Environmental Data Management and Visualization

ENVS 410/510 | Winter 2018 | Klamath 5 | MW 10:00 - 11:50 am | 4 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).

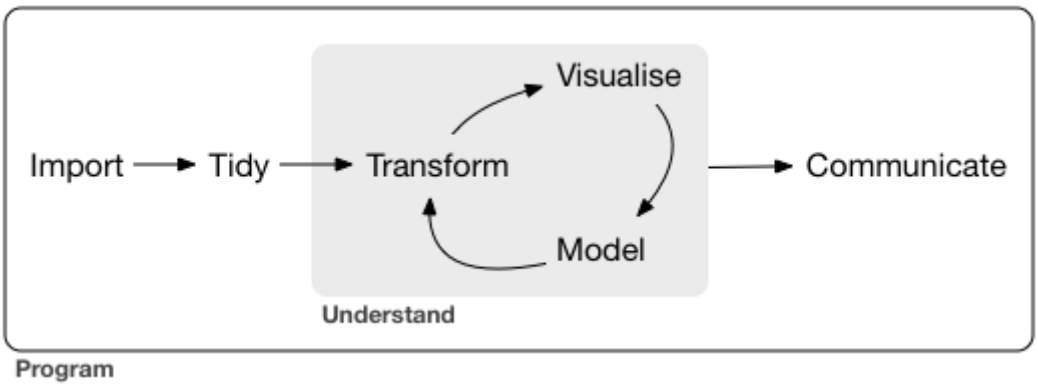

Figure 1: The data life-cycle (figure from Grolemund and Wickham)

This course satisfies the "Analytical Approaches" requirement for Environmental Science majors. MATH 252 or equivalent is a prerequisite.

Instructor: **Dr Lauren Hallett** is a plant community ecologist specializing in ecological restoration.
Email: hallett@uoregon.edu *Please include ENVS410/510 in email subject lines*
Office: 288 Onyx Bridge
Office hours: M 1-2 pm, Tu 11-12 am, or by appointment

GTF: **Geoffrey Johnson** studies long-term environmental change and ecosystem management policies.
Email: gmj@uoregon.edu
Office: 247 Columbia
Office hours: Fr 2-3 pm, or by appointment

Canvas site: Our website is accessible via the UO Canvas server, use your UO email and password to access the site. Problem sets will be distributed and submitted via Canvas. https://canvas.uoregon.edu/

How we will contact you: Our communication to you outside of class will take place via Canvas 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. We will post other readings to Canvas. Please make sure to read book sections and papers before the class in which they are assigned.

Objectives

By completing this course, students will be able to:
1) Interpret figures in scientific papers and popular media
2) Locate data relevant to 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.

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. Problem sets will typically include designing and implementing code and interpreting code and figures. Comfort in the R environment comes with regular use. Consequently, problem sets will be assigned on a weekly basis. They will be posted to Canvas after class Wednesday and will be due on Canvas prior to class the following Wednesday.

B. Final project The focal experience of the class be to develop a research project that addresses an 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. In general, students will be using public data, but if you have a data project of your own (particularly graduate students) we can discuss its suitability for this requirement.

Grade allocation

Grading will be based on a total of 200 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
9 Problem sets (10 pts each)	90
Final project	
Part I: Literature review and proposed workflow	30
Part II: Peer review	20
Part III: Final paper	40
Participation	20

Policies

1) Please note that problem sets are due Wednesday before class and progress toward the final project is generally due Mondays before class. If an assignment is late, we will deduct 10% of the total points allocated to that assignment, and we will deduct 10% for each additional late day.
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. We expect all students to actively participate in exercises and discussions.

4) We will follow school policy of plagiarism and academic dishonesty. All students need to be familiar with the Student Conduct Code (https://policies.uoregon.edu/vol-3-administration-student-affairs/ch-1-conduct/student-conduct-code).

5) Mac laptops will be provided for in-class use. The laptops wipe all personal data between classes and whenever you are logged off. Please don't log off the computers, and make sure to back up your work at the end of class. If you wish to use your own laptop please make sure to have R, RStudio and TeX installed. Most UO computers in computing labs have these programs installed, but please talk to us if you have trouble accessing computers or software out of class.

Deadlines

Problem sets are due weekly before class on Wednesday. Final project due dates are:
M 1/22 Topic proposed
M 2/5 Data sources identified
M 2/26 Literature review and workflow plan
M 3/5 Peer review
W 3/21 Final project

Course topics and tentative schedule

The topics on the tentative outline are subject to change. This is a guess, but we will take as long as needed on each lesson. Topics and lessons generally correspond to the noted chapter numbers in the book, additional readings will be posted to Canvas and emailed the week prior to when they should be read.

Day	Date	Topic	Lesson
1	M 1/8	Visualize	Plotting before analyzing
2	W 1/10	General	Overview of R, R Studio
-	*M 1/15*	*MLK day*	*no class*
3	W 1/17	Visualize	ggplot2: aesthetic mapping and facets 3.1-3.5
4	M 1/22	Visualize	ggplot2: geometric objects 3.6-3.7
5	W 1/24	Visualize	ggplot2: coordinate systems and scales 3.8-3.10
6	M 1/29	General	Environmental data
7	W 1/31	Transform	Rearranging data 5.1-5.5
8	M 2/5	Transform	Grouping and summarizing data 5.6-5.7
9	W 2/7	Transform	Workflows and the pipeline 6.1-6.3, 18.1-18.3
10	M 2/12	Transform	Relational data and joins 13.1-13.7
11	W 2/14	Wrangle	Importing data 11.1-11.6
12	M 2/19	Wrangle	Tidy data 12.1-12.2, 12.7
13	W 2/21	Wrangle	Spreading and gathering 12.3
14	M 2/26	Wrangle	Separating and uniting 12.4-12.6
15	W 2/28	Communicate	Rmarkdown 27.1-27.7, 29.1-10, 30
16	M 3/5	Communicate	Good vs bad graphs 1
17	W 3/7	Communicate	Good versus bad graphs 2
18	M 3/12	Communicate	ggplot2: labels and color 28.1-28.3
19	W 3/14	Communicate	ggplot2: scales and themes 28.4-28.8