Twitter sentiment analysis 1

Ziyad Abdulaziz

13/07/2020

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#connecting to twitter

consumer_key <- "jHHSo1xVw0mhp1cyFTac86nK3"
consumer_secret <- "sEMM8ed0DVd8j9DaKnta8UkzEaMLdzwoYAet8mf3MWiitVewQp"
access_token <- "244645993-QrRcQWJXOb5SGBI8W6jxIL0uYMk686TIIGKn9a7Q"
access_secret <- "mZ1aa2TsvWDqfBVMU2GUEHUiglTjkmQQ7KUimGFGQOilP"
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

```
## [1] "Using direct authentication"
```

```
#Obtaining the first set of tweets
peter <- userTimeline("PeterLBrandt", n=1000)
josephburns <- userTimeline("SJosephBurns", n=1000)
elerian <- userTimeline("elerianm", n = 1000)
ibd <- userTimeline("IBDinvestors", n = 1000)
bespoke <- userTimeline("bespokeinvest", n = 1000)
marketwatch <- userTimeline("MarketWatch", n = 1000)
appletweet <- searchTwitter("$AAPL", n=1000, lang = 'en')
msfttweet <- searchTwitter("$MSFT", n = 1000,lang = 'en')
amazontweet <- searchTwitter("$AMZN", n=1000, lang = 'en')
googletweet <- searchTwitter("$GOOG", n = 1000, lang = 'en')
nastweet <- searchTwitter("$NASDAQ", n = 1000, lang = 'en')
```

```
## Warning in doRppAPICall("search/tweets", n, params = params, retryOnRateLimit =
## retryOnRateLimit, : 1000 tweets were requested but the API can only return 790
```

```
#Transforming to dataframe
tweets1 <- tbl_df(map_df(c(peter, josephburns, elerian, ibd, bespoke, marketwatch, appletweet, msfttwee
```

```
## Warning: `tbl_df()` is deprecated as of dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```r
#Saving into CSV
write.csv(tweets1, file = "tweets1.csv", row.names = FALSE)
```

```r
#setworking directory and read in data

setwd("C:/Users/ziyad/Desktop/Data Analytics capstone")
tweets1 <- read.csv("tweets1.csv")
```

```r
#twitter data cleanup for dataset1

twittercorpus1 <- Corpus(VectorSource(tweets1$text))
inspect(twittercorpus1[1:10])
```

```
## <<SimpleCorpus>>
## Metadata:  corpus specific: 1, document level (indexed): 0
## Content:  documents: 10
##
##  [1] @eo1989 I agree. The Porsche situation was a bird of a different color.
##  [2] @TMPTRADING I can relate. This is how I have lived for 45 years.d
##  [3] $TSLA I am personally long Tesla, and not in any hurry to be flat. But a part of me wonders if
##  [4] Thanks Barry. This is very insightful. In the meanwhile, bears will be bears. https://t.co/r9M3
##  [5] Silver undergoing completion of major long-term chart bottom. Targets are 2610 and 2767. $SI_F
##  [6] @CryptoJamesG @RobinhoodApp I'll never tell
##  [7] Congrats to all the @RobinhoodApp\n Gen Ms and Gen Zs who caught this wild beast. With the volu
##  [8] Watching grandson play LL baseball https://t.co/r81ORDpblB
##  [9] @JohanDXB LOL.\nDoes your broker have the authority to cover your position without your knowled
## [10] @TAwithBA @RobinhoodApp Nope. <U+0001F44E> many mid cap tech stocks going parabolic
```

```r
twittercorpus1 <- tm_map(twittercorpus1, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1, content_transformer(tolower)):
## transformation drops documents
```

```r
twittercorpus1 <- tm_map(twittercorpus1, removeWords, stopwords("en"))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1, removeWords, stopwords("en")):
## transformation drops documents
```

```r
twittercorpus1 <- tm_map(twittercorpus1, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1, removeNumbers): transformation
## drops documents
```

```r
twittercorpus1 <- tm_map(twittercorpus1, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1, removePunctuation):
## transformation drops documents
```

```
removeURLhttp1 <- function(x) gsub ("http[[:alnum:]]*", "", x)
twittercorpus1 <- tm_map(twittercorpus1, content_transformer(removeURLhttp1))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1,
## content_transformer(removeURLhttp1)): transformation drops documents
```

```
removeURLedua1 <- function(x) gsub ("edua[[:alnum:]]*", "", x)
twittercorpus1 <- tm_map(twittercorpus1, content_transformer(removeURLedua1))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1,
## content_transformer(removeURLedua1)): transformation drops documents
```

```
removeNonAscii <- function(x) textclean::replace_non_ascii(x)
twittercorpus1 <- tm_map (twittercorpus1,content_transformer(removeNonAscii))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1,
## content_transformer(removeNonAscii)): transformation drops documents
```

```
twittercorpus1 <- tm_map(twittercorpus1, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1, stripWhitespace): transformation
## drops documents
```

```
twittercorpus1 <- tm_map(twittercorpus1, removeWords, c('cnbc','aapl','goog','msft','amzn','appl','appl
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1, removeWords, c("cnbc", "aapl", :
## transformation drops documents
```

```
mystopwords1 <- c(stopwords("en"), 'cnbc','aapl','goog','msft','amzn','appl','apple','googl','microsoft
inspect(twittercorpus1[1:10])
```

```
## <<SimpleCorpus>>
## Metadata:  corpus specific: 1, document level (indexed): 0
## Content:  documents: 10
##
##  [1] eo agree porsche situation bird different color
##  [2] tmptrading can relate lived yearsd
##  [3]  personally long tesla hurry flat part wonders volkswagen vwaga wi...
##  [4] thanks barry insightful meanwhile bears will bears
##  [5] silver undergoing completion major longterm chart bottom targets sif slv
##  [6] cryptojamesg robinhoodapp never tell
##  [7] congrats robinhoodapp gen ms gen zs caught wild beast volume spikes confirma...
##  [8] watching grandson  ll baseball
##  [9] johandxb lol broker authority cover position without knowledge order manage risk curious
## [10] tawithba robinhoodapp nope ufe many mid cap tech stocks going parabolic
```

```
#term document matrix for corpus 1

twittercorpus1 <- tm_map(twittercorpus1, stemDocument)
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus1, stemDocument): transformation
## drops documents
```

```
dtm1 <- TermDocumentMatrix(twittercorpus1)
dtm1
```

```
## <<TermDocumentMatrix (terms: 8972, documents: 7256)>>
## Non-/sparse entries: 65011/65035821
## Sparsity           : 100%
## Maximal term length: 38
## Weighting          : term frequency (tf)
```

```
termmatrix1 <- as.matrix(dtm1)
termmatrix1[1:10, 1:20]
```

```
##           Docs
## Terms      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
##   agre     1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  1  0  0  0
##   bird     1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   color    1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   differ   1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   porsch   1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   situat   1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   can      0 1 0 0 0 0 0 0 0  0  0  0  0  0  1  0  0  0  0  0
##   live     0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   relat    0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   tmptrade 0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
```

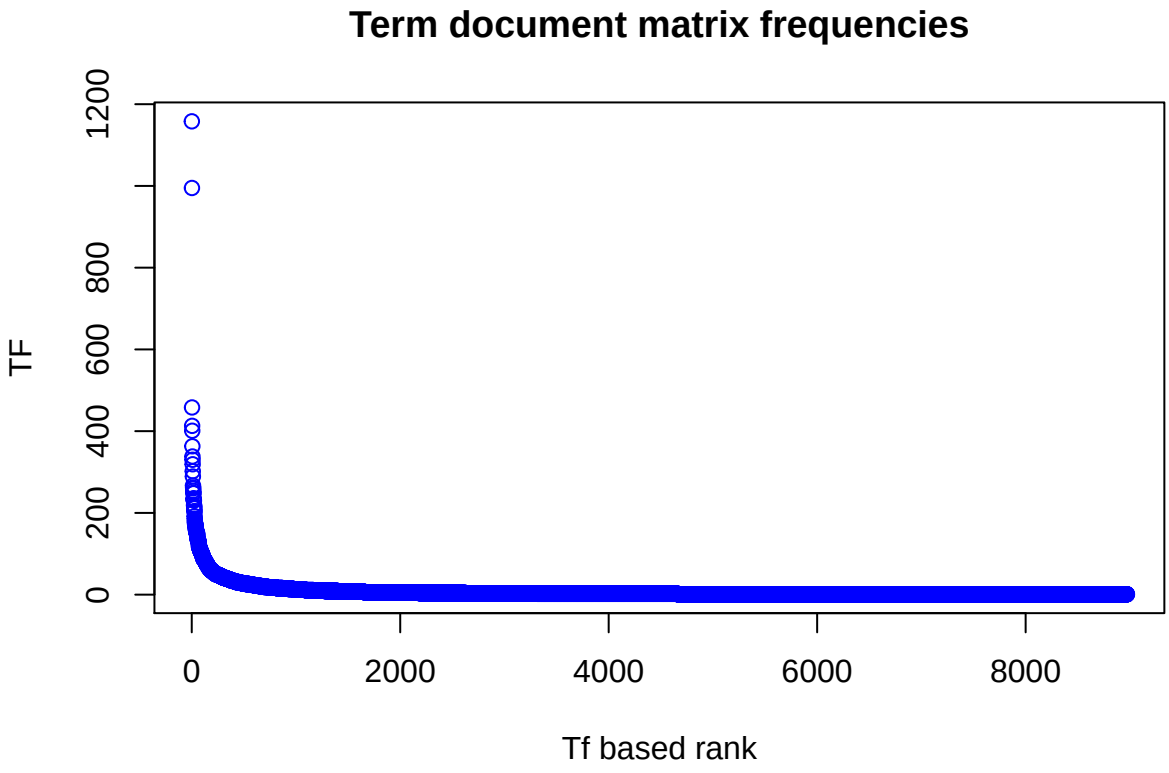
```
freq=rowSums(as.matrix(termmatrix1))
head(freq,10)
```

```
##     agre     bird    color   differ   porsch   situat      can     live
##       26        2        2       34        2        7      172       87
##    relat tmptrade
##       26        1
```

```
tail(freq,10)
```

```
##  andyheyward       cle...          knc       affect manufactu...          toe
##            1            1            2            1            1            1
##         vest       cleari    nasdaqwkh      thunder
##            1            1            1            1
```

```
plot(sort(freq, decreasing = T),col="blue",main="Term document matrix frequencies", xlab="Tf based rank
```




```
tail(sort(freq),n=10)
```

```
##   btc today   now   spx stock   new  nflx trade   spi   ...
##   302   319   330   338   363   401   413   458   995  1158
```

```
#detailed term frequency barplot

bp1 <- rowSums(termmatrix1)
bp1 <- subset(bp1, bp1>=25)
bp1
```

```
##             agre           differ              can             live
##               26               34              172               87
##            relat             long           person            tesla
##               26              139               30               97
##           wonder             bear            thank             will
##               31               47               78              260
##            chart          complet            major           silver
##              254               28               48               25
##              slv           target            never             tell
##               31               97               63               39
##            volum            watch              lol            manag
##               46               60               26               58
##            order            posit             risk              cap
##               49               80               66               91
```

```
##            mani           stock            tech           pleas
##              69             363             217              38
##             yes             ...           total          trader
##              27            1158              34             130
##          remain            c...            find             get
##              27              31              46             214
##            hous            just            need             way
##              28             219              86              60
##           great            make             one           trend
##              99             116             170              57
##             bet            done            o...             buy
##              28              27              39             232
##            keep           money          second            take
##              50             114              34             103
##             bro            even            lift             big
##              26              94              27             192
##         correct            love          school             see
##              28              48              26             180
##           peopl            said             say           trade
##             100              42             173             458
##           anoth           short           becom            like
##             111             170              31             236
##         possibl         raoulgmi            stay            year
##              28              36              33             232
##            home            mean            real            hope
##              30              38              53              43
##        interest            want          declin            drop
##              72              84              27              39
##             end           littl           might            note
##             137              26              32              56
##             top           happen            t...             two
##             248              40              65              42
##             yet          higher         twitter            well
##              34             115              41              62
##            work           small           futur             btc
##             152              25             115             302
##             eth            gain            near            bull
##             267             151              47              34
##           china            game          market             new
##              63              31             131             401
##          bought            sold             set           still
##              33              44              84             107
##          believ           tweet            file            open
##              42              33              46              82
##          profit             tax             etf           group
##              74              84              28              43
##            list             use            best           start
##              61              72              90              97
##            last         thought            term          global
##             132              28              32              65
##             dow           sampp             day           right
##             204             131             262             127
##            time           wrong             bad          recent
##             230              29              30              42
```

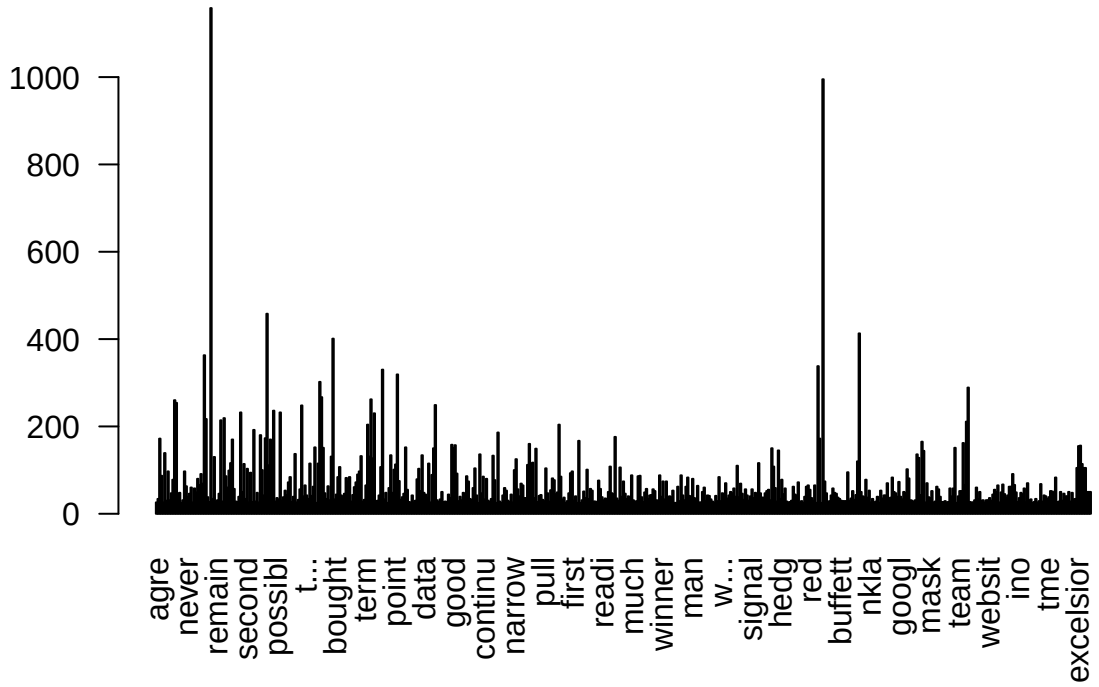
```
##           share             now             hey         pattern
##             107             330              33              48
##       robinhood           thing            call            made
##              25              59             134              48
##           point          expect           today             may
##             101             113             319              75
##         increas             ask            feel            look
##              32              38              46             152
##             lot          massiv          reason         bullish
##              55              27              26              42
##           india         support            grow            next
##              26              30              79             103
##            stop            back            data           video
##              52             134              48              44
##            issu            juli              p...            post
##              37             115              26              89
##          report            week          repres         surpris
##             150             249              31              27
##          demand             let         student         histori
##              33              50              27              28
##             per         special           studi            come
##              27              44              36             158
##            good           covid            free            lose
##             151             157              92              29
##         perform          presid           don't            give
##              39              31              48              39
##            news             job            idea         opinion
##              86              74              34              32
##          strong             hit         strategi            life
##              59             104              42              25
##           trump         everyon         continu        contract
##             136              47              85              33
##         product           creat            hard          realli
##              80              30              27              34
##            gold           month         comment           price
##             133              77              25             186
##          resist            step            sure           world
##              28              25              42              59
##           worth       yesterday          includ            line
##              53              30              27              44
##          narrow           close          record         current
##              26             100             125              56
##             gcf            hold            june            sale
##              40              69              65              36
##       sentiment         compani         million            wait
##              36             112             160              40
##            full           everi          invest          welcom
##             117              36             149              27
##            must            read            pull            loan
##              41              43              32              29
##            earn            head            huge            loss
##             104              47              28              54
##          artist            know         financi           offer
##              81              77              45              49
```

```
##         high       action       friday        learn
##          204           85           33           37
##         past     volatil          nqf        think
##           25           29           47           93
##        first     outsid        entri        swing
##           97           30           40           25
##          amp       enough         noth         spot
##          167           25           31           51
##        alway         sell         sign       number
##           36          101           34           57
##          run         zone     question       actual
##           53           28           28           27
##      economi         also        readi           b...
##           76           57           38           31
##         deal        earli         book          put
##           35           34           50          108
##       monday     american         case        daili
##           47           45          176           35
##        three         move         debt       follow
##           30          106           39           74
##        state        retir         word     technolog
##           46           34           40           33
##         much        death        quick       financ
##           88           31           25           29
##        check     analysi      technic           s...
##           86           87           38           27
##         hour        level        reach        claim
##           49           57           34           42
##         link       member        break          cut
##           43           56           52           31
##     opportun          win       winner          gld
##           30           88           39           74
##      generat         sinc         view         soon
##           33           74           37           40
##         rate          ath     momentum       pretti
##           38           53           25           25
##        ralli         ever          fed         rais
##           62           29           88           42
##        lower       better        offic         area
##           31           62           83           41
##          man        updat         york          tri
##           34           80           36           30
##         fund     everyth     california         nice
##           64           26           26           50
##        indic        clear         lead         surg
##           60           29           42           41
##          add        pick      elonmusk      analyst
##           33           28           25           41
##         wave        cost           w...         fall
##           29           84           29           47
##        smart        busi        chang          gap
##           30           70           35           42
##         rule      almost         show      biggest
##           50           47           58           39
```

```
##          option          equiti            rise          econom
##             110              38              69              46
##           cloud            miss           activ            amaz
##              32              56              42              30
##          signal             low            seem          public
##              40              56              36              48
##            test          return           minut            valu
##              41             116              47              37
##         premium        trillion           index          experi
##              37              47              52              55
##           power      coronavirus             ndx         billion
##              31             150             108              59
##            cash         announc            hedg          revers
##              28             145              34              78
##          reopen             via           upsid          import
##              45              58              28              26
##            went           phone             far             dip
##              27              30              62              44
##             oil             vix          season          health
##              42              72              25              29
##           night            wall            plan             got
##              25              39              62              65
##             red           green           bounc             box
##              55              36              30              65
##         alreadi             spx             esf        intraday
##              25             338             172              25
##             spi            morn            turn           estim
##             995              74              47              25
##           focus          growth          pandem            push
##              29              42              58              48
##            save           ahead         buffett          warren
##              46              37              34              30
##      candlestick          cheat           sheet          youtub
##              29              29              29              95
##        trendspid         netflix             bac             dal
##              30              35              52              32
##             jnj             jpm            nflx         florida
##              43             120             413              50
##          averag          diverg        investor           uufef
##              40              38              78              37
##            nkla          vaccin       portfolio           court
##              53              26              34              26
##          jobless           uber             ufa         chamath
##              28              39              25              53
##           march            key          introduc         weekend
##              35              42              42              70
##         tomorrow           face           ubufef        antitrust
##              33              37              83              50
##             ceo       facebook           googl           harmon
##              47              39              73              39
##          quarter           bynd             dia             dis
##              33              50              48             102
##             iwm           bank            warn   rampcapitalllc
##              81              31              30              27
```

```
##             snap             twtr             join           spread
##               31              136              128               43
##             flow            click          largest             intc
##              165              144               26               70
##             mask          appreci             djia             bond
##               38               36               52               26
##              ufc          discuss           sector      stockmarket
##               28               62               56               39
##           enhanc              ytd        ibdirusha investingwithibd
##               27               27               29               26
##            score             docu             twlo           featur
##               27               58               40               58
##             baba             dxcm             team             okta
##              151               25               40               52
##             tdoc             shop              wmt              amd
##               52              162               66              211
##             nvda           nvidia             "mr       wonderful"
##              289               27               26               26
##            faang             pypl              adb           revenu
##               31               60               34               50
##           morgan         oversold              jul            fauci
##               30               31               26               31
##           websit              ccl         alphabet              inc
##               31               36               30               44
##             ftse            regul              nio        communiti
##               55               50               65               26
##            music             gspc              aal          daytrad
##               49               67               45               37
##              zbf          discord             bidu             roku
##               42               62               62               91
##              pfe          deitaon              ino              usa
##               66               49               27               36
##            alert          oversea              rev            fb...
##               48               45               58               26
##             idex             dkng          madison        winstapro
##               70               31               33               25
##              xom             spce             shll             blue
##               27               34               26               29
##         musicnew          channel              wkhs             sedg
##               68               28               42               40
##              tme              uff              rti              rut
##               34               38               52               51
##             itox          analyt             nclh              mil
##               26               83               27               28
##           htt...              blu           cap...        iamcardib
##               50               30               46               40
##        lnpservic              dax   tradingchannel    harmonicchart
##               40               50               28               48
##              rts      triggertrad        excelsior      geniusbrand
##               33               34              105              155
##        gnusbrand         nasda...   stanleeunivers    therealstanle
##              156              114              105              105
##   kartoonchannel          kellogg      llamallama
##               50               50               50
```

```
barplot(bp1, las = 2, col = rainbow(100))
```




```
#Tf-IDF matrix for corpus 1

tfidf1 <- TermDocumentMatrix(twittercorpus1, control = list(weighting = weightTfIdf, stopwords = mystopw
```

```
## Warning in weighting(x): empty document(s): 110 1072 1078 1087 1173 1178 1184
## 1247 1250 1254 1259 1260 1270 1293 1327 1346 1360 1388 1441 1469 1560 1568 1663
## 1722 1826 2602 2605 4470 4513 4515 4535 4555 4567 4597 4663 4691 4732 4928 5102
## 5564 5859 6027
```

```
tfidf1
```

```
## <<TermDocumentMatrix (terms: 8308, documents: 7256)>>
## Non-/sparse entries: 60657/60222191
## Sparsity           : 100%
## Maximal term length: 38
## Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
```

```
inspect(tfidf1[1:10,1:20])
```

```
## <<TermDocumentMatrix (terms: 10, documents: 20)>>
## Non-/sparse entries: 12/188
```

```
## Sparsity           : 94%
## Maximal term length: 8
## Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
## Sample             :
##           Docs
## Terms            1        15       17        2 3 4 5 6 7 8
##   agre     1.354087 0.0000000 0.624963 0.000000 0 0 0 0 0 0
##   bird     1.970826 0.0000000 0.000000 0.000000 0 0 0 0 0 0
##   can      0.000000 0.5449919 0.000000 1.089984 0 0 0 0 0 0
##   color    1.970826 0.0000000 0.000000 0.000000 0 0 0 0 0 0
##   differ   1.296761 0.0000000 0.000000 0.000000 0 0 0 0 0 0
##   live     0.000000 0.0000000 0.000000 1.293481 0 0 0 0 0 0
##   porsch   1.970826 0.0000000 0.000000 0.000000 0 0 0 0 0 0
##   relat    0.000000 0.0000000 0.000000 1.636221 0 0 0 0 0 0
##   situat   1.669601 0.0000000 0.000000 0.000000 0 0 0 0 0 0
##   tmptrade 0.000000 0.0000000 0.000000 2.564992 0 0 0 0 0 0
```

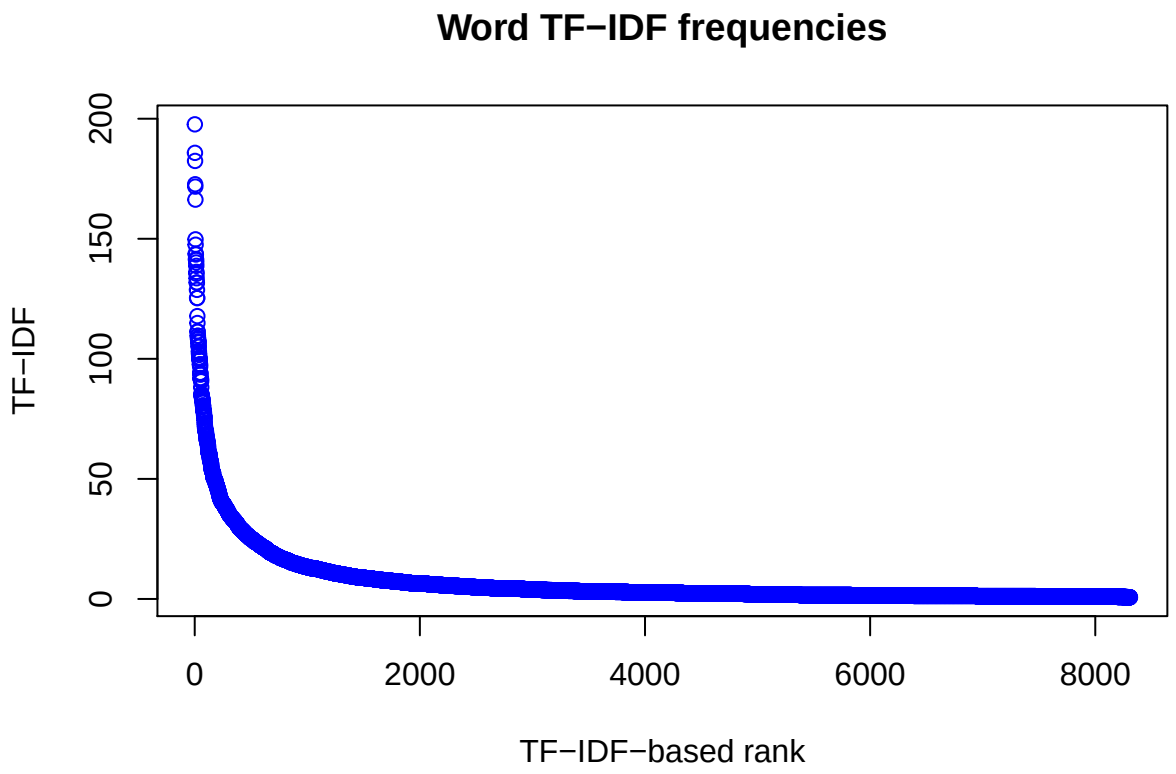
```
freq=rowSums(as.matrix(tfidf1))
head(freq,10)
```

```
##       agre       bird      color     differ     porsch     situat        can
##  32.730882   3.941653   3.941653  35.318573   3.941653   8.201281 109.610080
##       live      relat   tmptrade
##  68.711566  26.330909   2.564992
```

```
tail(freq,10)
```

```
## andyheyward         cle         knc      affect   manufactu         toe
##    1.832137    1.165905    2.553116    1.603120    1.603120    1.424995
##        vest      cleari   nasdaqwkh     thunder
##    1.424995    1.603120    2.137493    2.137493
```

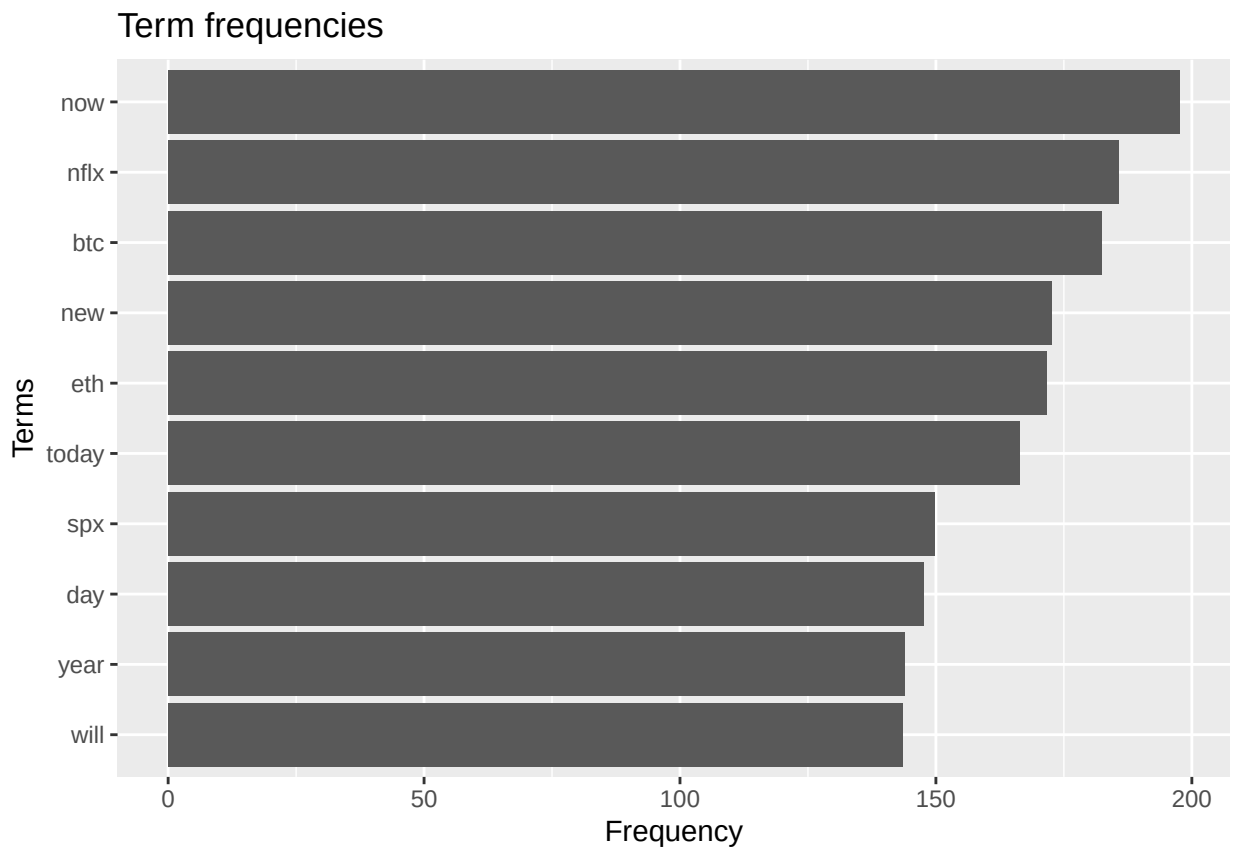
```
plot(sort(freq, decreasing = T),col="blue",main="Word TF-IDF frequencies", xlab="TF-IDF-based rank", yla
```




```
tail(sort(freq),n=10)
```

```
##     will     year      day      spx    today      eth      new      btc
## 143.4327 143.8118 147.5209 149.7566 166.2986 171.7264 172.6959 182.4114
##     nflx      now
## 185.7860 197.6486
```

```
high.freq=tail(sort(freq),n=10)
hfp.df=as.data.frame(sort(high.freq))
hfp.df$names <- rownames(hfp.df)

ggplot(hfp.df, aes(reorder(names,high.freq), high.freq)) +
  geom_bar(stat="identity") + coord_flip() +
  xlab("Terms") + ylab("Frequency") +
  ggtitle("Term frequencies")
```

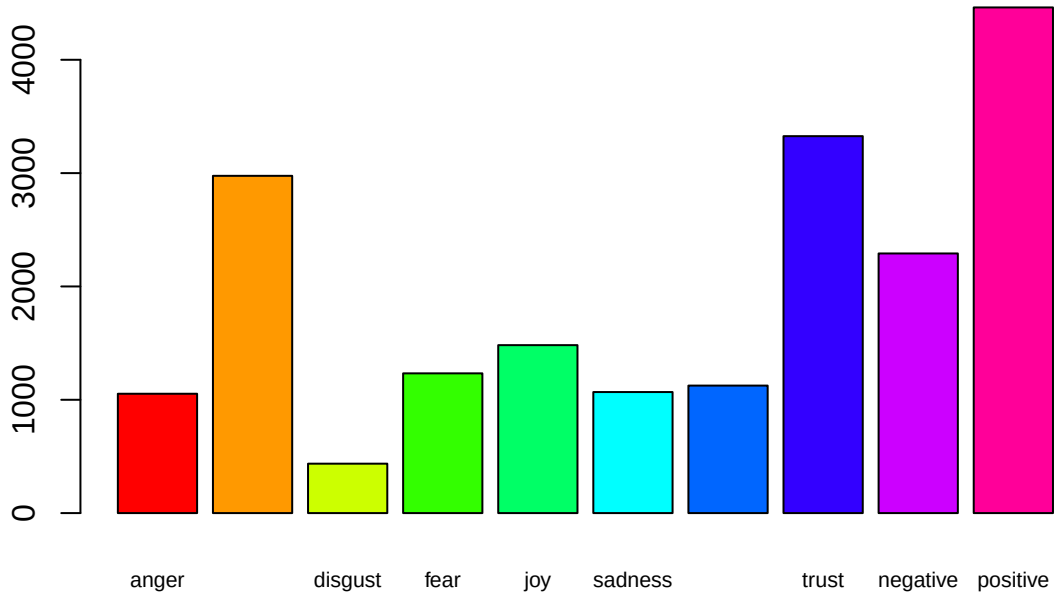



```
#Creating the wordcloud

cloud1 <- sort(rowSums(termmatrix1), decreasing = TRUE)
cloud1 <- data.frame(names(cloud1), cloud1)
colnames(cloud1) <- c('word', 'freq')
wordcloud2(cloud1, size = 0.5, shape = 'circle', rotateRatio = 0.5, minSize = 1)
```


```
#sentiment analysis

emotions1 <- get_nrc_sentiment(twittercorpus1$content)
```

```
## Warning: ‘filter_()‘ is deprecated as of dplyr 0.7.0.
## Please use ‘filter()‘ instead.
## See vignette('programming') for more help
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_warnings()‘ to see where this warning was generated.

## Warning: ‘group_by_()‘ is deprecated as of dplyr 0.7.0.
## Please use ‘group_by()‘ instead.
## See vignette('programming') for more help
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_warnings()‘ to see where this warning was generated.

## Warning: ‘data_frame()‘ is deprecated as of tibble 1.1.0.
## Please use ‘tibble()‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_warnings()‘ to see where this warning was generated.
```

```
barplot(colSums(emotions1),cex.names = .7, col = rainbow(10), main = "sentiment scores for tweets1")
```


sentiment scores for tweets1
0
1000
2000
3000
4000
anger
disgust
fear
joy
sadness
trust
negative
positive