

# What You Should Know About R

*A primer on the industry's open-source statistical analysis language*



CHRIS CHAPMAN

[cchapman@google.com](mailto:cchapman@google.com)

**N**early every time I'm in a group of researchers, I speak with someone who wants to learn R, the industry-wide, free statistical analysis language. If you're one of them, here's what you need to know.

## *Why R Gets So Much Attention*

R is the single most important platform for the development of new statistical methods and the practice of data science. Whether you are looking at the latest machine-learning methods, new statistical algorithms, or new data visualization techniques, the best bet is that they are implemented in R. Today, R is a common language for data scientists.

There are several reasons for this. As I describe below, it is easy to extend R to do any data analysis or statistics task you can dream up. It costs nothing, so universities, students and startups encounter few barriers to adopting it. Also, there are more than 6,000 add-on packages to R—nearly all available for free—that extend its utility. These packages make available the latest methods in statistics and provide bases for further development.

## *Think Programming, Not Statistics*

R is a programming language designed to work with data and implement statistical algorithms. An obvious implication is that if you don't enjoy programming, you won't enjoy working with R. R does not emphasize menus and data grids like an office software suite; instead, it uses a command line that interfaces with developer tools such as an integrated development environment.

This means that learning R requires significant, ongoing investment. No language—human or computer—can be learned by taking a short workshop or memorizing a few words and language structures. Instead, to become fluent, you have to practice the language often and use it for important tasks.

An implication of this is that R becomes progressively more valuable as you develop skill. With traditional statistical software, you can do no more than the developers have implemented and can only interact with the menus and interfaces that they have provided. In R, as you learn more, you'll be able to implement your own procedures to perform analyses and report results exactly as you need them. Because R is

extensible, there is no predetermined limit on what you can do or how much you can streamline your work.

## *How to Learn R*

You should learn R like any language: through immersion. You must choose a project and force yourself to complete it in R. This will be frustrating, but R is designed to be powerful and flexible, not to minimize frustration.

If you use R only to compute descriptive statistics and perform routine analyses such as t-tests, it may remain frustrating and require undue effort. Although R is especially suitable for projects that involve statistical learning algorithms, newly developed techniques, Bayesian analyses, bootstrapping and iterated processes in general, those may be too difficult to tackle when you are just learning it.

What is a good learning project? Look for a task with the following attributes: you understand the statistical methods; you have plenty of time and no urgent deadline; it involves some degree of graphics but ones that are not too complex; and it is a kind of analysis that you do often. These factors will ensure that you'll be able to focus just on R (not on new methods), you'll get something unique (from the graphics), and the learning will be reusable.

Such a project might be one of the following: analysis of a tracking survey; creating a map of sales by store or region; breaking out statistics by known customer segments; or fitting a linear regression model to a well-defined data set. For any of these, the key is to use a real data set that you care about. Keep chipping away at it until your analysis is complete.

When you use R, there is a large online community of enthusiasts and offline community of authors who are willing to assist. When you post

reproducible code snippets online, you'll often find expert assistance quickly.

### Where to Go Next

There are many tutorials online, as well as numerous books that teach R. Users who are migrating from SPSS or SAS will be interested in Robert Muenchen's *R for SAS and SPSS Users*. There are many other texts specific to various methods and domains, including a few for marketing.

Workshops and online courses in R are offered by several universities, and there are several R conferences. The annual User! Conference occurs each summer and is the largest gathering of R users. Another meeting of interest is the Effective

Applications of the R Language (EARL) conference, which occurs in fall 2015 in Boston. Many cities, universities and organizations host R meetups.

Whichever you might choose, remember that a tutorial, book or workshop is not as important as having a problem to solve. Find real data and a problem to tackle, and persist with analysis in R until you've completed the task. Then run it by a few R users to see if they have tips to do things more efficiently. You'll find that R users are a remarkably friendly bunch.

To learn R is to embark on a journey requiring commitment and persistence.

If you're looking to solve problems quickly or to have something showy to demonstrate, R is not the right choice. On the other hand, if you enjoy programming and are interested in building a skill set over time that can be applied with increasing power and effectiveness to many analyses, R is arguably the most powerful tool currently available.

Happy coding, and may your models always converge! **MI**

---

◆ **CHRIS CHAPMAN** is an R enthusiast, a senior researcher at Mountain View, Calif.-based Google, the incoming president of the AMA Marketing Insights Council, and the co-author of *R for Marketing Research and Analytics*.



# Sawtooth Software

The survey software of choice



## Online Surveys | SSI WEB SOFTWARE

Create unlimited online and offline surveys with standard question types, sophisticated piping, and exceptional flexibility to customize your questions.



## Conjoint Analysis | PREDICTING BEHAVIOR

Understand and predict buyer behavior using discrete choice analysis (CBC). Use MaxDiff (best-worst) scaling to assess what's important and segment the market.



## Support | WE'RE THERE FOR YOU

Responsive, expert technical support that consistently exceeds expectations. Let us take care of you!

Our survey and analytical tools are used by the world's most successful companies and consultants.

Give us a call and ask for a tour, or try our software for free today!

+1 801 477 4700

[www.sawtoothsoftware.com](http://www.sawtoothsoftware.com)