ESM 244: Advanced Environmental Data Analysis

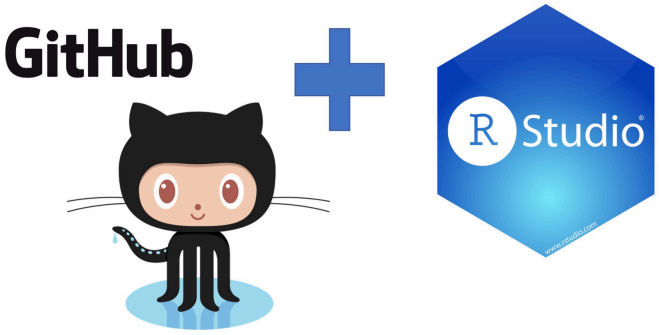
Bren School of Environmental Science & Management

Lectures: Tuesdays and Thursdays, 2:00pm - 3:15pm, BH 1414
Labs: Thursdays, 3:30pm - 4:45pm, BH 1414

Course forum: Slack (sign up here)
Assignments, resources, grading: Canvas

Instructor: Nathaniel Grimes (nggrimes@ucsb.edu)
Office Hours: Wednesday 1:00 - 3:00 pm Pine Room (Bren 3526)
Teaching Assistant: Yutian Fang (yutianfang@bren.ucsb.edu)
Office Hours: Monday 10am - 12:00 pm MSI 1208

Overview: ESM 244 is a survey course in advanced topics in environmental data analysis, including (but not limited to): basics of machine learning, logistic regression, bootstrapping, intro to wrangling and analyzing time series data, spatial data visualization and analysis, principal components analysis, hierarchical cluster analysis, and basic text mining. Focus is on building conceptual understanding and applied skills using real-world environmental datasets. Students will also learn modern methods for publication in data science, such as Shiny apps in R. Throughout, we will reinforce skills for data management and organization, reproducible workflows and collaboration in R, RStudio and GitHub.



ESM 244 LEARNING OBJECTIVES: Students will be able to...

- Work with novel data formats including spatial data, time series data, and text data
- Apply advanced methods for analyzing environmental data including logistic regression, clustering techniques, and nonlinear least squares
- Implement functional programming in R, including loops, functions, and iteration, to more proficiently wrangle, reshape, and visualize data
- Apply principles and best practices for data organization, conventions and code habits, project management, and reproducible workflows
- Leverage R Studio + Github to communicate analyses using Shiny (and optionally distill, gh-pages)
- Confidently debug and troubleshoot code errors
- Independently find and implement new data science skills to complete assignments and term projects

SCHEDULE (TENTATIVE):

Week	Lecture topics	Lab skills	Assignments & Projects
1	Course overview ESM 206 review Intro to supervised vs. unsupervised machine learning	Git/GitHub intro and setup with R Studio Intro to Quarto (video) Refresh your workflow, wrangling, and ggplot basics	
2	Model selection, test and training data, cross validation Intro to apps with Shiny	Cross validation to compare competing models Functions, loops, and purrr Shiny continued (video/tutorial)	
3	Binary logistic regression	Generalized linear models, binary logistic regression Cross validation with purrr in overdrive Build your first Shiny app (video/tutorial)	
4	Multivariate data, unsupervised machine learning Principal component analysis (PCA)	Multivariate exploration, PCA, biplots Git collaboration and merge conflicts More Shiny basics (video/tutorial)	Assignment 1 due Shiny app: group and topic
5	Partition-based (k-means) and hierarchical cluster analysis Intro to AI in coding	K-means clustering, hierarchical clustering, dendrograms Shiny continued (video/tutorial) Copilot introduction	
6	Random Forests	Cross Validation and Tidymodels	Assignment 2 due Shiny app: basic structure/theme
7	Introduction to spatial data (types, projections, CRS, applications) Spatial data interpolation (kriging)	Working with vector spatial data, visualizing spatial data, variograms and kriging (optional: self-guided raster lab)	
8	Spatial point pattern analysis (quadrat, nearest neighbor, test for CSR) Introduction to text mining and analysis workflows, sentiment analysis	Spatial point pattern analysis Working with text data, parsing documents to get tokens; stop words and sentiment analysis	Assignment 3 due Shiny app: mostly functional
9	Intro to time series: decomposition, autocorrelation Time series forecasting basics & concerns	Forecasting time series Wrangling time series	
10	Automating data from the internet using APIs;	Accessing data through APIs	Assignment 4 due (finals week) Present Shiny Apps to class

COMMUNICATION AND LAB REQUIREMENTS:

In general, course communications will be conducted through **Slack** - Join the ESM 244 W2025 Slack workspace here! This will be the easiest way to quickly contact (and be contacted by) the instructor and TA, and will be very helpful for group work and collaboration throughout the course.

All labs will be in person, in BH 1414, unless unforeseen circumstances strike (e.g., pandemic, fire, alien invasion). You will need to bring your laptop to BH1414 to follow along each week. You will need:

- Recent versions of **R (v. 4.3.0 or higher)** and **RStudio (at least v2024.x)**. *If you do **not** have a laptop that will happily run this software, I encourage you to contact the Bren Compute Team (compute@bren.ucsb.edu) to get set up with an RStudio Server account.*
    - Click here for instructions on installing R and RStudio.
- A **GitHub account**, with **Git** configured for working between RStudio & GitHub, including a valid Personal Access Token (PAT).
    - Click here for instructions on setting up Git and GitHub.
    - Click here for instructions on configuring your Git and setting up a Personal Access Token (PAT). We will walk you through this at the start of the first lab session.

All lab materials (data, keys, links, etc.) will be linked from Canvas.

COURSE RESOURCES:

There is **no textbook** for ESM 244. The resources we'll use in class are free, open-source books and online tutorials. Several essentials for this class are:

- R for Data Science by Garrett Grolemund and Hadley Wickham
- R Markdown: The Definitive Guide by Yihui Xie, J. J. Allaire, Garrett Grolemund
- Geocomputation with R by Robin Lovelace
- Forecasting: principles and practice by Rob Hyndman and George Athanasopoulos
- Tidy text mining in R by Julia Silge and David Robinson
- Happy git with R by Jenny Bryan and UBC's stat545 TAs
- Shiny tutorials from RStudio

GRADING BREAKDOWN:

- Assignments (4): 60%
- Participation: 10%
- Term project (functional Shiny app): 30%

ASSIGNMENTS:

Assignments will be completed in Quarto, and you will submit the .Qmd and knitted HTML files for grading throughCanvas. Some assignments may ask you to respond to a reading or resource on the Slack **#discussions** channel.

TERM PROJECT:

See guidelines for term projects HERE.

For your ESM 244 term project, you will design a Shiny app, created in groups of 2 - 3 people, that you'll present to the class during Week 10. Over the course of the quarter, each assignment will have a task related to the term project to help you stay on track.

CODE OF CONDUCT:

All students are expected to read and comply with the UCSB and ESM 244 course code of conduct.

From our ESM 244 code of conduct: All enrolled students, auditors, and course visitors are expected to comply with the following code of conduct. We expect cooperation from all members to help ensure a safe and welcoming environment for everybody.

We are determined to make our courses welcoming, inclusive and harassment-free for everyone regardless of gender, gender identity and expression, race, age, sexual orientation, disability, physical appearance, body size, or religion (or lack thereof). We do not tolerate harassment of class participants, teaching assistants, or instructors in any form. Derogatory, abusive, demeaning or sexual language and imagery is not appropriate or acceptable.

DISABLED STUDENTS:

Students with disabilities and alternative learning needs are encouraged to work with the Disabled Students Program at UCSB to ensure we can best accommodate you.

COURSE POLICIES:

- Assignments submitted late will only be accepted within one week of the due date, and will be worth 50% of the original score. **Homework submitted more than one week after the original due date will not be accepted.** We always welcome individual students to reach out to us for extra support and to discuss options as challenges arise.
- Assignments are due at the beginning of lecture on due dates.
- Cheating/plagiarism (including in R code) will result in 0 points awarded for the assignment or midterm/quiz/exam and disciplinary action according to UCSB policy. We **do** expect that you will pull from examples and tutorials by the R community but do not simply copy-and-paste - think through the examples and tutorials and adapt the code to your current context.
- If you are worried about your overall grade and/or passing the class, please reach out to us to discuss your concerns.

AI POLICY:

Using AI will be another arrow in the skill quiver of future data scientists. In this class we will teach students how to use GitHub CoPilot, an AI tool specifically dedicated to improve the coding process. CoPilot will make students faster at coding, but it can disrupt the learning process. Students must turn off CoPilot during labs and lectures to think critically and develop their skills. Students are free to use code directly prompted from CoPilot, and to seek input from other AI services such as ChatGPT or Gemini on homework assignments.
Class motto: Always bring your brain to the party!