User Complaints Mining

gor Baranov, Michael Parravani, Ariana Biagi, Grace Tsai

February 4, 2019

Init libraries

Preparing data

Seting directory to current script (works in R Studio only)

```
#this.dir <- dirname(rstudioapi::getActiveDocumentContext()$path)
#setwd(this.dir)
```

Load and clean the customer complaints data

The data should be downloaded from Kaggle's Consumer Complaints Database.

Loading and removing rows with no complaint narrative and unnecessary colimns:

```
df <- read_csv(file="../data/Consumer_Complaints.csv.zip",col_names = TRUE)
df <- df[!is.na(df[,"Consumer complaint narrative"]),-c(1,7:18)]
df
```

```
## # A tibble: 199,970 x 5
##    Product           `Sub-product`   Issue `Sub-issue` `Consumer complaint ~
##    <chr>             <chr>           <chr> <chr>       <chr>
##  1 Credit reporting <NA>             Inco~ Account st~ I have outdated info~
##  2 Consumer Loan    Vehicle loan     Mana~ <NA>        I purchased a new ca~
##  3 Credit reporting <NA>             Cred~ Inadequate~ "An account on my cr~
##  4 Debt collection  Other (i.e. p~ Disc~ Not given ~ This company refuses~
##  5 Debt collection  Credit card     Impr~ Talked to ~ "This complaint is i~
##  6 Mortgage         Conventional ~ Sett~ <NA>        Started the refinanc~
##  7 Mortgage         Conventional ~ Appl~ <NA>        In XXXX, I and my ex~
##  8 Credit reporting <NA>             Cred~ Problem wi~ I have disputed seve~
##  9 Mortgage         Conventional ~ Loan~ <NA>        "Mortgage was transf~
## 10 Credit card      <NA>             Othe~ <NA>        "Was a happy XXXX ca~
## # ... with 199,960 more rows
```

Converting all but narrative columns to factors:

```
df$Product <- as.factor(df$Product)
df$`Sub-product` <- as.factor(df$`Sub-product`)
df$Issue <- as.factor(df$Issue)
df$`Sub-issue` <- as.factor(df$`Sub-issue`)
```

Feature engineering

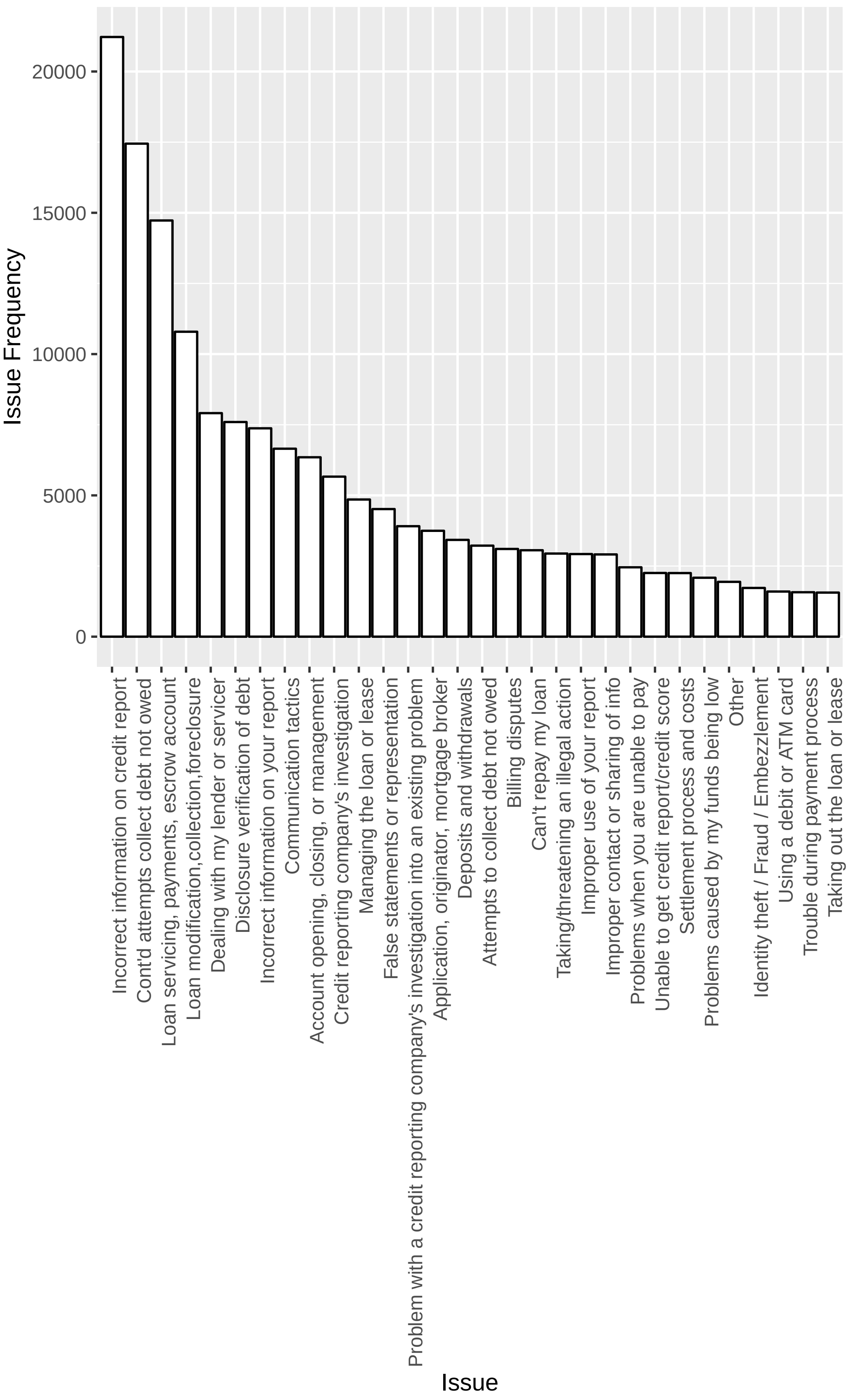
Creating 'complaints' dataframe having 30 most frequest "Issues":

```r
most_freq_issues <- levels(fct_infreq(df$Issue))[1:30]
complaints <- df[df$Issue %in% most_freq_issues,]
complaints[,c('Issue','Consumer complaint narrative')]
```

```
## # A tibble: 161,767 x 2
##    Issue                                    `Consumer complaint narrative`
##    <fct>                                    <chr>
##  1 Incorrect information on credit report   I have outdated information o~
##  2 Managing the loan or lease               I purchased a new car on XXXX~
##  3 Credit reporting company's investigation "An account on my credit repo~
##  4 Disclosure verification of debt          This company refuses to provi~
##  5 Improper contact or sharing of info      "This complaint is in regards~
##  6 Settlement process and costs             Started the refinance of home~
##  7 Application, originator, mortgage broker In XXXX, I and my ex-husband ~
##  8 Credit reporting company's investigation I have disputed several accou~
##  9 Loan servicing, payments, escrow account "Mortgage was transferred to ~
## 10 Taking/threatening an illegal action     "I am writing to request your~
## # ... with 161,757 more rows
```

Plotting distribution of the most frequent "Issues":

```r
ggplot() + aes(fct_infreq(complaints$Issue))+
  geom_histogram(colour="black", fill="white", stat = "count")+
  ylab("Issue Frequency")+
  xlab("Issue") + theme(axis.text.x = element_text(angle =90, hjust = 1))
```


Issue Frequency
20000
15000
10000
5000
0
Incorrect information on credit report
Cont'd attempts collect debt not owed
Loan servicing, payments, escrow account
Loan modification,collection,foreclosure
Dealing with my lender or servicer
Disclosure verification of debt
Incorrect information on your report
Communication tactics
Account opening, closing, or management
Credit reporting company's investigation
Managing the loan or lease
False statements or representation
Problem with a credit reporting company's investigation into an existing problem
Application, originator, mortgage broker
Deposits and withdrawals
Attempts to collect debt not owed
Billing disputes
Can't repay my loan
Taking/threatening an illegal action
Improper use of your report
Improper contact or sharing of info
Problems when you are unable to pay
Unable to get credit report/credit score
Settlement process and costs
Problems caused by my funds being low
Other
Identity theft / Fraud / Embezzlement
Using a debit or ATM card
Trouble during payment process
Taking out the loan or lease
Issue