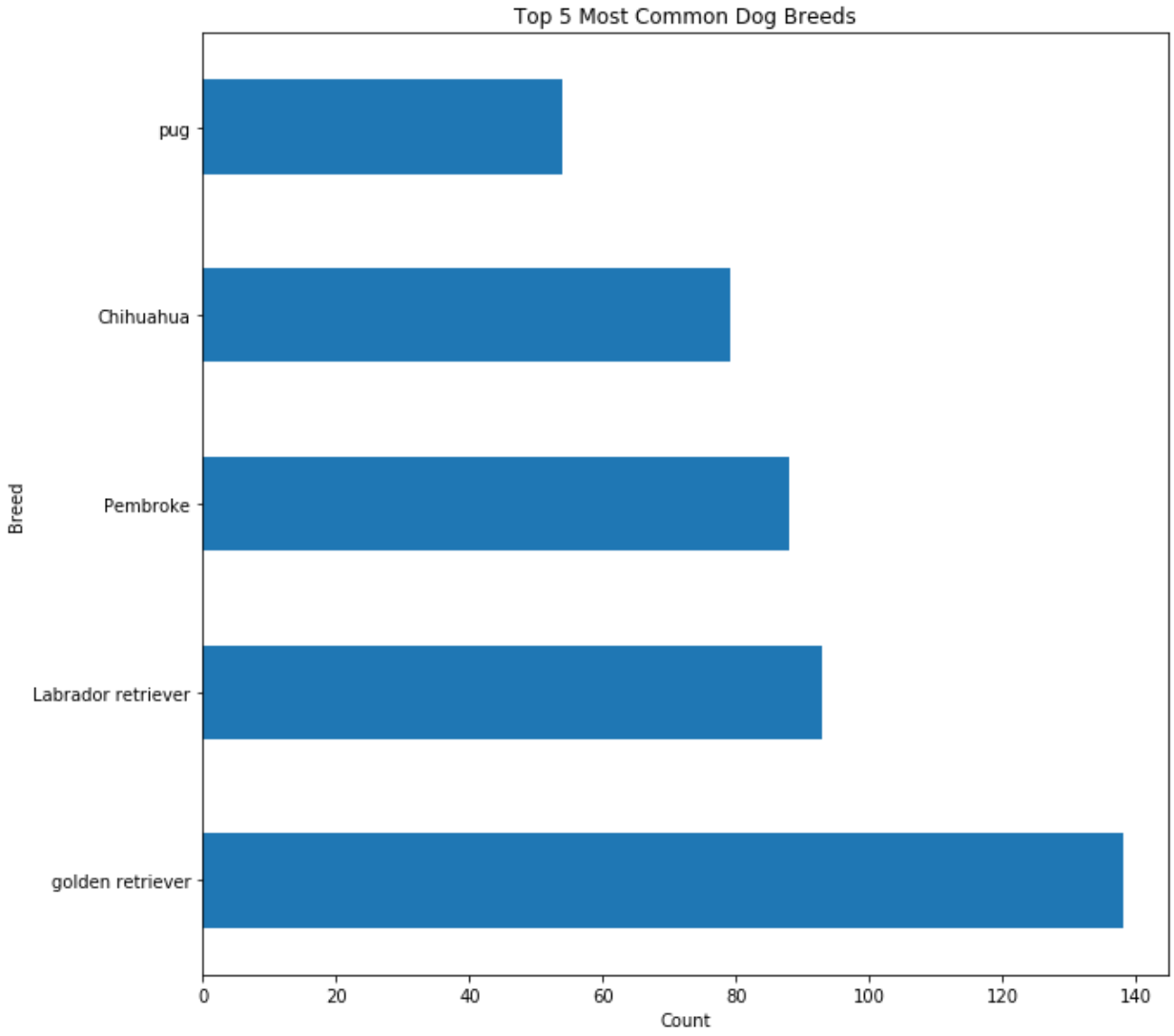
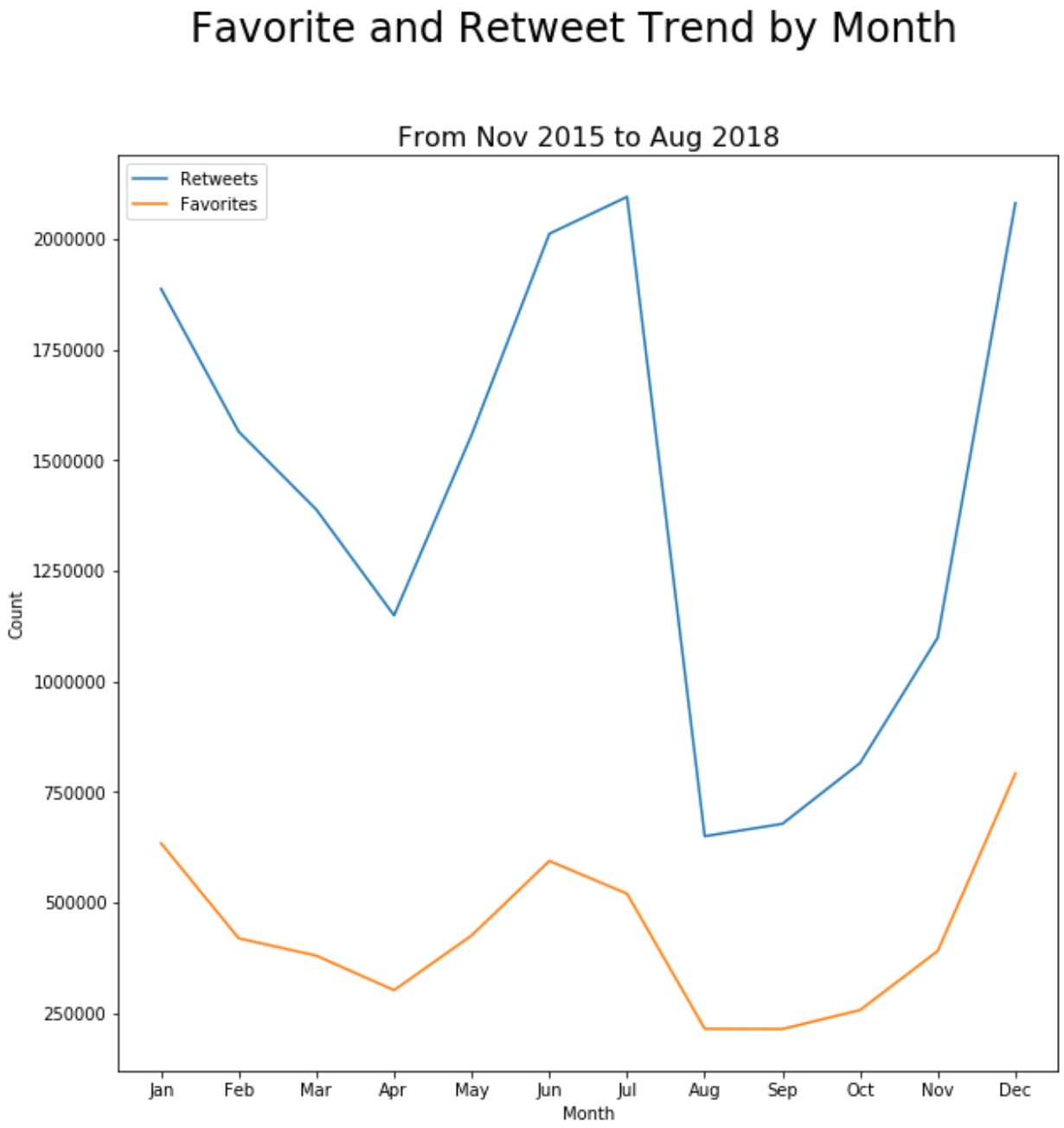
Analysis and Insights into the WeRateDogs Twitter Data



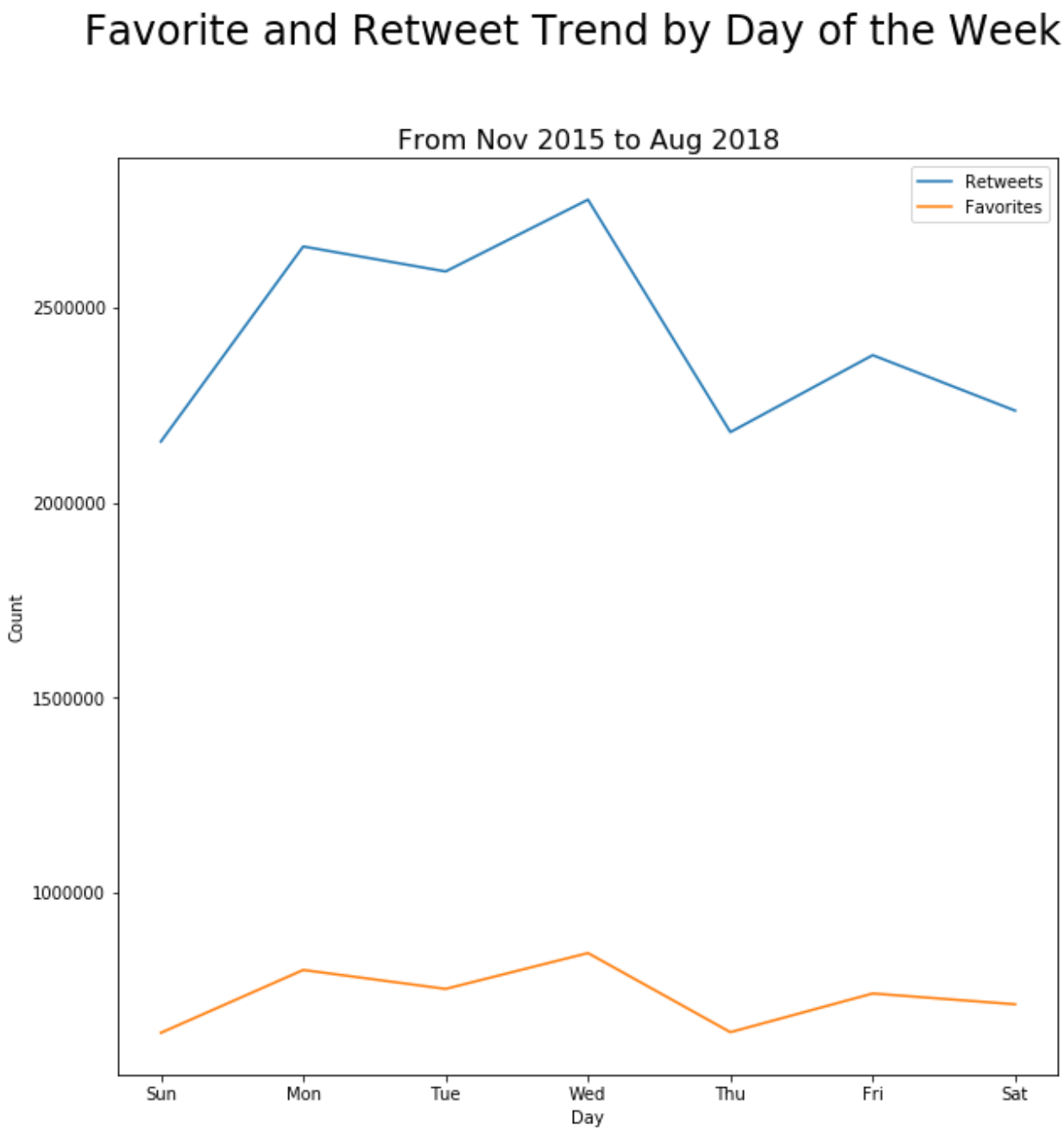
The golden retriever is the most common dog breed.

Some of the tweets in the archive didn't have dogs, so I isolated all the tweets with dogs and picked out the top 5 breeds with the highest counts. The most common dog breed happens to be the golden retriever (138). Following at 2nd place is the Labrador retriever at 93, then the Pembroke at 88, the Chihuahua at 79, and the pug at 54.



Retweet counts peak in July and December, while favorite counts peak in December. Alternatively, both retweet counts and favorite counts trough in August.

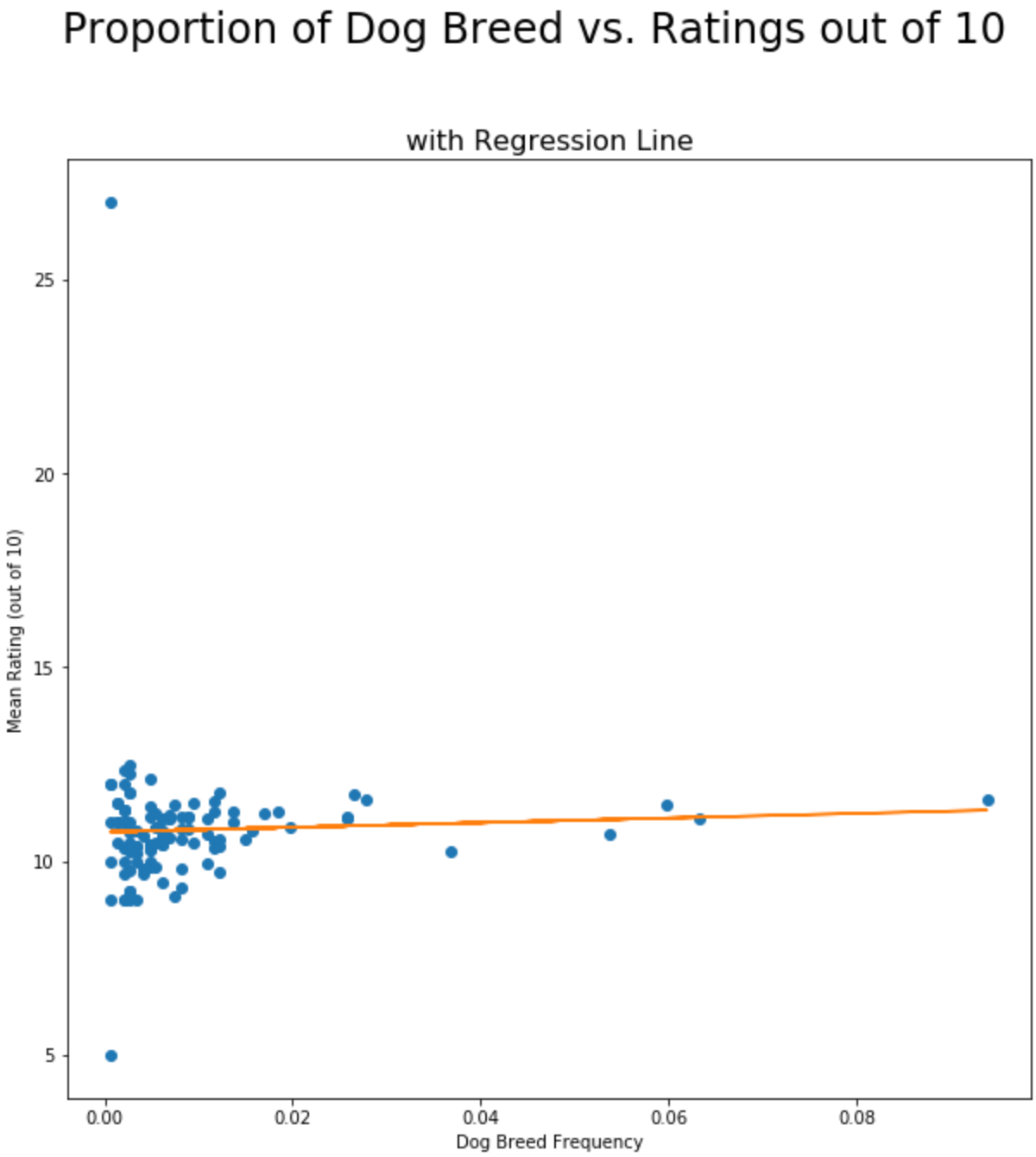
It appears that December is the single month in common where both retweet and favorite counts peak at their highest. Both trends depict a quartic pattern, which implies there is no linearity between the month and number of retweets & favorites.



Retweet counts and favorite counts both peak on Wednesdays. Alternatively, retweet and favorite counts trough on Sundays and Thursdays.

I feel it's no coincidence that these numbers peak during the middle of the week—it *is* hump day, after all. Additionally, it is hard to classify what polynomial trend these 2 series follow.

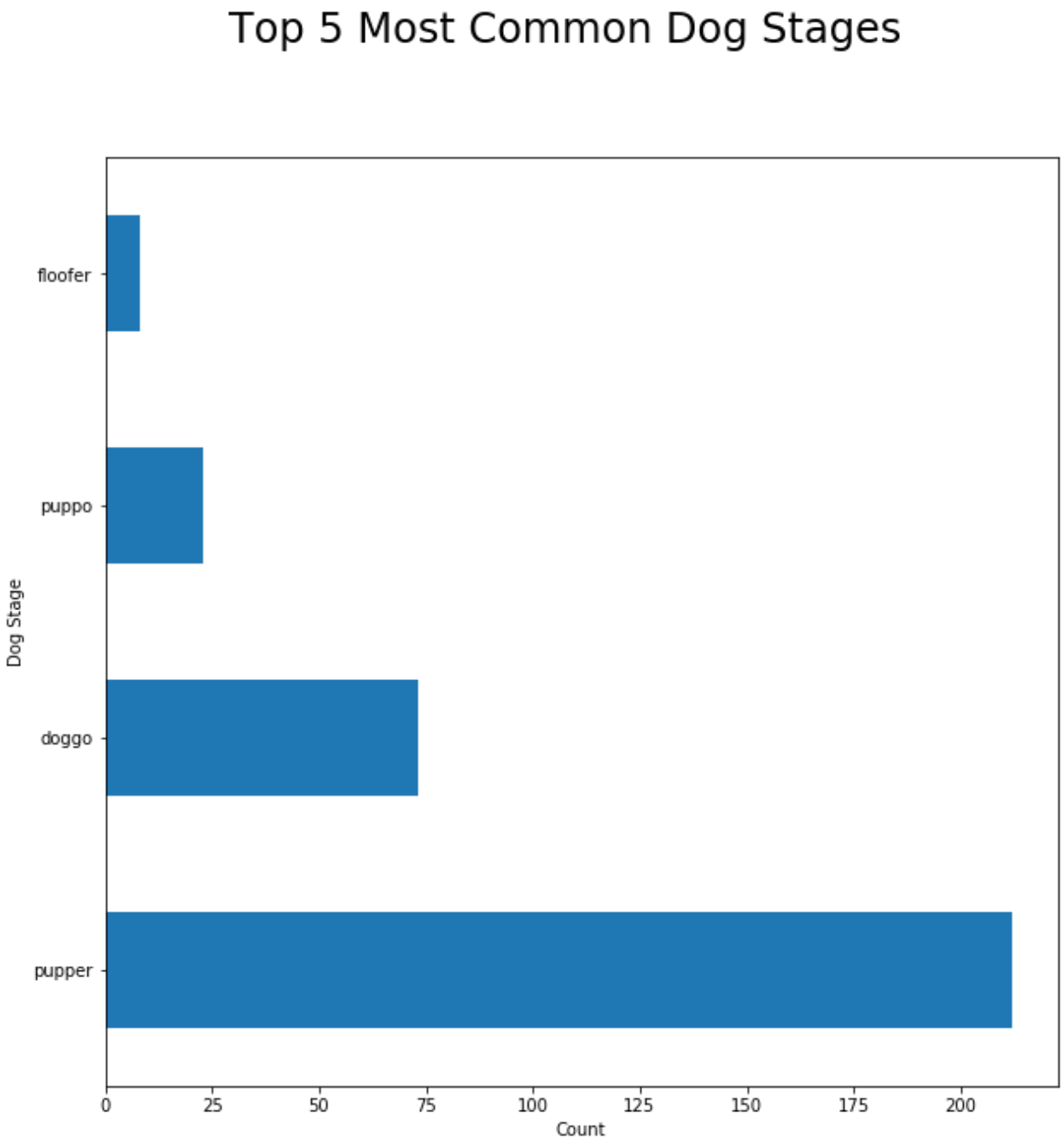
However, the trends appear to be much more steady and predictable than the previous trends for months vs. favorite & retweet counts.



There is no significant correlation between a dog breed's popularity and its rating. I initially thought, "if a dog breed is posted more often, then obviously more people must like it, and thus it will receive higher ratings on average." The regression line of this scatterplot has a slope of 5.93, suggesting that these 2 variables indeed are positively related. However, the p-value obtained from our regression model was 0.648, which suggests this correlation is not significant.

This makes sense largely because the points are largely clustered to the left of the scatterplot around the 0.00 to 0.01 mark. There is an obvious outlier where 1 uncommon dog breed received

a rating of over 25, whereas the most common dog breed (showing up at around 90%) received a mean rating that is within the range of all ratings received by the more rare dog breeds clustered to the left, excluding the outliers.



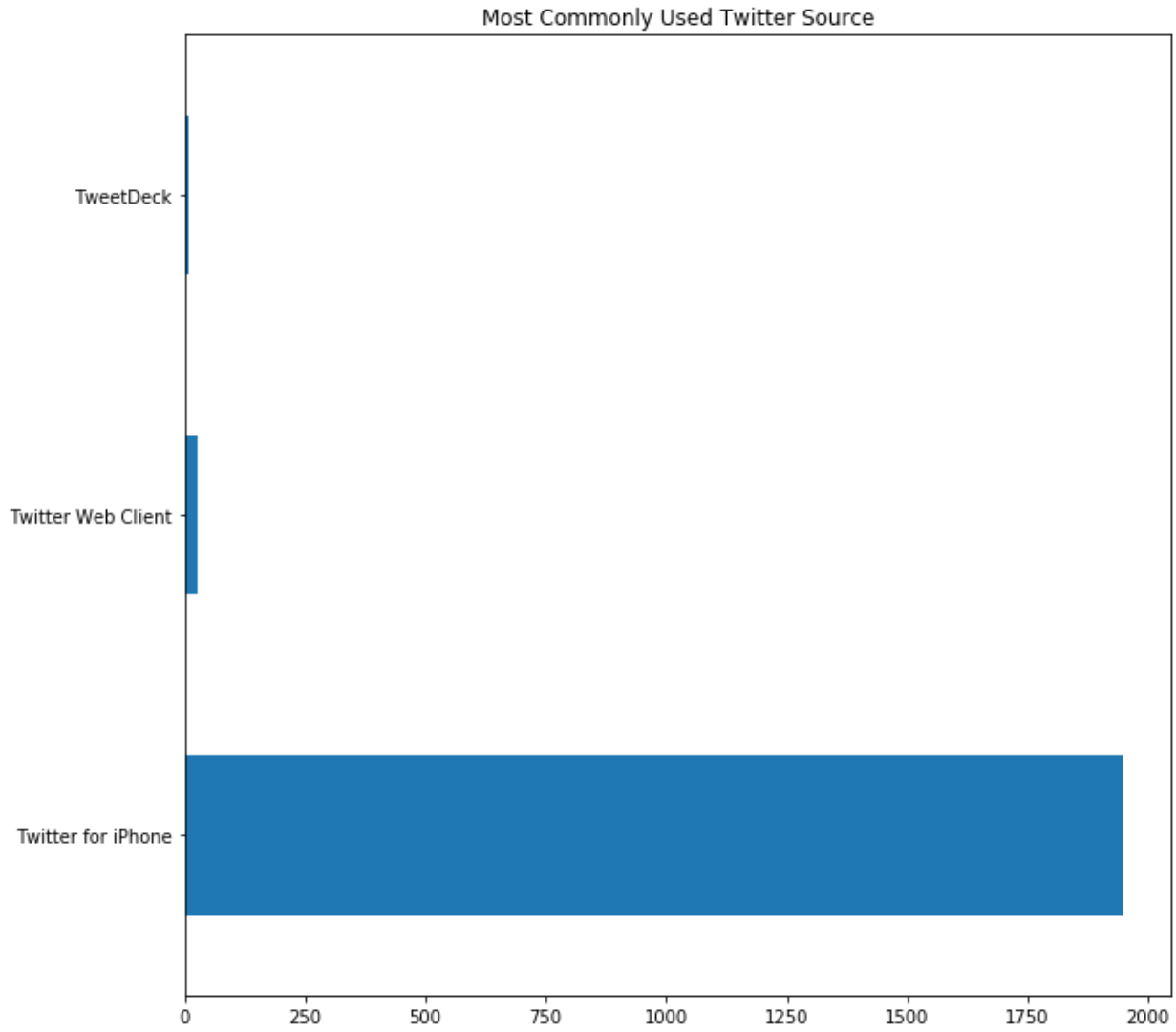
Out of all the dog maturity stages, pupper shows up the most often 212 times. Some of the tweets included more than 1 dog maturity stage, and those are included in the counts for this visualization.

In all tweets for both dogs and non-dogs, "good" is the adjective that occurs the most frequently. The first visual depicts the most common adjectives present in the tweets for dogs, while the second is for non-dogs.

Initially, I assumed that more negative adjectives would show up for non-dogs while more positive adjectives would show up for dogs. Although this appears to be mostly false (with the exception of "average" in the second visual), there is a fundamental concept I had not considered when analyzing these words.

I only kept track of the number of "occurrences" that each adjective shows up in the text of each tweet. This implementation ignores a large part of semantics. For example, the word "good" shows up the most frequently but we don't keep track of the difference between "not good" and "h*ckin' good." Doing so would require adjective classification based on the words preceding/proceeding from the adjective, which would require exploring more into natural language processing. This would be far outside the scope of this project (but not of that for another day!).

Therefore, these words are counts of ***how often*** adjectives appeared in the tweets without context or anything in between. If the question were instead tailored to be, "Which adjectives most often ***describe*** dogs?" then that would be a different story.



Twitter for iPhone is the most highly-used source to post tweets. The proportion of Twitter for iPhone as a source is stark in comparison to Twitter Web Client and TweetDeck, where WeRateDogs posts tweets using the iPhone 98% of the time.