

Sentiment-Analysis.R

karag

2023-08-03

```
library(tidytext)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(janeaustenr)
library(stringr)
library(reshape2)
```

```
##
```

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

```
##
```

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

```
##
```

```
##      smiths
```

```
library(wordcloud)
```

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

```
sentiments
```

```
## # A tibble: 6,786 x 2
```

```
##   word      sentiment
```

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

```
## 1 2-faces   negative
```

```
## 2 abnormal negative
```

```
## 3 abolish  negative
```

```
## 4 abominable negative
```

```
## 5 abominably negative
```

```
## 6 abominate negative
```

```
## 7 abomination negative
```

```
## 8 abort     negative
```

```
## 9 aborted      negative
## 10 aborts      negative
## # i 6,776 more rows
```

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

```
## # A tibble: 6,786 x 2
##   word      sentiment
##   <chr>     <chr>
## 1 2-faces    negative
## 2 abnormal  negative
## 3 abolish   negative
## 4 abominable negative
## 5 abominably negative
## 6 abominate  negative
## 7 abomination negative
## 8 abort      negative
## 9 aborted   negative
## 10 aborts    negative
## # i 6,776 more rows
```

```
tidy_data <- austen_books() %>%
  group_by(book) %>%
  mutate(linenum = row_number(),
         chapter = cumsum(str_detect(text, regex("^chapter [\\divxlc]",
                                                ignore_case = TRUE)))) %>%
  ungroup() %>%
  unnest_tokens(word, text)
```

```
positive_senti <- get_sentiments("bing") %>%
  filter(sentiment == "positive")
```

```
tidy_data %>%
  filter(book == "Emma") %>%
  semi_join(positive_senti) %>%
  count(word, sort = TRUE)
```

```
## Joining with `by = join_by(word)`
```

```
## # A tibble: 668 x 2
##   word      n
##   <chr>   <int>
## 1 well    401
## 2 good    359
## 3 great   264
## 4 like    200
## 5 better  173
## 6 enough  129
## 7 happy   125
## 8 love    117
## 9 pleasure 115
## 10 right   92
## # i 658 more rows
```

```
bing <- get_sentiments("bing")
```

```
Emma_sentiment <- tidy_data %>%
  inner_join(bing) %>%
  count(book = "Emma" , index = linenumber %/% 80, sentiment) %>%
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative)
```

```
## Joining with `by = join_by(word)`
```

```
## Warning in inner_join(., bing): Detected an unexpected many-to-many relationship between `x` and `y`.
## i Row 435434 of `x` matches multiple rows in `y`.
## i Row 5051 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =`
##   "many-to-many" to silence this warning.
```

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



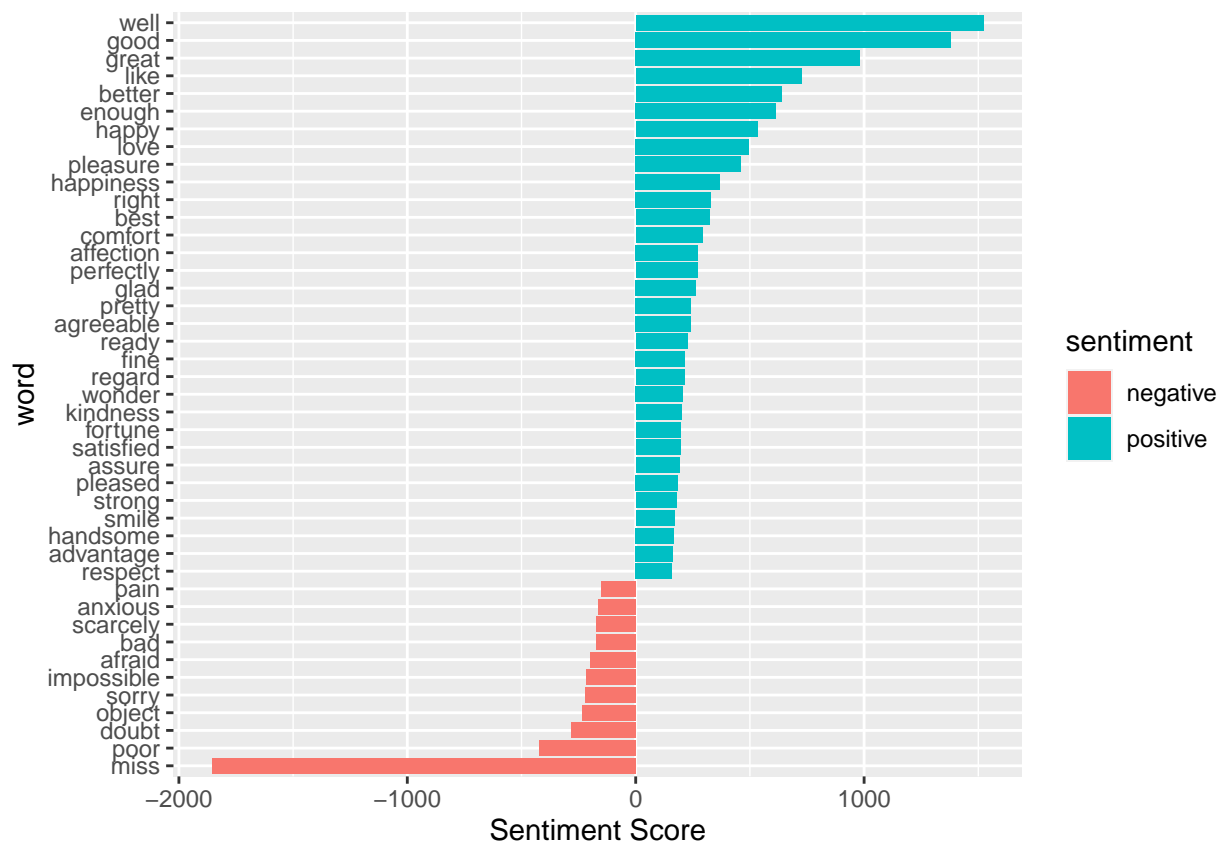
```
counting_words <- tidy_data %>%
  inner_join(bing, relationship = "many-to-many") %>%
  count(word, sentiment, sort = TRUE)
```

```
## Joining with `by = join_by(word)`
```

```
head(counting_words)
```

```
## # A tibble: 6 x 3
##   word      sentiment      n
##   <chr>    <chr>      <int>
## 1 miss     negative    1855
## 2 well     positive    1523
## 3 good     positive    1380
## 4 great    positive     981
## 5 like     positive     725
## 6 better   positive     639
```

```
counting_words %>%
  filter(n > 150) %>%
  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_data %>%
  inner_join(bing, relationship = "many-to-many") %>%
  count(word, sentiment, sort = TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("red", "dark green"),
    max.words = 100)
```

```
## Joining with `by = join_by(word)`
```

negative



positive