Netflix On Twitter

Melody Shaff

December 17, 2017

Figure 1:

Netflix: The Rationale

I chose to look at Netflix as a brand for a number of reasons. First, and foremost, I'm a fan of the platform and have known Reed Hastings since before Netflix blew up. I remember testing and providing feedback (primarily through my parents) when the company was just getting started. Second, as a provider of a wealth of content, I was curious to see how people interacted with and referenced Netflix on Twitter: were they exuberant about a new Netflix Original show? Were they disappointed at buffering speed? Were people primarily engaged with the brand itself or with the content it provided? Some of these questions proved exceptionally difficult to explore, given time constraints, but Netflix still served as an interesting case study for brand interaction on Twitter.

Netflix Tweets: Reading in the Data Part I

To read in my data, after an enormous amount of time trying to to use TwitteR's searchTwitter() to no avail, I decided to use filterStream. This required some maneuvering and brushing up with the ROAuth package, but ended up being much more compatible with the work I sought to do with the data. I focused on two different types of exploration and analysis: mapping and sentiment analysis. Because mapping requires location, which I used longitude and latitude for, I actually ended up using filterStream() a bit differently each time. For the tweets that I used for mapping, I streamed in tweets for 900 seconds, tracking "Netflix" and asking for tweets within the US. For my sentiment analysis tweets, however, I filtered for longer (1100 seconds) and only tracked for "Netflix" with no requirement for language, location, or any other variable. This allowed me to filter my tweets later, without inhibiting the initial stream or getting unrelated tweets (as unfortunately tracking, location parameters, etc. are "or" rather than "and" guidelines within filterStream()).

Netflix Tweeting and Mapping

- First, I put all of the tweets (collected using filterStream()) into a JSON file.
- Second, I parsed the tweets and created a data frame, selecting only the variables that I planned to analyze.
- Third, I renamed the columns for longitude and latitude.
- Fourth, I filtered the tweets for more accurate longitude and latitude of the United States (some global tweets managed to make it into my initial data frame which caused issues with setView later on).
- Fifth, I removed NA's (for longitude and latitude) as this data is being used specifically for mapping.
- Sixth, I sped up the later process by writing to a CSV and reading the data frame in from their, rather than parsing though a JSON file.

Netflix in the US: A Few Simple Statistics

A quick look at the popularity of those tweeting.

Variable	Minimum	Maximum	Mean	Median
Followers	0	1,163,554	2,312.6	551
Retweets	0	595,674	12,271.5	3,847

Netflix: We're putting it on the Map!

After I had cleaned up my data, I was ready to map the tweets I'd gathered. They were spread out very much like you would expect US tweets to be spread out. As you can see below, there are clusters around major cities and densely populated areas with the most tweets coming from the East and West Coast, the Chicano area, etc. with very few coming from the Midwest.

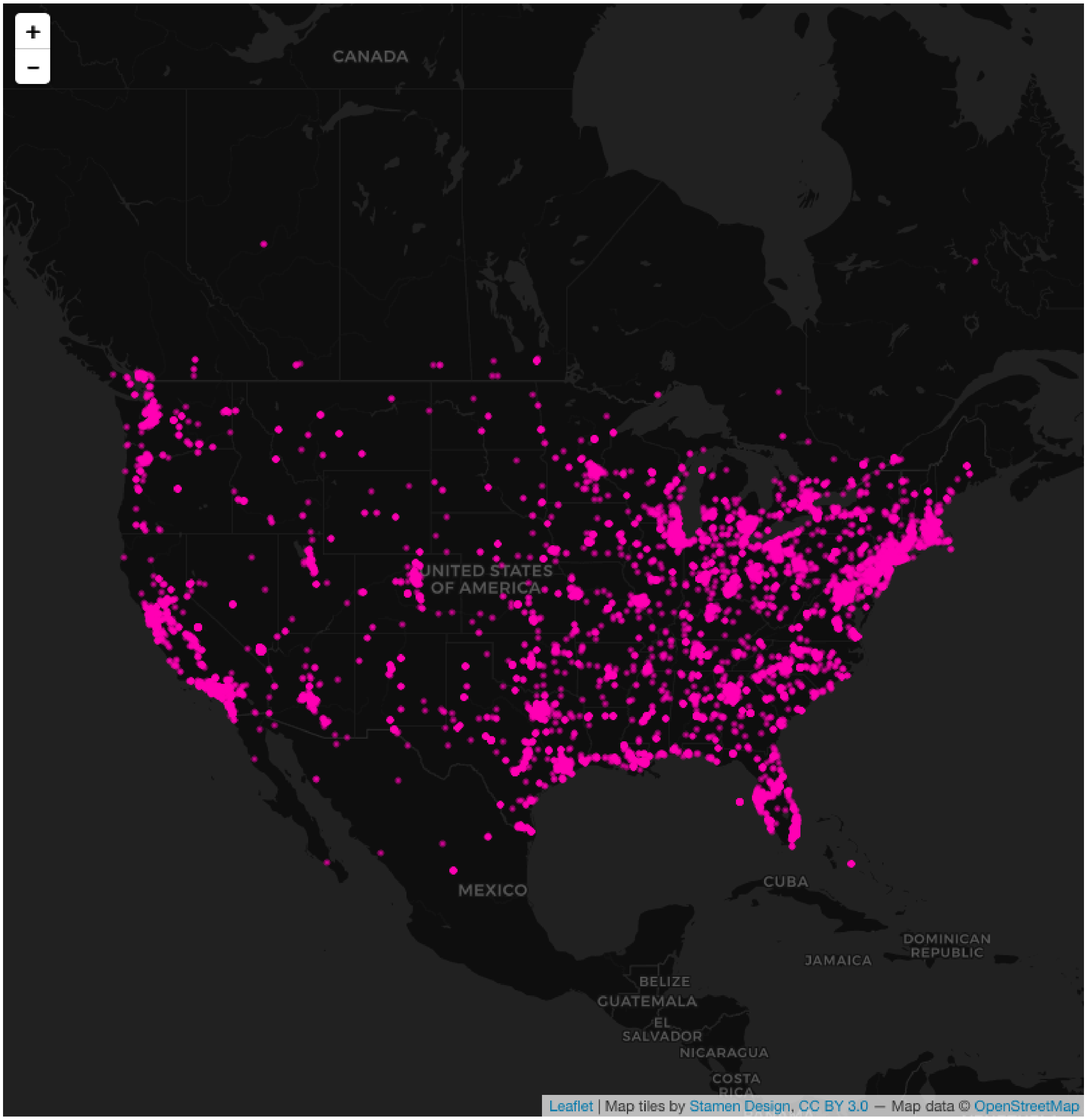


Figure 2:

Netflix Tweets for Sentiment Analysis

- First, I filtered in the tweets (again using filterStream(), but this time only tracking Netflix and not location or language). Also, because this had substantially less traffic, I opened the stream for longer (1100 seconds).
- Second, I parsed the tweets from a JSON file into a data frame for analysis.
- Third, because I did not filter by language earlier, I filtered out all of the tweets that were not in English (as these would hinder my ability to examine sentiment or just be filtered out later when the words didn't match sentiment words).
- Fourth, I established the data frame and made a new column "line" to mark which tweet words originated from (line is equal to tweet number, so I could assess as if it were a sentence in a book).
- Fifth, I stored the tweets in a new tibble, unnesting the words for their respective tweets.
- Sixth, I took out the stop words that got in the way of looking at most commonly used words of value.
- Seventh, I looked at the words sorted by the number of times they appeared. This allowed me to filter out junk words or character strings that have not relation or value in my project (ex. "http"), which I did next. > Now I could begin analyzing sentiment!

Netflix: Word and Tweet Sentiment

I went about looking at sentiment a few different ways.

1. Bing Quick Look: I looked at the words from all of the tweets about Netflix I collected and which of those overlapped with bing sentiment words (positive and negative binary sentiment of individual words). I put these in a new tibble. A few of the most common were:
   - Supported (n = 124)
   - Confusing (n = 123)
   - Happy (n = 121)
   - Boom (n = 112)

2. Word Cloud: I made a simple word cloud of the 100 most used words to talk about Netflix on Twitter.

3. Bing Comparison Cloud: I made a comparison cloud to look at the most common (bing) words used in the Netflix tweets, colored and split into negative and positive.

Figure 3:

4. Bing Sentiment: Because it can be hard to read the comparison cloud to draw any conclusions about the total bing sentiment (more positive vs. more negative) so I made a simple bar graph showing the total count of negative words and the total count of positive words.

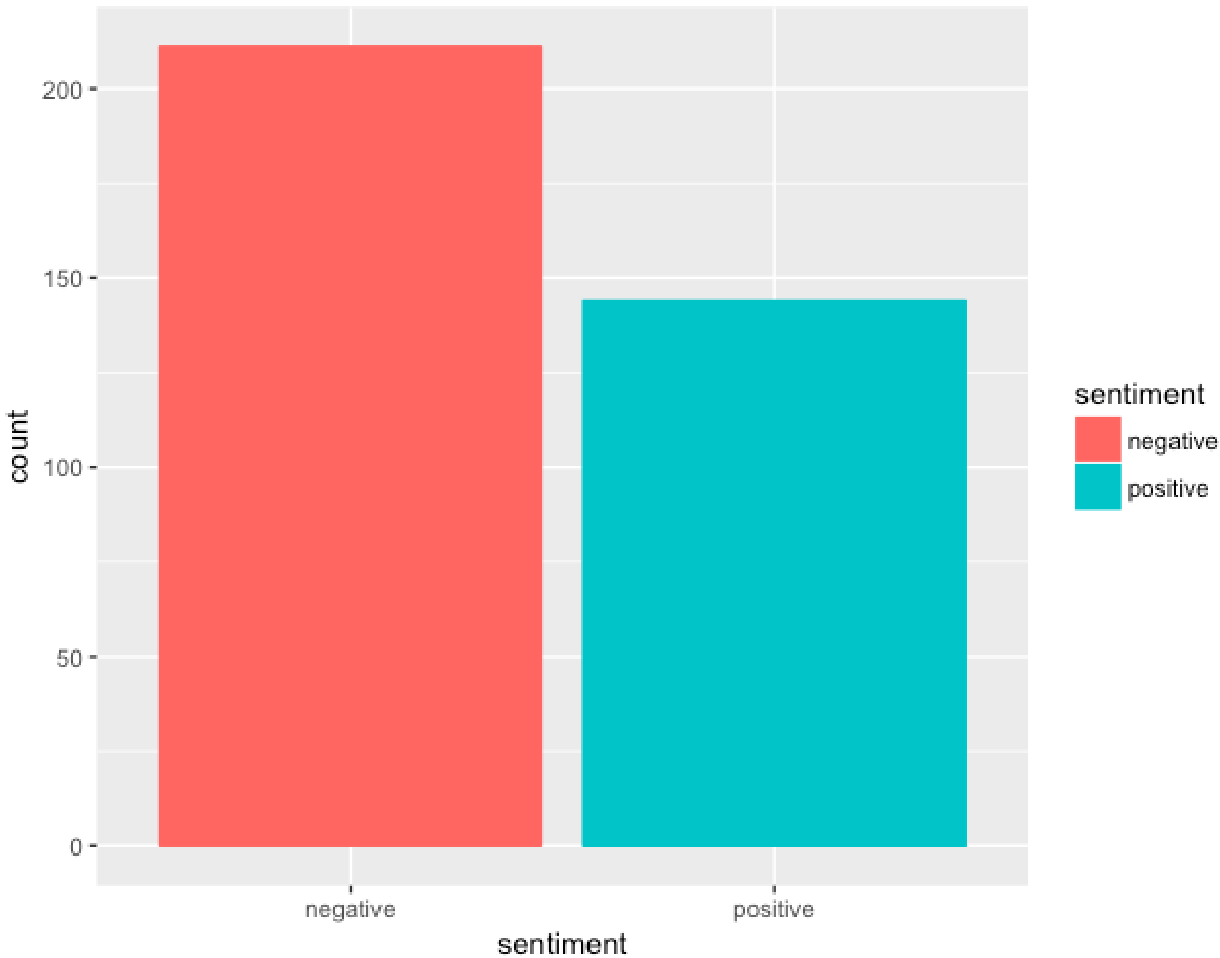


Figure 4:

5. Afinn Comparison Cloud: This time I made a comparison cloud looking at Afinn sentiment (which ranks sentiment positive to negative on a scale of -5 to 5).

Figure 5:

6. Afinn Bar Graph: A quick look at the count of each afinn score word used in tweets about Netflix.

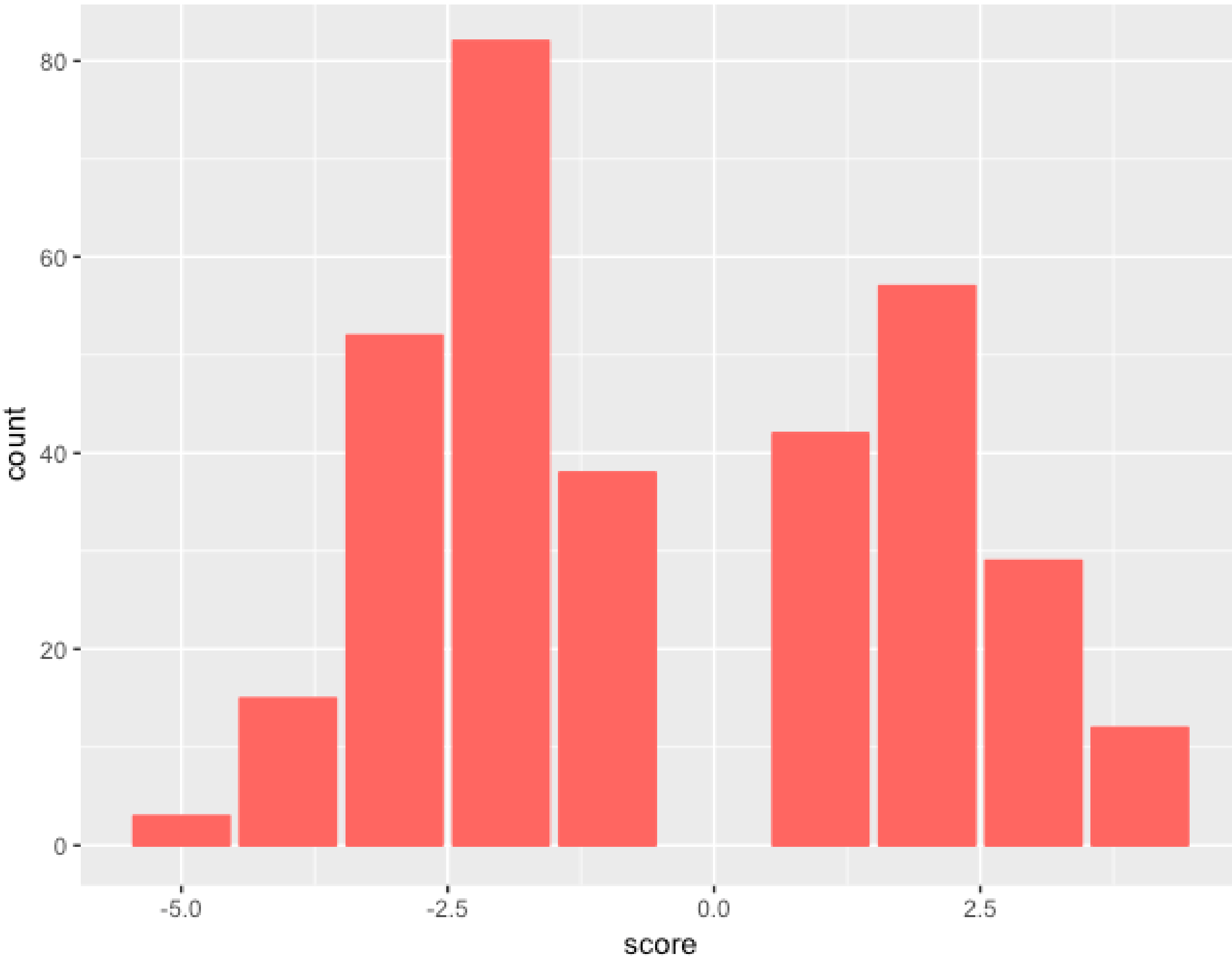
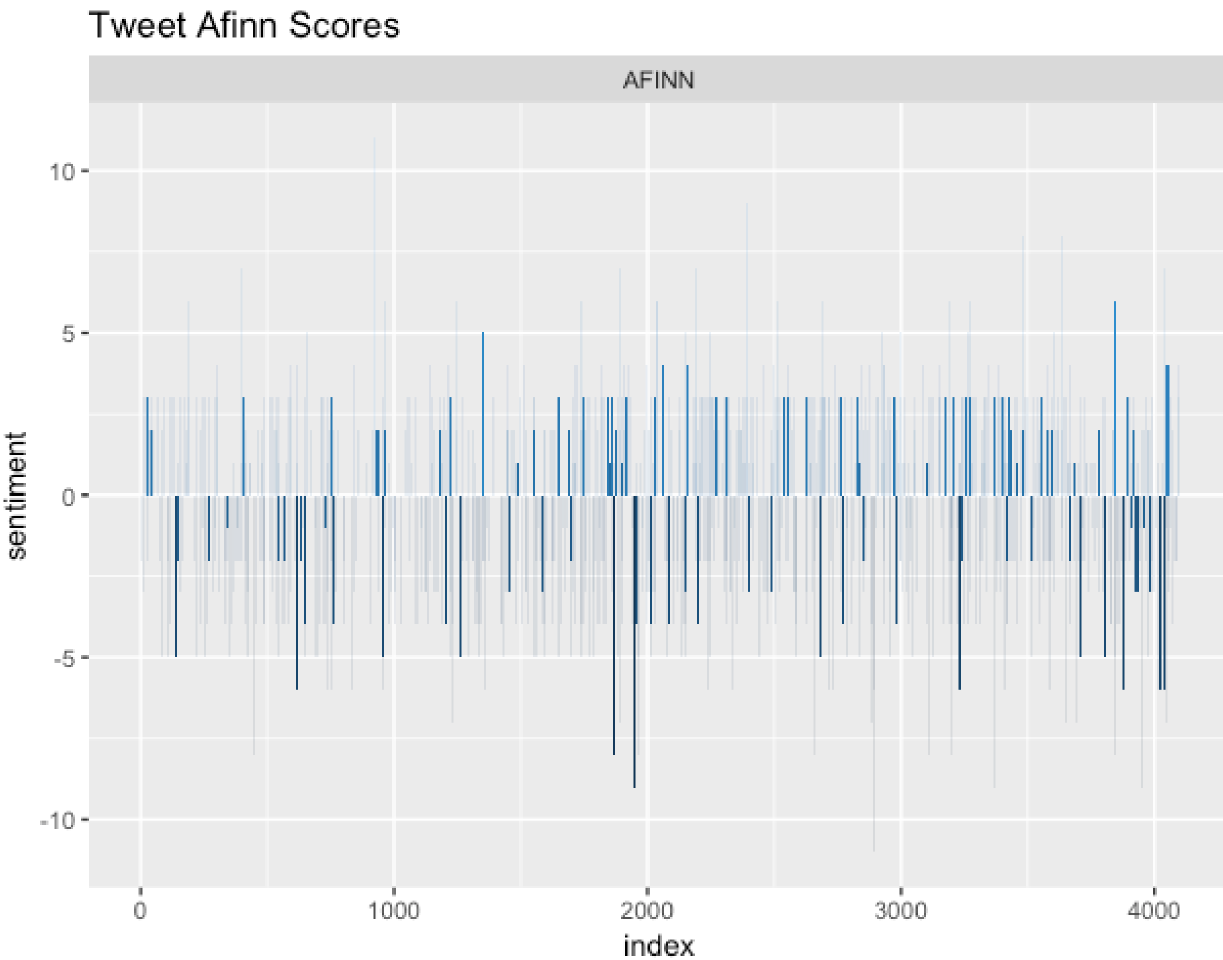


Figure 6:

7. Afinn: By Tweet. The plot below shows the sum afinn sentiment for each tweet (where a score of zero would be a balance of negative and positive sentiment within a tweet).



Shiny App

I also made a Shiny App wherein users could explore the counts of an afinn score (or range of scores) and below a table would display showing which words (that were used in tweets) fit the score criteria, and how many times each appeared.

Conclusions

The sentiment analysis was largely inconclusive. Though it is apparent that both the tweets and the individual words used to discuss Netflix on Twitter were negative, this does not necessarily reflect poorly on the brand. Given the timing of this assignment, and the appearance of "netneutrality" as one of the most common words used in tweets about Netflix, some of the negative sentiment may be closer linked to the recent vote in the FCC to lift regulations on Net Neutrality. Further, Netflix offers a wide array of content and it is more than likely that much of the Twitter discussions referencing Netflix are focused on television shows and movies watched via Netflix, rather than the service Netflix provides. This means that while the sentiment analysis is certainly interesting, results displaying negativity in relation to Netflix's brand on Twitter should be taken with a grain of salt as they likely do not reflect poorly on the brand itself.