WORKING WITH DATA

COURSE CONTENT

Course participants will be able to:

• import data from a wide variety of sources in the R-environment
• manipulate data to obtain a dataset that can be used for machine learning tasks
• visualize data using a systematic, grammar-based approach
• generate offline reports
• version control of projects using Git
• generate interactive reports using Shiny

The course will cover following topics:

Working with data:
• Importing data from various sources (SAS, SPSS, Excel) (‘haven’ package)
• Visualizing data using a “Grammar of Graphics” (‘ggplot2’ package)
• Modern data transform (‘dplyr’ package)
• Basics of relational data and “tidy” data (‘tidyr’ package)

Reproducibility:
• Why reproducible research is essential to good scientific practice
• RMarkdown, knitr, and Sweave for automatic report generation
• Version control using Git

Presentation:
• Offline reports (‘knitr’, ‘rmarkdown’ package)
• Interactive web presentations using Shiny

All parts include practical exercises. Please preinstall R Studio

SCHEDULE*
• Thursday 09:00 - 17:00, Friday 09:00 - 17:00, Saturday 09:00 - 12:30

PRE-REQUISITE
The participants must have basis knowledge in:
• R programming

REGISTRATION
• Deadline for registration is 3 weeks before.
• The fee for the course is € 645; discounted rate for affiliated with a university € 600.
• The courses may take place in parts or as a whole online (virtual conferences) if in-classroom teaching is not possible.

*subject to change