

# User Complaints Mining

*gor Baranov, Michael Parravani, Ariana Biagi, Grace Tsai*

*February 4, 2019*

## Preparing data

### 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 columns:

```
if(!file.exists("./user-complaints-mining/df.Rds")) {
  df <- read_csv(file = "../data/Consumer_Complaints.csv.zip", col_names = TRUE)
  df <- df[, -c(1, 7, 9:18)]
  df <- df[!is.na(df[, "Consumer complaint narrative"]),] #199,970
  df <- df[!is.na(df[, "Company"]),] # no NA's
  df <- df[!is.na(df[, "Product"]),] # no NA's
  df <- df[!is.na(df[, "Issue"]),] # no NA's
  df <- df[!is.na(df[, "Sub-product"]),] # 147,788 total left
  df <- df[!is.na(df[, "Sub-issue"]),] # 81,940 total left

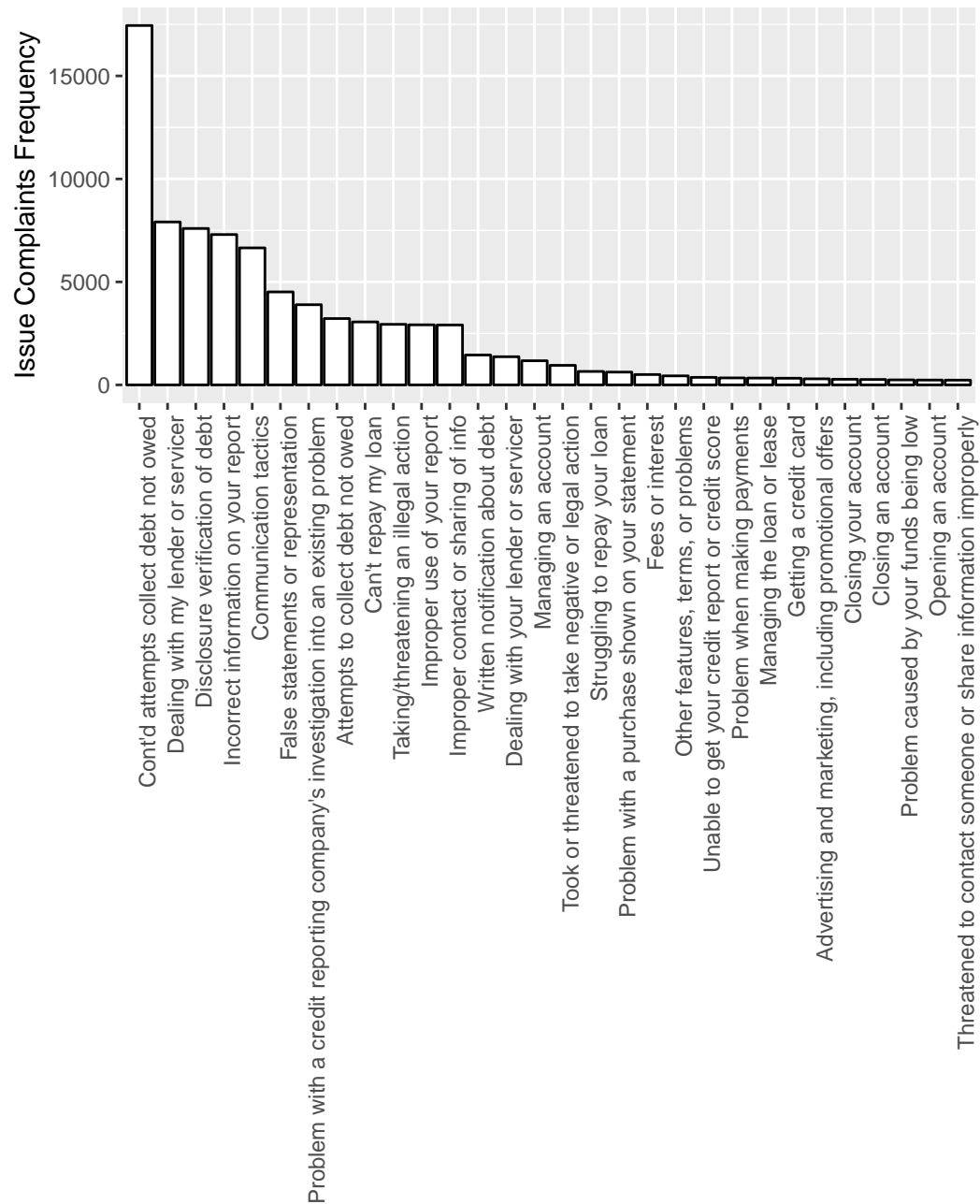
  # Converting all but narrative columns to factors
  df$Product <- as.factor(df$Product)
  df$`Sub-product` <- as.factor(df$`Sub-product`)
  df$Issue <- as.factor(df$Issue)
  df$`Sub-issue` <- as.factor(df$`Sub-issue`)
  df$Company <- as.factor(df$Company)
  saveRDS(df, file = "./user-complaints-mining/df.Rds")
  gc()
} else {
  df <- readRDS("./user-complaints-mining/df.Rds")
}
df
```

```
## # A tibble: 81,940 x 6
##   Product `Sub-product` Issue `Sub-issue` `Consumer complain Company
##   <fct>   <fct>         <fct> <fct>         <chr>         <fct>
## 1 Debt co~ Other (i.e. ph~ Discl~ Not given e~ This company refu~ The CBE~
## 2 Debt co~ Credit card     Impro~ Talked to a~ "This complaint i~ SQUARET~
## 3 Debt co~ Credit card     Takin~ Sued w/o pr~ "I am writing to ~ Selip &~
## 4 Debt co~ Other (i.e. ph~ Cont'~ Debt result~ My identity was s~ Southwe~
## 5 Student~ Federal studen~ Can't~ Can't get f~ "I was dropped fr~ AES/PHE~
## 6 Debt co~ Credit card     Discl~ Not given e~ The first communi~ Blatt, ~
## 7 Debt co~ Other (i.e. ph~ Commu~ Frequent or~ "My complaint is ~ AR Reso~
## 8 Debt co~ I do not know   False~ Attempted t~ In a clearance in~ SANTAND~
## 9 Student~ Non-federal st~ Can't~ Can't tempo~ XXXX University, ~ Navient~
## 10 Student~ Non-federal st~ Deal~ Received ba~ I had attended XX~ CITIZEN~
## # ... with 81,930 more rows
```

## Feature engineering

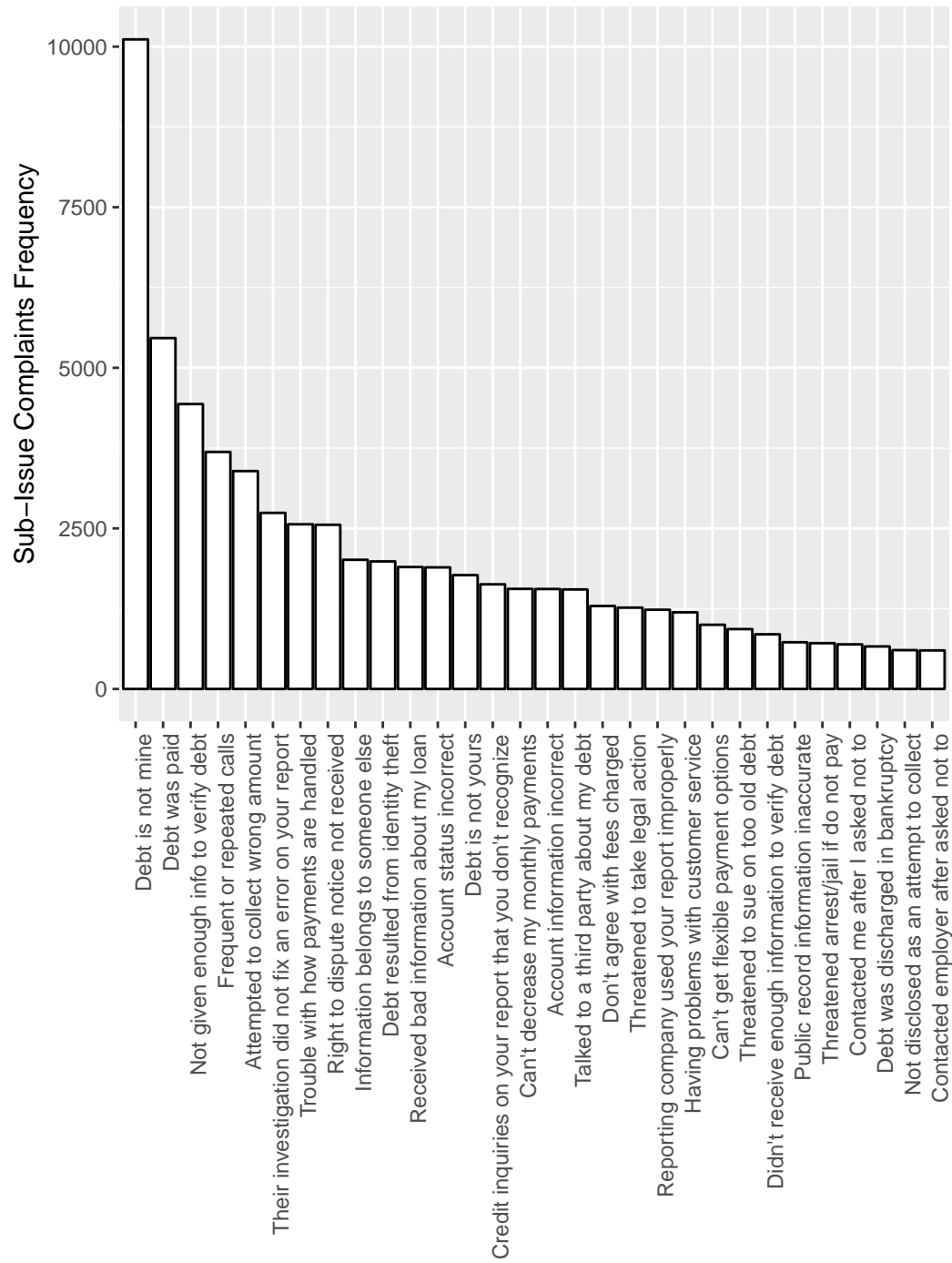
### Distribution of the most frequent “Issue” complaints

```
most_freq_issues_list <- levels(fct_infreq(df$Issue))[1:30]
ggplot() + aes(fct_infreq(df[df$Issue %in% most_freq_issues_list,]$Issue)) +
  geom_histogram(colour="black", fill="white", stat = "count") +
  ylab("Issue Complaints Frequency") + xlab("") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



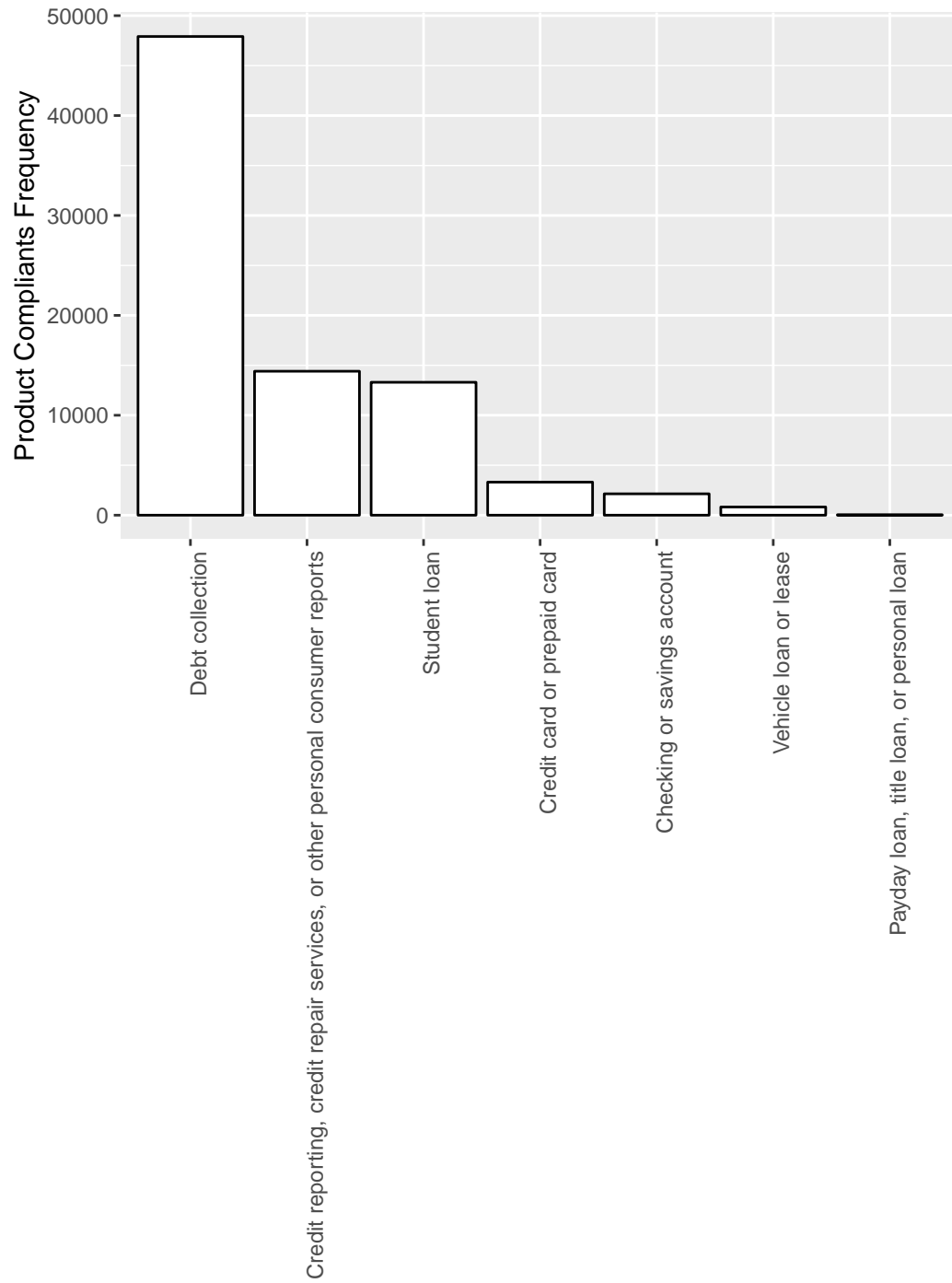
## Distribution of the most frequent “Sub-issue” complaints

```
most_freq_subissues_list <- levels(fct_infreq(df$`Sub-issue`))[1:30]
ggplot() + aes(fct_infreq(df[df$`Sub-issue` %in% most_freq_subissues_list,]$`Sub-issue`))+
  geom_histogram(colour="black", fill="white", stat = "count")+
  ylab("Sub-Issue Complaints Frequency") + xlab("")+
  theme(axis.text.x = element_text(angle =90, hjust = 1))
```



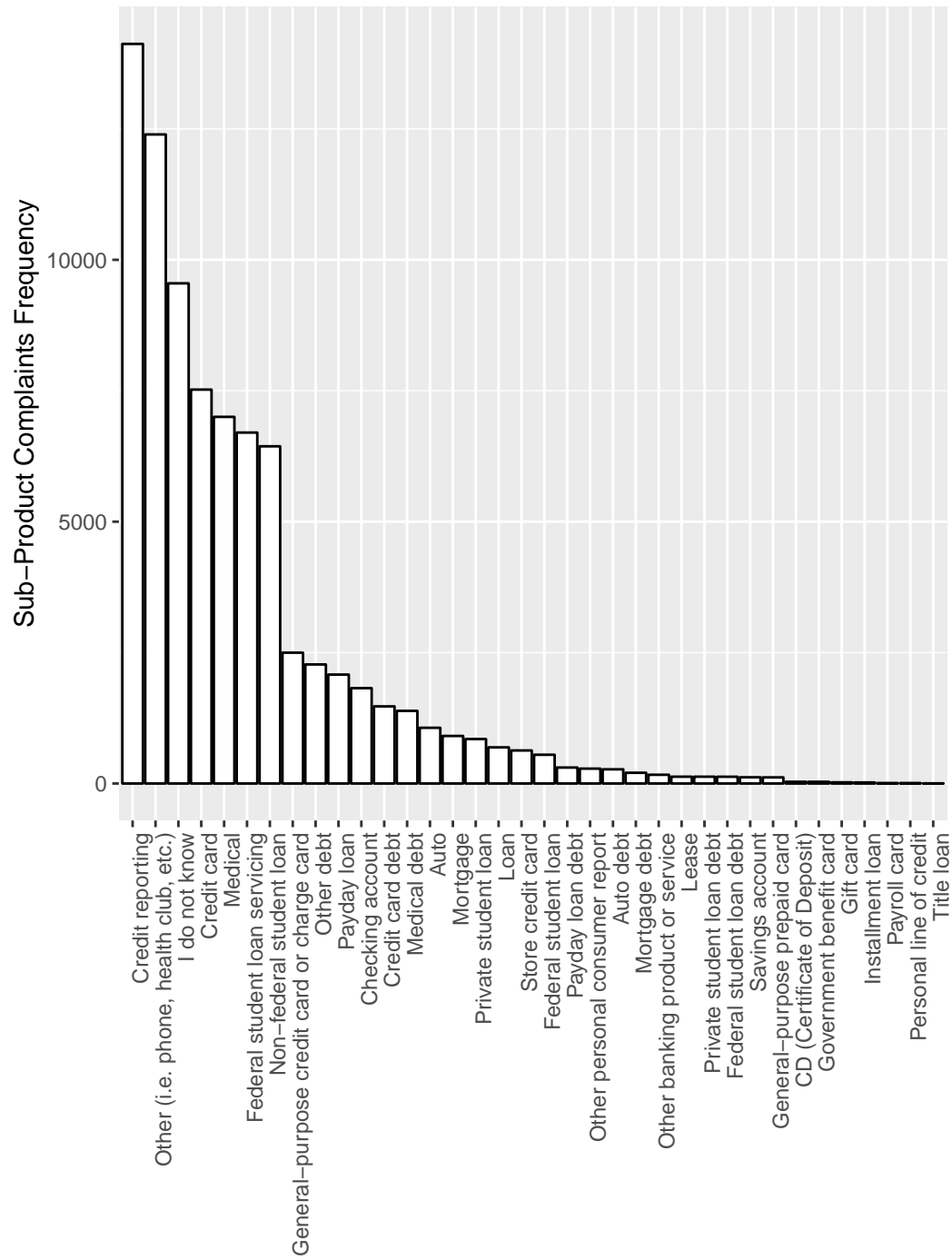
## Distribution of the most frequent “Product” complaints

```
most_freq_product_list <- levels(fct_infreq(df$Product))[1:30]
ggplot() + aes(fct_infreq(df[df$Product %in% most_freq_product_list,]$Product))+
  geom_histogram(colour="black", fill="white", stat = "count")+
  ylab("Product Compliants Frequency") + xlab("")+
  theme(axis.text.x = element_text(angle =90, hjust = 1))
```



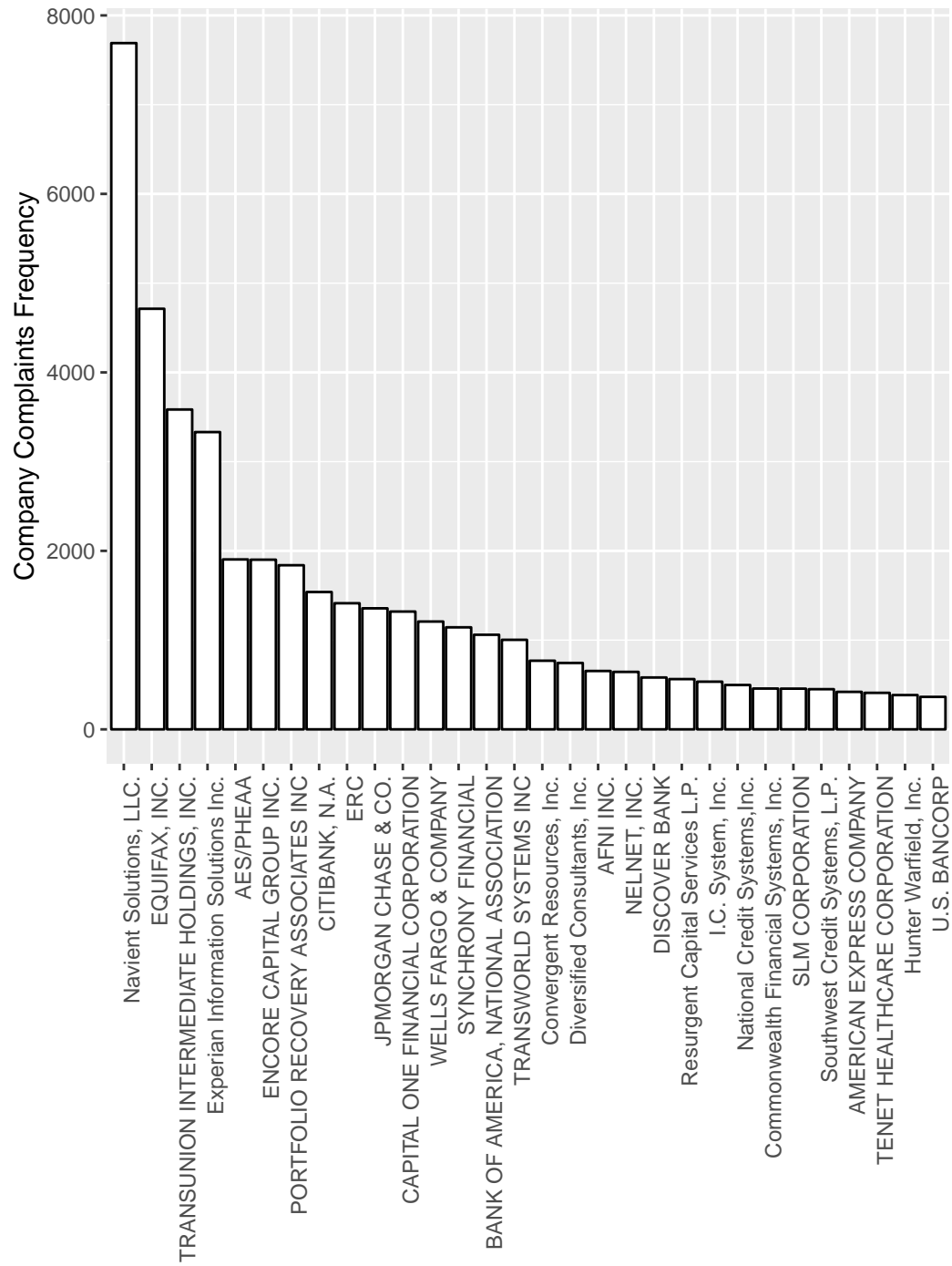
## Distribution of the most frequent “Sub-product” complaints

```
most_freq_subproduct_list <- levels(fct_infreq(df$`Sub-product`))
ggplot() + aes(fct_infreq(df[df$`Sub-product` %in% most_freq_subproduct_list,]$`Sub-product`))+
  geom_histogram(colour="black", fill="white", stat = "count")+
  ylab("Sub-Product Complaints Frequency") + xlab("")+
  theme(axis.text.x = element_text(angle =90, hjust = 1))
```



## Distribution of the most frequent “Company” complaints

```
most_freq_company_list <- levels(fct_infreq(df$Company))[1:30]
ggplot() + aes(fct_infreq(df[df$Company %in% most_freq_company_list,]$Company)) +
  geom_histogram(colour="black", fill="white", stat = "count") +
  ylab("Company Complaints Frequency") + xlab("") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



## Text Minig

### Split data into test and train sets

```
set.seed(123)
sample = sample.split(df$`Consumer complaint narrative`, SplitRatio = .5)
train = subset(df, sample == TRUE)
train$id <- c(1:nrow(train))
test = subset(df, sample == FALSE)
```

### Word analysis

Building a corpus, which is a collection of text documents VectorSource specifies that the source is character vectors. After that, the corpus needs a couple of transformations, including changing letters to lower case, removing punctuations/numbers and removing stop words. The general English stop-word list is tailored by adding some words specific to the documents in question.

```
if(!file.exists("./user-complaints-mining/myCorpus.Rds")) {
  myCorpus <- Corpus(VectorSource(train$`Consumer complaint narrative`))
  myCorpus <- tm_map(myCorpus, removePunctuation)
  myCorpus <- tm_map(myCorpus, removeNumbers)
  myCorpus <- tm_map(myCorpus, tolower)
  myStopwords <- c(stopwords(language="en", source="smart"),
    "xx", "xxxx", "xxxxxxxxxxxxxx", "xxxxxxxxxx",
    "told", "well", "month", "year"
  )
  myCorpus <- tm_map(myCorpus, removeWords, myStopwords)
  myCorpus <- tm_map(myCorpus, stripWhitespace)
  saveRDS(myCorpus, file = "./user-complaints-mining/myCorpus.Rds")
  gc()
} else {
  myCorpus <- readRDS("./user-complaints-mining/myCorpus.Rds")
}
print(myCorpus[1:5]$content)
```

```
## [1] " company refuses provide verification validation debt fdcpa debt mine"
## [2] " writing request assistance deceptive practices collection lawfirm appears
tactics violating consumer protection law debt collection practices depriving consumers
rights dispute received notice company day contacted offices instructed memo dated
instruct contact plaintiff attorney court instructions provided contacted plaintiff
attorney phone faxed letter disputing debt letter company responded letter dated sending
bill due date requested bill showing balance back made payment back disputing amount owed
disputing charges wrote back company faxed dispute letter continue dispute amount owed
company response furnished information disputing owe disputing balance inaccurate needed
proof charges activity account paragraph letter disputing amount send letter letter
company disputing balance requesting documents company responded letter communication
received letter copy default judgment filed court clerk office indicating failed respond
judgment facts judgment firm served original judgment attaching dispute letters showing
responded instructed office occasions plaintiff failed respond dispute furnish
information providedand unable obtain proof original debt credible legal procedure settle
debts utilized unfaithful dirty tactics violated rights court committed perjury law
filing false documents court defaulted judgment failed respond fact responded failed
```

```
furnish proof court house clerk office told company notify offices contact company told
court clerk office respond summons clerks office granted default judgement based false
information respond summons filed false affirmation clerk office"
## [3] " communication received debt collector court summons delivered mother laws home
received summons hand week summons stated alledgedly owed money debt collector advisement
dispute debt days demand debt collector validate debt attempted demand validation debt
online research certified letter days ago post office told today approximately certified
letter attempted delivered days mailed accepted debt collector today withoyut knowledge
information sufficient form opinion truth accuracy claim based deny generally
specifically claim debt collector"
## [4] " attended forced loan attend school loan interest rate make payments attended
school enrolled told classmates received reduction balance curious criteria receive
reduction"
## [5] "years ago harassed issue asked send application form signed understood charged
extra pet damages paid additional month pet covered damages wanted photos damages
property manager told walk left apartment thing scam money contacted requested
information"
```

## Steming

```
dictCorpus <- myCorpus
myCorpus <- tm_map(myCorpus, stemDocument)
print(myCorpus[1:5]$content)
```

```
## [1] "compani refus provid verif valid debt fdcpa debt mine"
## [2] "write request assist decept practic collect lawfirm appear tactic violat consum
protect law debt collect practic depriv consum right disput receiv notic compani day
contact offic instruct memo date instruct contact plaintiff attorney court instruct
provid contact plaintiff attorney phone fax letter disput debt letter compani respond
letter date send bill due date request bill show balanc back made payment back disput
amount owe disput charg wrote back compani fax disput letter continu disput amount owe
compani respons furnish inform disput owe disput balanc inaccur need proof charg activ
account paragraph letter disput amount send letter letter compani disput balanc request
document compani respond letter communic receiv letter copi default judgment file court
clerk offic indic fail respond judgment fact judgment firm serv origin judgment attach
disput letter show respond instruct offic occas plaintiff fail respond disput furnish
inform providedand unabl obtain proof origin debt credibl legal procedur settl debt util
unfaith dirti tactic violat right court commit perjuri law file fals document court
default judgment fail respond fact respond fail furnish proof court hous clerk offic told
compani notifi offic contact compani told court clerk offic respond summon clerk offic
grant default judgement base fals inform respond summon file fals affirm clerk offic"
## [3] "communic receiv debt collector court summon deliv mother law home receiv summon
hand week summon state alledg owe money debt collector advis disput debt day demand debt
collector valid debt attempt demand valid debt onlin research certifi letter day ago post
offic told today approxim certifi letter attempt deliv day mail accept debt collector
today withoyut knowledg inform suffici form opinion truth accuraci claim base deni
general specif claim debt collector"
## [4] "attend forc loan attend school loan interest rate make payment attend school
enrol told classmat receiv reduct balanc curious criteria receiv reduct"
## [5] "year ago harass issu ask send applic form sign understood charg extra pet damag
paid addit month pet cover damag want photo damag properti manag told walk left apart
thing scam money contact request inform"
```



## Building a Document-Term Matrix

This operation requires 64GB of RAM. To avoid calculation, the pre-build myDtm object will be loaded from the file system. To recalculate it needs to be removed from the file system first.

```
if(!file.exists("./user-complaints-mining/myDtm.Rds")) {
  myDtm <- TermDocumentMatrix(myCorpus, control = list(minWordLength = 1))
  rowTotals <- apply(myDtm, 1, sum) #Find the sum of words in each Document
  myDtm <- myDtm[rowTotals > 0, ] #remove all docs without words
  saveRDS(myDtm, file = "./user-complaints-mining/myDtm.Rds")
  gc()
} else {
  myDtm <- readRDS("./user-complaints-mining/myDtm.Rds")
}
inspect(myDtm)
```

```
## <<TermDocumentMatrix (terms: 27440, documents: 40970)>>
## Non-/sparse entries: 1585496/1122631304
## Sparsity          : 100%
## Maximal term length: 124
## Weighting          : term frequency (tf)
## Sample            :
##
##      Docs
## Terms  11425 11904 31399 31930 33039 34137 35502 35599 36932 9698
## account    26    21    60    62    18    64     5     8     0     6
## call        3    10     2     0    18     0     3     0     2     7
## collect     0     0     5     0     0     4     0     6     0    21
## credit      3     1     0     0     0     2    13     2     2     0
## debt        0     0     1     1     0     1     0     0     3    68
## inform     11     5     8     8     6     8    10    73     5     9
## loan        5     0     2     0    11     0     2     0     3     0
## payment     2    12    27    27    24    27     7     0     0     1
## receiv      4     2     0     0    25     0     1     0     8    10
## report      1     0    25    27     1    28    26     8     0     0
```

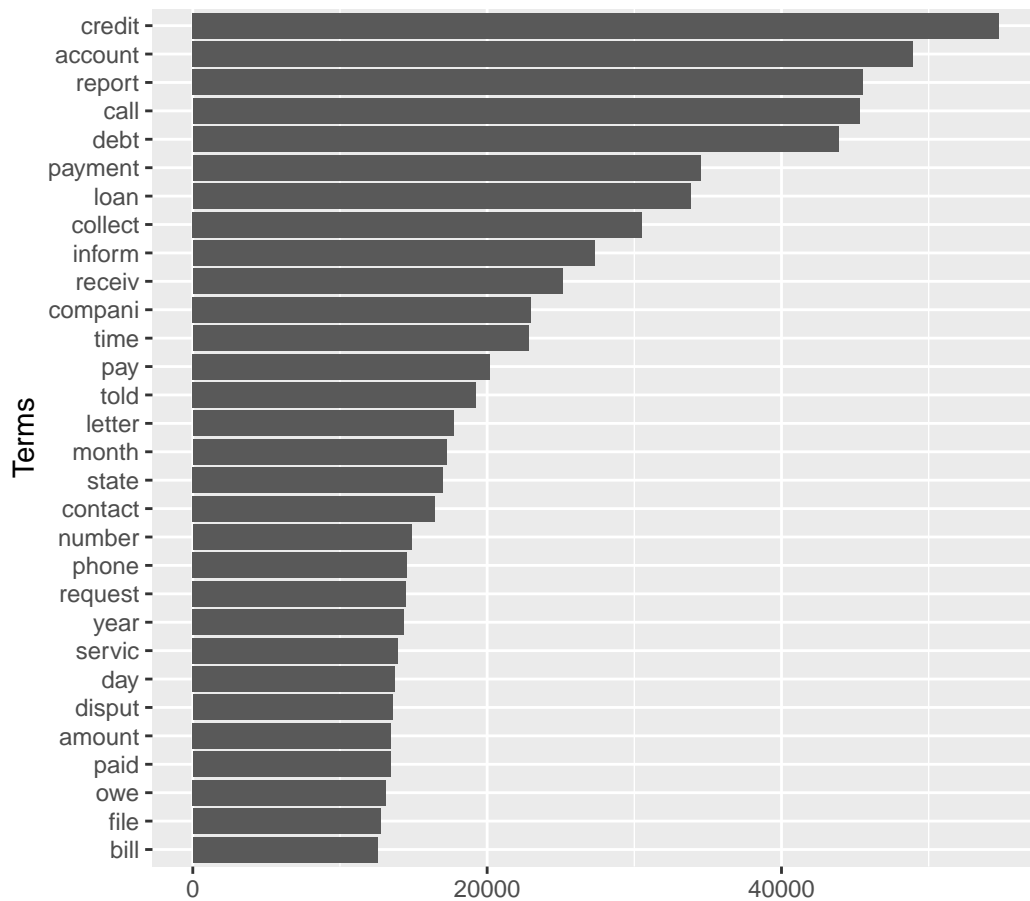


Figure 1: 30 Most Frequent Terms

## Frequent Terms and Association

```
freq.terms <- findFreqTerms(myDtm, lowfreq=5)
term.freq <- rowSums(as.matrix(myDtm))
term.freq <- subset(term.freq, term.freq >= 5)
```

```
dfTerms <- data.frame(term = names(term.freq), freq = term.freq)
ggplot(dfTerms[order(-dfTerms$freq),][1:30,], aes(x = reorder(term, freq), y = freq)) +
  geom_bar(stat = "identity") + xlab("Terms") + ylab("") + coord_flip()
```

## Which words are associated with term “loan”?

```
findAssocs(myDtm, c('loan'), 0.35)
```

```
## $loan
## student payment privat navient interest consolid
##      0.60      0.40      0.39      0.39      0.38      0.36
```

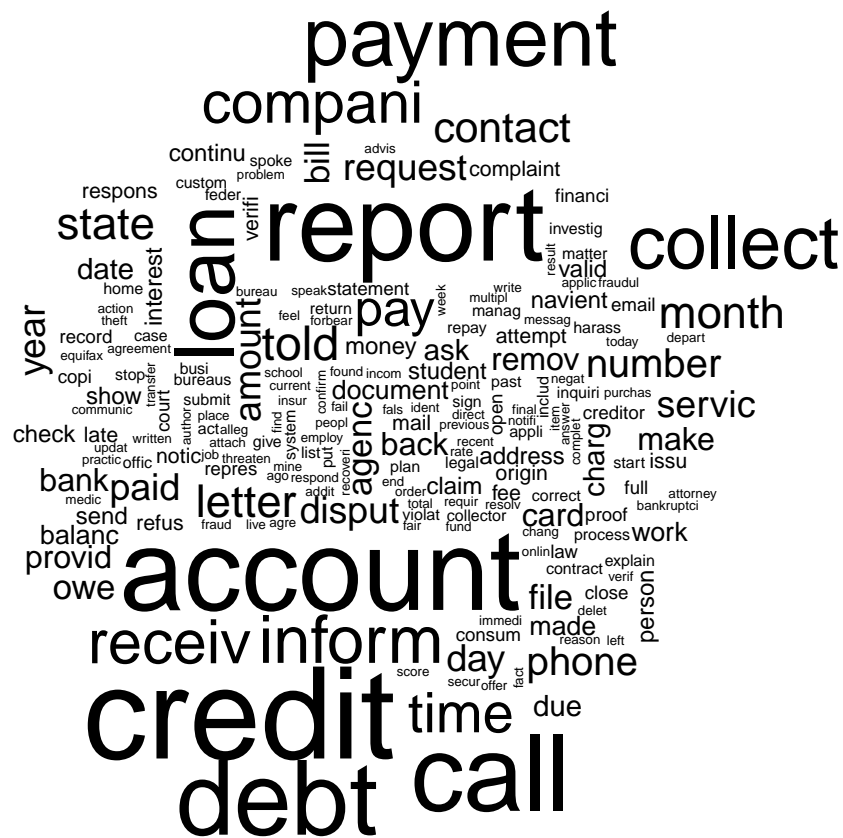


Figure 2: Words Cloud of Complaints

Building word cloud:

```
library(wordcloud)
m <- as.matrix(myDtm)
v <- sort(rowSums(m), decreasing=TRUE)
myNames <- names(v)
d <- data.frame(word=myNames, freq=v)
wordcloud(d$word, d$freq, min.freq=20, scale=c(4,.2), max.words = 200)
```

## Building LDA model for the ‘train’ set of complaints

```
library(topicmodels)
dtm <- as.DocumentTermMatrix(myDtm)
numTopics <- 30 # 7, 10 or 30
topicsFileName <- paste("./user-complaints-mining/lda",numTopics,".Rds", sep = "")
ui = unique(dtm$i) #array of unique row ids in 'train' set
if(!file.exists(topicsFileName)) {
  dtm.new = dtm[ui,]
  train.lda <- LDA(dtm.new, k = numTopics) # identify topics
  saveRDS(train.lda, file = topicsFileName)
  gc()
} else {
  train.lda <- readRDS(topicsFileName)
}
(train.terms <- terms(train.lda, 10)) # first terms of every topic
```

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
## [1,]	"debt"	"return"	"told"	"court"	"inform"	"loan"
## [2,]	"collector"	"advis"	"ask"	"file"	"person"	"student"
## [3,]	"alleg"	"cancel"	"contact"	"bankruptci"	"secur"	"navient"
## [4,]	"valid"	"purchas"	"inform"	"recoveri"	"number"	"repay"
## [5,]	"communic"	"offer"	"back"	"attorney"	"social"	"forbear"
## [6,]	"llc"	"charg"	"email"	"garnish"	"equifax"	"program"
## [7,]	"ceas"	"citi"	"spoke"	"law"	"data"	"privat"
## [8,]	"midland"	"receiv"	"repres"	"case"	"individu"	"defer"
## [9,]	"attempt"	"american"	"refus"	"judgment"	"breach"	"consolid"
## [10,]	"limit"	"express"	"speak"	"portfolio"	"access"	"default"
	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12
## [1,]	"account"	"disput"	"report"	"letter"	"servic"	"complaint"
## [2,]	"close"	"verifi"	"credit"	"show"	"process"	"attach"
## [3,]	"open"	"inform"	"remov"	"certifi"	"submit"	"file"
## [4,]	"number"	"delet"	"bureaus"	"apart"	"custom"	"record"
## [5,]	"activ"	"item"	"show"	"manag"	"form"	"state"
## [6,]	"mine"	"file"	"negat"	"leas"	"applic"	"cfpb"
## [7,]	"current"	"investig"	"bureau"	"send"	"chase"	"note"
## [8,]	"respons"	"day"	"updat"	"day"	"request"	"document"
## [9,]	"maci"	"inaccur"	"incorrect"	"move"	"time"	"action"
## [10,]	"onlin"	"fcra"	"list"	"detail"	"paperwork"	"claim"
	Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18
## [1,]	"call"	"pay"	"work"	"violat"	"address"	"compani"
## [2,]	"phone"	"work"	"harass"	"act"	"inquiri"	"request"
## [3,]	"number"	"time"	"compani"	"consum"	"author"	"provid"
## [4,]	"time"	"money"	"contact"	"fair"	"pull"	"document"
## [5,]	"day"	"job"	"messag"	"ident"	"file"	"valid"
## [6,]	"told"	"make"	"threaten"	"theft"	"inquir"	"proof"
## [7,]	"compani"	"school"	"stop"	"law"	"contact"	"contract"
## [8,]	"back"	"back"	"left"	"feder"	"compani"	"copi"
## [9,]	"answer"	"year"	"employ"	"inform"	"knowledg"	"send"
## [10,]	"person"	"financi"	"leav"	"practic"	"type"	"sign"
	Topic 19	Topic 20	Topic 21	Topic 22	Topic 23	Topic 24
## [1,]	"credit"	"collect"	"bill"	"receiv"	"charg"	"year"
## [2,]	"card"	"agenc"	"due"	"owe"	"fraud"	"home"
## [3,]	"score"	"contact"	"issu"	"state"	"money"	"ago"

##	[4,]	"discov"	"attempt"	"past"	"mail"	"fraudul"	"mortgag"
##	[5,]	"appli"	"debt"	"time"	"notic"	"claim"	"live"
##	[6,]	"affect"	"consult"	"correct"	"system"	"fund"	"husband"
##	[7,]	"drop"	"diversifi"	"wrong"	"notif"	"polic"	"hous"
##	[8,]	"limit"	"erc"	"problem"	"send"	"investig"	"sold"
##	[9,]	"point"	"notic"	"resolv"	"notifi"	"file"	"tax"
##	[10,]	"xxx"	"turn"	"error"	"correspond"	"acct"	"wife"
##		Topic 25	Topic 26	Topic 27	Topic 28	Topic 29	Topic 30
##	[1,]	"payment"	"insur"	"bank"	"interest"	"paid"	"date"
##	[2,]	"make"	"medic"	"check"	"month"	"amount"	"origin"
##	[3,]	"month"	"car"	"fee"	"payment"	"balanc"	"creditor"
##	[4,]	"late"	"financi"	"charg"	"rate"	"full"	"delinqu"
##	[5,]	"made"	"hospit"	"deposit"	"pay"	"pay"	"status"
##	[6,]	"pay"	"financ"	"debit"	"amount"	"statement"	"well"
##	[7,]	"day"	"vehicl"	"transact"	"year"	"agreement"	"fargo"
##	[8,]	"time"	"compani"	"cash"	"appli"	"settl"	"amount"
##	[9,]	"told"	"doctor"	"america"	"increas"	"agre"	"unpaid"
##	[10,]	"due"	"claim"	"day"	"lower"	"settlement"	"number"

## Show complaints correlation to topics in 'train' set

```
train.topics <- topics(train.lda, 5, threshold=.1)
for (i in c(1,55,500,333)) {
  print (paste(" "))
  print (paste("Complaint# ", i))
  print(paste("Topics found: ", train.topics[i]))
  print (train$`Consumer complaint narrative`[ui[i]])
}

## [1] " "
## [1] "Complaint# 1"
## [1] "Topics found: c(1, 18)"
## [1] "This company refuses to provide me verification and validation of debt per my
right under the FDCPA. I do not believe this debt is mine."
## [1] " "
## [1] "Complaint# 55"
## [1] "Topics found: 13"
## [1] "I have received calls from Midland Credit company since XX/XX/2015. They do not
ask for me by name but say they are trying to collect a debt. The same number of this
company has called me every day for the last several weeks. The number they used before
in XX/XX/XXXX was XXXX. They are now calling from XXXX. I spoke to a manager XXXX XXXX
and requested that they remove my number from their call list. This is the third time I
have ask them to quit calling my number. They have never complied. I HAVE NEVER GIVEN
THEM MY NAME AS THEY DONT ADDRESS ME BY NAME! All these people have is my phone number to
continue to harass me. They all sound like they are from XXXX."
## [1] " "
## [1] "Complaint# 500"
## [1] "Topics found: c(6, 14)"
## [1] "I am seeking information about my Genesis loans through XXXX in XXXX, PA. I have
heard about all of the loan forgiveness because of the XXXX XXXX settlement. I was one of
the many students who were talked into the lending scheme brought on by these schools and
Genesis lending company. I was charged horrible interest payments that I could n't
afford, while still in school. As well as, pulled out of class because I could not afford
payments. Horrible interest rates still haunt me, along with unreasonable monthly
payments. I am already on deferment, on which they still require me to make minimum
payments. I have a loan balance of {$21000.00} and have no idea how I will be able to
keep up with this. I feel as though I was a victim as well as other students who have
already had their loans forgiven. The way the company went about handing these loans out
to young students who know nothing about them, was wrong. My loan servicer XXXX told me
that Genesis has already handled the persons eligible for forgiveness and passed me along
to their company. So my question is, why was I considered not eligible? I attended school
in XX/XX/XXXX-XX/XX/XXXX and suffered just like everyone else who attended these scheming
schools. Please help!"
## [1] " "
## [1] "Complaint# 333"
## [1] "Topics found: c(1, 12, 14, 26)"
## [1] "Talked to someone not related to me about debt and left information about debt
with my name on front door where everyone could see it and read it. They are also listing
someone else in the document I do n't know, does n't exist and/Or has nothing to do with
this debt. The documents were so cigarette smoke infiltrated they made me sick."
```

## Topics related to most frequent issues

### Function definitions

```
prepareCorpus <- function(textArr) {
  myCorpus <- Corpus(VectorSource(textArr))
  myCorpus <- tm_map(myCorpus, removePunctuation)
  myCorpus <- tm_map(myCorpus, removeNumbers)
  myCorpus <- tm_map(myCorpus, tolower)
  myStopwords <- c(stopwords(language="en", source="smart"),
                   "xx", "xxxx", "xxxxxxxxxxxxxx", "xxxxxxxxxx")
  myCorpus <- tm_map(myCorpus, removeWords, myStopwords)
  myCorpus <- tm_map(myCorpus, stripWhitespace)

  return (myCorpus)
}

# Stemming and reverse - long operation!
doStemming <- function(myCorpus) {
  #dictCorpus <- myCorpus
  myCorpus <- tm_map(myCorpus, stemDocument)
  #myCorpus <- tm_map(myCorpus, stemCompletion, dictionary=dictCorpus)
  return (myCorpus)
}

getBestTopicNums <- function(arr, threshold) {
  topNums <- sort(arr, index.return=TRUE, decreasing = TRUE)$ix
  arr <- arr[arr>threshold]
  return (topNums[1:length(arr)])
}
```

## Form a matrix of topic probability vectors for each issue

```
issueTopicsProbMat <- matrix(NA,nrow = 30, ncol = 30)

# for each issue
for (i in 1:30) {
  issueName <- most_freq_issues_list[i]
  print(issueName)

  # matrix of issue topics probabilities
  issueids <- subset(train[,c("id")], train$Issue == issueName)
  uiIds <- match(issueids$id, ui)
  topicsProbabMatr <- train.lda@gamma[uiIds,]
  meanVec <- colMeans(topicsProbabMatr, na.rm = T)
  #tops <- getBestTopicNums(meanVec, .05)
  #print(train.terms[,tops])
  #print (meanVec)
  issueTopicsProbMat[i,] <- meanVec
}

## [1] "Cont'd attempts collect debt not owed"
## [1] "Dealing with my lender or servicer"
## [1] "Disclosure verification of debt"
## [1] "Incorrect information on your report"
## [1] "Communication tactics"
## [1] "False statements or representation"
## [1] "Problem with a credit reporting company's investigation into an existing problem"
## [1] "Attempts to collect debt not owed"
## [1] "Can't repay my loan"
## [1] "Taking/threatening an illegal action"
## [1] "Improper use of your report"
## [1] "Improper contact or sharing of info"
## [1] "Written notification about debt"
## [1] "Dealing with your lender or servicer"
## [1] "Managing an account"
## [1] "Took or threatened to take negative or legal action"
## [1] "Struggling to repay your loan"
## [1] "Problem with a purchase shown on your statement"
## [1] "Fees or interest"
## [1] "Other features, terms, or problems"
## [1] "Unable to get your credit report or credit score"
## [1] "Problem when making payments"
## [1] "Managing the loan or lease"
## [1] "Getting a credit card"
## [1] "Advertising and marketing, including promotional offers"
## [1] "Closing your account"
## [1] "Closing an account"
## [1] "Problem caused by your funds being low"
## [1] "Opening an account"
## [1] "Threatened to contact someone or share information improperly"
```



## Identify issues of random 'test' complaints

```
distInfo <- function(vec, mat) {
  retVec <- vector(length = nrow(mat))
  for (i in 1:nrow(mat)) {
    retVec[i] <- dist(rbind(vec,mat[i,]))
  }
  retVec <- sort(retVec, index.return=TRUE)
  return (retVec)
}

numOfTests <- 10
testThreshold =0.1

set.seed(122)
randomIds <- sample(nrow(test),numOfTests)

test.corpus <- prepareCorpus(test$`Consumer complaint narrative`[randomIds])
test.corpus <- doStemming(test.corpus)
test.dtm <- DocumentTermMatrix(test.corpus, control = list(minWordLength = 1))
test.topics <- posterior(train.lda,test.dtm)

#test.topics$topics[2,]
for (i in 1:numOfTests) {

  issuesFound <- distInfo (test.topics$topics[i,],issueTopicsProbMat)

  tops <- getBestTopicNums(test.topics$topics[i,], testThreshold)
  print("-----")
  # print("Topics identified:")
  # print(train.terms[,tops])
  print (paste("Inferred Issue: ", most_freq_issues_list[issuesFound$ix[1:3]]))
  print("-----")
  print (paste("Original Issue: ", test[[randomIds[i],"Issue"]]))
  print("-----")
  print(paste("Complaint narrative:",test[[randomIds[i],"Consumer complaint narrative"]]))
}

## [1] "-----"
## [1] "Inferred Issue: Problem with a credit reporting company's investigation into an
existing problem"
## [2] "Inferred Issue: Incorrect information on your report"
## [3] "Inferred Issue: Attempts to collect debt not owed"
## [1] "-----"
## [1] "Original Issue: Incorrect information on your report"
## [1] "-----"
## [1] "Complaint narrative: I have recently obtained a copy of my credit file and there
are many mistakes in my rep XXXX/ XXXX XXXX XXXX - XXXX : Unverified account XXXX/ XXXX
XXXX XXXX XXXX XXXX : I have never paid late XXXX XXXX XXXX XXXX XXXXXXXX XXXX XXXX :
Unverified account"
## [1] "-----"
## [1] "Inferred Issue: Problem when making payments"
## [2] "Inferred Issue: Managing the loan or lease"
## [3] "Inferred Issue: Fees or interest"
```

```

## [1] "-----"
## [1] "Original Issue: Problem when making payments"
## [1] "-----"
## [1] "Complaint narrative: Addendum to CFPB Complaint XXXX additional Kohls credit
representative complaints : ( As CFPB does not allow modifying complaint once entered : )
After reviewing Kohls.com online payment documentation there were two additional problems
in my interaction with Kohls credit representative on XX/XX/XXXX : Kohls credit
representative said with full confidence that Kohls charge card payments only post same
day in the local store up until XXXX XXXX. Kohls.com online documentation states payments
will post the same day any time up until store close. Kohls credit representative said
with full confidence that Kohls charge payments made the same day online, even before
XXXX XXXX, will not post that same day. Kohls.com online documentation states that online
payments made before XXXX XXXX will post that same day. The agent was a absolutely
terrible representative for Kohls. I now have no faith whatsoever in any interaction with
Kohls credit personnel, including supervisors. I think I have to go through CFPB on
everything account related, to get a straight story. XXXX"
## [1] "-----"
## [1] "Inferred Issue: Attempts to collect debt not owed"
## [2] "Inferred Issue: Written notification about debt"
## [3] "Inferred Issue: Cont'd attempts collect debt not owed"
## [1] "-----"
## [1] "Original Issue: Cont'd attempts collect debt not owed"
## [1] "-----"
## [1] "Complaint narrative: I was a victim of identity theft and noticed a collection
from Central Financial Control I 've attempted to dispute multiple times however no
response on their end. Not sure what else to do. Police department told me to contact the
federal trade commision which I wrote a formal complaint against Central Financial
Control in regards to this matter. The FTC also instructed me to complete a id theft
affidavit through identity theft.gov. Attached is my document. Thank you"
## [1] "-----"
## [1] "Inferred Issue: Managing the loan or lease"
## [2] "Inferred Issue: False statements or representation"
## [3] "Inferred Issue: Took or threatened to take negative or legal action"
## [1] "-----"
## [1] "Original Issue: Disclosure verification of debt"
## [1] "-----"
## [1] "Complaint narrative: Financed A Car in 2011 XXXX XXXX XXXX ... XXXX XXXX ( Buy
Here Pay Here ) For XXXX dollars had It almost Paid Off Down to XXXX Dollars ... Car
Started Giving Me Multiple Problems Called My Dealership Asking If They Could Have It
Fixed Or Put Me In Another Vechicle They Basically Told Me No It Wasent Thier Problem ...
I lost Out On Money & Also Had To Find Rides Back & Fourth To Work..Called Them Up To
Come & Get The Car & They Wouldnt Take It Back ... Financed From Them Before in Good
Standing ... They Also Went out Of Business Shortly After ..."
## [1] "-----"
## [1] "Inferred Issue: Cont'd attempts collect debt not owed"
## [2] "Inferred Issue: Improper contact or sharing of info"
## [3] "Inferred Issue: False statements or representation"
## [1] "-----"
## [1] "Original Issue: Communication tactics"
## [1] "-----"
## [1] "Complaint narrative: IQ Data ( DBA Rent Collect Global ) called 6 times within 3
days at my work place. They requested I pay an outstanding utility and move out fee. I
requested proof of the bill via mail. Caller was a bully and insinuated my credit would
be negatively affected and I would start paying an interest penalty if I did not

```

voluntarily pay today. The thing is -- I 've already attempted to pay the bill 10 days after the Landlord billed me. Key dates below : XX/XX/XXXX - Moved out of Apartment XX/XX/XXXX - Received final bill utility bill of {\$120.00} from LL with instructions to mail to Rent Collect Global XX/XX/XXXX - Mailed {\$120.00} to Rent Collect Global XXXX - Rent Collect Global RETURNED and voided my check ( I have scanned/saved a copy of the returned envelope, check, and bill ). Rent Collect Global stated I had no account on file with them XX/XX/XXXX - Received 4 calls at work. Debt collector stated I owed {\$320.00}, not the original {\$120.00} I was billed. I asked for proof. Debt collect stated none would probably be given. \nXX/XX/XXXX - Received another call & voicemail XX/XX/XXXX - Received another call and this guy stated he would start charging interest on my outstanding debt where there was no proof of the additional {\$200.00} of outstanding debt."

## [1] "-----"

## [1] "Inferred Issue: Written notification about debt"

## [2] "Inferred Issue: Disclosure verification of debt"

## [3] "Inferred Issue: Attempts to collect debt not owed"

## [1] "-----"

## [1] "Original Issue: Cont'd attempts collect debt not owed"

## [1] "-----"

## [1] "Complaint narrative: This company has continued to report the referenced collection account on my credit report, while failing to provide me with proper validation that I request from them in mailed correspondence dated XX/XX/2015. This is a flagrant violation of the FDCPA"

## [1] "-----"

## [1] "Inferred Issue: Took or threatened to take negative or legal action"

## [2] "Inferred Issue: False statements or representation"

## [3] "Inferred Issue: Taking/threatening an illegal action"

## [1] "-----"

## [1] "Original Issue: Improper contact or sharing of info"

## [1] "-----"

## [1] "Complaint narrative: Please refer to original complaint # XXXX. I am attaching more documentation for your records"

## [1] "-----"

## [1] "Inferred Issue: Threatened to contact someone or share information improperly"

## [2] "Inferred Issue: Improper contact or sharing of info"

## [3] "Inferred Issue: Communication tactics"

## [1] "-----"

## [1] "Original Issue: Communication tactics"

## [1] "-----"

## [1] "Complaint narrative: XXXX XXXX XXXX HIRED CAVALRY TO SKIP TRACE ME AND FAMILY MEMEBERS TO LOCATE ESTRANGED BROTHER AND SISTER -IN-LAW, WHILE VIOLATING OUR PRIVACY AND HARASSMENT VIA PHONE, ( SEE ATTACHED INFO ) ALL ARE IN VIOLATION OF FDCPA AND TCPA."

## [1] "-----"

## [1] "Inferred Issue: Dealing with my lender or servicer"

## [2] "Inferred Issue: Dealing with your lender or servicer"

## [3] "Inferred Issue: Struggling to repay your loan"

## [1] "-----"

## [1] "Original Issue: Dealing with my lender or servicer"

## [1] "-----"

## [1] "Complaint narrative: NAVIENT TOOK ABOUT EXTRA {\$110.00} WITH RESPECT TO THEIR OWN MAILED INVOICE AUTOMATICALLY FROM MY CHECKING ACCOUNT, \" SINCE XX/XX/2014 ''. THEY REFUSED TO REFUND THAT MONEY TO ME. I WROTE TO THEIR CEO AND GOT NO ANSWER. THIS IS A FRAUD. NAVIENT ALSO REPORTED 60/150 DAYS DELINQUENT IN FROM ONE LOAN PROGRAM TO ANOTHER LOAN PROGRAMONE STUDENT LOAN. NAVIENT ALSO REPORTED 90 DAYS DELINQUENT FOR CHANGING FROM

ONE LOAN PROGRAM TO ANOTHER LOAN PROGRAM"

## [1] "-----"

## [1] "Inferred Issue: Problem when making payments"

## [2] "Inferred Issue: Dealing with your lender or servicer"

## [3] "Inferred Issue: Dealing with my lender or servicer"

## [1] "-----"

## [1] "Original Issue: Dealing with my lender or servicer"

## [1] "-----"

## [1] "Complaint narrative: I have been placed on REPAY, and was informed that when I made a good faith payment that it had taken care of my payment in order to get my loans in a better payment option. I have been informed by Fedloan ( PHEAA ) now that I have to make another payment of the same amount because my \" good faith ' ' payment that was already made was satisfying the month previous to the payment, even though I was on a forbearance. This is not the first time that FedLoan has mistakenly informed me of regulations in regards to my loan repayment options. I am not sure that I will actually be forgiven at the end of the PSLF because it seems that they are n't sure of what their supposed to be doing."