Homework 8 – Due Monday November 9, 2015

Cheryl Calhoun

11/02/2015

```
## Setting up the work environment and twitteR API
##setwd("C:/Users/07001412/OneDrive/Education/UF/2015/Fall/EDF6938/Week 10")
library("twitteR") ##, lib.loc="~/R/win-library/3.2")
library("base64enc") ##, lib.loc="~/R/win-library/3.2")
library("dplyr") ##, lib.loc="~/R/win-library/3.2")
##download.file(url="http://curl.haxx.se/ca/cacert.pem", destfile="cacert.pe
m") ##-- this was suggested if you are running a Windows machine.

## Loading twitter API and Access keys.
source ("twitter-access-keys.R")
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)
```

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

```
## Setting search patterns.
usertag <- "@[A-z_0-9]+"
hashtag <- "#[A-z0-9]+"
```

Gathering data on Football related tweets at the University of Florida

Context: Every member of the class was assigned a school in the Southeastern Conference (SEC) as specified in the spreadsheet located at: [https://docs.google.com/spreadsheets/d/1IKRXc0hN1C9e5S845LmgY-rlLdCQ1xOQrI3thxKIgVA/edit?usp=sharing (https://docs.google.com/spreadsheets/d/1IKRXc0hN1C9e5S845LmgY-rlLdCQ1xOQrI3thxKIgVA/edit?usp=sharing)]

For this exercise, data will be captured from twitter each week, saved to a file, and reload from that file as needed. This will ensure both reproducibility and ease of use on Twitter’s servers.

The original plan was to use geolocation to find our tweets. To do this, we need to find the latitude and longitude of our school on Google Maps. (This is a manual process of looking up the coordinates.) The latitude and longitude of Ben Hill Griffin stadium at the University of Florida is: 29.649898, -82.348429. Now, we can use these coordinates to extract a sample of tweets from near that location. For this exercise, we will use a 5 mile radius.

Unfortunately, geolocation does not seem to work, so we will use the most common hasgtag associated with Florida Sports and Florida Football. After some initial data exploration, the #GoGators hashtag is selected.

There are four remaining games this season. They are:

Game #	Date	Opponent	Location
Game 9	Nov 7	Vanderbilt	Gainesville
Game 10	Nov 14	S. Carolina	Columbia
Game 11	Nov 21	Fl. Atlantic	Gainesville
Game 12	Nov 28	Florida State	Gainesville

```
## Collect weekly tweets beginning Thursday before game and ending on Sunday after the game. Store tweets in a .csv file.  This code will be executed on a weekly basis until we have gathered data for the remaining 4 games of the season.

##The original geolocation code.
## Florida <- searchTwitter('', geocode='29.649898,-82.348429,5mi', since="2015-11-05", until="2015-11-09", n=10000)

##The updated #GoGators hashtag code.
##Florida <- searchTwitter("#GoGators", since="2015-11-05", until="2015-11-09", n=10000)
##Florida.df <- rbind_all (lapply (Florida, function(rr) rr$toDataFrame()))
##write.csv(Florida.df, file = "Vanderbilt.csv")

## Collect data for game 10
##Florida <- searchTwitter("#GoGators", since="2015-11-12", until="2015-11-16", n=10000)
##Florida.df <- rbind_all (lapply (Florida, function(rr) rr$toDataFrame()))
##write.csv(Florida.df, file = "Carolina.csv")

## Collect data for game 11
##Florida <- searchTwitter("#GoGators", since="2015-11-19", until="2015-11-23", n=10000)
##Florida.df <- rbind_all (lapply (Florida, function(rr) rr$toDataFrame()))
##write.csv(Florida.df, file = "FLAtl.csv")

## Collect data for game 12
##Florida <- searchTwitter("#GoGators", since="2015-11-26", until="2015-11-30", n=10000)
##Florida.df <- rbind_all (lapply (Florida, function(rr) rr$toDataFrame()))
##write.csv(Florida.df, file = "FloridaSt.csv")

#Read previously stored data from data file.
Game9 <- read.csv("Vanderbilt.csv")
## Game10 <- read.csv("Carolina.csv")
## Game11 <- read.csv("FLAtl.csv")
## Game12 <- read.csv("FloridaSt.csv")
```

Game 9: Vanterbilt

Evaluating HashTags

Determining the number of hashtags used in each tweet and overall in the sample.

```
## Find hashtags for Game 9. Add a column for hashtags and a column for number
of hashtags.
Game9 <- mutate (Game9, hashtags=regmatches(Game9$text, gregexpr(hashtag, Game9
$text)))
Game9 <- mutate (Game9, HQty=as.integer(lapply(Game9$hashtags, function(x) leng
th(x))))
```

The number of tweets at each hashtag level

```
## Create a Hashtags per Tweet table.
HashtagsperTweet = table(Game9$HQty)
HashtagsperTweetTable = as.data.frame(HashtagsperTweet)
names(HashtagsperTweetTable)[1] = 'Number of Hashtags'
names(HashtagsperTweetTable)[2] = 'Number of Tweets'
HashtagsperTweetTable
```

```
##    Number of Hashtags Number of Tweets
## 1                   0               21
## 2                   1             4760
## 3                   2             3478
## 4                   3             1102
## 5                   4              365
## 6                   5              153
## 7                   6               58
## 8                   7               34
## 9                   8               19
## 10                  9                7
## 11                 10                1
## 12                 12                1
## 13                 14                1
```

The total number of hashtags in the data set.

```
## Determine total number of hashtags.
all.hashtags <- unlist(regmatches(Game9$text, gregexpr (hashtag, Game9$text)))
## all.hashtags
length(all.hashtags)
```

```
## [1] 18084
```

The total number of unique hashtags in the data set.

```
## Determine total number of unique hashtags.
unique.hashtags <- unique(all.hashtags <- unlist(regmatches(Game9$text, gregexp
r (hashtag, Game9$text))))
length(unique.hashtags)
```

```
## [1] 1109
```

Finding the most frequently used hashtags.

```
## Count the hashtag usage.
hashtags.df <- data.frame(cbind(all.hashtags))
hashcount <- count(hashtags.df, all.hashtags)
hashcount <- hashcount[order(-hashcount$n),]
hashcount
```

```
## Source: local data frame [1,109 x 2]
##
##          all.hashtags    n
## 1           #GoGators 9068
## 2          #VANDYvsUF 1477
## 3                #SEC  747
## 4           #gogators  645
## 5  #GatorsHeismanDay  486
## 6        #GatorNation  363
## 7             #Gators  330
## 8          #BeatVandy  273
## 9  #SECEastChampions  266
## 10            #TurnUp  240
## ..                ...  ...
```

Evaluating UserTags

Determining the number of users tagged in each tweet.

```
## Find users for game 9.Add a column for usertags and a column for number of u
ser tags in the tweet.
##Game9 <- mutate (Game9, usertags=regmatches(Game9$text, gregexpr(users, Game9
$text)))
##Game9 <- mutate (Game9, UQty=as.integer(lapply(Game9$users, function(x) lengt
h(x))))

Game9 <- mutate (Game9, usertags=regmatches(Game9$text, gregexpr(usertag, Game9
$text)))
Game9 <- mutate (Game9, UQty=as.integer(lapply(Game9$usertags, function(x) leng
th(x))))
```

The number of users tagged per tweet.

```
## Create a Hashtags per Tweet table.
UserTagsperTweet = table(Game9$UQty)
UserTagsperTweetTable = as.data.frame(UserTagsperTweet)
names(UserTagsperTweetTable)[1] = 'Number of Users'
names(UserTagsperTweetTable)[2] = 'Number of Tweets'
UserTagsperTweetTable
```

```
##   Number of Users Number of Tweets
## 1               0             2944
## 2               1             5893
## 3               2              865
## 4               3              212
## 5               4               26
## 6               5               59
## 7               6                1
```

The total number of usertags in the data set.

```
## Determine total number of hashtags.
all.usertags <- unlist(regmatches(Game9$text, gregexpr (usertag, Game9$text)))
## all.hashtags
length(all.usertags)
```

```
## [1] 8664
```

The total number of unique usertags in the data set.

```
## Determine total number of unique hashtags.
unique.usertags <- unique(all.usertags <- unlist(regmatches(Game9$text, gregexp
r (usertag, Game9$text))))
length(unique.usertags)
```

```
## [1] 788
```

Finding the most commonly used usertags.

```
## Count the hashtag usage.
usertags.df <- data.frame(cbind(all.usertags))
usercount <- count(usertags.df, all.usertags)
usercount <- usercount[order(-usercount$n),]
usercount
```

```
## Source: local data frame [788 x 2]
##
##      all.usertags    n
## 1       @GatorsFB 1923
## 2  @FloridaGators 1643
## 3  @CoachMcElwain  524
## 4    @AlbertGator  467
## 5  @GatorsGameday  419
## 6     @ImShmacked  263
## 7             @UF  251
## 8       @ufalumni  244
## 9   @GatorsSoccer  120
## 10  @Jakeallen_14   81
## ..            ...  ...
```

Gathering Data for Games 10-12 - Coming soon…

Collect a set of 10000 tweets for each of the last five ***Thursday-Saturday*** blocks on which a football game was played by the University of Florda. Repeat the hashtag collection and user tagging exercises for each of these samples. Do the same users appear to produce the same volume each week? Do the same secondary hashtags appear?

```
## Games 10-12 coming soon...
```

The following 25 users tweeted most frequently during the game 9 period. The usertags were selected after eliminating users that were clearly accounts for the team proper or their PR department. That is, we're trying to find a fan community for your team that tweets about their team's games on a regular basis. This community will form the basis for the project looking ahead.

```
## Count the hashtag usage.
frequenttweeters <- summarise(group_by(Game9, screenName), count=n())
frequenttweeters <- frequenttweeters[order(-frequenttweeters$count),]
head(frequenttweeters, 25)
```

```
## Source: local data frame [25 x 2]
##
##         screenName count
## 1        SECstagram    79
## 2      HotCorner_10    78
## 3     FloridaGators    41
## 4      gogators1974    30
## 5     MrMarques2727    26
## 6           Trippo7    25
## 7          Tebow815    23
## 8       CliffWilkin    22
## 9   _Whoa_itsPayge_    20
## 10        GatorsSRH    16
## ..              ...   ...
```

The top 25 tweeters are:

SECstagram, HotCorner_10, gogators1974, MrMarques2727, Trippo7, Tebow815, CliffWilkin, *Whoa_itsPayge*, jhern2498, TampaBaySRH, 2103_amber, bmanning96, major1029, ClaireabellGatr, CVan_Huss, hunterlynch29, krizielyvonne, REGarrison, ktharris, rebeccavelaz17, TOTODUVALDIVA, BrookinsOneil, ChristophersZen, Emmy_ArmadaFC, Gatorfan187