Project · Bring Your Own
Due Friday, April 8, 2016 (5 PM)

Project description

The assignment for this project is simple: pose an interesting question; collect a relevant data set; and use the data, in conjuction with the tools we have learned in class, to answer the question you have posed. Make sure to quantify any uncertainty that arises in answering your question, and to address any shortcomings in the answer provided by your data and analysis.[1]

This assignment is purposely open-ended, allowing you considerable freedom to follow a path dictated by your own intellectual curiosity. Strive to write something that a statistically literate person of wide-ranging interests (for example, a future employer) would find engaging and impressive. You will be evaluated both on the technical correctness and the overall intellectual quality of your presentation. Be impressive.

To turn in

You should turn in the following three items. As with the homeworks, you may work in groups of 4 people or fewer, or you may turn in your own project. If you work in a group, only one set of these items (bearing all of your names) needs to be turned in.

1. A written project report that describes your question, your data sources, your methodological approach, and your conclusions.[2]

2. The data set itself, in .csv format.

3. The R script used to analyze your data. If your analysis and plots are not 100% reproducible, you will not receive a passing grade.

The first item should be submitted as a hard copy, either in person or to my office (CBA 6.478) or mailbox (CBA 5.202). All three items should be e-mailed to the drop box at `statdropbox@gmail.com`. The subject line of your e-mail should be: "Project: (names)," where you fill in the blank with the full names of all your group members.

[1] Where can one find data? Everywhere! Here's a short list of sources you might consider: major newspapers, academic journals, the Economist, Twitter, the World Bank, ESPN.com, Craigslist, Amazon, EBay, the Bureau of Labor Statistics, Facebook, the World Economic Forum, the OECD Factbook, the CIA World Factbook, the Securities and Exchange Commission, Yahoo finance, Google Public Data Explorer, the census, the Federal Reserve, your own vital signs, your own experiment or survey, your favorite blogs, your other classes, and your friends. As template questions, you should recall some of those we've used for in-class activities so far (house prices versus size and neighborhood; GDP growth regressions; the price of a gallon of milk; SAP customers versus their industry peers; and so forth). These are good templates. You might even consider recapitulating a similar analysis using a different different set.

[2] A reasonable length here would be 4–6 type-written pages including figures, but treat this only as a rough guideline rather than an absolute quota or limit.