Techniques of Political Analysis:
Visualizing and Interpreting Data

Political Science 4781

Syllabus

August 23, 2012
Course description

Many, if not most, of the major debates in modern political science revolve around questions that can be addressed with data. The sources of voting behavior, the correlates of war, the determinants of development, political economy, psychology, institutions, and conflict—all are issues that are amenable to data-based analysis.

At the same time, the amount of available data and the number of publicly-available open-source tools for cleaning, transforming, analyzing and visualizing it have increased exponentially since the turn of the millennium. With a few clicks students can compare word frequencies in books over time or construct elaborate size-weighted wordclouds—tasks that would have taken scholars weeks if not months of effort in the past.

This course introduces students to those tools and the principles behind their use in the context of applications in political science. It marries the substance of political theory to the methodologies of data visualization and exploratory data analysis. It neither requires nor imparts any statistical background: it is designed to serve either as a standalone course or as a gateway to Political Science 4782, Research Methods in Political Science.

Requirements

Students will attend all lectures. There will not be a traditional midterm or final exam. Rather, occasional short assignments will make up 60% of the grade, and the remaining 40% will come from a final project in which the student finds a dataset in his or her area of interest that is not already used in the course, analyzes it to assess the structure of the data, and works through the most appropriate, succinct, and informative summaries and visualizations.

Books

Three books are required for the class. There is no coursepack. The books are:

- Hrishi, Mittal. *R Graph Cookbook* (Kindle edition; Packt Publishing)

Much of the course material will be found on the internet and will be referenced in the individual weeks’ lectures.

GE Goals and Expected Learning Outcomes

Goals

Students develop skills in quantitative literacy and logical reasoning, including the ability to identify valid arguments, and use mathematical models.
Expected Learning Outcomes

*Basic Computation*—Students demonstrate computational skills and familiarity with algebra and geometry, and apply these skills to practical problems.

*Mathematical or Logical Analysis*—Students comprehend mathematical concepts and methods adequate to construct valid arguments, understand inductive and deductive reasoning, and increase their general problem solving skills.

Academic Misconduct

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term “academic misconduct” includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3355-5-487). For additional information, see the Code of Student Conduct (http://studentaffairs.osu.edu/resource_csc.asp).

All students believe that they know how not to plagiarize. Many of them are wrong. Every year, many of them find that out the hard way. Don’t be one of them.

The short version is that passing off another person’s work or ideas as your own is plagiarism. That includes the unacknowledged word-for-word use or paraphrasing of another person’s work or ideas. It is not enough, for example, simply to copy and paste a passage and then cite the source at the end. If the passage is taken word-for-word, it must be in quotes as well to indicate that fact.

There is an excellent video at http://hdl.handle.net/1811/46848, if you have any doubts. You should be crystal clear, as the University’s policies exist to ensure fairness, and violators of University regulations on academic integrity will be dealt with severely.

Disability Services

Students with disabilities that have been certified by the Office for Disability Services will be appropriately accommodated, and should inform the instructor as soon as possible of their needs. The Office for Disability Services is located in 150 Pomerene Hall, 1760 Neil Avenue; telephone 292-3307, TDD 292-0901; http://www.ods.ohio-state.edu/.
Lecture Schedule

I: Principles

Lecture 1: **Introduction.** Nature of the course, its relationship to other courses and students' plans of study; how and why data visualization can be useful.

Lecture 2: **A Few of Our Favorite Things.** Exemplary representations of data from our field and others, with a discussion of what makes them so compelling. (Tufte, ch. 1; Yau, pp. 2–12)

Lecture 3: **DataViz 911.** Ideas that seemed good at the time, why they really aren’t, how we can fix them, and what general principles they imply. (Tufte, ch. 2; Yau, 13–20)

Lecture 4: **Truth.** Data visualization is inherently a reduction of information. The main imperative, in reducing information, is integrity: representing the data as honestly as possible. We look at some good cases and some not-so-good cases. (Tufte, ch. 3)

Lecture 5: **Beauty.** The difference between a fairly good data visualization and an amazing one often lies in the application of a few straightforward graphical principles. (Tufte, chs. 4–6)

II: Tools

Lecture 6: **All-In-One Tools.** Open-source solutions that require little or no additional information or effort to produce compelling results.

Lecture 7: **Big, Clean Data.** If we want to go beyond pre-packaged data, we need to be able to find what we're looking for and make sure it contains few if any errors. (Yau, pp. 21–43)

Lecture 8: **Exploring User-Supplied Data.** These versatile tools allow you to upload, explore, and visualize your own data. (Yau, pp. 54–62)

Lecture 9–11: **R.** A powerful (and free) general statistical package, R is capable of producing a wide range of graphs. In these lectures we demonstrate how to download and install the package, add libraries, import data, and create visualizations. (Yau, pp. 71–75.)
Lecture 12: Time. Different ways of examining variables over time. (Yau, ch. 4)

Lecture 13: Space. Less common, but more eye-catching, are ways of looking for patterns across space, especially in maps. (Yau, pp. 80–89)

Lecture 14: Proportions. Data on the parts of a whole, either on their own or across time, can often tell interesting stories as well. (Yau, ch. 5)

Lecture 15: Relationships. Learn to compare the values of multiple variables.

Lecture 16: Example—OkCupid. Bring together the tools used so far to analyze data from the popular dating website OkCupid.

Lectures 17–18: Research Design. Connecting data to theories and hypotheses about how the world works.

III: Practice

Lectures 19–28: Workshops. Exploration of a variety of questions from across the subfields in political science, utilizing the tools and principles explored in previous lectures.

Lecture 29: How to Lie Without Statistics. What are the limits of inference via data visualization? What more can data tell us that we can’t access visually or intuitively? Worse, how can apparently obvious visual patterns mislead us? In short, why do people ever take statistics courses?