

PH 718 Data Management and Visualization in **R**

Part 0: Syllabus Review & Introduction to **R**

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

2025/01/23 21:59:58

Contact

- Instructor: Zhiyang Zhou
 - Email: zhou67@uwm.edu
 - Homepage: zhiyanggeezhou.github.io
- Tutor: Don Cramer
 - Email: dkramer@uwm.edu
- Lectures
 - Tue/Thur 16:00–17:15 via Zoom
- 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%)
 - TBD
- Bonus points (TBD)
 - 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)
 - [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**

Course expectations

1. Understand given **R** code trunks and error messages (if any).

2. Be able to modify existing **R** code to meet specific requirements.
3. Develop **R** code from scratch, starting with basic structures and building up.

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 on 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 (reliable)
- Online resources (not always reliable)
 - Search engines: Google, etc.
 - Q&A communities: Stack Overflow, Reddit, Posit Community, etc.

- Large language models: ChatGPT, Claude, etc.

Installation

- Base **R**: <https://cran.r-project.org>
- **RStudio**: <https://posit.co/download/rstudio-desktop/>