1824 to 1964 STM Models

Charles J. Gomez, Harshvardhan Singh

2023-08-29

library(stringr)  
library(ggplot2)  
library(wordcloud)

## Loading required package: RColorBrewer

library(stm)

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

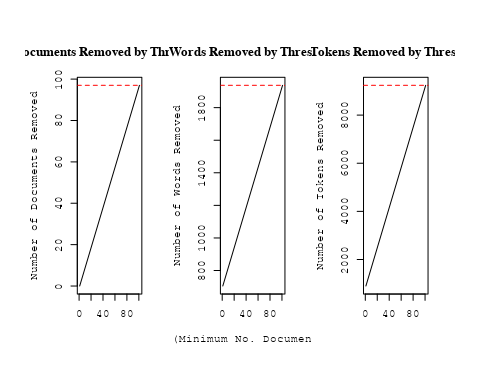
#Read csv file  
data = read.csv("preprocessed\_data\_Jul14.csv")  
  
# Define the publication intervals  
pub\_intervals <- c("pub\_interval\_1995\_1999",  
 "pub\_interval\_1985\_1994", "pub\_interval\_1975\_1984", "pub\_interval\_1965\_1974") #,"pub\_interval\_1900\_1964", "pub\_interval\_1824\_1899")  
  
  
# Define K values for each interval  
K\_values <- c(44, 45, 43, 40) #16,2  
  
  
   
# Filter data for the publication interval 1900 to 1964  
data\_filtered <- data[data[["pub\_interval\_1900\_1964"]] == 1, ]  
  
# Save the original title data for future use  
data\_filtered$original\_concatenated\_title\_abstract <- data\_filtered$concatenated\_title\_abstract  
  
#pre-processing the titles using textProcessor from the stm package  
processed\_text <- textProcessor(data\_filtered$concatenated\_title\_abstract, metadata = data\_filtered)

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

# Further prepare the data by removing low-frequency terms  
out\_text <- prepDocuments(processed\_text$documents, processed\_text$vocab, processed\_text$meta)

## Removing 705 of 1937 terms (705 of 6576 tokens) due to frequency   
## Your corpus now has 97 documents, 1232 terms and 5871 tokens.

docs\_text <- out\_text$documents  
vocab\_text <- out\_text$vocab  
meta\_text <- out\_text$meta  
  
  
#Prepare data  
plotRemoved(processed\_text$documents, lower.thresh = seq(1, 200, by = 100))



out\_text <- prepDocuments(processed\_text$documents, processed\_text$vocab, processed\_text$meta, lower.thresh = 5)

## Removing 1634 of 1937 terms (3217 of 6576 tokens) due to frequency   
## Your corpus now has 97 documents, 303 terms and 3359 tokens.

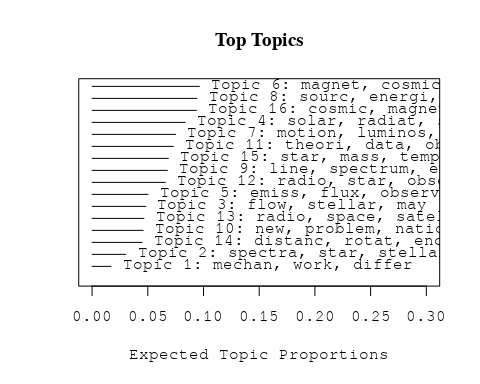
# Initialize an empty formula string  
prevalence\_formula\_str <- "~US "  
  
  
# Add each country variable to the formula string  
for (i in 10:27) {  
 prevalence\_formula\_str <- paste(prevalence\_formula\_str, "+", names(data)[i])  
}  
  
  
# Convert the string to a formula  
prevalence\_formula <- as.formula(prevalence\_formula\_str)  
print(prevalence\_formula)

## ~US + IN + DE + CH + GB + CN + FR + IT + RU + CA + NL + AU +   
## JP + ES + IL + Americas + Europe + Africa + AsiaAndOceania

# Run STM model  
Research\_topics <- stm(documents = out\_text$documents,   
 vocab = out\_text$vocab,   
 K = 16,   
 prevalence = prevalence\_formula,   
 data = out\_text$meta,   
 init.type = "Spectral",  
 max.em.its = 1000,  
 gamma.prior = 'L1')

## Beginning Spectral Initialization   
## Calculating the gram matrix...  
## Finding anchor words...  
## ................  
## Recovering initialization...  
## ...  
## Initialization complete.  
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 1 (approx. per word bound = -5.127)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 2 (approx. per word bound = -4.844, relative change = 5.511e-02)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 3 (approx. per word bound = -4.743, relative change = 2.081e-02)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 4 (approx. per word bound = -4.695, relative change = 1.008e-02)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 5 (approx. per word bound = -4.667, relative change = 6.156e-03)   
## Topic 1: mechan, work, differ, rather, detail   
## Topic 2: spectra, star, physic, theori, spectrum   
## Topic 3: flow, stellar, may, dynam, distanc   
## Topic 4: solar, sun, observ, obtain, bright   
## Topic 5: emiss, flux, observ, nebula, appear   
## Topic 6: cosmic, magnet, ray, near, storm   
## Topic 7: motion, luminos, star, observ, proper   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, spectrum, emiss, solar, region   
## Topic 10: temperatur, problem, new, surfac, nation   
## Topic 11: field, theori, data, observ, can   
## Topic 12: radio, observ, use, star, record   
## Topic 13: satellit, space, frequenc, radio, earth   
## Topic 14: distanc, star, rotat, object, enorm   
## Topic 15: star, mass, temperatur, energi, binari   
## Topic 16: cosmic, magnet, intens, ray, variat   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 6 (approx. per word bound = -4.649, relative change = 3.831e-03)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 7 (approx. per word bound = -4.636, relative change = 2.678e-03)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 8 (approx. per word bound = -4.627, relative change = 2.030e-03)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 9 (approx. per word bound = -4.620, relative change = 1.575e-03)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 10 (approx. per word bound = -4.614, relative change = 1.283e-03)   
## Topic 1: mechan, work, differ, rather, detail   
## Topic 2: spectra, star, continu, physic, spectrum   
## Topic 3: flow, stellar, may, distanc, dynam   
## Topic 4: solar, sun, observ, polar, bright   
## Topic 5: emiss, flux, observ, record, nebula   
## Topic 6: cosmic, magnet, ray, near, storm   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, spectrum, emiss, solar, region   
## Topic 10: new, problem, temperatur, surfac, nation   
## Topic 11: field, theori, data, observ, can   
## Topic 12: radio, observ, star, use, light   
## Topic 13: radio, space, satellit, frequenc, orbit   
## Topic 14: distanc, rotat, enorm, object, find   
## Topic 15: star, mass, temperatur, energi, binari   
## Topic 16: cosmic, magnet, intens, ray, variat   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 11 (approx. per word bound = -4.608, relative change = 1.202e-03)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 12 (approx. per word bound = -4.602, relative change = 1.266e-03)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 13 (approx. per word bound = -4.597, relative change = 1.228e-03)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 14 (approx. per word bound = -4.593, relative change = 7.030e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 15 (approx. per word bound = -4.591, relative change = 4.795e-04)   
## Topic 1: mechan, work, differ, rather, detail   
## Topic 2: spectra, star, continu, stellar, physic   
## Topic 3: flow, stellar, may, distanc, dynam   
## Topic 4: solar, sun, observ, polar, bright   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: cosmic, magnet, ray, near, storm   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, spectrum, emiss, solar, radio   
## Topic 10: new, temperatur, problem, surfac, nation   
## Topic 11: theori, field, data, observ, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, object, find   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, intens, magnet, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 16 (approx. per word bound = -4.589, relative change = 3.922e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 17 (approx. per word bound = -4.588, relative change = 3.173e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 18 (approx. per word bound = -4.587, relative change = 2.881e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 19 (approx. per word bound = -4.585, relative change = 2.859e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 20 (approx. per word bound = -4.584, relative change = 2.488e-04)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, continu, stellar, spectrum   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, sun, time, observ, polar   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: cosmic, magnet, ray, near, field   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, spectrum, emiss, solar, radio   
## Topic 10: new, problem, nation, equilibrium, surfac   
## Topic 11: theori, field, data, observ, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, find, object   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, intens, magnet, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 21 (approx. per word bound = -4.583, relative change = 2.013e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 22 (approx. per word bound = -4.582, relative change = 1.755e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 23 (approx. per word bound = -4.582, relative change = 1.614e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 24 (approx. per word bound = -4.581, relative change = 1.593e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 25 (approx. per word bound = -4.580, relative change = 1.577e-04)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, spectrum   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, sun, time, observ, polar   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, emiss, spectrum, solar, radio   
## Topic 10: new, problem, nation, equilibrium, surfac   
## Topic 11: theori, field, data, observ, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, find, object   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, intens, magnet, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 26 (approx. per word bound = -4.579, relative change = 1.584e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 27 (approx. per word bound = -4.579, relative change = 1.729e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 28 (approx. per word bound = -4.578, relative change = 1.734e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 29 (approx. per word bound = -4.577, relative change = 1.750e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 30 (approx. per word bound = -4.576, relative change = 1.815e-04)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, spectrum   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, sun, time, observ, polar   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, surfac   
## Topic 11: theori, field, data, observ, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, find, object   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, intens, magnet, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 31 (approx. per word bound = -4.575, relative change = 1.926e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 32 (approx. per word bound = -4.574, relative change = 2.077e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 33 (approx. per word bound = -4.574, relative change = 1.720e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 34 (approx. per word bound = -4.573, relative change = 1.176e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 35 (approx. per word bound = -4.573, relative change = 9.390e-05)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, spectrum   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, sun, time, observ, polar   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, surfac   
## Topic 11: theori, data, field, observ, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, find, object   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, intens, magnet, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 36 (approx. per word bound = -4.572, relative change = 8.572e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 37 (approx. per word bound = -4.572, relative change = 7.591e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 38 (approx. per word bound = -4.572, relative change = 6.426e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 39 (approx. per word bound = -4.571, relative change = 6.113e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 40 (approx. per word bound = -4.571, relative change = 6.178e-05)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, spectrum   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, sun, time, radiat, observ   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, surfac   
## Topic 11: theori, data, field, observ, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, find, can   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, intens, magnet, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 41 (approx. per word bound = -4.571, relative change = 6.207e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 42 (approx. per word bound = -4.570, relative change = 6.827e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 43 (approx. per word bound = -4.570, relative change = 7.198e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 44 (approx. per word bound = -4.570, relative change = 8.062e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 45 (approx. per word bound = -4.569, relative change = 9.091e-05)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, sun, radiat, time, observ   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, intens   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, field, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, find   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, intens, magnet, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 46 (approx. per word bound = -4.569, relative change = 1.103e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 47 (approx. per word bound = -4.568, relative change = 1.077e-04)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 48 (approx. per word bound = -4.568, relative change = 9.378e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 49 (approx. per word bound = -4.568, relative change = 8.795e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 50 (approx. per word bound = -4.567, relative change = 6.036e-05)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, radiat, sun, time, observ   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: radiat, sourc, energi, inform, distribut   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, field, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, find   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, intens, magnet, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 51 (approx. per word bound = -4.567, relative change = 6.196e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 52 (approx. per word bound = -4.567, relative change = 8.660e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 53 (approx. per word bound = -4.566, relative change = 9.759e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 54 (approx. per word bound = -4.566, relative change = 9.527e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 55 (approx. per word bound = -4.565, relative change = 7.571e-05)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, radiat, sun, time, observ   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: sourc, radiat, energi, inform, distribut   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, field, can   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, find   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, magnet, intens, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 56 (approx. per word bound = -4.565, relative change = 4.835e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 57 (approx. per word bound = -4.565, relative change = 3.565e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 58 (approx. per word bound = -4.565, relative change = 3.068e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 59 (approx. per word bound = -4.565, relative change = 3.134e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 60 (approx. per word bound = -4.565, relative change = 3.218e-05)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, radiat, sun, time, observ   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: sourc, energi, radiat, inform, distribut   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, can, field   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, find   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, magnet, intens, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 61 (approx. per word bound = -4.564, relative change = 2.970e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 62 (approx. per word bound = -4.564, relative change = 2.962e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 63 (approx. per word bound = -4.564, relative change = 3.200e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 64 (approx. per word bound = -4.564, relative change = 4.171e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 65 (approx. per word bound = -4.564, relative change = 5.178e-05)   
## Topic 1: mechan, work, differ, star, detail   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, radiat, sun, time, intens   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: sourc, energi, radiat, inform, distribut   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, can, field   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, find   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, magnet, intens, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 66 (approx. per word bound = -4.563, relative change = 6.190e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 67 (approx. per word bound = -4.563, relative change = 4.806e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 68 (approx. per word bound = -4.563, relative change = 3.167e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 69 (approx. per word bound = -4.563, relative change = 2.441e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 70 (approx. per word bound = -4.563, relative change = 2.437e-05)   
## Topic 1: mechan, work, differ, star, rather   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, radiat, sun, time, intens   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: sourc, energi, inform, radiat, distribut   
## Topic 9: line, emiss, spectrum, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, can, field   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, densiti   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, magnet, intens, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 71 (approx. per word bound = -4.563, relative change = 2.712e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 72 (approx. per word bound = -4.563, relative change = 2.238e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 73 (approx. per word bound = -4.563, relative change = 1.679e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 74 (approx. per word bound = -4.562, relative change = 1.646e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 75 (approx. per word bound = -4.562, relative change = 1.626e-05)   
## Topic 1: mechan, work, differ, star, rather   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, distanc, star   
## Topic 4: solar, radiat, sun, time, intens   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: sourc, energi, inform, radiat, distribut   
## Topic 9: line, spectrum, emiss, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, can, field   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, densiti   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, magnet, intens, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 76 (approx. per word bound = -4.562, relative change = 1.822e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 77 (approx. per word bound = -4.562, relative change = 2.352e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 78 (approx. per word bound = -4.562, relative change = 3.602e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 79 (approx. per word bound = -4.562, relative change = 5.652e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 80 (approx. per word bound = -4.561, relative change = 6.731e-05)   
## Topic 1: mechan, work, differ, star, rather   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, star, distanc   
## Topic 4: solar, radiat, sun, time, intens   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: sourc, energi, inform, radiat, distribut   
## Topic 9: line, spectrum, emiss, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, can, field   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, densiti   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, magnet, intens, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 81 (approx. per word bound = -4.561, relative change = 6.338e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 82 (approx. per word bound = -4.561, relative change = 5.834e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 83 (approx. per word bound = -4.561, relative change = 4.137e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 84 (approx. per word bound = -4.561, relative change = 3.188e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 85 (approx. per word bound = -4.560, relative change = 2.720e-05)   
## Topic 1: mechan, work, differ, star, rather   
## Topic 2: spectra, star, stellar, continu, physic   
## Topic 3: flow, stellar, may, star, distanc   
## Topic 4: solar, radiat, sun, intens, time   
## Topic 5: emiss, flux, observ, radio, record   
## Topic 6: magnet, cosmic, ray, field, near   
## Topic 7: motion, luminos, star, observ, distanc   
## Topic 8: sourc, energi, inform, radiat, distribut   
## Topic 9: line, spectrum, emiss, solar, theori   
## Topic 10: new, problem, nation, equilibrium, unit   
## Topic 11: theori, data, observ, can, field   
## Topic 12: radio, star, observ, use, light   
## Topic 13: radio, space, satellit, frequenc, research   
## Topic 14: distanc, rotat, enorm, can, densiti   
## Topic 15: star, mass, temperatur, model, energi   
## Topic 16: cosmic, magnet, intens, ray, latitud   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 86 (approx. per word bound = -4.560, relative change = 2.655e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 87 (approx. per word bound = -4.560, relative change = 1.986e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 88 (approx. per word bound = -4.560, relative change = 1.191e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 89 (approx. per word bound = -4.560, relative change = 1.109e-05)   
## .................................................................................................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Model Converged

# Plot the STM model summary  
plot(Research\_topics, type = "summary", xlim = c(0, 0.3))



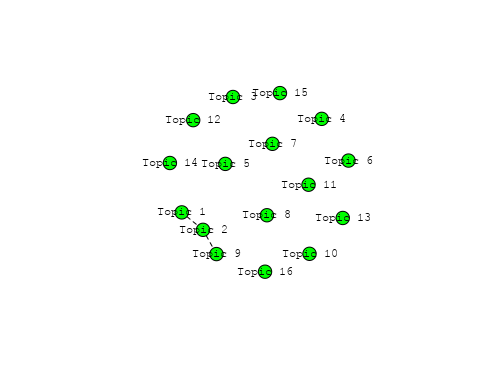
# Print the top 10 labels for each topic  
topic\_labels <- labelTopics(Research\_topics, n=10)  
print(topic\_labels)

## Topic 1 Top Words:  
## Highest Prob: mechan, work, differ, star, rather, detail, one, relat, type, paper   
## FREX: mechan, differ, work, rather, relat, detail, toward, order, paper, one   
## Lift: mechan, differ, rather, work, explan, toward, clear, detail, relat, systemat   
## Score: mechan, differ, work, detail, star, explan, rather, relat, light, toward   
## Topic 2 Top Words:  
## Highest Prob: spectra, star, stellar, continu, physic, spectrum, recent, suggest, author, problem   
## FREX: spectra, continu, author, recent, physic, stellar, wide, suggest, star, problem   
## Lift: spectra, author, believ, continu, wide, curv, certain, product, recent, physic   
## Score: spectra, author, star, physic, continu, stellar, temperatur, problem, publish, curv   
## Topic 3 Top Words:  
## Highest Prob: flow, stellar, may, star, distanc, dynam, extend, integr, main, region   
## FREX: flow, integr, dynam, stellar, equat, extend, main, may, radial, radius   
## Lift: flow, integr, equat, dynam, radial, radius, carri, properti, stellar, veloc   
## Score: integr, flow, dynam, stellar, distanc, radius, radial, equat, may, veloc   
## Topic 4 Top Words:  
## Highest Prob: solar, radiat, sun, intens, time, observ, bright, polar, larg, result   
## FREX: solar, sun, radiat, polar, bright, time, direct, sunspot, intens, obtain   
## Lift: polar, solar, bright, sunspot, great, atmospher, direct, sun, expect, time   
## Score: polar, solar, radiat, sun, sunspot, bright, obtain, time, variat, eclips   
## Topic 5 Top Words:  
## Highest Prob: emiss, flux, observ, radio, record, ionospher, sourc, appear, intens, even   
## FREX: flux, emiss, phase, record, even, number, track, sunspot, ionospher, signal   
## Lift: flux, phase, signal, emiss, track, sunspot, total, rang, interv, number   
## Score: flux, emiss, record, sunspot, nebula, phase, interv, total, signal, track   
## Topic 6 Top Words:  
## Highest Prob: magnet, cosmic, ray, field, near, storm, earth, particl, electron, forc   
## FREX: magnet, near, storm, ray, particl, cosmic, disturb, earth, electron, field   
## Lift: storm, near, disturb, note, forc, magnet, particl, ray, basi, electron   
## Score: storm, ray, magnet, cosmic, near, particl, earth, field, note, forc   
## Topic 7 Top Words:  
## Highest Prob: motion, luminos, star, observ, distanc, proper, caus, magnitud, characterist, two   
## FREX: motion, luminos, proper, caus, magnitud, characterist, thus, determin, absolut, system   
## Lift: caus, proper, motion, luminos, magnitud, thus, characterist, absolut, relat, galact   
## Score: caus, luminos, motion, proper, distanc, absolut, magnitud, thus, express, system   
## Topic 8 Top Words:  
## Highest Prob: sourc, energi, inform, radiat, distribut, high, develop, initi, observ, produc   
## FREX: inform, sourc, initi, develop, distribut, high, produc, fluctuat, energi, communic   
## Lift: initi, survey, develop, inform, nuclear, wide, sourc, fluctuat, technic, distribut   
## Score: initi, energi, sourc, inform, communic, high, fluctuat, survey, radiat, technic   
## Topic 9 Top Words:  
## Highest Prob: line, spectrum, emiss, solar, theori, radio, region, use, effect, radiat   
## FREX: line, emiss, spectrum, solar, express, region, veloc, strong, absorpt, theori   
## Lift: line, express, emiss, spectrum, strong, latter, approach, veloc, radial, ioniz   
## Score: line, emiss, solar, express, spectrum, absorpt, envelop, latter, radio, ceti   
## Topic 10 Top Words:  
## Highest Prob: new, problem, nation, equilibrium, unit, surfac, temperatur, provid, far, area   
## FREX: new, equilibrium, nation, unit, problem, surfac, area, far, provid, term   
## Lift: equilibrium, new, surfac, area, unit, materi, nation, problem, term, far   
## Score: equilibrium, new, nation, unit, problem, materi, far, area, provid, temperatur   
## Topic 11 Top Words:  
## Highest Prob: theori, data, observ, can, field, compon, ionospher, cosmic, analysi, shown   
## FREX: data, compon, analysi, theori, possibl, examin, electron, ionospher, consist, can   
## Lift: data, analysi, examin, compar, featur, agre, correl, compon, consist, possibl   
## Score: data, examin, analysi, compon, cosmic, ionospher, featur, variat, can, electron   
## Topic 12 Top Words:  
## Highest Prob: radio, star, observ, use, light, ceti, made, record, correct, spectrum   
## FREX: ceti, radio, light, correct, made, state, record, previous, use, observ   
## Lift: correct, ceti, previous, state, light, oper, telescop, diamet, uniform, astronomi   
## Score: correct, ceti, radio, light, star, record, previous, made, oper, observatori   
## Topic 13 Top Words:  
## Highest Prob: radio, space, satellit, frequenc, research, receiv, ionospher, purpos, earth, region   
## FREX: space, satellit, frequenc, research, purpos, radio, orbit, receiv, communic, ionospher   
## Lift: satellit, space, purpos, research, numer, frequenc, second, requir, interfer, orbit   
## Score: satellit, space, radio, frequenc, orbit, purpos, communic, research, earth, ionospher   
## Topic 14 Top Words:  
## Highest Prob: distanc, rotat, enorm, can, densiti, find, object, seem, nebula, given   
## FREX: rotat, enorm, find, densiti, distanc, seem, object, evid, nebula, structur   
## Lift: enorm, rotat, find, densiti, seem, actual, appar, object, nebula, evid   
## Score: enorm, distanc, rotat, object, nebula, find, mass, densiti, seem, evid   
## Topic 15 Top Words:  
## Highest Prob: star, mass, temperatur, model, energi, binari, effect, central, stellar, envelop   
## FREX: binari, model, mass, central, eclips, envelop, fraction, temperatur, star, process   
## Lift: binari, fraction, envelop, eclips, central, model, process, rate, mass, studi   
## Score: binari, mass, star, temperatur, eclips, envelop, model, energi, fraction, physic   
## Topic 16 Top Words:  
## Highest Prob: cosmic, magnet, intens, ray, latitud, variat, energi, particl, horizont, spectrum   
## FREX: latitud, intens, magnet, horizont, variat, cosmic, ray, amplitud, certain, particl   
## Lift: horizont, latitud, certain, amplitud, repres, consequ, relationship, variat, assum, intens   
## Score: horizont, ray, cosmic, magnet, latitud, variat, intens, relationship, particl, energi

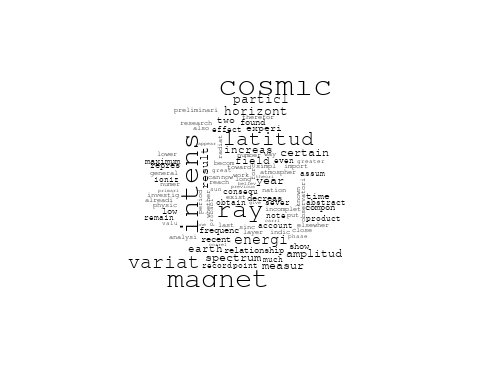
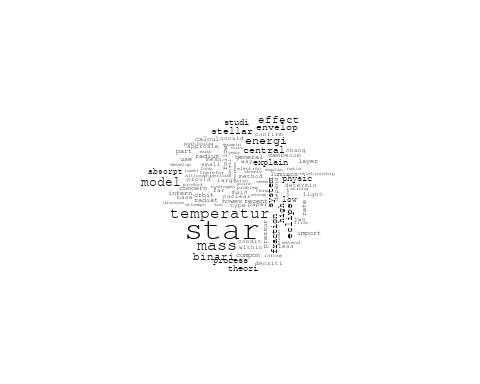
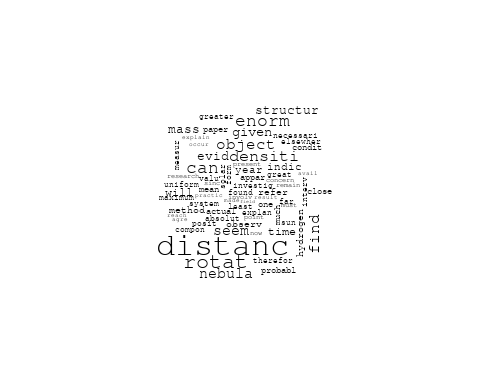
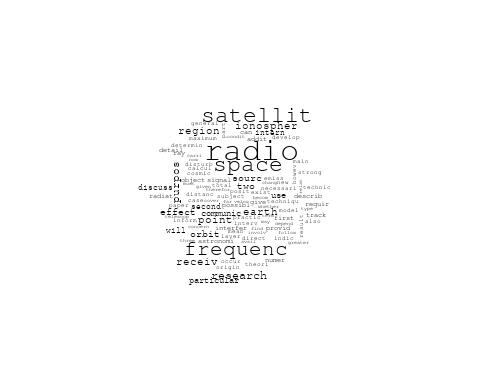
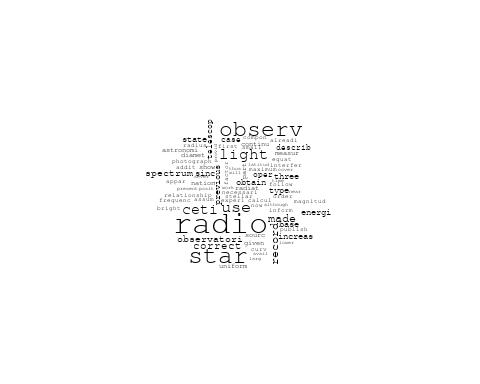
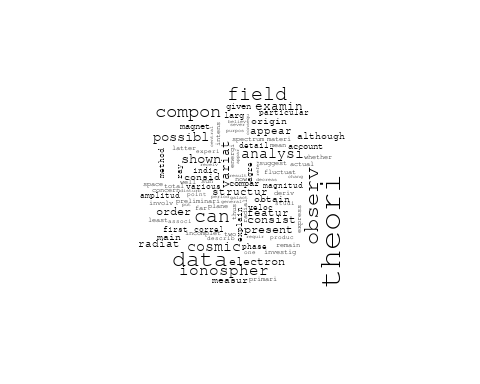
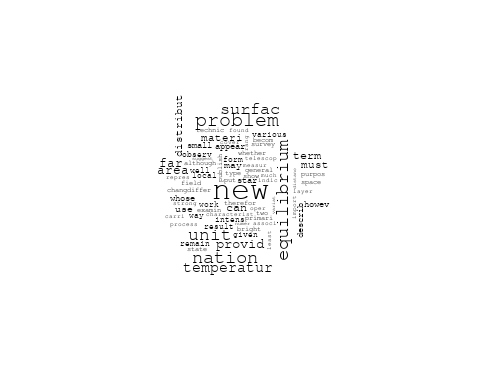
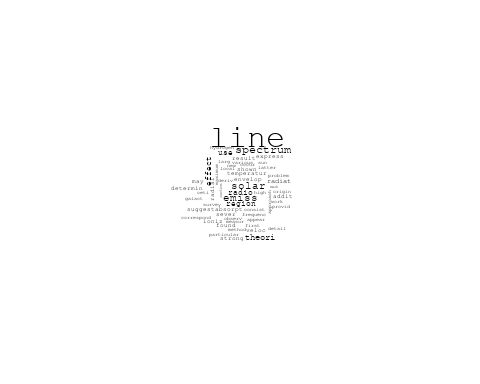
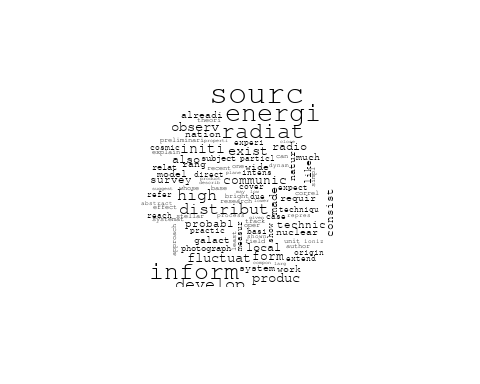
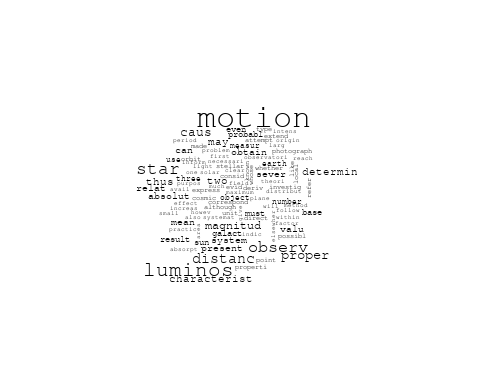
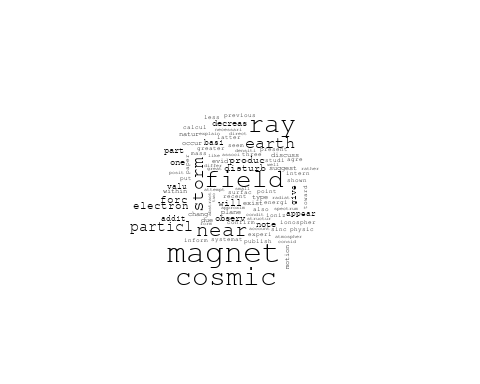
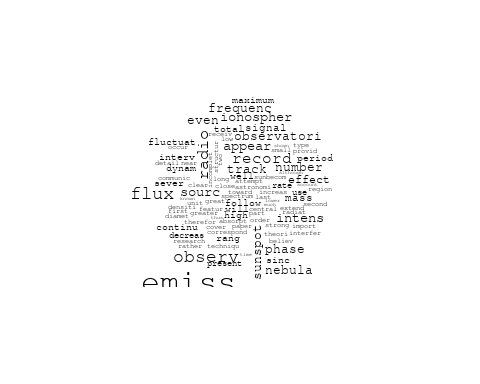
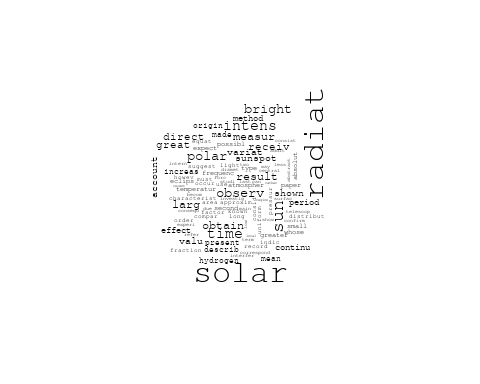
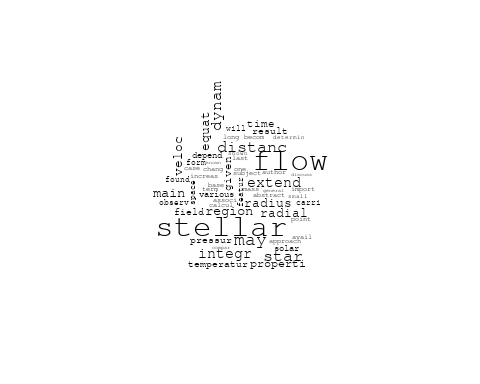
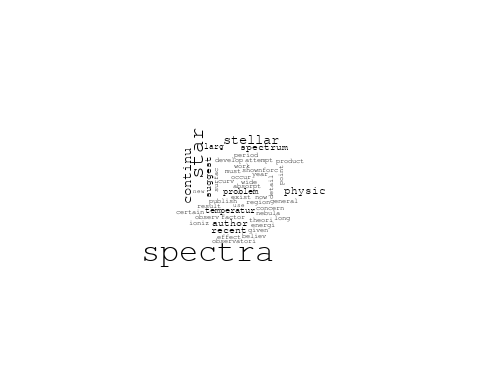
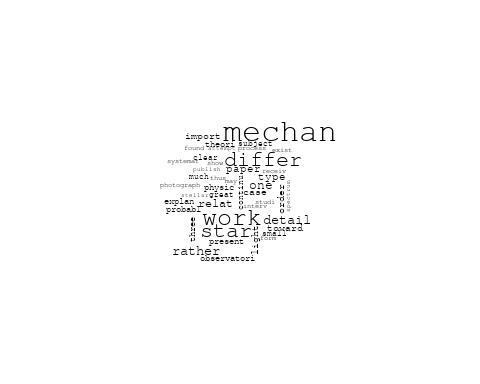
# Match the processed documents with the original titles  
matched\_titles <- out\_text$meta$original\_concatenated\_title\_abstract  
  
# Print top 5 documents for each topic  
top\_docs <- findThoughts(Research\_topics, texts = matched\_titles, n = 5)$docs[[1]]  
print(top\_docs)

## [1] "On a Physical Theory of Stellar Spectra The present paper embodies an attempt towards a physical explanation of the ordered gradation in the spectra of stars—a subject in which pioneering work was done by the late Sir Norman Lockyer, but which was worked up with systematic thoroughness at the Harvard College Observatory, under the lead of the late Prof. E. C. Pickering and Miss A. J. Cannon. During this interval the spectra of more than 100,000 stars have been photographed, classified, and published with full details in the Henry Draper Memorial Catalogue. The most noteworthy facts which have been brought to light from these monumental studies have thus been summarised by H. N. Russell. \"The spectra of the stars show remarkably few radical differences in type. More than 99 per cent. of them fall into one or the other of the six great groups which during the classic work of the Harvard College Observatory were recognised as of fundamental importance, and received as designations, by the process of the survival of the fittest, the rather arbitrary letters B, A, F, G, K, M. That there should be so few types is noteworthy, but much more remarkable is the fact that they form a continuous series. Every degree of gradation between the typical spectra denoted by B and A may be found in different stars, and the same is true to the end of the series, a fact recognised in the familiar decimal classification, in which B5A, for example, denotes a spectrum half-way between the typical examples B and A. The series is not merely continuous, it is linear. There exists slight difference between the spectra of different stars of the same spectral class, such as Ao, but these relate to minor details. Almost all the stars of the small outstanding minority fall into three other classes (or rather four), denoted by the letters P, O, N, R. Of these, O undoubtedly precedes B at the head of the series, while R and N, which grade one into the other, come probably at its other end, though in this case the transition stages are not clearly worked out.”"   
## [2] "On the Theory of Coincidence Experiments on Cosmic Rays It is now generally agreed that the cosmic radiation consists of two groups called the soft and the hard component. This classification refers to the penetrating power of the particles in question. The soft component consists of ordinary electrons and light quanta, whereas it now seems very probable that the hard group consists of \"heavy electrons\". The phenomena connected with the soft component are described by the cascade theory of showers put forward by Bhabha and Heitler (1937) and Carlson and Oppenheimer (1937). It is the purpose of the present paper to work out the consequences of the theory in greater detail in order to make possible a more quantitative comparison between theory and experiment than has hitherto been possible. This is desirable for several reasons: In the first place it is important to ascertain whether the phenomena connected with the soft component can be accountant for completely by the cascade theory or whether some of the facts require an explanation outside the present relativistic quantum mechanics."   
## [3] "Laboratory Spectra of the Ultraviolet Oxygen Airglow Laboratory spectra of the ultraviolet oxygen night airglow - atomic oxygen excitation mechanisms"   
## [4] "THE DISTRIBUTION OF DENSITY WITHIN THE STARS To find out anything about the distribution of matter inside a star demands the use of an “analytical boring machine ”-as Eddington says-and it is only under rather special circumstances that such a machine can get to work. Three different methods, however, have proved available, not merely “in principle’’ but in practical application. Two of these are of a “classical” nature, and depend essentially upon gravitation; the third involves modern astrophysics. I will speak of this first,-briefly, since it does not really come under our morning’s title. The equilibrium of the gaseous matter inside a star is now pretty well understood. At least, almost all the principal investigators have been led by their researches to agree upon the main principles. The principal observable characteristics of a star are its mass, its radius, and its luminosity. If we know these, we can find the value of gravity at its surface, the surface temperature, and the rate at which heat flows out per unit area. We can then work inward from the surface, calculating the density, pressure and temperature in deeper layers, provided that we know two things about the material, the mean molceular weight and the opacity, which determines the rate of rise of temperature with depth. Fortunately for us, both these quantities, under stellar conditions, depend mainly on the percentage of hydrogen which is present. The relative proportions of other kinds of atoms do not matter much (though helium has some influence). Chandrasekhar has developed an effective way of doing this, by which one can figure out with good approximation how deep we would have to go to leave behind us a given fraction (say one-tenth) of the whole mass of the star. The remaining 90 per cent of the mass will then be enclosed in a sphere whose radius is a given fraction (which Chandrasekhar calls E\*) of the whole star. For a star of uniform density, E\* would be 0.965. Smaller values indicate that the density increases toward the center. Applying his method to Capella, Chandrasekhar finds that if there is no hydrogen at all inside, 4\* = 0.68, indicating a small central condensation. With 30 per cent of hydrogen it drops to 0.50; with 80 per"  
## [5] "THE ABSORPTION COEFFICIENT AND THE LUMINOSITY CURVE IN THE GALACTIC PLANE In this paper the coefficient of absorption is derived for the case that this coefficient is a constant and that the stellar density is uniform throughout a thin layer along the galactic plane. Such a simple theory, although known not to apply in detail to our galaxy, ie likely to give a reasonably good average value, if applied to observations averaged over all galactic longitudes. The results are based on the counts of stars of a given apparent magnitude and since these include stars at widely different distances, the results are not apt to be influenced-by irregularities in the density extending over distances of the order of ten or even a hundred parsecs. The fact that some observers find a local cluster and that others find indication of a local region of avoidance seems to justify the stated case at the present stage of our knowledge. On the basis of a rotational symmetry in the stellar system the “uniform” theory could be expected to correspond to the actual case for directions 90° away from the galactic center. If we adopt as unit of distance the distance a t which a star undergoes an absorption of one magnitude and if we denote the absolute magnitude relative to this distance by M we can compute the relative numbers of stars brighter than m whose absolute magnitude is M. Putting 5 log T = p, we have m M = p + 10°.2p (1) and F ( m L M ) = 10°.op has been computed for values of m M from 15 to + 15 (See TABLE 1). By differencing F we obtain"

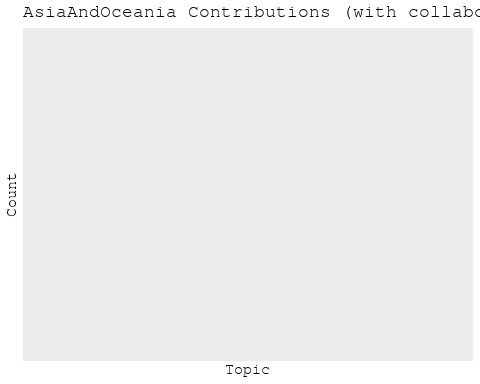
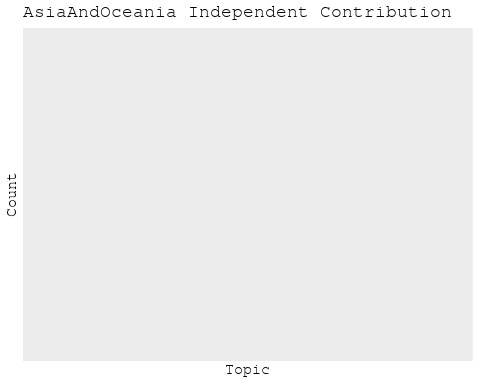
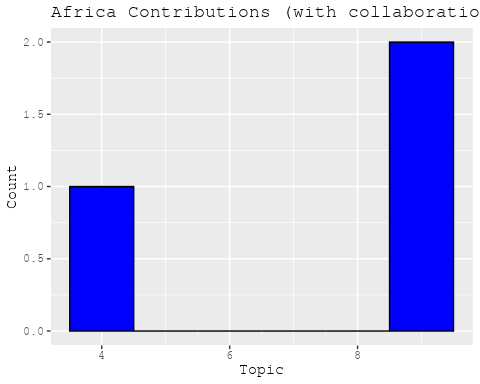
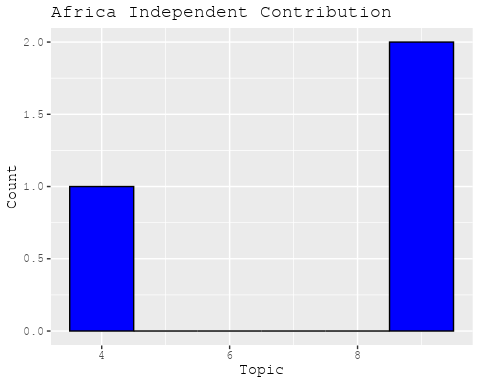
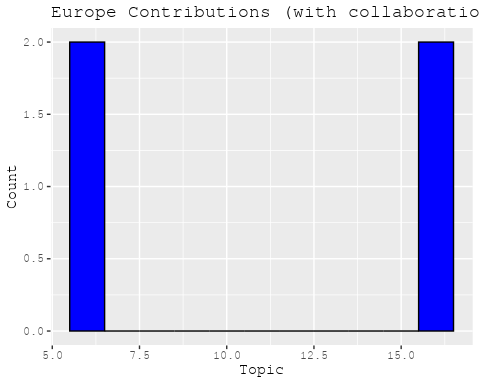
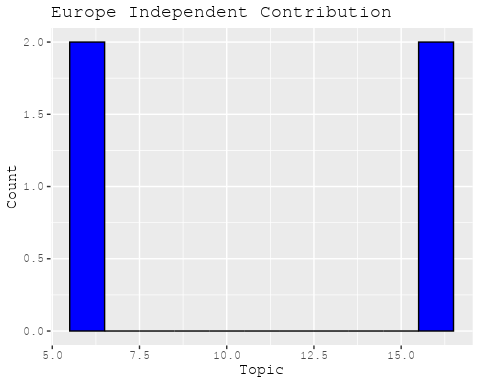
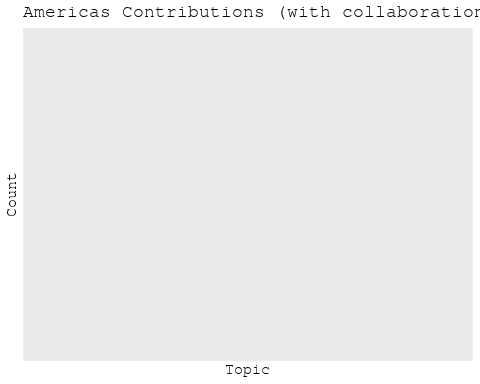
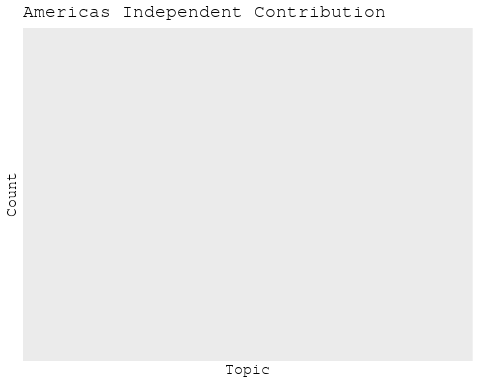
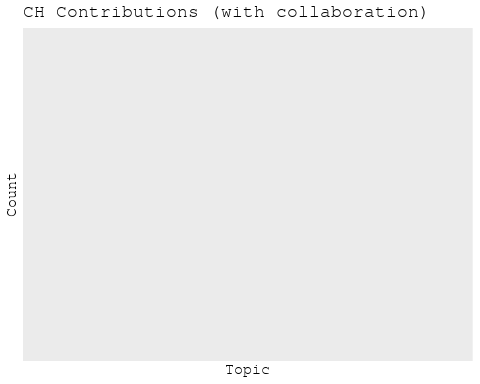
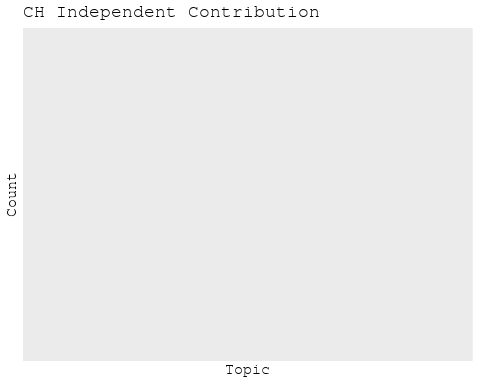
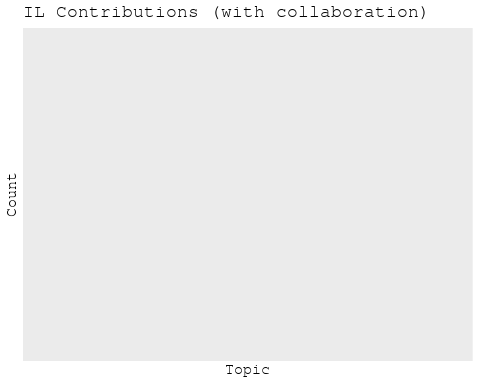
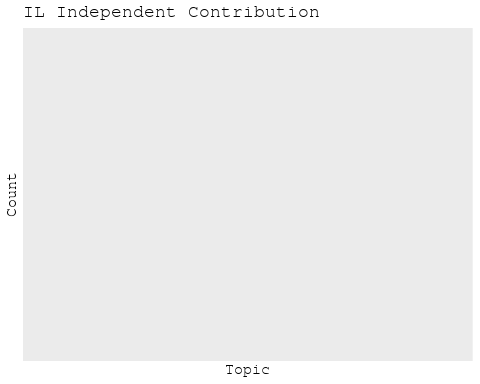
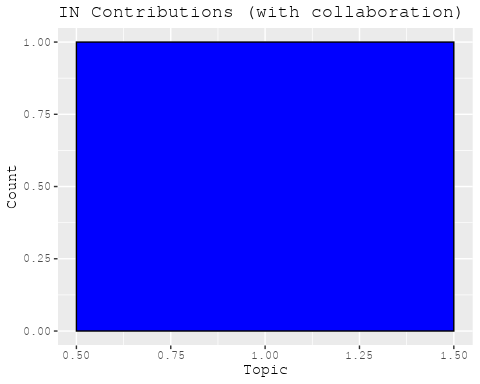
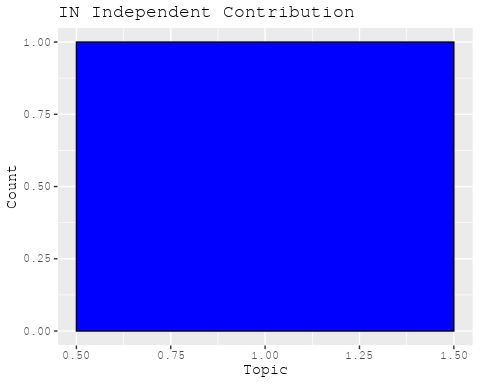
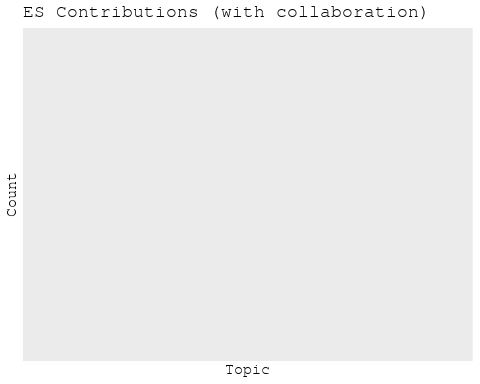
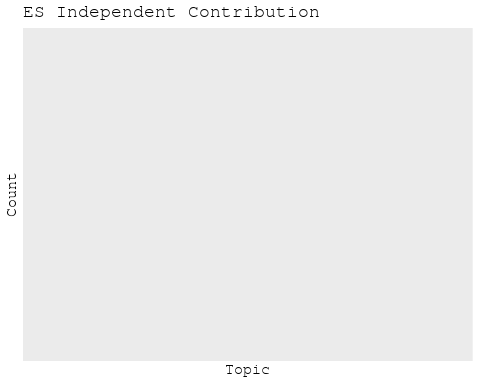
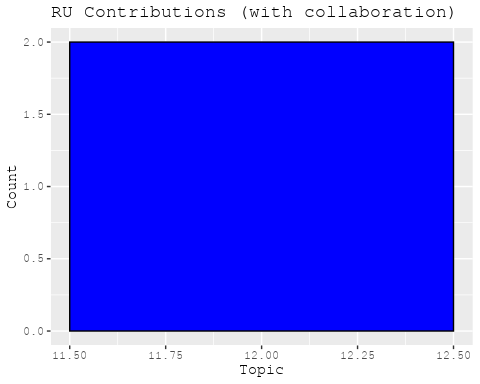
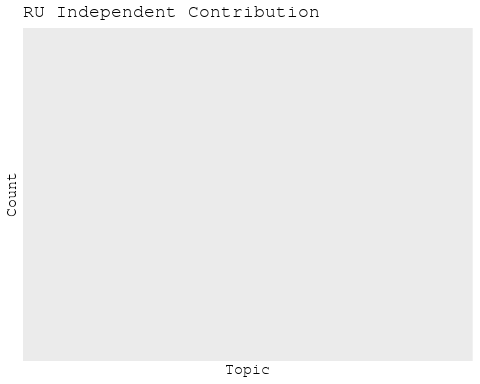
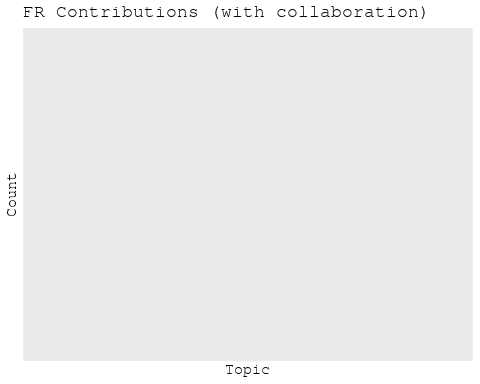
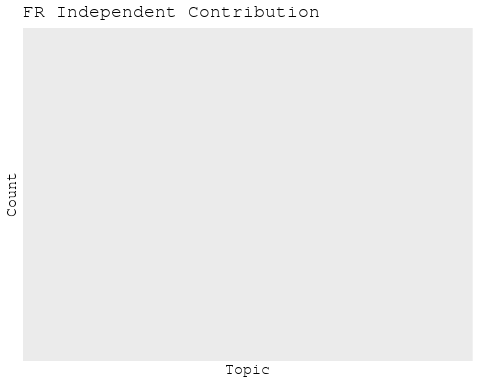
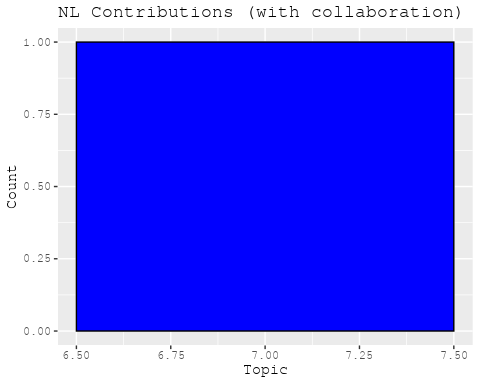
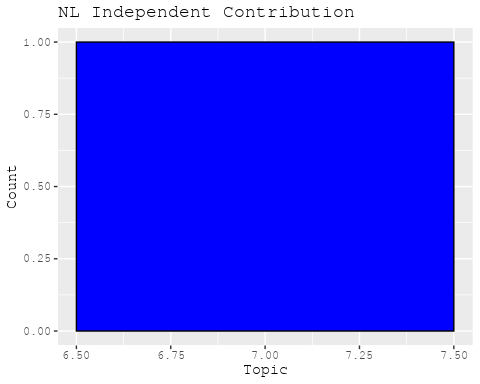
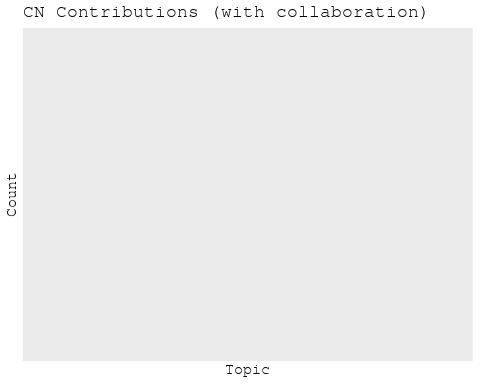
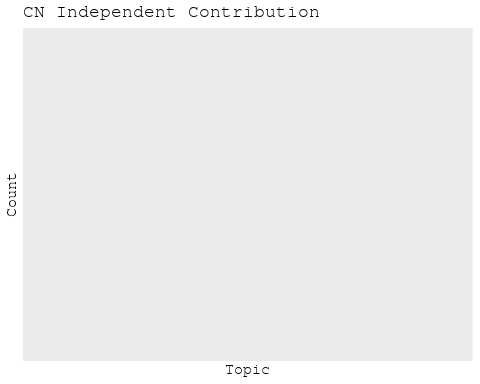
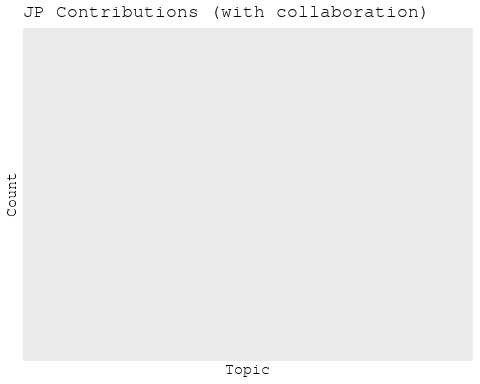
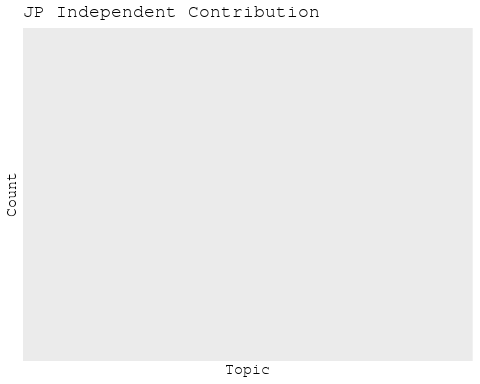
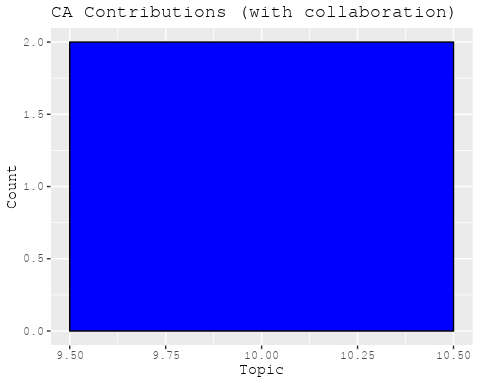
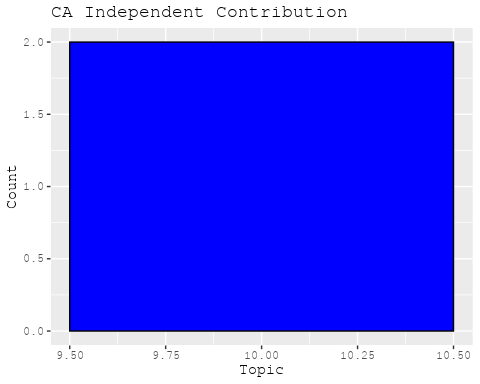
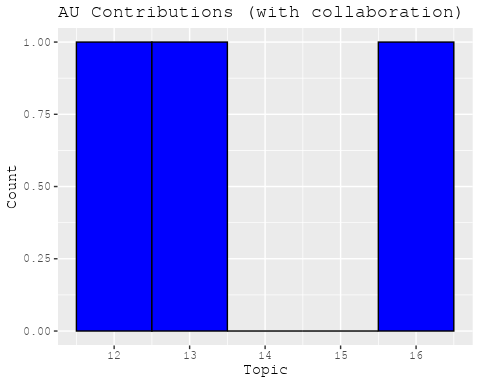
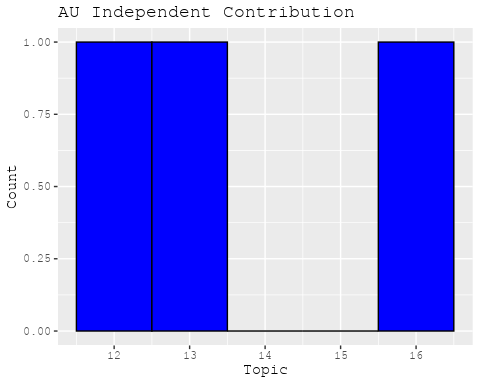
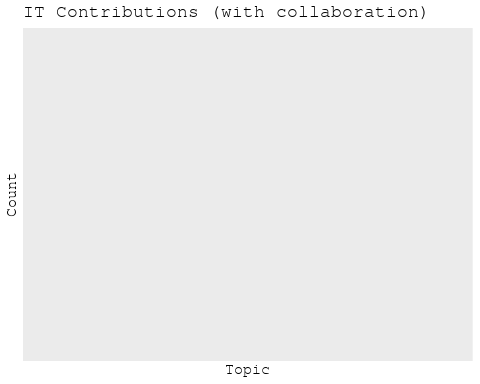
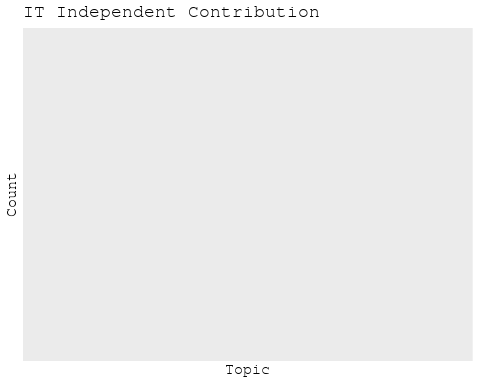
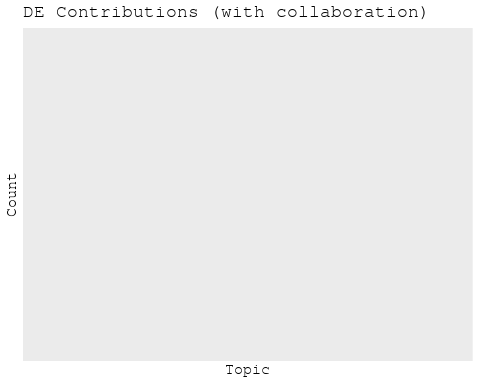
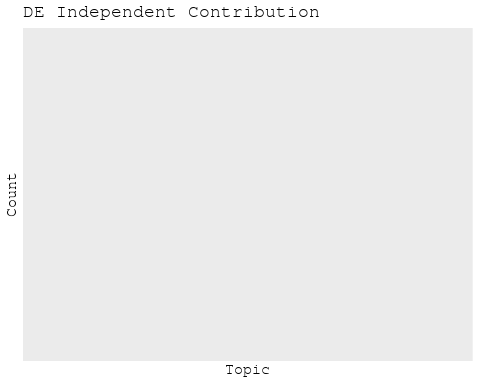
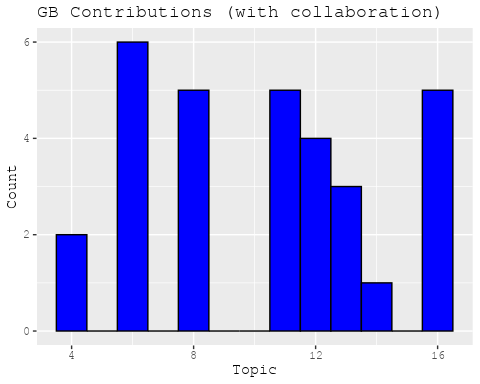
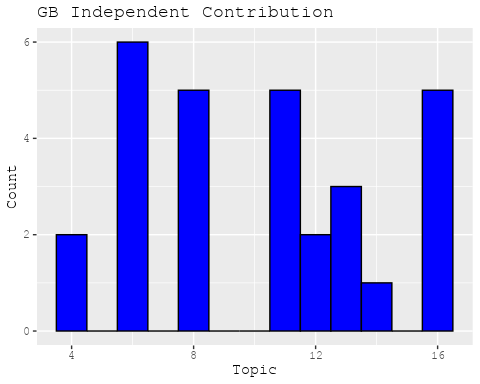
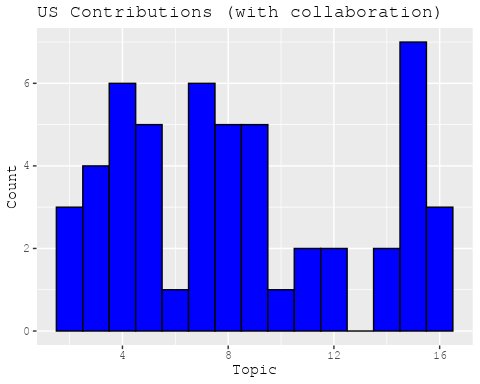
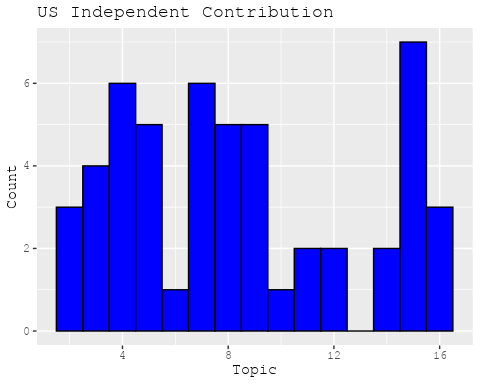
# Calculate and plot the correlation between topics  
mod.out.corr <- topicCorr(Research\_topics)  
plot(mod.out.corr, cex = 1.5)



# For each topic  
for (topic\_num in 1:16) {  
 # Plot the word cloud  
 cloud(Research\_topics, topic = topic\_num, scale = c(2, 0.25))  
 Sys.sleep(2)  
}



# Get the topic proportions for each document  
topic\_proportions <- Research\_topics$theta  
  
# Find the index of the topic with the highest proportion for each document  
# This will be the topic that each document is most likely to belong to  
max\_topic\_idx <- apply(topic\_proportions, 1, which.max)  
  
# Add this as a new column to your data  
data\_filtered$topic <- max\_topic\_idx  
  
  
# Define the countries  
countries <- c("US", "GB", "DE", "IT", "AU", "CA", "JP", "CN", "NL", "FR", "RU", "ES", "IN", "IL", "CH", "Americas", "Europe", "Africa", "AsiaAndOceania")  
  
  
  
# Create function to generate histogram  
make\_histogram <- function(country\_code) {  
   
 # Filter data for country with contribution of 100  
 data\_100 <- data\_filtered[data\_filtered[[country\_code]] == 100,]  
   
 # Filter data for country with contribution not equal to 0  
 data\_not\_0 <- data\_filtered[data\_filtered[[country\_code]] != 0,]  
   
 # Make the first histogram  
 hist\_100 <- ggplot(data\_100, aes(x = topic)) +  
 geom\_histogram(binwidth = 1, fill = "blue", color = "black") +  
 xlab("Topic") +  
 ylab("Count") +  
 ggtitle(paste(country\_code, "Independent Contribution"))  
   
 # Make the second histogram  
 hist\_not\_0 <- ggplot(data\_not\_0, aes(x = topic)) +  
 geom\_histogram(binwidth = 1, fill = "blue", color = "black") +  
 xlab("Topic") +  
 ylab("Count") +  
 ggtitle(paste(country\_code, "Contributions (with collaboration)"))  
   
 # Return list of plots  
 return(list(hist\_100, hist\_not\_0))  
}  
  
# Loop over country codes and generate plots  
for (country in countries) {  
 plots <- make\_histogram(country)  
   
 # Print the plots  
 print(plots[[1]])  
 print(plots[[2]])  
}



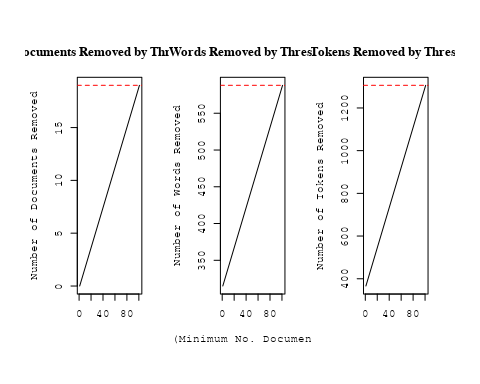
# Filter data for the publication interval 1824 to 1899  
data\_filtered <- data[data[["pub\_interval\_1824\_1899"]] == 1, ]  
  
# Save the original title data for future use  
data\_filtered$original\_concatenated\_title\_abstract <- data\_filtered$concatenated\_title\_abstract  
  
#pre-processing the titles using textProcessor from the stm package  
processed\_text <- textProcessor(data\_filtered$concatenated\_title\_abstract, metadata = data\_filtered)

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

# Further prepare the data by removing low-frequency terms  
out\_text <- prepDocuments(processed\_text$documents, processed\_text$vocab, processed\_text$meta)

## Removing 315 of 588 terms (315 of 1038 tokens) due to frequency   
## Your corpus now has 19 documents, 273 terms and 723 tokens.

docs\_text <- out\_text$documents  
vocab\_text <- out\_text$vocab  
meta\_text <- out\_text$meta  
  
  
#Prepare data  
plotRemoved(processed\_text$documents, lower.thresh = seq(1, 200, by = 100))



out\_text <- prepDocuments(processed\_text$documents, processed\_text$vocab, processed\_text$meta, lower.thresh = 2)

## Removing 497 of 588 terms (679 of 1038 tokens) due to frequency   
## Your corpus now has 19 documents, 91 terms and 359 tokens.

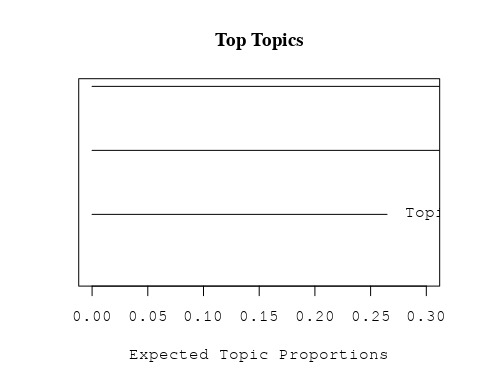
# Initialize an empty formula string  
prevalence\_formula\_str <- "~US "  
  
  
# Add each country variable to the formula string  
for (i in 10:27) {  
 prevalence\_formula\_str <- paste(prevalence\_formula\_str, "+", names(data)[i])  
}  
  
  
# Convert the string to a formula  
prevalence\_formula <- as.formula(prevalence\_formula\_str)  
print(prevalence\_formula)

## ~US + IN + DE + CH + GB + CN + FR + IT + RU + CA + NL + AU +   
## JP + ES + IL + Americas + Europe + Africa + AsiaAndOceania

# Run STM model  
Research\_topics <- stm(documents = out\_text$documents,   
 vocab = out\_text$vocab,   
 K = 3,   
 prevalence = prevalence\_formula,   
 data = out\_text$meta,   
 init.type = "Spectral",  
 max.em.its = 1000,  
 gamma.prior = 'L1')

## Beginning Spectral Initialization   
## Calculating the gram matrix...  
## Finding anchor words...  
## ...  
## Recovering initialization...  
##   
## Initialization complete.  
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 1 (approx. per word bound = -4.299)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 2 (approx. per word bound = -4.081, relative change = 5.064e-02)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 3 (approx. per word bound = -4.048, relative change = 8.128e-03)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 4 (approx. per word bound = -4.035, relative change = 3.224e-03)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 5 (approx. per word bound = -4.012, relative change = 5.637e-03)   
## Topic 1: result, proof, line, present, paper   
## Topic 2: given, star, may, mean, method   
## Topic 3: observ, measur, three, radiat, account   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 6 (approx. per word bound = -3.984, relative change = 6.988e-03)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 7 (approx. per word bound = -3.974, relative change = 2.476e-03)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 8 (approx. per word bound = -3.969, relative change = 1.233e-03)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 9 (approx. per word bound = -3.966, relative change = 8.329e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 10 (approx. per word bound = -3.964, relative change = 6.074e-04)   
## Topic 1: result, proof, star, present, upon   
## Topic 2: given, method, may, mean, power   
## Topic 3: observ, measur, three, line, radiat   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 11 (approx. per word bound = -3.962, relative change = 4.419e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 12 (approx. per word bound = -3.961, relative change = 3.190e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 13 (approx. per word bound = -3.960, relative change = 2.359e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 14 (approx. per word bound = -3.959, relative change = 1.868e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 15 (approx. per word bound = -3.958, relative change = 1.673e-04)   
## Topic 1: result, proof, star, present, upon   
## Topic 2: given, method, may, mean, power   
## Topic 3: observ, measur, three, line, radiat   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 16 (approx. per word bound = -3.958, relative change = 1.845e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 17 (approx. per word bound = -3.957, relative change = 2.598e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 18 (approx. per word bound = -3.955, relative change = 4.028e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 19 (approx. per word bound = -3.953, relative change = 5.001e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 20 (approx. per word bound = -3.951, relative change = 4.727e-04)   
## Topic 1: result, proof, star, present, upon   
## Topic 2: given, method, may, mean, power   
## Topic 3: observ, measur, three, line, radiat   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 21 (approx. per word bound = -3.950, relative change = 4.047e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 22 (approx. per word bound = -3.948, relative change = 3.184e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 23 (approx. per word bound = -3.947, relative change = 2.181e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 24 (approx. per word bound = -3.947, relative change = 1.613e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 25 (approx. per word bound = -3.946, relative change = 1.628e-04)   
## Topic 1: result, proof, star, present, upon   
## Topic 2: given, method, mean, may, power   
## Topic 3: observ, measur, three, line, radiat   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 26 (approx. per word bound = -3.945, relative change = 1.772e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 27 (approx. per word bound = -3.945, relative change = 1.470e-04)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 28 (approx. per word bound = -3.944, relative change = 9.417e-05)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 29 (approx. per word bound = -3.944, relative change = 5.801e-05)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 30 (approx. per word bound = -3.944, relative change = 3.849e-05)   
## Topic 1: result, proof, star, present, upon   
## Topic 2: given, method, mean, may, power   
## Topic 3: observ, measur, three, line, account   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 31 (approx. per word bound = -3.944, relative change = 2.845e-05)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 32 (approx. per word bound = -3.944, relative change = 2.235e-05)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 33 (approx. per word bound = -3.944, relative change = 1.822e-05)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 34 (approx. per word bound = -3.944, relative change = 1.491e-05)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 35 (approx. per word bound = -3.944, relative change = 1.265e-05)   
## Topic 1: result, proof, present, star, upon   
## Topic 2: given, method, mean, may, power   
## Topic 3: observ, measur, three, line, account   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Completing Iteration 36 (approx. per word bound = -3.944, relative change = 1.017e-05)   
## ...................  
## Completed E-Step (0 seconds).   
## Completed M-Step.   
## Model Converged

# Plot the STM model summary  
plot(Research\_topics, type = "summary", xlim = c(0, 0.3))



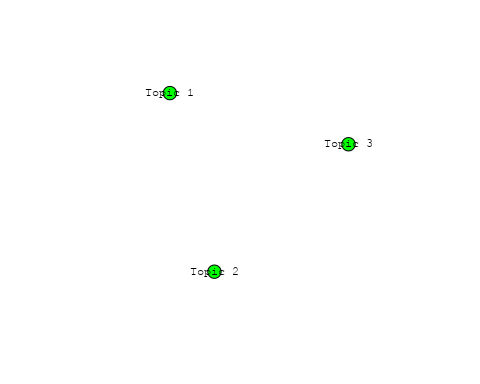
# Print the top 10 labels for each topic  
topic\_labels <- labelTopics(Research\_topics, n=10)  
print(topic\_labels)

## Topic 1 Top Words:  
## Highest Prob: result, proof, present, star, upon, line, investig, materi, probabl, one   
## FREX: proof, present, investig, materi, probabl, result, upon, attempt, caus, heat   
## Lift: attempt, caus, heat, investig, materi, number, presenc, probabl, can, gase   
## Score: show, proof, probabl, present, materi, caus, investig, gase, read, suppli   
## Topic 2 Top Words:  
## Highest Prob: given, method, mean, may, power, theori, star, obtain, fix, eclips   
## FREX: method, given, eclips, power, may, theori, mean, follow, unworthi, lunar   
## Lift: eclips, follow, unworthi, make, similar, lunar, method, power, given, correspond   
## Score: eclips, follow, unworthi, mean, method, may, power, lunar, theori, similar   
## Topic 3 Top Words:  
## Highest Prob: observ, measur, three, line, account, radiat, inch, made, star, paper   
## FREX: measur, three, account, radiat, observ, inch, line, made, cell, forc   
## Lift: account, measur, radiat, cell, forc, stellar, three, vol, inch, solar   
## Score: cell, measur, radiat, vol, three, stellar, forc, account, royal, inch

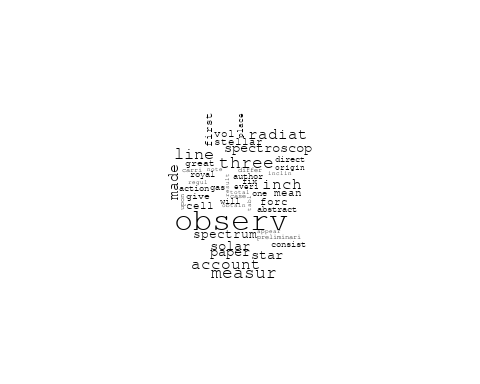
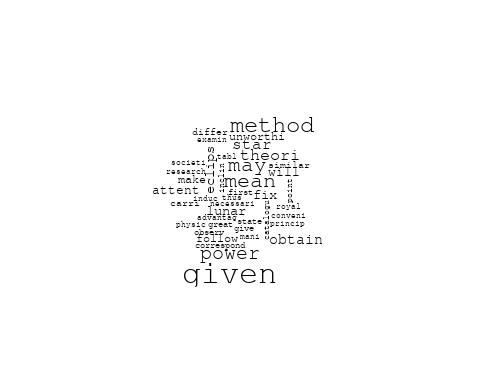
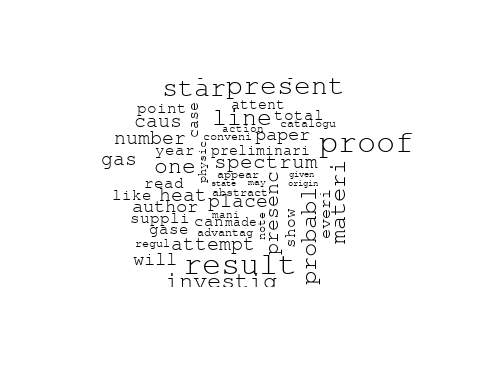
# Match the processed documents with the original titles  
matched\_titles <- out\_text$meta$original\_concatenated\_title\_abstract  
  
# Print top 5 documents for each topic  
top\_docs <- findThoughts(Research\_topics, texts = matched\_titles, n = 5)$docs[[1]]  
print(top\_docs)

## [1] "On the Physical Constitution of the Sun and Stars. An attempt is made in the memoir of which this is an abstract to take advantage of the insight we have gained within the last few years into the molecular constitution of gases, and the laws which regulate the exchanges of heat that take place between bodies placed in presence of one another, and to apply these new materials to the interpretation of the phenomena of the photosphere of the sun, the appearances presented during total eclipses, and the information about both sun and stars given by the spectroscope. In an inquiry like this, where we are obliged to put up with such proofs as the materials at our disposal can supply, we must be content to accept results of every variety of probability, from that degree, bordering upon certainty, which commands an unhesitating assent, to that of which the chief scientific value is that it prompts to further investigation and points out a path. Those who read the memoir itself will best judge of the probability of each conclusion from the proofs laid before them ; but in this sketch of its contents it may not be useless to indicate what is the value put upon each result by the author, since the proofs must in many cases be entirely omitted. It will be convenient to do this by numbers."  
## [2] "On the Physical Constitution of the Sun and Stars. An attempt is made in the memoir of which this is an abstract to take advantage of the insight we have gained within the last few years into the molecular constitution of gases, and the laws which regulate the exchanges of heat that take place between bodies placed in presence of one another, and to apply these new materials to the interpretation of the phenomena of the photosphere of the sun, the appearances presented during total eclipses, and the information about both sun and stars given by the spectroscope. In an inquiry like this, where we are obliged to put up with such proofs as the materials at our disposal can supply, we must be content to accept results of every variety of probability, from that degree, bordering upon certainty, which commands an unhesitating assent, to that of which the chief scientific value is that it prompts to further investigation and points out a path. Those who read the memoir itself will best judge of the probability of each conclusion from the proofs laid before them ; but in this sketch of its contents it may not be useless to indicate what is the value put upon each result by the author, since the proofs must in many cases be entirely omitted. It will be convenient to do this by numbers."  
## [3] "The Cause of Luminosity in the Flames of Hydrocarbon Gases In a paper read before the Chemical Society in 1893, I showed that in the inner non-luminous zone of a flame of ordinary illuminating gas, the hydrocarbons originally present in the gas, and consisting of ethylene, butylene, benzene, methane, and ethane, became converted by the baking action of the walls of flame between which they had to pass into acetylene, and that at the moment when luminosity commenced, over 80 per cent. of the total unsaturated hydrocarbons present consisted of this compound. The presence of acetylene at the point where luminosity commenced naturally suggested that it was in some way due to actions in which the acetylene played the principal part—either that it split up into carbon and hydrogen under the influence of heat, and so supplied the flame with the solid particles necessary, according to Sir Humphry Davy’s theory of the cause of luminosity, or else by its polymerisation it formed the dense vapours required by Dr. E. Frankland’s more recent hypothesis."   
## [4] "Note on the Unknown Chromospheric Substance of Young In the preliminary catalogue of the bright lines in the spectrum of the chromosphere published by Young in 1861, he calls special attention to the lines numbered 1 and 82 in the catalogue, remarking that “they are very persistently present, though faint, and can be distinctly seen in the spectroscope to belong to the chromosphere, as such, not being due, like most of the other lines, to the exceptional elevation of matter to heights where it does not properly belong. It would seem very probable that both these lines are due to the same substance which causes the D3 line.”"   
## [5] "I. On a gas showing the spectrum of helium, the reputed cause of D3, one of the lines in the coronal spectrum. Preliminary note In the course of investigations on argon, some clue was sought for, which would lead to the selection of one out of the almost innumerable compounds with which chemists are acquainted, with which to attempt to induce argon to combine. A paper by W. F. Hillebrand, “On the Occurrence of Nitrogen in Uraninite, &c.” (‘Bull, of the U. S. Geological Survey,’ No. 78, p. 43), to which M r. Miers kindly directed my attention, gave the desired clue. In spite of Hillebrand’s positive proof that the gas he obtained by boiling various samples of uraninite with weak sulphuric acid was nitrogen (p. 55)— such as formation of ammonia on sparking with hydrogen, analysis of the platinichloride, vacuum-tube spectrum , &c.— I was sceptical enough to doubt that any compound of nitrogen, when boiled with acid, would yield free nitrogen. The result has justified the scepticism."

# Calculate and plot the correlation between topics  
mod.out.corr <- topicCorr(Research\_topics)  
plot(mod.out.corr, cex = 1.5)



# For each topic  
for (topic\_num in 1:3) {  
 # Plot the word cloud  
 cloud(Research\_topics, topic = topic\_num, scale = c(2, 0.25))  
 Sys.sleep(2)  
}



# Get the topic proportions for each document  
topic\_proportions <- Research\_topics$theta  
  
# Find the index of the topic with the highest proportion for each document  
# This will be the topic that each document is most likely to belong to  
max\_topic\_idx <- apply(topic\_proportions, 1, which.max)  
  
# Add this as a new column to your data  
data\_filtered$topic <- max\_topic\_idx  
  
  
# Define the countries  
countries <- c("US", "GB", "DE", "IT", "AU", "CA", "JP", "CN", "NL", "FR", "RU", "ES", "IN", "IL", "CH", "Americas", "Europe", "Africa", "AsiaAndOceania")  
  
  
  
# Create function to generate histogram  
make\_histogram <- function(country\_code) {  
   
 # Filter data for country with contribution of 100  
 data\_100 <- data\_filtered[data\_filtered[[country\_code]] == 100,]  
   
 # Filter data for country with contribution not equal to 0  
 data\_not\_0 <- data\_filtered[data\_filtered[[country\_code]] != 0,]  
   
 # Make the first histogram  
 hist\_100 <- ggplot(data\_100, aes(x = topic)) +  
 geom\_histogram(binwidth = 1, fill = "blue", color = "black") +  
 xlab("Topic") +  
 ylab("Count") +  
 ggtitle(paste(country\_code, "Independent Contribution"))  
   
 # Make the second histogram  
 hist\_not\_0 <- ggplot(data\_not\_0, aes(x = topic)) +  
 geom\_histogram(binwidth = 1, fill = "blue", color = "black") +  
 xlab("Topic") +  
 ylab("Count") +  
 ggtitle(paste(country\_code, "Contributions (with collaboration)"))  
   
 # Return list of plots  
 return(list(hist\_100, hist\_not\_0))  
}  
  
# Loop over country codes and generate plots  
for (country in countries) {  
 plots <- make\_histogram(country)  
   
 # Print the plots  
 print(plots[[1]])  
 print(plots[[2]])  
}

