Data Processing & Visualization QMSS G4063

Navid Hassanpour, [nh2519@columbia.edu](mailto:nh2519@columbia.edu)

Due Thu. March 24, 2016

Alexandra Plassaras (amp2261)

Assignment 3

Please submit your assignment on Courseworks and include links to your 1) code and 2) web-based visualizations in the report. If you have static visualizations, embed them in your submission file. There are bonus points for conducting the challenge part (note: it is optional) at the end of the assignment and uploading your Shiny app online. Do not upload large data files, all coding should be executable on the data files made available to you online, no need for large file uploads.

**Counting Geolocated Tweets in U.S. States,** (5×2 visualizations, 1000 words, optional link to online app) At this link, please access a daily collection of tweets pertaining to the Primaries.

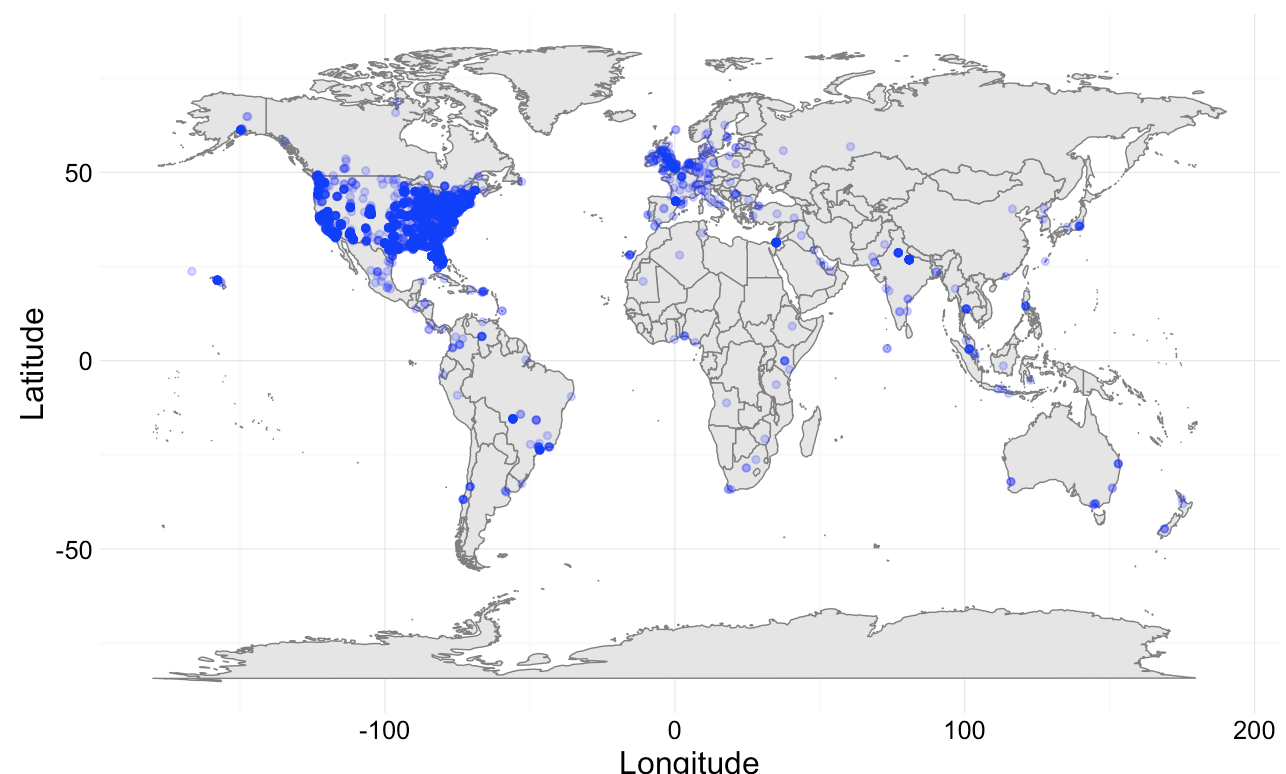
Download

tweets.02.09.2016.summary.json, tweets.02.20.2016.summary.json, tweets.02.23.2016.summary.json, tweets.02.27.2016.summary.json, tweets.03.01.2016.summary.json, tweets.03.05.2016.summary.json, tweets.03.06.2016.summary.json, tweets.03.08.2016.summary.json, tweets.03.15.2016.summary.json, (tweets.03.22.2016.summary.json).

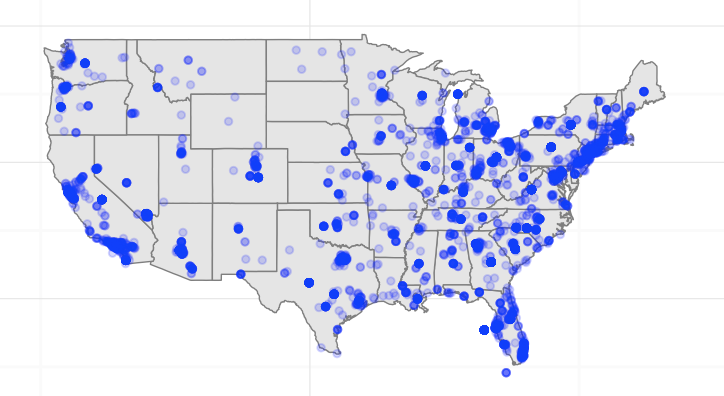
These are summary tweets for the dates on which primaries and caucuses were held. Using the streamR parsing command and what you learned in assignment 1, change the .json file into a r data frame and divide the results into five subsets for each of the five viable candidates in the race: Clinton, Cruz, Rubio, Sanders, Trump. 1 Then using GIS techniques in R, generate the following two sets of visualizations, 5 each:

* **Generate five maps showing the origins of tweets on each of the candidates. Do most of them come from inside the U.S.?**

Hillary Clinton – World Tweets

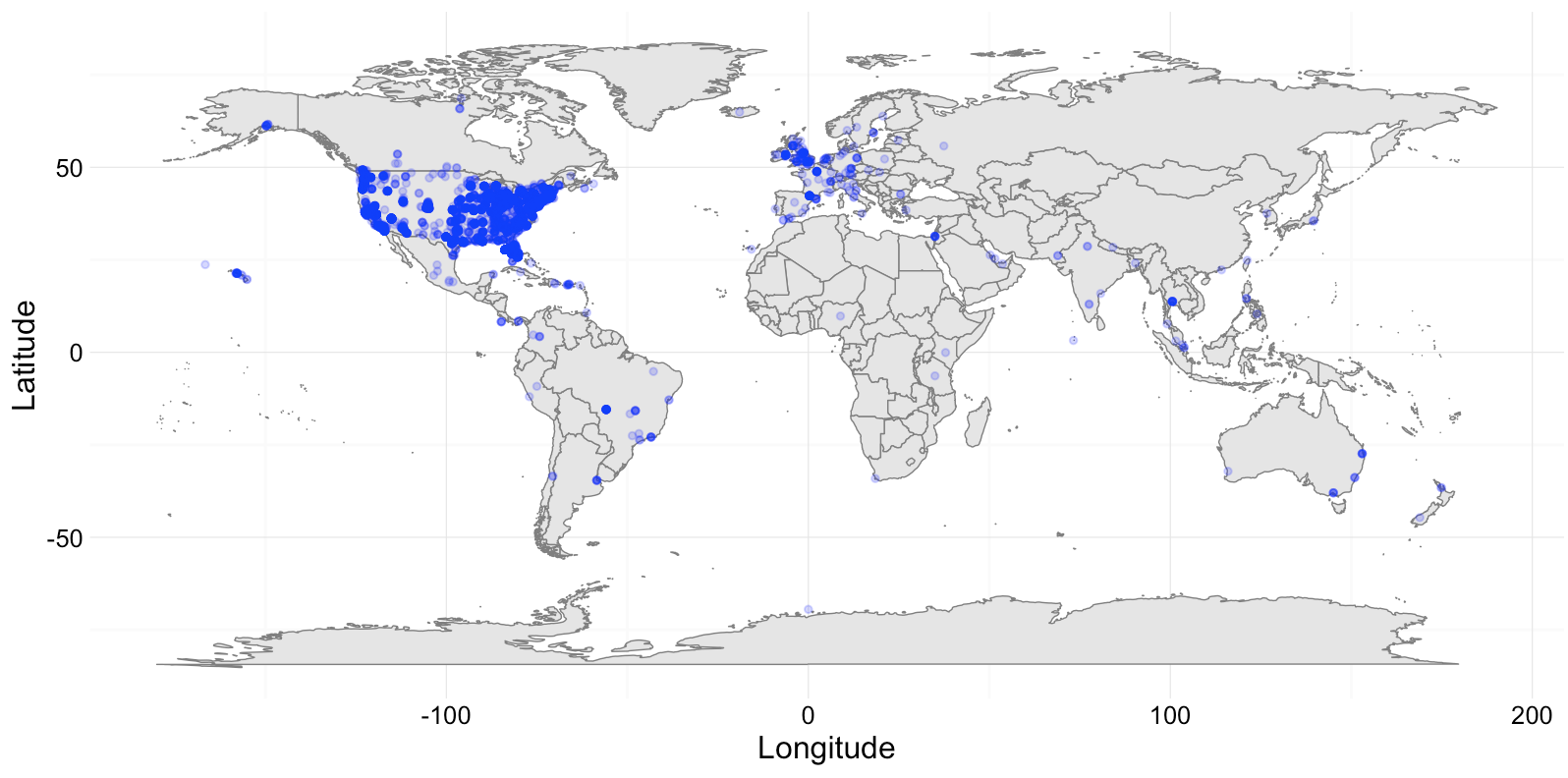


Hillary Clinton – US Tweets

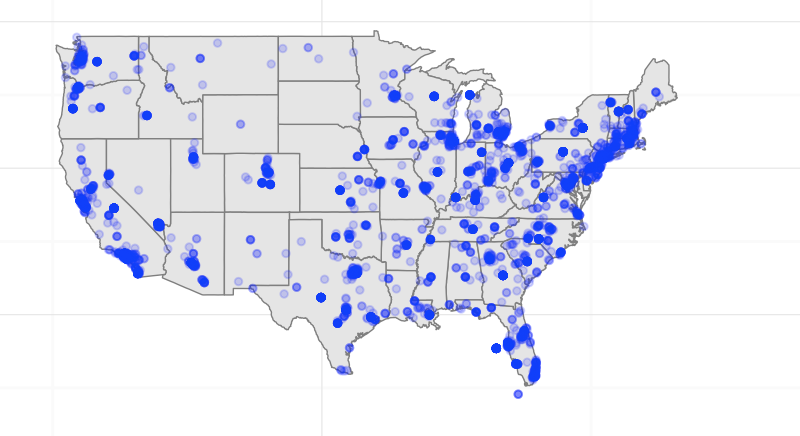


From the two visualizations above it appears that the majority of tweets mentioning Hillary Clinton are originating from the US. There are also many tweets referring to her across Europe and other parts of the world. Within the US tweets mentioning her are originating from both the East and West Coast as well as eastern central states. The low amount of tweets coming from Nebraska, Wyoming, and South Dakota may be due to the low amount of tweet accounts from those states. Globally, Clinton’s reach may be in part to her global popularity as well as the fact that was is a former Secretary of State for the US.

Bernie Sanders – World Tweets

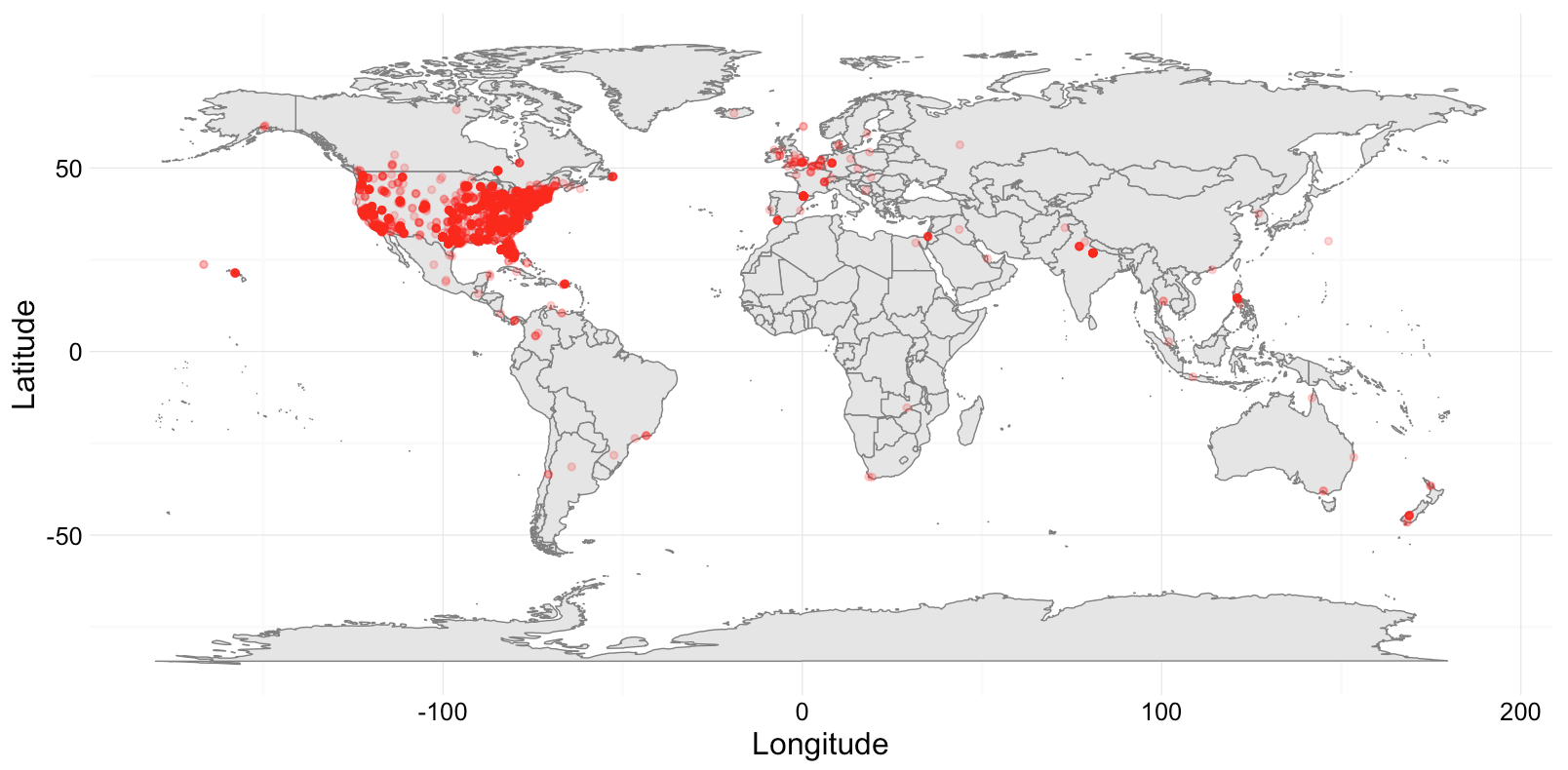


Bernie Sanders – US Tweets

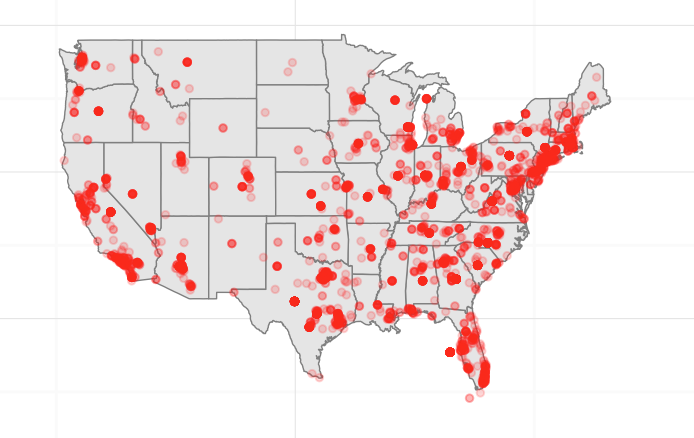


As with Hillary Clinton, Bernie Sanders appears to have the majority of his tweets originating from the US which is no surprise given that all of these individuals are US Presidential candidates. While Sanders also has a small cluster of tweets originating from Europe, tweets from other countries and regions outside of the US and Europe appear to be less than Clinton’s. In the US, Sanders’ appears to have more tweets mentioning his name in the East Coast and the North Central. Again, low tweets from the West Central may be due to the low volume of twitter accounts from people living in those states.

Ted Cruz – World Tweets

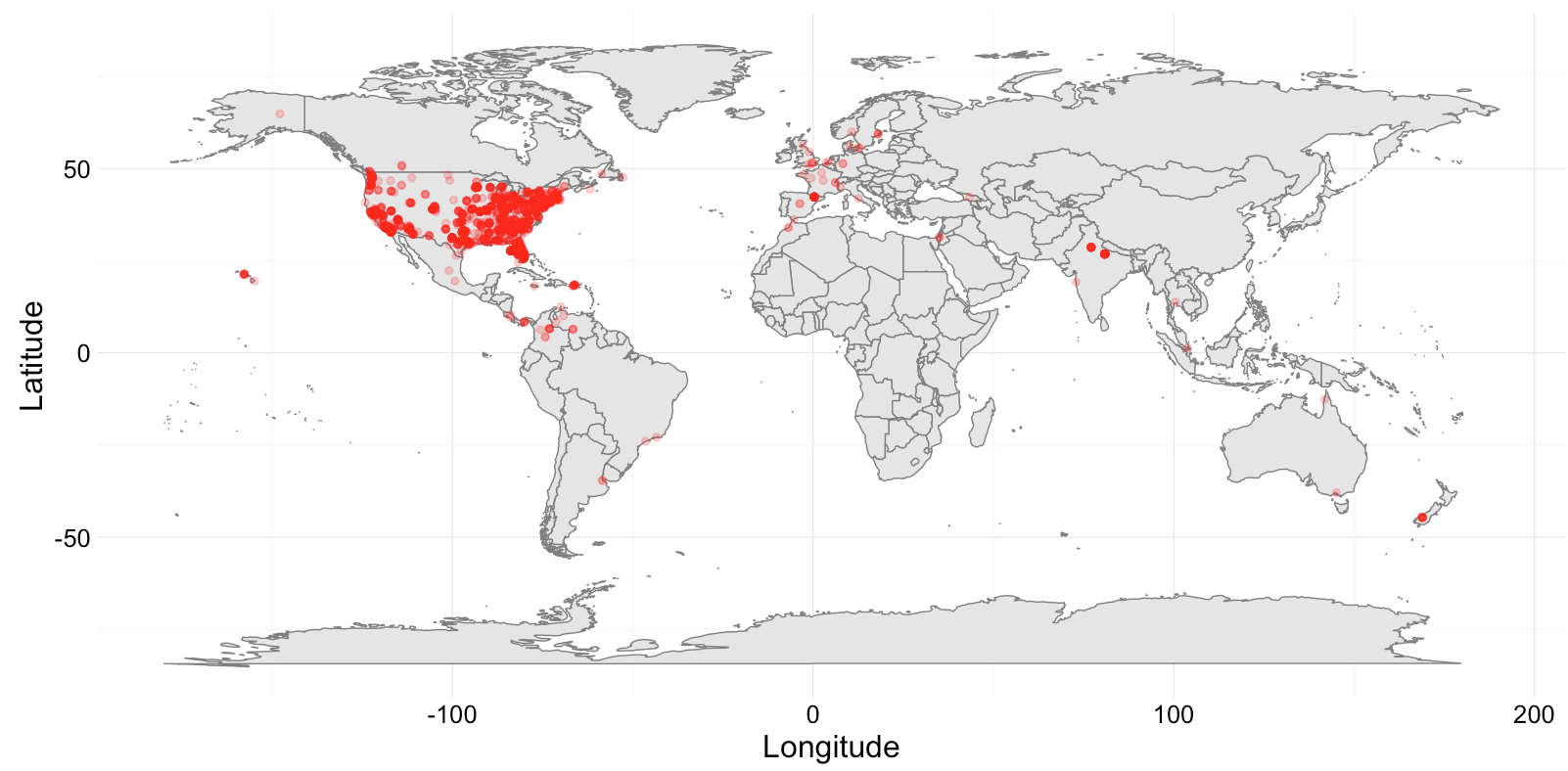


Ted Cruz – US Tweets

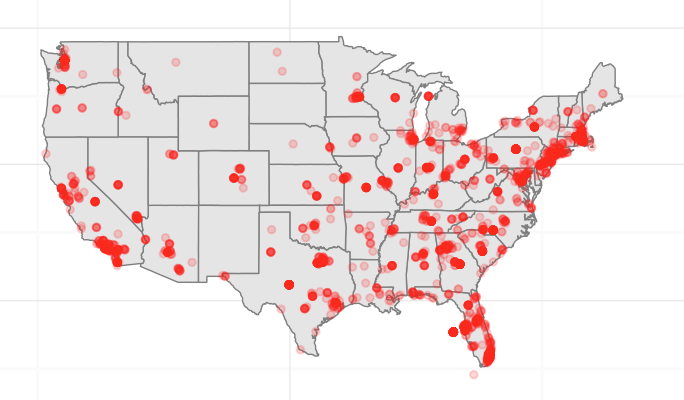


The majority of tweets mentioning Ted Cruz originate within the US with Europe coming in as second place for tweets mentioning Cruz. Compared to Clinton and Sanders, Cruz’s global twitter reach is smaller than the first two candidates. Within the US, Cruz’s tweets appear to be less than Clinton’s and Sanders’ throughout the country. Within his own tweets though, the East Coast appears to have more tweets mentioning his name.

Marco Rubio – World Tweets

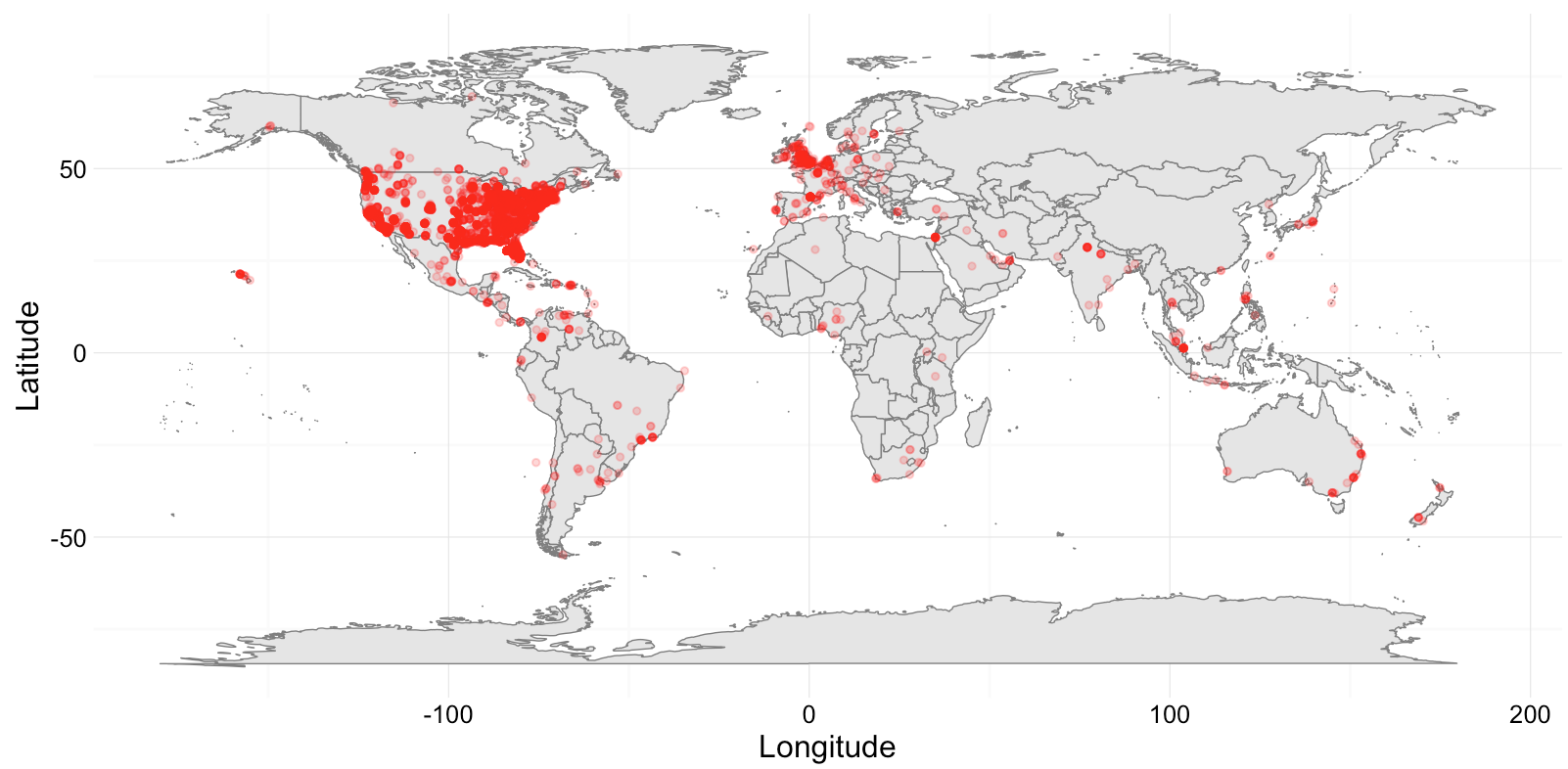


Marco Rubio – US Tweets

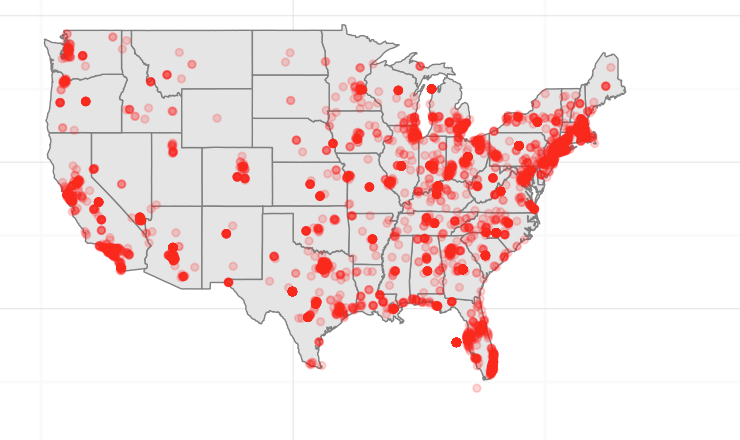


When compared to every other candidate, Marco Rubio appears to have the smallest amount of tweets in general as well as the lowest amount of tweets with a global reach. Outside of a handful of tweets in Europe, Rubio has only a few tweets in India, New Zealand, Colombia and Israel. Within the US the East Coast appears to have more tweets mentioning Rubio than the West Coast and central regions.

Donald Trump – World Tweets



Donald Trump – US Tweets



Donald Trump’s twitter following has a large global following that is on or around par with Clinton’s. Europe has a large amount of twitter mentions of Trump as does Central and South America, parts of Africa and the Middle East and Southeast Asia. An unproven hypothesis is that he is being mentioned so prevalently around the world because of his inflammatory words against the global Muslim community. Within the US, Trump mentioning is occurring heavily on the East Coast, North Central region and in California. He also has a large amount of mentions scattered throughout the South.

* **Count the number of geolocated tweets from each U.S. state on each of the five candidates. Based on the counts produce five U.S. maps, one for each candidate, with states color-coded for showing the proportion of tweets coming from each state.**

|  |  |
| --- | --- |
| Hillary Clinton | |
| Tweet Count by State  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 9.10.47 PM.png  **Top 5 States:** California, New York, Florida, Texas, Ohio | Tweets Per Capita  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 10.19.46 PM.png  **Top 5 states:** DC, Nevada, New York, Rhode Island, Oregon |
| Bernie Sanders | |
| Tweet Count by State  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 9.11.00 PM.png  **Top 5 States:** California, New York, Texas, Florida, Illinois | Tweets Per Capita  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 10.35.23 PM.png  **Top 5 States:** Vermont, DC, Nevada, New York, Oregon |
| Ted Cruz | |
| Tweet Count by State  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 9.11.15 PM.png  **Top 5 States:** Texas, California, New York, Georgia, Florida | Tweets Per Capita  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 10.36.00 PM.png  **Top 5 States:** DC, Nevada, Georgia, Texas, Kansas |
| Marco Rubio | |
| Tweet Count by State  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 9.11.37 PM.png  **Top 5 States:** Florida, California, Texas, Georgia, New York | Tweets Per Capita  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 10.37.50 PM.png  **Top 5 States:** DC, Nevada, Florida, Georgia, South Carolina |
| Donald Trump | |
| Tweet Count by State  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 9.11.26 PM.png  **Top 5 States:** California, Texas, New York, Florida, Georgia | Tweets Per Capita  Macintosh HD:Users:alexandraplassaras:Desktop:Screen Shot 2016-03-24 at 10.38.36 PM.png  **Top 5 States:** DC, Nevada, Arizona, Georgia, Kentucky |

**Describe and interpret what you see in the plots. Is it true that there were more tweets in states holding primaries? Can you see any difference among those states with primaries, and those without? (1000 words).**

Based on the counts produced for each candidate, the states that have the highest counts of tweets tended to be the states that were most highly populated – California, New York, Texas and Florida. Looking at the top five states that mentioned a candidate, Clinton’s fifth state was Ohio and Sanders’ was Illinois. For all three Republican’s the fifth state that had the highest amount of mentioned for the candidates was Georgia. To gain better insights on tweet mentions by candidate the right-hand maps depict the rate of tweets per capita. In these maps, DC, Nevada, New York and Oregon all had high tweet per capita ratios for Clinton and Sanders. Rhode Island was Clinton’s next highest state while Vermont was Sanders (which is his home state).It is interesting to note that New York and Oregon have not yet had their primaries yet have high rates of tweets per capita. DC’s primaries were on March 12, which was not a day in which our tweets were scraped but it does fall within the date range of collection.

For the Republicans, DC, Nevada, Georgia were states that had high rates of tweets per capita for all three candidates. These states were also states in which primaries were held within the collection date range. Cruz’s other top two states were Texas (his home state) and Kansas (both of which had had primaries within the collection range). Rubio’s other top two were Florida(his home state) and South Carolina. As for Trump, his other top two states were Arizona and Kentucky. His high rate in Arizona might be explained by his rhetoric on building a wall between the US and Mexico. In regards to the Republicans, all of the top five states with the highest tweet per capita were from states that had had primaries within the data collection range. The Democrats however had two states with high rates of tweets per capita that had not yet had their primaries.

Primary dates:

* 02.09.2016 – New Hampshire
* 02.20.2016 – Nevada , South Carolina
* 02.23.2016 - Nevada
* 02.27.2016 – South Carolina
* 03.01.2016 – Alabama, Alaska, Arkansas, Colorado, Georgia, Massachusetts, Minnesota, Oklahoma, Tennessee, Texas, Vermont, Virginia, American Samoa
* 03.05.2016 – Kansas, Kentucky, Louisiana, Maine, Nebraska
* 03.06.2016 – Maine, Puerto Rico
* 03.08.2016 – Hawaii, Idaho, Michigan, Mississippi
* 03.15.2016 – Illinois, North Carolina, Ohio, Florida, Missouri, Northern Mariana Islands
* 03.22.2016 – Arizona, Utah, Idaho

**Extra points: Use the D3 examples in the following two links (Link1 and Link2) to generate a static tweet production choropleth for each of the five candidates. This plot would be similar to the one you generated in R. Now using daily (or hourly) tweet count choropleth plots, produce an animated choropleth for each candidate, transitioning at each day (or hour). Interpret your dynamic plots.**

I attempted to work on the D3 examples but I was unable to generate successful code. I have attached my attempts to my homework submission.