Positive and Negative Words in Dracula

Who? Charles Redmond

When? November 15, 2017

Outline

Install and Load Libraries

```
library(gutenbergr)
```

Install and Load Libraries

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

Install and Load Libraries

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

Install and Load Libraries

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

Install and Load Libraries

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

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] "gutenberg_id" "word"

dracula_words[500,]

## # A tibble: 1 x 2
##   gutenberg_id  word
##          <int> <chr>
## 1          345  have
```

The Bing Lexicon

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

## [1] "word"      "sentiment"

bing[500,]

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

The Inner Join

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

## Joining, by = "word"

dracula_words$gutenberg_id<-NULL
colnames(dracula_words)

## [1] "word"      "sentiment"
```

Top Ten Positive Words I

```
dracula_pos<-dracula_words%>%
filter(sentiment=='positive')%>%
group_by(word)%>%
summarize(count=n(),sentiment=first(sentimen
arrange(count)%>%
top_n(10,wt=count)
```

Top Ten Positive Words II

```
  dracula_pos

## # A tibble: 10 x 3
##       word count sentiment
##      <chr> <int>     <chr>
##  1  sweet    66  positive
##  2  ready    71  positive
##  3 better    77  positive
##  4   love    84  positive
##  5  right    99  positive
##  6   work   146  positive
##  7  great   183  positive
##  8   well   245  positive
##  9   good   258  positive
## 10   like   292  positive
```

Top Ten Negative Words

```
## # A tibble: 10 x 3
##          word count sentiment
##         <chr> <int>     <chr>
##  1      hard    49  negative
##  2   trouble    53  negative
##  3      fell    59  negative
##  4      dark    77  negative
##  5   strange    90  negative
##  6     death    94  negative
##  7 terrible   100  negative
##  8      dead   109  negative
##  9      fear   137  negative
## 10      poor   193  negative
```

The Comparison Barplot I

```
dracula_pos$word<-factor(dracula_pos$word,
                         levels=dracula_pos$wo
dracula_neg$word<-factor(dracula_neg$word,
                         levels=dracula_neg$wo

dracula_comp<-rbind(dracula_pos,dracula_neg)

plot<-ggplot()+
geom_bar(data=dracula_comp,aes(x=word,y=count,
fill=sentiment, color=sentiment),stat='identit
coord_flip()+
facet_wrap(~sentiment,scales='free_y')+
scale_fill_manual(values=c('black','#ea6205'))+
scale_color_manual(values=c('#ea6205','black'))
```

The Comparison Barplot II