comparison Barplots

Shraddha Dubey

November 20, 2017

Outline

install and load libraries

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

install and load libraries

- ```
  library(dplyr)
  ```

- ```
  library(tidytext)
  ```

install and load libraries

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

install and load libraries

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

install and load libraries

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

Access Project Gutenbergr

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

## [1]   345 10150

df$title

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

Download Dracula

```
drac<-gutenberg_download(345)

## Determining mirror for Project Gutenberg from
http://www.gutenberg.org/robot/harvest
## Using mirror http://aleph.gutenberg.org

colnames(drac)

## [1] "gutenberg_id" "text"

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

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

Unpack the words

```
drac_words<-drac%>%
  unnest_tokens(word,text)
colnames(drac_words)

## [1] "gutenberg_id" "word"

drac_words$word[498:500]

## [1] "fail" "to"   "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

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

## Joining, by = "word"

drac_words$gutenberg_id<-NULL
drac_words[498:500,]

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

Positive Darcula Words

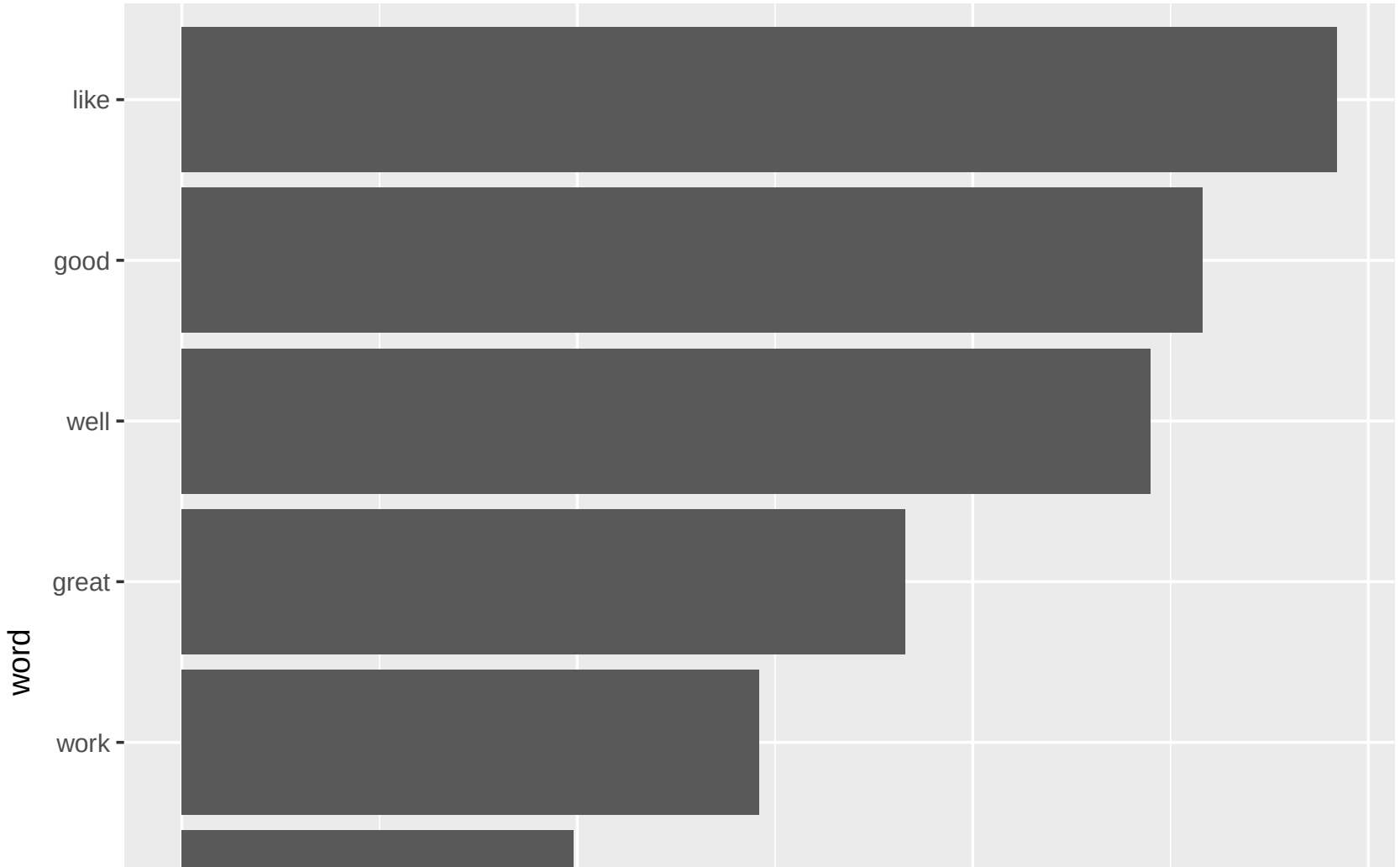
```
drac_pos<-drac_words%>%
  filter(sentiment=='positive')%>%
  group_by(word)%>%
  summarize(count=n(),sentiment=first(sentiment))%>%
  arrange(count)%>%
  top_n(8,wt=count)

drac_pos$word<-factor(drac_pos$word,level=drac_pos$word)
drac_pos[4:6,]

## # A tibble: 3 x 3
##      word count sentiment
##    <fctr> <int>     <chr>
## 1    work   146  positive
## 2   great   183  positive
## 3    well   245  positive
```

Graph of Positive Darcula Words

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



Negative Darcula Words

```
drac_neg<-drac_words%>%
  filter(sentiment=='negative')%>%
  group_by(word)%>%
  summarize(count=n(),sentiment=first(sentiment))%>%
  arrange(count)%>%
  filter(word!='miss')%>%
  top_n(8,wt=count)
drac_neg$word<-factor(drac_neg$word,level=drac_neg$word)
drac_neg[1:4,]
```

```
## # A tibble: 4 x 3
##       word count sentiment
##     <fctr> <int>     <chr>
## 1     fell    59  negative
## 2     dark    77  negative
## 3 strange    90  negative
## 4    death    94  negative
```

Graph of Negative Darcula Words

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