Data Science Reflection: Neighborhoods, Culture and Poverty, Winter 2020

Winter semester of sophomore year, I decided to take classes that I would not normally take and that were outside my major. There were two reasons for this, the first being that the semester before had been a STEM heavy semester, which made me want to take a break from content heavy classes. The second being that I was eager to get as many graduation requirements completed before junior year. For these reasons, I decided to take a sociology class called “Neighborhoods, Culture and Poverty”. What started out as a class used to take a break and meet requirements ended up as a springboard for me to discover a new career path and skillset unknown to me at the time.

The first thing that I learned were basic statistical terms and tests as I had not taken a statistics course by that point. Basics like linear regression, ANOVA tests, t-tests and correlation were taught as these were processes commonly discussed in lecture the literature. From there, R was introduced as well as the accompanying work environment, RStudio. This was my first introduction to any programming language which made for a steep learning curve. I also learned how to import different packages into RStudio and read documentation for code implementation. Obtaining datasets and cleaning them for only relevant information was also taught using the dplyr package. Lastly, data visualization was another skill gained which involved learning how to use packages such as leaflet and ggplot2. These packages allowed me to plot data on maps and create scatterplots as well.

Using these skills, I explored racial inequality in Los Angeles County between the white and black community. The first difference that was explored was concentrated disadvantage in the different cities within the county. By plotting calculated disadvantage on a map of the county, and after various literature review, it was easy to see that cities with a higher black population tend to experience more concentrated disadvantage. Using regression, we also found that there was little to no change in improvement of concentrated disadvantage for these cities. Following up on the disadvantage between the two communities we sought to find whether members of the college community were obtaining college education at the same rate as their white counterparts. We found that this was not the case and that there was a significant difference between both groups. All these figures culminated in a thirty-page report on the issues in Los Angeles County.