

Why You Should Become a UseR: A Brief Introduction to R

February 28, 2017

If you are reading this article, you've likely heard of the programming language R but may be avoiding it. R is a programming language for statistical computing and graphics that you can use to clean, analyze, and graph your data. It is widely used by researchers from diverse disciplines to estimate and display results and by teachers of statistics and research methods. It's free, making it an attractive option, but does rely on programming code — instead of drop down menus or buttons — to get the job done. Programming languages can be intimidating. Maybe you like the comfort and familiarity of whatever statistics program you've been working with. Maybe you don't have the time to learn a new skill. Maybe you just don't know where to start. These are all valid reasons for putting off using R. But we use R for research and teaching, and we believe that the benefits far outweigh the time and effort needed to start. We are here not only to convince you to use R, but to provide you with some resources to do so.

Reasons to become a useR

One of the most powerful characteristics of R is that it is open-source, meaning anyone can access the underlying code used to run the program and add their own code for free. This means that R:

1. will always be able to perform the newest statistical analyses as soon as anyone thinks of them;
2. will fix its bugs quickly and transparently; and
3. has brought together a community of programming and stats nerds (a.k.a., useRs) that you can turn to for help.

Anyone can write their own R code, which means anyone can add to the huge list of R's tools. Programmers submit their code to R in the form of "packages." Some packages specialize in specific kinds of analyses, while other packages are much broader. For example, the "pwr" package by Stephane Champely specializes in conducting power analyses. In contrast, the "psych" package by APS Fellow William R. Revelle can do anything from descriptive statistics to item-response theory to mediation analyses. At the start of 2017, there are just under 10,000 packages available. And as soon as a new statistical approach is developed, someone will create a new package or add new tools to an existing package.

Moreover, anyone can look at the code used in a package. And there are lots of useRs who know what they're doing and can recognize programming bugs when they appear. Package authors will tell you that their email inboxes are flooded with emails from fellow R aficionados who have come across errors in their code. This means that bugs are found quickly and fixed quickly. As a useR, you don't need to wait a year for a new version of a package to be released; new updates are available as soon as authors makes changes to their packages. And those updates are published, making the entire process transparent.

This dynamic between typical useRs who want to examine data and package authors who want to make

new techniques available is incredibly collaborative — so much so that R users find themselves entering a community of researchers and programmers. For some, this interaction is limited to asking for help (it's often as simple as Googling your question). For those who believe their soul mate is another useR (there are many of us), there are meetup groups around the country and entire conferences organized around R.

Now the question remains: What should you use R for? Everything. No seriously, *everything*. Toss out SPSS, SAS, and STATA, because R can do all the descriptive analyses, regression equations, (M)AN(C)OVA, and hierarchical linear modeling you want. No need to buy MPlus, because R has structural equation modeling covered. Don't bother opening Excel, because merging data sets, cleaning data, identifying important rows or columns, and even updating your gradebook can be done in R. Save money on colored pencils, because R will create whatever plot or graphic you can imagine, even if it's 3D or interactive or both. R can be used with text processors like LaTeX, so you can integrate your results right into the manuscript itself. Stuck using Microsoft Word because your collaborators like track changes? R will create APA formatted tables, complete with significance stars and horizontal lines and export them as .doc files for your convenience. R can do both frequentist *and* Bayesian statistics. R can make use of your multi-core processor and run analyses in parallel. Search for a “bit of fun with R” and learn how to make a winking elephant. R can bootstrap, simulate, randomize, resample, multiply, impute, and park your car. Well, R can't park your car — yet.

On a global scale, R can address many of the challenges of performing reproducible research. A particular study may fail to replicate for a variety of reasons, but one of the simplest being that we often forget exactly what we did to our data to get our results. How did you create scores from your items — averaging, summing, reverse-scoring, or item-response theory? Did you center variable two? Which participants did you exclude and based on what criteria? We often come back to our own data and say, “Wait, what did I do here?” R can remedy these issues because you're using scripting to perform your analyses. Scripting means you write code, which is later run to manipulate data, perform analyses, and make graphics. In other words, using R involves writing a document that contains everything you did, in the order you did it, as you analyze your data. Theoretically, you can share your code and data with literally anyone in the world, and they can use that code and that data to reproduce your results, statistics, and plots with no extra work or thought on their part. This ability to share your analyses has been augmented by online databases like the Open Science Framework in which you can publicly share your analysis scripts and data from your research projects.

A final reason you should become a useR is that R is increasingly being used as an industry standard in the realm of data analytics, also known as “data science.” Many companies (e.g., Facebook, Merck, Pfizer) that hire psychology PhD students recruit candidates who have a solid grasp of both statistics and programming. Learning R will make you a more attractive candidate if you apply for nonacademic jobs, and teaching R will provide your students with more career options.

How to actually become a useR

By now you may be thinking, “R sounds great, but I have absolutely no programming experience. How do I even get started with R?” Never fear! Here, we provide some concrete launching points to start your trajectory toward becoming an expert useR:

Install R and RStudio. The first step of become a userR is installing the correct software onto your computer. In the old days (technically before 2012), the learning curve for R was incredibly steep because the only graphical window that you could interface with was a large empty white console — the kind of blank slate that fills any psychologist’s heart with trepidation. Some really great engineers decided that this was terribly inefficient and developed a graphical user interface (GUI) called RStudio. This made R more user-friendly for individuals without a programming background. We strongly recommend that you install RStudio in addition to R, as it will make your life exponentially easier.

Learn the basics. There are some great tutorials that are freely available online and are great introductory tools for getting you started on your journey to R mastery. We have searched far and wide (across the Internet), and have identified a handful of useful resources, such as “Learning Statistics with R” by Dan Navarro and “YaRrr: A Pirate’s Guide to R” by Nathaniel D. Phillips (for full story, see p. 22). You can even learn R with accompanying cat GIFs. All of these tutorials appear in our extensive list of R Resources, [available online](#).

Explore the advanced techniques. At this point, your uses for R will depend on your research program and your own teaching needs. In our resources list, we’ve pointed to some packages that we like to use on a regular basis, and we have included some packages that are useful for advanced statistical and graphing techniques. Start exploring those, dipping into topics and tools that sound interesting to you. After a while, you’ll come across new packages on your own. Keep an eye on the R-Bloggers.com website to stay on top of new trends (e.g., the new fivethirtyeight package released by Andrew Flowers, the quantitative editor of FiveThirtyEight.com). The more you use R, the more you will get out of it. Furthermore, if you become savvy enough with the R language yourself, you can even write your own functions and packages and distribute them to the public for general use.

We hope that this brief introduction has provided you with the tools and momentum to get started using R for your analyses. R is an incredibly flexible and complex research tool, but once you have mastered it, you can do (almost) anything.

APS Fellow William R. Revelle will speak at the 2017 APS Annual Convention, May 25–28, 2017, in Boston, Massachusetts.