# The Geography of Super Tuesday In Ohio through Twitter

## Abstract

Twitter over the past 2 years has grown 1382% totaling over 175 million users. Over 140 million tweets are sent per day. An interaction of this magnitude has created a new geography, the geography of social media.

## Data Retrieval

1. Obtain Ohio counties from US Census Bureau
2. Determine the centroid of each county in order to generate a 3 km buffer at those centroids
3. Generate a list of X,Y coordinates for each centroid. Twitter search in "R"
4. Write a script that can use these X,Y coordinates to search Twitter for relevant content.

## Twitter Applications

The relationships found in this research between primary voting, Twitter, and the demographics of voters, highlight the important role of communication through social media. This analysis opens a new market that is highly important in today's society. A 140 character Tweet has proven to be a valuable expression of voter preference which can be related to election outcomes. This research demonstrates a new Geography of elections through Twitter and social media. This study is important to political Geographers and politicians alike, opening a new door into the minds of voters. Because of the clear prevalence of "Geotagged" social media the formalized definition of place can be re-evaluated and re-defined to incorporate cyberspace into the normal idea of place.

## Regression Results

Using Pearson's Correlation and scatterplot analysis, a positive correlation was found between the percent of votes for Romney and the percent of Tweets for Romney. This relationship is demonstrated with the positive slope of the trend line (Figure 2). Although the R value is not considered high at 0.429, a 2000 significance value indicates there is a linear relationship between the number of tweets for Romney and the percent of primary votes cast for him across Ohio counties.

A negative correlation was found between the percent of votes for Santorum and the percent of Tweets for Santorum. This relationship is demonstrated with a negative slope of the trend line (Figure 3). While the R value of 0.224 was weaker than Romney's correlation, there is still a significant linear relationship with a 0.036 probability value.

<table>
<thead>
<tr>
<th>County</th>
<th>Romney Votes</th>
<th>Santorum Votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamilton</td>
<td>20,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Butler</td>
<td>15,000</td>
<td>25,000</td>
</tr>
<tr>
<td>Mercer</td>
<td>10,000</td>
<td>15,000</td>
</tr>
</tbody>
</table>

## Tables

Table I illustrates the Ordinary Least Squares Regression analysis for the number of Tweets per candidate and their percentage of the vote cast per county. In both instances, weak R values were influenced by the large number of counties with no Tweets. To account for this limitation, Spatial Regression or Spatial Lag was used; results shown in Table II. The Spatial Regressions resulted in dramatically improved model performance. This improvement is indicated by the large increases in the R-squared (coefficient of determination), and the Akaike and Schwarz information criteria for each model. Romney outperformed Santorum in counties with Tweets mentioning Romney more often; while, Santorum performed worse in counties that mentioned Santorum more often. This discrepancy may be accounted for by the content of the Tweets, i.e., negative Tweets, and or by the demographics of Santorum's constituents and possible lack of Twitter activity.

The results of Stepwise Regression characterize the demographics of “Romney and Santorum Voters” and are shown in Figures 2 and 3, as well as Tables I-II. Romney outperformed Santorum in counties with increased proportions of persons with higher education, minority populations, and older populations. The strongest variable defining Romney constituents was percent bachelor's degree as indicated by the larger Beta Value. In contrast, Santorum performed better in counties with smaller proportions of bachelor's degree, minority populations, and younger populations. As with Romney, the proportion of bachelor's degree has the strongest Beta Value. All in all, Romney's greatest support was in urban areas, while Santorum had stronger support in rural areas of Ohio. This pattern is corroborated by the clustering evident in the LISA Maps (Figures 4 and 5).

## LISA Maps

LISA Maps illustrate the spatial clustering of candidate constituents in Ohio counties. Voter patterns in Ohio demonstrate Table’s Law, which states that nearer places are more similar than places that are further away. Romney constituents are concentrated in the Northeast and Southeast corners of the state, both larger urban counties (Figure 4). Santorum’s constituents are also clustered, but in opposite regions as Romney’s (Figure 5). These maps highlight the spatial polarity between candidate constituents. Moran’s I, a statistic which measures the degree of spatial autocorrelation, reflects this clustering.