An Analysis of U.S. PERM Immigration Data with an Emphasis on Employment Trends

Cole Stenger, Alex Swenson, Tong Wu
University of Nebraska at Omaha
Introduction to Data Science

September 28, 2017

Abstract

Immigration is a major component of American history and still proves to be a divisive topic today. While some immigration debates revolve around customs and culture, many others focus on jobs and the workplace. The Office of Foreign Labor Certification (OFLC) within the US Department of Labor publishes raw data that provides many details about foreign workers in the country. For example, in 2015 there were approximately 1 million people who obtained lawful permanent resident status. Two of the most useful sets of data provided by the OFLC are PERM and Prevailing Wage. We intend to explore this data and uncover insights that will shed some light on the status of current immigration. Additionally, the US Department of Homeland Security and the Refugee Processing Center publish valuable data. Analysis of the data from these two sources will reveal even more intriguing details and complement the PERM and PW findings. Overall, this project aims to provide some understanding of where immigrants are coming from, where they are settling, and what jobs they are doing there.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Warning: package 'stringr' was built under R version 3.4.2
```

Introduction

Immigration is an incredibly broad topic which continues to be heavily scrutinized in today's society. One aspect of immigration which is given considerable attention is its effect on the labor force. The Office of Foreign Labor Certification (OFLC) and the Department of Homeland Security are among several government agencies which make raw immigrant data available for public analysis. While there are countless ways to approach and learn from this information, we have chosen to focus on the data connected to employment. This will give some perspective on the current state of foreign workers in the US and characteristics of their employment.

PERM Data

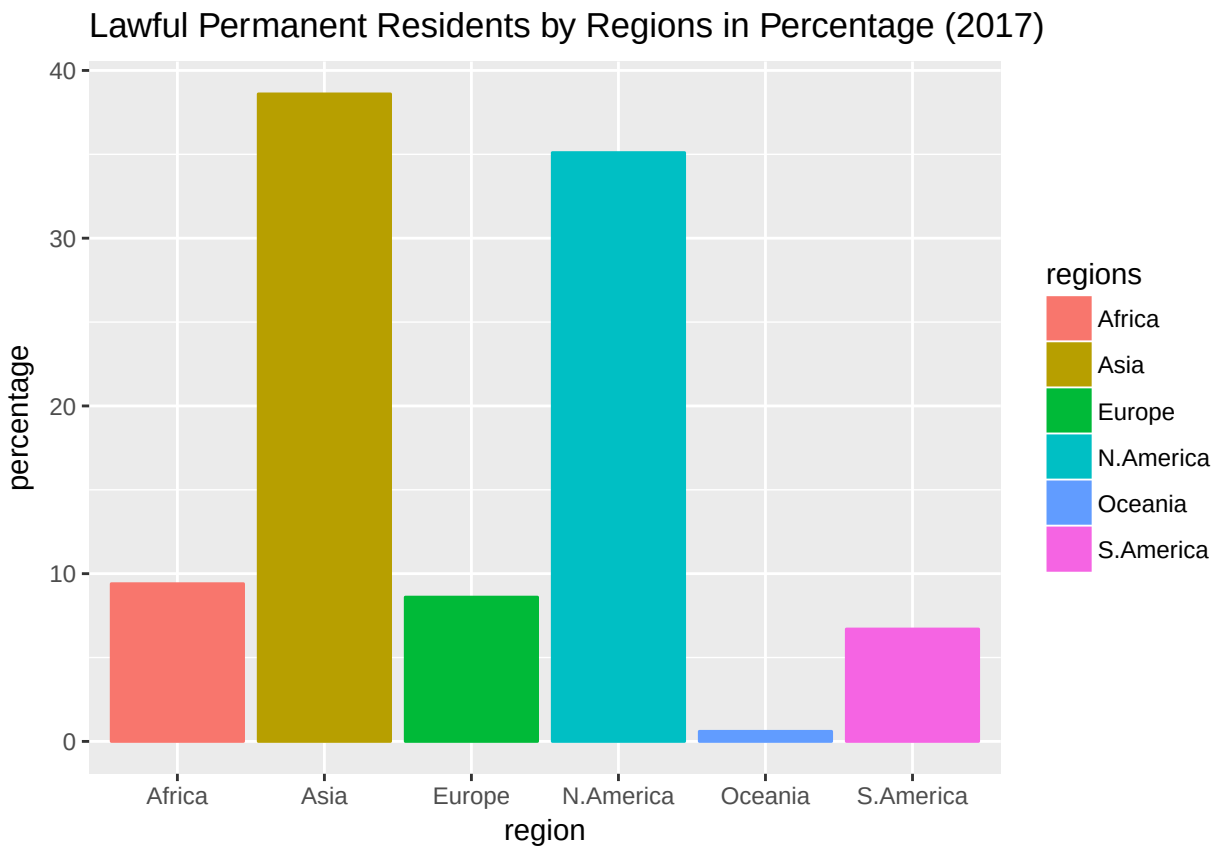
One of the main data sets available from the OFLC website is the PERM data file. This data contains information about individual foreign workers who are filing for permanent residency status in the United States. An immense amount of information can be found within the PERM data such as employment

location, country of origin of the worker, and certification status. In the sections below, analysis of various aspects of the data can be found along with a description of what was done.

Regional Analysis of PERM Approvals

Information related to a foreign worker's home country or region can be very interesting and insightful. The plot shown below represents what percent of lawful permanent residents come from each region of the world. It is intriguing because we can see that the Asian residents and North American residents represent the two highest percentages. There are comparitvely less LPR's from Europe. This data is from the Department of Homeland Security and relates to the PERM data. Plotting this data was straightforward and just required making a bar chart with ggplot.

```
regions <- c("Asia","N.America","Africa","Europe","S.America","Oceania")
y1 <- c(38.6,35.1,9.4,8.6,6.7,0.6)
df1 <- data.frame(regions,y1)
# We were unable to get the excel file to read correctly so we had to resort to making our own data fra
ggplot(df1,aes(regions,y1,color=regions,fill=regions)) + geom_bar(stat = "identity") + xlab("region") +
```

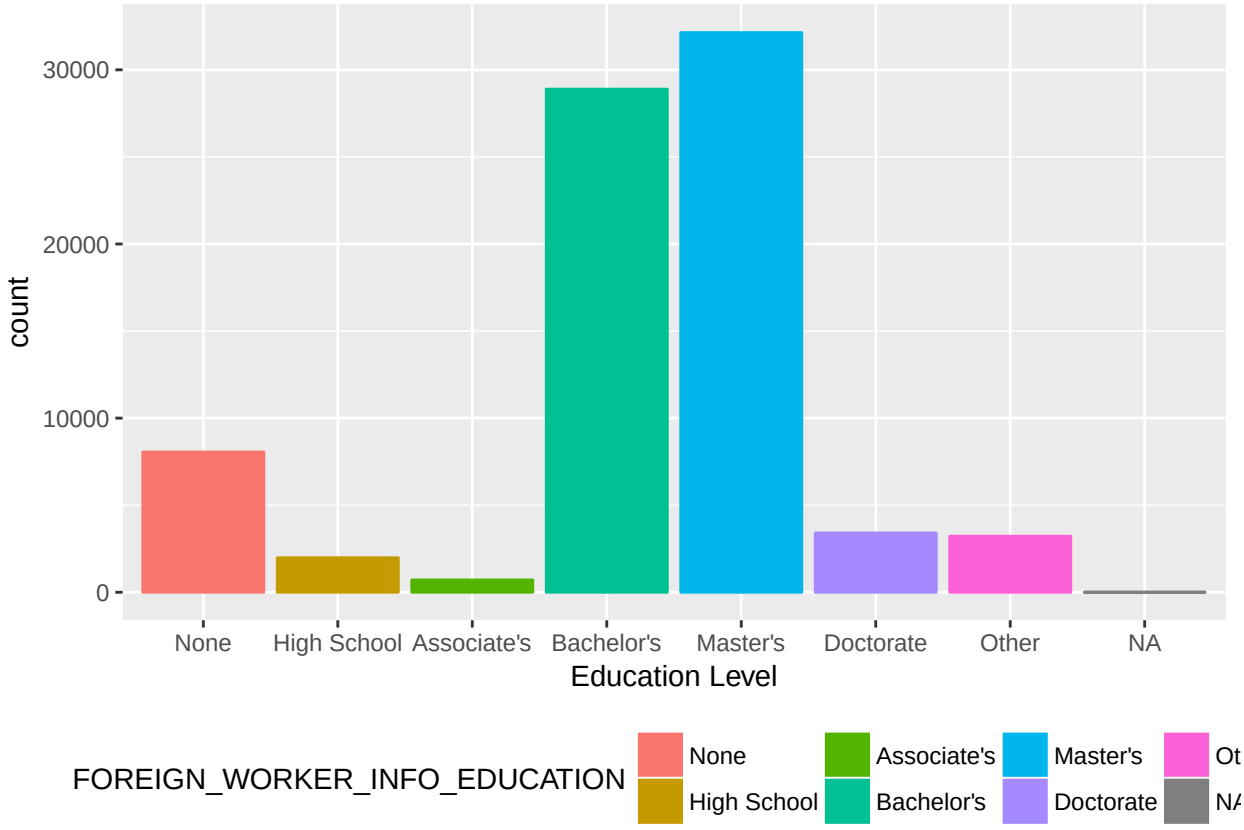


Education Level of Foreign Workers

While immigration is certainly a fixture on the news, the education levels of the foreign workers are rarely discussed. This valuable information can be found in the PERM data from the OFLC. As we can see in the following plot, a significant amount of the foreign workers applying for permanent residency possess a bachelor's degree and/or a master's degree. This has ramifications on what types of jobs these workers will

seek and hold. The only manipulation of the data that was done was adding levels to the education column. This was done to improve the ordering of the graphic.

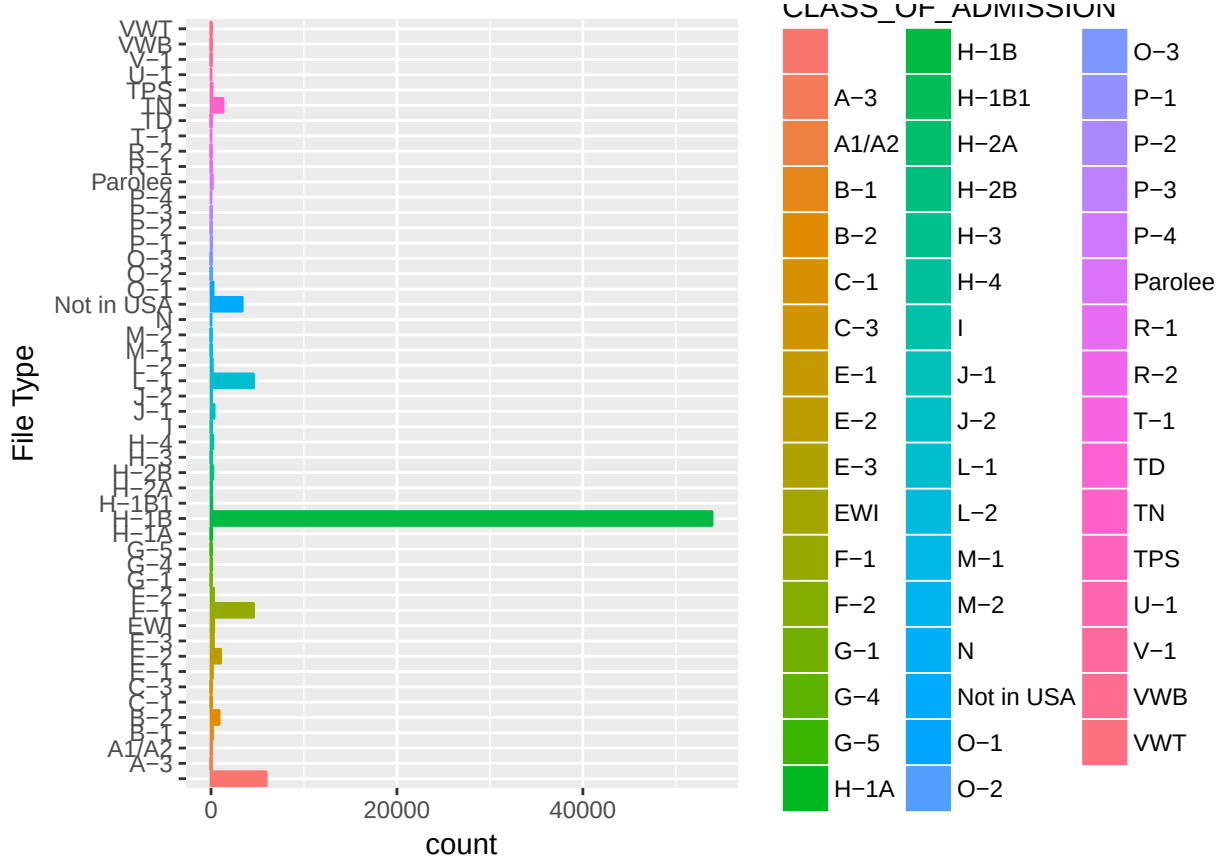
```
myDat$FOREIGN_WORKER_INFO_EDUCATION <- factor(myDat$FOREIGN_WORKER_INFO_EDUCATION, levels=c("None", "Hig
g2 = ggplot(myDat,aes(FOREIGN_WORKER_INFO_EDUCATION,color=FOREIGN_WORKER_INFO_EDUCATION,fill=FOREIGN_WO
g2+geom_bar() + xlab("Education Level") + theme(legend.position = "bottom")
```



Path to Application

Many applicants for permanent residency are already within the United States. This is another area of interest which does not receive enough attention. The PERM data records whether the applicant is already within the country or not. There are numerous ways that the applicant may have already been certified to work in the US and those are displayed in the graph below. As we can see, there are far more PERM applicants with a previous H1B certification than any other group. H1B is discussed in more detail below. Creating this graph simply required reading the data from the file and creating a bar chart of the class of admission column with ggplot.

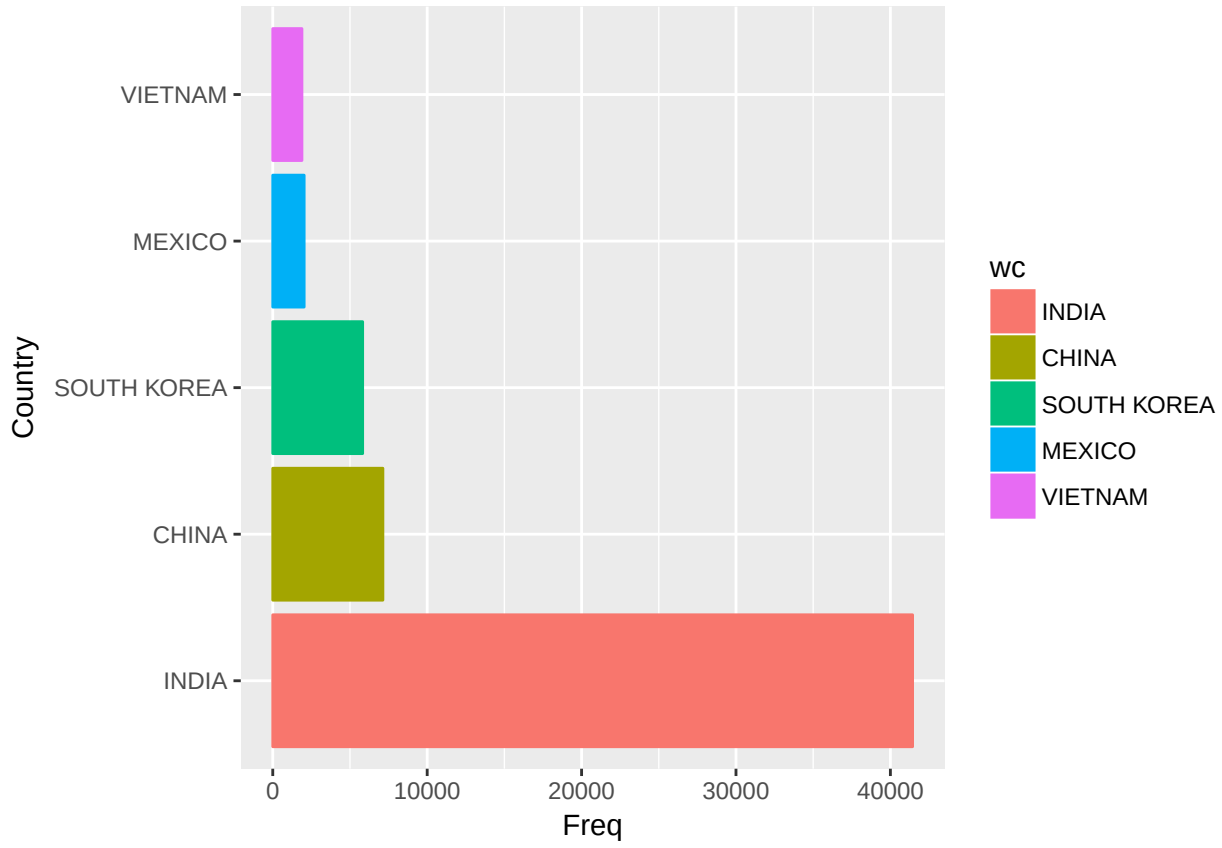
```
g5 = ggplot(myDat,aes(CLASS_OF_ADMISSION,color=CLASS_OF_ADMISSION,fill=CLASS_OF_ADMISSION))
g5 + geom_bar() + xlab("File Type") + coord_flip()
```



Origin of Foreign Worker

As was mentioned earlier, seeing where foreign workers are primarily coming from can be important and insightful. A column within the PERM data details the applicant's country of birth. Several things were done to the data before it was plotted. Initially, the column was read and used to create a frequency table. After that, it was sorted and the 15 most frequently recorded countries were stored in a new object. Then, a data frame was created using the 5 most frequent countries. Lastly, this data frame was used in ggplot to create a bar chart. Interesting observations from this include the fact that India has such a large number of applicants compared to all other countries and that four of the top five are Asian countries.

```
wc = myDat$FW_INFO_BIRTH_COUNTRY
wc2 = table(wc)
wc3 = sort(wc2,decreasing=TRUE)[0:15]
wc4 = data.frame(wc3[0:5])
g4 = ggplot(wc4,aes(x=wc,y=Freq,color=wc,fill=wc))
g4+geom_bar(stat = "identity")+coord_flip() + xlab("Country")
```



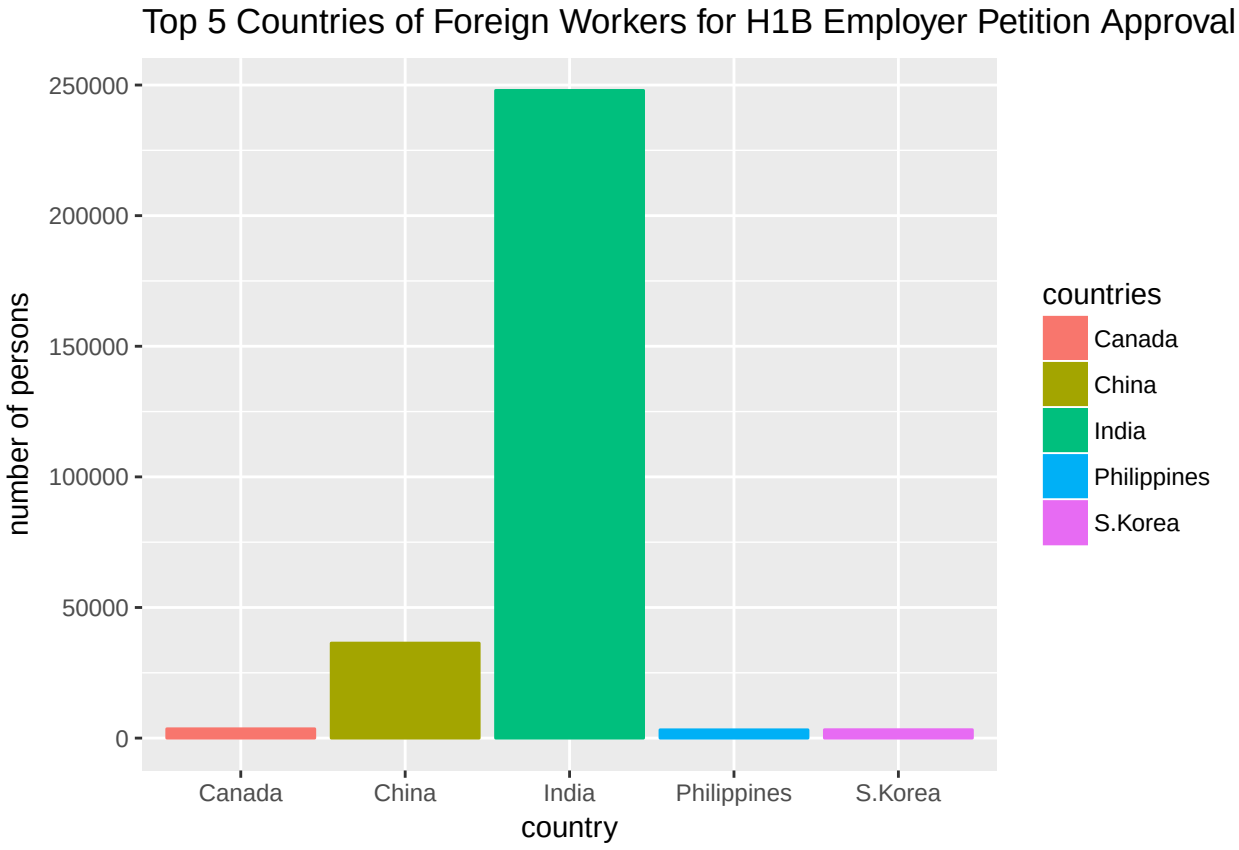
H1B Workers

H1B is a type of visa in the United States. This visa can be used by companies to employ foreign workers in certain occupations. In order to be considered a valid occupation for the H1B, it must require highly specialized knowledge in the field. The H1B is notable because it applies to a situation when a foreign student graduates from a university and then wants to continue being able to work in the US for an extended period. Additionally, there are only a limited number of H1B visas that are approved each year. The following sections look at where H1B applicants are coming from and what type of work they are doing.

Origin of H1B Workers

The plot below shows the number of petitions for an H1B visa by employers in 2017 for workers from certain countries. It details the top five. The most notable observation is that there are far more petitions for Indian workers than the other groups. This graphic is very similar to the one above related to PERM data and countries of origin. It can be useful to compare the two. While quite similar, Canada and the Philippines replaced Vietnam and Mexico in the top five. This might suggest something about the age groups of the foreign workers.

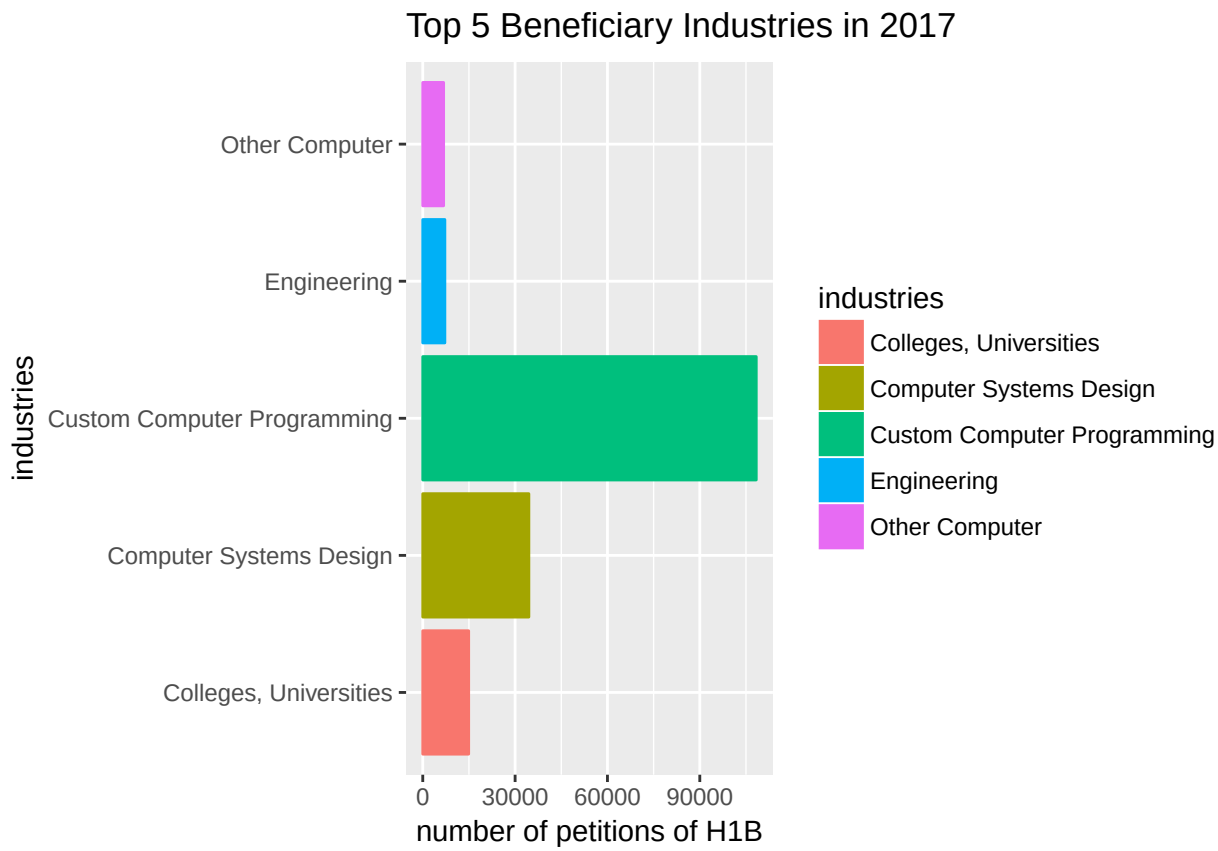
```
countries <- c("India","China","Philippines","S.Korea","Canada")
y <- c(247927,36362,3161,3203,3551)
df <- data.frame(countries,y)
# We were unable to get the excel file to read correctly so we had to resort to making our own data fra
ggplot(df,aes(countries,y,color=countries,fill=countries)) + geom_bar(stat = "identity") + xlab("countr
```



H1B Prevalence in Various Industries

What types of jobs foreign workers are doing is certainly worth analyzing. A component of the H1B data details the type of industry that the employer belongs to. The following plot represents the top five industries that employers petitioning for H1B visas belong to. We can see that STEM areas, specifically computer-related fields, represent a large percentage of the H1B applications. This is an important insight because it reveals what industries many foreign workers are employed in. It also may suggest characteristics about the US labor force or identify industries that are facing a shortage of workers.

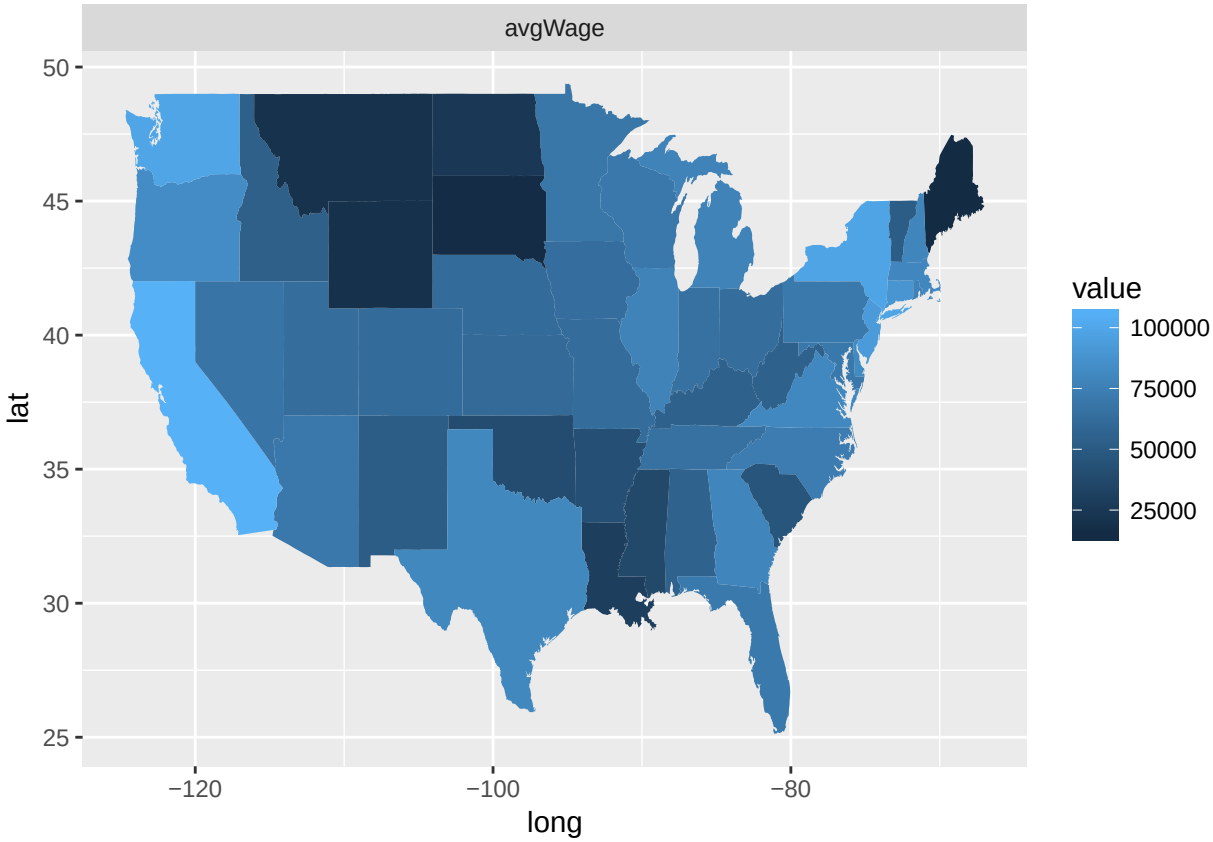
```
industries <- c("Custom Computer Programming","Computer Systems Design","Colleges, Universities","Engin
y2 <- c(108332,34462,14913,7169,6763)
df2 <- data.frame(industries,y2)
# We were unable to get the excel file to read correctly so we had to resort to making our own data fra
ggplot(df2,aes(industries,y2,color=industries,fill=industries)) + geom_bar(stat = "identity") + coord_f
```



Prevailing Wage Data

A final area of interest is wage information for foreign workers. The OFLC also provides prevailing wage data which can be analyzed. A graphic comparing wage data across the US as well as a discussion of what was done to the data to create the data product can be found below.

```
PrevailingWage_15<- read.csv("PrevailingWage'15.csv", header=TRUE)
PV15<-tolower(colnames(PrevailingWage_15))
colnames(PrevailingWage_15)<-PV15
subPrevailingWage_15<-PrevailingWage_15%>%
select(prevail_wage,wage_level,visa_class,worksite_state,business_name)%>%
group_by(worksite_state)%>%
summarise(avgWage=mean(prevail_wage,na.rm = TRUE))
mdat<-map_data("state")
stateNames<-tolower(subPrevailingWage_15$worksite_state)
subPrevailingWage_15$worksite_state<-stateNames
mapSWages<-merge(mdat,subPrevailingWage_15,by.x = "region",by.y = "worksite_state",all = TRUE)
index<-order(mapSWages$order)
mapSWages<-mapSWages[index,]
mSWages<-subset.data.frame(mapSWages,select=c(region,long,lat,group,avgWage))
mapdat<-melt(mSWages,id.vars = c("long","lat","group","region"))
avgWageMap<-ggplot(mapdat,aes(long,lat,group=group))+geom_polygon(aes(fill=value))+facet_wrap(~variable
avgWageMap
```



Above, we took the data set collected from the US Department of Foreign Labor, on prevailing wages for the fiscal year 2015. Initially, many columns were filled with NA's and therefore useless. So, through piping, we subsetted the data by the columns we wanted and, with the goal of mapping avg wages by state, we grouped the set by the worksite state names. For this set, we wanted to particularly compare the aggregate wages amongst states. The prevailing wage may not be what the laborer is paid but for their industry and region it is the average so the laborer will be paid based on that figure. Since all the columns and text were capitalized, we converted everything 'tolower' to compare to the map data later. Then before merging with the map we summarized the prevailing wages by state name. We then collected the map data and merged it with our new PV subset by state names. Lastly before plotting, we melted our map/PV data so we just have our state variables and avg wage values. The plotting is done using the long, lat map data and each state variable is filled to a certain degree by the avgWage value. From interpolating out map, we gathered that for 2015 the prevailing wages were the highest in California, Minnesota, Virgina, Indiana, Oregon and the northern east coast; while the lowest is an exception to the east coast being Maine, some of the other lowest averages are in Arkansas, Mississippi, Lousianna, Montana, and the Dakotas. This data gathered is pretty consistent with the average wage for all Americans by state, which is shown below. The maps both show the highest/lowest values in the east coast with the lowest in Maine and also the highest in California, Virginia and Oregon. More consistent lows are in Oklahoma, Mississippi, and Montana.

Conclusion

It is a little too early for us to have a definite conclusion on this data. It is still a work in progress. This is the location and section for the conclusion in our final draft.

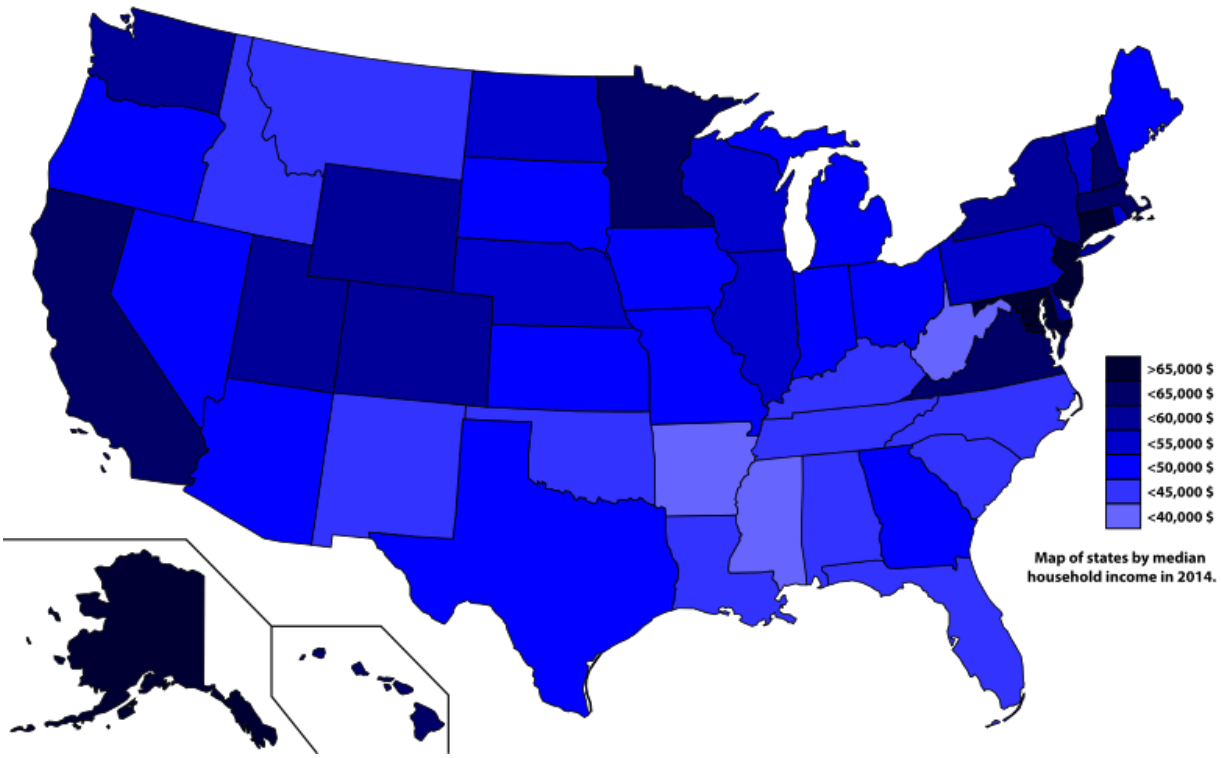


Figure 1: Figure 1: Comparison of State Average Wages

References

- http://www.flcdatacenter.com/CaseH1B.aspx
- https://www.foreignlaborcert.doleta.gov/performancedata.cfm
- https://www.dhs.gov/immigration-statistics/visualization/2015
- https://www.uscis.gov/tools/reports-studies/immigration-forms-data