KSooklall_Homework10 DATA 607

Kenan Sooklall

4/16/2021

The code in this report are taken from the textbook: Text Mining with R Chapter 2 - Sentiment analysis with tidy data

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

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
library(wordcloud)
```

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

Load sentiment data sets

```r
get_sentiments("afinn")
```

```
## # A tibble: 2,477 x 2
##    word       value
##    <chr>      <dbl>
##  1 abandon       -2
##  2 abandoned     -2
##  3 abandons      -2
##  4 abducted      -2
##  5 abduction     -2
##  6 abductions    -2
##  7 abhor         -3
##  8 abhorred      -3
##  9 abhorrent     -3
## 10 abhors        -3
## # ... with 2,467 more rows
```

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

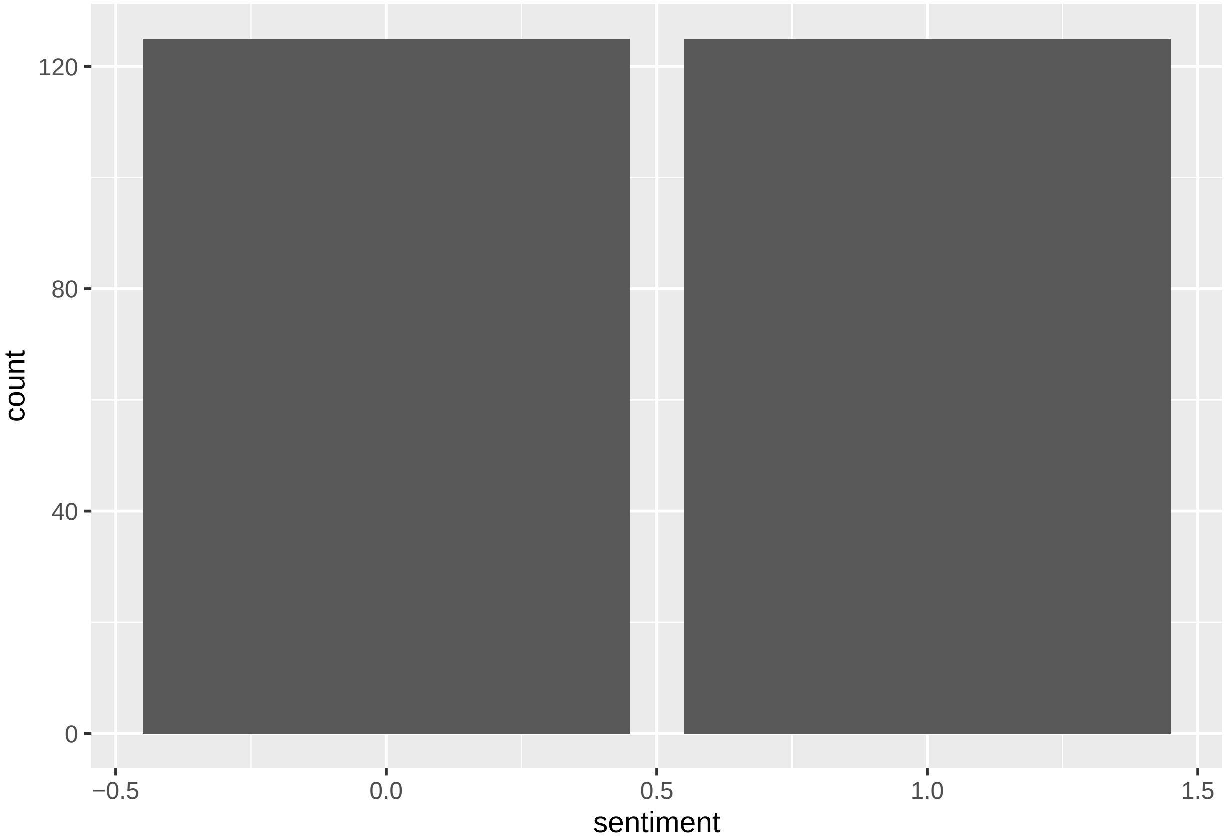
```
## # A tibble: 6,786 x 2
##    word        sentiment
##    <chr>       <chr>
##  1 2-faces     negative
##  2 abnormal    negative
##  3 abolish     negative
##  4 abominable  negative
##  5 abominably  negative
##  6 abominate   negative
##  7 abomination negative
##  8 abort       negative
## 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

```
get_sentiments("nrc")
```

```
## # A tibble: 13,901 x 2
##    word        sentiment
##    <chr>       <chr>
##  1 abacus      trust
##  2 abandon     fear
##  3 abandon     negative
##  4 abandon     sadness
##  5 abandoned   anger
##  6 abandoned   fear
##  7 abandoned   negative
##  8 abandoned   sadness
##  9 abandonment anger
## 10 abandonment fear
## # ... with 13,891 more rows
```

Read in the data and plot the ratio of good vs bad sentiment

```
df = read.csv('https://raw.githubusercontent.com/ksooklall/CUNY-SPS-Masters-DS/main/DATA_607/homework/h
df %>% ggplot(aes(x=sentiment)) + geom_bar()
```



Unnest the review column and remove stop_words. Stop_words are words that carry no sentiment like [i, the, a, able, about …]

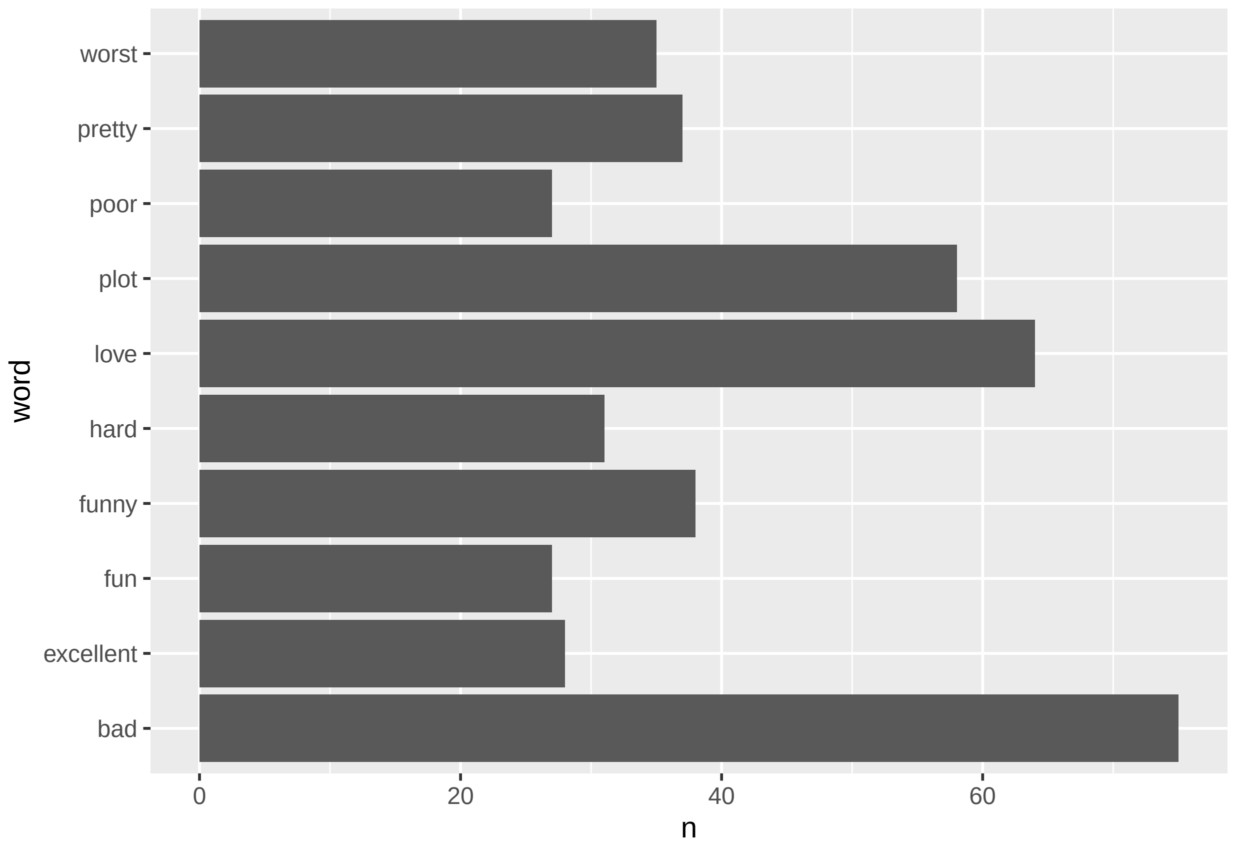
```
text <- df %>% mutate(linenum = row_number()) %>% unnest_tokens(word, review) %>% anti_join(stop_words)
```

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

View the distribution of different sentiment words

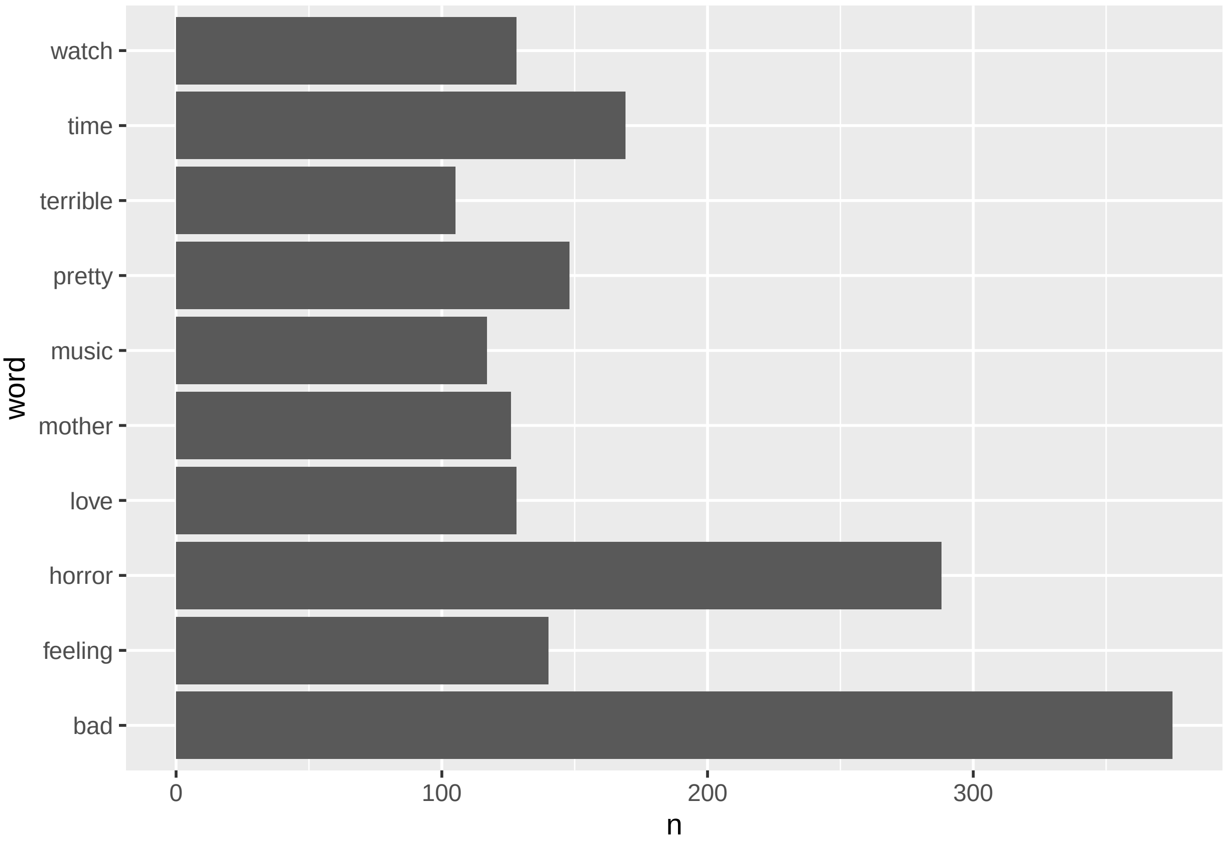
```
text %>% select(word) %>%
    inner_join(get_sentiments("bing")) %>%
    count(word, sort = TRUE) %>% top_n(n, n=10) %>% ggplot(aes(x=word, y=n)) + geom_col() + coord_flip(
```

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



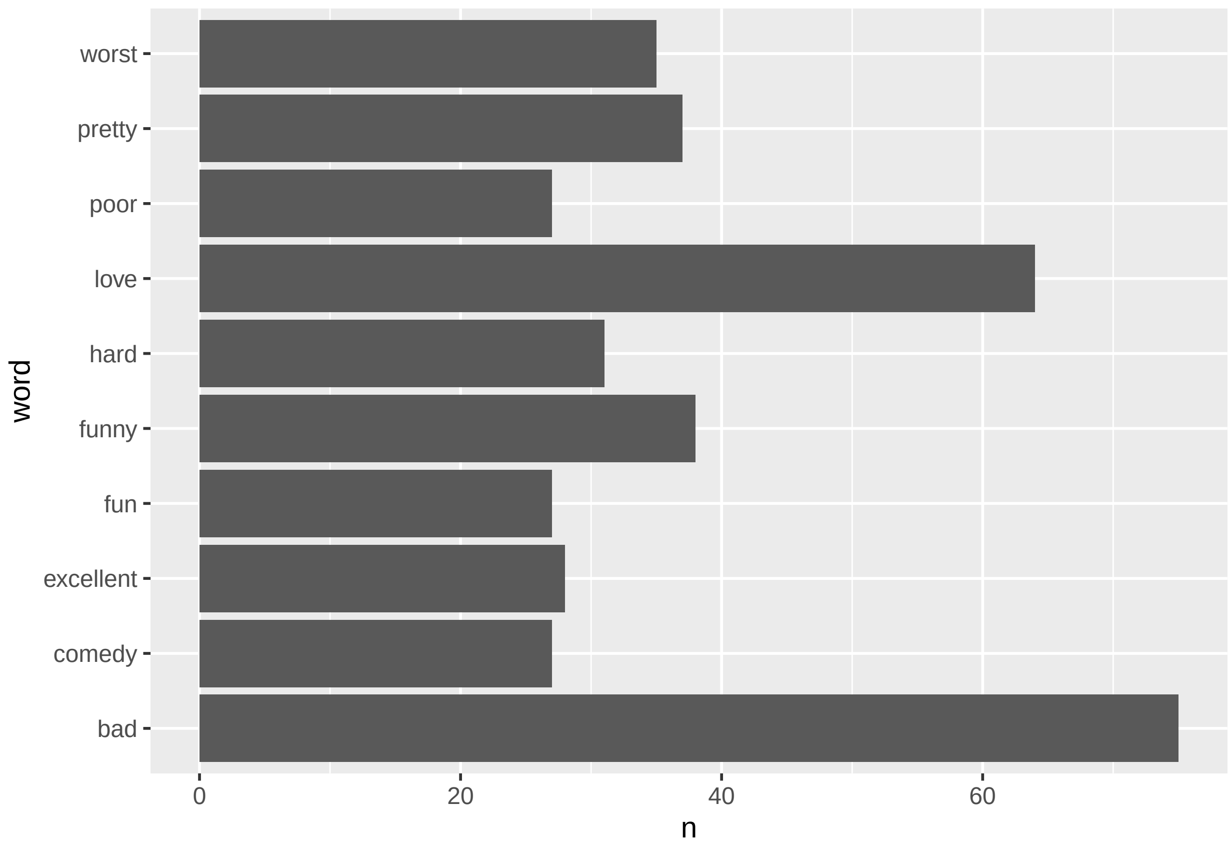
```
text %>% select(word) %>%
    inner_join(get_sentiments("nrc")) %>%
    count(word, sort = TRUE) %>% top_n(n, n=10) %>% ggplot(aes(x=word, y=n)) + geom_col() + coord_flip(
```

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



```
text %>% select(word) %>%
    inner_join(get_sentiments("afinn")) %>%
    count(word, sort = TRUE) %>% top_n(n, n=10) %>% ggplot(aes(x=word, y=n)) + geom_col() + coord_flip(
```

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



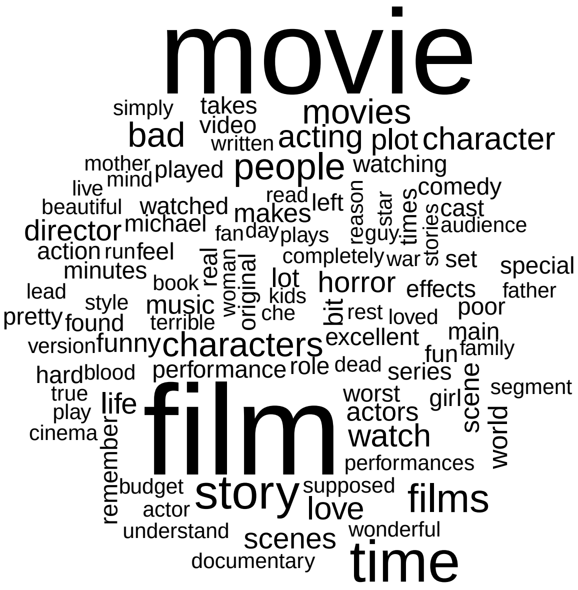
Word clouds

```
text %>% pull(word) %>% wordcloud(min.freq = 10, max.word=100)
```

```
## Loading required namespace: tm
```

```
## Warning in tm_map.SimpleCorpus(corpus, tm::removePunctuation): transformation
## drops documents
```

```
## Warning in tm_map.SimpleCorpus(corpus, function(x) tm::removeWords(x,
## tm::stopwords())): transformation drops documents
```

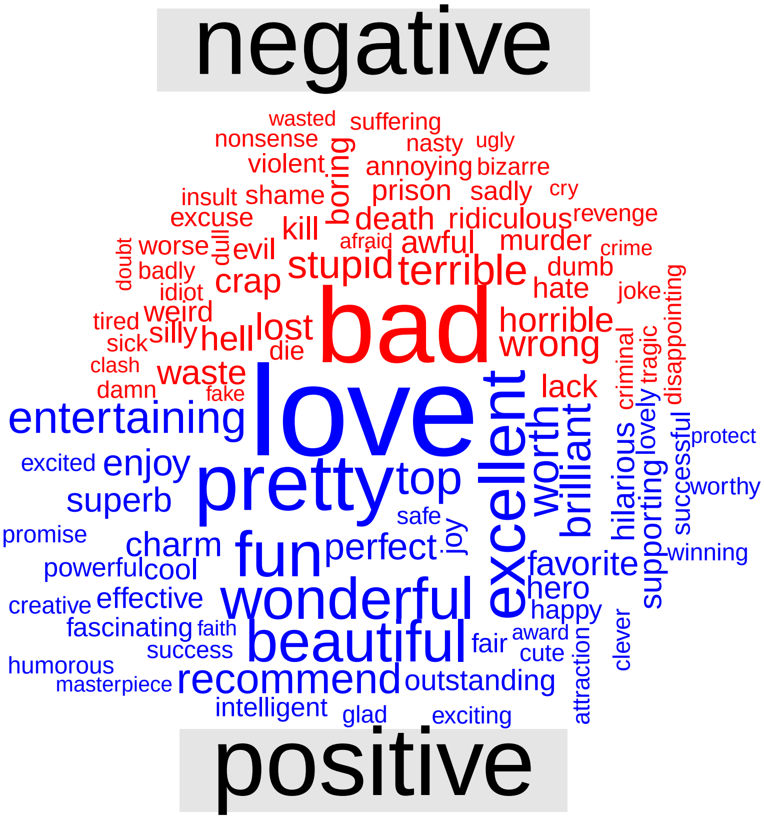


```
library(reshape2)
```

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

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

```
text %>% select(word) %>% inner_join(get_sentiments("afinn")) %>%inner_join(get_sentiments("nrc")) %>%i
  comparison.cloud(colors = c("red", "blue"),
                   max.words = 100)
```

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

## Joining, by = "word"

## Joining, by = c("word", "sentiment")
```


negative
wasted
suffering
nonsense
nasty
ugly
violent
boring
annoying
bizarre
insult
shame
prison
sadly
cry
excuse
kill
death
ridiculous
revenge
doubt
worse
dull
evil
afraid
awful
murder
crime
badly
stupid
terrible
dumb
idiot
crap
hate
joke
disappointing
tired
weird
bad
horrible
sick
silly
hell
lost
wrong
criminal
tragic
clash
die
waste
lack
damn
fake
entertaining
love
excellent
worth
brilliant
hilarious
lovely
successful
protect
excited
enjoy
pretty
top
supporting
worthy
superb
safe
joy
promise
charm
fun
perfect
winning
powerful
cool
favorite
creative
effective
wonderful
hero
happy
fascinating
faith
beautiful
clever
success
fair
award
cute
attraction
humorous
masterpiece
recommend
outstanding
intelligent
glad
exciting
positive