Sentiment Analysis

Han Nguyen, Carlos Echeverri, Nathan Mokhtarzadeh

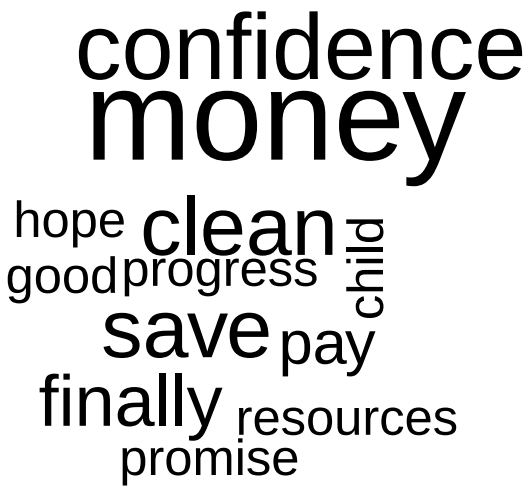
April 5, 2018

```
library(readtext)
library(tidytext)
library(dplyr)
library(stringr)
library(tidyr)
library(ggplot2)
library(wordcloud)
library(reshape2)

# read in text
jfk <- readtext("JFK first state of the union.txt") %>%
  unnest_tokens(word, text)

obama <- readtext("obama first state of the union.txt") %>%
  unnest_tokens(word, text)

# joy, anger, anticipation
nrc_joy <- get_sentiments("nrc") %>%
  filter(sentiment == "joy")

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

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

Obama

Joy: 52 words

```
obama_joy <- obama %>%
  inner_join(nrc_joy) %>%
  count(word, sort = TRUE)

obama_joy_wordcloud <- obama_joy %>%
  with(wordcloud(word, n))
```

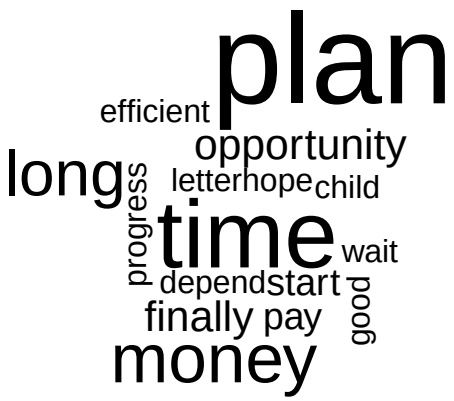


Anger: 51 words

```
obama_anger <- obama %>%
  inner_join(nrc_anger) %>%
  count(word, sort = TRUE)

obama_anger_wordcloud <- obama_anger %>%
  with(wordcloud(word, n))
```

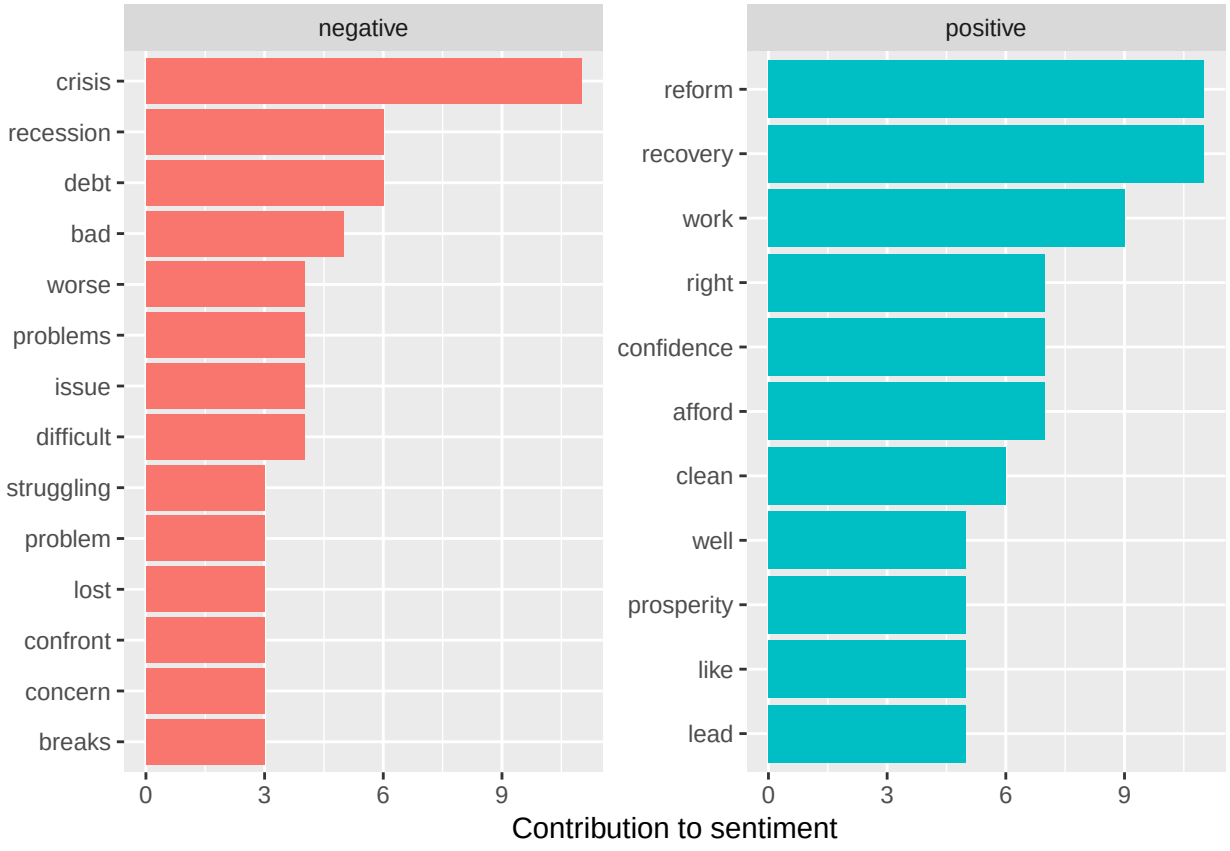
Anticipation: 77 words

```
obama_anticipation <- obama %>%
  inner_join(nrc_anticipation) %>%
  count(word, sort = TRUE)

obama_anticipation_wordcloud <- obama_anticipation %>%
  with(wordcloud(word, n))
```



Positive vs. Negative Words

```
bing_word_counts <- obama %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()

bing_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(y = "Contribution to sentiment",
       x = NULL) +
  coord_flip()
```




```
obama %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("gray20", "gray80"),
                   max.words = 100)
```


negative
terrible
peril
inaction
ordeal
overdue
extremism
lose
failure
doubt
terrorism
oversight
hardships
slowly
wasteful
struggling
problem
crushing
threats
harm
waste
concern
problems
fearful
ignore
worse
issue
lost
decline
costly
cold
recession
critical
suffering
burden
failed
fear
bad debt
difficult
anger
abuse
strain
blame
crisis
hard
fraud
vice
clean
reform
easy
lead
recovery
great
like
good
well
work
support
tough
progress
helping
safe
best
assure
afford
right
promise
benefits
enough
secure
revive
confidence
prosperity
free
protect
affordable
bless
fastest
positive
savings
trust
healthy
helped
efficient
commitment
threat
doubtful
breaks
confront

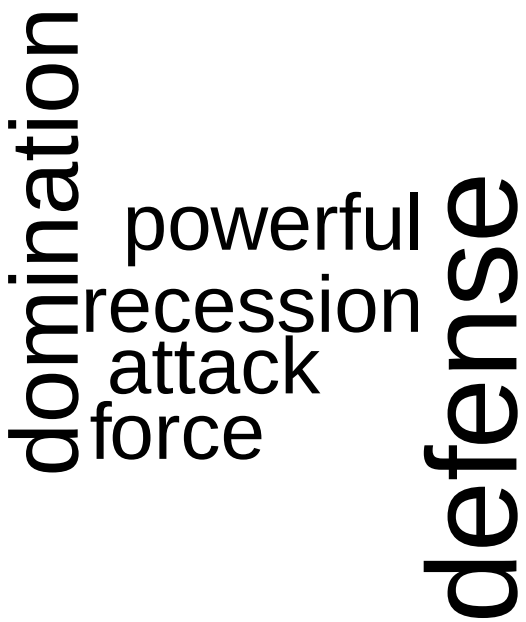
JFK

Joy: 62 words

```
jfk_joy <- jfk %>%
  inner_join(nrc_joy) %>%
  count(word, sort = TRUE)

jfk_joy_wordcloud <- jfk_joy %>%
  with(wordcloud(word, n))
```

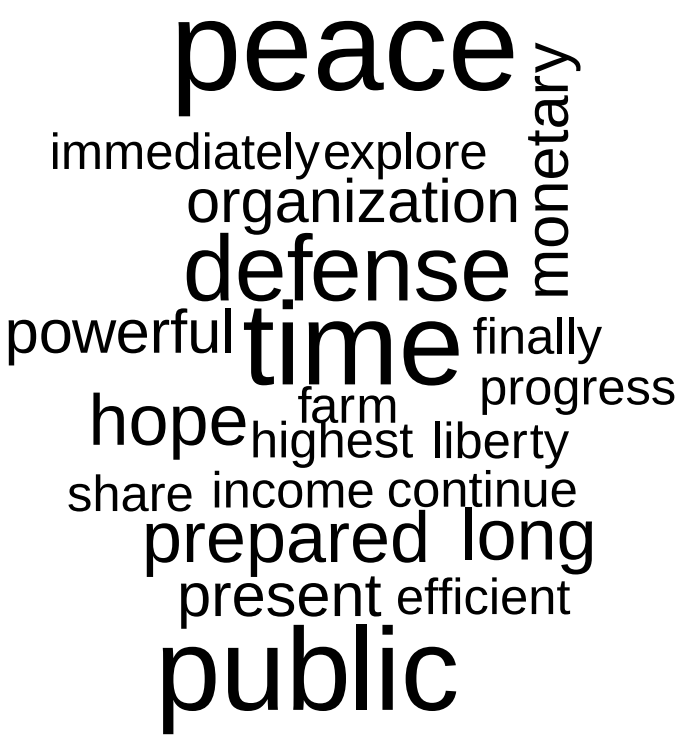
Anger: 51 words

```
jfk_anger <- jfk %>%
  inner_join(nrc_anger) %>%
  count(word, sort = TRUE)

jfk_anger_wordcloud <- jfk_anger %>%
  with(wordcloud(word, n))
```



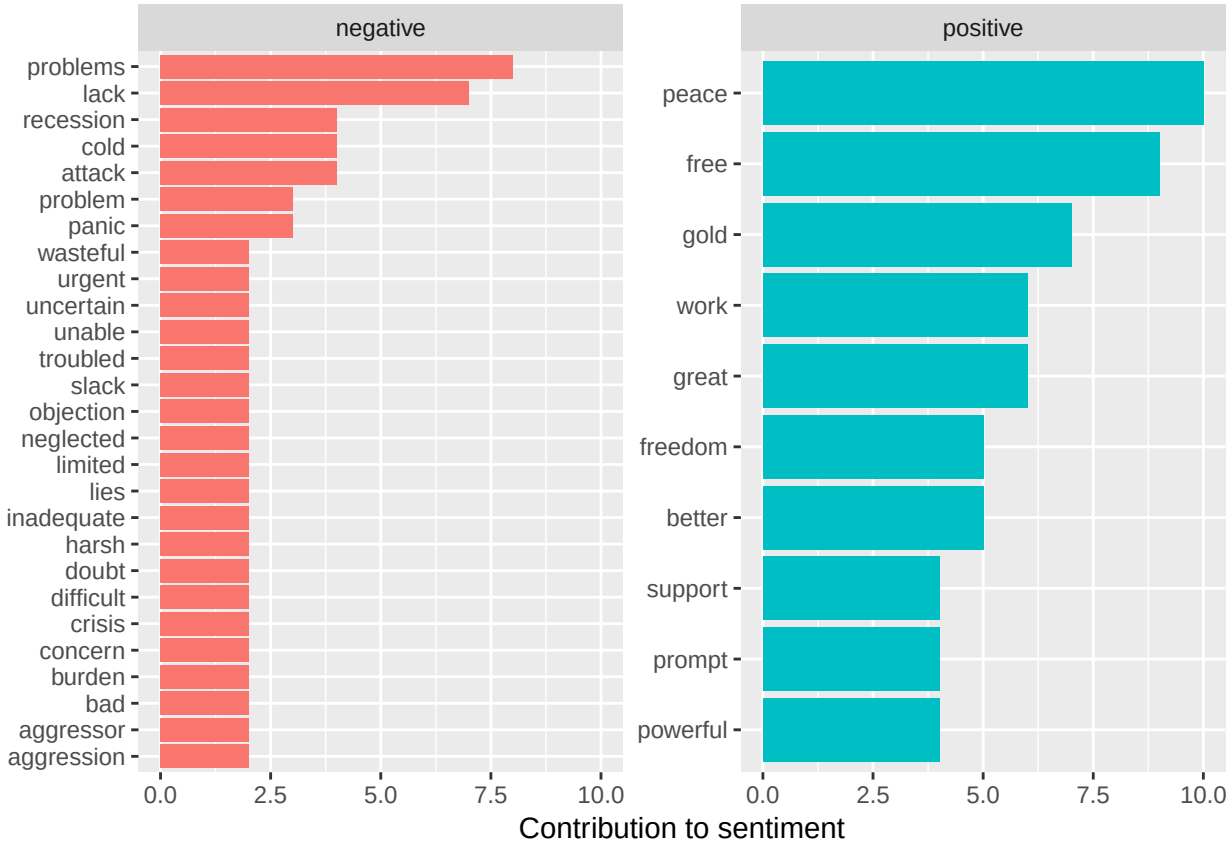
Anticipation: 89 words

```
jfk_anticipation <- jfk %>%
  inner_join(nrc_anticipation) %>%
  count(word, sort = TRUE)

jfk_anticipation_wordcloud <- jfk_anticipation %>%
  with(wordcloud(word, n))
```



Positive vs. Negative Words

```
bing_word_counts <- jfk %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()

bing_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(y = "Contribution to sentiment",
       x = NULL) +
  coord_flip()
```




```
jfk %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("gray20", "gray80"),
                   max.words = 100)
```


dwindling
negative
discourage
controversy
urgent
objection
discrimination
conflict
concern
aggression
neglected
depressed
absence
unable
slack
problem
recession
lies
risk
vice
peril
crisis
panic
attack
impose
burden
lag
difficult
aggressor
doubt
bad
cold
lack
denial
destroy
problems
powerful
limited
harsh
ease
enough
peace
work
clear
liberty
free
pride
healthy
better
gold
well
honor
secure
best
prompt
freedom
great
support
effectively
available
positive
clearly
efficient

Comparison Overall

*Both presidents use the same number of anger words, but JFK uses more joy and anticipation words than Obama.

*Joy: hope, resources are in both word clouds

*Anger: recession is in both word clouds

*Anticipation: time, hope, efficient, finally, money, are in both word clouds