**Computing and Research Methods for Climate Data Science**

Subject Code/Course Number (CLMT5045)

Meeting Days/Times (TuTh 11am-12:45pm)

Location (Chandler 401)

**Instructor**

Name Dhruv Balwada

Email address dbalwada@ldeo.columbia.edu

Office hours: days/times: TBD

**Teaching Assistant**

Name: Chhavi Dixit

Email address: cd3496@columbia.edu

Office hours: days/times: TBD

**Course Description**

Computing and data analysis have become an indispensable tool for researchers and industry professionals working in virtually any aspect of the modern world. This course will introduce students to the fundamental concepts and methods that are broadly applicable to any data science project, with a thematic focus on climate and environmental data. This includes an introduction to Unix, programming, common data formats, analysis, and visualization. The primary focus will be to teach students the foundations of Python in a climate data science context, which is of the most widely used and accessible programming languages today. Students will also be introduced to cloud computing, which will be the primary tool for in class assignments and projects.

The course is designed to be accessible for any students with an interest in being able to ask and answer questions using data. This course will also be invaluable for those looking to interact with scientists and engineers, manage scientific projects, and develop policies in the realm of climate science and sustainability.

**Course Objectives**

By the end of the course, students will:

* Get a flavor of data science applied to climate science.
* Develop a foundation in Python, with a focus on data analysis.
* Learn common approaches for finding, accessing, and visualizing public datasets.
* Get introduced to modern data science practices, including how to use cloud computing.
* Design and take a data science project from inception to finish.

**Required and Recommended Readings**

No required textbook.

**All the class material will be posted on** [**https://earth-ds-ml.github.io/summer\_2025/intro.html**](https://earth-ds-ml.github.io/summer_2025/intro.html) **as the semester progresses.**

The students may also refer to the following online material available here: <https://earth-env-data-science.github.io/intro.html> and <https://foundations.projectpythia.org/> to read ahead.

**Prerequisites**

Prerequisites include familiarity with some quantitative methods, e.g. basics of statistics and probability, introductory understanding of data analysis, and ability to comprehend graphs and plots. Successful completion of “Quantitative methods for climate applications” or a similar course in a different domain will suffice. Prior experience with some coding or algorithmic thinking may be beneficial, but is not essential.

**Method of Evaluation**

| **Activity** | **Weight** | **Week** | **Due Date** |
| --- | --- | --- | --- |
| Attendance | 10% | All weeks | Continuous |
| Weekly coding assignments | 50% | Week 1-5 | 1 week after assigned. |
| Final project code | 30% | Week 6 | 4 July |
| Final project presentation | 10% | Week 6 | 3 July |

**Attendance (10%)**

Attendance at lectures is mandatory, and students are expected to attend and actively participate in class. In most classes, the first half of class will be devoted to a lecture covering new material, while the second half will focus on hands-on coding exercises and assignments. We will use the LEAP Pangeo Platform (<https://leap.columbia.edu/leap-pangeo-3/>) to access data sets and code using Jupyter notebooks. Students are expected to bring a laptop to class to follow along with lectures and to participate in coding exercises.

Unexcused absences will count against the attendance score. Excused absences include a severe medical issue, family emergency, or significant career-related activity. All excused absences are at the professor’s discretion.

**Coding assignments (50%)**

Coding assignments will be assigned on a weekly basis following material presented in class. There will be 5 coding assignments, each assigned at the beginning of weeks 1 – 5. These exercises will be due by Tuesday at midnight the following week (e.g. the coding assignment assigned at the beginning of week 2 will be due by Tuesday of week 3).

The time commitment will depend to some degree on the prior experience of the students, including their familiarity with coding, algorithmic thinking, and data analysis. Generally, students may expect to spend around 3-4 hours every week on coding assignment.

**Final project (30%)**

The final project should be centered towards addressing a climate science relevant question of the student’s interest, which can be approached using publicly available data. The project will be submitted as a reproducible Jupyter notebook documenting the analysis that was undertaken and the insights that were drawn from the data.

Examples of past projects and the submission process can be found on <https://earth-env-data-science.github.io/projects.html>

Topics will be submitted to the instructor by the end of the 3rd week of class for approval. Final projects will be due at the end of the summer session, and will be evaluated in lieu of a final exam.

**Final project presentation (10%)**

The last class of the semester will be used for presentations from all the projects, where every presentation would be 5 minutes long.

**Late Assignment Policy**

Coding assignments that are turned in late will start to incur a 10% penalty for every 2 days of delay. Example assignment due on Tuesday submitted on Wed or Thu will incur 10% penalty, and if submitted on Friday will incur a 20% penalty and so on.

**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.

**Policy on use of ChatGPT and other AI Tools**

In this course, we recognize that tools like ChatGPT and other AI assistants can support learning, especially in exploring new concepts, debugging code, and clarifying questions. However, it is essential that you develop your own skills and understanding.

Permitted Use:

- You may use AI tools like ChatGPT for certain assignments where explicitly allowed—such as coding practice, project brainstorming, or exploratory data analysis.

- When you do use AI tools, you must document how and where you used them (e.g., in a code comment or footnote). A simple statement like “Used ChatGPT to help debug the plotting code in section 2” is sufficient.

Ultimately, you are responsible for understanding and being able to explain all work you submit. If you cannot explain how a piece of code or analysis works, that is a signal to revisit the material or use the AI agent to help you learn the steps, not rely more heavily on AI to solve for you.

**Course Schedule**

| **Date** | **Topics Covered** | **Description and Readings** | **Items Due** |
| --- | --- | --- | --- |
| Week 1 (26-30 May) | Overview of the course and the interactive computing environment. | Introduction to climate data science, and basics of Unix.  *Following sections\* on:*   * *Intro to JuputerLab* * *Intro to Unix* * *Intro to Git and Github* * *All about Data* |  |
| Week 2 | Introduction to Core Python | Python basics and version control.  *Following sections on:*   * *Python fundamentals* * *Python functions and classes* * *Organization of python projects* | DUE: HW 1 |
| Week 3 | Scientific Python Fundamental | Python for quantitative tasks and visualization.  *Following sections on:*   * *Numpy and Matplotlib* | DUE: HW 2 |
| Week 4 | High level data analysis | Working with tabular and multidimensional data.  *Following sections on:*   * *Pandas* * *Xarray for multidimensional gridded data* | DUE: HW 3 |
| Week 5 | Geoscience and climate model data analysis | Working with observations or climate model data.  *Following sections on:*   * *Working with Earth Science Model Output.* * *Working with output from many different CMIP6 climate models.* | DUE: HW 4 |
| Week 6 (30 June – 4 July) | Big environment data | Advanced plotting and maps  *Following sections on:*   * *Dask for parallel computing in Python* | DUE: HW5 and Final Project |

\*Sections mentioned above can also be found on the <https://earth-env-data-science.github.io/> website in the left panel.

**COLUMBIA CLIMATE SCHOOL POLICIES**

***Inclusive Excellence***

The Climate School aims to advance knowledge and educate leaders for equitable solutions to climate and sustainability challenges. Using an interdisciplinary approach, the School encourages engagement of diverse perspectives on climate variability and its impacts. The classroom should be a welcoming environment where a variety of perspectives and lived experiences enable us to collaboratively define challenges and identify solutions.

The Climate School has adopted the Inclusive Excellence framework, which helps to build institutional capacity and effectiveness in promoting equal access, belonging, and success throughout our community.  The Office of Inclusive Excellence supports student access, success, education, and scholarship.

Should you wish to discuss the School's environment or your engagement inside or outside of the classroom, please do not hesitate to contact Willie Williams (Assistant Dean for Inclusive Excellence) at 332.277.3697 or[wwilliams@climate.columba.edu](mailto:wwilliams@climate.columba.edu).

***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 Climate School 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 solutions will be identified.

***Student Wellness Message***

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](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.

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](http://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](mailto:disability@columbia.edu). If you have already registered with ODS and would like your professor(s) notified of your recommended accommodations, contact Natalie Unwin-Kuruneri ([natalie@climate.columbia.edu](http://natalie@climate.columbia.edu/))

 who serves as the ODS liaison to the Climate School.

***Academic Integrity***

The Columbia Climate School holds each member of its community responsible for understanding and abiding by the [Academic Integrity and Community Standards](https://gsas.columbia.edu/content/academic-integrity-and-responsible-conduct-research.). Ignorance of policies and standards concerning academic dishonesty shall not be a defense in any disciplinary proceedings.

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) are particularly serious offenses and will be dealt with severely under [Dean’s Discipline](https://gsas.columbia.edu/content/deans-discipline). It is essential to the academic integrity and vitality of this community that graduate students 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.

Use of Artificial Intelligence (AI) is also subject to these integrity standards and must never be employed for a use that would be considered plagiarism if that material were sourced from a human author. For example, submitting work under your name that you did not personally write is not allowed, regardless of whether it was written by another author or by an AI tool.  Individual faculty may have different policies on other ways students may use or not use AI in a particular class. **Students should assume all use of AI is forbidden unless told otherwise.**

***Copyright Policy***

Please note -- Due to copyright restrictions, online access to this material is limited to instructors and students currently registered for this course. Please be advised that by clicking the link to the electronic materials in this course, you have read and accept the following:

*The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction.  One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.*

**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](https://www.college.columbia.edu/core/uwp/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](https://www.college.columbia.edu/core/uwp/writing-center). You can also contact the center at [writingcenter@columbia.edu](mailto:writingcenter@columbia.edu) with questions.