May 29, 2022

The goal of the project is to have a map that displays the languages. Ideally, I would want it to be interactive and color coded. I also want to see if there is any correlation between latitude/longitude and the family/genus of languages and see if there is a way I can guess the family from the language. I am planning to use the World Atlas of Language Structures which was suggested by the Professor. I downloaded the csv from Kaggle and uploaded it to the 'Project' folder I have on my Jupyter Notebooks.

June 5, 2022

I have deleted some columns in the dataset. I am trying to delete columns that are mostly blank or are only for one or two specific geographic areas. Also, seemingly arbitrary columns like 'iso_code' and 'wals_code' are being deleted. I have not identified the specific columns I will use to identify universals yet. I want to assign each macroarea a number which I think will be useful in analysis but I do not know how to do that yet. In the upcoming week, I hope to learn and also hope to plot my first map and reduce the number of columns to 10-20 for ease of analysis. 200 seems excessive. The columns that I know I want to use are clean in terms of the values. There are no random or useless punctuations etc. I have deleted rows which were blank for certain columns as I think if the languages do not have those values, they are very niche and because of that, outliers in the analysis I want to do. For example, if a language does not have a macroarea listed, the language might not be researched well enough to help in the analysis as macroarea should be an easy column to fill out. I hope to report more code and progress for next week's check in.

June 16, 2022

I found a better way to drop the columns. I dropped columns which have at least 1,250 null values. This really trimmed down the number of columns to work with. I was left with columns that have a lot to do with the order of subjects, verbs, nouns, etc. The relevance of these systems in languages is important from what I remember in English so I am satisfied with the columns I have left. I also found a way to plot the latitude and longitude on a world map. I used GeoPandas and Shapely for it. I am glad I was able to do it, because it was the first time I have plotted a map on Python.

June 23, 2022

I managed to complete almost all the goals I had set out for myself. The map is not interactive, but maybe one day I can learn a tool like Tableau that can help with that. I managed to make a simple way to guess the macroarea via the language. At some point, I would like to make this a better system. I learned a lot of new functions, played with models such as KMeans, and managed to plot a map which was color coded in a relevant way. To run the code, place the csv to the same folder in Jupyter Notebooks as the python file. Everything should run well afterwards. Remember to install and import the packages!