House’s Debate Analysis

Xinyuan Li & Philipp Weisenburger

2023-12-07

## Data Cleaning

#House's Debate Analysis  
rm(list = ls())  
  
#Install the stm package  
#install.packages("stm")  
  
#Load the package   
library(stm)

## stm v1.3.6.1 successfully loaded. See ?stm for help.   
## Papers, resources, and other materials at structuraltopicmodel.com

#Load the data  
setwd("C:/Users/scene/Documents/R/GIRI/project")  
  
# read text and basic cleaning  
speeches\_paths <- c("data/hein-daily/speeches\_107.txt",  
 "data/hein-daily/speeches\_107.txt",  
 "data/hein-daily/speeches\_108.txt",  
 "data/hein-daily/speeches\_108.txt",  
 "data/hein-daily/speeches\_109.txt",  
 "data/hein-daily/speeches\_110.txt",  
 "data/hein-daily/speeches\_110.txt",  
 "data/hein-daily/speeches\_110.txt",  
 "data/hein-daily/speeches\_111.txt",  
 "data/hein-daily/speeches\_112.txt")  
  
maps\_paths <- c("data/hein-daily/107\_SpeakerMap.txt",  
 "data/hein-daily/107\_SpeakerMap.txt",  
 "data/hein-daily/108\_SpeakerMap.txt",  
 "data/hein-daily/108\_SpeakerMap.txt",  
 "data/hein-daily/109\_SpeakerMap.txt",  
 "data/hein-daily/110\_SpeakerMap.txt",  
 "data/hein-daily/110\_SpeakerMap.txt",  
 "data/hein-daily/110\_SpeakerMap.txt",  
 "data/hein-daily/111\_SpeakerMap.txt",  
 "data/hein-daily/112\_SpeakerMap.txt")  
speech\_list <- list()  
map\_list <- list()  
data\_list <- list()  
speech\_line <- list(c(25349,89),  
 c(113464,203),  
 c(22078,231),  
 c(131692,266),  
 c(14950,182),  
 c(9917,115),  
 c(127081,32),  
 c(170707,65),  
 c(69547,64),  
 c(22559,231)  
 )  
  
  
# House speeches for ten fiscal years  
for (i in 1:3){  
 column\_names <- read.table(speeches\_paths[i], sep = "|", nrows = 1, header = FALSE)  
 speech\_list[[i]] <- read.table(speeches\_paths[i], sep = "|",   
 skip = speech\_line[[i]][1], nrows = speech\_line[[i]][2], header = FALSE)  
 colnames(speech\_list[[i]]) <- as.character(unlist(column\_names))  
 map\_list[[i]] <- read.delim(maps\_paths[i], sep = "|", header = TRUE)  
 data\_list[[i]] <- merge(speech\_list[[i]], map\_list[[i]], by = "speech\_id", all.x = TRUE)  
}  
  
column\_names <- read.table(speeches\_paths[3], sep = "|", nrows = 1, header = FALSE)  
speech3a <- read.table(speeches\_paths[3], sep = "|",   
 skip = 20215, nrows = 220, header = FALSE)  
colnames(speech3a) <- as.character(unlist(column\_names))  
map3a <- read.delim(maps\_paths[3], sep = "|", header = TRUE)  
data3a <- merge(speech3a, map3a, by = "speech\_id", all.x = TRUE)  
  
for (i in 4:10){  
 column\_names <- read.table(speeches\_paths[i], sep = "|", nrows = 1, header = FALSE)  
 speech\_list[[i]] <- read.table(speeches\_paths[i], sep = "|",   
 skip = speech\_line[[i]][1], nrows = speech\_line[[i]][2], header = FALSE)  
 colnames(speech\_list[[i]]) <- as.character(unlist(column\_names))  
 map\_list[[i]] <- read.delim(maps\_paths[i], sep = "|", header = TRUE)  
 data\_list[[i]] <- merge(speech\_list[[i]], map\_list[[i]], by = "speech\_id", all.x = TRUE)  
}  
  
#merge  
  
data <- rbind(  
 cbind(data\_list[[1]], year = "2002", crisis='non-crisis'),  
 cbind(data\_list[[2]], year = "2003", crisis='non-crisis'),  
 cbind(data\_list[[3]], year = "2004", crisis='non-crisis'),  
 cbind(data3a, year = "2004", crisis='non-crisis'),  
 cbind(data\_list[[4]], year = "2005", crisis='non-crisis'),  
 cbind(data\_list[[5]], year = "2006", crisis='crisis'),  
 cbind(data\_list[[6]], year = "2007", crisis='crisis'),  
 cbind(data\_list[[7]], year = "2008", crisis='crisis'),  
 cbind(data\_list[[8]], year = "2009", crisis='crisis'),  
 cbind(data\_list[[9]], year = "2010", crisis='crisis'),  
 cbind(data\_list[[10]], year = "2011", crisis='non-crisis'))  
  
data<- na.omit(data)  
  
saveRDS(data, file = "data/data\_house.rds")  
  
  
#Clean text data  
data <- readRDS(file = "data/data\_house.rds")  
processed<- textProcessor(data$speech, #the column that has the text   
 metadata = data, #the name of data set  
 lowercase=TRUE, removestopwords=TRUE, removenumbers=TRUE,   
 removepunctuation=TRUE, stem=TRUE,   
 wordLengths=c(3,Inf), #remove words shorter than 3 letters  
 sparselevel=1, language="en", verbose=TRUE,   
 onlycharacter= FALSE, striphtml=FALSE,   
 customstopwords=c("economic","economics","economical","economies","economy",  
 "america","america’s","american","americans",  
 "speaker","can","minute", "minutes","gentleman","gentlewoman",  
 "just","now","time","madam","miss",  
 "will","that","ladies","gentlemen",  
 "take","what","there","across",  
 "one","year","want","back","say",  
 "yea","nay","get","rollcall","aye","noe",  
 "yield","chairman","committee","committees",  
 "amendment","amendments","amend","amends","amending",  
 "budget","member","members", "vote","votes",  
 "voter","voters", "house", "rule",  
 "come","let","resolution", "debate", "debates",   
 "thank", "ask", "asks", "asking",   
 "distinguish", "distinguished", "may",   
 "question","questions", "work", "works", "working",   
 "bill","bills","congress", "congressional",  
 "process", "senate","senator","senators",  
 "side","sides", "friend","friends",   
 "know","known","knowing","knows",   
 "colleague","colleagues", "think","thinks","thinking",  
 "unanimous","unanimously",  
 "caucus","caucuses", "make","makes")) #custom stopwords

## Building corpus...   
## Converting to Lower Case...   
## Removing punctuation...   
## Removing stopwords...   
## Remove Custom Stopwords...  
## Removing numbers...   
## Stemming...   
## Creating Output...

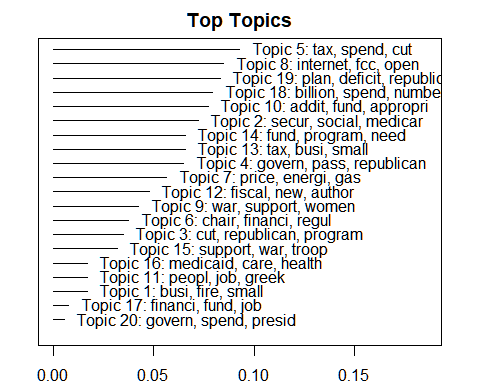
saveRDS(processed, file = "data/processed\_house.rds")

## STM

#House's Debate Analysis  
rm(list = ls())  
  
#Install the stm package  
#install.packages("stm")  
  
#Load the package   
library(stm)  
  
#Load the data  
setwd("C:/Users/scene/Documents/R/GIRI/project")  
processed <- readRDS(file = "data/processed\_house.rds")  
  
#Now we separate our data into neat bits for our analysis   
out <- prepDocuments(processed$documents, processed$vocab, processed$meta)

## Removing 5328 of 10375 terms (5328 of 81833 tokens) due to frequency   
## Removing 2 Documents with No Words   
## Your corpus now has 846 documents, 5047 terms and 76505 tokens.

docs <- out$documents  
vocab <- out$vocab  
meta <-out$meta  
  
# Find the best K value  
set.seed(2008)  
findingk <- searchK(out$documents, out$vocab,   
 K = c(19:21),   
 prevalence =~ crisis \* party, #topic prevalence   
 data = meta, verbose=FALSE)  
  
  
#Fit a model with our given k and topic prevalence and content equations  
  
set.seed(2008)  
First\_STM <- stm(documents = out$documents, vocab = out$vocab,  
 K = 20,   
 prevalence =~ crisis \* party,   
 max.em.its =75, data = out$meta,init.type = "Spectral",   
 verbose = FALSE)  
  
#We built the model!  
saveRDS(First\_STM, file = "data/First\_STM\_house.rds")  
#Now we analyze it  
First\_STM <- readRDS(file = "data/First\_STM\_house.rds")  
  
#First, we identify topics and interpret them  
  
  
#Plot the most prevalent topics in this model  
par(mar=c(2,2,2,2))  
plot(First\_STM)



sageLabels(First\_STM,n=10)

## Topic 1:   
## Marginal Highest Prob: busi, fire, small, servic, nation, peopl, state, also, plan, legisl   
## Marginal FREX: fire, forest, speci, interior, colorado, endang, cleveland, steve, island, manag   
## Marginal Lift: buildup, roadless, wildemess, chairmani, gradual, mall, prestigi, mid, ski, speci   
## Marginal Score: chairmani, hid, kennard, forest, madam, cleveland, speci, celebr, ski, interior   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 2:   
## Marginal Highest Prob: secur, social, medicar, fund, republican, tax, cut, year, drug, surplus   
## Marginal FREX: medicar, social, prescript, drug, trust, surplus, senior, secur, raid, cbo   
## Marginal Lift: fiveyear, dav, doughnut, echo, fantasi, freefal, giveback, hast, hid, horizon   
## Marginal Score: hast, hid, nonmedicar, medicar, prescript, drug, social, raid, trust, surplus   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 3:   
## Marginal Highest Prob: cut, republican, program, billion, veteran, million, nation, tax, year, care   
## Marginal FREX: slash, david, spratt, blueprint, veteran, wealthiest, cut, valu, child, back   
## Marginal Lift: clever, hay, heath, oregonian, prioritiesa, scrip, illadvis, toughen, sbas, skew   
## Marginal Score: hay, hid, nonmedicar, medicaid, david, spratt, slash, veteran, cut, blueprint   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 4:   
## Marginal Highest Prob: govern, pass, republican, peopl, spend, shut, today, that, shutdown, last   
## Marginal FREX: shutdown, shut, that, cant, yesterday, pass, adjourn, govern, parti, reid   
## Marginal Lift: adjourn, zeppelin, unveil, reid, scout, freshmen, circul, hemorrhag, shes, steni   
## Marginal Score: adjourn, hid, kennard, nonmedicar, shutdown, shut, reid, cant, didnt, what   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 5:   
## Marginal Highest Prob: tax, spend, cut, balanc, blue, govern, percent, growth, increas, dog   
## Marginal FREX: blue, dog, wast, growth, balanc, relief, tax, spend, budget, dig   
## Marginal Lift: dilig, toomey, tricar, rsc, phasein, dog, blue, undertax, ammunit, runaway   
## Marginal Score: dilig, hid, kennard, nonmedicar, tax, dog, blue, toomey, spend, rsc   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 6:   
## Marginal Highest Prob: chair, financi, regul, market, requir, end, swap, madam, risk, deriv   
## Marginal FREX: swap, deriv, chair, user, risk, exchang, madam, market, exempt, financi   
## Marginal Lift: adler, cras, neglig, columbia, euronext, issuer, kanjorski, liabl, mccarthi, nrsros   
## Marginal Score: columbia, kennard, swap, madam, deriv, bailout, regul, collater, overthecount, cfpa   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 7:   
## Marginal Highest Prob: price, energi, gas, oil, bay, democrat, chesapeak, countri, new, today   
## Marginal FREX: chesapeak, gas, ill, bay, oil, price, drill, gasolin, watertrail, pump   
## Marginal Lift: continent, gag, americanmad, cellulos, gallup, ill, illconsid, offlimit, pellet, ret   
## Marginal Score: hid, ill, kennard, nonmedicar, chesapeak, gas, drill, oil, watertrail, bay   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 8:   
## Marginal Highest Prob: internet, fcc, open, regul, consum, rule, innov, provid, compani, broadband   
## Marginal FREX: fcc, internet, broadband, fccs, innov, parliamentari, disapprov, communic, content, open   
## Marginal Lift: comcast, deregulatori, dna, ebay, googl, netflix, walden, wireless, amazon, ancillari   
## Marginal Score: hid, kennard, nonmedicar, parliamentari, internet, fcc, broadband, fccs, regul, cra   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 9:   
## Marginal Highest Prob: war, support, women, men, troop, peopl, countri, world, unit, nation   
## Marginal FREX: men, inquiri, love, soldier, women, mission, uniform, hussein, gulf, son   
## Marginal Lift: chronicl, demean, gear, guidanc, inquiri, matthew, proverb, speedi, teeth, energet   
## Marginal Score: hid, inquiri, kennard, nonmedicar, saddam, hussein, men, war, love, women   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 10:   
## Marginal Highest Prob: addit, fund, appropri, million, earmark, legisl, major, includ, provid, billion   
## Marginal FREX: addit, earmark, omnibus, obey, appropri, staff, contain, hous, subcommitte, research   
## Marginal Lift: headquart, meth, withhold, airdrop, explanatori, nabor, addit, archiv, chastis, darfur   
## Marginal Score: addit, hid, kennard, nonmedicar, earmark, obey, omnibus, subcommitte, formula, fund   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 11:   
## Marginal Highest Prob: peopl, job, greek, state, mani, greec, countri, independ, year, nation   
## Marginal FREX: greek, greec, nay, independ, ireland, ancient, polic, celebr, nuclear, northern   
## Marginal Lift: gothic, ancestor, britain, coffer, colonist, cyprus, fledgl, greek, greekamerican, hellen   
## Marginal Score: hid, kennard, nay, nonmedicar, greek, greec, ireland, turkish, celebr, ancient   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 12:   
## Marginal Highest Prob: fiscal, new, author, outlay, michigan, vote, res, concurr, consider, year   
## Marginal FREX: michigan, outlay, author, fiscal, new, concurr, consider, roll, detain, vote   
## Marginal Lift: michigan, rolical, nos, poster, outlay, detain, appro, herebi, author, propriat   
## Marginal Score: hid, kennard, michigan, nonmedicar, outlay, fiscal, author, detain, rolical, new   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 13:   
## Marginal Highest Prob: tax, busi, small, ask, day, thing, pay, peopl, like, person   
## Marginal FREX: ask, code, person, owner, sell, small, sometim, busi, death, ration   
## Marginal Lift: dispens, lehigh, patienc, comedian, imposit, layer, lobbyist, perkiomen, dread, outhous   
## Marginal Score: dispens, hid, kennard, nonmedicar, tax, suspens, quorum, restaur, farm, sell   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 14:   
## Marginal Highest Prob: fund, program, need, educ, provid, increas, veteran, nation, support, billion   
## Marginal FREX: frank, child, educ, highway, program, grant, fund, behind, cbc, black   
## Marginal Lift: postur, frank, elev, inmat, pertain, selfsuffici, socioeconom, americorp, reintegr, dismay   
## Marginal Score: frank, hid, kennard, nonmedicar, cbc, fund, educ, highway, teacher, veteran   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 15:   
## Marginal Highest Prob: support, war, troop, famili, nation, iraq, men, women, presid, forc   
## Marginal FREX: brave, tonight, saddam, god, prayer, command, arm, bless, troop, hussein   
## Marginal Lift: commenc, disrespect, routin, sympathi, tyrant, unanim, hampshir, tile, aleadi, badger   
## Marginal Score: hampshir, hid, kennard, nonmedicar, saddam, hussein, prayer, iraq, tonight, gratitud   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 16:   
## Marginal Highest Prob: medicaid, care, health, cut, state, program, peopl, billion, million, cost   
## Marginal FREX: medicaid, illinoi, coverag, uninsur, nurs, care, health, hospit, insur, vulner   
## Marginal Lift: awak, bottl, illinoi, payer, rangel, renaiss, tort, uncompens, threequart, gold   
## Marginal Score: hid, illinoi, kennard, nonmedicar, medicaid, uninsur, coverag, medicar, rangel, prescript   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 17:   
## Marginal Highest Prob: financi, fund, job, peopl, billion, spend, million, democrat, republican, includ   
## Marginal FREX: dealer, financi, yea, bailout, tarp, loan, system, arbitr, manufactur, fed   
## Marginal Lift: franchis, statutorili, yea, arbitr, chrysler, takeov, fanni, freddi, yemen, hook   
## Marginal Score: hid, kennard, nonmedicar, yea, dealer, bailout, tarp, arbitr, regul, madam   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 18:   
## Marginal Highest Prob: billion, spend, number, money, defens, year, peopl, confer, pay, need   
## Marginal FREX: wisconsin, confer, defens, iowa, shell, number, correct, game, billion, discretionari   
## Marginal Lift: bypass, flatlin, longrang, macroeconom, mug, placehold, riverboat, tradeoff, wisconsin, committeereport   
## Marginal Score: hid, kennard, nonmedicar, wisconsin, shell, scare, conceal, omb, chairman, baselin   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 19:   
## Marginal Highest Prob: plan, deficit, republican, veteran, cut, tax, democrat, billion, care, spend   
## Marginal FREX: plan, texa, deficit, veteran, south, carolina, show, trillion, substitut, surplus   
## Marginal Lift: dumb, hoyer, raw, sidelin, swing, principi, fascist, icit, undemocrat, revert   
## Marginal Score: hid, hoyer, kennard, nonmedicar, veteran, deficit, surplus, substitut, tax, spratt   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 20:   
## Marginal Highest Prob: govern, spend, presid, cut, peopl, republican, countri, need, shut, fund   
## Marginal FREX: parenthood, shut, shutdown, agreement, abort, tea, cancer, rider, boehner, awar   
## Marginal Lift: cervic, addon, awar, contracept, defund, libya, distract, boehner, parenthood, bewar   
## Marginal Score: awar, hid, kennard, nonmedicar, shutdown, internet, parenthood, shut, boehner, libya   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##

sink("output/sageLabels-selected.txt", append=FALSE, split=TRUE)  
print(sageLabels(First\_STM,n=10))

## Topic 1:   
## Marginal Highest Prob: busi, fire, small, servic, nation, peopl, state, also, plan, legisl   
## Marginal FREX: fire, forest, speci, interior, colorado, endang, cleveland, steve, island, manag   
## Marginal Lift: buildup, roadless, wildemess, chairmani, gradual, mall, prestigi, mid, ski, speci   
## Marginal Score: chairmani, hid, kennard, forest, madam, cleveland, speci, celebr, ski, interior   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 2:   
## Marginal Highest Prob: secur, social, medicar, fund, republican, tax, cut, year, drug, surplus   
## Marginal FREX: medicar, social, prescript, drug, trust, surplus, senior, secur, raid, cbo   
## Marginal Lift: fiveyear, dav, doughnut, echo, fantasi, freefal, giveback, hast, hid, horizon   
## Marginal Score: hast, hid, nonmedicar, medicar, prescript, drug, social, raid, trust, surplus   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 3:   
## Marginal Highest Prob: cut, republican, program, billion, veteran, million, nation, tax, year, care   
## Marginal FREX: slash, david, spratt, blueprint, veteran, wealthiest, cut, valu, child, back   
## Marginal Lift: clever, hay, heath, oregonian, prioritiesa, scrip, illadvis, toughen, sbas, skew   
## Marginal Score: hay, hid, nonmedicar, medicaid, david, spratt, slash, veteran, cut, blueprint   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 4:   
## Marginal Highest Prob: govern, pass, republican, peopl, spend, shut, today, that, shutdown, last   
## Marginal FREX: shutdown, shut, that, cant, yesterday, pass, adjourn, govern, parti, reid   
## Marginal Lift: adjourn, zeppelin, unveil, reid, scout, freshmen, circul, hemorrhag, shes, steni   
## Marginal Score: adjourn, hid, kennard, nonmedicar, shutdown, shut, reid, cant, didnt, what   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 5:   
## Marginal Highest Prob: tax, spend, cut, balanc, blue, govern, percent, growth, increas, dog   
## Marginal FREX: blue, dog, wast, growth, balanc, relief, tax, spend, budget, dig   
## Marginal Lift: dilig, toomey, tricar, rsc, phasein, dog, blue, undertax, ammunit, runaway   
## Marginal Score: dilig, hid, kennard, nonmedicar, tax, dog, blue, toomey, spend, rsc   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 6:   
## Marginal Highest Prob: chair, financi, regul, market, requir, end, swap, madam, risk, deriv   
## Marginal FREX: swap, deriv, chair, user, risk, exchang, madam, market, exempt, financi   
## Marginal Lift: adler, cras, neglig, columbia, euronext, issuer, kanjorski, liabl, mccarthi, nrsros   
## Marginal Score: columbia, kennard, swap, madam, deriv, bailout, regul, collater, overthecount, cfpa   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 7:   
## Marginal Highest Prob: price, energi, gas, oil, bay, democrat, chesapeak, countri, new, today   
## Marginal FREX: chesapeak, gas, ill, bay, oil, price, drill, gasolin, watertrail, pump   
## Marginal Lift: continent, gag, americanmad, cellulos, gallup, ill, illconsid, offlimit, pellet, ret   
## Marginal Score: hid, ill, kennard, nonmedicar, chesapeak, gas, drill, oil, watertrail, bay   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 8:   
## Marginal Highest Prob: internet, fcc, open, regul, consum, rule, innov, provid, compani, broadband   
## Marginal FREX: fcc, internet, broadband, fccs, innov, parliamentari, disapprov, communic, content, open   
## Marginal Lift: comcast, deregulatori, dna, ebay, googl, netflix, walden, wireless, amazon, ancillari   
## Marginal Score: hid, kennard, nonmedicar, parliamentari, internet, fcc, broadband, fccs, regul, cra   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 9:   
## Marginal Highest Prob: war, support, women, men, troop, peopl, countri, world, unit, nation   
## Marginal FREX: men, inquiri, love, soldier, women, mission, uniform, hussein, gulf, son   
## Marginal Lift: chronicl, demean, gear, guidanc, inquiri, matthew, proverb, speedi, teeth, energet   
## Marginal Score: hid, inquiri, kennard, nonmedicar, saddam, hussein, men, war, love, women   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 10:   
## Marginal Highest Prob: addit, fund, appropri, million, earmark, legisl, major, includ, provid, billion   
## Marginal FREX: addit, earmark, omnibus, obey, appropri, staff, contain, hous, subcommitte, research   
## Marginal Lift: headquart, meth, withhold, airdrop, explanatori, nabor, addit, archiv, chastis, darfur   
## Marginal Score: addit, hid, kennard, nonmedicar, earmark, obey, omnibus, subcommitte, formula, fund   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 11:   
## Marginal Highest Prob: peopl, job, greek, state, mani, greec, countri, independ, year, nation   
## Marginal FREX: greek, greec, nay, independ, ireland, ancient, polic, celebr, nuclear, northern   
## Marginal Lift: gothic, ancestor, britain, coffer, colonist, cyprus, fledgl, greek, greekamerican, hellen   
## Marginal Score: hid, kennard, nay, nonmedicar, greek, greec, ireland, turkish, celebr, ancient   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 12:   
## Marginal Highest Prob: fiscal, new, author, outlay, michigan, vote, res, concurr, consider, year   
## Marginal FREX: michigan, outlay, author, fiscal, new, concurr, consider, roll, detain, vote   
## Marginal Lift: michigan, rolical, nos, poster, outlay, detain, appro, herebi, author, propriat   
## Marginal Score: hid, kennard, michigan, nonmedicar, outlay, fiscal, author, detain, rolical, new   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 13:   
## Marginal Highest Prob: tax, busi, small, ask, day, thing, pay, peopl, like, person   
## Marginal FREX: ask, code, person, owner, sell, small, sometim, busi, death, ration   
## Marginal Lift: dispens, lehigh, patienc, comedian, imposit, layer, lobbyist, perkiomen, dread, outhous   
## Marginal Score: dispens, hid, kennard, nonmedicar, tax, suspens, quorum, restaur, farm, sell   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 14:   
## Marginal Highest Prob: fund, program, need, educ, provid, increas, veteran, nation, support, billion   
## Marginal FREX: frank, child, educ, highway, program, grant, fund, behind, cbc, black   
## Marginal Lift: postur, frank, elev, inmat, pertain, selfsuffici, socioeconom, americorp, reintegr, dismay   
## Marginal Score: frank, hid, kennard, nonmedicar, cbc, fund, educ, highway, teacher, veteran   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 15:   
## Marginal Highest Prob: support, war, troop, famili, nation, iraq, men, women, presid, forc   
## Marginal FREX: brave, tonight, saddam, god, prayer, command, arm, bless, troop, hussein   
## Marginal Lift: commenc, disrespect, routin, sympathi, tyrant, unanim, hampshir, tile, aleadi, badger   
## Marginal Score: hampshir, hid, kennard, nonmedicar, saddam, hussein, prayer, iraq, tonight, gratitud   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 16:   
## Marginal Highest Prob: medicaid, care, health, cut, state, program, peopl, billion, million, cost   
## Marginal FREX: medicaid, illinoi, coverag, uninsur, nurs, care, health, hospit, insur, vulner   
## Marginal Lift: awak, bottl, illinoi, payer, rangel, renaiss, tort, uncompens, threequart, gold   
## Marginal Score: hid, illinoi, kennard, nonmedicar, medicaid, uninsur, coverag, medicar, rangel, prescript   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 17:   
## Marginal Highest Prob: financi, fund, job, peopl, billion, spend, million, democrat, republican, includ   
## Marginal FREX: dealer, financi, yea, bailout, tarp, loan, system, arbitr, manufactur, fed   
## Marginal Lift: franchis, statutorili, yea, arbitr, chrysler, takeov, fanni, freddi, yemen, hook   
## Marginal Score: hid, kennard, nonmedicar, yea, dealer, bailout, tarp, arbitr, regul, madam   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 18:   
## Marginal Highest Prob: billion, spend, number, money, defens, year, peopl, confer, pay, need   
## Marginal FREX: wisconsin, confer, defens, iowa, shell, number, correct, game, billion, discretionari   
## Marginal Lift: bypass, flatlin, longrang, macroeconom, mug, placehold, riverboat, tradeoff, wisconsin, committeereport   
## Marginal Score: hid, kennard, nonmedicar, wisconsin, shell, scare, conceal, omb, chairman, baselin   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 19:   
## Marginal Highest Prob: plan, deficit, republican, veteran, cut, tax, democrat, billion, care, spend   
## Marginal FREX: plan, texa, deficit, veteran, south, carolina, show, trillion, substitut, surplus   
## Marginal Lift: dumb, hoyer, raw, sidelin, swing, principi, fascist, icit, undemocrat, revert   
## Marginal Score: hid, hoyer, kennard, nonmedicar, veteran, deficit, surplus, substitut, tax, spratt   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##   
## Topic 20:   
## Marginal Highest Prob: govern, spend, presid, cut, peopl, republican, countri, need, shut, fund   
## Marginal FREX: parenthood, shut, shutdown, agreement, abort, tea, cancer, rider, boehner, awar   
## Marginal Lift: cervic, addon, awar, contracept, defund, libya, distract, boehner, parenthood, bewar   
## Marginal Score: awar, hid, kennard, nonmedicar, shutdown, internet, parenthood, shut, boehner, libya   
##   
## Topic Kappa:   
## Kappa with Baseline:   
##

sink()  
  
#Now that we understand topics, we can see if their prevalence differs based on certain factors  
  
  
#Specify a model that assesses the prevalence of topics   
prep <- estimateEffect(c(1:20) ~ crisis \* party,  
 First\_STM,meta = meta, uncertainty = "Global")  
saveRDS(prep, file = "data/prep\_house.rds")

## Analysis

#House's Debate Analysis  
rm(list = ls())  
  
#Install the stm package  
#install.packages("stm")  
#install.packages("igraph")  
  
#Load the package   
library(stm)  
library(igraph)

## Warning: 程辑包'igraph'是用R版本4.3.2 来建造的

##   
## 载入程辑包：'igraph'

## The following objects are masked from 'package:stats':  
##   
## decompose, spectrum

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

library(ggplot2)  
library(dplyr)

##   
## 载入程辑包：'dplyr'

## The following objects are masked from 'package:igraph':  
##   
## as\_data\_frame, groups, union

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

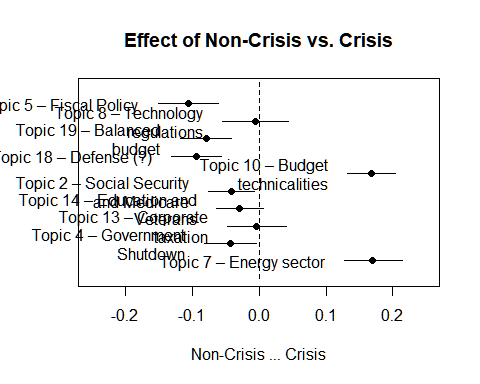
set.seed(2008)  
#Load the data  
setwd("C:/Users/scene/Documents/R/GIRI/project")  
First\_STM <- readRDS(file = "data/First\_STM\_house.rds")  
prep <- readRDS(file = "data/prep\_house.rds")  
processed <- readRDS(file = "data/processed\_house.rds")  
  
#Now we separate our data into neat bits for our analysis   
out <- prepDocuments(processed$documents, processed$vocab, processed$meta)

## Removing 5328 of 10375 terms (5328 of 81833 tokens) due to frequency   
## Removing 2 Documents with No Words   
## Your corpus now has 846 documents, 5047 terms and 76505 tokens.

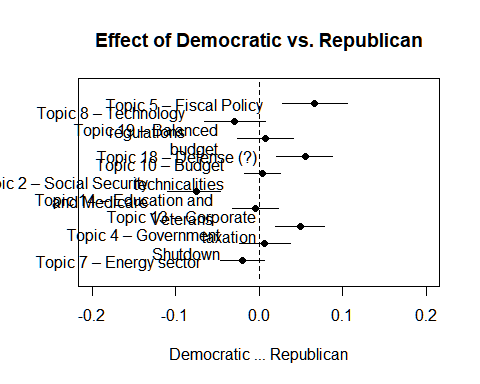
docs <- out$documents  
vocab <- out$vocab  
meta <-out$meta  
meta$year <- as.integer(meta$year)  
  
topic\_number = c(5,8,19,18,10,2,14,13,4,7)  
topic\_labels = c("Topic 5 – Fiscal Policy",   
 "Topic 8 – Technology regulations",   
 "Topic 19 – Balanced budget",   
 "Topic 18 – Defense (?)",   
 "Topic 10 – Budget technicalities",   
 "Topic 2 – Social Security and Medicare",   
 "Topic 14 – Education and Veterans",   
 "Topic 13 – Corporate taxation",   
 "Topic 4 – Government Shutdown",   
 "Topic 7 – Energy sector" )  
  
#Now we plot the effect of CRISIS  
Sys.setlocale(locale="en\_US.UTF-8")

## [1] "LC\_COLLATE=en\_US.UTF-8;LC\_CTYPE=en\_US.UTF-8;LC\_MONETARY=en\_US.UTF-8;LC\_NUMERIC=C;LC\_TIME=en\_US.UTF-8"

plot(prep,   
 covariate = "crisis",   
 topics = topic\_number,  
 model = First\_STM, method = "difference",   
 cov.value1 = 'crisis',   
 cov.value2 = 'non-crisis',   
 xlab = "Non-Crisis ... Crisis",   
 main = "Effect of Non-Crisis vs. Crisis",   
 xlim = c(-.25, .25),  
 labeltype = "custom",   
 custom.labels = topic\_labels  
 )



#Now we plot the effect of PARTY  
plot(prep,   
 covariate = "party",  
 topics = topic\_number,  
 model = First\_STM, method = "difference",   
 cov.value1 = 'R',   
 cov.value2 = 'D',   
 xlab = "Democratic ... Republican",   
 main = "Effect of Democratic vs. Republican",  
 xlim = c(-.2, .2),  
 labeltype = "custom",   
 custom.labels = topic\_labels  
 )



#Trend of top topics  
topic\_distributions <- First\_STM$theta  
topic\_distribution\_with\_year <- cbind(meta$year, topic\_distributions)  
topic\_distribution\_df <- as.data.frame(topic\_distributions)  
topic\_distribution\_df$year = meta$year  
  
for (i in seq\_along(topic\_number)) {  
 topic\_index <- topic\_number[i]  
 topic\_label <- topic\_labels[i]  
 yearly\_distribution <- topic\_distribution\_df %>%  
 group\_by(year) %>%  
 summarise(Average = mean(get(paste0("V", topic\_index))))  
 p <- ggplot(yearly\_distribution, aes(x = year, y = Average)) +  
 geom\_line() +  
 geom\_point() +  
 scale\_x\_continuous(breaks = yearly\_distribution$year) +  
 labs(x = "Fiscal Year", y = "Average Topic Distribution",  
 title = paste("Trend of", topic\_label)) +  
 theme\_classic()  
 ggsave(filename = paste0("t", topic\_index, ".png"), plot = p, width = 6, height = 4)  
}  
  
#Speech Count by States  
meta\_count <- meta %>% count(state) %>% filter(n > 15) %>% arrange(desc(n))  
ggplot(meta\_count, aes(x = reorder(state, n), y = n)) +  
 geom\_bar(stat = "identity") +  
 labs(x = "State", y = "Speech Count (>15)") +  
 coord\_flip() +  
 theme\_classic()

