Photo by Conny Schneider on Unsplash

Spring 2026
ENV H 490/590

Department of Environmental & Occupational Health Sciences

Data for Action:
Data Science in Public Health

Learn Applied Data Science Skills

Looking to master data science methods used in public health?

ENVH 490/590 is a hands-on course designed to move you beyond spreadsheets and into R programming, data visualization, and reproducible research.

Whether you are an undergraduate building a foundation or a graduate student ready to analyze your own datasets, this course provides both background knowledge of the method and practical experiences to sharpen your skills.

This course provide you with an introduction to R programming, and covers the essentials of developing Data Science analysis "pipelines", including (1) importing data, (2) data management (quality checks, cleaning, and the concept of tidy data), (3) data engineering (data feature design and dimension reduction), (4) data modeling, (5) data communication (visualization using tables and figures), and (6) data/code sharing to improve the rigor and reproducibility of data analyses. Data privacy and other data ethical concerns will also be covered in the the course.

The course labs will provide hands-on exercises involving common analyses in bioinformatics, epidemiology, environmental exposure assessment, natural language processing, and solving classification and regression problems.

While the course's case studies utilize real-world public health datasets, the skills are broadly applicable to other fields.

Learn reproducible analysis pipelines for:

- Data import and management
- Data visualization
- Basic statistical tests
- Multivariate regression models
- Machine Learning methods
- Generative AI tools

Instructors

Edmund Seto
Professor

Department of Environmental & Occupational Health Sciences

Tzu-Hsin Karen Chen
Assistant Professor

Department of Urban Design and Planning

Department of Environmental & Occupational Health Sciences