Untitled

President Sentiment Analysis

Lydon B Johnson gave his state of the union address in 1964 and Ford in 1975. According to the data, LBJ had a 59% positivty rate (139/233 tokenized words) with his most frequent words used being 'faith' (8 times), 'good' (6 times) and 'create' (4 times). Likewise, Ford had a 61% positivity rate (230/378 tokenized words) with his most used words being 'public' (11), 'faith' (8) and 'time' (7). The analysis for both presidents had a fairly positive sentiment rate (Even though they both were not elected!).

```
library(tidytext)
```

```
## Warning: package 'tidytext' was built under R version 3.4.4
```

```
library(tidyverse)
library(ggplot2)

#get txt file
LBJ <- data_frame(name="LBJ", text = read_lines("LBJ.txt"))
Ford <- data_frame(name="Ford", text = read_lines("Ford.txt"))

#tokenize data
tokenized1 <- LBJ %>% unnest_tokens(word, text)
tokenized2 <- Ford %>% unnest_tokens(word, text)

nrc_anticipation <- get_sentiments("nrc") %>%
  filter(sentiment == "anticipation")

nrc_anger <- get_sentiments("nrc") %>%
  filter(sentiment == "anger")

nrc_joy <- get_sentiments("nrc") %>%
  filter(sentiment == "joy")
```

LBJ Word Counts

```
joy <- tokenized1 %>%
  filter(name == "LBJ") %>% inner_join(nrc_joy) %>% count(word, sort = TRUE)
joy
```

```
## # A tibble: 47 x 2
##    word         n
##    <chr>    <int>
##  1 faith        8
##  2 good         6
##  3 create       4
##  4 peace        4
##  5 food         3
##  6 progress     3
##  7 achieve      2
##  8 finally      2
##  9 freedom      2
```

```
## 10 hope         2
## # ... with 37 more rows
```

```
anticipation <- tokenized1 %>%
  filter(name == "LBJ") %>% inner_join(nrc_anticipation) %>% count(word, sort = TRUE)
anticipation
```

```
## # A tibble: 56 x 2
##    word            n
##    <chr>       <int>
##  1 public         11
##  2 faith           8
##  3 time            7
##  4 good            6
##  5 peace           4
##  6 seek            4
##  7 develop         3
##  8 opportunity     3
##  9 progress        3
## 10 continue        2
## # ... with 46 more rows
```

```
anger <- tokenized1 %>%
  filter(name == "LBJ") %>% inner_join(nrc_anger) %>% count(word, sort = TRUE)
anger
```

```
## # A tibble: 38 x 2
##    word               n
##    <chr>          <int>
##  1 poverty            9
##  2 cutting            3
##  3 hate               3
##  4 aggression         2
##  5 attack             2
##  6 defense            2
##  7 devastating        2
##  8 discrimination     2
##  9 fight              2
## 10 vote               2
## # ... with 28 more rows
```

```
# get the sentiment from the LBJ:
sentiment <- tokenized1 %>%
  inner_join(get_sentiments("bing")) %>%
  count(sentiment) %>%
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative)
sentiment
```

```
## # A tibble: 1 x 3
##   negative positive sentiment
##      <dbl>    <dbl>     <dbl>
## 1     94.0      139      45.0
```

Ford Word Counts

```
# get the sentiment from the Ford:
sentiment2 <- tokenized2 %>%
  inner_join(get_sentiments("bing")) %>% # pull out only sentiment words
  count(sentiment) %>% # count the # of positive & negative words
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative) # # of positive words - # of negative owrds
sentiment2
```

```
## # A tibble: 1 x 3
##   negative positive sentiment
##      <dbl>    <dbl>     <dbl>
## 1      148      230      82.0
```

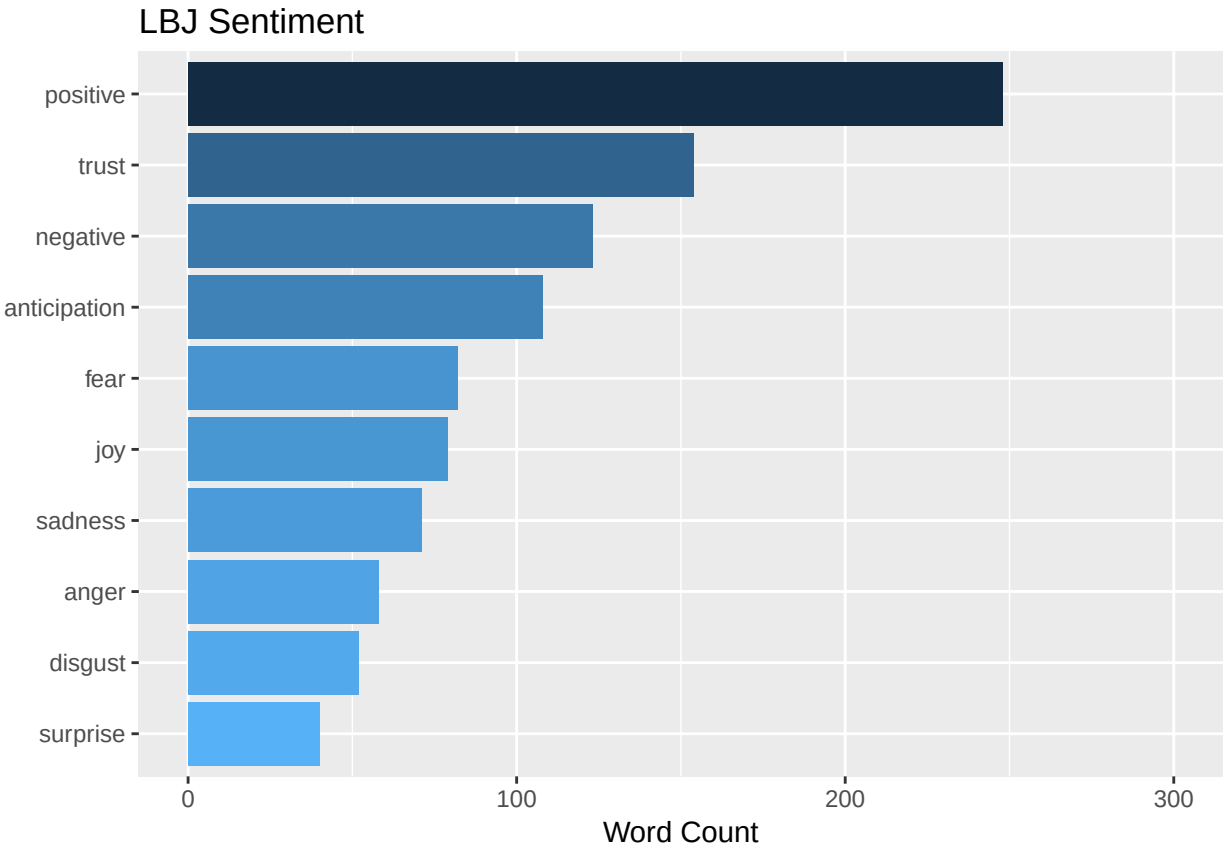
```
# sentiments for every word in file
tidy_pres <- tokenized1 %>%
  inner_join(get_sentiments("nrc"))
#tidy_pres

tidy_pres2 <- tokenized2 %>%
  inner_join(get_sentiments("nrc"))
#tidy_pres2
```

LBJ Sentiment

```
nrc_plot <- tidy_pres %>%
  group_by(sentiment) %>%
  summarise(word_count = n()) %>%
  ungroup() %>%
  mutate(sentiment = reorder(sentiment, word_count)) %>%

  ggplot(aes(sentiment, word_count, fill = -word_count)) +
  geom_col() +
  guides(fill = FALSE) +
  labs(x = NULL, y = "Word Count") +
  scale_y_continuous(limits = c(0, 300)) +
  ggtitle("LBJ Sentiment") +
  coord_flip()
nrc_plot
```

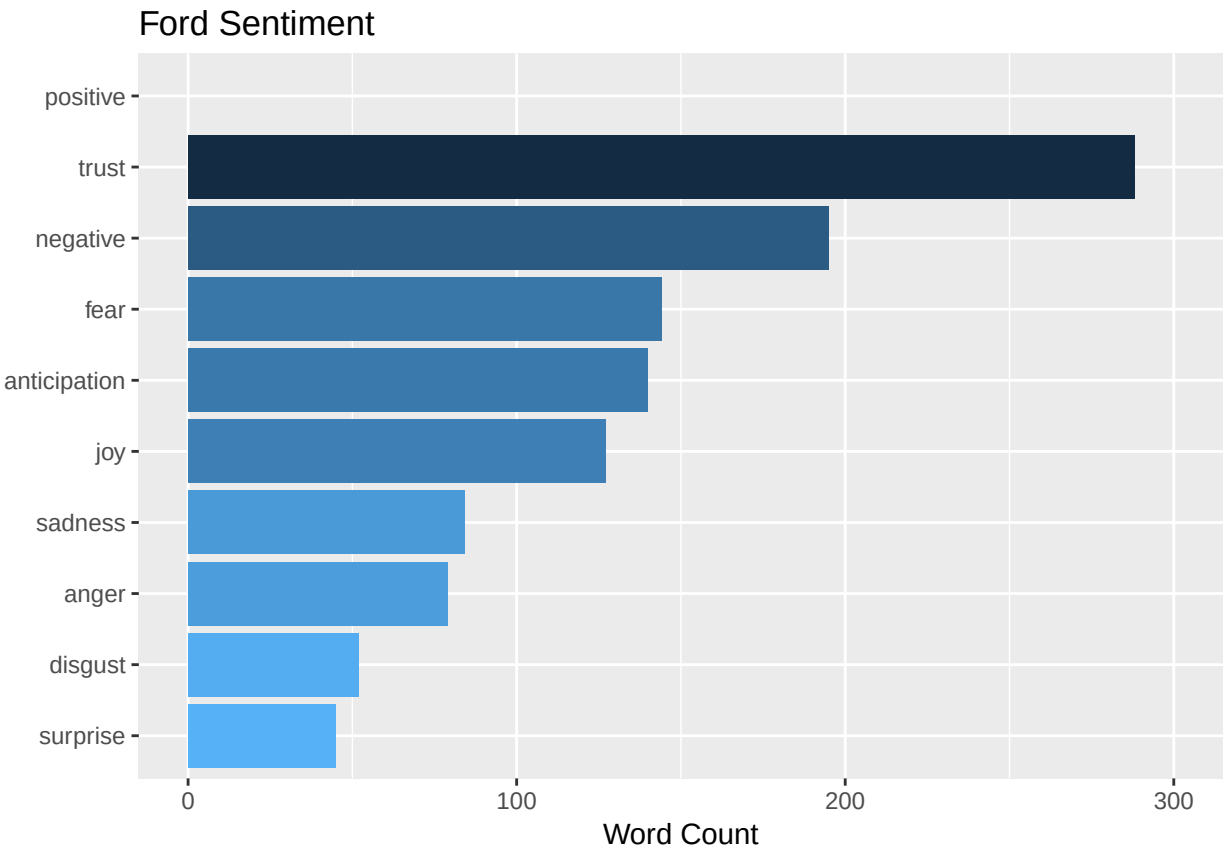


Ford Sentiment

```
#barplot for Ford
nrc_plot2 <- tidy_pres2 %>%
  group_by(sentiment) %>%
  summarise(word_count = n()) %>%
  ungroup() %>%
  mutate(sentiment2 = reorder(sentiment, word_count)) %>%

  ggplot(aes(sentiment2, word_count, fill = -word_count)) +
  geom_col() +
  guides(fill = FALSE) +
  labs(x = NULL, y = "Word Count") +
  scale_y_continuous(limits = c(0, 300)) +
  ggtitle("Ford Sentiment") +
  coord_flip()
nrc_plot2
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```


Ford Sentiment
positive
trust
negative
fear
anticipation
joy
sadness
anger
disgust
surprise
0
100
200
300
Word Count