# POLISCI 3325G Data Science for Political Science

Winter 2025 Tuesdays 1:30-3:30pm

Instructor: Shanaya Vanhooren Email: svanhoo5@uwo.ca Office Hours: Immediately after lecture or email to set up an appointment Teaching Assistants: Noah Vanderhoeven (nvande43@uwo.ca) & Jan Eckardt (jeckardt@uwo.ca)

> Version date: January 6, 2024. This syllabus is subject to change ahead of course start.

# **Course Description**

Social science research has been transformed by the increasing availability of both traditional forms of data (e.g. survey data) and non-traditional forms of data (e.g. social media data). As engaged citizens and consumers of political information, students benefit from knowing how to interpret quantitative measures of political activity. In addition, data analysis skills are essential for careers in advocacy, the public service, journalism, academia, and related fields. In this course, students will develop basic skills for collecting, formatting, wrangling, and describing data using the R software environment. Students do not require any background knowledge in computer science or statistics to succeed in this course. Instead, the 'hands-on' approach to learning will expose students to various real-world examples of political data in a way that is accessible and practical.

## 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.

## Learning Outcomes

Upon completion of this course:

- 1. Successful students will be critical consumers of quantitative data on political life. Successful students will recognize how quantitative data can advance and obscure understandings of a complex political world.
- 2. Successful students will be able to load, wrangle, and conduct basic analyses of datasets in the R software environment. They will do so effectively by drawing on knowledge of best practices in data visualization, summary, and coding learned throughout the course.
- 3. Successful students will be able to effectively communicate and interpret findings from exploratory data analysis for a wide variety of audiences.
- 4. Successful students will have a greater understanding of how aspects of gender, race, class, and ethnicity are measured and described (or, in some cases, fail to be) in quantitative research in political science.

## **Evaluations**

## Grade Distribution:

Evaluation	Percent	Due Date
Attendance	5%	Ongoing (see description)
Pop Quizzes	10%	Ongoing (see description)
Written Data Assignment 1	15%	March 6 @ $5:00$ pm
Midterm Test	25%	February 11 (in class)
Written Data Assignment 2	15%	March 27 @ 5:00pm
Final Exam	30%	TBD by registrars office

## 1. Attendance

- worth 5% of overall grade
- Attendance will be taken (on a pass/fail basis) for every lecture except weeks 1 and 6 (the midterm). Students must stay for the duration of the lecture in order to be awarded a pass.
- Why bother? This is an in-person course that is synchronous, and attending all classes throughout the term is essential for students to succeed in this course. In this course, you will learn about how social scientists handle quantitative data more generally AND how to code in R (which is like learning a whole new language.. eek!). This is a lot to learn on your own. Attending lectures will give you the opportunity to ask questions and seek support from the professor, TAs, and your peers in the class.
- Students who cannot attend a lecture because they are sick or face excruciating circumstances must seek accommodation through Academic Counselling. See the section titled "Late Policy, Academic Accommodations and Considerations" below for more information about missing a lecture and seeking consideration.

## 2. Pop Quizzes

- worth 10% of overall grade
- Throughout the term, I will randomly administer "pop quizzes" during lectures, which are five to ten multiple choice questions based on readings and concepts discussed in lecture. The quizzes will be administered on Brightspace. Only students who are present in the classroom at the time that the quiz becomes available will be able to complete the quiz. You will need a laptop in order to complete the quiz, please be sure to bring your laptop to class or inform the instructor ASAP if you do not have a laptop that you can bring to class.
- Why bother? Pop quizzes motivate students to stay on top of the readings and assess student understanding of course concepts, which the instructor can use to adjust their teaching style and fill in gaps in understanding.
- If a student cannot attend class for a legitimate reason and misses a pop quiz, they should seek consideration through Academic Counseling. See the section titled "Late Policy, Academic Accommodations and Considerations" below for more information about missing a pop quiz and seeking consideration.

## 3. Written Data Assignments (2)

- Two assignments worth 15% each, in total worth 30% of overall grade
- The first assignment will be due on March 6 @ 5:00pm. The second assignment will be due March 27 @ 5:00pm.
- Why bother? The assignments provide students the opportunity to test their R skills (e.g., open data in R, describe and/or visualize variables) and apply their understanding of theoretical considerations discussed in lectures (e.g., valid measurement of concepts etc.).
- Assignment instructions and a rubric will be posted at least two weeks prior to the assignment's due date.
- Students are expected to work independently on the assignment. Be sure to review the course policy on the use of Generative AI below. If you require an extension on the assignment, please refer to the "Late Policy, Academic Accommodations and Considerations" section of the syllabus below for specific instructions.

## 4. Midterm Test

- worth 25% of overall grade
- A midterm for the course will be held during regularly scheduled lecture on February 11, 2024. More information about the midterm structure will be discussed in weeks leading up to the test.
- Why bother? The midterm helps assess students understanding of course material. Studying for the midterm will help students remember important concepts that will be applied in written assignments.
- If you are unable to attend the midterm due to illness or some other extenuating circumstance, you are required to submit documentation to Academic Counselling. See the section below titled "Late Policy, Academic Accommodations and Considerations". If a student's request for academic consideration is accepted, the student will be asked to attend a make up exam. There will be only one make-up exam scheduled and having another class or tutorial scheduled during this time will not be accepted as a reason to miss it, unless the other class is holding an exam at that time. Be sure to communicate with the Professor if this is the case.
- If a student does not submit a request for academic consideration after missing the midterm, or their submission is denied, the student will not be able to write a make-up exam and will receive a zero on the midterm.

## 5. Final Exam

- worth 30% of overall grade
- A culminating exam will be scheduled by the Registrar's office during the exam period. Please note that the instructor does not have any control over the date and location of the exam.
- Why bother? The final exam is used to assess student understanding of course materials. Studying for the final exam will ensure that you walk away from this course with a clearer understanding of data science, and how to assess quantitative research and reports.

## Course Schedule

The weekly coverage might change depending on the progress of the class. However, you are required to keep up with assigned readings (both and R-related readings and resources). This means completing readings assigned for lecture before attending the lecture.

The following readings are referred to by the author's(s') last name(s) and date in the schedule below:

Llaudet, E., & Imai, K. (2022). Data analysis for social science: A friendly and practical introduction. Princeton University Press.

Brancati, D. (2018). Social scientific research. SAGE Publications.

Haan, M. & Godley, J. (2017). An introduction to statistics for Canadian social scientists. (Third ed.) Oxford University Press.

Berdahl, L., & Roy, J. (2021). Explorations: Conducting Empirical Research in Canadian Political Science (Fourth ed.). Oxford University Press.

#### Week 1 (Jan. 7): Introduction to the course

- Substantive Readings:
  - Casaburi, L., & Troiano, U. (2016). Ghost-House Busters: The Electoral Response to a Large Anti-Tax Evasion Program. The Quarterly Journal of Economics, 131(1), 273–314.
- R-related Readings and Resources: N/A

#### Week 2 (Jan. 14): What is data and where does it come from?

- Substantive Readings:
  - Brancati. (2018). Chapter 16, pgs. 231-240.
- *R*-related Readings and Resources:
  - Llaudet and Imai. (2022). Chpt 1, pgs. 1-2, 6-13.
  - Instructions on how to install R and RStudio for Mac, Windows, Linux.
  - MarinStatsLectures- R Tutorial 1.1 What is RStudio and Why Should You Download It?
  - Healy, Kieran. (2018). "Chapter 2: Get Started" from Data Visualization: A Practical Introduction.

#### Week 3 (Jan. 21): Datasets, variables, and observations: Examples from polisci

- Substantive Readings:
  - Haan & Godley. (2017). pp. 11-15. (starting at section titled "Data, Variables, and Observations)"
  - Schedler, A. (2012). "Judgment and Measurement in Political Science." Perspectives on Politics, 10(1), 21-36.
  - Beckwith, K. (2007). "Numbers and newness: The descriptive and substantive representation of women". Canadian Journal of Political Science, 40(1), 27–49.
- *R*-related Readings and Resources:
  - Wickham, H., Cetinkaya-Rundel, M. & Grolemund, G. (2023). "Tidy Data" in R for Data Science. O'Reilly Media.

# Week 4 (Jan. 28): From concepts to variables: Practical examples and measurement challenges

- Substantive Readings:
  - Berdahl & Roy. (2021). pp.55-66.
  - Bittner, A., & Goodyear-Grant, E. (2017). "Sex isn't Gender: Reforming Concepts and Measurements in the Study of Public Opinion." *Political Behavior*, 39(4), 1019–1041.
  - Bennett, D. S. (2006). "Exploring Operationalizations of Political Relevance." Conflict Management and Peace Science, 23(3), 245-261.
- *R*-related Readings and Resources:
  - Wickham, H. 2019. "Style Guide." In Advanced R. Chapman and Hall.
  - VIDEO: How To Knit In R Markdown in Word/PDF Format?! from Shainu on YouTube

#### Week 5 (Feb. 4): Exploring and describing a single variable

- Substantive Readings:
  - Bilodeau, A., Turgeon, L., & Karakoç, E. (2012). "Small worlds of diversity: Views toward immigration and racial minorities in Canadian provinces." *Canadian Journal of Political Science*, 45(3), 579–605. \*focus on pp. 579-591
  - Visualizing the data: Women's representation in society. UN Women
- *R*-related Readings and Resources:
  - Llaudet and Imai. (2022). pp.57-63 (from start of section titled "3.3.2 Frequency tables" to pg. 62)

Week 6 (Feb. 11): Midterm Exam No assigned readings this week. February 18 - reading week

#### Week 7 (Feb. 25): Visualizing a variable: Introduction to ggplot2

- Substantive Readings:
  - Haan & Godley. (2017). "Chapter 6: Measures of Central Tendency and Dispersion". pp.44-53.
  - Employment and Social Development Canada. (2023). "Blueprint for Transformation. The 2023 Report of the National Advisory Council on Poverty." focus on plots and tables, e.g. pp.26, 29, 32, 35
- *R*-related Readings and Resources:
  - Introduction to ggplot2 from Tidyverse
  - A simple introduction to ggplot2 from R for Ecology

# Week 8 (Mar. 4): Describing relationships between two variables & what is "big data"?

- Substantive Readings:
  - Llaudet and Imai. (2022). pp.62-65 (from start of section titled "3.4.2 Two-Way Frequency Tables")
  - Grimmer, J. (2014). "We are all social scientists now: How big data, machine learning, and causal inference work together." PS Political Science and Politics, 48(1), 80–83.
    \*focus on pp.80-81
- *R*-related Readings and Resources: Instructor will provide R code.

#### Week 9 (Mar. 11): Visualizing relationships between two variables

- Substantive Readings:
  - Haan & Godley. (2017). pp. 188-197.
- *R*-related Readings and Resources:
  - Healy, K. (2018). "1. Look at Data" In *Data Visualization: A Practical Introduction*. Princeton University Press.

#### Week 10 (Mar. 18): Relationships between more than two variables

- Substantive Readings:
  - Haan & Godley. (2017). pp. 198-210.
- *Resources:* 
  - How to Read a (Quantitative) Journal Article, from introsocsite: Introduction to Sociology

- Sevi, S. (2001). "A Guide to Interpreting Regression Tables"

#### Week 11 (Mar. 25): Text-as-data

- Substantive Readings:
  - Grimmer, J., & Stewart, B. M. (2013). "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis*, 21(3), 267–297.
  - Lawlor, A. (2015). "Framing Immigration in the Canadian and British News Media." Canadian Journal of Political Science, 48(2), 329–355.
- *R*-related Readings and Resources:
  - Quanteda Quick Start Guide
  - Benoit, K., Müller, S. & Obeng, A. "Example: textual data visualization"

## Week 12 (Apr. 1): Wrap up

- The importance of data analysis in public and private sector jobs:
  - Government of Canada. (2024). 2023–2026 Data Strategy for the Federal Public Service.
  - Li, V., Kwok, T., & Hamza, M. (2023). The Skills Algorithm: Digital Skills Demand Across Canada's Labour Market. *The Dais.*
- *R*-related Readings and Resources:

- N/A

## **Course Policies**

#### Software Requirements

- Students will be required to complete assignments using R (a free, open-source programming software). We will work in an application called RStudio, an interface that helps us look at data and keep track of our work. RStudio operates by running R in the background (both R and RStudio need to be installed on the computer being used to complete assignments).
- Since many of the examples discussed in class will be shown in R, it is best that students bring a laptop to lecture to be able to follow along in real time. We will also take time in the second half of most lectures to help students troubleshoot any issues encountered in R.
- It is recommended that students try to download and install R and RStudio before the second lecture, although there will be class time allotted to this. Download R here. Download RStudio here.

## Late Policy, Academic Accommodations and Considerations

#### Academic Accommodations:

Students who believe that they require ongoing academic accommodations due to a disability are encouraged to request accommodation through Accessible Education. More information can be found here: Accessible Education

## Late Work:

• Late assignments are unacceptable and will be penalized at a rate of five percent (5%) per day (including weekends). Students should get ahead of due dates by starting to work on assignments early. Assignments later than 10 days will not be accepted and will receive a grade of zero. If you require an extension due to an illness or other extenuating circumstance, please read the policy outlined below. I also recognize that sometimes life just "happens" and things get in the way of you completing your work. For this reason, I have also included a "Life Happens" clause (see below) to allow every student ONE opportunity throughout the term to receive a short extension on a written assignment (excludes lecture attendance and pop quizzes).

### Life Happens Clause for Written Assignments:

- Students are allowed a 3-day extension for ONE assignment in this course. This is the "Life Happens" clause (credit to Dr. Alexis Lerner for establishing this "clause"). In order to use the "Life Happens" clause, students must email their assigned TA with the subject line "Life Happens". In the body of the email, the student needs to note the assignment that they are using the clause for (e.g., Written Data Assignment 1) and the new deadline for the assignment. The student does **not** need to provide an explanation as to why they are utilizing the clause. TAs will respond within 48 hours to confirm they've received the email. Students are allowed only three days to complete the assignment without penalty.
- After three days since the original deadline passes, penalization occurs at 5% per day. The clause can only be used once during the term and it is **not** retroactive (meaning it cannot be applied after a deadline has passed or after a student submits a late assignment). The clause **CANNOT** be used for attendance, pop quizzes, the midterm or final exam.

#### Requests for extensions and missed lectures or exams:

- If you are sick or face other extenuating circumstances that prohibit you from submitting an assignment on time, attending class, or writing a test, the circumstance must be documented according to Western's Academic Consideration Policy. The process for granting consideration/accommodation for a range of issues is managed through Academic Counsellors in the student's home faculty. Students are required to use the Student Absence Portal to submit requests for Academic Consideration. More information about the portal and the University's Academic Consideration policy can be found here: Office of the Registrar Academic Consideration
- All requests for medical or compassionate extensions must go through Academic Counseling only. Please do not email medical documentation or other information to the Professor or TAs for the course. Accommodation requests received by Academic Counselling

are sent to professors shortly after, and it is the student's responsibility to follow up with professors and make the appropriate arrangements if approved.

• If you are unable to attend lecture, it is highly recommended that you find a student in the class who is willing to share their notes. The professor will not share their lecture notes.

### Policy specific to attendance and pop quizzes:

• If your request for Academic Consideration for missing a lecture (which contributes to your attendance grade) or a pop quiz is approved, the distribution of your grade across subsequent components of the evaluation will simply be re-weighted. For instance, a student who missed a lecture and received Academic Consideration for their absence will have their attendance grade re-weighted such that it is now a score out of 9 lectures, instead of 10 lectures.

#### Make up work and grade appeals

- There are no opportunities to "make up" any assignments where a student received a grade that they felt was not satisfactory.
- If you are concerned that your assignment was not graded fairly or correctly, you must wait 72 hours before contacting the appropriate TA/instructor that graded the paper to appeal your grade. To appeal your grade, you must email a 1-page written explanation to your TA stating why you think your assignment was improperly graded. The TA/instructor will respond to your appeal in writing. Should you be unhappy with the TA's reassessment of your paper, you can appeal to the course instructor (if your assignment was graded by the course instructor, they will pass the paper to a TA to re-grade). To do so, please email the course instructor your original 1-page explanation to the TA, the TA's response, a new 1-page written explanation detailing why a further appeal is necessary, and a clean copy of your paper. The instructor (or TA) will then re-grade the paper. If the grades are sufficiently similar, the original grade will hold.
- This policy is meant to promote fairness for students, as well as respect for the TAs and course instructor.

## Statement on the Use of Generative AI

- Students are expected to learn how to code in R through a process of trial and error. Running course materials through GenAI or using GenAI to complete a course assessment is considered unoriginal work, and therefore, plagiarism.
- It is not acceptable to continually consult generative AI tools when learning how to code and GenAI tools should not be used to produce a student's written work in any form. It can be destructive to use GenAI when learning to code, as it increases your dependence on the tool and decreases your ability to remember or comprehend the coding language. Furthermore, students are expected to rely mostly on R libraries and functions discussed in class and the instructor will provide sufficient sample code students can refer to when completing assignments.
- Consulting ChatGPT and other generative AI tools can be helpful when a student is feeling particularly "stuck" while de-bugging code after the student has tried several times on their

own to find a solution. Use of GenAI in this way is acceptable in this course, as long as it is minimally consulted for the purpose of debugging code. If a student uses Generative AI (e.g. ChatGPT), they must include an appendix at the end of the assignment stating how and why it was used in order to demonstrate that they are complying with course policy.

# **University Policies**

## Statement on Academic Offenses

- Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following Web site: Scholastic Discipline Policy (Undergrad)
- All required papers 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 papers 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.

## Statement on Use of Electronic Devices

• Students are allowed and encouraged to use their laptop during class time, however, they are expected to refrain from scrolling social media and looking at other non-class related content. Students should avoid using cellphones during lectures.

## Links to University Support Services

- Registrarial Services
- USC Services
- Students who are in emotional/mental distress should refer to: Mental Health@Western for a complete list of options about how to obtain help.

## Statement on Gender-Based Violence

• Western is committed to reducing incidents of gender-based and sexual violence (GBSV) and providing compassionate support to anyone who is going through or has gone through these traumatic events. If you are experiencing or have experienced GBSV (either recently or in the past), you will find information about support services for survivors, including emergency contacts at the following website here. To connect with a case manager or set up an appointment, please contact support@uwo.ca.