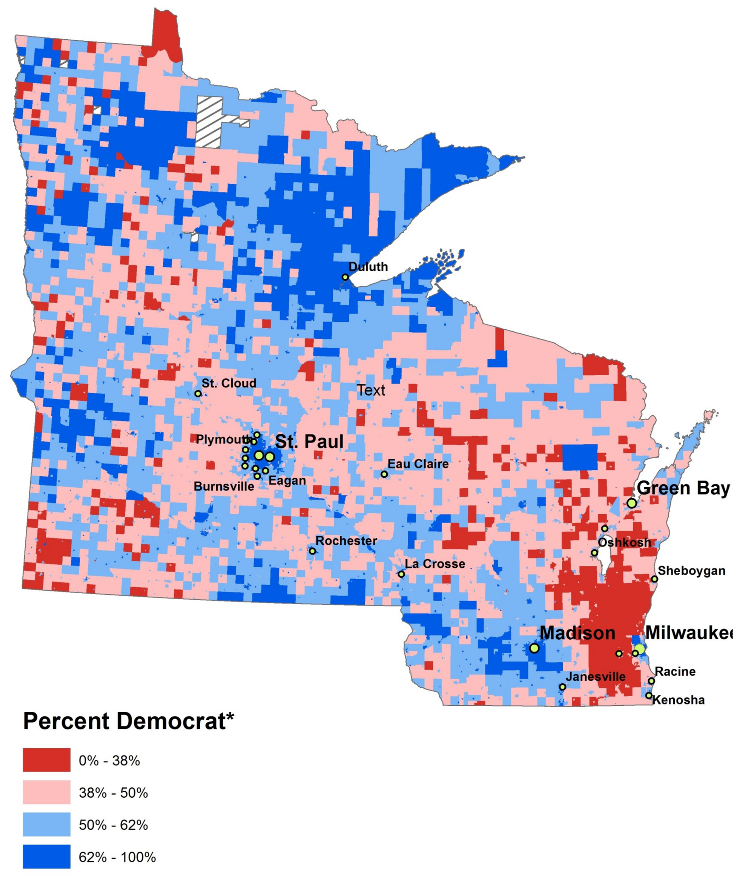
The Link Between Geographic Polarization and the Tone of Social Media Interaction

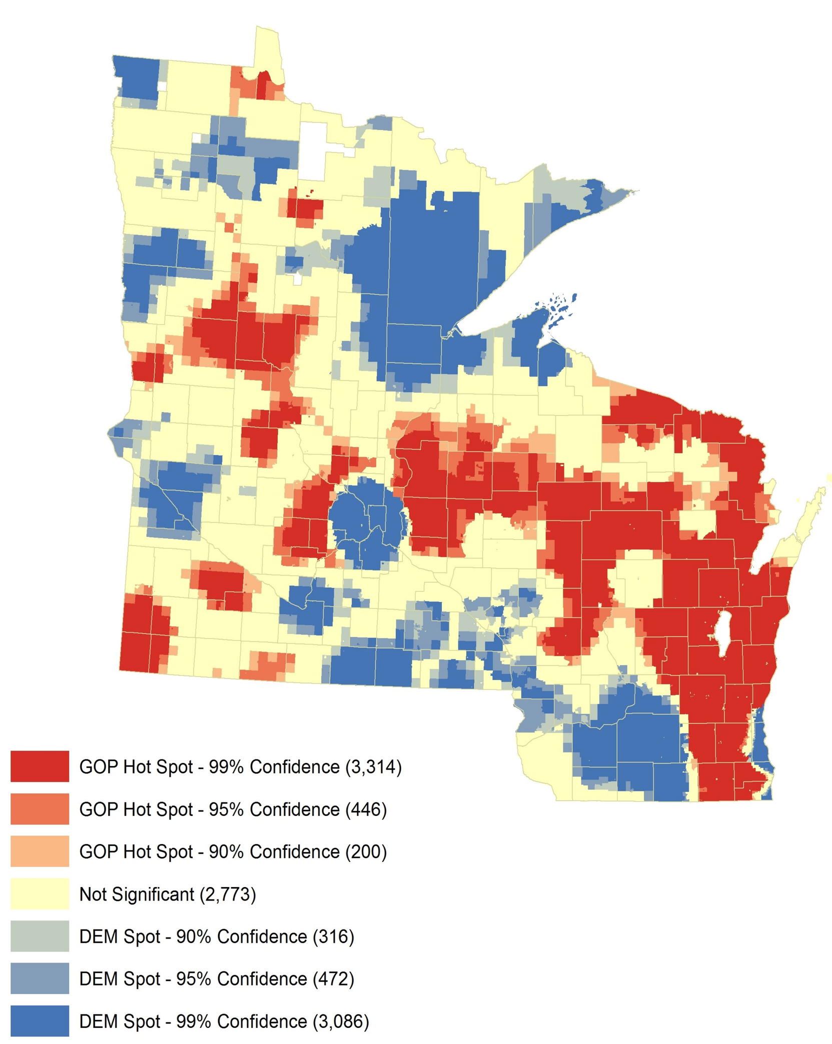
Cartographer: Jon Atkins

GIS 101 - May 12th, 2016

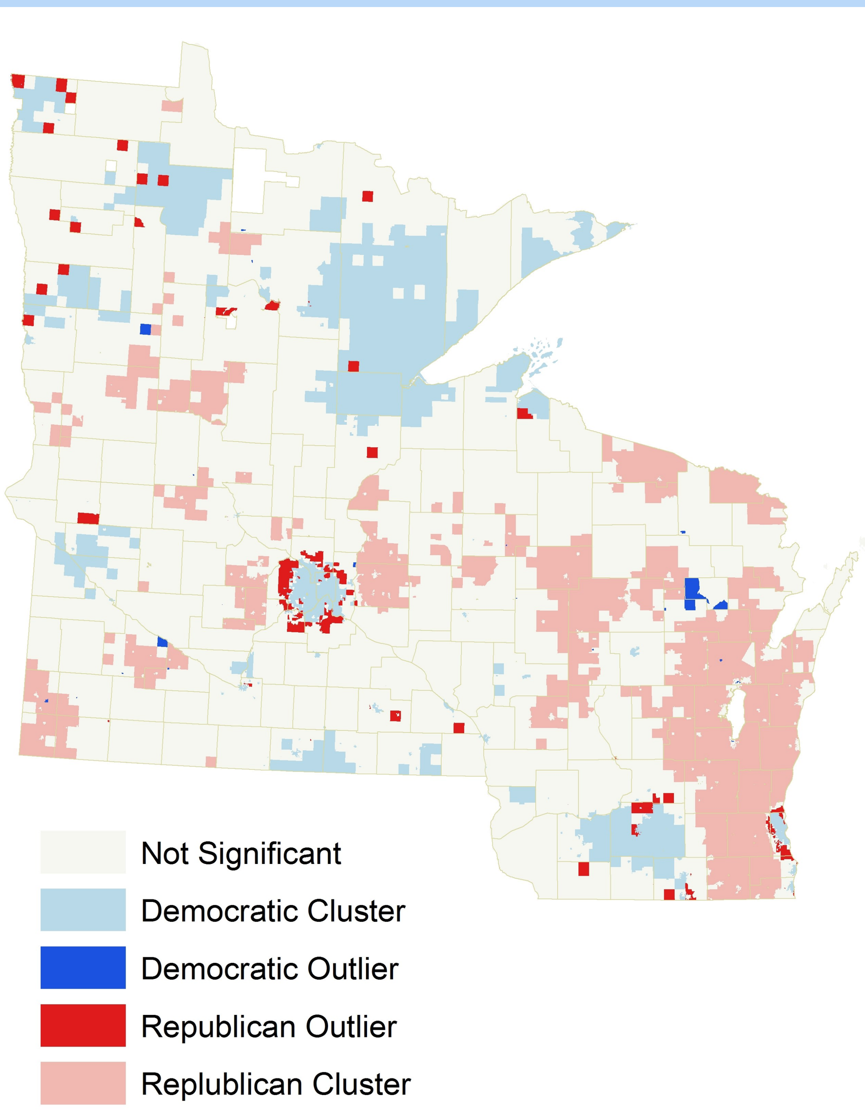
2012 Election Results by Precinct



2012 Hotspot Analysis by Precinct



Precinct Clusters Using Local Moran's I



Background

Twitter, like many other social media platforms, has become an increasingly important medium in which political discourse occurs. This is never more present than in the election cycle. In the United States election cycles have become points of social tension with social media as the arena in which many political debates rage, commentary arises and is shared.

Many studies have shown that politics and government in America have become increasingly polarized over the past few decades, but can the same be said for political discourse among citizens? States are large geographic entities, often with diverse political interests. Can these political differences be represented and this regional diversity categorized? With the rise of social media and natural language processing, we have a way to record, measure and quantify tone of conversation between people across the world. While the tone of social media interactions about politics can often take on a confrontational tone, can this tone be quantified and can it be related to this political and geographic diversity? Does the tone of political discourse change between those who come into regular contact with those who disagree with them and those that do not?

This Study

In the last full national election in 2012, the neighboring Wisconsin and Minnesota both voted for victorious Democratic president Barack Obama and Democratic Senators (Tammy Baldwin and Amy Klobuchar) by significant majorities. However, both states hold significant conservative minorities and a closer look at the results reveals two very divided Midwestern states. There are huge clusters of highly Democratic regions and highly Republican regions. Many of the people in these regions can feel like they are living in different states, rarely coming into contact with the opposition in person. However, that geographic divide is in theory removed in social media. Thus the research question for this study concerns whether the tone of political conversation on social media changes for those that live on the borders of these partisan strongholds or in outlier areas.

Methods (Precincts)

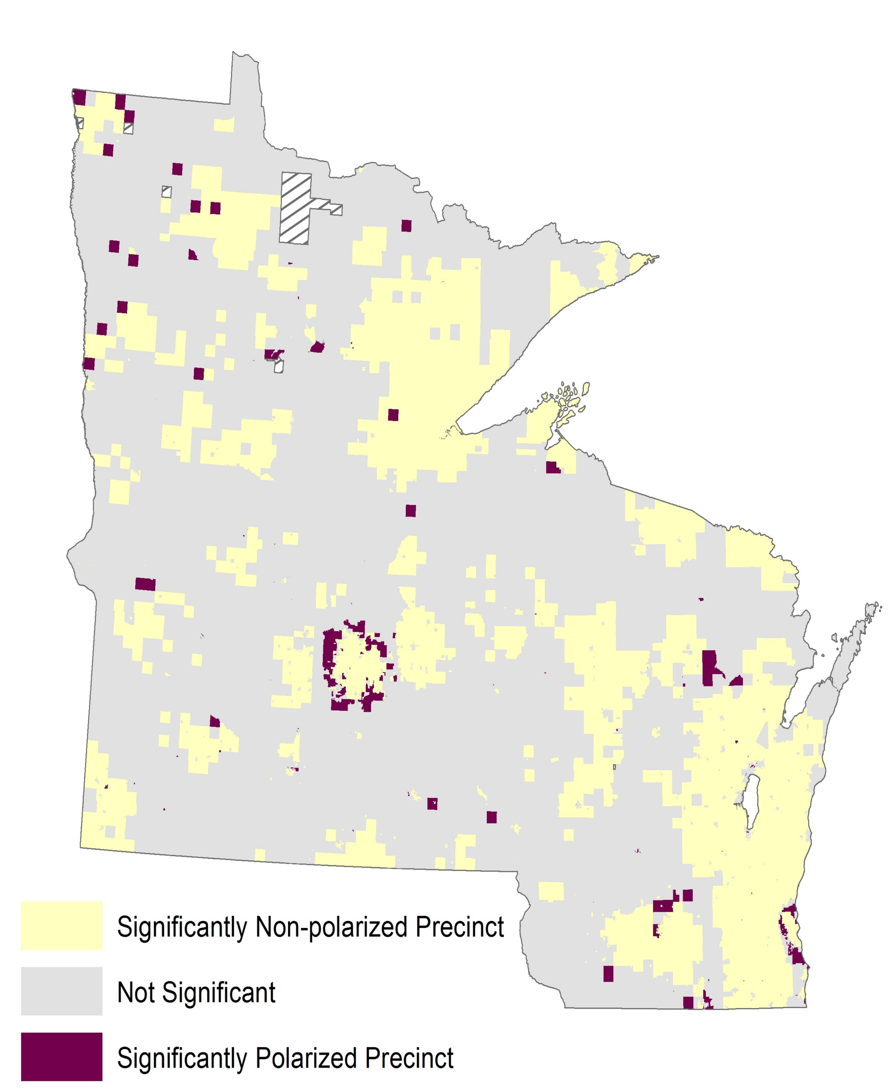
Analysis for this study was based on precinct delineations and precinct-level data from the 2012 general election. Data was procured and analyzed according to a 'Percent Democratic' variable:

Percent Democratic is an attempted generalized, binary representation of a precinct's partisan preferences. It is an attempt to both remove the impact of non-major party candidates as well as the complicating factors of specific candidates. The formula is as follows:

((DEM_PRES_VOTES/TOT_PRES_VOTES) + 1 - (GOP_PRES_VOTES/TOT_PRES_VOTES)
+ (DEM_SEN_VOTES/TOT_SEN_VOTES) + 1 - (GOP_SEN_VOTES/TOT_SEN_VOTES)) / 4

From there, two different spatial analyses were performed. **Hotspot Analysis** (left) was used to find the most strongly Democratic and Republican areas, as a preliminary measure and visual representation of divides. Next, **Local Moran's I Cluster Analysis** (below and left) was performed to find outlier precincts which were mapped and combined without regard for the partisan lean of the polarization.

Polarized Locations



Analysis

Wisconsin is a notoriously politically divided state. Simultaneously capable of electing Tammy Baldwin and voting for Bernie Sanders as re-electing Scott Walker and once supporting Joseph McCarthy. A precinct-level analysis of the state shows this divide starkest between the extremely Democratic Milwaukee proper and the fiercely Conservative suburbs. This divide then extends again as you travel further west to the Democratic stronghold of Madison, near the University.

Minnesota shows a similar, if less intense trend. The Democratic stronghold of the Twin Cities quickly gives way to the most Republican of suburbs that surround the urban area in a ring. Furthermore, like a smaller Madison, remote Duluth merges the liberal tendencies of a college town and mid-size city to create another political pole in the state.

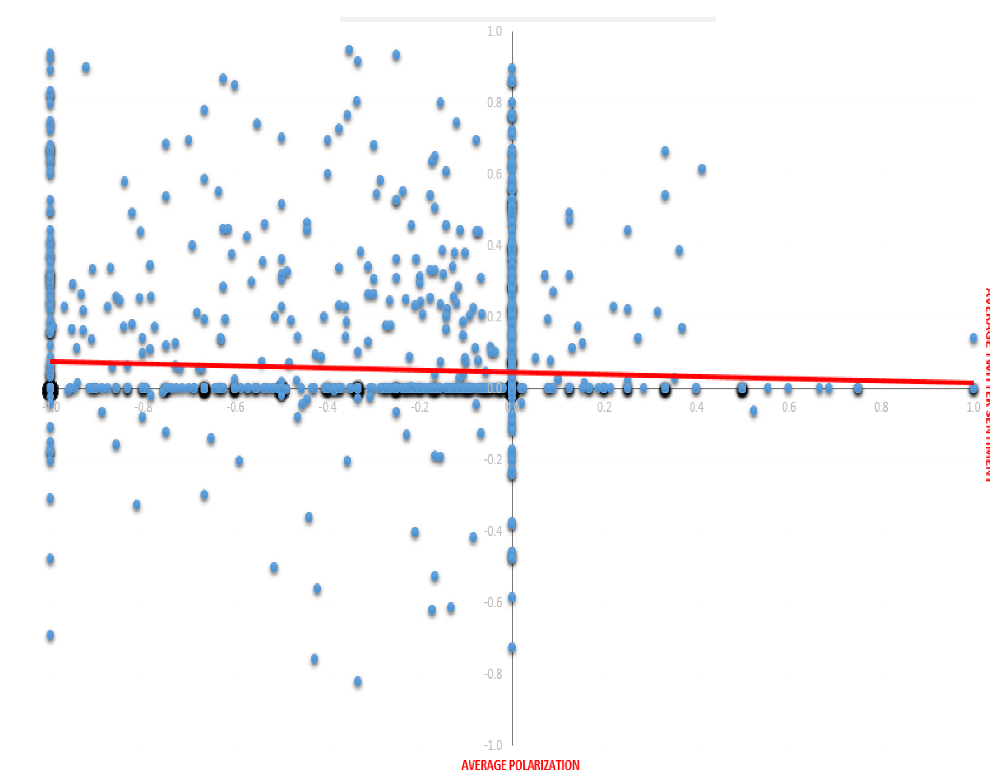
Precincts Analysis: As can be seen (left), the most intensely polarized locations exist mainly in suburbs of these major cities and remote locations away from population centers.

Twitter: Because most of the tweets were not geocoded beyond the Town level (mostly based on user-listed locations which rarely if ever specify the precinct), both data was dissolved into county and official town (where applicable) layers and then batched and analyzed accordingly.

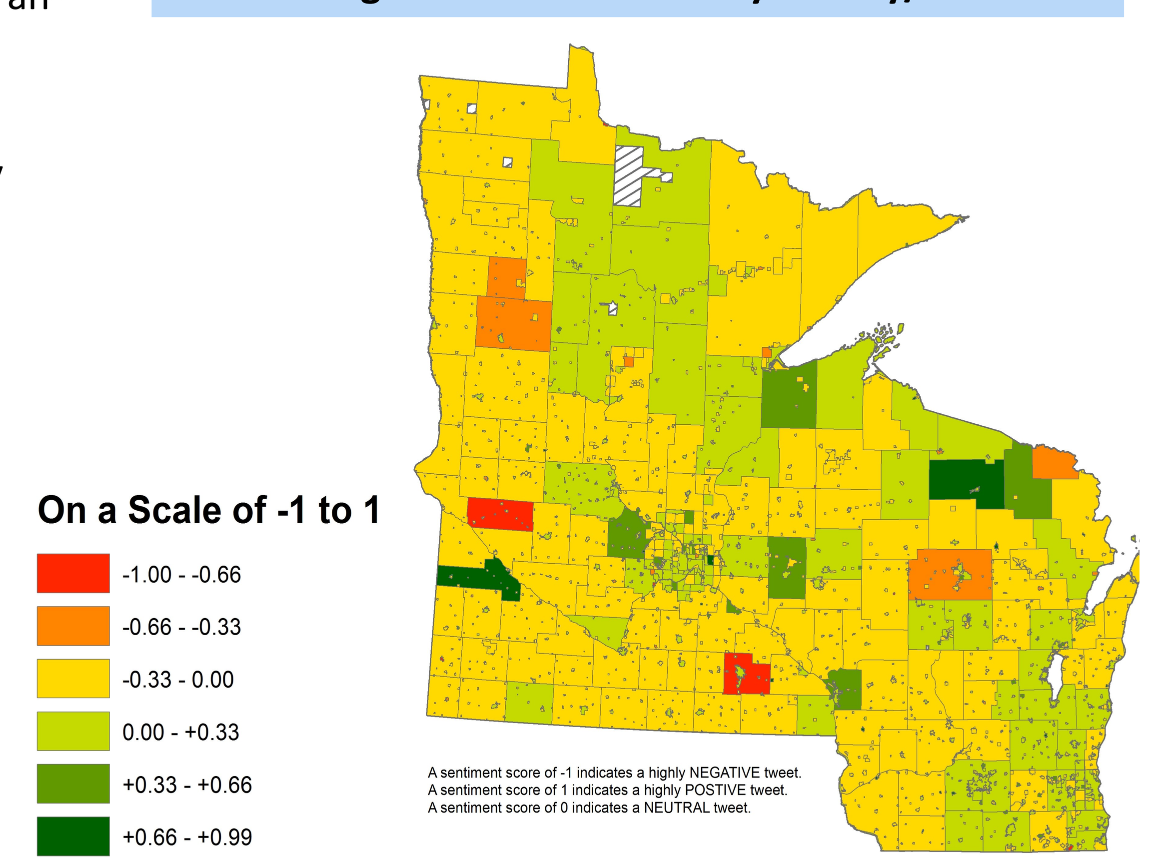
The overall results showed a slight positive tone (.060 in the tweets, but drastic differences based on the search keyword (see far right). As for geographic results, the tweets were fairly spread across the country but unsurprisingly tended to cluster around the areas of high population.

For the geographic correlation, results, as can be seen in the graph (above right), did not show a strong correlation (-.068) between the variables of geographic political polarization and positivity of tweets. Thus it appears, at least for these states and this week, there is little verifiable link between geographic political polarization and the tone of political discourse on social media.

Tweet Sentiment vs. Place Polarization



Average Tweet Sentiment by County/Town



Data sources

Twitter.com, Wisconsin Secretary of State, Minnesota Secretary of State, Aaron Strauss Github

Projection/Coordinate System

NAD 1983 StatePlane Minnesota; GCS WGS 1984

References

(Vader Sentiment Analysis) Hutto, C.J. & Gilbert, E.E. (2014). Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eight International Conference on Weblogs and Social Media(ICWSM-14). Ann Arbor, MI, June 2014

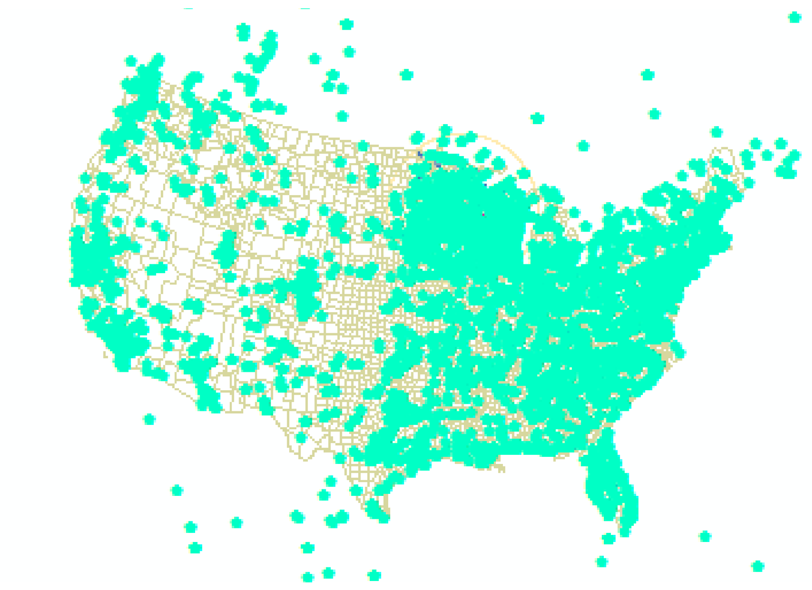
Libraries/APIs Used

Pandas, TwitterSearch, VaderSentiment, Twitter Search API, Google Geocoder API, Python Geocoder, Openpyxl,

Acknowledgments

Dr. Sumeeta Srinivasan for being a great teacher and for her incredible help through this endeavor. Dani Sandoval for their advice and references in Twitter-based geo-data analysis. Harsha Anaravadi for her emotional support and help with the poster design.

Tweet Sample Collected*



*Not shown, more tweets collected from other parts of the world

Methods (Twitter):

To gather tweets a number of steps were taken. First a dictionary of election-related terms was created. These 57 keywords ranged from generic "election" or "politics" to key issues like 'immigration' and "social security." It also included each of the presidential candidates' names, many variations on the party names and even common hashtags like 'feelthebern' & 'nevertrump.'

From there, a smallest possible circle was drawn around the two states (as seen above), and the origin and radius were used to establish a query using Twitter's search_by_location feature.

As the tweets were collected for the week from May 3rd-10th 2016, vaderSentiment Analysis was used to find the tone (on a scale from –1 - 1) of the tweet's text and, where possible, the coordinate or place information were recorded. For the vast majority of tweets that included no specific location information, the 'location' given in the user profile was used and geocoded. In all, around 50,000 tweets were able to be geocoded (above), of which 26,538 were within the sample area. These tweets were mapped and coded (below).

Search Keyword	Average Sentiment
vicepresident	0.742
reps	0.214
social security	0.212
campaign	0.181
vote	0.180
democrats	0.161
rep	0.160
student loans	0.143
primary	0.135
taxes	0.112
debate	0.106
election2016	0.102
convention	0.101
Bernie	0.100
republican	0.097
republican	0.094
president	0.093
Sanders	0.089
congress	0.084
feeltheBern	0.082
senator	0.080
medicare	0.069
dem	0.068
election	0.064
democrat	0.063
(blank)	0.060
inequality	0.056
Bush	0.054
nevertrump	0.046
tedcruz	0.040
conservative	0.036
gop	0.031
republicans	0.030
dems	0.029
Clinton	0.021
Hillary	0.017
Donald	0.017
government	0.016
primaryday	0.015
liberals	0.014
liberal	0.009
Trump	0.001
socialist	-0.001
capitalism	-0.002
Cruz	-0.010
politics	-0.018
Kasich	-0.043
medicaid	-0.067
immigration	-0.071
nazi	-0.085
immigrants	-0.098
welfare	-0.098
senate	-0.150
gun control	-0.264
illegalimmigration	-0.273
fascist	-0.342
terrorism	-0.648
Total	**0.0601454**

Tweet Locations by Sentiment