

H-1B Job Market Explorer

Abstract

The H-1B Job Market Explorer allows users to explore H-1B jobs in the U.S. Users can traverse possible openings by worksite location, job type, company and combinations thereof. It highlights wage information and number of available positions by job type and company. Using the Explorer we found regional hotspots for particular occupation types, major employers in each region, as well as general wage distributions within companies and regions.

Introduction

H-1B visas allow United States companies to employ foreign workers in specialty occupations (Immigration and Nationality Act, section 101(a)(15)(H)). Given the current political climate, H-1Bs are key to allowing foreign workers into the United States. Furthermore, if an H-1B holder quits or is terminated from a sponsoring employer, they must either be granted a change of state, find another H-1B employer or leave the country. As such, for those seeking to come to the U.S. or change jobs, knowing where to find these jobs is crucial. The H-1B Job Market Explorer allows users to explore national and regional H-1B job markets, the types of jobs available and the major employers in the area.

Dataset

The source for this dashboard is a tabular dataset from the U.S. Department of Labor Office of Foreign Labor Certification with details of employers' Labor Condition Applications for the 2017 fiscal year. The original dataset contained 54 features (columns) and 624,650 records (rows). To reduce size, records with nulls were dropped and data filtered on various columns.

Figure 1. Data Filters

Attribute	Condition
All	Not Null
CASE_STATUS	'Certified'
WAGE_UNIT_OF_PAY	'Year'
SOC_CODE	Is Valid
VISA_CLASS	'H-1B'
WORKSITE_STATE	Is Valid

Specifically, we only considered records classified as 'Certified', and where the visa type was 'H-1B'. Additionally, records for worksite locations outside of the fifty United States were excluded, as were records for irregular job types (ones that could not be mapped back to

https://www.bls.gov/oes/current/oes_nat.htm). Lastly, only rows with nonzero reported wages and annual salaries were considered. After data cleaning, we were left with 486,755 records. We also filtered our dataset by column, reducing our dimensionality from 54 features to 8. Doing so brought us from 33,731,100 to 3,894,040 values, reducing our data size by 89%.

The original dataset includes data related to application statuses, employer information, as well as worksite and job information. As mentioned, the H-1B Job Market Explorer focuses on a subset of these features, detailed below.

Figure 2. H-1B Job Market Explorer's Attributes, Descriptions and Attribute Types

Attribute	Description	Type
TOTAL_WORKERS	Total number of foreign workers requested by the Employer(s)	Sequential quantitative
WORKSITE_STATE	City information of the foreign worker's intended area of employment	Categorical, Geographical
WORKSITE_COUNTY	County information of the foreign worker's intended area of employment.	Categorical, Geographical
SOC_CODE	Occupational code associated with the job being requested for temporary labor condition, as classified by the Standard Occupational Classification (SOC) System	Sequential quantitative
SOC_NAME	Occupational name associated with the SOC_CODE	Sequential quantitative
WAGE_RATE_OF_PAY_FROM	Employer's proposed wage rate	Sequential quantitative
WAGE_RATE_OF_PAY_TO	Maximum proposed wage rate	Sequential quantitative
EMPLOYER_NAME	Name of employer submitting labor condition application	Categorical

Tasks

The targets for the H-1B Job Market Explorer are distributions of H-1B positions and their average wages around the United States, within specific companies and within job categories. Our stakeholders are anyone interested in understanding the H-1B job market, including but not limited to current H-1B visa holders interested in switching employers, foreign workers hoping to come to America and foreigners . Analysis is expected to be discovery and general

consumption of raw data. Search with the visualization includes looking through known attributes for unknown patterns and outliers. Additionally, users can search the visualization to locate specific records. Queries are expected to be jobs that are plentiful and pay the highest wages. As such, users can summarize regional employers and jobs markets and identify specific positions and associated pay.

Figure 3. Actions and Targets

Actions
Analyze → Consume → Discover
Search → Lookup, Locate, Browse
Query → Identify, Summarize
Targets
All Data → Outliers
Attributes → One → Distribution
Attributes → Many → Similarity
Spatial Data → Shape

Solution

We build the H-1B Job Market explorer as a modular dashboard. Using that flexibility we were able to sample different idioms and their ability to guide users through exploring this dataset. Even just the three dimensions (location, job type, company) and two attributes (average wage, number of certified positions) added up to thousands of potential search combinations displayed in 3 different views:

Figure 4. H-1B Job Explorer's Dimensions and Derived Attributes

Dimension	Derived Attributes
Company	Average Wage Number of Positions
Job Type	Average Wage Number of Positions
Location	Number of Positions

Early attempts at displaying all of this information at once led to a crowded, jumbled visualization. While the effectiveness of position and length as encoding channels is high, with so many different scales in a small space there was a lot of interference.

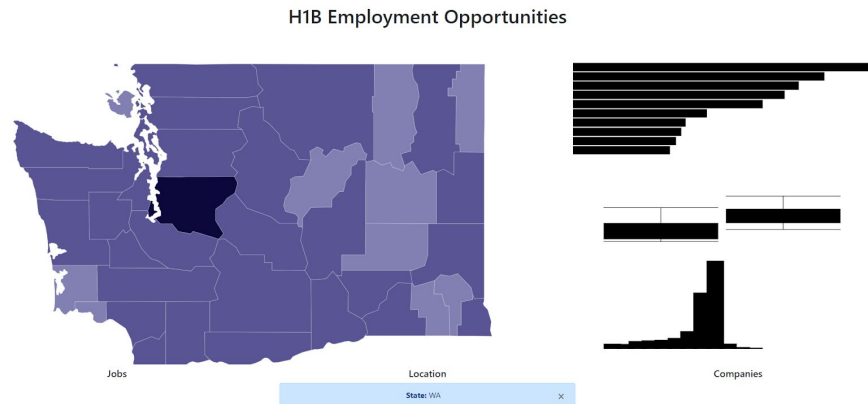
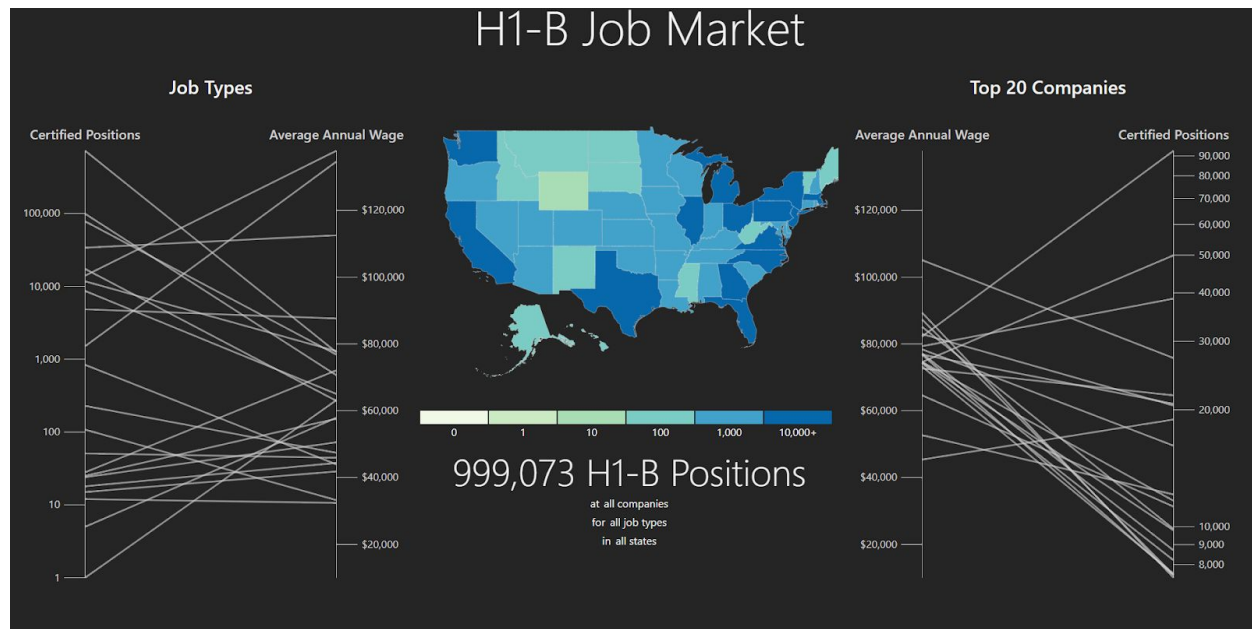


Figure 5. Early prototype of the H-1B Job Explorer

The solution was to follow Tufte's advice - to find one good design and multiply it. We chose the slopegraph as an effective way to display about 20 data points at a time of bivariate data. We created a slopegraph for both the Job Type and Company views and further simplified the display by sharing the Average Wage length scale between the views. We attempted to use a single scale for Number of Positions, however the high contrast in these values between the Job Type and Company views prevented us from using a single length scale. Even within a single view, logarithmic scales were often needed to adequately separate the plots. We ended up constructing the Number of Positions length scale dynamically, using a linear scale if the range of values was less than 100,000 and a logarithmic scale if they were greater.

Figure 6. Attribute Scales

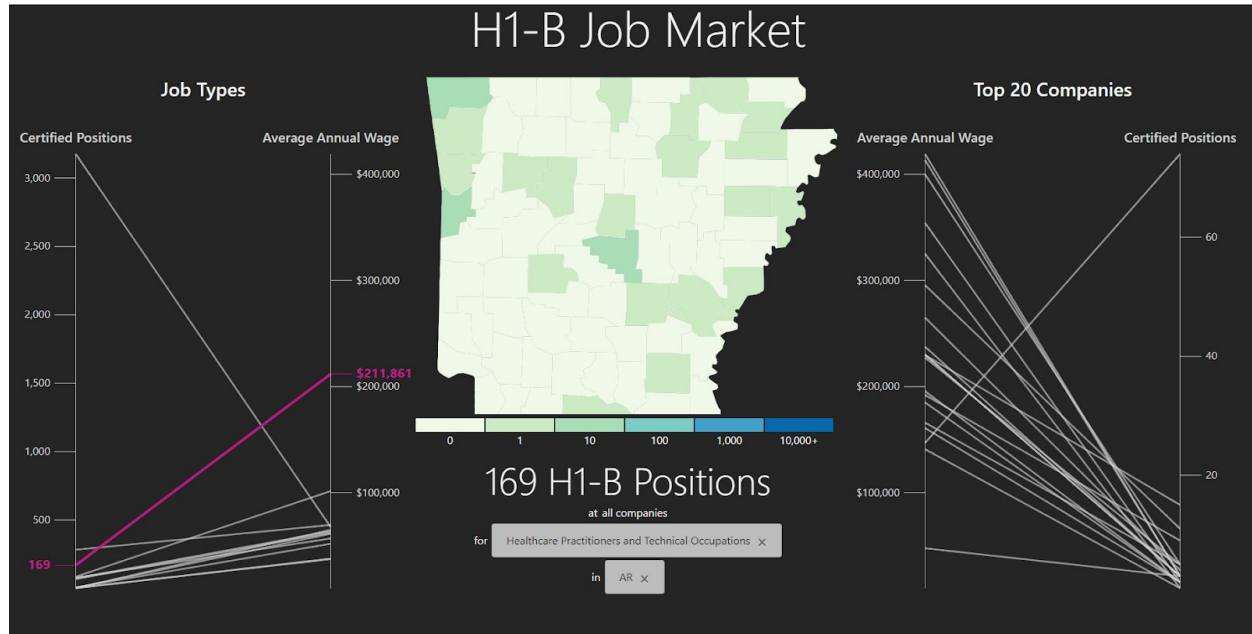
Attribute Scales		
Dimension	Average Wage	Number of Positions
Job Type	Length - Dynamic	Length - Dynamic - Logarithmic
Company	Length - Dynamic	Length - Dynamic - Logarithmic
Location	-	Color - Static - Logarithmic

Figure 7. Finalized H-1B Job Market Explorer

The location data naturally fit into a geographic visualization. We allowed zooming and panning to help mitigate the limiting view size and implemented two different map contexts, the first showing states, and the second showing counties when a specific state is selected. Data can then ultimately be filtered by county. To appropriately discretize the derived Number of Positions attribute for the map's use, we used a logarithmic scale with a multi-hue sequential color scheme from ColorBrewer. Each bucket covered an order of magnitude. The multi-hue color scheme was justified because those regions with job numbers under 100 could reasonably be considered a different category than those with 10s of 1000s.

With as much information as possible packed into two derived features, three consistent views, and four different scales, the last piece of functionality needed for useful exploration was interactivity. Each view allows selection of one element which is then used to cross filter the results on the other two views. For example, select a Job Type and the map and company views update with that Job Type as a filter. Select a company, the map and job views update. Select a location and all three views update.

Figure 8. H-1B Job Market Explorer filters by Region (Arkansas) and Job Type (Healthcare)



The dashboard was built primarily using the d3 javascript library. Other javascript libraries included underscorejs for filtering and data manipulation, momentjs for timestamp handling, and topojson for geo data wrangling. Styling was done with Bootstrap and Colorbrewer. Data cleaning and preprocessing was done with pandas in Python. Dynamic events were handled using a subscription-based model.

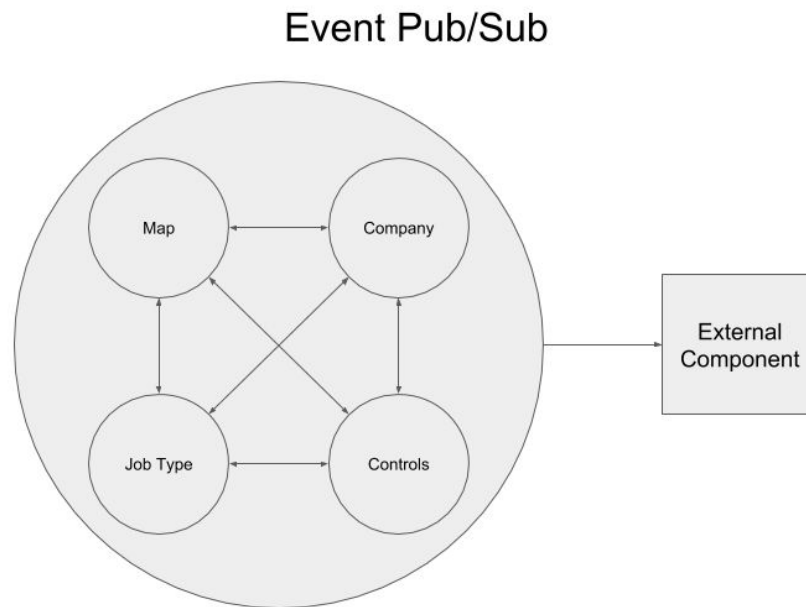


Figure 9. Event Handling Model

Results

All attributes detailed in the table above are encoded in the visualization, summarizing H-1B job counts across regions, sectors and companies. States and counties are encoded as geographic spatial regions, the most effective categorical channel. Job counts (total workers) are aggregated across sector and company and encoded using position on an aligned scale, and aggregated across geographic region and encoded using luminance. Average wages are aggregated across the same dimensions and also encoded using position on an aligned scale. There is a third attribute, the ratio between the wage and job count attributes, that is encoded in the angle of the line between the two points - an effective bonus channel as a result of the slopegraph idiom.

This is not to say that the Explorer is without limitations. Our right side panel only includes Top 20 Companies by number of jobs. It fails to show positions at small businesses. The reason for this is usability and readability. The figure below shows Top 100 companies, and although it catches the eye, it is extremely difficult for users to select or differentiate the lines. Additionally, the Job Types and associated wages are aggregated. The Healthcare Practitioners and Technical Occupations class includes everything from neurosurgeons to ultrasound technicians, which are very different jobs with very different earnings. Our dashboard groups both of these jobs together and average their salaries. Also, the dashboard does not allow for direct comparisons. Given more time and resources, we would include a drilldown options for the job types and facetings to allow for direct comparisons. To better show small companies, we would add filtering options (range of number of jobs per company) and navigation.

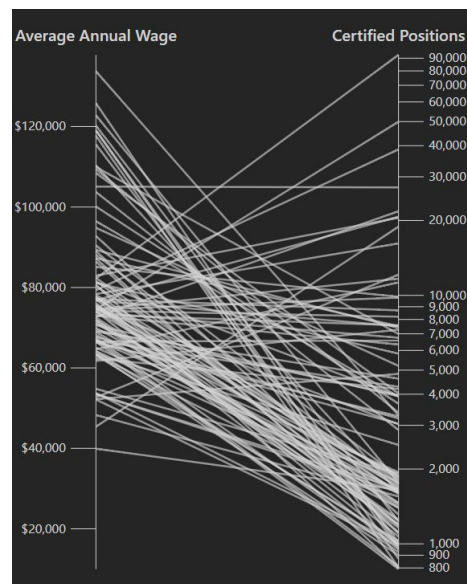


Figure 10. Explorer Side Panel Showing Top 100 Companies

The described views, channels, and interactions combine to give the user a flexible and quickly understandable overview of the H-1B job market across regions, disciplines and companies. For example you can quickly navigate to Washington state and see that while Amazon is the largest H-1B employer, on average Facebook outpays them. Or, if you're in the health and medical services field, it may be useful to know there are fewer but, in some cases higher paying jobs in the south. Or, while the high demand for lawyers in New York City may not be surprising, it could be useful to know the biggest three firms are White & Case LLP, Skadden, Arps, Slate, Meagher & Flom LLP, and Paul, Weiss, Rifkind, Wharton & Garrison LLP and that White & Case may have more entry level positions.

All of these patterns are brought out by the repeated slopegraph idiom and three-way crossfilter. With the data in a text table or other static display, the tasks described would remain out of reach. The H-1B Job Market Explorer makes the information in the data set more accessible.

Bibliography

Bootstrap. Version 4.0.0. <https://getbootstrap.com/>

Bureau of Labor Statistics - https://www.bls.gov/oes/current/oes_nat.htm

D3. Version 4.13. <https://d3js.org/>

Momentjs. Version 2.21 - <https://momentjs.com/>

Munzner, Tamara. *Visualization Analysis & Design*. CRC Press, Taylor & Francis Group, 2015.

Topojson Data - <https://github.com/deldersveld/topojson>

Topojson. Version 3.02 - <https://github.com/topojson/topojson>

Tufte, Edward R. *The Visual Display of Quantitative Information*. Graphics Press, 2011.

Underscorejs. Version 1.83 - <http://underscorejs.org/>