Introduction to Data Analysis with R

Day 1: Welcome to the workshop

Selina Baldauf

Freie Universität Berlin - Theoretical Ecology

October 15, 2023

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

.... workflow automation

...

Who are you?

Associations between behaviour and gut microbiota in laying hens

How did past amber forests look like in Asia?

Cell assay development and urinary excretion studies

Biodegradable polymers

Emotional processing in the brain

Does the reuse of pineapple residue through soil incorporation in pineapple cultivation effective to improve nutrient cycling and C sequestration in the soil?

What effect has air pollution on skin?

Nitrogen inefficiencies, nitrate leaching caused by different organic and mineral fertilization systems.

Influence of various pharmaceuticals on human UDP-glucuronosyltransferase activity

Chicken immunology

How does the hepatic thyroid hormone status change on modulating hepatic deiodinase enzyme?

Dissect epidermal cell development in Arabidopsis via single-cell technology

What influence do shocks and crisis situations have on individual food behavior?

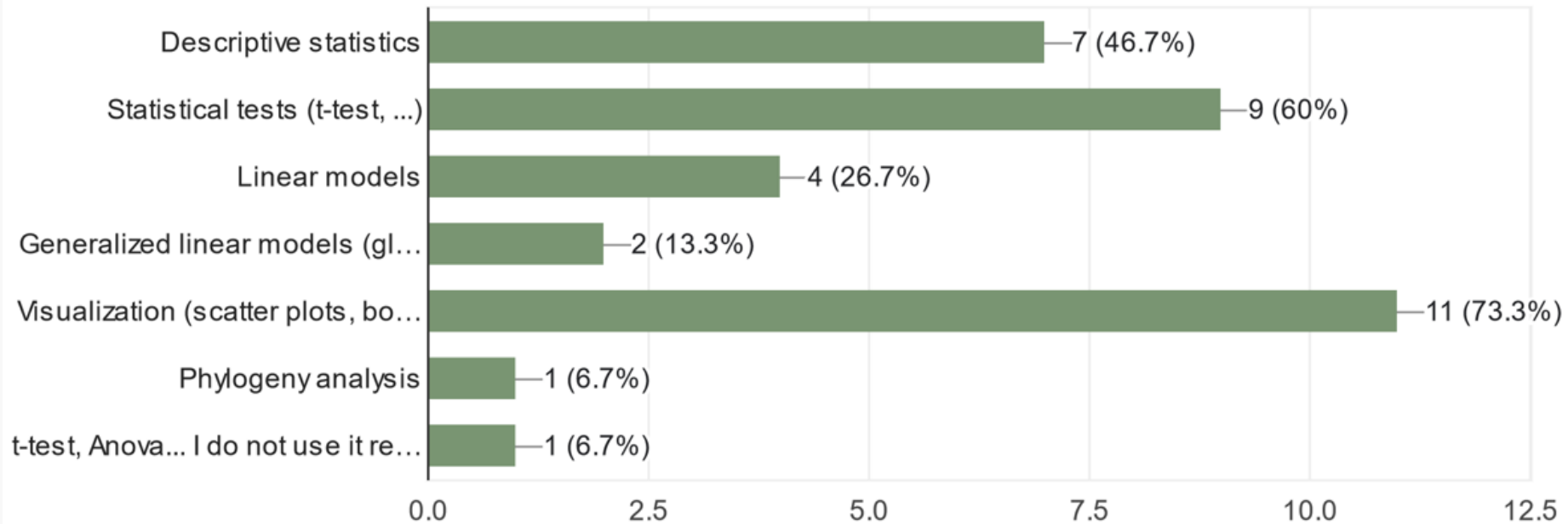# Who are you?

Which tools did you use for data analysis so far?

15 responses

Tool	Responses
Excel	13 (86.7%)
R	6 (40%)
Python	0 (0%)
Prism	3 (20%)
SPSS	1 (6.7%)
Origin	2 (13.3%)
TNT, BEAST	1 (6.7%)
Matlab	1 (6.7%)
I was trying with SAS, but rece…	1 (6.7%)
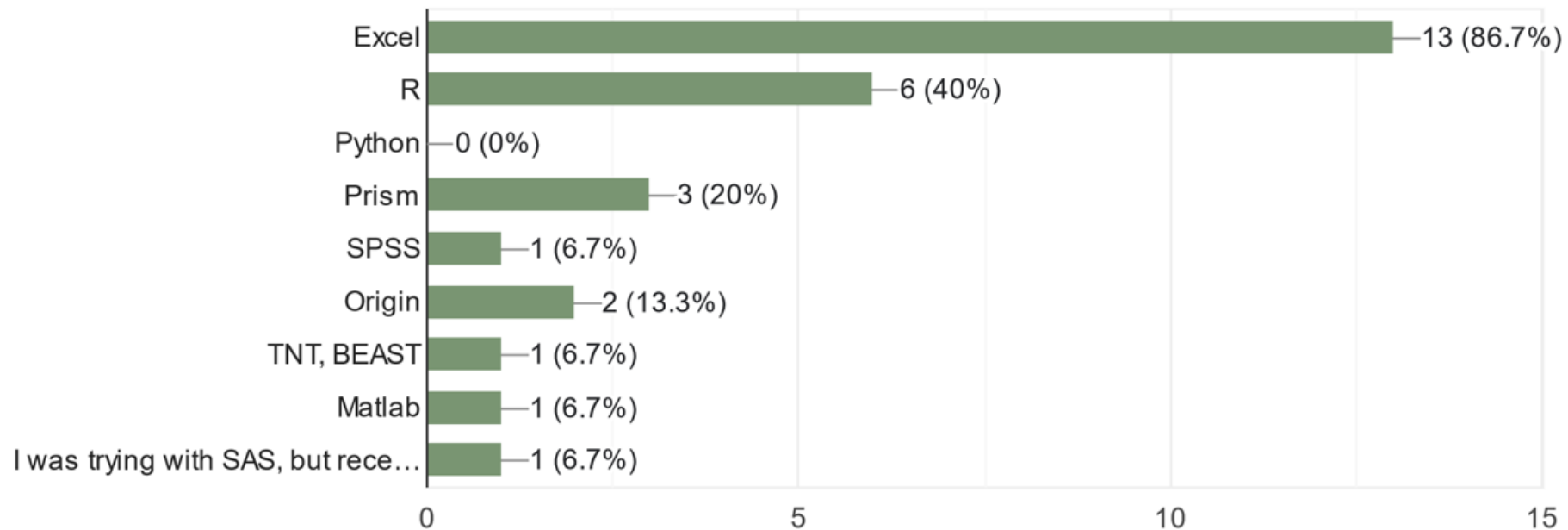
Who are you?

Which of the following data analysis methods do you use regularly?

15 responses

Method	Count (Percentage)
Descriptive statistics	7 (46.7%)
Statistical tests (t-test, ...)	9 (60%)
Linear models	4 (26.7%)
Generalized linear models (gl…	2 (13.3%)
Visualization (scatter plots, bo…	11 (73.3%)
Phylogeny analysis	1 (6.7%)
t-test, Anova… I do not use it re…	1 (6.7%)

Who are you?

Which type of plot do you use regularly?

15 responses

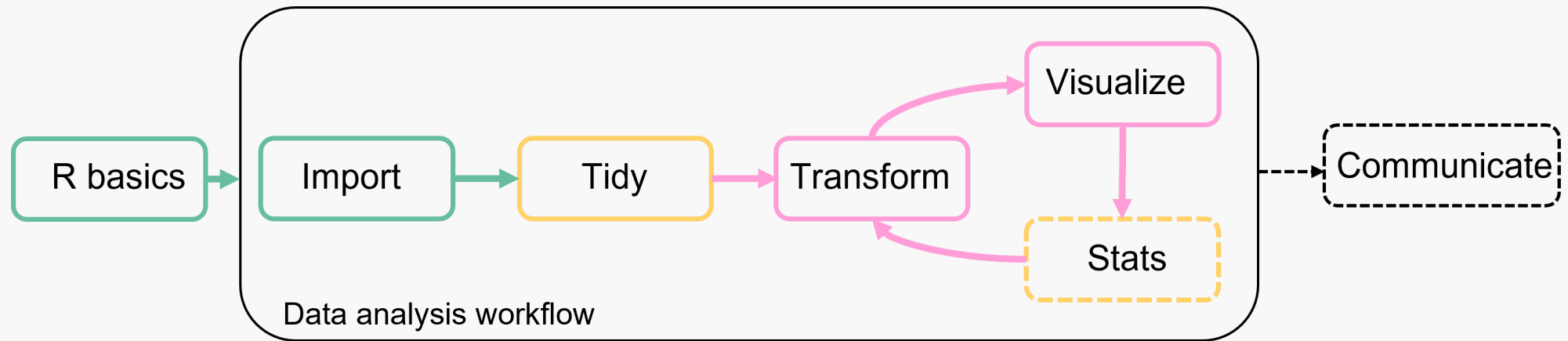| Plot type	Count
Boxplots	8 (53.3%)
Barplots	8 (53.3%)
Histograms	6 (40%)
Scatterplots	6 (40%)
Time series	3 (20%)
Maps	1 (6.7%)
Heat maps	2 (13.3%)
Trees	1 (6.7%)
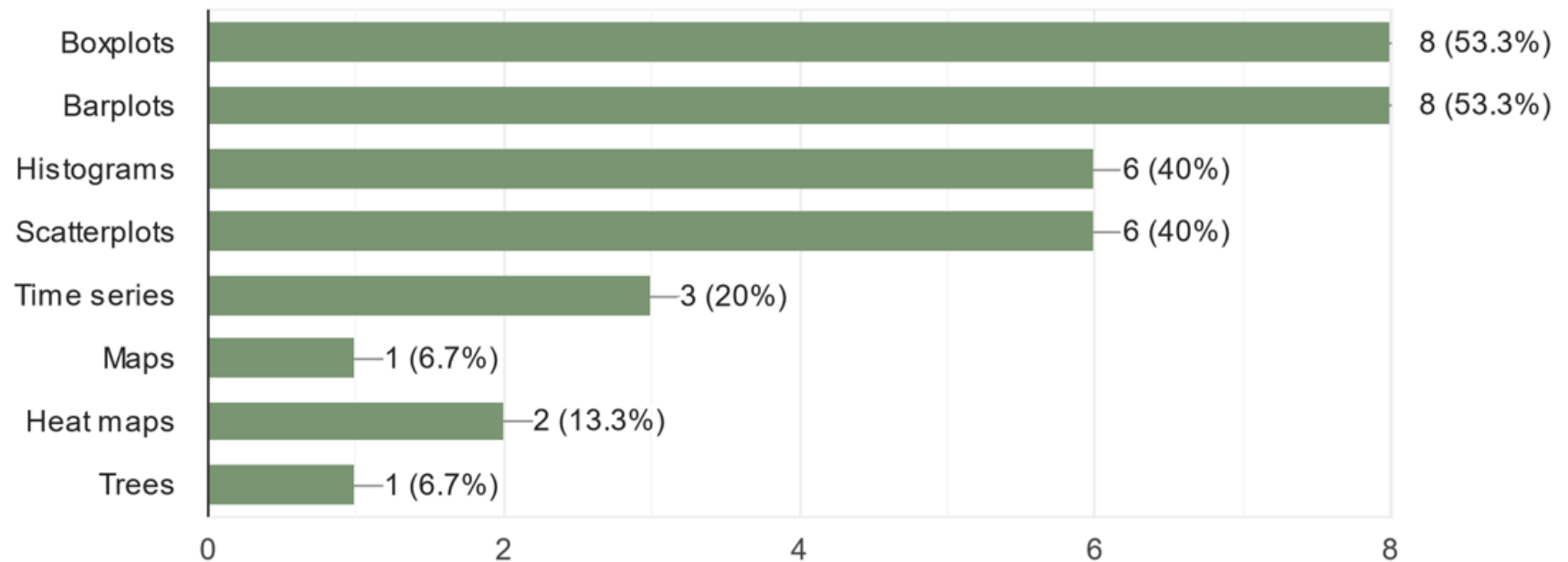
Workshop topics

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 and statistical tests

Day 4 Bring your own data

Schedule and Organization

📅 16.10.2023 - 17.10.2023 from 🕘 9 a.m. - 4 p.m.
📅 30.10.2023 - 31.10.2023 from 🕘 9 a.m. - 4 p.m.
🍜 ~ 12 a.m. - 1 p.m., ☕ in between
📌 We will meet in the `General` meeting on Webex

Input sessions

- Presentation and demonstration of a topic

- Some examples

Tasks

- Solve them in small groups

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

Bring your own data

- Keep the last workshop day in mind during the next days
  - Take a mental note if you learn something that is applicable to your data
  - Think of questions that you would like to answer for your data set
- 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
- All questions and comments are welcome
- Feedback is welcome
  - Evaluation at the end of the workshop

Before we get started II

How to use Webex teams

- `General` channel for our joint meetings and chat

- `Groups 1-4` for group work

    - Group spaces have their own chat and meetings

    - Use the chat in the groups for questions

    - Meet in your groups and solve the tasks jointly

- Please always have a look at the general chat during the tasks

Now you (5 min)

Let's try the Webex groups

- **Go into your groups** and **start a meeting** there
- **Say hello** to your team briefly
    - What do you study and what types of data analysis do you do?
- **Come back to the general meeting** when I write in the general chat

Before we get started III

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