Immigration Tracking

Group Members: Luke Staib, Arnaud Harmange

Motivation/Goals

Primary Goal: To create a visualization of the movement of immigrants and non-native United States residents over time.

Secondary Goal: To create an ancestry-related prediction model for persons currently living in the United States.

Data Collection

We used data from:

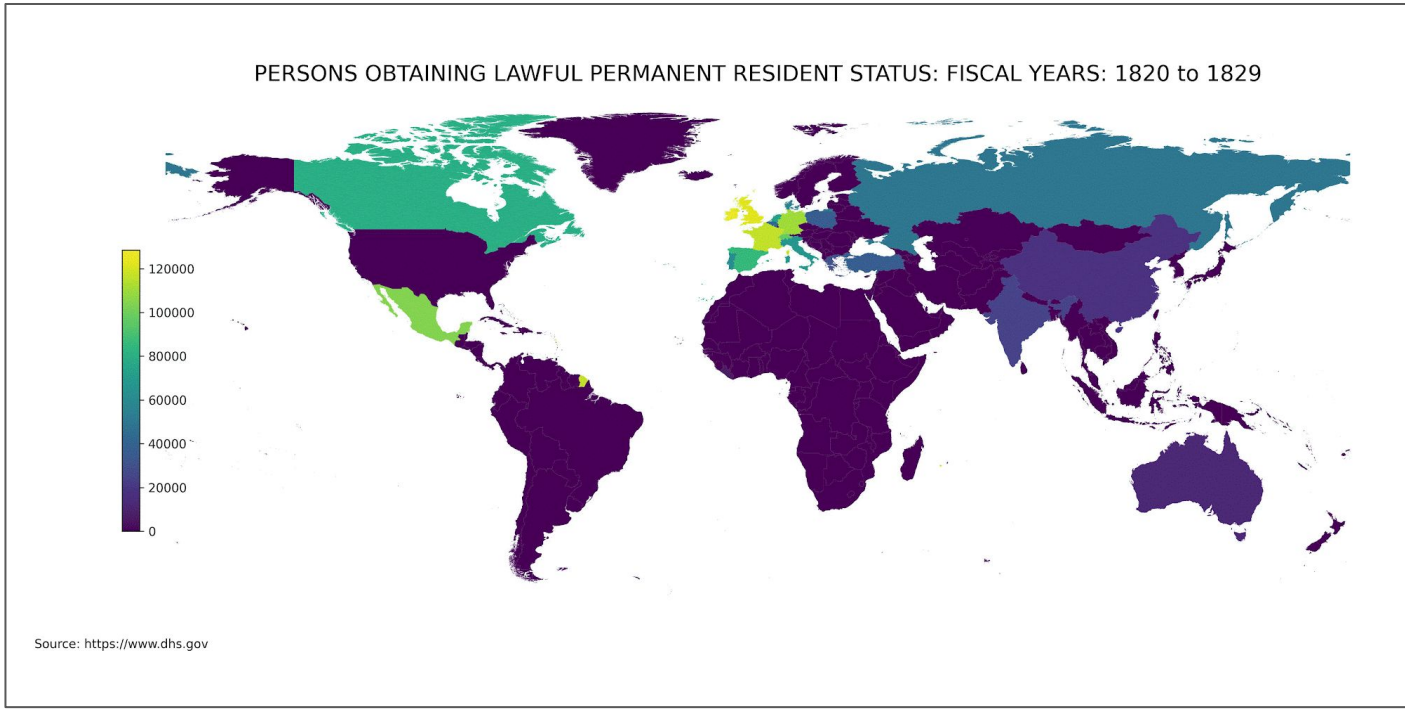
- The US Census: *https://www.census.gov/data/datasets.html*
- The US Department of Homeland Security (DHS): *https://www.dhs.gov/immigration-statistics/yearbook*

Challenges, Limitations:

- Lack of older data, lack of relevant data/missing fields
  - Ancestry is not always included in the datasets that we worked with.
- Missing country data
  - Some countries were not included in the data that we were able to find. Also, it was difficult to deal with data from countries that have disbanded/changed (example: Austria-Hungary).
- Older data (pre-dating 1990) is unorganized

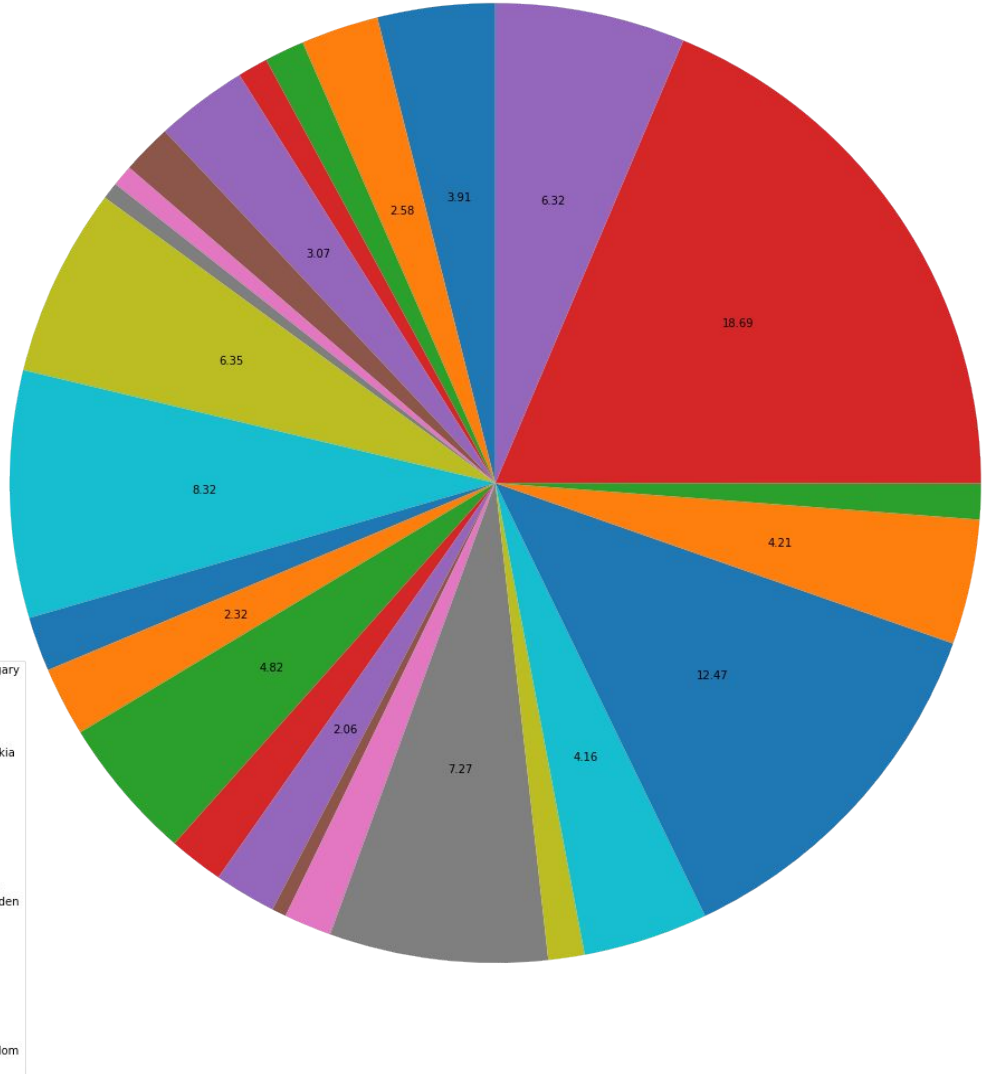
Data Analysis

- Timelapse of people emigrating to the US by country per decade (1820-2020)



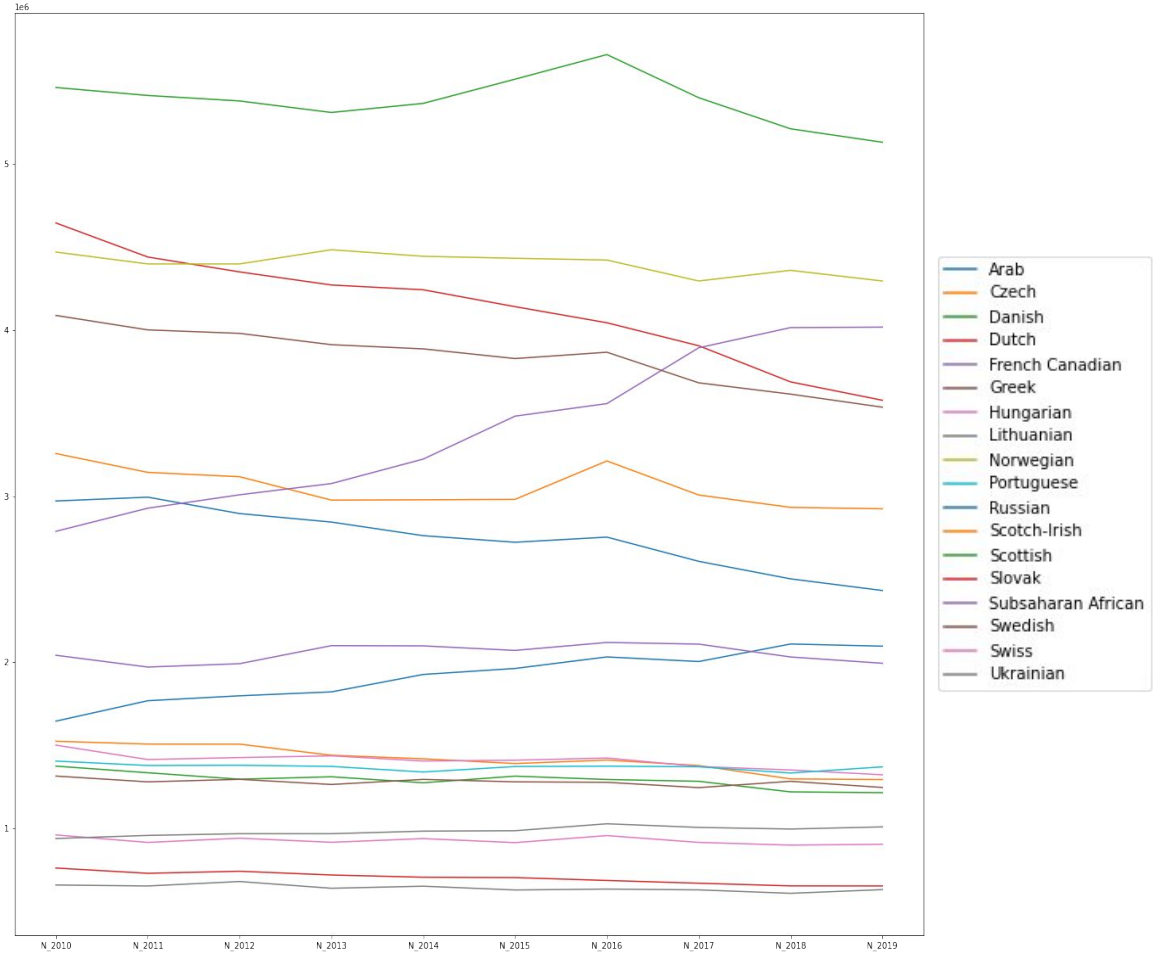
Immigration: Case Study

- We can take the data that we used to create our immigration world map to take a deeper dive into certain geographical areas
- Example: European immigration from 2010-2020



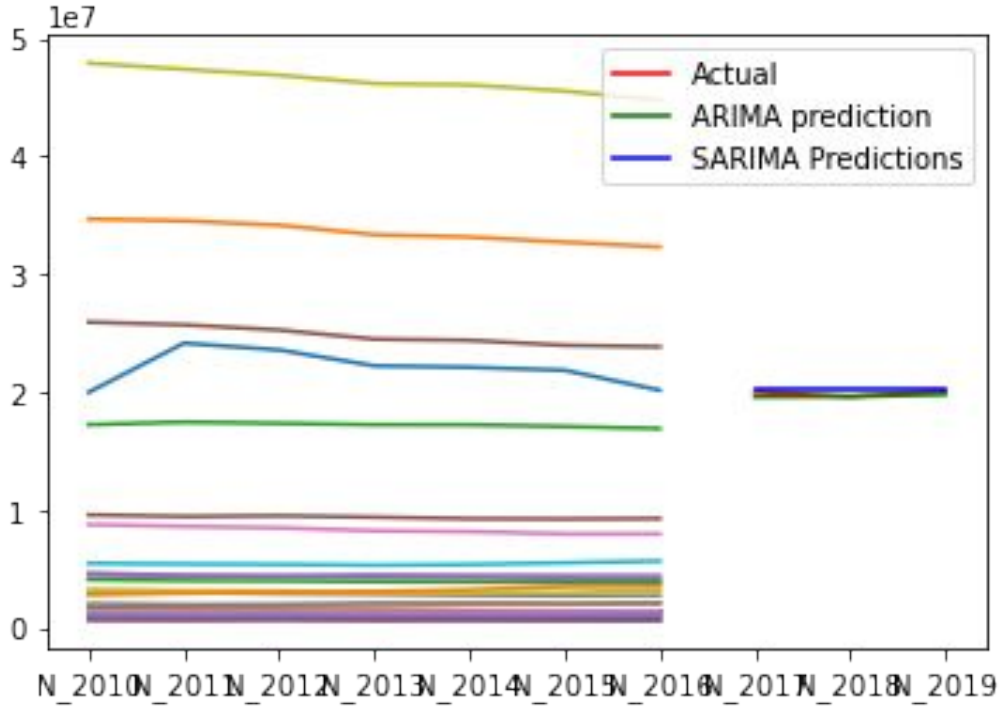
Data Analysis

- Ancestry of people living in the US from 2010 to 2020



Prediction Model

- Our rough prediction model attempts to estimate ancestry based on past years:
- We used ARIMA modeling and SARIMA modeling to work with time-series data



Possible Next Steps

- Improve the ancestry prediction model that we created
- Find supplementary data that we could use to fill in some of the missing gaps related to older data (pre-1990)
- Account more for political and socio-economical events that could skew our model