

TWITTER ANALYSIS ON RESERVATION

OVERVIEW

- Twitter.com is a popular microblogging website.
- Each tweet is 140 characters in length
- Tweets are frequently used to express a tweeter's emotion on a particular subject.
- There are firms which poll twitter for analysing sentiment on a particular topic.
- The challenge is to gather all such relevant data, detect and summarize the overall sentiment on a topic.

Classification Tasks and Tools:

- Polarity classification – positive or negative sentiment
- 3-way classification – positive/negative/neutral
- 10,000 unigram features – baseline
- 100 twitter specific features
- A tree kernel based model
- A combination of models.
- A hand annotated dictionary for emoticons and acronyms

About twitter and structure of tweets:

- 140 charactes – spelling errors, acronyms, emoticons, etc.
- @ symbol refers to a target twitter user
- # hashtags can refer to topics
- 11,875 such manually annotated tweets
- 1709 positive/negative/neutral tweets – to balance the training data

Problem Statement

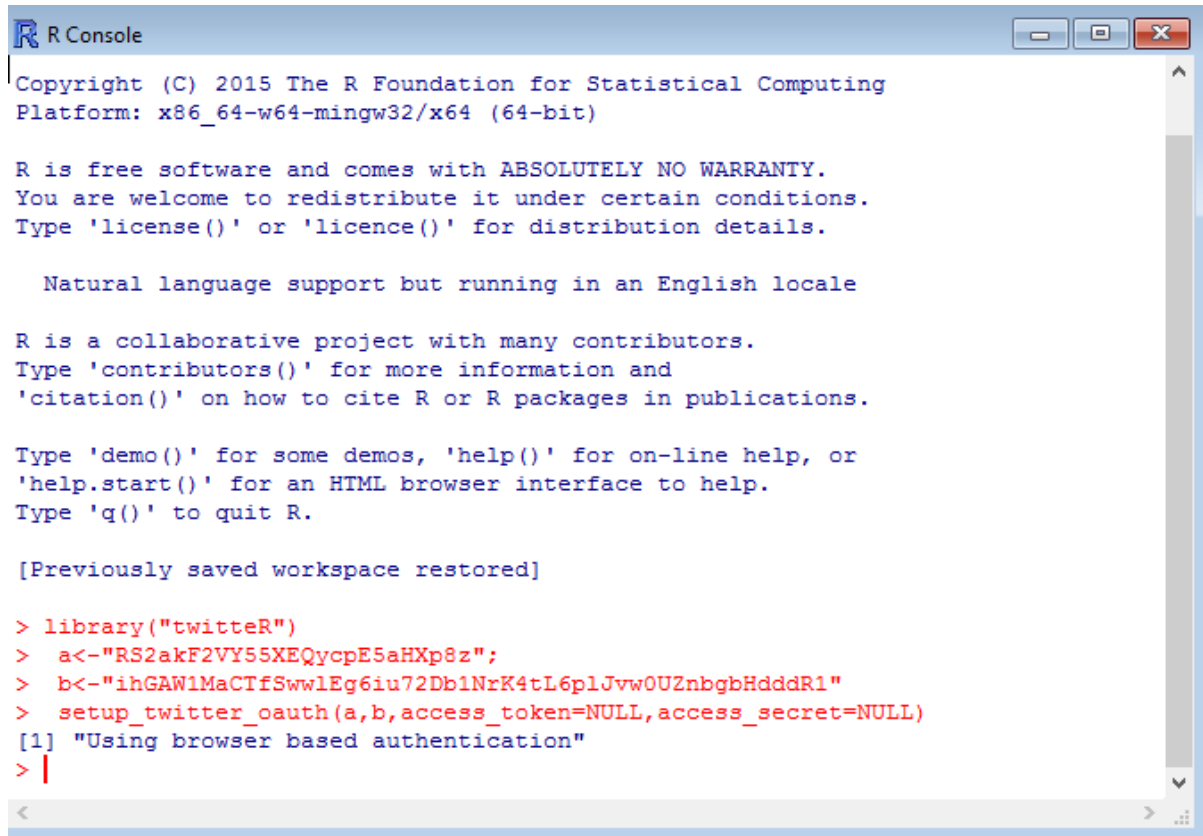
The problem In sentiment analysis is classifying the polarity of a given text. Whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative , neutral.

- ▶ Sentiment analysis over Twitter offer organisations a fast and effective way to monitor the publics' feelings towards their brand, business, directors, etc. A wide range of features and methods for training sentiment classifiers for Twitter datasets have been researched in recent years with varying results.
- ▶ To implement an algorithm for automatic classification of text into positive ,negative , neutral tweets.
- ▶ Analysis is used for determining the attitude of mass towards the topic of intrest.

CODE

TWITTER ANALYSIS

```
setwd("D:/LSDB");  
library('twitter')  
a<-"RS2akF2VY55XEQycpE5aHXp8z";  
b<-"ihGAW1MaCTfSwwlEg6iu72Db1NrK4tL6pIJvw0UZnbgbHdddR1"  
setup_twitter_oauth(a,b,access_token=NULL,access_secret=NULL)
```



```
R Console

Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

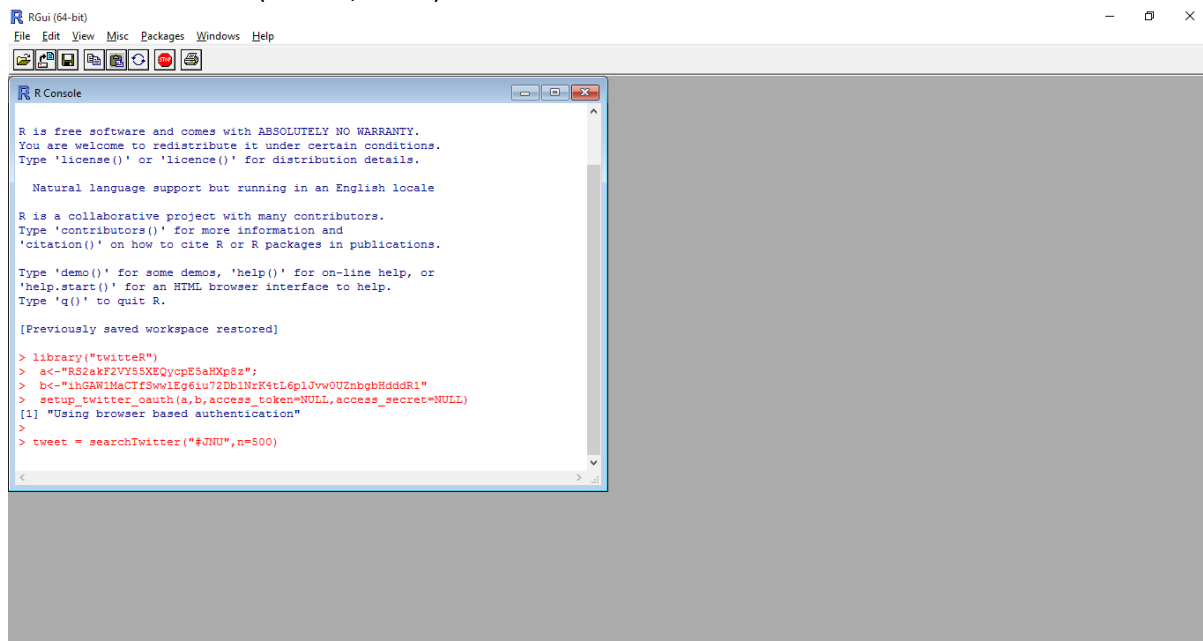
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> library("twitterR")
> a<-"RS2akF2VY55XEQycpE5aHXp8z";
> b<-"ihGAW1MaCTfSwwlEg6iu72Db1NrK4tL6plJvw0UZnbgbHdddR1"
> setup_twitter_oauth(a,b,access_token=NULL,access_secret=NULL)
[1] "Using browser based authentication"
> |
```

tweet=searchTwitter("#JNU",n=500)



```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> library("twitterR")
> a<-"RS2akF2VY55XEQycpE5aHXp8z";
> b<-"ihGAW1MaCTfSwwlEg6iu72Db1NrK4tL6plJvw0UZnbgbHdddR1"
> setup_twitter_oauth(a,b,access_token=NULL,access_secret=NULL)
[1] "Using browser based authentication"
>
> tweet = searchTwitter("#JNU",n=500)
```

```
R Console
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> library("twitterR")
> a<-"RS2akF2VY55XEQycpE5aHXp8z";
> b<-"ihGAW1MaCTfSwWlEg6iu72Db1NrK4tL6p1Jvw0UZnbgbHdddR1"
> setup_twitter_oauth(a,b,access_token=NULL,access_secret=NULL)
[1] "Using browser based authentication"
>
> tweet = searchTwitter("#JNU",n=500)
> length(tweet)
[1] 500
> |
```

```
tweets <- searchTwitter("#nba", n=1499, cainfo="cacert.pem", lang="en")
tweets.text <- sapply(tweets, function(x) x$getText())
```

```
R Console
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> library("twitterR")
> a<-"RS2akF2VY55XEQycpE5aHXp8z";
> b<-"ihGAW1MaCTfSwWlEg6iu72Db1NrK4tL6p1Jvw0UZnbgbHdddR1"
> setup_twitter_oauth(a,b,access_token=NULL,access_secret=NULL)
[1] "Using browser based authentication"
>
> tweet = searchTwitter("#JNU",n=500)
> length(tweet)
[1] 500
> write.csv(tweets, file = "jnu1.csv")
Error in as.data.frame.default(x[[i]], optional = TRUE) :
  cannot coerce class "structure("status", package = "twitterR")" to a data.frame
```

```
tweets = lapply(tweet, function(t) t$toDataFrame())
```

```

R Console
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
>
> library("plyr")

Attaching package: 'plyr'

The following object is masked from 'package:twitterR':

    id

> tweets = ldply(tweet, function(t) t$toDataFrame())
>

```

write.csv(tweets, file = "jnuharsh.csv")

	A	B	C	D	E	F	G	H	I	J	K	L	M
		text	statusSource	screenName	retweetCount	longitude	latitude						
1		RT @bholeswar7799: @ArvindKejriwal	<a href="http://twitter.coi.someshwar_dilip		22	NA	NA						
2		2 Aaj #JNU ke social science and other intellectual	<a href="http://twitter.coi.Ashok_Magadh		0	NA	NA						
3		3 Why the Law on #Sedition Should Be Repealed?	<a href="http://publicize.\\mediaamalendu		0	NA	NA						
4		4 RT @ABVPKonkan: #TruthOfJNU yesterday orag	<a href="http://twitter.coi.vikasgahlot315		22	NA	NA						
5		5 RT @bholeswar7799: @ArvindKejriwal	<a href="http://twitter.coi.Ajay649Ajay		22	NA	NA						
6		6 RT @ZeeNews: Sedition row fallout?#JNU sees	<a href="http://twitter.coi.balajiraj		45	NA	NA						
7		7 RT @bholeswar7799: @ArvindKejriwal	<a href="http://twitter.coi.VINEET4BJP		22	NA	NA						
8		8 RT @AskRaushan: Meanwhile @BDUTT	<a href="http://twitter.coi.NeelamZindgi		56	NA	NA						
9		9 RT @bholeswar7799: @ArvindKejriwal	<a href="http://twitter.coi.barotchetan99		22	NA	NA						
10		10 RT @bholeswar7799: @ArvindKejriwal	<a href="http://twitter.coi.NeelamZindgi		22	NA	NA						
11		11 RT @sudhirchaudhary: When #JNU admissions b	<a href="http://twitter.coi.VIBHORJ72521372		1283	NA	NA						
12		12 RT @SirJadeja: Meanwhile In #Pak ,#BAN	<a href="http://twitter.coi.PraveenaKetees		241	NA	NA						
13		13 RT @sudhirchaudhary: When #JNU admissions b	<a href="http://twitter.coi.007Ravi		1283	NA	NA						
14		14 #indVsWI #wi	<a href="http://twitter.coi.ChaitanyaAdvait		0	NA	NA						
15		15 RT @AskRaushan: Meanwhile @BDUTT	<a href="http://twitter.coi.singh4bahrat		56	NA	NA						
16		16 RT @bholeswar7799: @ArvindKejriwal	<a href="http://twitter.coi.rahulmishrakm		22	NA	NA						
17		17 RT @SirJadeja: Meanwhile In #Pak ,#BAN	<a href="http://twitter.coi.Salman1o1		241	NA	NA						
18		18 RT @Mamtatiripathi03: <U+0935><U+093E><U+09	<a href="http://twitter.coi.mukeshscsjsnu		2	NA	NA						
19		19 RT @AskRaushan: Meanwhile @BDUTT	<a href="http://twitter.coi.royprateek2010		56	NA	NA						
20		20 RT @sudhirchaudhary: When #JNU admissions b	<a href="http://twitter.coi.harshadspatel9		1283	NA	NA						
21		21 RT @bholeswar7799: @ArvindKejriwal	<a href="http://twitter.coi.YeslamSaffron		22	NA	NA						
22		22 RT @sudhirchaudhary: When #JNU admissions b	<a href="http://twitter.coi.sanjaydas11		1283	NA	NA						

after writing the data into csv file now we extract the text written by the

user by the below code

```
library('plyr')
```

```
tweets.text=lapply(tweet,function(t)t$getText())
```

```
< tweet.text=lapply(tweet,function(t)t$getText())
```

cleaning the text

hello.text<- iconv(hello.text, 'UTF-8', 'ASCII') this is used for removing the emoji
else tolower function cannot be used

```
.  
> tweet.text=lapply(tweet,function(t)t$getText())  
> hello.text<- iconv(tweet.text, 'UTF-8', 'ASCII')  
> asciemojii<- iconv(tweet.text, 'UTF-8', 'ASCII')  
>  
>  
>  
>
```

gsub() function replaces all matches of a string, if the parameter is a string vector, returns a string vector of the same length and with the same attributes (after possible coercion to character). Elements of string vectors which are not substituted will be returned unchanged (including any declared encoding).

```
gsub(pattern, replacement, x, ignore.case = FALSE, perl = FALSE,  
fixed = FALSE, useBytes = FALSE)
```

- pattern: string to be matched
- replacement: string for replacement
- x: string or string vector
- ignore.case: if TRUE, ignore case

...

```
> x <- "R Tutorial"
```

```
> gsub("ut","ot",x)
```

```
[1] "R Totorial"
```

Case insensitive replace:

```
> gsub("tut","ot",x,ignore.case=T))
```

```
[1] "R otorial"
```

#convert all text to lower case

```
tweets.text <- tolower(tweets.text)
```

```
>
>
>
> tweet.text=lapply(tweet,function(t)t$getText())
> hello.text<- iconv(tweet.text, 'UTF-8', 'ASCII')
> asciemojii<- iconv(tweet.text, 'UTF-8', 'ASCII')
>
>
>
>
>
> tweettolower <- tolower(asciemojii)
> |
```

Replace blank space ("rt")

```
tweets.text <- gsub("rt", "", tweets.text)
```

Replace @UserName

```
tweets.text <- gsub("@\\w+", "", tweets.text)
```

Remove punctuation

```
tweets.text <- gsub("[[:punct:]]", "", tweets.text)
```

Remove links

```
tweets.text <- gsub("http\\w+", "", tweets.text)
```

Remove tabs

```
tweets.text <- gsub("[ \\t]{2,}", "", tweets.text)
```

Remove blank spaces at the beginning

```
tweets.text <- gsub("^ ", "", tweets.text)
```

Remove blank spaces at the end

```
tweets.text <- gsub(" $", "", tweets.text)
```

COUNTING THE NUMBER OF FOLLOWERS FOR THE USERS

for (i in 500)

{

g=read.csv("names.csv",sep="")

supply(g, function(x)

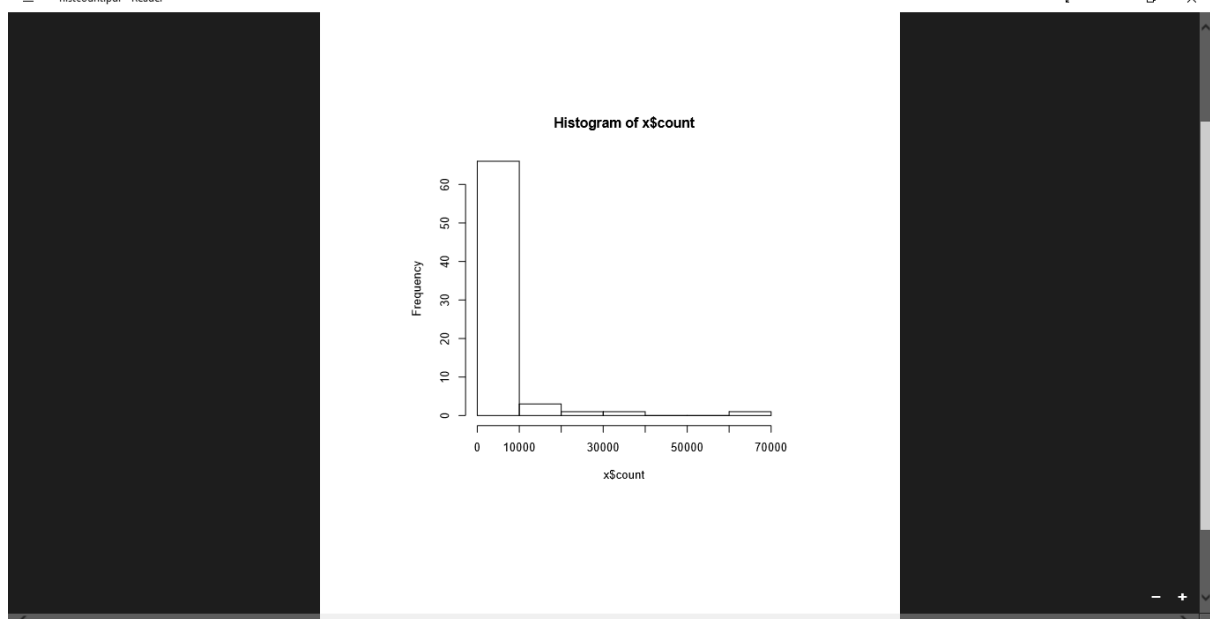
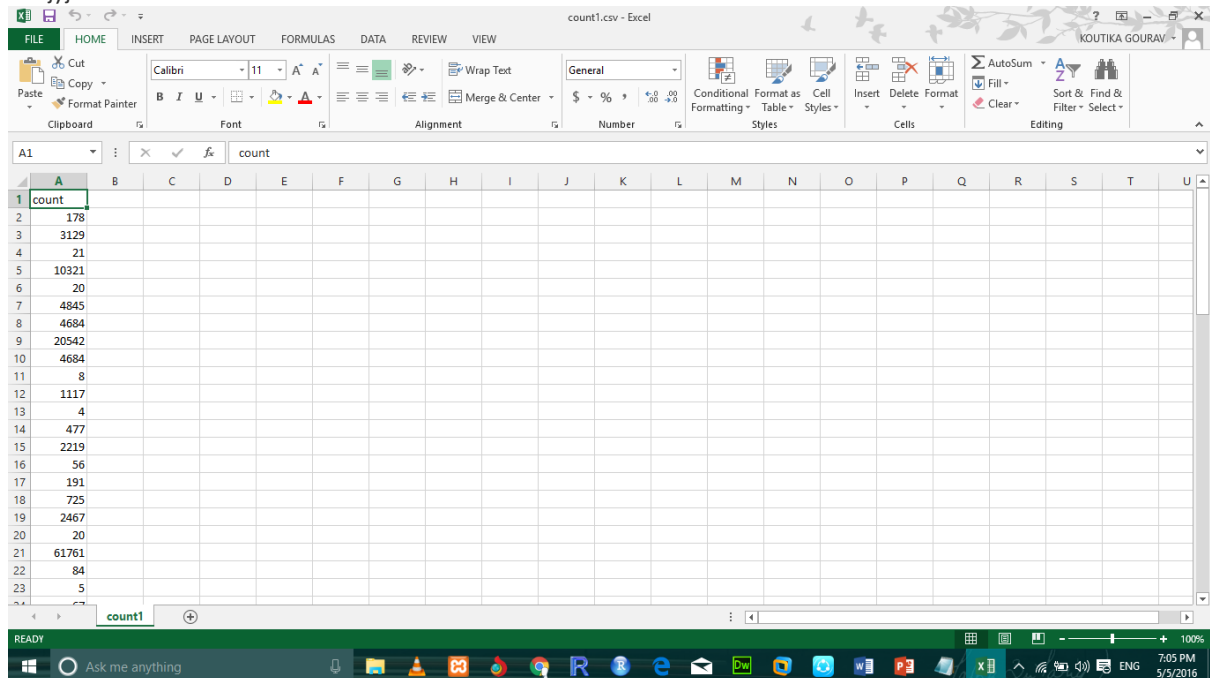
{

user <- getUser(x)

followers<- user\$followersCount

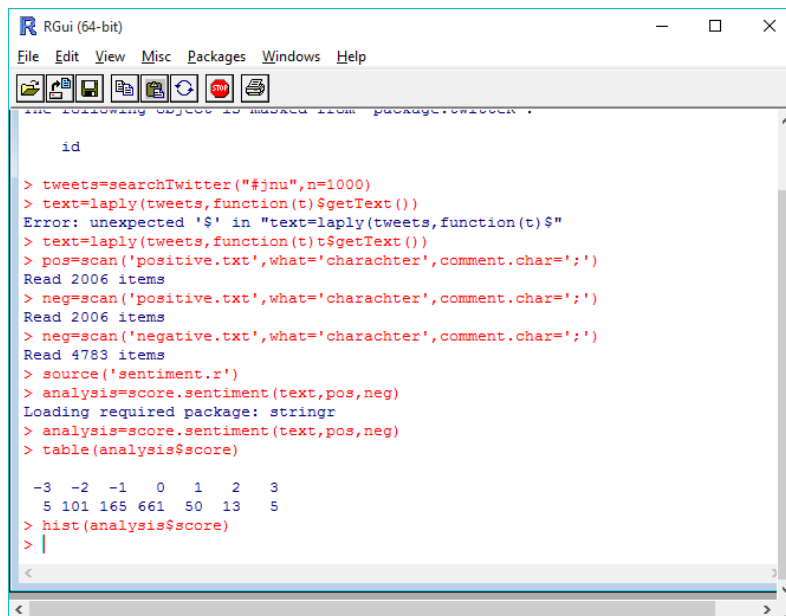
cat(followers,file="count.csv",sep="/n",append=TRUE)

}}}



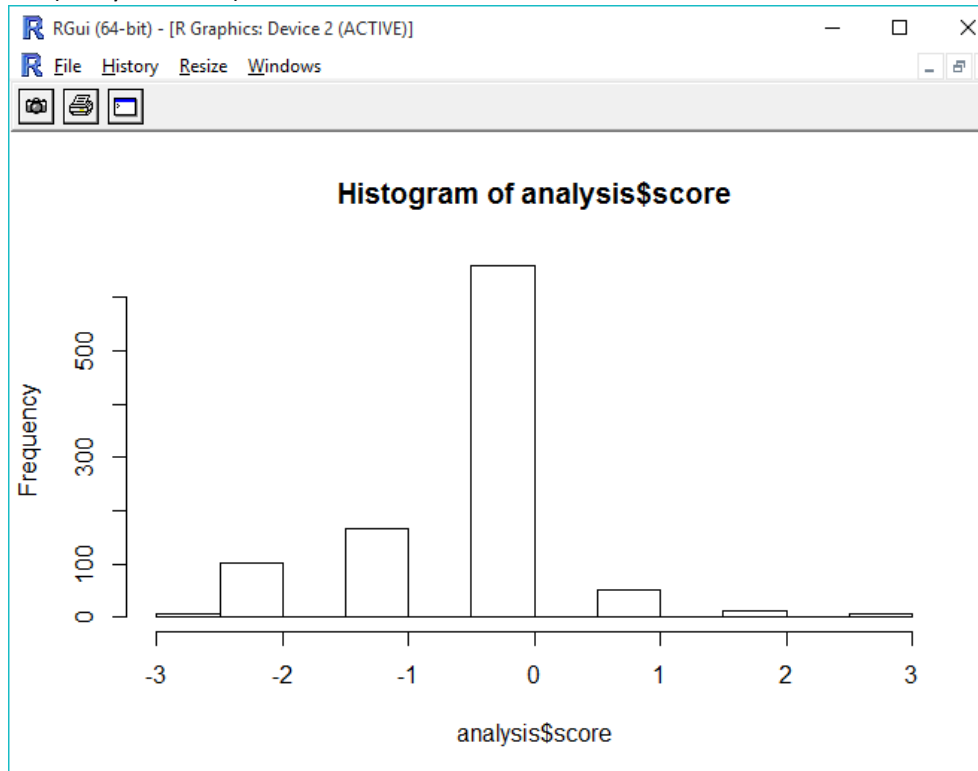
Calculating score


```
text=lapply(tweets,function(t)t$getText())  
pos=scan('positive.txt',what='character',comment.char=';')  
neg=scan('negative.txt',what='character',comment.char=';')  
source('sentiment.r')  
analysis=score.sentiment(text,pos,neg)  
table(analysis$score)
```



```
RGui (64-bit)  
File Edit View Misc Packages Windows Help  
The following object is masked from 'package:twitter':  
id  
  
> tweets=searchTwitter("#jnu",n=1000)  
> text=lapply(tweets,function(t)$getText())  
Error: unexpected '$' in "text=lapply(tweets,function(t)$"  
> text=lapply(tweets,function(t)t$getText())  
> pos=scan('positive.txt',what='character',comment.char=';')  
Read 2006 items  
> neg=scan('positive.txt',what='character',comment.char=';')  
Read 2006 items  
> neg=scan('negative.txt',what='character',comment.char=';')  
Read 4783 items  
> source('sentiment.r')  
> analysis=score.sentiment(text,pos,neg)  
Loading required package: stringr  
> analysis=score.sentiment(text,pos,neg)  
> table(analysis$score)  
  
-3 -2 -1 0 1 2 3  
5 101 165 661 50 13 5  
> hist(analysis$score)  
> |
```

```
hist(analysis$score)
```



EORD CLOUD IN R

```
corpus=Corpus(VectorSource(text))
```

```
inspect(corpus[1])// to look the tweet
```

```
corpus=tm_map(corpus,removePunctuation)
```

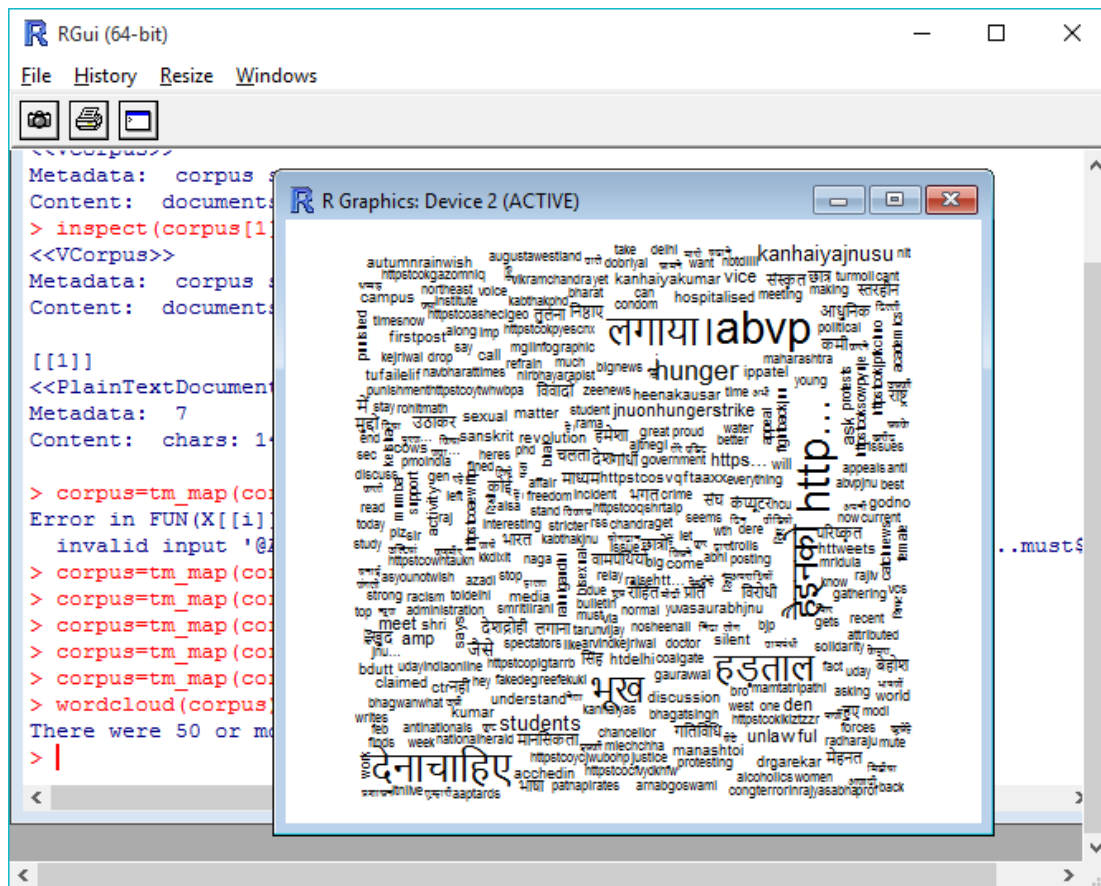
```
corpus=tm_map(corpus,content_transformer(tolower))
```

```
corpus=tm_map(corpus,removeWords,stopwords("english"))
```

```
corpus=tm_map(corpus,removeNumbers)
```

```
corpus=tm_map(corpus,stripWhitespace)
```

```
wodcloud(corpus)
```



wordcloud(corpus, random.order=F, max.words=50, scale=c(3,0.5), colors=rainbow(50))

