Country Music Project

Matt Shu

2022-04-28

## Prereqs

library(igraph)  
library(tidyverse)  
library(stm)  
library(RSQLite)  
library(RecordLinkage)  
library(stringdist)

## Preprocessing

conn <- dbConnect(RSQLite::SQLite(), "files/22-04-21-playback-fm-top-country.db")  
dfSongs <- dbGetQuery(conn, 'SELECT \* FROM lyrics')  
dfArtists <- dbGetQuery(conn, 'SELECT \* FROM artists')  
dbDisconnect(conn)

## Dataset Visualizations

dfArtists[dfArtists == "nan"] <- NA  
dfArtistsInterest <- dfArtists %>%  
 dplyr::select(artist\_id, mb\_id, type, area.name, gender, life\_span.begin, life\_span.ended) %>%  
 # the objects need to be class "data frame"   
 as.data.frame()

dfSongsArtists <- merge(dfSongs,dfArtistsInterest,by="artist\_id")

cleaned\_df <-dfSongsArtists %>%  
 # first, remove observation with missing values of the meta variables  
 filter(!is.na(lyrics)) %>%  
 # first, remove observation with missing values of the meta variables  
 filter(!is.na(artist)) %>%  
 as.data.frame()  
cleaned\_df$lyrics <- str\_replace\_all(cleaned\_df$lyrics,"[\\s]+", " ")

## Create Artist ID Hash

cleaned\_df$artist\_id <- as.character(as.numeric(as.factor(cleaned\_df$artist)))  
cleaned\_df$song\_id <- as.character((10000 + as.numeric(as.factor(cleaned\_df$track))))

## Filter out Mismatches

dim(cleaned\_df)

## [1] 7094 18

cleaned\_df$cleaned\_lyrics <-   
 str\_replace\_all(cleaned\_df$lyrics, 'Chap\\. [0-9]', NA\_character\_) %>%  
 str\_replace\_all(., 'Listening Log', NA\_character\_) %>%  
 str\_replace\_all(., 'Favorite Songs Of', NA\_character\_) %>%  
 str\_replace\_all(., 'Chapter [0-9]', NA\_character\_) %>%  
 str\_replace\_all(., 'New Music ', NA\_character\_) %>%  
 str\_replace\_all(., 'Nominees', NA\_character\_) %>%  
 str\_replace\_all(., 'Best Songs of ', NA\_character\_) %>%  
 str\_replace\_all(., "[0-9]+ U S", NA\_character\_) %>% # Court Cases  
 str\_replace\_all(., "[0-9]+ U.S", NA\_character\_) %>% # Court Cases  
 # keep only alphabet letters and numbers ("al" and "num")  
 str\_replace\_all(., "[^[:alnum:]]", " ") %>%  
 # make multiple spaces into one space  
 str\_replace\_all(.,"[ ]+", " ") %>%  
 str\_replace(., ".\*Lyrics", "")  
cleaned\_df <- cleaned\_df %>%  
 filter(!is.na(cleaned\_lyrics)) %>%  
 filter(levenshteinSim(track, str\_match(lyrics, "(.\*)Lyrics")[,2]) > .5) %>% # There are some false positives, when there are other languages  
 as.data.frame()  
dim(cleaned\_df)

## [1] 6371 19

## Preprocessing (and STM exploration)

# Dataframe containing the text  
docs\_df <- cleaned\_df %>%  
 dplyr::select(track\_id, cleaned\_lyrics) %>%  
 # first, remove observation with missing values of the meta variables  
 filter(!is.na(cleaned\_lyrics)) %>%  
 # the objects need to be class "data frame"   
 as.data.frame()

# Dataframe containing (sample) documents' metadata of interest  
meta\_df <- cleaned\_df %>%  
 dplyr::select(track\_id, rank, artist, track, year) %>%  
 # the objects need to be class "data frame"   
 as.data.frame()

processed\_docs\_1 <- textProcessor(documents = docs\_df$cleaned\_lyrics,   
 metadata = meta\_df,   
 lowercase = TRUE,   
 removestopwords = TRUE,   
 removenumbers = TRUE,   
 removepunctuation = TRUE,   
 ucp = TRUE,  
 stem = TRUE,   
 striphtml = TRUE,   
 wordLengths = c(3, Inf),  
 language = "en")

meta <- processed\_docs\_1$meta  
vocab <- processed\_docs\_1$vocab  
docs <- processed\_docs\_1$documents  
keep <- !is.na(meta$artist) && !is.na(meta$rank)  
meta <- meta[keep,]  
docs <- docs[keep]

prepped\_data <- prepDocuments(docs,   
 vocab,   
 meta,  
 # the lower threshold value means that only words  
 # that appear more times than the value (in this   
 # example the value = 3) will be retained; this is   
 # another researcher decision  
 lower.thresh = 2)

Old code for removing unusual mismatch with no words despite past filters

length(docs\_df$cleaned\_lyrics) # original documents  
length(prepped\_data$meta$track\_id) # off from the preceding count  
dif <- setdiff(docs\_df$track\_id, # original vector of documents  
 prepped\_data$meta$track\_id) # list of documents after prepDocuments  
tmp <- docs\_df  
tmp2 <- tmp[!tmp$track\_id %in% dif,]  
tmp\_doc <- tmp2 %>%  
 select(track\_id, cleaned\_lyrics)  
length(tmp\_doc$track\_id)  
length(prepped\_data$meta$track\_id)  
  
# View the track ids that were removed for some reason (often other language)  
tmp3 <- tmp[tmp$track\_id %in% dif,]  
tmp3

See Cleaned Sample!

head(cleaned\_df)

## artist\_id level\_0 index track\_id year artist track  
## 1 805 0 0 0 1944 Red Foley Smoke On The Water  
## 2 805 405 405 405 1950 Red Foley Sunday Down In Tennessee  
## 3 805 506 506 506 1951 Red Foley Hobo Boogie  
## 4 805 376 376 376 1950 Red Foley Birmingham Bounce  
## 5 805 386 386 386 1950 Red Foley Cincinnati Dancing Pig  
## 6 805 374 374 374 1950 Red Foley Chattanoogie Shoe Shine Boy  
## rank link  
## 1 1 /charts/country/video/1944/red-foley-smoke-on-the-water  
## 2 33 /charts/country/video/1950/red-foley-sunday-down-in-tennessee  
## 3 55 /charts/country/video/1951/red-foley-hobo-boogie  
## 4 3 /charts/country/video/1950/red-foley-birmingham-bounce  
## 5 13 /charts/country/video/1950/red-foley-cincinnati-dancing-pig  
## 6 1 /charts/country/video/1950/red-foley-chattanoogie-shoe-shine-boy  
## lyrics  
## 1 Smoke On The Water LyricsThere will be a sad day comin' For the foes of all mankind They must answer to the people And it’s troubling their mind Everybody who must fear them Will rejoice on that great day When the powers of dictators Shall be taken all away There’ll be smoke on the water On the land and the sea When our Army and Navy overtakes the enemy There’ll be smoke on the mountains Where the Heathen Gods stay And the sun that is risin’ Will go down on that day For there is a great destroyer Made of fire and flesh and steel Rollin’ toward the foes of freedom They’ll go down beneath its wheels There’ll be nothing left but vultures To inhabit all that land When our modern ships and bombers Make a graveyard of Japan Hirohito ‘long with Hitler Will be ridin’ on a rail Mussolini’ll beg for mercy As a leader he has failed But there’ll be no time for pity When the Screamin’ Eagle flies That will be the end of Axis They must answer with their lives Embed  
## 2 Sunday Down In Tennessee Lyrics Oh I wake in the morning with my head on the floor Where I left it Saturday night Makin' up my mind, I never said no more To start my Sunday right Oh the bells are a-ringin', and a, dingin' a-dangin' Get along, they're chimin' at me Gonna hurry to the meetin' before the shot begins On a Sunday down in Tennessee Oh brother take me, by the hand I'll lead you to the promised land They gather like a honeysuckle vine They just keep clingin', all the time Early in the evenin' when the sun goes down By the weepin' willow tree I get a great big kiss from a little miss On a Sunday down in Tennessee Oh I wake in the morning with my head on the floor Where I left it Saturday night Makin' up my mind, I never said no more To start my Sunday right Oh the bells are a-ringin', and a, dingin' a-dangin' Get along, they're chimin' at me Gonna hurry to the meetin' before the shot begins On a Sunday down in Tennessee Oh brother take me, by the hand I'll lead you to the promised land They gather like a honeysuckle vine They just keep clingin', all the time Early in the evenin' when the sun goes down By the weepin' willow tree I get a great big kiss from a little miss On a Sunday down in Tennessee I'm gonna walk that aisle with my head up high Look that a-possum right-a, in the eye Be shoutin' hallelujah 'til the day I die On Sunday, Sunday down in TennesseeEmbed  
## 3 Hobo Boogie Lyrics For a sandwich or a quarter, or a spider something wet We will play that old piano like you have never heard it yet The hobo boogie, that what he calls his melody Got the rhythm in a boxcar ridin' on a bumpy rail Wrote the tune from start to finish on the walls of the county jail The hobo boogie, that's just the way it seem to be He tires of faces, everyday He tires of places, wants to get away He gets the urge to hit the steel To Asbill, Nashville, Chattanooga, or Mobil You can hear the wheels a-clangin' and the rails just seem to hum It'll make you feel like shoutin' "Hey there Birmingham, here I come!" The hobo boogie, you got to love that melody He tires of faces, everyday He tires of places, wants to get away He gets the urge to hit the steel To Asbill, Nashville, Chattanooga, or Mobil You can hear the wheels a-clangin' and the rails just seem to hum It'll make you feel like shoutin' "Hey there Birmingham, here I come!" The hobo boogie, you got to love that melodyEmbed  
## 4 Birmingham Bounce Lyrics In the heart of dixie down in Alabam There's a place we love called Birmingham Everybody starts a-rockin' and a-shufflin' their feet When a drum starts playin' that solid beat Now everybody's dancin' and jumpin' too When the music starts a-rockin', nobody's blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham Bounce Now some like a horn with a valve in the middle And some like the beat of a big bare spiddle If you're a-dancin' in the moonlight or a-dancin' in the dark We all like to dance to an old french harp Now everybody's dancin' and a-jumpin' too When the music starts a-rockin', nobody's blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham Bounce If you like your music going eight to the bar I know a way to make it best before If the beat gets you movin' and you need a good start Let's everybody dance to a steel guitar Now everybody's dancin' and jumpin' too When the music starts a-rockin', nobody's blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham Bounce Now whether it's hot or fine or sweet If it's down in dixie, brother, it's got that beat We don't have the moon or even the star But here's that guitar going eight to the bar Now everybody's dancin' and jumpin' too When the music starts a-rockin', nobody's blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham Bounce When you come down south and you want a good time Just come to a dance when you're feelin' fine Now that's all it takes to spend a happy time Let's all dance together and have a wonderful time! Now everybody's dancin' and jumpin' too When the music starts a-rockin', nobody's blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham BounceEmbed  
## 5 Cincinnati Dancing Pig Lyrics Cincinnati's dancin' pig, he's the barnyard's mister big Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Children yell and clap and sing When he does his buckin' wing Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Dancin' bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain't-a seen nothin' at all From Duluth to Birmingham, soo wee He's the porkchop Dapper Dan He's the keenest hog white am, Cincinnati's dancin' pig With his riggidy-jiggidy-jiggidy-jig-jig-jig Cincinnati's dancin' pig, he's the barnyard's mister big Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Children yell and clap and sing When he does his buckin' wing Cincinnati's dancin' pig, with his Riggidy-jiggidy-jiggidy-jig, jigga-jig-jig Dancin' bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain't-a seen nothin' at all From Duluth to Birmingham, soo wee He's the porkchop Dapper Dan He's the keenest hog white am, Cincinnati's dancin' pig With his riggidy, with his jiggidy, jig-jig-jigEmbed  
## 6 Chattanoogie Shoe Shine Boy LyricsHave you ever passed the corner of Forth and Grand? Where a little ball o' rhythm has a shoe-shine stand People gather 'round and they clap their hands He's a great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy He charges you a nickel just to shine one shoe He makes the oldest kind o' leather look like new You feel as though you wanna dance when he gets through He's a great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy It's a wonder that the rag don't tear The way he makes it pop You ought to see him fan the air With his hoppity-hippity-hippity-hoppity-hoppity-hippity-hop He opens up for business when the clock strikes nine He likes to get up early when they're feelin' fine Everybody gets a little rise 'n shine With the great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy It's a wonder that the rag don't tear The way he makes it pop Just listen to him fan the air Here he goes! He opens up for business when the clock strikes nine He likes to get up early when they're feelin' fine Everybody gets a little rise 'n shine With the great big bundle o' joy He pops the boogie woogie rag The Chattanoogie shoe-shine boy The Chattanoogie shoe-shine boyEmbed  
## artist\_appearances mb\_id type area.name  
## 1 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 2 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 3 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 4 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 5 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## 6 33 aff932c2-ec30-4ee9-9125-5f761aae61a4 Person United States  
## gender life\_span.begin life\_span.ended song\_id  
## 1 male 1910-06-17 true 14521  
## 2 male 1910-06-17 true 14780  
## 3 male 1910-06-17 true 11892  
## 4 male 1910-06-17 true 10587  
## 5 male 1910-06-17 true 10833  
## 6 male 1910-06-17 true 10810  
## cleaned\_lyrics  
## 1 There will be a sad day comin For the foes of all mankind They must answer to the people And it s troubling their mind Everybody who must fear them Will rejoice on that great day When the powers of dictators Shall be taken all away There ll be smoke on the water On the land and the sea When our Army and Navy overtakes the enemy There ll be smoke on the mountains Where the Heathen Gods stay And the sun that is risin Will go down on that day For there is a great destroyer Made of fire and flesh and steel Rollin toward the foes of freedom They ll go down beneath its wheels There ll be nothing left but vultures To inhabit all that land When our modern ships and bombers Make a graveyard of Japan Hirohito long with Hitler Will be ridin on a rail Mussolini ll beg for mercy As a leader he has failed But there ll be no time for pity When the Screamin Eagle flies That will be the end of Axis They must answer with their lives Embed  
## 2 Oh I wake in the morning with my head on the floor Where I left it Saturday night Makin up my mind I never said no more To start my Sunday right Oh the bells are a ringin and a dingin a dangin Get along they re chimin at me Gonna hurry to the meetin before the shot begins On a Sunday down in Tennessee Oh brother take me by the hand I ll lead you to the promised land They gather like a honeysuckle vine They just keep clingin all the time Early in the evenin when the sun goes down By the weepin willow tree I get a great big kiss from a little miss On a Sunday down in Tennessee Oh I wake in the morning with my head on the floor Where I left it Saturday night Makin up my mind I never said no more To start my Sunday right Oh the bells are a ringin and a dingin a dangin Get along they re chimin at me Gonna hurry to the meetin before the shot begins On a Sunday down in Tennessee Oh brother take me by the hand I ll lead you to the promised land They gather like a honeysuckle vine They just keep clingin all the time Early in the evenin when the sun goes down By the weepin willow tree I get a great big kiss from a little miss On a Sunday down in Tennessee I m gonna walk that aisle with my head up high Look that a possum right a in the eye Be shoutin hallelujah til the day I die On Sunday Sunday down in TennesseeEmbed  
## 3 For a sandwich or a quarter or a spider something wet We will play that old piano like you have never heard it yet The hobo boogie that what he calls his melody Got the rhythm in a boxcar ridin on a bumpy rail Wrote the tune from start to finish on the walls of the county jail The hobo boogie that s just the way it seem to be He tires of faces everyday He tires of places wants to get away He gets the urge to hit the steel To Asbill Nashville Chattanooga or Mobil You can hear the wheels a clangin and the rails just seem to hum It ll make you feel like shoutin Hey there Birmingham here I come The hobo boogie you got to love that melody He tires of faces everyday He tires of places wants to get away He gets the urge to hit the steel To Asbill Nashville Chattanooga or Mobil You can hear the wheels a clangin and the rails just seem to hum It ll make you feel like shoutin Hey there Birmingham here I come The hobo boogie you got to love that melodyEmbed  
## 4 In the heart of dixie down in Alabam There s a place we love called Birmingham Everybody starts a rockin and a shufflin their feet When a drum starts playin that solid beat Now everybody s dancin and jumpin too When the music starts a rockin nobody s blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham Bounce Now some like a horn with a valve in the middle And some like the beat of a big bare spiddle If you re a dancin in the moonlight or a dancin in the dark We all like to dance to an old french harp Now everybody s dancin and a jumpin too When the music starts a rockin nobody s blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham Bounce If you like your music going eight to the bar I know a way to make it best before If the beat gets you movin and you need a good start Let s everybody dance to a steel guitar Now everybody s dancin and jumpin too When the music starts a rockin nobody s blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham Bounce Now whether it s hot or fine or sweet If it s down in dixie brother it s got that beat We don t have the moon or even the star But here s that guitar going eight to the bar Now everybody s dancin and jumpin too When the music starts a rockin nobody s blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham Bounce When you come down south and you want a good time Just come to a dance when you re feelin fine Now that s all it takes to spend a happy time Let s all dance together and have a wonderful time Now everybody s dancin and jumpin too When the music starts a rockin nobody s blue A funny little rhythm with a solid sound A boogie and a jumper called the Birmingham BounceEmbed  
## 5 Cincinnati s dancin pig he s the barnyard s mister big Cincinnati s dancin pig with his Riggidy jiggidy jiggidy jig jigga jig jig Children yell and clap and sing When he does his buckin wing Cincinnati s dancin pig with his Riggidy jiggidy jiggidy jig jigga jig jig Dancin bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain t a seen nothin at all From Duluth to Birmingham soo wee He s the porkchop Dapper Dan He s the keenest hog white am Cincinnati s dancin pig With his riggidy jiggidy jiggidy jig jig jig Cincinnati s dancin pig he s the barnyard s mister big Cincinnati s dancin pig with his Riggidy jiggidy jiggidy jig jigga jig jig Children yell and clap and sing When he does his buckin wing Cincinnati s dancin pig with his Riggidy jiggidy jiggidy jig jigga jig jig Dancin bears and kangaroos have a lot of the ball But until you seen that remarkable pig You ain t a seen nothin at all From Duluth to Birmingham soo wee He s the porkchop Dapper Dan He s the keenest hog white am Cincinnati s dancin pig With his riggidy with his jiggidy jig jig jigEmbed  
## 6 Have you ever passed the corner of Forth and Grand Where a little ball o rhythm has a shoe shine stand People gather round and they clap their hands He s a great big bundle o joy He pops the boogie woogie rag The Chattanoogie shoe shine boy He charges you a nickel just to shine one shoe He makes the oldest kind o leather look like new You feel as though you wanna dance when he gets through He s a great big bundle o joy He pops the boogie woogie rag The Chattanoogie shoe shine boy It s a wonder that the rag don t tear The way he makes it pop You ought to see him fan the air With his hoppity hippity hippity hoppity hoppity hippity hop He opens up for business when the clock strikes nine He likes to get up early when they re feelin fine Everybody gets a little rise n shine With the great big bundle o joy He pops the boogie woogie rag The Chattanoogie shoe shine boy It s a wonder that the rag don t tear The way he makes it pop Just listen to him fan the air Here he goes He opens up for business when the clock strikes nine He likes to get up early when they re feelin fine Everybody gets a little rise n shine With the great big bundle o joy He pops the boogie woogie rag The Chattanoogie shoe shine boy The Chattanoogie shoe shine boyEmbed

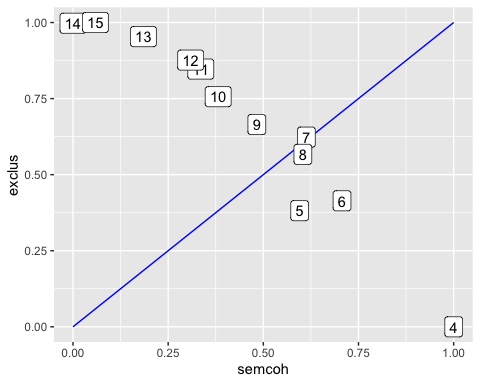
### Find K

k\_seq = seq(4, 15, 1)

## You can "watch" the algorithm model topics in the console  
searched = searchK(prepped\_data$documents,  
 prepped\_data$vocab,  
 K = k\_seq,  
 data = prepped\_data$meta,   
 seed = 183654)  
# saveRDS(searched, file = "22-04-22-searchK-4-15.RData")

### Show K

searched <- readRDS("22-04-22-searchK-4-15.RData")  
# Get values from `searchK` output  
semcoh <- unlist(searched$results$semcoh)  
exclus <- unlist(searched$results$exclus)  
  
# Max/min semantic cohesion  
max\_sc <- max(semcoh)  
min\_sc<-min(semcoh)  
  
# Max/min exclusivity  
max\_ex<-max(exclus)  
min\_ex<-min(exclus)  
  
# Min-max normalization is (value - min)/(max - min)  
x\_vals <- (semcoh-min\_sc)/(max\_sc-min\_sc)  
y\_vals <- (exclus-min\_ex)/(max\_ex-min\_ex)  
# add semantic cohesion and exclusivity together weighted evenly  
ids = k\_seq  
search\_plot\_df <- tibble(id = ids,   
 semcoh = x\_vals,  
 exclus = y\_vals,   
 combine = x\_vals\*0.5 + y\_vals\*0.5)  
  
# Plot  
ggplot(search\_plot\_df, mapping = aes(x = semcoh, y = exclus)) +  
 xlim(0,1) +  
 ylim(0,1) +  
 annotate("segment", x = 0, xend = 1, y = 0, yend = 1, color = "blue") +  
 geom\_label(aes(label=id))



### Model Work

# 6 topics seems to also work nice, with a strong "Country" category  
num\_topics <- 7 # Chosen after above search and some playing around  
out\_covariates\_7 <- stm(prepped\_data$documents,  
 prepped\_data$vocab,  
 K = num\_topics,  
 prevalence = ~ rank \* year,  
 max.em.its = 500,  
 data = prepped\_data$meta,  
 seed = 592669)

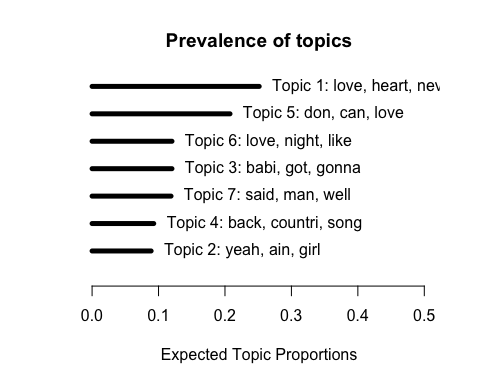
terms = labelTopics(out\_covariates\_7, n = 10)  
terms$prob # rows are topics; columns are most probable words (in order)

## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]   
## [1,] "love" "heart" "never" "one" "time" "now" "say" "still" "just"   
## [2,] "yeah" "ain" "girl" "like" "good" "got" "man" "just" "littl"  
## [3,] "babi" "got" "gonna" "time" "one" "littl" "come" "night" "now"   
## [4,] "back" "countri" "song" "roll" "get" "old" "road" "town" "like"   
## [5,] "don" "can" "love" "know" "want" "let" "just" "make" "like"   
## [6,] "love" "night" "like" "day" "dream" "eye" "blue" "sweet" "rain"   
## [7,] "said" "man" "well" "old" "home" "big" "daddi" "boy" "mama"   
## [,10]   
## [1,] "will"   
## [2,] "ooh"   
## [3,] "right"   
## [4,] "rock"   
## [5,] "feel"   
## [6,] "heaven"  
## [7,] "just"

terms$frex # rows are topics; columns are most FREX words (in order)

## [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]   
## [1,] "cri" "goodby" "fool" "tear" "true" "hurt" "lie" "memori"  
## [2,] "ooh" "boo" "huh" "gimm" "whoa" "yeah" "lovin" "girl"   
## [3,] "bye" "honki" "tonk" "honey" "shake" "babi" "drinkin" "gonna"   
## [4,] "boogi" "countri" "hillbilli" "jone" "santa" "cowboy" "crank" "cha"   
## [5,] "want" "need" "don" "hold" "let" "feel" "easi" "fall"   
## [6,] "angel" "heaven" "sail" "shine" "wing" "sea" "rain" "sky"   
## [7,] "mom" "dad" "hero" "wife" "twenti" "daddi" "famili" "blah"   
## [,9] [,10]   
## [1,] "darl" "still"   
## [2,] "bit" "woah"   
## [3,] "gotta" "thinkin"  
## [4,] "claus" "tractor"  
## [5,] "enough" "give"   
## [6,] "storm" "fli"   
## [7,] "father" "momma"

# Parameters modified from: https://milesdwilliams15.github.io/Better-Graphics-for-the-stm-Package-in-R/  
par(bty="n",lwd=5)  
plot(out\_covariates\_7,  
 type = "summary",  
 main = "Prevalence of topics")

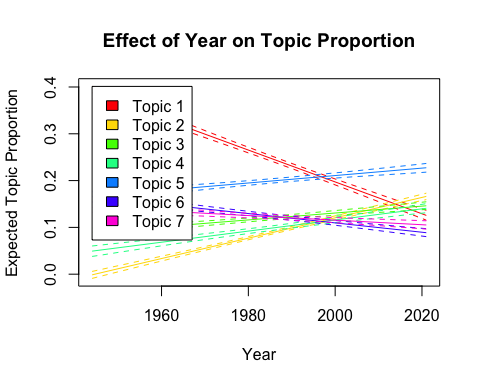


docs\_examples\_covar <- findThoughts(out\_covariates\_7,  
 texts = tmp\_doc$track\_id,  
 n = 10,  
 topics = c(1:num\_topics))  
  
for(topic\_num in c(1:num\_topics)) {  
 print(paste("Topic ", topic\_num))  
 for(track in docs\_examples\_covar$docs[[topic\_num]]) {  
 print(cleaned\_df$track[cleaned\_df$track\_id == track])  
 }  
 print("")  
}

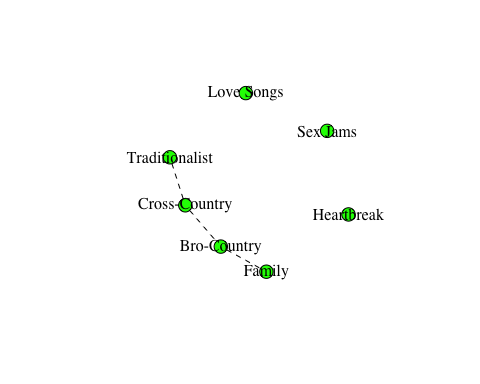
## [1] "Topic 1"  
## [1] "Something Old, Something New"  
## [1] "One Promise Too Late"  
## [1] "All Alone in This World without You"  
## [1] "Sweetheart You Done Me Wrong"  
## [1] "Fool Fool Fool"  
## [1] "Am I Losing You"  
## [1] "Happy Journey"  
## [1] "Am I Losing You"  
## [1] "When You Are Lonely"  
## [1] "Is It Wrong? (For Loving You)"  
## [1] ""  
## [1] "Topic 2"  
## [1] "Desperate Man"  
## [1] "Gimmie That Girl"  
## [1] "Gimmie That Girl"  
## [1] "Just The Way"  
## [1] "Just The Way"  
## [1] "Just the Way"  
## [1] "You Broke Up with Me"  
## [1] "You Broke Up with Me"  
## [1] "Uh-Huh--Mm"  
## [1] "Uh-Huh-mm"  
## [1] ""  
## [1] "Topic 3"  
## [1] "Swing"  
## [1] "Honky Tonkin'"  
## [1] "Heartache Medication"  
## [1] "Heartache Medication"  
## [1] "Honky Tonkin'"  
## [1] "Trademark"  
## [1] "Penny Arcade"  
## [1] "If You've Got The Money I've Got The Time"  
## [1] "If You've Got The Money I've Got The Time"  
## [1] "It's A Little Too Late"  
## [1] ""  
## [1] "Topic 4"  
## [1] "Teenage Boogie"  
## [1] "Redneck Yacht Club"  
## [1] "Cincinnati Dancing Pig"  
## [1] "The Rhumba Boogie"  
## [1] "Long Live"  
## [1] "She Cranks My Tractor"  
## [1] "Ragtime Cowboy Joe"  
## [1] "Hula Rock"  
## [1] "Mule Train"  
## [1] "Pan American Boogie"  
## [1] ""  
## [1] "Topic 5"  
## [1] "Don't Be Stupid (You Know I Love You)"  
## [1] "Don't Be Stupid (You Know I Love You)"  
## [1] "I Can't Get Close Enough"  
## [1] "I Can't Get Close Enough"  
## [1] "Losing Sleep"  
## [1] "Losing Sleep"  
## [1] "Love Lessons"  
## [1] "It Matters To Me"  
## [1] "It Matters to Me"  
## [1] "Fall Into Me"  
## [1] ""  
## [1] "Topic 6"  
## [1] "Ring Of Fire"  
## [1] "Sweet Summer Lovin'"  
## [1] "Your Name Is Beautiful"  
## [1] "Mockin' Bird Hill"  
## [1] "It's A Little More Like Heaven"  
## [1] "A Fallen Star"  
## [1] "The Red Strokes"  
## [1] "Wings Of A Dove"  
## [1] "Would You Lay With Me (In A Field Of Stone)"  
## [1] "Kentucky Waltz"  
## [1] ""  
## [1] "Topic 7"  
## [1] "(Margie's At) The Lincoln Park Inn"  
## [1] "Deck Of Cards"  
## [1] "No Charge"  
## [1] "Life Of A Poor Boy"  
## [1] "History Repeats Itself"  
## [1] "What's Your Mama's Name"  
## [1] "Poor, Poor Pitiful Me"  
## [1] "Po' Folks"  
## [1] "Shiftwork"  
## [1] "Sawmill"  
## [1] ""

# Topic 1: Heartbreak Songs  
# Topic 2: Cross-Country (Country Rock/Pop)  
# Topic 3: Traditionalist Country (Pardi, Hank Williams)  
# Topic 4: Bro-Country  
# Topic 5: Sex Jams  
# Topic 6: Love songs  
# Topic 7: Family  
topic\_labels <- c("Heartbreak", "Cross-Country", "Traditionalist", "Bro-Country", "Sex Jams", "Love Songs", "Family")

eff <- estimateEffect(formula = c(1:num\_topics) ~ year,  
 # the line above matches the model specification we used  
 stmobj = out\_covariates\_7,  
 meta = prepped\_data$meta,  
 uncertainty = "Global")  
  
# Second, plot the results  
plot(eff,  
 covariate = "year",  
 topics = c(1:num\_topics),  
 model = out\_covariates\_7,  
 method = "continuous",  
 xlab = "Year",  
 main = "Effect of Year on Topic Proportion")



plot(topicCorr(out\_covariates\_7),   
 vlabels = topic\_labels, vertex.label.cex = 1.0)

 Topics 3, 2, 4, 7 are all related. This is an interesting finding! This suggests that traditionalist country especially seems related to both country rock/pop songs Topic 2?: Country Rock/Pop Topic 3: Traditionalist Country Topic 4: Bro-Country Topic 7: Family

## More on Topic Models

### Questions/Interests

* How would I see where individual artists fell in terms of topics?
* In general, seeing prevalence of certain
* Would it be, taking the top x documents for different topics and counting from there? ### More to Do?
* Plot covariate interaction!
  + Particularly interested in tracking gender \* year interactions!

## ConText

library(quanteda)

## Package version: 3.2.1  
## Unicode version: 13.0  
## ICU version: 69.1

## Parallel computing: 10 of 10 threads used.

## See https://quanteda.io for tutorials and examples.

library(conText)

cr\_glove <- read.csv("/Users/mattshu/Code/Country/files/glove.6B/glove.6B.50d.txt", sep = " ", quote = "")

df\_gendered <- cleaned\_df %>%  
 filter(gender == "male" | gender == "female")

artist\_meta\_df <- df\_gendered %>%  
 dplyr::select(track\_id, rank, artist, track, year, gender) %>%  
 # the objects need to be class "data frame"   
 as.data.frame()  
par\_corpus <- quanteda::corpus(df\_gendered$cleaned\_lyrics, docvars = artist\_meta\_df)

print(par\_corpus)

## Corpus consisting of 5,291 documents and 6 docvars.  
## text1 :  
## "There will be a sad day comin For the foes of all mankind Th..."  
##   
## text2 :  
## " Oh I wake in the morning with my head on the floor Where I ..."  
##   
## text3 :  
## " For a sandwich or a quarter or a spider something wet We wi..."  
##   
## text4 :  
## " In the heart of dixie down in Alabam There s a place we lov..."  
##   
## text5 :  
## " Cincinnati s dancin pig he s the barnyard s mister big Cinc..."  
##   
## text6 :  
## "Have you ever passed the corner of Forth and Grand Where a l..."  
##   
## [ reached max\_ndoc ... 5,285 more documents ]

# From Vignette, modified  
# tokenize corpus removing unnecessary (i.e. semantically uninformative) elements  
toks <- tokens(par\_corpus, remove\_punct=T, remove\_symbols=T, remove\_numbers=T, remove\_separators=T)  
  
# clean out stopwords and words with 2 or fewer characters  
toks\_nostop <- tokens\_select(toks, pattern = stopwords("en"), selection = "remove", min\_nchar=3)  
  
# only use features that appear at least 3 times in the corpus (follow Week 8 lecture)  
feats <- dfm(toks\_nostop, tolower=T, verbose = FALSE) %>% dfm\_trim(min\_termfreq = 3) %>% featnames()  
  
# leave the pads so that non-adjacent words will not become adjacent  
toks <- tokens\_select(toks\_nostop, feats, padding = TRUE)

searchConText <- function(term) {  
 # # Build a tokenized corpus of contexts surrounding the target term   
 target\_toks <- tokens\_context(x = toks, pattern = "american", window = 6L)  
 # Build document-feature matrix (documents x context counts)  
 target\_dfm <- dfm(target\_toks)  
 # Construct document-embedding-matrix  
 target\_dem <- dem(x = target\_dfm, pre\_trained = cr\_glove\_subset, transform = TRUE,  
 transform\_matrix = cr\_transform, verbose = TRUE)  
}

# # Build a tokenized corpus of contexts surrounding the target term   
 target\_toks <- tokens\_context(x = toks, pattern = "american", window = 6L)

## 103 instances of "American" found.

# Build document-feature matrix (documents x context counts)  
 target\_dfm <- dfm(target\_toks)  
 # Construct document-embedding-matrix  
 target\_dem <- dem(x = target\_dfm, pre\_trained = cr\_glove\_subset, transform = TRUE,  
 transform\_matrix = cr\_transform, verbose = TRUE)

## the following documents could not be embedded due lack of overlap with pre-trained embeddings provided:   
## text5 text6 text14

head(docvars(target\_toks))

## pattern track\_id rank artist track year gender  
## 1 American 456 5 Hank Snow The Rhumba Boogie 1951 male  
## 2 American 456 5 Hank Snow The Rhumba Boogie 1951 male  
## 3 American 456 5 Hank Snow The Rhumba Boogie 1951 male  
## 4 American 2104 87 Dave Dudley There Ain't No Easy Run 1968 male  
## 5 American 2104 87 Dave Dudley There Ain't No Easy Run 1968 male  
## 6 American 2104 87 Dave Dudley There Ain't No Easy Run 1968 male

Now, embed the *contexts* in a pre-trained GloVe embedding space, getting a document-embedding matrix (DFM), or the embeddings of the *context* words

# Average embeddings for each group  
target\_wv\_group <- dem\_group(target\_dem,  
 groups = target\_dem@docvars$gender)

head(target\_wv\_group)

## 2 x 300 sparse Matrix of class "dgCMatrix"  
##   
## female 0.3030909 1.095349 -0.1755629 0.2329034 -0.4276386 -0.3083108 -0.6891217  
## male 0.9948873 1.122779 -0.2938416 0.1991825 -0.3281815 -0.8596758 -0.4971838  
##   
## female -0.9509797 0.640981 -0.6930542 1.2220827 -1.0485098 -0.7874636  
## male -0.5045496 0.467691 -0.3572709 0.7962255 -0.5113317 -0.6245418  
##   
## female -0.02591848 -0.2403192 0.93363922 -0.3263275 -0.36711970 -1.2563151  
## male -0.32784504 -0.3270963 -0.08611797 -0.9854426 0.04674982 -0.7307498  
##   
## female 0.91927279 0.6025674 0.03430842 -0.2071317 0.1687569 -0.1269114  
## male -0.02567004 -0.1901489 0.36241646 -0.9725345 0.4722875 0.1727238  
##   
## female -0.09032906 -0.2432363 0.4885883 0.08246122 -0.5595866 0.7205333  
## male 0.38863356 -0.6742723 0.1022031 -0.35247226 0.3936855 0.6808569  
##   
## female -0.93231644 0.1358051 -0.1791563 0.2827617 0.7595642 0.7327959  
## male -0.01468286 0.6601954 0.3496799 0.4220466 -0.1982212 -0.6019303  
##   
## female -0.2470729 -0.05838938 0.6678931 0.7196711 -0.1412269 -0.6495611  
## male -0.1757042 0.91638349 0.6865761 0.4332839 -0.2259804 -0.1642327  
##   
## female -0.2900546 -0.3907649 0.1399875 0.006399555 -0.5888902 0.6142218  
## male -0.1869556 0.8768652 -0.5506700 0.317018698 -0.4032789 0.7640179  
##   
## female 0.1768595 0.3457654 0.54484610 -0.8518642 -0.1478387 0.1912305  
## male 0.5536685 0.2124449 -0.04180836 0.2468672 -0.2128598 0.1518218  
##   
## female 0.9539514 0.6249391 -0.1591675 1.0042349 -0.5220514 0.03167331  
## male 0.1746726 -0.6382304 0.1983151 0.2955813 -1.1593093 -0.93916002  
##   
## female -0.07382302 0.6128713 -0.3267310 -0.05560593 -0.07532631 -0.6171553  
## male 0.02308379 0.4615594 -0.1774827 -0.17730385 -0.42475302 -0.3157373  
##   
## female -0.03661868 -0.3056548 -0.1294318 -0.5627023 -0.3962115 0.6177157  
## male -0.28458059 -0.5511863 -0.5303206 0.1939632 -1.1194814 0.1884615  
##   
## female -0.7982959 -0.462917 0.1574375 0.2154107 1.2052324 1.1697635 0.4684341  
## male -0.2841658 -0.472473 0.1247484 -0.4589524 -0.6043632 0.8887876 0.1946389  
##   
## female -1.161013 0.1789774 -0.7261436 -0.7068245 0.3070260 -0.5613230  
## male -1.134047 0.4132999 -1.3234160 -0.2536346 0.2445854 0.2548178  
##   
## female 1.1125152 -0.02769463 1.05514176 -0.26439236 -0.02403834 -0.2504286  
## male -0.1112482 0.89478401 0.05912343 0.04004097 -0.25022205 0.9192010  
##   
## female -0.2610992 0.3361751 -0.3963647 0.69925001 1.2069009 -0.2132067  
## male -0.4050782 1.0683911 0.6934883 -0.04432086 0.2985139 0.1433412  
##   
## female 0.01703106 0.3204761 -0.8602922 -0.71512295 0.79626452 -0.1419600  
## male 0.23581409 0.3438267 -1.0671419 -0.02706779 0.01693759 -0.3021458  
##   
## female 0.54759137 -0.3693321 -0.1098982 0.7268921 0.7807348 1.0596712  
## male -0.06730507 0.3492217 0.2888474 0.8040947 0.2942350 0.1710342  
##   
## female -0.3993035 0.5191703 -0.008042155 0.3810279 0.4239872 -0.1438582  
## male -1.1067761 0.6337433 0.551254922 0.9908779 0.4327272 -0.5142067  
##   
## female 0.2542464 0.7268061 0.9369834 0.6712257 0.09419796 0.6181391  
## male 0.3253349 0.7125235 0.1873192 -0.2732803 1.35539275 -0.5701575  
##   
## female -0.6994757 -0.3039099 0.01591601 -0.5344610 -0.2023332 -1.93933142  
## male -0.2163223 0.2483328 -0.04144557 0.1526602 0.2093871 -0.05338367  
##   
## female -0.3993044 0.06128639 0.6858845 0.3372417 0.1099619 -0.8569912  
## male -0.3449885 -0.45237117 0.2275099 0.5854036 -0.7559131 -0.4487358  
##   
## female 0.002000921 0.8080723 0.4321891 -0.2216565 -0.5067833 0.4122594  
## male -0.370159767 1.0019060 -0.7519621 0.5164404 -1.6256129 1.0178920  
##   
## female 0.4521103 0.2025188 0.1109616 -0.2791882 -0.01480319 -0.04465292  
## male 0.4738183 0.4542873 0.9072039 -0.3078951 -0.34425320 0.51682419  
##   
## female 1.1097895 0.5944275 0.9156496 -0.54722091 0.9497904 -0.77345138  
## male -0.1655715 0.7154960 0.5634079 -0.09556115 -0.4883035 -0.08437827  
##   
## female 0.1083365 0.9347559 0.3610195 -0.2461823 0.8965778 -0.7837595 0.312891  
## male 0.1012345 1.4105583 -0.7846625 -0.8849748 0.2327694 0.3295551 0.538410  
##   
## female 0.2086129 1.5651416 0.4063604 -1.0778357 0.2009019 -0.2030924 1.002941  
## male 1.4622175 -0.2757673 0.2968384 -0.9213939 0.2897461 0.4375941 1.084287  
##   
## female -0.1280502 -1.4592766 -0.10639470 0.4268171 0.3791798 -0.4842116  
## male 0.2090131 0.5120551 -0.04703201 0.2066111 0.3201747 -0.9787080  
##   
## female -1.0739718 -0.41181822 0.1846654 -0.09586019 0.08830802 0.26361955  
## male -0.1234892 -0.05238269 -0.1253780 -0.35969831 -0.30345801 -0.01919878  
##   
## female -0.42164057 -0.1218892 -0.06790991 0.6817599 -0.6242030 0.8261044  
## male -0.01514695 0.2500888 -0.18401846 0.1046412 -0.4372776 0.3188515  
##   
## female 0.76571173 0.7618712 -0.2332402 0.304836024 -0.4408048 1.0236763  
## male -0.04023457 0.2164937 -0.1122758 0.001638749 -0.5112973 0.5650177  
##   
## female 0.05076858 0.3668671 0.9053034 -0.1959840 -0.5859493 0.03307819  
## male 0.38874093 -0.5275153 0.4218907 0.2828624 -0.7245550 0.64441635  
##   
## female -0.03698659 -0.3377106 0.9150146 0.1116754 0.08266165 -0.006515221  
## male 0.25172401 -0.2249751 1.4233764 0.1829509 0.43632633 -0.664869757  
##   
## female 1.0800543 0.05345666 -0.7608237 -1.2884977 0.5944412 0.03055868  
## male 0.9714248 -0.40438198 -0.9628398 -0.3784598 0.5301931 0.38135612  
##   
## female -0.7094871 -0.27882187 0.54056035 0.19208960 -0.3358380 0.2131709  
## male -0.8174883 0.02961984 0.04096301 -0.07432882 -0.2660526 -0.5253395  
##   
## female 1.2595078 0.8990623 -0.09116556 1.142337 -0.0115831 0.28532794  
## male -0.4075378 -0.5095028 -0.59768002 1.133257 0.3075430 -0.00723112  
##   
## female 0.2432806 -0.3869904 0.1471348 1.0473533 0.07425627 -1.3533383  
## male 0.3553993 0.3607626 0.3284679 0.6808087 -0.59569523 -0.1513053  
##   
## female 0.06482483 0.3158898 -0.59290976 -0.2627614 0.3401162 0.08433392  
## male 0.03203143 0.1892784 0.03889327 0.3886355 -0.1137537 -0.25461783  
##   
## female -0.1242984 0.3751857 -0.4912555 0.4157222 0.01829567 0.8654644  
## male 0.1090990 0.2293638 0.1171273 0.5573219 -0.04766781 0.3653586  
##   
## female -1.10551135 -0.7603215 0.26888710 -1.0511110 -0.2621506 0.33021057  
## male 0.05320585 0.1546775 0.04411519 0.6110022 0.3937269 0.06073714  
##   
## female 0.23727749 -0.2604444 -0.0649688 -1.0048686 -0.1095281 0.2969898  
## male -0.02074022 0.0302774 -0.3196637 0.3286932 0.2743457 0.1112674  
##   
## female -0.9141487 0.2694757 0.06530885 0.6437093 0.9188469 -0.9053765  
## male 0.1582143 0.1305673 0.44189830 -0.8046526 0.3013798 -0.3283913  
##   
## female 1.1637446 0.4434989 -0.67703316 -0.1918490 0.02824104 -0.2222472  
## male 0.5387935 0.2349625 0.04911226 0.3648267 -0.08045425 -0.3780695  
##   
## female -0.3407362 0.71578371 0.2448074 -0.4510240 0.1363442 -0.71205350  
## male 0.6028140 0.06212063 -0.2448839 0.7336483 1.2866514 0.02207333  
##   
## female 0.1679339 -0.7133195 -0.3146768 -0.1850570 -0.3106618 0.2511258  
## male -0.3101437 0.2194559 -0.3362269 -0.1332822 0.2453832 -0.5563728  
##   
## female -1.0198783 -0.15306368 0.6686412 0.17917552 -0.08843675 -0.005473811  
## male -0.6231355 0.07728708 0.2187502 0.07209923 0.09868763 -0.580589947  
##   
## female -0.4669486 -0.2185142 0.2988355 1.14229200 -0.3895302 -0.2151374  
## male -0.1997393 -0.5065985 0.8293952 -0.06285638 -0.5800934 0.7231014  
##   
## female -0.2867562 -0.1369735 -0.8308772 -0.6196471 -0.9718702 -0.3814613  
## male 0.4015373 0.3689133 -0.2019204 -1.4262848 -0.3668824 -0.4352654  
##   
## female 0.03797288 0.1078091 0.06783057 -1.17176260 -0.16324525 0.107630  
## male 0.19573380 -0.6533899 0.33870565 -0.06544488 -0.03192761 -1.042785  
##   
## female 0.2093034 -0.6116888  
## male 0.9052546 -1.1742455

lives\_nns <- nns(target\_wv\_group, pre\_trained = cr\_glove\_subset, N = 5,  
 candidates = target\_wv\_group@features, as\_list = TRUE)

# Results for nationalists  
lives\_nns[["female"]]

## # A tibble: 5 × 4  
## target feature rank value  
## <chr> <chr> <int> <dbl>  
## 1 female little 1 0.314  
## 2 female good 2 0.272  
## 3 female hard 3 0.242  
## 4 female got 4 0.180  
## 5 female big 5 0.172

# Results for Christians  
lives\_nns[["male"]]

## # A tibble: 5 × 4  
## target feature rank value  
## <chr> <chr> <int> <dbl>  
## 1 male american 1 0.518  
## 2 male america 2 0.397  
## 3 male stand 3 0.376  
## 4 male young 4 0.362  
## 5 male kids 5 0.357

target\_sim <- cos\_sim(target\_wv\_group, pre\_trained = cr\_glove\_subset,  
 features = c("america", "american"), as\_list = TRUE)

target\_sim[["america"]]

## target feature value  
## 3 female america 0.06839516  
## 4 male america 0.39679960

target\_sim[["american"]]

## target feature value  
## 1 female american 0.1602421  
## 2 male american 0.5177563

## Questions

* Can’t get cr\_glove to work?
* WHere did lives\_dem come from?s