# The Effect of Twitter Exposure on Presidential Elections

John Lucas Garofalo, Daniel Hernandez, Matt Lollis, Dr. Alexander Herzog

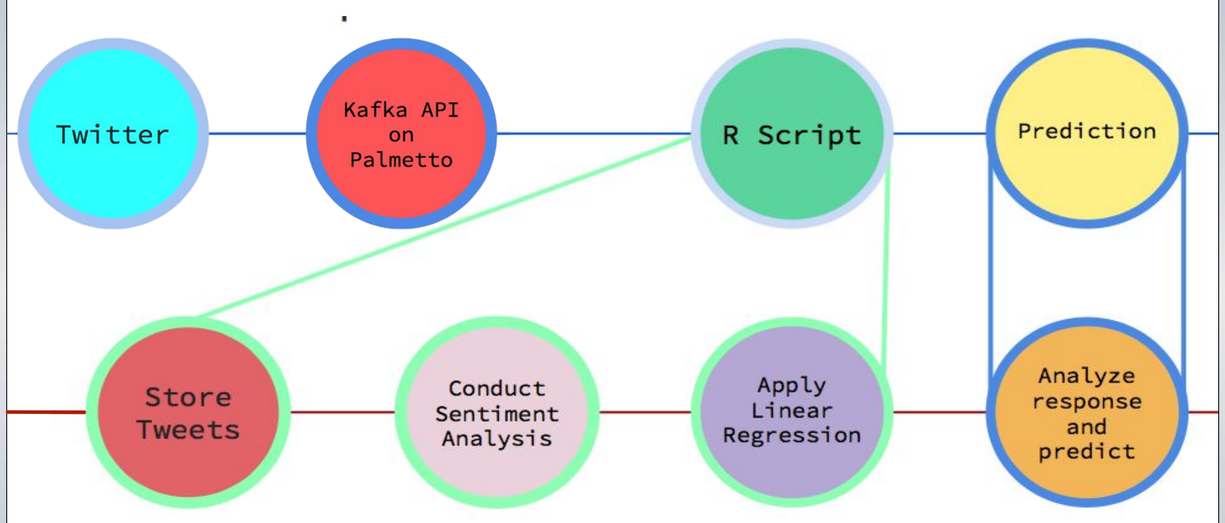
*Clemson University*

## Introduction

*"All publicity is good publicity."*

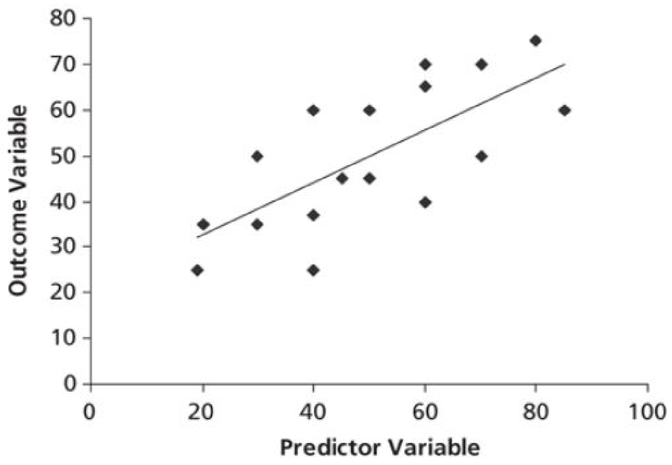
Our goal was to train a statistical model that analyzes the sentiment of tweets that mention presidential candidates and determine how Twitter exposure, both positive and overall, affects the results of preliminary opinion polls and state primary elections. We performed this analysis on presidential candidates Donald Trump, Ted Cruz, Hillary Clinton, and Bernie Sanders.

## Methods & Model



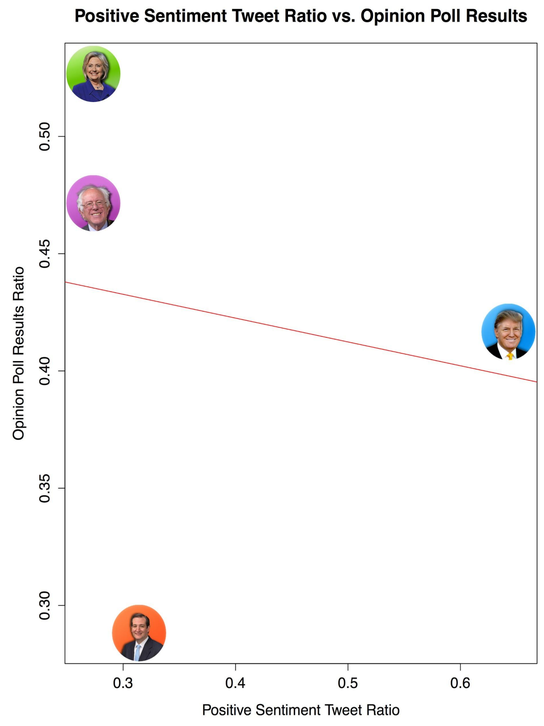

We chose "Simple Linear Regression" as the method to analyze our results. If you recall, linear regression is a simple supervised learning method that assumes a linear relationship between the predictor Y and the features, $X_1$,..., $X_f$.

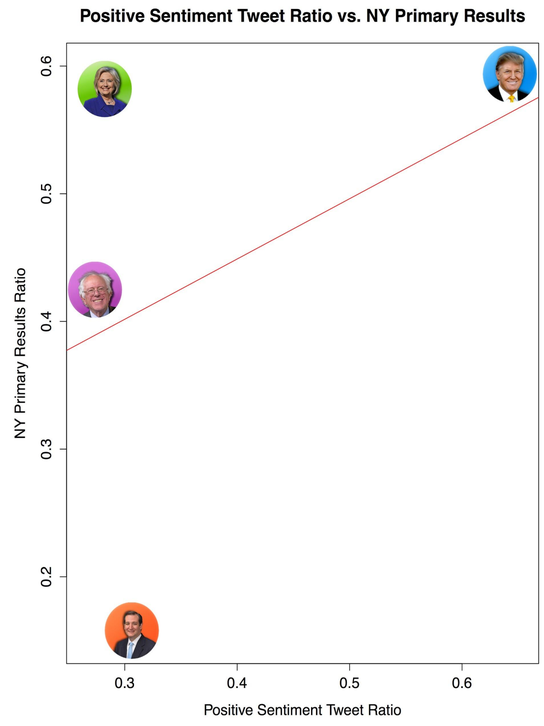
We used linear regression to produce a quantitative response that represents a relative score for each candidate. We use this quantitative response to predict popularity (by percentage) in both preliminary popular opinion polls and state primary elections.

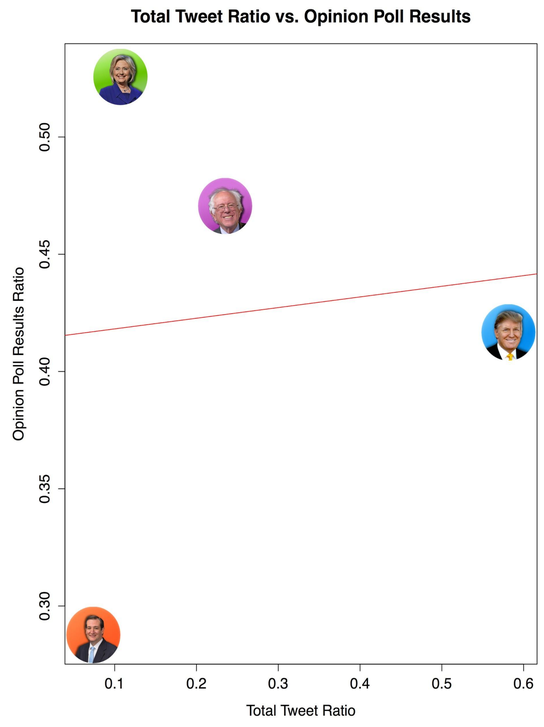


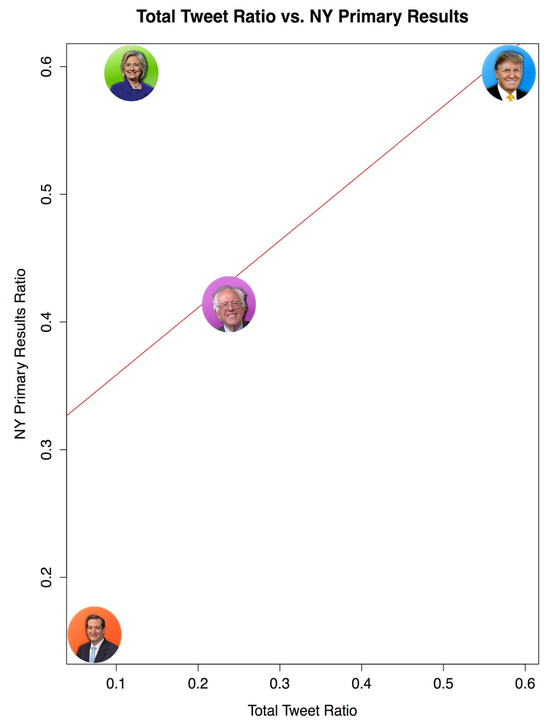

We collected approximately 1.2 million tweets for our dataset, filtered by each candidate's last name, over the span of four days leading up to the New York Primary. We plan to run it for a longer period of time leading up to the next primary to see how our model holds up.

## Results




## Conclusions

While Twitter exposure isn't reliable enough to be an autonomous election predictor, it is viable as a factor in a multiple linear regression method prediction.

For example, democratic candidate Bernie Sanders and republican candidate Donald J. Trump had the highest total tweet ratios and fell right along our regression line. Republican candidate Ted Cruz had very little Twitter exposure and also did very poorly in the preliminary opinion polls as well as the New York primary.

In contrast, democratic candidate Hillary Clinton did very well in the preliminary opinion polls and the New York state primary despite lack of Twitter exposure, positive or otherwise.

We decided that positive and total Twitter exposure would be a viable weighted predictor in a multiple linear regression model but isn't perfect as a simple linear regression prediction method to predict poll results.