Project 5

Kelly Markor

5/14/2020

Load necessary libraries

library(tm)

## Loading required package: NLP

library(XML)

Read data file

sbaLoc <- URLencode('http://www.historyplace.com/lincoln/emanc.htm')  
doc.html <-htmlTreeParse(sbaLoc, useInternal=TRUE)  
doc.html

## <!DOCTYPE html>  
## <html>  
## <head>  
## <meta charset="UTF-8">  
## <title>The History Place - Abraham Lincoln: Emancipation Proclamation</title>  
## <meta name="Description" content="Emancipation Proclamation - Part of our Lincoln Timeline.">  
## <meta name="KeyWords" content="emancipation proclamation, abraham lincoln, lincoln, civil war, history, history place">  
## <!-- Global site tag (gtag.js) - Google Analytics --><script async src="https://www.googletagmanager.com/gtag/js?id=UA-29013255-1"></script><script>  
## window.dataLayer = window.dataLayer || [];  
## function gtag(){dataLayer.push(arguments);}  
## gtag('js', new Date());  
##   
## gtag('config', 'UA-29013255-1');  
## </script>  
## </head>  
## <body text="#000000" bgcolor="#333333" link="#0000EE" vlink="#FF0000" alink="#0000FF">  
## <table width="975" border="1" align="center" cellpadding="10" bordercolor="#999999" bgcolor="#FFFFFF"><tr>  
## <td valign="top">  
## <center><img src="thplinc.gif" alt="The History Place - Abraham Lincoln" width="356" height="77"></center>  
##   
## <center>  
## <p><img src="thin-blue-365.jpg" width="365" height="1"></p>  
## </center>  
##   
## <center>  
## <p><i><font color="#0000FF" size="+2">Emancipation Proclamation</font></i>  
## </p>  
## </center>  
##   
## <center><p>BY THE PRESIDENT OF THE UNITED STATES OF AMERICA:<br>  
## A PROCLAMATION.  
## </p></center>  
##   
## <blockquote><blockquote>  
## <p>Whereas, on the twentysecond day of September, in the year of our Lord  
## one thousand eight hundred and sixty two, a proclamation was issued by  
## the President of the United States, containing, among other things, the  
## following, towit: </p>  
##   
## <p>"That on the first day of January, in the year of our Lord one  
## thousand eight hundred and sixty-three, all persons held as slaves within  
## any State or designated part of a State, the people whereof shall then  
## be in rebellion against the United States, shall be then, thenceforward,  
## and forever free; and the Executive Government of the United States, including  
## the military and naval authority thereof, will recognize and maintain the  
## freedom of such persons, and will do no act or acts to repress such persons,  
## or any of them, in any efforts they may make for their actual freedom.  
## </p>  
##   
## <p>"That the Executive will, on the first day of January aforesaid,  
## by proclamation, designate the States and parts of States, if any, in which  
## the people thereof, respectively, shall then be in rebellion against the  
## United States; and the fact that any State, or the people thereof, shall  
## on that day be, in good faith, represented in the Congress of the United  
## States by members chosen thereto at elections wherein a majority of the  
## qualified voters of such State shall have participated, shall, in the absence  
## of strong countervailing testimony, be deemed conclusive evidence that  
## such State, and the people thereof, are not then in rebellion against the  
## United States." </p>  
##   
## <p>Now, therefore I, Abraham Lincoln, President of the United States, by  
## virtue of the power in me vested as Commander-in-Chief, of the Army and  
## Navy of the United States in time of actual armed rebellion against authority  
## and government of the United States, and as a fit and necessary war measure  
## for suppressing said rebellion, do, on this first day of January, in the  
## year of our Lord one thousand eight hundred and sixty three, and in accordance  
## with my purpose so to do publicly proclaimed for the full period of one  
## hundred days, from the day first above mentioned, order and designate as  
## the States and parts of States wherein the people thereof respectively,  
## are this day in rebellion against the United States, the following, towit:  
## </p>  
##   
## <p>Arkansas, Texas, Louisiana, (except the Parishes of St. Bernard, Plaquemines,  
## Jefferson, St. Johns, St. Charles, St. James, Ascension, Assumption, Terrebonne,  
## Lafourche, St. Mary, St. Martin, and Orleans, including the City of New-Orleans)  
## Mississippi, Alabama, Florida, Georgia, South-Carolina, North-Carolina,  
## and Virginia, (except the fortyeight counties designated as West Virginia,  
## and also the counties of Berkley, Accomac, Northampton, Elizabeth-City,  
## York, Princess Ann, and Norfolk, including the cities of Norfolk &amp;  
## Portsmouth); and which excepted parts are, for the present, left precisely  
## as if this proclamation were not issued. </p>  
##   
## <p>And by virtue of the power, and for the purpose aforesaid, I do order  
## and declare that all persons held as slaves within said designated States,  
## and parts of States, are, and henceforward shall be free; and that the  
## Executive government of the United States, including the military and naval  
## authorities thereof, will recognize and maintain the freedom of said persons.  
## </p>  
##   
## <p>And I hereby enjoin upon the people so declared to be free to abstain  
## from all violence, unless in necessary self-defense; and I recommend to  
## them that, in all cases when allowed, they labor faithfully for reasonable  
## wages. </p>  
##   
## <p>And I further declare and make known, that such persons of suitable  
## condition, will be received into the armed service of the United States  
## to garrison forts, positions, stations, and other places, and to man vessels  
## of all sorts in said service. </p>  
##   
## <p>And upon this act, sincerely believed to be an act of justice, warranted  
## by the Constitution, upon military necessity, I invoke the considerate  
## judgment of mankind, and the gracious favor of Almighty God. </p>  
##   
## <p>In witness whereof, I have hereunto set my hand and caused the seal  
## of the United States to be affixed. </p>  
##   
## <p>Done at the City of Washington, this first day of January, in the year  
## of our Lord one thousand eight hundred and sixty three, and of the Independence  
## of the United States of America the eighty-seventh. </p>  
##   
## <div align="right"><p>ABRAHAM LINCOLN </p></div>  
##   
## <p>By the President: <br>  
## WILLIAM H. SEWARD, Secretary of State. </p>  
##   
## </blockquote></blockquote>  
## <center>  
## <p><img src="thin-blue-785.jpg" width="785" height="1"></p>  
## </center>  
##   
## <p><a href="lincpix/emanc-proc.jpg">See Page One of the Emancipation Document</a><br><a href="index.html#emance">Return to The History Place - Lincoln Timeline</a><br><a href="../index.html">The  
## History Place Main Page</a></p>  
## </td>  
## </tr></table>  
## </body>  
## </html>  
##

Remove HTML code from downloaded file:

sba <- unlist(xpathApply(doc.html,'//p',xmlValue))  
head(sba)

## [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(sba)  
words.corpus <- Corpus(words.vec)   
  
str(sba)

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

str(words.vec)

## List of 18  
## $ encoding: chr ""  
## $ length : chr "Emancipation Proclamation\n"  
## $ position: chr "BY THE PRESIDENT OF THE UNITED STATES OF AMERICA:\n A PROCLAMATION.\n"  
## $ reader : chr "Whereas, on the twentysecond day of September, in the year of our Lord\none thousand eight hundred and sixty tw"| \_\_truncated\_\_  
## $ content : chr "\"That on the first day of January, in the year of our Lord one\nthousand eight hundred and sixty-three, all pe"| \_\_truncated\_\_  
## $ NA: chr "\"That the Executive will, on the first day of January aforesaid,\nby proclamation, designate the States and pa"| \_\_truncated\_\_  
## $ NA: chr "Now, therefore I, Abraham Lincoln, President of the United States, by\nvirtue of the power in me vested as Comm"| \_\_truncated\_\_  
## $ NA: chr "Arkansas, Texas, Louisiana, (except the Parishes of St. Bernard, Plaquemines,\nJefferson, St. Johns, St. Charle"| \_\_truncated\_\_  
## $ NA: chr "And by virtue of the power, and for the purpose aforesaid, I do order\nand declare that all persons held as sla"| \_\_truncated\_\_  
## $ NA: chr "And I hereby enjoin upon the people so declared to be free to abstain\nfrom all violence, unless in necessary s"| \_\_truncated\_\_  
## $ NA: chr "And I further declare and make known, that such persons of suitable\ncondition, will be received into the armed"| \_\_truncated\_\_  
## $ NA: chr "And upon this act, sincerely believed to be an act of justice, warranted\nby the Constitution, upon military ne"| \_\_truncated\_\_  
## $ NA: chr "In witness whereof, I have hereunto set my hand and caused the seal\nof the United States to be affixed. "  
## $ NA: chr "Done at the City of Washington, this first day of January, in the year\nof our Lord one thousand eight hundred "| \_\_truncated\_\_  
## $ NA: chr "ABRAHAM LINCOLN "  
## $ NA: chr "By the President: \nWILLIAM H. SEWARD, Secretary of State. "  
## $ NA: chr ""  
## $ NA: chr "See Page One of the Emancipation DocumentReturn to The History Place - Lincoln TimelineThe\nHistory Place Main Page"  
## - attr(\*, "class")= chr [1:3] "VectorSource" "SimpleSource" "Source"

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)

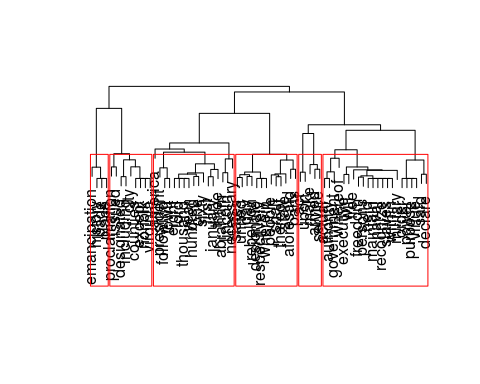
## states united day shall one people   
## 21 15 8 7 6 6

Perform Cluster Analysis

library(cluster); library(lsa)

## Loading required package: SnowballC

m1 <- t(m)  
m2 <- m1[,colSums(m1) > 1]  
  
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

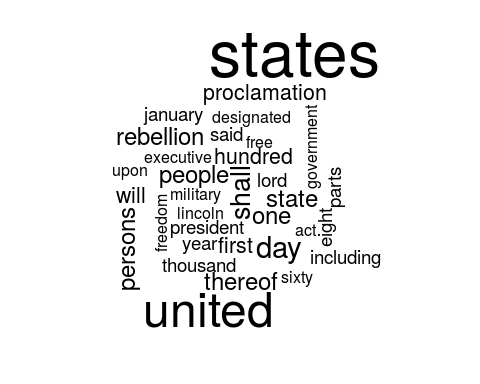


Create simple word cloud using (Use “words” / “word counts” to create data frame for word clouds (cloudFrame))

library(wordcloud)

## 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'))

 library(devtools) install.packages(“tm.plugin.sentiment”, repos = “<http://R-Forge.R-project.org>”) install\_github(“mannau/tm.plugin.sentiment”) library(tm.plugin.sentiment) my.polarity <- polarity(tdm, positive = posterms\_GI(), negative = negterms\_GI()) mean(my.polarity)

```

Create sentiment colored word cloud.

```{r}

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)

