AI Tools for R

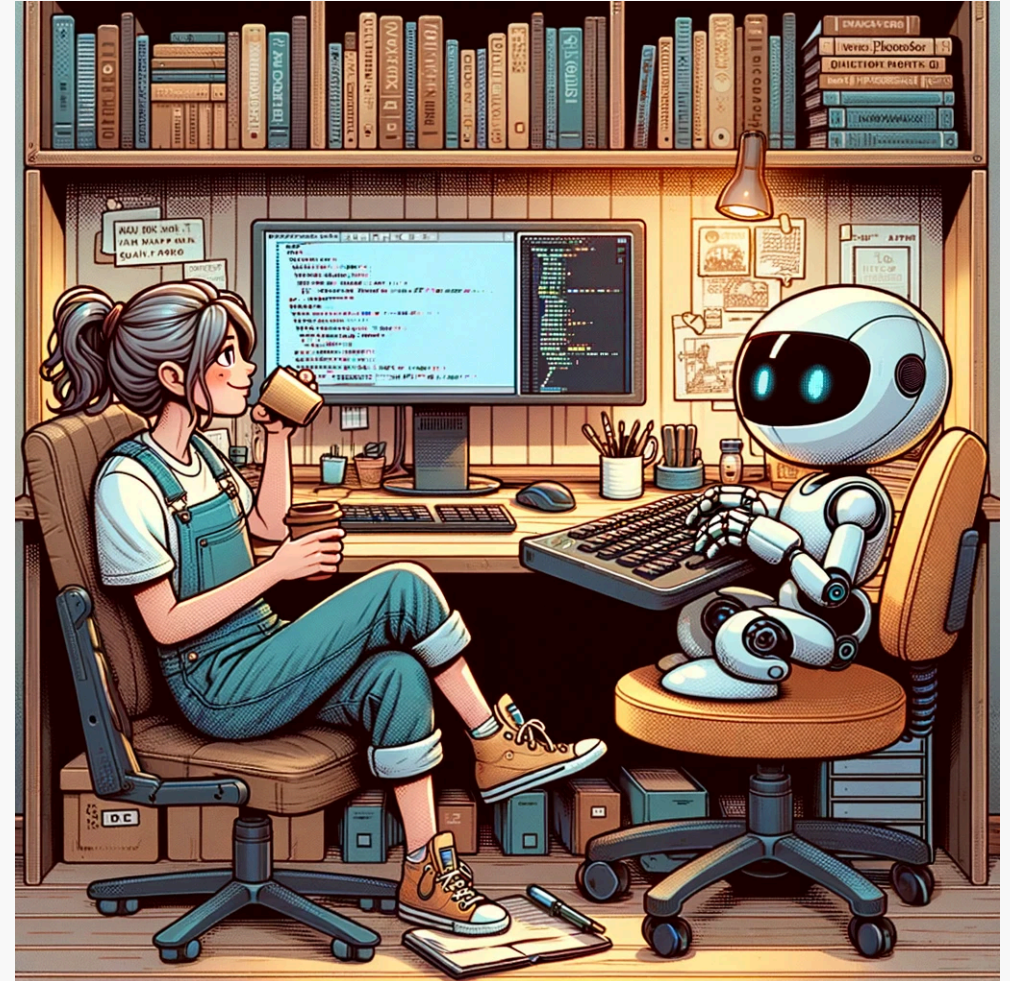## Day 1 - Introduction to Data Analysis with R

Selina Baldauf

Freie Universität Berlin - Theoretical Ecology

March 4, 2025

Motivation

- AI tools assist programmers with
    - Coding
    - Debugging
    - Learning
    - ...
- Higher productivity and efficiency
- More motivation
- But careful: You still need to understand what's going on!

Overview of tools

- Browser-based chat bots (ChatGPT, Bard, …)
  - General-purpose

- Data-analysis tools (Julius AI, RTutor, …)
  - Upload data and ask questions about it
  - Can also execute code
  - Chat with your data

- Integrated AI tools (GitHub Copilot, Codium AI, …)
  - Integrated directly in programming environment
  - Real-time suggestions, chat, debugging, …
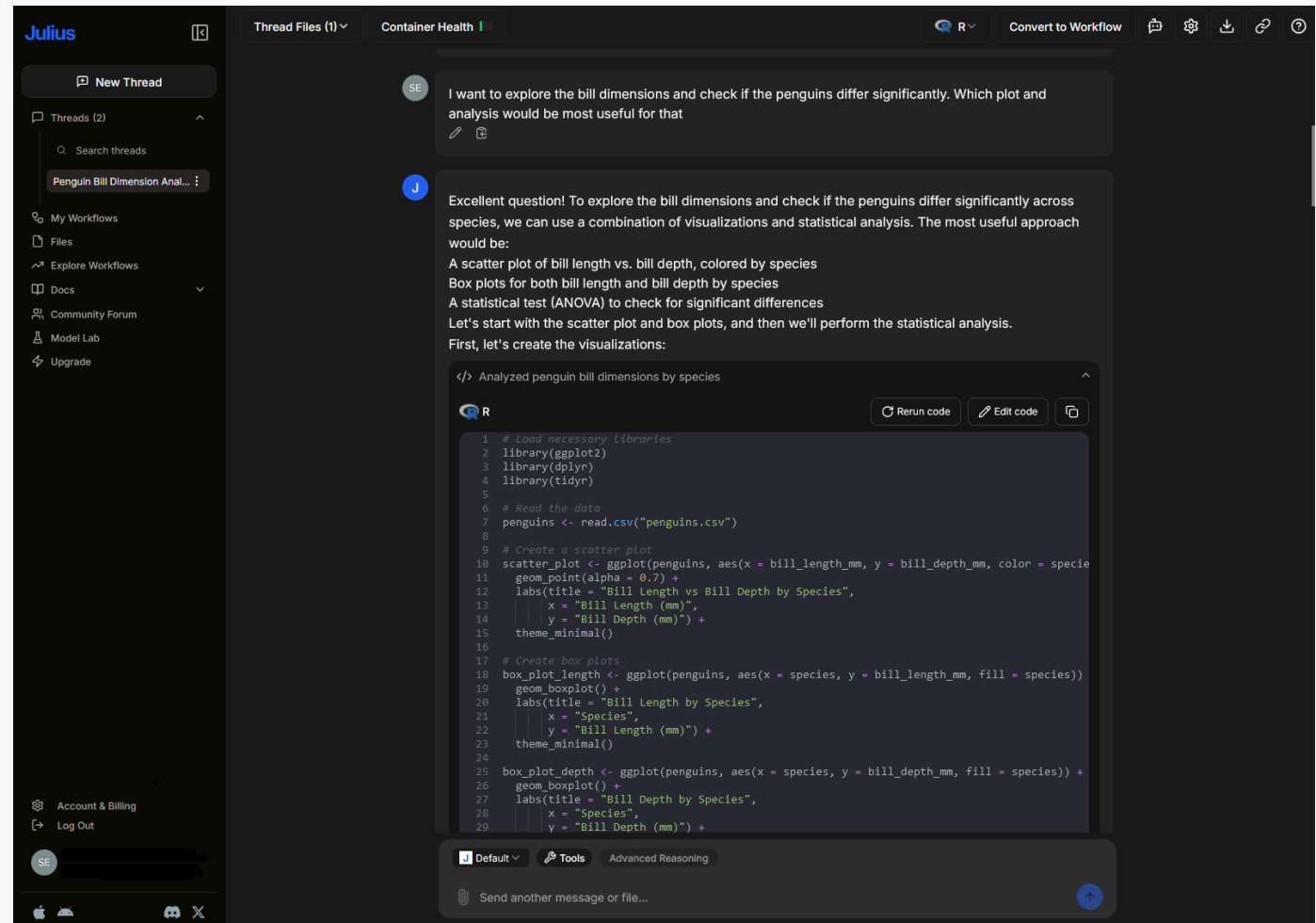
Find the tools that best fit your workflow!

R Tutor

- https://rtutor.ai/

- Free browser tool

- Upload data and ask questions about it

- Use the demo data

Julius AI

- https://julius.ai/

- Try for free

- Basic plan ~20€ per months (-50% academic discount)

- Upload data and ask questions about it

Github Copilot

- https://github.com/features/copilot

- Model based on GPT-4 and OpenAI's Codex

  - Specifically trained on source code

- Basic idea: Integrate directly into R Studio (or other IDEs)

- Works best for well-represented languages (Python, JS, ...)

How to get GitHub Copilot

See this website for step-by-step guide and more information.

It's really easy, but you need:

- GitHub Account

- Active GH Copilot subscription (10$ per month)
  - Get it for free as an academic with an educational account

- IDE that supports Copilot
  - Full support: Visual Studio (Code), Vim, Neovim, JetBrains IDEs (e.g. PyCharm)
  - Limited support: RStudio, ?

GH Copilot: Inline code suggestions

- Copilot tries to predict what you want to do next

- Suggestions are based on the context

  - Previous code

  - Comments

  - Variable and function names

  - ...

```r
fibonacci <- function(n) {
  if (n == 0) {
    return(0)
  } else if (n == 1) {
    return(1)
  } else {
    return(fibonacci(n - 1) + fibonacci(n - 2))
  }
}
```

Get better suggestions

- Provide context
  - Open other files
  - Add top level comments explaining the purpose of the script
  - Name variables and functions properly
  - Copy-paste sample code and delete it later

- Be consistent
  - "Garbage in, garbage out"
  - Have a nice and consistent coding style

Nice side effect of using Copilot: More good-practice coding

Chat with GH copilot in R Studio

- Available through the `chattr` package

- Chat with Copilot in the sidebar

- Also supports other LLMs (e.g. GPT4o, …)

Concerns to consider

- Privacy
  - Chose whether your prompts and suggestions will be used by Github (`Github -> Seetings -> Copilot -> Policies`)
  - Check privacy guidelines before you upload data
- Plagiarism
  - Block suggestions matching public code (`Github -> Seetings -> Copilot -> Policies`)
- Ethical concerns
  - For-profit tool trained on open-source
- Environmental concerns
  - Water and energy usage

Usage guidelines

- No definite guidelines, but see examples listed here

- **Responsibility**

  - You are responsible for your scientific output

  - Stay critical, double-check

- **Transparency**

  - Make clear for which tasks you used which AI

- **Know relevant guidelines**

  - Journals

  - Your university

- **Still understand what is happening!**