HW7B

Milad Tatari

11/3/2019

This tutorial shows you how to use the Twitter Streaming API to get tweet data using R. You can retrieve Twitter tweet data in JSON format through the Twitter Web API external application after registering the application with Twitter. This code did not work, so I am going to provide 8 solutions that I tried and it did not work, I mean hanshake function. Finally, I will extract tweeter data by searchTweeter function.

```
setwd("C:/Users/milad/OneDrive/Desktop/data science/DA5020/HW7B")

library(RCurl)
```

```
## Loading required package: bitops
```

```
#install.packages("ROAuth")
library(ROAuth)
#install.packages("streamR")
library(streamR)
```

```
## Loading required package: rjson
```

```
## Loading required package: ndjson
```

```
#install.packages("twitteR")
library(twitteR)

library(rjson)

options(RCurlOptions = list(cainfo = system.file("CurlSSL", "cacert.pem", package = "RCurl")))

   download.file(url="http://curl.haxx.se/ca/cacert.pem",ssl.verifypeer = FALSE, destfile="cacert.pem")

download.file(url="http://curl.haxx.se/ca/cacert.pem", destfile="cacert.pem")

outFile <- "tweets_sample.json"
requestURL        <- "https://api.twitter.com/oauth/request_token"
accessURL         <- "https://api.twitter.com/oauth/access_token"
authURL           <- "https://api.twitter.com/oauth/authorize"
consumerKey       <- "XrRQO2YYQTRNGpTendMVIAeqE"
consumerSecret    <- "E074LwAYzK8AyayMcP32Iuov1sW1wOUqgmELfmHNC6v0Yb1vnG"
accessToken       <- "1186757776556482560-CJAXaIcSSZ2OWxy0V9RMVqJinlI1xd"
accessSecret <- "V7PXqzr6o4oZnn8PPqtKsnguJWmOVTYylrdFAgpjfYfjE"

#obtain oauth by handshaking and save the oauth to the local disk for future connections
myoauth <- OAuthFactory$new( consumerKey=consumerKey,
                             consumerSecret=consumerSecret,
                             requestURL=requestURL,
                             accessURL=accessURL,
```

```
                                authURL=authURL,needsVerifier=FALSE)

setup_twitter_oauth(consumerKey ,consumerSecret,accessToken ,accessSecret)
```

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

```
#myoauth$handshake(cainfo = system.file("CurlSSL", "cacert.pem", package = "RCurl"))


#tweets <- twitteR::searchTwitter("#trump",n =12,lang ="en",since = '2018-11-08')
#strip retweets
#strip_retweets(tweets)

#sampleStream( file=outFile, oauth=my_oauth, tweets=100 )




#options(RCurlOptions = list(cainfo = system.file("CurlSSL", "cacert.pem", package = "RCurl")))


#my_oauth$handshake(cainfo="cacert.pem")


library("openssl")
library("httpuv")
#setup_twitter_oauth(consumer_key, consumer_secret, access_token , access_secret)
#origop <- options("httr_oauth_cache")
#options(httr_oauth_cache = TRUE)

#CD_Timeline <- userTimeline(user = "ChancellorDav", n = 3200, includeRts = FALSE, excludeReplies = TRU
#toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))

#Sapply Approach
#CD_tweets <- sapply(CD_Timeline, function(x) { strsplit(gsub("[^[:alnum:] ]", "", x$text), " +")[[1]] }

#Overview of Data Scraped
#head(CD_tweets)
#tail(CD_tweets)
#str(CD_tweets)

# returns the oauth
#myoauth$handshake()



# installing/loading the latest installr package:
#install.packages("installr")
#library(installr) # install+load installr
#updateR() # updating R.
```

Step 1: I am getting an error as this:

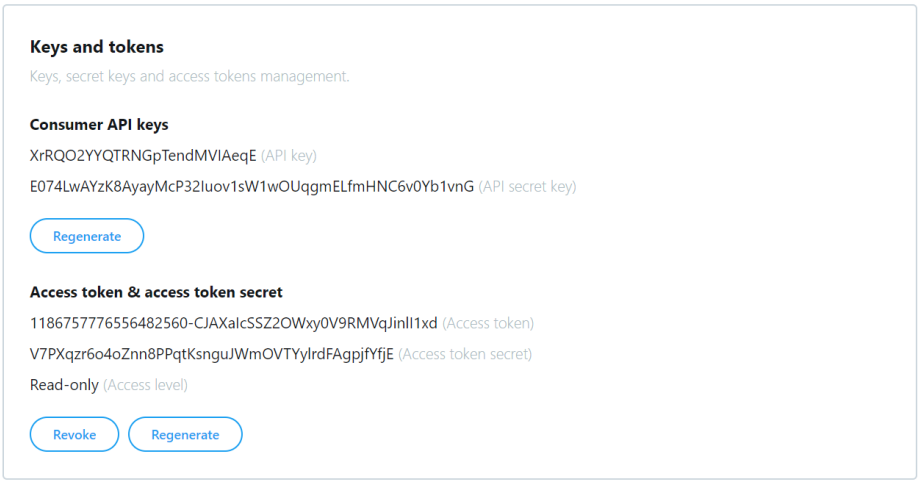

Figure 1: Accesing to tweeter

```
> myoauth$handshake(cainfo = system.file("CurlSSL", "cacert.pem", package = "RCurl"))
Error in function (type, msg, asError = TRUE)  :
  Unknown SSL protocol error in connection to api.twitter.com:443
```

Figure 2: Error in handshake function

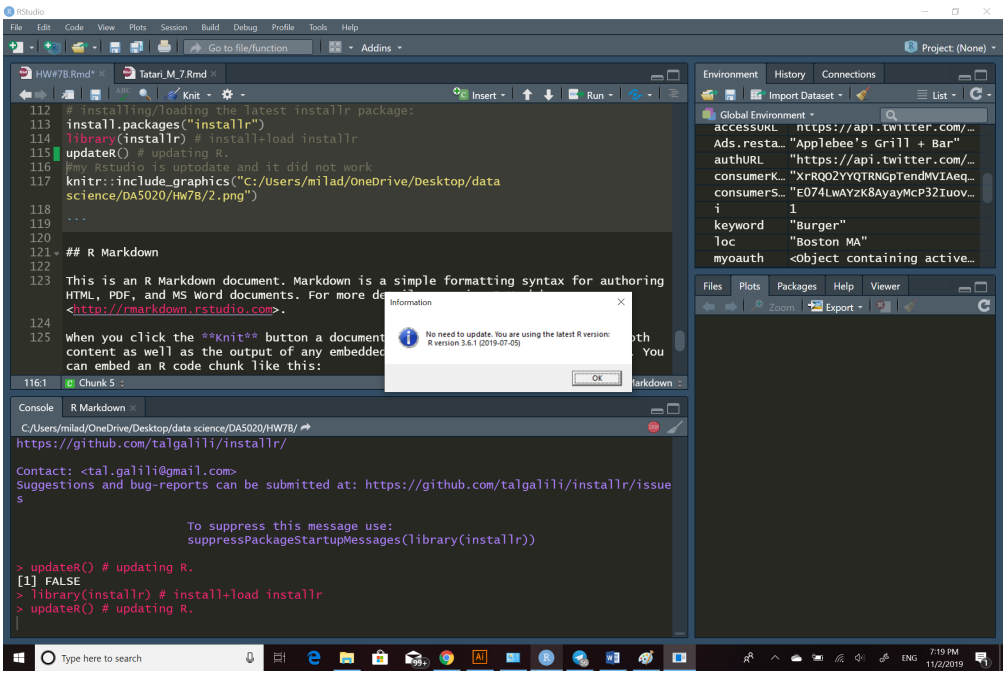

Figure 3: Solution 1,Error

Step2: I struggled a lot, searched a lot to find a solution that works to connect with twitter using handshake, I could not do it. I have to mention that I was able to connect with twitter with other functions that will be mentioned later. Here are a couple of solutions that did not work for the handshake.

Solution 1: is to update R

Solution 2: Running without oauth setup, sth like this:

Solution 3: Getting appropriate libraries: still did not work

Solution 4: doing authentication in cmd, it did not work:

Solution 5: Getting the following libraries, but still it did not work

Solution 6: Installing StreamR from github library(devtools) devtools::install_github("pablobarbera/streamR/streamR")

Soltion 7: Updating all the packages:

Solution 8: I have read a lot of forums and discusions, I notice that it is a very common probles, I did my best solve the issure also using other suggestions in the litrature, but it did not solve. It includes installation of other packages as well. I also did authentication of streamR.

Finally, this is the working solution: we want to extract the tweeter of President Trump for example:

```
#install.packages("rtweet")
library(rtweet)
```

```
##
## Attaching package: 'rtweet'
```

```
## The following object is masked from 'package:twitteR':
##
##     lookup_statuses
```

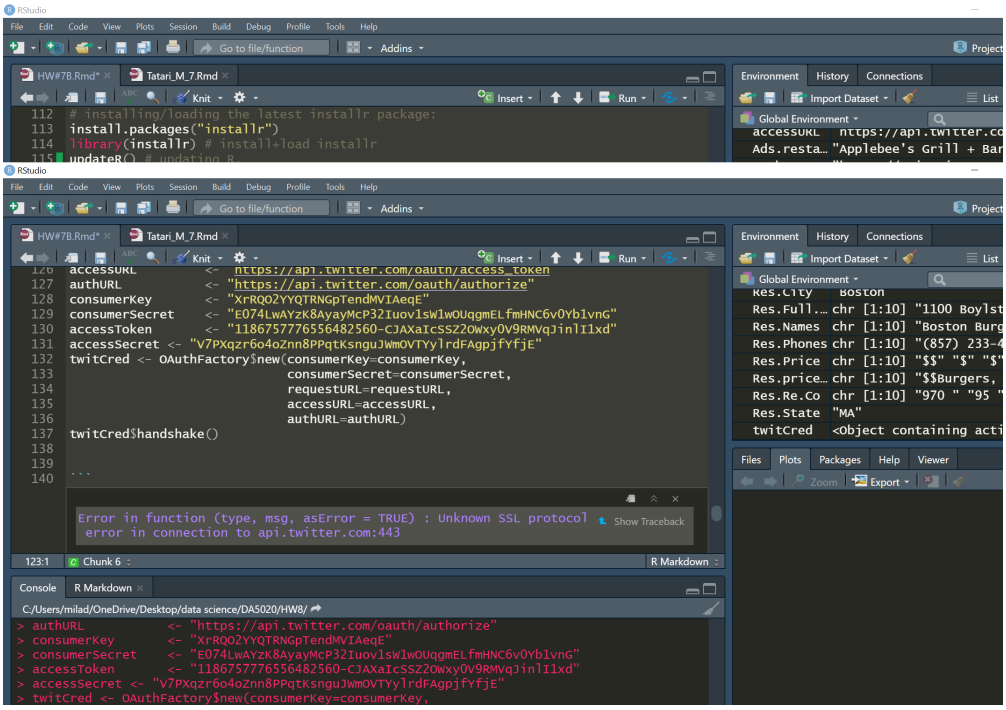

Figure 4: Solution 2,Error

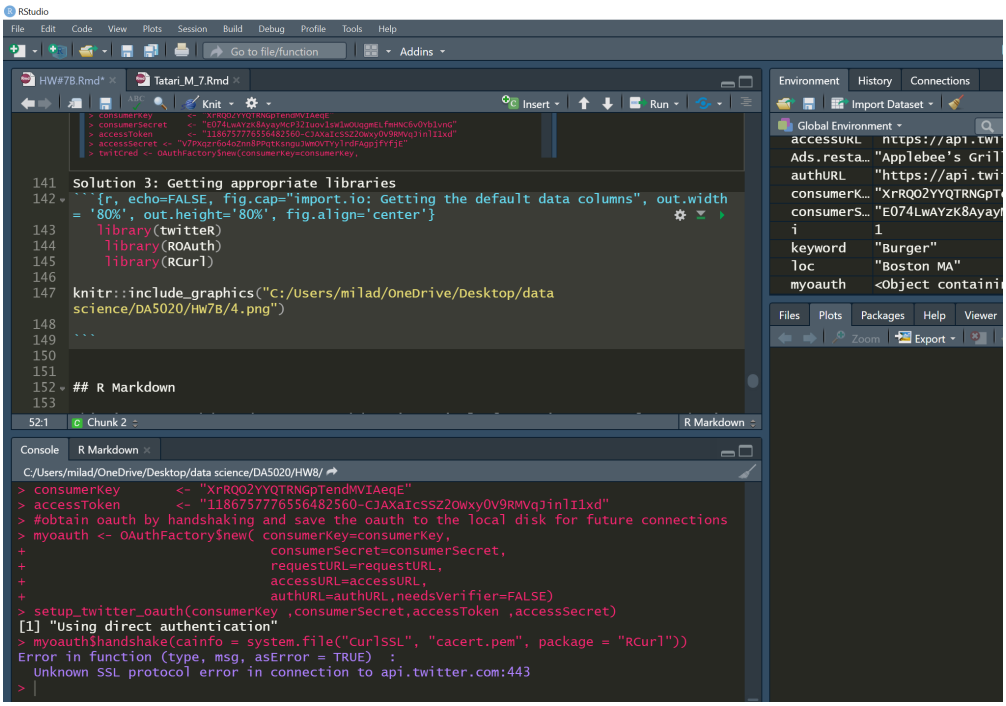

Figure 5: Solution 3,Error

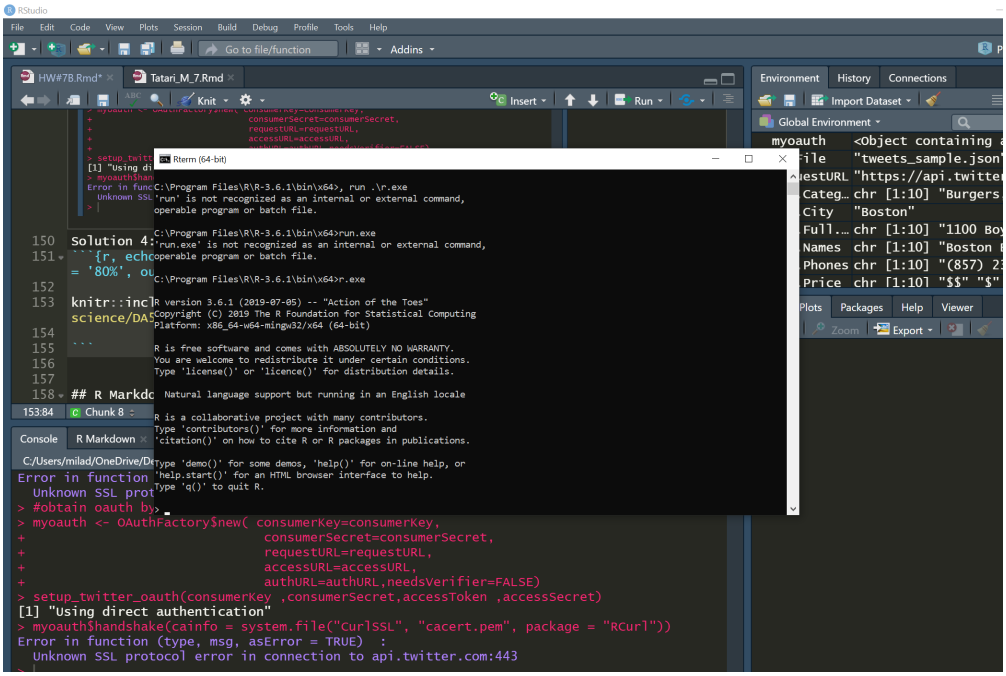

Figure 6: Solution 4,Error

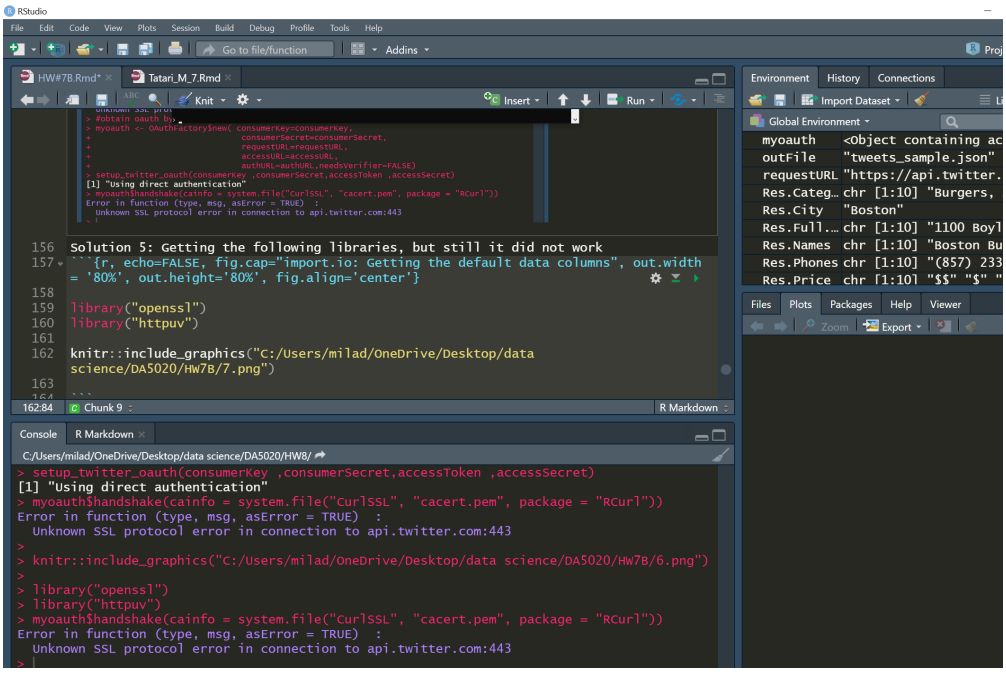

Figure 7: Solution 5,Error

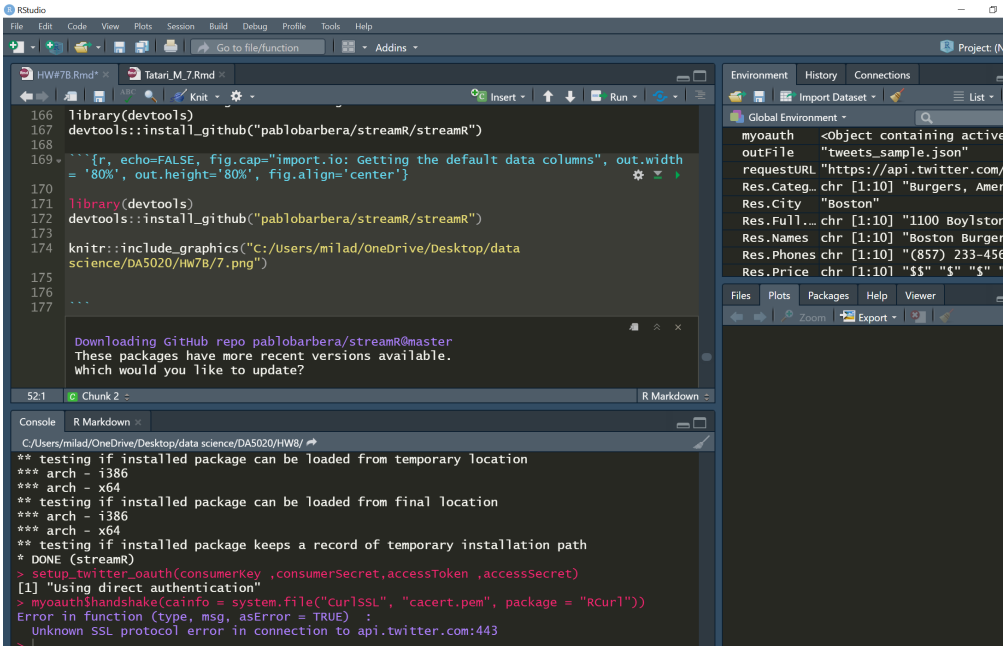

Figure 8: Solution 6,Error

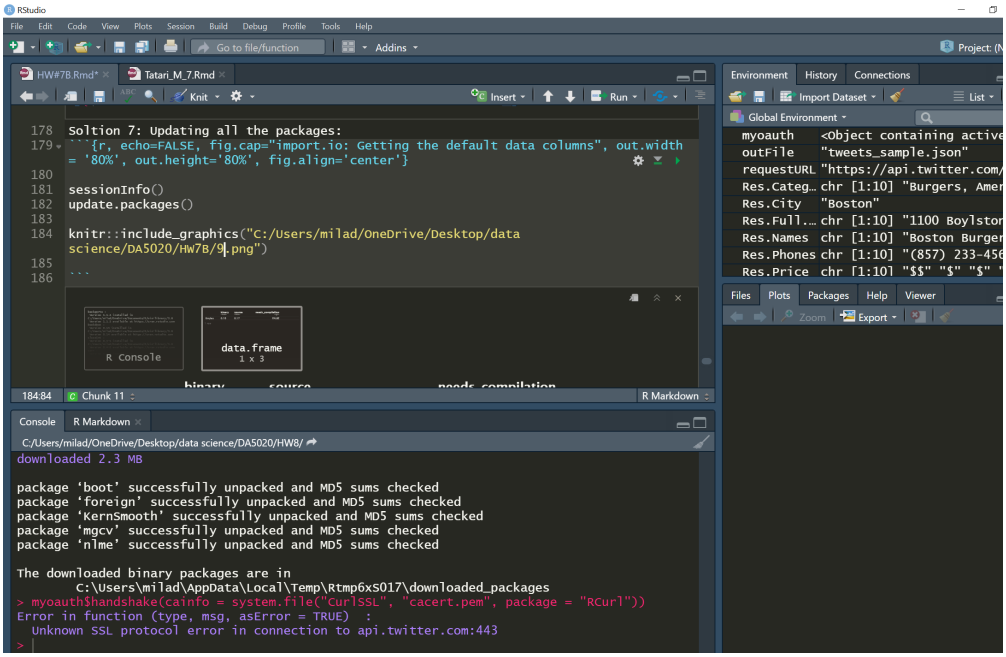

Figure 9: Solution 7,Error

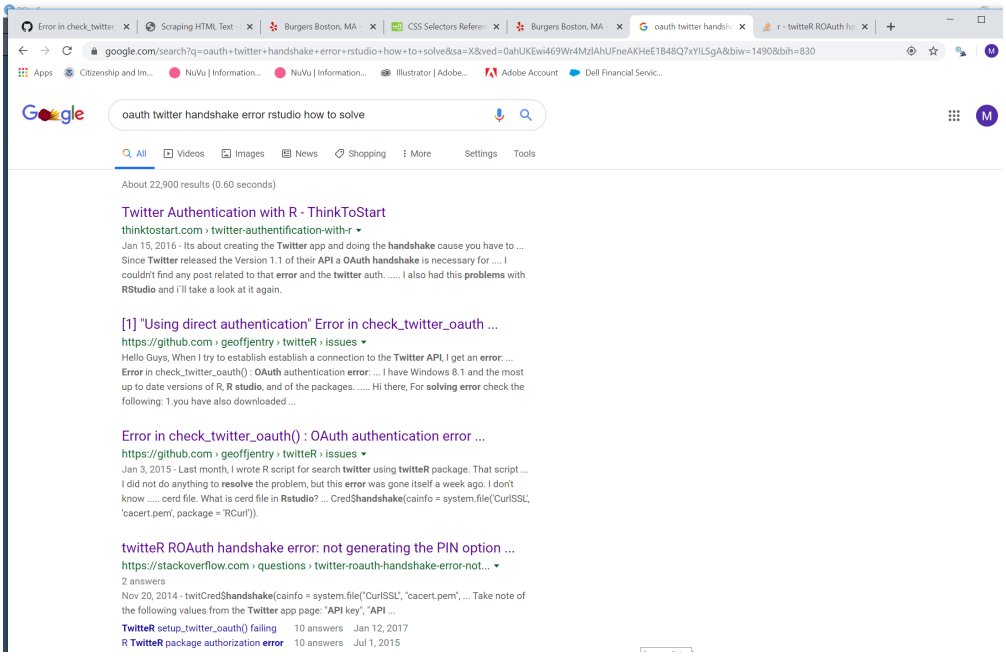

Figure 10: Solution 8,Other searches

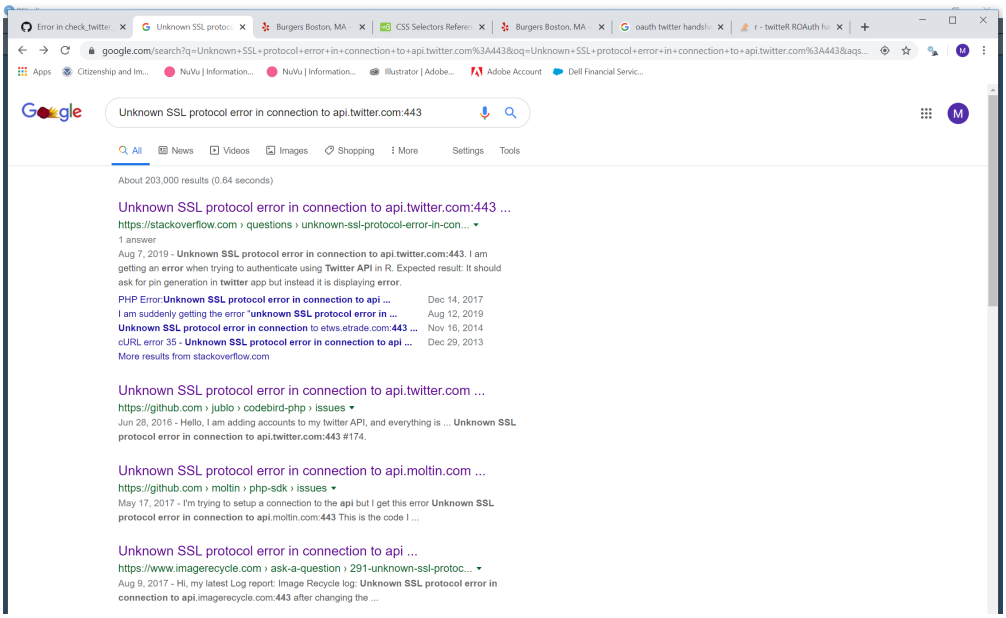

Figure 11: Solution 8,Other searches

```
## The following object is masked from 'package:ndjson':
##
##     flatten
```

```
# plotting and pipes - tidyverse!
library(dplyr)
```

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

```
## The following objects are masked from 'package:twitteR':
##
##     id, location
```

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

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

```
# text mining library
#install.packages("tidytext")
library(tidytext)
options(RCurlOptions = list(cainfo = system.file("CurlSSL", "cacert.pem", package = "RCurl")))

   download.file(url="http://curl.haxx.se/ca/cacert.pem",ssl.verifypeer = FALSE, destfile="cacert.pem")

download.file(url="http://curl.haxx.se/ca/cacert.pem", destfile="cacert.pem")

outFile <- "tweets_sample.json"
requestURL        <- "https://api.twitter.com/oauth/request_token"
accessURL         <- "https://api.twitter.com/oauth/access_token"
authURL           <- "https://api.twitter.com/oauth/authorize"
consumerKey       <- "XrRQO2YYQTRNGpTendMVIAeqE"
consumerSecret    <- "E074LwAYzK8AyayMcP32Iuov1sW1wOUqgmELfmHNC6v0Yb1vnG"
accessToken       <- "1186757776556482560-CJAXaIcSSZ2OWxy0V9RMVqJinlI1xd"
accessSecret <- "V7PXqzr6o4oZnn8PPqtKsnguJWmOVTYylrdFAgpjfYfjE"


myoauth <- OAuthFactory$new( consumerKey=consumerKey,
                             consumerSecret=consumerSecret,
                             requestURL=requestURL,
                             accessURL=accessURL,
                             authURL=authURL,needsVerifier=FALSE)

setup_twitter_oauth(consumerKey ,consumerSecret,accessToken ,accessSecret)
```

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

```r
Tweets<- searchTwitter("trump", n=10,lang = "en")
head(Tweets)
```

```
## [[1]]
## [1] "virraj: @nytimes posits you can reach the figure of 66 million Twitter followers from a base of
##
## [[2]]
## [1] "Jon11996642: RT @charliekirk11: The Ukraine Call:\n\n1,990 words\n\n121 sentences\n\nOver 10 mi
##
## [[3]]
## [1] "FoxMcElroy: @Scallywag1967 @whtgld @itsJeffTiedrich @realDonaldTrump @GavinNewsom Except they d
##
## [[4]]
## [1] "29361RMSM: RT @CREWcrew: No matter how much the Trump sons repeat themselves saying otherwise,
##
## [[5]]
## [1] "Kathrynjeansays: RT @charliekirk11: Election Day is in one year\n\nA historically successful Pr
##
## [[6]]
## [1] "Stewar1W: Trump didn’t wright this piece, it sound like someone with half a brain wrote it! So
```