

Data integration, pre-processing and data fusion in R

§ Name of instructor

Domenico Vistocco, University of Naples Federico II, Italy

§ Short description

Real-world data is messy. Data preparation and wrangling are crucial tasks to prepare real-world data for analysis. The lecture will focus on the technique useful in R to obtain tidy data, to clean data, to deal with missing values, to create new variables, and to recode existing variables.

§ Schedule

- Tidy data and the main different types of messy data
- Pivoting data, Rectangling data and Nesting data
- Splitting and combining data
- Joining data
- Recode and manipulate variables
- Basic treatment of missing variables
- Cleaning dirty data

§ Introductory background

- Tidyverse: <https://www.tidyverse.org/>
- Janitor: <https://sfirke.github.io/janitor/>

§ Facilities Required

- Software: R and RStudio, open source
- R packages: tidyverse and janitor
- Course Material. All course materials, including the data and R scripts for the examples, will be made available for course participants.