Tweets

References

- https://rstudio-pubs-static.s3.amazonaws.com/265713__cbef910aee7642dc8b62996e38d2825d.html
- http://www.sthda.com/english/wiki/text-mining-and-word-cloud-fundamentals-in-r-5-simple-steps-you-should-know

```r
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```r
library(tm)
```

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

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr)
library(wordcloud)
```

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

```r
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:NLP':
##
##     annotate
```

```r
library(tidytext)
library(SentimentAnalysis)
```

```
##
## Attaching package: 'SentimentAnalysis'

## The following object is masked from 'package:base':
##
##     write
```

```r
library(syuzhet)
library(readr)
library(purrr)
```

Analyzing Twitter Data

```
tweetsDF <- readr::read_csv("./../DataExtract/Data/Top Tweets.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   Tweet_Timestamp = col_double(),
##   Tweet_Number_of_Reviews = col_double(),
##   Comment_Timestamp = col_double(),
##   Comment_Number_of_Replies = col_double(),
##   Comment_Number_of_Retweets = col_double(),
##   Comment_Number_of_Likes = col_double(),
##   User_Timestamp = col_double(),
##   User_Number_of_Replies = col_double(),
##   User_Number_of_Retweets = col_double(),
##   User_Number_of_Likes = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
tweetsDF <- clean_names(tweetsDF)
tweetsDF$tweet_number_of_likes <- gsub("," , "" , tweetsDF$tweet_number_of_likes )
tweetsDF$tweet_number_of_likes <- as.numeric(tweetsDF$tweet_number_of_likes )
```

```
## Warning: NAs introduced by coercion
```

```
head(tweetsDF)
```

```
## # A tibble: 6 x 27
##   category keyword web_page_url tweet_website author_name author_web_page~
##   <chr>    <chr>   <chr>        <chr>         <chr>       <chr>
## 1 Top      hair d~ https://twi~ https://twit~ Sam Bhatt ~ https://twitter~
## 2 Top      hair d~ https://twi~ https://twit~ god's litt~ https://twitter~
## 3 Top      hair d~ https://twi~ https://twit~ spooky dai~ https://twitter~
## 4 Top      hair d~ https://twi~ https://twit~ dy          https://twitter~
## 5 Top      hair d~ https://twi~ https://twit~ Anxious bi~ https://twitter~
## 6 Top      hair d~ https://twi~ https://twit~ Megan Mari~ https://twitter~
## # ... with 21 more variables: tweet_timestamp <dbl>, tweet_content <chr>,
## #   tweet_image_url <chr>, tweet_video_url <chr>,
## #   tweet_number_of_likes <dbl>, tweet_number_of_retweets <chr>,
## #   tweet_number_of_reviews <dbl>, commenter_name <chr>,
## #   commenter_web_page_url <chr>, comment_timestamp <dbl>, comment <chr>,
## #   comment_number_of_replies <dbl>, comment_number_of_retweets <dbl>,
## #   comment_number_of_likes <dbl>, user_name <chr>,
## #   user_web_page_url <chr>, user_timestamp <dbl>, user_content <chr>,
## #   user_number_of_replies <dbl>, user_number_of_retweets <dbl>,
## #   user_number_of_likes <dbl>
```

Tweet content analysis

```
## [1] "Mr Ghafoor you are in a #FantasyLand. By the time you wake up u will definitely need some hair c
## [2] "hair dye companies be like our customers have tiny little women hands uwu just the smallest litt
## [3] "Hair dye suggestions please. pic.twitter.com/KcBCGAlyZk"
## [4] "i talked about how i'm looking for dark blue hair dye and this bitch really,,, pic.twitter.com/
## [5] "My birthday week, amma spoil myself to a new piercing and maybe a temporary hair dye. Wanna swi
```

```
## [6] "I was getting color prep to remove my hair dye and I find this! Haha! pic.twitter.com/fnklvJk3Z

## [1] "Bottom me"
## [2] "Is ur natural blonde?"
## [3] "i told you that you look naturally blonde"
## [4] "can the blond hair dye dispear so we can have the brunet u ? pery@ anyways all versions of u ca
## [5] "pic.twitter.com/kgZinsb9Yk"
## [6] "pic.twitter.com/nur9IhkF5Y"
```

Converting to corpus

```r
corpus <- Corpus(VectorSource(documents))
corpus <- tm_map(corpus, content_transformer(tolower))
```

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

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

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

```r
corpus <- tm_map(corpus, removeWords, stopwords("english"))
```

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

```r
#corpus <- tm_map(corpus, stemDocument)

as.character(corpus[3])
```

```
## [1] "hair dye suggestions please pictwittercomkcbcgalyzk"
## [2] "list(language = \"en\")"
## [3] "list()"
```

Word Frequencies

```r
dtm <- DocumentTermMatrix(corpus , control = )
freq <- colSums(as.matrix(dtm))

wf <- data.frame(word=names(freq), freq=freq)
rownames(wf) <- seq(1:nrow(wf))

wf <- wf %>% arrange(desc(freq))

# remove hair and dye
wf <- wf %>% filter(!(wf$word %in% c("hair" , "dye")))

head(wf)
```
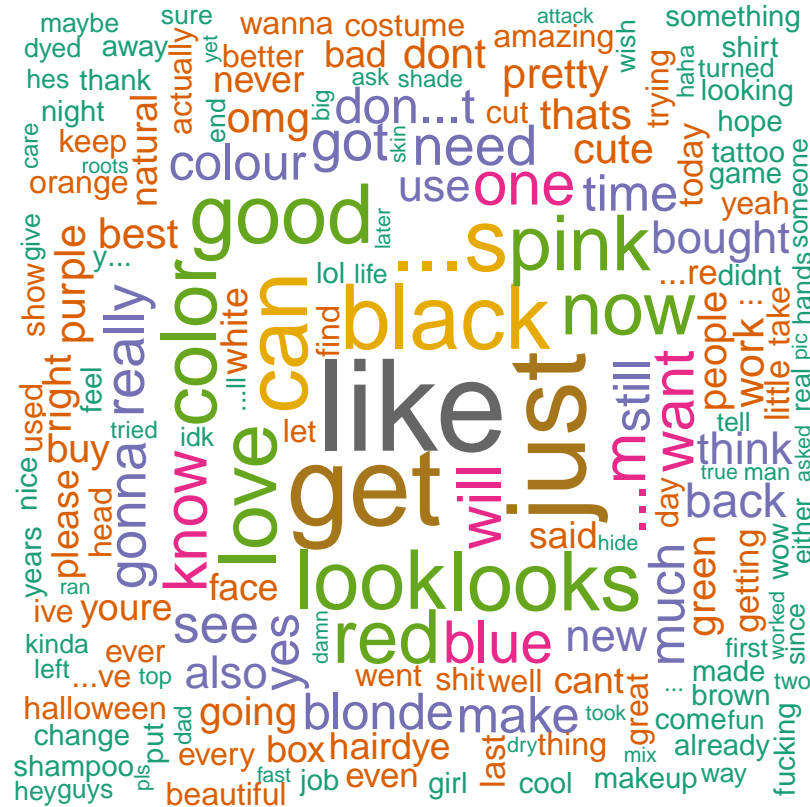
```
##     word freq
## 1   like  106
## 2    get   90
## 3   just   88
## 4  black   70
## 5    can   68
## 6     's   68
```
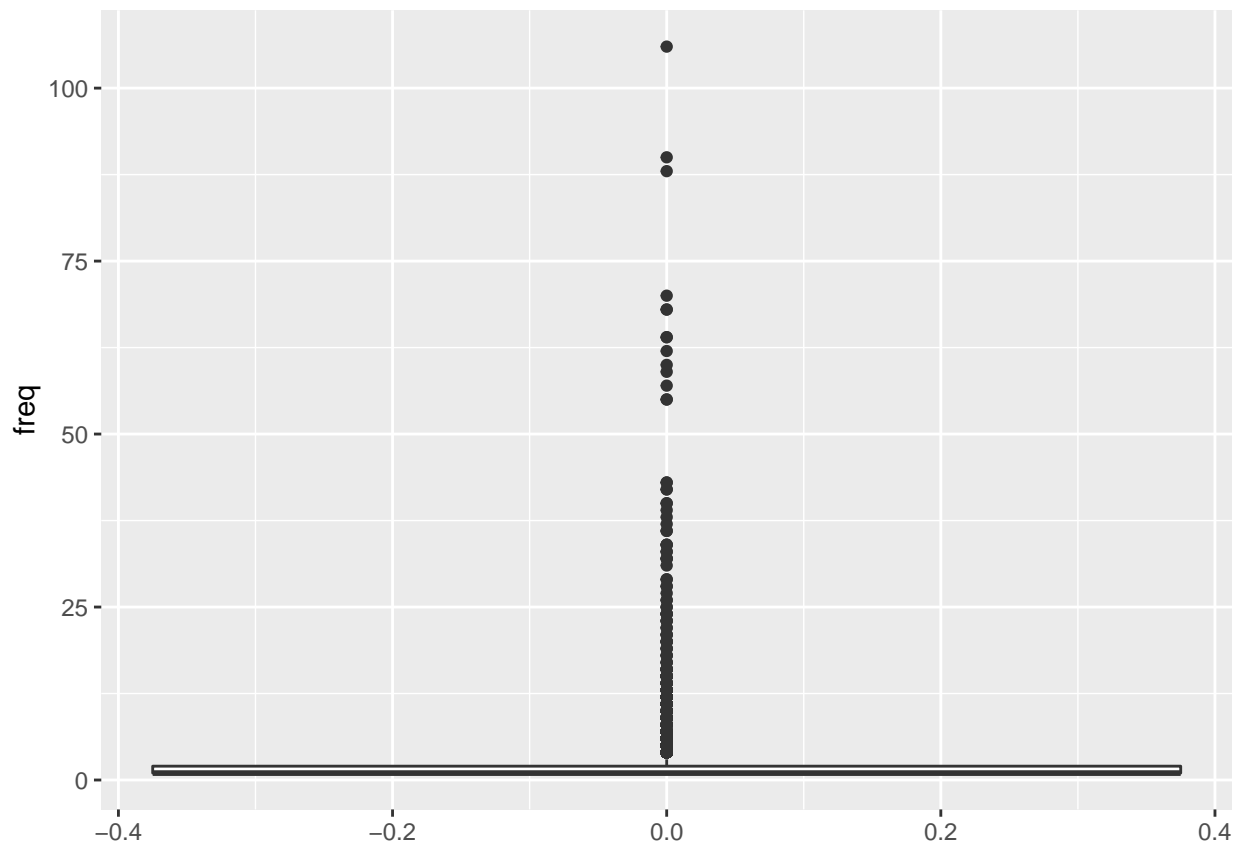
3

Generating Word Cloud

```r
set.seed(1234)
wordcloud(words = wf$word, freq = wf$freq, min.freq = 1,
          max.words=300, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```


Freq distribution

```r
ggplot(data = wf , aes(y = freq)) +
  geom_boxplot()
```

Rare words

```r
head(wf %>% filter(freq > 5 & freq < 10 ))
```

```
##        word freq
## 1 piercing    9
## 2   making    9
## 3    manic    9
## 4     fuck    9
## 5   doesnt    9
## 6    ready    9
```

Waiting ? Wanting ? Thinking?

```r
head(wf %>% filter(word %in% c( "want" , "raw" , "deserve" , "need","buy" ,"wait" )))
```

```
##       word freq
## 1    want   40
## 2    need   39
## 3     buy   22
## 4    wait   10
## 5 deserve    8
```

Brand Counts

```r
head(wf %>% filter(word %in% c("garnier" , "l'oreal" , "raw")))
```

```
##       word freq
```

```
## 1 garnier    1
```

- Not helpful

Color Counts

```r
colorCounts <- wf %>% filter(word %in% c("black" , "brown" , "red" , "purple" , "orange" , "blonde" , "g
colorCounts
```

```
##     word freq
## 1  black   70
## 2    red   55
## 3 blonde   33
## 4 purple   24
## 5  green   22
## 6 orange   15
## 7  brown   12
## 8 silver   10
```

Shade Count

```r
wf %>% filter(word %in% c("dark" , "light" , "medium" , "fade","highlight"))
```

```
##    word freq
## 1  dark   10
## 2  fade    3
## 3 light    3
```

Sentiments

```r
sent <- analyzeSentiment(tweets)
sent <- sent[,1:4]
sent <- as.data.frame(sent)
sent <- cbind(sent , tweet = tweets)
sent <- clean_names(sent)
summary(sent)
```

```
##    word_count      sentiment_gi       negativity_gi      positivity_gi
##  Min.   : 2.00   Min.   :-0.40000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.: 7.00   1st Qu.:-0.06250   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :11.00   Median : 0.00000   Median :0.08000   Median :0.1000
##  Mean   :12.63   Mean   : 0.02273   Mean   :0.08769   Mean   :0.1104
##  3rd Qu.:16.00   3rd Qu.: 0.11111   3rd Qu.:0.14286   3rd Qu.:0.1667
##  Max.   :36.00   Max.   : 0.50000   Max.   :0.40000   Max.   :0.5000
##
##
##  "Dear Noah, indications are very strong that the Antifa super soldiers have the town surrounded. I
##  "There's 50% off blue hair dye at Boots!!!" https://twitter.com/CerianJenkins/status/11844900560122
##  "We can get caught, the last thing you need is a scandal because of me." He looked at Taehyung and
##  "What are you looking for?" Taehyung asked watching him. He took out some ripped jeans, tossed them
##  "YOOOOO I LOVE YOUR CASUAL DEKU HAHA WHERE'D YOU GET THOSE GLASSES HAHA BRO CAN I GET A PICTURE??? 
##  "YOU ARE SHORT JUST LIKE MS. HUSEMAN!!!!" True. "WHY IS YOUR HAIR YELLOW IF HERS IS BROWN." Hair dy
##  (Other)
```

```r
head(sent %>% filter(negativity_gi > 0.14286) %>% arrange(desc(negativity_gi)) %>% select(tweet))
```

```
##                                                                          tweet
## 1                    Are you getting the most life out of your box hair dye?https://rach.tv/2BTPqEc
## 2 how do you get rid of hair dye https://twitter.com/millieknight333/status/1185697425605955584 …
## 3              i just need to get:- sword- hair dye- cool forehead scar pic.twitter.com/l32XKAZpY2
## 4                      Chan: wait what who stole my hair dyeStays:pic.twitter.com/7p4A0t9VZs
## 5                                                        sad. hitting the hair dye aisle
## 6               just waiting for that post hair dye regret to hit pic.twitter.com/sUWfDV54um
```

Emotion analysis

```r
emotions <- lapply(documents , get_nrc_sentiment)
emotionsdf <- cbind(unlist(emotions  , documents))

emotionsdf <- reduce(emotions , rbind , .init = data.frame())
emotionsdf$tweet <- documents

head(emotionsdf)
```

```
##    anger anticipation disgust fear joy sadness surprise trust negative
## 1      0            1       0    0   0       0        0     0        0
## 2      0            0       0    0   0       0        0     0        0
## 3      0            0       0    0   0       0        0     0        0
## 4      1            0       1    1   0       3        0     0        1
## 5      0            1       1    0   1       0        1     0        1
## 6      1            0       0    1   0       1        0     0        1
##    positive
## 1         1
## 2         0
## 3         0
## 4         0
## 5         1
## 6         0
##
## 1 Mr Ghafoor you are in a #FantasyLand. By the time you wake up u will definitely need some hair dye
## 2          hair dye companies be like our customers have tiny little women hands uwu just the smalles
## 3
## 4                                              i talked about how i'm looking for dark
## 5                              My birthday week, amma spoil myself to a new pie
## 6                              I was getting color prep to rem
```
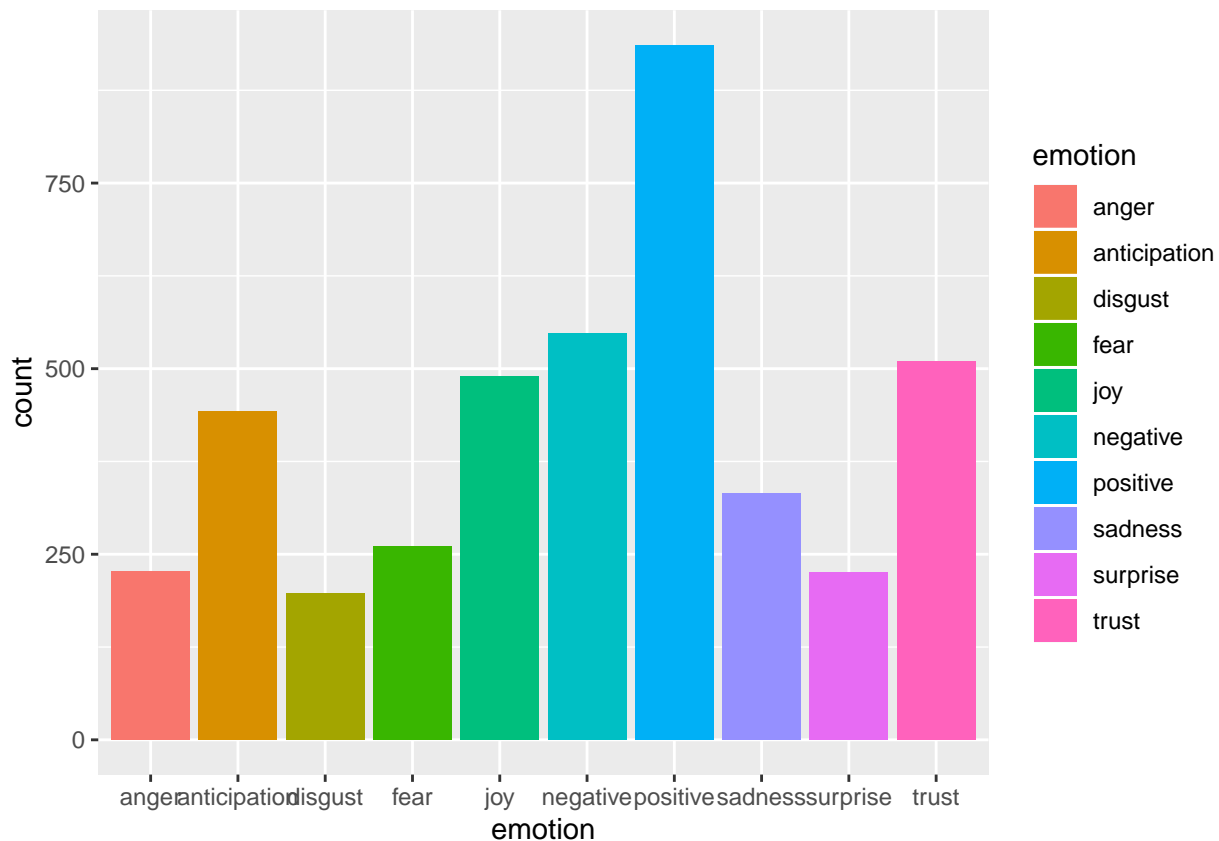
```r
meltedEmotions <- emotionsdf %>% gather(key="emotion" , value="count" , -tweet)

ggplot(data=meltedEmotions, aes(y = count ,x= emotion,fill=emotion))+
  geom_col()
```

Joy

```r
getWordCloud <- function(documents , count=200){
  corpus <- Corpus(VectorSource(documents))
  corpus <- tm_map(corpus, content_transformer(tolower))
  corpus <- tm_map(corpus, removePunctuation)
  corpus <- tm_map(corpus, removeWords, stopwords("english"))
  #corpus <- tm_map(corpus, stemDocument)
  dtm <- DocumentTermMatrix(corpus , control = )
  freq <- colSums(as.matrix(dtm))

  wf <- data.frame(word=names(freq), freq=freq)
  rownames(wf) <- seq(1:nrow(wf))

  wf <- wf %>% arrange(desc(freq))

  # remove hair and dye
  wf <- wf %>% filter(!(wf$word %in% c("hair" , "dye")))
  wordcloud(words = wf$word, freq = wf$freq, min.freq = 1,
          max.words=count, random.order=FALSE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
}

selcdf <- emotionsdf %>% filter(joy > 2)  %>% arrange(desc(joy)) %>% select(tweet)
getWordCloud(selcdf,100)
```

Negative

```
selcdf <- emotionsdf %>% filter(negative > 2)  %>% arrange(desc(negative)) %>% select(tweet)
getWordCloud(selcdf,100)
```

Trust

```
selcdf <- emotionsdf %>% filter(trust > 2)  %>% arrange(desc(trust)) %>% select(tweet)
getWordCloud(selcdf,100)
```

Positive

```
selcdf <- emotionsdf %>% filter(positive > 2)  %>% arrange(desc(positive)) %>% select(tweet)
getWordCloud(selcdf,100)
```

please youre temporary old want
word real youre game home spent
getting water will stop people friend
pink absolute got lee new dark forget rapping
short glad song always see thats girls
good one like job natural best
think pretty can box red
blonde love just now
dear yet don...t ...s really
never baby fun get cute ...m ask made side fuck
bald looks
man two much black feeling
guess blue also make bad using staff
logos dis etc buy costume immediately
haircuts skin wanna can...t dress fan every mutual top