Case_Study.R

rahul

2020-09-17

```r
setwd("C:\\Users\\rahul\\OneDrive\\Desktop\\WPI\\03. Fall 2020")

tweets <- read.csv("my_data.csv")



## Import Libraries

library(dplyr)
library(tidytext)
library(stringr)
library(ggplot2)

# Remove http elements manually
tweets$stripped_text <- gsub("http.*","",  tweets$content)
tweets$stripped_text <- gsub("https.*","", tweets$content)

## Tokenize

tweets_clean <- tweets %>%
  dplyr::select(stripped_text) %>%
  unnest_tokens(word, stripped_text)


## Word counts without removing stop words



tweets_clean %>%
  count(word, sort = TRUE) %>%
  top_n(30) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(x = word, y = n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip() +
  labs(x = "Count",
       y = "Unique words",
       title = "Count of unique words found in tweets")
```
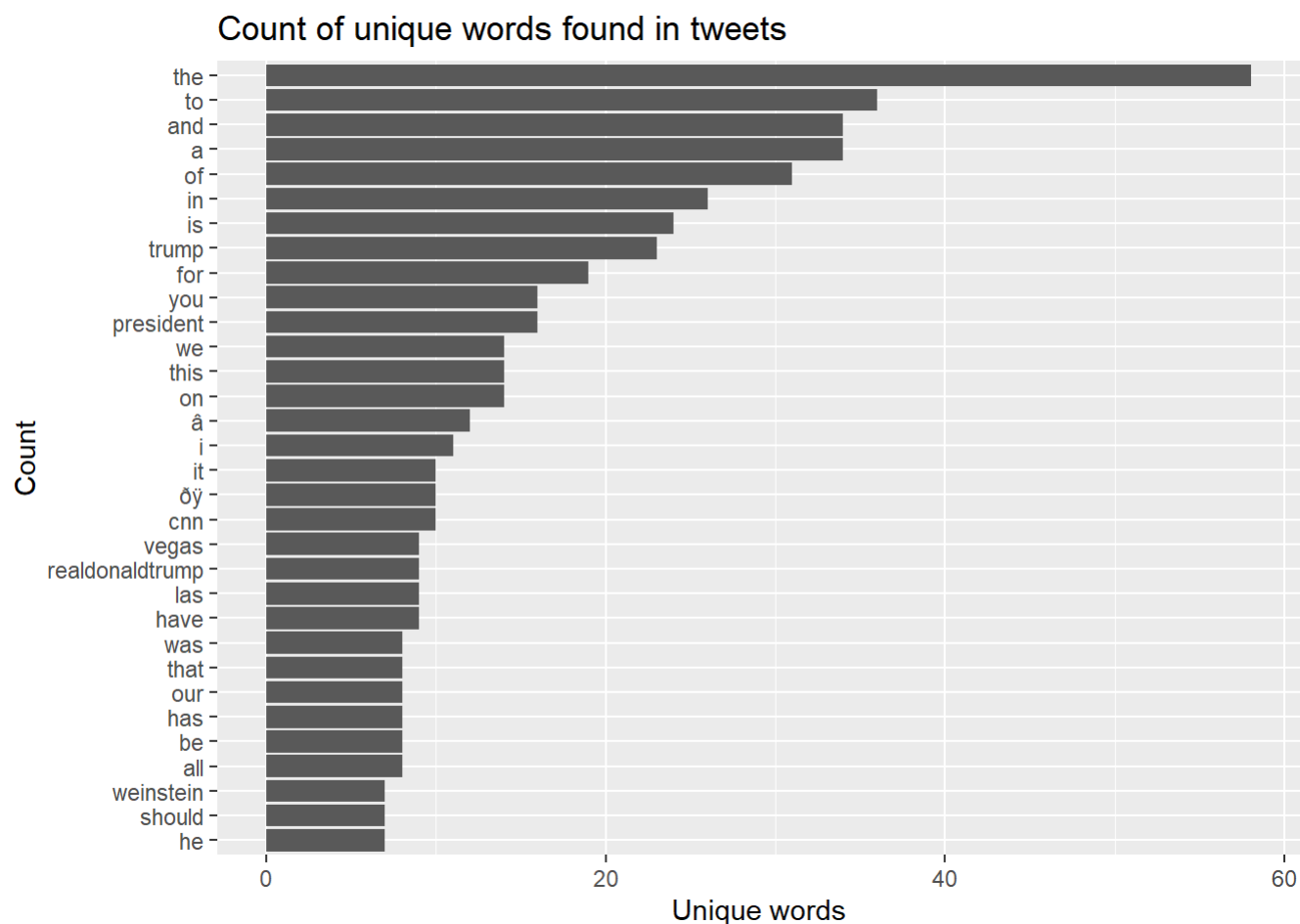
```
## Selecting by n
```

Count of unique words found in tweets


```
## Remove stop words

cleaned_tweet_words <- tweets_clean %>%
  anti_join(stop_words)
```

```
## Joining, by = "word"
```

```
cleaned_tweet_words %>%
  count(word, sort = TRUE) %>%
  top_n(30) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(x = word, y = n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip() +
  labs(x = "Count",
       y = "Unique words",
       title = "Count of unique words found in tweets")
```
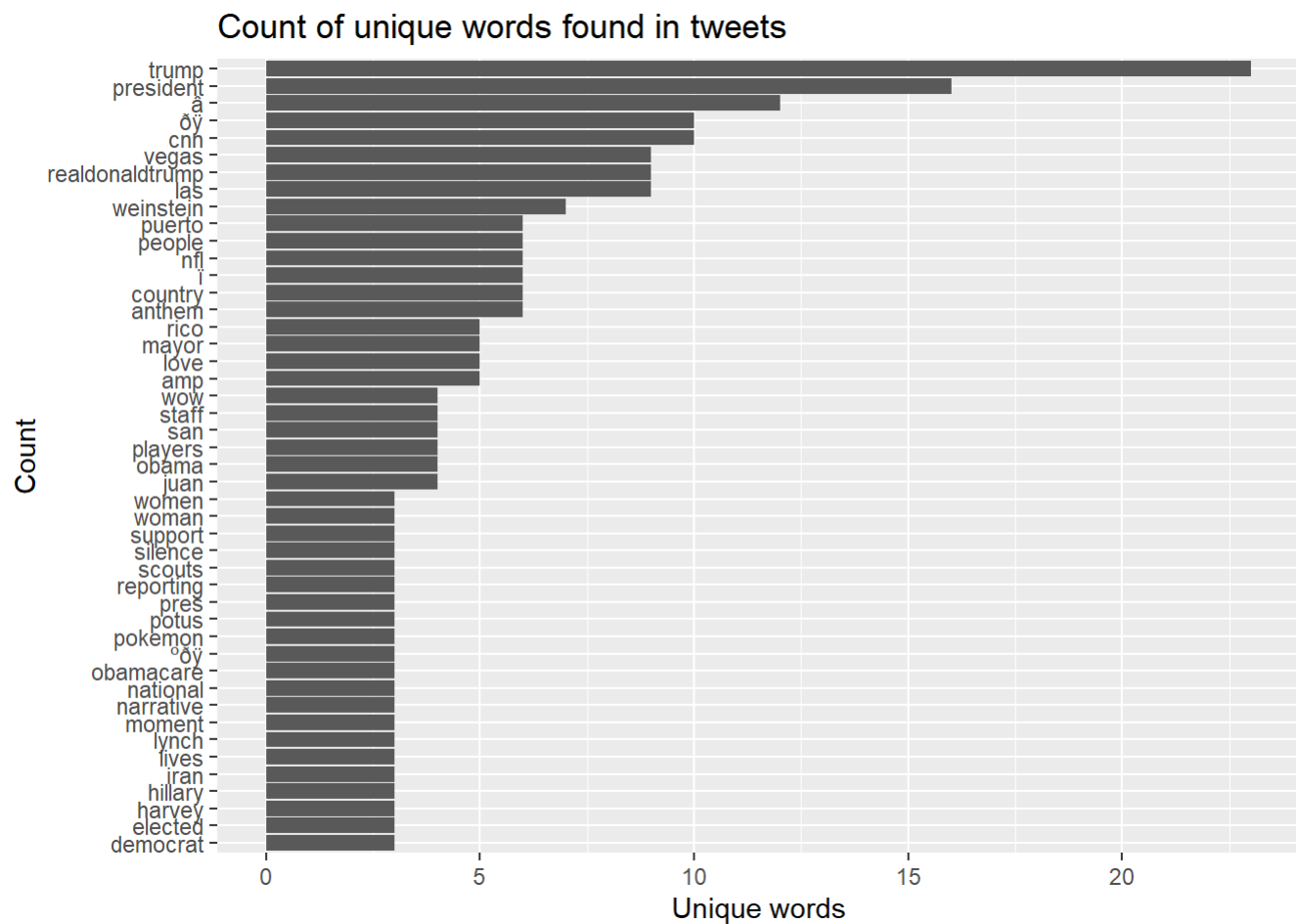
```
## Selecting by n
```

Count of unique words found in tweets


```
output <- cleaned_tweet_words %>%
  count(word, sort = TRUE) %>%
  top_n(30)
```

```
## Selecting by n
```

```
output[1:30,] %>% as.data.frame()
```

```
##                word   n
## 1             trump  23
## 2         president  16
## 3                 â  12
## 4               cnn  10
## 5                ðÿ  10
## 6               las   9
## 7   realdonaldtrump   9
## 8             vegas   9
## 9         weinstein   7
## 10           anthem   6
## 11          country   6
## 12                ï   6
## 13              nfl   6
## 14           people   6
## 15           puerto   6
## 16              amp   5
## 17             love   5
## 18            mayor   5
## 19             rico   5
## 20             juan   4
## 21            obama   4
## 22          players   4
## 23              san   4
## 24            staff   4
## 25              wow   4
## 26         democrat   3
## 27          elected   3
## 28           harvey   3
## 29          hillary   3
## 30             iran   3
```

```
## Ordering tweets by retweet

ordered_tweets <- tweets[order(-tweets$retweet),]

ordered_tweets[1:30,c("content","retweet")] %>% as.data.frame()
```

##
content
3   Daughter of fallen Navy Sailor delivers powerful m
onologue on anthem protests, burns her NFL packers gear. #BoycottNFL https://t.co/qDlFBGMeag
5
19,000 RESPECTING our National Anthem! #StandForOurAnthemðŸ‡ºðŸ‡¸ https://t.co/czutyGaMQV
7
ðŸ\220\235ðŸ\220\235ðŸ\220\235 https://t.co/MorL3AQW0z
18   BREAKING: The audio of sexual predator Harvey Weinstein admitting to
assaulting model Ambra Gutierrez has leaked.   I am truly disgusted. https://t.co/ka1nhd0ReR
21   Because he hates women s
o much, President Trump just nominated another woman, Kirstjen Nielsen, to be in his Cabinet.
24   I am an Eagle Scout   Only boys
should be in Boy Scouts   Only girls should be in Girl Scouts   Don't change things that work
26
Do you agree with the Boy Scouts for allowing girls to join?
29   Network news has become so partisan, distorte
d and fake that licenses must be challenged and, if appropriate, revoked. Not fair to public!
31   To be clear, when we say "Pok
Ã©mon Go was used" we mean "there is no evidence Pokemon Go was used" https://t.co/R5uvyhZIE5
34   So wait, did t
he Russians also purchase Hillary Clinton? "Pokemon Go to the polls"! https://t.co/2I1ayPLMBu
35   Students describe what itâ\200\231s like being a conservative o
n a college campus. The described behavior is inexplicable and inexcusable. https://t.co/OBEZKBT
amL
36
Eminem is a Trump supporter?!   ðŸ\230® https://t.co/JnCUJ7sqnU
41
Is this real? https://t.co/QUTSZc3quW
43
oh https://t.co/XU3MBQ5f2K
44   President Trump: #Iran was Economically BANKRUPT until Obama gifted them 1.
7 BILLION in CASH & infused 150 BILLION into their economy https://t.co/GocTNG4KyL
46   All of the good guys I know wanted us out
of Iran deal and all of the globalists wanted us in it. Look at whose mad today - deep state.
47   NFL having to write out how players conduct themselves during the anthe
m is HILARIOUS. I learned how to honor the flag in GRADE SCHOOL #MAGA https://t.co/pe7ijVpslF
48   Dear CNN: When will you be
interviewing one of the mayors in Puerto Rico's other 77 municipios? https://t.co/MKNFP3IHZe
50   .@Joy_Villa: The Hollywood swamp is starting to crumble. If Weinstein i
s getting pulled to the surface, all these other smaller players will https://t.co/9j2wgxnMwH
61   Health Insurance stocks, which have gone thro
ugh the roof during the ObamaCare years, plunged yesterday after I ended their Dems windfall!
62   #FakeNews MSM has been out to get President Trump s
ince he was a candidate. Be warned--We The People are fighting back & you're going down.
65   Liberal's are the most hate filled, racist, hypocritical, bigoted, irration
al & mentally unstable creatures on earth. A danger to the world. https://t.co/qpzuVTlhAL
66   Photo of NFL player
s today protesting domestic violence, sexual assault, and child abuse https://t.co/jkLONjGZJu
69
Ignorant, irrelevant and exploitative https://t.co/6XWGQxcDPt
72   Thousands of Americans are
waiting in line to donate blood in Las Vegas. #LasVegasStrongðŸ‡ºðŸ‡¸ https://t.co/aRNromHfb5

```
## 77                                       Will players â\200œtake a kneeâ\200\235 tonight
to protest the lives saved by Las Vegas police officers? Will ESPN spit on first responders yet
again?
## 79                    Crybaby Marshawn Lynch of the Oakland Raiders sitting during the ant
hem, hiding behind staff. What a joke the NFL has become. #BoycottNFL https://t.co/tuhU59v9My
## 80 The truth about Puerto Rico ðŸ‡µðŸ‡· &amp; it's not the propaganda Mayor of #SanJuan has b
een spewing w/MSM &amp; cluless celebritiesâ\231¦ï¸\217RTâ\231¦ï¸\217  #TakeAKnee https://t.co/g
RYqsM9g8M
## 84                              '@CNN in hysterics as Puerto Rico FULLY EMBRACES Presiden
t Trump w/ Love and Cheers of Praise !!  Fake News Narrative CRUSHED https://t.co/Mn6Jt7IO9B'
## 85                             A couple hours before the Las Vegas Mass Shooting -
the crowd proudly sang "God Bless America" ðŸ‡ºðŸ‡¸ beautiful video.  https://t.co/awQ1Uu17qH
##     retweet
## 3         1
## 5         1
## 7         1
## 18        1
## 21        1
## 24        1
## 26        1
## 29        1
## 31        1
## 34        1
## 35        1
## 36        1
## 41        1
## 43        1
## 44        1
## 46        1
## 47        1
## 48        1
## 50        1
## 61        1
## 62        1
## 65        1
## 66        1
## 69        1
## 72        1
## 77        1
## 79        1
## 80        1
## 84        1
## 85        1
```

```
## Hashtags and mentions

hashtag <- function(x){
  a <- paste(str_extract_all( x,"#\\S+")[[1]], collapse = ", ")
  return(a)
}

mention <- function(x){
  a <- paste(str_extract_all( x,"@\\S+")[[1]], collapse = ", ")
  return(a)
}


tweets$hashtags <- "NA"
tweets$mentions <- "NA"

for (i in 1:100) {
  if (length(hashtag(tweets$content[i])) >0) {
    tweets$hashtags[i] <- hashtag(tweets$content[i])
  }

  if (length(mention(tweets$content[i])) >0) {
    tweets$mentions[i] <- mention(tweets$content[i])
  }

}




# View top 10 hashtags

all_hashtags <- paste(tweets$hashtags, collapse = ", ")

hashs <- str_extract_all( all_hashtags,"#\\S+")[[1]] %>% table %>% sort(decreasing = T)

hashs[1:10] %>% as.data.frame()
```

```
##                      . Freq
## 1          #BoycottNFL,    2
## 2            #FakeNews,    2
## 3      #LasVegasShooting,    2
## 4          #VegasStrong,    2
## 5         #HipHopAwards,    1
## 6                #Iran,    1
## 7            #LasVegas,    1
## 8           #LasVegas.,    1
## 9  #LasVegasStrongðŸ‡ºðŸ‡,,    1
## 10               #MAGA,    1
```

```r
# view top 10 mentions

all_mentions <- paste(tweets$mentions, collapse = ", ")

ments <- str_extract_all( all_mentions,"@\\S+")[[1]] %>% table %>% sort(decreasing = T)

ments[1:10] %>% as.data.frame()
```

```
##                     . Freq
## 1  @realDonaldTrump,     9
## 2             @CNN,      3
## 3   @MichelleObama,,     2
## 4            @POTUS,     2
## 5       @PrisonPlanet,   2
## 6       @AP_Politics,    1
## 7        @Breaking911,   1
## 8  @CarmenYulinCruz,    1
## 9          @Cernovich,   1
## 10          @FLOTUS!,    1
```