title: “Russian Tweets - October 2018 Data”
output:
html_document: default
Load libraries

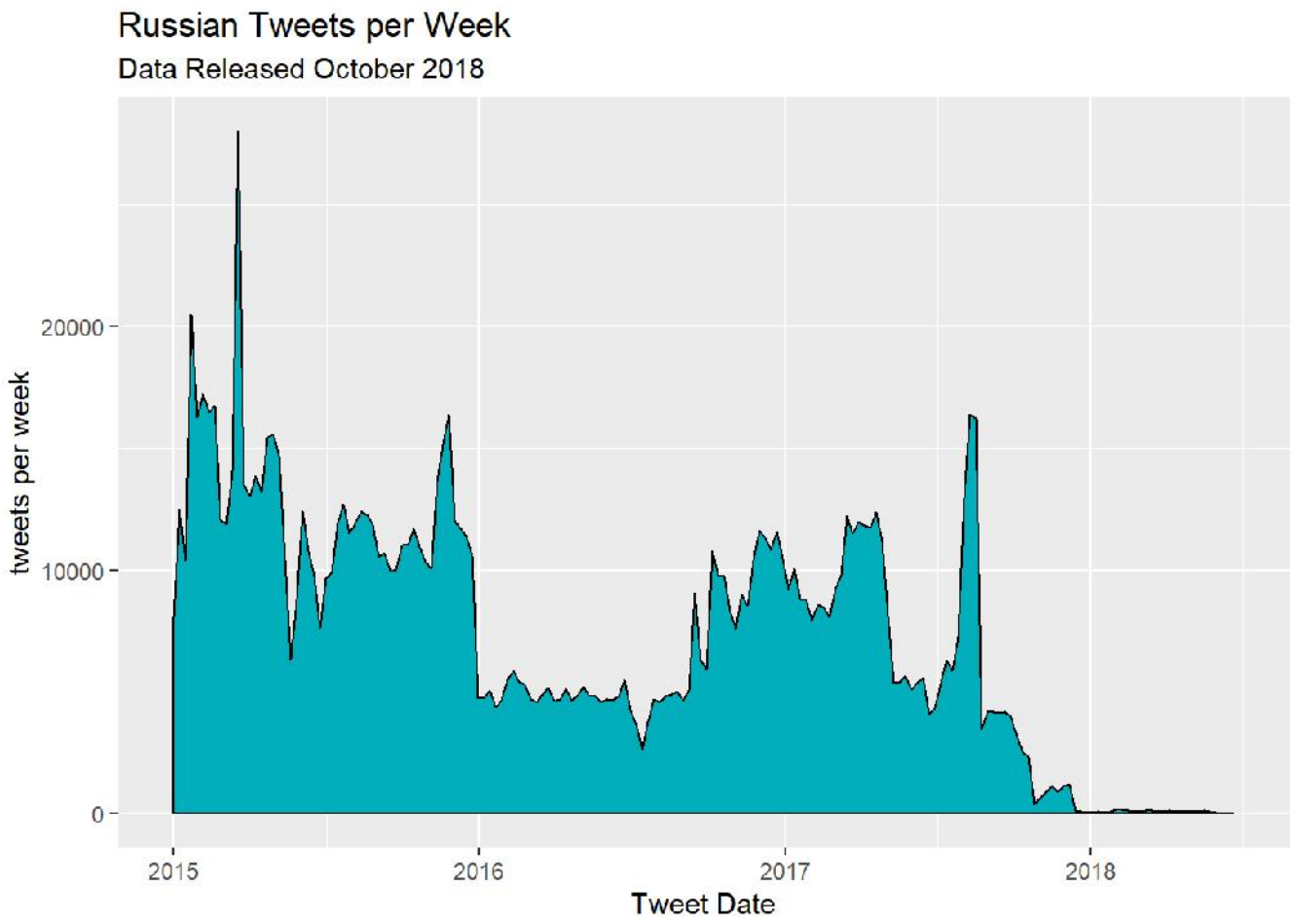
```
library(rmarkdown)
library(knitr)
library(dplyr)
library(tidyverse)
library(lubridate)
library(stringr)
library(tidytext)
library(ggthemes)#Themes for formating
library(grid) #Add grid line
library(wordcloud2)
```

Read original tweet archive from October 2018

```
IRA_users <- read.csv("ira_users_csv_hashed.csv", encoding = "UTF-8")
IRA_tweets <- read_csv("ira_tweets_csv_hashed.csv")
```

```
# trend over time
IRA_tweets <- IRA_tweets %>%
  mutate(tweet_date=date(tweet_time)) %>%
  filter(tweet_date > ymd("2014-12-31"))

ggplot(data = IRA_tweets, mapping = aes(tweet_date)) + geom_area(stat = "bin",  fill = "#00AFBB"
, color = "black", binwidth=7) +
  labs(title="Russian Tweets per Week", subtitle = "Data Released October 2018", x = "Tweet Dat
e", y= "tweets per week")
```




```
# likes and retweets
IRA_tweets %>%
  group_by(is_retweet) %>%
  summarize(n=n()) %>%
  mutate(prop = n/sum(n))
```

```
## # A tibble: 2 x 3
##   is_retweet      n  prop
##   <chr>       <int> <dbl>
## 1 false      744108 0.567
## 2 true       568395 0.433
```

```
# most common hashtags
hashtag_summary0 <- IRA_tweets %>%
  select(hashtags, userid, user_display_name, account_language) %>%
  filter(hashtags != "[]", account_language == "en")  %>% # get rid of tweets with no hashtags
  mutate(hashtags = str_sub(hashtags,2,str_length(hashtags)-1)) %>% # remove first and last char
acters (brackets)
  mutate(hashtags = str_split(hashtags,",")) # separate multiple hashtags into a list

hashtag_list0 <- unnest(hashtag_summary0, hashtags) # restructure so that there is 1 hashtag per
 record

hashtag_freq0 <- hashtag_list0 %>%
  group_by(hashtags) %>%
  summarize(n=n()) %>%
  filter(n >= 200) %>%
  rename(word = hashtags, freq = n)

# word clouds

wordcloud2(data=hashtag_freq0)
```

eliminate most common words

```
hashtag_freq1 <- hashtag_freq0 %>%
  filter(freq < 4000)

wordcloud2(data=hashtag_freq1)
```