Assignment 1, Data Processing & Visualization QMSS G4063
Navid Hassanpour, nh2519@columbia.edu
Due February 16, 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.

1. Tweet collection and RShiny app, **30%** (Two visualizations, 200 words, link to the online app)

Use the code you write in R or Python to collect 30 minutes (or longer if you wish) worth of tweets based on a keyword filter. You keywords should be (case insensitive)

Hillary Clinton, Bernie Sanders, Jeb Bush, Marco Rubio, Ted Cruz, Donald Trump, John kasich, Ben Carson, Chris Christie, Carly Fiorina, Hillaryclinton, Jebbush, Berniesanders, Marcorubio, Chrischristie, Bencarson, Johnkasich, Donaldtrump, Tedcruz, CarlyFiorina

Produce two visualizations based on what we covered in the class, embed them in an interactive R Shiny app, and include a link to your app in your report. Write a 200 words description of what you depict.

2. D3 Visualization using JSON dataset, **70%** (Interactive and animated visualization, comparison among candidates, 800 words, link to the .html output online)

Access the JSON files of Primaries tweets at the link you received via email. The files contain tweets on the Primaries from February 7 (you can use the smaller file if your computer can not read the larger file into the memory). Use the D3 techniques you have learned and sample code you find online to produce the following visualizations:

1. Tweet frequencies per candidates (at least four) per hour for the 24 hours included in the JSON files you use.

2. Use the above mentioned vectors of size 24 to arrange transitions between candidate pairs you choose, after a click on the diagram.

3. Use the total frequencies per candidate to produce diagrams for comparing the frequencies. Choose a type of comparative diagram that fits your data the best. Feel free to consult samples at this link.

For part 2 of the assignment you can produce as many visualization as you wish (up to 8) different candidates' temporal and cross-sectional relations. Use animation (transition) and user-interaction techniques to generate a dynamic profile of the data. Interpret what you see.