FOR 875: R Programming for Data Science

Cross-listed: STT

Opening remarks and background

Welcome to Forestry 875! My name is Andy Finley and I'll be your instructor for this online class. You can find my contact information in the course syllabus available in the "Syllabus" directory on our course's D2L home page (https://d2l.msu.edu). The syllabus is reviewed in a subsequent video available in the "Getting started" D2L directory.

This document provides a bit of context for the course and gets you started in the right direction.

A bit about me—I'm an Associated Professor at Michigan State University (MSU) with a joint appointment in the Departments of Forestry and Geography, and am adjunct in the Department of Statistics & Probability. I am also a member of the interdisciplinary Ecology, Evolutionary Biology, and Behavior (EEBB) Graduate Program faculty.

I hold degrees are in Forestry and Statistics. My research interests are in developing methodologies for monitoring and modeling environmental processes, Bayesian statistics, spatial statistics, and statistical computing. A central theme in my research is the use of models to integrate information from disparate sources in order to improve inference and prediction. In terms of application areas, my research focuses on spatial-temporal modeling of changing ecosystem components and systems. My recent interest is in improving frameworks for modeling exposure to pollutants, climate change, and health outcomes (ecosystem and public).

My first programming experiences were as an undergraduate at Penn State, where I had an internship with a team of career programmers who were working on software to model environmental data as part of a NASA-funded study. The thrill I felt learning to write UNIX shell scripts, C, and Fortran got me hooked on computing and data analysis! Throughout the remainder of my undergraduate program and subsequent graduate degrees, I sought out opportunities, classes, and research topics that would allow me to build new skills in programming, high-performance computing, and data analysis.

I started learning R during my PhD at the University of Minnesota in 2003. Until then, most of my programming experience was with lower-level languages such as C, C++, and Fortran. I had worked with higher-level languages such as PHP, Perl, and Python, but R's superior syntax, graphics, statistical analysis capabilities, and extendability captivated my attention and desire to learn all I could about this remarkable tool. Over the 14 years since I first typed "R" on a command line, few days have gone by where I've not worked with R in some way.

My hope is that, starting with this course, you also come to deeply enjoy working with R to tackle interesting and pressing data science questions and that you are able to use these tools to do exciting things and make impactful contributions to your field!

How this course is organized, grading, and keeping on schedule

I purposely kept the structure and variety of tasks simple throughout this course. Your course book (available on D2L) is your guide. Read Chapters 1-10 in order. At the end of each chapter there is a list of one or more exercises for you to complete. The exercises along with a brief introductory video and

supporting material are in the “Exercises” directory on D2L. Complete the exercises and submit them to their corresponding “dropbox” for grading.

As noted in the course syllabus, 95 percent of your semester grade is based on the completion of 12 exercises. Two of these 12 exercises are due each week on Friday 5pm EST. Table 1, below, is a guide that will keep you on track to submit everything on time.

Also, as noted in the syllabus, your lowest score on the 12 exercises will not be included in the calculation of your final grade. It is fine if you choose to just “skip” turning in one exercise (it will just serve as your lowest grade); however, know that the material in the first 10 chapters builds on itself, so even if you don’t turn in an exercise, be sure you understand the material covered in the corresponding chapter.

Exercise solutions will be posted in the given exercise directory within the hour after the exercise is due. Hence, no late exercises will be accepted.

You are, of course, welcome to work through the material, submit exercises, and complete the course at a faster pace.

Week	Read chapters	Complete exercises (due date)
1	1 & 2	No exercises are due. However, get RStudio installed on your computer. See Chapter 2 for details.
2	3	1 & 2 (7/13)
3	4	3 & 4 (7/20)
4	4 & 5	5 & 6 (7/27)
5	6.1 - 6.4	7 & 8 (8/3)
6	6.5 & 7	9 & 10 (8/10)
7	Choose two from 8 – 14	Choose two from 11 – 17 (8/17)

Table 1: Exercise due dates and recommend scheduled of readings. You are welcome to submit exercises prior to due dates.

The remain 5 percent of your grade will be awarded based on your participation in one or more online surveys that will be given throughout the semester. These are not tests; rather, they are a tool I’ll use to adjust the material and learning experience as we progress. More details on this to come.

How to do well in this class

As noted in the previous section, the formula to get a good grade in this class is simple—correctly answer questions and code statements for 12 exercises and participate in a few online surveys. This will likely require that you closely read the chapters corresponding to the exercises and, in some cases, work through the external material noted in the exercises and chapters.

The effort and time you need to expend to get a good grade in this course will likely be substantially less than that needed to maximize your knowledge and programming proficiency. See Section 1.6 in the course book for tips on getting the most out of this course.

How to get help

There are two modes of communication:

1) email me directly,
2) Piazza.

I want you to use Piazza for questions about course material, logistics, and exercises. You will receive an invitation to join our course's Piazza site via your MSU email (let me know if you haven't receive this invitation). Please accept this invitation and review how to use Piazza through Piazza's information pages and video. Also, please read a document called "FOR875-Piazza.pdf" in the "Getting stared" directory on our course's D2L page.

In addition to the exercise-specific Piazza directories, I've also include "course_logistics" and "general_coding" folders where you can post questions about the course and coding topics that don't fit in any given exercise folder, respectively. Please feel very free to use the anonymous setting to ask your questions on Piazza.

You can always email me directly with questions, concerns, difficulties, etc.. However, if it is an exercise or learning question, I strongly prefer that you submit it to Piazza so the question/answer can benefit all participants (in fact, if you email me a such a question, I'll likely ask you to post it on Piazza for me to answer).

Although there are a lot of learning materials currently posted on our D2L site, I view this as a living course and hence will generate new learning material as needed and in response to your questions. For example, if there seem to be a lot of questions about some aspect of a chapter or topic, I'll create a video that works through the material with examples and explanations.

Technical difficulty with D2L should be directed to MSU help channels, see Section "Technical Assistance" in the syllabus for contact information.

How to get started

Log onto D2L and find our course site (https://d2l.msu.edu). Go to the "Getting started" directory where you will find this document and a Piazza howto. The course syllabus is in the D2L "Syllabus" directory. It is your responsibility to read and understand the syllabus and the course expectations. Please let me know if you have any questions.

Next go to the "Course book" directory, download the course book, and start reading. The exercises you'll need to complete are in the "Exercises" directory. Aside from the videos that introduce each exercise (contained in each exercise's directory), there is a "Videos" directory which contains a video that briefly covers RStudio, which you will read about in Chapter 2. As the course progresses, additional videos may be added to the "Videos" directory.

Final note about course material

My colleague Dr. Vince Malfi and I are actively working on this book. We hope the book is helpful to you. I can almost guarantee there are typos and mistakes. Please point them out to me when you come across them, and I'll correct them (add them to the "book_typos_ideas" directory on Piazza, again feel free to use the anonymous setting). Also, please let me know if text or code is confusing. Like the course, the book is in active development; hence, corrections and updates can be made. I'll keep a running list of correction/additions in the "Course book" directory on D2L. Like the book, the exercises also might contain typos or confusing text, so please let me know if you see anything, and I'll post the correction or clarification on Piazza.