

# Wordcloud of negative and positive words

Jerrin Joe Varghese

December 6, 2017

# Table of Contents

Install and Load libraries

Hands on Project Gutenberg

Download Dracula the Book

Unpack the Words

The Bing Lexicon for sentiments

The Inner Join

Calculate Frequency for Wordcloud

Wordcloud

# Install and Load libraries

► `library(dplyr)`

# Install and Load libraries

- ▶ `library(dplyr)`
- ▶ `library(gutenbergr)`

# Install and Load libraries

- ▶ `library(dplyr)`
- ▶ `library(gutenbergr)`
- ▶ `library(tidytext)`

# Install and Load libraries

- ▶ `library(dplyr)`
- ▶ `library(gutenbergr)`
- ▶ `library(tidytext)`
- ▶ `library(ggplot2)`

# Install and Load libraries

- ▶ `library(dplyr)`
- ▶ `library(gutenbergr)`
- ▶ `library(tidytext)`
- ▶ `library(ggplot2)`
- ▶ `library(stringr)`

# Install and Load libraries

- ▶ `library(dplyr)`
- ▶ `library(gutenbergr)`
- ▶ `library(tidytext)`
- ▶ `library(ggplot2)`
- ▶ `library(stringr)`
- ▶ `library(wordcloud)`



# Install and Load libraries

- ▶ `library(dplyr)`
- ▶ `library(gutenbergr)`
- ▶ `library(tidytext)`
- ▶ `library(ggplot2)`
- ▶ `library(stringr)`
- ▶ `library(wordcloud)`
- ▶ `library(wordcloud2)`

# Access Project Gutenberg

```
df<-gutenberg_works(str_detect(title,'Dracula'))  
df$gutenberg_id
```

```
## [1] 345 10150
```

```
df$title
```

```
## [1] "Dracula" "Dracula's Guest"
```

# Download Dracula

```
dracula<-gutenberg_download(345)
colnames(dracula)

## [1] "gutenberg_id" "text"

substr(dracula$text[500],1,21)

## [1] "my own disappointment"
```

# Unpack the Words

```
dracula_words<-dracula%>%  
  unnest_tokens(word,text)  
colnames(dracula_words)  
  
## [1] "guttenberg_id" "word"  
  
dracula_words[498:500,]  
  
## # A tibble: 3 x 2  
##   guttenberg_id  word  
##           <int> <chr>  
## 1           345  fail  
## 2           345   to  
## 3           345  have
```

# The Bing Lexicon

```
bing = get_sentiments('bing')
colnames(bing)

## [1] "word"      "sentiment"

bing[498:500,]

## # A tibble: 3 x 2
##       word sentiment
##   <chr>    <chr>
## 1 bereave negative
## 2 bereavement negative
## 3 bereft   negative
```

# The Inner Join

```
dracula_words <- inner_join(dracula_words,bing)
```

```
## Joining, by = "word"
```

```
dracula_words$gutenberg_id <- NULL  
dracula_words[498:500,]
```

```
## # A tibble: 3 x 2  
##       word sentiment  
##   <chr>    <chr>  
## 1  great  positive  
## 2   love  positive  
## 3 crowded negative
```

# Frequency of words

```
dracula_words<-dracula_words%>%  
group_by(word)%>%  
summarise(freq=n(),sentiment=first(sentiment))  
dracula_words[25:30,]
```

```
## # A tibble: 6 x 3  
##       word    freq sentiment  
##   <chr> <int>    <chr>  
## 1  aghast      1  negative  
## 2   agony      8  negative  
## 3  alarm       3  negative  
## 4 alarmed      5  negative  
## 5 alarming     1  negative  
## 6  aloof       1  negative
```

## Wordcloud

