Research Computing in Earth Science
EESC GR6901
Tuesdays and Thursdays 1:10 – 2:25 pm
558 EXT Schermerhorn Hall

Instructor

Prof. Daniel M. Westervelt
danielmw@ldeo.columbia.edu
Office hours: By appointment: https://calendly.com/dan-westervelt (on zoom)
Github: dmw2166

Teaching Assistant

Garima Raheja
garima.raheja@columbia.edu
Office hours: Thursday, 2:30-3:30 PM in 551 Schermerhorn
Github: graheja

Unless it is an emergency, please expect email responses in 24-48 hours.

Course Description

Computing has become an indispensable tool for nearly all Earth and Environmental Scientists, and for many careers inside and out of academia, but it doesn't often appear in our curricula. This course focuses on *data analysis*, a subset of computing in which the data already exist, e.g. from observations or from the output of a simulation, but have to be transformed into understanding. There are many different ways to gain understanding, but most workflows often boil down to:

- read data files
- perform some analysis operations, from very simple (e.g. take the mean) to very complex (e.g. train a deep neural network)
- visualize the output in a plot

The goal of this course is to teach the basic foundations of Earth and Environmental Data Science which are often overlooked. The material is designed to be accessible for Earth Science graduate students in any discipline, with no prerequisites.

This course is intended to introduce new graduate students to modern computing software, programming tools and best practices that are broadly applicable to the analysis and visualization of Earth and Environmental data. This includes an introduction to Unix, version control (git and GitHub), and basic programming in the open-source Python language. The bulk of the content is devoted to in-depth exploration of the numerical analysis and visualization packages which comprise the modern Scientific Python ecosystem, including Numpy, Scipy, Matplotlib, Pandas, Xarray, using real Earth and Environmental datasets.

Course Objectives

- Use Linux/Unix commands to work with files and directories
- Navigate the JupyterLab Environment effectively
- Identify common geoscience data formats and the Python packages which can load them
- Perform basic exploratory data analysis on Earth and Environmental data, distinguishing between
    - *Tabular data*: rows and columns
    - *Gridded data*: multidimensional numerical arrays
- Use visualization to enhance interpretation of data, including maps and interactive visualizations
- Construct complete, well-structured programs in Python
- Practice collaborative, reproducible research through version control and GitHub

Course website

Courseworks: https://courseworks2.columbia.edu/courses/191326

Course Slack - link

The Slack channel can be used to communicate with your peers and instructors about course material.

Please use this channel to collaborate and debug your homework together! Note that this Slack is considered to be an extension of the classroom, so you must follow the normal standards of professional communication.

Method of Evaluation

Activity	Weight	Due Date
Attendance and Participation	10%	Each week
Assignments	40%	Approx. every week
Final Project	50%	End of semester

Course Format

Officially, this is a lecture-based course. However, slideshow presentation is not the most effective modality for learning to use data science tools such as python. The course will include some (not many!) traditional slideshow presentations and otherwise will be real-time coding demonstrations. The best way to learn coding and data science is hands-on, so students will have the most success if they are following along in real time. Therefore, be sure to bring your laptop to each class meeting. Tuesday meetings of class will consist of lectures and/or real-time coding demonstrations. **Thursday meetings will be more informal and will be used for either: finishing up some concepts from Tuesday, collaborative work time on assignments with classmates, and/or open office hours.**

Textbook

There is a free textbook available online, originally developed by Prof. Ryan Abernathy.
https://earth-env-data-science.github.io/

Participation (10%)

Students are expected to come to class prepared and to actively participate in discussion and activities. If traditional participation in class has been challenging for you previously, we encourage you to meet with the instructor and/or TA to develop a plan for sustainable, meaningful participation in this class. Participation grades will be assessed based on both the frequency and quality of students' engagement in class and/or on Slack. A good rule of thumb is asking at least one thoughtful question during a lecture-style class, and collaborating with at least one other student during work session-style class. Quality is not about being "correct" or "knowing all the answers", but about the extent to which students' comments reflect thoughtful engagement with the assigned materials and the comments of others. If you are sick or unable to come to class in person, please email the instructor and TA at least 1 hour before class, and we will do our best to set up a Zoom.

Assignments (40%)

There will be approximately 10 assignments throughout the semester. Assignments are due by 2:00pm on Fridays. Assignments will be posted on Courseworks. To turn in assignments, you will simply push your code/notebooks to an assignments repository on your github, and you will make the professor (Github: dmw2166) and TA (Github: graheja) collaborators so we can check your work. You can keep pushing changes to your assignments up through the deadline. We will grade the last push before the deadline.

We encourage assignments to be collaborative and will be setting aside dedicated time for collaborative work! However, it is important that every student submit their own homework assignment that accurately represents their own understanding of the topics of the assignment. Spot-checking code, asking for help from classmates if stuck, comparing answers, etc. are all encouraged. In your submission, you MUST list your collaborators in a commented line in the first cell of your Jupyter notebook. Copying one's code directly verbatim would be an example of an unacceptable level of group work. Please reach out to the professor or the TA if you would like any clarification.

Final Project (50%)

A final project will be submitted by each student. As with assignments, student collaboration is encouraged, as long as students submit an independent, unique project accurately representing their understanding. The project will synthesize many topics we cover over the course of the semester. Broadly, students will be asked to pick a dataset, and use this dataset to investigate a scientific question or hypothesis. They will use the coding skills to create figures (in a Jupyter notebook) that

provides insight regarding the scientific question or hypothesis. There will also be a 5-minute presentation on the project. Students are welcome to use datasets related to their PhD research. More details will be forthcoming later in the semester.

Late Assignment Policy

All assignments should be submitted via CourseWorks by the due date and time. That said, we understand when illnesses, burnout, research travel, job interviews, or other things get in the way of work and school. Each student will have up to 5 total days of "no questions asked" extensions that they may use on any assignment. This is 5 cumulative days, meaning students could use a 1-day (24 hour) extension 5 times, or a 5-day extension one time, or anything in between. Note that weekends don't count towards the 5-day tally. For example, a 24-hour extension of an assignment due Friday afternoon would extend to Monday afternoon. Extensions are given in 1-day increments, so submitting a homework 1 minute late uses up 1 day of your extension. You do not need to ask or wait for permission or give a reason in order to use the extension days. However, when you have turned in the late assignment, you must email both the instructor and the TA with the words "Late Assignment Submission" in the subject line to inform them of your submission. Of course, extensions will always be given for medical or family emergencies and extended professional obligations, outside of these 5-day extensions. For this, please communicate with the instructor and the TA as soon as reasonably possible.

A Note on Late Work and Plagiarism

Plagiarism often happens not by desire but in desperation. Since Columbia has strict zero-tolerance policies regarding plagiarism, even one incident can have long-standing effects on your academic career. If you are struggling to understand the concepts to the point of being unable to do the assignments, or if you are submitting work late because the assignments are taking you significantly longer than the allotted time, then we want to hear about it, and we will take your feedback as we grow our teaching in the future. Often students wait until the last few weeks of the semester to reach out – there is not much we can do then! We hope to create a supportive and inclusive environment, so please reach out and **communicate proactively** so that we can help! Please also refer to the Academic Integrity statement towards the end of this document.

Grades: Definitions and Scale

Final grades are assigned to course average scores based on the following scale:

A+ Rare performance. Reserved for highly exceptional achievement.
A Excellent. Outstanding achievement.
A- Excellent work, close to outstanding.
B+ Very good. Solid achievement expected of most graduate students.
B Good. Acceptable achievement.
B- Acceptable achievement, but below what is generally expected of graduate students.
C+ Fair achievement, above minimally acceptable level.
C Fair achievement, but only minimally acceptable.
C- Very low performance.
F Failure. Course usually may not be repeated unless it is a required course.

Course Schedule

Date	Topics Covered	Description	Items Due on Friday 2 PM
Week 0, Jan 16 and 18	Survey	No class!	Week 0 Survey
Week 1, Jan 23 and 25	Intro	What is data science, big data? Motivation and inspiration for research computing	N/A
Week 2, Jan 30 and Feb 1	Unix, JupyterLab	Terminal, command-line, shell scripting, interactive computing environment	Assignment 0

Date	Topics Covered	Description	Items Due on Friday 2 PM
Week 3, Feb 6 and 8	Git, github	Version control	Assignment 1
Week 4, Feb 13 and 15	Python fundamentals	Data structures, functions, classes, dictionaries, lists, loops	Assignment 2
Week 5, Feb 20 and 22	Numpy and Scipy	Importing libraries, arrays, indexing, reduction operations	Assignment 3
Week 6, Feb 27 and 29	Numpy, Scipy, and Matplotlib	Plots, subplots, labeling plots, customizing	Assignment 4
Week 7, March 5 and 7	Pandas Part 1	Dataframes, resampling, reading files	Assignment 5
Week 8, March 12 and 14	SPRING BREAK!	*No class*	None
Week 9, March 19 and 21	Pandas Part 2	Plots, groupby, merging	Assignment 6
Week 10, March 26 and 28	Gridded data	Xarary, map projections, interpolation	Assignment 7
Week 11, April 2 and 4	Scikit-learn	Machine learning, classification, regression, clustering, dimensionality reduction	Assignment 8
Week 12, April 9 and 11	Machine learning in earth research	Big data, deep learning	Assignment 9
Week 13, April 16 and 18	Final project work time	Open office hours in class, time for working on final projects	None
Week 14, April 23 and 25	Final projects	5-min in-class presentations	Final project

COMMITMENT TO DIVERSITY, EQUITY, INCLUSION, AND ANTIRACISM

Diversity, equity, inclusion, and antiracism are critical values in education, but they must be supported by a commitment to action. The instructors are committed to creating an inclusive classroom environment. They expect students to work collaboratively to create an equitable classroom environment that actively engages all students in meaningful and relevant learning, values the contributions of students' diverse backgrounds, and acknowledges systemic and institutional challenges.

Names/Pronouns

You deserve to be addressed in a manner that reflects your identity. Please see https://universitylife.columbia.edu/pronouns for more information on university policies and services.

Religious Observance

It is the policy of the Columbia to respect its members' observance of their major religious holidays. Students should notify instructors at the beginning of the semester about their wishes to observe holidays on days when class sessions are scheduled. Where academic scheduling conflicts prove unavoidable, no student will be penalized for absence due to religious reasons, and alternative means will be sought for satisfying any course requirements.

Student Wellness
Mental and physical health is important while undergoing graduate studies. While maintaining good health is a priority, occasional stress is an unavoidable part of a graduate education. Please know that there are resources on campus to help you cope with the pressures of student life. These include support groups and workshops on managing stress, healthy eating and life management. For resources for coping with stress, anxiety, and other specific health related concerns, please visit Columbia Health at https://www.health.columbia.edu/.

Accessibility
Columbia is committed to providing equal access to qualified students with documented disabilities. A student's disability status and reasonable accommodations are individually determined based upon disability documentation and related information gathered through the intake process. For more information regarding this service, please visit the University's Health Services website: http://health.columbia.edu/services/ods/support.

In order to receive disability-related academic accommodations, students must first be registered with the Office of Disability Services (ODS). More information on the ODS registration process is available online at www.health.columbia.edu/ods. Please allow for at least two weeks to complete the ODS registration process. Students are encouraged to contact ODS at the start of the semester. Students who have, or think they may have, a disability are invited to contact ODS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or by email at disability@columbia.edu.

COLUMBIA UNIVERSITY POLICIES
https://www.gsas.columbia.edu/content/university-policies

Academic Integrity
Students should be aware that academic dishonesty (for example, plagiarism, cheating on an examination, or dishonesty in dealing with a faculty member or other university official) or the threat of violence or harassment are particularly serious offenses and will be dealt with severely under Dean's Discipline. Graduate students are expected to exhibit the high level of personal and academic integrity and honesty required of all members of an academic community as they engage in scholarly discourse and research. It is essential to the academic integrity and vitality of this community that individuals do their own work and properly acknowledge the circumstances, ideas, sources, and assistance upon which that work is based. Academic honesty in class assignments and exams is expected of all students at all times.

Students in this course should review and abide by the standards adopted by the Graduate School of Arts and Sciences: https://gsas.columbia.edu/content/academic-integrity-and-responsible-conduct-research. You should read these standards within the first few days of class. Ignorance of policies and standards concerning academic dishonesty shall not be a defense in any disciplinary proceedings.

ADDITIONAL RESOURCES

Columbia University Library
Columbia's extensive library system ranks in the top five academic libraries in the nation, with many of its services and resources available online: https://library.columbia.edu/.

Columbia Writing Center
The Columbia Writing Center aids undergraduate and graduate students with writing consultations for specific writing projects, writing productivity sessions, and workshops. The Writing Center offers these services for free to current Columbia students. Read more about their services on their website. You can also contact the center at writingcenter@columbia.edu with questions.