Tableau Workshop Part 1

By Zeyad Kelani

Graduate Fellow

Murty Sunak Quantitative and Computing Lab

Claremont Mckenna College

Summer 2019

Agenda

- Getting Started: What is Tableau?
- Tableau Environment
- Visualizing Data
- Marks & Customizing the View
- Intro to calculations & mapping
- 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
    - Basic maps with Tableau

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	Table 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`

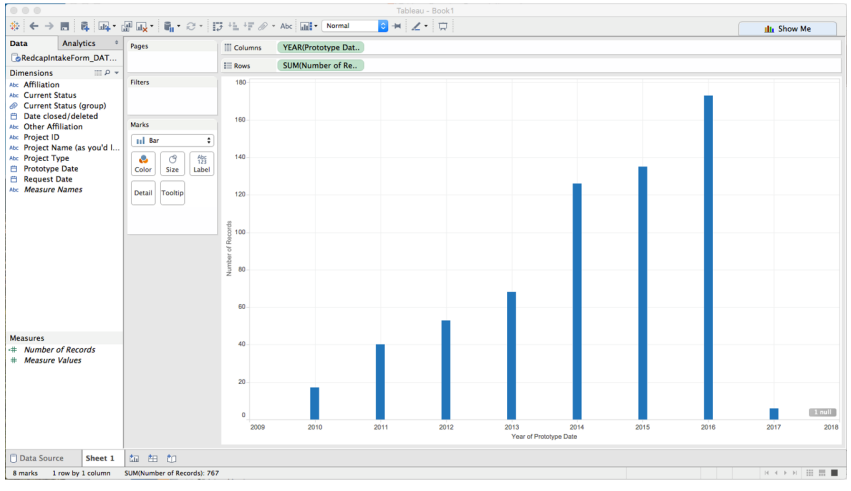
Installing Tableau

- 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)

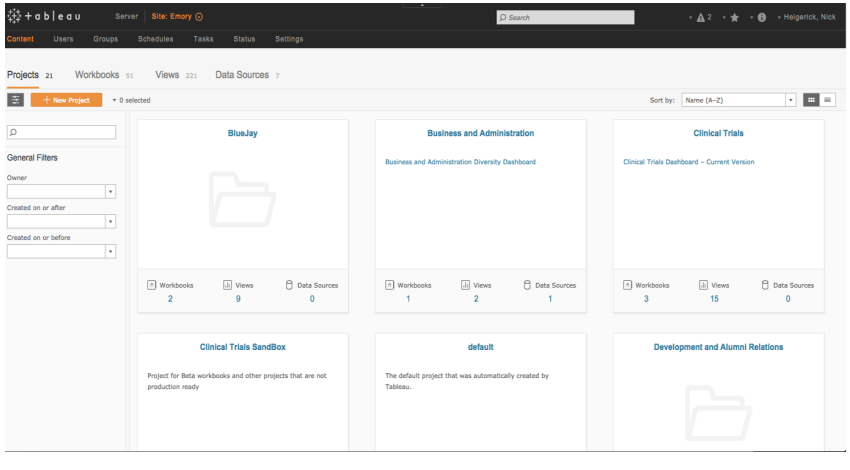
Tableau File Types

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 use realistic but not real data and see how can visualization using tableau can be really helpful.
- For this session, we will use SyntheaTM, which is an open-source, synthetic patient generator that models the medical history of synthetic patients.
- https://synthetichealth.github.io/synthea/#about-landing (https://synthetichealth.github.io/synthea/#about-landing)
- You can also download the data through the following link:
    - `<INSERT BOX HYPER-LINK>`

Data Overview:

Synthea data contains -realtistic but unreal- complete medical history, including medications, allergies, medical encounters, and social determinants of health. This data can be used without concern for legal or privacy restrictions.

- From Synthea Datasets, we will use:
    - patients.csv
    - observations.csv

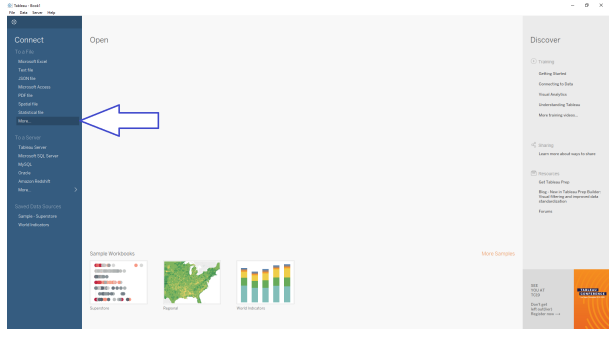
Notes on Variables

- **patients dataset contains**: (wide-format)
    - patient ID, birth date, death date, marital, race, ethnicity, gender, city, state, zip.
- **observations dataset contians**: (long-format)
    - date, patient ID, encounter ID, code, description, value, units, type.

Connecting to Data:

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

- Connect to Data > More… > observations.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.

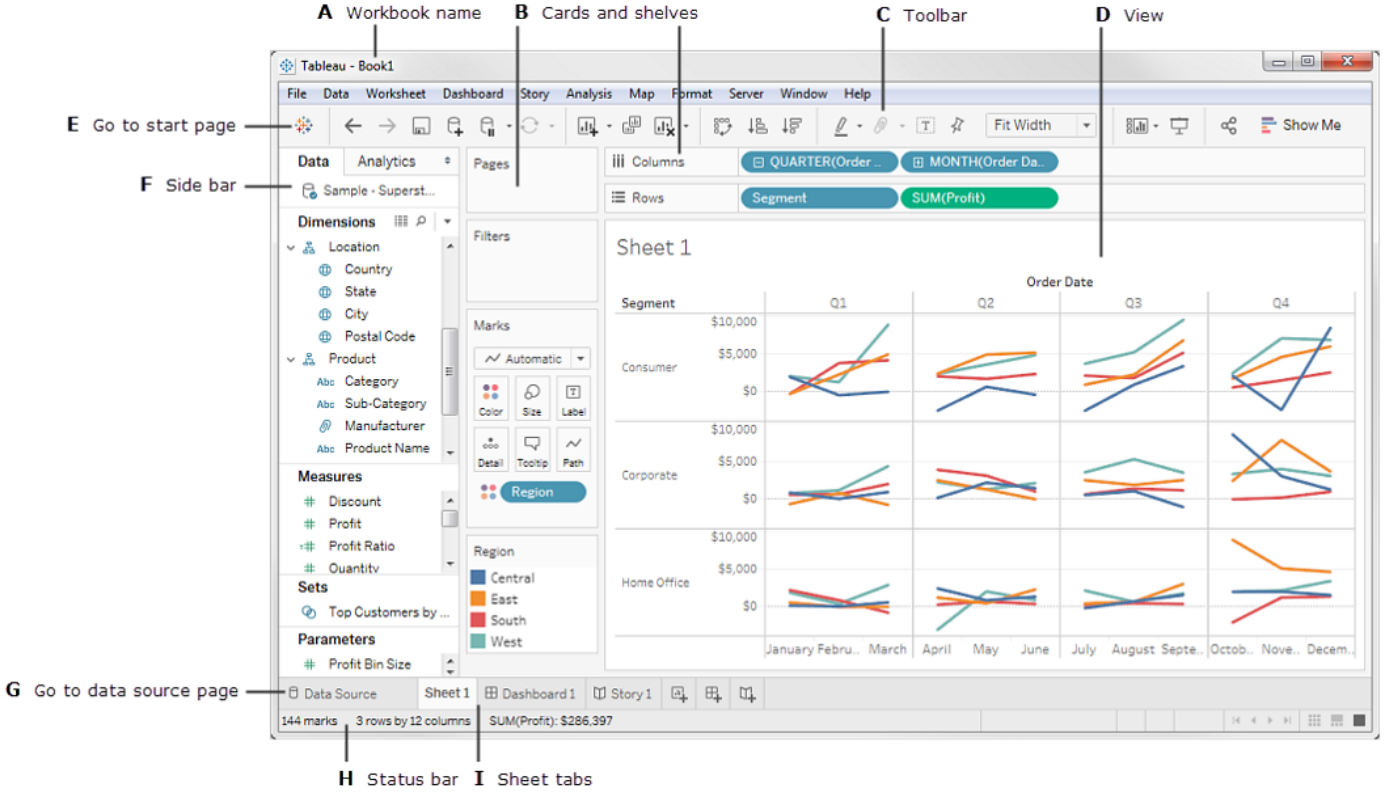
Workbooks and Sheets

- 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

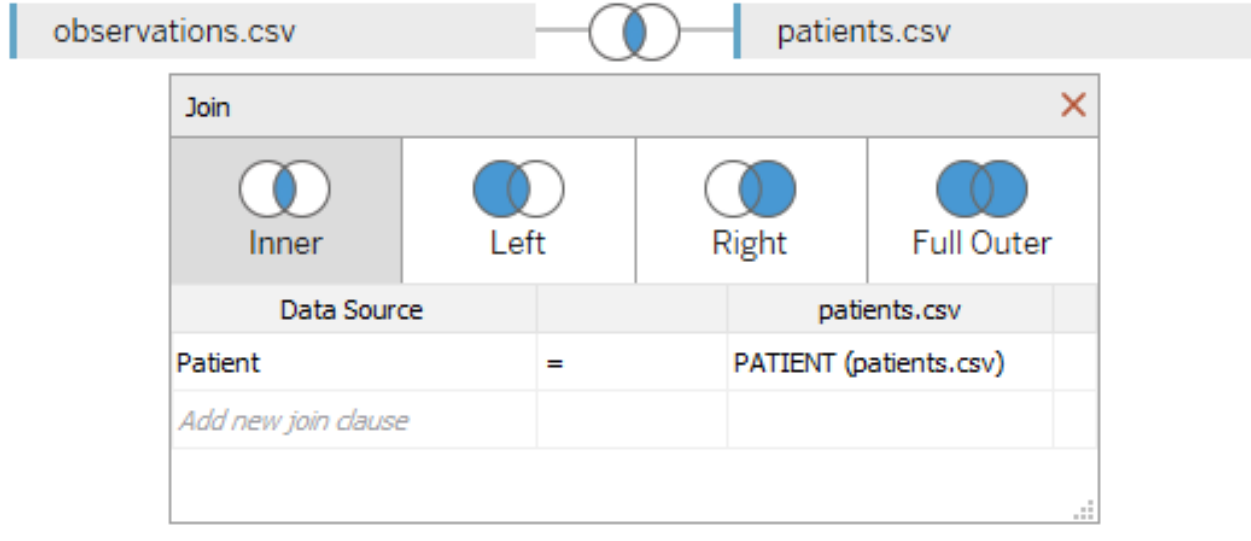
Worksheet UI



Worksheet Components

Add More Data:

- Tableau allows for simple merging based on unique identifiers, in this case, it will be "patient id."
    - `Action` drag *patients.csv* to data part and choose "inner join."



Inner Join in Tableau

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 *Gender* & *Race* into Columns and **Number of Records** into Rows and then click on the histogram in the **Show Me** tab.
- NOTE: You can change the bin size by going to size and change size.

Scatter Plot

- A scatter plot is good at showing relationships between two numerical variables (measures).
- `Action` Drag **Description** into **Filters** and Right-click and choose **Diastolic Blood Pressure**
- `Action` Add *Birthdate* into columns and *Number of Recirds* into Rows, and choose shape.
- Is there any relationship between birthdate and blood pressure?

Text Tables

- Text tables aren't the most interesting way to visualize data but they have their time and place.
- `Action` Add **Gender** into Columns, **Race & Ethnicity & Marital** into Rows.
- `Action` Then add Number of Records into the **Text** box in the **Marks** area.
- `Action` Right-click on *Marital* and unmark *Null*
- what is the racial/ethnic features of our sample?

Bar Chart / Stacked Bars / Pie Chart

- Pie charts are probably the most simple and effective way of presenting categorical data.
- `Action` Drag **Description** into **Filters** and Right-click and choose **Body Mass Index**
- `Action` Add *Description* into Rows self.

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 **Race** dimension onto Color

More on Customizing the View

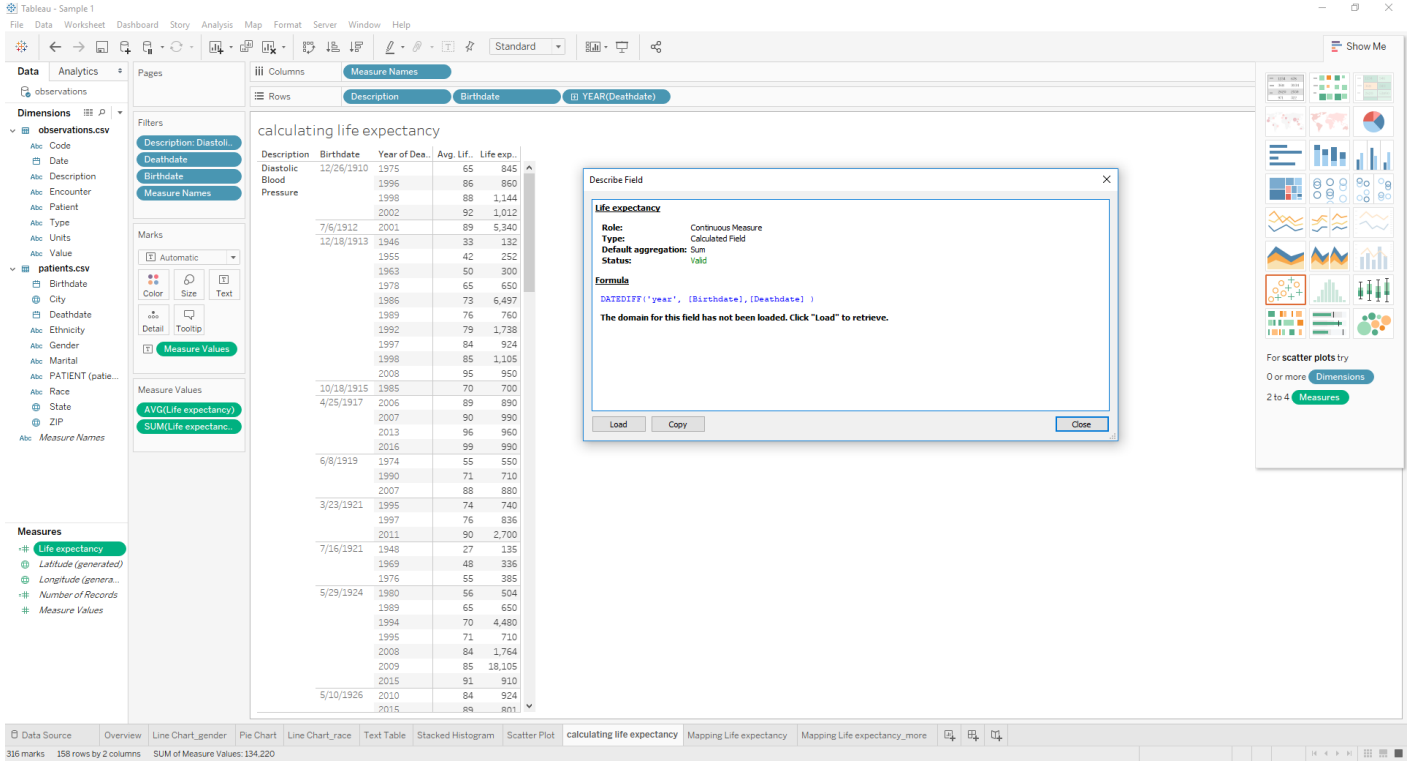
- Size and Shape act the same way as color. You can change them by clicking on them and you can make them change based on a variable by dragging that variable on them.
- Play around with these until you get a scatterplot that you like.
- Label adds a text label to each mark
- Tooltip changes what you 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).
- Ethnicity varibale can be confusing, you can rename levels through *Right-Click* on *Race* variable and click on **Aliases**, and start editing.
    - `Action` Right click *Race* and rename.
- 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.

Simple Calcluation:

- let's calculate life expectancy from our sample data



Life Expectancy Calculation:

- `Action` choose *Diastolic blood pressure* from **Description** filter.
- `Action` move *Measure Names* into columns and *Description* & *Birthdate* & *Deathdate* into Rows
- `Action` add *Measure Values* into **Label** in *Marks* area.
- **From Analysis** `Action` Choose **Calculated field** and type the following:
    - DATEDIFF('year', [Birthdate],[Deathdate])
- `Action` add **Life Expectancy** to Rows.

Mapping Life Expectancy

- `Action` move *Longitude and Latitude* to columns and Rows, then choose *Map* from **Show Me**
- `Action` move *Life Expectancy* into **Color & Size** and *City* to **Label** in *Marks*
- From Size, chnage size of your circle and same for color.
- For **Treemap** visualization, choose *Treemaps* from **Show Me** section.

Thank you

For questions & feedback:

zelkelani@cmc.edu