Journal Report 4
9/23/19-9/29/19
Sarah Luthra
Computer Systems Research Lab
Period 2, White

Daily Log

Monday September 23

I realized that the workspace on RStudio was not connected to console. I tried to fix this problem with different commands and I read two tutorials on why this might be happening, but I was not able to solve this problem. I just wrote my code in the workspace and ran it in the console directly to still get some results.

Tuesday September 24

I continued trying to figure out why my console was not connecting with my workspace. I eventually then ended up researching if CO2 emissions could be directly linked to agricultural output and if life expectancy could be linked to crime rates. I read some articles online and one research paper that all seemed to indicate that because more resources were used by farmers to grow and sustain larger quantities of crops, CO2 emissions were good indicators of relative agricultural output. I also learned that life expectancy was influenced heavily by homicide and crime rates. I found data for CO2 emissions and life expectancy and ran updated regressions on this data, as my previous sets had many holes in them.

Thursday September 26

I reran my remaining 3 regressions with better data I found from the World Bank website to get updated p-values for all of the factors. None of the results of whether a factor was significant or not had changed, but the p-values did shift. For example, for GDP, the p-value went down significantly lower, indicating that it was a lot more important than I had originally imagined. I also researched the cor() and chisq() functions to see how I could model my data.

Timeline

Date	Goal	Met
Sept 9	Finalize factor list and finish a script in R to run a multi-linear regression to determine p values	Yes, but now I need to do this same thing with my actual immigration factors.
Sept 16	Run a regression on my factors and get p values for each factor. Run this by Mr. White to see if this makes sense. Start researching the best way to make my predictive analytics model	Yes, I received p values for the factors I tested.
Sept 23	Run a chi-squared goodness of fit test on the factors using R and obtain chi squared values for each factor. Determine weights based on these factors.	No, I did not ending up running a chi squared goodness of fit test since I tried to troubleshoot some RStudio problems, fixed gaps in my data,
Sept 30	Run a chi-squared goodness of fit test on the factors using R and obtain chi squared values for each factor. Determine weights based on these factors.	
Oct 7	Check with Mr.White and a statistics teacher to make sure that my results make sense, and start to write my predictive analytics model.	

Reflection

I ran into a few problems this week, one of which I still don't have the solution to. I hope to resolve the console connection issue this week- I think it's a simple problem to fix, but none of the websites I found online explain how to fix it. I hope that next week is the last week that I work on this factor weight aspect of my project so that I can begin my predictive analytics model as well, since I believe that will take a good amount of time as well.

I am glad I got updated p values for all of my factors and found more complete data sets for the factors I am examining, though, so I can better determine each factor's relative importance. Since the linear regressions also return R values, I can also determine R squared values. From what it looks like, though, the r squared values are pretty low for the small p-values I recieved. I'd like to investigate how that happens and come up with a reasonable explanation for that.