Department of Forestry, College of Agriculture and Natural Resources

Forestry 875: R Programming for Data Sciences

Cross-listed: STT

Summer Semester 2019 Syllabus

Part 1: Course Information

Instructor Information

Instructor: Andrew Finley
Office: 211 Natural Resources Building
Office hours: By appointment (an audio and/or video connection will be provided)
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 documenting workflow, 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 Data Science Salary Survey[1] 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: 7/1/2019-8/15/2019

Textbooks & Course Materials

Required materials:
- Wickham, H. (2014) Advanced R. Chapman and Hall/CRC. http://catalog.lib.msu.edu/record=b10629576~S39a Note, this text is available for free download via MSU library link using valid MSU NetID.
- All other required materials will be provided by the instructor via D2L.

Other course materials:

---

1   King, J. and Magoulas, R. 2014 Data Science Salary Survey: Tools, Trends, What Pays (and What Doesn't) for Data Professionals.

- Wickham, H. (2009) Ggplot2. Springer. http://catalog.lib.msu.edu/record=b7232787~S39a
  Note, this text is available for free download via MSU library link using valid MSU NetID.
- Peng, R.D. (2015) Exploratory Data Analysis with R. https://leanpub.com/exdata

Course Requirements

- Internet connection (DSL, LAN, or cable connection desirable)
- Access to Desire2Learn (D2L)

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 various readings and related links, online lectures, and software writing assignments. 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 online quizzes, exams, and assignments.

Technical Assistance

If you need technical assistance 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. This number should be added to your cell phone directory.
- 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, markdown languages, and version control

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

- readings
- online lectures
- links to online presentations and programming forums from non-MSU sources
- assignment and exam preparation

Course Topic Schedule

List of topics covered in lecture and supporting learning material (dates and additional specifics will be added upon course approval):

1. History and overview of R

2. Install and configuration of R programming environment

3. Basic language elements and data structures

4. R+Knitr+Markdown+GitHub

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. Statistical simulation

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.