Political Data Analysis

POS 6737 –Class Number 14676 (M.A. course)
Department of Political Science, University of Florida
Spring 2026

PROF. SUZANNE M. ROBBINS
205 ANDERSON HALL
SUZANNE.ROBBINS@UFL.EDU

OFFICE HOURS: MONDAY & FRIDAY 12-3 PM
CLASS MEETS: WEDS 11:45AM-2:45 PM, ROGERS FRAZIER HALL, ROOM 129

1 COURSE DESCRIPTION

This course introduces the statistical tools most commonly used to process, analyze, and visualize data in the social sciences. We will explore descriptive statistics, statistical inference, hypothesis testing, and linear regression, with a brief introduction to logistic regression. Using the statistical software package R, students will learn to transform, visualize, and analyze data with a strong emphasis on interpreting results for real-world applications. Each topic will cover methodology, including underlying assumptions, the mechanics of the analysis, and appropriate interpretation of results. Throughout the course, we will work with real political data to ground our learning in practical examples.

2 COURSE OBJECTIVES

This course is designed to achieve three main goals:

1. Build a strong foundation in statistical concepts to prepare students for advanced methods and analyses.
2. Equip students with statistical skills to understand and critically engage with political science research that employs statistical analysis.
3. Enable students to analyze real-world political data and communicate findings effectively.

Weekly class meetings will follow a lecture-based workshop format. Proficiency in R for statistical computing is essential for success in this course.

Specific goals this semester include:

- Develop testable hypotheses based on political science research questions.
- Enhance statistical literacy by learning to:
  - Summarize and display data effectively.
  - Compute and interpret descriptive statistics.
  - Construct confidence intervals and conduct hypothesis tests for numerical variables (e.g., t-tests).
  - Prepare contingency tables and conduct hypothesis tests for categorical variables (e.g., Chi-square tests).
  - Build, interpret, and draw inferences from bivariate and multivariate linear regression models.
  - Evaluate statistical findings in academic research and popular media.

- Present statistical analyses professionally, with clear visualizations and narratives.
- Develop modern statistical computing skills, including:
  - Reading, cleaning, and organizing data.
  - Practicing good coding habits and etiquette.
  - Conducting reproducible analyses.

3 REQUIRED MATERIALS

3.1 REQUIRED READING MATERIALS

1. Holbrook, Thomas M., 2024. *An Introduction to Political and Social Data Analysis (with R).* Sage Publications. Available via UF All Access (via Canvas!)

3.2 RECOMMENDED READING MATERIALS

1. Thulin, Mans. 2024-07-01. *Modern Statistics with R, 2nd Edition.* Available free online: https://www.modernstatisticswithr.com/
2. Grolemund, G., and H. Wickham. 2023. *R for Data Science*, *2nd Edition.* Available free online: https://r4ds.hadley.nz/
3. Matloff, N. nd. fasteR: Fast Lane to Learning R!. Available free online: https://faster-site.netlify.app/

3.3 TECHNOLOGY REQUIREMENTS

All models in this class will be estimated using R software packages and a standard computer. The following are required to complete the course:

1. A laptop computer with wifi access
2. R statistical computing software and related packages (free and open source; we will set this up together the first day of class)

4 ASSIGNMENTS/ASSESSMENT

4.1 OVERVIEW

Student progress will be measured using multiple methods. The class consists of homework assignments, an in-class written midterm exam, and a final project.

Work diligently and persistently. Attend classes. Read carefully before the seminar meets. Do the work on time. Practice R. Communicating your results to others is as important as getting good results in the first place. Every assignment – homework, exam, paper - requires interpretation and is as important as getting the correct result. ***Professionalism matters: do not submit raw computer output as you will not receive credit.***

4.2 WEEKLY PROBLEM SETS (30%):

We will have problem sets almost every week. The problem sets incorporate both the material from the lecture and what we are learning in R. The more you practice, the better you get – and the more you learn. We will start each problem set in class and you will complete it on your own before the next class (and submit to canvas).

While group discussion and work are explicitly encouraged, you are required to submit your own computer code and final write-up of the answers. DO NOT simply copy computer code or answers from your classmates.

All work in the class should be professional: a well-formatted, electronic format (e.g., R Markdown). Late homework is penalized 5% per day; no homework will be accepted more than five days late except in extenuating (documented) circumstances.

The homework may be difficult, especially in the beginning. Keep trying, do not give up.

Problem Set Grading Rubric

Homework assignments are graded on a percentage scale (0–100%) and converted to letter grades based on the following criteria:

- **A (93–100%)**:
  The problem set is **exceptional**. Every question is answered, all responses are correct, and interpretations are clear and insightful. The document is professional and easy to follow. Code is well-written, efficient, and demonstrates excellent attention to detail.
- **A- (90–92%)**:
  The problem set is **outstanding**, with most questions answered and the vast majority correct. Answers are thoughtful, and code is clear and functional, with only minor issues in execution or presentation. *This is the expected level of performance*
- **B+ (87–89%)**:
  The problem set demonstrates **very good performance**. Most questions are completed, with generally correct answers and interpretations. Code and presentation are effective but may contain minor errors or lack polish.
- **B (84–86%)**:
  The problem set is **solid but has noticeable flaws**. While most of the work is complete and correct, errors in interpretation, coding, or clarity suggest the need for more careful attention to detail.
- **C Range (70–83%)**:
  The problem set shows **adequate to below-average performance**. Completion and accuracy vary, with significant gaps in understanding, coding, or interpretation. Additional effort is needed to meet course expectations.
- **D Range (60–69%)**:
  The problem set is **incomplete or incorrect**. Significant issues with completion, accuracy, or effort indicate a lack of understanding. Students receiving this grade are encouraged to seek assistance immediately.
- **E (<60%)**:
  The problem set is **failing**. It is incomplete, contains pervasive errors, or shows minimal effort. This level of work is unacceptable and requires immediate improvement.

4.3 MIDTERM EXAM (35%):

The in-class midterm will include problem sets and key definitions covering material from the first part of the course. You will work with a small dataset to calculate descriptive and inferential statistics. Additionally, you will be evaluated on the quality of your interpretations, demonstrating a strong understanding of the concepts.

4.4 FINAL PROJECT (35%):

The final project will require you to analyze data and present professional results to a client (the instructor). The goal is to create a polished report that showcases your ability to apply the knowledge gained throughout the semester. Some problem sets will include questions directly related to your project, providing opportunities for feedback and skill development (scaffolding). You will not need to find your own data; all required datasets will be provided. A well-crafted final project can make a significant difference in demonstrating your mastery of the material. Additional details are available on Canvas.

4.5 GRADING SCALE:

I will use the following grading scale for those assignments receiving a letter grade as well as your overall grade. Please note that I do not round grades until the final course grade.

A	93-100	B	84-86	C	74-76	D	64-66
A-	90-92	B-	80-83	C-	70-73	D-	60-63
B+	87-89	C+	77-79	D+	67-69	E	<60

5 COURSE POLICIES

5.1 ATTENDANCE/LATE WORK:

Attendance is required. You should come prepared to work. Do not be late to class. I note who attends class/is late, even if it doesn't count toward your course grade. If you must miss class, you will need to get notes from a classmate. Missing even a single class can be difficult to overcome, so do your best to avoid it.

Requirements for **class attendance** are consistent with the attendance policy stated in the Graduate Catalog Regulations found here: http://gradcatalog.ufl.edu/content.php?catoid=6&navoid=1219.

Late work is penalized 5% per day and not accepted after five days, documented extenuating circumstances notwithstanding.

5.2 FALLING BEHIND:

This class will move quickly. At first you may feel comfortable with the material, or you may feel intimidated walking in. I am here to help you. If you fall behind, I cannot help you unless you come see me.

5.3 Statement on the Use of AI Tools

In this graduate statistics course, the primary goal is to develop your analytical and interpretive skills. The use of AI tools such as ChatGPT or other coding assistance is allowed with specific restrictions:

1. **AI for Coding**: You may use UF Navigator or other AI tools to assist with coding, such as generating or troubleshooting R code. However, you must **document** the use of AI by providing a description of the prompt(s) used and citing the AI tool. Include this information in your code or homework/project report to ensure transparency. Note that some tools generate needlessly complex R code.
2. **AI for Writing and Interpretation**: AI tools **may not be used to assist with writing**, including completing homework assignments or the final project, **nor should they be used to interpret** data or statistical results. All writing and interpretation must reflect your original work, demonstrating your understanding and critical analysis of the material.
3. **Critical Evaluation**: Any content generated by AI must be critically evaluated. AI tools can be inaccurate, incomplete, or fail to capture the depth of analysis required for graduate-level work. Always cross-check AI outputs with the assigned readings, class materials, and your own reasoning to ensure accuracy and depth. Note that AI tools can generate incorrect code.
4. **Privacy and Data Collection**: While you may use AI tools of your choice, **UF's Navigator** is recommended, as it ensures that no personal data is collected.
5. **Academic Integrity**: All work submitted must comply with UF's Academic Integrity policies. Any use of AI that violates these policies—such as using AI for writing, interpretation, or failing to properly document AI assistance—will be considered academic misconduct.

By following these guidelines, you can utilize AI tools to support your learning while maintaining academic integrity and ensuring your work reflects your own skills and understanding.

5.4 Communications and Logistics:

Almost all our work will take place in class, on your own time, and within the Canvas environment. Please feel free to email me at any time. While I try to answer email quickly, I sometimes need 24-48 hours to do so. Please keep in mind that I am not always available via email evenings and weekends.

To protect your federally protected privacy rights, I cannot discuss your grade(s) over the phone, via email, or in front of class. Please make an appointment.

5.5 Other:

This course complies with all UF academic policies. For information on those polices and for resources for students, please see this link." (The direct link is https://syllabus.ufl.edu/syllabus-policy/uf-syllabus-policy-links/.)

6 COURSE SCHEDULE

Date	Topic	Reading/Problem Set:
14-Jan	Research Process; Ethics The R Environment	Holbrook: Chapter 1, through page 14 Holbrook: Chapter 2
21-Jan	Levels of Measurement, Causality Frequency Distributions RStudio, Workflow, Quarto, Markdown	Holbrook: Chapter 1, remainder Holbrook: Chapter 3, through page 60 Holbrook: Appendix B
28-Jan	Measures of Central Tendency Data Visualization Data Preparation	Holbrook: Chapter 5 Holbrook: Chapter 3, remainder Holbrook: Chapter 4 *Problem Set 1 due*
4-Feb	Measures of Dispersion More Data Prep & Viz	Holbrook: Chapter 6 *Problem Set 2 due*
11-Feb	Probability Sampling & Inference R Review and Qs (every week going forward)	Holbrook: Chapter 7 Holbrook: Chapter 8 *Problem Set 3 due*
18-Feb	Hypothesis Testing with 1 or 2 groups	Holbrook: Chapters 9 & 10 *Problem Set 4 due*
25-Feb	Hypothesis Testing with Multiple Groups	Holbrook: Chapter 11 *Problem Set 5 due*
4-Mar	Hypothesis Testing with Non-Numeric Variables (cross-tabs) Review	Holbrook: Chapter 12
11-Mar	Midterm	
18-Mar	Spring Break	
25-Mar	Measures of Association Correlation & Scatterplots	Holbrook: Chapter 13 Holbrook: Chapter 14
1-Apr	Regression I	Holbrook: Chapter 15 *Problem Set 6 Due*
8-Apr	Regression II	Holbrook: Chapter 16 *Problem Set 7 due*
15-Apr	Regression III	Holbrook: Chapter 17
22-Apr	Regression IV	Holbrook: Chapter 18 *Problem Set 8 due*
29-Apr	Final Project Due	