

## Content

R has become a leading statistical programming language in data science and statistics. The R software is free, highly extensible, and allows to easily produce publication-quality plots. The aim of this course is to equip participants with the basic skills to effectively use the software R, embedded in the user interface RStudio. These skills will cover data import, processing and data exploration, graphical representations with ggplot2, basic concepts of statistical inference and linear regression. All topics will be demonstrated in working examples and practiced with hands on examples.

The course will be split in two parts. In the first two days we are going to teach the basic functionality of R, RStudio, and ggplot2. Day three and four are dedicated to the application of statistical tests, which will be extended to linear mixed-effects models.

## Specific Topics

### Part I (April 23-24)

- Introduction to R and RStudio
- Data types (vectors, data frames, matrices, lists, factors)
- Explorative data analysis
- Creating functions and loops
- Creating and customizing plots with ggplot2

### Part II (April 25-26)

- Recap of the principles of statistical testing
- Linear regression models
- Linear mixed-effects models
- Model evaluation and model selection

## Aim

The workshop should prepare participants to read, clean, and explore their own datasets. They will become familiar with the basic workflow of statistical analyses and should become capable to conduct their own analyses, even if depending on methods that were not taught during this course. In short, this course should allow students to confidently work with the R language to apply their own analyses.

## Methodology

The workshop will consist of lectures, working examples and practical exercises, each covering roughly one third of the course time. Day one and two are dedicated to the handling of R, RStudio and ggplot2. After participants have acquired the basic functionality the days three and four will then be dedicated to linear regression and model selection.

Kindly note:

For students already familiar with statistics it might thus be possible to only participate in the first half of the course (Part I) whereas students who are already familiar with R might wish to participate only on day three and four (Part II).

## Target Group

Students, PhD students and Postdocs which want to become acquainted with the R language and the basics of linear regression and model selection. No previous knowledge in R or statistics will be required.

## Term and Application

- April 23-26, 2019, each day 9am – 4pm
- New** Venue: 06120 Halle (Saale), von-Seckendorff-Platz 4
- Group size: maximum 20.
- Workshop-Language: English.
- **Application period between March 18 and April 18**  
email to: [koordination@ingra.uni-halle.de](mailto:koordination@ingra.uni-halle.de)  
*with specification of the required participation on Part I and / or Part II*
- Applications are considered in chronological order of receipt. You get information.

**23.04. - seminar room 1.34**

**24.04. - seminar room 3.21**

**25.04. - seminar room 3.21**

**26.04. - seminar room 1.34**

## Trainer

Dr. Stephan Kambach, PostDoc at the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig

Christian Ristok, M.Sc. Biology, PhD student at the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig and Scientific employee at the Bioinformatics Unit of iDiv and the Martin Luther University Halle-Wittenberg

<a href="http://www.ingra.uni-halle.de">http://www.ingra.uni-halle.de</a> – <a href="http://www.ingra.uni-halle.de/courses/">http://www.ingra.uni-halle.de/courses/</a>
---