Thank you for coming to attend this **introductory Tableau workshop**! Here, you will learn the advantages of the data visualization tool Tableau public and important concepts and techniques needed to display and analyze data. You’ll learn to make and convey data-driven decisions or recommendations for your work through hands-on activities.

Hi, my name is Nopphiphat Suraminitkul (or you can call me by Nopi). I am a graduate student pursuing my degree in Computer Science, and my role in the Gelman library is statistical consultant.

So, let’s begin with the **flow of the workshop**. We will start with what are we going to accomplish by the end of the workshop, basic introduction to data visualization, reason for us to use Tableau, materials we will be working with, data visualization process walk-through, and ends with some practices

The goal for the end of today’s session is that you should be able to understand different ways to handle different data types, methods for creating good visualizations, and gain some hands-on experience with Tableau Public.

What is data visualization? It is the **pictorial representation** of dataset(s) and information using **visual elements** like line, shape, color to give us a framework to make sense of masses of information, see and understand trends, detect hidden patterns, identify outliers, and develop new insights in data, which can then be used to **make data driven decision**. These are the basic visual elements which can be combined to make complex visualizations like graphs, charts, and maps to **show relationships between variables**.

Tableau is one of the visualization tools that **allows anyone to quickly learn and make beautiful visualization** using drag-and-drop method on large data files of most file types, e.g. Excel, CSV, PDF, and SQL to name a few.

The **drag-and-drop method** just means that you can just drag any elements from dimensions and measures under data pane and put them into the columns, rows, marks, and views to create visualizations within minutes. Just to answer what you guys may be wondering, **dimensions contain qualitative values**, such as names, dates, or geographical data, which can be used to categorize, segment, and reveal the details in your data, whereas **measures contain numeric, quantitative values** that you can measure. By default, when measure is dragged into the view, Tableau applies an aggregation to that measure.

These are some type of the visualizations that can be created using Tableau, dual-axis bar charts, time series, scatter plot, and density maps.

Just a show of hands, who haven’t installed this software? For those who raised their hands, please navigate to this link, enter your email, and install the software. While your software is being installed, on the same website, sign up for an account on Tableau Public. For those who have already signed up, hover your mouse on the profile icon on top-right icon and click “Settings”, and ensure that both “Allow viz data to be downloaded” and “Set my vizzes to be hidden” are ticked.

Next is the dataset that we will be working with. Navigate to the link and click ‘download’ to download it your local desktop. For those who haven’t worked with Kaggle before, you might need to register before you are eligible to download this dataset.

Now, the main content, the data visualization process. For every data visualization, we will go through these steps, loading your data, knowing your data, determining your data, pre-processing your data, visualizing your data, and exporting your data. So, lets open the software.

**Data Visualization Process**

**Step 1) Loading your data**

* Open Tableau Public
* Click on “File” -> “Open”
* Navigate to the location of the file (e.g. Desktop)
* “Select” the file

Once this is done, the selected data will be loaded into the format that Tableau can work with.

**Step 2) Knowing your data: Understand data definition**

If there is only 1 sheet in the data file, the data will be automatically loaded into the table for us to see. However, for Excel file with multiple sheets, we have to select the sheets that we want to work with. We can view the data in any sheets by hovering the mouse over the sheet name and click on the icon to the right of the name. Once we have a good idea on which data sheets we will be working with, then we will drag them into the canvas.

* [Optional] Under “Files” in the left column, [Tick] Cleaned with Data Interpreter, which is designed to help take care of cleaning and transforming the data into the proper format for analysis in Tableau. This can be useful for extracting data from table in excel where column names not on the first row.

Once data is being displayed in the data connection page, we need to look at the data, see what we have, as well as check to see if the data types (String, Integer, Date, Geographic, Boolean, Cluster) are being displayed correctly.

* **String values** (Text): Abc Icon; **Integer values** (Numbers): Hash Icon; **Date values** (DD/MM/YYYY): Calendar Icon;

**Geographic values** (Regions, Postal code, etc.): Globe Icon; **Boolean values** (True or False; relational): T|F Icon;

**Cluster group** or mixed values

What do we have?

* show\_id: Unique ID for every movie or tv show
* type: Identifier – a movie or tv show
* title: Title of the movie or tv show
* country: Country where the movie or show was produced
* date\_added: Date it was added on Netflix
* release\_year: Actual release year of the movie or show
* rating: TV rating of the movie / show
* duration: Total duration – in minutes or number of seasons -> strings but we would want their numeric numbers
* listed\_in: Genre

**Step 3) Determining what to visualize: What answer are we looking for**

* What do we not already know?

(Quick overview of data)

* Can the information be represented without visualization?

(Not all data needs to be visualized)

* Does the audience need to see pattern to understand the data?

(Changes in preferences over time)

**Step 4) Pre-processing your data**

* This part normally consists of renaming and splitting data values, changing data types, replacing nulls, etc. which we wouldn’t go in-depth in this workshop.

Example 1: duration (Abc -> Integers)

* Click on the inverted triangle icon on the top right when hovering mouse over duration column header
* Click ‘Custom Split’, input space “ ” in the section for use the separator, and split off “All”
* Rename the newly created column “Duration - Split 1” to “Minutes”
* Click on the “Abc” icon and change it to “Number (whole)”

**Step 5) Visualizing your data**

**Few Tricks:** In case of error, press “Ctrl + Z” to return to the state before the current action is made. However, if you want to clear the whole view, then click “Worksheet”, “Clear”, and “Sheet” or click on the “Clear Sheet” icon.

* **Basic Pie Chart**: Content Type on Netflix
  + Rename “Sheet 1” to “Content Type on Netflix”
  + Drag “Type” under Dimensions to “Rows”
  + Drag “Measure Values” to “Text” under Marks
  + From the “Show Me” drop-down menu, select “Pie Chart”
  + Select display “Entire View”
  + Drag “Type” to “Label” under Marks
  + Drag “Measure Values” to “Label” under Marks
  + Click on the inverted triangle icon on the green pill “SUM(Number of Records)” under “Measure Values” and click “Quick Table Calculation” and select “Percent of the total”
  + Modify the sizes of the “Pie Chart” by click on the “Size” under “Marks”
* **Basic Line Chart**: Content added on Netflix over the year by Type
  + Drag “Date Added” under Dimensions to “Columns”
  + Drag “Type” under Dimensions to “Rows”
  + Drag “Number of Records” to “Text”
  + From the “Show Me” drop-down menu, select “Lines (continuous)”
  + Drag “Number of Records” to “Label”
  + Click on “Label” and click “Select” under “Marks to Label”
  + Right-click on the points you want to show label, and “Mark Label” and “Always Show”
  + Right-click on the points you want to show drop lines, and “Drop Lines” and “Show Drop Lines”
  + Right-click on the points again, and then “Edit Drop Lines”, Select Labels to “Automatic”
  + Drag “Date Added” to “Filters”, and select “Range of dates”, and click “OK”
  + Click on “Date Added” under Filters, and select “Show Filter”

Here, also point out how data points are being limited, may be because the source wasn’t doing great with data sourcing.

* **Reference Lines**
  + Click on “Analytics” beside the “Data Pane”
  + Hold and Drag “Average Line” to the “Table”
  + Right-click on the line and click “Edit”
  + Under Line, select “Custom” for “Label”, and click the next arrow “>”, select “Value”
  + Type the meaningful text like “Average:” in front of <Value> or “Movies added” after the <Value>
  + Under Formatting, select “- - - -” for “Line”, and select color for “Fill Below”
* **Basic Bar Chart**: Rating Histogram
  + Drag “Rating” under Dimensions to “Columns”
  + Drag “Number of Records” to “Text” under Marks
  + Select “Horizontal Bar” from the “Show Me” menu
  + Select “Pivot” Icon to modify the bar chart to vertical bar chart
  + Sort the data in descending order by clicking the “Sort Descending” icon
  + Formatting the background, fonts, etc.
  + Add Caption
  + Modify title to accommodate the information about the data
* **Exercise: Add Reference Lines**
  + Add “Average Line” to the View
  + Right Click on the “Average Line”
  + Select “Custom” under “Label”
  + Press on the “>” sign and select “<Value>”, type in “records”
  + Select any color under “Fill Below” to shows the better contrast
* **Tree Maps:** Rating on Netflix
  + Drag “Rating” to “Rows”
  + Drag “Number of Records” to “Text” under “Marks”
  + Click on “Null” and select hide
  + From the “Show Me” drop-down menu, select “Treemaps”
  + Drag “Number of Records” to “Label” under “Marks”

Simple Tree Maps (1 or more dimensions, 1 or more measures)  
Value in dimension define structure, Measure represent the size  
Colored Tree Maps  
Interactive Tree Maps

* **Packed Bubbles**: Rating on Netflix
  + Duplicate the previous sheet
  + From the “Show Me” drop-down menu, select “Packed Bubbles”
  + Sort the data in descending order
  + Drag “Number of Records” to “Label” under “Marks”

Circle represents the dimension, Size represents the measures

* **Creating dashboard**
  + Click on creating new dashboard icon below
  + Drag the sheets that are being created into the dashboard
  + Modify the size to fit the screen
  + Play around with the filtering

**Step 6) Exporting your data**

* Click “Save to Tableau Public” button under “File”
* Go to your Tableau Public account
* Your saved works will appear under the “Vizzes”
* Click on “View” to view the visualization created
* Click the “Download” icon on the bottom-right and the type of format you want your file to be saved as

Exporting the visualization is way is inconvenient when we dealing with many visualizations. So, this is where Tableau Desktop comes in. Tableau Desktop is a paid version for Tableau, which has many useful features that Tableau Public do not have. Example of it is the “pivot functions” on the actual data, which can be used to create the accurate visualization like this map. Luckily, if you are a student, you can request 1 free access to Tableau Desktop for your usage.

To conclude this session, I just want to restate the goal of this class, which is to provide you guys with the knowledge about data visualization and some of the methods to create your own visualization using Tableau Public. If you want to continue learning on your own, there is a resource that is available for GWU students, which is linkedin learning. An online learning resource that allows you to search and learn new skills at your own pace.

If you guys have any problems regarding your work with visualization, you can schedule an appointment with statistical consultant through the Academic Commons. Thank you for coming. See you again.