````
---
title: "Week8Lab"
author: "Frank Shen"
date: "2019/3/14"
output:
  html_document: default
  pdf_document: default
---

```{r}
library(ROAuth)
library(twitteR)
library(wordcloud)
# Declare Twitter API Credentials
consumer_key = "7JEX5oJVhow2gNa0hQfX0lLnA"
consumer_secret = "fGZ77gbFhNB0fQdTt3rQqub2hpnSfD6hx7E7ByCnUnrkntk1Bh"
access_token = "1589845460-VZAJEXduKoqFuPRRXdEUng9Svjjdtg6YOnLolad"
access_secret = "ohu83GNXf3cUPRJbaH241PAdw3K2UnqvxtcM8UsH6gWL7"

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

```{r}
##Question 1 a)
dstweets = userTimeline("SFU", n = 3200)
length(dstweets)
save(dstweets, file = "Tweets_week8.Rdata")
##Question 1 b)
dstweetsDF = twListToDF(dstweets)
names(dstweetsDF)
min(dstweetsDF$created)
max(dstweetsDF$created)
##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)

```

```{r}
(my_datetime = Sys.time())
format(my_datetime, tz = "America/New_York")
format(my_datetime, tz = "America/Vancouver")
format(my_datetime, tz = "Asia/Shanghai")
as.POSIXct(as.integer(my_datetime), origin = "1970-01-01",tz = "America/New_York")
as.POSIXct(as.integer(my_datetime), origin = "1970-01-01",tz = "Asia/Shanghai")
as.POSIXct(as.integer(my_datetime), origin = "1970-01-01",tz = "Asia/Shanghai")

##Qeustion 2 a)
grep(OlsonNames(),pattern="Montreal",value=TRUE)
grep(OlsonNames(),pattern = "Tokyo", value = TRUE)
grep(OlsonNames(),pattern = "Dubai", value = TRUE)
##Question 2 b)
userlocate = getUser("SFU")
userlocate$toDataFrame()$location
##Question 2 c)
grep(OlsonNames(),pattern = "Vancouver", value = TRUE)
````
````

```
format(my_datetime, tz ="America/Vancouver")
dstweetsDF$created = as.POSIXct(as.numeric(dstweetsDF$created), origin = "1970-01-
01", tz ="America/Vancouver")
```
```{r}
##Question 3 a)
my_datetime
as.numeric(difftime(my_datetime, trunc(my_datetime,"days"), Sys.timezone(),
"hours"))
trunc(my_datetime, "days")
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$timezone))
```
```{r}
x = c("asfef", "qwerty", "yuiop[", "b", "stuff.blah.yech")
nchar(x)
##Question 4 a)
plot(density(nchar(dstweetsDF$text)), main = "Distribution of Number of
characters", ylab = "Number of text ", xlab = "characters of text")
##Question 4 b)
dstweetsDF$text[which (nchar(dstweetsDF$text) > 140 & dstweetsDF$created < "2017-
09-26")]
```
```{r}
my_datetime
cut(my_datetime, "weeks")
cut(my_datetime, "months")
cut(my_datetime, "quarters")
suppressPackageStartupMessages(library(dplyr))
str(iris)
by_species = group_by(iris, Species)
summary_stat = summarise(by_species, Avg.Petal.Length =
mean(Petal.Length),Avg.Petal.Width = mean(Petal.Width))
summary_stat
##Question 5 a)
dstweets1 = userTimeline("Translink", n = 3200)
save(dstweets1, file = "Tweets_week.Rdata")
load("Tweets_week.Rdata")
dstweetsDF1 = twListToDF(dstweets1)
source("getSentimentScore.R")
library(dplyr)
pos = scan("positive-words.txt", what = "character", comment.char = ";")
neg = scan("negative-words.txt", what = "character", comment.char = ";")
length(pos)
length(neg)
neg = c(neg, "wtf")
sresult = getSentimentScore(dstweetsDF1$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)
##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)

```
```{r}
##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)
```
```
```