Course Syllabus
Analysis of Geologic Data (GLY 429/529)
University at Buffalo

Instructor: Dr. Yifan Cheng
Email: ycheng46@buffalo.edu
Office: Cooke 429
Lectures: Tuesday/Thursday 9:30AM-10:50AM**,** Hochstetter 430
Dr. Cheng Office Hours: Tuesday 1-1:30PM, Thursday 1-1:30PM, or by appointment

Course Description:

This course focuses on understanding statistical approaches to problems applied in the field of geology. Like most natural systems the assumptions made by techniques in introductory statistics are often violated and poor descriptors of a complex world. Further, geology has a historical aspect (we have only one history of Earth from which to draw data) which requires specific statistical approaches. Over the course of the semester we will explore probabilistic thinking, nonparametric resampling, time series analysis, and spatial statistics using real-world data. We will be performing these analyses in the Python programming environment and students will be taught how to write their own analyses over the course of the semester. In the process of applying statistical techniques we will also learn how to manipulate data inside Python and manage the import and export of that information.

Course Objectives and Learning Outcomes:

This course is intended to give students a broad base on knowledge in statistical theory and application through the Python programming environment. As this is a geology course we will be focuses on techniques and test datasets which are most likely to be encountered in that field. However, whether you intend to pursue a career in geology, science broadly, or something else entirely the general approaches in this course and skillsets are applicable to almost every conceivable field (one of my former advisors is an applied statistician and has worked on projects from the use of bite marks to identify suspects in criminal proceedings to shape change in fishes to risk management for a banking company). Similarly learning to program is an incredibly generalizable skill and once you know one programming language others are easier to pick up as well.

By the end of the course students should expect to fulfill the learning outcomes below.

1) Become familiar with nonparametric approaches to hypothesis-testing such as bootstrap, jackknife, and permutation
2) Be able to identify what statistical approaches are appropriate when presented with a new dataset and problem to explore/test
3) Develop fluency in the Python programming language for statistical analyses, but also best practices for producing reproducible code

Course Materials:

There are <u>no required textbooks</u> for the class, instead I will be drawing from across the literature and the department for real data to analyze (with citations). However, Python is an open programming language and as such

has a number of free online resources that may be helpful if you find yourself struggling with particular aspects of programming. The link to the MIT Open Course about Python Programming is below.

https://ocw.mit.edu/courses/6-0001-introduction-to-computer-science-and-programming-in-python-fall-2016/pages/syllabus/

Lectures slides will be made available on Course GitHub page the night before the lecture will be given under the "slides" folder.

All coding exercises and homework will be performed on Github CodeSpaces. Detailed instructions will be provided in Week 1&2 lectures.

Grading Policy: (see course schedule at the end of this document)

Final grades are letter-based (see here for more information on UB grade policy) and are a weighted average of attendance, quizzes, weekly exercises, and a final project over the course of the semester. The formula for calculating your grade is below. I use this weighted approach because it allows me to be flexible in altering the number or character of assignments if needed during the semester.

Overall percent grade = (attendance percent * 0.1) + (problem sets * 0.35) + (exams*0.30) + (final project * 0.25)

Percentage of weight for individual assessments

Weighting	Assessment / Assignment
10%	Attendance
35%	Weekly Problem Sets (3.5% each)
30%	Exams (15% each)
25%	Final Project

Letter grades, quality points and corresponding percentage ranges.

Grade	Quality Points	Percentage
A	4.0	93.0-100
A-	3.67	90.0 – 92.9
B+	3.33	87.0 – 89.9
B	3.00	84.0 – 86.9
B-	2.67	80.0 – 83.9
C+	2.33	77.0 – 79.9
C	2.00	74.0 – 76.9
C-	1.67	70.0 – 73.9
D+	1.33	67.0 – 69.9

Grade	Quality Points	Percentage
D	1.00	60.0 – 66.9
F	0	< 60.0

Attendance: Attendance will be taken each day as simply being present during class correlates to outcomes. However, as there are always life events which may prevent you from attending at least a few days each student is allowed **up to 4 days of unexcused absences** across the semester. This policy covers all kinds of absences (including illness). If you have extended issues that mean you are unable to make it to 24 total lectures you should contact me (Dr. Boyle).

Weekly Problem Sets: Each Thursday (except on exam days) an assignment will be released on UBLearns in that week's folder. The task will typically be writing your own Python code to analyze a dataset based on what we have covered that week in class. The assignments will be due a week later by the start of the following Thursday's class as commented code. GRADUATE STUDENTS ENROLLED IN THE COURSE WILL TYPICALLY HAVE A FEW ADDITIONAL QUESTIONS IN EACH PROBLEM SET FOCUSING ON EXPLORING TOPICS OF DATA EXPLORATION AND UNCERTAINTY.

Instructions for these assignments will be detailed at the start of the semester and progressively less detailed. As an example, let's say you are given a dataset of rock's mineral proportions across a volcanic field and your task is to identify two most common minerals across all samples and plot the variation in those elements. Early in the semester the instructions might tell you to 1) import the data as a csv file, 2) calculate the average proportion of each mineral, and 3) create a scatterplot using the plot function specifying the x and y variables as the elements with the largest averages. Later in the semester the instruction for the same task might just be "produce a plot showing the variation in the two elements which are, on average, the most common in the associated dataset". This means that early in the semester grading will be focused on demonstrating competence in coding practices and Python syntax and then shift toward demonstrating competence of problem-solving in applying appropriate statistical techniques.

Exams: The two exams will be released on February 26th and April 9th. These are take-home exams due a week later (March 5th and April 16th) through UBLearns. On the class days when the exams are released the whole period is dedicated to starting the exam and ask questions. The take-home portion of the exam will be similar to the format of the weekly problem sets but be a deeper dive into a single dataset.

Final Project: Upon returning from Spring Recess (i.e. March 23th) I will present an overview of several datasets to choose among as well as the requirements for each of them. Each student may choose from among any of the presented datasets. GRADUATE STUDENTS ENROLLED IN THE COURSE MAY ALSO PROPOSE THEIR OWN DATASET AND GOALS IF DESIRED WITH MY APPROVAL. There will be two checkpoints, all of which contribute to your project grade as shown in the table below. GRADUATE STUDENTS ENROLLED IN THE COURSE WILL PRESENT ON THEIR RESULTS ON MAY 5TH AS A ~15-MINUTE TALK. All submissions for the project are through UBLearns.

Important dates for the final project:

Checkpoint	Due Date	Project Grade	Outcome
Pick topic	March 26, 11:59PM	5%	Approval
Draft Section 1	April 23, 11:59PM	15%	Draft comments
Presentation (grads only)	May 5 Class	15%	Awesome Presentations
Project Due	May 12, 11:59PM	80%/65%	Final Project Grade

Bonus Exercises: For many students this course is their first college-level introduction to coding, statistics, or both. For both these subjects (but particularly coding) familiarity through repetition is the most reliable way to get the basics down so we can build on more tools through the semester. So, each week I will be releasing a series of simple exercises associated with the material we are covering at that time which are focused purely on the coding techniques or statistics, not an trying to use any of that to understand a dataset. These exercises will remain available through the end of exam week (May 8th). Each fully completed bonus exercise will be worth 0.5% towards your total grade up to a total of +5% if all ten bonus exercises are completed.

Make-up Policy:

The due dates for assignments are to provide structure and allow for quick feedback in the course. However, I am far less concerned with specific deadlines being met than with people engaging with the material. If you are unable to make a deadline for some reason (life happens) reach out and providing some kind of extension is usually possible, though it may mean fewer opportunities for discussion/revisions.

Academic Integrity:

Academic integrity is a fundamental university value. Through the honest completion of academic work, students not only advance their educational objectives, they sustain the integrity of the university and facilitate the transmission of knowledge and culture based upon the generation of new and innovative ideas. The Academic Integrity Policy provides additional information about what UB considers to be academic dishonesty and the possible consequences for violating UB's policies on academic integrity. In particular, you should be sure that you are aware of what UB considers to be academic dishonesty and that you understand how to avoid academic dishonesty. If you are unsure about the meaning of any of this information please talk to me or your academic advisor about them and we will try to clarify our expectations.

Use of AI and LLMs:

In the last year there has been an explosion of publicly available tools such as ChatGPT that use scraped information from the internet to generate content in a way that has not been accessible before. I personally have serious misgivings about the use of these for any kind of factual information summaries (as we will talk about in class computers are dumb in a particular way). The reasons my misgivings are many but mostly comes down to the fact that use of AI/LLM tools leads to less engagement and actual learning of material by students as the discovery and processing of information itself is incredibly important to problem-solving and retention of information. Second, these models do not produce truth, instead they are language-prediction algorithms which scrape the internet for text and then produce sentences that are similar to what it has found. It's just a more complex version of your smartphone's suggested next word for a text or the prompts for responses in Gmail. This means that these models often get basic facts correct, but details of specific topics are wrong inventing statements and citations that sound OK but are in fact complete fabrications. **However, one of the areas where these tools do appear to give more reliable answers is for coding. In this course we may explore the use of AI tools later in the semester but I ask that students refrain from using these tools for the first half of the semester as we are building up our coding skills in R**.

Accessibility Resources:

If you have any disability which requires reasonable accommodations to enable you to participate in this course, please contact the Office of Accessibility Resources, 60 Capen Hall, 645-2608, and also the instructor of this course. The Office of Accessibility Resources will provide you with information and review appropriate arrangements for reasonable accommodations.

Student Wellness:

As a student you may experience a range of issues that can cause barriers to learning or reduce your ability to participate in daily activities. These might include strained relationships, anxiety, high levels of stress, alcohol/drug problems, feeling down, health concerns, or unwanted sexual experiences. Counseling, Health Services, and Health Promotion are here to help with these or other issues you may experience. You learn can more about these programs and services by contacting:

Counseling Services:	120 Richmond Quad (North Campus), phone 716-645-2720 202 Michael Hall (South Campus), phone: 716-829-5800
Health Services:	Michael Hall (South Campus), phone: 716- 829-3316
Health Promotion:	114 Student Union (North Campus), phone: 716- 645-2837

Available Resources on Sexual Assault:

UB is committed to providing an environment free of all forms of discrimination and sexual harassment, including sexual assault, domestic and dating violence and stalking. You may call UB's Office of Equity, Diversity and Inclusion at (716) 645-2266 for more information or visit their website.

Visa Concerns:

For international students, please see the information provided by the International Student Services. If you have any questions, they have drop-in advising hours.

Campus Safety:

If you re the victim of a crime (including sexual assault) or witnessed a crime, report it here. You can call campus police at 716-645-2222 (emergencies). You can use this number for on-campus incidents, including possible drug/alcohol overdose.

Bias-Related Incidents:

Please report any bias-related incidents here. You can also contact the Office of Equity, Diversity, and Inclusion, located at 406 Capen Hall, 716-645-2266

Food Insecurity:

Access to food is a basic human right. Outside my office (Cooke 457), I have granola bars, and menstruation products, available for free. There is also an on-campus virtual food pantry available for free for all students, called Blue Table.

Bathroom Locator:

Looking for a single-use (gender-neutral) bathroom? Check out this list. Note: UB community members are entitled to use restroom facilities that are consistent with their gender identity.

Name Change:

Please click here to learn how to change your preferred name.

Childcare Services:

If your childcare plans fall through, you are more than welcome to bring them to class. Also, check out UB's childcare center.

Emergency Funds:

Some students facing hardships may be eligible for emergency funds. See the UB Emergency Fund page for more details.

If you find yourself struggling with course-related issues, or any other issues regardless of the reason, please don't hesitate to contact me so I can help you resolve the difficulty or direct you to some other resource who can.

Course schedules:

Dates	Week	Tuesday	Thursday
1/19-1/23	1		A
1/26-1/30	2	B	C
2/3-2/6	3	D	E
2/10-2/13	4	F	G
2/16-2/20	5	H	I
2/23-2/27	6	J	K
3/2-3/6	7	L	M
3/8-3/13	8	N	O
3/15-3/20	9	Spring Recess	
3/22-3/27	10	P	Q
3/29-4/3	11	R	S
4/6-4/10	12	T	U
4/13-4/17	13	V	W
4/20-4/24	14	X	Y
4/27-5/1	15	Z	AA
5/4-5/5	16	AB	

Key for sequence of lecture topics. Blue is the material covered up to the first exam (stats theory and basic coding), green material up to the second exam (practical application), and purple class time for the final project.

Course contents (subject to changes):

ID	Content
A	Introduction (get Python working)
B	Crash course in Python
C	Statistics I (central tendency, matrices, & vectors)
D	Statistics II (confidence, plotting with errors)
E	Probabilistic Thinking (beyond p-value, Bayes theorem)
F	Distributions (normal, uniform, Poisson, bimodal)
G	Complex Python syntax (loops, if else, packages)
H	Resampling techniques (bootstrap, jackknife, Monte Carlo)
I	Correlation, GLMs, & ANOVA I
J	Uncommon distributions and model choice
K	**Exam I**
L	Intro to Time Series
M	Time Series II (detecting signal)
N	Time Series III (Loess & Moving Average)
O	Reducing Dimensionality I
P	Final Project Introduction
Q	Reducing Dimensionality II
R	Maps & Spatial Data
S	Spatial Statistics I

ID	Content
T	Spatial Statistics II
U	**Exam II**
V	Final Project Work
W	Final Project Work
X	Final Project Work
Y	Final Project Work
Z	Final Project Work
AA	Final Project Work
AB	Final Project Presentation