# Final Project Indeed

Maggie Sha

12/13/2020

## Introduction

This is a practice of text mining by R.

An interactive shinyapp is also created.

The main question is to find how companies describe the data science jobs when they post these job positions.

Other questions:

```
* Are the descriptive words different from various geographical areas?
* What are the commonly used Job titles for data science positions?
* Are the top descriptive words on different job boards also different?
```

Data comes from Indeed job board and Github Jobs.

See:

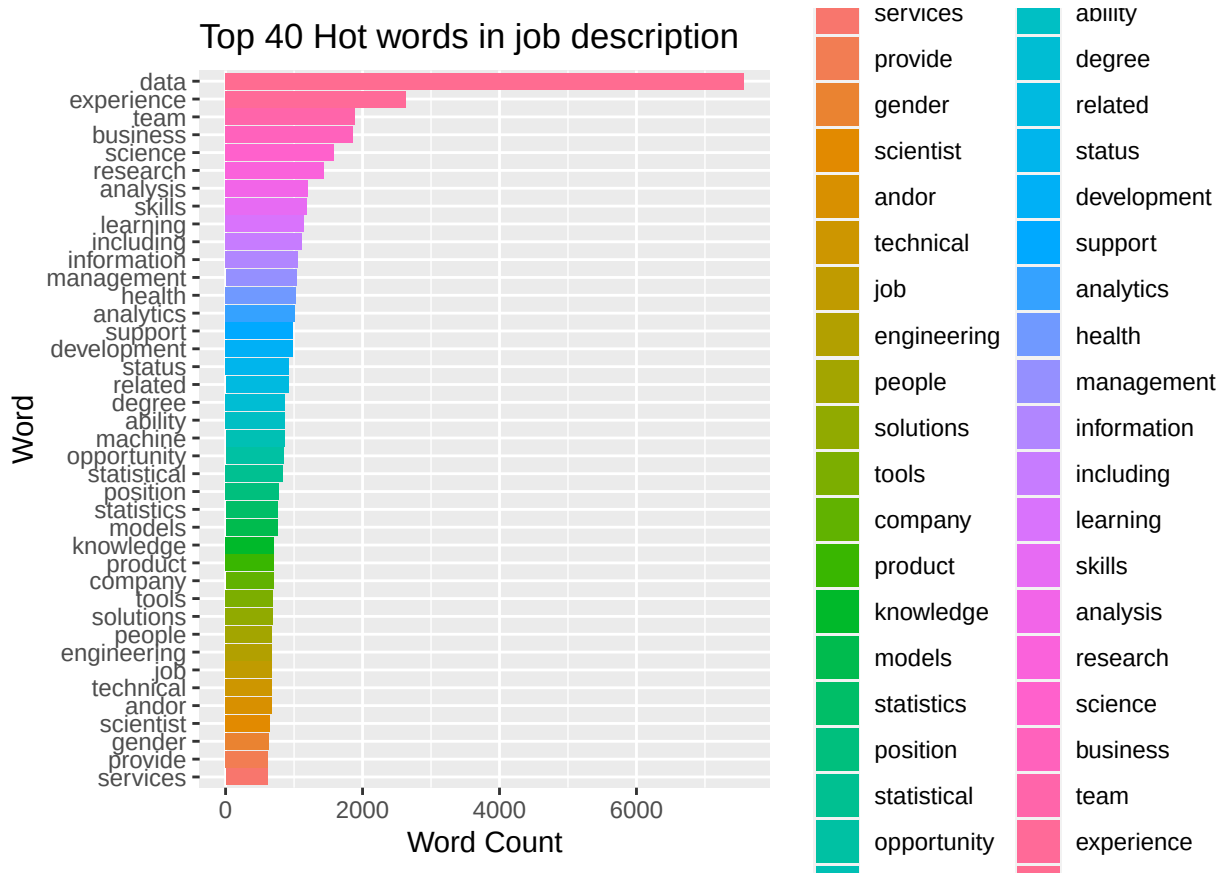
- https://www.indeed.com/hire
- https://jobs.github.com/api

## Indeed Data

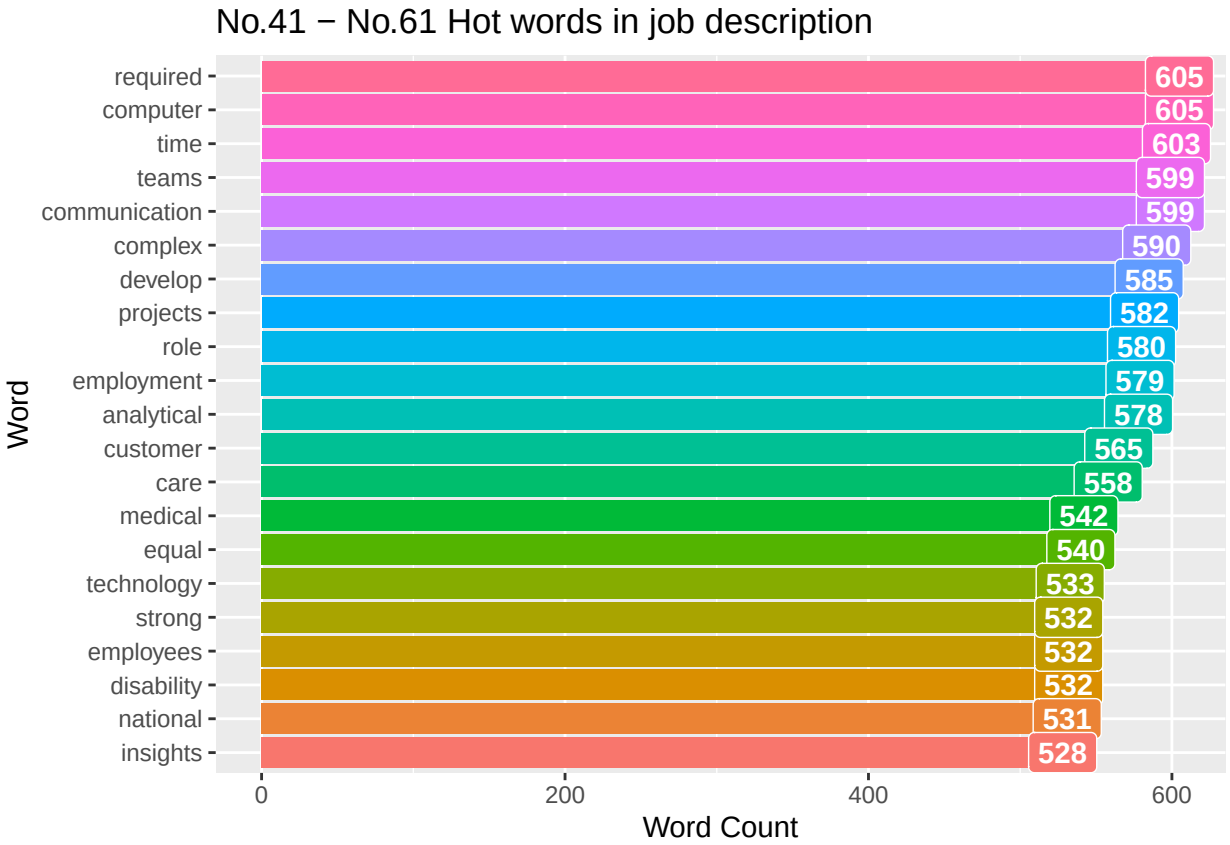
```
##   X                                          title                    company
## 1 1                                Data Scientist        Tradeweb Markets LLC
## 2 2 Junior Data Scientist â\200" Performance Analytics Marina Maher Communications
##       location
## 1 New York, NY
## 2 New York, NY
##                                                                                summary
## 1 3+ years of experience working in data science at a financial, technology or media firm.
## 2                               1-3 of experience in data analytics or data science role.
##
## 1 https://www.indeed.com/pagead/clk?mo=r&ad=-6NYlbfkN0CMUDaL693WNRk3nrnFOjU9lIdT9MBrzyndVEc7SiTvGZio
## 2
##                          uniqueid
## 1 a1bc5c50e2daa4a6e8d0c0636cd51be7
## 2 70823013c23be8c72e4ec2b1a24e474f
##
## 1
## 2 About MMCMMC is not your average PR agency We are a datadriven nextgen communications company that
```

## Plot top description words

```
## Joining, by = "word"
```




```
## Joining, by = "word"
```

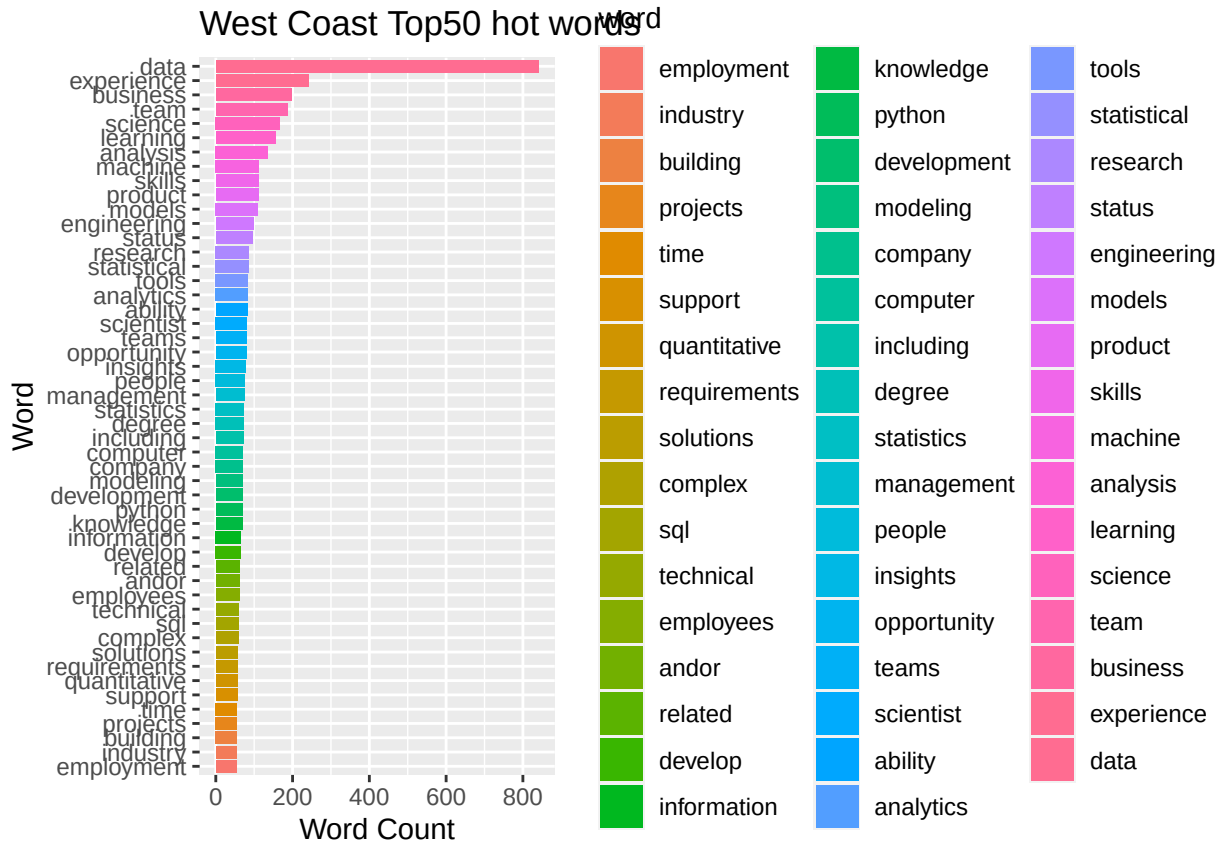



There are many words we can include in our resume or to describe our skills, such as machine learning, python, communication...

## Comparison of top words in the west and the east

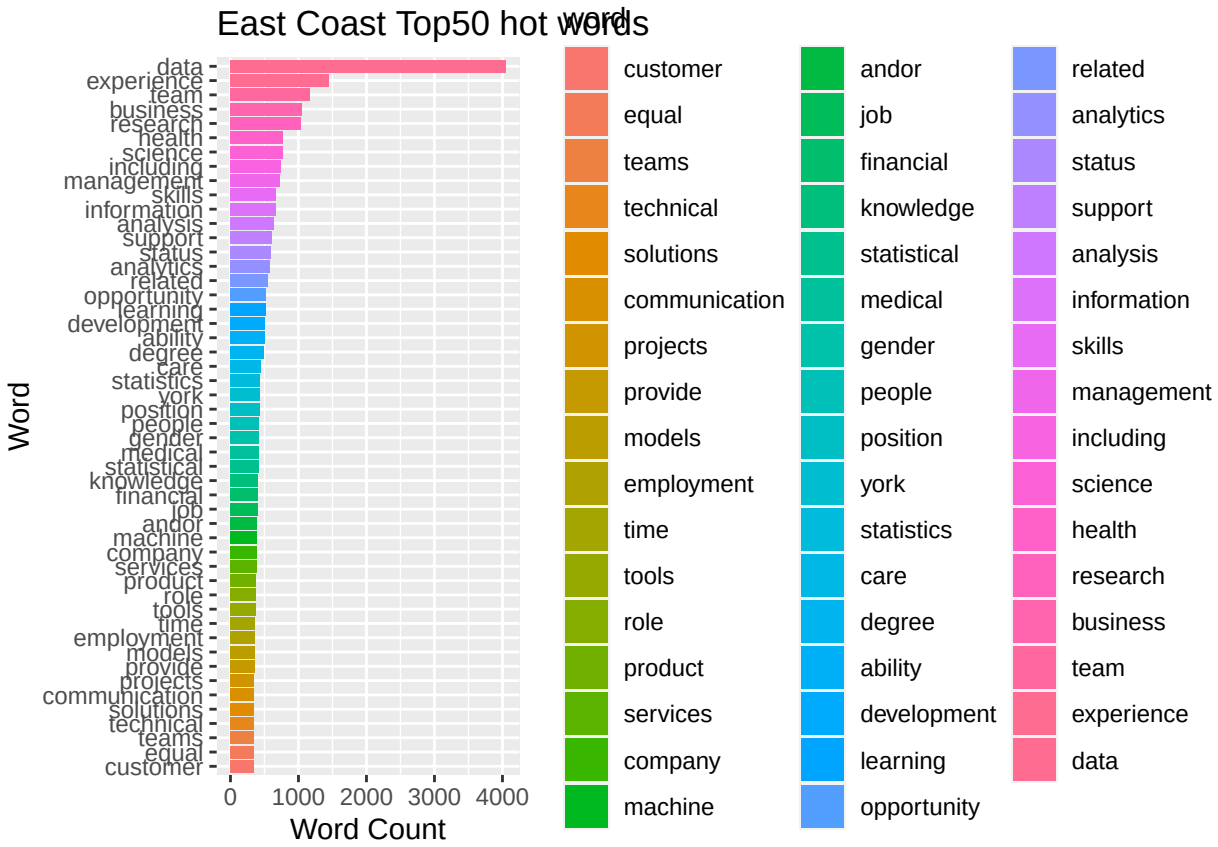
```
f(west_city)[1:50,] %>%
  ggplot(aes(x = n, y = word, fill=word, label=n)) +
  geom_col() +
  labs(title = "West Coast Top50 hot words", x = "Word Count", y = "Word")
```

```
## Joining, by = "word"
```




```
f(east_city)[1:50,] %>%
  ggplot(aes(x = n, y = word, fill=word, label=n)) +
  geom_col() +
  labs(title = "East Coast Top50 hot words", x = "Word Count", y = "Word")
```

```
## Joining, by = "word"
```




The leading words are very similar in the west and the east. However, I found some interesting difference, such as 'financial' is a trending word in th east while "engineering" is hot in the west. In the west it shows "SQL" is also a hot word, so we can consider describe our SQL skills.

## Common used Job titles

```
## Warning in wordcloud(words = title_count$title, freq = title_count$n, min.freq
## = 2, : Scientist II - Biopharma Ingredients - Tarrytown, NY could not be fit on
## page. It will not be plotted.
```


For the data scientist position, many companies choose to call it Data Scientist directly. Many companies also call it Data Analyst, some other titles are Data Engineer, Budget Analyst...

## Comparison of tops words on Github Jobs and Indeed

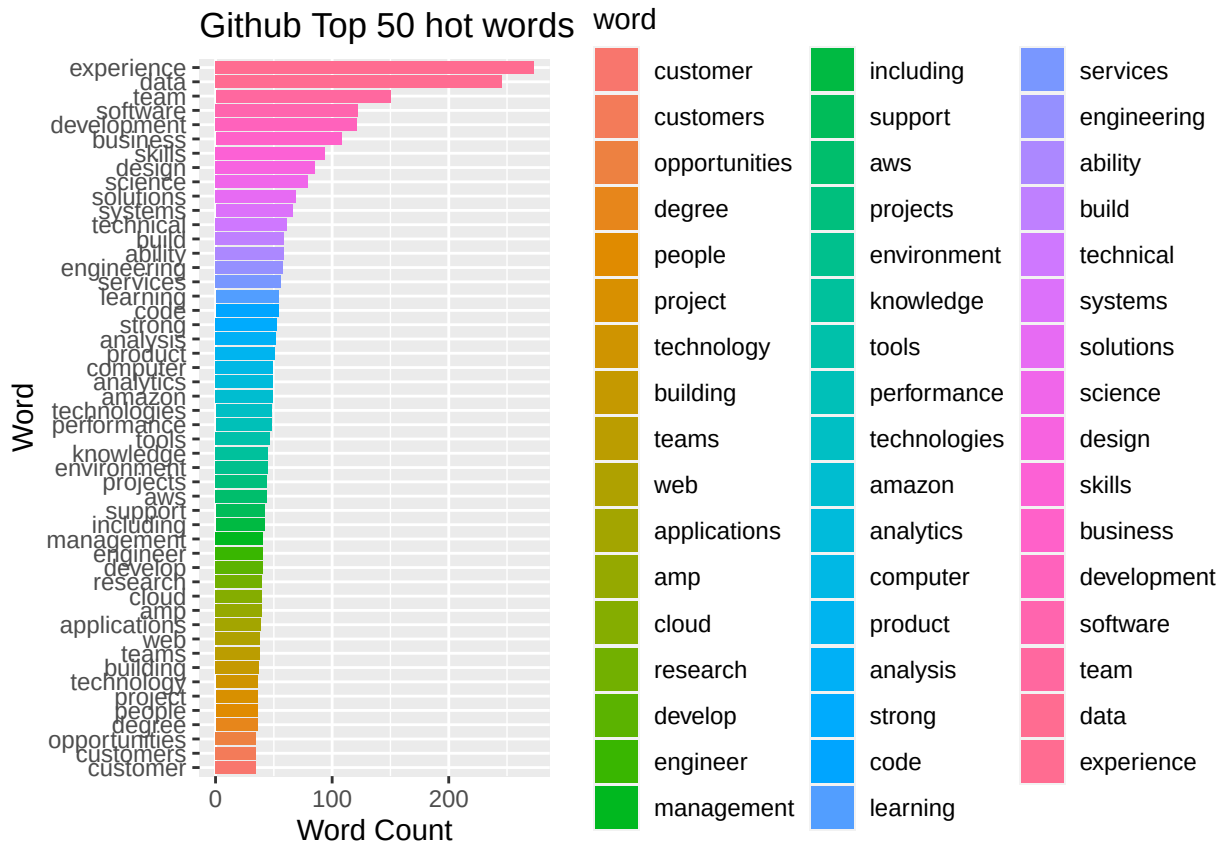
### Github Data

```
##                                     id      type
## 1 efe2509d-a7c0-4172-8e9a-f38fea2ff9ad Full Time
## 2 619735aa-7ba6-473b-bc9e-8e5b84d7a1b5  Contract
##                                                                  url
## 1 https://jobs.github.com/positions/efe2509d-a7c0-4172-8e9a-f38fea2ff9ad
## 2 https://jobs.github.com/positions/619735aa-7ba6-473b-bc9e-8e5b84d7a1b5
##                     created_at           company               company_url
## 1 Mon Dec 14 20:54:25 UTC 2020 Spear Technologies http://www.spear-tech.com
## 2 Mon Dec 14 16:01:42 UTC 2020          Fullstack   https://gofullstack.com
##   location                                    title
## 1   remote         Experienced Power Apps developer
## 2   Remote Fullstack Mobile Engineer (React Native)
##
## 1
## 2 Fullstack is looking for an experienced, passionate, and motivated React Native engineer who will
##
## 1 <p>Email your resume to Kate Taylor at <a href="mailto:ktaylor@spear-tech.com">ktaylor@spear-tech.
## 2                    <p>Email your resume to <a href="mailto:jobs@gofullstack.com">jobs@gofullstack.
```

```
##
## 1
## 2 https://jobs.github.com/rails/active_storage/blobs/eyJfcmFpbHMiOnsibWVzc2FnZSI6IkJBaHBBcjJTIiwiZXh
```

```
f(Github)[1:50,] %>%
  ggplot(aes(x = n, y = word, fill=word, label=n)) +
  geom_col() +
  labs(title = "Github Top 50 hot words", x = "Word Count", y = "Word")
```

```
## Joining, by = "word"
```




By comparison with Indeed, I found the job description on Github Jobs emphasize more on the software by the words such as “software” and “code”. Also, it may contains many jobs in Amazon since those words “amazon” and “aws” are mentioned many times.