

Reliable Sources - Analyzing Open-Ended Responses

Patrick Kraft

March 2, 2020

Load data

```
### Packages
library(here)
library(readr)
library(tidyverse)
library(stm)

### Load cleaned data, combine oe responses, omit empty responses
df <- read_csv(here("data/immigration_20191219_clean.csv")) %>%
  mutate(college = as.numeric(educ %in% c("Completed post-graduate or professional school, with degree",
    "Graduated 2-year college", "Graduated 4-year college")),
    college_pid = college * pid_rep,
    oe = paste(taxes_oe, jobs_oe),
    assigned = as.numeric(condition == "assigned"),
    choice = as.numeric(condition == "choice")) %>% # combine both items
  filter(!is.na(age) & !is.na(college) & !is.na(pid_rep))
```

Estimate Structural Topic Model

```
### fit structural topic model

## process for stm
processed <- textProcessor(df$oe, metadata = df[,c("age", "college", "pid_rep", "college_pid",
  "assigned", "choice")])
out <- prepDocuments(processed$documents, processed$vocab, processed$meta)

## stm fit with 49 topics
stm_fit <- stm(out$documents, out$vocab, prevalence = as.matrix(out$meta),
  K = 0, init.type = "Spectral", seed=12345)
```

Topic overview

```
summary(stm_fit)
```

```
## A topic model with 56 topics, 598 documents and a 992 word dictionary.
## Topic 1 Top Words:
##   Highest Prob: servic, use, immigr, social, health, other, also
##   FREX: servic, use, health, social, resid, other, afraid
##   Lift: afraid, servic, use, resid, health, incred, nose
##   Score: servic, use, afraid, health, social, resid, other
```

```

## Topic 2 Top Words:
##   Highest Prob: immigr, job, get, system, take, citizen, illeg
##   FREX: system, group, get, doesnt, citizen, stori, avail
##   Lift: automat, group, candid, total, degre, doesnt, huge
##   Score: automat, group, system, citizen, illeg, doesnt, stori
## Topic 3 Top Words:
##   Highest Prob: will, probabl, may, live, abl, often, mani
##   FREX: will, probabl, may, abl, success, often, wont
##   Lift: bias, feet, somewhat, will, probabl, happi, forth
##   Score: will, probabl, bias, may, abl, wont, happi
## Topic 4 Top Words:
##   Highest Prob: economi, immigr, add, contribut, mani, break, way
##   FREX: economi, add, break, contribut, show, dollar, way
##   Lift: break, economi, add, show, fee, subsidi, self
##   Score: economi, break, add, show, contribut, fee, dollar
## Topic 5 Top Words:
##   Highest Prob: often, amount, anyon, ever, food, shop, howev
##   FREX: ever, environ, huge, hold, shop, amount, groceri
##   Lift: environ, brick, cart, facilit, wic, energi, notic
##   Score: brick, ever, environ, cart, hold, steal, wic
## Topic 6 Top Words:
##   Highest Prob: theyr, skill, labor, lot, low, manual, worker
##   FREX: theyr, skill, manual, labor, exist, low, let
##   Lift: central, theyr, manual, clean, refus, unless, skill
##   Score: theyr, central, skill, manual, labor, let, doctor
## Topic 7 Top Words:
##   Highest Prob: job, immigr, countri, wage, idea, employ, perform
##   FREX: perform, constant, idea, wage, plus, net, arriv
##   Lift: constant, perform, pattern, agricultur, price, exploit, commod
##   Score: constant, perform, interest, bottom, idea, scale, background
## Topic 8 Top Words:
##   Highest Prob: peopl, countri, probabl, your, born, live, just
##   FREX: your, probabl, born, stay, corpor, yes, countri
##   Lift: corpor, forward, your, vet, move, toss, problem
##   Score: corpor, your, probabl, stay, big, move, born
## Topic 9 Top Words:
##   Highest Prob: job, peopl, will, live, come, time, immigr
##   FREX: time, often, difficult, live, peopl, popul, someon
##   Lift: creator, rare, difficult, janitori, circumst, tax-pay, sum
##   Score: creator, often, time, difficult, popul, rare, circumst
## Topic 10 Top Words:
##   Highest Prob: immigr, fair, share, come, job, citizen, need
##   FREX: fair, share, open, establish, small, citizen, hold
##   Lift: cut, share, fair, sell, prefer, tough, opt
##   Score: share, fair, cut, hold, small, establish, open
## Topic 11 Top Words:
##   Highest Prob: money, spend, earn, nice, increas, economi, back
##   FREX: money, nice, spend, earn, increas, expans, stimul
##   Lift: expans, cycl, nice, expand, money, spend, decis
##   Score: money, nice, spend, earn, cycl, polici, expans
## Topic 12 Top Words:
##   Highest Prob: want, els, one, get, need, averag, give
##   FREX: want, els, one, averag, nobodi, get, give
##   Lift: democrat, didnt, want, nobodi, els, vote, strang

```

```

##      Score: want, els, democrat, nobodi, averag, give, one
## Topic 13 Top Words:
##      Highest Prob: nation, countri, peopl, mani, develop, usa, live
##      FREX: nation, develop, usa, research, noth, unemploy, america
##      Lift: develop, nation, research, lowest, fabric, us-born, confirm
##      Score: develop, nation, usa, taxat, noth, research, addict
## Topic 14 Top Words:
##      Highest Prob: good, just, differ, immigr, requir, els, though
##      FREX: good, differ, requir, just, though, bad, els
##      Lift: differ, good, judg, indirect, bad, mix, requir
##      Score: good, differ, just, requir, bad, judg, els
## Topic 15 Top Words:
##      Highest Prob: base, experi, inform, need, peopl, skill, social
##      FREX: experi, base, inform, environment, instagram, invest, skill
##      Lift: environment, experi, base, inform, instagram, bigger, becam
##      Score: base, experi, inform, environment, instagram, unsur, indic
## Topic 16 Top Words:
##      Highest Prob: even, expens, can, also, use, social, need
##      FREX: expens, even, popul, enter, secur, purchas, social
##      Lift: expens, past, poverti, truli, cloth, racist, treasuri
##      Score: expens, even, past, poverti, truli, popul, purchas
## Topic 17 Top Words:
##      Highest Prob: new, servic, lot, work, hard, peopl, use
##      FREX: new, servic, lot, hard, idea, variet, time
##      Lift: fals, variet, new, ahead, produc, idea, nich
##      Score: new, fals, servic, lot, variet, idea, hard
## Topic 18 Top Words:
##      Highest Prob: tax, peopl, pay, immigr, countri, job, fill
##      FREX: sale, fill, back, well, receiv, incom, kind
##      Lift: absolut, file, mess, gap, wasnt, cater, warehous
##      Score: file, sale, absolut, well, run, tax, mess
## Topic 19 Top Words:
##      Highest Prob: pay, citizen, job, alway, take, healthcar, money
##      FREX: alway, healthcar, sure, citizen, away, flood, fair
##      Lift: flood, everyday, promot, desir, ton, meanwhil, damag
##      Score: flood, healthcar, alway, sure, choos, recent, everyday
## Topic 20 Top Words:
##      Highest Prob: american, immigr, job, live, back, mani, take
##      FREX: american, back, send, home, grant, compani, around
##      Lift: appl, grant, orchard, risk, keep, determin, around
##      Score: grant, american, home, send, risk, around, back
## Topic 21 Top Words:
##      Highest Prob: less, lot, farm, will, way, free, hire
##      FREX: farm, less, migrant, free, hire, lot, green
##      Lift: green, migrant, summer, facil, hour, farm, farmer
##      Score: migrant, green, less, farm, free, summer, facil
## Topic 22 Top Words:
##      Highest Prob: immigr, illeg, effect, feel, major, benefit, legal
##      FREX: effect, gut, major, went, feel, follow, net
##      Lift: gut, went, effect, vast, respect, essenti, damag
##      Score: gut, effect, feel, went, vast, illeg, major
## Topic 23 Top Words:
##      Highest Prob: busi, mani, start, help, also, open, employ
##      FREX: busi, start, mani, help, open, also, end

```

```

##      Lift: harvest, start, busi, costum, goe, hurt, open
##      Score: busi, start, open, harvest, mani, help, hire
## Topic 24 Top Words:
##      Highest Prob: media, use, news, social, can, thing, usag
##      FREX: media, usag, news, consumpt, retir, individu, hsa
##      Lift: hsa, usag, media, retir, user, consumpt, indic
##      Score: media, usag, hsa, consumpt, news, user, retir
## Topic 25 Top Words:
##      Highest Prob: pay, tax, help, amount, low, will, govern
##      FREX: pay, tax, amount, low, outweigh, help, side
##      Lift: hustl, pay, tax, properti, outweigh, side, rent
##      Score: pay, tax, hustl, outweigh, amount, properti, side
## Topic 26 Top Words:
##      Highest Prob: immigr, come, take, job, put, lot, illeg
##      FREX: educ, high, healthcar, lower, depend, simpl, right
##      Lift: india, simpl, teenag, ten, promis, argument, immedi
##      Score: india, educ, healthcar, simpl, high, border, teenag
## Topic 27 Top Words:
##      Highest Prob: will, immigr, bring, think, mani, can, come
##      FREX: bring, equal, enter, add, may, continu, enrich
##      Lift: enrich, inspir, continu, human, extra, creativ, strive
##      Score: inspir, enrich, extra, human, bring, enter, continu
## Topic 28 Top Words:
##      Highest Prob: read, articl, news, answer, previous, still, report
##      FREX: read, articl, news, choos, report, answer, still
##      Lift: instruct, read, articl, choos, portion, degre, stat
##      Score: read, news, articl, instruct, choos, report, answer
## Topic 29 Top Words:
##      Highest Prob: creat, put, job, addit, general, balanc, definit
##      FREX: creat, addit, put, key, general, balanc, job
##      Lift: key, addit, creat, put, necessarili, general, balanc
##      Score: creat, put, addit, key, general, balanc, job
## Topic 30 Top Words:
##      Highest Prob: day, question, answer, time, increas, busi, econom
##      FREX: day, question, answer, 're, increas, knowledg, time
##      Lift: knowledg, 're, day, write, loan, respons, age
##      Score: day, knowledg, answer, question, 're, loan, increas
## Topic 31 Top Words:
##      Highest Prob: job, take, make, one, say, amount, immigr
##      FREX: say, amount, own, overal, one, make, lean
##      Lift: lean, theyd, landscap, undesir, own, exact, favor
##      Score: lean, own, say, amount, make, overal, actual
## Topic 32 Top Words:
##      Highest Prob: realli, sure, noth, already, anyon, probabl, seem
##      FREX: realli, noth, sure, already, anyon, seem, cours
##      Lift: clue, love, noth, realli, cours, sure, liter
##      Score: realli, noth, sure, love, clue, seem, anyon
## Topic 33 Top Words:
##      Highest Prob: legal, immigr, illeg, countri, believ, talk, howev
##      FREX: legal, illeg, talk, drain, immigr, howev, whether
##      Lift: mexico, specifi, legal, talk, demonstr, split, drain
##      Score: legal, illeg, mexico, talk, drain, specifi, whether
## Topic 34 Top Words:
##      Highest Prob: citizen, immigr, countri, posit, econom, work, requir

```

```

##      FREX: citizen, requir, econom, posit, result, monetari, boost
##      Lift: monetari, boost, lifestyl, caus, cheat, suppli, entri
##      Score: monetari, citizen, econom, posit, requir, taxat, boost
## Topic 35 Top Words:
##      Highest Prob: immigr, undocu, social, also, will, secur, tax
##      FREX: undocu, near, secur, sale, nativ, espec, howev
##      Lift: near, low-wag, point, undocu, inelig, heavili, black
##      Score: near, undocu, sale, point, low-wag, nativ, secur
## Topic 36 Top Words:
##      Highest Prob: better, life, person, just, fill, american, normal
##      FREX: life, better, person, normal, littl, fill, just
##      Lift: left, normal, life, better, fit, vari, person
##      Score: normal, better, life, person, hope, just, littl
## Topic 37 Top Words:
##      Highest Prob: peopl, work, come, hard, countri, yes, famili
##      FREX: peopl, work, come, hard, yes, odd, countri
##      Lift: odd, peopl, work, come, hard, mind, creativ
##      Score: work, peopl, come, hard, odd, countri, yes
## Topic 38 Top Words:
##      Highest Prob: think, opinion, equal, can, tend, immigr, workforc
##      FREX: opinion, think, equal, tend, can, lack, bigger
##      Lift: opinion, think, equal, candid, bigger, lack, workforc
##      Score: think, opinion, equal, lack, candid, tend, bigger
## Topic 39 Top Words:
##      Highest Prob: feel, come, immigr, major, take, live, mani
##      FREX: feel, major, come, generat, look, welfar, must
##      Lift: combin, origin, feel, enterpris, owe, dare, self
##      Score: feel, origin, major, must, combin, generat, come
## Topic 40 Top Words:
##      Highest Prob: thing, immigr, work, bring, just, hard, peopl
##      FREX: thing, bring, tri, return, hard, labor, grow
##      Lift: plant, vital, havent, demograph, return, built, thing
##      Score: plant, thing, bring, tri, return, havent, vital
## Topic 41 Top Words:
##      Highest Prob: make, contribut, get, just, tax, welfar, product
##      FREX: rest, product, make, welfar, save, contribut, pool
##      Lift: pool, save, fruit, rest, suitabl, hide, broken
##      Score: pool, contribut, welfar, rest, make, product, smaller
## Topic 42 Top Words:
##      Highest Prob: believ, contribut, immigr, will, take, societi, countri
##      FREX: believ, contribut, societi, case, quit, whole, away
##      Lift: quit, believ, toss, throughout, prove, privileg, pass
##      Score: believ, contribut, quit, societi, abil, earner, whole
## Topic 43 Top Words:
##      Highest Prob: immigr, work, pay, tax, worker, contribut, illeg
##      FREX: drive, worker, avoid, guess, wage, work, usual
##      Lift: recruit, scare, crimin, histori, surviv, whenever, hesit
##      Score: recruit, drive, scare, ever, guess, poorer, avoid
## Topic 44 Top Words:
##      Highest Prob: dont, know, much, give, matter, answer, anyth
##      FREX: dont, know, matter, refer, anyth, give, much
##      Lift: refer, dont, comfort, melt, anywher, wasnt, matter
##      Score: dont, refer, know, feel, comfort, give, answer
## Topic 45 Top Words:

```

```

##      Highest Prob: wage, work, immigr, will, illeg, lower, american
##      FREX: wage, lower, kid, illeg, tabl, minimum, worker
##      Lift: scam, penni, minimum, clear, wage, standard, fall
##      Score: wage, scam, lower, kid, illeg, tabl, minimum
## Topic 46 Top Words:
##      Highest Prob: immigr, job, lot, busi, usual, first, tax
##      FREX: usual, first, minor, level, assist, lot, race
##      Lift: serious, miami, race, visit, incent, mention, minor
##      Score: serious, miami, minor, base, race, usual, consid
## Topic 47 Top Words:
##      Highest Prob: state, unit, tax, mani, pay, incom, immigr
##      FREX: unit, state, incom, supplement, famili, assist, everybodi
##      Lift: shortag, wast, unit, state, supplement, spent, crap
##      Score: unit, state, shortag, supplement, land, estim, incom
## Topic 48 Top Words:
##      Highest Prob: immigr, way, either, depend, can, abl, also
##      FREX: either, depend, subject, done, way, abl, guess
##      Lift: similar, subject, fulfil, complex, grasp, wide, anti-immigr
##      Score: similar, subject, either, depend, attempt, process, capabl
## Topic 49 Top Words:
##      Highest Prob: immigr, work, lot, get, american, benefit, welfar
##      FREX: healthcar, lot, get, welfar, work, benefit, free
##      Lift: simpli, got, ride, construct, stop, unfair, factori
##      Score: simpli, lot, healthcar, welfar, work, free, get
## Topic 50 Top Words:
##      Highest Prob: job, take, american, away, immigr, qualifi, spot
##      FREX: take, job, away, american, spot, qualifi, immigr
##      Lift: spot, take, away, job, qualifi, american, fruit
##      Score: take, job, away, american, spot, qualifi, immigr
## Topic 51 Top Words:
##      Highest Prob: immigr, tend, countri, run, open, long, busi
##      FREX: run, tend, long, small, assum, open, owner
##      Lift: took, percentag, run, presenc, face, outsid, profil
##      Score: took, run, tend, small, long, owner, assum
## Topic 52 Top Words:
##      Highest Prob: like, trickl, rate, fee, immigr, need, just
##      FREX: like, trickl, fee, rate, said, there, workforc
##      Lift: trickl, like, fee, rate, immigr, need, just
##      Score: like, trickl, fee, rate, need, just, immigr
## Topic 53 Top Words:
##      Highest Prob: get, illeg, countri, other, know, one, make
##      FREX: other, half, wall, treat, intent, know, illeg
##      Lift: half, wall, treat, vagu, lie, technolog, feed
##      Score: wall, half, treat, illeg, intent, other, steal
## Topic 54 Top Words:
##      Highest Prob: tax, pay, immigr, like, just, fill, citizen
##      FREX: wealthi, tax, littl, deduct, fill, pay, advantag
##      Lift: deduct, wealthi, regular, suck, natural-born, agre, pull
##      Score: wealthi, deduct, suck, tax, regular, citi, town
## Topic 55 Top Words:
##      Highest Prob: peopl, that, job, immigr, even, tax, wish
##      FREX: that, wish, stori, ive, obvious, wont, even
##      Lift: wish, that, fewer, mention, survey, retail, count
##      Score: wish, that, stori, ive, fewer, thousand, wont

```

```
## Topic 56 Top Words:
##   Highest Prob: job, women, immigr, mani, pay, come, peopl
##   FREX: job, women, cheap, workforc, member, regardless, grate
##   Lift: women, job, immigr, mani, pay, come, need
##   Score: women, job, immigr, pay, come, mani, need
```

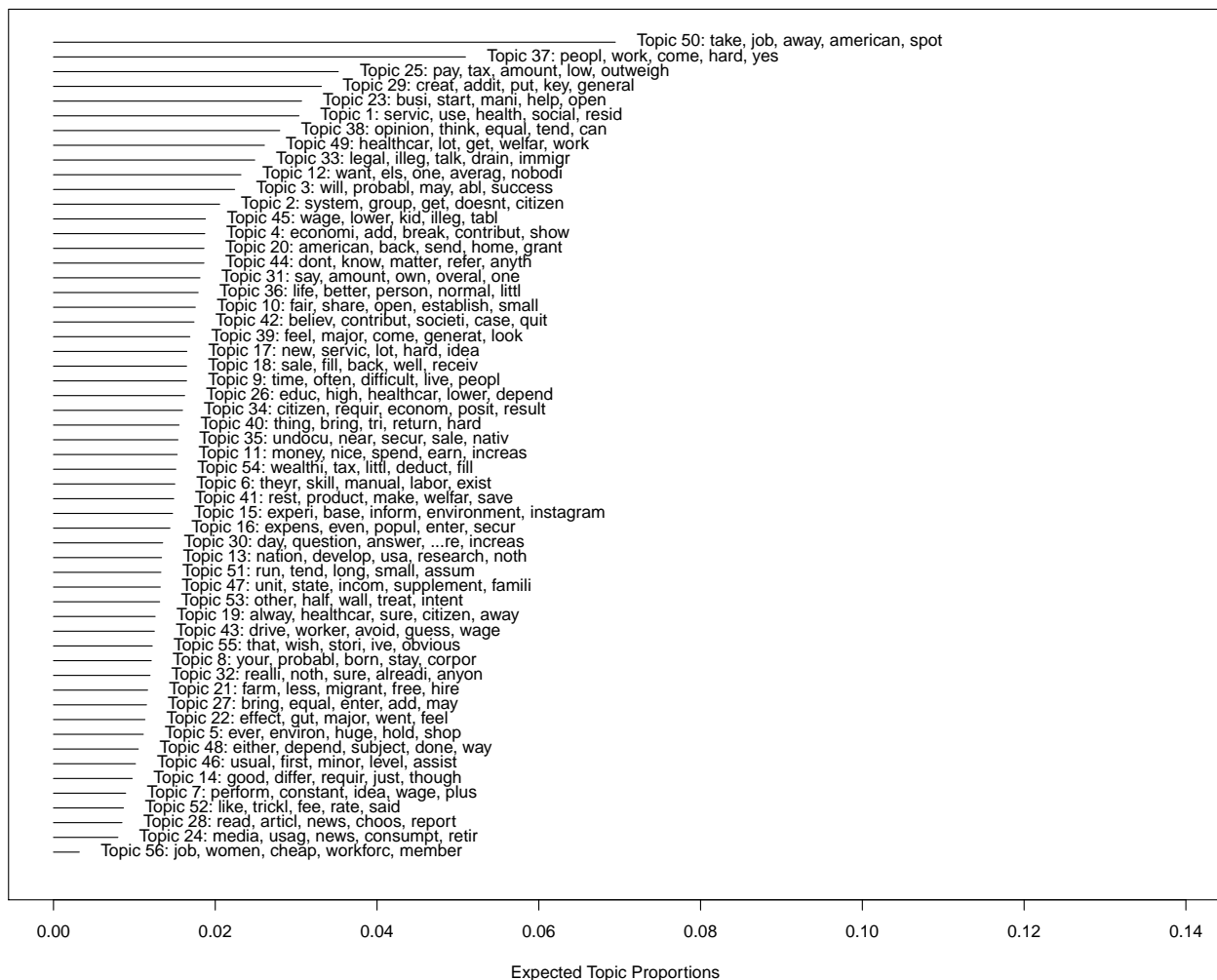
```
par(mar=c(4.2,0.5,2.5,0.5))
plot(stm_fit
     , main=paste0("k = ",stm_fit$settings$dim$K,""),collapse = "")
     , n=5, labeltype = "frex", text.cex = 1)
```

```
## Warning in text.default(frequency[invrnk[i]] + min(2 * max(frequency), :
## conversion failure on 'Topic 30: day, question, answer, 're, increas' in
## 'mbcsToSbcs': dot substituted for <e2>
```

```
## Warning in text.default(frequency[invrnk[i]] + min(2 * max(frequency), :
## conversion failure on 'Topic 30: day, question, answer, 're, increas' in
## 'mbcsToSbcs': dot substituted for <80>
```

```
## Warning in text.default(frequency[invrnk[i]] + min(2 * max(frequency), :
## conversion failure on 'Topic 30: day, question, answer, 're, increas' in
## 'mbcsToSbcs': dot substituted for <99>
```

k = 56)



Sample topics

```
par(mar=c(0.5,0.5,2.5,0.5))
plot(stm_fit, "labels", topics=c(50,27,25,29),
     labeltype = "frex",
     main="Sample Topics (MTurk)")
```

Sample Topics (MTurk)

Topic 50: take, job, away, american, spot, qualifi, immigr, done, fruit, costum, crimin, cheat, menial, close, profil, workforc, cheap, member, payer, there
Topic 27: bring, equal, enter, add, may, continu, enrich, inspir, human, situat, other, extra, accept, made, will, can, depend, cultur, previous, posit
Topic 25: pay, tax, amount, low, outweigh, help, side, properti, hustl, industri, sometim, anyon, govern, incom, collect, sale, communiti, children, school, still
Topic 29: creat, addit, put, key, general, balanc, job, definit, result, necessarili, includ, yes, everyth, cheap, grate, goal, regardless, workforc, english, special

Topic Differences between control and treatment conditions

```
## estimate topic prevalence effects
prep <- estimateEffect(~ age + college + pid_rep + college_pid + assigned + choice
                      , stm_fit, meta = out$meta, uncertainty = "Global")

## select topics with largest treatment differences
tmp_assigned <- tibble(estimate = sapply(summary(prepare)[[3]], function(x) x["assigned", "Estimate"])
                      , topics = prepare$topics) %>% arrange(estimate)
topics_assigned <- c(head(tmp_assigned$topics), tail(tmp_assigned$topics))

tmp_choice <- tibble(estimate = sapply(summary(prepare)[[3]], function(x) x["choice", "Estimate"])
                      , topics = prepare$topics) %>% arrange(estimate)
topics_choice <- c(head(tmp_choice$topics), tail(tmp_choice$topics))

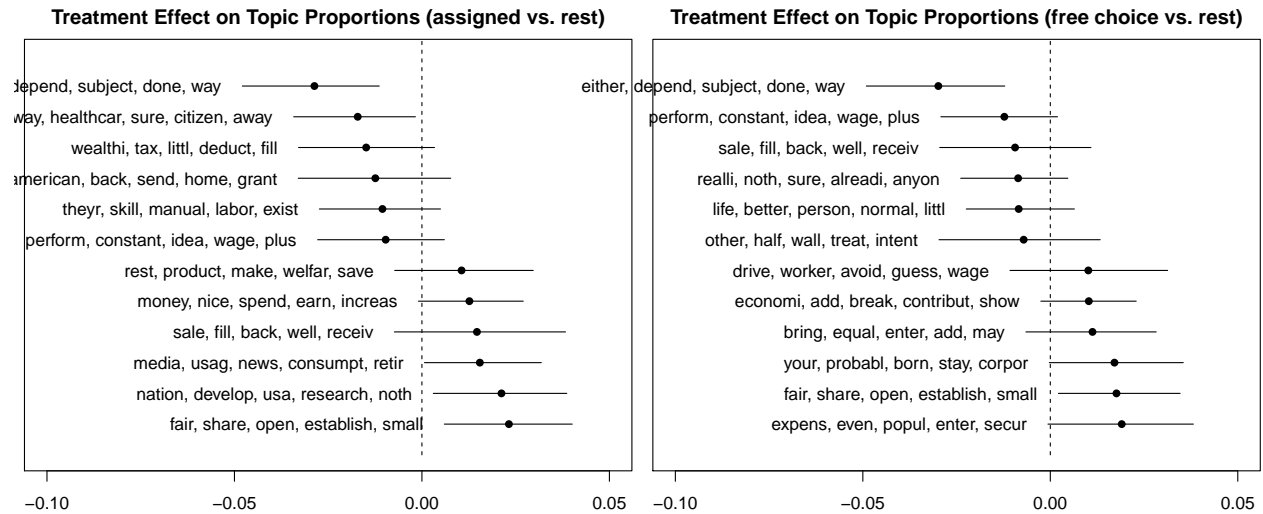
## plot topic differences
par(mfrow=c(1,2), mar=c(2.2,0.5,2.2,0.5))
plot.estimateEffect(prepare, covariate = "assigned", topics = topics_assigned, model = stm_fit)
```



```

, xlim = c(-.1,.05), method = "difference", cov.value1 = 1, cov.value2 = 0
, labeltype = "frefx", n=5, verbose.labels = F, width=50
, main = "Treatment Effect on Topic Proportions (assigned vs. rest)")
plot.estimateEffect(prepare, covariate = "choice", topics = topics_choice, model = stm_fit
, xlim = c(-.1,.05), method = "difference", cov.value1 = 1, cov.value2 = 0
, labeltype = "frefx", n=5, verbose.labels = F, width=50
, main = "Treatment Effect on Topic Proportions (free choice vs. rest)")

```



Next steps

- sentiment analysis?