R Version 2.0.0

Kurt Hornik¹ and Friedrich Leisch²
(on behalf of the R Development Core Team)

¹ Institut für Statistik und Mathematik, Wirtschaftsuniversität
Wien, Augasse 2-6, 1090 Wien, Austria
² Institut für Statistik und Wahrscheinlichkeitstheorie, Technische
Universität Wien, Wiedner Hauptstrasse 8–10/1071, 1040
Wien, Austria

R, also known as "GNU S", is a language and environment for statistical
computing and graphics. R implements a dialect of S, a very high level
language which has been developed at Bell Laboratories for about 30 years.
In 1998, the Association for Computing Machinery (ACM) presented its
Software System Award to John M. Chambers, the principal designer of S,
for "the S system, which has forever altered the way people analyze, visualize,
and manipulate data ...". The R project was started in 1992 in Auckland,
New Zealand, by Ross Ihaka and Robert Gentleman (?). Since 1997, R
is being developed by an international team, the R Core Team, currently
consisting of 16 members and including John Chambers.

An R distribution provides a run-time environment with graphics, a de-
bugger, access to to certain system functions, and the ability to run programs
stored in script files, and contains functionality for a large number of statis-
tical procedures. This "base system" is highly extensible through so-called
packages which can contain R code and corresponding documentation, data
sets, code to be compiled and dynamically loaded, and so on. In fact, the R
distribution itself provides its functionality via "base" packages such as base,
stats, graphics, and methods.

The release of R 1.0.0 on 2000-02-29 indicated that R had reached a level
of stability and maturity making it suitable for production use, and that the base language and the API for extension writers would remain stable for the foreseeable future. In addition, R 1.0.0 provided a reference implementation of S version 3 as documented in the so-called “Blue” and “White” books by John Chambers and coauthors (??). Since then, R has developed steadily, with two “minor” releases per year, in particular adding reference implementations of the new paradigm for object oriented programming (OOP) introduced in S version 4 and described in the so-called “Green Book” (?). The advent of R 2.0.0 (?), the “next generation of R”, mostly indicates the common view of the R developers that R has moved substantially beyond R 1.0.0 and S 4.

Key innovations in R 2.0.0 include the new OOP paradigm, generalized input/output via connection objects (?), much improved garbage collection (?), exception handling constructs, and significant performance improvements via “lazy loading” of R code in packages (?). Name spaces allow package authors to control how global variables in their code are resolved (?). The package management facilities provide efficient bundling, quality control, distribution, and installation of extensions. Sweave allows to embed the R code for complete data analyses in \LaTeX{} documents (?).

These enhancements greatly increase the potential and power of R for developing and deploying large-scale data analysis solutions, such as the Bioconductor environment (?). The graphics facilities have been substantially enhanced via support for mathematical (\LaTeX{} style) annotation for plots (?) and the new “grid” graphics engine (?), which offers increased flexibility compared to the traditional, but still viable, ink-on-paper model. The R Windows GUI has been further refined, and there is now a Mac OS X port using the Aqua interface (?). A cross-platform method for developing GUI interfaces is provided via Tcl/Tk (?), work on alternate interfaces and use of other interface toolkits is ongoing and in various stages of development. Some of these key innovations are described in more detail in (?).

The packages in the R distribution were substantially reorganized, with a much smaller base providing the part of the R language implemented in R itself, and stats collecting all core statistical functionality. In addition, several recommended packages, maintained outside the R Core Team but deemed indispensable in a comprehensive environment for statistical computing and graphics, are now bundled with R. The data analytic techniques described in such popular books as ?, ?, or ? have corresponding recommended R packages (MASS, nlme, and survival). In addition, the recommended packages include foreign with import/export for other statistical systems, and lattice (?), which reimplements Bill Cleveland’s Trellis graphics and extends it to structured multiframe layouts, building on the new grid graphics engine.

R has increased dramatically in popularity in recent years, in teaching and academic research, notably in bioinformatics, and also in industry and finance. Currently, more than 400 packages are available via the Comprehen-