

Structural Topic Models

Loren Collingwood

The structural topic model incorporates document metadata into the correlated topic model. In addition, the stm R packages provides a great suite of functions to better understand your data, models, and how to present the findings.

```
options(scipen = 999, digits = 4)

#####
#      Packages      #
#####

#install.packages("stm")
library(stm)

## stm v1.3.6 successfully loaded. See ?stm for help.
## Papers, resources, and other materials at structuraltopicmodel.com

#install.packages("wordcloud")
library(wordcloud)

## Loading required package: RColorBrewer
```

Step 1

Read in and preprocess your data.

```
# Read in Data #
data <- read.csv("~/Dropbox/collingwood_research/posc_fall_20/POSC-207/data/poliblogs2008.csv")

# Check Dimensions; n=13,246
dim(data)

## [1] 13246      6

# Look at column names, and you can see the metadata (we will use rting and day)
colnames(data)

## [1] "X"          "documents" "docname"   "rating"    "day"       "blog"

# Process the text real good #
processed <- textProcessor(data$documents,
                           metadata = data)

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

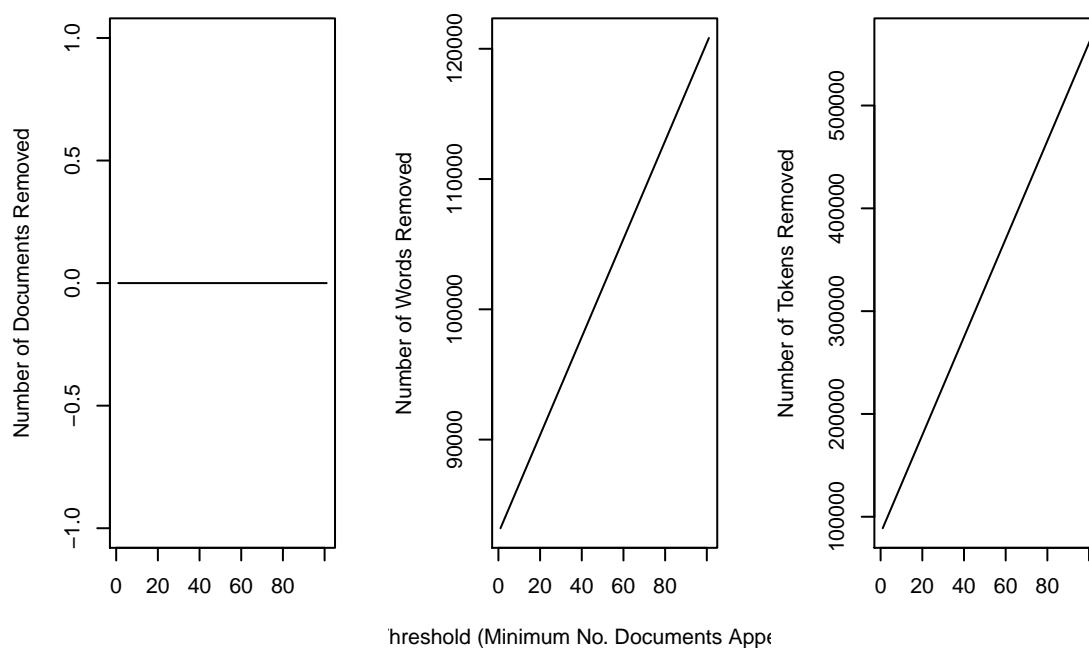
```
## Stemming...
## Creating Output...

# Prep the Documents too #
out <- prepDocuments(processed$documents,
                     processed$vocab,
                     processed$meta)

## Removing 83198 of 123990 terms (83198 of 2298953 tokens) due to frequency
## Your corpus now has 13246 documents, 40792 terms and 2215755 tokens.

# Remove more Terms
# This plot is hard to understand, good luck! #
plotRemoved(processed$documents,
            lower.thresh = seq(1, 200, by = 100))
```

Documents Removed by Thresh Words Removed by Thresho Tokens Removed by Threshc



```
# Prep the Documents some More then... #
out <- prepDocuments(processed$documents,
                     processed$vocab,
                     processed$meta,
                     lower.thresh = 15)

## Removing 114230 of 123990 terms (220223 of 2298953 tokens) due to frequency
## Your corpus now has 13246 documents, 9760 terms and 2078730 tokens.
```

Step 2

Estimate the model(s) – in this case we estimate a $K = 20$ topic, topic model. The prevalence covariates act as additional document-specific covariates. We can also plot out a variety of outcomes to help us make sense of our data.

```
poliblogPrevFit <- stm(out$documents,
                      out$vocab,
                      K=20,
                      prevalence=~rating+s(day),
                      max.em.its=75,
                      data=out$meta,
                      init.type="Spectral",
                      seed=8458159)
```

```
## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##   .....
##   Recovering initialization...
##   .....
## Initialization complete.
## .....
## Completed E-Step (10 seconds).
## Completed M-Step.
## Completing Iteration 1 (approx. per word bound = -7.677)
## .....
## Completed E-Step (10 seconds).
## Completed M-Step.
## Completing Iteration 2 (approx. per word bound = -7.520, relative change = 2.045e-02)
## .....
## Completed E-Step (10 seconds).
## Completed M-Step.
## Completing Iteration 3 (approx. per word bound = -7.483, relative change = 4.961e-03)
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 4 (approx. per word bound = -7.467, relative change = 2.095e-03)
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 5 (approx. per word bound = -7.459, relative change = 1.116e-03)
## Topic 1: obama, mccain, poll, voter, democrat
## Topic 2: peopl, one, like, think, polit
## Topic 3: tax, will, mccain, economi, econom
## Topic 4: one, 're, get, 'll, like
## Topic 5: vote, elect, voter, state, republican
## Topic 6: obama, mccain, campaign, barack, john
## Topic 7: media, news, stori, time, like
## Topic 8: democrat, senat, bill, republican, legisl
## Topic 9: will, price, oil, govern, energi
## Topic 10: court, law, case, right, will
## Topic 11: mccain, campaign, john, million, oil
## Topic 12: iran, nuclear, will, iranian, russia
## Topic 13: palin, sarah, governor, report, state
## Topic 14: israel, terrorist, attack, will, kill
## Topic 15: school, obama, will, educ, work
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, climat, chang
## Topic 18: bush, presid, said, administr, hous
```

```

## Topic 19: hillari, clinton, democrat, obama, will
## Topic 20: will, allah, muslim, say, muhammad
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 6 (approx. per word bound = -7.454, relative change = 6.643e-04)
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 7 (approx. per word bound = -7.451, relative change = 4.272e-04)
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 8 (approx. per word bound = -7.448, relative change = 2.932e-04)
## .....
## Completed E-Step (11 seconds).
## Completed M-Step.
## Completing Iteration 9 (approx. per word bound = -7.447, relative change = 2.069e-04)
## .....
## Completed E-Step (14 seconds).
## Completed M-Step.
## Completing Iteration 10 (approx. per word bound = -7.446, relative change = 1.550e-04)
## Topic 1: obama, mccain, poll, voter, democrat
## Topic 2: peopl, like, think, polit, one
## Topic 3: tax, will, economi, econom, plan
## Topic 4: one, 're, get, don't, like
## Topic 5: vote, elect, voter, republican, state
## Topic 6: obama, mccain, campaign, barack, john
## Topic 7: media, news, stori, time, report
## Topic 8: democrat, senat, republican, bill, vote
## Topic 9: will, oil, energi, price, govern
## Topic 10: court, law, case, right, rule
## Topic 11: mccain, campaign, million, john, report
## Topic 12: iran, nuclear, will, iranian, russia
## Topic 13: palin, sarah, governor, state, alaska
## Topic 14: terrorist, israel, attack, kill, pakistan
## Topic 15: school, work, union, educ, obama
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, chang, climat
## Topic 18: bush, presid, said, administr, hous
## Topic 19: hillari, clinton, will, democrat, obama
## Topic 20: will, muslim, allah, islam, say
## .....
## Completed E-Step (14 seconds).
## Completed M-Step.
## Completing Iteration 11 (approx. per word bound = -7.445, relative change = 1.199e-04)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 12 (approx. per word bound = -7.444, relative change = 9.612e-05)
## .....
## Completed E-Step (11 seconds).
## Completed M-Step.
## Completing Iteration 13 (approx. per word bound = -7.443, relative change = 7.983e-05)

```

```

## .....
## Completed E-Step (11 seconds).
## Completed M-Step.
## Completing Iteration 14 (approx. per word bound = -7.443, relative change = 6.583e-05)
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 15 (approx. per word bound = -7.443, relative change = 5.530e-05)
## Topic 1: obama, mccain, poll, voter, state
## Topic 2: peopl, think, like, polit, dont
## Topic 3: tax, will, econom, economi, plan
## Topic 4: one, 're, like, get, don't
## Topic 5: vote, elect, voter, republican, state
## Topic 6: obama, mccain, campaign, barack, john
## Topic 7: media, news, stori, time, report
## Topic 8: democrat, senat, republican, bill, vote
## Topic 9: oil, will, energi, price, drill
## Topic 10: law, court, case, right, rule
## Topic 11: mccain, campaign, million, money, john
## Topic 12: iran, nuclear, will, world, iranian
## Topic 13: palin, sarah, governor, state, polit
## Topic 14: terrorist, attack, israel, kill, govern
## Topic 15: school, work, union, educ, famili
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, gay, chang
## Topic 18: bush, presid, said, administr, hous
## Topic 19: hillari, clinton, will, democrat, obama
## Topic 20: will, muslim, allah, christian, god
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 16 (approx. per word bound = -7.442, relative change = 4.620e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 17 (approx. per word bound = -7.442, relative change = 3.901e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 18 (approx. per word bound = -7.442, relative change = 3.390e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 19 (approx. per word bound = -7.441, relative change = 3.039e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 20 (approx. per word bound = -7.441, relative change = 2.869e-05)
## Topic 1: obama, poll, mccain, voter, state
## Topic 2: peopl, think, like, dont, polit
## Topic 3: tax, will, econom, economi, plan
## Topic 4: one, 're, like, get, don't
## Topic 5: vote, elect, voter, republican, ballot
## Topic 6: obama, mccain, campaign, barack, john

```

```

## Topic 7: media, news, time, stori, report
## Topic 8: democrat, senat, republican, bill, vote
## Topic 9: oil, energi, will, price, drill
## Topic 10: law, court, case, right, rule
## Topic 11: mccain, campaign, million, money, report
## Topic 12: iran, nuclear, world, will, nation
## Topic 13: palin, sarah, governor, biden, state
## Topic 14: terrorist, attack, israel, kill, govern
## Topic 15: school, work, famili, union, educ
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, gay, chang
## Topic 18: bush, presid, said, administr, hous
## Topic 19: clinton, hillari, will, democrat, obama
## Topic 20: will, muslim, christian, god, allah
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 21 (approx. per word bound = -7.441, relative change = 2.824e-05)
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 22 (approx. per word bound = -7.441, relative change = 2.951e-05)
## .....
## Completed E-Step (7 seconds).
## Completed M-Step.
## Completing Iteration 23 (approx. per word bound = -7.441, relative change = 3.048e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 24 (approx. per word bound = -7.440, relative change = 2.877e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 25 (approx. per word bound = -7.440, relative change = 2.876e-05)
## Topic 1: obama, poll, mccain, voter, state
## Topic 2: peopl, think, like, dont, polit
## Topic 3: tax, will, econom, economi, plan
## Topic 4: one, like, get, 're, don't
## Topic 5: vote, elect, voter, republican, ballot
## Topic 6: obama, mccain, campaign, barack, john
## Topic 7: media, news, time, stori, report
## Topic 8: democrat, senat, republican, bill, vote
## Topic 9: oil, energi, will, price, drill
## Topic 10: law, court, case, rule, right
## Topic 11: campaign, million, mccain, money, report
## Topic 12: iran, nuclear, world, will, nation
## Topic 13: palin, biden, sarah, governor, obama
## Topic 14: terrorist, attack, israel, kill, govern
## Topic 15: school, work, famili, year, union
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, gay, chang
## Topic 18: bush, presid, said, administr, hous
## Topic 19: clinton, hillari, will, democrat, primari
## Topic 20: will, muslim, christian, god, allah

```

```

## .....
## Completed E-Step (7 seconds).
## Completed M-Step.
## Completing Iteration 26 (approx. per word bound = -7.440, relative change = 2.832e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 27 (approx. per word bound = -7.440, relative change = 2.620e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 28 (approx. per word bound = -7.440, relative change = 2.425e-05)
## .....
## Completed E-Step (7 seconds).
## Completed M-Step.
## Completing Iteration 29 (approx. per word bound = -7.439, relative change = 2.315e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 30 (approx. per word bound = -7.439, relative change = 2.253e-05)
## Topic 1: obama, poll, mccain, voter, state
## Topic 2: peopl, think, like, dont, know
## Topic 3: tax, will, econom, economi, govern
## Topic 4: one, like, get, 're, don't
## Topic 5: vote, elect, voter, republican, ballot
## Topic 6: obama, mccain, campaign, barack, john
## Topic 7: media, news, time, stori, report
## Topic 8: democrat, senat, republican, bill, vote
## Topic 9: oil, energi, will, price, drill
## Topic 10: law, court, case, rule, state
## Topic 11: campaign, million, money, report, mccain
## Topic 12: iran, nuclear, world, will, nation
## Topic 13: palin, biden, sarah, governor, obama
## Topic 14: terrorist, attack, israel, kill, govern
## Topic 15: school, work, famili, american, year
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, gay, chang
## Topic 18: bush, presid, said, administr, hous
## Topic 19: clinton, hillari, will, democrat, primari
## Topic 20: church, will, muslim, christian, god
## .....
## Completed E-Step (7 seconds).
## Completed M-Step.
## Completing Iteration 31 (approx. per word bound = -7.439, relative change = 2.208e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 32 (approx. per word bound = -7.439, relative change = 2.111e-05)
## .....
## Completed E-Step (7 seconds).
## Completed M-Step.
## Completing Iteration 33 (approx. per word bound = -7.439, relative change = 2.041e-05)
## .....
## Completed E-Step (8 seconds).

```

```

## Completed M-Step.
## Completing Iteration 34 (approx. per word bound = -7.439, relative change = 2.042e-05)
## .....
## Completed E-Step (7 seconds).
## Completed M-Step.
## Completing Iteration 35 (approx. per word bound = -7.438, relative change = 2.029e-05)
## Topic 1: obama, poll, mccain, voter, state
## Topic 2: peopl, think, like, dont, know
## Topic 3: tax, will, econom, economi, govern
## Topic 4: one, like, get, 're, don't
## Topic 5: vote, elect, voter, ballot, republican
## Topic 6: obama, mccain, campaign, barack, john
## Topic 7: media, news, time, stori, report
## Topic 8: democrat, senat, republican, vote, bill
## Topic 9: oil, energi, will, price, drill
## Topic 10: law, court, case, rule, state
## Topic 11: million, campaign, money, report, new
## Topic 12: iran, nuclear, world, will, nation
## Topic 13: palin, biden, sarah, governor, obama
## Topic 14: terrorist, attack, israel, kill, govern
## Topic 15: school, american, famili, work, year
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, gay, chang
## Topic 18: bush, said, presid, administr, hous
## Topic 19: clinton, hillari, will, democrat, primari
## Topic 20: church, will, muslim, christian, god
## .....
## Completed E-Step (7 seconds).
## Completed M-Step.
## Completing Iteration 36 (approx. per word bound = -7.438, relative change = 1.963e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 37 (approx. per word bound = -7.438, relative change = 1.873e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 38 (approx. per word bound = -7.438, relative change = 1.849e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 39 (approx. per word bound = -7.438, relative change = 1.840e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Completing Iteration 40 (approx. per word bound = -7.438, relative change = 1.788e-05)
## Topic 1: obama, poll, mccain, voter, state
## Topic 2: peopl, think, like, dont, know
## Topic 3: tax, will, econom, economi, govern
## Topic 4: one, like, get, 're, don't
## Topic 5: vote, elect, voter, ballot, republican
## Topic 6: obama, mccain, campaign, barack, john
## Topic 7: media, news, time, stori, report
## Topic 8: democrat, senat, republican, vote, bill

```



```

## Topic 9: oil, energi, will, price, drill
## Topic 10: law, court, case, state, rule
## Topic 11: million, campaign, money, report, new
## Topic 12: iran, nuclear, world, will, nation
## Topic 13: palin, biden, obama, sarah, governor
## Topic 14: terrorist, attack, israel, kill, govern
## Topic 15: american, school, famili, america, work
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, gay, chang
## Topic 18: bush, said, presid, hous, administr
## Topic 19: clinton, hillari, will, democrat, primari
## Topic 20: church, will, muslim, christian, god
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 41 (approx. per word bound = -7.438, relative change = 1.809e-05)
## .....
## Completed E-Step (14 seconds).
## Completed M-Step.
## Completing Iteration 42 (approx. per word bound = -7.437, relative change = 1.782e-05)
## .....
## Completed E-Step (9 seconds).
## Completed M-Step.
## Completing Iteration 43 (approx. per word bound = -7.437, relative change = 1.631e-05)
## .....
## Completed E-Step (10 seconds).
## Completed M-Step.
## Completing Iteration 44 (approx. per word bound = -7.437, relative change = 1.477e-05)
## .....
## Completed E-Step (11 seconds).
## Completed M-Step.
## Completing Iteration 45 (approx. per word bound = -7.437, relative change = 1.350e-05)
## Topic 1: obama, poll, mccain, voter, state
## Topic 2: think, peopl, like, dont, know
## Topic 3: tax, will, econom, economi, govern
## Topic 4: one, like, get, 're, don't
## Topic 5: vote, elect, voter, ballot, state
## Topic 6: obama, mccain, campaign, barack, john
## Topic 7: media, news, time, stori, report
## Topic 8: democrat, senat, republican, vote, bill
## Topic 9: oil, energi, will, price, drill
## Topic 10: law, court, case, state, legal
## Topic 11: million, money, campaign, report, new
## Topic 12: iran, nuclear, world, will, nation
## Topic 13: palin, biden, obama, sarah, governor
## Topic 14: terrorist, attack, israel, kill, govern
## Topic 15: american, school, america, famili, year
## Topic 16: iraq, war, iraqi, troop, militari
## Topic 17: global, warm, abort, gay, chang
## Topic 18: bush, said, presid, hous, administr
## Topic 19: clinton, hillari, will, democrat, primari
## Topic 20: church, will, muslim, christian, god
## .....
## Completed E-Step (13 seconds).

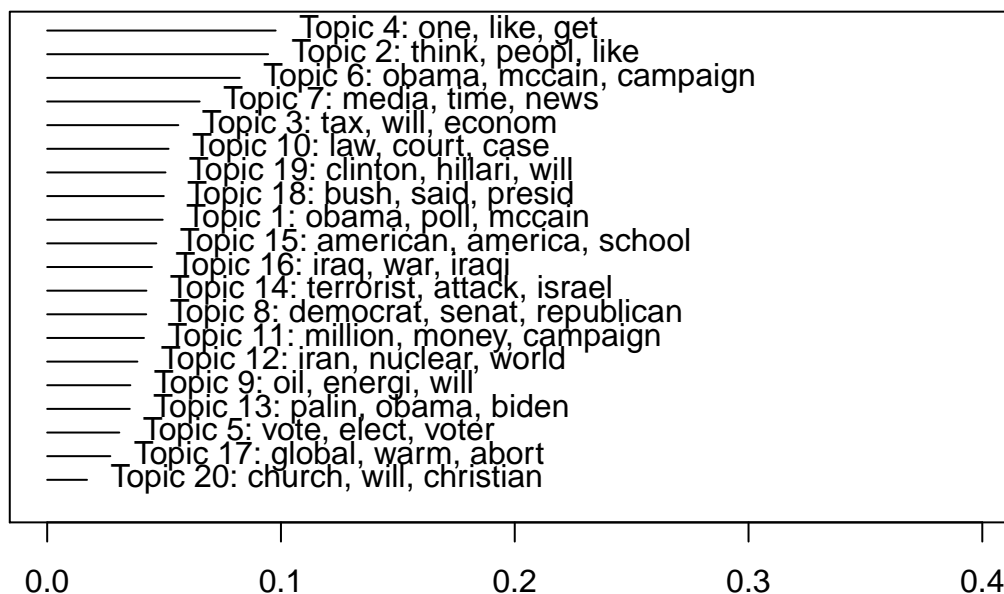
```

```
## Completed M-Step.
## Completing Iteration 46 (approx. per word bound = -7.437, relative change = 1.172e-05)
## .....
## Completed E-Step (10 seconds).
## Completed M-Step.
## Completing Iteration 47 (approx. per word bound = -7.437, relative change = 1.067e-05)
## .....
## Completed E-Step (8 seconds).
## Completed M-Step.
## Model Converged
```

```
# Plot out a short summary of the topics and the proportions of each
```

```
plot(poliblogPrevFit,
     type="summary",
     xlim=c(0,.4))
```

Top Topics



Expected Topic Proportions

```
# PLOT more in depth with more 'top' words #
```

```
plot(poliblogPrevFit,
     type="labels",
     topics=c(1,6,13,18))
```

```
## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'senat, call, mccain', question, support,
## candid, wright, speech, today' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'senat, call, mccain', question, support,
## candid, wright, speech, today' in 'mbcsToSbcs': dot substituted for <80>

## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'senat, call, mccain', question, support,
## candid, wright, speech, today' in 'mbcsToSbcs': dot substituted for <99>
```

```
## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'know, yesterday, today, former, don't, claim,
## 're, ask, rove' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'know, yesterday, today, former, don't, claim,
## 're, ask, rove' in 'mbcsToSbcs': dot substituted for <80>

## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'know, yesterday, today, former, don't, claim,
## 're, ask, rove' in 'mbcsToSbcs': dot substituted for <99>

## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'know, yesterday, today, former, don't, claim,
## 're, ask, rove' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'know, yesterday, today, former, don't, claim,
## 're, ask, rove' in 'mbcsToSbcs': dot substituted for <80>

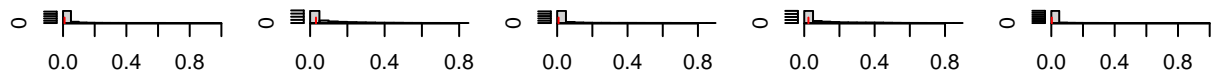
## Warning in text.default(n/2, (i + 0.5), string, family = family, cex =
## text.cex): conversion failure on 'know, yesterday, today, former, don't, claim,
## 're, ask, rove' in 'mbcsToSbcs': dot substituted for <99>
```

<p>Topic 1:</p> <p>a, poll, mccain, voter, state, democrat, lead, percent, vote, among, new, point, number, win, elect, show, republican, will, week, two</p> <p>-----</p> <p>Topic 6:</p> <p>a, mccain, campaign, barack, john, said, say, polici, attack, issu, debat, nat, call, mccain..., question, support, candid, wright, speech, today</p> <p>-----</p> <p>Topic 13:</p> <p>n, obama, biden, sarah, governor, joe, state, will, senat, polit, barack, chicago, alaska, blagojevich, run, report, illinoi, gov, appoint, rezko</p> <p>-----</p> <p>Topic 18:</p> <p>sh, said, presid, hous, sen, white, administr, report, say, think, watch, know, yesterday, today, former, don...t, claim, ...re, ask, rove</p>
--

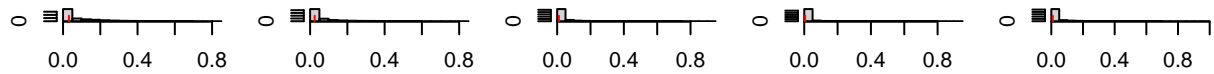
```
# Plot out Topic Distribution (i.e., Topic 2, 6, 7, 15 most prevalent)
plot(poliblogPrevFit, type="hist")
```

Distribution of MAP Estimates of Document–Topic Proportions

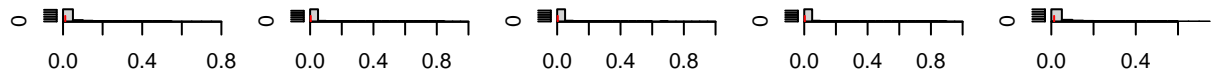
Topic 1: obama, poll, mppic 2: think, peoplpic 3: tax, will, econTopic 4: one, like, gpic 5: vote, elect, v



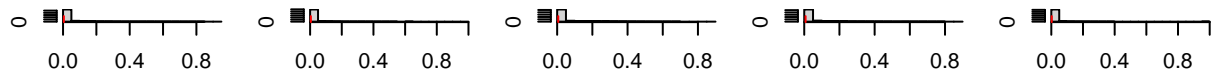
Topic 6: obama, mccain, cpic 7: media, time, r democrat, senat, rTopic 9: oil, energi, wpic 10: law, court, c



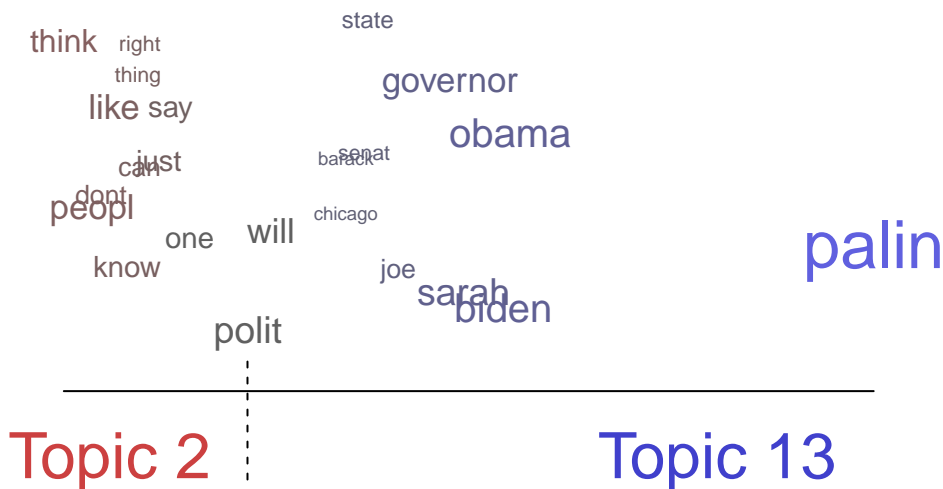
Topic 11: million, money, cic 12: iran, nuclear, ic 13: palin, obama, Topic 14: terrorist, attackTopic 15: american, americ



Topic 16: iraq, war, iric 17: global, warm, ic 18: bush, said, pic 19: clinton, hillarTopic 20: church, will, ch



```
# Plot to see how two topics are discussed differently #
plot(poliblogPrevFit, type="perspectives", topics=c(2,13))
```



Step 3

Select the appropriate number of topics. Note this is the inevitable question and is not answerable. The `manyTopics()` function lets you estimate K topics at a time then take a look to what may work best. The `searchK()` function, with argument $K=0$ will estimate the best fitting topic model based on a mathematical criterion; however, in my experience the estimated number of topics is large and often generates many useless topics.

```
# Topic K Selection; Estimate and then look at, this can take a while #
#storage <- manyTopics(out$documents,
#                        out$vocab,
#                        K=c(5,10,15,20,25),
```

```

#           prevalence=~rating+s(day),
#           data=meta,
#           runs=5)

#storageOutput1 <- storage$out[[1]] # For example, choosing the model with 5 topics

# Then plot it out real nice #
#plot(storageOutput1)

# Lee and Mimmo (2014) Topic Selection strategy #
# Set to 0 and you can then get a guess as to the number of topics using
#storage2 <- searchK(out$documents,
#           out$vocab,
#           K = 0,
#           prevalence =~ rating + s(day),
#           data = meta)

```

Step 4

Interpret the outputs by examining what words tend to pop in what topics. Documents are not always about just one topic but you can print out the top x number of words associated with each topic. This usually provides fairly clear topics unless you have too many filler non-descript words. In which case, go back and further thin down your corpus/DTM text.

```

#####
# Further Interpretation: Print out top words per topic #
#####

# Look at the top words for each Topic #
labelTopics(poliblogPrevFit)

```

```

## Topic 1 Top Words:
##   Highest Prob: obama, poll, mccain, voter, state, democrat, lead
##   FREX: pennsylvania, poll, virginia, elector, percent, margin, survey
##   Lift: crosstab, composit, five-point, rasmussen', zogbi, outlier, quinnipiac
##   Score: poll, obama, mccain, voter, composit, percent, rasmussen
## Topic 2 Top Words:
##   Highest Prob: think, peopl, like, dont, know, just, can
##   FREX: dont, linktocommentspostcount, postcounttb, that, there, cant, hes
##   Lift: alinski, nixonland, digbyi, fuck, perlstein, shit, digbyim
##   Score: dont, linktocommentspostcount, postcounttb, that, hes, alinski, republican
## Topic 3 Top Words:
##   Highest Prob: tax, will, econom, economi, govern, plan, american
##   FREX: tax, mortgag, fanni, economi, insur, econom, billion
##   Lift: cra, gramm', scriptalreadyrequest, acflruncont, allowfullscreentru, allowscriptaccessalw
##   Score: tax, mortgag, bailout, billion, economi, fanni, loan
## Topic 4 Top Words:
##   Highest Prob: one, like, get, 're, don't, doesn't, just
##   FREX: 'll, film, can't, doesn't, eastern, isn't, didn't
##   Lift: bookshelf, cinemat, film', handychristma, hartley, live-chat, log-
##   Score: 're, film, 'll, don't, 've, obama', doesn't
## Topic 5 Top Words:
##   Highest Prob: vote, elect, voter, ballot, state, republican, democrat

```

```

##      FREX: franken, coleman, ballot, registr, chambliss, acorn, absente
##      Lift: franken, lunsford, mahoney, r-co, saxbi, wrongly-reject, brunner
##      Score: franken, coleman, ballot, registr, acorn, voter, vote
## Topic 6 Top Words:
##      Highest Prob: obama, mccain, campaign, barack, john, said, say
##      FREX: mccain, wright, obama, mccain', barack, debat, john
##      Lift: goolsbe, oct, schieffer, mcsame, austan, town-hal, wright
##      Score: mccain, obama, oct, campaign, wright, barack, mccain'
## Topic 7 Top Words:
##      Highest Prob: media, time, news, stori, report, press, new
##      FREX: media, journalist, coverag, matthew, blog, blogger, editor
##      Lift: riell, enquir, tabloid, matthews', dobb, corsi, newsroom
##      Score: media, stori, riell, matthew, fox, journalist, coverag
## Topic 8 Top Words:
##      Highest Prob: democrat, senat, republican, vote, bill, hous, legisl
##      FREX: legisl, reid, pelosi, lieberman, nanci, bipartisan, fisa
##      Lift: clotur, hoyer, legisl, pelosi', reid, steni, chairmanship
##      Score: legisl, republican, lieberman, vote, senat, reid, democrat
## Topic 9 Top Words:
##      Highest Prob: oil, energi, will, price, drill, gas, product
##      FREX: drill, oil, energi, mugab, gas, price, food
##      Lift: mugab, opec, refinери, zimbabwean, ethanol, mbeki, mdc
##      Score: oil, drill, mugab, energi, price, gas, sadc
## Topic 10 Top Words:
##      Highest Prob: law, court, case, state, legal, rule, use
##      FREX: court, legal, illeg, justic, detaine, prosecut, law
##      Lift: taser, boumedien, leonard, souter, habeas, scalia, firearm
##      Score: court, tortur, detaine, taser, law, justic, attorney
## Topic 11 Top Words:
##      Highest Prob: million, money, campaign, report, group, new, fund
##      FREX: donat, fundrais, donor, lobbi, lobbyist, financ, davi
##      Lift: hess, sulzberg, vip, golf, coffer, fcc, bundler
##      Score: hess, donor, fundrais, donat, lobbyist, million, campaign
## Topic 12 Top Words:
##      Highest Prob: iran, nuclear, world, will, nation, state, foreign
##      FREX: russian, russia, iran, iranian, nuclear, ahmadinejad, weapon
##      Lift: dmitri, abkhazia, iaea, ossetia, saakashvili, elbaradei, kremlin
##      Score: iran, iranian, nuclear, russia, russian, elbaradei, israel
## Topic 13 Top Words:
##      Highest Prob: palin, obama, biden, sarah, governor, joe, state
##      FREX: palin, blagojevich, biden, sarah, palin', rezko, governor
##      Lift: wasilla, blago, blagojevich, monegan, palin, "palin, "sarah
##      Score: palin, blagojevich, sarah, biden, palin', alaska, rezko
## Topic 14 Top Words:
##      Highest Prob: terrorist, attack, israel, kill, govern, said, terror
##      FREX: hama, palestinian, musharraf, pakistani, pakistan, laden, taliban
##      Lift: alvaro, aq', ayman, baitullah, benazir, hama, imad
##      Score: hama, israel, fatah, palestinian, isra, pakistan, pakistani
## Topic 15 Top Words:
##      Highest Prob: american, america, school, famili, year, work, women
##      FREX: school, student, educ, children, teacher, parent, young
##      Lift: voucher, bernardin, dohrn, faculti, naacp, cac, student
##      Score: voucher, school, ayer, student, educ, children, chicago
## Topic 16 Top Words:

```

```
## Highest Prob: iraq, war, iraqi, troop, militari, will, forc
## FREX: iraqi, petraeus, iraq, baghdad, troop, maliki, sadr
## Lift: anbar, drawdown, mahdi, moqtada, sadr, al-dabbagh, al-maliki'
## Score: iraq, iraqi, troop, maliki, sadr, afghanistan, militari
## Topic 17 Top Words:
## Highest Prob: global, warm, abort, chang, gay, climat, right
## FREX: abort, warm, climat, gay, scientif, scientist, scienc
## Lift: embryo, fetus, contracept, embryon, ipcc, condom, heterosexu
## Score: abort, embryon, global, gay, climat, warm, marriag
## Topic 18 Top Words:
## Highest Prob: bush, said, presid, hous, sen, white, administr
## FREX: cheney, bush', r-az, rove, perino, digg, waterboard
## Lift: addington, perino, cia', fratto, information", mcclellan, torture"
## Score: addington, perino, bush, sen, cia, rove, bush'
## Topic 19 Top Words:
## Highest Prob: clinton, hillari, will, democrat, primari, obama, campaign
## FREX: hillari, romney, clinton, huckabe, deleg, primari, mitt
## Lift: super-del, super-deleg, bylaw, romney, superdeleg, mcauliff, huckabe
## Score: hillari, clinton, romney, deleg, huckabe, superdeleg, obama
## Topic 20 Top Words:
## Highest Prob: church, will, christian, muslim, god, religi, allah
## FREX: allah, church, vers, muhammad, qur', christian, hage
## Lift: arabicenglish, "blog, allah", ibn, kathir, muhammad', qur'
## Score: qur', allah, sura, muhammad, church, ibn, kathir
```

```
# Thoughts: Find the documents 'most' related to the topic #
```

```
thoughts1 <- findThoughts(poliblogPrevFit,
                          texts = data$documents,
                          n = 4,
                          topics = c(1, 2, 5))
```

```
#####
# Set up a plot for like a paper to convey the key text of the topic #
#####
```

```
two_sent <- c("A new set of Rasmussen polls, all conducted yesterday in the
middle of John McCain's post-convention bounce, suggests that this
race remains close on the state-by-state level",
"Here's our final daily composite of the six major national
tracking polls. Many national polls have coalesced around a
projected 52% popular-vote share for Barack Obama, and it shows in
our composite")
```

```
plotQuote(two_sent,
          width =30,
          main = "Topic 1: 2008 Polling Horse Race")
```

Topic 1: 2008 Polling Horse Race

A new set of Rasmussen polls, all conducted yesterday in the middle of John McCain's post-convention bounce, suggests that this race remains close on the state-by-state level. Here's our national daily composite of the six major national tracking polls. Many national polls have coalesced around a projected 52% popular-vote share for Barack Obama, and it shows in our composite

Step 5

Estimate Metadata Covariate Effects on Topic Selection.

```
# Convert Rating into a Factor #
out$meta$rating <- as.factor(out$meta$rating)

prep <- estimateEffect(1:20 ~ rating , poliblogPrevFit,
                      meta = out$meta, uncertainty = "Global")

# Look at the effects of the metadata 'rating' variable on influence on each topic
summary(prep)
```

```
##
## Call:
## estimateEffect(formula = 1:20 ~ rating, stmobj = poliblogPrevFit,
##   metadata = out$meta, uncertainty = "Global")
##
##
## Topic 1:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.05217    0.00178   29.29 <0.0000000000000002 ***
## ratingLiberal -0.00181    0.00270   -0.67         0.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 2:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
```



```

## (Intercept)    0.05585    0.00178    31.4 <0.0000000000000002 ***
## ratingLiberal  0.07183    0.00263    27.4 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 3:
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.05080   0.00174   29.23 < 0.0000000000000002 ***
## ratingLiberal  0.01191   0.00264    4.51    0.0000066 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 4:
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.13935   0.00209   66.8 <0.0000000000000002 ***
## ratingLiberal -0.11288   0.00292  -38.7 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 5:
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.02391   0.00147   16.3 <0.0000000000000002 ***
## ratingLiberal  0.02573   0.00222   11.6 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 6:
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.06196   0.00151   40.9 <0.0000000000000002 ***
## ratingLiberal  0.02899   0.00242   12.0 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 7:
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.06566   0.00146   44.88 < 0.0000000000000002 ***
## ratingLiberal -0.01242   0.00212   -5.85    0.0000000051 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

##
##
## Topic 8:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.03818    0.00127   30.10 < 0.0000000000000002 ***
## ratingLiberal 0.01135    0.00203    5.58    0.000000024 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 9:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.04612    0.00159   29.04 <0.0000000000000002 ***
## ratingLiberal -0.02130    0.00214   -9.96 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 10:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.05015    0.00169   29.66 <0.0000000000000002 ***
## ratingLiberal 0.00458    0.00250    1.83    0.067 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 11:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.03874    0.00119   32.67 <0.0000000000000002 ***
## ratingLiberal 0.00609    0.00185    3.29    0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 12:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.05201    0.00180   28.83 <0.0000000000000002 ***
## ratingLiberal -0.02309    0.00252   -9.17 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 13:
##

```

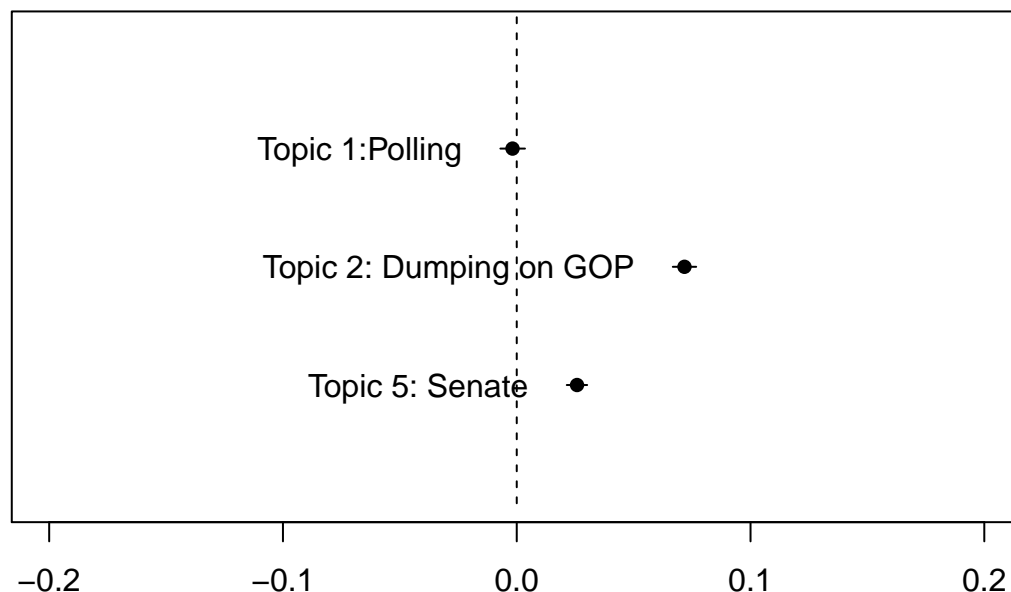
```

## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.04208    0.00147   28.59 <0.0000000000000002 ***
## ratingLiberal -0.00557    0.00220   -2.53      0.011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 14:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.06373    0.00164   39.0 <0.0000000000000002 ***
## ratingLiberal -0.04346    0.00241  -18.1 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 15:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.05282    0.00132   40.05 <0.0000000000000002 ***
## ratingLiberal -0.01714    0.00189   -9.07 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 16:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.04357    0.00170   25.62 <0.0000000000000002 ***
## ratingLiberal  0.00832    0.00265    3.14      0.0017 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 17:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.03090    0.00118   26.14 <0.0000000000000002 ***
## ratingLiberal -0.00142    0.00190   -0.75      0.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 18:
##
## Coefficients:
##           Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.01345    0.00138    9.78 <0.0000000000000002 ***
## ratingLiberal  0.08825    0.00271   32.58 <0.0000000000000002 ***

```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 19:
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.05189    0.00162   32.01 <0.0000000000000002 ***
## ratingLiberal -0.00372    0.00247   -1.51      0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Topic 20:
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)   0.02666    0.00102   26.06 <0.0000000000000002 ***
## ratingLiberal -0.01440    0.00151   -9.54 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# Plot out an Effects Plot the rating variable has on probability of each topic
plot(prepare,
      covariate="rating",
      topics = c(1, 2, 5),
      method="difference",
      model = poliblogPrevFit,
      cov.value1 = "Liberal",
      cov.value2 = "Conservative",
      xlab="More Conservative ... More Liberal",
      main="Effect of Liberal vs. Conservative",
      xlim = c(-.20, .20),
      labeltype = "custom",
      custom.labels=c('Topic 1:Polling', 'Topic 2: Dumping on GOP', 'Topic 5: Senate')
    )
```

Effect of Liberal vs. Conservative



More Conservative ... More Liberal

Step 6 Look at

inter topic correlation.

```
mod.out.corr <- topicCorr(poliblogPrevFit)
```

```
head(mod.out.corr$cor)
```

```
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
## [1,]  1.00000 -0.06783 -0.06333 -0.05717  0.02724  0.02701 -0.12323 -0.07176
## [2,] -0.06783  1.00000 -0.04805 -0.22791 -0.06082  0.01963  0.04414 -0.02912
## [3,] -0.06333 -0.04805  1.00000 -0.11612 -0.06460 -0.04777 -0.15959  0.05747
## [4,] -0.05717 -0.22791 -0.11612  1.00000 -0.10072 -0.08989  0.12701 -0.11349
## [5,]  0.02724 -0.06082 -0.06460 -0.10072  1.00000 -0.11245 -0.09610  0.00000
## [6,]  0.02701  0.01963 -0.04777 -0.08989 -0.11245  1.00000  0.02530 -0.04169
##          [,9]      [,10]     [,11]     [,12]     [,13]     [,14]     [,15]     [,16]
## [1,] -0.07663 -0.14832 -0.06142 -0.10813 -0.03249 -0.11847 -0.11138 -0.10121
## [2,] -0.10644 -0.05121 -0.12428 -0.07791 -0.05094 -0.11768  0.04320 -0.09494
## [3,]  0.08336 -0.10588  0.06666 -0.09687 -0.06417 -0.13063 -0.07895 -0.10256
## [4,] -0.06607 -0.09049 -0.10439 -0.06159 -0.03909 -0.06621  0.00000 -0.08389
## [5,] -0.04357  0.00000  0.01478 -0.08826 -0.02510 -0.07507 -0.05708 -0.09123
## [6,] -0.13907 -0.19683 -0.01167 -0.06523  0.01155 -0.14178 -0.06204 -0.06137
##          [,17]     [,18]     [,19]     [,20]
## [1,] -0.07550 -0.13356  0.19665 -0.06922
## [2,] -0.01322 -0.15553 -0.02865  0.01061
## [3,] -0.04740  0.00000 -0.12794 -0.08828
## [4,] -0.04686 -0.18788  0.00000  0.00000
## [5,] -0.05750 -0.08757 -0.03737 -0.05377
## [6,] -0.11713 -0.06976  0.06247 -0.03262
```

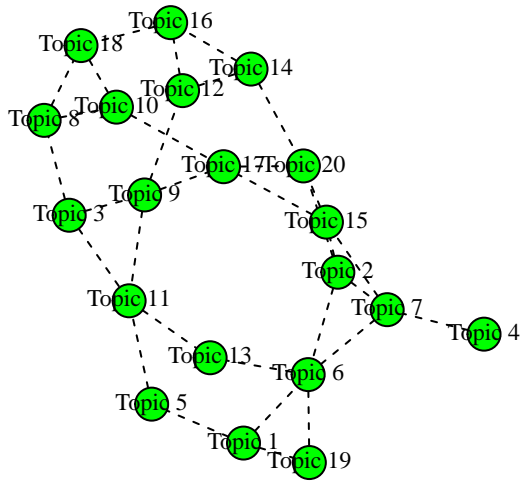
For example Topic 1 and 19 are correlated more than most correlations.

Plot it out real nice

But I then also go and look at some articles from Topic 1 and 19 using the

```
# findThoughts() function above to make sense of the correlation...
```

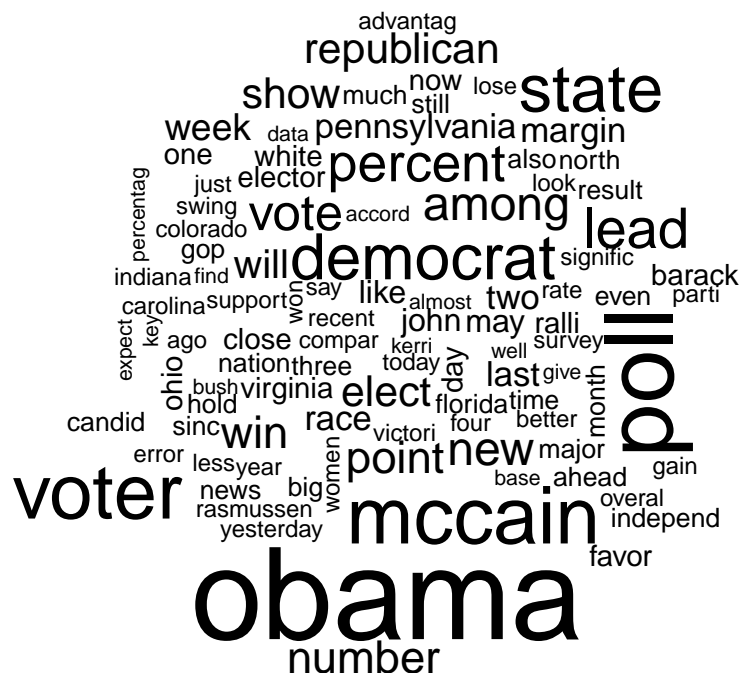
```
plot(mod.out.corr)
```



Step 7

Always gotta present a Wordcloud to impress everyone!

```
cloud(poliblogPrevFit, topic=1)
```



```
# Content Plot -- how 'libs' and 'conservs' use words for a particular topic #
# Re-estimate STM now add in a content argument
```

```
poliblogContent <- stm(out$documents,
  out$vocab,
```

```

K=20,
prevalence=~rating+s(day),
content=~rating,
max.em.its=75,
data=out$meta,
init.type="Spectral",
seed=8458159)

```

```

## Beginning Spectral Initialization
##   Calculating the gram matrix...
##   Finding anchor words...
##   .....
##   Recovering initialization...
##   .....
## Initialization complete.
## .....
## Completed E-Step (10 seconds).
## .....
## Completed M-Step (40 seconds).
## Completing Iteration 1 (approx. per word bound = -7.677)
## .....
## Completed E-Step (10 seconds).
## .....
## Completed M-Step (42 seconds).
## Completing Iteration 2 (approx. per word bound = -7.507, relative change = 2.214e-02)
## .....
## Completed E-Step (11 seconds).
## .....
## Completed M-Step (44 seconds).
## Completing Iteration 3 (approx. per word bound = -7.465, relative change = 5.668e-03)
## .....
## Completed E-Step (11 seconds).
## .....
## Completed M-Step (39 seconds).
## Completing Iteration 4 (approx. per word bound = -7.446, relative change = 2.527e-03)
## .....
## Completed E-Step (10 seconds).
## .....
## Completed M-Step (40 seconds).
## Completing Iteration 5 (approx. per word bound = -7.435, relative change = 1.366e-03)
## Topic 1: gallup, rcp, zogbi, mason-dixon, quinnipiac
## Topic 2: negro, baptiz, oti, slaveri, vatican
## Topic 3: gdp, rebat, "tax, recess, earner
## Topic 4: film', studio, cameo, chad, flick
## Topic 5: franken, absente, coleman, canvass, minnesotan
## Topic 6: goolsbe, mcsame, corsi, austan, schieffer
## Topic 7: matthews', chelsea, enquir, extramarit, riell
## Topic 8: hoyer, clotur, boehner, steni, pelosi'
## Topic 9: unrest, tariff, refineri, africa', musharraf'
## Topic 10: habea, guantánamo, anthrax, detent, boumedien
## Topic 11: iseman, keat, bundler, vicki, greenberg
## Topic 12: georgian, iranian, ossetia, ossetian, elbaradei
## Topic 13: wasilla, blagojevich, anchorag, palin', juneau
## Topic 14: fatah, gaza, khalidi, islamist, palestinian

```

```

## Topic 15: voucher, chrysler, tuition, toyota, uaw
## Topic 16: anbar, nour, al-maliki, petraeus, maliki
## Topic 17: dioxid, embryo, embryon, hansen, nasa
## Topic 18: mcclellan, plame, yoo, interrog, waterboard
## Topic 19: deleg, superdeleg, bylaw, hillari, mcauliff
## Topic 20: mose, biblic, unbeliev, vers, prostrat
## Aspect 1: hoft, instapundit, reportsth, stench, itch
## Aspect 2: ddaythi, itbi, dday, said, ddayth
## .....
## Completed E-Step (9 seconds).
## .....
## Completed M-Step (38 seconds).
## Completing Iteration 6 (approx. per word bound = -7.429, relative change = 8.091e-04)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (36 seconds).
## Completing Iteration 7 (approx. per word bound = -7.426, relative change = 5.106e-04)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (35 seconds).
## Completing Iteration 8 (approx. per word bound = -7.423, relative change = 3.533e-04)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (35 seconds).
## Completing Iteration 9 (approx. per word bound = -7.421, relative change = 2.538e-04)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (36 seconds).
## Completing Iteration 10 (approx. per word bound = -7.420, relative change = 1.852e-04)
## Topic 1: gallup, rcp, zogbi, mason-dixon, quinnipiac
## Topic 2: negro, slaver, falwel, oti, atheist
## Topic 3: gdp, earner, rebat, recess, tax
## Topic 4: film', album, punchlin, heh, flick
## Topic 5: franken, absente, coleman, registr, chambliss
## Topic 6: mcsame, schieffer, goolsbe, austan, wes
## Topic 7: tabloid, enquir, matthews', russert, dowd
## Topic 8: boehner, pelosi', clotur, steni, hoyer
## Topic 9: unrest, refiner, wheat, tariff, grid
## Topic 10: habeas, detaine, anthrax, boumedi, detent
## Topic 11: bundler, lobbyist, iseman, keat, vicki
## Topic 12: georgian, saakashvili, ossetian, dmitri, warhead
## Topic 13: blago, wasilla, palin, monegan, blagojevich
## Topic 14: fatah, gaza, khalidi, islamist, rashid
## Topic 15: voucher, pension, tuition, grandpar, inequ
## Topic 16: anbar, nour, shiit, baghdad, petraeus
## Topic 17: emiss, dioxid, embryo, nasa, embryon
## Topic 18: mcclellan, plame, bolten, interrog, mcclellan'
## Topic 19: deleg, dnc, superdeleg, bylaw, penn
## Topic 20: biblic, unbeliev, mose, prostrat, decre
## Aspect 1: hoft, instapundit, reportsth, stench, itch

```



```

## Aspect 2: ddaythi, unfound, said, long-run, brisl
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (35 seconds).
## Completing Iteration 11 (approx. per word bound = -7.419, relative change = 1.329e-04)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (36 seconds).
## Completing Iteration 12 (approx. per word bound = -7.418, relative change = 9.895e-05)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (36 seconds).
## Completing Iteration 13 (approx. per word bound = -7.417, relative change = 8.199e-05)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (35 seconds).
## Completing Iteration 14 (approx. per word bound = -7.417, relative change = 6.110e-05)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (36 seconds).
## Completing Iteration 15 (approx. per word bound = -7.417, relative change = 4.906e-05)
## Topic 1: gallup, rcp, zogbi, mason-dixon, outlier
## Topic 2: falwel, slaveri, yanke, sublimin, oti
## Topic 3: gdp, earner, tax, recess, economist
## Topic 4: album, film', studio, flick, barry'
## Topic 5: franken, absente, coleman, registr, chambliss
## Topic 6: schieffer, mcpeak, mcsame, goolsbe, austan
## Topic 7: dowd, msnbcs, tabloid, enquir, russert
## Topic 8: boehner, pelosi', clotur, steni, reid
## Topic 9: unrest, grid, musharraf', wheat, shipment
## Topic 10: detent, antonin, anthrax, habea, detainee
## Topic 11: lobbyist, bundler, keat, rupert, coffer
## Topic 12: georgian, ossetian, saakashvili, iranian, warhead
## Topic 13: blago, wasilla, palin', palin, sarah
## Topic 14: fatah, khalidi, islamist, anti-semit, madrid
## Topic 15: voucher, pension, teacher, naacp, grandpar
## Topic 16: nour, shiit, baghdad, al-maliki, petraeus
## Topic 17: dioxid, emiss, embryo, scientif, embryon
## Topic 18: mcclellan, plame, bolten, interrog, mcclellan'
## Topic 19: deleg, superdeleg, dnc, penn, hillari
## Topic 20: biblic, jesus, almighti, decre, mose
## Aspect 1: hoft, instapundit, reportsth, stench, powerlin
## Aspect 2: unfound, said, long-run, brisl, ddaythi
## .....
## Completed E-Step (9 seconds).
## .....
## Completed M-Step (39 seconds).
## Completing Iteration 16 (approx. per word bound = -7.416, relative change = 4.088e-05)
## .....

```

```

## Completed E-Step (8 seconds).
## .....
## Completed M-Step (39 seconds).
## Completing Iteration 17 (approx. per word bound = -7.416, relative change = 3.721e-05)
## .....
## Completed E-Step (9 seconds).
## .....
## Completed M-Step (38 seconds).
## Completing Iteration 18 (approx. per word bound = -7.416, relative change = 3.072e-05)
## .....
## Completed E-Step (9 seconds).
## .....
## Completed M-Step (40 seconds).
## Completing Iteration 19 (approx. per word bound = -7.416, relative change = 2.334e-05)
## .....
## Completed E-Step (8 seconds).
## .....
## Completed M-Step (37 seconds).
## Completing Iteration 20 (approx. per word bound = -7.416, relative change = 1.888e-05)
## Topic 1: gallup, rcp, arg, zogbi, mason-dixon
## Topic 2: schiavo, slaveri, sublimin, empathi, oti
## Topic 3: earner, gdp, tax, economist, recess
## Topic 4: album, mug, film', lyric, snob
## Topic 5: franken, absente, coleman, registr, chambliss
## Topic 6: goolsbe, schieffer, mcsame, mcpeak, factcheck
## Topic 7: dowd, msnbcs, enquir, tabloid, airtim
## Topic 8: pelosi', boehner, clotur, pelosi, steni
## Topic 9: unrest, musharraf', grid, shipment, wheat
## Topic 10: detent, detaine, antonin, anthrax, habea
## Topic 11: lobbyist, bundler, coffer, keat, rupert
## Topic 12: georgian, ossetian, iranian, saakashvili, warhead
## Topic 13: blago, palin', wasilla, palin, sarah
## Topic 14: fatah, anti-semit, khalidi, madrid, aviv
## Topic 15: voucher, grandpar, inequ, uaw, teacher
## Topic 16: nouri, shiit, baghdad, al-maliki, petraeus
## Topic 17: emiss, dioxid, gase, embryo, scientif
## Topic 18: mcclellan, mcclellan', bolten, plame, interrog
## Topic 19: deleg, hillari, dnc, superdeleg, bylaw
## Topic 20: biblic, decre, god", jesus, mose
## Aspect 1: hoft, instapundit, reportsth, stench, powerlin
## Aspect 2: said, unfound, long-run, bristl, abysm
## .....
## Completed E-Step (7 seconds).
## .....
## Completed M-Step (39 seconds).
## Completing Iteration 21 (approx. per word bound = -7.415, relative change = 1.745e-05)
## .....
## Completed E-Step (9 seconds).
## .....
## Completed M-Step (39 seconds).
## Completing Iteration 22 (approx. per word bound = -7.415, relative change = 1.641e-05)
## .....
## Completed E-Step (8 seconds).
## .....

```

```
## Completed M-Step (40 seconds).
## Completing Iteration 23 (approx. per word bound = -7.415, relative change = 1.157e-05)
## .....
## Completed E-Step (9 seconds).
## .....
## Completed M-Step (38 seconds).
## Model Converged
```

```
# Plot out how liberal/conservative blogs speak about same topic #
plot(poliblogContent,
     type="perspectives",
     topics=2,
     main = "Dumping on GOP")
```

