SEIS 631 Project Draft

John Muller

05/03/2022

Introduction

What

Project is about COVID deaths compared to total deaths. The project will calculate the percentage of deaths that are related to COVID. It will run other statistics based on gender or age. It will also populate GGPLOT and provide a visualization for the project/results

Why

I choose this topic because it is all fresh in our mind and I knew we could all relate to COVID. Data is readily available. I enjoy doing visualizations and this will get me more familiar with importing csv data files to create plots.

How

I knew there would be online resources readily available for using ggplot. So I am utilizing google searches and how to videos to specifically design ggplot to what I want. Importing a csv file full of data has also been new to myself. So doing some google searches and debugging the code has been important.

Body

Everyone knows that the Covid Pandemic has killed many people in the world. We have all relied on data from the News, Websites, or friends. I thought it would be interesting to look into the data myself and see if I can portray it in a manner that I could customize. The data set I have found includes the following columns

Sex Age Years Total Deaths COVID-19 Deaths

Sex and Age Years are categorical variables than I can run statistics on. Plus, for the purpose of learning, it should be a smooth transition to creating plots in ggplot.

ggplot is interesting because you can modify the plots in many ways. You can alter the X and Y axis, modify the table, and change colors. . . etc

It is important to understand how much COVID has impacted our lives. The goal of this project is to see what percentage of deaths are now related to COVID. If you grow older, how much more likely are you to die from COVID??? We will be able to run statistics on male/female deaths or how age relates to the fatality rate when exposed to COVID. For example, if you are less than 20 years old, if you were to die, how likely would it be if it were from COVID. what about if you are 85?

Topics from Class

Rmarkdown

We are using Rmarkdown to generate a PDF of our results and findings. It will generate headers, texts, and plots. You can place R code into the Rmarkdown file and it will run the code and place it into the PDF file. It is convenient since it stores your code so you can run it in the future.

Github

This project draft will be posted to Github. It allows others to view and work together on projects anymore. If we were sharing code, we could load code up to the site and make it open so anyone could grab it. In our case, we are posting our PDF's to the website so that other classmates can view our projects.

Bar Plots and Categorical Variables

A bar plot is a common way to display a single categorical variable. Categorical variables that are included in this dataset are Sex and Years Age. It would be interesting to show how the categorical variables of sex, if males or females have more deaths related to Covid. And how age is related to deaths.

Numerical Values

This data set has a range of numerical variables. The Tot_death field is the numerical variable I focused on for this project. We can take the average, add, subtract. How many deaths of women compared to men are related to covid?

Skewness

This distribution set for COVID deaths is left skewed. As you look at the dataset, you have two variables that you can measure. You can relate age and the amount of covid deaths. As you are a younger age, the less likely you are to die from COVID. And as you increase in Age, the more likely you are to die from COVID. This data set is not normally distributed

Conclusion

My knowledge before working on this project was limited. I believe we talked about ggplot in class but most of my learning and debugging came from using interest research and walk through videos. I first had to tackle how to import a csv file. The files in class seemed to be easier to load. As for ggplot, I have learned more about how to structure a r code to customize a plot. There are plenty of ways to customize the table and I have learned it is critical to customize the plot because the basic one that is provided doesn't fit your expectations. Perhaps you would want different labels, or different colors, or show different values on your x and y axis.

Sample Data

```
##   Data.as.of Start.Date  End.Date    Sex    Age.Years Total.deaths
## 1  4/27/2022  1/1/2020  4/23/2022 Female  0-05 Months        18051
## 2  4/27/2022  1/1/2020  4/23/2022   Male  0-05 Months        22197
```

```
## 3  4/27/2022   1/1/2020 4/23/2022 Female 06-11 Months          1572
## 4  4/27/2022   1/1/2020 4/23/2022   Male 06-11 Months          1976
## 5  4/27/2022   1/1/2020 4/23/2022 Female     01 Year           1436
## 6  4/27/2022   1/1/2020 4/23/2022   Male     01 Year           1757
##   COVID.19.Deaths
## 1             97
## 2            106
## 3             27
## 4             36
## 5             24
## 6             37
```

Plot

If you die...% you will die from COVID