Environmental Data Analysis and Visualization

ENVS 496/696 | Spring 2021 | W 11:20 AM - 2:10 PM | 3 credits

Overview

This course covers the non-statistical aspects of the data life cycle, including how to store, clean, visualize and communicate data (Figure 1). It is intended as a complement to statistics courses - we will cover how to get your data into shape for analysis, and how to communicate your findings visually. It is primarily a methods class and will be taught in R (but there is no expectation that students know R coming in).

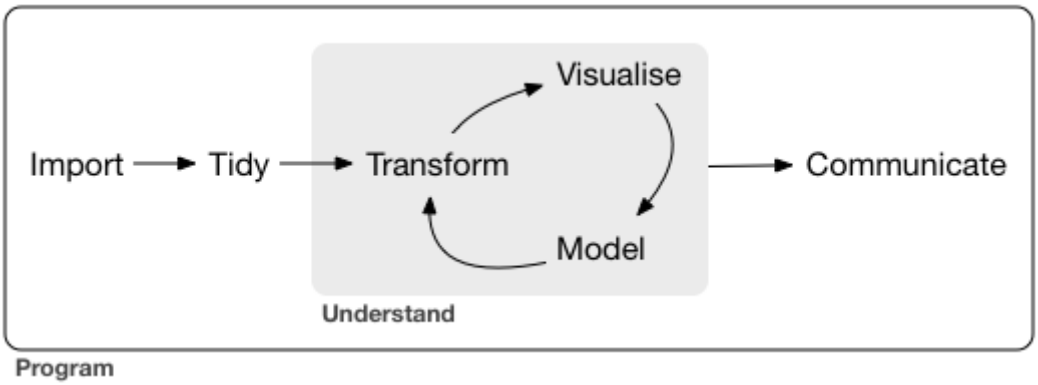

Figure 1: The data life-cycle (figure from Grolemund and Wickham)

Instructor: Dr. Jessica Gephart is an environmental scientist focused on global seafood trade

Email: jgephart@american.edu

Office: Hall of Science 322

Office hours: M 3:00-5:00 pm (https://american.zoom.us/my/jgephart) and by appointment (https://doodle.com/mm/jessicaagephart/calendar)

Hacky hour W 2:10-3:00 - Hang around after class to work on code individually, in groups or ask questions

How I will contact you: My communication to you outside of class will take place via announcements and email.

Course material

Our primary course material will be **R for Data Science** by Garrett Grolemund and Hadley Wickham, which is available for free online (http://r4ds.had.co.nz/). I will post other readings to Canvas in that week's reading section. Readings should be viewed as a resource to accompany the exercises in class and to support your final project. It will be helpful for you to review the readings ahead of the class session and return to readings as needed. This book includes code that you may find useful for additional practice.

Objectives

By completing this course, students will be able to:
1) Interpret figures in scientific papers and popular media
2) Locate data relevant to biological and environmental questions
3) Understand the steps linking raw data to communicated findings
4) Create exploratory and publication-worthy graphs

Structure of the course

This class combines lectures on topics in data science and in-class exercises.The course is centered around an individual final project where you will demonstrate the data visualization skills you gain through this course.

Class assignments and requirements

There are two main components required for successful completion of the course.

A. Problem sets Problem sets are designed to develop the skills you learn in class and to gain comfort in the R environment through practice. Problem sets will typically include designing and implementing code and interpreting code and figures. There will be four problem sets, to be submitted on Canvas before class on the day they are due. Students are encouraged to collaborate on problem sets, but must complete their own work (i.e., you need to write and fully understand your own code at all points, as the skills and concepts will build upon each other throughout the course).

B. Final project The focal experience of the class will be to develop a research project that addresses a biological or environmental question with data. Students will be expected to identify a question, contextualize the question with a literature review, locate data relevant to answering the question, and interpret and communicate that data with a workflow in R. It is expected that students integrate data from multiple sources. You can work with me to identify public data sources or you can work on your own data, which may be part of a thesis project.

Grade allocation

Grading will be based on a total of 250 points, where 90% of the points will earn an A, 80% a B, etc. Participation will reflect attendance and involvement in discussion and in-class exercises. The breakdown by assignments is as follows:

Assignment	Points
5 Assignments (20 pts each)	100
Final project	120 total (breakdown below)
I. Topic proposal	5
II. Data source identification	10
III. Literature review and workflow plan	15
IV. Peer review	15
V. Final report	50
VI. Final presentation	25
Participation	30

Policies

1) Please note that assignments are due to Canvas before class. It is important to turn assignments in on time because skills and concepts will build upon each other throughout the course. You therefore need to practice the skills from each class to prepare for the next class. If an assignment is late, I will deduct 10% of the total points allocated to that assignment, and I will deduct 10% for each additional late day, up to a deduction of 50% of the possible points.

2) All missed classes need to be approved with the instructor prior to the start of class. Unexcused absences will result in the deduction of participation points.

3) This class includes frequent in-class exercises and periodic workdays. I expect all students to actively participate in exercises and discussions. Lack of engagement will result in the deduction of participation points.

4) We will follow school policy of plagiarism and academic dishonesty. All students need to be familiar with the Student Conduct Code (https://www.american.edu/policies/students/student-conduct-code.cfm).

Deadlines

Assignments are due by the end of the day on:

2/10 Problem set 1: Intro to R

2/17 Problem set 2: ggplot problem set

2/24 Data visualization critiques

3/17 Problem set 3: Tidy data problem set 1

3/24 Problem set 4: Tidy data problem set 2

Final project due dates are:

2/17 Topic proposal

3/03 Data sources identification

3/17 Literature review and workflow plan

3/31 First draft for peer review (bring to class)

4/07 Peer review

4/21 Final report and final presentations

Course topics and tentative schedule

Topics and lessons generally correspond to the noted chapter numbers in the book.

Week	Date	Topic	Lesson & Readings
-	**W 1/20**	**Inauguration Day**	**No class**
1	W 1/27	General	Course introduction
2	W 2/03	General	Getting started in R
3	W 2/10	Visualize	Visualizing data with `ggplot2`
4	W 2/17	Visualize	What makes a good data visualization? (asynchronous)
5	W 2/24	Visualize	RMarkdown
6	W 3/03	Tidy	Tidy data part 1
7	W 3/10	Tidy	*Wellness week work day*

Week	Date	Topic	Lesson & Readings
8	W 3/17	Tidy	Tidy data part 2
9	W 3/24	Visualize	Mapping
10	W 3/31	Visualize	Odds and ends
11	W 4/07	Transform	*Work day*
12	W 4/14	Wrangle	R Shiny
13	W 4/21	Wrangle	Final presentations