Animal Crossing Text Analysis

Josh Kong

8/10/2020

Loading the data

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------- t

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.4
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

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

```
library(tidytext)
critic <- readr::read_tsv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/20
```

```
## Parsed with column specification:
## cols(
##   grade = col_double(),
##   publication = col_character(),
##   text = col_character(),
##   date = col_date(format = "")
## )
```

```
user_reviews <- readr::read_tsv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/da
```

```
## Parsed with column specification:
## cols(
##   grade = col_double(),
##   user_name = col_character(),
##   text = col_character(),
##   date = col_date(format = "")
## )
```

Exploring the critic data

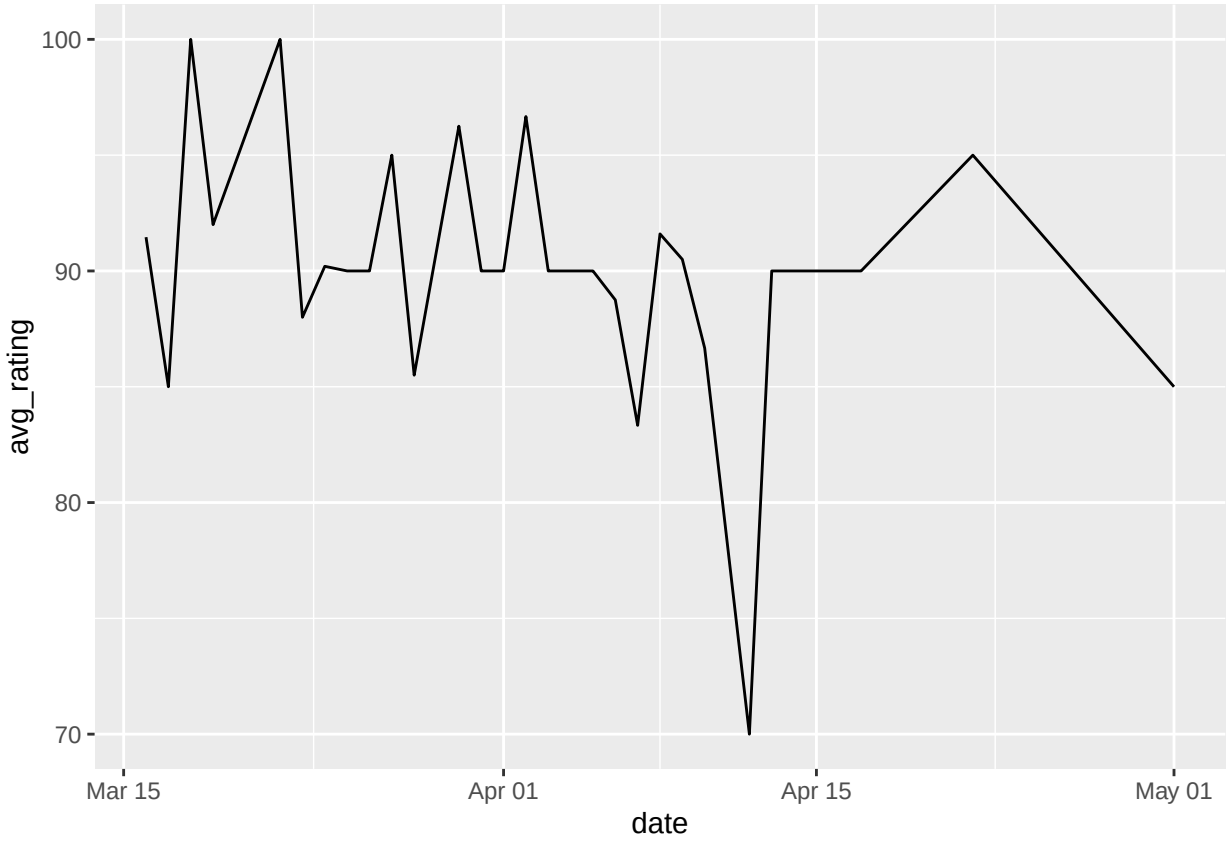
```r
head(critic)
```

```
## # A tibble: 6 x 4
##   grade publication    text                                              date
##   <dbl> <chr>          <chr>                                             <date>
## 1   100 Pocket Gamer ~ Animal Crossing; New Horizons, much like its ~ 2020-03-16
## 2   100 Forbes         Know that if you're overwhelmed with the worl~ 2020-03-16
## 3   100 Telegraph      With a game this broad and lengthy, there's m~ 2020-03-16
## 4   100 VG247          Animal Crossing: New Horizons is everything I~ 2020-03-16
## 5   100 Nintendo Insi~ Above all else, Animal Crossing: New Horizons~ 2020-03-16
## 6   100 Trusted Revie~ Animal Crossing: New Horizons is the best gam~ 2020-03-16
```

```r
critic %>%
  summarise(avg = mean(grade))
```

```
## # A tibble: 1 x 1
##     avg
##   <dbl>
## 1  90.6
```

```r
#seems like critics really enjoyed this game

critic %>%
  group_by(date) %>%
  summarise(avg_rating = mean(grade)) %>%
  ungroup() %>%
  ggplot(aes(date,avg_rating)) +
  geom_line()
```




```
#how the ratings changed as the dates changed


critic_unnested <- critic %>%
  unnest_tokens(word,text) %>%
  add_count(publication) %>%
  anti_join(stop_words, by = "word")

#taking a look at the most viewed words
critic_unnested %>%
  count(word, sort = TRUE)
```

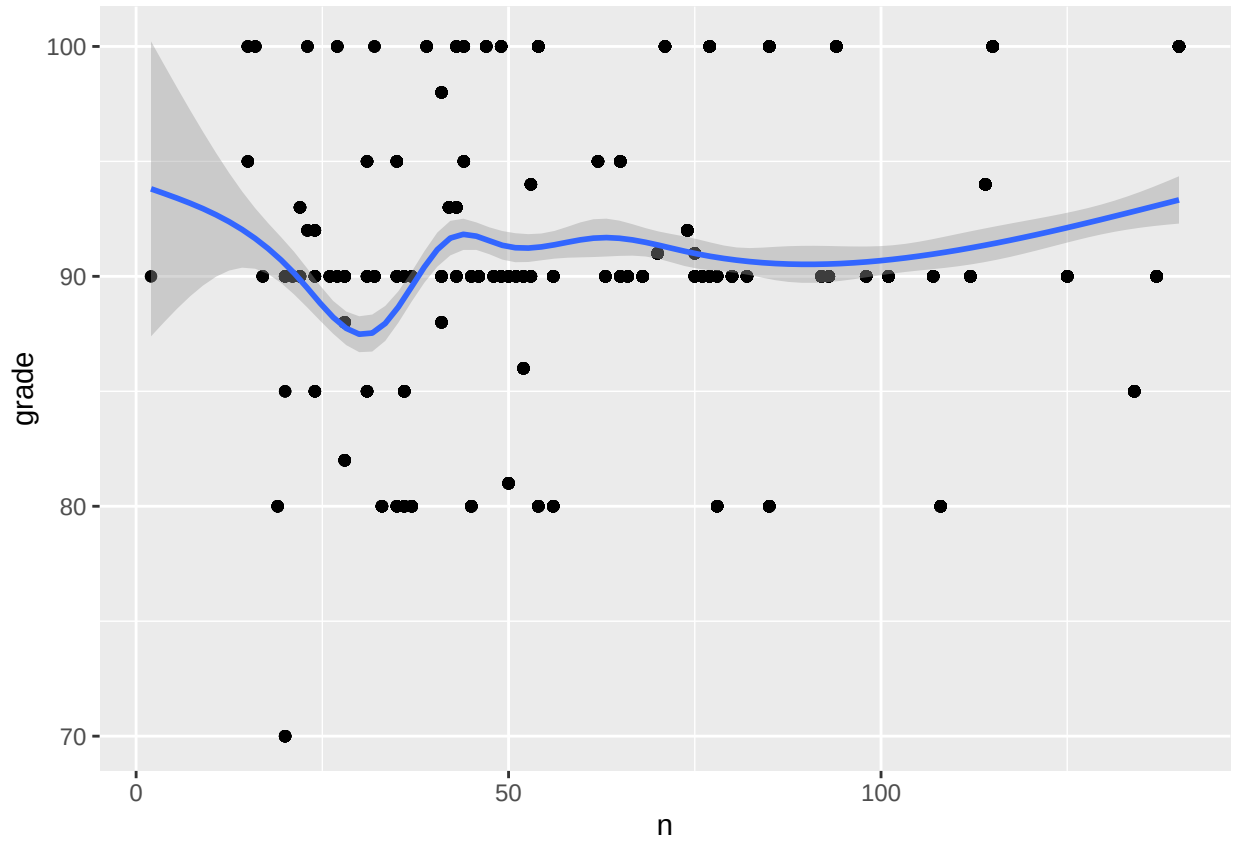
```
## # A tibble: 1,061 x 2
##    word           n
##    <chr>      <int>
##  1 animal       111
##  2 crossing     109
##  3 horizons      94
##  4 game          61
##  5 series        35
##  6 island        31
##  7 time          26
##  8 nintendo      22
##  9 life          19
## 10 experience    18
## # ... with 1,051 more rows
```

```r
#taking a look at which of the words had the most positive impact on the game
critic_unnested %>%
  group_by(word) %>%
  summarise(count = n(),
            avg_rating = mean(grade)) %>%
  filter(count > 5) %>%
  arrange(desc(avg_rating))
```

```
## # A tibble: 45 x 3
##    word     count avg_rating
##    <chr>    <int>      <dbl>
##  1 player       6       95
##  2 world       15       94.1
##  3 life        19       93.8
##  4 feel         8       93.8
##  5 perfect     13       93.5
##  6 home         6       93.3
##  7 players     10       93.2
##  8 games       13       92.8
##  9 day         16       92.6
## 10 nintendo    22       92.5
## # ... with 35 more rows
```

```r
#is there a correlation between the length of the review and the rating?
critic_unnested %>%
  group_by(publication) %>%
  ggplot(aes(n,grade)) +
  geom_point() +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



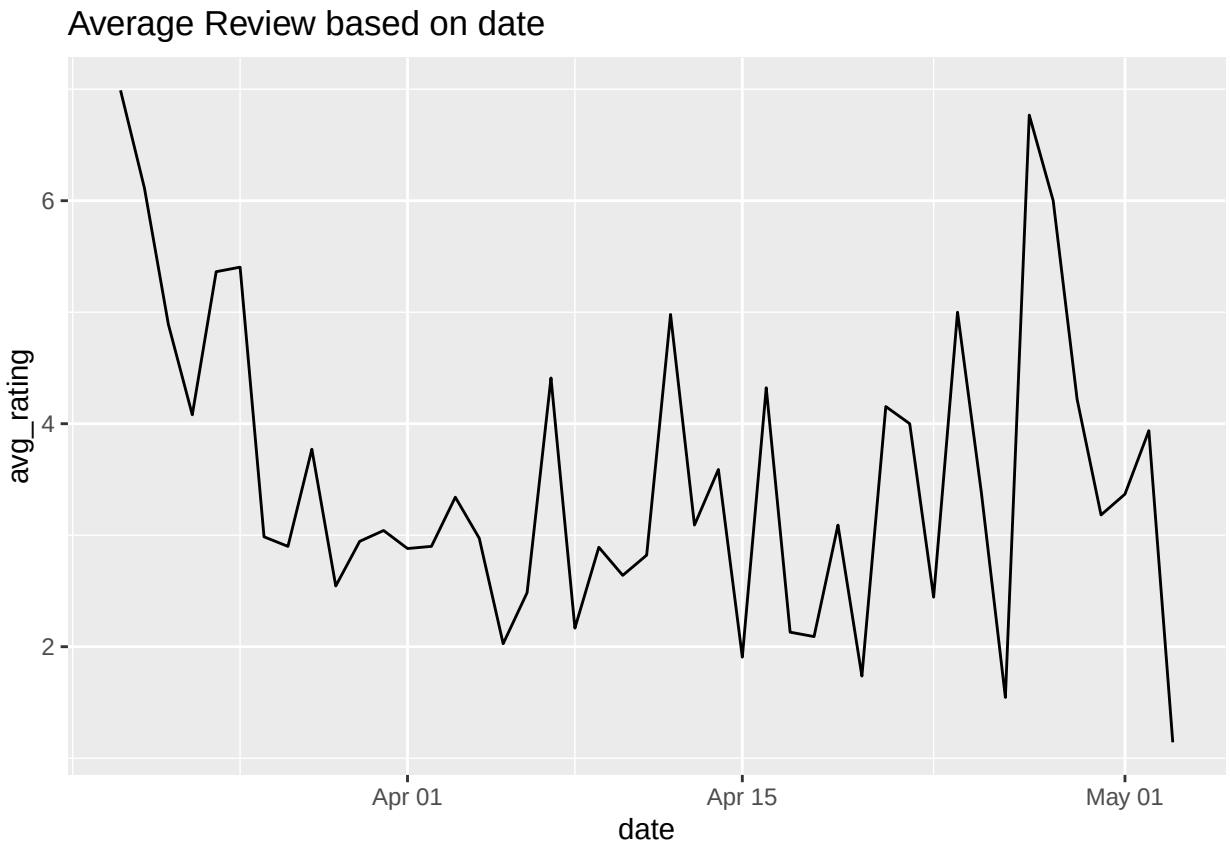

```
#there doesn't seem to be a correlation between length of the review and the grade
```

Analyzing User Reviews

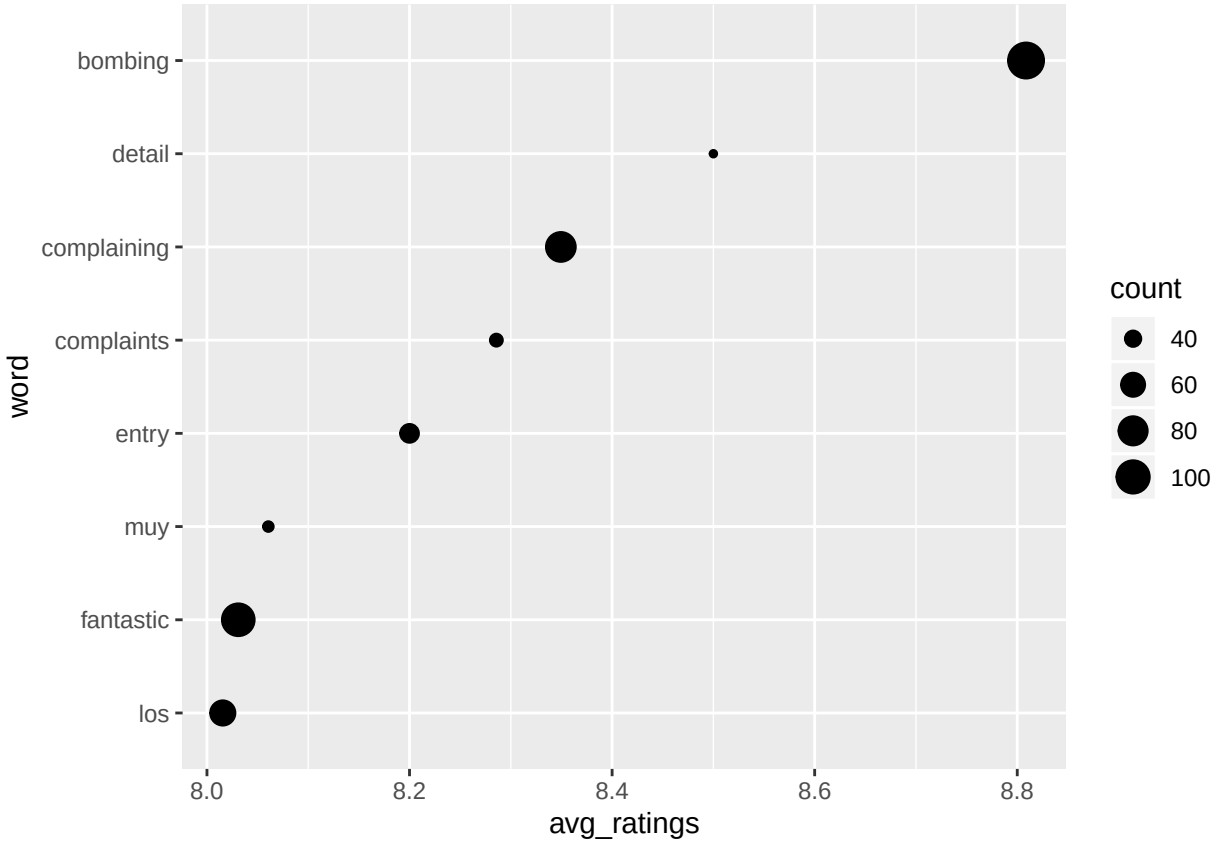
```
user_reviews %>%
  summarise(avg = mean(grade))
```

```
## # A tibble: 1 x 1
##     avg
##   <dbl>
## 1  4.22
```

```
#the average rating the users gave is a lot lower than the critics


user_reviews %>%
  group_by(date) %>%
  summarise(avg_rating = mean(grade)) %>%
  ungroup() %>%
  ggplot(aes(date,avg_rating)) +
  geom_line()+
  labs(title = "Average Review based on date")
```



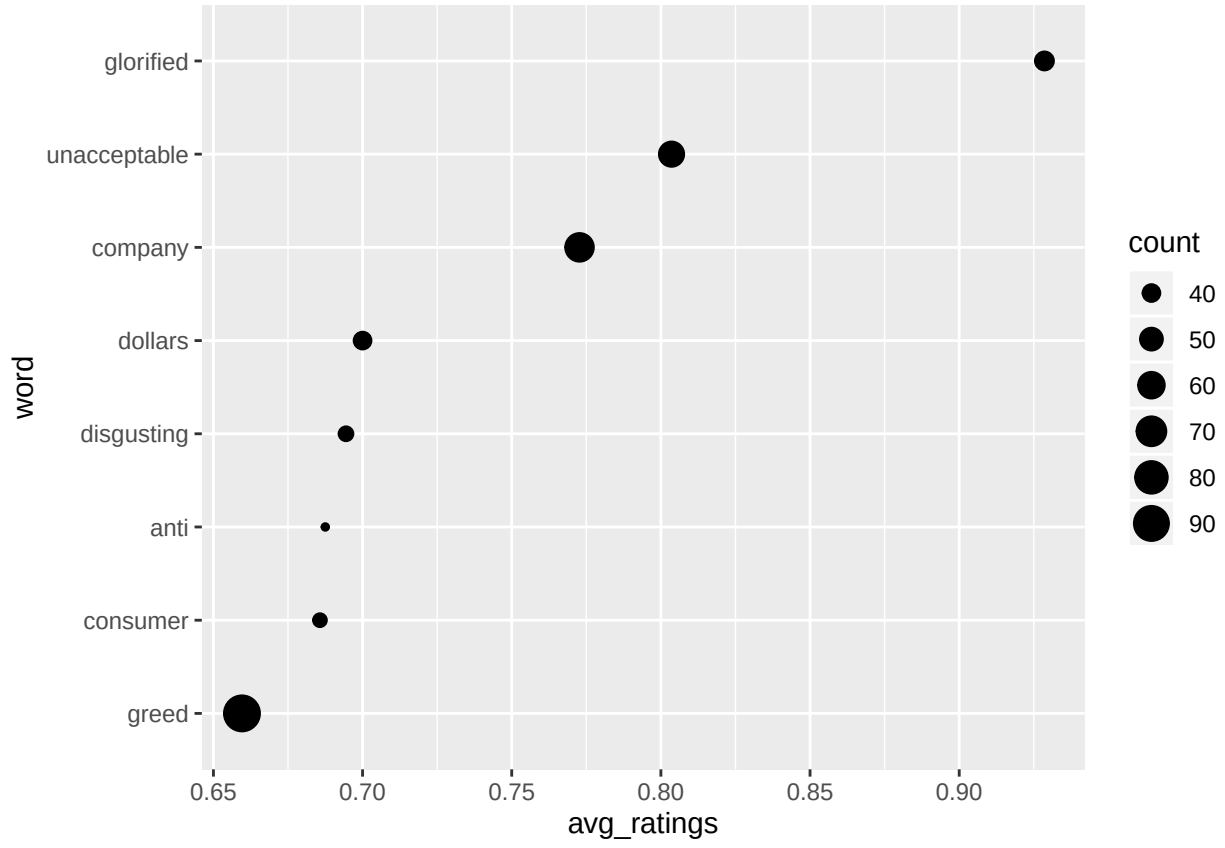

```
user_unnested <- user_reviews %>%
  unnest_tokens(word,text) %>%
  add_count(user_name) %>%
  anti_join(stop_words, by = "word") %>%
  filter(n > 30)

#which words were correlated with a positive impact on the game?
user_unnested %>%
  group_by(word) %>%
  summarise(count = n(),
            avg_ratings = mean(grade)) %>%
  filter(count > 30) %>%
  arrange(desc(avg_ratings)) %>%
  head(8) %>%
  mutate(word = fct_reorder(word,avg_ratings)) %>%
  ggplot(aes(word, avg_ratings, size = count)) +
  geom_point()+
  coord_flip()
```




```
#the word bombing is highly correlated with a positive rating. let's take a closer look at that
user_bombing <- user_unnested %>%
  filter(word == "bombing") %>%
  count(user_name, sort = TRUE) %>%
  pull(user_name)

user_reviews %>%
  filter(user_name %in% user_bombing)
```

```
## # A tibble: 57 x 4
##    grade user_name    text                                                  date
##    <dbl> <chr>        <chr>                                                 <date>
##  1    10 CDizzness    Best Animal Crossing game hands down, graphics~ 2020-03-20
##  2    10 Sorites      I haven't played animal crossing since Wild Wo~ 2020-03-20
##  3    10 Earthquakem~ Pretty fun game. Relaxing, adorable, and simpl~ 2020-03-21
##  4    10 vgriscom     Soo much fun!! People arent rating the game fa~ 2020-03-24
##  5     9 E_Z_E        This is a very fun game if you are playing sin~ 2020-03-24
##  6     9 Oldskoolberk I know about the rating bombing of Animal Cros~ 2020-03-24
##  7    10 FantasyHarm~ I've been playing Animal Crossing since the fi~ 2020-03-24
##  8    10 ACNH         Don't trust the people review bombing this gam~ 2020-03-24
##  9    10 Jadedflames  I'd probably normally give it a 9, since I res~ 2020-03-24
## 10    10 cbland       Absolutely wonderful to play. Anyone who is ne~ 2020-03-24
## # ... with 47 more rows
```

```
#people are usually saying "don't worry about the review bombing" which is something that seems to be h


#which words are correlated with a low review in the game
user_unnested %>%
  group_by(word) %>%
  summarise(count = n(),
            avg_ratings = mean(grade)) %>%
  filter(count > 30) %>%
  arrange(avg_ratings) %>%
  head(8) %>%
  mutate(word = fct_reorder(word,avg_ratings)) %>%
  ggplot(aes(word, avg_ratings, size = count)) +
  geom_point()+
  coord_flip()
```



Lasso Regression model

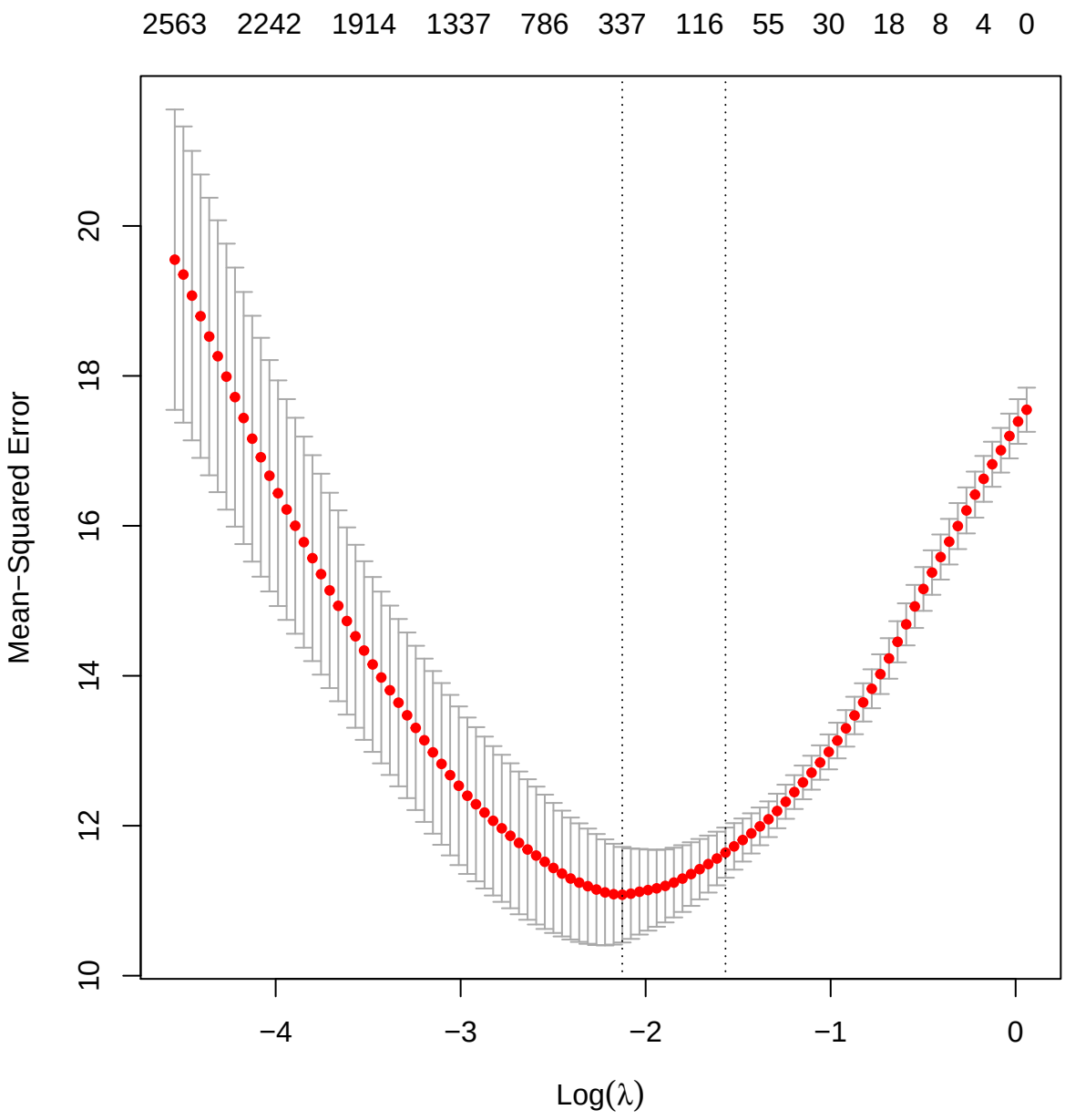
```
library(glmnet)
```

```
## Loading required package: Matrix
```

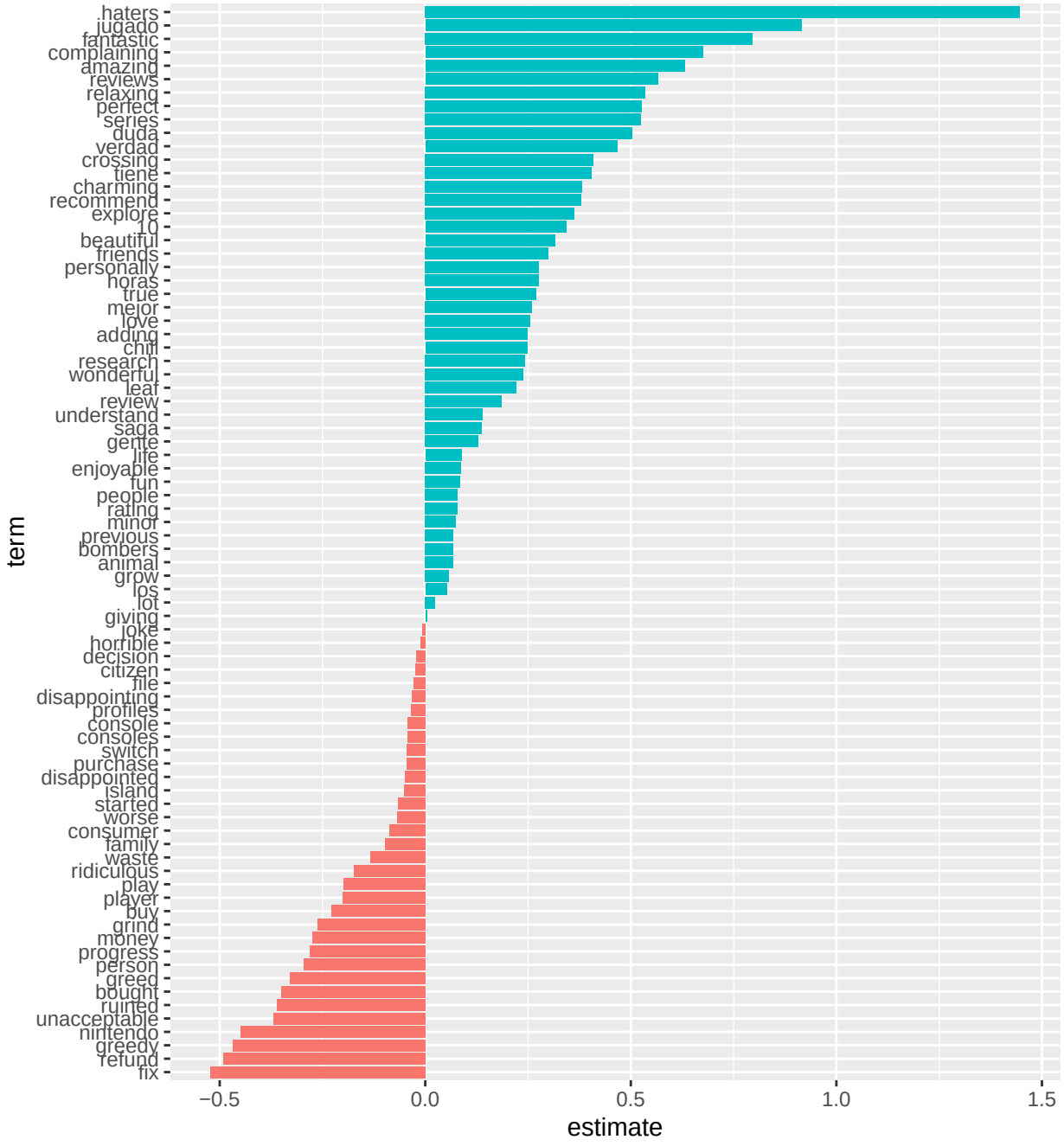
```
##
```

```
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

## Loaded glmnet 3.0-2
```

```
library(Matrix)
library(broom)
text_matrix <- user_unnested %>%
  count(user_name, word) %>%
  cast_sparse(user_name,word,n)

grade <- user_reviews$grade[match(rownames(text_matrix),user_reviews$user_name)]
lasso_model <- cv.glmnet(text_matrix,grade)
plot(lasso_model)
```




```
tidy(lasso_model$glmnet.fit) %>%
  filter(lambda == lasso_model$lambda.1se,
         term != "(Intercept)") %>%
  mutate(term = fct_reorder(term,estimate)) %>%
  ggplot(aes(term,estimate,fill = estimate > 0))+
  geom_col() +
  coord_flip()+
  theme(legend.position = "none")
```


haters
jugado
fantastic
complaining
amazing
reviews
relaxing
perfect
series
duda
verdad
crossing
tiene
charming
recommend
explore
10
beautiful
friends
personally
horas
true
mejor
love
adding
chill
research
wonderful
leaf
review
understand
saga
genre
life
enjoyable
fun
people
rating
minor
previous
bombers
animal
grow
lps
lot
giving
joke
horrible
decision
citizen
file
disappointing
profiles
console
consoles
switch
purchase
disappointed
island
started
worse
consumer
family
waste
ridiculous
play
player
buy
grind
money
progress
person
greed
bought
ruined
unacceptable
nintendo
greedy
refund
fix
term
−0.5
0.0
0.5
1.0
1.5
estimate