Practicum

Nam Hee Kim

02-07-2020

Consumer Complaints

```
library(data.table)
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------
```

```
## v tibble  2.1.3     v dplyr   0.8.4
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
## v purrr   0.3.3
```

```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

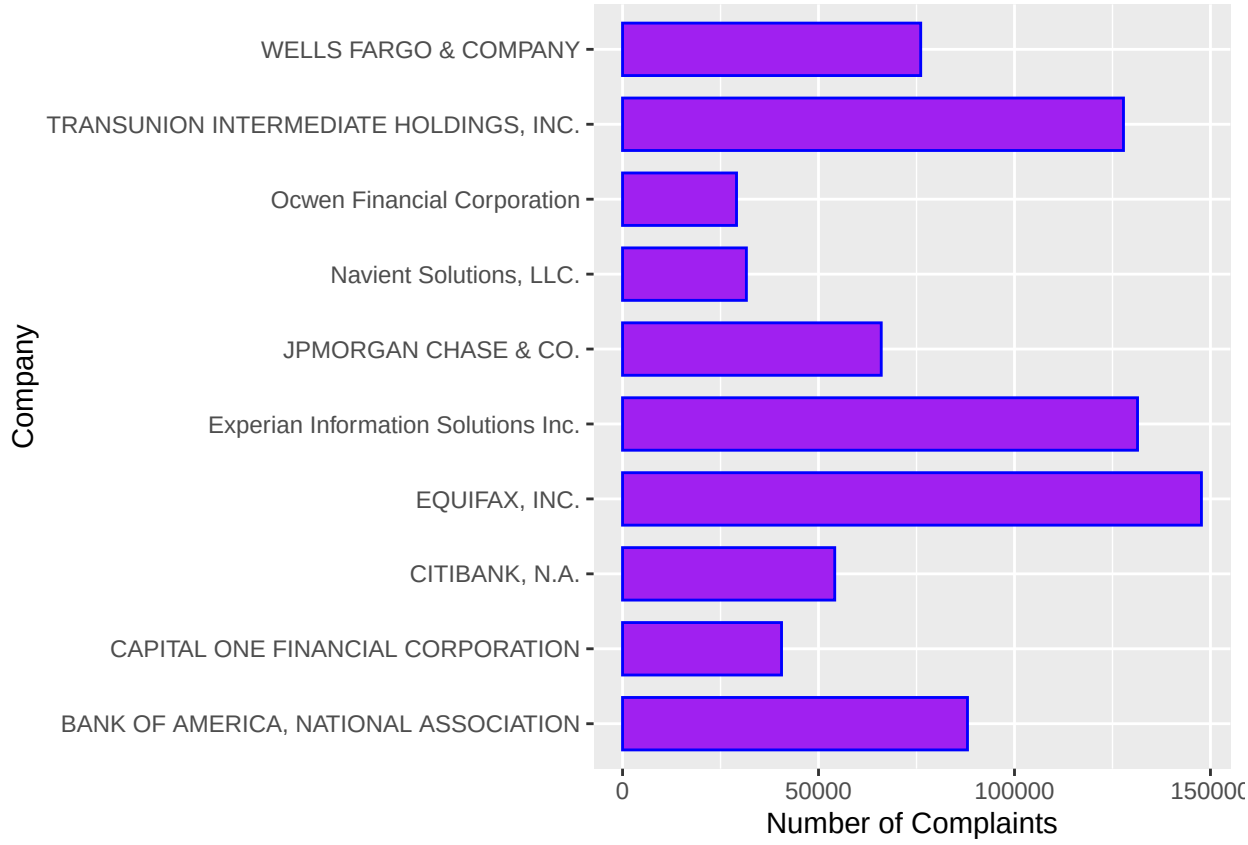
```
library(dplyr)
```

```
complaints <- fread("complaints.csv", header = T, sep = ',')
```

Data Visualization 1 - Top X companies with highest number of complaints (Bar Plot)

```
dv1 <- function(number) {
  complaints[complaints == ""] <- NA
  com.comp <- table(complaints$Company)
  com.comp2 <- as.data.frame(com.comp)
  com.comp3 <- com.comp2[tail(order(com.comp2$Freq), number), ]
  com.comp4 <- ggplot(com.comp3, aes(x = Var1, y = Freq)) + geom_col(width = 0.7, color = "blue", fill =
    labs(x = "Company", y = "Number of Complaints")
  com.comp4
  }
```

```
dv1(10) #Top 10 companies that have the most complaints
```



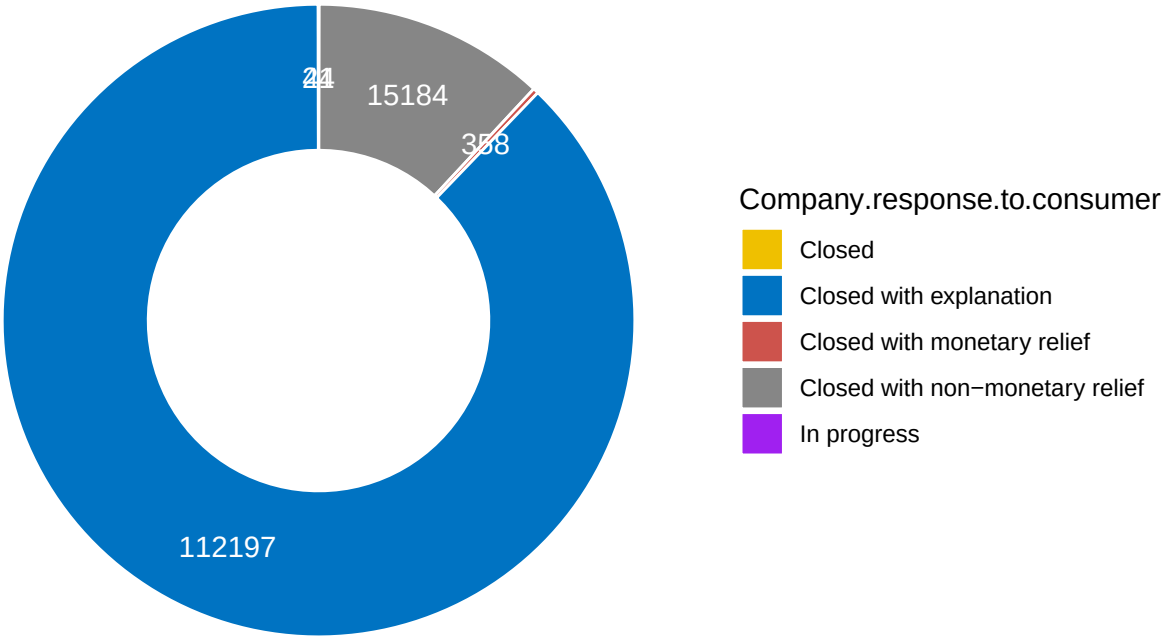
Data Visualization 2 - Response to consumers by certain company (Pie chart)

```
dv2 <- function(company) {
    a <- select(complaints, "Company", "Company response to consumer")
    b <- filter(a, Company %in% c(company))
    c <- as.data.frame(table(b))
    d <- c %>% arrange(desc(Company.response.to.consumer)) %>% mutate(lab.ypos = cumsum(Freq) - 0.5*Freq

    mycols <- c("#EFC000FF", "#0073C2FF", "#CD534CFF","#868686FF" , "purple")
    ggplot(d, aes(x = 2, y = Freq, fill = Company.response.to.consumer)) +
    geom_bar( stat = "identity", color = "white") +
    coord_polar("y", start = 0) +
    geom_text(aes(y = lab.ypos, label = Freq), color = "white") +
    scale_fill_manual(values = mycols) +
    theme_void() + xlim(0.5, 2.5) + labs(title = company)
    }
```

```
dv2("TRANSUNION INTERMEDIATE HOLDINGS, INC.") #Response to consumers by TRANSUNION INTERMEDIATE HOLDING
```

TRANSUNION INTERMEDIATE HOLDINGS, INC.



3 Early Takeaways

1. It seems that the company with most commplaints is Equifax followed by Experian Information Solutions. Because there were a lot of companies, it was hard to display all the companies in one bar chart. Therefore, it might be important to find a way to dispaly the companies or even subset the companies into sub categories for better visualization.

2. For the 2nd visualisation, I wanted to find out how a certain company responds to consumer in what ways. I found that TRANSUNION INTEREMEDIATE HOLDINGS, INC. responds mostly with explanation and sometimes with non-monetary relief. I am thinking it might be useful to compare how each companies respond to consumers and whether complaints are lower for those companies who answer with explanation for later analysis.

3. For later analysis, I am preparing to perform text analysis of the complaints. I will first have to clean the data and come up with a way to distinguish the level of anxiety each complaints contains. In addition, I am planning to come up with a way to visualize such complaints.