

Short Paper

Huanye Liu

```
library(readr)
library(haven)
library(dplyr)

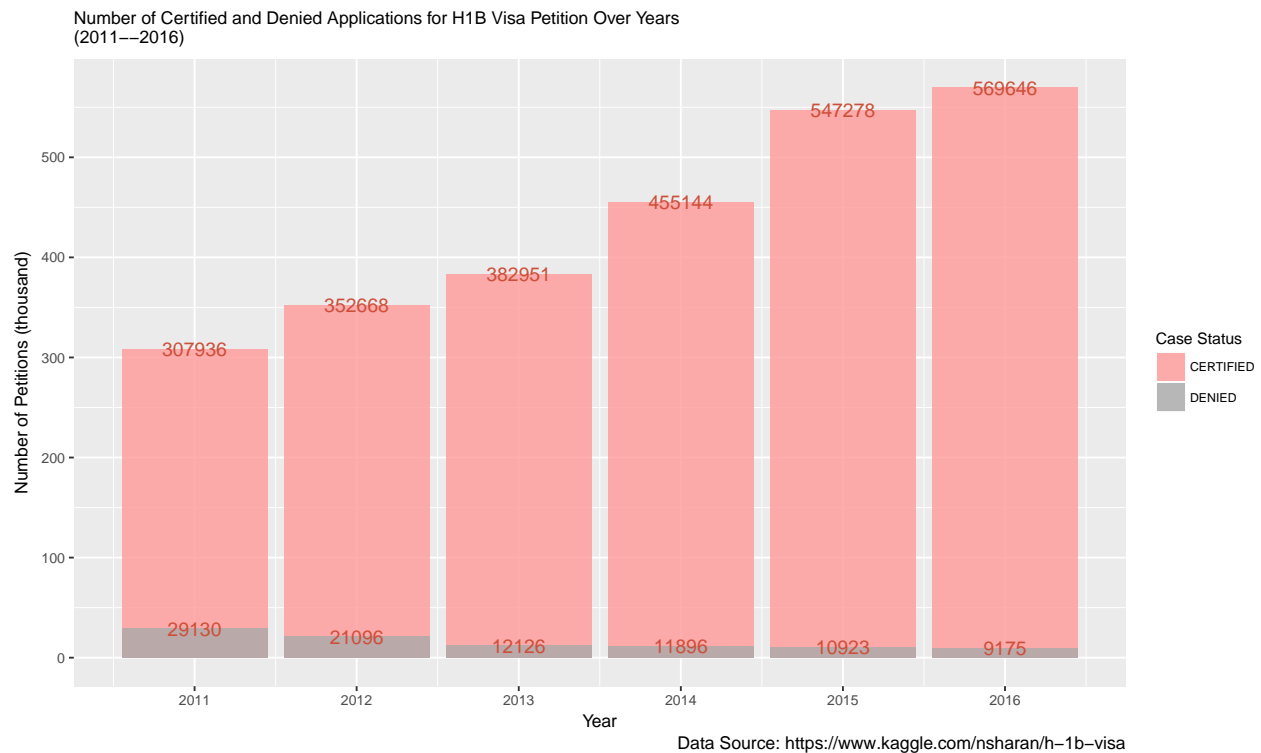
## Warning: package 'dplyr' was built under R version 3.4.2
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(tidyr)
library(stringr)
library(ggplot2)
library(maps)
library(viridis)

## Loading required package: viridisLite
# It may take 2-3 minutes to read the csv file.
h1b = read.csv("h1b_kaggle.csv",header = T,row.names=1)
```

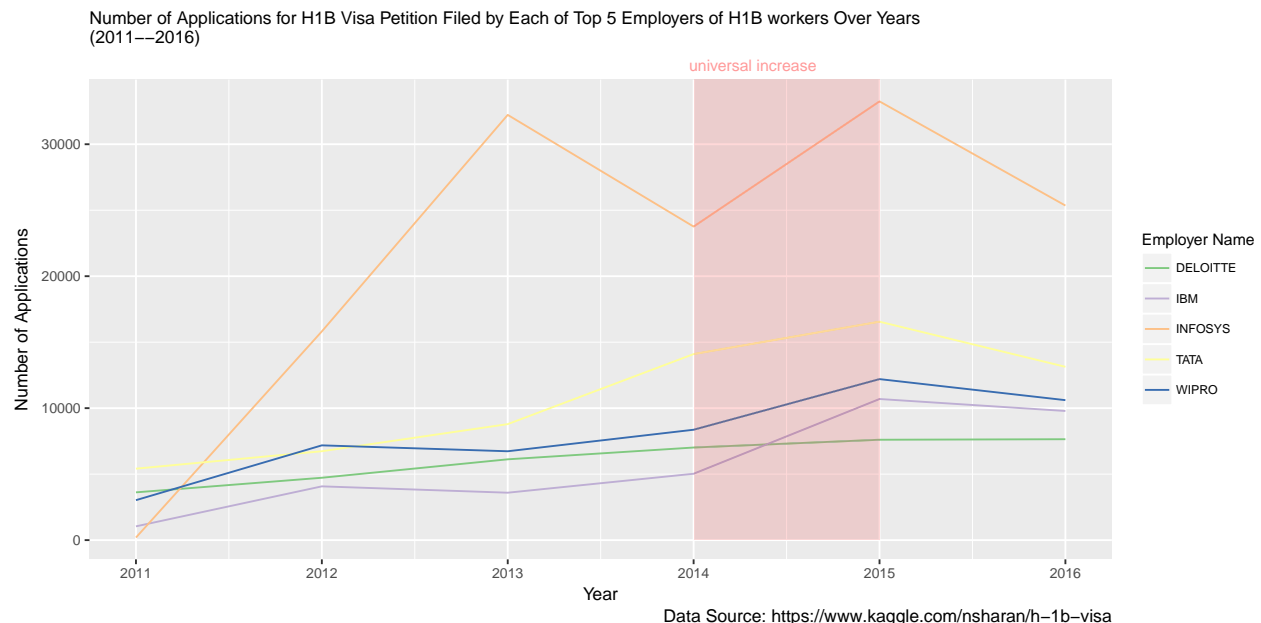
To petition the U.S. Citizenship and Immigration Services (CIS) for a H1B visa, employers must first obtain labor certification through the U.S. Department of Labor (DOL). The graph below shows trends of number of certified and denied applications for H1B visa petition from year 2011 to 2016. We can see a clear and steady rise in the number of certified applications for H1B petitions accompanied by a moderate decrease in the number of denied ones, which signifies an increasing supply of foreign works who are willing to temporally work in the United States. A closer examination may lead us to the largest increase in certified H1B petitions from year 2014 to 2015, a jump from 455144 to 547278, which could be related to the highest GDP growth rate from 2014 to 2015 in this period.

```
h1b%>%
  group_by(YEAR,CASE_STATUS)%>%
  summarize(number=n())%>%
  filter(CASE_STATUS %in% c("CERTIFIED","DENIED"))%>%
  ggplot(aes(x=YEAR,y=number/1000,fill=CASE_STATUS))+
  geom_bar(position="identity",stat="identity",na.rm = T,alpha=0.8)+
  geom_text(aes(label=number),color="tomato3")+
  scale_fill_manual(values = c("#FF9999","darkgray"))+
  scale_x_continuous(breaks=c(2011:2016))+
  scale_y_continuous(breaks=100*c(0:6))+
  theme(axis.text.y = element_text(size=8),
        axis.text.x = element_text(size=8),
        axis.title = element_text(size=10),
        plot.title= element_text(size=10),
        legend.text = element_text(size=7),
        legend.title = element_text(size=9))+
  labs(title = "Number of Certified and Denied Applications for H1B Visa Petition Over Years\n(2011--2016)",
       caption = "Data Source: https://www.kaggle.com/nsharan/h-1b-visa",
       x = "Year", y = "Number of Petitions (thousand)",
       fill="Case Status")
```



Next, we take a look at who are the largest employers of H1B workers. We can see from the graph that the top 5 are Infosys, Tata, Wipro, IBM and Deloitte, all providing IT services and technology consulting as their main business, and the biggest three among them, Infosys, Tata and Wipro, all have Indian origin and headquartered in India with a great deal of outsourcing services. So we may infer that Indian workers must have been a large ethnic group among all H1B applicants these years. We can also observe a universal increase of all 5 companies in number of applications from year 2014 to 2015, which matches the sharp rise shown on the first graph well.

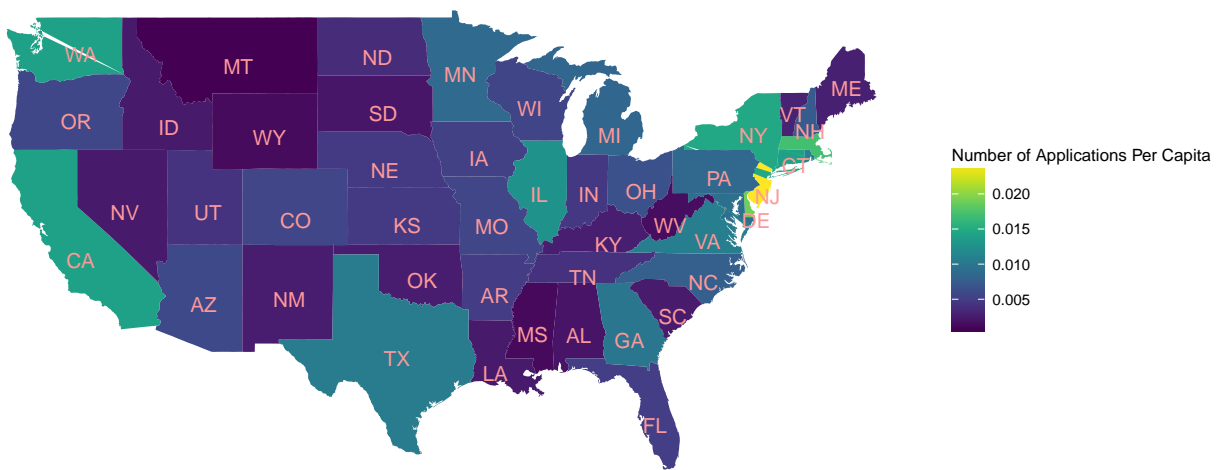
```
h1b %>%
  group_by(EMPLOYER_NAME) %>%
  summarise(NUMBER = n()) %>%
  arrange(desc(NUMBER)) %>%
  head(5)%>%
  right_join(h1b,by="EMPLOYER_NAME")%>%
  filter(!is.na(NUMBER))%>%
  group_by(EMPLOYER_NAME, YEAR)%>%
  summarise(N = n())%>%
  mutate(EMPLOYER = str_extract(EMPLOYER_NAME, '^[^ ]+'))%>%
  ggplot(aes(x=YEAR,y=N))+
  geom_line(aes(color=EMPLOYER))+
  geom_rect(xmin=2014,xmax=2015,ymin=0,ymax=35000,fill='#FF9999',alpha=0.01)+
  theme(axis.text.y = element_text(size=8),
        axis.text.x = element_text(size=8),
        axis.title = element_text(size=10),
        plot.title= element_text(size=10),
        legend.text = element_text(size=7),
        legend.title = element_text(size=9),
        plot.subtitle = element_text(size=9,hjust=0.67,color='#FF9999'))+
  labs(title = "Number of Applications for H1B Visa Petition Filed by Each of Top 5 Employers of H1B workers Over Years",
        subtitle = "universal increase",
        caption = "Data Source: https://www.kaggle.com/nsharan/h-1b-visa",
        x = "Year", y = "Number of Applications")+
  guides(color=guide_legend(title="Employer Name"))+
  scale_color_brewer(palette = "Accent")
```



We are also curious about the worksite distribution of all applications for H1B visa petitions among different states in the US. However, the total number of applications whose worksites are located within a state is a so-called “spatial extensive” measurement, which needs to be converted to its number per capita counterpart using the state population. The resulting map below shows that New Jersey, Massachusetts, New York, California and Washington are the states attracting the largest number of foreign workers.

```
h1b%>%
  mutate(state = str_to_lower(str_extract(WORKSITE, '\\b[~,]+$')))%>%
  group_by(state)%>%
  summarise(number = n())%>%
  right_join(map_data("state")[1:5], by=c("state"="region"))%>%
  mutate(state = str_to_title(state))%>%
  left_join(bind_cols(fname=state.name, abbrev=state.abb, xcenter=state.center$x,
                     ycenter=state.center$y), by=c("state"="fname"))%>%
  left_join(population, by=c("state"="State"))%>%
  rename(pop = `2017 Population`)%>%
  mutate(num_per_capita = number/pop)%>%
  ggplot()+
  geom_polygon(aes(long,lat,group = abbrev,fill=num_per_capita))+
  geom_text(aes(x=xcenter,y=ycenter,label=abbrev),color = '#FF9999',na.rm=TRUE,
            check_overlap = T,size=4,vjust=1)+
  theme_classic()+
  theme(axis.line=element_blank(),
        axis.text=element_blank(),
        axis.title=element_blank(),
        axis.ticks = element_blank(),
        plot.title= element_text(size=10),
        legend.text = element_text(size=8),
        legend.title = element_text(size=9))+
  labs(title = "Geographic Distribution of Number of Applications for H1B Visa Petition Per Capita by State",
       fill = "Number of Applications Per Capita",
       caption = "Data Source: https://www.kaggle.com/nsharan/h-1b-visa") +
  scale_color_brewer(palette = "Accent")+
  scale_fill_viridis(option = "viridis")+
  coord_quickmap()
```

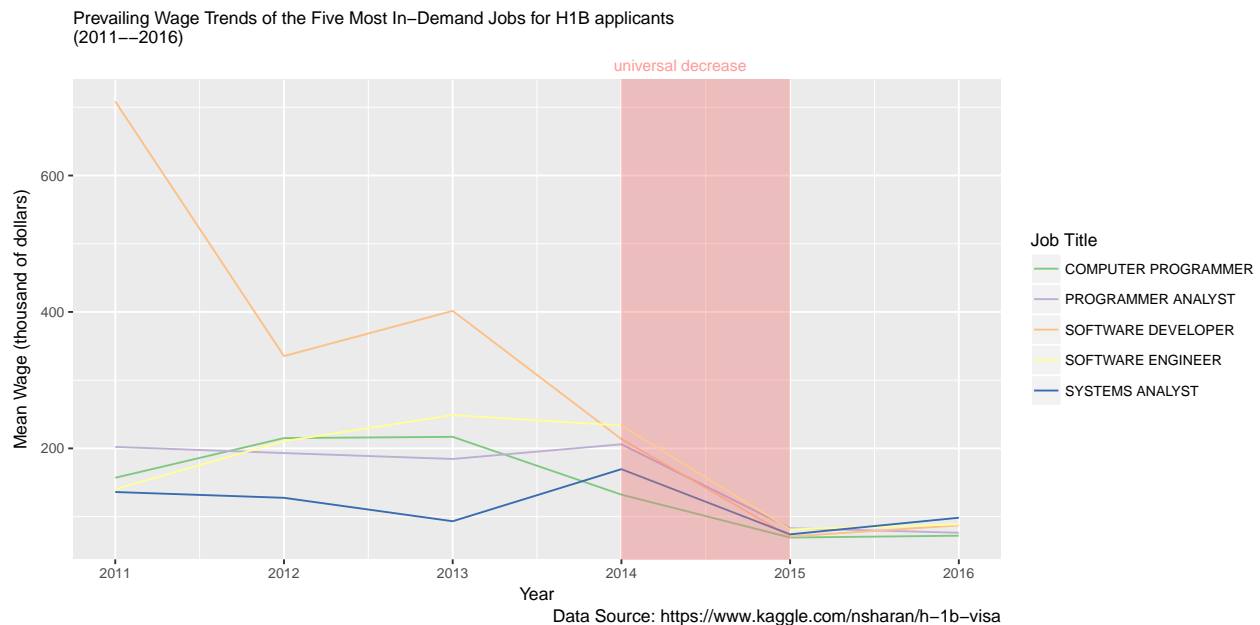
Geographic Distribution of Number of Applications for H1B Visa Petition Per Capita by State (summation over period 2011—2016)



Data Source: <https://www.kaggle.com/nsharan/h-1b-visa>

Now we turn to wage levels of H1B jobs over years. As the graph below shows, the overall wage level went down from year 2011 to 2016 as represented by those of the 5 most in-demand jobs, and the sharpest decline happened to software developers, although the extremely high wage level for them in 2011 could be an outlier. Besides, a universal decrease can be observed from year 2014 to 2015 with wage levels of all 5 types of jobs dropping below \$100,000, which may find its counterpart of the increase in number of applications on the first and the second graph. Dose it mean that the higher number of labor supply has pushed down the wage levels? More evidence is needed.

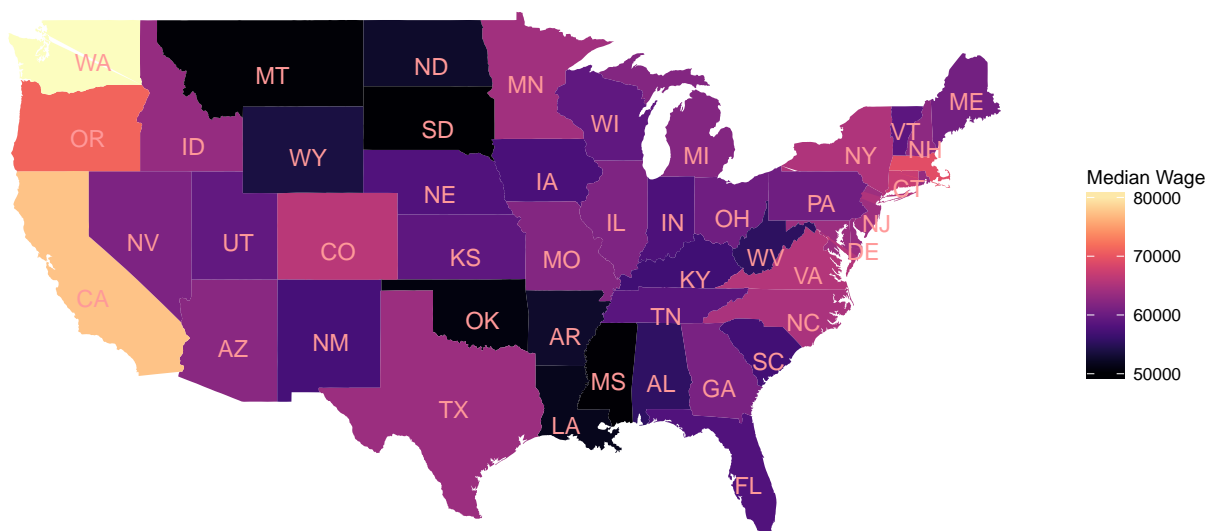
```
h1b%>%
  group_by(JOB_TITLE) %>%
  summarise(NUMBER = n()) %>%
  arrange(desc(NUMBER)) %>%
  head(5)%>%
  right_join(h1b,by="JOB_TITLE")%>%
  filter(!is.na(NUMBER))%>%
  group_by(JOB_TITLE, YEAR)%>%
  summarise(mean_wage_year=mean(PREVAILING_WAGE,na.rm=TRUE))%>%
  ggplot(aes(YEAR,mean_wage_year/1000))+
  geom_line(aes(color = JOB_TITLE),na.rm=TRUE)+
  geom_rect(xmin=2014,xmax=2015,ymin=0,ymax=750000,fill='#FF9999',alpha=0.02)+
  scale_y_continuous(labels = scales::comma)+
  labs(title = "Prevailing Wage Trends of the Five Most In-Demand Jobs for H1B applicants\n(2011--2016)",
       subtitle = "universal decrease",
       caption = "Data Source: https://www.kaggle.com/nsharan/h-1b-visa",
       x = "Year", y = "Mean Wage (thousand of dollars)")+
  theme(axis.text.y = element_text(size=8),
        axis.text.x = element_text(size=8),
        axis.title = element_text(size=10),
        plot.title= element_text(size=10),
        plot.subtitle = element_text(size=9,hjust=0.68,color='#FF9999'),
        legend.text = element_text(size=8),
        legend.title = element_text(size=10))+
  guides(color=guide_legend(title="Job Title"))+
  scale_color_brewer(palette = "Accent")
```



```
h1b%>%
  mutate(state = str_to_lower(str_extract(WORKSITE, '\\b[~,]+$')))%>%
  group_by(state)%>%
  summarise(median_wage = median(PREVAILING_WAGE,na.rm=T))%>%
  right_join(map_data("state")[1:5],by=c("state"="region"))%>%
  mutate(state = str_to_title(state))%>%
  left_join(bind_cols(fname=state.name,abbrname=state.abb,xcenter=state.center$x,
                      ycenter=state.center$y),by=c("state"="fname"))%>%

  ggplot()+
  geom_polygon(aes(long,lat,group = abbrname,fill=median_wage))+
  geom_text(aes(x=xcenter,y=ycenter,label=abbrname),color = '#FF9999',na.rm=TRUE,
            check_overlap = T,size=4,vjust=1)+
  theme_classic()+
  theme(axis.line=element_blank(),
        axis.text=element_blank(),
        axis.title=element_blank(),
        axis.ticks = element_blank(),
        plot.title= element_text(size=10),
        legend.text = element_text(size=8),
        legend.title = element_text(size=9))+
  labs( title = "Geographic Distribution of Prevailing Wage Level for H1B applicants by State\n(median over period 2011—2016)",
        fill = "Median Wage",
        caption = "Data Source: https://www.kaggle.com/nsharan/h-1b-visa")
  scale_fill_viridis(option = "magma")+
  coord_quickmap()
```

Geographic Distribution of Prevailing Wage Level for H1B applicants by State
(median over period 2011—2016)



Data Source: <https://www.kaggle.com/nsharan/h-1b-visa>

Again, we also want to know the wage level distribution at the state level based on the job worksite of each H1B applicant. As the map above shows, the three states on the west coast shows the highest median wage levels, where high-tech companies are concentrated. Another reason may be the higher living costs there which urge employers there to offer higher wages to attract foreign workers. Moreover, the graph below shows a further look at the wage ranges offered by the biggest employers of H1B applicants compared with those offered by 5 famous high-tech companies all of which are headquartered on the west coast. We can see that

the median wage levels of the 5 high-techs all exceed those of the 5 biggest hirers of H1B workers.

```
h1b%>%
mutate(EMPLOYER = str_extract(EMPLOYER_NAME, '^[^ ]+'))%>%
filter(EMPLOYER %in% c("GOOGLE", "FACEBOOK", "AMAZON", "LINKEDIN", "MICROSOFT", "IBM", "WIPRO",
                        "TATA", "INFOSYS", "DELOITTE") & !is.na(PREVAILING_WAGE))%>%
ggplot(aes(x=factor(EMPLOYER), y=PREVAILING_WAGE/1000, fill=factor(EMPLOYER), alpha=0.4), na.rm=T)+
geom_boxplot(aes(x=reorder(EMPLOYER, PREVAILING_WAGE, median)), na.rm = T)+

theme(axis.text.y = element_text(size=8),
      axis.text.x = element_text(size=8),
      axis.title = element_text(size=10),
      plot.title= element_text(size=10),
      legend.text = element_text(size=7),
      legend.title = element_text(size=9))+
labs(title = "Prevailing Wage Ranges by Some Employers of H1B workers\n(accumulated over 2011--2016)"
     , caption = "Data Source: https://www.kaggle.com/nsharan/h-1b-visa",
     , x = "Employer", y = "Prevailing Wage (thousand of dollars)",
     , fill="Employer")+
scale_y_continuous(labels=scales::comma, limits = c(0,300), breaks = 50*c(0:6))+
scale_color_brewer(palette = "Accent")+
coord_flip()
```

