

# POLS 641: Introductory Analysis of Political Data

Dr. Ches Thurber

*Fall 2021*

E-mail: [cthurber@niu.edu](mailto:cthurber@niu.edu)

Office Hours: M/T 9:30-11:00 a.m.

Office: 414 Zulauf Hall

Web: [www.chesthurber.com](http://www.chesthurber.com)

Class Hours: Mon/ Wed 2:00-3:15 p.m.

Class Room: 464 DuSable

---

## Course Description

This is the first course in the NIU Political Science Department's quantitative methods sequence. It is required of all graduate students and is intended as an introduction to the methodological tools used by social scientists in analyzing numeric data. The primary goal of the course is to increase students' comfort in engaging with and thinking about quantitative analysis. It will emphasize the linkages between concept and measurement, the fundamentals of distributions, randomness and uncertainty, and the challenges of statistical inference. The emphasis is on fundamental concepts and programming proficiency, with relatively little formal mathematics. Students should leave the course with the ability to engage with quantitative empirical research, even if they do not fully understand the methods being employed, and with the basic statistical and computer programming tools to pursue more advanced study.

## Mode of Delivery

This course will be conducted in person, twice a week. At least for now, Students are expected to bring a laptop to class capable of running statistical software. Generally, we will focus on substantive topics on Mondays and programming on Wednesdays. If you are not comfortable attending in person, I will make it fully possible to attend class via Zoom. If a deteriorating public health situation requires us to switch fully to remote learning, we will do so.

## A Note on Learning Amidst Pandemic

I was hoping to be able to delete this section this year. But the virus had other ideas. These are still crazy times. We will be proceeding with this class the best we can and to some degree I hope the process of reading, writing, and talking about topics we are all passionate about helps provide you with some sense of normalcy. But sometimes that can feel like absurdity. Please know that you are not alone in feeling it and that we are all in this together. I will be as accommodating as I can this semester in light of these circumstances and I hope that we can all agree to engage each other in a spirit of mutual patience, empathy, and understanding. If you or somebody you know is struggling with anxiety or other issues, do not hesitate to reach out. Resources available include the DRC, Student Counseling Services or call 815-306-2777.

## Office Hours

I will be holding office hours on Mondays and Tuesdays between 9:30 and 11:00. Office hours appointments are available either in-person in my office or online. I strongly encourage you to take advantage of this; since we don't have class in person, this is the best chance I have to get to know you! Sign up for an appointment at [this link](#). The system will send you a Zoom link that we can use for our meeting. Or you can just show up at my office at that time.

## Texts

The two required textbooks for the course are:

- John Gerring and Dino Christensen, *Applied Social Science Methodology: an Introductory Guide*. Cambridge University Press, 2017.
- Michael A. Bailey. *Real Stats: Using Econometrics for Political Science and Public Policy*. Oxford University Press, 2020.

I also strongly recommend the following new book on data visualization using R and GGPlot. You can buy a hard copy online, or use a free online version available [here](#).

- Kieran Healy. *Data Visualization: A Practical Introduction*. Princeton University Press, 2018.

And the following book, available for free online, offers helpful R code to implement all of the analyses, tables, and figures, presented in the Bailey textbook:

- Carilli, Tony. *An R Companion to Real Econometrics*.

Finally, the short book by Stephen Van Evera below was a valuable tool to me as a graduate student, and may be to you as well.

- Stephen Van Evera, *Guide to Methods for Students of Political Science*. Cornell University Press, 1997.

Additional texts will be available on the Blackboard site. This includes both some additional readings on the syllabus, as well as supplementary texts. Statistics can be hard and confusing and sometimes seeing the same thing presented in different ways can be helpful.

## Computer Programming in R

As part of this course, you will learn to use the statistical programming language R. Why R? Because it's free. And also because it's the one your instructor uses in his own research (Why? See reason 1...) We will talk about some of the other statistical programming languages in class and their relative advantages and weaknesses. Whatever software is being used however, learning to program, especially for the first time, can be difficult. In fact, even for advanced users, dealing with coding errors is time-consuming, mentally exhausting, and often extremely frustrating. We will therefore proceed at a slow pace, walk through everything in class, and I will encourage you to work in groups on homework problems. I only ask that your final submitted work be your own.

## Evaluation

*Class attendance, preparation and participation (20%):* The time we spend in class is for me the most important part of this course. As such, punctual attendance is mandatory. But more important than just being present at a desk is that you are actively engaged. I expect that you have done the readings and that you try to participate in discussion each and every class section. I also expect that you follow along with programming demonstrations and work with your classmates to complete in-class exercises.

*Problem Sets and Data Camp(40%):* You will have a number of problem sets to complete over the course of the semester. I encourage you to work in groups to complete the problem sets. However, I ask that the final work submitted be your own. This can be a little ambiguous, but to me means that 1) group work is conducted in a good-faith collaborative effort in which all participants are actively seeking to contribute; and 2) written work, including code, is produced by the individual student, not copied and pasted from other students' work. I will also assign some lessons through the website Data Camp. I will check for completion and these will be worth less than the problem sets.

*Final Project (40%):* You will produce a final project in written and presentational form. This project will involve an independent quantitative analysis of some type in which you develop a hypothesis or more, present relevant data and provide visualizations and basic descriptive measures, and evaluate your hypothesis with appropriate statistical tests including (but not limited to) linear regression. You will write-up a 10-page (or so) mini-paper and deliver a 10-minute conference-style presentation on the final day of class.

## Grading Standards

Grades in graduate school are a little bit silly, in my opinion, as you are at the point in your careers where it is your written work itself that matters (and how it is received by a broader scholarly community) more than the grade arbitrarily assigned by a single cantankerous professor. Nevertheless, they can serve as a helpful signaling device for your own self-assessment, for departmental funding decisions, and to admissions committees at PhD programs (for MA students planning to continue on...). The scheme below is taken from Prof. Kyle Beardsley in the political science department at Duke University with some modifications to adapt to our departmental norms and my own personal views. It may also be helpful to think about the grading distributionally. In the past, I have generally awarded between 1 and 3 straight As per graduate seminar. The modal grade has been an A-, with a few Bs and B+s.

- **A** : *Exceptional Performance.* Outstanding work on all course-related tasks at a level that distinguishes the student from other members of the class. A comprehensive and incisive command of the issues, literature, and substantive information relevant to the course. The ability to master and integrate large amounts of factual material and abstract theories.
- **A-** : *Very Good Performance.* Consistently strong work on all course-related tasks. A command of the issues, literature, and substantive information relevant to the course. Understands well and can integrate the relevant factual and theoretical material central to the course.
- **B+** : *Good Performance.* Solid work on all course-related tasks. A good grasp of the issues, literature, and substantive information relevant to the course. A very good command of factual and theoretical material, and some capacity to integrate the two.

- **B** : *Decent Performance*. Generally consistent work on most course-related tasks. A general understanding of the issues, literature, and substantive information relevant to the course. An acceptable understanding of factual and theoretical material, but limited evidence of the capacity to integrate the two.
- **B-** : *Barely Satisfactory Performance*. Mostly satisfactory work on course-related tasks, but with notable deficiencies. A general understanding of the issues, literature, and substantive information relevant to the course. Understands at a basic level the facts and theories related to the course, but with clear gaps, errors, or incomplete work. Grades lower than this represent unsatisfactory work.

### Course Policies

- Successful completion of this course requires consistent, punctual attendance, and active participation in both class discussion and class exercises. If you must miss class due to a conflicting responsibility or an emergency, it is your responsibility to contact me to discuss the situation. I will generally require a short written assignment to be submitted in lieu of class attendance. Any more than 1 absence without prior notification and make-up work will affect your final grade.
- If you think you need to delay submitting a problem set, you should talk to me, generally at least two weeks before the due date. Such extensions are never guaranteed, but will be based on university policy, exigency of the circumstances, and timeliness of the request (the earlier the better). Any missed written assignment submitted after the deadline without prior authorization from me will not be accepted and a grade of 0 will be assigned.)
- All students are expected to adhere to the highest levels of academic integrity. Violations of university, departmental, and disciplinary standards will not be tolerated and will lead to an F for the course. This includes forms of plagiarism as well as honesty and transparency in data analysis. We will talk at length in class about what this means.
- If you need an accommodation for this class, please contact the Disability Resource Center as soon as possible. The DRC coordinates accommodations for students with disabilities. It is located on the 4th floor of the Health Services Building, and can be reached at 815-753-1303 or [drc@niu.edu](mailto:drc@niu.edu). Also, please contact me privately as soon as possible so we can discuss your accommodations. Please note that you will not be required to disclose your disability, only your accommodations.
- It is my personal policy to allow graduate students to call me by my first name, “Ches.” This reflects the idea that I view you all as colleagues-in-training. Please let me know how you prefer to be addressed, both in name and pronoun, if it differs from what is in the college directory. I will make every effort to address you in the way you wish to be addressed. Please try and do the same for your fellow classmates, as well as for other faculty in the department.
- I am committed to your success in this class – if you feel that you are not performing to your expectations, please come and see me. I am available to answer any questions you may have about course assignments, requirements or content. I generally answer e-mails within 24 hrs on weekdays, and would be happy to schedule an appointment to meet with you if you are unavailable during my posted office hours.

## Class Schedule

Students are expected to read the indicated assignments and complete the assigned Data Camp exercises prior to the indicated class.

### Week 1: Introduction

*Monday 08/23: What Are We Doing Here?*

- ASSM: Preface-Ch. 2
- [Healy, Ch. 2](#)

*Wednesday 08/25: An Introduction to R*

- DATA CAMP: Intro to Basics, Vectors, Matrices

### Week 2: Describing the Social World

*Monday 08/30: Concepts, Measures, and Analyses*

- ASSM: Chs. 3-4

*Wednesday 09/01: Data Frames in R*

- DATA CAMP: Factors, Data Frames, Lists

### Week 3: Data Visualization

*Monday 09/06: LABOR DAY - NO CLASS*

- Optional: [Healy, Ch. 1](#)

*Wednesday 09/08: Intro to the Tidyverse*

- DATA CAMP: Data Wrangling, Data Visualization

### Week 4: Describing Data

*Monday 09/13: Univariate Statistics*

- ASSM: Ch. 18

*Wednesday 09/15: Describing Data in R*

- DATA CAMP: Grouping and summarizing, Types of Data Visualizations

### Week 5: Probability

*Monday 09/20: Probability Distributions*

- ASSM: Ch. 19

*Wednesday 09/22: Bringing it all Together in R*

- PROBLEM SET #1 DUE

## **Week 6: Statistical Inference**

*Monday 09/27: Bivariate Statistics*

- ASSM: Chs. 20-21

*Wednesday 09/29: Hypothesis Testing*

- DATA CAMP: Yum That Dish Tests Good, Letting the Categoricals Out of the Bag

## **Week 7: Regression**

*Monday 10/04: Intro to Regression*

- ASSM: Ch. 22

*Wednesday 10/06: Regression in R*

- DATA CAMP: Simple Linear Regression, Predictions and Model Objects

## **Week 8: Causation**

*Monday 10/11: Causal Inference*

- ASSM: Ch. 23
- RS: Ch. 1

*Wednesday 10/13: Catch-Up Day*

- PROBLEM SET #2 DUE

## **Week 9: Digging Deeper into Bivariate OLS**

*Monday 10/18: Mathematics and Assumptions*

- RS: Ch. 3

*Wednesday 10/20: Assessing Fit and Assumptions*

- DATA CAMP: Assessing Model Fit

## **Week 10: Statistical Inference and Regression**

*Monday 10/25: Hypothesis Testing and Interval Estimation*

- RS: Ch. 4

*Wednesday 10/27: Putting It All Together*

- Problem Set #3 Due

## **Week 11: Controlling for Confounding Variables**

*Monday 11/01: Multivariate OLS*

- RS: Ch. 5

*Wednesday 11/03: Multivariate in Practice*

- DATA CAMP: Parallel Slopes

## **Week 12: Categorical Inputs in Multivariate OLS**

*Monday 11/08: Dummy Variables*

- RS: Ch. 6

*Wednesday 11/10: Interactions*

- DATA CAMP: Interactions, Multiple Linear Regression

## **Week 13: Specifying Models**

*Monday 11/15: Transformations*

- RS: Ch. 7

*Wednesday 11/17: Mixed Methods*

- PROBLEM SET #4 DUE

## **Week 14: Thanksgiving Week**

- No Class, work on your final projects, Zoom meetings

## **Week 15: Final Presentations**

*Monday 11/29: Presentation Session 1*

*Wednesday 12/01: Presentation Session 2*

**Final Papers Due : *Wednesday 12/08: 5:00pm***