Freie Universität Berlin

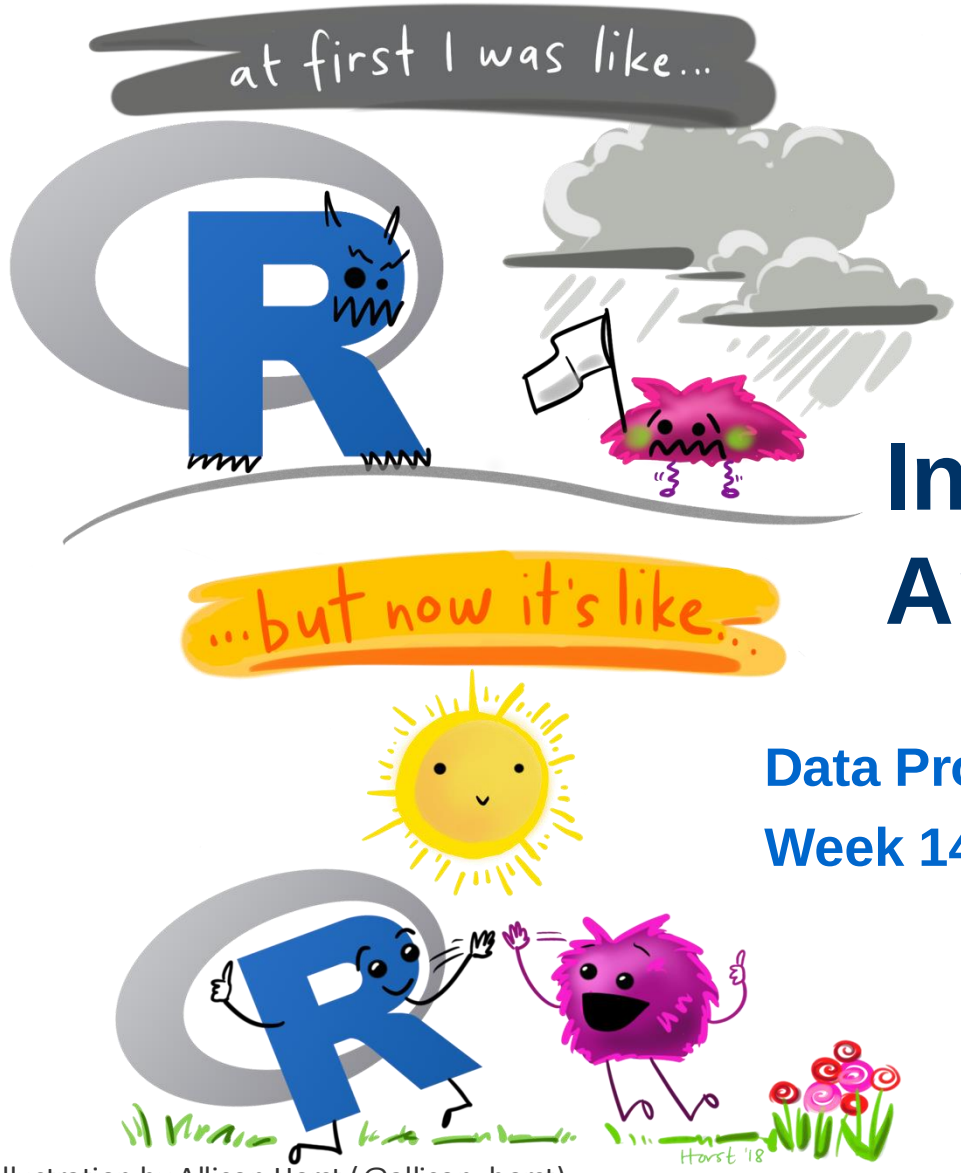

Illustration by Allison Horst (@allison_horst)

Introduction to Data Analysis with R

Data Project –
Week 14, 30.01.25

Assignment 3 (Start: 30.01.; Deadline: 15.02.)

1. Ideally, form a group of 3-4 students.
2. Find a dataset that interests you.
   - Suggestions: ESS, WVS, CSES, SOEP…
   - You can also search for datasets here:
     - https://www.icpsr.umich.edu/web/pages/index.html
3. Look at the codebooks and think about a research question that you can answer with your dataset.
   - Pick a dependent variable from your dataset and think about other variables that could have an relationship with this variable.
   - Your main hypothesis should be a correlation between at least two variables.
   - Think about additional variables that can explain your effect and add these! Think about WHY these could affect your relationship.
4. Clean your data, present your descriptives, run fitting linear regressions and visualize your data however you like with plots and tables.
5. Document how you run your analysis and what you have found.

Assignment 3 (Start: 30.01.; Deadline: 15.02.)

What you need to submit (1 submission per group):

- **R script**
  - Script you use for your analysis
  - Should include some comments to make it understandable
- **Documentation**
  - Includes your research question and hypothesis
  - Includes notes about your used variables and data
    - What data do you use and which variables and why?
    - Include visualizations to show interesting relationships
  - What do you find when you look at the descriptives of the data?
    - Include a summary table and describe it
  - Run a linear regression with your chosen variables
    - Include a regression table and describe your findings
    - Can you accept your hypotheses? What can you say about your research question?

How to form Research Questions and Hypotheses

- **Research Question:**
  - Should be a broad question about your study object of interest
  - Can be used for quantitative and qualitative studies
  - **Example:**
    - **RQ:** How does social media influence young adults own body image?

- **Hypotheses:**
  - Implies a statistical relationship with a certain expected direction
  - **Example:** Let's assume we have survey data about social media and body image…
    - **H1: The higher** the social media consumption the lower the self-esteem.
    - **H2:** Young people **are more likely** to agree that their beauty standard is inspired by social media compared to older people.
    - **H3:** TikTok influences the body image **more than** other social media.

Recommended Schedule 30.01.

1. Get together with your group (Suggestion: 3-4 people).
2. Figure out **what you want to research** and **find a fitting dataset**.
3. Formulate a **research question**.
4. Think about potential results and **formulate some hypotheses**, that you can test later. Try to formulate at least two.
5. Think about control variables you would like to add.
6. Start loading in your data, selecting your variables and clean them by sorting out missing values.
7. If still time: Start exploring your data through data visualization.

Don’t forget to document your decisions and steps.

Recommended Schedule 06.01.

1. Look at the descriptives of your variables, describe them and create a summary table.
2. If not already done: Play around with graphs and visualize your data.
3. Found anything interesting?
   - Take a look at your dependent and independent variables as well! Look at their distribution!
4. Run your analysis as planned with your hypothesis. Don't forget your control variables.
   - Do you find what you expected?
   - Describe your findings! → Create scatterplots and tables!
5. What did your findings contribute to answer your RQ?

Don't forget to document your decisions and steps.