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.

- **Questions**

  Here are some of the questions and how they have evolved with the progress of the project:

  1. Our visualisation tries to answer the question of how immigrants are moving into United States, at what rate, from what part of country and how has this been changing across a time period from 2000-2016.
  2. Our visualisation also answers the question of which are countries that top the charts in migration.
  3. We also try to analyse the movement of people for all the countries across these years and how that rate has been changing for each country.
  4. Also we try make some interesting observations while visualising the data like one observation was about how people have migrated less from countries that were involved in the Cold War with USA whereas from most of of the world there has been an increase.

- **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.

- **Exploratory Data Analysis**

  ➔ **What visualizations did you use to initially look at your data?**

Our initial idea was to do a world immigration analysis but searching for datasets and evaluating the feasibility we decided to stick to visualising the trends and patterns of immigration in United States. We also had in mind to visualise the the inflow and outflow of people coming from and going to a country. This idea was discarded when we discarded the idea of world immigration because that is essentially what would capture the inflow and outflow. Other than this we thought of implementing the animation to show the flow of people from a country to united states.

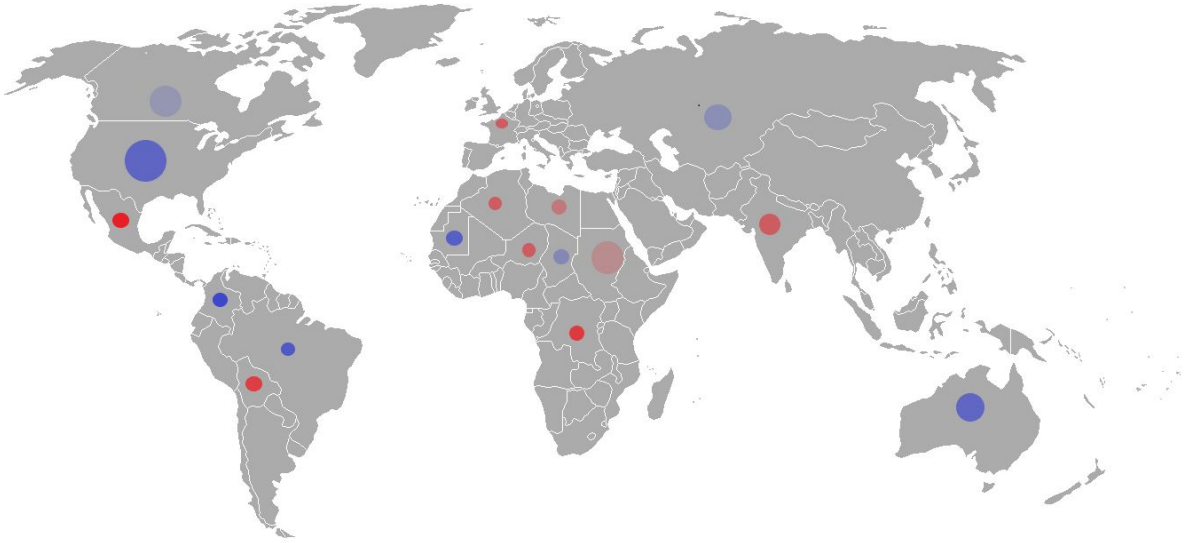
➔ **What insights did you gain? How did these insights inform your design?**
These visualisations helped us in getting ideas for our visualisation. Our animation is inspired by Max on All the world's Immigration Visualized in one map. He shows the estimated net immigration (inflows minus outflows) by origin and destination country between 2010 and 2015. We also got a few other design which made us realise a few things that we should not be doing like the Chord Diagram for immigration between limited countries. We had also thought of using circles as elements representing a particular country and the color of circle determining if the immigration in the country is more than the migration. All these were the initial ideas. Other than this we decide to have trend charts to visualise trends of countries where migration is happening the most and what are the variations for a country across all the years. We decide to move ahead with the bar chart and line chart to represent these ideas.

● 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.

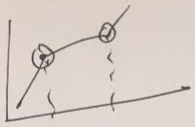
At first, we thought of visualizing net migrating for each country by the radius of the circle and rate of population inflow by changing opacity of the circle as we move the time slider. However, we realized that change in radius is more pronounced than change in opacity so we thought of switching the 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)

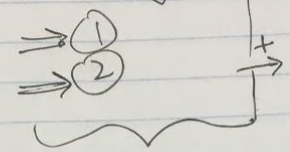


Plan A, Plan B, Plan C

3 weeks + 2 weeks (Data Cleanup)

(US) | country + inflows + outflows

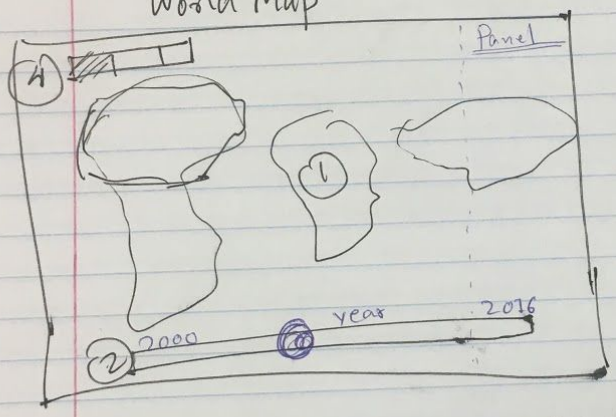
Immigration in
1. US
2. Data Cleanup
3. Animation



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 (

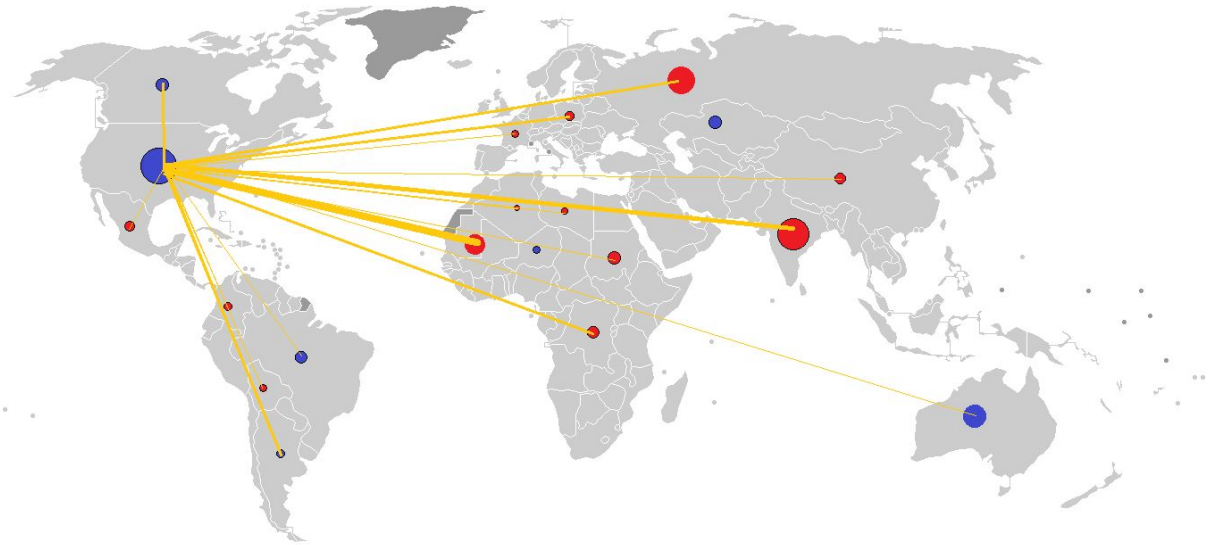
Data Cleanup + Pre-processing of dataset

(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.
→

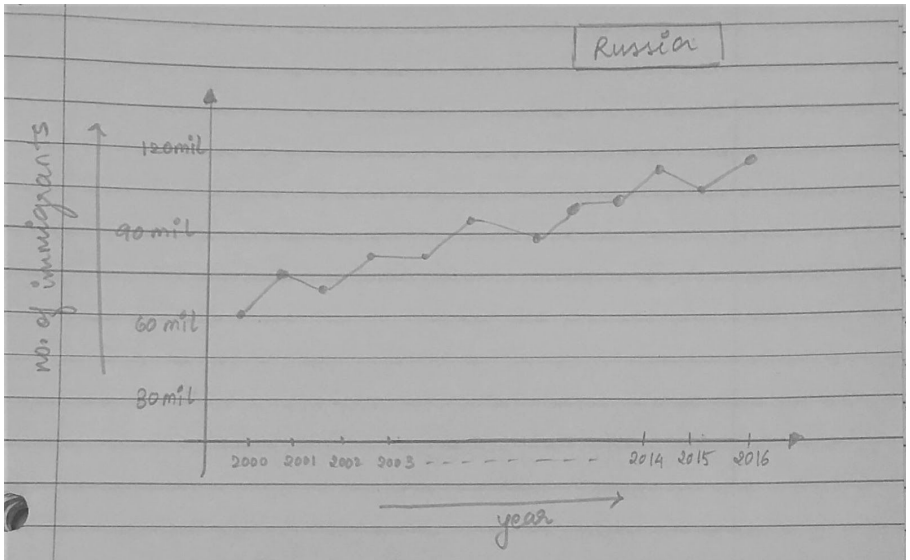
➔ **Realisation design:**

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.

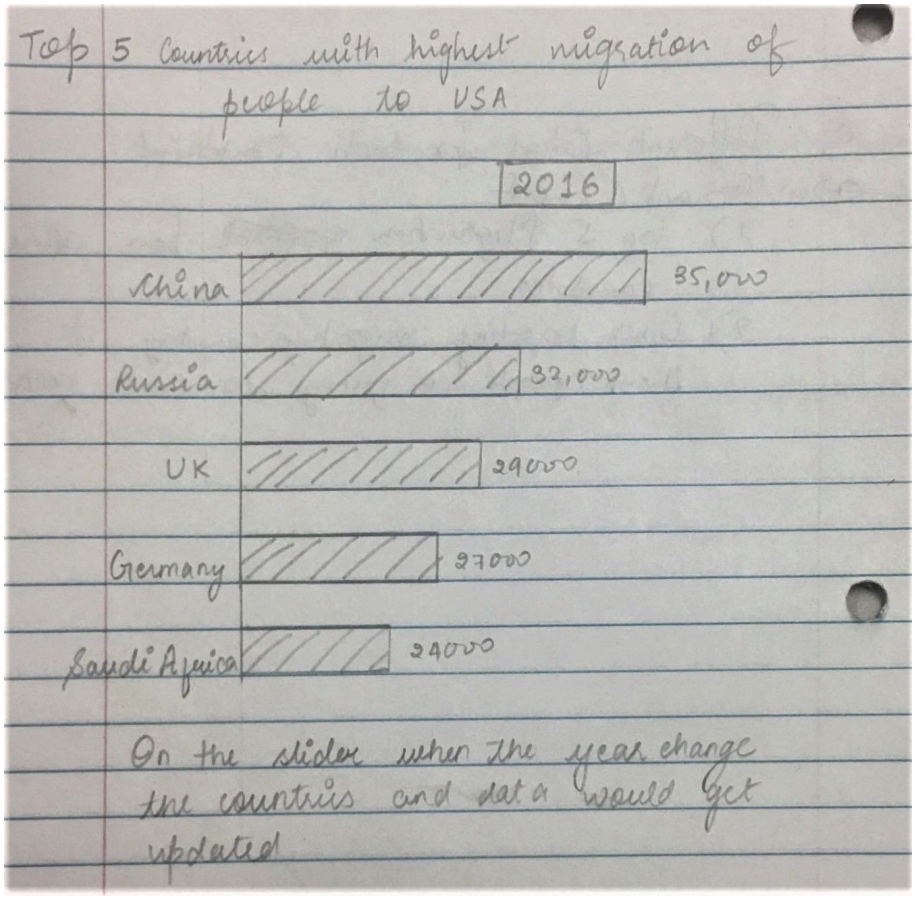


Final Draft.

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



- **Implementation**

  The objective of this project is to get a detailed overview of the immigration patterns in the US. This has been successfully implemented by various features incorporated in our visualisation design. To show the incoming population to USA for a particular year the visualisation shows through animation no. of people (represented as particles) coming into USA. Timeline is the main driver that drives the entire visualisation. Initially we start the visualisation of the 2016's migration patterns. The world map includes the animation for 2016 Immigration patterns and the bar chart shows the top 5 countries from where the maximum no. of people migrated. Initially our line chart represents the trends for the country that has the maximum immigrants throughout all years.
  Following is the detailed description about the implementation of each element-

  1. **TimeLine:** The timeline ranges from the year 2000 to 2016 based on the dataset we have obtained. The slider has a knob that can be used to update the immigration data for all the other elements based on the current year which is selected. We have implemented an year scale to assist the user while sliding the

knob of the timeline tool. The scale reflects all the years from 2000 to 2016. The map, the bar chart and the trend chart get updated when the year value gets changed.

2. **Tool-tip:** The tool-tip is implemented using the **"d3-tip.min.js"** library. The tool-tip gets displayed whenever a user hovers over a country. The information that gets displayed in the tool-tip is the country name and the number of migrants moving to the United States of America in during a selected year.

3. **Data Parser:** The data parser is an external batch file that takes the input csv files with the raw data obtained for all countries. We had to first extract only the immigration data for individuals migrating to the US from other countries. Next we had to clean up and pad the data for countries which do not have immigration data for all the years from 2000 to 2016. The parser outputs the data in a format where the immigration data for each country is present in individual columns ranging from 2000 to 2016. The parser has been written in C++.

4. **World Map:** The world map is implemented using the **d3.*geoEquirectangular()*** projection. We used same projection for transforming animated particles. Although, this type of projection does not give accurate world view around the equator, we were having issues calculating centroids for other projections.
5. **Bar Chart:** The Bar Chart captures the top five countries for which the no. of people migrating for a chosen year is the highest. The numbers in front of the bar represents the exact count of people for the country for a particular year. Initially the bar chart shows the data for year 2016.
6. **Line Chart:** The line chart shows the trends of a chosen country from 2000-2016. The country could be selected from the World Map and the data will update accordingly. The Line Chart also has red colored markers which indicates the countries data for the selected year which gives a better understanding of the change in the following years. Initially the line charts shows the data of Mexico as it is the country that has topped the charts for all the years.
7. **Map Legend:** The map legend is implemented as an svg element and has been added to the World Map section. It signifies the number of individuals from foreign countries immigrating to the US on a per unit basis.

➔ **Following are the images of implementation at various stages-**

1. Milestone 1 Submission

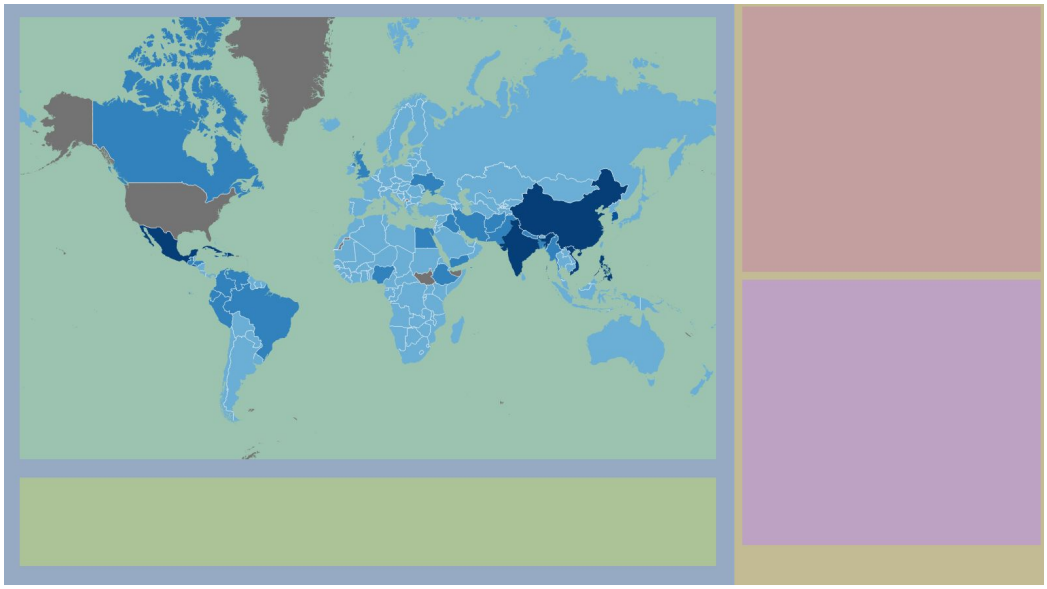


Fig. Milestone 1 submission

2. Implementation including Animation

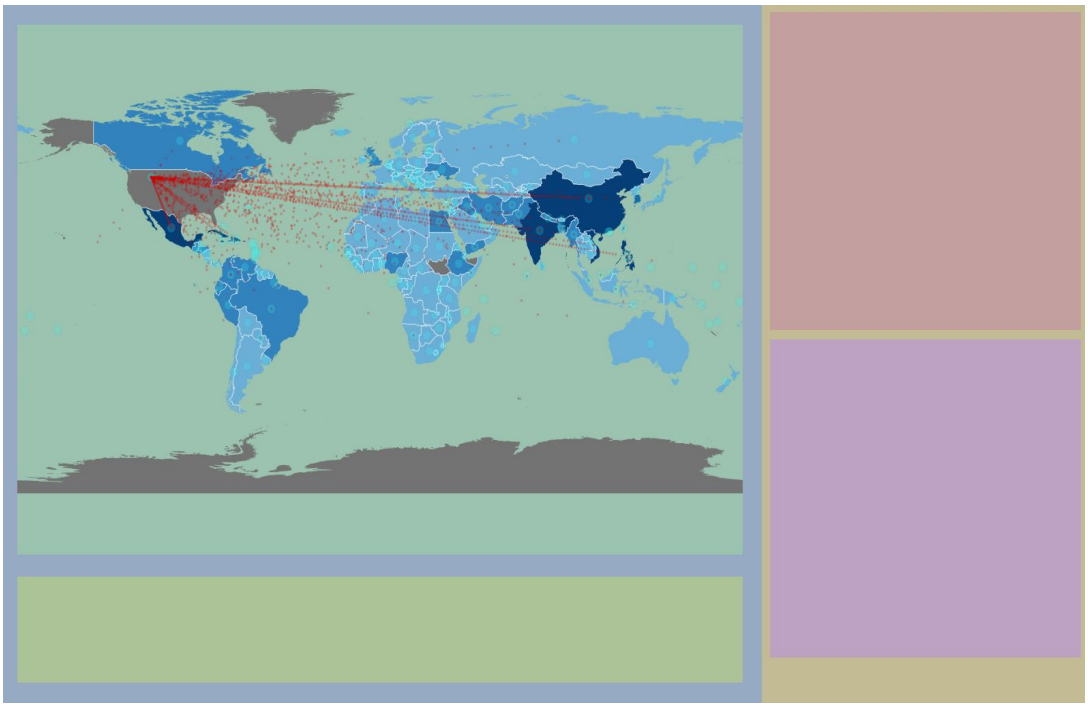


Fig. Animation indication flow of population

3. Timeline Integration

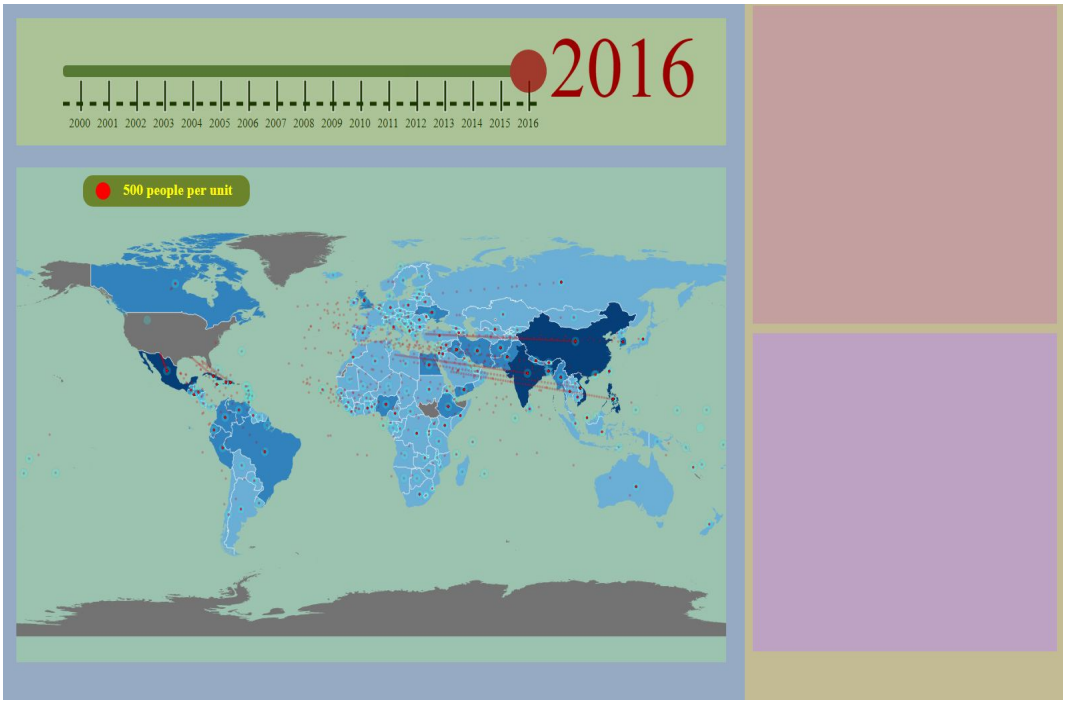


Fig. TimeLine Integration

4. Trend charts Integration

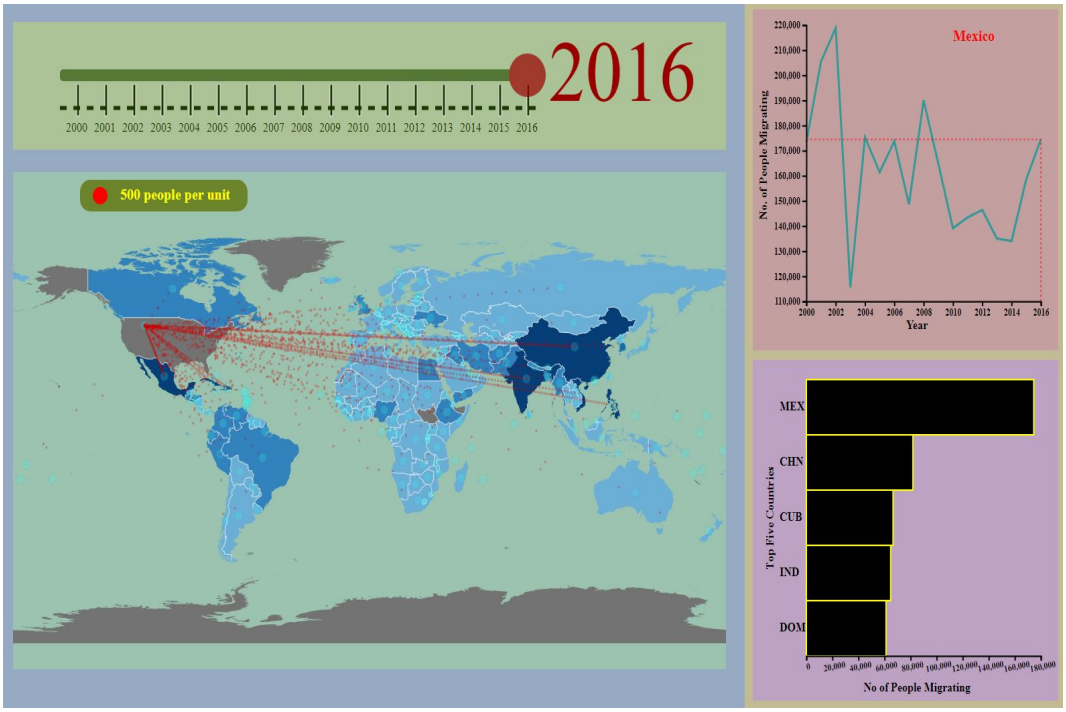


Fig. Line Chart and Bar Chart Integration

5. Final Implementation

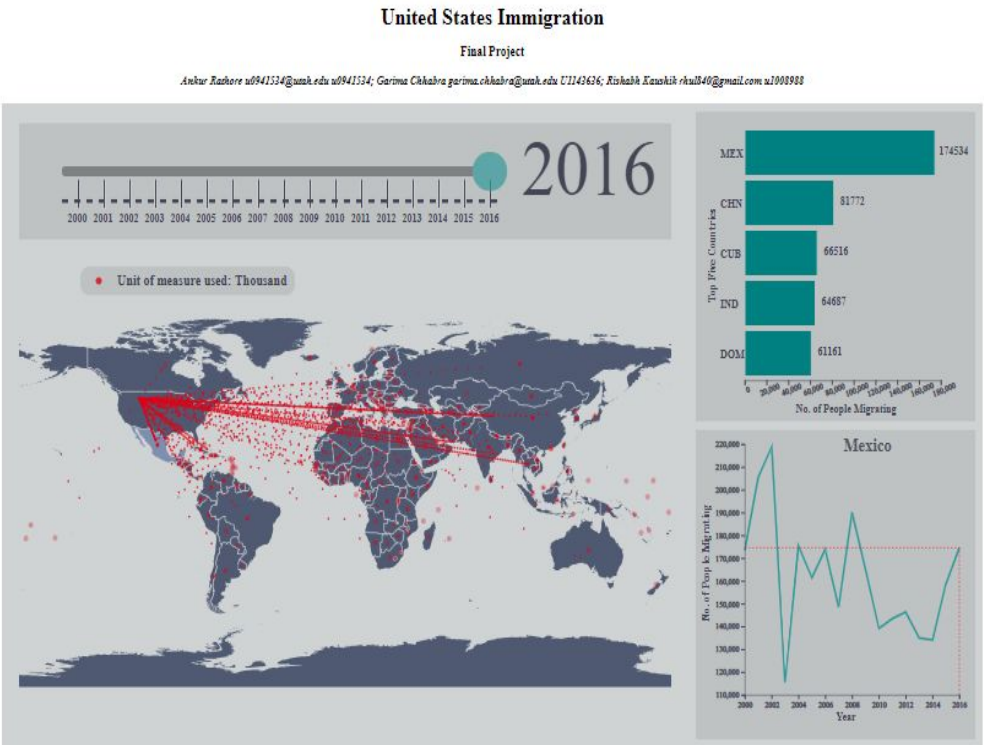


Fig. Final Implementation

- **<u>Challenges so far</u>**

  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

Once we had our initial structure working, we were able to visualize our dataset for year 2016. At this point we did not have functionality to change timescale and reflect the new data set on our world map so we color-scaled the values based on population inflow for each country.

The biggest hurdle was to implement animations to represent inflow from all countries to United States. We started by streamlining the data structure that we are supposed to pass to animation particles. When it came to animating particles, we looked into possible ways of implementing it. We couldn't find dedicated external libraries for animations which could solve our problem so we started looking into available transition and timer features of d3.

Our initial idea was to show particles with some opacity moving from countries to United States representing population inflow. We had to keep some variables constant because if we vary too much at once, the animation might not look consistent. Before we could start the animation we had to figure out the origin of every country. The data that we had had no direct way of getting that information. This suddenly became a huge problem of calculating centroid of convex and concave polygons. Fortunately, we found the answer in one of d3 functions.

We had centroid of all countries available where we could instantiate particles which can translate towards United States with varying velocity. We decided to keep the time that each country takes to reach United States constant and just vary the number of particles that each country will spawn at a fixed amount of time. This way, country with more outflow will spawn more particles in a fixed amount of time in comparison to country with less outflow giving an effect of swinging density of particles forming a path.

When visualizing all countries with one particle representing at least 500 thousand of population, there is a point when the system becomes slightly unresponsive. It usually happens when we instantiate the particles which means it also happens whenever we change timeline because that's when we update all particles. We tried solving the issue by recycling or caching the particles, by not using transparent particles to avoid overdraw issue but it did not solve the issue.

- **Evaluation**

➔ **What did you learn about the data by using your visualizations?**

Here is the list of things that we learned about the data while visualising it-

1. The most important thing that we learned about data was that Mexico always tops the chart in terms of the number of people moving out of the country.
2. Also the top five countries which have the most no. of people migrating mostly remain the same.
3. We also observed that a general trend for all the countries was that the number of people migrating along the years increases but the exception to this are the countries that were in Cold War with USA (example: Japan, Russia etc.).

➔ **How did you answer your questions?**

1. Using animation we were able to capture the the inflow of immigrants to United States from the entire world. Each particle represents upto five hundred thousand people coming to USA.
2. We analyse the countries that have topped the charts of immigration to united states using a bar chart for every year.
3. We have analysed the trends for individual countries using a line chart where the marker indicates the data of chosen country for a selected year.

➔ **How well does your visualization work, and how could you further improve it?**

We were able to achieve all the objectives that we thought of at the initial brainstorming stage, we were able to add few additional features to be make our visualisation better. We initially wanted to generate a visualisation that through the means of animation displays the migration from a country to USA. Also we wanted to visualise some top countries for a chosen year and how those countries changed over time. We were able to showcase this using our Bar Chart which shows the data of top 5 countries which gets updated every time you choose a different year. Data corresponding to every country for a chosen year can be obtained by hovering over it. One can also analyse the trends of a particular country across all 17 years by clicking over the country whose trends would be reflected in line chart with a marker set on year that is being chosen in the timeline.

Ways to improve the visualisation are as follows-

- ➢ A hovering effect could be added on bar chart too so that the hovering over the bar chart for a particular country country would get reflected in the world map. This would be helpful because not everyone in audience will have a good sense of geography.

- We could have categorized the world map by using the same color for the countries belonging to the same continent which would help in analysing the trends across various continents.
- Similar to the above suggestion we can make the color of the particles showing animation from a particular continent same as that of continent as it would assist in the above points evaluation.