

POLS 3030 H01: Data Science For Politics

Fall 2023

Lecture: Tuesday & Thursday 1:00 p.m. – 2:15 p.m.
Classroom: Smith 325
Professor: Dr. Yaoyao Dai
Email: yaoyao.dai@charlotte.edu (include “POLS3030” in subject line)
Office: Fretwell 450B
Office Hours: Tuesday 2:30 p.m. – 3:30 p.m.; Thursday 2:30 – 3:15 p.m.
book through <https://yaoyaodai.youcanbook.me/>

This syllabus contains the policies and expectations I have established for POLS 3030 H01. Please read the entire syllabus carefully before continuing in this course. I reserve the right to modify the standards and requirements outlined in this syllabus at any time. I will notify you of such changes by announcements in class and posting a new syllabus on Canvas.

1 Course Description

POLS 3030 Data Science for Politics provides an introduction to data science for social data and problems. In this course, we will learn the way data science has changed our political and daily lives. In addition, students will learn through hands-on projects the basic and fundamental tools of data science and apply them to a wide range of political and policy-oriented questions. Prior knowledge of programming and statistics is not required.

The **Objectives** of this course are twofold:

- This course will provide you with the knowledge to *understand* data science and how advances in data science impact society. To this end, we’ll read the most recent social science papers evaluating the impacts of algorithms on individual behavior and political outcomes.
- This course will provide you with data science tools and skills. To this end, we’ll learn basic programming in Python, data visualization, intuitions behind machine learning algorithms, and applied machine learning practice.

2 Computation

We will teach the course using Python and code using Google Colab. Python is an open-source general-purpose programming language. It is the most popular programming language in data science. Colab is a Google service that requires no setup to use Python Jupyter Notebook and provides free access to computing resources. The best and only way of learning to program is through practicing. Therefore, we will use DataCamp for out-of-class practice for weeks without lab assignment: <https://learn.datacamp.com>.

3 Course Requirements

Attendance and **Participation** in all lectures and lab sessions are **required** unless otherwise noted by the professor. Attendance in 10 random lectures/labs will be graded. To facilitate class participation, there will be various in-class activities, such as polls, quizzes, and discussions. We will use **Poll Everywhere** for in-class polls and quizzes. Students can access Poll Everywhere using a phone, laptop, or tablet to answer in-class polls and quizzes. Please check <https://www.polleverywhere.com/> and <https://teaching.uncc.edu/academic-technologies/student-response-system/student-guides> for more information.

Readings. All readings should be completed by the date indicated in the syllabus. Unless otherwise noted, this means that weekly readings should be completed before the first lecture each week. It is very important that you come to class prepared so you can get the most out of the lectures and are ready to ask (and answer) questions about the readings. I may call on students at random with questions from time to time.

Lab Assignments. There will be four problem oriented lab assignments. The lab assignments usually include both programming questions (problem-solving) and short answer questions (interpretation of results or story-telling). You will have one week to complete the lab assignment. Late lab assignments will be accepted and graded. However, they will be subject to a 1% score per day (including weekends) late penalty. Exceptions will be made on a case-by-case basis and should be discussed in advance with the professor or lab instructor.

DataCamp Assignments. There is only one way of learning programming – practice! Therefore, for weeks without a lab assignment, I will assign DataCamp tutorials/exercises. Those will be graded based on completion. Late completion will also be subject to a 1% score per day (including weekends) late penalty. Exceptions will be made on a case-by-case basis and should be discussed in advance with the professor or lab instructor.

Hackathon. As part of the data science experience, we will hold two hackathons during classes. Those are group problem-solving coding events. Each hackathon will include 1-2 classes of group coding, and one group presentation. Each member of the team that wins the Hackathon will receive one extra credit. Showcase your coding and teamwork skills in those fun coding events! Detailed instructions will be announced on Canvas and during Hackathon.

Exam. There will be one (final) exam. Since it is a highly applied class, the final exam will NOT be a quiz. It will be problem-solving project that you complete in Google Colab. Consider it as a timed lab assignment that you need to complete by yourself. You are not allowed to communicate with your peers, but you are allowed to check your notes and the internet.

4 Grading Policy

Lecture Participation	10%
(Almost) Weekly DataCamp Assignments	15%
Four Lab Assignments	20%
Hackathon I	15%
Hackathon II	20%
Final Exam	20%
	100% possible out of 100% total

Points	Letter Grade
90% – 100%	A
80% – 90%	B
70% – 80%	C
60% – 70%	D
0% – 60%	F

5 General Policies and Expectations

RESPECT. In this course, we are learning challenging statistical and computational knowledge. We will also read and discuss political science research on sensitive and controversial topics. Everyone comes to this course with a different background and knowledge in both the statistics and the substantive topics. It is

important that we all treat each other with the utmost respect. The conflict of ideas is encouraged and welcome. However, you should base your opinion on logic and empirical evidence, instead of beliefs and stereotypes. I will exercise my responsibility to manage the classroom so that ideas and arguments can proceed in an orderly fashion. You should expect that if your conduct during class seriously disrupts the atmosphere of mutual respect, you will not be permitted to participate further. Unwelcome conduct directed toward another person based upon that person's actual or perceived race, actual or perceived gender, color, religion, age, national origin, ethnicity, disability, or veteran status, or for any other reason, may constitute a violation of University Policy 406, The Code of Student Responsibility. Any student suspected of engaging in such conduct will be referred to the Office of Student Conduct.

WORKING TOGETHER. I encourage you to work together on the course materials and assignments. We learn from each other's questions and experiences. Moreover, there is no better way to master the materials than to explain to your peers. However, every keystroke that you type and every mark you make with a pen or pencil must be your own work. You **cannot** collaborate on the final exam.

LATE ASSIGNMENTS. Assignments not submitted by the assigned due date and time are late. All assignment deadlines are based on the Eastern Time Zone (ET), not your local time zone. Please be sure to pay close attention to any time differences if you are located in a different time zone. Given potential technical difficulties that may arise with things like your internet connection, I strongly recommend that you do not wait until the last moment to submit your assignments. Late submissions will be accepted; however, they will be subject to a 1% scores per day (including weekends) late penalty.

EXTENSIONS. Extensions will be granted in only the most severe circumstances. If you foresee the need for an extension, one needs to be requested and granted at least 24 hours before the due date. No one is entitled to an extension; they will be offered only at my discretion.

REGARDING GRADES. I do not give grades. You earn grades. If, at any point, you are unsure of your current standing in the course, please come to my office hours. Do not wait until grades are posted and then ask how your grade could be improved. At that point, barring a mathematical error on my part, it cannot be.

ACADEMIC INTEGRITY. All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Definitions and examples of plagiarism are set forth in the Code. The Code is available from the Dean of Students Office or online at <https://legal.uncc.edu/policies/up-407>.

All course work by students is to be done on an individual basis unless an instructor clearly states that an alternative is acceptable. The instructor may ask students to produce identification at examinations and may require students to demonstrate that graded assignments completed outside of class are their own work. Any reference materials used in the preparation of any assignment must be explicitly cited. Students uncertain about proper citation are responsible for checking with their instructor.

DISABILITY ACCOMMODATIONS. Students in this course seeking accommodations for disabilities must first consult with the Office of Disability Services and follow the instructions of that office for obtaining accommodations. I will do everything I can to meet the requested accommodations.

PREFERRED GENDER PRONOUN. This course affirms people of all gender expressions and identities. If you prefer to be called a different name than what is indicated on the class roster, please let me know. Feel free to correct me on your preferred gender pronoun. If you have any questions or concerns, please do

not hesitate to contact me.

6 Tentative Course Schedule

This schedule should be treated as tentative and flexible. It may take us more or less time for a particular topic than I have allotted here. We will adapt accordingly. In the event that deviations from this schedule are necessary, they will be announced in class and updated on Canvas. Additional readings to the textbook will be uploaded to Canvas at least one week before class.

Week 1 (Aug.21 - Aug.27): Introduction to Class. What is Data Science?.....

- **Reading:**
 - UC Berkeley, “What is Data Analytics:” <https://ischoolonline.berkeley.edu/data-science/what-is-data-analytics/>
 - UC Berkeley, “What is Data Science:” <https://ischoolonline.berkeley.edu/data-science/what-is-data-science/>
 - Karlijn Willems, *The Data Science Industry: Who Does What* (Infographic).

Week 2 (Aug.28 - Sep.03): No Class

- **Reading:**
 - Welcome to Colab: https://colab.research.google.com/?utm_source=scs-index

Week 3 (Sep.04 - Sep. 10): What is (Big) Data? Python Basics & Working with Structured Data in Python.....

- **Reading:**
 - Welcome to Colab: https://colab.research.google.com/?utm_source=scs-index
 - Brady, Henry E. "The challenge of big data and data science." *Annual Review of Political Science* 22 (2019): 297-323.

Week 4 (Sep. 11 - Sep.17): Data Visualization.....

- **Reading:**
 - Schwabish, Jonathan A. “An economist’s guide to visualizing data.” *Journal of Economic Perspectives* 28.1 (2014): 209-234.
 - Acemoglu, Daron, and Pascual Restrepo. “Robots and jobs: Evidence from US labor markets.” *Journal of political economy* 128.6 (2020): 2188-2244.
- **Lab Project I Due: Sep.14**

Week 5 (Sep. 18 - Sep. 24): Data Visualization II.....

- **Reading:**
 - Guess, Andrew M., et al. “How do social media feed algorithms affect attitudes and behavior in an election campaign?.” *Science* 381.6656 (2023): 398-404.
 - Fun data visualization video and channel by political scientist, Global exports animated: <https://www.youtube.com/watch?v=nJZ10BJREUQ>

Week 6 (Sep. 25 - Oct. 01): Introduction to Text as Data.....

- **Reading:**
 - Wilkerson, John, and Andreu Casas. “Large-scale computerized text analysis in political science: Opportunities and challenges.” *Annual Review of Political Science* 20 (2017): 529-544.
- **Lab Project II Due: Sep.28**

- Week 7 (Oct. 02 - Oct. 08): Hackathon I**.....
- Week 8 (Oct. 9 - Oct. 15): Hackathon Presentation & Introduction to Machine Learning**
- Week 9 (Oct. 16 - Oct. 22): Unsupervised Learning. Topic Modeling**.....
- **Reading:**
 - Topic Modeling in Python: Latent Dirichlet Allocation (LDA): <https://towardsdatascience.com/end-to-end-topic-modeling-in-python-latent-dirichlet-allocation-lda-35ce4ed6b3e0>
 - Pan, Jennifer, and Kaiping Chen. “Concealing corruption: How Chinese officials distort upward reporting of online grievances.” *American Political Science Review* 112.3 (2018): 602-620.
- Week 10 (Oct. 23 - Oct. 29): Supervised Learning.**
- **Reading:**
 - Dai, Yaoyao, and Alexander Kustov. “When do politicians use populist rhetoric? Populism as a campaign gamble.” *Political Communication* 39.3 (2022): 383-404.
 - **Lab Project III Due: Oct. 29**
- Week 11 (Oct. 30 - Nov. 05): Neural Network & Word Vector**
- **Reading:**
 - Garg, Nikhil, et al. “Word embeddings quantify 100 years of gender and ethnic stereotypes.” *Proceedings of the National Academy of Sciences* 115.16 (2018): E3635-E3644.
- Week 12 (Nov. 06 - Nov. 12): Large Language Models**
- **Reading:**
 - Explore: <https://huggingface.co/>
 - Spitale, Giovanni, Nikola Biller-Andorno, and Federico Germani. “AI model GPT-3 (dis) informs us better than humans.” arXiv preprint arXiv:2301.11924 (2023).
- Week 13 (Nov. 13 - Nov. 19): Basic Web-scraping or website making**
- **Lab Project IV Due: Nov. 19**
- Week 14 (Nov. 20 - Nov. 26): Thanksgiving. No Class**
- Week 15 (Nov. 27 - Dec. 3): Hackathon II**
- Week 16 (Dec. 04 - Dec. 06): Hackathon Presentation**.....
- Final Exam: TBD**

7 Additional Information

Counseling and Mental Health Services. It is common for college students to experience challenges that may interfere with academic success such as academic stress, sleep problems, juggling responsibilities, life events, relationship concerns, or feelings of anxiety, hopelessness, or depression. If you or a friend is struggling, we strongly encourage you to seek support. Helpful, effective resources are available on campus at no additional cost.

Visit the Counseling and Psychological Services website at caps.uncc.edu for information about the broad range of confidential on-campus mental health services, online health assessments, hours, and additional information.

Call CAPS at (704) 687-0311 if interested in scheduling an appointment with a counselor. After-hours crisis support is also available through this phone number.

Title IX Statement. UNC Charlotte is committed to providing an environment free of all forms of discrimination and sexual harassment, including sexual assault, domestic violence, dating violence, and stalking. If you (or someone you know) has experienced or experiences any of these incidents, know that you are not alone. UNC Charlotte has staff members trained to support you in navigating campus life, accessing health and counseling services, providing academic and housing accommodations, helping with legal protective orders, and more.

Please be aware that many UNC Charlotte employees, including all faculty members, are considered Responsible Employees who are required to relay any information or reports of sexual misconduct they receive to the Title IX Coordinator. This means that if you tell me about a situation involving sexual harassment, sexual assault, dating violence, domestic violence, or stalking, I must report the information to the Title IX Coordinator. Although I have to report the situation, you will still have options about how your case will be handled, including whether or not you wish to pursue a formal complaint. Our goal is to make sure you are aware of the range of options available to you and have access to the resources you need.

If you wish to speak to someone confidentially, you can contact any of the following on-campus resources, who are not required to report the incident to the Title IX Coordinator: (1) University Counseling Center (counselingcenter.uncc.edu, 7-0311); (2) Student Health Center (studenthealth.uncc.edu, 7-7400); or (3) Center for Wellness Promotion (wellness.uncc.edu, 7-7407). Additional information about your options is also available at titleix.uncc.edu under the “Students” tab.

Intellectual Property. My lectures and course materials, including presentations, tests, exams, outlines, and similar materials, are protected by copyright. I am the exclusive owner of copyright in those materials I create. I encourage you to take notes and make copies of course materials for your own educational use. However, you may not, nor may you knowingly allow others to reproduce or distribute lecture notes and course materials publicly without my express written consent. This includes providing materials to commercial course material suppliers such as CourseHero and other similar services. Students who publicly distribute or display or help others publicly distribute or display copies or modified copies of an instructor’s course materials may be in violation of University Policy 406, The Code of Student Responsibility. Similarly, you own copyright in your original papers and exam essays. If I am interested in posting your answers or papers on the course website, I will request your written permission.

Video and/or audio recording is not permitted during class unless the student obtains permission from the instructor. If permission is granted, any distribution of the recording is prohibited. Students with specific electronic recording accommodations authorized by the Office of Disability Services do not require instructor permission; however, the instructor must be notified of any such accommodation prior to recording. Any distribution of such recordings is prohibited.