ENV 872 - Environmental Data Analytics

Nicholas School of the Environment - Duke University
Spring 2022

Instructors

Luana Marangon Lima
Gross Hall, Suite 101
Office hours: TBD

John Fay
3112 Grainger Hall
Office hours: Monday 1:00-2:00pm & by appt

Teaching Assistants

Camille DeSisto
Eva May
Aislinn McLaughlin
Tay Holliday
Yingfan Zeng

Lectures:

Lectures for this class are pre-recorded and made available on Sakai. You should watch them prior to your section's in-person session.

Lab Sections:

Section 1 (Lima): Mon 10:15 - 11:30am
Section 2 (Fay): Tues 8:30 - 9:45am
Section 3 (Lima): Wed 10:15 - 11:30am
Section 4 (Fay): Thur 8:30 - 9:45am

Course Description

Given the growing focus of environmental disciplines on quantitative approaches, students entering the environmental workplace have a need to face new challenges related to data. Data analytics encompasses not only statistics and data visualization but also puts those practices in context of the acquisition, exploration, processing, and reporting of data. In this course, we work through contemporary data analyses while developing skills to integrate software applications, manage data, and effectively report results. Students will develop reproducible workflows to analyze real environmental datasets from start to finish. The setting of this course is a hands-on lab, where students will work through a series of lessons, assignments, and a final course project.

Course Objectives

- Gain proficiency in the language and application of R software
- Synthesize information from datasets, working from start to finish in the data pipeline

- Develop skills to identify and apply appropriate statistical and graphical approaches for environmental datasets, incorporating the guidelines of experimental design and interpretation of output
- Integrate multiple technological applications involved in contemporary data analysis, workflow, and management

---

Course Platforms and Communications

We will use R and RStudio to develop our codes. Recorded classes, additional resources and announcements will be posted on Sakai. We will use a Slack workspace for communication. That way your are just a text message away from instructors and TAs. Please click here join us at ENV872-2022. We will use GitHub to share the lessons, scripts developed in lectures and lab sections as well as Assignments. Here is the link to the GitHub class repository: https://github.com/ENV872/Environmental_Data_Analytics_2022.

Course Prerequisites

Prior R programming is preferable but not mandatory.

Course Resources

- R for Data Science, Garrett Grolemund and Hadley Wickham. Open source online textbook: https://r4ds.had.co.nz/.
- RStudio Cheatsheets https://rstudio.com/resources/cheatsheets/.
- Additional texts and articles will be made available on Sakai.

Course Format and Grading

The course consists of lectures at which we will discuss theory and applications using R packages. Each module will be accompanied by an assignment. Grades will be based on:872

- 11 Homework assignments (85%)
- 1 final project (15%)

Homework Assignments

The assignments involve applying concepts and tools learned in class to an specific data set or problem. Students might work together and help each other. However, the assignments are to be submitted individually.

Final Project

The final project could take several forms. If you have an interesting dataset, you may choose to work with it using existing methods and software tools to run your data analysis. Students are encouraged to work in teams of two or three for a project. Students may only work with students in the same Lab section.

There will be one short presentations to be held during the Lab sections where you will show the class the main results obtained throughout the analysis. Aside from the presentations, you are required to submit a final report where you describe the data sets, tools used and results.

	Sec. 1 (M)	Sec. 2 (T)	Sec. 3 (W)	Sec 4 (Th)
Project Presentation	TBA	TBA	TBA	TBA

Class Proposed Schedule

The schedule below is subject to change. The instructors may modify it throughout the semester if extra time is needed for some particular topics. We will provide updates in class, via Sakai and/or Slack.

Week	Module/Lessons	Recorded Videos Post Date	Sec. 1 (M)	Sec. 2 (Tu)	Sec. 3 (W)	Sec. 4 (Th)	Due Date
1	What is Data Analytics? Data Structures, File formats, languages, Intro to R and RStudio, GitHub	1/5	1/5	1/11	1/12	1/13	1/17
2	RStudio basics, R Markdown, Packages, Reproducibility, Coding basics	1/16	1/10	1/18	1/19	1/20	1/24
3	Data exploration, Importing Data Sets to R, View summary, data formats, data frame structure, missing observations and NAs	1/23	1/24	1/25	1/26	1/27	1/31
4	Data wrangling, creating tidy data sets, package `dplyr`, filtering data, pivoting data, ordering data, `lubridate` package	1/30	1/31	2/1	2/2	2/3	2/7
5	Data visualization, package `ggplot`, improving plots, labels, legend, saving plot to file	2/6	2/7	2/8	2/9	2/10	2/14
6	Modeling - GLMs, interpreting the results from `lm()`, ANOVA, correlation	2/13	2/14	2/15	2/16	2/17	2/28
7	Modeling - continued		2/21	2/22	2/23	2/24	
8	Time Series Analysis, stationarity, trends, seasonality, tests *Spring break*	2/20	2/28	3/1	3/2	3/3	3/14
9	Spatial Data Analysis & Visualization	2/28	3/14	3/15	3/16	3/17	3/21
10	Data Scraping, web-scraped data, manual copy & paste, API, DOM parsing	3/20	3/21	3/22	3/23	3/24	3/28
11	Crafting Reports, Dashboards, R shiny	3/27	3/28	3/29	3/30	3/31	4/2
12	Python from an R perspective	4/3	4/4	4/5	4/6	4/7	NA
13	Class wrap-up	4/10					

Class Etiquette

You should take responsibility for your education. We expect students to attend every class and get to class on time. If you must enter the class late, please do so quietly. Refrain from using phones and tablets for social media during class. Most classes will involve coding . We expect you to focus on the assignment and refrain from any web browsing that may disrupt the progress of your work.

Your classmates deserve your respect and support. We will likely have students from many different backgrounds and countries in this class and you should all feel comfortable and make each other comfortable while participating.

Nicholas School Honor Code

All activities of Nicholas School students, including those in this course, are governed by the Duke Community Standard, which states:

> Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and nonacademic endeavors, and to protect and promote a culture of integrity. To uphold the Duke Community Standard:
>
> - I will not lie, cheat, or steal in my academic endeavors;
> - I will conduct myself honorably in all my endeavors; and
> - I will act if the Standard is compromised.

Land Acknowledgment

> What is now Durham was originally the territory of several Native nations, including Tutelo (TOOtee-lo) and Saponi (suh-POE-nee) - speaking peoples. Many of their communities were displaced or killed through war, disease, and colonial expansion. Today, the Triangle is surrounded by contemporary Native nations, the descendants of Tutelo, Saponi, and other Indigenous peoples who survived (suh-POE-nee), and Occaneechi (oh-kuh-NEE-chee) Band of Saponi. North Carolina's Research Triangle is also home to a thriving urban Native American community who represent Native nations from across the United States. Together, these Indigenous nations and communities contribute to North Carolina's ranking as the state with the largest Native American population east of Oklahoma."

Final Remarks

The instructors would like to acknowledge that significant part of this course material was developed by Kateri Salk (kateri.salk@duke.edu).