EESC BC3017 Environmental Data Analysis (Section 2)

Fall 2024
Mondays/Wednesdays 2:10PM–4:00PM
222 Milbank Hall

Mike He
Course Instructor
zh2263@columbia.edu
Office: 528 Building, Room 706 (Teachers College)
Office Hours: Mondays 4–6pm and by appointment

Elizabeth Cook
Course Instructor (Section 1)
ecook@barnard.edu
Office: 528 Building, Room 735 (Teachers College)
Office Hours: Tuesdays 4–6pm and by appointment (https://tinyurl.com/ApptsElizabeth)

Shuo Huang
Teaching Assistant
sh4208@columbia.edu
Office: 528 Building, Room 735 (Teachers College)
Office Hours: Wednesdays 4:15–6:15pm

Sachi Patel
Computing Fellow
sdp2173@barnard.edu
Office: Milstein 502 (CS Help Room)
Office Hours: Mondays 4:30–6pm

Course Overview

Environmental Data Analysis focuses on data acquisition, analysis, interpretation and presentation of environmental data. We will use publicly available and student-generated data sets to explore case studies of urban ecology, water quality, climate change impacts on the environment, and environmental justice in New York City. We will cover basic principles of working with data, statistics, and exploring spatial and temporal variability through labs and practical exercises with Excel, R, and the QGIS software.

Prerequisites: One year of college science or EESC V2100 or permission of the instructor.

Purpose of the Course

Data are the backbone of environmental science and decision-making. In professional, as well as in everyday life, we have to evaluate and interpret data that others provide, as well as generate and analyze data. This course uses case studies to teach various concepts and tools for analyzing and communicating with environmental data.

In this course, we will learn and practice a workflow from data collection and management to analysis and visualization of data with Excel, R, and QGIS; basic statistics; concepts of Geographic Information Systems (GIS) and spatial analyses; and how to interpret data, results, and communicate about your data.

The goal is to understand diverse methods for analyzing, interpreting, and visualizing data for use in your own research and especially your senior thesis!

Learning Objectives

At the end of this course, students will be able to:

- Manipulate, analyze, and visualize temporal and spatial environmental datasets in Excel, R, and QGIS
- Interpret basic statistical analyses and identify when they are appropriate for use in different cases
- Demonstrate basic statistical coding competency in R and spatial analyses in Geographic Information Systems (GIS)
- Formulate hypotheses and interpret data and results from environmental case studies
- Illustrate best practices for science communication and data visualization through scientific writing in lab reports, presentations, and data tables and figures.

Organization and Structure

The course follows a practical approach with a focus on learning through case studies. We will work through several case studies that integrate environmental (e.g., water quality), climate (e.g., temperature), and social (e.g., demographic) data.

(1) Hudson River water quality (descriptive statistics, uncertainty, significance tests)
(2) Changes of Hudson River water characteristics (time series, regression)
(3) Air pollution epidemiology and exposure assessment (generalized linear models, health impact assessment)
(4) Ecological amenities (urban green space) and dis-amenities (coastal flooding) in NYC communities (GIS, spatial analysis)

Each class is a combination of lecture and practical lab exercises. During the lecture, we will discuss basic concepts and theory, which will then be applied to the case study using Excel, R, and/or QGIS. The lab exercises offer practice and guidance in analyzing data for each case study.

Computers, Programs, and Data

Each student will need a desktop or laptop computer on which they have access to **Excel, R, RStudio, and QGIS**. R, RStudio, and QGIS are free programs that can be installed on both PC and Mac computers.

Our classroom (Milbank 222) has computers with all the required programs for student use during class, or you may work on your own laptop. For work outside of the classroom, the Empirical Reasoning Center (https://erc.barnard.edu/) and the Barnard Student Computer Labs (https://barnard.edu/student-computer-labs-printing) may be available for completing homework and quizzes.

R: A free, open-source statistical data analysis software package. https://www.r-project.org/

RStudio: A free, open-source integrated development environment to work with R. https://posit.co/products/open-source/rstudio/

QGIS: A free, open-source Geographic Information System (GIS). https://www.qgis.org/en/site/

Case studies, lecture slides, lab instructions, and the relevant publicly available data will be uploaded to course modules throughout the semester.

Course Materials

No textbook is required for this course. All readings will be available through the Barnard/Columbia CLIO library system or posted as a PDF in CourseWorks. We will also rely on online tutorials.

- McKillup (2012). *Statistics explained: An Introductory Guide for Life Sciences.* Cambridge University Press. 413 pages. ***Good explanation of statistics*****;** CU Library Link: https://clio.columbia.edu/catalog/13571179
- Jones (2022). *The R Book*, 3rd Edition. Wiley-Blackwell. 880 pages. ***Good explanation of statistics and introduction to R basics***. CU library link: https://clio.columbia.edu/catalog/16963931
- Dalgaard (2008). *Introductory statistics with R*, 2nd Edition. Springer Science + Business Media, LLC. 363 pp. ***Explanation of statistics, and good introduction into R basics****;* CU Library Link: https://clio.columbia.edu/catalog/7861320
- Qian (2017). *Environmental and Ecological Statistics with R,* 2nd Edition. CRC Press. ***Good R introduction, plus focus on basic and more advanced statistical analyses;*** CU Library link: https://clio.columbia.edu/catalog/14716539

Additional recommended references/resources (not required):

- Bolstad (2022). *GIS Fundamentals: A first text on Geographical Information Systems*, 7th Edition. Xanedu Press. 764 pages. ***Good explanation of GIS basics, plus QGIS exercises.***
- Heywood, Cornelius, and Carver (2012). *An Introduction to Geographical Information Systems*, 4th Edition. Prentice Hall. 480 pages. ***Good explanation of GIS fundamentals and basics.***
- Grolemund and Wickham (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd Edition. O'Reilly Media. 576 pages. ***Good introductory R textbook***

for data cleaning and basic data analysis + many free resources on the website. https://r4ds.hadley.nz/

Assessment and Grading

The final grade of the course will be based of your fulfillment of each of the following requirements:

Attendance and participation (10%): Students are expected to have read all the required readings before class, attend every class, and participate in lecture discussions and small groups during lab sessions. We follow a **strict attendance policy**: more than two absences will result in a grade reduction; each absence after the second absence will result in a 1% grade deduction, and more than four absences will result in a failing grade unless there are extenuating circumstances, such as an extended illness, family emergency, or observance of a religious holiday.

- If you will not be able to attend class, please notify me in advance. Unexcused absences will result in the deduction of participation points.
- Significant lateness will be considered an absence.
- If you register during the add/drop period, you are responsible for missed assignments.

That being said, if you are feeling ill, **please stay home and contact me as soon as possible** about your absence. You are responsible for content missed. Schedule virtual office hours to catch up as needed. In the case of prolonged absence(s) due to illness, we will work together to develop a plan to ensure your ability to complete the class work successfully.

Labs and take-home quizzes (60%): Each lecture will be accompanied by a lab posted online at the beginning of class. Labs will include a set of tasks and problems set to practice and gain skills in the software and statistical analyses covered throughout the semester.

- Lab sessions will be the second half of the class period. Students will be assigned to small groups and are encouraged to work together, share advice and tips, and discuss the assignment.
- While labs will be in small groups, **each student is responsible for completing and recording their own work for every assignment**.
- Lab assignments will not be submitted and graded in full, but questions from each lab will appear on the take home quiz. You will need to keep up with the lab assignments to successfully complete the take home quizzes.

Take-home quizzes will be posted Thursday afternoon and due the following Monday at 2pm ET. Quizzes will include approximately half of the problems from the in-class lab assignments (i.e., already completed during the week) and half new problems.

- **Quizzes must be completed individually**. Once the quiz has been opened on Canvas, working together is not allowed.
- Quizzes will focus on the recently covered topics and skills. The final quiz will be comprehensive and is designed to showcase what you have learned and how your scientific, data analysis, and critical reasoning skills have advanced.
- There will be six total quizzes throughout the course. You will have the opportunity to drop the lowest quiz grade or choose not to take the final quiz.

Data analysis report and presentation (30%): As part of the course, you are required to complete one lab report and an oral presentation.

The **data analysis report** will focus on Hudson River water quality and characteristics. It will be submitted first as a draft (75 points) with an opportunity for peer feedback (25 points) before the final draft (100 points) is submitted. The report will focus primarily on methods, results, and data communication through tables and figures to highlight your analyses and results.

For the **oral presentation**, students will have the opportunity to work in small groups to formulate a research question related to flood risk and environmental justice in NYC and conduct further independent analyses to address their question. Each small group will draft a one-page outline (20 points) and present their analyses during 10-minute class presentations (100 points), and there will also be an opportunity for peer evaluation (20 points).

More details and guidelines for both assignments will be provided later in the semester.

Late Submission Policy: All assignments and quizzes are expected to be submitted on the due date. For every day after the submission date, 10% of the maximum grade will be deducted from the score.

Statement on Academic Integrity

Use of Generative AI: AI is a collaborator and partner. "AI tools are useful to find background information about a topic, clarify challenging material, draft outlines, or improve grammar and style" (https://ctl.columbia.edu/resources-and-technology/resources/incorporating-generative-ai-teaching/). AI can also be a useful tool in data analysis and coding.

However, it is not a replacement for learning data analysis, learning to code, critical thinking, and you must be aware of the limitations. Be cautious in using AI for research assistance, as all existing AI apps 'fabricate' and supply false information and often operate at high levels of generality. In other words, don't trust everything the system says. Unless you already know the answer or can verify and corroborate with trusted sources (e.g., peer reviewed research papers), assume the AI answer is biased, lacks accountability, and often provides inaccurate information. In addition, please be aware that AI uses immense computing power and hence has an environmental footprint that we are just beginning to understand.

After you graduate, the workforce expects you to know how to use AI to increase efficiency and innovation. Therefore, we feel that college is the time to build your AI literacy. However, using AI can reduce the skill development inherent in independent thinking, reading, writing, and data analysis. **Using AI to complete lab or quiz assignments is not allowed, unless explicit instructions include AI.** Doing so would violate the college's Honor Code. Think for yourself and do not cheat yourself.

Barnard and Columbia have developed useful guides on generative AI and how to engage with it and maintain academic integrity. We encourage you to review the resources **here** and are happy to discuss them with you.

Barnard Student Honor Code: Approved by student body in 1912 and updated in 2020. The Code states: *We, the students of Barnard College, resolve to uphold the honor of the College by engaging with integrity in all of our academic pursuits. We affirm that academic integrity is the honorable creation and presentation of our own work. We acknowledge that it is our responsibility to seek clarification of proper forms of collaboration and use of academic resources in all*

assignments or exams. We consider academic integrity to include the proper use and care for all print, electronic, or other academic resources. We will respect the rights of others to engage in pursuit of learning in order to uphold our commitment to honor. We pledge to do all that is in our power to create a spirit of honesty and honor for its own sake.

The Honor Code governs all aspect of academic work. If a violation should arise, it will be reported to the Dean of Studies for appropriate action. Honor Board Guidelines on the procedures for implementing the Honor System and acting on charges of dishonesty can be found in the Student Handbook. Remember that Barnard students reaffirm their acceptance of the Honor Code by signing their registration form. Columbia students commit themselves to the Honor Code upon registering for a Barnard course.

Because the Honor Code is not entirely specific and contains qualifications and exceptions, such as "authorized by the instructor" or "approved by the instructor", please read the summary below of *What behaviors constitute academic dishonesty?*

What behaviors constitute academic dishonesty?

- Cheating on examinations, quizzes, tests, or other assignments: the giving of assistance to another or the receiving of assistance from another person, another examination paper, other written material, or any source not explicitly permitted by the instructor, is cheating. Thus, you may not look at another's paper or answers; you may not show your paper or answers to another or leave your paper or answers around for others to look at; and, you may not verbally read or reveal your answers to another in any way. It is also cheating to have access, without the instructor's approval, to examination, quiz, or test questions prior to the administration of the examination, quiz, or test.
- Plagiarism: the submission or presentation of ideas or work in any form that are not one's own without appropriate acknowledgement of the source(s). Even with the acknowledgement, close paraphrasing can constitute plagiarism. You may quote the work of others if properly attributed. Close paraphrasing also requires attribution; close paraphrasing is, however, a gray area on a slippery slope, and the slope tends to become steeper and more slippery with the length of the paraphrase.
- Submission of the same work for more than one course without the explicit permission of the instructors involved.
- Falsification or misrepresentation of data in any coursework.
- Altering, defacing, or concealing library materials.
- Participating in the academic dishonesty of another student by offering assistance or advice that encourages such behavior.
- Misrepresentation of one's state of health or personal situation to gain deferrals of examinations or extensions of academic deadlines.
- Forgery of a signature on any document or form related to a student's academic life, including the adviser's signature on a program, drop/withdrawal slip, or petition.
- Except for the above I encourage and expect students to share and work together, to ask questions, and to receive help from instructors and other students. Admittedly, there are gray areas but these gray areas will not be an issue if the intent of the foregoing is understood. If there are questions about the foregoing, it is prudent to ask your instructor for clarification.

Academic and Personal Well-being Resources

The College provides many resources to help students achieve academic excellence and navigate the pressures, burdens, and stressors you may be facing—whether personal, emotional, physical, financial, mental, or academic. We, as a community, urge you to make yourself your priority throughout the semester and your career. ***Sleep, exercise, and eat well!*** Below are resources to help you navigate your academic and personal needs. Please make use of them or contact me for additional guidance.

- **Barnard-Columbia Libraries:** https://library.barnard.edu/; https://library.columbia.edu/
- **Empirical Reasoning Center:** https://erc.barnard.edu/
- **Peer-to-Peer tutoring:** https://barnard.edu/academic-support/peer-to-peer-tutoring
- **Remote learning resources:** https://cep.barnard.edu/learning-resources-students
- **Center for Accessibility Resources and Disabilities Services (CARDS):** https://barnard.edu/disability-services
  For any student who needs academic accommodations, you are welcome to meet with me privately. All conversations will be kept confidential. Students who would like to request accommodations should meet with CARDS. CARDS will conduct an intake and, if appropriate, provide an academic accommodation letter. At that point, we will discuss the accommodations. Accommodations are not retroactive, so it is best to register with CARDS early each semester to access your accommodations.
- **Barnard Primary Care Health Service:** https://barnard.edu/primarycare
- **Barnard Rosemary Furman Counseling Center:** https://barnard.edu/furman-counseling/about-counseling
- **Barnard Feel Well Do Well:** https://barnard.edu/health-wellness

Note: The information in this syllabus is subject to change.

Course Schedule and Topics

This is a tentative schedule. It may change depending on how we progress.

	Date	Lecture Topic, *Lab topic*, and ***Due dates***
1	4 September, Wednesday	**Introduction** *Lab: Computer set-up & Excel basics*
2	9 September, Monday	**Units and Numbers** *Lab: Unit conversions and sea level rise*
3	11 September, Wednesday	**Hudson River Water Quality** *Lab: NYC Hudson River water quality data*
	13 Sep, Fri	***Deadline to add Fall classes***
	16 Sep, Mon	***Quiz 1 due – Monday, September 16, 2pm ET***
4	16 September, Monday	**Descriptive Statistics** *Lab: Hudson River descriptive statistics*
5	18 September, Wednesday	**Urban Street Tree & Air Quality (Field Sampling)** *Lab: Field measurements, recording data sheets*
6	23 September, Monday	**Data Analysis Report, Data Visualization - Tables & Figures** *Lab: Questions, hypotheses & making tables and figures; Download R & R Studio*
7	25 September, Wednesday	**Introduction to R** *Lab: Introduction to R coding*
	30 Sep, Mon	***Data Analysis Report first draft due – Monday, September 30, 2pm ET***
8	30 September, Monday	**Probability and Standard Error** *Lab: Calculating error on Hudson River data*
9	2 October, Wednesday	**Distributions and Confidence Intervals** *Lab: Distributions & testing assumptions; Lab report peer review*
	7 Oct, Mon	***Quiz 2 due – Monday, October 7, 2pm ET***
10	7 October, Monday	**Significance Tests: T-tests** *Lab: T-tests and ANOVAS with Hudson River data*
	8 Oct, Tues	***Deadline to drop Fall classes***
11	9 October, Wednesday	**Significance Tests: ANOVA and Time Series** *Lab: Correlations and time-series with Hudson River data*
	14 Oct, Mon	***Data Analysis Report peer review due– Monday, October 14, 2pm ET***
12	14 October, Monday	**Significance Tests: Correlations and Regressions** *Lab: Regressions with Hudson River data*
13	16 October, Wednesday	**Significance Tests: Multiple Regressions** *Lab: Multiple regression with Hudson River data*
	21 Oct, Mon	***Quiz 3 due – Monday, October 21, 2pm ET***
14	21 October, Monday	**Geographic Information System (GIS) Introduction** *Lab: Introduction to QGIS*

15	23 October, Wednesday	**Spatial: Making Maps, Projections; Data, Tables and Symbols** *Lab: Create a map*
16	28 October, Monday	**Guest Lecture: Shuo Huang** *Lab: NYC area and population*
17	30 October, Wednesday	**Air Pollution Modeling** *Lab: Air pollution predictions using machine learning*
	1 Nov, Fri	***Data Analysis Report final draft due – Friday, November 1, 2pm ET***
	4 – 5 Nov	***Election Day – No Class***
18	6 November, Wednesday	**Air Pollution and Health** *Lab: Air pollution health impact assessment*
19	11 November, Monday	**Spatial: Data Analyses & Manipulation – Part 1** *Lab: Spatial analysis – Part 1 – Clip, Erase, Split, Union*
20	13 November, Wednesday	**Spatial: Data Analyses & Manipulation – Part 2** *Lab: Spatial analysis – Part 2 - Proximity*
	14 Nov, Thurs	***Last day to withdraw & elect P/D/F***
	18 Nov, Mon	***Quiz 4 due – Monday, November 18, 11:59pm ET***
21	18 November, Monday	**Spatial: Raster Data – Introduction** *Lab: Maps with raster data*
22	20 November, Wednesday	**Spatial: Raster Analysis Tools** *Lab: Zonal statistics analysis*
	25 Nov, Mon	***Presentation Outline due – Monday, November 25, 2pm ET***
23	25 November, Monday	**Spatial: Interpolations** *Lab: Interpolation analyses*
	27 – 29 Nov	***Thanksgiving Holiday – No Class***
	2 Dec, Mon	***Quiz 5 due – Friday, December 6, 2pm ET***
24	2 December, Monday	**Introduction to Spatial Analysis in R** *Lab: Spatial analysis with R*
25	4 December, Wednesday	**Work on Final Presentations / Make-up Class**
	9 Dec, Mon	***Final Presentations due – Sunday, December 8, 11:59pm ET***
26	9 December, Monday	**Final Presentations and Review**
	13 – 20 Dec	**Exam Period** ***Quiz 6 due – Monday, December 16, 2pm ET***