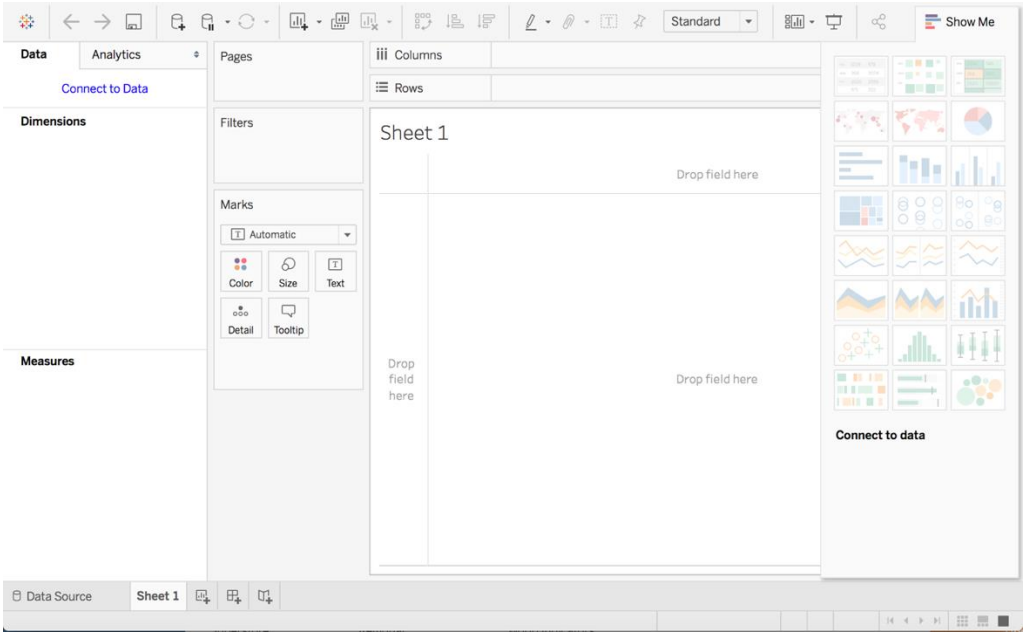
Data Visualization with Tableau

1 Introduce of Tableau

1.1 What is Tableau?

Tableau is a Business Intelligence Tool used for data visualization that enables you to create interactive and apt visualizations in form of dashboards and worksheets to gain business insights. It connects easily to several data sources, and allows for rapid insight by transforming data into dashboards that look amazing. Dashboard creation can be done with simple drag and drop of data fields, which make it the possible to pick up right away and start exploring and building dashboards. The advanced capabilities require experience and training, and one can learn more about Tableau from here: https://www.tableau.com/learn/training

1.2 Tableau Interface



- Above is the sheet of Tableau 10.2 in OS X, which is where we can build visualizations. We'll go over the various areas of the screen.
- The menus are not shown in this screenshot because of the layout on a Mac. The menus contain many useful and powerful options that we'll use in the future.
- The logo button on the very top left can bring us back to the start experience, where we can connect to safe databases, open recent workbooks, etc.
- The very left of the sheet is the data pane, which lists all open data sources, and are broken out to dimensions and measures of a chosen dataset below.
- If we click on the analytics tab, we can bring out pieces of the analysis directly as drag and drop elements.

- New sheets, dashboards and stories can be added by clicking the plus button on the bottom of the sheet.
- Finally, the most important part is shelves on the right, which can be built directly from the data pane by drag and drop on to the canvas, or onto the shelves. On the screenshot, it contains the columns and rows shelf, pages shelf, filters shelf and marks card. Legends such for colors, size and shape will automatically be generated, and could be customized and removed.

2 Creating Visualizations with Tableau

Before we start creating data visualizations with Tableau, we first need to figure out who is our audience, what is the flow of the report, and what kinds of layout should be used. We'll used a case study later to show these parts step-by-step.

2.1 Who is my audience?

Here we need to consider about several aspects. First, who is the one that make the decisions and what decisions do they make? People with different role will have different requirements for a report. Second, how comfortable are they with data? Some high level managers are not familiar with very specific techniques and data, but your senior colleague will be very comfortable with data. Third, how will they use the information and how much time do they have? When they only need a very general result, keep the repot short and highlight the important parts. Finally, how familiar are they with the key performance metrics? An introduction of the metrics maybe helpful.

2.2 Flow

In general, we'll put the most important and essential images on the top left.

2.3 Layout

The page should be well organized with just the right details. The color and typography is also important, for example, you can use high contrast color scheme to highlight and bold key words. Also, choose the right charts for representing clearly about what you want to show. Add proper legends and descriptions to clarify any ambiguity.

2.4 Case Study – Sample Superstore

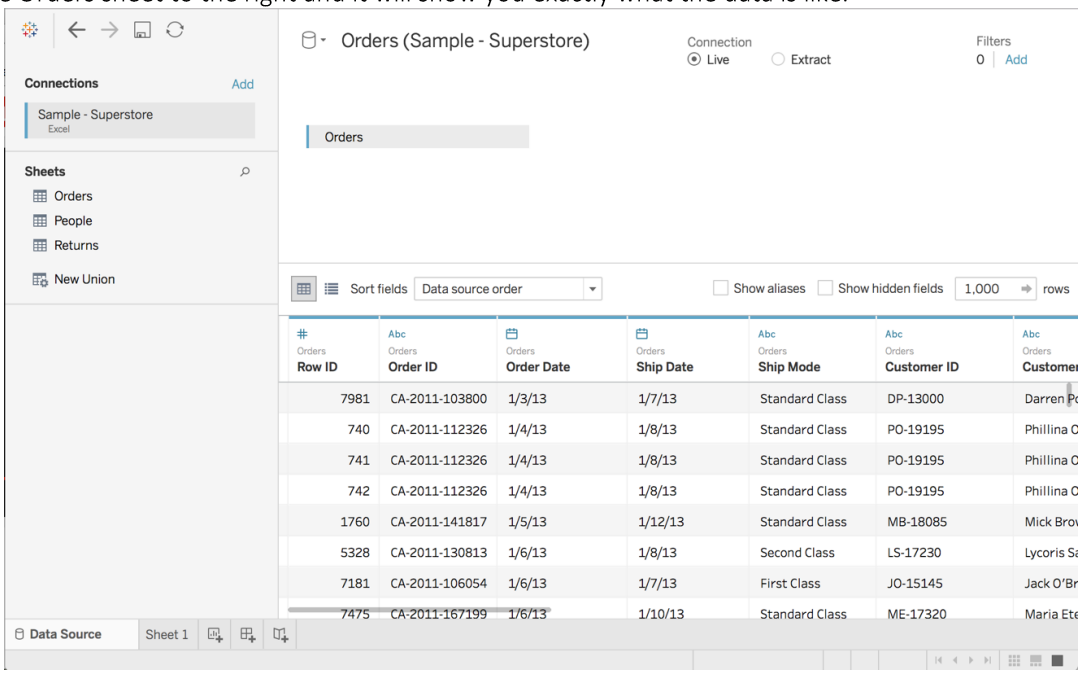
In this section, we'll focus on show you how to do visualization with Tableau step-by-step with Sample Superstore case study. This will cover the basic and essential usages of Tableau.

3 Cast Study

3.1 Sample Superstore Over View

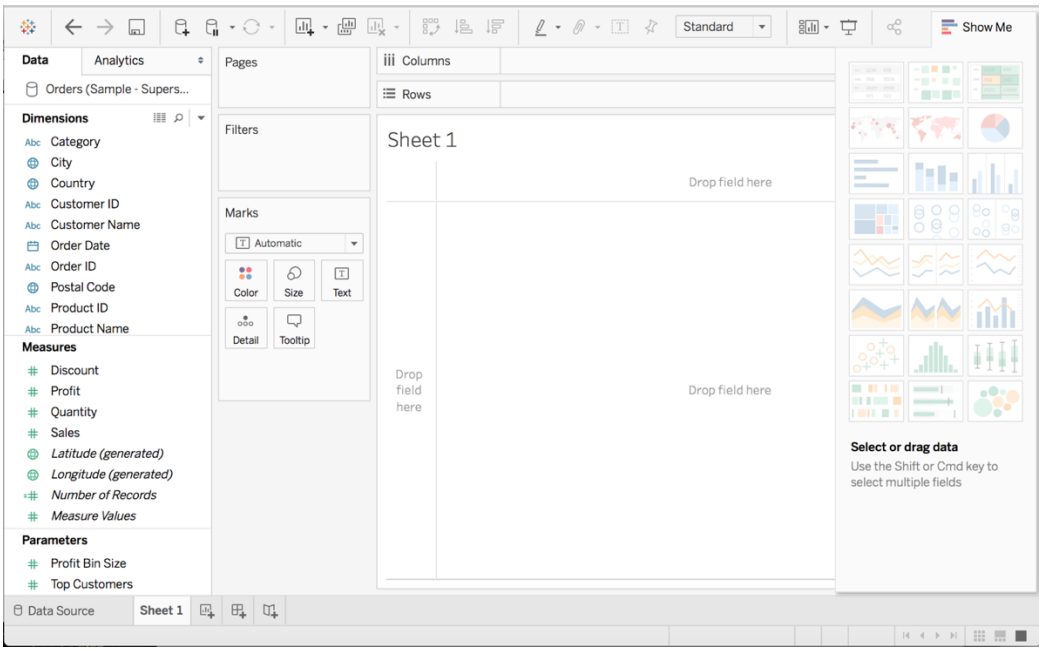
The Superstore dataset is the sample dataset come, which is very good for us to start with. Starting from the home page, we can choose to connect a server or file. In this Superstore case, just click the excel button under connect to file section and choose Sample - Superstore dataset to connect. You can also download the dataset from here:
https://community.tableau.com/docs/DOC-1236.
Then click on open and wait the query.

The Superstore dataset contains 3 sheets, orders, People and Returns. We'll use Orders sheet today. So simply drag the Orders sheet to the right and it will show you exactly what the data is like.



To go to the worksheet, click on Sheet 1 on the bottom and the worksheet is as following. In the Data pane, there are two parts: Dimensions and Measures. After connected to the dataset, Tableau will assign any fields that contain discrete categorical information (for example, fields where the values are strings or Boolean values) to the Dimensions area, and assign any fields that contain quantitative, numerical information (that is, fields where the values are numbers) to the Measures area in Data pane. Any fields you drag from the Dimensions area will initially be discrete when you add it to a view, with a blue background. And after you drag a field from the Measures area to Rows or Columns, Tableau creates a continuous axis. After you drag a dimension to Rows or Columns, you can change the field to a measure just by clicking the field and covert to Measure. Any field that you drag from the Measures area will initially be continuous when you add it to the view, and so its background will be green. But if you then click the field and choose Discrete, the values become column headers.

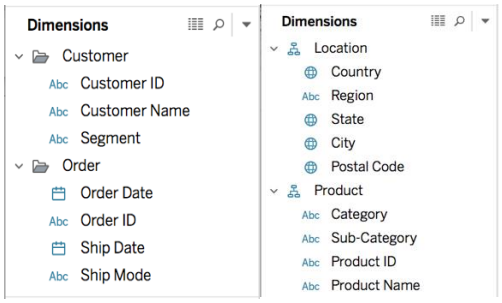
Several things that we need to pay attention: there is a Measure Names variable on the bottom of Dimensions, also Latitude, Longitude, Number of Records, Measure Values on the bottom of Measures. These variables are not from the original dataset, which are added automatically after loading the dataset. We'll talk about the usage of theses variables later in detail.



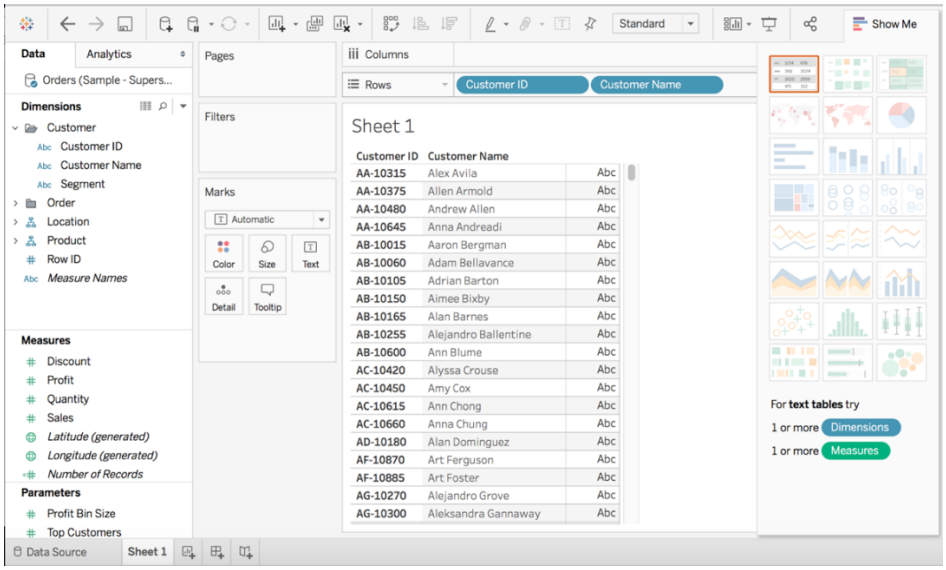
3.2 Group your variables

To make the relationships between the variables more clearly, a good way is to group the variables or make a hierarchy. In the Dimensions of our Superstore data, we can find that there are variables related to the orders, the ones about customers, etc. For example, we can put the ones related to customers in one group, which are Customer ID, Customer Name and Segment. Click on the little triangle appeared after Customer Name, choose Folder under Group by, which turns the Folders option form grey to black. Then we can click on Folders to create a folder called Customer. After that, this folder will appear in Dimensions as following. And then simply drag Customer ID and Segment to the Customer folder. Also Create another Order folder with Order Date, Order ID, Ship Date and Ship Mode with the same method as we just showed.

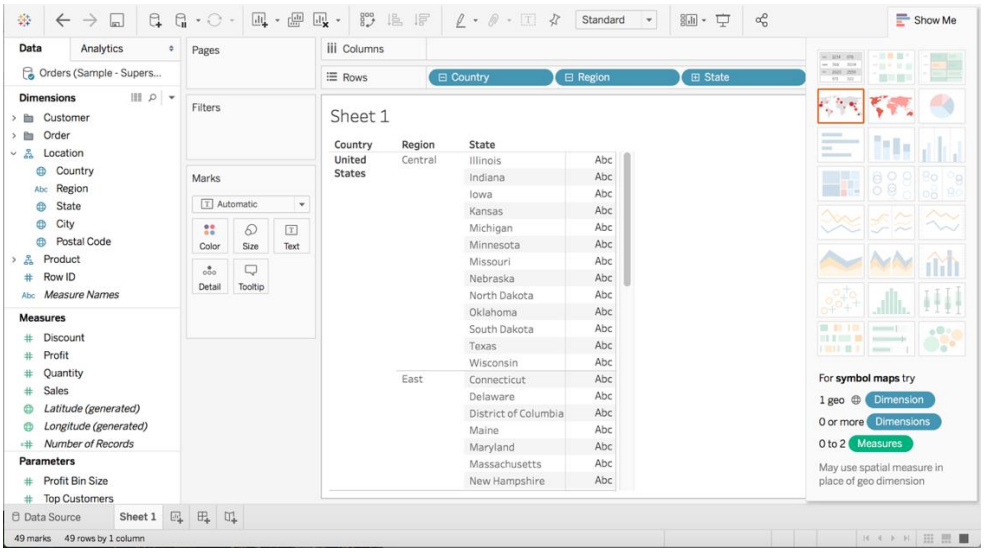
To create a hierarchy, choose create hierarchy from the hierarchy option. We'll put Country, Region, State, City, Postal Code into Location hierarchy with this specific order. The Category variable and subcategory variable would fit a hierarchy very well. So we create another hierarchy called product with these variables in order: Category, Sub-Category, Product ID and Product Name.



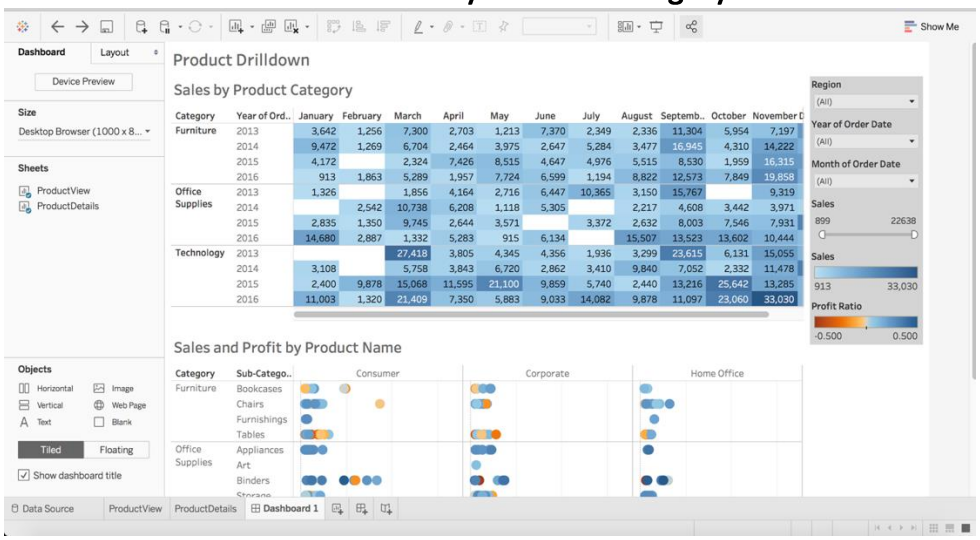
For the difference between group and hierarchy, if we drag Customer ID in Customer group to Rows, then drag Customer Name to Rows after Customer ID, there will be two rows in Sheet 1 as following.



Then clear sheet by clicking the clear sheet button on the top. If we drag Country in Location hierarchy to Rows, click the plus button on Country button, which will give you Region after Country Row. You can also skip one of the variables by dragging the variable away, like you can drag Region away and only leave the Country and State variables. The Sheet is as following.



3.3 Dashboard: Product Drilldown – Sales by Product Category

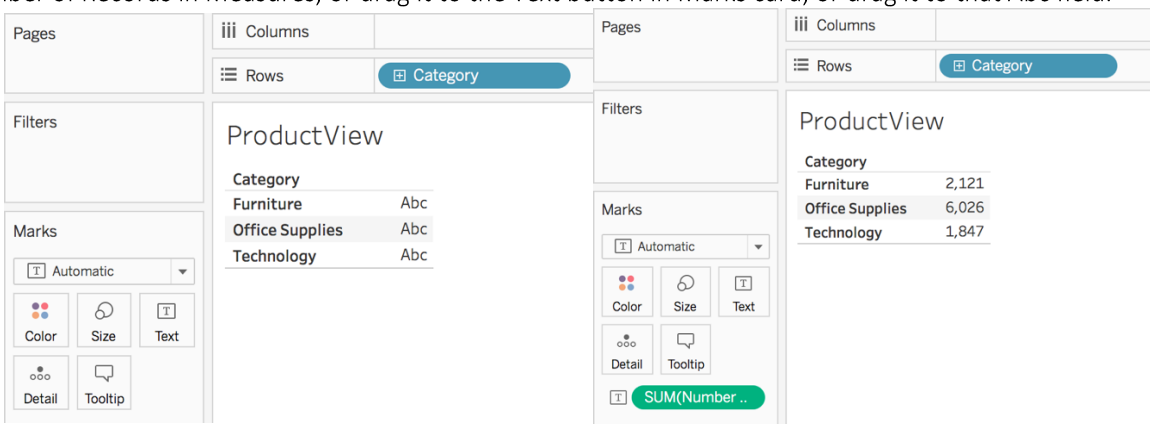


This is the dashboard we'll have in the end, and we'll need a ProductView worksheet and a ProductDetails worksheet. So we'll start with the ProductView worksheet, which is the graph on the top.

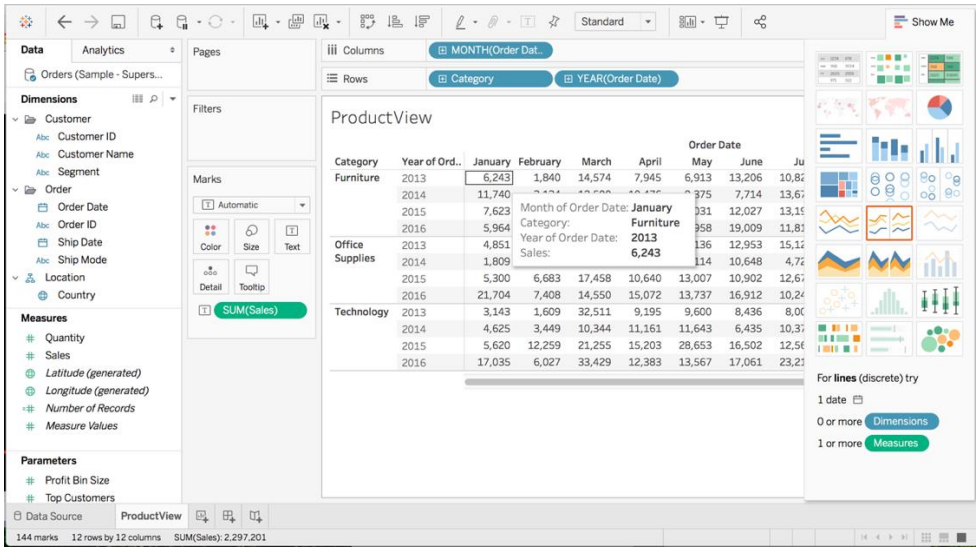
3.4 ProductView worksheet

In ProductView worksheet, the first level row is Category and the second level row is year. The columns are months from January to December. If you stop the hove over one of the entry, a small pop-up box contains Profit and Sales. Profit is not shown in the table, but Sales is shown by the number of each entry.

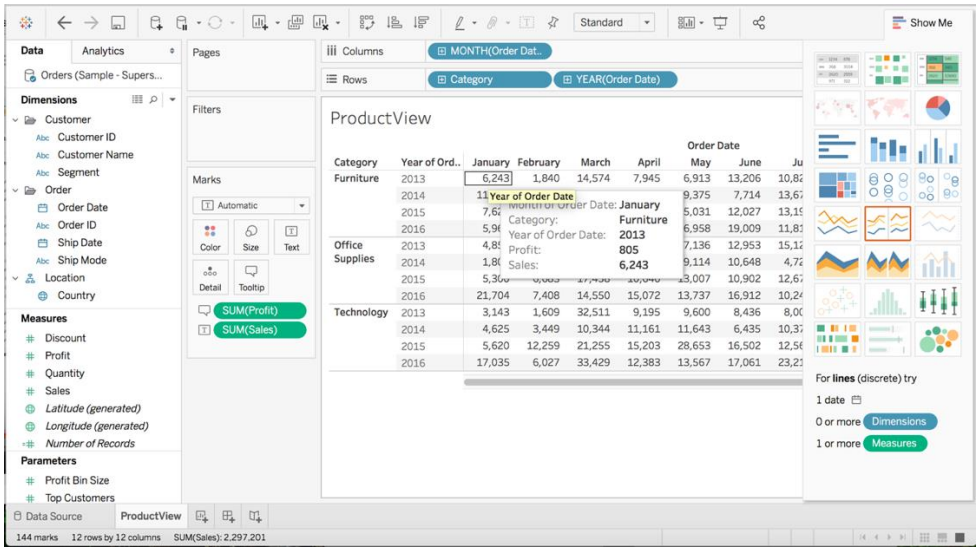
So we should put Category from Product hierarchy to Row shelve. The Abc shown after furniture is matched with Text in Marks card. For example, suppose we want to add Number of Records. You can double-click Number of Records in Measures, or drag it to the Text button in Marks card, or drag it to that Abc field.



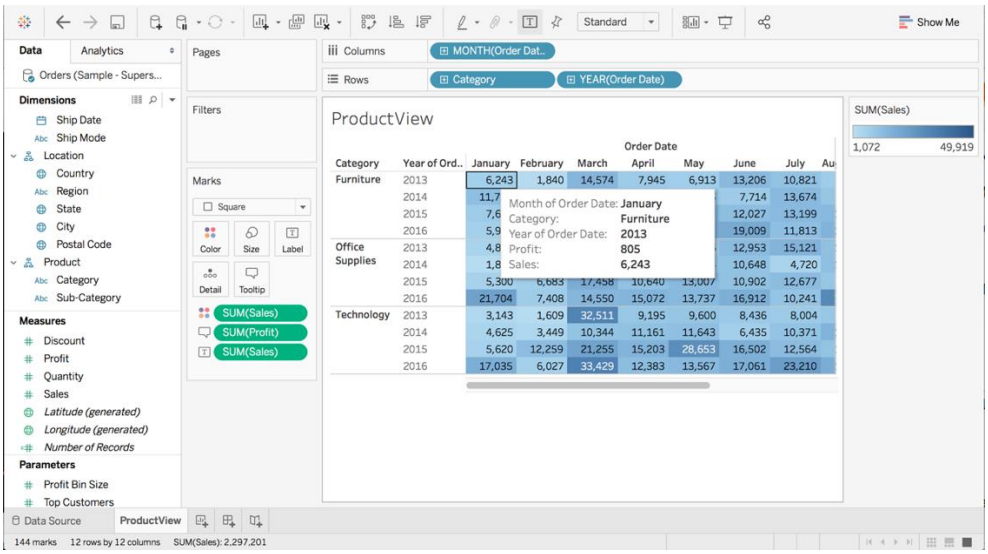
Then put Order Date after Category, and it will show the year of order. For columns, we also need to put Order Date, but this time we need to change the default Year to month by clicking the small triangle and choose Month. Sales will be the number in the table, so we simply drag sales Measure to the table.



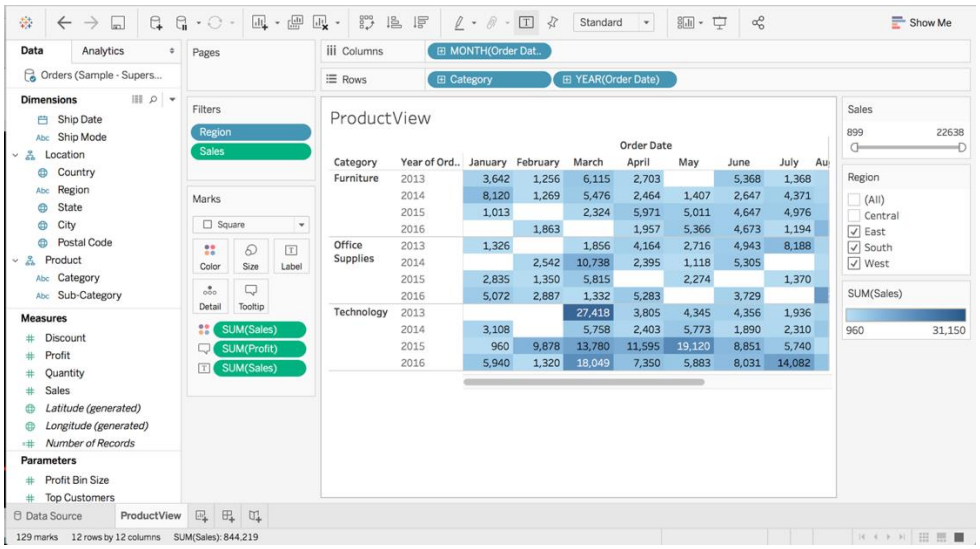
Now when we hove the cursor over the number, the pop-up box doesn't contain profit. The Tooltip in Markers card will add info to the table without change the table. So we can drag profit to Tooltip.



We also hope to use color to indicate the value of sales. Drag Sales to color option in Markers card and edit the color scheme. Here we change the automatic to square so that the number itself is not colored but the background.



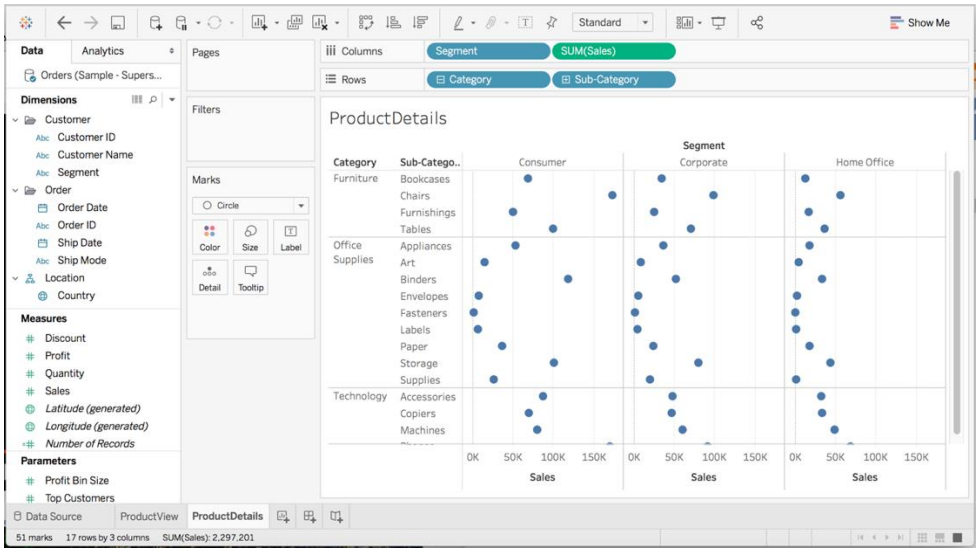
To apply a filter, drag the Region to filter shelve and select all the regions. Then you can right click the region in filter and choose show filter, so that the filters will show on the right. Now you can choose the specific region you want. For continuous variables, it can be added as filter with range of values.



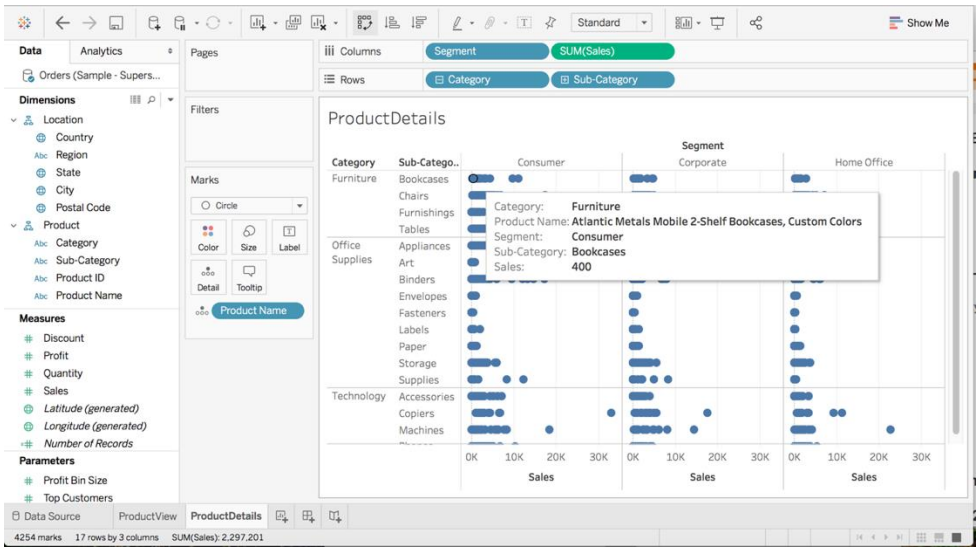
3.5 ProductDetails works

In ProductDetails worksheet, the first level row is Category and the second level row is Sub-Category. The column shelve is Segments. The pop-up box contains Profit, Sales and Profit Ratio.

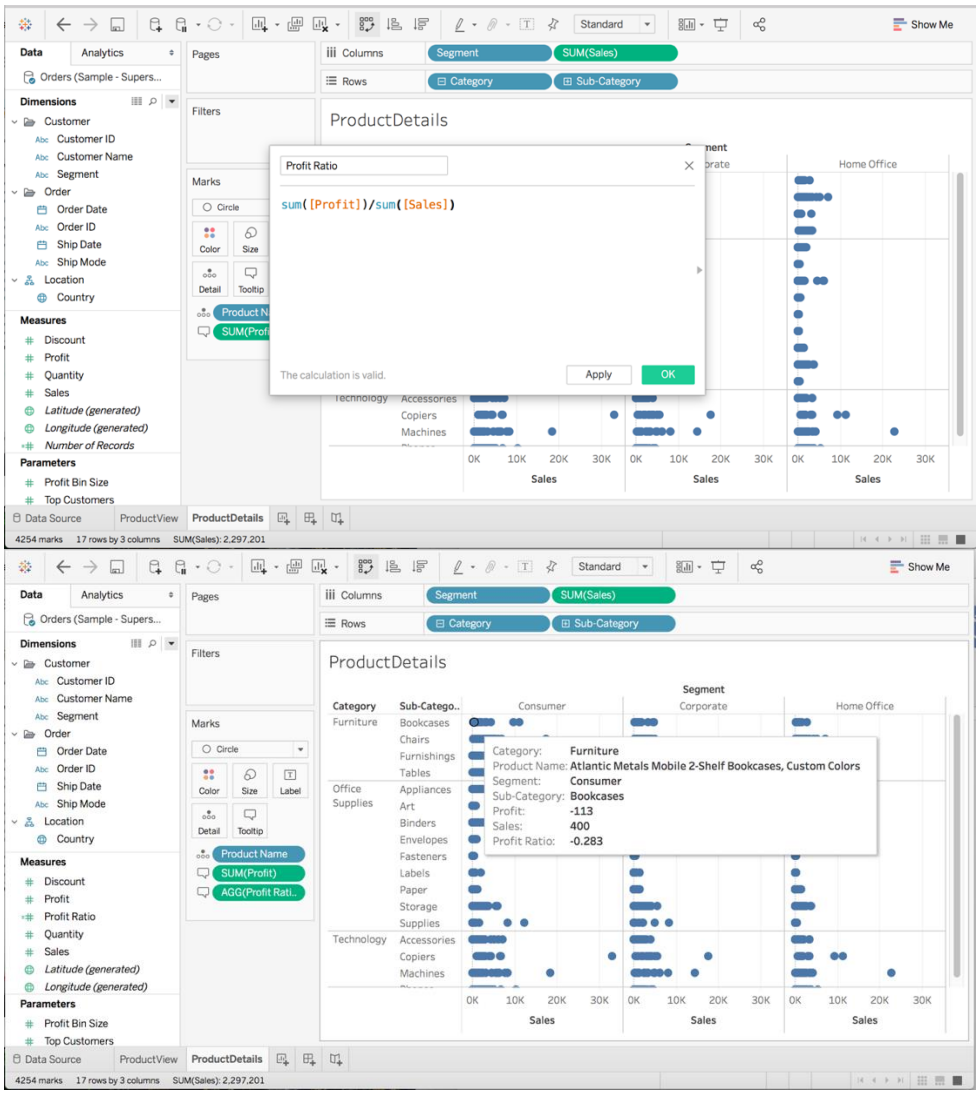
First, create the basic row shelve and column shelve by putting Category and Sub-Category to Row, Segment to Columns, and Sales to Text. But at this time we don't have the round points indicating the sales values. We will drag the Sales in Marks card to Columns, so it will take sales as a dimension. To change the bar table to circle, click on automatic in Marks card and choose circle.



But now another dimension, Product Name is still missing. It appears in info box but not in row or columns. So we'll drag Product Name to Detail in Marks card, which will not change the row and column, but only separate the value in each entry.

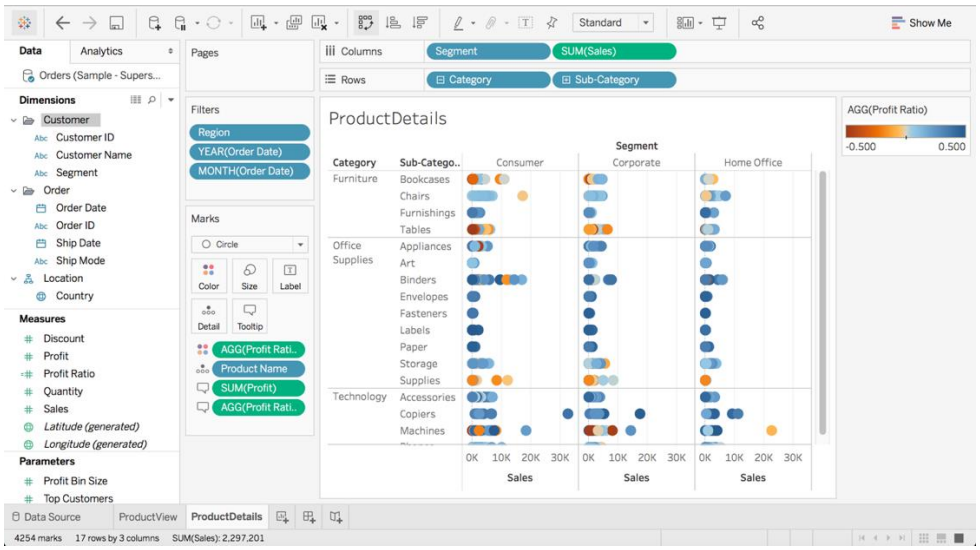


As in ProductView worksheet, drag Profit and Sales to Tooltip to add them to the pop-up box.
For Profit Ratio, we need to create a new variable in Dimensions by clicking create calculated field and name it as Profit Ratio. Then drag it to Tooltip to put it in the info box.



Also change the color of the points by dragging Profit Ratio to color button. To make the color palette symmetric, click on advanced and set the start with -0.5 and end with 0.5. The size can also be changed by size button.

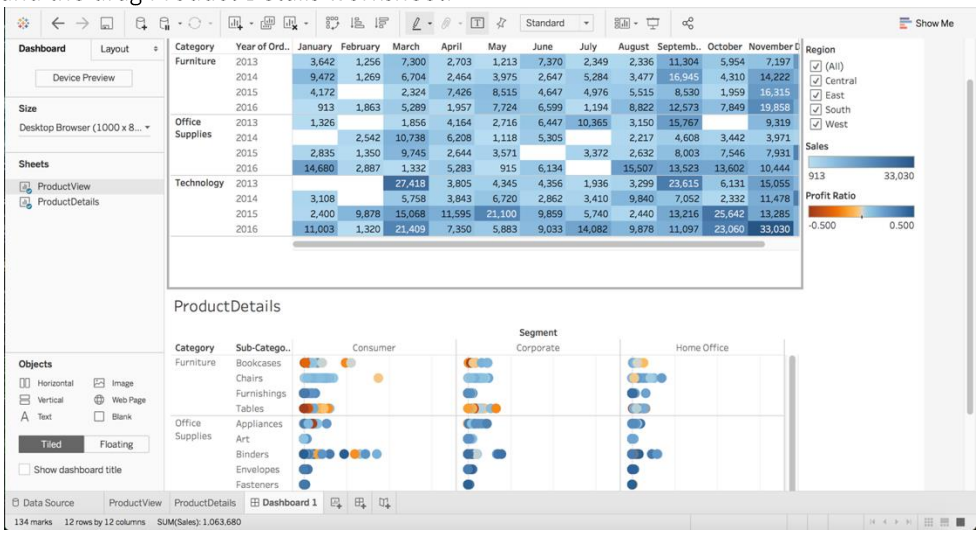
You can still add filters like Region, Order Date. The final table is as following.



contains exactly *K* objects with that feature, wherein each draw is either a success or a failure. In contrast, the binomial distribution describes the probability of *k* successes in *n* draws with replacement.

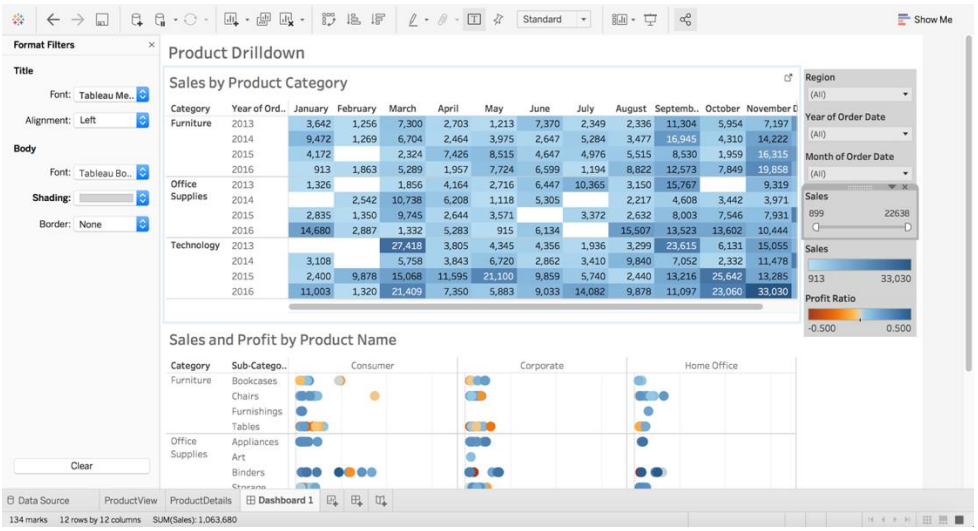
3.6 Product Dashboard

Now we have finished building the two worksheets we need, we can create a new dashboard by clicking the new dashboard button on the bottom. You can change the dashboard size. Drag the ProductView worksheet to the right and the drag Product Details worksheet.

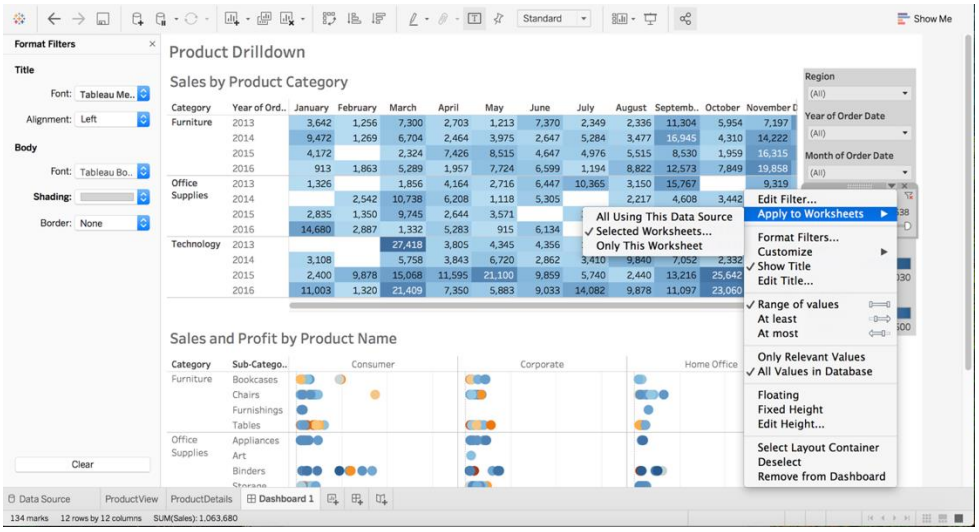


We'll change the dashboard title by first show the dashboard title, then change the default name (the name of the dashboard) to Product Drilldown. Similar for the name of Productview and ProductDetails worksheet.

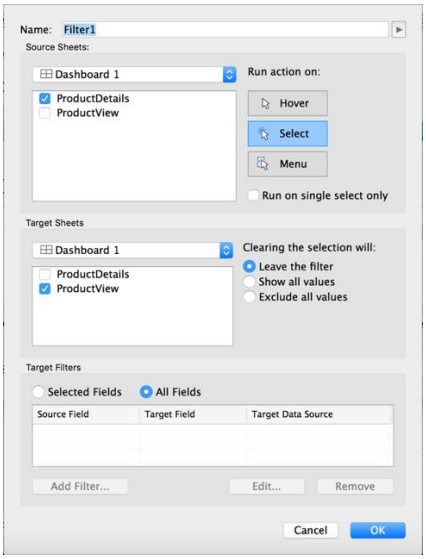
The field name for columns are hided, and the space between the two worksheet are decreased. Also you can change the look of the filters to single drop down. For the filters that are not shown, you can click the small triangle button to add certain filters. The filters format can also be changed, like add light grey shade to them.



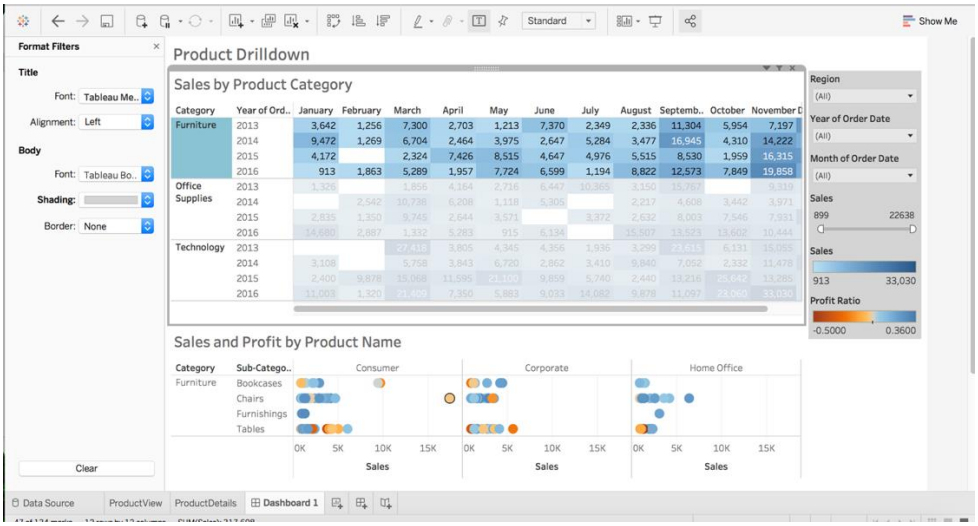
The filters are now only working for one worksheet, and we can make it work for both of the worksheets. Click on the small triangle and change only this worksheet to selected worksheets.



Now we hope that when clicking on furniture in the first table, the second one can also only show the furniture. Click the Dashboard option on the top and choose action, add action for filter. Leave ProductDetails in source sheets and ProductView in target sheets.

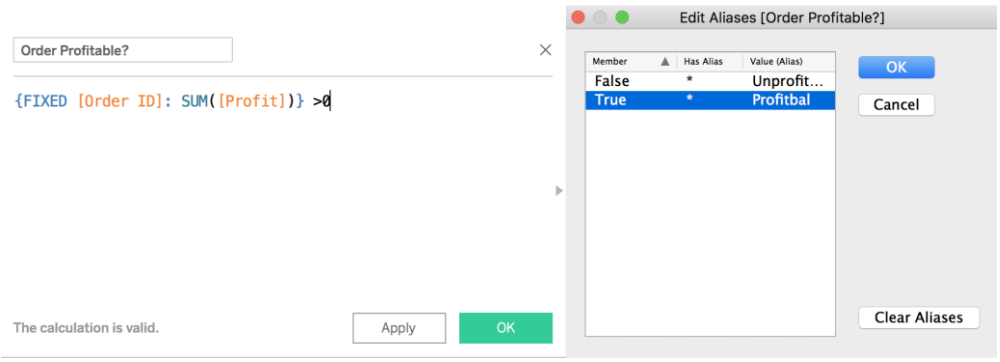


Then when you select furniture in ProductDetails, this filter will also apply for ProductView. Also add a ProductView to ProductDetails action.



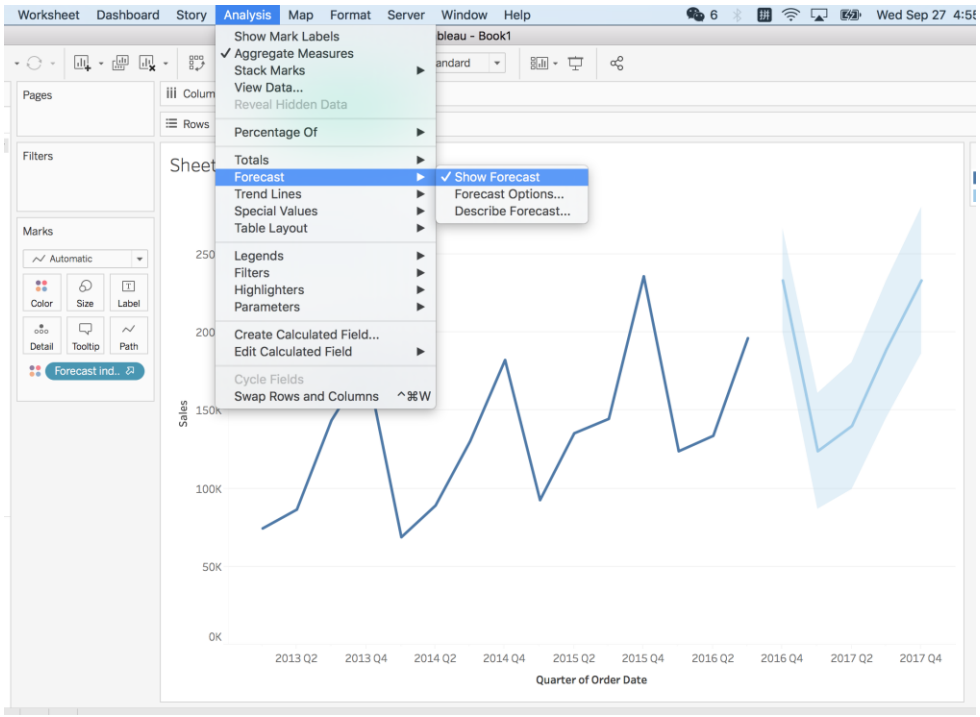
3.7 Create Dimension Code

Now, we will know how to create a dimension code. For example, we want to get a result that show the category is profitable or not. We can click “Dimensions”, choose “Create Calculate Field”, then, give a name and make code. After that, we can beautify the format. Click the dimension that we just created, and choose “Aliases”. Double click any member, and can change the “Value” (Name). I change False to Unprofitable, and True to Profitable in here.

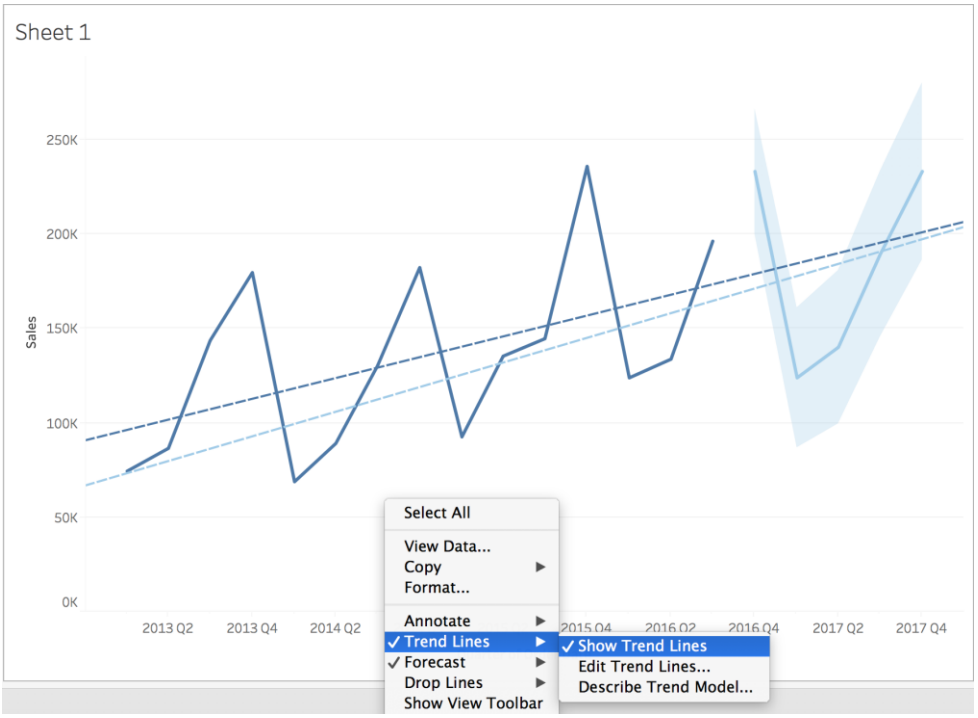


3.8 Make a Prediction of Sales

We can make a straightforward prediction in Tableau. For this part, we practice with our Super Store data, and we try to predict the sales value. After we drop Order data to Column and choose Sales for Label, we can click "Analysis", and choose "Forecast" to see the prediction by one click.



After that, we can see the trend lines in Tableau. Right click in the graph, and choose "Show Trend Lines" in "Trend Lines". If you choose any trend lines, you can see the details about statistics result.


Sheet 1
250K
200K
150K
100K
50K
0K
Sales
2013 Q2
2013 Q4
2014 Q2
2015 Q4
2016 Q2
2016 Q4
2017 Q2
2017 Q4
Select All
View Data...
Copy
Format...
Annotate
Trend Lines
Forecast
Drop Lines
Show View Toolbar
Show Trend Lines
Edit Trend Lines...
Describe Trend Model...