

Project Report Draft 1

Arnaud Harmange

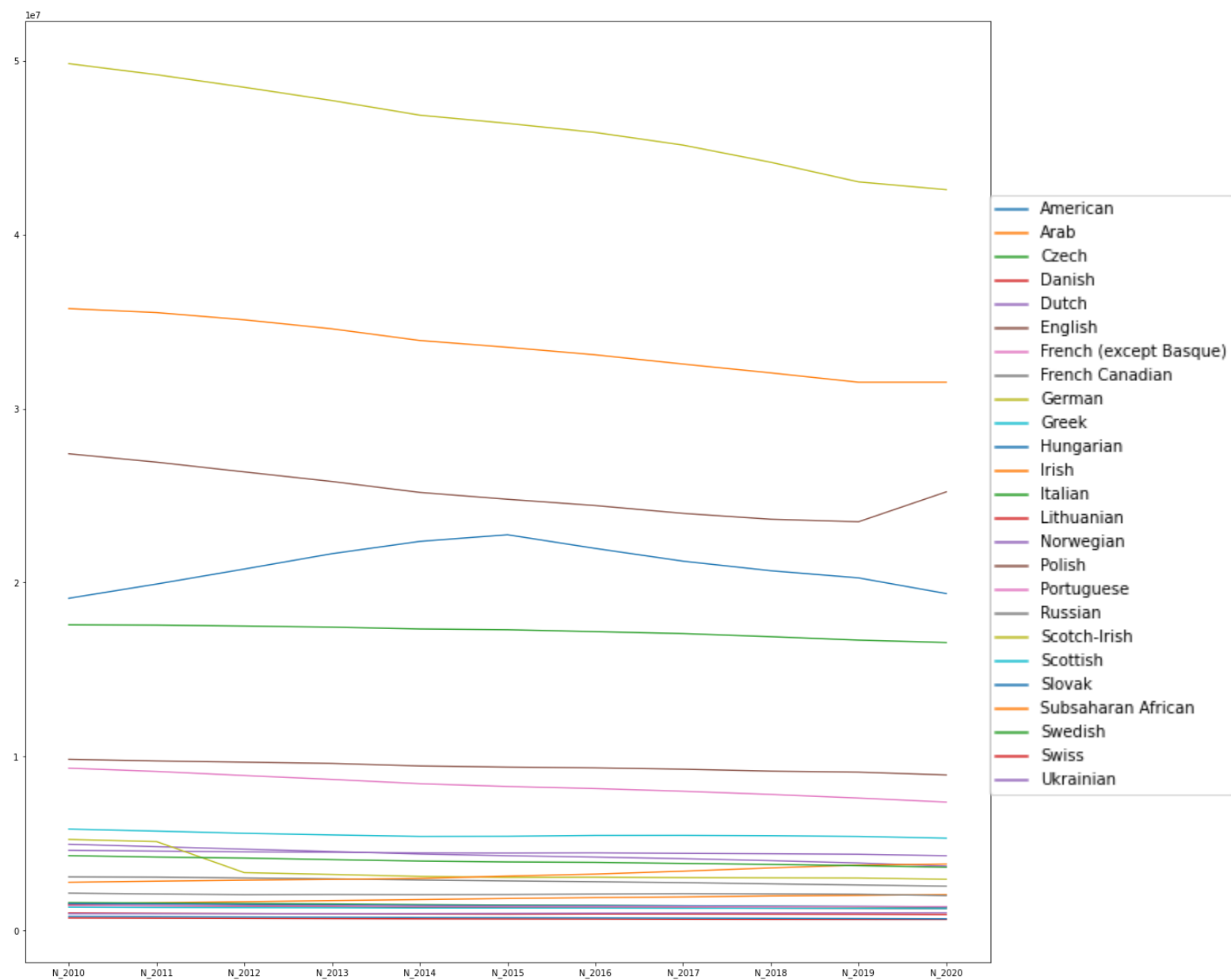
Luke Staib

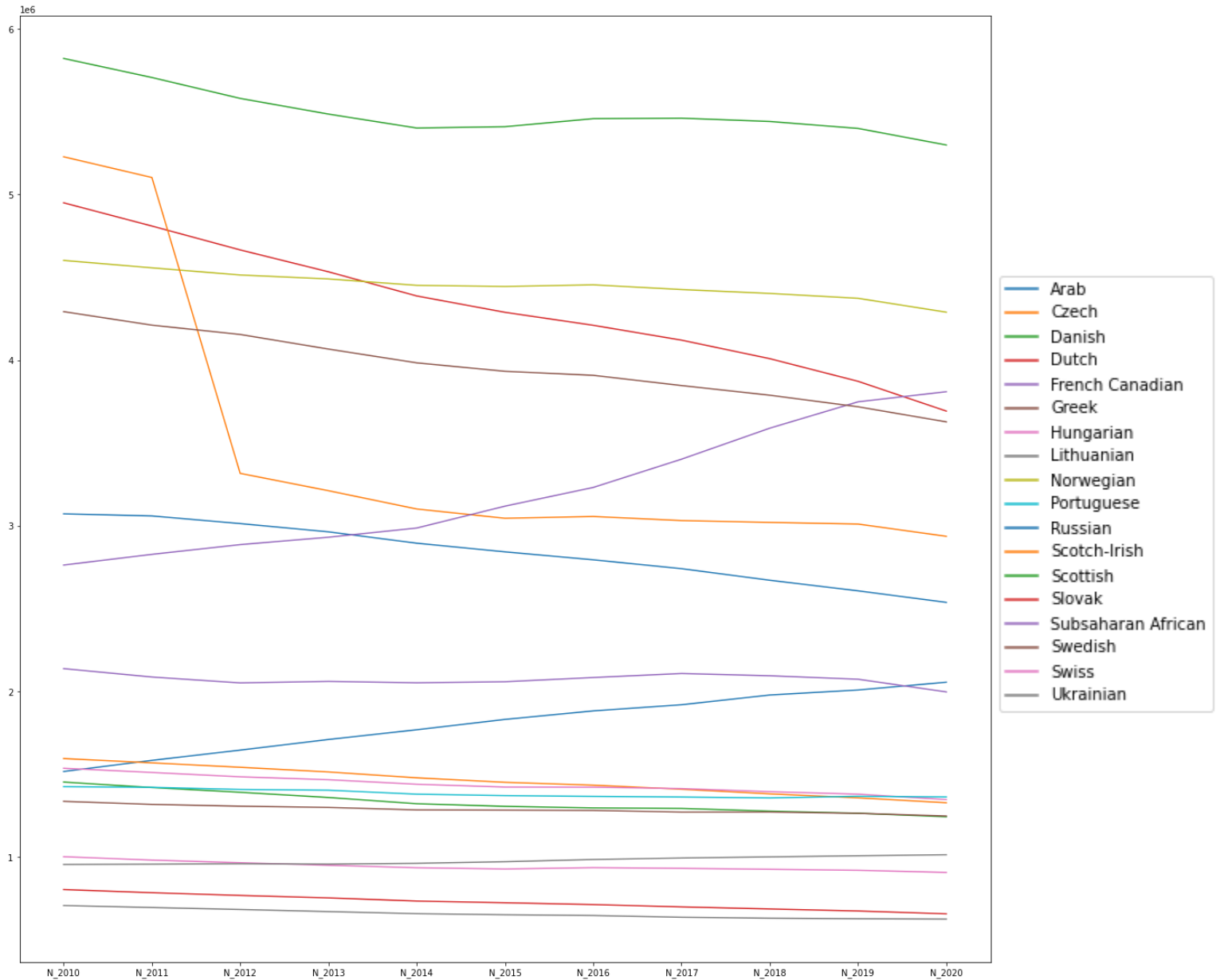
04/18/2022

The United States is often portrayed as a “melting pot” of people and cultures coming together from across the globe, and forming unique communities throughout the country. However, for as long as people have immigrated to the United States from other countries, current citizens and residents of the United States often find reasons to complain about incoming populations. In recent years especially, claims of massive immigration and especially claims of immigration from specific countries have been used to polarize politics and create discord amongst people. This project aims to provide definitive answers to the following questions: What are current immigration patterns globally to the United States? Do these current trends differ from historical trends? Based on historical data, how might these trends continue looking forward? In order to answer these questions, this project will make use of data from the US Census and the US Department of Homeland Security (DHS) that details the origin country of individuals obtaining lawful permanent residence status in the United States from 1820 to 2020. The goal is to then utilize this dataset and create a prediction model utilizing techniques learned in class to help predict possible immigration trends in the future.

In working on this project, a significant amount of time has been spent researching the best means of tracking immigration trends over time. The first instinct was to look into Census data, which should have provided historical data on immigration reaching back past the 1900s. However, while data for the time period of 2010 to 2020 was readily available, it was difficult to find the same information for years preceding this time period. Fortunately, the Department of Homeland Security provided an acceptable substitute for the Census data. The Department of Homeland Security maintains a dataset detailing the origin region in the world from which people who obtain legal permanent residence status in the United States come from. This should allow for a more accurate model to be built, since the dataset will be much larger and should provide a better training dataset than the smaller, ten year span that was covered by the census data.

While the DHS data is more comprehensive from a time perspective, it does lack the resolution that the census data has. The census data provides a much more detailed picture of immigrants' origin, since instead of simply breaking down the immigrant origins by region, the census data breaks them down by country. Despite the census data not covering the same time span as the DHS data, because of its more detailed nature, it can still be used to gain insight into immigration trends from specific countries in the last ten years. Below is a graph of the Immigration trends from the period of 2010 to 2020 according to US census data gathered in that time. Following the first figure is a second version of the same graph, but some of the countries have been removed so that the countries with smaller immigration numbers have more visibility and trends can better be analyzed. The X axis intervals are each year from 2010 to 2020, and the Y axis is the number of immigrants in millions.



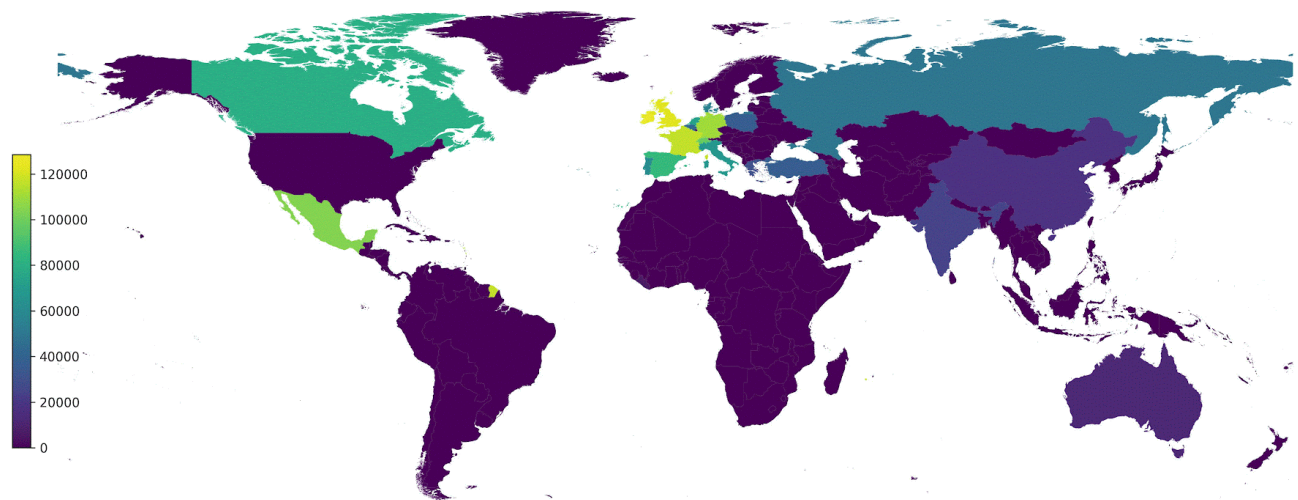


These figures show that immigration to the United States has not seen a massive increase, but rather immigration from most countries has remained fairly constant in the last ten years aside from some exceptions such as immigration from the Scotch-Irish dropping significantly, and immigration from Arabic and sub saharan African countries increasing. These clear visualizations are helpful in understanding the current trends in immigration to the United States, but historic trends can provide further context to this picture. There appears to be a means of obtaining census data like the data displayed above for years dating back to about 1900, but unfortunately the format in which it is made available made it difficult to work with, and

ultimately there was not enough time for the data to be aggregated, parsed, and formatted to work for this particular project. In the future, it would be interesting to revisit these same goals from this project, but aim to utilize the historical census data and see if any notable changes in immigration from certain countries are more apparent, and whether this data would improve the accuracy of the prediction model.

In order to see the trends in immigration over a longer period of time, the DHS data was used. This allows for a visualization of immigration by region to the United States to be created, which should help understand the trends in immigration over time from 1820 to present day. The visualization below was created using geopandas to create the individual images, and PIL to transform the multitude of images into an animated GIF for a more intuitive and engaging visualization. The separate images can be viewed in the project repository along with the animated GIF. The animation shows darker regions around the globe indicating fewer people from those regions acquired legal permanent residence status, and lighter areas on the map indicate regions from which more people acquired legal permanent residence status in the United States. By viewing the animation, it is clear that immigration to the United States over time has changed, and while in earlier years immigration seems to come almost entirely from European countries, Canada, and Mexico, as time progresses, these patterns change. While immigration from European countries and Canada continues to remain relatively high, Immigration from Mexico drops and immigration from China begins to rise. Over time, immigration from Russia, Australia, and some South American countries begins to increase as well. For some time, it appears that Canada, Europe, and Russia remain the largest contributors to United States immigration, while immigration from South America and Mexico remain stable. Over time, interestingly, Immigration becomes less concentrated in certain regions and is much more global, especially from 1960 onwards. From this point on, immigration from all over South America is higher, Immigration from India, China, and Middle Eastern countries increases. Finally, in 1990 to 2020 there is an increase in immigrants from countries in the African continent such as Egypt, South Africa, Ethiopia, and Morocco.

PERSONS OBTAINING LAWFUL PERMANENT RESIDENT STATUS: FISCAL YEARS: 1820 to 1829



The findings and research that we have presented in this report directly answer the questions that we presented earlier in this report: What are current immigration patterns globally to the United States? Do these current trends differ from historical trends? Based on historical data, how might these trends continue looking forward? With the ancestral and immigration graphics that we created, we are able to give a general estimate of historical and modern immigration-related trends. We saw that with the census data we are able to observe and analyze specific ancestral data. Likewise, with DHS data we are able to discover progressions related to immigration data into the US by country.

Project Deliverable 3 (v1 Final Report)

All data should have been collected. All project questions should have been reviewed, answered, and submitted in a written document outlining findings as a PR. You will also be asked to submit the associated data and a README explaining what each label/feature in your dataset represents. Your team should meet with the client before this deliverable.

Checklist

1. All data is collected
2. Refine the preliminary analysis of the data performed in PD1&2
3. Answer another key question
4. Attempt to answer overarching project question
5. Create a draft of your final report
6. Refine project scope and list of limitations with data and potential risks of achieving project goal
7. Submit a PR with the above report and modifications to original proposal

Background, motivation, exploration, and a good part of the analysis should be somewhat complete at this point. Would like to see a rough draft given the information you have so far. There's obviously still time left for analysis but at this point you should have a good 80% of the project completed - enough to make a first draft where maybe some parts are more complete than others.