R4DS - Unit 1: Tidyverse

Andrea Gilardi
February 29, 2024

Outline and main concepts

- The objective of this course is to introduce a set of effective and modern tools for data science, version-control and R packages' development.
- At the beginning of this class, I will briefly mention a set of (**opinionated**) views that may improve your experience when developing R code.
- The examples are based on the Rstudio IDE (version 2023.12.1.402), but similar considerations hold for other Rstudio versions and different IDEs.

Outline and main concepts (cont)

- Then, we are going to briefly present the `tidyverse` and some of its most important packages via several examples.
- These practical examples we will based on a series of datasets shared by the Department for Transport: https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data
- We are not going to review the basics of the R language, but if you have any question feel free to ask!

But first, my favourite analogy!

Source: https://www.youtube.com/watch?v=7oyiPBjLAWY&t=448s&ab_channel=RConsortium

What They Forgot (WTF!!!)

- Always start R with a blank state!
- Adopt a project-oriented workflow.
- Practice safe "paths".
- Work and share examples in a reproducible environment, a so-called `reprex` (see Unit 2).

The examples are taken from `https://rstats.wtf/`. Another course on similar topics (not R-based and slightly more advanced): `https://missing.csail.mit.edu/`.

The tidyverse!

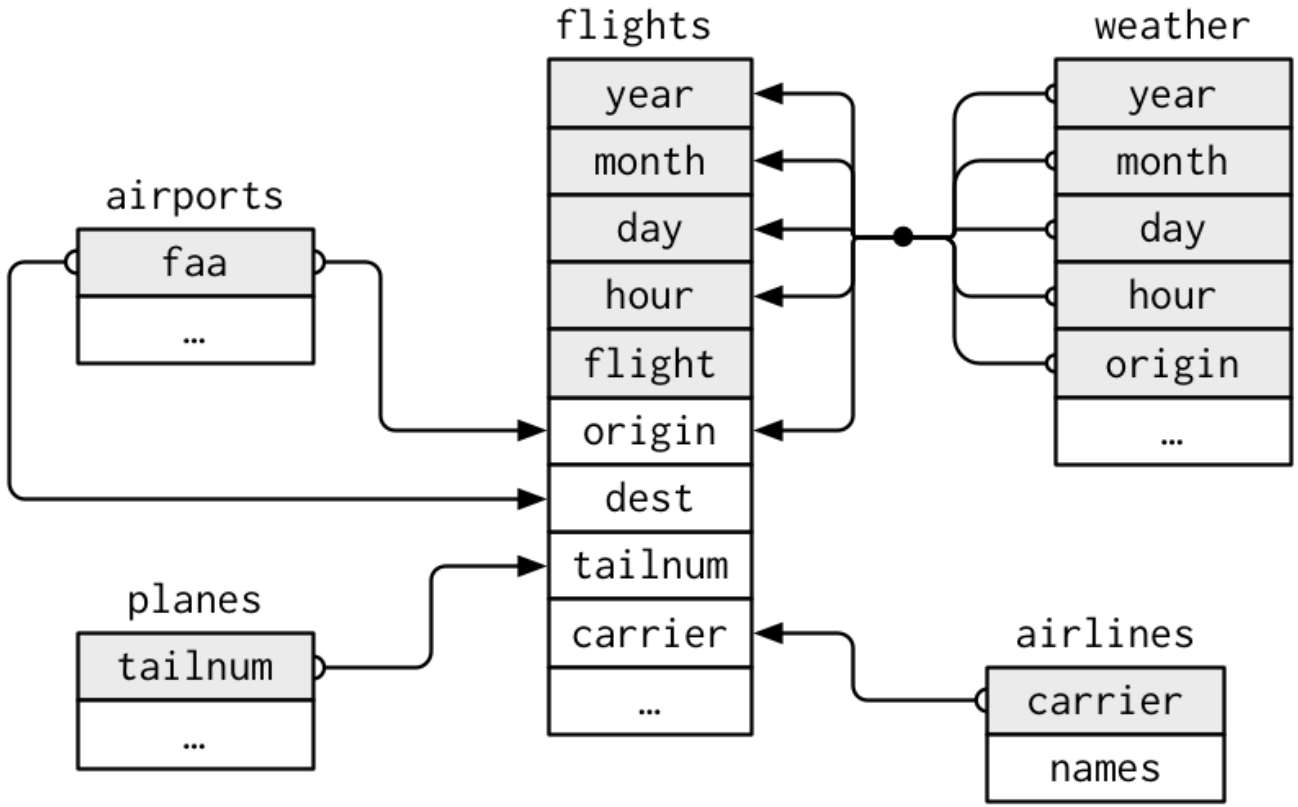
The tidyverse is a set of packages that work in harmony because they share common data representations and API design.

---

https://tidyverse.tidyverse.org/

EDA with the Tidyverse

We are going to briefly showcase the tidyverse toolkit using a series of relational dataset obtained from here.



Source: https://r4ds.had.co.nz/relational-data.html.

Enough theory, let's start coding 🥳