

pd1

Robert Trąbczyński

Sunday, March 06, 2016

Poniższy raport zawiera informacje o postach umieszczonych na serwisie *Twitter.com* związanych z hasłem “oskary2016”. Użyte biblioteki:

```
library(twitterR)
library(streamR)
library(ROAuth)
library(streamR)
library(maps)
library(dplyr)
library(stringi)
```

Wybrane Tweety zapisuję do pliku *oskary2.json*, który załączam w mailu.

```
requestURL <- "https://api.twitter.com/oauth/request_token"
accessURL <- "https://api.twitter.com/oauth/access_token"
authURL <- "https://api.twitter.com/oauth/authorize"
consumerKey <- "jIbHMPbOQUdHyv0io71PLxvRI"
consumerSecret <- "UdfwBZZTAUWwAr3iohNo8DESdwF6w7oq98ZqIFdeUX74YQQA2"
oauthKey <- "RTrabczynski"
oauthSecret <- "2941401575"

# proces autoryzacji jest trzykrokowy
paczka <- OAuthFactory$new(consumerKey = consumerKey, consumerSecret = consumerSecret,
                           oauthKey = oauthKey, oauthSecret = oauthSecret,
                           requestURL = requestURL, accessURL = accessURL, authURL = authURL)

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

setwd("d:/Desktop/pw/Ribd")

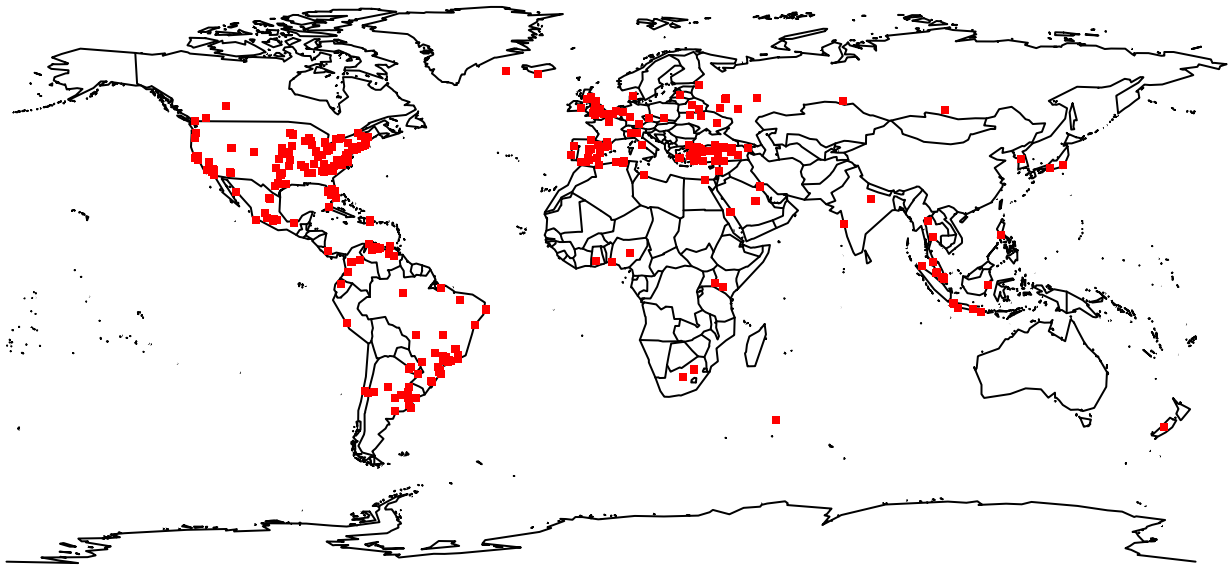
filterStream( file="oskary2.json",
              track=c("oskary2016"),
              timeout=1*60, oauth=paczka,
              locations=c(-180,-90,180,90))

setwd("d:/Desktop/pw/Ribd")
parsedTweets <- parseTweets("oskary2.json", simplify = FALSE, verbose = TRUE)
```

2851 tweets have been parsed.

W ciągu minuty wyszukano 2851 postów. Poniżej znajduje się mapa obrazująca rozmieszczenie lokalizacji wysyłania Tweetów:

```
map(mar=c(0,0,0,0))
points(parsedTwees$lon, parsedTwees$lat, col="red", pch=".", cex=4)
```



Za pomocą zapytań *SQL* wyszukałem 10 krajów, z których wysyłanych było najwięcej Tweetów:

```
najczestsze<-count(parsedTwees,country_code )
arrange(top_n(najczestsze, 10,n ),desc(n))
```

```
## Source: local data frame [10 x 2]
##
##   country_code      n
##   (chr) (int)
## 1          US    839
## 2          BR    359
## 3          AR    210
## 4          GB    160
## 5          TR    145
## 6          ES    102
## 7          FR     83
## 8          SA     77
## 9          MX     61
## 10         JP     50
```

Sprawdźmy również, jaki procent Amerykanów wysłało swojego Tweeta z iPhone'a:

```
a<-stri_count_regex(filter(parsedTwees, country_code=="US")$source,"iphone")
b<-length(a)
paste(round(sum(a)/b , 2)*100, "%")
```

```
## [1] "62 %"
```

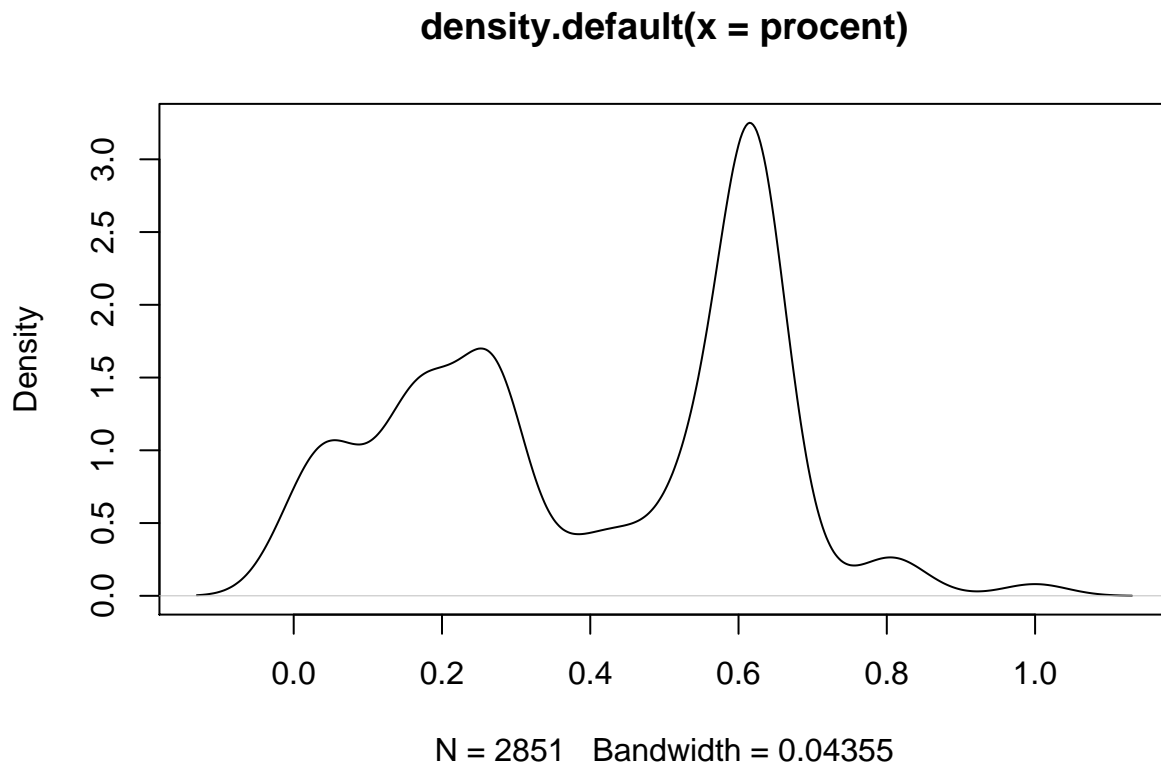
Policzę procent użytkowników Iphone'a dla wszystkich krajów:

```
kraje<-parsedTwees$country_code
ile<- length(kraje)
procent<-rep(0,ile)

for(i in 1:ile){
a<-stri_count_regex(filter(parsedTwees, country_code==kraje[i])$source,"iphone")
b<-length(a)
if (b!=0){procent[i]<-sum(a)/b }
}
```

Sprawdźmy, czy rozkład użytkowników Iphone'a jest zbliżony do normalnego:

```
plot(density(procent))
```



```
shapiro.test(procent)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  procent  
## W = 0.90058, p-value < 2.2e-16
```

P-value jest bardzo małe, więc nie mamy podstaw sądzić, że jest on normalny.