

PH 718 Data Management and Visualization in R

Part 0: Syllabus Review & Introduction to R

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2025/12/25 15:22:22

Contact

- Instructor: Zhiyang Zhou
 - Email: zhou67@uwm.edu
 - Homepage: zhiyanggeezhou.github.io
- Lectures
 - Tue/Thur 16:00–17:15
- Office Hours
 - By appointment

Grading

- Assignments (60%)
 - Submitting digital copies
 - Attaching (if applicable) both outputs and source codes
 - Including necessary interpretation
 - Organized in a CLEAR and READABLE way
 - Accepting NO late submission
- Final project (40%)
 - Refer to the project guideline
- Bonus points (TBD)
 - Potentially irregular quiz held on Canvas

Materials

- Reading list
 - [R4DS] H. Wickham, M. Cetinkaya-Rundel, & G. Grolemund. (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*, 2nd Ed. Sebastopol: O'Reilly Media. (Accessible at r4ds.hadley.nz)
 - [ModernDive] Chester Ismay, Albert Y. Kim, & Arturo Valdivia. (2025). *Statistical Inference via Data Science: A ModernDive into R and the Tidyverse*, 2nd Ed. Boca Raton: CRC Press. (Accessible at moderndive.com/v2)
 - [ISL] G. James, D. Witten, T. Hastie, & R. Tibshirani. (2021). *An Introduction to Statistical Learning: with Applications in R*, 2nd Ed. New York: Springer. (Accessible at www.statlearning.com)
- Lecture notes and beyond
 - Posted at Canvas and zhiyanggeezhou.github.io
 - Subject to update **without prior notice**

Why using R?

- Open source: Freely accessible to everyone.
- Powerful for data analysis: Extensive libraries created and maintained by statisticians; built-in methods for advanced statistical modeling.
- Data visualization: High-quality plots with libraries like *ggplot2*.
- Cross-platform: Working with Windows, macOS, and Linux.

Comparing R, JMP, and SPSS

- Cost
 - R: Free and open-source.
 - JMP: Proprietary software with a paid license.
 - SPSS: Proprietary software, requires a paid license.
- Ease of use
 - R: Requires programming; steep learning curve; graphical interface like RStudio improves usability.
 - JMP: User-friendly interface with drag-and-drop capabilities.
 - SPSS: Point-and-click interface for beginners.
- Statistical analysis capabilities
 - R: Extremely powerful for basic, advanced, and cutting-edge statistical methods; extensive free packages support complex modeling.
 - JMP: Robust statistical capabilities with a strong focus on exploratory data analysis; less effective for advanced machine learning compared to R.
 - SPSS: Suitable for traditional statistical methods; limited support for advanced analytics unless additional modules are purchased; less customizable compared to R unless using SPSS Syntax (a programming language that is unique to SPSS).
- Visualization
 - R: Best-in-class visualization capabilities via free packages; allows highly customized, publication-quality visualizations.
 - JMP: Excels in dynamic and interactive graphics; provides instant visual feedback as data and models are explored; limited customization options compared to R.
 - SPSS: Basic charting and visualization options; less customizable and visually appealing compared to R and JMP.

Be careful when using R

- NO quality control: Packages developed by small groups without extensive testing

How to learn R?

- Self-learning with regular practice
- Be sophisticated in statistics

How to find solutions when using R?

- Help manual (mostly reliable)
- Online resources (less reliable)
 - Search engines: Google, etc.
 - Q&A communities: Stack Overflow, Reddit, Posit Community, etc.
 - AI tools: ChatGPT, Claude, Copilot, Grok, Cursor, Gemini, DeepSeek, etc.

Setup

- Base R: <https://cran.r-project.org>
- RStudio: <https://posit.co/download/rstudio-desktop/>
 - Old versions of RStudio: <https://dailies.rstudio.com/release/>

Using GitHub Copilot in RStudio

- Refer to <https://docs.posit.co/ide/user/ide/guide/tools/copilot.html>
- Prerequisites
 - Internet access
 - RStudio version 2023.09.0 or later (newer versions recommended)
 - A GitHub account with active GitHub Copilot subscription
 - * The basic GitHub Copilot is free
 - * The GitHub Copilot Pro is free too for verified students <https://github.com/education>
- Setup
 1. Go to **Tools > Global Options > Copilot**.
 2. Check **Enable GitHub Copilot**.
 3. If prompted, download and install the Copilot Agent components.
 4. Click **Sign In**.
 5. Copy the displayed Verification Code.
 6. Open the provided link in your browser (or go to GitHub's device login page), paste the code, and click **Continue**.
 7. Authorize the "GitHub Copilot Plugin".
 8. Return to RStudio: it will show your signed-in GitHub user.
- Using GitHub Copilot
 - Autocomplete-style suggestions for coding: light grey "ghost text"
 - * Accept: Press **Tab**.
 - * Dismiss: Continue typing or press **Enter**.
 - * Status indicator shows waiting, received, or no suggestions.
 - While Copilot is primarily intended to generate code, it can also be used to directly answer simple questions.
 - When using Copilot to solve a broad problem, it is a good practice to provide a high-level goal at the beginning of the file.

R basic syntax

- Refer to R script `UWM_PH718_2026Spring_Part00.r`

Course expectations

1. (Expected to) understand given R code and error messages (if any) with the assistance of AI tools.
2. (Expected to) be able to modify existing R code to meet specific requirements with the assistance of AI tools.
3. (Not required to) develop R code from scratch, starting with basic structures and building up.