Journals: Introducing a Mock Dataset

Resul Umit

15 March 2021

Summary

This document introduces a mock dataset, which is generated with the *fabricatr* package (Blair et al., 2019). It includes nine variables on 1,091 fictitious journals. The dataset can be used for an imaginary exploration of these journals' Google Scholar rankings.

Variables

The dataset includes nine variables, as presented in Table 1.

Table 1: Variables in the dataset

Variables	Explanations
name	journals (1090 random titles)
origin	geographic origins (five continents)
branch	major discipline of journals (four branches)
since	time of first publication (years)
h5_index	H5 Index (integers)
h5_median	H5 Median (integers)
english	English (1) vs. other-language (0) journals
subfield	subfield (1) vs. generalist (0) journals
issues	number of issues published per year (integers)

Observations

There are 1,091 journal-level observations in the dataset. Figure 1 plots these journals' h5-index and h5-median — two popular metrics by Google Scholar. the figure divides the journals along the four branches of science (with facets; *formal*, *life*, *physical*, *social*) and two categories of scope (with colour; *generalist* and *subfield*).

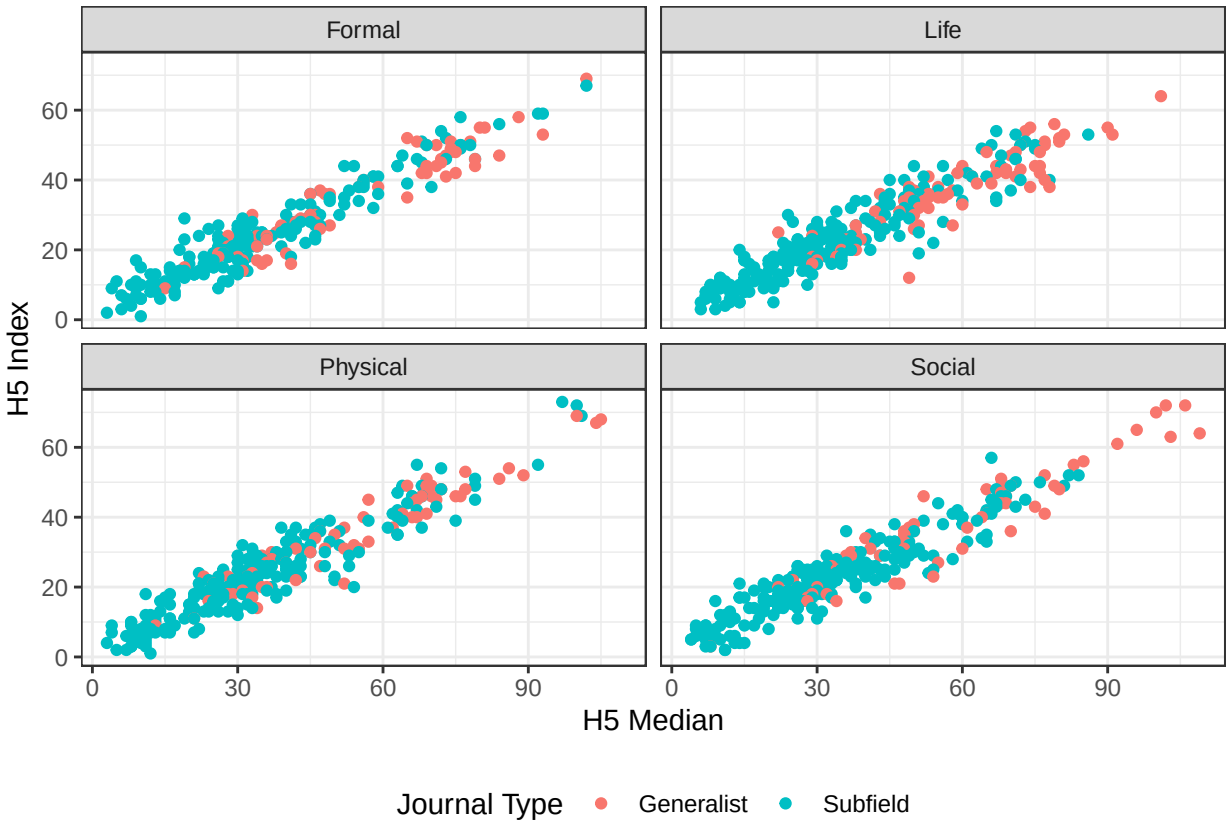

Figure 1: Google Scholar metrics.

Based on this figure, we can make infer that:

- life is the branch with the most journals, and
- generalist journals have higher rankings than specialist journals in all four branches.

References

Blair, G., Cooper, J., Coppock, A., Humphreys, M., Rudkin, A. and Fultz, N. (2019). *fabricatr: Imagine your data before you collect it*. R package, version 0.10.0.