Computational Methods: Final Problem Set

Simón Hicks-Prieto

11/21/2019

```
knitr::opts_chunk$set(echo = TRUE)
```

Part 1: Preprocessing/General NLP

Problem 1: Loading Data & Libraries

```
library(tinytex)
library(tm)
library(amerika)
library(wordcloud)
library(wordcloud2)
library(tidyverse)
library(skimr)
library(readtext)
library(readr)
library(textdata)
library(tidyr)
library(tidyverse)
library(tidytext)
library(glue)
library(stringr)
library(dplyr)
library(psych)
library(expss)
library(topicmodels)
library(raster)
library(ggplot2)

#loading platforms csv
platforms <- read.csv(("~/Downloads/problem-set-3-master 2/platforms.csv"))

#loading each party's platform, and creating Corpus (VectorSource)
newd16 <- readtext("~/Downloads/problem-set-3-master 2/Party Platforms Data/d16.txt")
newr16 <- readtext("~/Downloads/problem-set-3-master 4/Party Platforms Data/r16.txt")
View(newr16)
View(newd16)
democratdocs1 <- VCorpus(VectorSource(newd16))
republicandocs1 <- VCorpus(VectorSource(newr16))
```

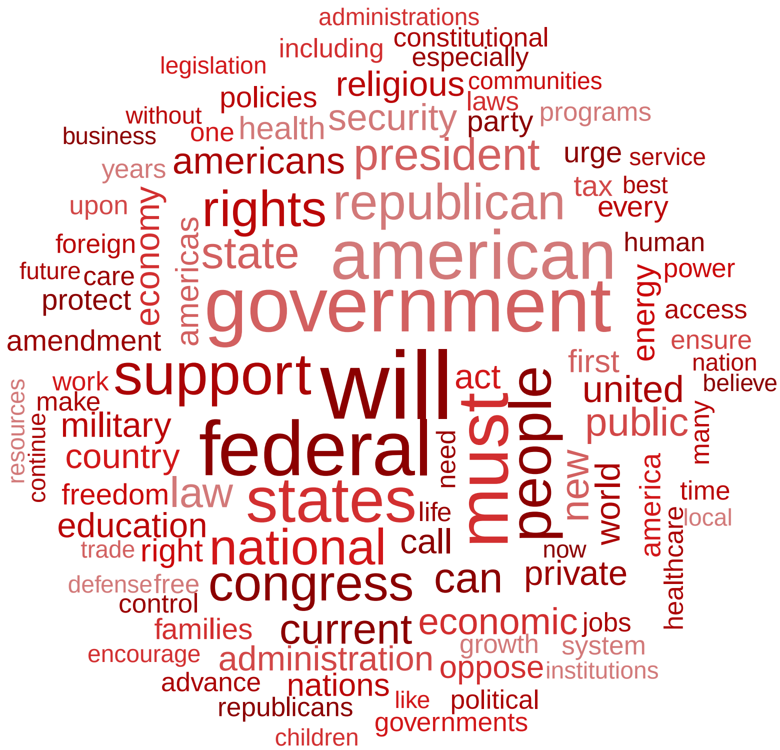
Problem 2: Preprocessing + Creating DTM for Each Party

```
republicandocs1 <- republicandocs1 %>%
  tm_map(stripWhitespace) %>%
  tm_map(removeNumbers) %>%
  tm_map(removePunctuation) %>%
```

```
  tm_map(content_transformer(tolower)) %>%
  tm_map(removeWords, stopwords("english")) %>%
  tm_map(removeWords, c("also"))
republicandocs1 <- tm_map(republicandocs1, PlainTextDocument)


democratdocs1 <- democratdocs1 %>%
  tm_map(stripWhitespace) %>%
  tm_map(removeNumbers) %>%
  tm_map(removePunctuation) %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removeWords, stopwords("english")) %>%
  tm_map(removeWords, c("also"))
democratdocs1 <- tm_map(democratdocs1, PlainTextDocument)


#stemming and storing in new Corpus
stem_democrat <- tm_map(democratdocs1,stemDocument)
stem_democrat<- tm_map(stem_democrat, PlainTextDocument)

stem_republican <- tm_map(republicandocs1, stemDocument)
stem_republican<- tm_map(stem_republican, PlainTextDocument)

dtm_democrat <- DocumentTermMatrix(stem_democrat)
dtm_republican <- DocumentTermMatrix(stem_republican)
```

Problem 3: Word Clouds and Analysis

```
#Republican Party
set.seed(2345)
wordcloud(scale=c(3, .3),
          republicandocs1,
          max.words = 100,
          colors = amerika_palette(n = 10, name = "Republican", type = "continuous"),
          random.color = TRUE,
          random.order=FALSE)
```




```
#Democratic Party
set.seed(2345)
wordcloud(scale=c(5, .35),
          democratdocs1,
          max.words = 100,
          colors = amerika_palette(n = 10, name = "Democrat", type = "continuous"),
          random.color = TRUE,
          random.order=FALSE)
```

At this early stage, the most commonly used words of each party suggest some important similarities and distinctions. For instance, the Democrats use the word "will" the most, a word that implies a commitment or a promise. Unsurprisingly, "will" is also the most frequently used word for the Republican platform. This similarity perhaps reinforces the notion that politicians, regardless of their political affiliation, frequently make promises to their voter bases to shore up support.

This word cloud also reveals notable differences between the parties. The most used word for Republicans, however, is "government", and this is followed closely by "federal". These words alone indicate that the federal government is a highly salient topic for the Republican party. In contrast, the word "government" does not even appear in the Democrats' top 25 words. Instead, "democrats" is the second most frequently used word for the Democratic Party's platform, which implies that the Democrats place themselves as the subject, while Republicans place the federal government as the subject of conversation.

Another interesting note is that the Republicans use the words "must" and "rights" with a significantly higher relative frequency than Democrats. These two words, in contrast to popular words among Democrats such as "support" and "believe", convey different tones– with the former conveying a sense of entitlement, and the latter carrying more aspirational meanings.

Part 2: Sentiment Analysis

Problem 4: Calculating Sentiment Using Bing and Afinn Dictionaries

Problem 4 (A): Democrat Platform Sentiment

```
#Tokenizing
demtokens <- data_frame(text = as.character(democratdocs1)) %>%
  unnest_tokens(word, text)
```

```
## Warning: `data_frame()` is deprecated, use `tibble()`.
## This warning is displayed once per session.
```

```
#Bing Sentiment Calculations
bing_demtokens <- demtokens %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE)
```

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

```
bing_demtokens
```

```
## # A tibble: 607 x 3
##    word       sentiment     n
##    <chr>      <chr>     <int>
##  1 support    positive    123
##  2 work       positive     72
##  3 protect    positive     46
##  4 right      positive     37
##  5 clean      positive     33
##  6 affordable positive     27
##  7 well       positive     25
##  8 strong     positive     24
##  9 trump      positive     24
## 10 better     positive     21
## # ... with 597 more rows
```

```
#Afinn Sentiment Calculations
demtokens %>%
  inner_join(get_sentiments("afinn"))
```

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

```
## # A tibble: 2,315 x 2
##    word      value
```

```
##    <chr>      <dbl>
##  1 thanks         2
##  2 hard          -1
##  3 great          3
##  4 recession     -2
##  5 growth         2
##  6 positive       2
##  7 gained         2
##  8 best           3
##  9 wealth         3
## 10 dream          1
## # ... with 2,305 more rows
```

```
#Total Sentiment Calculations
demtokens %>%
  inner_join(get_sentiments("bing")) %>%
  count(sentiment)
```

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

```
## # A tibble: 2 x 2
##   sentiment     n
##   <chr>     <int>
## 1 negative    811
## 2 positive   1372
```

Problem 4 (B): Repubican Platform Sentiment

```
#Tokenizing
reptokens <- data_frame(text = as.character(republicandocs1)) %>%
  unnest_tokens(word, text)

#Bing Sentiment Calculations
bing_reptokens <- reptokens %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE)
```

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

```
bing_reptokens
```

```
## # A tibble: 898 x 3
##    word      sentiment     n
##    <chr>     <chr>     <int>
##  1 support   positive    100
##  2 right     positive     46
##  3 oppose    negative     43
##  4 freedom   positive     42
##  5 protect   positive     38
##  6 free      positive     37
##  7 work      positive     37
##  8 encourage positive     30
##  9 best      positive     29
## 10 like      positive     28
## # ... with 888 more rows
```

```r
#Afinn Sentiment Calculations
reptokens %>%
  inner_join(get_sentiments("afinn"))
```

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

## # A tibble: 2,652 x 2
##    word      value
##    <chr>     <dbl>
##  1 united        1
##  2 defender      2
##  3 happiness     3
##  4 limited      -1
##  5 freedom       2
##  6 freedom       2
##  7 freedom       2
##  8 freedom       2
##  9 peril        -2
## 10 united        1
## # ... with 2,642 more rows
```

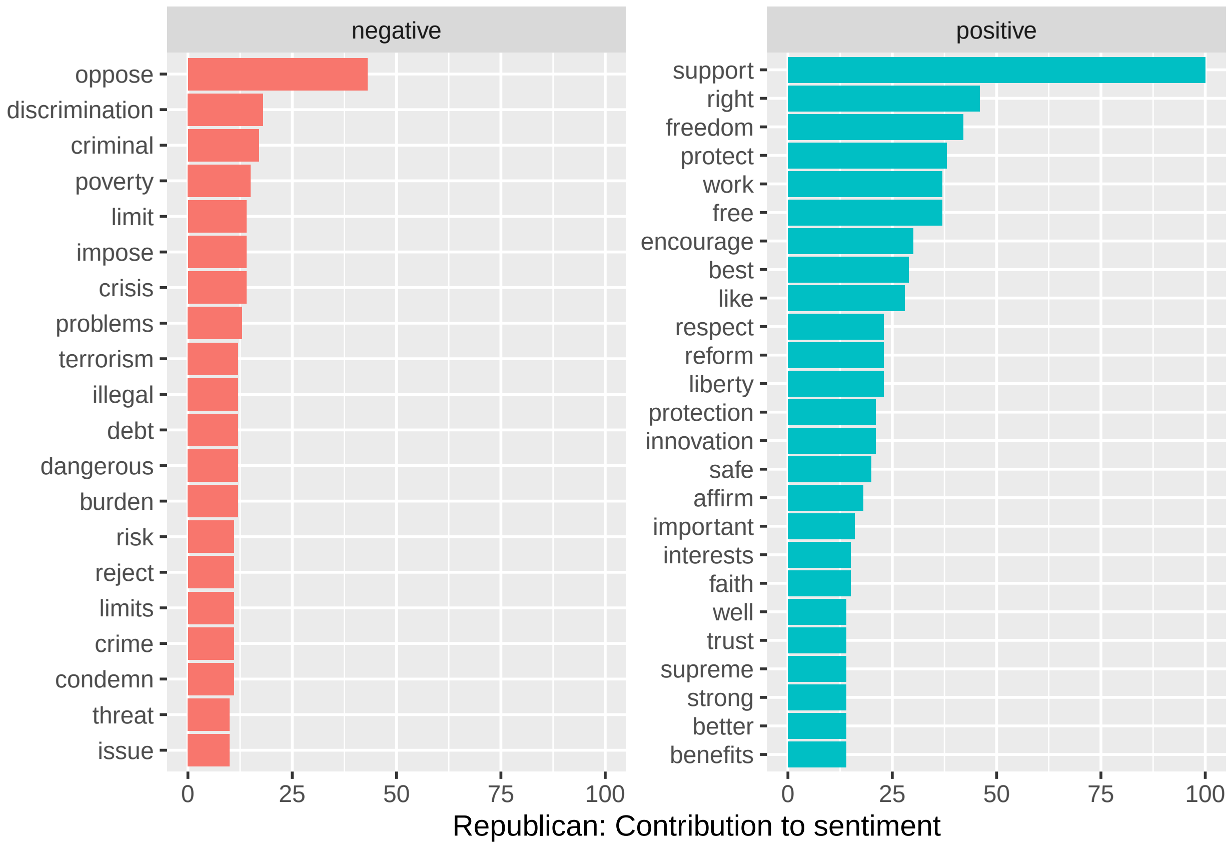
```r
#Total Sentiment Calculations
reptokens %>%
  inner_join(get_sentiments("bing")) %>%
  count(sentiment)
```

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

## # A tibble: 2 x 2
##   sentiment     n
##   <chr>     <int>
## 1 negative   1245
## 2 positive   1577
```

Problem 4 (C): Visual Comparison of Both Party's Sentiments

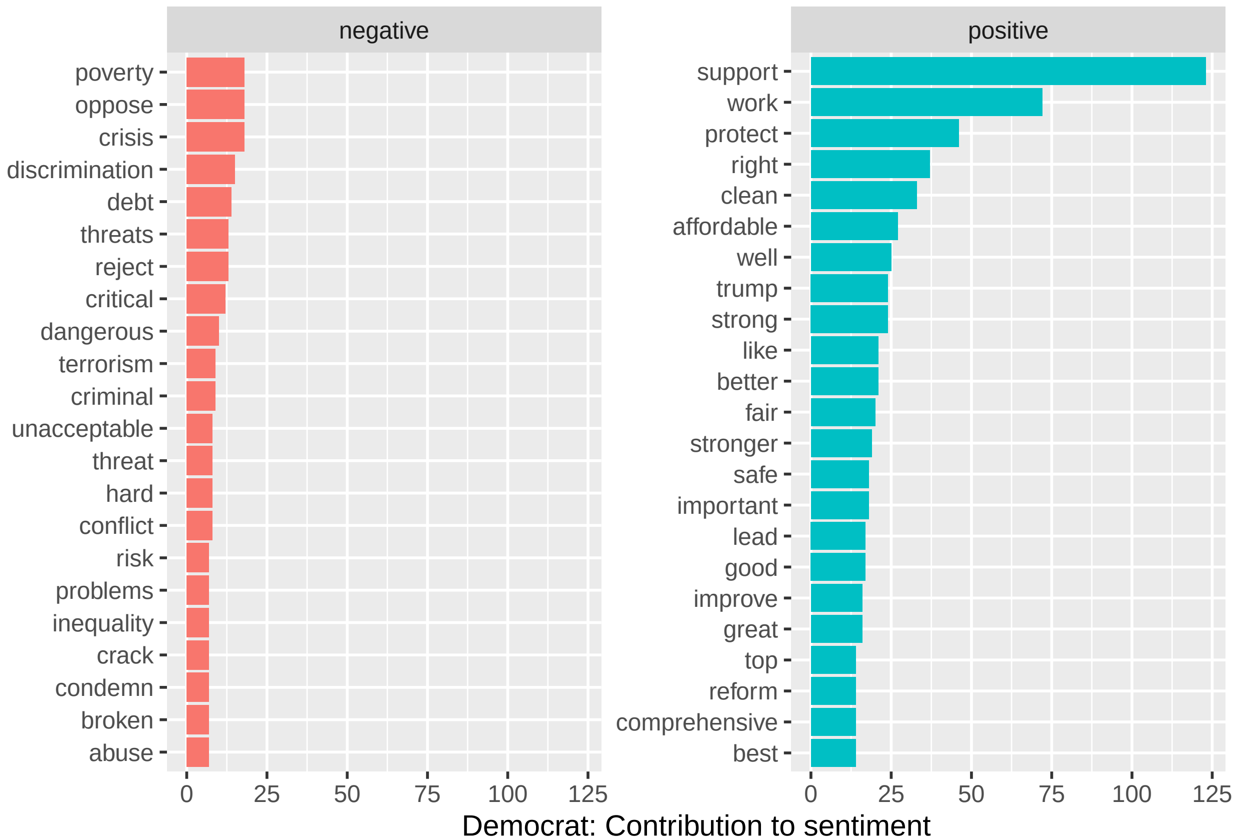
```r
bing_reptokens %>%
  group_by(sentiment) %>%
  top_n(20) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~sentiment, scales = "free_y") +
  labs(y = "Republican: Contribution to sentiment",
       x = NULL) +
  coord_flip()
```

```
## Selecting by n
```




```
bing_demtokens %>%
  group_by(sentiment) %>%
  top_n(20) %>%
  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 = "Democrat: Contribution to sentiment",
       x = NULL) +
  coord_flip()
```

```
## Selecting by n
```



Problem 5: Analyzing Differences & Similarities

Overall, the Republicans appear to be more negative in their platform than the Democrats, as negatively connotated words account for approximately 44% of the Republican platform (1245/2822), yet only 37% of words (811/2183) in the Democratic platform are negative.

In regards to how each country perceives the future, both parties appear to be very similar in their outlook. The two graphs above, which outline the 20 most used positive and negative words for each party, are nearly identical. Indeed, the top 3 most frequently used positive words for the Democratic party are also the most frequently used positive words for the Republican party. The same is true for the top 3 most frequently used negative words. This finding suggests that both parties may be discussing the future of the country and the problems that the people face today with a similar level of optimism.

However, despite the significant overlap between each party's most frequently used positive and negative words, there are some distinctions that are worth noting. For instance, the terms "reform" and "comprehensive" are the 18th and 19th most frequently used positive words for the Democratic party, respectively. These words imply structural change, progress, and growth. The Republican party, on the other hand, has neither of these words in their top 20. From this small distinction, a possible implication is that the Democrats are generally more ambitious in the kind of change that they want to see in the country.

Since my perception of each party is that the Democratic party has been one of progressive reform and the Republican party has generally stood for increasing limitations on the government, these results comport with my understanding of these parties.

Part 2: Topic Models

Problem 6: Topic Models at K=5 & Bar Graphs

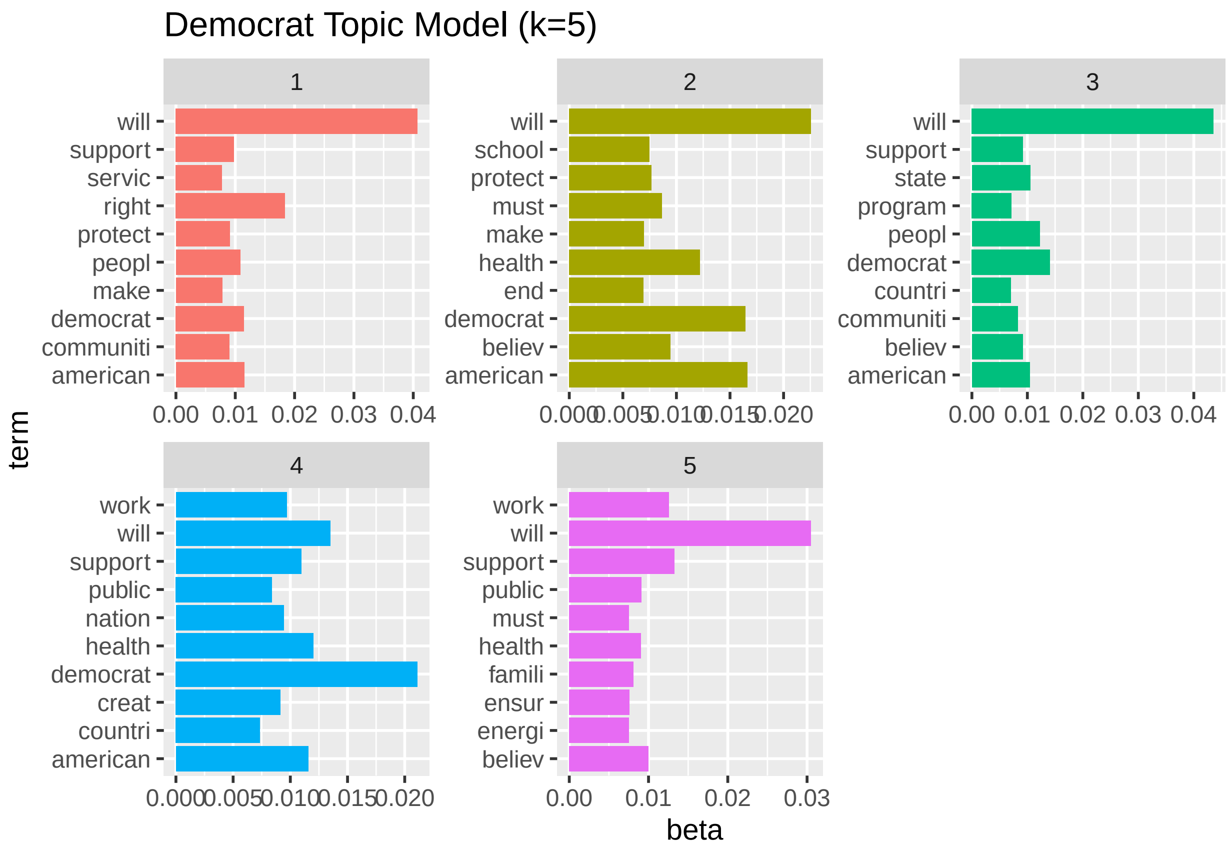
```
dem_lda <- LDA(dtm_democrat, k = 5, control = list(seed = 1234))
dem_lda
```

```
## A LDA_VEM topic model with 5 topics.
```

```
cleandem_lda <- tidy(dem_lda, matrix = "beta")
cleandem_lda
```

```
## # A tibble: 13,575 x 3
##    topic term          beta
##    <int> <chr>        <dbl>
##  1     1 -whether 0.0000907
##  2     2 -whether 0.0000232
##  3     3 -whether 0.0000412
##  4     4 -whether 0.0000880
##  5     5 -whether 0.0000704
##  6     1 -without 0.0000596
##  7     2 -without 0.0000714
##  8     3 -without 0.0000854
##  9     4 -without 0.0000186
## 10     5 -without 0.000107
## # ... with 13,565 more rows
```

```
demtopten_5 <- cleandem_lda %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

demtopten_5 %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  ggtitle("Democrat Topic Model (k=5)") +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```




```r
rep_lda <- LDA(dtm_republican, k = 5, control = list(seed = 1234))
rep_lda
```

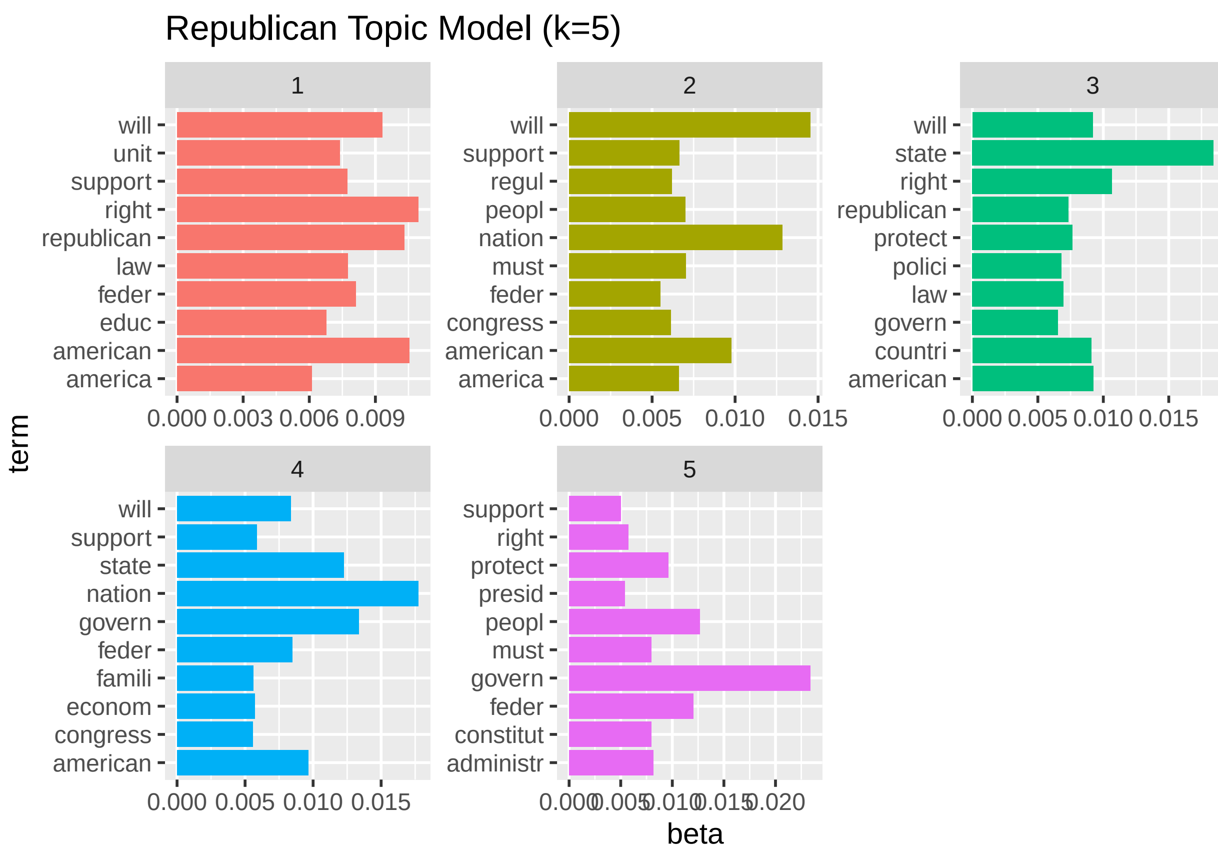
```
## A LDA_VEM topic model with 5 topics.
```

```r
cleanrep_lda <- tidy(rep_lda, matrix = "beta")
cleanrep_lda
```

```
## # A tibble: 17,375 x 3
##    topic term            beta
##    <int> <chr>          <dbl>
##  1     1 -activ   0.0000812
##  2     2 -activ   0.00000207
##  3     3 -activ   0.0000276
##  4     4 -activ   0.0000954
##  5     5 -activ   0.0000604
##  6     1 -billion 0.0000407
##  7     2 -billion 0.0000815
##  8     3 -billion 0.0000405
##  9     4 -billion 0.0000309
## 10     5 -billion 0.0000558
## # ... with 17,365 more rows
```

```r
reptopten_5 <- cleanrep_lda %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)
```

```r
reptopten_5 %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  ggtitle("Republican Topic Model (k=5)") +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```



Problem 7: Analysis of Trends at K=5

When comparing the Democrat and Republican Topic Models at k=5, there are some notable differences between the policy areas that each party highlights in its platform. For instance, for the Democratic Party, the combination of the words "public", "support", and "health" are in 3 out of their 5 topic models. In contrast, "health" does not appear once in the topic models of the Republican Party. This indicates that health (and more specifically, the government's role in ensuring the health of the public) are policy areas that are crucial questions for Democrats yet are not central talking points for Republicans.

Another interesting example is that the combination of terms "protect", "nation/people", "government" are found in many of the topic models of the Republican party. This suggests that a highly salient issue for the Republican party is ensuring that the government does not infringe on the rights of the people (ie. they appear determined to "protect" people from these infringements).

Lastly, the Republican party's frequent use of "nation", and the Democratic party's regular use of "community" may reveal an important distinction between the two platforms. For Republicans, there is not a single use of the word "community" in these 5 models, while the Democratic party uses it in the majority of their models. In contrast, the Democratic party never uses the word "nation" in these 5 models, while this word appears on multiple models of the Republican party. This finding indicates that Democrats have a more community-based approach when communicating with the American people, while the Republican party may tend to be more nationalistic in their communication with voters.

Problem 8: Topic Models at K=10 & K=25

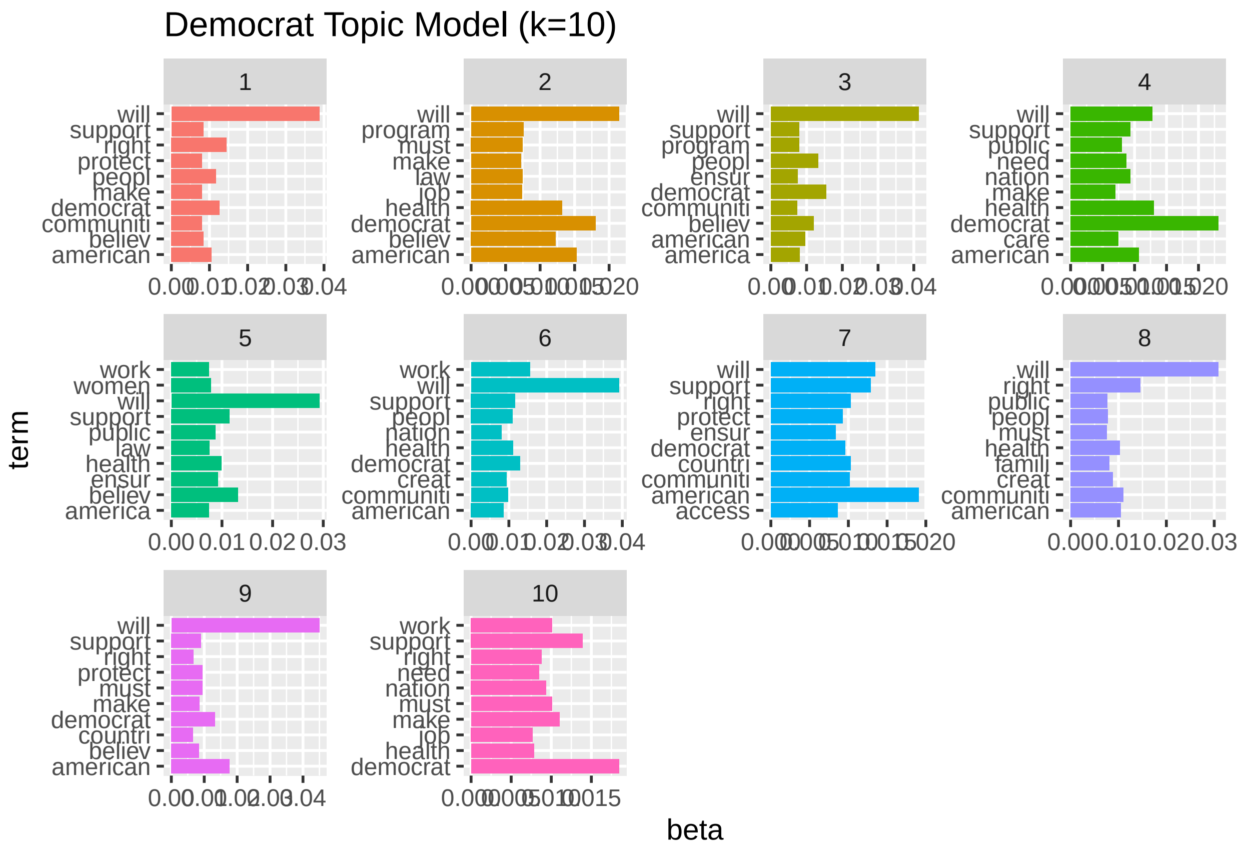
```r
dem_lda10 <- LDA(dtm_democrat, k = 10, control = list(seed = 1234))
dem_lda10
```

```
## A LDA_VEM topic model with 10 topics.
```

```r
cleandem_lda10 <- tidy(dem_lda10, matrix = "beta")
cleandem_lda10
```

```
## # A tibble: 27,150 x 3
##    topic term          beta
##    <int> <chr>        <dbl>
##  1     1 -whether 0.0000675
##  2     2 -whether 0.0000172
##  3     3 -whether 0.0000306
##  4     4 -whether 0.0000653
##  5     5 -whether 0.0000531
##  6     6 -whether 0.0000519
##  7     7 -whether 0.0000807
##  8     8 -whether 0.0000765
##  9     9 -whether 0.000113
## 10    10 -whether 0.0000977
## # ... with 27,140 more rows
```

```r
demtopten_10 <- cleandem_lda10 %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

demtopten_10 %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  ggtitle("Democrat Topic Model (k=10)") +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```




```
rep_lda10 <- LDA(dtm_republican, k = 10, control = list(seed = 1234))
rep_lda10
```

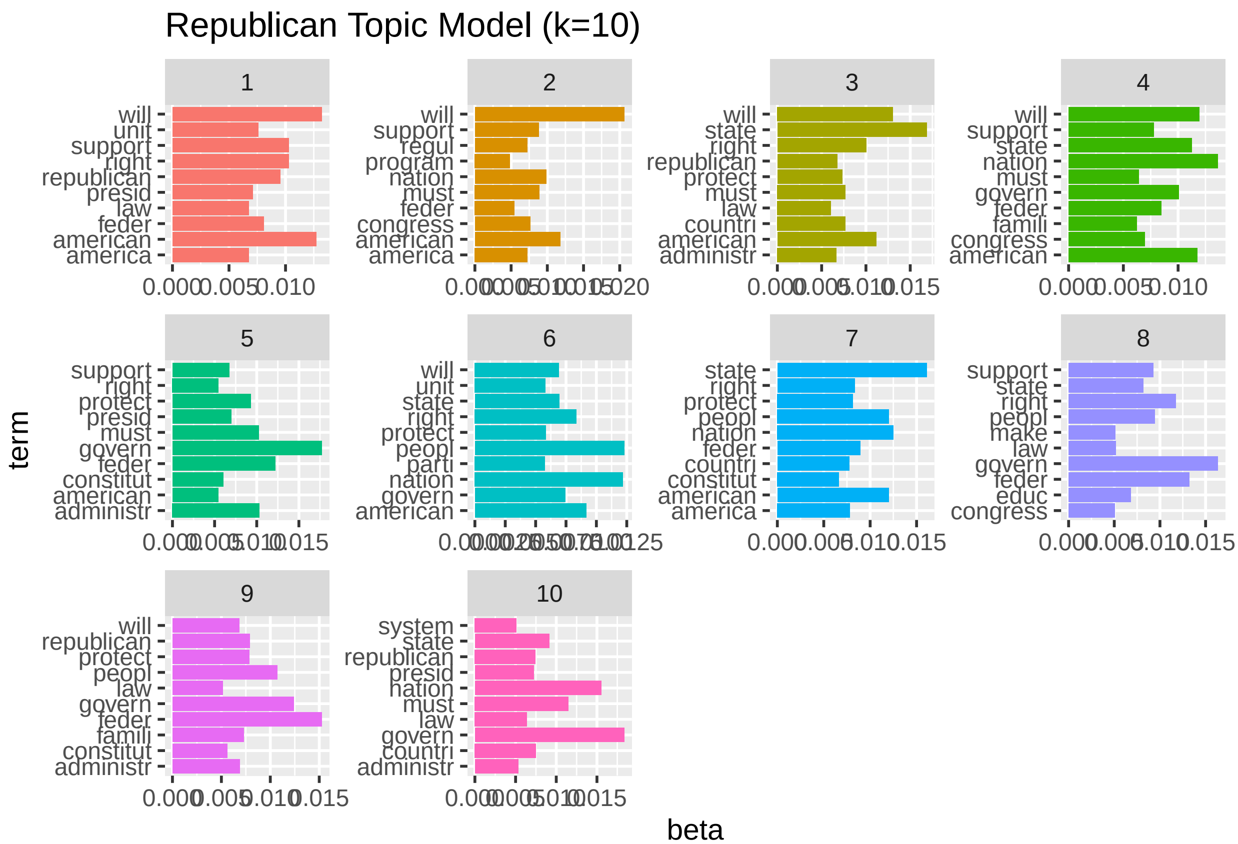
```
## A LDA_VEM topic model with 10 topics.
```

```
cleanrep_lda10 <- tidy(rep_lda10, matrix = "beta")
cleanrep_lda10
```

```
## # A tibble: 34,750 x 3
##    topic term        beta
##    <int> <chr>      <dbl>
##  1     1 -activ 0.0000534
##  2     2 -activ 0.00000136
##  3     3 -activ 0.0000182
##  4     4 -activ 0.0000633
##  5     5 -activ 0.0000407
##  6     6 -activ 0.0000494
##  7     7 -activ 0.000103
##  8     8 -activ 0.0000555
##  9     9 -activ 0.0000948
## 10    10 -activ 0.0000471
## # ... with 34,740 more rows
```

```
reptopten_10 <- cleanrep_lda10 %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)
```

```r
reptopten_10 %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  ggtitle("Republican Topic Model (k=10)") +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```




```r
dem_lda25 <- LDA(dtm_democrat, k = 25, control = list(seed = 1234))
dem_lda25
```

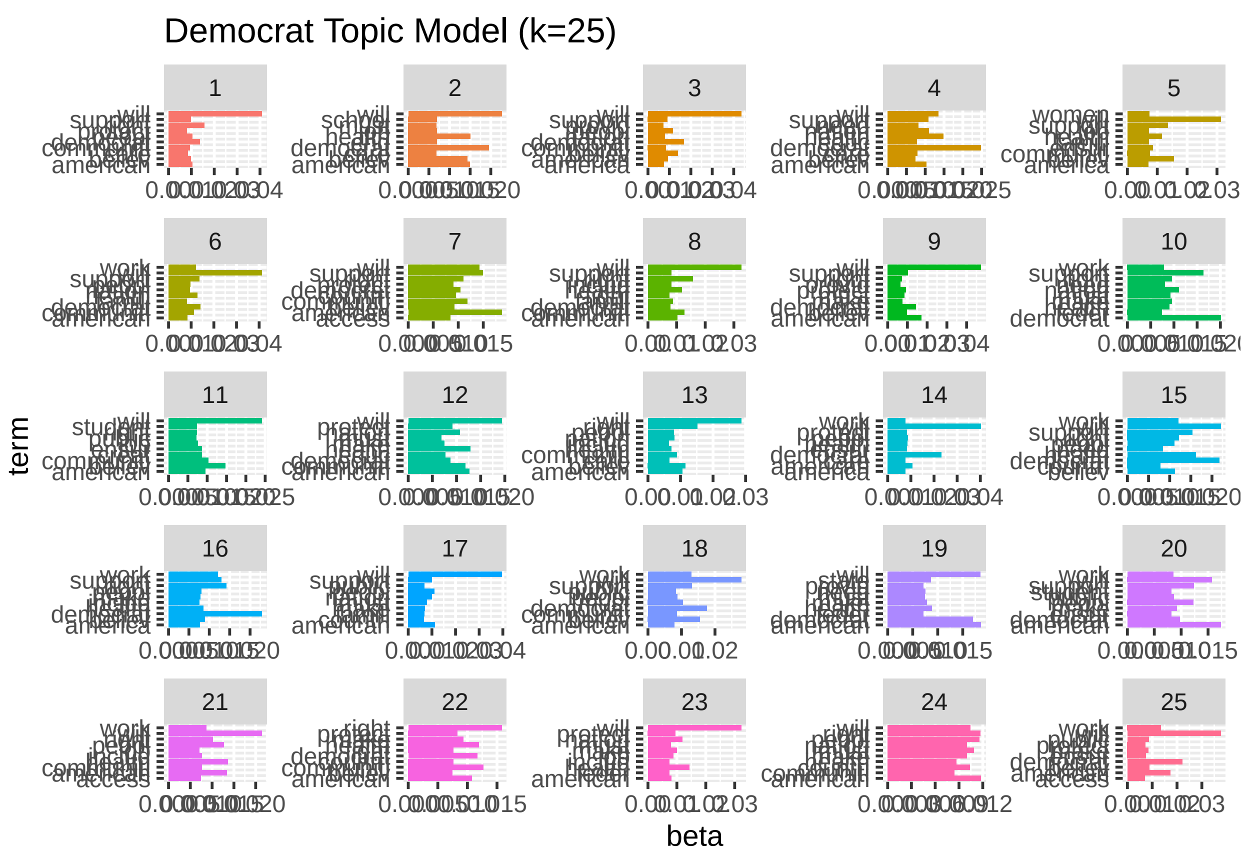
```
## A LDA_VEM topic model with 25 topics.
```

```r
cleandem_lda25 <- tidy(dem_lda25, matrix = "beta")
cleandem_lda25
```

```
## # A tibble: 67,875 x 3
##    topic term          beta
##    <int> <chr>        <dbl>
##  1     1 -whether 0.0000846
##  2     2 -whether 0.0000217
##  3     3 -whether 0.0000384
##  4     4 -whether 0.0000816
##  5     5 -whether 0.0000672
##  6     6 -whether 0.0000650
##  7     7 -whether 0.000102
##  8     8 -whether 0.0000966
##  9     9 -whether 0.000141
## 10    10 -whether 0.000124
```

```
## # ... with 67,865 more rows
```

```
demtopten_25 <- cleandem_lda25 %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

demtopten_25 %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  ggtitle("Democrat Topic Model (k=25)") +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```




```
rep_lda25 <- LDA(dtm_republican, k = 25, control = list(seed = 1234))
rep_lda25
```

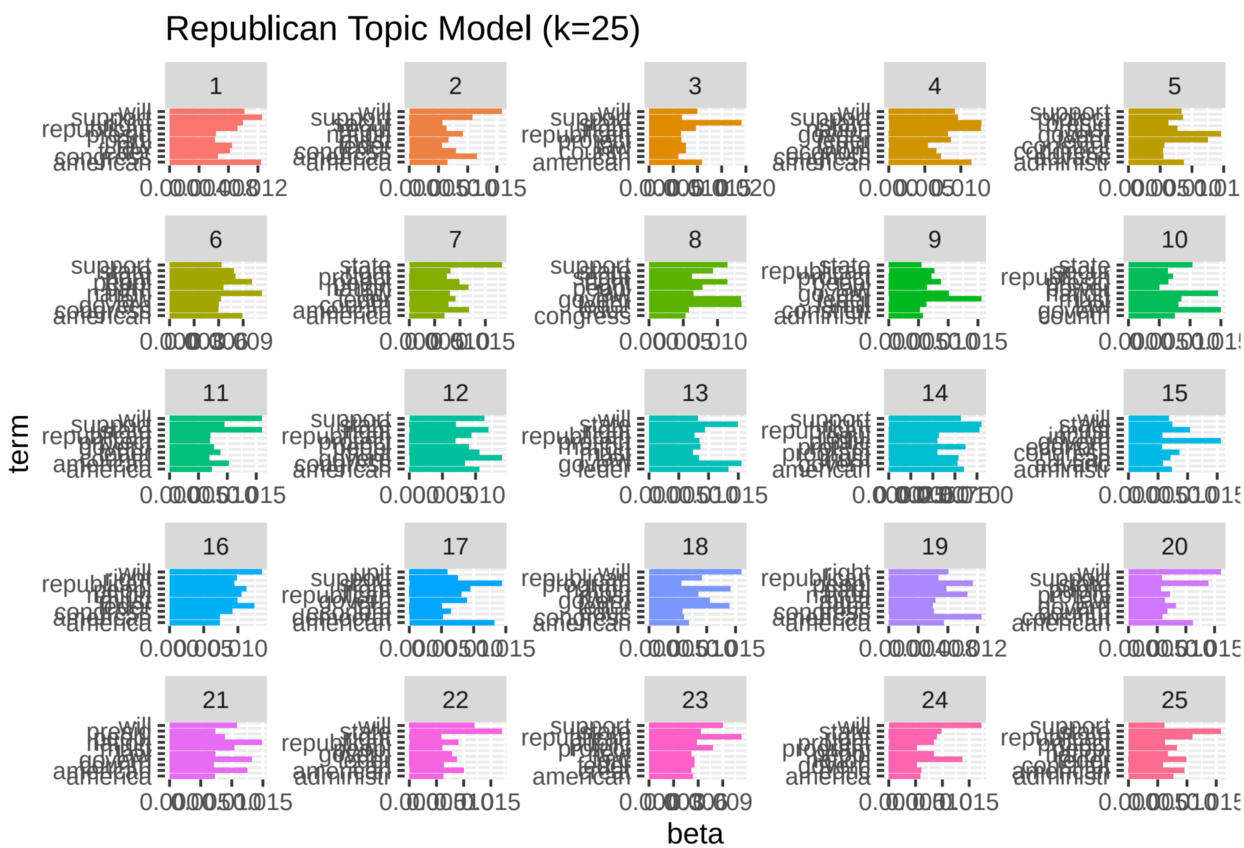
```
## A LDA_VEM topic model with 25 topics.
```

```
cleanrep_lda25 <- tidy(rep_lda25, matrix = "beta")
cleanrep_lda25
```

```
## # A tibble: 86,875 x 3
##    topic term         beta
##    <int> <chr>       <dbl>
##  1     1 -activ 0.0000462
##  2     2 -activ 0.00000119
##  3     3 -activ 0.0000157
```

```
## 4      4 -activ 0.0000551
## 5      5 -activ 0.0000355
## 6      6 -activ 0.0000426
## 7      7 -activ 0.0000892
## 8      8 -activ 0.0000485
## 9      9 -activ 0.0000827
## 10    10 -activ 0.0000410
## # ... with 86,865 more rows
```

```r
reptopten_25 <- cleanrep_lda25 %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)

reptopten_25 %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  ggtitle("Republican Topic Model (k=25)") +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```



Problem 9: Perplexity Calculations

```r
#re-loading data -- changing from vectorsource to dirsource:
d.texts <- file.path("~", "Downloads", "problem-set-3-master 2", "Party Platforms Data")
dir(d.texts)
```

```
## [1] "d16.txt"
```

```r
democratdocs <- VCorpus(DirSource(d.texts))

r.texts <- file.path("~", "Downloads", "problem-set-3-master 4", "Party Platforms Data")
dir(r.texts)
```

```
## [1] "r16.txt"
```

```r
republicandocs <- VCorpus(DirSource(r.texts))

#preprocessing
republicandocs <- republicandocs %>%
  tm_map(stripWhitespace) %>%
  tm_map(removeNumbers) %>%
  tm_map(removePunctuation) %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removeWords, stopwords("english")) %>%
  tm_map(removeWords, c("also"))
republicandocs <- tm_map(republicandocs, PlainTextDocument)

democratdocs <- democratdocs %>%
  tm_map(stripWhitespace) %>%
  tm_map(removeNumbers) %>%
  tm_map(removePunctuation) %>%
  tm_map(content_transformer(tolower)) %>%
  tm_map(removeWords, stopwords("english")) %>%
  tm_map(removeWords, c("also"))
democratdocs <- tm_map(democratdocs, PlainTextDocument)

#re-stemming and converting to DTM:
newstem_democrat <- tm_map(democratdocs,stemDocument)
newstem_democrat<- tm_map(newstem_democrat, PlainTextDocument)

newstem_republican <- tm_map(republicandocs, stemDocument)
newstem_republican<- tm_map(newstem_republican, PlainTextDocument)

newdtm_democrat <- DocumentTermMatrix(newstem_democrat)
newdtm_republican <- DocumentTermMatrix(newstem_republican)

#calculating perplexity:
newdem_lda <- LDA(newdtm_democrat, k = 5, control = list(seed = 1234))
newdem_lda10 <- LDA(newdtm_democrat, k = 10, control = list(seed = 1234))
newdem_lda25 <- LDA(newdtm_democrat, k = 25, control = list(seed = 1234))

perplexity(newdem_lda)
```

```
## [1] 988.5193
```

```r
perplexity(newdem_lda10)
```

```
## [1] 989.3493
```

```r
perplexity(newdem_lda25)
```

```
## [1] 992.3199
```

```
newrep_lda <- LDA(newdtm_republican, k = 5, control = list(seed = 1234))
newrep_lda10 <- LDA(newdtm_republican, k = 10, control = list(seed = 1234))
newrep_lda25 <- LDA(newdtm_republican, k = 25, control = list(seed = 1234))

perplexity(newrep_lda)
```

```
## [1] 1370.349
```

```
perplexity(newrep_lda10)
```

```
## [1] 1370.672
```

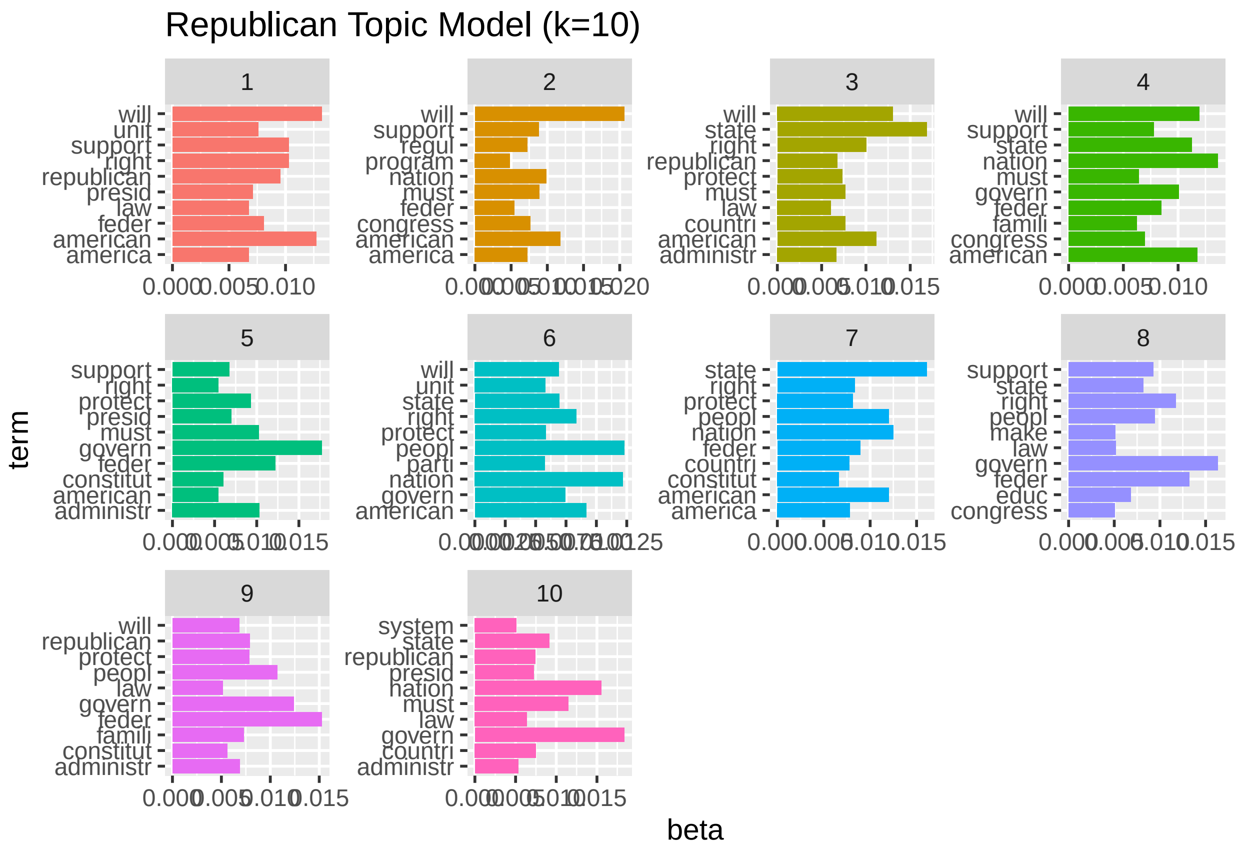
```
perplexity(newrep_lda25)
```

```
## [1] 1374.588
```

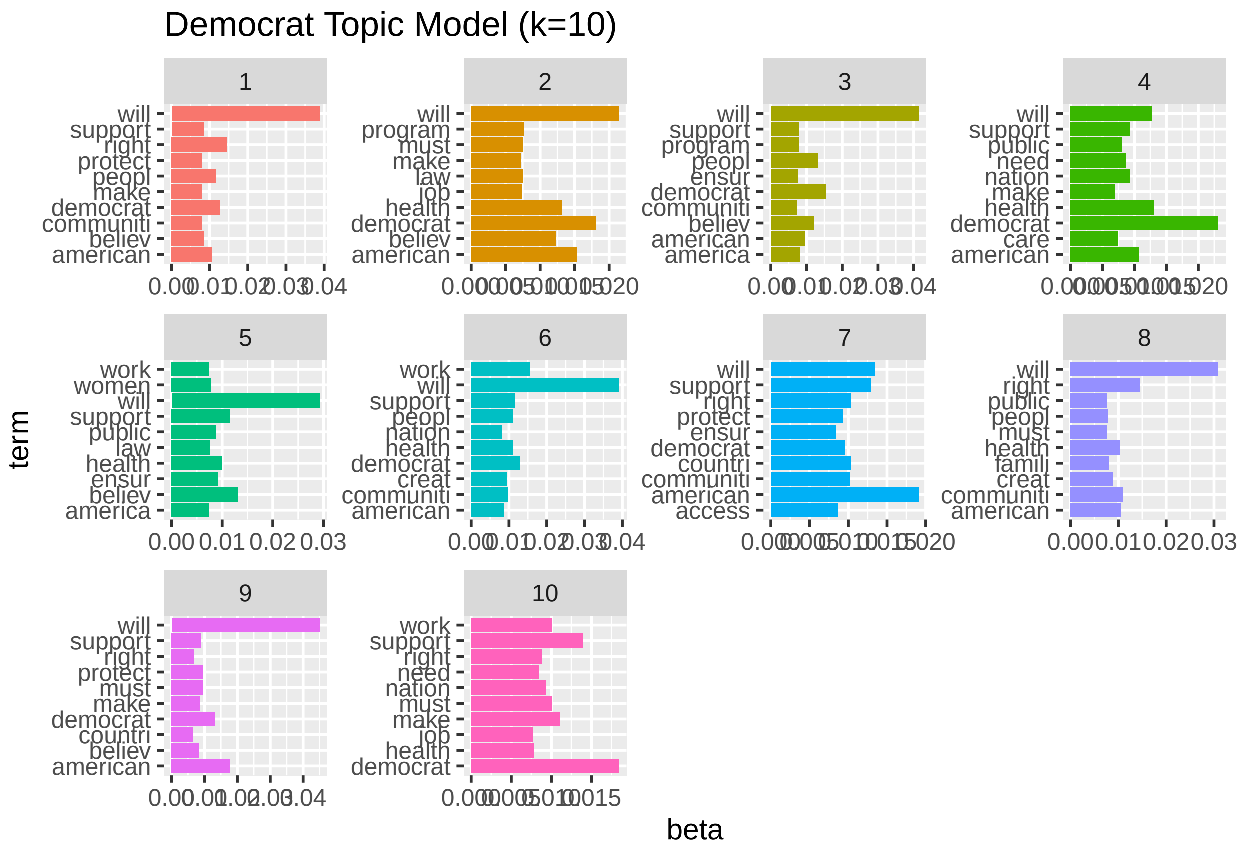
For both Democrats and Republicans, the topic model at k=5 fits slightly better than the other options. For Democrats, the 988.5193 score is lower than the next best topic model of k=10 (989.3493). The same is true for Republicans, as the perplexity score for their topic model at k=5 is approximately 1.2 lower than the next best topic model of k=10. Ultimately, the difference in fit does not appear to be dramatic between the different model iterations, yet in both parties, k=5 is the iteration that technically fits the best.

Problem 10: Barplot at K=10 & Discussion

```
reptopten_10 %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  ggtitle("Republican Topic Model (k=10)") +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```




```
demtopten_10 %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  ggtitle("Democrat Topic Model (k=10)") +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```



For Republicans, a trend that emerges when k=10 is the frequent use of "state". While this trend was present at k=5, seeing that 6, 7, 8, and 10 all contain the word state cements this pattern. Interestingly, the term state is often found alongside the word "right", which suggests that the topic of state's rights is highly salient for the Republican party. This trend, in combination with the trend outlined earlier of "protection" from "federal government", now clearly indicates that the Republican party is strongly pro-state when it comes to the state vs. federal government debate.

In addition, a strong emphasis on the constitution represents an emerging trend when these topic models at set to k=10, as 3 of the last 6 topic models contain the word "constitut-". This finding suggests that the Republican party is adamant about adhering to the constitution, and may often refer to this legal document to bolster their argument of protecting the public from the federal government's infringement.

Another trend that becomes more clear at topic models run at k=10 is that Democrats are constantly talking about themselves. Out of the 10 topic models of the Democratic party, a whopping 8 of them included the term "Democrat". In comparison, the Republican topic models only refer to themselves in 4 out of the 10 models. This demonstrates an inclination on behalf of the Democrats to discuss their achievements and record when communicating with voters.

While the Democrats place themselves at the center of their conversations, Republicans often place the administration. While the Democrats do not reference the president or administration directly in any of their 10 models, the Republican party uses either "administr-" or "presi-" 7 times. Given that these manifestos were being written during a period in which Obama soon stepping out of office, the strong focus that the Republican party placed on the executive office was likely a reflection of their attempts to point out Obama's perceived flaws and failures, so that these ideas were in the voters mind when it came time for the November election. Given the extent to which new trends emerged and previous trends were reinforced, it does appear that the topic models of each party at k=10 more efficiently distinguishes the parties from one another.

Conclusion

Problem 11: Which Party These Results Would Lead Me to Support

This multi-pronged approach has led me to the following conclusions for Republicans: first, the word cloud of the most frequently used words of the Republican platform show that there is a very strong focus on the role of the federal government, and frequently words such as "protect" suggest that limiting the power of the government is one of the party's priorities; second, the topic models reinforce my previous point, as many of these topic models appear to involve protecting the American people from the federal government's reach and advocating for state rights; third, the topic models also reveal that Republicans tend to communicate in a more nationalistic way, as their key policy areas often invoke the term "nation".

For the Democratic party, these are my key takeaways: (1) the word cloud reveals that "health" is a highly emphasized talking point, as is public "support"; (2) the topic models reinforce my previous point, since public support and health programs appear to be policy areas that are central to the Democratic platform; (3)the word cloud also shows that the Democrats tend to be ambitious and aspirational in their beliefs, since the word "believe" is very frequently used; (4) despite both parties generally having similar sentiments in their platform, the graph shows that terms like "comprehensive" and "reform" are often used, which supports the previous point that Democrats are advocating for structural change and progress to a degree that is not seen in the Republican party's platform.

Given these results, I can firmly say that I would vote for the Democratic party in the 2020 election cycle. Even though I am putting my political affiliation aside in this project, I am assuming that in this hypothetical world I would still carry the same core values. For me, some of these core values would be that: a) I believe in dismantling the centuries of laws and regulations that engendered—and still maintain—white supremacy, with the federal government leading the fight. b) I believe that one's race or ethnicity should not be strong determinants of one's long term health and the health of one's family. c) Perhaps most relevantly to this discussion, I believe that progress on the issues listed above should be made at a national level, so that living in a certain state does not exclude you from taking part in this change.

Such ideals are not compatible with those that appear in our analysis of the Republican Party, which the results suggest has a somewhat nationalistic party, slightly gloomier future outlook, and aims to limit the role of the federal government in favor of state rights.

In contrast, these core values along with the results acquired from our statistical analysis, it's apparent that I should support the Democratic party since this output of our analyses demonstrate that they advocate for comprehensive reform, prioritize health, and believe that the government should have a significant role in ensuring the public receives support.
