

Introduction to R and the tidyverse

Paolo Crosetto

Lecture 0: getting started

The course

This course will be a *24 hour* brief introduction to R, from the point of view of the *tidyverse*, a set of packages developed by Hadley Wickham.

We will be dealing with the basics of data handling:

- ▶ data input & output
- ▶ data wrangling, selecting; filtering, mutating (with *dplyr*)
- ▶ data reshaping and merging (with *tidyr*)
- ▶ data plotting (with *ggplot*)
- ▶ tidy data analysis
- ▶ web crawling & scraping (with *rvest* and *RCrawler*)

Slides will be in English. **All** R documentation is in English. Out there you will be needing help, and it is good to be taught in the language you're going to get helped with.

Lecture plan: week 1

- ▶ **Lecture 0:** (tue 10th) setup, tools, workflow
- ▶ **Lecture 1:** (wed 11th) basic plotting
- ▶ **Lecture 2:** (thu 12th) data IO & basic data wrangling
- ▶ **Lecture 3:** (fri 13th) advanced data wrangling

Lecture plan: week 2

- ▶ **Lecture 4:** (tue 17th) tidy statistical analysis
- ▶ **Lecture 5:** (wed 11th) advanced plotting
- ▶ **Lecture 6:** (thu 12th) crawling data

Assignment: week 4

In **Lecture 7** (tue oct 3rd) each of you will present a **statistical report** on a topic of choice (list provided at the end of L6).

This report will be

- ▶ done within an .Rmd file
- ▶ hosted on your gitHub page, on your repository
- ▶ a mixture of text, code, results and plots

Example: https://jtanwk.github.io/us-solar/#how_have_solar_panel_costs_changed

Contacts

Paolo Crosetto

- ▶ paolo.crosetto@gmail.com
- ▶ Office 413, 4th floor, BATEG
- ▶ do not hesitate to pop by or send an email if needed