University of Wisconsin – Madison
Econ 690 [Fall 2018], Data analytics for economists
Project information

The project is an important part of this class — and your grade. The project is also a great opportunity to showcase what you have learned both to me and to future employers. The project is made up of two parts.

1. **Code and data.** You will submit a jupyter notebook and any other files needed to make the code work. This notebook should explain what you are doing (and have many useful comments) but does not contain your questions/conclusions. Think of this as the document you would give a coworker who would like to replicate your results and understand what you have done to get to those results.
2. **Project report.** This is a professional report to be turned in as a pdf. This report clearly lays out the question(s) you are addressing, discusses the analysis you performed in the notebook and draws a conclusion. This report should be no more than three pages long (not including figures). This is not a technical document, but you should discuss any technical difficulties and any assumption you made in your analysis.

Due dates

November 26 (end of class) Project proposal. The proposal is a two page (max) document that clearly states what question(s) you would like to address and what data you are using to address it. By the time you submit your proposal, you should have already found the data and loaded into a notebook. The proposal should include a discussion of which variables you will be using.

The proposal should consist of three sections, labeled 1) Question 2) Data 3) Expected results

You should expect your question(s) and expected results to change a bit as you study the data more and learn new things. That is how research works. The proposal is to ensure that you and I agree that the project is off on the right foot.

December 14 (midnight) Final project files. Sent by email to me at ruhl2@wisc.edu with the subject 'Econ 690 Project files'. The email must contain a zipped file with the project report, the jupyter notebook, and any other files needed to run the code. The zipped file must be named 'Last_First_Econ_690' where 'Last' and 'First' are your last and first names.

Grading

Projects will be graded on both the quality and readability of the code and the project report. Some of the things I will be looking for include:

1. Project idea: Is the question clearly stated? Are the data appropriate? Is the approach appropriate?
2. Code: Does it run? Is it commented? Is it readable? Are complicated parts of the code explained?
3. Project report: Is the report clearly written? Are the graphs labeled? Do they display graphical excellence?

Generating ideas

Many coding or statistics classes that require projects assign project ideas to students. Not here! Part of the data mentality that we are developing is being able to ask good questions. Like any skill, this takes some practice.

- Read. Blogs, newspapers, magazines,... almost anything can spark a question. Reading publications that are not afraid to show some data are even better.
- Think about other classes you have taken. What questions came up in labor class? Industrial organization? Sociology?
- What interests you outside of economics? Sports, theater, food? What questions and data are out there?
- Keep an idea journal. Mine is a text file on my computer. Anytime an idea strikes — no matter how crazy it might sound at the time — I write it down. Periodically I go over my journal see what questions I might be able to make some progress on.
- The course resources page (http://kimjruhl.com/data-analytics-resources) is the beginning of a list of data sources and writing about data. Send me more ideas!

Some examples

This class was first developed at NYU's Stern School of Business by a group of economists, of which I was a part. Since this is the first time it has been offered at UW, I am linking to some projects from past versions of the course at NYU. Next year, some of **your projects** will be listed here!

These projects represent a range of difficulties, topics, data, and approaches. They are meant to give you some idea of what **you** can do!

- This project uses data from the American Community Survey to study the geography of income inequality in New York. Nice use of map visualizations. (https://goo.gl/iKJDRm)
- A project about NFL quarterbacks. Includes some more formal regression analysis. (https://goo.gl/pgZuGN)
- This project looks at the causes of death, and how they change over time. (https://goo.gl/WrDesa)
- A project using NOAA data to analyze weather patterns. (https://goo.gl/DDhRHX)

Keep in mind that your project format is different from the ones above. They have mixed their project report and their code together. Outside of data science blogs, people tend not to communicate that way. Code is written for other coders and reports are written for people who may not know how to code, but are interested in questions and answers.