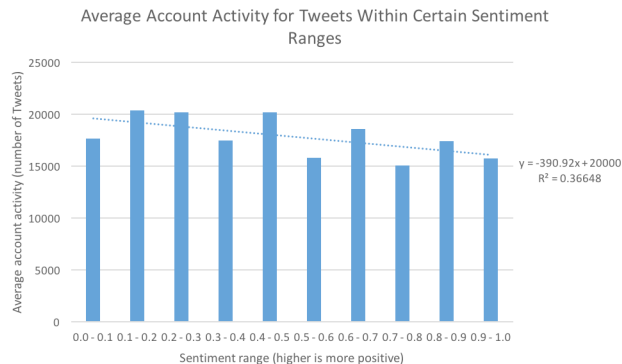


What Does Twitter Activity Reveal About Post-Election Sentiment?

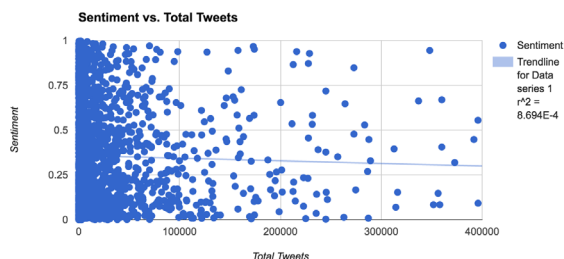
By Cole Carlin and Yunhua Zhao

We collected three main types of data from Twitter using its live streaming API:
Location (Time Zone) • **User Activity Level** (Number of Tweets) • **Account Age** (Days Since Creation)

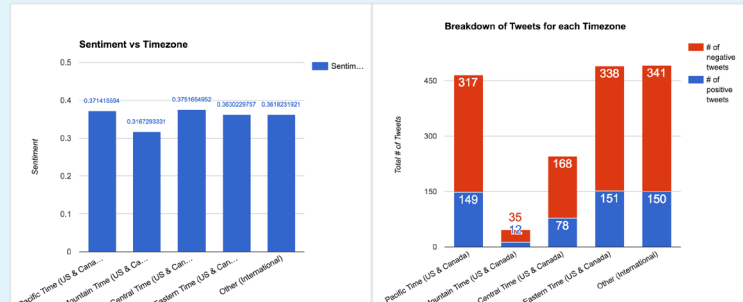
Are pro-Trump users less active on Twitter?



There appears to be a slight negative correlation between account activity and pro-Trump tweets.



How does location affect Trump support on Twitter?



Note: This data was collected separately, as it required location data, which not every tweet has.

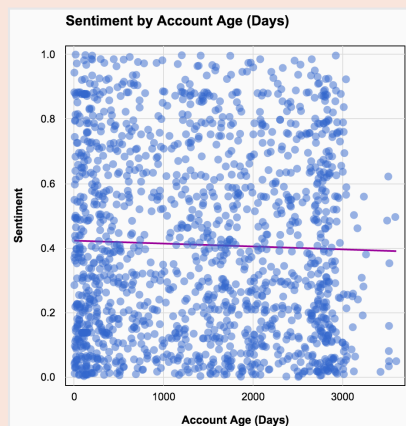
Sentiment toward Trump on Twitter does not appear to follow general geopolitical trends.

...And just how reliable is our data?

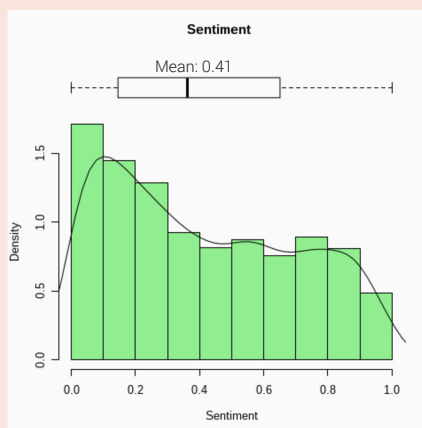
Obviously, Indico's machine learning API (which we used) is not perfect. So we performed several tests to make sure our dataset is statistically significant. To do these, we make several assumptions:

- 1) Twitter and the actual US population have the same variance
- 2) The populations are normally distributed
- 3) Our sample is independent from the population

Is account age correlated with pro-Trump tweets?



There does not appear to be a general linear correlation between account age and sentiment toward Trump. However, a closer look at the data reveals some interesting findings.



This is a fat-tailed distribution, suggesting that Twitter introduces more responses toward either extreme than a ordinary normal distributions, but especially (in this case) negatively

For the dataset with an associated account age:
Mean sentiment: 0.410
Standard deviation: 0.292
Sample size: 1497
95% Confidence Interval: (.395215, .424785)

By contrast, Politico and Huffington Post both have Trump's favorability rating at 47%, slightly higher than our average sentiment, which is reasonable given the internet's well-documented liberal bias.

Control sample:

To make sure Indico's sentiment analysis API was not inherently biased, we performed a two-sample two-tailed t-test against tweets collected without a filter.
Control sample statistics:
Mean sentiment: 0.497
Standard deviation: 0.265
Sample size: 1497

Null hypothesis: $\mu_c - \mu_s = 0$
Alternative hypothesis: $\mu_c - \mu_s \neq 0$

95% Confidence Interval for $\mu_c - \mu_s$:
(0.0666, 0.1066)
Two-tailed p-value: less than 0.0001

Thus, the difference between filtered tweets and unfiltered tweets is very statistically significant.