

FOR 875: R Programming for Data Science

Cross-listed: STT

Summer Semester 2017 Syllabus

(last modified: 7/1/18)

Part 1: Course Information

Instructor Information

Instructor: Andrew Finley
Office: 211 Natural Resources Building
Office telephone: 517-898-5970
Email: finleya@msu.edu

Course Description

Programming in R and use of associated Open Source tools. Addressing practical issues in workflow documentation, data management, and scientific computing.

Course Overview

R has emerged as a preferred programming language in a wide range of data intensive disciplines (e.g., O'Reilly Media's 2014 Data Science Salary Survey found that R is the most popular programming language among data scientists). The goal of this course is to teach applied and theoretical aspects of R programming for data sciences. Topics will cover generic programming language concepts as they are implemented in high-level languages such as R. Course content focuses on design and implementation of R programs to meet routine and specialized data manipulation/management and analysis objectives. Attention will also be given to mastering concepts and tools necessary for implementing reproducible research.

Course Location/Time

Classroom: Online

Course dates: 7/2/2018-8/16/2018

Textbooks & Course Materials

Required materials: All required materials will be provided by the instructor via Desire2Learn (D2L).

Other course materials: The following texts are available for free download via Michigan State University (MSU) library link using valid MSU NetID. They are not required but are useful references.

- Wickham, H. (2014) Advanced R. Chapman and Hall/CRC.
<http://catalog.lib.msu.edu/record=b10629576~S39a>
- Wickham, H. (2009) Ggplot2. Springer. <http://catalog.lib.msu.edu/record=b7232787~S39a>

Course Requirements

- Internet connection (DSL, LAN, or cable connection desirable)
- Access to D2L
- A computer desktop or laptop with Mac, Windows, or Linux operating system on which you can install Open Source software

Course prerequisites/co-requisites: None.

Course Structure

This course will be delivered entirely online through the course management system D2L. You will need your MSU NetID to login to the course from the D2L home page (<https://d2l.msu.edu>).

In D2L, you will access online lessons, course materials, and additional resources. Activities will consist of readings and related links, videos, and software writing exercises. This course is asynchronous, meaning that there is not a set time when everyone needs to be online for course activities. Performance assessment is based on exercises and participation in online surveys.

We will be using Piazza for questions and discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza.

Find our class page at: <https://piazza.com/msu/summer2018/for875/home>

Technical Assistance

If you need technical assistance with MSU D2L at any time during the course or to report a problem you can:

- Visit the Distance Learning Services Support Site here www.lib.msu.edu/dls
- Call the Desire2Learn hotline at 517-355-2345. They are very helpful.
- Visit the Desire2Learn Help Site here <http://help.d2l.msu.edu>

Part 2: Course Objectives

Upon completion of this course participants should be able to:

- establish an efficient scientific computing environment
- identify and use available R packages and associated Open Source software to meet given scientific objectives
- design and write efficient programs using R (and similar high-level languages) to perform routine and specialized data manipulation/management and analysis tasks
- document, share, and collaborate on code development using a suite of Open Source standards and tools
- document analytical workflow using R and markdown languages

Participants will meet the learning objectives listed above through a combination of the following course activities:

- readings
- online videos
- exercises

- links to online presentations and programming forums from non-MSU sources

Course Topic Schedule

List of topics covered in the learning material:

1. Overview of R
2. Install and configuration of R programming environment
3. Basic language elements and data structures
4. R+Knitr+Markdown
5. Data input/output
6. Data storage formats
7. Subsetting objects
8. Vectorization
9. Control structures
10. Functions
11. Scoping Rules
12. Loop functions
13. Graphics and visualization
14. Grammar of data manipulation (dplyr and related tools)
15. Debugging/profiling
16. Spatial and textual
17. R and C++ integration
18. Development of online interfaces and data visualization
19. R and external database integration
20. Data structures for specialized data types, e.g., spatial and signal data

Important: If necessary, this syllabus will be modified. Any modifications to the syllabus will be posted on D2L and email notification will be distributed to course participants.

Course Requirements and Evaluation

1. Individual-based exercises (95%) – 12 exercises will be graded. You are required to complete exercises 1-10, then you choose two additional to complete among exercises 11-17 (see D2L for the list of exercises and submission instructions). Your lowest exercise grade will be dropped from your final grade calculation (i.e., your final grade will be based on only your 11 best exercise grades). Exercise due dates are posted on the D2L calendar and in the “Getting started” document on D2L.
2. Participant feedback via online surveys and assignment comments (5%).

Final grade assignment is based on the following table.

Grade Percentage

- 4.0 – 90 to 100%
- 3.5 – 85 to 89%
- 3.0 – 80 to 84%
- 2.5 – 75 to 79%
- 2.0 – 70 to 74%
- 1.5 – 65 to 69%
- 1.0 – 60 to 64%

0 – 0 to 59%

Late Exercises and Solutions

Late exercises will not be accepted. Exercise solutions will be posted shortly after the due date. I would strongly encourage you not to wait until the last moment to start on your exercises—give yourself plenty of time to complete and submit your work.

****If you experience technical difficulties with D2L that prevent you from completing an assignment on time, seek help from the 24-hour Help Desk. Record the Ticket # from your contact****

Viewing Grades

Grades for all exercises can be viewed in the D2L Gradebook. Grades for all exercises will generally be posted within one week of the due date.

Part 5: Course Policies

Building Rapport

If you find that you have any trouble keeping up with exercises or other aspects of the course, make sure you let your instructor know as early as possible. As you will find, building rapport and effective relationships is key to becoming an effective professional. Make sure that you are proactive in informing your instructor when difficulties arise during the semester so that we can help you find a solution.

Complete Exercises

All exercises for this course will be submitted electronically through D2L. Exercises must be submitted by the given deadline or special permission must be requested from instructor before the due date. Extensions will not be given beyond the next assignment except under extreme circumstances.

Understand When You May Drop This Course

It is the student's responsibility to understand when they need to consider un-enrolling from a course. Refer to the Michigan State University Office of the Registrar for important dates and deadlines.

Inform Your Instructor of Any Accommodations Needed

Michigan State University is committed to providing equal opportunity for participation in all programs, services and activities. If you have a documented disability and verification from the Resource Center for Persons with Disabilities (RCPD), and wish to discuss academic accommodations, please contact your instructor as soon as possible. It is the student's responsibility to provide documentation of disability to RCPD and meet with an RCPD specialist to request special accommodation before classes start.

Once your eligibility for an accommodation has been determined, you will be issued a verified individual services accommodation ("VISA") form. Please present this form to the instructor at the start of the term and/or two weeks prior to the accommodation date (test, project, etc). Requests received after this date will be honored whenever possible.

RCPD is located in 120 Bessey Hall, near the center of the Michigan State University campus, on the southwest corner of Farm Lane and Auditorium Road. RCPD may be contacted by phone at (517) 884-7273 (884-RCPD), or via their website (<http://www.rcpd.msu.edu>).

Commit to Integrity

Academic Honesty

Article 2.3.3 of the Academic Freedom Report states that "The student shares with the faculty the responsibility for maintaining the integrity of scholarship, grades, and professional standards." In addition, the Department of Forestry adheres to the policies on academic honesty as specified in General Student Regulations 1.0, Protection of Scholarship and Grades; the all-University Policy on Integrity of Scholarship and Grades; and Ordinance 17.00, Examinations. (See Spartan Life: Student Handbook and Resource Guide and/or the MSU Web site: www.msu.edu.)

You are expected to develop original work for this course; therefore, you may not submit course work you completed for another course or content created by other students. Students who violate MSU academic integrity rules may receive a penalty grade, including a failing grade on the assignment or in the course. Contact your instructor if you are unsure about the appropriateness of your course work. (See also <https://ombud.msu.edu/academic-integrity/student-faq.html>)