Exploring Environmental Data (ENS-215)

Winter 2023

Course Information

Term: Winter 2023
Location: Olin 306 (*Class*) and Olin 306 (*Lab*)
Meeting times:

- **Class:** M/W/F, 9:15-10:20 AM
- **Lab:** M, 1:50-4:50 PM

Professor: Mason Stahl (https://sites.google.com/union.edu/masonstahl/home) (stahlm@union.edu (mailto:stahlm@union.edu))
TA: Jacob Abbott (abbottj@union.edu (mailto:abbottj@union.edu))
Office hours: Olin 102A – Mondays 10:30-11:30 AM and Thursdays 9:30-11:00 AM

Course description

> *We are drowning in information, while starving for wisdom. The world henceforth will be run by synthesizers, people able to put together the right information at the right time, think critically about it, and make important choices wisely.*
> – Edward O. Wilson

Understanding how the Earth and environment works requires the careful analysis and interpretation of scientific data. Increasingly, the limitations to our understanding lie not in the availability of data, but rather in our ability to analyze and find meaning in it. Deriving insight from environmental data, in particular large and complex datasets, requires new tools, methods, and ways of thinking. In this class we are going to learn how to code in the programming language R and use it to analyze environmental data in order to better understand the Earth's systems. This course will feature a hands-on classroom with programming and data analysis occurring interactively during the class. Students will learn how to analyze and visualize large datasets ("big data") and how to write code, while also covering interesting components of environmental and Earth sciences. The datasets used throughout the term will be research quality (e.g., from peer-reviewed publications, government agencies, your thesis data) and will address a range of global challenges including climate change, water resources, biodiversity loss, ecological and environmental degradation, energy resources, and agricultural production. The term will culminate with a "research conference" where each student will deliver a presentation on the environmental data analysis project they conducted over the course of the term. These projects involve the in-depth analysis of research quality datasets (e.g., data from NASA, USGS, EPA, NOAA) to address a research question(s) posed by each student and are an opportunity for each student to develop and conduct an original research project on an environmental topic of their interest.

Course premise

This course will be taught as a **"flipped"** class. This means that you will learn much of the basic material outside of class – through online assignments/tutorials, readings, and watching videos – while you will "learn by doing" through hands on programming exercises during class.

At the start of each class you will be given a packet with the days work. You should keep these handouts as they will serve as part of your course notes. During class you will work through the project laid out in the packet and at the end of class you will have an R Notebook that has all of your code, results, and notes. You will then apply what you learned in class to your weekly homework assignments. You can not rush through the class work as care and thought are required to learn the concepts needed for the assignments and later classes.

Resources

Textbooks

The three texts that we will primarily rely on are listed below. All three are freely available on the web. Print versions of all three books are also available.

- **R for Data Science** (https://r4ds.had.co.nz/): Excellent textbook on using R for data science.
- **ModernDive** (https://moderndive.com/): Introductory textbook on data science and stats in R, that presents the material in a clear and concise manner. This book is freely available online.

- **Geocomputation with R** (https://geocompr.robinlovelace.net/): Excellent textbook on geospatial analysis using R.

Course sites

- **This website** (https://stahlm.github.io/ENS_215/Winter_2023/Home.html): The site you are on now will be updated throughout the term. In particular the schedule will have material added on a weekly basis.
- **Nexus:** (https://nexus.union.edu) You will submit all of your assignments through our Nexus class site
- **DataCamp:** (https://app.datacamp.com/groups/exploring-environmental-data-w-2023) The online assignments/tutorials can be found on our DataCamp class site

Assignment templates and guidelines

- **Template for lab and homework assignments: (html)** (https://stahlm.github.io/ENS_215/Homework/Template_Assignments.html) **(Rmd)** (https://raw.githubusercontent.com/stahlm/stahlm.github.io/master/ENS_215/Homework/Template_Assignments.Rmd)
- **Lab report guidelines** (https://stahlm.github.io/ENS_215/Labs/Lab_Guidelines.html)

Useful links/resources

- **Google** (https://www.google.com): This may seem pretty obvious, but Googling your R question almost always yields a helpful result/example.
- **stackoverflow.com** (https://stackoverflow.com/): Website where people post questions related to programming. This is an extensive resource with a huge database of questions and answers related to R. Generally it is best to Google your question and often stackoverflow will be one of the first hits.
- **YouTube** (https://www.youtube.com/): Lots of instructional videos in R are available on Youtube.
- **RStudio Cloud** (https://rstudio.cloud): A free service (up to 25 hrs/month) from the developers of RStudio, which allows you to run RStudio right in your browser. You will install R and RStudio on your personal computer for this class, however this is an additional way to run R on any computer that has an internet connection without the need to install software. Note that the free use has a monthly time cap, so this service can be used as a supplement but cannot replace your personal install (which is 100% free).
- **R Cheatsheets** (https://www.rstudio.com/resources/cheatsheets/): Very helpful 1-2 page summary sheets that clearly show you how to code in R. There are a number of different cheat sheets ranging from a basic R tutorial to creating plots in R.
- **R Webinars** (https://resources.rstudio.com/webinars): These instructional web videos cover a range of topics in R.
- **STDHA.com** (http://www.sthda.com/english/): Excellent resource for help with R.
- **USGS R Training Materials** (https://owi.usgs.gov/R/training-curriculum/intro-curriculum/): Excellent R tutorial with some cool examples USGS data from the USGS.
- **Fundamentals of Data Visualization** (https://serialmentor.com/dataviz/): Excellent resource for learning how to visually display data in an attractive and informative way. FYI, this entire book was produced in R Markdown!
- **R Style Guide** (https://style.tidyverse.org/): Advice on how to write code that is neatly organized and easy to read.
- **Guide to Using R Markdown for Class Reports** (https://stahlm.github.io/ENS_215/Resources/R_Markdown_for_Reports.html): Very helpful guide on how to create reports using R Markdown.

Assignments

Online tutorials/homework

These tutorials/homework are required and are a very helpful and important part of teaching you how to code in the short timeframe of one term. The tutorials help to make learning to code a much more enjoyable and efficient process and also serve as an excellent refresher once you've learned coding. There will typically be a homework assignment for every class topic. If you come to class and work diligently you should be able to complete the homework. As the homework builds off of the in-class material, coming to class is necessary. I have set up a class group on the site where the tutorials are hosted and I will be able to monitor and track your progress. Successful completion of the tutorial will result in a 100% grade for that assignment. If the tutorial is not fully completed then you will receive a score equal to the percent that you completed at the time the assignment was due. Note that the there is an option to take hints during the tutorials. Taking hints does affect your final score (however a single hint typically has a minimal effect) so use them sparingly.

Labs

In lab you will be given a dataset from the Earth and environmental sciences with the goal of exploring and discovering interesting features and illuminating important scientific concepts. You will be given background information and the data you will need to successfully complete the lab, though successful completion of the lab will require independent and creative exploratory analysis. The goal of the lab assignments is to apply the tools and concepts learned in class towards asking and answering interesting scientific questions. These labs are intended to improve your skills in data analysis, learn about concepts from Earth and environmental sciences, and importantly to develop skills as a creative and independent researcher. These skills will serve you will in future classes, grad school, and jobs.

Final project

Each student will produce a report/presentation based on their analysis of an environmental/Earth science dataset. You will develop questions/hypotheses about your dataset and use the skills we have learned throughout the term to tackle these questions. You can analyze a dataset that you are interested and find on your own or I can point you to an interesting dataset. If you are doing a senior thesis or independent research project, this is a great opportunity to start analyzing that data. During the last week of class we will have a mini conference session where each student will present their work. Additional details on the final project will be provided later in the term.

Course policies

Grading

Assignment	Weight
Homework and online tutorials	32.5%
Lab	25%
Final project	32.5%
Class participation	10%

As discussed in the **Honor code** section below, all submitted work must be your own. If you use any references you must cite your sources.

Evaluation rubric

Where applicable a rubric will be applied to graded assignments.

Topic	Excellent: 3	Good: 2	Needs work: 1
Code style and strategy	Above and beyond what was expected/required. Code is very organized, well structured, and well commented.	Code runs reasonably well but may have a few errors, acceptable level of commenting, and reasonably organized and structured.	Many errors in style/structure and limited effort to make the code human readable
Presentation of results	Figures and tables are above and beyond what is expected. Clear and easy to read. Well chosen graphs that clearly illustrate the intended point/concept.	Well made figures and tables with some minor issues (e.g. labels, symbol sizes, design not ideal)	Results poorly conveyed. Poor formatting and/or choose of method for conveying results.
Achievement, mastery, creativity	Above and beyond what is expected. Excellent use of tools/concepts and sophisticated application of concepts from this course. Creative approach to problem solving and analysis.	Good application of tools/concepts learned in this class. Competent approach, though not necessarily very creative.	Student did not display mastery of the concepts/tools learned in this class. Worked lacked sufficient depth.
Discussion	High quality discussion and interpretation of the results. Clear understanding of the scientific concepts under examination in the assignment.	Good discussion and demonstrates reasonable understanding of the scientific concepts.	Limited discussion that does not demonstrate a good understanding of the concepts being examined in the assignment.
Ease of access for instructor	Code runs on my computer with no problems. Assignment is named according to specified naming convention. Proper headers (e.g. name, date).	Satisfactory	No effort made to make code easy to run and assignment easy to interpret.

Grade mapping

Grade	Percentage
0	0%
1	50%
2	85%
3	100%

Intermediate points are allowed, and follow a linear scale in between each point.

My Expectations

I want and expect the class to be an engaging and enjoyable environment for all students. We should all work together in a positive and respectful manner and every student should feel comfortable contributing to class discussions and asking questions. All students must comply with the Union College honor code. This includes but is not limited to issues of cheating and plagiarism. If you have any doubts about whether you are in compliance with the honor code you should ask me for guidance prior to submitting the material in question.

Computer use

We will use computers in all of our class and lab meetings. If you have a personal laptop you can use it or you can use the computers in the classroom. **Using your computer for anything not related to the class is unacceptable and will adversely impact your participation grade**. Among other things, this means that you should not check email, Instagram, Twitter, news, or any other content not directly related to the coursework during class/lab meetings. Your undivided attention and focus is important to your success in this class.

Class attendance and participation

You are expected to attend every class and to arrive on time. I will have a sign-in sheet that you will need to sign as you enter class. Missing class or arriving late will significantly impact your class participation grade. This course is very hands-on and you will learn by doing, so missing class will make it very difficult to succeed. Furthermore, this class is very interactive and you will learn from and teach your classmates so attendance is important to supporting the best

learning environment for all. While some people like to wear headphones when coding, this is not allowed during class or lab as it interferes with participation/collaboration. If you have to miss class for a legitimate reason let me know ahead of time.

Due-dates

All assignments must be turned in on time. Homework is due prior to the start of class on the date it is due. Labs are due prior to the start of lab on the day the assignment is due. In the case of online tutorials (e.g. DataCamp assignments) you will recieve a grade corresponding to the percent of the tutorial successfully completed at the time the assignment was due (e.g. if you successfully completed 85% of the tutorial at the time it was due, then you would receive and 85%). For other assignments, I will use the time-stamp on Nexus to determine if the assignment was turned in on time. Late assignments (*not including online tutorials*) will lose 12.5 pts/day and will not be accepted if > 3 days late.

Email policy

I try to respond to emails in a reasonably timely manner, however do not expect rapid replies late at night or on weekends. If you have a question about coding you should first try Googling it, checking your textbook, R help files, my office hours, etc. before emailing. Please state your question as clearly as possible. Also, please make sure to reply to confirm that you've received my response (a simple Thanks, got it is often all that's needed), as this is a good habit to develop in your academic and professional careers.

Collaboration

I encourage collaboration and an interactive environment. Talking problems over with classmates is an excellent way to overcome challenges and to teach and learn from one another. However, it is critical that all work you submit is your own. This means that discussing approaches, techniques, things to look out for are all great but copying someone else's code is absolutely not allowed and constitutes plagiarism.

Honor code

Union College recognizes the need to create an environment of mutual trust as part of its educational mission. Responsible participation in an academic community requires respect for and acknowledgement of the thoughts and work of others, whether expressed in the present or in some distant time and place. Matriculation at the College is taken to signify implicit agreement with the Academic Honor Code, available at: honorcode.union.edu. It is each student's responsibility to ensure that submitted work is his or her own and does not involve any form of academic misconduct. Students are expected to ask their course instructors for clarification regarding, but not limited to, collaboration, citations, and plagiarism. **Ignorance is not an excuse for breaching academic integrity**.

Disabilities

Academic accommodations are available for students with disabilities who are registered with the Office of Student Support Services (8785). Students in need of disability accommodations should schedule an appointment with me early in the semester to discuss any accommodations for this course that have been approved by the college, as indicated on your OSSS ID card.

Mason Stahl (2023)
Union College