**H1B Approval Analysis**

R Final Project 2019 Fall

# Introduction

Our R shiny app is to analyze the certification of H1B visa in the US. According to the survey of our classmates, more than xx percent of them heard about H1B visa and need it in the future. In this app, we will first see H1B certified jobs distribution among the US.

What’s more, many of our classmates pay attention to find jobs about data analyst and data scientist. In the Shiny app, there is data visualization of certified H1b cases on map by utilizing GIS; a bar chart of top H1b applied jobs for each year from 2011-2016; a table of top 10 companies by certified data analyst number; a line chart showing average prevailing wage for data analysts and data scientists related jobs.

# Data preparation

We choose H1B Visa Petitions 2011-2016 from Kaggle (<https://www.kaggle.com/nsharan/h-1b-visa>). This dataset contains five years’ worth of H-1B petition data, with approximately 3 million records overall. The columns in the dataset include case status, employer name, worksite coordinates, job title, prevailing wage, occupation code, and year filed.

# App development

1. **GIS**

To show the geographic distribution of certified H1B cases, we decide to use GIS. We use DT, leaflet and ggmap packages to complete the map. The map includes all certified cases as well as a drop-down menu for choosing different years. In order to do that, we used latitude and longitude for each job title to graph.

1. **Top H1b applied jobs**

As international students who plan to look for jobs in the US, we are interested in what are the jobs that have the largest number of applications for H1b and what is the proportion of certified H1b cases for each job.

In the panel, we set two parameters to get the results. One is parameter N, which sets the number of top job categories and its value is between 5 and 20. The other parameter is Year, which sets the year of data that is of interest. To use these parameters, we first set a Year (e.g., 2016) to extract applications for H1b in a specific year, and then we select a number N (e.g., 5) to get the top N job types sorted by number of applications.

1. **Table of top 10 certified data**

Since the dataset includes over 3 million observations, we cleaned the dataset to create a new table by selecting specific salary range and year. In this panel, there is a slider for filtering max and min salary, and a drop-down menu for choosing different year. In the table we created, the first column shows top 10 companies which have the largest amount of certified data analysts’ case in corresponding salary range; the second and third columns display the frequency of all certified H1B cases and data analyst cases respectively of the corresponding company.

1. **Average Prevailing Wage**

We all know that applicants’ salary is one of the requirements for H1b visa approval. We think it will be interesting to see the trend of average prevailing salaries for data analysts and data scientists related jobs. Because our dataset is huge, different companies put different job titles for similar positions, therefore, our first step of cleaning the data is to extract similar jobs under different titles and group them together. The second step is to visualize our data on a line chart by years. The interesting fact is that although the average prevailing salary of data scientists is still much higher than that of data analysts, we can clearly see that the average prevailing salary of data analysts increases over the years, and becomes closer to the level of data scientists.