Twitter Analysis

Tony

Sun Oct 01 20:32:53 2017

Introduction

This report compares the volume, behavior, and content of the tweets made by RealSkipBayless and stephenasmith.

3200 tweets were orignally collected for RealSkipBayless. 3195 tweets were originally collected for stephenasmith.

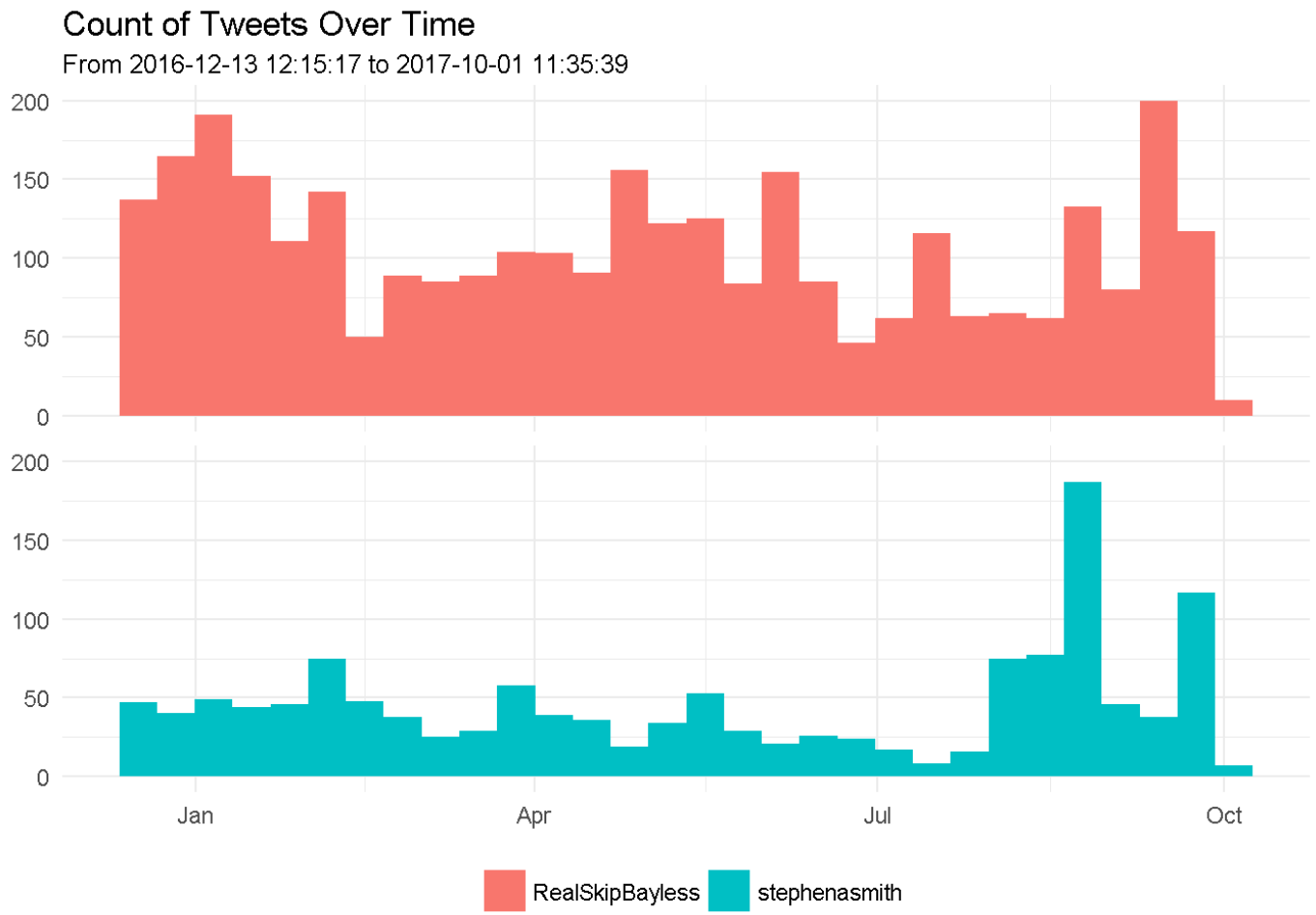
The oldest tweet collected from RealSkipBayless is from 2016-12-13 12:15:17 and the oldest tweet from stephenasmith is from 2016-04-21 17:24:52. The most recent tweets are from 2017-10-01 12:59:30 and 2017-10-01 11:35:39

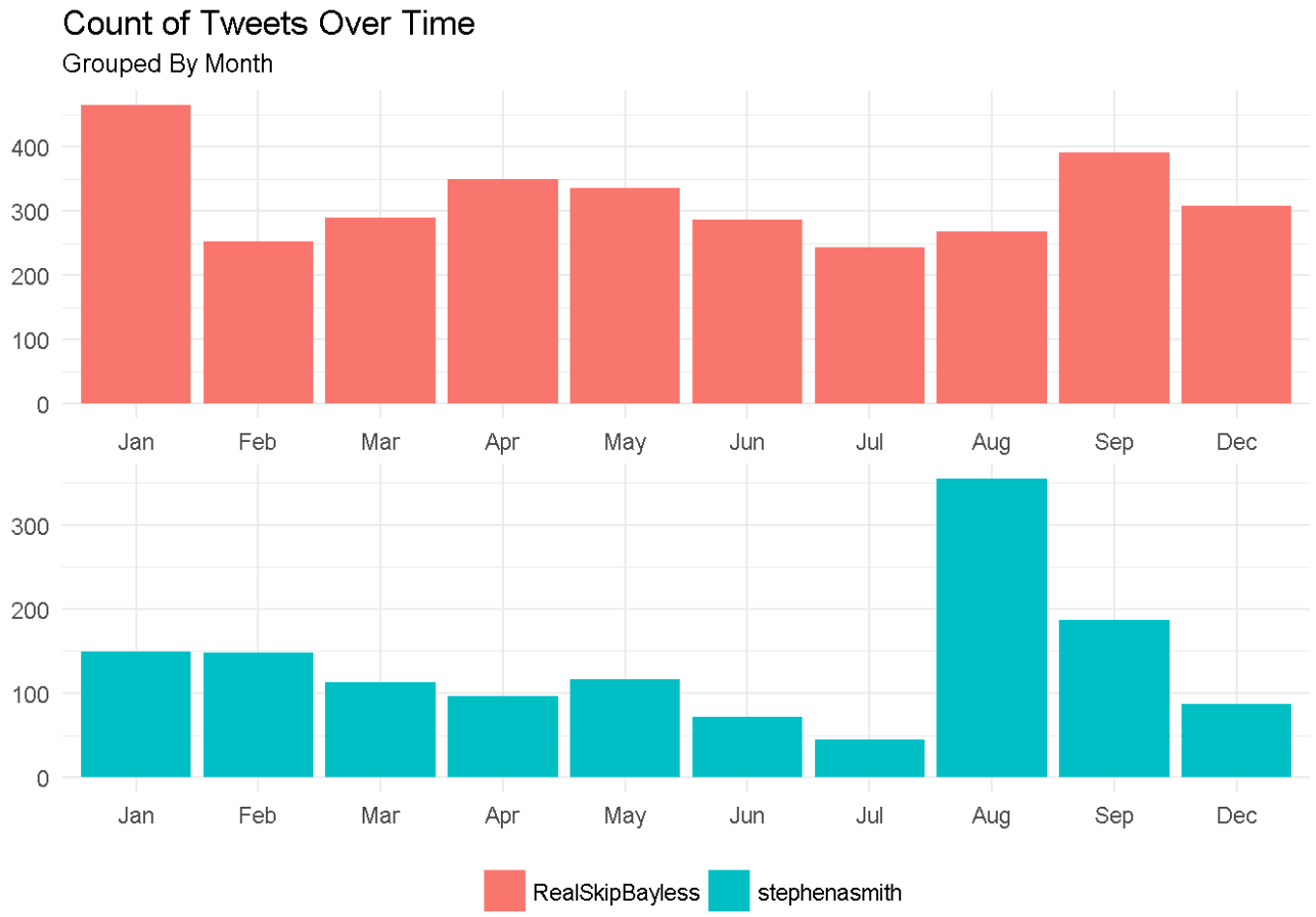
The two sets of tweets were trimmed to 3190 and 3194 tweets respectively in order to align the dates of the last collected tweets.

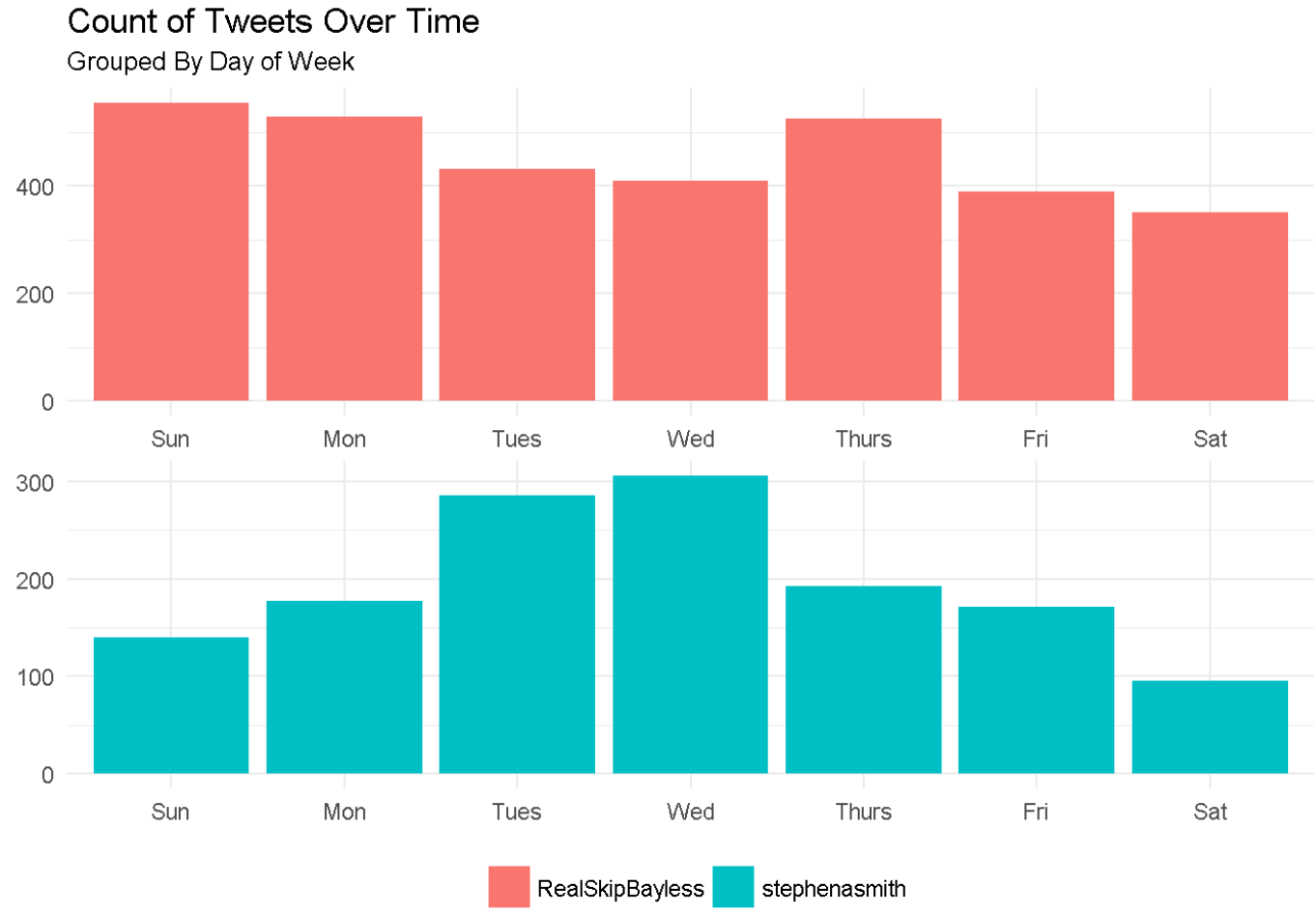
Because the number of tweets for at least one of the people is less than the threshhold 3000, the data sets were resized such that they cover the same periods of time. The number of tweets from RealSkipBayless and stephenasmith were reduced to 3190 and 1368.

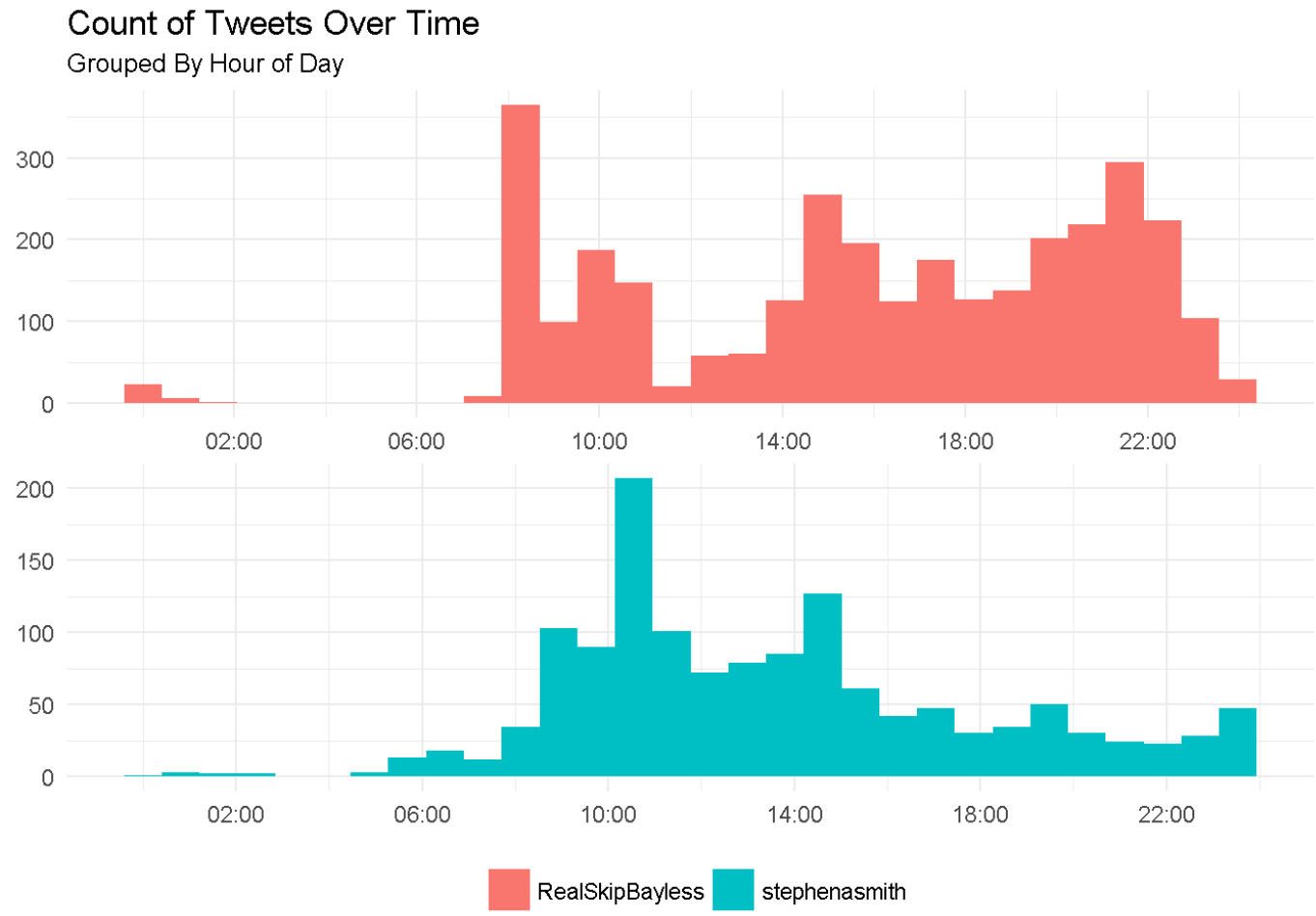
Tweet Volume

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


Count of Tweets Over Time
From 2016-12-13 12:15:17 to 2017-10-01 11:35:39
200
150
100
50
0
200
150
100
50
0
Jan
Apr
Jul
Oct
RealSkipBayless
stephenasmith


Count of Tweets Over Time
Grouped By Month
400
300
200
100
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Dec
300
200
100
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Dec
RealSkipBayless
stephenasmith





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 thatn some threshold value (e.g. 0.05), then I

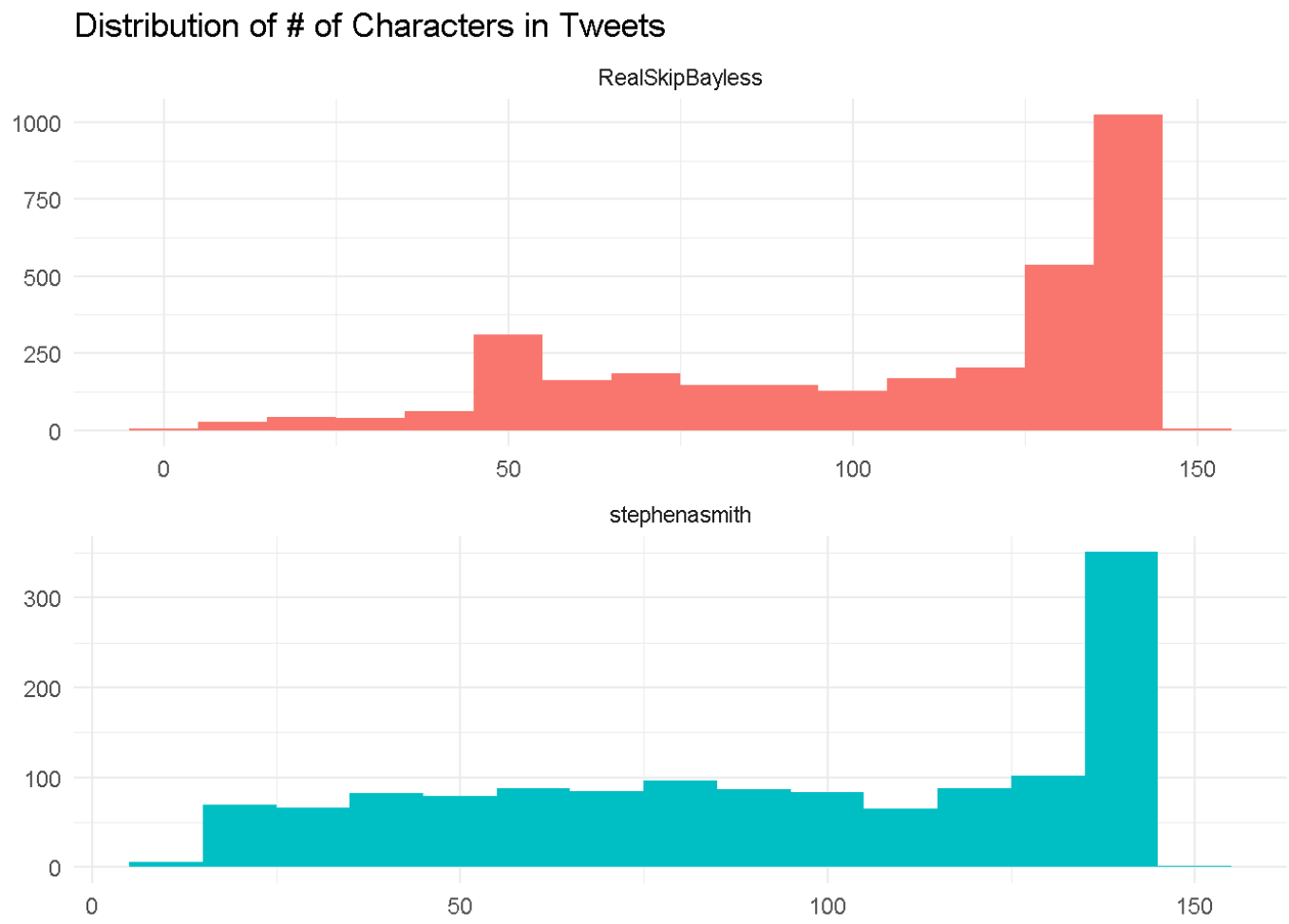
can deduce that the the null hypothes (that the distribution is uniform) is invalid. In fact, it appears that our tweet volume does differ depending on the month and day of the week.

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)?

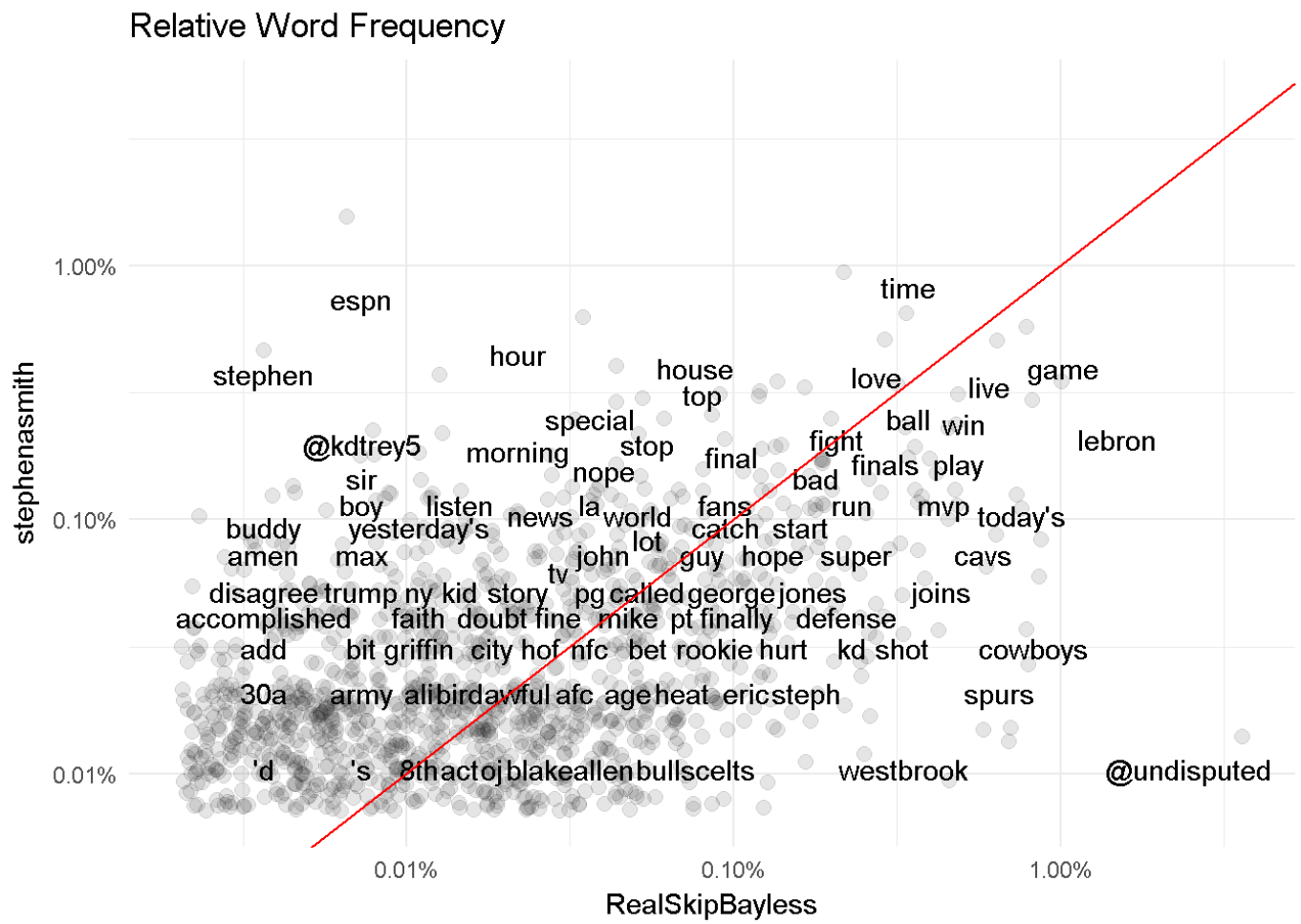
Tweet Content

How long are the tweets?

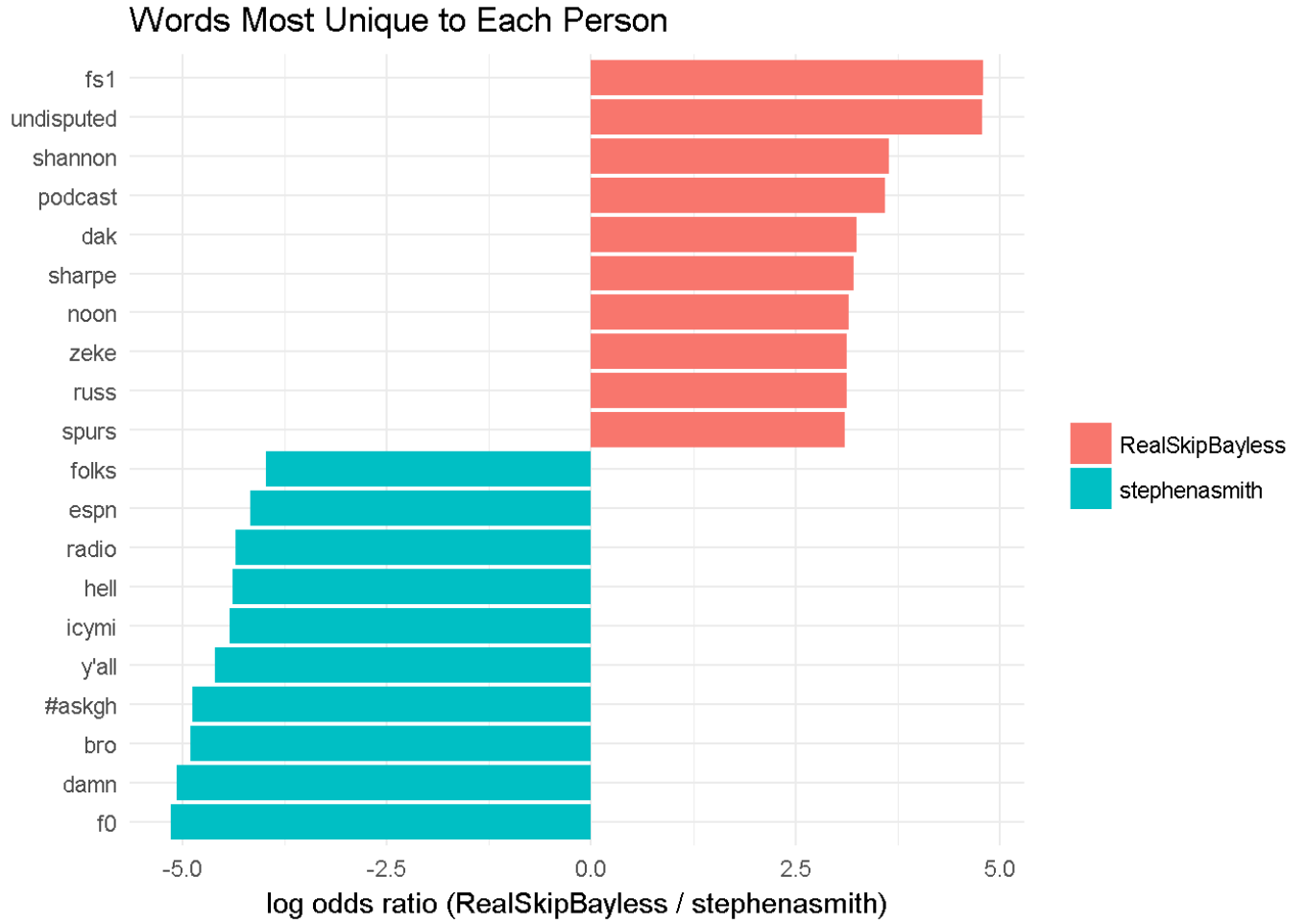


Word Frequency and Usage

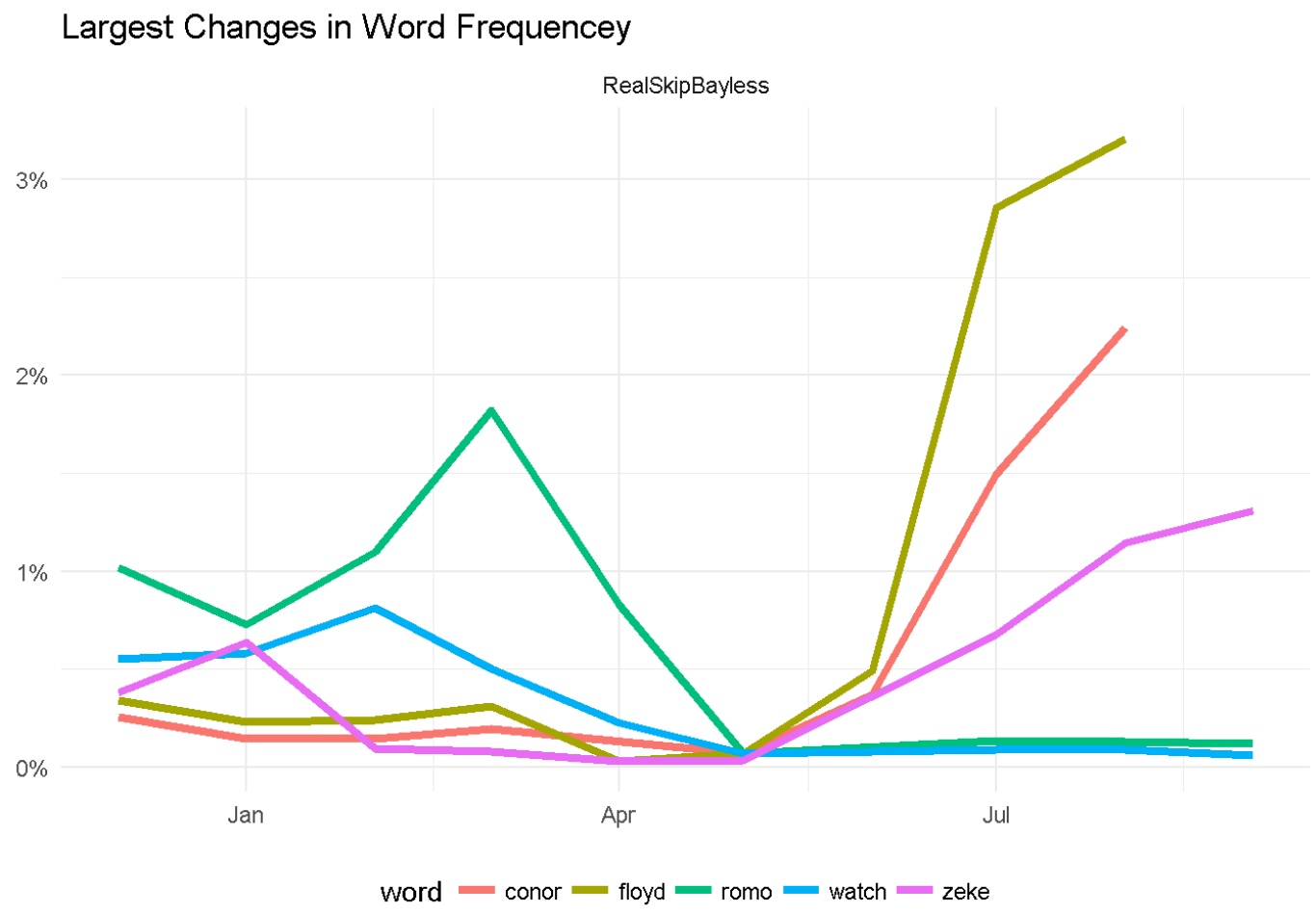
Which words are used most frequently?



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

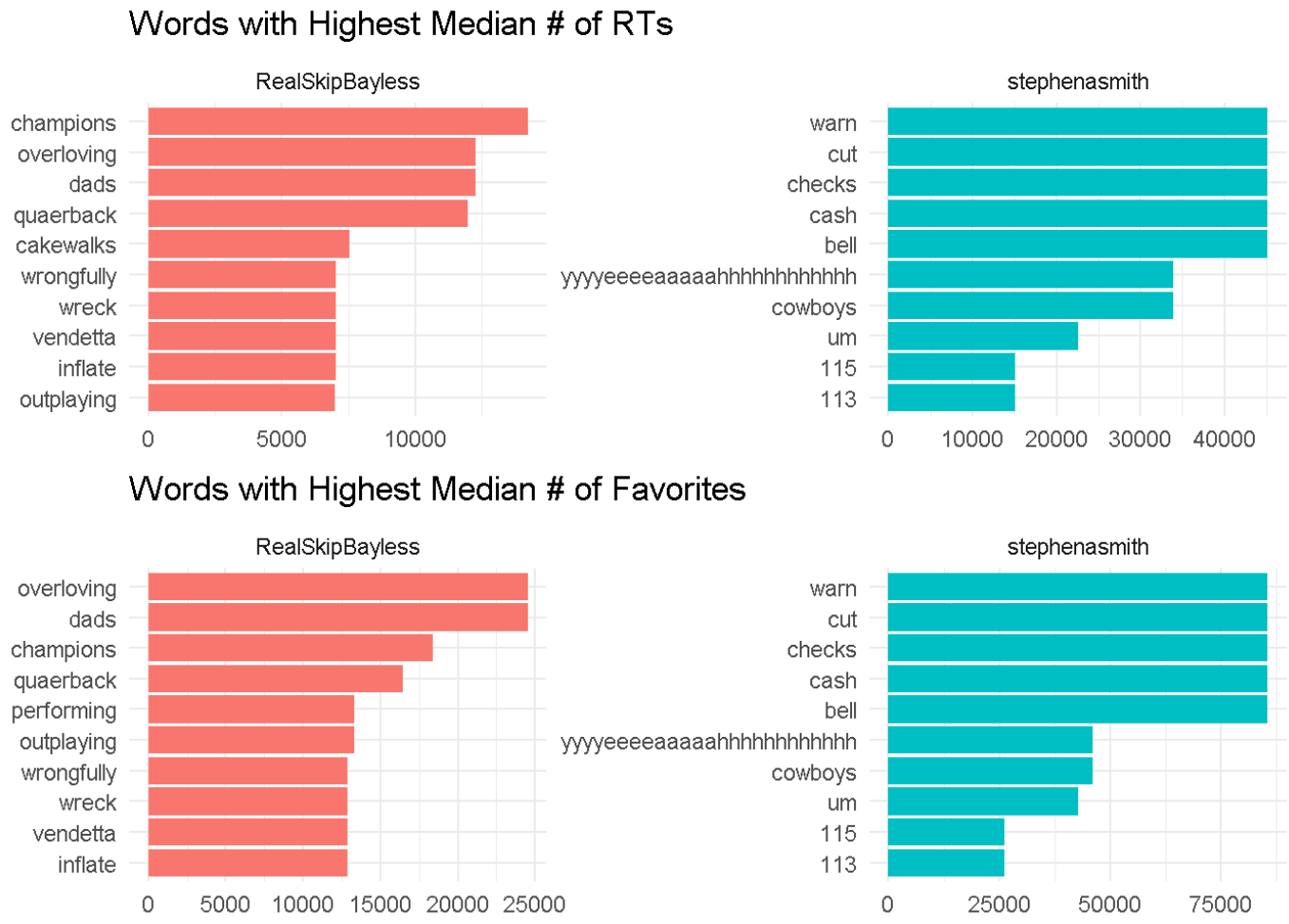


Which words have have been used more and less frequently over time?



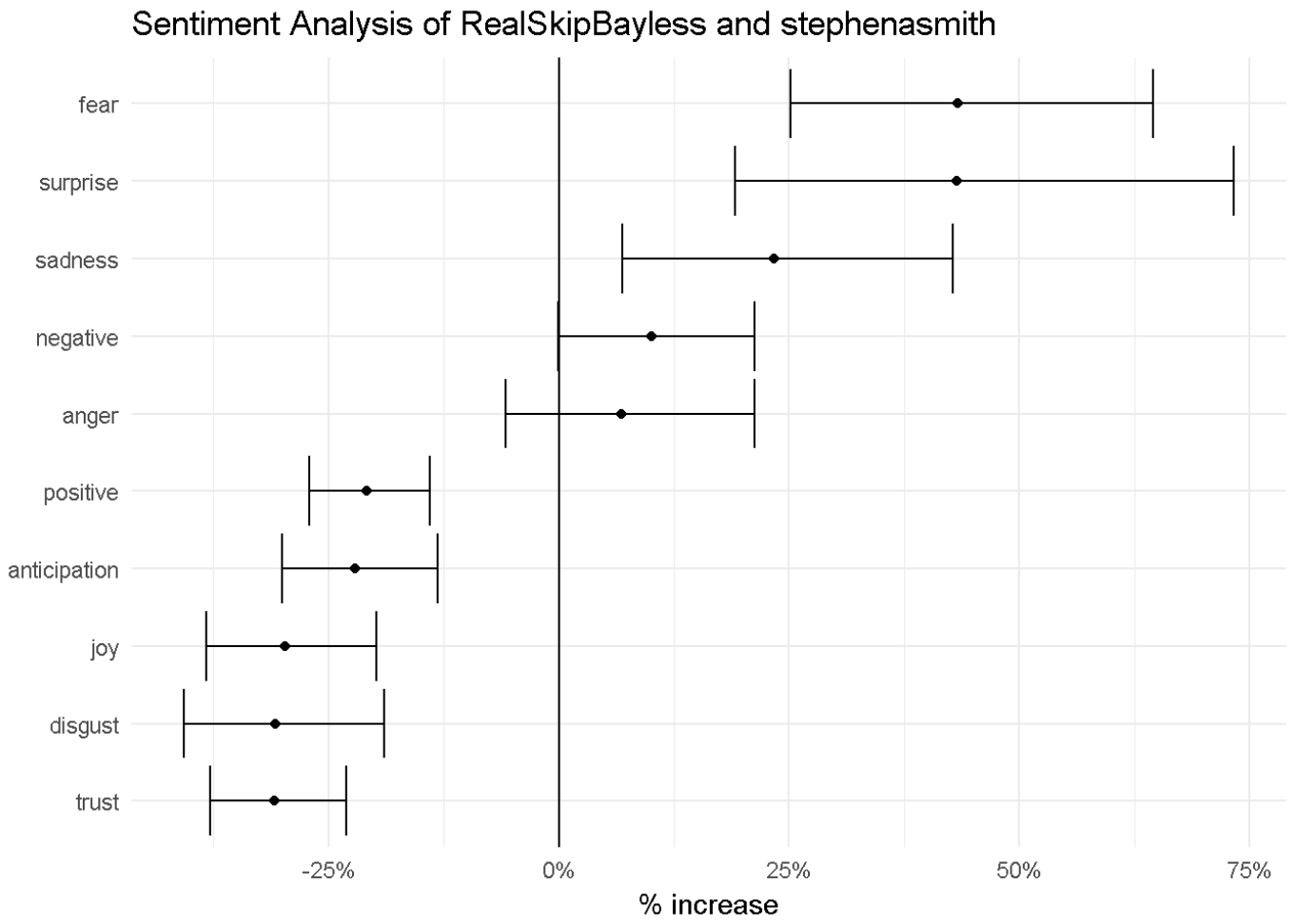
Tweet Popularity

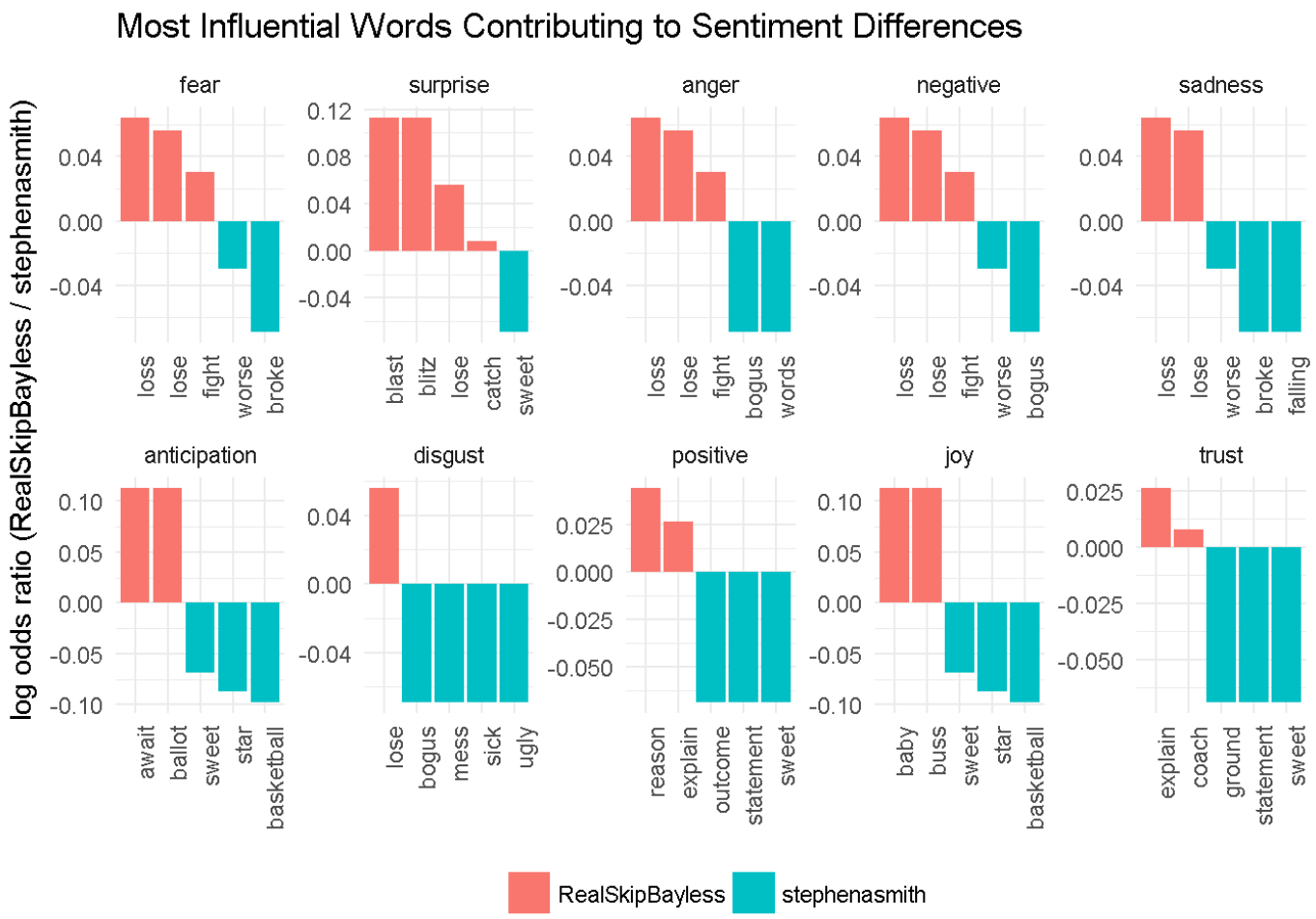
How often do the original tweets get liked/favorited/retweeted?



Sentiment Analysis

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


Sentiment Analysis of RealSkipBayless and stephenasmith
fear
surprise
sadness
negative
anger
positive
anticipation
joy
disgust
trust
-25%
0%
25%
50%
75%
% increase


Most Influential Words Contributing to Sentiment Differences
log odds ratio (RealSkipBayless / stephenasmith)
fear
loss lose fight worse broke
surprise
blast blitz lose catch sweet
anger
loss lose fight bogus words
negative
loss lose fight worse bogus
sadness
loss lose worse broke falling
anticipation
await ballot sweet star basketball
disgust
lose bogus mess sick ugly
positive
reason explain outcome statement sweet
joy
baby buss sweet star basketball
trust
explain coach ground statement sweet
RealSkipBayless
stephenasmith

Conclusion

That’s it!