Data Visualization with Tableau – Level 1

QCL Workshop Series

#Data #Visualization and #Tableau

Instructor: Dr. Jeho Park, Director, Murty Sunak QCL

Make sure you are signed-in

https://bit.ly/3gJrYhl

After this workshop, you will be able to…

- Understand fundamentals of data visualization
- Connect various data sources to Tableau
- Create different worksheets to visualize various insights of your data
- Create Tableau Dashboard showing meaningful data visualization
- Share Tableau Dashboard on Tableau Public/Online
- (Stretch Topics) Filtering, Calculation, Analytics

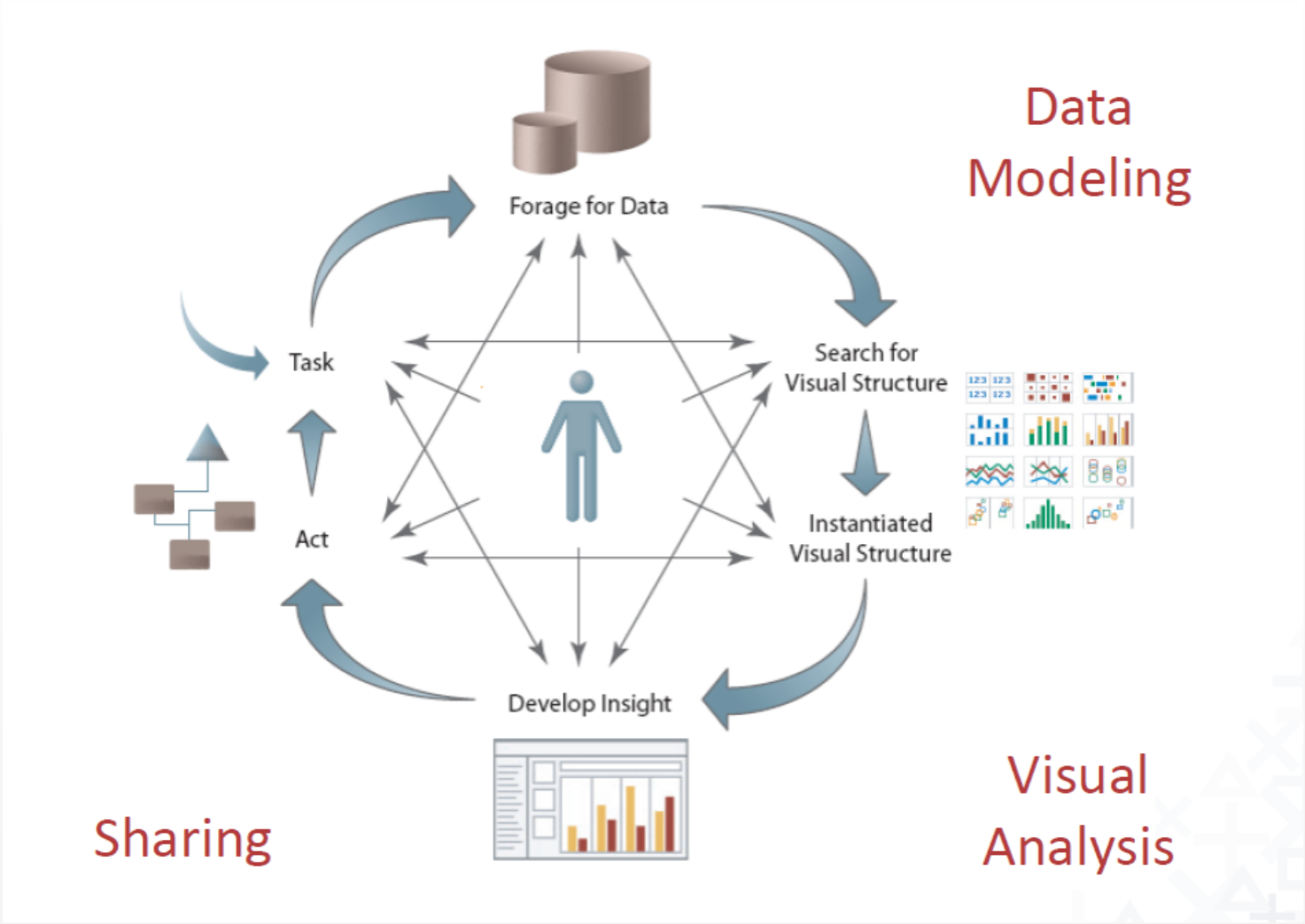
What is Tableau?

- Tableau is one of the fastest evolving business intelligence and data visualization tools.
- Tableau software was founded in 2003

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms

Source: Gartner (February 2021)

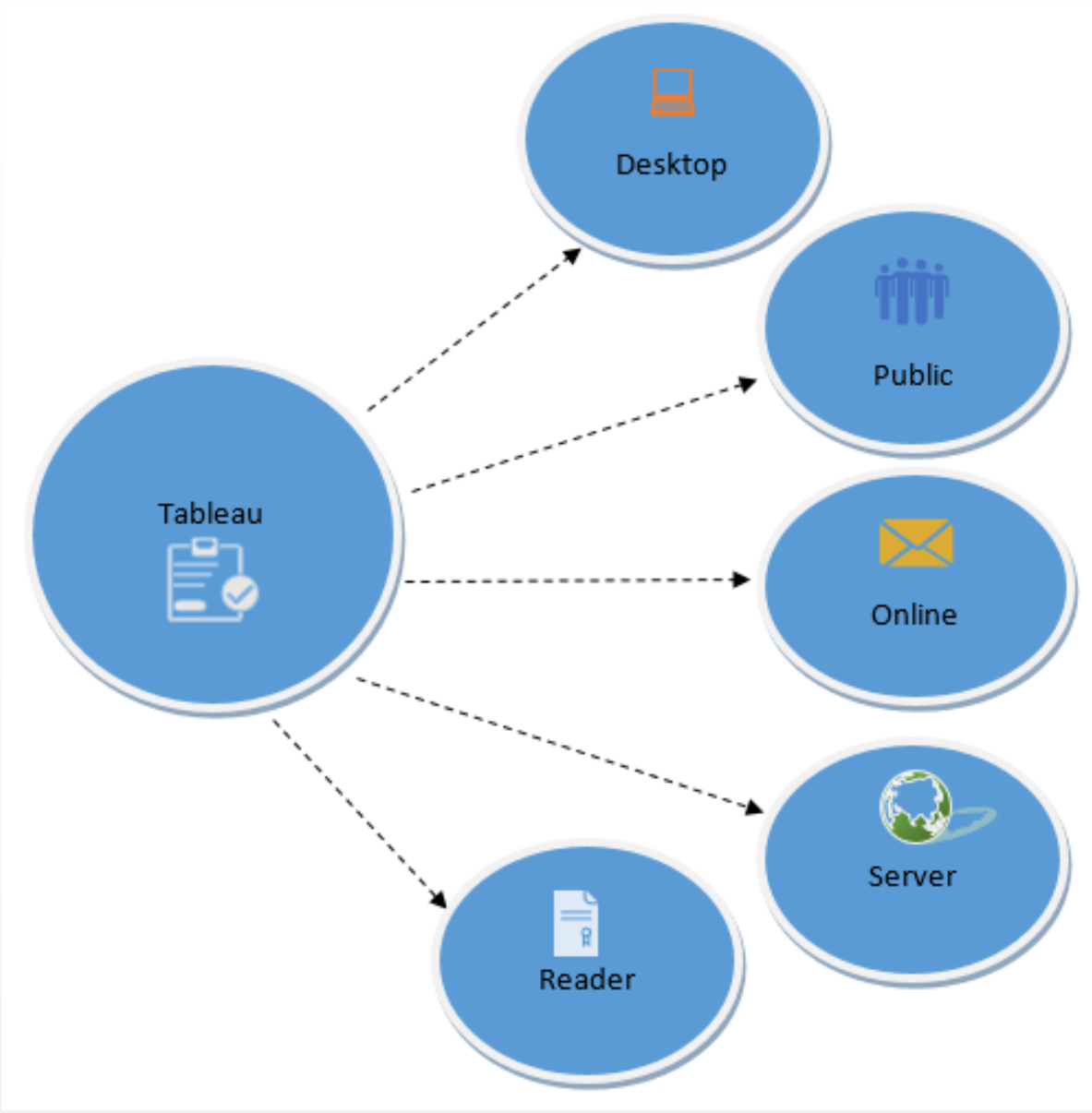
Cycle of Visual Analytics



Source: Dynamic Workload Driven Data Integration in Tableau (Proceedings of the ACM SIGMOD International Conference on Management of Data)

Tableau Product Suite

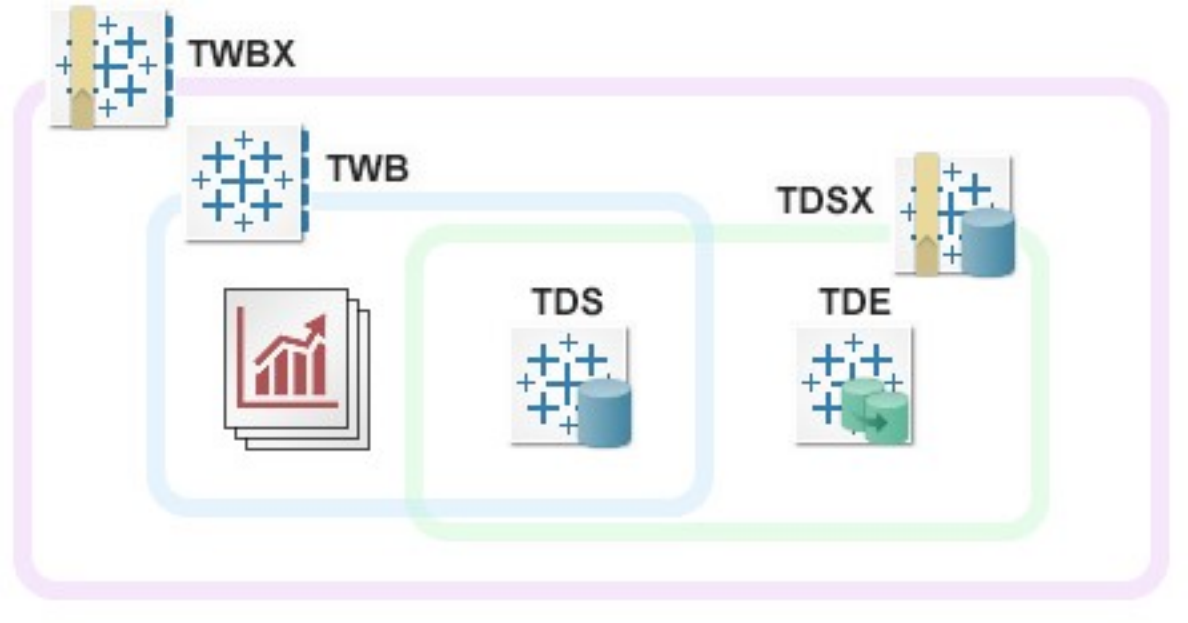
- Tableau Desktop
- Tableau Public
- Tableau Online
- Tableau Server
- Tableau Reader



Source: guru99.com

Tableau File Types

TWB	TWBX	TDE
Tableau Workbook File	Tableau Packaged Workbook	Tableau Data Extract
XML file with visualization	Zip file	Compressed data sources
Does not conatin data & Cannot open data files	Contains TWB file & data	



Source: https://interworks.com/blog/rcurtis/2016/09/15/alphabet-soup-understanding-tableau-10-file-extensions/

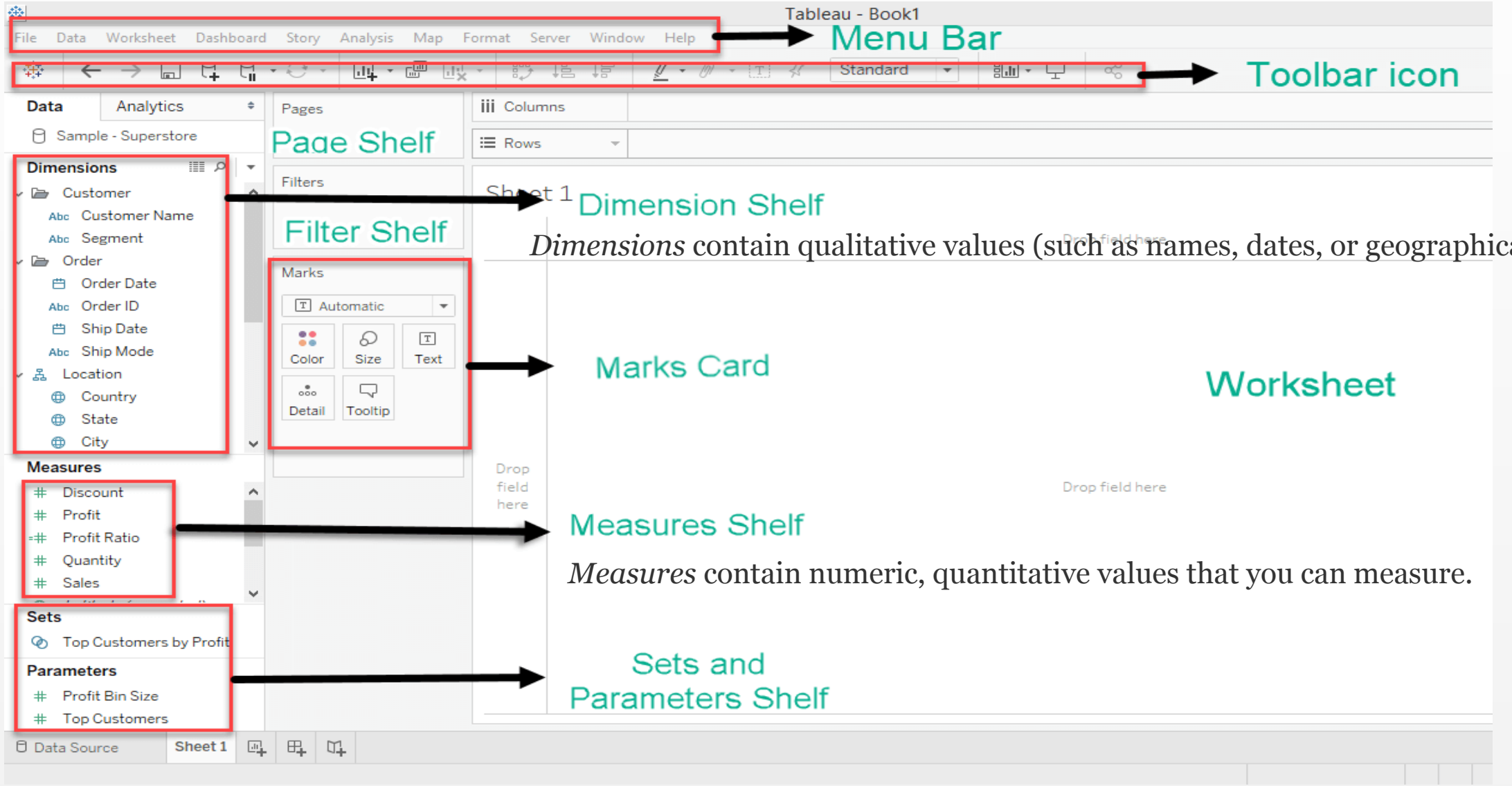
What can you do with Tableau?

https://public.tableau.com/en-us/s/gallery/analyzing-ums?gallery=votd

Installing Tableau

- Install the free trial of Tableau here:
  - https://www.tableau.com/products/trial
- To continue using Tableau after the trial students and instructors can apply for a free license here:
  - https://www.tableau.com/academic

Tableau Desktop Workspace



Dimensions contain qualitative values (such as names, dates, or geographical data).

Measures contain numeric, quantitative values that you can measure.

Tableau Desktop Workspace

Dimensions

contain qualitative values (such as names, dates, or geographical data).

Measures

contain numeric, quantitative values that you can measure.

Tableau Desktop Workspace Cont’d

- **Menu Bar:** It consists of menu options such as File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Server, and Windows. The options in the menu bar include features such as file saving, data source connection, file export, table calculation options, and design features for creating a worksheet, dashboard, and storyboard.
- **Toolbar Icon:** Toolbar icon present below the menu bar can be used to edit the workbook using different features such as undo, redo, save, new data source, slideshow and so on.
- **Dimension Shelf:** The dimensions present in the data source can be viewed in the dimension shelf.
- **Measure Shelf:** The measures present in the data source can be viewed on the measure shelf.
- **Sets and Parameters Shelf:** The user-defined sets and parameters can be viewed in the sets and parameter shelf. It can also be used to edit the existing sets and parameters.
- **Page Shelf:** Page shelf can be used to view the visualization in video format by keeping the relevant filter on the page shelf.
- **Filter Shelf:** The filters that can control the visualization can be placed on the filter shelf, and the required dimensions or measures can be filtered in.
- **Marks Card:** Marks card can be used to design the visualization. The data components of the visualization such as color, size, shape, path, label, and tooltip used in the visualizations can be modified in the marks card.
- **Worksheet:** The worksheet is the place where the actual visualization can be viewed in the workbook. The design and functionalities of the visual can be viewed in the worksheet.

Tableau Navigation

Current Sheet: Current Sheet can be viewed with the name of the sheet. All the sheets, dashboards and story board present in the workbook can be viewed here.

New Sheet: The new sheet icon present in the tab can be used to create a new worksheet in the Tableau Workbook.

New Dashboard: The new dashboard icon present in the tab can be used to create a new dashboard in the Tableau Workbook.

New Storyboard: The new storyboard icon present in the tab can be used to create new storyboard in the Tableau Workbook.

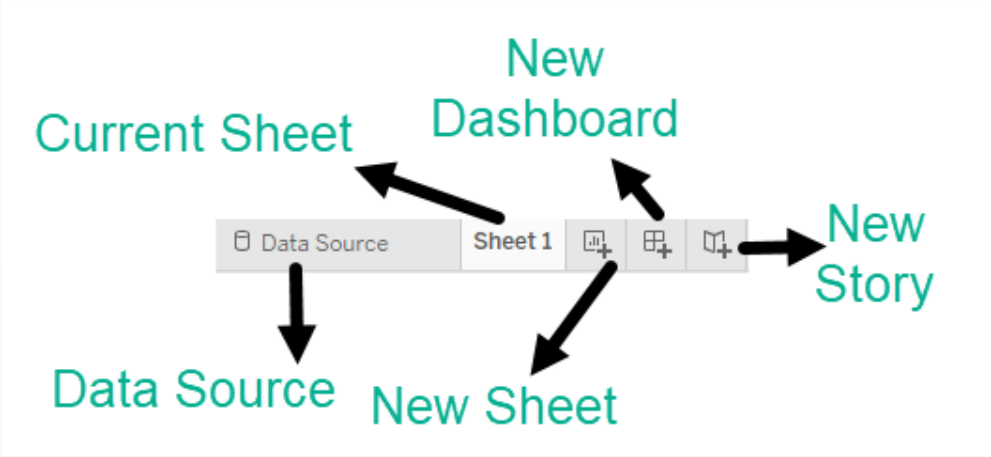
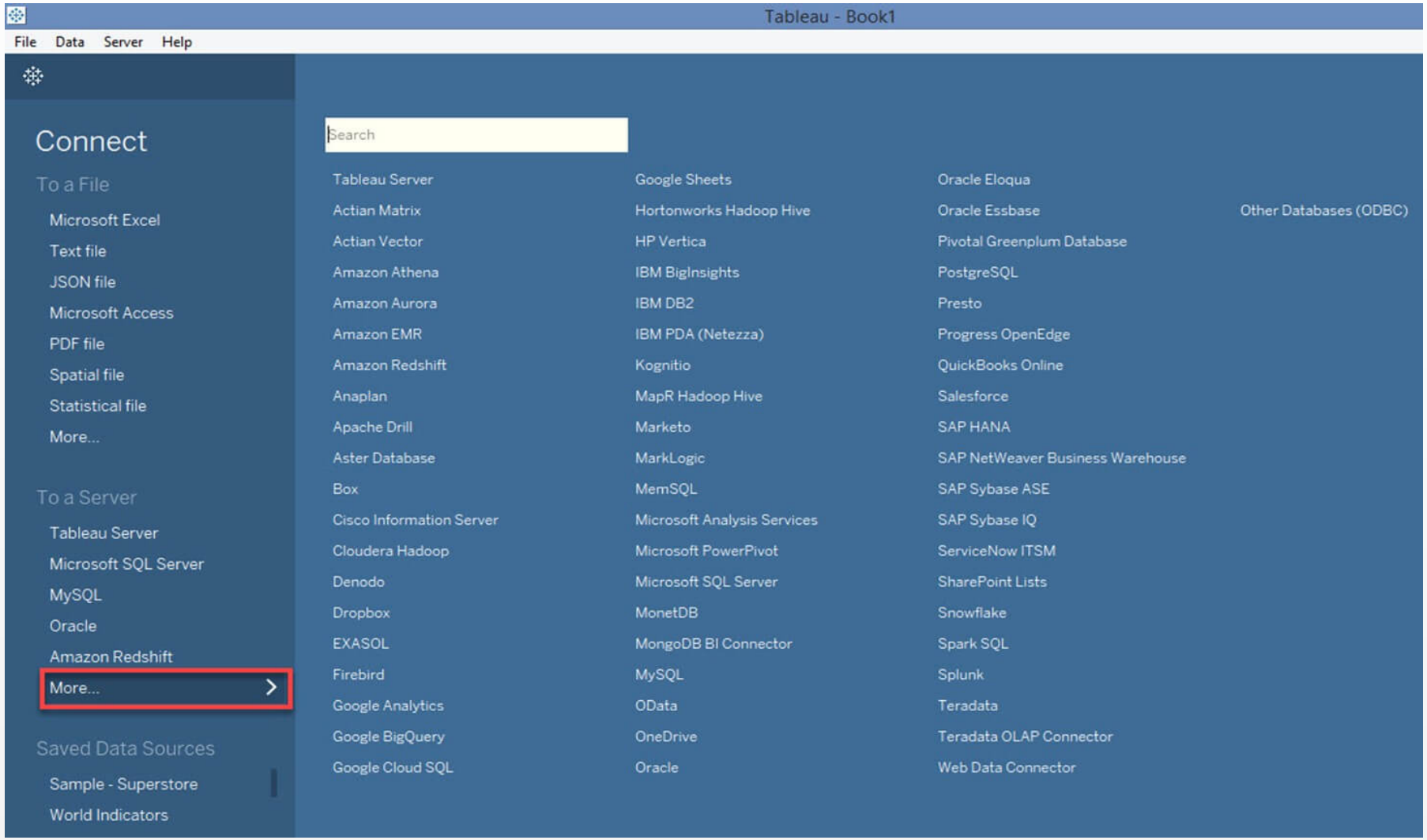


Tableau Data Connections



Data Repo

- http://bit.ly/qcl-tableau-Level1-data

CREDIT: COURTESY OF PARAMOUNT PICTURES

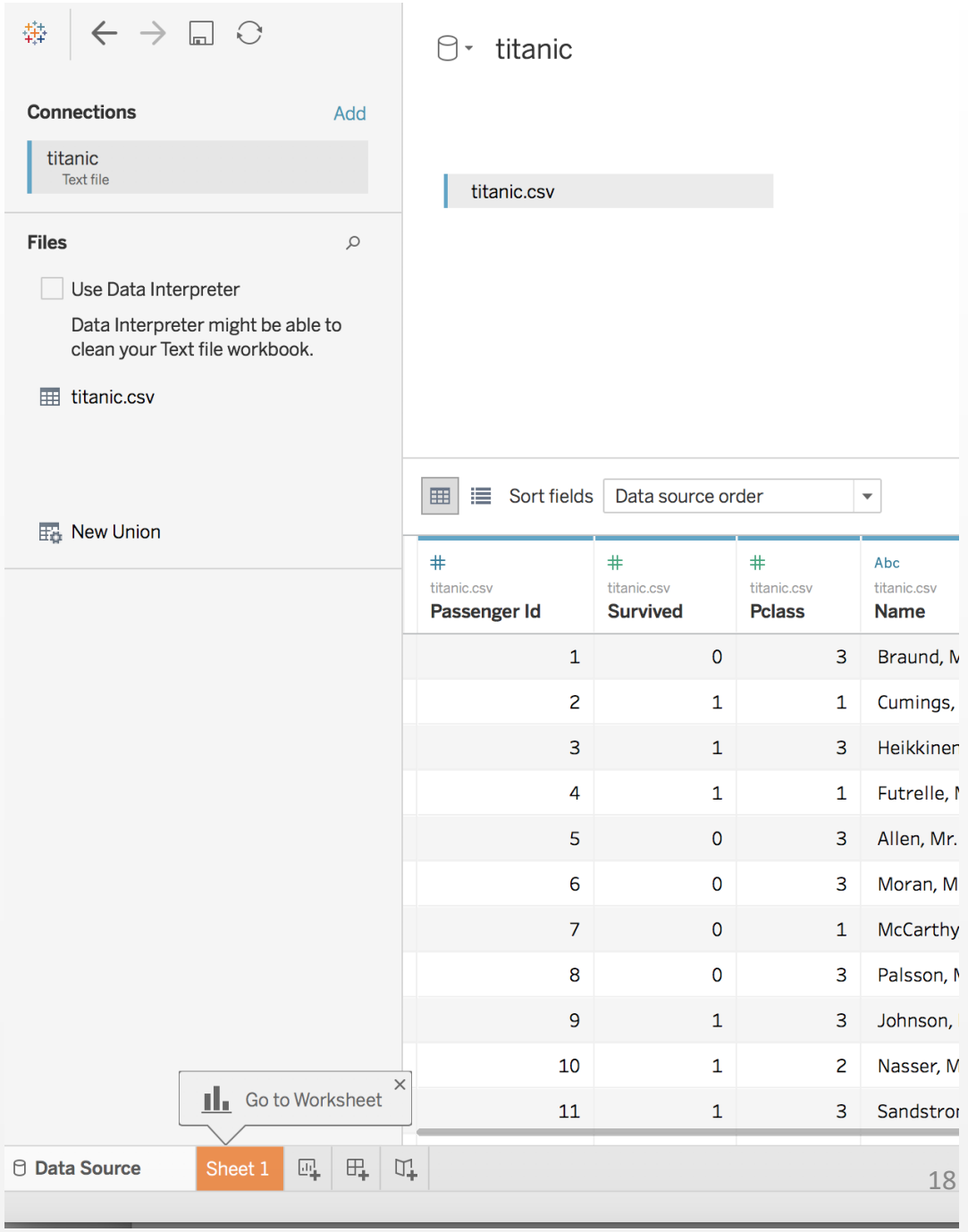
By Willy Stöwer - Magazine Die Gartenlaube, en:Die Gartenlaube and de:Die Gartenlaube

Connecting to Data

- Tableau can connect to many filetypes
  - Excel, csv, spatial, statistical
- **Connect to Data > More... > titanic.csv**
- This is only a section of the full titanic dataset.

Source: https://www.kaggle.com/c/titanic/data

Different ways to connect your data:
E.g. Google Sheet integration (demo)



Connecting to Data

- Tableau can connect to many filetypes
  - Excel, csv, spatial, statistical
- **Connect to Data > More... > titanic.csv**
- This is only a section of the full titanic dataset.

Source: https://www.kaggle.com/c/titanic/data

Different ways to connect your data:
E.g. Google Sheet integration (demo)

Variable	Definition	Key
survived	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	male | female
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	

Visualizing Data

The Chart Chooser From Abela's Extreme Presentation Method

What would you like to show?

COMPARISON

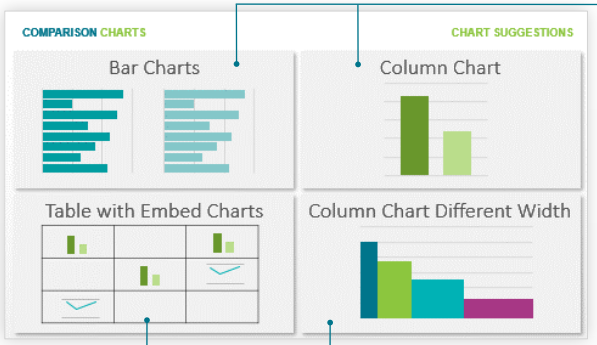
Among Items

Many Items / Few Items

Few Categories

Many Categories

Two Variables



Many Categories

Two Variables

Over Time

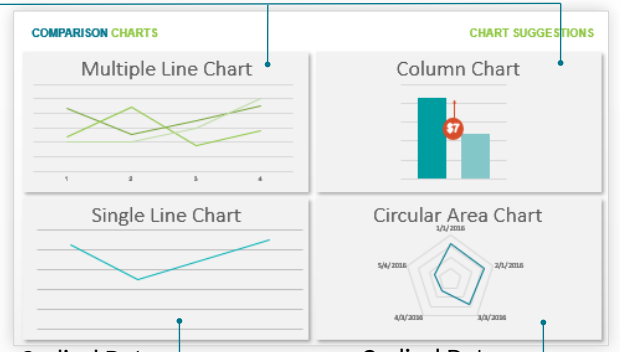
Many Categories / Few Categories

Many Periods

Few Periods

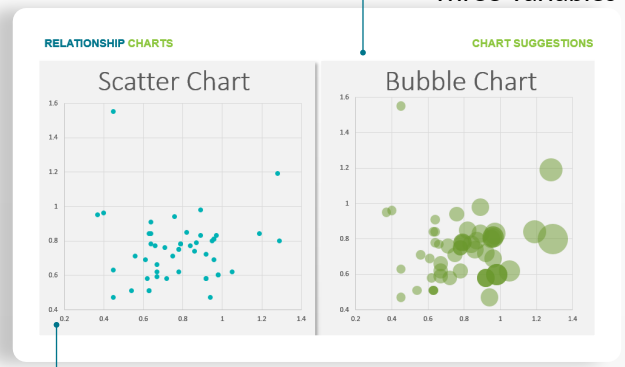
Non-Cyclical Data

Cyclical Data



RELATIONSHIP

Three Variables



Two Variables

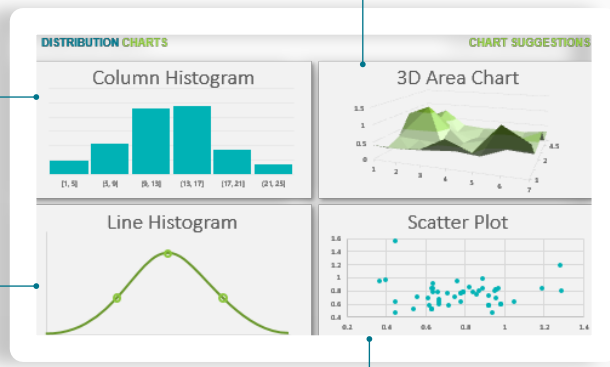
DISTRIBUTION

Three Variables

Few Points

Single Variable

Many Points



Two Variables

COMPOSITION

Change over time

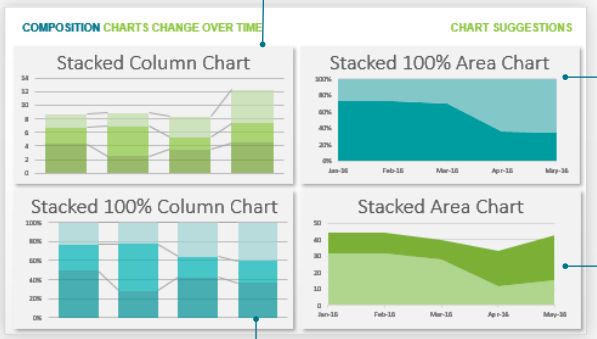
Few Periods – Relative and Absolute Difference

Many Periods

Relative Difference

Relative and Absolute Difference

Few Periods – Relative Difference

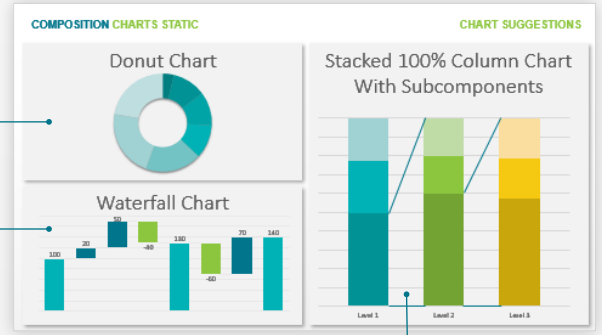


Static over time

Simple Share of Total

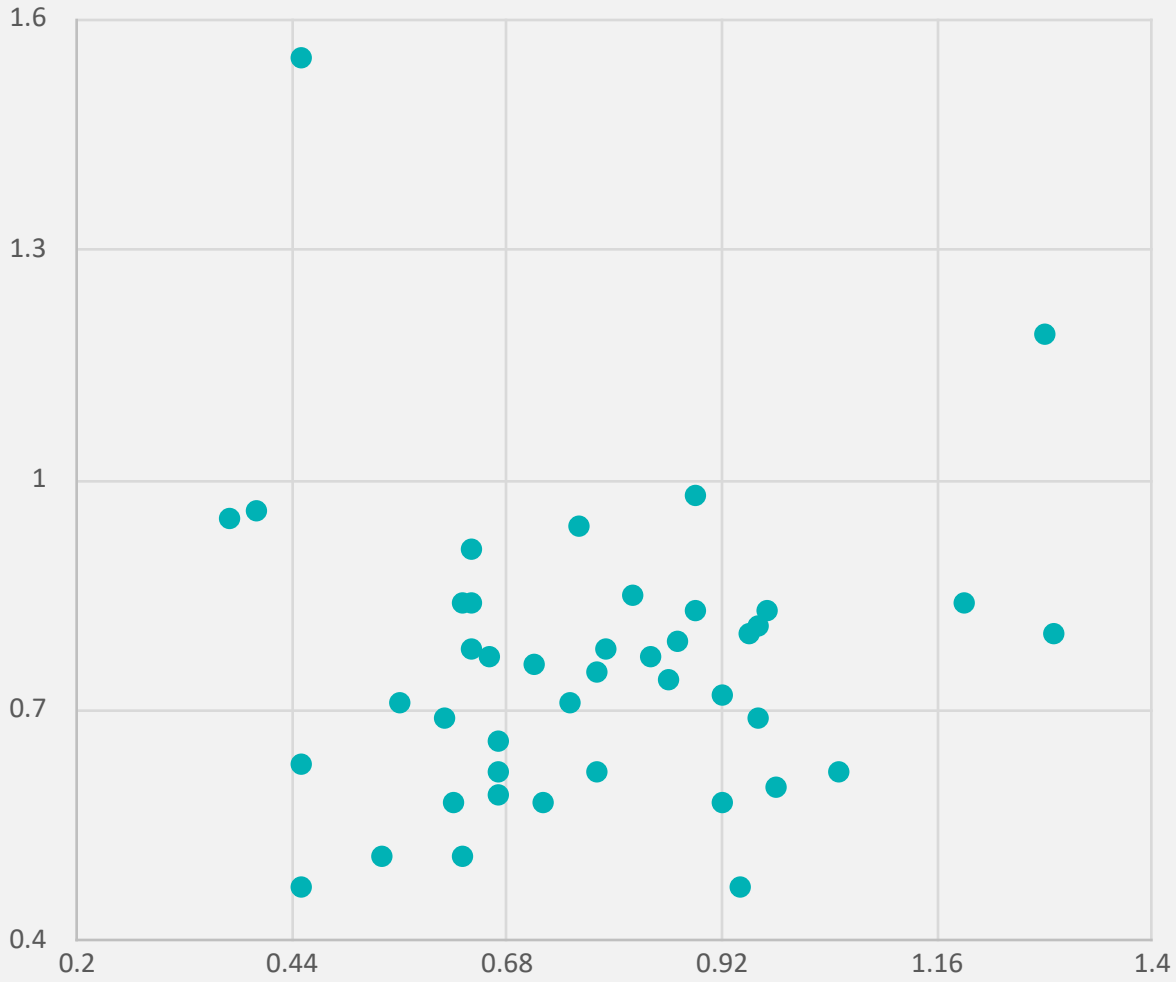
Accumulation Subst. of Total

Components of Components

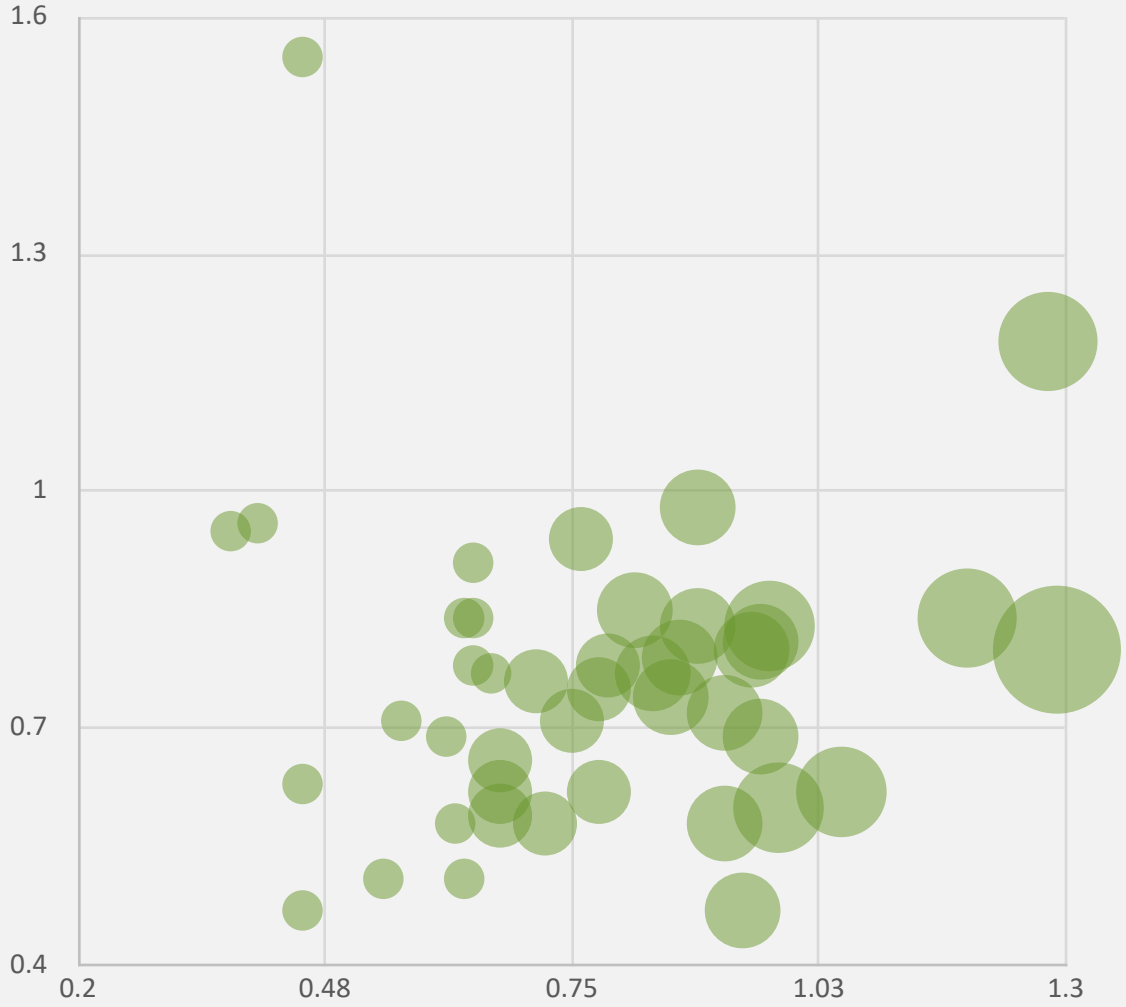


Source: Vinay Prajapati https://www.techprevue.com/decision-tree-perfect-visualisation-data/

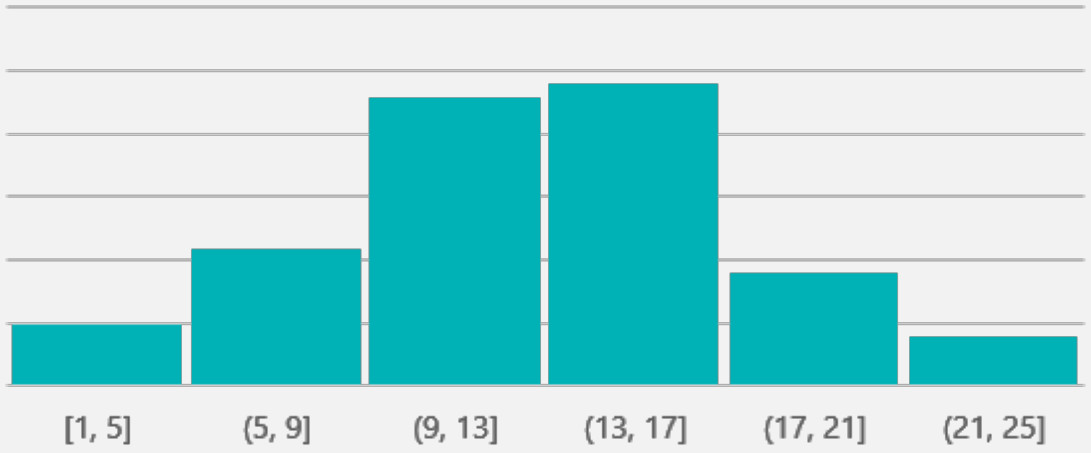
Scatter Chart



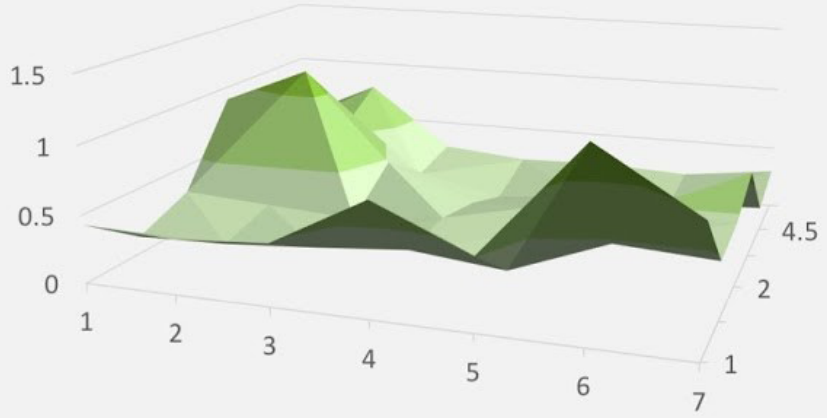
Bubble Chart



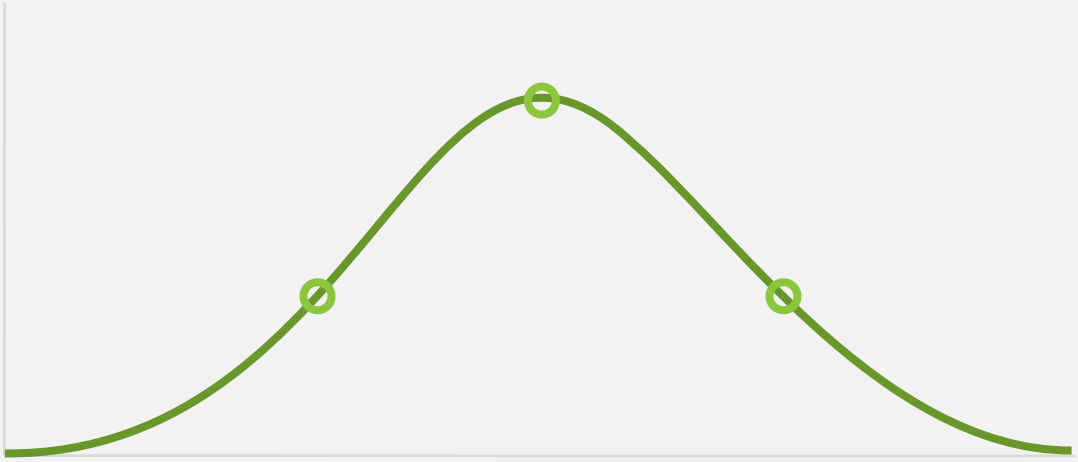
Column Histogram



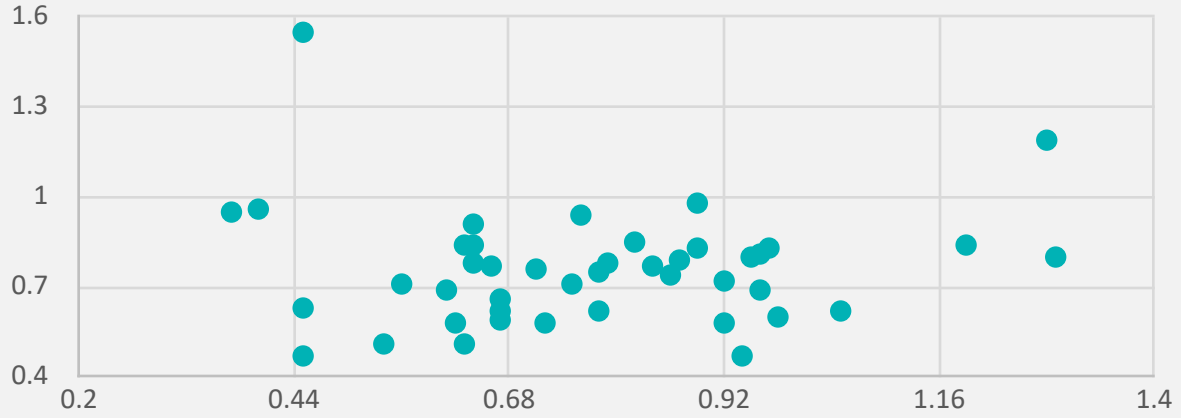
3D Area Chart



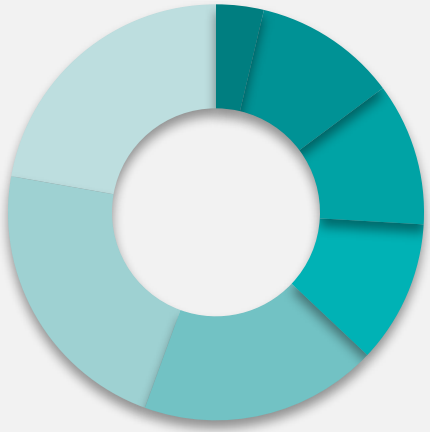
Line Histogram



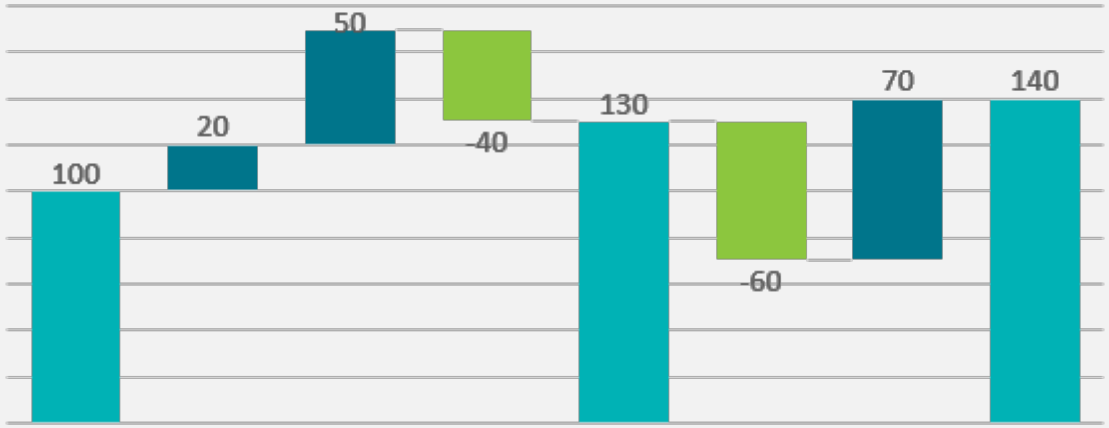
Scatter Plot



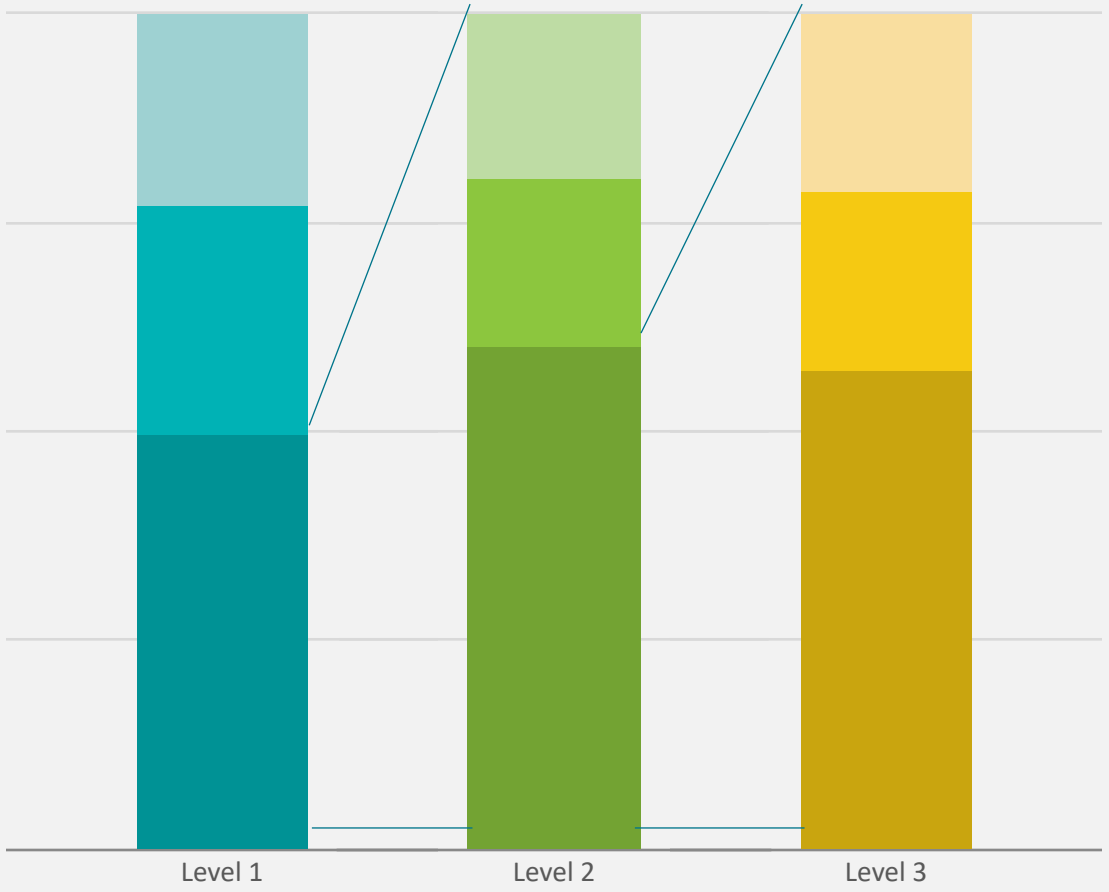
Donut Chart



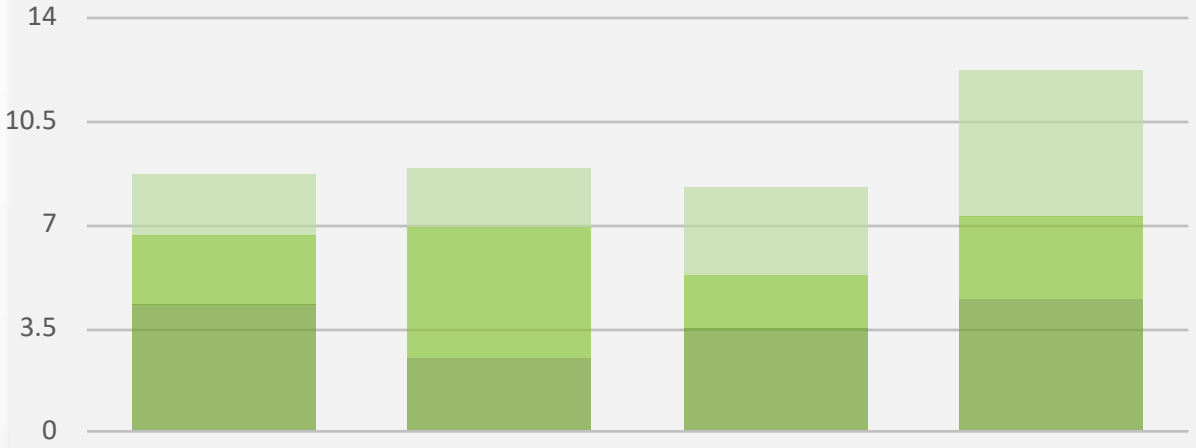
Waterfall Chart



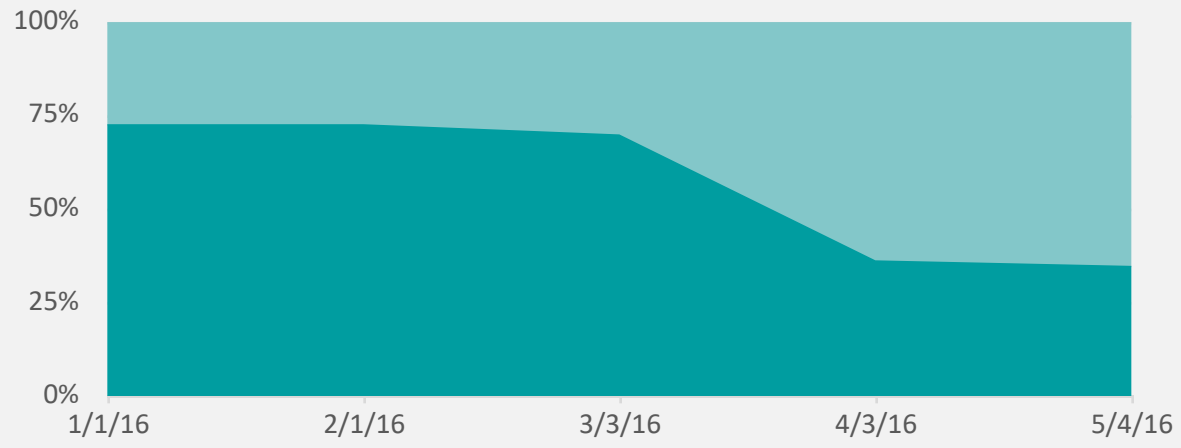
Stacked 100% Column Chart With Subcomponents



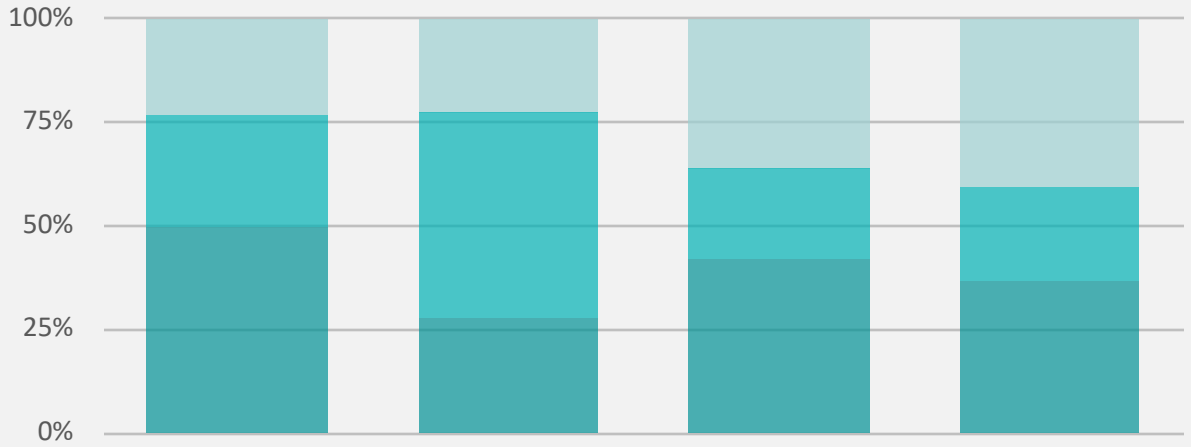
Stacked Column Chart



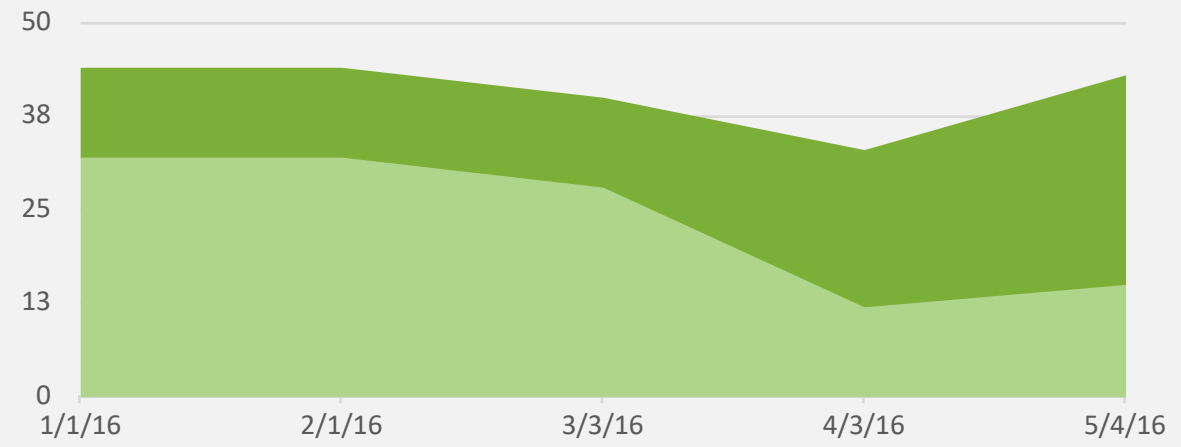
Stacked 100% Area Chart



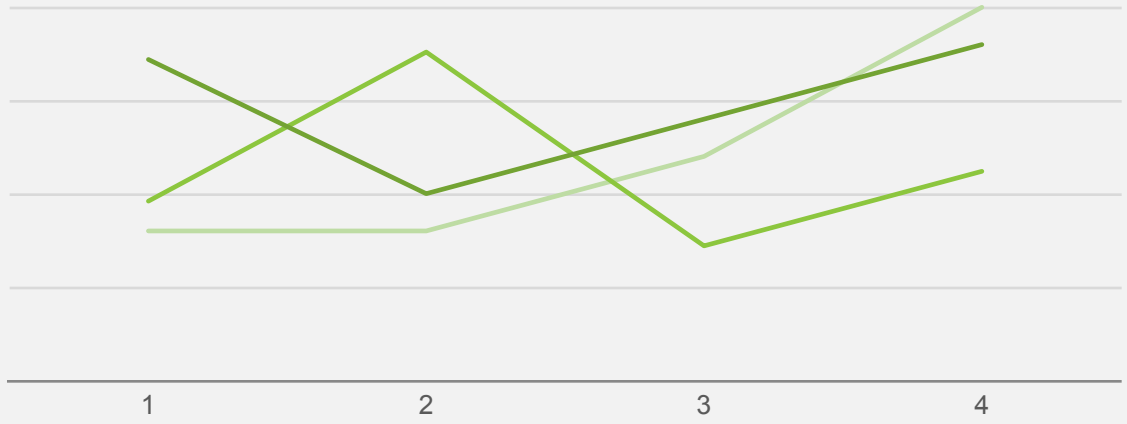
Stacked 100% Column Chart



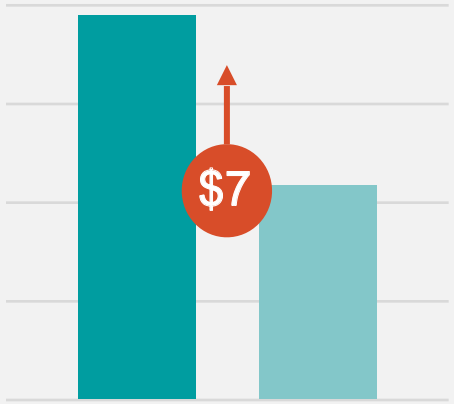
Stacked Area Chart



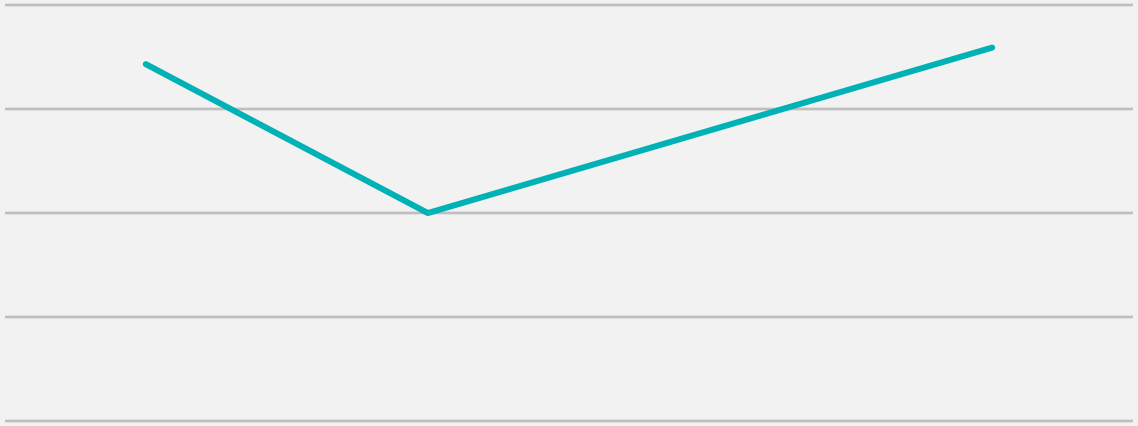
Multiple Line Chart



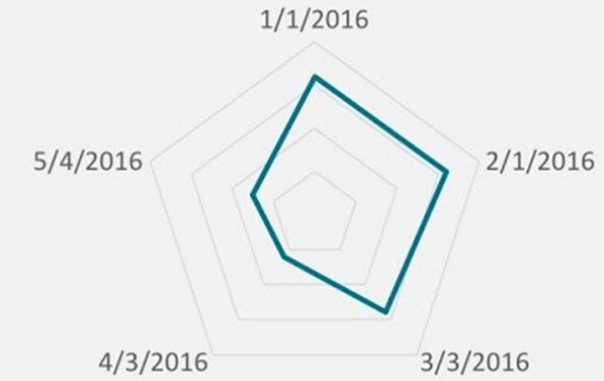
Column Chart



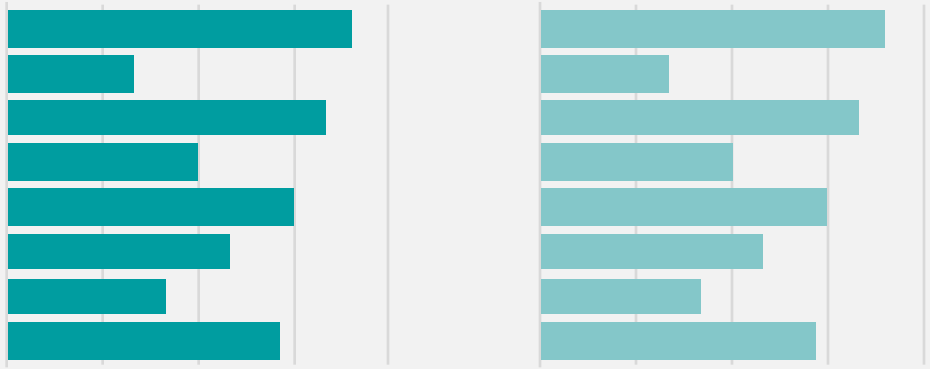
Single Line Chart



Circular Area Chart



Bar Charts



Column Chart

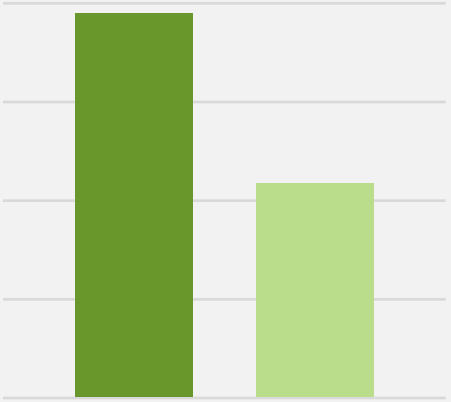
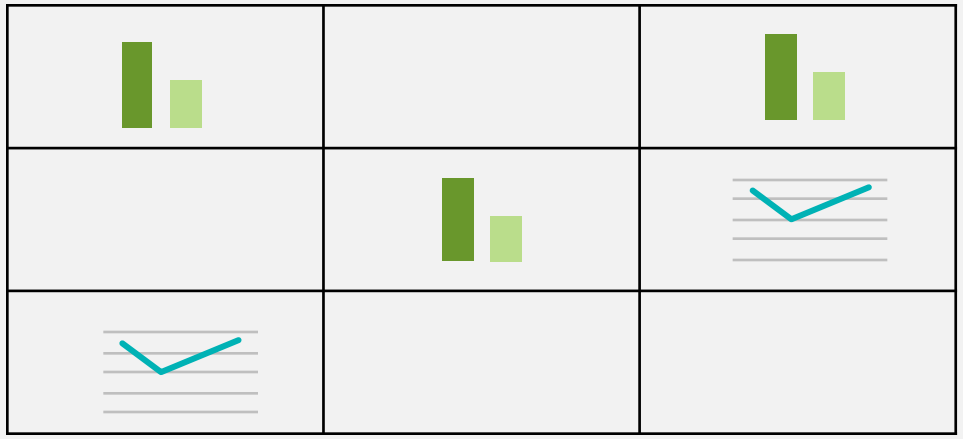


Table with Embed Charts



Column Chart Different Width

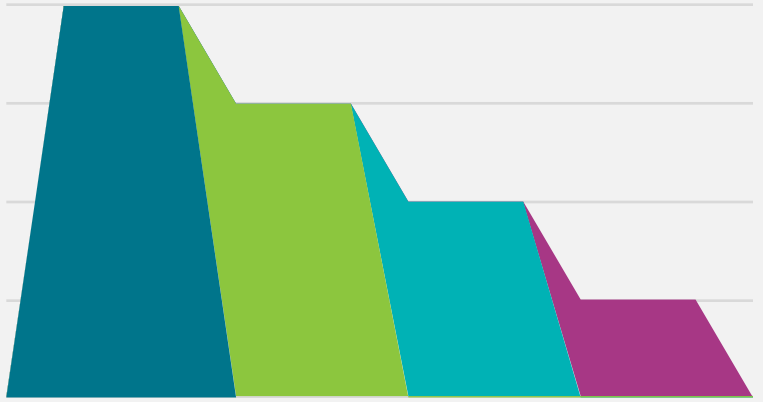


Tableau Environment

>> Data types, Workbooks, Sheets, and interface

Dimensions and Measures

- Tableau assigns any fields to ***Dimensions*** if they cannot be aggregated. (e.g. categorical data in strings or Booleans)
- Tableau assigns any fields to ***Measures*** if they can be measured, aggregated, or used for mathematical operations. (e.g., numbers)
- Ordinal data is a categorical, statistical data type where the variables have ordered categories like school grades (1st year for 1, 2nd year for 2, etc.). Tableau will import these as measures but often they make more sense as dimensions.
    - [Action] Convert *Pclass* and *Survived* to dimensions

Questions

(1) Tableau distinguishes data (fields) into two kinds. What are they?
(2) Say, in your data set, one field contains zip code. Would the zip code field be dimension or measure?

Workbooks and Sheets

- Tableau uses a workbook and sheet file structure, much like Microsoft Excel. A workbook contains sheets in three different kinds:
  - A worksheet contains a single view along with shelves, cards, legends, and the Data pane.
  - A dashboard is a collection of views from multiple worksheets.
  - A story contains a sequence of worksheets or dashboards that work together to convey information.

Workbooks and Sheets (cont.)

- The difference between a workbook (.twb) and packaged workbook (.twbx) is that a packaged workbook is meant for sharing and includes the data source and any other files used to make the workbook.
- Sharing Tableau workbook
  - Via Tableau file
  - Via Tableau Server
  - Via Tableau Public

Action

- Check if you can login to Tableau Public (if you don’t have an account, you may want to create one now.)

Visualizing Numerical Data

- Tableau has a collection of charts you can use to visualize numerical data.
  - These include histograms, scatterplots, box-and-whisker plots, and bullet graphs.

Action and Question

You wonder how the age distribution of the Titanic passengers was like.

- Show the age distribution of the passengers in a worksheet
- In what age range were most Titanic passengers? Use groups of 10 years (e.g., 0-9, 10-19, 20-29, etc.)

Histogram

- A histogram helps represent the distribution of numerical data.
  - It is similar to a bar chart but it is used to plot frequency of a continuous variable that is divided into bins.
- [Action] Drag *Age* into Columns or Rows, and then click on the histogram in the **Show Me** tab.
- In what age range were most Titanic passengers?
- NOTE: You can change the bin size by going to the *Age (bins)* dimension and clicking Edit...

Action and Question

Your colleague says that the older the passenger, the richer he/she would be. So there would be more aged people in the first class. You wonder if that's true or not.

- How can you show this?
- Is there any relationship between age and fare?

Scatter Plot

- A scatter plot is good at showing relationships between two numerical variables (measures).
- [Action] Add *Age* and *Fare* into Columns and Rows shelf
- By default Tableau aggregates the data which is normally what you want to do. However in a scatter plot you want to see each individual record. So you need to go to **Analysis > Aggregate Measures** to disaggregate the measures.
- Is there any relationship between age and fare?

Action and Question

In your dashboard, you want to include a table showing the number of survivors and deaths for each ticket class.

- Create a table showing the number of survivors (data value = 1) and deaths (data value = 0) per different ticket class on each row.
- In what class were most of the passengers that died?
- Change the low and column labels to ??

Text Tables

- Text tables aren't the most interesting way to visualize data but they have their time and place.
- [Action] Add *Pclass* and *Survived* into Rows and Columns
- [Action] Then add *titanic.csv (Count)* into the **Text** box in the **Marks** area.
    - I'll explain in more detail about what the Marks area is for later.
- In what class were most of the passengers that died?

Action and Question

- Show how many people died and how many people survived in a simple bar chart.

They should give a priority to women and children to get off the sinking ship.

- Can you show if sex played a part into survival?

Bar Chart / Stacked Bars / Pie Chart

- Bar charts are probably the most simple and effective way of presenting categorical data.
- [Action] Add *Survived* and *titanic.csv (Count)* into Columns and Rows shelf.
- What if you also want to see how sex played a part into survival?
  - [Action] Add *Sex* into Columns shelf.
  - This can also be represented as a Stacked Bar Chart or a Pie Chart using the **Show Me** tab.

Marks & Customizing the View

Color

- *Color* can be a useful tool in data visualization.
- [Action] Return to your histogram.
  - To change the color of this chart simply click the **Color** button in the **Marks** area and choose a new one.
- [Action] Return to your scatter plot.
  - Changing the color of a chart is nice but changing color based on a variable can be exceptionally helpful
  - [Action] Drag the *Pclass* dimension onto **Color**
  - [Action] Click on **Color** and you can change which color represents which *Pclass*.

More on Customizing the View

- *Size and Shape* act the same way as color. You can change them by clicking on them and you can make them change based on a variable by dragging that variable on them.
- Play around with these until you get a scatterplot that you like.
- Label adds a text label to each mark
- Tooltip changes what you read on the tooltip when you hover over a mark.
- Detail adds other variables to the tooltip.

Aliases & Formatting

- [Action] Return to your text table (Sheet 3).
- *Survived* 0,1 and *Pclass* 1,2,3 could be confusing to someone who doesn't know the data. However, we can change how this is labeled without changing the original dataset.
    - For numerical axes you can edit the title of the axes but for categorical axes you have to change the name of the dimension
    - [Action] Right click *Pclass* and rename to *Ticket Class*.
    - Next for the data itself you need to edit aliases to something more interpretable.
        - [Action] For example you can right click 0 and edit the alias to "No" and change 1 to "Yes".
        - [Action] Then you could change the ticket classes to "1st", "2nd", "3rd"
- Finally, you can click the title to give this chart a name.
    - You could name this Survival by Ticket Class
    - If you rename the sheet at the bottom it will automatically change the title as well.

Filtering, Calculations, & Analytics (Stretch Topic)

Filtering

- [Action] Make a copy of your text table worksheet (**Right-Click > Duplicate**)
- Perhaps you only want this view to show for adult women.
  - [Action] Drag *Sex* into filters and choose female.
  - [Action] Then drag *Age* into filters and select at least and then enter 18
- There are many ways to filter by variable and it's important when presenting data to explain what parts of the data are filtered out to avoid being misleading.

Analytics

- The Analytics pane provides quick and easy access to common analytic features in Tableau.
  - You can use this pane to add average lines, trend lines, totals and more.
  - [Action] Try adding an average line to your scatter plot
  - [Action] Try adding totals to your text table.

Dashboard

- Tableau Dashboard allows you to set up your visual output to fit specific device dimension.
  - Go to Dashboard
  - [Action] Select any worksheet
  - [Action] Try changing dimensions and check out device preview

Sharing

Sharing

- To share or present your findings the easiest way is to just export an image.
  - On any worksheet go to **Worksheet > Export > Image...**
  - It will give you options on Titles, Legends, Captions, and more.
- If you want to show a collection of visualizations you will be better off creating a dashboard and/or a story and exporting that.
- If you are collaborating with another Tableau user you can share the .twbx file or just a .twb if they have the data.
- If you need your visualizations to be interactive you can publish sheets to a Tableau Server. You can make a Tableau Public account to share your visualizations publically for free.

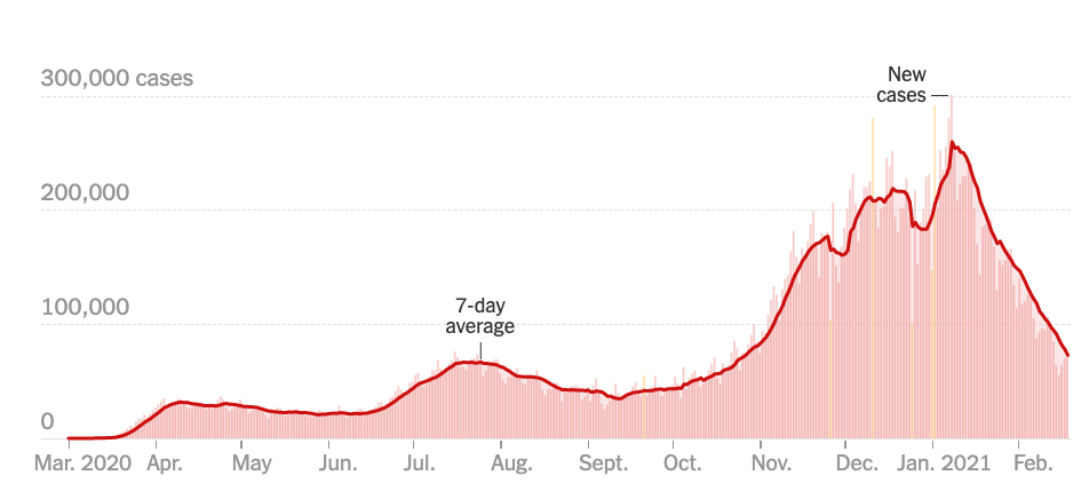
Tableau Public

- Example: https://public.tableau.com/profile/jeho.park4543#!/vizhome/DataVizwithTableau_0/Dashboard1
- Example: http://publichealthintelligence.org/content/global-overview-magnitude-disparities-and-trend-infant-mortality-world-1950-2011

Hands-On

COVID-19 Daily New Cases Visualization

- See New York Times Coronavirus in the US: https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html
- Download the county level data in CSV format from GitHub: https://github.com/nytimes/covid-19-data
- Create a viz like the NYT's daily new cases chart.



Further Learning

Further Learning: Topics

- Formatting
- Data Cleaning
- Advanced Calculations and Expressions
- Advanced Chart Types
- R Integration
- Geospatial Analysis
- Forecasting / Trend Lines / Regression

Further Learning: Resources

- Learning Resources:
  - Tableau.com/learn
  - Explore Tableau Public
  - Tutorialspoint
  - YouTube / Blog Posts
- Data Sources:
  - Kaggle
  - Data.gov
  - Data.world
  - Los Angeles Open Data (https://data.lacity.org)
  - UCI Machine Learning Repository

Thank you!