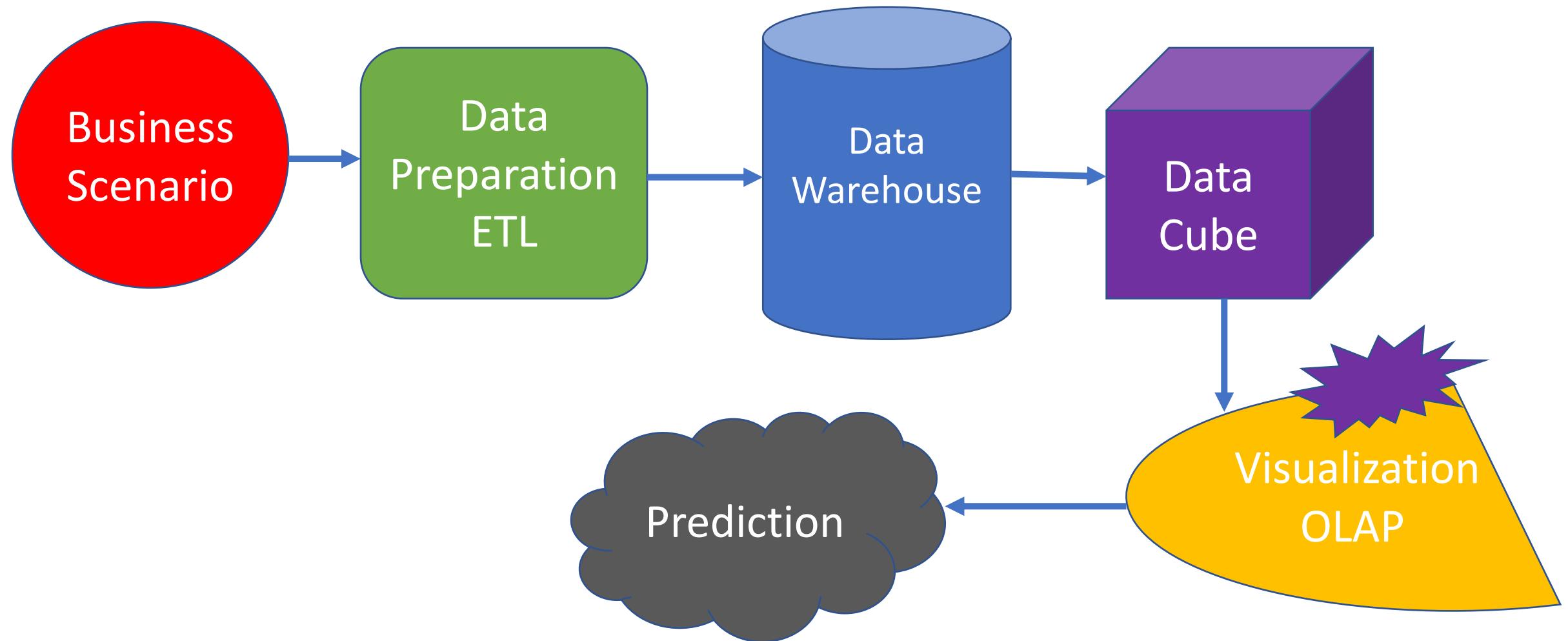


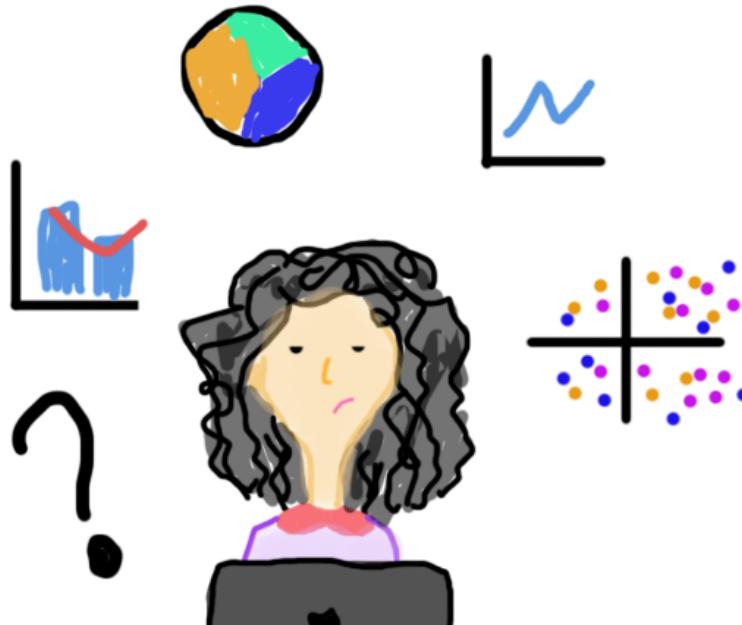
Data Visualization

Trishla Shah

BI Scenario (Useful for your final Project!)



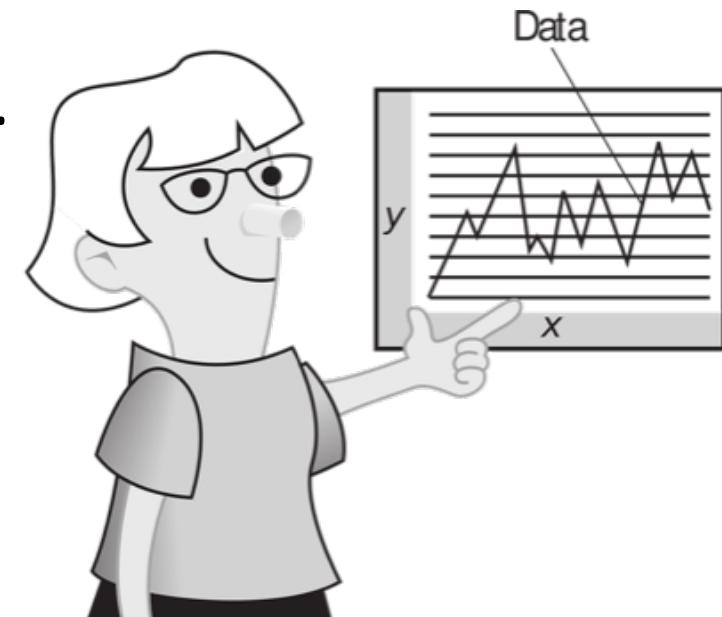
What is Data Visualization?



Data Visualization

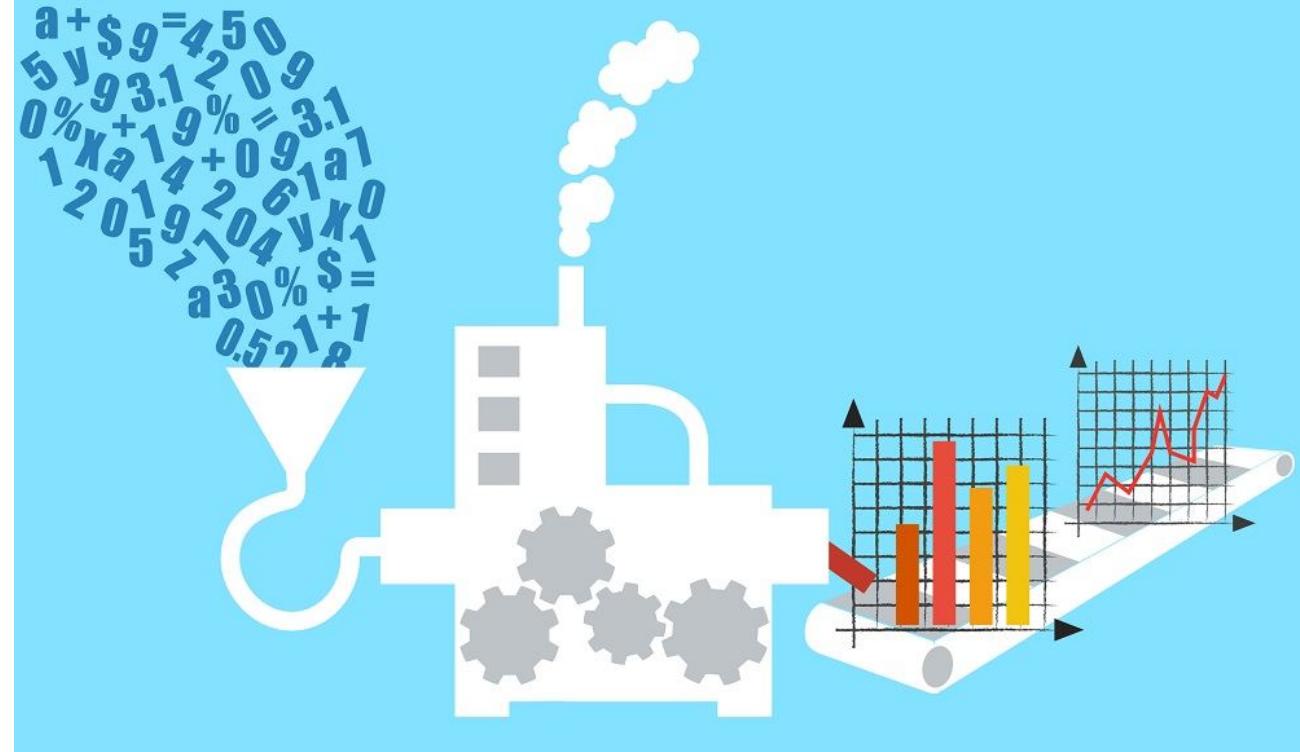
Graphical representation of information and data

By using visual elements like charts, graphs, and maps, data visualization is an accessible way to see and understand trends, outliers, and patterns in data.



We need to know what Big Data says

- Visualization is an increasingly key tool to make sense of the trillions of rows of data generated every day.
- A good visualization tells a story, removing the noise from data and highlighting the useful information.



Advantages of Data Visualization

- Enhanced Assimilation of Business Information
- Quick Access to Relevant Business Insights
- Better Understanding of Operational & Business Activities
- Rapid Identification of Latest Trends
- Accurate Customer Sentiment Analysis
- Direct Interaction with Data
- Predictive Sales Analysis
- Drill-Down Sales Analysis
- Easy Comprehension of Data



Why data visualization is important for Industry?

- It's hard to think of a professional industry that doesn't benefit from making data more understandable.
- Every industry benefits from understanding data such as:
 - Government
 - Finance
 - Marketing
 - History
 - Consumer good
 - Service industries
 - Education
 - Health
 - Sports, and so on.



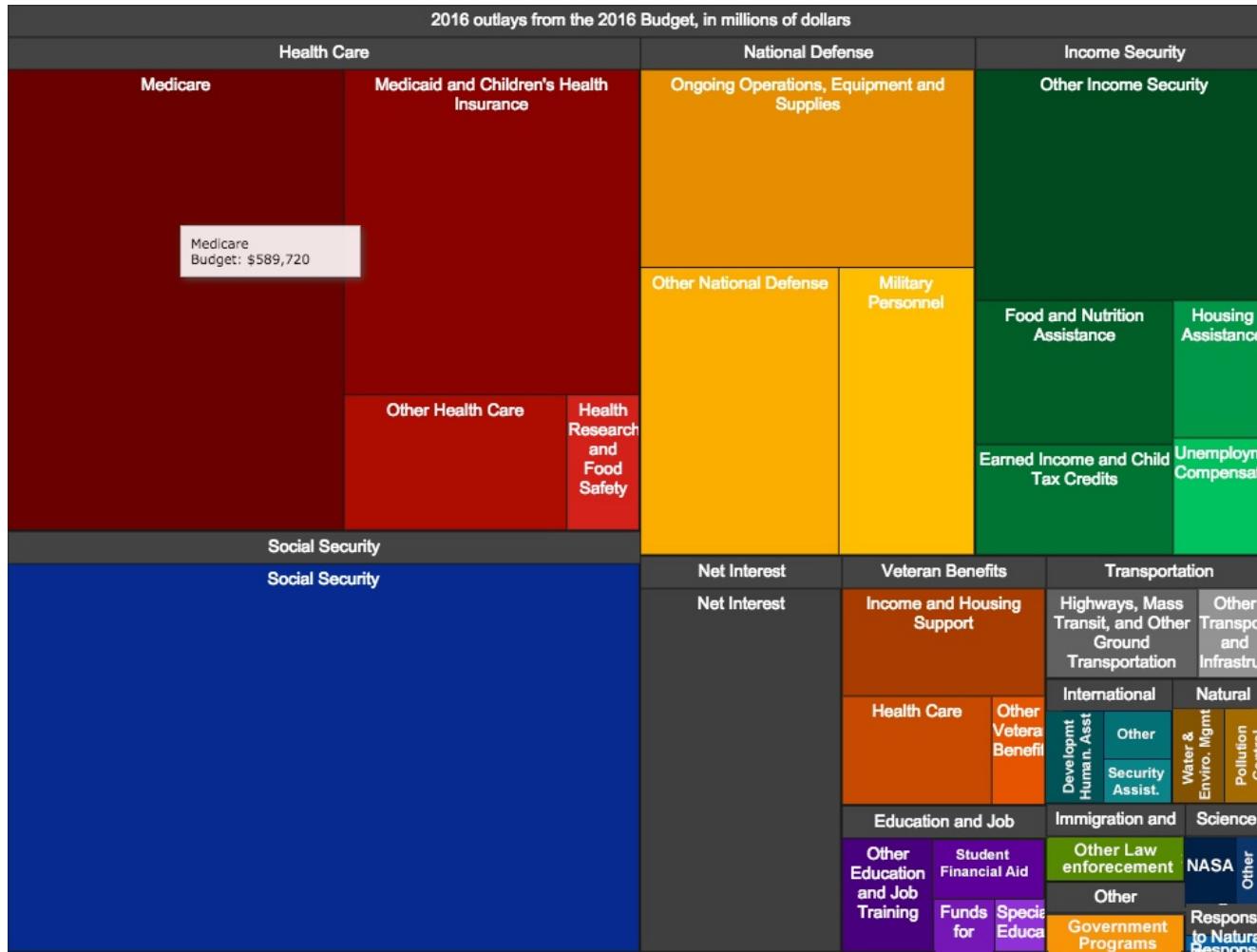
I EXPECTED A LITTLE
MORE OUT OF THIS
DATA VISUALIZATION
WORKSHOP.



Examples of Visualizations



Interactive Government Budget



Link: <https://obamawhitehouse.archives.gov/node/320071>

more examples



PLEASE

NEXT AMERICA

Percent of U.S. Population by Age Group, 1950-2060

■ Baby Boomers

MALE

1950

FEMALE

85+

75-79

50-54

25-29

0

4% 2% 0% 2% 4%

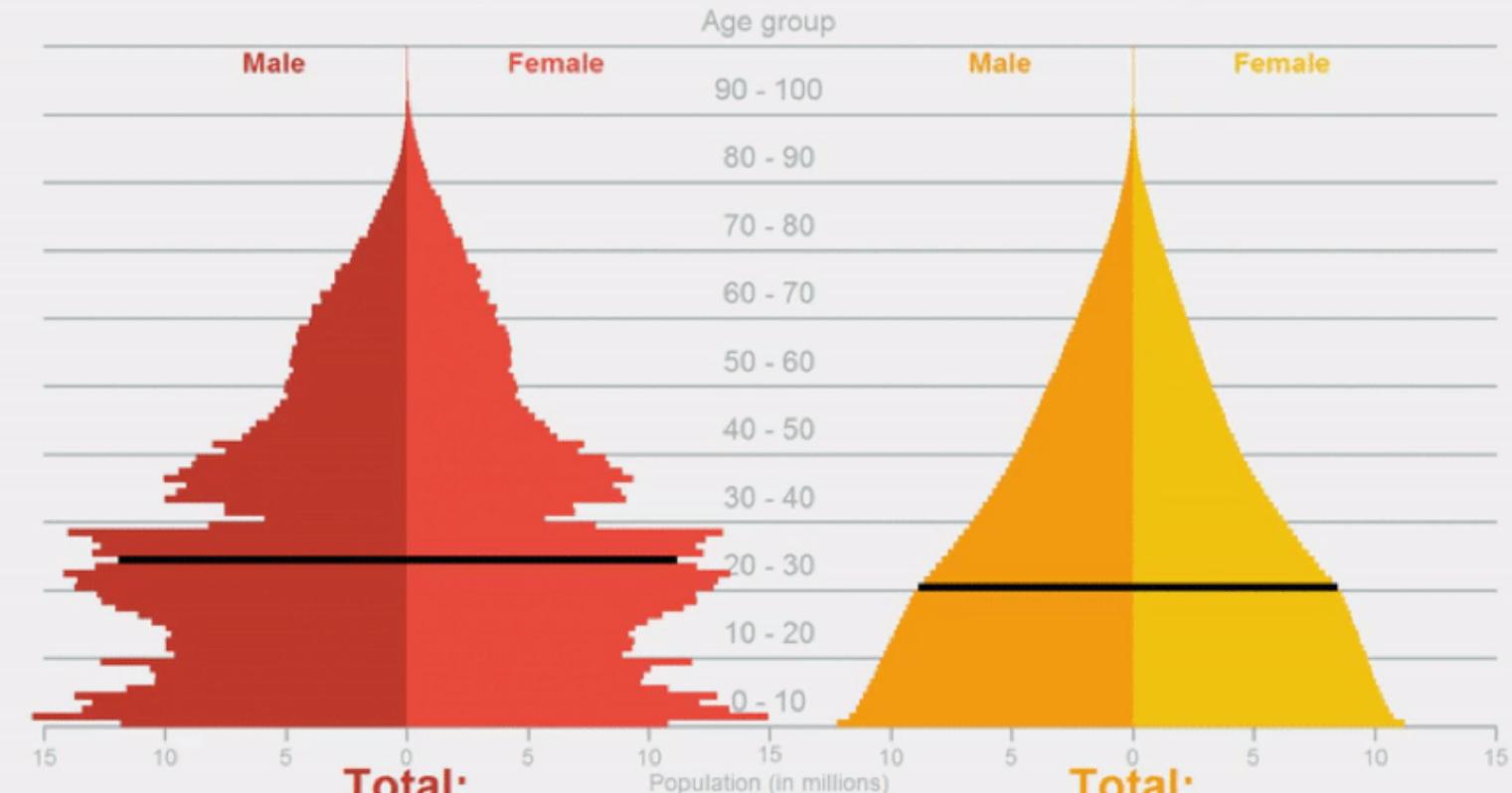
PEW RESEARCH CENTER

Link: <http://www.pewresearch.org/next-america/>

Population projection 1991

China

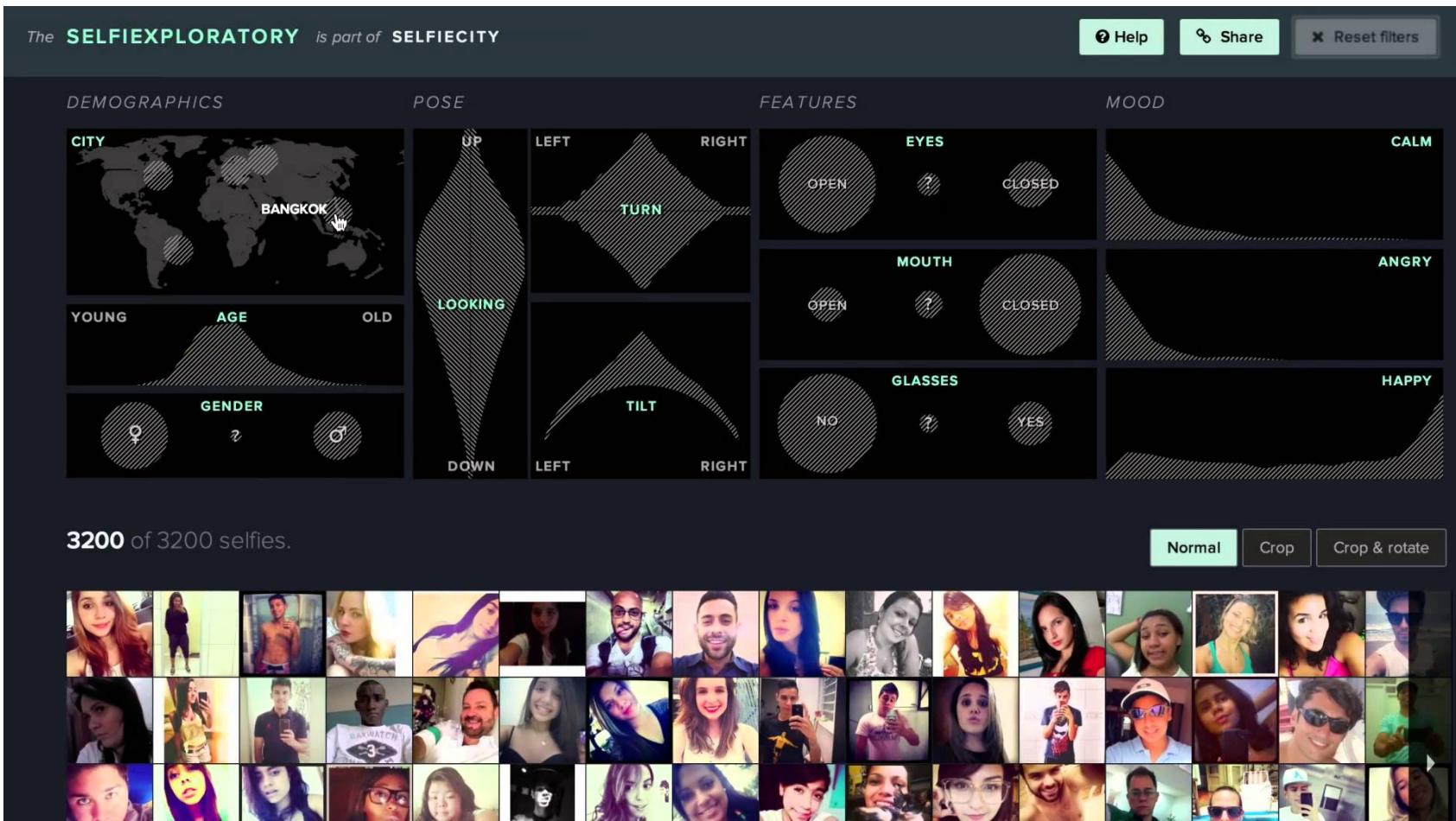
India



■ Median age

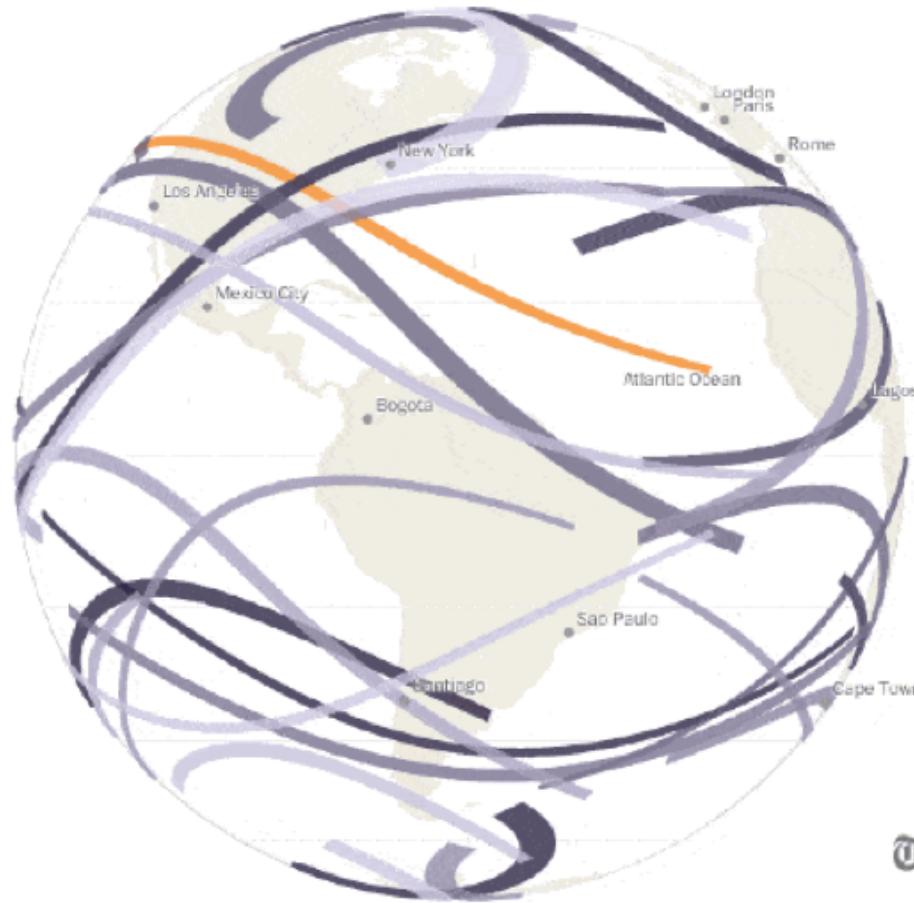
Data source: US Census Bureau IDB
Visualization by @aronstrandberg

SELFIECITY



Link: <http://selfiecity.net/>

Every Upcoming Solar Eclipse (until 2080)



The Washington Post

Link:

https://www.washingtonpost.com/graphics/national/eclipse/?utm_term=.5e035337d0d2

Question!



Think and share any unique Visualization Project.

Why is the speedometer
stuck on 35?

The car only collects
speed data once
a year.



Current Tools in Industry



Tableau



Microsoft Power BI



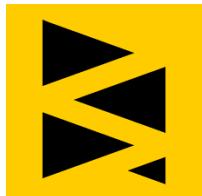
Qlik



SAP Lumira



SAS visual Analytics



Sisense



Tibco Spotfire



ClearStory Data

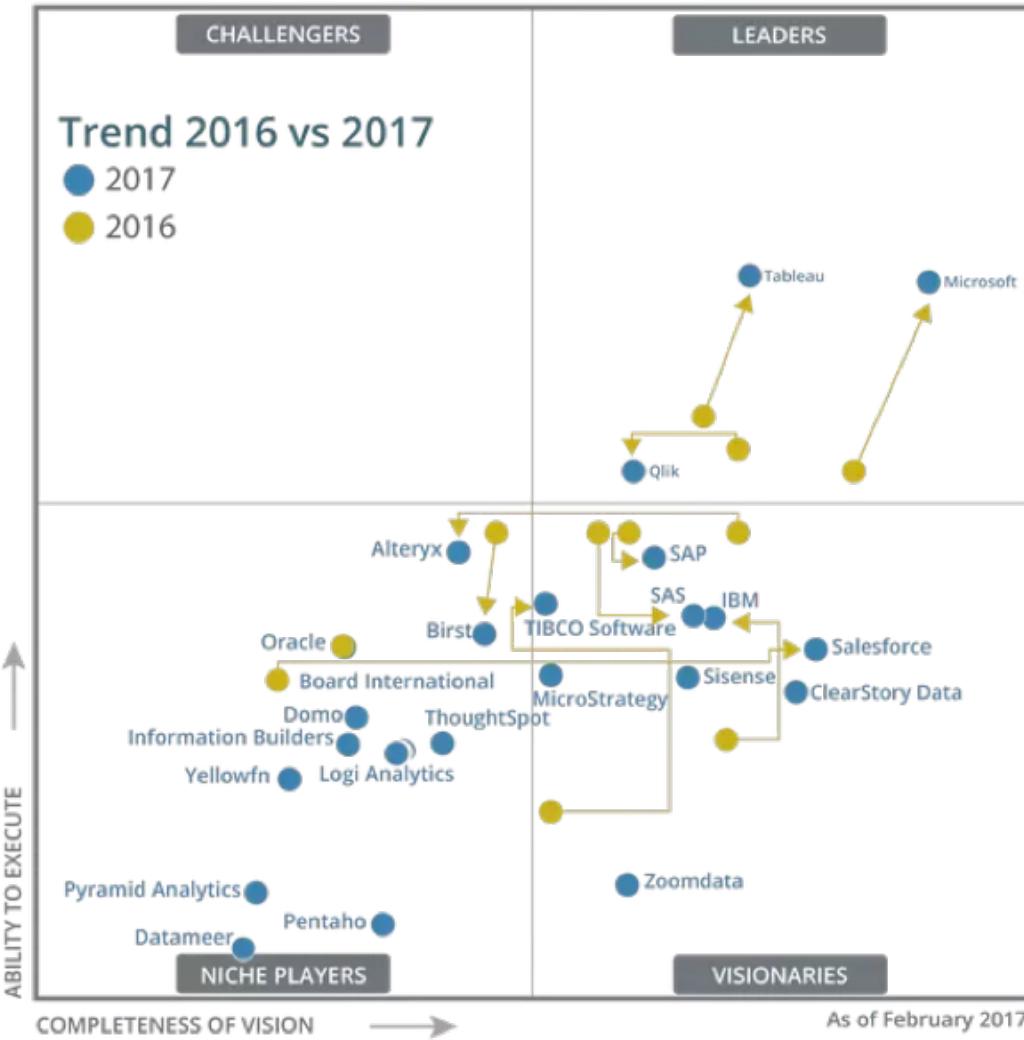


IBM Cognos

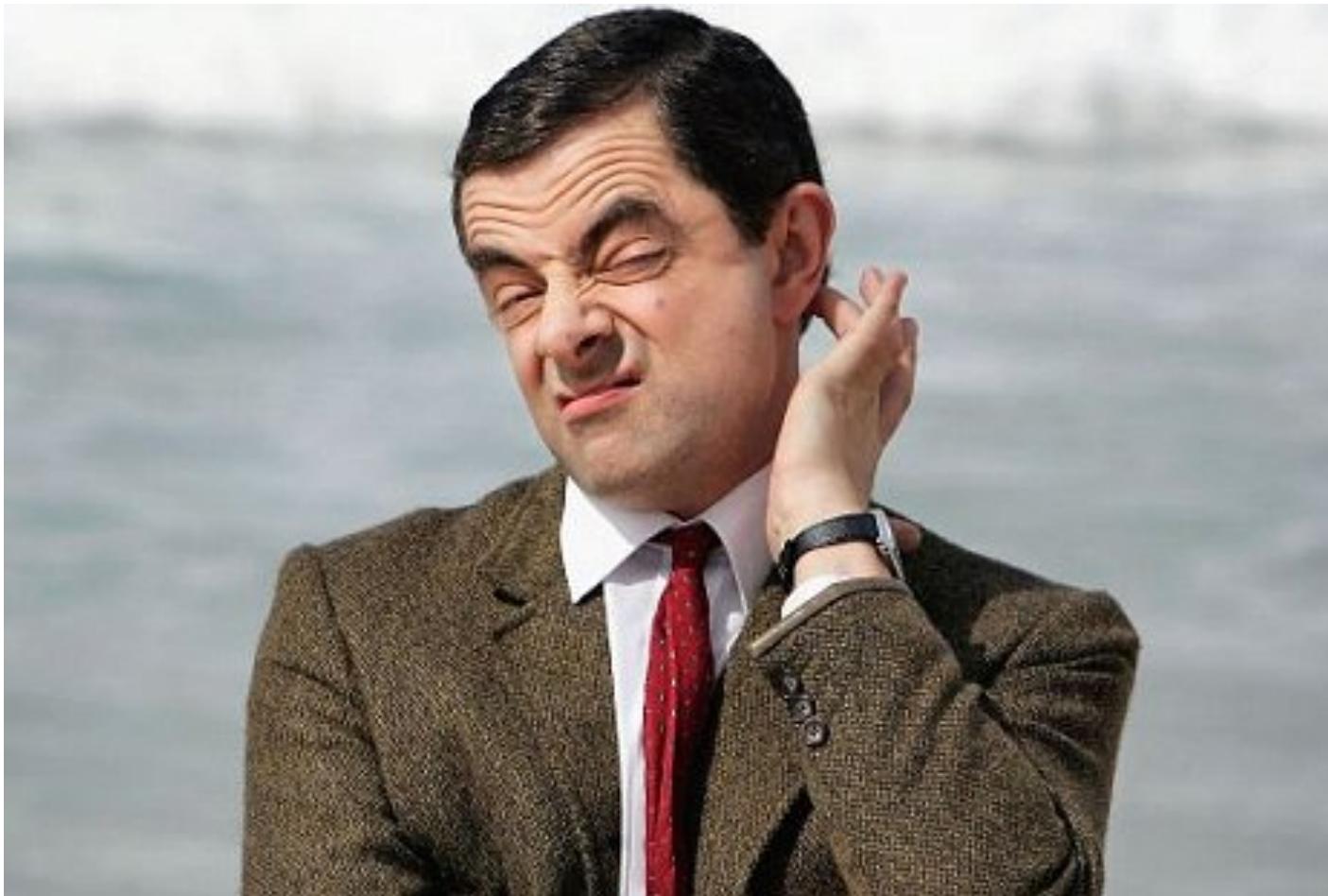


MicroStrategy Analytics
Desktop

Trends



How to Choose Data Visualization Tools?



Don't try this!



What type of questions are your business users asking?

- *Are you trying to find trends?*
- *Do you even know what the questions are?*
- *How much data are you dealing with (terabytes of transactional data or gigabytes of aggregate data)?*

What technical resources do you have ready access to?

- *Do you have a separate business analyst team?*
- *How big is your user base?*
- *What flavor of database do you have (Relational databases or Big Data)?*
- *Do you have legacy applications that you need to integrate into your new solution?*

How much data do you have and in what condition is it?

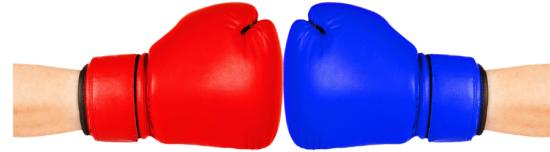
- *Are you going to be reporting on terabytes of transactional data, or do you have data aggregated to a reporting layer?*
- *Does your data need to go through a transformational layer before getting to the reporting users?*
- *Do you have many different data source types (i.e. flat files, databases, web services)?*

The answers to these questions should drive you toward a single tool, or at least help you narrow the field.

Other Factors

- Visualization
- Cost of Ownership
- Integration
- Data Management
- Functional Comparison

Comparison of Tableau and Power BI



Factors	Power BI	Tableau
Visualization	Custom Visualization	Pure Visualization
Cost	Around \$10 per month initial cost, Costly in the long run	Around \$30 per month initial cost, Cost efficient in long run
Integration	Azure, SharePoint, Office365, Power Flow, Power Apps, Excel	Some Options
Data Management (Data Shaping & Modeling)	Query Editor, Flexible, M Language, Data Cube technology	Few Options

Tableau is too expensive. Instead
we spend hundreds of hours trying
to make Excel work somewhat
"Tableau-like."

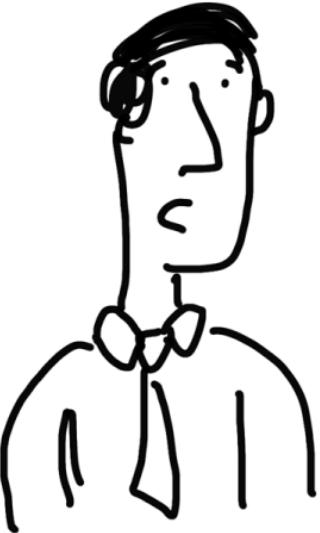


Lysydesign

When to use Excel (not Tableau)

- Less Single user cost
- Less setup time
- Able to utilize VBA, where you can code straightforward macros to control information
- Excel works well for creating quick, one-off reports.
- Compatibility with Microsoft Products





We decided to build
our dashboard in Excel.



You don't know the new
tools, so you decided
to spend lots of time
developing an inferior
product?

When to use Tableau (Not Excel)

- Working with data at moderately large scale (Never use Excel for very large dataset)
- Building interactive visual dashboards
- Faster speed to analysis
- Maps (Tableau instantly recognizes geographical data)

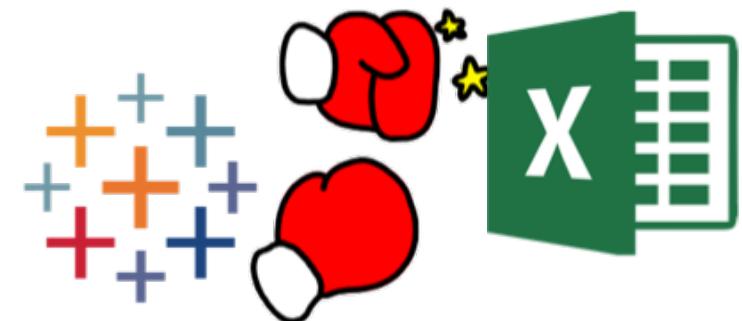


Tableau Products



Tableau Desktop

Called “the gold standard” in visual analytics, Tableau Desktop upended the business intelligence industry and ushered in a new paradigm of self-service insight.

[LEARN MORE →](#)



Tableau Prep

Tableau Prep empowers more people to get to analysis faster by helping them quickly and confidently combine, shape, and clean their data.

[LEARN MORE →](#)



Tableau Online

Want the sharing and collaboration of Server, but without having to actually manage a server? Then you want Tableau Online. Secure. Scalable. And Look Ma—
No hardware to maintain!



Tableau Server

Share your data and dashboards to multiply your impact. Whether you keep your Server deployment on-prem or deploy to the public cloud you can keep the management of your server in your hands.

Link:

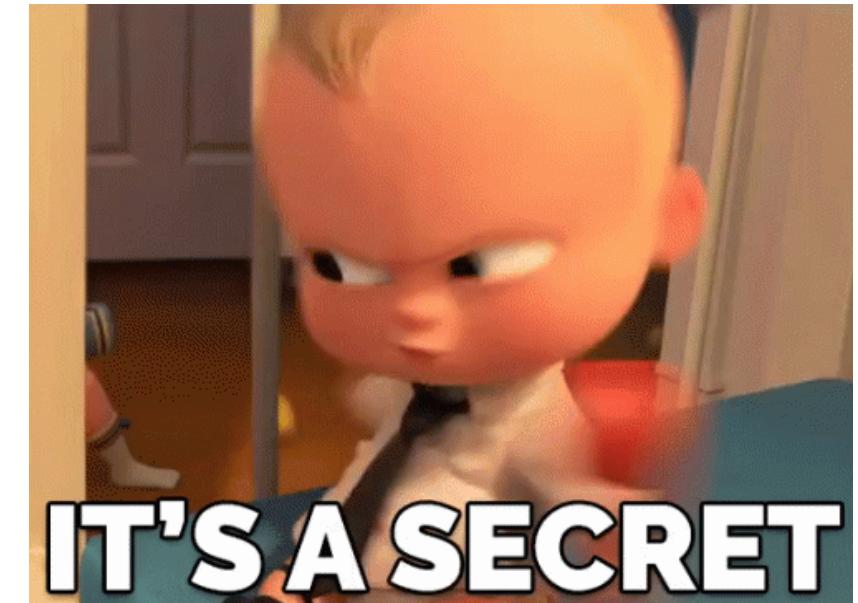
<https://www.tableau.com/products>

Tableau Installation

- Students can get 1 year free license for Tableau for Desktop and Tableau Prep.

Link to Download:

<https://www.tableau.com/academic/students>



Video: <https://www.youtube.com/watch?v=GMqHKMHm6zM>

Data Connectors

Connect

Connect

To a File

Microsoft Excel

Text file

JSON file

PDF file

Spatial file

Statistical file

More...

To a Server

Tableau Server

Microsoft SQL Server

MySQL

Oracle

Amazon Redshift

More... >

Saved Data Sources

Sample - Superstore

World Indicators

Search

Tableau Server

Amazon Athena

Amazon Aurora

Amazon EMR

Amazon Redshift

Anaplan

Apache Drill

Aster Database

Box

Cloudera Hadoop

Denodo

Dropbox

EXASOL

Firebird

Google Analytics

Google BigQuery

Google Cloud SQL

Google Sheets

Hortonworks Hadoop Hive

HP Vertica

Kognitio

MapR Hadoop Hive

Marketo

MemSQL

Microsoft SQL Server

MongoDB BI Connector

MySQL

OData

Aster Database

OneDrive

Oracle

Oracle Eloqua

Pivotal Greenplum Database

PostgreSQL

Presto

QuickBooks Online

Salesforce

SAP HANA

ServiceNow ITSM

SharePoint Lists

Snowflake

Spark SQL

Web Data Connector

Other Databases (ODBC)

Connect to Microsoft SQL Server

The screenshot shows the Tableau Connect interface. On the left, a sidebar lists connection options: 'To a File' (Microsoft Excel, Text file, JSON file, PDF file, Spatial file, Statistical file, More...), 'To a Server' (Tableau Server, Microsoft SQL Server, MySQL, Oracle, Amazon Redshift, More...), and 'Saved Data Sources' (Sample - Superstore, World Indicators). A search bar is at the top. In the center, a modal dialog for 'Microsoft SQL Server' is open. It has fields for 'Server:' and 'Database:' (labeled 'Optional'). Below these are instructions: 'Enter information to sign in to the database:' followed by two radio button options: 'Use Windows Authentication (preferred)' (selected) and 'Use a specific username and password'. There are also optional checkboxes for 'Require SSL' and 'Read uncommitted data'. At the bottom of the dialog are 'Initial SQL...' and 'Sign In' buttons. The background shows other available connectors like Tableau Server, Amazon Aurora, and Google Sheets.

Connect

To a File

- Microsoft Excel
- Text file
- JSON file
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

- Tableau Server
- Microsoft SQL Server
- MySQL
- Oracle
- Amazon Redshift
- More... >

Saved Data Sources

- Sample - Superstore
- World Indicators

Search

Microsoft SQL Server

Server: _____

Database: Optional

Enter information to sign in to the database:

Use Windows Authentication (preferred)

Use a specific username and password

Username: domain\username -- Optional

Password: Optional

Require SSL

Read uncommitted data

Initial SQL...

Sign In

Tableau Server

Amazon Aurora

Amazon EMR

Amazon Redshift

Anaplan

Apache Drill

Aster Database

Box

Cloudera HDFS

Denodo

Dropbox

EXASOL

Firebird

Google Analytics

Google BigQuery

Google Cloud Storage

Google Sheets

Hortonworks Hadoop Hive

HP Vertica

Kognitio

Snowflake

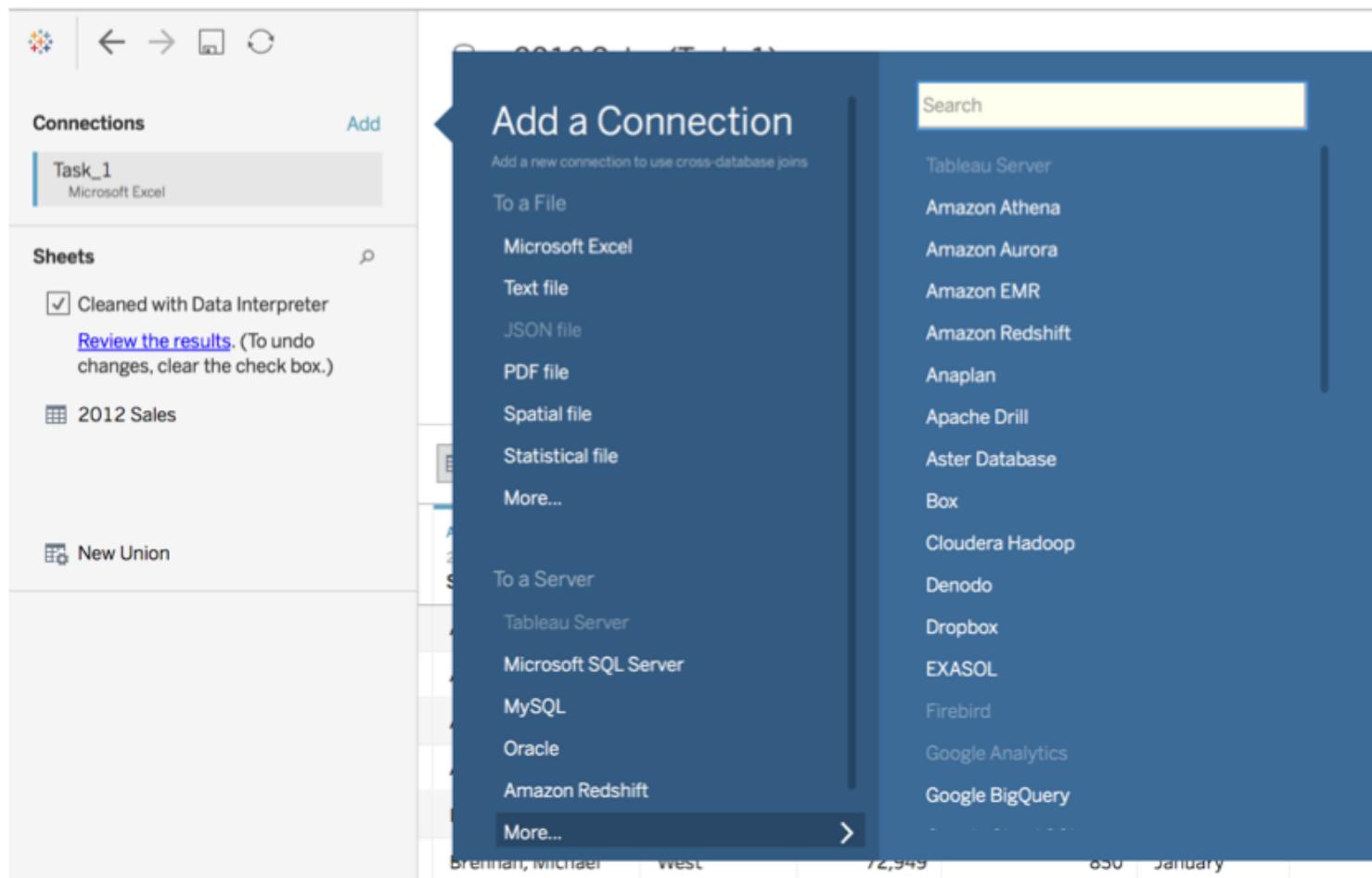
Spark SQL

Teradata

Web Data Connector

Other Databases (ODBC)

Add More than one Connection



Add More than one Connection

The screenshot shows the Power BI desktop interface. On the left, the 'Connections' pane is open, displaying two connections: 'Task_1' (Microsoft Excel) and 'Task_2' (Microsoft Excel). A red box highlights this pane. Below it, the 'Sheets' pane lists 'Customer Info', 'Order Info', 'Payment Info', and 'New Union'. In the center, under the heading '2012 Sales (Multiple Connections)', there is a data view titled '2012 Sales'. At the top of this view are buttons for 'Sort fields' and 'Data source order'. The data table has columns: Salesperson, Region, Account, Order Amount, and Month. The first three rows of data are:

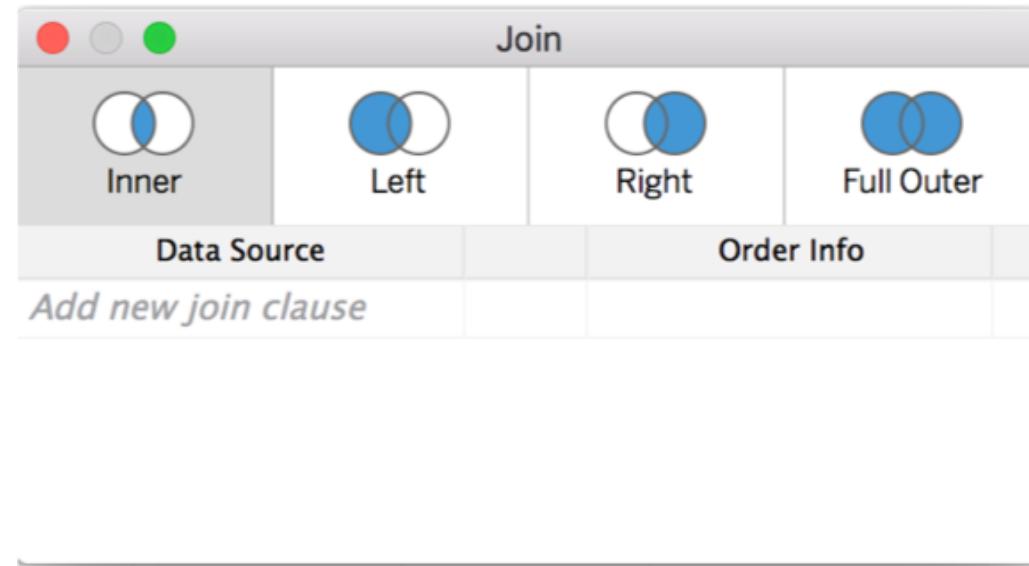
Salesperson	Region	Account	Order Amount	Month
Albertson, Kathy	East	29,386	925	January
Albertson, Kathy	East	74,830	875	February
Albertson, Kathy	East	90,099	500	February

Load and Prepare Data

- Dataset: Fruit Wholesalers (Available on Brightspace)
- Video: <https://www.youtube.com/watch?v=ApyfHmsVJHA>

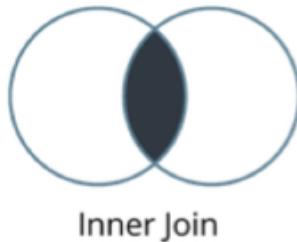
Overview of Join Types

- Inner Join
- Left Join
- Right Join
- Full Outer

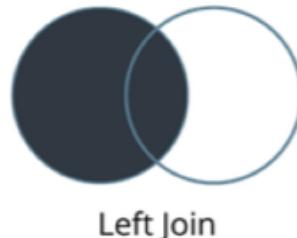


Overview of Join Types

- **Inner Join:** When you use an inner join to combine tables, the result is a table that contains values that have matches in both tables.

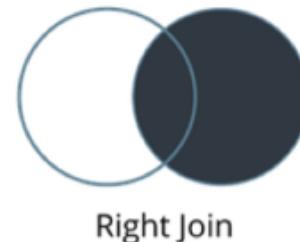


- **Left Join:** When you use a left join to combine tables, the result is a table that contains all values from the left table and corresponding matches from the right table.

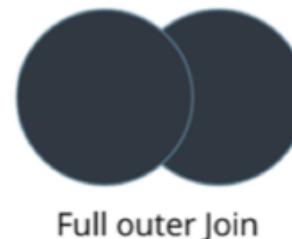


Overview of Join Types

- **Right Join:** When you use a right join to combine tables, the result is a table that contains all values from the right table and corresponding matches from the left table.



- **Full Outer Join:** When you use a full outer join to combine tables, the result is a table that contains all values from both tables.



Join Calculation

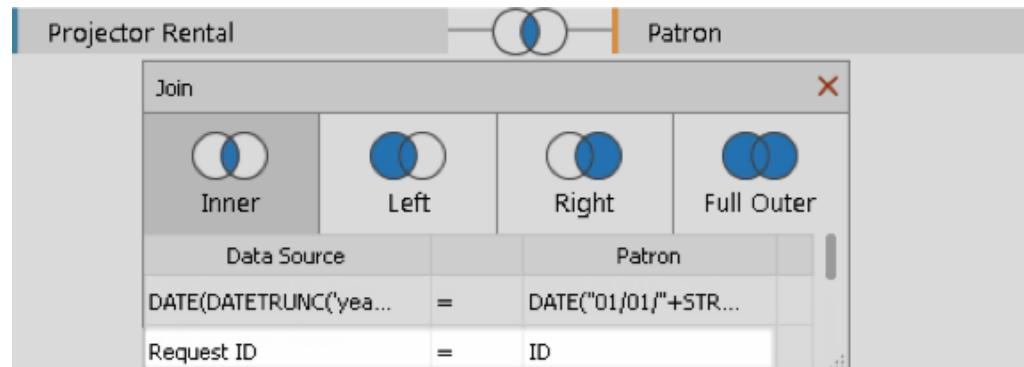
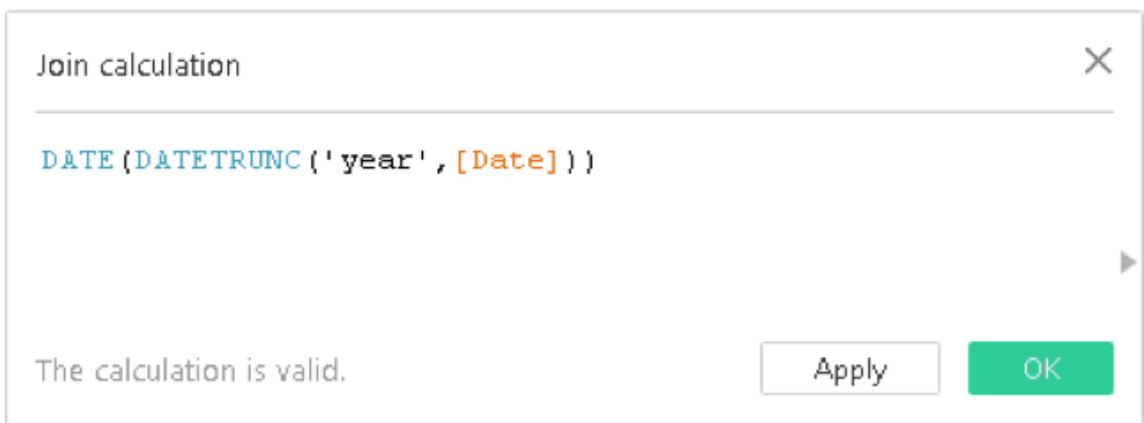
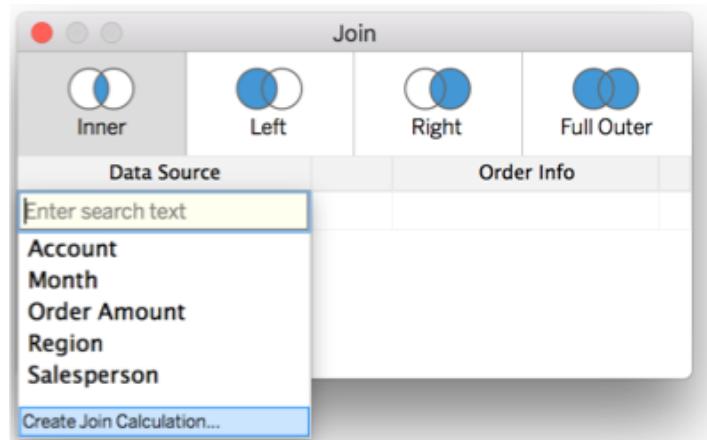
Projector rental

Date	Reservation type	Requester ID
1/1/2000	Individual	233445589
1/28/2002	K-12	365948999
1/29/2002	Non-profit	233448888
12/5/2002	K-12	365948999
5/5/2003	Non-profit	334015476
3/12/2004	Non-profit	334015476
3/15/2006	City	211896980
7/8/2007	K-12	334015476
1/4/2008	Individual	560495523
3/8/2009	Non-profit	233445566
2/14/2014	Non-profit	233445566
12/21/2015	Non-profit	233445566
2/10/2016	Non-profit	233445566

Patron

ID	First name	Last name	Branch	Member since	Units borrowed	Fees	Suggested limit
454613981	Adam	Davis	West	2012	25	0	10
232502870	Alan	Wang	North	2000	1	0	15
298000916	Amanda	Smith	North	2001	54	6.4	15
233448978	Andrew	Smith	North	2000	36	3.50	15
233445566	Ashley	Garcia	South	2000	243	11.30	15
900005122	Brian	Frank	East	2011	12	.10	10
921491769	Elizabeth	Johnson	West	2010	19	.5	10
233445589	Fred	Suzuki	North	2000	52	.90	15
344556677	Henry	Wilson	South	2005	3	.2	15
939502870	Jane	Johnson	West	2017	0	0	10

Join Calculation





Different Graphs in Tableau

Video: <https://www.youtube.com/watch?v=8iYpxiipNHQ>

At the dataviz expert ugly sweater party.

Did we all go with
pie charts again?

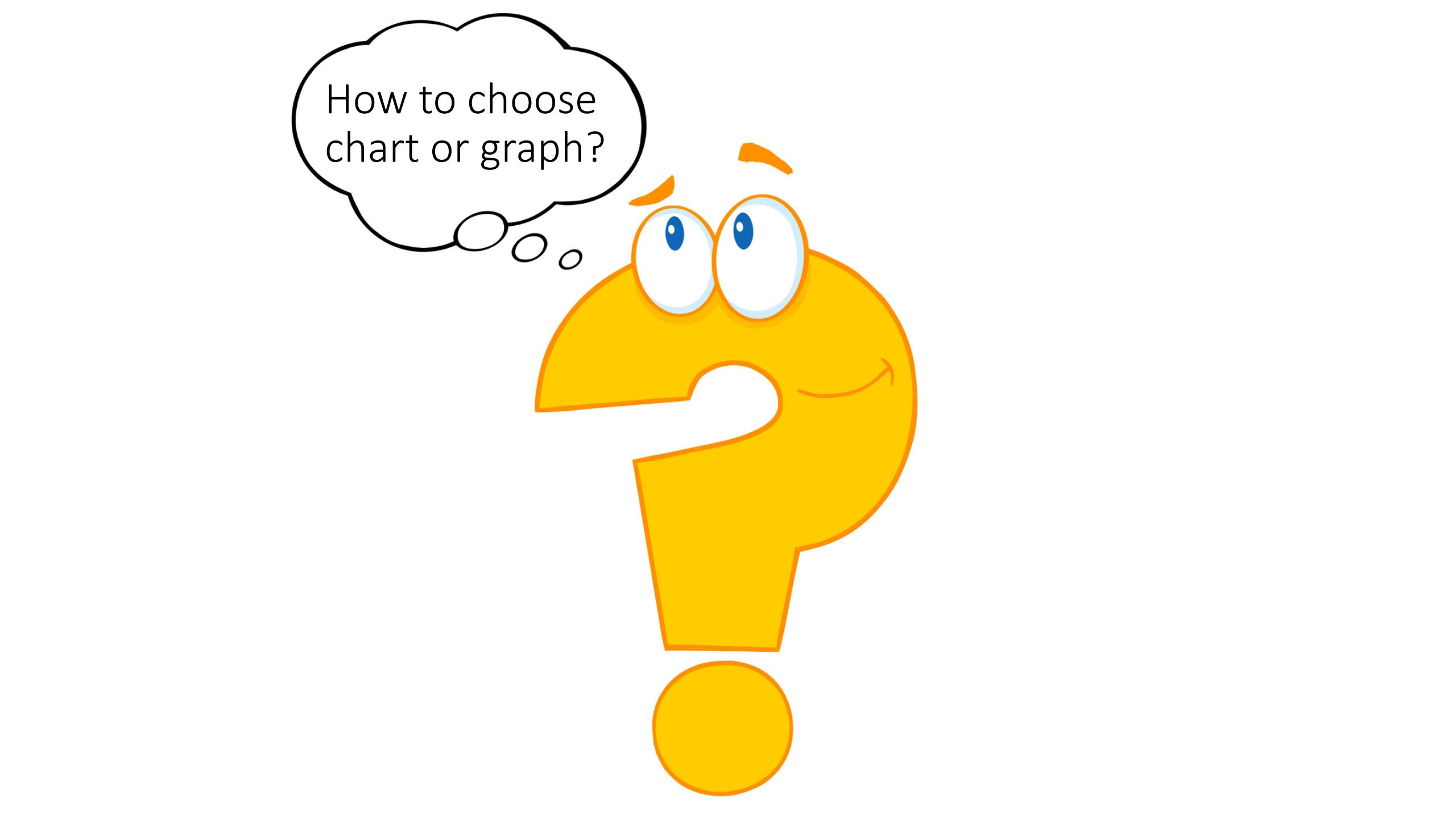


Types of charts or graphs in Tableau

- Bar Chart
- Line Chart
- Pie Chart
- Map
- Scatter Plot
- Bubble Chart
- Histogram Chart
- Bullet Chart
- Heat Maps
- Highlight Table
- Treemap
- Box-and-whisker Plot



SO MANY CHOICES...



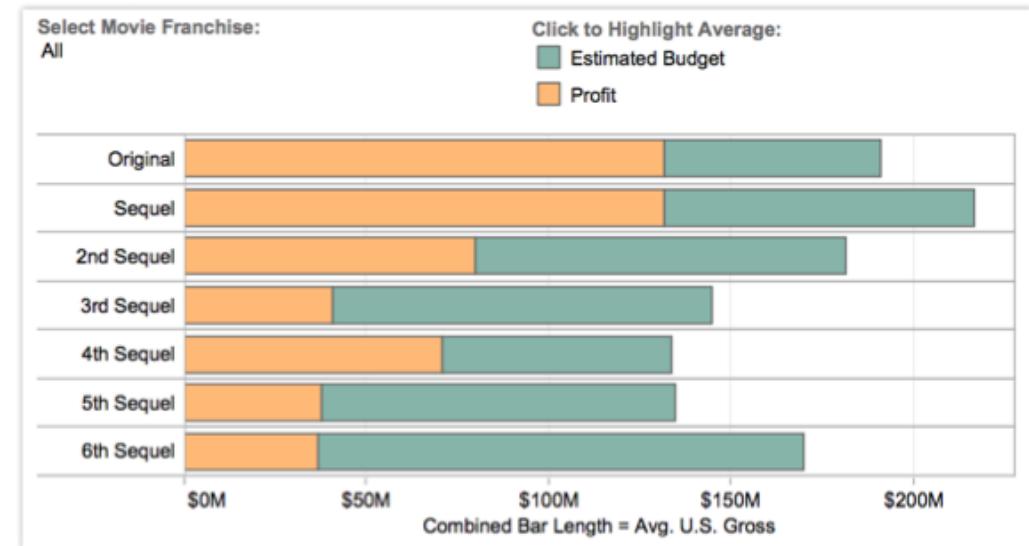
How to choose
chart or graph?

Bar Chart

- Comparing data across categories.

Examples:

- Volume of shirts in different sizes
- Website traffic by origination site
- Percent of spending by department

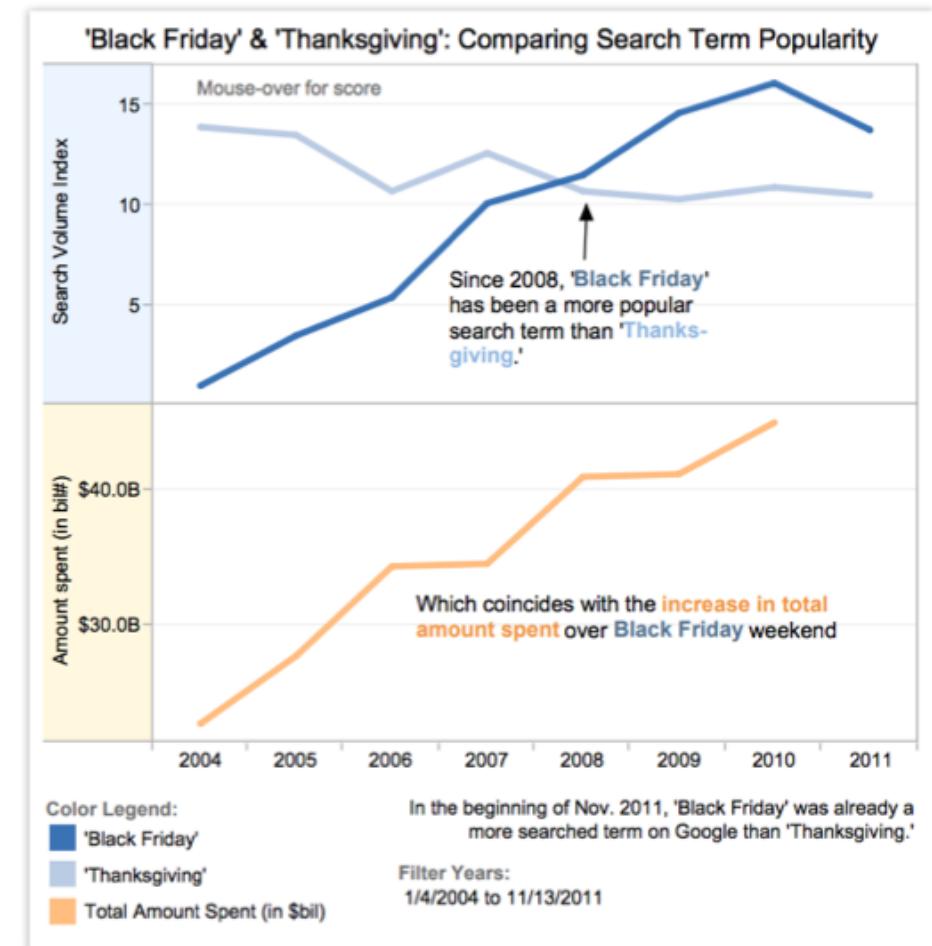


Line Chart

- Viewing trends in data over time.

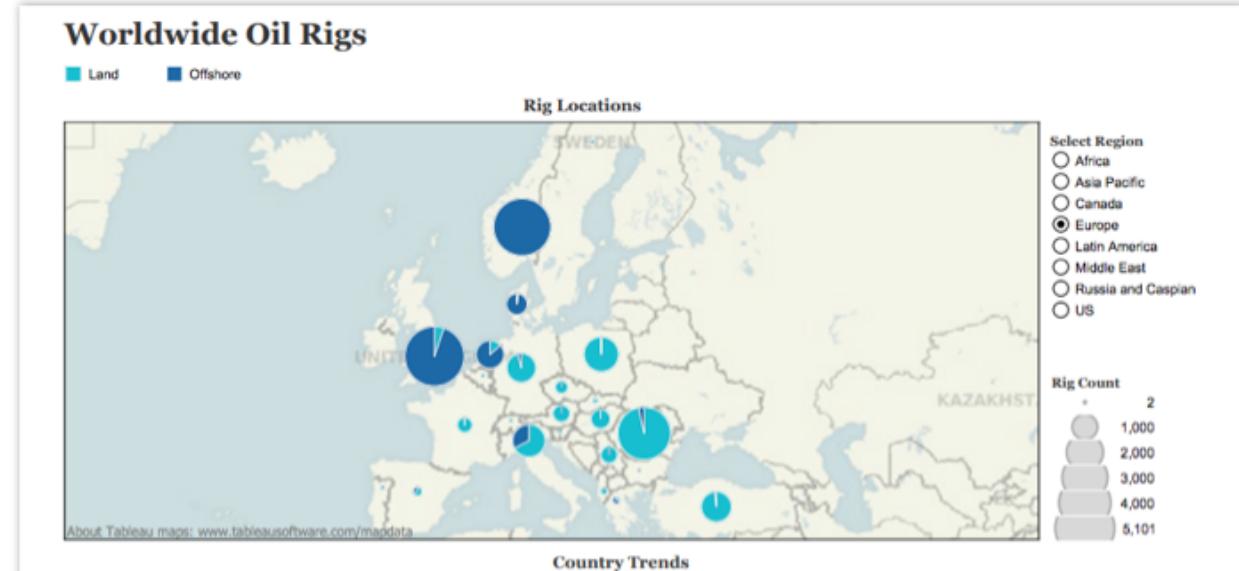
Examples:

- stock price change over a five year period
- website page views during a month
- revenue growth by quarter



Pie Chart

- Showing proportions.



Examples:

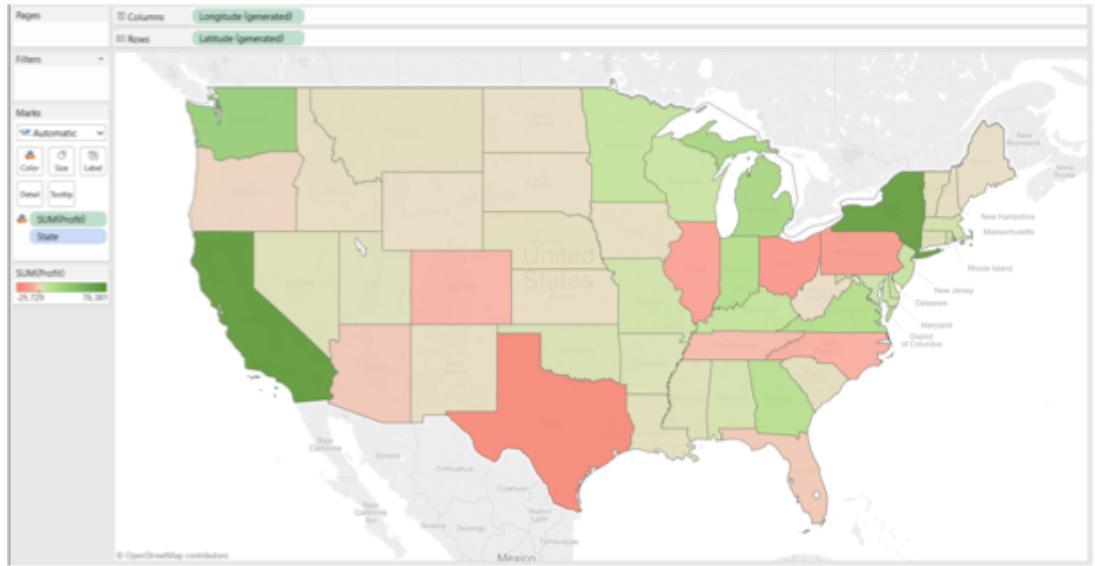
- Percentage of budget spent on different departments
- Response categories from a survey
- Breakdown of how Americans spend their leisure time

Map

- Showing geocoded data.

Examples:

- Insurance claims by state
- Product export destinations by country
- Car accidents by zip code
- Custom sales territories

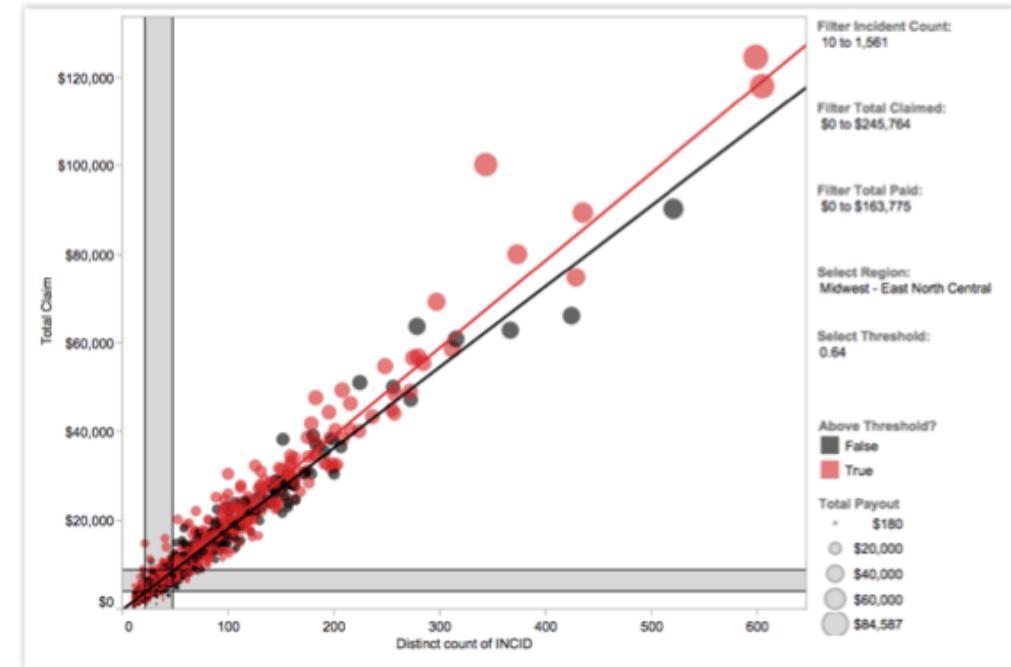


Scatter plot

- Investigating the relationship between different variables.

Example:

- Male versus female likelihood of having lung cancer at different ages
- Technology early adopters' and laggards' purchase patterns of smart phones
- Shipping costs of different product categories to different regions.

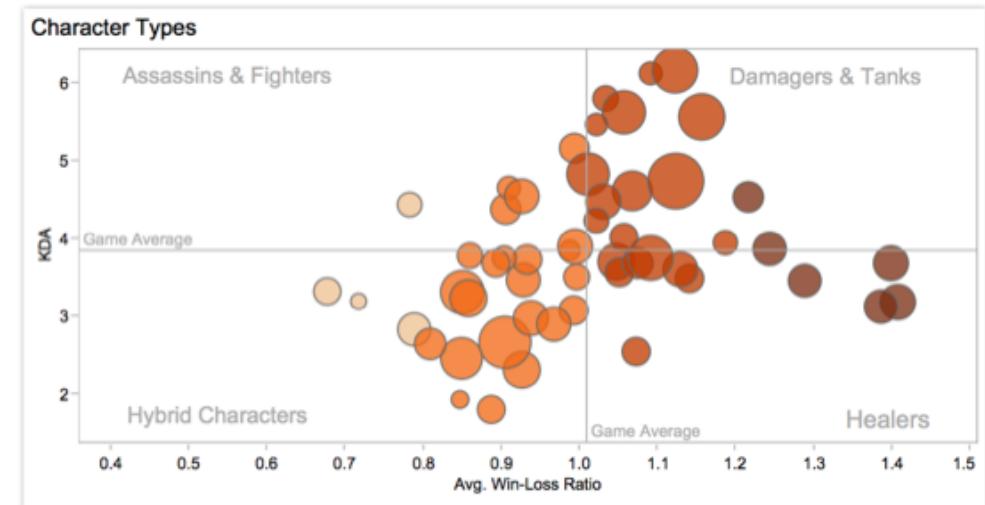


Bubble chart

- Showing the concentration of data along two axes.

Examples:

- Sales concentration by product and geography
- Class attendance by department
- Time of day

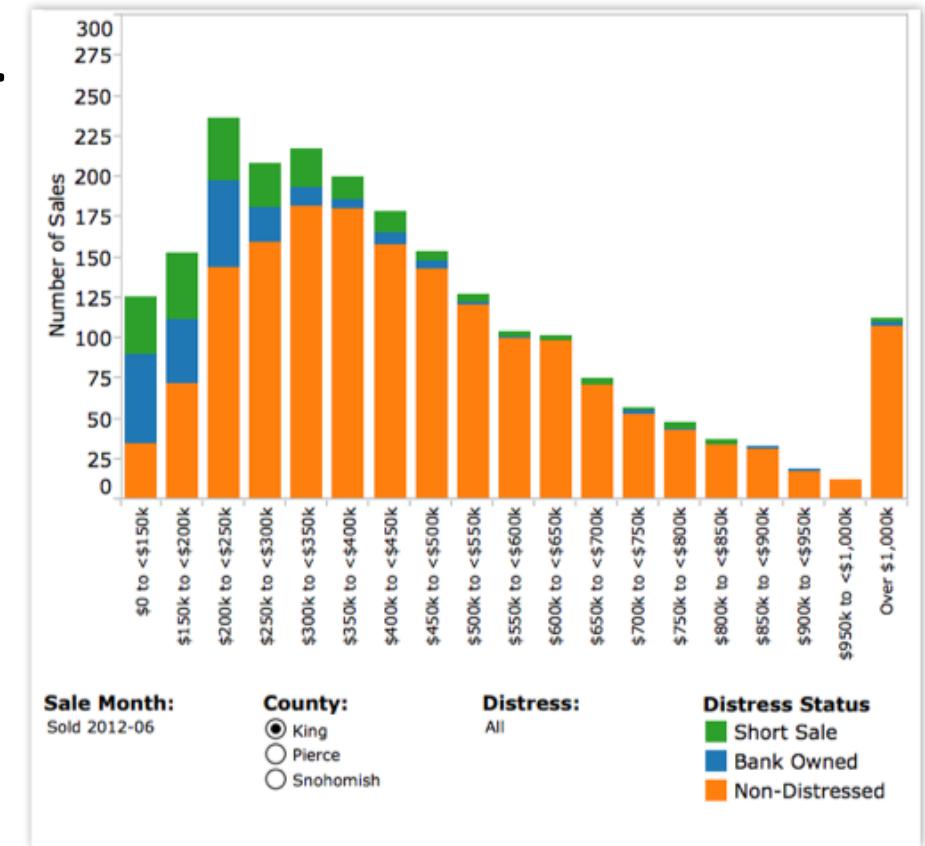


Histogram chart

- Understanding the distribution of your data.

Examples:

- Number of customers by company size
- Student performance on an exam
- Frequency of a product defect

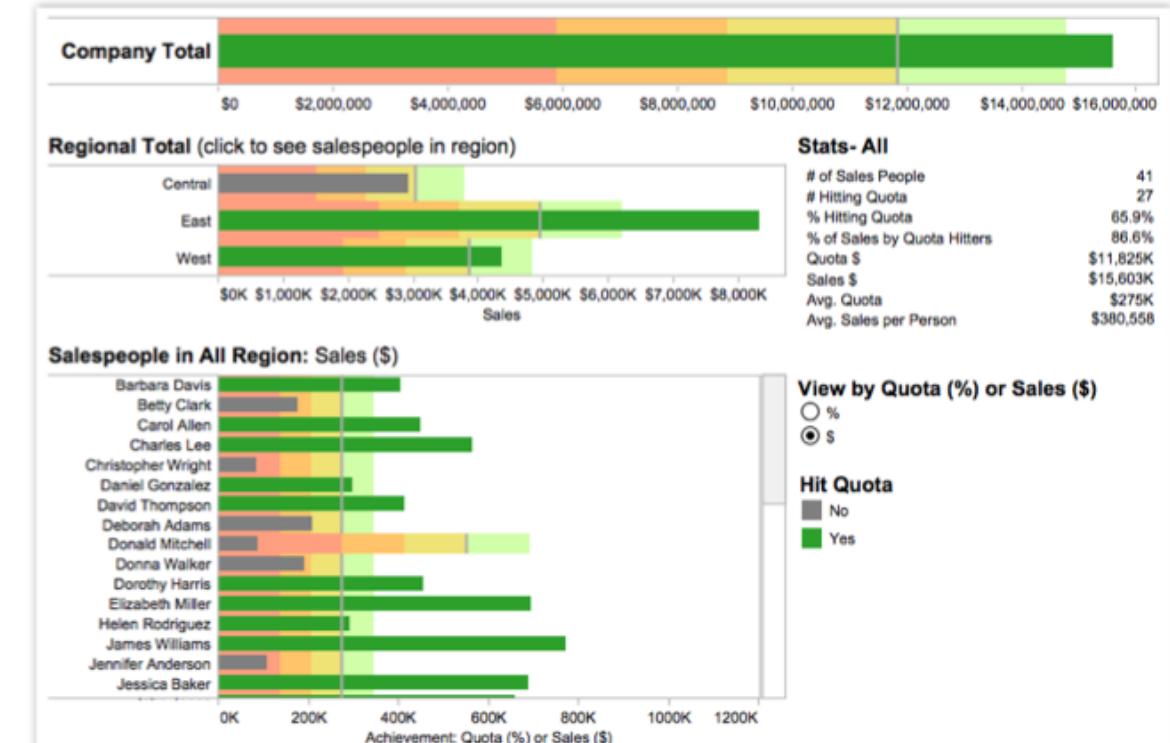


Bullet chart

- Evaluating performance of a metric against a goal

Examples:

- Sales quota assessment
- Actual spending vs. budget
- performance spectrum (great/good/poor).



Heat maps

- Showing the relationship between two factors.

Examples:

- Segmentation analysis of target market
- Product adoption across regions
- Sales leads by individual rep.

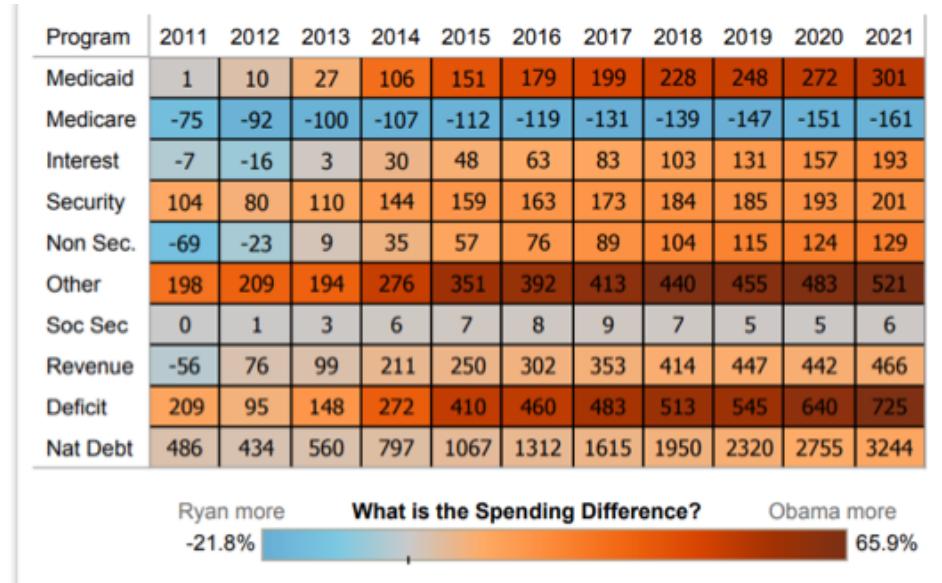


Highlight table

- Providing detailed information on heat maps.

Examples:

- The percent of a market for different segments
- Sales numbers by a reps in a particular region
- Population of cities in different years

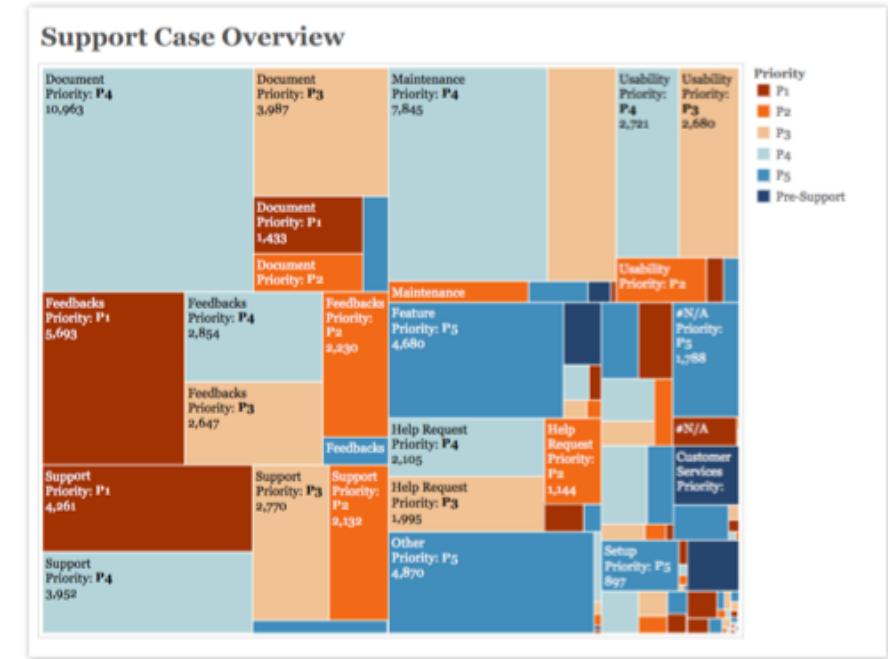


Treemap

- Showing hierarchical data as a proportion of a whole.

Examples:

- Storage usage across computer machines
- Managing the number and priority of technical support cases
- Comparing fiscal budgets between years

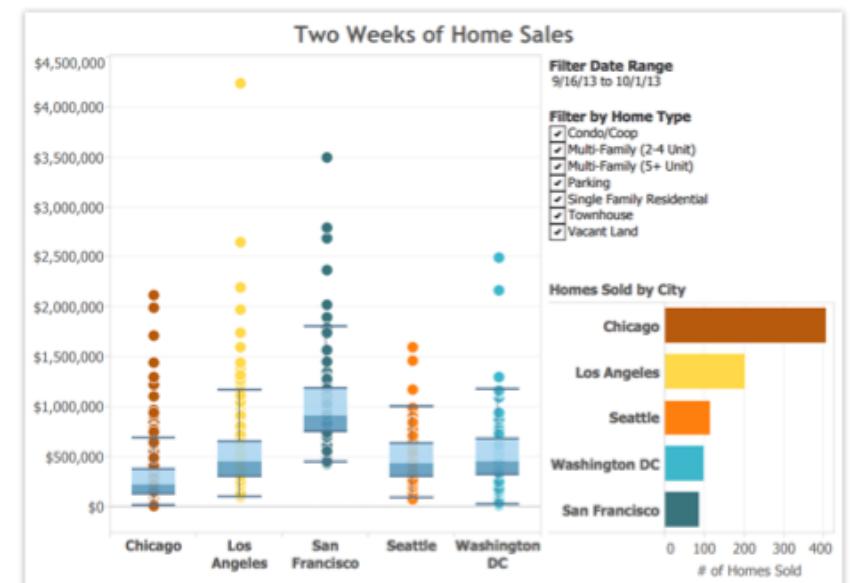


Box-and-whisker Plot

- Showing the distribution of a set of a data

Examples:

- Understanding your data at a glance
- Seeing how data is skewed towards one end
- Identifying outliers in your data.

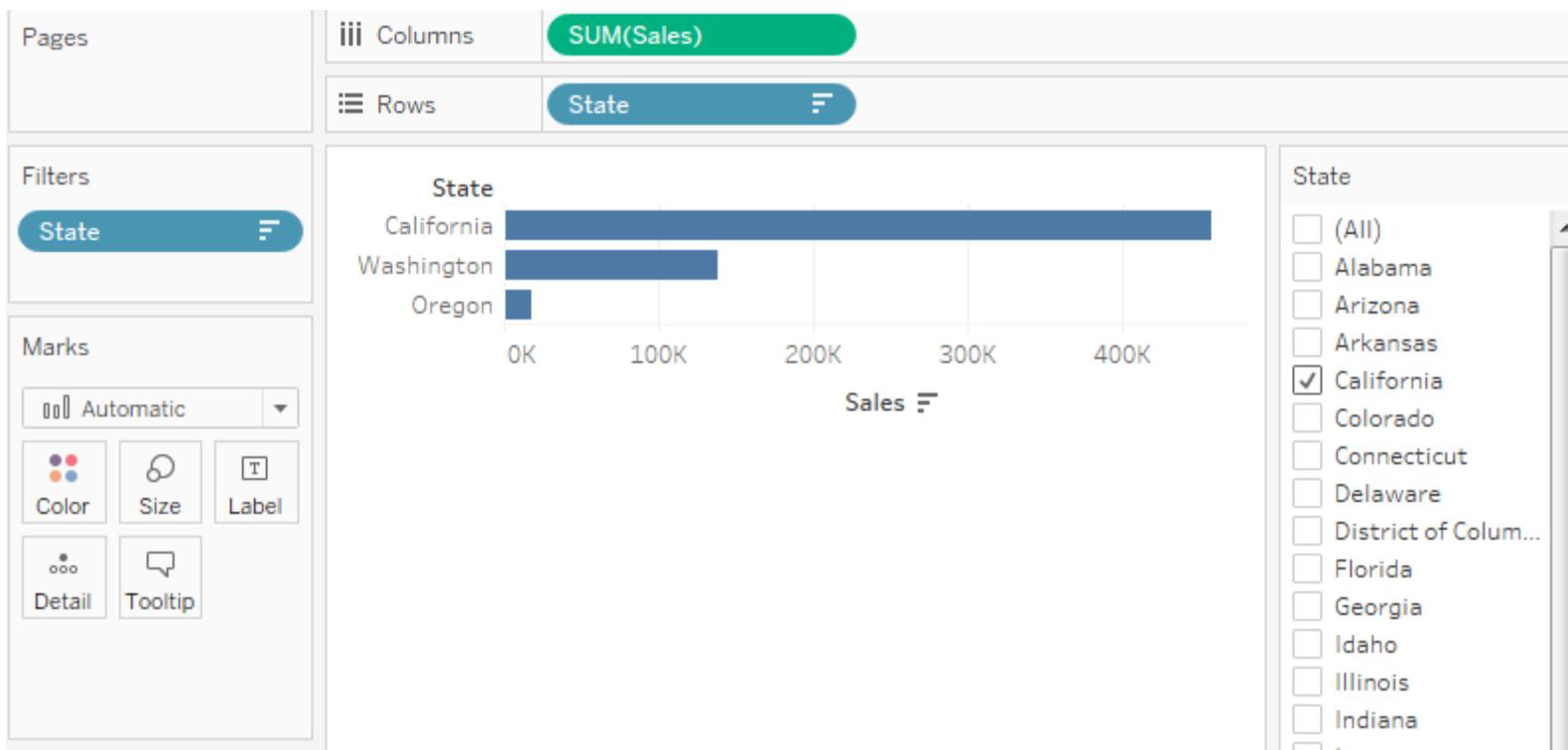


Question!

- What is the difference between Heat map and Treemap

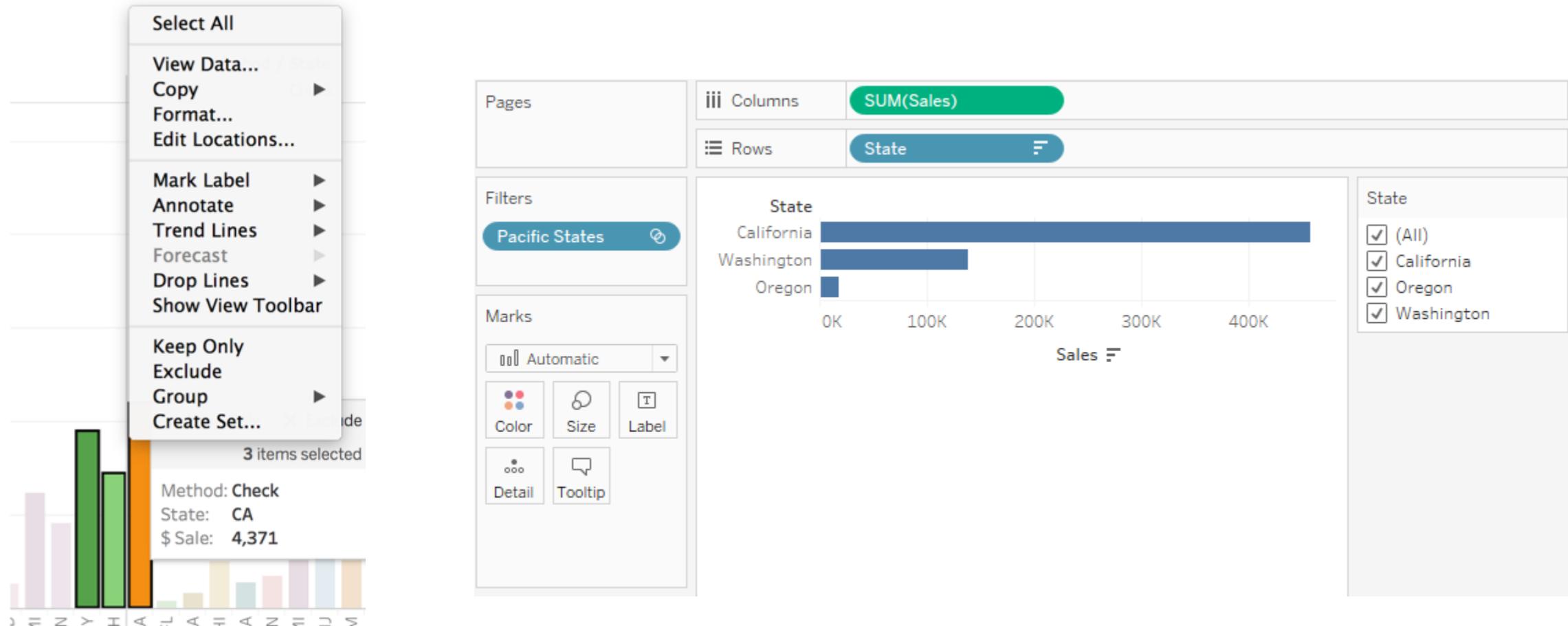
Filters in Tableau

- Filters reduce the data that is visible either in our view or in our workbook depending on where your filters are applied.



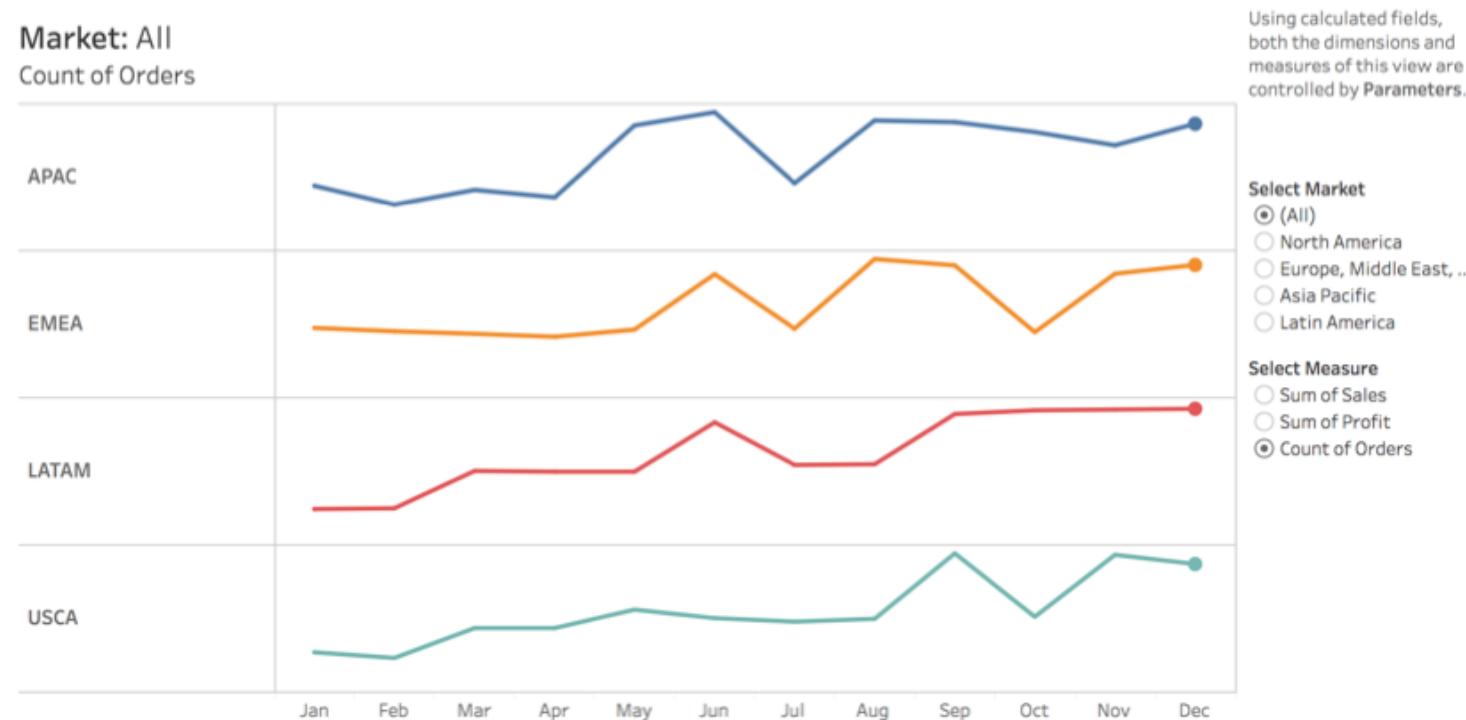
Sets in Tableau

- Within my set, I will only choose states border the Pacific Ocean. I'll name my new set "**Pacific States**" and then apply it as a filter to get the following result



Parameters in Tableau

- Users can choose and change the value of the parameter similar in functionality and appearance to a filter.
- Parameter controls are interactive and allow the report consumer to affect your visualization.

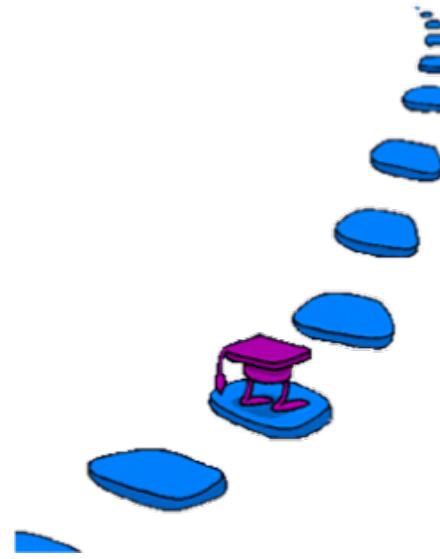


Parameter Values can be:

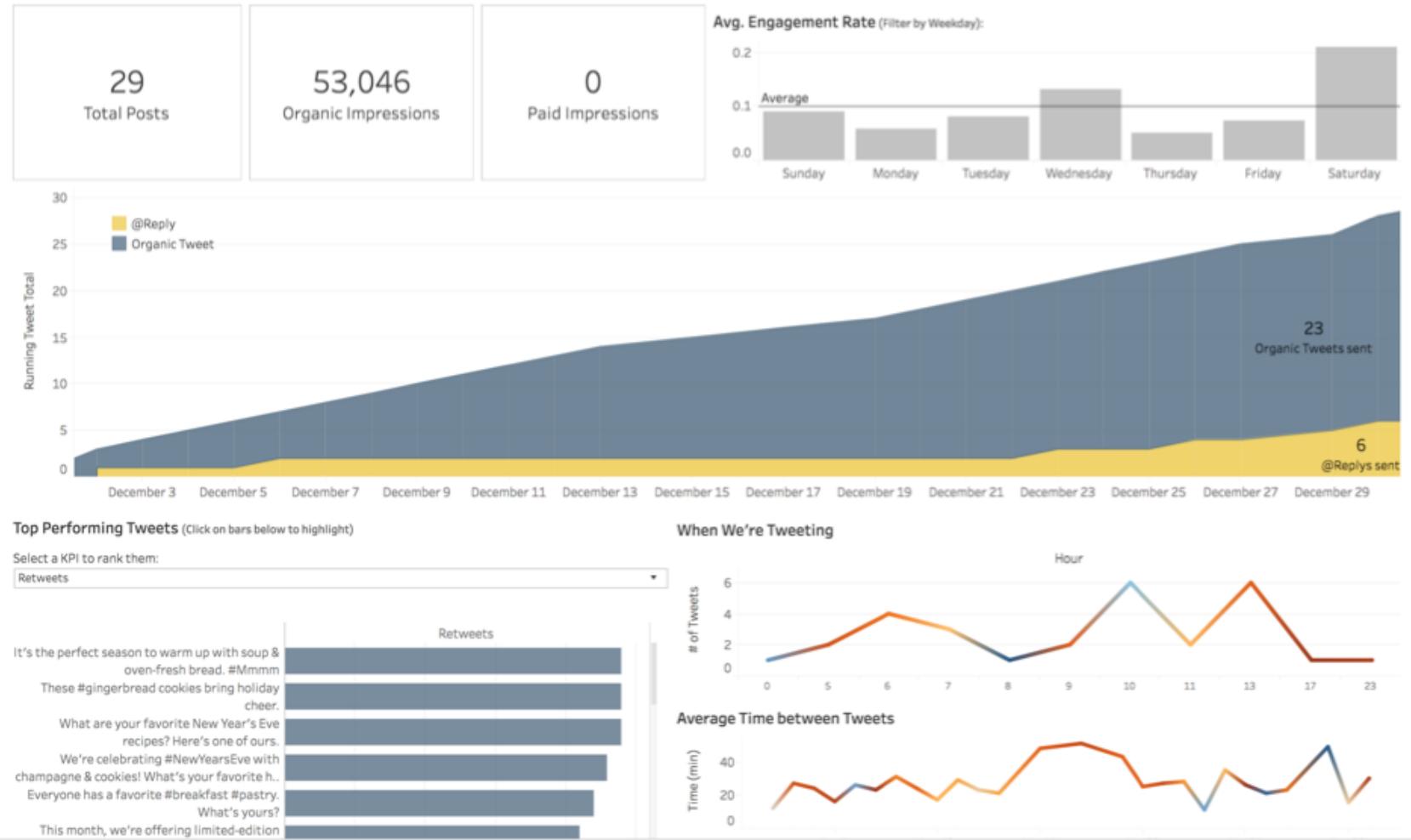
- Defined by the Desktop user
- Entered as an input by the report consumer
- Populated with the values of a field from the data source

Steps to create parameter

- Build Your Parameter
- Show Parameter Control
- Use Your Parameter in Calculation
- Use Your Calculation in Your View



Twitter Dashboard



Link: <https://www.tableau.com/solutions/workbook/get-twitter-analytics-you-need-about-your-own-brand-or-build-interactive-viral-content>

Tableau Twitter Connector



- Video: <https://youtu.be/PDct4VBo45Q>
- Tableau Twitter Connector Link:
<http://files.tableaujunkie.com/twittersearch/twitterwebconnect.html>

If you want to create your own twitter Web Data connector

- Reference Link:

<http://tableaujunkie.com/post/119681578798/creating-a-twitter-web-data-connector>

Analyze Facebook Data



- Use data of your page (if you have) otherwise download it from Brightspace.
- Video:<https://youtu.be/VAbiRcdIBk>