REVOLUTION ANALYTICS

White Paper

R Is *Still* Hot–and Getting Hotter

Looking Back on Four Years of Rapid Growth, Increasing Popularity, and Enhanced Functionality

By David Smith

When I wrote a white paper titled "R Is Hot" about four years ago, my goal was to introduce the R programming language to a larger audience of statistical analysts and data scientists. As it turned out, the timing couldn't have been better: R has now blossomed into the language of choice for data scientists worldwide.

Today, R is widely used by scientists, researchers, and statisticians for modeling data and solving problems quickly and effectively. When people ask me which factors are driving the broader adoption of R among data analysts, I usually offer two key points:

1. R was designed specifically for statistical analysis, which means that analytics written in R typically require fewer lines of code (and hence less work) than analytics written in Java, Python, or C++.

2. R is an open source project, which means it is continually improved, upgraded, enhanced, and expanded by a global community of incredibly passionate developers and users.

Let's take a quick look at the many ways in which R has proven its value in real-world scenarios:

- R is used frequently by *The New York Times* to produce consistently great infographics.
- Google uses R to calculate the ROI on advertising campaigns.
- Ford Motor Company uses R to improve the design of its vehicles.
- The Rockefeller Institute of Government uses R to develop models for simulating the finances of public pension funds.
- The Human Rights Data Analysis Group uses R in their efforts to quantify the impact of war.
- The U.S. National Weather Service uses R to predict severe flooding.

Paul Butler, a data scientist and product developer who interned at Facebook, used R to create one of the most spectacular data visualizations in recent memory–the worldwide map of Facebook friends. What's even more amazing, especially to coders who aren't familiar with R, is that Paul produced the iconic image with about 150 lines of R code. He truly demonstrated the power and flexibility of R.

Image credit: Paul Butler, Facebook

In addition to *The New York Times,* Google, and Facebook, many other respected organizations routinely use R for data analysis. The list of users includes Twitter, Kickstarter, eHarmony, Accenture, Deloitte, Coursera, Zillow, Trulia, DataSong, *The Economist,* RealScientists, ANZ, Credit Suisse, Lloyds Bank, Nationwide Insurance, Bank of America, Ford, John Deere, Nordstrom, Uber, and Etsy.

Not bad for an initially esoteric language created by two academics in New Zealand in the early 1990s.

REVOLUTION ANALYTICS

More Than a Programming Language

Unlike legacy analytic software products, R is a fully fledged programming language. But it's more than just a language: R represents a radically different approach to the challenges posed by increasingly larger and more complex sets of data. In that respect, R is something of a cultural phenomenon.

Usage of R has increased steadily since 2007, according to a Rexer Analytics 2013 Data Miner Survey. R is also among the world's top 10 most popular programming languages as ranked by *IEE Spectrum.* On a practical note, R programmers were among the highest paid developers in recent surveys by Dice and O'Reilly Media.

As an open source project, R depends on a worldwide community of active developers to grow and evolve. Like Linux, the most famous open source project, R isn't "owned" by any single person or entity. R is maintained and supported by thousands of individual users who also contribute to its ongoing development.

The members of this global community serve as R's guardians and custodians—and they take their responsibilities seriously. Like doting parents, they take pride in the achievements of their offspring—and they are quick to leap in when they perceive a problem.

R Is Ready for Business

The effort to commercialize R is serious and ongoing. In fact, commercializing R has been a key part of the mission here at Revolution Analytics since our founding. We are honored and privileged to play a role in the R movement, and we understand our responsibility to the greater R community.

Revolution Analytics recently unveiled Revolution R Open, an enhanced downstream distribution of R from the R Foundation for Statistical Computing. Revolution R Open comes with enhancements focused on performance and reproducibility to improve the R experience.

For example, Revolution R Open is linked with multi-threaded math libraries. These replace the standard BLAS/LAPACK libraries to improve the performance of R, especially on multi-core hardware. You don't need to modify your R code to take advantage of the performance improvements.

Additionally, Revolution R Open comes with the Reproducible R Toolkit. The default CRAN repository in each new version is a recent static snapshot of CRAN, so you won't be surprised by package updates that break your code. You can also reliably access newer R packages with the pre-installed checkpoint package. These changes make it easier for you to share R code with other R users, confident they'll get the same results you got when you wrote the code.

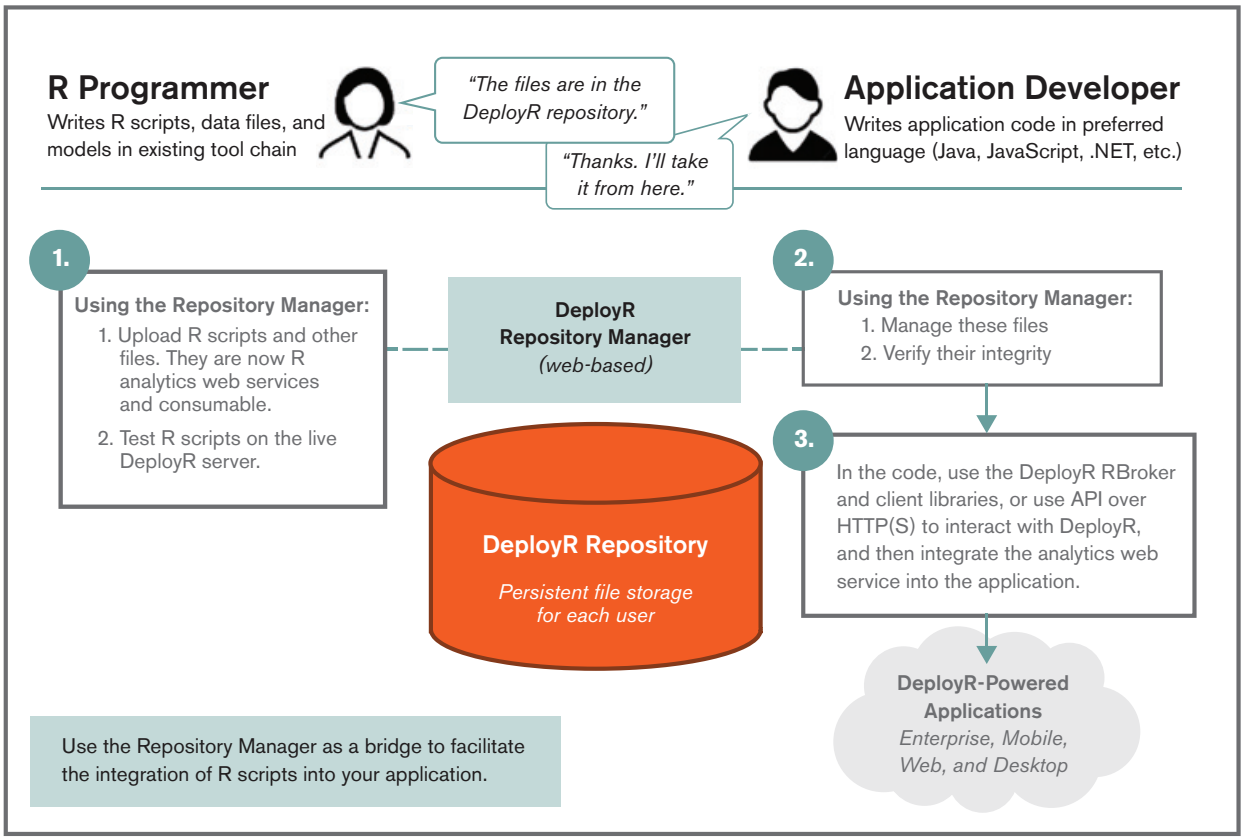
We've also introduced the new Managed R Archive Network website, MRAN, where you can find information about R, Revolution R Open, and R packages. MRAN includes tools for exploring R packages by description or by topic, making it easy to find packages to extend R's capabilities. MRAN is updated daily.

Revolution R Open is available for download now. Visit mran.revolutionanalytics.com/download for binaries for Windows, Mac, Ubuntu, CentOS/Red Hat Linux, and (of course) the GPLv2 source distribution.

Embedding Results of R Functions

If you need to embed the results of R functions—data, charts, or even a single calculation—into other applications, then you might want to take a look at DeployR Open. DeployR Open is an open source, server-based framework for R that makes it easy to call out to the server to run R code in real time.

The workflow is simple: R programmers develop an R script (using their standard R tools) and publish that script to the DeployR server. Once published, R scripts can be executed by any authorized application using the DeployR API. We provide native client libraries in Java, JavaScript, and .NET to simplify making calls to the server. The R results returned on these calls can be embedded, displayed, or processed in any way your application needs.



REVOLUTION ANALYTICS

Here are several useful features of the architecture:

- The application developers don't need to know any R. They use a consistent API in their chosen application programming language to execute any R code, passing parameter values as needed.
- The server, not the application, manages all aspects of the R session lifecycle. This restricts the complexity to the server, so the application developer can stay focused on business logic.
- Only specific R functions are exposed beyond the server "firewall." If you connect to R directly, you expose all of R's functions (including functions that can modify the file system) to the outside world. By contrast, DeployR Open only exposes selected functions, via RScripts, accessible through the API to serve a specific task. With the Repository Manager, you can manage which RScripts are published, as well as who can access them.
- The server enforces deployment-wide and user-specific security policies—set in the Administration Console—at runtime to ensure fair resource allocation and help prevent malicious R code execution.

DeployR Open offers a number of ways for client applications to integrate with the server:

- The simplest way is to use the RBroker Framework. If you want to quickly execute an independent R calculation—or if you are considering building a sophisticated R-based, real-time scoring engine—use this framework.
- If your application needs more control over the full R session lifecycle, such as might be required by a graphical user interface (GUI) to R, then consider using the Java, JavaScript, or .NET client libraries.
- If for any reason the RBroker Framework and client libraries do not meet your needs, you can always integrate directly on the DeployR API (application programming interface).
- We provide examples of R integration using DeployR within Excel, Jaspersoft, and QlikView.

DeployR Open is 100% open source and includes many features previously available only as part of Revolution R Enterprise DeployR. If you'd like technical support for DeployR Open, phone and email support is included with a Revolution R Plus subscription.

The DeployR Open server is deployed as a single node and, as such, mainly designed for prototyping, building, and deploying applications when the expected load on the server is low or moderate. If you anticipate a need to scale to multiple server resources to handle increased workload or improved throughput—or if you want to enjoy seamless integration with popular enterprise security solutions such as single sign-on (SSO), Lightweight Directory Access Protocol (LDAP), Active Directory, or Pluggable Authentication Modules (PAM)—consider upgrading to Revolution R Enterprise DeployR.

If you'd like to learn more, check out the DeployR Open website. There, you'll find developer documentation and administrator documentation, and you can download and install DeployR Open for free.

If You're Not Using R Already, You Probably Will Be Soon

With thousands of contributors and more than 2 million users worldwide, R is a truly global phenomenon. Unlike traditional commercial software for data analysis, R is both flexible and extensible. Supported by an active community of users and developers, R is also constantly changing to meet the needs of a rapidly shifting global economy.

The popularity of R is no fluke or fad. R has become the common language of data analysis because it was designed–from the ground up–as a practical system for handling the real-world challenges of complex data sets. R-based programs have many uses and are routinely applied to solve problems in finance, real-time trading, risk assessment, forecasting, biotechnology, drug development, and data visualization.

The wide acceptance of R as the lingua franca of statistics is based on the language's unique ability to change, transform, and evolve. When new techniques in statistical analysis are discovered, they tend to first emerge as R packages–years before those innovations are incorporated into traditional enterprise software products.

Thanks to its open source roots, R has undergone a viral spread across the map. It has become both ubiquitous and indispensable. The R community supports development, innovation, and continuous improvement. New players are welcome and encouraged. The R ecosystem has become a fertile breeding ground for novel ideas and original ways of thinking about numbers.

No one can foretell the future of quantitative analytics, but it's a safe wager that a good deal of it will be written in R. Whether the job involves production or analysis, R remains a highly logical choice.

About David Smith

Chief Community Officer David Smith leads the open source solutions group at Revolution Analytics. He draws on his data science background daily to write about applications of R and predictive analytics for the Revolutions blog, and he was named a top-10 influencer on the topic of big data by *Forbes.*

He co-authored (with Bill Venables) the tutorial manual *An Introduction to R* and was also one of the originating developers of the ESS (Emacs Speaks Statistics) project. Prior to joining Revolution Analytics, David was the director of product management for S-PLUS at Insightful Corporation.

David studied statistics at the University of Adelaide in South Australia, as well as at Lancaster University in the UK. He lives in Chicago with his husband, Jay. Follow David on Twitter: @revodavid.

About Revolution Analytics

Since 2007, Revolution Analytics has worked to foster growth and learning within the R community, as well as actively supported the growing ranks of commercial users and their needs.

Our flagship suite of Revolution R software and services is the innovation leader in big data analytics for the enterprise. Powered by the R language, Revolution R products are used by enterprises dealing with massive data, performance, and multi-platform requirements, along with the need to drive down big data costs.

The suite includes Revolution R Open, a free and enhanced open source R distribution. Revolution R Plus adds a package for subscription-based technical support and indemnification. Revolution R Enterprise delivers parallelized analytics to the heart of big data platforms, including data warehouses and Hadoop. Revolution R Cloud is available on demand through AWS Marketplace.

We partner with our customers to develop the skills they need today, and we help them design the programs that fulfill the promise of tomorrow. Our AdviseR consulting services and technical support deliver the expertise necessary for quick and effective installation, development, and deployment of advanced, R-based analytics within even the most complex and diverse data architectures. Revolution Analytics AcademyR training services help your people to help your program–from updating your experts on legacy platforms to solidifying the skillsets of your newest employees.

Please visit us at revolutionanalytics.com.