Mining Text Data

Hazim Fitri

2025-01-05

Contents

Unstructured Data Analytic

NLP

- Computer procedure to interpret unstructured data to understand human language.

Objective :

- Identify trends, popular topics and themes relate to some particular issue

- Extract sentiment and people's emotions towards some particular issue

Text corpus

- Text need to be converted into corpus format for analysis
- Data mining analysis, hypothesis testing, checking occurrences, validating linguistic rules can be done on corpus format.

Data Cleaning

- Remove special character (/, @, |)
- Convert text to lower case
- Remove numbers
- Remove stopwords (the, is, at, on)
- Remove punctuation
- Eliminate extra unnecessary spaces

Word tokenization

Text stemming

Document-term matrix

Word cloud

Word association

Sentiment analysis

Emotion classification

- Built on the NRC Word-Emotion Association Lexicon
- 8 basic emotions :
    - Anger
    - Fear
    - Anticipation
    - Trust
    - Surprise
    - Sadness
    - Joy
    - Disgust

Steps of cleaning text data.

1. Remove special characters from the text, where a symbol such as; /, @ and | will be replace by a space.
2. Convert the text to lower case.
3. Remove numbers
4. Remove the stopwords in text data. Example: Stopwords in English are "the, is, at, on". There is no single universal list of stopwords used by all NLP tools.
5. Remove punctuation.
6. Eliminate extra unnecessary spaces in the text.

```r
library(tm)
```

```
## Loading required package: NLP
```

```r
library(SnowballC)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```r
library(RColorBrewer)
library(syuzhet)
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:NLP':
##
##     annotate
```

```r
# warning occurs due to the file does not end with newline
text = readLines('./Data/text.txt')
```

```
## Warning in readLines("./Data/text.txt"): incomplete final line found on
## './Data/text.txt'
```

```r
class(text)
```

```
## [1] "character"
```

```r
# after we add new line at the end, no warning will appear
readLines('./Data/newtext.txt')
```

```
##  [1] ""
##  [2] "And so even though we face the difficulties of today and tomorrow, I still have a dream. It is
##  [3] " "
##  [4] "I have a dream that one day this nation will rise up and live out the true meaning of its cree
##  [5] " "
##  [6] "We hold these truths to be self-evident, that all men are created equal."
##  [7] " "
##  [8] "I have a dream that one day on the red hills of Georgia, the sons of former slaves and the son
##  [9] " "
## [10] "I have a dream that one day even the state of Mississippi, a state sweltering with the heat of
## [11] " "
## [12] "I have a dream that my four little children will one day live in a nation where they will not
## [13] " "
## [14] "I have a dream today!"
## [15] " "
## [16] "I have a dream that one day, down in Alabama, with its vicious racists, with its governor havi
```

```
## [17] " "
## [18] "I have a dream today!"
## [19] " "
## [20] "I have a dream that one day every valley shall be exalted, and every hill and mountain shall be
## [21] " "
## [22] "This is our hope, and this is the faith that I go back to the South with."
## [23] " "
## [24] "With this faith, we will be able to hew out of the mountain of despair a stone of hope. With th
## [25] " "
## [26] "And this will be the day, this will be the day when all of God s children will be able to sing
## [27] " "
## [28] "My country  tis of thee, sweet land of liberty, of thee I sing."
## [29] "Land where my fathers died, land of the Pilgrim s pride,"
## [30] "From every mountainside, let freedom ring!"
## [31] "And if America is to be a great nation, this must become true."
## [32] "And so let freedom ring from the prodigious hilltops of New Hampshire."
## [33] "Let freedom ring from the mighty mountains of New York."
## [34] "Let freedom ring from the heightening Alleghenies of Pennsylvania."
## [35] "Let freedom ring from the snow-capped Rockies of Colorado."
## [36] "Let freedom ring from the curvaceous slopes of California."
## [37] " "
## [38] "But not only that:"
## [39] "Let freedom ring from Stone Mountain of Georgia."
## [40] "Let freedom ring from Lookout Mountain of Tennessee."
## [41] "Let freedom ring from every hill and molehill of Mississippi."
## [42] "From every mountainside, let freedom ring."
## [43] "And when this happens, when we allow freedom ring, when we let it ring from every village and
## [44] "Free at last! Free at last!"
## [45] " "
## [46] "Thank God Almighty, we are free at last!"
```

Transform into Corpus

```
docs = Corpus(VectorSource(text))
#inspect(docs)
```

```
class(docs)
```

```
## [1] "SimpleCorpus" "Corpus"
```

Replace special characters with space

Remove special characters

```
toSpace = content_transformer(function(x, pattern)
  gsub(pattern, '', x))
```

replace

```
docs2 = tm_map(docs, toSpace, '!')
```

```
## Warning in tm_map.SimpleCorpus(docs, toSpace, "!"): transformation drops
## documents
```

```
docs3 = tm_map(docs2, toSpace, ':')
```

```
## Warning in tm_map.SimpleCorpus(docs2, toSpace, ":"): transformation drops
## documents
```

```
docs4 = tm_map(docs3, toSpace, ',')
```

```
## Warning in tm_map.SimpleCorpus(docs3, toSpace, ","): transformation drops
## documents
```

Convert Capital letter to Small letter

```
docs5 = tm_map(docs4, content_transformer(tolower))
```

```
## Warning in tm_map.SimpleCorpus(docs4, content_transformer(tolower)):
## transformation drops documents
```

Remove number

```
docs6 = tm_map(docs5, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(docs5, removeNumbers): transformation drops
## documents
```

Remove stopwords (the, is, at, on)

```
docs7 = tm_map(docs6, removeWords, stopwords('english'))
```

```
## Warning in tm_map.SimpleCorpus(docs6, removeWords, stopwords("english")):
## transformation drops documents
```

Remove punctuation

```
docs8 = tm_map(docs7, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(docs7, removePunctuation): transformation drops
## documents
```

Remove extra spaces

```r
docs9 = tm_map(docs8, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(docs8, stripWhitespace): transformation drops
## documents
```

Text stemming

Find the root word

```r
docs10 = tm_map(docs9, stemDocument)
```

```
## Warning in tm_map.SimpleCorpus(docs9, stemDocument): transformation drops
## documents
```

Document term matrix

```r
dtm = TermDocumentMatrix(docs10)

m = as.matrix(dtm)

m
```

```
##               Docs
## Terms          1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
##   american     0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   deepli       0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   difficulti   0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   dream        0 3 0 1 0 0 0 1 0  1  0  1  0  1  0  1  0  1  0  1  0  0  0  0
##   even         0 1 0 0 0 0 0 0 0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   face         0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   root         0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   still        0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   though       0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   today        0 1 0 0 0 0 0 0 0  0  0  0  0  1  0  0  0  1  0  0  0  0  0  0
##   tomorrow     0 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   creed        0 0 0 1 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   day          0 0 0 1 0 0 0 0 1  0  1  0  1  0  0  0  2  0  0  0  1  0  0  1
##   live         0 0 0 1 0 0 0 0 0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0
##   mean         0 0 0 1 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   nation       0 0 0 1 0 0 0 0 0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  1
##   one          0 0 0 1 0 0 0 0 1  0  1  0  1  0  0  0  2  0  0  0  1  0  0  1
##   rise         0 0 0 1 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   true         0 0 0 1 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   will         0 0 0 1 0 0 0 0 1  0  1  0  2  0  0  0  1  0  0  0  2  0  0  4
##   creat        0 0 0 0 0 1 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

```
## equal       0 0 0 0 0 1 0 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## hold        0 0 0 0 0 1 0 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## men         0 0 0 0 0 1 0 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## selfevid    0 0 0 0 0 1 0 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## truth       0 0 0 0 0 1 0 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## abl         0 0 0 0 0 0 1 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  3
## brotherhood 0 0 0 0 0 0 1 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  1
## former      0 0 0 0 0 0 2 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## georgia     0 0 0 0 0 0 1 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## hill        0 0 0 0 0 0 1 0   0   0  0  0  0  0  0  0  0  0  1  0  0  0  0
## owner       0 0 0 0 0 0 1 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## red         0 0 0 0 0 0 1 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## sit         0 0 0 0 0 0 1 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## slave       0 0 0 0 0 0 2 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## son         0 0 0 0 0 0 2 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## tabl        0 0 0 0 0 0 1 0   0   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## togeth      0 0 0 0 0 0 1 0   0   0  0  0  0  0  0  0  0  0  1  0  0  0  5
## freedom     0 0 0 0 0 0 0 0   1   0  0  0  0  0  0  0  0  0  0  0  0  0  1
## heat        0 0 0 0 0 0 0 0   2   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## injustic    0 0 0 0 0 0 0 0   1   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## justic      0 0 0 0 0 0 0 0   1   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## mississippi 0 0 0 0 0 0 0 0   1   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## oasi        0 0 0 0 0 0 0 0   1   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## oppress     0 0 0 0 0 0 0 0   1   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## state       0 0 0 0 0 0 0 0   2   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## swelter     0 0 0 0 0 0 0 0   2   0  0  0  0  0  0  0  0  0  0  0  0  0  0
## transform   0 0 0 0 0 0 0 0   1   0  0  0  0  0  0  0  0  0  0  0  0  0  1
## charact     0 0 0 0 0 0 0 0   0   0  1  0  0  0  0  0  0  0  0  0  0  0  0
## children    0 0 0 0 0 0 0 0   0   0  1  0  0  0  0  0  0  0  0  0  0  0  0
## color       0 0 0 0 0 0 0 0   0   0  1  0  0  0  0  0  0  0  0  0  0  0  0
## content     0 0 0 0 0 0 0 0   0   0  1  0  0  0  0  0  0  0  0  0  0  0  0
## four        0 0 0 0 0 0 0 0   0   0  1  0  0  0  0  0  0  0  0  0  0  0  0
## judg        0 0 0 0 0 0 0 0   0   0  1  0  0  0  0  0  0  0  0  0  0  0  0
## littl       0 0 0 0 0 0 0 0   0   0  1  0  0  0  2  0  0  0  0  0  0  0  0
## skin        0 0 0 0 0 0 0 0   0   0  1  0  0  0  0  0  0  0  0  0  0  0  0
## alabama     0 0 0 0 0 0 0 0   0   0  0  0  0  0  2  0  0  0  0  0  0  0  0
## black       0 0 0 0 0 0 0 0   0   0  0  0  0  0  2  0  0  0  0  0  0  0  0
## boy         0 0 0 0 0 0 0 0   0   0  0  0  0  0  2  0  0  0  0  0  0  0  0
## brother     0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## drip        0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## girl        0 0 0 0 0 0 0 0   0   0  0  0  0  0  2  0  0  0  0  0  0  0  0
## governor    0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## hand        0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## interposit  0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## join        0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## lip         0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## nullif      0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## racist      0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## right       0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## sister      0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## vicious     0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## white       0 0 0 0 0 0 0 0   0   0  0  0  0  0  2  0  0  0  0  0  0  0  0
## word        0 0 0 0 0 0 0 0   0   0  0  0  0  0  1  0  0  0  0  0  0  0  0
## crook       0 0 0 0 0 0 0 0   0   0  0  0  0  0  0  0  0  0  1  0  0  0  0
```

```
## everi        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   2 0 0 0 0
## exalt        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## flesh        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## glori        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## lord         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## low          0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## made         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   3 0 0 0 0
## mountain     0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 1
## place        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   2 0 0 0 0
## plain        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## reveal       0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## rough        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## see          0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## shall        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   4 0 0 0 0
## straight     0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## valley       0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   1 0 0 0 0
## back         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 1 0 0
## faith        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 1 0 3
## hope         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 1 0 1
## south        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 1 0 0
## beauti       0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## despair      0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## discord      0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## free         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## hew          0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## jail         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## jangl        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## know         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## pray         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## stand        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## stone        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## struggl      0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## symphoni     0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## work         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 1
## god          0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## new          0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## sing         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## countri      0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## land         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## liberti      0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## sweet        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## thee         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## tis          0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## die          0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## father       0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## pilgrim      0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## pride        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## let          0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## mountainsid  0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## ring         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## america      0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## becom        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## great        0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
## must         0 0 0 0 0 0 0 0 0   0 0 0 0 0 0 0 0 0 0   0 0 0 0 0
```

```
##   hampshir     0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   hilltop      0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   prodigi      0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   mighti       0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   york         0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   allegheni    0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   heighten     0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   pennsylvania 0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   colorado     0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   rocki        0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   snowcap      0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   california   0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   curvac       0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   slope        0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   lookout      0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   tennesse     0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   molehil      0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   allow        0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   cathol       0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   citi         0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   gentil       0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   hamlet       0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   happen       0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   jew          0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   negro        0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   old          0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   protest      0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   speed        0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   spiritu      0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   villag       0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   last         0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   almighti     0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##   thank        0 0 0 0 0 0 0 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##            Docs
## Terms       25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
##   american   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   deepli     0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   difficulti 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   dream      0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   even       0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   face       0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   root       0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   still      0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   though     0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   today      0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   tomorrow   0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   creed      0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   day        0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0
##   live       0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   mean       0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   nation     0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   one        0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   rise       0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   true       0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

```
##   will         0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0
##   creat        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   equal        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   hold         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   men          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0
##   selfevid     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   truth        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   abl          0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0
##   brotherhood  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   former       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   georgia      0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
##   hill         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
##   owner        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   red          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   sit          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   slave        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   son          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   tabl         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   togeth       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   freedom      0 0 0 0 0 1 0 1 1 1 1 1 0 0 1 1 1 1 1 0 0
##   heat         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   injustic     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   justic       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   mississippi  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
##   oasi         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   oppress      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   state        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   swelter      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   transform    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   charact      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   children     0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   color        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   content      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   four         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   judg         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   littl        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   skin         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   alabama      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   black        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   boy          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   brother      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   drip         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   girl         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   governor     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   hand         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   interposit   0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   join         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   lip          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   nullif       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   racist       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   right        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   sister       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   vicious      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   white        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
```

```
## word           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0
## crook          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## everi          0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  1  1  4  0  0
## exalt          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## flesh          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## glori          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## lord           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## low            0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## made           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## mountain       0  0  0  0  0  0  0  0  1  0  0  0  0  0  1  1  0  0  0  0  0
## place          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## plain          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## reveal         0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## rough          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## see            0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## shall          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## straight       0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## valley         0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## back           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## faith          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## hope           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## south          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## beauti         0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## despair        0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## discord        0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## free           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2  0
## hew            0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## jail           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## jangl          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## know           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## pray           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## stand          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## stone          0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0  0  0
## struggl        0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## symphoni       0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## work           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## god            0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0
## new            0  1  0  0  0  0  0  1  1  0  0  0  0  0  0  0  0  0  0  0  0
## sing           0  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  0  0
## countri        0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## land           0  0  0  1  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## liberti        0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## sweet          0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## thee           0  0  0  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## tis            0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## die            0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## father         0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## pilgrim        0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## pride          0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## let            0  0  0  0  0  1  0  1  1  1  1  1  0  0  1  1  1  1  1  0  0
## mountainsid    0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  1  0  0  0  0
## ring           0  0  0  0  0  1  0  1  1  1  1  1  0  0  1  1  1  1  2  0  0
## america        0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## becom          0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

```
##   great        0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   must         0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   hampshir     0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
##   hilltop      0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
##   prodigi      0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
##   mighti       0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
##   york         0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
##   allegheni    0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
##   heighten     0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
##   pennsylvania 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
##   colorado     0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
##   rocki        0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
##   snowcap      0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
##   california   0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
##   curvac       0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
##   slope        0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
##   lookout      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
##   tennesse     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
##   molehil      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
##   allow        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   cathol       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   citi         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   gentil       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   hamlet       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   happen       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   jew          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   negro        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   old          0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   protest      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   speed        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   spiritu      0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   villag       0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
##   last         0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0
##   almighti     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##   thank        0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##             Docs
## Terms      46
##   american    0
##   deepli      0
##   difficulti  0
##   dream       0
##   even        0
##   face        0
##   root        0
##   still       0
##   though      0
##   today       0
##   tomorrow    0
##   creed       0
##   day         0
##   live        0
##   mean        0
##   nation      0
##   one         0
```

```
##     rise           0
##     true           0
##     will           0
##     creat          0
##     equal          0
##     hold           0
##     men            0
##     selfevid       0
##     truth          0
##     abl            0
##     brotherhood    0
##     former         0
##     georgia        0
##     hill           0
##     owner          0
##     red            0
##     sit            0
##     slave          0
##     son            0
##     tabl           0
##     togeth         0
##     freedom        0
##     heat           0
##     injustic       0
##     justic         0
##     mississippi    0
##     oasi           0
##     oppress        0
##     state          0
##     swelter        0
##     transform      0
##     charact        0
##     children       0
##     color          0
##     content        0
##     four           0
##     judg           0
##     littl          0
##     skin           0
##     alabama        0
##     black          0
##     boy            0
##     brother        0
##     drip           0
##     girl           0
##     governor       0
##     hand           0
##     interposit     0
##     join           0
##     lip            0
##     nullif         0
##     racist         0
##     right          0
##     sister         0
```

```
##   vicious        0
##   white          0
##   word           0
##   crook          0
##   everi          0
##   exalt          0
##   flesh          0
##   glori          0
##   lord           0
##   low            0
##   made           0
##   mountain       0
##   place          0
##   plain          0
##   reveal         0
##   rough          0
##   see            0
##   shall          0
##   straight       0
##   valley         0
##   back           0
##   faith          0
##   hope           0
##   south          0
##   beauti         0
##   despair        0
##   discord        0
##   free           1
##   hew            0
##   jail           0
##   jangl          0
##   know           0
##   pray           0
##   stand          0
##   stone          0
##   struggl        0
##   symphoni       0
##   work           0
##   god            1
##   new            0
##   sing           0
##   countri        0
##   land           0
##   liberti        0
##   sweet          0
##   thee           0
##   tis            0
##   die            0
##   father         0
##   pilgrim        0
##   pride          0
##   let            0
##   mountainsid    0
##   ring           0
```

```
##    america          0
##    becom            0
##    great            0
##    must             0
##    hampshir         0
##    hilltop          0
##    prodigi          0
##    mighti           0
##    york             0
##    allegheni        0
##    heighten         0
##    pennsylvania     0
##    colorado         0
##    rocki            0
##    snowcap          0
##    california       0
##    curvac           0
##    slope            0
##    lookout          0
##    tennesse         0
##    molehil          0
##    allow            0
##    cathol           0
##    citi             0
##    gentil           0
##    hamlet           0
##    happen           0
##    jew              0
##    negro            0
##    old              0
##    protest          0
##    speed            0
##    spiritu          0
##    villag           0
##    last             1
##    almighti         1
##    thank            1
```

```
dim(m)
```

```
## [1] 162  46
```

There are 162 unique words in 46 line of text

Word cloud

```
v = sort(rowSums(m), decreasing = T)
```

```
v
```

```
##         will   freedom      ring     dream       day        let
##           17        13        12        11        11         11
##         everi       one       abl    togeth  mountain    nation
##            9         8         8         7         5          4
##         shall     faith      free     today       men       hill
##            4         4         4         3         3          3
##         state  children     littl     black     white       made
##            3         3         3         3         3          3
##          god       new      sing      land      last       even
##            3         3         3         3         3          2
##         live      mean      true brotherhood   former   georgia
##            2         2         2         2         2          2
##        slave       son      heat mississippi  swelter transform
##            2         2         2         2         2          2
##      alabama       boy      girl      hand      join       word
##            2         2         2         2         2          2
##        place      hope     stone      thee mountainsid american
##            2         2         2         2         2          1
##       deepli difficulti    face      root     still     though
##            1         1         1         1         1          1
##      tomorrow     creed      rise     creat     equal       hold
##            1         1         1         1         1          1
##      selfevid     truth     owner       red       sit       tabl
##            1         1         1         1         1          1
##      injustic     justic      oasi   oppress   charact      color
##            1         1         1         1         1          1
##      content      four      judg      skin   brother       drip
##            1         1         1         1         1          1
##     governor interposit      lip    nullif    racist      right
##            1         1         1         1         1          1
##       sister   vicious     crook     exalt     flesh      glori
##            1         1         1         1         1          1
##         lord       low     plain    reveal     rough        see
##            1         1         1         1         1          1
##     straight    valley      back     south    beauti    despair
##            1         1         1         1         1          1
##      discord       hew      jail     jangl      know       pray
##            1         1         1         1         1          1
##        stand   struggl  symphoni      work   countri    liberti
##            1         1         1         1         1          1
##        sweet       tis       die    father   pilgrim      pride
##            1         1         1         1         1          1
##      america     becom     great      must  hampshir    hilltop
##            1         1         1         1         1          1
##      prodigi    mighti      york allegheni  heighten pennsylvania
##            1         1         1         1         1          1
##     colorado     rocki   snowcap california   curvac      slope
##            1         1         1         1         1          1
##       lookout  tennesse   molehil     allow    cathol       citi
##            1         1         1         1         1          1
##       gentil    hamlet    happen       jew     negro        old
##            1         1         1         1         1          1
##      protest     speed   spiritu    villag  almighti      thank
##            1         1         1         1         1          1
```

```
d = data.frame(word = names(v), freq = v)

d
```

```
##                     word freq
## will               will   17
## freedom         freedom   13
## ring               ring   12
## dream             dream   11
## day                 day   11
## let                 let   11
## everi             everi    9
## one                 one    8
## abl                 abl    8
## togeth           togeth    7
## mountain       mountain    5
## nation           nation    4
## shall             shall    4
## faith             faith    4
## free               free    4
## today             today    3
## men                 men    3
## hill               hill    3
## state             state    3
## children       children    3
## littl             littl    3
## black             black    3
## white             white    3
## made               made    3
## god                 god    3
## new                 new    3
## sing               sing    3
## land               land    3
## last               last    3
## even               even    2
## live               live    2
## mean               mean    2
## true               true    2
## brotherhood brotherhood    2
## former           former    2
## georgia         georgia    2
## slave             slave    2
## son                 son    2
## heat               heat    2
## mississippi mississippi    2
## swelter         swelter    2
## transform     transform    2
## alabama         alabama    2
## boy                 boy    2
## girl               girl    2
## hand               hand    2
## join               join    2
## word               word    2
```
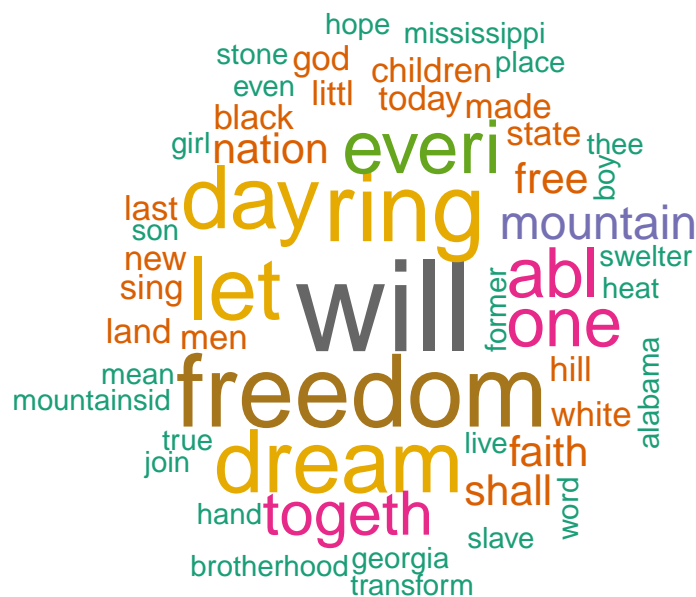
```
## place             place        2
## hope               hope        2
## stone             stone        2
## thee               thee        2
## mountainsid   mountainsid      2
## american       american       1
## deepli             deepli      1
## difficulti     difficulti      1
## face               face        1
## root               root        1
## still             still        1
## though           though        1
## tomorrow       tomorrow        1
## creed             creed        1
## rise               rise        1
## creat             creat        1
## equal             equal        1
## hold               hold        1
## selfevid       selfevid        1
## truth             truth        1
## owner             owner        1
## red                 red        1
## sit                 sit        1
## tabl               tabl        1
## injustic       injustic        1
## justic           justic        1
## oasi               oasi        1
## oppress         oppress        1
## charact         charact        1
## color             color        1
## content         content        1
## four               four        1
## judg               judg        1
## skin               skin        1
## brother         brother        1
## drip               drip        1
## governor       governor        1
## interposit     interposit      1
## lip                 lip        1
## nullif           nullif        1
## racist           racist        1
## right             right        1
## sister           sister        1
## vicious         vicious        1
## crook             crook        1
## exalt             exalt        1
## flesh             flesh        1
## glori             glori        1
## lord               lord        1
## low                 low        1
## plain             plain        1
## reveal           reveal        1
## rough             rough        1
## see                 see        1
```

```
## straight      straight     1
## valley        valley       1
## back          back         1
## south         south        1
## beauti        beauti       1
## despair       despair      1
## discord       discord      1
## hew           hew          1
## jail          jail         1
## jangl         jangl        1
## know          know         1
## pray          pray         1
## stand         stand        1
## struggl       struggl      1
## symphoni      symphoni     1
## work          work         1
## countri       countri      1
## liberti       liberti      1
## sweet         sweet        1
## tis           tis          1
## die           die          1
## father        father       1
## pilgrim       pilgrim      1
## pride         pride        1
## america       america      1
## becom         becom        1
## great         great        1
## must          must         1
## hampshir      hampshir     1
## hilltop       hilltop      1
## prodigi       prodigi      1
## mighti        mighti       1
## york          york         1
## allegheni     allegheni    1
## heighten      heighten     1
## pennsylvania pennsylvania  1
## colorado      colorado     1
## rocki         rocki        1
## snowcap       snowcap      1
## california    california   1
## curvac        curvac       1
## slope         slope        1
## lookout       lookout      1
## tennesse      tennesse     1
## molehil       molehil      1
## allow         allow        1
## cathol        cathol       1
## citi          citi         1
## gentil        gentil       1
## hamlet        hamlet       1
## happen        happen       1
## jew           jew          1
## negro         negro        1
## old           old          1
```

```
## protest          protest     1
## speed              speed     1
## spiritu           spiritu     1
## villag             villag     1
## almighti         almighti     1
## thank              thank     1
```

```r
set.seed(12)
wordcloud(words=d$word, freq = d$freq, min.freq = 2, max.words = 150,
          random.order = F, colors = brewer.pal(8, 'Dark2'))
```


Word Association

Words associate with freedom

```r
findAssocs(dtm, terms = 'freedom', corlimit = 0.3)$freedom
```

```
##          let        ring    mountain mississippi    transform        stone
##         0.89        0.86        0.40        0.34         0.34         0.34
## mountainsid       state       everi
##         0.34        0.32        0.32
```

relation of words that appear at least 10 times

```
findAssocs(dtm, terms=findFreqTerms(dtm, lowfreq = 10), corlimit = 0.3)
```

```
## $dream
##   american     deepli difficulti       face       root      still     though
##       0.74       0.74       0.74       0.74       0.74       0.74       0.74
##   tomorrow       even      today        one
##       0.74       0.67       0.67       0.46
##
## $day
##        one        abl   children      littl      black      white
##       0.78       0.60       0.56       0.56       0.56       0.56
##       mean       hand       join       word    alabama        boy
##       0.52       0.52       0.52       0.52       0.51       0.51
##     brother       drip       girl   governor  interposit        lip
##       0.51       0.51       0.51       0.51       0.51       0.51
##      nullif     racist      right     sister    vicious        god
##       0.51       0.51       0.51       0.51       0.51       0.39
##        sing       live brotherhood  transform     nation
##       0.39       0.31       0.31       0.31       0.30
##
## $will
##        abl     togeth     beauti    despair    discord        hew
##       0.80       0.68       0.62       0.62       0.62       0.62
##       jail      jangl       know       pray      stand    struggl
##       0.62       0.62       0.62       0.62       0.62       0.62
##    symphoni       work   children        one       faith brotherhood
##       0.62       0.62       0.60       0.58       0.57       0.52
##   transform     nation       mean       hope      stone        god
##       0.52       0.49       0.40       0.40       0.40       0.39
##        sing   mountain      everi
##       0.39       0.33       0.32
##
## $freedom
##    mountain mississippi  transform      stone mountainsid      state
##       0.40       0.34       0.34       0.34       0.34       0.32
##       everi
##       0.32
##
## $let
## mountainsid      everi
##       0.38       0.36
##
## $ring
##       everi      allow     cathol       citi     gentil     hamlet
##       0.57       0.53       0.53       0.53       0.53       0.53
##      happen        jew      negro        old    protest      speed
##       0.53       0.53       0.53       0.53       0.53       0.53
##     spiritu     villag        men       hand       join       word
##       0.53       0.53       0.45       0.32       0.32       0.32
## mountainsid
##       0.32
```

Sentiment Analysis

```
library(sentimentr)
```

```
## Warning: package 'sentimentr' was built under R version 4.4.2
```

```
##
## Attaching package: 'sentimentr'
```

```
## The following object is masked from 'package:syuzhet':
##
##      get_sentences
```

```
x = 'Sentiment analysis is super fun'
```

```
sentiment(x)
```

```
## Key: <element_id, sentence_id>
##    element_id sentence_id word_count sentiment
##         <int>       <int>      <int>     <num>
## 1:          1           1          5 0.6708204
```

0.67 which is positive shows that the sentiment is positive

```
y = 'sentiment analysis is super boring. But I do love working with R'
sentiment(y)
```

```
## Key: <element_id, sentence_id>
##    element_id sentence_id word_count  sentiment
##         <int>       <int>      <int>      <num>
## 1:          1           1          5 -0.1118034
## 2:          1           2          7  0.8504201
```

```
sentiment_text = get_sentiment(text, method='syuzhet')
```

```
sentiment_text
```

```
##  [1]  0.00 -0.25  0.00  0.75  0.00  0.00  0.00 -0.50  0.00  1.00  0.00  0.85
## [13]  0.00  0.25  0.00 -0.70  0.00  0.25  0.00  1.60  0.00  0.50  0.00  2.15
## [25]  0.00  1.40  0.00  1.85 -0.25  0.75  1.00  2.05  2.30  0.75  0.75  0.75
## [37]  0.00  0.00  0.75  0.75  0.75  0.75  2.10  0.50  0.00  2.00
```
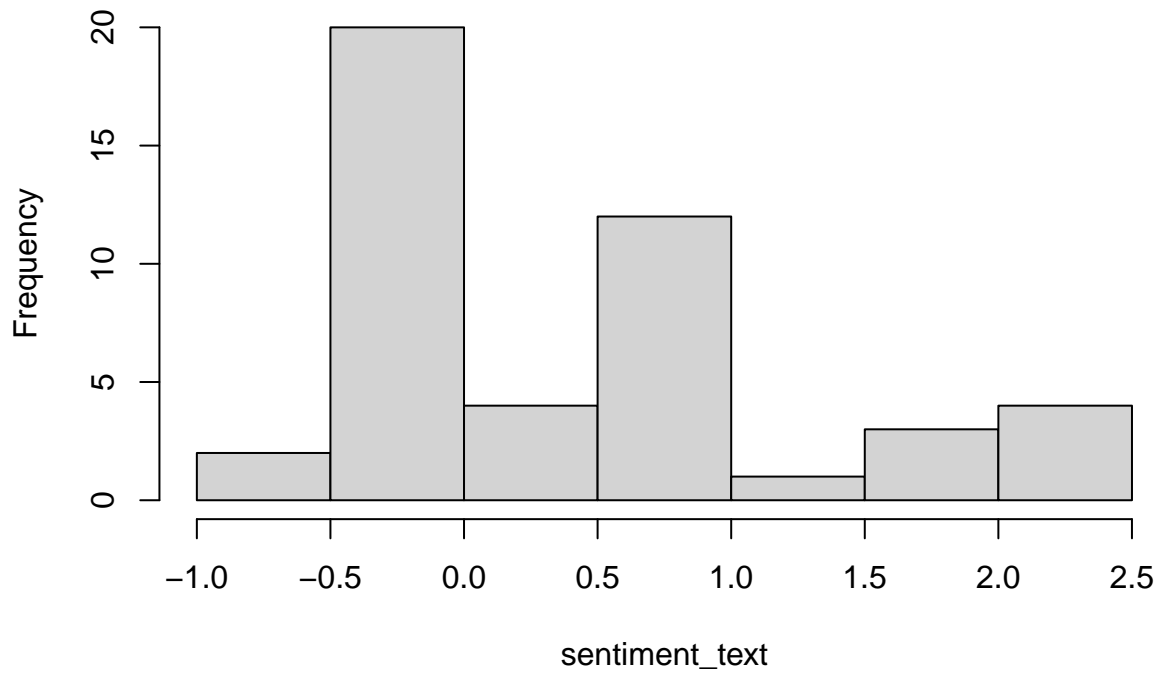
```
hist(sentiment_text)
```

Histogram of sentiment_text


```r
summary(sentiment_text)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.7000  0.0000  0.2500  0.5402  0.7500  2.3000
```

Emotion classification

The NRC Emotion Lexicon is a list of English correspond to eight basic emotions:

1. Anger
2. Fear
3. Anticipation
4. Trust
5. Surprise
6. Sadness
7. Joy
8. Disgust

```r
d2 = get_nrc_sentiment(text)

d2
```

```
##    anger anticipation disgust fear joy sadness surprise trust negative positive
```

```
## 1     0        0        0   0   0       0       0   0        0        0
## 2     0        1        0   0   0       1       0   1        1        1
## 3     0        0        0   0   0       0       0   0        0        0
## 4     0        0        0   0   0       0       0   1        0        0
## 5     0        0        0   0   0       0       0   0        0        0
## 6     0        0        0   0   0       0       0   1        0        1
## 7     0        0        0   0   0       0       0   0        0        0
## 8     1        0        0   1   0       1       0   1        1        1
## 9     0        0        0   0   0       0       0   0        0        0
## 10    2        1        1   1   2       1       0   3        2        3
## 11    0        0        0   0   0       0       0   0        0        0
## 12    0        0        0   0   1       0       0   2        0        1
## 13    0        0        0   0   0       0       0   0        0        0
## 14    0        0        0   0   0       0       0   0        0        0
## 15    0        0        0   0   0       0       0   0        0        0
## 16    2        1        1   0   1       1       0   2        3        2
## 17    0        0        0   0   0       0       0   0        0        0
## 18    0        0        0   0   0       0       0   0        0        0
## 19    0        0        0   0   0       0       0   0        0        0
## 20    0        2        2   0   2       0       0   3        1        3
## 21    0        0        0   0   0       0       0   0        0        0
## 22    0        2        0   0   2       0       1   2        0        2
## 23    0        0        0   0   0       0       0   0        0        0
## 24    3        5        1   4   6       3       2   6        4        8
## 25    0        0        0   0   0       0       0   0        0        0
## 26    0        2        0   1   2       1       0   2        0        2
## 27    0        0        0   0   0       0       0   0        0        0
## 28    0        3        0   0   3       1       2   3        0        4
## 29    0        0        0   0   1       0       0   0        0        2
## 30    0        0        0   0   1       0       0   1        0        1
## 31    0        0        0   0   0       0       0   1        0        0
## 32    0        0        0   0   1       0       0   1        0        2
## 33    1        0        0   1   2       0       0   2        0        2
## 34    0        0        0   0   1       0       0   1        0        1
## 35    0        0        0   0   1       0       0   1        0        1
## 36    0        0        0   0   1       0       0   1        0        1
## 37    0        0        0   0   0       0       0   0        0        0
## 38    0        0        0   0   0       0       0   0        0        0
## 39    1        1        0   0   1       0       0   1        1        1
## 40    0        1        0   0   1       0       0   1        0        1
## 41    0        0        0   0   1       0       0   1        0        1
## 42    0        0        0   0   1       0       0   1        0        1
## 43    1        3        0   1   4       3       0   4        3        5
## 44    0        0        0   0   0       0       0   0        0        0
## 45    0        0        0   0   0       0       0   0        0        0
## 46    0        1        0   1   1       0       0   1        0        2
```

```r
td = data.frame(t(d2))
td_new = data.frame(rowSums(td))

td_new
```

```
##               rowSums.td.
## anger                  11
```

```
## anticipation          23
## disgust                5
## fear                  10
## joy                   36
## sadness               12
## surprise               5
## trust                 44
## negative              16
## positive              49
```

```r
names(td_new)[1] = 'count'

td_new = cbind('sentiment'=rownames(td_new), td_new)

rownames(td_new) = NULL

qplot(sentiment, weight = count, data=td_new, geom = 'bar',
      fill = sentiment,
      ylab = 'count') + ggtitle('Sentiment Score')
```

```
## Warning: 'qplot()' was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```