

H1b Petition (2011 – 2016)

Akash Sethiya

Computer Science Graduate – Spring 2017

Mississippi State University

Final Project Report – Data Visualization with R

Introduction:

H1b visa is the business visa issued by US government to international employees gives permission to work in USA. I have found this dataset on <http://www.kaggle.com>. This dataset contains information of all H1b applications filed by US companies for their employees from year 2011 to year 2016. Data is quite good and already in tidy format but still it requires some tidying.

Here is some info about dataset:

- Total number of records in dataset: 3002458 (Over 3 million)
- Columns:

```
[1] "X"                "CASE_STATUS"      "EMPLOYER_NAME"  
[4] "SOC_NAME"         "JOB_TITLE"        "FULL_TIME_POSITION"  
[7] "PREVAILING_WAGE"  "YEAR"             "WORKSITE"  
[10] "lon"              "lat"
```

- Info about columns:
 - **X** – Serial Number
 - **CASE_STATUS**: Status of the application – There are total 6 status in dataset
 - **CERTIFIED** – It means that applicant is certified for H1b visa
 - **CERTIFIED-WITHDRAWN** – Applicant is certified but application is withdrawn
 - **DENIED** – Denied the application by US government
 - **PENDING** – It's pending and still in process
 - **INVALID** – Invalid application
 - **WITHDRAWN** – Application is withdrawn by employer
 - **EMPLOYER_NAME**: Name of the employer who is filing for h1b application.
 - **SOC_NAME**: Standard Occupational Qualification
 - **JOB_TITLE**: Job title of employee for which application is filed
 - **FULL_TIME_POSITION**: Yes/no
 - **PREVAILING_WAGE**: Annual salary of the applicant
 - **WORKSITE**: Location of the employer. In format, (City,State)
 - **Lon**: longitude of the location
 - **Lat**: latitude of the location

Tidying data:

Sampling:

Since actual dataset has 3M+ records, so simplify the visualization and amplify the processing of graph, I used data sampling. Data sampling is the method to sample data such as take out subset of data up to x%. I extracted out 10% of the data with random sampling.

Records after sampling: 300K

Removal of unused columns and NA values:

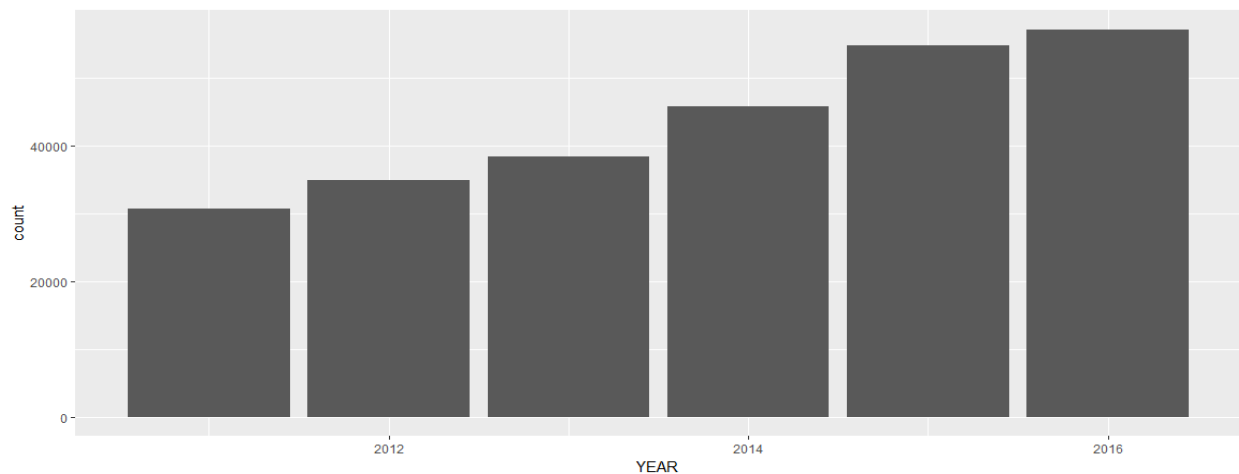
In the visualization of this dataset, we don't require X, lon, lat columns therefore it will be removed from the sampled dataset. Also, the column WORKSITE has information of city and state which we require visualizing the data in deeper manner, therefor, WORKSITE column is splitted into two columns such as STATE and CITY.

Columns after removing and splitting:

```
[1] "CASE_STATUS"      "EMPLOYER_NAME"    "SOC_NAME"         "JOB_TITLE"
[5] "FULL_TIME_POSITION" "PREVAILING_WAGE"  "YEAR"             "CITY"
[9] "STATE"
```

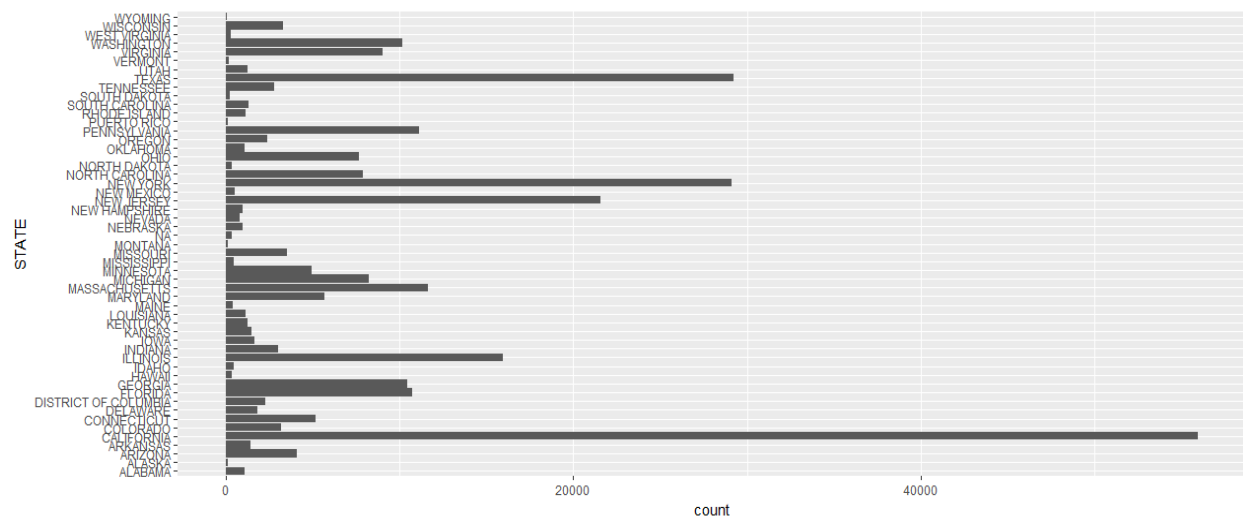
Analysis of the dataset:

First, we consider the total application approved over the year.



From the above graph, we can see that the number of approved applications increases over the year. It's probably because every year companies file more number of h1b applications.

To get to know about the dataset and h1b application, we need to have information that which state has highest number of application filed:



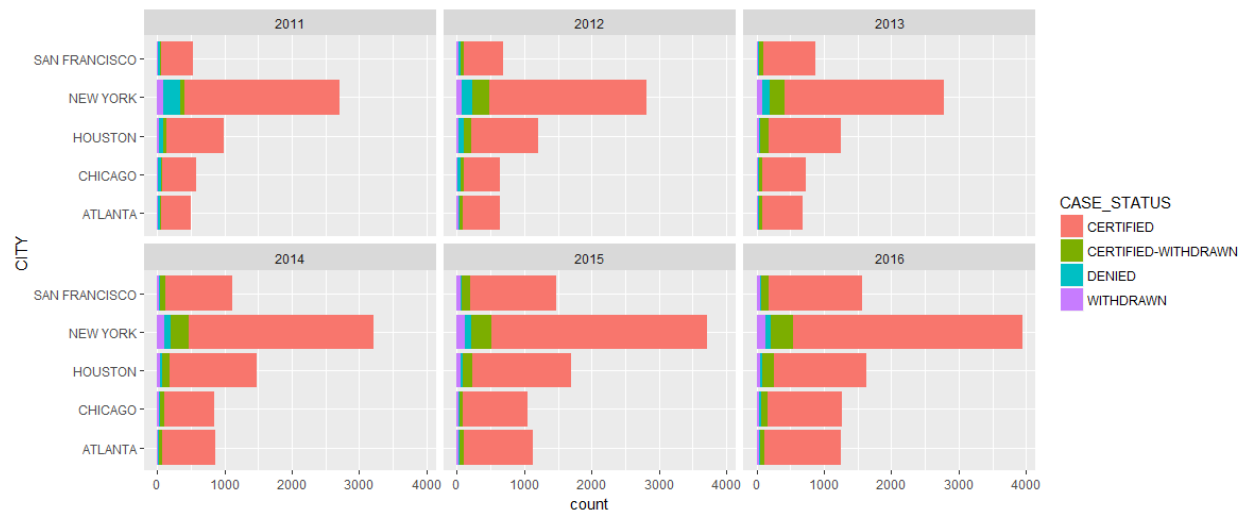
From the bar chart, it is clearly shown that California has highest number of h1b applications followed by Texas and New York. The reason behind California at the top is obvious, SILICON VALLEY and San Francisco, where most of the companies have their offices.

To get more understanding that which are those top 5 states which has highest number of h1b application filed:



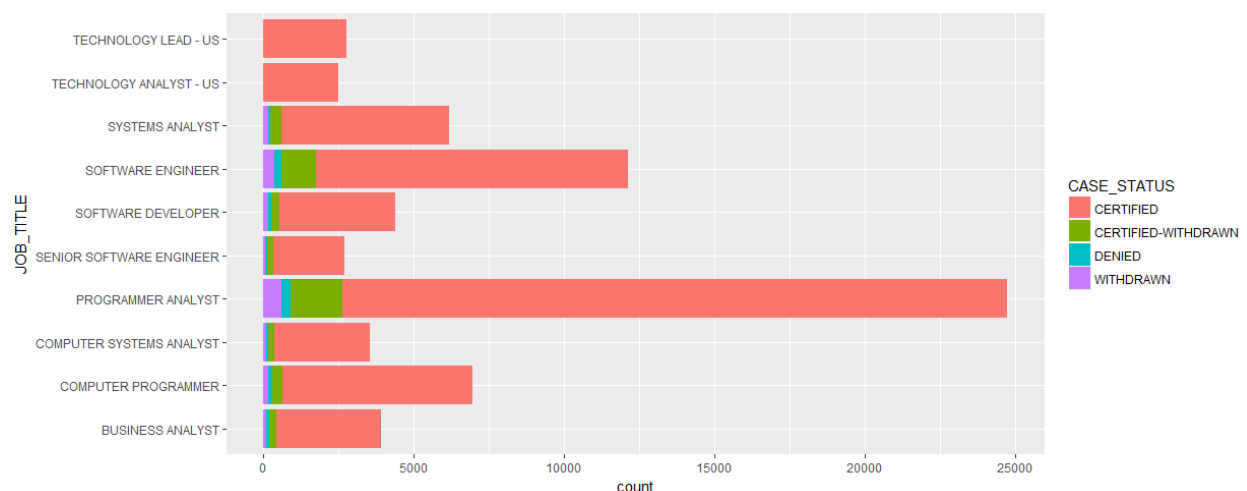
From the faceted graph by year, we can say that California has highest number of applications, but if we notice Texas and New York from year 2011 to 2016, application from Texas is increasing significantly than New York. The reason behind this, because of vast number of international students in Texas.

We have come to know about States so far, but to get to know it specifically that which city of the state has highest number of applications.

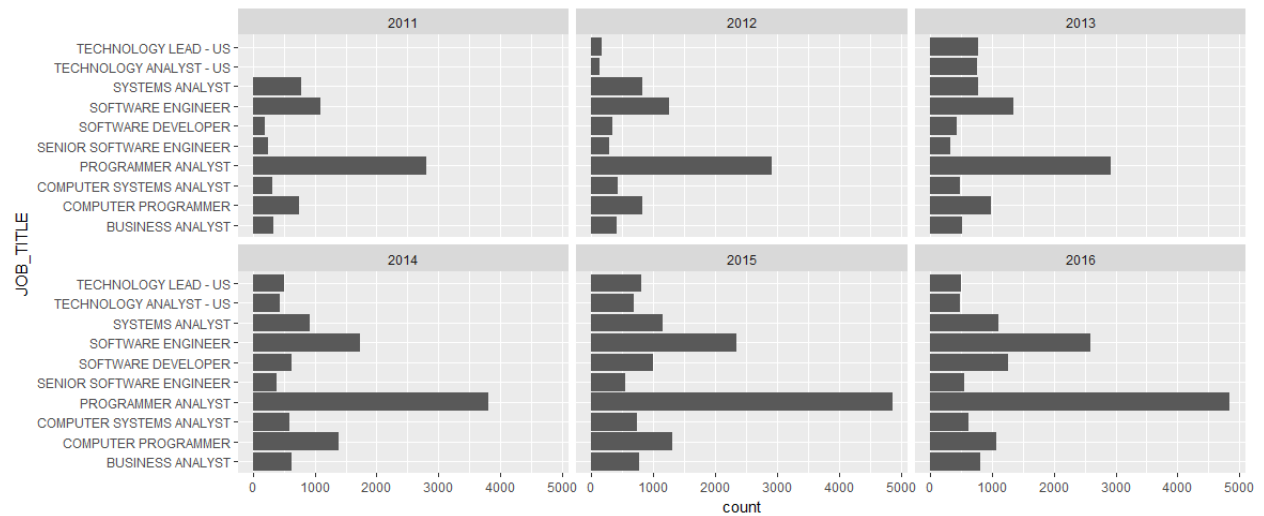


We have found that California is the top state but from the above faceted plot, it is clearly shown that New York City has highest number of applications. We might be thinking that this could be one of the city of California, but it's not. The reason might be that, In California, there are so many regions(city) where IT sector is there, not largely at one city. But in New York, New York city is heart of USA and most of the companies have their office in NYC only in New York State. That's why, New York City has highest number of applications.

So far, we have come to know that California is the top state and New York City is the top city for highest number of applications, the next important factor is the position, For which position most H1b application is filed.

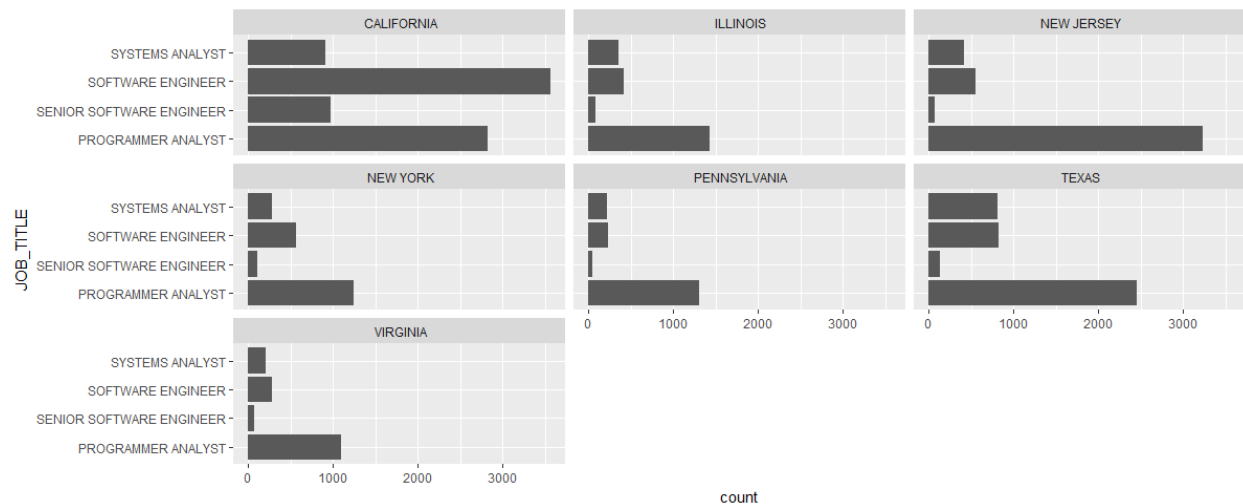


From the above analysis, Programmer Analyst is the most demanding position in USA for H1b immigrants followed by Software Engineer and Software Developer.



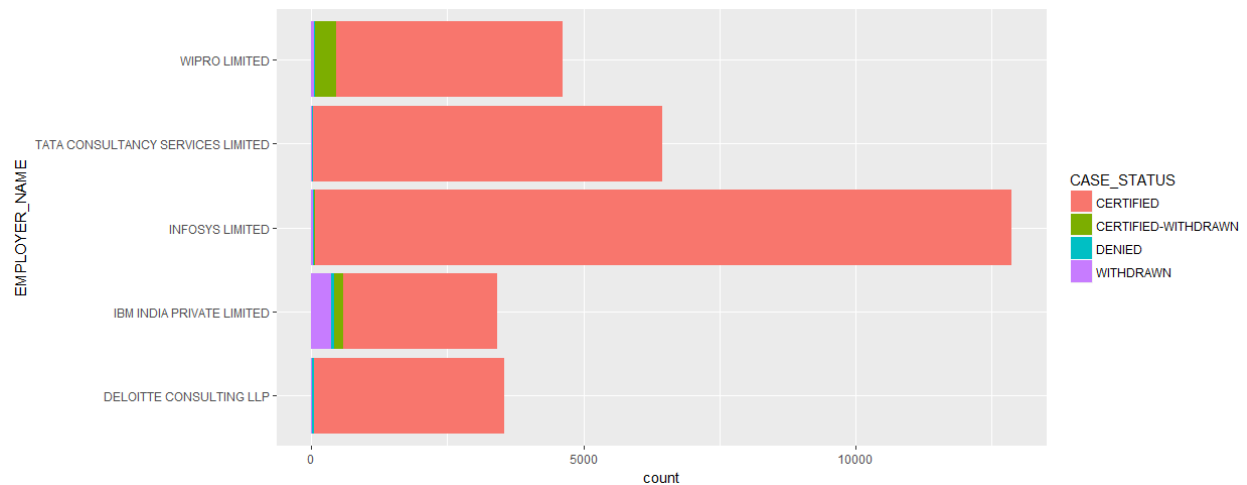
Above graph gives us the information about application by job title over the year. It is clearly shown in the faceted plot that programmer analyst is at top. The strange thing is, in 2011, job title “Technology Lead – US and Technology Analyst – US” is not demanding but over the period of 6 years, in 2016, it increases significantly, the reason may be because over this period, so many new technology comes in market and companies need leaders to lead their projects or migrating their projects from old to new technology.

The interesting thing to know about the job title vs state is that which state demands which role?



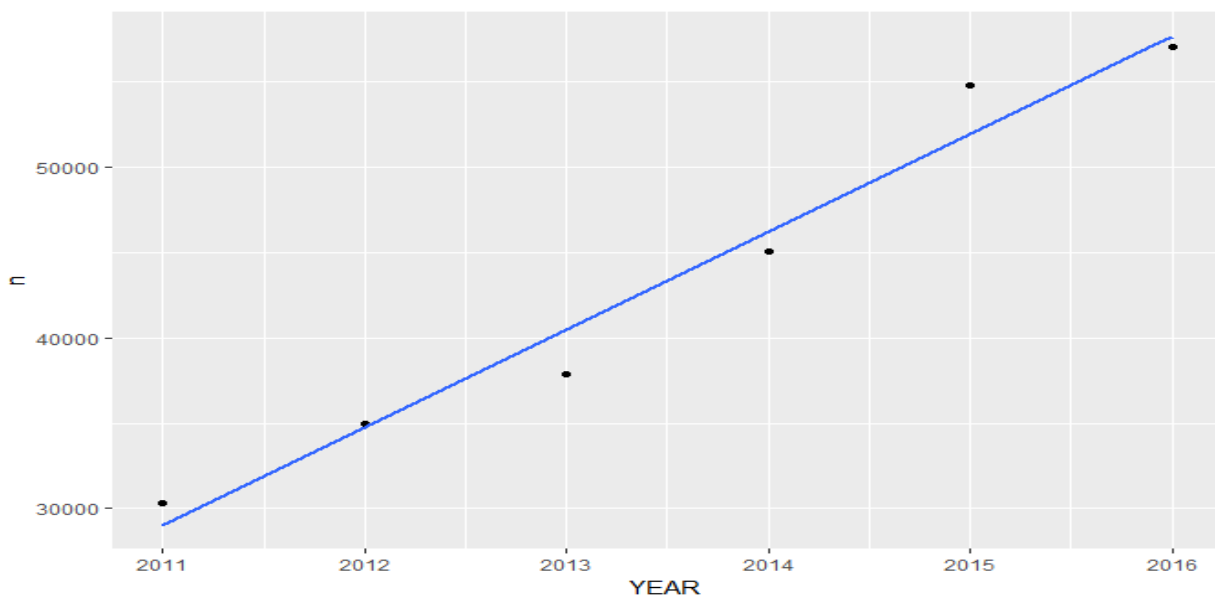
From the above analysis, it gives us the same result that Programmer Analyst is highly demanding but if we look at California graph, we will find out that in California, there is a need of Software Engineer more than Programmer Analyst. This is because of Silicon Valley and more technical companies in California region who requires employee who can implement their projects flawlessly.

The most crucial factor to determine in this dataset for International immigrants is the list of employers who filed most h1b applications.

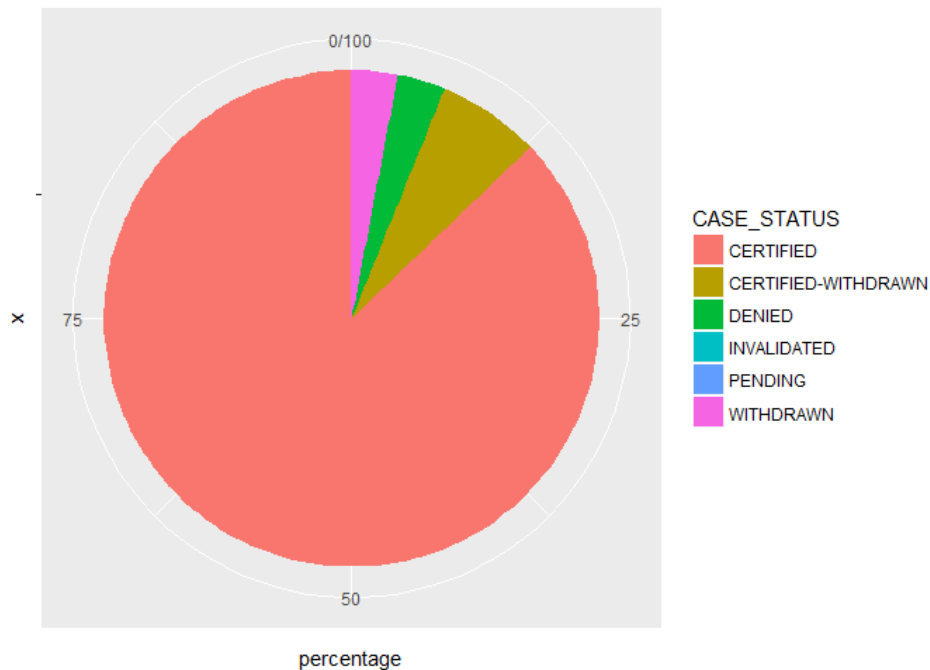


Above graph describe the relationship between Employers and no of applications filed by them. We all in the impression that big IT giants like Google, Facebook, Microsoft etc. might filed highest h1b applications, but it's not, Its Infosys Limited, which filed highest h1b applications followed by Tata Consultancy Service. The reason could be that both are Indian companies and they require their employees from India here in USA to provide best services to their clients.

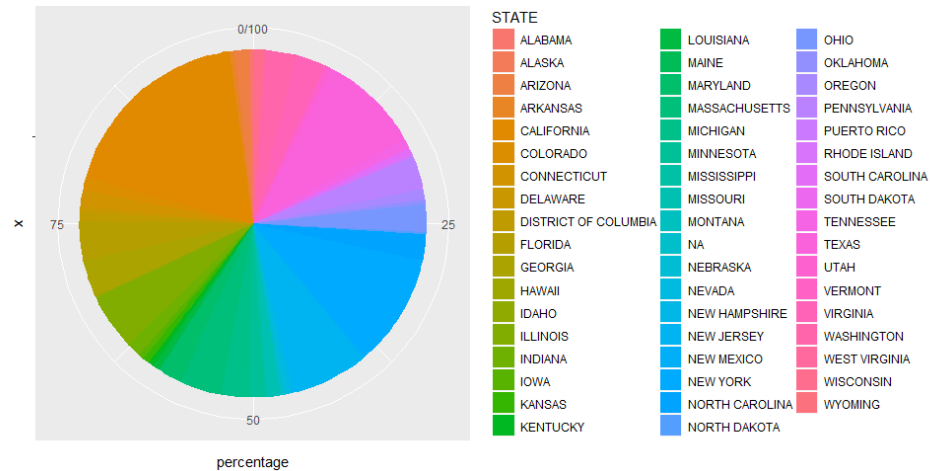
From year 2011-2016, we have got information about State, City, Job Title, Employers etc. Therefore, there is a linear model which describe the relationship between no of applications approved over the year and predicted the number of approved applications by year.



It gives the information that over the year, applications approval gets increased and it's increasing.



Following Pie chart gives us the information that about 87% of the applications gets approved and around 4% of the applications denied.



Also, there are rejections as well, so above pie chart gives us the information about rejections of the applications by State. And we can see that clearly, California has highest rate of application rejections. The reason is simple, it has highest number of applications as well.

So, from the above analysis, we get the information about h1b application that which state and city has highest approval rate and which employer and job title is safer for approval of applications.

Role of H1b Immigrants in USA Economy:

According to my research on internet, I have found out that h1b Immigrants plays a crucial role in growth of USA economy. It can be understood by following figures:

- GDP of USA in 2011 - \$15.52 trillion
- GDP of USA in 2016 - \$18.56 trillion
- Total H1b applications approved – 2615623
- % change in economy – 19.587% ▲

By looking at the figure, 19.587% increase in GDP in 5 years, it is clearly shown that H1b immigrants plays a crucial role in growth of USA GDP.

President Trump policy proposal for H1b visa applications:

Recently, in news we have watched that President Mr. Trump has proposed a policy for the application of h1b that companies can apply to h1b of only those employees who has yearly income of $\geq 130k$. And according to the data of last 5 years (2011-2016), out of 3M+ applications, only 87095 applicants has yearly wages more than 130K. If this proposal gets approved then only 2.903% will be eligible for H1B.

References:

- R for data science - <http://r4ds.had.co.nz/>
- <http://www.stackoverflow.com>
- <https://stats.stackexchange.com/>
- https://www.google.com/publicdata/explore?ds=d5bncppjof8f9_&met_y=ny_gdp_pcap_cd&idm=country:USA:RUS:GBR&hl=en&dl=en
- <http://www.livemint.com/Politics/711r4SV1bhUssgqz9CUcl/H1B-visa-reform-bill-introduced-in-US-minimum-pay-doubled-t.html>