Journal Report 13
12/09/19-12/13/19
Sarah Luthra
Computer Systems Research Lab
Period 2, White

Daily Log

Monday December 9

I started using Atom instead of PyCharm, which was easier since Atom already had a virtual environment set up for me to work in.

Tuesday December 10

I changed the set-up of the code from more of a DFFNN structure to a regression NN with keras. I put the models file in another directory so the code was easier to follow. I tested my code on sample test data online (house data; and it worked!) and then finalized my refugee data to match the same file format.

Thursday December 12

I ran my code (trained and then tested) with my data, and got some results! I've attached the output to this journal. It's not that accurate, but I have something to work with now. I think the problem is mainly a lack of data, which is something I can't really change. But, I can definitely continue to play with the network architecture, which will definitely change the accuracy of my results.

Timeline

Date	Goal	Met
Nov 25	Try to fix the version control issues.	No, I started working on it with miniconda though.
Dec 2	Fix all data imports and run the model to see if I produce any sort of intelligible result. If I am getting errors, try to fix those so I can rerun it the following week.	Yes, but now I am trying to fix the data so I can actually get good results and also fix the version control problem.
Dec 9	Get results to be within 40 percent of the actual results	Yes, the percent error of my code was 39.56 percent. So I just barely made it!
Dec 16	Get results to be within 30 percent of the actual results.	
Dec 23 or winter break	Get results to be within 20 percent of the actual results.	

Reflection

I think I finally made a break through this week with the model. Changing the structure really helped me think about it differently. Next week, I also want to investigate more into how I should incorporate the statistics of my results from the script in R that I wrote. Also, I acknowledge that the results of my code won't ever really be super good due to the limited amount of data that I have found. Later, I want to go back to see if I can find more data that will improve the accuracy of my model. I'd like to think I can drastically improve the percent error, but I just don't know if that's completely possible even with the best possible NN, given my sample size restriction. For now, though, I can keep trying the improve the accuracy and finally move on to the simulation part of the project. I've attached the output for this week:

```
[INFO] training model...
[INFO] predicting refugee numbers...
[INFO] avg. refugee number: 151348
[INFO] mean: 39.56%
```