Advanced R Programming: BIOL 6064

Fall 2023

Instructors:

Abby Lewis (she/her): aslewis@vt.edu
Dr. Cayelan Carey (she/her): cayelan@vt.edu

Class location and time: We will meet in Derring 1065 from 11:30–12:20 on Thursdays

Office Hours: by appointment; please email or ask in class to set a time to meet.

GitHub repository: https://github.com/abbylewis/Advanced_R

Canvas: https://canvas.vt.edu/courses/175241

Welcome! We are excited to explore advanced
R topics with you all this semester!

Overview

In this seminar, we will discuss advanced topics in data science using the R statistical programming language, with biological and ecological applications. Topics will vary based on student interests, but will likely include code efficiency, functional programming, and reproducible workflows.

Learning objectives

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

- Demonstrate an advanced understanding of R fundamentals (e.g., object types, environments, data storage)
- Predict which operations in R will be slow or memory-intensive, and describe ways of detecting and addressing these bottlenecks
- Identify technologies that enable collaborative coding and reproducibility
- Collaboratively create best practices for code efficiency, reproducibility, and documentation
- Explore and understand other advanced R topics as desired based on student interest

Is this class a good fit for you?

The vast majority of R users are not computer scientists. For many of us in environmental science, R is used as a powerful tool to accomplish data analysis tasks, using datasets of relatively small/moderate size. Optimizing efficiency or memory allocation can often take more time than it saves!

In this class, we want to bring together a community of graduate students who are interested in going a step or two further. This class is targeted to students who are interested in thinking about some of the computer science underpinning data analysis in R, in order to increase the efficiency and reproducibility of their code. Consequently, this class will assume a basic level of comfort with R skills already covered in courses that introduce students to R (e.g., basic data wrangling, figure making, and statistics).

Examples of classes in the VT Biological Sciences department that could provide this foundational R training include *Ecosystem Dynamics*, *Quantitative Community Ecology*, and *Working with Ecological and Evolutionary Data*. We also recognize that a lot of R learning happens outside of the classroom, so we note that these Biological Sciences courses are not necessarily prerequisites for this class.

When deciding if this class would be a good fit for your interests and expertise, here are a few questions to consider:

- Are you comfortable loading data and working with data frames in R (e.g., applying transformations to columns, adding new columns, summarizing based on a grouping variable)?
- Are you comfortable using if statements and for loops and writing your own functions?
- If you come across a function you are not familiar with, are you comfortable reading and using the documentation for that function?
- Have you run into cases where a certain R analysis was excessively time- or memory-intensive? Are you interested in learning about how R works under the hood (e.g., computer memory allocation), to make your code more efficient?
- Are you interested in sharing code as part of collaborative projects? Do you want to learn and develop best practices for doing this efficiently?

If you answered yes to most of these questions, then this class should be a good fit for you!

Respect and welcome for all

We ask that you all work with us to encourage a climate that supports the learning and growth of every individual. All students are entitled to a safe, welcoming, and respectful environment free from prejudice, bullying, discrimination, harassment, and bias. By participating in this course, you agree to be respectful of diversity across race, color, religion, nationality, culture, gender, age, gender identity, sexual orientation, disability, socio-economic condition, and all other aspects of individual identity and experience. If any student feels this respect has been violated due to these attributes, they should notify both instructors immediately.

Services for students with disabilities

If you need accommodation because of a disability (learning disability, attention deficit disorder, psychological, physical, etc.), please send Cayelan and Abby an email with SSD documentation during the first week of class—we will work with you to ensure that the class provides an inclusive learning environment around your needs.

Well-Being Commitment

Virginia Tech is committed to protecting the health and safety of all members of its community. By participating in this class, all students agree to abide by the Virginia Tech Wellness principles. See: https://www.vt.edu/public-health.html for more information. Following university policy, please do not attend class if you test positive for COVID or have COVID symptoms ; as long as you email your instructors, your absence will be excused.

Attendance

We have taken a number of steps to minimize out-of-class responsibilities for students in this course as a 1-credit graduate seminar. *Consequently, it is imperative that you attend and participate during our class time together.*

Your participation grade in this course is calculated from both your engagement in class (questions, participation in group work, etc.) as well as the number of unexcused absences. The rubric for your class participation grades will be posted on Canvas under “Assignments.” An excused absence (see dos.vt.edu/our_services.html) will NOT harm your lecture or lab participation grades. However, unexcused absences will decrease your participation grade.

We encourage you to reach out to your instructors if you have any questions, and we will work with students on a case-by-case basis if unique personal situations arise.

CLASS LOGISTICS

Texts

This class will be based around the *Advanced R* textbook, written by Hadley Wickham, which is a great resource for diving in depth into some advanced concepts in R. The book is freely available online here (with solutions available here). You are welcome to purchase and use a physical copy (second edition) if you would like, but it is not required for the course. Throughout the semester, we will likely also draw from other resources, depending on student interest.

Other useful resources related to the material in this course:

- Course on R debugging and robust programming by Laurent Gatto & Robert Stojnic
- Data Challenge Lab by Hadley Wickham
- R for Data Science by Garrett Grolemund & Hadley Wickham, and some solutions
- R packages by Hadley Wickham
- Efficient R programming by Colin Gillespie & Robin Lovelace
- R Programming for Data Science by Roger D. Peng
- Mastering Software Development in R by Roger D. Peng, Sean Kross and Brooke Anderson

GitHub

Class materials and resources will be hosted on GitHub here.

Honor System

We trust and expect that you will all be honest about the work you have completed, as the goal of this course is to further your own learning. Students enrolled in this course are responsible for abiding by the Honor Code. If you have doubts about how the Honor Code applies to any assignment, you are responsible for obtaining specific guidance from the course instructors *before* submitting the assignment. Ignorance of the rules does not exclude any member of the University community from the requirements and expectations of the Honor Code. For additional information about the Honor Code, please visit: https://www.honorsystem.vt.edu/

Please ask us if you have any questions about balancing independent vs. group work for any assignments. We have a zero-tolerance policy for cheating and will give a zero to any assignment in which a student tries to pass off someone else's work as their own. *Importantly, the use of generative AI tools (e.g., ChatGPT) is not permitted in this class unless specifically discussed with the instructors in advance. Use of generative AI may be considered a violation of the Honor Code, as the work is not your own.*

Course Schedule

Draft course schedule, subject to change based on student interests and R experience, availability of guest speakers, and other factors.

A note on readings: recognizing that this is a 1-credit seminar, it is not required that you do the readings before coming to class. However, these readings may be beneficial, especially if you are not familiar with a week's topic. Readings will form the basis for in-class activities, and are included as references for students who want to reinforce learning done in class.

Week #	Date	Contents	Optional reading
1	24 Aug	*In lieu of meeting in the first week, please complete this survey by 5:00PM on August 25th to assess interests and experience.*	
2	31 Aug	Course introduction and expectations Git basics, RStudio integration Warm-up exercise	
3	7 Sep	Memory	Wickham Ch. 2
4	14 Sep	Memory	
5	21 Sep	Functions (*Class check-in survey*)	Wickham Ch. 6 W3 schools tutorial
6	28 Sep	Functions, environments	Wickham Ch. 7 For more advanced topics: Wickham Ch. 9–11
7	5 Oct	Debugging, functional programming	For more advanced topics: Wickham Ch. 9–11, Ch. 22
8	12 Oct	Parallel processing, brainstorm end-of-semester presentation ideas	Peng Ch. 22
9	19 Oct	*tidyverse* and *ggplot2*	Basics: Wickham ggplot2 Ch. 2 Advanced topics: Wickham ggplot2 Ch. 18–21

10	26 Oct	Code review introduction, GitHub project management	
11	2 Nov	Code review exercise	
12	9 Nov	R Shiny	
13	16 Nov	Student presentations on additional topics of your choice	
	23 Nov	– no class, Thanksgiving break –	
14	30 Nov	Student presentations on additional topics of your choice	

Assessment

As a 1-credit graduate seminar, this course is primarily centered around in-class participation. We expect students to be present and participate in class discussions every week. Additionally, the course will include in-class coding assignments to practice the skills we discuss, a code review assignment (week 10), and presentations on an advanced R topic of your choice (weeks 13–14). Rubrics for all assessments will be provided on Canvas.

Overall course grades will be calculated as below:

Assessment	**Percent of overall course grade**
Participation	80%
Code review assignment (week 10)	5%
Final presentation (weeks 13 and 14)	15%
Total	*100%*

Grading modality

This seminar course can either be taken as A–F or P/F, depending on student preference. The default for the course on HokieSpa is A–F, but due to the participation-based format of the course, we strongly encourage you to consider switching to P/F. Before the start of the semester, this can be done from the “Schedule and Options” tab of the course registration website. After the start of the semester, you need to contact the Office of the University Registrar. Please reach out to Abby and Cayelan if you have any questions.

Feedback

If you have feedback on the materials, activities, or approaches used in this class, please let us know so we can take steps to ensure the course meets your needs! Because we need to keep the interest of all students in mind, we cannot promise that we will change the course, but we do promise to listen and consider your suggestions.