Name: Swathi Reddy Konatham Z-Id: Z1864290

**Final Project Report**

**H1B Visa Petitions 2014-2016**

**Link to the dataset:**

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

**Context:**

H1B is a non-immigrant visa which allows the foreign students or professionals to work in United States. For a foreign worker to apply for H1-B visa, a US employer must offer them a job and submit a petition for an H-1B visa to the US immigration department. This project work illustrates and visualizes few trends in the number of H1B application filed each year by each employer in the United States.

**Content:**

Kaggle dataset contains H-1B petition data form the years 2011 to 2016, with approximately 3 million records overall with a size of about 492.3MB. Below is the list of attributes and its descriptions that are listed in the dataset.

Description of Columns:

|  |  |
| --- | --- |
| **Column** | **Description** |
| CASE\_STATUS | Status of the Petition. The values that are used are:  a. Certified b. Certified-Withdrawn c. Denied d. Withdrawn |
| EMPLOYER\_NAME | Name of the Employer who submits the H1B visa application. |
| # PREVAILING\_WAGE | Prevailing Wage of the person for the job being requested for temporary labor condition (LCA) |
| # YEAR | Year in which H1B petition is filed |
| WORKSITE | The location details of the area of employment of the Foreign workers |
| LON | Longitude coordinates information |
| LAT | Latitude coordinates information |

**Attributes used:**

**Categorical:** CASE\_STATUS, EMPLOYER\_NAME

**Ordinal:** YEAR

**Quantitative:** PREVAILING\_WAGE, LAT, LON

**Filtered Dataset:**

Kaggle dataset has huge data which was difficult to load and visualize, so the data has been filtered to consider 2014, 2015 and 2016 years H1B petitions.

As the CSV file contains approximately 3 million records, the file couldn’t be opened in Excel and Numbers. To preview the entire file I have used **“TAD”** tool. The tool is also used to filter the data for the required years (2014 - 2016) and the state of ‘Illinois’. The filtered CSV file is reduced to 14MB with approximately 93,000 records, it is loaded in GIT and used as reference in the code.

**Questions my visualization answers:**

1. Who are the top Employers (top 10, top 15 and top 20) that have filed H1B petitions in all the three years (2014 and 2015 and 2016) or in a particular selected year. - This is visualized using **Bubble Chart**.
2. Identify the trends in the H1B petition filed by an Employer in comparison with all the years. - This is visualized using **Grouped Bar Chart**.
3. For each of the top employer, identify how many H1B petitions are Certified, Withdrawn, Denied and Certified-Withdrawn status. - This is visualized using **Sunburst Chart**.
4. What are the Worksite locations of the employers that have filed H1B petitions. - This is visualized using **Choropleth map*.***
5. How many H1B petitions are filed from a particular worksite location - This is visualized using **Choropleth map*.***

**Design:**

***Iteration 1:***

The number of H1B Petitions filed by each employer is calculated using d3.js in **createVis** method by iterating over the entire data and calculating the count of rows for an employer. To get the count of H1 Petitions for a selected employer, the key is used as combination of Year + Employer. This will be used as a quantitative attribute to plot Bubble chart and Grouped bar chart.

* 1. **Year drop-down:**
     1. On page load, a drop-down menu is displayed for the user to select a particular year.
     2. By default, the initial value of the drop-down is set to *‘-All-‘* which indicates the data is calculated for the combination of the 3 years (2014-2016).
     3. The drop down has the values 2014, 2015 and 2016. On change of the year the entire visualization is recalculated and loaded for that selected year.
  2. **Bubble Chart:**

**Marks:** Point at x position and y position

**Channels:** Size(Area), Color Hue

* + 1. The bubble chart is plotted with y-axis as the count of H1B petitions and the x-axis as the prevailing wage allocated by the Employer.
    2. On page load, the bubble chart displays the top 10 Employers calculated for all the years 2014, 2015 and 2016.
    3. The top 10 Employers are calculated based on the count of number of H1B petitions by sorting and slicing the list either for a selected year or for the combination of the three years in **getTopEmployers** function.
    4. Each ***bubble represents an Employer*** which is a categorical attribute hence ***Color Hue*** is used to identify each employer.
    5. The ***size of the bubble is determined by the count of number of H1B petitions***.
    6. For each year selected using the drop-down menu, the bubble chart is updated with the top 10 employers in that selected year.
    7. Legend is displayed matching each employer name with the bubble in the bubble chart.
    8. On selection of different years from the dropdown, the bubble chart axis is ***not re-scaled*** providing good viewpoint to the user to follow the change.
  1. **Choropleth Map (Illinois State):**

**Marks:** Point, Area

**Channels:** Color Hue, Position

* + 1. The Choropleth map uses Illinois state projections.
    2. On page load, the Illinois state map has all the worksite locations from which the employers have filed H1 petitions is plotted.
    3. The worksite locations are identified using circles which by default has black color fill.
    4. On page load, the top 10 employers which are plotted in the bubble chart are highlighted with the ***same color encoding*** in the choropleth map.
  1. **Grouped Bar Chart:**

**Marks:** Line

**Channels:** Color Saturation, Spatial Region, Position

* + 1. The Grouped Bar Chart is drawn by grouping each employer for the three years (2014, 2015 and 2016) along x-axis and the number of H1B petitions along y-axis.
    2. As the years data are Ordered attributes, ***Color Saturation*** is used to identify each year.
    3. To identify the employer, each employer name is labeled accordingly along x-axis using ***Spatial Region*** as the channel along with the color encoding.
    4. ***The bar chart will be resized and adjusted to fit in the same screen space for every multiple employers being selected.***

**Interactions:**

1. The bubble chart has **multi-select feature**. On click of the employer on Bubble chart:
   1. The worksite location related to the employer is highlighted in the choropleth map using the same color encoding as that of the employer. For user feedback, the highlighted circle has increased radius and white stroke with stroke width increased. If multiple employers are selected then all the worksites related to the employers will be highlighted.
   2. A Grouped bar chart will be created with that employer comparing the number of H1 petitions filed in 2014, 2015 and 2016. If multiple employers are selected, then the grouped bar chart will be appended with all the selected employers.

2. On **double-click** the employer will be deselected:

* 1. The worksite which was highlighted earlier, will know be de-highlighted with reduced radius and stroke width in choropleth map.
  2. The grouped bar chart will be updated to remove the deselected employer. If all the employers are deselected then the entire grouped bar chart will be removed.

3. MouseOver and MouseLeave:

* 1. On mouseover on the employer in bubble chart, the employer name and the count of number of H1B petitions filed by the employer is displayed as tooltip.
  2. On mouseover of the worksite location in choropleth map, the worksite name and the count of number of H1B petitions filed by all the employers located in that particular worksite is displayed as tooltip.
  3. On mouse leave of bubble chart and choropleth map, the tooltip is hidden.

***Iteration 2:***

* 1. **Sunburst Chart:**

**Marks:** Area

**Channels:** Angle, Color Hue

* + 1. The sunburst chart displays the status of H1B petitions filed by each selected employer. The status values include - Certified, Denied, Withdrawn and Certified-Withdrawn.
    2. The inner circle represents the employer and the outer circle represents the status of petitions.
    3. The arc angle and area is determined by the count of petitions.
    4. ***The color of the employer and status is same as the color encoding used for employers in bubble chart.***
  1. **Count of Top Employers dropdown:**
     1. A dropdown menu is provided to enable users to select top 10, top 15 and top 20 employers.
     2. On change of top employers dropdown, the bubble chart and choropleth map are updated with the selected number of employers.

**Interactions:**

1. On click of the employer in bubble chart:
   1. The sunburst chart is created for that employer displaying the status of the H1petitions. If multiple employers are selected then the sunburst will be updated with all the selected employers.
   2. On double-click of employer in bubble chart, the sunburst chart will be updated to remove the deselected employer.
   3. If all the employers are deselected, then the entire sunburst chart will be removed along with the label present below the sunburst being hidden.
2. On MouseOver and MouseLeave:
   1. On mouseover on the arcs of the sunburst, the opacity of the arc is reduced (differentiating form the other) and the related employer name, status and count of H1 Petitions are displayed beside the sunburst chart.
   2. On mouse leave the text is removed.

***Iteration 3:***

As per the presentation feedback, changes done to Grouped Bar chart to use the ***same color encoding*** as of that the Bubble chart for the bars which are used to group the Employers, the color saturation is calculated for each color hue using the brightness of that categorical color.

**Future Work:**

1. Analyze the prevailing wage of the employers, to find the top employers with more wage.
2. Instead of visualizing only Illinois data, extend the visualization for all the states in USA.

**References:**

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

<https://www.foreignlaborcert.doleta.gov/performancedata.cfm>

<https://www.foreignlaborcert.doleta.gov/docs/Performance_Data/Disclosure/FY15-FY16/H-1B_FY16_Record_Layout.pdf>