

Utilizing Social Media to Predict the 2016 Elections

Nhi Chung, Hans Marasigan, Richard Ko

UC Santa Barbara, Department of Geography

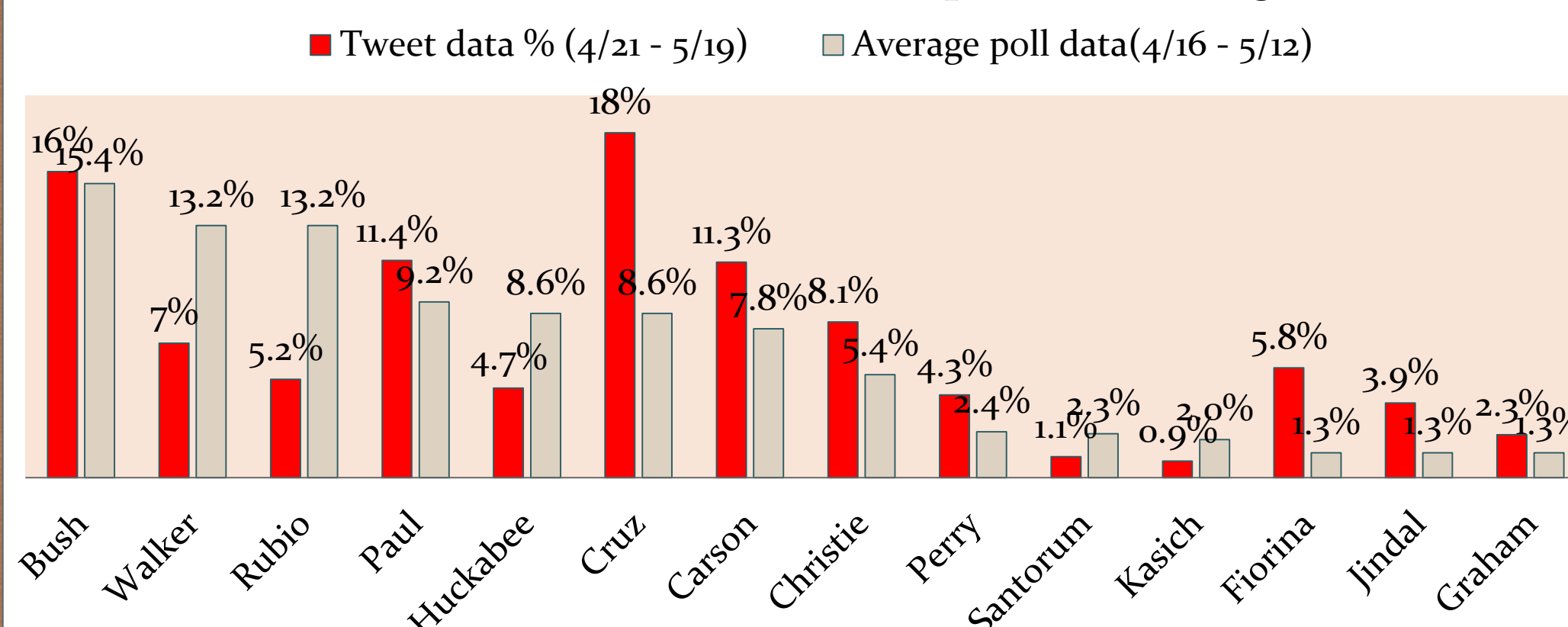
Introduction:

- With the growing availability of big data, studying trends in social media and making predictions from those trends has become increasingly popular among data scientists
- Twitter , Facebook , and Instagram  are some of the most popular social media sites in the United States with more than hundreds of millions of people using them regularly
- Goal: utilize Twitter data to predict the 2016 primaries and presidential elections

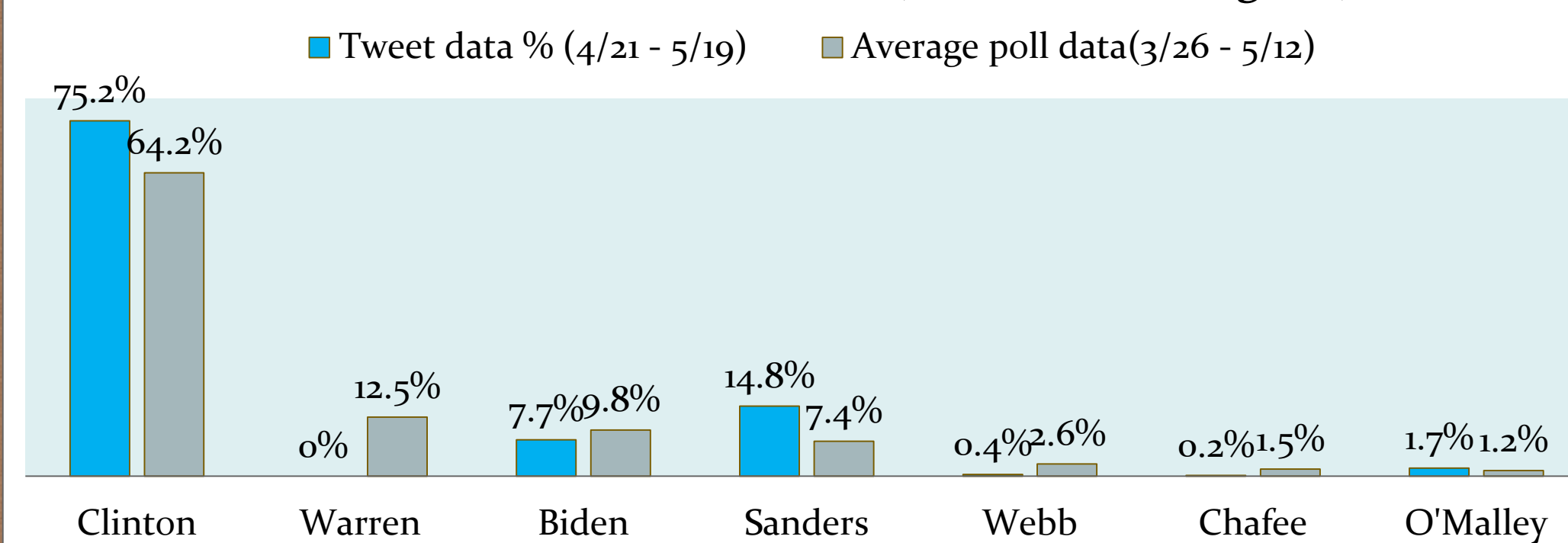
Data and Methods:

- Data: used Java programming language & Twitter4j API to gather geotagged Twitter data streamed from the public “garden hose” 24/7 for 4 weeks
- Methods: compared data with polls, sentiment analysis, flow map, network graph, temporal trends, spatial visualization

Twitter Total Versus Poll Total (Republican Average %)

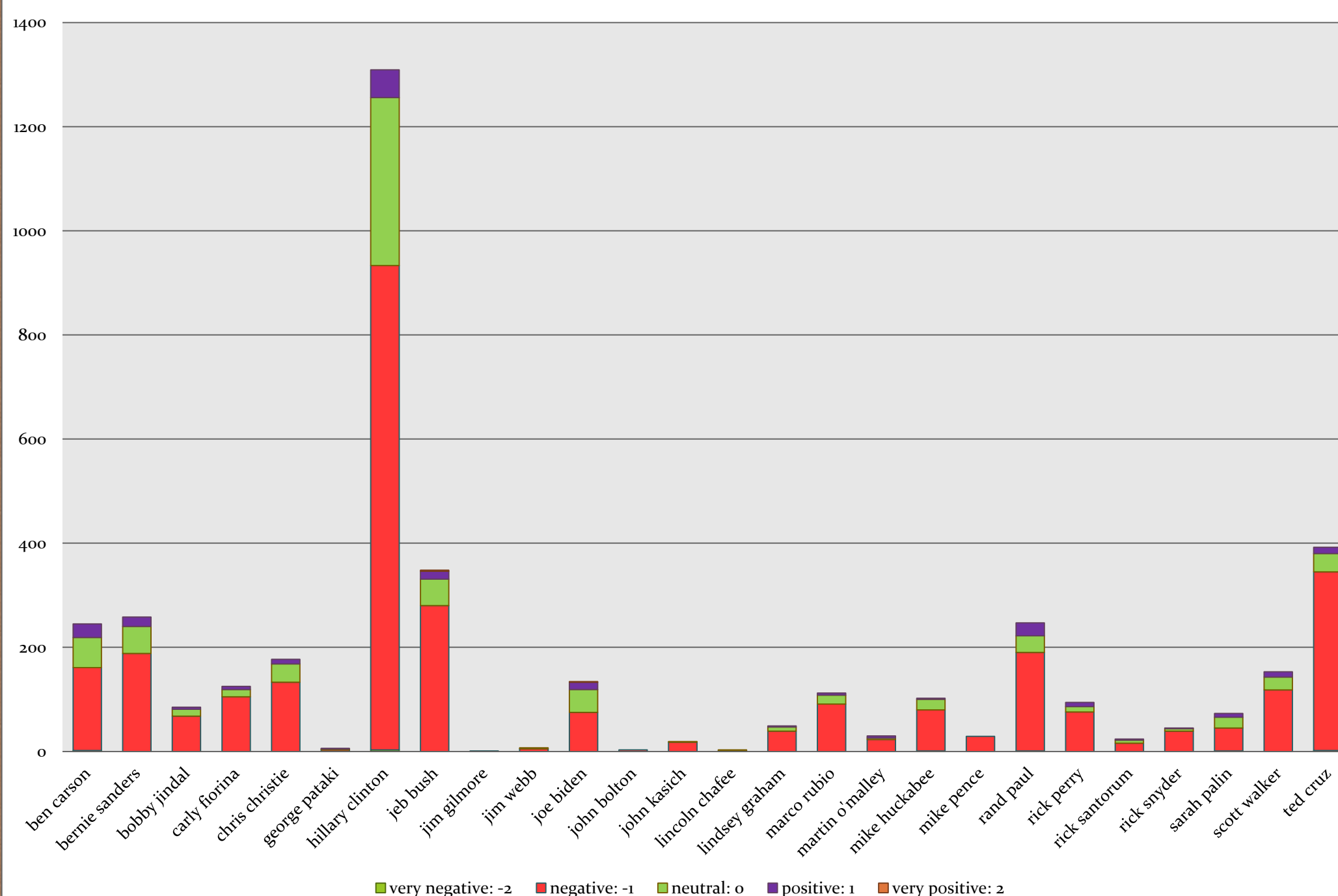


Twitter Total Versus Poll Total (Democrat Average %)



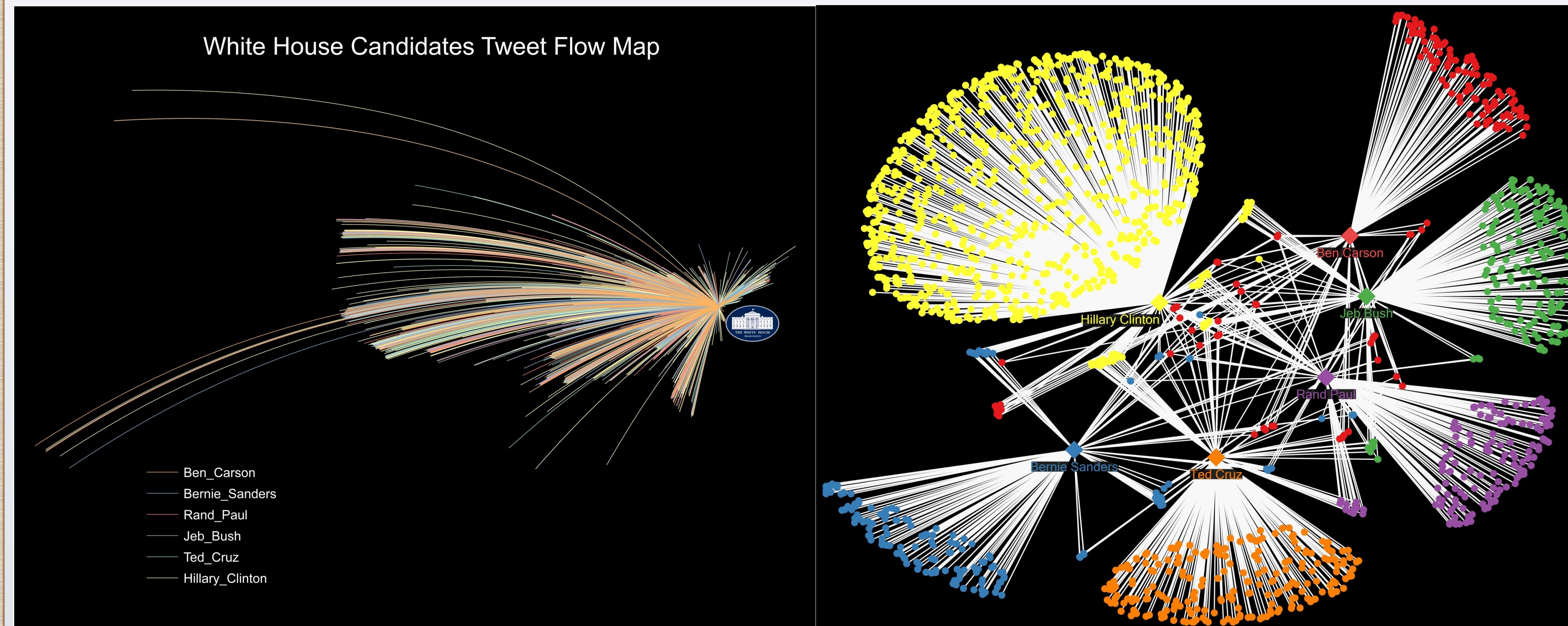
- Figure 1: shows each candidate's share of tweets vs. their poll results

Sentiment Analysis Results Per Candidate



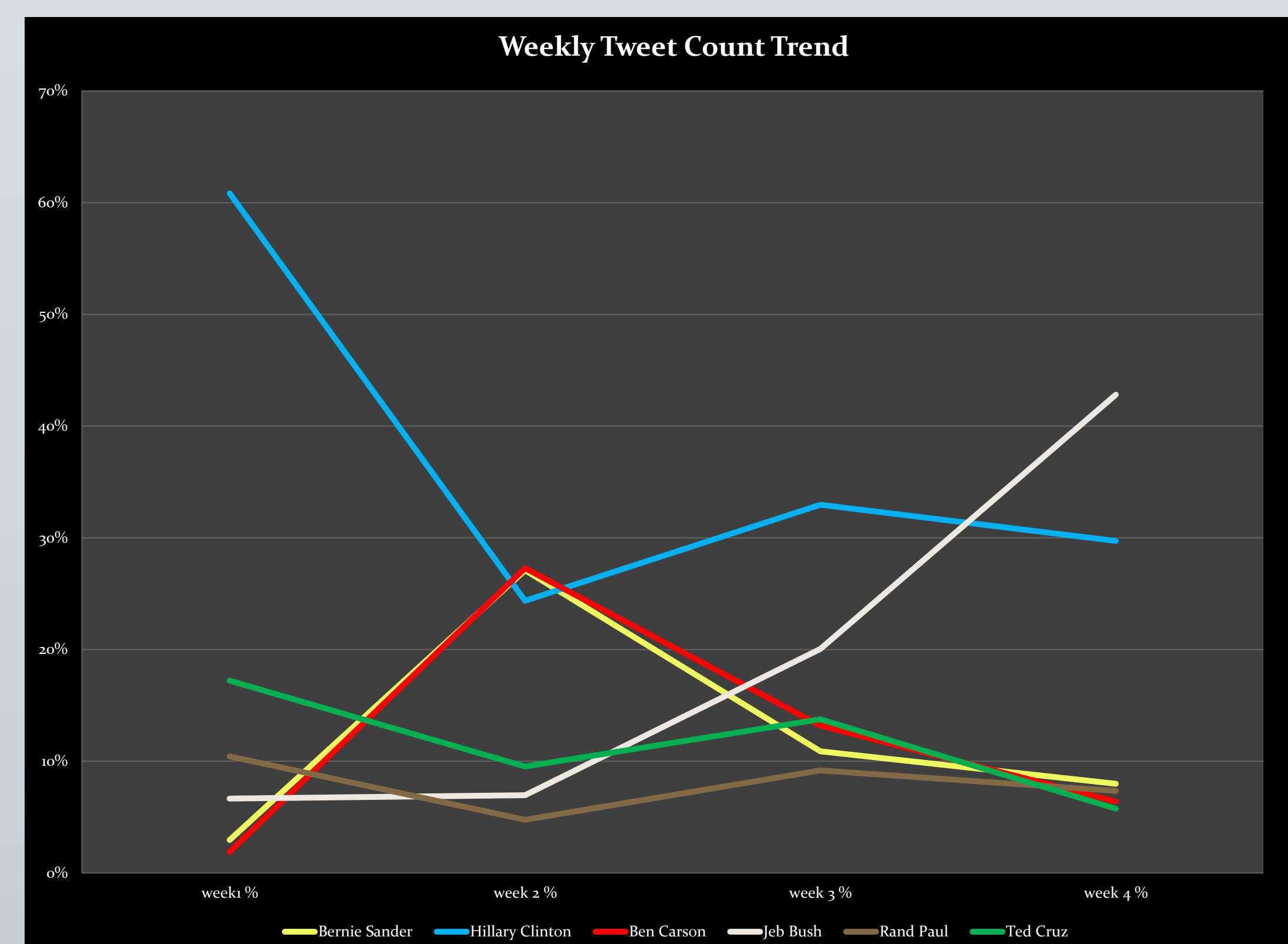
- Figure 2: shows each candidate's count of positive, neutral, or negative tweets (analysis performed with the Stanford Natural Language Processing tool)

Interesting Visualizations



- The flow map above was made with the ArcGIS tool
- Each of the top 6 candidates is represented by a different colored line, all heading to the White House
- Although there are no boundaries drawn, you can still make out the shape of the US, similar to a well known Facebook connection map of the world

- The graph above was made with the NodeXL tool in Excel
- Unlike the previous map where it is difficult to distinguish between lines due to overlapping, this one clearly shows the volume of tweets using nodes and edges
- Nodes with multiple edges represent users tweeting about different candidates

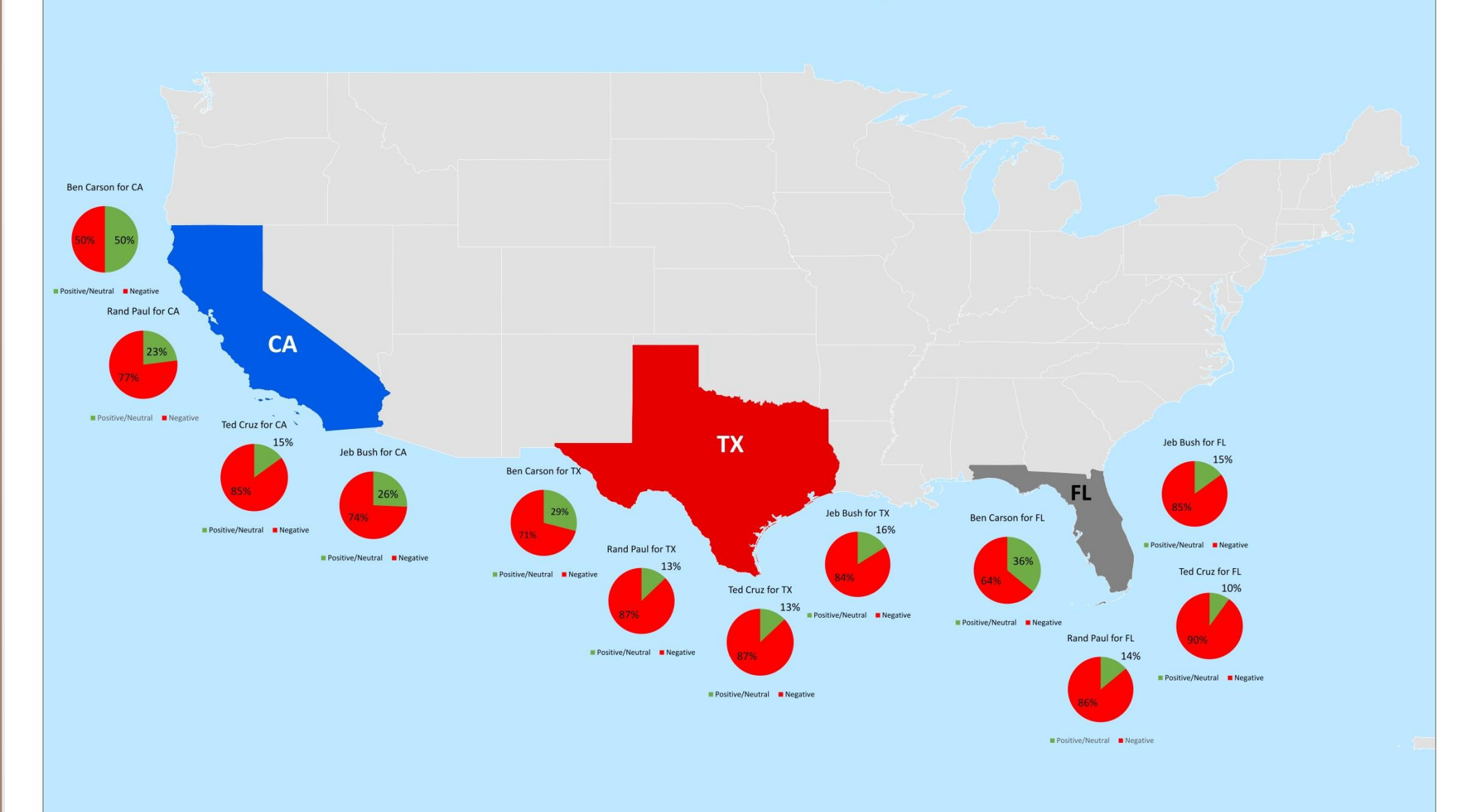


- Figure 5: shows the temporal aspect of the data set
- Events like candidate announcements, controversial speeches, or media attention are marked by gains
- Others remain constant or show no significant changes

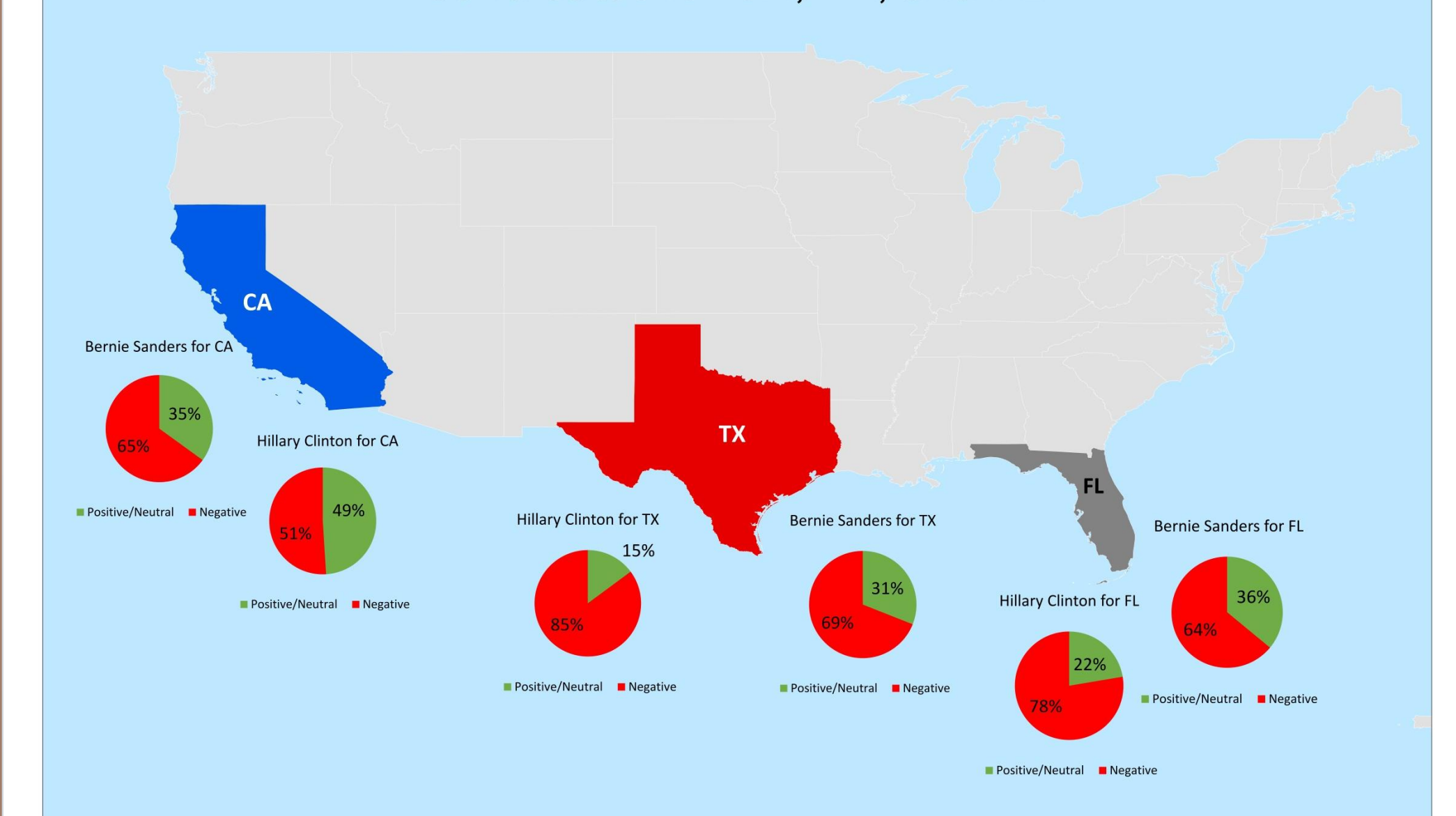
Conclusion:

- Interesting pattern: majority of tweet contents were negative; we speculated that people express negative thoughts more readily
- Challenges we faced: the majority of people who tweet were young and young people tend to vote Democrat, this lowered the number of Republican candidates' tweets making it difficult to see a clearly favored Republican candidate
- Two frontrunners: Hillary Clinton for the Democratic Party, due to her high tweet count and high positive sentiment score; Ben Carson for Republican party due his positive sentiment score.
- For the national presidential election, we predict that Hillary Clinton will become the next president since she had a much higher count than any other candidate and a better sentiment score than most.

Sentiment Analysis of Top 4 Republican Candidates for CA, TX, and FL



Sentiment Analysis of Top 2 Democratic Candidates for CA, TX, and FL



- Figure 6: shows the spatial aspect of the sentiment analyzed data for California (Democrat state), Texas (Republican state), and Florida (Swing state)

- On average the tweets about the Republic candidates are much more negative than the Democratic candidates within these 3 states
- The tweets gathered from Texas are the most negative, tweets from California are the most positive, and Florida has the most neutral tweets for the top 6 candidates

Acknowledgements:

- Special thanks to Song Gao, Grant McKenzie, Daniel Ervin, Carlos Baez, Krzysztof Janowicz, Werner Kuhn

References:

- Percentage of U.S. internet users who use Twitter in September 2014, by age group (2014). In Statista. Retrieved May 26
- The Generation Gap and the 2012 Election (2012, November 11). In PewResearchCenter. Retrieved May 26, 2015
- www.people-press.org
- www.realclearpolitics.com