GLY429/529 – Analysis of Geologic Data

Instructor Name: Dr. James Boyle
Office: Cooke 453
Student Hours: Tu/F 11AM-12PM or by appointment (over zoom available as well by request)
Email: jamesboy@buffalo.edu

Course Location: Hochstetter 430
Course Time: Tu/Th 9:30-10:50AM

COVID-19 continues to be a threat in our community with the current variant XBB.1.5 (aka 'Kraken') being highly contagious. Please follow sensible health precautions, including masking in order to make the classroom more inclusive to all, especially individuals who have, live with, or care for those with compromised immune systems. Everybody must follow the UB and New York State health guidelines and remember that those rules are to protect others as well as yourself.

Public Health Compliance in Classroom setting:

The most current health guidelines can be found here (https://www.buffalo.edu/coronavirus/health-and-safety/health-safety-guidelines.html). Completion of the primary vaccination sequence is **required** for all students and those found not to be in compliance will be blocked from registering in future terms.

If you are unable to attend class you should contact the professor (Dr. Boyle) about accommodations to complete any work you might miss.

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 R 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 R and manage the import and export of that information.

Student Learning Outcomes:

1. Be able to identify what statistical approaches are appropriate when presented with a new dataset and problem to explore/test
2. Develop fluency in the R programming language for statistical analyses, but also best practices for producing reproducible code

Course Materials:

Materials: There are no required textbooks for the class, instead I will be drawing from across the literature and the department for real data to analyze (with citations). However, R is an open programming language and as such has a number of free online "textbooks" that may be helpful if you find yourself struggling with particular aspects of programming. The links to two such resources are found below.

https://cran.r-project.org/doc/manuals/R-intro.pdf
https://r4ds.had.co.nz/

Lecture: Lectures slides will be made available on UBLearns the night before the lecture will be given under the "Course Documents/Lectures" folder.

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, discussion board assignments, and exams throughout the semester. There is no final exam during final exam week at the end of the semester.

Learning assessments will be graded based on rubric criteria and weighted according to the following break-down.

Weighting	Assessment / Assignment
10%	Attendance
35%	Weekly Problem Sets (3.5% each)
40%	Exams (3.33% each)
15%	Final Project

Attendance: Attendance will be taken each day. As there are always life events which may prevent you from attending at least a few days each semester you will receive the full 10% portion of the attendance grade if you attend 21 lectures (=75% of all classes excluding the drop/add period and exam days). For every lecture missed after that you will lose 1% of your total grade to a maximum loss of 10%. This policy covers all kinds of absences (including illness). If you have extended issues that mean you are unable to make it to 21 total lectures you should contact me (Dr. Boyle).

Weekly Problem Sets: Each Thursday (except on exam days) an assignment will be released on UBLearns under the "Assignments" tab. The task will typically be writing your own R code to analyze a dataset based on what we have covered that week in class. The instructions for these assignments will be detailed at the start of the semester and progressively less detailed (similar to the technical challenge of the Great British Bakeoff). 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, and 4) export the scatterplot using the pdf() function. 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 R syntax and then shift toward demonstrating competence of problem-solving in applying appropriate statistical techniques. The problem sets (set of R code, outputs and plots) will be due by the start of following Tuesday's class. The tenth "problem set" is being present for both presentation days (May 9th and May 11th) on the last week of the semester.

Exams: The two exams will include both an in-class and a take-home component. The in-class portions will take place on March 9th and April 20th. The take-home portion will be released those days and due through UBLearns a week later (March 16th and April 27th) by 11:59PM. The in-class portion will be focused on how you would approach solving a particular data analysis issue and annotating or interpreting code. This section will be due at the end of the class period. 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: The last three weeks of the course will be dedicated to preparing and, for the graduate students, presenting analyses of datasets. Undergraduate students will be given a choice of datasets to work with and a general set of hypotheses to test and explore. For undergraduates the final product will be a written report describing the dataset, analyses, and results (including plots) as well as the R code they have written in association with the project.

For graduate students, they are expected to use data relevant to their own project or field of interest for analyses. The proposed goals of the final project must be discussed with and approved by me (Dr. Boyle) prior to April 25th. In addition to the written report graduate students will also give a presentation, in the style of a conference talk (~15 minutes), to the class on either May 9th or May 11th.

.

Make-up Policy:

If you are unable to complete an assignment before the due date due to unavoidable circumstances (ex. car accident) email me (jamesboy@buffalo.edu) when you can safely do so (i.e. do not worry about emailing me the day of an exam if you are in a car accident and have to go the hospital).

Final Grades:

Grade	Quality Points	Percentage
A	4.0	94.0+
A-	3.67	90 – 93
B+	3.33	87 - 89
B	3.00	84 – 86
B-	2.67	80 - 83
C+	2.33	77 – 79
C	2.00	74 – 76
C-	1.67	70 - 73
D+	1.33	67 – 69
D	1.00	65 - 66
F	0	< 65

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 Undergraduate 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.

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.

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 as soon as possible. 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:	4350 Maple Road, phone: 716- 829-3316
Health Promotion:	114 Student Union (North Campus), phone: 716- 645-2837

Course Schedule

Dates	Week	Tuesday	Thursday
1/30-2/3	1	A	B
2/6-2/10	2	C	D
2/13-2/17	3	E	F
2/20-2/24	4	G	H
2/27-3/3	5	I	J
3/6-3/10	6	K	L
3/13-3/17	7	M	N
3/20-3/24	8	**Spring Recess**	
3/27-3/31	9	O	P
4/3-4/7	10	Q	R
4/10-4/14	11	S	T
4/17-4/21	12	U	V
4/24-4/28	13	W	X
5/1-5/5	14	Y	Z
5/8-5/12	15	AA	BB

Key for sequence of lecture topics. Blue is the material covered up to the first exam, green material up to the second exam, and purple class time for the final project.

A	Introduction (course & get R working)
B	Excel, importing data, and R syntax
C	Basic Statistics I (central tendency, matrices & vectors)
D	Basic Statistics II (confidence, plotting with error)
E	Probability Thinking (beyond p-value, bayes theorem)
F	Distributions (normal, uniform, poisson, bimodal)
G	Complex R syntax (loops, ifelse, packages)
H	Resampling techniques (bootstrap, jackknife, Monte Carlo)
I	Correlation, GLMs, & ANOVA I
J	Correlation, GLMs, & ANOVA II
K	Complex Distributions and model choice
L	**Exam I (in-person portion)**
M	Intro to Time Series
N	Loess & Moving Average
O	Stationary Time Series
P	Spectral & Wavelet Analysis
Q	Importing and plotting maps in R
R	Manipulating spatial data in R
S	Spatial statistics
T	Spatial interpolation I
U	Spatial interpolation II
V	**Exam II (in-person portion)**
W	Final Project Work
X	Final Project Work
Y	Final Project Work
Z	Final Project Work
AA	Final Project Presentation
BB	Final Project Presentation