Taras Courmet Carmels

David Jackson

2020-01-17

YouTube Source: https://youtu.be/eQU8Zd1B9tM

Import and Clean Amazon Product Reviews

- Change Log:
- Imported Amazon review data
- Cleaned up column names
- Removed unsed columns
- Extracted date from date column.

Tara's All Natural Handcrafted Gourmet Sea Salt Caramel

- OLD FASHIONED CRAFT QUALITY – Tara's caramels are kettle cooked in small batches from a century old recipe. Grandma's were so good, we didn't change them.
- SIMPLE, PREMIUM NATURAL INGREDIENTS – We use real butter, fresh heavy cream, and Madagascar vanilla. We use Grandma's same recipe from 100 years ago
- NATURALLY RICH CARAMEL FLAVOR – Tara's handmade gourmet caramels have the homemade taste and color you can only get from carefully crafting in small batches. No artificial flavors or colors added – GLUTEN FREE and KOSHER
- SOFT, CREAMY, BUTTERY TEXTURE – Individually wrapped high quality caramel means a soft bite that doesn't stick to your teeth
- INGREDIENTS - Corn syrup, sweetened condensed milk, heavy cream, invert sugar, butter, brown sugar, sugar, natural flavor, sea salt, soy lecithin (http:// Amazon.com)

```
taras_caramels <- read_csv("./taras_caramels.csv") %>%
  janitor::clean_names() %>% select(-id,-profile_name,-images)
taras_caramels$text <- tolower(taras_caramels$text)
```

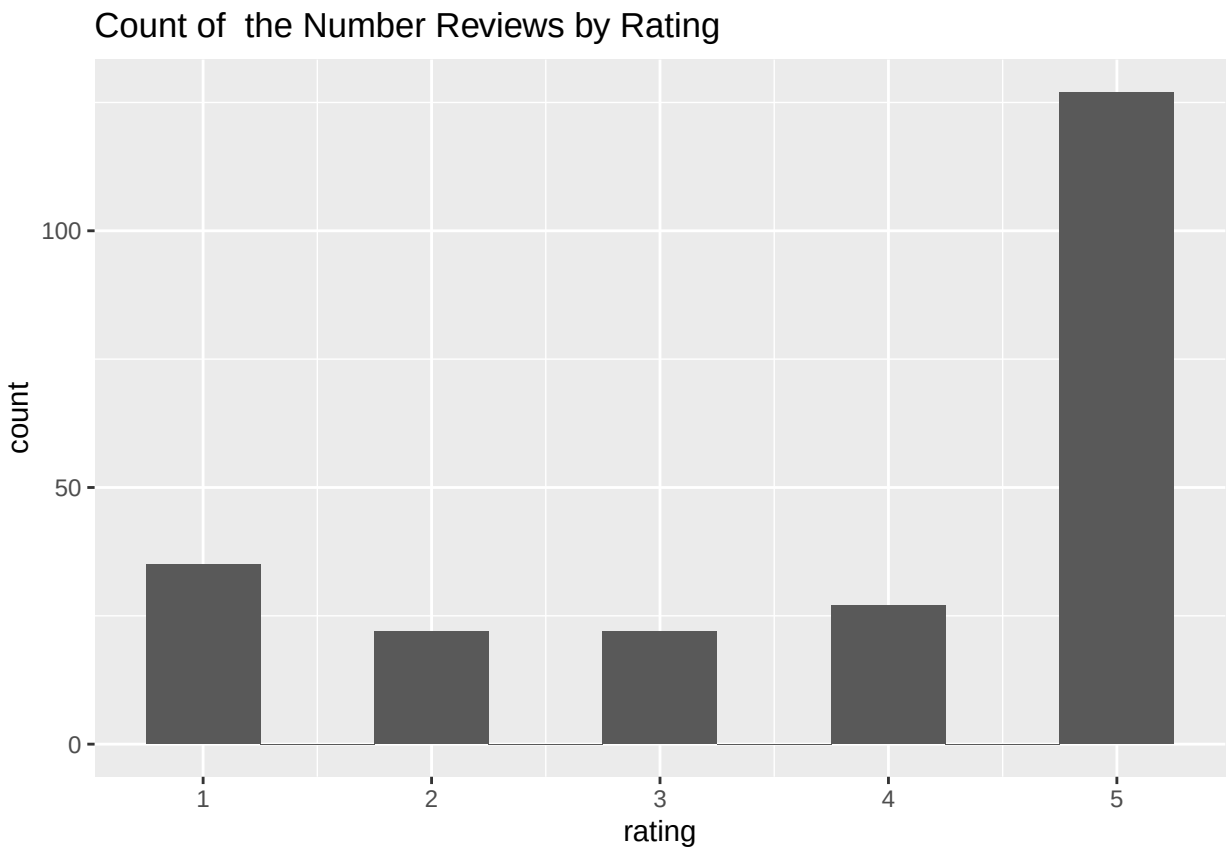
```
# Extract Dates from test column
taras_caramels$Ymd <- lubridate::mdy(taras_caramels$date)
taras_caramels <- taras_caramels %>% select(Ymd,title:helpful)  %>%
  rename(review = title)
```

```
summary(taras_caramels)
```

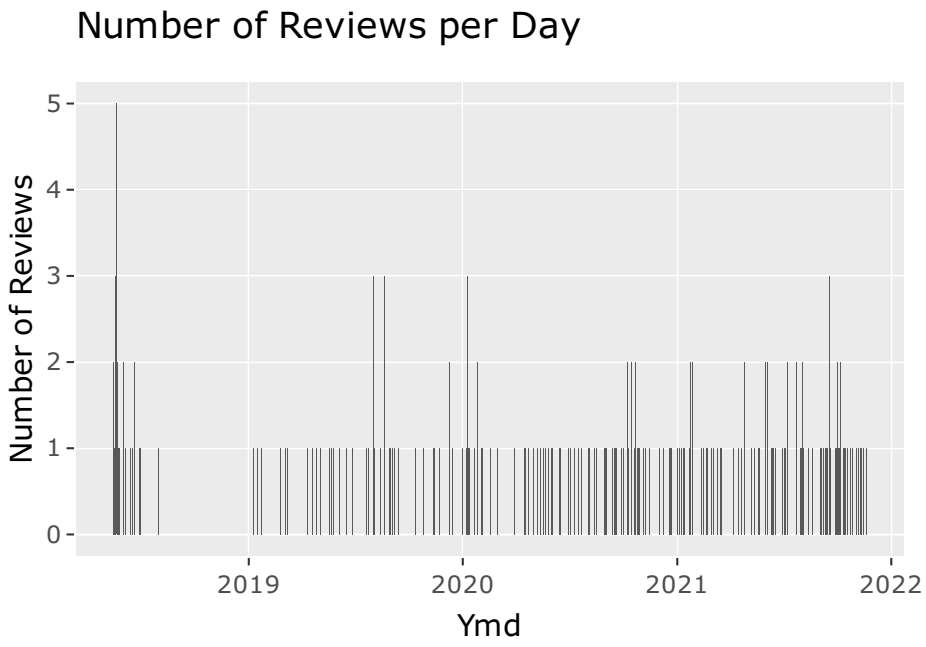
```
##       Ymd                review             rating         helpful
##  Min.   :2018-05-16   Length:233         Min.   :1.000   Min.   :  0.0000
##  1st Qu.:2019-09-04   Class :character   1st Qu.:3.000   1st Qu.:  0.0000
##  Median :2020-09-28   Mode  :character   Median :5.000   Median :  0.0000
```

```
##  Mean   :2020-06-27                    Mean   :3.811   Mean   :  0.9828
##  3rd Qu.:2021-06-01                    3rd Qu.:5.000   3rd Qu.:  0.0000
##  Max.   :2021-11-19                    Max.   :5.000   Max.   :118.0000
```

```
ggplot(taras_caramels) + geom_histogram(aes(x=rating),bins = 9) +
  labs(title = "Count of  the Number Reviews by Rating")
```




```
p1 <-taras_caramels %>% count(Ymd) %>% ggplot() + geom_col(aes(x=Ymd,y=n)) +
  labs(title = "Number of Reviews per Day",y="Number of Reviews")
ggplotly(p1)
```



Analyize Words Using "sentimentr" Package

```
head(sentiment(taras_caramels$review),25)
```

Commands from Video (see above)

```
##     element_id sentence_id word_count   sentiment
##  1:          1           1          4 -0.25000000
##  2:          1           2          8  0.35355339
##  3:          1           3          5  0.44721360
##  4:          2           1          5  0.22360680
##  5:          3           1         11  0.00000000
##  6:          4           1         10  0.04743416
##  7:          5           1          9  0.61683333
##  8:          6           1          2  0.53033009
##  9:          7           1          3  0.28867513
## 10:          8           1          4  0.30000000
## 11:          9           1          1  1.00000000
## 12:         10           1          5 -0.50311529
## 13:         11           1          6  0.36742346
## 14:         12           1          1 -0.50000000
## 15:         12           2          3 -0.57735027
```

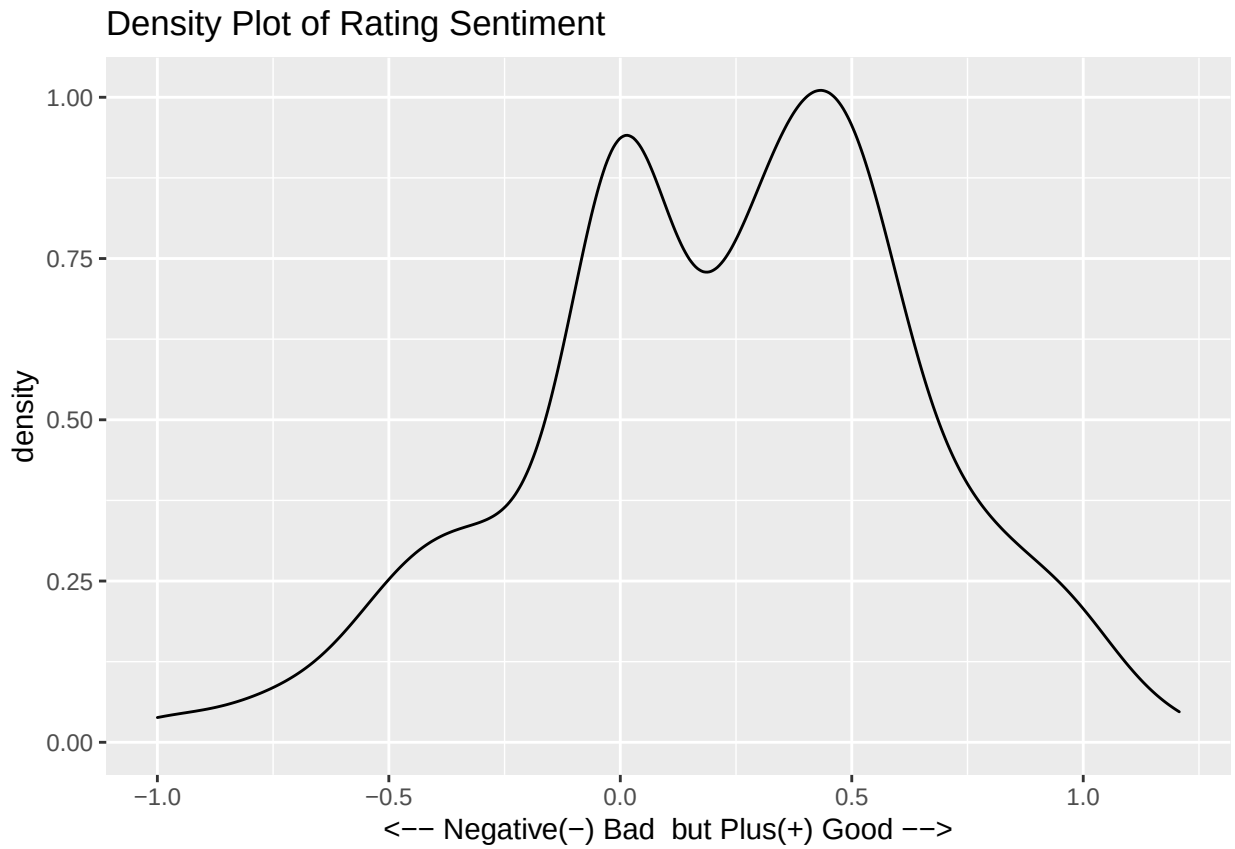
```
## 16:          13           1          1  0.00000000
## 17:          13           2          3  0.00000000
## 18:          14           1          1  0.50000000
## 19:          15           1          3  0.43301270
## 20:          16           1          4 -0.50000000
## 21:          17           1          1  0.50000000
## 22:          17           2          1  0.00000000
## 23:          17           3          1  0.00000000
## 24:          18           1          6  0.00000000
## 25:          19           1          6 -0.30618622
##     element_id sentence_id word_count   sentiment
```

```
sentiment_by(taras_caramels$review)
```

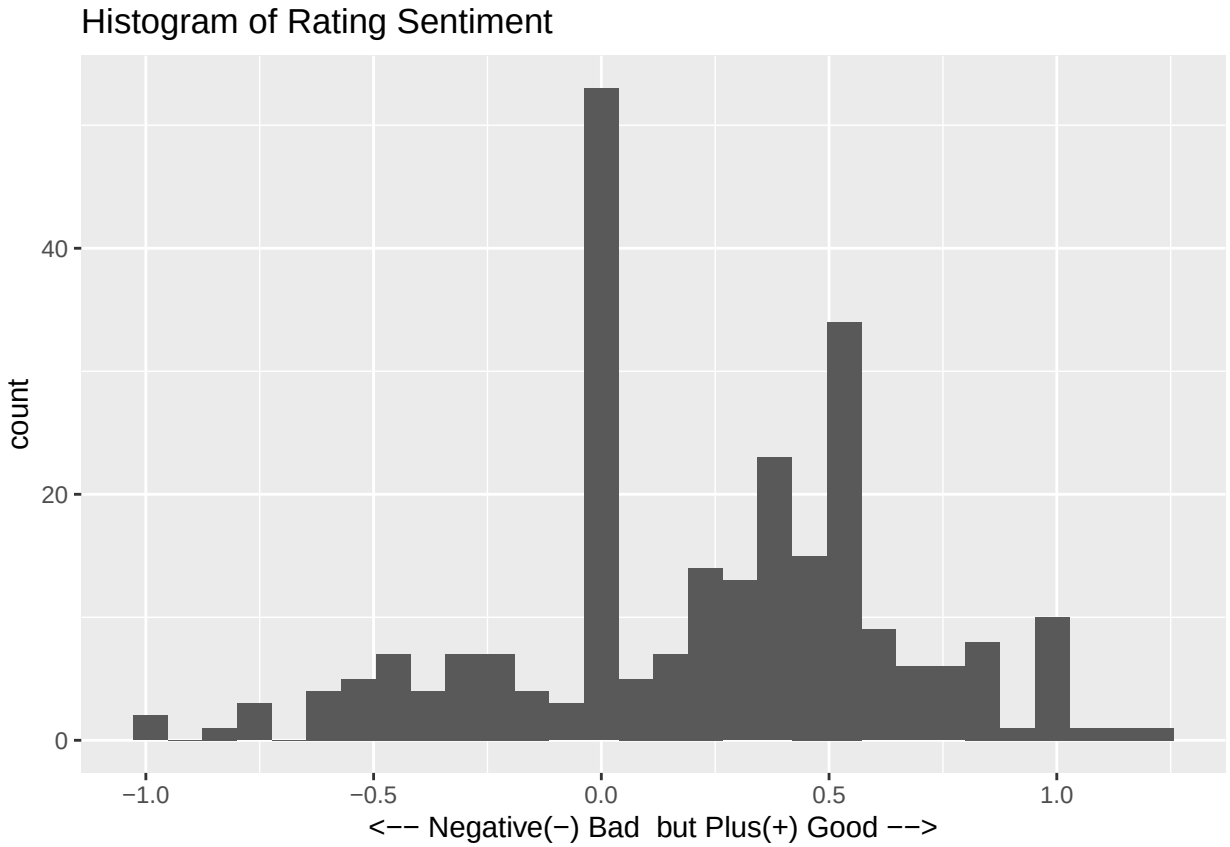
```
##      element_id word_count       sd ave_sentiment
##   1:          1         17 0.378408    0.18358900
##   2:          2          5       NA    0.22360680
##   3:          3         11       NA    0.00000000
##   4:          4         10       NA    0.04743416
##   5:          5          9       NA    0.61683333
##  ---
## 229:        229         14       NA    0.06681531
## 230:        230          1       NA    0.50000000
## 231:        231          6       NA   -0.20412415
## 232:        232          3       NA    0.08660254
## 233:        233          6       NA    0.24494897
```

Sentence Structure

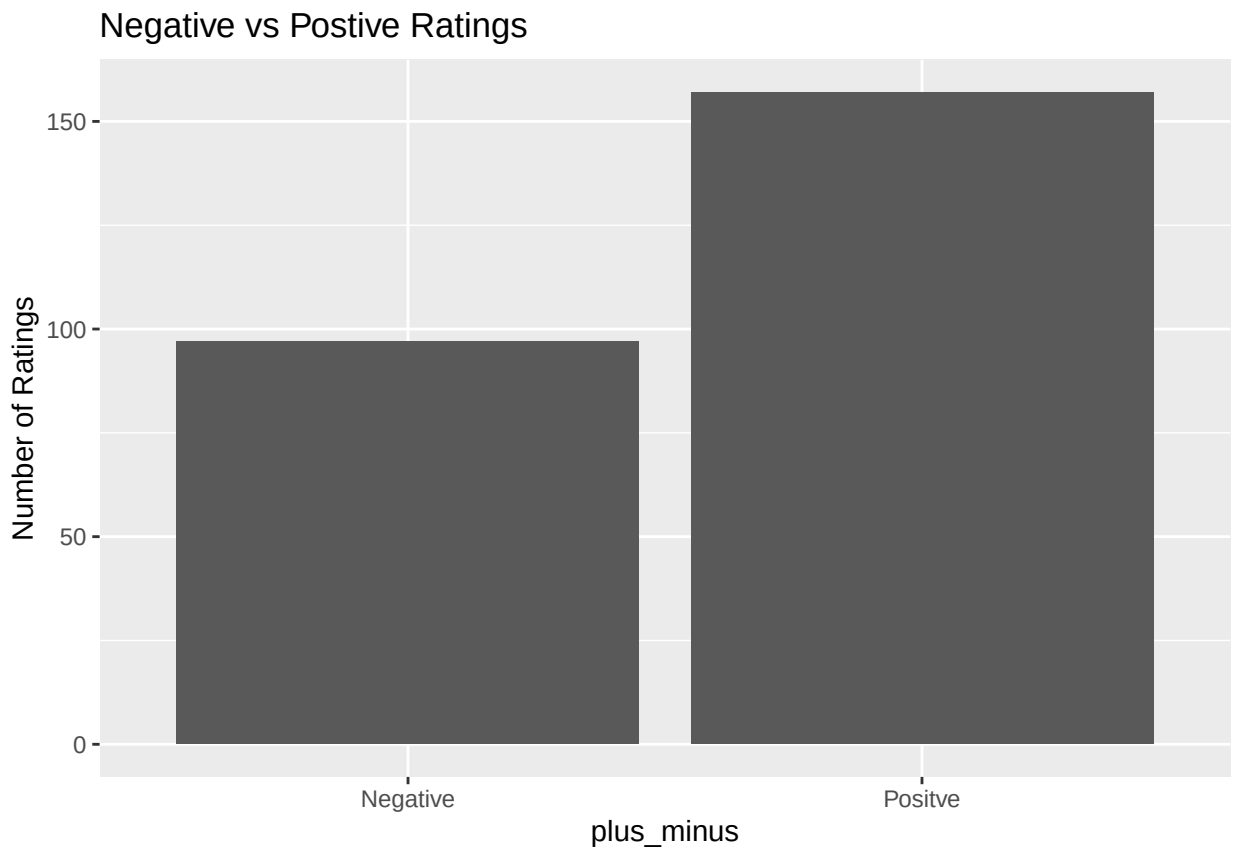
```
taras_sentence <- taras_caramels %>% select(review) %>%
  get_sentences() %>%
  sentiment()

taras_sentence %>% ggplot() + geom_density(aes(x=sentiment)) +
  labs(title="Density Plot of Rating Sentiment",
       x=" <-- Negative(-) Bad  but Plus(+) Good --> ")
```




```
taras_sentence %>% ggplot() + geom_histogram(aes(x=sentiment)) +
  labs(title="Histogram of Rating Sentiment",
       x=" <-- Negative(-) Bad  but Plus(+) Good --> ")
```



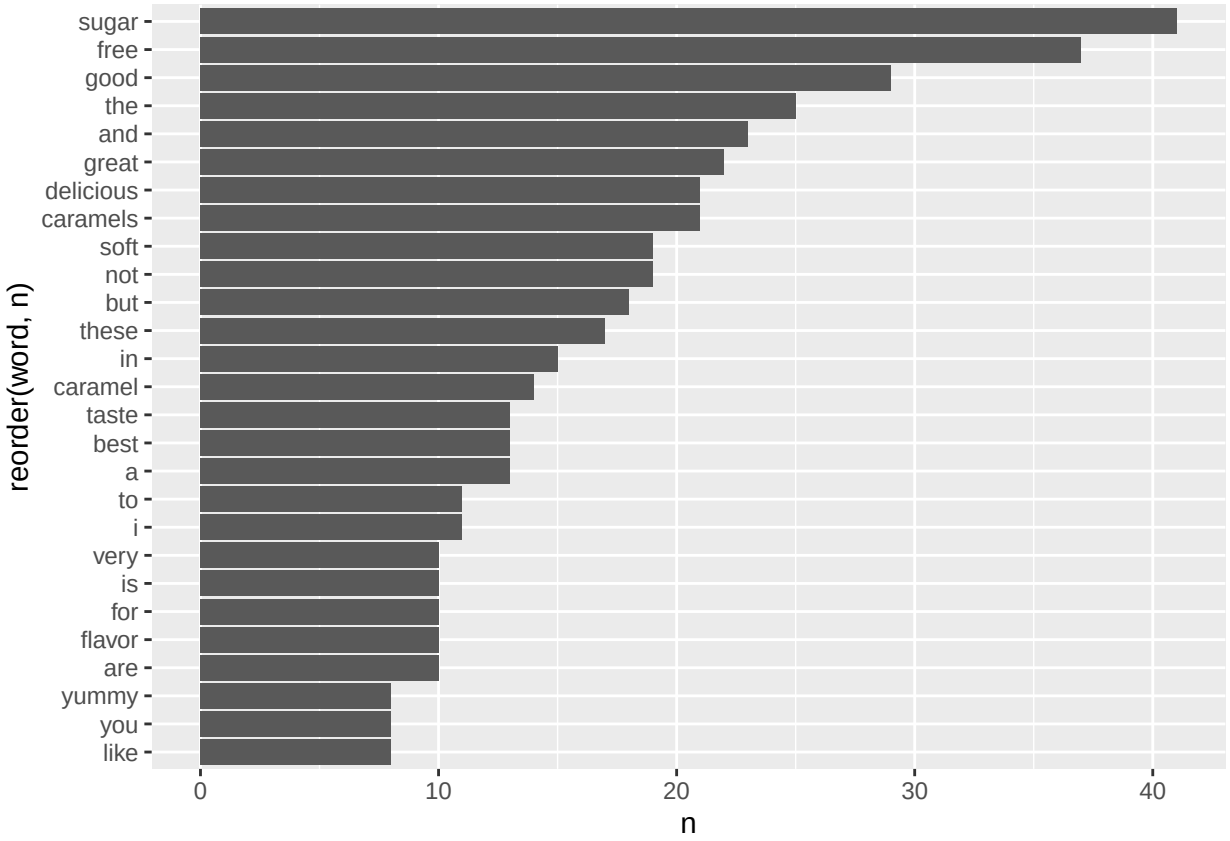

```
taras_sentence <- taras_sentence %>%
  mutate(plus_minus = if_else(sentiment >0,"Positve","Negative"))

taras_sentence %>% count(plus_minus,sort =T) %>%
  ggplot(aes(x=plus_minus,y=n)) + geom_col() +
  labs(title = "Negative vs Postive Ratings",
       y="Number of Ratings")
```



Begin TidyText Package Analysis

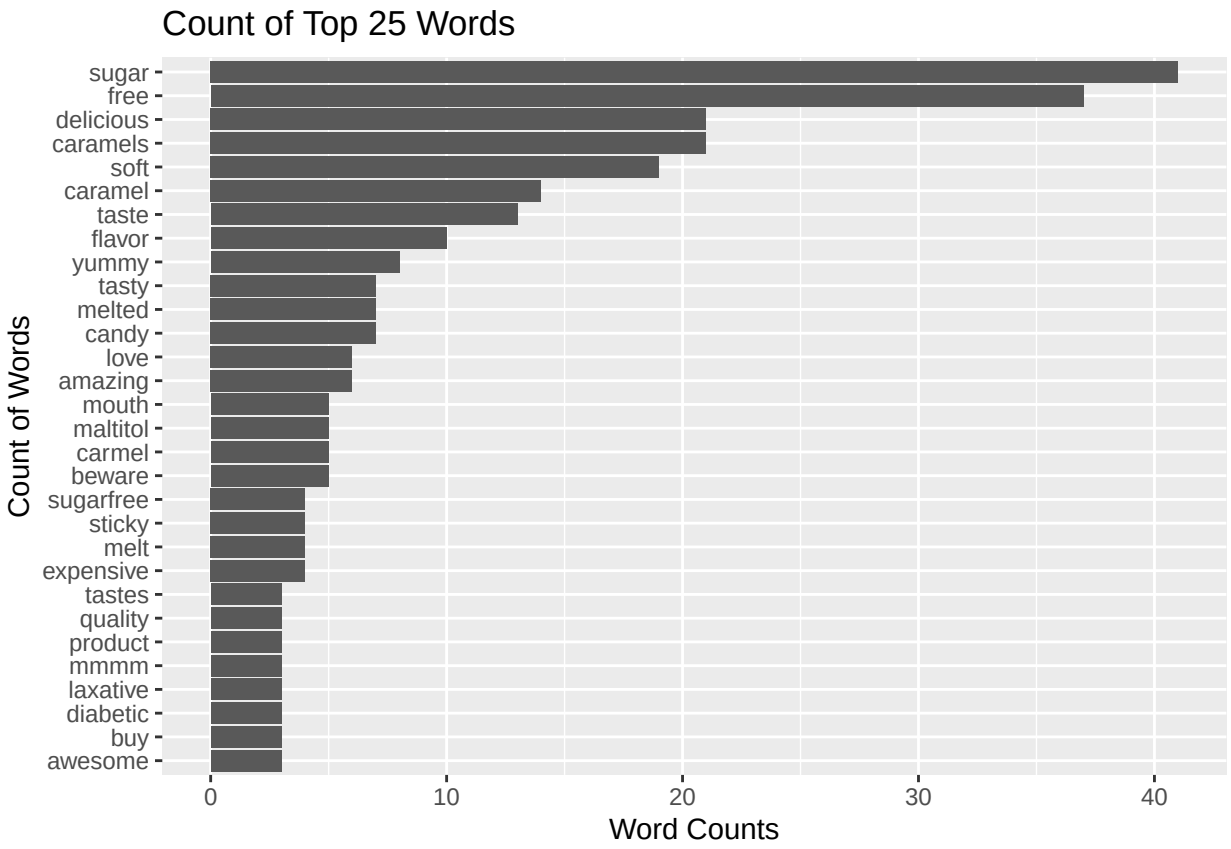
```
taras_tokens <- taras_caramels %>% select(review) %>%
  unnest_tokens(word, review)
```

```
taras_tokens  %>% count(word,sort =T ) %>% top_n(25) %>%
  ggplot(aes(x=reorder(word,n),y=n)) + geom_col() + coord_flip()
```




```
data(stop_words)
taras_clean <- taras_tokens %>%
  anti_join(stop_words)
```

```
taras_clean %>% count(word,sort =T ) %>% top_n(25) %>%
  ggplot(aes(x=reorder(word,n),y=n)) + geom_col() + coord_flip() +
  labs(title ="Count of Top 25 Words",
       x="Count of Words","Token Words",y="Word Counts")
```



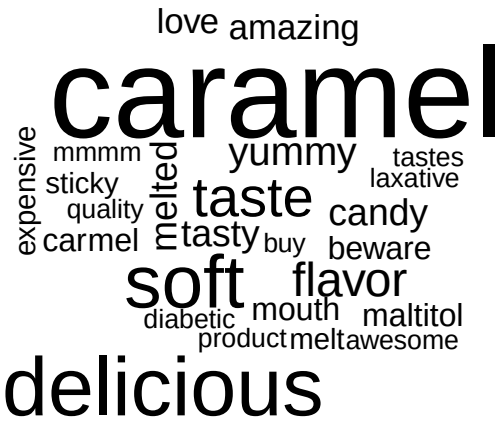
Construct a Word Cloud

```
taras_clean <- taras_clean %>% mutate(word = if_else(word =="free","sugarfree",word))
taras_clean <- taras_clean %>% mutate(word = if_else(word =="sugar","sugarfree",word))

taras_clean <- taras_clean %>% mutate(word = if_else(word =="caramals","caramel",word))

taras_clean <- taras_clean %>% mutate(word = if_else(word =="caramels","caramel",word))

taras_clean <- taras_clean %>% mutate(word = if_else(word =="caramel's","caramel",word))
```

```
taras_clean %>% filter(word =="sugarfree") %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 100))
```

sugarfree

```
taras_clean %>% filter(word !="sugarfree") %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 100,random.color=TRUE))
```


love
amazing
caramel
expensive
mmmm
melted
yummy
tastes
laxative
sticky
taste
candy
quality
carmel
tasty
buy
beware
soft
flavor
diabetic
mouth
maltitol
product
melt
awesome
delicious