Daily Log

Monday September 16

I put in some basic commands into RStudio to graph the factors against emmigration data to see the best fit lines for each of the six factors. I also researched the different types of tests I could run to obtain p values or correlation coefficeints that could tell me the relative importance of each factor on emmigration data.

Tuesday September 17

I wrote a regression model script for my data. This involved importing my csv files, reading in the data, creating and running the model, and outputting results. I ran this at home, since it did take a while to run. I got some p value results and showed Mr. White, and I observed that GDP was an especially important factor in determining whether people would emmigrate from a country or not.

Thursday September 19

I looked up information about the chi-squared test, which I think is what I will implement next week, since those results make sense in the context of my project. I learned the syntax in R for chi-squared regression tests, and tried a couple of tutorials that explained how to obtain and interpret data from a chi squared output table and plot.

Timeline

Date	Goal	Met
Sept 2	Download and process data sets, start researching and creating lists of factors for emigration to put through a multi linear regression. Downloaded R packages to write the multi linear regression	Yes, I made lists for 2 countries and finished the regression.
Sept 9	Finalize factor list and finish a script in R to run a multi-linear regression to determine p values	Yes, and I worked on doing the same thing for the 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 created a linear model for the data and got associated p values for each factor.
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.	
Sept 30	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 learned a lot this week- primarily that the p values I obtained did not make very much sense in context of the project, even though they are some indicator of how influential the factors are on emmigration numbers. After researching the different types of statistical tests, I have decided to also run a chi-squared goodness of fit test to determine the chi-squared values for all of the factors. The larger the chi-squared value is, that means that the factor was either very influential or very unimportant to the outcome of emmigration data. The smaller the chi-squared values, the closer it is to the expected value for how influential is it influential (16 percent, since I have 6 factors).

Below is a screenshot of one of the outputs I used when I ran a regression on an R package of sample data.

```
Residuals:
     Min       1Q    Median        3Q       Max
-158.205  -41.667   -6.248    57.741   118.810

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         1798.4      899.2   2.000  0.05861 .
Interest_Rate        345.5      111.4   3.103  0.00539 **
Unemployment_Rate   -250.1      117.9  -2.121  0.04601 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 70.56 on 21 degrees of freedom
Multiple R-squared:  0.8976,    Adjusted R-squared:  0.8879
F-statistic: 92.07 on 2 and 21 DF,  p-value: 4.043e-11
```