The Strange Case of Dr. Jekyll and Mr. Hyde in Wordclouds

Elizabeth Bross

December 11, 2017

Outline

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(tm)`

Access Project Gutenberg

```
df<-gutenberg_works(str_detect(title,
'Strange Case of Dr. Jekyll and Mr. Hyde'))

df$gutenberg_id

## [1] 42

df$title

## [1] "The Strange Case of Dr. Jekyll and Mr. Hyde"
```

Download The Strange Case of Dr. Jekyll and Mr. Hyde

```
JandH<-gutenberg_download(42)
colnames(JandH)

## [1] "gutenberg_id" "text"

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

## [1] "of the will?\" But he "
```

Unpack the Words

```
JandH_words<-JandH%>%
  unnest_tokens(word,text)
colnames(JandH_words)

## [1] "gutenberg_id" "word"

JandH_words[498:500,]

## # A tibble: 3 x 2
##   gutenberg_id   word
##          <int>  <chr>
## 1           42   well
## 2           42     it
## 3           42 seemed
```

The Bing Lexicon

```
bing<-get_sentiments('bing')
bing_neg<-bing%>%
  filter(sentiment=='negative')
bing_pos<-bing%>%
  filter(sentiment=='positive')
```

The Inner Join

```
JandH_words_neg<-inner_join(JandH_words,bing_neg)

## Joining, by = "word"

JandH_words_pos<-inner_join(JandH_words,bing_pos)

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

Prepare for the Wordclouds

```
JandH_words_neg<-JandH_words_neg%>%
group_by(word)%>%
summarize(count=n())
```

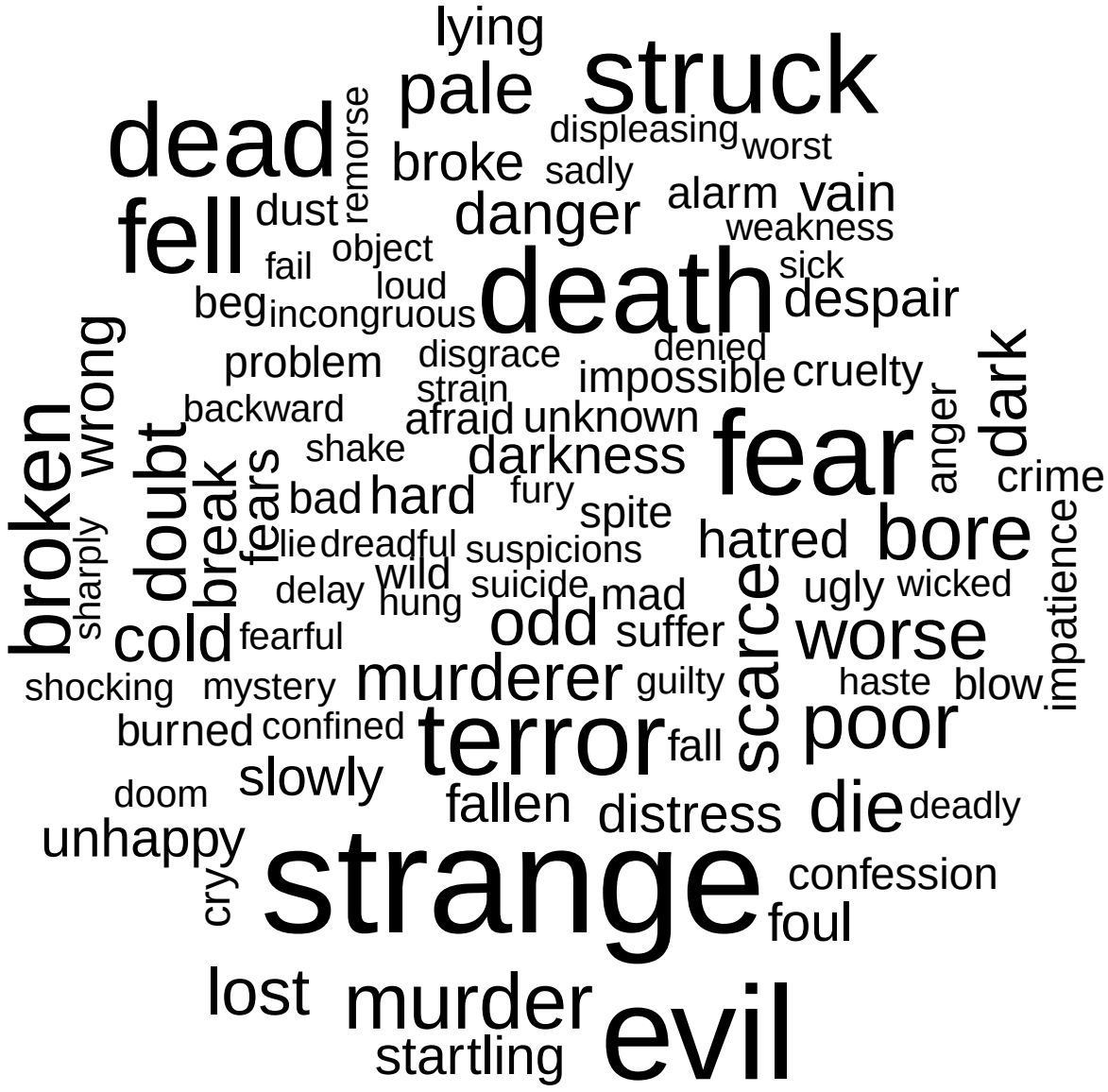
```
JandH_words_pos<-JandH_words_pos%>%
group_by(word)%>%
summarize(count=n())
```

Wordcloud Creation

```
wordcloud(JandH_words_neg$word,JandH_words_neg$count,
          min.freq=3)
```

```
wordcloud(JandH_words_pos$word,JandH_words_pos$count,
          min.freq=2)
```

Negative Word Results



Positive Word Results