DATA ANALYTICS FOR ORGANIZATION

Prepared by:
Dr. Nor Azizah Ali
Dr. Noorfa Haszlinna Mustaffa
Faculty of Computing, UTM
2022/2023 – Sem. 1

Session 2:

Application of Data Analytics with Tableau

(Part 1)

Outline:

PART 1:

- Getting Started
- Connecting to Data
- Overview of Tableau – Basics and Concepts
- Creating Basic Visualization Charts
- Organizing Data

PART 2:

- Analytics Tab
- Mapping the Data
- Data Blending
- Dashboard and Stories

Introduction

"**Tableau** is a trending and market-leading BI tool used to visualize and analyze your data in an easily digestible format. It allows you to work on live data-set and spend more time on data analysis rather than data wrangling."

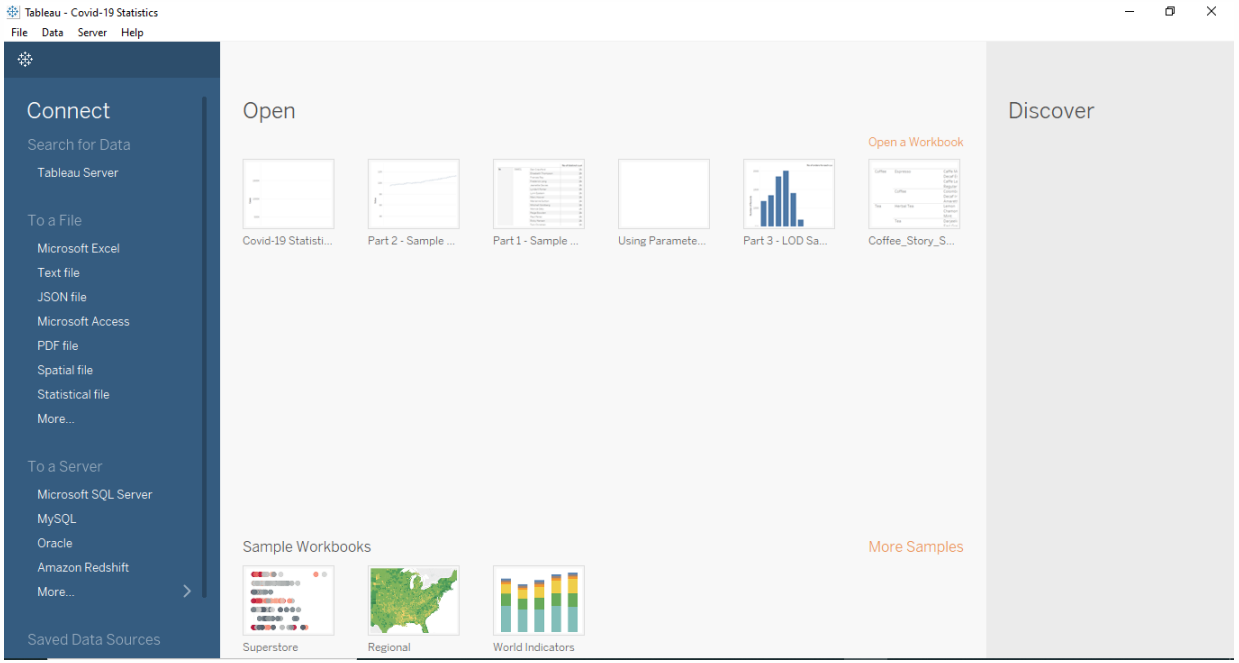
Tableau Product

Tableau Product (cont'd)

Tableau Desktop	Tableau Desktop Public
Paid (free for 14 days)	Free
Supports all kinds of data connections including local files and databases.	Supports only local file connections.
Workbook file can be saved locally as *.twb or *.twbx format.	Workbook file cannot be saved locally, need to be saved and uploaded to Tableau Public Server
Privacy	No privacy
Can work offline	Must work online in order to save the analysis file

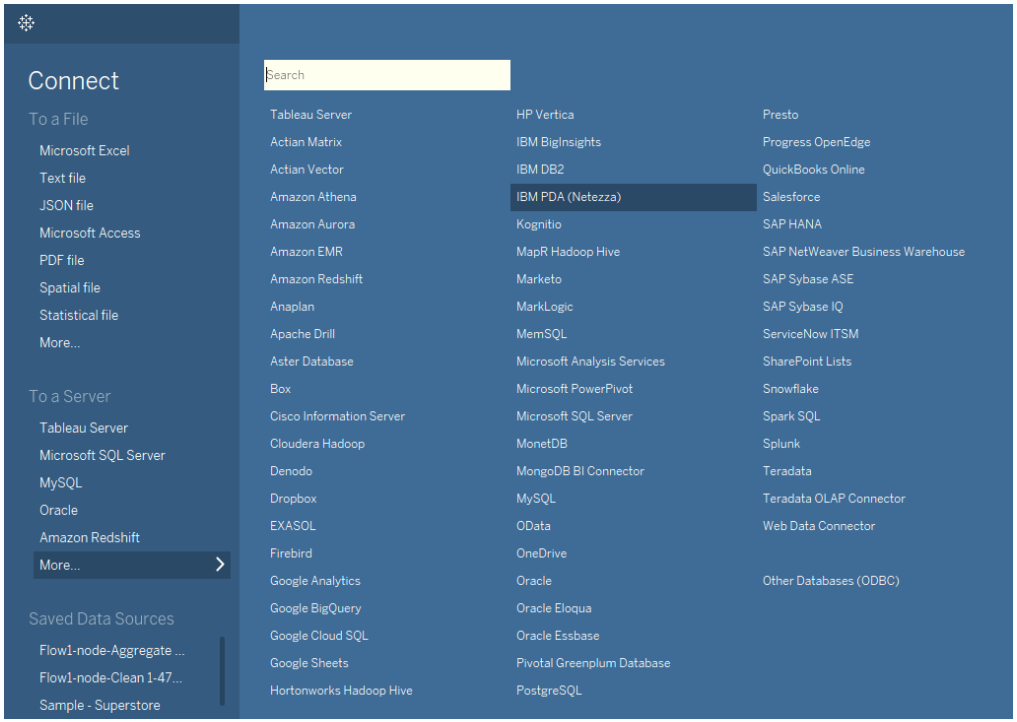
Tableau's Homepage

Open the Tableau by clicking on the Tableau icon . You will see a Tableau's Homepage , where you have the option to choose your data connection or open the existing file.



Connect To Data

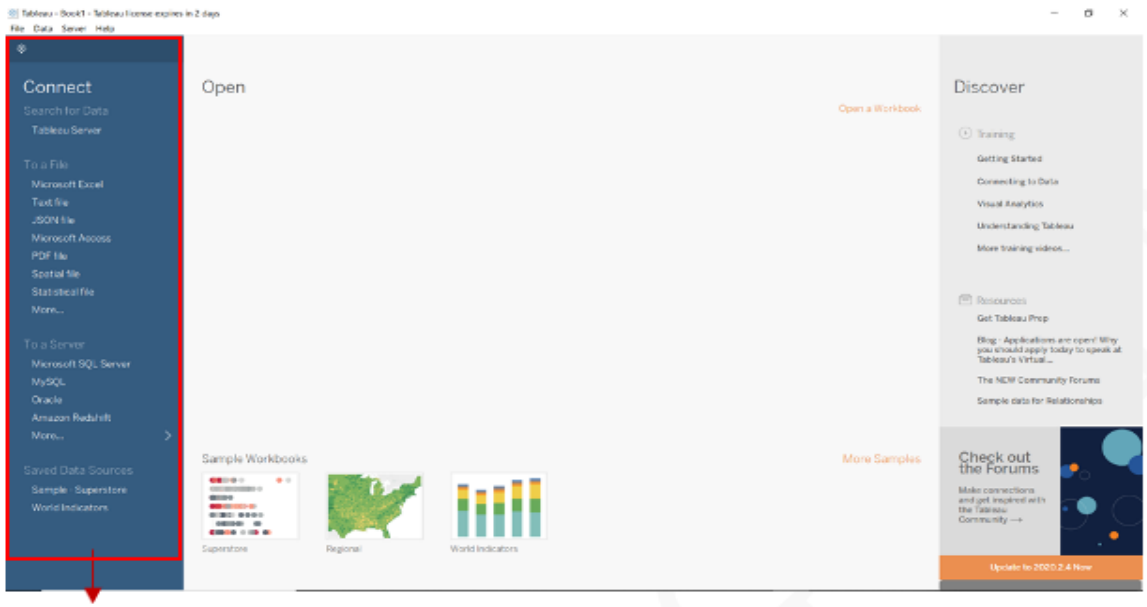
- Tableau can connect to many data sources regardless of whether they're local files or databased on-premises or in the cloud.
- A workbook can have multiple data source.
- Joining Tables and Unions



Connect To Data (cont'd)

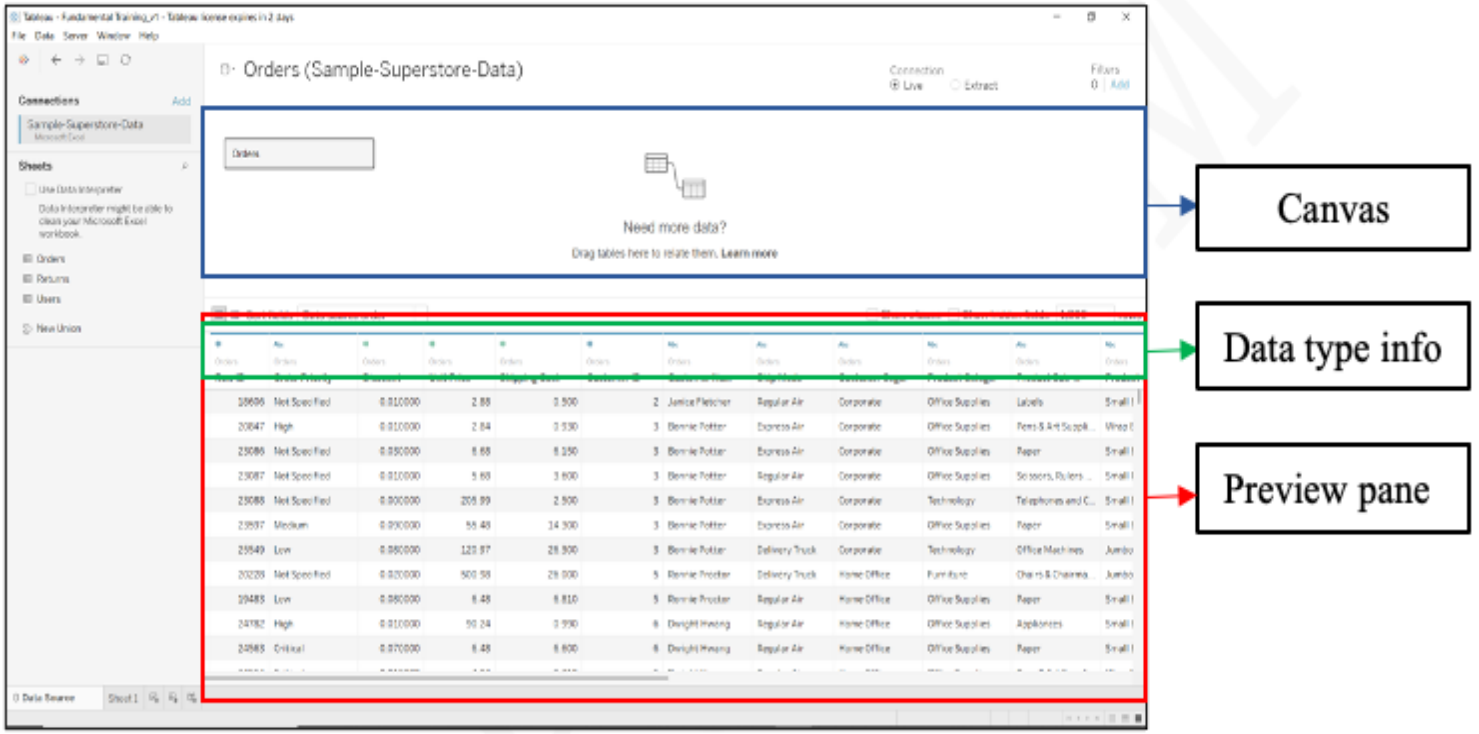
1. On the Connect Pane, click Microsoft Excel.

2. Navigate to the file on your machine (desktop/laptop) and open a file: **US APAC – Superstore Spreadsheet.xlsx**.



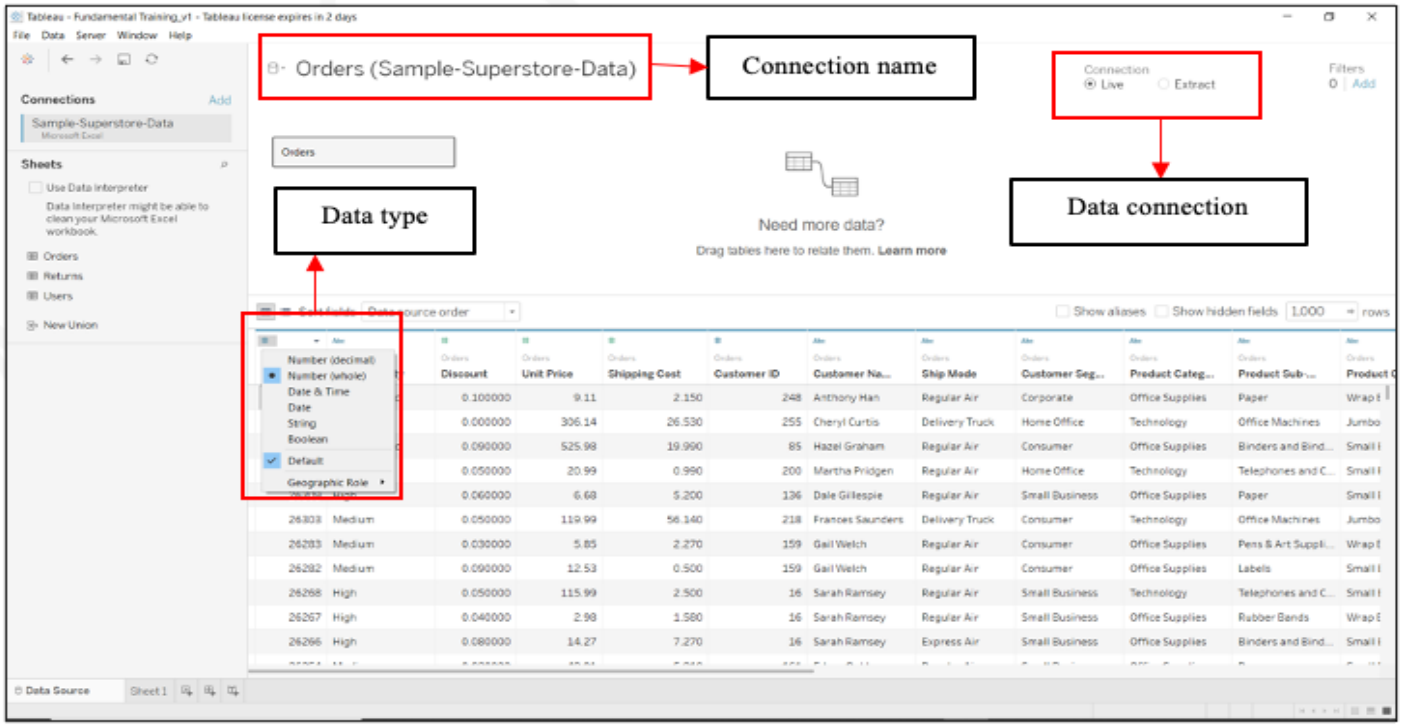
Connect To Data (cont'd)

3. Click and drag out the sheet named **Order** into the canvas (*Drag tables here*). Part of the data can be viewed on the preview pane.



Connect To Data (cont'd)

4. Once Tableau has read the data, it will assign the data type to each field. If it wasn't correctly assigned, you can make changes on the **Data Sources** page.



Connect To Data (cont'd)

Data Type

Icon	Data type
Abc	Text (string) values
	Date values
	Date & Time values
#	Numerical values
T|F	Boolean values (relational only)
	Geographic values (used with maps)

Connect To Data (cont'd)

Data Connection (Live Vs Extract)

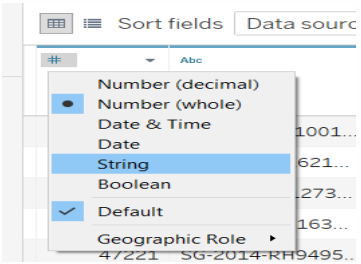
Live	Extract
Maintains a live connection with the data source.	Extract a copy of the data into Tableau environment.
Any changes in data source will be reflected in Tableau.	No changes unless data extract is being refreshed.
Performance as fast as your database.	Performance depends on your PC's resources.
Secure – User would need to reconnect the data source with credentials.	Not secure because data are not encrypted.
Need to be online if not connecting to local data.	Convenience - can be used offline.

Connect To Data (cont'd)

Data Preparation

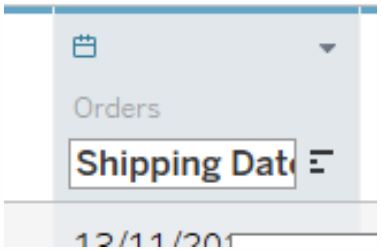
- **Changing Data Type**

You can change the data type of a column by clicking the data type icon at the top of the column.
For example, Row ID, although it is a number, but it is not numerical. Change the Row ID to String, by simply click on the icon and choose String.



- **Changing Name of Data Field**

Double clicking on its name and the name of data field can be renamed. For example, double click on the Ship Date, and change the name to Shipping Date. Repeat the same steps for Ship Mode.



Connect To Data (cont'd)

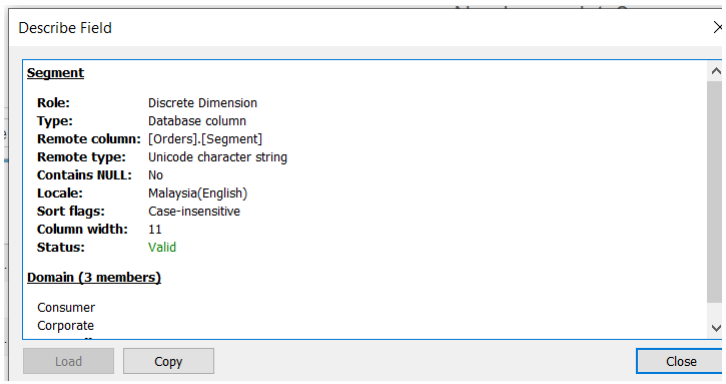
Data Preparation

- **Split**

Right-click on the data field and select Split. For example Order ID .Tableau will detect the hyphen as the separator and split the column into 4 new data fields.

- **EDA (Exploratory Data Analysis)**

Right-click on the data field and select Describe. You can see what is the role has been assigned by Tableau to this data field and the distinct values this column contains. For example select describe for segment field.

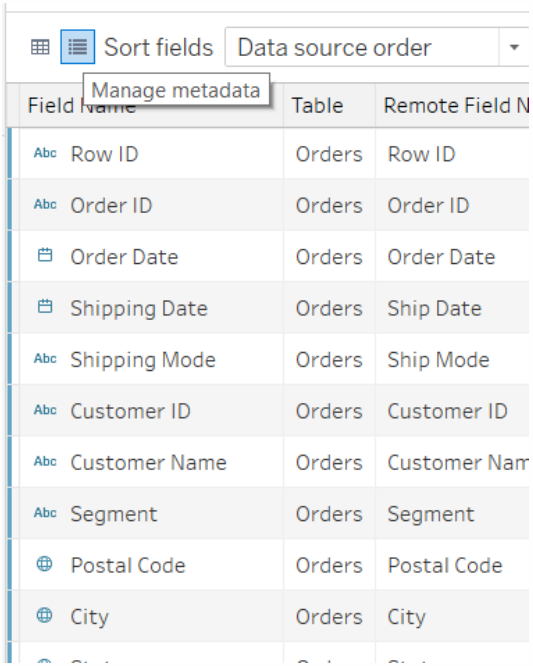


Connect To Data (cont'd)

Data Preparation

- **Metadata Grid View**

This can be a useful view, as the vertical layout can be easier to navigate, especially if you have large number of data fields. Useful when tables have been joined, to know where does each of the data field are actually coming from before joining.



Connect To Data (cont'd)

Data Preparation

- **Metadata Grid View**

This can be a useful view, as the vertical layout can be easier to navigate, especially if you have large number of data fields. Useful when tables have been joined, to know where does each of the data field are actually coming from before joining.

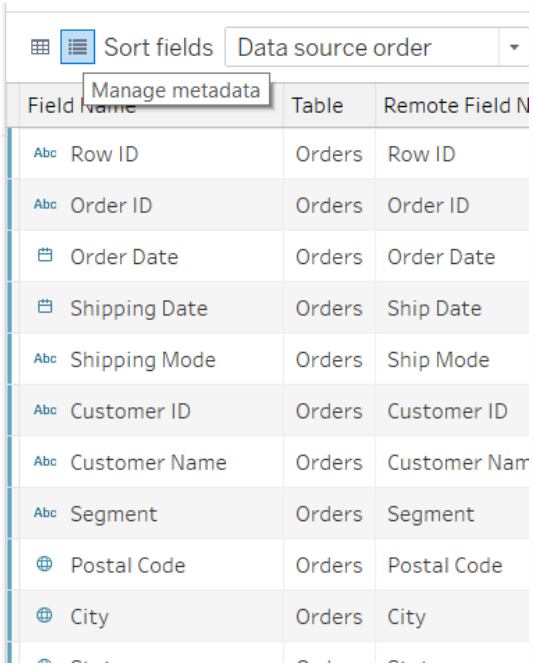


Tableau - Basics & Concepts

Click on **Sheet1** at the bottom left. The new worksheet interface will be prompted..

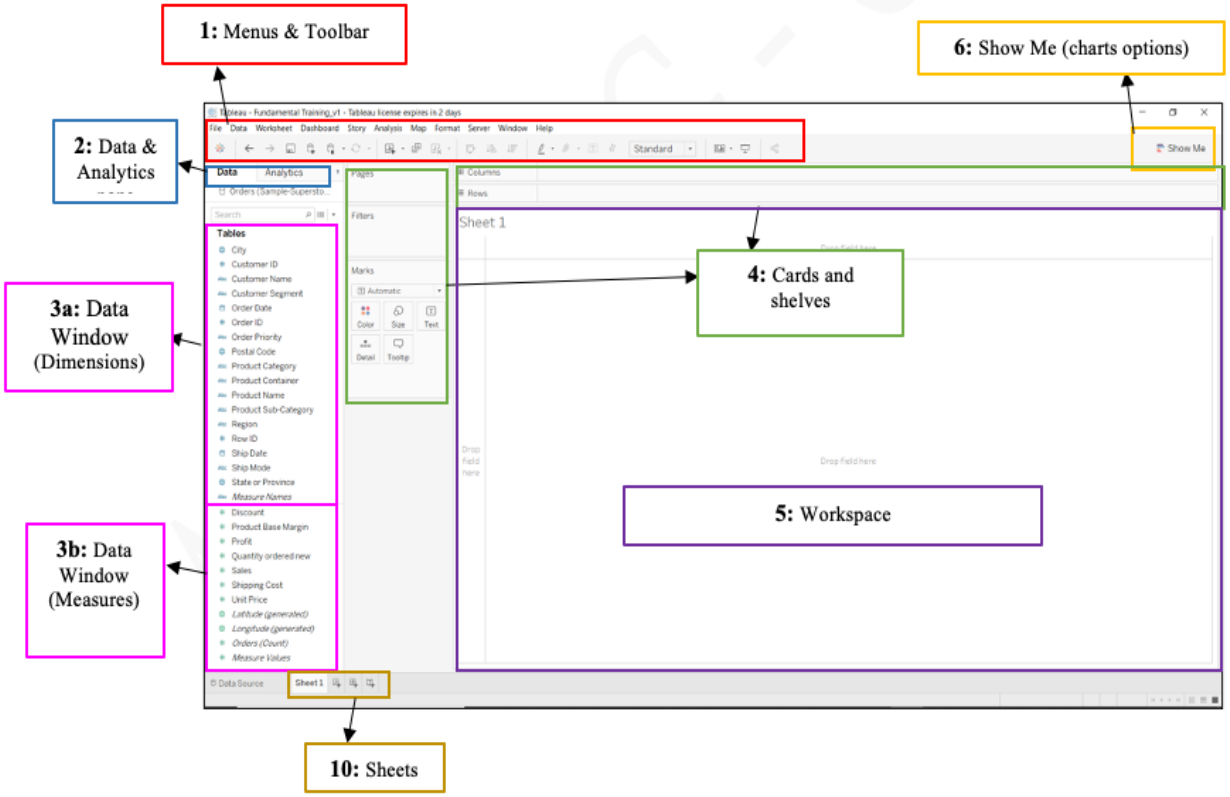


Tableau - Basics & Concepts (cont'd)

Dimension vs Measure

Dimension	Measure
Qualitative (categorical) data	Quantitative (numerical) data
String, Boolean, ID, date and time	Numerical data, latitude and longitude
Usually discrete (blue)	Usually continuous (green)
No aggregation will be applied when the data field	Aggregation will be applied automatically, by default - sum

Tableau - Basics & Concepts (cont'd)

Discrete vs Continuous

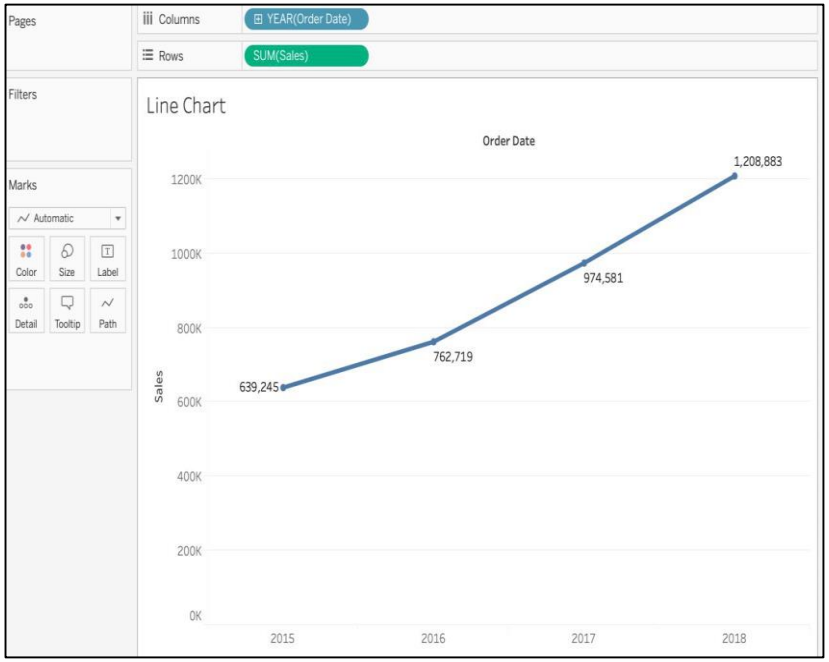
Discrete (blue)	Continuous (green)
Individual, separate values that do not contain values in between.	Can take on any value at any point within the range of values
Is counted	Is measured
Example: No of items in a transaction, no of person in a class	Example: Height, Windspeed
In Tableau, by default, all dimension values are discrete.	In Tableau, by default, all measure values are continuous.

Creating Basic Visualization Charts

Line Chart

Analysis 1: What are the trend of sales over time?

1. In the **Data** pane, select **Sales** and **Order Date**.
2. Go to **Show Me**, click on the recommended chart (highlighted).
3. In the **Marks** card, click **Label** and tick **Show mark labels**.
4. Rename sheet as "Sales Over Time".
5. Save your workbook.



Bar Chart

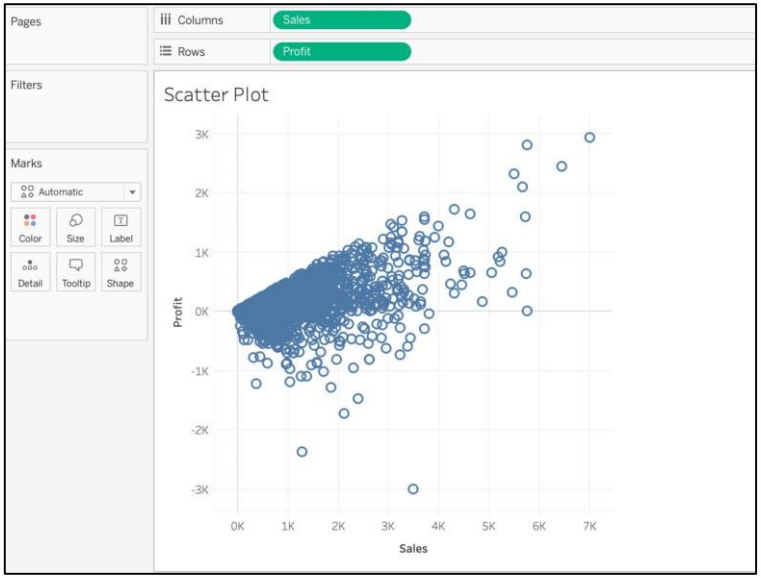
Analysis 2: Which product category has the highest sales?

1. Create a new worksheet.
2. In the **Data** pane, drag **Sales** to **Rows** shelf.
3. Drag **Category** to **Columns** shelf.
4. Go to Sales axis, untick Show Header.
5. In the **Marks** card, click **Label** > tick **Show mark labels.**
6. Optional: To make the visualization fit into your screen, click drop down menu at **Standard** toolbar, select **Entire View**.
7. Rename sheet as "Sales by Category"

Scatter plot

Analysis 3: Is the highest sales has highest profit?

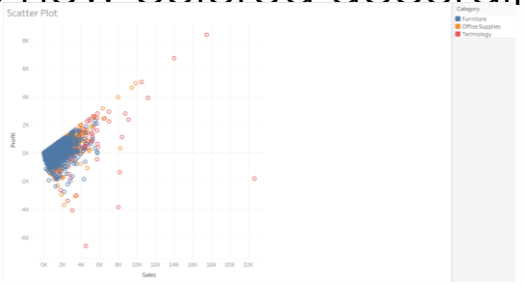
1. Create a new worksheet.
2. In the Data pane, highlight **Profit** and **Sales**.
3. Go to **Show Me**, click on **scatter plots**.
4. Go to **Analysis** menu, untick *Aggregate Measures*.



Scatter plot

Analysis 4: What is the Sales and Profit relationship based on category?

1. Drag **Category** data field and drop it in the Color marks card.
   The data points are now colored according to their respective Category.

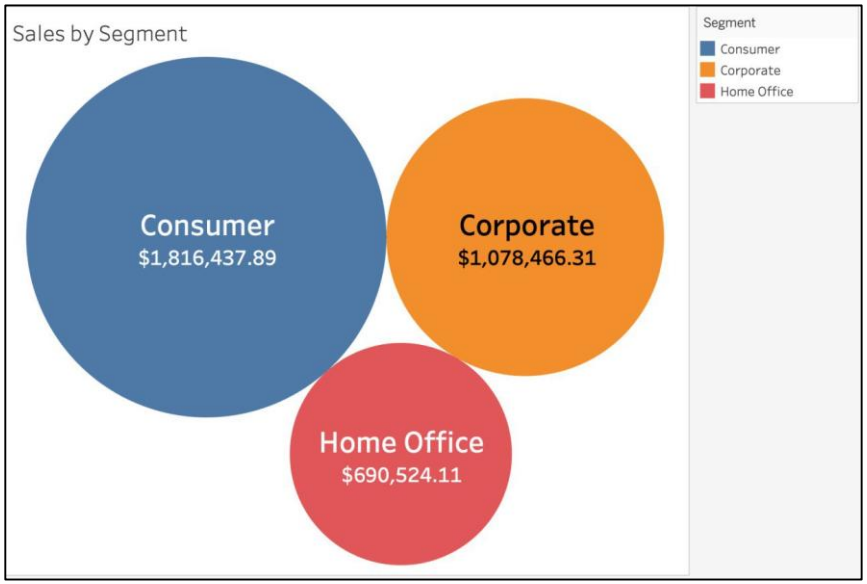


2. The Scatter Plot also can be break down into 3 scatter plot, each for a Category by dragging the Category field and drop it to the Column shelf

Packed-Bubbles

Analysis 5: Which segment has the highest sales?

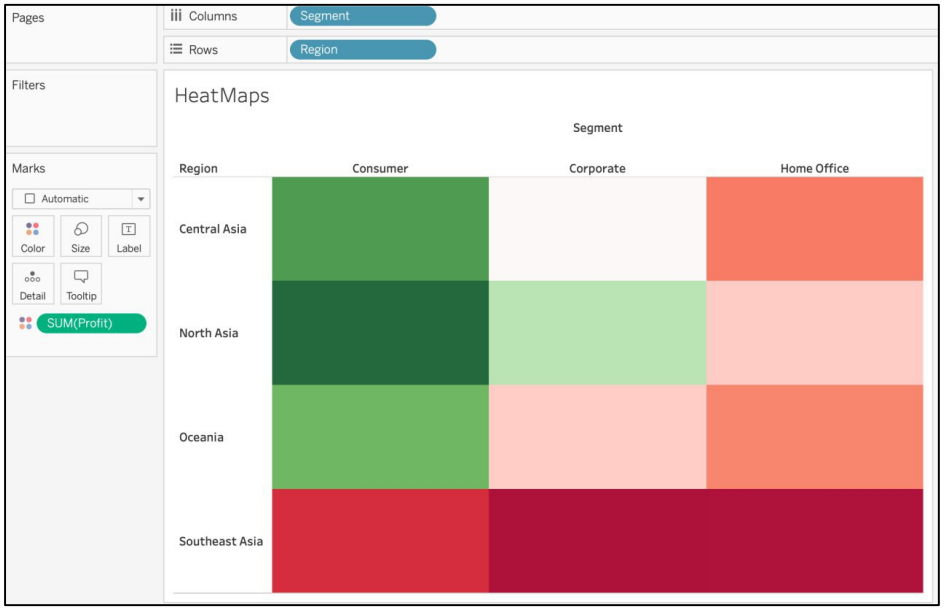
1. Create a new worksheet and rename as "Sales by Segment".
2. In the **Data** pane, highlight **Sales** and **Segment**.
3. Go to **Show Me**, click on packed bubbles chart.
4. Drag **Sales** to **Label** on **Marks** card.
5. Optional: Click on **Presentation Mode** toolbar or hit function F7 on your keyboard.



Heat Map

Analysis 6: How the profit varies across regions and customer segments?

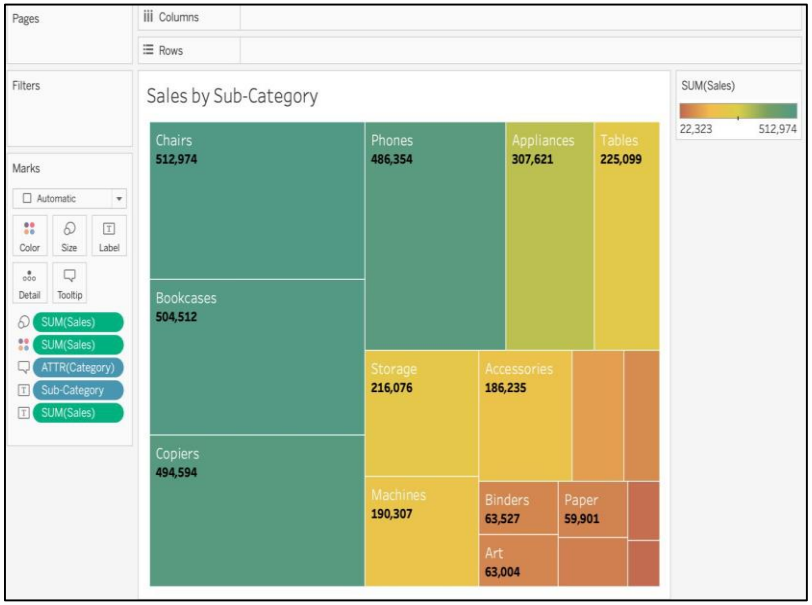
1. Create a new worksheet.
2. Drag **Segment** to **Columns** shelf.
3. Drag **Region** to **Rows** shelf.
4. Drag **Profit** to **Color** on the **Marks** card.
5. Change **Standard** view to **Entire View**.



Tree Map

Analysis 7: Which product sub-category has the highest sales?

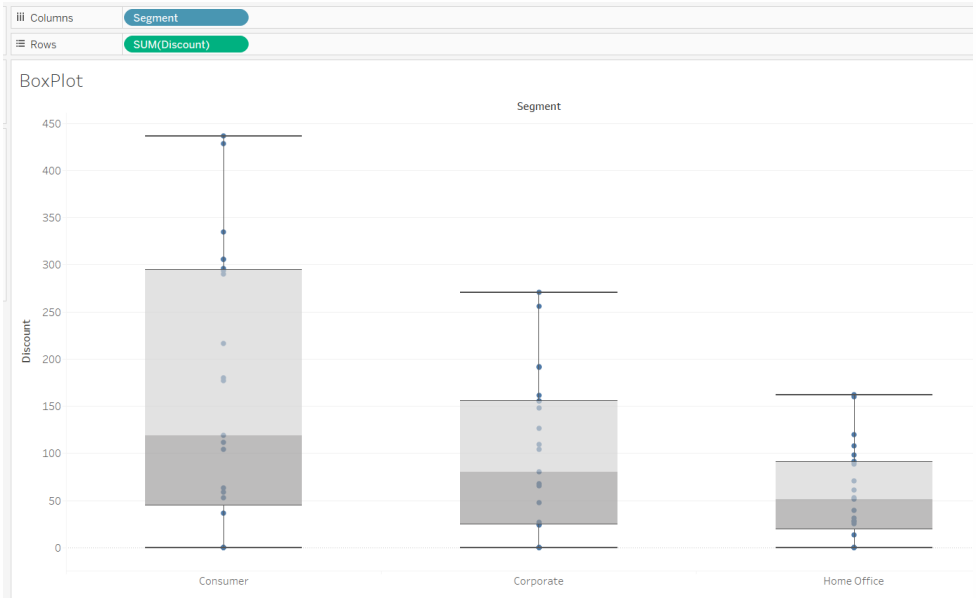
1. Create a new worksheet.
2. In **Data** pane, highlight **Sales** and **Sub-Category**.
3. Go to **Show Me**, click on **treemaps.**
4. Optional: Drag **Category** to **Tooltip** on **Marks** card.
5. Optional: Drag **Sales** to **Label** on **Marks** card.
6. Rename your sheet as "Sales by Sub-Category".



Box Plot

Analysis 8: How the discount varies across regions and customer segments?

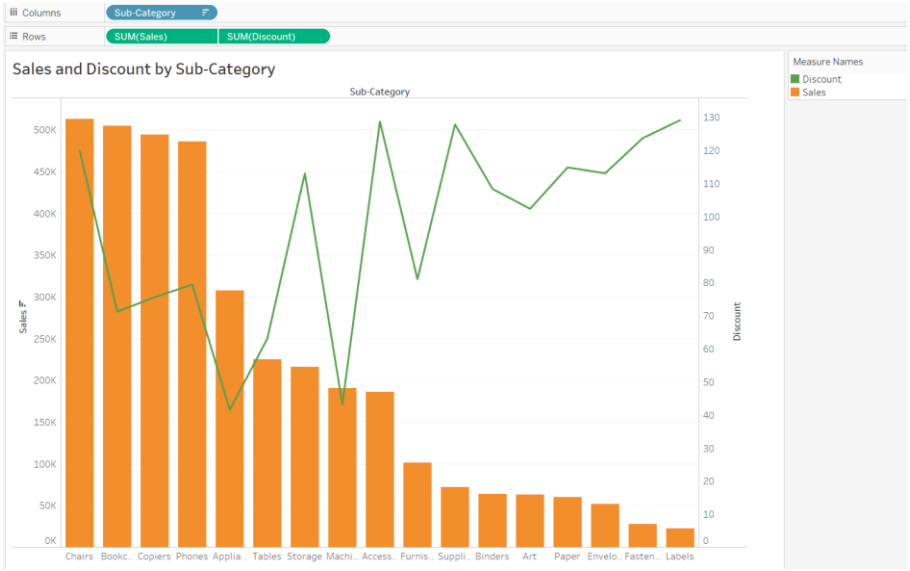
1. Create a new worksheet.
2. Drag **Segment** to **Columns** shelf. Then, drag **Region** to **Columns** shelf and drop it to the right of **Segment.**
3. Drag **Discount** to **Rows** shelf.
4. Click **Show Me** in the toolbar, then select **box-and-whisker plots** chart type.



Combination Charts

Analysis 9: Compare sales and discount by sub-category.

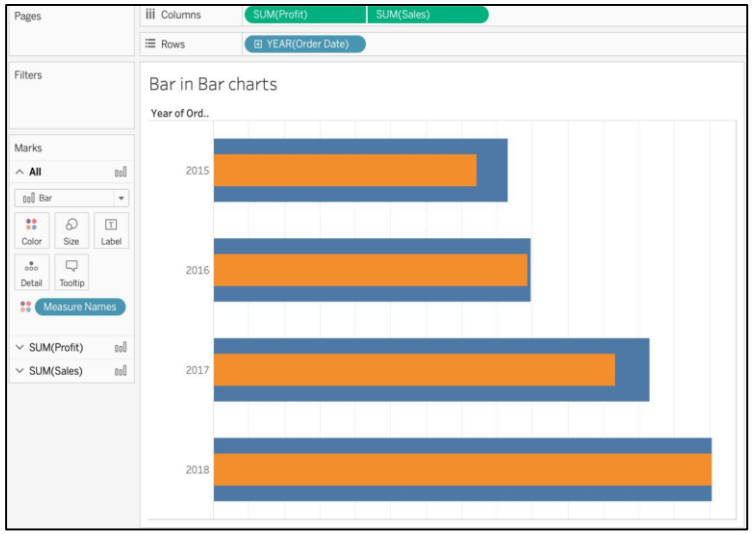
1. Create a new worksheet. Rename it as "Sales and Discount by Sub-Category"
2. On **Data** pane, drag **Sub-Category** to **Column** shelf.
3. Drag **Sales** and **Discount** to **Row** shelf.
4. Click on **SUM(Discount)** pill, select **Dual Axis**.
5. Click on **SUM(Sales)** at **Marks** card and change the chart type to **Bar**.
6. Click on **SUM(Discount)** at **Marks** card and change the chart type to **Line**.
7. Click on the sort icon at the **Sales** axis to sort the data descending by sales.



Overlapped Bar Chart

Analysis 10: Compare sales and profit in one chart.

1. Create a new worksheet. Rename as "Bar in Bar charts".
2. On **Data** pane, drag **Order Date** to **Columns** shelf.
3. Drag **Sales** and **Profit** to the **Rows** shelf.
4. Right-click the second measure on the **Rows** shelf, and select **Dual Axis**
5. On the **Marks** card labeled **All**, set the mark type to **Bar** in the dropdown menu.
6. On the **Marks** card labeled **SUM(Sales)**, click **Size** and then adjust the slider to change the width.
7. Repeat step 6 on the Marks card labelled **SUM(Profit)**.



Organizing Data

Hierarchies and Drill-Down

1. In the **Data** pane, drag **Sub-Category** and drop it directly on top of **Category** field.
2. When prompted, enter a name for the hierarchy and click **OK**.
3. Drag additional fields into the hierarchy as needed. You can also re-order fields in the hierarchy by dragging them to a new position.

Sorting

1. Go to *Sales by Category* sheet and duplicate it. Rename as "Sales by Sub-Category".
2. In the **Data** pane, drag **Sub-Category** to **Columns** shelf and drop it after **Category** pill.
3. Swap the columns to row by clicking the swap icon 
4. Right-click on **Sub-Category** pill. Select **Sort**. The Sort dialog box will be prompted. Do configuration as follows: **Sort By** > select **Field**; **Sort Order** > **Descending**

Grouping

1. Open *Sales by Sub-Category* sheet.
2. Hover to **Office Supplies**. Click on **Binders**.
3. Hit **shift** button on your keyboard, then click the last item, i.e., **Labels**.
4. Right-click on highlighted items, select **Group**.
5. To rename the group, right-click on that grouped items, click Edit Alias.
6. In the Edit Alias dialog box, type new name for the group. For example, "Small Items".

Filtering

We use filter to select a subset of the data in the view to focus on that data or to exclude data from the view.

Sets

Analysis 11: How many percent of customers that has sum of sales over $5000?

1. Create a new worksheet. Rename it as "Set".
2. In the **Data** pane, right-click on **Customer Name** > select **Create** > select **Set.**
3. In the Create Set dialog box,
    a. Type a name for the set. Let's put "Customers A".
    b. On **General** tab, click on **Use all** button. This is to make sure the condition always applies to all values even when new customers are added.
    c. Hit the **Condition** tab. Click **By field** button. Then, define a condition that only includes customers when sum of sales is greater than 5000.
    d. Click **OK**.
4. In the **Data** pane, drag the **Customers A** to **Row** shelf. Then, drag **Sales** to **Columns** shelf.
5. On the **SUM(Sales)** pill, right-click and select **Quick Table Calculation** > **Percent of Total**.
6. Click **Label** on the **Marks** card, tick **Show marks label**.

Combine Sets

Analysis 12: Determine the number of customers who purchased in both year, 2017 and 2018.

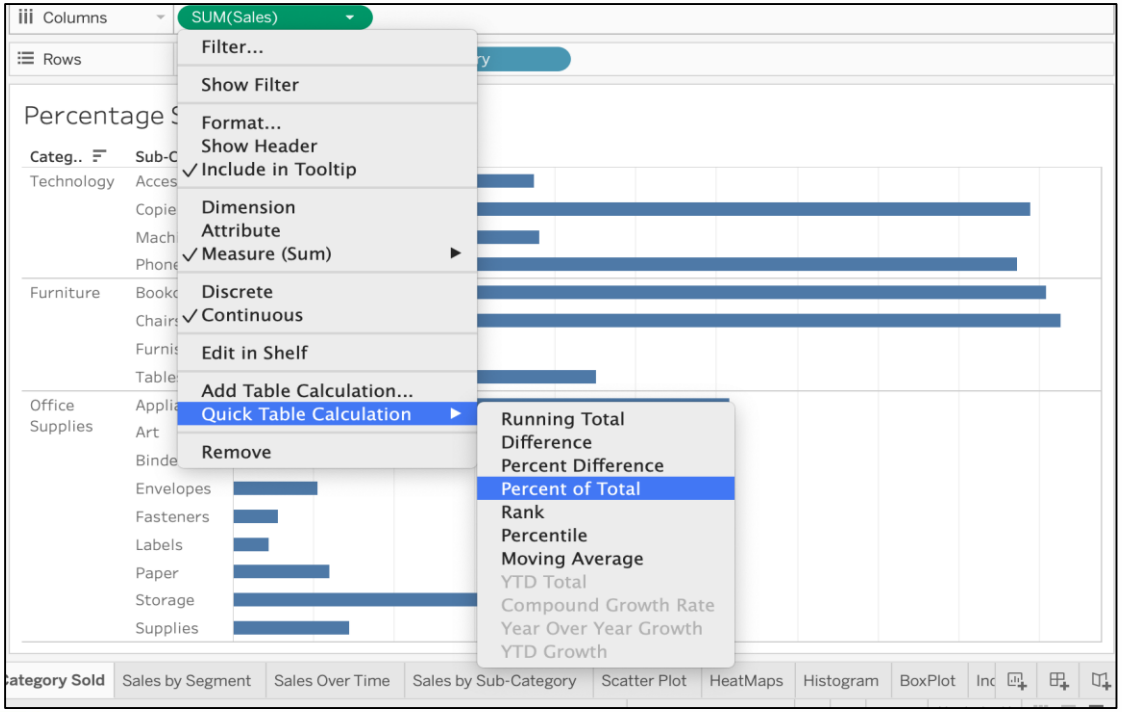
1. Create a new worksheet. Rename as "Combine Set"
2. In Data pane, drag **Customer Name** to **Row** shelf.
3. Drag **Order Date** to **Filters** card. In the *Filter Field* dialog box, select **Year** > click **Next** > tick on 2018 > click **OK**.
4. Back to the view, press key **Ctrl+A** on your keyboard to select all of the customers.
5. Right-click on the selection and select **Create Set**.
6. In the *Create Set* dialog box, type a name for the set (example: Customers 2018). Then, click **OK**.
7. In the **Filters** card, right-click **Order Date** pill and select **Edit Filter**.
8. In the **Filter Field** dialog box, change the year to 2017 instead of 2018, then click **OK**.
9. Repeat step 4 - 6.

Combine Sets (cont'd)

10. In the **Data** pane, select both sets: **Customers 2017** and **Customers 2018** by holding the **Ctrl** key on your keyboard as you select. Right-click the selection and click on **Create Combined Set**.
11. In the **Create Set** dialog box, type a name for the new set (example: Customers 2017 & 2018). Next, select the option: *Shared members in both sets* (**Note:** This option is depending on your analysis). Then, click **OK**.
12. Clear the worksheet by click  at the Toolbar.
13. In **Data** pane, drag **Customer Name** to **Row** shelf. Then, right-click the **Customer Name** pill > select **Measure** > select **Count (Distinct)**.
14. Drag the **Customers 2017 & 2018** field to **Filters** card.
15. In the **Marks** card, click **Label** > tick *Show mark label*. You will see the number of customers who purchased in 2017 and 2018 are 605.

Quick Table Calculation

1. Go to *Sales by Sub-Category* sheet. Duplicate this sheet. Rename as "Percentage of Sub-Category Sold"
2. Right-click the **SUM(Sales)** pill, select **Quick Table Calculation** > select **Percent of Total**. The marks label will be changed from aggregation sum to percentage.



Session 3:

Application of Data Analytics with Tableau

(Part 2)

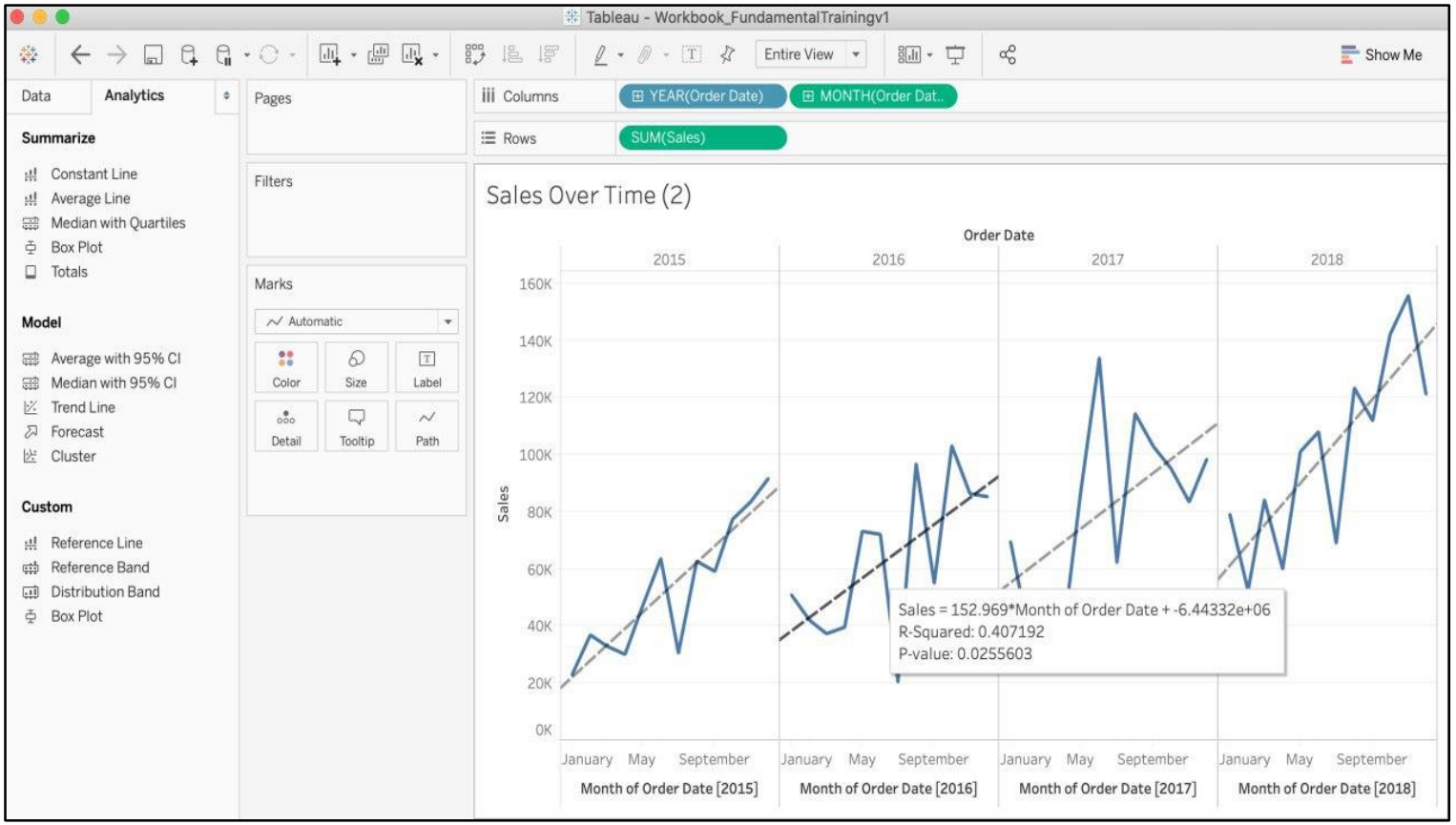
Analytics Tab

Trend Line

Analysis 13: How the relationship of sales over time?

1. Go to *Sales Over Time* sheet. Duplicate and rename as "Sales Trend Line".
2. Click '**+**' on the **Year(Order Date)** pill to drill-down till **Month(Order Date)**. Bring out the **Quarter (Order Date)**.
3. In the **Analytics** pane, drag the **Trend Line** to the view and a few possible trend line types to choose will be turned on (Linear, Logarithmic, Exponential, Polynomial, or Power model types).
4. Drop it on the **Linear** type to create a linear trend line in the line chart.
5. Hover over the trend line to display the statistics information including the equation, R-squared and P-values.

Trend Line (cont'd)



Forecasting

Analysis 14: What is the sales forecast value for the next 6 months?

1. Go to **Sales Over Time** sheet. Duplicate this sheet and rename as "Forecast".
2. Right-click on the trend line and untick **Show Trend Lines**.
3. In the **Analytics** pane, drag Forecast to the canvas. You will see the line chart will be expanded with the forecast values shown by a light blue shaded color. Tableau will automatically forecast up to next 13 months.
4. If you want to change the default configuration for the forecast, right-click on the forecast line, select **Forecast** > click on **Forecast Options.**
5. In the **Forecast Options** dialog box, do the configuration as you wish.

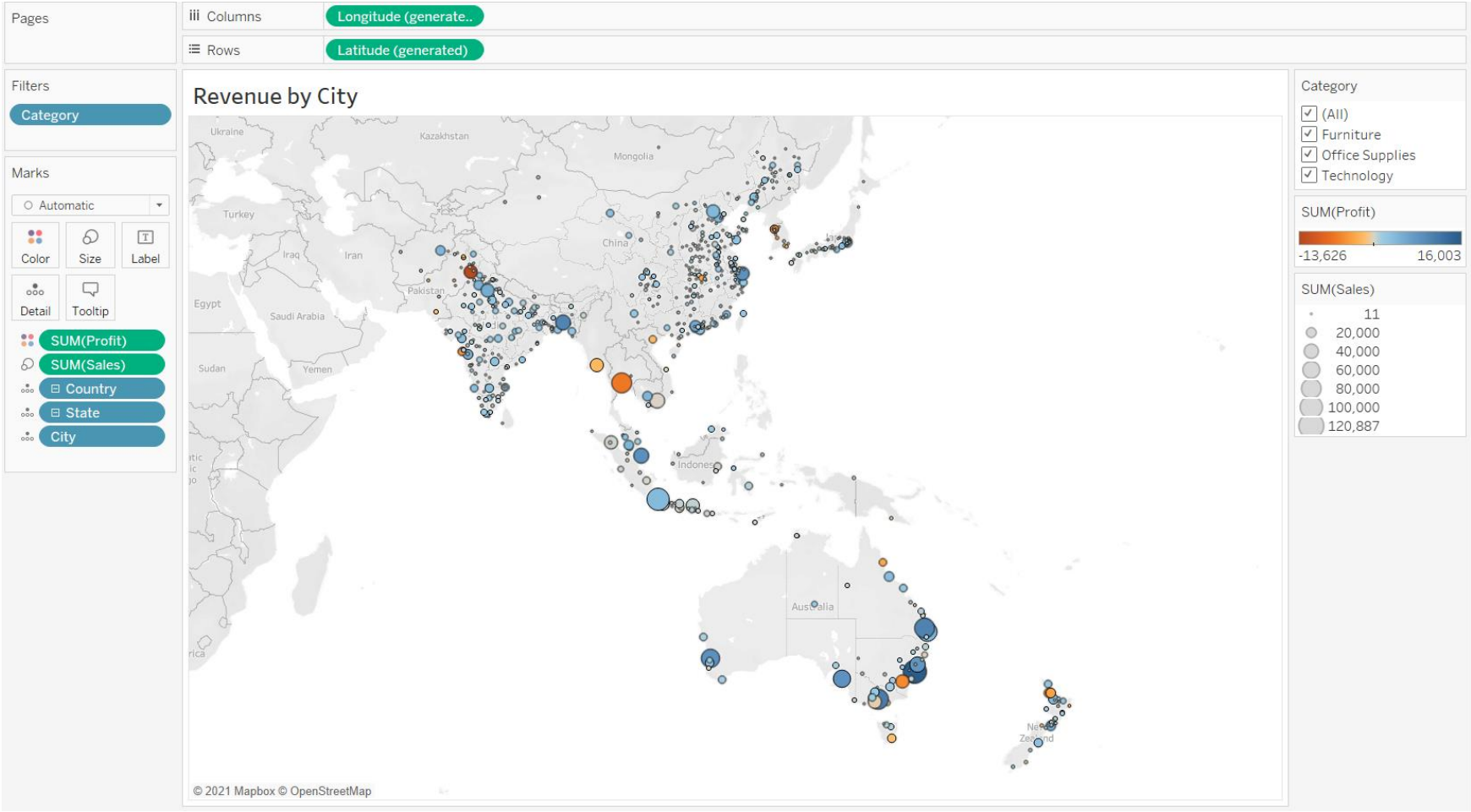
Mapping the Data

Map Data Geographically

Analysis 15: Where is data coming in from? Which state has the highest lost?

1. Create a new worksheet.
2. In the **Data** pane, double click on **State**. Note that, the generated **Latitude** and **Longitude** were automatically plotted on the **Rows** and **Columns** shelf.
3. Drag **Profit** to **Color** on **Marks** card**.** This should create a filled map colored on a gradient of 'SUM(Profit)'. Note that the default aggregation is SUM.
4. Hover over the darkest orange state to see which state has the highest lost.
5. Then, drag **Sales** to **Size** on the **Marks** card.
6. Resize the shape of the circle by click **Size** and then adjust the slider to change the width.
7. Optional: To put border on the circle, click on **Color** at the **Marks** card, select **Border** > choose color.
8. Optional: You can extend you analysis by adding another measure to **Filter** card. For example, Category. On the **Data** pane, right-click on **Category** field and select **Show Filter**.
9. Optional: You can expand **State** to **City** by clicking the '**+**' symbol on **State** pill on **Marks** card.
10. Rename this sheet as "Revenue by City".

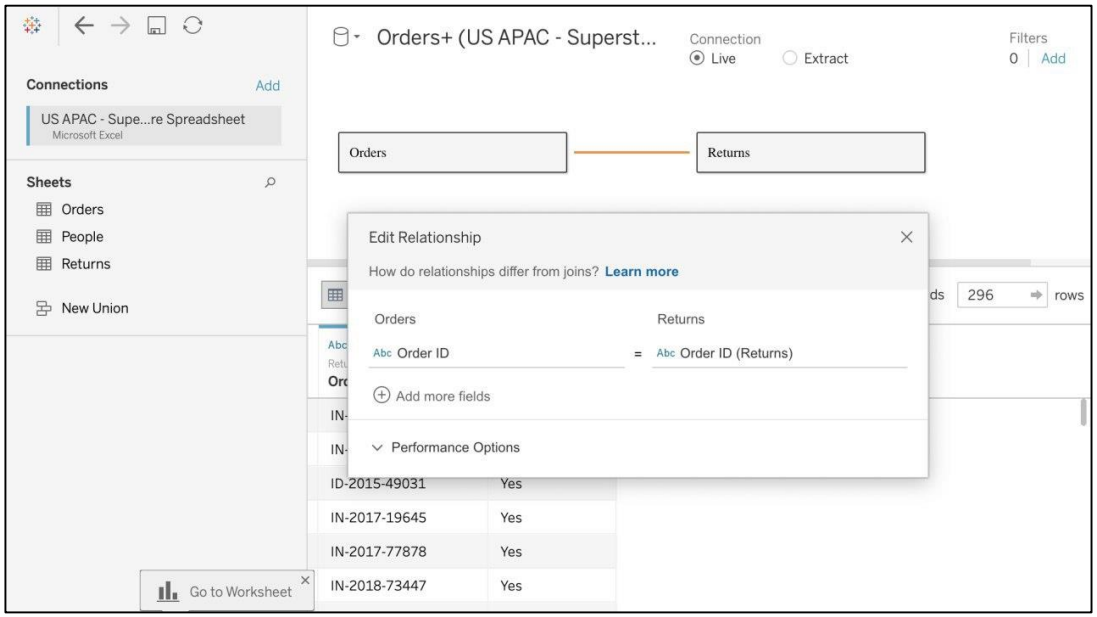
Map Data Geographically (cont'd)



Data Blending

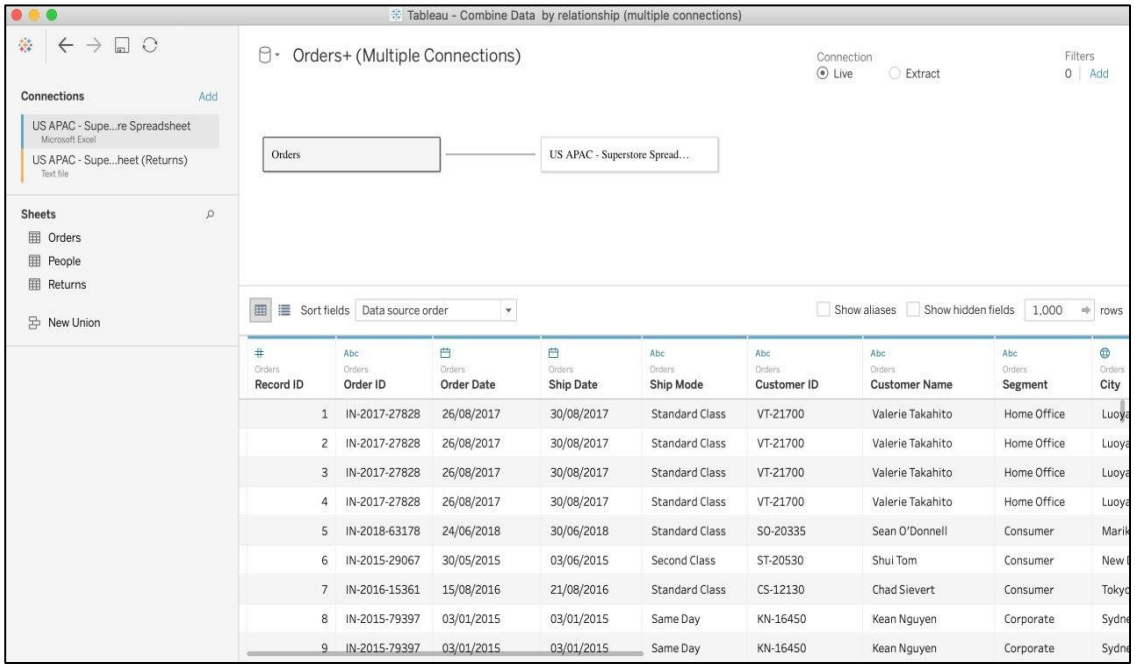
Single Connection

1. On your workbook, go to **Data Source** window.
2. In the **Sheets** pane, drag another table (i.e., **Returns sheet**) to the canvas. When you see the "noodle" between the two tables, drop that table.



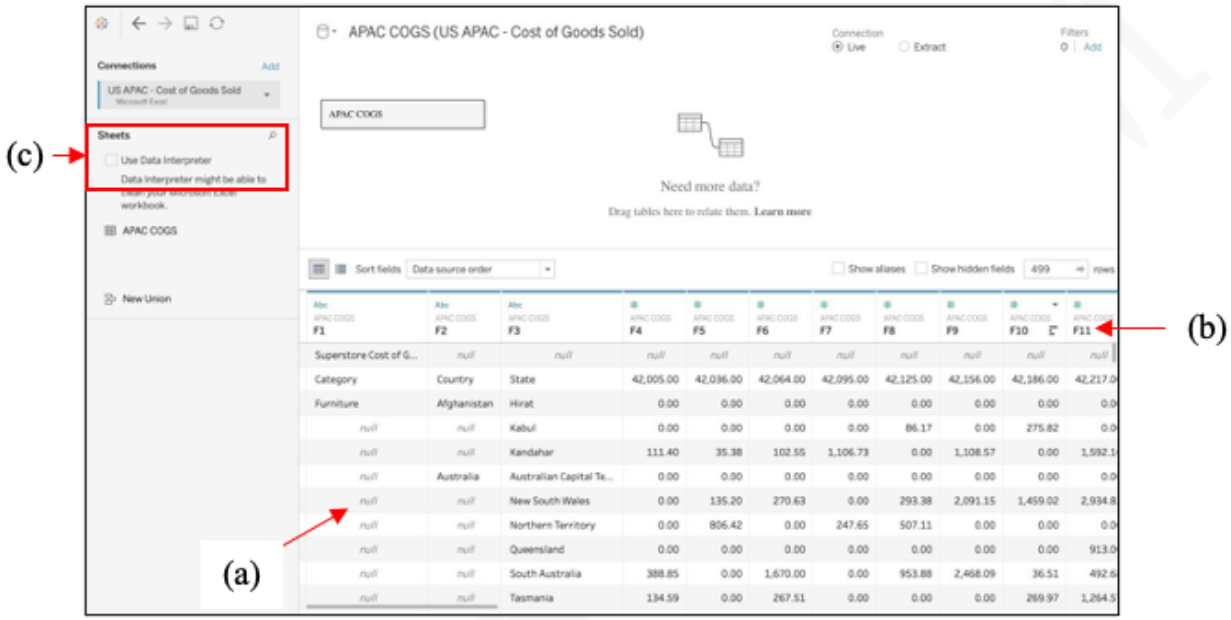
Multiple Connections

1. On your workbook, go to **Data Source** window.
2. In the **Connections** pane, click **Add**.
3. Connect to another data source (i.e., APAC Superstore Spreadsheet (Returns).csv).



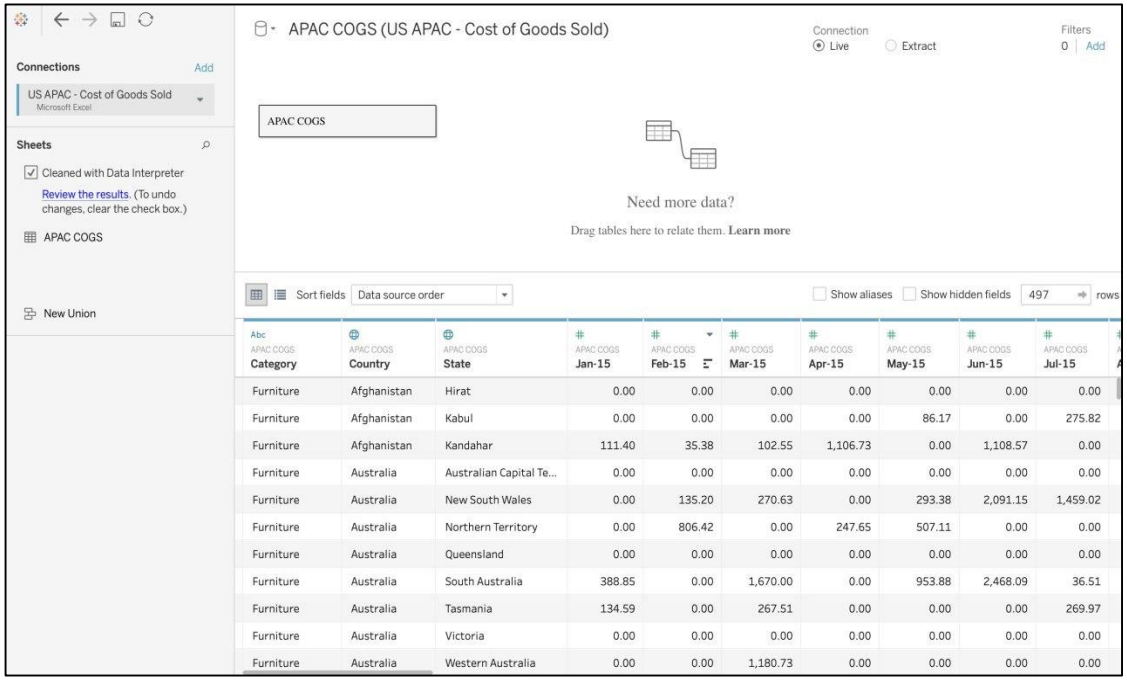
Blend

1. Create a new worksheet.
2. Drag **Profit** to Rows shelf. Then, drag **Category** to **Color** on **Marks** card.
3. To add another data source, click on icon in the Toolbar.
4. Choose **More..**, and navigate to the M Excel file: *US APAC - Cost of Goods Sold*, and click **Open**. The preview of data will be prompted as follows.



Blend (cont'd)

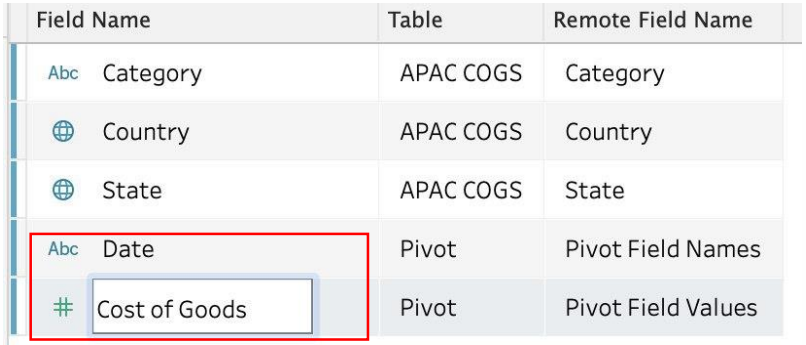
5. Based on the view, we can see that there is (a) *null* in some of the cells and, (b) no appropriate column name. This data requires data cleansing. In Tableau, we can use (c) - **Use Data Interpreter** to fix this problem. After you tick on the button, your data will be look like as shown below.



Category	Country	State	Jan-15	Feb-15	Mar-15	Apr-15	May-15	Jun-15	Jul-15
Furniture	Afghanistan	Hirat	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Furniture	Afghanistan	Kabul	0.00	0.00	0.00	0.00	86.17	0.00	275.82
Furniture	Afghanistan	Kandahar	111.40	35.38	102.55	1,106.73	0.00	1,108.57	0.00
Furniture	Australia	Australian Capital Te...	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Furniture	Australia	New South Wales	0.00	135.20	270.63	0.00	293.38	2,091.15	1,459.02
Furniture	Australia	Northern Territory	0.00	806.42	0.00	247.65	507.11	0.00	0.00
Furniture	Australia	Queensland	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Furniture	Australia	South Australia	388.85	0.00	1,670.00	0.00	953.88	2,468.09	36.51
Furniture	Australia	Tasmania	134.59	0.00	267.51	0.00	0.00	0.00	269.97
Furniture	Australia	Victoria	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Furniture	Australia	Western Australia	0.00	0.00	1,180.73	0.00	0.00	0.00	0.00

Blend (cont'd)

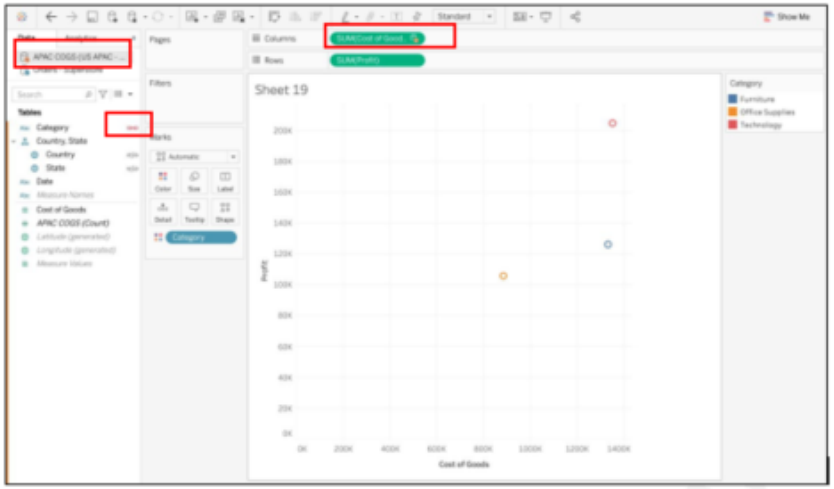
6. Now, the problem with null data has been solved. However, the cost values are separated by date and the date have become column names. We need to fix this by changing it into a single column instead of rows.
7. Click or (**Manage metadata** icon). Then, click on **Jan-15** column. On your keyboard, hit and hold the **Shift** button, and click on **Dec-18** column.
8. Right-click on the highlighted area, and select **Pivot**.
9. After pivoting, the highlighted columns will be transformed into two columns only. Rename Pivot Field Names as **Date**, and Pivot Field Values as **Cost of Goods**.



Field Name	Table	Remote Field Name
Abc Category	APAC COGS	Category
Country	APAC COGS	Country
State	APAC COGS	State
Abc Date	Pivot	Pivot Field Names
# Cost of Goods	Pivot	Pivot Field Values

Blend (cont'd)

10. Go back to preview data source by clicking on the icon
11. Next, go back to the most recent sheet. Look on the Data tab, you can see there are two different data sources. The one with blue tick is the primary data source for the analysis.
12. In the Data pane (under APAC COGS), drag **Cost of Goods** to **Columns** shelf.



You will see there is an orange tick at APAC COGS data source, which indicate this is a secondary data source in this analysis.

Both of these data sources are linked by one unique key. In this case, Tableau detect both data source contains Category field, hence it automatically creates link using this field.

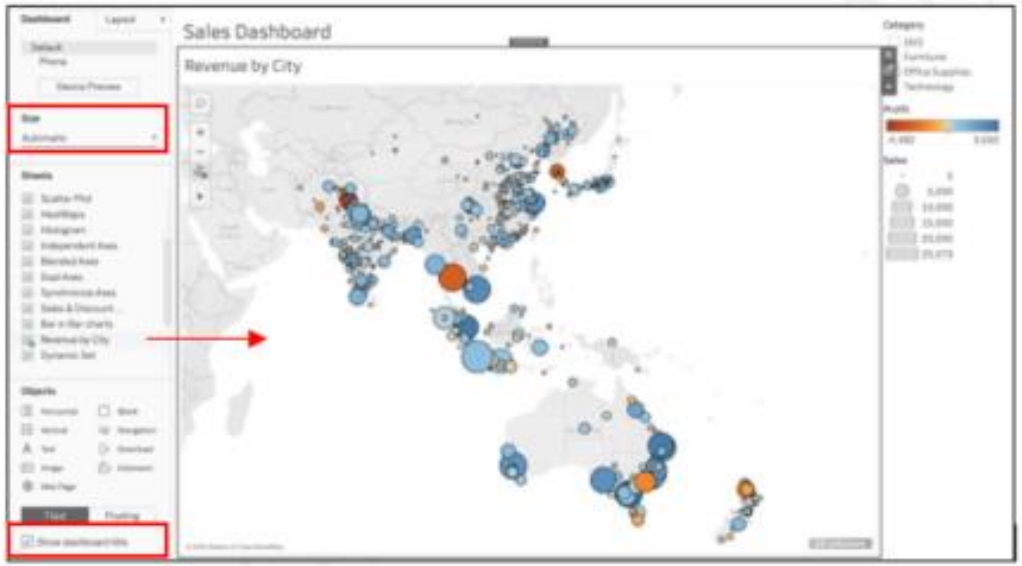
Blend (cont'd)

13. Change back into Orders – Superstore data source. Drag **State** into center of the canvas.
14. Do some adjustment by changing the mark type to **Circle**. Then, alter the color by reducing the opacity and select black color for border.
15. Rename this sheet as "Cost vs Profit".

Dashboard & Stories

Create Dashboard

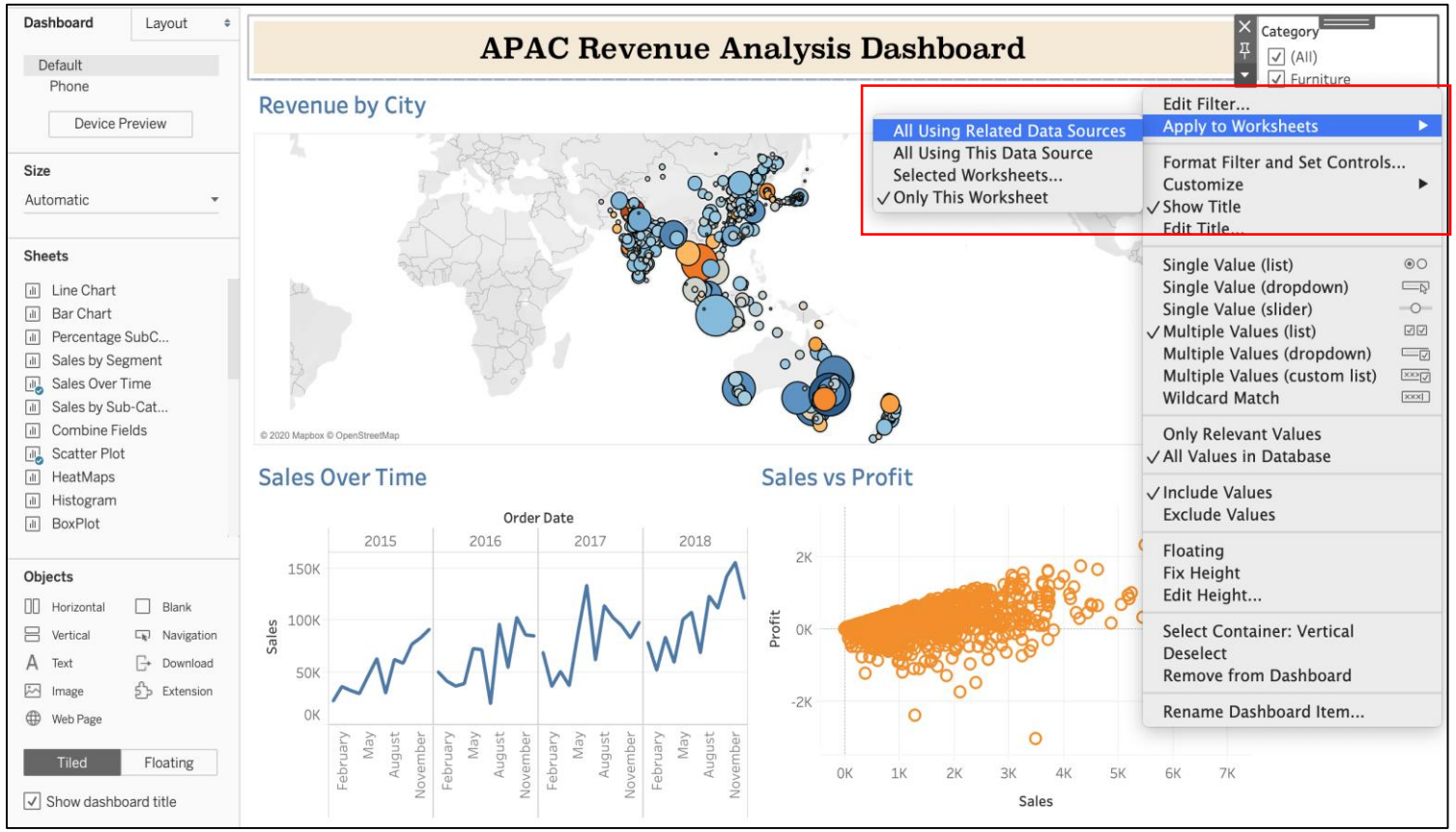
1. Click the **New Dashboard** icon located at the bottom of the workbook. Rename as "Sales Dashboard".
2. In the **Dashboard** pane, tick on **Show dashboard title**. (Optional) Change size to **Automatic**.



Create Dashboard (cont'd)

3. From the **Sheets** list at left, drag **Revenue by City** to your dashboard at right.
4. Next, drag **Sales Over Time** sheet to the bottom. Note: As you drag the sheet around the dashboard, a grey shaded area indicates the various places you can drop it.
5. Drag **Scatter Plots** to the bottom right. Rename the title as "Sales vs Profit".
6. To add interactivity to dashboard to enhance users' data insights, we can use filter as one the options. To configure this, click the caret on **Category** card to bring up the menu. Then, select **Apply to Worksheets** > select **All Using Related Data Sources**

Create Dashboard (cont'd)



Create Dashboard (cont'd)

Before we proceed to the next sub-topic, kindly create another dashboard that contains the following worksheet:

a) Sales by Category (use Bar Chart worksheet).
b) Sales and Discount by Sub-Category
c) Percentage of Sub-Category Sold.

Set Sales by Category as a filter. Rename dashboard as "Category Dashboard".

Create Stories

1. Click the **New Story** icon located at the bottom of the workbook.
2. In the **Story** workspace, right-click on the Story 1 title, select **Edit title**.
3. In the Edit Title dialog box, highlight the whole <*Sheet Name*> and change to "APAC Superstore Revenue Analysis".
4. In the **Story** pane, drag **Sales Dashboard** to the story at right.
5. Click on the **Navigator box** to add a caption on the point. For this example, write "This is the overall sales analysis in APAC region". You can re-format the font type, size, etc. by clicking the **Story** menu, select **Format**.
6. Add new story point by clicking **Blank** at Story pane.
7. Drag **Category Dashboard** to the canvas. Add a caption to tell the story.