Obdurate Movie Review Sentiment Analysis - R

Chris Selig

June 24, 2017

# Summary

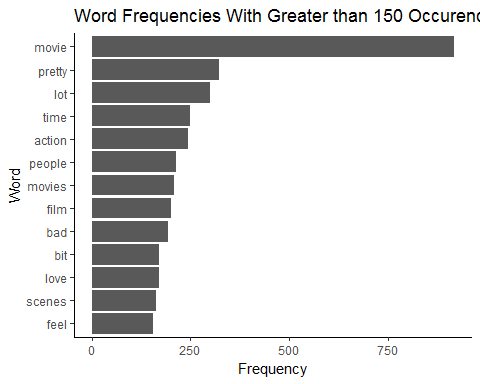
The goal of this analysis was to jump into my first foray into text analysis using data from someone I was pretty familiar with, my brother (aka Obdurate). Whenever we see movies, especially together, we like to discuss our thoughts and our ratings for the movies to see how we agree/disagree with the other's comments.

Using the data from his [Criticker account](https://www.criticker.com/profile/Obdurate), I explored his 1900+ movie reviews using [Sentiment Analysis](https://en.wikipedia.org/wiki/Sentiment_analysis) in R. Sentiment analysis is used to identify the overall sentiment in a text data set.

The raw data was accessed March 29th, 2017 and can be found on my [Github](https://github.com/chrisselig/ObdurateSentimentAnalysis) page, along with the associated R code so you can reproduce my analysis.

## Sentiment Analysis

First, let's take a look at the individual words used in Obdurate's reviews, and their frequency counts. I used tokenization (create one word per row) using the [tidytext package](https://cran.r-project.org/web/packages/tidytext/index.html), and removed unnecessary words (stop words) like "the" and "of" from the dataset because they are not useful or interesting for my purposes.



The plot shows the highest frequency words used, that have greater than 150 occurrences within the reviews. Not surprisingly, the most popular word by far was "movie", and similar words like "movies" and "film" are also very popular. Just from looking at the most frequent words, you can already see some positive words like "pretty" and "love" and even a negative word "bad."

Now for a comparison between the positive and negative words. The lexicon used for this analysis was generated through crowdsourcing, or through the labor of one of the authors that created the tidytext package used in the sentiment analysis. The construct was then validated against another form of crowdsourcing like restaurant reviews.

Plotted below is a word cloud that plots the top 100 words, and colors based on whether they are positive or negative. The larger the word, the more times it was used. Note, the size of a word in one category cannot be directly compared to the size of a word in the other category.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : recommendation could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : fascinating could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : genuine could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : engaging could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : charm could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : perfectly could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : strong could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : effective could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : gorgeous could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : intelligent could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : satisfying could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : beauty could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : believable could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : convincing could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : fans could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : amusing could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : daring could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : surreal could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : astounding could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : humorous could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : intriguing could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : appealing could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : consistently could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : inventive could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : magic could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : meaningful could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : praise could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : worthy could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : disgusting could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : appeal could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : classy could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : humor could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : impressive could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : masterpiece could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : nicely could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : poetic could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : positive could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : pure could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : superb could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : break could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : challenging could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : criticism could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : deny could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : die could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : difficult could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : doubt could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : gangster could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : gruesome could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : inconsistent could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : menacing could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : mindless could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : oddly could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : offensive could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : uncomfortable could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : wasted could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : attractive could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : greatness could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : stunning could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : annoying could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : assassin could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : crappy could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : disaster could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : dreadful could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : flawed could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : gross could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : ignore could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : impossible could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : insult could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : joker could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : killing could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : messy could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : painfully could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : pointless could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : sadness could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : struggle could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : zombie could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : brilliantly could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : expertly could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : obsession could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : respect could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : silent could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : accomplishment could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : cohesive could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : easier could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : flashy could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : foremost could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : intelligence could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : logical could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : marvel could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : master could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : modern could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : proves could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : redemption could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : romantic could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : succeed could not be fit on page. It will not be plotted.

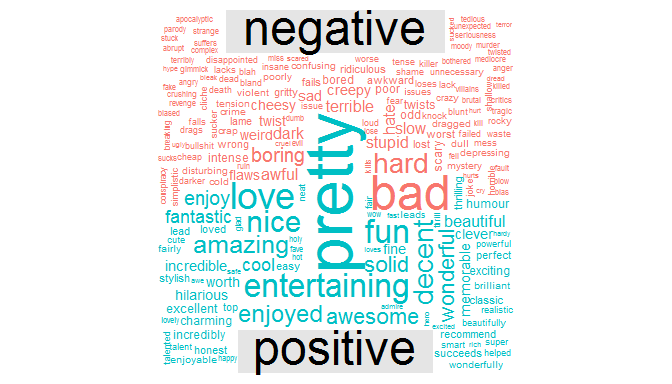
## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : thrilled could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : complaints could not be fit on page. It will not be plotted.

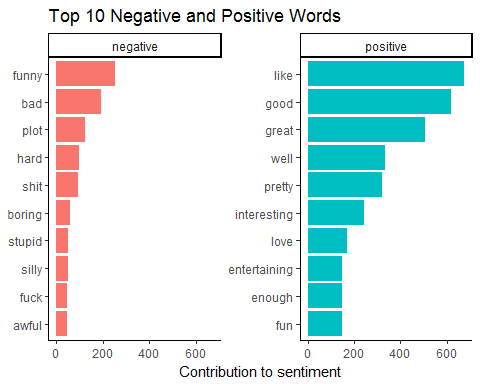
## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : conflicted could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : drag could not be fit on page. It will not be plotted.

## Warning in comparison.cloud(., colors = c("#F8766D", "#00BFC4"), max.words  
## = 300, : struggling could not be fit on page. It will not be plotted.



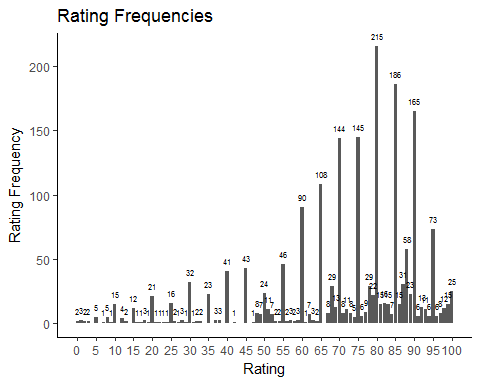
Taking a closer look, here are the top 10 negative and positive words.



There are some anomalies in the data, for example, "funny" and "plot", which are classified as negative words. Plot would be classified as a negative word because the lexicon thinks it's a secret plan by a group of people to do something illegal, instead of the main events of a movie. This is probably because the lexicon I used was not specifically for movie reviews, but rather restaurant reviews. On the other side, the positive words all look legitimate.

Looking at the frequencies of the positive and negative words, it appears that Obdurate's reviews are mostly positive. For fun, let's test this theory by plotting the frequency of his review scores, but first a quick blurb on how Obdurate review scoring system works.

Obdurate reviews on a 100-point scoring system, but the methodology of assigning a score has changed over time. He used to assign scores such as "82" but realized he could not really describe what makes an "82" worse than an "83", and has abandoned it. The simplified version scores uses increments of 5. So, the "82" would most likely be an "80" while "83" would turn into an "80" or "85"



As suspected, the majority of Obdurate's ratings are positive with a median score of 77, and a mean of 71. Note, Obdurate's ratings between 50-59 Obdurate describes as "on the fence", which translates to he likes it but probably would not recommend it to others, so every score 50 or both can be classified as "positive" or that he "liked" the movie.