Introduction to Statistical Computing

Instructor: Dr. Hongxiao Zhu
                  Office: Hutcheson Hall 403H
                  Email: hongxiao@vt.edu
                  https://hongxiaozhu.github.io/

Modality: Online synchronous
Time: Monday 6:00 PM–9:00 PM
Office hour: By appointment
Zoom link: https://virginiatech.zoom.us/j/86156411455
Canvas site: This course has a canvas site through Virginia Tech's Canvas system (http://canvas.vt.edu). All course materials will be posted on canvas site. Homework will be collected and graded through canvas.

Course description: This course helps develop necessary R programming skills for statistical analysis and data science. Students will learn the basic syntax and functions of R, data organization and visualization, programming, as well as more advanced tools for analyzing modern datasets. Materials will be taught together with examples and case studies in applied statistics, data science, and machine learning.

Topics covered in this class include:

Module I: R basics

        Week 1: Setup; directory, workspace, scripts; packages; get help. Rmarkdown and reporting.

        Week 2: Sequences and vectors; basic plot. Rmarkdown and reporting.

        Week 3: Data types and structure (arrays, data frame, list); data import/export.

        Week 4. Visualization with ggplot; pipe operator.

        Week 5. Data processing with tidyverse (part I)

        Week 6. The apply function; descriptive statistics and visualization.

Module II: R programming

        Week 7. Programming: if/for statement.

        Week 8. Programming: while, function.

        Week 9. A case study on programming.

Module III: More advanced topics

        Week 10. Data processing with tidyverse (part II)

        Week 11. Data scraping

        Week 12. Text mining

        Week 13. Optimization

        Week 14. Intro to Python

Prerequisites: Graduate standing.

Textbook: There is no textbook required. Students are recommended to refer to the following books:

- *Introduction to Scientific Programming and Simulation Using R*, by Owen Jones, Robert Maillardet and Andrew Robinson, 2014, CRC Press, Taylor & Francis Group.
- *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data by Wickham & Grolemund.*
https://r4ds.had.co.nz/

Software: R and R studio.

Hardware: This course requires to use a computer.

Grading:

Homework: 90%
Class attendance: 10%. Attendance will be recorded for every lecture.

Homework and class attendance add up to 100%. Scores are first scaled to percentages and then weighted using the corresponding weights in percentages.

Class attendance: Class attendance is required unless there are medical conditions or other reasonable obligations. If you have a medical condition or an obligation that will result in missing a class, let me know ahead of time.

Course materials: The canvas site will be used to post class materials, lecture videos, and collect/grade homework and exams.

Homework policy: A due date is shown on each homework assignment. Late homework is only accepted when it is submitted no later than three days after the deadline. Late homework is subject to an extra deduction of 20% per late day.

Honor code: Graduate students enrolled in this course are responsible for abiding by the Graduate Honor Code. For additional information about the Honor Code, please visit: http://graduateschool.vt.edu/academics/expectations/graduate-honor-system.html.

Diversity and inclusion: Virginia Tech strive to inspire and nurture all in our community - understanding that each person brings distinct life experiences to the research and education. For statement of diversity and inclusion of Department of Statistics, please refer to https://www.stat.vt.edu/about/Diversity.html