

Why you should transition from Excel to R

R is excellent for those who already are used to working with data in spreadsheets.

Over time, the open-source statistical programming language has consistently grown in popularity among those who work with numbers, with thousands of user-created libraries to expand on its power.

Though it was first created primarily to make it easier to create statistical models and output very basic visuals to explore data, it's expanded to the point that people can use R to do so many advanced processes such as scrape websites, communicate with APIs, and publish beautiful interactive charts and maps.

All with just a few lines of code.

The practice of Reproducible Research has been spreading outside the world of academia to other areas, like non-profits and journalism. It's the idea that analyses should be published with the original data, as well as the methodology or software code so that others can verify or build on them.

Others can reproduce your practice, but the big draw is that for other projects, so can you.

In Excel, a user might run a formula, do some sorting, create a couple of pivot tables— but the work you did on that data cannot be replicated quickly on another spreadsheet with a similar data structure. The whole process has to be repeated step by step.

The point of doing data analysis in R is that a user can write a script to slice up and analyze a spreadsheet, then it can be saved, and then brought back to be used on another spreadsheet with just a few tweaks in the code.