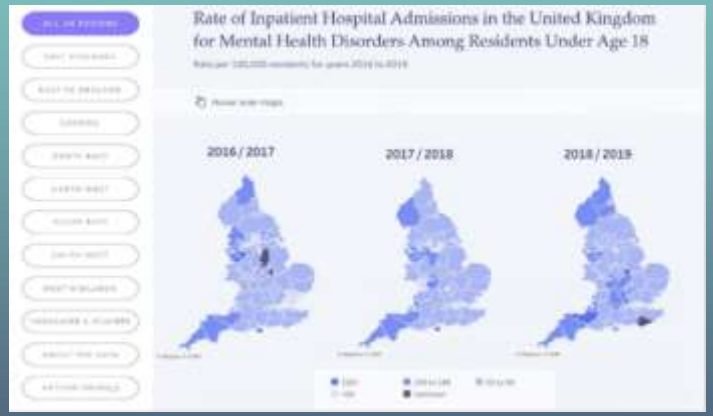
Introduction to Tableau

Your new data analysis software

Data Visualization with Tableau

Tableau Public Gallery



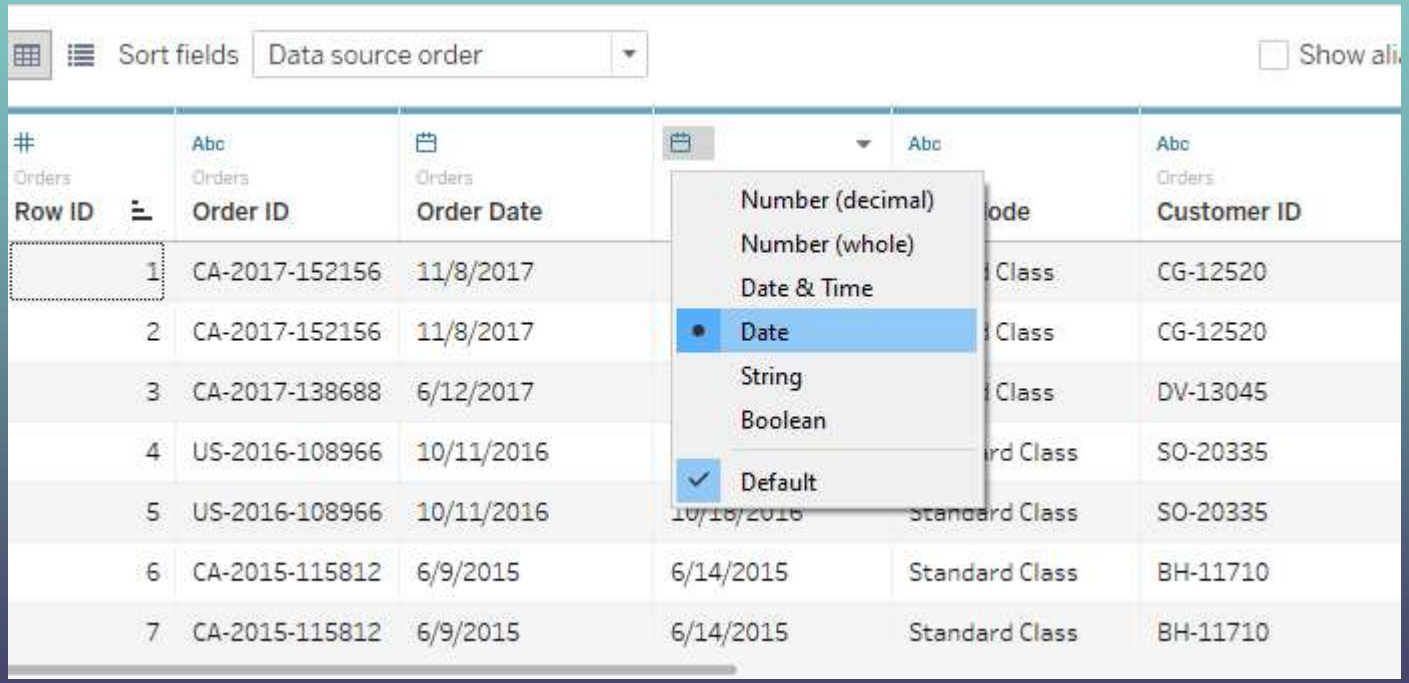
Import Data

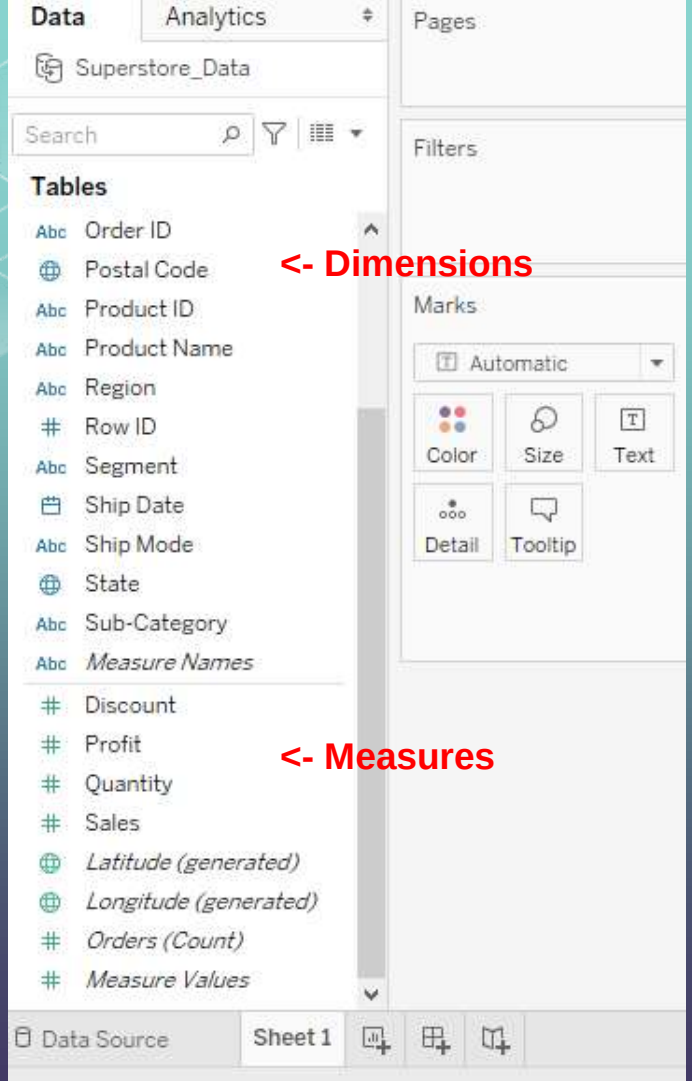
Download the sample datafile “SuperStore_Data” and open it with your Tableau.

You can choose different sheets from the excel file, e.g., “Orders”.

Once the file is open, check the data type first.

Check Data Type

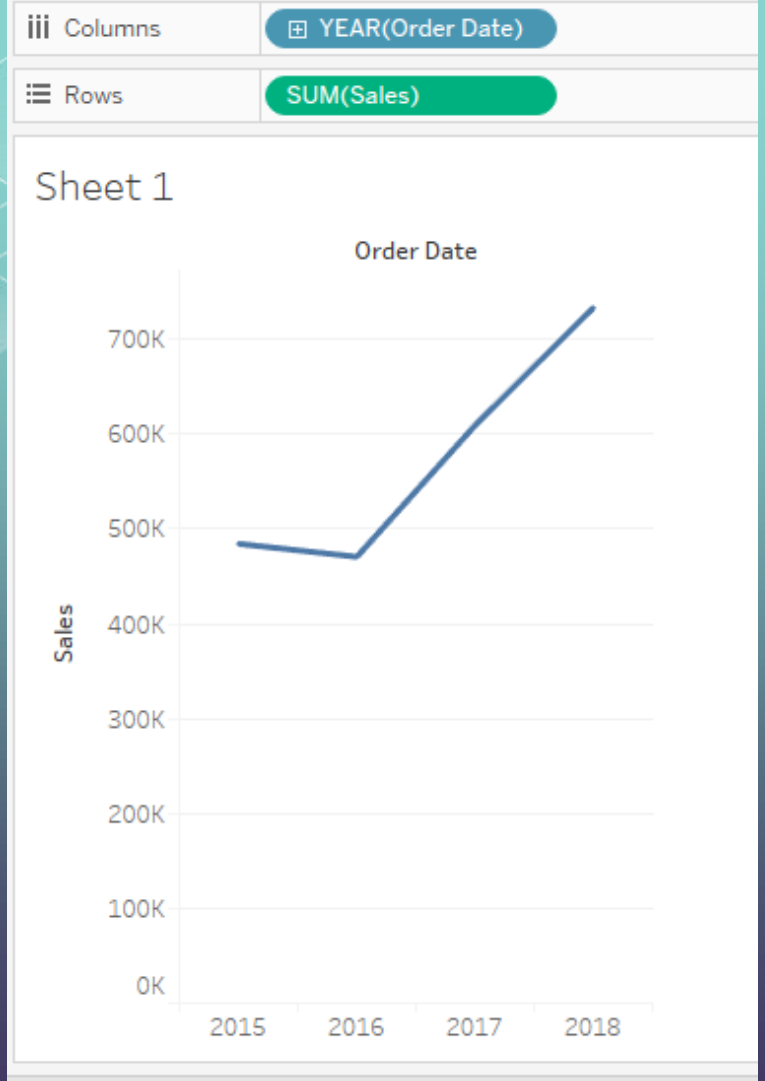




Open “Sheet 1” and you will get to the following page.

Measures are variables that can be calculated.

Dimensions are used to partition your measures (e.g., partition the data based on geographic locations).

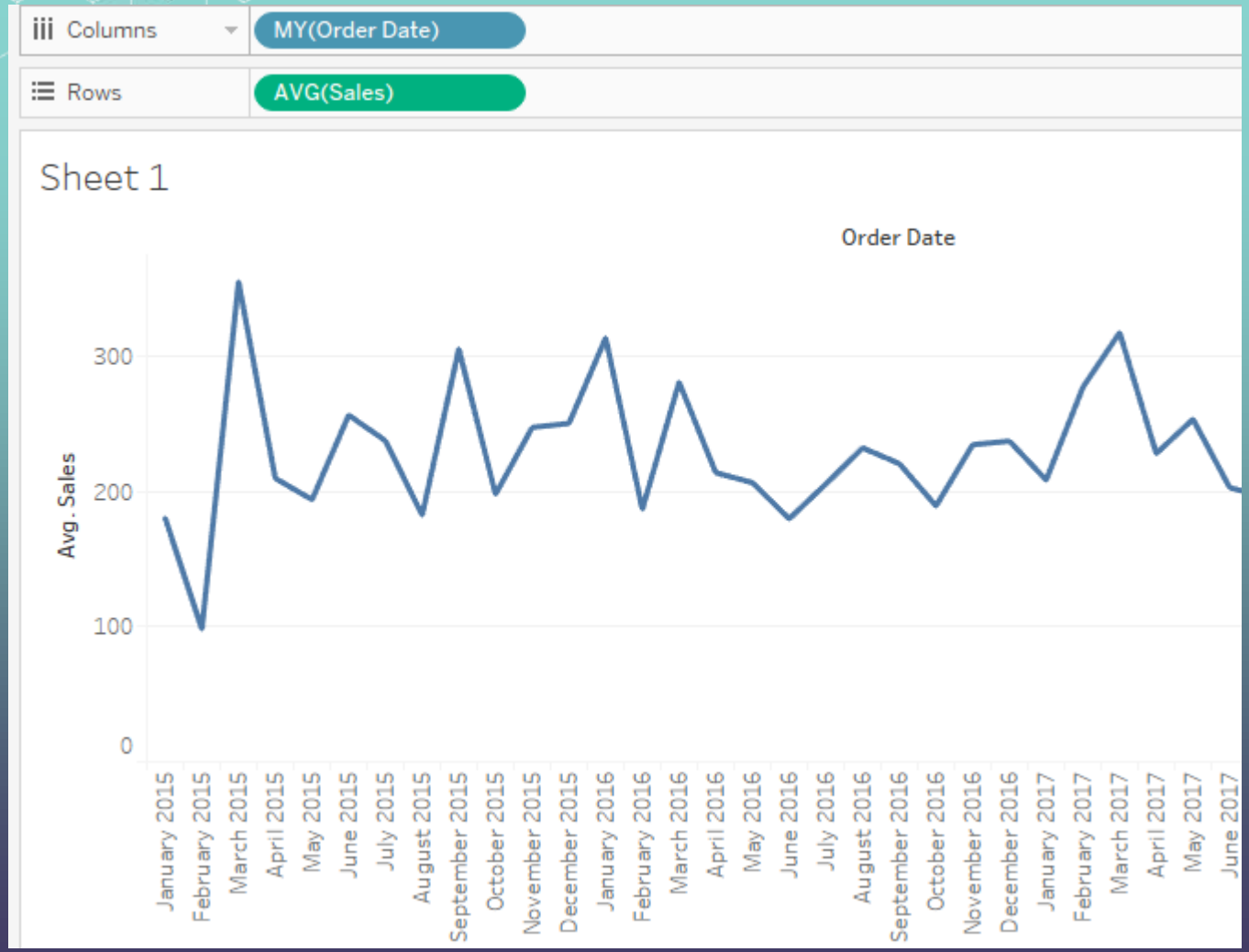


Here, Sales is a measure.

Order date is a dimension.

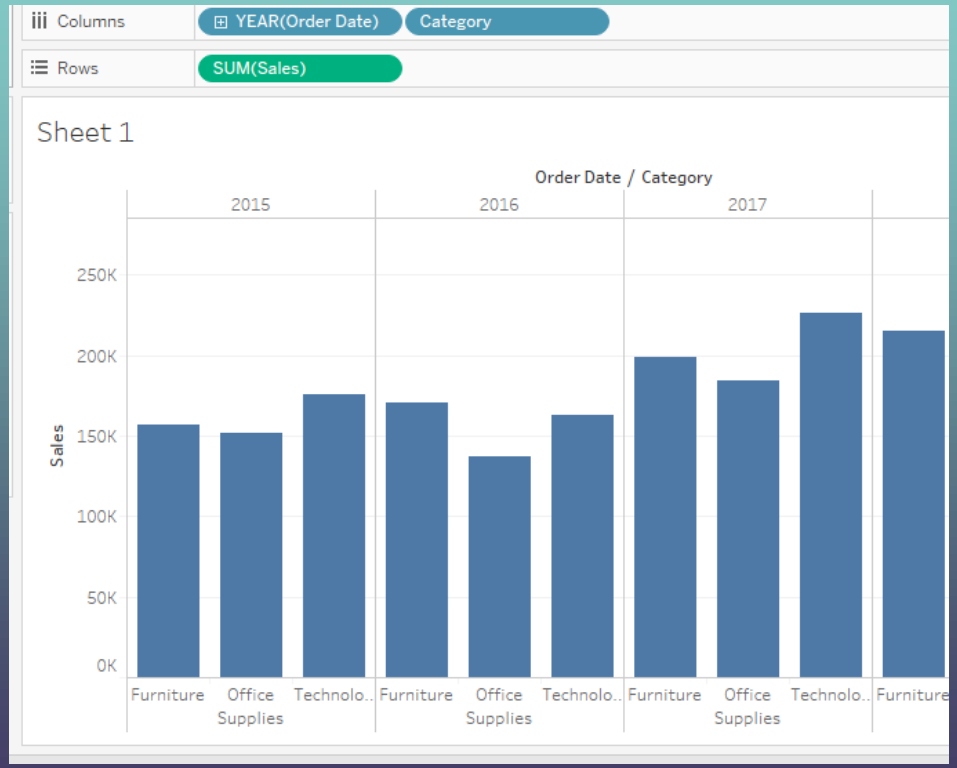
We partition sales based on year of order.

You can also change the properties of your rows and columns.



Here, we use average sales instead of total sales, and use month/year partition instead of year partition.

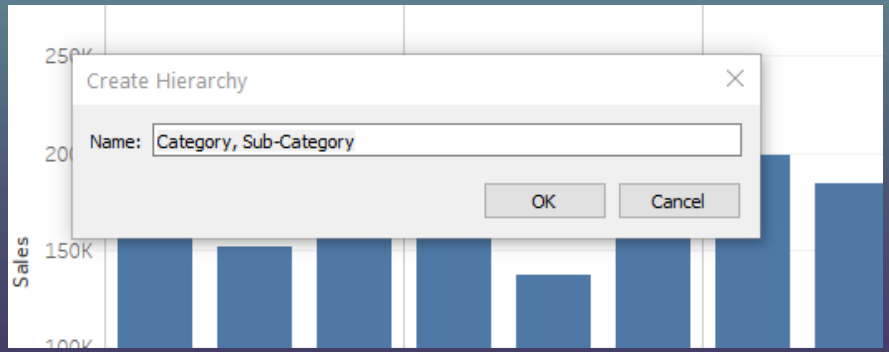
Adding Category As Another Classifier



Create Hierarchy

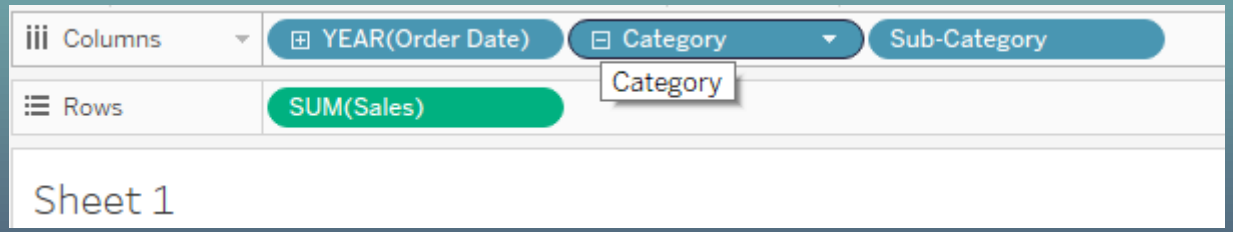
There are several subcategories within each product category (e.g., there are many kinds of furniture with the furniture category), and this information is kept as “subcategory”.

Now, drag “subcategory” to “category” and rename it as “products”.

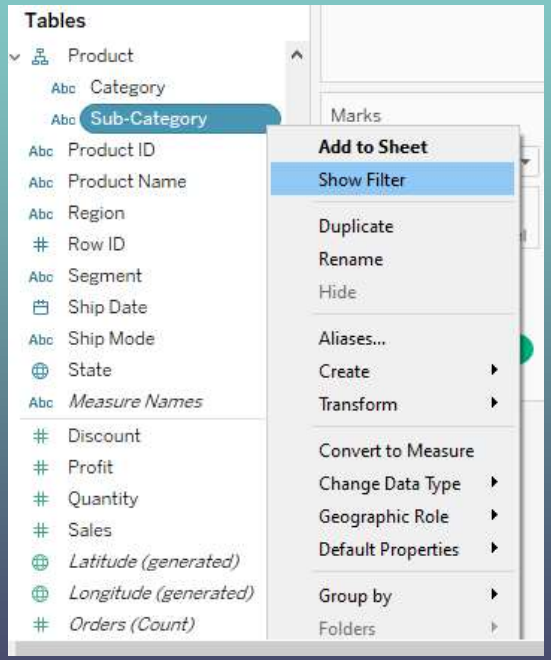


Create Hierarchy

Then, you can display each product subcategories.



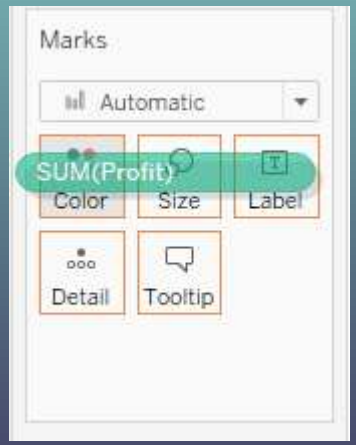
Apply Filters



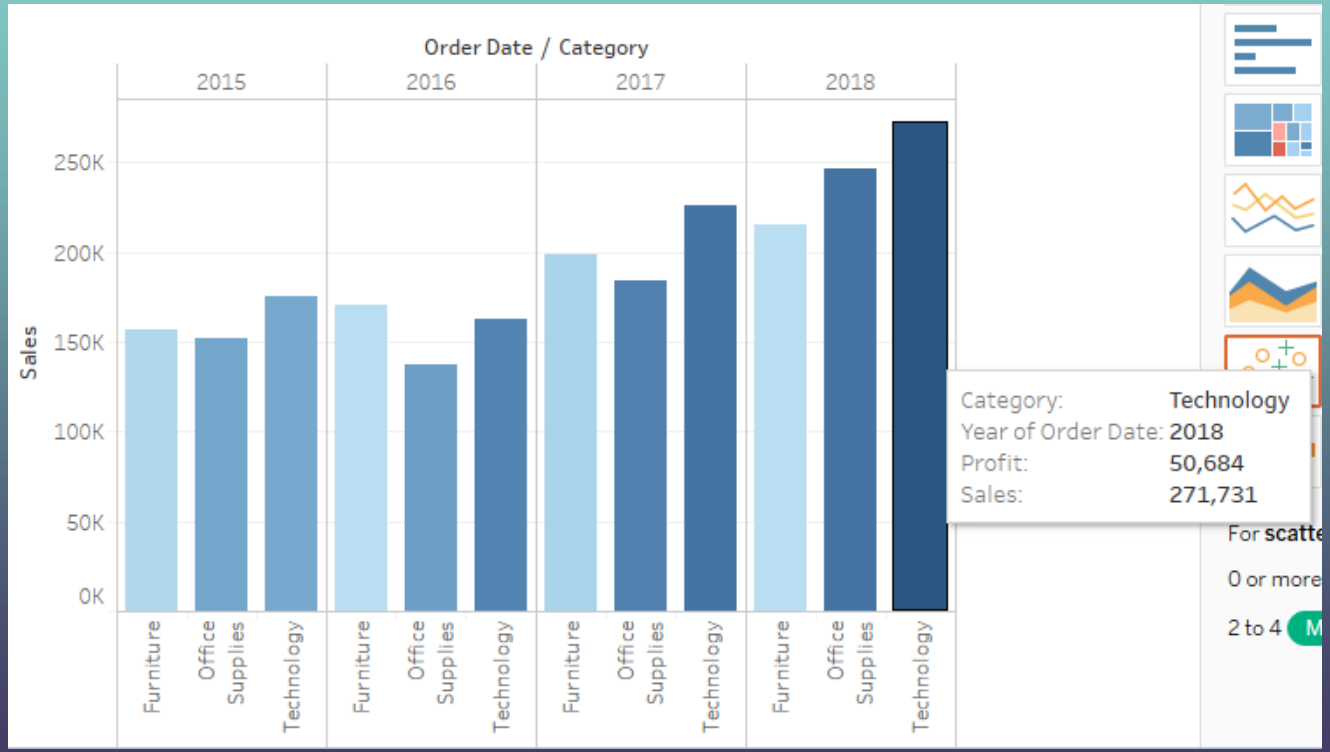
Color Your Output

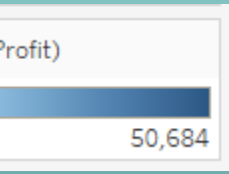
Suppose that you want to color your figure based on the profit of your products.

Drag profit to color.

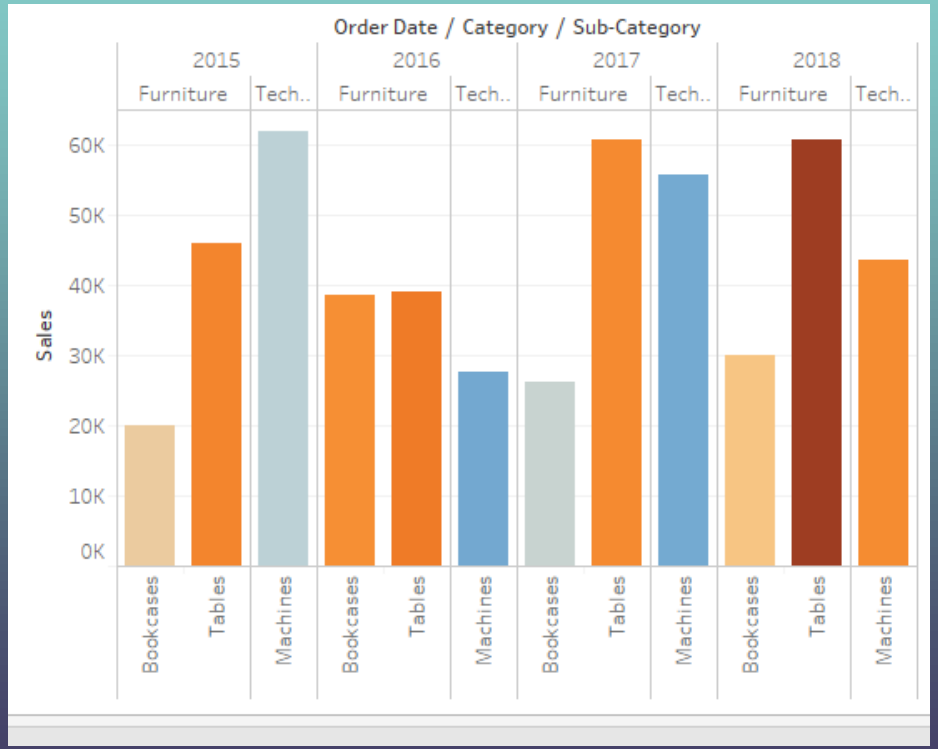


Color Your Output

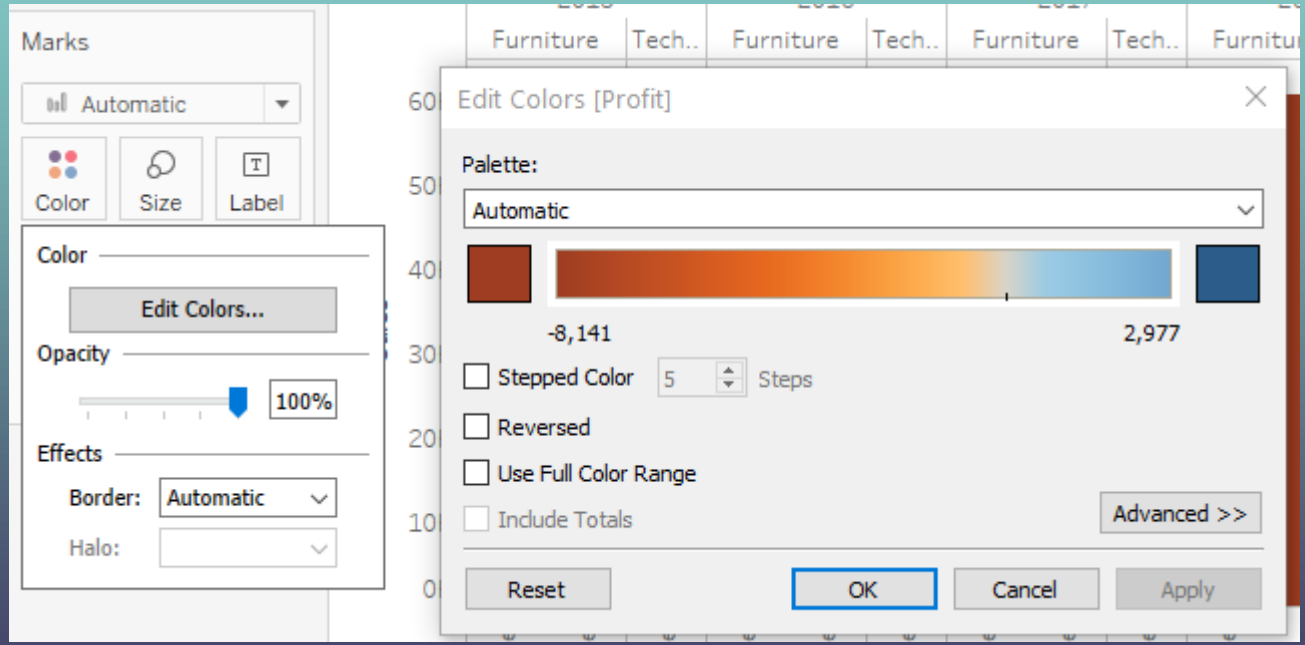




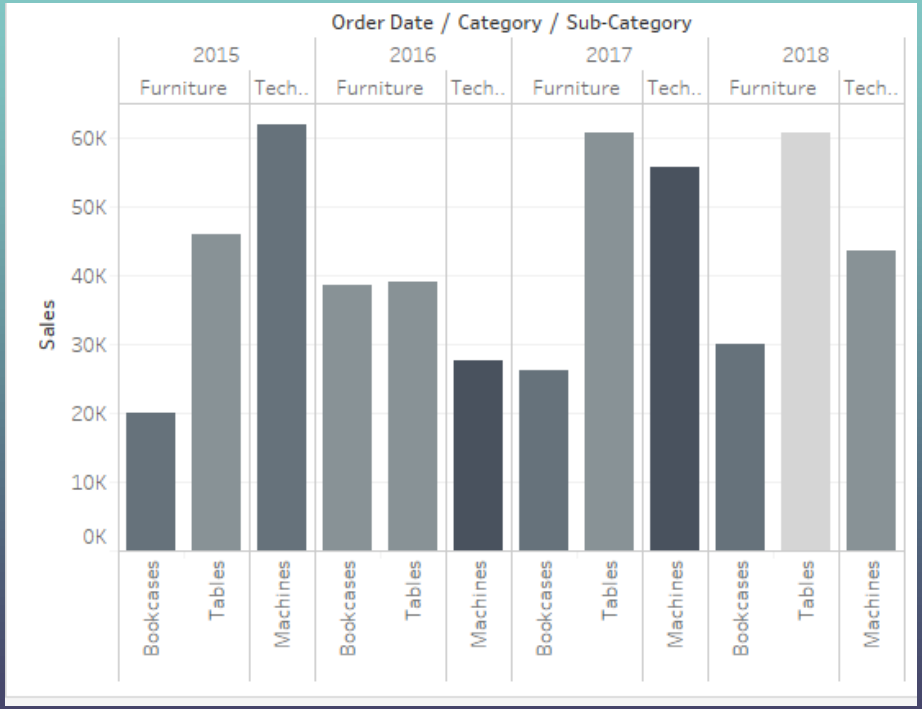
Apply Filters



Edit Colors

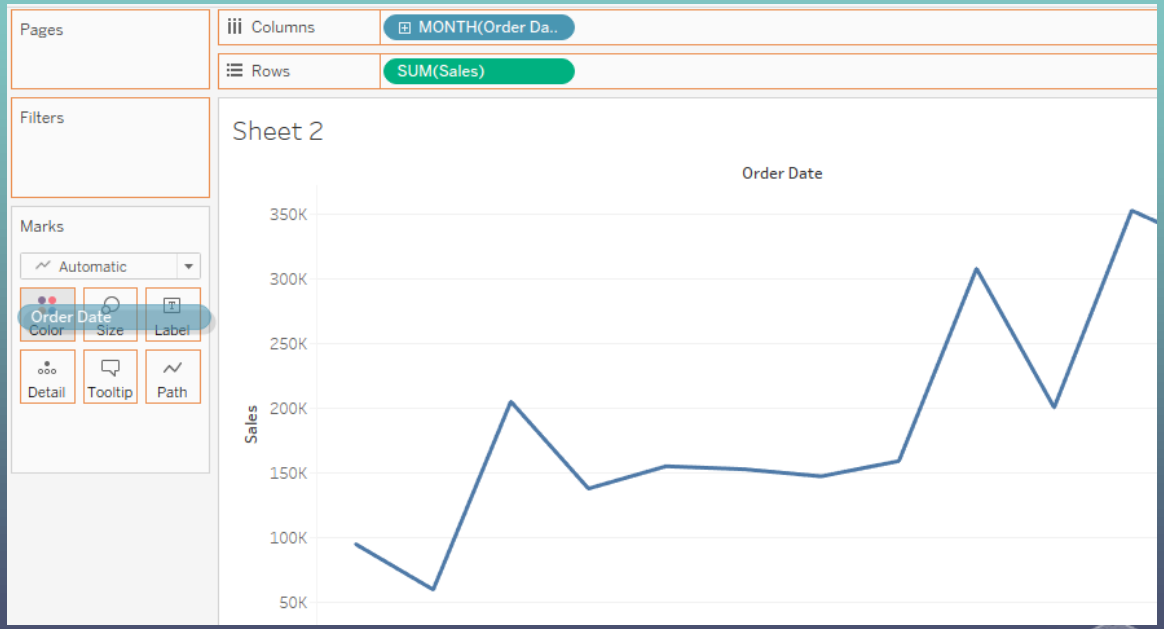

Marks
Automatic
Color
Size
Label
Color
Edit Colors...
Opacity
100%
Effects
Border:
Automatic
Halo:
Furniture
Tech..
Furniture
Tech..
Furniture
Tech..
Furnitu
60
50
40
30
20
10
0
Edit Colors [Profit]
Palette:
Automatic
-8,141
2,977
Stepped Color
5
Steps
Reversed
Use Full Color Range
Include Totals
Advanced >>
Reset
OK
Cancel
Apply

Edit Colors

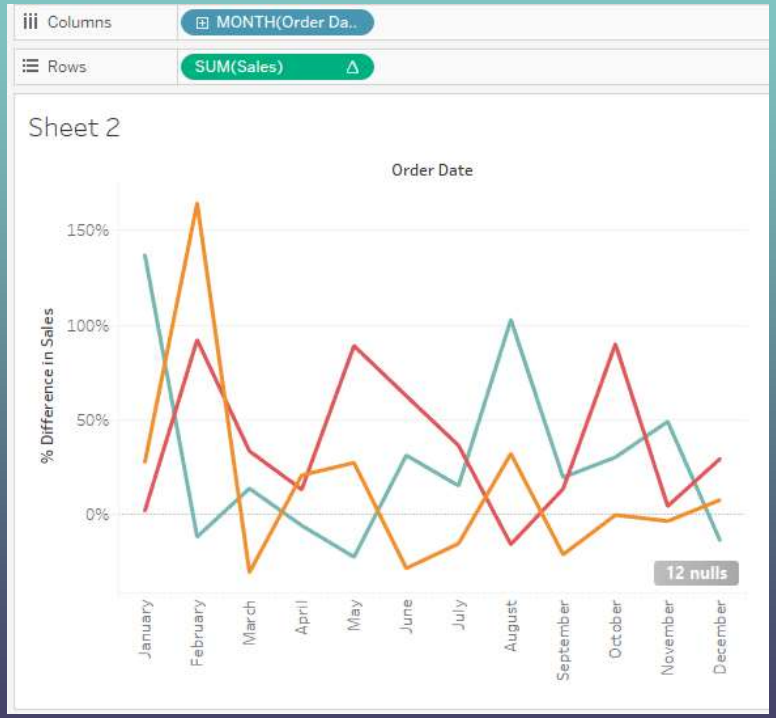

Order Date / Category / Sub-Category
2015
2016
2017
2018
Furniture
Tech..
Sales
60K
50K
40K
30K
20K
10K
0K
Bookcases
Tables
Machines

Contrast Sales Data by Year

Drag Order Date to Color



YoY Growth by Editing Rows

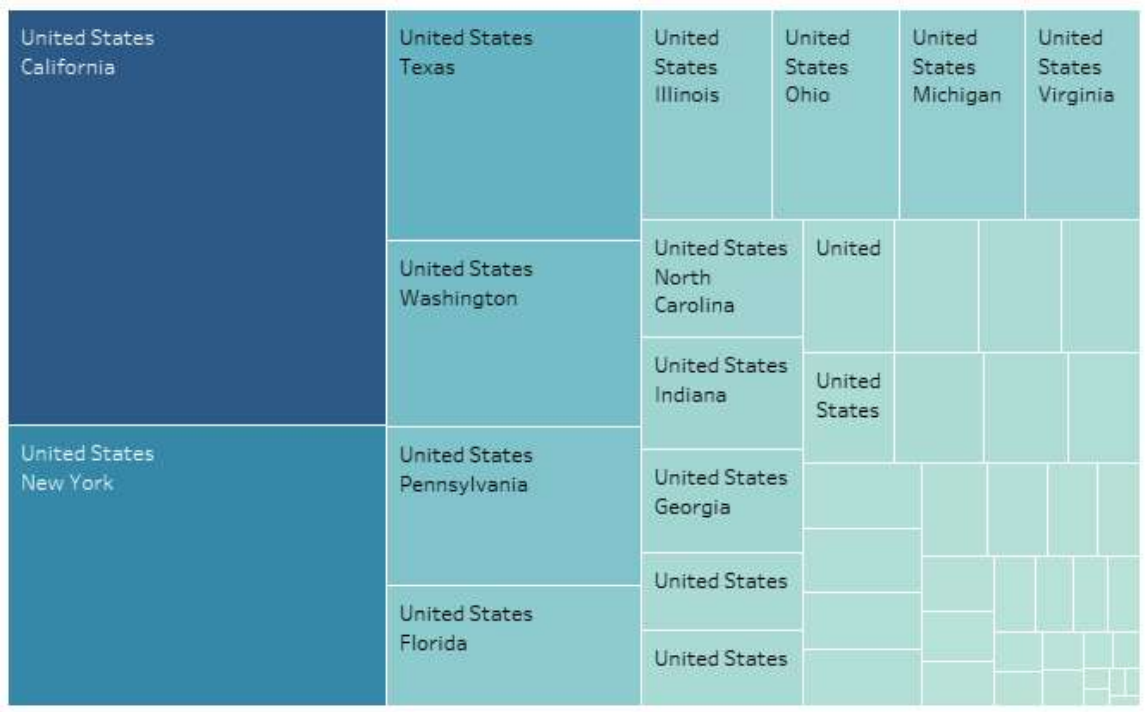


Map Representation

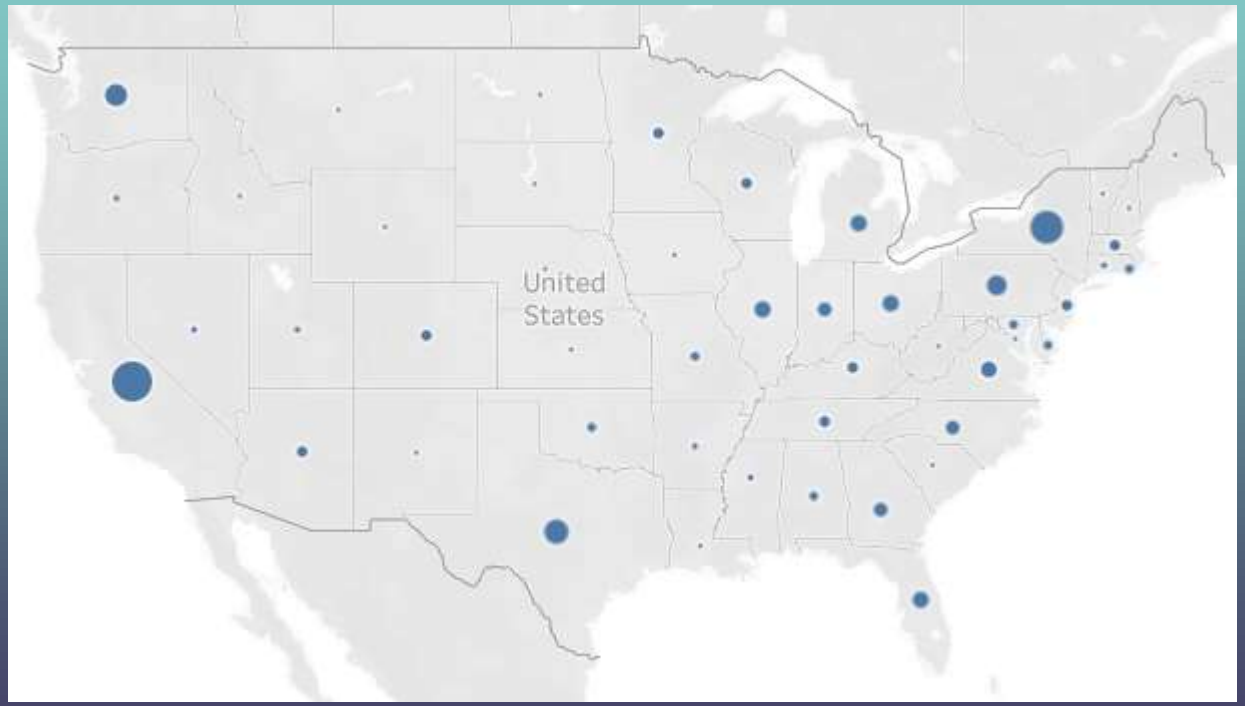
Use Control Button to select multiple attribute “Country”, “State” and “Sales”.

Try different functions in show me panel on the right-hand side.

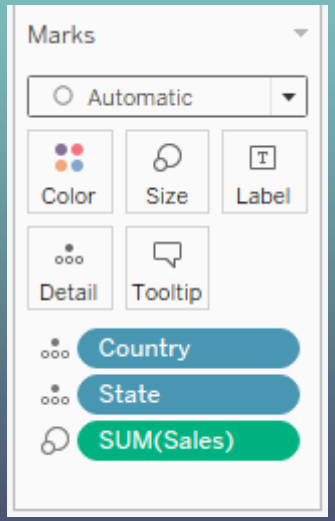
Map Representation



Map Representation



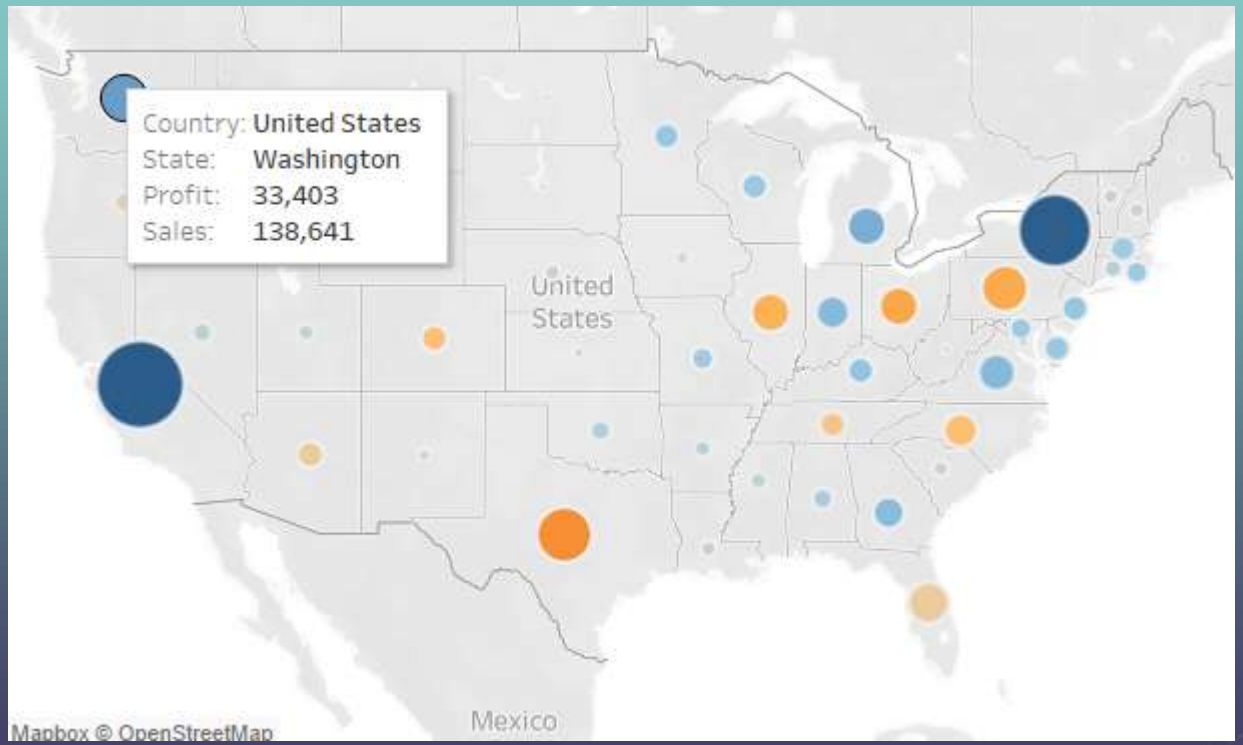
Map Representation

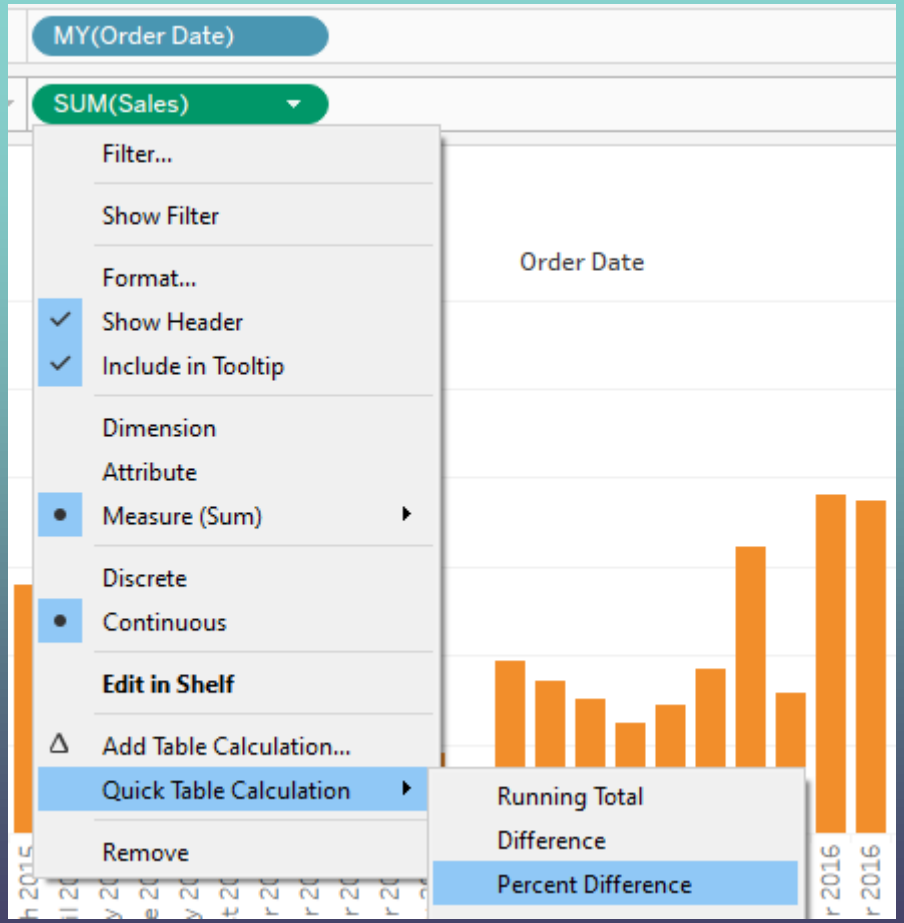


We can use profit to select the color of the circles in the map.

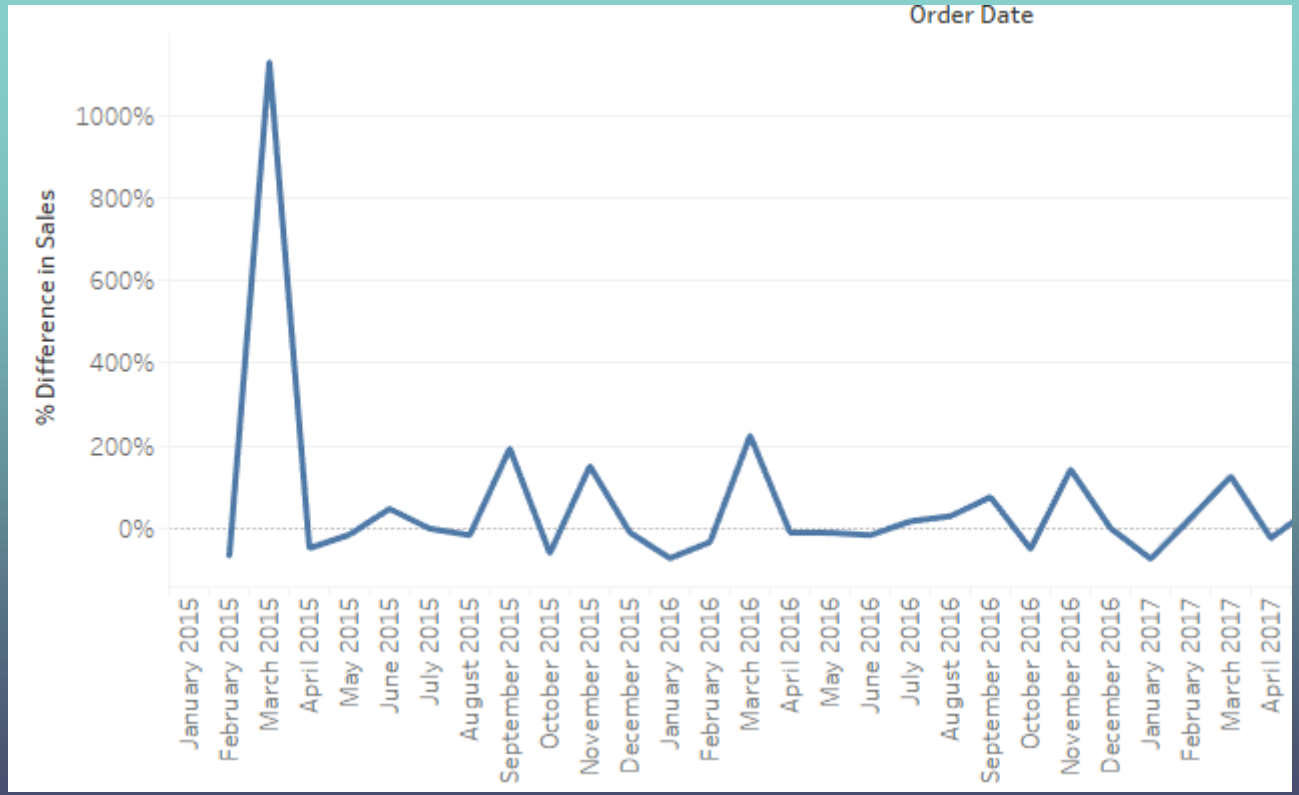
Moreover, we adjust the size of the circles.

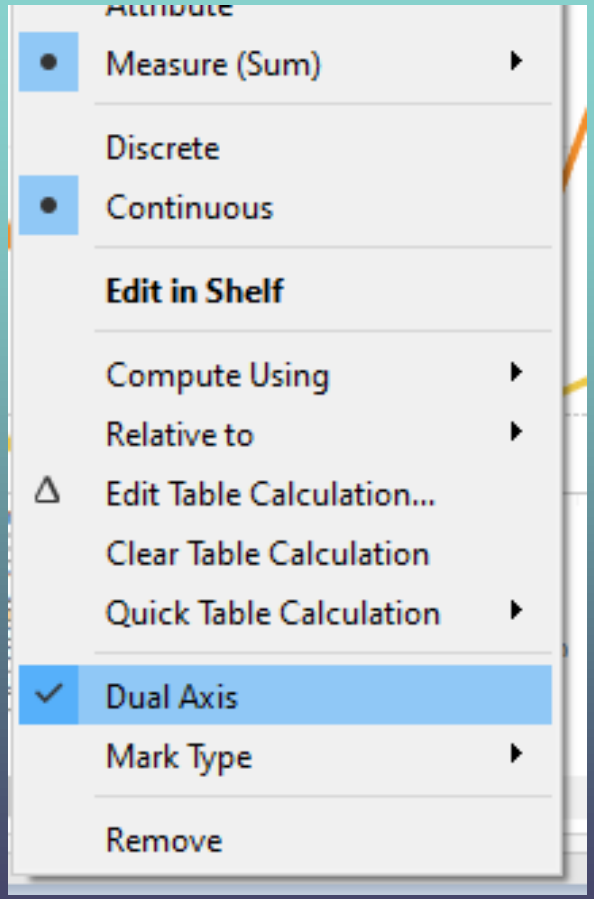
Map Representation



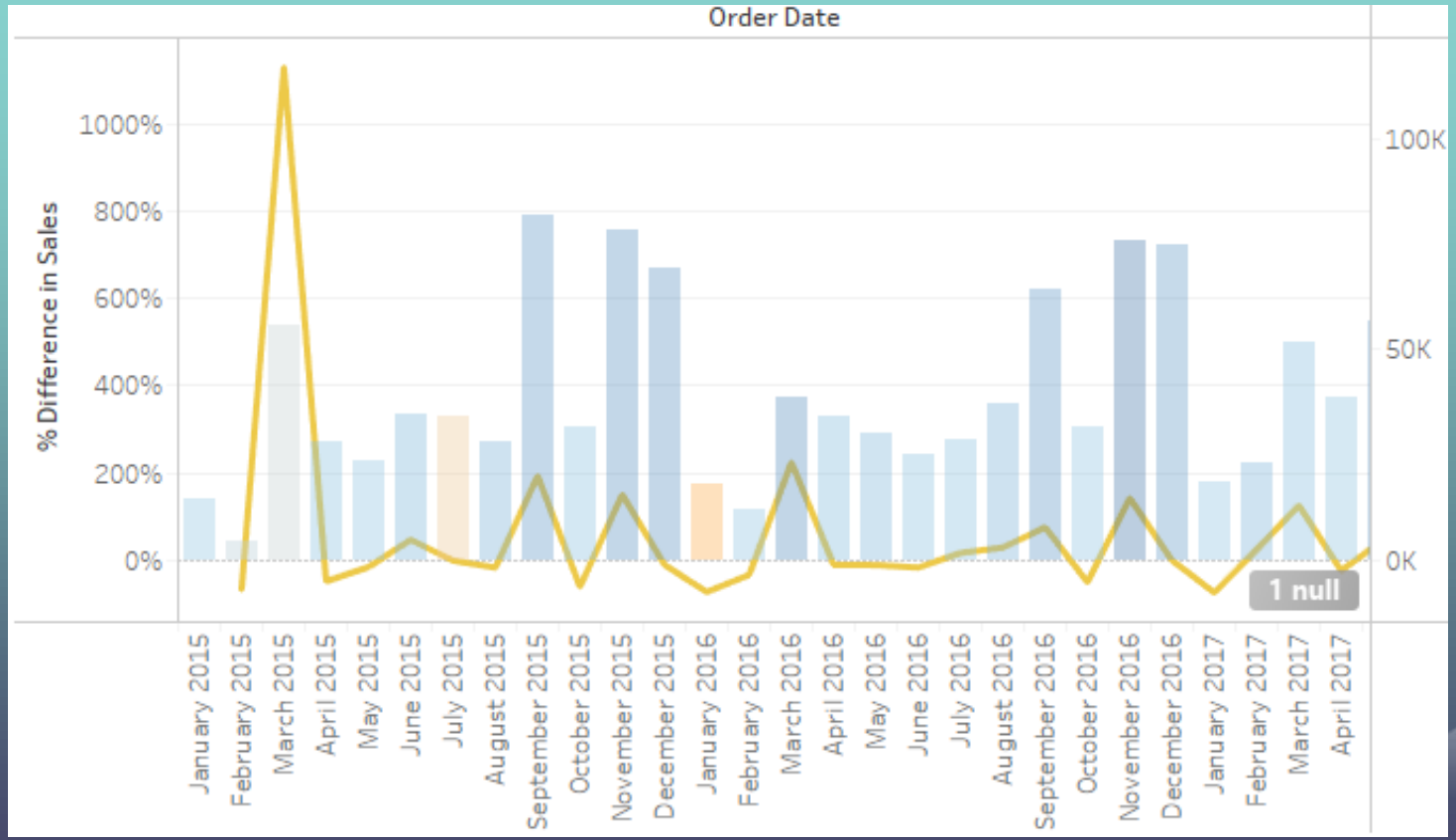


Here, we want to calculate how the sales change over time.


Order Date
% Difference in Sales
1000%
800%
600%
400%
200%
0%
January 2015
February 2015
March 2015
April 2015
May 2015
June 2015
July 2015
August 2015
September 2015
October 2015
November 2015
December 2015
January 2016
February 2016
March 2016
April 2016
May 2016
June 2016
July 2016
August 2016
September 2016
October 2016
November 2016
December 2016
January 2017
February 2017
March 2017
April 2017



We want to put both sales and sales change in a single plot.


Order Date
% Difference in Sales
1000%
800%
600%
400%
200%
0%
100K
50K
0K
1 null
January 2015
February 2015
March 2015
April 2015
May 2015
June 2015
July 2015
August 2015
September 2015
October 2015
November 2015
December 2015
January 2016
February 2016
March 2016
April 2016
May 2016
June 2016
July 2016
August 2016
September 2016
October 2016
November 2016
December 2016
January 2017
February 2017
March 2017
April 2017

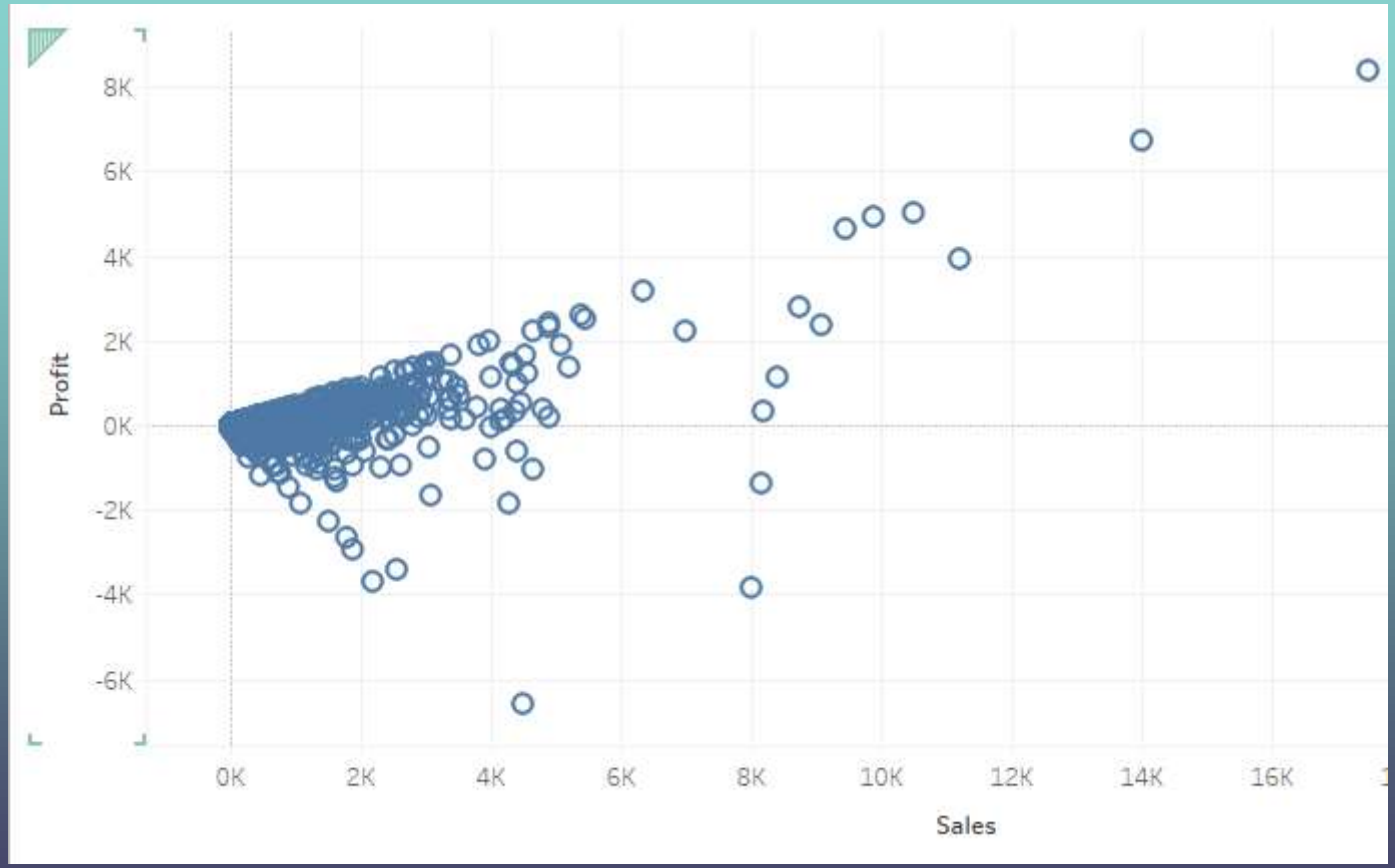
Regression Analysis

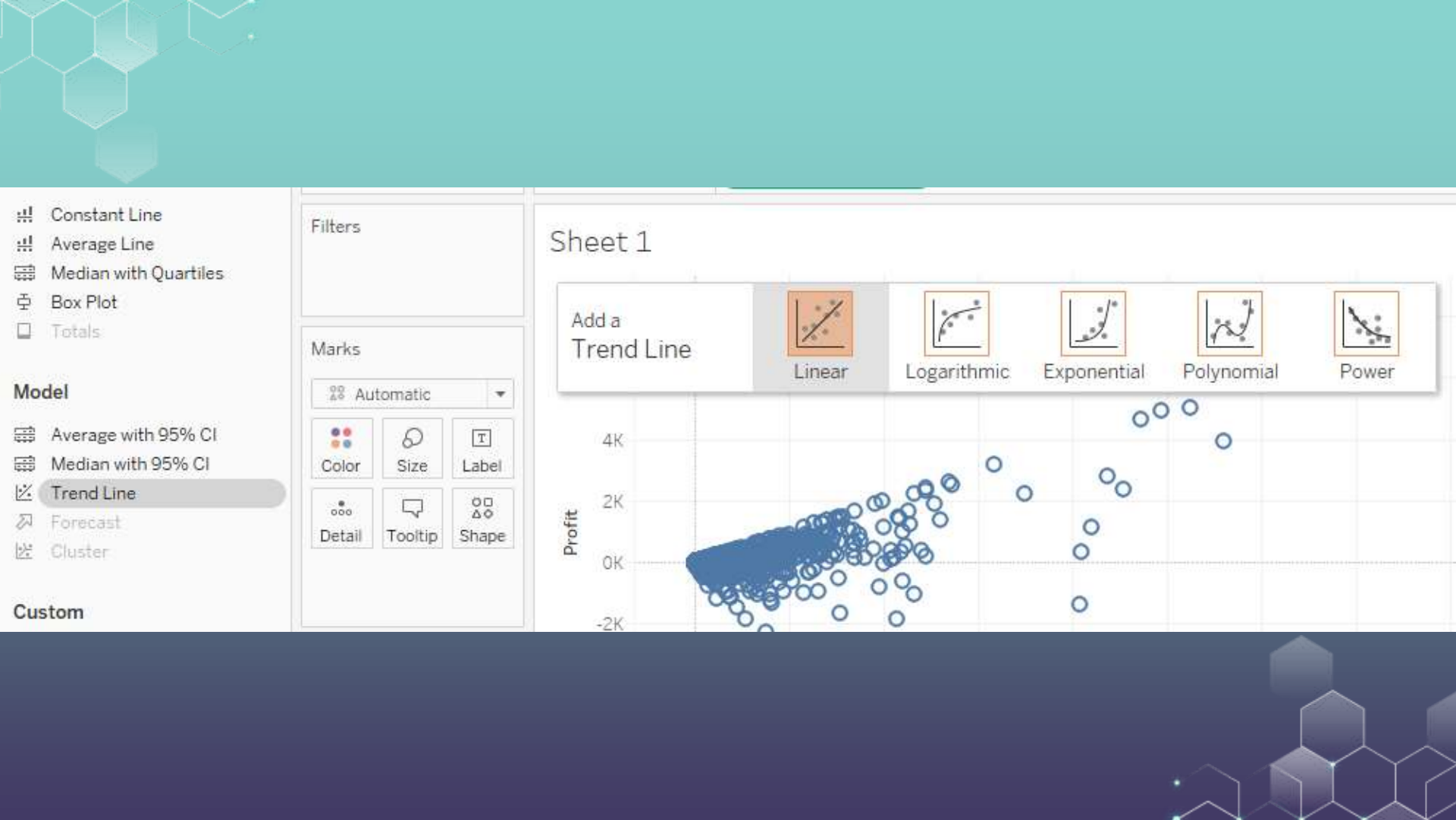
Suppose that we want to run the following regression:

$$\text{Profit} = a + b\ \text{Sales}$$

Go to analytics page. Use sales as columns and profit as rows.

Set “dimensions” for each variable.


8K
6K
4K
2K
0K
-2K
-4K
-6K
Profit
0K
2K
4K
6K
8K
10K
12K
14K
16K
Sales


Constant Line
Average Line
Median with Quartiles
Box Plot
Totals
Model
Average with 95% CI
Median with 95% CI
Trend Line
Forecast
Cluster
Custom
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Shape
Sheet 1
Add a
Trend Line
Linear
Logarithmic
Exponential
Polynomial
Power
4K
2K
0K
-2K
Profit


Profit = 0.182356*Sales + -17.7608
R-Squared: 0.22892
P-value: < 0.0001

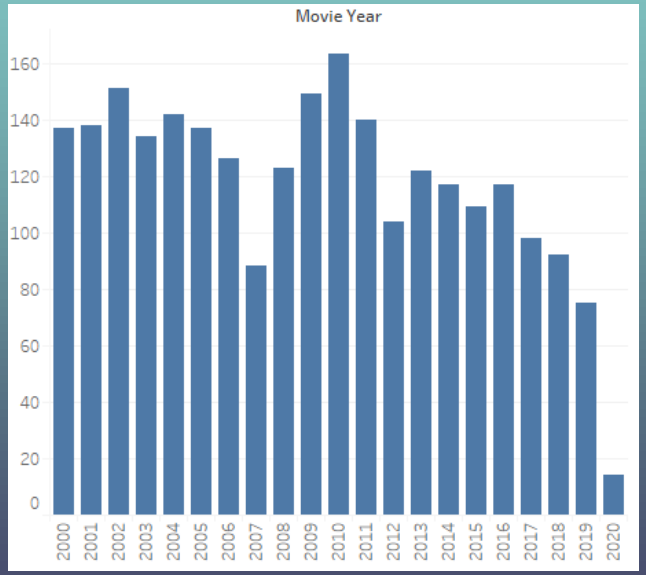
THE MOVIE DATASET

THE MOVIE DATASET

This dataset contains information for 2,476 movies for the last 20 years, including their budget, domestic box office, international box office, actors, genre etc.

The original dataset can be found here.

Number of Movies by Year



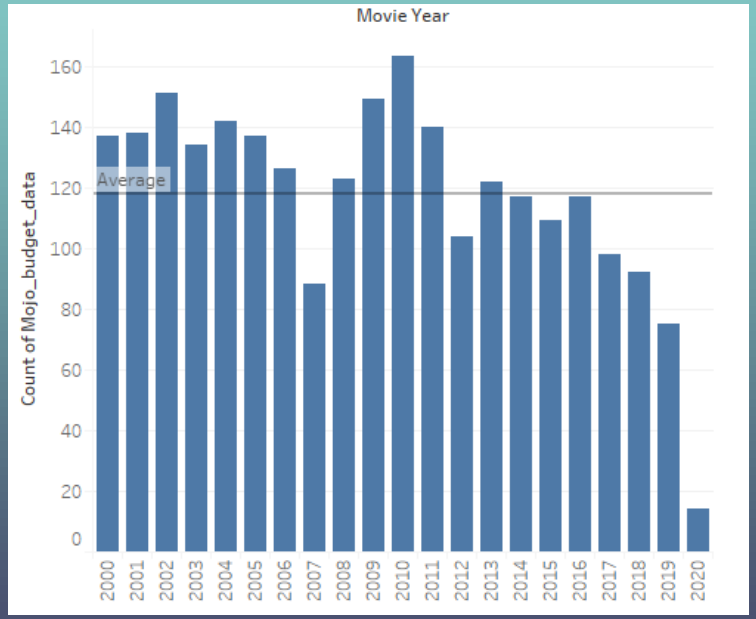
Columns: Movie Year

Set “Movie Year” as discrete

Rows: CNT(Mojo_budget_data)

Set “Marks” to be “Bar”

Number of Movies by Year

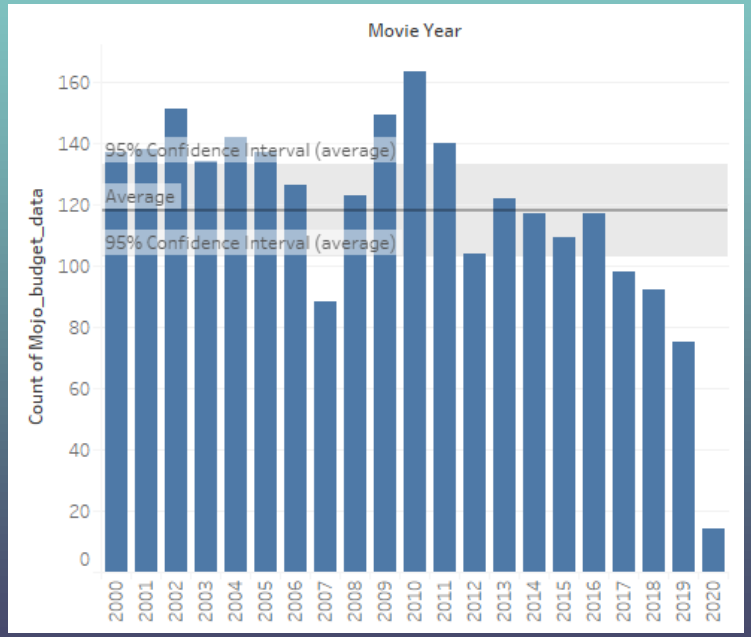


Go to Analytics Menu.

Drag "Reference Line" to the Plot.

Similarly, you can add median to your plot.

Number of Movies by Year

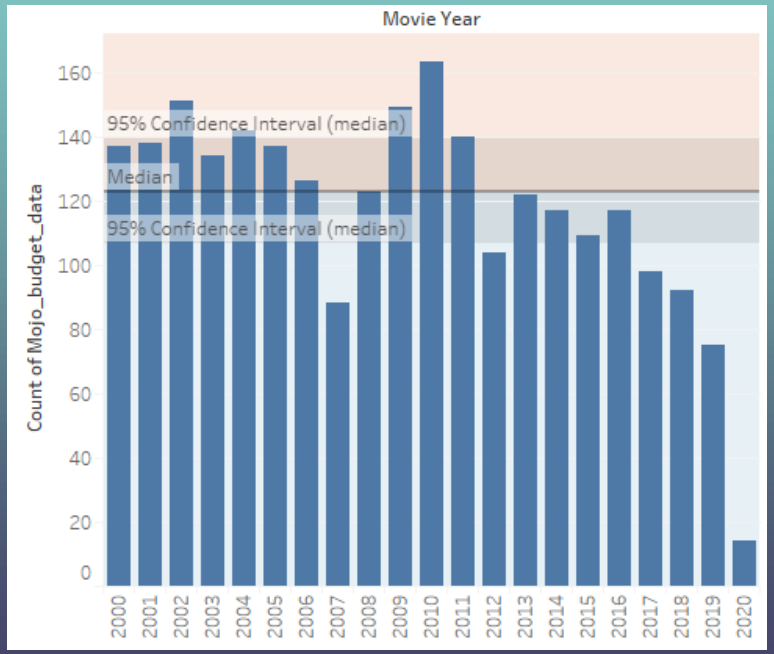


Now we plot the confidence interval of the mean.

After dragging "Reference Line" to the Plot, change "Line only" to "Line and confidence interval".

You can specify your CI.

Number of Movies by Year

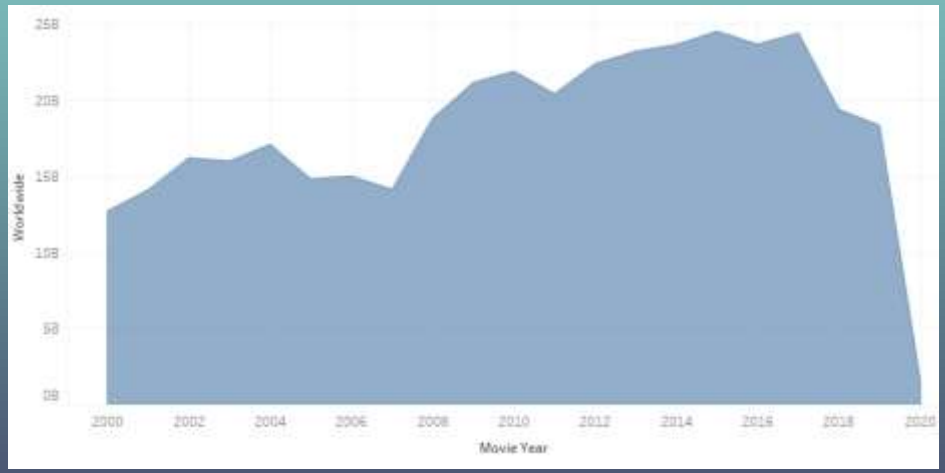


Now we add colors to the plot.

After dragging "Reference Line" to the Plot, you can change colors under "Formatting": "Fill above" and "Fill below".

The color in the confidence band is darker than that outside.

Area Chart

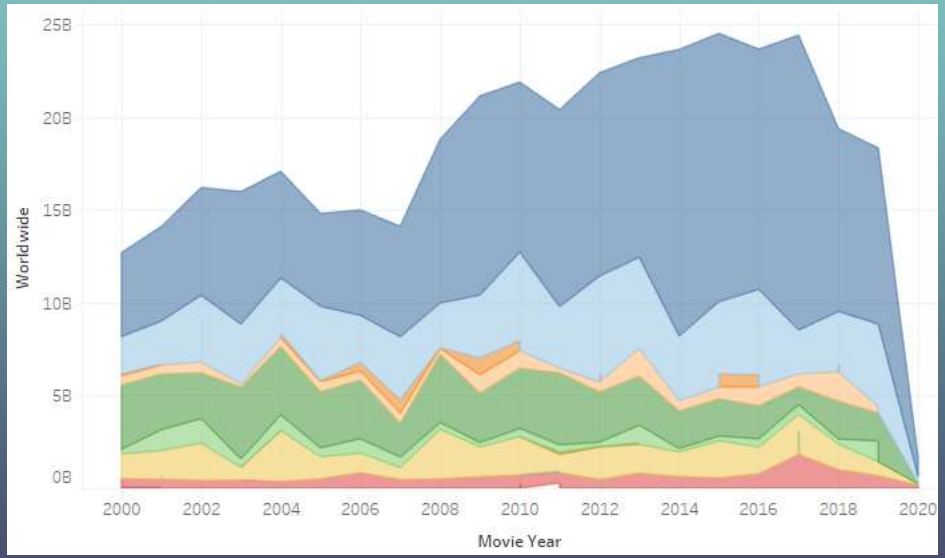


Columns: Movie Year

Rows: Worldwide (SUM)

Show me:

Area Chart

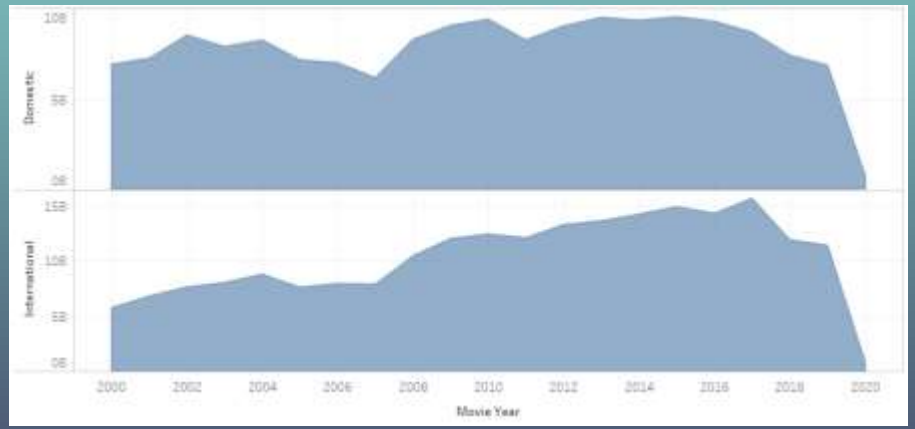


Columns: Movie Year

Rows: Worldwide (SUM)

Drag "Genre 1" to Color under Marks Menu.

Area Chart

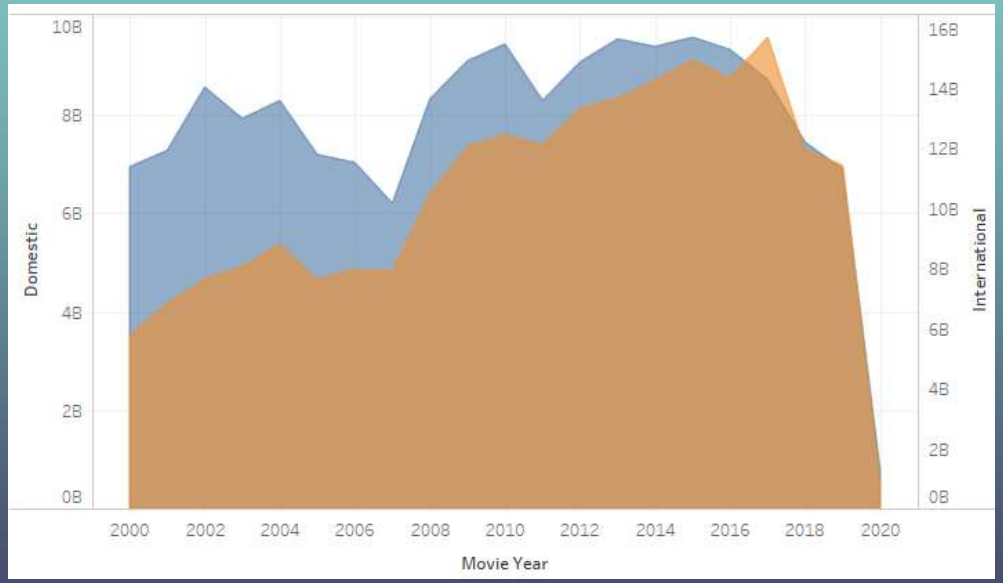


Columns: Movie Year

Rows: Domestic (SUM) and International (SUM)

Select Area Chart

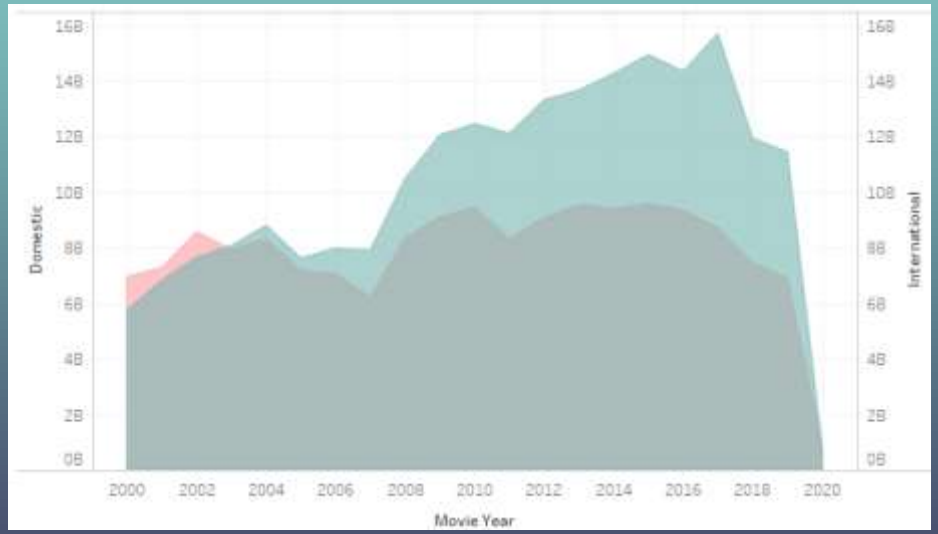
Area Chart



Right Click International (Second Chart)

Choose "Dual Axis"

Area Chart

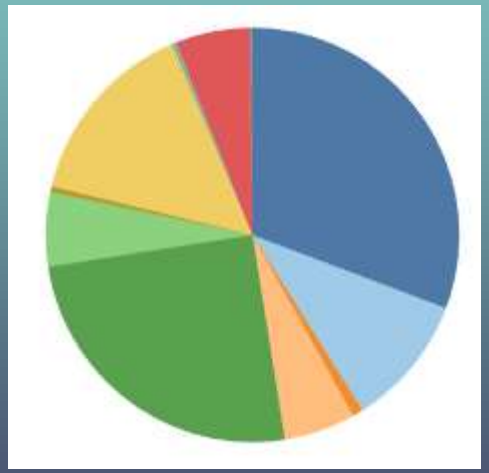


Right Click Domestic

Choose "Synchronize Axis"

Update Color on the right-hand side if you want

Genre Pie Chart



Columns: Genre 1

Rows: CNT(Mojo_budget_data)

Show me: 

Change “standard” (top) to “entire view”

Genre by Year

	Movie Year								
Genre 1	2000	2001	2002	2003	2004	2005	2006	2007	2008
Action	33	34	42	39	37	32	25	22	38
Adventure	15	6	15	15	9	14	18	8	13
Animation	1	2			1	1	1	2	1
Biography	4	3	5	7	7	9	6	7	7
Comedy	51	52	33	40	47	43	36	20	34
Crime	7	12	18	11	8	9	7	10	5
Documenta..				1	1			1	
Drama	19	21	32	14	27	22	22	9	19
Family					1				
Fantasy		1		1				1	
Horror	6	6	6	5	4	7	11	7	6
Music									
Mystery	1	1		1					
Romance									
Sci-Fi								1	

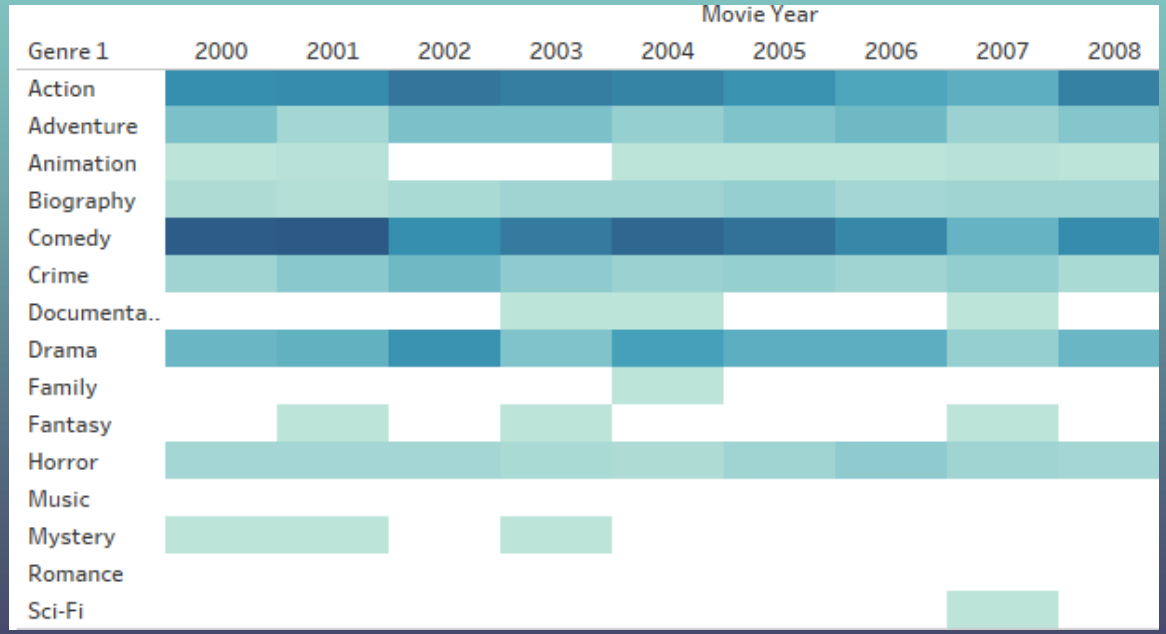
Columns: Movie Year (Discrete)

Rows: Genre 1 (you can choose genre 2…as well)

Set “Marks” to Text

Drag Mojo_budget_data to Labels (or Text) under the Marks box

Genre by Year

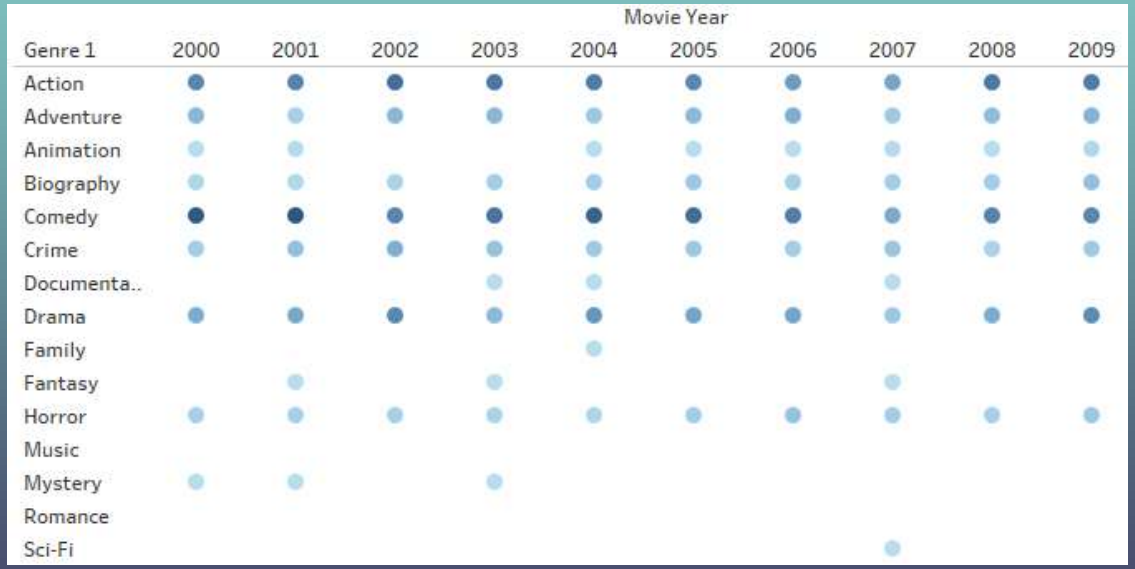


Columns: Movie Year

Rows: Genre 1 (you can choose genre 2…as well)

Drag Mojo_budget_data to Color under the Marks box

Genre by Year

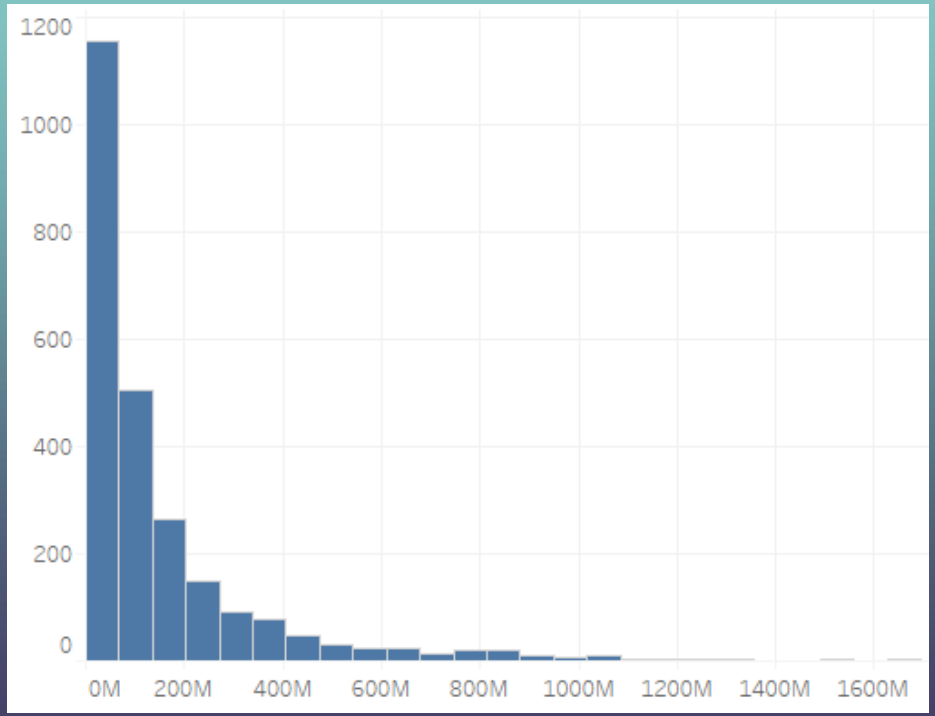


Columns: Movie Year

Rows: Genre 1 (you can choose genre 2…as well)

You can explore the settings under Marks (here we use circles).

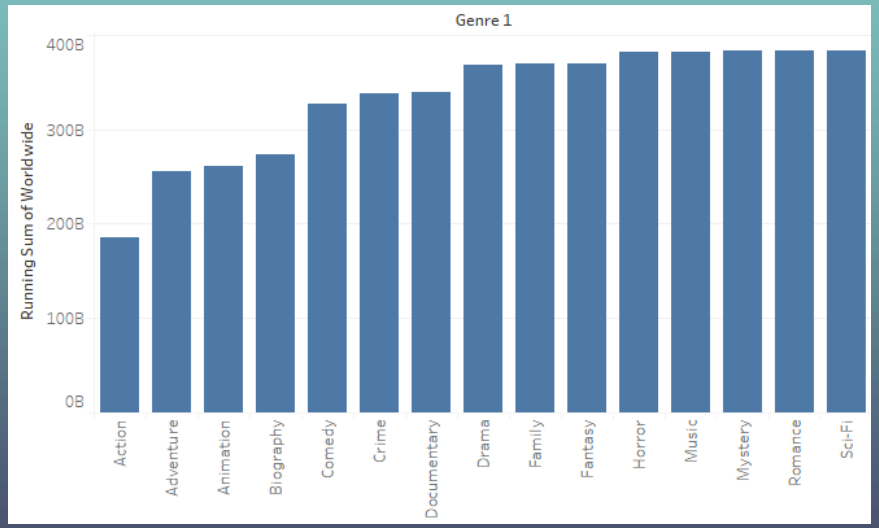
Worldwide Box Office Histogram



Columns: Worldwide

Choose  in “Show me”.

Waterfall

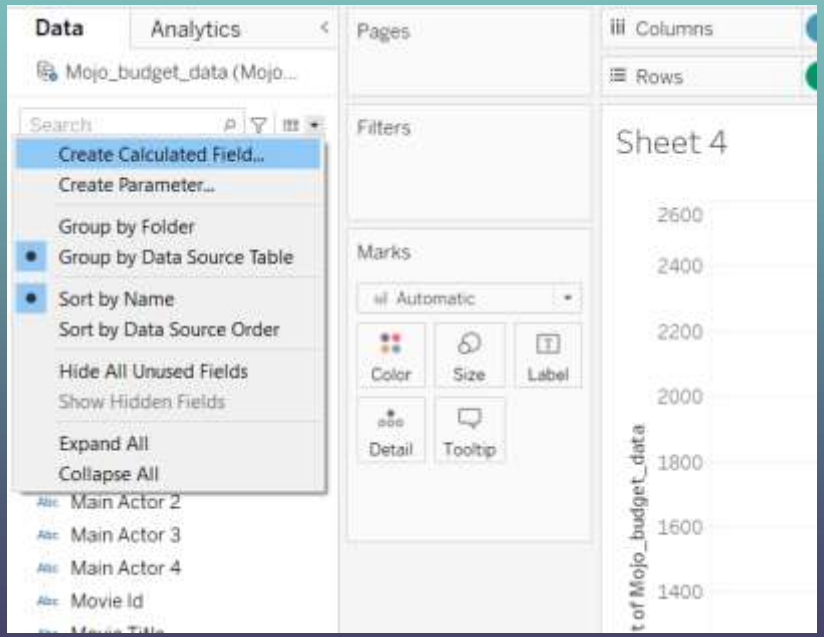


Columns: Genre 1

Rows: Worldwide

Right click Worldwide, select Add Table Calculation. Then under Calculation Type, choose Running Total.

Waterfall

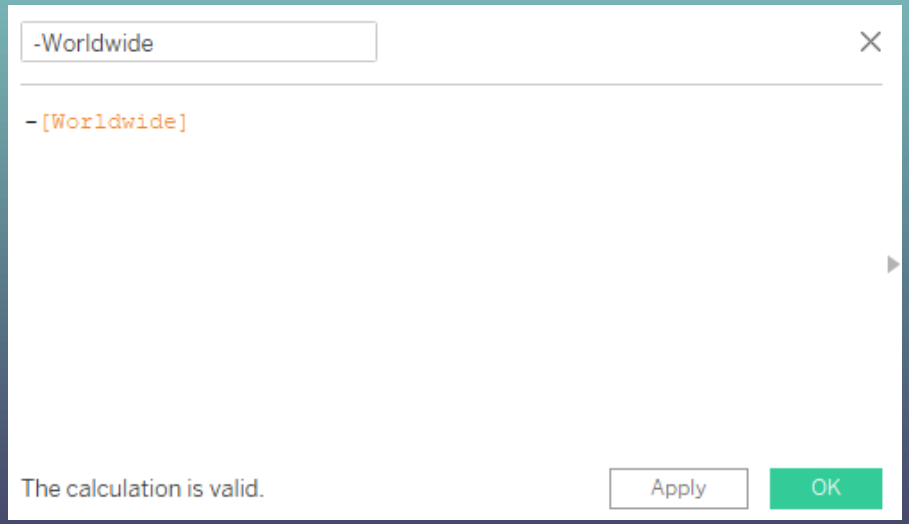


Under Data Menu, choose Create Calculate Field

Input the following calculation.

Click OK to proceed.

Waterfall

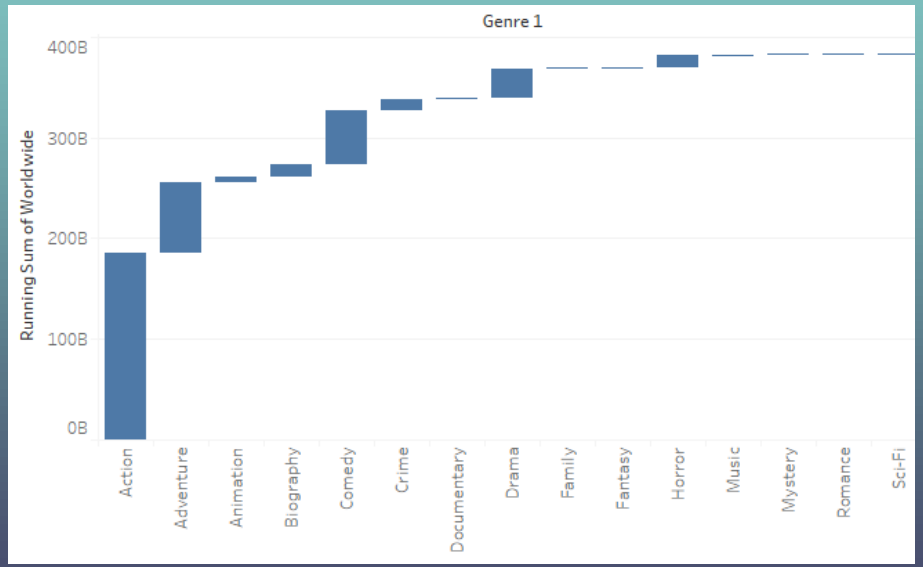


Under Analysis Menu, choose Create Calculate Field

Input the following calculation.

Click OK to proceed.

Waterfall

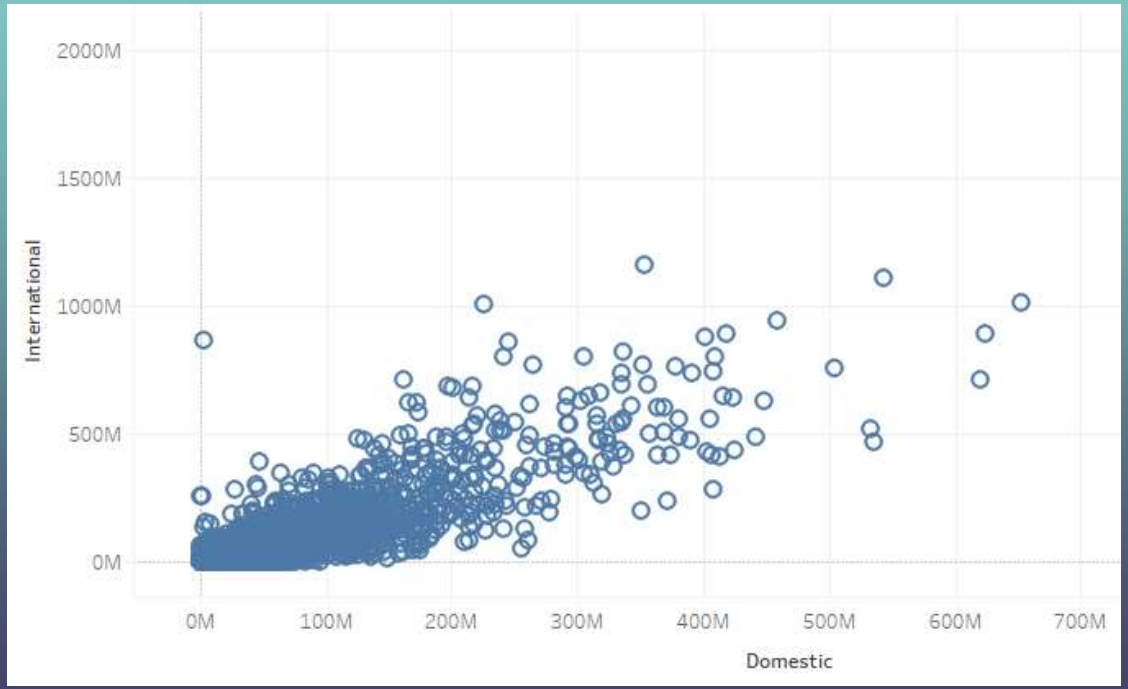


There is a new variable "-Worldwide".

Under Marks, change Automatic to Gantt Bar.

Drag new variable "-Worldwide" to "Size" under Marks.

International vs. Domestic Box Revenue

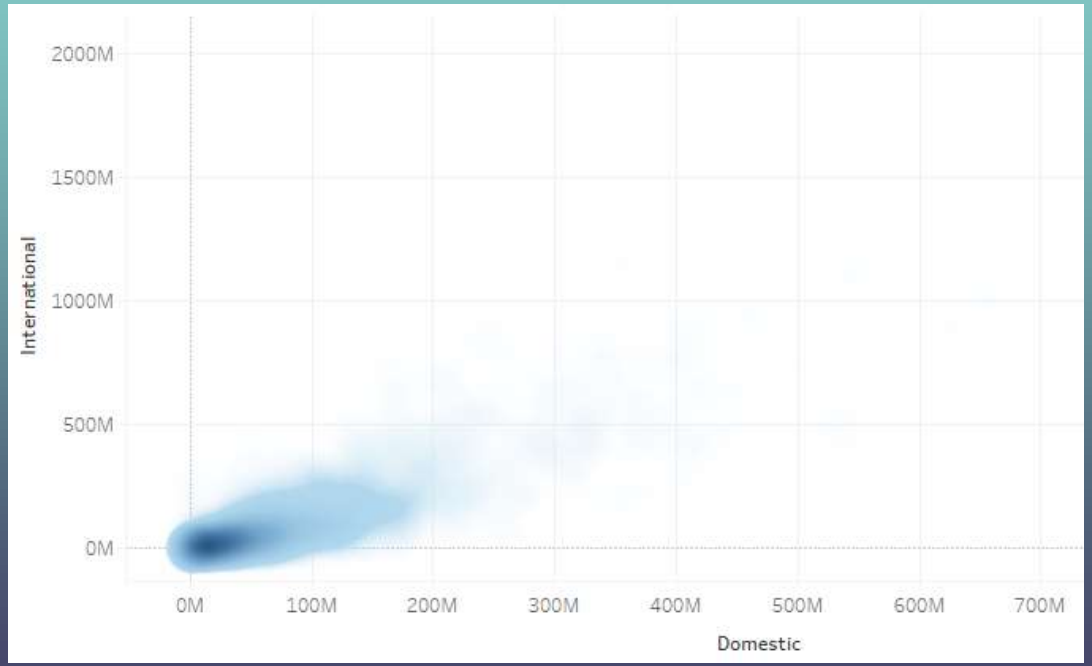


Columns: Domestic

Rows: International

Set both variables as Dimension

International vs. Domestic Box Revenue



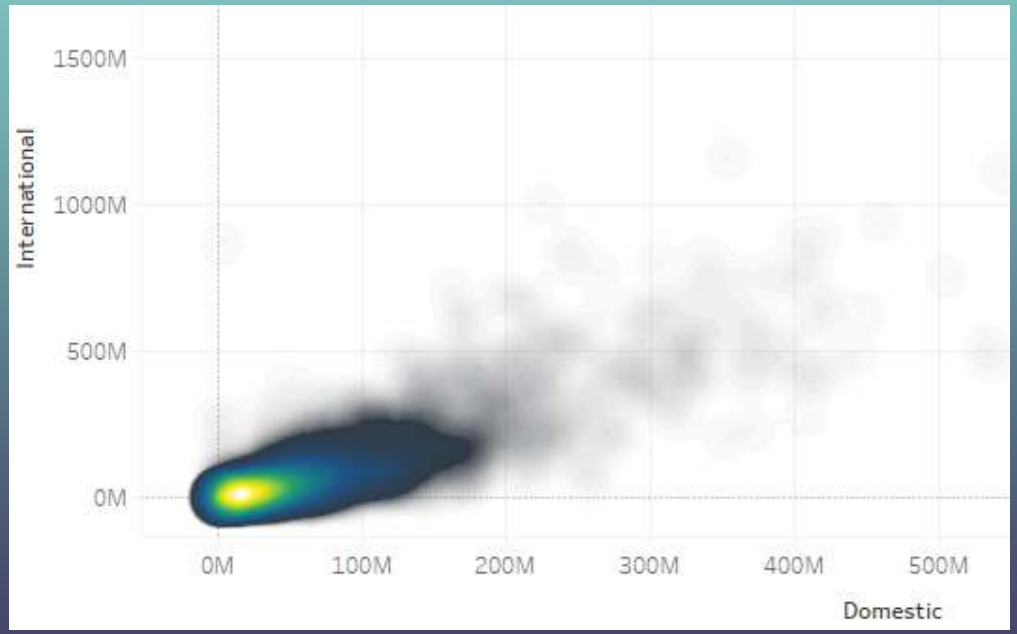
Columns: Domestic

Rows: International

Set both variables as Dimension

Change "Automatic" to "Density" under Marks

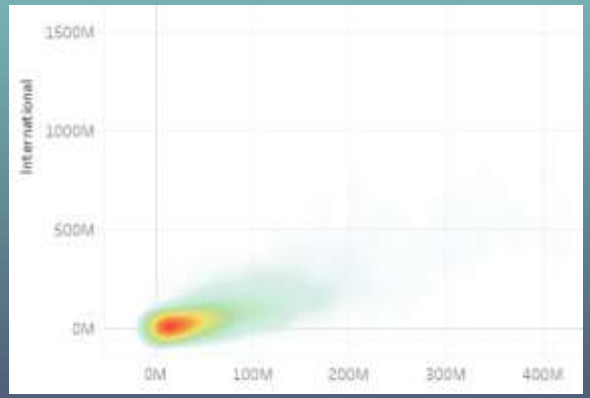
International vs. Domestic Box Revenue



Change "Automatic" to "Density" under Marks

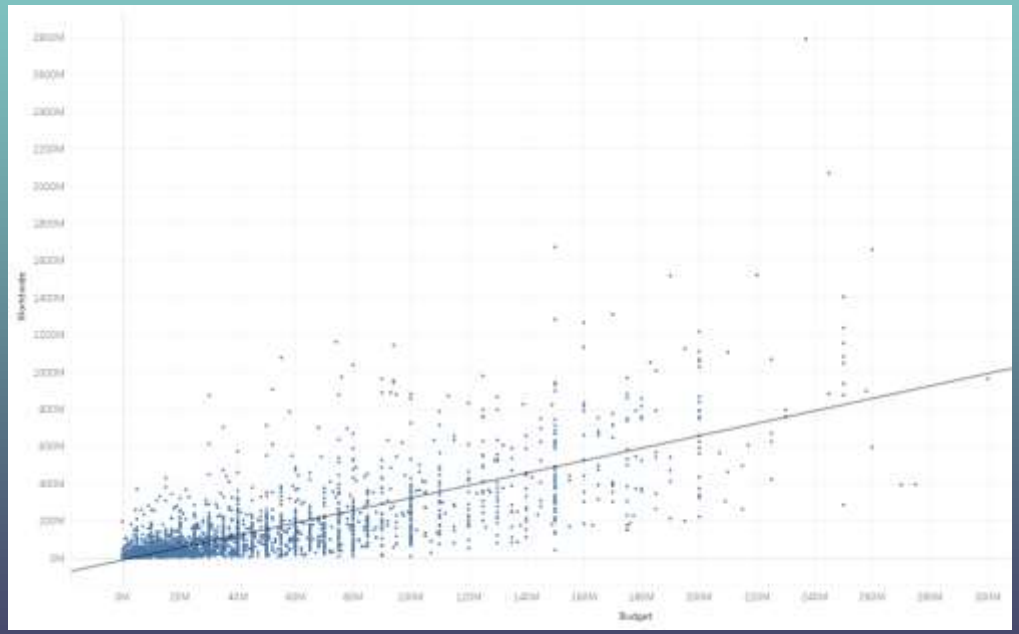
Set Colors to "Density-Multicolor" under Marks

Density Heatmap



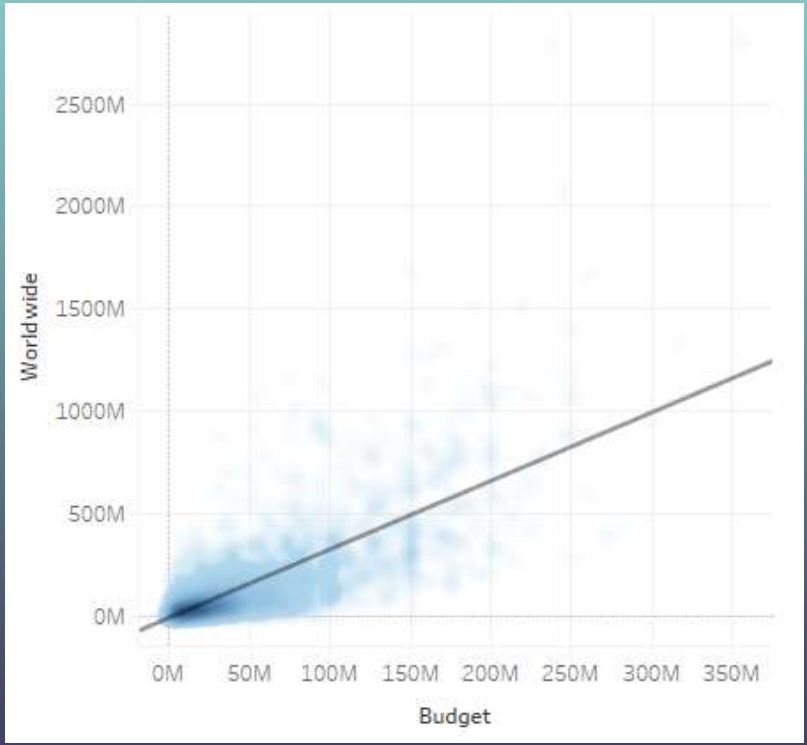
In a density heatmap, in places where the observations are dense (i.e., many points in the area), the heatmap displays a warm color. In places where observations are sparse, the heatmap displays a cold color.

Regression Line



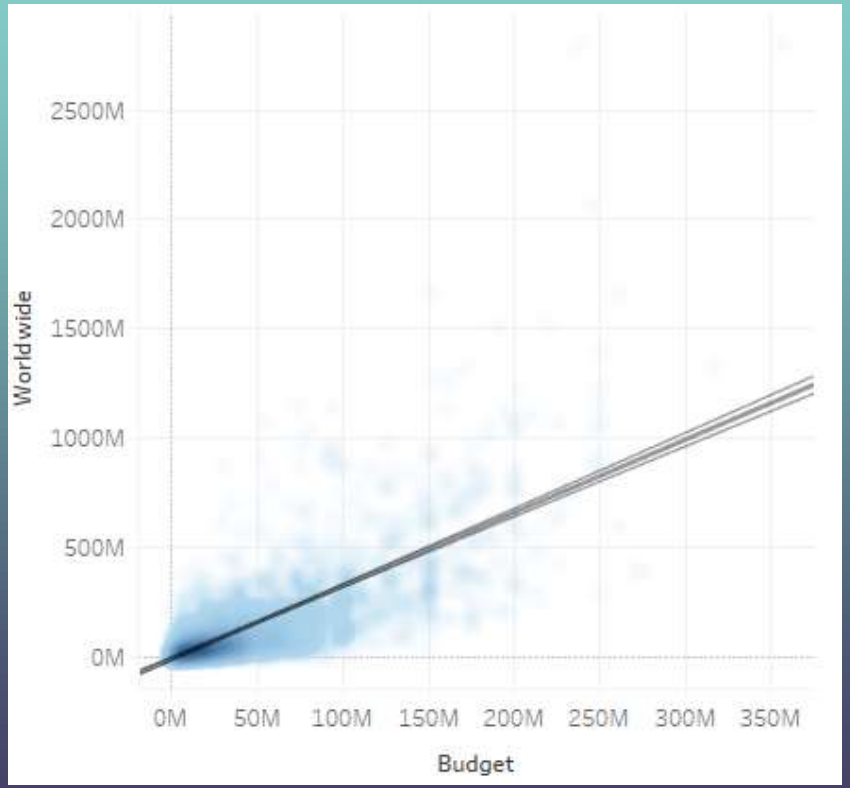
Here, we use budget as the independent variable (i.e., Column) and worldwide box office as the dependent variable (i.e., Row) and draw the regression line.

Regression Line



Then, we change the figure to a density heatmap to make it look better.

Regression Line



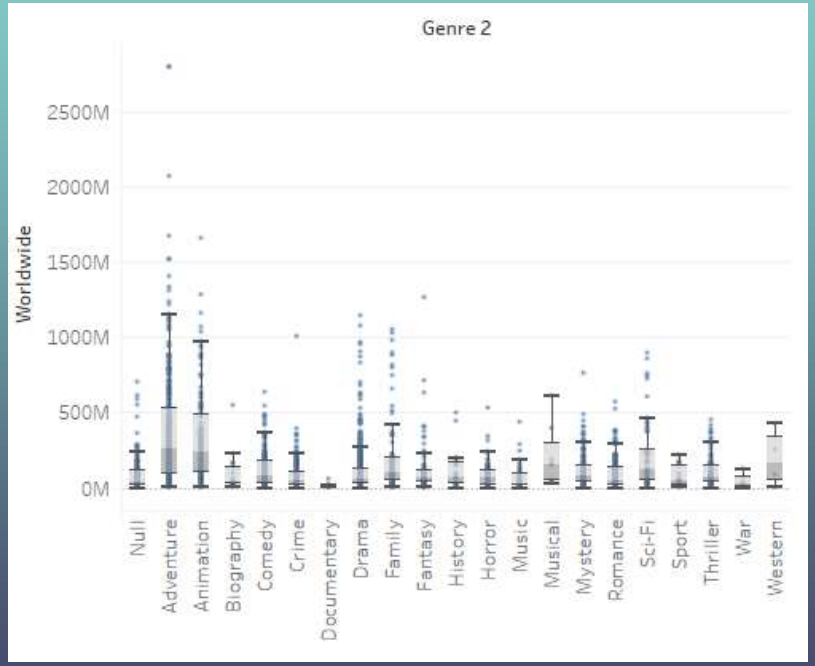
Click and edit the regression line.

Under Options menu, check "show confidence bands".

Confidence Bands

Tableau confidence bands show upper and lower 95% confidence lines. That is, with probability 95%, your regression line falls within your confidence bands. You can visit the Wikipedia for the detailed description of the confidence bands (click here to visit Wikipedia page).

Box and Whisker Plot



Box and Whisker Plot

Meaning of the box: 50% of the observations fall within the box (25% of the data are greater than the box limit and 25% are smaller than the box limit).

Line within the box: The median of your data.

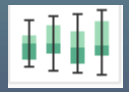
The other two lines, called upper and lower Whisker, are more complex. For details, please refer the Wikipedia on this topic here.

Box and Whisker Plot

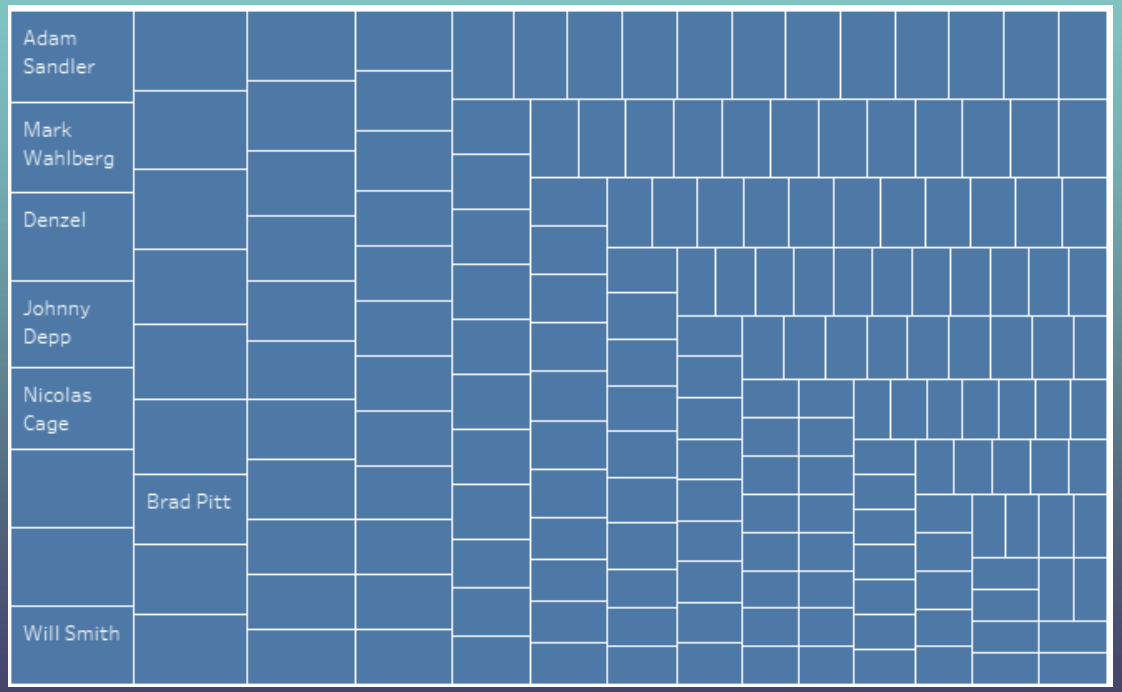
How to draw the Box and Whisker Plot?

Use Genre 2 as rows and Worldwide as columns.

Under Analysis (top bar), *uncheck* "Aggregate Measures".

Under Show me, choose 

Actor/Actress Treemap



Columns: Main Actor 1

Rows: "Mojo_budget_data (Count)" to Size

Check Treemap

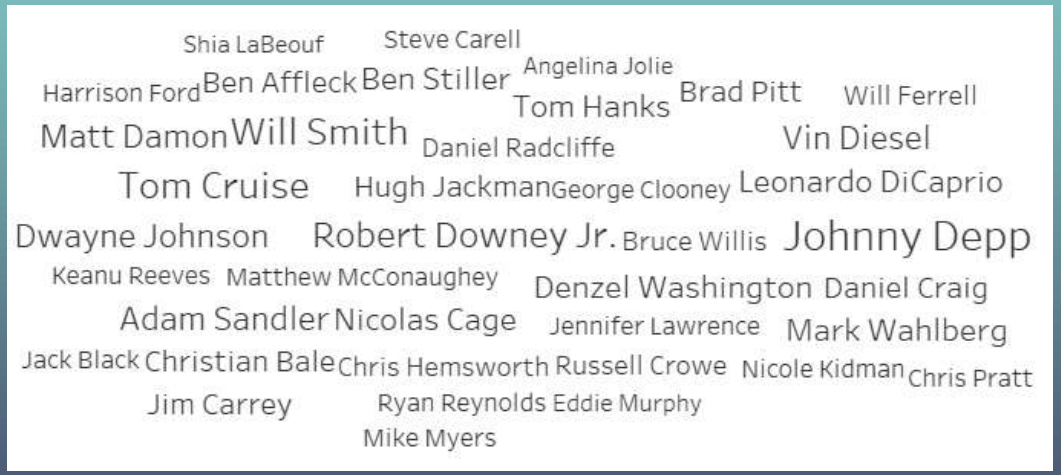
You can filter the figure by Mojo_budget_data (Count)

Actor/Actress Word Cloud

Following the previous step, you can change "Automatic" to "Text" under the Marks box.

Again, you can use filters to only keep the frequent names.

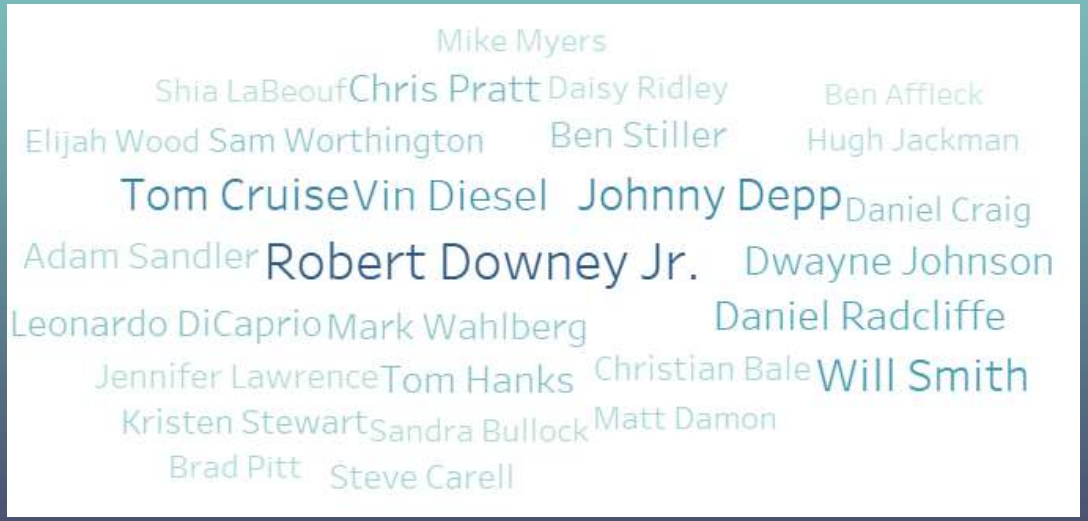
Actor/Actress Word Cloud



Now, replace "Count(Mojo_budget_data)" by "Worldwide".

Then, a bigger name means the actor/actress brings more box office worldwide.

Actor/Actress Word Cloud



And you can also color the actor/actress by his/her box office worldwide.

Drag "Worldwide" to color and you will get this cloud.

Word Cloud by Tagul (Optional Topic)

Dituhui.com Creating Maps for China (Optional Topic)


Tableau 2022
2022
New Year
Big Data
Algorithms
Tableau
Marketing
Big Data
Marketplaces
Tableau
Christmas
Marketing
2022
New Year
Marketing
2022 Christmas