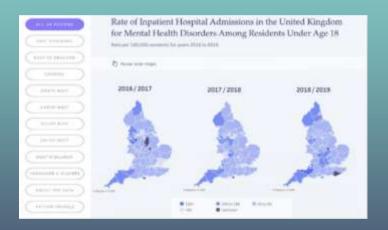
Introduction to Tableau

Your new data analysis software

Data Visualization with Tableau



Tableau Public Gallery





Import Data

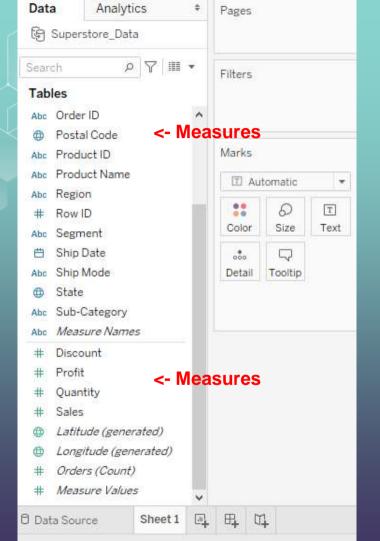
Download the sample datafile "SuperStore_Data" and open it with your Tableau.

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

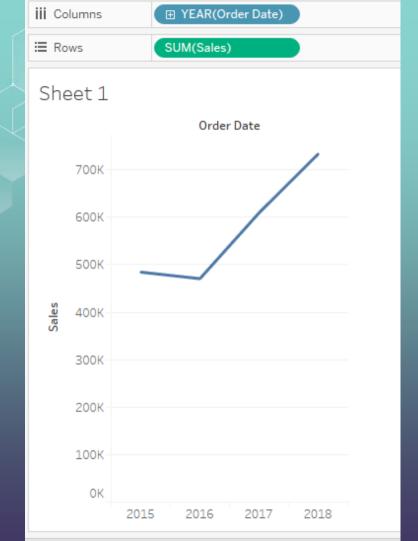
#		Abc	曲		*	Abc		Abo
Row ID =		Order ID	Order Date		Number (decimal)		ode	Customer ID
	1	CA-2017-152156	11/8/2017		Number (who Date & Time	le)	Class	CG-12520
	2	CA-2017-152156	11/8/2017		Date		Class	CG-12520
	3	CA-2017-138688	6/12/2017		String Boolean		Class	DV-13045
	4	US-2016-108966	10/11/2016	_			ird Class	SO-20335
	5	US-2016-108966	10/11/2016	150	Default	stan	oard Class	SO-20335
	6	CA-2015-115812	6/9/2015	6/14	1/2015	Stan	dard Class	BH-11710
	7	CA-2015-115812	6/9/2015	6/14	1/2015	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).

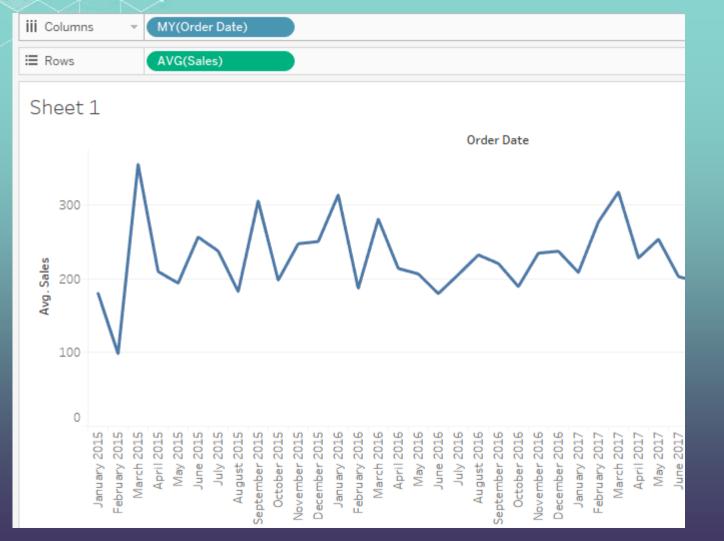


Here, Sales is a measure.

Order data is a dimension.

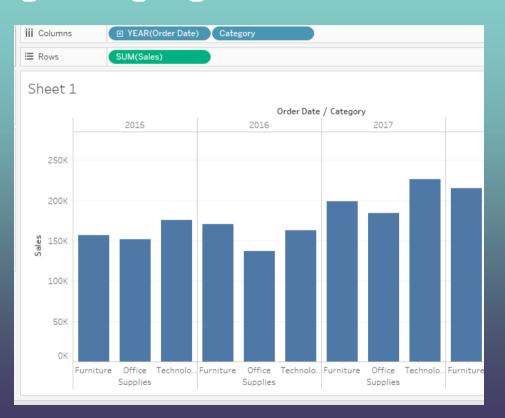
We partition sales based on year of order.

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.

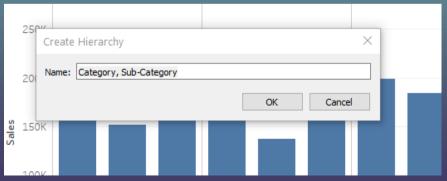
Adding Category As Another Classifier



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

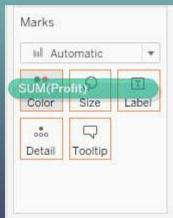
Then, you can display each product subcategories.

iii Columns	▼ (☐ Category	▼ Sub-Category	
		Category		
≔ Rows	SUM(Sales)			
Sheet 1				
SHOELT				

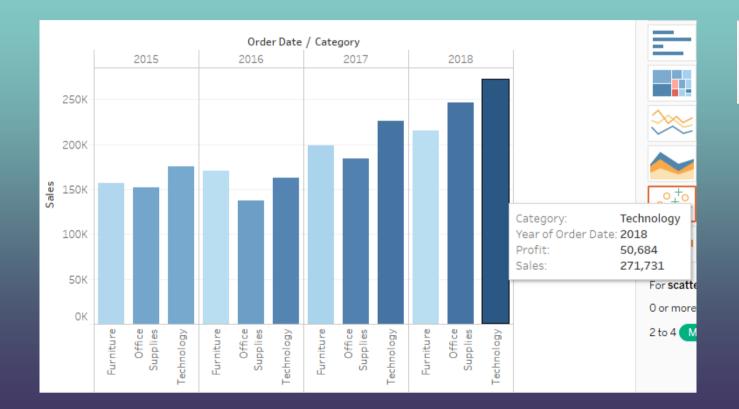
Color Your Output

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

Drag profit to color.



Color Your Output



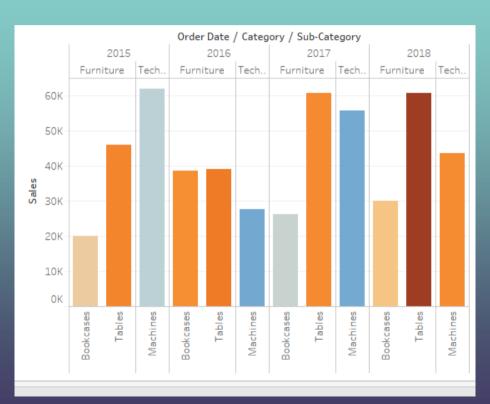


Apply Filters

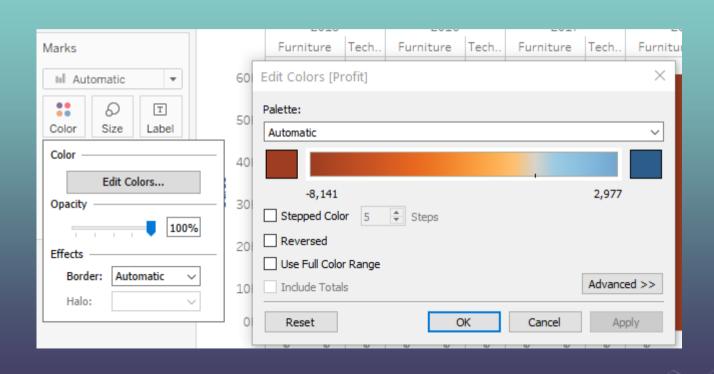
Tab	les					
~ 品	Product	^				
A	be Category					
A	Bub-Category	Marks				
Abc	Product ID	Add to SI	neet			
Abc	Product Name	Show Filte	er			
Abc	Region	Duplicate				
#	Row ID	Rename Hide				
Abo	Segment					
	Ship Date					
Abc	Ship Mode	Aliases				
#	State	Create	*			
Abc	Measure Names	Transform	ı +			
#	Discount	Convert	o Measure			
#	Profit					
#	Quantity	Change D				
#	Sales	Geograph	100000000000000000000000000000000000000			
#	Latitude (generated)	Default Properties				
0	Longitude (generated)	Group by	•			
#	Orders (Count)	Folders	>			

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

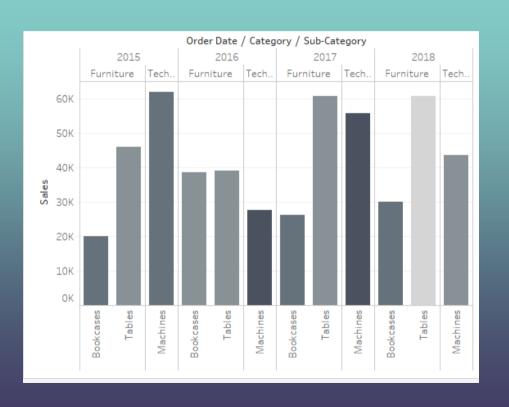
Apply Filters



Edit Colors

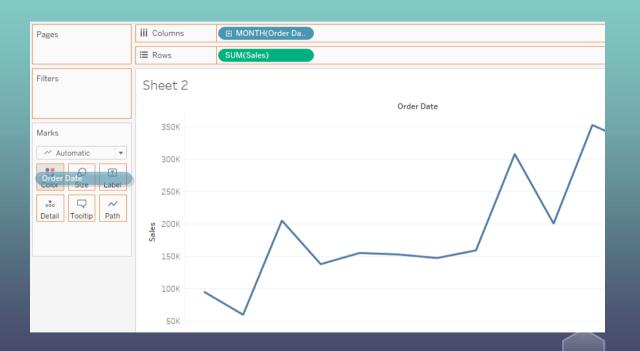


Edit Colors



Contrast Sales Data by Year

Drag Order Date to Color



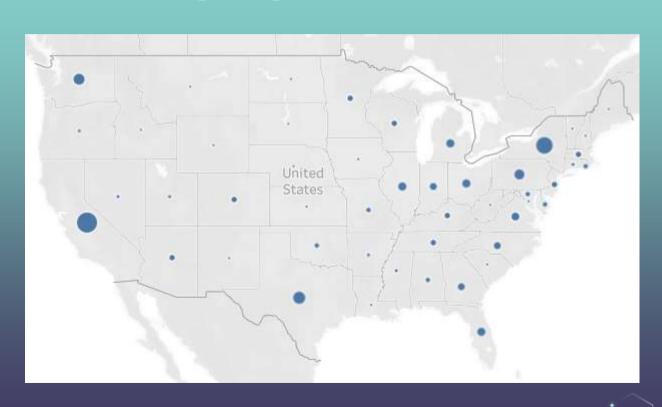
YoY Growth by Editing Rows

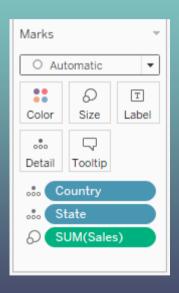


Use Control Button to select multiple attribute "Country", "State" and "Sales".

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

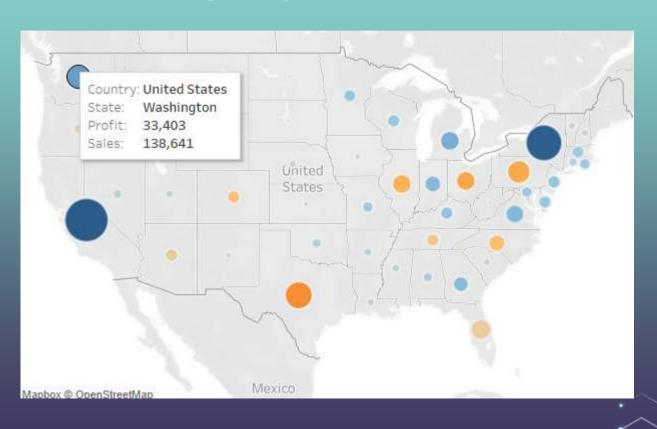
United States California	United States Texas United States Washington	United States Illinois	United States Ohio	United States Michigan	United States Virginia
		United Star North Carolina	tes United		
		United Star	tes United States		
United States New York	United States Pennsylvania	United Sta Georgia	tes		
	United States Florida	United Sta	tes		
		United Sta	tes		

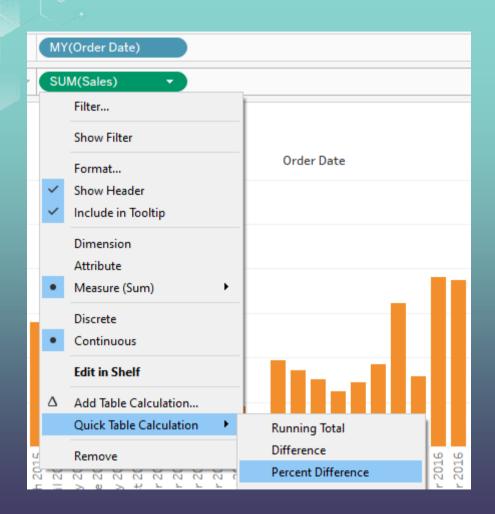




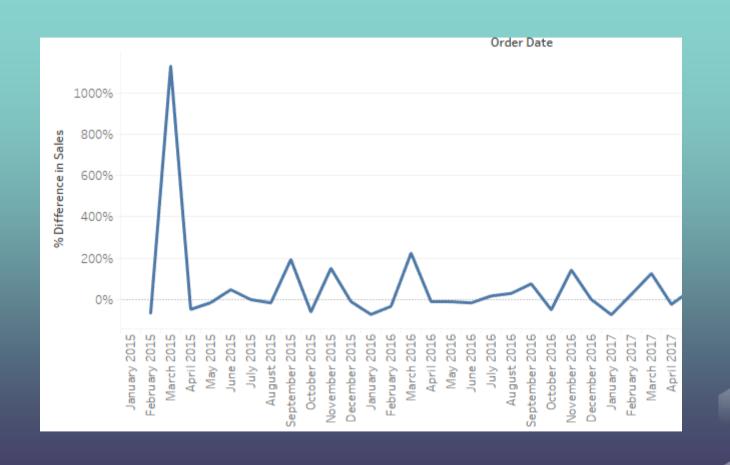
We can use profit to select the color of the circles in the map.

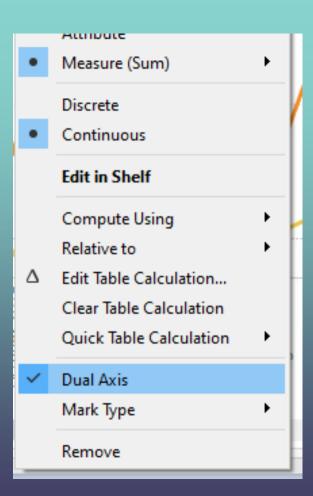
Moreover, we adjust the size of the circles.



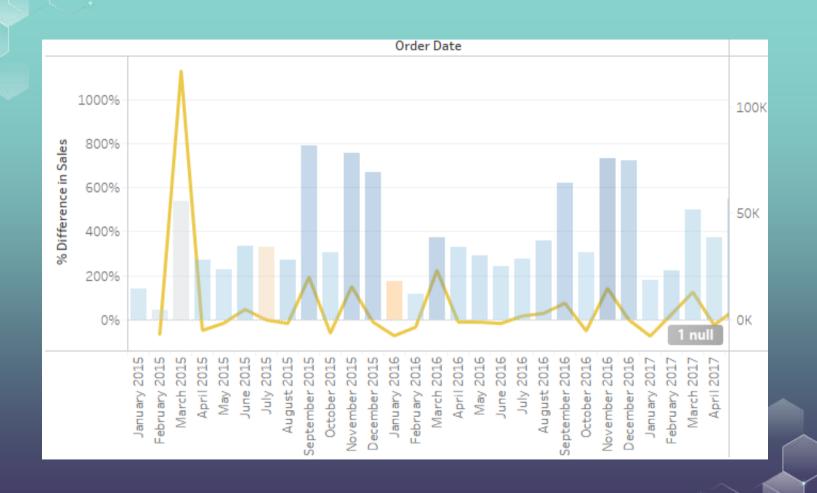


Here, we want to calculate how the sales change over time.





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



Regression Analysis

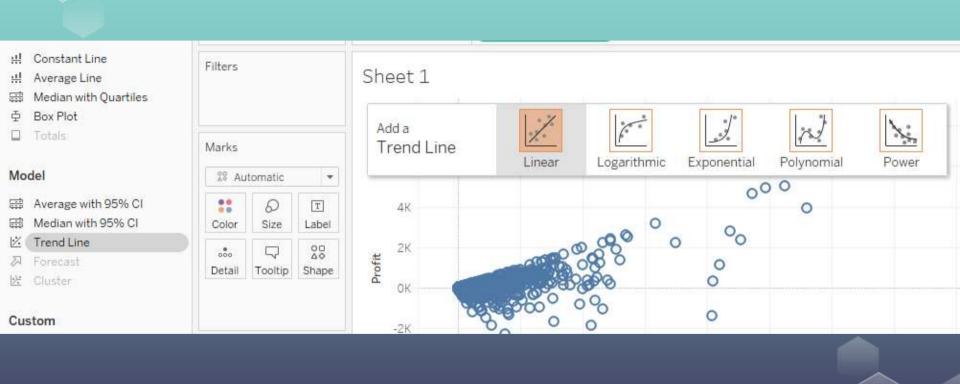
Suppose that we want to run the following regression:

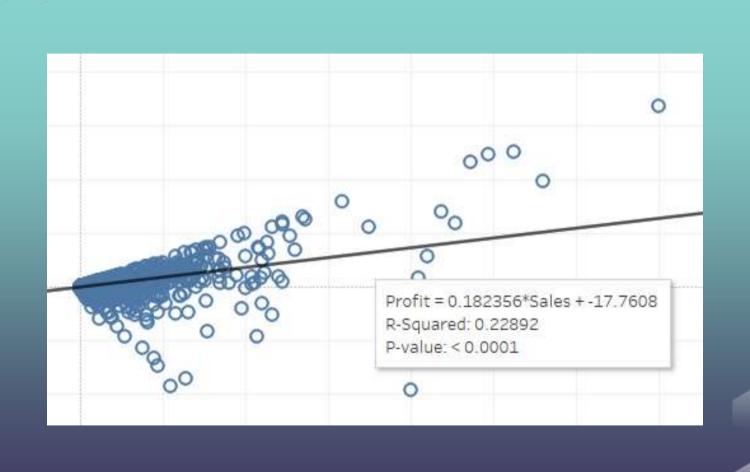
Profit =
$$a + b$$
 Sales

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

Set "dimensions" for each variable.









THE MOVIE DATASET

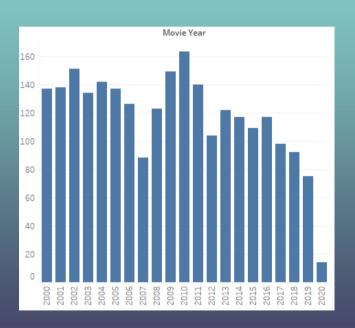


THE MOVIE DATASET

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

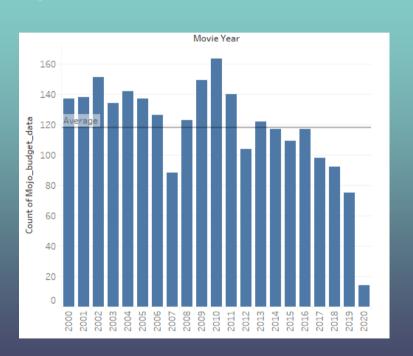


Columns: Movie Year

Set "Movie Year" as discrete

Rows: CNT(Mojo_budget_data)

Number of Movies by Year

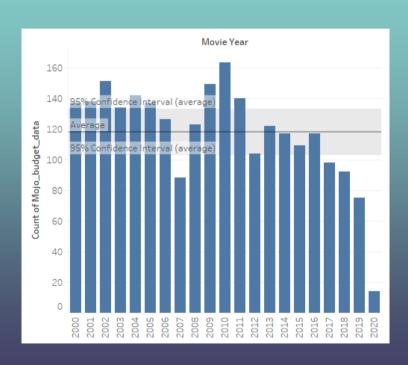


Go to Analytics Menu.

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.

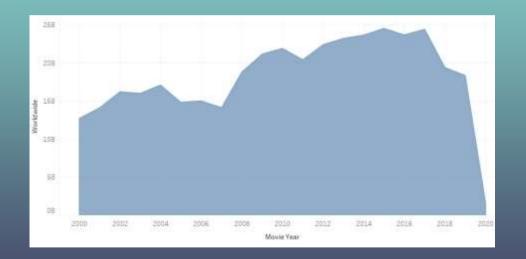
Number of Movies by Year



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.

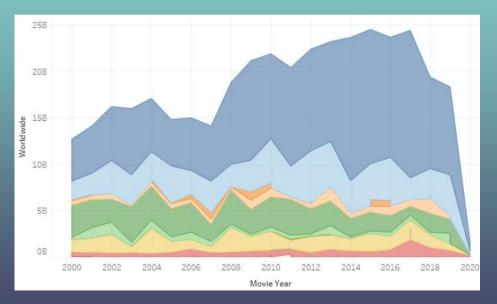


Row: Movie Year

Column: Worldwide (SUM)

Show me:





Row: Movie Year

Column: Worldwide (SUM)

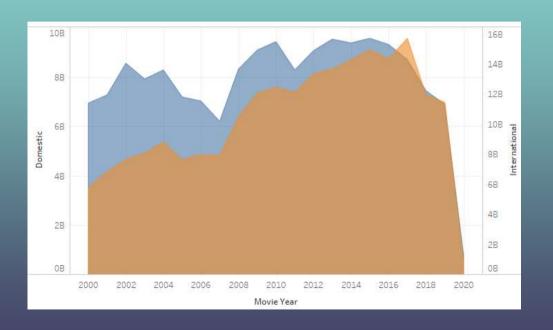
Drag "Genre 1" to Color under Marks Menu.



Row: Movie Year

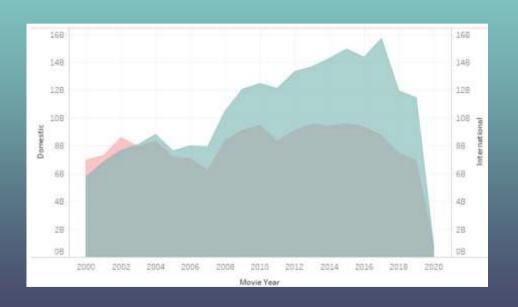
Column: Worldwide (SUM) and Domestic (SUM)

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" to "entire view"

Genre by Year

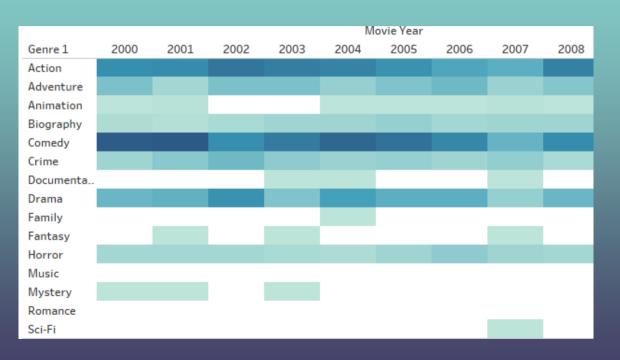
						1	Movie Year			
Genre 1	2000	2001	2002	2003	2004	2005	2006	2007	2008	
Action	33	34	42	39	37	32	25	22	38	
Adventure	15	6	15	15	9	14	18	8	13	
Animation	1	2			1	1	1	2	1	
Biography	4	3	5	7	7	9	6	7	7	
Comedy	51	52	33	40	47	43	36	20	34	
Crime	7	12	18	11	8	9	7	10	5	
Documenta				1	1			1		
Drama	19	21	32	14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

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

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

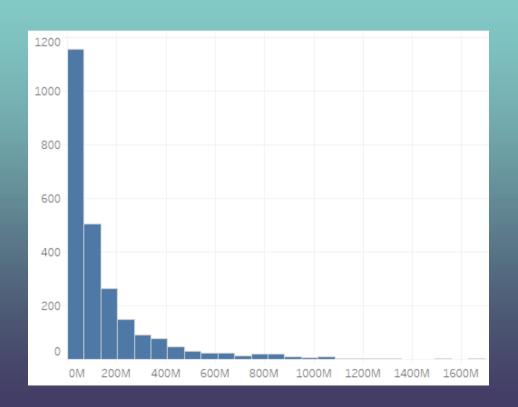
Genre 1	Movie Year									
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Action	0	0	0	0	0	0		0	0	0
Adventure		0								
Animation	0	0			0	0	0	0	0	0
Biography	0	0	0							
Comedy	•	•	0	0	•	•	0	0	0	0
Crime										
Documenta				0	0			0		
Drama										
Family					0					
Fantasy		0								
Horror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

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



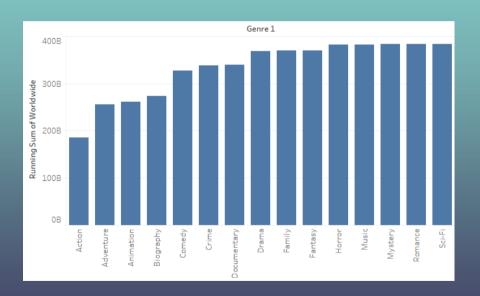
Columns: Worldwide

Choose



in "Show me".

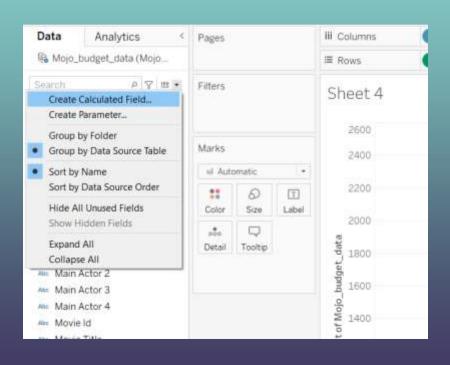




Columns: Genre 1

Rows: Worldwide

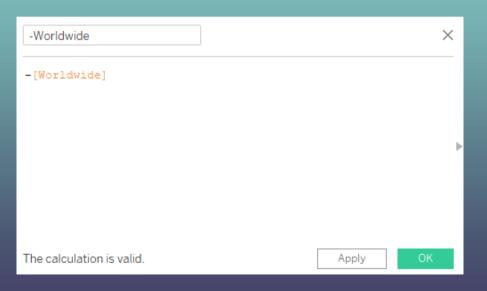
Right click Worldwide, select Add Table Calculation. Then under Calculation Type, choose Running Total.



Under Data Menu, choose Create Calculate Field

Input the following calculation.

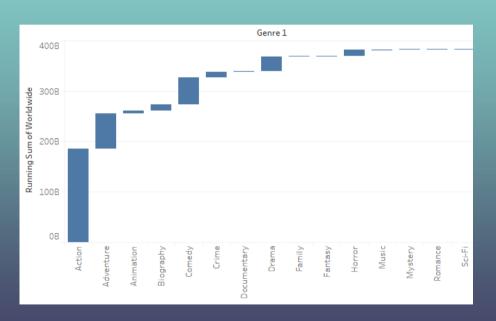
Click OK to proceed.



Under Analysis Menu, choose Create Calculate Field

Input the following calculation.

Click OK to proceed.

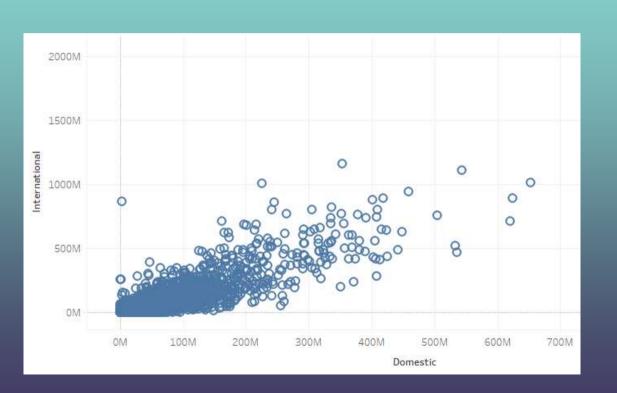


There is a new variable "-Worldwide".

Under Marks, change Automatic to Gantt Bar.

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

International vs. Domestic Box Revenue

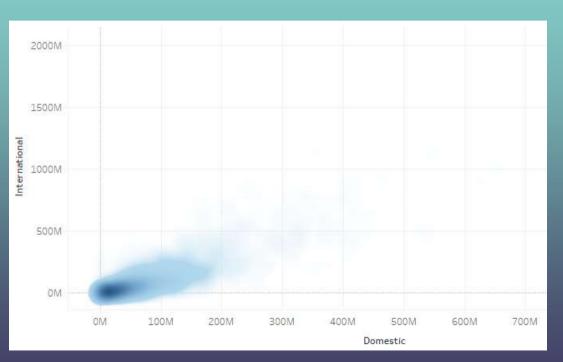


Columns: Domestic

Rows: International

Set both variables as Dimension

International vs. Domestic Box Revenue



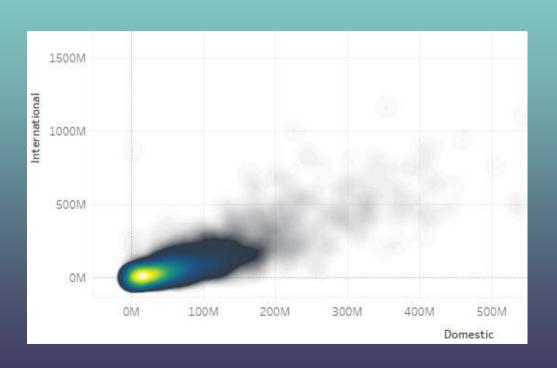
Columns: Domestic

Rows: International

Set both variables as Dimension

Change "Automatic" to "Density" under Marks

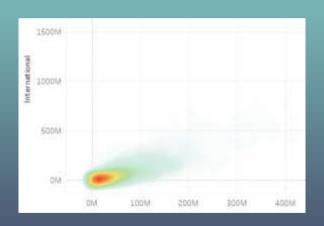
International vs. Domestic Box Revenue



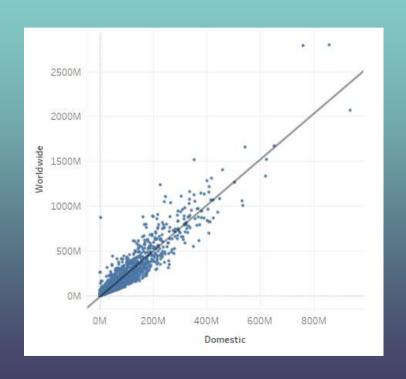
Change "Automatic" to "Density" under Marks

Set Colors to "Density-Multicolor" under Marks

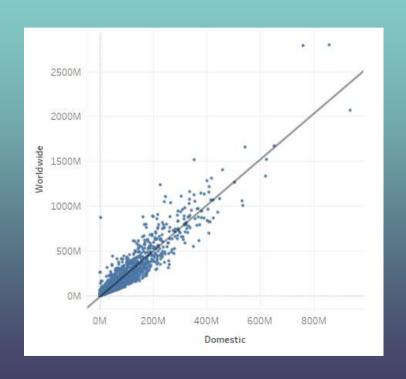
Density Heatmap



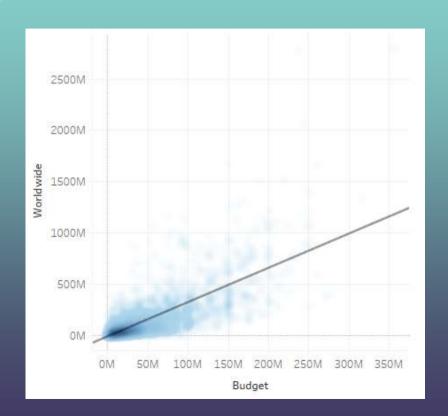
In a density heatmap, in places where the observations are dense (i.e., many points in the area), the heatmap displays a warm color. In places where observations are sparse, the heatmap displays a cold color.



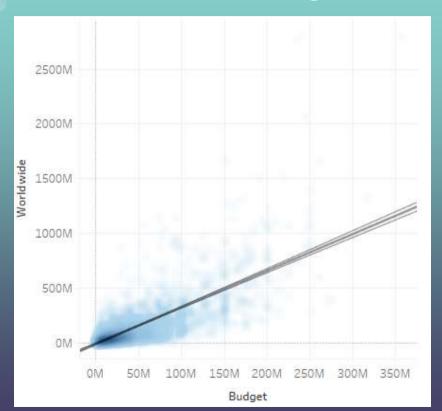
Here, we use budget as the independent variable (i.e., Column) and worldwide box office as the dependent variable (i.e., Row) and draw the regression line.



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.



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



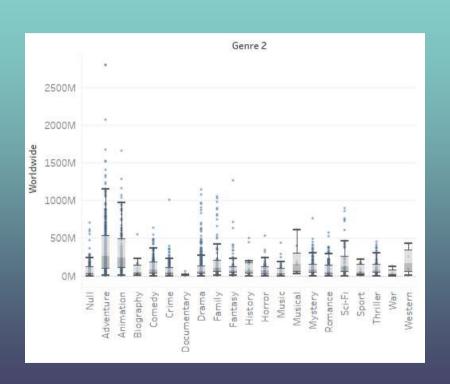
Click and edit the regression line.

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 <u>here</u>.

Box and Whisker Plot

How to draw the Box and Whisker Plot?

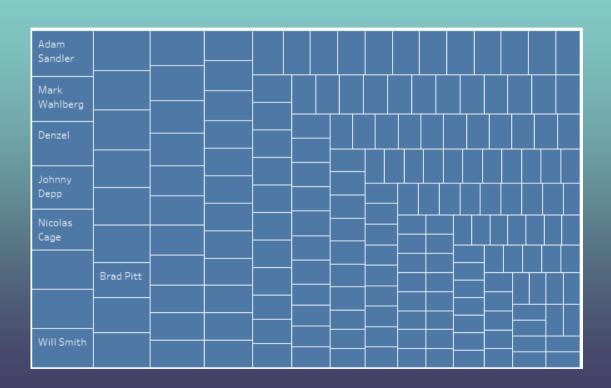
Use Genre 2 as rows and Worldwide as columns.

Under Analysis, uncheck "Aggregate Measures".

Under Show me, choose



Actor/Actress Treemap



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

Milla Invovich Anna Faris Mel Gibson Harrison Ford Ashton Kutcher Shia LaBeouf Anna Kendrick James McAvoy Charlise Theron Sean Penn John Travolta Scarlett Johansson Anthony Hopkins Kevin Hart Jack Black Natalie Portman Steve Martin 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 in Ciba Matt Damon Robert Downey Jr. Jake Gyllenhaal Vin Diesel Tom Hanks Jennifer Lawrence Jennifer Lonez Ryan Reynolds Ewan McGregor Jim Caviezel Kevin Spacey Jason BatemanCuba Gooding Jr. Sylvester Stallone Eddie Murphy Gerard Butler Michael Fassbender John Cusack Billy Bob Thornton George Clooney Kristen Stewart Edward Norton Daniel Radcliffe Jim Carrey Melissa McCarthy Jackie Chan Helen Mirren Mike Myers Ryan Gosling Jennifer AnistonJesse Eisenberg Meryl Streep Patrick Stewart Kate Beckinsale Chris Pine

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 Tom Hanks Brad Pitt Will Ferrell
Matt DamonWill 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
Elijah Wood Sam Worthington Ben Stiller Hugh Jackman
Tom Cruise Vin Diesel Johnny Depp Daniel Craig

Adam Sandler Robert Downey Jr. Dwayne Johnson

Leonardo DiCaprio Mark Wahlberg Daniel Radcliffe

Jennifer LawrenceTom Hanks Christian Bale Will Smith Kristen StewartSandra 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>

