

POL 478H1-F: Data Management and Visualization for Political Scientists

FALL 2020

PREREQUISITES: POL232H1/POL242Y1/POL322H1/equivalent

Lecture: Monday, 10am –12 pm, online–synchronous

Instructor: Professor Olga Chyzh, olga.chyzh@utoronto.ca

Office Hours: by appointment, <https://chyzh.youcanbook.me/>

Overview and Objectives

Real data, such as election results, location and type of terrorist attacks, or international trade flows, rarely come in nice, convenient to manage spreadsheets. Most data are not stored in spreadsheets at all. The task of collecting and cleaning the data is frequently upon the data analyst. Time-consuming and challenging as it is, this task need not be tedious and frustrating.

This course introduces students to the fundamental computational tools for cleaning, managing, storing, analyzing, and presenting political data. The material will enable students to become proficient enough to actively implement the methods and tools in their own research.

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 analysis of network data. For each model covered, the objectives are that students will:

- Learn how to deal with complex, messy, real data
- Use graphics to explore, understand, and present data
- Gain familiarity with basic data collection, storage and management
- Reshape data into the most convenient form for analysis or reporting
- Automate cleaning and analysis in R.

Course Website

Materials for the course are posted on the course website pol478.netlify.edu and Quercus. All course videos will be posted on Quercus (allow up to 24 hours for processing).

Software

R <https://www.r-project.org/>

RStudio <https://rstudio.com/products/rstudio/download/>

Reading

The textbook that I recommend for the class is "ggplot2: Elegant Graphics for Data Analysis" by Hadley Wickham. Note: it's just a recommendation, not a requirement!

Homework

Students will have a weekly homework, due before each class (10 total). For each homework, students will apply the methods covered that week to a dataset made available by the instructor. Each student will submit their homework as an Rmd and an html file. Each homework will take 1–5 hours to complete.

Coding Sessions

For each method covered we will run through applications in R during class. Students are strongly encouraged to follow along during class and review/run through these examples after class. Students will be provided with data, but may also use their own datasets.

Grading Scale

Grades will be assigned based on performance on 10 homework assignments (each contributing 10% to the final grade). Each assignment is posted shortly after class and is due at the beginning of the following class. No late assignments are accepted. 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 the following grading scheme to the sum of homework grades.

A+	≥ 90
A	≥ 85
A-	≥ 80
B+	≥ 77
B	≥ 73
B-	≥ 70
C+	≥ 67
C	≥ 63
C-	≥ 60
D+	≥ 57
D	≥ 53
D-	≥ 50
F	< 50

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 treating 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

Week 1: Introduction to R, Rstudio, and RMarkdown

Week 2: Introduction to Data Visualization

Week 3: Visualization of Categorical Data

Week 4: Introduction to Data Management

Week 5: Advanced Data Management

Week 6: Loops and Functions

Week 7: Advanced Data Visualization

Week 8: Working with Dates and Time

Week 9: Visualizing Network Data

Week 10: Automating Data Collection

Week 11: Drawing Maps