



MONASH  
University

FIT5147

# Data visualisation and exploration

ASSIGNMENT -2  
RACE FOR EMPLOYEE SATISFACTION  
VISUALISATION TOOL

Reyash Kadyan  
29895405

10<sup>th</sup> June, 2019

Tutor  
Jie (Lewis) Liu

## Table of Contents

<i>INTRODUCTION</i> .....	3
Intended Message .....	3
Audience .....	3
<i>DESIGN</i> .....	4
SHEET – 1: IDEAS.....	4
SHEET – 2: RATINGS BY COUNTRY.....	5
SHEET – 3: FREQUENT WORDS OF COMPANIES .....	5
SHEET – 4: RATINGS WITH ANNUAL PROFITS .....	6
SHEET – 5: FINAL DESIGN .....	7
<i>IMPLEMENTATION</i> .....	7
UNDERSTANDING THE PLOTS .....	8
1. RATINGS BY COUNTRIES IN CHOROPLETH MAP .....	8
2. RATINGS BY YEAR IN SANKEY DIAGRAM .....	8
3. TOP 5 WORDS ASSOCIATED WITH FIRMS IN RADIAL BAR CHART.....	8
4. 10 FREQUENT WORDS IN DONUT CHART.....	9
5. RATINGS WITH ANNUAL PROFITS IN MOTION BUBBLE CHART .....	9
6. ANNUAL PROFITS OF FIRMS IN LINE CHART .....	9
<i>USER GUIDE</i> .....	9
1. Introduction.....	9
2. Ratings.....	10
3. Words .....	11
4. Profits .....	12
5. Summary.....	13
<i>CONCLUSION</i> .....	14
<i>REFERENCES</i> .....	14

# INTRODUCTION

In today's corporate world, every employee wants different things out of their professional career, be it opportunities of growth, cultural belonging, or financial benefits, and with multitude of companies in business, choosing the firms suiting one's requirements poses a challenge to all. According to a survey, more than 50% employees in Australia are dissatisfied with their jobs. [1]

## Intended Message

Major proportion of the freshly graduated students from all field struggles to find a company, best suiting their needs. In this project we will look at the making and operation of a **visualization tool to choose the best firm to suit the values and requirements of any prospective employee**, amongst the leader firms (in terms of stock prices) of the market, which are Apple, Amazon, Microsoft, Google, Netflix and Facebook. We will build this tool by using the only thing through which one can firm learn to make wise choices i.e., experience! We will use the data about the reviews provided by both, current and previous employees of the aforementioned firms, on 'Glassdoor' from all over the world, combined with the Annual gross profit data, used as an indicator for financial performance of these firms.

## Audience

This tool will aid all the **students, working professionals or any other prospective employee**, to find a firm that shares the same value as themselves, and a working environment where they feel they can thrive.

# DESIGN

Five design sheet, methodology is selected to conceptualise the ideas and build a basic layout of my tool. Following are the sheets made during the brainstorming process:

## SHEET – 1: IDEAS

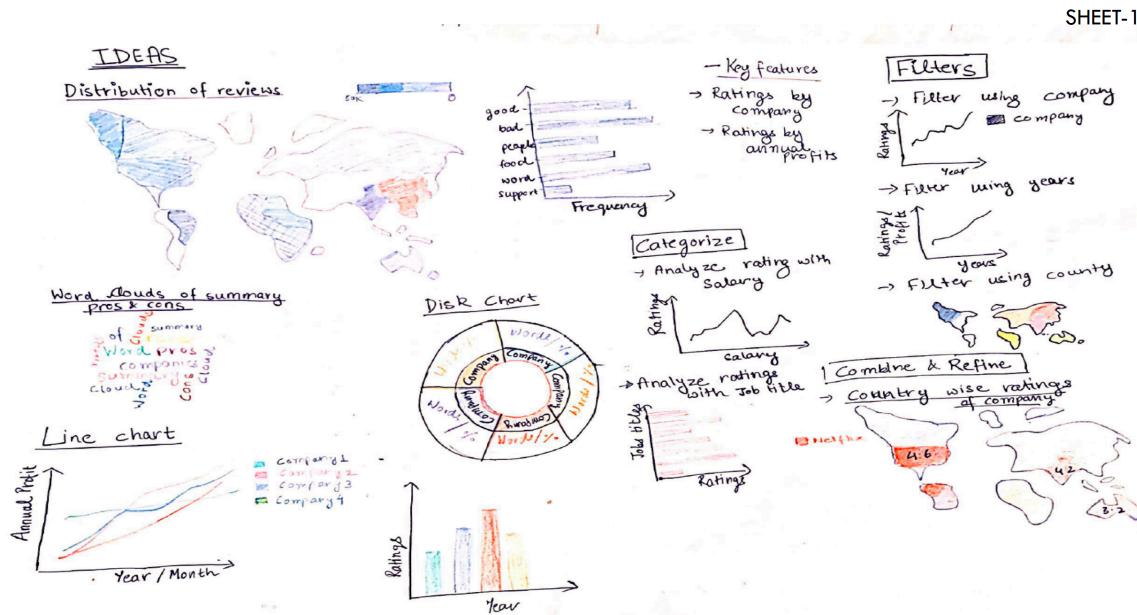


Fig 1. Ideas Sheet

Many ideas were selected initially to visualise many key findings from the combined data. Following are the few explored above :

- Choropleth map to visualise distribution of reviews
- Bar charts to visualise ratings of companies
- Word clouds of text in reviews
- Donut chart for most frequent words in reviews of firms
- Line and bar chart to visualise changes in ratings with changes in annual profits of firm, and to find any correlation among both of them.
- Bar chart to observe the frequencies of most common words

Reviews can be filtered on, company, year, and country values.

I chose to combine, choropleth map and ratings, to avoid basic bar chart and make the map more interactive. I also chose to combine donut chart of companies and its frequently associated words and frequency chart of common words.

Some key features decided on initially were, 'Ratings by Country' and 'Ratings with Annual Profits'.

## SHEET – 2: RATINGS BY COUNTRY

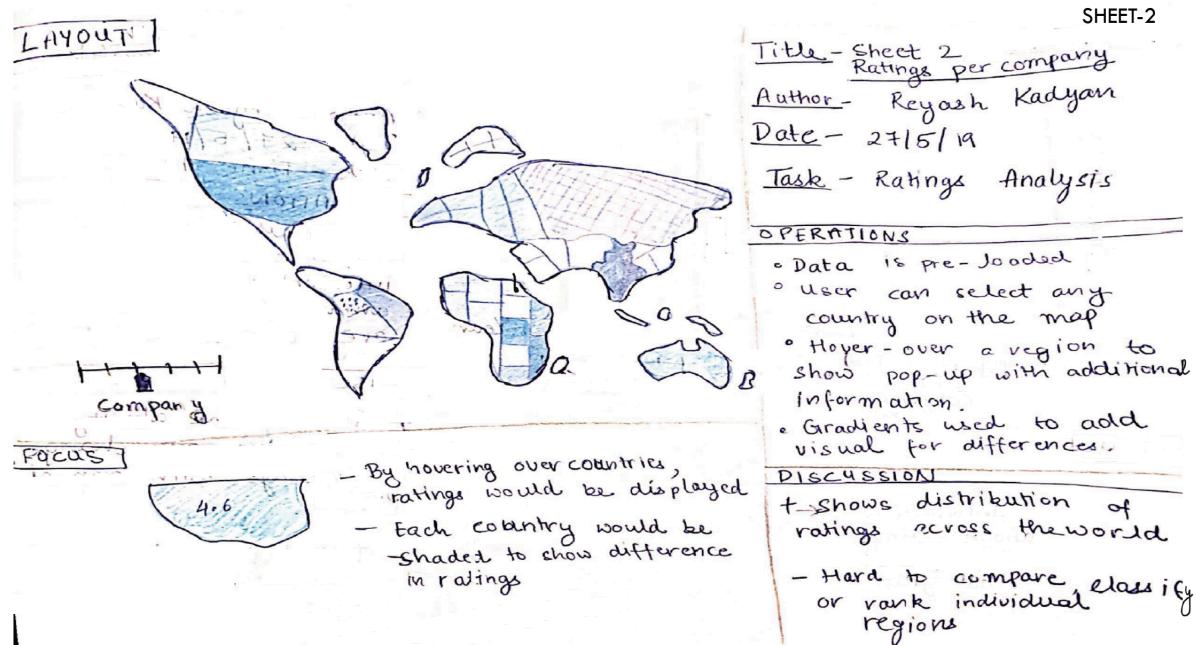


Fig 3. Sheet – 2: Choropleth Map

This map will enable user to visualise average ratings for selected company for all country where the reviews came from. It will allow user to locate the best country to work in for any particular company.

## SHEET – 3: FREQUENT WORDS OF COMPANIES

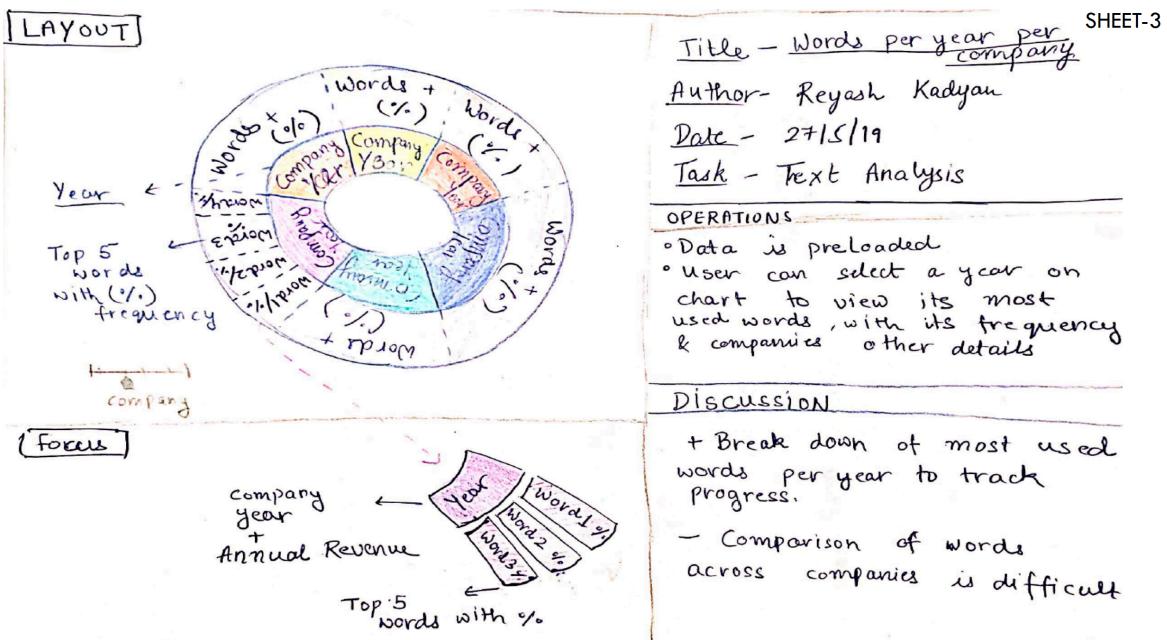


Fig 3. Sheet – 3: Donut Chart

This plot will enable user to visualise most frequent words associated with any firm, both positive and negative, and will allow them to select a company which shares the same values as them.

#### SHEET – 4: RATINGS WITH ANNUAL PROFITS

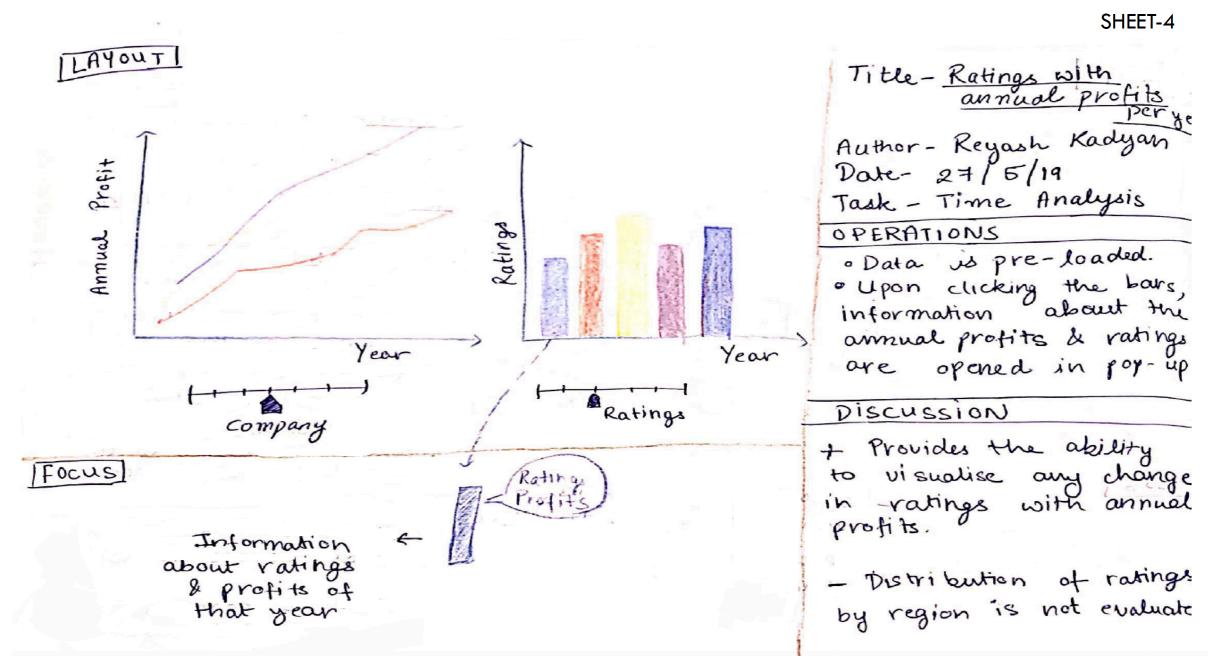


Fig 4. Sheet – 4: Bar chart and Line chart

These plots will enable user to visualise the changes in ratings with annual profits, to observe any correlation between firm's profits and employee satisfaction.

## SHEET – 5: FINAL DESIGN

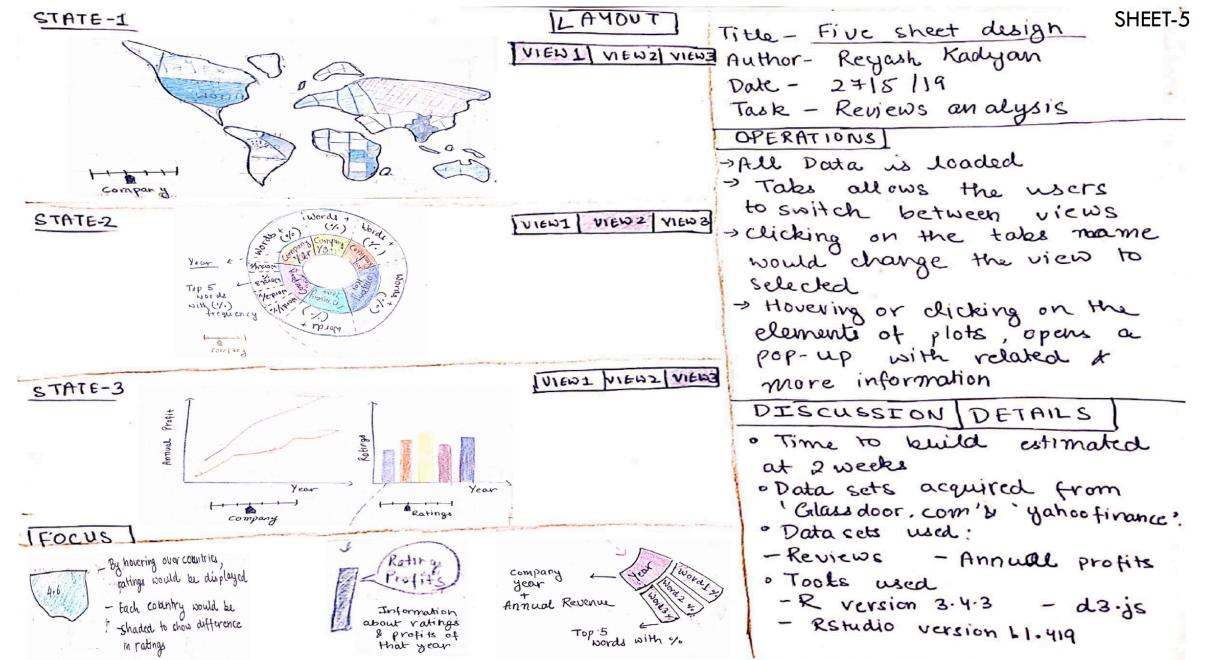


Fig 3. Sheet – 5: Five design sheet

R-Shiny is chosen to make an interactive app with three tabs, which are,  
 TAB – 1: To visualise ratings of companies for countries  
 TAB – 2: To visualise most common words mentioned in company's reviews  
 TAB – 3: To visualise annual profits of a company with different type of ratings provided by employees.

## IMPLEMENTATION

During the implementation of the 'Employee reviews analysis' web based application in R shiny, following libraries are used:

1. **shiny**: Library used to build a visually appealing and interactive web-based application with use of R programming.
2. **shinythemes**: Library used to add themes to the application (I choose Sandstone).
3. **ggplot2**: Library used to create plots and visualizations by adding layers of graphics one on another. It is used make radial bar chart.
4. **plotly**: Library used to make interactive plots. It allows us to interact with graphs and visualize more information on the entities in graphs, with the use of actions like hovering, zoom, filter, tooltips and etc. It is used to make motion chart and donut chart.
5. **dplyr**: Library used for data wrangling and manipulation. I have used this library for 'group\_by', 'select', 'filter', 'mutate' and 'summarize' functions.

6. **tidy**: It is used write tidy code and cleaning data. I used '%>%' pipe operator provided in the library.
7. **leaflet**: It is a widely used open-source JavaScript library used to make user-friendly interactive maps.
8. **rgdal**: It is a 'Geospatial' Data Abstraction Library ('GDAL'). It is used to add polygons on the leaflet maps, according to the boundaries of the countries.
9. **geojsonio**: It is a library used to convert data from 'GeoJSON' to 'TopoJSON'. I used it to convert Geospatial data in JSON format into format understandable by leaflet.
10. **googleVis**: Library used to build interactive and beautiful visualisations, to better understand the data. I used it to build a 'sankey' diagram.
11. **rlang**: Library used to work with R conditional panels. I used it to convert strings from UI in shiny app to be used as textual part of code in server.
12. **tidyR**: It is used to write neat and understandable code, and also to wrangle data.
13. **tm**: Library used for text mining and processing. It is used to convert the text from reviews into matrix, removing stop words, and perform other manipulations on textual corpus.

## UNDERSTANDING THE PLOTS

### 1. RATINGS BY COUNTRIES IN CHOROPLETH MAP

As initially planned, to depict the overall ratings of companies by country on a map. 'Leaflet' map is used to make an interactive choropleth map. I used the Geospatial data about the boundaries of countries to add polygons by 'addPloygons()' functions in leaflet. Average ratings for every company is computed and for all countries, from where the reviews has been provided. These average ratings are then added to the geospatial data about countries and are used to add labels on the popups for each polygon(representing country on map) added.

### 2. RATINGS BY YEAR IN SANKEY DIAGRAM

This plot was included at a later stage, to satisfy the need of visualising changes in different types of ratings over the years. It enables user to visualise the changes in ratings and speculate the priorities of company and changes in those priorities over the years. Reviews are grouped by company and the year of the review, and average ratings are computed for each group, for the rating type selected by user. First layer of nodes represents the companies, the second layer represents the year, and the links width is based on the average of ratings provided. By hovering over each link average rating values is displayed. 'googleVis' has been used to plot it. Reviews are grouped by company and year(using 'group\_by' function), and average values of ratings(using 'summarise' function) are computed for each group. These values are used by 'gvisSankey' function to render the sankey diagram.

### 3. TOP 5 WORDS ASSOCIATED WITH FIRMS IN RADIAL BAR CHART

This is a combined version of frequency bar chart of common words and donut chart of words associated with each firm for a particular year selected by user. It is selected over two-layered donut chart with companies (in inner layer), and words associated with them (in outer layer),

since it is more comprehensive and informative than donut chart, as user can also visualise the use of word with the length of the bar. ‘ggplot2’ has been used to plot it. Reviews are processed and converting into count matrix using ‘tm’ library. Frequencies of each word, after removal of stop words, are computed and word with maximum occurrences are displayed.

#### 4. 10 FREQUENT WORDS IN DONUT CHART

It is an extension of the radial bar chart above, it shows top 10 words associated with any firm, for a particular year. It is an interactive donut chart, which displays the percentage of occurrence of that word in the reviews (summary, pros or cons), when hovered upon. ‘Plotly’ is used to make it, with all text pre-processing handled by ‘tm’ library same as above.

#### 5. RATINGS WITH ANNUAL PROFITS IN MOTION BUBBLE CHART

Bubble chart is selected, instead of basic bar chart selected initially, to visualise ratings, since it is more engaging and appealing. It allows user to inspect the changings in overall ratings of a company, with changes in annual profits over years. Radius of bubbles on the graph, varies proportionately with annual profits of firm. An animation is used to observe the changes in aforementioned attributes together, over the years. It helps user to identify relationship between employees’ satisfaction and annual profits of firms, and figure out whether company invest in their employees when profits increases. This chart is made using ‘Plotly’. A tooltip is popped when mouse is hovered over any bubble, displaying the average rating and annual profit for particular year selected by user. Reviews are grouped by company and year(using ‘group\_by’ function), and average values of ratings(using ‘summarise’ function) are computed for each group. These values are used for ‘plot\_ly’ function to render the bubble chart.

#### 6. ANNUAL PROFITS OF FIRMS IN LINE CHART

Interactive line chart is used to visualise annual profits of selected firms over the years. This chart along with bubble chart above, helps user to infer the value each company place on their employees, by observing whether a company chooses to share its profits with its employees’. This chart is also made using ‘Plotly’.

## USER GUIDE

The application can be opened in browser of your choice or R shiny window after clicking the button Run application in the R script.

### 1. Introduction

This page is static and provides an overview of the visualization tool, with the brief introductions to plots and expectations from the tool.

## Overview

In today's corporate world, every employee wants different things out of their professional career, be it opportunities of growth, cultural belonging, or financial benefits, and with multitude of companies in business, choosing the firms suiting one's requirements poses a challenge to all. According to a survey, more than 50% employees in Australia are dissatisfied with their jobs.

Major proportion of the freshly graduated students from all field struggles to find a company, best suiting their needs. This visualization will enable you to choose the best firm to suit your requirements, amongst the leader firms (in terms of stock prices) of the market, which are Apple, Amazon, Microsoft, Google, Netflix and Facebook.

### Ratings

This tab consists of a choropleth map which will enable you to visualise average ratings for selected company for all country where the reviews came from. It will allow you to locate the best country to work in for any particular company.

### Words

This tab consists of stacked bar chart will enable you to visualise most frequent words associated with any firm, both positive and negative, and will allow you to select a company which shares the same values as them.

### Profits

This tab consists of motion chart, which will enable you to visualise the changes in ratings with annual profits, and to observe any correlation between firm's profits and employee satisfaction.

Fig 6. Introduction page in visualisation tool

## 2. Ratings

This navigation panel shows you Overall ratings of the companies. It will display two graphs, which are:

- a) Choropleth map, depicting '**Overall Ratings per Country**'. User can select the company and the for the year ,for which they want to see the ratings of, through a drop down and slider, respectively. Tap on any country on the map, to see a tooltip displaying country's average overall ratings value on the map.
- b) Sankey diagram, depicting '**Ratings of Companies over the years**'. User has an option to select the type of rating they want to look, among 'Work balance stars', 'Career opportunities stars', 'Senior Management stars' and 'Company Benefit stars'. This plot needs a running internet connection. Hover over any link to see the average value of overall ratings for that link. Hovering over any node, will highlight all the connected links.

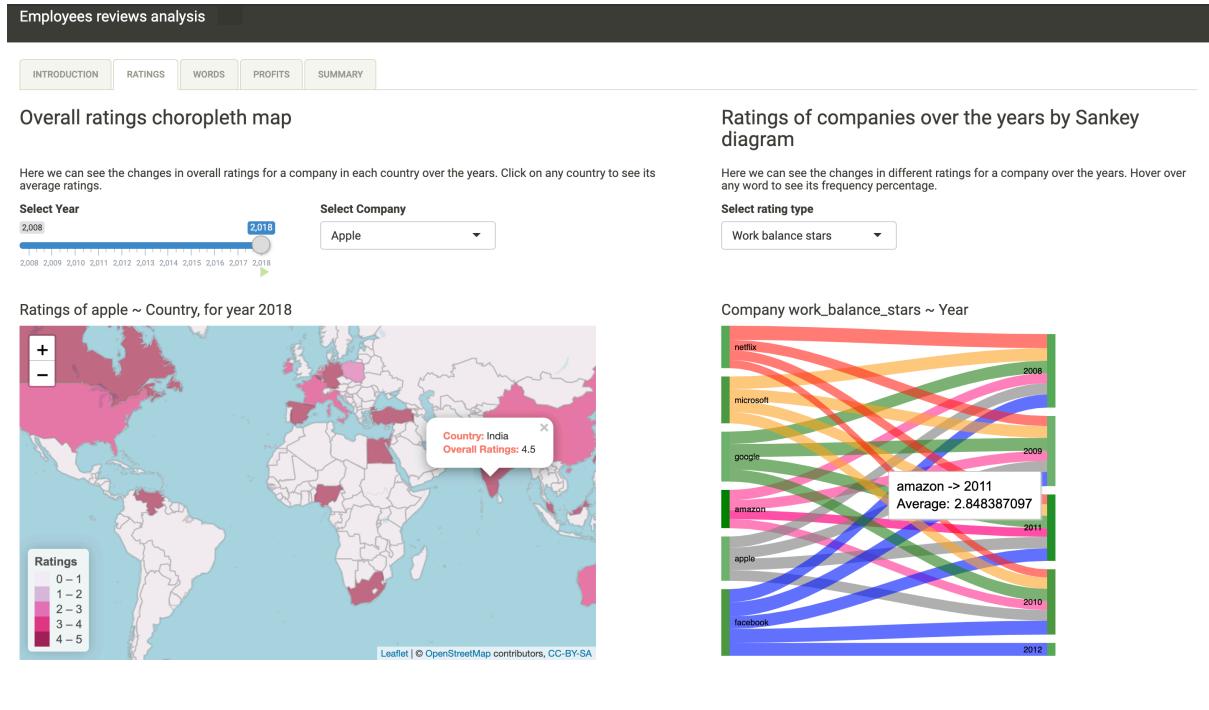


Fig 7. Ratings by country and year in visualisation tool

NOTE: This page takes a little time to compute, because of the large dataset (67k rows originally). Please be patient, waiting is worth the while! In case, you see the stuck map, select some/same option again. This will reload the computation, and clear the wrongly cached data.

### 3. Words

This navigation panel shows you Overall ratings of the companies. It will display two graphs, which are:

- Radial Bar chart, depicting '**Top 5 words frequently used with all companies**'. User can select the type of review to use from 'Pros', 'Cons' and 'Summary' provided in reviews, to find most frequent words. User can also select the year for which they want to see the top words. Length of a bar word is proportional to the frequency percentage of that word, and is also displayed on each bar.
- Donut chart, depicting '**Top 10 words associated with a company**'. User can select the company, for which they want to see the top 10 words, for the year selected in the radial bar chart. Hovering over each word opens a tooltip with the frequency percentage of that word.

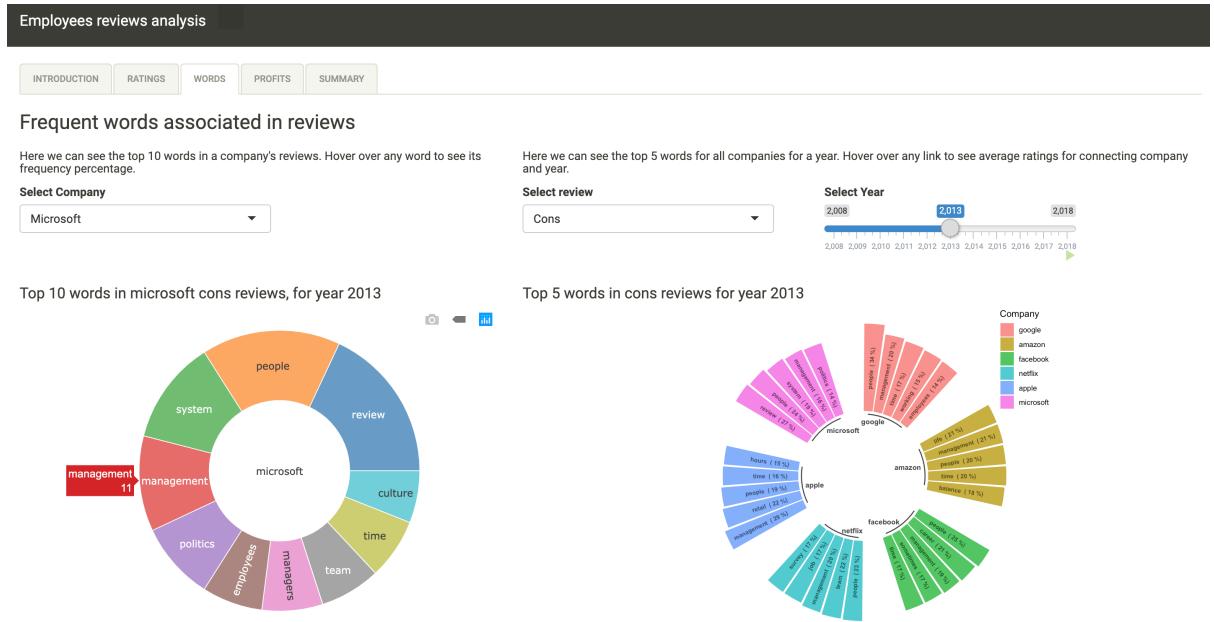


Fig 7. Frequent words associated with company(s) and year in visualisation tool

#### 4. Profits

This navigation panel shows you Impact of profits on the ratings of the companies. It will display two graphs, which are:

- Motion Bubble chart, depicting the changes of '**Ratings and Profits**' over the years. User can filter on the company, for which they want to visualise the changes. Hovering over any Bubble displays a tooltip on screen, displaying the values of average ratings and annual profits of that company for and for that particular year.
- Line chart, depicting the '**Annual profits of companies**' over the years. User can filter on the company, for which they want to visualise the profits. Hovering over line displays a tooltip on screen, displaying the annual profits of that company for and for that particular year.

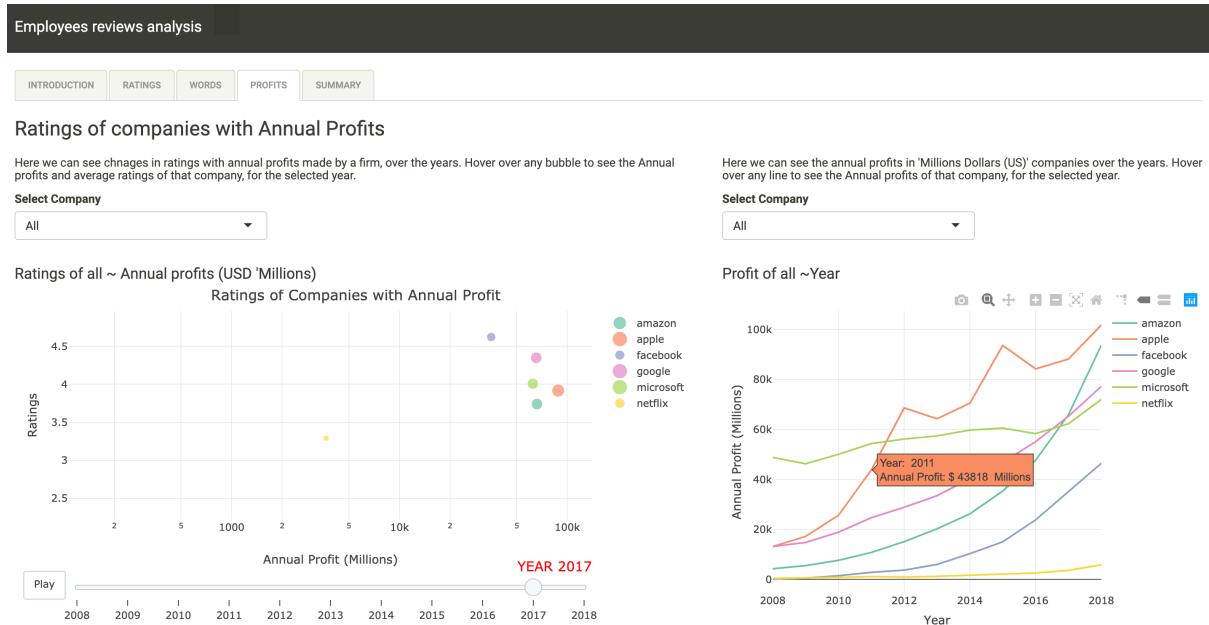


Fig 8. Ratings of company(s) with Annual profits in visualisation tool

## 5. Summary

This tab is static and concludes the observed trends and results from all visualisations.

INTRODUCTION RATINGS WORDS PROFITS SUMMARY

Some inferences...

### Amazon

Amazon seems to be investing moderately towards providing more benefits and opportunities to its employees. Employees seem to be neutral towards senior management. It seems that people in Amazon are not contented with their jobs, as the word 'bad' pops up a lot. However, one seems to 'learn' a lot while working, and are paid good!

### Apple

Apple seems to follow aggressive approach when profit goes down. Bumps in ratings are in sync with profits, and increases and decreases proportionately with annual profits, which indicates strictness of management when profits go down! Apple ratings rise when the profits rise. Employees seem to be happy about their 'pay'. It seems that 'family' like 'culture' is prevalent.

### Facebook

Facebook seem to invest highly towards providing more career opportunities and other benefits to its employees. Work-Balance seems to be decreasing for Facebook. They are losing track of their employees' work load. Cultural values are on a setback, with its eyes on the price!

### Google

Google seem to invest highly towards providing more career opportunities and other benefits to its employees. Work-Balance seems to be constant. Cultural values are on a setback, with its eyes on the price! Google employees seem to be very contented with their jobs as words 'amazing' and 'awesome' pops up a lot in their reviews.

### Microsoft

Microsoft has a fairly constant profits over year, and they don't seem to increase company benefits and work load over employees. This steady flow seems to satisfy the employees, as the ratings increase with the passage of time. Employee seems to be fed up of office 'politics' and 'bureaucracy' of 'management', consistently after year 2008.

### Netflix

Netflix ratings seems to vary a lot, with dips in 2010, 2012, and 2015, in overall increasing graphs. This indicate to unstable working environment, where employees aren't valued enough. Employees seem to be working in the fear of being fired at any moment!

Fig 6. Summary page in visualisation tool

# CONCLUSION

Through this data visualization project, I got the essence of application development, with a power tool of R Shiny, and developed appealing and user-friendly dashboards. The dashboard created using Shiny are intuitive and simple to use. With the use of this developed tool, students and working professionals will be able to benefit, by choosing a firm which shares the same values as themselves. Companies might also benefit from the use of this tool, by monitoring their ratings and investing in the areas where their employees feel lagged.

## Reflection

This assignment touched the various techniques and methods of data manipulation and visual analysis, by following tasks:

- **Data Wrangling & Manipulation:** I got in-depth knowledge of data manipulation in our chosen tool of wrangling. Since different graphs needed inputs in different formats, I repeatedly used '*dplyr*' and other libraries with it, namely '*rlang*' and '*tidyR*', to group, summarize, aggregate and reshape the data.
- **Data Visualization:** I got rigorous knowledge about the techniques of visualization. I learned about various graphs and plots, and the data they are best used to describe. For example, '*Sankey diagram*', '*Bubble chart*', '*Choropleth map*' and etc.
- **Application development:** I learned about various intricacies of User-Interface (UI) development and got to experience the basics of application development, where I thought about the use cases of my application and altered my application for better results.

## Improvements

1. I would have used D3 to produce more indulging and appealing plots and dashboards, if there were no time constraint.
2. Data about more companies could have been collected to broaden the scale of market, and allowing user to choose best from more.

# REFERENCES

- Hate my job. (2019). Retrieved from: [https://www.huffingtonpost.com.au/2016/03/22/hate-my-job\\_n\\_9519688.html](https://www.huffingtonpost.com.au/2016/03/22/hate-my-job_n_9519688.html)
- googleVis examples. (2019). Retrieved from: [https://cran.r-project.org/web/packages/googleVis/vignettes/googleVis\\_examples.html](https://cran.r-project.org/web/packages/googleVis/vignettes/googleVis_examples.html)
- Leaflet for R - Introduction. (2019). Retrieved from: <https://rstudio.github.io/leaflet/>
- plotly. (2019). Retrieved from: <https://plot.ly/r/>
- Kaggle.com. (2019). *Google, Amazon and more Employee Reviews*. Retrieved from: <https://www.kaggle.com/petersunga/google-amazon-facebook-employee-reviews>
- Macrotrends.net. (2019). Retrieved from: <https://www.macrotrends.net/stocks/stock-screener>