

ESM 244: Advanced Environmental Data Analysis

Bren School of Environmental Science & Management

Live weekly check-in: 30 minutes (time TBD)

Live weekly discussions: Each student will attend a 50-minute live discussion weekly (times TBD)

Lectures: Pre-recorded (~75 - 90 minutes total each week)

Labs: Pre-recorded (~60 minutes each week)

Course forum: Slack (you should have received an email invite))

Instructor: Allison Horst (ahorst@bren.ucsb.edu)

Office Hours: TBD

Teaching Assistant: Casey O'Hara (cohara@bren.ucsb.edu)

Office Hours: TBD

Overview: ESM 244 is a survey course in advanced topics in environmental data analysis, including: 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, including by building a website and Shiny app 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:

- Building on statistics and regression basics from ESM 206, learn advanced methods for analyzing environmental data including logistic regression, time series data, and spatial analysis
- Grow skills for data wrangling, reshaping, and visualization with functional programming in R
- Continue building principles for data organization, conventions and code habits, project management, and reproducible workflows
- Develop skills for communication tools in R (including distill, gh-pages and Shiny)
- Increase confidence and skills for troubleshooting code errors
- Independently find, learn and implement new skills to complete assignments and term projects

SCHEDULE (TENTATIVE & LIKELY TO CHANGE):

Week	Lecture topics	Lab skills
1	ESM 206 review, multivariate data, principal components analysis (PCA) intro	Refresh your workflow, wrangling, and dataviz skills in R, meet blogdown
2	PCA continued, intro to apps with Shiny	Multivariate exploration, PCA, biplots, build your first Shiny app
3	Logistic regression (binary & ordinal)	Generalized linear models, binary logistic regression, Shiny continued
4	Working with time series data, exploration & decomposition, autocorrelation, forecasting basics & concerns	Wrangling time series with <i>lubridate</i> , <i>feasts</i> , and <i>fable</i> , tsibbles, moving average, and a forecasting example
5	Introduction to spatial data (types, projections, CRS, applications)	Working with spatial points, lines and polygons with <i>sf</i> , visualizing spatial data with <i>ggplot</i> and <i>tmap</i>
6	Spatial data interpolation (kriging) and point pattern analysis (quadrat, nearest neighbor, test for CSR)	Raster example (<i>raster</i>), variograms and kriging in R, making web pages with gh-pages in GitHub
7	Intro to partition-based (k-means) and hierarchical cluster analysis	Spatial point pattern analysis, k-means, hierarchical cluster analysis & dendrograms
8	Bayes theorem & Bayesian thinking / case studies, introduction to text mining and analysis workflows, sentiment analysis	Working with text data, parsing documents to get tokens, stemming, stop words, and sentiment analysis
9	Nonlinear models, bootstrapping	Nonlinear least squares, bootstrapping confidence intervals
10	Becoming a versatile data scientist in other languages and platforms	Comparing R with Python (<i>reticulate</i>), querying SQL databases in RStudio

LAB REQUIREMENTS:

Labs will be pre-recorded. You are required to follow along with labs each week.

- Recent versions of R (v. 3.6.1 or higher) and RStudio (v. 1.2 or higher). *If you do **not** have a laptop that will happily run this software, I encourage you to contact the Bren Compute Team (request@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. **Note:** if you are using a different Git client, that's fine, but we might not be able to help you if you get stuck. [Click here](#) for instructions on installing and configuring Git.

All lab materials (data, keys, links, etc.) will be shared on the course website.

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
- [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): 40%
- Slack engagement: 5%
- Term project 1 (personal data science website): 25%
- Term project 2 (functional Shiny app): 30%

ASSIGNMENTS:

Assignments will be completed in R Markdown, and you will submit the .Rmd and knitted HTML files for grading through GauchoSpace or GitHub

TERM PROJECTS: See guidelines for term projects [HERE](#)

For ESM 244 you will complete two term projects:

1. A personal website/blog highlighting 3 data science projects (made in R)
2. A Shiny app, created in groups of 1 - 3, that you'll present to the class during Week 10

CODE OF CONDUCT:

All students are expected to read and comply with the [UCSB](#) and [206/244 course code of conduct](#).

From our ESM 206/244 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
- If you are worried about your overall grade and/or passing the class, please reach out to us to discuss your concerns