TDS3401 Project

Group Members

1. Khor Kia Kin - 1142701883
2. Kirthivaasan Puniamurthy - 1142702268

Exploring Google Careers

This project explores several dimensions of Google careers and comprises of four different visualizations. These are:

- A bar chart which shows the popularity of various programming languages used in Google.
- A bar chart that shows the number of jobs available for various job categories with respect to years of experience required.
- A word cloud enumerating some of the most popular words for the preferred qualifications of a job applicant.
- A map visualization visualizes the number of available jobs in Google according to country.

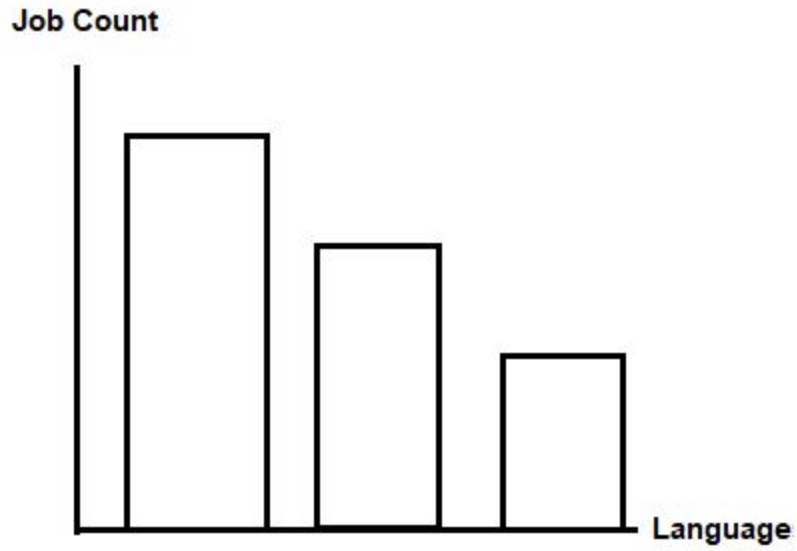
Google Job Skills dataset contains scraped data from Google Careers website. It is available for download at Kaggle. The columns are:

Column Name	**Description**
Title	The title of the job
Category	Category of the job
Location	Location of the job
Responsibilities	Responsibilities for the job
Minimum Qualifications	Minimum Qualifications for the job

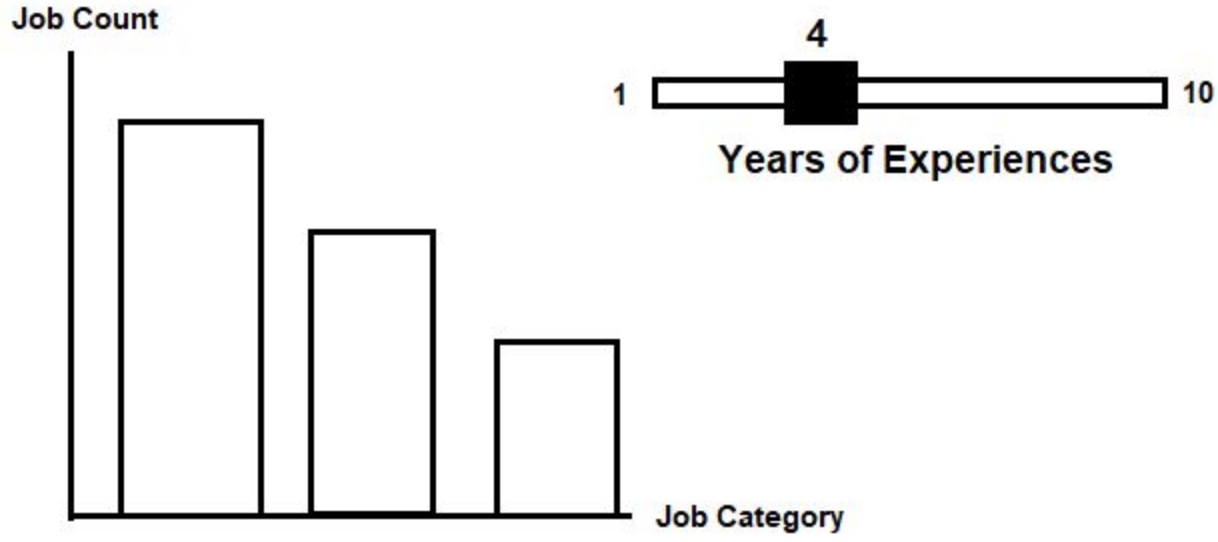
Preferred Qualifications	Preferred Qualifications for the job (extra qualifications)

Storyboard

(1) Popularity of programming languages

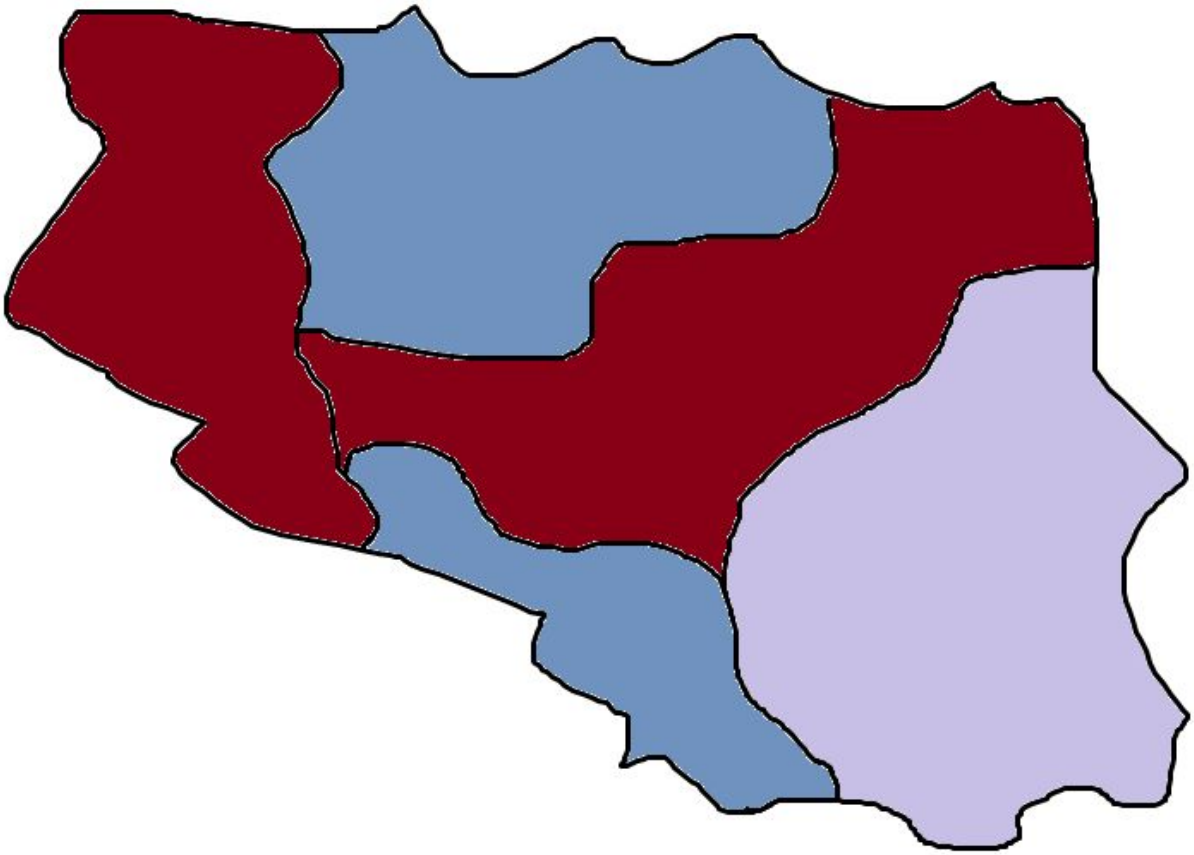


(2) Number of jobs available according to job category



(3) Top keywords in qualifications

(4) Number of jobs available in a particular location



(Preset view buttons)
When hover to certain location, the page should display location name and job count in that area.
We will use color gradient to indicate the degree of number of jobs available. For illustration purposes, we use red for higher number of jobs, blue for lesser number of jobs, light purple for no jobs available.

Changes between Storyboard and the Final Implementation

(1) Popularity of Programming Languages

- The axes' labels are updated. Job count is changed to number of jobs while language is changed to programming language.
- The bars are blue by default, and orange when hovered.
- The job counts are shown on top of the bars.
- Since the list of programming languages is hardcoded, it may not be exhaustive.

(2) Number of Jobs available for Job Categories with respect to Years of Experience

- Instead of displaying the bar chart vertically, the bar chart is displayed horizontally to accommodate the long text of each job category.
- Job count (now a x-axis label) is replaced with number of jobs.
- The bars are blue by default, and orange when hovered.
- The job counts are shown on right of the bars.
- The title is updated to include years of experience.
- Some of the job offers did not specify the years of experience required, so we have added an option called *Not Stated* to the leftmost of slider.

- We also add an option called *ALL* to the rightmost of slider which displays everything regardless of years of experience. This is the default selection.
- When user interacts with the slider, a simple transition on the bars is shown.

(3) Top Keywords in Preferred Qualifications

- A word cloud is generated to visualize the top keywords in the preferred qualifications column. An external library by Jason Davies was used for its implementation.
- The frequencies of each of the words were counted.
- The common words such as ‘the’ and ‘a’ were filtered from the results as they do not mean much.
- The word cloud includes tooltip functionality to view the exact count of each word displayed.
- Filters for the words extracted from preferred qualifications were not applied. Context would be required to accurately make filters work. For instance if the person finds the most frequent words by job category this will not make very much sense with just unigrams (single words). It would require more advanced natural language processing techniques to extract meaning from the sentences.

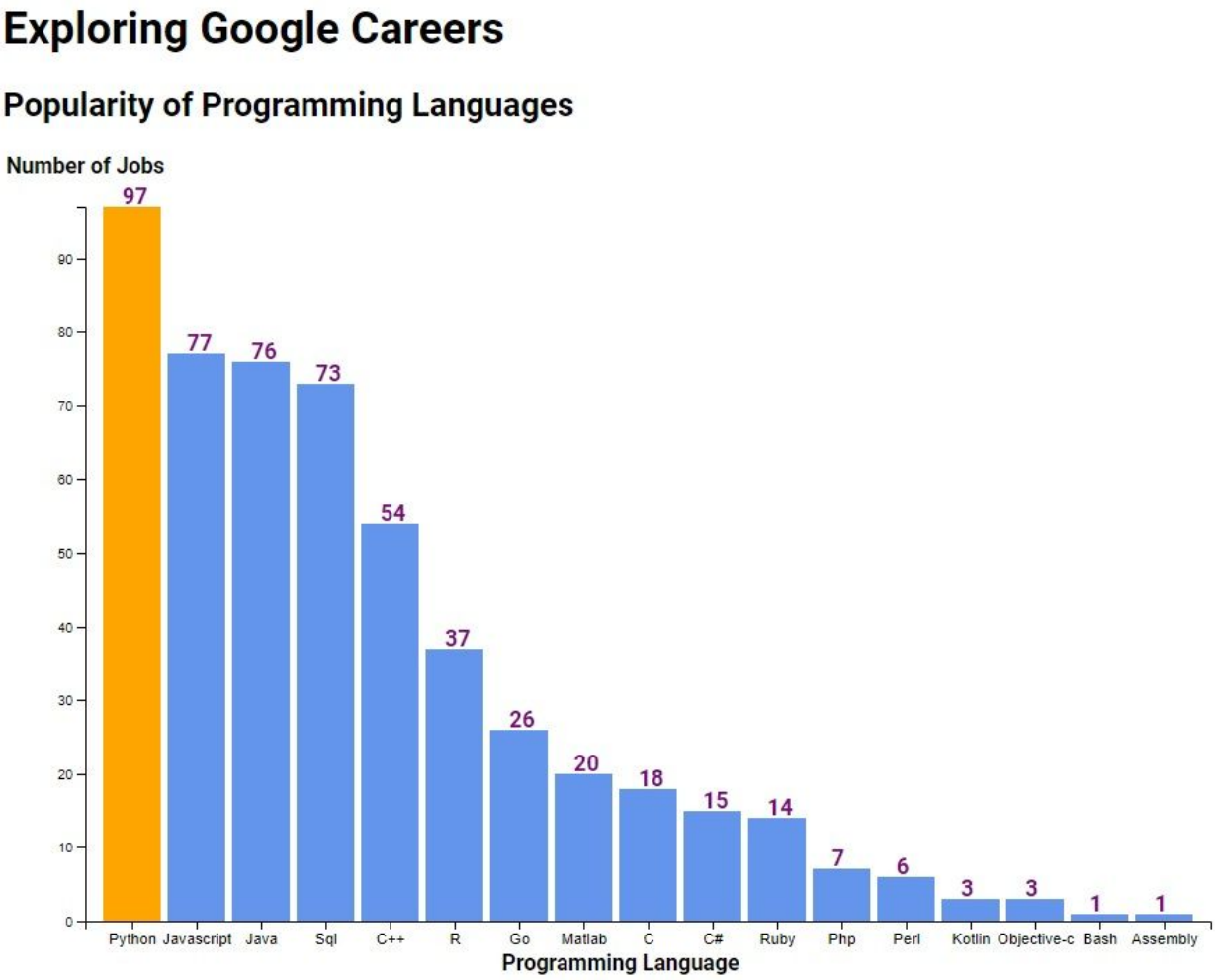
(4) Number of Jobs by Country

- A choropleth was created to visualize the job density in different countries.
- A tooltip appears when hovering over a given country, displaying the country name as well as the number of jobs found there.
- Legend for the map was added.
- Zoom is limited to a certain scale, so that the user does not get lost in the map.

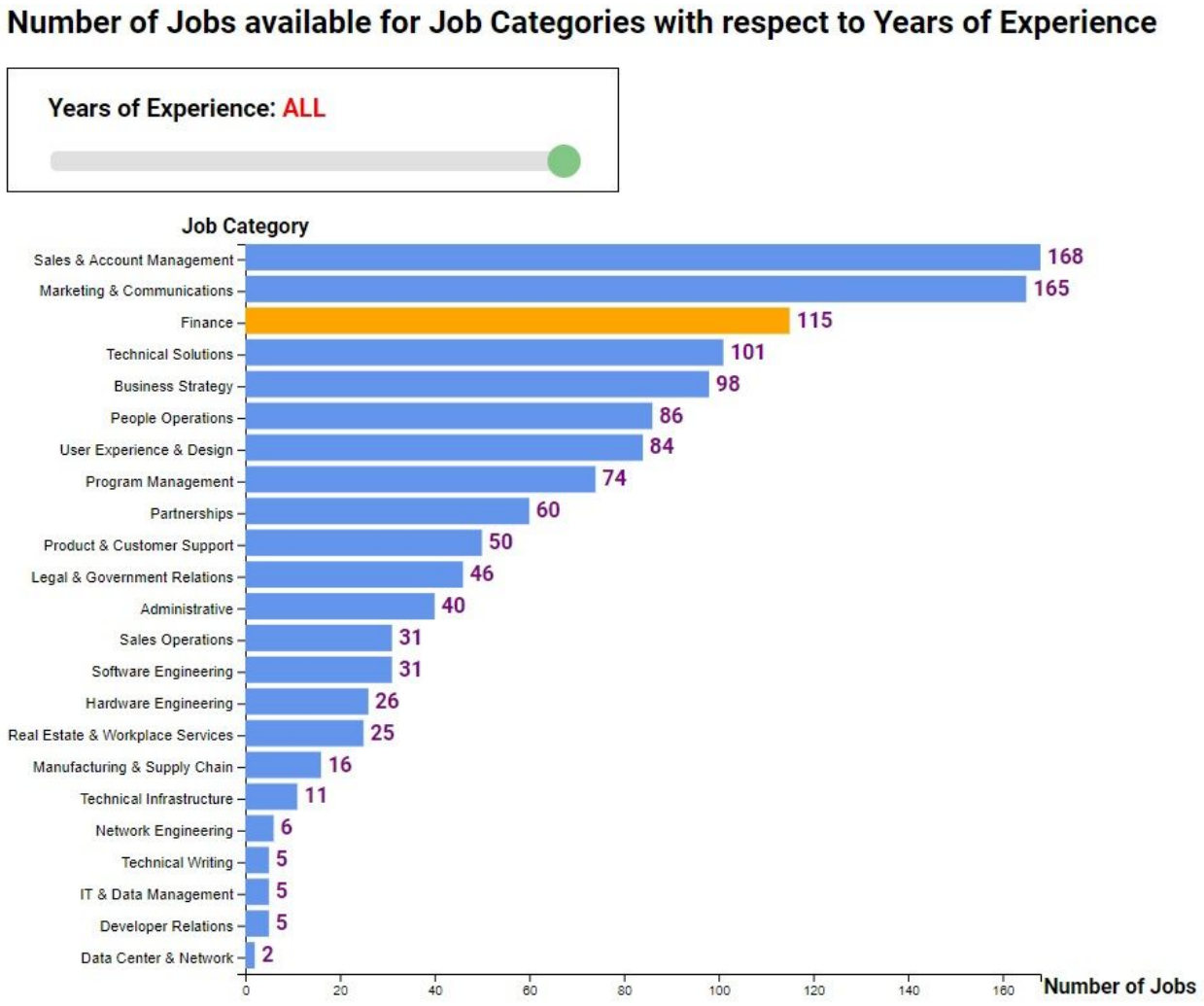
- Preset view button is not implemented since it does not focus on one specific country, but multiple countries. Originally it was assumed that only few countries would be displayed and this would be important. However this was not the case.

The Final Implementation

(1) Popularity of Programming Languages

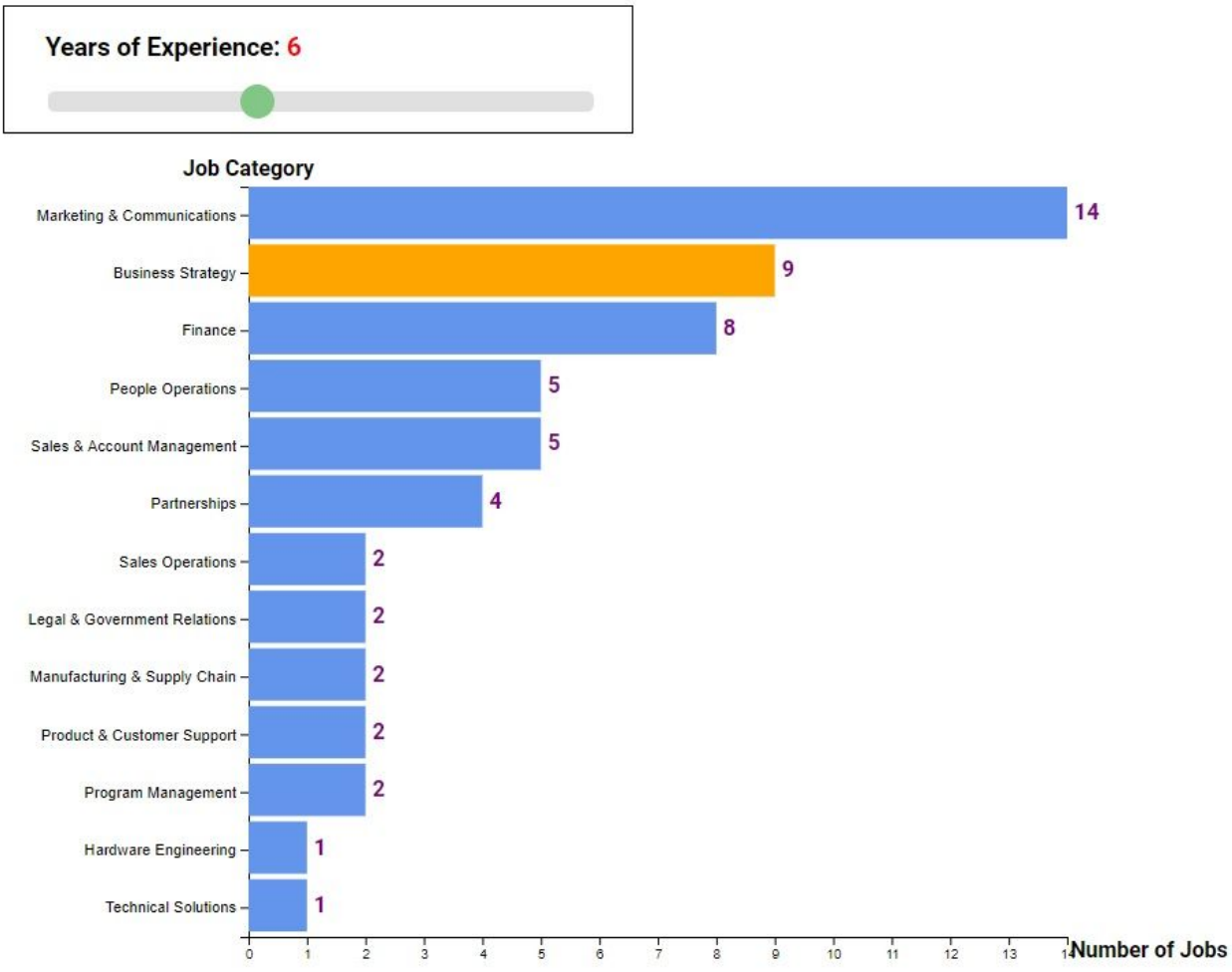


(2) Number of Jobs available for Job Categories with respect to Years of Experience



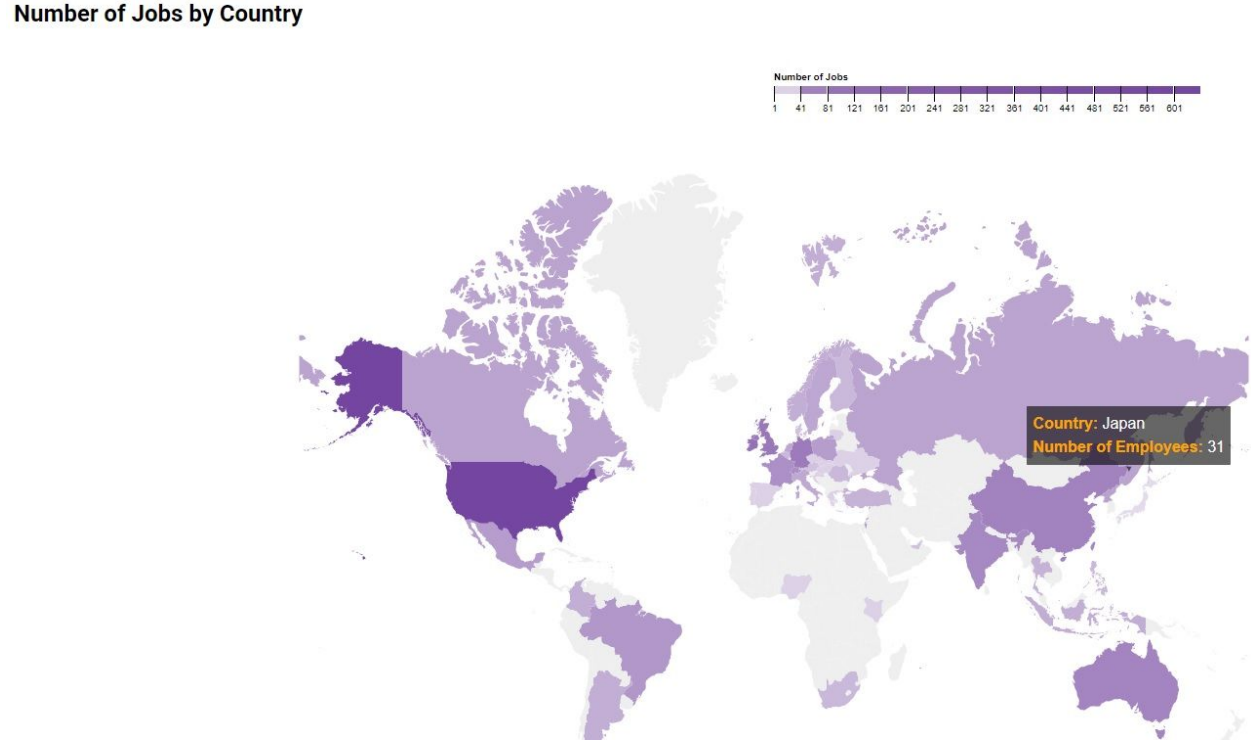
After interacted with slider:

Number of Jobs available for Job Categories with respect to Years of Experience



(3) Top Keywords in Preferred Qualifications

(4) Number of Jobs by Country



Development Process

Kia Kin worked on the first two visualizations whereas Kirthi worked on the latter two visualizations.

(1) Popularity of Programming Languages

The barchart was implemented by mainly referring to Mike Bostock's block. By referring to other blocks online, axes' labels and showing actual job count on top of the bars were implemented. Overall, it took a couple of hours to implement the barchart. Some time was spent on understanding how to use regular expressions in JavaScript to process the text.

(2) Number of Jobs available for Job Categories with respect to Years of Experience

Initially, the code from previous visualization was reused. Again, some time was spent on regular expressions to process the text. However, the text labels for different job categories were overlapping with each other. So a horizontal barchart was implemented instead of a vertical barchart, which took around two hours of tweaking. Next, a couple of hours were spent on implementing the slider, its associated event listener and extending the slider to include *Not Stated* and *ALL* options.

(3) Top Keywords in Preferred Qualifications

The visualization for this section was the word cloud. Initially, the task was attempted without the aid of a library. The results were fairly non hideous, but it was soon decided that the visualization will be more effectively done with the utility of a library and proper algorithm. The library that was used is by Jason Davies. Around 4 hours was spent experimenting with procedural methods to generate word clouds. However when this was abandoned and

the library was adopted for word cloud, around 2 hours was needed for integration.

(4) Number of Jobs by Country

The map visualization took around 2 days to complete. The first day was dedicated to resolving names in the topojson file and in the dataset. The problems mainly arose from the difference in naming convention and topojson files. The issue with the topojson files was the level of granularity. Singapore and Hong Kong were missing from the map and these are important locations for Google and cannot be omitted (41 jobs in Singapore, 8 in Hong Kong). Hence, a map with better precision was used. However the topojson of the more detailed map had names like People's Republic of South Korea instead of just South Korea. Hence a table was created translating each country appearing in the Google dataset with its corresponding name on the map (topojson). This was done for around 80 countries. The next day was spent creating the choropleth, integrating tooltip and adding a flexible legend.