

Twitter Analysis

Tony

Sun Oct 01 20:37:57 2017

- [Introduction](#)
- [Tweet Volume](#)
- [Tweet Behavior](#)
- [Tweet Content](#)
 - [Word Frequency and Usage](#)
- [Tweet Popularity](#)
- [Sentiment Analysis](#)
- [Conclusion](#)

Introduction

This report compares the volume, behavior, and content of the tweets made by BarstoolBigCat and PFTCommenter.

3195 tweets were originally collected for BarstoolBigCat. 3195 tweets were originally collected for PFTCommenter.

The oldest tweet collected from BarstoolBigCat is from 2017-08-13 18:56:08 and the oldest tweet from PFTCommenter is from 2017-03-16 11:32:41. The most recent tweets are from 2017-09-30 17:41:31 and 2017-09-30 14:57:24

The two sets of tweets were trimmed to 3189 and 3194 tweets respectively in order to align the dates of the last collected tweets.

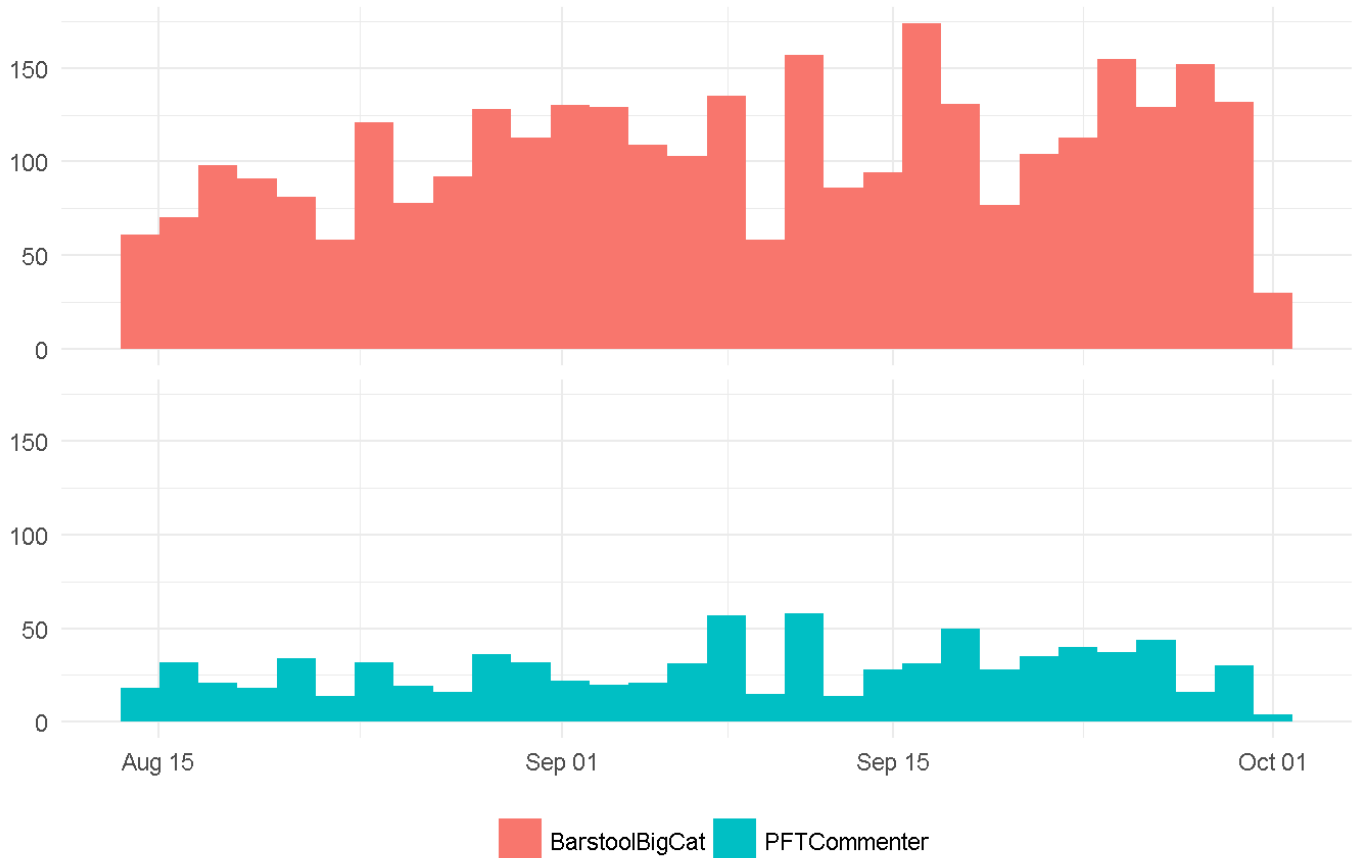
Because the number of tweets for at least one of the people is less than the threshold 3000, the data sets were resized such that they cover the same periods of time. The number of tweets from BarstoolBigCat and PFTCommenter were reduced to 3189 and 853.

Tweet Volume

How often do BarstoolBigCat and PFTCommenter tweet? Does the volume of tweets look different for temporal periods?

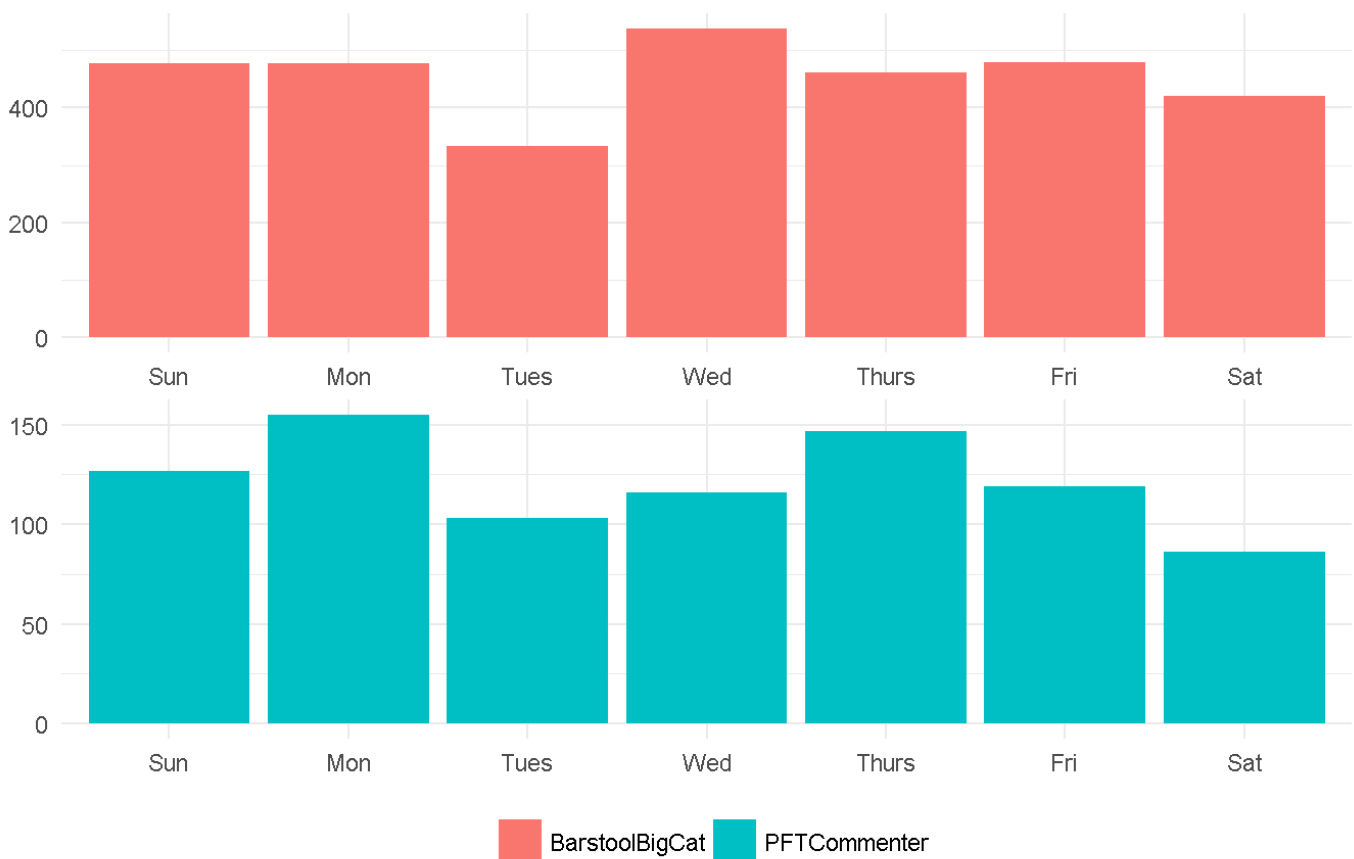
Count of Tweets Over Time

From 2017-08-13 18:56:08 to 2017-09-30 14:57:24



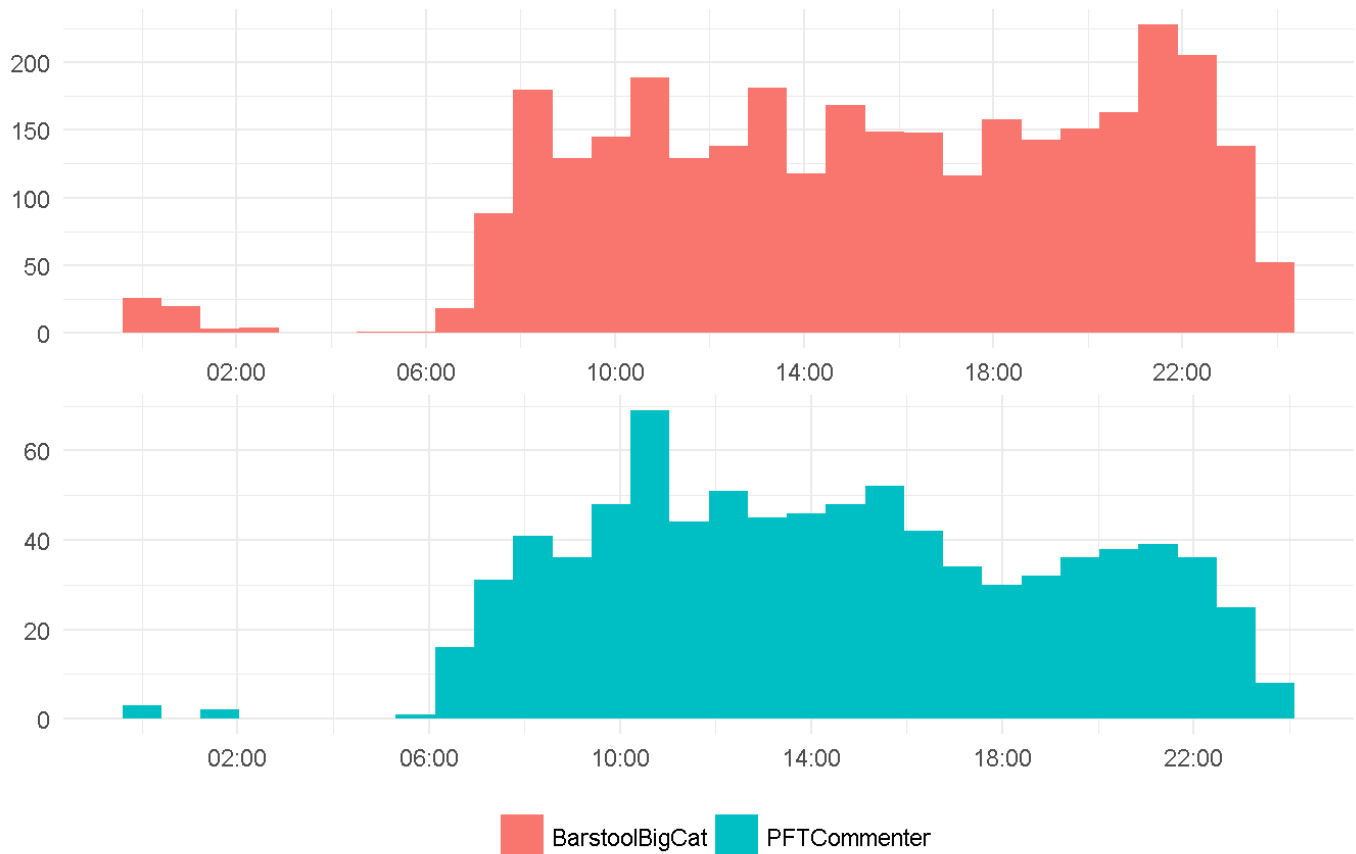
Count of Tweets Over Time

Grouped By Day of Week



Count of Tweets Over Time

Grouped By Hour of Day



Is the distribution of our volume of tweets given a certain temporal period statistically significant? Here, I use the Chi-Squared Test. If the p-value is calculated to be less than some threshold value (e.g. 0.05), then I can deduce that the null hypothesis (that the distribution is uniform) is invalid. In fact, it appears that our tweet volume does differ depending on the month and day of the week.

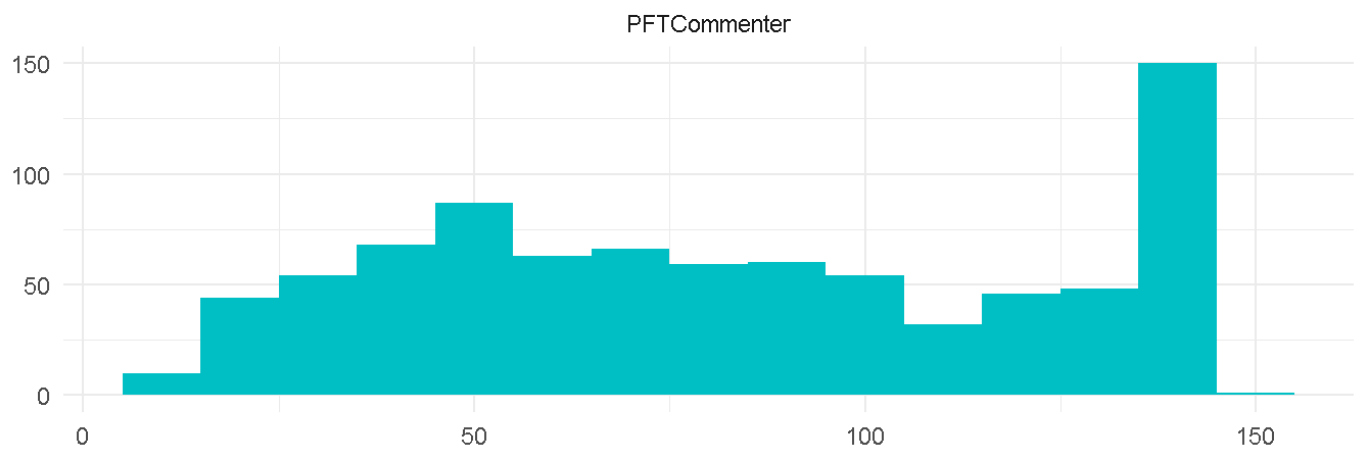
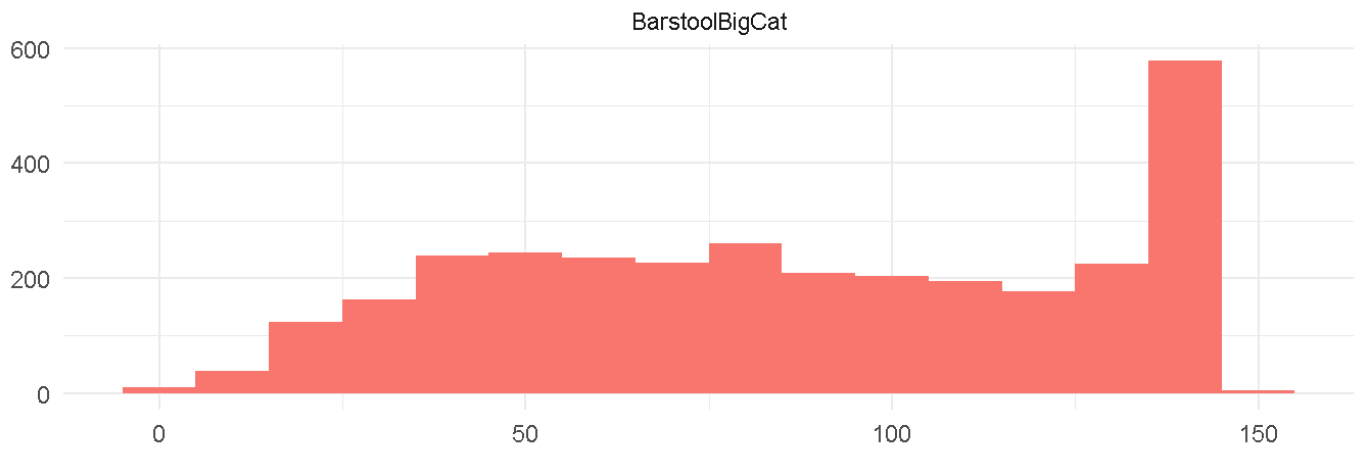
Tweet Behavior

What proportion of tweets include more than just plain text (e.g. hashtags, links, etc.)? What proportion are not undirected, self-authored tweets (i.e. RTs or replies)?

Tweet Content

How long are the tweets?

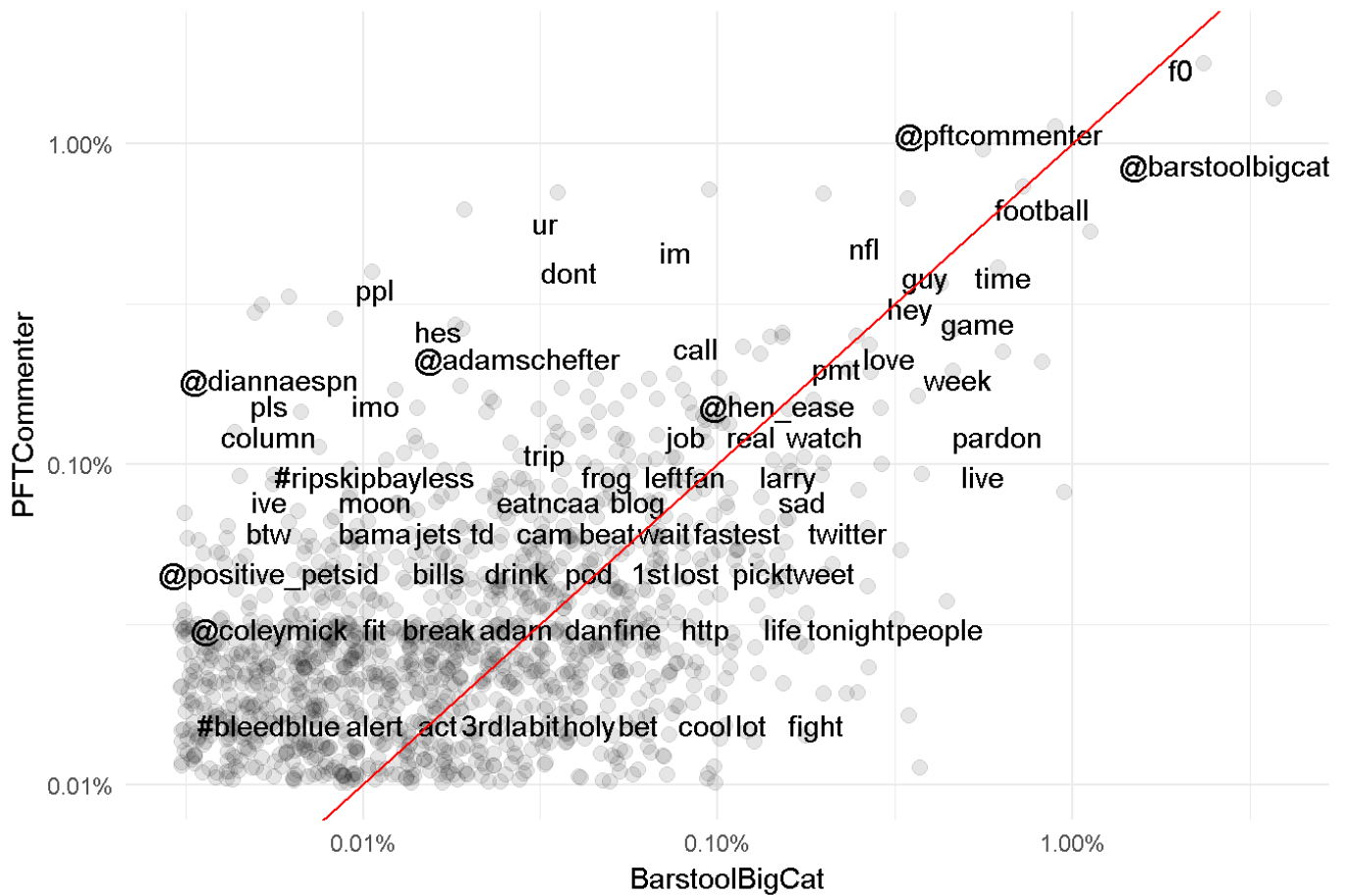
Distribution of # of Characters in Tweets



Word Frequency and Usage

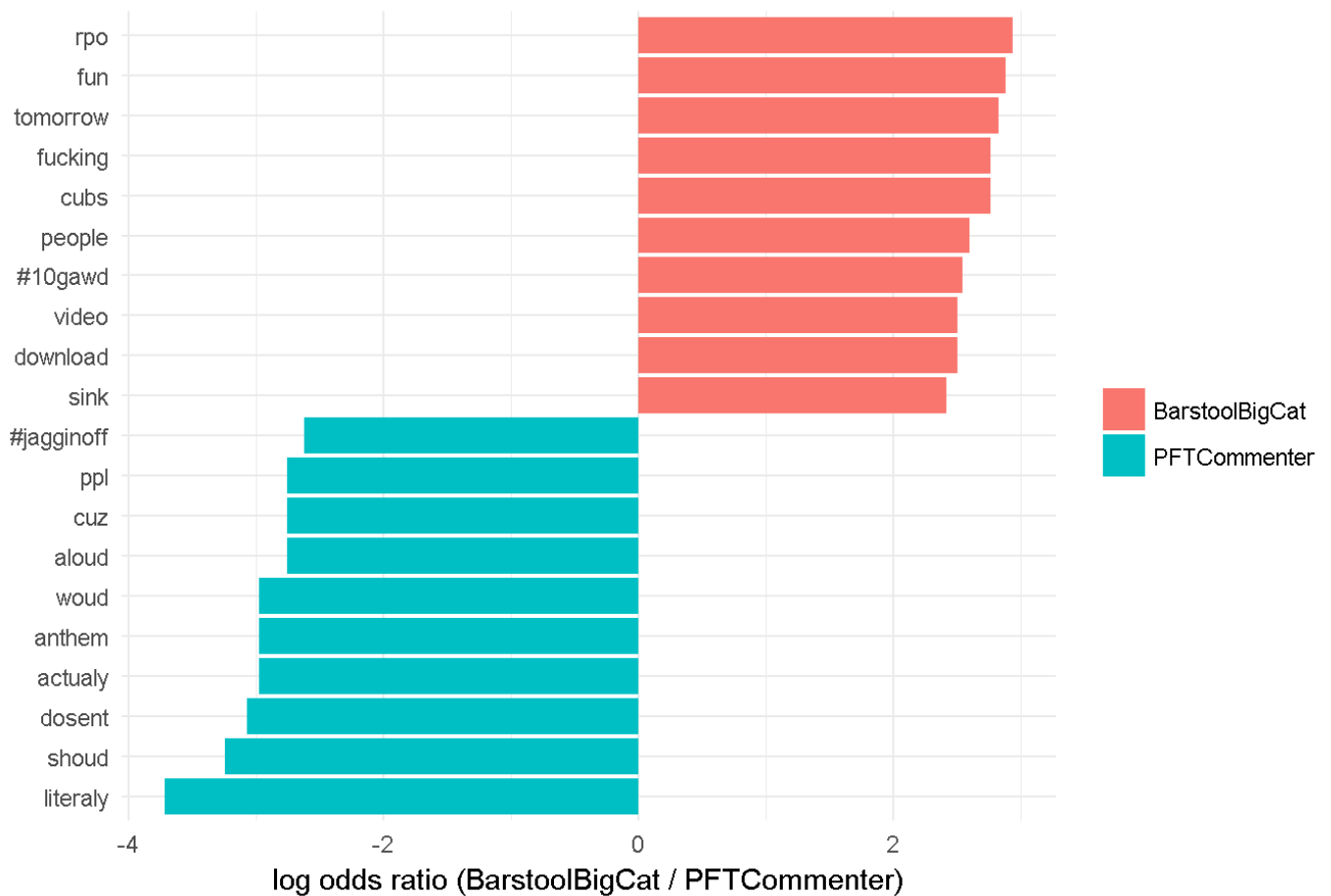
Which words are used most frequently?

Relative Word Frequency



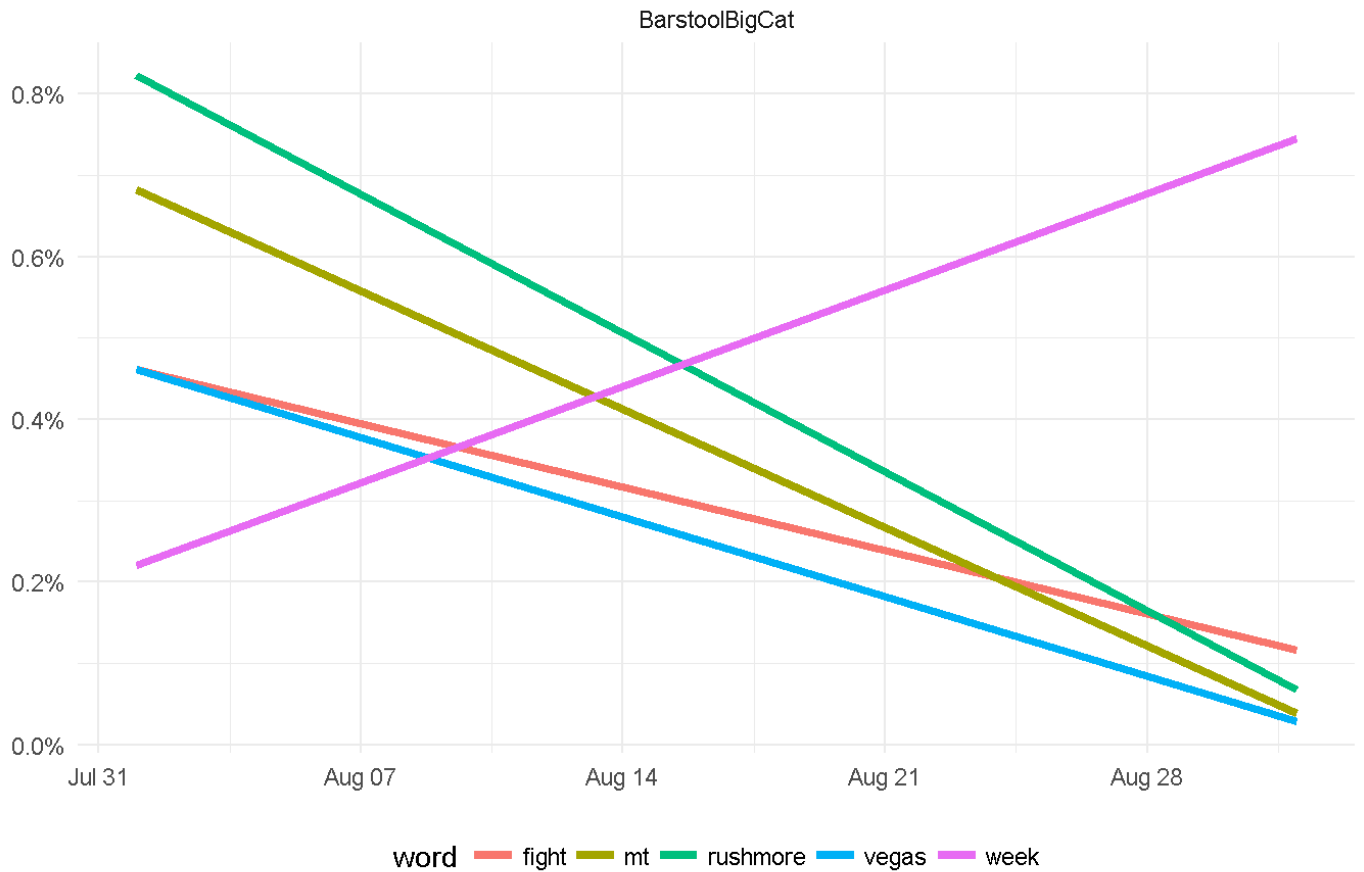
Which words are most likely to be used by one person compared to the other?

Words Most Unique to Each Person



Which words have have been used more and less frequently over time?

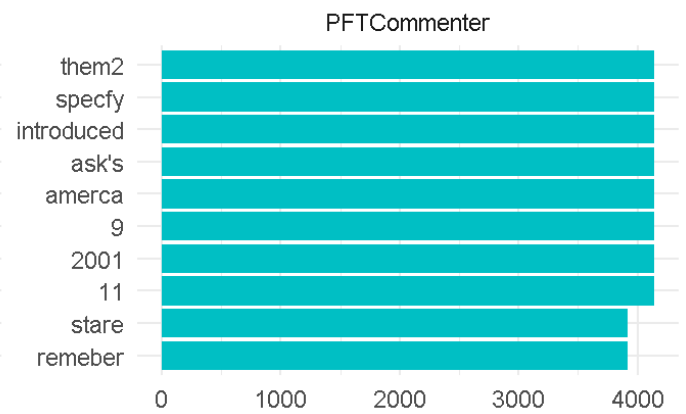
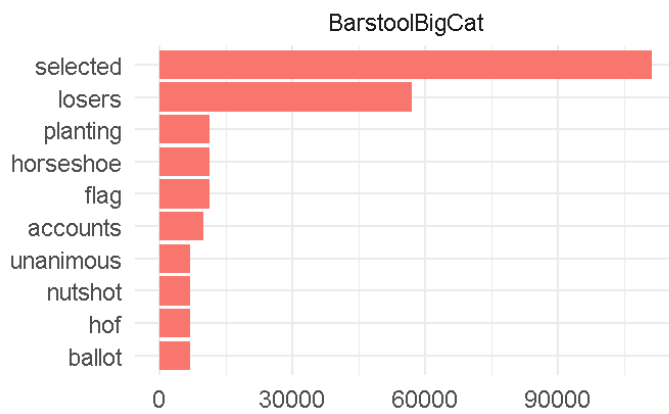
Largest Changes in Word Frequency



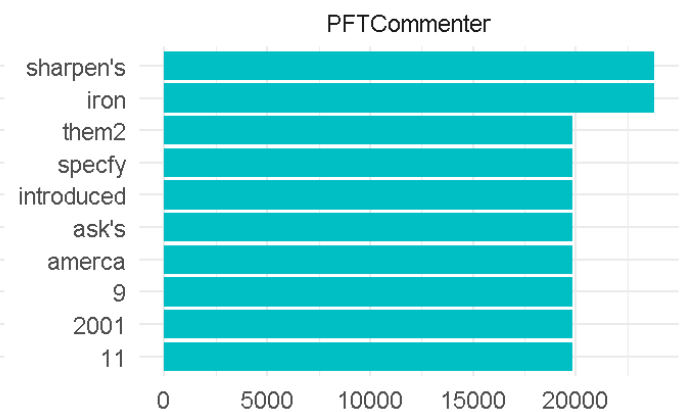
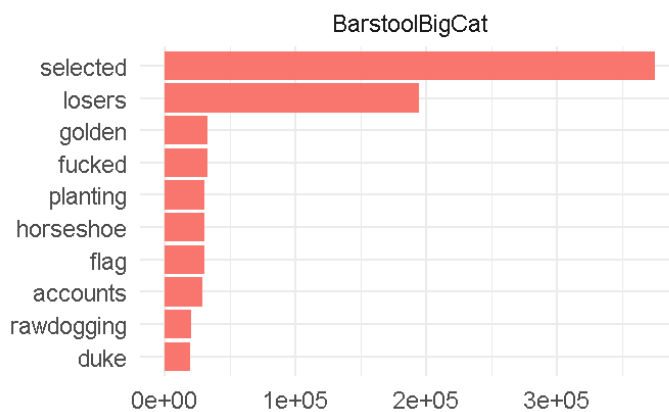
Tweet Popularity

How often do the original tweets get liked/favorited/retweeted?

Words with Highest Median # of RTs



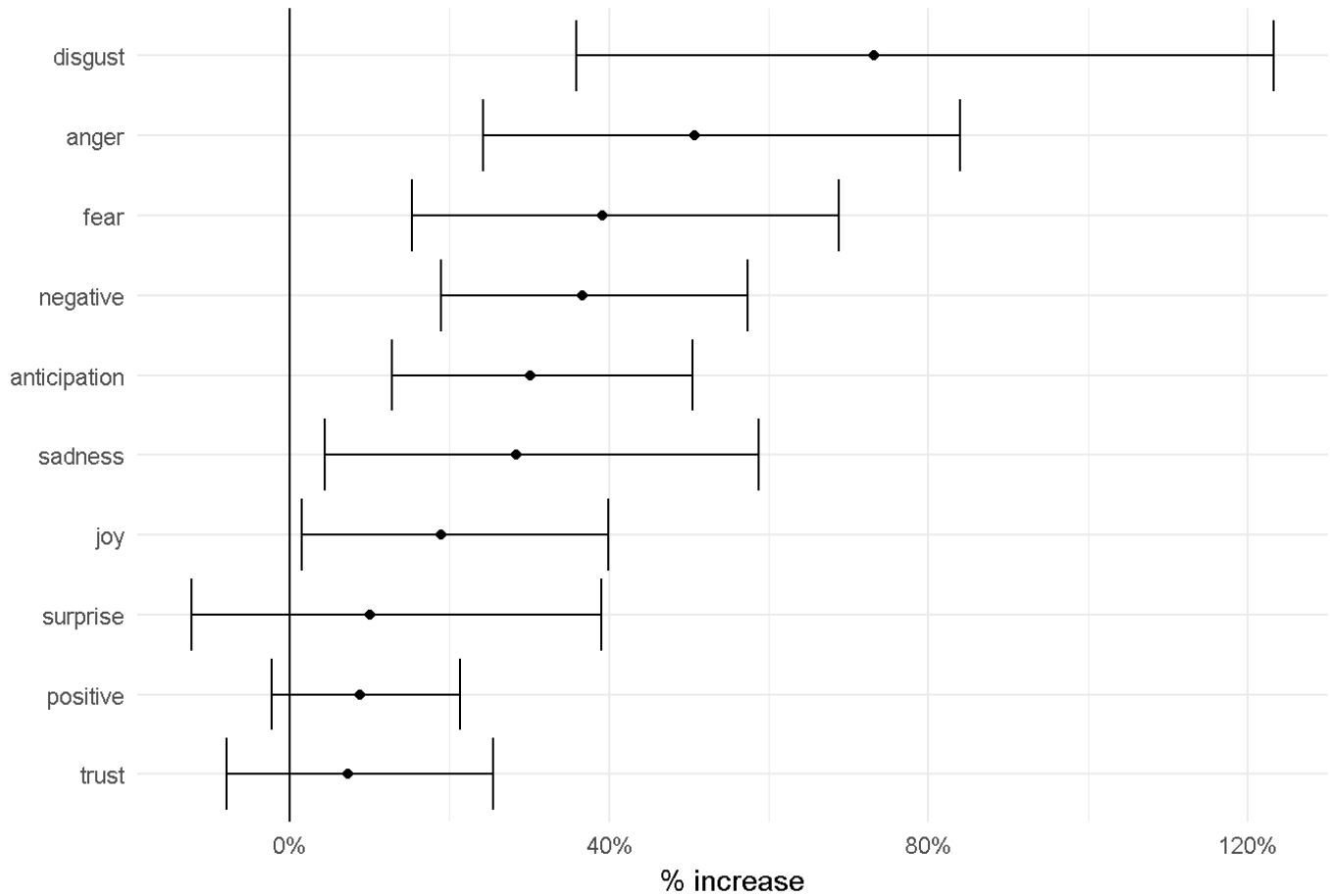
Words with Highest Median # of Favorites



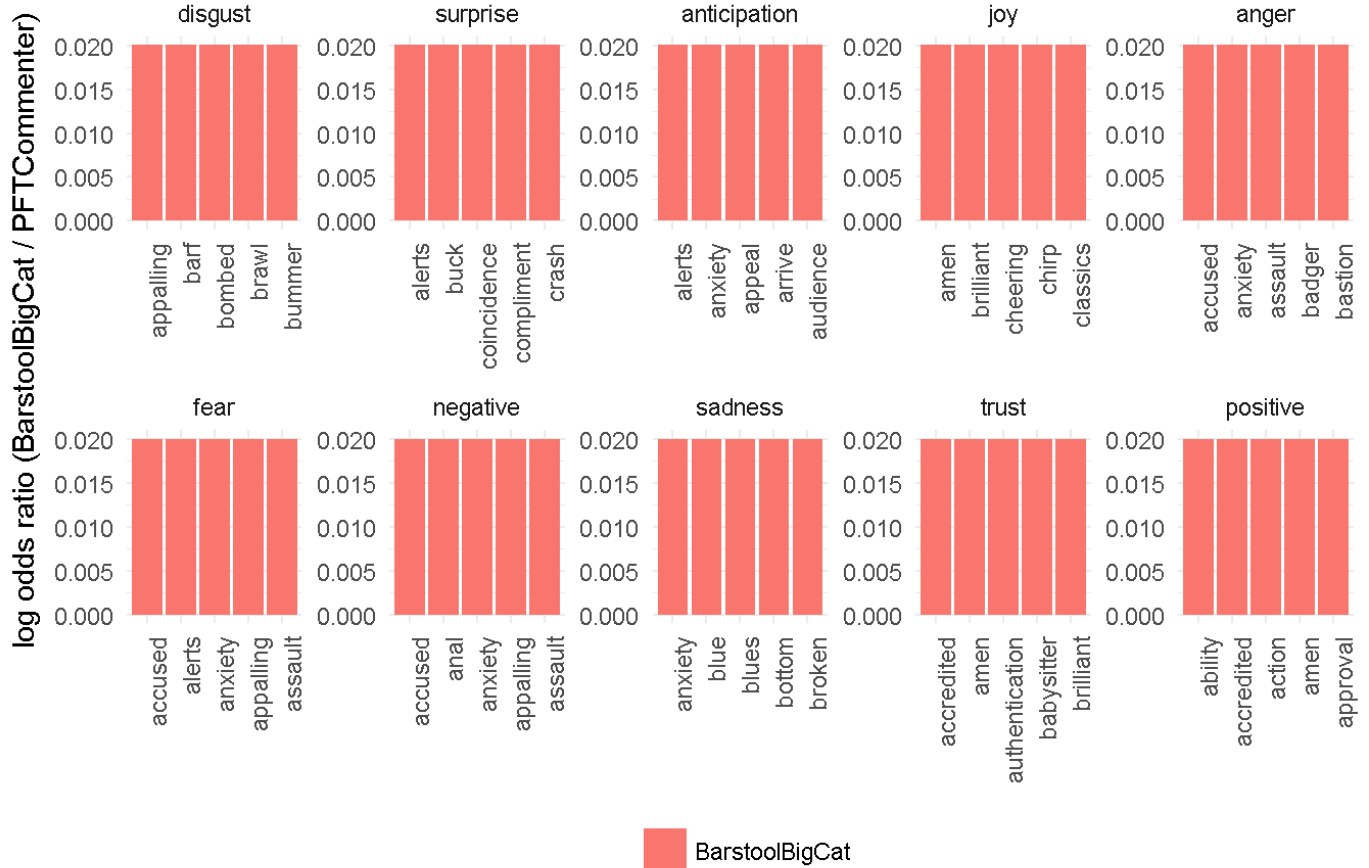
Sentiment Analysis

What is the sentiment (i.e. "tone") of the tweets?

Sentiment Analysis of BarstoolBigCat and PFTCommenter



Most Influential Words Contributing to Sentiment Differences



Conclusion

That's it!