Individual-Project-#2

Angel Janica Marie De Jesus

2022-12-12

#De Jesus #Individual-Project-2

Instructions:

#Individual Project #2 #Extract 10000 tweets from Twitter using twitteR package including retweets. #Subset the retweets and the original tweets into a separate file #Plot the retweets and the original tweets using bar graph in vertical manner. #Include legends

#install and load packages install.packages('twitteR') install.packages("RCurl") install.packages("rtweet") install.packages("tinytex") install.packages("plotly") install.packages("RColorBrewer") install.packages("stringr") install.packages("magrittr") install.packages("tm") install.packages("wordcloud") install.packages("wordcloud2")

library("twitteR") library("RCurl") library(tinytex) library(rtweet) library(ggplot2) library(RColorBrewer) library(tm) library(dplyr) library(wordcloud) library(wordcloud2) library(stringr)

#install.packages("syuzhet") #for sentiment analysis install.packages("syuzhet") library(syuzhet)

install.packages("rdfp") library("rdfp") library(magrittr)

#RESTART R session! install.packages(c("devtools", "rjson", "bit64", "httr"))

#declare tokens and keys

```
CONSUMER_KEY <- "rNOC9AXlOmMwltDU6fxZGJvoc"
CONSUMER_SECRET <- "LMC48sUCCPib785PAYb1pzwjDiWueoJZthbpdpxiYUh2gxa4on"
ACCESS_TOKEN <- "1596020398365151234-Ttupby9xwEH2IPfF2aIs1SdiXHtPgo"
ACCESS_SECRET <- "yBeaMXTih50GXUOhiZBF2o2xHNuqUA9w3cUCFS0Oqt0U6"
```

#Connect to twitter app

```
library("twitteR")
```

```
## Warning: package 'twitteR' was built under R version 4.2.2
```

```
setup_twitter_oauth(consumer_key = CONSUMER_KEY,
                    consumer_secret = CONSUMER_SECRET,
                    access_token = ACCESS_TOKEN,
                    access_secret = ACCESS_SECRET)
```

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

#extract 10k tweets including retweets #Netflix

```
  searchTwt <- searchTwitter("#Wednesday",
                             n = 10000,
                             lang = "en",
                             since = "2022-12-03",
                             until = "2022-12-10",
                             retryOnRateLimit = 120)
```

#converet to dataframe

```
  library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 4.2.2
```

```
##
## 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
```

```
  TwtDF <- twListToDF(searchTwt)
  class(TwtDF)
```

```
## [1] "data.frame"
```

```
  names(TwtDF)
```

```
##  [1] "text"          "favorited"     "favoriteCount" "replyToSN"
##  [5] "created"       "truncated"     "replyToSID"    "id"
##  [9] "replyToUID"    "statusSource"  "screenName"    "retweetCount"
## [13] "isRetweet"     "retweeted"     "longitude"     "latitude"
```

```
  View(TwtDF)[1:5]
```

```
## NULL
```

```
  head(TwtDF$text)[1:5]
```

```
## [1] "RT @maik_check: gomez addams #wednesday https://t.co/TD2lkuyLyF"
## [2] "RT @maik_check: gomez addams #wednesday https://t.co/TD2lkuyLyF"
## [3] "RT @sleepyash013: They're my comfort ship now\n#wenclair #wednesday #wednesdayaddams #enidsincla
## [4] "RT @maik_check: gomez addams #wednesday https://t.co/TD2lkuyLyF"
## [5] "RT @maik_check: gomez addams #wednesday https://t.co/TD2lkuyLyF"
```

```
  save(TwtDF,file = "WedTwtDF.Rdata")
```

#load dataset

```
  load(file = "WedTwtDF.Rdata")
```

#Subset the retweets and the original tweets into a separate file

```
  library(dplyr)
  trendSubset <- TwtDF %>% select(text, screenName, created, isRetweet) %>%
    filter(isRetweet == FALSE)

  trendSubset1 <- TwtDF %>% select(text, screenName, created, isRetweet) %>%
    filter(isRetweet == TRUE)
```

#Plot the retweets and the original tweets using bar graph in vertical manner.

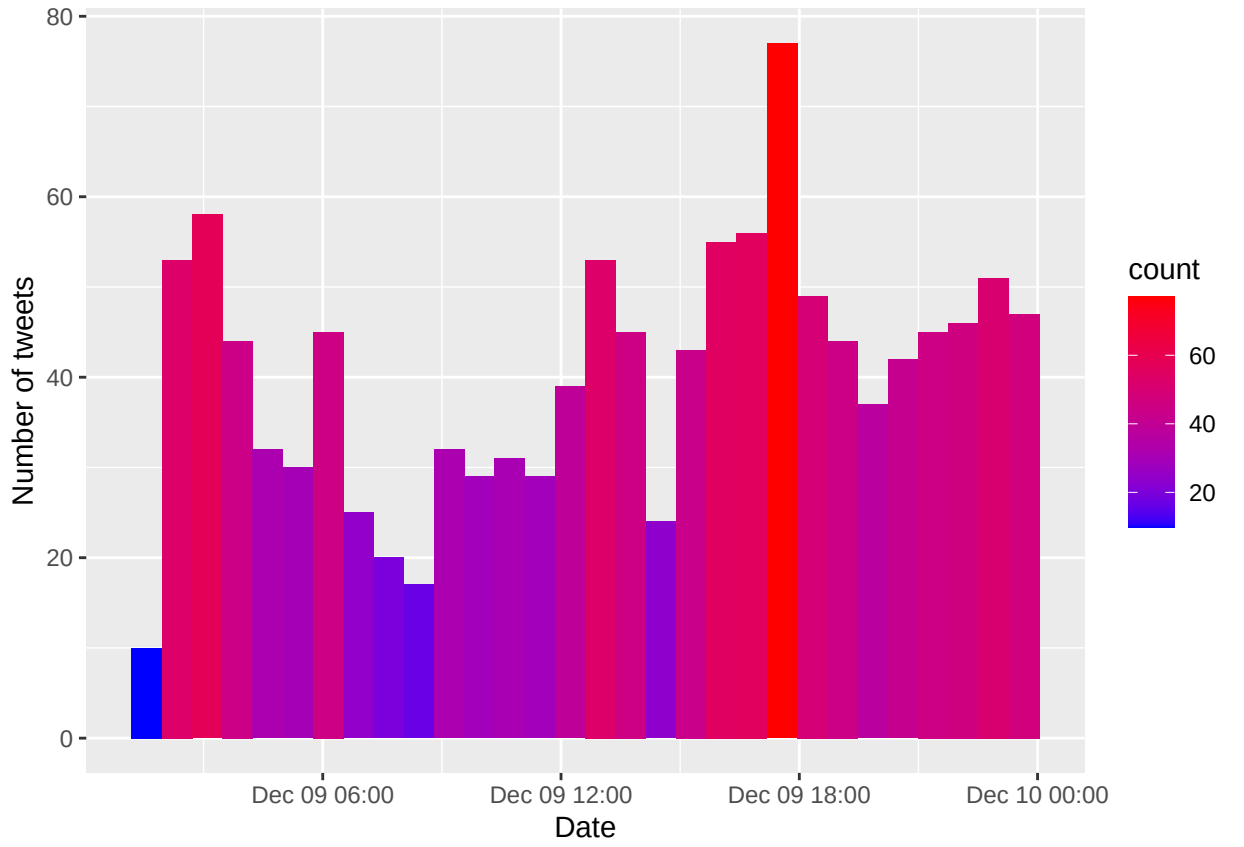
```
  library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
  ggplot(data = trendSubset, aes(x = created)) +
    geom_histogram(aes(fill = ..count..)) +
    xlab("Date") + ylab("Number of tweets") +
    scale_fill_gradient(low = "blue", high = "red")
```

```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(count)` instead.

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

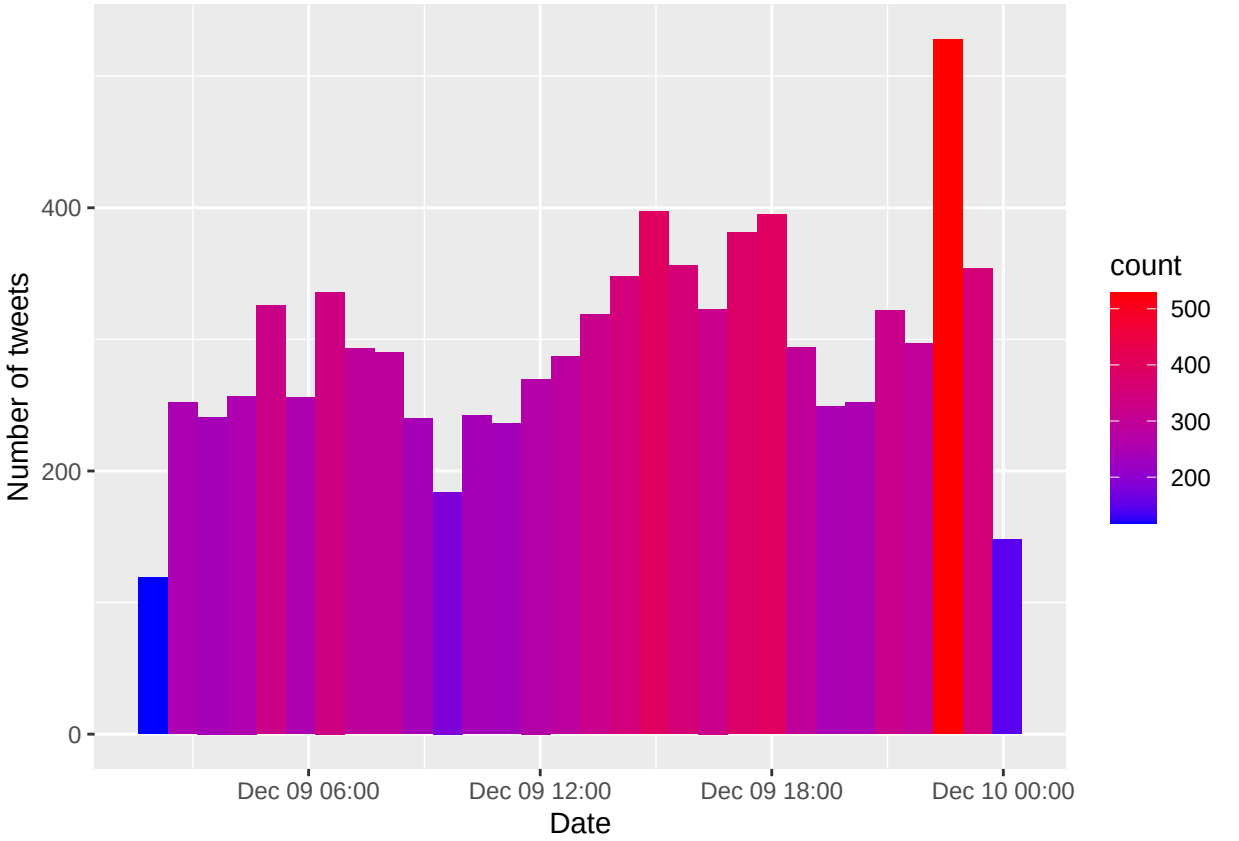



```
  theme(legend.position = "topleft")
```

```
## List of 1
##  $ legend.position: chr "topleft"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

```
  library(ggplot2)
  ggplot(data = trendSubset1, aes(x = created)) +
    geom_histogram(aes(fill = ..count..)) +
    xlab("Date") + ylab("Number of tweets") +
    scale_fill_gradient(low = "blue", high = "red")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
  theme(legend.position = "topleft")
```

```
## List of 1
##  $ legend.position: chr "topleft"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```