

- The signature for the predefined S4 method for `as.character()` was missing
- `readBin(raw_vector)` could read beyond the end of the vector when size-changing was involved.
- The C entry point `PrintValue` (designed to emulate auto-printing) would not find `show()` for use on S4 objects, and did not have the same search path (for `show()`, `print()` and `print()`

methods) as auto-printing. Also, auto-printing and `print()` of S4 objects would fail to find `show()` if the methods namespace was loaded but the package was not attached (or otherwise not in the search path).

- `print()` (and auto-printing) now recognize S4 objects even when `methods` is not loaded, and print a short summary rather than dump the internal structure.

Changes on CRAN

by Kurt Hornik

New contributed packages

ARES Allelic richness estimation, with extrapolation beyond the sample size. Generates an allelic richness accumulation curve. This curve shows the expected number of unique alleles in a population when taking a sample of individuals. The function `aresCalc` takes a binary data matrix as input, showing the presence of alleles per individual, and gives an accumulation curve (mean with 95% confidence bounds) back. The function `aresPlot` can be used to plot the output from `aresCalc`. By Emiel van Loon and Scott Davis.

BayHaz Bayesian Hazard Rate Estimation: a suite of R functions for Bayesian estimation of smooth hazard rates via Compound Poisson Process (CPP) priors. By Luca La Rocca.

BiasedUrn Biased Urn model distributions. Statistical models of biased sampling in the form of univariate and multivariate non-central hypergeometric distributions, including those of Wallenius and Fisher (also called extended hypergeometric distribution). By Agner Fog.

BootCL Bootstrapping test for chromosomal localization. By Eun-Kyung Lee, Samsun Sung, and Heeal Kim.

Broddingnag Very large numbers in R. Real numbers are held using their natural logarithms, plus a logical flag indicating sign. The package includes a vignette that gives a step-by-step introduction to using S4 methods. By Robin K. S. Hankin.

CCA Canonical correlation analysis. Provides a set of functions that extend the `cancor` function with new numerical and graphical outputs. It also include a regularized extension of the canonical correlation analysis to deal with data

sets with more variables than observations. By Ignacio González and Sébastien Déjean.

CreditMetrics Functions for calculating the CreditMetrics risk model. By Andreas Wittmann.

DAAGxtras Data sets and functions additional to **DAAG**, used in additional exercises for the book "Data Analysis and Graphics Using R" by J. H. Maindonald and W. J. Brain (2007, 2nd edn.), and for laboratory exercises prepared for a 'Data Mining' course. By John Maindonald.

GenABEL genome-wide association analysis between quantitative or binary traits and SNPs. By Yurii Aulchenko.

GeoXp Interactive exploratory spatial data analysis. A tool for researchers in spatial statistics, spatial econometrics, geography, ecology etc., allowing to link dynamically statistical plots with elementary maps. By Christine Thomas-Agnan, Yves Aragon, Anne Ruiz-Gazen, Thibault Laurent, and Laurianne Robidou.

HydroMe Estimation of the parameters in infiltration and water retention models by curve-fitting method. The models considered are those that are commonly used in soil science. By Christian Thine Omuto.

IPSUR Data sets and functions accompanying the book "Introduction to Probability and Statistics Using R" by G. Andy Chang and G. Jay Kerns (in progress). By G. Jay Kerns with contributions by Theophilus Boye, adapted from the work of John Fox et al.

InfNet Simulation of epidemics in a network of contacts. The simulations consider SIR epidemics with events in continuous time (exponential inter-event times). It can consider a structure of local networks and has an option to visualize it with the animator called SoNIA (<http://www.stanford.edu/group/sonia/>). By Lilia Ramirez Ramirez and Mary Thompson.

- JointModeling** Joint modeling of mean and dispersion through two interlinked GLMs or GAMs. By Mathieu Ribatet and Bertrand Iooss.
- NMMAPSLite** NMMAPS Data Lite. Provides remote access to daily mortality, weather, and air pollution data from the National Morbidity, Mortality, and Air Pollution Study for 108 U.S. cities (1987–2000); data are obtained from the Internet-based Health and Air Pollution Surveillance System (iHAPSS). By Roger D. Peng.
- NRAIA** Data sets from the book “Nonlinear Regression Analysis and Its Applications” by D. M. Bates and D. G. Watts (1998), John Wiley and Sons. By Douglas Bates.
- PKtools** Unified computational interfaces for pop PK. By M. Suzette Blanchard.
- PhySim** Phylogenetic tree simulation based on a virth death model. Functions are provided to model a lag-time to speciation and extract sister species ages. By Jason T. Weir, Dolph Schluter.
- PresenceAbsence** Presence-absence model evaluation. Includes functions for calculating threshold dependent measures such as confusion matrices, pcc, sensitivity, specificity, and Kappa, and produces plots of each measure as the threshold is varied. Can also optimize threshold choice, and plot the threshold independent ROC curves along with the associated AUC By Elizabeth Freeman.
- RJDBC** Provides access to databases through the JDBC interface, implementing R’s DBI interface using JDBC as a back-end. This allows R to connect to any DBMS that has a JDBC driver. By Simon Urbanek.
- RLadyBug** Analysis of small scale infectious disease data using stochastic Susceptible-Exposed-Infectious-Recovered (SEIR) models. By Michael Hoehle and Ulrike Feldmann.
- RaschSampler** A Rasch sampler for sampling binary matrices with fixed margins. By Reinhold Hatzinger, Patrick Mair, and Norman Verhelst.
- Rcmdr.HH** Rcmdr support for the introductory course at Temple University, which spends time on several topics that are not yet in the R Commander. By Richard M. Heiberger, with contributions from Burt Holland.
- Rserve** A binary R server which acts as a socket server (TCP/IP or local sockets) allowing binary requests to be sent to R. Every connection has a separate workspace and working directory. Client-side implementations are available for popular languages such as C/C++ and Java, allowing any application to use facilities of R without the need of linking to R code. Supports remote connection, user authentication and file transfer. A simple R client is included as well. By Simon Urbanek.
- SLmisc** Miscellaneous functions for analysis of gene expression data at SIRS-Lab GmbH. By Matthias Kohl.
- SMPracticals** Data sets and a few functions for use with the practicals outlined in Appendix A of the book “Statistical Models” by A. Davison (2003), Cambridge University Press. The practicals themselves can be found at <http://statwww.epfl.ch/davison/SM/>. By Anthony Davison.
- SNPmaxsel** Asymptotic methods related to maximally selected statistics, with applications to SNP data. By Anne-Laure Boulesteix.
- Snowball** Snowball stemmers. By Kurt Hornik.
- SpherWave** Spherical wavelets and SW-based spatially adaptive methods. Carries out spherical wavelet transform developed Li (1999) and Oh (1999), and implements wavelet thresholding approaches proposed by Oh and Li (2004). By Hee-Seok Oh and Donghoh Kim.
- Synth** Causal inference using the synthetic control group method of Abadie and Gardeazabal (2003) and Abadie, Diamond, Hainmueller (2007). By Alexis Diamond and Jens Hainmueller.
- TRAMPR** Matching of terminal restriction fragment length polymorphism (TRFLP) profiles between unknown samples and a database of knowns. Facilitates the analysis of many unknown profiles at once, and provides tools for working directly with electrophoresis output through to generating summaries suitable for community analyses with R’s rich set of statistical functions. Also resolves the issues of multiple TRFLP profiles within a species, and shared TRFLP profiles across species. By Rich FitzJohn and Ian Dickie.
- VGAM** Vector generalized linear and additive models, and associated models (Reduced-Rank VGLMs, Quadratic RR-VGLMs, Reduced-Rank VGAMs). Fits many models and distribution by maximum likelihood estimation (MLE) or penalized MLE. Also fits constrained ordination models in ecology. By Thomas W. Yee.
- WWGbook** Functions and data sets for the book “Linear Mixed Models: A Practical Guide Using Statistical Software” by B. T. West, K. B. Welch, and A. T. Galecki (2006), Chapman

- Hall / CRC Press. By Shu Chen and Andrzej Galecki.
- WeedMap** Map of weed intensity. Compute spatial predictions from exact count and a covariate. By Gilles Guillot.
- ZIGP** Adapt and analyze Zero Inflated Generalized Poisson (ZIGP) regression models. By Vinzenz Erhardt.
- adimpro** Adaptive smoothing of digital images via the Propagation Separation approach by Polzehl and Spokoiny (2006). By Karsten Tabe-low and Joerg Polzehl.
- agricolae** Statistical Procedures for Agricultural Research. These functions are currently utilized by the International Potato Center Research (CIP), the Statistics and Informatics Instructors and the Students of the Universidad Nacional Agraria La Molina Peru, and the Specialized Master in "Bosques y Gestión de Recursos Forestales" (Forest Resource Management). Contains functionality for the statistical analysis of experimental designs applied specially for field experiments in agriculture and plant breeding. By Felipe de Mendiburu.
- analogue** Analogue methods for palaeoecology. Fits modern analogue technique transfer function models for prediction of environmental data from species data. Also performs analogue matching, a related technique used in palaeoecological restoration. By Gavin L. Simpson.
- arm** Data analysis using regression and multi-level/hierarchical models. By Andrew Gelman, Yu-Sung Su, Jennifer Hill, Maria Grazia Pittau, Jouni Kerman, and Tian Zheng.
- arrayMissPattern** Exploratory analysis of missing patterns for microarray data. By Eun-kyung Lee and Taesung Park.
- bbmle** Modifications and extensions of the **stats4** **mle** code. By Ben Bolker, modifying code by Peter Dalgaard and other R-core members.
- bcp** Bayesian Change Point. An implementation of the Barry and Hartigan (1993) product partition model for the standard change point problem using Markov Chain Monte Carlo. By Chandra Erdman.
- binGroup** Evaluation and experimental design for binomial group testing. By Frank Schaarschmidt.
- cgh** Analysis of microarray comparative genome hybridization data using the Smith-Waterman algorithm. By Tom Price.
- classify** Exploration of classification models in high dimensions, implementing methods for understanding the division of space between groups. See <http://had.co.nz/classify> for more details. By Hadley Wickham.
- clustTool** GUI for clustering data with spatial information. By Matthias Templ.
- clusterGeneration** Random cluster generation (with specified degree of separation). By Weiliang Qiu and Harry Joe.
- clusterSim** Searching for optimal clustering procedure for a data set. By Marek Walesiak Andrzej Dudek.
- cobs99** Constrained B-splines – outdated 1999 version. By Pin T. Ng and Xuming He; R port by Martin Maechler.
- corrgram** Plot correlograms. By Kevin Wright.
- distrDoc** Documentation for packages **distr**, **distrEx**, **distrSim**, and **distrTEst**. By Florian Camphausen, Matthias Kohl, Peter Ruckdeschel, and Thomas Stabla.
- drm** Likelihood-based marginal regression and association modeling for repeated, or otherwise clustered, categorical responses using dependence ratio as a measure of the association. By Jukka Jokinen.
- elrm** Exact Logistic Regression via MCMC. By David Zamar, Jinko Graham, and Brad McNeney.
- emdbook** Data sets and auxiliary functions for the book "Ecological Models and Data" by B. Bolker (in progress). By Ben Bolker.
- epicalc** Epidemiological calculator. By Virasakdi Chongsuvivatwong.
- fEcofin** Selected economic and financial data for teaching financial engineering and computational finance. By Diethelm Wuertz and many others.
- fame** Interface for the FAME time series database. Includes FAME storage and retrieval function, as well as functions and S3 classes for time indexes and time indexed series, which are compatible with FAME frequencies. By Jeff Hallman.
- filehashSQLite** Simple key-value database using SQLite as the backend. By Roger D. Peng.
- gamlss.dist** Extra distributions for GAMLSS modeling. By Mikis Stasinopoulos and Bob Rigby with contributions from Calliope Akantziliotou and Raydonal Ospina.

- glm** Fits generalized linear models where the parameters are subject to linear constraints. The model is specified by giving a symbolic description of the linear predictor, a description of the error distribution, and a matrix of constraints on the parameters. By Sanjay Chaudhuri, Mark S. Handcock, and Michael S. Rendall.
- heplots** Visualizing Tests in Multivariate Linear Models. Represents sums-of-squares-and-products matrices for linear hypotheses and for error using ellipses (in two dimensions) and ellipsoids (in three dimensions). By John Fox, Michael Friendly, and Georges Monette.
- homtest** A collection of homogeneity tests and other stuff for hydrological applications. By Alberto Viglione.
- ibdreg** A method to test genetic linkage with covariates by regression methods with response IBD sharing for relative pairs. Accounts for correlations of IBD statistics and covariates for relative pairs within the same pedigree. By Jason P. Sinwell and Daniel J. Schaid.
- kernelPop** Spatially explicit population genetic simulations. Creates an individual-based population genetic simulation in discrete time, where individuals have spatial coordinates, with dispersal governed by mixtures of Weibull and normal pdfs. Simulates diploid and haploid inheritance. Allows the development of null distributions of genotypes for complex demographic scenarios. By Allan Strand and James Niehaus.
- klin** Implements efficient ways to evaluate and solve equations of the form $Ax = b$, where A is a Kronecker product of matrices. Also includes functions to solve least squares problems of this type. By Tamas K Papp.
- knnflex** A more flexible KNN implementation which allows the specification of the distance used to calculate nearest neighbors (euclidean, binary, etc.), the aggregation method used to summarize response (majority class, mean, etc.) and the method of handling ties (all, random selection, etc.). Additionally, continuous responses are handled. By Atina Dunlap Brooks.
- languageR** Data sets and functions for the book "Analyzing Linguistic Data: A practical introduction to Statistics" by R. H. Baayen (2006), Cambridge University Press. By R. H. Baayen.
- lss** Accelerated failure time model to right censored data based on least-squares principle. By Lin Huang and Zhezhen Jin.
- mFilter** Several time series filters useful for smoothing and extracting trend and cyclical components of a time series. Currently, Christiano-Fitzgerald, Baxter-King, Hodrick-Prescott, Butterworth, and trigonometric regression filters are included in the package. By Mehmet Balçilar.
- meboot** Maximum entropy density based dependent data bootstrap for time series. An algorithm is provided to create a population of time series (ensemble) without assuming stationarity. By Hrishikesh D. Vinod and Javier López-de-Lacalle.
- memisc** Miscellaneous tools for data management, including tools for preparing (especially social science) survey data, conducting simulation studies, and presentation of results of statistical analyses. By Martin Elff.
- mixstock** Functions for mixed stock analysis. By Ben Bolker.
- mixtools** Tools for analyzing mixture models. By Ryan Elmore, Tom Hettmansperger, David Hunter, Hoben Thomas, Fengjuan Xuan, and Derek Young.
- mlCopulaSelection** Use numerical maximum likelihood to choose and fit a bivariate copula model (from a library of 40 models) to the data. By Jesus Garcia and Veronica Gonzalez-Lopez.
- msgcop** Semiparametric Bayesian Gaussian copula estimation, treating the univariate marginal distributions as nuisance parameters in Hoff (2006). Also provides a semiparametric imputation procedure for missing multivariate data. By Peter Hoff.
- muS2RC** S-PLUS to R compatibility for package **muStat**. By Knut M. Wittkowski and Tingting Song.
- muStat** Prentice Rank Sum Test and McNemar Test. Performs Wilcox rank sum test, Kruskal rank sum test, Friedman rank sum test and McNemar test. By Knut M. Wittkowski and Tingting Song.
- muUtil** A collection of utility functions for package **muStat**. By Knut M. Wittkowski and Tingting Song.
- np** Nonparametric kernel smoothing methods for mixed data types. By Tristen Hayfield and Jeffrey S. Racine.
- nsRFA** Tools for objective (non-supervised) applications of the Regional Frequency Analysis methods in hydrology. By Alberto Viglione.

- pARtial** (Partial) attributable risk estimates, corresponding variance estimates and confidence intervals. By Andrea Lehnert-Batar.
- pmml** Generate PMML for various models. The Predictive Modeling Markup Language (PMML) is a language for representing models, in an application independent way. Such models can then be loaded into other applications supporting PMML, such as Teradata Warehouse Miner and IBM's DB2. By Graham Williams.
- popbio** Construct and analyze projection matrix models from a demography study of marked individuals classified by age or stage. Covers methods described in the books "Matrix Population Models" by Caswell (2001) and "Quantitative Conservation Biology" by Morris and Doak (2002). By Chris Stubben and Brook Milligan.
- portfolioSim** Framework for simulating equity portfolio strategies. By Jeff Enos and David Kane, with contributions from Kyle Campbell.
- prim** Patient Rule Induction Method (PRIM) for estimating highest density regions in high-dimensional data. By Tarn Duong.
- qgen** Analysis of quantitative genetic data, especially helpful to perform parametric resampling of quantitative genetic data sets. By Thomas Fabbro.
- rcdk** Interface to the CDK libraries, a Java framework for cheminformatics. Allows to load molecules, evaluate fingerprints, calculate molecular descriptors and so on. By Rajarshi Guha.
- rcompgen** Generates potential completions for a given command snippet based on circumstantial evidence. By Deepayan Sarkar.
- rcompletion** Pseudo-intelligent TAB completion for R when run from a terminal, using the GNU Readline library. By Deepayan Sarkar.
- relaxo** Relaxed Lasso, a generalization of the Lasso shrinkage technique for linear regression. Both variable selection and parameter estimation is achieved by regular Lasso, yet both steps do not necessarily use the same penalty parameter. By Nicolai Meinshausen.
- rggm** Fitting Gaussian Graphical Models with robustified methods. By Masashi Miyamura.
- rjacobi** Compute Jacobi polynomials and Gauss-Jacobi quadrature related operations. By Paulo Jabardo.
- sciplot** Scientific graphs with error bars for data collected from one-way or higher factorial designs. By Manuel Morales.
- signal** A set of generally Matlab/Octave-compatible signal processing functions. Includes filter generation utilities, filtering functions, resampling routines, and visualization of filter models. By Tom Short.
- smoothtail** Smooth Estimation of the Generalized Pareto distribution shape parameter. By Kaspar Rufibach and Samuel Mueller.
- snp.plotter** Creates plots of p -values using single SNP and/or haplotype data, with options to display a linkage disequilibrium (LD) plot and the ability to plot multiple data sets simultaneously. By Augustin Luna and Kristin K. Nicodemus.
- spsurvey** Algorithms required for design and analysis of probability surveys such as those utilized by the U.S. Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP). By Tom Kincaid and Tony Olsen, with contributions from Don Stevens, Christian Platt, Denis White, and Richard Remington.
- staRt** Inferenza classica con TI-83 Plus. Una libreria per utilizzare con semplicità le tecniche di statistica inferenziale presenti sulla calcolatrice scientifica grafica TI-83 Plus. By Fabio Frascati.
- stashR** A set of tools for administering shared repositories. By Sandy Eckel, and Roger D. Peng.
- stochmod** Stochastic Modeling: learning and inference algorithms for a variety of probabilistic models. By Artem Sokolov.
- stream.net** Functions with example data for creating, importing, attributing, analyzing, and displaying stream networks represented as binary trees. Capabilities include upstream and downstream distance matrices, stochastic network generation, segmentation of network into reaches, adding attributes to reaches with specified statistical distributions, interpolating reach attributes from sparse data, analyzing auto-correlation of reach attributes, and creating maps with legends of attribute data. Target applications include dynamic fish modeling. By Denis White.
- svcm** Fir 2-d and 3-d space-varying coefficient models to regular grid data using either a full B-spline tensor product approach or a sequential approximation. By Susanne Heim, with support from Paul Eilers, Thomas Kneib, and Michael Kobl.

tdm A tool for therapeutic drug monitoring. Can estimate individual pharmacokinetic parameters with one or more drug serum/plasma concentrations obtained from a single subject or multiple subjects, calculate a suggested dose with the target drug concentration or calculate a predicted drug concentration with a given dose. By Miou-Ting Chen, Yung-Jin Lee.

titan GUI to analyze mass spectrometric data on the relative abundance of two substances from a titration series. By Tom Price.

tkrgl TK widget tools for package **rgl**. By Duncan Murdoch and Ming Chen.

tm A framework for text mining applications within R. By Ingo Feinerer.

trip Access and manipulate spatial data for animal tracking. By Michael D. Sumner.

wombsoft Wombling Computation. Analyses individually geo-referenced multilocus genotypes for the inferences of genetic boundaries between populations. Based on the wombling method which estimates the systemic function by looking for the local variation of the allele frequencies. By Ameline Crida.

yaImpute Perform k -NN imputation. By Nicholas L. Crookston and Andrew O. Finley.

Other changes

- Package **micEcdat** was moved from the main CRAN section to the Archive.
- Package **SAGx** was removed from CRAN.

Kurt Hornik
Wirtschaftsuniversität Wien, Austria
 Kurt.Hornik@R-project.org

R Foundation News

by *Kurt Hornik*

Donations and new members

Donations

Peter M Maksymuk (USA)

New benefactors

Schröder Investment Management Ltd., London, GB
 Prediction Company, Santa Fe, New Mexico, USA

New supporting institutions

Ef-prime Inc., Tokyo, Japan

New supporting members

David Vonka (CZ)
 Olivier Celhay (France)
 Katarzyna Kopczewska (Poland)

Kurt Hornik
Wirtschaftsuniversität Wien, Austria
 Kurt.Hornik@R-project.org

R News Referees 2006

by *Torsten Hothorn*

R News articles are peer-reviewed by external and independent experts. The editorial board members would like to take the opportunity to thank all referees who read and commented on submitted manuscripts during the last year. Much of the quality of R News publications is due to their invaluable and timely service. Thank you!

- Murray Aitkin
- Doug Bates
- Axel Benner

- Adrian Bowman
- Patrick Burns
- Philip Dixon
- Romain Francois
- Thomas Gerds
- Jos Hageman
- Frank Harrell
- David James
- Sigbert Klinke