SMI 606 (Intro. to Quant. Data Analysis)
Dr. Hartman

Assessment 3

To help you complete this assessment, you may use resources (e.g., books, lecture notes, online guides, etc.) and/or work with other students in the class, but each person must hand in their own work (i.e., you are not allowed to copy text from someone else—that is plagiarism!).

R Markdown

You should upload your R Markdown file directly to MOLE, which will provide me with your R script, results, and any notes that you include to answer each question.

To create an R Markdown file, you should install the rmarkdown package via 'pacman' from the command line:

```
> pacman::p_load(rmarkdown)
```

And then render your R script from the command line (NOT from the R script you wish to run!):

```
> rmarkdown::render("your R script file location and name here.R")
```

Make sure you comment out (place a # in front of) any lines that will cause errors for the markdown process (e.g., 'setwd()', 'file.choose()', 'rmarkdown::render()', comments, etc.).

You should start your R script with the following header:

```
#' ---
#' title: "SMI 606 Assessment 2"
#' author: "Anonymous"
#' date: "`r format(Sys.Date())`"
#' ---
```

Remember, that at its most basic form, R Markdown is just an R script in which comments # are replaced by #' as in the following:

```
#' This is an example of a heading comment using the R Markdown formatting.
```

Discovery Using Textual, Network, or Spatial Data

For this assessment, you'll develop a research question (or questions) related to your field of study (or the social sciences, more generally) and explore it (them) using textual, network, and/or spatial data from the web. You must do some essential things described below, but you're free to explore and analyze the data as you see fit (or conduct additional analyses, visualizations, etc.). Basically, this is your opportunity to show me what you've learned this semester.

1. Describe your research question, as well as why you think it's interesting. What do you hope to learn from the data?

2. Describe your data. What type of data will you be using? Where did you get it from? How many observations are in the dataset? What are your key variables or measures, and how are they scaled/coded? What do low/high values on the scale mean? How well does this variable capture your outcome of interest? Did you create the variables or were they already contained in the dataset? Did you merge data from multiple sources or use a single dataset? Be detailed.

   In addition to databases that may be useful such as the UK Data Archive, Data.Gov.uk, Michigan's ICPSR, and Harvard's Dataverse, there are other open repositories like Kaggle, which may be of interest. And, of course, there's always Google. Below are some specific suggestions regarding where to find data:

   Textual Data: Project Gutenberg (books), Manifesto Project (political documents), LexisNexis Academic (TV and radio news transcripts), (screenplays)

   Network Data: UCINET, Network Repository, UC Irvine Network Data Repository

   Spatial Data: No specific datasets (search for geocoded or spatial data)

   You should not use the QSS datasets for this task. You may use data from any of the lab exercises we did in class (provided you do not repeat the exact same analysis detailed in the handout; e.g., you can map election data but not the Brexit vote).

3. Describe your analytical approach. What analyses did you conduct and why? What are the strengths/weaknesses of this approach? If you used more than one method, then discuss how these methods compare to each other.

4. Substantively interpret your results. Make sure that you thoroughly discuss your findings and compare different outputs/statistics across approaches. For instance, if you analyze network data, you should discuss various statistics that measure centrality. Overall, what do the results tell you?

5. Visualize your results. Again, discuss how your figure(s) help illustrate your main findings. Make sure that you label your figures with titles, axis labels, etc. Note some figures may not show up in your R Markdown file and will need to be manually uploaded to TurnItIn. Let me know if this is a problem.