Data Science Reflection: Data Science: Visualizing and Exploring Big Data, Fall 2020

After being exposed to RStudio and seeing the analysis you could do with data, I decided my junior year that I wanted to become a data science minor. Once I came to realize this, I then realized I had to begin my journey learning R the correct way. This prompted me to take an intro level data science class called “Data Science: Visualizing and Exploring Big Data”. Not only was this class a requirement for the minor, but also essential for me to learn R from the beginning instead of just the essentials needed for a class project or research.

The class taught me several things, some of which I knew before however I never understood how they worked. Clearing datasets, indexing columns were actions I relearned from previous classwork. ggplot2 was a package I revisited and learned all the features it could do. The most important skill I learned in this class though was how to build a shiny app. Shiny is a package that allows for the building of interactive web applications. Not only did I learn how to create a server and user interface for the app, but how to customize it and add different data visualizations.

Using these skills, my partner and I explored racial inequality between the white and black community in the States. The first disparity we wanted to display was life expectancy between both groups as well as by gender over time. Afterward we looked at income disparity between the two groups over time. In both visualizations, it was clear to that those who identified as Black lagged in life expectancy and income obtained when compared to their white counterparts. We attempted to plot life expectancy by US counties over different time periods, with the idea being to see if life expectancy has increased over time, especially in counties known to be home to low-income residents. There was a slight increase in life expectancy in all counties over time, which was impressive to see. We then had a page where we could have users download the raw data we used in case they wanted to create their own shiny app. For the improvement I did on the shiny app, I used Python to examine college attainment for Black and White residents in Virginia counties. I then ran regression on the dataset I used to determine if there was a correlation in counties between percentage of residents that identify as black and percent of college attainment. The same correlation was sought with white residents and it seemed there was a negative correlation between county percentage of Black residents and college attainment while there was no correlation between county percentage of White residents and college attainment.