PSC 3780: Data Visualization and Data Literacy

The Ohio State University

Syllabus: Fall 2017

Instructor:	Drew Rosenberg	Class location:	Hagerty Hall 186
Office:	2120 Derby Hall	Class time:	TBD
Email:	*rosenberg.1108@osu.edu*	Office hours:	TBD

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 a more advanced data-analytics class.

GE Goals and Expected Learning Outcomes

Goals

Students develop skills in drawing conclusions and critically evaluating results based on data.

Achieving Expected Learning Outcomes

Students understand basic concepts of statistics and probability, comprehend methods needed to analyze and critically evaluate statistical arguments, and recognize the importance of statistical ideas.

The course will fulfill the General Education (GE) requirement in Data Analysis by helping students develop skills in drawing conclusions and critically evaluating arguments based on data. It will introduce students to basic concepts in statistics and probability, including sampling, data distributions, and the Central Limit Theorem, and it will teach students how to use iterated simulation and re-sampling (i.e., Monte Carlo simulation and bootstrapping) to obtain estimates of unknown probabilistic outcomes and to assign measures of accuracy to sample estimates. It will tie these elements together with the logic of research design in order to give students the ability to evaluate statistical arguments, and it will show them examples of how to do so using data on such topics as American partisanship and elections, drone strikes, and international conflict.

Assessment of Expected Learning Outcomes

The effectiveness of this course in achieving the expected learning outcomes outlined above will be determined in three ways.

1. A problem set will be assigned in class every week in weeks 2–13 to help students internalize the material delivered in the online lectures. The problem sets will be graded, students' mastery of the relevant skills will be assessed, and they will be informed of problem areas, if any.
2. During the last two weeks of class, students will work on an independent project of their choosing, one that involves answering questions using the skills they have acquired. This final assignment will count for 40% of the course grade.
3. Students will be urged to fill in the online Student Evaluation of Instruction (SEI) reports to assess the amount that they have learned in this course relative to others.

Required text and materials

Texts

Three books are required for the class. All are available either as a Kindle book, which can be read on Kindle tablet or PC apps, or as a PDF.

[1] Teetor, Paul. R Cookbook. New York: O'Reilly. (PDF, also available for Kindle) (T)

[2] Yau, Nathan. 2011. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics.*. New York: Wiley. (Y)

[3] Carsey, Thomas M., and Jeffrey J. Harden. 2014. *Monte Carlo Simulation and Resampling Methods for Social Science*. New York: O'Reilly. (Kindle edition, rentable) (CH)

Materials

Video Lectures

The lectures are online and can be accessed via iTunes U. or at http://tinyurl.com/iTunesU-dataviz. You should listen to and view the lectures prior to class.

The lectures are recorded in Quicktime format, which should be viewable on nearly any computer or mobile device. One advantage to the format is that, on most devices, lectures can be sped up to 1.4x to 1.5x; beyond about 2x, most of them become unintelligible. Students are advised to take advantage of this feature, as humans can generally understand speech at a higher rate than they can produce it. Be advised, though, that higher speeds generally require more focused attention, as important details are easier to miss. It's also possible to rewind and slow the recording back down if a particular section moves quickly or is difficult to understand.

Also, note that this course was originally developed as a variant of Political Science 4781, and most of the lectures were recorded before it had its new number, so don't be surprised to hear the wrong course number at the beginning of each lecture.

Software

In this course, we will use the statistical software called R (see `https://www.r-project.org`). We will also use RStudio (see `https://www.rstudio.com`), a text editor that will make it easy for you to interact with R. Both R and RStudio are free, and you should download each on your own computer (Mac or PC). Versions will also be available on classroom computers.

What is R and why use it?

- Widely-used in academia and industries
- Open-source and free
- Power and flexibility
- Graphical capabilities
- Learning R = learning basic programming

The *New York Times* described R as

> a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca [...] whether being used to set ad prices, find new drugs more quickly or fine-tune financial models. Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it. [...] "The great beauty of R is that you can modify it to do all sorts of things," said Hal Varian, chief economist

at Google. "And you have a lot of prepackaged stuff that's already available, so you're standing on the shoulders of giants."

Resources for Using R

Students invariably find learning R to be the most challenging part of this course. It is also, in the long run, the most rewarding, in that it's a valuable and marketable skill that also prepares you for more advanced classes. The learning curve is very steep, however, and it is best to be prepared. Toward that end, students should take advantage of the following:

- *Statistical consulting at PRISM.* The Department of Political Science offers help through the Program In Statistics and Methodology (PRISM). You may contact the PRISM Fellows to setup a brief appointment for assistance using the following email: ps-prism@lists.osu.edu. You can also email the Junior PRISM Fellow, Daniel Kent, directly:

  kent.249@osu.edu

  Adam can arrange to meet with you in person, or answer questions over email.

- *Google is your friend.* Simply google your coding question, such as "how do I make a bar chart in R," and you are bound to find many useful resources on the web.

- *Online tutorials and books.* You should download Shidan Murphy's excellent R Instructor app for iPhone or Android and bookmark, or just download and print out, Tom Short's fantastic R Reference Card. You may also wish to consider Paul Teetor's *R Cookbook* from O'Reilly. You might also consider the following book: Hrishi, Mittal. *R Graph Cookbook* (Kindle edition; Packt Publishing)

Course requirements

There will not be a traditional midterm or final exam. Rather, weekly 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. Students will be given the last 2-3 weeks of recitation sections to work in-class on final projects, with the Professor and the TA present to assist. The final course grade will be determined based on the following breakdown and grading scale:

				A	93–100%	A−	90–92%
Problem Sets	60%	B+	87–89%	B	84–86%	B−	80–83%
Final Project	40%	C+	77–79%	C	74–76%	C−	70–73%
		D+	67–69%	D	64–66%	D−	60–63%
		E	< 60%				

Problem Sets

Weekly in-class assignments will be posted prior to recitation section on Monday and will be due the following Friday by 15:00 (5:00 p.m). Assignments can be completed in groups of up to three people. The grade assigned to a group project will be assigned to each individual in the group. The final project will be based on many of the skills obtained in the group projects, so there is little sense in joining a group and not being an active contributor.

Final Project

The final project is designed to be the foundation of an independent research project that will fulfill the requirements for a senior thesis. Students who are eligible to graduate with research distinction who are interested in writing a senior thesis should discuss doing so with their advisor. Applications are due early in the fall semester. Funding for thesis research is available from the University. Applications for the Undergraduate Research Scholarship (URS) and the International Research Grant for undergraduates in the Arts and Sciences are available at http://aschonors.osu.edu/opportunities/scholarships/undergrad.

The deadline for applications and project advisor recommendations is in mid-September, and selections will be made by mid-November. Applicants will compete for approximately 50 scholarships awarded in amounts ranging from $500 to $6,000. University regulations require that the funds be used to meet tuition and fee expenses, with any remaining amount available directly to recipients (provided they have not already exceeded the maximum allowable amount for financial aid).

Policies and procedures

Communication and logistics

Outside of class, the primary method of communication will be email. *Your University issued email will be used, so please be sure that you have access to that account and that you check it regularly.* I will try to answer any emails within 24 hours during the week, and 48 hours over the weekend. Thus, you should always prepare to write to me well in advance if you have questions about the course. I will make it a priority to respond as quickly as possible to emergencies and other extreme issues. As always, formal communication models are expected. You do not need to call me "Mr. Rosenberg;" however, I expect you to

use respectful, professional language, as well as proper grammar, spelling, and syntax. In addition, I hold two hours of office hours per week, but you may arrange a meeting outside of those hours if you are unavailable during this time. Please make use of office hours, as that is the time I allocate to be 100% available to you. If you have any questions or are having difficulty completing course requirements, please come see me as soon as possible.

Accommodations

I encourage you to begin assignments well ahead of time, as I am aware of the heavy workload you will be facing with other courses. Please be aware that I will not make any concessions regarding workload (such as dispensing when you miss class or excusing you from having done assignments). You are responsible for distributing your time according to the time you have every week. I am, however, well aware of the potential negative effect of exogenous factors on your learning or the possibility that you may be facing a tough time or an illness. I will require written confirmation from the counsellor or your physician to engage in dispensation and accommodation, but I will be more than happy to help in any way I can, within reasonable limits. Specific policies for assignments and the final project are below.

Assignment dispensation policy

Assignments must be submitted *on the day they are due (by 17:00—5:00pm).* If a student is unable to complete an assignment, they will be allowed to turn it in late only if the absence is due to a *documented* medical, family, or similar serious emergency, observance of religious holy days (which requires written notification to the instructor at least 14 days prior to the requested absence date), or properly documented University-sponsored planned activities. *Incomplete assignments in all other cases will result in a score of zero.* If you become aware that you will not be able to complete an assignment or final project ahead of time, please contact the instructor and seek permission for an extension as soon as possible.

Grade disputes

Grade disputes will be considered only if they adhere to this policy. Grade disputes must be made in writing (TYPED!). You must wait at least 2 full days after you receive your grade to submit a grade dispute ("cooling off period"); you may wait no more than 2 weeks after you receive your grade to submit a dispute. Your written dispute must contain a documented logic for why you believe your answer for each disputed item was incorrectly marked—you must cite specific passages in the texts and/or lectures and explain why you thought they applied to the item in question. The instructor will then review your dispute and issue a decision within one week. Failure to comply with this procedure will result in forfeiture of your ability to dispute your grade.

Grade adjustment policy

I do not tolerate emails asking me to round up grades or "find points" that otherwise do not exist. When you ask me to do either of these things, particularly at the end of the term, you are putting me in a position where I feel pressure to treat you differently from your colleagues. The policies above are designed to eliminate ambiguity in this regard. In addition, I have two policies on rounding: 1) I do not round on individual assignments, and 2) I round all final grades to the *nearest tenth*. For example, if you finish the term with a 79.88% in the course, I will round the grade to 79.9%. This policy is not an attempt to be mean. On the contrary, my goal is to hold all students in equal standing.

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 3335–5–487). For additional information, see the Code of Student Conduct (http://studentaffairs.osu.edu/resource_csc.asp).

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

Date	Subject	Readings	Assignments (Fri.)
August 28	Class Intro		
September 4	Labor Day—No Class		
September 18	Lectures 1 & 2	Y, ch. 2	PS 1
September 11	Lectures 3 & 4	Y, ch. 1 (to "Design")	
September 25	Lectures 5 & 6	Y, rest ch. 1 & Y, ch. 3 (to "Illustration")	PS 2
October 2	Lectures 7 & 8	T, chs. 1–4.	
October 19	Lectures 9 & 10	T, chs. 5–7.	PS 3
October 16	Lectures 11 & 12	Y, ch. 4 & T, chs. 8–10.	
October 23	Lectures 13 & 14	Y, pp. 80–89.	PS 4
October 30	Lectures 15 & 16	Y, ch. 5.	
November 6	Lectures 17 & 18		PS 5
November 13	Lectures 19 & 20	CH, 4.1–4.4.3.	
November 20	Lecture 21	CH, 8.1–8.2, 8.4.	PS 6
November 27	In-class work on projects		
December 4	In-class work on projects		
December 8			**Final Projects Due**

Description of Lectures

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: Data are All Around Us! Terminology, sources, and advice on what to do if you can't find a preexisting dataset.

Lecture 3: Data Tools, part I. Open-source solutions that require little or no additional information or effort to produce compelling results.

Lecture 4: The Good. Exemplary representations of data from our field and others, with a discussion of what makes them so compelling.

Lecture 5: The Bad and The Ugly. Ideas that seemed good at the time, why they really aren't, how we can fix them, and what general principles they imply.

Lecture 6: Data Tools, part II. These versatile tools allow you to upload, explore, and visualize your own data.

Lecture 7: 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.

Lecture 8: Introduction to `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.

Lecture 9: Beauty. The difference between a fairly good data visualization and an amazing one often lies in the application of a few straightforward graphical principles.

Lecture 10: `R`, part II. Data structures and data transformations.

Lecture 11: Time. Different ways of examining variables over time.

Lecture 12: Plots in `R`. Scatterplots, line graphs, boxplots, dot charts, and more.

Lecture 13: APIs. The application program interface: a wonderful way for people to make their data available to you.

Lecture 14: Maps. Less common, but more eye-catching, are ways of looking for patterns across space, especially in maps.

Lecture 15: Careers in Data I. Aaron Schwartz describes his work at Community Research Partners.

Lecture 16: Space and Time. Representing relationships and trends over both space and time.

Lecture 17: Research Design. Designs for causal inference using observational data, including natural experiments, regression discontinuity, and differences in differences.

Lecture 18: Careers in Data II. Jill Riepenhoff and Mike Wagner discuss their work at the Columbus *Dispatch*.

Lecture 19: Simulated Worlds Using simulation to predict the outcome of the 2012 Presidential election.

Lecture 20: Workshop—Partisanship. Is American politics more partisan than it was 50 years ago? If so, why?

Lecture 21: Simulations and Inference Monte Carlo techniques for understanding uncertainty.