

<http://doi.org/10.1126/science.aam6116>

Article type: LETTER

**Title: Academia needs data science**

**Authors:** Francisco Rodríguez-Sánchez<sup>1,\*</sup>, Ben Marwick<sup>2,3</sup>, Ed Lazowska<sup>2</sup>, Jake VanderPlas<sup>2</sup>

**Affiliations:**

<sup>1</sup> Integrative Ecology Group, Estación Biológica de Doñana (CSIC), E-41092 Sevilla, Spain.

<sup>2</sup> UW eScience Institute, University of Washington, USA.

<sup>3</sup> Department of Anthropology, University of Washington, USA.

\*Correspondence to: frodriguez.work@gmail.com.

**Main text:**

We applaud the recommendations of Stodden *et al.* (1) for improving computational reproducibility in science. We believe many of their recommendations can be fulfilled if researchers embrace modern statistical and computational ("data science") practices. But some changes are necessary to accelerate this process.

Academia has proven slow to adapt best practices and retain many brilliant quantitative scientists. This is bad for science, as prevalent data analyses are often suboptimal, data-driven discoveries remain underdeveloped, and highly-skilled people, who could use their unique expertise to move forward the most pressing scientific questions, are lost to industry.

We propose three reasons why this is happening. First, there is a problem of incentives. Good programming and statistical competence are still not valued enough in most scientific disciplines. In the current 'publish or perish' culture, many scientists are reluctant to invest in learning data science skills that could lead to more powerful and robust research because of the potential impact on narrow measures of productivity (number of papers). To remedy this, we ask our peers to value reproducible research and software as first class research products.

Second, proper training is rare. Despite the online courses and data science programs in some leading universities, most academic centers worldwide do not provide good enough (or even any) training on data science skills. These should be considered part of the students' core training.

Finally, many highly-skilled data scientists leave academia to work on business or industry where they may get better job stability or working conditions. Academia must find career pathways to support these researchers.

In all, this results in a slow uptake of best data science practices and a 'brain drain' out of academia. Some research institutions are already taking measures to ensure this does not slow scientific progress. Others should take note, lest they be left behind.

## References

1. V. Stodden *et al.*, *Science* **354**, 1240 (2016).