Module 4 Bonus Assignment - Sentiment Analysis

## Gil Graybill

## MIS 506

## April 10, 2021

As I was looking through the amazonbook\_reveiw.csv dataset I saw that 836 of the reviewers had the name of “Amazon Customer” out of 7045. Apparently this is the default name if the reviewer doesn’t pick a name. If a reviewer isn’t picking a name, is it because they want to remain anonymous or not be contacted about their review? Are they more negative or positive than those who choose a name? Are they more prone to use certain words?

# Two questions to answer

1. Which words contribute to the sentiment scores for these reviews?
2. Do reviews by “Amazon Customer” differ in sentiment than the rest of the reviews.

### Load the required libraries

### Read in the data

read-in the following data set in R: amazonbook\_review.csv

### Modify the dataset to make columns more readable

book\_reviews <- book\_reviews %>%   
 select (ReviewerID, name = ReviewerName, review = ReviewerText)

### Remove contractions,

### Tokenize and preprocess text, remove whitespace

### Add whether or not the reviewer had “Amazon Customer” in the name, and get the counts for later

tidy\_book\_reviews <- tidy\_book\_reviews %>%  
 mutate(amazon\_reviewer = ifelse(grepl('Amazon Customer', name), "True", "False"))  
  
amazon\_reviewer\_count = as.numeric(count(tidy\_book\_reviews %>% filter(amazon\_reviewer == "True")))  
non\_amazon\_reviewer\_count = as.numeric(count(tidy\_book\_reviews %>% filter(amazon\_reviewer == "False")))  
total\_reviewer\_count = as.numeric(count(tidy\_book\_reviews))

### Using bing lexicon, get the sentiments for the reviewers whose names are “Amazon Customer”. Add a column “prop” which is the proportion of the uses of the words divided by the number of reviews in that catetory. The number itself is not necessarily helpful but will be useful when comparing

bing<-get\_sentiments("bing")  
  
joined\_tidy\_book\_review\_true <- tidy\_book\_reviews %>%  
 filter(amazon\_reviewer == "True") %>%  
 inner\_join(bing) %>%  
 count(word,sentiment,sort = TRUE) %>%   
 # Group by sentiment  
 group\_by(sentiment)%>%  
 # Take the top 10 words for each sentiment)%>%  
 top\_n(10) %>%  
 ungroup() %>%  
 mutate(word = reorder(word, n), amazon\_reviewer = "True") %>%  
 mutate(prop = n/amazon\_reviewer\_count)

## Joining, by = "word"

## Selecting by n

### Using bing lexicon, get the sentiments for the reviewers whose names are NOT “Amazon Customer”. Add a column “prop” which is the proportion of the uses of the words divided by the number of reviews in that catetory. The number itself is not necessarily helpful but will be useful when comparing

bing<-get\_sentiments("bing")  
  
joined\_tidy\_book\_review\_false <- tidy\_book\_reviews %>%  
 filter(amazon\_reviewer == "False") %>%  
 inner\_join(bing) %>%  
 count(word,sentiment,sort = TRUE) %>%   
 # Group by sentiment  
 group\_by(sentiment)%>%  
 # Take the top 10 words for each sentimentiment)%>%  
 top\_n(10) %>%  
 ungroup() %>%  
 mutate(word = reorder(word, n), amazon\_reviewer = "False", prop = n/non\_amazon\_reviewer\_count)

## Joining, by = "word"

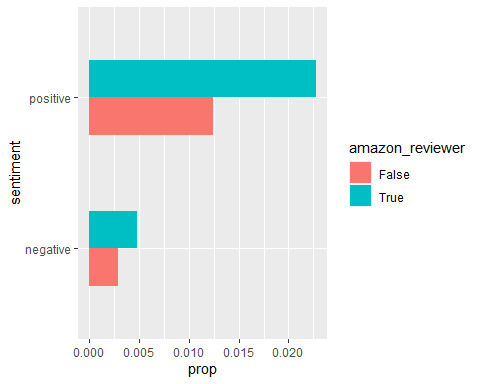
## Selecting by n

### Join the two tables together.

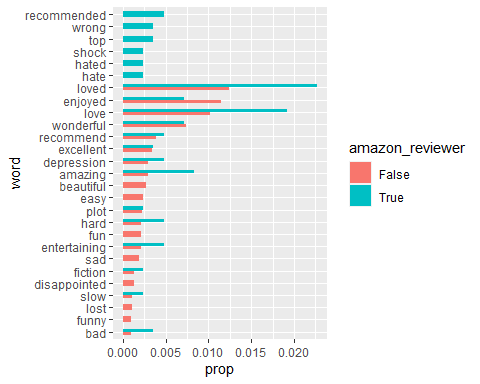
### Graph by sentiment and group by reveiwer type.

### Graph by word and group by reveiwer type.

all\_joined\_tidy\_book\_reviews <-   
 rbind(joined\_tidy\_book\_review\_false, joined\_tidy\_book\_review\_true)  
  
all\_joined\_tidy\_book\_reviews %>%  
 ggplot(aes(x = sentiment, y = prop, fill = amazon\_reviewer)) +  
 geom\_bar(stat = "identity", position = "dodge", width = 0.5) +   
 coord\_flip()



all\_joined\_tidy\_book\_reviews %>%  
 ggplot(aes(x = word, y = prop, fill = amazon\_reviewer)) +  
 geom\_bar(stat = "identity", position = "dodge", width = 0.5) +   
 coord\_flip()



## Answer to questions

It looks like those reviewers who do not choose a name are both more positive and more negative than those who do. They seem to want to express themselves more forcefully either in a positive or negative way. They use more words that trigger sentiment in the Bing Lexicon.

The use of the words “Love” and “loved” lead the list and account for big differnces. Anonymous reviewers use the words “Amazing”, “hard”, and “entertaining” more. The words “wonderful”, “recommend”, and “excellent” are used in equal proportion across all users.