TEXAS A&M UNIVERSITY

Geological Data Analysis (Gr)

Course title and number	GEOL 489/689 Geological Data Analysis
Term	Fall 2019
Meeting times and location	Monday/Wednesday at 11:00 AM – 12:30 PM, Room TBA

Course Description and Prerequisites

Description: The course covers scientific programming and statistical methods commonly used in geology and paleontology. Course material will involve applying univariate and multivariate statistical analyses to geological data, and writing short programming scripts for R. Students will gain practical expertise in evaluating statistical approaches and solving methodological obstacles.
Prerequisite: None

Learning Outcomes

Graduate students who complete this course should be able to:

- Manipulate relevant datasets of environmental, historical and spatial observations in a programmatic environment, including loading, cleaning and transforming the data.
- Combine functions for loading, manipulating and analyzing datasets into a project workflow.
- Troubleshoot error messages and other problems with their code.
- Identify appropriate analyses for a given question or hypothesis, with a particular dataset.
- Apply familiar statistical tests to new datasets.
- Teach themselves an unfamiliar statistical test to apply to their own data.
- Interpret statistical results in terms of a given study system, and suggest possible future directions.
- Intuit possible biases or issues with an analytical approach.
- Devise novel analytical approaches when necessary, using statistical theory, simulations, etc.

Instructor Information

Name	Dr. David Bapst
Email address	dwbapst@tamu.edu
Office hours	TBA
Office location	Halbouty 169

Necessary Equipment and Resource Materials

Students need their own computer, with R, RStudio, and a spreadsheet program installed. We will cover installation of R and RStudio in class. Although I do not recommend this as a long-term solution, students and staff may borrow laptops for four hours at a time from the TAMU libraries: http://library.tamu.edu/services/media_services/equipment.html#laptops

This class has no mandatory textbook, but some optional suggestions for reference are:

Davis, 2003. Statistics and Data Analysis in Geology. Detailed discussions of many methods we'll apply, particularly multivariate methods, with application to geology, despite some aged content.

McKillup and Dyar, 2010. Geostatistics Explained. Clear explanations on how probability and statistics work, and about geospatial and temporal analyses specialized to geosciences.

Hammer and Harper, 2006. Paleontological Data Analysis. A 'cookbook' reference to data analysis methods, with simplified, introductory summaries on many methods, but without explanation of how methods work. This book is very specialized for paleontology, but will be useful to others as well.

Course Calendar

		IN CLASS	ASSIGNMENTS
WEEK 1	Aug 26	Course Introduction, Questions and Methods in Science	
	Aug 28	Structure of Geological Datasets Using the R Terminal & RStudio	Install R and RStudio on laptop before class
WEEK 2	Sept 2	Object Types: Matrices, Arrays, Lists, Factors Reading Data Files into R	**Grad Student Project Ideas Due** **Assignment Begins:** Load This File
	Sept 4	Logicals, If-Else Controls and Subsetting Datasets	**Assignment Due:** Load This File
WEEK 3	Sept 9	Writing Loops, Function and Scripts	**Assignment Begins:** Ostracod Size Differences
	Sept 11	**Project Pitches (Graduate Students Present)** Writing an R Routine to Clean a Dataset	**Due: Project Pitches from Grad Students**
WEEK 4	Sept 16	Sampling, Betting Odds, Venn Diagrams, and Useful Probability Distributions	**Assignment Due:** Ostracod Size Differences
	Sept 18	Observables: Visualizing and Summarizing Datasets Plotting and Summary Statistics	**Assignment Begins:** Modelling Earthquakes with Waiting Time Distributions
WEEK 5	Sept 23	Deterministic Models, Stochastic Simulations and The Structure of Frequentist Tests	
	Sept 25	Bootstrapping, Jack-knifing, Rarefaction for Dealing with Odd Distributions and Weird Summary Statistics in Geology	**Assignment Due:** Modelling Earthquakes with Waiting Time Distributions
WEEK 6	Sept 30	Modelling Incomplete Sampling of Sedimentary and Historical Records	**Assignment Begins:** Sampling and Extinction in the Durations of Fossil Plankton
	Oct 2	Parametric versus Non-Parametric Statistics: Mann-Whitney Tests and Others	
WEEK 7	Oct 7	Planning Analyses in Historical Sciences and Headaches: Multiple Comparisons, Spurious Correlations, Non-Independence	**Assignment Due:** Sampling and Extinction in the Durations of Fossil Plankton
	Oct 9	Statistical Tests for Univariate Datasets: T-tests, F-Tests, ANOVA (1-Way and 2-Way)	**Assignment Begins:** Disentangling Measurement Methods in Digital Field Geology
WEEK 8	Oct 14	Tests for Discrete Data: Binomial test, Chi-Squared Test	
	Oct 16	Analyzing Bivariate Data: Correlations and Linear Regressions	**Assignment Due:** Disentangling Measurement Methods in Digital Field Geology
WEEK 9	Oct 21	Looking Under the Hood: Regression in a Spreadsheet	**Assignment Begins:** Changes in Grain Size Distributions
	Oct 23	Multiple Regressions for Analyzing Environmental Variables: Forecasting and Predictions	
WEEK 10	Oct 28	**Project Presentations – Progress Reports on Data & Methods (All Students)**	**Assignment Due:** Changes in Grain Size Distributions **Progress Report Presentations**
	Oct 30	Frequentist, Maximum Likelihood and other Inferential Schools of Thought – Case Study in Sedimentation Rate Modeling	**Assignment Begins:** Predicting Groundwater Quality from Geochemical Data
WEEK 11	Nov 4	Information Criteria, Model Selection Methods, and Stepwise Regression Models in Multiple Regression	**Due: Draft Introduction and Methods Sections**
	Nov 6	Decomposition of Complex Datasets with Principle Components Analysis	**Assignment Due:** Predicting Groundwater Quality from Geochemical Data
WEEK 12	Nov 11	Irregular Time-Series, Trends and Autocorrelation Thru Geologic Time; Correlation and Causality in Phanerozoic Nutrients	**Assignment Begins:** Analyzing Ammonite Shell Septa with PCA and Spectral Analysis
	Nov 13	Finding Patterns in Paleoclimate with Spectral Analysis, Eccentricity and Mass Extinctions	
WEEK 13	Nov 18	Troubleshooting Projects – Discussion, Open Help	**Assignment Due:** Analyzing Ammonite Shell Septa with PCA and Spectral Analysis
	Nov 20	The Use of Morphometrics in Geology: Maps, Trilobites, and Sediment Grains	**Due: Project Paper Draft**
WEEK 14	Nov 25	**Final Project Presentations**	**Due: Final Project Presentations**
FINAL TERM PAPER DUE LAST DAY OF CLASSES, DECEMBER 3RD			

Assignments and Grading Policies

The final course grade will be based upon:

Project Paper	**40%**
Assignment Reports	**20%**
Intro + Methods & Full Draft	**15%**
Project Presentations	**10%**
Submit Project Pitch	**5%**
Participation	**10%**

Grades in this class will be based on in-class presentations, participation, a series of individual assignments (which must be submitted as individual reports, but students can work on them as a group) and a final project. See the next page for more details on the grading of this paper and related drafts.

Students will give an in-class progress report on their final project, covering their question, datasets and research methods mid-semester, and give a final presentation on the last day of class that summarizes their findings and interpretations. These presentations should be short, maybe 10-15 minutes in length, and will followed by taking questions from the audience. A presentation may involve projector slides, which should be supplied to the instructor before class, or the whiteboard can be used instead for any necessary visuals. Students in the audience will also write comments and feedback on the talks they see, but those responses will not the basis for student grades these presentations.

The mid-semester presentation should cover the motivating research question, and describe the data and statistical approach to be used. Students should try to be as specific as possible on these details, to the degree that they know at that time. Current questions or concerns about how they will analyze their data should be discussed as part the presentation, as well as explaining any difficulties encountered.

Students are also expected to participate in live, in-class coding exercises throughout the semester, as well as in-class discussions regarding lecture material and their proposed final project topic, as well as communicating problems they've encountered (technical, theoretical, etc). Participation in live-coding activities and in-class discussions will be graded.

Grading Scale and Attendance

Standard Letter Grading Scale: A = 90-100%, B = 80-89%, C = 70-79%, D = 60-69%, F = <60%

Attendance is mandatory with allowances for excused absences with proper documentation in accordance with http://student-rules.tamu.edu/rule07.

Americans with Disabilities Act (ADA)

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you have a disability requiring an accommodation, please contact Disability Services, currently located in the Disability Services building at the Student Services at White Creek complex on west campus or call 979-845-1637. For additional information, visit http://disability.tamu.edu.

Reporting Requirements for Instructors Under Title IX

All TAMU employees are mandatory reporters of sexual harassment, per Texas A&M System Regulation 08.01.01. Instructors cannot maintain confidentiality regarding reports of sexual harassment, due to their obligation to report certain issues that jeopardize the health and safety of our community. If you would like to discuss events in a more confidential setting, please visit Counseling and Psychological Services: http://scs.tamu.edu

Academic Integrity

For additional information please visit: http://aggiehonor.tamu.edu

"An Aggie does not lie, cheat, or steal, or tolerate those who do."

Term Project

This class requires a project-based term paper, with undergraduates organized into groups by the instructor, with a question and dataset provided by consultation, while graduate students are required to propose their own research question within the earth sciences, with undergrads potentially joining a graduate student's working group. Graduate students are also required to locate and obtain an appropriate dataset --- perhaps from a publicly-available source, such as supplemental material for an article, or collected as part of their own or their advisor's research. Datasets must be large enough and detailed enough to address a worthwhile research question. Addressing these research questions will require applying various analyses from throughout the course, and possibly beyond the covered course material.

Graduate students must submit two short paragraph-long descriptions of potential research projects they might do for their term paper during Week 2, followed by short pitches to the class at large on their chosen project idea during Week 3, potentially recruiting undergraduates to join their project group. Graduate students are expected to provide mentoring and leadership to those undergraduates who join their research team. Undergraduates who do not join a graduate student's group will instead be grouped together by the course instructor, and a range of potential projects suggested.

Interim Drafts of the term paper are due at two points during the semester. The first draft due is expected to contain relatively detailed Introduction and Methods sections, with proper in-text citations and formatted references, including an appendix of current data, while the second draft due should be relatively complete (see below). Each draft will be graded relative to the final paper rubric (attached), and grades for complete drafts (as expected for the second draft) will be locked in as 'minimum' grades, upon which students can only improve on successive submissions (including the final submission).

Expected Term Paper Structure

The length of your paper depends on how much you have to say. Each member of a final project team will need to submit a separate term paper, formatted like a scholarly manuscript, covering the methods, and results of their project. A complete paper draft will have the following required sections:

- **Abstract (REQUIRED)**
  - A concise, 1 paragraph summary of your study (e.g. questions, data, methods, results).
- **Introduction (REQUIRED)**
  - Provide scientific context for your research question and background a reader would need to understand your study system and the importance of answering your research question. State your questions clearly and introduce how you will investigate them.
- **Methods (REQUIRED)**
  - A complete description of the data and analyses you use, with justifications for why you chose those specific methods. Include references that support your decisions.
- **Results (REQUIRED)**
  - A report of the results of <u>your</u> analyses, including reporting all necessary statistical values, and necessary supporting figures and tables. Interpretation of results should be minimized.
- **Discussion (REQUIRED)**
  - An interpretation of your results in the context of the hypothesis you set out to test with reference to the published results of other workers. Address any analytical challenges, alternative interpretations, related your findings to the work of others, implications of your results to the field, and suggestions for future work that could improve your ability to answer your research questions.
- **Conclusions (OPTIONAL)**
  - A concise summary of the "take-home" results, interpretations, and implications of your research. Similar to the abstract, but does not restate the study's setup.
- **References (REQUIRED)**
  - Choose a consistent journal format. If uncertain which to use, use Paleobiology style.
- **Data Appendix (REQUIRED)**
  - A file or table that contains all data you used for your paper. You must submit a data appendix with your Intro + Methods draft, but you should update this appendix as you revise or add to your data over the course.