Tableau Workshop Part 0.1

By Zeyad Kelani

Graduate Fellow

Murty Sunak Quantitative and Computing Lab

Claremont McKenna College

Summer 2019

Sign-in Form: https://bit.ly/2K5CX5f (https://bit.ly/2K5CX5f)

Agenda

- Getting Started: What is Tableau?
- Tableau Environment
- Visualizing Data
- Marks & Customizing the View
- Intro to calculations
- All Hands-On

Learning Objectives

- Participants will be able to walk away with the following:
  - Connecting to data
  - Perfrom data merger in Tableau
  - Difference between measures and dimensions
  - Data Visualization "Viz"
  - Functions and Calculations

Getting Started with Tableau

What is Tableau?

Tableau is is one of the fastest evolving business intelligence and data visualization tools.

- It is a tool that connect to a data source, create visualization ("viz"), and combine/generate multiple visualizations into an interactive dashboard ("dash").

Figure 1 - Magic Quadrant for Analytics and Business Intelligence Platforms - Source: Gartner (February 2018)

What Can you do with Tableau?

- List of public projects availabe for anyone -without data sources- for idea generation.

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

Tableau Products

| Tableau Desktop	Tableau Online Server	Tableau Public	Tableau Public Server
Private	**Private**	**Public**	**Public**
Individual Computer License	`https://tableau.schoolname.edu`	`https://public.tableau.com`	Public Hosting for Tableau Files
Creat and edit visualization	Customized security	Data is stored/accessed through website	Read and Interact with Vizzes
`License - can grant one year for Academics`	`Institutional Level`	`Free`	`Free and Open Access`

Installing Tableau

- Install Free Trial Version of Tableau here:
  - https://www.tableau.com/products/trial (https://www.tableau.com/products/trial)
- Instructors and Researchers
  - Free Desktop license for a year (renewable) https://www.tableau.com/academic/teaching/course-licenses (https://www.tableau.com/academic/teaching/course-licenses)
- Students
  - Free Desktop license for a year (renewable) https://www.tableau.com/academic/students (https://www.tableau.com/academic/students)

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	

Tableau Environment
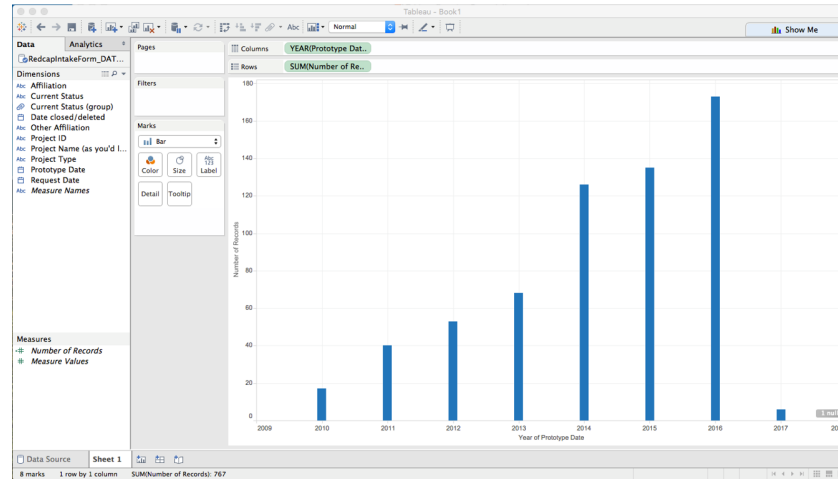
Desktop
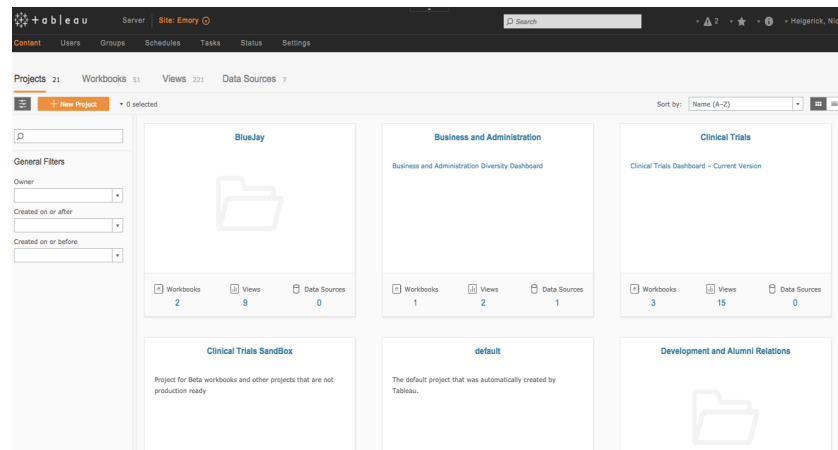
Server

Let's Start with Tableau

- Let's start by using Titanic dataset from Kaggle: The sinking of the RMS Titanic is one of the most infamous shipwrecks in history. On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships. One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.Let's use realistic but not real data and see how can visualization using tableau can be really helpful.
- We use data offered through a Kaggle competition:- Kaggle WebPage: `https://www.kaggle.com/c/titanic/data`

Make sure to download the data through the following link:

https://bit.ly/2SPtMZt

Data Overview:

- This section entails variables included in used dataset.

Variable	Definition	Key
survival	Surviva	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
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	
embarked	Port of Embarkation C = Cherbourg, Q = Queenstown, S = Southampton	

Connecting to Data:

- Tableau can connect to many filetypes
    - Excel, csv, spatial, statistical
- Download the dataset

- Connect to Data > More… > titanic.csv
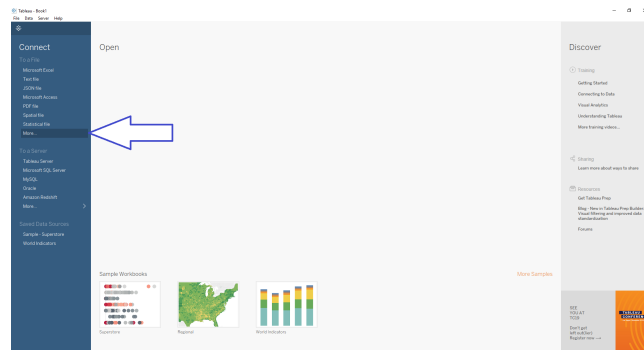
- Different ways to connect your data:

    - E.g. Box.com integration (demo)

Two main data types: 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.

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

Worksheet UI
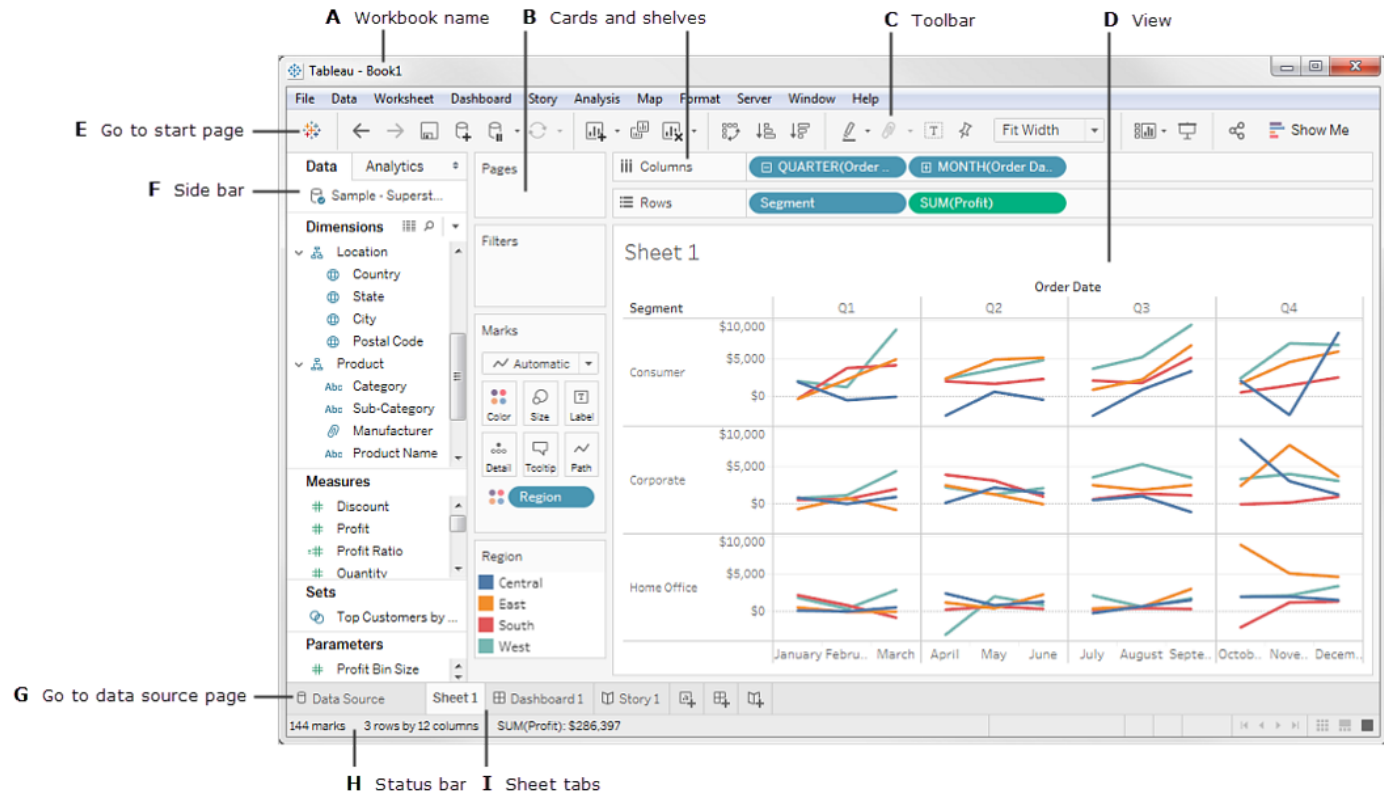
Worksheet Components

Visualizing Data

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.

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 and then click on the histogram in the **Show Me** tab.
- NOTE: You can change the bin size by going to Age (bins) dimension and clicking Edit...

Scatter Plot

- A scatter plot is good at showing relationships between two numerical variables (measures).
- `Action`Add **Age & Fare** into Columns and Rows self
- 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?

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 Number of Records 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?

Bar Chart / Stacked Bars / Pie Chart

- Pie charts are probably the most simple and effective way of presenting categorical data.
- `Action` Add **Survived and Number of Records** into Columns and Rows self.
- What if you also want to see how sex played a part into survival?
    - Add Sex into Columns self.
    - 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
  - 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

- `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 women.
  - `Action` Tne 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.

Calculations:

- Just summing up the data isn't always the best way to view it. Tableau allows for you to do many different mathematical calculations on the data.
- `Action` Return to Sheet 4
  - Right-click on the *Number of Records* in Rows self **> Quick Table Calculation > Percent of Total**
  - If you double click on the measure in Rows self, it will show the calculation that Tableau uses to get the result: ***SUM([Number of Records]) / TOTAL(SUM([Number of Records]))***
    - Tableau offers many functions that are very similar if not the same as the ones in SQL or Excel.
    - http://onlinehelp.tableau.com/current/pro/desktop/en-us/functions_all_categories.html (http://onlinehelp.tableau.com/current/pro/desktop/en-us/functions_all_categories.html)
- What percentage of passengers were male survivors?

Analaytics

- 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.

Thank you

For questions & feedback:

zelkelani@cmc.edu