Hong Kong Baptist University

Department of Computer Science

COMP 7990 Principles and Practices of data analytics (2022-23)

Lab 3b: Data Visualization using Tableau

Introduction

What is data visualization?

Data Visualization is one of the most important part of data analysis. You can present the data in an understandable and visually appealing format. E.g. pictorial forms: graphs, bars or diagrams.

What is Tableau?

Tableau is one of the most powerful data visualizations tool which is focused on business intelligence. Many organizations use Tableau to get more value from their data and make better decisions. We can query relational databases, OLAP cubes, cloud databases, and spreadsheets and then generates different graphs or charts by using Tableau. It is fast and easy, and it can work with hundreds of data sources. It provides interactive data visualizations and dashboards to the users. (A dashboard is a visual display of the most important information needed to achieve one or more objectives)

Why Tableau?

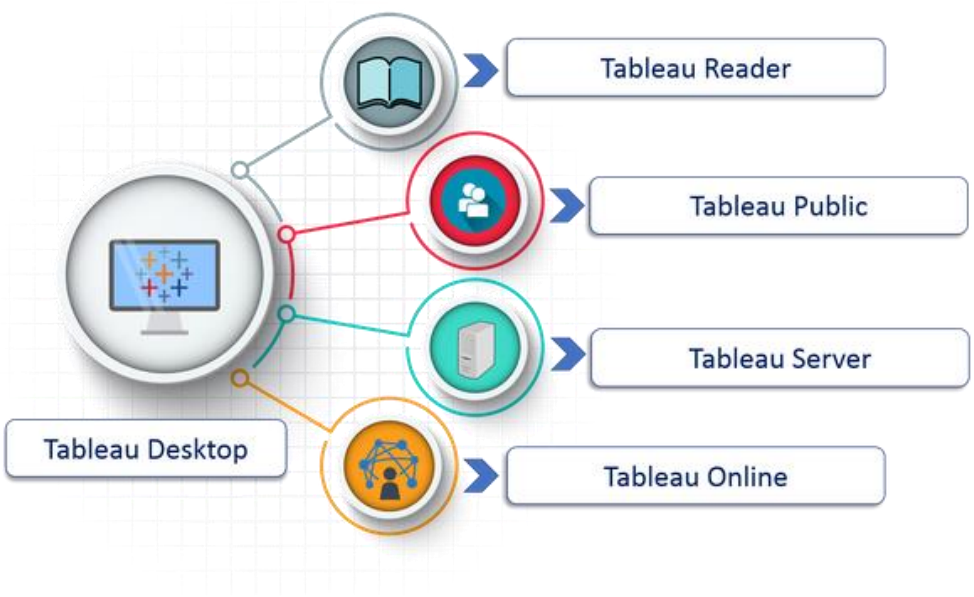
It is considered as one of the 10 technical skills with explosive growth in job demand.

Technical Skills With The Biggest Increases In Demand:

1. Big Data (Information Technology): 3,977%
2. Node.js (Design): 2,493%
3. Tableau (Research and Analysis): 1,581%
4. NoSQL (Information Technology): 1,002%
5. Apache Hadoop (Information Technology): 704%
6. HTML5 (Information Technology): 612%
7. Python (Research and Analysis): 456%
8. Oracle (Sales): 382%
9. JSON (Information Technology): 318%
10. Salesforce CRM (Sales): 292%

(Source: https://www.forbes.com/sites/jeffkauflin/2017/01/08/the-10-technical-skills-with-explosive-growth-in-job-demand/#75584d754f5c)

Tableau Family



1. **Tableau Desktop**: It *offers all the full features* of software and allows you the connect to different file types, create extracts of the data sources and save your Tableau workbooks locally.

2. **Tableau Public**: It is a free version of Tableau visualization software. You can create visualizations and connect to CSVs, Text and Excel documents. However, you are *not allowed to save your workbooks locally*. Everyone can see your data as it is saved on cloud.

3. **Tableau Reader:** It is a free desktop application that allows you to read the Tableau file types. You can open the data visualizations built in Tableau Desktop. You can filter, drill down data but you cannot edit or perform other kind of interactions.

4. **Tableau Server**: It allows users to save workbooks securely across the organization using a secure server. It is browser-based and mobile-based. You can publish dashboards with Tableau Desktop and share them with web-based Tableau server.

5. **Tableau Online**: It is a hosted version of Tableau server which helps makes business intelligence faster and easier than before.

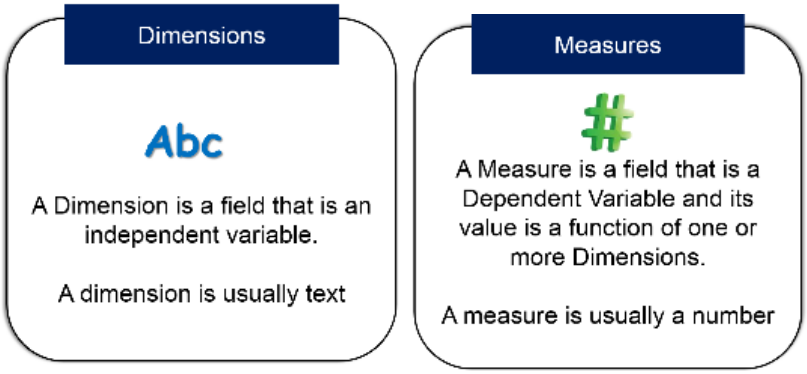
Advantages of Tableau

1. Tableau is capable of analyzing of hundreds of millions of rows and give answer in seconds.
2. It is easy to use because of its drag and drop feature.
3. The dashboard of tableau gives dynamic and interactive results.
4. Data access is easy. You can get data from spreadsheet to Hadoop.

5. The dashboard can be published and shared on web or mobile devices so that you can get the results quickly.

Measures and Dimension

Tableau divides the data in two main types: **dimensions** and **measures**.
Measures are the fields that can be measured, aggregated, or used for mathematical operations. Generally, it contains numeric value.
Dimensions are usually those fields that cannot be aggregated, they are the fields by which you want to summarize. E.g. Category, Country, City etc
Suppose you want to calculate the sum of order amount by country. Order amount will be the measure and country will be a dimension.

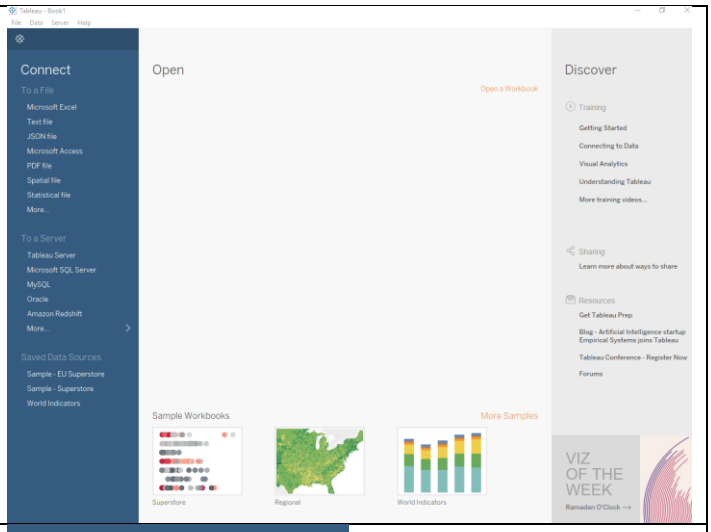
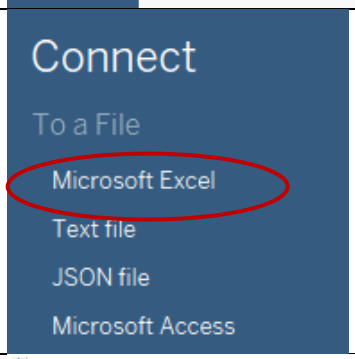
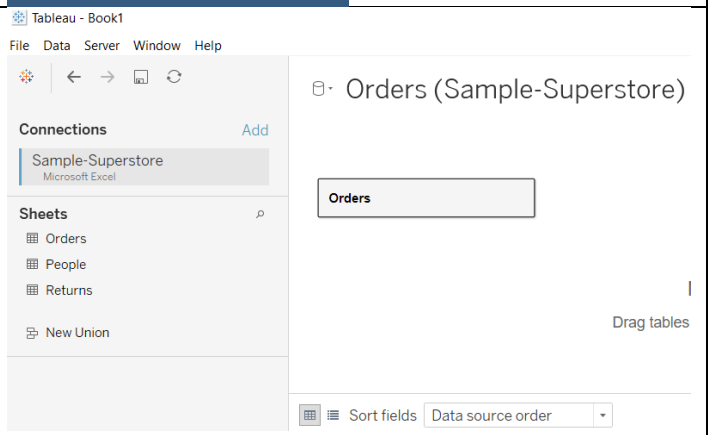


Learning Outcome

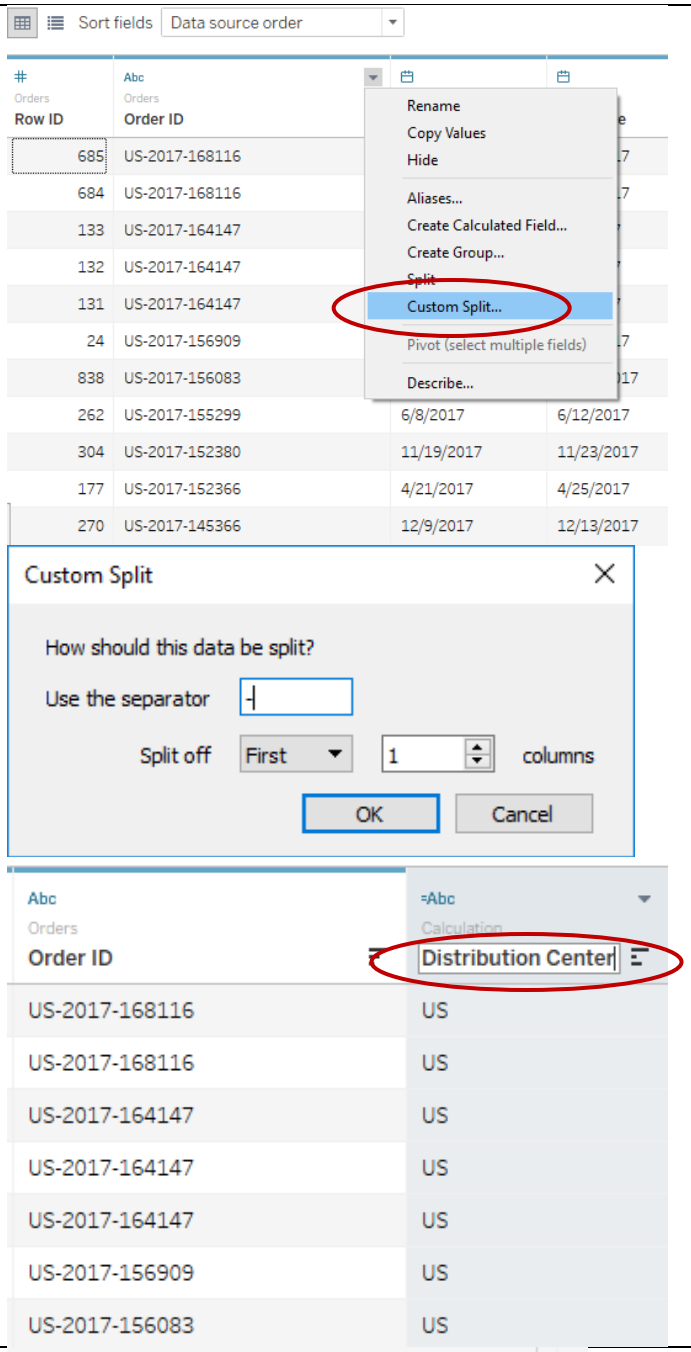
1. How to connect to the data?
2. Understand Tableau Family and the advantages of Tableau
3. Data visualization by building bar chart, column chart, pie chart, line chart etc
4. Create calculated field

I. Data Connections

The first thing to do in Tableau is to connect to your data. Tableau easily connects to nearly any data source, be it corporate Data Warehouse, Microsoft Excel or web-based data.

1. Open the software **Tableau Desktop.** You can see the **Start screen**. You can connect to new data or open recently used workbooks.	
2. Under the **Connect** Pane, select **Microsoft Excel**.	
3. Download **Lab3b-Tableau.zip**. Unzip and open the file **Sample-Superstore.xlsx**. This data set contains transactions of customers purchasing specific products. 4. On the **Data Source** Page, you can choose which sheets or tables you would like to use. Drag the sheet **Orders** to the canvas on the right. 5. You can add or remove other sheets. Or add other connections (in csv). Tableau can create join automatically.	

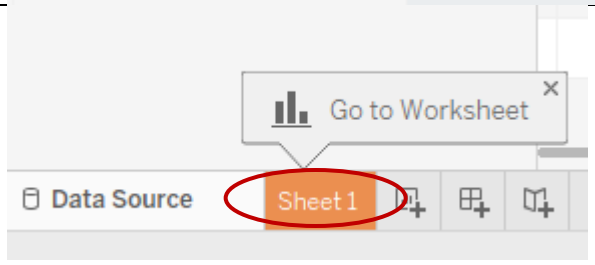
6. You can split a certain field (Order ID) by using **Custom Split**. Type hyphen (-) as separator and get the first string. **Rename** the new column to **Distribution Center**.

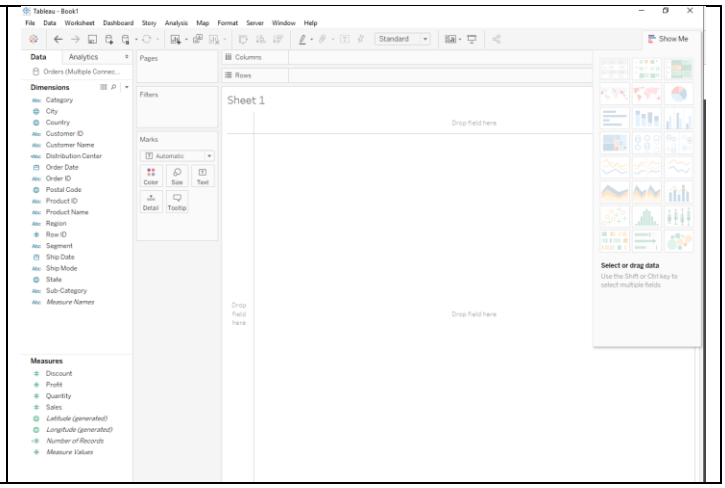


7. Click on the sheet tab **Sheet 1** at the bottom.

8. In the worksheet, the columns from your data source are shown as fields on the left side in the Data pane

9. Choose **File ➔ Save** to save the workbook with the name **lab3b-inclass-ans1.twbx**.

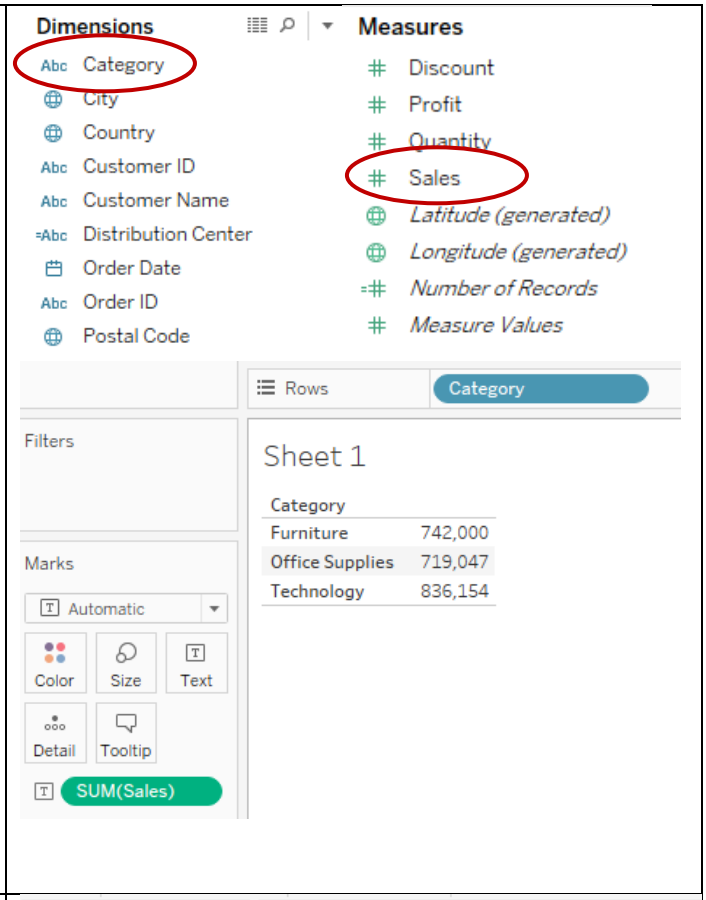




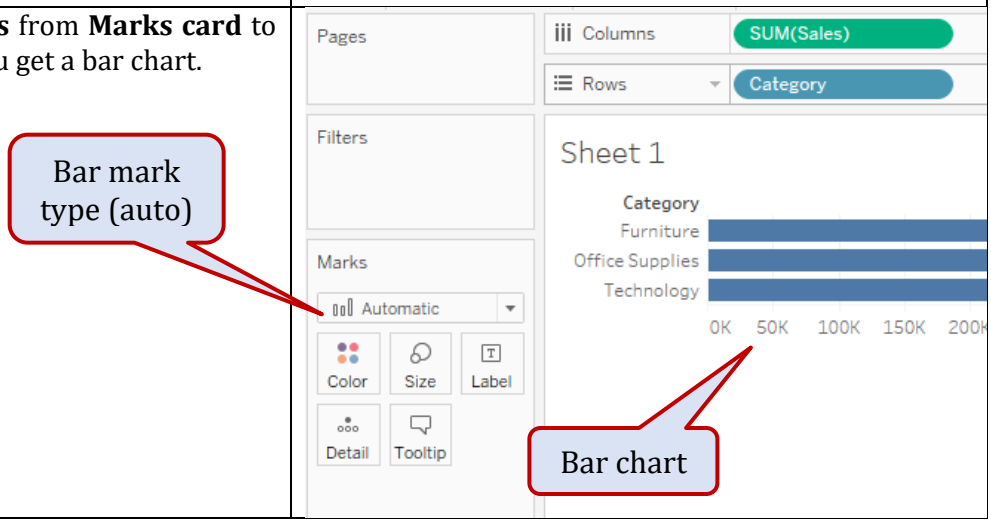
II. Building Bar chart and Column chart

Bar charts and column charts are used to compare data across categories. **Bar** mark type is used to compare measures across categories, or to break data down into stacked bars.

1. **Double click Category** under **Dimensions** Area and **double click Sales** under **Measures** Area. It shows the sum of sales for different categories.
   (Measures are usually continuous; they are the fields you want to analyze, which are in green color)



2. Drag the measure **Sales** from **Marks card** to **Columns** shelf, now you get a bar chart.



3. Select the button  to swap fields in **Rows** shelf and **Columns** shelf. It changes to a **column chart**.

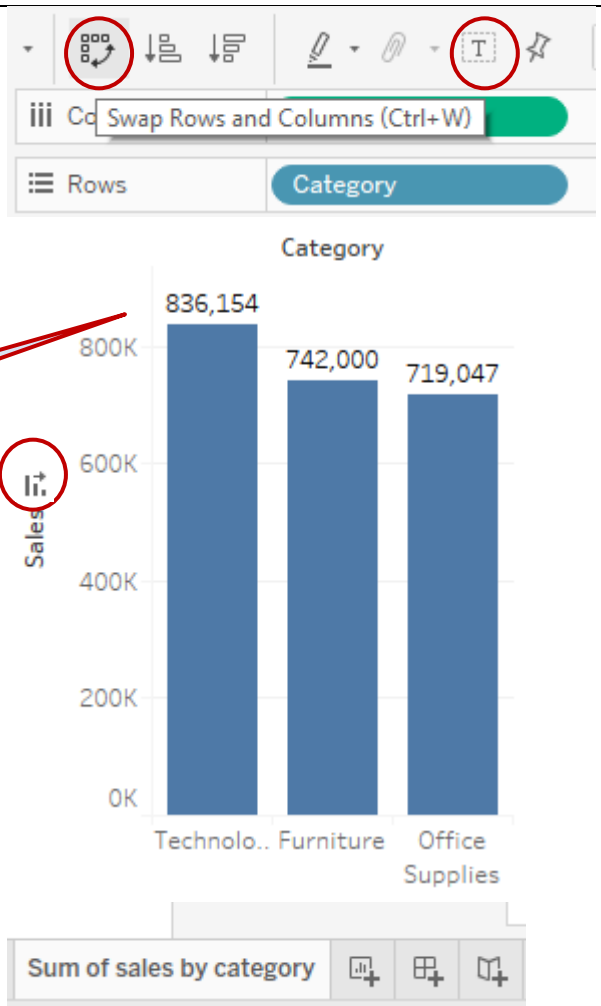
4. **Sort** the Sum of Sales in **descending** order by clicking  near the y axis label **Sales**.

5. Show the mark labels (data labels) by clicking **Show Mark Labels**  .

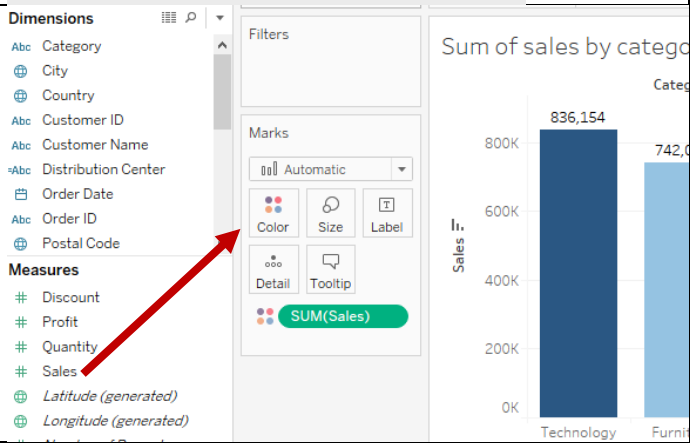
Column chart with data labels

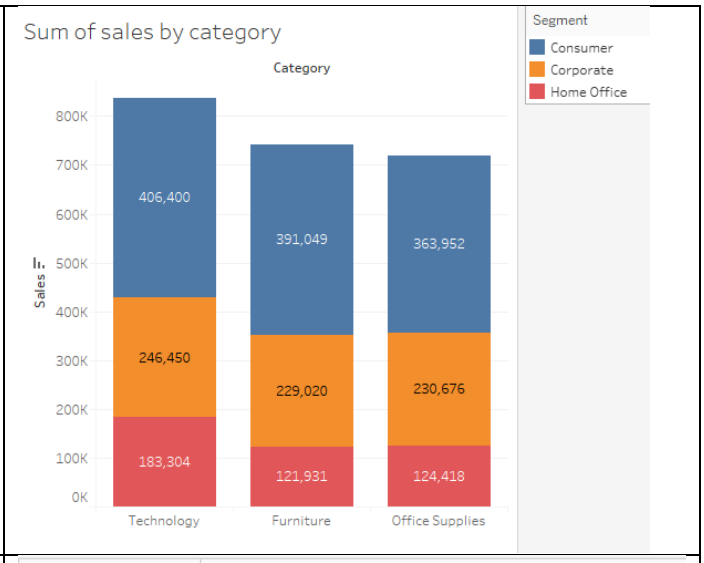
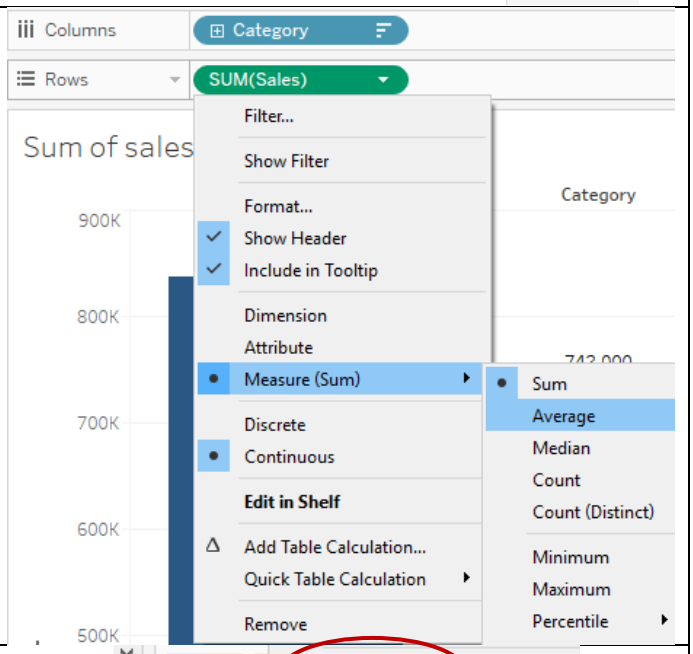
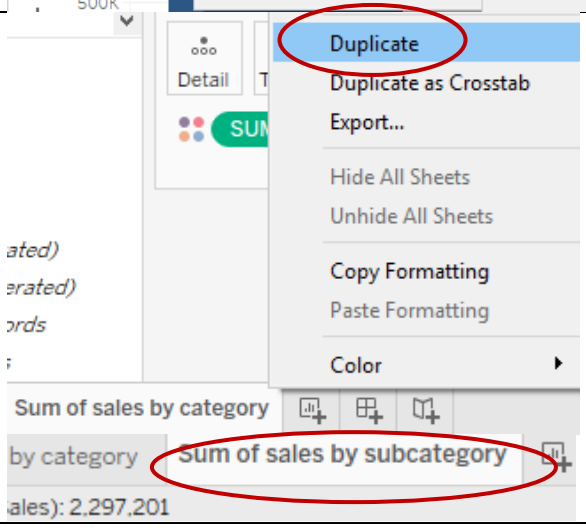
6. Select **Entire View** to increase the column width.

7. **Rename** the worksheet to **Sum of sales by category**. The chart title is changed as well.

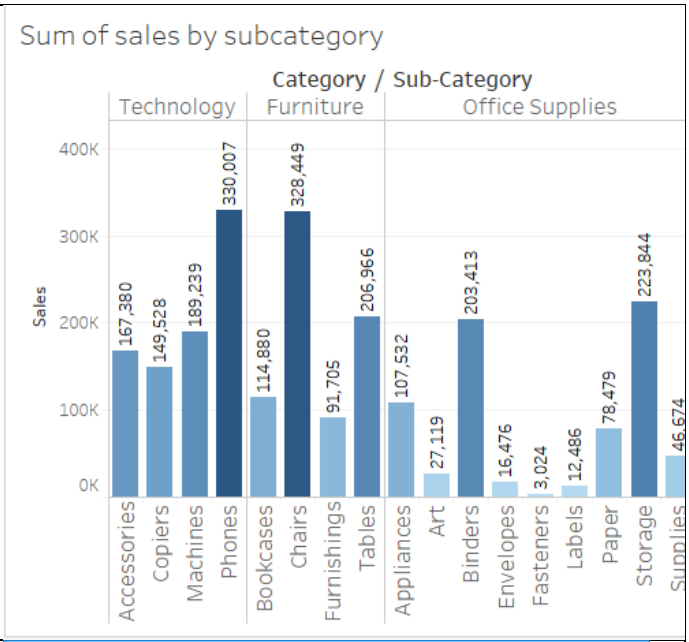


8. Highlight the largest sales by using darker color, just drag the **Sales** measure to **Color**.

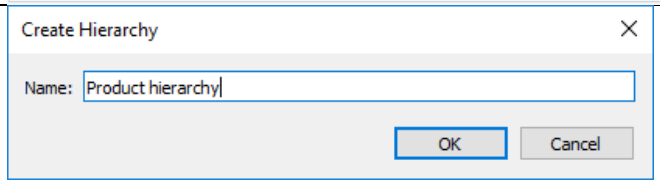


9. Drag the **Segment** dimension to **Color** to break data down into *stacked column* Undo the action by pressing **Ctrl+Z**.	
10. Click on **SUM (Sales)** measure in **Rows** shelf, choose **Measure (Sum) → Average** to display average sales. [Sum of sales/Number of records] **Undo** the action by pressing **Ctrl+Z**.	
11. **Duplicate** the worksheet and rename the new worksheet to **Sum of sales by subcategory**.	

12. **Double click Sub-Category** dimension. The result will be similar to this one.



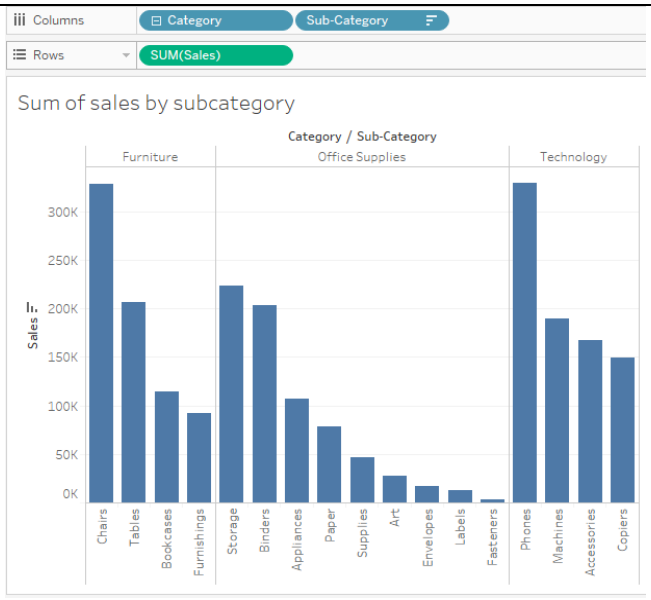
13. **Clear** the worksheet by clicking  , create a hierarchy by dragging **Sub-Category** to **Category. Rename** the hierarchy to **Product hierarchy**. Press **OK**.



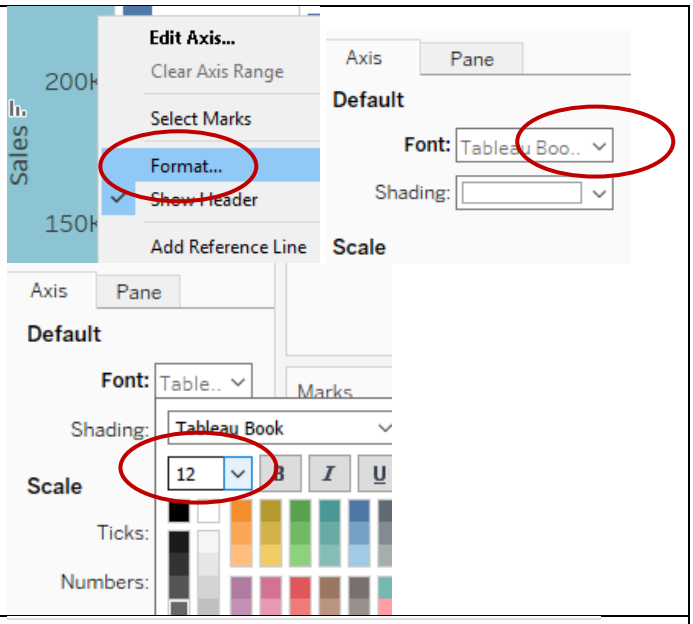
14. Drag the **Product hierarchy** to **Columns** shelf, and drag the **Sales** measure to **Rows** shelf. Click on **+ sign** to expand the product hierarchy to show subcategory sales.

15. **Sort** the **sum of sales** in **descending** order within each category.

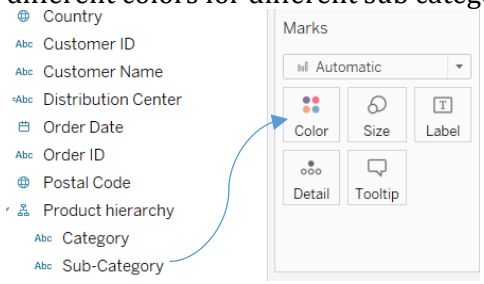
16. **Size** property in **Marks card** allows you to make the columns wider or narrower.

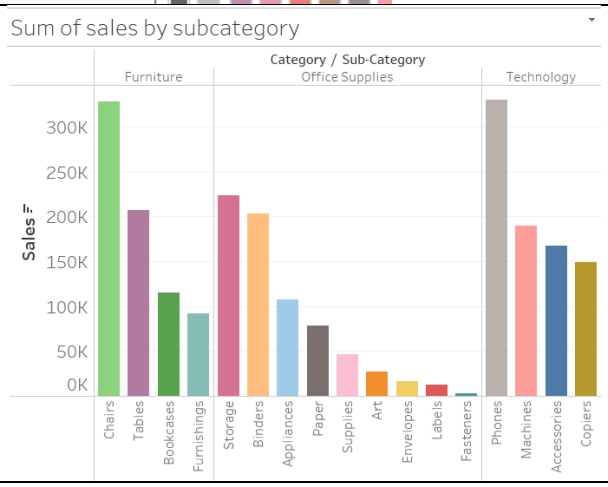


17. **Right click** the **y axis**, choose **Format**, and **Increase** the **font size to 12** for y axis labels.
18. Close the **Format** pane.



19. Drag the **Sub-Category** dimension from **Data Pane** to **Color** in **Marks card** to generate different colors for different sub categories.



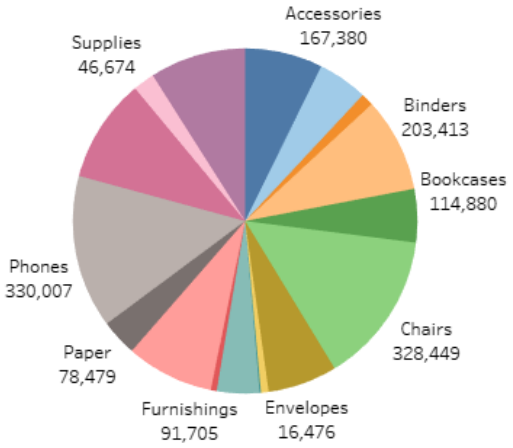


20. Save the workbook **lab3b-inclass-ans1.twbx**.

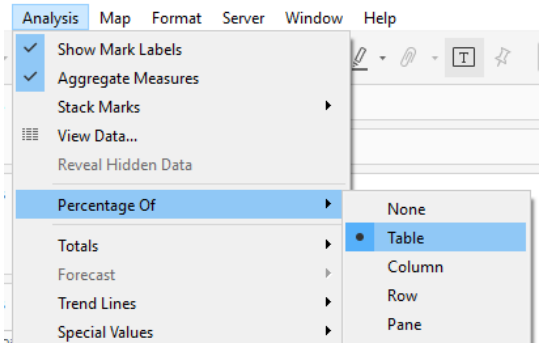
III. Building Pie chart

Pie charts are used to show proportions of a whole. Pie mark is used to show proportions.

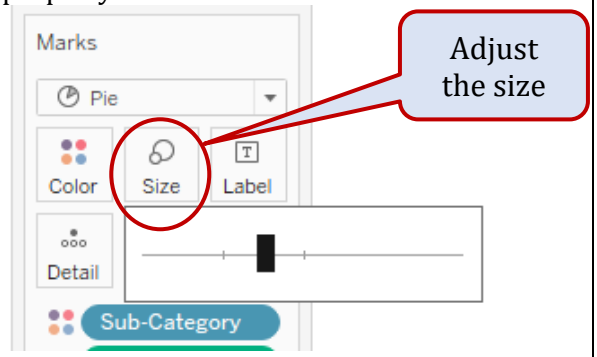
1. Create a new worksheet named **Pie chart**.
2. Drag the **Sales** measure to **Columns**, drag the **Sub-Category** dimension to **Rows**.
3. Click **Show Me** on the toolbar and select **pie chart** type. Select **Entire View** to make the chart bigger.
4. Drag the **Sub-Category** dimension to **Label** on the **Marks card**. Drag **Sales** measure to **Label** too.



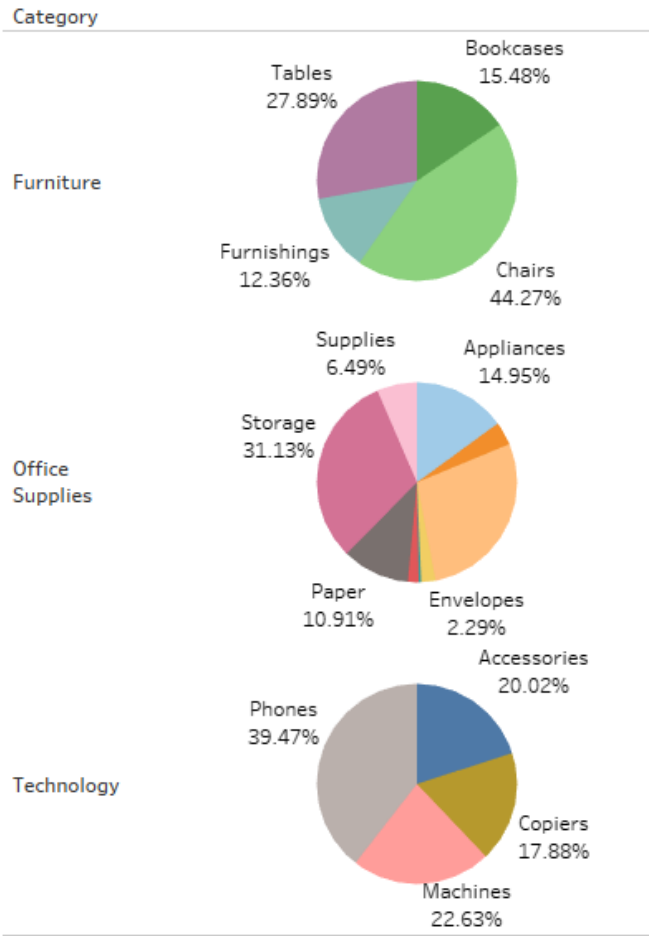
5. Select **Analysis** tab → choose **Percentage Of** → **Table** to show percentage as a whole.



6. Adjust the size of the chart by using **Size** property in **Marks card**.



7. Drag the **Category** dimension to **Rows** to display a pie chart showing percentage of sales for different subcategories, one for each category. Change the percentage so as to show the distribution of each sub-category within each category. (as shown here)



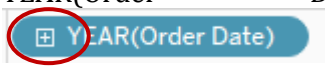
8. Save the file **lab3b-inclass-ans1.twbx**

IV. Building Line chart

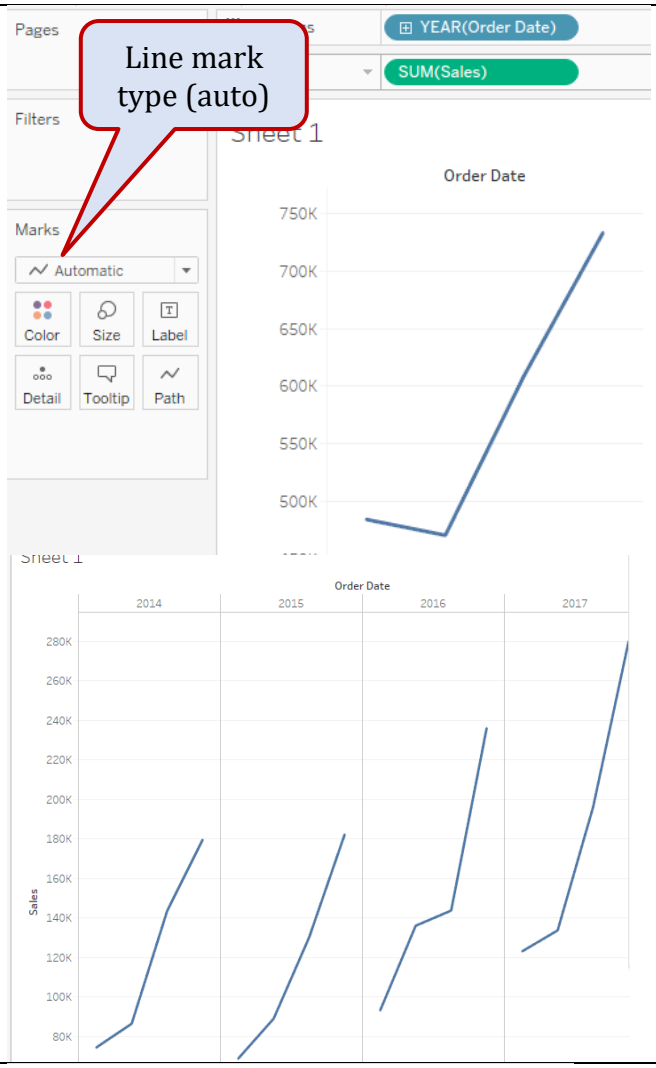
Line charts are used to visualize a sequence of values and they are useful when you want to see trends over time, or to forecast future values. The line mark type is useful when you want to see trends in data over time, your data are ordered, or interpolation makes sense.

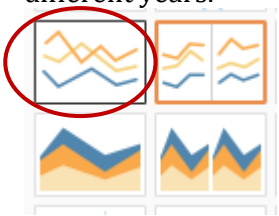
1. Create a new worksheet named **Line chart**.

2. Drag the **Order Date** dimension to **Columns**, drag the **Sales** measure to **Rows** to show the sum of sales for different years. [default date level is YEAR]

3. **Click** the **plus sign** on the left side of the field YEAR(Order Date) field 

to show quarter sales. The new dimension divides the view into separate panes for each year. It is a nested table because it displays multiple headers.

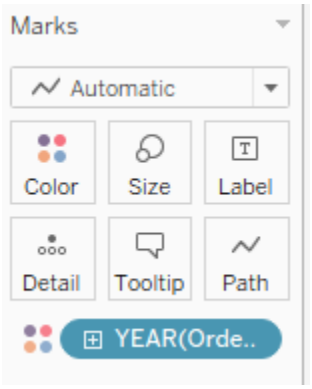


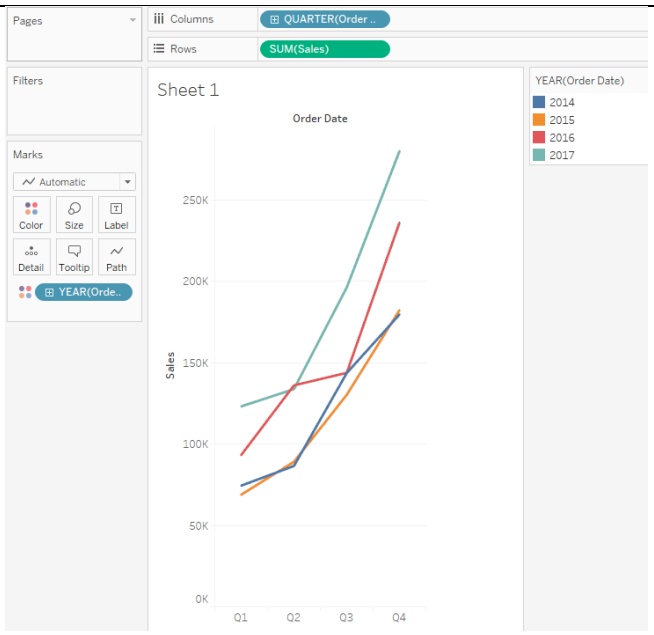
4. Change the chart type to **Continuous line** chart by clicking **Show Me**. The line becomes continuous across different quarters in different years.


5. **Undo** the action and change the chart type to **discrete line** chart.

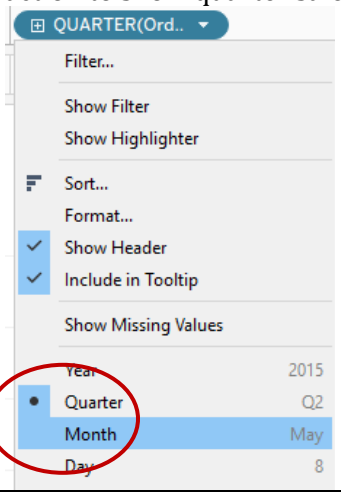
6. Compare the growth by quarter across the years. Move the **Year** field to **Color**.

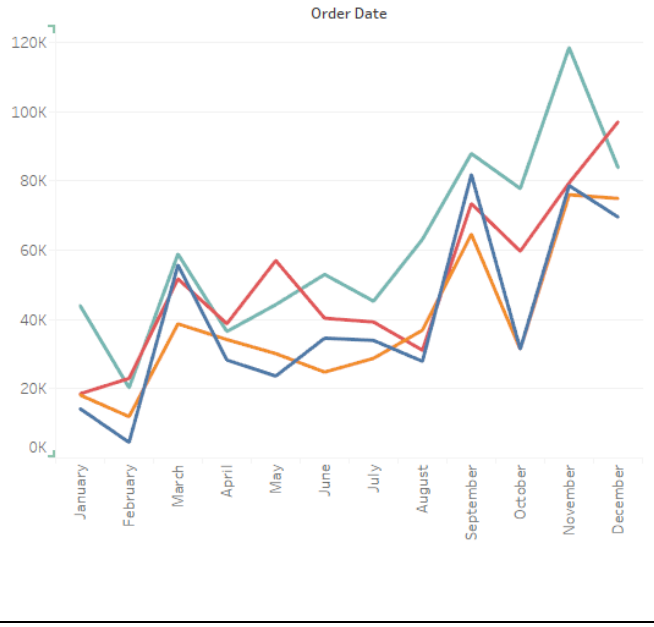
   [Quarter 4 shows more sales for every year.]



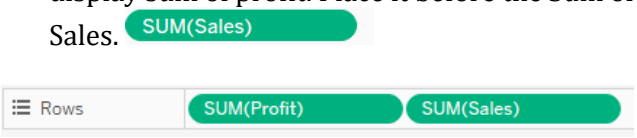


7. Click on **Quarter** in **Columns** shelf → select **Month** option as shown below to show monthly sales for different years. **Undo** the action to show quarter sales again.

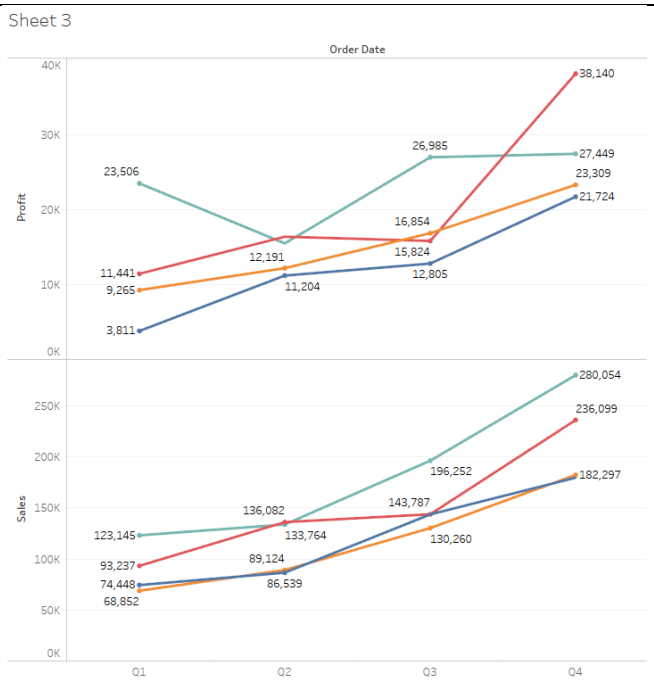




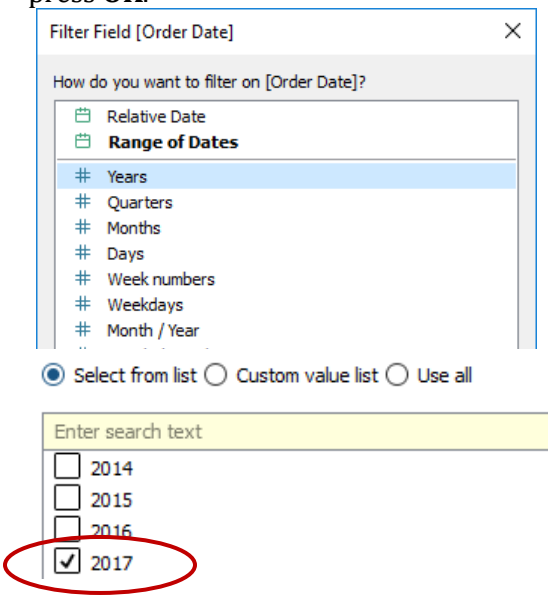
8. Drag the **Profit** measure to **Rows** shelf to display sum of profit. Place it before the Sum of Sales.



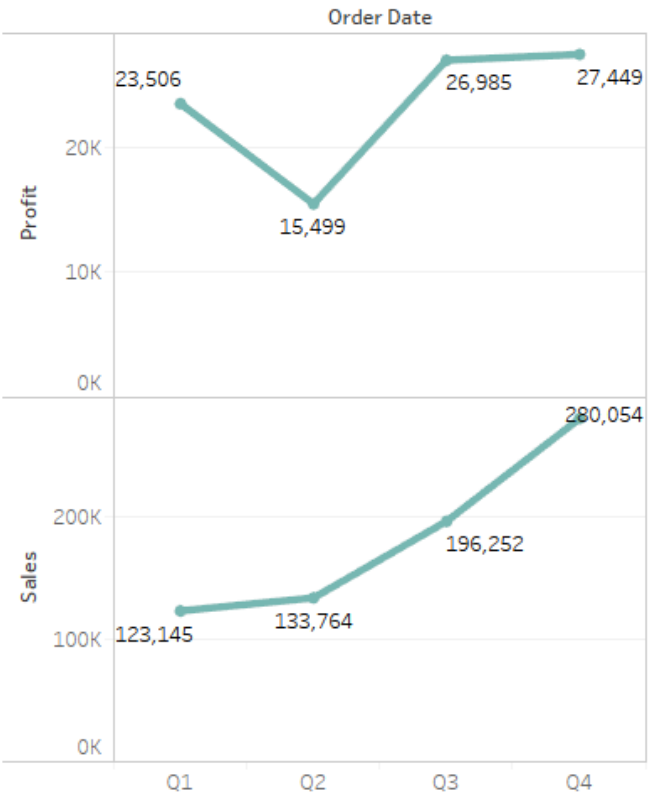
9. **Show** the **data labels** to view the details. Select **Entire View** to enlarge the chart.



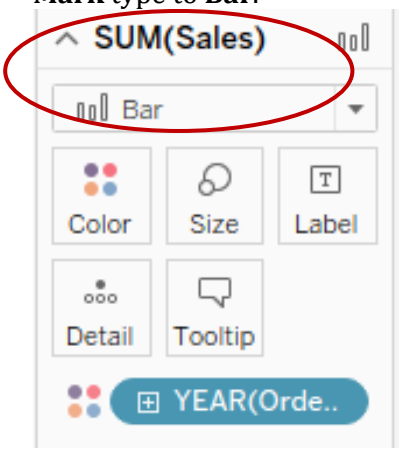
10. Drag the **Order Date** measure to **Filters** shelf, select **Years**, click **Next**. Select only **2017**. Then press **OK**.

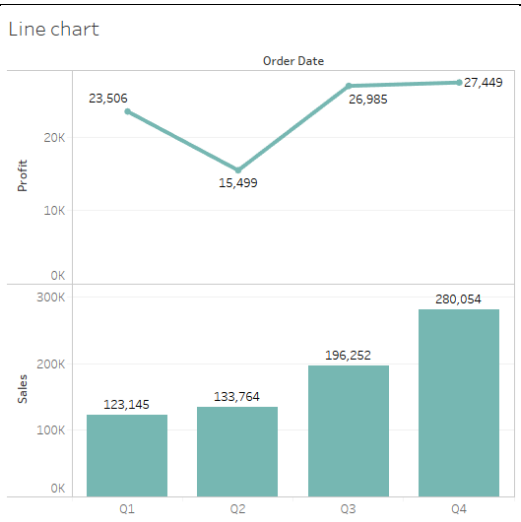


11. Adjust the **Size** property in **Marks card** to make the lines thicker.



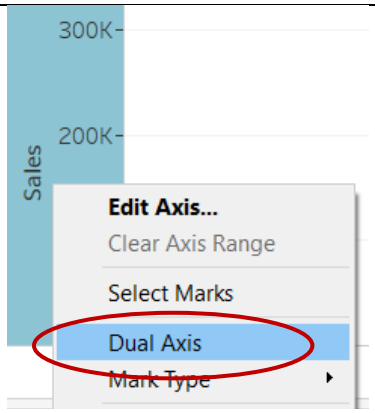
12. For the second chart **SUM(Sales),** change the **Mark** type to **Bar**.

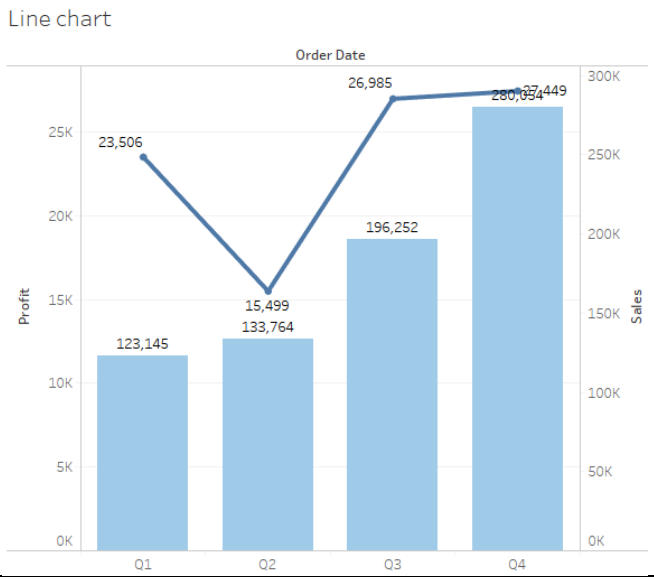




13. **Right click y-axis** for the column chart. Select **Dual axis**. It combines sales and profit by using dual axis and show both measures in one chart.

14. Save the file **lab3b-inclass-ans1.twbx**



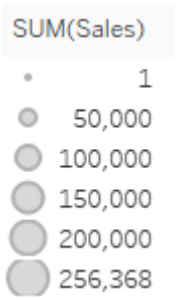


V. Geographic Data Analysis

1. Create a new worksheet named **Map**.

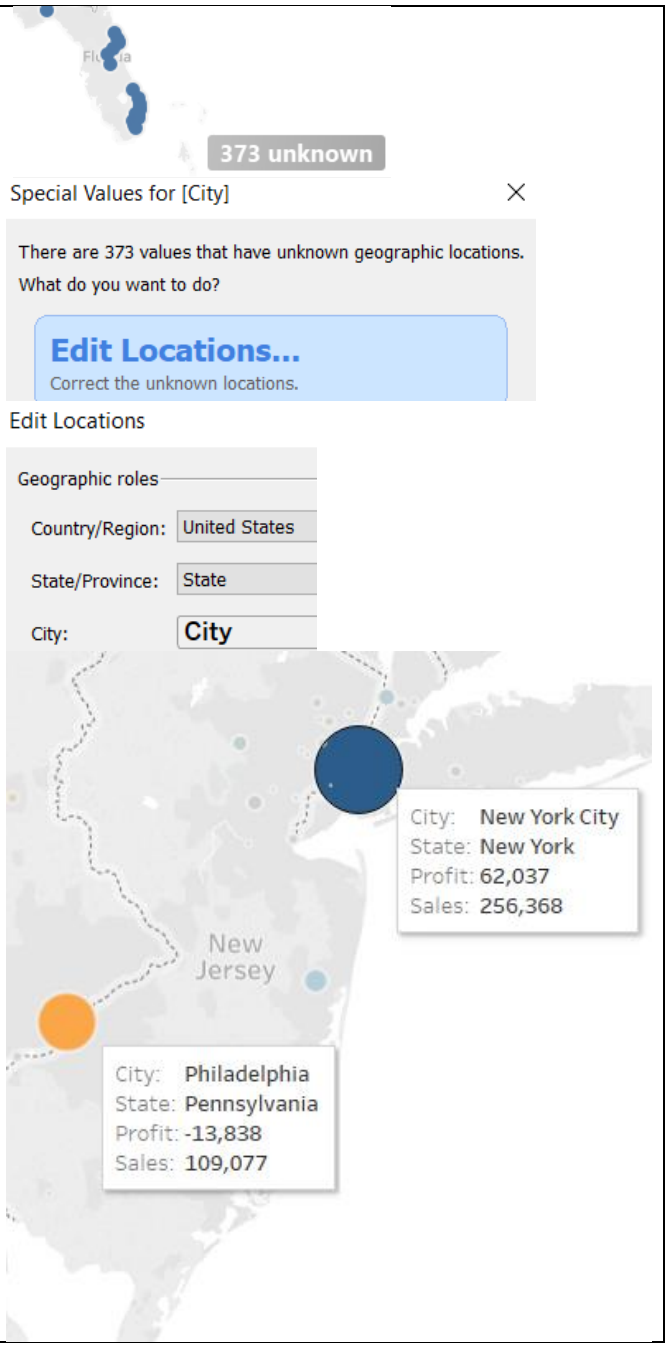
2. **Double click City** dimension. It shows a map of United States. Click **373 unknown** at the bottom right corner and choose **Edit Locations**. Change the **State/Province** option to **State** and change **Country/Region** option to **United States**. Then press **OK**.

3. Drag **Sales** measure to **Size** in **Marks card** to show which city has sales. The bigger the size of the bubbles, the larger the sales. You may adjust the **Size**.



4. Drag **Profit** to **Color** in **Marks card**. It shows the profit in Philadelphia is low, it is suggested to close the store there. (Profit and Sales in nearby city [NY] is high)

5. Save the file **lab3b-inclass-ans1.twbx**

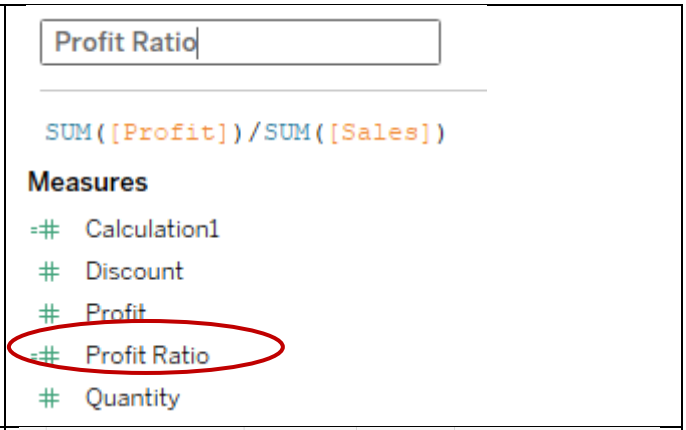


VI. Create the calculated field

1. Create a new worksheet named **Profit Ratio**.

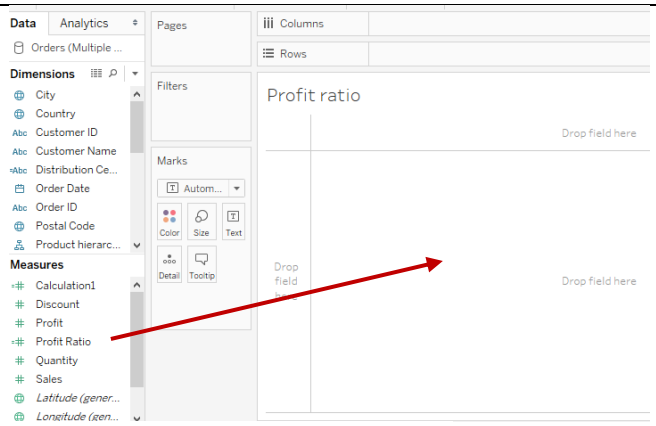
2. Select **Analysis** menu → **Create Calculated Field**, type **Profit Ratio** as calculated field name, enter the formula **SUM([Profit])/SUM([Sales])** then press **OK**.

3. The new calculated field is added to the **Data** pane. If the new field computes quantitative data, it is **added to Measures**. If it computes qualitative data, it is added to Dimensions.



4. Drag the new measure **Profit Ratio** to **Drop field here** area on the right. Then drag the **Product hierarchy** to **Rows**, expand the level to show profit ratio for different sub-categories.

6. **Save** the file **lab3b-inclass-ans1.twbx** and **close** the program **Tableau Desktop**.

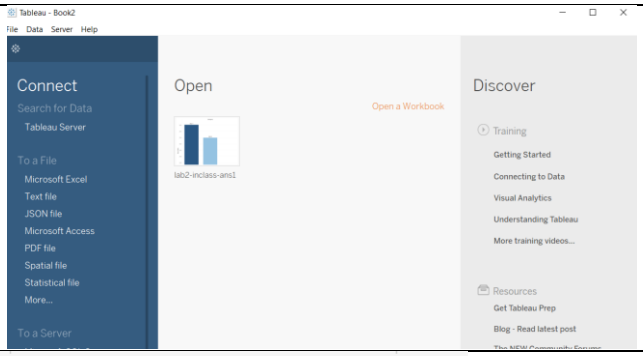
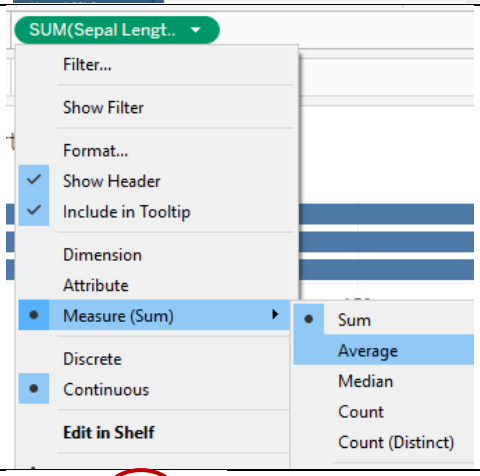
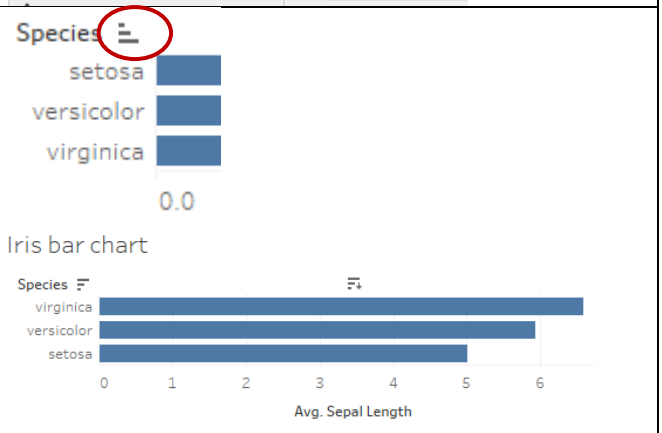
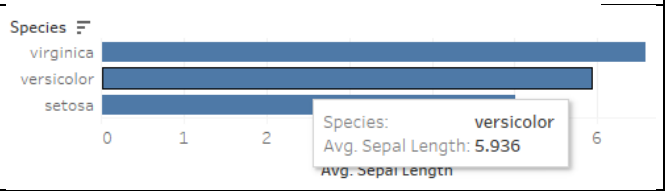


Profit ratio

Category	Sub-Catego..	
Furniture	Bookcases	-0.0302
	Chairs	0.0810
	Furnishings	0.1424
	Tables	-0.0856
Office Supplies	Appliances	0.1687
	Art	0.2407
	Binders	0.1486
	Envelopes	0.4227
	Fasteners	0.3140
	Labels	0.4442
	Paper	0.4339
	Storage	0.0951
	Supplies	-0.0255
Technology	Accessories	0.2505
	Copiers	0.3720
	Machines	0.0179
	Phones	0.1349

VII. Building bar chart

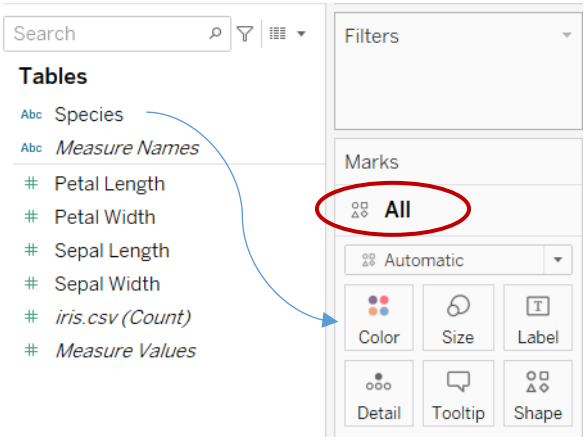
We can use bar charts to compare data across categories.

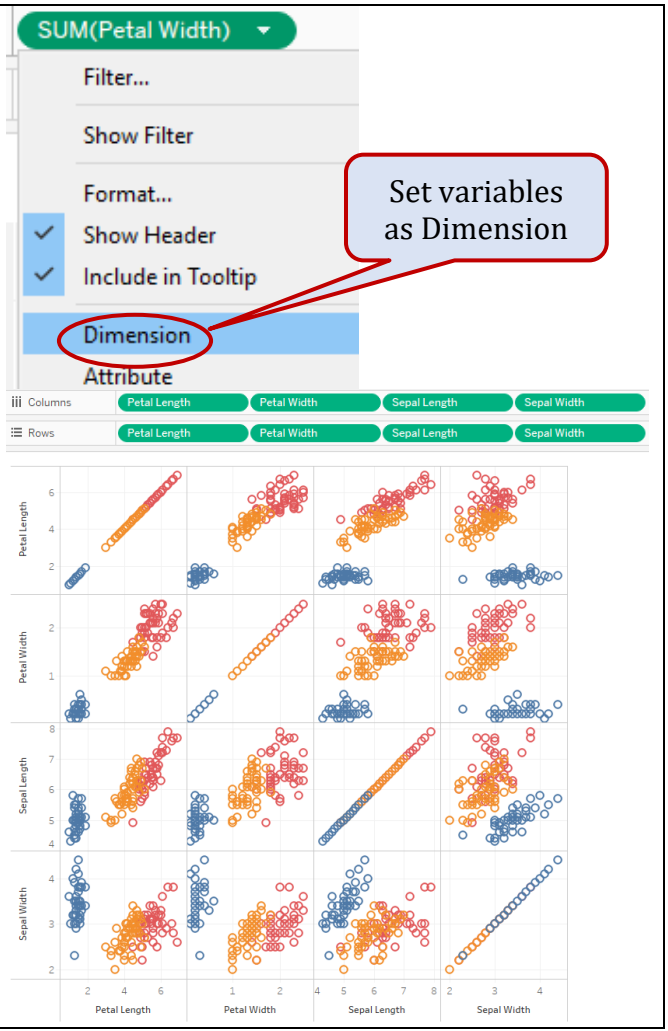
1. Open the program **Tableau Desktop.** Under the **Connect** Pane, select **Text file**. And open the file **iris.csv**. **Save** the tableau file with the name **lab3b-inclass-ans2.twbx**	
2. **Rename** the worksheet **Sheet1** as **Iris bar chart** 3. Drag **Species** to **Rows** and **Sepal Length** to **Columns**. 4. Click on **SUM (Sepal Length)** measure, choose **Measure (Sum)** → **Average** to display average sepal length.	
5. Click on the **Sort** button (next to **Species**) to sort the bars by name or by value.	
6. You can highlight the selected elements when you hover your mouse to the element	

VIII. Building interactive Scatter Plots

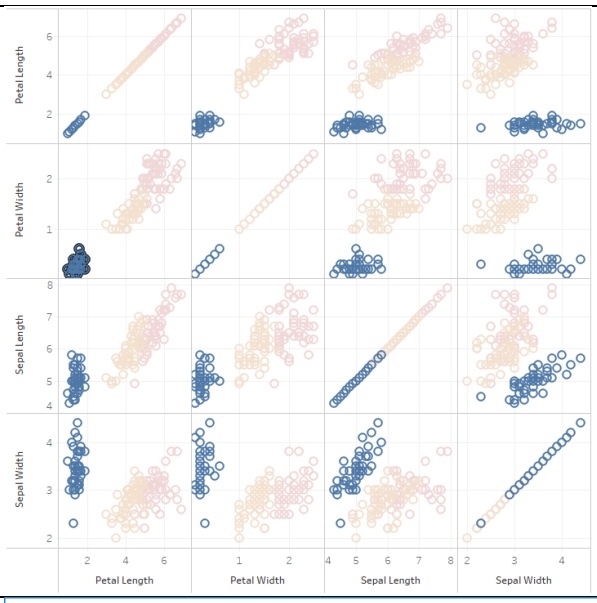
Scatter plots are used to visualize relationships between numerical variables. You need to place at least one measure on the **Columns** shelf and at least one measure on the **Rows** shelf. By default, scatter plot uses **shape** mark type, which is useful when you want to clearly see individual data points while also viewing categories associated with those points.

1. Create a new worksheet named **Scatter plot**.
2. Drag the **Petal Length, Petal Width, Sepal Length** and **Sepal Width** measures to **Columns**.
3. Drag the **Petal Length, Petal Width, Sepal Length** and **Sepal Width** measures to **Rows**.
4. Click on **SUM (Petal Length)** measure on **Columns**, select **Dimension**. Repeat the step for other measures on Columns and Rows.
5. Drag the **Species** dimension to **Color** on the **Marks card**. (under **All**, as shown below)

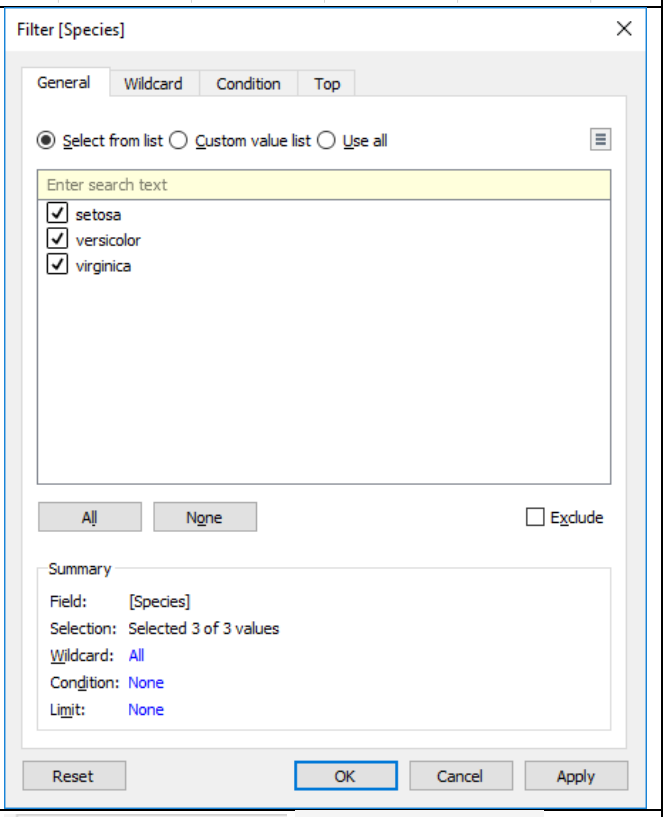




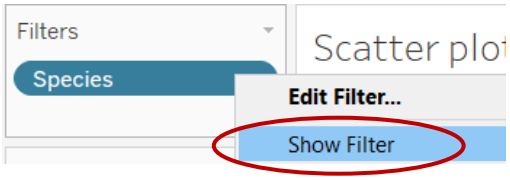
6. Items that interactively selected in one cell in scatterplot matrices are immediately highlighted in other cells.



7. Drag the **Species** dimension to **Filters** shelf to filter the data by species name. Press **OK**.

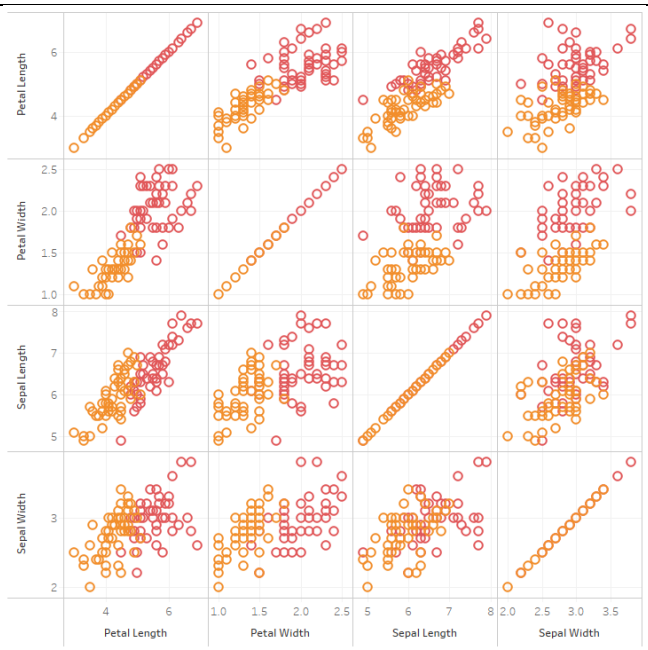


8. Right click on **Species** dimension in **Filters** shelf and choose **Show Filter**



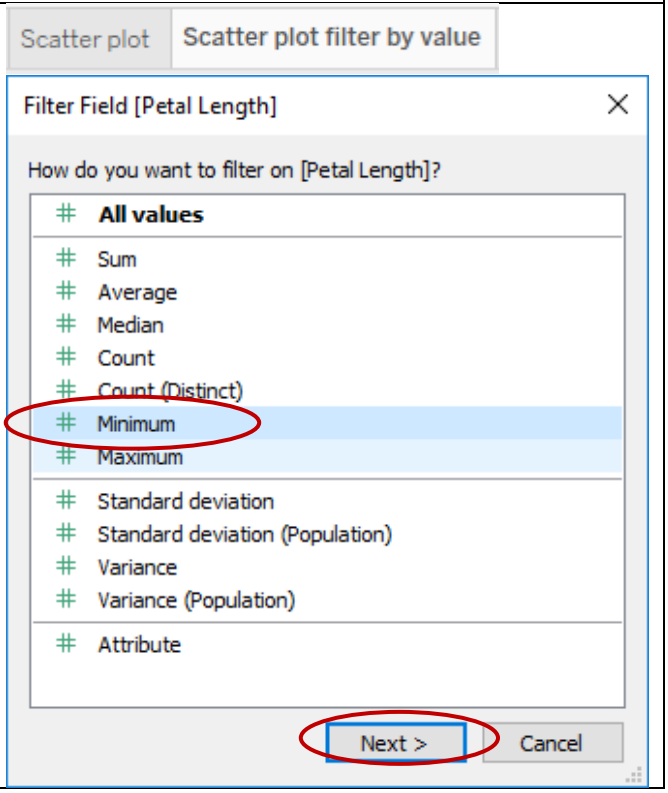
9. A filter is added to your visualization (on the right). Only show the data of 'versicolor' and 'virginica'.

10. **Save** the workbook **lab3b-inclass-ans2.twbx**

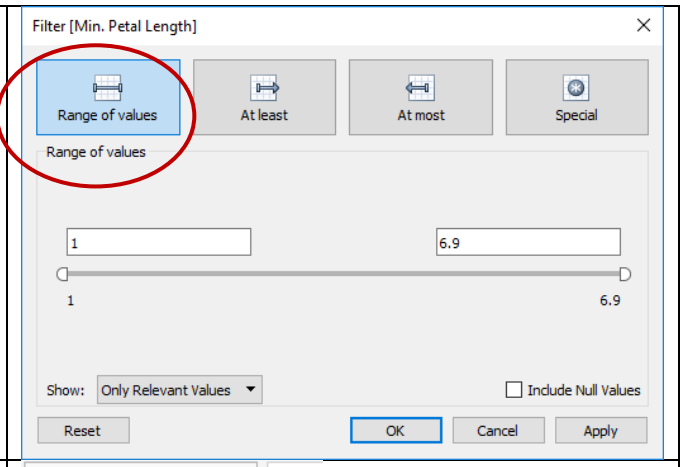


11. **Duplicate** the worksheet **Scatter plot** and named the new worksheet as **Scatter plot filter by value**.

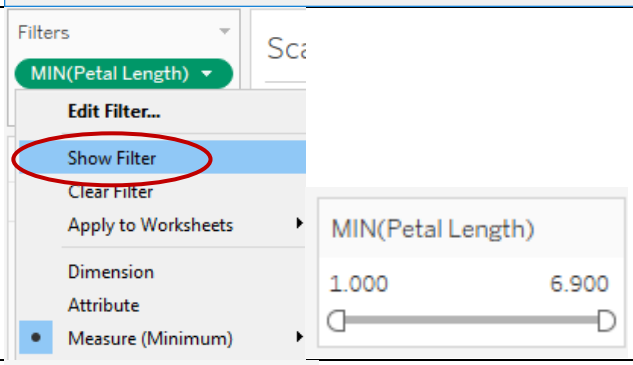
12. Remove the **Species** Dimension in **Filter** shelf. And add **Petal Length** to **Filters**, select **Minimum** and press **Next**



13. Filter by a **Range of values.** Press **OK.**

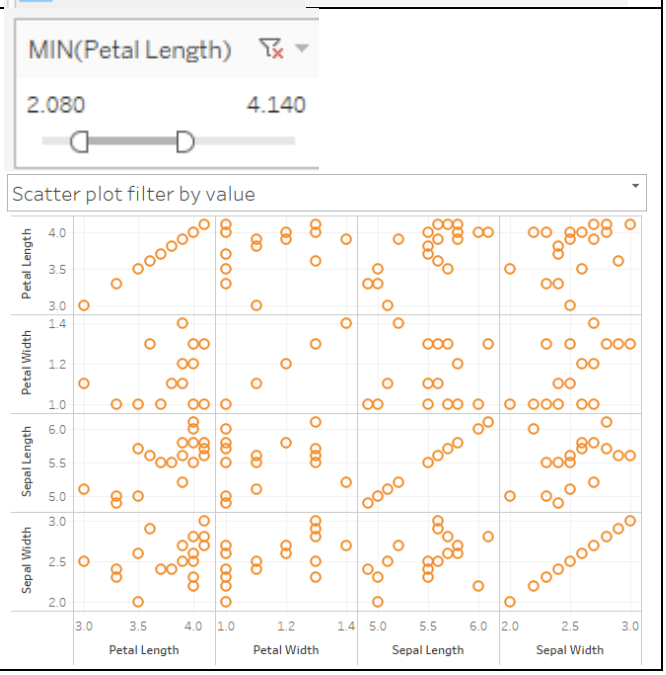


14. Select **Show Filter** in **Filters** shelf. A slider for filtering numeric values is added to your visualization.



15. Only show the data with petal length in the range of **[2.08, 4.14]**

16. **Save** the workbook **lab3b-inclass-ans2.twbx**

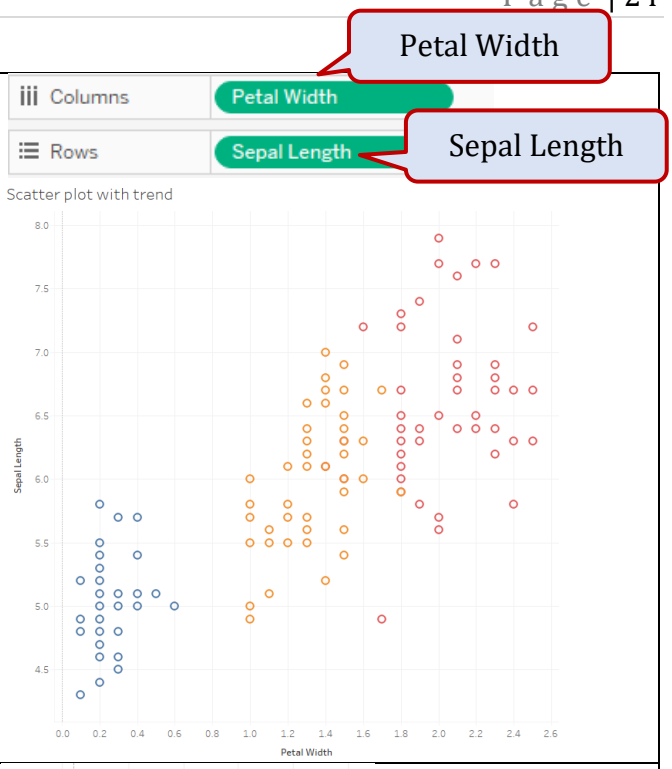


17. Create a new worksheet named **Scatter plot with trend**.

18. Drag the **Petal Width** measure to **Columns**. Drag the **Sepal Length** measure to **Rows**.

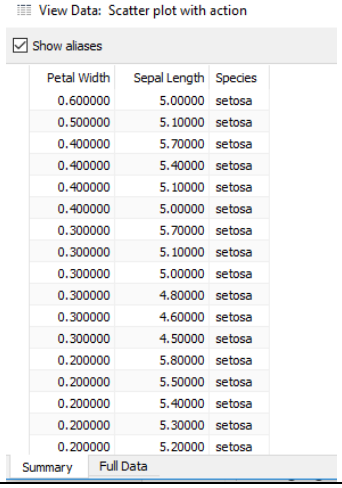
19. Click on **SUM (Petal Width)** measure on **Columns**, select **Dimension**. Repeat the step for **Sepal Length**.

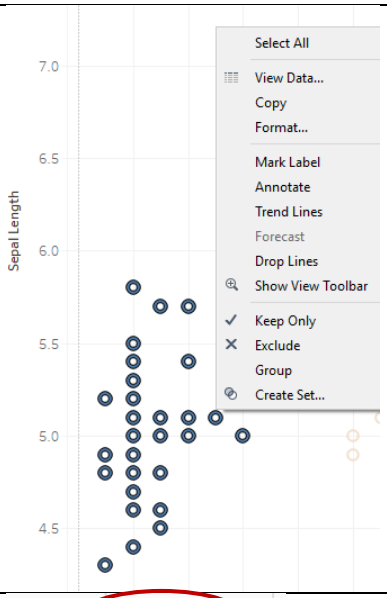
20. Drag the **Species** dimension to **Color** on the **Marks card**.



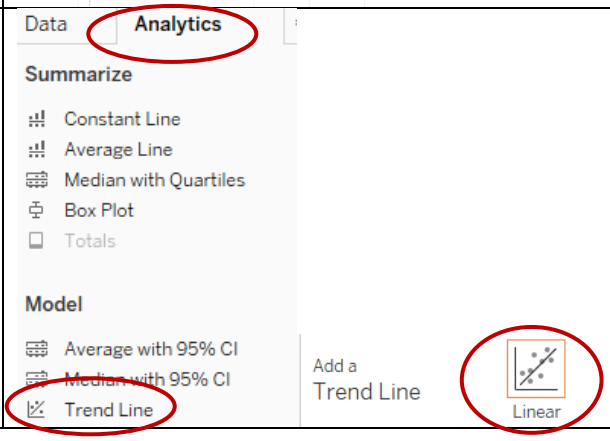
21. Select all setosa items and **right click**, choose **View Data**

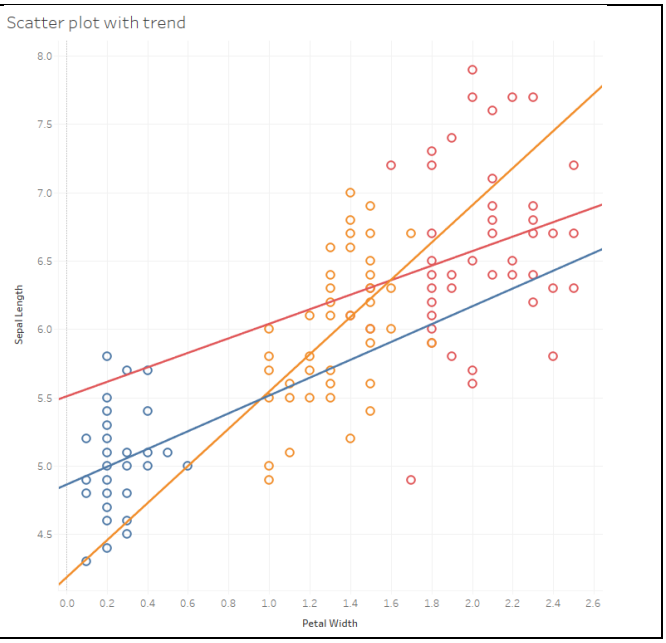
22. Close the dialog box.



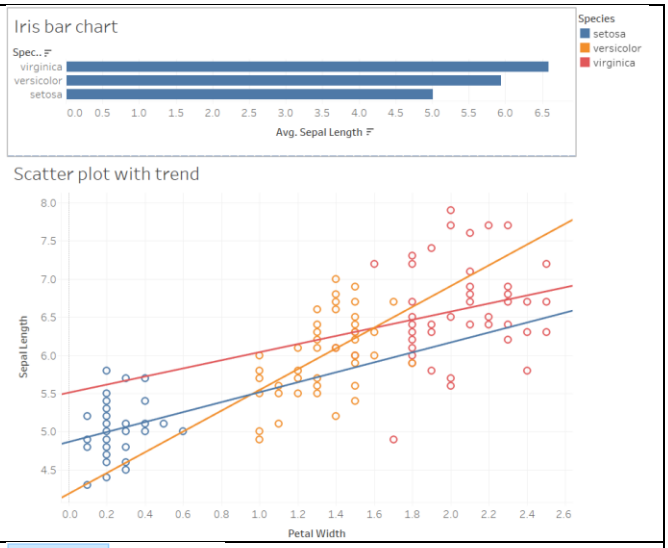
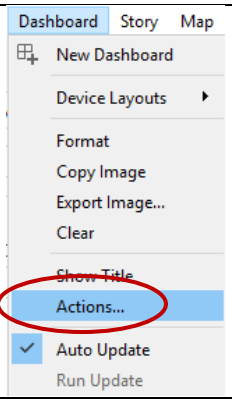


23. Go to the **Analytics** pane, drag and drop the **Trend Line** on the model type **Linear**. A trend line can provide a statistical definition of the relationship between two numerical values. Hover the cursor over the trend lines to see statistical information to create the line.

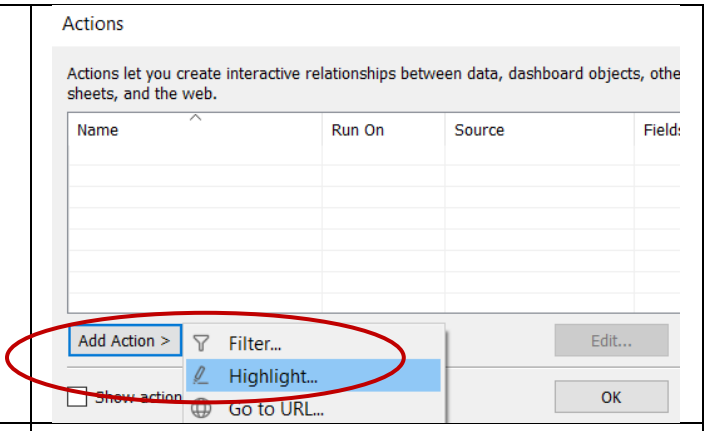


24. The chart will be similar to this. **Save** the file **lab3b-inclass-ans2.twbx**	

IX. Create Dashboard

1. Create a new **Dashboard** named **Interactive Dashboard** by using  2. Add two worksheets **Iris bar chart** and **Scatter plot with trend**. The result will be like this:	
3. Select **Dashboard** menu → **Actions**	

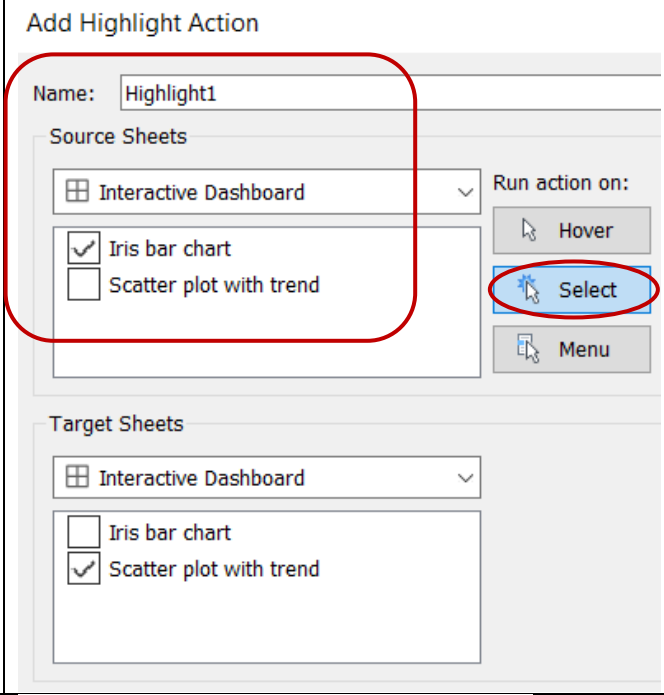
4. Choose **Add Action > → Highlight**



5. There are different options to launch the action:
   - **Hover**: Rest the pointer over a mark in the view to run the action.
   - **Select**: Click on a mark in the view to run the action
   - **Menu**: Right-click a selected mark in the view to run the action.

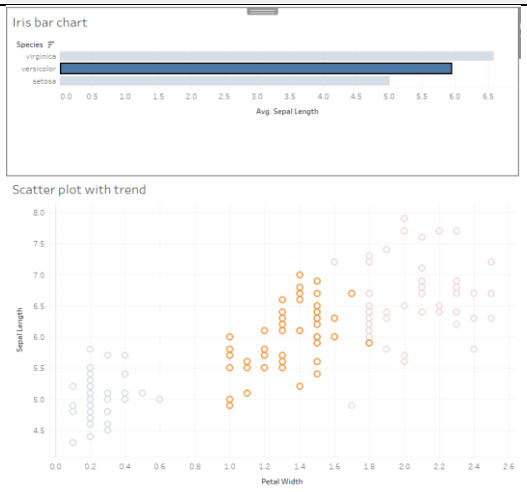
6. Select **Iris bar chart** as **Source Sheet**. **Select Scatter plot with trend** as **Target Sheet**.

7. Choose **Select** under **Run action on**, press **OK** and **OK**



8. **Select** a species like **versicolor** in **Iris bar chart**. Corresponding items in **scatter plot with trend** will be highlighted.

9. Save the file **lab3b-inclass-ans2.twbx**



Take home assignment

Open the files **Financial Sample.xlsx** and **lab3b-assignment-ans.docx** and complete the exercises.

Submission

Submit the following files to buelearning website:

- lab3b-inclass-ans1.twbx, (In-class exercise)
- lab3b-inclass-ans2.twbx, (In-class exercise)
- lab3b-assignment-ans.docx, (Take home assignment)
- lab3b-assignment-ans.twbx (Take home assignment)

References

- Tableau Tutorial | Step by Step Guide to Learn Tableau | Edureka
- What is Tableau | Data Visualization Using Tableau | Edureka
- https://data-flair.training/blogs/tableau-tutorial/