CS 6630 - Process Book

- **Basic Info:**

  Project Title: Immigration Pattern for United States of America

  Repository: https://github.com/lordawak3n/dataviscourse-pr-WorldMigration.git.

  Team Members:

Name	Email	UID
Ankur Rathore	u0941534@utah.edu	u0941534
Garima Chhabra	garima.chhabra@utah.edu	u1143636
Rishabh Kaushik	rhul840@gmail.com	u1008988

- **Overview and Motivation:**

  The objective of this project is to get a detailed overview of the immigration patterns in the US. It is a well-known fact that the United States of America (US) is a country whose foundation is based on immigration of people from all across the world. With the help of visualization techniques we want to understand the immigration trend in US from all other countries over the course of years. The motivation comes from all of us being immigrants ourselves, who have moved to the US in pursuit of better career opportunities.
  Through this project we plan on accomplishing a way to have a better understanding of how people immigrate to US over the years. This will help in regression analysis for anyone whose business model depends on immigration. This would be achieved by creating representations such as inflows pattern of a country to the US on a per year basis and elements like a trend chart that showcases the pattern of increase or decrease of outflows out of any country to the US. The representation through animation demonstrates the rate of inflow into the US over the years.

- **Related Work**

  While going through works done relevant to the Immigration and Migration around the globe we found out many things that had been done in this regard. A few that we found fascinating included work by Max on All the world's Immigration Visualized in one map. He shows the estimated net immigration (inflows minus outflows) by origin and destination country between 2010 and 2015. We found a video of his work on youtube which showed a zoom-in on the U.S., the U.K., Australia, and Syria, four places where immigration has been closely linked to current events.
  There were a few others (which have been discussed in the initial brainstorm) which had work relevant to this area but this one was the most effective. We thought some ideas were worth implementing and trying out different possibilities with it which could serve as the inspiration for our visualization choices.

- **All About Data And Its Processing**

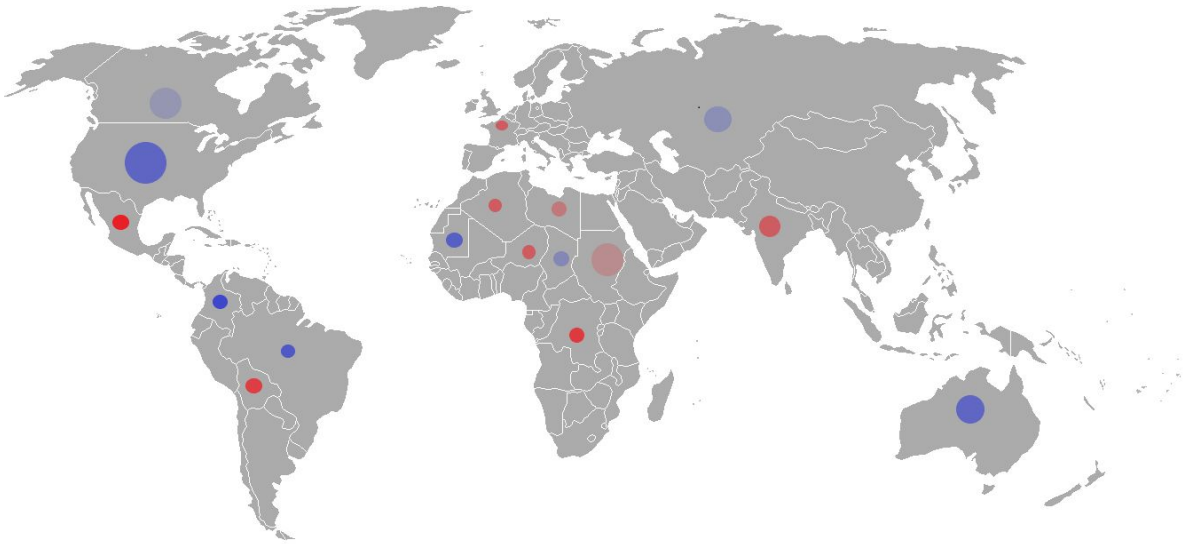
  We will be using the dataset supplied by the Organisation for Economic and Co-operation and development (OECD). It provides the information of inflows of foreign population into the US by nationality. The dataset can be found at: https://stats.oecd.org/Index.aspx?DataSetCode=MIG and then selecting the variable from the dropdown list as: Inflows of foreign population by nationality. The data has to be preprocessed to extract attributes which we want.
  The data is available in a .csv format and did not require a lot of cleanup. However, in order to get data on a per year basis we need to parse the data within an years range. Additionally, our dataset is a composition of many countries so in order to get the inflow data for the US we have to parse and cleanup the data for US only.

- **Design Evolution**

  Brainstorm and initial design:

  Our initial draft establishes the major visualizations elements that will be present in the scene. Although, other visualization exists to capture the immigration inflow for multiple years but none explicitly visualizes the rate of inflow or outflow during a period of time. Thus, we wanted to make sure that our visualization does not just represent rate of inflow implicitly but user should be able to comprehend the difference in inflow during a period of time easily and compare it with other countries at the same time.

  Initially we had drafted the visualization of population inflow as opacity. We felt that it wasn't really effective in establishing movement as a form of rate of flow of people. Although, rate of flow can be visualized as a line chart, it would not be able to capture different rates from different countries at the same time. Also, we wanted our visualization to be on the map so that user can gather secondary information such as geographical location, distance as a factor in immigration etc. at the same time. Since we are trying to visualize change in data for a period of time, we knew there has to be a time slider.



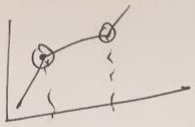
Discarded design 1. Representing net migration by radius and rate of population inflow by changing opacity.

At first, we thought of visualizing net migrating for each country by the radius of the circle and rate of population inflow by changing opacity of the circle as we move the time slider. However, we realized that change in radius is more pronounced than change in opacity so we thought of switching the role. Even with these adaptations, this approach of visualizing had one more shortcoming. With this visualization, user would have difficult time comparing rate of inflow for countries that are not near each other, even if we vary radius to represent rate. Which is why we decided to actually show connecting lines between migrating countries and try to apprehend the rate of population inflow by animating line density or by varying speed of animation.

We also considered visualizing our immigration using wrap up chord diagrams. It was a good choice as it can show both migration and immigration between multiple countries in a single diagram but it suffers from same limitations as line charts and could easily become cumbersome as we increase number of datapoints. We thought of animating the diagram as we change the selected year. Cord's width will then increase or decrease based on the rate of inflow but chord diagram is difficult to grasp as it is and with animations if would make it very difficult to understand. So we decided to let go of this idea and focus on visualizing data on the map.

Discarded design 2. Wrap up Chord Diagram for immigration between limited countries.

(33)

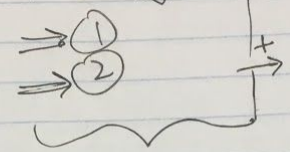


Plan A, Plan B, Plan C

3 weeks + 2 weeks (Data Cleanup)

(US) | country + inflows + outflows

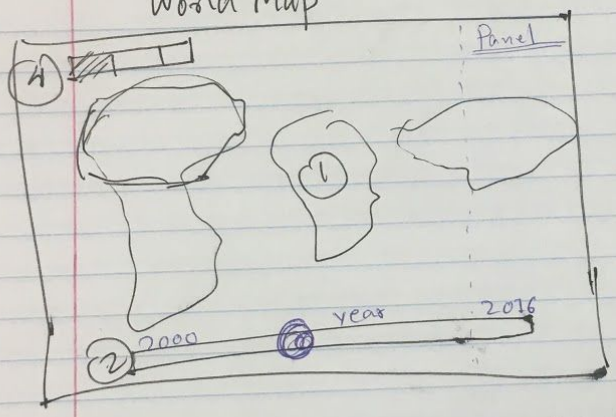
Immigration in
1. US
2. Data Cleanup
3. Animation



Optional
1. Color Continent
2. Bar chart
3. Trend chart
4. Circles

2 ~~weeks~~ weeks

World Map



World Map
1. Map
2. Slider → Year
3. Animation
4. Legend (

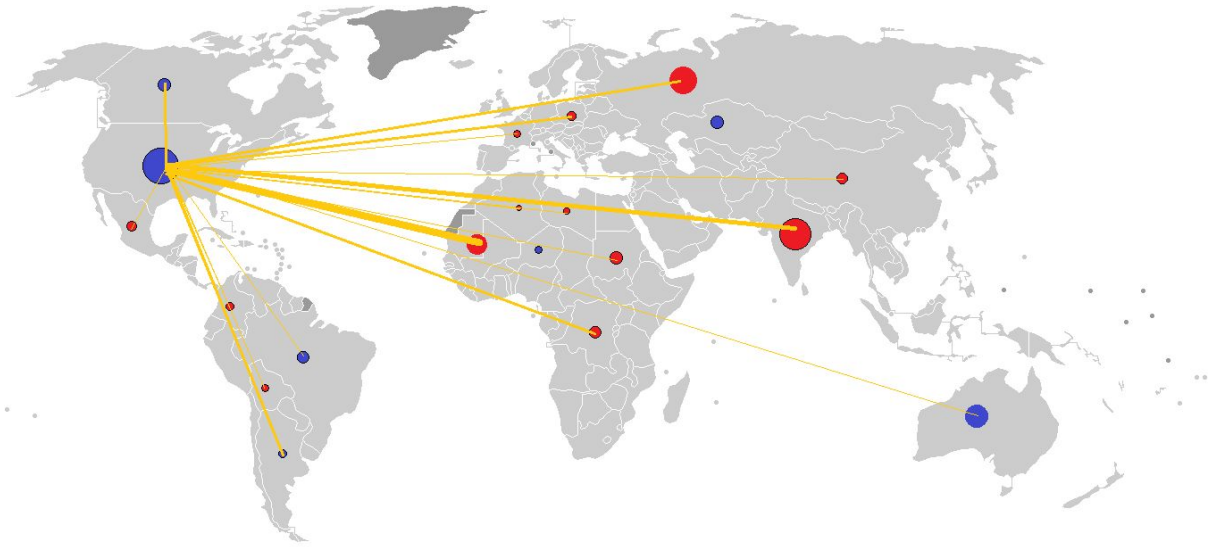
Data Cleanup + Pre-processing of dataset

(UN data)
(~~1980~~ – 2013) ⇐ (1) Get only attributes which we are interested in.
(~~2000~~-2016) ⇐ (2) Oecd.stat

→ US Immigration Visualization from 2000 - 2016
→ Visuali
→ Time Slider ~~which~~ with dynamic animation data.
→

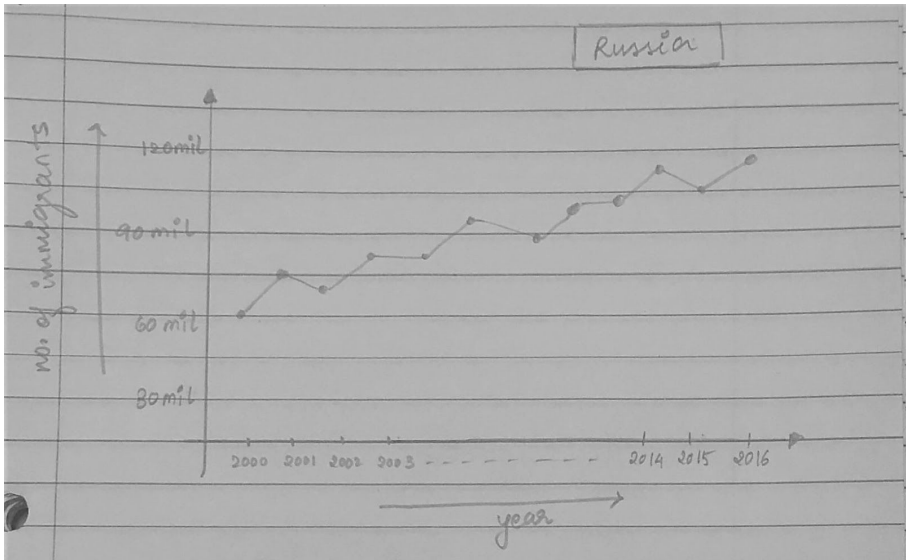
Realisation design:

We decided to actually show connecting lines between migrating countries and try to apprehend the rate of population inflow by animating line density or by varying speed of animation. As seen in the diagram below, the radius of circle represents the net migration for that country on the selected year. The color represents whether population inflow is more than population outflow. As we slide right the time-slider, the lines connecting migrating countries will animate to show the rate of population inflow.

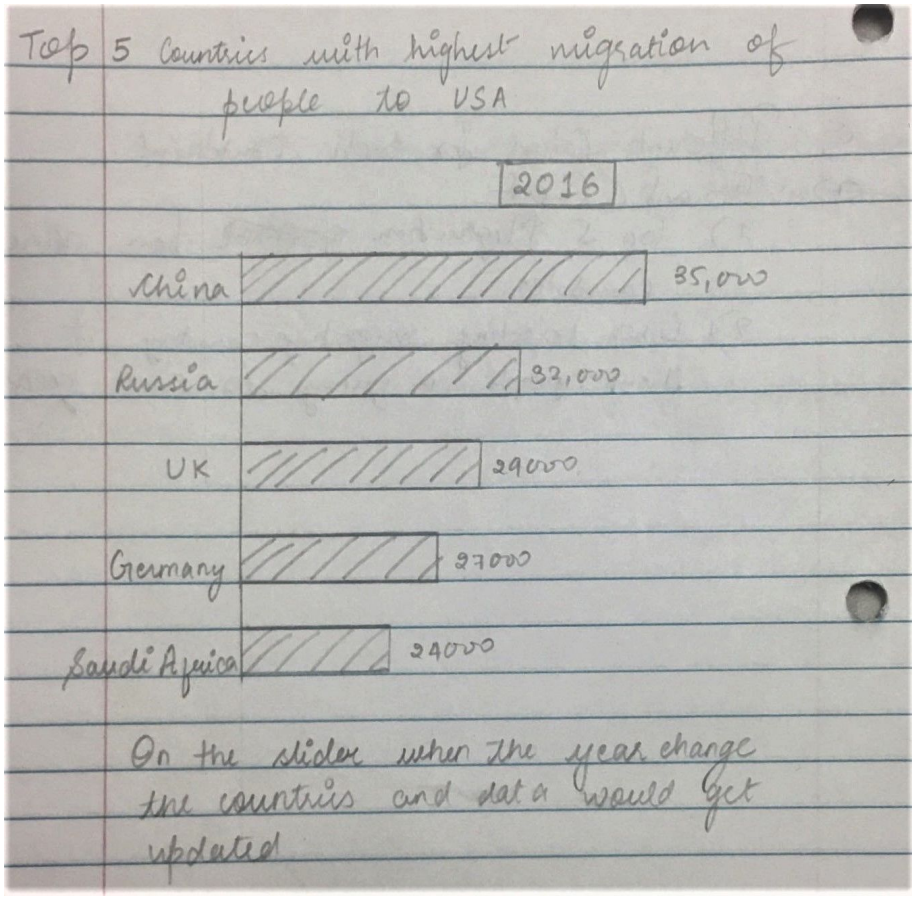


Final Draft.

On hover we can show migrating populating from selected country to United States while on the side we can show trend charts. Line chart would represent the trend of migration for the selected country. The data shown in the line chart would be the trends of migration over all the years for a given country. While the bar chart can show the top 5 countries from which the migration to US has been maximum for the selected year.



For country selected the entire trend across all years is shown.
This is a good way to analyze the change across the years of the migrants going out from their country to USA



- **<u>Implementation</u>**

  For this project our target was to present a atleast a map which visualises the ratio of the inflow and outflow from a country and try to analyse the migration and immigration of different countries across the globe using circles of different diameters and shades of 2 colors whose shade and diameter would represent whether the country is belongs to an immigrant dominant category or migrant dominant category. For this we required a data processing which requires to ratio of total count of inflow to outflow ratio corresponding to each country for every year. But due to the time constraints it wasn't possible for us to processes our dataset into this form and thus dropped the idea of presenting this visualization.

  For now we are presenting a color mapping scheme which would provide the information of the outflow of people from that country to United states.

- **Challenges so far**

  While implementing we faced a few challenges, we started of with data processing in which we decided to have a field relevant to country name and rest of the fields would be the years from 2000 to 2016. What we realized while implementing was that we needed some kind of id which would map our world map drawn using world.json to our data. We had to restructure our data to include the country id in our dataset so as to enable the mapping.

  We obtained the dataset with a lot of unwanted attributes such as variable indices, flag codes, flags, gender, “Outflows of foreign population by nationality”, “Stock of foreign-born population by country of birth” and “Inflows of asylum seekers by nationality”. Moreover the dataset doesn’t have immigration data for certain years. We had to do post processing and cleanup of the dataset. We created a parser in C++ (v14) to extract attributes such as country code (a unique three letter alphabetic code), country name and year wise "Inflows of foreign population by nationality" (from the year 2000 to 2016). For the records which didn’t have immigration data for specific years we have used the value 0. The input and output format is in a .csv format.

  Here is a link to the parser and the parsed dataset: https://goo.gl/d41brS