Politics

2022-10-04

Packages and data to load

```
library(rtweet)
library(dplyr)
library(lubridate)
library(stringr)
library(readr)
library(tidytext)
library(rvest)
library(httr)
data("stop_words")
```

Twitter

Accessing tweets from all the political parties holding seats in the UK House of Commons.

These are:

- Conservative Party
- Labour Party
- Scottish National Party (Scotland)
- Liberal Democrats
- Democratic Union Party (Northern Ireland)
- Plaid Cymru (Wales)
- Social Democratic Party (Northern Ireland)
- Alba Party (Scotland)
- Alliance Party (Northern Ireland)
- Green Party
- Sinn Féin (Northern Ireland, non-sitting, Twitter profile includes Republic of Ireland too)

Some tweets from Plaid Cymru will be in Welsh. Also a possibility of Scots or Irish.

First, accessing the data from the API. This will not be run whilst knitting the document as the data has now been saved in a CSV file.

```
# Read in tweets once, and then save as a file for speed

con_tweets <- get_timeline("Conservatives", n = 3200)
lab_tweets <- get_timeline("UKLabour", n = 3200)
snp_tweets <- get_timeline("theSNP", n = 3200)
libdem_tweets <- get_timeline("LibDems", n = 3200)
dup_tweets <- get_timeline("duponline", n = 3200)
plaid_tweets <- get_timeline("Plaid_Cymru", n = 3200)
sdp_tweets <- get_timeline("SDPhq", n = 3200)
alba_tweets <- get_timeline("AlbaParty", n = 3200)
alliance_tweets <- get_timeline("allianceparty", n = 3200)
green_tweets <- get_timeline("TheGreenParty", n = 3200)
```

```
sf_tweets <- get_timeline("sinnfeinireland", n = 3200)

# Creating a single data frame for storage
data_list <- list('con_tweets', 'lab_tweets', 'snp_tweets', 'libdem_tweets', 'dup_tweets', 'plaid_tweets

hoc_working <- tibble(created_at = POSIXct(),
                      party = character(),
                      text = character(),
                      favorite_count = integer(),
                      retweet_count = integer())

for (x in data_list) {
  add_tweets <- get(x) %>%
    filter(retweeted == "FALSE") %>%
    mutate(party = str_remove(x, '_tweets')) %>%
    select(created_at, party, text, favorite_count, retweet_count)

  hoc_working <-
    hoc_working %>%
    bind_rows(add_tweets)
}

hoc_tweets <-
  hoc_working %>%
  filter(str_starts(text, 'RT', negate = T)) %>% # Not sure why retweet flag didn't work
  # Handling tweets in Welsh or Gaelic
  mutate(lang = cld2::detect_language(text)) %>%
  filter(lang == 'en') %>%
  select(-lang) %>%
  mutate(party = case_when(party=='con' ~ 'Conservatives',
                           party=='lab' ~ 'Labour',
                           party=='libdem' ~ 'Liberal Democrats',
                           party=='snp' ~ 'Scottish National Party',
                           party=='dup' ~ 'Democratic Unionist Party',
                           party=='plaid' ~ 'Plaid Cymru',
                           party=='sdp' ~ 'Social Democratic Party',
                           party=='alba' ~ 'Alba',
                           party=='alliance' ~ 'Alliance',
                           party=='green' ~ 'Green',
                           party=='sf' ~ 'Sinn Féin'))

# Save data to a CSV
readr::write_csv(hoc_tweets, 'hoc_tweets_221003.csv')
```

The main parties all returned 3200 tweets from the request. Some of the smaller parties had fewer tweets in the initial sample eg. Alba Party as was only founded in 2021. All retweets have then been removed.

Then processing the data to look at the most frequent words by party. This uses the data that has already been saved in a CSV.

```
hoc_tweets <- readr::read_csv('hoc_tweets_221003.csv')

clean_hoc_tweets <-
  hoc_tweets %>%
  mutate(
```

```
    # identify tweets with hashtags
    has_tag = str_detect(text, "#\\w+"),
    # how many at-mentions are there?
    n_at = str_count(text, "(^|\\s)@\\w+"),
    # extract first url
    url = str_extract(text, "(https?://\\S+)"),
    # remove at-mentions, tags, and urls
    clean_text =
      str_replace_all(text, "(^|\\s)(@|#|https?://)\\S+", " ") %>%
      str_replace_all("\\W+", " "))

tokens <-
  clean_hoc_tweets %>%
  unnest_tokens(word, clean_text) %>%
  select(party, word) %>%
  anti_join(stop_words) %>%
  count(word, party, sort = TRUE)

top_tokens <-
  tokens %>%
  group_by(party) %>%
  slice_max(order_by=n, n=10) %>% # Keep top 10 tokens by party
  ungroup()

tweets_by_party <-
  clean_hoc_tweets %>%
  count(party)

tweets_by_party
```

```
## # A tibble: 11 x 2
##    party                          n
##    <chr>                      <int>
##  1 Alba                        1080
##  2 Alliance                    1335
##  3 Conservatives                420
##  4 Democratic Unionist Party   1140
##  5 Green                       1138
##  6 Labour                      1315
##  7 Liberal Democrats           1982
##  8 Plaid Cymru                  787
##  9 Scottish National Party     1219
## 10 Sinn Féin                   2379
## 11 Social Democratic Party     1403
```

```
top_tokens %>% filter(party=='Conservatives') # Most common tokens used by the Conservatives
```

```
## # A tibble: 12 x 3
##    word      party             n
##    <chr>     <chr>         <int>
##  1 uk        Conservatives    77
##  2 people    Conservatives    54
##  3 jobs      Conservatives    43
##  4 plan      Conservatives    41
```

```
##  5 labour    Conservatives    40
##  6 000       Conservatives    29
##  7 support   Conservatives    26
##  8 ve        Conservatives    24
##  9 day       Conservatives    20
## 10 economy   Conservatives    20
## 11 secretary Conservatives    20
## 12 trade     Conservatives    20
```

```
top_tokens %>% filter(party=='Labour')        # Most common tokens used by Labour
```

```
## # A tibble: 10 x 3
##    word          party      n
##    <chr>         <chr>  <int>
##  1 labour        Labour   755
##  2 people        Labour   341
##  3 government    Labour   290
##  4 britain       Labour   250
##  5 energy        Labour   232
##  6 conservatives Labour   216
##  7 plan          Labour   214
##  8 bills         Labour   212
##  9 crisis        Labour   198
## 10 tories        Labour   196
```

```
top_tokens %>% count(word) %>% filter(n>2)
```

```
## # A tibble: 7 x 2
##   word           n
##   <chr>      <int>
## 1 crisis         3
## 2 energy         3
## 3 government     8
## 4 minister       3
## 5 party          3
## 6 people        10
## 7 uk             3
```

The Conservatives twitter account produces far fewer original tweets than any other party. Of the 3200 tweets downloaded, only 420 were not retweets. Sinn Féin produce significantly more original tweets than any other party. This could potentially be related to the frequency with which their individual politicians use Twitter. For the Conservatives, some of their original Tweets may be published using official Government Twitter accounts and they then retweet those tweets.

The appearence of '000' in the most popular words for the Conservatives is due to their use of large numbers written in full eg. 100,000.

Many of the parties have the Conservatives ('conservatives', 'tories', 'tory', and sometimes 'government') referenced in their 10 most frequent words. The Conservatives are an exception to this, but do include mention of Labour ('labour'). 'people' was in the most popular words for 10 of the 11 parties (only not appearing for the DUP).