Twitter Sentiment towards three South African politicians

Tyryn

13/03/2022

In this short report I explore Twitter sentiment towards three prominent South African politicians: Cyril Ramaphosa, Julius Malema, and Helen Zille. Cyril Ramaphosa is the current President of South Africa and the African National Congress (ANC), a political party that has been in power since 1994. Julius Malema is the leader of the Economic Freedom Fighters (EFF), a political party formed in 2013 that has quickly grown to be the country's third largest. Helen Zille was the leader of the second largest party, the Democratic Alliance (DA), until 2015 but has since remained as the DA's arguably most outspoken member.

Opinion is strongly divided on all three politicians. Ramaphosa leads a party that has failed to deliver on many of its promises and has endured an endless number of corruption scandals. That said, the ANC remains the country's most popular at the ballot with it receiving 57.5% of the vote in the last election. Malema's party, the EFF, carries a far-left populist message that has resonated with many disaffected young black South Africans but simultaneously, it has brought the ire of those who've deemed his messaging as 'anti-white' racism. Lastly, Zille's DA has drawn criticism for primarily serving the interests of privileged white South Africans but moreover, Zille has drawn much backlash against some of the views she holds. All three politicians use Twitter as a tool to spread their political messages.

Collecting tweets with R

For this article I used the rtweet package to collect tweets that mentioned any of the three politicians in the first week of March 2022. The package allows you to scrape tweets up to 6 days before (unless you pay for a premium Twitter account).

After some processing I managed to scrape:

- 28405 tweets that mention Ramaphosa
- 24462 tweets that mention Malema
- 798 tweets that mention Zille

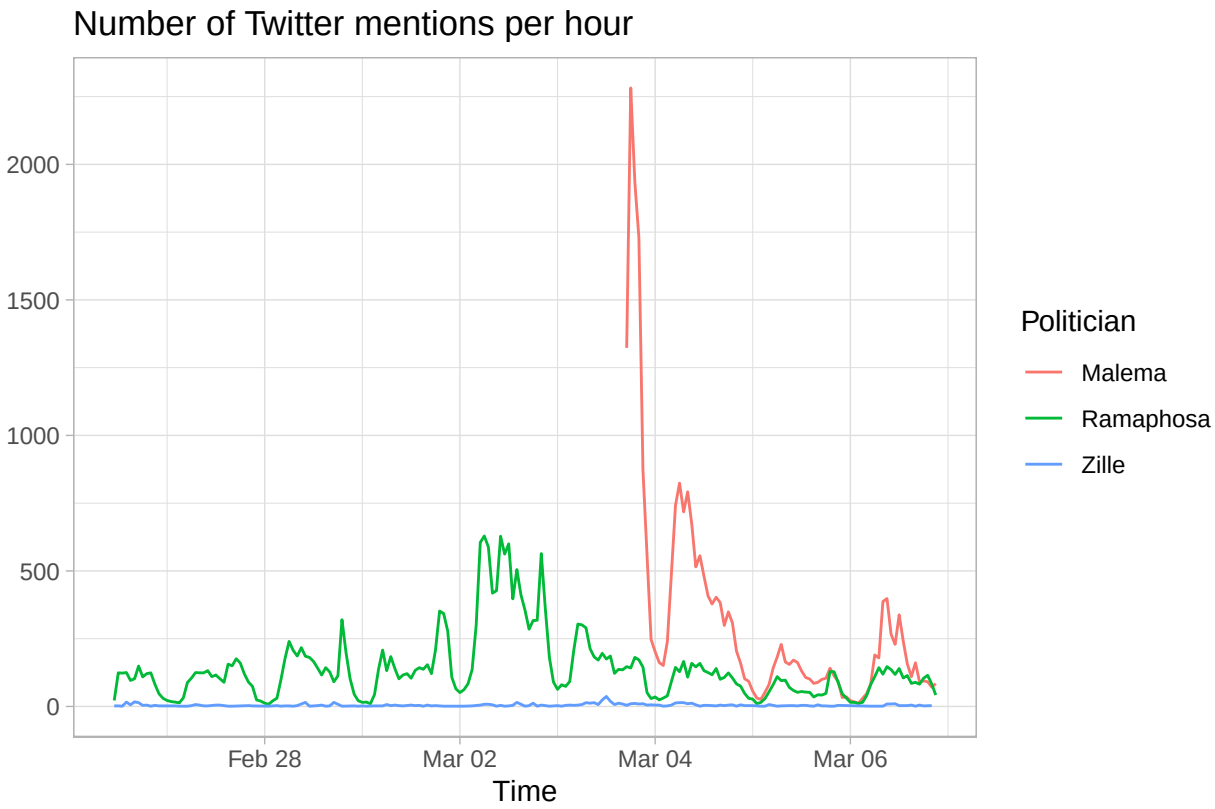
The code chunk below describes the process I used to gather the data:

```
library(tidyverse)
library(rtweet)

# Function to pre-process text
clean.text = function(x) {
  # convert to lower case
  x = tolower(x)
  # remove rt
  x = gsub("rt", "", x)
  # remove at
  x = gsub("@\\w+", "", x)
```

```
  # remove punctuation
  x = gsub("[[:punct:]]", "", x)
  # remove numbers
  x = gsub("[[:digit:]]", "", x)
  # remove links http
  x = gsub("http\\w+", "", x)
  # remove tabs
  x = gsub("[ |\t]{2,}", "", x)
  # remove blank spaces at the beginning
  x = gsub("^ ", "", x)
  # remove blank spaces at the end
  x = gsub(" $", "", x)
  # some other cleaning text
  x = gsub('https://','',x)
  x = gsub('http://','',x)
  x = gsub('[^[:graph:]]', ' ',x)
  x = gsub('[[:punct:]]', '', x)
  x = gsub('[[:cntrl:]]', '', x)
  x = gsub('\\d+', '', x)
  x = str_replace_all(x,"[^[:graph:]]", " ")
  return(x)
}

# The names of the politicians
politicians <- c("Zille", "Ramaphosa", "Malema")

# Scrape the data (takes a while)
tweet_list <- lapply(politicians, function(x) search_tweets(x,
  retryonratelimit = TRUE, n=18000
))

# Clean up the text using a user written cleaning function (clean.text)
tweet_list_preprocess <- lapply(tweet_list, function(x){
  df <- x %>%
    mutate(text = clean.text(text))
  return(df)
})

# Name the list elements
names(tweet_list_preprocess) <- politicians

# Number of tweets per hour (for graphing)
tweet_list_time <- lapply(tweet_list_preprocess, function(x){
  df <- x %>%
    mutate(Created_at_round = round.POSIXt(created_at, units = "hours"))
  return(df)
})

# Below to create single dataframe
tweet_df <- Map(cbind, tweet_list_time, Politician = names(tweet_list_time)) %>%
  bind_rows(.)
```

Unpacking the data

Between the three politicians there are notable differences in how they come across on Twitter. First, as shown in the graph below, there were substantially more tweets per hour mentioning Malema than the other two politicians. In fact, Zille got very few mentions per hour. On average, Malema received list(mean_tweets = 318) mentions per hour, Ramaphosa list(mean_tweets = 140) mentions per hour and Zille just list(mean_tweets = 5) mentions per hour.



Malema's mentions reached the 18 000 tweet limit on 3 March.

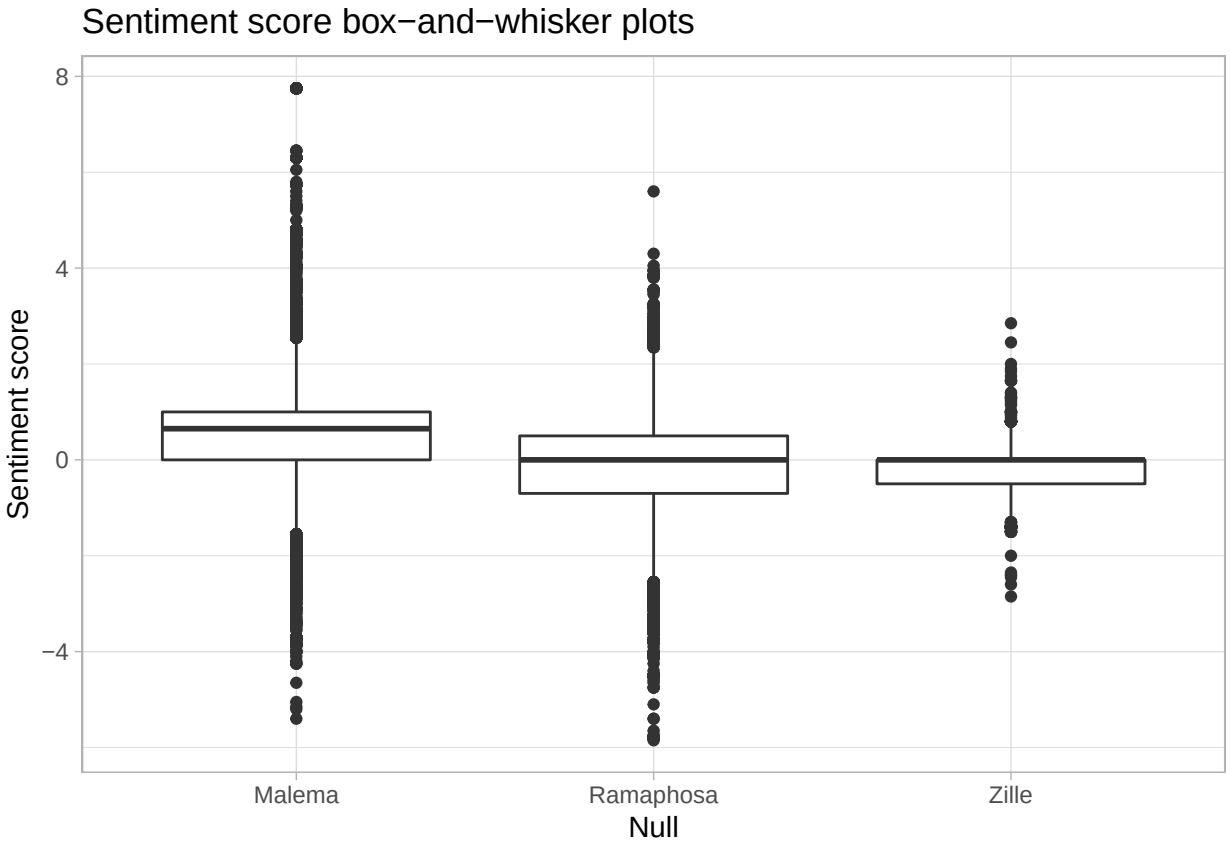
While its now clear that Malema is the most mentioned on Twitter, we haven't yet got an impression of the contents of tweets mentioning the three politicians. As a starting point I've produced a word cloud for each politician. Word clouds provide a visual representation of free-form text data by showing the most commonly mentioned terms.

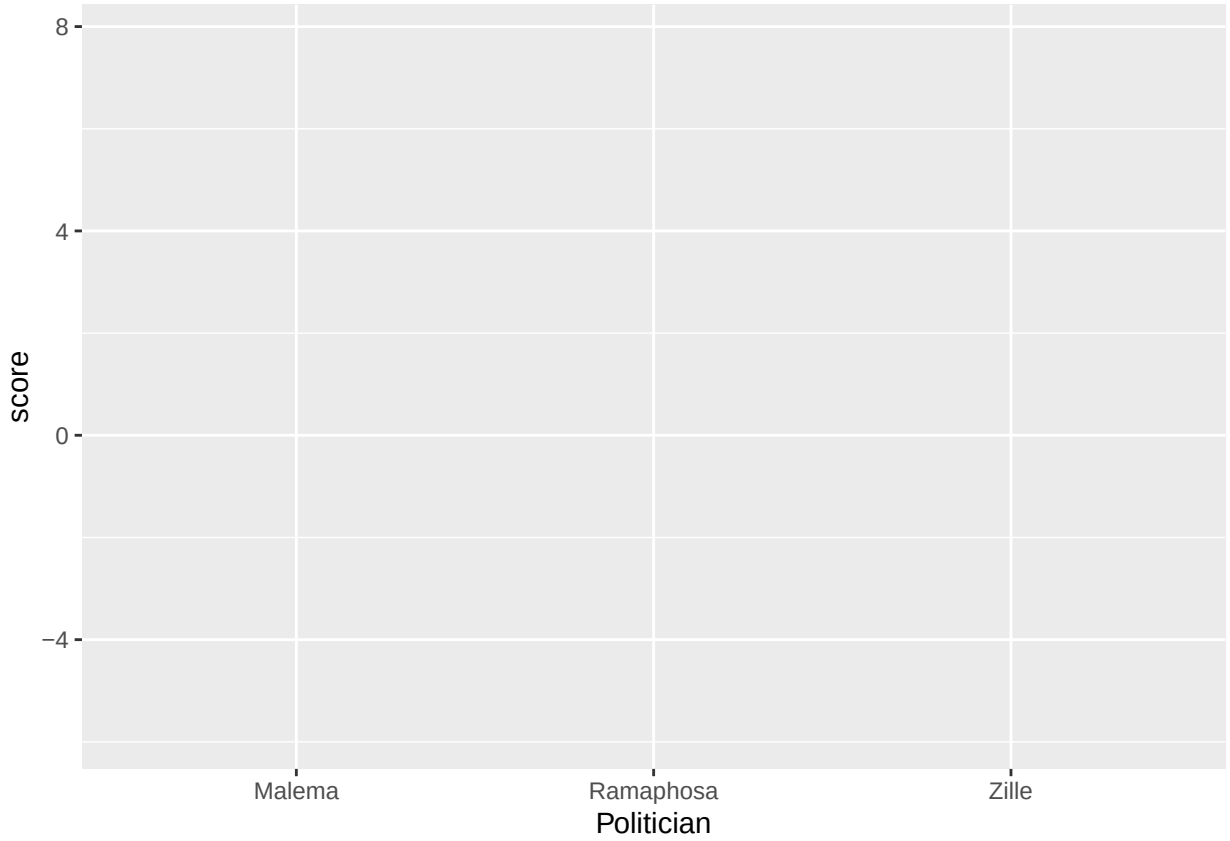
Ramaphosa's tweets seemed to most commonly relate to the Ukraine-Russia war, for Malema they most commonly related to his birthday (which was 3 March and also probably explains his large number of mentions), while for Zille tweets often seemed to relate to her controversial opinions on 'PC culture'.

Tweet sentiment

As a final step I use the syuzhet package to generate sentiment scores from the tweets. The syuzhet package offers a few different algorithms, each taking a different approach to sentiment scoring. It also does emotion scoring based upon the nrc algorithm. The code chunk below shows how the sentiment scores are generated.

The box-and-whisker plots below show that tweets mentioning Malema generally have the most positive sentiment whereas those that mention Zille generally have the most negative sentiment.


Sentiment score box−and−whisker plots
8
4
0
−4
Sentiment score
Malema
Ramaphosa
Zille
Null


8
4
0
−4
score
Malema
Ramaphosa
Zille
Politician