

Sentiment Analysis

```
library(tidytext)
```

```
## Warning: package 'tidytext' was built under R version 3.5.3
```

```
sentiments
```

word	sentiment
<chr>	<chr>
2-faces	negative
abnormal	negative
abolish	negative
abominable	negative
abominably	negative
abominate	negative
abomination	negative
abort	negative
aborted	negative
aborts	negative

1-10 of 6,786 rows

Previous123456...679Next

```
get_sentiments("bing")
```

word <chr>	sentiment <chr>
2-faces	negative
abnormal	negative
abolish	negative
abominable	negative
abominably	negative
abominate	negative
abomination	negative
abort	negative
aborted	negative
aborts	negative
1-10 of 6,786 rows	Previous 1 2 3 4 5 6 ... 679 Next

```
library(janeaustenr) # textual data of Jane Austen's books
```

```
## Warning: package 'janeaustenr' was built under R version 3.5.3
```

```
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 3.5.3
```

```
library(tidytext) # efficient text analysis on our data
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
tidy_data <- austen_books() %>%  
  group_by(book) %>%  
  mutate(linenumber = row_number(),  
         chapter = cumsum(str_detect(text, regex("^chapter [\\divxlc]",  
                                                ignore_case = TRUE)))) %>%  
  ungroup() %>%  
  unnest_tokens(word, text) #convert the text of our books into a tidy format
```

```
positive_sentiment <- get_sentiments("bing") %>%  
  filter(sentiment == "positive")  
tidy_data %>%  
  filter(book == "Pride & Prejudice") %>%  
  semi_join(positive_sentiment) %>%  
  count(word, sort = TRUE)
```

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

word	n
<chr>	<int>
well	224
good	200
great	142
enough	106
better	92
love	92
pleasure	92
happy	83
like	77
happiness	72

1-10 of 592 rows

Previous [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) ... [60](#) [Next](#)

```
negative_sentiment <- get_sentiments("bing") %>%
  filter(sentiment == "negative")
tidy_data %>%
  filter(book == "Pride & Prejudice") %>%
  semi_join(negative_sentiment) %>%
  count(word, sort = TRUE)
```

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

word	n
<chr>	<int>
miss	283

word	n
<chr>	<int>
object	48
scarcely	45
impossible	44
poor	38
afraid	37
doubt	37
sorry	34
lost	29
pain	28

1-10 of 838 rows

Previous [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) ... [84](#) Next

```
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

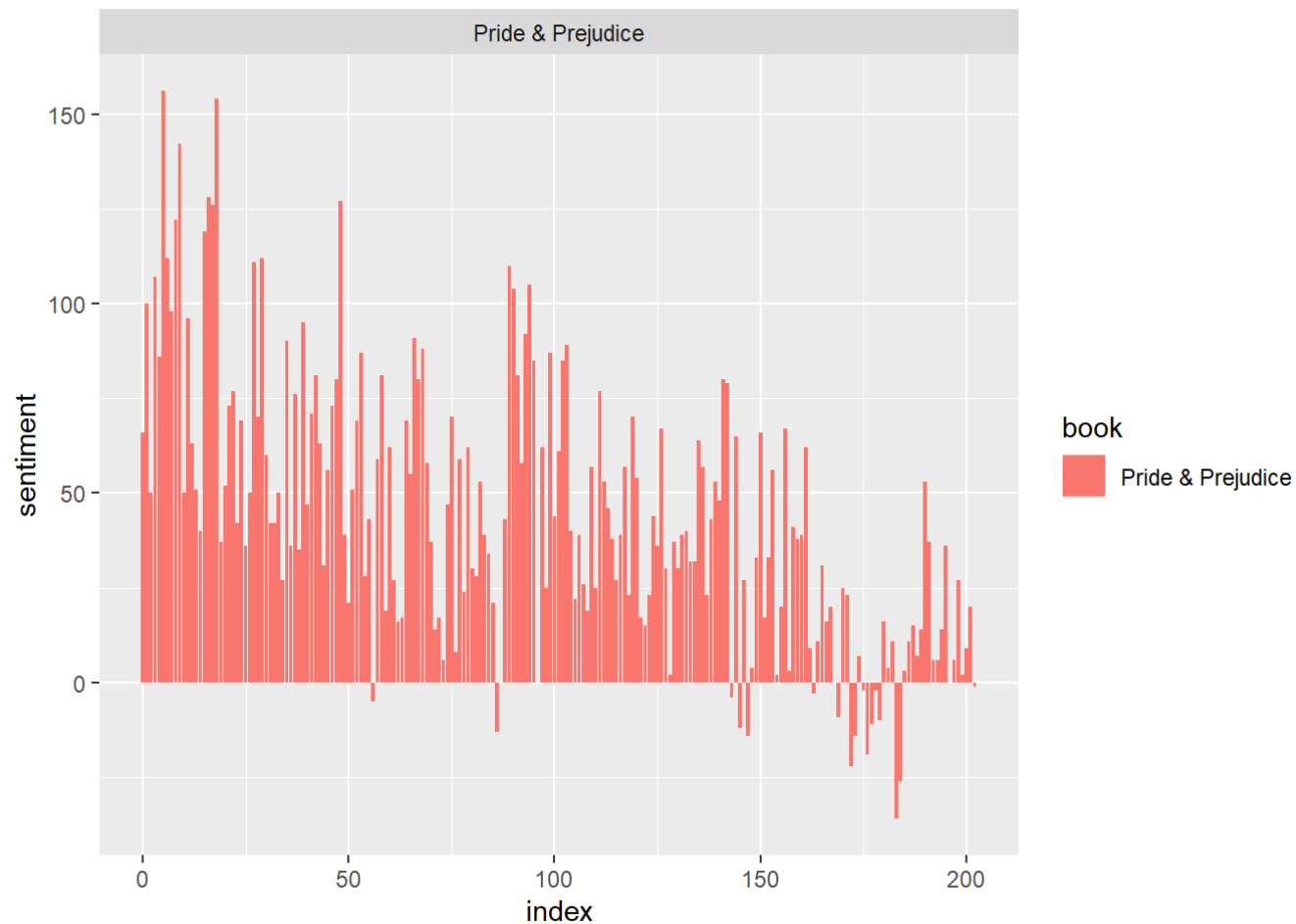
```
bing <- get_sentiments("bing")
Pride_Predjudice_sentiment <- tidy_data %>%
  inner_join(bing) %>%
  count(book = "Pride & Prejudice" , index = linewidth %/% 80, sentiment) %>%
  spread(sentiment, n, fill = 0) %>% #separate columns by sentiment
  mutate(sentiment = positive - negative)#calculate total sentiment
```

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

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
ggplot(Pride_Predjudice_sentiment, aes(index, sentiment, fill = book)) +  
  geom_bar(stat = "identity", show.legend = TRUE) +  
  facet_wrap(~book, ncol = 2, scales = "free_x")
```



```
counting_words <- tidy_data %>%
  inner_join(bing) %>%
  count(word, sentiment, sort = TRUE)
```

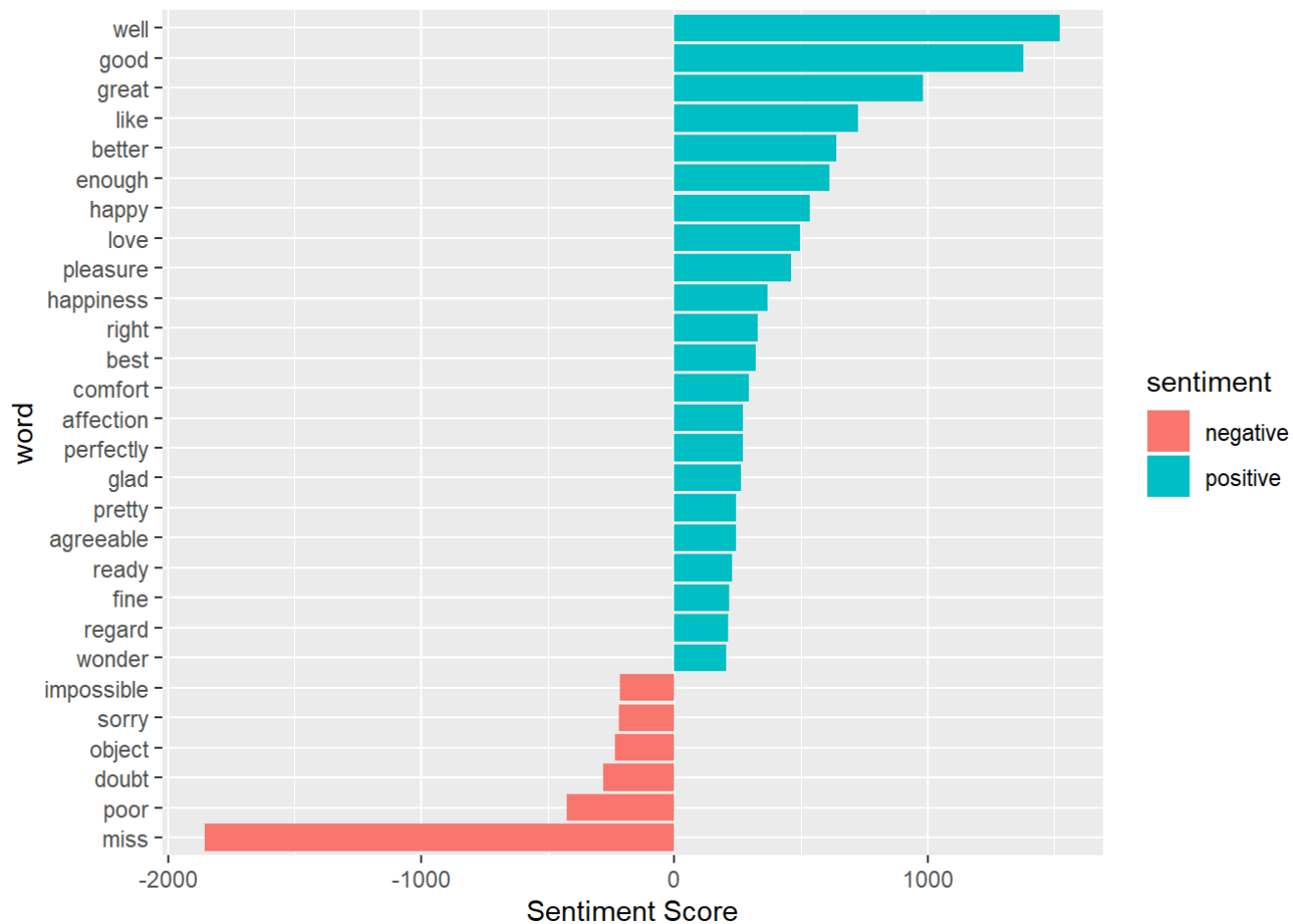
```
## Joining, by = "word"
```

```
head(counting_words)
```

word <chr>	sentiment <chr>	n <int>
miss	negative	1855
well	positive	1523
good	positive	1380
great	positive	981
like	positive	725
better	positive	639

6 rows

```
counting_words %>%
  filter(n > 200) %>%
  mutate(n = ifelse(sentiment == "negative", -n, n)) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment))+
  geom_col() +
  coord_flip() +
  labs(y = "Sentiment Score")
```



```
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
## smiths
```



```
library(wordcloud)
```

```
## Warning: package 'wordcloud' was built under R version 3.5.3
```

```
## Loading required package: RColorBrewer
```

```
tidy_data %>%  
  inner_join(bing) %>%  
  count(word, sentiment, sort = TRUE) %>%  
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%  
  comparison.cloud(colors = c("red", "blue"),  
    max.words = 100)
```

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

negative



positive

```
library(gutenbergr)
```

```
## Warning: package 'gutenbergr' was built under R version 3.5.3
```

```
dickens <- gutenbergr_download(c(98, 1400, 46, 730, 786))
```

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

```
## Using mirror http://aleph.gutenberg.org
```

```
tidy_dickens <- dickens %>%  
  unnest_tokens(word, text) %>%  
  anti_join(stop_words)
```

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

```
bing_word_counts <- tidy_dickens %>%  
  inner_join(get_sentiments("bing")) %>%  
  count(word, sentiment, sort = TRUE) %>%  
  ungroup()
```

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

```
bing_word_counts
```

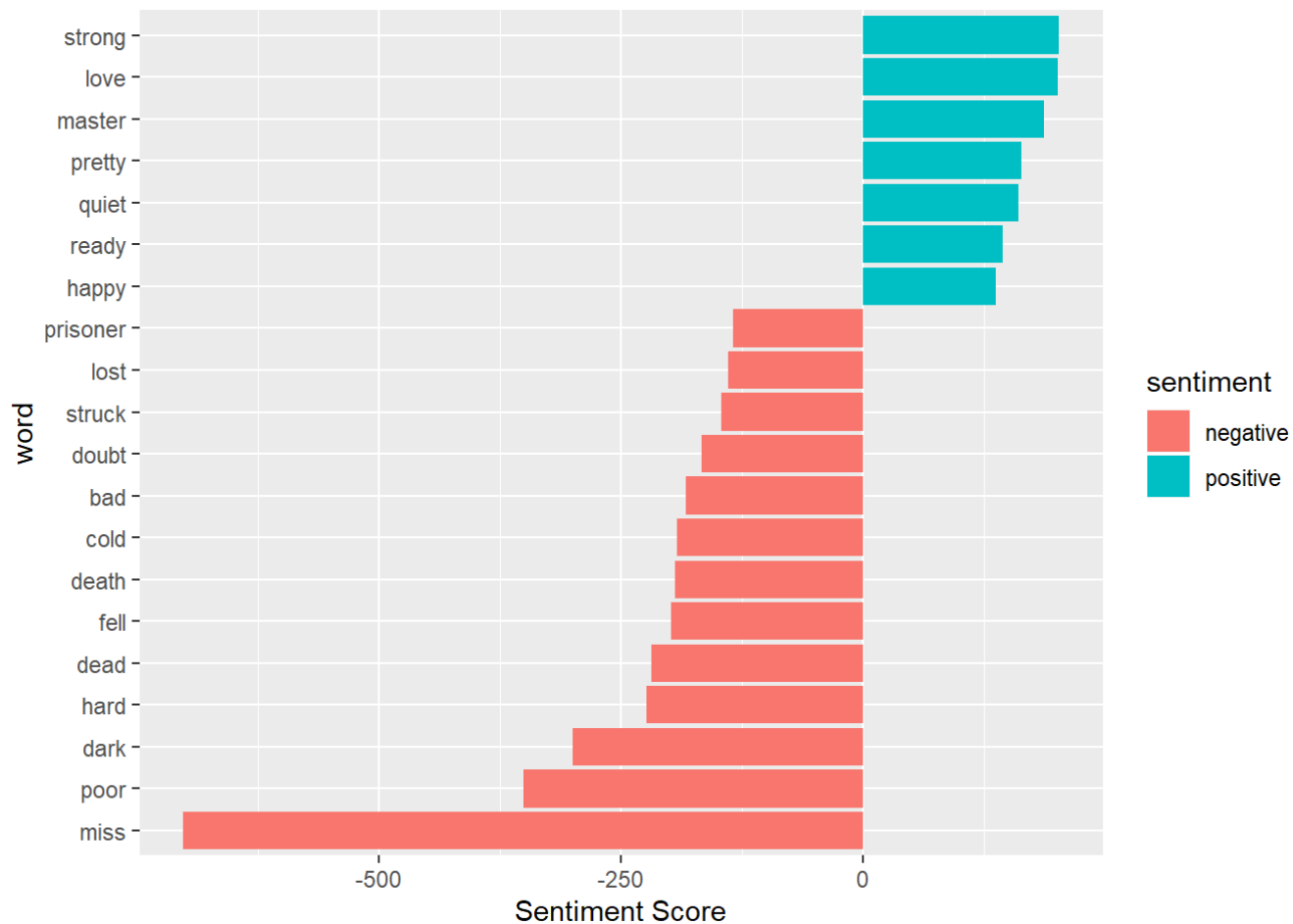
word <chr>	sentiment <chr>	n <int>
miss	negative	702
poor	negative	350
dark	negative	299
hard	negative	223
dead	negative	218
strong	positive	203
love	positive	202

word <chr>	sentiment <chr>	n <int>
fell	negative	198
death	negative	194
cold	negative	192

1-10 of 3,145 rows

Previous **1** 2 3 4 5 6 ... 315 Next

```
bing_word_counts %>%
  filter(n > 130) %>%
  mutate(n = ifelse(sentiment == "negative", -n, n)) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment))+
  geom_col() +
  coord_flip() +
  labs(y = "Sentiment Score")
```



```
tidy_dickens %>%
  inner_join(bing) %>%
  count(word, sentiment, sort = TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("red", "blue"),
    max.words = 100)
```

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

negative

