Dr hab. January Weiner
Berlin Institute of Health
Charité – Universitätsmedizin Berlin

Reproducible statistical analyses with R

Incorrect and poorly documented statistical analyses and a widely spread lack of transparency of statistical methods in published works are thought to be a major reason for the current reproducibility crisis in science. The statistical programming language R allows a working implementation of Donald E. Knuth's "literate programming" paradigm, in which the human readable documenation of statistical procedures and the programming code used to perform them are one. This approach allows to generate transparent, well documented and reproducible statistical analysis and even ready to publish manuscripts.

This course serves as an introduction to both R and reproducible science with R and Rmarkdown. While it is not possible to give a thorough training in a programming language in such a short time, after the course the students will have learned the basics of using Rstudio, loading and viewing data, performing simple statistical analyses and creating a reproducible analysis protocol.

5 meetings:
6th, 13th, and 20th of May (6pm-9pm)
11th and 18th of May (5pm-8pm)

Detailed description of the course:

1. Reproducibility crisis and statistics: how data wrangling, p-hacking and cherry picking will make you fool yourself.
    a. Practical part: introduction to R
    b. Practical part: functional programming principles in R
    c. On-hands demonstration: simulating research with R demonstrates the importance of transparency and reproducibility in research
2. Reproducibility, replicability and repeatability: they are all different
    a. Practical part: basic plotting in R
    b. Practical part: creating functions
    c. On-hands demonstration: from data to publication
3. Literate programming: the concept of computer language as means to scientific communication
    a. Practical part: basic statistics with R
    b. Practical part: introduction to markdown Rmarkdown
    c. Practical part: generating a simple analysis with Rmarkdown
4. The garden of forking paths: how planning and choosing the correct statistical analysis influences the outcome
    a. Practical part: basic linear modeling with R
    b. On-hands demonstration: example data analysis workflow in R markdown

c. On-hands demonstration: using nlme / lme4 for mixed effects models in R

5. Making your research transparent: tools and procedures for publishing your workflows, analyses, findings and data

a. On-hands demonstration: using git and github

b. Practical part: creating and using a git repository