ERHS 535: R PROGRAMMING FOR RESEARCH

Fall 2023

Instructor: Rachel Severson	**Time:** M/W 10:00 am – 11:00 am
Email: rachel.severson@colostate.edu	**Place:** Microbiology B120

Online Coursebook: The online coursebook for this course is available at:
https://geanders.github.io/RProgrammingForResearch/

Office Hours: M/W 11:00 am – 12:00 pm in Microbiology B120, F 12:00 pm – 1:00 pm in AZE 120

Time and place: This course meets in Room B120 of the Microbiology Building on Mondays and Wednesdays, 10:00 am – 11:00 am. Exceptions to these meeting times are:

- There will be no meeting on Labor Day.
- There will be no meetings the week of Thanksgiving.

Detailed Schedule: A detailed schedule for this course is available on the online coursebook, in the chapter on "Course Information."

Grading: Course grades will be determined by the following five components:

Assessment component	Percent of grade
Final group project	30
Weekly in-class quizzes	25
Homework	25
Attendance and class participation	10
Weekly in-course group exercises	10

Attendance and class participation: Because so much of this class is through interactive work in class, it is critical that you come to class. Out of a possible 10 points for class attendance, you will get:
- **10 points** if you miss two or fewer classes
- **8 points** if you miss three classes
- **6 points** if you miss four classes
- **4 points** if you miss five classes

- **2 points** if you miss six classes
- **0 points** if you miss seven or more classes

Exceptions:
- Attendance on the first week of class will not be counted.
- If you miss classes for "University-sanctioned" activities. These can include activities like attending a conference or travel to collect data for your dissertation. For these absences, you must inform me prior to the date that you will be absence. No points will be lost for attendance if you provide a signed letter from your research advisor by Dec. 13, 2023 (start of finals week), and you can make arrangements with me to make up any missed work. For more details, see CSU's Academic Policies on Course Attendance (http://catalog.colostate.edu/general-catalog/academic-standards/academic-policies/).
- If you have to miss class for a serious medical issue (e.g., operation, sickness severe enough to require a doctor's visit, Covid), the absence will be excused.

Lecture videos: The lectures for each class are available through online videos. You will be assigned specific videos to watch before each class (see the detailed schedule in the online coursebook). These are posted on YouTube so they will be available for you to revisit even after you complete the course.

Weekly in-course group exercises: The in-person part of each class will be spent doing in-course group exercises. Ten points of your final grade will be based on your participation in these exercises. As long as you are in class and participate in these exercises, you will get full credit for this component. If you miss a class, to get credit towards this component of your grade, you will need to turn in a one-page document describing what you learned from doing the in-course exercise on your own time. All in-class exercises are included in the online course book at the end of the chapter on the associated material.

In-class quizzes: You will have nine total in-class quizzes (see the detailed schedule online for dates). There will be at least 10 questions per quiz (usually closer to 13 or 14). You will get 1/3 point for each correct answer. If you do the math, you can get full credit for this if you get at least about 75% of your answers right. You cannot get more than the maximum of 25 points for this component—once you reach 25 points on quizzes, you will have achieved full credit for the quiz component of the course grade.

All quiz questions will be multiple choice, matching, or some other form of "close-answered" question (i.e., no open-response-style questions). You cannot make up a quiz for a class period you missed unless you are absent for an approved reason. You can still get full credit on your total possible quiz points if you miss a class, but it means you will have to work harder and get more questions right for days you are in class.

Because grading format for these quizzes allows for you to miss some questions and still get the full quiz credit for the course, I will not ever re-consider the score you got on a previous quiz, give points back for a wrong answer on a poorly-worded question, etc. However, if a lot of people got a particular question wrong, I will be sure to cover it in the next class period. Also, especially if a question was poorly worded and caused confusion, I will work a similar question into a future quiz—in addition to the 10 guaranteed questions for that quiz—so every student will have the chance to get an extra 1/3 point of credit for the question.

The "Vocabulary" appendix of our online book has the list of material for which you will be responsible for each quiz. Most of the functions and concepts will have been covered in class, but some may not. You are responsible for going through the list and, if there are things you don't know or remember from class, learning them. To do this, you can use help functions in R, Google, Stack Overflow, books on R, ask a friend, and any other resource you can find.

The final version of the Vocabulary list you will be responsible will be posted by the Wednesday evening before each quiz.

In general, using R frequently in your research or other coursework will help you to prepare and do well on these quizzes.

Homework: There will be six homework assignments, starting a few weeks into the course and then due approximately every two weeks (see the detailed schedule in the online course book for exact due dates).

The first homework (HW #1) should be done individually. For other homeworks, feel free to talk with other students as you work to complete them.

For later homeworks, a subset of the full set of questions will be selected for which we will do a detailed grading of the code itself, with substantial feedback on coding. All other questions in the homework will be graded for completeness and based on the final answer produced.

Homework is due by the start of class on the due date. Your grade will be reduced by 10 points for each day it is late, and will receive no credit if it is late by over a week.

Final Group Project: For the final project, you will work in small groups (3-4 people) on an R programming challenge. The final grade will be based on the resulting R software, as well as on a short group presentation and written report describing your work. You will be given a lot of in-class time during the last third of the semester to work with your group on this project, and you will also need to spend some time working outside of class to complete the project. More details on this project will be provided later in the semester.

Course set-up: Please be sure you have the latest version of R and RStudio (Desktop version, Open Source edition) installed. Both are free for anyone to download. Also, be sure to sign up for a free GitHub account.
Here are useful links for this set-up:
- R: https://cran.r-project.org
- RStudio: https://www.rstudio.com/products/rstudio/#Desktop
- Sign-up for a GitHub account: https://github.com

Online coursebook: The coursebook will serve as the only required textbook for this course. I am still in the process of editing and adding to this book, so content may change somewhat over the semester. We typically cover about a chapter of the book each week of the course.
This coursebook includes:
- Links to video lectures for each class
- Links to the slides presented in the lectures for each topic
- In-course exercises, typically including links to the data used in the exercise
- An appendix with homework assignments
- A list of vocabulary and concepts that should be mastered for each quiz

If you find any typos or bugs, or if you have any suggestions for how the book can be improved, feel free to post it on the book's GitHub Issues page (https://github.com/geanders/RProgrammingForResearch/issues).

This book was developed using Yihui Xie's phenomenal bookdown (https://bookdown.org) framework. The book is built using code that combines R code, data, and text to create a book for which R code and examples can be re-executed every time the book is re-built, which helps identify bugs and broken code examples quickly. The online book is hosted using GitHub's free GitHub Pages (https://pages.github.com). All material for this book is available and can be explored at the book's GitHub repository (https://github.com/geanders/RProgrammingForResearch).

Other helpful books (not required): The best book to supplement the coursebook and lectures for this course is R for Data Science (http://r4ds.had.co.nz), by Garrett Grolemund and Hadley Wickham. The entire book is freely available online through the same format at the coursebook. You can also purchase a paper version of the book (published by O'Reilly) through Amazon, Barnes & Noble, etc., for around $40. This book is an excellent and up-to-date reference by some of the best R programmers in the world.

There are a number of other useful books available on general R programming, including:
- R for Dummies (https://colostate.primo.exlibrisgroup.com/discovery/fulldisplay?vid=01COLSU_INST:01COLSU&tab=Everything&offset=0&docid=alma991000341369703361&query=any,contains,r%20for%20dummies&context=L&adaptor=Local%20Search%20Engine&sortby=rank&lang=en&search_scope=MyCampus_FC_CI_PU_P)
- R Cookbook (https://rc2e.com/)
- R Graphics Cookbook (https://r-graphics.org/)
- Roger Peng's Leanpub books (https://leanpub.com/u/rdpeng)
- Various books on bookdown.org (www.bookdown.org)

The R programming language is used extensively within certain fields, including statistics and bioinformatics. If you are using R for a specific type of analysis, you will be able to find many books with advice on using R for both general and specific statistical analysis, including many available in print or online through the CSU library.