Free Beginner's R Workshop

27th of October 2025

Pia Benedetti Vallenari and Orlin Todorov (TIA)

Why this R workshop is different

- Focus on teaching what is actually practical for researchers
- Not trying to teach *everything* about R
- We'll introduce the tools we like and use
- We'll learn how to responsibly use GenAI as a coding assistant
- Start from messy data (because that's realistic)
- No deep dive into statistical theory

What is R and RStudio?

R

- Programming language used for data analysis, statistics, and visualisation
- You can use R on its own but it's not very user-friendly

RStudio

- An *Integrated Development Environment (IDE)* for R
- Interface to write, run, and visualise R code

Both need to be installed, but you will only run RStudio

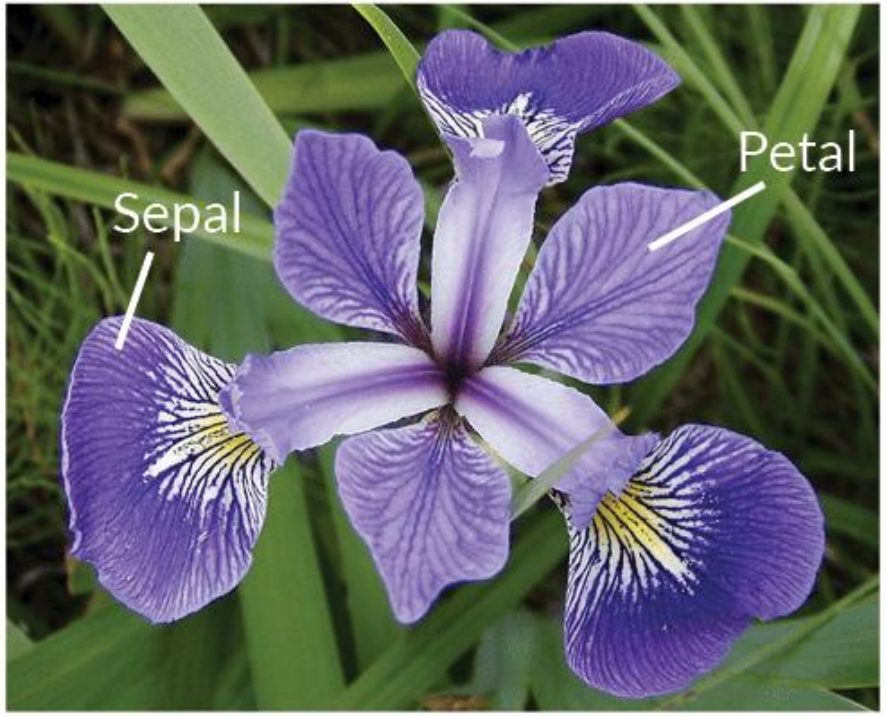
Why use R?

- Reproducible
- Can clean, analyse and plot data all in one
- Clean, analyse, and plot all in one
- Flexible
- Free and open-source (large community and online support)

Why clean data using R?

- Keeps your raw data intact
- Cleaning steps can be re-run anytime
- Easier to describe in your methods section
- Good practice!

iris dataset



Iris Versicolor

Iris Setosa

Iris Virginica

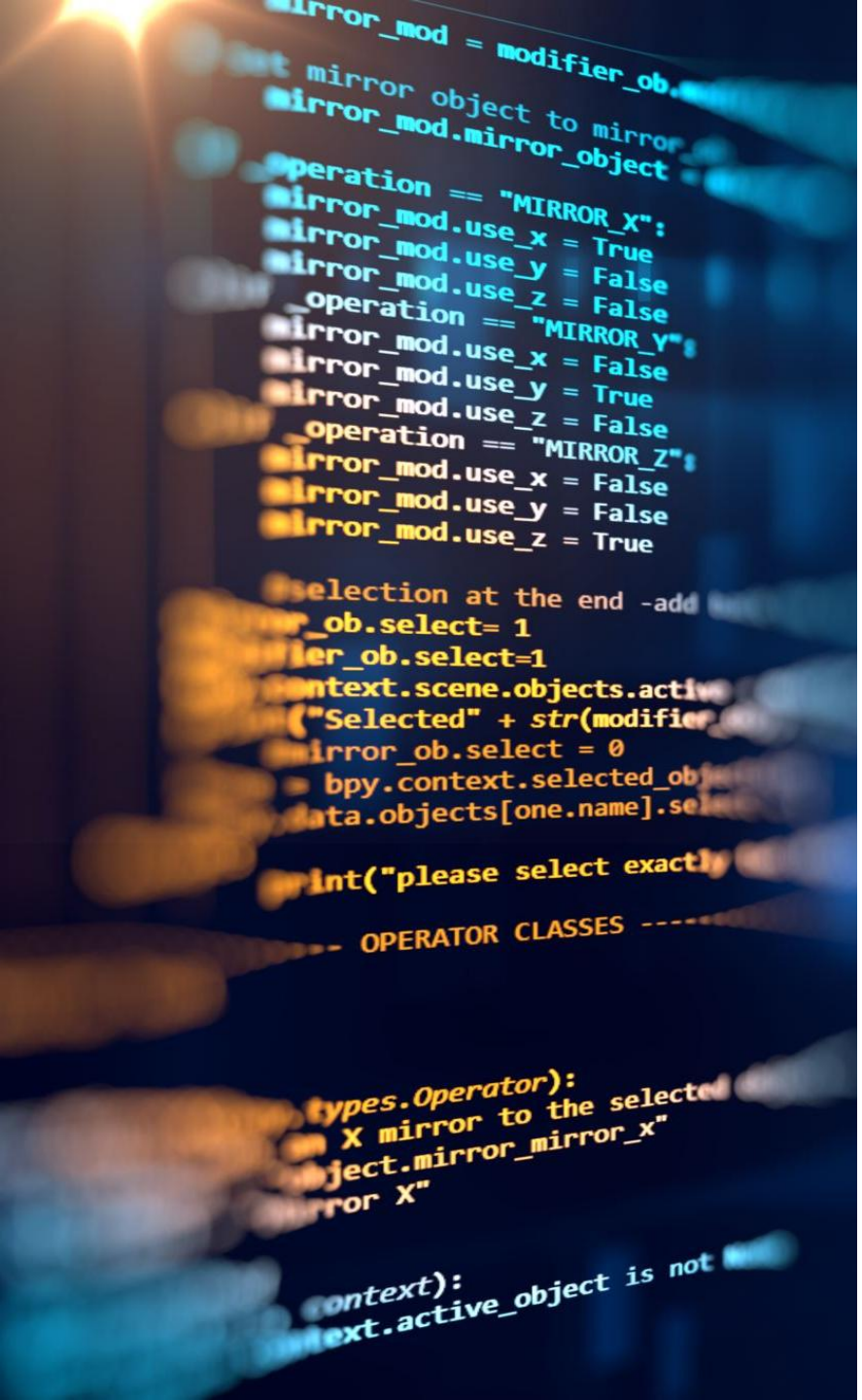
https://www.datacamp.com/tutorial/machine-learning-in-r

How I've made the iris data realistic ("messy")

	A	B	C	D	E
1	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
2	5.1	3.5	1.4	0.2	setosa
3	4.9	3	1.4	0.2	setosa
4	4.7	3.2	1.3	0.2	setosa
5	4.6	3.1	1.5	0.2	setosa
6	5	3.6	1.4	0.2	setosa

	A	B	C	D	E	F
1	Sepal length	sepal_Width	PetalLength	petal width(cm)	Species	Notes
2	50	3.5	1.4	0.2	setosa	check later
3	?	3	1.4	10	setosa	check later
4					setosa	check later
5	4.6	3.1	1.5	0.2	setosa	check later
6	5	3.6	1.4	0.2	setoosa	check later

99	6.2
100	5.1
101	five
102	6.3
103	5.8



Using AI to help write code

- Many researchers now use ChatGPT, Copilot, etc. to write or debug code
  - “Vibe Coding”
- AI tools are now integrated into IDEs like RStudio
- But you must still:
  - Understand what your code is doing and why
  - Check your outputs
  - Keep your data confidential

AI for Coding: Opportunities and Risks

“AI as a tutor” mindset

- Understanding error messages
- Get explanations of unfamiliar syntax
- Be introduced to new packages or functions ⚠️
- Learn by comparison
- Iterate interactively
- Translate plain English to code (and vice versa)

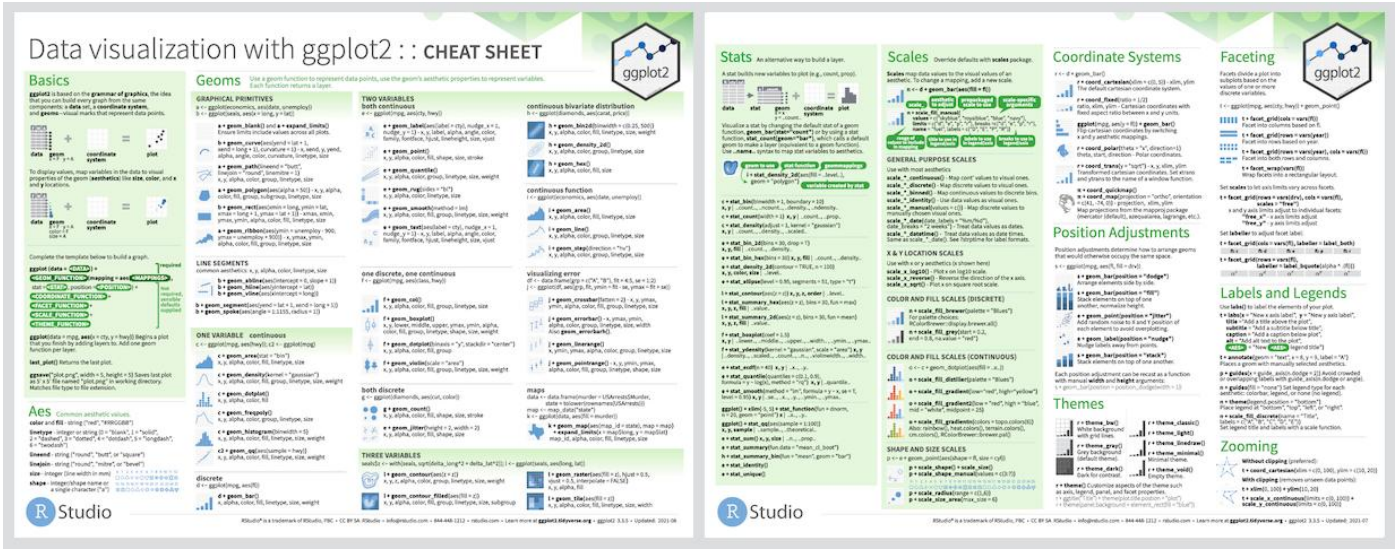
Limitations

- Code may look correct but is wrong
  - Doesn’t have entire context
- Hallucination of functions and packages
- AI can get stuck
- Can (confidently) recommend the wrong statistical approach
- Assumes your data is clean and passes all assumptions

ggplot2

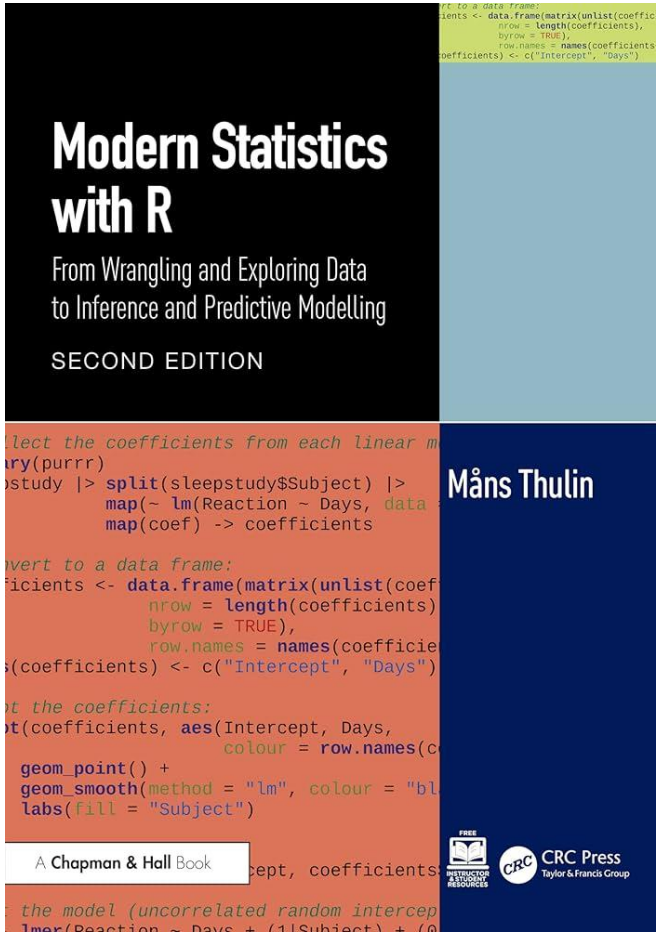
Why a ggplot2 example?

- *Low stakes*
- *Easy to prompt in natural language – i.e., vibe coding*
- *ggplot2 syntax is notoriously difficult to comprehend!*

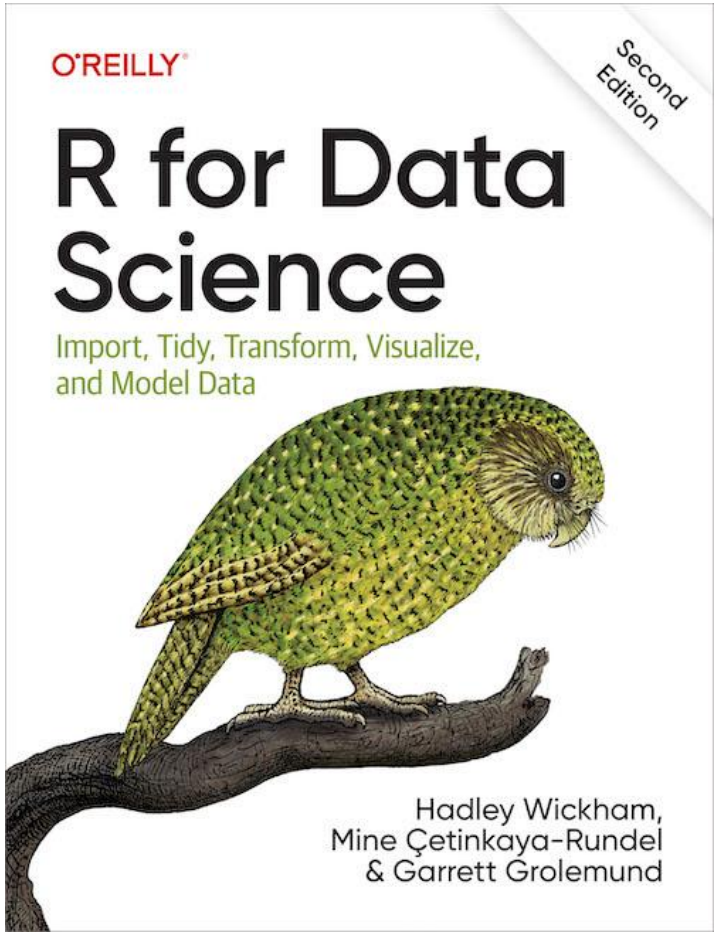


https://ggplot2.tidyverse.org/

Free Online R Textbooks



https://www.modernstatisticswithr.com/



https://r4ds.hadley.nz/

Other Free Resources

- Cheat sheets: https://posit.co/resources/cheatsheets/
- R code snippets: https://posit.cloud/learn/recipes
- Free R courses, for example: https://pll.harvard.edu/subject/r
- YouTube
- Other free online books:
  - Learning Statistics with R: https://learningstatisticswithr.com/
  - Data Visualisation: https://socviz.co/