Narratives

Emily Maloney

March 19, 2019

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------------- tidyverse 1.2.
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0
```

```
## -- Conflicts -------------------------------------------------------------- tidyverse_conflicts()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(textnets)
```

```
## Loading required package: udpipe
```

```
## Loading required package: ggraph
```

```
## Loading required package: networkD3
```

```
## Warning: replacing previous import 'dplyr::union' by 'igraph::union' when
## loading 'textnets'
```

```
## Warning: replacing previous import 'dplyr::as_data_frame' by
## 'igraph::as_data_frame' when loading 'textnets'
```

```
## Warning: replacing previous import 'dplyr::groups' by 'igraph::groups' when
## loading 'textnets'
```

```
library(tidytext)
library(readxl)
library(stm)
```

```
## stm v1.3.3 (2018-1-26) successfully loaded. See ?stm for help.
##  Papers, resources, and other materials at structuraltopicmodel.com
```

```
library(topicmodels)
```

Load in data

```
d <- read_xlsx("walkaway_narratives.xlsx")
d <- d %>% mutate(Race = ifelse(Race == "White", "White", "Non-White"))

#get into tidy format
tidy <- d %>%
    select(Title, Text) %>%
    unnest_tokens("word", Text)

#preprocessing
 data("stop_words")
    tidy <-tidy %>%
      anti_join(stop_words)
```

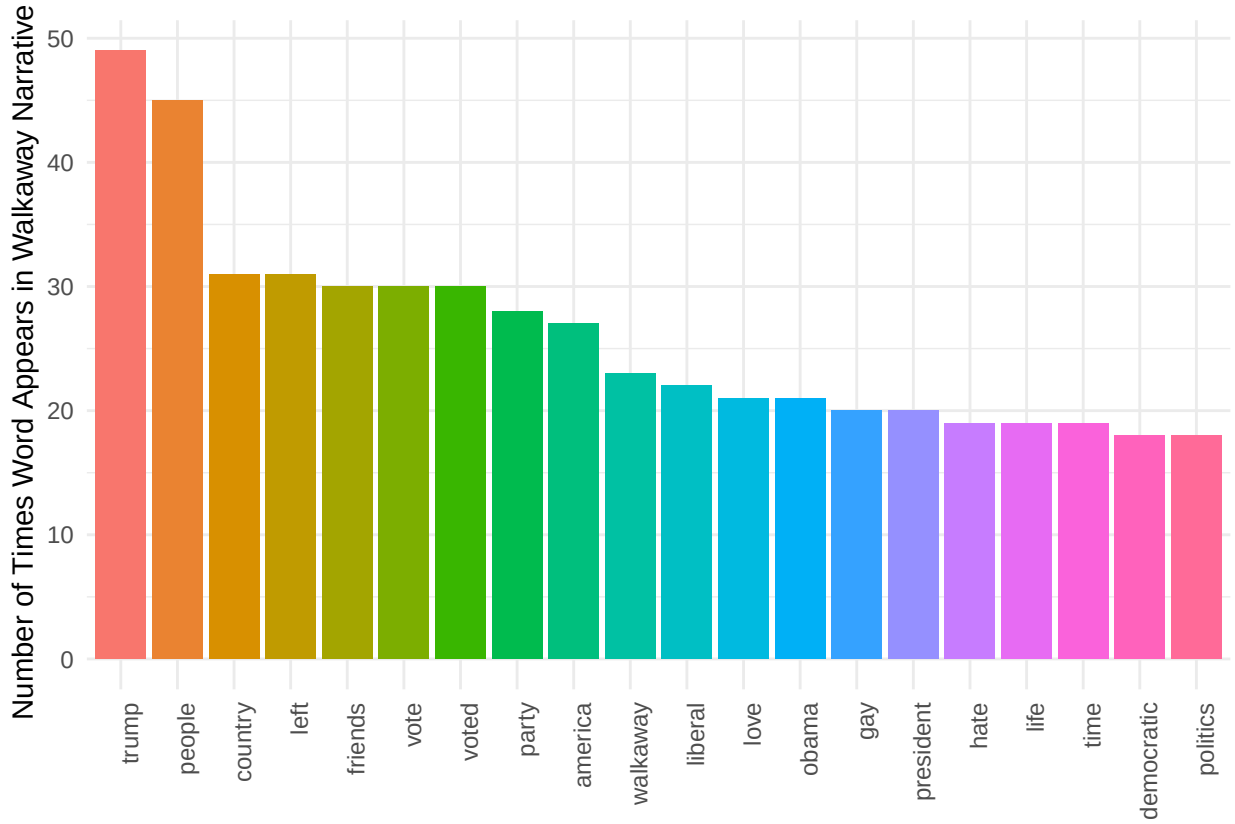
```
## Joining, by = "word"
```

Basic stuff

```
tidy_top <- tidy %>% count(word) %>% arrange(desc(n)) %>% slice(1:20)

#create factor variable to sort by frequency
tidy_top$word <- factor(tidy_top$word, levels = tidy_top$word[order(tidy_top$n,decreasing=TRUE)])


ggplot(tidy_top, aes(x=word, y=n, fill=word))+
  geom_bar(stat="identity")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  ylab("Number of Times Word Appears in Walkaway Narratives")+
  xlab("")+
  guides(fill=FALSE)
```



Trying out structural topic modeling:

```
processed <- textProcessor(d$Text, metadata = d)
```

```
## Building corpus...
## Converting to Lower Case...
## Removing punctuation...
## Removing stopwords...
## Removing numbers...
## Stemming...
```

```
## Creating Output...
```

```
out <- prepDocuments(processed$documents, processed$vocab, processed$meta)
```

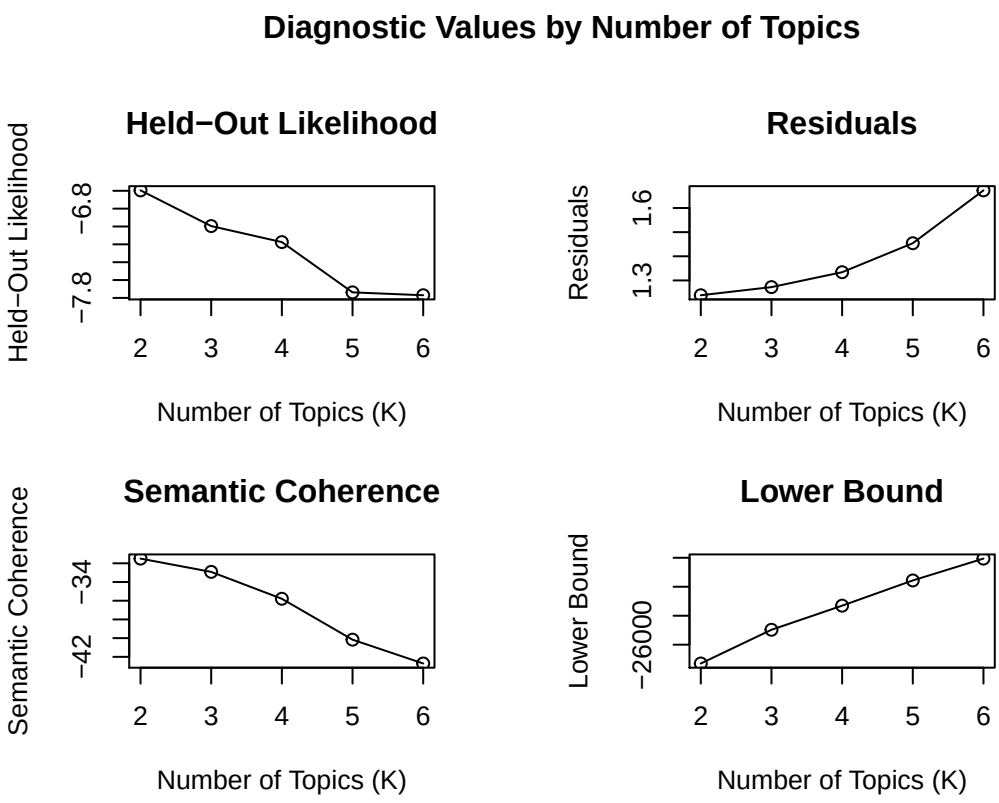
```
## Removing 1148 of 1927 terms (1148 of 4597 tokens) due to frequency
## Your corpus now has 35 documents, 779 terms and 3449 tokens.
```

```
docs <- out$documents
vocab <- out$vocab
meta <-out$meta
meta <- meta %>% mutate(Gender = as.factor(Gender),
                        Race = as.factor(Race),
                        Age = as.factor(Age))

findingk <- searchK(out$documents, out$vocab, K = c(2:6),
 prevalence =~ Gender + Age + Race, data = out$meta, verbose=FALSE)
```

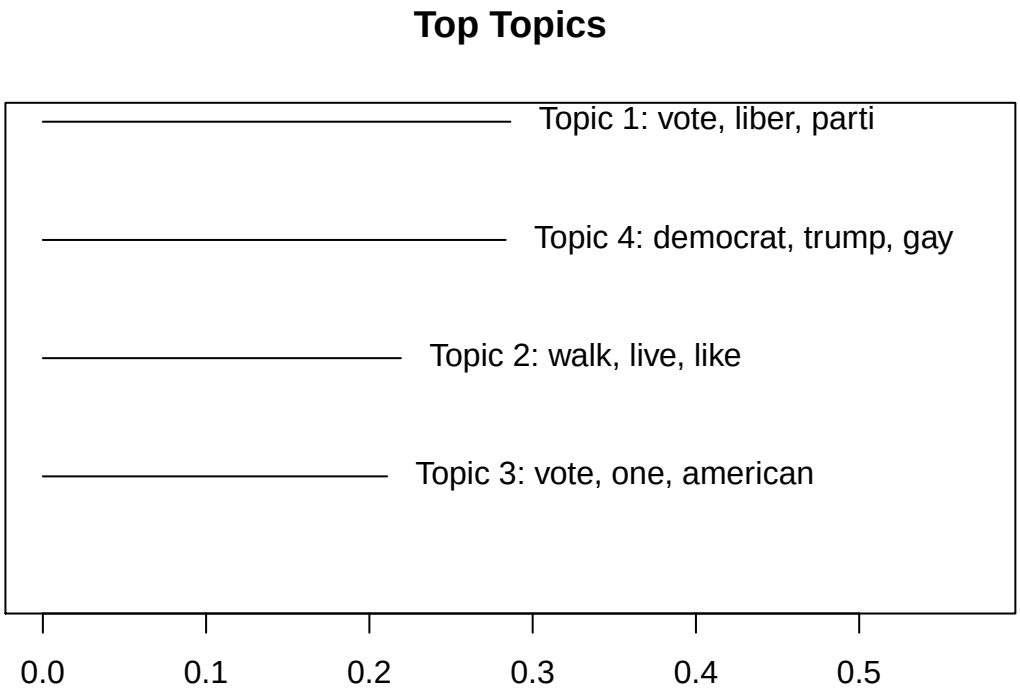
```
## Warning in stm(documents = heldout$documents, vocab = heldout$vocab, K
## = k, : K=2 is equivalent to a unidimensional scaling model which you may
## prefer.
```

```
plot(findingk)
```




```
First_STM <- stm(documents = out$documents, vocab = out$vocab,
              K = 4, prevalence =~ Gender + Age + Race,
              max.em.its = 75, data = out$meta,
              init.type = "Spectral", verbose = FALSE)
```

```
plot(First_STM)
```




```
#anti-left?
findThoughts(First_STM, texts = d$Text,
     n = 2, topics = 1)
```

```
##
##  Topic 1:
##       I never was a democrat but I was at one time indifferent. In my adolescent years where I felt s
##      I am a UAW member and for the record I am NOT anti-Union. I am Anti CORRUPT Union. I was raised
```

```
findThoughts(First_STM, texts = d$Text,
     n = 2, topics = 2)
```

```
##
##  Topic 2:
##       I saw people were doing introductions so I thought I would do one too. My name is Rachel and I
##      I am loving these stories and mine starts kind of recent. I was born in Virginia and moved to No
```

```
#most like typical conversion narrative?
findThoughts(First_STM, texts = d$Text,
     n = 2, topics = 3)
```

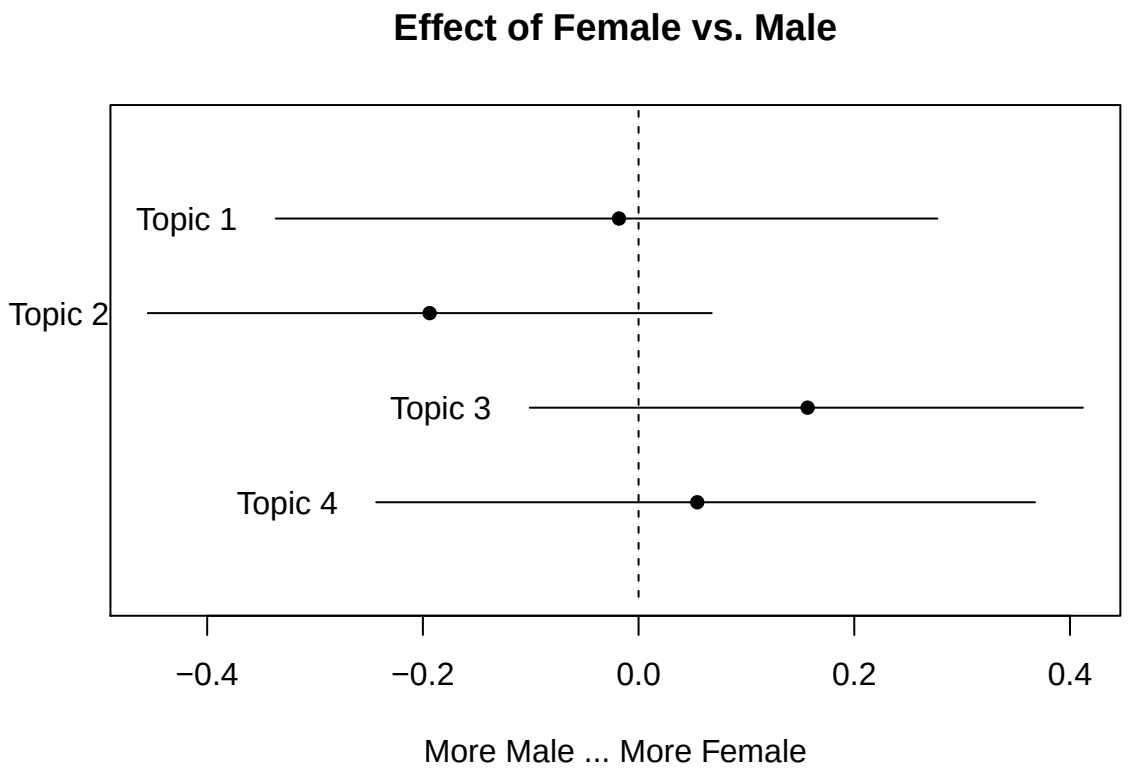
```
##
##  Topic 3:
##       I want to share my story coming from the perspective of a 25 year old millennial. My generation
##      Are you kidding me? Four days ago I asked to join this group! Two days ago I was accepted and po
```

```
#positive attitude
findThoughts(First_STM, texts = d$Text,
     n = 2, topics = 4)
```

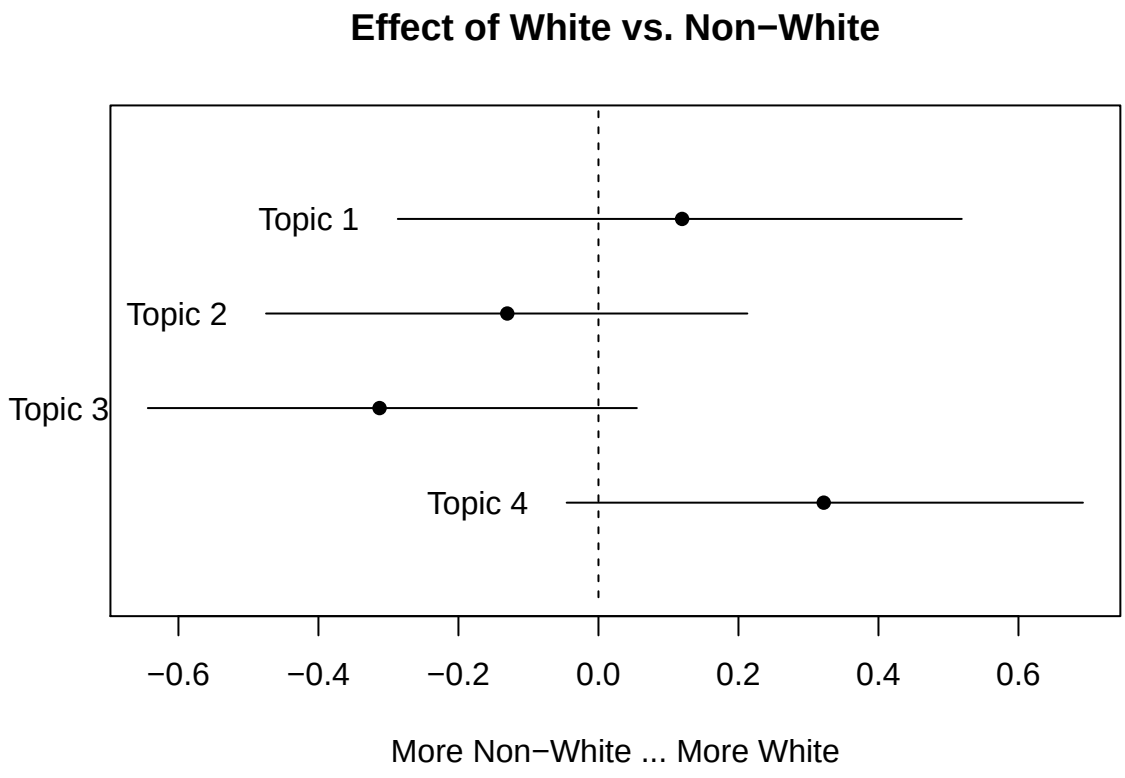
```
##
##  Topic 4:
##       I'm an artist and musician from Orlando who grew up democrat, and only ever really knew other
##      Hi folks! My name is Danny. First and foremost I am a proud patriotic American. I'm so happy to
```

```
predict_topics<-estimateEffect(formula = 1:4 ~ Gender + Race + Age, stmobj = First_STM, metadata = out$
```

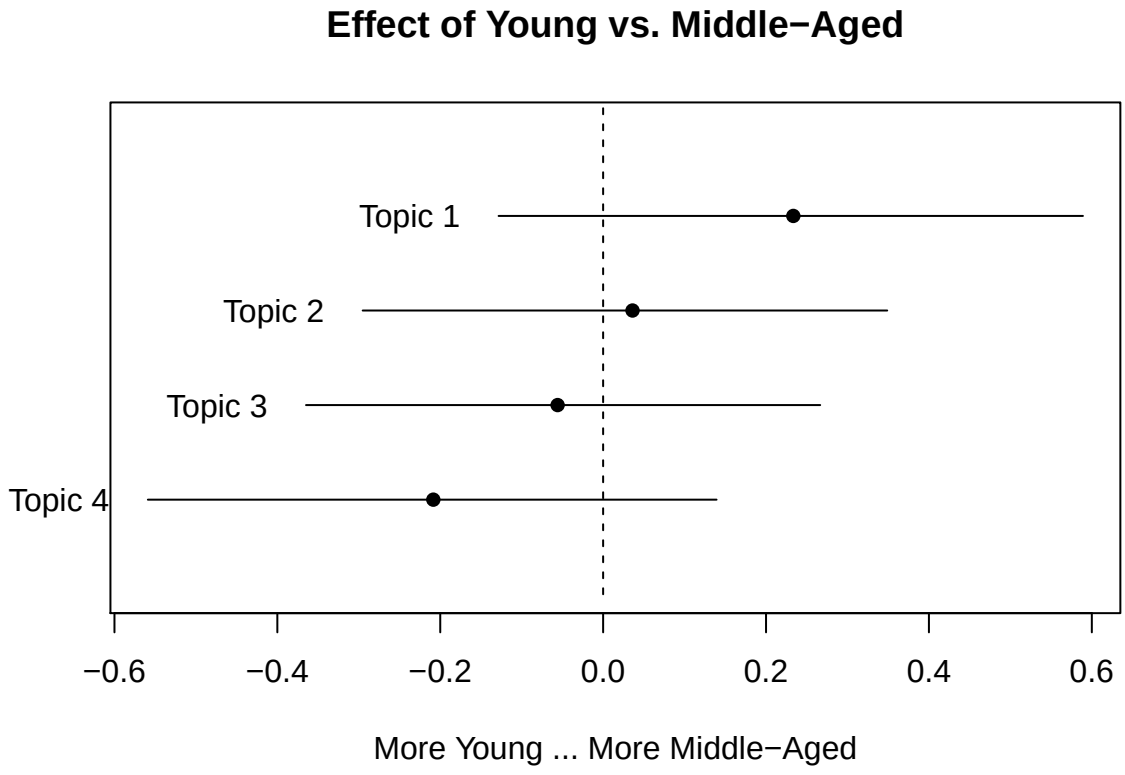
```
plot(predict_topics, covariate = "Gender", topics = 1:4,
 model = First_STM, method = "difference",
 cov.value1 = "Female", cov.value2 = "Male",
 xlab = "More Male ... More Female",
 main = "Effect of Female vs. Male",labeltype = "custom",
 custom.labels = c('Topic 1', 'Topic 2','Topic 3', 'Topic 4'))
```



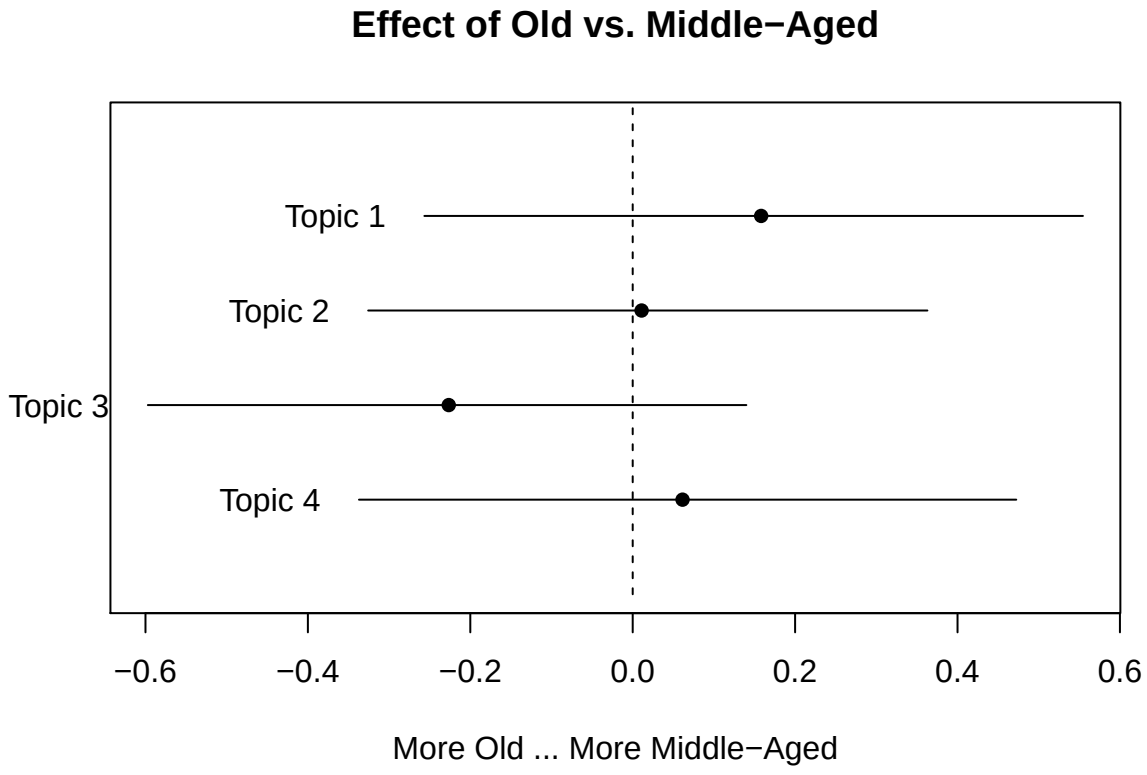

```
plot(predict_topics, covariate = "Race", topics = 1:4,
 model = First_STM, method = "difference",
 cov.value1 = "White", cov.value2 = "Non-White",
 xlab = "More Non-White ... More White",
 main = "Effect of White vs. Non-White",labeltype = "custom",
 custom.labels = c('Topic 1', 'Topic 2','Topic 3', 'Topic 4'))
```




```
plot(predict_topics, covariate = "Age", topics = 1:4,
 model = First_STM, method = "difference",
 cov.value1 = "Middle", cov.value2 = "Young",
 xlab = "More Young ... More Middle-Aged",
 main = "Effect of Young vs. Middle-Aged",labeltype = "custom",
 custom.labels = c('Topic 1', 'Topic 2','Topic 3', 'Topic 4'))
```



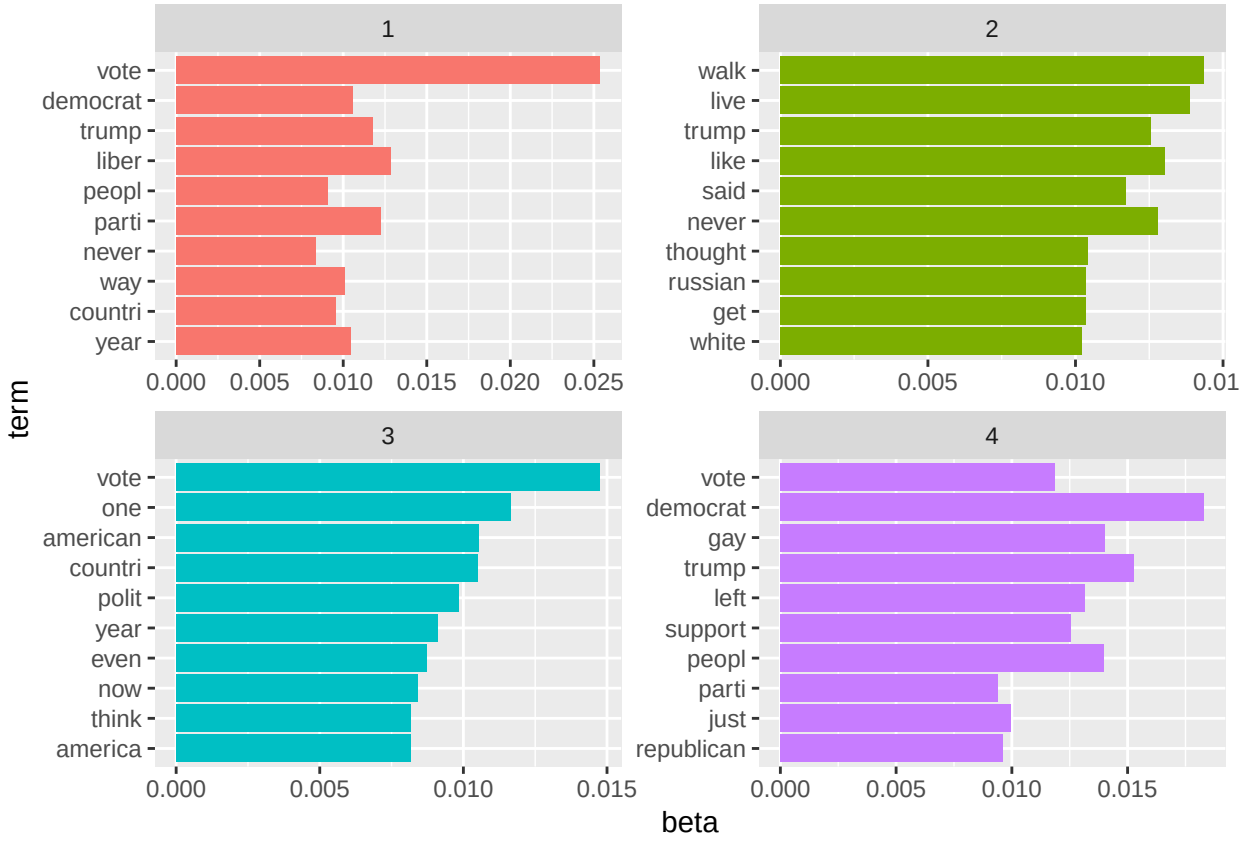

```
plot(predict_topics, covariate = "Age", topics = 1:4,
 model = First_STM, method = "difference",
 cov.value1 = "Middle", cov.value2 = "Old",
 xlab = "More Old ... More Middle-Aged",
 main = "Effect of Old vs. Middle-Aged",labeltype = "custom",
 custom.labels = c('Topic 1', 'Topic 2','Topic 3', 'Topic 4'))
```




```
topics <- tidy(First_STM, matrix = "beta")

top_terms <-
  topics %>%
  group_by(topic) %>%
  top_n(10, beta) %>%
  ungroup() %>%
  arrange(topic, -beta)


top_terms %>%
  mutate(term = reorder(term, beta)) %>%
  ggplot(aes(term, beta, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  coord_flip()
```



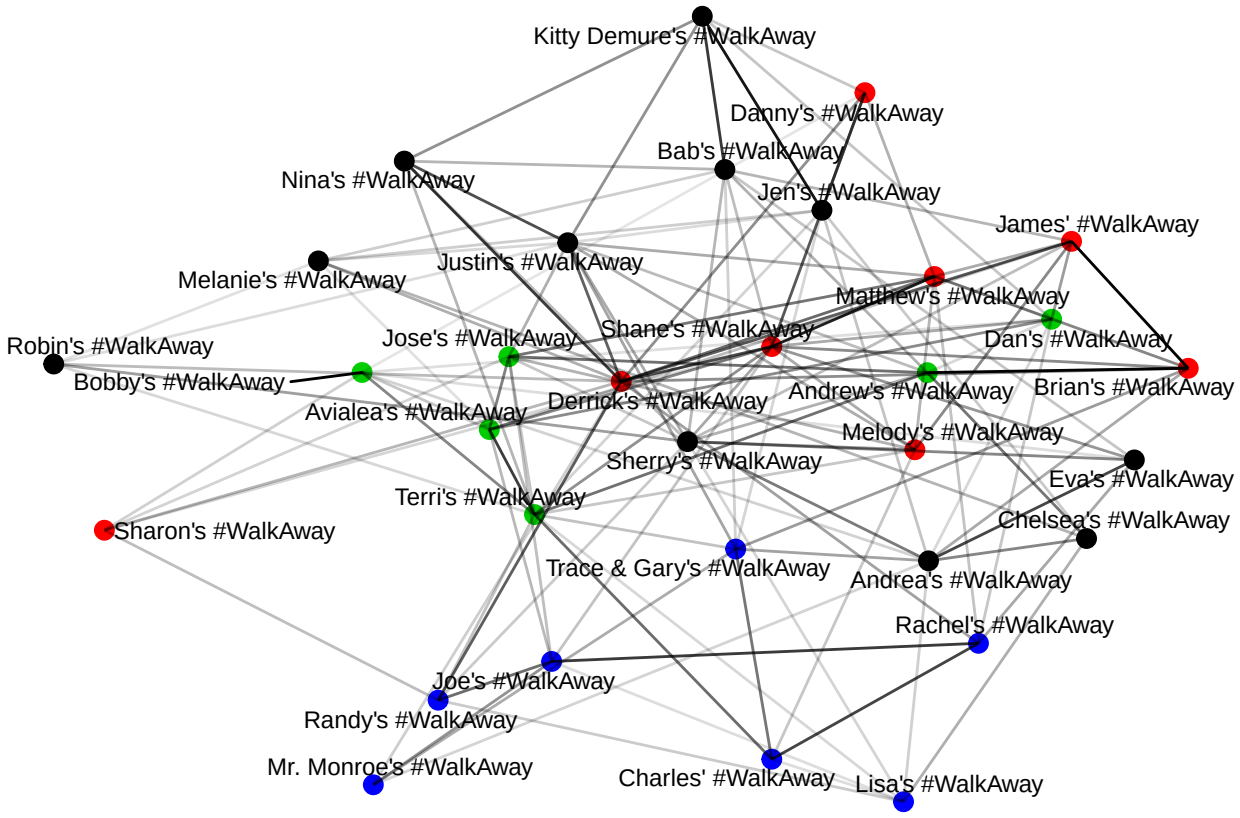
TextNets

```
#with narratives as nodes
prepped_walk <- PrepText(d, groupvar = "Title", textvar = "Text", node_type = "groups", tokenizer = "wo
```

```
## Downloading udpipe model from https://raw.githubusercontent.com/jwijffels/udpipe.models.ud.2.3/maste
```

```
walk_network <- CreateTextnet(prepped_walk)
VisTextNet(walk_network, label_degree_cut = 0)
```

```
## Using `nicely` as default layout
```




```
#with words as nodes
prepped_walk2 <- PrepText(d, groupvar = "Title", textvar = "Text", node_type = "words", tokenizer = "wo
```

```
## Downloading udpipe model from https://raw.githubusercontent.com/jwijffels/udpipe.models.ud.2.3/maste
```

```
walk_network2 <- CreateTextnet(prepped_walk2)

walk_communities <- TextCommunities(walk_network2)
```

```
##
## Attaching package: 'igraph'
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
## The following objects are masked from 'package:purrr':
##
##     compose, simplify
## The following object is masked from 'package:tidyr':
##
##     crossing
## The following object is masked from 'package:tibble':
##
##     as_data_frame
## The following objects are masked from 'package:stats':
##
```

```
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

```
walk_communities %>% group_by(modularity_class) %>% slice(1:10)
```

```
## # A tibble: 190 x 2
## # Groups:   modularity_class [19]
##    group       modularity_class
##    <chr>       <chr>
##  1 70          1
##  2 aca         1
##  3 adult       1
##  4 airport     1
##  5 anchor      1
##  6 anesthesia  1
##  7 ar          1
##  8 arm         1
##  9 arm dog tag 1
## 10 attention   1
## # ... with 180 more rows
```

```
text_centrality <- TextCentrality(walk_network2)


head(text_centrality)
```

```
##                      betweenness_centrality closness_centrality
## '                                  8.070588        0.0003098297
## #                                  0.000000        0.0002564591
## # loveamerica                      0.673913        0.0002654971
## # notabot # walkaway             488.187500        0.0004126684
## # redpill letâ                   119.661290        0.0003944178
## # walkaway                       112.000000        0.0003891888
```