INFSCI 2415: Information Visualization

University of Pittsburgh

School of Computing and Information

Final Project Report

Submitted by:

Name	Student ID	Email Address
Goutham Uppalapati	4374886	sru6@pitt.edu
Piu Mallick	4374215	pim16@pitt.edu
Shubhrika Sehgal	4375489	shs2532@pitt.edu
Shruti Gupta	4374956	shg104@pitt.edu

Table of Contents
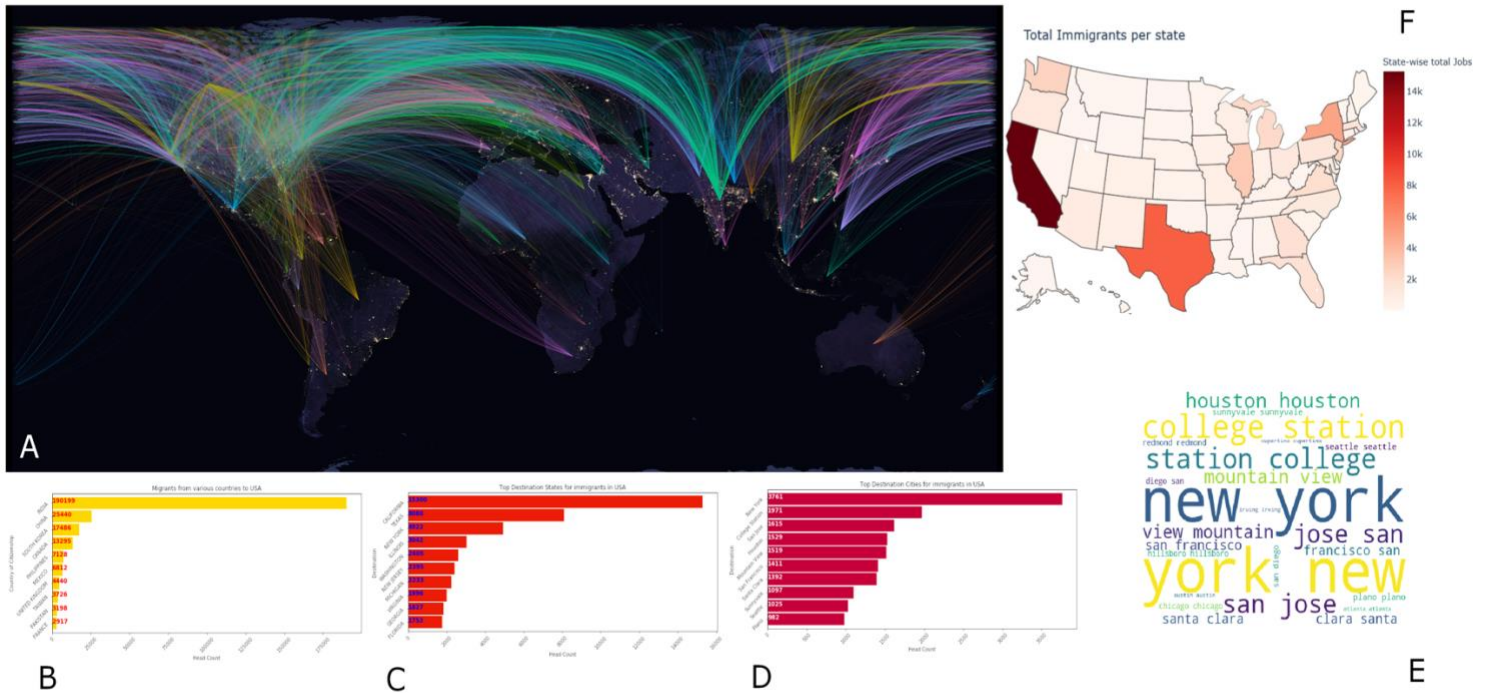
The Crumbling American Dream

Fig. 1: The Crumbling American Dream

Legend

Two datasets in total have been used in this analysis.

I. The main source of data is '***us_perm_visas.csv***', to find people who migrate to U.S. with H1-B visa (more details on H1-B available in '**Background and Significance**' and '**Discussion**' section).

II. The second source of dataset is '***lat.csv***', which contains the location details of the places; latitude and longitude to be specific.

Please see '**Data'** section for more information on how the datasets were used and their respective significance in the analysis and the visualization.

The above figure(s) describe/portray the following:

A. Visualizing a world map in 'night-mode/NASA night lights image' (which is the main figure) with migration path from various countries, keeping U.S. as the final destination.

For each location, it can be seen that a unique travel route has been generated from various countries. The denser the lines, it is an indication of a greater number of people travelling from that country. As can be seen from the figure, few countries like India (green lines), China (yellow lines), South Korea, Canada, etc. has the greater number than other countries. The unique line colors representing each home country have been generated randomly.

B. Envisioning the top-10 countries and their respective count of people migrating to US.

The horizontal bar graph shows a clear pictorial representation (bar line along with the headcount) of the fact that the following countries are leading the race:

India, China, South Korea, Canada, Philippines, Mexico, United Kingdom, Taiwan, Pakistan and France.

The interesting fact being India is way more ahead than other countries when it comes to immigration. For the other countries, a gradual decrease rate can be noticed.

C. Depicting the top-10 states in the U.S. where the migrating people make their destinations.

This graph also shows a pictorial representation (bar line along with the headcount) of the U.S. states where the immigrants settle in for job purposes. California is the states that beats the other states in this sprint; Texas, New York, Illinois, Washington, New Jersey, Michigan and few more come right after the hot favorite California. The graph color (red and blue with white background) signifies the flag color of U.S.

D. Breaking down further from 'figure C', showing the top-10 cities in U.S. which the people choose as their destination.

Figure d is actually a breakdown from figure C, which shows the hot favorite U.S. cities that the immigrants choose for staying in. However, in this case, the ranking of the cities differs a bit from their respective states. New York City (NY) stays at number one position, followed by College Station (TX), San Jose (CA), Houston (TX), Mountain View (CA), and others.

E. A word-cloud explaining figure D further.

The wordcloud in figure E shows the names of the most highly populated U.S. cities with H1-B applicants. Some of the names which appear bold and bigger in size hold the higher population – New

York, Houston, College Station, San Jose, Mountain View and Santa Clara can be seen very clearly in the figure.

F. **A U.S. map showing the total count of immigrants per state**.

This figure is a U.S. heat map which shows the breakdown of the total immigrants (H1-B applicants) by states; the states with darker shades of red depict that they hold the most immigrant population and as the shades of red become lighter, they indicate that the immigration population is less as compared to the darker ones. For example, it can be seen from the figure that California is dark red in color, showing immigration population exceeding 14,000. The second state in the list is Texas, nearing a population of around 10,000 to 12,000. The other states following the leading states, where the population of H1-B applicants ranges from 4,000 to 6,000 are New York, Florida, Washington, Illinois, Oregon, Georgia, North Carolina, Virginia, etc.

Background and Significance

A permanent labor certification issued by the **Department of Labor** (**DOL**) allows an employer to hire a foreign worker to work permanently in the United States. In most instances, before the U.S. employer can submit an immigration petition to the Department of Homeland Security's **U.S. Citizenship and Immigration Services** (USCIS), the employer must obtain a certified labor certification application from the DOL's **Employment and Training Administration** (ETA). The DOL must certify to the USCIS that there are not sufficient U.S. workers able, willing, qualified and available to accept the job opportunity in the area of intended employment and that employment of the foreign worker will not adversely affect the wages and working conditions of similarly employed U.S. workers. That's why, the **H-1B visa program** was launched in **1990**, when **President George H.W. Bush** signed the "**Immigration Act of 1990**." It is intended to help American firms deal with labor shortages in rapidly growing fields that demand specialized skills, such as research, engineering and computer programming.

The research project responds to the ongoing large-scale migration of skilled laborers to the United States of America from across various parts of the world.

Findings

1. Majority of the specialty occupation immigrants are coming from India with China and South Korea being a distant second and third consecutively.
2. The most favorite state of immigration is California with Texas being the second and New York the third.
3. However, city wise, New York (NY) easily tops the chart as the city with highest density of immigrants with College Station (TX) coming up second and San Jose (CA) a close third.
4. In the top ten cities in immigrant destination, even though New York is at top, it is the only city from the east coast making the list. West coast takes the top position with the mention of total six cities with five of them from California. Central takes second spot with three cities all from Texas.
5. The density of immigration of H1B applicants is more in the Western and Eastern U.S. as compared to the Central U.S.

Data

1. The main dataset is available in Kaggle, under the heading US Permanent Visa Applications.
   This dataset has been used in this research with the file name '***us_perm_visas.csv***', which includes the visa information of 3,74,362 people migrating to the U.S. from their birth country or country of citizenship. The data is then filtered out with irrelevant information – dropping out **NaN** (not a number) and **missing values** and concentrating only on the people who migrate to U.S. with H1-B visa.
   The dataset covers information about peoples' US Visa Applications for the years ranging from **2010 to 2017**, and some of the important features include:

Feature/Attribute Name	Description
case_status	The case status of the applicants, like 'Certified', 'Denied', etc. For the research purpose, the 'Certified' candidates have been taken into consideration.
class_of_admission	Admission class of the applicants. For the research purpose, only H1-B applicants have been taken into consideration.

country_of_citizenship	The home country or the birth country of the H1-B applicants.
employer_city	The employer city of the H1-B applicants.
employer_postal_code	The postal code or the zip code of the employer of the H1-B applicants.
employer_state	The employer state of the H1-B applicants.
job_info_work_city	Job information of the work city of the H1-B applicants.
job_info_work_state	Job information of the work state of the H1-B applicants.

2. The second dataset is available in the following website: https://simplemaps.com/data/us-cities
    This dataset has been used in this research with the file name '***lat.csv***' and contains 28889 unique rows.
    It covers the details of U.S. places, like country, state, city, longitude, latitude, timezone, zip codes, etc.

Method

Initial Exploration and Data Pre-processing (in **Python**):

1. Checked for the null/NaN values and missing values and dropping them.

2. Filtered out the data of the immigrants who are H1-B applicants and have their case status as 'Certified'.

3. Defined a Python function called '**geolocate**', which inputs city and country, or just country and returns the latitude/longitude coordinates of either the city if possible, if not, then returns latitude/longitude of the center of the country. '**geolocator.geocode**' does the trick here.

4. Obtained the latitudes and longitudes for the home countries.

5. Merged the geocodes (information about latitudes and longitudes of the home countries obtained from Step 4) with the original dataset.

6. Split the latitudes and longitudes into different attributes.

7. Studied the whole dataset carefully and finding out intricate details like:

    -   Checked for the overall distribution of the people from home countries travelling to USA. (using **matplotlib** library)

    -   Checked for the overall distribution for the destination states and cities in USA. (using **matplotlib** library)

- Built a word-cloud to get a clear picture for the destination places in USA. (using **wordcloud** library)

- Showed a map view for a better pictorial representation of the immigrant counts per US states. (using **Choropleth** map, a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map).

Building the main map (the world map) in **R:**

1. Narrowed down to 6 main attributes from building the main graph (plotting the travel lines from home country to destination country). The main attributes are:

Attribute Name	Description
country_of_citizenship	the home country of the immigrants
job_info_work_city	places where the immigrants are employed
home_lat	Latitude of their home country
home_long	Longitude of their home country
lat	Latitude of the destination country, i.e. USA
lng	Longitude of the destination country, i.e. USA

2. The new dataset contains 60,000 rows approximately, which means there are the same number of travel lines in the main map.
3. Loaded the necessary R libraries like – **geosphere, tidyverse, magrittr, ggplot2, grid, jpeg, dplyr**.
4. Downloaded the NASA night lights image and loaded the picture and render.
5. Counted for each unique connection and ordered according to importance.
6. Created a complete dataframe with the points for all the lines.
7. Finally, plotted the graph with **ggplot** and saved the image.

Discussion

"**What is the status of the current U.S. H1B visa program and what can we look forward to in the new year?**" – This is the most buzzing question going around right now as the Trump Administration is trying hard to enforce new and stringent laws to the U.S. H1B visa program.

There is no denying the fact that foreigners view the H1B visa as the golden gate to furthering their career, starting a family and building a new life in America. The visa is particularly popular because people can also transition from it to a green card. Many applicants are international students trying to transition from F-1 student status to H1B work visas on route to a green card. According to the generic statistics, each year, the **U.S. Citizenship and Immigration Service** (USCIS) opens a total of 85,000 such visas — 65,000 for those applicants who have a Bachelor's degree or equivalent, and 20,000 for those obtaining a Master's degree or higher from the U.S.

Hence, it becomes worth considering the fact where individuals have passed through this H1B visa gateway and what they have contributed to American society. Nearly 50% of the Fortune 500 companies were founded by the 1st or 2nd generation immigrants please see '**Data'** section for more information. According to a recent report, U.S. tech giants like Amazon, Microsoft, Google and Intel are among top 10 employers of approved H1-B applicants. U.S. educators have established without doubt that the entry of foreign-born scientists and engineers remains an unmatched source of "strength and vitality" for the country's tech industry. Under these circumstances, one would expect the American government to support the H1-B visa program. Instead, it is trying hard to introduce a number of policies that appear to impede its effectiveness.

One thing has happened for sure - securing a work visa for foreign nationals accomplished in the math and engineering fields fit for high tech has become more difficult under President Trump's executive order "**Buy America, Hire American**". The policy directs the Department of Homeland Security to issue H-1B visas to only the most-skilled foreigners or highest-paid beneficiaries. According to new data acquired by the National Foundation for American Policy (NFAP), the USCIS has begun to increase H-1B visa denials and the number of Requests for Evidence issued to H-1B visa applicants. According to an NFAP source, "***Employers report the time lost due to the increase in denials and Requests for Evidence has cost millions of dollars in project delays and contract penalties, while aiding competitors that operate exclusively outside the United States***." According to an article published in Forbes, it can be noticed that the H1-B denial rate has surged to a greater extent, particularly in the Trump era. The figure below depicts this fact very clearly.

Table 1: Denial Rate for H-1B Petitions for Initial (New) Employment

FISCAL YEAR	DENIAL RATE
FY 2019	24%
FY 2018	24%
FY 2017	13%
FY 2016	10%
FY 2015	6%

Story

We have researched and found out earlier that people have migrated to the U.S. from various parts of the world for better job opportunities, better life and hoping that they would have live the "**American dream**". The facts and figures, shown above, also prove the same. Even, we as current graduate students (being STEM students and holding Master's degree from a reputed university in U.S.), also aspire to find better employment opportunities after completing graduation and live the similar dream. U.S. has always welcomed talented people across the globe which in turn boosts the U.S. economy and as a result can claim itself to be world's number one country. However, looking at the current situation, we can say that the current situation may not be as good as it used to be earlier and the migration of people to U.S. might have affected to a certain extent. We need to wait and watch what the American government unfolds for us in future!

References

1. **On H1B visa details**: https://www.forbes.com/sites/andyjsemotiuk/2019/01/02/recent-changes-to-the-h1b-visa-program-and-what-is-coming-in-2019/#17c6fab14a81

2. **On H1B denial rate statistics**: https://www.forbes.com/sites/stuartanderson/2019/10/28/latest-data-show-h-1b-visas-being-denied-at-high-rates/#7559175354c3