

32.6: Doing Reproducible Data Analysis

So far we have focused on the ability to replicate other researchers' findings in new experiments, but another important aspect of reproducibility is to be able to reproduce someone's analyses on their own data, which we refer to a *computational reproducibility*. This requires that researchers share both their data and their analysis code, so that other researchers can both try to reproduce the result as well as potentially test different analysis methods on the same data. There is an increasing move in psychology towards open sharing of code and data; for example, the journal *Psychological Science* now provides "badges" to papers that share research materials, data, and code, as well as for pre-registration.

The ability to reproduce analyses is one reason that we strongly advocate for the use of scripted analyses (such as those using R) rather than using a "point-and-click" software package. It's also a reason that we advocate the use of free and open-source software (like R) as opposed to commercial software packages, which will require others to buy the software in order to reproduce any analyses.

There are many ways to share both code and data. A common way to share code is via web sites that support *version control* for software, such as [Github](#). Small datasets can also be shared via these same sites; larger datasets can be shared through data sharing portals such as [Zenodo](#), or through specialized portals for specific types of data (such as [OpenNeuro](#) for neuroimaging data).

This page titled [32.6: Doing Reproducible Data Analysis](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.