Introduction to R and the tidyverse

– getting started –

Paolo Crosetto

Know your teacher

The teacher

It's `Paolo`:

- a researcher at INRAE
- experimental economics, food consumption, risk
- passionate about econ, stats & R
- paolo.crosetto@inrae.fr or @gmail.com
- Office 432, 4th floor, BATEG (seldom there)

Know your course

Language

Slides will be in **English**.

- **All** R documentation is in English.
- Out there you will be needing help.
- Help with mostly be in English.

> `Best get used to it`.

The course

This course will be a *24 hour* introduction to R, using the *tidyverse*, a set of packages[1] developed by Hadley Wickham & people at Posit.

- data `input` & `output`
- data `wrangling`, selecting; filtering, mutating
- data `reshaping` and `merging`
- data visualization using `plots`
- data visualization using `tables`

Our take on R: in general

- **R is a programming language for statistics.**
- It can do **many** things:
  - these *slides* (and others too)
  - *websites* (e.g.)
  - *dashboards* (e.g.)
  - ...
- here we will focus on `descriptive data analysis`

Our take on R: in practice

> In practice, we will deal with `data.frames`

> We will manipulate, extract, analyze, plot, more or less complex `data.frames`

> We will **not** do high-level `R` programming or low-level `R` commands.

Schedules & lectures

Lecture plan (*no TD/CM distinction*)

- **L1** (2/9, 3h): setup, `tools`, workflow
- **L2** (3/9, 3h): basic data `wrangling`
- **L3** (18/9, 3h): basic `plotting`
- **L4** (17/9, 2h): `advanced` data wrangling
- **L5** (19/9, 3h): `advanced` plotting
- **L6** (1/10, 2h): tidy statistical `analysis`
- **L7** (2/10, 2h): basic & advanced `tables`
- **L8** (3/10, 3h): stat report `practice`

Evaluation plan

- **Report** (till 10/10): take-home `stat report`
- **Pres** (10/10, 3h): `presentation` of statistical report
- **Exam** (12/11, 2h): individual live coding `exam`

Material

All slides are written in `R` itself, using `quarto`[1].

For each lecture, you have (or will have):

- a `.qmd` file (source)
- the rendered `.html`
- a `notes_L[x]` file by me in `/Paolos Notes`
- an `exer_L[x]` file in `/Exercises`

All slides are available at the gitlab repository of the course

How lectures work

Me:

- there are **slides** (in `/Lecture X`)
- I take notes in a **notes file** (in `/Paolos_Notes`)
- I will provide exercises (in `/Exercises`)

You:

- you have a **personal folder** `/Student_Area/SURNAME`
- you take notes, make changes… **exclusively** there
- at home, copy exercises in your folder and solve them

Exam

Exam

The exam is composed of **three** parts

1. You `write` a **statistical report** at home on a data set. 1/3 of the final mark.

2. You `present` the report **in class** in front of peers. 1/3 of the final mark.

3. You **analyze** fresh data in an `individual live coding` session. 1/3 of the final mark.

Exam/1: report

You will have `until Oct 10th` to prepare a **statistical report** on a topic of choice.

- done within a `quarto .qmd` file
- hosted on the `gitlab` page of our course
- a mixture of `text`, `code`, `results` and `plots`

Example from elsewhere: here

Example from this course: `/Exam/exemple_2023_nobel_prize.html`

Exam/2: presentation

You will `present` the report to the class on `Oct 18th`. I will be asking questions.

- 5 minutes (sharp!) presentation
- 3 minute Q&A with questions from me
- no need to prepare slides (present your report)

Exam/3: individual live coding session

`Nov 12th` you will analyze data live, individually, in front of me

- You will get a `fresh dataset` (never seen before)
- 15 minutes to answer `two questions`: one easy, one hard
- You create a .qmd data analysis `on the fly`
- We discuss what you do live.
- You can use google, stackoverflow, LLMs… but within limits

LLMs

LLMs: the problem

> LLMs are very good coding assistants.
>
> `You will learn and code faster if you let them help you`.

> LLMs are very good at replacing you.
>
> `You will not learn anything if you use them full time`.

LLMs: the solution

> LLMs `as an assistant`? **No problem**

> LLMs `as a substitute`? **No way**

LLMs: the solution

You **can** use LLMs. But you will have to `understand` what they're doing.

`You` are the master, not the AI. This is why:

- You will be `presenting` your report in class, and I'll be `asking questions`.
  Don't know what your code does and how? `No points`.

- You will be `coding live` in front of me.
  Want to ask LLM for a particular piece of code? `Please do`.
  Want the LLM to solve the problem for you? `No points`.

Questions?

Ready? Go!