Sentiment Analysis for Amazon Phone Reviews

Andrew Zhang

12/3/2018

Abstract

Amazon is a huge marketplace for a variety of products, ranging from school material to beauty care. With such a huge customer base, each of the products has an option for reviews. Looking specifically at phones, we are interested in any patterns associated with mobile phone reviews. In addition, we want to see whether or not the number of words per review follows the benford distribution. Following our analysis, we were able to conclude that there is a stark difference between negative and positive reviews in terms of word counts and word types.

Background

Amazon has been expanding their market reaches to almost every industry in the world and the mobile device industry is no exception. Amazon has been marketing cellular devices from a wide variety of companies such as Apple and Samsung. What makes Amazon unique is the fact that every product has readily available reviews from previous customers, whereas, company sites such as Apple, do not include such reviews. However, we are curious as to how useful these reviews are and whether or not they are genuine. For this project, we utilize sentiment analysis to explore the general characteristics of reviews and apply Benford analysis to determine the validity of these reviews.
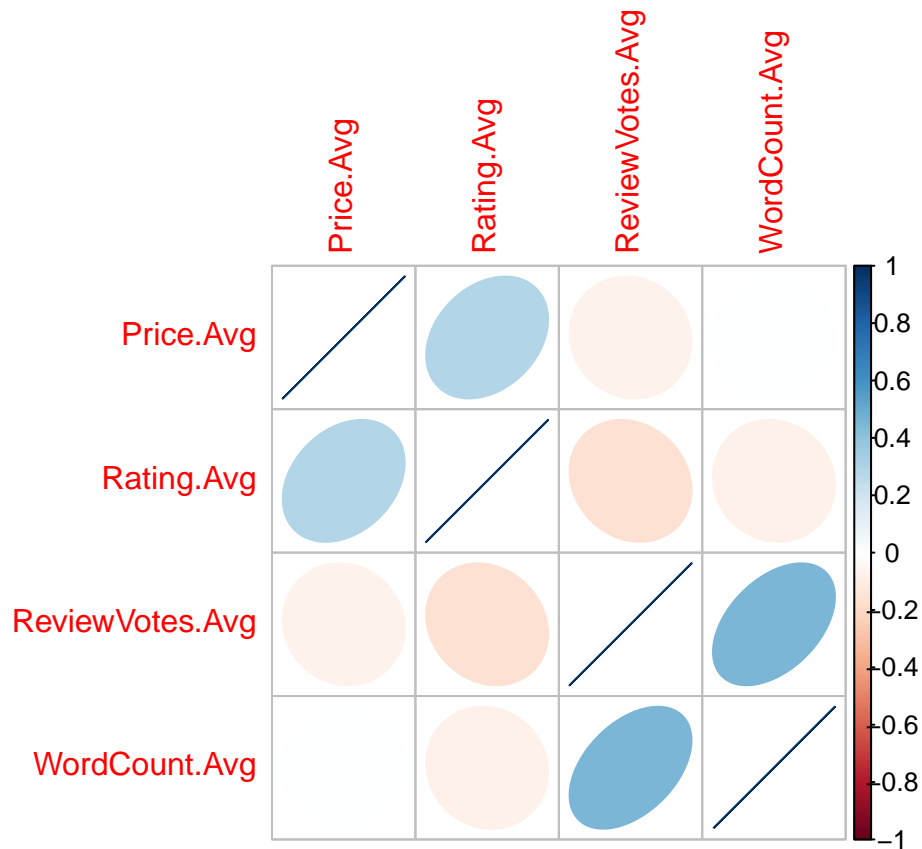
Variables

Table 1: Variables

variable_names	descriptions
Product.Name	Name of the Product
Brand.Name	Company that created the product
Price	Amount of money for the product
Rating	Rating a customer gives the product(Out of 5)
Reviews	Written comments regarding the product
Review.Votes	Number of votes a review gets

Data Cleaning/Exploration

First and foremost, we looked at the structure of the data to ensure each variable was the proper type. Then we proceeded to remove any missing observations. Then, utilizing sentiment analysis, we split the reviews into words, removed stop words, and counted up the words per review. We then, added positive and negative sentiments to determine the general characteristics for both types of sentiments. For the sake of exploration, we aggregated each of the observations by review number in order to make exploratory data analysis easier. Finally, because there were so many words being tokenized, we filtered the data to include only the biggest brands or brands with reviews that counted more than 800.

We want to look at the variables to see which of the variables are correlated with one another, whether it be trivially or not. Above, we can see that Ratings and Prices are slightly correlated while Word Counts and Review Votes are also correlated. Being positively correlated means that there is a positive relationship where an increase in one variable would mean an increase in the other. Meanwhile, Ratings and Review Votes are negatively correlated, meaning an increase in ratings would mean a decrease in number of review votes. This is something we should pay attention and will discuss further in the analysis.

Next, we will look at some exploratory plots for the aggregated data.

From the graphs above, it is difficult to make any conclusions between any of these variables as, even though the regression line points in a certain direction, there is too much noise and variance in the datapoints to soundly conclude a positive or negative relationship between variables.
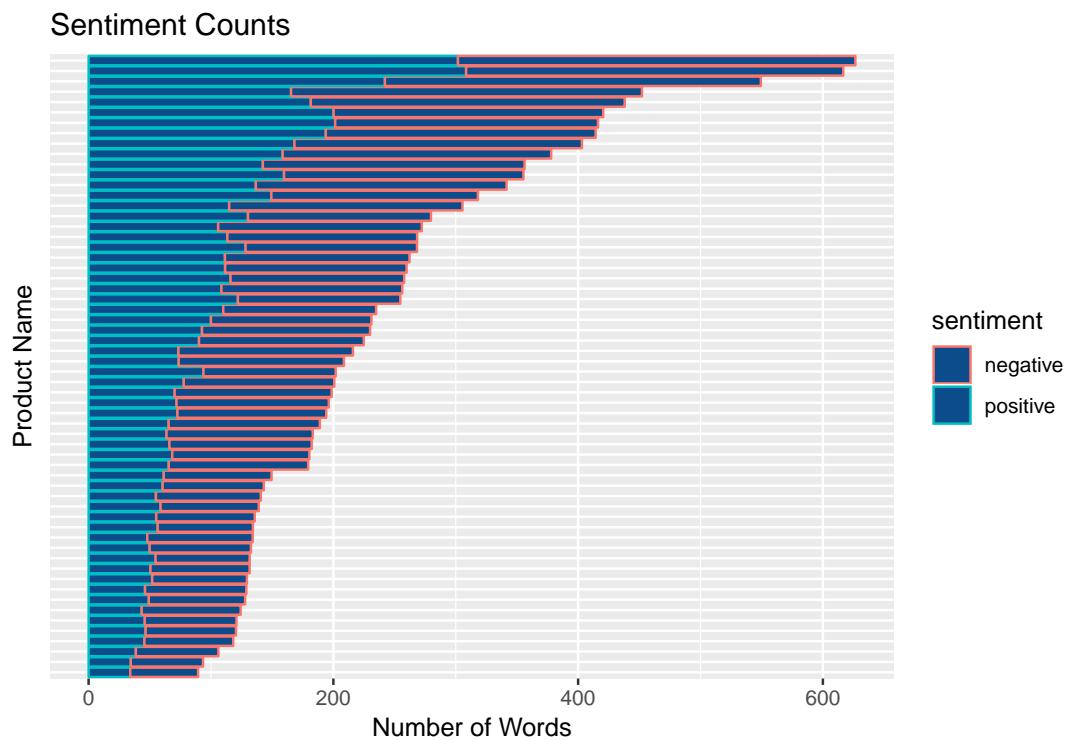

We can see that, looking at the number of negative and positive words, reviews typically have more words for negative reviews than for positive reviews.

Pattern Analysis

Ratings vs. Price
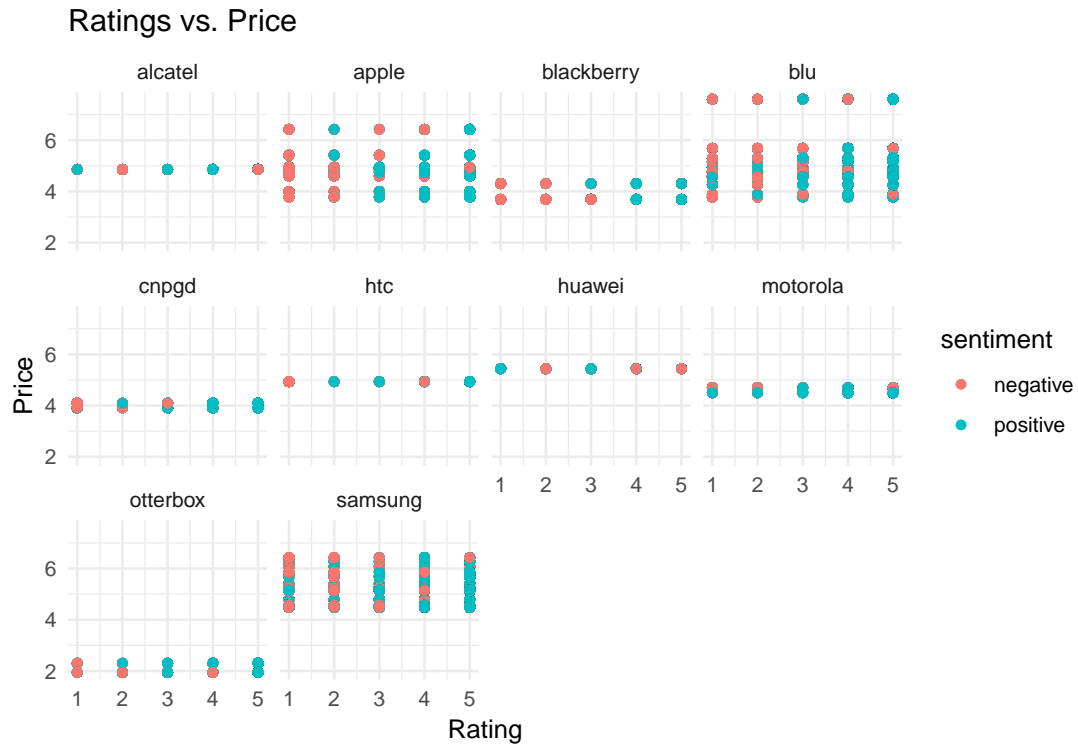
Continuing on our exploratory plots from earlier in the report, it's still hard to find clear patterns within the data. Looking at the correlated variables within the full dataset, it's hard to decipher any relationship between the variables as many of the points are the same. Examining the Ratings vs. Prices plot, we can see that the majority of the negative sentiment is on the lower end of the rating scale while the positive sentiment is on the higher end of the rating scale, indicating more positive reviews mean higher ratings while more negative reviews mean lower ratings.

Word Count vs. Review Votes
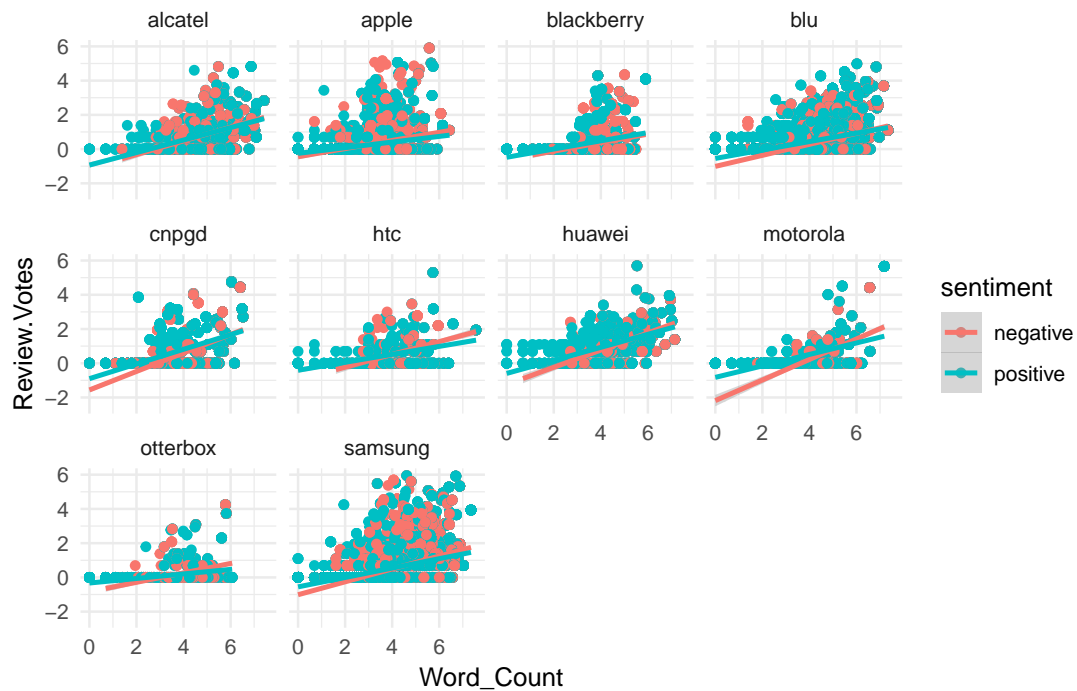
Next, if we look at the plots for the Review Votes against the Word_Counts, we can see that both negative and positive sentiments have an upward trend, which signify that more words in a review can be associated with a higher number of votes for a particular review. We also know that a larger number of words indicates a lower rating or more negative sentiment, so we can tentatively say that people vote for more negative reviews.

Ratings vs. Review.Votes
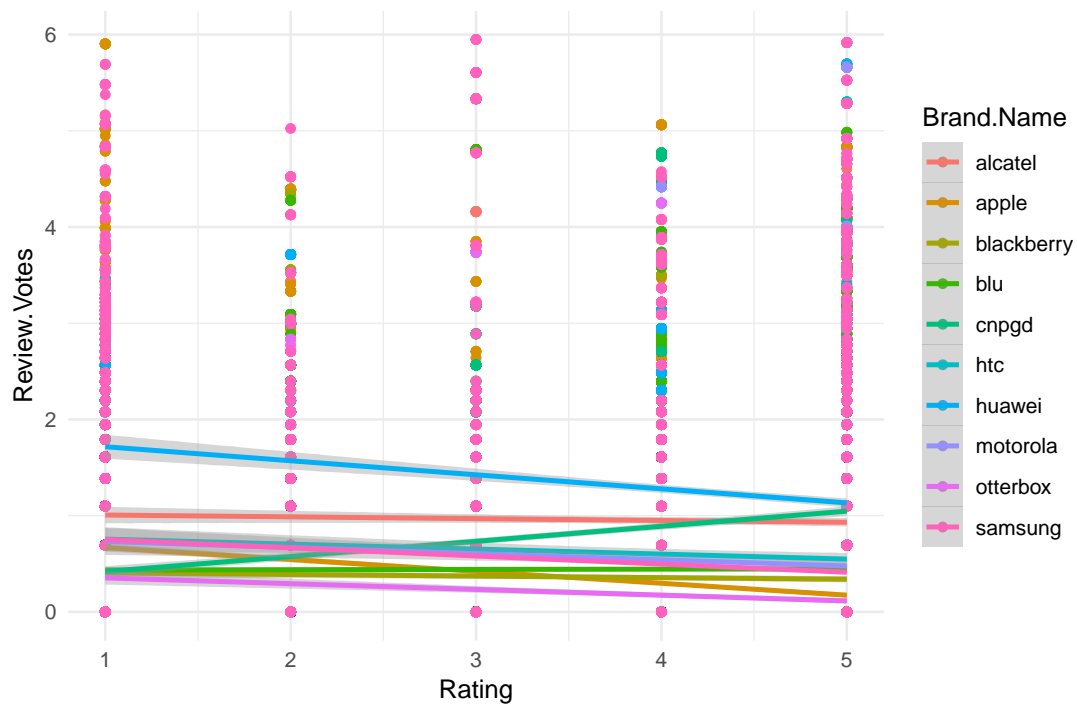
Examining the relationship between ratings and number of review votes, we can see, for the most part, that as the ratings increased, the number of review votes decreased. From the previous plot with the relationship

between word counts and the review votes, the inverse relationship here indicates a potential relationship between the number of words and ratings.

Positive vs. Negative Distribution
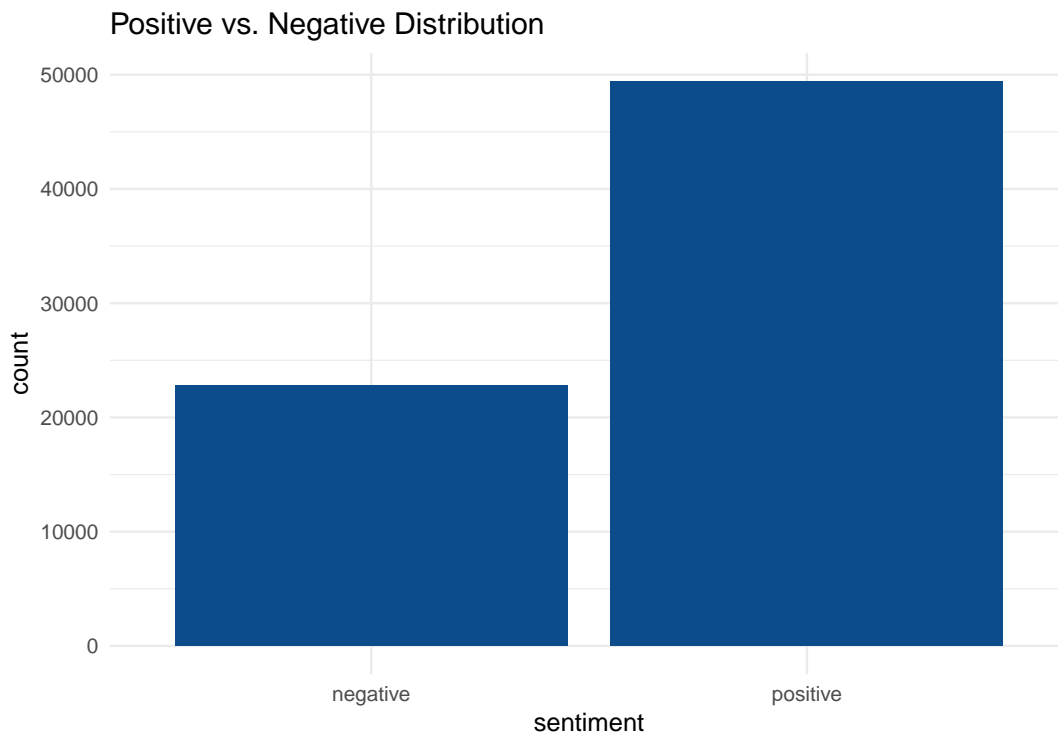

We can see that the majority of the sentiments in the reviews are positive, which is reflective of the average rating of the reviews, 3.87/5.

Distribution of Word Counts
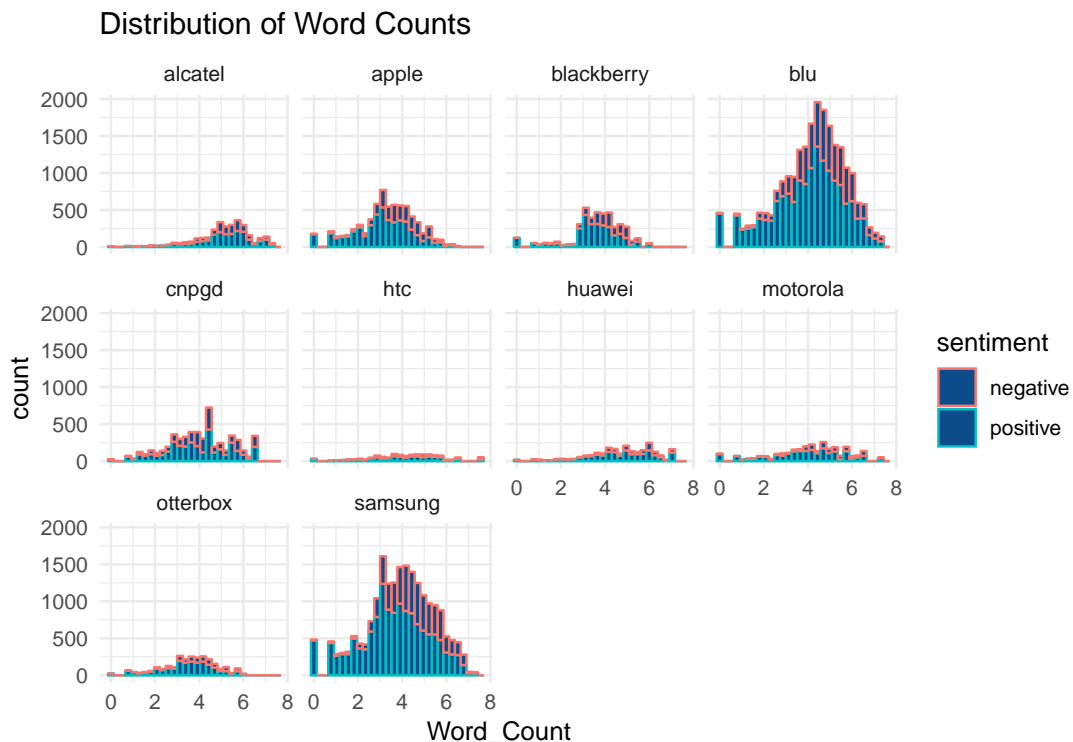
From this particular plot, we can see that the number of words is significantly more for negative sentiments than for positive sentiments. In addition, we can see that Samsung and Blu have the most reviews amongst

the groups. This could be attributed to company popularity.
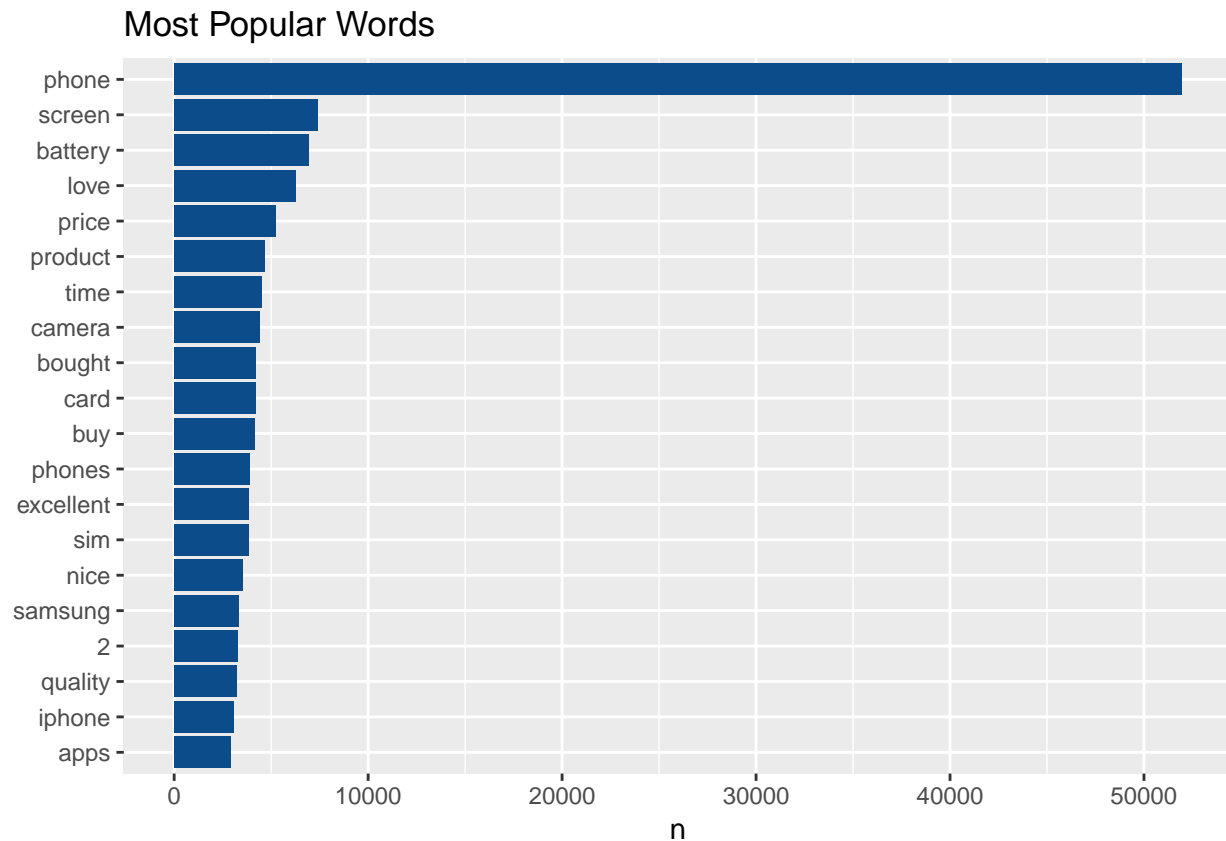
Most Popular Words

Looking at the most popular words, we can see that the word *phone* is easily the most used word form the figure above. However, if we look at the words in this list of 20, we can see lots of the words are reflective of the features that people look for prior to phone purchases. These words include:

- *Screen*
- *Battery*
- *Camera*
- *Sim*
- *apps*

Interestingly, we can also see the word *iphone* and *samsung* on the list. Although we filtered the brands to only reflect the most popular brands, we can see that two of the most popular brands are Apple and Samsung as many people include the product/brand name in their reviews.

Selecting by n
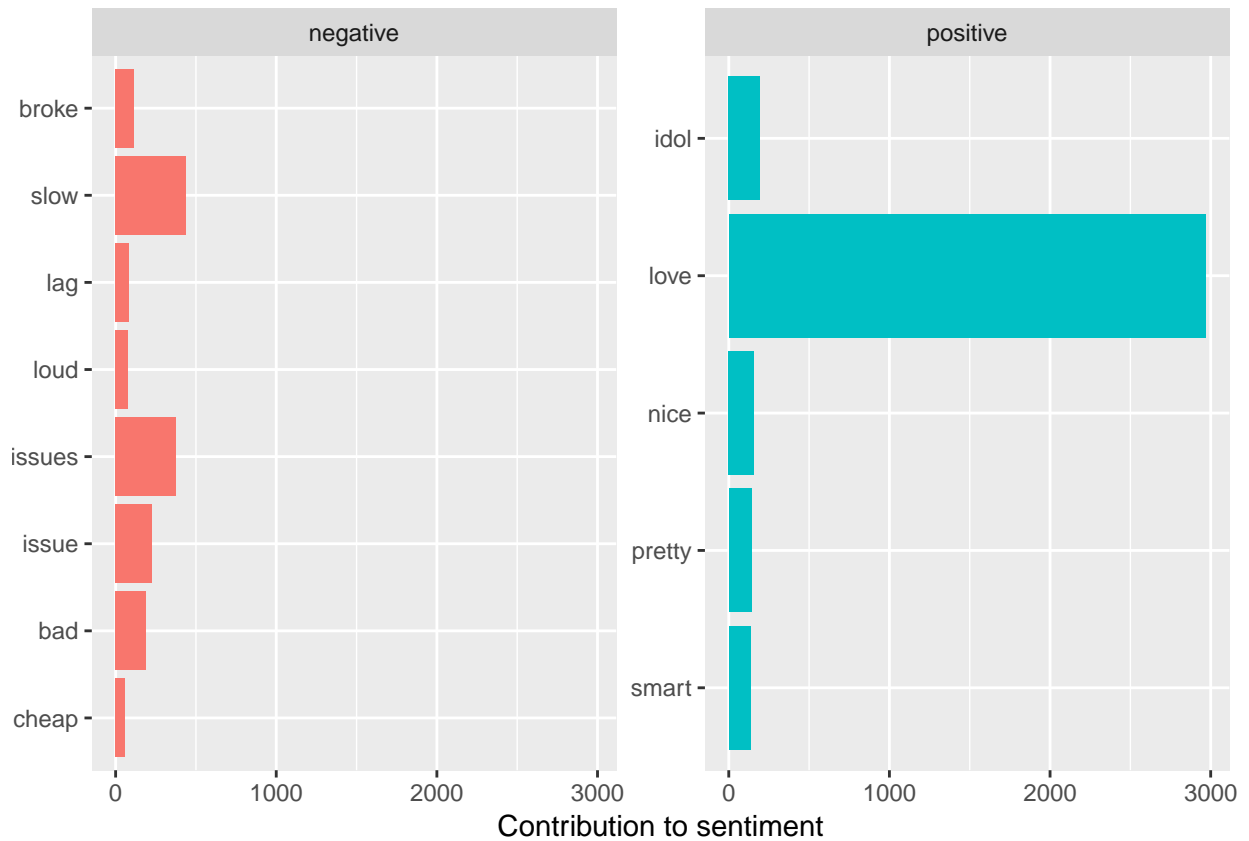
From the figure above, we look at the words that contribute the most to the bing sentiment. Looking at the positive graph, we can see that, most of these words are subjective towards the phone; love, nice, pretty, etc. If we look at the negative graph, most of these words are related to technical aspects of these phones; slow, lag, issues, cheap, etc. This helps show the focus of both positive and negative reviews, where positive reviews are often subjective and emotionally driven, while negative reviews are harsh criticisms on the device features.

Benford Analysis

Benford Analysis for Product Names

Digits Distribution

Digits Distribution Second Order Test

Summation Distribution by digi

Chi−Squared Difference

Summation Difference

Legend Dataset: revcounts
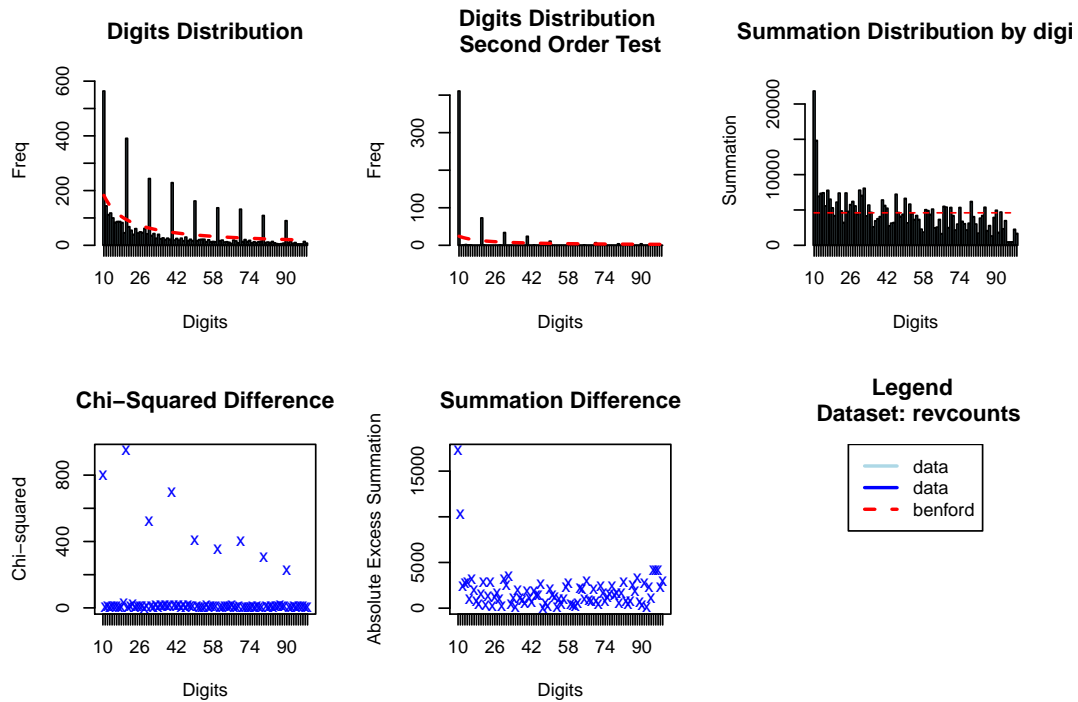
- data
- data
- benford

Benford Analysis for Brand Names

Digits Distribution

Digits Distribution Second Order Test

Summation Distribution by digi

Chi−Squared Difference

Summation Difference

Legend Dataset: revcounts2
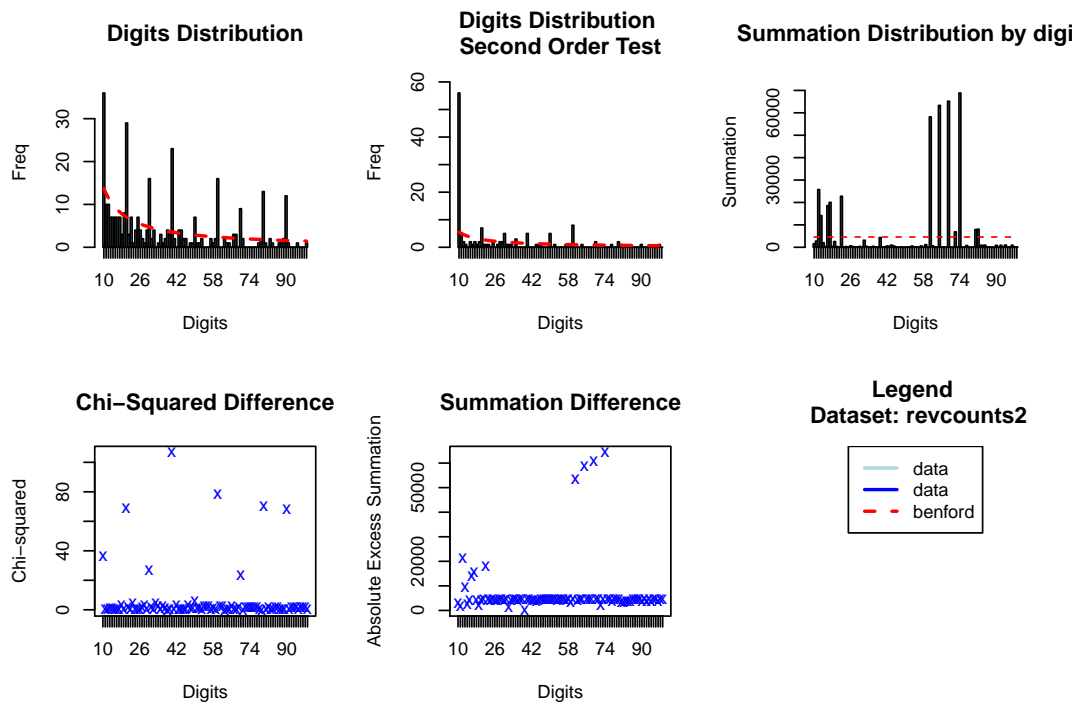
- data
- data
- benford

Looking at the Benford analysis, we can see that the data does not follow a benford distribution as there are

spikes throughout the data. It is possible that falsely generated reviews or "meme" reviews were written for these devices.

Conclusion

Utilizing the initial correlation plot, we discovered potential positive relationships between ratings/price and review votes/word count, while there is a potential negative relationship between ratings/review votes. Looking at the general trends, we were able to conclude several things about the data.

First, the number of words in a negative review is significantly more than the number of words in a positive review. If we think back to all of the times we browsed amazon reviews, most of the time, people have more to say about why a product is dysfunctional and why they don't particularly enjoy it, whereas, positive reviews are typically along the lines of "Great product" or "I love it", rather than a detailed explanation of why the individual enjoys the product.

Second, we also discovered that the number of review votes were associated with the word counts of reviews. The more words a review had, the more votes the particular review had. Since we know that negative reviews had more words per review than positive reviews, we can associate negative reviews with a larger number of review votes. This makes sense as, when people scroll thorugh reviews looking for sincere reviews regarding products, many are not going to take "I love it" seriously. However, an individual who comments on how poor the camera quality of the phone might be or how poor the battery life is, will likely get more votes as others will find the review more helpful.

Lastly, utilizing the previous two conclusions regarding number of review votes being associated with word counts and negative ratings having more words per review, we can draw the conclusion that negative ratings have more words per review, which follows the negative correlation displayed in the correlation plot.

On top of negative vs. positive reviews, we can look at company specific reviews. If we look at the distribution of word counts divided by companies in the plot earlier in the report, we can see there is a significant difference in how reviews are distributed. We expect larger companies like Apple and Samsung to dominate the board, however, we really only see Samsung and Blu. Apple doesn't a significantly large number of mobile reviews even though the iPhone is one of the most popular phones in the world. Doing some research, both Samsung and Blu have had issues with their phones recently. As we all know, Samsung had the exploding phone issue not too long ago, while Blu has recently been dealing with public backlash for chinese spyware on their phones. With such issues, we might come to the conclusion that people are filling more and more complaints against these companies for any tiny issues with their mobile devices.

In regards to the Benford analysis, we can see that per both number of reviews per product and number of reviews per brand, neither follow a Benford distribution. Now, if we think about it, going through reviews, often times, we can find reviews that are insincere or fake, where people either write something as a joke or there is a computer automated response that makes no sense. It is possible that these sorts of responses skew the distribution of the data slightly. Something like this is worth looking into more detail about.

Future Works

One of the things we can potentially look at is a more in-depth review of each of the phone brands. Companies like Samsung, Apple, and LG are bound to have significantly different reviews and types of reviews compared to one another.

Another aspect worth looking into is creating a model for predicting either ratings or prices based on a combination of variables such as review votes, word counts, sentiment score, etc. It would be interesting to see if an universal model exists for how different brands price their phones and whether or not they take into consideration the public opinion on their devices.

Acknowledgement

Special thank you to Professor Wright for spending so much time teaching and creating so many unique assignments and projects this semester.

Shiny App

```
library(shiny)

## Warning: package 'shiny' was built under R version 3.4.4

library(ggpubr)

# Define UI ----
ui <- fluidPage(
    titlePanel("Sentiment Reviews per Phone"),
     # Input: Slider for the number of bins ----
    sidebarLayout(
      sidebarPanel(
        selectInput(inputId = "Top.Brand.Name", label = "Mobile Phone Brands",
                    choices = unique(amazon_sub$Brand.Name),
                helpText("Choose which brand you'd like to observe"),
                multiple = FALSE)
      ),
    mainPanel(
      plotOutput("plot")
      )
    )
)

# Define server logic ----
server <- function(input, output, session) {
  afinn4 <- reactive(tidy_words %>%
  filter(Brand.Name == input$Top.Brand.Name) %>%
  inner_join(get_sentiments("afinn")) %>%
  group_by(index = id) %>%
  summarize(sentiment = sum(score)) %>%
  mutate(method = "AFINN"))

  bing_and_nrc4 <- reactive(bind_rows(tidy_words %>%
                               filter(Brand.Name == input$Top.Brand.Name) %>%
                               inner_join(get_sentiments("bing")) %>%
                               mutate(method = "Bing et al."),
                            tidy_words %>%
                               inner_join(get_sentiments("nrc") %>%
                                           filter(sentiment %in%
                                                   c("positive",
                                                     "negative"))) %>%
                               filter(Brand.Name == input$Top.Brand.Name) %>%
                               mutate(method = "NRC")) %>%
    count(method, index = id, sentiment) %>%
    spread(sentiment, n, fill = 0) %>%
```

```r
    mutate(sentiment = positive - negative))
  output$plot <- renderPlot(
    ggplot(bind_rows(afinn4(), bing_and_nrc4()), aes(index, sentiment, fill =
                                                      method)) +
      geom_col(show.legend = FALSE) +
      facet_wrap(~method, ncol = 1, scales = "free_y")
    )
}

# Run the app ----
shinyApp(ui = ui, server = server)
```

PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, please

Shiny applications not supported in static R Markdown documents