Sentiment Analysis 2

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:

```r
#connecting to twitter

consumer_key <- "jHHSo1xVw0mhp1cyFTac86nK3"
consumer_secret <- "sEMM8ed0DVd8j9DaKnta8UkzEaMLdzwoYAet8mf3MWiitVewQp"
access_token <- "244645993-QrRcQWJXOb5SGBI8W6jxILOuYMk686TIIGKn9a7Q"
access_secret <- "mZ1aa2TsvWDqfBVMU2GUEHUiglTjkmQQ7KUimGFGQOilP"
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

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

```r
#obtaining the second set of tweets

appleonly <- searchTwitter("#AAPL", n = 1000, lang = 'en')
```

```
## Warning in doRppAPICall("search/tweets", n, params = params, retryOnRateLimit =
## retryOnRateLimit, : 1000 tweets were requested but the API can only return 872
```

```r
applemoney <- searchTwitter("#$AAPL", n = 1000, lang = 'en')
msftonly <- searchTwitter("#microsoft", n = 1000, lang = 'en')
msftmoney <- searchTwitter("#$msft", n = 1000, lang = 'en')
amazononly <- searchTwitter("#amazon", n = 1000, lang = 'en')
amazonmoney <- searchTwitter("#$amzn,", n = 1000, lang = 'en')
googleonly <- searchTwitter("#google", n = 1000, lang = 'en')
googlemoney <- searchTwitter("#$goog", n = 1000, lang = 'en')
nasdaqonly <- searchTwitter("#nasdaq", n = 1000, lang = 'en')
nasdaqmoney <- searchTwitter("#$ndaq", n = 1000, lang = 'en')
```

```
## Warning in doRppAPICall("search/tweets", n, params = params, retryOnRateLimit =
## retryOnRateLimit, : 1000 tweets were requested but the API can only return 172
```

```
#Transforming to dataframe
tweets2 <- tbl_df(map_df(c(appleonly, applemoney, msftonly, msftmoney, amazononly, amazonmoney, googleor
```

```
## 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.
```

```
#Saving into CSV
write.csv(tweets2, file = "tweets2.csv", row.names = FALSE)
```

```
#setworking directory and read in data
setwd("C:/Users/ziyad/Desktop/Data Analytics capstone")
tweets2 <- read.csv("tweets2.csv")
```

```
#twitter data cleanup for dataset2

twittercorpus2 <- Corpus(VectorSource(tweets2$text))
inspect(twittercorpus2[1:10])
```

```
## <<SimpleCorpus>>
## Metadata:  corpus specific: 1, document level (indexed): 0
## Content:   documents: 10
##
##  [1] 55 pips with #AUDJPY trade!<U+0001F525><U+0001F525>\n\nhurry get Free signals : https://t.co/Ep
##  [2] $AAPL Big doji, is this a reversal signal or pause? - #AAPL chart https://t.co/mOxU3gxxA2
##  [3] The world is changing \n\nImagine investing into the next Amazon or Apple, except this time it'
##  [4] #AAPL Has reached the next potential target. The price has been pushed back to retest the recen
##  [5] The single largest bearish inverted Hammer I have ever seen on #AAPL in 25 years. Very rare for
##  [6] Closed at $381.91, down 1.77. Wild. @Apple #AAPL
##  [7] RT @tq_equities: $AAPL closes within previous day's range: https://t.co/mzmYImiHSA\nSentiment: 
##  [8] $AAPL closes within previous day's range: https://t.co/mzmYImiHSA\nSentiment: Neutral #AAPL #Ap
##  [9] #AAPL hit the 1.618 #Fibonacci projection of the #GVC range today, and sure couldn't do much wi
## [10] $AAPL #AAPL worth dipping foot into some calls
```

```
twittercorpus2 <- tm_map(twittercorpus2, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2, content_transformer(tolower)):
## transformation drops documents
```

```
twittercorpus2 <- tm_map(twittercorpus2, removeWords, stopwords("en"))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2, removeWords, stopwords("en")):
## transformation drops documents
```

```
twittercorpus2 <- tm_map(twittercorpus2, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2, removeNumbers): transformation
## drops documents
```

```
twittercorpus2 <- tm_map(twittercorpus2, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2, removePunctuation):
## transformation drops documents
```

```
removeURLhttp2 <- function(x) gsub ("http[[:alnum:]]*", "", x)
twittercorpus2 <- tm_map(twittercorpus2, content_transformer(removeURLhttp2))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2,
## content_transformer(removeURLhttp2)): transformation drops documents
```

```
removeURLedua2 <- function(x) gsub ("edua[[:alnum:]]*", "", x)
twittercorpus2 <- tm_map(twittercorpus2, content_transformer(removeURLedua2))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2,
## content_transformer(removeURLedua2)): transformation drops documents
```

```
removeNonAscii2 <- function(x) textclean::replace_non_ascii(x)
twittercorpus2 <- tm_map (twittercorpus2,content_transformer(removeNonAscii2))
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2,
## content_transformer(removeNonAscii2)): transformation drops documents
```

```
twittercorpus2 <- tm_map(twittercorpus2, removeWords, c("amp","...","ufef","ufeff","ufeft","uufefuufefu
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2, removeWords, c("amp", "...", :
## transformation drops documents
```

```
mystopwords2 <- c(stopwords("en"),"cnbc","join","today","signal","level","amp","...","ufef","ufeff","uf
twittercorpus2 <- tm_map(twittercorpus2, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2, stripWhitespace): transformation
## drops documents
```

```
inspect(twittercorpus2[1:10])
```

```
## <<SimpleCorpus>>
## Metadata:  corpus specific: 1, document level (indexed): 0
## Content:  documents: 10
##
##  [1] pips audjpy tradeufuf hurry signals stocks crypto eurchf ...
##  [2]  doji reversal signal pause chart
##  [3] world changing imagine investing next except time outside ...
##  [4]  reached next potential target price pushed back retest recent breakout failure...
##  [5] single largest bearish inverted hammer ever seen years rare pattern...
##  [6] closed wild
##  [7] tqequities closes within previous days range sentiment neutral
##  [8]  closes within previous days range sentiment neutral
##  [9]  fibonacci projection range today sure much believe...
## [10]  worth dipping foot calls
```

```
#term document matrix for corpus 2

twittercorpus2 <- tm_map(twittercorpus2, stemDocument)
```

```
## Warning in tm_map.SimpleCorpus(twittercorpus2, stemDocument): transformation
## drops documents
```

```
dtm2 <- TermDocumentMatrix(twittercorpus2)
dtm2
```

```
## <<TermDocumentMatrix (terms: 11493, documents: 9044)>>
## Non-/sparse entries: 75543/103867149
## Sparsity           : 100%
## Maximal term length: 40
## Weighting          : term frequency (tf)
```

```
termmatrix2 <- as.matrix(dtm2)
termmatrix2[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
##   ...     1 0 1 0 0 0 0 0 0  0  0  0  1  0  0  1  1  0  0  0  1
##   audjpi  1 0 0 0 0 0 0 0 0  0  0  0  0  1  0  0  0  0  0  0
##   crypto  1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   eurchf  1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   hurri   1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  1
##   pip     1 0 0 0 0 0 0 0 0  0  0  0  0  1  0  0  0  1  1  0
##   signal  1 1 0 0 0 0 0 0 0  0  0  0  0  1  1  1  0  1  1  1
##   stock   1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  1  0
##   tradeufuf 1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
##   chart   0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
```

```
freq=rowSums(as.matrix(termmatrix2))
head(freq,10)
```

```
##        ...     audjpi     crypto     eurchf      hurri        pip     signal      stock
##       2814         56         13          8         16        354        333        441
## tradeufuf      chart
##         6        159
```

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

```
##     lollllll coinbase...     scrutin      schw...      defici     receipt
##            2           2           1           2           1           1
##      disjoint    felixleon  expected...     brokerag
##            2           1           1           1
```

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

Term document matrix frequencies
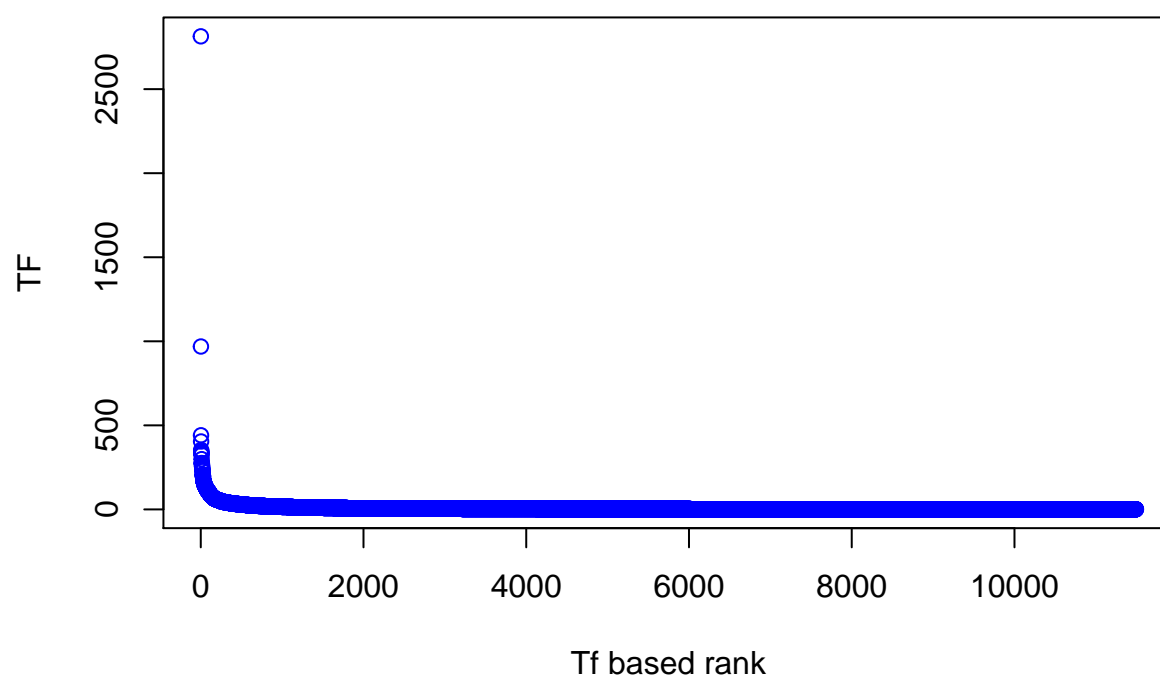

```r
tail(sort(freq),n=10)
```

```
##    high signal   join   will  trade    pip  today  stock  googl    ...
##     324    333    334    340    346    354    404    441    969   2814
```

```r
#detailed term frequency barplot

bp2 <- rowSums(termmatrix2)
bp2 <- subset(bp2, bp2>=25)
bp2
```

```
##              ...          audjpi             pip          signal
##             2814              56             354             333
##            stock           chart          revers           chang
##              441             159             112              35
##           invest            next          outsid            time
##              274             203              30             265
##            world            back        breakout           price
##               85             119              25             239
##            reach          recent          target            ever
##               44              53             276              57
##             year           close             day       sentiment
##              236             143             144              39
##           within       fibonacci            much         project
##               52              82             105              35
```

```
##           sure            today             call           worth
##             41              404              137              48
##          becom            futur             miss            spce
##             48               91               87              32
##          bubbl           profit             best         channel
##             49              255              124              34
##        premium           result            thing           watch
##             44               63               45             201
##         eurusd             join           audnzd          eurcad
##             31              334               25              35
##         gbpaud  profitufuubufef      coronavirus            earn
##             32               27               56              80
##          juli            report             will            also
##            146               86              340              55
##          start            tesla            trade           trump
##            106              108              346              92
##          faang            learn           second           video
##             31               87               36             116
##          daili            forex          bitcoin         checkout
##            146              158               45              45
##           info             open            peopl           short
##             98              119              110             175
##         xauusd            avail            group            hour
##             37              115               77              82
##         member             view             know            done
##             63               58              103              53
##       stocks...     profitufuuf           usdjpi          analysi
##             32               29               37              59
##           live           record             room            high
##             65               69               29             324
##        netflix        forexsign            usoil            t...
##             51               33               41              84
##          ahead           review             week          economi
##             34               63              256              93
##         higher           market           almost            m...
##            106              161               46              26
##         number          current            money           print
##             35               54               59              44
##           show            first            minut            sinc
##             30              104               33              82
##           soon             morn            break       elliottwav
##             56               76               44              37
##        forex...           earli             like           indic
##             35               58              199              72
##           long           bought         opportun            rais
##             95               31               31              60
##           keep             rise           better            even
##             48               45               59              47
##           pick            anoth           releas          leader
##             41               71               65              29
##        predict             tech           silver            love
##             29              275               26             109
##            run             ufuf             must          biggest
##             27               32               37              37
```

```
##          make           move           cost        product
##           130             82             68             108
##         servic          still         usdcad           fast
##            67            122             36              26
##          point         vaccin         gbpusd          manag
##            65             28             36              53
##           sell          think           look          readi
##           132             64            173              44
##           stop        success          updat        overview
##            31             42            139              40
##        technic           f...       facebook       robinhood
##            32             30             92              43
##           wait           past         return          track
##            40             27            127              43
##          great        compani         friday        interest
##           103            136             32              75
##           take         option           play           nice
##           111             81             76              47
##         chrome         comput           just           quit
##            34             55            229              41
##           blue            buy        perfect           book
##            42             44             42             244
##           baba          googl          share           step
##           163            969            129              38
##         trader           post         sector          three
##            55             60             64              27
##           busi           cycl           gold goldswitzerland
##           144             35            142              26
##    proprietari            guy           rate          ubufef
##            27             27             30              50
##       investor           twlo          level            low
##            35             36            229              87
##            min       spiderman         experi          littl
##           118             45             72              33
##           come           ebay           o...           shop
##           171             32             40             206
##           twtr          final           g...           huge
##           166             30             30              31
##           nvda        bullish         season          thank
##           299             28             29             112
##        everyon           talk         featur          month
##            48             41             29              91
##           valu           save        discord         xtrade
##            57             76             73              25
##          volum        account          click           link
##           114             65             60              97
##        support     trading...          activ          index
##           101             25             34              39
##           cnbc          sampp         growth           note
##           134            115             34              55
##         expect           head          trial           find
##           121             42             28              73
##           idea            put           sold           went
##            50             45             43              29
```

```
##         spot    trillion        gift        bank
##           60          46          37          45
##         good    antitrust        data        plan
##          170          56          70          54
##       continu    oversold      happen        want
##           66          34          70         120
##         bull        code        educ     program
##           34          50          39          40
##   stockmarket     develop      launch        tool
##          108          55          74          38
##        enter        last       entri     increas
##           45         147          47          74
##      analyst         say     million       total
##           53          63          72          32
##         news         hit     control    tomorrow
##          128          38          28          37
##        everi       retail        lead    technolog
##           65          39          45          54
##         case       covid       notic        stay
##          108          98          32          43
##        never       realli      expert        need
##           46          31          30         120
##          tri       offer      provid        real
##           30          60          29          63
##        alert        djia       ralli      health
##           59          36          87          33
##        offic        loss      action       pleas
##          134          38          42          61
##       strong        tell        idex       might
##           39          32          60          28
##      startup        hard     billion       posit
##           35          27         126          42
##         near        life        june     session
##           33          43          32          35
##         care       analyt      discov       media
##           29          91          32          30
##       public       trend     generat        fall
##           39          34          36          33
##    afternoon      direct       right         see
##           36          31          83          36
##         mark        help        w...        work
##           33         112          33         194
##        boost      demand        jump     samsung
##           32          26          26          33
##      alreadi         get       candl       score
##           27         115          33          41
##         base        c...    communiti     suggest
##           29          28          35          43
##        state        fund        hold        hope
##           33          87         103          45
##       august        pass         win        sale
##           40          28          43          38
##       stream       remot     everyth       estim
##           34          26          29          28
```

8

```
##            held          user   themotleyfool          gspc
##              26            67              31            76
##           heart          mani       sandropow         skill
##              31            63              32            52
##       brilliant       madison           studi           use
##              28            39              25            83
##            team         equiti           cmcsa          roku
##              97            46              31           105
##            snap          gain          follow          made
##              30           130              43            56
##         confirm        countri            home          read
##              26            28              33           140
##            game          test         earlier          sign
##             101            41              44            32
##         discuss        monday          beauti         quick
##              40            86              28            25
##           china         latest             end          size
##              29            53              54            42
##            shot          drop         announc          full
##              45            83             153           127
##            tdoc        deitaon            docu        spread
##              56            56              59            33
##            bidu       sedgpypl        musicnew          list
##              81            39              66            59
##            word        cnbc...            deal          well
##              39           169             152            64
##          articl         improv            bynd       perform
##              59            37              51            39
##             adb         upgrad           power          pypl
##              39            26              77            28
##           creat           soni          enhanc        websit
##              52            43              29            48
##            fire           lyft            uber      platform
##              30            25              33            33
##        composit  investingcom        microso...         nkla
##              36            26              28            43
##            line           intc           train       back...
##              30            60              35            28
##            meli         welcom            boom        system
##              25            33              42            33
##          narrow       thursday            feel       thought
##              25            30              56            33
##            bear          digit           often         bring
##              26           147              26            47
##          wonder        oversea             rev        revenu
##              38            45              56            46
##          offici         winner           y...          dkng
##              46            39              32            28
##            turn          green     "wonderful"          okta
##              47            32              25            49
##            file           part         android         order
##              73            34              42            63
##           coupl           cent            nyse         store
##              27            38              30            46
```

```
##           flow    winstapro       reason       averag
##            190           25           26           26
##           star        check         flat         ship
##             40          166           39           26
##           grow     consider        simul     introduc
##             65           39           50           42
##         multipl        chanc         meet         wkhs
##             43           27           26           28
##         financi       defend         said         bond
##             27           30           26           25
##         detect         lrcx        music        build
##             48           26           63           42
##           face        white        begin        alway
##             34           35           32           40
##          award      partner       exclus         xbox
##             28           66           38          110
##           azur        excel        onlin        cloud
##            168           34           64          137
##      iamnetwork         date       detail      collabor
##             33           42           34           28
##           page       advanc        stori      protect
##             54           35           52           44
##         storag       threat          app       xbox...
##             29           33           33           29
##         design       flight        secur      contact
##             29           51           67           25
##         complet          ban       youtub       window
##             30           35          135          111
##          solut       applic          set        smart
##             27           25           29           30
##       transform        write        devic        engin
##             26           51           27           38
##          class     research       summer        linux
##            105           55           56           28
##         surfac     infograph       search       artist
##             31           27           76           77
##       alphabet  sundarpichai  deadarticgam         seri
##             35           57           57           79
##          found         edit     preorder      signific
##             27           34           35           44
##          child      version       import        ebook
##             25           54           28           57
##         school         qcom      weekend        heard
##             25           28           33           32
##          major         rank        india        steal
##             26           38          208           44
##           pack         roof        night         mask
##             40           46           29           33
##        wayfair      present        visit        novel
##             47           28           28           31
##         author        kindl        shelv       freebi
##             62           83           36           34
##         romanc         card          tax        women
##             35           28           48           35
```

```
##               iartg        internet             copi         mysteri
##                  25              40               27              26
##              amazon         giveaway       amazonprim         adventur
##                  25              29               25              29
##            birthday           spend          "china          ezralev
##                  25              25               30              29
##     rebelnewsonlin           virus"            polic          dingber
##                  28              29               39              27
##       googleforindia         fatkidd          uffauff            upper
##                  41              35               35              30
##               regul    digitalmarket            crore         google...
##                  50              28               30              28
##           daysofcod            html           pichai           sundar
##                  46              32               29              30
##    growthhacking...     nitishsharma           onpag          chamath
##                  26              25               26              47
##           iamcardib         lnpservic         elerianm        intrasess
##                  37              37               34              35
##         etfprofessor  financialservic            ndaq
##                  44              44              154
```

```r
barplot(bp2, las = 2, col = rainbow(100))
```
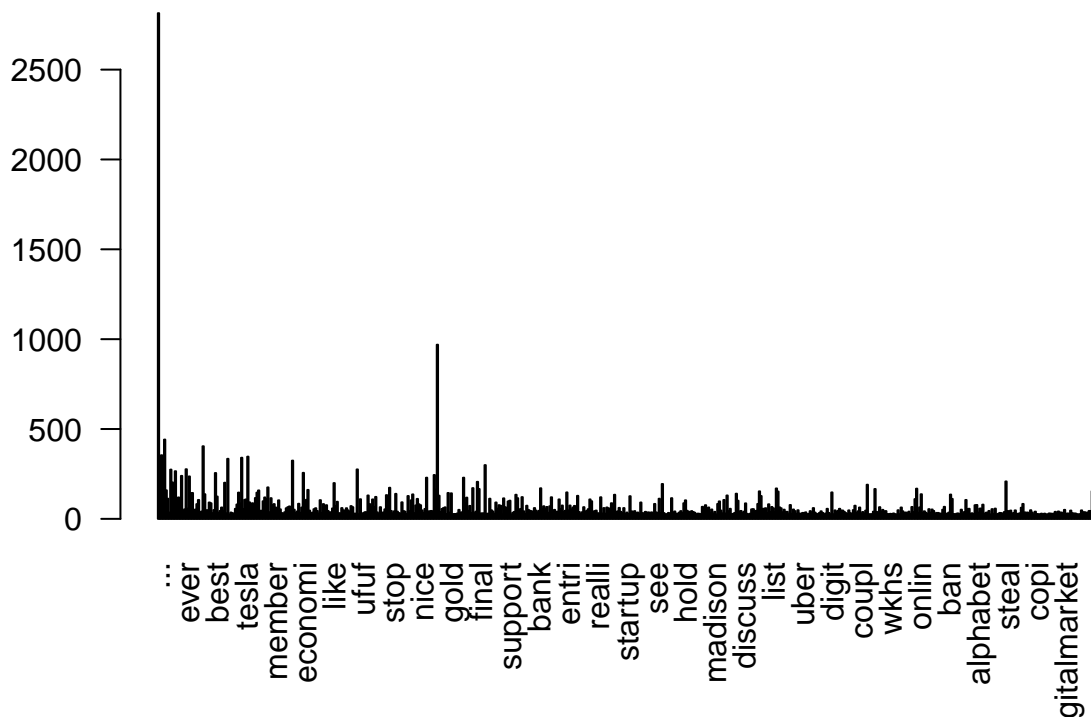
```r
tfidf2 <- TermDocumentMatrix(twittercorpus2, control = list(weighting = weightTfIdf, stopwords = mystop
```

```
## Warning in weighting(x): empty document(s): 55 164 694 1391 1434 1609 1832 2985
## 3028 3030 3050 3070 3082 3112 3177 3185 3205 3246 3343 3442 3460 3537 3616 3774
## 4761 6562 7009 7066 7100 7304 7472 7642 8016 8051 8091 8130
```

```
tfidf2
```

```
## <<TermDocumentMatrix (terms: 10866, documents: 9044)>>
## Non-/sparse entries: 68256/98203848
## Sparsity           : 100%
## Maximal term length: 40
## Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
```

```r
inspect(tfidf2[1:10,1:20])
```

```
## <<TermDocumentMatrix (terms: 10, documents: 20)>>
## Non-/sparse entries: 12/188
## Sparsity           : 94%
## Maximal term length: 9
## Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
## Sample             :
##           Docs
## Terms              1        14        2        20         3 4 5 6 7 8
##   audjpi    1.467078 1.833848 0.000000 0.000000 0.000000 0 0 0 0 0
##   chang     0.000000 0.000000 0.000000 0.000000 1.001683 0 0 0 0 0
##   chart     0.000000 0.000000 1.508555 0.000000 0.000000 0 0 0 0 0
##   crypto    1.778964 0.000000 0.000000 0.000000 0.000000 0 0 0 0 0
##   doji      0.000000 0.000000 2.493205 0.000000 0.000000 0 0 0 0 0
##   eurchf    1.936663 0.000000 0.000000 0.000000 0.000000 0 0 0 0 0
##   hurri     1.828549 0.000000 0.000000 1.306106 0.000000 0 0 0 0 0
##   paus      0.000000 0.000000 2.889446 0.000000 0.000000 0 0 0 0 0
##   revers    0.000000 0.000000 1.583848 0.000000 0.000000 0 0 0 0 0
##   tradeufuf 2.111557 0.000000 0.000000 0.000000 0.000000 0 0 0 0 0
```

```r
freq=rowSums(as.matrix(tfidf2))
head(freq,10)
```

```
##     audjpi     crypto     eurchf      hurri  tradeufuf      chart       doji
##  73.822553  24.475188  15.939391  24.485868  11.362185 127.919872  12.692680
##       paus     revers      chang
##   7.430003  96.625445  37.989362
```

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

```
##   lolllll   coinbase    scrutin        schw     defici    receipt   disjoint felixleon
##  3.083872   2.698388   1.877535    3.469356   2.190458   2.190458   2.563468  1.314275
##  expected    brokerag
##  1.460305   1.642843
```

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

## Word TF–IDF frequencies
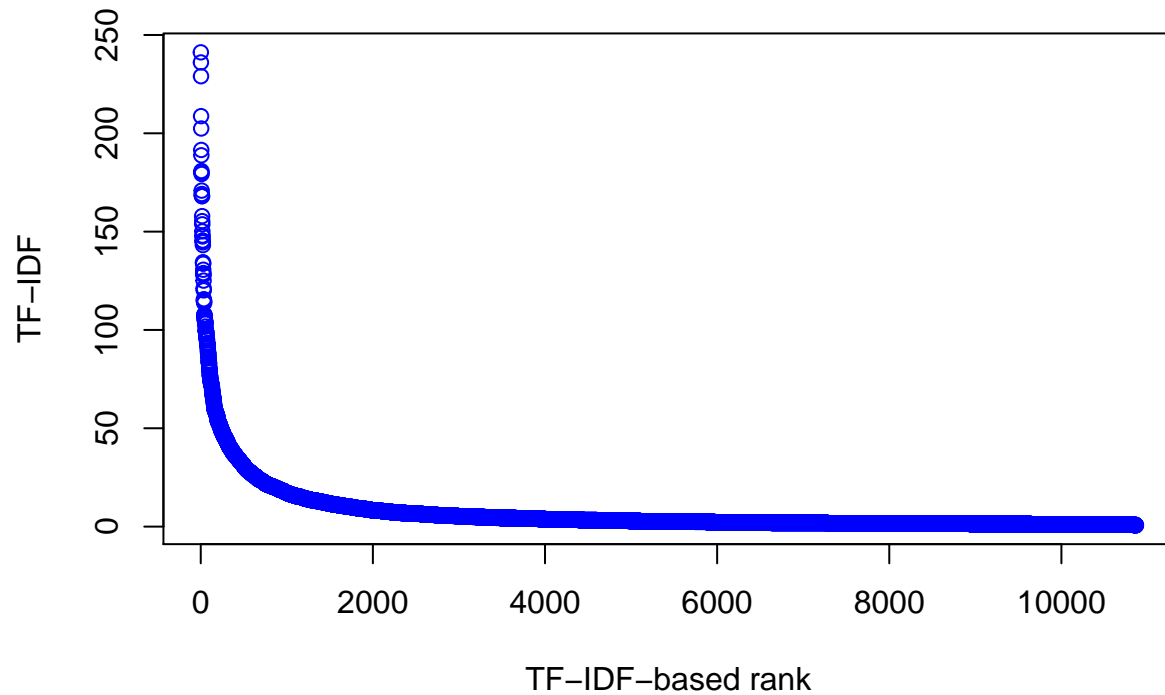

```r
tail(sort(freq),n=10)
```

```
##      time      tech     price      shop    invest      will    profit      nvda
## 179.3050 180.0987 180.8174 188.8136 191.5315 202.4379 208.7226 229.0171
##    target      high
## 236.0255 241.1619
```

```r
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")
```

Term frequencies
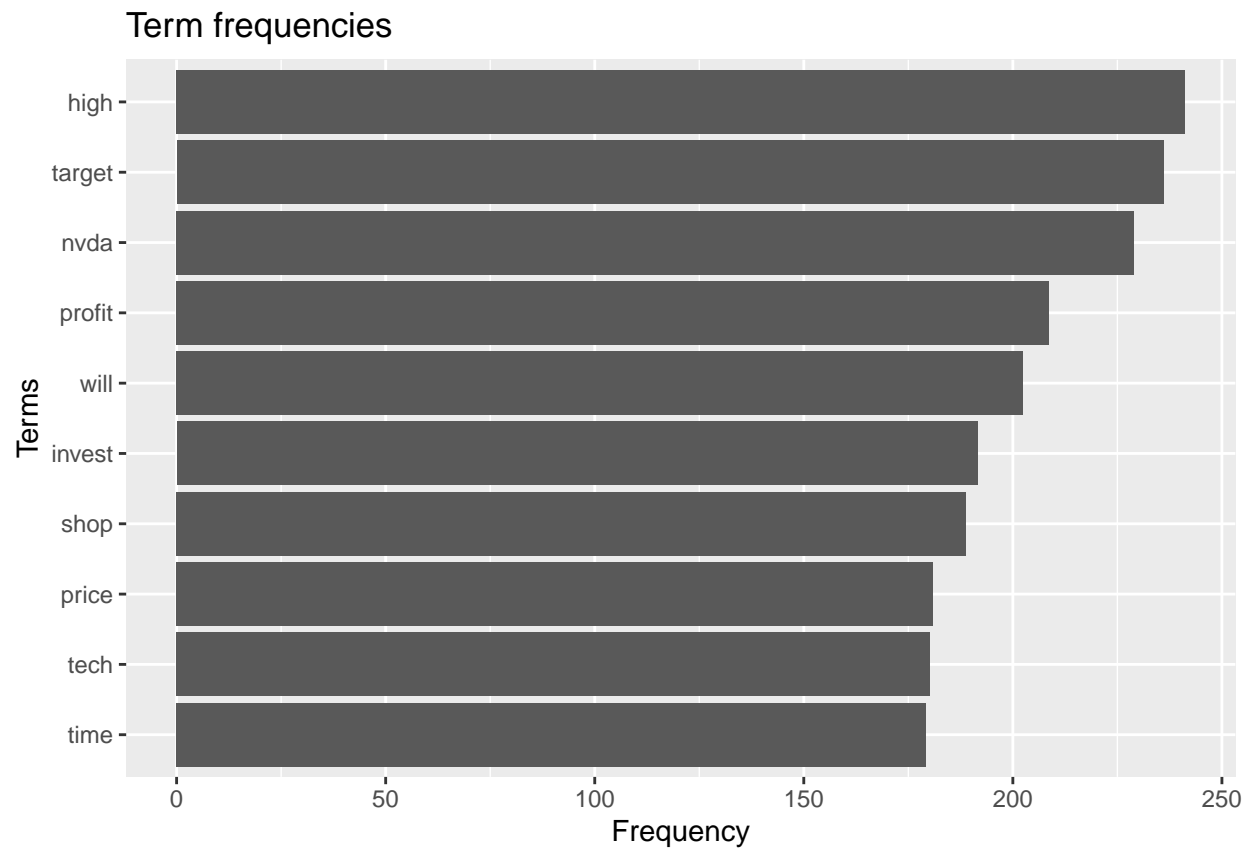

```r
#Creating the wordcloud for the tdm

cloud2tdm <- sort(rowSums(termmatrix2), decreasing = TRUE)
cloud2tdm <- data.frame(names(cloud2tdm), cloud2tdm)
colnames(cloud2tdm) <- c('word', 'freq')
wordcloud2(cloud2tdm, size = 1, shape = 'circle', rotateRatio = 0.5, minSize = 1)
```

]

```
#Sentiment analysis for the second set of tweets

emotions2 <- get_nrc_sentiment(twittercorpus2$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(emotions2),cex.names = .7, col = rainbow(10), main = "sentiment scores for tweets2")
```

sentiment scores for tweets2