ESM 206: Introduction to Statistics and Data Analysis in Environmental Science and Management (Fall 2020)

Bren School of Environmental Science & Management

Instructor: Allison Horst
Email: ahorst@ucsb.edu
Remote Open Hours: Mondays 3:00 - 4:30 PST (Zoom link)

TA: Jessica Couture
Email: jcouture@bren.ucsb.edu
Remote Open Hours: TBD

TA: Juan Carlos Villaseñor-Derbez
Email: juancarlos@ucsb.edu
Remote Open Hours: TBD

Overview: In ESM 206, you will build conceptual, technical and communication skills to investigate and answer environmental questions using data. Topics include: data wrangling, exploratory data analysis, descriptive statistics, hypothesis testing, describing uncertainty, and data visualization. Skills for data exploration, analysis and project management will be developed through analysis of real-world datasets using R, RStudio and GitHub, while building tools and mindsets for computational reproducibility and collaboration.

COURSE COMPONENTS:

All course materials will be posted to GauchoSpace (and to allisonhorst.github.io for public use & access). Enrolled students should refer to GauchoSpace as the primary course resource.

ESM 206 has 6 components that you will interact with and complete weekly:

Weekly learning is through:

- **Lecture videos (pre-recorded):** ~5 videos / wk, 10 - 15 min each
- **Computer labs (pre-recorded):** Recorded code-along lessons (3 - 4 /wk, 15 - 20 min each)
- **Reading (posted):** Read posted material before your discussion section each week

Weekly practice and assessment is through:

- **Assignments:** Posted Wednesday each week, due the following Wednesday

Weekly synchronous engagement & participation is through:

- **Discussion (live):** Attend one 60-minute synchronous discussion (discussion times TBD)
- **Slack workspace:** The go-to place for questions, discussion, feedback, suggestions, etc.

Other: Live guest lectures (2 - 3 during the quarter)

- There will be several live guest visits / presentations that we'll ask you to call in for synchronously during the quarter. We'll announce these far in advance so you can plan ahead.

Note: *This looks like an intimidating list to keep track of. You can be assured that everything will be carefully organized on GauchoSpace in a weekly "to-do" list for the course.*

BEFORE THE FIRST WEEK OF CLASS:
You need to install R and RStudio, and the tidyverse package, before the first week of class. **All software for this class is free and open source.**

Click here for instructions on installing R, RStudio and the tidyverse

GRADE BREAKDOWN:

Weekly discussion participation: 15%

- You are expected to attend all 10 weekly discussions
- You can miss one discussion without impacting your grade
- You **must attend at least 7 discussions** to get a passing grade in the course
- We understand that things come up. You can attend a discussion section other than the one you signed up for (please let us know beforehand if you're switching)

Forum engagement: 10%

- Part of your weekly assignment will be to post to the forum. This may include posting a data visualization that you find in the news, sharing a new coding skill or tool that you learned from

the R community, responding to weekly readings, investigating and sharing examples of bias in data science, etc.

Weekly assignments: 60%

- Each week you will complete an assignment to practice conceptual, computational, and critical thinking skills needed for environmental data science. You will submit your assignment via GauchoSpace.

Final exam: 15%

- You will complete an individual, take-home final exam (think of this as a final assignment that aggregates everything you've learned during the quarter).

Evaluation will be by a combination of:

- Automatic submission & score recording (in GauchoSpace)
- Peer review & feedback
- Self-assessment
- Instructor/TA evaluation

CODE OF CONDUCT:

All students are expected to read and comply with the UCSB and 206/244 course 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 welcoming and inclusive 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, or religion (or lack thereof). We do not tolerate harassment of class participants, teaching assistants, or instructors in any form. Derogatory, abusive, or demeaning language or imagery will not be tolerated.

GAUCHOSPACE PRIVACY:

GauchoSpace logs student interaction information (e.g. last login-time, if you've downloaded documents or clicked on links, etc.). In ESM 206 we have a policy of NOT looking at GauchoSpace user logs or using logged information in assessments or final grades, but you should know that the data is stored and made available to instructors & TAs and may be used in other courses.

COURSE POLICIES:

- You are expected to read and comply with the ESM 206/244 Code of Conduct
- You are expected to watch all lectures and code along with all labs
- You must attend at least 7 discussion sections to pass the course
- While we do encourage collaboration and peer teaching, cheating/plagiarism (including code) will result in 0 points awarded for the assessment or activity and disciplinary action according to UCSB policy
- If you are worried about your overall grade and/or passing the class, please reach out to us to discuss your concerns

STUDENT SUPPORT:

We understand that ongoing crises impact students differently based on experiences, identities, living situations and resources, family responsibilities, and unforeseen challenges. We encourage you to prioritize your well-being. We are here to help you reach your learning and career goals. You are always welcome to reach out to our teaching team so that we can best support you. Please see the UCSB Campus Resource Guide for campus student support and services.

DISABLED STUDENTS PROGRAM:

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

TENTATIVE COURSE SCHEDULE AND TOPICS:

Week	Lecture topics	Lab skills
1	Course intro and motivation; what is data science?; computational reproducibility & reproducible workflows; introduction to R	Meet R/RStudio; project management in R projects; creating organized, well-annotated and reproducible scripts; reading in CSVs; basic data wrangling with {dplyr}; {ggplot2} intro
2	Tidy data; exploring data; "Good enough practices" in data science; engaging and responsible data visualization	R Markdown intro; initial cleaning for messy(ish) data; data wrangling with {dplyr} and {tidyr} continued; exploratory data analysis; dataviz with {ggplot2} continued
3	Version control with Git & GitHub, metadata, population, sampling, bias, basic summary statistics (central tendency & data spread)	Basics of working between GitHub and RStudio; wrangling and data viz continued; {lubridate} for dealing with dates and times
4	Exploring differences between populations (effect size, actual differences, difference in context, intro to null hypothesis statistical testing)	Frequency tables with {count}, {tally} and {n}; summary statistics; exploring normality; t-tests; project management with GitHub and R projects
5	Hypothesis testing pros & cons, t-tests, ANOVA, pitfalls of null hypothesis statistical testing, errors, risk & bias, communicating outcomes of hypothesis tests (and why the p-value should be the least interesting thing you present)	GitHub continued; t-tests; ANOVA; responsibly visualizing differences between groups; reading in data from different sources/file types (URLs, .xlsx, etc.)
6	Ordinary least squares, correlation, thinking critically about relationships	Linear regression
7	Types of bias; chi-square; multiple linear regression	Chi-square; advanced wrangling and cleaning; dataviz customization; exporting graphs and tables; GitHub collaboration intro; multiple linear regression intro & assumptions
8	Multiple linear regression continued; interpreting coefficients; interaction terms; communicating regression outcomes	Tidy model outputs; diagnostics; assessing fit; reporting results; {broom} and {stargazer};
9	Logistic regression intro; rank-based tests	Binary logistic regression, ordinal logistic regression intro; Mann-Whitney U; Wilcoxon Signed-Rank
10	Intro to spatial data viz in R; thinking beyond R; course review	Spatial data viz and exploration with {sf}, {tmap}