

Data Visualization Using Tableau



Connect Data

- Open Tableau
- Click on connect and type of file
 - Various types which can be connected

The screenshot shows the 'Connect' dialog box from Tableau. At the top, there's a 'To a File' section with a 'More...' button. Below it is a 'To a Server' section with 'OData' and 'More...' buttons. A right-pointing arrow is located in the top right corner. The main area lists various data source types:

- Microsoft Excel
- Text file
- JSON file
- Microsoft Access
- PDF file
- Spatial file
- Statistical file

Working on Excel Workbook

- If workbook has several worksheets
 - Drag the worksheet to canvas to start
 - Click on Sheet1 to start the analysis

The screenshot shows the Microsoft Power BI Data Editor interface. On the left, under 'Connections', 'Sample - Superstore Excel' is selected. Below it, the 'Sheets' section lists 'Orders', 'People', 'Returns', and 'New Union'. A curved arrow points from the 'Orders' sheet name to the 'Orders' column header in the main table. The main table has four columns: 'Order ID', 'Order Date', 'Ship Date', and 'Ship Mode'. The first two rows of data are visible.

Order ID	Order Date	Ship Date	Ship Mode
CA-2013-152156	11/9/2013	11/12/2013	Second Class
CA-2013-152156	11/9/2013	11/12/2013	Second Class

Data Pane

- Columns from data are shown as fields on the left side in **Data pane**
- Fields are divided into :
 - Dimensions
 - Categorical data eg. Country, Product ID, Order date
 - They are blue in colour
 - Measures
 - Holds numeric data eg. Sales, profit
 - They are green in colour

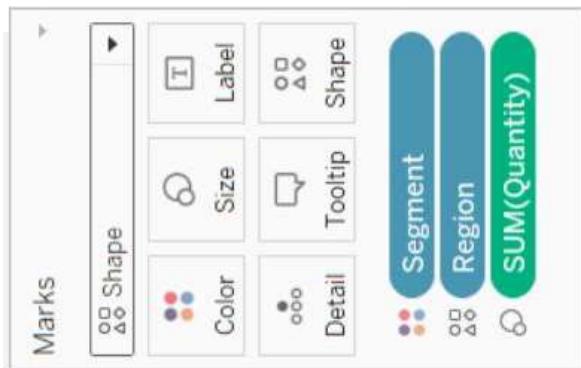
Ways to Build View

- Drag fields from **Data pane** and drop them onto the cards and shelves
- Double-click one or more fields in the **Data pane**
- Select one or more fields in the **Data pane** and then choose a chart type from **Show Me**
- Drop a field on the **Drop field here** grid, to start creating a view



Different Working Areas

- Marks – helps to encode data with color, size, shape, text & detail
- Shelves - **Columns shelf** creates columns & **Rows shelf** creates rows of a table
- Filters – helps to get specific data from fields
 - Filter can be done based on list, wild cards, Conditions, Top/Bottom rules



Filter [Region]

X

General Wildcard Condition Top

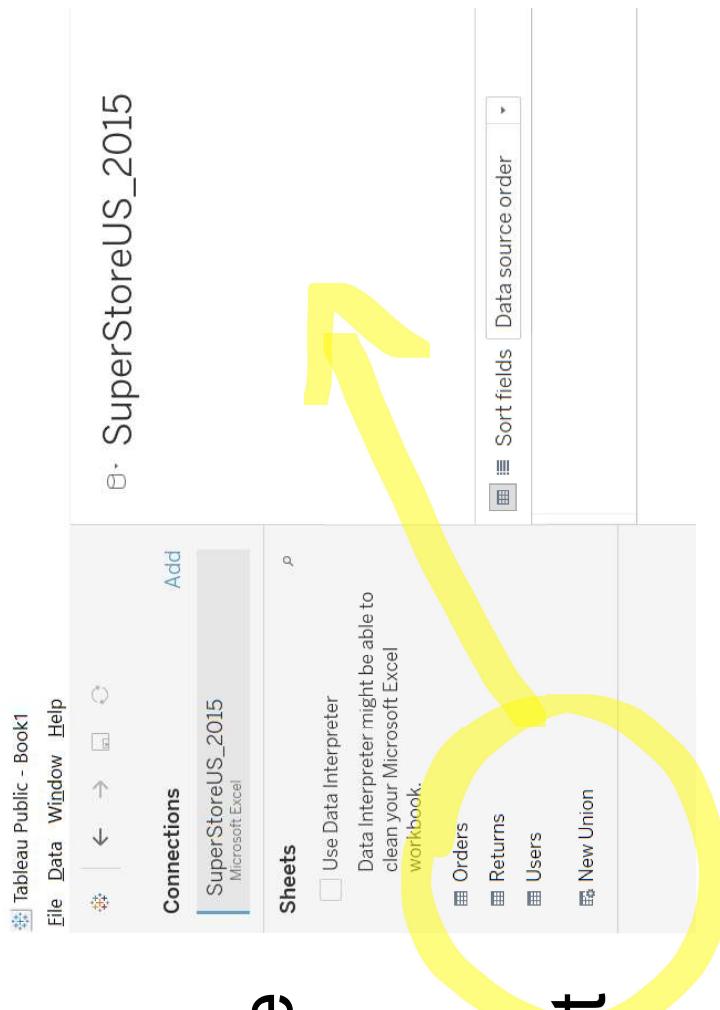
Select from list Custom value list Use all

Enter search text

Central East West

Working with Data

- Connect Excel workbook
- Sheet come on the left side
- To work with a dataset, drag the worksheet on right



WORKING IN DATA SOURCE

Data Source #1

- The data of the sheet dragged shows at the bottom
- Tableau categorizes the data as:
 - Numerical
 - Date
 - Text
 - Date/Time
 - Boolean
 - Geographical
- Change type of data by clicking on the symbol

Data Source #2

- Before analysis in tableau one can

1. Filter data

2. Sort column

3. Split data

4. Hide Column

5. Rename Col

6. Calculated field

The screenshot shows the Tableau Data Source pane for a data source named "Orders (SuperStoreUS_2015)". The pane includes a toolbar with filters, add, sort fields, data source order, and various context menu options like rename, copy values, hide, aliases, and create calculated field. A legend indicates the current state of columns: Abc (Visible), Abc (Hidden), and Orders (Calculated). The main area displays a table with columns: Row ID, Order Priority, # Orders, Discount, Unit Price, # Orders, Shipping Cost, and Customer ID. Step 1 is circled in orange around the filters section. Step 2 is circled in orange around the sort fields dropdown. Step 3 is circled in orange around the split data icon. Step 4 is circled in orange around the hide column icon. Step 5 is circled in orange around the rename column icon. Step 6 is circled in orange around the calculated field icon.

#	Abc Orders	Abc Orders	# Orders	Discount	Unit Price	# Orders	Shipping Cost	Customer ID
20847	High		0.010000		2.84		0.938	
20228	Not Specified		0.020000		500.98		26.000	
21776	Critical		0.060000		9.48		7.290	
24844	Medium		0.090000		78.69		19.990	
24846	Medium		0.080000		3.28		2.310	
24847	Medium		0.050000		3.28		4.200	

Why Filter Data?

- Helps to work with data required
- Large amount of data slow the system
- Optimization of data before analysis

Function in Tableau

<https://onlinehelp.tableau.com/current/pro/desktop/en-us/functions.html>

Exercise

- Connect with dataset – Super Store Sales Data
- In Order table, show data pertaining to
 - Countries – India, China, Japan, Bangladesh
 - Positive profit
 - In year 2012 and 2013 (order date)
- Create a new column with the last 5 digit of the Order ID

Dimension and Measures

- Dimensions –
 - Categorical data
 - Eg. Customer ID, Region
 - Blue colour
 - Placed in column
- Measure
 - Fields which can be measured, aggregated, or perform calculation on
 - Eg. Sales, profit
 - Green colour
 - Placed in row

WORKING IN SHEETS

Sheet

- This area is used to create charts using dimensions and measures
- It has
 - Chart area
 - Marks card
 - apply different effect to charts
- Filter
 - helps to apply filter on charts
- Show me
 - change different types of charts

Measures

- The mathematical calculation on measures can be changed

- Right click

- Measure

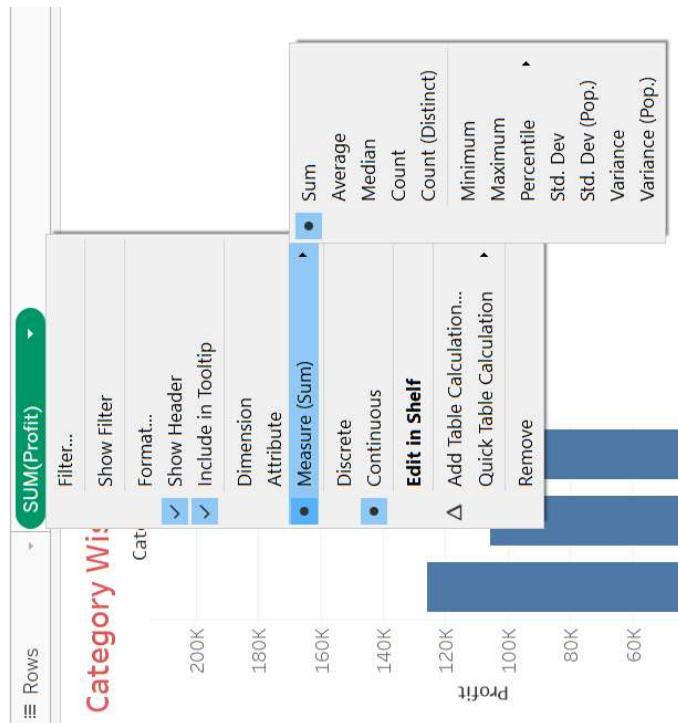
- Sum

- Average

- Median

- Count

- Etc..



Marks Card

- Helps to make changes in the graph created
 - Color
 - Drop dimension on color – this show items of dimension with different colors on chart
 - Size
 - Drop dimension on Size – this show items of dimension with different sizes on chart
 - Label
 - Works like data label. Typically measures are dropped
- The marks (line, shape, column etc) can be changed by drop down under Marks

Adding Colours

- To change the colour of the chart from default → click on colour and select one colour from the panel
- We can colour the graph based on the dimension → drag on drop the dimension on the color Mark card
 - We can drag and down from column also by pressing Ctrl and dragging
 - This is easy as we don't have to look for the dimension in the list on the right

Working With Date

- Date when dropped in column and measure in row creates a timeline
 - Dates gets aggregated to Year
 - To add granularity to data
 - Change the year to month

Exercise

- Use Super Store data set
- Show the Month wise and Year wise profit for various category

Calculated Field

- Office Supplies Data
- Analyze Region Wise Sales Rep performance
 - Find dollar value based on unit and unit price
- Calculated field are created using existing field
- They have small equal sign in front to depict they are calculated
- Show label and formatting of label

Working in Geographical Data

- SuperStore data
- Show Country wise Average Profit
 - Double click on Country
 - From show me pick the world map
 - Drop profit in size
- Add date filter with Year and Month

Exercise

- Use the Super Store data set
- Divide into groups of 4 students
- Come up with 4 question for analysis as:
 - CEO of the company
 - Country Head
 - Customer Relationship Manager

JOINING DATA IN TABLEAU

Types of Joins

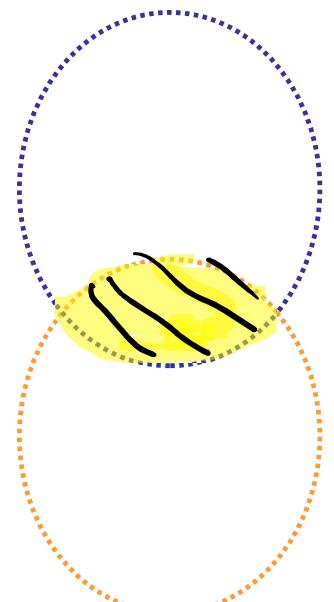
Table
A

Customer	Gender	Age
Adam	Male	24
Ben	Male	32
Jack	Male	34
Nick	Male	29
Susan	Female	24

Table
B

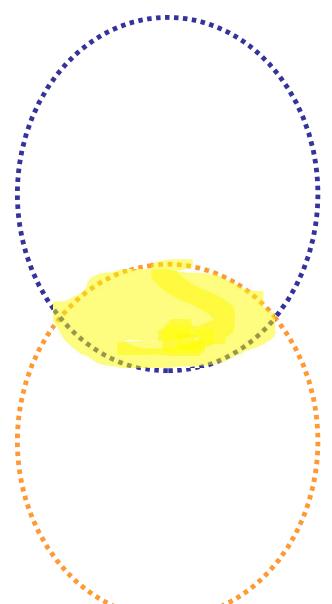
Employee	Title	Wage
Jack	Associate	15 \$/hr
John	Mgr	20 \$/hr
Mary	Mgr	22 \$/hr
Susan	Associate	17 \$/hr

Inner Join #1



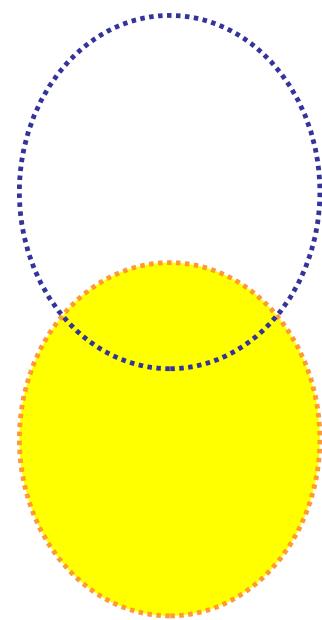
Customer	Gender	Age	Employee	Title	Wage
Adam	Male	24	Jack	Associate	15 \$/hr
Ben	Male	32	John	Mgr	20 \$/hr
Jack	Male	34	Mary	Mgr	22 \$/hr
Nick	Male	29	Susan	Associate	17 \$/hr
Susan	Female	24			

Inner Join #2



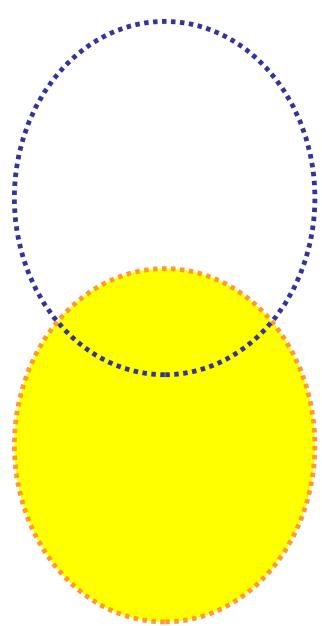
Customer	Gender	Age	Employee	Title	Wage
Jack	Male	34	Jack	Associate	15 \$/hr
Susan	Female	24	Susan	Associate	17 \$/hr

Left Join #1



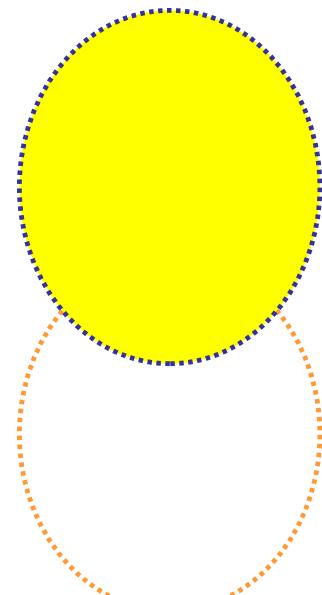
Customer	Gender	Age	Employee	Title	Wage
Adam	Male	24	Jack	Associate	15 \$/hr
Ben	Male	32	John	Mgr	20 \$/hr
Jack	Male	34	Mary	Mgr	22 \$/hr
Nick	Male	29	Susan	Associate	17 \$/hr
Susan	Female	24			

Left Join #2



Customer	Gender	Age	Employee	Title	Wage
Adam	Male	24			
Ben	Male	32			
Jack	Male	34	Jack	Associate	15 \$/hr
Nick	Male	29			
Susan	Female	24	Susan	Associate	17 \$/hr

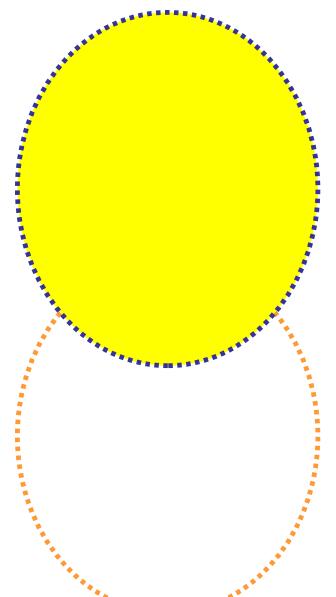
Right Join #1



Employee	Title	Wage
Jack	Associate	15 \$/hr
John	Mgr	20 \$/hr
Mary	Mgr	22 \$/hr
Susan	Associate	17 \$/hr

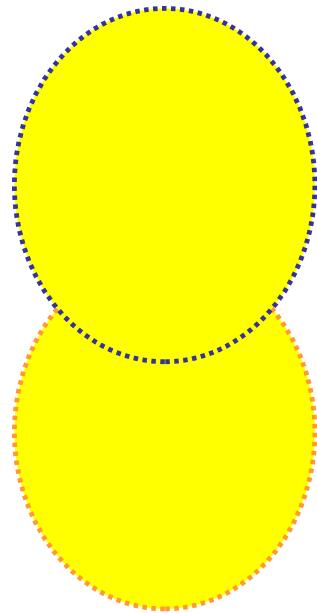
Customer	Gender	Age
Adam	Male	24
Ben	Male	32
Jack	Male	34
Nick	Male	29
Susan	Female	24

Right Join #1



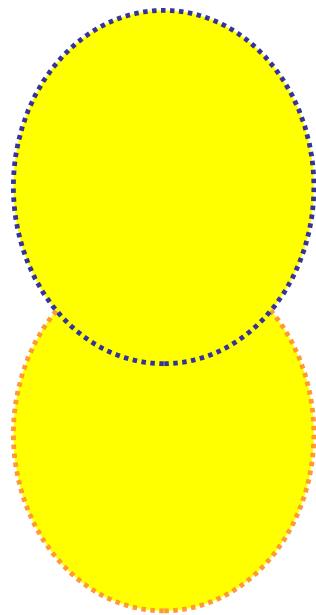
Customer	Gender	Age	Employee	Title	Wage
Jack	Male	34	Jack	Associate	15 \$/hr
			John	Mgr	20 \$/hr
			Mary	Mgr	22 \$/hr
Susan	Female	24	Susan	Associate	17 \$/hr

Full Outer Join #1



Customer	Gender	Age	Employee	Title	Wage
Adam	Male	24	Jack	Associate	15 \$/hr
Ben	Male	32	John	Mgr .	20 \$/hr
Jack	Male	34	Mary	Mgr .	22 \$/hr
Nick	Male	29	Susan	Associate	17 \$/hr
Susan	Female	24			

Full Outer Join #2



Customer	Gender	Age	Employee	Title	Wage
Adam	Male	24			
Ben	Male	32			
Jack	Male	34	Jack	Associate	15 \$/hr
			John	Mgr	20 \$/hr
Nick	Male	29			
			Mary	Mgr	22 \$/hr
Susan	Female	24	Susan	Associate	17 \$/hr

Joining Data

- Data **joining** is a very common requirement in any data analysis
- May **join** data from multiple sources or **join** data from different tables in a single source
- Tableau provides the feature to **join** the table by using the data pane
 - Joins on basis of a common field
 - Inner and Left Join are used commonly
 - 32 tables can be joined
 - Can't join tables from two different data sources

Assignment

- Which customer has the highest sales in each region?
- Which sub category is giving max average profit for each category? Analyze based on regions. Make sure the customer can pick only one region at a time.
- Which countries have consistently high returns in all the years?
- For any one of the country identified above to find the sub-category where return is more than 10 units.

Joining on Multiple Fields #1

Table
A

Table
B

Store	Order #	Customer
North	001	Mike
North	002	Jack
South	001	Susan

Store	Order #	Item	Sales
North	001	Laptop	\$999
North	001	Mouse	\$49
North	002	Monitor	\$155
South	001	Camera	\$97

Joining on Multiple Fields #2

Perform Left Join on
Table A and Table B
based on Store



Store	Order #	Customer
North	001	Mike
North	002	Jack
South	001	Susan

Store	Order #	Item	Sales
North	001	Laptop	\$999
North	001	Mouse	\$49
North	002	Monitor	\$155
South	001	Camera	\$97

Exercise

- Dataset – Join More Col
- Use this data for joining the tables

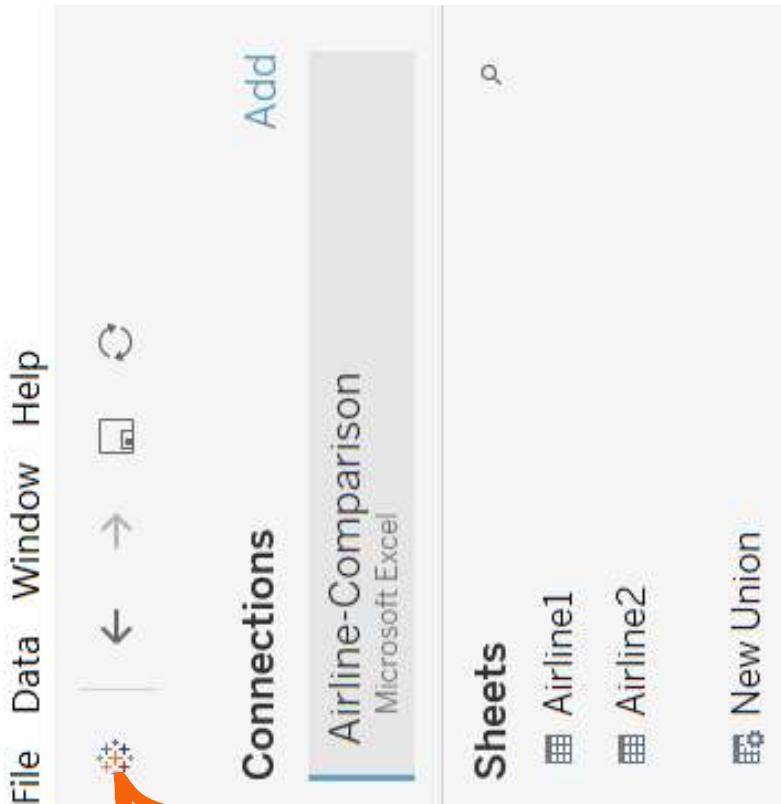
DATA BLENDING

Data Blending

- What is Blending?
 - It is used when there is related data in multiple data sources, which you want to analyze together in a single view. As an example, consider the Sales data is present in a relational database and Sales Target data in an Excel spreadsheet.
- When to Blend vs Join?
 - Combining data from different sources
 - Data is at different levels of detail (granularity)
 - Data preparation is required

Data Blending Process

- Connect to one file
- Pick table from the file
- Connect to another file
- Pick table from the second file



Join Vs Blend

- Join
 - Data gets combines based on row level
 - If the granularity of two dataset is not the same then it create unnecessary data
- Blend
 - Tableau sends queries to both the table
 - Aggregates based on field
 - Then join them

Marks Card – Blended Data

- Marks card is there to control both together
- There is individual marks card also



Multiple Field

- Usually blending is done on one field
- In order to change the relationship
 - Go to Data menu
 - Edit Relationship
 - Custom
 - Add
 - Select the fields to connect
 - OK
- When field connected is removed from visualization, the link is disabled
- Link can be enabled/disabled manually by clicking on it
- Blending done on individual sheet basis
- Blending performs a LEFT Join



Exercise

- Dataset – Data_Airline_Comparison
- Compare the revenue of two airlines based on year and region

Assignment

- Dataset – AmazingMartEU2
- Show Monthly Sales by Segment
- Show How Actual Sales Is Compared To Target Sales

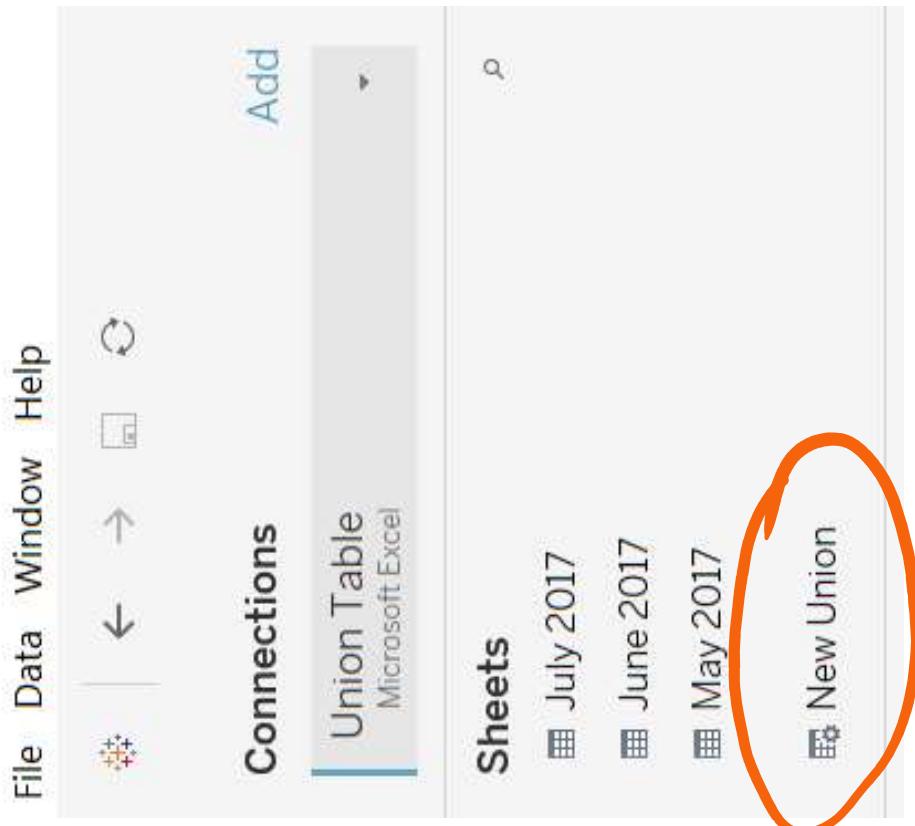
UNION

Union

- Used for combining data by appending rows of one table onto another table
- Union tables should have same number of fields, same field names, and fields are same data type

Process for Union

- Click on New Union
- When dialogue box open, drag and drop tables in it
- Union table appear at bottom
- Rename Union – double click on Union
 - Add or drop tables from Union – click on arrow → edit union



DUAL AXIS CHARTS

Dual Axis

- The charts which have two vertical axis
- Why they are required?
 - When there is large difference in the magnitude of field to be plotted
- This is also called as Secondary Axis

Creating Dual Axis

- Create the charts
- In the second chart created
 - Right click on the axis
 - Select dual axis
- Make sure you change the chart type of values plotted on secondary axis

DEALING WITH UNIDENTIFIED GEOGRAPHICAL DATA

Unidentified Geographical Location

- Sometime location is taken as text rather than geographical location
- Change it by:
 - Right click → geographic role
- If they are not able to map they show as unknown at right bottom
 - Click on unknown
 - Edit location
 - Pick right country

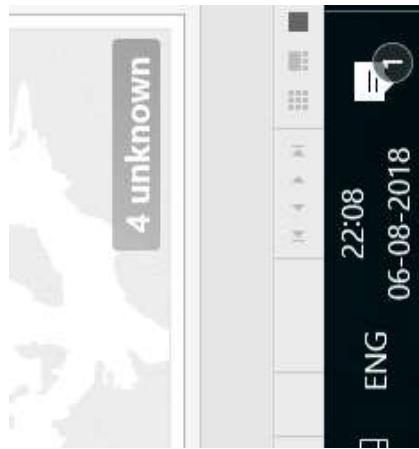


TABLE CALCULATION

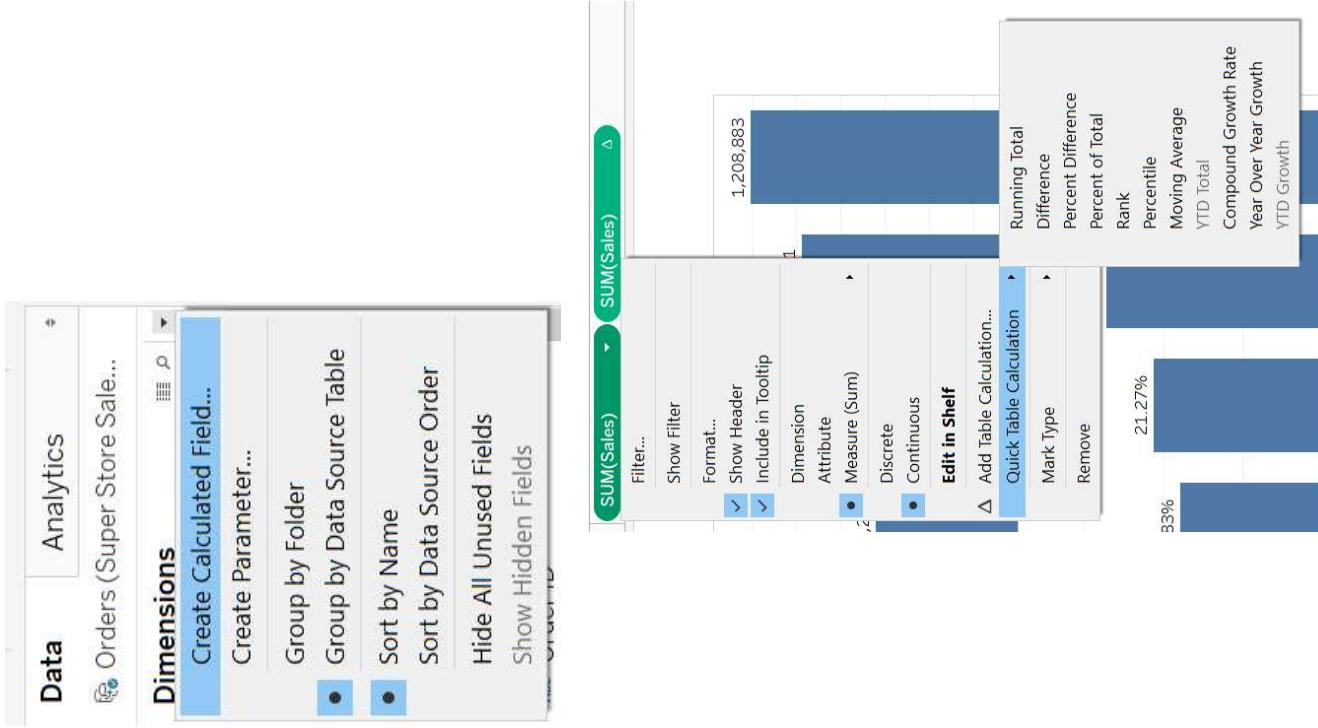
Why Table Calculation?

- It allows to perform computations on values in table
 - Eg. calculate percent of total
- Table Calculation show small triangle on field

YEAR(Order Date)	SUM(Sales)	SUM(Sales) △
------------------	------------	--------------

Database vs Table Calculations

- Database calculations
 - Directly related to database
 - Eg. Difference between two fields
- Table calculations
 - Calculated based on output in Tableau



Exercise

- Use Amazing Mart Data Set
- Percentage contribution of each segment toward total sales
- Month wise running total

Assignment

- Dataset - Super Store Sales data
- Show total sales are each year as well as percentage contribution of each year towards total sales

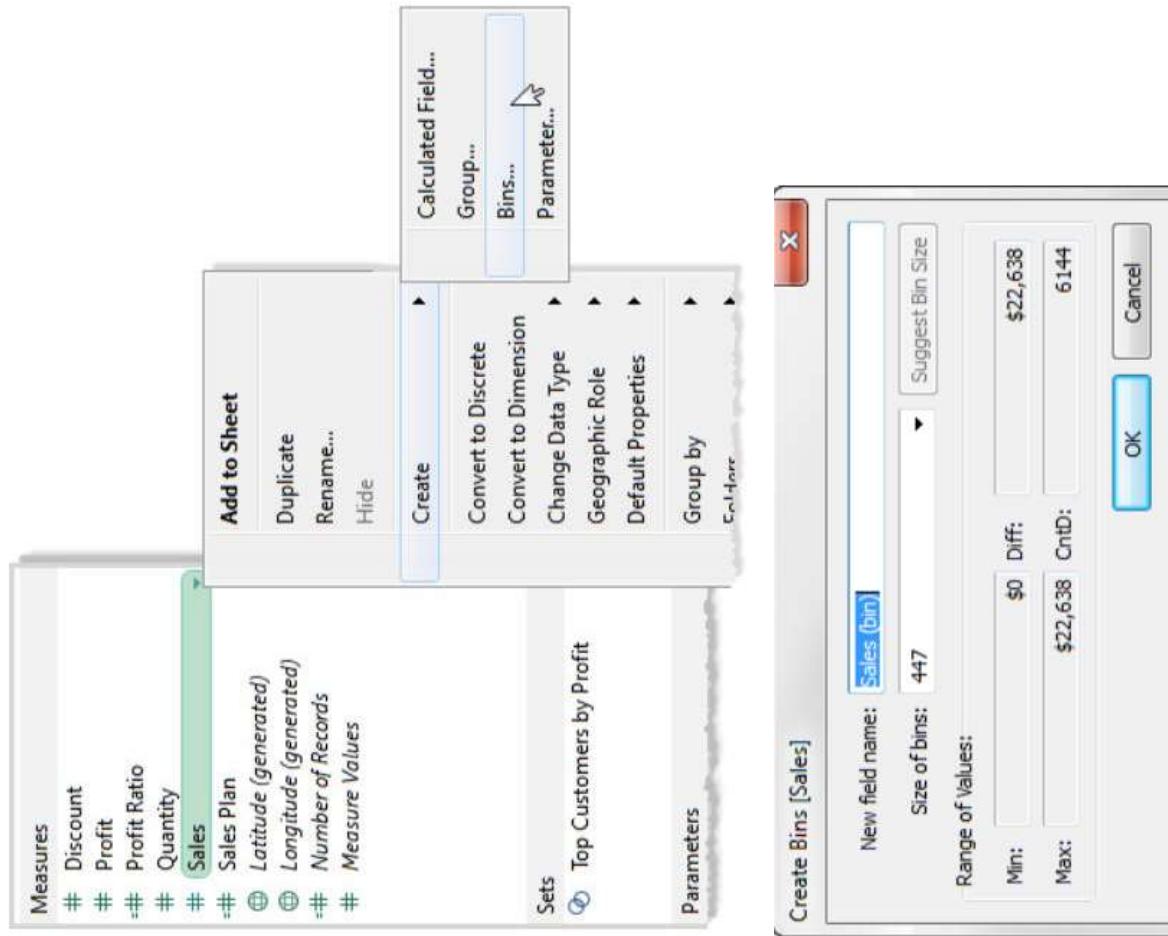
CREATING BINS

Bins

- Helps to group the numerical data for more useful analysis
 - Eg. Age, sales
- It can be created for a continuous measure
- Once bin is created, it converts to dimension
 - Because you are creating a field with a limited and discrete set of possible values

How to Create Bins?

- In the Data pane, right-click a measure and select Create > Bins.
- In the Create Bins dialog box, accept the proposed New field name or specify a different name for the new field.
- Either enter a value in the Size of bins field or have Tableau calculate a value



Assignment

- Amazing mart
- Show what percentage of sales happened in the bins:
 - 0-1999
 - 2000-3999
 - 4000-5999
 - 6000-7990

PARAMETERS

Why Parameters?

- It is used to dynamically change the visualization
- Important when we want to add interactivity and flexibility to a report
- Things important to make them useful:
 - They need to be used in calculations
 - The parameter control needs to be displayed so that viewers can interact with it

Ways Parameter Can Be Used

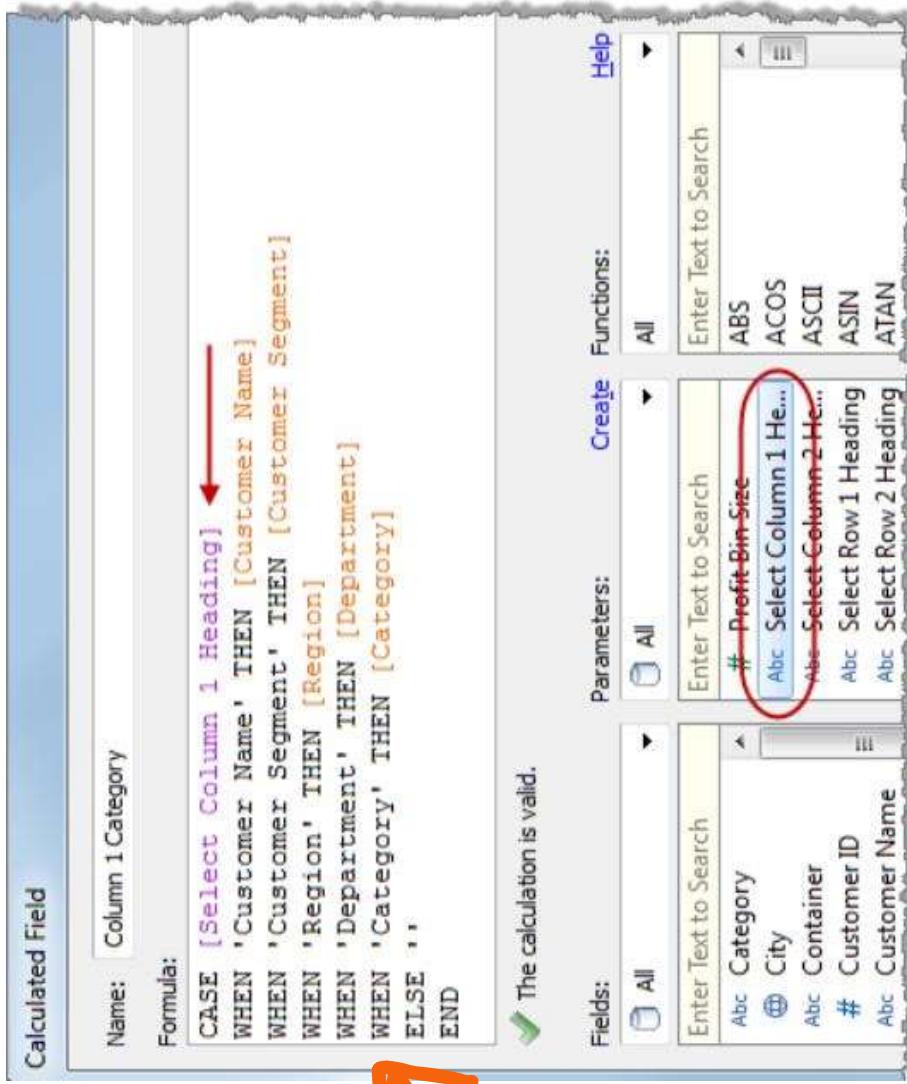
- User-controlled threshold
- What-if analysis
- Dynamic field, axis, titles
- Top N

Steps to Create Parameter #1

1. Select **Analysis > Create Calculated Field**
2. In the Calculated Field dialog box, for **Name**, type **Column 1 Category**
3. Next to **Parameters**, click **Create**, and in the **Create Parameter** dialog box
 - Name parameter such that user understands
 - Select data type
 - For **Allowable Values**, select **List**, type **None** as the first value in the list, and then press **Enter**
 - Complete the list by typing the names of the additional dimension fields

Steps to Create Parameter #2

4. Repeat the previous step to create the following additional parameters
5. In the Calculated Field dialog box, build for **Formula**, build the following calculation



Steps to Show Parameters

- In Parameters pane, right-click parameter and **Select Show Parameter Control**
- From Dimensions pane, drag calculated fields you created to the Columns and Rows
- From the Measures pane, drag a measure to the view
- Test your parameters by selecting fields in the parameter controls

Steps to Use Parameter

1. Create parameter
2. Use parameter in either calculated field, reference line, set or filter

Exercise

- Dataset - Super Store Sales data
- Create chart with sales and forecasted sales for each category
 - Forecasted sales → sales * (1+% increase)
- Create a parameter such that user can adjust the forecasted sales

Exercise

- Database - UK Bank Customer
 - Create a parameter to change the view of Avg balance based on gender, region, Job Classification
- Database – Amazing mart
 - Create a Top N parameter to see the City wise profit

Steps to Make Sort Parameter

- Create Parameter
 - Name it e.g. Sort By
 - Data Type will be String
 - Allowable value – List
 - Give List values
- Create Calculated field with Case
 - Case Parameter Name (Sort By)
When 'Sales' Then Sum(Sales)
When 'Profit' Then Sum(profit)
 - End
- Please note the list values are case sensitive
- Select dimension in shelf on which sort is to be applied
 - Sort → Select Sort Order
 - Sort By → Field → Select the parameter

Steps to Make Top N Parameter

- Select dimension in shelf to apply Top N
- Select Filter
 - In General Tab select Use All
 - Go to Top Tab
 - By Field
- In front of Top instead of value select Create parameter
- Create the top N parameter

Dynamic Dimension

- Create parameter
 - Name it e.g. Select a Dimension
 - Data Type will be String
 - Allowable value – List
 - Give List values
- Create Calculated field with Case
 - Case Select a Dimension
 - When ‘Country’ Then Country
 - When ‘Customer’ Then Customer
 - End
- Drop Calculated Field create in shelf

Date Range Parameter

- Create 2 Parameter
 - Start Date
 - End Date
- Create Calculated Field
 - Order date<=Start date and Order date<=End date
- Drag the calculated field created in filter
 - Select True

Exercise

- Dataset – Super Store Sales data
 - Date range parameter
 - Create dimension parameter for Customer, Sub Category and Category
 - Create Top N parameter
 - Create measure parameter for Sales and profit

Context Filters

- Filters are computed independently
- It disregard other filters
- Context filters helps to filter the data based on another filter
 - Example – cities have to be filtered based on country

GROUPING IN TABLEAU

Groups

- They are created to combine related members in a field
- Working in one dimension
- They are useful for both correcting data errors
 - e.g., combining CA, Calif., and California into one data point

Process of Making Group

- Create a group by selecting data in the view
- Create a group from a field in the Data pane
- Include an Other Group
 - Eg. We can see high performing, low performing, and others. We want to focus on high and low only.

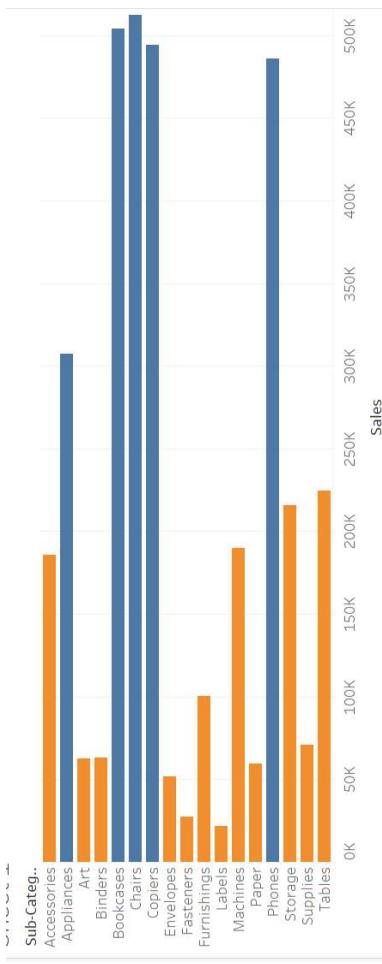
Show Group Within Another Field

- If normal hierarchy is used, the groups don't appear under the other field
- In order to show the groups inside another field put the grouped field in the hierarchy



Grouping Measure

- Measures can also be created as group
 - Can be dropped in colour to see the grouping
 - Eg. Sales of Categories with total more than 3,00,000
- Process
 - Create a calculated field
 - IF sum(sales) >=300000 then True
 - Else false
 - End
 - Drop the calculated field in color card
- Can create parameter to control
 - Limit value in calculated field
- Drop Group field in filter to filter By true/false



Exercise

- Dataset – SuperStore
- In Sub Category – make a group of the following items:
 - Bookcases
 - Chairs
 - Tables
 - Furnishing
- Mark rest of the items as ‘Others’
- Compare the sales for the two groups

Exercise

- On same dataset
- Create Different regions based on the countries
 - Region 1 – India, Pakistan, Afghanistan, Nepal and Bangladesh
 - Region 2 – China and Mongolia
 - Region 3 – Australia and New Zealand
 - Regions 4 – rest of the countries
- Show
 - No. of customers, Avg Sales, Total Profit for each group

Exercise

- Dataset – UK Customer
- Create the bins for Age (10)
- Show the total Balance for each age bin
- Divide balance into two part (visually) based on the value of Avg Balance passed by the user

SETS

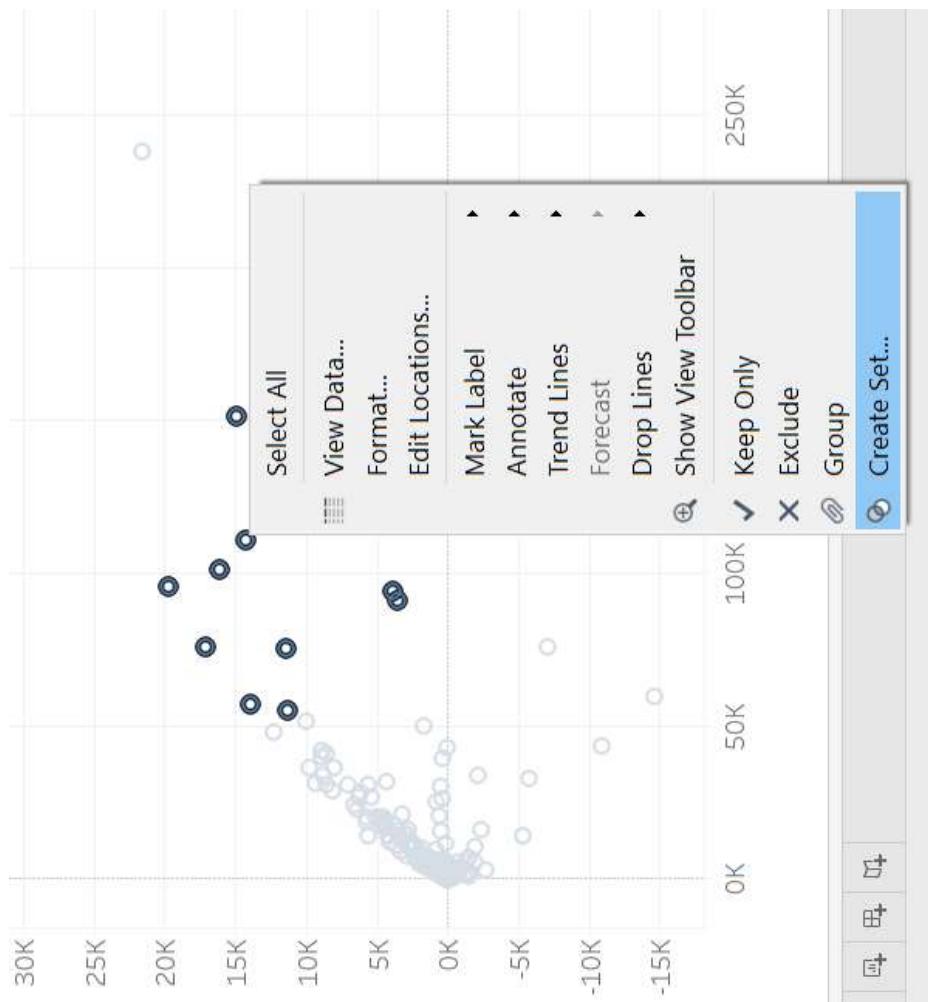
Difference Between Set & Group

- Group – it is done on one dimension only
- Sets – They can be created on various dimensions and measures based on criteria
 - They can be static as well as dynamic
 - Static mean selecting elements of set manually
 - Dynamic means selecting elements through formulas
 - Dynamic sets will adapt to any change in the data
- Set created can be used in other visualization

- Identity the city which have sales $>50k$ and profit more than $4k$

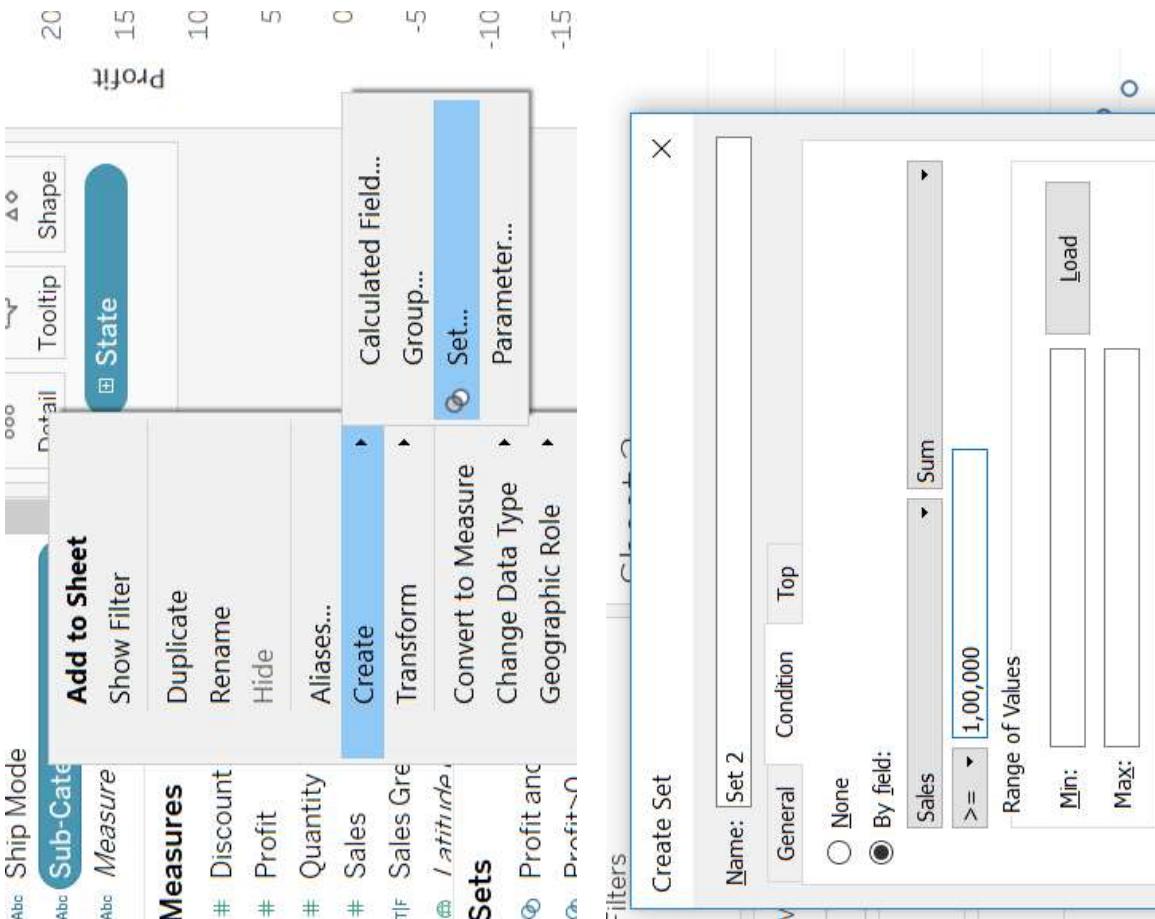
Process of Creating Static Set

- Create visualization
- Manually Select elements to create set
- Right click and create set
- Drop set created in colour to see it visually
- Output of set is IN and OUT



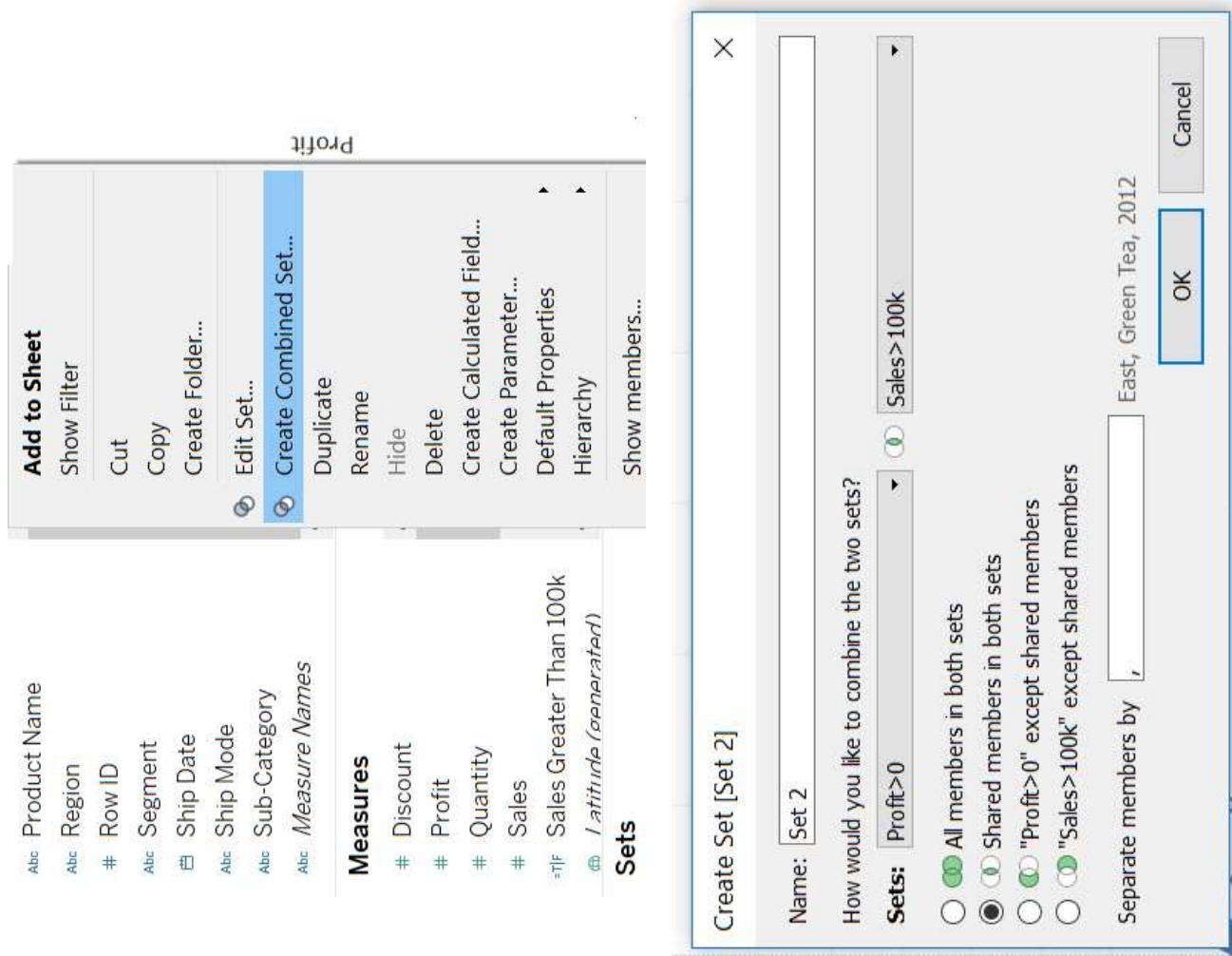
Process of Creating Dynamic Set

- Right click on dimension to create set
- Give condition for creating
- Drop set in colour card to see in visualization



Combined Sets

- Two sets can be combined
- Right click on any one set and select Create Combined Set
- Select sets in window which opens
- Select type of join between two



Exercise

- Dataset – Startup Expansion
- Create a dynamic set such that Cities with Revenue more than 50000 and Marketing Spend of less than 2500 are captured in the set
- Identify the cities where the revenue will increase with increase in marketing spend.

HIERARCHY

Hierarchy

- Create hierarchy
 - To create it drag a field and drop it directly on top of another field in Data Pane
 - Drag and drop in order that you want to view
- Drill up or down in a hierarchy
 - When you add a field from a hierarchy to the visualization, you can quickly drill up or down in the hierarchy to add or subtract more levels of detail
 - Done by pressing the + sign on the left of field in shelf

Exercise

- Create a new Hierarchy
 - Category, Sub-category
 - Show category and under it Sub-Category

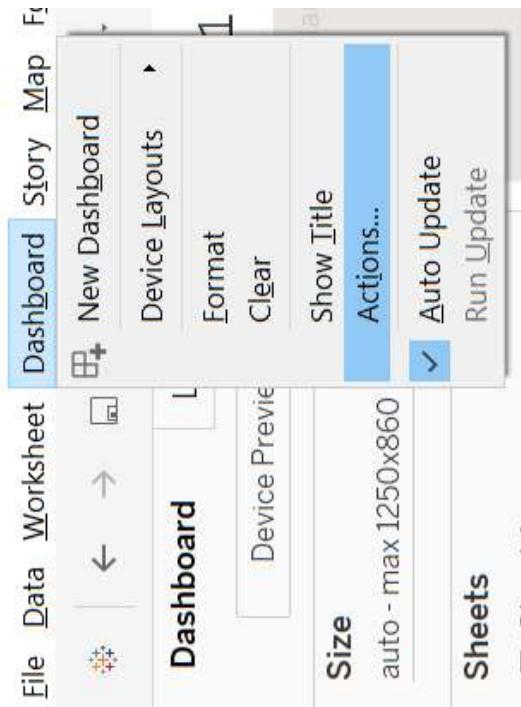
CREATING DASHBOARD

Creating Dashboard

- Dashboard is created after all the view are ready
- Click on New Dashboard Icon at Bottom
- **Dashboard** area appears on the left and lists the sheets in your workbook
- Add views
 - Drag and drop sheets to be included in dashboard
 - Adjust various views to make it presentable

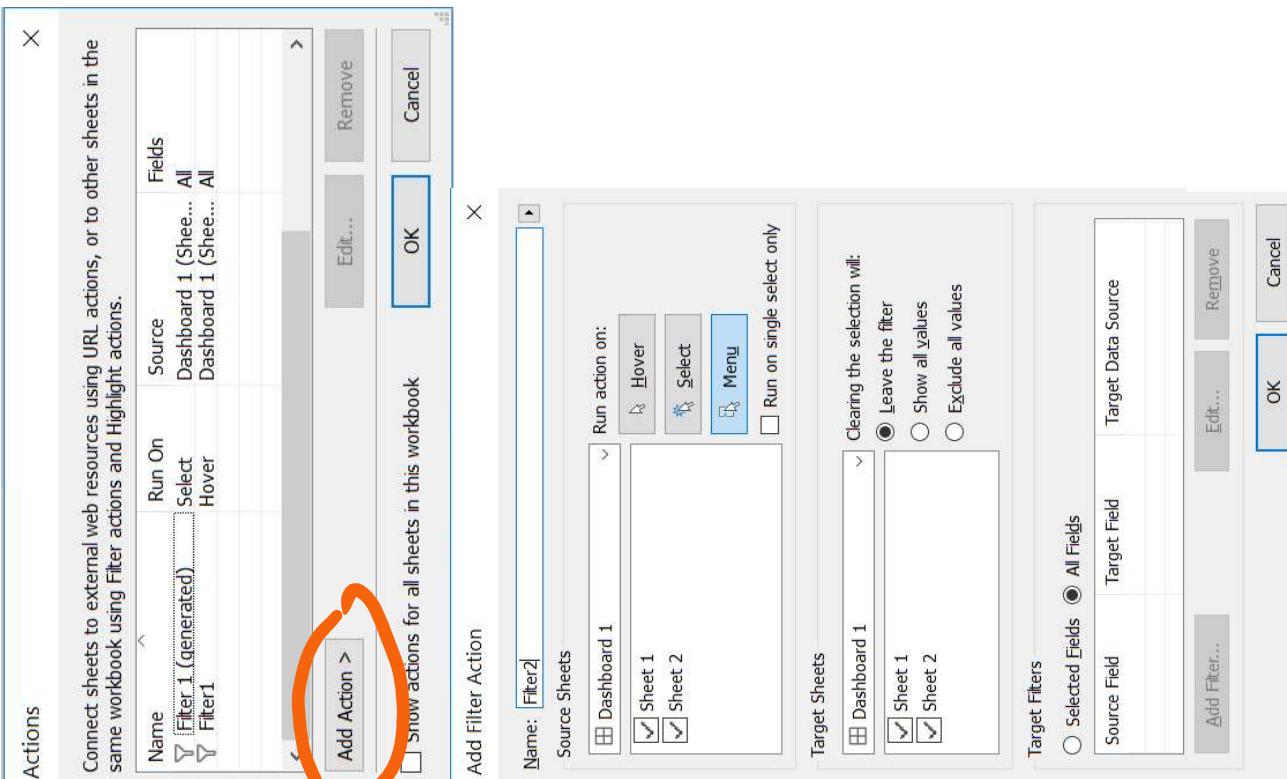
Add Interactivity In Dashboard #1

- Power of dashboards is that you can set up filters and interactivity to associate the different views on your sheet and enhance your users' analysis
- Use the **Use as Filter** option to make one of the views act as a filter on all the other views in the dashboard
- Go to menu
 - Dashboard → Actions



Add Interactivity In Dashboard #2

- Click on Add Action
- Select Filter
- Select which view will act like filter and will have effect on which other views
- In order to apply filter role on different view click filter symbol on right top corner of view



Best Practices for Effective Dashboard

- Know your purpose and audience
- Leverage the most-viewed spot
 - most important view should occupy at upper-left corner of your dashboard
- Format dashboard for size and colour coordination
- Limit number of views on one dashboard
 - Not more than 4
- Add interactivity to encourage exploration

Exercise

- Dataset – UK-Bank-Customer
- Create a dashboard to show Customer Segmentation based on
 - Region
 - Age
 - Balance
 - Gender

Exercise

- Dataset – SuperStoreUS_2015
- Create a dashboard where Top 5 cities come for the state selected.
- If filter removed the Top 5 cities should not appear.
i.e. It should become blank.

Conclusion

- Northern Ireland
 - Females
 - 20-30 age
 - 0-10k balance
- Wales
 - Female
 - 30-40 age
 - 20-40k Balance
- England
 - Male
 - 20-30 age

DATA PREPARATION

Problems in Data #1

- Unstructured Data
 - This data is okay for us
 - However, machines find it difficult to read
 - Ideal structure would have 3 columns
 - Age
 - Period
 - Gender
 - No total row or columns

Age	Period	Men	Women	Total
16 to 19 years	Jan-16	428976	440256	869232
20 to 24 years	Jan-16	534965	133632	668597
25-29 years	Jan-16	263258	425796	689054
30-34 years	Jan-16	112832	478311	591143
35-40 years	Jan-16	450867	450880	901747
16 to 19 years	Feb-16	166238	26091	192329
20 to 24 years	Feb-16	364622	373925	738547
25-29 years	Feb-16	273910	505049	778959
30-34 years	Feb-16	510184	140923	651107
35-40 years	Feb-16	523696	354092	877788
16 to 19 years	Mar-16	432107	531984	964091
20 to 24 years	Mar-16	583873	388756	972629
25-29 years	Mar-16	424666	85901	510567
30-34 years	Mar-16	534354	296156	830510
35-40 years	Mar-16	444650	559994	1004644

Problems in Data #2

- Total row and columns

Region	Sales 2015	Sales 2016	Total
East	123000	200000	323000
West	45000	60000	105000
North	400000	1000000	1400000
Total	168000	260000	3228000

- Problems

- It is difficult to run the analysis based on the years
- Totals will show up as a record in tableau

Problems in Data #3

- Missing values

Age	Period	Men	Women	Total
16 to 19 years	Jan-16	428976	440256	869232
20 to 24 years		534965	133632	668597
25-29 years		263258	425796	689054
30-34 years		112832	478311	591143
35-40 years		450867	450880	901747
16 to 19 years	Feb-16	166238	26091	192329
20 to 24 years		364622	373925	738547
25-29 years		273910	505049	778959
30-34 years		510184	140923	651107
35-40 years		523696	354092	877788
16 to 19 years	Mar-16	432107	531984	964091
20 to 24 years		583873	388756	972629
25-29 years		424666	85901	510567
30-34 years		534354	296156	830510
35-40 years		444650	559994	1004644

Problems in Data #4

- Tables with text on top
 - Unnecessary row on top of data should be removed
 - Unnecessary columns before the data should be removed

The diagram shows a table with several rows and columns. Three orange arrows highlight specific parts of the table:

- A vertical arrow points to the first row, which contains the text "Population Data".
- A horizontal arrow points to the second row, which contains the text "Created By : Krishna".
- A double-headed horizontal arrow points between the "Age" column and the "Period" column.

Population Data					
Created By : Krishna					
Age	Period	Men	Women	Total	
16 to 19 years	Jan-16	428976	440256	869232	
20 to 24 years		534965	133632	668597	
25-29 years		263258	425796	689054	
30-34 years		112832	478311	591143	
35-40 years		450867	450880	901747	
16 to 19 years	Feb-16	166238	26091	192329	
20 to 24 years		364622	373925	738547	
25-29 years		273910	505049	778959	
30-34 years		510184	140923	651107	
35-40 years		523696	354092	877788	
16 to 19 years	Mar-16	432107	531984	964091	
20 to 24 years		583873	388756	972629	
25-29 years		424666	85901	510567	
30-34 years		534354	296156	830510	
35-40 years		444650	559994	1004644	

Data Interpreter

- New feature of Tableau 9
- What it can do?
 - It detects things like titles, notes, footers, empty cells, and so on and bypass them to identify the actual fields and values in your data set
 - It can even detect additional tables and sub-tables so that you can work with a subset of your data independently of the other data

Pivot

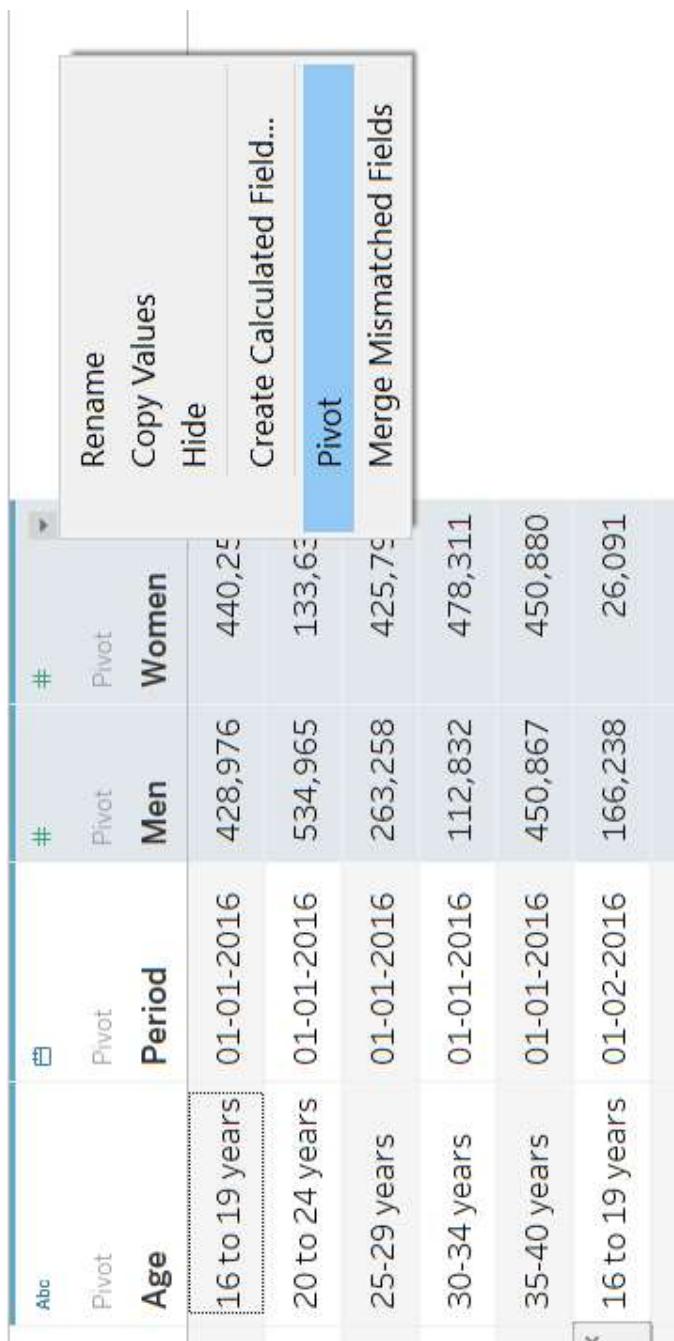
- It helps to convert the row data into columns



Age	Period	Men	Women	Abc		Pivot		Abc		#	
				Pivot	Age	Period	Gender	Pivot1	Balance	Pivot1	Balance
16 to 19 years	Jan-16	428976	440256			25-29 years	01-03-2016	Men		424,666	
20 to 24 years	Jan-16	534965	133632			30-34 years	01-03-2016	Men		534,354	
25-29 years	Jan-16	263258	425796			35-40 years	01-03-2016	Men		444,650	
30-34 years	Jan-16	112832	478311			16 to 19 years	01-01-2016	Women		440,256	
35-40 years	Jan-16	450867	450880			20 to 24 years	01-01-2016	Women		133,632	
16 to 19 years	Feb-16	166238	26091			25-29 years	01-01-2016	Women		425,796	
20 to 24 years	Feb-16	364622	373925			30-34 years	01-01-2016	Women		478,311	
25-29 years	Feb-16	273910	505049			35-40 years	01-01-2016	Women		450,880	
30-34 years	Feb-16	510184	140923			16 to 19 years	01-01-2016	Women		26,091	
35-40 years	Feb-16	523696	354092			20 to 24 years	01-01-2016	Women			
16 to 19 years	Mar-16	432107	531984			25-29 years	01-01-2016	Women			
20 to 24 years	Mar-16	583873	388756			30-34 years	01-01-2016	Women			
25-29 years	Mar-16	424666	85901			35-40 years	01-01-2016	Women			
30-34 years	Mar-16	534354	296156			16 to 19 years	01-02-2016	Women			
35-40 years	Mar-16	444650	559994			20 to 24 years	01-02-2016	Women			

Steps for Pivoting Data

- Select the columns to pivot
 - To select multiple columns at one time press Ctrl while selecting the columns
- Click on drop down on top of any field



Age	Period	#	Pivot	#	Pivot	Women	Rename
16 to 19 years	01-01-2016	428,976		440,25		Create Calculated Field...	
20 to 24 years	01-01-2016	534,965		133,63	Pivot	Merge Mismatched Fields	
25-29 years	01-01-2016	263,258		425,79			
30-34 years	01-01-2016	112,832		478,311			
35-40 years	01-01-2016	450,867		450,880			
16 to 19 years	01-02-2016	166,238		26,091			

Exercise

- Dataset – Arrest Premier Location offence
- Find the top 5 teams based on number of Public disorder Offence

METADATA GRID

Meta Data Grid

- What is it ?
 - It is used when we want to isolate the data and only focus on the headers
- Why we use it?
 - It becomes useful when we have a lot columns
 - Helps to understand which insights can be achieved based on the given data
- Process
 - Click on Manage metadata button on top left

Field	Remote Field Name	Pivot	Age
Abc	Age	Pivot	Age
Period	Period	Pivot	Period
# Men	Men	Pivot	Men
# Women	Women	Pivot	Women

Grid Explanation

- Field name
 - Name of field in tableau
 - Can be same of source or renamed by user
 - Name of field can be changed in grid also
- Table
 - The table name from where the data is picked
- Remote field name
 - Name of the field in the original table

MIS.C.

Creating Custom Territories

- Two methods for creating:
 - *Method 1*
 - Drop locations on the map
 - Manually select the regions for marking one territory
 - Right click
 - Group
 - *Method 2*
 - Change the data type of the region to geographic role
 - The field would come under the country, state etc hierarchy
 - Through this process tableau perform aggregate function on associated measures

Cross-Database Joins

- Tableau 10 allows user to create join between 2 different data source
- Process :
 - Open first database
 - Click on Add
 - Choose other database
 - Tableau will create a Join
 - Check the column used for join

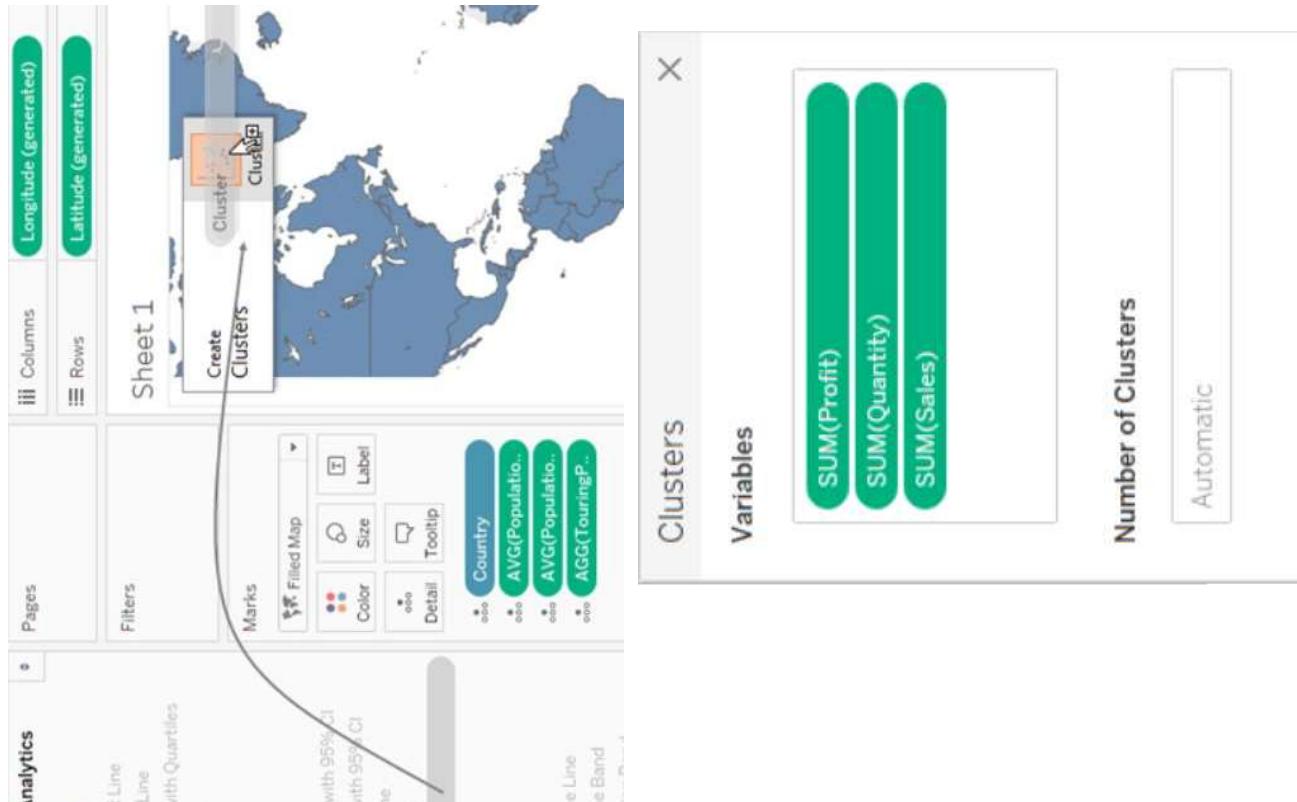


ANALYTICS TAB

Clusters

- It is a feature of Tableau 10
- Allows you to easily group similar dimension members
- Helps to create statistically-based segments which provide insight into how different groups are similar as well as how they are performing compared to each other
- Tableau uses the *k-means algorithm* for clustering

Create Cluster



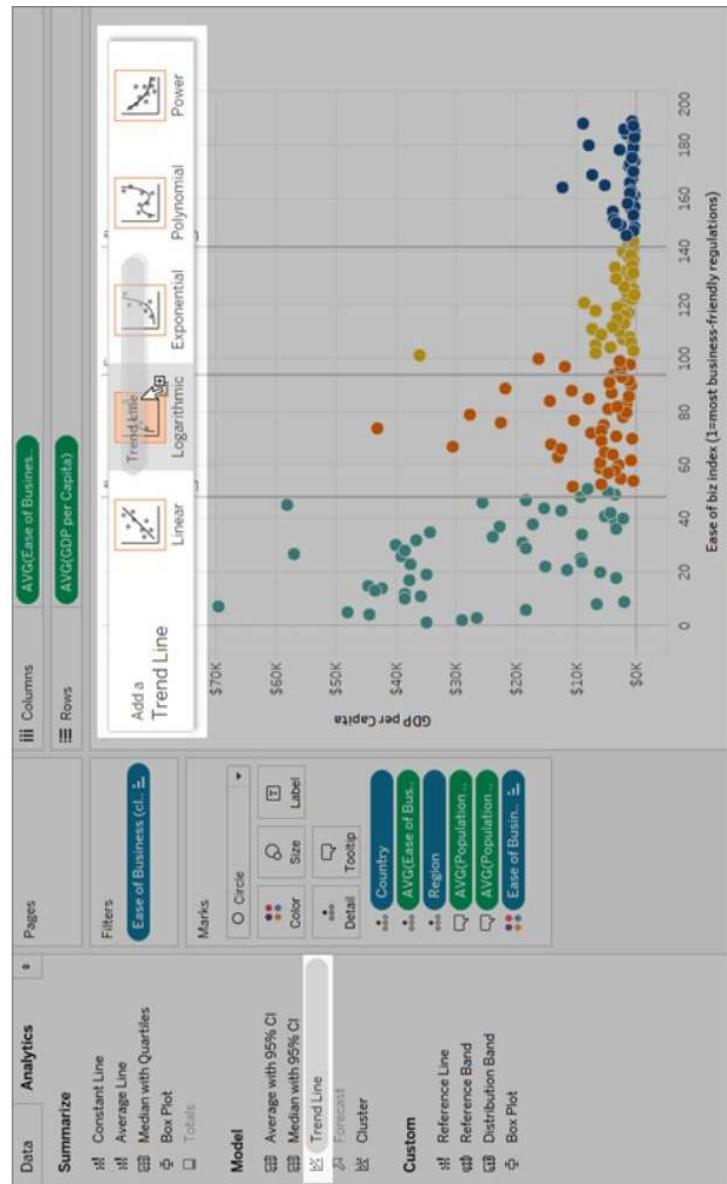
- Create a View
- Drag Cluster from Analytics pane into View
- In Cluster dialogue box drag the field based on which we want the cluster
 - Cluster can be edited to add or remove a field
 - Cluster can be *saved to be used later by dragging cluster to Dimensions*

Trend Line

- It is a line which indicates the general course or tendency of something
- It helps us to forecast the behaviour
- To draw Trend line on view **both axes must contain numerical field**
- Trend line *gives us the correlation between two fields in view*

Create Trend Line

- Create a View
- Drag Trend Line from Analytics pane into View
- Choose which type



TYPES OF CHARTS IN TABLEAU

Cheat Sheet

- Refer to PDF in google classroom for one page cheat sheet
 - Chart Type - Cheat Sheet

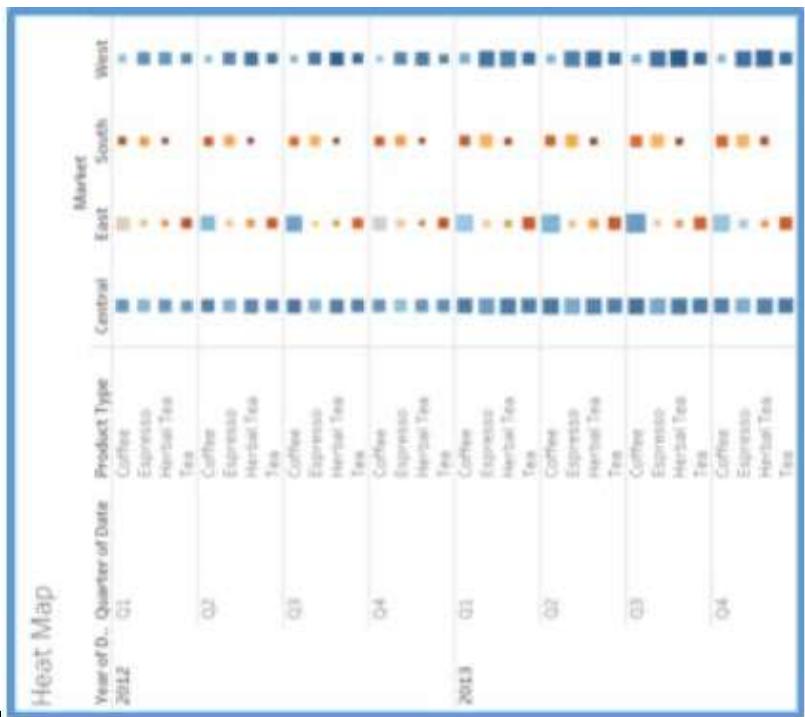
Text Table (Crossstab)

- **When to Use:** Similar to an Excel table. Use this if you want to see your data in rows and columns without any extra visual cues.
- **Minimum Requirements:** 1 or more dimensions, 1 or more measures

Text Table			
Year of Sale	Quarter -	Product Type	Market Unit
2012	Q1	Coffee	\$1,000
		Espresso	\$1,000
		Hot Tea	\$1,000
		Tea	\$1,000
Q2		Coffee	\$1,000
		Espresso	\$1,000
		Hot Tea	\$1,000
		Tea	\$1,000
Q3		Coffee	\$1,000
		Espresso	\$1,000
		Hot Tea	\$1,000
		Tea	\$1,000
Q4		Coffee	\$1,000
		Espresso	\$1,000
		Hot Tea	\$1,000
		Tea	\$1,000
2013	Q1	Coffee	\$1,000
		Espresso	\$1,000
		Hot Tea	\$1,000
		Tea	\$1,000
Q2		Coffee	\$1,000
		Espresso	\$1,000
		Hot Tea	\$1,000
		Tea	\$1,000
Q3		Coffee	\$1,000
		Espresso	\$1,000
		Hot Tea	\$1,000
		Tea	\$1,000
Q4		Coffee	\$1,000
		Espresso	\$1,000
		Hot Tea	\$1,000
		Tea	\$1,000

Heat Map

- **When to Use:** It is an effective way to compare categories using color and size
- **Minimum Requirements:** 1 or more dimensions, 1 or 2 measures



Highlight Table

- **When to Use:** Similar to an Excel table but the cells are colored (similar to conditional formatting in Excel). Can be used to compare values across rows and columns.
- **Minimum Requirements:** 1 or more dimensions, 1 measure

Highlight Table					
Year of B.	Product Type	Central	East	Market	West
2012	Office	\$33,464	\$27,746	\$10,104	\$13,304
	Supply	\$33,724	\$23,472	\$12,504	\$13,392
	Healthcare	\$33,213	\$20,241	\$12,304	\$13,392
	Total	\$33,464	\$23,472	\$12,304	\$13,392
2013	Office	\$35,272	\$28,904	\$11,816	\$12,564
	Supply	\$36,444	\$24,725	\$12,304	\$13,392
	Healthcare	\$36,612	\$21,128	\$13,112	\$13,392
	Total	\$36,495	\$26,627	\$12,304	\$13,392

Symbol Map

- **When to Use:** Uses to tell a story that has geographical data in it. Can highlight where you have the most sales, or to identify concentration of customers in a location.
Use size and color to make the visual pop.
- **Minimum Requirements:** 1 geo dimension, 0 or more dimensions, 0 to 2 measures



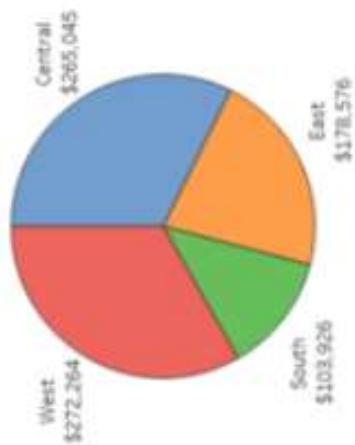
Filled Map

- **When to Use:** Similar to the symbol map discussed above, however, instead of symbols, you use color to fill the geographical region in order to tell the story.
- **Minimum Requirements:** 1 geo dimension, 0 or more dimensions, 0 to 2 measures



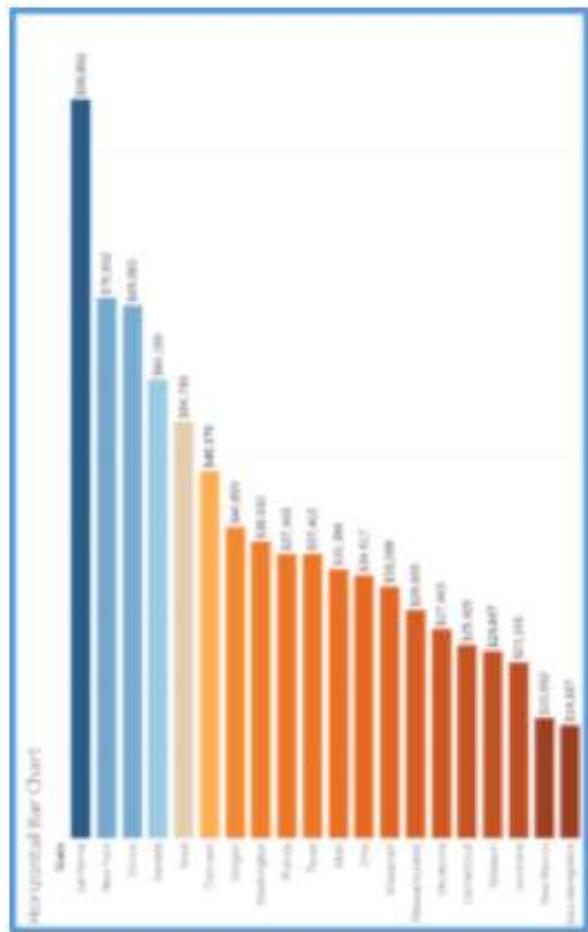
Pie Chart

- **When to Use:** They are best suited to show proportional or percentage relationships. Tableau recommends that users limit pie wedges to six. If you have more than six proportions to communicate, consider a bar chart. It becomes too difficult to meaningfully interpret the pie pieces when the number of wedges gets too high.
- **Minimum Requirements:** 1 or more dimensions, 1 or 2 measure



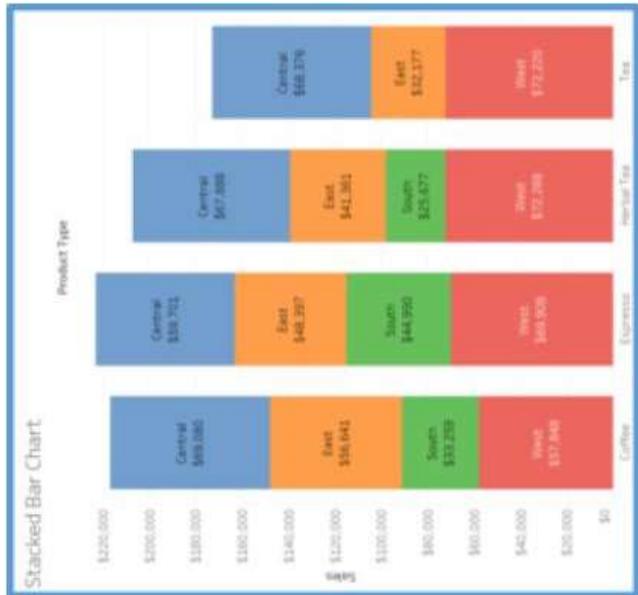
Horizontal Bar Chart

- **When to Use:** Comparison. Mostly used chart.
- **Minimum Requirements:** 0 or more dimensions, 1 or more measures



Stacked Bar Chart

- **When to Use:** Similar to the horizontal bar chart, you can use the stacked bar chart to show data in categories that are also stratified into sub-categories. It allows us to see more details than the regular bar chart would provide.
 - **Minimum Requirements:** 1 or more dimensions, 1 or more measures



Treemap

- **When to Use:** You can use a treemap to show hierarchical (tree-structured) data and part-to-whole relationships. Treemapping is ideal for showing large amounts of items in a single visualization simultaneously. This view is very similar to a heat map, but the boxes are grouped by items that are close in hierarchy.
- **Minimum Requirements:** 1 or more dimensions, 1 or 2 measures

