Presidential Inauguration Speeches

We examine some common facotrs in presidential inauguration speeches since the founding of United States.

Part 0: Install and load libraries

```r
packages.used=c("tm", "dplyr", "tidytext", "ggplot2", "topicmodels")

# check packages that need to be installed.
packages.needed=setdiff(packages.used,
                        intersect(installed.packages()[,1], packages.used))
# install additional packages
if(length(packages.needed)>0){
  install.packages(packages.needed, dependencies = TRUE)
}

# load packages
library(tm)
library(dplyr)
library(tidytext)
library(topicmodels)
library(ggplot2)
source("../lib/TextProcessing.R")
```
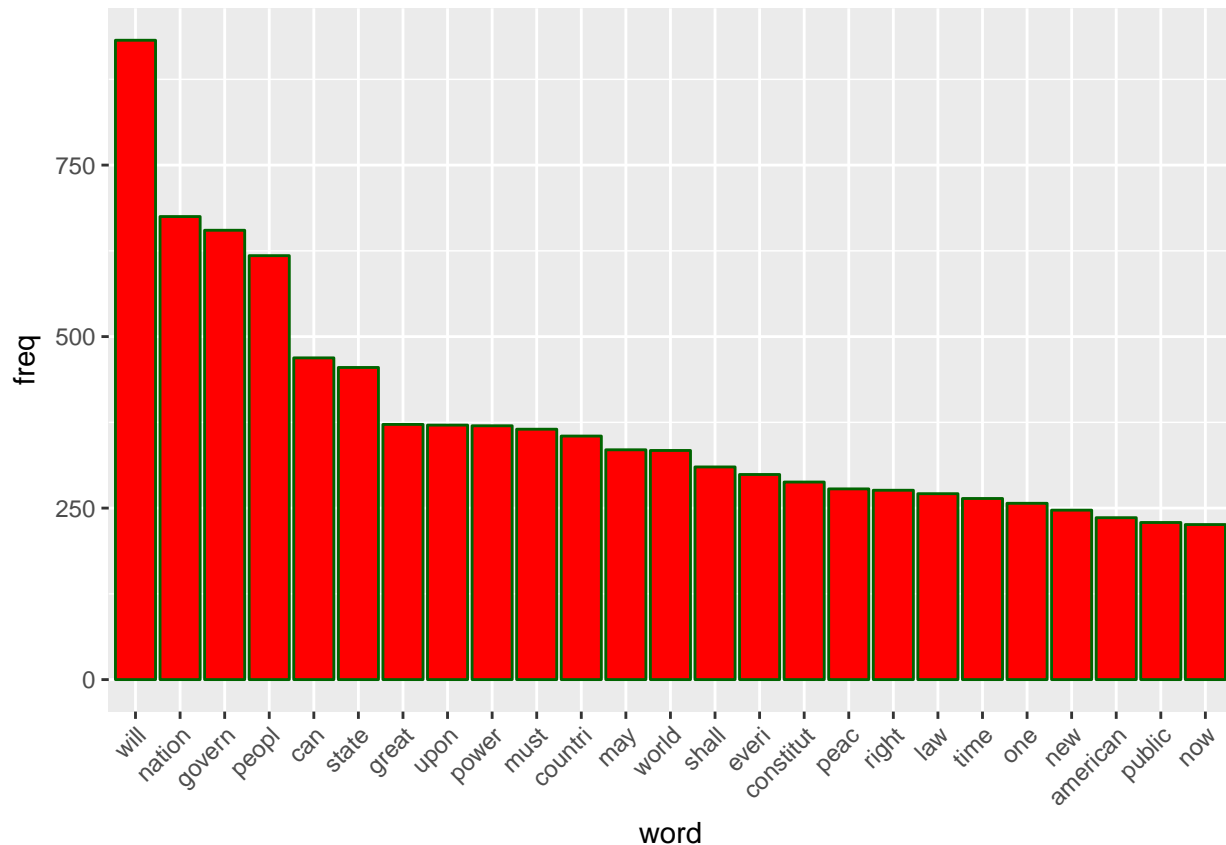
Part 1: Most-used words

We first examine the most frequently used words in presidential inauguration speeches. Below is the code to process all speeches and plot the 25 of the most frequently used words (stemmed) across all speeches. The graph generated is stored in the figs folder.

```r
folder.path="../data/InauguralSpeeches/"
docs <- getSpeeches(folder.path)

# get most frequent words
tdm <- TermDocumentMatrix(docs)
freq <- sort(rowSums(as.matrix(tdm)), decreasing=TRUE)
wf <- data.frame(word=names(freq), freq=freq)
wf$word <- factor(wf$word, levels = wf$word[order(wf$freq, decreasing=TRUE)])

# build histogram
p <- ggplot(subset(wf, freq>=head(freq, 25)[[25]]), aes(x=word, y=freq))
p <- p + geom_bar(stat="identity", fill="red", colour="darkgreen")
p <- p + theme(axis.text.x=element_text(angle=45, hjust=1))
ggsave("../figs/most-used-words.png", plot=p, device="png")
p
```

Not suprisingly, the most used word is "will" ("shall" is also on the list) since inauguration speeches generally lays out a plan-of-attack for the upcoming term of the president. Moreover, we can see that words that are used to convey ideologies like "must" and "every" are also on the list. Finally, words like "constitution" and "people" that define the American government also appear on the list.

Part 2: Topic modeling

We used LDA to model the topics in all inaugural speeches. The LDA model initialized to 3 topics and is seeded with 3 random integers for reproducibility. The generated tables are stored into txt files in the output folder.

```r
folder.path="../data/InauguralSpeeches/"
# empiricially removing words that does not convey info in these termsspeeches
extra.words <- c("will", "can", "may", "upon", "let", "must", "shall",
                 "every", "one", "now", "today", "time", "day", "new",
                 "united", "states", "country", "nation")
docs <- getSpeeches(folder.path, extra.words)
dtm <- DocumentTermMatrix(docs)
lda <- LDA(dtm, k=3,
           method="Gibbs",
           control=list(nstart=3, seed=list(272, 166, 966), delta=0.1,
                        best=TRUE, burnin=0, iter=2000, thin=2000))
topics <- as.matrix(topics(lda))
terms <- as.matrix(terms(lda, 12))
write.table(topics, file = "../output/topic-speech.txt", quote=FALSE, sep = "\t")
write.table(terms, file = "../output/term-topic.txt", quote=FALSE, sep = "\t")
```

```
terms
```

```
##       Topic 1      Topic 2     Topic 3
##  [1,] "govern"     "law"       "world"
##  [2,] "power"      "govern"    "peopl"
##  [3,] "constitut"  "nation"    "america"
##  [4,] "right"      "peopl"     "american"
##  [5,] "great"      "secur"     "freedom"
##  [6,] "interest"   "peac"      "work"
##  [7,] "public"     "civil"     "life"
##  [8,] "union"      "administr" "live"
##  [9,] "peopl"      "polit"     "know"
## [10,] "principl"   "polici"    "peac"
## [11,] "duti"       "question"  "make"
## [12,] "execut"     "purpos"    "hope"
```

The 1st topic contains words like "duty", "public", "people", and "right", suggesting a moral aspect of the topic. The 2nd topic contains words like "law", "civil", "govern", and "policy", suggesting a legal inclination. The 3rd topic contains words like "freedom", "life", "work", "hope", which is typical for the American ideology. There are also some interesting insights from the topic/speech table. For example, Obama and Trump's speeches falls into the same topic. This can be a result of the financial crisis of 2007/2008 and political crisis in the most recent election. As a result, both presidents need to rekindle hope for the people by emphasizing the ideologies of work and life for Americans. Another thing that is interesting to note is that the topics for speeches across terms for the same president are relatively consistent, showing that there is certain level of validity of this model.

Part 3: Some conclusions

Inauguration speeches are not designed to be surprises and they generally cover similar topics with different emphasis. We can see from part 1 that none of the most frequently used words are too out of our expectation. In addition, the topics in part 2 also generally adhere to the American tradition. Since these speeches are generally written with strong intention and thesis, we should not expect anything out of the ordinary from mining them.