Assignment 2, Data Processing & Visualization QMSS G4063
Navid Hassanpour, nh2519@columbia.edu
Due Thu. March 3, 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.

Twitter Conversation Network Visualization and Analysis, (five+ visualizations, 1000 words, optional link to the online app)

Go to the link. Download the 8 files therein, four `edges_[candidate].csv` and four `nodes_[candidate].cs` along one one larger file `edges_all.csv`. These are data for mention networks associated with tweet capture based on the name of each contender. `edges_all.csv` was harvested using a combination of all election-related keywords.

1. Generate the following five D3 visualizations:

   - A force layout visualization of mention network using `edges_all.csv`, and code for straight link depiction of the network. What do you see? Describe in 200 words
   - Now use the remaining edge and node .csv files (4+4) to reconstruct twitter conversation network for four candidates (Clinton, Sanders) and (Cruz, Trump). Ensure you have node labels (twitter handles) added to the visualization using the code and instructions provided in lecture 6 (accessible via course website). These visualizations are going to be *force layouts* with curved, directed edges, and labeled nodes.

   How do you compare the 4 networks above? Who are the main mentioned tweeters? Who are the main mentioning tweeters? (Remember, the graph is directed.) What do you confer from the comparison between Clinton and Sanders networks? What about the Cruz and Trump comparison? Write 400 words.

2. Import the files in the proper format into Gephi and calculate network parameters such as average degree, average clustering coefficient, network diameter, and average path length. Do you see meaningful difference between the four candidates? in 400 words describe the network statistics and noteworthy comparisons you find.

3. (bonus points, to be added to the student's grade for assignment 1) For each of the four aforementioned networks, generate force layout network visualizations with *straight* lines (see M Bostock's code, linked on lecture6.md) *with node labels*. If you are able to produce the labeled plots, upload them as a part of an RShiny app on shinyapps.io.