Introduction to R Statistical Analysis Software
Summer – 2024

Instructor Information

Name: Shunkei Kakimoto
Office: 316a Ruttan
Email: kakim002@umn.edu
Office Hours: After class, in-person or on Zoom

Class Information

Dates: August 19 – 23
Time: 1:00 pm – 4:00 pm
Classroom: Ruttan 135B (The Land O Lakes Classroom)

Course Description

This is a one-week course designed as an introduction to R statistical software for incoming graduate students. R is the leading software used in Ph.D.-level econometric classes in APEC (APEC 8211-8214). For example, we use R to conduct Monte Carlo simulations to understand the properties of statistical tools. In this course, we will go through the basics of R programming. Although the topics we can cover within five days are limited, I selected essential topics that would be directly helpful to establish a solid foundation for students' R skills not only for first-year Ph.D. Econometrics classes, but also for personal research and future development of R skills.

By completing the course, students will be able to

- create an R project and know how to access and save data.
- handle different types of base R objects (e.g. list, vector, matrix, and data.frame).
- use basic data wrangling skills with the data.table package (e.g. subset rows, select and compute on columns, rename columns, perform aggregations by group, merge multiple datasets, and reshape wide-to-long and long-to-wide, respectively).
- write their own R functions.
- write code for simple Monte Carlo Simulations with loop.
- use basic ggplot functions to visualize data.

For students who are interested in learning R programming further, I recommend that you take Programming for Econometrics (APEC8221) and Big Data Methods in Economics (APEC8222).

Before the First Day of the Class

- Download and Install R and R studio from this website
- Get UCard access to Ruttan Hall:
  - Get a U card so you can access the building.
  - Request advanced access to the building (Ruttan Hall) through this form.
  - If you have questions, reach out to Melissa Isle (webe0342@umn.edu)
- Finish the Survey.
- Bring your laptop to the class.

Recommended Textbook

This course does not require any textbooks. Below is the recommendation for reference. All the materials are freely available online.

- R for Data Science 2nd edition and Solutions to Exercise
  - This book mostly uses the dplyr package, which we do not cover in this course (Instead, we use data.table package.) But still, some chapters are useful for learning the basics of R. Specifically, take a look at the following chapters:
    - Chapter 2 Workflow: basics
    - Chapter 6 Workflow: scripts and projects
    - Chapter 25 Functions
    - Chapter 26 Iteration
    - Chapter 27 A field guide to base R
- Introduction to data.table
  - I highly recommend this as an introduction to data.table syntax.
- ggplot2: Elegant Graphics for Data Analysis (3e)
- Modern Data Visualization with R
- Matloff, Norman. The art of R programming: A tour of statistical software design. No Starch Press, 2011.
  - The Programming for Econometrics class uses this book.

Syntax Cheat Sheet

- Base R Cheat Sheet
- Data Transformation with data.table

- Data visualization with ggplot2

Class Style

Each lecture will be divided into three sessions, where each session consists of a 50-minute lecture and a 10-minute break.

By just looking at code, you cannot acquire the skills for coding. The only way to learn coding is to do it yourself! At the end of each topic, we will work on small quizzes to check your understanding level. In addition, there will be exercise problems at the end of the slides. Some of the problems are adapted from the textbook R for Data Science. These exercise problems require you to combine multiple operations you learned in the lecture. Of course, I do not expect you to solve the quizzes and exercise problems immediately. I want you to go through the problems many times until you feel comfortable. The solutions are included in the slides.

Tentative Class Schedule

Date	Topic
8/19	Introduction to R and R studio interface, Basic operation of R
8/20	Data wrangling with data.table
8/21	Data visualization with ggplot2 package
8/22	Regression analysis with R, and Monte Carlo simulation
8/23	Write functions, Introduction to Quarto, Review