

Class Sessions

Instructor

Teaching Fellow

Teaching Assistant

Guest Lecturers:

Course Description

Prerequisites

Course Learning Objectives

Assessment and Grading Policy

Course Requirements

Class Components:

Class Flow

Homework Assignments:

Final Assignment:

Syntax Inventory:

Data Sources

Academic Integrity

Disability Access

Course Software

Course Reading and Textbooks

Course Schedule

Analysis of Environmental Health Data (P6360)

Spring 2021

Class Sessions

All class sessions will be held online using zoom.

<https://columbiacuimc.zoom.us/j/95808076529>
(<https://columbiacuimc.zoom.us/j/95808076529>)

passcode: 887110

Instructor

Dr. Julie Herbstman, PhD ScM (jh2678@cumc.columbia.edu (mailto:jh2678@cumc.columbia.edu))

- (now: home in NJ)
- Office hours by appointment; will set up zoom link

Teaching Fellow

Ahlam Abuawad, MPH (aka2170@cumc.columbia.edu (mailto:aka2170@cumc.columbia.edu))

Teaching Assistant

Seonyoung (Shannon) Park (sp3804@cumc.columbia.edu (mailto:sp3804@cumc.columbia.edu))

- TA Office Hours:

TBD

Guest Lecturers:

Dr. Miranda Spratlen, PhD MHS (mjs2376@cumc.columbia.edu (mailto:mjs2376@cumc.columbia.edu))

Dr. Annie Nigra, PhD MHS (aen2136@cumc.columbia.edu (mailto:aen2136@cumc.columbia.edu))

Course Description

Data is the foundation of all research and becoming comfortable reading about, describing, analyzing, interpreting, summarizing and presenting EHS data is critical for the success of all environmental health scientists. Therefore, all EHS MPH graduates should know how to work with the types of data they will typically encounter in EHS-related research settings. This course will teach students how to work with these data at a fundamental level which includes a focus on data configuration, organization, and preparation as well as learning to apply techniques to generate, describe, summarize, and conduct basic analyses relevant to all EHS certificates.

Classroom data are like teddy bears; real data are like a grizzly with salmon blood dripping out its mouth. (@JennyBryan)

We will use real datasets to illustrate analytic techniques, common problems, and approaches typical to environmental/molecular epidemiology, toxicology, EHS policy, air pollution and climate time-series.

Prerequisites

Completion of the MPH Core (Semester 1) or equivalent

Course Learning Objectives

Students who successfully complete this course will be able to:

- Identify key steps and “best practices” in analyzing various forms of data typically encountered in EHS-related settings
- Configure, organize, and prepare molecular epidemiology, toxicology, EHS policy, and climate-related data for analysis.
- Apply the appropriate analytic techniques to describe, summarize, and conduct basic analyses relevant to these research areas.

Assessment and Grading Policy

- Assignments (weekly; 13 total, highest 12 will be counted): 65%
- Weekly “learning check”: 10%
- Final Assignment: 15%
- Syntax Inventory: 10%

Course Requirements

Class Components:

For each module (week), there will be:

Lecture (ASYNCHRONOUS): this provides the foundation for and the context for the new material. This will be recorded and posted on courseworks at least 5 days prior (Thursday) before the Monday class.

Learning Check (ASYNCHRONOUS) after each lecture, there will be a “learning check” which can be found in the “quizzes” link on courseworks. There will be 3-5 short questions that capture the essence of the material from the lecture. Learning Checks are due by noon each Monday.

Recap and Demonstration (SYNCHRONOUS): this will begin at 2p on Mondays via zoom and can be expected to last ~90 minutes.

Homework (ASYNCHRONOUS): weekly homework assignments have been designed to allow students to practice and apply the concepts and skills presented during the demonstration. Homework is due by noon each Monday

Homework answers (ASYNCHRONOUS): at 1p on Mondays, the “answers” to the homework file will be posted, as well as a recording of us going through it. This is an opportunity for you to compare your answers/approaches to ours.

Class Flow

By Thursday:

Lecture and Learning Check posted (via courseworks)

On Mondays:

12 noon: Homework due (via courseworks); Learning Check due (via courseworks)

1p: Homework “answers” released (via courseworks: recording, RMD, and HTML file will be posted); you should review before 2p.

2p: Live session begins (via zoom)

2p-2:10p: Housekeeping issues; answer any questions about homework that was due

2:10p-2:30p: Review weekly learning check and go over main points of new material (from recorded lecture)

2:30p-3:30p: Live demonstration of new material

~3:30p: Live session ends; new homework released

Homework Assignments:

Homework exercises will cover the material introduced during the class session. Students are expected to submit written responses to the homework exercise (via Courseworks) by noon of the next class session. Typically, no late assignments will be accepted; however, given the circumstances of the pandemic, please reach out to me (Julie) about exceptions, which will be made on a case-by-case basis.

For each homework assignment, students will be expected to submit an HTML file (generated using R Markdown). This file will not render if there are errors in your code! For each assignment, students will receive 1 point for generating and .html file. If you cannot fix all the errors (and therefore cannot render an .html file) but would like partial credit, you may submit your R-markdown script (.rmd) and also a word file (.doc) with the responses to the HW questions. [More detail will be provided in Class #1]. The remainder of each homework grade will be based on responses to the questions on a 5-point scale. Partial credit may be assigned.

Late Assignment Policy:

Late submission of homework assignments will typically not be accepted without approval by course instructor (see note above).

Students may work together for all assignments but homework must be prepared and submitted independently (i.e., must be your own work)

How to submit assignments:

All course exercises are to be turned in electronically via the course website:

<https://courseworks2.columbia.edu/courses/78801>

(<https://courseworks2.columbia.edu/courses/78801>)

The *final assignment* and *syntax inventory* (see below) must be submitted both through courseworks and through GIT.

Please remember to include your name in the filename on each submission (“herbstman_hw1.html”)

Final Assignment:

This assignment will build off of the previous homework assignments. Students will have 2 weeks to complete it. Students will be asked to conduct an analysis using the CDC NHANES data. Students will be guided through this analysis by answering a discrete set of questions that will be provided. The format and grading schematic of this assignment will be similar to that of the homeworks.

Syntax Inventory:

At the end of this course, you will have 12 markdown .html files (from homework assignments). These are meant to serve as a resource to you in the future. Therefore, you should go over your homeworks, make corrections, additional notes, etc. and collect them in a folder. You will be expected to create a searchable syntax inventory (with examples) that describe each of the new commands used each week (this will constitute 10% of your final grade). The goal is for you to have a self-created reference that you can use in the future.

It is strongly recommended that you correct mistakes in your own homework each week and work on the inventory each week throughout the semester. You can do it at the end but it will be a lot more work!

Data Sources

New York City Health and Nutrition Survey (NYC HANES):

<http://www.nyc.gov/html/doh/html/hanes/datasets.shtml>

(<http://www.nyc.gov/html/doh/html/hanes/datasets.shtml>)

NYC Neighborhoods Asthma and Allergy Study (NYC NAAS; courtesy of Dr. M. Perzanowski) (Relevant de-identified data will be posted on Courseworks)

Prospective Yugoslavia data (Yugo; courtesy of Dr. J. Graziano and Dr. P. Factor-Litvak) (Relevant de-identified data will be posted on Courseworks)

Toxicological Data (Tox; courtesy of TBD) (Relevant data will be posted on Courseworks)

The National Morbidity, Mortality, and Air Pollution Study (NMMAPS)

<http://www.ihapss.jhsph.edu/index.htm> (<http://www.ihapss.jhsph.edu/index.htm>)

National Health and Nutrition Survey (NHANES), final assignment

<http://www.cdc.gov/nchs/nhanes.htm> (<http://www.cdc.gov/nchs/nhanes.htm>)

MAILMAN SCHOOL POLICIES AND EXPECTATIONS Students and faculty have a shared commitment to the School's mission, values and oath. <http://mailman.columbia.edu/about-us/school-mission/> (<http://mailman.columbia.edu/about-us/school-mission/>)

Academic Integrity

Students are required to adhere to the Mailman School Honor Code, available online at <http://mailman.columbia.edu/honorcode> (<http://mailman.columbia.edu/honorcode>).

Disability Access

In order to receive disability-related academic accommodations, students must first be registered with the Office of Disability Services (ODS). 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, please speak to your instructor to ensure that s/he has been notified of your recommended accommodations by Nina Kulacki and/or the School's liaison to the Office of Disability Services.

Course Software

As described in EHS department and Mailman orientation materials, all EHS students are expected to have laptop computers. This course will primarily use the software R, which is freely available and can be downloaded from the internet: <http://cran.r-project.org> (<http://cran.r-project.org>) and the user interface RStudio, which is also freely available and can be downloaded from the internet: <http://www.rstudio.com/download/> (<http://www.rstudio.com/download/>). We will also use Microsoft Excel (which can be downloaded for free with your UNI from the CUIT website) and some other web-based program (e.g, Qualtrics), that are available with your UNI login.

Course Reading and Textbooks

There are no required textbooks for this course. There are many, many resources available online to assist with R. It is anticipated that the majority of students' course-related effort outside the classroom will be spend completing assignments.

Course Schedule

1/11: Class 1

- Introduction/Orientation

(tidyverse, R Markdown, ggplot)

1/18: No class: Martin Luther King Jr. Holiday

1/25: Class 2

- Data Wrangling I

(dplyr: filter/arrange/rename)

2/1: Class 3

- Data Wrangling II

(dplyr: group_by/mutate/summarize)

2/8: Class 4

- Lab Data; Log-transformation, Linear regression, merging datasets
-

2/15: Class 5

Please note that it is President's Day but class will meet (per OSA)

- Survey Data; using qualtrics (Julie)
-

2/22: Class 6

Lead Instructor: Dr. Miranda Spratlen

- Missingness and Simple Imputation
-

3/1: Class 7

Please note that Spring Break begins on Tuesday 3/2 (per OSA)

- Sampling, Re-sampling & Power
-

3/8: No class: Spring Break

3/15: Class 8

Lead Instructor: Ahlam Abuawad

- Toxicology
-

3/22: Class 9**Lead Instructor: Dr. Annie Nigra**

- Survey Weights
-

3/29: Class 10

- Measures of Associations
 - Focus on adjustment (multivariate model, dummy variable/reference group)
-

4/5: Class 11

- Measures of Associations
 - Focus on logistic regression, interaction
-

4/12: Class 12

- Time series
-

4/19: Class 13

- Identifying anomalies