Text Analysis- Project 5

Jack Fitzgibbons

Load necessary libraries

library(tm)

## Warning: package 'tm' was built under R version 4.1.3

## Loading required package: NLP

library(XML)

Read data file from website - Emancipation proclamation

emp <- URLencode('http://www.historyplace.com/lincoln/emanc.htm') # emancipation proclamation  
junk <- htmlTreeParse(emp, useInternal=TRUE)  
  
# doc.html # un-comment to see html file

Remove HTML code from downloaded file:

abe <- unlist(xpathApply(junk,'//p',xmlValue))  
head(abe)

## [1] ""   
## [2] "Emancipation Proclamation\n"   
## [3] "BY THE PRESIDENT OF THE UNITED STATES OF AMERICA:\n A PROCLAMATION.\n"   
## [4] "Whereas, on the twentysecond day of September, in the year of our Lord\none thousand eight hundred and sixty two, a proclamation was issued by\nthe President of the United States, containing, among other things, the\nfollowing, towit: "   
## [5] "\"That on the first day of January, in the year of our Lord one\nthousand eight hundred and sixty-three, all persons held as slaves within\nany State or designated part of a State, the people whereof shall then\nbe in rebellion against the United States, shall be then, thenceforward,\nand forever free; and the Executive Government of the United States, including\nthe military and naval authority thereof, will recognize and maintain the\nfreedom of such persons, and will do no act or acts to repress such persons,\nor any of them, in any efforts they may make for their actual freedom.\n"   
## [6] "\"That the Executive will, on the first day of January aforesaid,\nby proclamation, designate the States and parts of States, if any, in which\nthe people thereof, respectively, shall then be in rebellion against the\nUnited States; and the fact that any State, or the people thereof, shall\non that day be, in good faith, represented in the Congress of the United\nStates by members chosen thereto at elections wherein a majority of the\nqualified voters of such State shall have participated, shall, in the absence\nof strong countervailing testimony, be deemed conclusive evidence that\nsuch State, and the people thereof, are not then in rebellion against the\nUnited States.\" "

Create corpus. Note: usually a corpus is comprised of a large collection of documents. Here we analyze a single document (broken into 16 parts by “Corpus”).

words.vec <- VectorSource(abe)  
words.corpus <- Corpus(words.vec)   
  
str(abe)

## chr [1:18] "" "Emancipation Proclamation\n" ...

str(words.vec)

## Classes 'VectorSource', 'SimpleSource', 'Source' hidden list of 5  
## $ encoding: chr ""  
## $ length : int 18  
## $ position: num 0  
## $ reader :function (elem, language, id)   
## $ content : chr [1:18] "" "Emancipation Proclamation\n" "BY THE PRESIDENT OF THE UNITED STATES OF AMERICA:\n A PROCLAMATION.\n" "Whereas, on the twentysecond day of September, in the year of our Lord\none thousand eight hundred and sixty tw"| \_\_truncated\_\_ ...

words.corpus

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 18

Do some data wrangling:

words.corpus <- tm\_map(words.corpus,content\_transformer(tolower))

## Warning in tm\_map.SimpleCorpus(words.corpus, content\_transformer(tolower)):  
## transformation drops documents

words.corpus <- tm\_map(words.corpus,content\_transformer(removePunctuation))

## Warning in tm\_map.SimpleCorpus(words.corpus,  
## content\_transformer(removePunctuation)): transformation drops documents

words.corpus <- tm\_map(words.corpus,content\_transformer(removeNumbers))

## Warning in tm\_map.SimpleCorpus(words.corpus,  
## content\_transformer(removeNumbers)): transformation drops documents

words.corpus <- tm\_map(words.corpus,removeWords,stopwords('english'))

## Warning in tm\_map.SimpleCorpus(words.corpus, removeWords, stopwords("english")):  
## transformation drops documents

words.corpus

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 18

Create a term-document matrix:

tdm <- TermDocumentMatrix(words.corpus)  
tdm

## <<TermDocumentMatrix (terms: 223, documents: 18)>>  
## Non-/sparse entries: 345/3669  
## Sparsity : 91%  
## Maximal term length: 16  
## Weighting : term frequency (tf)

inspect(tdm)

## <<TermDocumentMatrix (terms: 223, documents: 18)>>  
## Non-/sparse entries: 345/3669  
## Sparsity : 91%  
## Maximal term length: 16  
## Weighting : term frequency (tf)  
## Sample :  
## Docs  
## Terms 10 11 12 14 4 5 6 7 8 9  
## day 0 0 0 1 1 1 2 3 0 0  
## one 0 0 0 1 1 1 0 2 0 0  
## people 1 0 0 0 0 1 3 1 0 0  
## persons 0 1 0 0 0 3 0 0 0 2  
## rebellion 0 0 0 0 0 1 2 3 0 0  
## shall 0 0 0 0 0 2 4 0 0 1  
## state 0 0 0 0 0 2 3 0 0 0  
## states 0 1 0 1 1 2 5 6 0 3  
## thereof 0 0 0 0 0 1 3 1 0 1  
## united 0 1 0 1 1 2 3 4 0 1

Examine a simple term-document matrix representation by converting to a matrix. (Note: this transformation removes the sparse representation found in the original TDM and therefore may potentially use a large amount of memory!)

m <- as.matrix(tdm); m

## Docs  
## Terms 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18  
## emancipation 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1  
## proclamation 0 1 1 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0  
## america 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## president 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 1 0 0  
## states 0 0 1 1 2 5 6 0 3 0 1 0 1 1 0 0 0 0  
## united 0 0 1 1 2 3 4 0 1 0 1 0 1 1 0 0 0 0  
## among 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## containing 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## day 0 0 0 1 1 2 3 0 0 0 0 0 0 1 0 0 0 0  
## eight 0 0 0 1 1 0 1 0 0 0 0 0 0 1 0 0 0 0  
## following 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## hundred 0 0 0 1 1 0 2 0 0 0 0 0 0 1 0 0 0 0  
## issued 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## lord 0 0 0 1 1 0 1 0 0 0 0 0 0 1 0 0 0 0  
## one 0 0 0 1 1 0 2 0 0 0 0 0 0 1 0 0 0 1  
## september 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## sixty 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0  
## things 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## thousand 0 0 0 1 1 0 1 0 0 0 0 0 0 1 0 0 0 0  
## towit 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## twentysecond 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## two 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## whereas 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
## year 0 0 0 1 1 0 1 0 0 0 0 0 0 1 0 0 0 0  
## act 0 0 0 0 1 0 0 0 0 0 0 2 0 0 0 0 0 0  
## acts 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## actual 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0  
## authority 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0  
## designated 0 0 0 0 1 0 0 1 1 0 0 0 0 0 0 0 0 0  
## efforts 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## executive 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0  
## first 0 0 0 0 1 1 2 0 0 0 0 0 0 1 0 0 0 0  
## forever 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## free 0 0 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0  
## freedom 0 0 0 0 2 0 0 0 1 0 0 0 0 0 0 0 0 0  
## government 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0  
## held 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0  
## including 0 0 0 0 1 0 0 2 1 0 0 0 0 0 0 0 0 0  
## january 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 0 0  
## maintain 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0  
## make 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0  
## may 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## military 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 0  
## naval 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0  
## part 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## people 0 0 0 0 1 3 1 0 0 1 0 0 0 0 0 0 0 0  
## persons 0 0 0 0 3 0 0 0 2 0 1 0 0 0 0 0 0 0  
## rebellion 0 0 0 0 1 2 3 0 0 0 0 0 0 0 0 0 0 0  
## recognize 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0  
## repress 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## shall 0 0 0 0 2 4 0 0 1 0 0 0 0 0 0 0 0 0  
## sixtythree 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## slaves 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0  
## state 0 0 0 0 2 3 0 0 0 0 0 0 0 0 0 1 0 0  
## thenceforward 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
## thereof 0 0 0 0 1 3 1 0 1 0 0 0 0 0 0 0 0 0  
## whereof 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0  
## will 0 0 0 0 2 1 0 0 1 0 1 0 0 0 0 0 0 0  
## within 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0  
## absence 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## aforesaid 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0  
## chosen 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## conclusive 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## congress 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## countervailing 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## deemed 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## designate 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0  
## elections 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## evidence 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## fact 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## faith 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## good 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## majority 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## members 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## participated 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## parts 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0  
## qualified 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## represented 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## respectively 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0  
## strong 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## testimony 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## thereto 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## voters 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
## wherein 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0  
## abraham 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0  
## accordance 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## armed 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0  
## army 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## commanderinchief 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## days 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## fit 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## full 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## lincoln 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 1  
## measure 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## mentioned 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## navy 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## necessary 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0  
## now 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## order 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0  
## period 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## power 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0  
## proclaimed 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## publicly 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## purpose 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0  
## said 0 0 0 0 0 0 1 0 2 0 1 0 0 0 0 0 0 0  
## suppressing 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## therefore 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## three 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0  
## time 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## vested 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## virtue 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0  
## war 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0  
## accomac 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## alabama 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## also 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## ann 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## arkansas 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## ascension 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## assumption 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## berkley 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## bernard 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## charles 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## cities 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## city 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0  
## counties 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0  
## elizabethcity 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## except 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0  
## excepted 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## florida 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## fortyeight 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## georgia 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## james 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## jefferson 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## johns 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## lafourche 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## left 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## louisiana 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## martin 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## mary 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## mississippi 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## neworleans 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## norfolk 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0  
## northampton 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## northcarolina 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## orleans 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## parishes 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## plaquemines 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## portsmouth 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## precisely 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## present 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## princess 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## southcarolina 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## terrebonne 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## texas 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## virginia 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0  
## west 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## york 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0  
## authorities 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0  
## declare 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0  
## henceforward 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0  
## abstain 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## allowed 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## cases 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## declared 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## enjoin 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## faithfully 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## hereby 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## labor 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## reasonable 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## recommend 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## selfdefense 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## unless 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## upon 0 0 0 0 0 0 0 0 0 1 0 2 0 0 0 0 0 0  
## violence 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## wages 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
## condition 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## forts 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## garrison 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## known 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## man 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## places 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## positions 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## received 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## service 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0  
## sorts 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## stations 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## suitable 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## vessels 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0  
## almighty 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## believed 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## considerate 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## constitution 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## favor 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## god 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## gracious 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## invoke 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## judgment 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## justice 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## mankind 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## necessity 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## sincerely 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## warranted 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0  
## affixed 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## caused 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## hand 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## hereunto 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## seal 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## set 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## witness 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0  
## done 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## eightyseventh 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## independence 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## washington 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0  
## secretary 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0  
## seward 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0  
## william 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0  
## documentreturn 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1  
## history 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2  
## main 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1  
## page 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2  
## place 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2  
## see 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1  
## timelinethe 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1

wordCounts <- rowSums(m)   
myNames <- names(wordCounts)   
  
checkWordCounts <- sort(wordCounts,decreasing=TRUE)  
head(checkWordCounts)

## states united day shall one people   
## 21 15 8 7 6 6

These were the results of taking the TDM and converting it to a matrix. As you can see, “states”, “united”, “day”, “shall”, “one”, and “people” were the most common word. This is interesting since the overall message of the Emancipation Proclamation is essentially that the “United States shall one day be one people”.

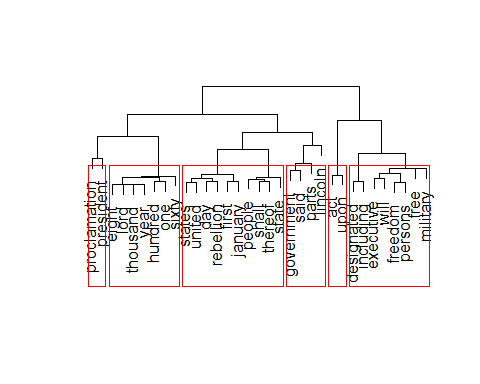
Perform Cluster Analysis

library(cluster); library(lsa)

## Warning: package 'lsa' was built under R version 4.1.3

## Loading required package: SnowballC

m1 <- t(m)  
m2 <- m1[,colSums(m1) > 2]  
  
mycos <- as.dist(1-cosine(m2)) # Cosine Distance Matrix Between Terms  
  
agnes.out <- agnes(mycos, method ="ward",stand = TRUE ) # Agglom . Clust .  
  
pltree(agnes.out, main =" ", ylab =" ", xlab =" ", yaxt ="n", sub=" ") # Create the Dendogram and   
rect.hclust(agnes.out, k=6 , border ="red") # draw boxes around groups



Using cluster analysis to measure the distance between words by looking at the vector associated with each word shows the above results. For m2, I chose to look at column sums for my matrix with a word at occurred in at least 2 of my documents. I found that choosing 1 here resulted in too much clutter in the final chart. Looking at the first cluster, president and proclamation occur close to each other twice as the president of the United States issued the proclamation. For example in the first sentence, “a proclamation was issued by the president of the Untied States”. The second and third cluster also make sense, as the date is mentioned three times, both the date Abraham Lincoln wrote the statement, and the date the proclamation goes into effect (January 1, 1863) The fourth cluster makes sense as well since Lincoln addresses the “parts of States” that this is applicable to, as well as the “said” poeple Lincoln is referring to. Lastly, the fifth and sixth cluster address the message Lincoln is stating. It is an “executive” order for those “designated persons”, and he declares their freedom and how they are allowed to be included in any facet of the military.

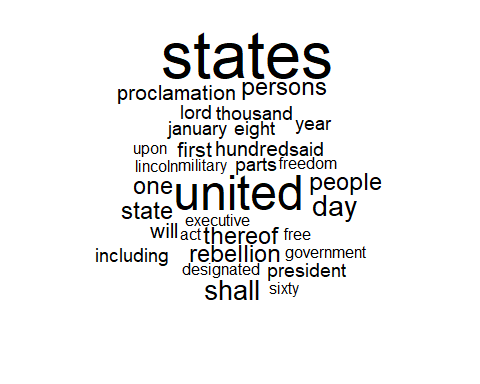
Create simple word cloud using (Use “words” / “word counts” to create data frame for word clouds (cloudFrame))

library(wordcloud)

## Warning: package 'wordcloud' was built under R version 4.1.3

## Loading required package: RColorBrewer

cloudFrame <- data.frame(word = myNames, freq=wordCounts)  
wordcloud(cloudFrame$word,cloudFrame$freq)



Create wordcloud with arbitrary colors:

wordcloud(cloudFrame$word, cloudFrame$freq, min.freq=2, max.words=50, rot.per=0.35, colors=brewer.pal(8,'Dark2'))



Examining the two word clouds, one being simple and the other being with arbitrary colors, we can see United States is obviously the most frequent. Following the United States, “day” seems to be the next biggest, which is in line with our tdm results above.The simple cloud is much harder to identify the next most common words. However, with the arbitrary colored cloud we see that those words in blue/purple come after “day”, followed by dark orange, and then orange and green.

Determine if emancipation proclamation speech was generally positive or negative:

Note: posterms\_GI = Postitive words from General Inquirer negterms\_GI = Negative words from General Inquirer

library(devtools)

## Warning: package 'devtools' was built under R version 4.1.3

## Loading required package: usethis

## Warning: package 'usethis' was built under R version 4.1.3

#install\_github("mannau/tm.plugin.sentiment")  
#install.packages("tm.plugin.sentiment", repos = "http://R-Forge.R-project.org")  
library(tm.plugin.sentiment)

## Loading required package: quantmod

## Warning: package 'quantmod' was built under R version 4.1.3

## Loading required package: xts

## Warning: package 'xts' was built under R version 4.1.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.1.3

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: TTR

## Warning: package 'TTR' was built under R version 4.1.3

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

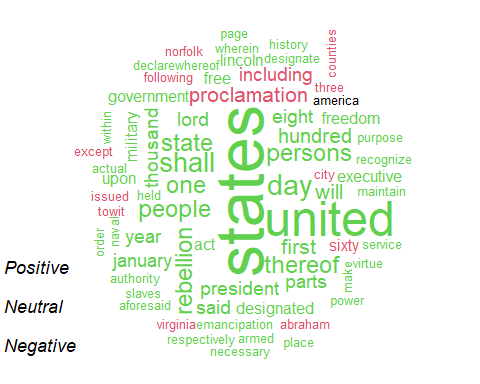
# tm\_tag\_score <- tm\_term\_score  
my.polarity <- polarity(tdm, positive = posterms\_GI(), negative = negterms\_GI())  
mean(my.polarity)

## [1] 0.2802469

According to the tm package, the words in this tdm for the Emancipation Proclamation was generally positive at a rate of .28. I did not add any positive or negative words, but just used the original positive and negative sentiment vectors.

Create sentiment colored word cloud.

my.sentiments <- my.polarity %\*% t(m) # agg over documents  
my.colors <- rep(99,length(my.sentiments))  
  
my.colors[my.sentiments < 0] <- 2 # Negative = 2 = red  
my.colors[my.sentiments == 0] <- 1 # Neutral = 1 = black  
my.colors[my.sentiments > 0] <- 3 # Positive = 3 = green  
  
sentDat <- data.frame(myNames = myNames, my.colors = my.colors, v=wordCounts)  
  
wordcloud(sentDat$myNames ,freq = sentDat$v, colors = sentDat$my.colors, random.order = FALSE,   
 ordered.colors = TRUE , scale = c(4,.5) , min.freq=2)  
  
legend(-.4, .4, c("Positive","Neutral","Negative"), col = c(3 ,1 ,2), lty = 1, ncol = 1, lwd = 5, cex = 1.1 , bty ="n", text.font = 3)



Looking at the results for the sentiment word cloud, “United” and “States” are again clearly the most frequent, and are both positive (green). As for the negative words in red, “proclamation” and “including” stand out as the largest. Perhaps most proclamations result in negative reactions, which is why it is most often viewed as negative, but in this case, proclamation should generally be positive. “America” can be seen as the only term that is neutral, which is interesting. Other negative terms such as “Virginia” makes sense in this case since Abraham Lincoln states that some states, including Virginia, are in rebellion against the United States. “Abraham” being viewed as negative could be true for some people in regards to the time period, but other than that, doesn’t really have any reason to be negative in my opinion. Also, “rebellion” being positive is interesting as I would typically view that in a negative light. Overall, an interesting outcome in regards to sentiment, but there could definitely be an argument for swapping some terms.