Text as Data Problem Set 2

Ma Adelle Gia Arbo

18 October 2022

Contents

I. Research question 1
II. Data collection and processing 1
III. Topic modeling 3
1. Latent Dirichlet Allocation 3
2. Structural Topic Modeling 12
V. Conclusion 25

I. Research question

The objective of this report is to answer the research question, "How did the intentions, motives, and views between Democrats and Republicans evolve in the United States using electoral manifestos in 2012, 2016, and 2020?"

Specifically, this report wants to answer the following questions: 1. What are the prevailing topics in the United States electoral manifestos between the parties across the years? 2. What are the differences and similarities of the prevailing topics between the parties across the years?

These questions can be answered using topic modeling, an unsupervised machine learning method for classifying a set of documents, detecting words and phrases, and clustering word groups to best describe a set of documents.

To answer the research question, two types of topic modeling methods will be used. First is the Latent Dirichlet Allocation (LDA), treats each document as a mixture of topics and each topic as a mixture of words. Structural topic modeling will also be implemented which allows correlations between topics. The results of each method will be compared to each other to assess whether using one approach is better than the other for this task.

With topic modeling, the prevailing and dominant topics between Democrats and Republicans can be observed over time. The differences and similarities between the parties' views can also be identified.

II. Data collection and processing

For this topic modeling task, the raw dataset used is the United States manifestos obtained from the WZB. The final dataset used consists of 12,121 documents between the Democratic Party and Republican Party across last three electoral years (2012, 2016, 2020).

```r
# loading packages
if (!require("pacman")) install.packages("pacman")
pacman::p_load(manifestoR, quanteda, tidyr, purrr, ggplot2,
               tidytext, httr, rvest, readr, xml2, reshape2,
               stringr, stringi, dplyr, tibble, lexicon,
               NMF, topicmodels, LDAvis, stm)
```

In the code below, the manifestos of each party per year are downloaded and saved into a dataframe. For ease of use, the dataset from 2000 to 2020 is already saved in the data folder so as to avoid downloading the same data repeatedly.

```r
# loop data collection

collect = FALSE

if (collect == TRUE) {
  mp_setapikey("manifesto_apikey.txt")

  mpds <- mp_maindataset()
  my_corpus <- mp_corpus(countryname == "United States" &
                           edate > as.Date("2000-11-07")) #change country year
  sample <- mpds %>%
    filter(countryname == "United States" &
             edate > as.Date("2000-11-07"))

  us_df <- tibble()

  for (i in names(my_corpus)) {
    doc <- my_corpus[[i]]
    doc_df <- as_tibble(as.data.frame(doc))
    doc_df$id <- i
    us_df <- rbind(us_df, doc_df)
  }

  us_df <- us_df %>%
    mutate(party = as.numeric(sub("_.*", "", id)),
           date = as.numeric(sub(".*_", "", id)))

  us_manifestos <- us_df %>%
  left_join(sample[,1:9], by = c("party"="party", "date"="date"))

  # save data
  write.csv(us_manifestos, "./data/us_manifestos.csv")
} else {
  us_manifestos <- read.csv("./data/us_manifestos.csv")
}
```

Initially, five electoral years were considered but unfortunately, the data for the Democratic Party in 2008 only contained one document. Due to this, the analysis will only focus on the three recent electoral years.

```r
print(table(us_manifestos$edate, us_manifestos$partyname))
```

```
            Democratic Party Republican Party
  2004-11-02             1000             1989
  2008-11-04                1             1154
  2012-11-06             1395             1793
  2016-11-08             1593             2289
  2020-11-03             2762             2289
```

A total of 12,121 documents will be processed and analyzed, with 5750 documents for the Democratic Party and 6371 documents from the Republican Party, from the last three electoral years.

```
df <- us_manifestos %>%
  filter(countryname == "United States" & edate >= as.Date("2012-11-06")) %>%
  filter(!is.na(text))

print(table(df$edate, df$partyname))
```

```
            Democratic Party Republican Party
  2012-11-06             1395             1793
  2016-11-08             1593             2289
  2020-11-03             2762             2289
```

```
print(table( df$partyname))
```

```
Democratic Party Republican Party
            5750             6371
```

III. Topic modeling

1. Latent Dirichlet Allocation

The document-feature matrix is created for each party and electoral year. Each dfmat is generated using the set of documents which are tokenized by removing English stopwords and punctuations, omitting obvious words like "united", "democrat", and "republican", stemming words, and setting the minimum term frequency as 10.

The resulting 6 dfmats are appended into a single list, `dfmat_list`. The LDA model processes each each dfmat in the `dfmat_list` in the loop. Each document-topics matrix (gamma) and topic-words matrix (beta) are converted into a dataframe and are tagged to the corresponding party and year. To create a single dataframe of `doc_topics_df` and `topic_words_df`, each converted dataframe is also appended inside the loop.

With LDA, `doc_topics_df`is a dataframe in which every document for a particular year and party is a mixture of topics. Each document may contain words from several topics in particular proportions. For instance, if the number of topics is set as 2, document 1 can be 80% topic1 and 20% topic2 while document 2 can be 40% topic1 and 60% topic2.

Meanwhile, `topic_words_df`is a dataframe in which every topic for a specific year and party is a mixture of words. For example, a two-topic model of US manifestos of the Republican Party in 2020 can have one topic about economic growth and another on healthcare. The most common words for the economic growth topic can be "GDP", "development", "inflation", while the healthcare topic can be "covid", "pandemic", and "health". In LDA, words can be shared between topics.

```r
# loop create dfmat and LDA by year and party

LDA = TRUE

if (LDA == TRUE) {
  dfmat_list <- list()
  doc_topics_df <- tibble()
  topic_words_df <- tibble()
  omit_words <- c("united", "state", "democrat", "republican",
                    "american", "america", "u.s")

  for (i in unique(df$edate)) {
    for (j in unique(df$partyname)) {
      df1 <- df %>%
      filter(partyname == j & edate == as.Date(i))

      dfmat <-  df1$text %>%
        tokens(remove_punct = T) %>%
        tokens_remove(pattern=stopwords("en")) %>%
        tokens_remove(omit_words) %>%
        tokens_wordstem() %>%
        dfm()  %>%
        dfm_trim(min_termfreq = 10)

      raw.sum=apply(dfmat,1,FUN=sum)
      dfmat=dfmat[raw.sum!=0,]

      print(sprintf("Created (%s, %s) dfmat using %s manifestos of the %s",
                  dim(dfmat)[1], dim(dfmat)[2], format(as.Date(i), format = "%Y"), j))

      dfmat_list <- append(dfmat_list, dfmat)


      # LDA model
      print(sprintf("Starting LDA model using %s manifestos of the %s",
                   format(as.Date(i), format = "%Y"), j))

      lda <- LDA(dfmat, control=list(seed=28), k=10) # change num of topics

      W <- lda@gamma # document-topic
      H <- lda@beta # topic-term

      doc_topics <- tidy(lda, matrix="gamma") %>%
        mutate(date = i, partyname = j)
      doc_topics_df <- rbind(doc_topics_df, doc_topics)


      topic_words <- tidy(lda, matrix="beta") %>%
        mutate(date = i, partyname = j)
      topic_words_df <- rbind(topic_words_df, topic_words)

      print(sprintf("Finished LDA model using %s manifestos of the %s",
                   format(as.Date(i), format = "%Y"), j))
```

```
    }
  }
  write.csv(doc_topics_df, "./data/doc_topics.csv")
  write.csv(topic_words_df, "./data/topic_words.csv")
} else {
  doc_topics_df <- read.csv("./data/doc_topics.csv")
  topic_words_df <- read.csv("./data/topic_words.csv")
}
```

```
[1] "Created (1386, 354) dfmat using 2012 manifestos of the Democratic Party"
[1] "Starting LDA model using 2012 manifestos of the Democratic Party"
[1] "Finished LDA model using 2012 manifestos of the Democratic Party"
[1] "Created (1760, 444) dfmat using 2012 manifestos of the Republican Party"
[1] "Starting LDA model using 2012 manifestos of the Republican Party"
[1] "Finished LDA model using 2012 manifestos of the Republican Party"
[1] "Created (1562, 384) dfmat using 2016 manifestos of the Democratic Party"
[1] "Starting LDA model using 2016 manifestos of the Democratic Party"
[1] "Finished LDA model using 2016 manifestos of the Democratic Party"
[1] "Created (2238, 516) dfmat using 2016 manifestos of the Republican Party"
[1] "Starting LDA model using 2016 manifestos of the Republican Party"
[1] "Finished LDA model using 2016 manifestos of the Republican Party"
[1] "Created (2721, 599) dfmat using 2020 manifestos of the Democratic Party"
[1] "Starting LDA model using 2020 manifestos of the Democratic Party"
[1] "Finished LDA model using 2020 manifestos of the Democratic Party"
[1] "Created (2238, 516) dfmat using 2020 manifestos of the Republican Party"
[1] "Starting LDA model using 2020 manifestos of the Republican Party"
[1] "Finished LDA model using 2020 manifestos of the Republican Party"
```

For this task, the LDA model is set to return 10 topics. The number of topics is a hyperparameter, which affects the resulting topics in such a way that a higher the number of topics can identify more diverse and specific topics. In contrast, by decreasing this hyperparameter, the model can identify more general topics given the set of documents.

Also, it must be noted that the LDA model returns different results unless the `contol` is fixed by setting a similar seed every run.

After investigating the document-topics matrix, using a 10-topic model identifies that each topic account for more or less than 10% per document. For example, the table below shows the topic share of the 10 topics using Democrats' manifesto in 2012.

```
doc_topics_df %>%
  arrange(document, date, partyname) %>%
  mutate(topic_share = round(gamma,3)) %>%
  select(date, partyname, document, topic, topic_share) %>%
  head(10)
```

```
# A tibble: 10 x 5
   date       partyname        document topic topic_share
   <chr>      <chr>            <chr>    <int>       <dbl>
 1 2012-11-06 Democratic Party text1        1       0.099
 2 2012-11-06 Democratic Party text1        2       0.099
 3 2012-11-06 Democratic Party text1        3       0.1
 4 2012-11-06 Democratic Party text1        4       0.102
```

```
 5 2012-11-06 Democratic Party text1        5       0.1
 6 2012-11-06 Democratic Party text1        6       0.1
 7 2012-11-06 Democratic Party text1        7       0.1
 8 2012-11-06 Democratic Party text1        8       0.1
 9 2012-11-06 Democratic Party text1        9       0.099
10 2012-11-06 Democratic Party text1       10       0.1
```

Now, to investigate the underlying topics in each party's manifestos in the three recent electoral years, the results are plotted showing the top 10 terms with the highest per-topic-per-word probabilities in topic.

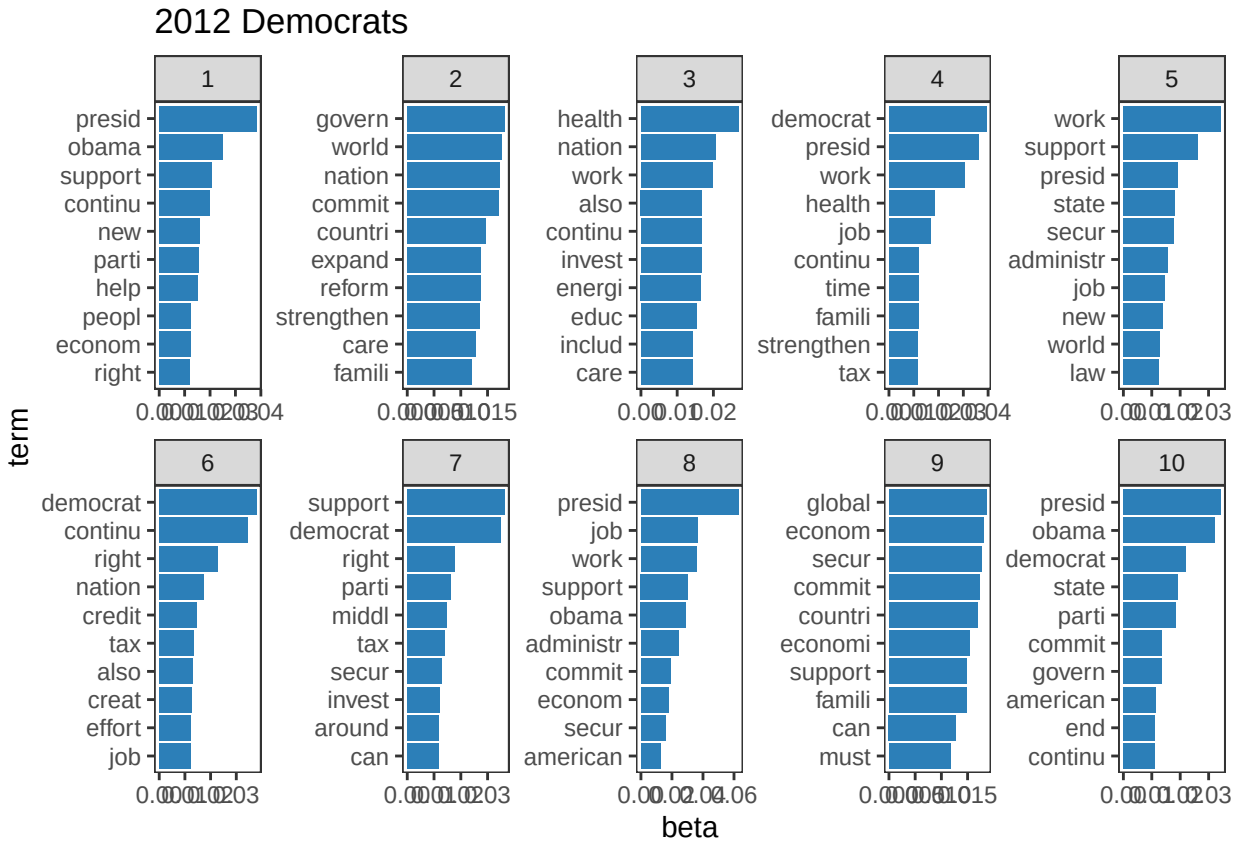
```
topic_words_df1 <- topic_words_df %>%
  mutate(year = format(as.Date(date), format = "%Y")) %>%
  mutate(party = ifelse(partyname == "Republican Party", "Republicans", "Democrats")) %>%
  mutate(group = paste0(year, " ", party),
         color = ifelse(party == "Democrats", "#2c7fb8", "#de2d26")) %>%
  group_by(topic, date, year, partyname) %>%
  slice_max(beta, n = 10) %>%
  ungroup() %>%
  arrange(date, topic, partyname, -beta)
```

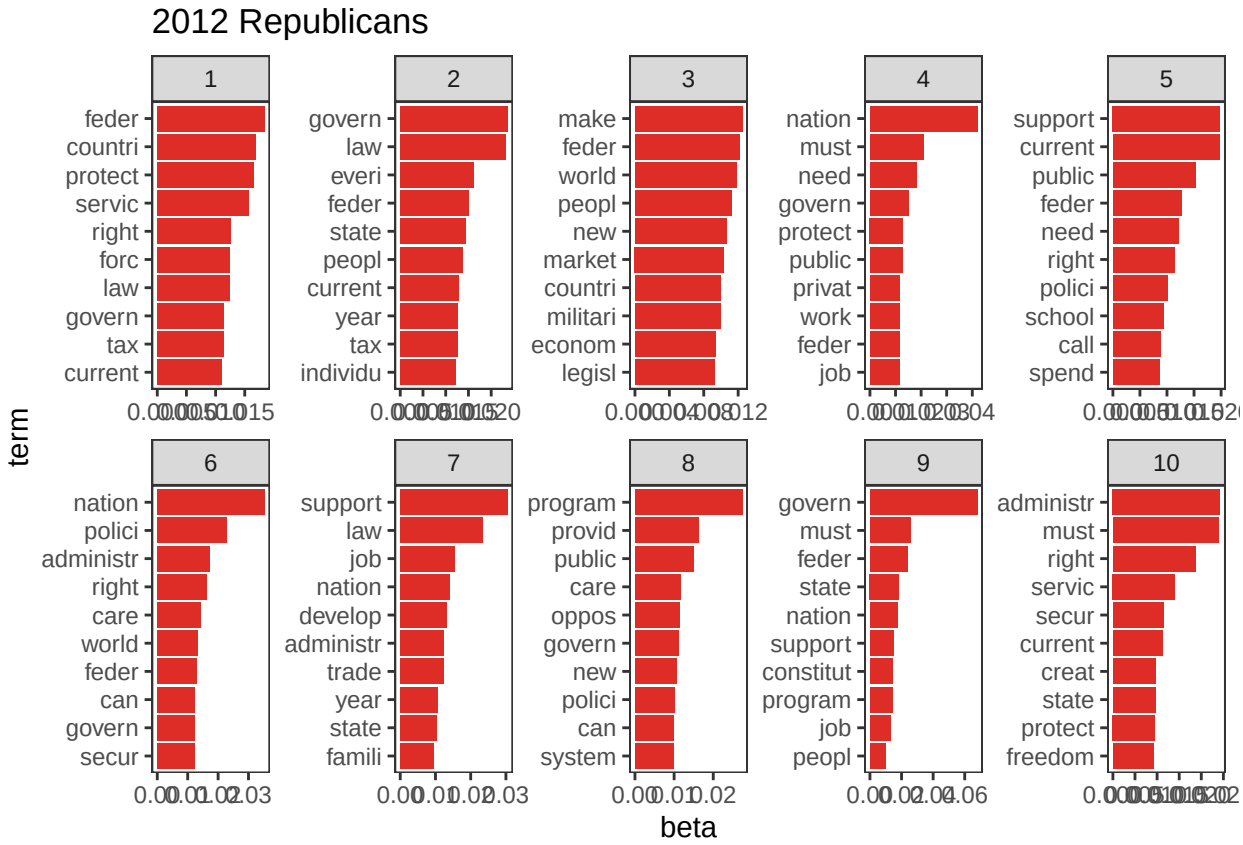
In 2012, Democrats' manifesto appear to show Obama's focus on health care or Obamacare in topic 3, and about the middle class and tax in topic 7.

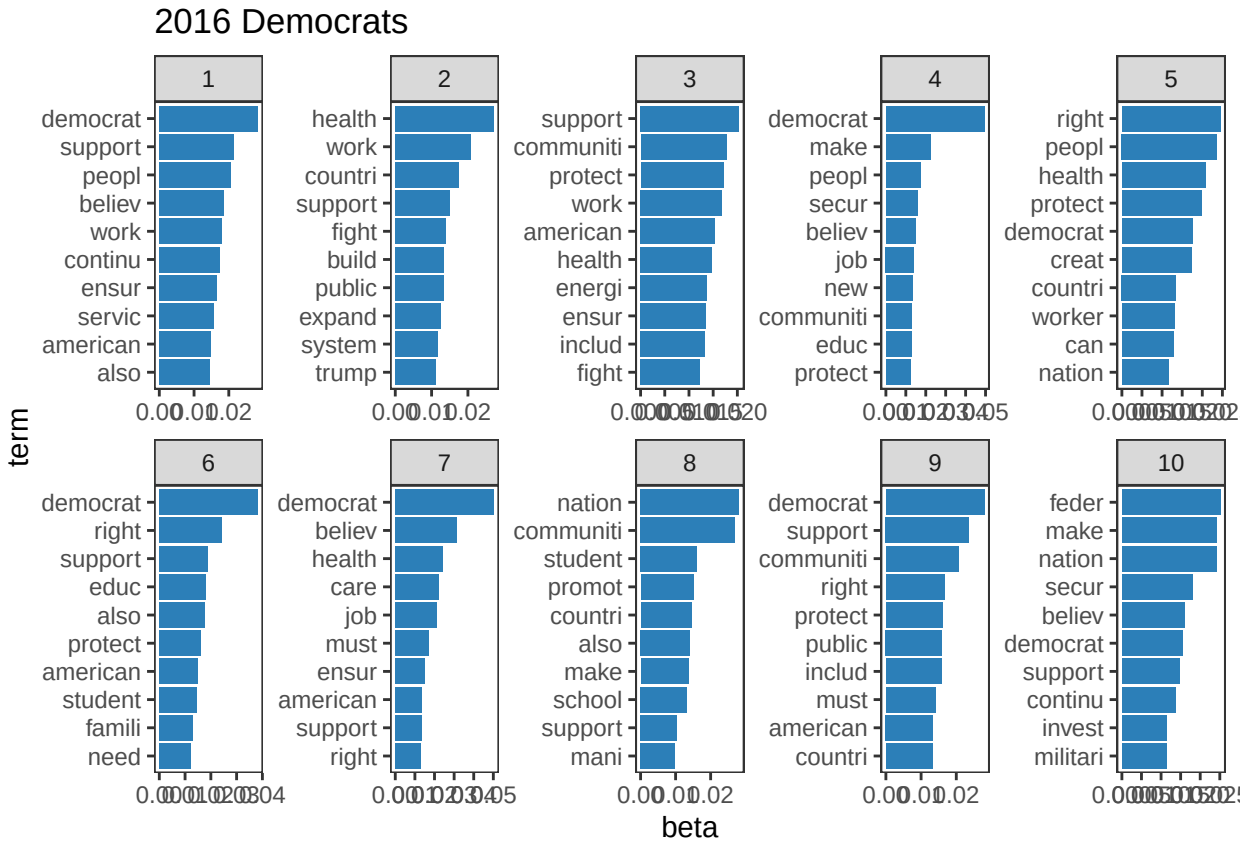
In 2016, the topics from Democrats' manifesto are also about healthcare in topic 7.

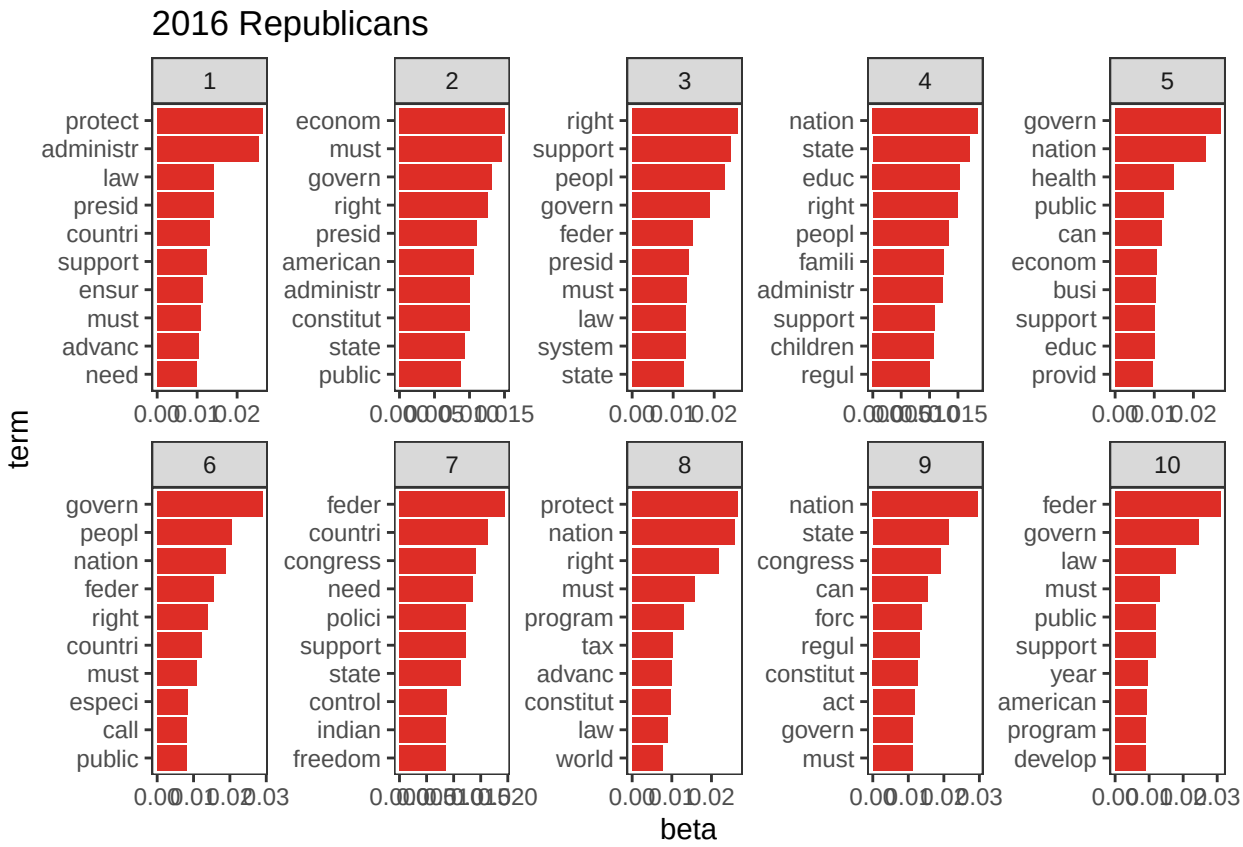
In 2020, most of the topics are about healthcare for the Democrats, for instance, topic 10 consists words like "health workers" and "pandemic". Interestingly, topic 4 contains the word "trump", "secur", and "protect" together.

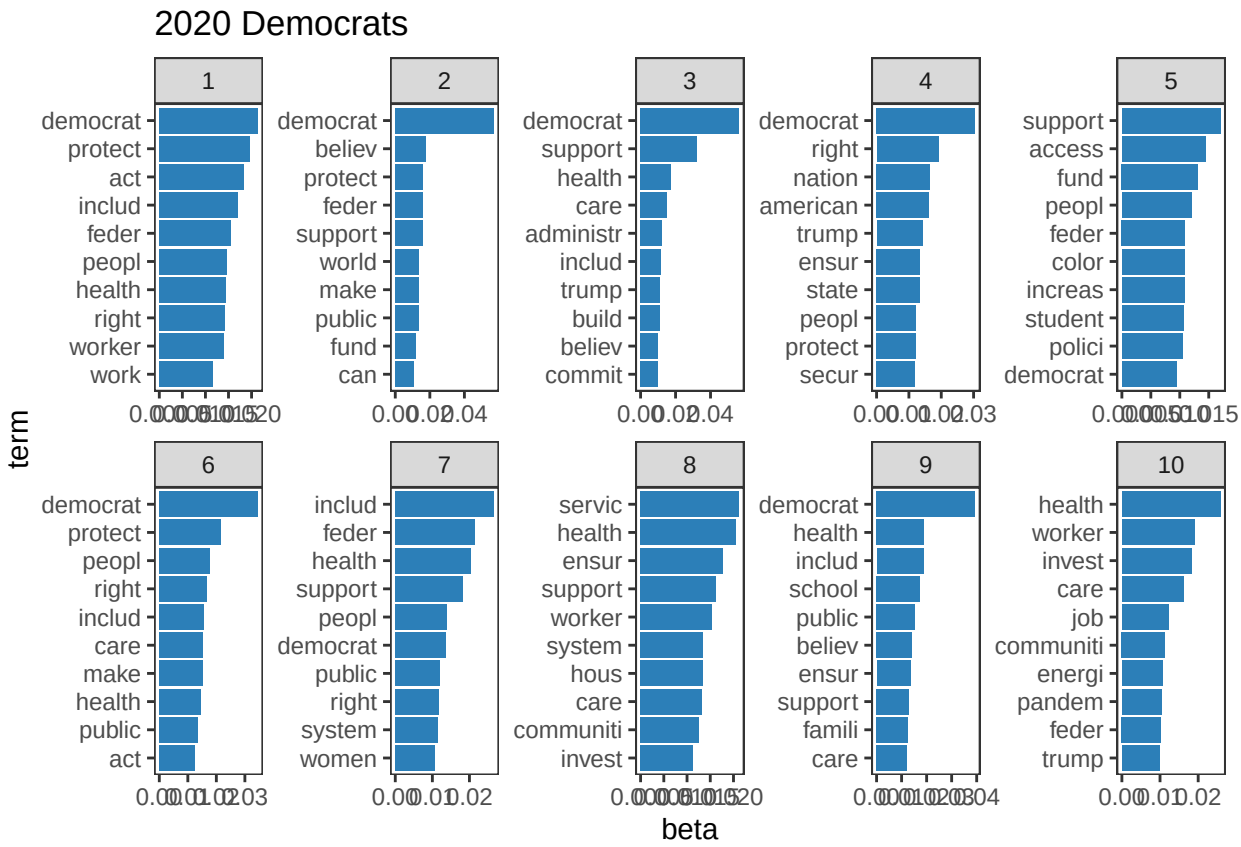
```
for (i in unique(topic_words_df1$group)){
    p<-topic_words_df1 %>%
      filter(group == i) %>%
      mutate(term = reorder_within(term, beta, topic)) %>%
      ggplot(aes(beta, term, fill = color)) +
      geom_col(show.legend = FALSE) +
      facet_wrap(~ topic, scales = "free", ncol=5) +
      scale_y_reordered() +
      labs(title = i) +
      scale_fill_identity() +
      theme_test()
    print(p)
    ggsave(paste0("./plots/lda_",i, ".png"))
}
```

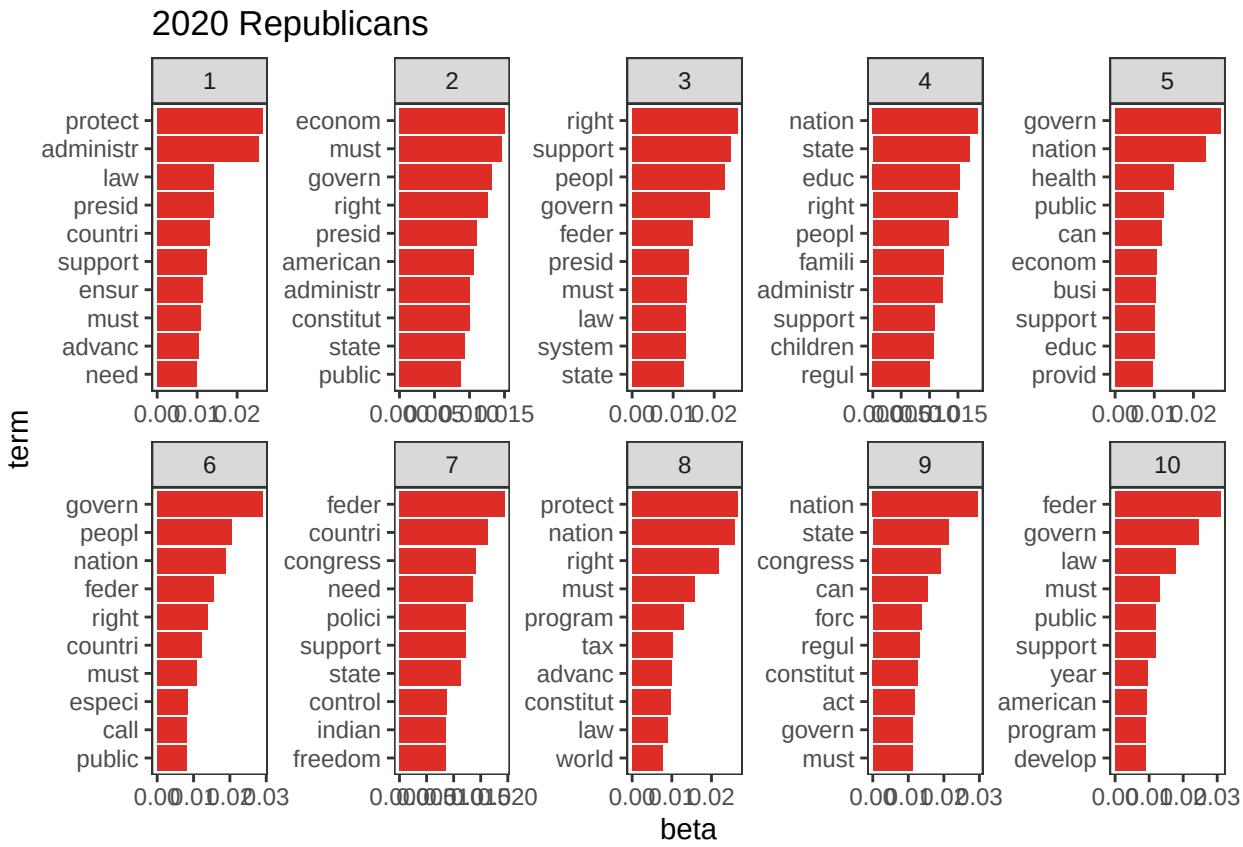

2012 Democrats
1
presid
obama
support
continu
new
parti
help
peopl
econom
right
2
govern
world
nation
commit
countri
expand
reform
strengthen
care
famili
3
health
nation
work
also
continu
invest
energi
educ
includ
care
4
democrat
presid
work
health
job
continu
time
famili
strengthen
tax
5
work
support
presid
state
secur
administr
job
new
world
law
6
democrat
continu
right
nation
credit
tax
also
creat
effort
job
7
support
democrat
right
parti
middl
tax
secur
invest
around
can
8
presid
job
work
support
obama
administr
commit
econom
secur
american
9
global
econom
secur
commit
countri
economi
support
famili
can
must
10
presid
obama
democrat
state
parti
commit
govern
american
end
continu
term
beta


2012 Republicans
1
feder
countri
protect
servic
right
forc
law
govern
tax
current
2
govern
law
everi
feder
state
peopl
current
year
tax
individu
3
make
feder
world
peopl
new
market
countri
militari
econom
legisl
4
nation
must
need
govern
protect
public
privat
work
feder
job
5
support
current
public
feder
need
right
polici
school
call
spend
6
nation
polici
administr
right
care
world
feder
can
govern
secur
7
support
law
job
nation
develop
administr
trade
year
state
famili
8
program
provid
public
care
oppos
govern
new
polici
can
system
9
govern
must
feder
state
nation
support
constitut
program
job
peopl
10
administr
must
right
servic
secur
current
creat
state
protect
freedom
term
beta


2016 Democrats
1
democrat
support
peopl
believ
work
continu
ensur
servic
american
also
2
health
work
countri
support
fight
build
public
expand
system
trump
3
support
communiti
protect
work
american
health
energi
ensur
includ
fight
4
democrat
make
peopl
secur
believ
job
new
communiti
educ
protect
5
right
peopl
health
protect
democrat
creat
countri
worker
can
nation
6
democrat
right
support
educ
also
protect
american
student
famili
need
7
democrat
believ
health
care
job
must
ensur
american
support
right
8
nation
communiti
student
promot
countri
also
make
school
support
mani
9
democrat
support
communiti
right
protect
public
includ
must
american
countri
10
feder
make
nation
secur
believ
democrat
support
continu
invest
militari
term
beta


2016 Republicans
1
protect
administr
law
presid
countri
support
ensur
must
advanc
need
2
econom
must
govern
right
presid
american
administr
constitut
state
public
3
right
support
peopl
govern
feder
presid
must
law
system
state
4
nation
state
educ
right
peopl
famili
administr
support
children
regul
5
govern
nation
health
public
can
econom
busi
support
educ
provid
6
govern
peopl
nation
feder
right
countri
must
especi
call
public
7
feder
countri
congress
need
polici
support
state
control
indian
freedom
8
protect
nation
right
must
program
tax
advanc
constitut
law
world
9
nation
state
congress
can
forc
regul
constitut
act
govern
must
10
feder
govern
law
must
public
support
year
american
program
develop
0.00
0.01
0.02
0.03
0.05
0.10
0.15
0.20
term
beta


2020 Democrats
1
democrat
protect
act
includ
feder
peopl
health
right
worker
work
2
democrat
believ
protect
feder
support
world
make
public
fund
can
3
democrat
support
health
care
administr
includ
trump
build
believ
commit
4
democrat
right
nation
american
trump
ensur
state
peopl
protect
secur
5
support
access
fund
peopl
feder
color
increas
student
polici
democrat
6
democrat
protect
peopl
right
includ
care
make
health
public
act
7
includ
feder
health
support
peopl
democrat
public
right
system
women
8
servic
health
ensur
support
worker
system
hous
care
communiti
invest
9
democrat
health
includ
school
public
believ
ensur
support
famili
care
10
health
worker
invest
care
job
communiti
energi
pandem
feder
trump
term
beta



2. Structural Topic Modeling

In this section, structural topic model is implemented using the library `stm`. STM offers more flexibility in the analysis such as explicitly modeling which variables influence the prevalence of topics. It also allows for topic correlation which will be shown below.

To start STM, each document-feature matrix which were created in the code above is converted to an stm format then STM is applied to identify 15 topics in the for loop. In this exercise, the number of topics is arbitrarily chosen to be 15 since decreasing the number of topics results to longer time for the model to converge. The results of the model for each set of documents is saved as a list in `model_list`, so that it can easily be called later on.

The `stm` library also offers more functions than `lda`. For example, the top words of the 2012 Democratic Party's manifesto is printed below by the highest conditional probability for each topic. Meanwhile, FREX shows which words are comparatively common for a topic and exclusive for that topic compared to other topics

```
labelTopics(model_list[[1]])
```

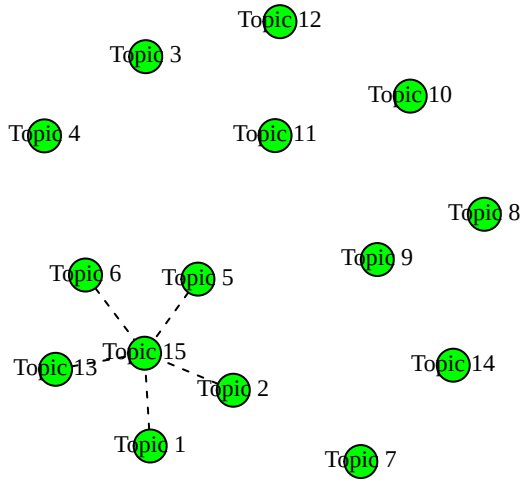
```
Topic 1 Top Words:
     Highest Prob: peopl, right, democrat, also, nation, american, parti
     FREX: right, peac, peopl, understand, recogn, live, univers
     Lift: univers, understand, recogn, peac, israel, scienc, live
     Score: univers, right, understand, peopl, peac, recogn, israel
Topic 2 Top Words:
     Highest Prob: presid, obama, rule, strengthen, govern, financi, economi
```

```
     FREX: rule, strengthen, took, wall, play, financi, place
     Lift: wall, transpar, rule, place, account, took, strengthen
     Score: wall, rule, everyon, obama, presid, took, strengthen
Topic 3 Top Words:
     Highest Prob: world, middl, class, make, polici, back, us
     FREX: middl, polici, class, world, progress, foreign, us
     Lift: foreign, polici, progress, middl, class, allianc, world
     Score: foreign, fail, class, middl, world, polici, us
Topic 4 Top Words:
     Highest Prob: law, women, act, support, effort, protect, oppos
     FREX: law, oppos, women, violenc, enforc, act, equal
     Lift: enforc, equal, violenc, oppos, respect, law, freedom
     Score: enforc, law, women, equal, violenc, act, oppos
Topic 5 Top Words:
     Highest Prob: help, must, ensur, strong, worker, free, work
     FREX: strong, free, must, help, poverti, ensur, labor
     Lift: labor, free, strong, maintain, citi, must, poverti
     Score: labor, must, strong, help, free, safeti, poverti
Topic 6 Top Words:
     Highest Prob: support, invest, continu, reform, provid, democrat, fight
     FREX: invest, reform, institut, infrastructur, provid, support, standard
     Lift: financ, standard, process, institut, infrastructur, train, invest
     Score: financ, support, reform, invest, infrastructur, provid, institut
Topic 7 Top Words:
     Highest Prob: secur, new, america, chang, region, build, reduc
     FREX: chang, america, secur, climat, region, partnership, build
     Lift: climat, chang, practic, emerg, africa, natur, partnership
     Score: climat, chang, secur, global, trade, america, region
Topic 8 Top Words:
     Highest Prob: health, care, famili, busi, access, work, veteran
     FREX: care, health, credit, insur, veteran, access, busi
     Lift: abl, compani, mortgag, care, insur, credit, health
     Score: abl, health, care, credit, insur, busi, compani
Topic 9 Top Words:
     Highest Prob: job, creat, year, million, program, student, home
     FREX: creat, million, job, student, year, good, program
     Lift: loan, million, creat, student, thousand, good, job
     Score: loan, job, creat, million, student, year, save
Topic 10 Top Words:
     Highest Prob: commit, presid, obama, end, war, nuclear, romney
     FREX: war, end, commit, mitt, weapon, romney, iraq
     Lift: hiv, mitt, war, weapon, end, iraq, romney
     Score: hiv, commit, war, end, weapon, mitt, romney
Topic 11 Top Words:
     Highest Prob: presid, tax, cut, work, obama, administr, pay
     FREX: cut, pay, tax, first, relief, bill, strategi
     Lift: strategi, pay, bill, cut, first, relief, sign
     Score: strategi, tax, cut, pay, first, obama, presid
Topic 12 Top Words:
     Highest Prob: state, time, move, forward, togeth, continu, step
     FREX: forward, move, time, togeth, puerto, status, face
     Lift: forward, status, puerto, rico, move, togeth, face
     Score: forward, move, status, puerto, rico, togeth, time
Topic 13 Top Words:
```

```
     Highest Prob: countri, communiti, need, econom, take, made, intern
     FREX: countri, need, communiti, great, restor, serv, take
     Lift: serv, centuri, 21st, countri, restor, contribut, great
     Score: serv, countri, communiti, need, safe, centuri, great
Topic 14 Top Words:
     Highest Prob: public, advanc, innov, organ, technolog, threat, nation
     FREX: organ, innov, advanc, network, public, technolog, traffick
     Lift: traffick, network, terrorist, drug, crimin, organ, innov
     Score: traffick, network, public, crimin, terrorist, organ, al-qaeda
Topic 15 Top Words:
     Highest Prob: can, believ, democrat, street, work, like, administr
     FREX: believ, can, street, democrat, like, work, administr
     Lift: street, believ, can, like, democrat, exist, parti
     Score: street, can, believ, democrat, like, parti, work
```

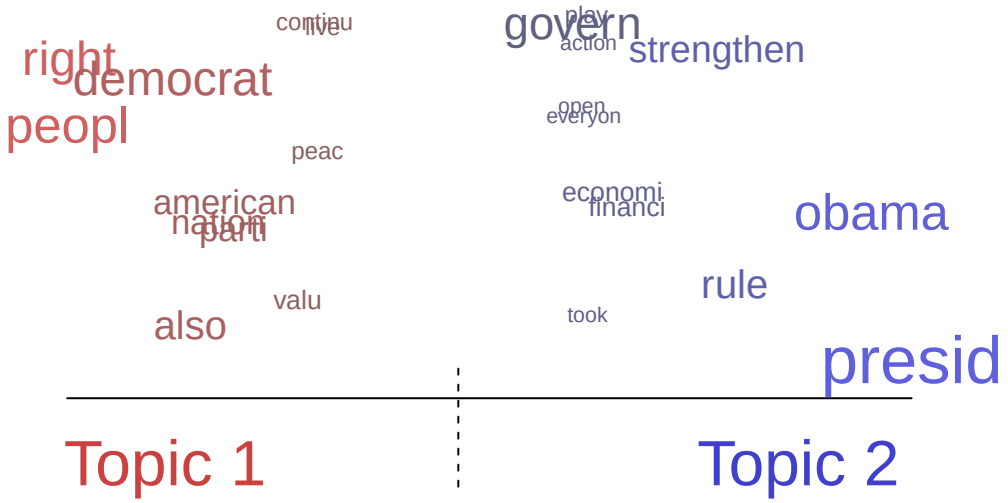
The correlation across the topics can also be visualized. Using the same manifesto above, topics 15, 1, 2, 6, 5 and 13 appear to be correlated. In addition, two topics (topic 1 and 2) can be compared wherein words are plotted with size proportional to their use within the topic and the x-axis shows how close the terms are to one topic over the other (y-axis is random).

```
topic_correlation<-topicCorr(model_list[[1]])
plot(topic_correlation)
```



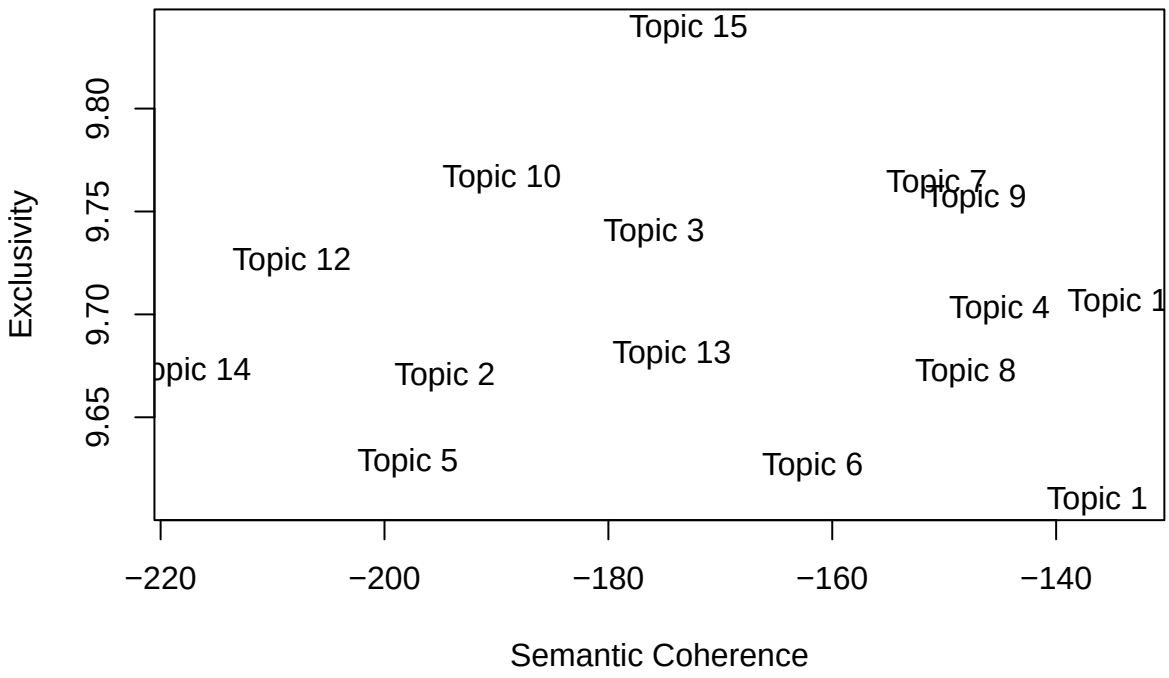

```
plot(model_list[[1]],
     type="perspectives",
```

```
    topics=c(1, 2),
    plabels = c("Topic 1","Topic 2"))
```



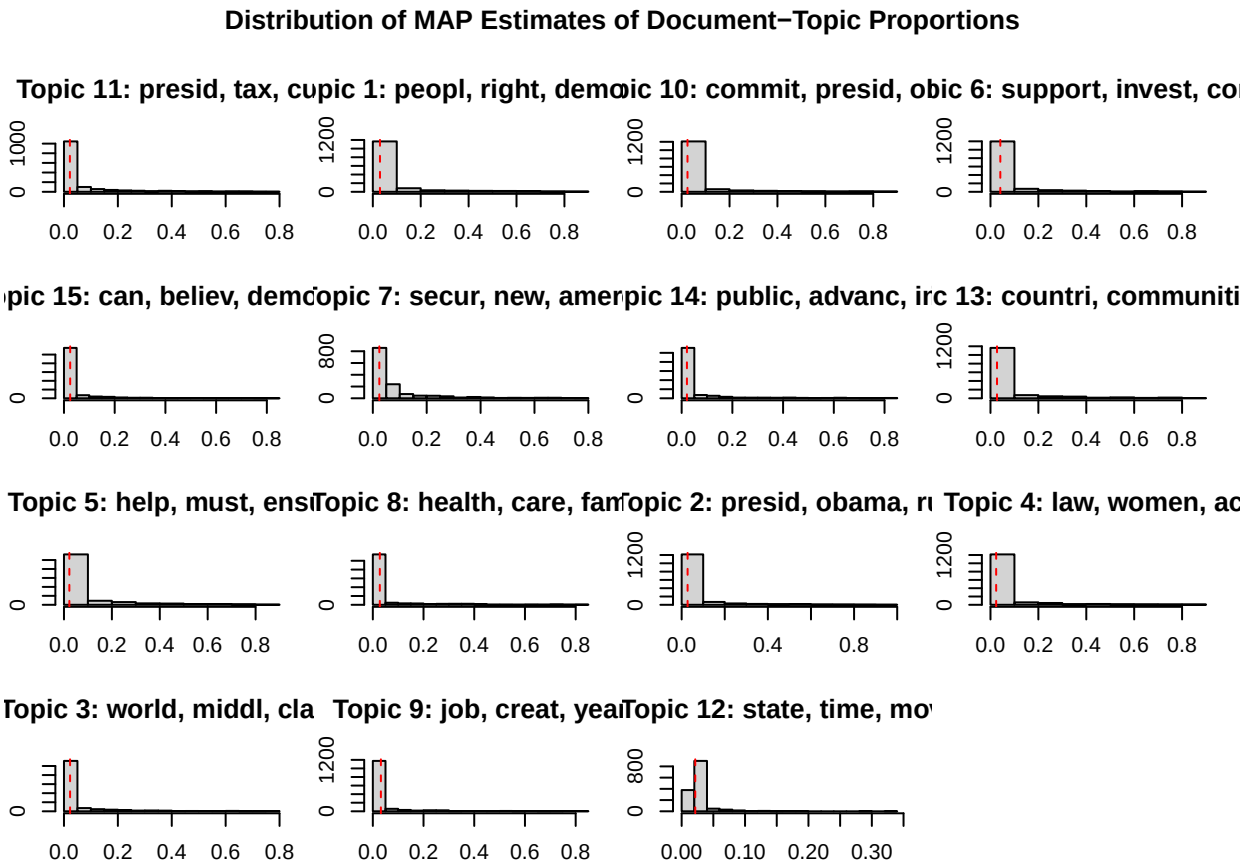

```
topicQuality(model=model_list[[1]], documents=dfm_stm_list[[1]]$documents)
```

```
 [1] -136.3100 -194.6037 -175.9260 -145.0520 -197.9109 -161.7569 -150.6700
 [8] -148.0860 -147.1607 -189.5163 -133.6732 -208.2623 -174.3291 -217.2114
[15] -172.8444
 [1] 9.609207 9.669494 9.739603 9.702083 9.627720 9.625868 9.763253 9.671788
 [9] 9.756122 9.765698 9.705767 9.725709 9.680164 9.672155 9.838986
```

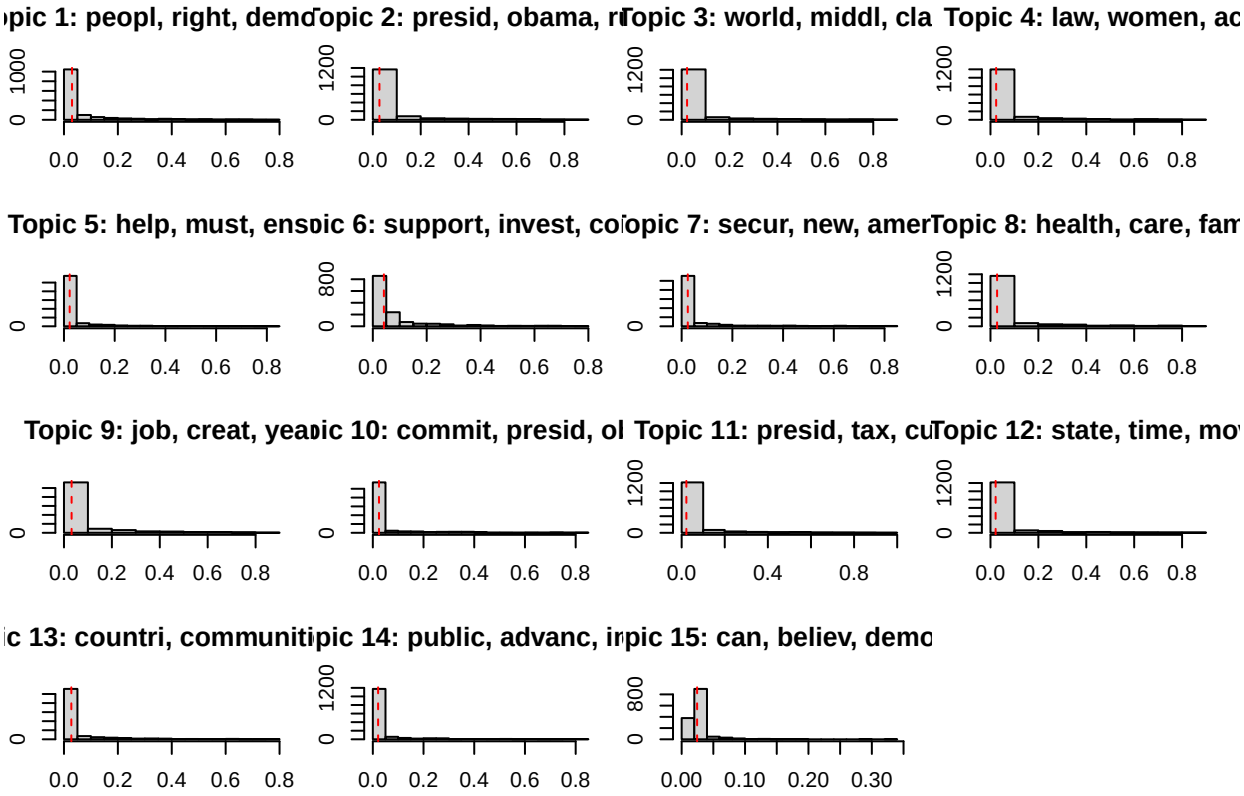


Below, the topic estimates of document-topic proportions is plotted using 2012 Democrats' manifesto. This plot basically tells us which topics are coming from which documents. As expected, each topic has no relation or very little relation with several documents.

```
plot(model_list[[1]], type = "hist", topics = sample(1:15, size = 15))
```

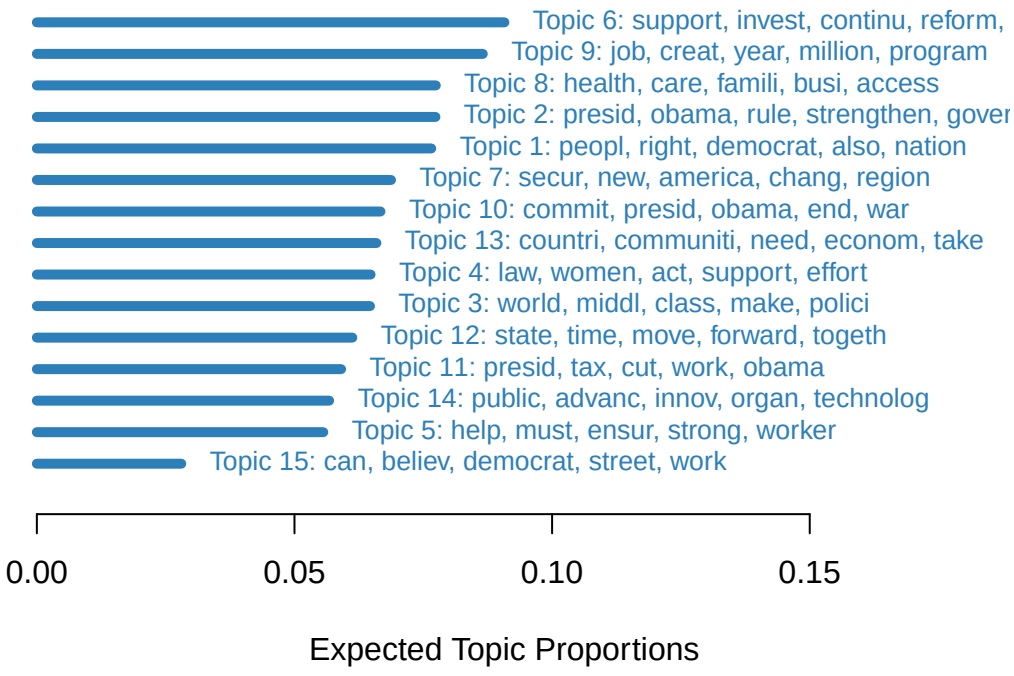



```
plot(model_list[[1]], type="hist")
```

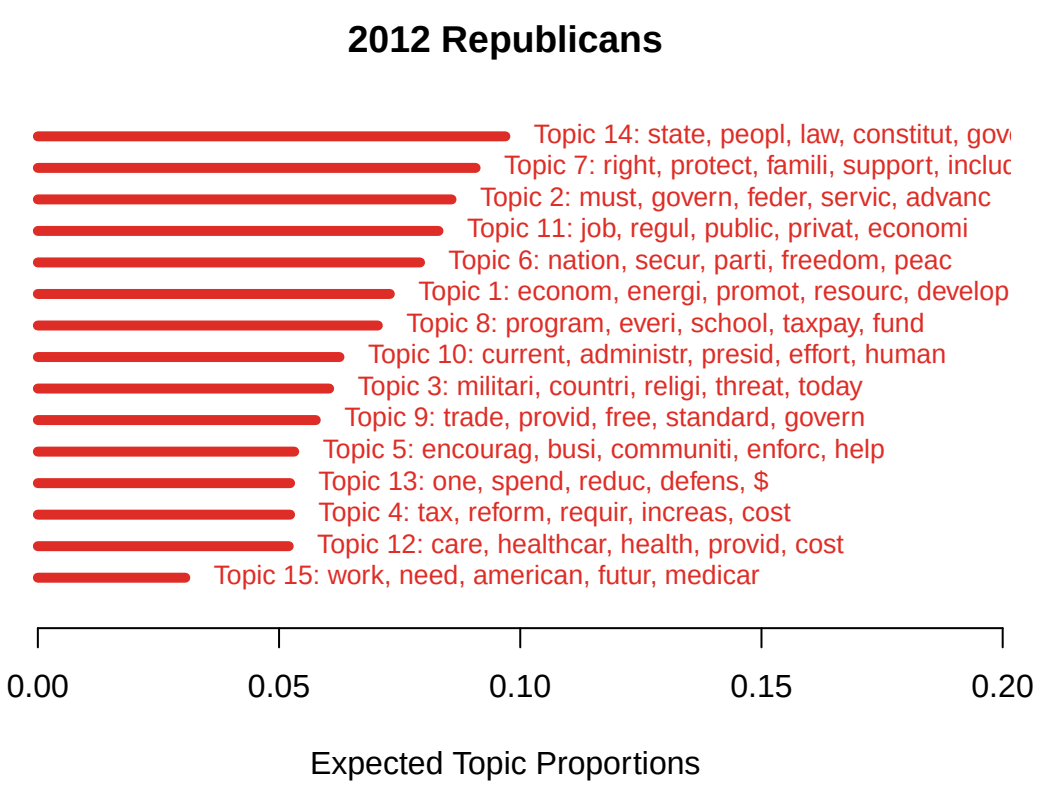


Another built-in function allows for topic exploration to easily analyze the results. Using `plot.STM`, the topic distribution is visualized, with most common words for each topic, between Democrats and Republicans across the three electoral years. This is basically similar to the topic exploration that was done with LDA but the plots are easier to interpret as it is arranged by the expected topic proportions and with the top common words already provided for each topic.

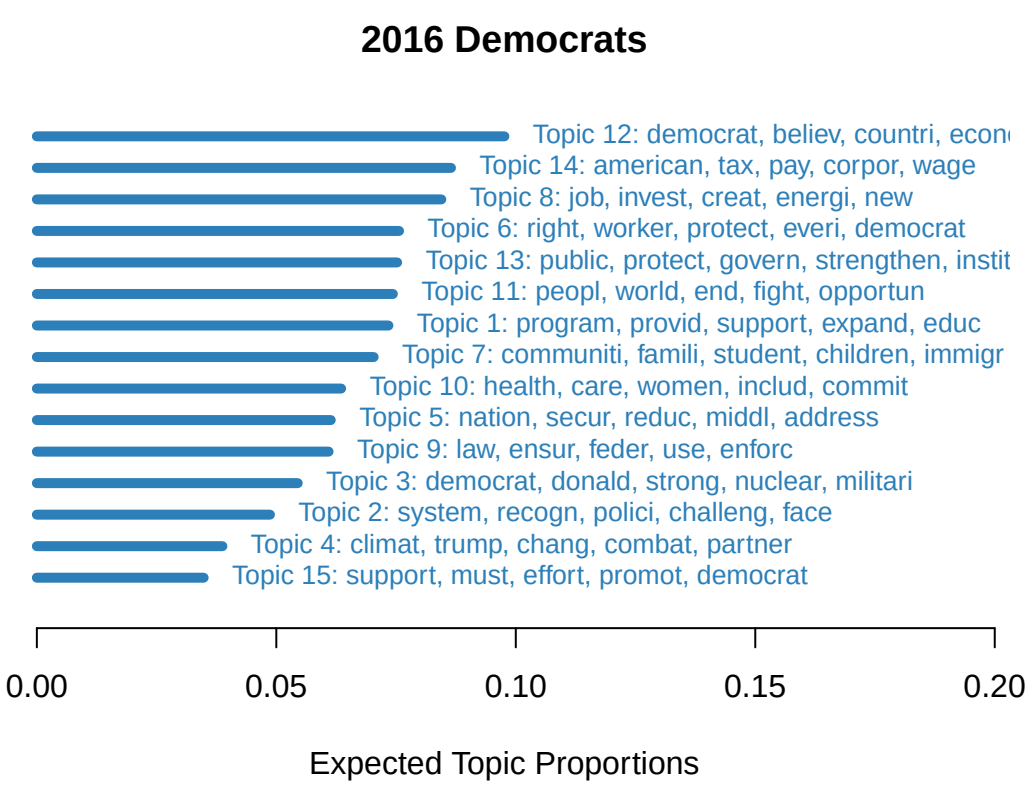
```
par(bty="n",col="#2c7fb8",lwd=5)
plot.STM(model_list[[1]],type="summary", n = 5, main = "2012 Democrats",
         width=50, text.cex=.8)
```



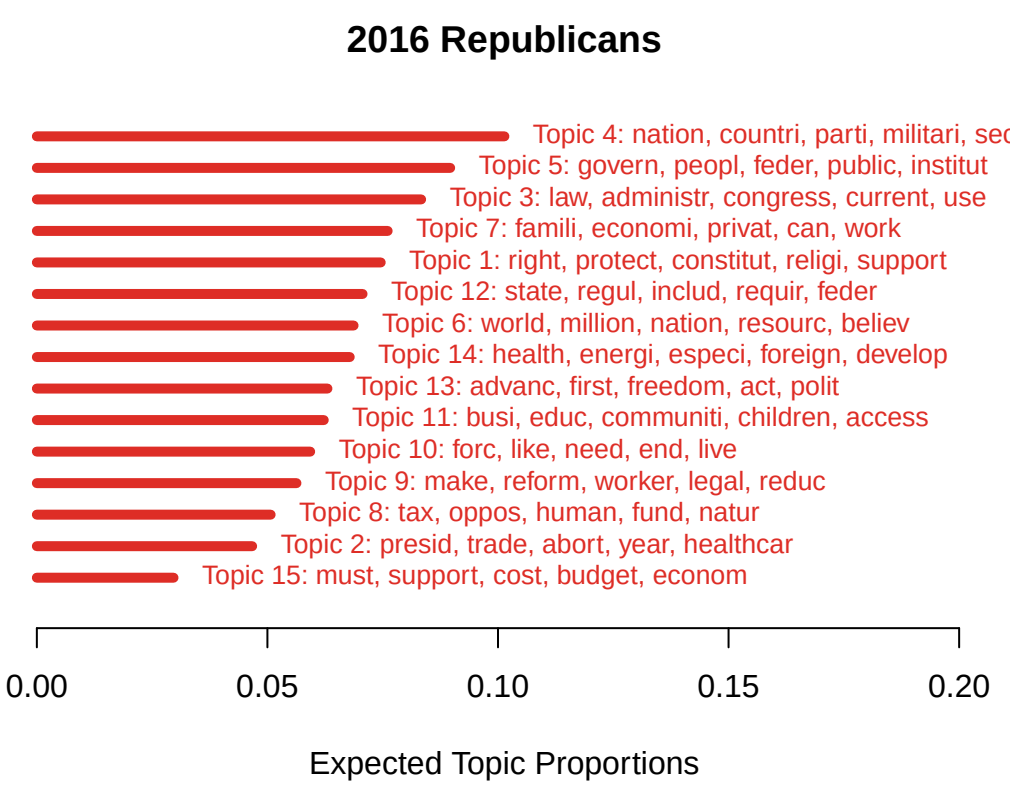

```
par(bty="n",col="#de2d26",lwd=5)
plot.STM(model_list[[2]],type="summary", n = 5, main = "2012 Republicans",
        width=50, text.cex=.8)
```



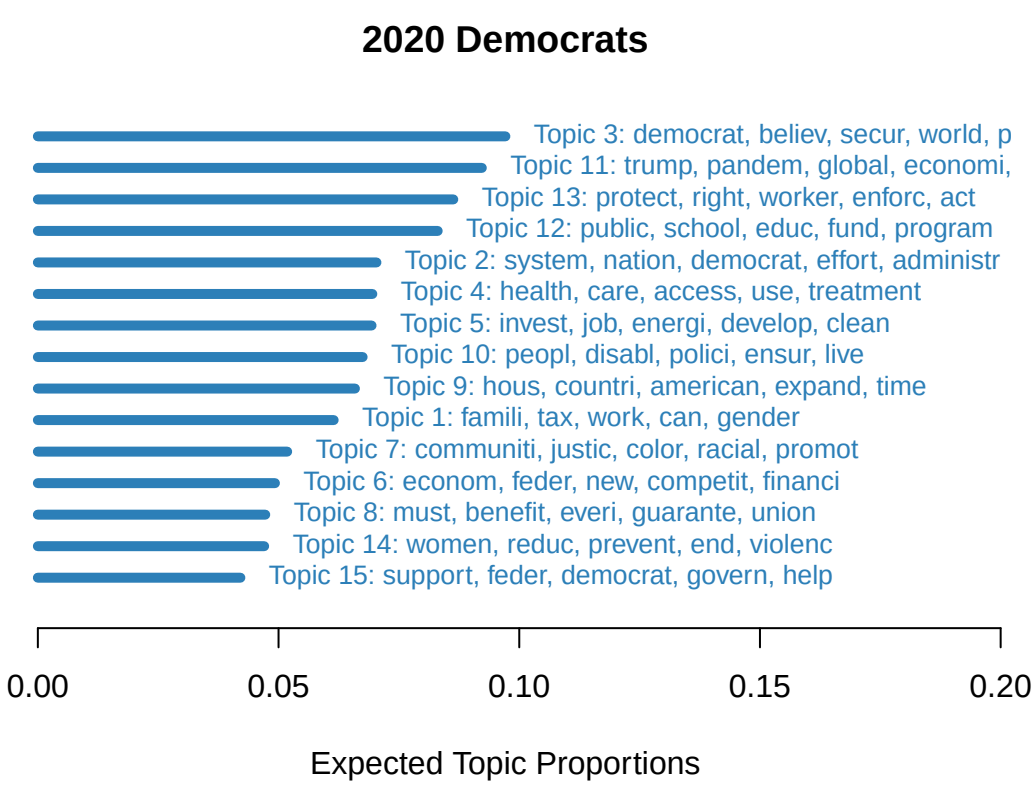

```
par(bty="n",col="#2c7fb8",lwd=5)
plot.STM(model_list[[3]],type="summary", n = 5, main = "2016 Democrats",
         width=50, text.cex=.8)
```



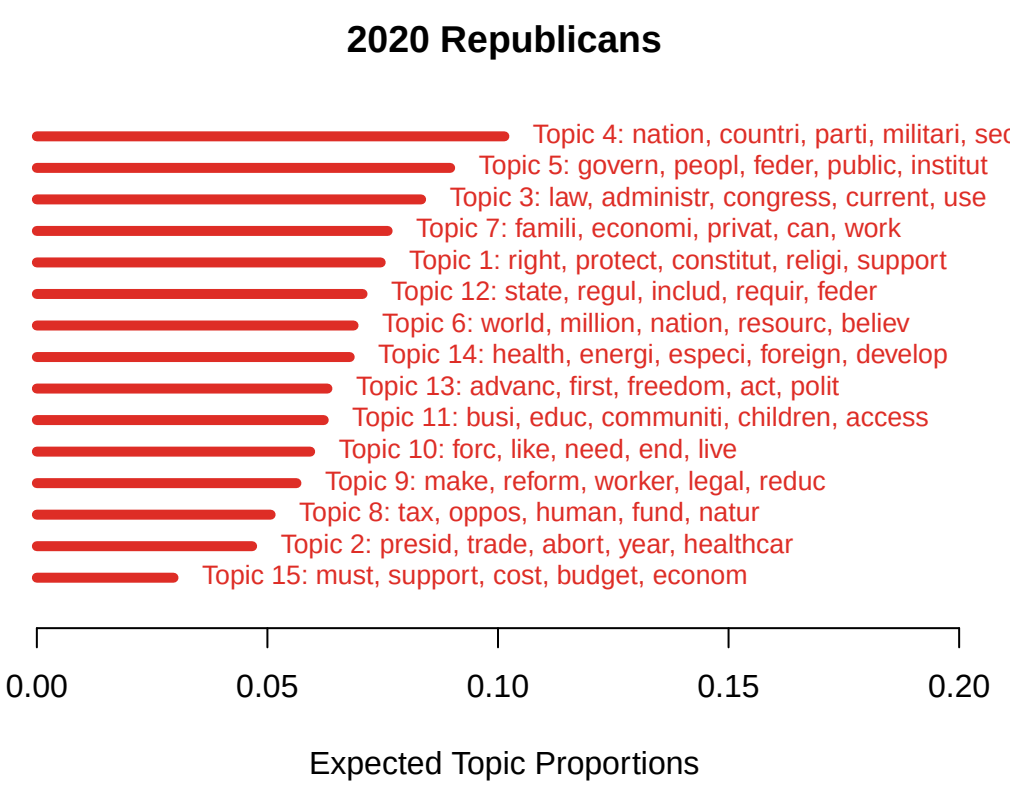

```
par(bty="n",col="#de2d26",lwd=5)
plot.STM(model_list[[4]],type="summary", n = 5, main = "2016 Republicans",
         width=50, text.cex=.8)
```



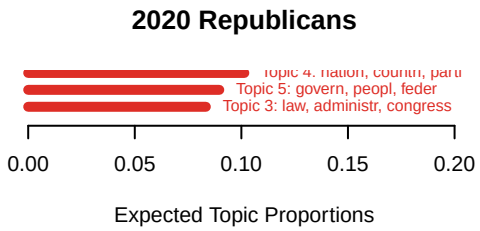

```
par(bty="n",col="#2c7fb8",lwd=5)
plot.STM(model_list[[5]],type="summary", n = 5, main = "2020 Democrats",
         width=50, text.cex=.8)
```

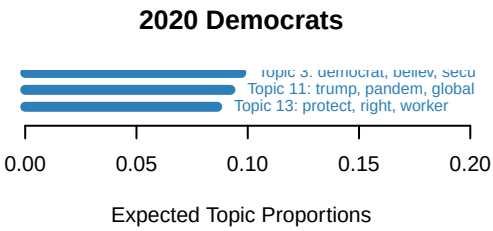


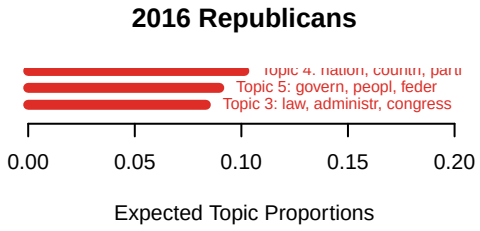

```
par(bty="n",col="#de2d26",lwd=5)
plot.STM(model_list[[6]],type="summary", n = 5, main = "2020 Republicans",
         width=50, text.cex=.8)
```

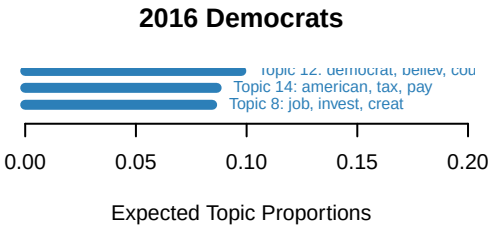


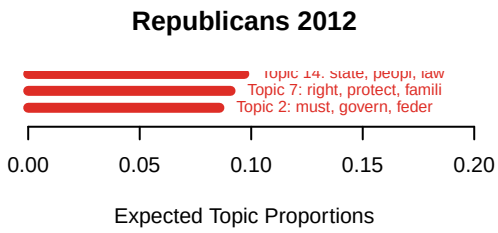

```
par(mfrow=c(3,2))
par(bty="n",col="#de2d26",lwd=5)
plot.STM(model_list[[6]],type="summary", n = 3, main = "2020 Republicans",
     topics=c(4,5,3), text.cex=.8, width=30)
par(bty="n",col="#2c7fb8",lwd=5)
plot.STM(model_list[[5]],type="summary", n = 3, main = "2020 Democrats",
     topics=c(3, 11, 13), text.cex=.8, width=30)
par(bty="n",col="#de2d26",lwd=5)
plot.STM(model_list[[4]],type="summary", n = 3, main = "2016 Republicans",
     topics=c(4,5,3), text.cex=.8, width=30)
par(bty="n",col="#2c7fb8",lwd=5)
plot.STM(model_list[[3]],type="summary", n = 3, main = "2016 Democrats",
     topics=c(12,14,8), text.cex=.8, width=30)
par(bty="n",col="#de2d26",lwd=5)
plot.STM(model_list[[2]],type="summary", n = 3, main = "Republicans 2012",
    topics=c(14,7,2), text.cex=.8, width=30)
par(bty="n",col="#2c7fb8",lwd=5)
plot.STM(model_list[[1]],type="summary", n = 3, main = "2012 Democrats",
     topics=c(6,9,8), text.cex=.8, width=30)
```

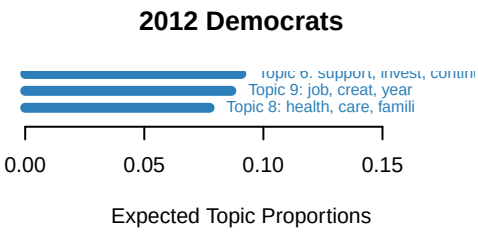












V. Conclusion

Use your topic model to answer your research question by showing plots or statistical results. Discuss the implications of what you find, and any limitations inherent in your approach. Discuss how the work could be improved upon in future research.