Environmental Data Analytics

ENV 872

Nicholas School of the Environment - Duke University

Spring 2021

Instructors

Luana Marangon Lima
luana.marangon.lima@duke.edu
Office hours: TBD

John Fay
john.fay@duke.edu
Office hours: TBD

Teaching Assistants

Abhishek Jain
abhisheksanjay.jain@duke.edu
Office hours: TBD

Sarah Roberts
sarah.m.roberts@duke.edu
Office hours: TBD

Lectures: On-line asynchronous (pre-recorded classes will be posted on Mondays)
Lab Sections:
Section 1 (Lima): Wed 10:15 - 11:30am
Section 2 (Fay): Thurs 8:30 - 9:45am
Section 3 (Lima): Mon 1:45 - 3:00am

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 env872-2021. 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.

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 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. The table below shows tentative post and due dates for the assignments.

#	Post Date	Due Date
Assignment 1	1/20/21	1/26/21
Assignment 2	1/27/21	2/2/21
Assignment 3	2/3/21	2/9/21
Assignment 4	2/10/21	2/16/21
Assignment 5	2/17/21	2/23/21
Assignment 6	2/24/21	3/2/21
Assignment 7	3/3/21	3/16/21
Assignment 8	3/17/21	3/23/21
Assignment 9	3/24/21	3/30/21
Assignment 10	3/31/21	4/6/21
Assignment 11	4/14/21	4/20/21

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 (W)	Sec. 2 (Th)	Sec. 3 (M)
Project Presentation	Apr-21	Apr-22	Apr-26
Final Report		Apr-23	

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	Lab Sec. 1 (W)	Lab Sec. 2 (Th)	Lab Sec. 3 (M)
1	What is Data Analytics? Data Structure, File formats, languages, Intro to R and RStudio, Github	20-Jan	20-Jan	21-Jan	25-Jan
2	RStudio basics, R Markdown, Packages, Reproducibility, Coding Basics	25-Jan	27-Jan	28-Jan	1-Feb
3	Data exploration, Importing Data Sets to R, View summary, data formats, data frame structure, missing observations and NAs	1-Feb	3-Feb	4-Feb	8-Feb
4	Data wrangling, creating tidy data sets, package *dplyr*, filtering data, pivoting data, ordering data, *lubridate* package	8-Feb	10-Feb	11-Feb	15-Feb
5	Data visualization, package *ggplot*, improving plots, labels, legend, saving plot to file	15-Feb	17-Feb	18-Feb	22-Feb
6	Modeling - GLMs, interpreting the results from *lm*(), ANOVA, correlation	22-Feb	24-Feb	25-Feb	1-Mar
7	Time Series Analysis, stationarity, trends, seasonality, tests	1-Mar	3-Mar	4-Mar	8-Mar
8	*Data challenges in the environmental field (Guest speaker: Kateri Salk)*	-	**10-Mar**	11-Mar	15-Mar

Week	Module/Lessons	Recorded Videos Post Date	Lab Sec. 1 (W)	Lab Sec. 2 (Th)	Lab Sec. 3 (M)
9	Spatial Visualization	15-Mar	17-Mar	18-Mar	22-Mar
10	Data Scraping, web-scraped data, manual copy & paste, API, DOM parsing	22-Mar	24-Mar	25-Mar	29-Mar
11	Crafting Reports, Dashboards, R shiny	29-Mar	31-Mar	1-Apr	5-Apr
-	*Break - No classes*	-	**7-Apr**	**8-Apr**	**12-Apr**
12	Machine Learning ?	12-Apr	14-Apr	15-Apr	19-Apr
13	Python ? (Final project presentations)	19-Apr	21-Apr	22-Apr	**26-Apr**

Class Etiquette

You should take responsibility for your education. I expect students to attend every class and get to class on time. If you must enter the class late, please do so quietly. Retain from using phones and tablets for social media during class. Some classes will involve coding on your laptop. I 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."*

Please add the following affirmation to the end of all assignments, and sign your name beside it: "I have adhered to the Duke Community Standard in completing this assignment."

Land Acknowledgment

"What is now Durham was originally the territory of several Native nations, including Tutelo (TOO-tee-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 early colonization. These nations include the Haliwa-Saponi (HALL-i-wa suh-POE-nee), Sappony (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).