# POLSCI 9590A: Intro to Methodology
# Fall, 2024

**Instructor:**
Dave Armstrong
Department of Political Science
University of Western Ontario
e: dave.armstrong@uwo.ca

**Teaching Assistant:**
William Poirier
e: wpoirier@uwo.ca

## 1 Overview and Course Objectives

This course is meant to get students thinking about how data are used to make social, political and economic decisions and how we can answer questions we care about with data. The time when political scientists could happily exist in a statistics-free world has long since passed. We live in a world where many decisions are made using data. This has a number of interesting implications.

1. It is incumbent on those of us who wish to be savvy consumers to understand how data are being used to make decisions and how to discriminate good from bad uses of data.

2. We have to have some basic understanding of research design and probability to evaluate advances in health, medicine, politics, economics, etc... When looking at media reports of studies, how can we figure out whether the study was done well or poorly? Does the study generalize? What actionable intelligence really exists?

3. As political scientists, we answer questions about policy effectiveness, the efficacy of citizen interactions with government, etc... with data. But, we need to know and understand how.

Throughout the course, you will be asked to engage, and in some cases produce, quantitative work.

One of the interesting aspects of how stats courses (particularly those in the social sciences) have evolved is that they often continue to be perceived as "math classes." One of my goals here is to (mostly) disabuse you of this notion. This is largely not a math class. That is to say, we will be *doing* (in terms of paper and pencil) very little math. We will

be *looking at* mathematical notation from time to time. We use mathematical notation because it is precise - with terms and operators defined, we could all look at a statistical function and understand its main components and how they fit together. Mathematical notation gives us a common foundation for shared scientific communication. Similarly, mathematical notation is concise - it allows us to represent some quite abstract and complicated ideas in relatively little space.

Instead of thinking about this as a math class, this is a class that part psychology, part story-telling, part visual design and part data analysis (obviously). Statistics in the social sciences are often (as in this course) focused on applications rather than complicated mathematical derivations.

# 2 Course Text(s)

The required text for the course is:

Hahn, Michael and Jenny Godley (2017) An Introduction to Statistics for Canadian Social Scientists. *3rd ed.* Oxford University Press.

Other readings will be made available electronically through the course's OWL site.

# 3 Requirements

The course has a number of requirements, including attendance, reading, quizzes, and written work of different sorts. Here is a breakdown of the different elements:

**Participation** 10 points

There are quizzes in the short videos. In total, there are about 2 hours of videos to watch across the whole semester. They are designed to make sure that you're understanding the material. I will use your participation in the quizzes as your participation grade. This should be done independently, without collaboration with your colleagues.

**Cheat Sheet** 10 points

It is important to keep track of what we've been doing and to have some understanding of how the pieces fit together. One thing I've found useful as an R user are Posit's cheat sheets (for example, here are the ones for the `dplyr` package.) You will be making your own cheat sheet (I will provide a suitable template in powerpoint and keynote). Each week, you will turn in your cheat sheet integrating the previous week's material. Turning in the cheat sheet with new, relevant information on it will earn you a point. Failing to turn it in or turning it in without updating will earn you 0 points for that week. You can discuss this with your colleagues, but each student's cheat sheet should be completed independently.

**Homework** 40 points

There will be roughly one homework assignment per week. These will be entirely applied - I will ask you to answer a set of questions using data I provide and the tools we develop in class each week. Each homework will be marked on an integer scale from 0 to 4 where 0 = Missing, 1 = Inadequate, 2 = Adequate, 3 = Good, 4 = Excellent. You will submit your homework as an RMarkdown document that William and I will be able to knit. It will provide all relevant input and output along with whatever interpretation we require. Homework assignments can be discussed with your colleagues, but each student should do and submit the assignments independently.

**In-class Final Exam** 30 points

The final exam is a non-collaborative, handwritten final. I will show you output from R and ask you questions about it. There may also be a few questions that test your ability to read R code (i.e., know what the different functions we talked about will do). You may bring your cheat sheet to use during the in-class final. I will provide some instructions about the physical dimensions of the cheat sheet closer to the exam. This is a way for me to ensure that you actually know how to engage the material without simply searching on the internet or using generative AI.

**Take Home Final** 10 points

The second component of the final exam will ask you to reproduce the output that was featured on the take-home component. For the take home component, you can use whatever resources you have at your disposal, including books, notes, internet, generative AI, etc... You will submit the final as an RMarkdown file that William and I will knit. The take home final is also not collaborative - while you have a wide array of reousrces available, you should complete the final without collaborating or discussion with anyone else.

# 4 Absences, Late Assignments and Makeups

I will follow the guidelines for accommodations posted at the Office of the Registrar.

# 5 Generative AI

Large Language Models (aka LLMs or generative AI) are a game changer when it comes to coding. A short set of well-formed questions can often produce code (especially for R) that will generate the answer you want. This is a tool that is not going away and if you like, you should become familiar with how it works and how it can assist you in your work. You can use it to help with homework and the take home component of the final. Ultimately, you are responsible for knowing how to accomplish the tasks we do in class and what the code means. While it is helpful, the opportunity exists to rely too heavily on the technology which could leave you unable to do what you want if you don't

know how some of the fundamentals work. All this is to say, try to be a conscientious user of generative AI.

# 6 Academic Offenses

I will follow the University's policies on plaigarism and other academic misconduct. The Department of Political Science has a useful discussion of those policies you can view if you like.

# 7 Course Outline

The course will meet from 9:30 - 11 AM on Mondays for lecture/discussion and 1-2:30 PM on Wednesdays for lab. Attendance at both class periods each week is mandatory. On Mondays, I will address some theoretical material and then walk through some of the important concepts in R (we do, you watch or play along if you like). On Wednesdays, William will provide some exercises that will help you get you working more independently on these ideas (you do, we help). Finally, the homeworks will allow you to try these tools out independently.

The outline below gives the readings and other assignments that should be done before coming to class that day.

Many of the course readings are identified below. From time to time, I may assign other readings that will be posted to the course's OWL site.

**Week1**

- Week1 (9/9-11): Introduction

**Week 2**

- Lecture (9/16): Why Learn About Statistics?/What are data?
  - Readings: Hahn and Godley: Chapters 1 & 2
  - Videos:
    - Chapter 02: 1. Mathematical Operations (06:22)
    - Chapter 02: 2. Data, Observations and Variables (09:08)
- Lab (9/18) Lab 1: Data Management.
  - Hadley Wickham *R for Data Science* Chapters 1,2,4,5 (sections 5.1-5.5)

**Week 3**

- Lecture (9/23): Univariate Statistics and Graphs
  - Readings: Hahn and Godley: Chapters 3;
  - Videos:
    * Chapter 03: 1. Frequencies (04:32)
    * Chapter 03: 2. Rates, Ratios and Percentiles (05:42)
    * Chapter 03: 3. Visualizing Data (05:21)
- Lab (9/25) Data @ Western Workshop (Kristi Thompson and Liz Hill, Data Library)

**Week 4**

- Lecture (9/30) No class - National Day for Truth and Reconciliation
- Lab (10/2) Lab 2: Graphs.
  - *R for Data Science* Chapter 3
  - Andersen and Armstrong Chapter 3 (sections 3.1-3.2)

**Week 5**

- Lecture (10/7): Univariate Statistics, Graphs and Probability
  - Readings: Hahn and Godley: Chapters 4 & 5.
  - Videos:
    * Chapter 04: Probabilities (10:36)
    * Chapter 05: Describing Distributions (05:21)
- Lab (10/9): Recoding, missing values

**Week 6**

- No Class - Fall Reading Week

**Week 7**

- Lecture (10/21): Measures of Central Tendency and Dispersion
  - Readings: Hahn and Godley: Chapters 6 & 7
  - Andersen and Armstrong Chapter 4 (section 4.2)
  - Videos:
    - Chapter 06: 1. Measures of Centre (07:49)
    - Chapter 06: 2. Measures of Spread (03:29)
    - Chapter 07: Probabilities Under the Normal Distribution (09:01)
- Lab (10/23) Lab 3: Summary Statistics and Reshaping Data

**Week 8**

- Lecture (10/28): Sampling and Generalization
  - Readings: Hahn and Godley: Chapters 8 & 9
  - Videos:
    - Chapter 09: 1. Sampling Distributions (04:38)
    - Chapter 09: 2. Confidence Intervals (08:14)
- Lab (10/30) Understanding Confidence Intervals

**Week 9**

- Lecture (11/4): Bivariate Statistics
  - Readings: Hahn and Godley: Chapters 10 & 11
  - Videos:
    - Chapter 10: Hypotheses and Hypothesis Testing (11:40)
- Lab (11/6) Presenting Bivariate Statistics.

**Week 10**

- Lecture (11/11): Bivariate Statistics for Nominal and Ordinal Data
  - Readings: Hahn and Godley: Chapters 12 & 13
  - Andersen and Armstrong Chapter 4 (Section 4.2)
  - Videos:
    - Chapter 12: Contingency Tables (10:03)
- Lab (11/13) Hypothesis Tests and Effect Sizes for Contingency Table Statistics.

**Week 11**

- Lecture (11/18): Bivariate Statistics for Interval/Ratio Data
  - Readings: Hahn and Godley: Chapters 14
  - Andersen and Armstrong Chapter 4 (section 4.3)
  - Videos:
    - Chapter 14: Pearson Correlation Coefficient (02:48)
- Lab (11/20) Correlations, hypothesis tests for correaltions.

**Week 12**

- Lecture (11/25): OLS Regression
  - Readings: Hahn and Godley: Chapter 16
  - Andersen and Armstrong Chapter 2 (section 2.3) and Chapter 5 (sections 5.1-5.6).
  - Videos:
    - Chapter 16: 1. OLS Regression (06:38)
    - Chapter 16: 2. Multiple OLS Regression (09:53)
- Lab (11/27) Linear Regression

**Week 13**

- Lecture (12/2): Sampling Weights and Final Exam Discussion
- Lab (12/4): Questions about Material for Final Exam.

**Final Exam**

- In-person exam: Monday 12/9, 9:30 - 11:30 AM
- Take home exam: Available Monday 12/9 12:00 PM (or directly after you take the in-person exam, whichever is later).