Assignment 3, Data Processing & Visualization QMSS G4063
Navid Hassanpour, `nh2519@columbia.edu`
Due Thu. March 24, 2016

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.03.01.2016.json` (or the smaller version, `tweets.03.01.2016.summary.json`, if your computer can not process the larger file). These are tweets from Super Tuesday. Using the streamR parsing command and what you learned in assignment 1, change the .json file into a r data frame and divide the file 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.?

- 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.

Describe and interpret what you see in the plots. Is it true that there were more tweets in states holding the Super Tuesday Primaries? Can you see any difference among those states with primaries on March 1, and those without? (1000 words).

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 *hourly* tweet count choropleth plots, produce an animated choropleth for each candidate, transitioning at each hour of March 1. Interpret your dynamic plots.

---

[1] Obviously you can choose to divide the .json file instead of importing it in R first.