

A Christmas Carol - Comparison Barplots

Dayana Moncada

December 11, 2017

Outline

Install and Load Libraries

Access Project Gutenberg

Download Carol

Unpack the Words

The Bing Lexicon

The Inner Join

Plotting Sentiments

Negative Words

Comparison Plot

Install and Load Libraries

▶ `library(dplyr)`

▶ `library(tidytext)`

▶ `library(gutenbergr)`

▶ `library(ggplot2)`

▶ `library(stringr)`

Access Project Gutenberg

```
df<-gutenberg_works(str_detect(title,'A Christmas Carol'))
df$gutenberg_id

## [1]      46 19337 30368 40729 41739

df$title

## [1] "A Christmas Carol in Prose; Being a Ghost Story of
## [2] "A Christmas Carol"
## [3] "A Christmas Carol\nThe original manuscript"
## [4] "\"Old Scrooge\": A Christmas Carol in Five Staves."
## [5] "A Christmas Carol; Or, The Miser's Warning!\r\n(Ada
```

Download Carol

```
carol<-guttenberg_download(46)
```

Unpack the Words

```
carol_words<-carol%>%  
  unnest_tokens(word,text)  
colnames(carol_words)  
  
## [1] "guttenberg_id" "word"
```

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

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

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

```
carol_words$gutenberg_id<-NULL
```

```
carol_words[498:500,]
```

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

```
##           word sentiment
```

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

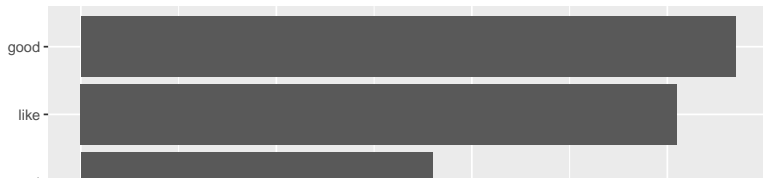
```
## 1  strangest    negative
```

```
## 2    bright    positive
```

```
## 3     clear    positive
```


Positive Words

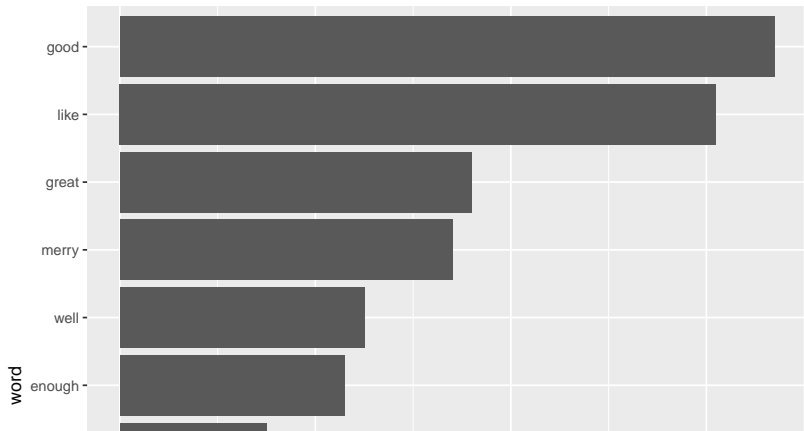
```
carol_pos<-carol_words%>%  
  filter(sentiment=='positive')%>%  
  group_by(word)%>%  
  summarize(count=n(),sentiment=first(sentiment))%>%  
  arrange(count)%>%  
  top_n(10,wt=count)  
  
carol_pos$word<-factor(carol_pos$word,levels=carol_pos$word  
  
ggplot()+  
  geom_bar(data=carol_pos,aes(x=word,y=count),stat='identit  
  coord_flip()
```



Plotting

```
carol_pos$word<-factor(carol_pos$word,levels=carol_pos$word)

ggplot()+
  geom_bar(data=carol_pos,aes(x=word,y=count),stat='identity')
  coord_flip()
```



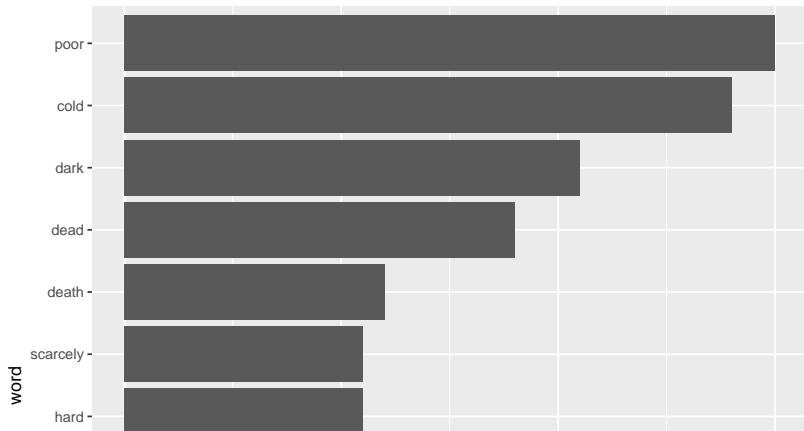
Negative Words

```
carol_neg<-carol_words%>%  
  filter(sentiment=='negative')%>%  
  group_by(word)%>%  
  summarize(count=n(),sentiment=first(sentiment))%>%  
  arrange(count)%>%  
  filter(word != 'miss')%>%  
  top_n(10,wt=count)
```

Plotting

```
carol_neg$word<-factor(carol_neg$word,levels=carol_neg$word)

ggplot()+
  geom_bar(data=carol_neg,aes(x=word,y=count),stat='identity')
  coord_flip()
```



Comparison Plot

```
carol_comp<-rbind(carol_pos,carol_neg)

ggplot()+
  geom_bar(data=carol_comp,aes(x=word,y=count,
                               fill=sentiment, color=sentiment),
  coord_flip()+
  facet_wrap(~sentiment,scales='free_y')+
  scale_fill_manual(values=c('green','red'))+
  scale_color_manual(values=c('green','red'))
```

