Introduction to Data Analysis with R

Day 1: Welcome to the workshop

Selina Baldauf
Freie Universität Berlin - Theoretical Ecology

September 29, 2025

Who am I?

- 💻 **Scientific programmer** @theoretical ecology group
- 🎓 PhD in **dryland ecology**
- 👩‍🏫 **Teaching** R, Git, good scientific practice, …

I'm using R a lot for ...

... data analysis

... ecological modelling

... writing documents/websites/presentations

.... workflow automation

...

Who are you?

What are the molecular basis of **learning and memory in the brain**?

Is this newly developped **sample preparation** technique suitable for use in **clinical diagnostics**?

Biophysics

How do **neurons connect** to their partners?

Biochemistry

Behavioral economics

To study the **biofilm formation** process in flow by analyzing the fluorescent images

Environmental ecology

Structure and Function study of the **Glycogen Debranching Enzyme**

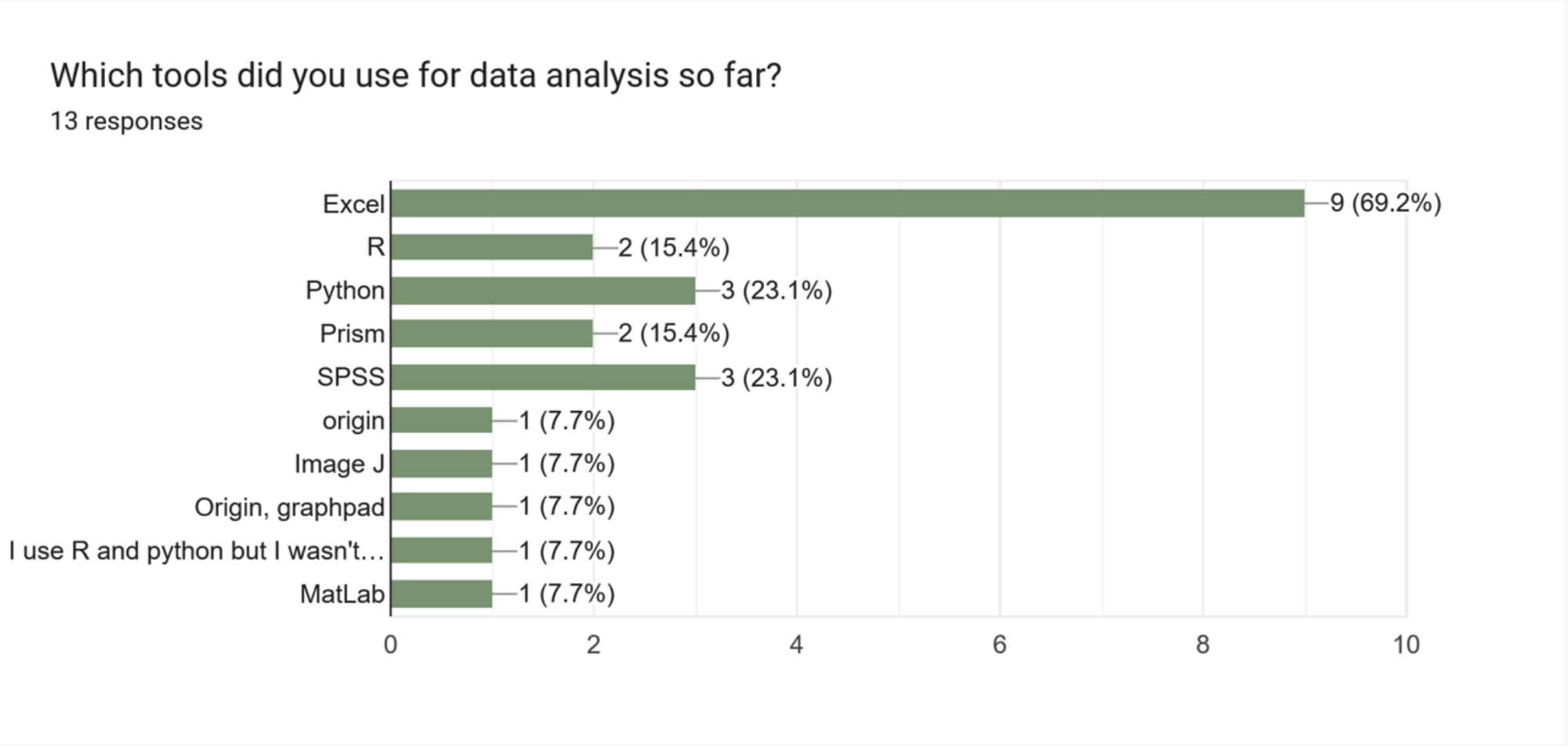
Department of Economics and Government of Agriculture, Food and Natural Resources

The impact of tripartite interactions among host immunity, microbiota, and pathogens on **pathogen virulence evolution**

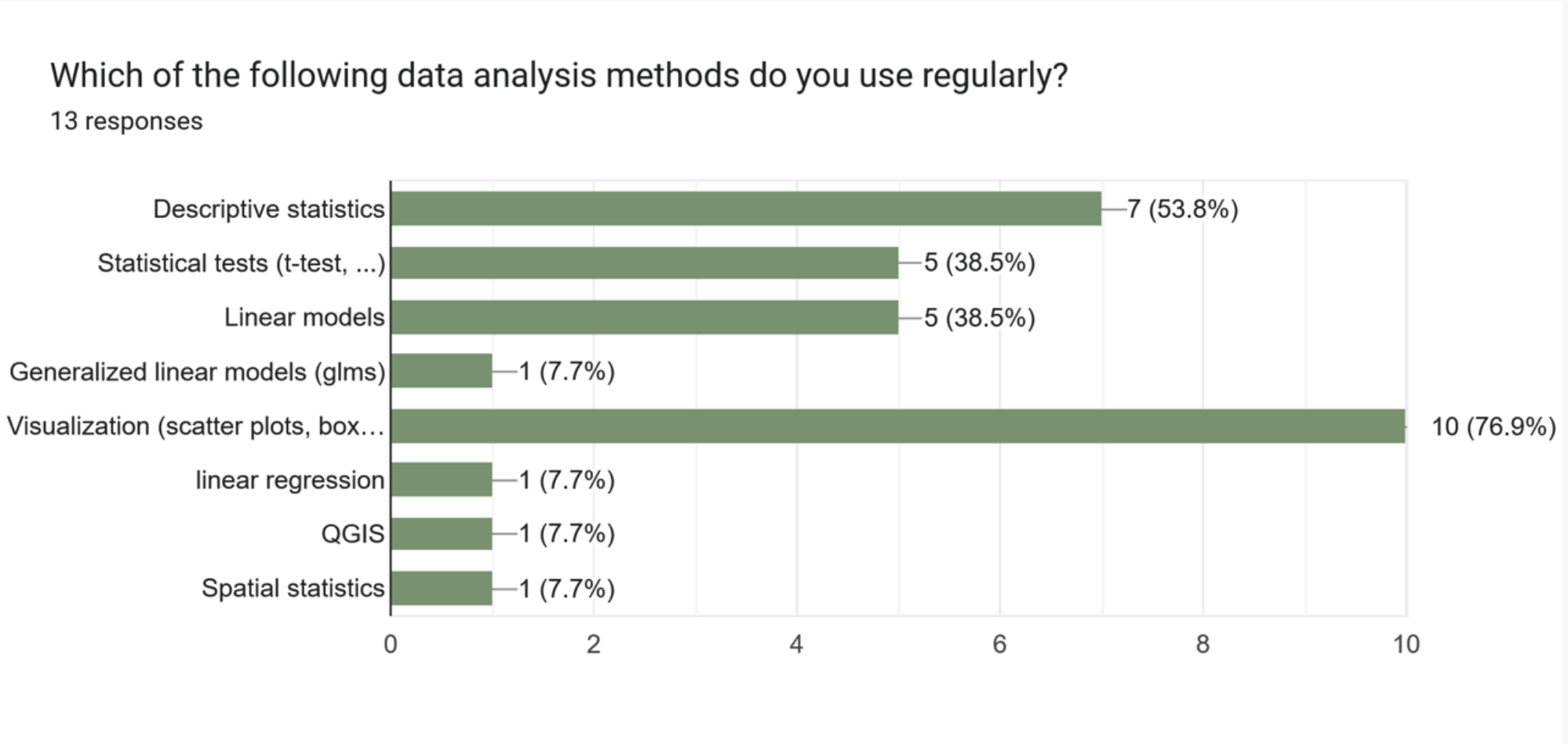
How can biochemical markers and comparative methods contribute to the monitoring of **ash dieback in Germany**?

The effect of community diversity on **bacterial spatial patterns in the phyllosphere**

Who are you?


Which tools did you use for data analysis so far?
13 responses
Excel
R
Python
Prism
SPSS
origin
Image J
Origin, graphpad
I use R and python but I wasn't…
MatLab
9 (69.2%)
2 (15.4%)
3 (23.1%)
2 (15.4%)
3 (23.1%)
1 (7.7%)
1 (7.7%)
1 (7.7%)
1 (7.7%)
1 (7.7%)
0
2
4
6
8
10

Who are you?


Which of the following data analysis methods do you use regularly?
13 responses
Descriptive statistics
7 (53.8%)
Statistical tests (t-test, ...)
5 (38.5%)
Linear models
5 (38.5%)
Generalized linear models (glms)
1 (7.7%)
Visualization (scatter plots, box...
10 (76.9%)
linear regression
1 (7.7%)
QGIS
1 (7.7%)
Spatial statistics
1 (7.7%)
0
2
4
6
8
10

Workshop topics

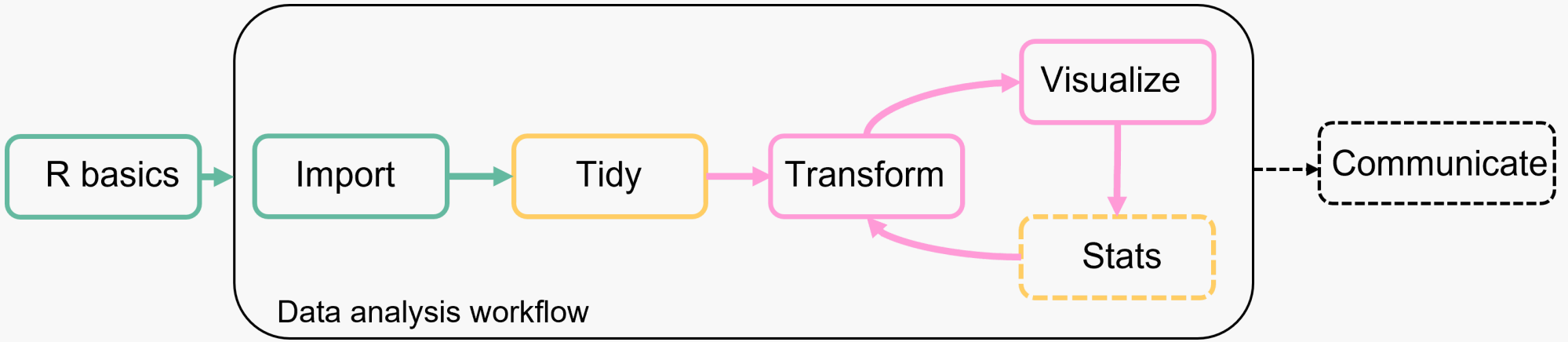


Image adapted from Wickham & Grolemund: R for Data Science

Day1 Introduction to R and RStudio and data import

Day2 Data transformation and visualization with the tidyverse

Day3 Cleaning data, statistical tests, good practice and AI tools

Day 4 Bring your own data

Day 5 Bring your own data + Q&A (optional)

Schedule and Organization

🗓️ 03.03.2025 - 04.03.2025 from 🕘 9 a.m. - 4 p.m.
🗓️ 10.03.2025 - 11.03.2025 from 🕘 9 a.m. - 4 p.m.
🗓️ 17.03.2025 🕘 9 a.m. - 12 a.m.
🍜 ~ 12 a.m. - 1 p.m., ☕ in between
📌 We will meet in the `General` meeting on Webex

1. Input sessions

- Presentation and demonstration of a topic + examples

2. Tasks

- Solve them in small groups

3. Joint discussion of tasks & questions

Material

- All material can be found on the workshop's website
  - Presentations, Tasks, Solutions, Additional resources
- You can download slides and R Scripts from there
- Website will stay online after the workshop

Bring your own data

On the last workshop day, you can **work with your own research data**. I will also provide some **real life data sets** from different topics.

Learning by doing

- Get started using R for your own analyses
- Use any of the methods from the course or try new things, ...
- Present/discuss your results, questions and problems at the end of the day

- Add your name and some details on what you plan to do in this joint table

Before we get started I

- **Help each other** if possible
- Have an eye on the chat and on the time
- All questions and comments are welcome
- If possible, please turn on your camera
- Feedback is welcome (Evaluation at the end of the workshop)

Before we get started II

Did anyone have problems installing R and RStudio?

Download and install R from https://cran.r-project.org

Download and install RStudio from https://www.posit.co

Before we get started III

How to use Webex teams

- `General` channel for our joint meetings and chat
- `Groups 1-4` for solving tasks jointly
  - Group spaces have their own chat and meetings
  - Ask questions in the chat, share screenshots, share your screen and talk

Now you (10 min)

> Get to know your team

- **Go into your groups** and **start a meeting** there
- **Introduce yourself** to your team (2 minutes for everyone)
  - What is your PhD about?
  - Which tools did you use so far for data analysis?
  - Which types of data analysis do you need?
- **Come back to the general meeting** after 15 mins