

STAT Final Project

Prof. Jeffrey Holt

April 30, 2018

What is the Internet Mad About?

By Hailey Ross, Matthew Foreman, and Charlie Trochlil

Twitter is a melting pot of diverse opinions, from crazy to relatively sane. For our project, we were interested in analyzing what exactly makes Tweeters upset or angry. To start, we harvested 10,010 tweets from random times of the day over the course of roughly a week and a half. Our original project proposal thought to mine tweets with multiple key phrases, but we settled on single criterion: each of our tweets contain the phrase “I hate.” There were no restrictions placed on location or any other factors. The following analysis seeks to draw conclusions from the data regarding how anger is communicated on Twitter, and online in general.

Methodology and Difficulties

One of the immediate challenges of our project was that the Twitter API cuts off text of tweets at 140 characters, which truncates many tweets that use the now-longer format. To work around this challenge, in the parts of our analysis that require reading the text of the tweet itself, we are sampling only the un-truncated tweets, a sample of 7,336 out of the 10,010 total. Figures using this smaller sample will be noted.

Another difficulty that somewhat obfuscated our analysis was twitter’s quote function, because the tweets we harvested were often not the actual location of the word ‘hate’ but rather a quote, or even a quote of a quote of the tweet that mentioned it. This is addressed in analysis on page 4.

Things Hate-Adjacent

The tables below explore, respectively, the overall regularity of selected words within the body of a tweet, and the word(s) that directly follow the word hate.

Words/Phrases	95% Confidence Interval for appearance
Love, loving	[0.0519, 0.0626]
Lol, lmao, haha	[0.0292, 0.0374]
Profanity (fuck, shit, ass, hell)	[0.1354, 0.1514]
She, her, girl, women	[0.0440, 0.0539]
He, him, boy, man	[0.2531, 0.2733]
Kanye, Yeezus, Yeezy, Ye	[0.0219, 0.0291]
Trump, 45, president	[0.0033, 0.0065]

It's worth noting that Kanye West's manic behavior and Trump support occurred across the time frame during which we were sampling, and this data would indicate that roughly 2% of the hate tweets at those times involved him in some way. It is interesting, then, that Trump himself is mentioned so rarely.

The disparity between use of male and female concepts and pronouns is one that we do not have an adequate explanation for, besides the grounds-less assertion that men make people angry more often.

Some of our tweets were flagged as sensitive under the "possibly_sensitive" dictionary. We were originally unclear on why tweets received this distinction. In total, only 48 of our 10,010 tweets were flagged as potentially sensitive.

Of the 48 tweets flagged possibly sensitive, only 4 met our criteria for containing profanity. Upon closer inspection of the individually flagged tweets, we determined that many of the tweets on the list were retweets that had not received the tag from other users. This would

suggest that twitter’s algorithm determines something ‘potentially sensitive’ based on the account, not the content of a singular tweet itself.

When parsing our data, we noticed that the users tended to explain the target of their hatred in their first few words following “hate”. We decided to limit the amount of words to two for our purpose. The table below depicts the top ten two word phrases, first words, and second words that followed “hate” within the 10,010 tweets we collected.

Phrase	# of Occurrences (out of 10,010)	First Word	# of Occurrences (out of 10,010)	Second Word	# of Occurrences (out of 10,010)
chewing it	356	when	685	it	373
when people	189	chewing	356	people	273
it when	93	that	294	I	249
when I	73	this	288	this	118
it	68	you	283	but	114
seeing this	65	it	274	so	111
you	56	myself	177	when	106
speech this	54	being	168	me	86
myself for	53	how	154	you	83
when you	51	me	123	my	80

*blank entries, “to”, “the”, “and”, and “for” have been omitted

Looking at the phrases, “chewing it” occurs the most in 356 tweets. See page 4 for explanation. “When people” occurs the second most, indicating that users are targeting hate towards the situations that others cause. The phrases, “It when,” “when I,” “myself for,” and “when you,” likely refer to situations that the user is affected by, which occurred a total of 459

times. It should be noted that “you” occurred 56 times, and this can indicate a specific user being targeted.

Within the “First Word” column, “when” occurs the most with 685 occurrences. It occurs in the “Second Word” column 106 times. This indicates that users are directing hate towards a specific instance, which could take the form of a true grievance or a joke tweet. “This” and “it” occurred within in the four and six spots in “First Word” and the first and fourth spots in “Second Word”, which may indicate nonhuman hatreds from users. “You” occurred 283 times in “First Word” and 83 times in “Second Word” signifying the targeting of other people. The most entries significant of the “First Word” column are “myself” and “me.” Without tweet context, this may indicate self-hatred that users have.

Analyzing Quotation

In addition to looking at what words were commonly used within the hate tweets, we also were interested in looking at what types of tweets--with regard to format--that we were dealing with. Not all of the tweets we pulled were originally written. In fact, we found that many of the tweets in our data set were either Retweets or tweets that had been Retweeted with quotes. We found a confidence interval of [0.6119, 0.6375] of our data set is retweets.

Of 1726 quoted tweets, 53% contained the word ‘hate’ within the tweet, while the other 47% were merely quotes (or in some cases, quotes of quotes) where the word hate was used. This was a significant lesson and a bit of a stumbling block to our analysis, as many of the tweets within the data set were not in fact representative of that user’s feelings.

The most retweeted status (pictured below) was retweeted a total of 356 times, meaning it alone made up about 3.5% of our data. It appears in sample batches from multiple days, evidencing the longevity of the humor’s engagement.

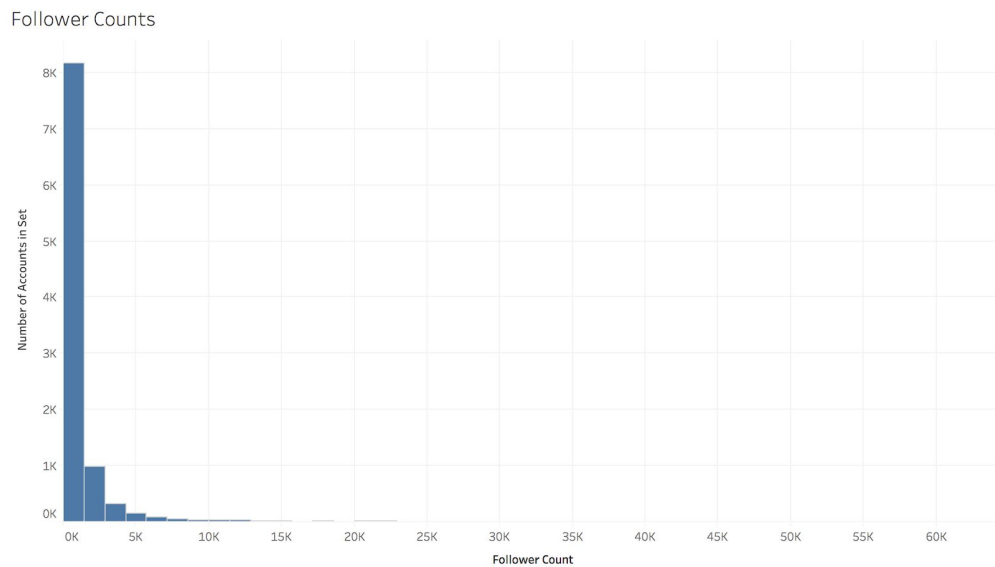


In the quoted tweet’s video, the instagram model pictured drinks from the water bottle and then looks at the camera, moving his lips as though he is chewing.

Analyzing Haters' Data

Even after dropping the outliers (24 of them, to be exact) from the follower counts of all the accounts in our data set, the histogram of the follower counts is heavily skewed to the right.

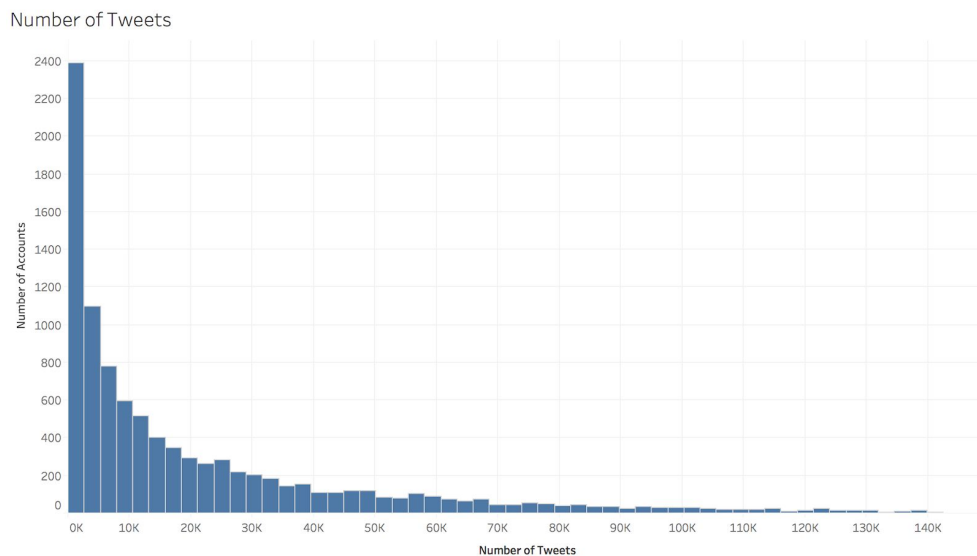
This is likely emblematic of the distribution of follower count across all twitter accounts: the overwhelming majority sit below 1000, but the number can reach the millions



The trend of count of F2 for F2 (bin).

Were we able to mine tweets from deeper in the timeline, it might have been interesting to determine whether popular accounts generally have more negative tweets.

Below is a histogram of the number of tweets from each account where a hate tweet originated from. As you can see, the figure is less skewed, but still indicates most twitter users in this set are less active than might be expected.



The trend of count of F2 for F2 (bin).

Most Productive Haters

In our quest to analyze hateful tweets, we found that the angriest user (in other words, the user that appeared most in our data set--9 times to be exact) was actually a bot trying to promote music. Another bot made the top ten contributors of hate, the aptly named @I_Hate_Bot, which “tweets senseless messages every 10 minutes.”

The most prolific human tweeter to appear on our list retweeted 6 tweets within 3 minutes, all of which were other users with very few followers tweeting that they hated Ed Sheeran.

Conclusions, Final Thoughts

Although our intention had been to examine the petty and sincere frustrations of Twitter users, the trends in the data pointed repeatedly to the fact that ‘hate’ is a weak word in the internet’s vocabulary, frequently juxtaposed with softer concepts or used for the sake of exaggeration in humor. If we were to attempt this type of analysis again, it would likely be more valuable to tighten our search criteria, although it’s unclear what language might be sought out in order to find only legitimate rage. It would also be worthwhile to collect a larger data set, so that tweets with certain peculiarities, such as the sensitive material tag, could be examined. Finally, removing retweets and quotes from the pool of tweets would allow the regularity tables to have more meaning, since users would be independently arriving at similar situations and syntax, rather than directly parroting one another.

NOTES:

of tweets: 10,010

of short tweets: 7336

of users with place listed: 6942

'love' or 'loving' comes up 606 times in the entire tweet list

(lcl and ucl): Love & Hate w/ short list: [0.05193548911010028, 0.06256832768379285]

(lcl and ucl): Love & Hate w/ full list: [0.055867517088542026, 0.06521140399037906]

(lcl and ucl): Retweets/Quotes w/ short list: [0.4892397266373085, 0.5121234140388093]

(lcl and ucl): Retweets/Quotes w/ full list: [0.6244286166292781, 0.6433036511029897]

(lcl and ucl): Sensitive Material w/ short list: [0.004698094287970599,

0.008388056202760045]

(lcl and ucl): Sensitive Material w/ full list: [0.0034418899603423985, 0.006148519630067192]

10 tweets at Kanye

4 at trump

1316 are replies

#####TOP TEN BELOW#####

chewing it 356

when people 189

it when 93

when I 73

it 68

seeing this 65

you 56

myself for 53

when you 51

when he 49

that being 48

*"speech This" has been omitted f

#####TOP TEN FIRST WORD#####

when	685
chewing	356
that	294
this	288
you	283
it	274
myself	177
being	168
how	154
me	123
dtype: int64	

*to and the have been omitted

#####TOP TEN SECOND WORD#####

it	373
people	273
I	249
this	118
but	114
so	111
when	106
i	86
you	83
my	80

*blank lists, and, and for have been omitted

QUOTES:

Of 1726 quoted tweets:

806 contain 'hate' within the tweet that was quoted

920 contain 'hate' within the text of the response

Most quoted tweet (356) is this: <https://twitter.com/vinnybrack/status/988530519175393281>