An Analysis of PERM Labor Certification and Labor Condition Applications from the United States Department of Labor

Arunkumar Ranganathan
Brian Detweiler
Jacques Anthony

December 15, 2016

Abstract

Foreign born workers make up 16.7% of the United States workforce. In 2014, nearly one million foreign nationals became lawful permanent residents in the United States. Of those one million, 140,000 came through visas which are allocated to employment based residency. Where are these workers, and what do the demographics look like? How does their compensation measure up? Here, we use statistical analysis to examine visa application data from the U.S. Department of Labor from 2008 to 2016. We analyze similar reports done in previous years to determine what changes, if any have taken place in the last ten years.

1 Introduction

The U.S. Department of Labor provides data for Labor Condition Applications and PERM Labor Certifications dating back to 2008. This data contains a wealth of job market information including prevailing wage, and the wages offered by particular companies to individuals with particular qualifications.

Foreign workers can work legally in the U.S. under temporary or immigrant visas. To employ foreign workers legally in the U.S., employers must submit requests to the Office of Foreign Labor Certification. This process keeps borders secure, while protecting foreign workers from unfair treatment by employers. Businesses are required to pay the visa holder the higher of the prevailing wage for the position or the prevailing wage for the occupation in the geographic region of employment.

The H-1B visa program allows businesses to hire skilled professionals with at least a bachelor's degree in areas of specialization. H-1B is the largest of the Labor Condition Application (LCA) programs, which also consists of H-1B1 for Singaporean and Chilean workers, and the E-3 program for Australian workers. Each year the federal government grants a maximum of 65,000 applications for the H-1B visa program and additional 20,000 applications for applicants with at least a U.S. master's degree. Permanent applications are limited to 140,000 per year. Applicants whose petitions are denied for permanent residency must leave the U.S. for one year but are eligible to reapply later on.

The maximum length for the H-1B visa is six years. For permanent residency, an individual under the H-1B program may be sponsored by the employer. Workers with remarkable ability in science, education, arts, business, and education - professors, researchers, or business executives - are eligible to apply for permanent residency without a labor certification. Some special exceptions to these rules include business investors who invest over half a million dollars in the US, employees of US foreign service posts, retired employees from international organizations, and other classes of foreigners may also petition for permanent residency with no labor certification.

Here, we look for disparities in wages for domestic and foreign workers if any exist, expanding on the work of Mukhopadhyay and Oxborrow, which investigated data from 1999 - 2008. [**Greencard**]

1.1 Document reproducibility

The entirety of this project is reproducible using R (version 3.2 and above) with the `Knitr` package. All code, including this document, is available through GitHub. [**GitHub**]

2 About the data

The Office of Foreign Labor Certification, under the Department of Labor provides data for PERM Labor Certification (LC) applications and Labor Condition Applications (LCA) via XLSX files. Data is available from 2008 onward. The iCERT system was implemented in 2009, so there are two files for LCA 2009. Each file is structured similar to the others, but there are differences which must be addressed.

PERM (Program Electronic Review Management) data is available from 2008 to the present. Years 2015 and 2016 have over 120 attributes while 2008 through 2014 have 27-30 attributes. For our purposes, we will be using 27 attributes common across 2008-2016, as well as an additional 14 attributes found only in 2015-2016, which may be useful for limited analysis. Some of the fields include employer and employee details, job function, salary, university, job city, number of years of experience, country of citizenship, and industry.

2.1 H-1B Data preparation

The H-1B [1] data is about 75% larger than the PERM data, and spreadsheet programs do not handle these well, so the first task was to export these to CSV formatted files so that they could be handled with better tools. When exporting, they were also given more uniform file names in the form of "H-1B_yyyy.csv" where "yyyy" is the year of the data. The one exception here is the 2009 iCERT data, which was named "H-1B_2009_icert.csv". Once in CSV format, we needed to identify common columns across all spreadsheets. The difficulty here, is that the columns do not have the same names across spreadsheets, even though they may be holding the same data.

Using the UNIX tool `head -n 10 *.csv > headers.txt`, we took the first ten rows of each file and put them into a separate file. Each of the CSVs first ten rows were then copied and pasted into another spreadsheet, and we undertook a manual effort to match columns of the same identity. We also discarded some excess information that we deemed to be unnecessary for our purposes.

It is also important to note that there was not always a match for the columns we had selected. For instance, we found some interesting information regarding the attorney used by the employer to file the H-1B application. This was only introduced in the 2015 and 2016 data sets, however, so prior years would have no data for this.

After determining the standard columns, we wrote an import script in R that made use of the function `data.table::fread`. This allowed us to not only quickly read in the file, but also select only the columns of interest, and rename them to the standard naming convention upon read.

Once the data was read into individual data frames, additional cleaning rules were applied. In some years, wage data contained invalid numeric characters such as dollar signs or a range of wages in a single column. To get around this, dollar signs were removed before converting to numeric, and ranges were split into a *from* and a *to* column. Ultimately, all wages were transformed into ranges. If there was no range for the wage, then the wage itself was used as the range.

Another normalization task was the wage unit. Some wages were represented as yearly salary, some as hourly, and others as monthly, weekly, and bi-weekly. There can be subtle differences in each type of pay, but these were normalized according to a yearly salary. Hourly wage was multiplied by 40 hours a week and 52 weeks a year. Monthly wage was multiplied by 12, weekly by 52, and bi-weekly by 26. This allows all wages to be treated roughly on the same scale.

With the exception of exporting data to CSV format, the rest of the steps have been combined into a single script, `csv_manipulation.R`, to produce a final data set of 3,993,851 records. [**GitHub**]

---

[1] It should be noted that the LCA data is referred to generically throughout this paper as H-1B data, although it also includes the H-1B1 and E-3 programs.

2.2 PERM Data preparation

PERM data from the Department of Foreign Labor Certification is in XLSX format. We can read and manipulate these using the `xlsx` and `dplyr` packages.

Transforming all character variables to upper case deemed as necessary to avoid duplicity or incorrectness when performing data manipulation.

In total there were 9 files downloaded, with each file ranging from 40,000 to 90,000 records.

Most column names from 2008-2014 were the same with minor differences, while 2015-2016 had a different format, but were similar to each other. Column names are standardized to the 2016 version for posterity and the assumption that in the short term future, the columns will remain the same. To standardize all columns, we have replaced spaces in column names with underscores and created all column names in upper case. The first step in preparing data was to select all the relevant columns and fill empty columns with `NA`s.

Once we have all the data available for each year, a final combined version is created by standardizing the columns and structure.

This data set can be used by anyone interested in H-1B and PERM hiring trends to expand on our work.

Once we loaded and cleaned the data, we selected only the required columns and exported this as an RDS file. In this process of creating final file, we further reduced the columns and kept only the columns that are essential for our data analysis.

Multiple XLSX files are initially filtered with the selected 41 columns and merged back into one large file containing 502,978 records.

Data cleaning specific to this project is also required not just for standardizing data but also to have the data in a specific format.

During the process of creating final file, most columns are read as text and a few columns such as 'decision_date' are read as date, while salary information was read as numeric. Decision date is first transformed to YYYY-MM-DD format for consistency within our database.

As with the H-1B data, wage units were converted to a unified format and wages were normalized according to an annual rate and added to a new column just with annual salary. The range of salary was also normalized in this step. Employer zip codes were trimmed to 5 digits to keep all the zip codes at the same level of depth. Employment state data was in abbreviated form as well as expanded form, so we needed to standardize to the expanded level for all 50 U.S. states as well as U.S. territories. Finally, a unique row ID was added to each row.

This cleaned data was saved as an RDS file before moving on to exploratory data analysis.

2.3 Shortcomings

As mentioned in the previous section, because the data is not homogeneous, there are bound to be disparities. Missing data - columns which are not found across all spreadsheets - is the biggest issue. We can make assumptions when there is sparse data, but it would not be prudent to make assumptions where there is no data. For this reason, we fully disclose the absence data where necessary.

All of the data has been entered by humans at some point, so there are likely many human-generated errors. Some of these can be seen as outliers. Particularly in the LCA 2008 and pre-iCERT 2009 data, the wage unit is most certainly incorrect in some spots. For example, some wages are listed at $500 per hour, but the intended unit may have been per week. It is not possible to fix this programmatically though, because there are, in fact, some jobs that pay $500 per hour (CEOs, for instance). This data must be dealt with in one of two ways. They can either be corrected by hand inspection of outliers, or outliers can be removed completely. This results in a slight loss of fidelity. Extremely high paying jobs, such as CEO or physician may not be displayed.

Another issue is the switch from U.S. Citizenship and Immigration Services Dictionary of Occupational Titles (DOT) codes in 2008 and pre-iCERT 2009 data, to the North American Industry Classification System (NAICS) codes. The DOT codes are three digits and fairly high-level, where as the NAICS codes are hierarchical, with the first two being the industry, and the specification of the job title narrowing with up to six digits. For this reason, it is difficult to get consistent job titles across years.

PERM data is comprehensive for 2015 and 2016, however the data for each year have inconsistencies. We had to make some educated judgement about some categories and ignore many cases with empty values.

Years 2008-2014 have 25 attributes, but most of them do not give us substantial information other than salary and employer name. The nature of this data made it hard to graph plots based of these categorical variables. Here is a sample of unique values in the categorical data:

53,011	Job title based on Perm data
46,410	Job Title based on Work
40,431	Study Major
135,955	Employers
213	Countries in the "other" category

3 Methods

The numeric data provided by the H-1B and PERM data sets are mostly in the form of wages, both the wage that the employer is offering and the prevailing wage.[2] Also of interest, are the number of workers an applicant is filing for, and the implicit number of applications faceted by status and year.

3.1 Data product

First we will look the number of applications processed since 2008. Using these statistics, we can gauge an increasing or decreasing trend in the number of applications processed. Although there are caps on the annual number of employment based immigration, an increasing trend or maximum use of employment based immigration would provide insight.

Next, we can look for immigration based on geography. In what regions are employers hiring? We also look at year-over-year data on employment trends as well as education and other background information.

4 Results
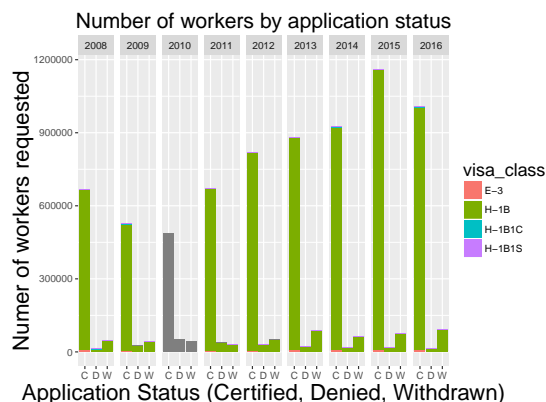
4.1 Labor Condition Agreement Applications

Figure 1: LCA status from 2008 - 2016

The number of workers requested via LCA has shown a marked increase since 2010 (Certified, Denied, and Withdrawn). It is also immediately clear, from Figure 1 that H-1B LCAs dwarf E-3 and H-1B1 applications, making up 98.94% of all LCA applications. [3]

---

[2]Prevailing wage is defined as the hourly wage, usual benefits and overtime, paid to the majority of workers, laborers, and mechanics within a particular geographic area.

[3]2010 data did not include the visa classification

The number of denied and withdrawn applications remains fairly constant, but the number of certified LCA workers tracks along with a strengthening economy and returning jobs. Compared side-by-side (Figure 2), it can be seen that LCA applications rise and fall with the job market. [**BLS**]

The astute reader may note that the number of certified LCA applications greatly exceeds the cap of 65,000 workers. It is important to make the distinction that an LCA does not necessarily lead to a visa certification. They are neccessary but not sufficient for obtaining a visa.
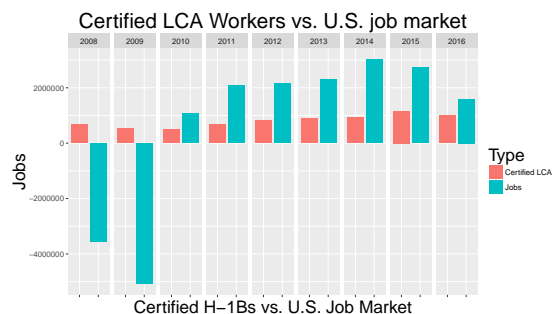
Figure 2: LCA applications vs. jobs lost or added
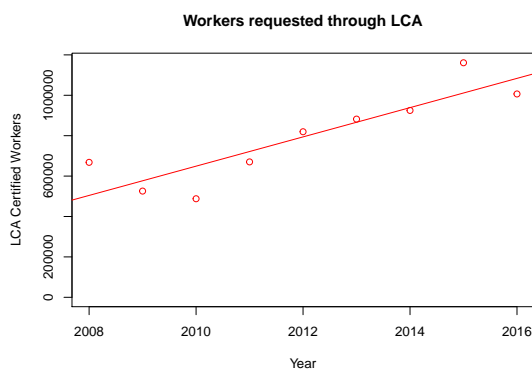
Figure 3: LCA application growth

```
## NULL
```

Figure 3 shows the demand for new temporary worker jobs year over year suggests a strong linear relation with $\rho = 0.7824673$ and a p-value of 0.0015341. We could estimate that if things continue on this course, in 2017 we should expect to see 1,156,143 LCA applications certified.

4.2 Prevailing Wage and Actual Wage

A study done by the Center for Immigration Studies, a non-profit conservative think tank made claims that for 2005 LCA data, employers were lowballing prevailing wages in order to bring in cheap under-skilled labor. The concept of skill level was defined as the Department of Labor's skill-based prevailing wage system, which doesn't appear in any of our data.

We attempted to recreate parts of this study for 2015 data to see if things had changed over a decade later.

The 2005 report, which focussed specifically on computer programmers, stated that 84% of H-1B prevailing wage claims fell below the median U.S. wage when compared with the same occupation in the same state. For the 2015 data, we found that 88.92% of prevailing wage claims fall below the median (see Figure 4). This
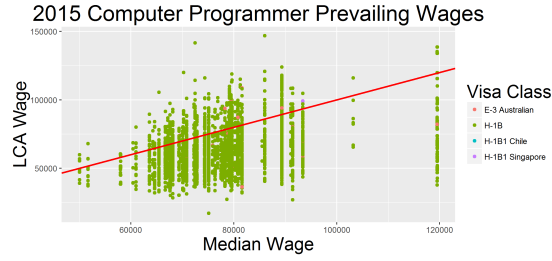
Figure 4: Prevailing wage claims for Computer Programmers compared with the US median wage (compared by state). The red line indicates where the prevailing wage claim equals the median wage.
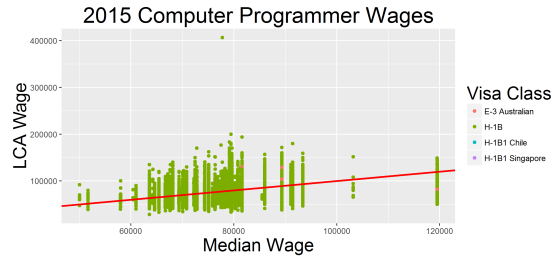

Figure 5: Actual visa wages for Computer Programmers compared with the US median wage (compared by state). The red line indicates where the visa wage equals the median wage.

seems to back up the claims of the CIS report, that employers are making dubious prevailing wage claims in order to hire cheaper programmers.
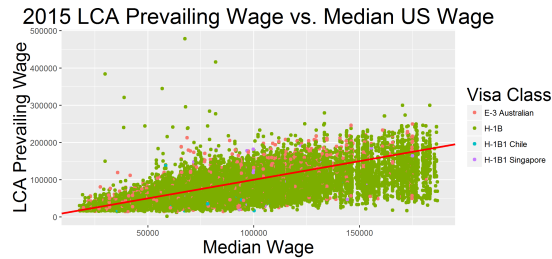

Figure 6: Claimed prevailing wage compared with the equivalent median wage by occupation and state. The red line indicates where the visa wage equals the median wage.

The CIS study claimed that in FY 2005, H-1B employer prevailing wage claims averaged $16,000 below the median wage for U.S. computer workers in the same location and occupation. Here, we look at computer programmers specifically in 2015, and we can see that the employer prvailing wage claims average about $17,510 below the median, which is $1,510 worse than 10 years ago, or roughly on par with the 21% inflation. [**BLSInflation**]

However, when looking at the H-1B program as a whole, the numbers improve significantly. We found that the prevailing wage claims in general were only 72.68% below the median on average. There is certainly room for improvement here, but this is a 17.32% difference in the claims made by the he CIS report. When comparing actual wages, 39.24% fell below the median, which is a 44.76% improvement over what was reported.

Wages in 2005 were $12,000 below the median on average, and in 2015, we found this had not really changed. In general, however, wages were only $5,861 below the median on average. The standard deviation
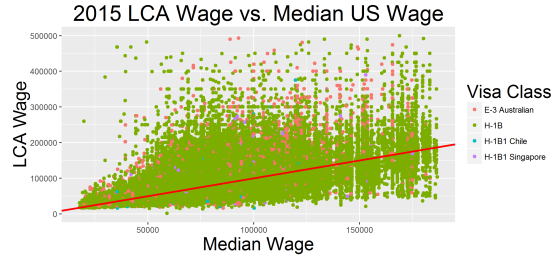
Figure 7: Actual visa wages compared with the US median wage by occupation and state. The red line indicates where the visa wage equals the median wage.

of prevailing wage claims was $23,159.

The case that the CIS report makes is that the H-1B program is being abused to bring in under-qualified entry-level workers to work for less than their US counterparts. While there does seem to be some merit to these claims when analysing computer programmers, as a whole, the H-1B program is far more complex, with a standard deviation of $31,895.

Table 1: PERM Median Pay and Number of Applications by Year

Year	Median Salary	Applications
2008	$63,000	61997
2011	$83,158	73207
2012	$89,150	66488
2013	$91,645	44152
2014	$93,995	70998
2015	$90,522	89299
2016	$95,000	96837

4.3 Permanent Resident Visas

Permanent residents (commonly known as green card holders), have shown a steady increase in median salary over the years, and the number of applications has increased as well.
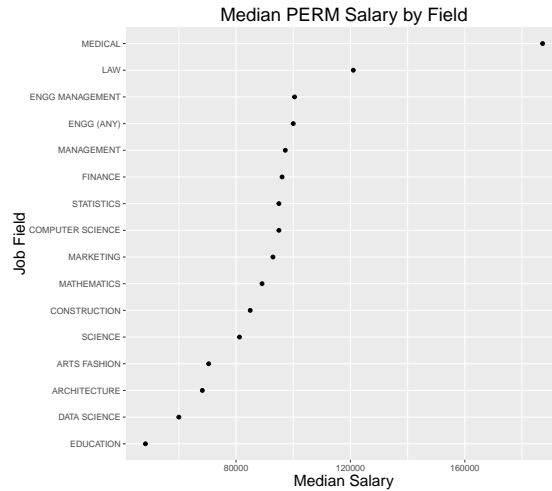

Figure 8: Median wage for PERM workers in Engineering Fields by State

For permanent workers, the medical field dominates in terms of salary, coming in with a median pay of $187,200. Lawyers come in at a distant second with a median salary of $121,000.

When looking geographically at engineering jobs, California, Washington, and New York came through as some of the highest paying states, with median salaries of over $100,000.

Finally, we examined which universities were most responsible for providing skilled labor by the number of applicants, as well as how well those applicants are paid. Carnegie Mellon University topped the list with alumni making a median salary of over $110,000. Jawaharlal Nehru Technological University and Anna University however, top the list of applications with over 3600 certified applicants each.

4.4 Attorneys

Something that may be of interest to employers is finding a good attorney to help navigate the legal process. The H-1B data lists the names and states of agent attorneys for fiscal years 2015 and 2016. By filtering and summarizing the data, we can see the top attorney for each state (and Washington, D.C.), based on how many Labor Condition Applications have been certified under their name.
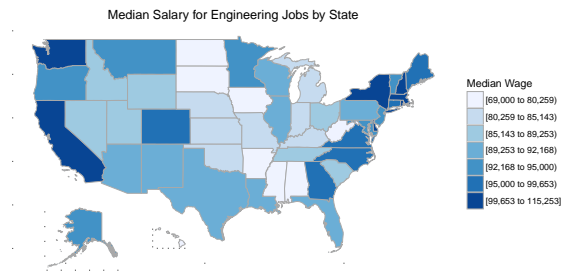
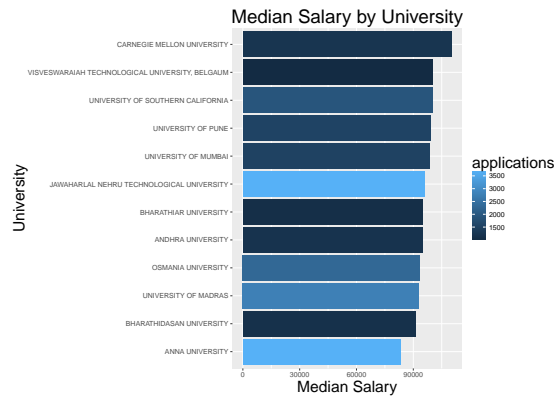Figure 9: Median wage for PERM workers in Engineering Fields by State

Figure 10: Median wages for PERM workers by University, showing universities with over 1,000 applications.
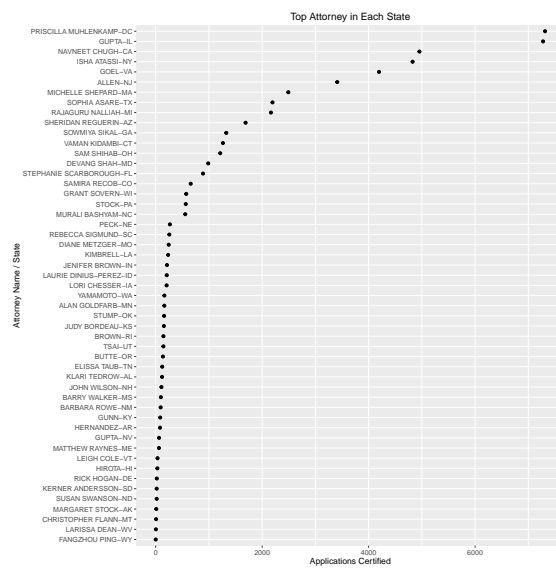

Figure 11: Top H-1B attorney in each state (and D.C.) by number of applications certified.

5   Conclusion

We analyzed two data sets, LCA applications for the H-1B program, and PERM applications. We replicated a study done on H-1Bs by the Center for Immigration Studies, and were able to conclude that their analysis was correct, but their focus was too narrow and did not generalize well the the H-1B program as a whole. It also neglected to mention that while prevailing wage claims tend to undershoot the median wage, actual H-1B pay is only slightly below the median on average, and in many cases, well above the median. It should be noted though, that H-1B workers may not receive benefits and other compensation seen by citizens or PERM workers.

It is clear that PERM workers receive 15% higher wages than H-1B workers on average.

To perhaps no one's surprise, the highest paying jobs can be found in the fields of medicine and law, followed by engineering, management, and finance jobs. Education jobs trailed far behind most other fields.

We found the most successful attorneys for processing H-1B LCA forms in each state. The top attorneys by far were Priscilla Muhlenkamp in Washington, D.C. and Oishika Gupta in Chicago, IL.

We also found that the highest paid graduates tend to come from Carnegie Mellon University.

6   Future Work

We were very ambitious when we started this project. Our end goal was to build a Shiny app that would allow users to filter jobs down to the individual lat-long coordinates and determine where the highest paying jobs are when accounting for cost of living.

We built a working Shiny app and had collected geographical data for each company in the PERM data set. However, we ran into resource limitations on ShinyApps.io. The data set is large, and even when removing unnecessary columns, the load time for the application exceeded the memory of the free accounts offered on ShinyApps.io. Aggregating the data reduced the usefulness of the app, making it cumbersome and inflexible.

As a future project, it would be ideal to have a paid host that could accommodate the processing requirements of this data.

In addition, it would be nice to be able to update this report year-over-year as new data comes out. This should be feasible and take minimal effort as long as the format of the data does not change.

References

[1] U.S. Department of Labor, Bureau of Labor Statistics, *News Release*, May 19, 2016 https://www.bls.gov/news.release/archives/forbrn$_0$5192016.*pdf*

[2] Argueta, Carla N. Congressional Research Service *Numerical Limits on Permanent Employment-Based Immigration: Analysis of the Per-country Ceilings*, July 28, 2016 https://fas.org/sgp/crs/homesec/R42048.pdf

[3] American Immigration Center *How Many Immigration Applications Are Filed Each Year?* https://www.us-immigration.com/how-many-immigration-applications-filed-each-year/

[4] Zong, Jie. Batalova, Jeanne Migration Policy Institute *Frequently Requested Statistics on Immigrants and Immigration in the United States*, April 14, 2016 http://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states

[5] U.S. Department of Labor, Office of Foreign Labor Certification Disclosure Data, https://www.foreignlaborcert.doleta.gov/performancedata.cfm

[6] Detweiler, Brian. GitHub *stat-8416-final-project* https://github.com/bdetweiler/stat-8416-final-project

[7] Bureau of Labor Statistics *Employment, Hours, and Earnings from the Current Employment Statistics survey (National)* http://data.bls.gov/timeseries/CES0000000001?output_view=net_1mth

[8] Bureau of Labor Statistics *May 2015 National Occupational Employment and Wage Estimates* http://www.bls.gov/oes/current/oes_nat.htm

[9] Kerr, William R. Lincoln, William F. *The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention.* National Bureau of Economic Research, http://www.nber.org/papers/w15768.pdf

[10] Mukhopadhyay, Sankar. Oxborrow, David. *The Value of an Employment-Based Green Card* PubMed, https://www.ncbi.nlm.nih.gov/pubmed/22161232

[11] Office of Foreign Labor Certification *Annual Report 2015* https://www.foreignlaborcert.doleta.gov/pdf/OFLC_Annual_R

[12] Bureau of Labor Statistics *Inflation Calculator* http://www.bls.gov/data/inflation_calculator.htm

[13] Miano, John *Low Salaries for Low Skills: Wages and Skill Levels for H-1B Computer Workers, 2005* Center for Immigration Studies http://cis.org/LowSalariesforLowSkills-H1B