FIT5147 – Visualisation Project

Introduction:

The data set which I have taken is **H1B USA Visa 2017** Case Status dataset. he H-1B is a visa in the United States under the Immigration and Nationality Act, section 101(a)(15)(H) which allows U.S. employers to employ foreign workers in specialty occupations [Wikipedia].

Description: This data has details of employers who employed H1B visa individuals and details related to their petition. This Website helps you to understand the different outcome of USA H1-B visa case status based on numerous criteria such as Location based, Employer based, Salary based and other factors like H1B dependent category.

Users will be able to select different conditions and compare the previous case status outcome and predict their chances of getting an H1-B visa.

Trying to convey: What location, job role and which company can increase their chances of getting their visa.

Target audience are employers who are going to apply H1B and who wants to try to visualize the previous case status for other employees who has similar background to visually see what chances there are of getting a USA visa and what job role's they need to get in to in order to increases their chances of getting H1B visa.

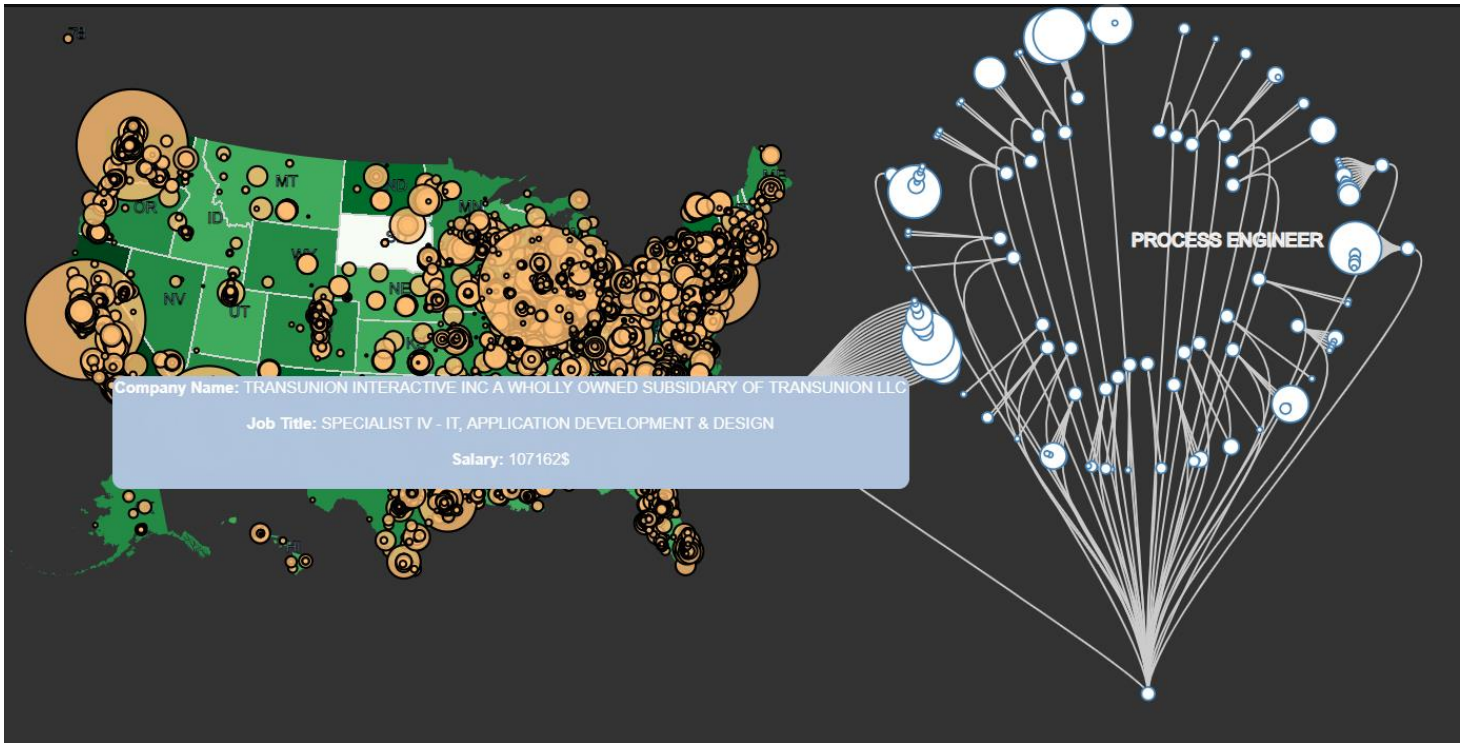
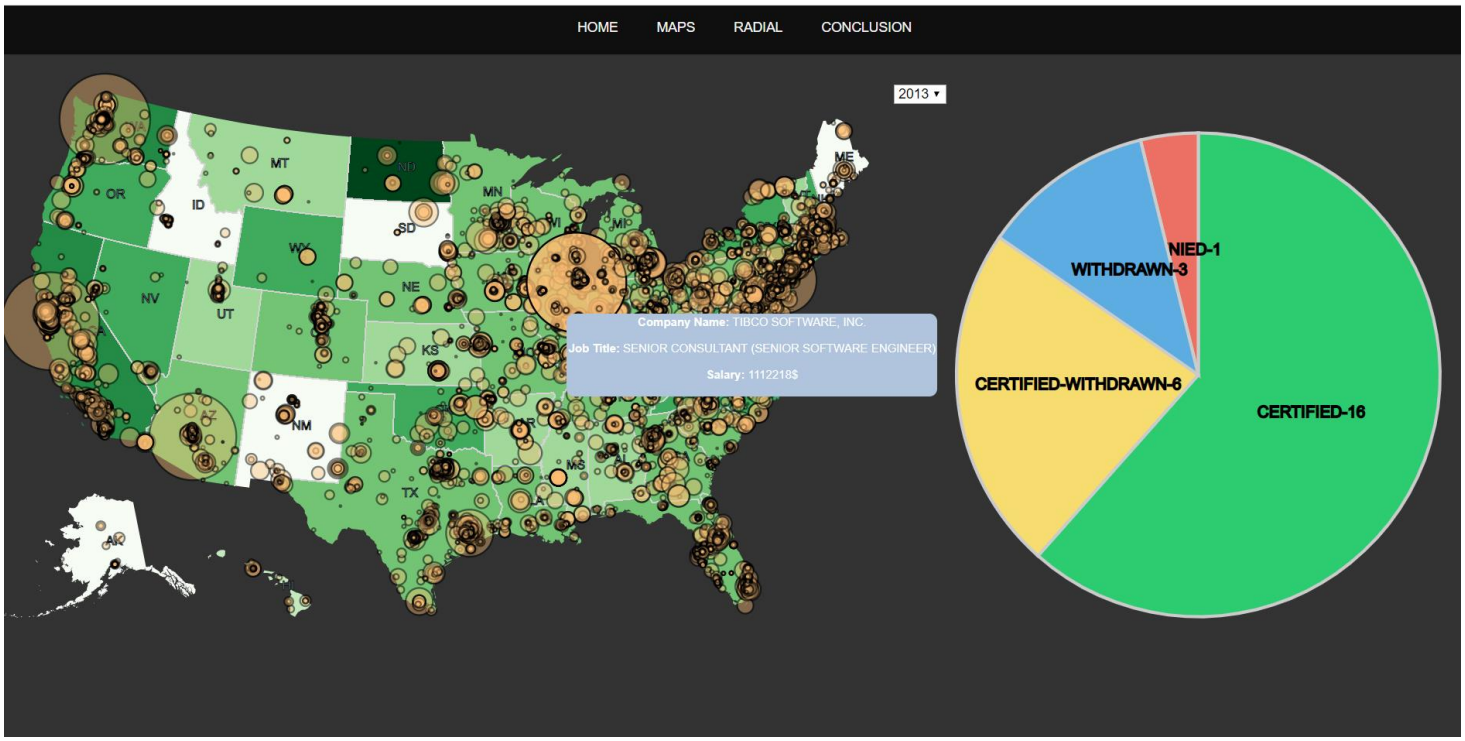


Image of Sheet 5

2 Design



This is my initial Design sheet Where I have layered out a USA Map and the points represent the worksite location of the employers who applied the visa for.

The Heat Map is generated from various calculations from the data set and grouped based on different states available in the US and the white areas where data is not available for that year / state.

The greener the are the more salary they get, and the circles size represents the salary size of that employee for that job type.

The year drop down changes the salary heat map only.

Thus, clicking on one of the circles will show the case status outcome of the particular company selected and where the pie chart colours represent each type of case status available and the count inside the pie chart is the total employees based on different category.

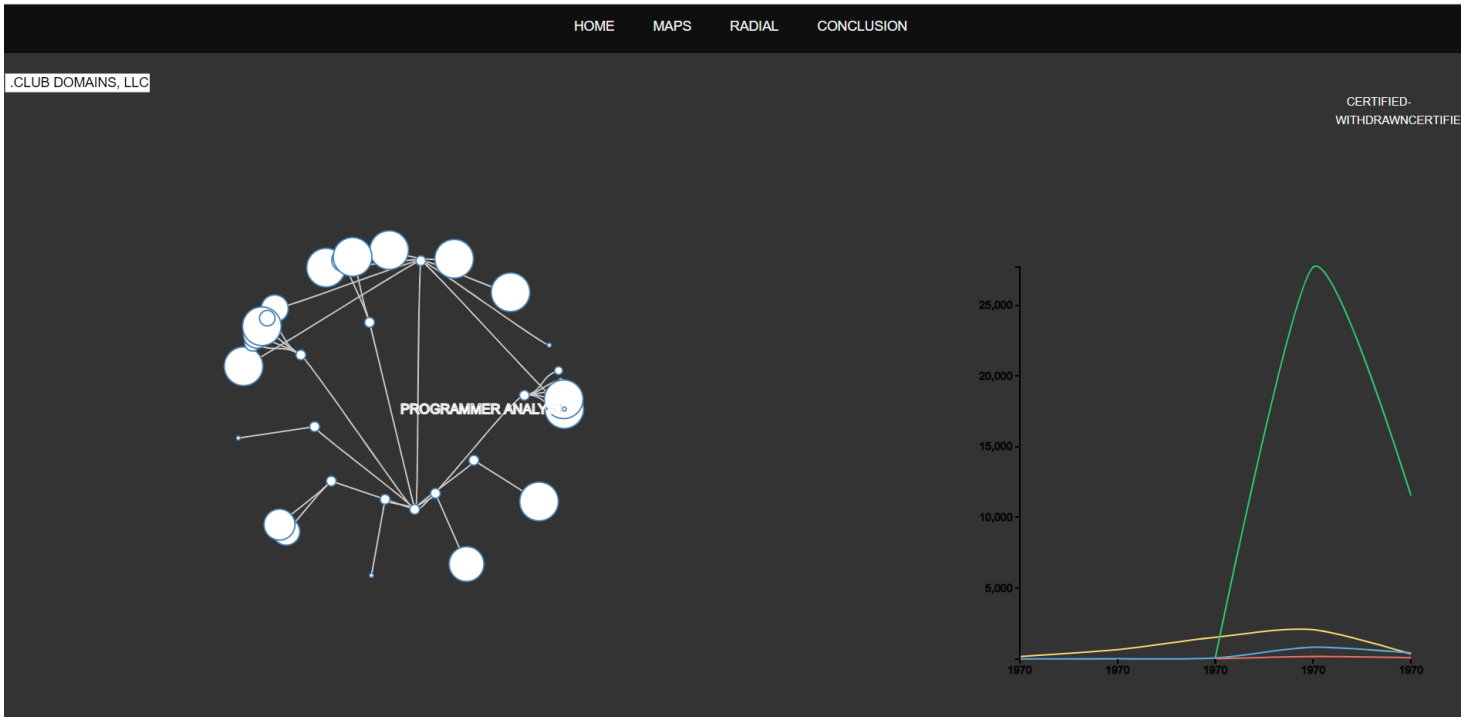
Other Ideas

I had other ideas like selecting only states map instead of entire US maps but giving a bigger picture is better.

Though of giving different colours to circle location could give a better picture but it just made it look uglier and har to find the data points.

A single stacked bar chart could give a better picture instead of a pie graph, but the pie was the better option to show the split.

Thought of, when a user selects a data point it should filter the company and the job title but it didn't give a idea about the company. Hence, I filtered only the company.



The second sheet talks about the job title the tree graph generated here is based on the company and the branches here are the top-level job type and the leaf nodes are the job roles and when clicking on the job roles the line graph generates. Each colour represents each case status and the height represent the no of employees and the width is the years from 2013 to 2017 and the graph changes based on the current company and current top job level and the job type finally.

The dropdown helps to select the company. Where the tree generates a different graph for each individual company.

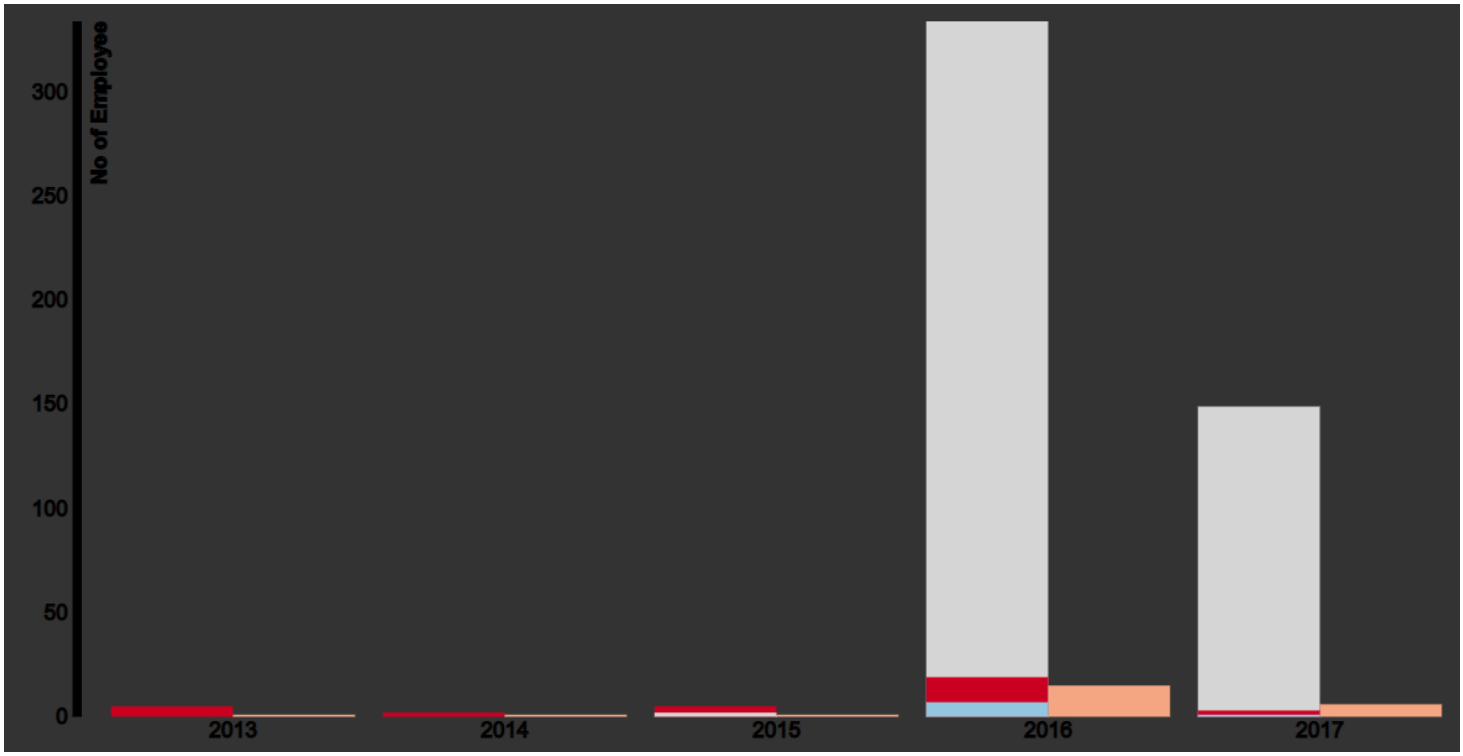
The line: Selects the job title and checks in data source for the same job title and get the out come for any company the selected job title. (Gives a good different ideas and patters to regonise.)

The circle represents the salary size and the length represent the no of employees for that particular case status.

Thus, giving a very good representation of tree.

Other Ideas:

1) Small node network: Instead this was a better option to narrow down a particular company if a person is interested in one company and one job type.
2) Line graphs give us the slop, but the pie or a bar chart doesn't show us the drop
3) Force bubble connected line was a good alternative, but tree is more structured and better in generating a great visual effect. I saw extremely beautiful trees for "Accenture"
4) Radial tidy tree could have helped but having just a circler way makes things harder and this is clickable and reduces the size if the tree is bigger.



In sheet 5 – Select the company from the map → Select the Job type from the Radial Tree → The case status appears.

Thus, in the design sheet 5 The case status is shown in a stacked bar chart way.

Implementation:

I have implemented this in D3. D3 is way better than R studio where R studio is more of a statistical tool where I have already done my calculation in R studio and wrangling in python already and got the necessary data set thus creating it in D3 gives a better visualisation effect and user visual friendly as well.

Libraries I have use:

d3.v3 – for importing D3 package

scale chromatic- for colour variation based on value

queue- to queue the stack till the process is completed

topojson – used for better visual maps

jquery – for manipulating html

filesaver- better file saving format

bootstrap- for better CSS.

Reasons for the implementation decisions for your narrative visualisation

Heat Map- Salary heat map for selecting the sate

Location points – Easy to locate

Location size – bigger the salary circles the better the options

Pie – shows the split perfectly

Radial Tree- Creates a beautiful tree structure and selects the leaf and the parent , intermediate nodes easily.

Line – shows the drop in particular case status.

Grouped bar chart - This is the best way to show the case status for one company → one Top level job role → one job type→ for different years.

5) User Guide:

User selects the "**Map's Section**" and Hoves over the code and the selected hover give a tool tip of the company and salary and job type.

Once clicked the pie graph generates based on the selected company

User selects the **"Radial Section"** and selects the drop down and once licked the tree is generated based on the company and the available job types. hover on top of it to get the job title. Once clicked on leaf nodes the line graph is generated based on the job title (leaf node)

Select the **conclusion section** Select the particular location (employer) and the tree is generated and selecting the job type from the tree gives the case status out put for different years based on case status.

Conclusion: I have learnt many ways to identify a problem from the data given and think of various ways to find different patterns and understand the data. From the design sheet ideas, I was able to narrow down exactly what I way to do and how I want to find patterns and then implementing to find amazing patters is the best way to visualise the data.

I have pretty much achieved most of the things from the 5 design sheets into real implementation.

I could have used Web works for optimising calculations and tree generations for faster web browser effects due to very less time given. I focus more on visualisation. (It's such a huge data set……. OMG!)

Reference:

https://www.youtube.com/watch?v=eQDD1DYxbvw

https://codepen.io/vlad-bezden/pen/OMEXJz?editors=0012

https://stackoverflow.com/questions/1960473/get-all-unique-values-in-a-javascript-array-remove-duplicates

https://stackoverflow.com/questions/5223/length-of-a-javascript-object

https://stackoverflow.com/questions/24232725/how-do-i-install-underscore-js

https://stackoverflow.com/questions/3390396/how-to-check-for-undefined-in-javascript

https://developer.mozilla.org/en-US/docs/Web/API/Web_Workers_API/Using_web_workers

https://stackoverflow.com/questions/43107109/build-nested-json-in-javascript

https://stackoverflow.com/questions/48243569/how-to-save-pandas-dataframes-rows-as-json-strings

https://bl.ocks.org/almccon/410b4eb5cad61402c354afba67a878b8

https://stackoverflow.com/questions/25053905/how-to-set-a-tooltip-width-dynamically

https://www.youtube.com/watch?v=EiPytIxrZtU

https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/How_the_Web_works

http://bl.ocks.org/enjalot/1203641

https://codepen.io/fernoftheandes/pen/pcoFz

http://jsfiddle.net/augburto/YMa2y/

https://stackoverflow.com/questions/18578388/html-dropdown-select-with-text-wrap-and-border-after-every-value-option

https://stackoverflow.com/questions/20797521/increasing-gap-between-nodes-of-my-d3-tree-layout

https://stackoverflow.com/questions/46977965/find-specific-nodes-in-d3js-tree

https://blockbuilder.org/bricedev/0d95074b6d83a77dc3ad

https://www.pexels.com/photo/close-up-photo-of-people-holding-usa-flaglets-1449057/

https://www.kaggle.com/jonamjar/h1b-data-set-2017

Appendix

FACTOR'S THAT INFLUENCE H-1B US VISA?

MAIN IDEA SHEET-1

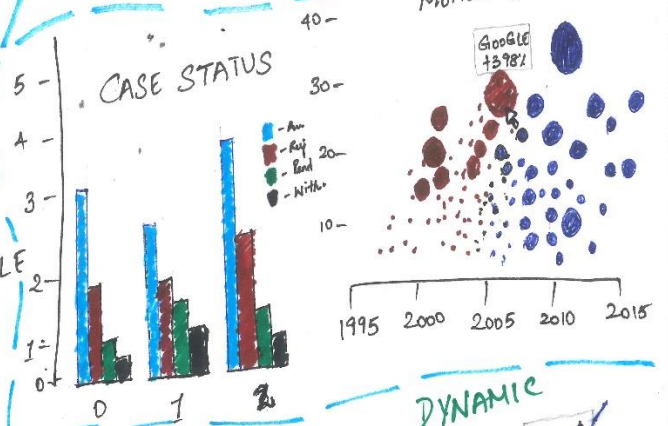
KEY THINGS

* TOP COMPANY SPONSORING
* NO OF APPLICANT PER STATE
* AVERAGE SALARY PER STATE
* AVERAGE SALARY PER EMP TYPE
* TIME TAKEN | BASED ON EMPLOYMENT TYPE & PAY

FEATURES TO VIZUALIZE

* COMPANY
* PER STATE
* JOB TYPE/TITLE
* SALARY
* CASE STATUS
* DATE

VIZUALIZATION'S IDEA'S



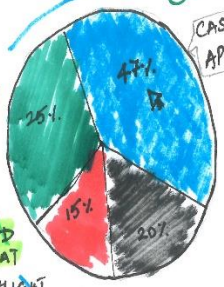
FILTER CONDITIONS

* FILTER BASED ON
  - JOB TITLE
  - NEW EMPLOYMENT
  - OLD EMPLOYMENT
  - FULL TIME EMP
  - DEPENDANT
  - COMPANY
  - STATE
* PIE CHART IS BETTER TO SHOW THE RATIO THAN
* SIDE BY SIDE BAR CHART IS BETTER WHEN COMPARED TO STACKED BAR CHART

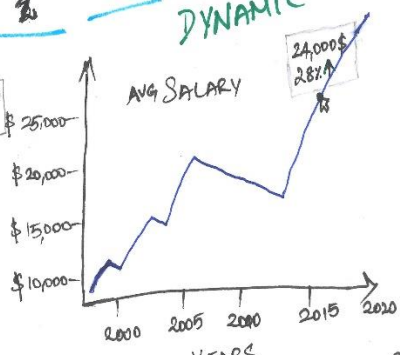
CATEGORIZE

* PUT TOGETHER THE BEST IDEAS. STATIC VS DYNAMIC
* CASE-STATUS - PIE, SBS BAR
* COMPANY SALARY - MOTION → LINE-STAT
* LOCATION - STATIC MAP - HIGHLIGHT

STATIC



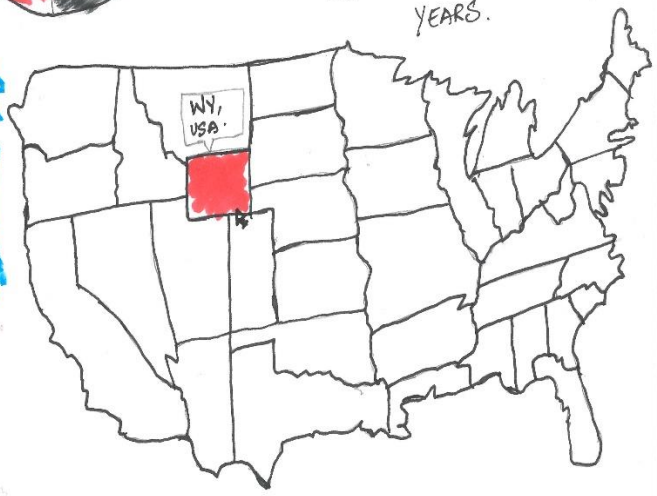
DYNAMIC

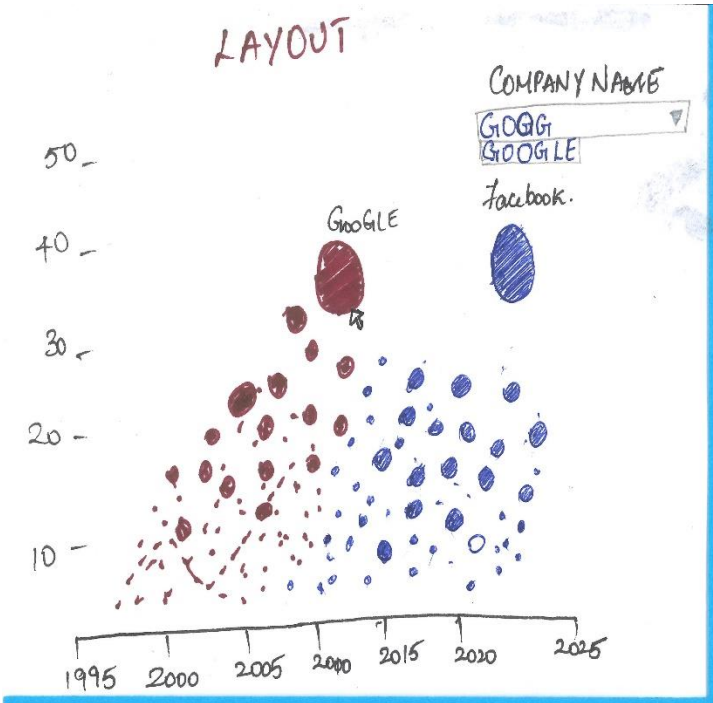


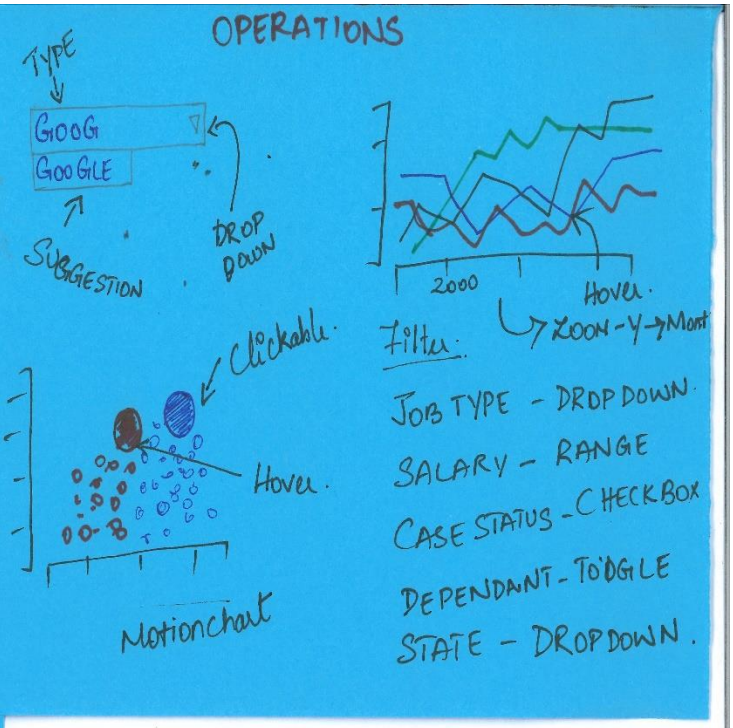
COMBINE AND REFINE

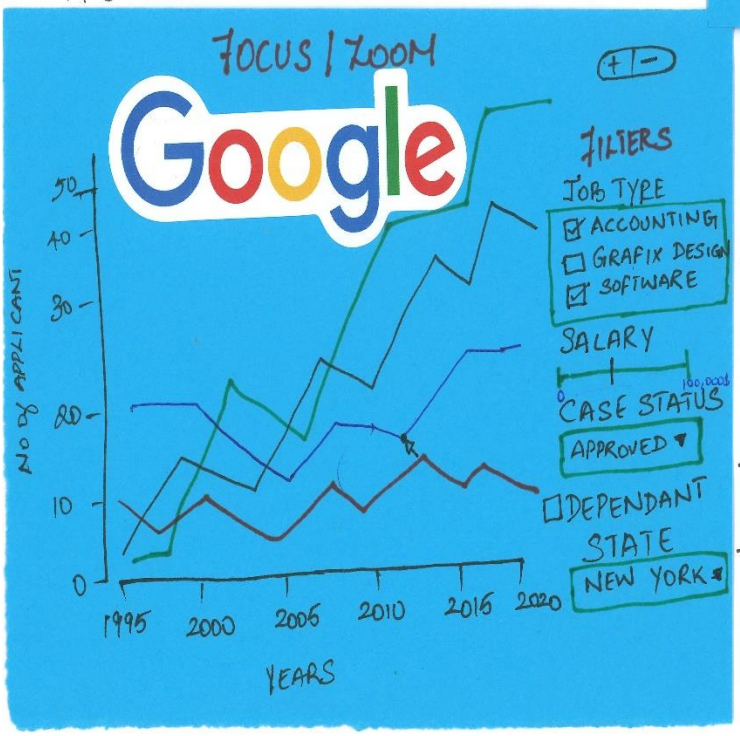
* MINI IDEAS BIGGER SOLUTION
* SHOW CO-RELATION
* SHOW TWO-GRAPHS FOR MULTIPLE VIEWS THAT DEMONSTRATE DIFFERENT ASPECT OF SAME INFORMATION

Questions?









DISCUSSIONS

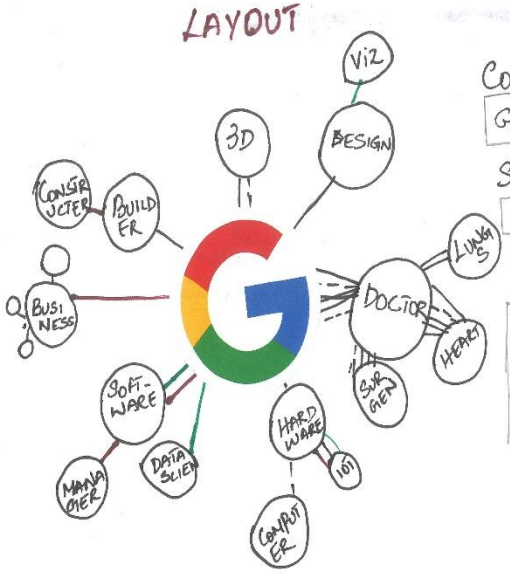
POSITIVES

- ABLE TO SEE THE TOP PAID COMPANY EASILY
- DETAIL FILTERABLE OPTIONS TO PREDICT THE US' VISA RATE

NEGATIVES

- DIFFICULT TO SEE THE SMALL COMPANY
- KEEP TRACK OF MOTION

LAYOUT



COMPANY NAME

GOOGLE

STATE

CALIFORNIA

LENGTH

TIME TAKEN FOR DESCISION

LENGTH OF EM

OPERATIONS

COMPANY NAME - DROP DOWN & AUTOFILL

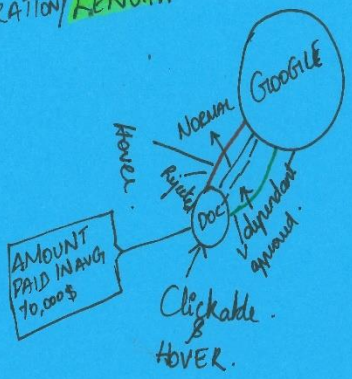
AND STATE.

DURATION/LENGTH

Circle Size - Salary

Length of the line. - Time Taken or Length of Employ ment.

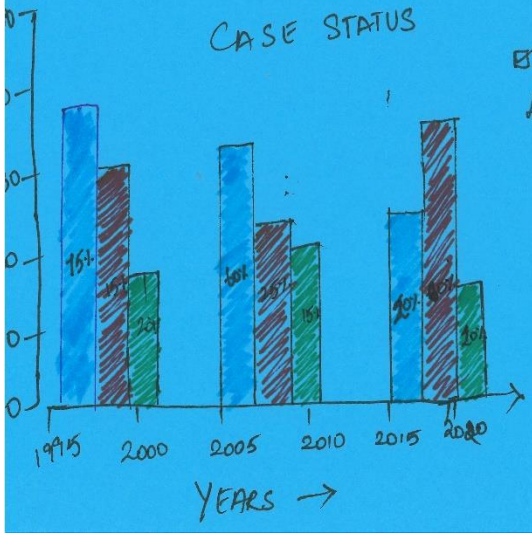
FORCE GRAPH.



DEPENDANT

* CHECK BOX FILTER.
* ZOOM - YEARS → MONTHS
* HOVER - BAR DETAILS

FOCUS/ ZOOM.



DEPENDANT LENGTH.

DISCUSSIONS

POSITIVES

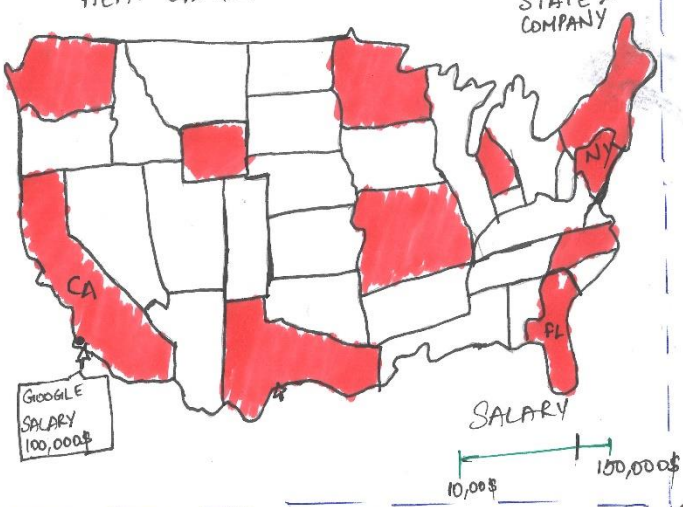
- VISUALY ABLE TO FIND TOP PAID EASILY
- ABLE TO IDENTIFY ALL MAJOR KEY DETAILS

NEGATIVES

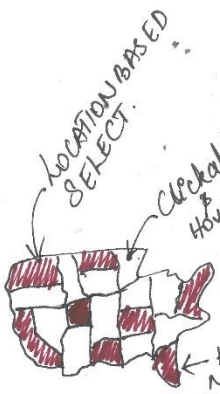
- NEED TO HOVER TO SEE THE EXACT SALARY
- DIFFICULT TO SEE ALL THE CONNECTIONS LINE AT A TIME

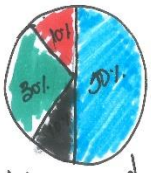
LAYOUT

HEAT - CHOROPLETH MAP BASED ON SALARY & STATE & COMPANY



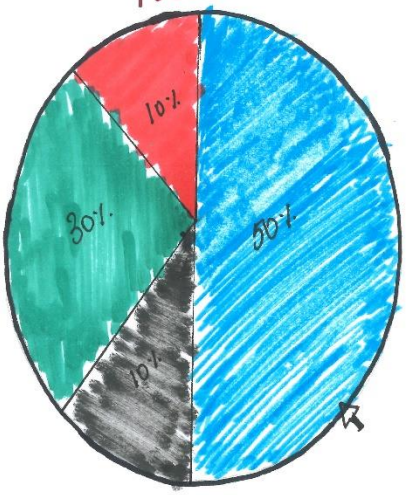
OPERATIONS





Filter Based on.
JobType - Dropdown.
Dependant - Toggle.
New/old - Toggle.
EMP

FOCUS / ZOOM.



- APPROVED.
- WITHDRAW
- PENDING
- REJECTED.

JOB TYPE

SOFTWARE

☑ DEPENDANT

☐ NEW/OLD

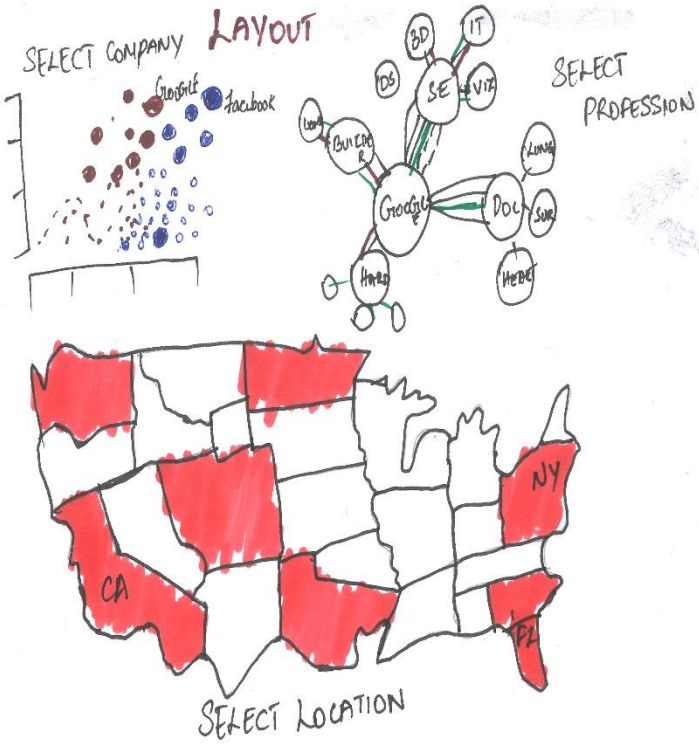
DISCUSSIONS

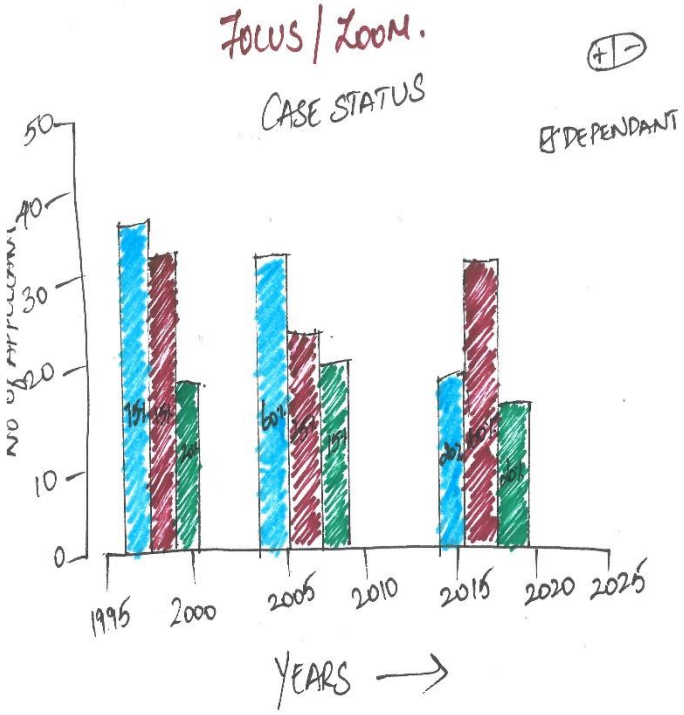
POSITIVE's

- HEAT MAP HELPS TO FIND THE LOCATION FASTER.
- PIE CHART GIVES A ACCURATE SLICE OF RATIO.

NEGATIVES

- DIFFICULT TO LOCATE THE COMPANY NAME IN MAP.
- HEAT MAP IS DIFFICULT FOR UNDERSTANDING.





OPERATIONS

* SELECT THE COMPANY
* THEN SELECT THE LOCATION OF THE COMPANY
* FINALLY, THE PROFESSION.
* The ZOOM/FOCUS AREA IS DISPLAYED
* The CASE STATUS CAN BE ACCESED EASILY
* FASTER AND BETTER OPERATIONS.
* PROVIDE OVERVIEW AT THE BEGINING.

DETAILS

ALGORITHMS - R, Script TO CATOGORIZE JOB TYPE

DEPENDENCIES - CASE STATUS PREDICTION TRANING AND TEST DATA.

TIME TO BUILD - 2 WEEKS

MATERIALS & QUANTITY REQUIREMENT - D3.js Site & No OF GRAPHS REQUIRED.

Difficulty:

I have draw the graph using topojson and manually put the map using diffenent shapes.

Map consist of heat map and location points – 2 different things in same one

Radial Tree graph data generation. – the most difficult part. Generating a 3 generations of tree

Grand Parent – parent- child and then generating tree.

Sampling the data. Since the data is huge I had to sample the entire dataset based on different sampling onditions using python and give weightage to the columns required (only for map).

Convert Java script to json for heatmap.

Calculations required for changing the size of a radial tree node and the size of the line should be based on the salary size was very difficult.

For every mistake I made on the code. Fixing also some what easy. wouldn't say easy but "for running the entire code with huge data" way way tooooooooo time consuming and took lot of my patience.

And many more internal parent child data generations for the second graph (pie, line and bar) generations was way too difficult. Because java script is call by refence and not call by value.

Hence, I learnt a lot and hope, I have done a lot so. HOPEFULLY.

Note:

I had two continues exam on Monday and Tuesday. Kindly requesting you to be lenient on marks allocation since I have put lot of efforts due to my time table and had many sleepless nights. ;(