CASE STUDY: PFIZER

"The goal of the S language … is 'to turn ideas into software, quickly and faithfully' … it is the duty of the responsible data analysts to engage in this process … the exercise of drafting an algorithm to the level of precision that programming requires can in itself clarify ideas and promote rigorous intellectual scrutiny." – Venables and Ripley (2000).

Background

R is an implementation of the S language which, "forever altered how people analyze visualize, and manipulate data" (excerpt from the citation accompanying the Association for Computing Machinery Award for Software Systems, awarded to John Chambers in 1998 for his development of the language). Pfizer current uses of R include the following:

1. Analysis of genetic data, including whole-genome SNP profiling of 100K+ SNPs (Molecular Profiling).

2. Microarray pre-processing (normalization, summarization, scaling) and analysis of gene expression data, analysis of RNAi/cellomic assays, analysis of rtPCR gene expression data, predictive models for method of action (Discovery and Safety Sciences)

3. Predictive modeling relating structure to outcomes such as activity and ADMET properties (Discovery Chemistry and Pharmaceutical Sciences).

4. Ad hoc visualization, exploratory data analyses, curve fitting, and mixed model analyses (Discovery Biology).

5. Monte Carlo simulation.

Moreover, Pfizer is **dependent** on R as a component of the following software systems:

1. Microarray data analysis system.
2. System for deploying statistical calculations to the web.
3. Database for Pharmacogenomics genetic information.
4. Analysis workflow tool.
5. In-silico model generator.
6. Biology analysis and data management system

REvolution computing
We do the math

CASE STUDY PFIZER cont.

The use of R at Pfizer now seems to have outpaced their ability to support that use. As R is provided freely, it comes with no installation support, no programming support, no guarantee that adequate testing has been applied to new releases, and so on. At times, significant Pfizer resources have been consumed by the installation and configuration process, both on servers and PCs. More than with commercially supported applications, it is often difficult to assess whether errors are the fault of the programmer, the installation technician, or the language itself. Also, programming support has informally become the responsibility of (perceived) in-house experts, who also have day jobs.

Value added by RPro

RPro is a value-added, commercially supported version of the R language offered by REvolution Computing. Advantages of RPro over public domain R include:

1. Contract assurance that industry best practices are being followed with respect to the software development and security. In addition, documentation of the processes used is made available to Pfizer at a very detailed level (e.g. including specific descriptions of unit tests performed).

2. Contract assurance that releases will be delivered according to an agreed upon schedule.

3. Pre-compiled binaries and installers for Windows, Linux, and Solaris (32 and 64 bit).

4. Bug fixes for current and previous releases. These would be “hot patches”, i.e. they would not require a complete reinstall as do the publicly available bug fixes.

5. Phone and email support during normal business hours. This includes installation and configuration support as well as programming support.

Summary

“De facto, R is already a significant component of Pfizer core technology. Access to a supported version of R will allow us to keep pace with the growing use of R in the organization, and provides a path forward to use of R in regulated applications.”

– James A. Rogers Ph.D. is an Associate Director in the Nonclinical Statistics group for Pfizer Global Research and Development

REvolution computing
We do the math