R, Quo Vadis?

Bill Venables[1,⋆]

1. CSIRO Division of Mathematics, Informatics & Statistics, Australia
⋆Contact author: Bill.Venables@CSIRO.au

Keywords: Programming, Data Analysis, Modelling, R

Language designers seem to regard *R* as an ugly, inefficient language and hence find its popularity mystifying. See, for example, [3, 4].

I will present four examples from my own work, two large data analyses problems in fisheries, and two more abstract programming examples. Hopefully these will be of intrinsic interest, but together they encapsulate why I think useRs find *R* so invaluable. My thesis is that good data analysis and modelling require the practitioner to engage *interactively* with data, and that at some level *programming* becomes essential to this. This is essentialy the same message as that presented in Chambers [1, 2], and the same idea implicitly underlies [5]. The popularity of *R* is primarily due to the way it provides support for this activity, making near optimal trade-offs. This view is mostly consistent with Cook [3] but there are some important differences. (The claim that *R* is *necessarily* ugly is also disputed!)

Although *R* may be well suited to meet many contemporary data analysis problems, it will not remain so indefinitely. I do not attempt to answer the existential question posed in the title, but rather suggest it as one we should be thinking about, now. I will present some thoughts on a SWOT assessment for *R*, and suggest ways we might prepare for a graceful transition to whatever becomes the next phase. Such a new phase, or phases, will inevitably come as data analysis itself rapidly evolves in both scope and scale.

References

[1] Chambers, J. M. (1998). *Programming with Data: A Guide to the S Language*. New York: Springer-Verlag.

[2] Chambers, J. M. (2008). *Software for Data Analysis: Programming with R*. New York: Springer-Verlag.

[3] Cook, J. D. (2012). Why and how people use *R*. http://channel9.msdn.com/Events/Lang-NEXT/Lang-NEXT-2012/Why-and-How-People-Use-R. See also http://lambda-the-ultimate.org/node/4503 and http://lambda-the-ultimate.org/node/4507.

[4] Morandat, F., B. Hill, L. Osvald, and J. Vitek (2012). Evaluating the design of the *R* language: Objects and functions for data analysis. http://www.cs.purdue.edu/homes/jv/pubs/ecoop12.pdf. An ECOOP 2012 paper.

[5] Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer-Verlag.