

Section 3

Getting and Cleaning Data

Practical Work

- Swirl : Interactive Learning environment for R
- Getting and Cleaning Data (4 hours)
 - Step 1: Install swirl
> install.packages("swirl")
 - Step 2: load swirl
> library("swirl")
 - Step 3: Install the interactive course
> swirl::install_course("Getting and Cleaning Data")
 - Step 4: Start the course
> swirl()

Getting & Cleaning Data Interactive Course

- Manipulating Data with dplyr
- Grouping and Chaining with dplyr
- Tidying Data with tidyr
- Dates and Times with lubridate

Getting Data

- Download files
- Reading Excel files
- Reading XML
- Reading JSON
- Reading from MySQL
- HDF5
- Web Scraping (tutorial)
- Working with API
- Reading from Other Sources

Data Cleaning R packages

- dplyr
- tidyr
- data.table
- lubridate

Data Manipulation

- Subsetting and Sorting
- Summarizing Data
- Creating New Variables
- Reshaping Data
- Merging Data
- Editing Text Variables
- Regular Expressions
- Working with Dates