GEORGETOWN UNIVERSITY PUBLIC POLICY 646
DATA VISUALIZATION
ASSIGNMENT 5: BUILDING YOUR PORTFOLIO: TWO MORE ORIGINAL VISUALIZATIONS IN R
Prof. Wesley Joe

For this assignment, please create two original data visualizations from your choice of data. You may use the same data set for each of the two graphs, but you are not required to do so. Please create both of the graphs using R and refine each graph in Adobe Illustrator.

For each visualization, please begin by writing a brief "statement of purpose" paragraph that explains the purpose of the graph. (This does not need to be more than a few sentences.) Do you want viewers to explore the data to answer their own questions? Or do you want the visualization to support a particular point that you want to make, perhaps to reinforce something that you explain in accompanying written narrative? Do you want to emphasize specific associations? Patterns? Trends? Exceptions?

Then, create your visualization. You can choose the types of visualization. To demonstrate your growing range of capabilities, each graph should consist of a separate type of plot. In other words, please avoid submitting something like two separate line graph visualizations. (Remember that the strongest final portfolios will demonstrate the author's capacity to use a variety of visualization types.)

There is no minimum number of variables required for a visualization. That freedom notwithstanding, remember that the strongest visualizations add value by revealing non-obvious insights from the data. At a minimum, a visualization should contribute a nontrivial insight that supports a point you would make in a presentation, a thesis, an article, a research report, or the like. Now that you are becoming more comfortable with the technical tools and basic presentation requirements (such as useful titles, annotation, axes, and so on), the strongest work will consist of visualizations convey insights that a reader could not easily discover by simply looking at a small table of the data, or from a few summary statistics based on the raw data.

Each visualization should include enough contextual information (such as titles, unit labels, axis labels, and so on) for your audience to easily find the main point(s) of the visualization. In the interest of creating cleaner, more efficient graphs, however, also try to minimize the presence of "non-data ink" in your graphs.

Also, each visualization should include annotation. For examples, see the *New York Times*, *Washington Post*, *Wall Street Journal*, Pew Research Center, and *The Economist* web sites. Each visualization's annotation should its readers quickly find each the key takeaway point(s). Finally, each visualization should be polished and aesthetically pleasing to a professional audience. When creating the annotation, assume that your

audience will have access *only* to the visualization. In other words, assume that your audience will *not* have access to the "statement of purpose" paragraph that you write for me. (That paragraph helps me to evaluate how well your graph fulfills your intentions for it.) Hence the data source credit, for example, should appear in the graph.

Please use Adobe Illustrator to add the annotation, indicate the source of the data, and otherwise refine each of your two visualizations. For each text element (for example the title, subhead, annotation, axis and tic mark labels, data source credit, etc.), the size/scale and style (for example plain, bold, italic) of the type should reflect the element's relative importance in a visual hierarchy. Strive generally to organize each visualization so that the reader's focus turns first to the elements that communicate the primary insight(s). Throughout the visualization, apply principles of sound visual hierarchy and Mackinlay's expressiveness and effectiveness criteria.

Please turn in a single PDF file that contains six pages:

- a brief statement of the purpose of the first original visualization
- the first original visualization
- your R code for the first original visualization
- a brief statement of the purpose of the second original visualization
- the second original visualization
- your R code for the second original visualization

Finally, a word about sharing code and appropriating code examples from the Internet. Learning to write code from scratch is an important learning objective of this course. You'll meet this objective only by writing your own code. Consequently, for assignments in this class, students cannot obtain code from any current or former McCourt student. You can, however, use online resources, such as StackOverflow or other StackExchange sites, RStudio Community, and other the like. If you use code from such sites, even as a model, please explicitly cite the source that you used. Please include the citation in the "statement of purpose" associated with the visualization for which you used the site. Any shared or recycled code that is used, even as a model, without attribution will be considered a violation of the academic honesty guidelines.

As before, I will deduct 0.2 points for PDFs that have pages out of this requested order. Thanks for your understanding! ☺