

# Sentiment Analysis with R

## Case Study: Twitter Data

Babatunde Adebayo

12/20/2019

## Goals of this session

By the end of this session you will:

- Understand the basics of the `rtweet` package and use it to fetch tweets
- tokenize tweets and clean it up for sentiment analysis
- Use one or two sentiment lexicons to analyze the sentiment of tweets
- Visualize the sentiments

## Packages

Below are the list of packages we are going to be using. If you don't have them installed please do so by using `install.packages()`

```
library(rtweet) # to fetch tweets from twitter using Twitter APIs
library(tidytext) # for making tidy text dataframe
library(tidyverse) # galaxy of data manipulation and visualization packages; ggplot2, dplyr, re
  adr, tidyr, tibble
library(wordcloud) # to creat cloud of words
library(RColorBrewer) # for nice color palette
```

## Getting Tweets

`rtweet` is a R client for accessing Twitter's REST and stream APIs.

With `rtweet` all you need is a Twitter account details (user name and password) and you can be up in running in minutes!

How is this possible, first starting with the following assumption (a) you are working in an interactive/live session of R and (b) you have installed the `{httpuv}` package. - All that is left for you to authorize the embedded `rstats2twitter` app (approve the browser popup), and your token will be created and saved/stored (for future sessions) for you!

But you can still choose to go the long way... Create a twitter app so as to obtain and use Twitter API access tokens checkout this link (<https://rtweet.info/articles/auth.html>).

FYI

Task	rstats2twitter application	Personal application
Work interactively	✓	✓
Create reproducible research	✗	✓
Search tweets	✓	✓
Stream tweets	✓	✓
Get friends	✓	✓
Get timelines	✓	✓
...	...	...
Post tweets	✗	✓
Run package tests	✗	✓

## Getting Tweets using the REST API via `get_timeline()`

For this demo, we are going to be getting tweets and retweets from the official pages of PDP and APC

```
## get most recent 3200 tweets posted by Official APC twitter account
apc_tweets <- get_timeline("APCNigeria", n = 3200)
pdp_tweets <- get_timeline("OfficialPDPNig", n = 3200)
```

## Load data

### Inspect the data

```
dim(apc_tweets)
```

```
## [1] 3198  90
```

```
dim(pdp_tweets)
```

```
## [1] 3200  90
```

```
colnames(apc_tweets)
```

```
## [1] "user_id"           "status_id"
## [3] "created_at"        "screen_name"
## [5] "text"              "source"
## [7] "display_text_width" "reply_to_status_id"
## [9] "reply_to_user_id"  "reply_to_screen_name"
## [11] "is_quote"          "is_retweet"
## [13] "favorite_count"    "retweet_count"
## [15] "quote_count"       "reply_count"
## [17] "hashtags"          "symbols"
## [19] "urls_url"          "urls_t.co"
## [21] "urls_expanded_url" "media_url"
## [23] "media_t.co"        "media_expanded_url"
## [25] "media_type"        "ext_media_url"
## [27] "ext_media_t.co"    "ext_media_expanded_url"
## [29] "ext_media_type"    "mentions_user_id"
## [31] "mentions_screen_name" "lang"
## [33] "quoted_status_id"  "quoted_text"
## [35] "quoted_created_at" "quoted_source"
## [37] "quoted_favorite_count" "quoted_retweet_count"
## [39] "quoted_user_id"    "quoted_screen_name"
## [41] "quoted_name"       "quoted_followers_count"
## [43] "quoted_friends_count" "quoted_statuses_count"
## [45] "quoted_location"   "quoted_description"
## [47] "quoted_verified"   "retweet_status_id"
## [49] "retweet_text"      "retweet_created_at"
## [51] "retweet_source"    "retweet_favorite_count"
## [53] "retweet_retweet_count" "retweet_user_id"
## [55] "retweet_screen_name" "retweet_name"
## [57] "retweet_followers_count" "retweet_friends_count"
## [59] "retweet_statuses_count" "retweet_location"
## [61] "retweet_description" "retweet_verified"
## [63] "place_url"         "place_name"
## [65] "place_full_name"   "place_type"
## [67] "country"           "country_code"
## [69] "geo_coords"        "coords_coords"
## [71] "bbox_coords"       "status_url"
## [73] "name"              "location"
## [75] "description"       "url"
## [77] "protected"         "followers_count"
## [79] "friends_count"     "listed_count"
## [81] "statuses_count"    "favourites_count"
## [83] "account_created_at" "verified"
## [85] "profile_url"       "profile_expanded_url"
## [87] "account_lang"      "profile_banner_url"
## [89] "profile_background_url" "profile_image_url"
```

```
# head(apc_tweets[,c('text', "is_retweet")])
pdp_tweets[1:5,] %>%
  mutate(hashtags2 = paste0(unlist(hashtags), sep=", ")) %>%
  select(user_id, created_at, screen_name, source, text, is_retweet, hashtags2) %>%
  knitr::kable() %>%
  kableExtra::kable_styling(bootstrap_options = "striped", full_width = F)
```

```
## Warning in kableExtra::kable_styling(., bootstrap_options = "striped",
## full_width = F): Please specify format in kable. kableExtra can customize either
## HTML or LaTeX outputs. See https://haozhu233.github.io/kableExtra/ for details.
```

user_id	created_at	screen_name	source	text	is_retweet	hashtags2
747064131166044160	2019-12-20 10:53:58	OfficialPDPNig	Twitter for iPhone	Media Interactive session with the @OfficialPDPNig NPS, @officialKolaO, live from Legacy House, Abuja <a href="https://t.co/9RI7JldYxz">https://t.co/9RI7JldYxz</a> ( <a href="https://t.co/9RI7JldYxz">https://t.co/9RI7JldYxz</a> )	FALSE	NA,
747064131166044160	2019-12-20 10:02:21	OfficialPDPNig	Twitter for iPhone	#UPDATE: Governor Ifeanyi Okowa taking the first lesson at the Festival of Nine Lessons and Carols at the Government House Chapel in Asaba. <a href="https://t.co/ZoKZFq5o1b">https://t.co/ZoKZFq5o1b</a> ( <a href="https://t.co/ZoKZFq5o1b">https://t.co/ZoKZFq5o1b</a> )	TRUE	UPDATE,
747064131166044160	2019-12-20 10:02:17	OfficialPDPNig	Twitter for iPhone	#UPDATE: The Corps Marshal of the Federal Road Safety Corps, Mr. Boboye Opeyemi (MFR), is now on a courtesy visit to Governor Ifeanyi Okowa to brief him about the plans of the Corps to ensure that the roads are safe this holiday season. <a href="https://t.co/1cEHcdSWXj">https://t.co/1cEHcdSWXj</a> ( <a href="https://t.co/1cEHcdSWXj">https://t.co/1cEHcdSWXj</a> )	TRUE	UPDATE,
747064131166044160	2019-12-19 18:57:20	OfficialPDPNig	Twitter for iPhone	"We are quite pleased with the way they have conducted themselves and we hope that other operations outside Delta State are able to emulate what they have done in order to ensure that there is an improved relationship between civilians and the military" Governor Okowa said. <a href="https://t.co/P39F8R0ft0">https://t.co/P39F8R0ft0</a> ( <a href="https://t.co/P39F8R0ft0">https://t.co/P39F8R0ft0</a> )	TRUE	NA,

user_id	created_at	screen_name	source	text	is_retweet	hashtags2
747064131166044160	2019-12-19 18:57:17	OfficialPDPNig	Twitter for iPhone	Governor Okowa who was represented by the Information Commissioner, Mr. Charles Aniagwu, gave the commendation when he paid an inspection visit to the Nigerian Army, 63 Brigade Super Camp in Asaba. <a href="https://t.co/mGcePcwqIU">https://t.co/mGcePcwqIU</a> ( <a href="https://t.co/mGcePcwqIU">https://t.co/mGcePcwqIU</a> )	TRUE	NA,

## Tidy Data

### Data Cleaning

Merging the two sets together (Just to carryout the cleaning process at once)

```
parties_tweet <- bind_rows(apc_tweets, pdp_tweets) %>%
  rename(party = screen_name, tweet = text) # Renaming screen_name to party and text to tweet
```

Extract the tweets with corresponding partys (APCNigeria and OfficialPDPNig)

```
tweets_df <- parties_tweet %>%
  filter(!is_retweet) %>% #Removing Retweets
  select(status_id, party, tweet, created_at) # status_id will serve as a unique identifier for each tweet
```

### Creating a tidy dataframe

Using `unnest_tokens()` to make a tidy data frame of all the words in our tweets dataframe.

```
remove_reg <- "&|<|>|"

tidy_tweet_df <- tweets_df %>%
  mutate(tweet = gsub("http.*", "", tweet)) %>% # removing url
  mutate(tweet = gsub("http.*", "", tweet)) %>% # removing url
  mutate(tweet = stringr::str_remove_all(tweet, remove_reg)) %>% # removing some foreign characters
  unnest_tokens(input = tweet, output = word) %>% # tokenization
  filter(!word %in% stop_words$word, #removing stopwords
         !word %in% str_remove_all(stop_words$word, ""),
         str_detect(word, "[a-z]")) # Removing non-alphanumeric character

knitr::kable(bind_rows(head(tidy_tweet_df), tail(tidy_tweet_df)))
```

status_id	party	created_at	word
1207633971384897537	APCNigeria	2019-12-19 12:09:10	vice
1207633971384897537	APCNigeria	2019-12-19 12:09:10	president

status_id	party	created_at	word
1207633971384897537	APCNigeria	2019-12-19 12:09:10	yemi
1207633971384897537	APCNigeria	2019-12-19 12:09:10	osinbajo
1207633971384897537	APCNigeria	2019-12-19 12:09:10	san
1207633971384897537	APCNigeria	2019-12-19 12:09:10	presides
1149740100353515520	OfficialPDPNig	2019-07-12 17:59:35	yokri
1149740100353515520	OfficialPDPNig	2019-07-12 17:59:35	roads
1149740100353515520	OfficialPDPNig	2019-07-12 17:59:35	ogulagha
1149740100353515520	OfficialPDPNig	2019-07-12 17:59:35	burutu
1149740100353515520	OfficialPDPNig	2019-07-12 17:59:35	lga
1149740100353515520	OfficialPDPNig	2019-07-12 17:59:35	delta

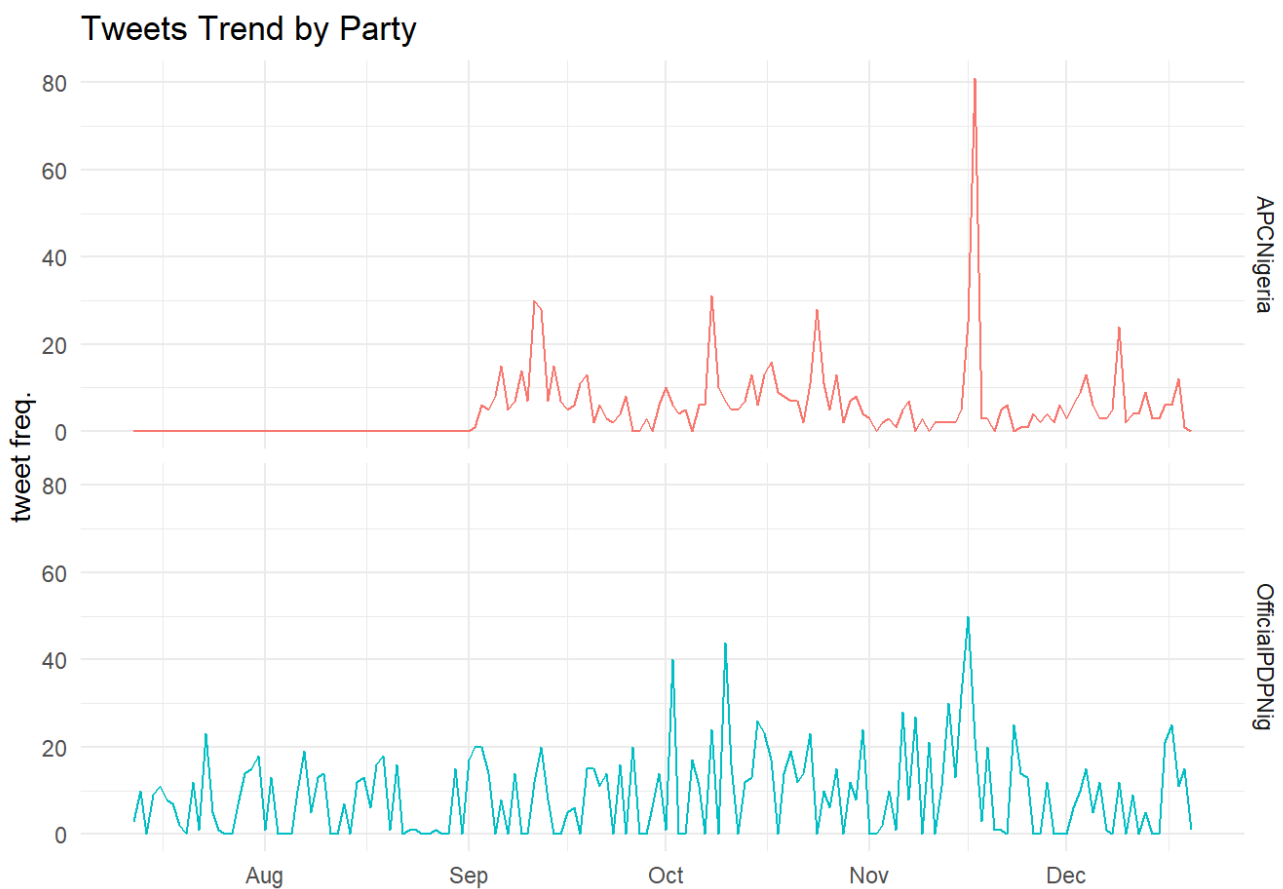
## Exploratory Analysis

### Tweet Trend

```

tweets_df %>%
  group_by(party) %>%
  ts_plot(by = "days") +
  facet_grid(party~.) +
  theme_minimal() +
  labs(title = "Tweets Trend by Party", x="", y = "tweet freq.") +
  theme(legend.position = "none")

```

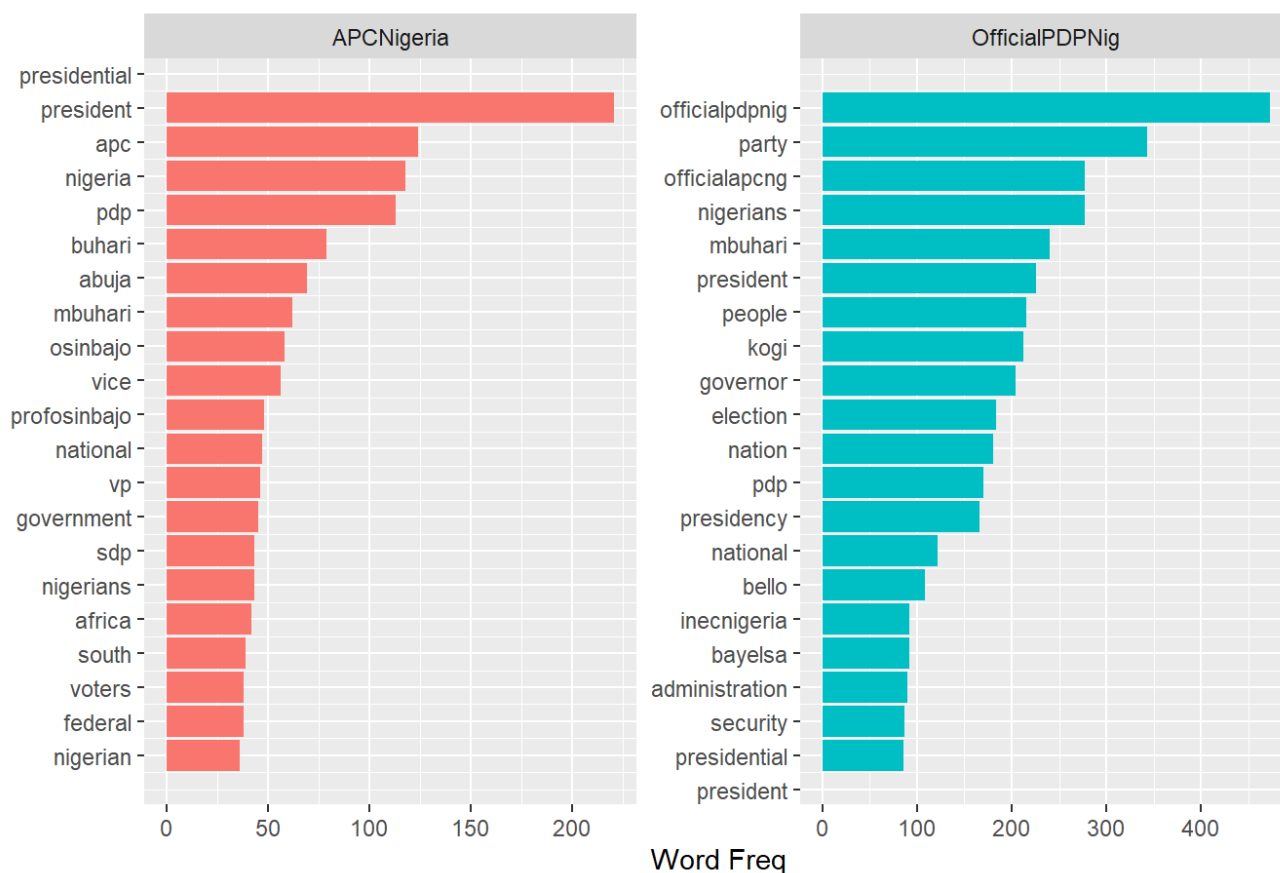


### Popular words

```
popular_words <- tidy_tweet_df %>%
  group_by(party) %>%
  count(word, party, sort = TRUE) %>%
  slice(seq_len(20)) %>% # top 10
  ungroup() %>%
  arrange(party, n) %>%
  mutate(row = row_number())

popular_words %>%
  ggplot(aes(row, n, fill = party)) +
  geom_col(show.legend = F) +
  labs(x = NULL, y = "Word Freq") +
  ggtitle("Popular Words by Political Parties") +
  # theme_lyrics() +
  facet_wrap(~party, scales = "free") +
  scale_x_continuous( # This handles replacement of row
    breaks = popular_words$row, # notice need to reuse data frame
    labels = popular_words$word) +
  coord_flip()
```

Popular Words by Political Parties



## Generate Wordcloud

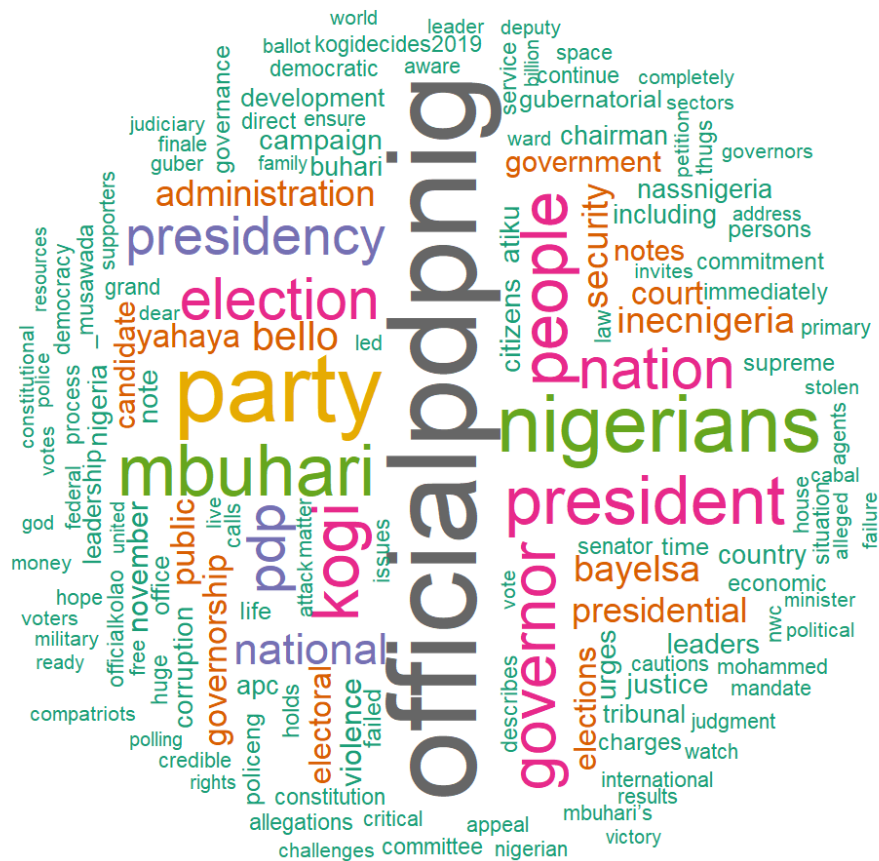
APC

```
apc_freq = tidy_tweet_df %>%
  filter(party == "APCNigeria") %>%
  count(word, name = "freq", sort = T) #>%

set.seed(1234) # for reproducibility
summary(apc_freq)
```







## Removing Some words

# Sentiment Analysis

For this we will be using the Bing Lexicon

```
head(get_sentiments(lexicon = "bing"))
```

```
## # A tibble: 6 x 2
##   word      sentiment
##   <chr>     <chr>
## 1 2-faces   negative
## 2 abnormal negative
## 3 abolish  negative
## 4 abominable negative
## 5 abominably negative
## 6 abominate negative
```

```
parties_tweet_senti_bing <- clean_tidy_tweet_df %>%
  inner_join(get_sentiments("bing"))
```

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

```
parties_tweet_senti_nrc <- tidy_tweet_df %>%
  inner_join(get_sentiments("nrc"))
```

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

```
parties_tweet_senti_bing
```

```
## # A tibble: 3,778 x 5
##   status_id      party    created_at      word      sentiment
##   <chr>         <chr>    <dtm>      <chr>     <chr>
## 1 1207446838321778688 APCNigeria 2019-12-18 23:45:34 supreme    positive
## 2 1207446838321778688 APCNigeria 2019-12-18 23:45:34 appeal     positive
## 3 1207282652979617794 APCNigeria 2019-12-18 12:53:09 breaking   negative
## 4 1207265437181980673 APCNigeria 2019-12-18 11:44:44 dilapidated negative
## 5 1207055721843560449 APCNigeria 2019-12-17 21:51:24 exemplary  positive
## 6 1207055721843560449 APCNigeria 2019-12-17 21:51:24 visionary  positive
## 7 1207055719943540736 APCNigeria 2019-12-17 21:51:24 celebrate  positive
## 8 1207055719943540736 APCNigeria 2019-12-17 21:51:24 exemplary  positive
## 9 1206973612357562368 APCNigeria 2019-12-17 16:25:08 glad       positive
## 10 1206933773985746944 APCNigeria 2019-12-17 13:46:50 kindly     positive
## # ... with 3,768 more rows
```

## Result

### Overall Sentiment

```

parties_tweet_senti_bing_res <- parties_tweet_senti_bing %>%
  group_by(status_id, party) %>%
  count(sentiment) %>% # count the # of positive & negative words
  spread(sentiment, n, fill = 0) %>% # made data wide rather than narrow
  mutate(sentiment_score = positive - negative) %>% # Calculate Sentiment Score
  mutate(sentiment = ifelse(sentiment_score < 0, "Negative", ifelse(sentiment_score > 0, "Positive", "Neutral"))) %>% # Calculate Sentiment Score
  ungroup()

table(parties_tweet_senti_bing_res$sentiment)

```

```

##
## Negative Neutral Positive
##      849      175      572

```

## Quick Check

```

knitr::kable(
  parties_tweet_senti_bing_res %>%
    slice(seq_len(20)) %>%
    left_join(tweets_df) %>%
    select(created_at, status_id, party, tweet, positive, negative, sentiment)
)

```

```
## Joining, by = c("status_id", "party")
```

created_at	status_id	party	tweet	positive	negative	sentiment
2019-07-13 18:08:50	1150104815260905472	OfficialPDPNig	Funke Olakunrin: The @OfficialPDPNig condemns in totality the gruesome killing of Funke Olakunrin, daughter of Chairman of Yoruba Socio-cultural group, Afenifere, Pa Reuben Fasoranti, by yet to be apprehended gunmen. <a href="https://t.co/bpZMJtwTX1">https://t.co/bpZMJtwTX1</a> ( <a href="https://t.co/bpZMJtwTX1">https://t.co/bpZMJtwTX1</a> )	0	3	Negative
2019-07-13 18:08:51	1150104820273078272	OfficialPDPNig	The @OfficialPDPNig also laments the unpardonable lethargy in apprehending Funke's killers since Friday. The party describes as heartrending that under the President @MBuhari administration, our nation has become a killing field and large funeral palour with insurgents, bandits and assassins having a field day waylaying and killing innocent compatriots, especially those making...	0	1	Negative
2019-07-13 18:08:52	1150104823557251079	OfficialPDPNig	The party describes as heartrending that under the President @MBuhari administration, our nation has become a killing field and large funeral palour with insurgents, bandits and assassins having a field day waylaying and killing innocent compatriots, especially those making...	0	2	Negative

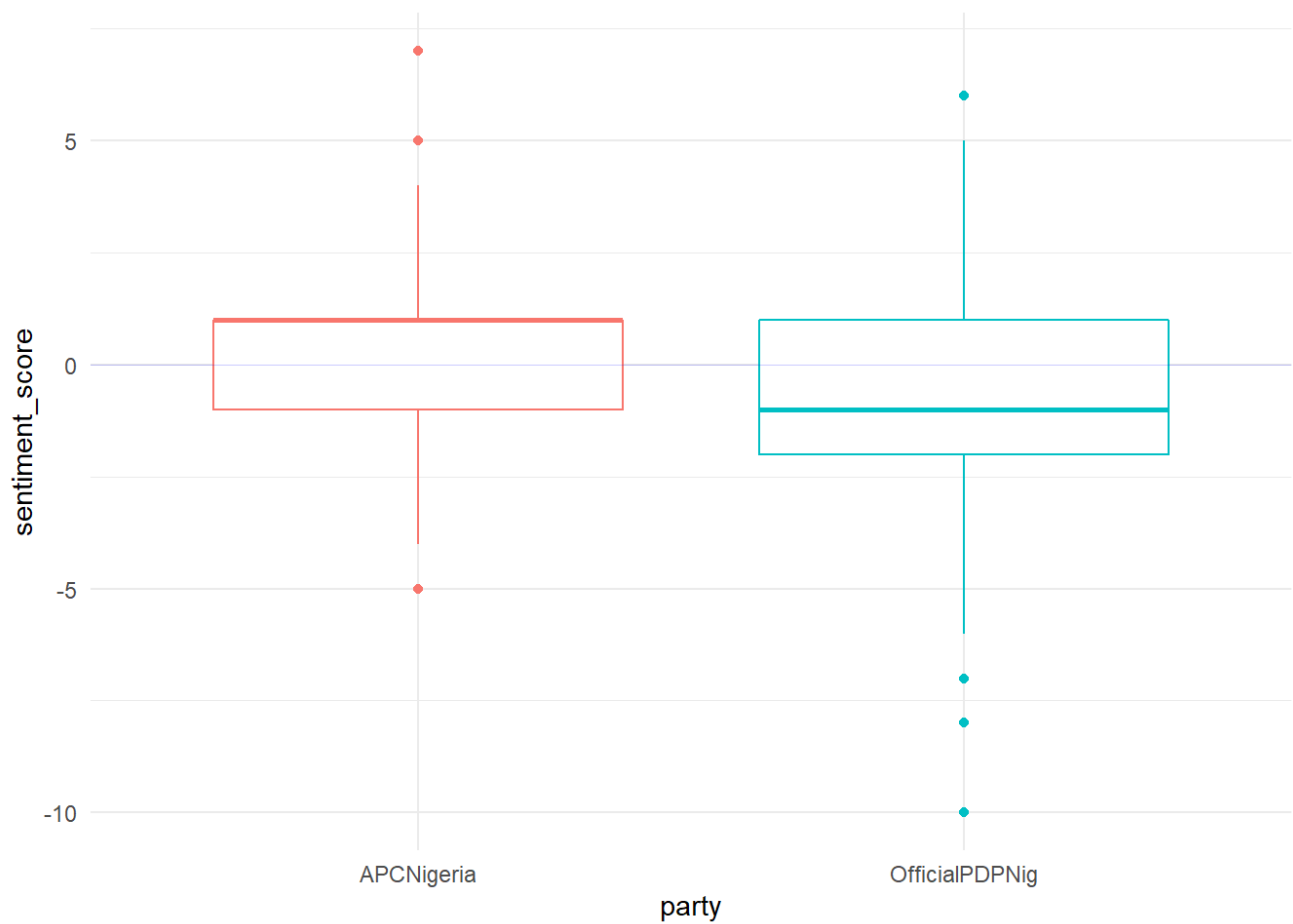
created_at	status_id	party	tweet	positive	negative	sentiment
2019-07-13 18:08:53	1150104829345316864	OfficialPDPNig	Our party is worried that security agencies are rather engaging in debates over the identity of the killers and hurriedly pointing to a case of abduction even before any extensive investigation has been carried out on the horrific and dastardly act.	1	3	Negative
2019-07-13 18:08:54	1150104832122003456	OfficialPDPNig	The party bewails that the killing of Funke Olakunrin marks yet another national loss occasioned by the failure of the Buhari administration to secure the nation beyond lip service, a development that has emboldened marauders in our country.	1	3	Negative
2019-07-13 18:08:55	1150104834525335552	OfficialPDPNig	The PDP posits that Funke Olakunrin must not die in vain. Our dear nation must not continue to lose her finest and brightest in the hands of bandits. Our party therefore charges the authorities to go beyond condolence...	2	3	Negative
2019-07-13 18:08:56	1150104841718513664	OfficialPDPNig	The @OfficialPDPNig, while condoling Pa Fasoranti, charges the Inspector General of Police, Adamu Mohammed, to take immediate and decisive steps, beyond the lip service of the @MBuhari administration, to track down the killers...	1	0	Positive
2019-07-13 18:08:57	1150104845191438341	OfficialPDPNig	... without further delay and make them face the full wrath of the law, no matter who they are.	0	2	Negative
2019-07-15 19:24:26	1150848616011042816	OfficialPDPNig	Insecurity: The @OfficialPDPNig says former President Olusegun Obasanjo's open letter to President @MBuhari on the worsening insecurity and unrestrained bloodletting in the country, vindicates its position and that of millions of Nigerians that ... <a href="https://t.co/DRSkxb00P3">https://t.co/DRSkxb00P3</a> ( <a href="https://t.co/DRSkxb00P3">https://t.co/DRSkxb00P3</a> ) ...President @MBuhari has failed as the nation's chief security officer.	0	3	Negative
2019-07-15 19:24:27	1150848618670252032	OfficialPDPNig	The @OfficialPDPNig describes the open letter as patriotic, momentous and captures the apprehensions of majority of Nigerians across the board.	2	2	Neutral

created_at	status_id	party	tweet	positive	negative	sentiment
2019-07-15 19:24:27	1150848620683497472	OfficialPDPNig	The party notes that President Obasanjo's letter reechoes its numerous caution against the violations, insensitivity, impunity, vindictiveness, divisiveness, intolerance and dangerous sectionalism that... ...characterize the @MBuhari administration and which are jeopardizing our national unity.	0	6	Negative
2019-07-15 19:24:28	1150848622570983424	OfficialPDPNig	The @OfficialPDPNig urges President @MBuhari to read between the lines and listen to the voice of reason from well-meaning Nigerians at this moment. Our party is deeply worried that our nation, which grew in leaps and bounds in all sectors before the @OfficialAPCNg took over in 2015,	1	0	Positive
2019-07-15 19:24:28	1150848624311578624	OfficialPDPNig	is now dangerously sliding on all fronts under President @MBuhari, who has shown crass incompetence in managing our national affairs. The @OfficialPDPNig maintains that had President @MBuhari heed wise counsel to rejig his parade and respect the wishes and aspirations of Nigerians in the national interest, our nation would not have been on this dangerous bend.	0	3	Negative
2019-07-15 19:24:28	1150848626098364418	OfficialPDPNig	The fact remains that President @MBuhari has not demonstrated the capacity to lead a nation as complex as our Nigeria. Events have shown that he cannot handle our diversities and ensure a ...	3	1	Positive
2019-07-15 19:24:29	1150848627792912385	OfficialPDPNig	...united and equitable country that guarantees fairness, justice and space to all citizens without fear or favour. The @OfficialPDPNig therefore counsels President @MBuhari to listen to the pulse of our nation;	1	1	Neutral
2019-07-15 19:24:29	1150848629667762179	OfficialPDPNig	accept his failure and seek help from other well-meaning Nigerians so as to take back our nation from this imminent precipice.	3	1	Positive
2019-07-15 19:24:30	1150848631488045059	OfficialPDPNig		0	1	Negative

created_at	status_id	party	tweet	positive	negative	sentiment
2019-07-16 12:51:24	1151112093040566272	OfficialPDPNig	Just In: The father of the Former Governor of Kaduna State & Fmr National Chairman of the PDP, Sen. Ahmed Mohammed Makarfi, has passed on. The body of his late father will be buried by 4:00Pm later today according to Islamic Rites. May his gentle soul rest in perfect peace, Amin. <a href="https://t.co/IN4Jpwg83T">https://t.co/IN4Jpwg83T</a> ( <a href="https://t.co/IN4Jpwg83T">https://t.co/IN4Jpwg83T</a> ) The national leadership, on behalf of the entire @OfficialPDPNig family, condoles with the former Chairman of the National Caretaker Committee of our great	3	0	Positive
2019-07-16 17:05:17	1151175985255145477	OfficialPDPNig	party and former Governor of Kaduna state, Sen. Ahmed Mohammed Makarfi over the death of his father, Malam Muhammad Makarfi <a href="https://t.co/CnB2hNs2UI">https://t.co/CnB2hNs2UI</a> ( <a href="https://t.co/CnB2hNs2UI">https://t.co/CnB2hNs2UI</a> ) The @OfficialPDPNig is deeply saddened by the news of the passing of Malam Makarfi, a devoted statesman and devout Muslim, who dedicated his life to the service of God, stability and growth of our dear nation and the good of humanity at large.	0	1	Negative
2019-07-16 17:05:18	1151175989046796288	OfficialPDPNig	Muslim, who dedicated his life to the service of God, stability and growth of our dear nation and the good of humanity at large.	3	0	Positive

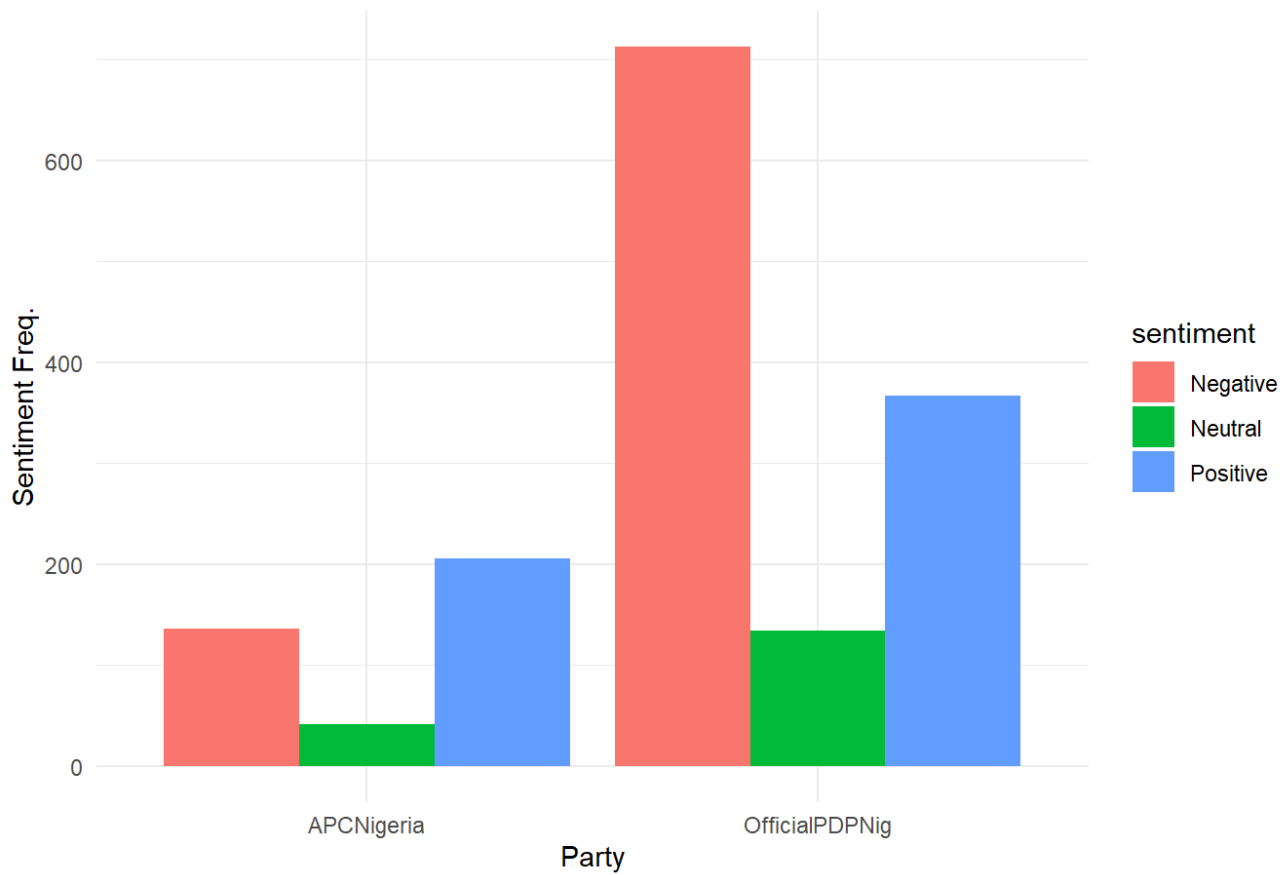
## Boxplot of Overall Sentiment

```
# plot of sentiment by party
ggplot(parties_tweet_senti_bing_res, aes(x = party, y = sentiment_score, color = party)) +
  geom_boxplot() + # draw a boxplot for each party
  theme_minimal() +
  theme(legend.position = "none") +
  geom_hline(yintercept = 0, size = .2, alpha = .2, color = "blue")
```



```
parties_tweet_senti_bing_res %>%
  group_by(party) %>%
    count(sentiment) %>% # count the # of positive & negative words
    ggplot(aes(x = party, y = n, fill = sentiment)) +
    geom_bar(stat = "identity", position = "dodge") +
    labs(title = "Overall Sentiment of APC and PDP tweets", x = "Party", y = "Sentiment Frequency") +
    theme_minimal()
```

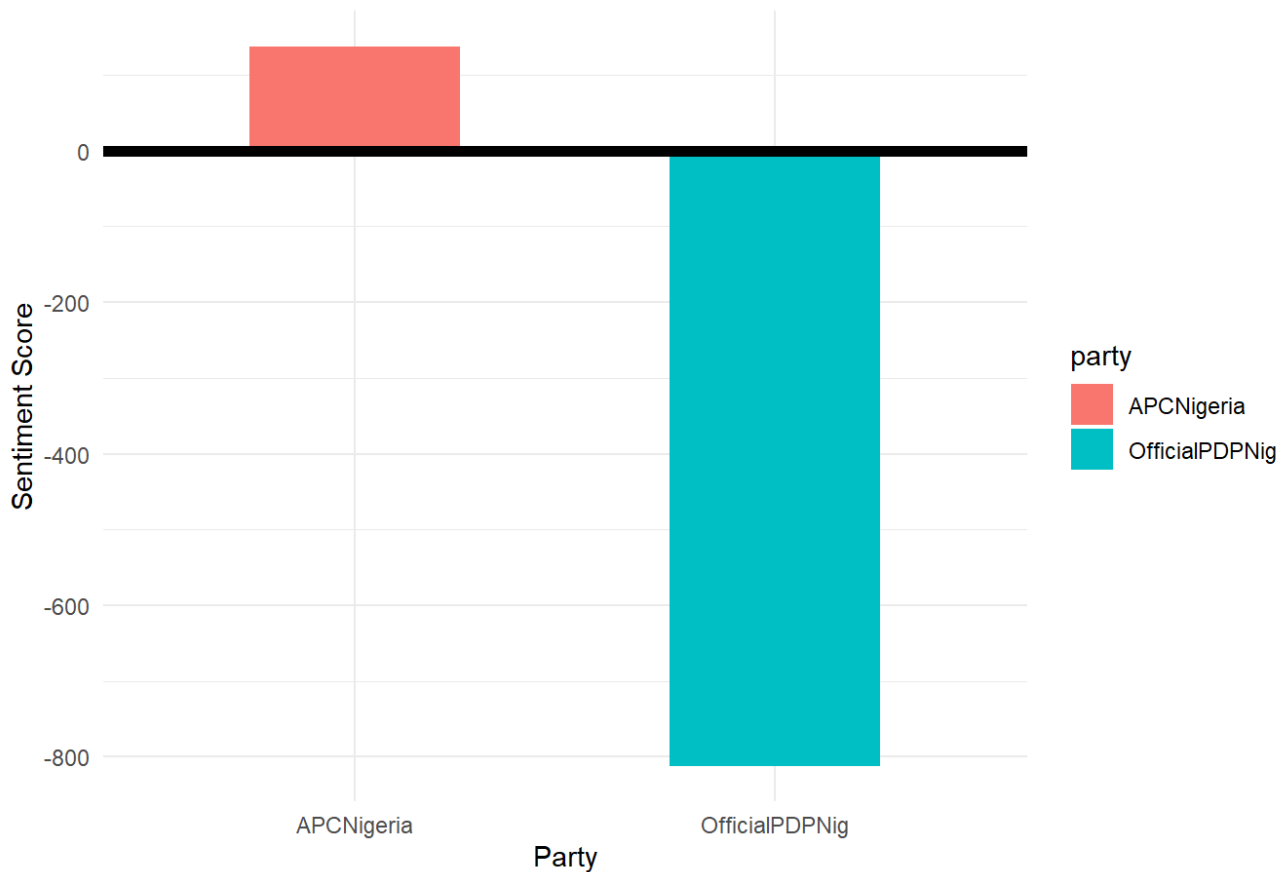
Overall Sentiment of APC and PDP tweets



```
parties_tweet_senti_bing_res %>%
  group_by(party) %>%
  summarise(all_sentiment = sum(sentiment_score)) %>%
  ungroup() %>%
  ggplot(aes(x = party, y = all_sentiment, fill = party)) +
  geom_bar(stat = "identity", width = .5) +
  geom_hline(yintercept = 0, size = 2) +
  # coord_flip() +
  theme_minimal() +
  labs(title = "Overall Sentiment of APC and PDP tweets", x = "Party", y = "Sentiment Score") +
  theme_minimal()
```



## Overall Sentiment of APC and PDP tweets

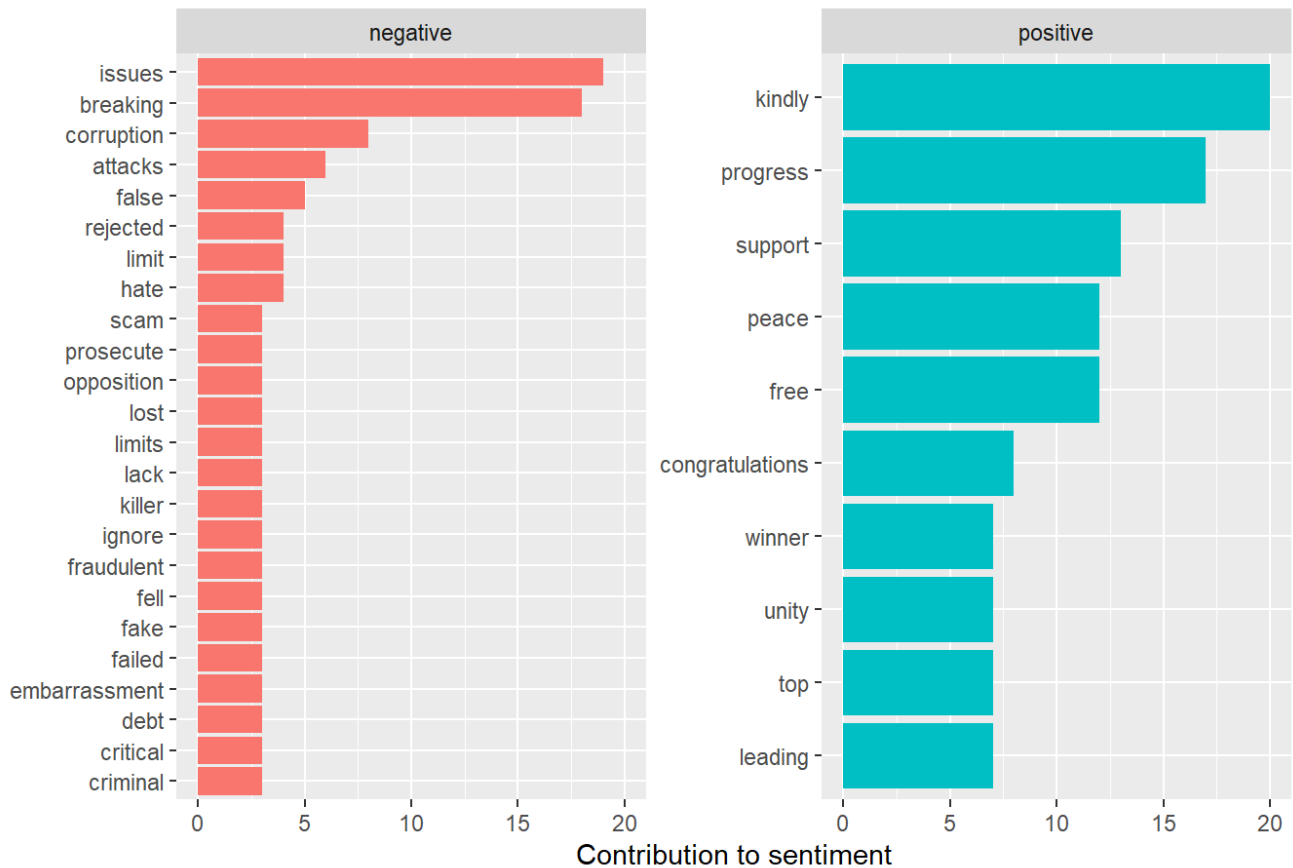


```
apc_word_counts <- parties_tweet_senti_bing %>%
  filter(party == "APCNigeria") %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()
```

```
apc_word_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(title = "APC - Top Polarize Words",
       y = "Contribution to sentiment",
       x = NULL) +
  coord_flip()
```

```
## Selecting by n
```

## APC - Top Polarize Words

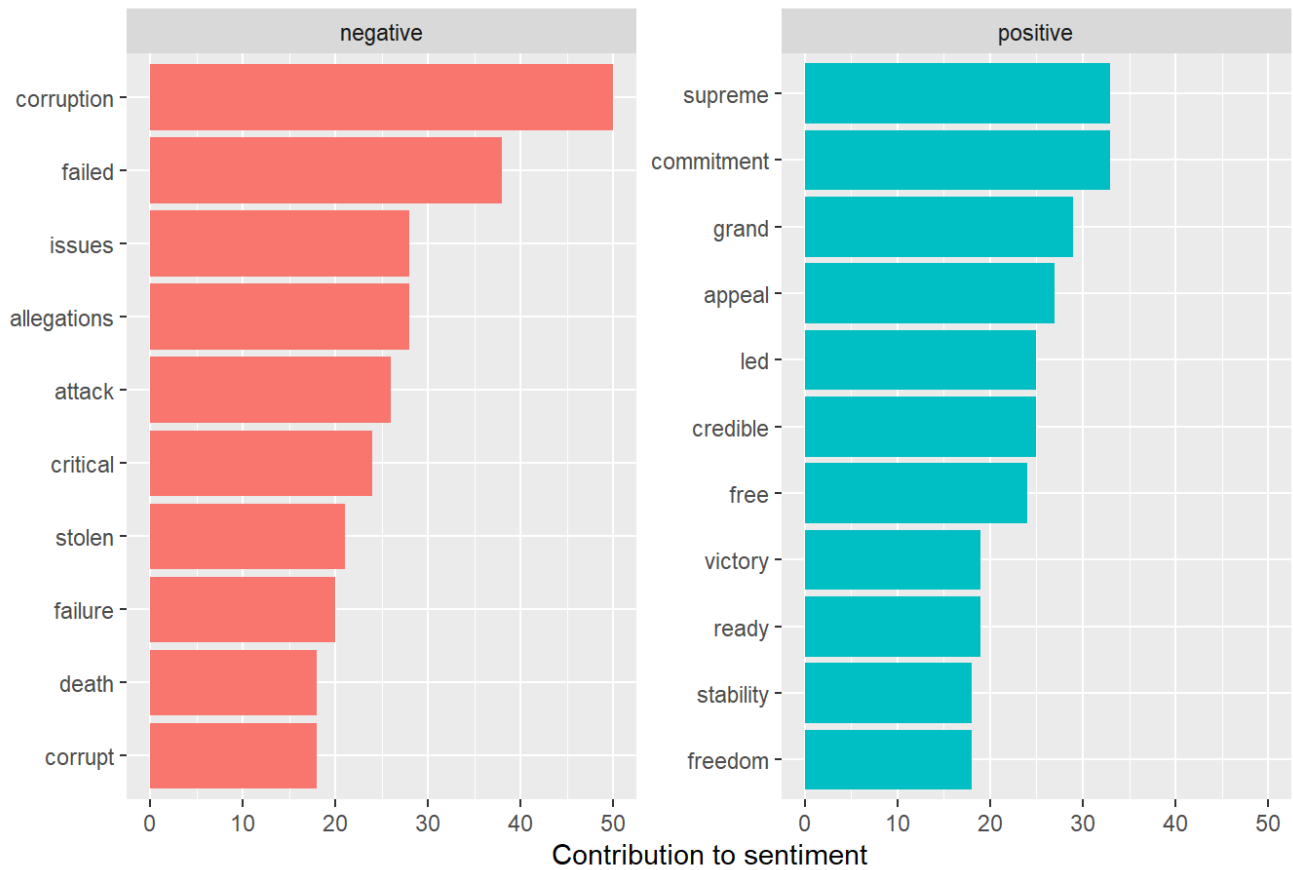


```
pdp_word_counts <- parties_tweet_senti_bing %>%
  filter(party == "OfficialPDPNig") %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()

pdp_word_counts %>%
  group_by(sentiment) %>%
  top_n(10) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(title = "PDP - Top Polarize Words",
       y = "Contribution to sentiment",
       x = NULL) +
  coord_flip()
```

```
## Selecting by n
```

## PDP - Top Polarize Words



## Comparison Word Cloud

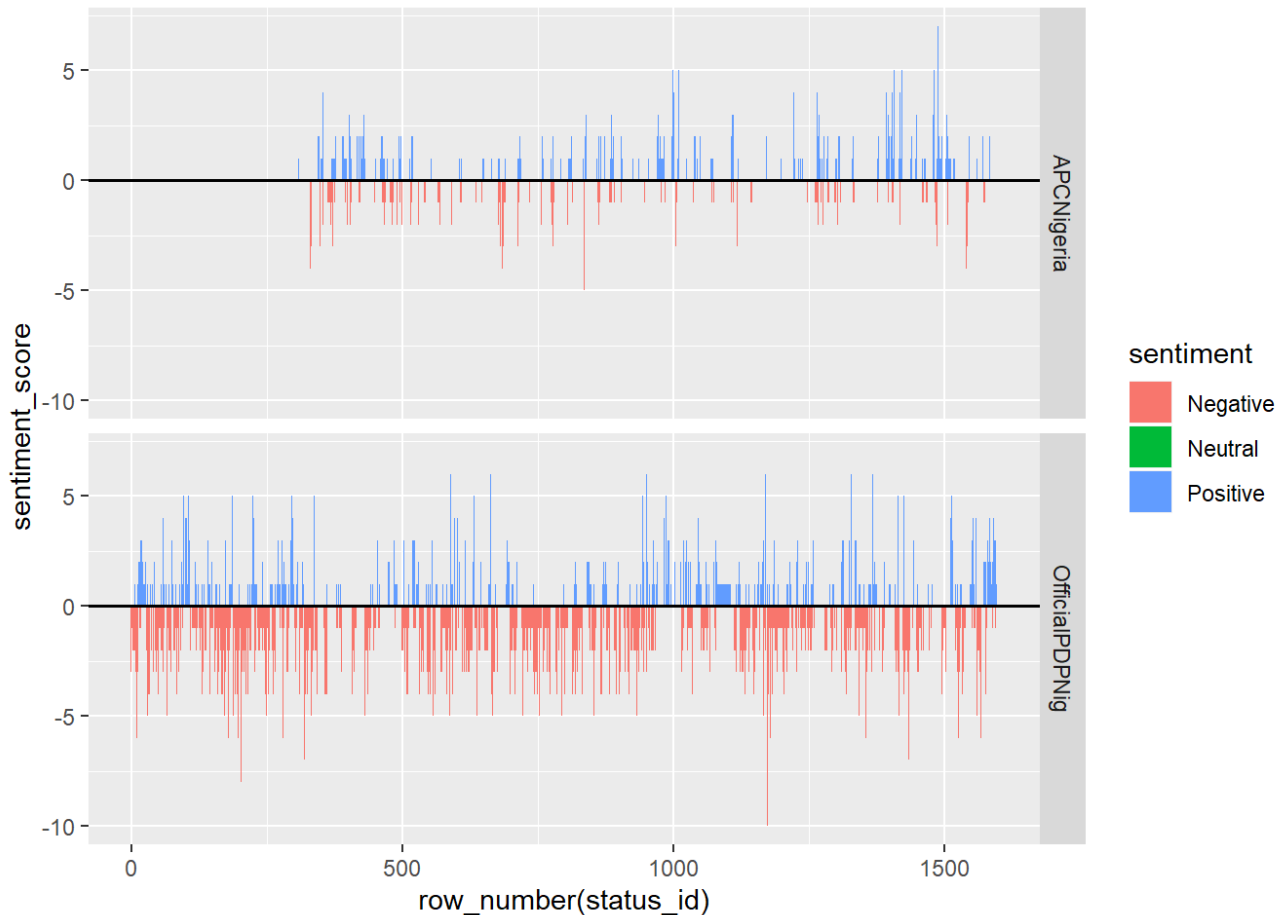
```
library(reshape2)

apc_word_counts %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("#f74d23", "green"),
    max.words = 100)
```



# Sentiment Per tweet

```
parties_tweet_senti_bing_res %>%
  ggplot(aes(x = row_number(status_id), y = sentiment_score )) +
  geom_col(aes(fill = sentiment)) +
  geom_hline(yintercept = 0, size = .7) +
  # geom_smooth(method = "auto") + # pick a method & fit a model
  facet_grid(party~.)
```



## Improving the Result

### 1. Normalizing for text length

How does normalizing for text length (for each tweet) change the outcome of the analysis?

### 2. Using a different sentiment lexicon

Hint (<https://www.kaggle.com/kakiac/sentiment-analysis-with-r-exercise-2>) # Does using a different lexicon result in a different outcome for your # analysis? What does this suggest about the original analysis?

### 3. Creating your own sentiment lexicon

Hint (<https://www.kaggle.com/kakiac/sentiment-analysis-with-r-exercise-3/data>) # How does this affect your analysis? Do you think it would have had a different # effect if you had annotated 500 words instead? 50,000? Would your new lexicon # be helpful in analyzing product reviews? Tweets?