

INTRODUCTION TO R STATISTICAL ANALYSIS SOFTWARE

Summer 2025

Instructor:	Qingyin Cai	Email:	cai00154@umn.edu
Date:	August 18 – 22	Time:	13:00 – 16:00 (Central Time)
Place:	Ruttan 135B and Zoom	Office Hours:	After class, 16:00 – 17:00

Course Description

This one-week course provides an introduction to the R statistical software for incoming graduate students. R is the primary statistical software used in Ph.D.-level econometrics courses in APEC and is widely used for a range of tasks, including simulation, data analysis, and visualization. We will cover the fundamentals of R programming, focusing on essential topics that lay a strong foundation for success in first-year econometrics, independent research, and continued skill development.

By completing this course, students will be able to:

- Code in an IDE (RStudio) using both R scripts and R Markdown.
- Handle different types of base R objects (e.g., lists, vectors, matrices, and data frames).
- Apply basic data wrangling skills using the `data.table` package.
- Use basic `ggplot2` functions to visualize data.
- Run regressions in R.
- Write simple functions and loops in R.

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 Class

- Download and install R and R Studio on your desktop from this [website](#).
- Get your [UCard](#) access to Ruttan Hall:
 - Request advanced access to the 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. APEC students can borrow a laptop with R and R Studio installed from [Magrath Library](#).

Recommended Textbook

This course does not require any textbooks. Recommended reference materials are listed below and are all freely available online.

- [R for Data Science \(2e\)](#) and [Solutions to Exercises](#)
 - This book mostly uses the `dplyr` package, which we will 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](#)
- [ggplot2: Elegant Graphics for Data Analysis \(3e\)](#)
- [The R Graph Gallery](#)
- [Modern Data Visualization with R](#)
- [Introduction to Econometrics with R](#)
- [Matloff, Norman. The art of R programming: A tour of statistical software design. No Starch Press, 2011.](#)

Syntax Cheat Sheet

- [Basic R Cheat Sheet](#)
- [Data Transformation with data.table](#)
- [Data Visualization with ggplot2](#)
- [R Markdown Cheat Sheet](#)

Class Style

Each lecture will be divided into three sessions, with each session consisting of a 50-minute lecture followed by a 10-minute break. Coding is a hands-on skill—you learn it by doing, not just by watching. To reinforce key concepts, we will work through in-class exercises together at the end of each topic. After-class exercise problems are included at the end of the slides, which are designed to help you understand and practice operations covered in the lectures. The problems are **optional**, and a typed answer key will be provided.

Tentative Class Schedule

Date	Topic
8/18	Introduction to R and R studio interface, basic operation of R
8/19	Data wrangling with <code>data.table</code>
8/20	Data visualization with <code>ggplot2</code> package
8/21	Regression analysis with R, and Monte Carlo simulation
8/22	Write functions, and Review