

ERHS 581A3: R PROGRAMMING FOR RESEARCH

Fall 2022

Instructor: Brooke Anderson

Email: brooke.anderson@colostate.edu

Time: M/W 10:00 am – 12:00 pm

Place: 120 Environmental Health

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

Office Hours: Fridays, 10:00–11:00 am, Room 120 of the Environmental Health Building.

Time and place: This course meets in Room 120 of the Environmental Health Building on Mondays and Wednesdays, 10:00 am–12:00 pm. Exceptions to these meeting times are:

- There will be no meeting on Labor Day (Monday, Sept. 3).

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
Weekly in-class quizzes	40
Homework	30
Attendance and class participation	10
Weekly in-course group exercises	20

Attendance and class participation: Because so much of the learning for 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 one or fewer classes
- **8 points** if you miss two classes
- **6 points** if you miss three classes
- **4 points** if you miss four classes
- **2 points** if you miss five classes
- **0 points** if you miss six or more classes

Exceptions:

- Attendance on the first day of class will not be counted.
- If you miss classes for “University-sanctioned” activities. These can include attending a conference, travel to collect data for your dissertation), For these absences, you must inform 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. 11, 2017 (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.

Weekly in-course group exercises: Part of each class will be spent doing in-course group exercises. Twenty 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 four 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 point for each correct answer. If you do the math, you can get full credit for this if you get at least 75% of your answers right. You can not get more than the maximum of 40 points for this component—once you reach 40 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 can not make up a quiz for a class period you missed. 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 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 this 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, StackOverflow, 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 two homework assignments (see the detailed schedule in the online course book for exact due dates).

The first homework (HW #1) should be done individually.

For the second homework, a subset of the full set of questions will be selected for which I 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. For homeworks with a heavy coding component, I will provide solution code for all questions.

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. To give you time to absorb the material before completing the homework, the final homework will be due after the last day that you attend class.

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: This 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 (particularly for the second half of the book, which is currently in a rawer draft than the beginning of the book). We typically cover about a chapter of the book each week of the course.

This coursebook includes:

- Links to the slides presented in class 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.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=01COLSU_ALMA51267598310003361&context=L&vid=01COLSU&lang=en_US&search_scope=Everything&adaptor=Local%20Search%20Engine&tab=default_tab&query=any,contains,r%20for%20dummies&sortby=rank&offset=0)
- R Cookbook (https://colostate-primo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=01COLSU_ALMA21203304500003361&context=L&vid=01COLSU&lang=en_US&search_scope=Everything&adaptor=Local%20Search%20Engine&tab=default_tab&query=any,contains,r%20cookbook&sortby=rank&offset=0)
- R Graphics Cookbook (http://www.amazon.com/R-Graphics-Cookbook-Winston-Chang/dp/1449316956/ref=sr_1_1?ie=UTF8&qid=1440997472&sr=8-1&keywords=r+graphics+cookbook)
- 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.