Does Twitter Political Engagement Imply Real-world Election Participation?

By Ao Wang, Jiayang Wu, Xinyu Song, Yicheng Shi

Introduction

Participation in the presidential election has a fundamental impact on American society. Twitter, as one of the largest social network platforms, has been claimed to play a vital role in presidential elections.

Therefore, our project aims to investigate the relationship between voter participation and Twitter engagement in 2020 election.

Hypothesis

Our general hypothesis is that **the 2020 presidential election participation is associated with Twitter engagement.**

Claim (1) Greater Twitter engagement leads to higher state-wise voter turnout.
Claim (2) The state-wise winner of the presidential election is related to the candidate who appeared the majority of times in the state-wise Tweets.

Data

In this project, we used two datasets.

1. A Tweets dataset from Kaggle. We cleaned and filtered out non-US location records.
2. We collected the voter turnout dataset from the Washington Post. The visualization of this dataset can be found in Fig.1.

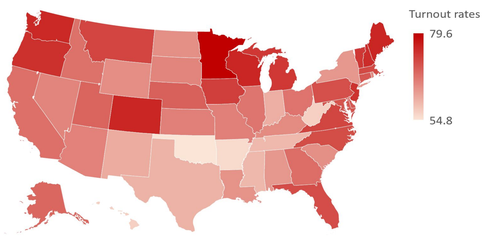

Figure 1. Presidential election turnout by state

Pearson Correlation & Single Linear Regression

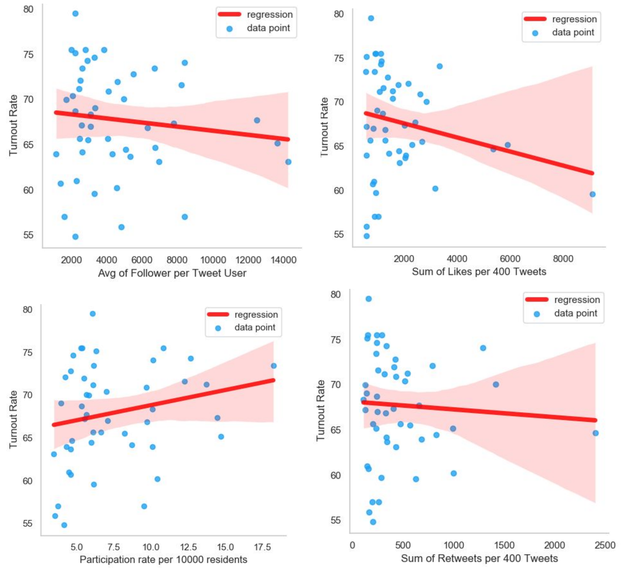

Figure 2. Single variable linear regression plots with 95% confidence band

Table 1. The result of Pearson Correlation Analysis

	Pearson Correlation	P-value
Likes	-0.17	0.25
Retweets	-0.03	0.81
Participation Cnt	0.22	0.13
Avg Followers	-0.11	0.46

To measure the engagement of Twitter users, we defined four metrics.

(1) Avg number of followers per user
(2) Sum of likes per 400 Tweet Samples
(3) Participation rate per 10K residents
(4) Sum of retweets per 400 Tweet Samples

We then conducted Pearson Correlation Analysis(results shown in Table 1) and built four single linear regression models using data in 5% - 95 % range to show relationships between voter turnout and Twitter engagement(results shown in Figure 2).

Pearson Correlation Analysis:

- All p-values > 0.05, our results are not statistically significant.

Single Variable Linear Regression:

- Most of the points are **out of** 95% confidence band, which means unfortunately our data doesn't observably have a linear correlation.

Multiple Linear Regression

We then ran a multiple regression to find the relationship between state-wise voting rate and four metrics defined before that indicates engagement levels. Results can be found in Table 2.

None of the p-values is < 0.05, which shows that **even when we expand to multiple variables, there is no evidence that state-wise voting rate can be explained by Twitter user engagement on political matters.**

Table 2. The result of Multiple Linear Regression

	Coefficients	P-value
Participation Cnt	0.011909109	0.807829
Like	-0.000552305	0.229485
Retweet	0.000595507	0.822347
Avg of Followers	0.000100375	0.638767

Chi-Squared Test

We used the Chi-Squared Test to validate Claim (2). Twitter winner is defined as the candidate whose number of mentions in hashtags outnumbers other candidates state-wise. Table 3 displays the number of states that falls into each category (winner of real-world v.s.Twitter).

Table 3. The result of Pearson Chi-Squared test

	Trump (Twitter winner)	Biden (Twitter winner)
Trump (Real)	17	8
Biden (Real)	21	5

Chi-squared Test results:
Chi-squared = 1.09417, df = 1, p-value = 0.295548

Conclusion

In conclusion, **we didn't find a significant correlation between Twitter engagement and voter participation, regardless of how the engagement is measured.**

However, other factors may affect our results:
1. Potential bias when generating Tweet data
2. Not having data on voting rates on the county level

Future Work

1. Come up with more well defined methods to measure users' Twitter engagement
2. Experiment on past elections to see if the results are the same

♥ Acknowledgement ♥

Thank you to Professor De Stefani, and our TA mentor **Tiffany Nguyen**, for their kind support and guidance throughout the course of this project!