

Assignment 8: Text Mining

Yiming Ge

12/7/2020

PART 1: Exploratory Data Analysis & Text Analytics load data

```
consumerRcomplaints<-read.csv("/Users/yiming/Downloads/Consumer_Complaints.csv",header = TRUE,na.strings
```

How many complaints have been generated?

```
library(quanteda)
```

```
## Package version: 2.1.2
```

```
## Parallel computing: 2 of 4 threads used.
```

```
## See https://quanteda.io for tutorials and examples.
```

```
##
```

```
## Attaching package: 'quanteda'
```

```
## The following object is masked from 'package:utils':
```

```
##
```

```
##      View
```

```
dim(consumerRcomplaints)
```

```
## [1] 257341      18
```

```
names(consumerRcomplaints)
```

```
## [1] "Date.received"      "Product"
## [3] "Sub.product"        "Issue"
## [5] "Sub.issue"          "Consumer.complaint.narrative"
## [7] "Company.public.response" "Company"
## [9] "State"              "ZIP.code"
## [11] "Tags"               "Consumer.consent.provided."
## [13] "Submitted.via"      "Date.sent.to.company"
## [15] "Company.response.to.consumer" "Timely.response."
## [17] "Consumer.disputed." "Complaint.ID"
```

```
head(consumerComplaints)
```

```
##      Date.received
## 1      01/01/2018
## 2      01/01/2018
## 3      01/01/2018
## 4      01/01/2018
## 5      01/01/2018
## 6      01/01/2018
##
##                                     Product
## 1                                     Debt collection
## 2                                     Mortgage
## 3 Credit reporting, credit repair services, or other personal consumer reports
## 4                                     Mortgage
## 5                                     Debt collection
## 6                                     Credit card or prepaid card
##
##                                     Sub.product
## 1                                     I do not know
## 2          Conventional home mortgage
## 3          Credit reporting
## 4          Conventional home mortgage
## 5          Mortgage debt
## 6 General-purpose credit card or charge card
##
##                                     Issue
## 1          Attempts to collect debt not owed
## 2          Struggling to pay mortgage
## 3 Problem with a credit reporting company's investigation into an existing problem
## 4          Incorrect information on your report
## 5          Attempts to collect debt not owed
## 6          Fees or interest
##
##                                     Sub.issue
## 1          Debt is not yours
## 2          <NA>
## 3 Difficulty submitting a dispute or getting information about a dispute over the phone
## 4          <NA>
## 5          Debt was already discharged in bankruptcy and is no longer owed
## 6          Charged too much interest
##
## 1
## 2
## 3
## 4
## 5
## 6 I took out a credit card from Capital One several years ago after experiencing a XXXX illness which
##
##                                     Company.public.response
## 1          <NA>
## 2          <NA>
## 3          <NA>
## 4          <NA>
## 5 Company has responded to the consumer and the CFPB and chooses not to provide a public response
## 6          <NA>
##
##                                     Company State ZIP.code      Tags
## 1          Viking Client Services      TX      761XX      <NA>
```

```
## 2 Alabama Housing Finance Authority AL 36869 <NA>
## 3 EQUIFAX, INC. GA 30339 <NA>
## 4 OCWEN FINANCIAL CORPORATION FL 32789 <NA>
## 5 OCWEN FINANCIAL CORPORATION FL 329XX Servicemember
## 6 CAPITAL ONE FINANCIAL CORPORATION WI <NA> <NA>
## Consumer.consent.provided. Submitted.via Date.sent.to.company
## 1 Consent provided Web 01/01/2018
## 2 Other Web 01/01/2018
## 3 Consent not provided Web 01/01/2018
## 4 Consent provided Web 01/01/2018
## 5 Consent provided Web 01/01/2018
## 6 Consent provided Web 01/01/2018
## Company.response.to.consumer Timely.response. Consumer.disputed. Complaint.ID
## 1 Closed with explanation Yes N/A 2768843
## 2 Untimely response No N/A 2768516
## 3 Closed with explanation Yes N/A 2768808
## 4 Closed with explanation Yes N/A 2768826
## 5 Closed with explanation Yes N/A 2768824
## 6 Closed with explanation Yes N/A 2768873
```

```
str(consumerRcomplaints)
```

```
## 'data.frame': 257341 obs. of 18 variables:
## $ Date.received : Factor w/ 365 levels "01/01/2018","01/02/2018",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Product : Factor w/ 9 levels "Checking or savings account",...: 4 6 3 6 4 2 4 6 4 6 ...
## $ Sub.product : Factor w/ 48 levels "Auto debt","CD (Certificate of Deposit)",...: 20 20 20 20 20 20 20 20 20 20 ...
## $ Issue : Factor w/ 81 levels "Advertising",...: 5 64 49 30 5 20 5 70 28 30 ...
## $ Sub.issue : Factor w/ 162 levels "Account information incorrect",...: 42 NA 56 NA 56 NA 56 NA 56 NA 56 ...
## $ Consumer.complaint.narrative: Factor w/ 110520 levels "- [ ] My business partner and I bought the ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Company.public.response : Factor w/ 9 levels "Company believes complaint caused principally by ...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Company : Factor w/ 3284 levels "1 STOP MONEY CENTERS, LLC",...: 3164 147 1081 1081 1081 1081 1081 1081 1081 1081 ...
## $ State : Factor w/ 61 levels "AA","AE","AK",...: 52 4 16 14 14 59 41 46 6 41 ...
## $ ZIP.code : Factor w/ 13317 levels "00601","00602",...: 9694 5693 4275 4750 4834 4834 4834 4834 4834 4834 ...
## $ Tags : Factor w/ 3 levels "Older American",...: NA NA NA NA 3 NA NA NA NA NA ...
## $ Consumer.consent.provided. : Factor w/ 5 levels "Consent not provided",...: 2 5 1 2 2 2 1 2 2 1 ...
## $ Submitted.via : Factor w/ 6 levels "Email","Fax",...: 6 6 6 6 6 6 6 6 6 6 ...
## $ Date.sent.to.company : Factor w/ 423 levels "01/01/2018","01/01/2019",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Company.response.to.consumer: Factor w/ 4 levels "Closed with explanation",...: 1 4 1 1 1 1 3 1 1 1 ...
## $ Timely.response. : Factor w/ 2 levels "No","Yes": 2 1 2 2 2 2 2 2 2 2 ...
## $ Consumer.disputed. : Factor w/ 1 level "N/A": 1 1 1 1 1 1 1 1 1 1 ...
## $ Complaint.ID : int 2768843 2768516 2768808 2768826 2768824 2768873 2768909 2768930 2768931 2768932 ...
```

Based on `dim()` output, we can say there have been 257341 complaints generated.

How many are unique or recurring?

```
unique(consumerRcomplaints$Product)
```

```
## [1] Debt collection
## [2] Mortgage
## [3] Credit reporting, credit repair services, or other personal consumer reports
## [4] Credit card or prepaid card
## [5] Money transfer, virtual currency, or money service
```

```
## [6] Checking or savings account
## [7] Payday loan, title loan, or personal loan
## [8] Student loan
## [9] Vehicle loan or lease
## 9 Levels: Checking or savings account ... Vehicle loan or lease
```

```
unique(consumerRcomplaints$Issue)
```

```
## [1] Attempts to collect debt not owed
## [2] Struggling to pay mortgage
## [3] Problem with a credit reporting company's investigation into an existing problem
## [4] Incorrect information on your report
## [5] Fees or interest
## [6] Trouble during payment process
## [7] Improper use of your report
## [8] Money was not available when promised
## [9] Managing an account
## [10] Problem with the payoff process at the end of the loan
## [11] Problem when making payments
## [12] Problem with fraud alerts or security freezes
## [13] Took or threatened to take negative or legal action
## [14] Problem with customer service
## [15] Written notification about debt
## [16] Closing an account
## [17] Charged fees or interest you didn't expect
## [18] Communication tactics
## [19] False statements or representation
## [20] Struggling to repay your loan
## [21] Other transaction problem
## [22] Other service problem
## [23] Other features, terms, or problems
## [24] Getting a loan
## [25] Dealing with your lender or servicer
## [26] Credit monitoring or identity theft protection services
## [27] Problem with a purchase shown on your statement
## [28] Problems at the end of the loan or lease
## [29] Unable to get your credit report or credit score
## [30] Advertising and marketing, including promotional offers
## [31] Getting a credit card
## [32] Confusing or missing disclosures
## [33] Struggling to pay your loan
## [34] Closing your account
## [35] Fraud or scam
## [36] Managing the loan or lease
## [37] Applying for a mortgage or refinancing an existing mortgage
## [38] Problem caused by your funds being low
## [39] Can't contact lender or servicer
## [40] Unexpected or other fees
## [41] Problem with a lender or other company charging your account
## [42] Threatened to contact someone or share information improperly
## [43] Closing on a mortgage
## [44] Struggling to pay your bill
## [45] Trouble using your card
## [46] Vehicle was repossessed or sold the vehicle
```

```
## [47] Opening an account
## [48] Loan payment wasn't credited to your account
## [49] Money was taken from your bank account on the wrong day or for the wrong amount
## [50] Getting a loan or lease
## [51] Managing, opening, or closing your mobile wallet account
## [52] Problem with a purchase or transfer
## [53] Unauthorized transactions or other transaction problem
## [54] Can't stop withdrawals from your bank account
## [55] Wrong amount charged or received
## [56] Problem with a company's investigation into an existing issue
## [57] Getting a line of credit
## [58] Trouble using the card
## [59] Was approved for a loan, but didn't receive the money
## [60] Problem getting a card or closing an account
## [61] Applying for a mortgage
## [62] Problem with additional add-on products or services
## [63] Lost or stolen check
## [64] Getting the loan
## [65] Confusing or misleading advertising or marketing
## [66] Incorrect exchange rate
## [67] Received a loan you didn't apply for
## [68] Vehicle was damaged or destroyed the vehicle
## [69] Identity theft protection or other monitoring services
## [70] Problem adding money
## [71] Lost or stolen money order
## [72] Advertising
## [73] Problem with overdraft
## [74] Excessive fees
## [75] Problem with cash advance
## [76] Credit limit changed
## [77] Property was sold
## [78] Property was damaged or destroyed property
## [79] Was approved for a loan, but didn't receive money
## [80] Problem with credit report or credit score
## [81] Overdraft, savings, or rewards features
## 81 Levels: Advertising ...
```

```
unique(consumerRcomplaints$Sub.issue)
```

```
## [1] Debt is not yours
## [2] <NA>
## [3] Difficulty submitting a dispute or getting information about a dispute over the phone
## [4] Debt was already discharged in bankruptcy and is no longer owed
## [5] Charged too much interest
## [6] Debt was result of identity theft
## [7] Credit inquiries on your report that you don't recognize
## [8] Account status incorrect
## [9] Fee problem
## [10] Their investigation did not fix an error on your report
## [11] Problem with personal statement of dispute
## [12] Problem during payment process
## [13] Threatened or suggested your credit would be damaged
## [14] Threatened to sue you for very old debt
## [15] Reporting company used your report improperly
```

[16] Funds not handled or disbursed as instructed
[17] Didn't receive enough information to verify debt
[18] Account information incorrect
[19] Can't close your account
[20] Frequent or repeated calls
[21] Funds not received from closed account
[22] Attempted to collect wrong amount
[23] Personal information incorrect
[24] Problem lowering your monthly payments
[25] Other problem
[26] Information belongs to someone else
[27] Problem using a debit or ATM card
[28] Banking errors
[29] Debt was paid
[30] Fraudulent loan
[31] Old information reappears or never goes away
[32] Trouble with how payments are being handled
[33] Billing dispute for services
[34] Didn't receive notice of right to dispute
[35] Credit card company isn't resolving a dispute about a purchase on your statement
[36] Problem related to refinancing
[37] Problem with fees
[38] Problem with rewards from credit card
[39] Received bad information about your loan
[40] Unexpected increase in interest rate
[41] Sued you without properly notifying you of lawsuit
[42] Card was charged for something you did not purchase with the card
[43] Problem with customer service
[44] Problem getting your free annual credit report
[45] Was not notified of investigation status or results
[46] Didn't receive advertised or promotional terms
[47] Card opened as result of identity theft or fraud
[48] Public record information inaccurate
[49] Investigation took more than 30 days
[50] Lender trying to repossess or disable the vehicle
[51] Don't agree with the fees charged
[52] Company closed your account
[53] Deposits and withdrawals
[54] You told them to stop contacting you, but they keep trying
[55] Called before 8am or after 9pm
[56] Problem with the interest rate
[57] Used obscene, profane, or other abusive language
[58] Can't get other flexible options for repaying your loan
[59] Information is missing that should be on the report
[60] Problem with fees charged
[61] Overdrafts and overdraft fees
[62] Confusing or misleading advertising about the credit card
[63] Transaction was not authorized
[64] Bounced checks or returned payments
[65] Fees charged for closing account
[66] Other problem getting your report or credit score
[67] Application denied
[68] Privacy issues
[69] Problem with paying off the loan

[70] Excess mileage, damage, or wear fees, or other problem after the lease is finish
[71] Impersonated attorney, law enforcement, or government official
[72] Talked to a third-party about your debt
[73] Unable to receive car title or other problem after the loan is paid off
[74] Confusing or misleading advertising
[75] Credit card company won't work with you while you're going through financial hardship
[76] Can't use card to make purchases
[77] Termination fees or other problem when ending the lease early
[78] Didn't receive terms that were advertised
[79] Billing problem
[80] Problem accessing account
[81] Keep getting calls about your loan
[82] Can't stop withdrawals from your account
[83] Can't temporarily delay making payments
[84] Confusing or missing disclosures
[85] Indicated you were committing crime by not paying debt
[86] Contacted your employer
[87] Cashing a check
[88] Loan balance remaining after the vehicle is repossessed and sold
[89] Need information about your loan balance or loan terms
[90] You never received your bill or did not know a payment was due
[91] Sent card you never applied for
[92] Problem with additional products or services purchased with the loan
[93] Seized or attempted to seize your property
[94] Non-sufficient funds and associated fees
[95] Charged for a purchase or transfer you did not make with the card
[96] Unable to open an account
[97] Problem with product or service terms changing
[98] Credit card company won't increase or decrease your credit limit
[99] Threatened to arrest you or take you to jail if you do not pay
[100] Problem canceling credit monitoring or identify theft protection service
[101] Delay in processing application
[102] Account sold or transferred to another company
[103] Add-on products and services
[104] Problem making or receiving payments
[105] Notification didn't disclose it was an attempt to collect a debt
[106] Credit denial
[107] Denied request to lower payments
[108] Received unsolicited financial product or insurance offers after opting out
[109] Account opened as a result of fraud
[110] Problem with balance transfer
[111] Problem with convenience check
[112] Problem after you declared or threatened to declare bankruptcy
[113] Loan sold or transferred to another company
[114] Trouble using the card to spend money in a store or online
[115] Money was taken from your account on the wrong day or for the wrong amount
[116] Trouble getting a working replacement card
[117] Contacted you after you asked them to stop
[118] Collected or attempted to collect exempt funds
[119] Information is incorrect
[120] Denied loan
[121] Trouble closing card
[122] Late or other fees
[123] Card company isn't resolving a dispute about a purchase or transfer

```

## [124] Changes in terms mid-deal or after closing
## [125] Confusing or misleading advertising or marketing
## [126] Problem using the card to withdraw money from an ATM
## [127] Problem getting a working replacement card
## [128] Deposits or withdrawals
## [129] Told you not to respond to a lawsuit they filed against you
## [130] Didn't receive services that were advertised
## [131] Trouble getting information about the card
## [132] Problem with direct deposit
## [133] High-pressure sales tactics
## [134] Report provided to employer without your written authorization
## [135] Overcharged for something you did purchase with the card
## [136] Problem while selling or giving up the vehicle
## [137] Problem with fees or penalties
## [138] Information that should be on the report is missing
## [139] Problem when attempting to purchase vehicle at the end of the lease
## [140] Filed for bankruptcy
## [141] Sued you in a state where you do not live or did not sign for the debt
## [142] Problem with cash advances
## [143] Confusing or misleading advertising about the card
## [144] Problem adding money
## [145] Received unwanted marketing or advertising
## [146] Changes in terms from what was offered or advertised
## [147] Overdraft charges
## [148] Trouble getting, activating, or registering a card
## [149] Was signed up for overdraft on card, but don't want to be
## [150] Don't want a card provided by your employer or the government
## [151] Problem with renewal
## [152] Credit card company forcing arbitration
## [153] Problem extending the lease
## [154] Problem with additional add-on products or services purchased with the loan
## [155] Trouble using the card to pay a bill
## [156] Problem with a trade-in
## [157] Problem with signing the paperwork
## [158] Contacted you instead of your attorney
## [159] Trouble using the card to send money to another person
## [160] Problem with a check written from your prepaid card account
## [161] Overcharged for a purchase or transfer you did make with the card
## [162] Qualified for a better loan than the one offered
## [163] Threatened to turn you in to immigration or deport you
## 162 Levels: Account information incorrect ...

```

Using `unique()` function, we found for different 9 products we totally have 81 unique issue complaints. If we count `sub.issue` as well, we have 163 different issue complaints in total.

Using “`Consumer.complaint.narrative`”, what can you say about the type of complaints in this report?

```

consumerRcomplaints$Consumer.complaint.narrative <- gsub("'", "", consumerRcomplaints$Consumer.complaint.narrative)
consumerRcomplaints$Consumer.complaint.narrative <- gsub("[:punct:]", " ", consumerRcomplaints$Consumer.complaint.narrative)
consumerRcomplaints$Consumer.complaint.narrative <- gsub("[:cntrl:]", " ", consumerRcomplaints$Consumer.complaint.narrative)
consumerRcomplaints$Consumer.complaint.narrative <- gsub("^[:space:]+$", "", consumerRcomplaints$Consumer.complaint.narrative)
consumerRcomplaints$Consumer.complaint.narrative <- gsub("[:space:]+$", "", consumerRcomplaints$Consumer.complaint.narrative)
consumerRcomplaints$Consumer.complaint.narrative <- gsub("[^a-zA-Z -]", " ", consumerRcomplaints$Consumer.complaint.narrative)
consumerRcomplaints$Consumer.complaint.narrative <- tolower(consumerRcomplaints$Consumer.complaint.narrative)

```



```
#create a corpus with metadata
comcorpus<- corpus(consumeRcomplaints$Consumer.complaint.narrative)
```

```
## Warning: NA is replaced by empty string
```

```
#explore the corpus
summary(comcorpus) #summary of corpus
```

```
## Corpus consisting of 257341 documents, showing 100 documents:
```

```
##
##      Text Types Tokens Sentences
##      text1      30      38         1
##      text2       0       0         0
##      text3       0       0         0
##      text4     110     216         1
##      text5     154     383         1
##      text6     167     421         1
##      text7       0       0         0
##      text8     253     683         1
##      text9      57     108         1
##     text10       0       0         0
##     text11      45      59         1
##     text12     106     209         1
##     text13       0       0         0
##     text14       0       0         0
##     text15       0       0         0
##     text16       0       0         0
##     text17       0       0         0
##     text18      66      97         1
##     text19     107     179         1
##     text20       0       0         0
##     text21       0       0         0
##     text22     274     656         1
##     text23     152     285         1
##     text24       0       0         0
##     text25       0       0         0
##     text26      92     181         1
##     text27      98     233         1
##     text28     100     195         1
##     text29       0       0         0
##     text30       0       0         0
##     text31       0       0         0
##     text32       0       0         0
##     text33       0       0         0
##     text34      65     127         1
##     text35      79     118         1
##     text36      57     100         1
##     text37       0       0         0
##     text38       0       0         0
##     text39       0       0         0
##     text40       0       0         0
##     text41      74     126         1
##     text42      99     233         1
```

##	text43	0	0	0
##	text44	0	0	0
##	text45	0	0	0
##	text46	81	135	1
##	text47	0	0	0
##	text48	24	39	1
##	text49	42	51	1
##	text50	121	252	1
##	text51	0	0	0
##	text52	54	78	1
##	text53	0	0	0
##	text54	67	97	1
##	text55	40	64	1
##	text56	146	361	1
##	text57	0	0	0
##	text58	40	47	1
##	text59	0	0	0
##	text60	117	237	1
##	text61	145	297	1
##	text62	0	0	0
##	text63	0	0	0
##	text64	0	0	0
##	text65	301	919	1
##	text66	0	0	0
##	text67	0	0	0
##	text68	7	8	1
##	text69	0	0	0
##	text70	0	0	0
##	text71	0	0	0
##	text72	36	61	1
##	text73	0	0	0
##	text74	152	349	1
##	text75	29	65	1
##	text76	80	119	1
##	text77	83	175	1
##	text78	77	126	1
##	text79	52	68	1
##	text80	17	18	1
##	text81	0	0	0
##	text82	63	114	1
##	text83	0	0	0
##	text84	79	124	1
##	text85	201	541	1
##	text86	302	900	1
##	text87	0	0	0
##	text88	58	85	1
##	text89	63	99	1
##	text90	67	97	1
##	text91	165	601	1
##	text92	57	74	1
##	text93	25	29	1
##	text94	0	0	0
##	text95	102	196	1
##	text96	92	165	1

```
##   text97      0      0      0
##   text98      0      0      0
##   text99     98    232      1
##   text100     0      0      0
```

```
#create document feature matrix from clean corpus + stem
dfm.stem<- dfm(comcorpus,
               remove = stopwords("english"),
               verbose=TRUE,
               stem=TRUE)
```

```
## Creating a dfm from a corpus input...
```

```
## ...lowercasing
```

```
## ...found 257,341 documents, 51,512 features
```

```
## ...removed 125 features
```

```
## ...stemming types (English)
```

```
## ...complete, elapsed time: 29.1 seconds.
```

```
## Finished constructing a 257,341 x 34,435 sparse dfm.
```

```
topfeatures(dfm.stem, n=50)
```

```
##      xxxx      xx account  credit  report payment  call  inform  loan  receiv
## 1307028 506963 206225 204217 177010 124563 110732 99813 84634 78042
##      time compani  debt  bank  told  card  month  state  day  letter
## 74291 70885 69368 62923 59418 58133 55445 54919 54709 54672
## request disput  pay collect contact never  get  sent  ask  remov
## 54222 50532 50421 50125 48823 48718 47508 47235 46913 45180
##      file  charg  back  provid  also  can  amount  date  servic  one
## 44942 44334 43105 42243 41819 41686 41482 41274 41038 40949
##      check  number  year  paid  said  made  due  balanc  phone  make
## 40577 40489 39179 38514 37885 36612 35904 35474 34070 33931
```

Since the top features are related to ‘account’, ‘credit’, ‘payment’, ‘loan’, ‘debt’, ‘bank’, we can get the result that most complaints are about their issue of credit including payment, loan and debt. The xxxx and xx in features refer to the personal information.

What type of product issues & complaints are the most frequent?

```
sort(table(consumeRcomplaints$Product), decreasing=TRUE)
```

```
##
## Credit reporting, credit repair services, or other personal consumer reports
##                                     111632
##                                     Debt collection
##                                     51195
##                                     Mortgage
##                                     24577
##                                     Credit card or prepaid card
```

```
##                                     24251
##                               Checking or savings account
##                                     21211
##                               Student loan
##                                     8780
##                               Vehicle loan or lease
##                                     5892
##                               Money transfer, virtual currency, or money service
##                                     5434
##                               Payday loan, title loan, or personal loan
##                                     4369
```

```
sort(table(consumerRcomplaints$Issue),decreasing = TRUE)
```

```
##
##                               Incorrect information on your report
##                                     70248
## Problem with a credit reporting company's investigation into an existing problem
##                                     25844
##                               Attempts to collect debt not owed
##                                     22501
##                               Improper use of your report
##                                     13675
##                               Managing an account
##                                     13416
##                               Written notification about debt
##                                     12354
##                               Trouble during payment process
##                                     11411
##                               Struggling to pay mortgage
##                                     8656
##                               Communication tactics
##                                     6026
##                               Problem with a purchase shown on your statement
##                                     5879
##                               Dealing with your lender or servicer
##                                     5731
##                               Took or threatened to take negative or legal action
##                                     4638
##                               False statements or representation
##                                     4320
##                               Other features, terms, or problems
##                                     3142
##                               Fees or interest
##                                     3138
##                               Problem when making payments
##                                     2831
##                               Unable to get your credit report or credit score
##                                     2337
##                               Getting a credit card
##                                     2268
##                               Closing an account
##                                     2257
##                               Struggling to repay your loan
```

##		2153
##	Managing the loan or lease	
##		2088
##	Struggling to pay your loan	
##		1936
##	Applying for a mortgage or refinancing an existing mortgage	
##		1933
##	Opening an account	
##		1929
##	Closing your account	
##		1868
##	Problem caused by your funds being low	
##		1812
##	Problem with fraud alerts or security freezes	
##		1781
##	Closing on a mortgage	
##		1641
##	Problem with a lender or other company charging your account	
##		1616
##	Fraud or scam	
##		1607
##	Advertising and marketing, including promotional offers	
##		1458
##	Threatened to contact someone or share information improperly	
##		1356
##	Charged fees or interest you didn't expect	
##		1243
##	Credit monitoring or identity theft protection services	
##		1209
##	Other transaction problem	
##		1156
##	Money was not available when promised	
##		1105
##	Problems at the end of the loan or lease	
##		1102
##	Trouble using your card	
##		906
##	Getting a loan or lease	
##		647
##	Struggling to pay your bill	
##		588
##	Problem with a purchase or transfer	
##		485
##	Unexpected or other fees	
##		410
##	Problem with the payoff process at the end of the loan	
##		382
##	Trouble using the card	
##		359
##	Unauthorized transactions or other transaction problem	
##		345
##	Getting the loan	
##		311
##	Managing, opening, or closing your mobile wallet account	

##		308
##	Other service problem	
##		301
##	Problem with a company's investigation into an existing issue	
##		250
##	Applying for a mortgage	
##		221
##	Problem with customer service	
##		216
##	Getting a line of credit	
##		176
##	Getting a loan	
##		173
##	Problem getting a card or closing an account	
##		161
##	Confusing or missing disclosures	
##		158
##	Identity theft protection or other monitoring services	
##		134
##	Wrong amount charged or received	
##		130
##	Confusing or misleading advertising or marketing	
##		112
##	Problem with additional add-on products or services	
##		107
##	Can't contact lender or servicer	
##		103
##	Received a loan you didn't apply for	
##		101
##	Can't stop withdrawals from your bank account	
##		87
##	Loan payment wasn't credited to your account	
##		63
##	Vehicle was repossessed or sold the vehicle	
##		57
##	Money was taken from your bank account on the wrong day or for the wrong amount	
##		56
##	Lost or stolen money order	
##		46
##	Advertising	
##		42
##	Lost or stolen check	
##		41
##	Problem adding money	
##		39
##	Was approved for a loan, but didn't receive the money	
##		38
##	Incorrect exchange rate	
##		24
##	Problem with cash advance	
##		19
##	Credit limit changed	
##		18
##	Excessive fees	

```
##
##                                     18
##                                     Vehicle was damaged or destroyed the vehicle
##                                     18
##                                     Problem with overdraft
##                                     10
##                                     Was approved for a loan, but didn't receive money
##                                     7
##                                     Overdraft, savings, or rewards features
##                                     4
##                                     Problem with credit report or credit score
##                                     2
##                                     Property was sold
##                                     2
##                                     Property was damaged or destroyed property
##                                     1
```

We can see most complaints issues are Incorrect information on Credit reporting, credit repair services, or other personal consumer reports.

Complete a sentiment analysis for all the types of complaint submissions observed during this year.

```
#Sentiment Analysis
mydict <- dictionary(list(payment = c("payment","pay","paid","sent"),
                                receive =c("receiv*","collect","get"),
                                credit=c("credit"),
                                loan = c("loan","debt")
                                ))
#differentiate the complaints based on payment, receive, credit and loan.

dfm.sentiment <- dfm(comcorpus,
                    remove = stopwords("english"),
                    verbose=TRUE,
                    dictionary = mydict,
                    stem=FALSE)
```

```
## Creating a dfm from a corpus input...

## ...lowercasing

## ...found 257,341 documents, 51,512 features

## ...applying a dictionary consisting of 4 keys
## ...removed 0 features
## ...complete, elapsed time: 25.9 seconds.
## Finished constructing a 257,341 x 4 sparse dfm.
```

```
topfeatures(dfm.sentiment)
```

```
## payment  credit    loan receive
## 203710  200621  133908 125284
```

```
head(dfm.sentiment)
```

```
## Document-feature matrix of: 6 documents, 4 features (58.3% sparse).
##           features
## docs    payment receive credit loan
## text1      1      1      0      1
## text2      0      0      0      0
## text3      0      0      0      0
## text4      0      0      5      0
## text5      0      2      2      2
## text6      9      0      7      3
```

PART 2: Steve Jobs Commencemnt Speech

```
library(readtext)
jobsSpeech<-readtext("https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BADData/JobsS
```

```
jobsSpeech$text <- gsub("'", "", jobsSpeech$text) # remove apostrophes
jobsSpeech$text <- gsub("[[:punct:]]", " ", jobsSpeech$text) # replace punctuation with space
jobsSpeech$text <- gsub("[[:cntrl:]]", " ", jobsSpeech$text) # replace control characters with space
jobsSpeech$text <- gsub("^[:space:]+", "", jobsSpeech$text) # remove whitespace at beginning of docum
jobsSpeech$text <- gsub("[[:space:]]+$", "", jobsSpeech$text) # remove whitespace at end of documents
jobsSpeech$text <- gsub("[^a-zA-Z -]", " ", jobsSpeech$text) # allows only letters
jobsSpeech$text <- tolower(jobsSpeech$text) # force to lowercase
```

```
#create a corpus with metadata
help(corpus)
jobscorpus<- corpus(jobsSpeech$text)
#explore the corpus
summary(jobscorpus) #summary of corpus
```

```
## Corpus consisting of 1 document, showing 1 document:
##
##   Text Types Tokens Sentences
## text1    687    2270         1
```

```
#create document feature matrix from clean corpus
dfm.jobs<- dfm(jobscorpus,
               remove = stopwords("english"),
               verbose=TRUE,
               stem=FALSE)
```

```
## Creating a dfm from a corpus input...
```

```
## ...lowercasing
```

```
## ...found 1 document, 687 features
```

```
## ...removed 97 features
```

```
## ...complete, elapsed time: 0.025 seconds.
```

```
## Finished constructing a 1 x 590 sparse dfm.
```



```
topfeatures(dfm.jobs, n=50)
```

```
##      t      life  college      s      one      years      apple
##      19      17      14      11      9      9      9
##      never      just  dropped      months      later      looking      don
##      8      8      7      7      7      7      7
##      want      first everything      didn      great      know      now
##      6      6      6      6      6      6      6
##      death      stay      ever      started      decided      parents      work
##      6      6      5      5      5      5      5
##      best      every      get      way      can      something      love
##      5      5      5      5      5      5      5
##      company      next      today      ve      story      dots      put
##      5      5      4      4      4      4      4
##      got      course      found      go      going      made      loved
##      4      4      4      4      4      4      4
##      turned
##      4
```

```
# create a custom dictionary
```

```
swlist = c("t","s","one","just","don","didn","now","can","ve","dots","got","get","put","something","made")
dfm.jobs<- dfm(jobscorpus,
               remove = c(swlist,stopwords("english")),
               verbose=TRUE,
               stem=FALSE)
```

```
## Creating a dfm from a corpus input...
```

```
## ...lowercasing
```

```
## ...found 1 document, 687 features
```

```
## ...removed 112 features
```

```
## ...complete, elapsed time: 0.026 seconds.
```

```
## Finished constructing a 1 x 575 sparse dfm.
```

```
topfeatures(dfm.jobs, n=50)
```

```
##      life  college  years      apple      never      dropped      months
##      17      14      9      9      8      7      7
##      later      looking      want      first everything      great      know
##      7      7      6      6      6      6      6
##      death      stay      ever      started      decided      parents      work
##      6      6      5      5      5      5      5
##      best      every      way      love      company      next      today
##      5      5      5      5      5      5      4
##      story      course      found      go      going      loved      turned
##      4      4      4      4      4      4      4
##      let      even      computer      connect      fired      heart      find
##      4      4      4      4      4      4      4
##      die      world      graduated      told      closest      tell      three
```

```
##          4          3          3          3          3          3          3
##      drop
##          3
```

```
#exploration in context
kwic(jobscorpus, "life", 4)
```

```
##
## [text1, 51]      three stories from my | life |
## [text1, 316]     to do with my | life |
## [text1, 345]     had saved their entire | life |
## [text1, 603]    practical application in my | life |
## [text1, 771]    something your gut destiny | life |
## [text1, 790]     the difference in my | life |
## [text1, 811]     to do early in | life |
## [text1, 955]     of my entire adult | life |
## [text1, 1133]    creative periods of my | life |
## [text1, 1257]    patient needed it sometimes | life |
## [text1, 1321]    large part of your | life |
## [text1, 1472]    last day of my | life |
## [text1, 1530]    the big choices in | life |
## [text1, 1933]    single best invention of | life |
## [text1, 1936]    of life it is | life |
## [text1, 1997]    living someone else s | life |
## [text1, 2099]    he brought it to | life |
##
## that s it no
## and no idea how
## so i decided to
## but years later when
## karma whatever this approach
## my second story is
## woz and i started
## was gone and it
## during the next five
## hits you in the
## and the only way
## would i want to
## because almost everything all
## it is life s
## s change agent it
## don t be trapped
## with his poetic touch
```

```
kwic(jobscorpus, "college", 3)
```

```
##
## [text1, 25] never graduated from | college | truth be told
## [text1, 39] gotten to a | college | graduation today i
## [text1, 74] out of reed | college | after the first
## [text1, 115] a young unwed | college | graduate student and
## [text1, 137] be adopted by | college | graduates so everything
## [text1, 221] never graduated from | college | and that my
```

```
## [text1, 257]      someday go to | college | and years later
## [text1, 265]      did go to | college | but i naively
## [text1, 271]      naively chose a | college | that was almost
## [text1, 292]      spent on my | college | tuition after six
## [text1, 321]      no idea how | college | was going to
## [text1, 493]      one example reed | college | at that time
## [text1, 648]      single course in | college | the mac would
## [text1, 721]      i was in | college | but it was
```

```
kwic(jobscorpus , "apple", 3)
```

```
##
## [text1, 816]      and i started | apple | in my parents
## [text1, 830]      and in years | apple | had grown from
## [text1, 883]      started well as | apple | grew we hired
## [text1, 1051]     of events at | apple | had not changed
## [text1, 1089]     getting fired from | apple | was the best
## [text1, 1194]     turn of events | apple | bought next i
## [text1, 1200]     i returned to | apple | and the technology
## [text1, 1213]     the heart of | apple | s current renaissance
## [text1, 1243]     been fired from | apple | it was awful
```

```
#specifying a correlation limit of 0.1
library(tm)
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

```
## The following objects are masked from 'package:quanteda':
##
##      meta, meta<-
```

```
##
## Attaching package: 'tm'
```

```
## The following objects are masked from 'package:quanteda':
##
##      as.DocumentTermMatrix, stopwords
```

```
dfm.tm<-convert(dfm.jobs, to="tm")
findAssocs(dfm.tm,
            'life',
            corlimit=0.1)
```

```
## $life
## numeric(0)
```

```
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
set.seed(88)    #keeps cloud' shape fixed  
dark2 <- brewer.pal(8, "Set1")  
freq<-topfeatures(dfm.jobs, n=500)  
  
wordcloud(names(freq),  
          freq, max.words=100,  
          scale=c(3, .1),  
          colors=brewer.pal(9, "Set1"))
```

