Trump Speech Analysis

Import the text from all 56 Donald Trump speeches into R. Tokenize the data into a tidy text data frame, using bigrams as tokens.Filter the data, removing bigrams where either word is a stop word or the word "applause", and removing bigrams where the first word is a negation word such as "never", "no", "not", or "without".plot the top 15 most common bigrams in Trump's speeches.

```
text <- read_lines("../hw6/full_speech.txt")

poem <- tibble(line = 1:74,text = text)

tidy_poem <-
  unnest_tokens(poem,bigram, text, token = "ngrams", n = 2 )


trump_bigrams <- tidy_poem %>%
  separate(bigram, c("word1", "word2"), sep = " ")

trump_bigrams_1 <- trump_bigrams %>%
  filter(!word1 %in% c(stop_words$word, "applause", "not", "no","never", "without")) %>%
  filter(!word2 %in% c(stop_words$word, "applause")) %>%
unite(bigram, word1, word2, sep = " ")


trump_bigrams_1 %>%
  count(bigram, sort = TRUE) %>%
  top_n(15) %>% mutate(bigram = reorder(bigram, n)) %>%
  ggplot() + geom_col(aes(bigram, n, fill = n)) + coord_flip() + ggtitle("Top15 trump bigrams")
```
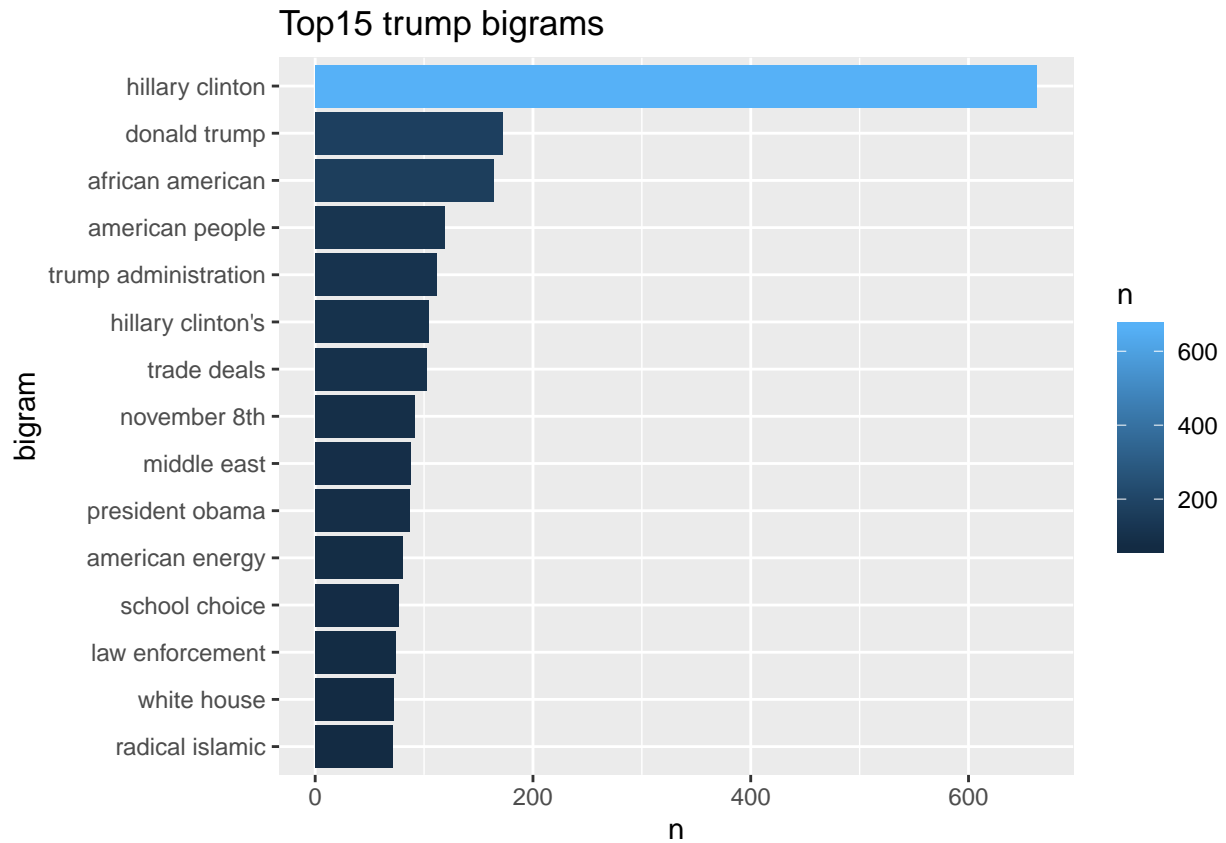
```
## Selecting by n
```

Top15 trump bigrams

Lets see the most commonly negated words in Donald Trump's speeches, and how they're negated. Filter the bigrams, keeping only bigrams where the first word is any of "not", "no", "never", or "without", and removing those where the second word is a stop word or "applause". Then visualize the most common (top ~5) words preceded (separately) by each of "never", "no", "not", and "without".

```
## Selecting by n
```

Top (~5) trump words starting with a negation
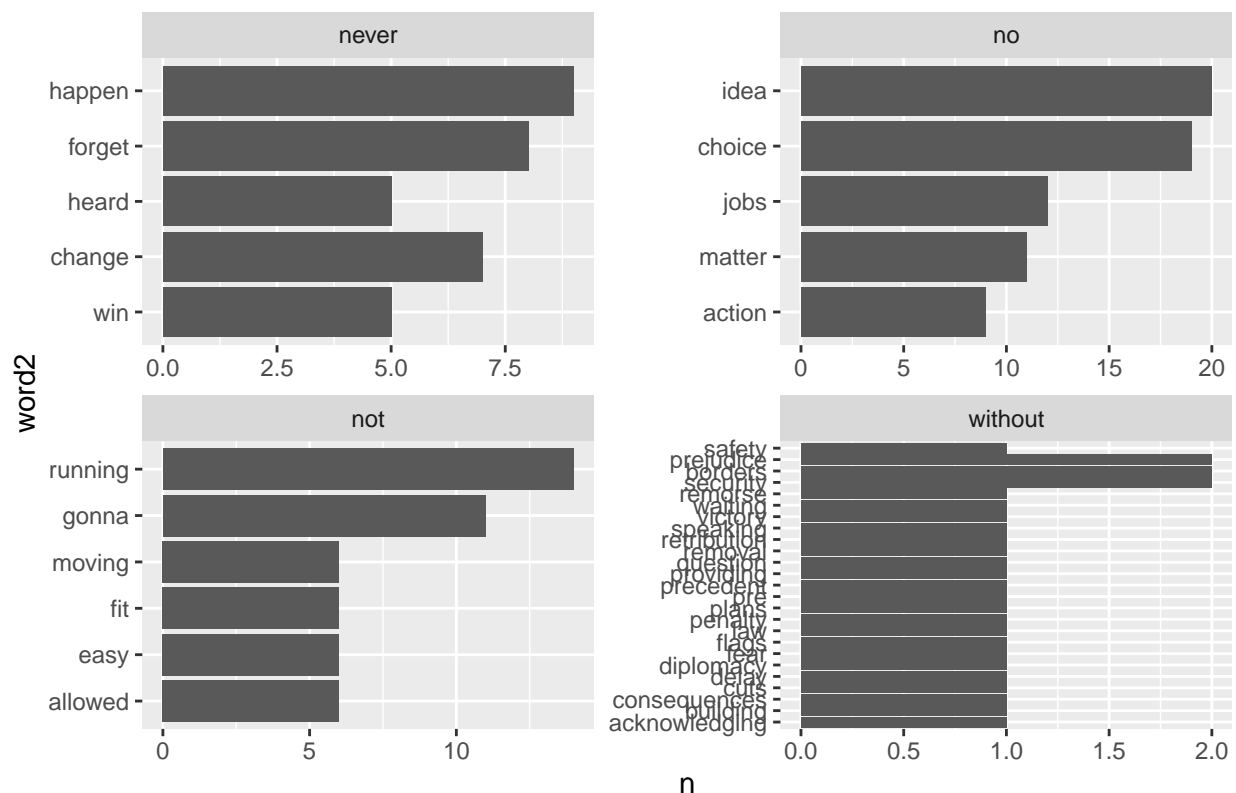
Lets do a sentiment analysis of Donald Trump's speeches. In order to make sure sentiments are assigned to appropriate contexts, first tokenize the speeches into bigrams, and filter out all bigrams where the first word is any of the words "not", "no", "never", or "without". Now consider only the second word of each bigram. Filter out all bigrams where the second word is a stop word or "applause". Then visualize the most common words (top ~5) in Trump's speeches that are associated with each of the 6 sentiments in the "loughran" lexicon.

```
trump_bigrams3 <- trump_bigrams %>%
  filter(!word1 %in% c("not", "no","never", "without")) %>%
  filter(!word2 %in% c(stop_words$word) )%>%
  filter(word2 != "applause")


senti <- trump_bigrams3 %>%
  inner_join(get_sentiments("loughran"), by = c("word2" = "word")) %>%
  count(word2,sentiment, sort = TRUE)




  senti %>%
    arrange(desc(n)) %>%
  mutate(word2 = reorder(word2, n)) %>%
  group_by(sentiment) %>%
    top_n(5) %>%
```

```
ggplot(aes(word2, n, fill = sentiment)) +
geom_col(show.legend = FALSE) +
facet_wrap(~sentiment, scales = "free")+
coord_flip() + ggtitle("Top(~5) words in each sentiment")
```

Selecting by n

Top(~5) words in each sentiment