sentiments\_rmd

Adarsh

30/07/2021

library(tidytext)  
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   
## # ... with 6,776 more rows

library("janeaustenr")  
library("stringr")  
library("dplyr")

##   
## 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

library("RMySQL")

## Loading required package: DBI

sentiments\_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)

positive\_senti <- get\_sentiments("bing") %>%  
 filter(sentiment == "positive")  
negative\_senti<- get\_sentiments("bing") %>%  
 filter(sentiment == "negative")

sentiments\_data %>%  
 filter(book == "Emma") %>%  
 semi\_join(positive\_senti) %>%  
 count(word, sort = TRUE)

## Joining, 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  
## # ... with 658 more rows

sentiments\_data %>%  
 filter(book == "Pride & Prejudice") %>%  
 semi\_join(negative\_senti) %>%  
 count(word, sort = TRUE)

## Joining, by = "word"

## # A tibble: 838 x 2  
## word n  
## <chr> <int>  
## 1 miss 283  
## 2 object 48  
## 3 scarcely 45  
## 4 impossible 44  
## 5 poor 38  
## 6 afraid 37  
## 7 doubt 37  
## 8 sorry 34  
## 9 lost 29  
## 10 pain 28  
## # ... with 828 more rows

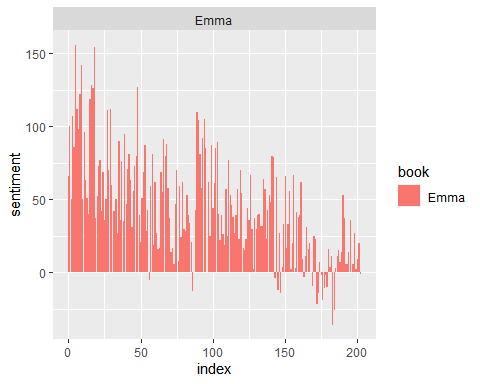
library("tidyr")  
bing <- get\_sentiments("bing")  
Emma\_sentiment <- sentiments\_data %>%  
 inner\_join(bing) %>%  
 count(book = "Emma" , index = linenumber %/% 80, sentiment) %>%  
 spread(sentiment, n, fill = 0) %>%  
 mutate(sentiment = positive - negative)

## Joining, by = "word"

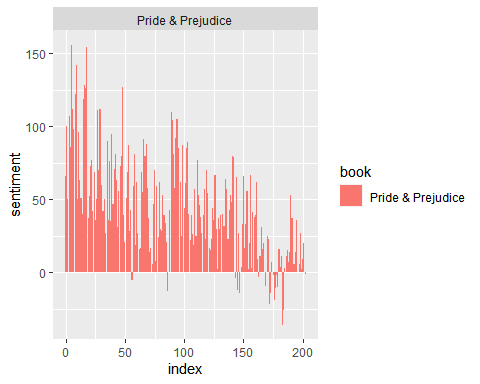
bing <- get\_sentiments("bing")  
Pride\_sentiment <- sentiments\_data %>%  
 inner\_join(bing) %>%  
 count(book = "Pride & Prejudice" , index = linenumber %/% 80, sentiment) %>%  
 spread(sentiment, n, fill = 0) %>%  
 mutate(sentiment = positive - negative)

## Joining, by = "word"

library(ggplot2)  
ggplot(Emma\_sentiment, aes(index, sentiment, fill = book)) +  
 geom\_bar(stat = "identity", show.legend = TRUE) +  
 facet\_wrap(~book, ncol = 2, scales = "free\_x")



ggplot(Pride\_sentiment, aes(index, sentiment, fill = book)) +  
 geom\_bar(stat = "identity", show.legend = TRUE) +  
 facet\_wrap(~book, ncol = 2, scales = "free\_x")



counting\_words <- sentiments\_data %>%  
 inner\_join(bing) %>%  
 count(word, sentiment, sort = TRUE)

## Joining, 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")

