



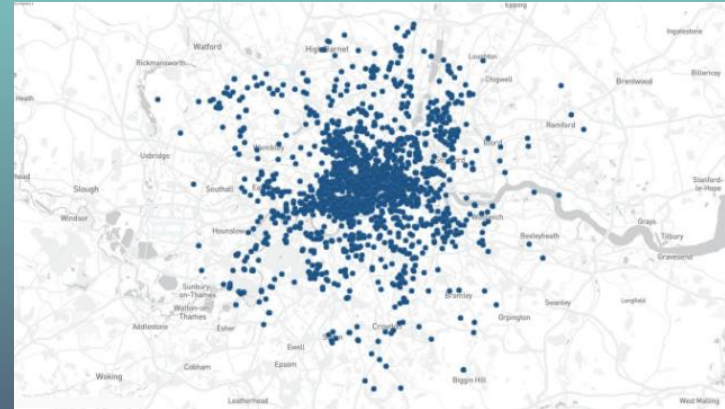
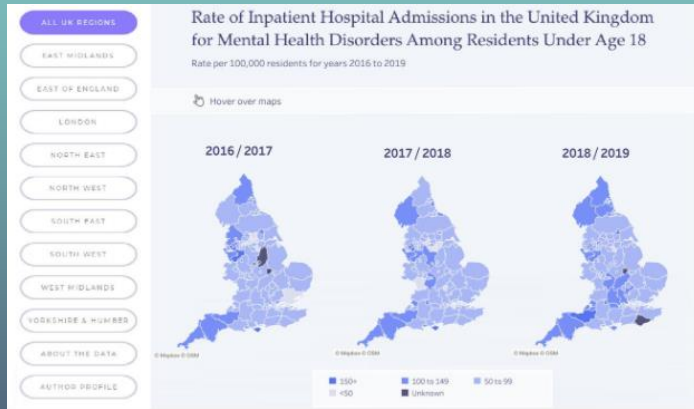
Introduction to Tableau

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

Sort fields Show all

#	Orders	Orders	Orders	Orders	Orders
Row ID	Order ID	Order Date	Order Date	Order Date	Customer ID
1	CA-2017-152156	11/8/2017	11/8/2017	11/8/2017	CG-12520
2	CA-2017-152156	11/8/2017	11/8/2017	11/8/2017	CG-12520
3	CA-2017-138688	6/12/2017	6/12/2017	6/12/2017	DV-13045
4	US-2016-108966	10/11/2016	10/11/2016	10/11/2016	SO-20335
5	US-2016-108966	10/11/2016	10/11/2016	10/11/2016	SO-20335
6	CA-2015-115812	6/9/2015	6/9/2015	6/14/2015	BH-11710
7	CA-2015-115812	6/9/2015	6/9/2015	6/14/2015	BH-11710

Number (decimal)
Number (whole)
Date & Time
• Date
String
Boolean
✓ Default

The screenshot shows the Tableau Desktop interface. On the left, the 'Data' pane is visible, displaying a list of fields under the 'Superstore_Data' data source. The fields are categorized into dimensions and measures. A red arrow points to the 'Postal Code' field, which is a dimension. Another red arrow points to the 'Profit' field, which is a measure. The 'Marks' card is set to 'Automatic'. The 'Filters' card is empty. The 'Columns' and 'Rows' shelves are also empty. The bottom of the interface shows the 'Data Source' and 'Sheet 1' tabs.

Dimensions

- Order ID
- Postal Code
- Product ID
- Product Name
- Region
- Row ID
- Segment
- Ship Date
- Ship Mode
- State
- Sub-Category
- Measure Names

Measures

- Discount
- Profit
- Quantity
- Sales
- Latitude (generated)
- Longitude (generated)
- Orders (Count)
- Measure Values

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

Measures are variables that can be calculated.

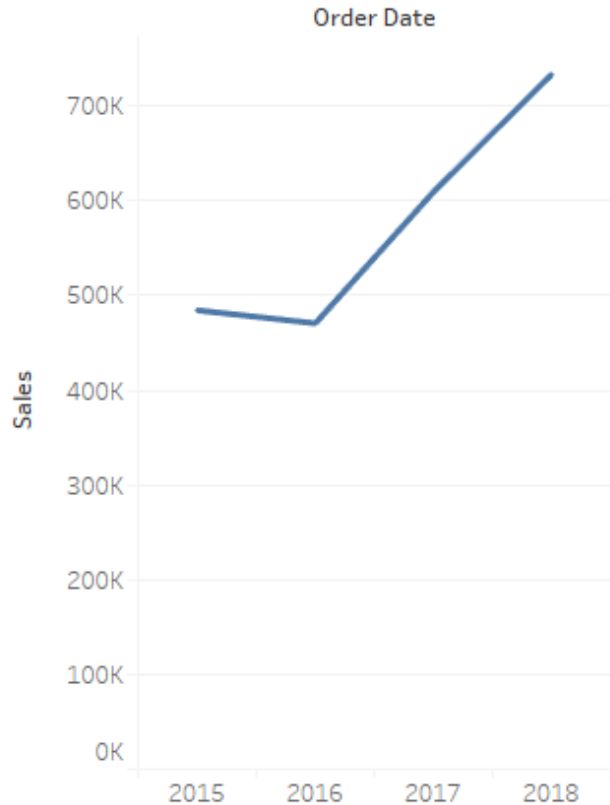
Dimensions are used to partition your measures (e.g., partition the data based on geographic locations).



Columns: YEAR(Order Date)

Rows: SUM(Sales)

Sheet 1



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.



Columns

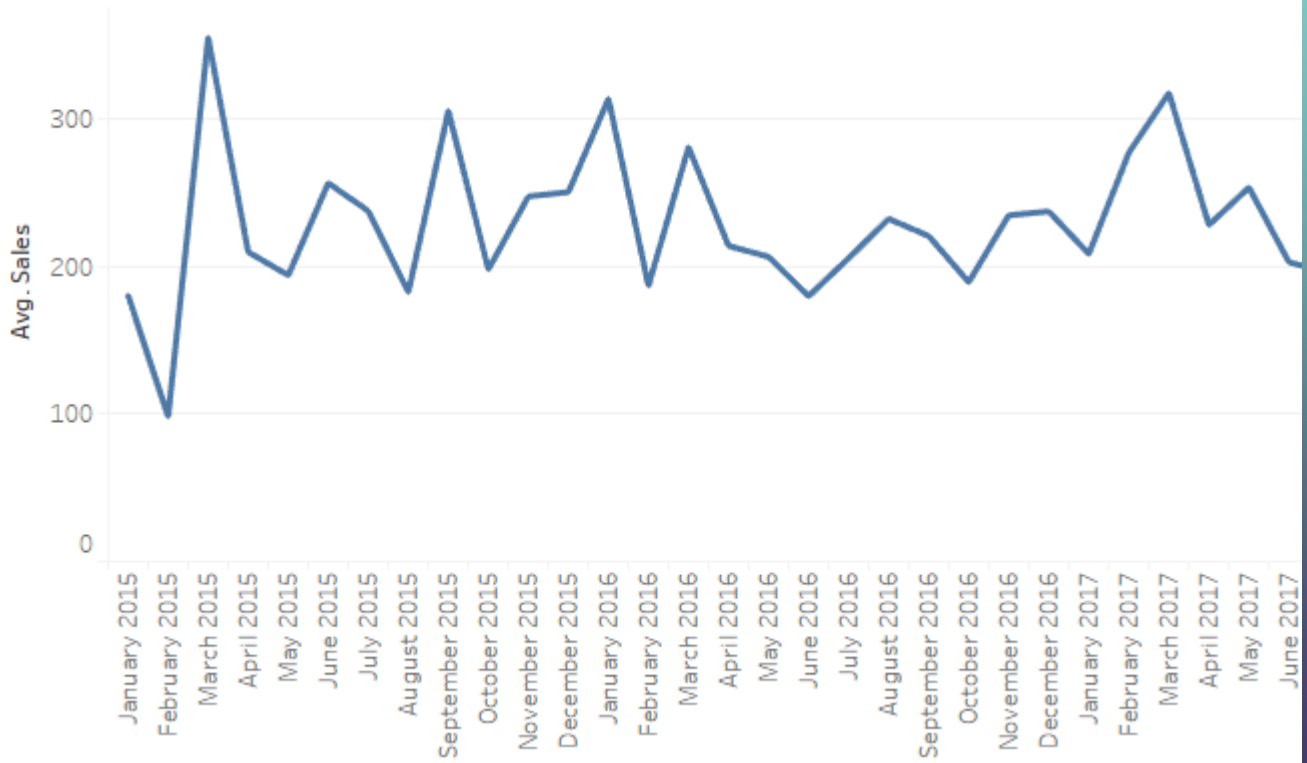
MY(Order Date)

Rows

AVG(Sales)

Sheet 1

Order Date



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



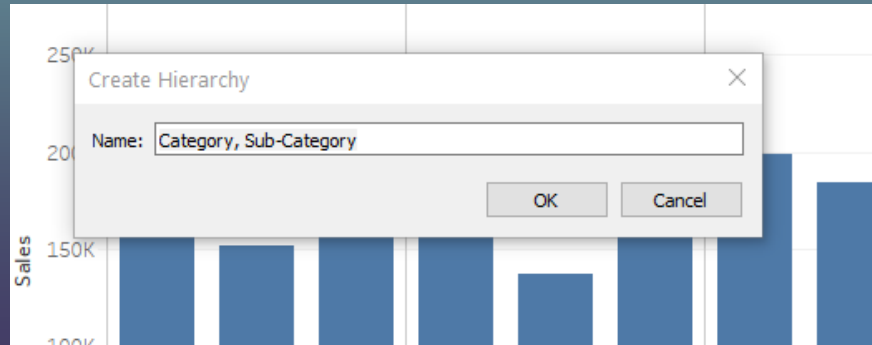
Adding Category As Another Classifier



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.

Columns	YEAR(Order Date)	Category	Sub-Category
Rows	SUM(Sales)		

Category

Sheet 1

Apply Filters

The screenshot shows a 'Tables' panel on the left with a list of fields. The 'Sub-Category' field is selected and highlighted with a blue bar. A context menu is open over this field, listing various actions. The 'Show Filter' option is highlighted in blue.

Table	Field
Product	Category
Product	Sub-Category
Product	Product ID
Product	Product Name
Product	Region
Product	Row ID
Product	Segment
Product	Ship Date
Product	Ship Mode
Product	State
Product	Measure Names
Product	Discount
Product	Profit
Product	Quantity
Product	Sales
Product	Latitude (generated)
Product	Longitude (generated)
Product	Orders (Count)

- Add to Sheet
- Show Filter
- Duplicate
- Rename
- Hide
- Aliases...
- Create
- Transform
- Convert to Measure
- Change Data Type
- Geographic Role
- Default Properties
- Group by
- Folders

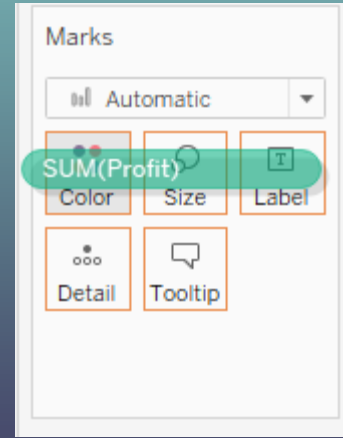
The screenshot shows a 'Sub-Category' filter panel. It contains a list of categories, each with a checkbox. The 'Bookcases' and 'Machines' categories are checked.

- (All)
- Accessories
- Appliances
- Art
- Binders
- Bookcases
- Chairs
- Copiers
- Envelopes
- Fasteners
- Furnishings
- Labels
- Machines
- Paper
- Phones
- Storage
- Supplies
- Tables

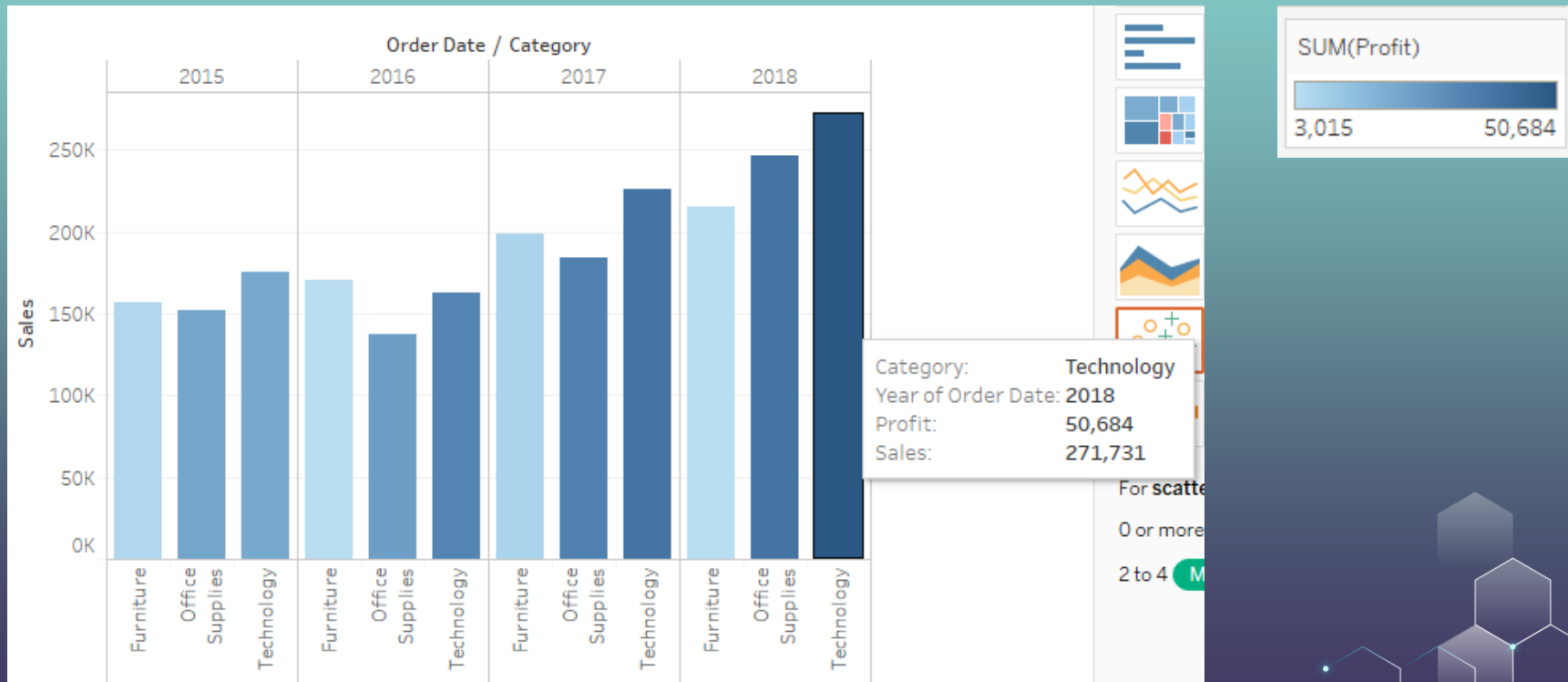
Color Your Output

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

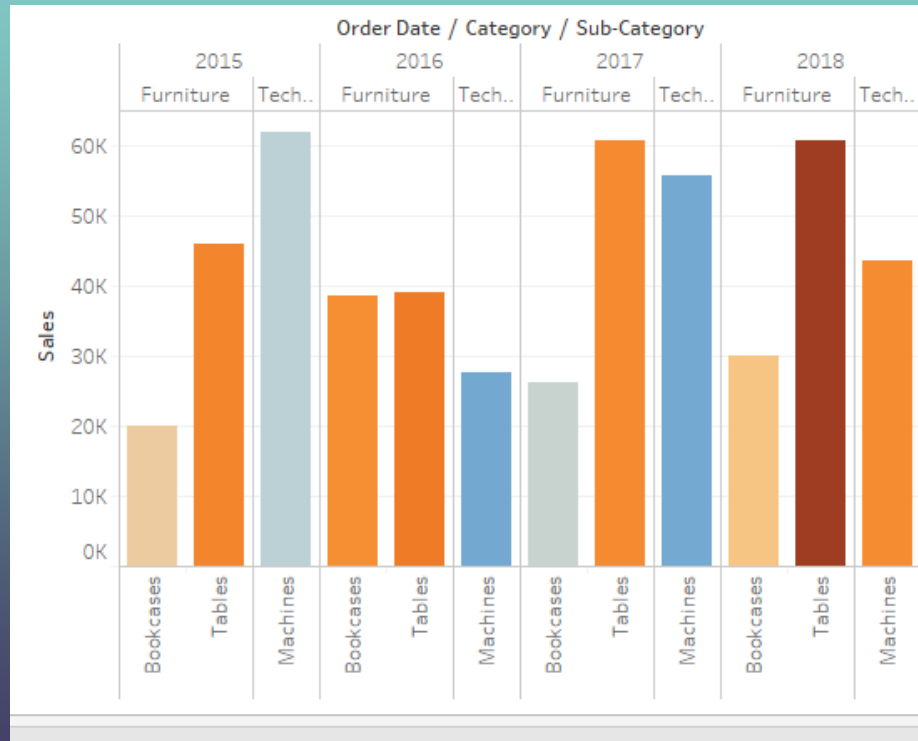
Drag profit to color.



Color Your Output



Apply Filters

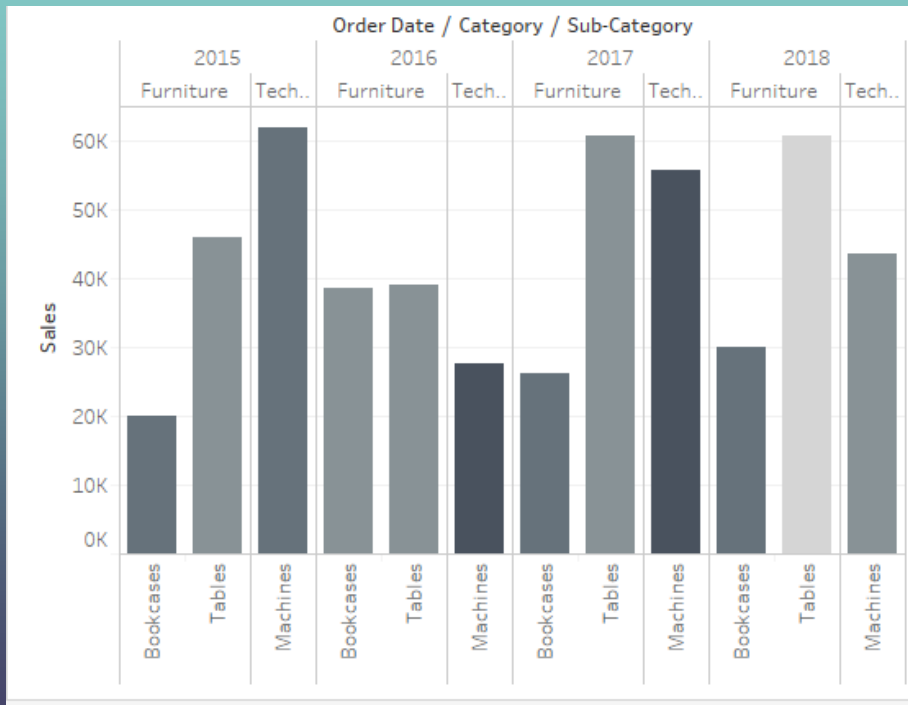


Edit Colors

The image displays a software interface with two overlapping dialog boxes. The background dialog, titled "Marks", features a dropdown menu set to "Automatic" and three buttons: "Color", "Size", and "Label". The "Color" button is active, showing a "Color" section with an "Edit Colors..." button, an "Opacity" slider at 100%, and an "Effects" section with "Border" and "Halo" dropdowns, both set to "Automatic".

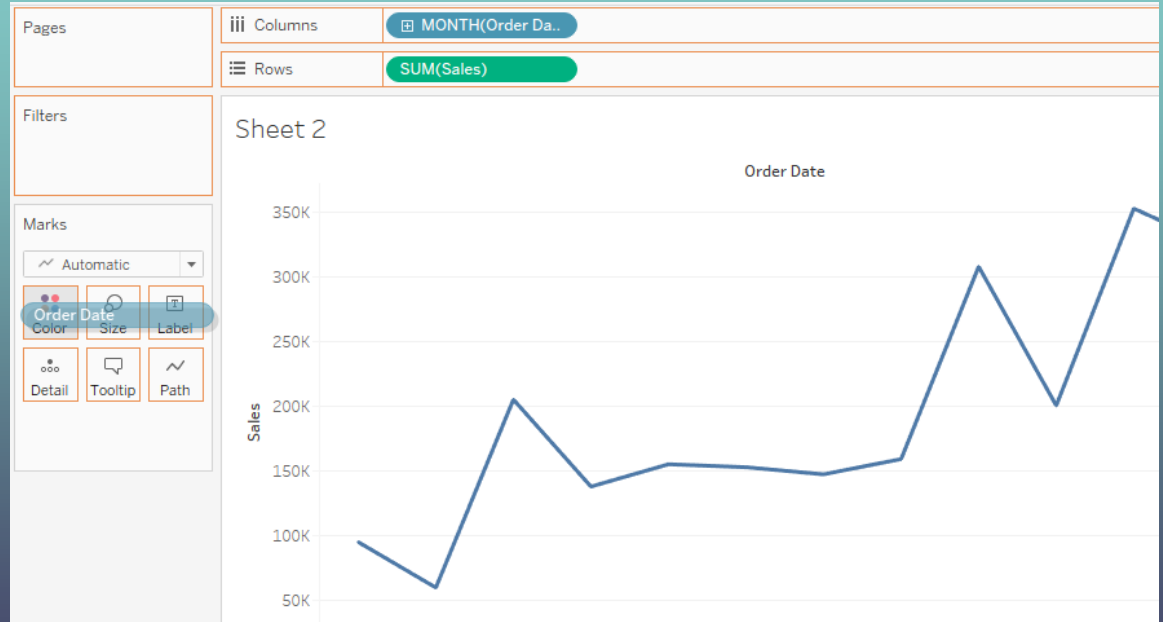
The foreground dialog, titled "Edit Colors [Profit]", has a "Palette" dropdown set to "Automatic". It displays a color gradient bar ranging from -8,141 (dark red) to 2,977 (dark blue). Below the bar are several options: Stepped Color (5 Steps), Reversed, Use Full Color Range, and Include Totals. An "Advanced >>" button is located at the bottom right of the dialog. At the very bottom are "Reset", "OK", "Cancel", and "Apply" buttons.

Edit Colors

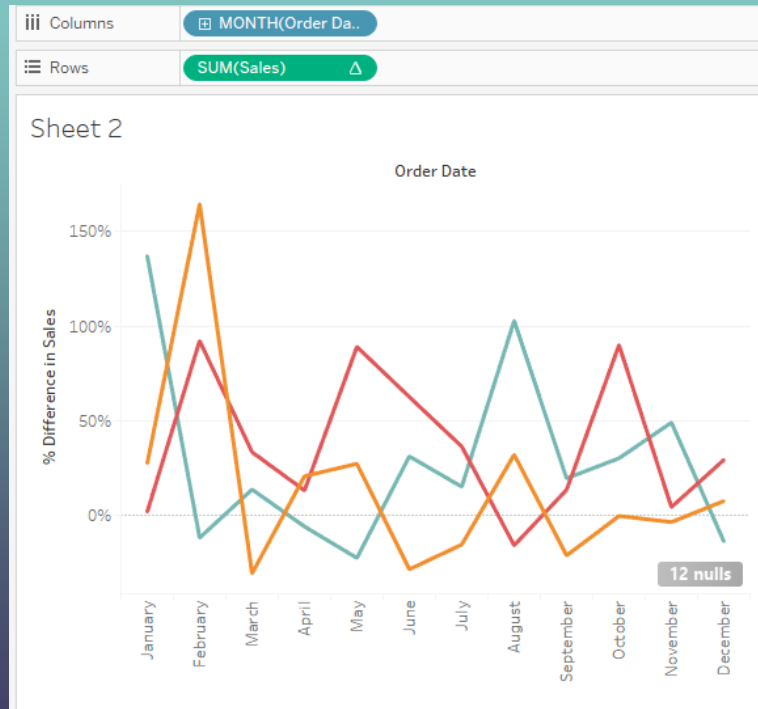


Contrast Sales Data by Year

Drag Order Date to Color



YoY Growth by Editing Rows





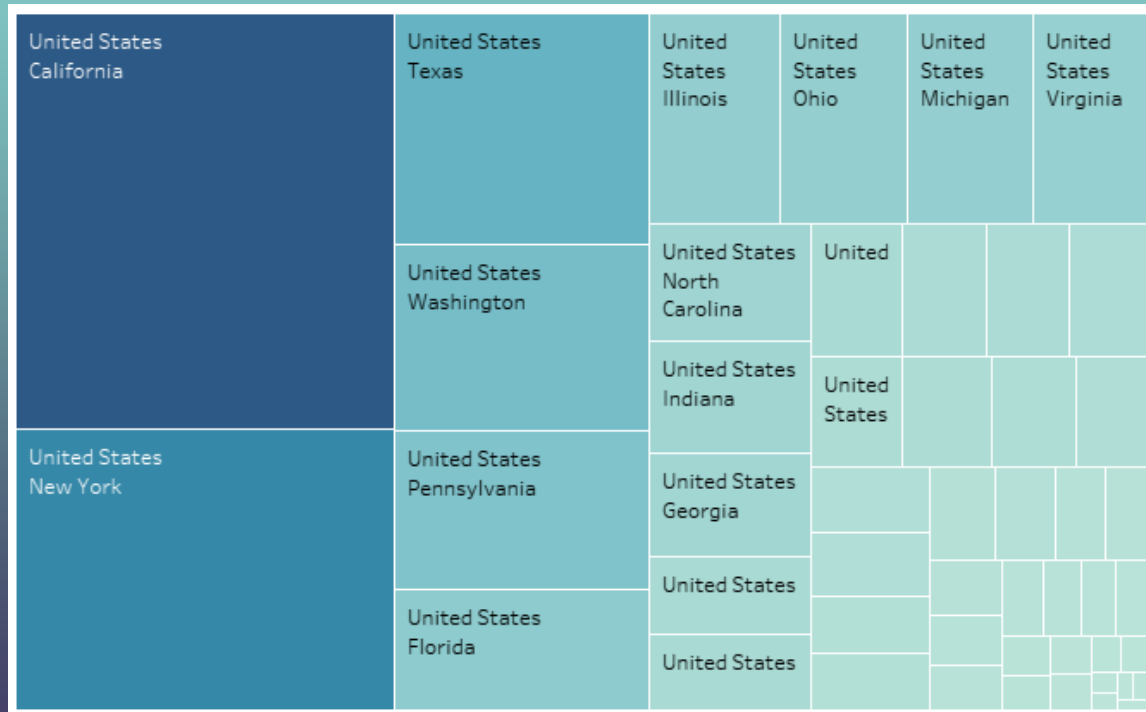
Map Representation

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

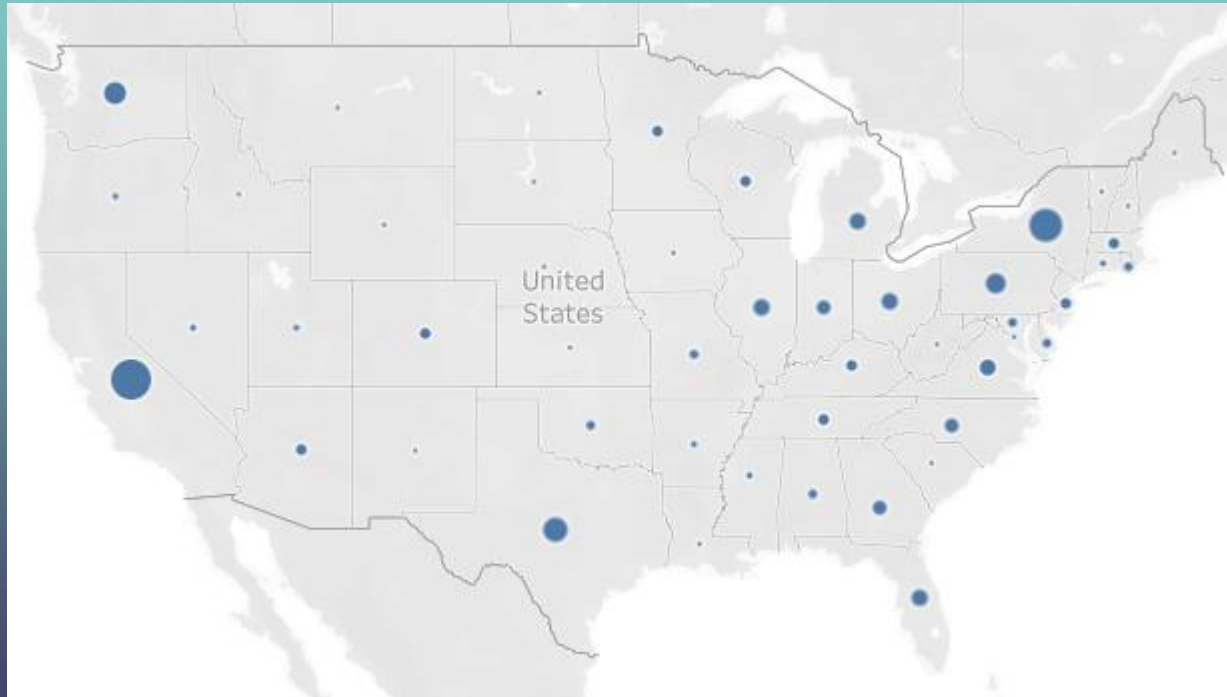
Try different functions in show me panel on the right-hand side.



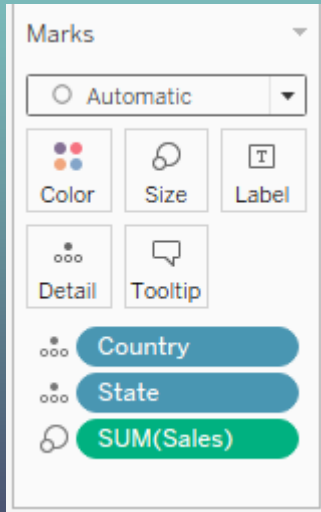
Map Representation



Map Representation



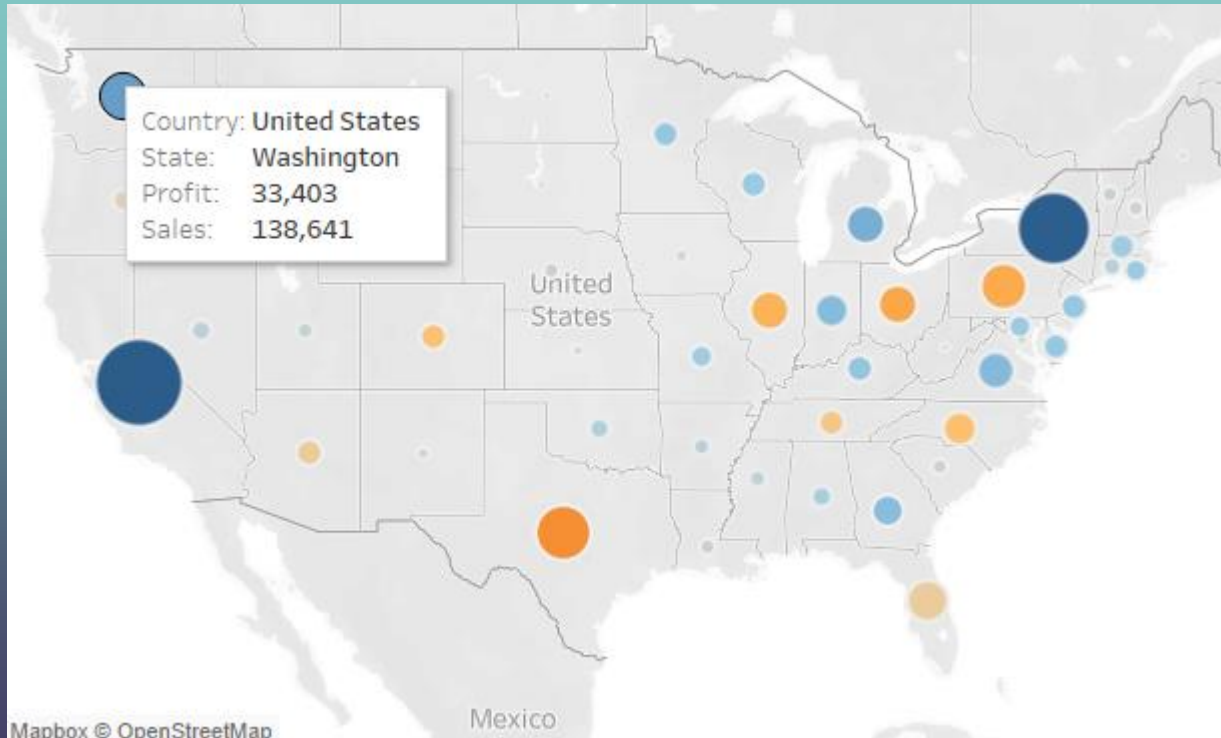
Map Representation

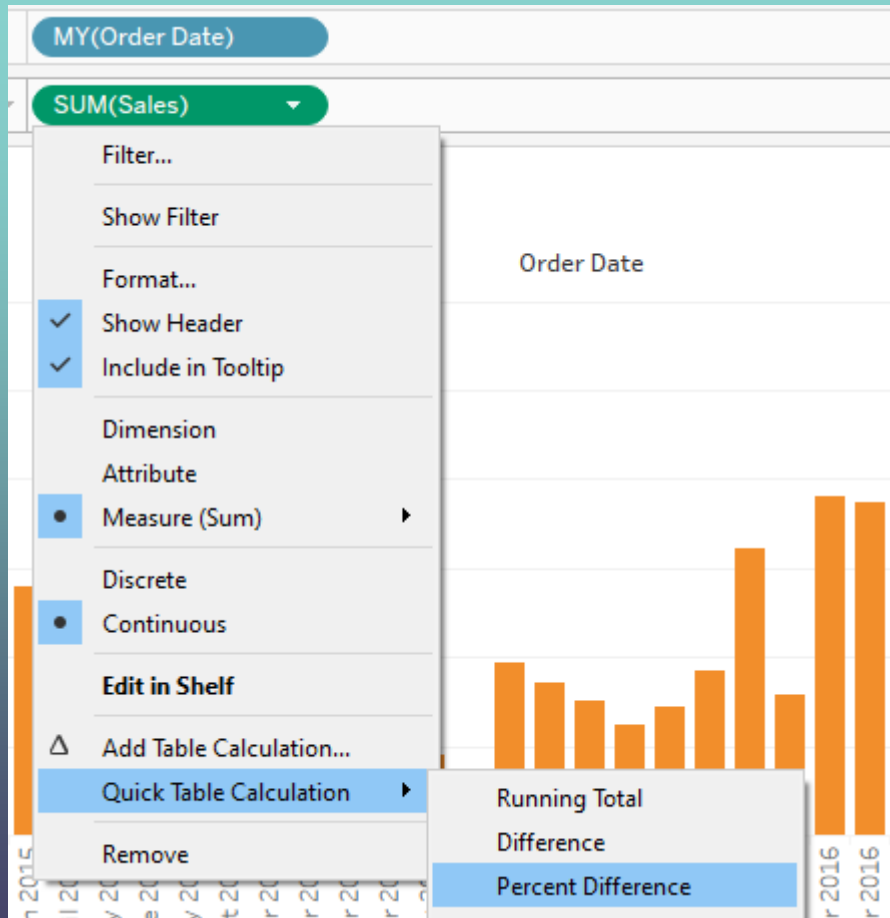


We can use profit to select the color of the circles in the map.

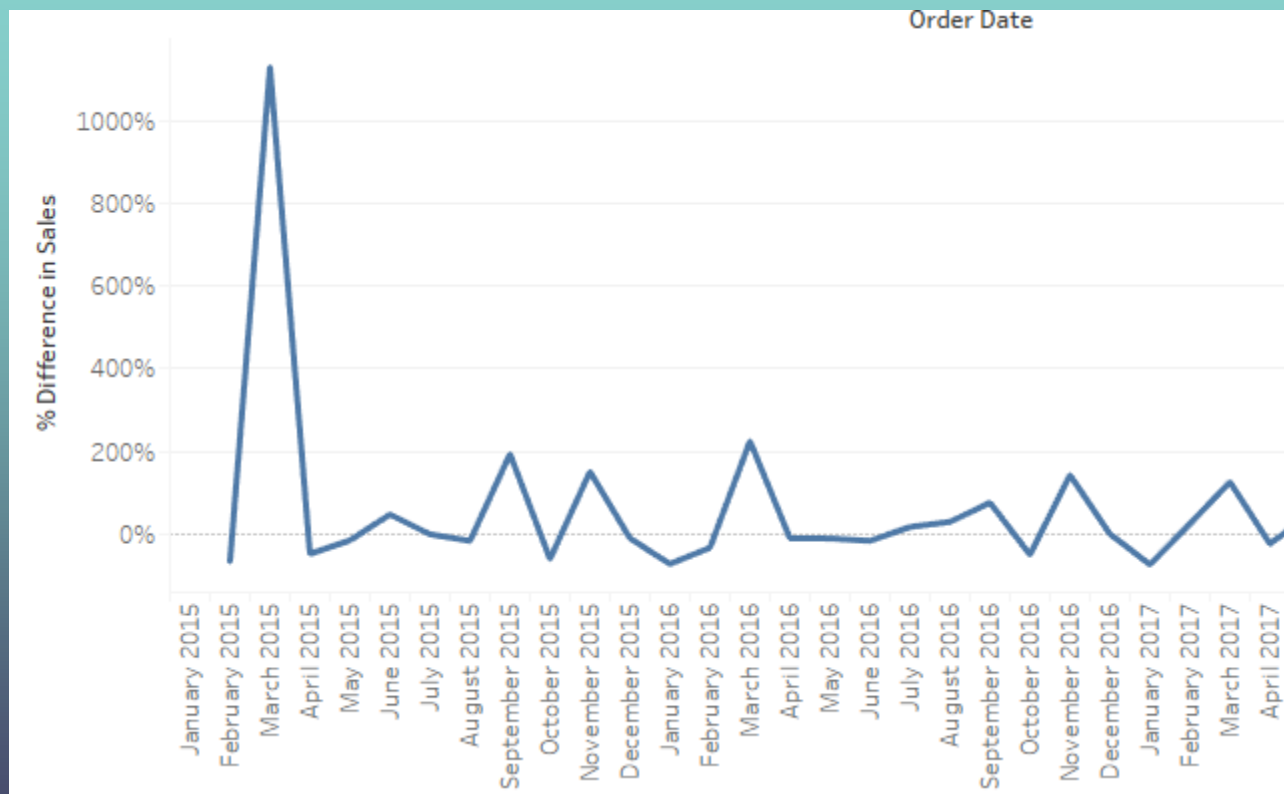
Moreover, we adjust the size of the circles.

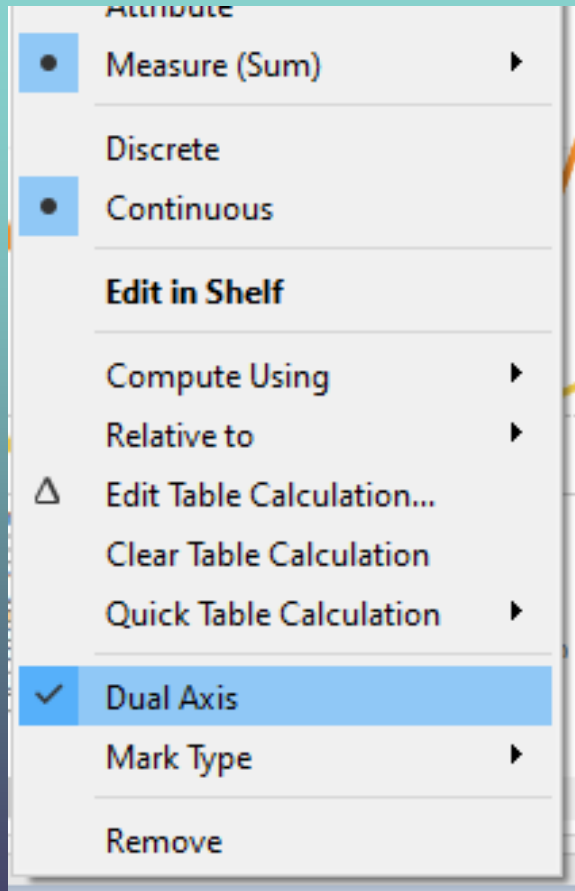
Map Representation





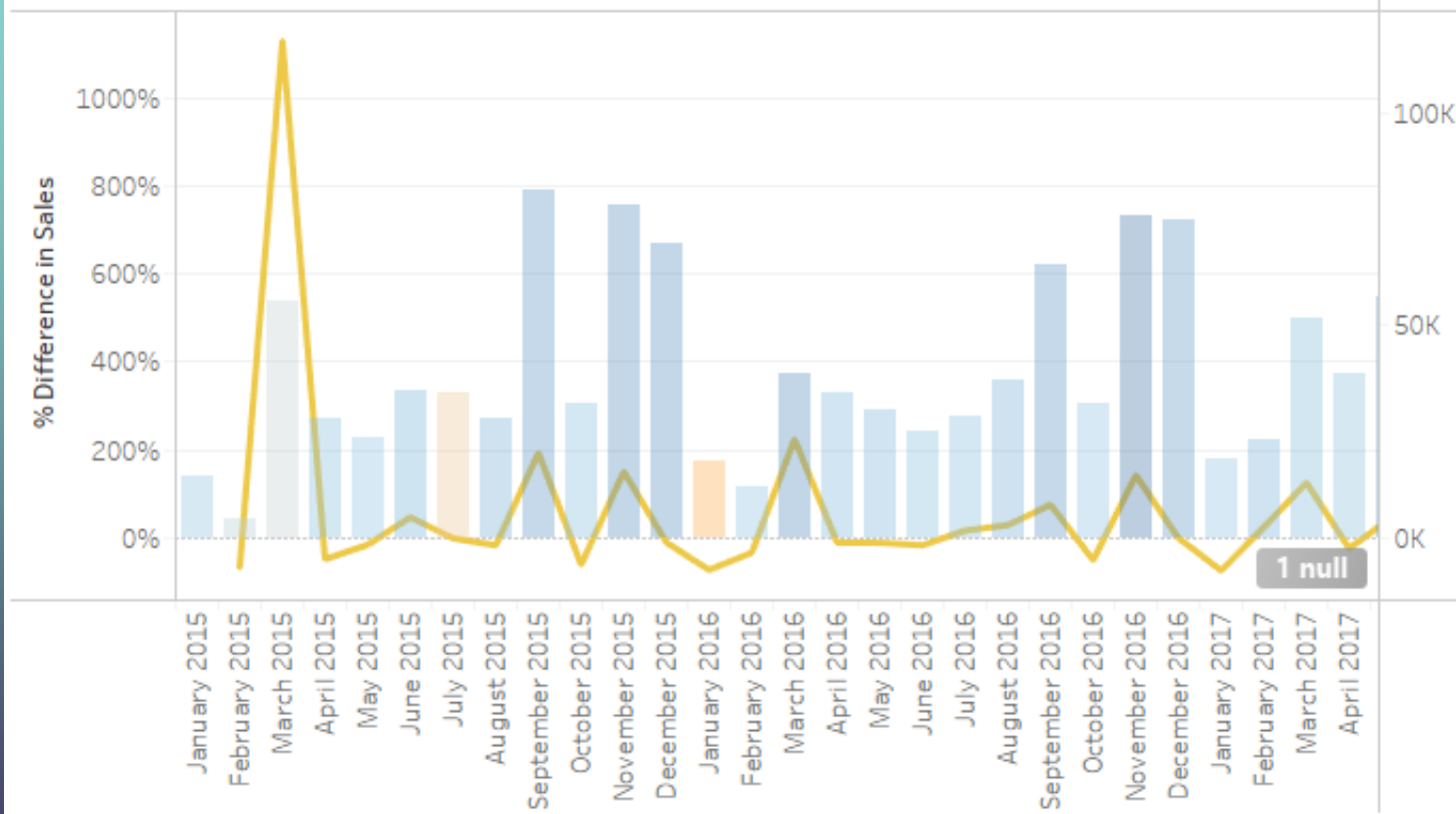
Here, we want to calculate how the sales change over time.





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

Order Date



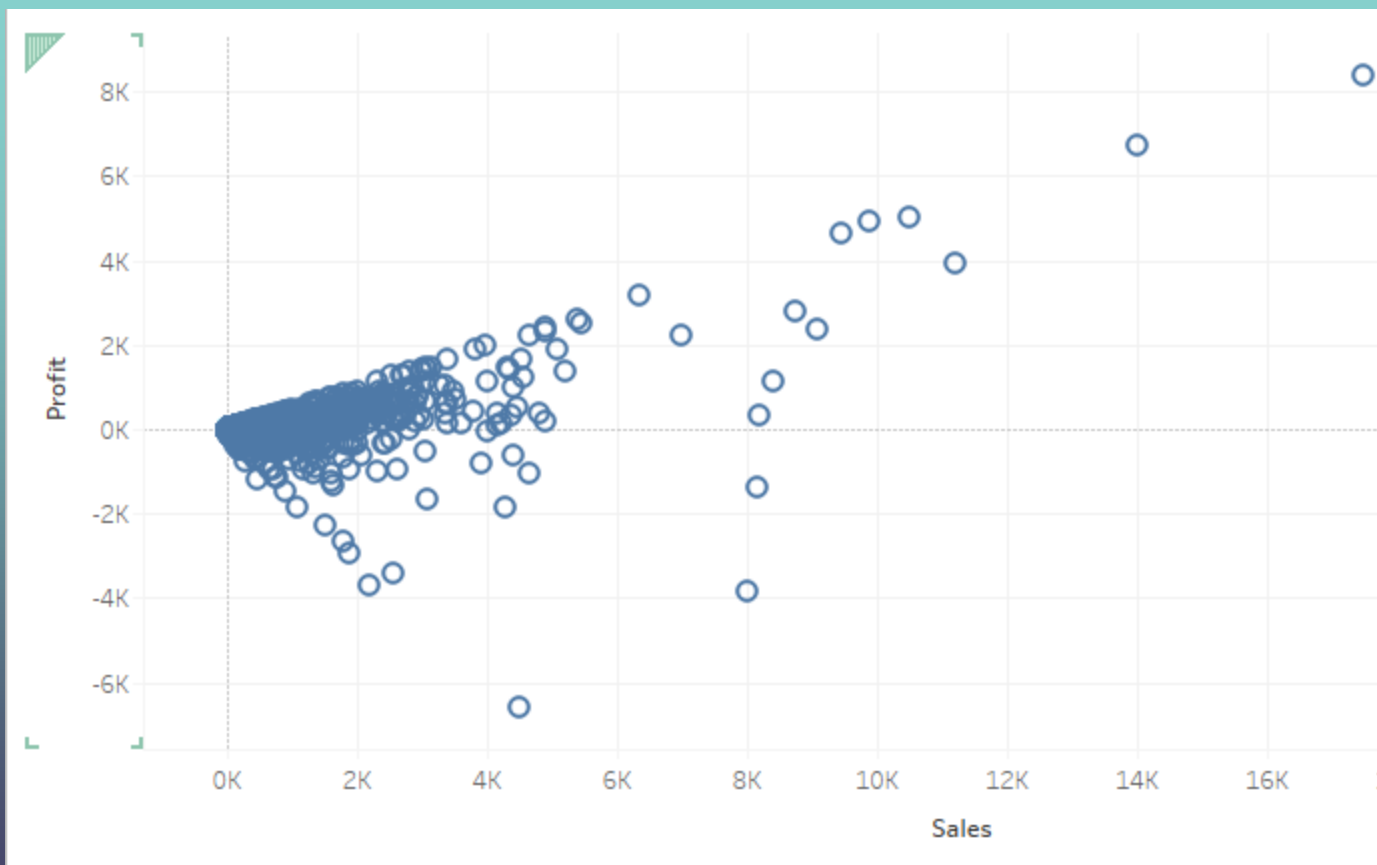
Regression Analysis

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.



- Constant Line
- Average Line
- Median with Quartiles
- Box Plot
- Totals

Filters

Marks

Automatic

Color Size Label

Detail Tooltip Shape

Model

- Average with 95% CI
- Median with 95% CI
- Trend Line
- Forecast
- Cluster

Custom

Sheet 1

Add a Trend Line



Linear



Logarithmic



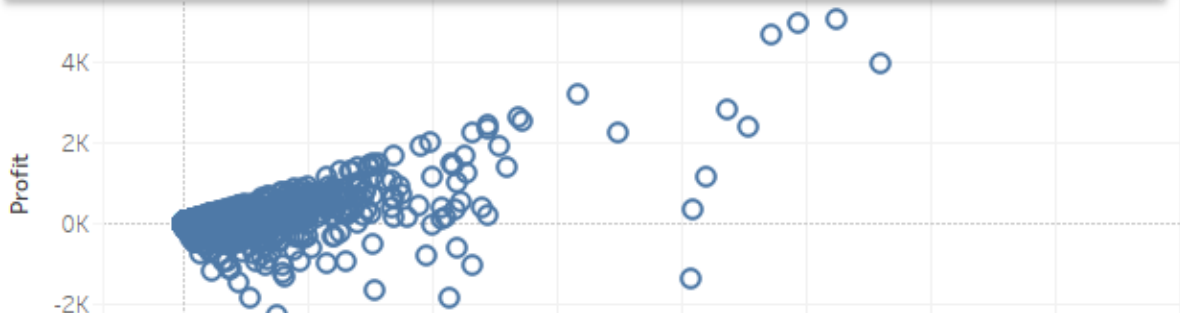
Exponential

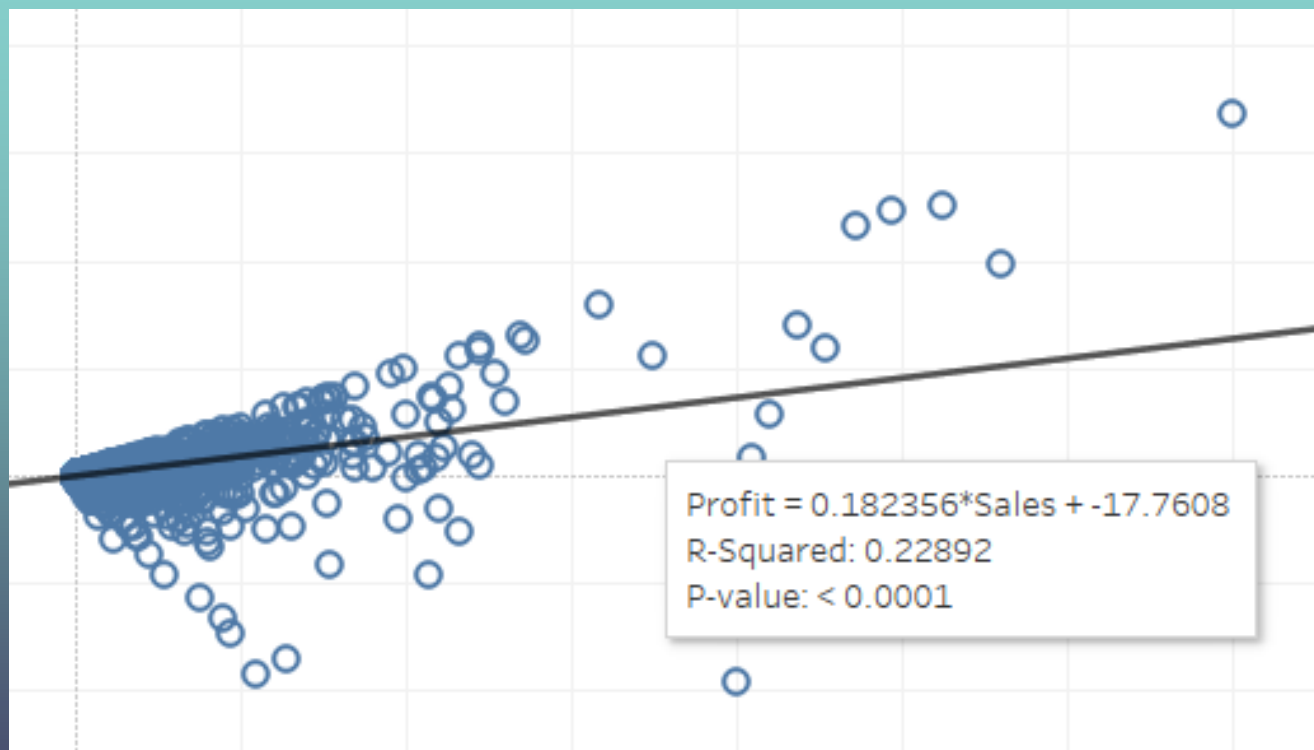


Polynomial



Power







THE MOVIE DATASET





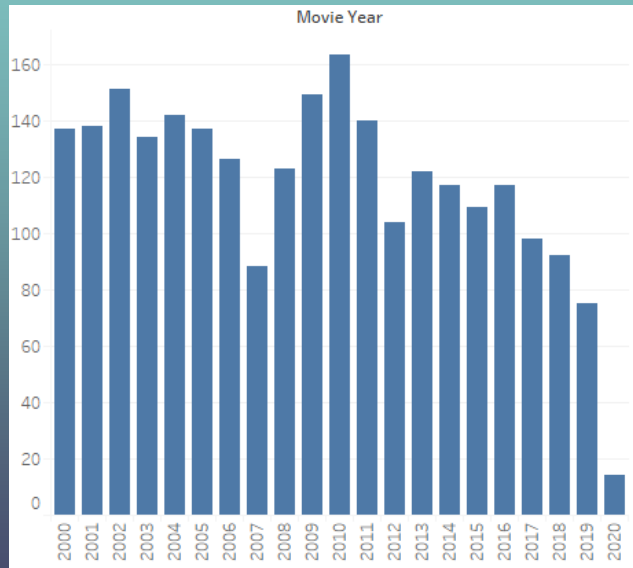
THE MOVIE DATASET

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



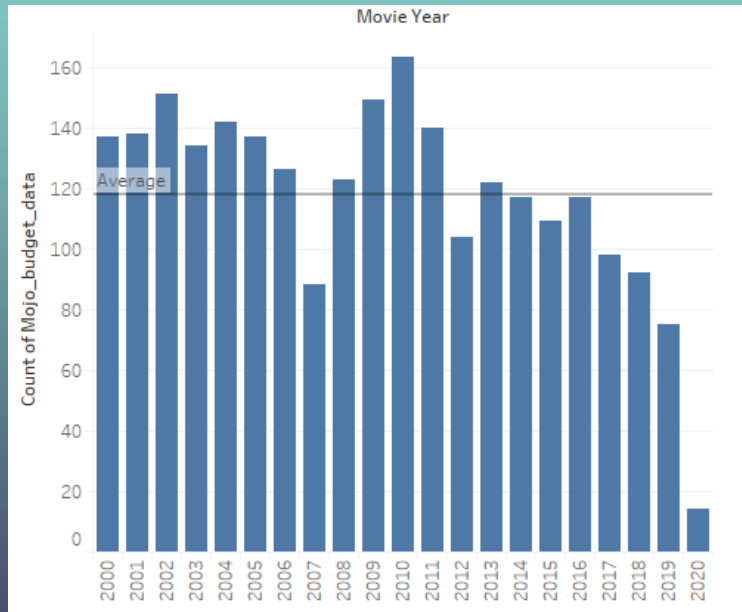
Columns: **Movie Year**

Set "Movie Year" as discrete

Rows: **CNT(Mojo_budget_data)**

Set "Marks" to be "Bar"

Number of Movies by Year

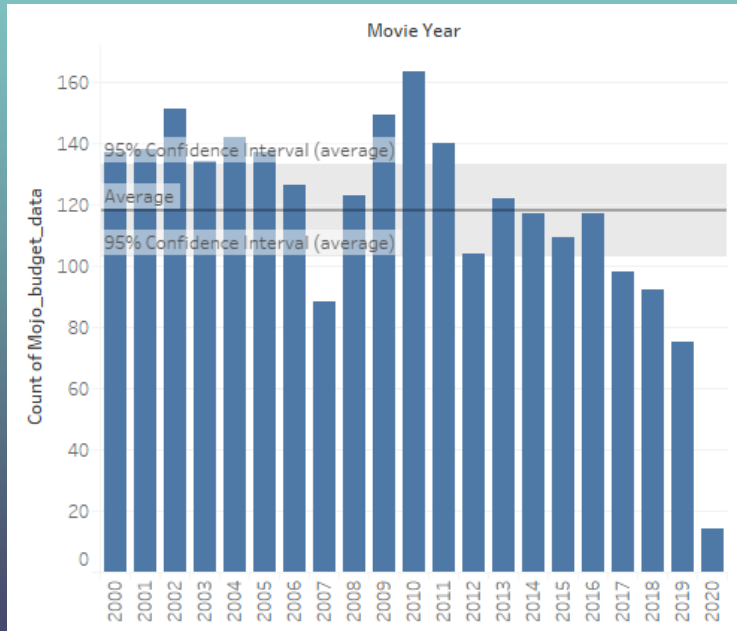


Go to **Analytics** Menu.

Drag "**Reference Line**" to the Plot.

Similarly, you can add median to your plot.

Number of Movies by Year

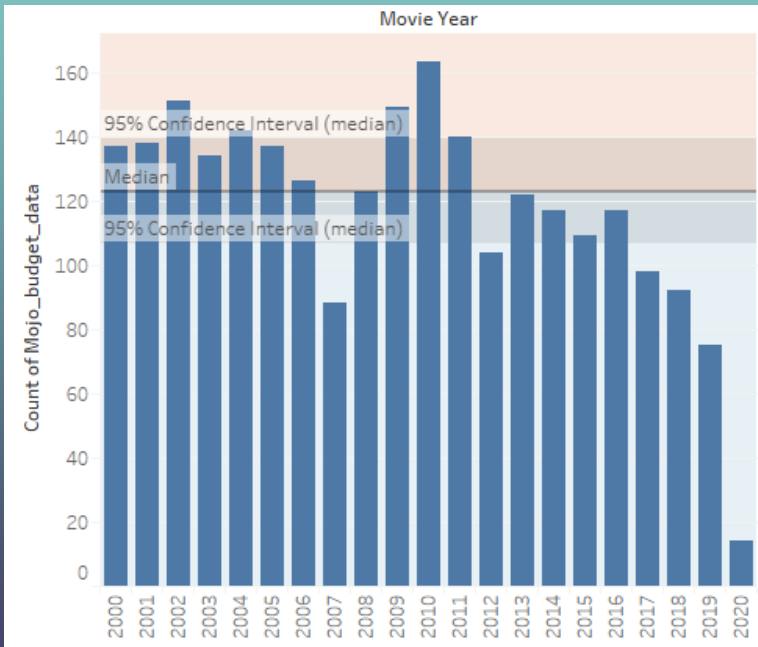


Now we plot the confidence interval of the mean.

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

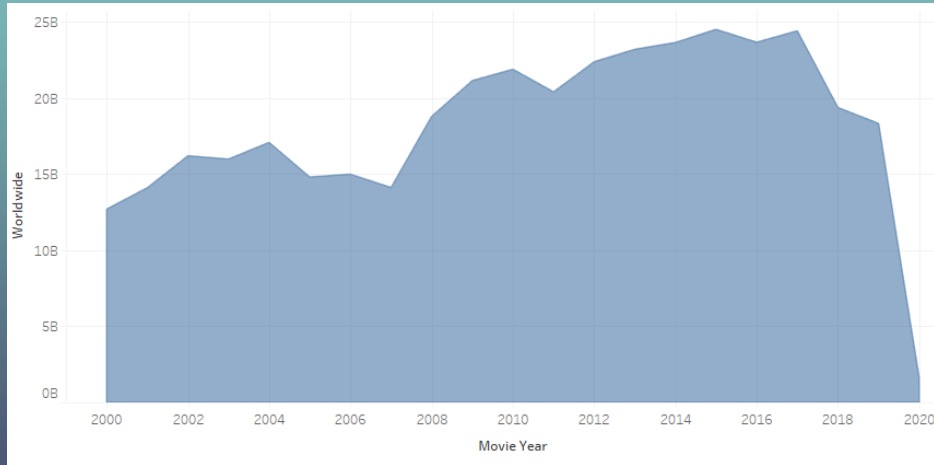


Now we add colors to the plot.

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



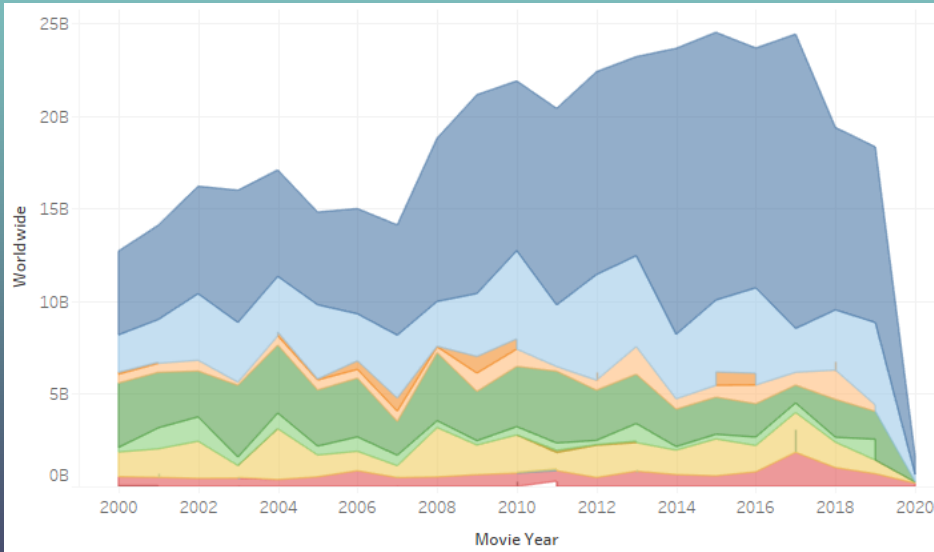
Columns: Movie Year

Rows: Worldwide (SUM)

Show me:



Area Chart

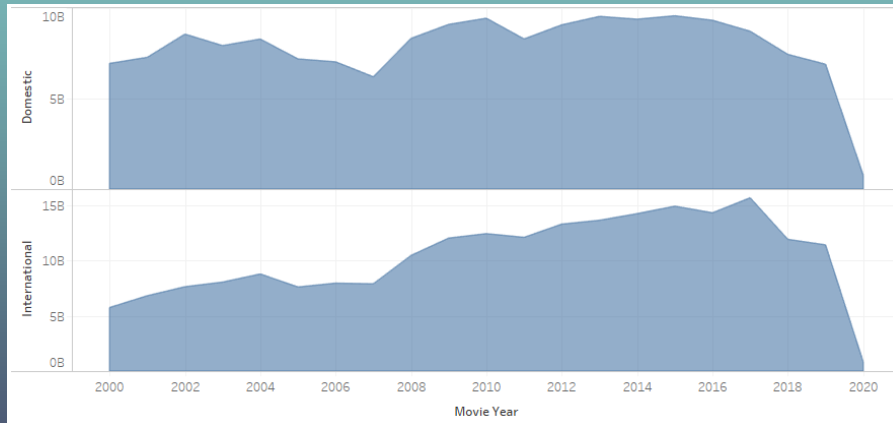


Columns: **Movie Year**

Rows: **Worldwide (SUM)**

Drag "**Genre 1**" to Color under **Marks** Menu.

Area Chart

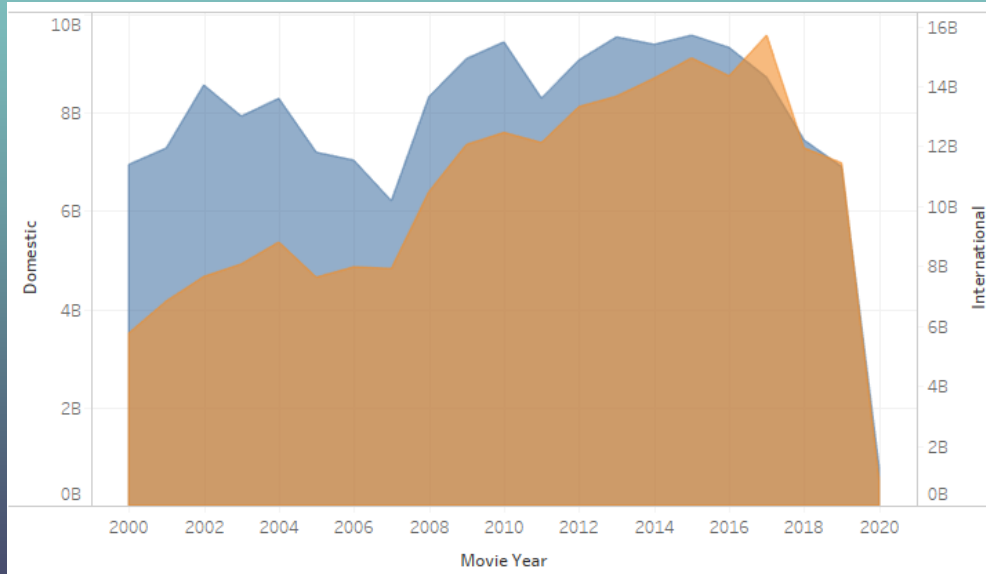


Columns: **Movie Year**

Rows: **Domestic (SUM) and International (SUM)**

Select **Area Chart**

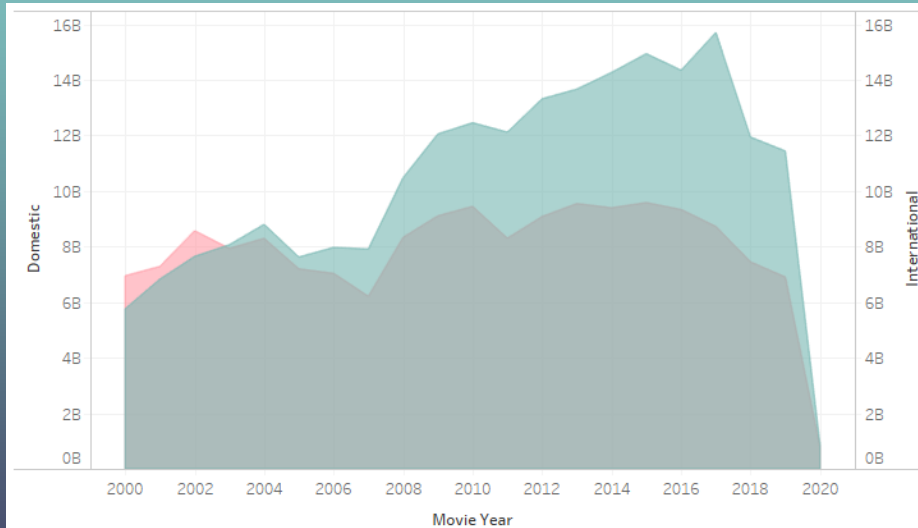
Area Chart



Right Click International
(Second Chart)

Choose **"Dual Axis"**

Area Chart

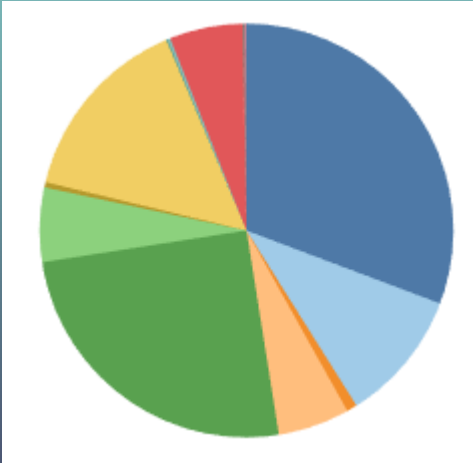


Right Click Domestic

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

Genre 1	Movie Year								
	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	

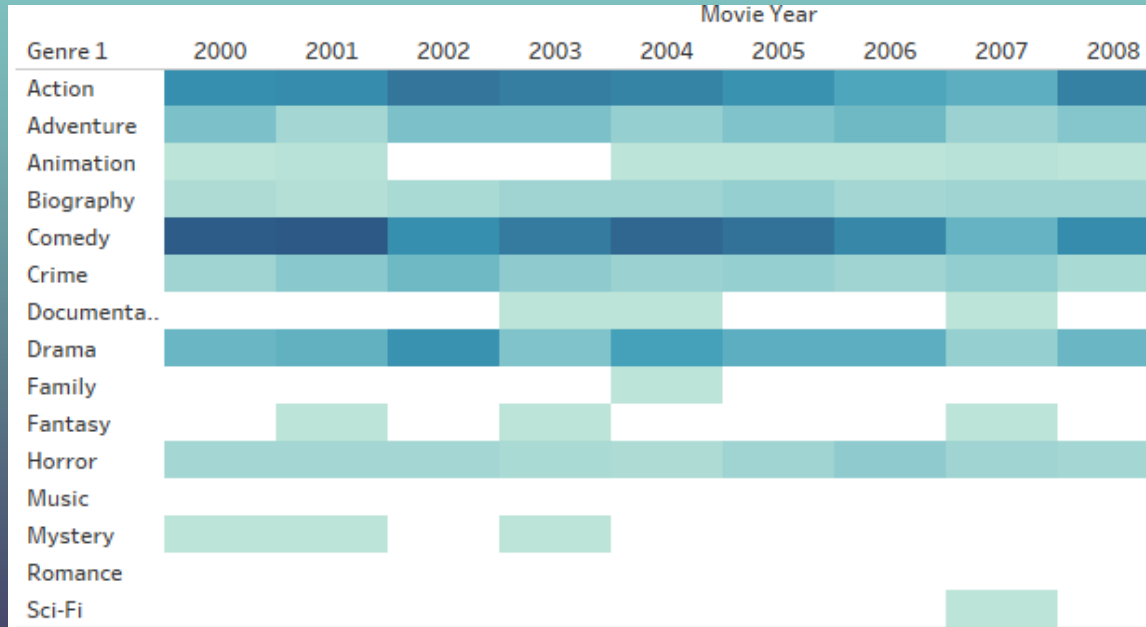
Columns: **Movie Year (Discrete)**

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

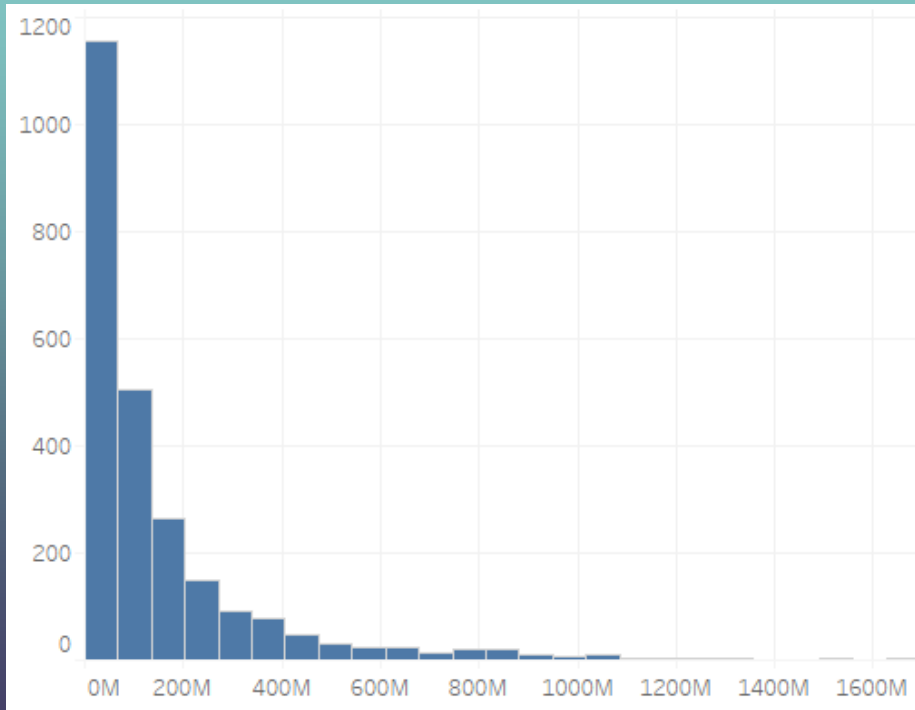
Genre 1	Movie Year									
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Action	●	●	●	●	●	●	●	●	●	●
Adventure	●	●	●	●	●	●	●	●	●	●
Animation	●	●		●	●	●	●	●	●	●
Biography	●	●	●	●	●	●	●	●	●	●
Comedy	●	●	●	●	●	●	●	●	●	●
Crime	●	●	●	●	●	●	●	●	●	●
Documenta..				●	●			●		
Drama	●	●	●	●	●	●	●	●	●	●
Family					●					
Fantasy		●		●				●		
Horror	●	●	●	●	●	●	●	●	●	●
Music										
Mystery	●	●		●						
Romance										
Sci-Fi								●		

Columns: **Movie Year**

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

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

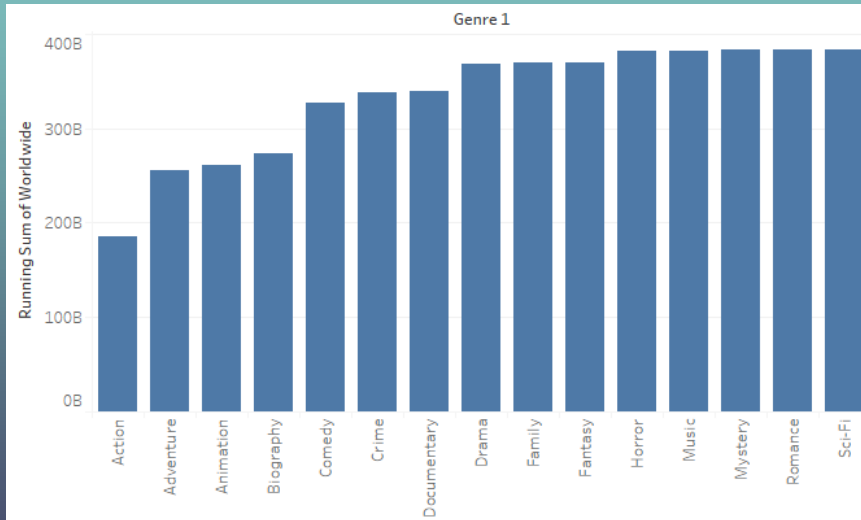
Worldwide Box Office Histogram



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

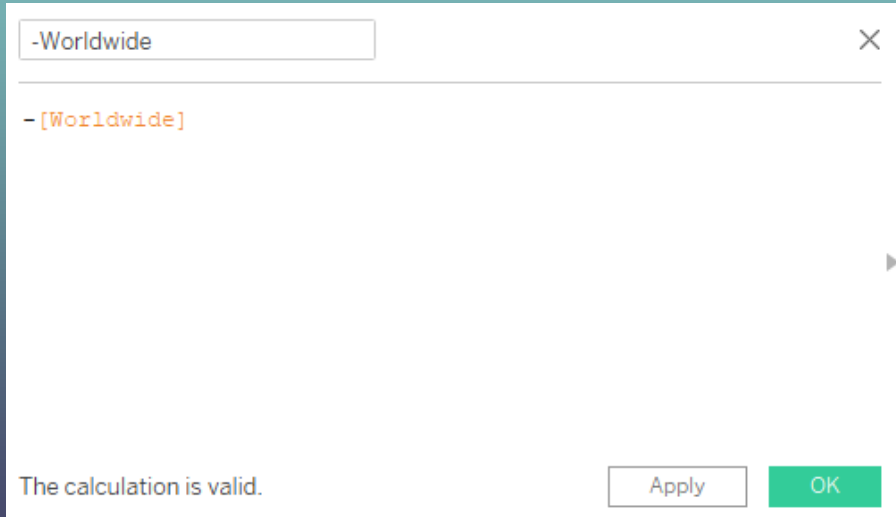
The screenshot shows the Tableau Desktop interface. The 'Data' menu is open, and 'Create Calculated Field...' is highlighted. The background shows a worksheet named 'Sheet 4' with a vertical axis labeled 'of Mojo_budget_data' and a scale from 1400 to 2600. The interface includes tabs for 'Data', 'Analytics', 'Pages', 'Columns', and 'Rows'. The 'Marks' card is set to 'Automatic', and the 'Detail' and 'Tooltip' cards are visible. The 'Data' menu options include: Search, Create Calculated Field..., Create Parameter..., Group by Folder, Group by Data Source Table (selected), Sort by Name (selected), Sort by Data Source Order, Hide All Unused Fields, Show Hidden Fields, Expand All, and Collapse All. Below the menu, a list of fields is visible: Main Actor 2, Main Actor 3, Main Actor 4, Movie Id, and Movie Title.

Under **Data** Menu, choose **Create Calculate Field**

Input the following calculation.

Click **OK** to proceed.

Waterfall



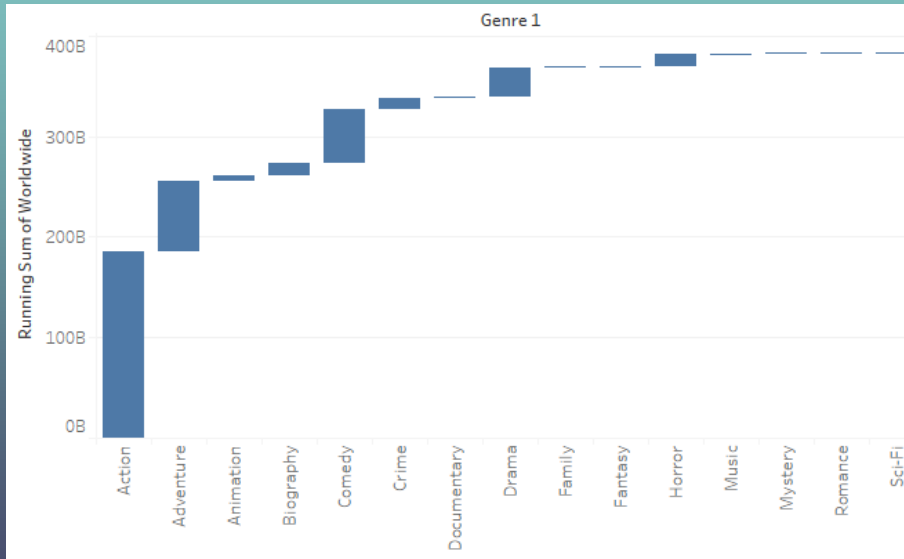
The screenshot shows a dialog box with a title bar containing a text field with the value "-Worldwide" and a close button (X). The main area of the dialog contains a text field with the value "-[Worldwide]". At the bottom left, there is a status message: "The calculation is valid." At the bottom right, there are two buttons: "Apply" and "OK".

Under **Analysis** Menu, choose **Create Calculate Field**

Input the following calculation.

Click **OK** to proceed.

Waterfall

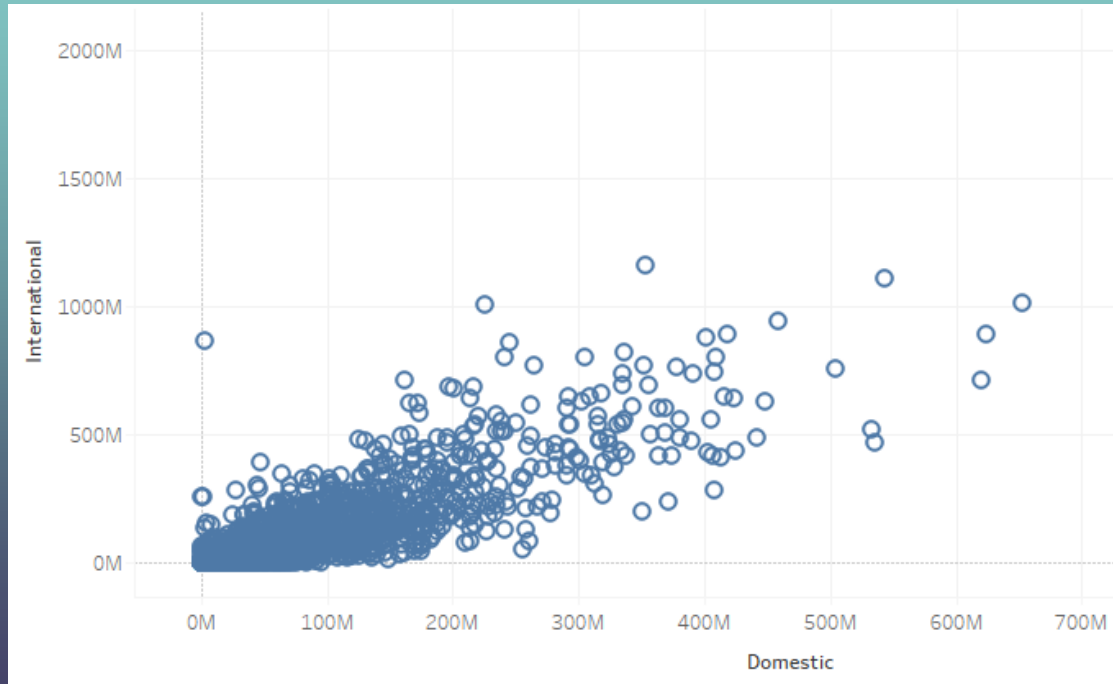


There is a new variable “-Worldwide”.

Under Marks, change **Automatic** to **Gantt Bar**.

Drag new variable “-Worldwide” to “Size” under **Marks**.

International vs. Domestic Box Revenue

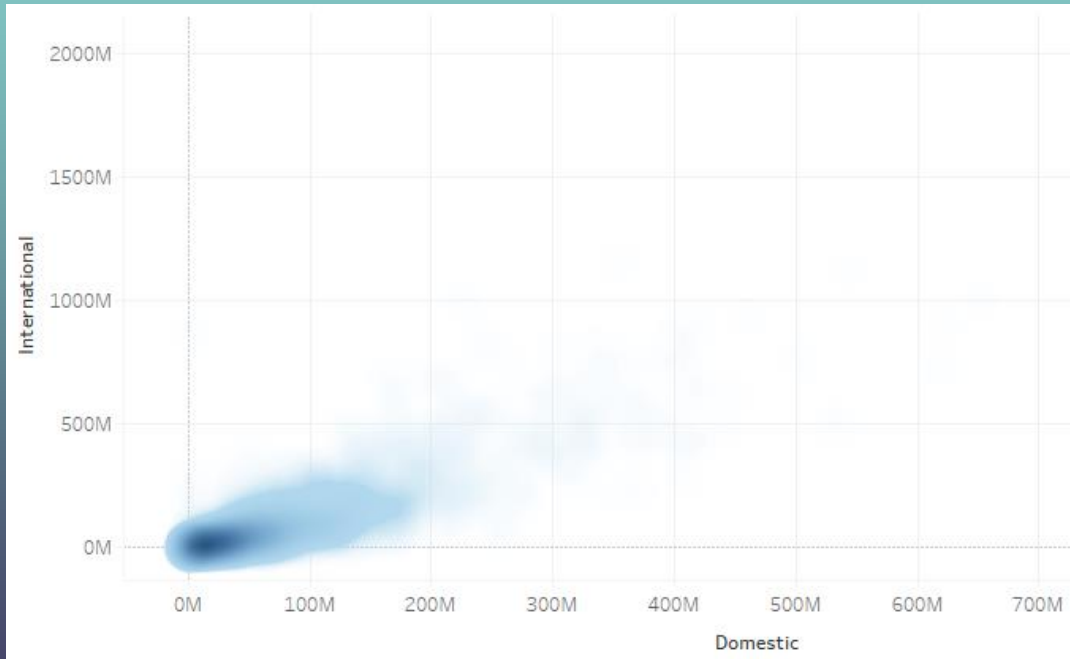


Columns: Domestic

Rows: International

Set both variables as
Dimension

International vs. Domestic Box Revenue



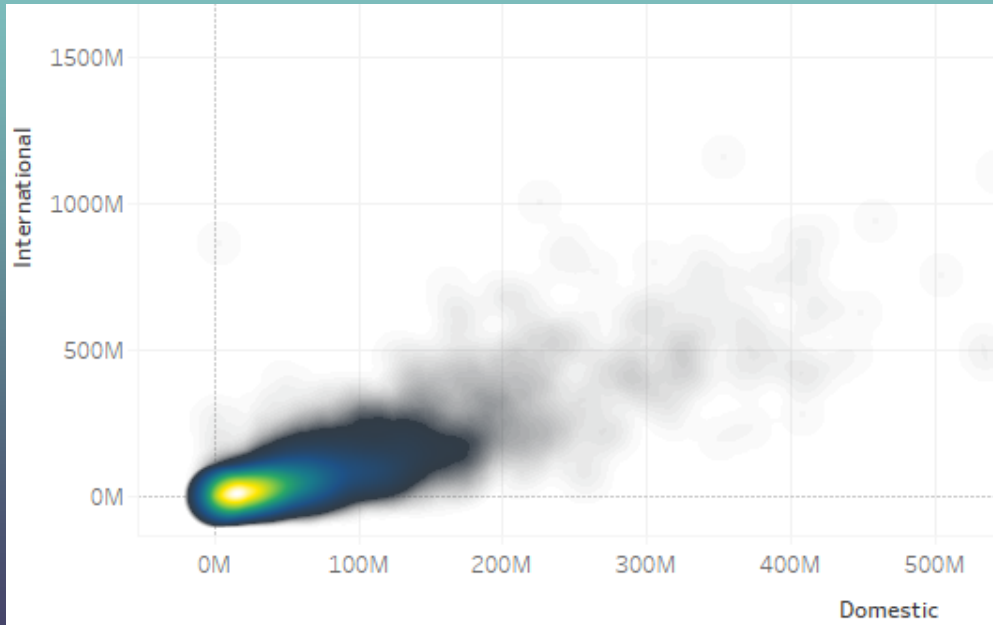
Columns: **Domestic**

Rows: **International**

Set both variables as
Dimension

Change “**Automatic**” to
“**Density**” under **Marks**

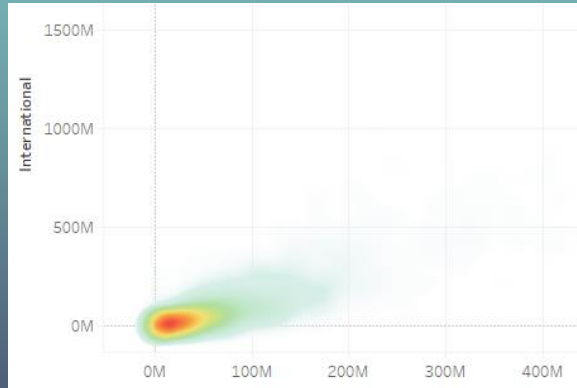
International vs. Domestic Box Revenue



Change “Automatic” to
“Density” under Marks

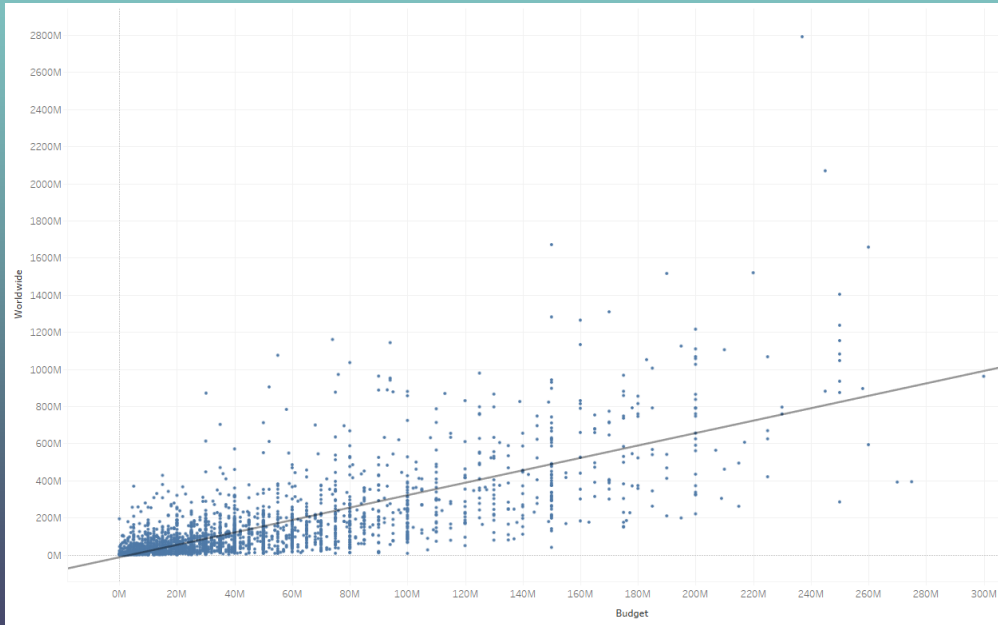
Set Colors to “Density-
Multicolor” under Marks

Density Heatmap



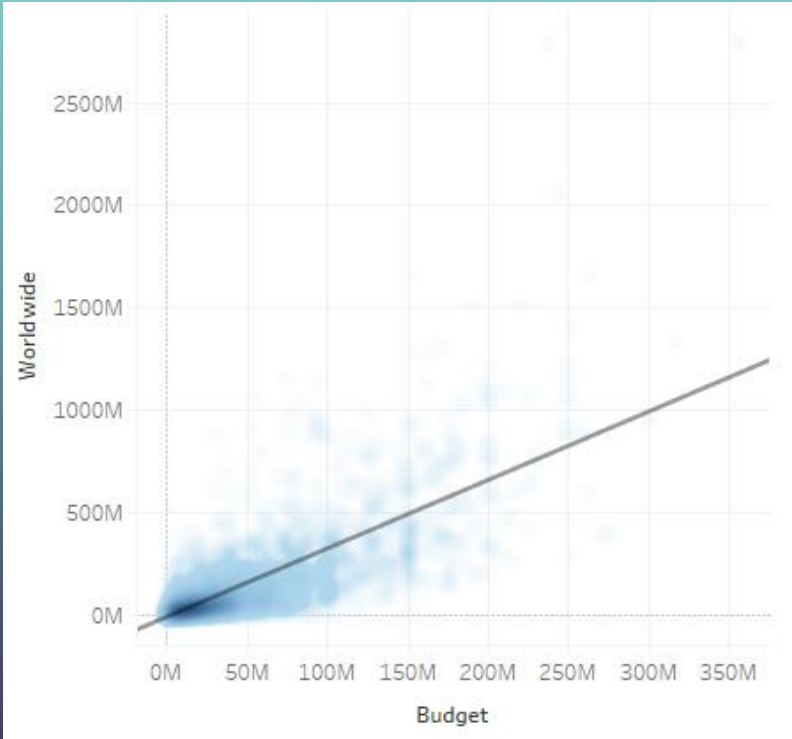
In a density heatmap, in places where the observations are dense (i.e., many points in the area), the heatmap displays a warm color. In places where observations are sparse, the heatmap displays a cold color.

Regression Line



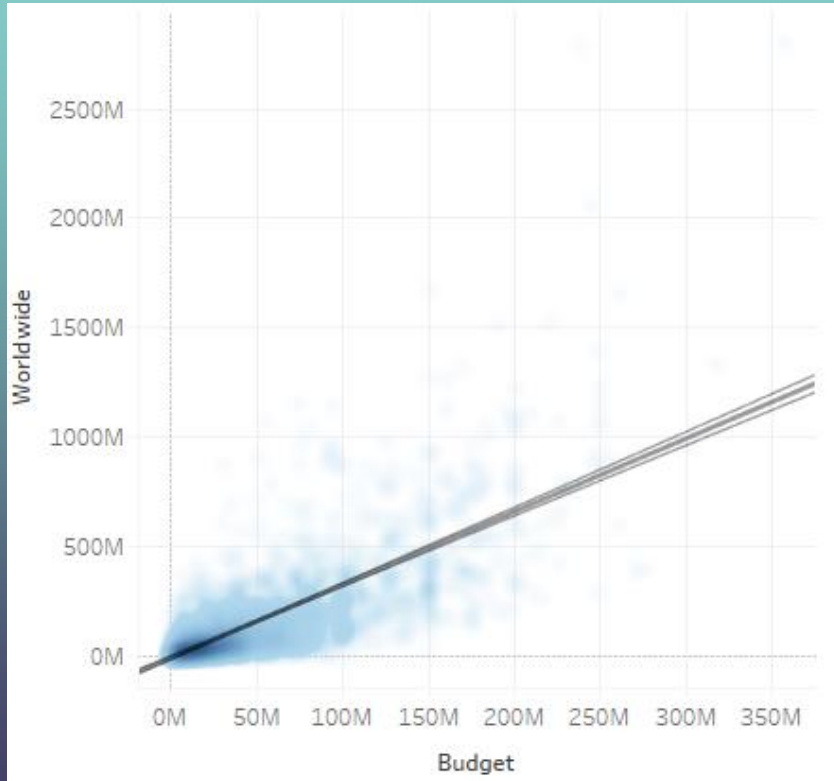
Here, we use budget as the independent variable (i.e., Column) and worldwide box office as the dependent variable (i.e., Row) and draw the regression line.

Regression Line



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

Regression Line



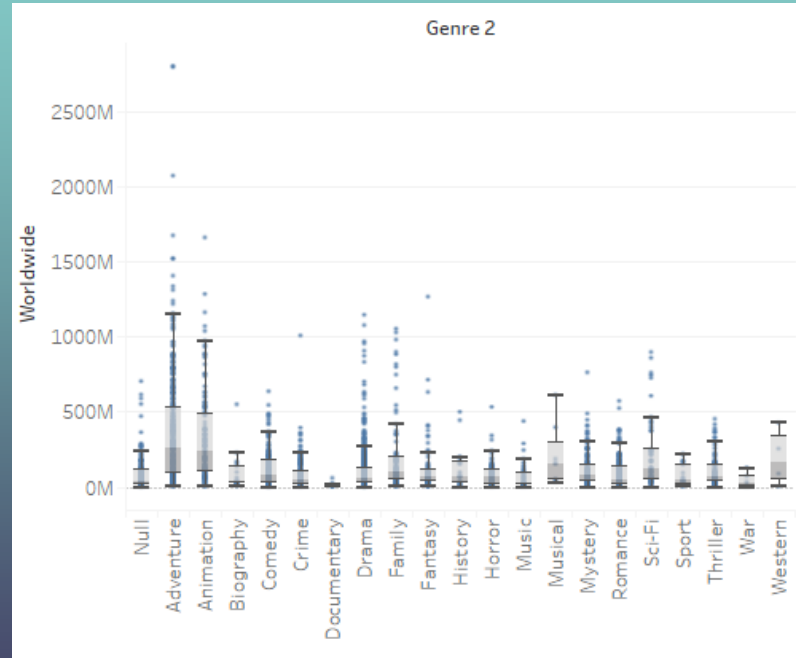
Click and edit the regression line.

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

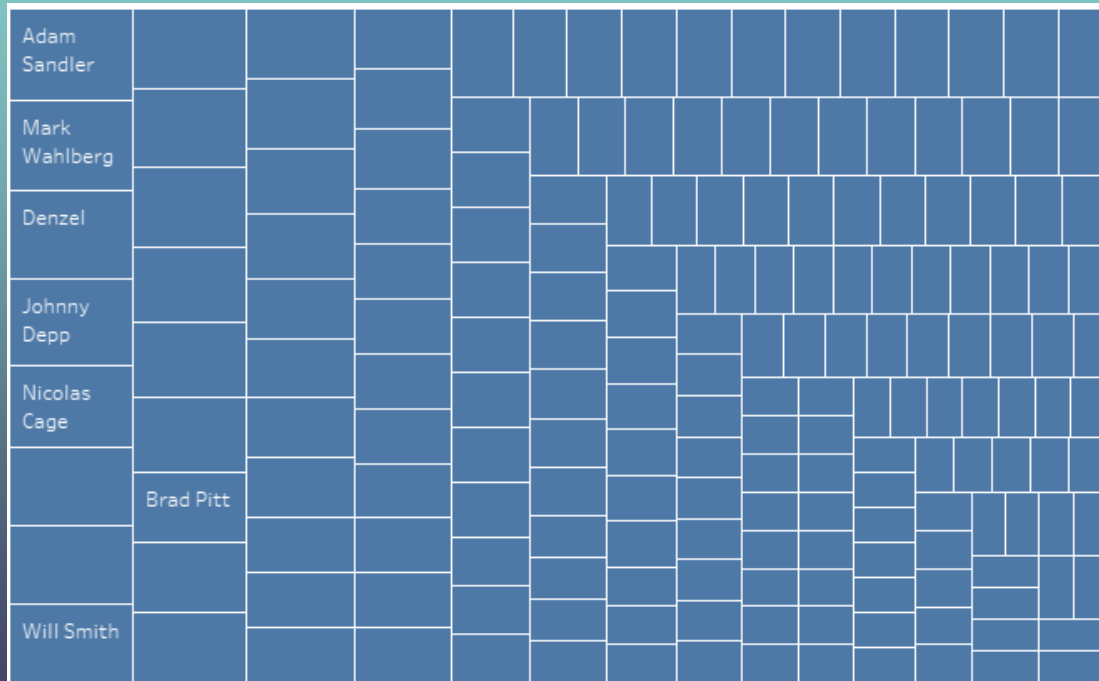
How to draw the Box and Whisker Plot?

Use **Genre 2** as rows and **Worldwide** as columns.

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

Under Show me, choose 

Actor/Actress Treemap



Check Aggregate Measures again!

Columns: **Main Actor 1**

Drag “**Main Actor 1**” to from **Columns** to **Text**

Drag “**Mojo_budget_data (Count)**” to **Size**

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

Shia LaBeouf Steve Carell
Harrison Ford Ben Affleck Ben Stiller Angelina Jolie
Tom Hanks Brad Pitt Will Ferrell
Matt Damon Will Smith Daniel Radcliffe Vin Diesel
Tom Cruise Hugh Jackman George Clooney Leonardo DiCaprio
Dwayne Johnson Robert Downey Jr. Bruce Willis Johnny Depp
Keanu Reeves Matthew McConaughey Denzel Washington Daniel Craig
Adam Sandler Nicolas Cage Jennifer Lawrence Mark Wahlberg
Jack Black Christian Bale Chris Hemsworth Russell Crowe Nicole Kidman Chris Pratt
Jim Carrey Ryan Reynolds Eddie Murphy
Mike Myers

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)

