BSOS 233: Data Science for Social Sciences
Tuesdays, Thursdays: ASY 1213 2pm - 2:50pm,
Wednesday Lab: ASY 1213

Instructor: Brian Kim (kimbrian@umd.edu)
Office: LeFrak 1218S
Office Hours: Tuesday/Thursdays 3-4pm, or when available by appointment

TA: Quinn Dang (qdang@umd.edu)
Office Hours: Tuesdays 1-2pm, Wednesday 2-3pm

An introduction to modern methods of data analysis for social scientists. This course emphasizes teaching students who have no previous coding experience how to analyze data and extract meaning in a social science context. Students will gain critical programming skills and learn inferential thinking through examples and projects with real-world relevance.

Learning Objectives

After completing this course you will be able to:

- Understand how Python can be used for data analysis.

- Think and reason about inferential statistics.

- Apply statistical methods to social issues.

Course Design

This course is designed to introduce you to Python programming for data science application in the social sciences. All of the Python programming material builds on itself over the course of the semester. This means that keeping up with all material is crucial to succeeding in this class. If you fall behind on the programming aspect, you may find it hard to catch up. **If you are feeling lost or having trouble keeping on the coding, please let the TA or me know early so we can work to get you up to speed**. All of the homework assignments, lab assignments, and projects will contain at least a little bit of coding in it, so you need to make sure you are diligent about not falling behind on coding concepts.

We will be doing all our coding in the JupyterHub cloud environment. You can access this at `https://bsos233.umd.edu`. You do not need anything installed on your computer for this. We will go over creation of accounts and accessing the environment on the first day of class.

Course Format

This is a **HyFlex (hybrid-flexible) course**. It is designed to give you the flexibility to choose how you want to attend. Each Tuesday/Thursday class will take place both in the classroom (ASY1213) and

on Zoom (link available through ELMS-Canvas). **You may choose whether to attend in-person or online each class.**

Regardless of modality, you are **expected to attend and participate actively in each class.** This includes following along in Jupyter, discussing ideas with peers, and asking questions.

All labs (Wednesdays) will take place in-person in ASY 1213.

Should I attend online or in-person?

You should choose to participate in the way that will best support your learning. Consider each of the following.

- **Logistics**: Does your roommate like to nap on Thursday afternoons? You'll want to attend in-person. Do you have another class across campus right before this one? You may want to attend online. If you choose to attend remotely, make sure you have the appropriate space and technology to allow you to actively participate in whole-class and small group discussions.

- **Distractions**: If you attend a few classes online but notice yourself opening other tabs and struggling to focus, consider coming in-person to minimize distractions.

- **Understanding**: If you are having trouble understanding the course content, you should attend in-person. This will allow me to better support you. Use the weekly check-in quizzes and your class conversations with peers to monitor your own understanding and progress. The course content builds on itself throughout the semester, so don't let yourself fall behind.

Recordings

Lectures will be recorded and posted in ELMS-Canvas (click Zoom then Cloud Recordings tab). If you have an excused absence (see attendance policy), you may view the recorded lecture as your make-up work. If you are struggling to understand something, you may want to re-watch that part of the lecture after class.

Course Materials

Required Readings

Completing the required reading for the class is essential to understanding the core statistical concepts. In order to learn, you must review the material multiple times. The required reading listed below in the course schedule is given for the main textbook used in the class. Any additional required or optional readings for this course will be posted on ELMS.

Required Textbook

- Ani Adhikari & John DeNero, *The Foundations of Data Science*. This book is available for free online under a Creative Commons License at: `https://www.inferentialthinking.com/`.

Course Activities

Weekly Check-in Quiz (10% of grade)

There will be a short quiz on Canvas every week that must be completed by Monday. The quiz will cover material in the reading for that week as well as possibly some material from previous weeks. You will be allowed to retake the quiz as many times as you want before the Monday deadline.

Lab Assignments (20% of grade)

Wednesdays will be lab sessions focused on learning how to code in Python. Part of this will include working on in-class lab assignments that will be due at the end of the day of the lab sessions. These lab assignments will graded mostly on completion.

Homework Assignments (25% of grade)

There will be a total of 5 assignments designed to assess your mastery of the topics and techniques covered in the lectures. The assignments will generally be assigned around two weeks before the due date.

Midterm Project (15% of grade)

You will have a midterm project due at the end of Week 9. More information about this project will be provided during Week 7.

Final Project(30% of grade)

This will be due during Finals week. More information about the Final Project will be provided in Week 12.

Grading

Grades will be assigned based on the total percent earned, using the following rubric. Grades will be rounded to the nearest 10th of a percent. Please come and talk to me early if you think that there might be a problem.

A+	97.0-100.0%	B+	87.0-89.9%	C+	77.0-79.9%	D+	67.0-69.9%	F	0-59.9%
A	93.0-96.9%	B	83.0-86.9%	C	73.0-76.9%	D	63.0-66.9%		
A-	90.0-92.9%	B-	80.0-82.9%	C-	70.0-72.9%	D-	60.0-62.9%		

Course Policies

Late Work

Timely submission of the completed assignments is essential. The due date of each assignment will be stated clearly in the assignment description. Late assignments will be subject to a 10% penalty for each day it is late, up to a maximum of 50%. Late assignments will not accepted more than seven days after the due date.

Campus Policies

Please visit `http://www.ugst.umd.edu/courserelatedpolicies.html` for the Office of Undergraduate Studies' full list of course related policies and follow up with me if you have questions. It is our shared responsibility to abide by the University of Maryland's policies.

Academic Dishonesty

Cheating in any form (copying, falsifying signatures, plagiarism, etc.) will not be tolerated. It will result in a referral to the Office of Student Conduct irrespective of scope and circumstances, as required by university rules and regulations. There are severe consequences of academic misconduct, some of which are permanent and reflected on the student's transcript. If you have any questions regarding the University's policies on scholastic dishonesty, please see `http://osc.umd.edu/OSC/Default.aspx`. It is very important that you complete your own assignments, and do not share files (excluding raw data), partial work or final work.

University of Maryland Code of Academic Integrity

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit `http://shc.umd.edu/SHC/Default.aspx`.

Accomodations

Please come and see me as soon as possible if you think you might need any special accommodations for disabilities. In addition, please contact the Disability Support Services (301-314-7682 or `http://www.counseling.umd.edu/DSS/`). Disability Support Services will work with us to help create appropriate academic accommodations for any qualified students with disabilities. If you experience psychological distress during the course of the semester you can get professional help at the Counseling Center (301-314-7651 or `http://www.counseling.umd.edu/`)

Tentative Course Schedule

All readings, assignments, and due dates are subject to change. Refer to ELMS for the most current version of activities and due dates.

Week	Date	Topic	Readings	Assignments
1	1/25	**Introduction and Course Overview** *Introduction to JupyterHub*	Ch. 1	*Welcome survey*
	1/26	**Cause and Effect** *Experiments*	Ch. 2	*Assignment 1*
2	1/31	**Data and Variables** *Population and Sample*	Ch. 3	
	2/1	**Lab 1: Tables** *Table Methods*	Ch. 4	
	2/2	**Data Basics** *Exploring Tables*		
3	2/7	**Python Operations** *Sequences and Tables*	Ch. 5	
	2/8	**Lab 2: Working with Data** *Arrays and Tables*	Ch. 6	
	2/9	**Describing Data** *Categorical and Numerical Data*		*Assignment 2*
4	2/14	**Visualizations** *Types of Visualizations*	Ch. 7	**Assignment 1 due**
	2/15	**Lab 3: Visualizations** *Groups and Graphs*		
	2/16	**Functions** *Writing Reusable Code*	Ch. 8.0	
5	2/21	**Apply** *Using Functions with Tables*	Ch. 8.1	
	2/22	**Lab 4: Functions** *Applying Functions*		
	2/23	**Groups**	Ch. 8.2	*Assignment 3*
6	2/28	**Pivots and Joins** *Cross-Classifying Data*	Ch. 8.3-8.4	**Assignment 2 due**
	3/1	**Lab 5: Cross-Classifying** *Pivots and Joins*		
	3/2	**Exploring Data** *Visualizations with Pivots*	Ch. 8.5	*Midterm Project*

7	3/7	**Conditionals and Iteration** *Sampling*	Ch. 9.1-9.2	
	3/8	**Lab 6: Iteration and Simulation**		
	3/9	**Chance** *Comparing Data to Models*	Ch. 9.3-9.5	
8	3/14	**Sampling and Models**	Ch. 10	**Assignment 3 due**
	3/15	**Lab 7: Empirical Distributions**		
	3/16	**Distributions** *Models and Shuffling*		
	3/21- 3/23	**Spring Break (No Class)**		
9	3/28	**Hypothesis Testing**	Ch 11	
	3/29	**Lab 8: Hypothesis Testing**		
	3/30	**Hypothesis Testing** *Comparing Two Samples*		*Assignment 4*
10	4/4	**Hypothesis Testing** *Proportions and Means*	Ch. 12	**Midterm Project Due**
	4/5	**Lab 9: Hypothesis Testing, Part 2**		
	4/6	**Hypothesis Testing** *Examples*		
11	4/11	**Confidence Intervals** *Bootstrapping*	Ch. 13	
	4/12	**Lab 10: Confidence Intervals**		
	4/13	**Confidence Intervals** *Examples and Interpretation*	Ch. 16	*Assignment 5*
12	4/18	**Prediction Models** *K-Nearest Neighbors*	Ch. 17.0-17.1	**Assignment 4 due**
	4/19	**Lab 11: Prediction**		
	4/20	**Regression Models** *Numerical and Categorical Outcomes*	Ch. 15	
13	4/25	**Machine Learning** *Train/Test Split, Performance Measures*	Ch. 17.1-17.5	
	4/26	**Lab 12: Machine Learning, Part 1**		
	4/27	**Machine Learning Models** *Decision Trees*	See Canvas	*Lab 12*
14	5/2	**Machine Learning Models**	See Canvas	

	5/3	**Lab 13: Machine Learning, Part 2**		
	5/4	**Machine Learning** *Examples*		**Assignment 5 due**
15	5/9	**Bias and Fairness**	See Canvas	
	5/10	**Final Project Review and Feedback**		
	5/11	**Data Privacy and Ethics**		
		Finals Week: Final Project Due		