Describing data set Sept 2021

Joost Bloos

21/11/2021

```
#install.packages("ISwR")
#install.packages("dplyr") # includes ggplot
#install.packages("ggplot")
#install.packages("ggplot2")
#install.packages("twitteR")
#install.packages("tidyr")
#install.packages("tidyverse")
#install.packages("ggmap")
#install.packages("sf")
#install.packages("mapview")
#install.packages("maps")
#install.packages("magrittr")
#install.packages("rgeos")
#install.packages("revgeo")
#install.packages("NLP")

#install.packages(c("cowplot", "googleway", "ggplot2", "ggplot", "ggrepel", "ggspatial", "libwgeom", "s


library(dplyr)
```

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

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

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

```
library(tm)
```

```
## Loading required package: NLP
```

```
library(ISwR)
library(twitteR)
```

```
##
## Attaching package: 'twitteR'
```

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

```
library(tidyr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x ggplot2::annotate() masks NLP::annotate()
## x dplyr::filter()     masks stats::filter()
## x twitteR::id()       masks dplyr::id()
## x dplyr::lag()        masks stats::lag()
## x twitteR::location() masks dplyr::location()
```

```
library(ggmap)
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(mapview)
library("rnaturalearth")
library("rnaturalearthdata")
library(devtools)
```

```
## Loading required package: usethis
```

```
library(devtools)
install_github('mhudecheck/revgeo')
```

```
## Skipping install of 'revgeo' from a github remote, the SHA1 (5a17dcbf) has not changed since last ins
##   Use `force = TRUE` to force installation
```

```
library(revgeo)
```

```
#getwd()

#setwd("C:/Ryerson University - Capstone project/Module 2/EIEEE - Large dataset/Combined")
```

```r
data1 <- read.csv("corona_tweets_544 Sept 2021", sep=",", stringsAsFactors = F, na.strings = c("","NA"))
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
## embedded nul(s) found in input
```

```r
#brief description of original hydrated data set May 2020:
str(data1) #shows number of observation out of 35 variables
```

```
## 'data.frame':    1456316 obs. of  35 variables:
##  $ coordinates              : chr  NA NA NA NA ...
##  $ created_at               : chr  "Mon Sep 13 04:25:41 +0000 2021" "Mon Sep 13 04:25:45 +0000 2021
##  $ hashtags                 : chr  "COVID19" NA NA NA ...
##  $ media                    : chr  NA NA NA NA ...
##  $ urls                     : chr  NA NA "https://www.tga.gov.au/media-release/new-restrictions-pres
##  $ favorite_count           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ id                       : num  1.44e+18 1.44e+18 1.44e+18 1.44e+18 1.44e+18 ...
##  $ in_reply_to_screen_name  : chr  NA "TizzyEnt" "jimiuorio" NA ...
##  $ in_reply_to_status_id    : num  NA 1.44e+18 1.44e+18 NA NA ...
##  $ in_reply_to_user_id      : num  NA 27933405 60622883 NA NA ...
##  $ lang                     : chr  "en" "en" "en" "en" ...
##  $ place                    : chr  NA NA NA NA ...
##  $ possibly_sensitive       : chr  NA NA "false" NA ...
##  $ quote_id                 : num  NA NA NA 1.44e+18 NA ...
##  $ retweet_count            : int  21 0 0 7 41 66 212 817 1040 39 ...
##  $ retweet_id               : num  1.44e+18 NA NA 1.44e+18 1.44e+18 ...
##  $ retweet_screen_name      : chr  "Jamz5251" NA NA "EMECONOMOU" ...
##  $ source                   : chr  "<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\"
##  $ text                     : chr  "Remember, total bed rest after recovering from any variant of #
##  $ tweet_url                : chr  "https://twitter.com/maashimellows/status/1437271241572769800" "
##  $ user_created_at          : chr  "Mon Mar 12 08:29:33 +0000 2018" "Wed Dec 16 02:50:16 +0000 2020
##  $ user_id                  : num  9.73e+17 1.34e+18 2.96e+08 1.21e+18 5.59e+08 ...
##  $ user_default_profile_image: chr  "false" "false" "false" "false" ...
##  $ user_description         : chr  "a whirlwind of many things ðŸŒ»" "Hi :)" "Just a Dad trying to 
##  $ user_favourites_count    : int  18936 5777 10202 19895 11835 4522 115065 83795 5046 6733 ...
##  $ user_followers_count     : int  5751 103 273 277 92 411 1581 192 119 125 ...
##  $ user_friends_count       : int  907 936 996 959 1029 749 1225 186 313 399 ...
##  $ user_listed_count        : int  7 3 3 1 0 4 2 4 0 2 ...
##  $ user_location            : chr  "Sri Lanka" "Orion Nebula " "Brisbane, Queensland" "Nashville, T
##  $ user_name                : chr  "Amashi." "Torrey Spinelli" "Craig Unthank" "Patriot DAWG fan" .
##  $ user_screen_name         : chr  "maashimellows" "Torrey42997369" "CraigUnthank" "OncoAdvocate" .
##  $ user_statuses_count      : int  21715 6731 1868 10845 735 4228 31610 29553 3197 4784 ...
##  $ user_time_zone           : logi  NA NA NA NA NA NA ...
##  $ user_urls                : chr  "https://medium.com/@maashimellows" NA NA "http://www.natera.com
##  $ user_verified            : chr  "false" "false" "false" "false" ...
```

```r
head(data1) # most informative
```

```
##   coordinates                     created_at hashtags media
## 1        <NA> Mon Sep 13 04:25:41 +0000 2021  COVID19  <NA>
## 2        <NA> Mon Sep 13 04:25:45 +0000 2021     <NA>  <NA>
## 3        <NA> Mon Sep 13 04:25:43 +0000 2021     <NA>  <NA>
## 4        <NA> Mon Sep 13 04:25:42 +0000 2021     <NA>  <NA>
## 5        <NA> Mon Sep 13 04:25:45 +0000 2021     <NA>  <NA>
## 6        <NA> Mon Sep 13 04:25:45 +0000 2021     <NA>  <NA>
##                                                                                     urls
## 1                                                                                   <NA>
## 2                                                                                   <NA>
## 3 https://www.tga.gov.au/media-release/new-restrictions-prescribing-ivermectin-covid-19
## 4                                                                                   <NA>
## 5                                                                                   <NA>
## 6                                                                                   <NA>
##   favorite_count           id in_reply_to_screen_name in_reply_to_status_id
## 1              0 1.437271e+18                    <NA>                    NA
## 2              0 1.437271e+18                TizzyEnt          1.437113e+18
## 3              0 1.437271e+18               jimiuorio          1.437078e+18
## 4              0 1.437271e+18                    <NA>                    NA
## 5              0 1.437271e+18                    <NA>                    NA
## 6              0 1.437271e+18                    <NA>                    NA
##   in_reply_to_user_id lang place possibly_sensitive     quote_id retweet_count
## 1                  NA   en  <NA>               <NA>           NA            21
## 2            27933405   en  <NA>               <NA>           NA             0
## 3            60622883   en  <NA>              false           NA             0
## 4                  NA   en  <NA>               <NA> 1.436738e+18             7
## 5                  NA   en  <NA>               <NA>           NA            41
## 6                  NA   en  <NA>               <NA>           NA            66
##     retweet_id retweet_screen_name
## 1 1.437269e+18            Jamz5251
## 2           NA                <NA>
## 3           NA                <NA>
## 4 1.437264e+18          EMECONOMOU
## 5 1.437002e+18      HINDU_hiteswar
## 6 1.436481e+18       FeistyLibLady
##                                                                                   source
## 1   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
## 2 <a href="http://twitter.com/download/android" rel="nofollow">Twitter for Android</a>
## 3              <a href="https://mobile.twitter.com" rel="nofollow">Twitter Web App</a>
## 4   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
## 5   <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
## 6              <a href="https://mobile.twitter.com" rel="nofollow">Twitter Web App</a>
##
## 1                                    Remember, total bed rest after recovering from any variant of #C
## 2
## 3
## 4
## 5 The No 1 World Class visionary CM of Odisha has found a way to control Corona 3rd wave and that is
## 6                  CDC studies show unðŸ’‰people were 11 times more likely to die of covid than fully
##                                                        tweet_url
## 1  https://twitter.com/maashimellows/status/1437271241572769800
## 2 https://twitter.com/Torrey42997369/status/1437271254818562049
## 3   https://twitter.com/CraigUnthank/status/1437271247130226692
## 4   https://twitter.com/OncoAdvocate/status/1437271244525776900
## 5  https://twitter.com/dillipswain87/status/1437271257397952514
## 6 https://twitter.com/gisellecbalido/status/1437271256894787588
##                  user_created_at      user_id user_default_profile_image
## 1 Mon Mar 12 08:29:33 +0000 2018 9.731138e+17                      false
## 2 Wed Dec 16 02:50:16 +0000 2020 1.339040e+18                      false
## 3 Mon May 09 11:51:27 +0000 2011 2.956319e+08                      false
## 4 Wed Jan 08 02:44:23 +0000 2020 1.214740e+18                      false
## 5 Fri Apr 20 18:57:13 +0000 2012 5.588658e+08                      false
## 6 Thu Jan 09 23:57:42 +0000 2020 1.215422e+18                      false
##
## 1
## 2
## 3 Just a Dad trying to navigate todays fraudulent and manipulated markets for his kids. Structural E
## 4                                                                                             America
## 5                                            à¬†à¬®à‡ à¬"à¬¡à¬¼à¬¿à¬†, à¬à¬¾à¬°à¬¿ à¬¬à¬¢à¬¿à¬†....R
## 6        Florida Editor @ https://t.co/zJdKit8Jr1 & @Floricua|Author of Cubantime|Global citizen|Stu
##   user_favourites_count user_followers_count user_friends_count
## 1                 18936                 5751                907
## 2                  5777                  103                936
## 3                 10202                  273                996
## 4                 19895                  277                959
## 5                 11835                   92               1029
## 6                  4522                  411                749
##   user_listed_count        user_location            user_name user_screen_name
## 1                 7            Sri Lanka              Amashi.    maashimellows
## 2                 3        Orion Nebula       Torrey Spinelli   Torrey42997369
## 3                 3 Brisbane, Queensland        Craig Unthank     CraigUnthank
## 4                 1        Nashville, TN     Patriot DAWG fan     OncoAdvocate
## 5                 0       Cuttack, India Dillip SwainðŸ‡®ðŸ‡³    dillipswain87
## 6                 4                 <NA>     Giselle C Balido   gisellecbalido
##   user_statuses_count user_time_zone                         user_urls
## 1               21715             NA https://medium.com/@maashimellows
## 2                6731             NA                              <NA>
## 3                1868             NA                              <NA>
## 4               10845             NA             http://www.natera.com
## 5                 735             NA        http://www.dillipswain.com
## 6                4228             NA                              <NA>
##   user_verified
## 1         false
## 2         false
## 3         false
## 4         false
## 5         false
## 6         false
```

```r
summary(data1) #not of much information as is mostly text in data set
```

```
##  coordinates         created_at          hashtags            media
##  Length:1456316     Length:1456316     Length:1456316     Length:1456316
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##      urls           favorite_count            id
##  Length:1456316     Min.   :     0.00   Min.   :1.437e+18
##  Class :character   1st Qu.:     0.00   1st Qu.:1.437e+18
##  Mode  :character   Median :     0.00   Median :1.437e+18
##                      Mean   :     3.69   Mean   :1.437e+18
##                      3rd Qu.:     0.00   3rd Qu.:1.437e+18
##                      Max.   :105850.00   Max.   :1.437e+18
##
##  in_reply_to_screen_name in_reply_to_status_id in_reply_to_user_id
##  Length:1456316          Min.   :2.167e+10     Min.   :1.200e+01
##  Class :character        1st Qu.:1.437e+18     1st Qu.:8.173e+07
##  Mode  :character        Median :1.437e+18     Median :1.338e+09
##                          Mean   :1.436e+18     Mean   :4.260e+17
##                          3rd Qu.:1.437e+18     3rd Qu.:1.051e+18
##                          Max.   :1.437e+18     Max.   :1.437e+18
##                          NA's   :1257946       NA's   :1249642
##      lang             place          possibly_sensitive    quote_id
##  Length:1456316     Length:1456316     Length:1456316     Min.   :3.191e+17
##  Class :character   Class :character   Class :character   1st Qu.:1.437e+18
##  Mode  :character   Mode  :character   Mode  :character   Median :1.437e+18
##                                                           Mean   :1.435e+18
##                                                           3rd Qu.:1.437e+18
##                                                           Max.   :1.437e+18
##                                                           NA's   :1133487
##  retweet_count      retweet_id       retweet_screen_name    source
##  Min.   :     0   Min.   :3.394e+17   Length:1456316      Length:1456316
##  1st Qu.:     1   1st Qu.:1.437e+18   Class :character    Class :character
##  Median :    58   Median :1.437e+18   Mode  :character    Mode  :character
##  Mean   :  2440   Mean   :1.437e+18
##  3rd Qu.:  1001   3rd Qu.:1.437e+18
##  Max.   :450213   Max.   :1.437e+18
##                   NA's   :441902
##      text            tweet_url         user_created_at       user_id
##  Length:1456316     Length:1456316     Length:1456316     Min.   :2.210e+02
##  Class :character   Class :character   Class :character   1st Qu.:3.646e+08
##  Mode  :character   Mode  :character   Mode  :character   Median :3.424e+09
##                                                           Mean   :5.475e+17
##                                                           3rd Qu.:1.198e+18
##                                                           Max.   :1.437e+18
##
##  user_default_profile_image user_description   user_favourites_count
##  Length:1456316             Length:1456316     Min.   :      0
##  Class :character           Class :character   1st Qu.:   4155
##  Mode  :character           Mode  :character   Median :  19463
##                                                Mean   :  55689
##                                                3rd Qu.:  64057
##                                                Max.   :2815954
##
##  user_followers_count user_friends_count user_listed_count  user_location
##  Min.   :       0     Min.   :      0    Min.   :     0.0   Length:1456316
##  1st Qu.:     126     1st Qu.:    254    1st Qu.:     0.0   Class :character
##  Median :     481     Median :    708    Median :     2.0   Mode  :character
##  Mean   :   12633     Mean   :   1909    Mean   :    75.6
##  3rd Qu.:    1742     3rd Qu.:   2022    3rd Qu.:    12.0
##  Max.   :55009637     Max.   :2094245    Max.   :211454.0
##
##   user_name         user_screen_name   user_statuses_count user_time_zone
##  Length:1456316     Length:1456316     Min.   :      1     Mode:logical
##  Class :character   Class :character   1st Qu.:   5178     NA's:1456316
##  Mode  :character   Mode  :character   Median :  19310
##                                        Mean   :  60789
##                                        3rd Qu.:  61866
##                                        Max.   :5145057
##
##   user_urls         user_verified
##  Length:1456316     Length:1456316
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

```
#number of record that include a value for fields: user_location,coordinates,place: This fields could b
length(data1$user_location)-length(which(is.na(data1$user_location)))
```

```
## [1] 925624
```

```
length(data1$coordinates)-length(which(is.na(data1$coordinates)))
```

```
## [1] 94
```

```
length(data1$place)-length(which(is.na(data1$place)))
```

```
## [1] 7403
```

```
#Ti inspect the appropriateness for strata building

#print(data1$user_location) #best option as has least amount of NA, but needs to clean up list city, co
head(data1$user_location)
```

```
## [1] "Sri Lanka"            "Orion Nebula "        "Brisbane, Queensland"
## [4] "Nashville, TN"        "Cuttack, India"       NA
```

```
#print(data1$coordinates) #cleanest list with data points
head(data1$coordinates)
```

```
## [1] NA NA NA NA NA NA
```

```
#print(data1$place)
head(data1$place)
```

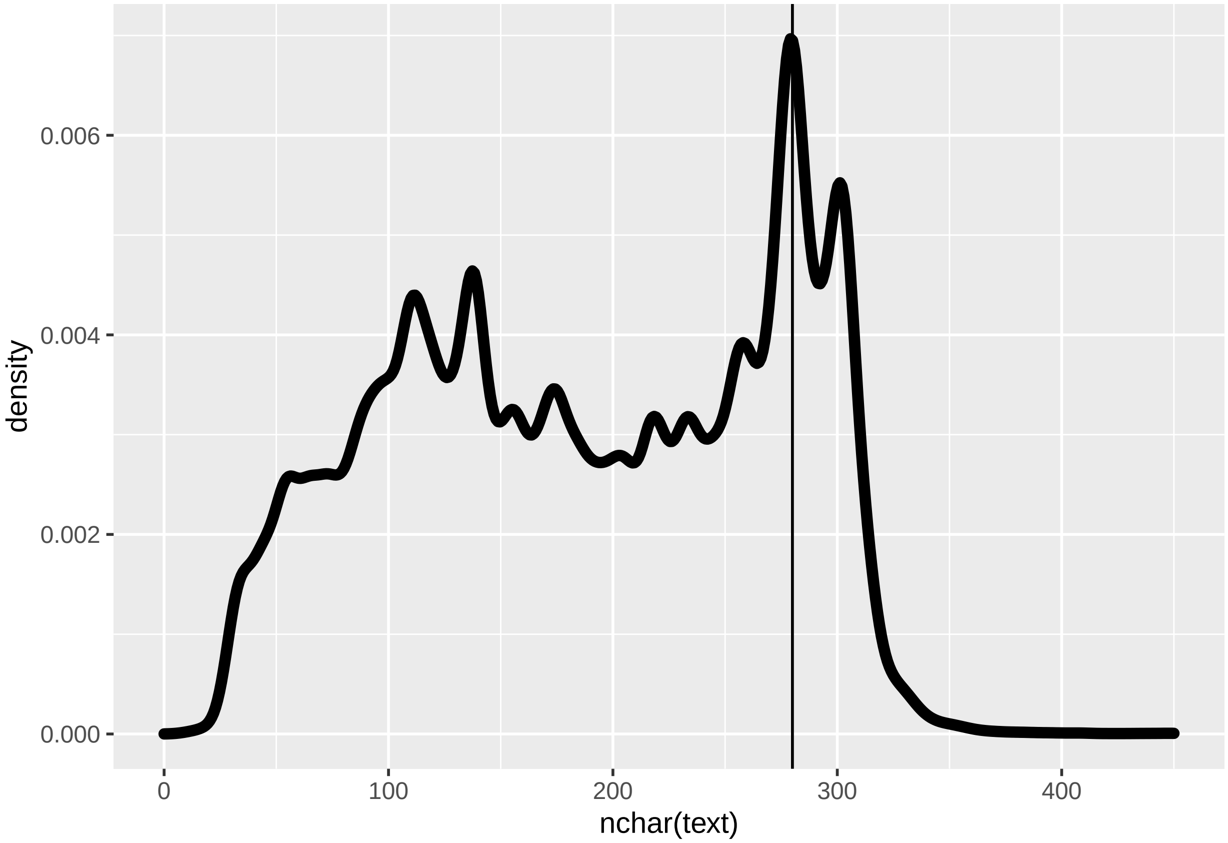
```
## [1] NA NA NA NA NA NA
```

```
#distribution of the number of characters in the data set attribute text / tweets content

ggplot(data = data1, aes(x = nchar(text))) + geom_density(size = 2) + geom_vline(xintercept = 280) + sca
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```

```
## Warning: Removed 1614 rows containing non-finite values (stat_density).
```



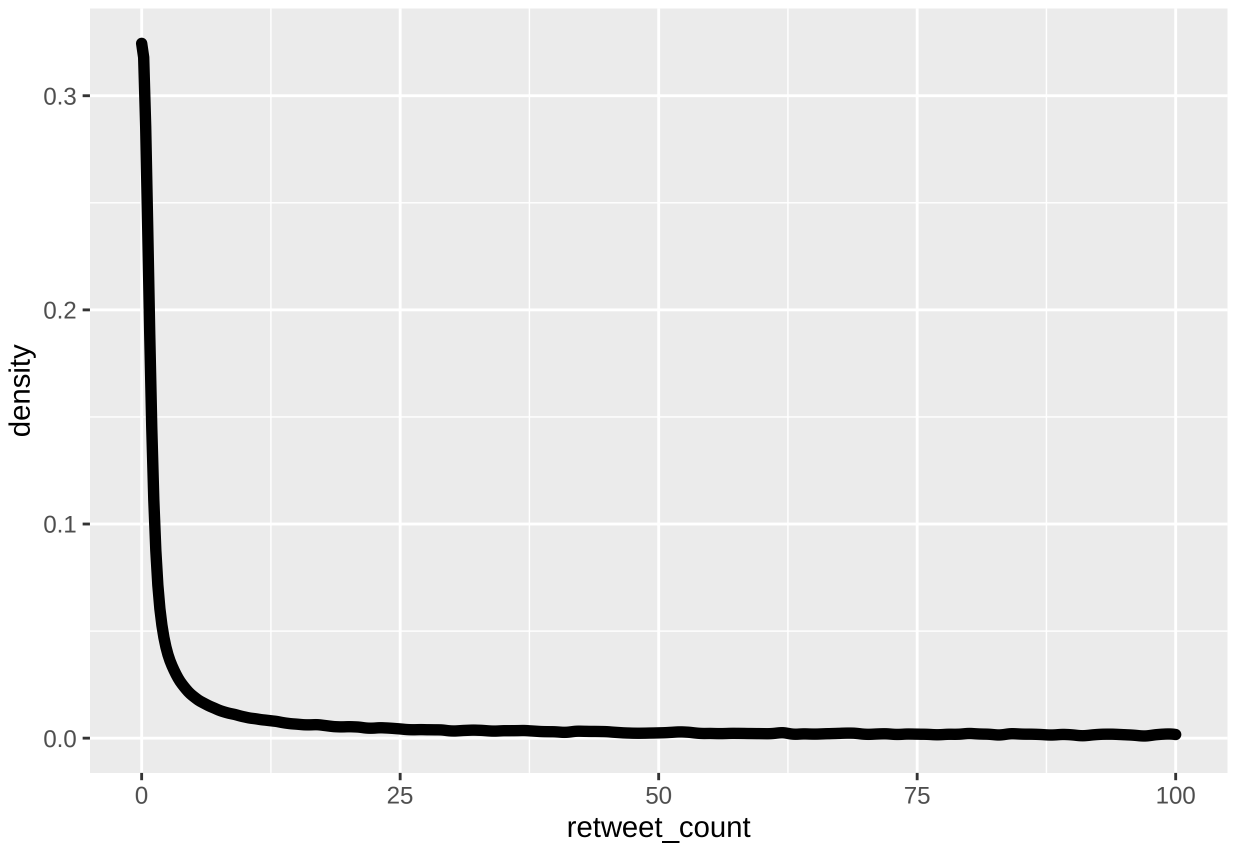
```
#This is a density graph : Computes and draws kernel density estimate, which is a smoothed version of t

#Conclusion: max number of characters per tweet is set at 280 by Twitter as can also been seen in the g

#Note: to remove scientific numbering , first create object p <- ggplot()
# p + scale_x_continuous(labels = function(x) format(x, scientific = FALSE))
```

```
# showing count of retweets in data set
ggplot(data = data1, aes(x = retweet_count)) + geom_density(size = 2) + xlim(0,100)
```

```
## Warning: Removed 665994 rows containing non-finite values (stat_density).
```



```
#Conslusion: only a few tweets are retweeted frequently.
```

```
#split attribute Coordinates into two columns
CoordinateDF <- data.frame(x = data1$coordinates)

SplitCoordinate <- CoordinateDF %>% separate(x, c("long","lat"), sep = "([,])")
```

```
#remove NAs
CoordinatesremoveNA <- na.omit(SplitCoordinate)

CoordinatesremoveNA$long <- as.numeric(CoordinatesremoveNA$long)
CoordinatesremoveNA$lat <- as.numeric(CoordinatesremoveNA$lat)
```

```
#building a world map of countries.
#Source: https://r-spatial.org/r/2018/10/25/ggplot2-sf.html#:~:text=This%20call%20nicely%20introduces%2

library(ggplot2)
theme_set(theme_bw())
library(sf)
```

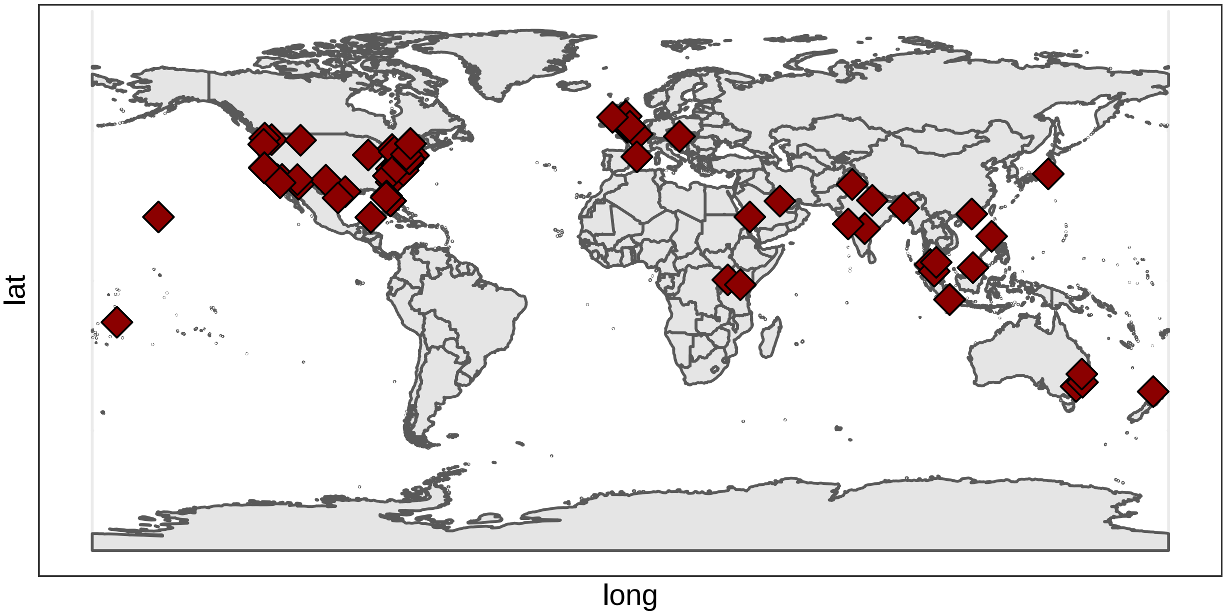
```
## Linking to GEOS 3.9.1, GDAL 3.2.1, PROJ 7.2.1
```

```
library("rnaturalearth")
library("rnaturalearthdata")
```

```
world <- ne_countries(scale = "medium", returnclass = "sf")
class(world)
```

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

```
#plotting data set to see geographical spread
ggplot(data = world) +
  geom_sf() +
  geom_point(data = CoordinatesremoveNA, aes(x = long, y = lat), size = 4,
             shape = 23, fill = "darkred")
```



```
# Zoom in by adding: + coord_sf(xlim = c(-88, -78), ylim = c(24.5, 33), expand = FALSE)

#save graph to PDF:
ggsave("mapdataset Sept2021.pdf")
```

```
## Saving 6.5 x 4.5 in image
```

```
#show table with country names using photon


#install.packages('revgeo')
```

```
#library(devtools)
#install_github('mhudecheck/revgeo')

#library(revgeo)

start <- Sys.time()
#This line do all the reverse geocoding using Photon as a provider
results<-revgeo(longitude=CoordinatesremoveNA$long,
                latitude=CoordinatesremoveNA$lat,
                provider = 'photon', output="frame")
```

```
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=149.0676&lat=-35
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-171.74671143&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-74.63516&lat=40
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=39.94367123&lat=2
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-79.7164011&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-122.4131575&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=120.812&lat=14.84
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-74.46848886&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-119.99490738&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-79.7164011&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=2.15168&lat=41.39
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-110.9708&lat=32
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=78.31565354&lat=1
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=32.5811&lat=0.313
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=106.8335573&lat=-
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=151.20797&lat=-33
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=150.98933&lat=-3
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-116.5161305&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=16.41560835&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-1.31348562&lat=5
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=114.1726015&lat=2
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-1.5881164&lat=55
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-72.66276246&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.56725453&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-1.07269049&lat=5
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=114.566667&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=100.295847&lat=5
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.9633593&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-122.30706871&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=49.997632&lat=26
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=106.7342665&lat=-
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=101.63504084&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=36.7667&lat=-1.3
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=78.380978&lat=17
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-101.878&lat=33.5
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=174.85938702&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=114.1726015&lat=2
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=139.76194909&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-122.675&lat=45.5
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-110.326&lat=46.9
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-1.31348562&lat=5
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=74.1833&lat=32.15
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.692001&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.692001&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-75.22486746&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-87.687672&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.692001&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.692001&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-157.85385272&la
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-77.7436&lat=39.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-117.868184&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-1.58224081&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-72.54854947&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=72.8381&lat=18.9
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-0.12485951&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=16.41560835&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-1.3911345&lat=5
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-80.031475&lat=3
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=80.91&lat=26.85"
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-95.3694&lat=29.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=102.283&lat=6.16
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-86.82475999&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-79.9581&lat=37.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=2.29493333&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-122.4396537&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-118.15649367&la
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-80.95853882&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-78.02873528&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-0.17988195&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.56725453&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-76.9815877&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-111.388&lat=33.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=2.15288&lat=41.3
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-78.9592&lat=43.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.56725453&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.56725453&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-75.9787&lat=36.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-81.5401&lat=28.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-78.2057&lat=37.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-74.13545781&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-97.7975&lat=27.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-80.24741765&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=91.37802124&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-122.41374814&la
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-80.24741765&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-80.24741765&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.56725453&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.8178769&lat=
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-117.1610838&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.941941&lat=4
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-74.9384&lat=42.
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-5.9811359&lat=5
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-73.56725453&lat
## [1] "Getting geocode data from Photon: https://photon.komoot.io/reverse?lang=en&lon=-81.717&lat=27.8
```

```
end <- Sys.time()

str(results)
```

```
## 'data.frame':    94 obs. of  6 variables:
##  $ housenumber: chr  "12" "House Number Not Found" "House Number Not Found" "House Number Not Found"
##  $ street     : chr  "Pecan Drive" "<U+0637><U+0631><U+064A><U+0642> <U+0627><U+0644><U+0623><U+0645
##  $ city       : chr  "Brampton" "City Not Found" "Apia" "District of Belconnen" ...
##  $ state      : chr  "Ontario" "Makkah Region" "Tuamasaga" "Australian Capital Territory" ...
##  $ zip        : chr  "L6P 2X4" "Postcode Not Found" "Postcode Not Found" "2617" ...
##  $ country    : chr  "Canada" "Saudi Arabia" "Samoa" "Australia" ...
```

```
#save object, results.
saveRDS(results, file = "resultsSept.Rds")
```

```
#getwd()
#setwd("C:/Ryerson University - Capstone project/Module 2/EIEEE - Large dataset/Combined")
```

```
#load object results


results <- readRDS(file = "resultsSept.Rds")
#str(results)
```

```
#Create list frequency by city

#install.packages("stats")

#aggregate(results$city, by=list(results$city), FUN=length)
res <- aggregate(results$city, by=list(results$city), FUN=length)
#head(res, 40)
#res[order(res$x, decreasing = TRUE),]
```

```
#Create a table and graph with more than 10 tweets per city
# save as dataframe, then plot frequency in ggplot
Locations <- data.frame(res[order(res$x, decreasing = TRUE),])
str(Locations)
```

```
## 'data.frame':    60 obs. of  2 variables:
##  $ Group.1: chr  "City Not Found" "Mascouche" "Elmont" "Loxahatchee Groves" ...
##  $ x      : int  17 6 4 3 3 2 2 2 2 2 ...
```

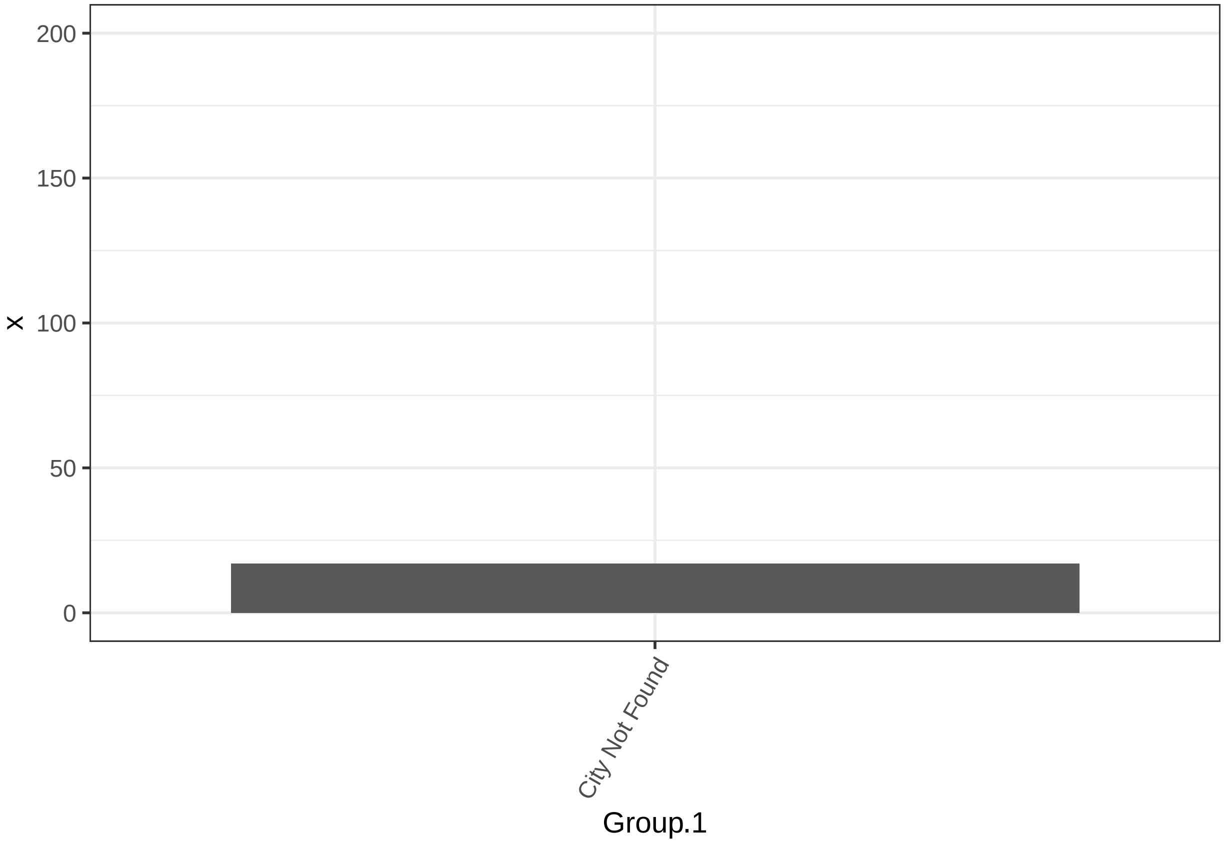
```
Locations$x = as.numeric(Locations$x)
length(Locations$x) #out of 1,332 coordinates (long,lat), only 571 returned with a city name including
```

```
## [1] 60
```

```
newdf <- subset(Locations, x > 10)
newdf
```

```
##           Group.1  x
## 10 City Not Found 17
```

```
ggplot(newdf,aes(x=Group.1, y=x)) + geom_bar(stat = 'identity') + scale_y_continuous(limits = c(0, 200))
```



```
#+ scale_x_discrete(name ='x')
```