api-twitter

Yarden Toren& Oran Shemesh

March 30, 2016

For this assignment we use Twitter api for getting live last 500 tweets from twitter that regards to Israel. The data is about twitter tweets, where each tweet composed from created_time, tweet_text, Author_name, tweet_location, etc.

installtions and libraries declaration:

```
#install.packages(c("devtools", "rjson", "bit64", "httr"))
#install_github("twitteR", username="geoffjentry")
#install.packages("wavelets")
#install.packages("fGarch")
#install.packages('tm')
#install("maps")
#install.packages("car", dependencies=TRUE)

library(devtools)
library(twitteR)
library(wavelets)
library(fGarch)
library(tm)
library(igraph)
library(RColorBrewer)
library(httr)
library(wordcloud)
library(maps)
library(mapproj)
require(ggplot2)
require(maps)
require(twitteR)
```

create connection with twitter-api:

```
api_key <- "dn5M2jEg1xAJTMFZR7PDjAG8X"
api_secret <- "ODBygdzQG0Ykk63phxvTKtSCoBt7lnpOETByyPZrXJxYpxA28i"
access_token <- "421905010-OMpoQR2JEwtPrR7kSHcqQOD4aHn9RrVsa3HCKMIo"
access_token_secret <- "e4zYMV44EoKRNcPUUqdoYF5WKT52fKL7crDKKRWBSNWyQ"

setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)
```

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

we desided to explore the behaver on twitter tweets relation to the word “israel”.

```
#data_mining_tweets <- searchTwitter('israel', lang='en', n=5000, #geocode='32.0508901,34.8444391,99999
data_mining_tweets <- searchTwitter('israel', lang='en', n=5000)
```

Data example:.

```
head(data_mining_tweets)
```

```
## [[1]]
## [1] "RNN_Bible: 24:1 And there went up the way to Bethany, that he should be given her of the childr
##
## [[2]]
## [1] "1amyisraelchai: RT @TMIJOS: Ugandan Jews get official recognition from The Jewish Agency https:/
##
## [[3]]
## [1] "AmericaDied: @davidjones720 Good thing Kerry got Iran to agree to never attack Israel !"
##
## [[4]]
## [1] "EgyptBulletinEn: Israel gives blessing to Egypt&amp;amp;#039;s return of Red Sea islands to Saudi A
##
## [[5]]
## [1] "StandWithUsUK: We welcomed Eurovision contestants to Israel this week for Israel Calling! They a
##
## [[6]]
## [1] "Ninaberry: Addressed a package of books won in a giveaway to Israel. Wish I could follow in my l
```

first, we explore the context on the text:

```
#wordCount plot
#list of words in text
LT <- list()
for(i in 1:length(data_mining_tweets)){
  LN <- list()
  LN <- strsplit(data_mining_tweets[[i]]$text," ", fixed=TRUE)
  for(j in 1:length(LN[[1]])){
    LT[[length(LT)+1]] <- LN[[1]][j]
  }
}

print(paste("there was found:",length(LT),"words in the 500 tweets"))
```

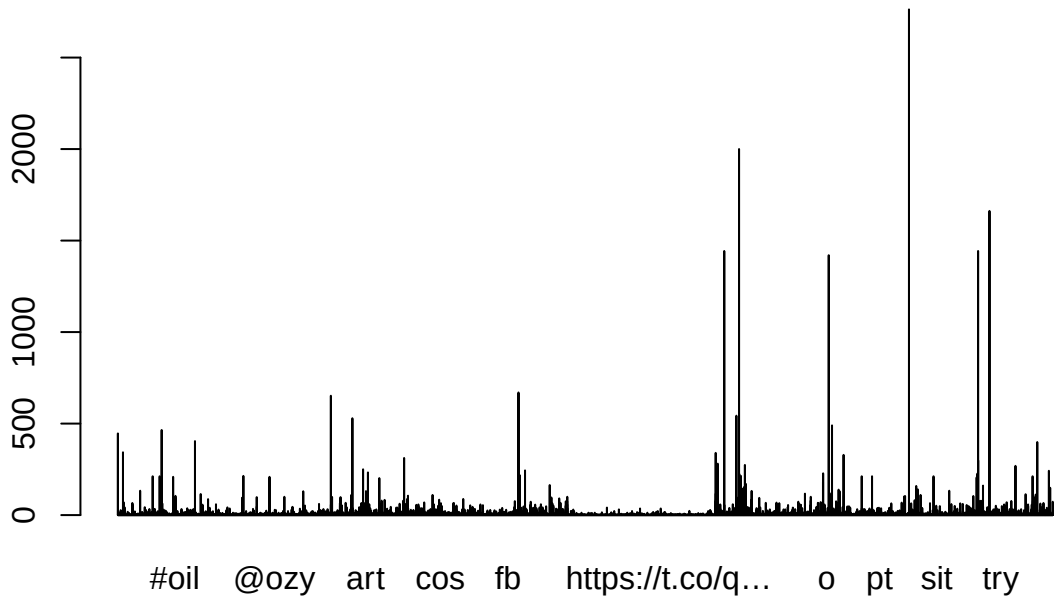
```
## [1] "there was found: 80696 words in the 500 tweets"
```

```
d<-as.data.frame(LT)
t<-as.vector(t(d))
print(paste("with:",length(table(t)),"distributed words"))
```

```
## [1] "with: 14857 distributed words"
```

Distributed as follows:

```
barplot(table(t),main="Distributed words in the 500 tweets")
```

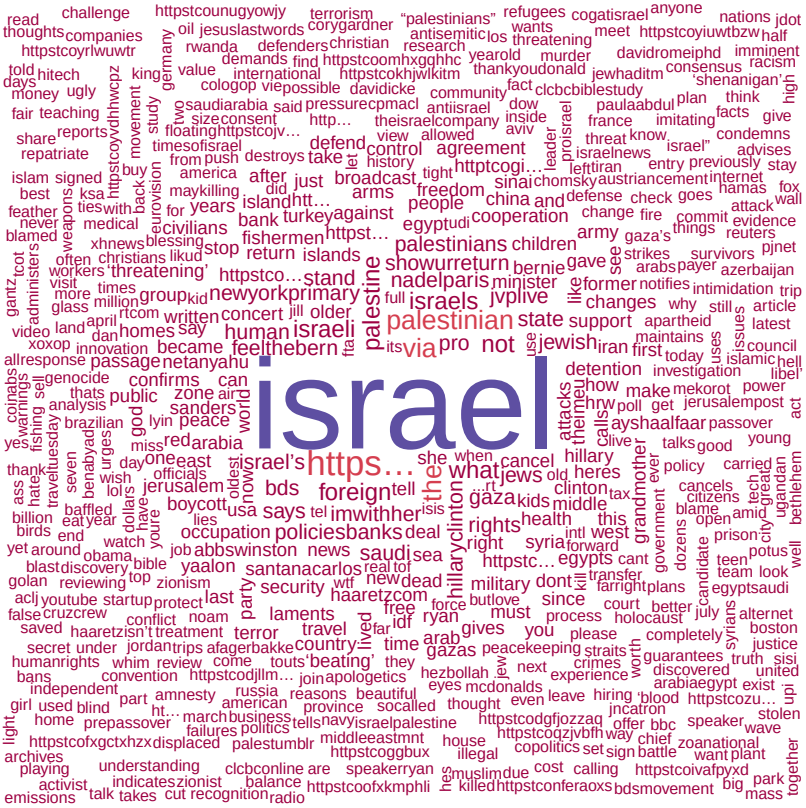


With such a large number of distributed words, we expected to have a lot of issues that talk about Israel (what actually didn't happened). Further, interest us to know what the main issues that are assigned to Israel in the social network.

wordcloud image with minimum 10 frequency:

```
#get list of all text
W <- lapply( data_mining_tweets , function(x) `$`( x , text)[[1]] )

#word count
#https://deltadna.com/blog/text-mining-in-r-for-term-frequency/
review_text <- paste(W, collapse=" ")
review_source <- VectorSource(review_text)
corpus <- Corpus(review_source)

# cleaning the text
corpus <- tm_map(corpus, removePunctuation)
corpus <- tm_map(corpus, stripWhitespace)
corpus <- tm_map(corpus, removeWords, stopwords("english"))

tdm =TermDocumentMatrix(
  corpus,
  control = list(
    removePunctuation = TRUE,
    stopwords = c("amp","http", stopwords("german")),
    removeNumbers = TRUE,
    tolower = TRUE)
```

```
)


m = as.matrix(tdm)
# get word counts in decreasing order
word_freqs = sort(rowSums(m), decreasing = TRUE)
# create a data frame with words and their frequencies
dm = data.frame(word = names(word_freqs), freq = word_freqs)
# word cloud
plotfile1 <- "krankheit_wordcloud.png" # optional
wordcloud(dm$word, dm$freq, min.freq=10, random.order = FALSE,
          colors = brewer.pal(11, "Spectral"))
```

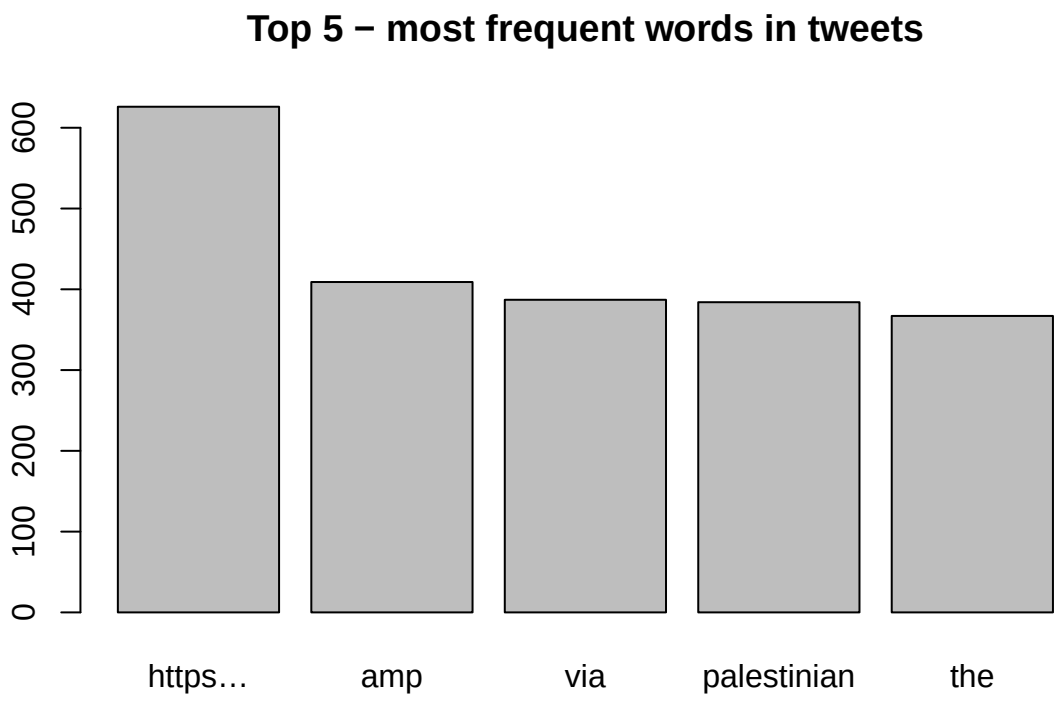



```
png(filename=plotfile1, width=1740, height=1740, units="px") # optional
dev.off() # optional
```

```
## pdf
##   2
```

lets make zoom in, and watch the top 5 most frequent words in tweets, after clean the data, and remove stopwords.

```r
#create the document-term matrix.
dtm <- DocumentTermMatrix(corpus)
dtm2 <- as.matrix(dtm)
frequency <- colSums(dtm2)
frequency <- sort(frequency, decreasing=TRUE)
barplot(frequency[2:6],main="Top 5 - most frequent words in tweets")
```



By analysis those graphs, we can find high correlation between israel-palastin, and israel-bds. It's something we expected, but we thought it would be not so exclusive.

Than, we wanted to know if there are abnormall active users, so we explored the distribution of users who wrote on Israel.

wordcloud image with minimum 3 frequency:

```r
#get list of all text
W <- lapply( data_mining_tweets , function(x) `$`( x , screenName)[[1]] )

#word count
#https://deltadna.com/blog/text-mining-in-r-for-term-frequency/
review_text <- paste(W, collapse=" ")
review_source <- VectorSource(review_text)
corpus <- Corpus(review_source)

tdm =TermDocumentMatrix(corpus, control = list(removePunctuation = TRUE,tolower = TRUE))

m = as.matrix(tdm)
```

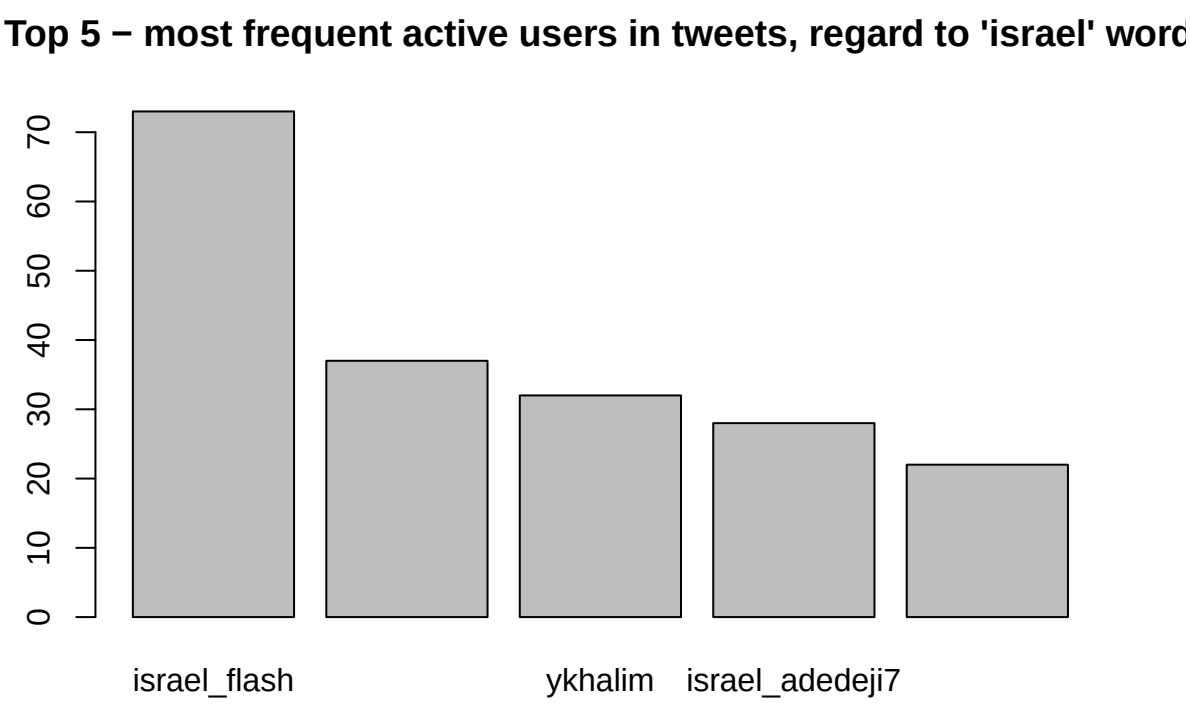
```
# get word counts in decreasing order
word_freqs = sort(rowSums(m), decreasing = TRUE)
# create a data frame with words and their frequencies
dm = data.frame(word = names(word_freqs), freq = word_freqs)
# word cloud
plotfile1 <- "krankheit_wordcloud.png" # optional
wordcloud(dm$word, dm$freq, min.freq=3, random.order = FALSE,
          colors = brewer.pal(11, "Spectral"))
```


```
png(filename=plotfile1, width=1740, height=1740, units="px") # optional
dev.off() # optional
```

```
## pdf
##   2
```

lets make zoom in, and watch the top 5 most frequent active users in tweets.

```
#create the document-term matrix.
dtm <- DocumentTermMatrix(corpus)
dtm2 <- as.matrix(dtm)
frequency <- colSums(dtm2)
frequency <- sort(frequency, decreasing=TRUE)
barplot(frequency[1:5],main="Top 5 - most frequent active users in tweets, regard to 'israel' word")
```

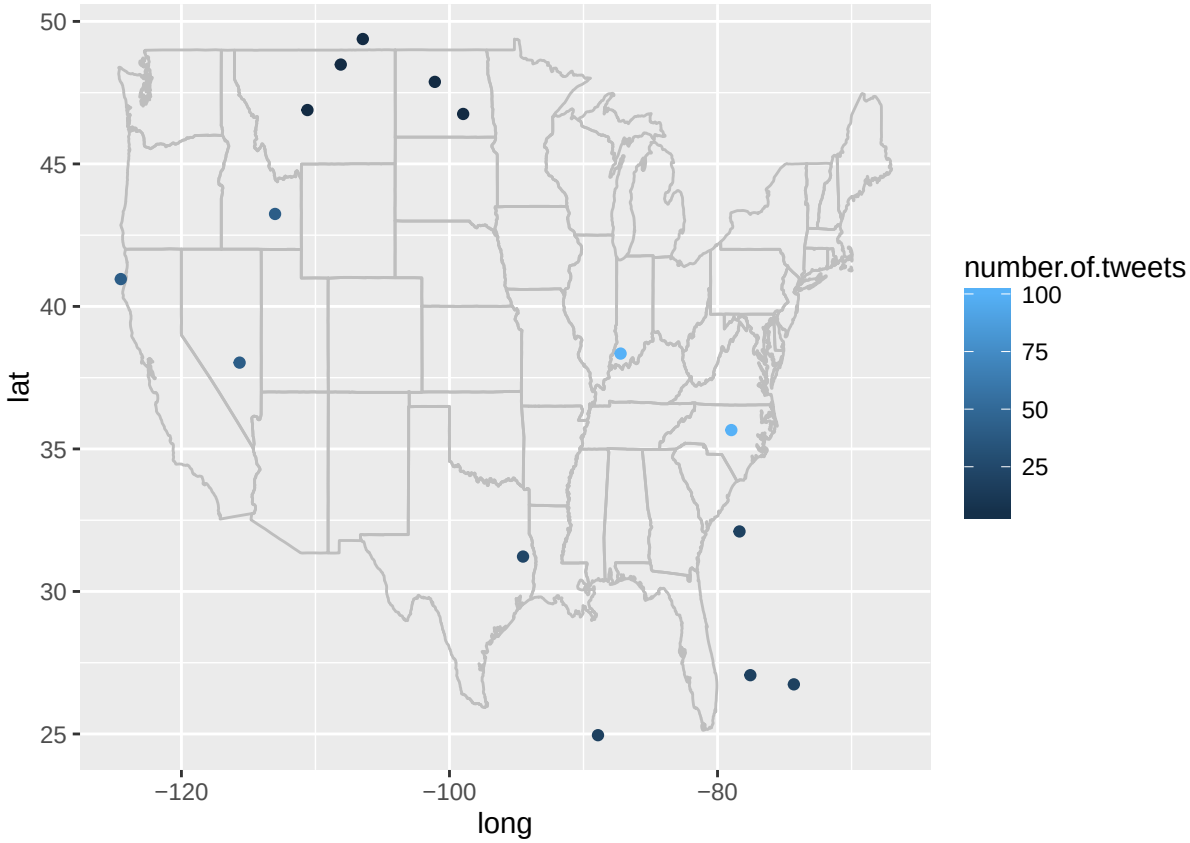


By analysis those graphs, we can find that there are 4 main active user that writes on Israel. Thats small number of decision leaders, which writing ferment tweets related Israel-Palestinians or Israel-Bds. This outrageous, because Israel has a lot to offer, and because of those decision leaders, Twitter full of messages related to topics less posetive on Israel.

Another analysis that interested us was "Where in the United States are the most talking about Israel".

```
locations=15
radius="1000000mi"
#radius from randomly chosen location
radius=radius
lat<-runif(n=locations,min=24.446667, max=49.384472)
long<-runif(n=locations,min=-124.733056, max=-66.949778)
#generate data fram with random longitude, latitude and chosen radius
coordinates<-as.data.frame(cbind(lat,long,radius))
coordinates$lat<-lat
coordinates$long<-long
#create a string of the lat, long, and radius for entry into searchTwitter()
for(i in 1:length(coordinates$lat)){
  coordinates$search.twitter.entry[i]<-toString(c(coordinates$lat[i],coordinates$long[i],radius))
}
# take out spaces in the string
coordinates$search.twitter.entry<-gsub(" ","", coordinates$search.twitter.entry,fixed=TRUE)

#Search twitter at each location, check how many tweets and put into dataframe
for(i in 1:length(coordinates$lat)){
  coordinates$number.of.tweets[i]<-
```

```
    length(searchTwitter(searchString="israel",n=100,geocode=coordinates$search.twitter.entry[i]))
}
#making the US map
all_states <- map_data("state")
#plot all points on the map
p <- ggplot()
p <- p + geom_polygon( data=all_states, aes(x=long, y=lat, group = group),colour="grey",fill=NA )

p<-p + geom_point( data=coordinates, aes(x=long, y=lat,color=number.of.tweets
)) + scale_size(name="# of tweets")
p
```



By analysis the map, we can find that the east side of United State is more envolve in Israel tweets. note: We didn't check the relative general activities of these regions.

Summary, Conclusions:

In this research we saw that the Israeli-Palestinian conflict plays a major role in the scene of Twitter. Also, we noticed that the Eastern United States more actively from the Western side related to Israel (We didn't check the relative general activities of these regions). As citizens of Israel, we didn't expect to get such results. Israel has many things to offer, such as tourism, innovation, high-tech, sports and more and I would expected that they would be reflected also in active environment like Twitter.

Recommendations:

Our Recommendations are that the Ministry of Foreign Affairs of the state Israel must come to the scene of the tweeter, to encourage people to write positive things about Israel and/or divert the public distracted to themes that presenters Israel in a different aspect.