

POLISCI 3325G - Data Science for Political Science

Spring 2023

Professor: Evelyne Brie

E-mail: ebrie@uwo.ca

Course Schedule:

Tuesday 12:30pm - 2:30pm

Spencer Engineering Building 2100

Office Hours:

Brie: Monday 3pm - 5pm (SSC 7233)

Mosannef (TA): Wednesday 1pm-3pm (SSC 7336)

Pereira (TA): Friday 10am-12pm (SSC 7328)

Course Description

The objective of this course is to provide undergraduate students with the necessary statistical tools to make inferences about politics. The ability to quickly and accurately find, collect, manage, and analyze data is now a fundamental skill for quantitative social science researchers. We will cover fundamentals of data analytics and visualization using the statistical programming language R. These topics will be discussed with an eye on applications to research questions in all subfields of political science. Leaving this course, students will be able to acquire, format, analyze, and visualize various types of data. This is an in-person course that is synchronous, and students are required to attend all classes during the semester.

Prerequisites

You must have completed 2325F - *Research Design in Political Science* to enroll in this class. Unless you have either the prerequisites for this course or written special permission from your Dean to enroll in it, you may be removed from this course.

Software Requirements

This course will be taught using R, a free object-oriented statistical programming language. We will also use RStudio, an interface which makes working with R substantially easier. You should download the latest version of R (<http://cran.r-project.org/>) and the latest version of RStudio (<https://www.rstudio.com/>) before our first class.

You will receive an invitation to a **Slack workspace** at the beginning of the semester. Feel free to share all your problem set-related, substantive or technical questions in the appropriate Slack channel (i.e. not via private messages). Note that questions must be written in such a way as

to not give away answers to the problem sets. Students are encouraged to answer each other's questions.

Evaluations

Evaluations will consist of four problems sets, two exams and attendance points. Please note that your final grade in this class might be curved.

Week	Date	Evaluation	Percent
3	24/01/2022	Problem Set 1	12.5%
5	07/02/2022	Problem Set 2	12.5%
6	14/02/2022	Midterm Exam	20%
10	14/03/2022	Problem Set 3	12.5%
12	28/03/2022	Problem Set 4	12.5%
13	04/04/2022	Final Exam	25%
2-13		Attendance	5%

1. Problem Sets (12.5% each – 50% in total)

You will complete 4 problem sets over the course of this class, each worth 12.5% of your final grade. Some of them will be difficult or ask you to expand on what we discussed in class. Please note that these challenges (and frustrations) are inherent to learning computational social science, and while we will work to minimize them, we cannot eliminate them.

Problem sets should be submitted **by the beginning of class** on their due date. The script used to solve the programming problems must be compiled as a PDF and submitted via OWL. An how-to guide is available on OWL (Resources → Problem Sets → SubmissionGuide.pdf). Not submitting your work in the appropriate PDF format will result in a 1 point penalty. Your problem sets will be evaluated on: the conciseness of your code, the targetedness of your output, whether you obtain the right answer and whether your code is commented.

You might find it helpful to form study groups to work on the problem sets with your peers. However, **each student must submit their own write-up**, and I strongly recommend that you attempt the problems on your own before getting together with your study group. This is the best way to prepare for the exams. One key point: do not share code in any fashion. You may discuss the problem sets conceptually, but the coding must be 100% your own. Please note that you must write on your copy the name of all other students you collaborated with.

2. Exams (45% in total)

There will be a midterm (20%) and a final exam (25%), which are worth 45% of your grade in total. Both the theoretical content and the practical programming content of the class will be

evaluated during these exams. Most of the questions will be variations or expansions of the material covered in the problem sets. TAs will hold an optional review session prior to each of the exams to discuss any questions you may have on the material. Exam answers will be evaluated on: the conciseness of your code, the targetedness of your output, whether you obtain the right answer and whether your code is commented.

Exams will be held in class. Your script should be compiled as a PDF and uploaded on OWL by the end of the evaluation. Not submitting your work in the appropriate PDF format will result in a 1 point penalty. You will have access to your computer during the exam. Accessing Google, personal notes and class material will be allowed, but any form of real-time communication (Slack, Discord, email, SMS, etc.) is prohibited and will result in disciplinary action.

3. Attendance (5% in total)

Class attendance is mandatory. Each student is allowed one unexcused absence, beyond which each additional unexcused absence will be penalized by 1%, up to a total of 5%. We will hold three optional tutorials over the course of the semester (on week 2, week 5 and week 12). Attending these tutorials is encouraged, but not mandatory.

Class Material

1. Slides, Lab Material, Problem Sets Material

Presentation slides, datasets and material for the problem sets will be made available on OWL. RMarkdown write-ups for our laboratories will be uploaded on the following web page: <https://www.evelynbrie.com/intro-to-data-science/>.

2. Readings

The mandatory coursebook is the following:

- Imai, K. and Webb Williams, N. *Quantitative Social Science: An Introduction in Tidyverse*. 2022. Princeton University Press

Please note that a variety of high-quality material relative to programming with R is freely available online, among others here:

- Wickham H. & Grolemund G. *R for Data Science*. 2017. O'Reilly. Available here: <https://r4ds.had.co.nz/>

I also recommend consulting the following book, which discusses how to conduct empirical research in political science using R:

- Berdahl R. *Explorations: Conducting Empirical Research in Canadian Political Science, 4th Edition*. 2021. Oxford University Press. Available here: https://learninglink.oup.com/access/berdahl4e-student-resources#tag_lab-manual

Late Problem Sets and Grading

All students will be granted one 24-hour automatic problem set extension during the semester, no questions asked. You do not need to inform us in advance when using your extension, as we will be keeping track of submission times. Beyond that, late problem sets will be penalized by 10 percentage points for each 24 hours that they are late. The material in this course is cumulative, so lateness on a specific assignment will have negative effects on your ability to learn the material.

If illness prevents you from coming to class or from turning in assignments on time, the illness must be documented according to Western's Policy on Accommodation for Illness (available [here](#)). To obtain any accommodation in the context of this class, you must email appropriate documentation to the Academic Counselling office at arts@uwo.ca. The Accommodation request is emailed to Professors shortly after, and it is the student's responsibility to follow up with professors and make the appropriate arrangements if approved.

You are not allowed to re-write any exams or problem sets once they have been handed in for marking, unless due to exceptional circumstances as assessed by the Dean's office. In the event that you do not submit some evaluations, you may not have your mark reweighted as to ignore your unsubmitted work unless you have written permission from the Academic Counsellor in your Faculty.

Office Hours and Tutorials

I will hold office hours to address any follow-up questions, to suggest further readings, to discuss assignments and to cover any other topics relevant to the class. These office hours are held in-person at **Social Science Centre 7233** or **via Zoom on Monday from 15:00 to 17:00** (you can make an appointment here to avoid waiting in line if there is a high turnout: https://calendly.com/e_brie). The teaching assistants will also hold office hours at the times and locations indicated on the first page of this syllabus.

Statement on Academic Offences

Scholastic offences are taken seriously and students are encouraged to read the appropriate policies, specifically, the definition of what constitutes a Scholastic Offence and the associated penalties [here](#).

Computer-marked tests and exams may be subject to submission for similarity review by softwares that will check for unusual coincidences in answer patterns. All assignments may be subject to submission for textual similarity review to the commercial plagiarism-detection software under license to the University for the detection of plagiarism. All assignments submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com.

Accommodations and Accessibility

When a course requirement conflicts with a religious holiday that requires an absence from the University or prohibits certain activities, students should request accommodation for their absence in writing at least two weeks prior to the holiday to the course instructor and/or the Academic Counselling office of their Faculty of Registration. Please consult University's list of recognized religious holidays (updated annually) [here](#).

Students with disabilities are encouraged to contact Accessible Education, which provides recommendations for accommodation based on medical documentation or psychological and cognitive testing. The policy on Academic Accommodation for Students with Disabilities can be found [here](#).

Use of Cell Phones and Computers

Students are required to refrain from using cell phones during the class. At no time are students to use computers for any non-class related purpose.

Schedule and learning goals

The learning goals below should be viewed as the key concepts you should grasp after each class. Please note that this schedule is tentative and subject to change.

Week 1

Tuesday, January 10th

- Concepts
 - Introduction to Data Science
 - Inputting and Managing Data
- Readings
 - Imai & Webb Williams: Chapter 1

Week 2

Tuesday, January 17th

- Concepts
 - Basic Descriptive Statistics
 - Subsetting, Merging and Manipulating Data

- Readings
 - Imai & Webb Williams: Chapter 2 (sections 2.1 and 2.2)

Optional tutorial: setting up R and compiling PDFs

- Wednesday January 18th, 9:30-10:30, SEB 2100
- Thursday January 19th, 12:30-1:30, SEB 2100

Week 3

Tuesday, January 24th**Submit Problem Set 1 (12.5%)**

- Concept
 - Causal Inference
- Readings
 - Imai & Webb Williams: Chapter 2 (sections 2.3 to 2.8)
 - King, Keohane and Verba. *Designing Social Inquiry*. 1994 (Chapter 3)

Week 4

Tuesday, January 31st

- Concepts
 - Univariate and Multivariate Data Visualization
 - Cluster Analysis
- Readings
 - Imai & Webb Williams: Chapter 3
 - Wickham. *ggplot2: Elegant Graphics for Data Analysis*. 2016: Chapter 2

Week 5

Tuesday, February 7th**Submit Problem Set 2 (12.5%)**

- Concepts
 - Loops
 - Linear Regression

- Readings
 - Imai & Webb Williams: Chapter 4 (sections 4.1 to 4.3)

Optional tutorial: midterm review session

- Wednesday February 8th, 9:30-10:30, SEB 2100
- Thursday February 9th, 12:30-1:30, SEB 2100

Week 6

Tuesday, February 14th

Midterm Exam (20%)

Week 7: Spring Break

Week 8

Tuesday, February 28th

- Concepts
 - Randomized Experiments
 - Regression Discontinuity Designs
- Readings
 - Imai & Webb Williams: Chapter 4 (sections 4.4 to 4.6)
 - Shang, Y. *Making Sense of Data with R*. 2022. (Sections 8 and 9 - available here: <https://bookdown.org/yshang/book/>)

Week 9

Tuesday, March 7th

- Concepts
 - Web Scraping
 - Textual Analysis
- Readings
 - Imai & Webb Williams: Chapter 5 (sections 5.1 and 5.5.1)

Week 10

Tuesday, March 14th

Submit Problem Set 3 (12.5%)

- Concepts
 - Sampling from Probability Distributions
 - Law of Large Numbers
 - Central Limit Theorem
- Readings
 - Imai & Webb Williams: Chapter 6 (sections 6.1 and 6.2 to 6.4)

Week 11

Tuesday, March 20th

- Concepts
 - Estimation
 - Introduction to p-values
- Readings
 - Imai & Webb Williams: Chapter 7 (sections 7.1 and 7.2)

Week 12

Tuesday, March 28th

Submit Problem Set 4 (12.5%)

- Concepts
 - Visualizing Spatial Data
 - Spatial Regression Models
- Readings
 - Imai & Webb Williams: Chapter 5 (section 5.3)

Optional tutorial: final review session

- Wednesday March 29th, 9:30-10:30, SEB 2100
- Thursday March 30th, 12:30-1:30, SEB 2100

Tuesday, April 4th

Final Exam (25%)