I would strongly recommend the book to scientists who have already had a regression or a linear models course and who wish to learn to use R. However, my recommendation has a few caveats. The first chapter of the book takes the reader through an introduction to R that has the potential to be a little frustrating for a reader with no prior R experience. For example, the first plotting command given is

```
plot(ACT ~ Year, data=ACTpop, pch=16)
```

The meaning of the `pch=16` option is described and the option `data=ACTpop` is self evident, but the syntax `ACT ~ Year` is not explained and is potentially confusing to an R beginner who does not automatically translate ~ into "is modeled by". Page 5 gives the advice to create a new workspace before experimenting with R functions, but provides no details on how one actually does this. Most examples of R code in the book do contain adequate descriptions, but there are a number of exceptions.

A second caveat is that the descriptions of statistical methods are an adequate refresher, but are inadequate as a primary source of information. The authors indicate clearly that the book is meant to complement other books in the presentation of, and the mathematical description of, statistical methods. I agree that the book would not work well as a stand-alone text book for a course on statistical modeling. However, it is also not short and I would hesitate to require students to buy it in addition to another comprehensive textbook. The scope of the book is greater than simply serving as a companion book for teaching R.

Despite my hesitation to use this book in teaching, I give it a strong recommendation to the scientist or data analyst who wishes an easy-to-read and an understandable reference on the use of R for practical data analysis.

Bret Larget
University of Wisconsin—Madison
larget@stat.wisc.edu

Changes in R 1.8.0

by the R Core Team

MacOS changes

- As from this release there is only one R port for the Macintosh, which runs only on MacOS X. (The 'Carbon' port has been discontinued, and the 'Darwin' port is part of the new version.) The current version can be run either as a command-line application or as an 'Aqua' console. There is a 'Quartz' device `quartz()`, and the download and installation of both source and binary packages is supported from the Aqua console. Those CRAN and BioC packages which build under MacOS X have binary versions updated daily.

User-visible changes

- The defaults for `glm.control(epsilon=1e-8, maxit=25)` have been tightened: this will produce more accurate results, slightly slower.
- sub, gsub, grep, regexpr, chartr, tolower, toupper, substr, substring, abbreviate and strsplit now handle missing values differently from "NA".
- Saving data containing name space references no longer warns about name spaces possibly being unavailable on load.
- On Unix-like systems interrupt signals now set a flag that is checked periodically rather than calling longjmp from the signal handler. This is analogous to the behavior on Windows. This reduces responsiveness to interrupts but prevents bugs caused by interrupting computations in a way that leaves the system in an inconsistent state. It also reduces the number of system calls, which can speed up computations on some platforms and make R more usable with systems like Mosix.

Changes to the language

- Error and warning handling has been modified to incorporate a flexible condition handling mechanism. See the online documentation of `tryCatch()` and `signalCondition()`. Code that does not use these new facilities should remain unaffected.
- A triple colon operator can be used to access values of internal variables in a name space (i.e. `a:::b` is the value of the internal variable `b` in name space `a`).
- Non-syntactic variable names can now be specified by inclusion between backticks `` `Like This` ``. The `deparse()` code has been changed to output non-syntactical names with this convention, when they occur as operands in expressions. This is controlled by a `backtick`

argument, which is by default `TRUE` for composite expressions and `FALSE` for single symbols. This should give minimal interference with existing code.

- Variables in formulae can be quoted by backticks, and such formulae can be used in the common model-fitting functions. `terms.formula()` will quote (by backticks) non-syntactic names in its `"term.labels"` attribute. [Note that other code using terms objects may expect syntactic names and/or not accept quoted names: such code will still work if the new feature is not used.]

New features

- New function `bquote()` does partial substitution like LISP backquote.
- `capture.output()` takes arbitrary connections for `file` argument.
- `contr.poly()` has a new `scores` argument to use as the base set for the polynomials.
- `cor()` has a new argument `method = c("pearson","spearman","kendall")'` as `cor.test()` did forever. The two rank based measures do work with all three missing value strategies.
- New utility function `cov2cor()` Cov -> Corr matrix.
- `cut.POSIXt()` now allows 'breaks' to be more general intervals as allowed for the 'by' argument to `seq.POSIXt()`.
- `data()` now has an `envir` argument.
- `det()` uses an LU decomposition and LAPACK. The `method` argument to `det()` no longer has any effect.
- `dev.control()` now accepts `enable` as well as `inhibit`. (Wishlist PR#3424)
- `*`, `-` and `/` work more generally on `"difftime"` objects, which now have a `diff()` method.
- `dt(*, ncp = V)` is now implemented, thanks to Claus Ekstroem.
- `dump()` only quotes object names in the file where necessary.
- `eval()` of a promise forces the promise
- `file.path()` now returns an empty character vector if given at least one zero-length argument.
- `format()` and hence `print()` make an effort to handle corrupt data frames, with a warning.
- `format.info()` now also works with 'nsmall' in analogy with `format.default()`.
- `gamma(n)` is very slightly more precise for integer n in 11:50.
- `?` and `help()` will accept more un-quoted arguments, e.g. `NULL`.
- The `?` operator has new forms for querying documentation on S4 methods. See the online documentation.
- New argument `frame.plot = axes (== TRUE)` for `filled.contour()`.
- New argument `fixed = TRUE` for `grep()` and `regexpr()` to avoid the need to escape strings to match.
- `grep(x, ..., value = TRUE)` preserves names of x.
- `hist.POSIXt()` can now pass arguments to `hist.default()`
- `legend()` and `symbols()` now make use of `xy.coords()` and accept a wider range of coordinate specifications.
- Added function `library.dynam.unload()` to call `dyn.unload()` on a loaded DLL and tidy up. This is called for all the standard packages in namespaces with DLLs if their namespaces are unloaded.
- `lm(singular.ok = FALSE)` is now implemented.
- Empty `lm()` and `glm()` fits are now handled by the normal code: there are no methods for classes `"lm.null"` and `"glm.null"`. Zero-rank fits are handled consistently.
- `make.names()` has improvements, and there is a new auxiliary function `make.unique()`. (Based on code contributed by Tom Minka, since converted to a `.Internal` function.) In particular `make.names()` now recognises that names beginning with a dot are valid and that reserved words are not.
- `methods()` has a print method which asterisks functions which are not user-visible. `methods(class = "foo")` now lists non-visible functions, and checks that there is a matching generic.
- `model.matrix()` now warns when it removes the response from the rhs of the formula: that this happens is now documented on its help page.

- New option `"locatorBell"` to control the confirmation beep during the use of `locator()` and `identify()`.
- New `option("scipen")` provides some user control over the printing of numbers in fixed-point or exponential notation. (Contributed by David Brahm.)
- `plot.formula()` now accepts `horizontal=TRUE` and works correctly when boxplots are produced. (Wishlist PR#1207) The code has been much simplified and corrected.
- `polygon()` and `rect()` now interpret density < 0 or NA to mean filling (by colour) is desired: this allows filling and shading to be mixed in one call, e.g. from `legend()`.
- The `predict()` methods for classes lm, glm, mlm and lqs take a 'na.action' argument that controls how missing values in 'newdata' are handled (and defaults to predicting NA). [Previously the value of `getOption("na.action")` was used and this by default omitted cases with missing values, even if set to 'na.exclude'.]
- `print.summary.glm()` now reports omitted coefficients in the same way as `print.summary.lm()`, and both show them as NAs in the table of coefficients.
- `print.table()` has a new argument 'zero.print' and is now documented.
- `rank(x, na.last = "keep")` now preserves NAs in 'x', and the argument 'ties.method' allows to use non-averaging ranks in the presence of ties.
- `read.table()`'s 'as.is' argument can be character, naming columns not to be converted.
- `rep()` is now a generic function, with default, POSIXct and POSIXlt methods. For efficiency, the base code uses `rep.int()` rather than `rep()` where possible.
- New function `replicate()` for repeated evaluation of expression and collection of results, wrapping a common use of `sapply()` for simulation purposes.
- `rev()` is now a generic function, with default and dendrogram methods.
- `serialize()` and `unserialize()` functions are available for low-level serialization to connections.
- `socketSelect()` allows waiting on multiple sockets.
- `sort(method = "quick", decreasing = TRUE)` is now implemented.
- `sort.list()` has methods `"quick"` (a wrapper for `sort(method = "quick", index.return = TRUE)` and `"radix"` (a very fast method for small integers). The default `"shell"` method works faster on long vectors with many ties.
- `stripchart()` now has 'log', 'add' and 'at' arguments.
- `strsplit(x, *)` now preserves `names()` but won't work for non-character 'x' anymore formerly used `as.character(x)`, destroying `names(x)`.
- `textConnection()` now has a local argument for use with output connections. `local = TRUE` means the variable containing the output is assigned in the frame of the caller.
- Using `UseMethod()` with more than two arguments now gives a warning (as R-lang.texi has long claimed it did).
- New function `vignette()` for viewing or listing vignettes.
- `which.min(x)` and `which.max(x)` now preserve names.
- `xy.coords()` coerces "POSIXt" objects to "POSIXct", allowing lines, etc. to be added to `plot.POSIXlt()` plots.
- `.Machine` has a new entry, sizeof.pointer.
- `.Random.seed` is only looked for and stored in the user's workspace. Previously the first place a variable of that name was found on the search path was used.
- Subscripting for data.frames has been rationalized:
  - Using a single argument now ignores any 'drop' argument (with a warning). Previously using 'drop' inhibited list-like subscripting.
  - `adf$name <- value` now checks for the correct length of 'value', replicating a whole number of times if needed.
  - `adf[j] <- value` and `adf[[j]] <- value` did not convert character vectors to factors, but `adf[,j] <- value` did. Now none do. Nor is a list 'value' coerced to a data frame (thereby coercing character elements to factors).
  - Where replicating the replacement value a whole number of times will produce the right number of values, this is always done (rather than some times but not others).

  - Replacement list values can include NULL elements.
  - Subsetting a data frame can no longer produce duplicate column names.
  - Subsetting with `drop=TRUE` no longer sometimes drops dimensions on matrix or data frame columns of the data frame.
  - Attributes are no longer stripped when replacing part of a column.
  - Columns added in replacement operations will always be named, using the names of a list value if appropriate.
  - `as.data.frame.list()` did not cope with list names such as 'check.rows', and formatting/printing data frames with such column names now works.
  - Row names in extraction are still made unique, but without forcing them to be syntactic names.
  - `adf[x] <- list()` failed if `x` was of length zero.

- Setting dimnames to a factor now coerces to character, as S does. (Earlier versions of R used the internal codes.)

- When coercion of a list fails, a meaningful error message is given.

- Adding to `NULL` with `[[ ]]` generates a list if more than one element is added (as S does).

- There is a new command-line flag '`--args`' that causes the rest of the command line to be skipped (but recorded in `commandArgs()` for further processing).

- S4 generic functions and method dispatch have been modified to make the generic functions more self-contained (e.g., usable in apply-type operations) and potentially to speed dispatch.

- The data editor is no longer limited to 65535 rows, and will be substantially faster for large numbers of columns.

- Standalone Rmath now has a `get_seed` function as requested (PR#3160).

- GC timing is not enabled until the first call to `gc.time()`; it can be disabled by calling `gc.time(FALSE)`. This can speed up the garbage collector and reduce system calls on some platforms.

Standard packages

- New package 'mle'. This is a simple package to find maximum likelihood estimates, and perform likelihood profiling and approximate confidence limits based upon it. A well-behaved likelihood function is assumed, and it is the responsibility of the user to gauge the applicability of the asymptotic theory. This package is based on S4 methods and classes.

- Changes in package 'mva':
  - `factanal()` now returns the test statistic and P-value formerly computed in the print method.
  - `heatmap()` has many more arguments, partly thanks to Wolfgang Huber and Andy Liaw.
  - Arguments 'unit' and 'hmin' of `plclust()` are now implemented.
  - `prcomp()` now accepts complex matrices, and there is `biplot()` method for its output (in the real case).
  - dendrograms are slightly better documented, methods working with "label", not "text" attribute. New `rev()` method for dendrograms.
  - `plot.dendrogram()` has an explicit 'frame.plot' argument defaulting to FALSE (instead of an implicit one defaulting to TRUE).

- Changes in package 'tcltk':
  - The package is now in a namespace. To remove it you will now need to use `unloadNamespace("tcltk")`.
  - The interface to Tcl has been made much more efficient by evaluating Tcl commands via a vector of Tcl objects rather than by constructing the string representation.
  - An interface to Tcl arrays has been introduced.
  - `as.tclObj()` has gained a 'drop' argument to resolve an ambiguity for vectors of length one.

- Changes in package 'tools':
  - Utilities for testing and listing files, manipulating file paths, and delimited pattern matching are now exported.
  - Functions

    ```
    checkAssignFuns()
    checkDocArgs()
    checkMethods()
    ```

  have been renamed to

  `checkReplaceFuns()`
  `checkDocFiles()`
  `checkS3methods()`

  to given better descriptions of what they do.

  - R itself is now used for analyzing the markup in the `\usage` sections. Hence in particular, replacement functions or S3 replacement methods are no longer ignored.
  - `checkDocFiles()` now also determines 'over-documented' arguments which are given in the `\arguments` section but not in `\usage`.
  - `checkDocStyle()` and `checkS3Methods()` now know about internal S3 generics and S3 group generics.
  - S4 classes and methods are included in the QC tests. Warnings will be issued from `undoc()` for classes and methods defined but not documented. Default methods automatically generated from nongeneric functions do not need to be documented.
  - New (experimental) functions

    `codocClasses()`
    `codocData()`

    for code/documentation consistency checking for S4 classes and data sets.

- Changes in package 'ts':
  - `arima.sim()` now checks for inconsistent order specification (as requested in PR#3495: it was previously documented not to).
  - `decompose()` has a new argument 'filter'.
  - `HoltWinters()` has new arguments 'optim.start' and 'optim.control', and returns more components in the fitted values. The plot method allows 'ylim' to be set.
  - `plot.ts()` has a new argument 'nc' controlling the number of columns (with default the old behaviour for plot.mts).
  - `StructTS()` now allows the first value of the series to be missing (although it is better to omit leading NAs). (PR#3990)

Using packages

- `library()` has a pos argument, controlling where the package is attached (defaulting to pos=2 as before).
- `require()` now maintains a list of required packages in the toplevel environment (typically, .GlobalEnv). Two features use this: `detach()` now warns if a package is detached that is required by an attached package, and packages that install with saved images no longer need to use `require()` in the .First as well as in the main source.
- Packages with name spaces can now be installed using '`--save`'.
- Packages that use S4 classes and methods should now work with or without saved images (saved images are still recommended for efficiency), writing `setMethod()`, etc. calls with the default for argument 'where'. The `topenv()` function and `sys.source()` have been changed correspondingly. See the online help.
- Users can specify in the DESCRIPTION file the collation order for files in the R source directory of a package.

R documentation format

- New logical markup commands for emphasizing (`\strong`) and quoting (`\sQuote` and `\dQuote`) text, for indicating the usage of an S4 method (`\S4method`), and for indicating specific kinds of text (`\acronym`, `\cite`, `\command`, `\dfn`, `\env`, `\kbd`, `\option`, `\pkg`, `\samp`, `\var`).
- New markup `\preformatted` for preformatted blocks of text (like example but within another section). (Based on a contribution by Greg Warnes.)
- New markup `\concept` for concept index entries for use by `help.search()`.
- Rdconv now produces more informative output from the special `\method{GENERIC}{CLASS}` markup for indicating the usage of S3 methods, providing the CLASS info in a comment.
- `\dontrun` sections are now marked within comments in the user-readable versions of the converted help pages.
- `\dontshow` is now the preferred name for `\testonly`.

Installation changes

- The zlib code in the sources is used unless the external version found is at least version 1.1.4 (up from 1.1.3).
- The regression checks now have to be passed exactly, except those depending on recommended packages (which cannot be assumed to be present).

- The target make check-all now runs R CMD check on all the recommended packages (and not just runs their examples).
- There are new macros DYLIB_* for building dynamic libraries, and these are used for the dynamic Rmath library (which was previously built as a shared object).
- If a system function log1p is found, it is tested for accuracy and if inadequate the substitute function in src/nmath is used, with name remapped to Rlog1p. (Apparently needed on OpenBSD/NetBSD.)

C-level facilities

- There is a new installed header file R_ext/Parse.h which allows R_ParseVector to be called by those writing extensions. (Note that the interface is changed from that used in the unexported header Parse.h in earlier versions, and is not guaranteed to remain unchanged.)
- The header R_ext/Mathlib.h has been removed. It was replaced by Rmath.h in R 1.2.0.
- PREXPR has been replaced by two macros, PREXPR for obtaining the expression and PRCODE for obtaining the code for use in eval. The macro BODY_EXPR has been added for use with closures. For a closure with a byte compiled body, the macro BODY_EXPR returns the expression that was compiled; if the body is not compiled then the body is returned. This is to support byte compilation.
- Internal support for executing byte compiled code has been added. A compiler for producing byte compiled code will be made available separately and should become part of a future R release.
- On Unix-like systems calls to the `popen()` and `system()` C library functions now go through R_popen and R_system. On Mac OS X these suspend SIGALRM interrupts around the library call. (Related to PR#1140.)

Utilities

- R CMD check accepts "ORPHANED" as package maintainer. Package maintainers can now officially orphan a package, i.e., resign from maintaining a package.
- R CMD INSTALL (Unix only) is now 'safe': if the attempt to install a package fails, leftovers are removed. If the package was already installed, the old version is restored.
- R CMD build excludes possible (obsolete) data and vignette indices in DCF format (and hence also no longer rebuilds them).
- R CMD check now tests whether file names are valid across file systems and supported operating system platforms. There is some support for code/documentation consistency checking for data sets and S4 classes. Replacement functions and S3 methods in `\usage` sections are no longer ignored.
- R CMD Rdindex has been removed.

Deprecated & defunct

- The assignment operator '_' has been removed.
- `printNoClass()` is defunct.
- The classic MacOS port is no longer supported, and its files have been removed from the sources.
- The deprecated argument 'white' of `parse()` has been removed.
- Methods `pacf/plot.mts()` have been removed and their functionality incorporated into `pacf.default/plot.ts()`.
- `print.coefmat()` is deprecated in favour of `printCoefmat()` (which is identical apart from the default for na.print which is changed from "" to "NA", and better handling of the 0-rank case where all coefficients are missing).
- `codes()` and `codes<-()` are deprecated, as almost all uses misunderstood what they actually do.
- The use of multi-argument `return()` calls is deprecated: use a (named) list instead.
- anovalist.lm (replaced in 1.2.0) is now deprecated.
- `-` and `Ops` methods for POSIX[cl]t objects are removed: the POSIXt methods have been used since 1.3.0.
- `glm.fit.null()`, `lm.fit.null()` and `lm.wfit.null()` are deprecated.
- Classes "lm.null" and "glm.null" are deprecated and all of their methods have been removed.
- Method `weights.lm()`, a copy of `weights.default()`, has been removed.
- `print.atomic()` is now deprecated.
- The back-compatibility entry point Rf_log1p in standalone Rmath has been removed.