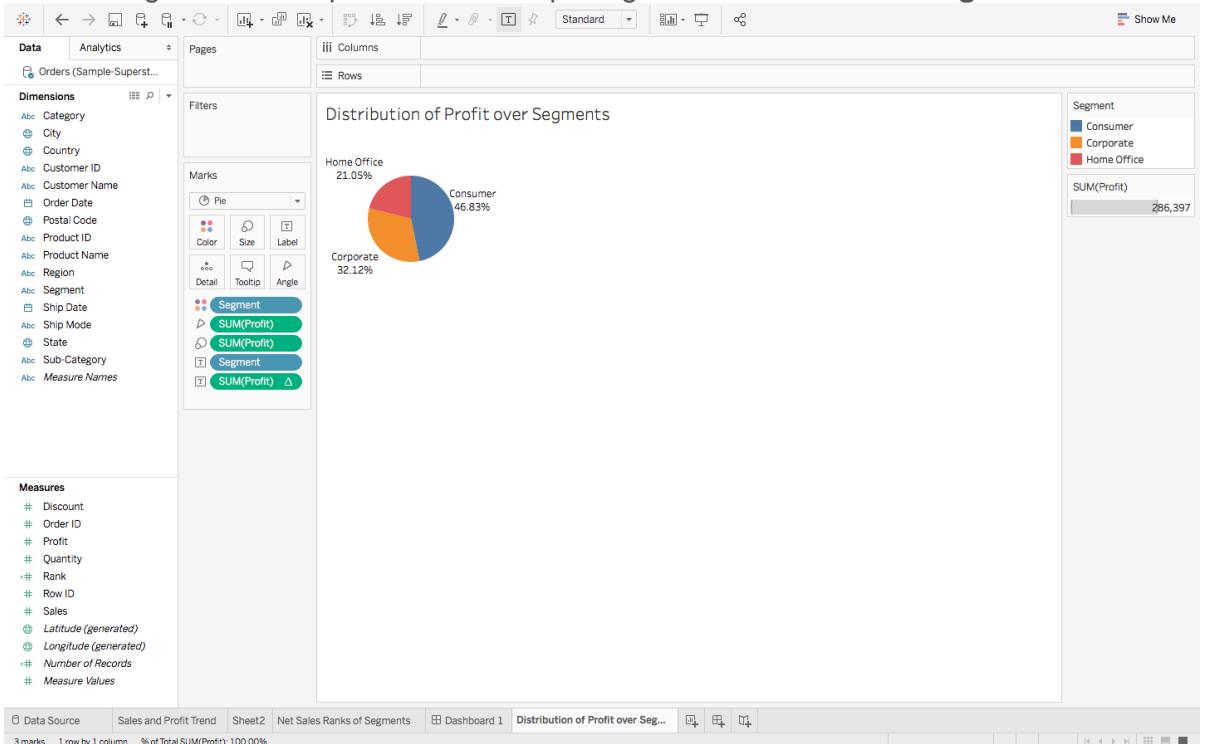
## 1.3 Donut Chart

A donut chart is yet another representation of an elementary chart. To put it candidly, its a pie chart with a hole in the middle, but it helps put more emphasis on the various segments, as you can see below:

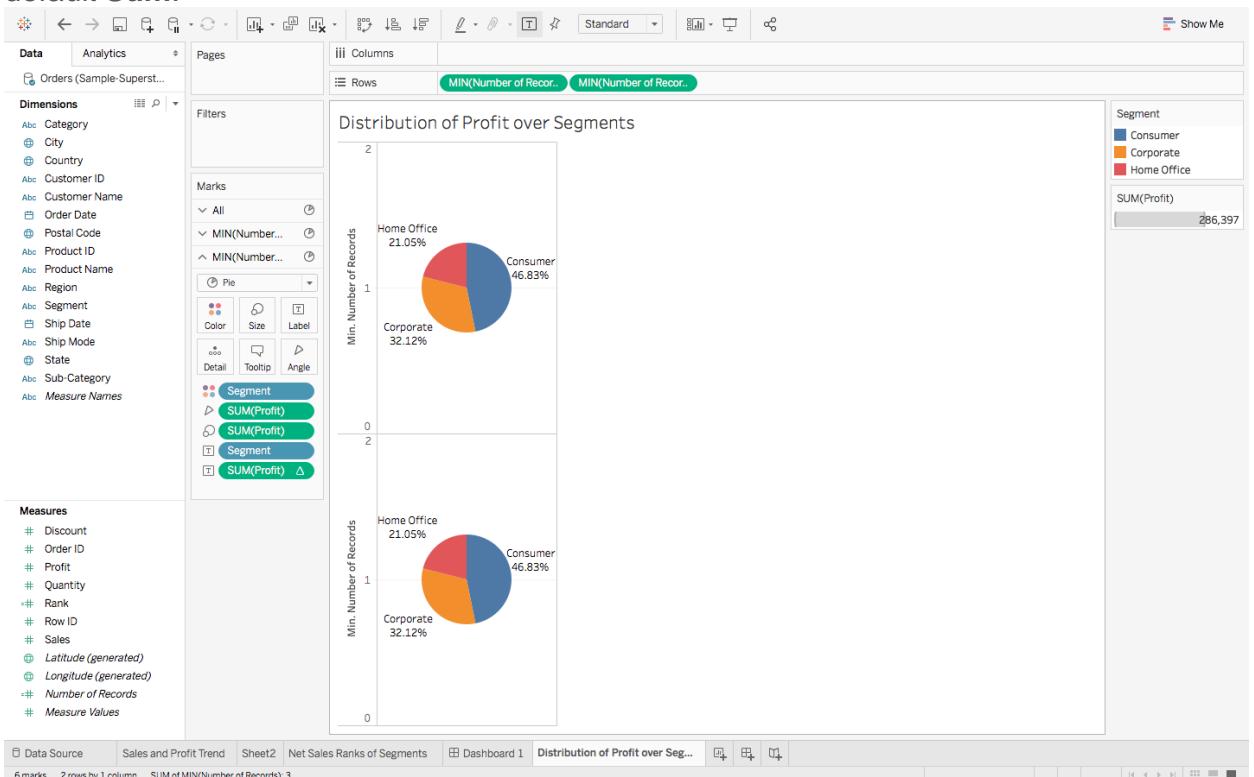


Let's understand the difference as we create this.

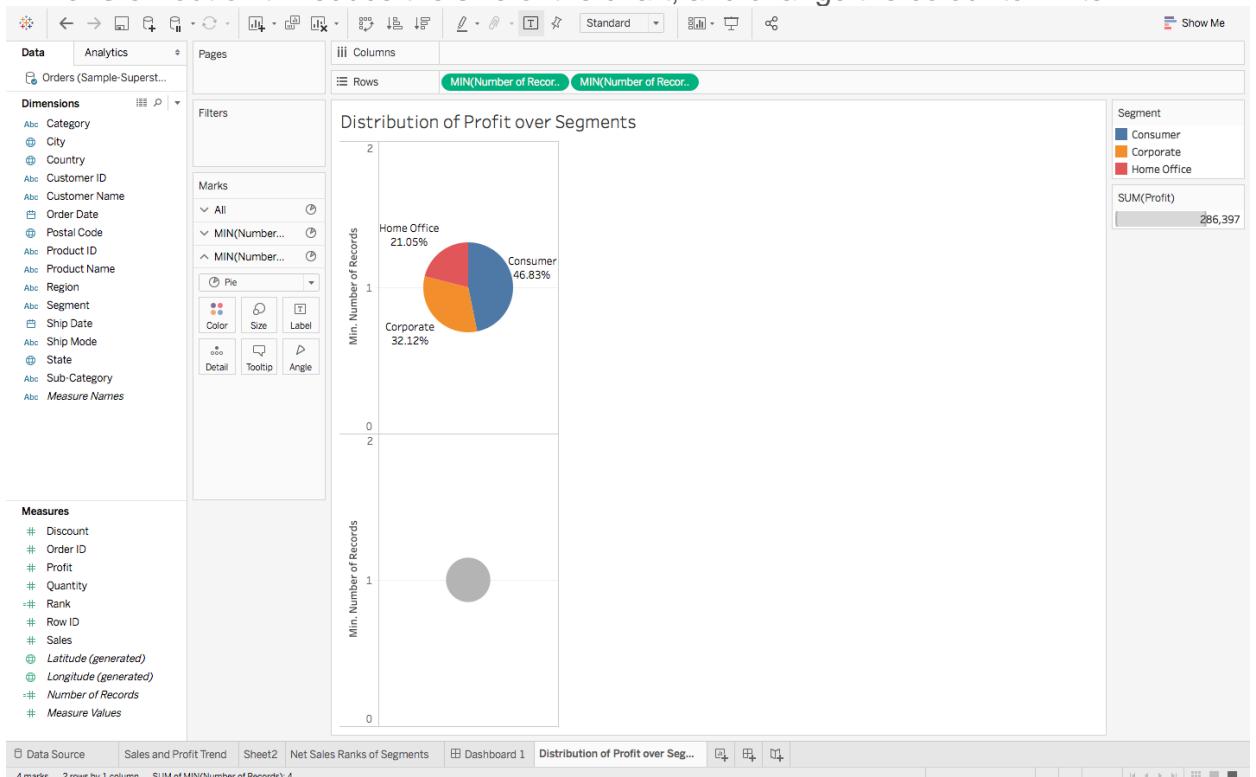
1. We will begin with a simple Pie Chart depicting the Profits of each Segment:



2. To create a Dual Axis for the Pie Chart, drag **Number of Records** from **Measures** over to the **Rows**, twice. Change the Measure of each green pill, by right clicking on them and choosing **Minimum** in place of the default **Sum**:



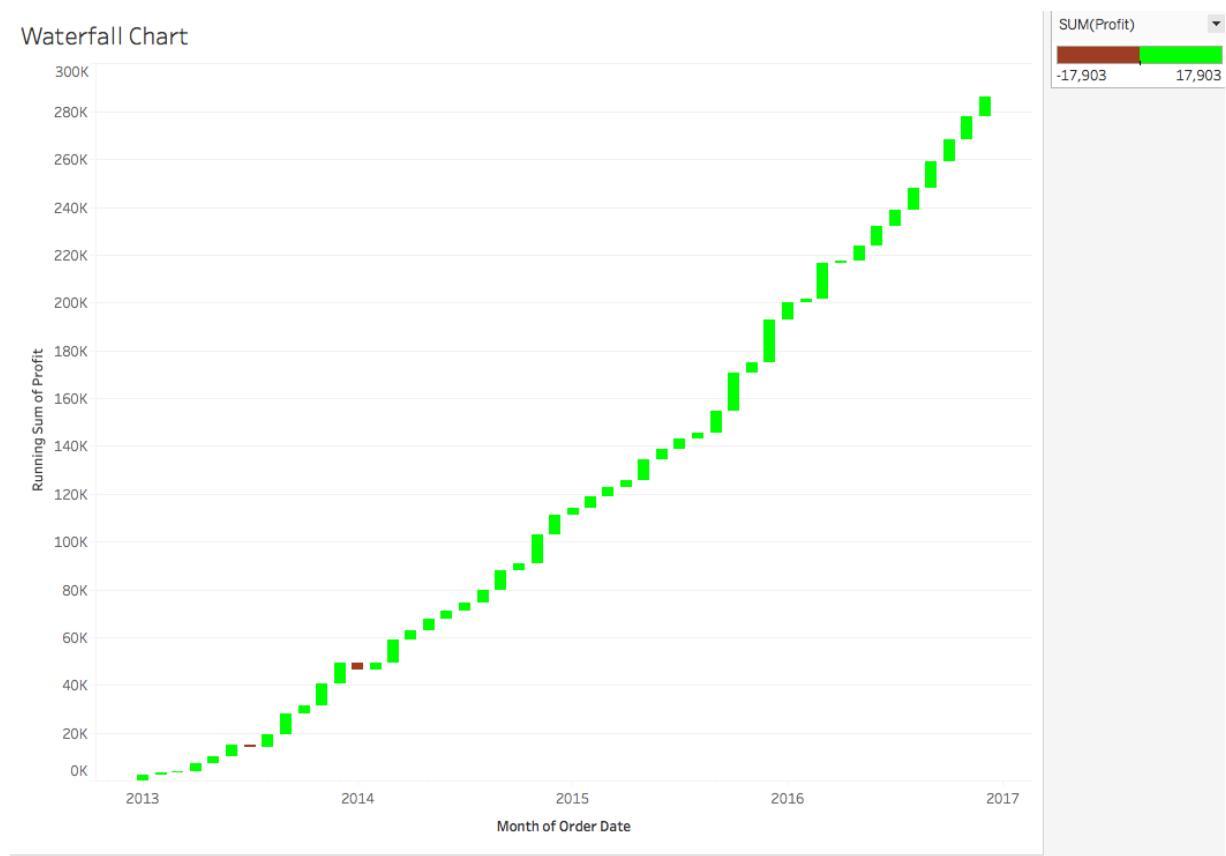
3. Choose the **second Pie Chart** in the Marks Pane, and drag every **Measure / Dimension** out of it. Reduce the size of the chart, and change the colour to white



4. To create the Dual axis, right click on the **second Pie Charts' Y – Axis**, and select the Dual Axis, to get your chart.

You must have understood by now that all the above charts, although different in their final looks, were all derived from the core graphs of the '**Show Me**' feature. But wait, its not over yet. I have more to show you.

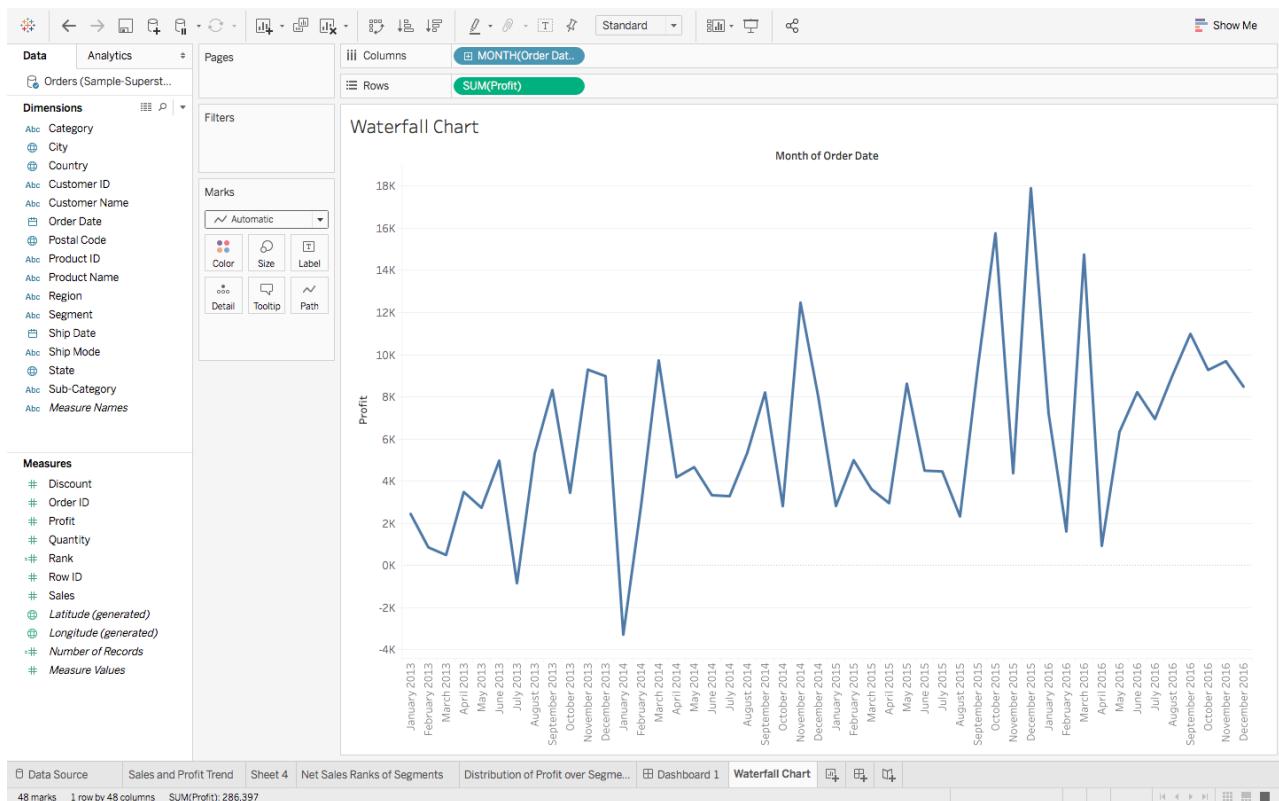
## 1.4 Waterfall Chart



A waterfall chart derives its name from its analogous orientation and flow. Here we have plotted the Running Sales of the Superstore over its years, and you can see the two small red areas in the middle of 2013 and the beginning of 2014, indicating that the Sales actually dipped and also the measure by how much.

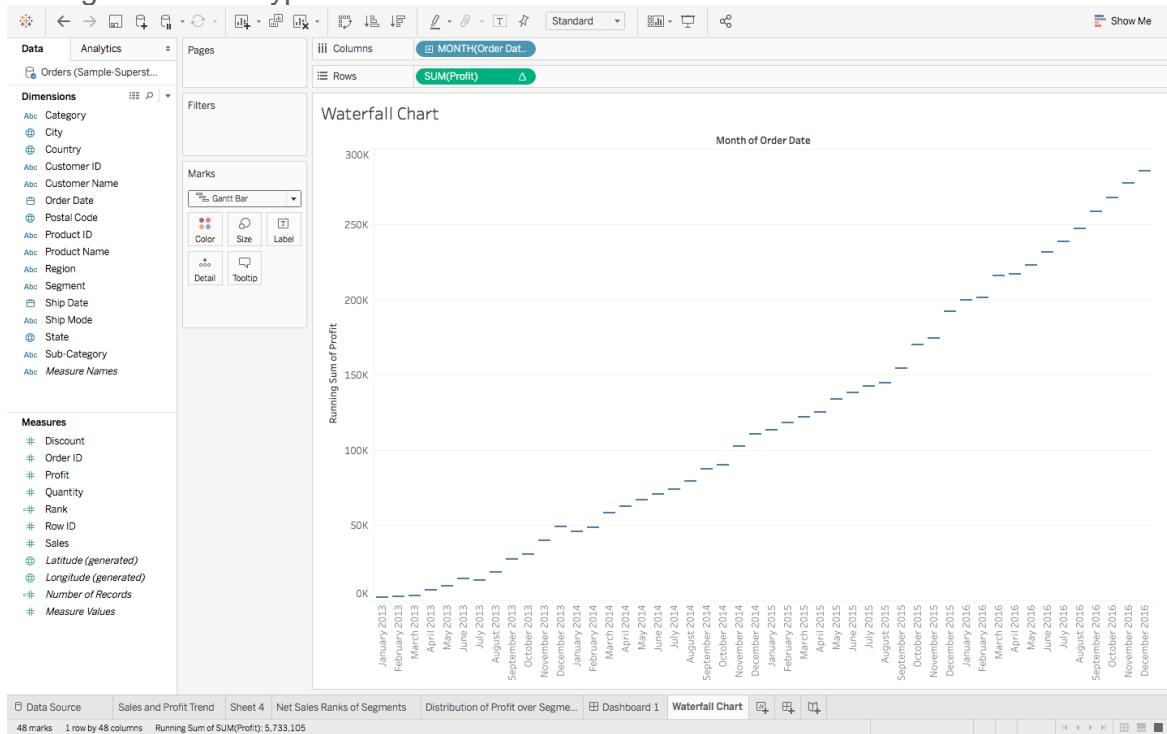
This implies that such charts are used to analyze the cumulative effect of a Measure, and see how it increases and decreases as a whole. To understand this better, let's visualize it.

A waterfall chart is a derivative of a **Line Chart**, so we will begin with this graph:



Note: Here the X-axis is **Order Date** (in Month-Year format and converted to Discrete). And the Y-axis is **Profit**.

1. Right-click on the green **Profit** Pill, and select Quick Table Calculation – > Running Total.
2. Change the Mark Type from Automatic to Gantt Bar:

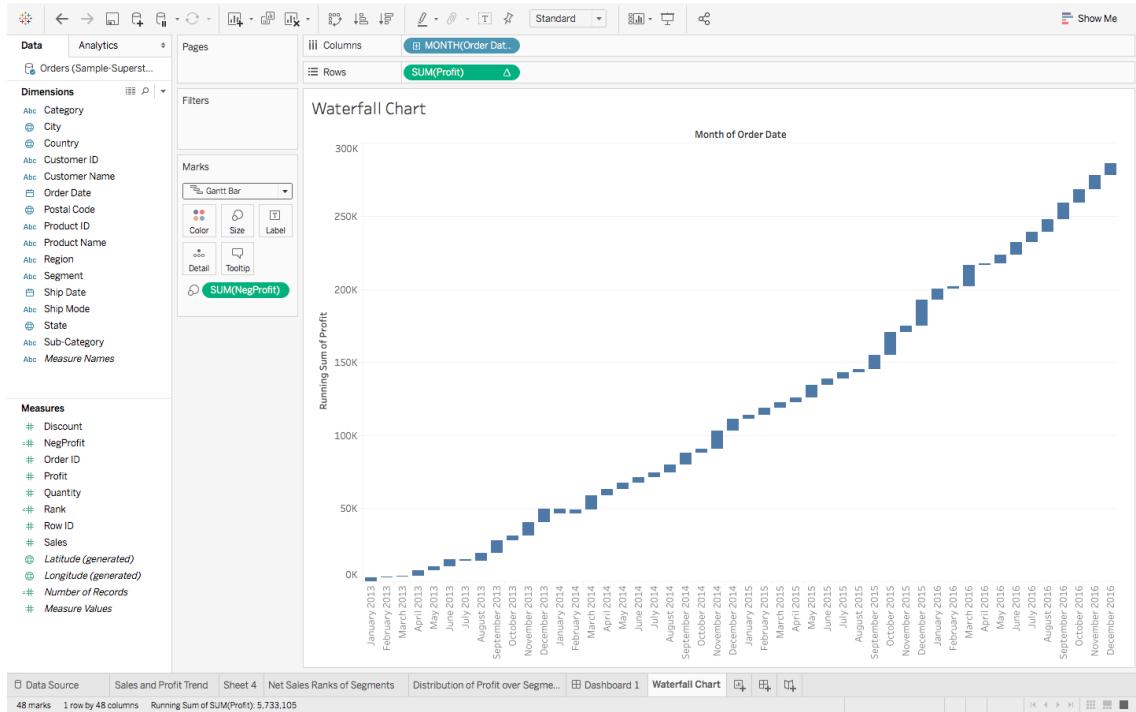


### 3. Create a Calculated Field called 'NegProfit':

The calculation is valid.

**ABS(number)**  
Returns the absolute value of the given number.  
Example:  $\text{ABS}(-7) = 7$

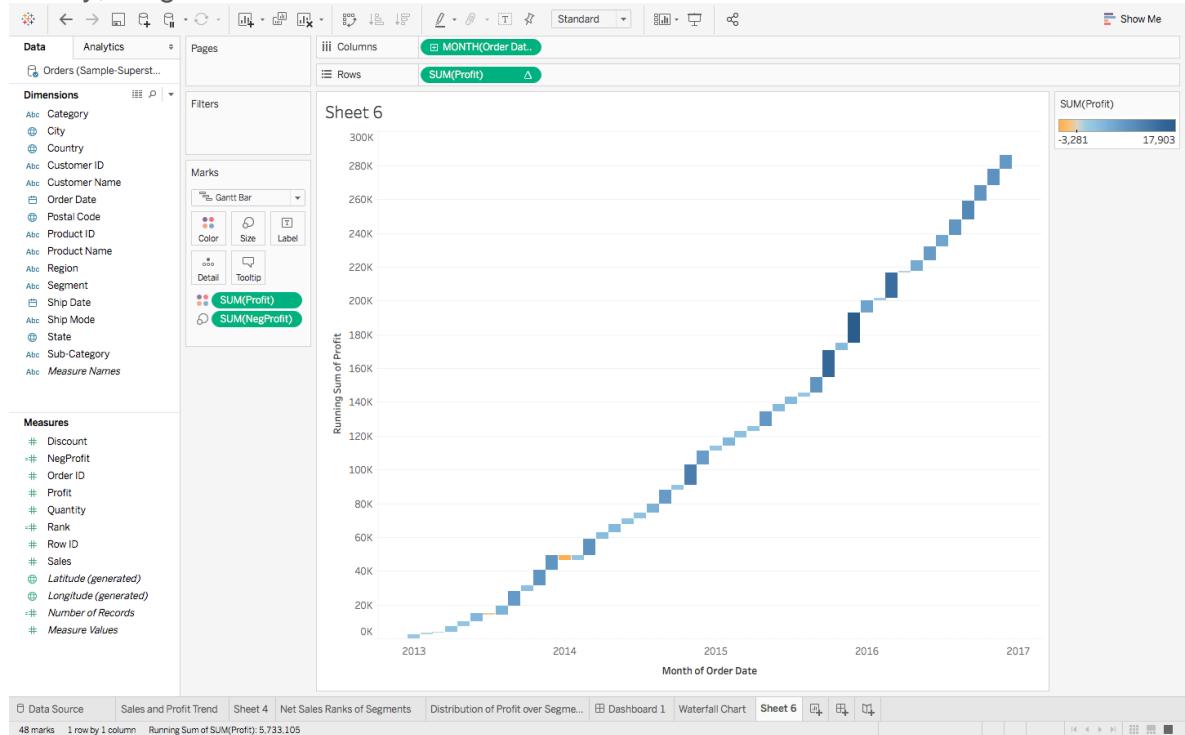
### 4. Drag this NegProfit over Size in the Marks' shelf to get:



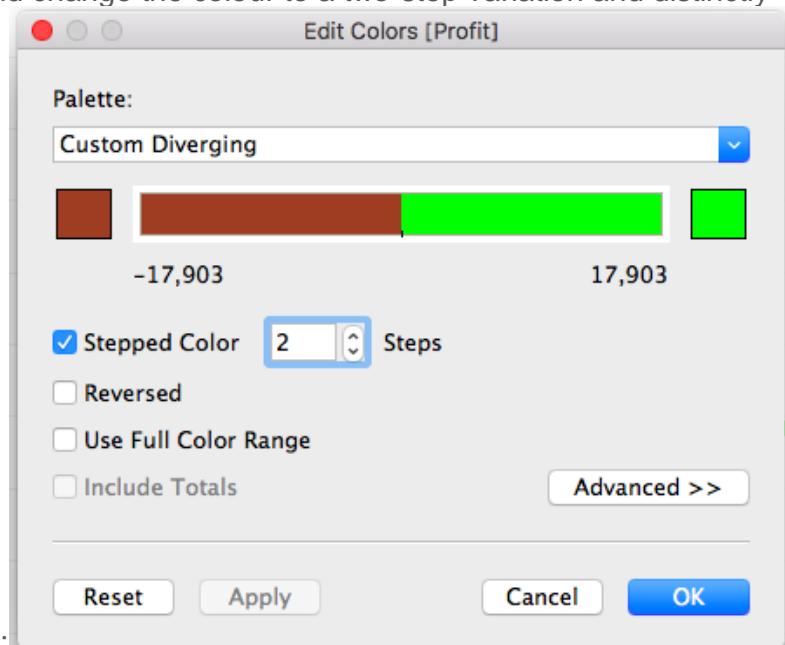
The calculated field was used to fill in the space in the Gantt Chart. A negative value in Profit would extend the bar downwards, whereas a positive one would extend it upwards.

The length of each small bar in the chart represents the amount of change in Profit from one month to the next.

5. Finally, drag Profit over to Colour :



6. You can go ahead and change the colour to a two-step variation and distinctly



view the rise and fall :

The graph that you will get could be very easily represented in the form of a Bar Chart as well. Do note that I have reversed the colors here, to make the anomalies stand out:

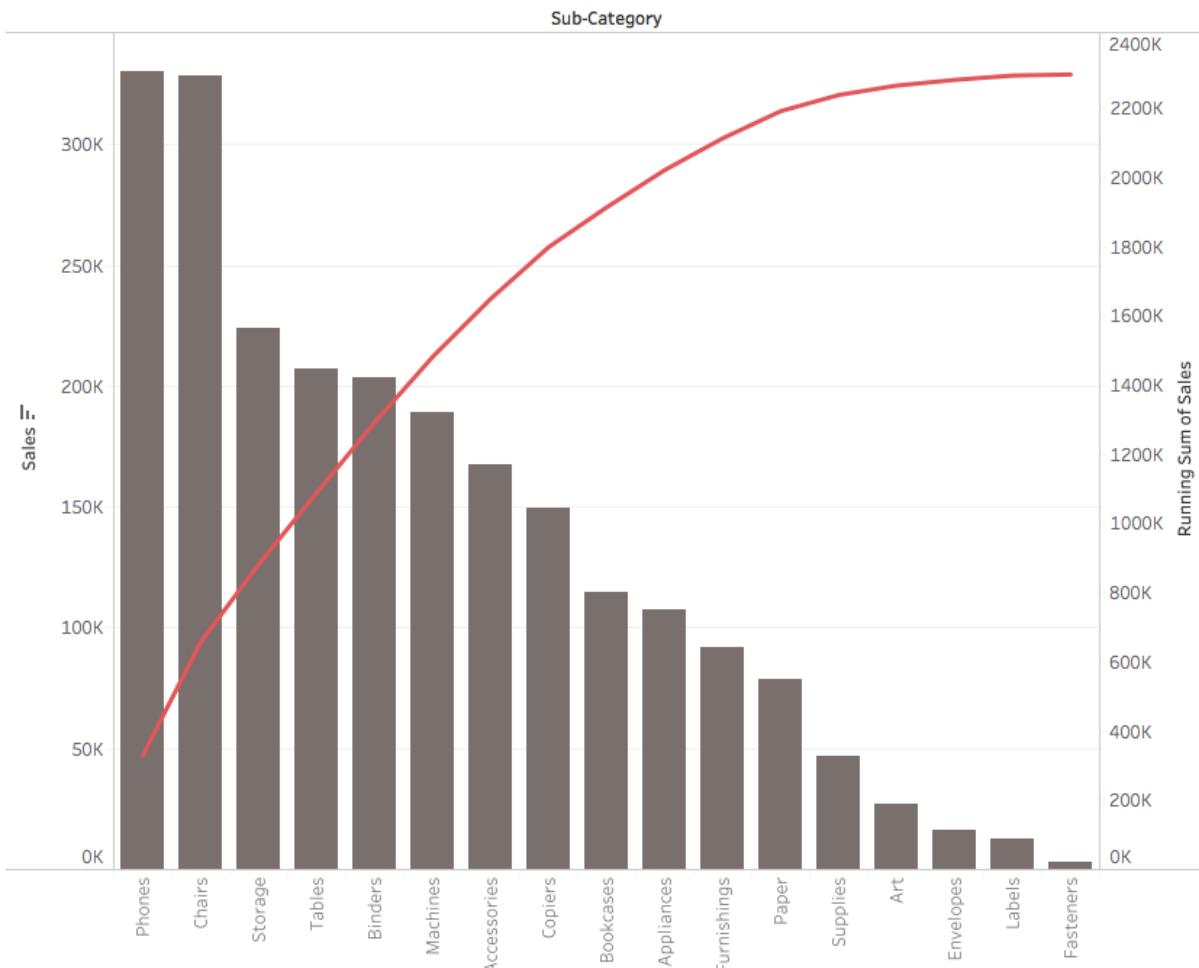


But I am sure you would agree that using a Waterfall chart was a more intuitive way of representing the data, especially to see the changes in Measures such as Sales and Profit over the years.

## 1.5 Pareto Chart

Below I have visualized a popular 80-20 principle of data analytics. If you have not heard of it, let me try and explain it with our example. It is often observed that the majority of the sales of a Superstore come from a select few products.

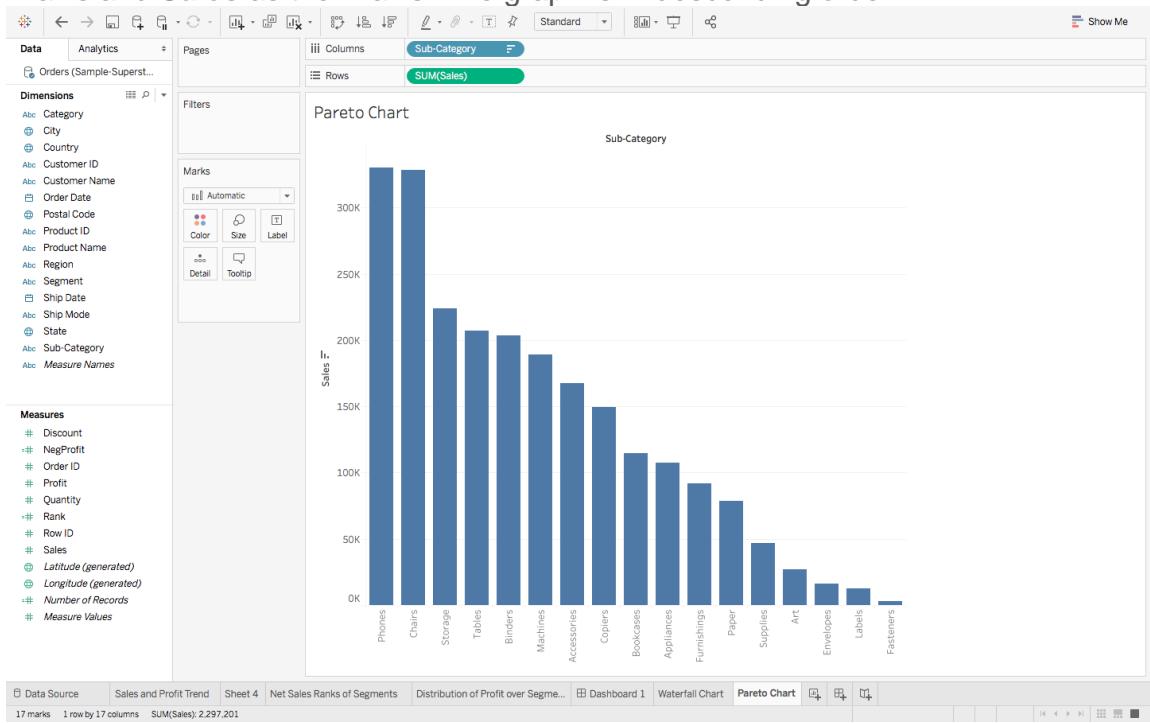
One cannot expect bread and eggs to have the same sales figures as cakes, right? This is officially termed as the 80-20 principle, meaning that 80% of the Sales come from 20% of the Products. In our Superstore, this principle can be observed in the below chart, where most of the sales are generated by Phones and Chairs :



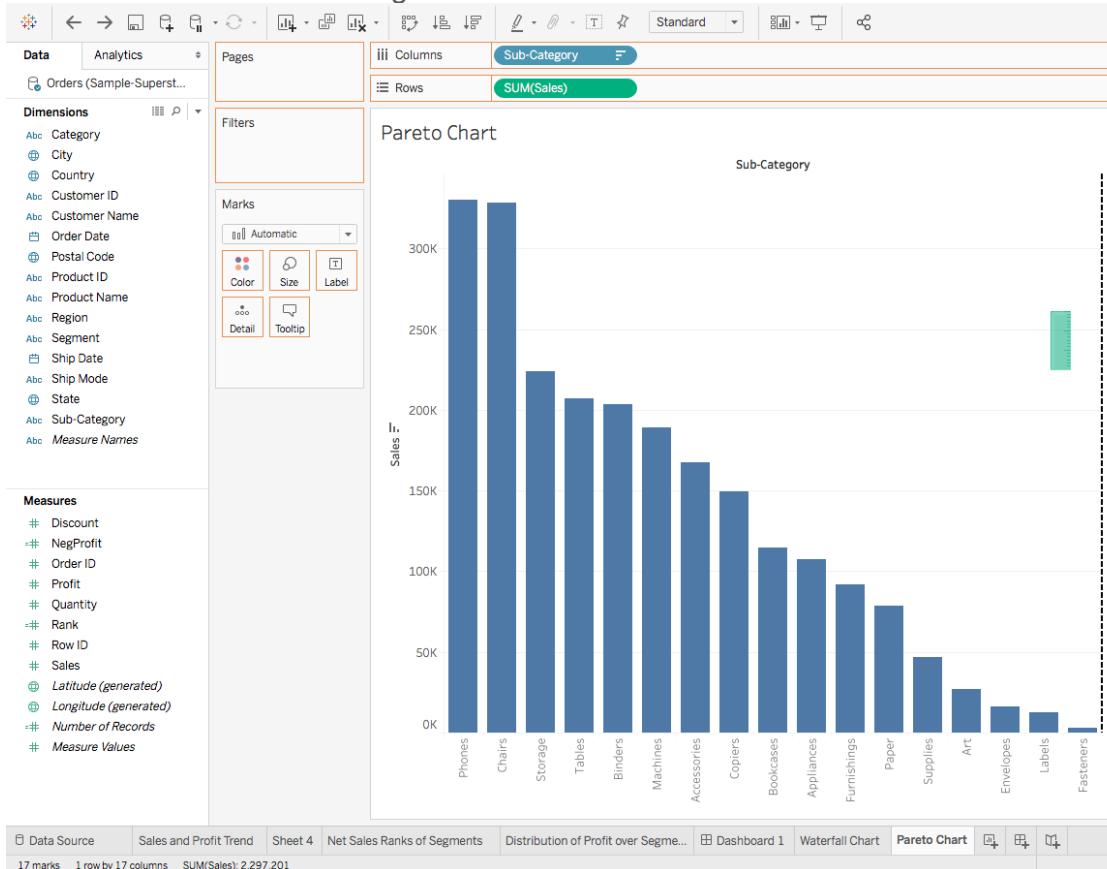
Quite a popular visualization, **Pareto charts are often used for Risk Management to determine the most common problems that are having the most negative impact on a project**; but as we will see, it can have other applications too.

Let's see how its done:

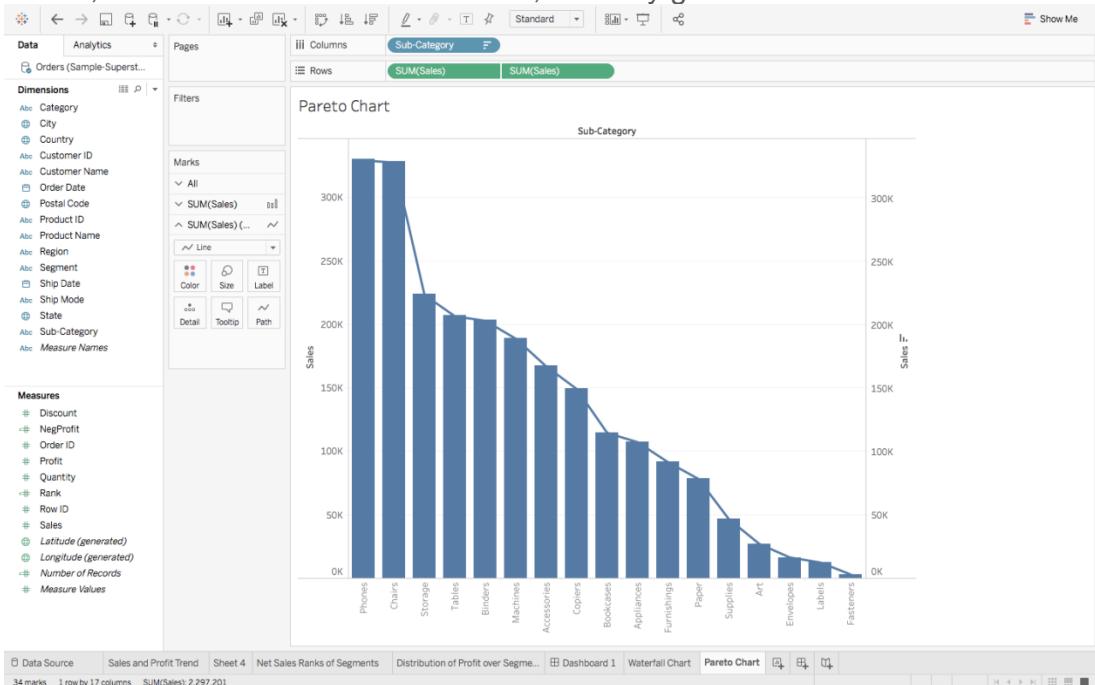
- We are going to start off with the following chart. This has **Sub Category** as the X-axis and **Sales** as the Y-axis. The graph is in descending order:



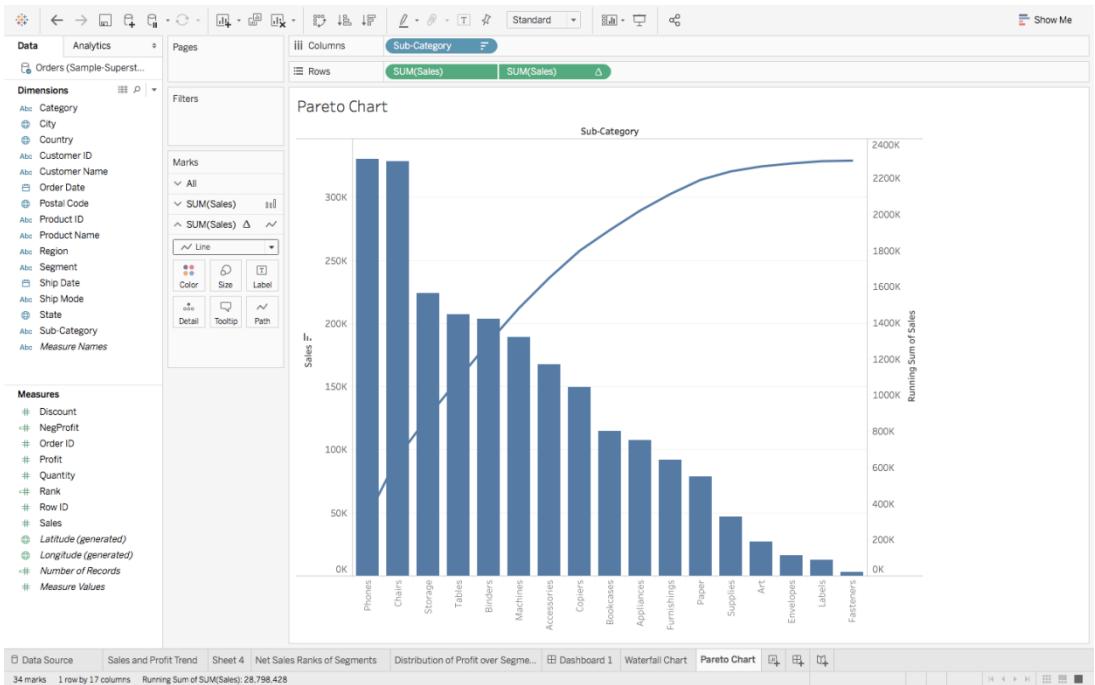
- Next, drag **Sales** over to the chart, until you see a green highlighted bar, and a dotted axis towards the far right :



3. Drop Sales here to create a **Dual Axis**. Change the Mark Type of the first chart to **Bar**, and of the second chart to **Line**, to finally get :



4. Right click on the second green **Sales Pill**, and add a **Running Total Calculation** to it:



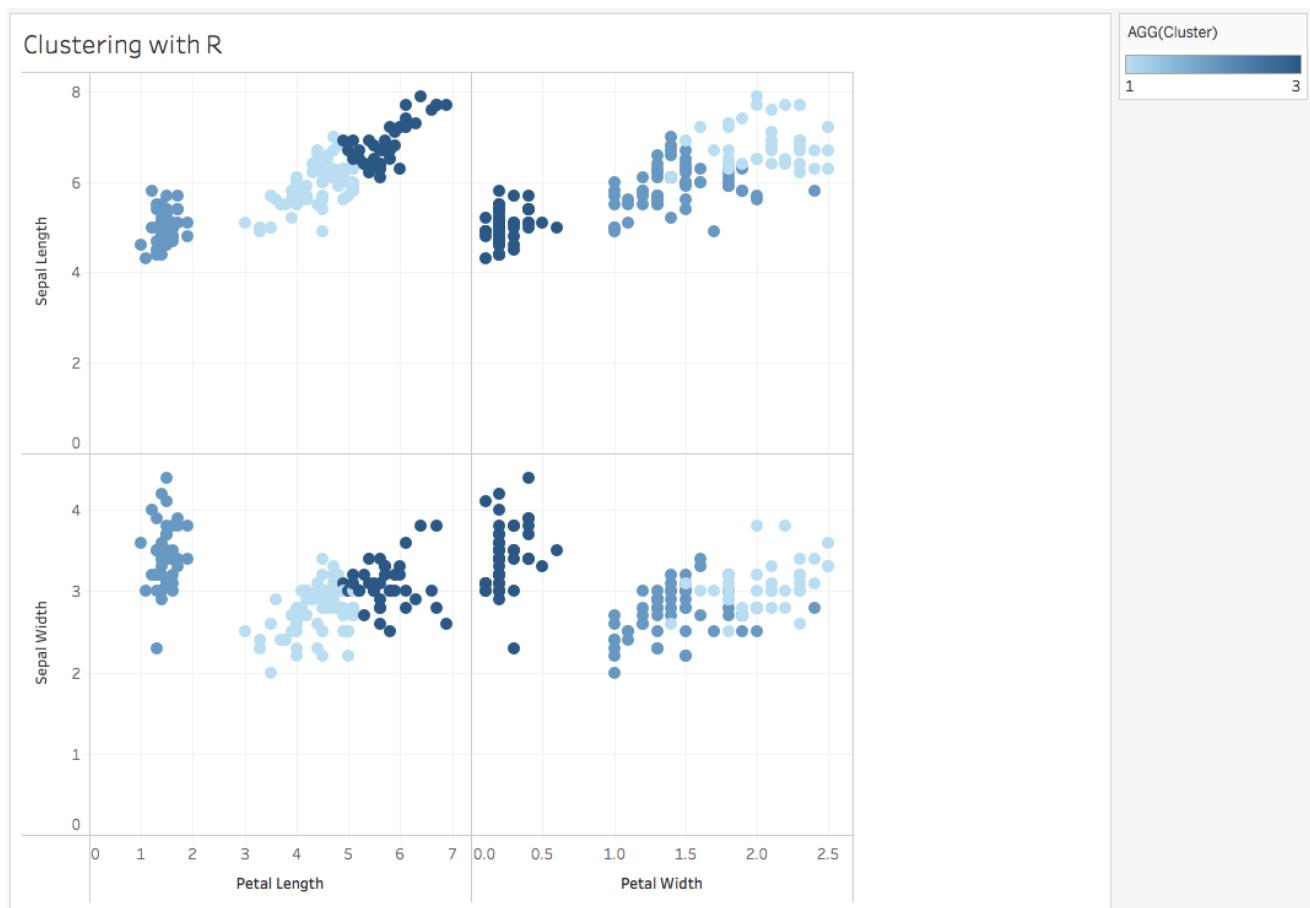
5. All that is left is to just change the colour schemes, and your Pareto Chart is ready!

## 2. Introducing R programming in Tableau

One thing I like about Tableau is that it's not just a tool meant to create pretty graphs with mere drag and drop actions. With the release of Tableau 8.1 in 2013 came a plethora of new functionalities.

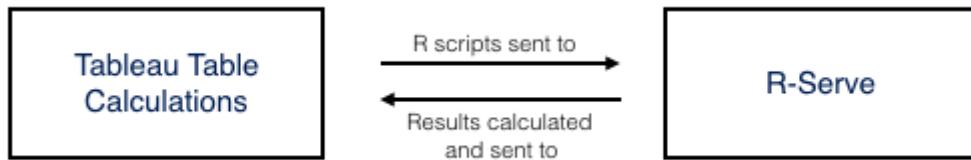
The introduction of R, to enable making richer and dynamic visualizations, was one of the predominant features. R can be used with Tableau for techniques like Clustering, Prediction and Forecasting, to name a few.

I wanted to start the exploration of R and Tableau through Clustering, so I used the ultra popular [Iris Dataset](#). It contains different features to distinguish between 3 types of flowers, namely Virginica, Setosa and Versicolor. As you can see in the below image, the R integration quite easily creates clusters of these 3 species:



Interested in making this yourself? First let's go through the basics and the installation process, before delving into the visualization!

The following depicts the flow of control between Tableau and R to make this integration possible:



**R scripts** are written in Tableau as **Table Calculations**, which are sent to the **R serve** package of R. Here the module carries out the necessary computations and **returns the result** to Tableau.

**Note:** To properly understand and thereby use this feature, you must possess some knowledge of R and its various syntaxes. For the same you may refer to the following tutorial:

[Learn Data Science in R from scratch](#)

Now let's look at the steps for this integration:

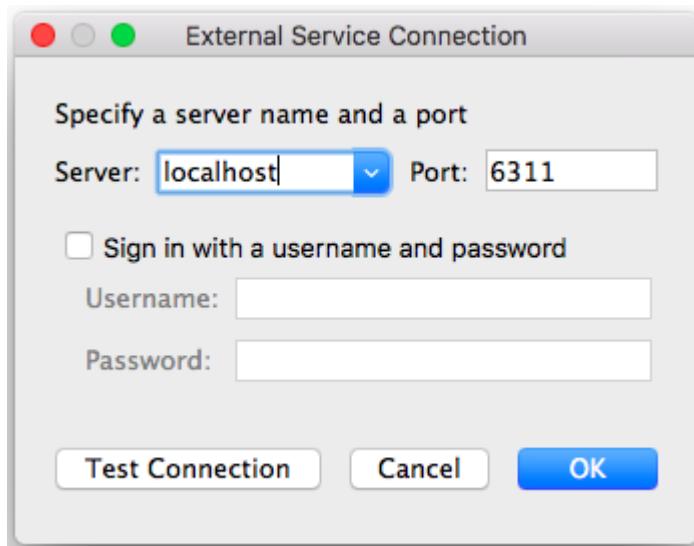
1.

1. [Install R](#)
2. Install the Rserve package
  - Run the following in the R command line:

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

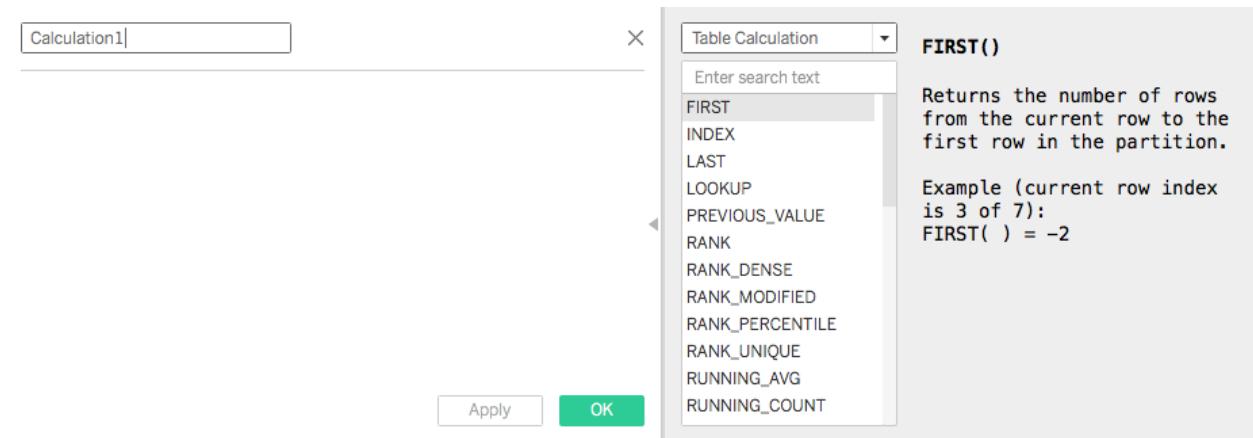
3. Configure Tableau to run in R

- Open Tableau – > Help – > Settings and Performance – > Manage R / External connection. Fill in the fields with the following default information and select Test Connection:



So now that you have the proper ingredients ready, let's start cooking!

As was shown in the image above, you make use of Tableau's Table Calculation to communicate with R :



If you scroll down the list of functions, you will come across the following four:

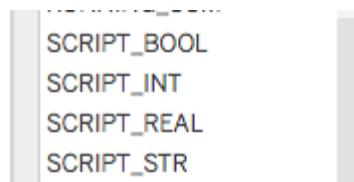
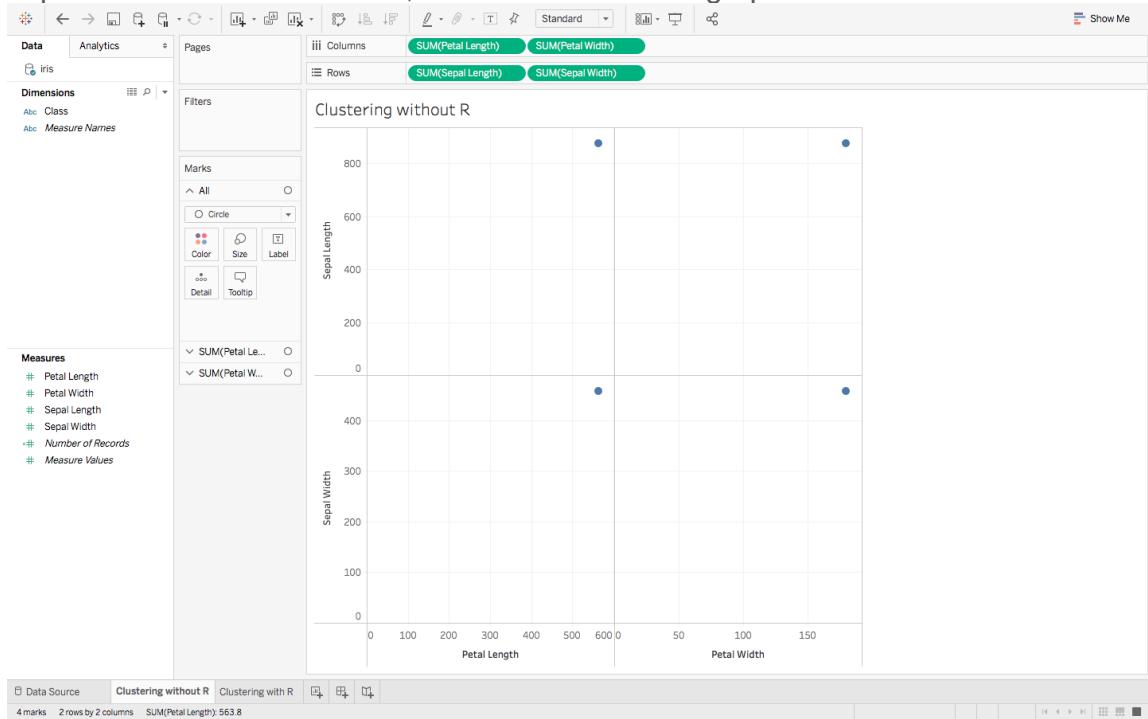


Tableau automatically understands that the script is meant for R when these functions are included in the calculation area.

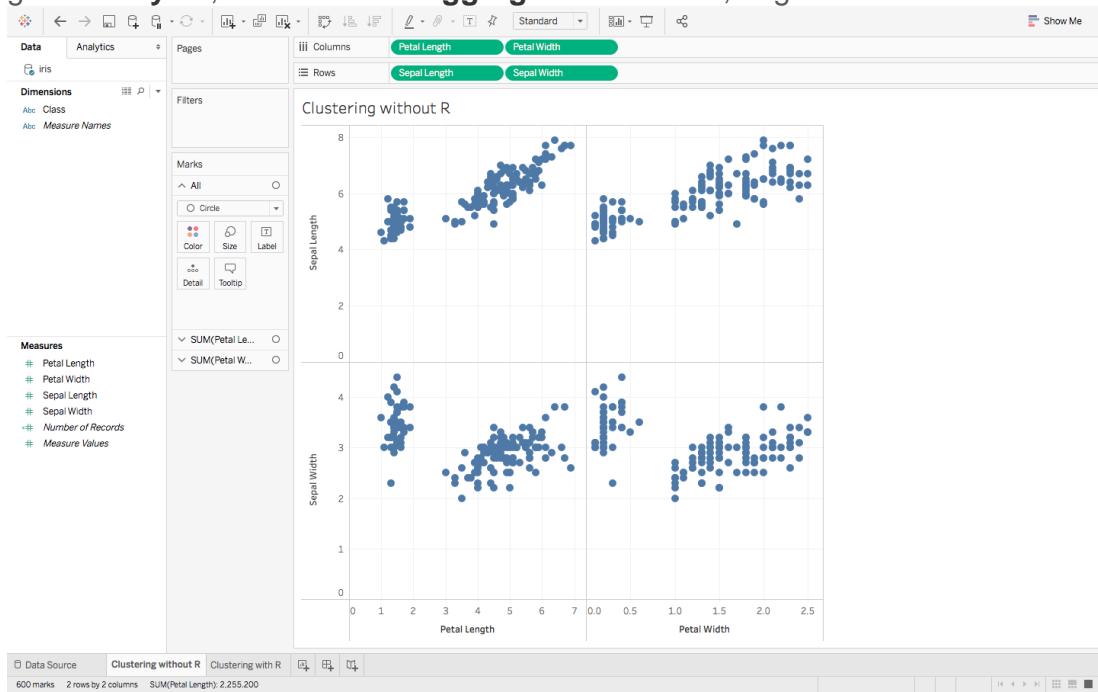
I hope that your initial excitement of making the clusters is still there! Let's proceed.

1. Download the Iris Dataset from [here](#).

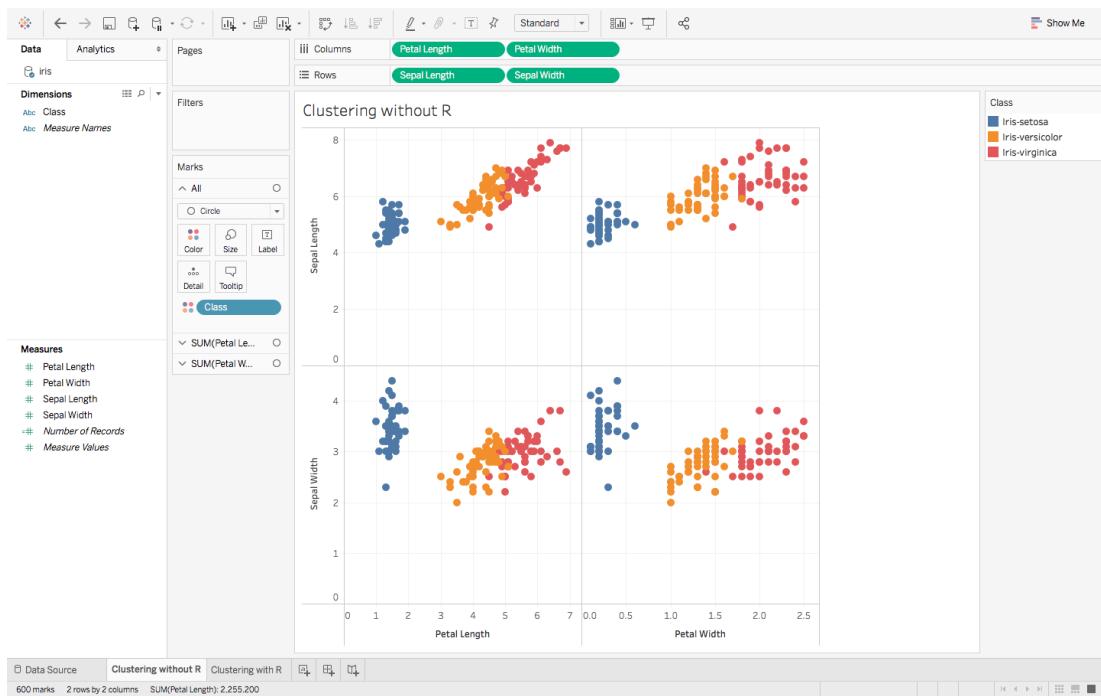
2. Import the dataset in Tableau, and make the below graph:



3. Here you are getting the Sum across different Measures. To get discrete values, go to **Analysis**, and uncheck **Aggregate Measures**, to get:



4. Finally, to form the clusters, drag the **Class** Dimension over **Color** in the **Marks** Pane:



What we have above is a Scatter Plot, which shows clusters of data points divided into 3 distinct clusters.

Let's try doing the same with R now, and compare the two visualizations that we will get. We will be using the most common clustering algorithm, K-Means:

1. Begin with the same scatter plot as **point 2** above.
2. Create a new Calculated Field and fill it with the following:

Cluster

```
SCRIPT_INT(
'result <- kmeans(data.frame(.arg1,.arg2,.arg3,.arg4), 3);result$cluster;',
SUM([Petal Length]), SUM([Petal Width]),SUM([Sepal Length]),SUM([Sepal Width]))'
```

The calculation is valid.

All

Enter search text

- SIGN
- SIN
- SIZE
- SPACE
- SPLIT
- SQRT
- SQUARE
- STARTSWITH
- STDEV
- STDEVP
- STR
- SUM**

Default Table Calculation

Apply    OK

For clarity, the above Calculation is :

3. **SCRIPT\_INT(**

4. **'result <- kmeans(data.frame(.arg1,.arg2,.arg3,.arg4), 3);result\$cluster;'**

```
SUM([Petal length]), SUM([Petal width]),SUM([Sepal length]),SUM([Sepal width]))
```

- Finally drag the newly formed Field **Cluster** to Color in the Marks Pane, to get your clusters ready!

Although there are a few overlaps, the two visualisations do appear to be quite accurate.

This was a small gist of the potential of integrating R with Tableau. Its applications are limitless, and I am sure you must have already started to think of the different ways you can interact with it.

## End Notes

It would be naive of me to say that this is all there is to Tableau. As new versions roll in, so do new functionalities.

Not only that, people are always experimenting and exploring Tableau, and coming up with new visuals. There are multiple blogs where people publish their experiments with data too. Do check them out.

You can also find new and gorgeous visualizations weekly on [Tableau's official Gallery page](#). I would definitely advise you to keep referring to these posts, creating your own visuals, and sharing it with the community.

Stay creative and all the best on your journey as a Data Explorer!