

# Week8Lab

Frank Shen

2019/3/14

```
library(ROAuth)
library(twitteR)
library(wordcloud)
```

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

```
# Declare Twitter API Credentials
consumer_key = "7JEX5oJVhow2gNa0hQfX0lLnA"
consumer_secret = "fGZ77gbFhNB0fQdTt3rQqub2hpnSfD6hx7E7ByCnUnrkntk1Bh"
access_token = "1589845460-VZAJEXduKogFuPRRXdeUng9Svjddtg6YOnLolad"
access_secret = "ohu83GNXf3cUPRJbaH241PAw3K2UnqvxtdCM8UsH6gWL7"

##Send R to requested site to authenticate
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

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

```
##Question 1 a)
dstweets = userTimeline("SFU", n = 3200)
length(dstweets)
```

```
## [1] 565
```

```
save(dstweets, file = "Tweets_week8.Rdata")
##Question 1 b)
dstweetsDF = twListToDF(dstweets)
names(dstweetsDF)
```

```
## [1] "text"          "favorited"      "favoriteCount"  "replyToSN"
## [5] "created"       "truncated"      "replyToSID"     "id"
## [9] "replyToUID"    "statusSource"   "screenName"     "retweetCount"
## [13] "isRetweet"     "retweeted"      "longitude"      "latitude"
```

```
min(dstweetsDF$created)
```

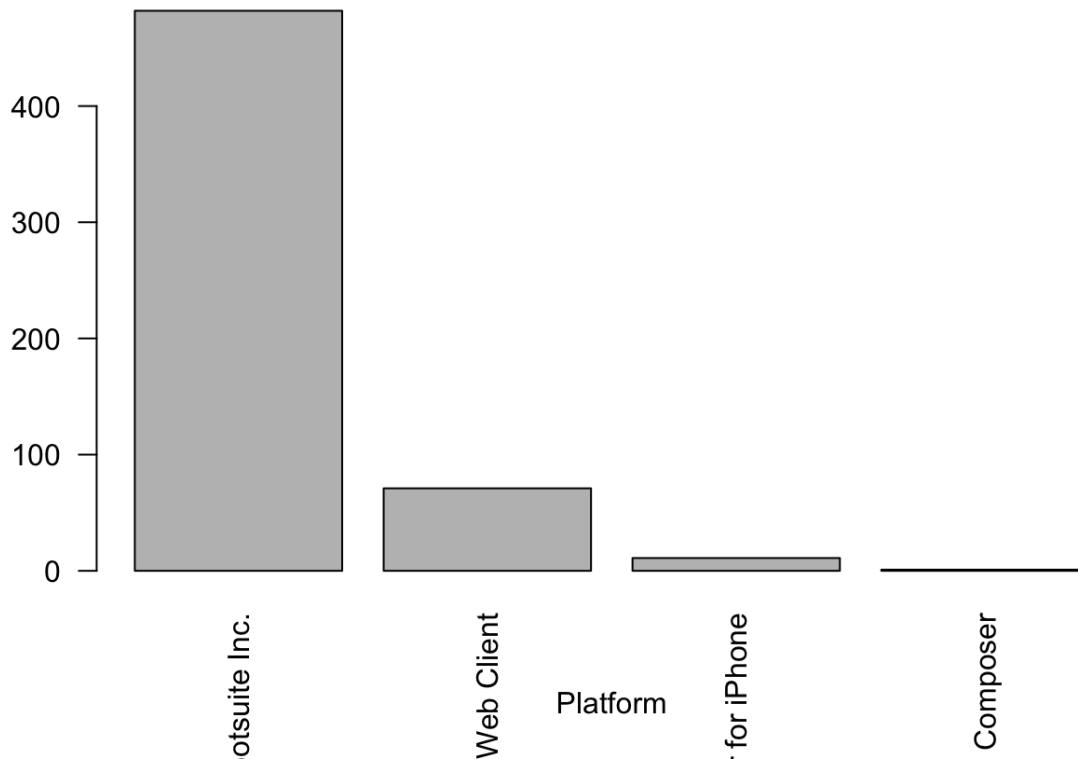
```
## [1] "2018-08-16 19:35:03 UTC"
```

```
max(dstweetsDF$created)
```

```
## [1] "2019-03-15 22:29:29 UTC"
```

```
##Question 1 c)
dstweetsDF$statusSource = substr(dstweetsDF$statusSource, regexpr(">", dstweetsDF$statusSource) + 1, regexpr("</a>", dstweetsDF$statusSource) - 1)
barplot(sort(table(dstweetsDF$statusSource), decreasing = TRUE), xlab = "Platform", las = 2)
```

This study source was downloaded by 100000797999271 from CourseHero.com on 04-24-2023 21:00:27 GMT -05:00



```
(my_datetime = Sys.time())
```

```
## [1] "2019-03-17 15:16:30 PDT"
```

```
format(my_datetime, tz = "America/New_York")
```

```
## [1] "2019-03-17 18:16:30"
```

```
format(my_datetime, tz = "America/Vancouver")
```

```
## [1] "2019-03-17 15:16:30"
```

```
format(my_datetime, tz = "Asia/Shanghai")
```

```
## [1] "2019-03-18 06:16:30"
```

```
as.POSIXct(as.integer(my_datetime), origin = "1970-01-01", tz = "America/New_York")
```

```
## [1] "2019-03-17 18:16:30 EDT"
```

```
as.POSIXct(as.integer(my_datetime), origin = "1970-01-01", tz = "Asia/Shanghai")
```

```
## [1] "2019-03-18 06:16:30 CST"
```

```
as.POSIXct(as.integer(my_datetime), origin = "1970-01-01", tz = "Asia/Shanghai")
```

This study source was downloaded by 10000679799271 from CourseHero.com on 04-24-2023 21:00:27 GMT -05:00

```
## [1] "2019-03-18 06:16:30 CST"
```

```
##Question 2 a)
grep(OlsonNames(),pattern="Montreal",value=TRUE)
```

```
## [1] "America/Montreal"
```

```
grep(OlsonNames(),pattern = "Tokyo", value = TRUE)
```

```
## [1] "Asia/Tokyo"
```

```
grep(OlsonNames(),pattern = "Dubai", value = TRUE)
```

```
## [1] "Asia/Dubai"
```

```
##Question 2 b)
userlocate = getUser("SFU")
userlocate$toDataFrame()$location
```

```
## [1] "British Columbia, Canada"
```

```
##Question 2 c)
grep(OlsonNames(),pattern = "Vancouver", value = TRUE)
```

```
## [1] "America/Vancouver"
```

```
format(my_datetime, tz ="America/Vancouver")
```

```
## [1] "2019-03-17 15:16:30"
```

```
dstweetsDF$created = as.POSIXct(as.numeric(dstweetsDF$created), origin = "1970-01-01", tz =
"America/Vancouver")
```

```
##Question 3 a)
my_datetime
```

```
## [1] "2019-03-17 15:16:30 PDT"
```

```
as.numeric(difftime(my_datetime, trunc(my_datetime,"days"), Sys.timezone(), "hours"))
```

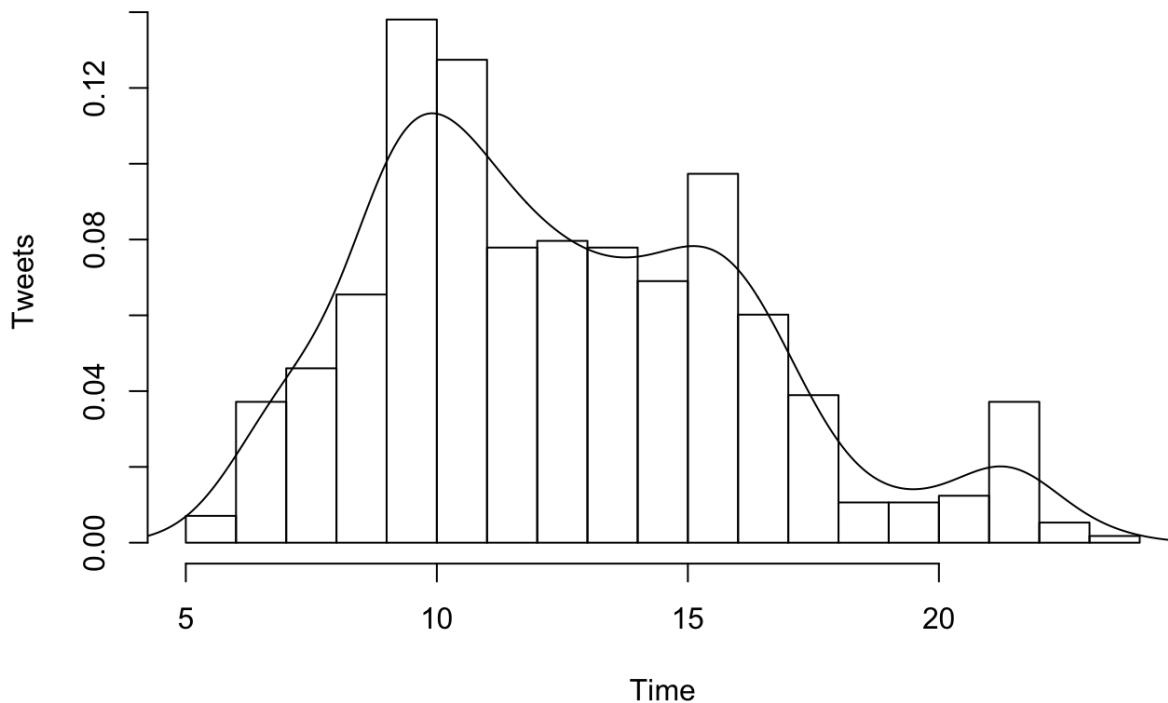
```
## [1] 15.27516
```

```
trunc(my_datetime, "days")
```

```
## [1] "2019-03-17 PDT"
```

```
dstweetsDF$timezone = as.numeric(difftime(dstweetsDF$created, trunc(dstweetsDF$created, "days"), Sys.timezone(), "hours"))
##Question 3 b)
hist(dstweetsDF$timezone, main = "Tweets time in the day from user SFU", breaks = 24, freq = FALSE, ylab = "Tweets", xlab = "Time")
lines(density(dstweetsDF$timezone))
```

### Tweets time in the day from user SFU

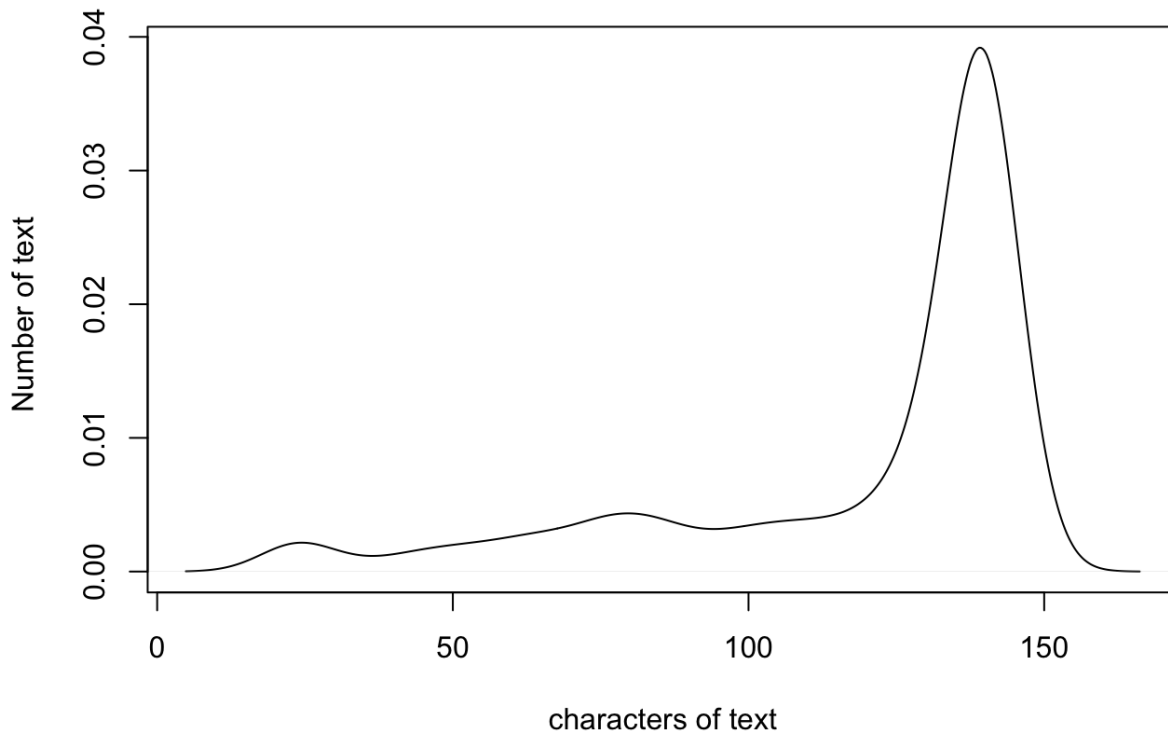


```
x = c("asfef", "qwerty", "yuiop[", "b", "stuff.blah.yech")
nchar(x)
```

```
## [1] 5 6 6 1 15
```

```
##Question 4 a)
plot(density(nchar(dstweetsDF$text)), main = "Distribution of Number of characters", ylab = "Number of text ", xlab = "characters of text")
```

## Distribution of Number of characters



```
##Question 4 b)
dstweetsDF$text[which (nchar(dstweetsDF$text) > 140 & dstweetsDF$created < "2017-09-26")]
```

```
## character(0)
```

```
my_datetime
```

```
## [1] "2019-03-17 15:16:30 PDT"
```

```
cut(my_datetime, "weeks")
```

```
## [1] 2019-03-11
## Levels: 2019-03-11
```

```
cut(my_datetime, "months")
```

```
## [1] 2019-03-01
## Levels: 2019-03-01
```

```
cut(my_datetime, "quarters")
```

```
## [1] 2019-01-01
## Levels: 2019-01-01
```

```
suppressPackageStartupMessages(library(dplyr))
str(iris)
```

This study source was downloaded by 100000797999271 from CourseHero.com on 04-24-2023 21:00:27 GMT -05:00

```
## 'data.frame': 150 obs. of 5 variables:
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
by_species = group_by(iris, Species)
summary_stat = summarise(by_species, Avg.Petal.Length = mean(Petal.Length), Avg.Petal.Width =
mean(Petal.Width))
summary_stat
```

```
## # A tibble: 3 x 3
## Species Avg.Petal.Length Avg.Petal.Width
## <fct> <dbl> <dbl>
## 1 setosa 1.46 0.246
## 2 versicolor 4.26 1.33
## 3 virginica 5.55 2.03
```

```
##Question 5 a)
dstweets1 = userTimeline("Translink", n = 3200)
save(dstweets1, file = "Tweets_week.Rdata")
load("Tweets_week.Rdata")
dstweetsDF1 = twListToDF(dstweets1)
source("getSentimentScore.R")
library(dplyr)
pos = scan("positive-words.txt", what = "character", comment.char = ";")
neg = scan("negative-words.txt", what = "character", comment.char = ";")
length(pos)
```

```
## [1] 2006
```

```
length(neg)
```

```
## [1] 4783
```

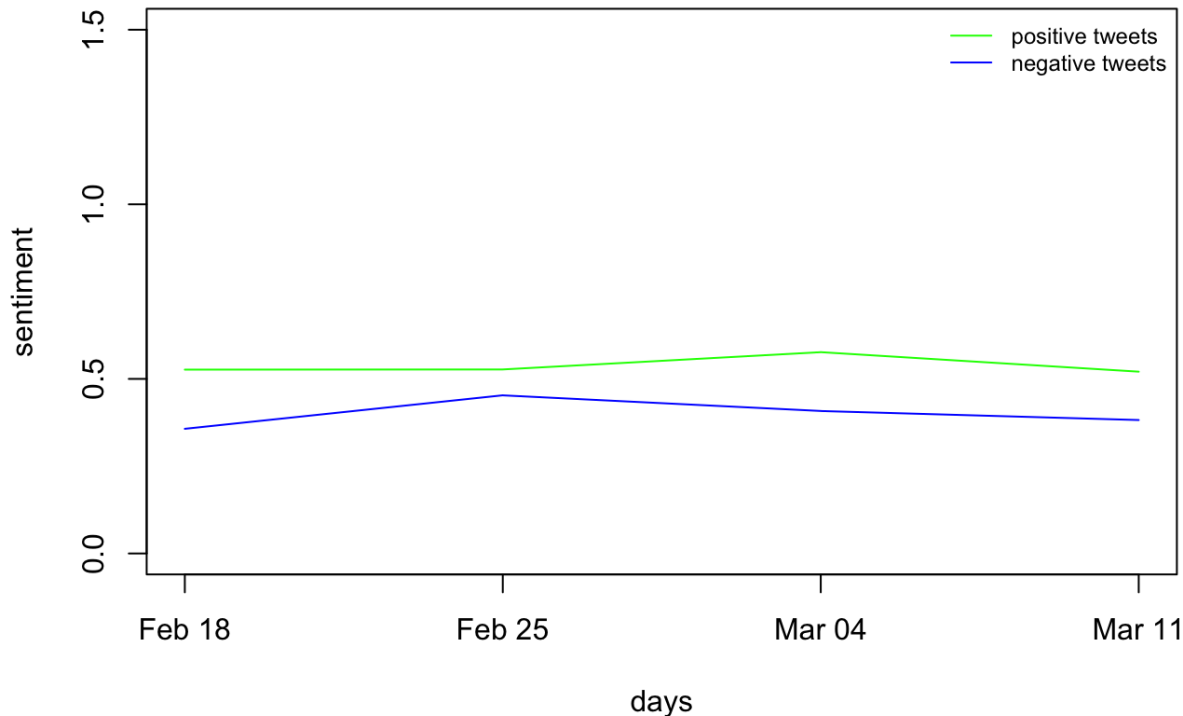
```
neg = c(neg, "wtf")
sresult = getSentimentScore(dstweetsDF1$text, neg, pos)
dstweetsDF1$pos_sentiment = sresult[,1]
dstweetsDF1$neg_sentiment = sresult[,2]
dstweetsDF1$sentiment = sresult[,3]
dstweetsDF1$week = as.Date(cut(dstweetsDF1$created, breaks = "week"))
week_avg = dstweetsDF1 %>% group_by(week) %>% summarise(meanpos = mean(pos_sentiment), meanneg = mean(neg_sentiment), meanSent = mean(sentiment))
head(week_avg, 5)
```

```
## # A tibble: 4 x 4
## week meanpos meanneg meanSent
## <date> <dbl> <dbl> <dbl>
## 1 2019-02-18 0.527 0.357 0.169
## 2 2019-02-25 0.527 0.453 0.0741
## 3 2019-03-04 0.577 0.408 0.169
## 4 2019-03-11 0.521 0.382 0.139
```

```
##Question 5 b)
```

```
plot(week_avg$week, week_avg$meanpos, type = "l", col = "green", ylim = c(0, 1.5), xlab = "days", ylab = "sentiment", main = "The user's positive and negative sentiment score over time")
lines(week_avg$week, week_avg$meanneg, col = "blue")
legend("topright", c("positive tweets", "negative tweets"), lty = 1, col = c("green", "blue"), bty = "n", cex = 0.75)
```

### The user's positive and negative sentiment score over time



```
##Question 6 a)
```

```
library(wordcloud)
text = gsub("http[s]?://[[:alnum:]]\\.\\[/]+", "", dstweetsDF1$text)
text = gsub("(?!(#|@))[[[:punct:]]]", "", text, perl = T)
text = gsub("[[:cntrl:]]", "", text)
words = unlist(strsplit(text, " "))
namehashtags = grep("^#\\w+", unlist(strsplit(text, " ")), value = T)
namemention = grep("^@\\w+", unlist(strsplit(text, " ")), value = T)
hashfreq = table(namehashtags)
menfreq = table(namemention)
pal <- brewer.pal(9, "YlGnBu")
pal <- pal[-(1:4)]
##Question 6 b)
wordcloud(names(hashfreq), hashfreq, scale = c(2, 0.5), min.freq = 5, random.order = FALSE,
  rot.per = 0.1, use.r.layout = FALSE, colors = pal)
```



##Question 6 c)

```
wordcloud(names(menfreq), menfreq, scale = c(2, 0.5), min.freq = 5, random.order = FALSE, rot.per = 0.1, use.r.layout = FALSE, colors = pal)
```

