

Don Quixote: Creating a Word Cloud

Jorge Fitzmaurice

November 27, 2017

Outline

- 1 Install and Load Libraries
- 2 Access Project Gutenberg
- 3 Download Don Quixote
- 4 Unpack the Words
- 5 The Bing Lexicon
- 6 The Inner Join
- 7 The Fear Words
- 8 The Word Cloud
- 9 The Word Cloud 2

Install and Load Libraries

- `library(dplyr)`

Install and Load Libraries

- `library(dplyr)`
- `library(tidytext)`

Install and Load Libraries

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

Install and Load Libraries

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

Install and Load Libraries

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

Install and Load Libraries

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

Install and Load Libraries

- `library(dplyr)`
- `library(tidytext)`
- `library(gutenbergr)`
- `library(wordcloud)`
- `library(stringr)`
- `library(wordcloud2)`
- `library(webshot)`

Access Project Gutenberg

```
df<-gutenberg_works(str_detect(title,'Don Quixote'))  
df$gutenberg_id[1:5]
```

```
## [1] 996 5903 5904 5905 5906
```

```
df$title[1:5]
```

```
## [1] "Don Quixote"
```

```
## [2] "The History of Don Quixote, Volume 1, Part 01"
```

```
## [3] "The History of Don Quixote, Volume 1, Part 02"
```

```
## [4] "The History of Don Quixote, Volume 1, Part 03"
```

```
## [5] "The History of Don Quixote, Volume 1, Part 04"
```

Download Don Quixote

```
quixote<-gutenberg_download(996)  
colnames(quixote)
```

```
## [1] "gutenberg_id" "text"
```

```
substr(quixote$text[1001],1,21)
```

```
## [1] "His idea of \"shortly\""
```

Unpack the Words

```
quixote_words<-quixote%>%  
  unnest_tokens(word,text)  
colnames(quixote_words)  
  
## [1] "guttenberg_id" "word"  
  
quixote_words[1000:1003,]  
  
## # A tibble: 4 x 2  
##   guttenberg_id    word  
##         <int>   <chr>  
## 1         996 quarrel  
## 2         996   that  
## 3         996    don  
## 4         996 quixote
```

The Bing Lexicon

```
nrc<-get_sentiments('nrc')
colnames(nrc)

## [1] "word"      "sentiment"

nrc[1000:1003,]

## # A tibble: 4 x 2
##       word sentiment
##   <chr>      <chr>
## 1 badness   negative
## 2 bailiff   fear
## 3 bailiff   negative
## 4 bailiff   trust
```

The Inner Join

```
quixote_words<-inner_join(quixote_words,nrc)
```

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

```
quixote_words$gutenberg_id<-NULL
```

```
quixote_words[500:502,]
```

```
## # A tibble: 3 x 2
```

```
##       word sentiment
```

```
##    <chr>      <chr>
```

```
## 1 failing    sadness
```

```
## 2 worse      fear
```

```
## 3 worse     negative
```

The Fear Words I

```
quixote_fear<-quixote_words%>%  
  filter(sentiment == 'fear')%>%  
  group_by(word)%>%  
  summarize(freq = n())%>%  
  filter(freq<=100)
```

The Fear Words II

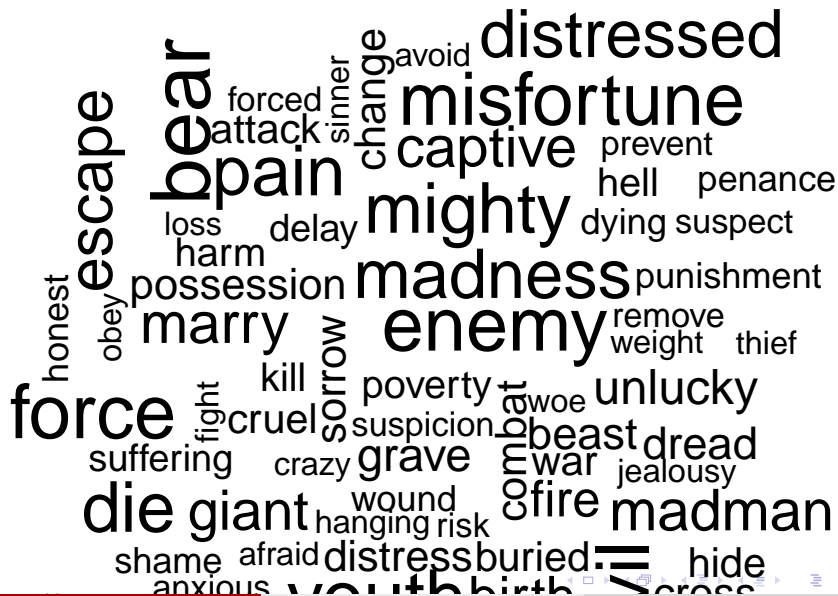
```
quixote_fear
```

```
## # A tibble: 712 x 2
##       word freq
##       <chr> <int>
## 1   abandon     5
## 2 abandoned     4
## 3    abhor      1
## 4 abominable     1
## 5   absence    22
## 6    abuse     17
## 7    abyss      6
## 8  accident      8
## 9  accursed     11
## 10 accusing      1
## # ... with 702 more rows
```


The Word Cloud I

```
plot<-wordcloud(quixote_fear$word,  
                quixote_fear$freq,min.freq = 25)
```

The Word Cloud II



The Word Cloud 2

```
wordcloud2(quixote_fear,color='random-dark')
```

