Twitter

Web Monitoring Group

# The Data

Using data already stored on computer

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.4.3

library(stringr)  
library(twitteR)  
  
register\_sqlite\_backend("../data/nesreanigeria.db")  
dframe <- load\_tweets\_db(as.data.frame = TRUE,  
 table\_name = "nesreanigeria\_tweets")  
twList <- load\_tweets\_db(as.data.frame = FALSE,  
 table\_name = "nesreanigeria\_tweets")

## Data cleaning

Remove unwanted and add extra variable columns

dframe <- dframe %>%  
 select(text, favoriteCount, created, statusSource:isRetweet) %>%  
 mutate(month = factor(format(as.Date(created), "%B"), levels = month.name, ordered = TRUE))  
  
  
glimpse(dframe)

## Observations: 3,439  
## Variables: 8  
## $ text <chr> "RT @ecomsaoauife: We specially appreciate @NESR...  
## $ favoriteCount <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 1, ...  
## $ created <dttm> 2016-12-29 11:55:58, 2016-12-29 11:33:01, 2016-...  
## $ statusSource <chr> "<a href=\"http://www.hootsuite.com\" rel=\"nofo...  
## $ screenName <chr> "NESREANigeria", "riverine\_child", "riverine\_chi...  
## $ retweetCount <dbl> 1, 0, 0, 5, 5, 0, 5, 2, 5, 1, 1, 2, 1, 1, 1, 2, ...  
## $ isRetweet <lgl> TRUE, FALSE, FALSE, TRUE, TRUE, FALSE, TRUE, TRU...  
## $ month <ord> December, December, December, December, December...

summary(dframe)

## text favoriteCount created   
## Length:3439 Min. : 0.000 Min. :2016-12-19 21:20:07   
## Class :character 1st Qu.: 0.000 1st Qu.:2017-02-20 07:19:29   
## Mode :character Median : 0.000 Median :2017-05-30 13:34:49   
## Mean : 0.549 Mean :2017-06-20 17:05:28   
## 3rd Qu.: 0.000 3rd Qu.:2017-11-24 05:05:30   
## Max. :22.000 Max. :2017-12-23 21:03:58   
##   
## statusSource screenName retweetCount isRetweet   
## Length:3439 Length:3439 Min. : 0.00 Mode :logical   
## Class :character Class :character 1st Qu.: 1.00 FALSE:1383   
## Mode :character Mode :character Median : 3.00 TRUE :2056   
## Mean : 18.38   
## 3rd Qu.: 7.50   
## Max. :30569.00   
##   
## month   
## December:948   
## February:794   
## March :334   
## June :321   
## May :235   
## October :150   
## (Other) :657

Remove non-text characters from the *text* variable and remove the URLs in the *statusSource*

dframe$text <- str\_replace(dframe$text, "[^[:graph:]]", " ")  
  
dframe$statusSource <- substr(dframe$statusSource,  
 start = regexpr(">", dframe$statusSource) + 1,  
 stop = regexpr("</a>", dframe$statusSource) - 1)  
  
head(dframe$statusSource)

## [1] "Hootsuite" "TweetDeck" "TweetDeck"   
## [4] "Twitter for iPhone" "Twitter for iPad" "Facebook"

tbl\_df(dframe)

## # A tibble: 3,439 x 8  
## text  
## <chr>  
## 1 RT @ecomsaoauife: We specially appreciate @NESREANigeria and @ncfnigeria fo  
## 2 .@riversstategov adding no certainty &amp; spreading misinformation. @NESRE  
## 3 .@airqualityindia @AirPollutionnew Hasn't improved. Dry season is making pr  
## 4 RT @ecomsaoauife: @AminaJMohammed @estherclimate @iujibril Special thanks t  
## 5 RT @ecomsaoauife: @AminaJMohammed @estherclimate @iujibril Special thanks t  
## 6 "Retweeted ECOMSA (@ecomsaoauife):\n\n@AminaJMohammed @estherclimate @iujib  
## 7 RT @ecomsaoauife: @AminaJMohammed @estherclimate @iujibril Special thanks t  
## 8 RT @ecomsaoauife: @AminaJMohammed @iujibril @FMEnvng @NESREANigeria Plans f  
## 9 RT @ecomsaoauife: @AminaJMohammed @estherclimate @iujibril Special thanks t  
## 10 RT @ecomsaoauife: @AminaJMohammed @iujibril @FMEnvng @NESREANigeria 4 PhD a  
## # ... with 3,429 more rows, and 7 more variables: favoriteCount <dbl>,  
## # created <dttm>, statusSource <chr>, screenName <chr>,  
## # retweetCount <dbl>, isRetweet <lgl>, month <ord>

mostfav <- dframe$text[which(dframe$favoriteCount == max(dframe$favoriteCount))]  
cat("Most favorited tweet(s):\n ", mostfav, "\n")

## Most favorited tweet(s):  
## You are invited for the 11th National Stakeholders' Forum and Celebration of NESREA@10 https://t.co/ucdVxlLxNi

rm(mostfav)

mostRts <- dframe$text[which(dframe$retweetCount == max(dframe$retweetCount))]  
cat("Most retweeted:\n", mostRts, "\n")

## Most retweeted:  
## RT @Pontifex: We must never forget that the natural environment is a collective good, the patrimony of all humanity and the responsibility<U+0085>

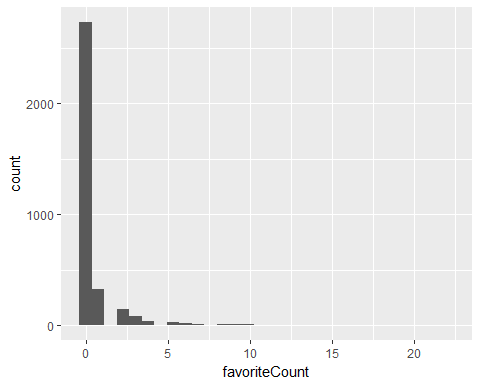
rm(mostRts)

sort(table(dframe$statusSource), decreasing = TRUE)

##   
## Twitter for Android Twitter for iPhone   
## 1679 558   
## Twitter Web Client Hootsuite   
## 369 330   
## Twitter for BlackBerry Twitter for iPad   
## 173 91   
## TweetDeck Twitter Lite   
## 77 49   
## Mobile Web (M2) Twitter for Windows Phone   
## 29 27   
## Twitter for Windows Mobile Web (M5)   
## 11 9   
## Facebook UberSocial for Android   
## 6 6   
## IFTTT Linkis: turn sharing into growth   
## 4 3   
## Put your button on any page! Twitter for BlackBerry®   
## 3 2   
## Twitterbot SEC amazing-news   
## 2 1   
## Commun.it Ecoexperttm   
## 1 1   
## Fair Trade Neeraj Jagga ki app   
## 1 1   
## Paper.li RoundTeam   
## 1 1   
## SDGsBot Storify   
## 1 1   
## Twitter for Android Tablets Unilag My Campus LifeRetweet Bot   
## 1 1

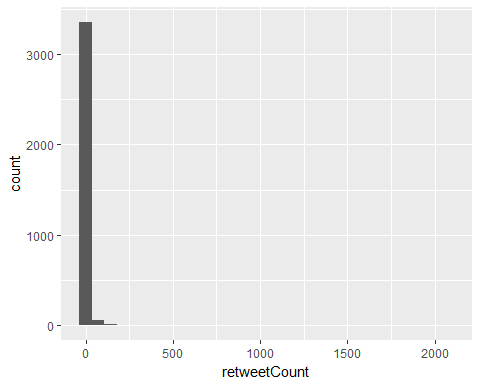
library(ggplot2)  
  
ggObj <- ggplot(dframe, aes(favoriteCount)) +  
 geom\_histogram()  
ggObj

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



minusPontifex <- dframe[which(dframe$retweetCount != max(dframe$retweetCount)), ]  
  
ggObj <- ggplot(minusPontifex, aes(retweetCount)) +  
 geom\_histogram()  
ggObj

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



rm(ggObj)

library(wordcloud)

## Loading required package: methods

## Loading required package: RColorBrewer

library(tm)

## Loading required package: NLP

##   
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':  
##   
## annotate

txt <- dframe %>%  
 filter(isRetweet == "FALSE") %>%  
 select(text) %>%  
 unlist() %>%  
 paste(collapse = " ") %>%  
 gsub("@[[:alnum:]]{3,}|amp", " ", .) %>%  
 str\_trim() %>%  
 stripWhitespace() %>%  
 removeWords(stopwords("english")) %>%  
 removePunctuation(preserve = TRUE) %>%  
 removeNumbers()   
  
oldpar <- par()  
par(margin = rep(.1, 4))

## Warning in par(margin = rep(0.1, 4)): "margin" is not a graphical parameter

wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2"))

## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## environmental could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## regnesrea could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## stakeholders could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## enforcement could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## thanks could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## nigeria could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## emission could not be fit on page. It will not be plotted.

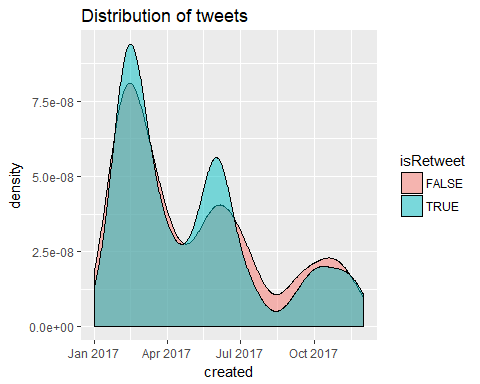
## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## harcourt could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, min.freq = 10, colors = brewer.pal(6, "Dark2")):  
## working could not be fit on page. It will not be plotted.

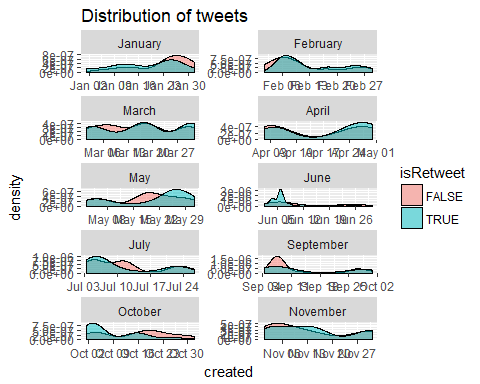


suppressWarnings(par(oldpar))  
rm(oldpar, txt)

dframe\_x <- dframe[dframe$month != "December", ]  
yearplot <- ggplot(dframe\_x, aes(created)) +  
 geom\_density(aes(fill = isRetweet), alpha = 0.5) +  
 ggtitle("Distribution of tweets")  
yearplot



yearplot +  
 facet\_wrap( ~ month, ncol = 2, scales = "free")



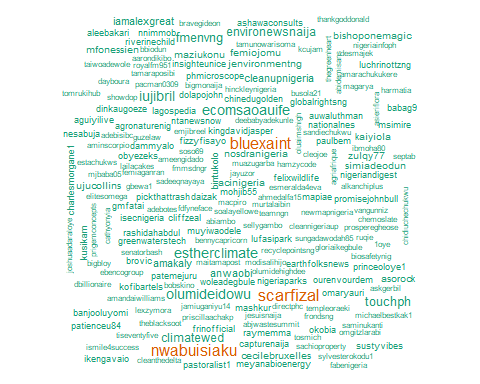
txt <- dframe$screenName %>%  
 paste(collapse = " ") %>%  
 str\_trim() %>%  
 stripWhitespace()  
  
wordcloud(txt, colors = brewer.pal(12, "Dark2"))

## Warning in brewer.pal(12, "Dark2"): n too large, allowed maximum for palette Dark2 is 8  
## Returning the palette you asked for with that many colors

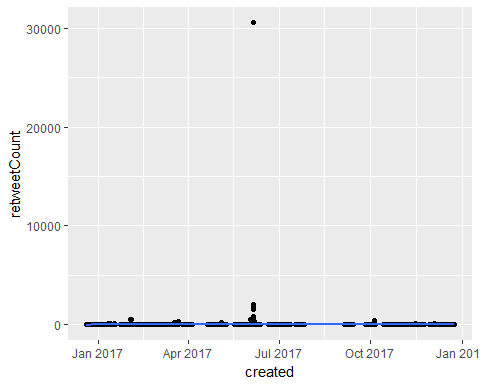
## Warning in wordcloud(txt, colors = brewer.pal(12, "Dark2")): nesreanigeria  
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, colors = brewer.pal(12, "Dark2")): aminajmohammed  
## could not be fit on page. It will not be plotted.

## Warning in wordcloud(txt, colors = brewer.pal(12, "Dark2")): merrymakerz12  
## could not be fit on page. It will not be plotted.

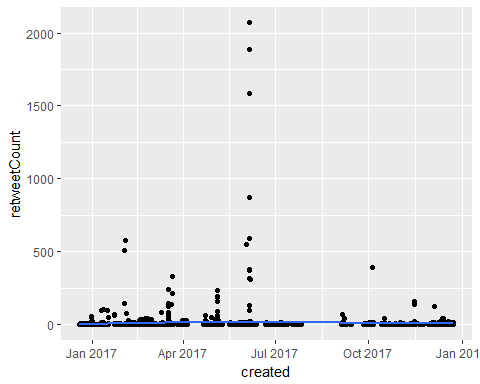


rtTrend <- ggplot(dframe, aes(created, retweetCount)) +  
 geom\_point(aes(y = retweetCount)) +   
 geom\_smooth(method = "loess")  
rtTrend

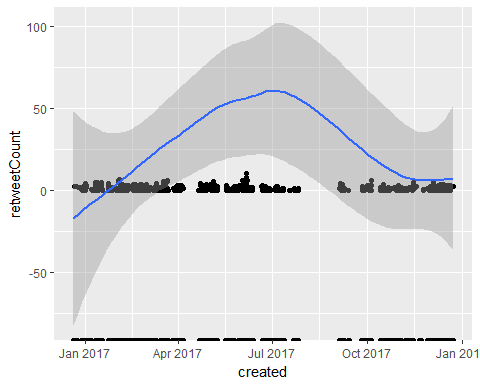


After removing the outlier

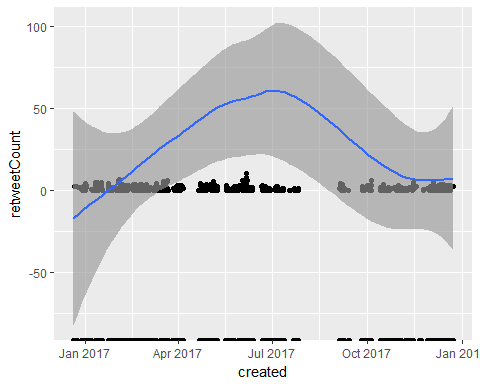
sansOutlier <- ggplot(minusPontifex, aes(created, retweetCount)) +  
 geom\_point(aes(y = retweetCount)) +  
 geom\_smooth(method = "loess")  
  
sansOutlier



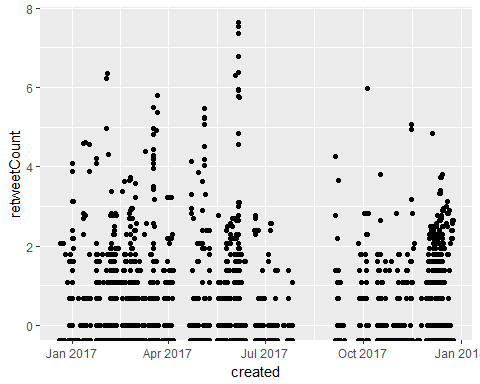
rtLogTrend <- ggplot(dframe, aes(created, retweetCount)) +  
 geom\_point(aes(y = log(retweetCount))) +  
 geom\_smooth(method = "loess")  
rtLogTrend



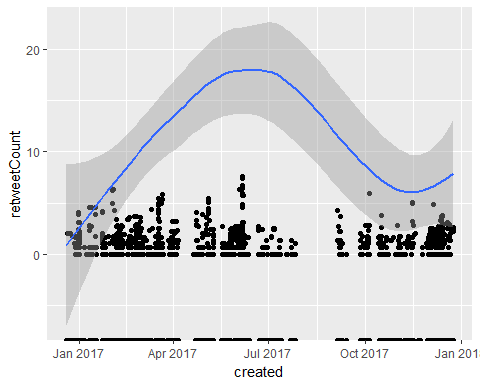
rtLogTrend +  
 geom\_smooth(method = "loess")



rtLogTrend2 <- ggplot(minusPontifex, aes(created, retweetCount)) +  
 geom\_point(aes(y = log(retweetCount)))  
  
rtLogTrend2

 Without the Pope’s tweet (outlier)

rtLogTrend2 +  
 geom\_smooth(method = "loess")



## User Data