## Introduction to Tableau

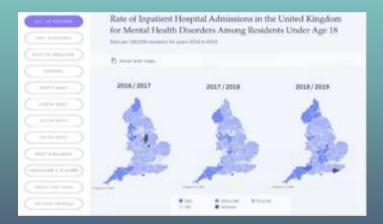
Your new data analysis software

#### **Data Visualization with Tableau**

# 



## **Tableau Public Gallery**





#### Import Data

Download the sample datafile "SuperStore\_Data" and open it with your Tableau.

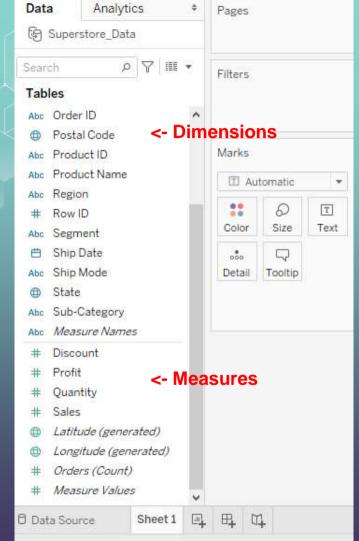
You can choose different sheets from the excel file, e.g., "Orders".

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



## Check Data Type

🖽 🔳 So	rt fi	elds Data sourc	e order 🔹	1				Show	
# Orders Row ID =		Abc Orders Order ID	Corders Order Date	Ø	Number (decimal)		ode	Abo Orders Customer ID	
	4	CA-2017-152156	11/8/2017		Number (who Date & Time	le)	l Class	CG-12520	
1	2	CA-2017-152156	11/8/2017	•	Date		Class	CG-12520	
	3	CA-2017-138688	6/12/2017		String Boolean		Class	DV-13045	
2	4	US-2016-108966	10/11/2016					ird Class	SO-20335
ļ	5	US-2016-108966	10/11/2016	10/1	Default 6/2016	stan	oard Class	SO-20335	
(	6	CA-2015-115812	6/9/2015	6/14	/2015	Stan	dard Class	BH-11710	
	7	CA-2015-115812	6/9/2015	6/14	/2015	Stan	dard Class	BH-11710	



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).

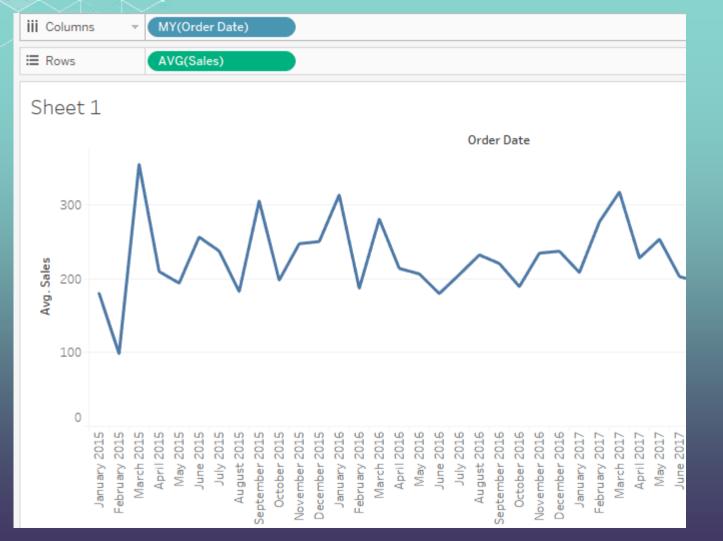


#### Here, Sales is a measure.

#### Order date is a dimension.

We partition sales based on year of order.

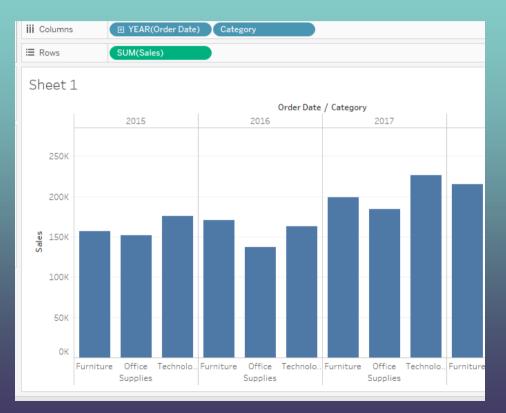
You can also change the properties of your rows and columns.



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



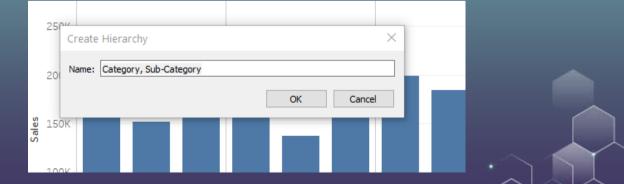
#### **Adding Category As Another Classifier**



#### **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".



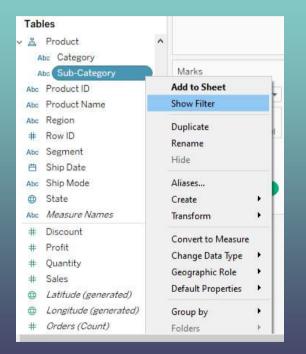
#### **Create Hierarchy**

#### Then, you can display each product subcategories.

iii Columns	▼ ( I YEAR(Order Date) 📄 Category - Sub-Category
E Rows	SUM(Sales)
Sheet 1	



## **Apply Filters**



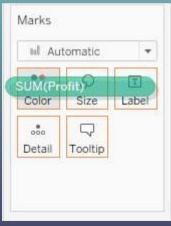
	(AII)
	Accessories
	Appliances
	Art
	Binders
1	Bookcases
	Chairs
	Copiers
	Envelopes
	Fasteners
	Furnishings
	Labels
1	Machines
	Paper
	Phones
	Storage
	Supplies
1	Tables

c

#### **Color Your Output**

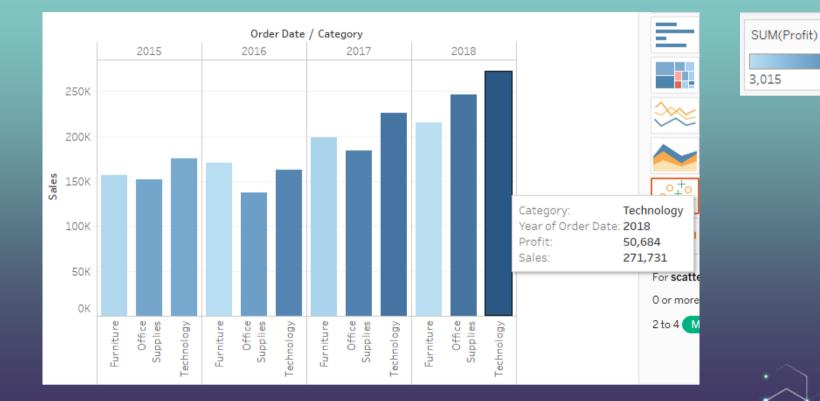
Suppose that you want to color your figure based on the profit of your products.

Drag profit to color.



#### **Color Your Output**

50,684



## **Apply Filters**

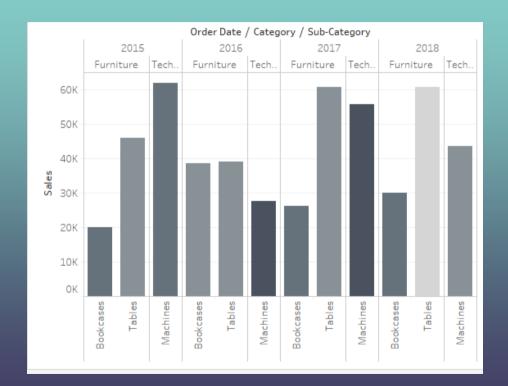


c

#### **Edit Colors**

		2010		2010				
Marks		Furniture	Tech	Furniture	Tech	Furniture	Tech	Furnitu
🗤 Automatic 👻	60 E	Edit Colors [Pi	rofit]					×
Color Size Label	50	Palette: Automatic						~
Color	40							
Opacity	30	-8,141					2,977	
100%		Stepped Cold	or 5	Steps				
Effects	20	Reversed						
Border: Automatic 🗸		Use Full Colo	_				Advance	4.5.5
Halo: V	10	Include Tota	S				Auvance	20 >>
	0	Reset		C	Ж	Cancel	App	ply
		w w	ų	v v	ų	ų ų	ų	

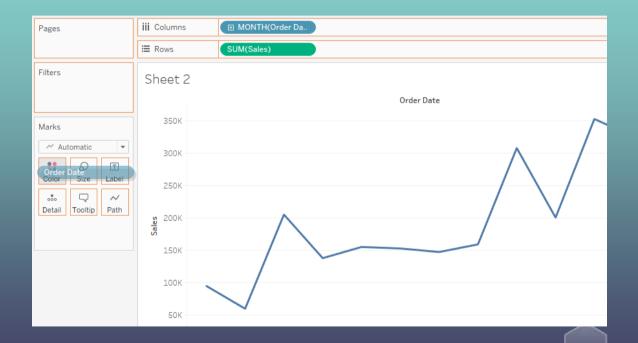
#### **Edit Colors**



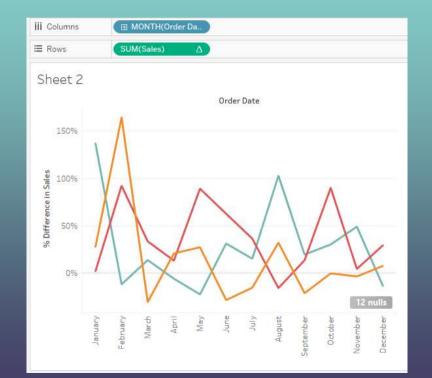
c

#### **Contrast Sales Data by Year**

#### Drag Order Date to Color



## YoY Growth by Editing Rows





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

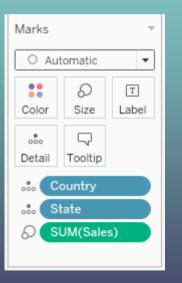
Try different functions in show me panel on the right-hand side.



United States California	United States Texas	United States Illinois	United States Ohio	United States Michigan	United States Virginia	
	United States Washington	United State North Carolina	s United			
		United State Indiana	United States			
United States New York	United States Pennsylvania	United State Georgia	:5			
	United States	United State	s			
	Florida	United State	5			



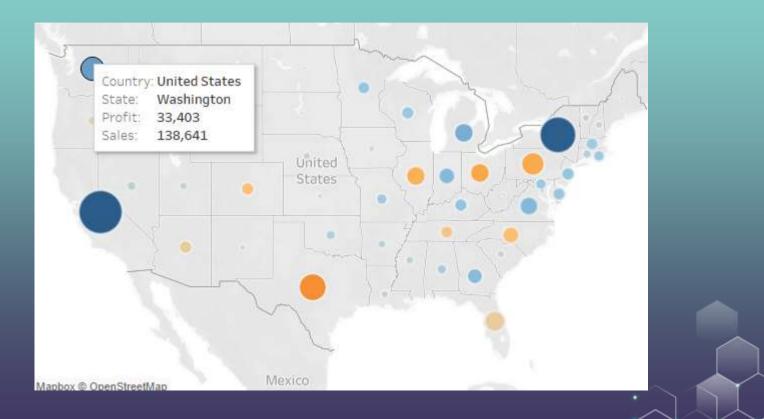
 $\sim$ 

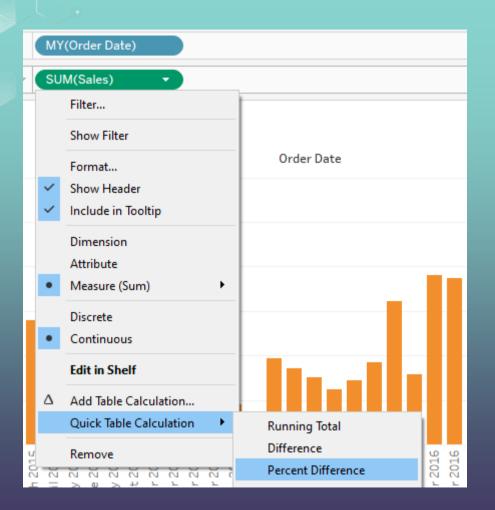


We can use profit to select the color of the circles in the map.

Moreover, we adjust the size of the circles.

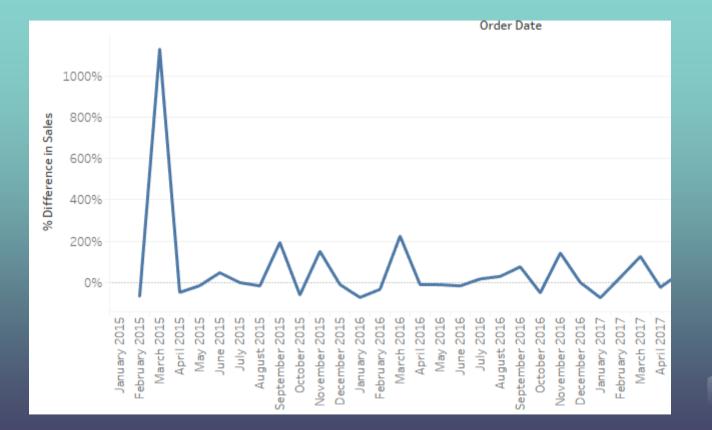






Here, we want to calculate how the sales change over time.



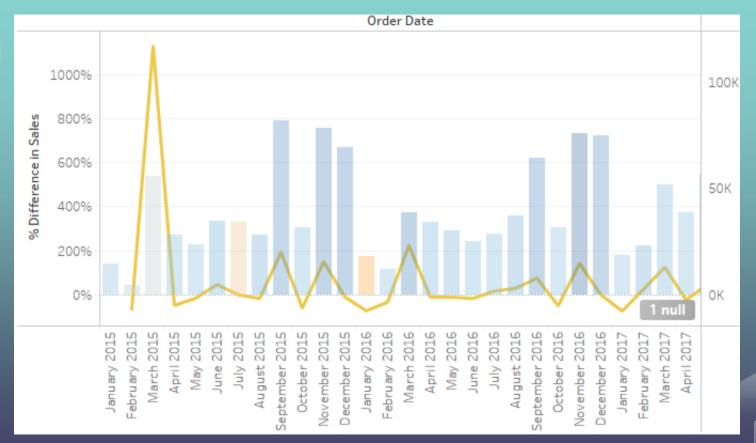


c

	Attribute	
•	Measure (Sum)	•
	Discrete	
•	Continuous	
	Edit in Shelf	
	Compute Using	•
	Relative to	•
Δ	Edit Table Calculation	
	Clear Table Calculation	
	Quick Table Calculation	•
/	Dual Axis	
	Mark Type	•
	Remove	

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





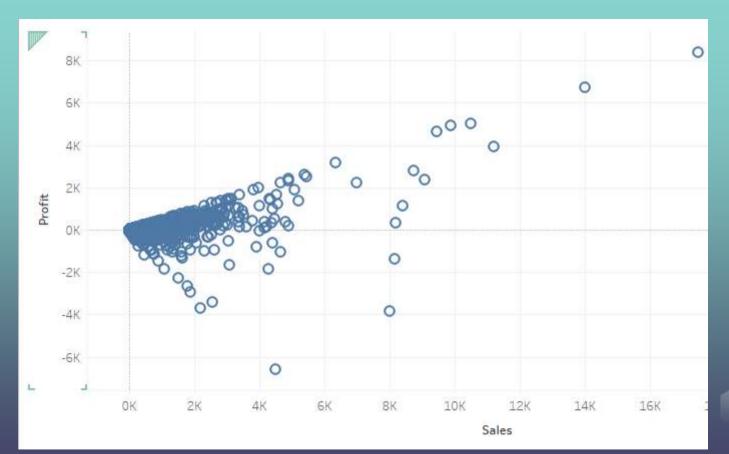
#### **Regression Analysis**

#### Suppose that we want to run the following regression:

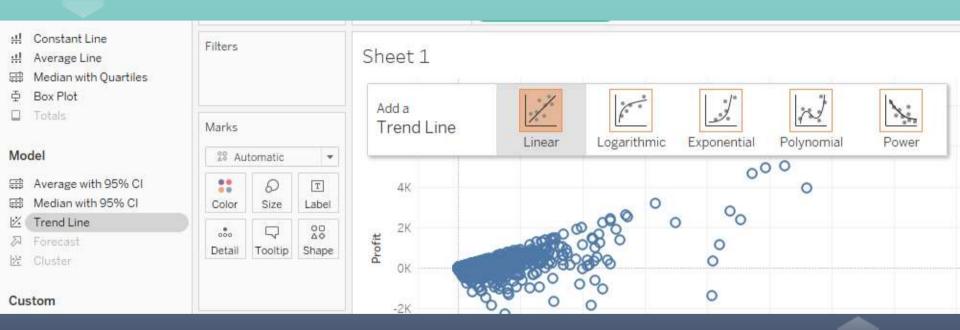
Profit = a + b Sales

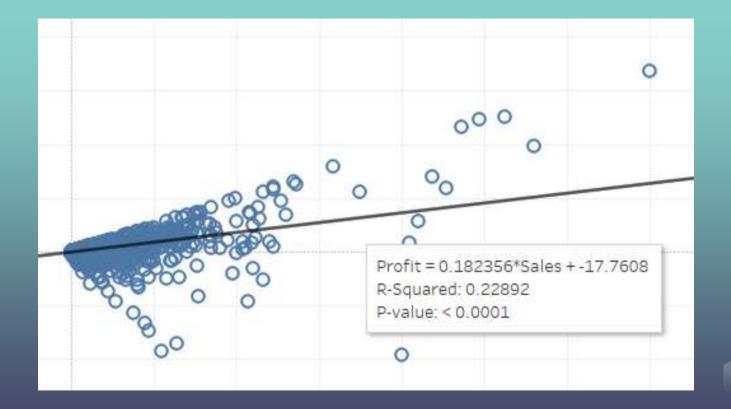
#### Go to analytics page. Use sales as columns and profit as rows.

Set "dimensions" for each variable.



~





## THE MOVIE DATASET



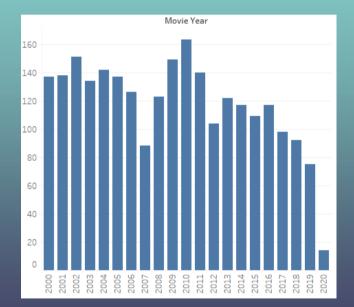
#### THE MOVIE DATASET

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 <u>here</u>.



#### Number of Movies by Year



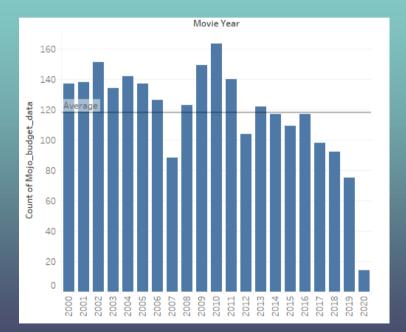
Columns: Movie Year

Set "Movie Year" as discrete

Rows: CNT(Mojo\_budget\_data)

Set "Marks" to be "Bar"

#### Number of Movies by Year

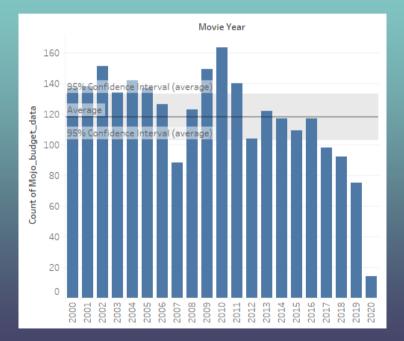


Go to Analytics Menu.

Drag "Reference Line" to the Plot.

Similarly, you can add median to your plot.

#### Number of Movies by Year

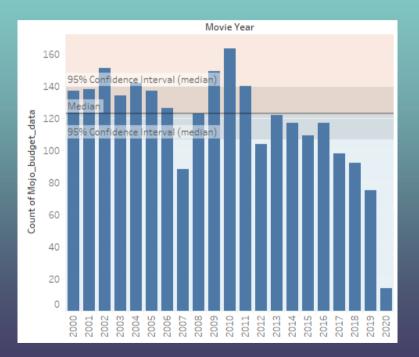


Now we plot the confidence interval of the mean.

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

You can specify your CI.

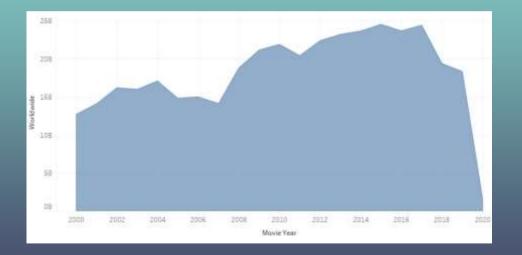
#### Number of Movies by Year



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.

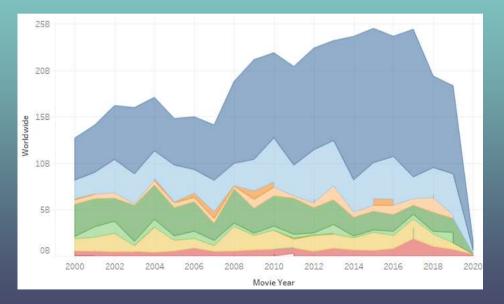


#### Columns: Movie Year

#### Rows: Worldwide (SUM)

Show me:

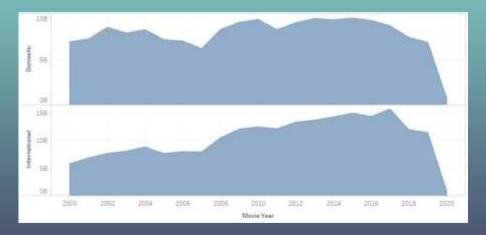




Columns: Movie Year

Rows: Worldwide (SUM)

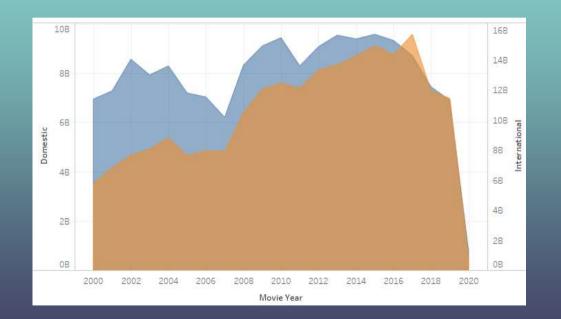
Drag "Genre 1" to Color under Marks Menu.



#### Columns: Movie Year

Rows: Domestic (SUM) and International (SUM)

Select Area Chart



#### Right Click International (Second Chart)

Choose "Dual Axis"



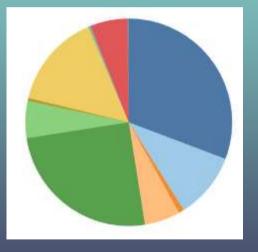


#### Right Click Domestic

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

						ſ	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	

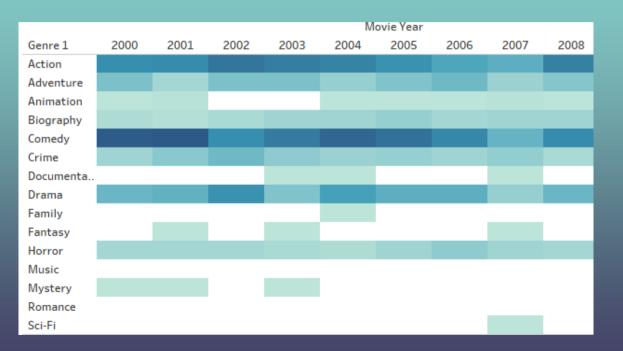
Columns: Movie Year (Discrete)

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

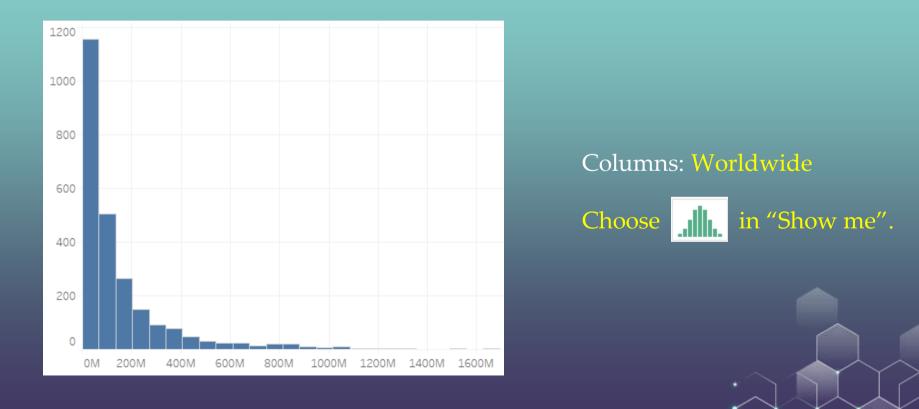
	Movie Year											
Genre 1	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009		
Action	0	0	0	0	0	0	0	۲	0	0		
Adventure		•		•		•			•	•		
Animation	0	0			0	۲	0	0	0	0		
Biography				•				•				
Comedy	•	•	0	0		•	0	0	0	0		
Crime									0	•		
Documenta				0	0			0				
Drama						•						
Family					0							
Fantasy								0				
Horror	0	•			0	0		0	0			
Music												
Mystery	0	0		0								
Romance												
Sci-Fi								۲				

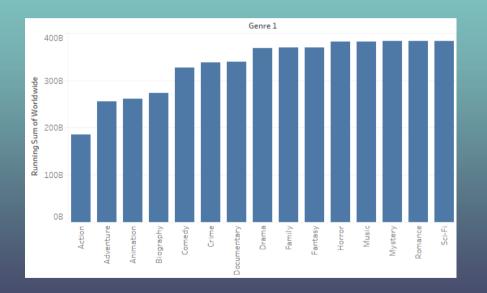
Columns: Movie Year

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

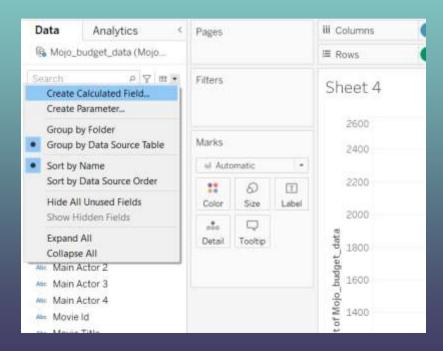




#### Columns: Genre 1

#### Rows: Worldwide

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



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

Input the following calculation.

Click OK to proceed.

-	w	or	С	WI	d	0
		<u> </u>	-		~	~

-[Worldwide]

 $\times$ 

Under Analysis Menu, choose Create Calculate Field

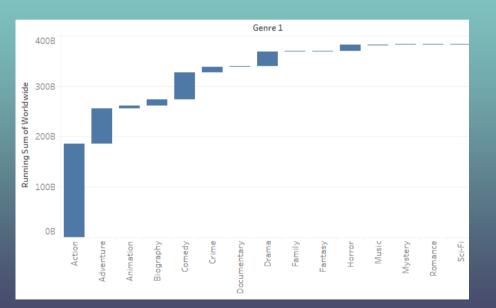
Input the following calculation.

Click OK to proceed.

The calculation is valid.





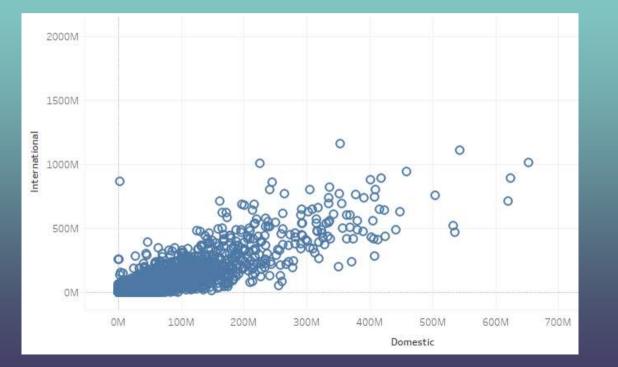


There is a new variable "-Worldwide".

Under Marks, change Automatic to Gantt Bar.

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

#### **International vs. Domestic Box Revenue**

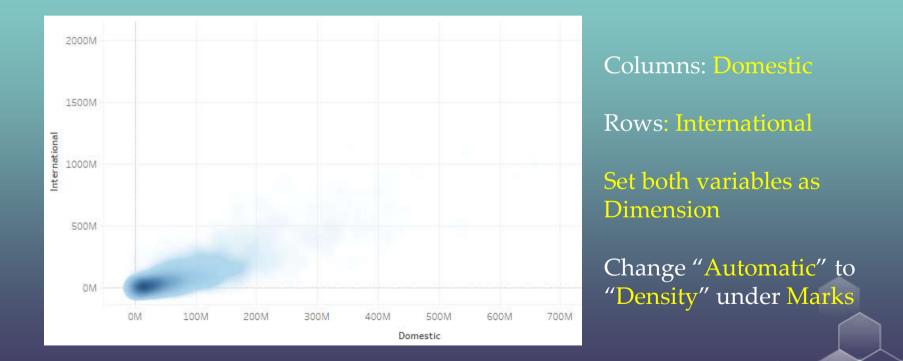


Columns: Domestic

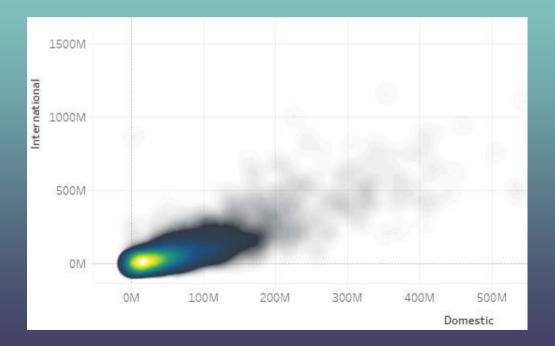
**Rows:** International

Set both variables as Dimension

#### **International vs. Domestic Box Revenue**



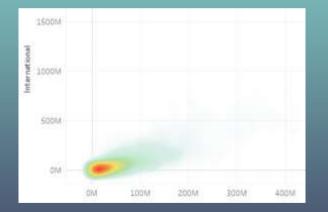
#### **International vs. Domestic Box Revenue**



Change "Automatic" to "Density" under Marks

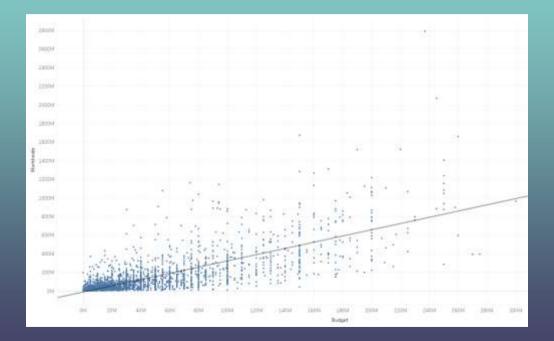
Set Colors to "Density-Multicolor" under Marks

## **Density Heatmap**



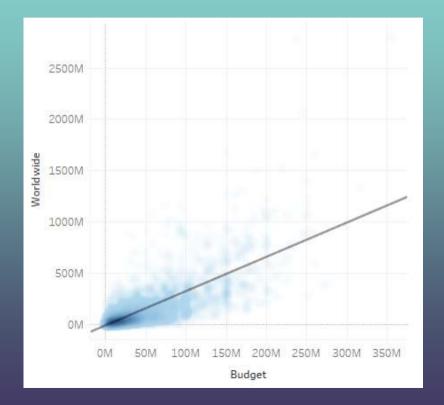
In a density heatmap, in places where the observations are dense (i.e., many points in the area), the heatmap displays a warm color. In places where observations are sparse, the heatmap displays a cold color.

# **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.

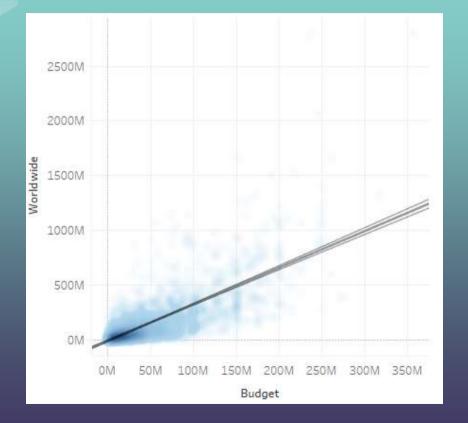
# **Regression Line**



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".

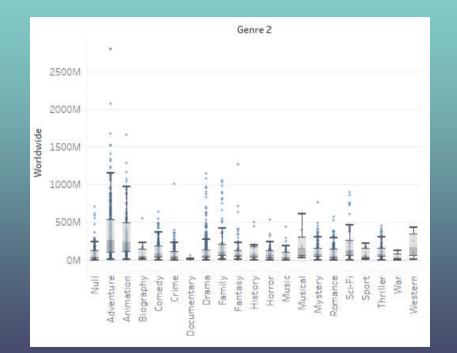


#### **Confidence Bands**

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 <u>here</u> 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 <u>here</u>.

## **Box and Whisker Plot**

How to draw the Box and Whisker Plot?

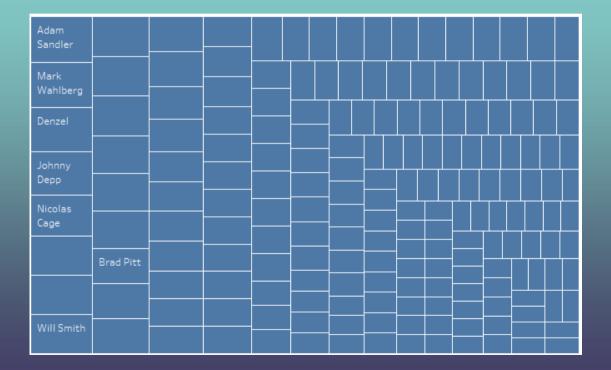
Use Genre 2 as rows and Worldwide as columns.

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

Julia Roberts Anna Faris Mel Gibson Harrison Ford Ashton Kutcher Shia LaBeouf Anna Kendrick James McAvoy Charlize Theron Sean Penn John Travolta Scarlett Johansson Anthony Hopkins Kevin Hart Jack Black Natalie Portman Steve Martin Colin Farrell Julianne Moore Jason Statham Brendan Fraser Steve Carell Daniel Craig Dennis Quaid Cameron Diaz Aaron Eckhart **Richard Gere** Clive Owen Owen Wilson Chris Hemsworth Ben Affleck Sandra Bullock Christian Bale Hugh Jackman Bradley Cooper Reese Witherspoon Will Ferrell Russell Crowe Zac Efron Morgan Freeman Ben Stiller Channing Tatum Angelina Jolie Johnny Depp Emily Blunt Joseph Gordon-Levitt Ethan Hawke Tom Cruise Liam Neeson Leonardo DiCaprio Matthew McConaughey Nicole Kidman Denzel Washington Dwavne Johnson Mark Wahlberg Brad Pitt Nicolas Cage Robert De Niro Will Smith Samuel L. Jackson Adam Sandler Inscite Matt Damon Robert Downey Jr. Jake Gyllenhaal Vin Diesel Tom Hanks Jennifer Lawrence Jennifer Lopez Ryan Reynolds Ewan McGregor Jim Caviezel Kevin Spacey Jason BatemanCuba Gooding Jr. Sylvester Stallone Bruce Willis Eddie Murphy Gerard Butler Michael Fassbender John Cusack Billy Bob Thornton George Clooney Kristen Stewart Edward Norton Daniel Radcliffe Jim Carrey Melissa McCarthy Keanu Reeves Ren??e Zellweger Jackie Chan Helen Mirren Mike Myers Ryan Gosling Jennifer AnistonJesse Eisenberg Meryl Streep Patrick Stewart Kate Beckinsale Seth Rogen Chris Pine Kevin Costner

Milla Invovich

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 Matt Damon Will Smith Daniel Radcliffe Vin Diesel Tom Cruise Hugh JackmanGeorge 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

Mike Myers Shia LaBeouf Chris Pratt Daisy Ridley Ben Affleck Elijah Wood Sam Worthington Ben Stiller Hugh Jackman Tom Cruise Vin Diesel Johnny Depp<sub>Daniel</sub> Craig Adam Sandler Robert Downey Jr. Dwayne Johnson Leonardo DiCaprio Mark Wahlberg Daniel Radcliffe Jennifer Lawrence Tom Hanks Christian Bale Will Smith Kristen Stewart Sandra Bullock Matt Damon Brad Pitt Steve Carell

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 <u>Tagul</u> (Optional Topic)



# Dituhui.com Creating Maps for China (Optional Topic)



