

BAE 495/590: R Coding for Data Management and Analysis

Fall 2020 | Online Course (asynchronous) | 3 credit hours



Description: This course provides students with foundational coding skills in R, an open-source statistical software environment, as well as instruction on best practices for data management. As this is an introductory course, prior programming experience is not required or expected. Coding approaches taught in the course will be targeted towards developing skills needed for summarizing data, creating data visualizations, and applying simple statistical models for data analysis.

R coding skills are applicable to a wide range of datasets, but most of the data we work with in the course will be related to environmental and agricultural sciences and engineering, and background information will place R coding skills in the context of environmental and agricultural systems analysis. The course is specifically targeted for students in biological and agricultural engineering, environmental engineering, and life sciences, but students from any discipline are welcome to register for the course.

Structure: The course is offered online and completely asynchronous, meaning that students have no real-time class meeting requirements. Learning activities include reading assignments, quizzes, videos, discussion forums, and an individual project.

Topics:

- Data visualization (ggplot2)
- Data transformation (dplyr)
- Data tidying (tidyr)
- Joins, factors, dates, and times (lubridate)
- Loops and user-defined functions (base R)
- Geospatial data visualization (sf)
- Regressions and correlations (base R)
- Text mining (stringr, tidytext)
- Literate programming (R Markdown)

Prerequisites: Introductory statistics (ST 370 or ST 515) or instructor permission.

Instructor: Natalie Nelson, Assistant Professor, Biological and Agricultural Engineering; Email: nnelson4@ncsu.edu