POL 332H1-S: Using Data to Understand Politics and Society
Fall 2025

Prerequisites: POL222H1 or equivalent

Lecture: Wednesday, 9–11 am
Instructor: Professor Olga Chyzh, olga.chyzh@utoronto.ca
Office Hours: Wednesday, 11–12 pm

Overview and Objectives

Unlike data we obtain from experiments, observational data rarely come nicely pre-packaged for hypothesis testing, and especially less so, for causal inference. Yet some of the most central political processes and outcomes, such as terrorism, repression, and other forms of political violence, are often not compatible with experimental research. For the study of many such processes, observational data are all we will ever have access to. Students will learn the research tools and methods for causal inference with observational data, the assumptions behind these tools, and what common pitfalls to watch out for when working with non-experimental data.

The course covers the principles of panel and temporal design, difference-in-difference design, linear regression, regression discontinuity design, as well as text and network analysis, spatial statistics, and time-series analysis. This course draws from topics in the study of international relations, economics, political behavior, and statistics to offer a diverse set of tools for processing and analyzing different types of data. Applications include war and conflict, terrorism, international trade, social media, elections, and representation.

Learning Outcomes

This course is designed as a series of weekly modules that build upon each other. Each module covers one or more state-of-the-art approaches to statistical data analysis. For each topic covered, the objectives are that students will:

- Learn the general mechanics of the specific method;

- Formulate theories and derive hypotheses testable using this method;

- Apply the method to extract/analyze real-world political and social data.

Course Materials

Materials for the course are posted on Quercus as well as the course website, `www.pol332.netlify.app`.

Software

R (latest version) `https://www.r-project.org/`
RStudio (latest version) `https://rstudio.com/products/rstudio/download/`

Textbook

Please complete all assigned readings prior to class.

- Kosuke Imai. Quantitative Social Science: An Introduction. Princeton: Princeton University Press, 2017.

R Tutorials

For each method covered we will run through applications in RStudio. Students will follow along with these tutorials and turn in their R script after each class. The tutorials will contain required and optional components. Student submissions will be graded based on the completion of the required component only.

Grading Scale

Students will demonstrate their mastery of the course materials by completing in-class tutorials due on Quercus by the end of each lab session (10% of the grade), 3 assignments (10%, 20%, and 50%). The remaining 10% will come from participation. Late assignments are penalized 10% per day for up to 7 days (including weekends), after which the assignment earns a grade of 0. Students who are experiencing extenuating circumstances that may prevent them from completing an assignment should contact the instructor as soon as possible. The final grade will be calculated using University of Toronto grading scheme.

Course Policies

Student Responsibilities in the Learning Process: Students are expected to complete any assigned readings prior to completing that topic's assessment. Students are also expected to complete all assessments on time. This means accessing the materials with sufficient time to complete assessments prior to deadlines. In the event that a student has questions concerning the material, they should formulate specific questions to ask the professor via office hours or email with sufficient time for a response prior to assessment deadlines (i.e. emailed questions should be sent at least 24 hours prior to a deadline, excluding weekends).

Classroom Conduct: Students are expected to participate in class in a thoughtful and respectful manner while in the pursuit of knowledge accumulation. Generally, this means engaging with one another's ideas and treating others as you would like to be treated as well as *not* treating others how you would *not* like to be treated. Please

see university policies on freedom of speech (`https://governingcouncil.utoronto.ca/secretariat/policies/freedom-speech-statement-may-28-1992`) and discrimination and harassment (`https://governingcouncil.utoronto.ca/secretariat/policies/harassment-statement-prohibited-discrimination-and-discriminatory-harassment`).

Accommodations: Please discuss any special needs with the instructor start of the semester, for example, to request reasonable accommodations if an academic requirement conflicts with your religious practices and/or observances. Those seeking accommodations based on disabilities should complete the appropriate documentation with Student Life Programs and Services (`https://studentlife.utoronto.ca/department/accessibility-services/`).

Academic Misconduct: All acts of dishonesty in any work constitute academic misconduct. The Student Disciplinary Regulations (`https://governingcouncil.utoronto.ca/secretariat/policies/code-behaviour-academic-matters-july-1-2019`) will be followed in the event of academic misconduct.

A special note on plagiarism: plagiarism is the act of representing directly or indirectly another person's work as your own. It can involve presenting someone's speech, wholly or partially, as your; quoting without acknowledging the true source of the quoted material; copying and handing in another person's work with your name on it; and similar infractions. Even indirect quotations, paraphrasing, etc., can be considered plagiarism unless sources are properly cited.

Copyright: Course materials, including recorded lectures and slides, are the instructor's intellectual property covered by the Copyright Act, RSC 1985, c C-42. Course materials posted on Quercus or the class website may not be posted to other websites or media without the express permission of the instructor. Unauthorized reproduction, copying, or use of online recordings will constitute copyright infringement.

Course Schedule

I reserve the right to modify the syllabus to reflect the pace of the course.

Course Outline

Introduction to Causality (Week 1)

- Reading: Chapter 1 (optional background) and Chapter 2 Sections 2.1, 2.3, and 2.4.

Natural Experiments (Week 2)

- Cross-Sectional, Temporal, and Difference-in-Difference Designs

- Reading: Chapter 2 Sections 2.5, 2.6 (optional review).

Measurement Bias (Week 3)

- Desirability Bias, Non-response Bias, Probability Sampling, List Experiments

- Reading: Chapter 3 Sections 3.1, 3.4.2.

- Assignment 1 is due on Friday, Sep. 19, at 5 pm.

Linear Regression (Week 4)

- Chapter 4 Sections 4.2 (optional review), 4.3.1, 4.3.2

Regression Discontinuity Design (Week 5)

- Chapter 4 Sections 4.3.4

Research Design (Week 6)

- Barbara Geddes. How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political analysis*, 2:131–150, 1990;

- Jason Seawright and John Gerring. Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political research quarterly*, 61(2):294–308, 2008;

- Scott L. Minkoff. A guide to developing and writing research papers in political science, 2016;

- Assignment 2 is due on Friday, Oct. 10, at 5 pm.

Text Analysis (Week 7)

- Jodie Archer and Matthew L. Jockers. *The Bestseller Code: Anatomy of the Blockbuster Novel.* St. Martin's Publishing Group, 2016. Chapter 4.

Web-scraping (Week 8)

Big Data (Week 9)

- Danah Boyd and Kate Crawford. Six provocations for big data. a decade in internet time. Symposium on the Dynamics of the Internet and Society., 2011.

- David Lazer et al. The parable of Google flu: Traps in big data analysis. *Science*, 343:1203–1205, 2014.

- Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186, 2017.

- Jevin D. West. How to improve the use of metrics: Learn from game theory. *Nature*, 465(17):870–872, 2010.

Network Analysis (Week 10)

- Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):l360–1980, 1973, **Trigger Warning:** This article uses racial and gendered language common at the time of its writing.

- John F. Padgett and Christopher K. Ansell. Robust action and the rise of the Medici, 1400-1434. *American Journal of Sociology*, 98(6):1259–1319, 1993.

Spatial Analysis: Maps (Week 11)

Instrumental Variables (Week 12)

- Matthew Potoski and R Urbatsch. Entertainment and the opportunity cost of civic participation: Monday night football game quality suppresses turnout in US elections. *The Journal of Politics*, 79(2):424–438, 2017.

- Emily Hencken Ritter and Courtenay R Conrad. Preventing and responding to dissent: The observational challenges of explaining strategic repression. *American Political Science Review*, 110(1):85–99, 2016.

- Final research paper is due on Monday, December 1 at 5 pm.