The workbench beyond the Carpentries

An analysis of the unofficial Carpentries lessons hosted on GitHub

Methods

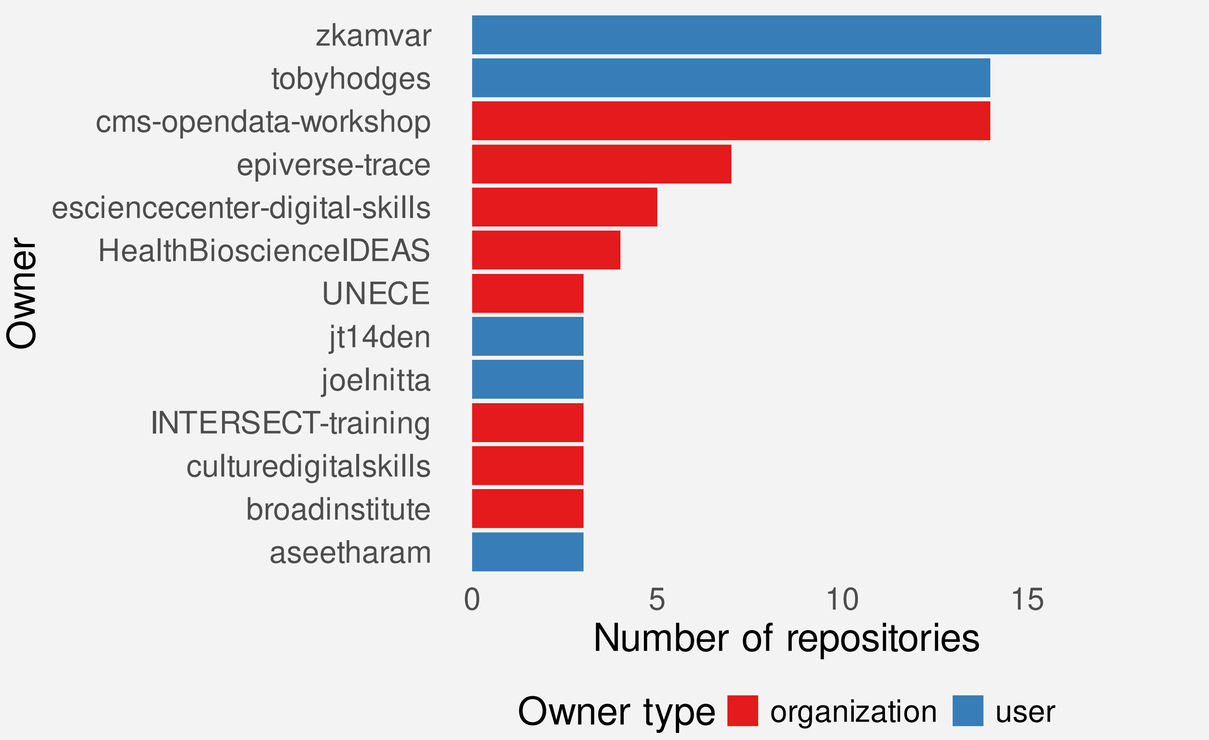
I used the GitHub API to:

1. Find repositories using the workbench via the presence of a sandpaper-version.txt or sandpaper-main.yaml file
2. Extract their config.yaml file for metadata (language, keywords, lifecycle, etc.)
3. Look for the presence of specifically worded PR or specific files

Who?

102 unique GitHub users or organizations host a total of **299 lessons** powered by the Carpentries workbench.

Individuals or organizations with more than 2 repositories



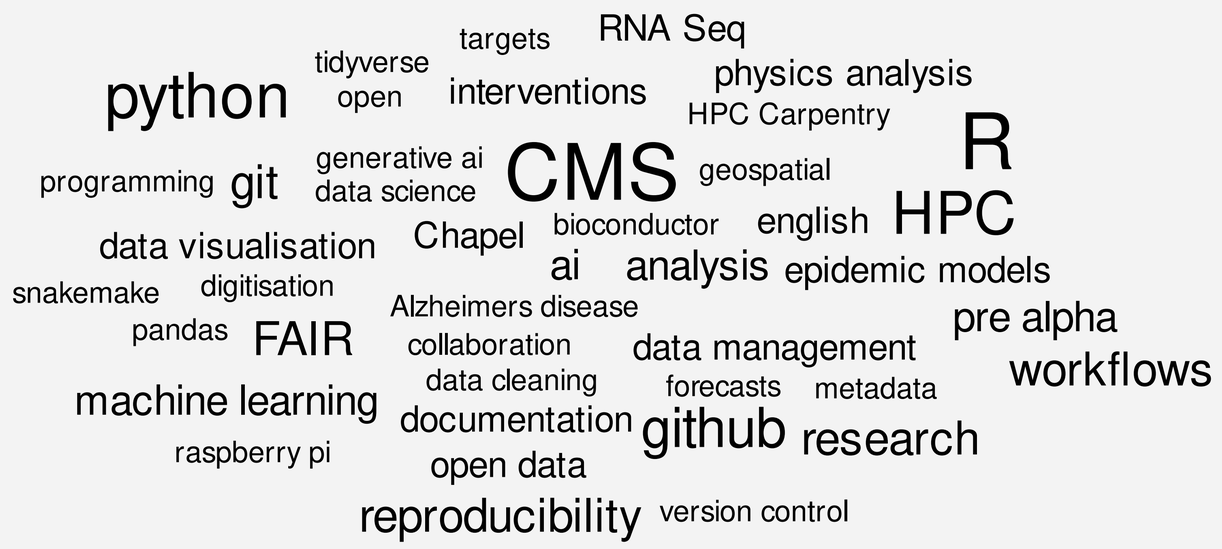
How?

- **69%** of the lessons use the Rmarkdown workbench structure (vs standard markdown standard)
- **40%** of the lessons use a sandpaper version from more than 1 year ago
- **45%** of the lessons use the dependency update bot
- **14%** use a custom fork of the workbench

For whom?

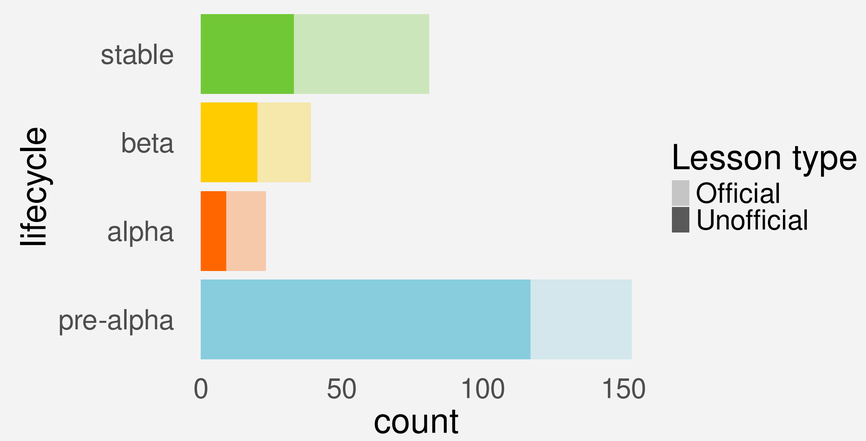
- **3 languages** are currently used in unofficial Carpentries lessons: English (76), Spanish (7) and German (2). Most lessons (214) do not explicitly document their language.

What? (keywords present more than once)



Unofficial lessons have the same distribution of lifecycles as official lessons.

Lifecycle of official and unofficial Carpentries lessons



Case study: Building the Epiverse-TRACE curriculum with the workbench

Benefits:

+ Embeds the Carpentries pedagogical approach
+ Open source (can be tweaked) and free to use
+ Accessibility validation
+ Active and responsive community
+ Plain text source files facilitate translations

Downsides:

- Custom theming requires maintaining a workbench fork. Until recently, we had to maintain 3 different forks.
- Decentralized deployment requires time (proportional to the number of lessons) and technical expertise from all lesson maintainers.

Content and values as of 2024-11-12

Live analysis available at https://hugogruson.fr/workbench-beyond-carpentries

Hugo Gruson
data.org