Open Reflection

This week we started using tidycensus and the census API to access datasets. I've worked with this before, but only for a month or so. I absolutely love the direct link to census data where I can clearly define what data I want, what data I don't want, and quickly import it into RStudio without having to download 15 csv files. In the past, I've done a lot of population analysis in excel, which isn't terrible, but it is time consuming and frustrating to simply get the data ready to look at. It takes time to search and locate the data you want, it takes time and space on your computer to download each individual file, and then it takes even more time to load the data into excel and begin tidying it up enough to digest what it is saying. I am very excited about the prospect of working with tidycensus and the API key because I can do it quickly and seamlessly.

One function that I learned this week is case_when(). I've never used that function before because I've always used if statements. My main programming experience has been in C++ and Python and both of those make use of a lot of if statements. I've had to adapt to R and RStudio and how nearly every function is built in. There is something simple and controllable that I love about C++ and the customizability in the language. RStudio has been a little difficult for me to use because I have to teach myself functions all the time instead of telling the program to do exactly what I want it to do. Luckily, I'm getting better at learning new functions and they've mostly proven to be useful.

I really like this week's discussion of how to display data and the different benefits of each method. We looked at showing growth in Chicago Community Areas based on categorical variables as well as continuous variables. The categorical variables are useful because they account for outliers. Continuous variables are useful because they show more distinct differences, however outliers can skew the whole classification system. These are things to keep in mind when showing the results of my own data.