
Data Visualization

Selam Tesfamariam

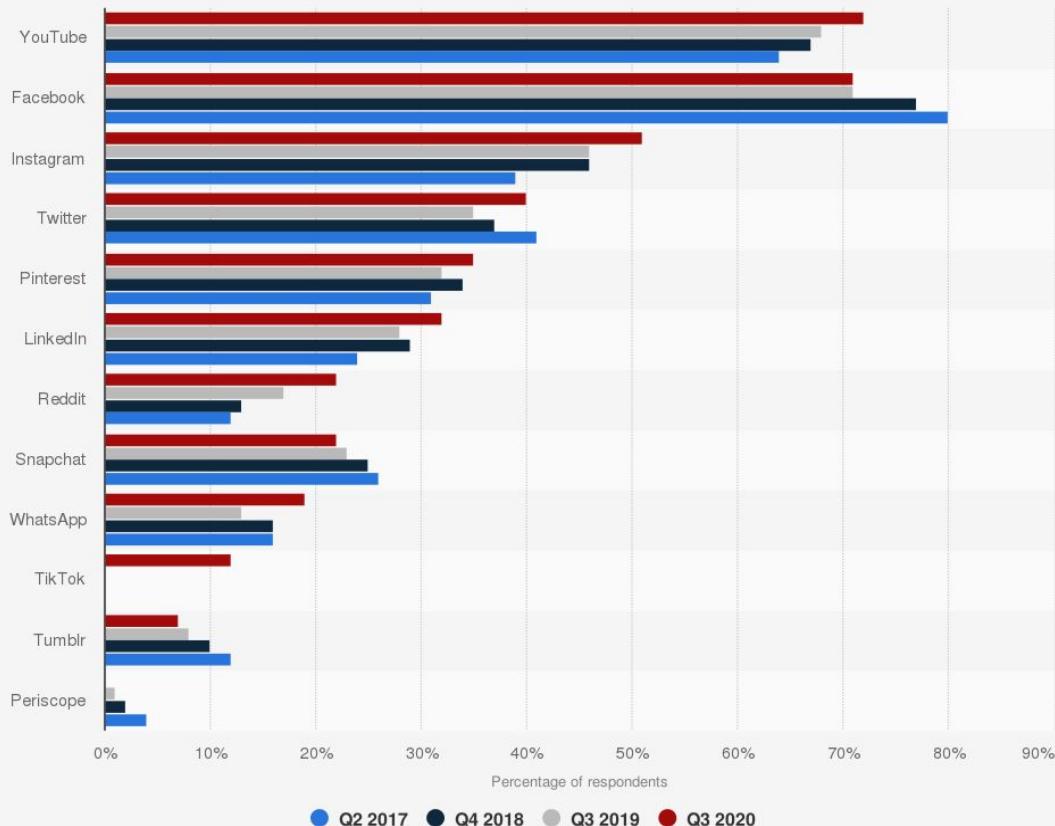
Research Statement

- To test the use of social media scraping to understand customer relations
 - Insurance companies: Geico, State Farm, Progressive
- Problem
 - Not enough tweets
 - The use of Twitter varied
- Test the different methods or ways twitter scraping can be used
 - #BLM
 - #VOTE
 - Starbucks
 - Southwest Airlines
 - Nike

Social Media Usage

- Social Media usage has increased in the past few years
 - Twitter Scraping
- Excel
- wordclouds.com
- Rstudio
 - ggplot
 - ggplot2
 - tidyverse

Share of online users in the United States who use selected social media platforms as of 3rd quarter 2020



Source
AudienceProject
© Statista 2020

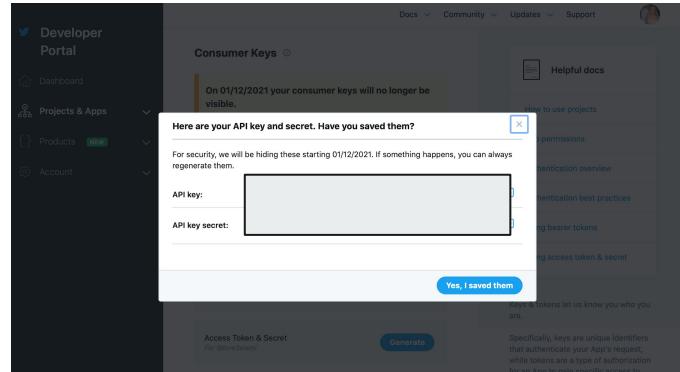
Additional Information:
United States; AudienceProject; Q2 2017, Q4 2018, Q3 2019 and Q3 2020; 2,000+; 15 years and older; Online panel



Twitter Scraping

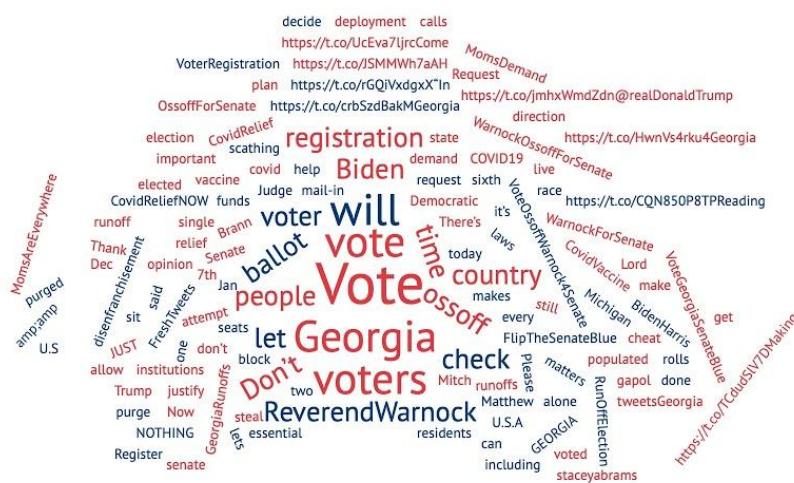
1. Create Developers Account
2. Obtain Consumer Keys
3. RStudio

```
1 ##TWITTER API
2
3 rm(list=ls())
4 install.packages(c("rtweet", "igraph", "tidyverse", "ggraph", "data.table"), repos = "https://cran.r-project.org")
5
6 library(rtweet)
7 library(igraph)
8 library(tidyverse)
9 library(ggraph)
10 library(data.table)
11 install.packages("widgetframe")
12 library(widgetframe)
13 install.packages("twitteR")
14 library(twitteR)
15 library("lubridate")
16 library("ggplot2")
17 install.packages(c("tm", "wordcloud"))
18 library("tm")
19 library("wordcloud")
20
21 token <- rtweet::create_token(
22   app = "Data Review"
23   consumer_key <- "
24   consumer_secret < "
25   access_token <- "
26   access_secret < "
27
28 ## Check token
29
```



#VOTE

- Extracted tweets
 - Created a dataframe
 - Exported table to Excel
 - Worldcloud.com



#BLM

- Extracted tweets
 - Created a dataframe
 - Exported table to Excel
 - Worldcloud.com



R Codes

```
##TRENDS ON TWITTER

blm <- rtweet::search_tweets(q = "#BLM", n = 5000, lang = "en")

dfblm <- data.frame(tweets = c("#ProLife people are hilarious when it comes to #KyleRittenhouse murdering 2 people! Why aren't they also protesting when cops murder people? You c
  "BLM-antifa repeatedly hurl degrading insults at a woman holding a pro-@OmahaPolice sign outside the police headquarters tonight. #BLM https://t.co.
  "@AP We live in a world where people raise $2Mil to free a killer from jail but 18yr old #SincerePierce, who was killed by the police last week, fa
  "It's a great day to remember silence is violence. #blm",
  "BLM-antifa repeatedly hurl degrading insults at a woman holding a pro-@OmahaPolice sign outside the police headquarters tonight. #BLM https://t.co.
  "Today I learned: In #america you can #murder #people and become a #celebrity. #KyleRittenhouse #WhiteSupremacy #BLM #thanks @realDonaldTrump for M
  "That one day we will truly be unified. #BLM https://t.co/0A7nI2ceWJ",
  "This is @cornell. After a summer of saying "#BLM" this is how Black students who speak up are treated. https://t.co/uyTdEE07tr",
  "Complicit Police do nothing. If these activists were having an armed #BLM protest, Police wouldn't be in attack mode. https://t.co/jdfNL72vI3"))

print(df)

install.packages("writexl")
library("writexl")
write_xlsx(dfblm,"/Users/selamtesfamariam/Desktop/EPPS 6356/data.xlsx")

vote <- rtweet::search_tweets(q = "#vote", n = 100, lang = "en")

dfvote <- data.frame(tweets = c("Georgia voters check your registration!! Don't allow them to purge & cheat. This senate race is too important ! #vote #VoteGeorgiaSenateBlu
  "@realDonaldTrump Thank the Lord that Biden was elected and Trump will have NOTHING to do with deployment of the #COVID19 vaccine #BidenHarris #Co
  "Making calls to #Georgia voters. The runoff election for U.S. will decide the direction of Georgia and our country. #vote #MomsDemand #MomsAreEv
  "Georgia, "lets get this done again" ~staceyabrams 🗳️ Register 🗳️ Request your ballot 🗳️ make a plan and #Vote 🗳️ @ReverendWarnock and @ossoff #gapo
  "Reading that people who JUST voted in #Georgia are being purged from the voter registration rolls If you live in Georgia check your #VoterRegistr
  "Georgia residents! There's still time to request your mail-in ballot for the essential 1/5 Senate runoffs! Please do it today. If those two seats
  "'In U.S.A. , this cannot justify the disenfranchisement of a single voter, let alone all the voters of its sixth most populated state," the scath
  "Come on, #Georgia! Don't sit this one out. Your #vote matters. https://t.co/JSMWh7aAH"))

write_xlsx(dfvote,"/Users/selamtesfamariam/Desktop/EPPS 6356/data2.xlsx")
```

Twitter + Business

- Market Trends Identified
- Target Market Segmentation
- Customer Sentiment Measurement
- Hashtags
 - Engage customers
 - Introduce and idea or product
 - Can be a success or fail

 Disney+  @disneyplus 

It. Is. Time. From Snow White and the Seven Dwarfs to The Mandalorian, check out basically everything coming to [#DisneyPlus](#) in the U.S. on November 12.

Pre-order in the U.S. at [DisneyPlus.com](#) today:
di.sn/600511FbJ



8:05 AM · Oct 14, 2019 

 66.9K  21.4K people are Tweeting about this



Old Spice

@OldSpice

9 Jul 12

Why is it that "fire sauce" isn't made with any real fire? Seems like false advertising.



Taco Bell

@tacobell

[Follow](#)

@OldSpice Is your deodorant made with really old spices?

10:52 PM - 9 Jul 2012

1,042 782

Wendy's @Wendys · 1h

It's go time! You want the roast? Then let's go!
#NationalRoastDay

5,516 1,218 7,117



Hooters @Hooters · 46m

Whatcha got?

9 7 201

Wendy's @Wendys

Replying to @Hooters

Uniforms our employees can wear in the winter.

9:40 PM · 04 Jan 19 · Twitter Web Client

87 Retweets 920 Likes



RUFFLES

@RUFFLES

Follow

When your roommate eats all the Ruffles.



1:58 PM - 29 Jul 2016

28 Retweets 49 Likes

2 28 49



Chipotle @ChipotleTweets · Mar 12

The moment you hold your baby for the first time.



962 1.7K



Tony X @soloudcity · Jun 11, 2018

so @Wendys u just gonna let @IHOB sell burgers on your block?
thought you were the og?



Wendy's

Not really afraid of the burgers from a place that decided
pancakes were too hard.

11:36 AM · Jun 11, 2018

280.1K 95.2K people are Tweeting about this

17h



@Wendys My friend wants to go McDonalds, what should I tell
him?



Wendy's
@Wendys

[Follow](#)

Find new friends.

8:46 PM - 3 Jan 2017

31 128



Follow

@Wendys good morning how did you sleep

4:03 PM - 3 Jan 2017

1



Wendy's
@Wendys

Restaurants don't sleep.

4:04 PM - 3 Jan 2017

2 26



@Wendys you could've just said good and left it at
that

4:08 PM - 3 Jan 2017

3



Wendy's
@Wendys

We aren't here to lie though

4:11 PM - 3 Jan 2017

2 35



Chipotle
@ChipotleTweets

Good morning to everyone, except people
who take our Tabasco bottles.

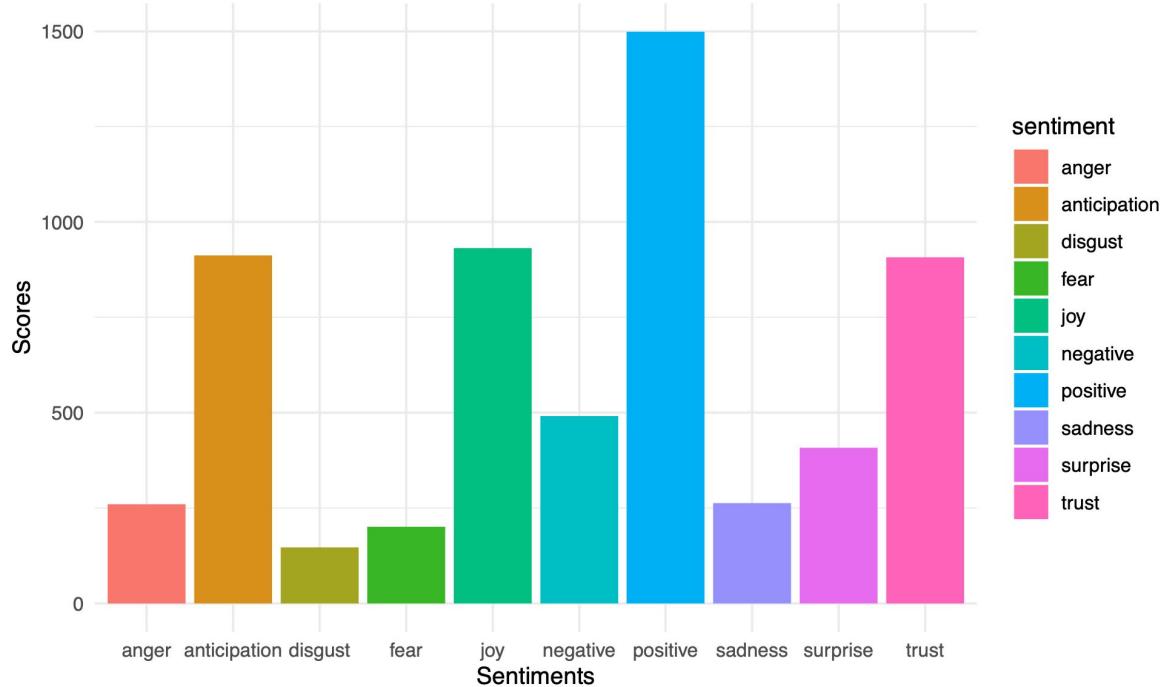
11:00 AM - 26 May 2018

26,612 Retweets 113,070 Likes

2.3K 27K 113K



Total Sentiment of Tweets under #Starbucks



- Starbucks utilized the platform to engage with customer
- 10.9M Followers
- Overall positive sentiment

Code for Starbucks Sentiment Graph

```
sb <- rtweet::search_tweets(q = "#starbucks", n = 1000, lang = "en")

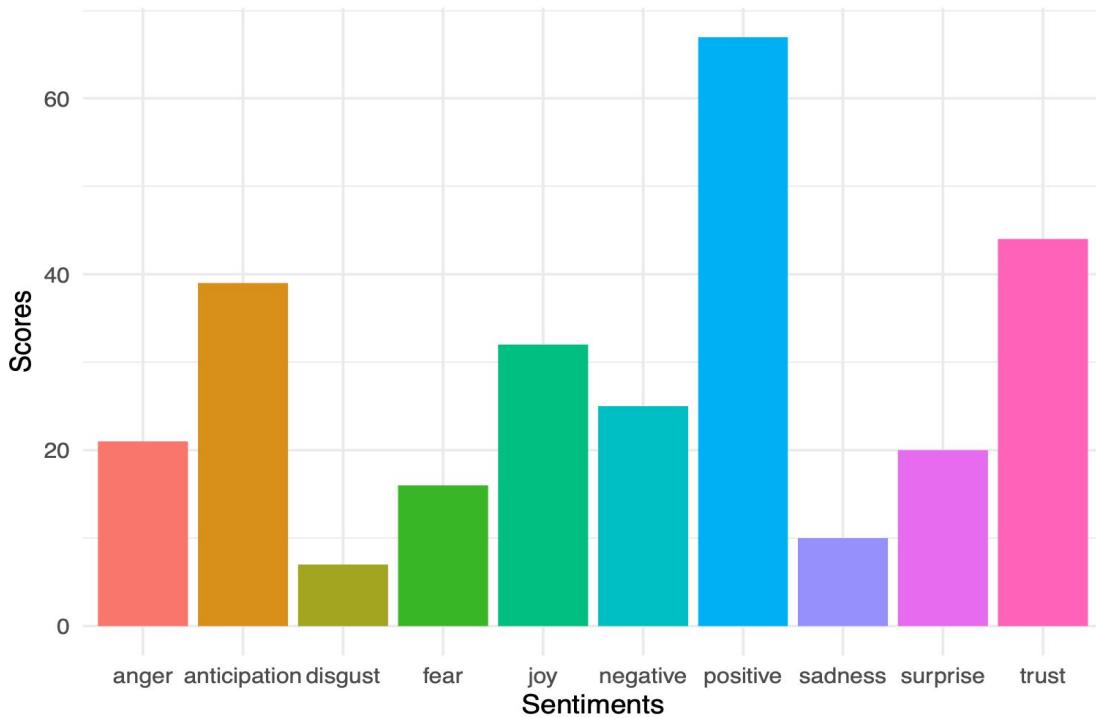
#sb$text

sb$text <- iconv(sb$text, from="UTF-8", to="ASCII", sub="")
sb$text<-gsub("(RT|via)((?:\\b\\w*@[\\w+]+)|", "", sb$text)

sb$text<-gsub("@\\w+", "", sb$text)
sb_sentiment<-get_nrc_sentiment((sb$text))
sentimentscores<-data.frame(colSums(sb_sentiment[,]))
names(sentimentscores) <- "Score"
sentimentscores <- cbind("sentiment"=rownames(sentimentscores),sentimentscores)
rownames(sentimentscores) <- NULL
ggplot(data=sentimentscores,aes(x=sentiment,y=Score))+  
  geom_bar(aes(fill=sentiment),stat = "identity")+
  theme(legend.position="none")+
  xlab("Sentiments")+ylab("Scores")+
  ggtitle("Total Sentiment of Tweets under #Starbucks")+
  theme_minimal()
```



Total Sentiment of Tweets under #SouthwestAirlines



- Southwest Airlines has gain popularity
 - Luggage
 - Open middle seat
 - Excellent customer service
- Reflected in this graph
 - Positive
 - Joy
 - Trust

Code for Southwest Airlines Sentiment Graph

```
#southwest
sw <- rtweet::search_tweets(q = "#southwestairlines", n = 100, lang = "en")

#sw$text

sw$text <- iconv(tweet_text, from="UTF-8", to="ASCII", sub="")
sw$text<-gsub("(RT|via)((?:\\b\\w*@[\\w+]+)|", "",sw$text)

sw$text<-gsub("@\\w+", "",sw$text)
sw_sentiment<-get_nrc_sentiment(sw$text)
sentimentscores<-data.frame(colSums(sw_sentiment[,]))
names(sentimentscores) <- "Score"
sentimentscores <- cbind("sentiment"=rownames(sentimentscores),sentimentscores)
rownames(sentimentscores) <- NULL
ggplot(data=sentimentscores,aes(x=sentiment,y=Score))+ 
  geom_bar(aes(fill=sentiment),stat = "identity")+
  theme(legend.position="none")+
  xlab("Sentiments")+ylab("Scores")+
  ggtitle("Total Sentiment of Tweets under #SouthwestAirlines")+
  theme_minimal()
```

Disclosure:

Twitter hashtags are an effective way to increase social media engagement. But, the combination of words, acronyms, context and timing of the hashtag should be closely monitored to ensure the success of the hashtag campaign.

Overlooking these factors could lead to a great deal of embarrassment and PR nightmares.

6 hours ago



capnmarrrrk Capn Marmik

Fingernail in my BigMac Once **#McDStories**, McDonald's Twitter Hashtag Promotion, Goes Horribly Wrong huff.to/y1cIBQ via **@cvbarnhart**

6 hours ago



johngarrettX John Garrett

So PETA and McDonalds got into it today on Twitter. I was surprised I didn't know there was actual meat at McDonalds. **#McDStories**

6 hours ago



PuppyPuncher Nick

Ordered a McDouble, something in the damn thing chipped my molar. **#McDStories**

54 minutes ago



Alice_2112 May

Hospitalized for food poisoning after eating McDonalds in 1989. Never ate there again and became a Vegetarian. Should have sued. **#McDStories**

57 minutes ago



JKKingArt James King

#McDStories I lost 50lbs in 6 months after I quit working and eating at Mcdonalds

2 hours ago



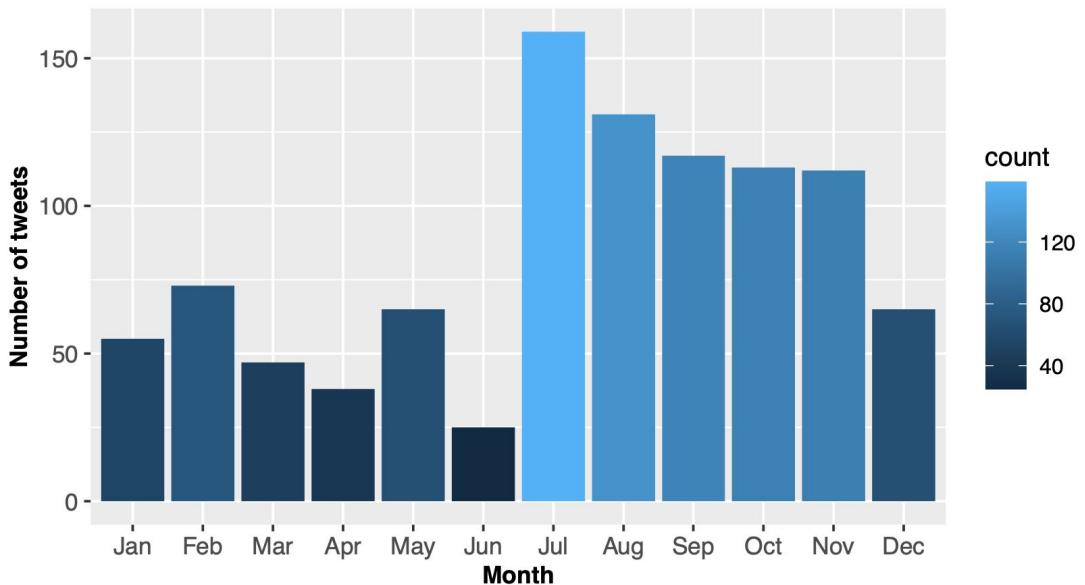
rolfarnold rolfarnold

#McDStories Learn all about McDonalds using pigs from gestation crates. **#McCruelty** chn.ge/r5Q4fA

1 hour ago



Tweets under @Nike



- Colin Kaepernick and Nike's 4th of July Sneaker Controversy
- New Nike AD - "We'll come together for change"
- Increased interaction on social media

"NIKE MADE THE DECISION TO HALT DISTRIBUTION OF THE AIR MAX 1 QUICK STRIKE FOURTH OF JULY BASED ON CONCERN THAT IT COULD UNINTENTIONALLY OFFEND."

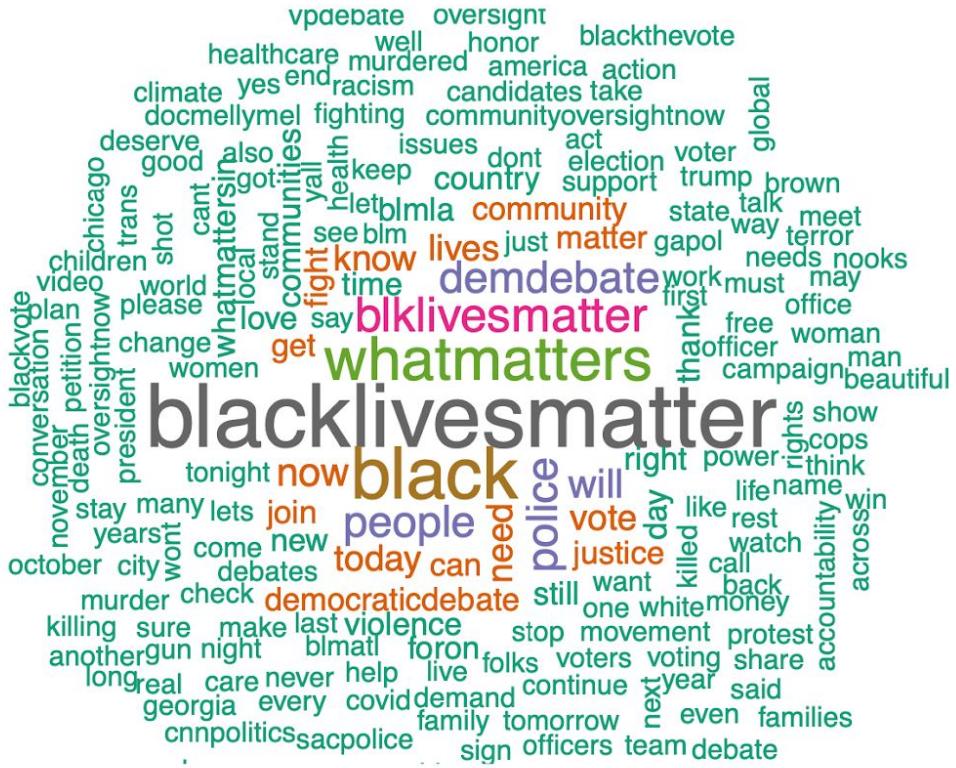
#GGPLOT
#nike

```
nike <- get_timeline("nike", n = 1000)
ggplot(data = nike, aes(x = month(created_at, label = TRUE))) +
  geom_bar(aes(fill = ..count..)) +
  xlab("Month") + ylab("Number of tweets") +
  ggtitle("Tweets under @Nike") +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=10,face="bold")) +
  theme(plot.title = element_text(size = 20, face = "bold"))
  theme(legend.text=element_text(size=10)) +
  scale_fill_gradient(low = "grey", high = "black")
```

MAP CLOUD

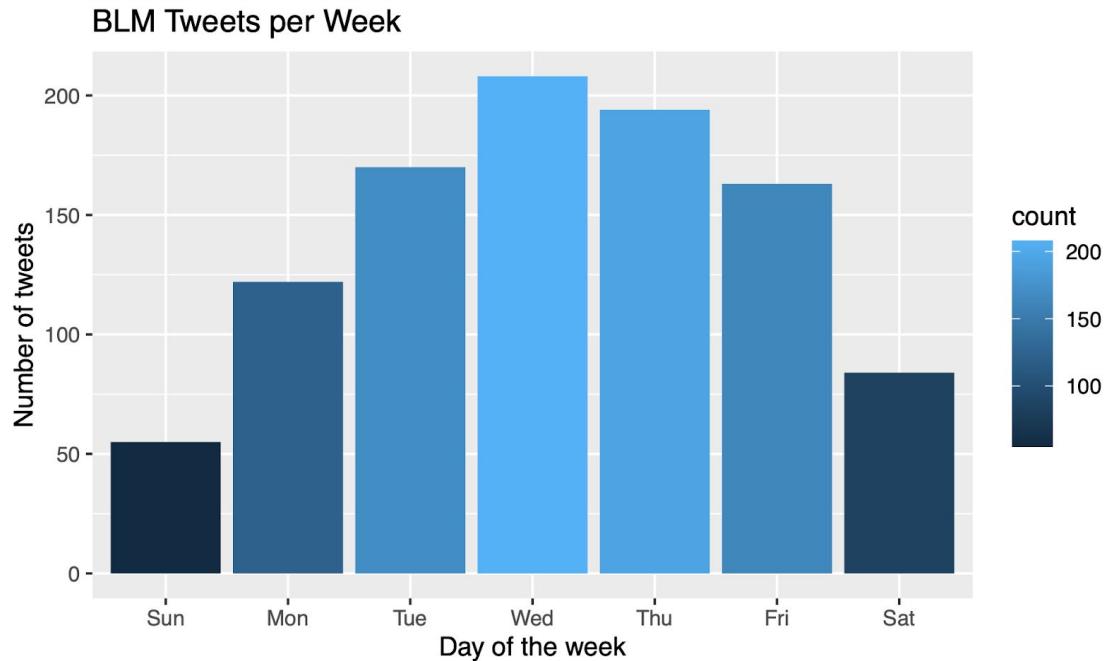
@BlkLivesMatter

```
205 #WORD CLOUD
206 tweet_text <- blm_tweets$text
207 #Removing numbers, punctuations, links and alphanumeric content
208 tweet_text<- gsub("[[:digit:]]+","", tweet_text)
209 tweet_text<- gsub("[[:punct:]]+","", tweet_text)
210 tweet_text<- gsub("http[[:alnum:]]*", "", tweet_text)
211 tweet_text<- gsub("([[:alpha:]]+)1+", "", tweet_text)
212
213 #creating a text corpus
214 docs <- Corpus(VectorSource(tweet_text))
215 # converting the encoding to UTF-8 to handle funny characters
216 docs <- tm_map(docs, function(x) iconv(enc2utf8(x), sub = "byte"))
217 # Converting the text to lower case
218 docs <- tm_map(docs, content_transformer(tolower))
219 # Removing english common stopwords
220 docs <- tm_map(docs, removeWords, stopwords("english"))
221 # Removing stopwords specified by us as a character vector
222 docs <- tm_map(docs, removeWords, c("amp"))
223 # creating term document matrix
224 tdm <- TermDocumentMatrix(docs)
225 # defining tdm as matrix
226 m <- as.matrix(tdm)
227 # getting word counts in decreasing order
228 word_freqs = sort(rowSums(m), decreasing=TRUE)
229 # creating a data frame with words and their frequencies
230 blm_wf <- data.frame(word.names(word_freqs), freq=word_freqs)
231 # plotting wordcloud
232 set.seed(1234)
233 wordcloud(words = blm_wf$word, freq = blm_wf$freq,
234           min.freq = 1, scale=c(1.8,.5),
235           max.words=200, random.order=FALSE, rot.per=0.15,
236           colors=brewer.pal(8, "Dark2"))
```





```
ggplot(data = blm_tweets, aes(x = wday(created_at, label = TRUE))) +  
  geom_bar(aes(fill = ..count..)) +  
  xlab("Day of the week") + ylab("Number of tweets") +  
  ggtitle("BLM Tweets per Week")  
  theme_minimal() +  
  scale_fill_gradient(low = "turquoise3", high = "darkgreen")
```



Thank you!

<https://github.com/selamtesfa>