



Journal of Statistical Software

May 2008, Volume 25, Book Review 2.

<http://www.jstatsoft.org/>

Reviewer: Gabor Grothendieck
GKX Associates Inc.

lattice: Multivariate Data Visualization with R

Deepayan Sarkar
Springer-Verlag, New York, NY, 2008.
ISBN 978-0-387-75968-5. 268 pp. USD 54.95 (P).
<http://dsarkar.fhcrc.org/lattice/>

The R **lattice** package is a powerful and complex system implementing William Cleveland's trellis graphics (Cleveland 1985) – a system focusing on the display of multivariate data in panels. Trellis was one of the many ideas that came from a particularly notable generation in the history of statistical developments at Bell Labs – a period that also included the S programming language (Becker, Chambers, and Wilks 1988). The attraction to **lattice** is derived from its inclusion in the R download, i.e., it need not be separately downloaded from the Comprehensive R Archive Network (CRAN), its Bell Labs heritage, its wide variety of displayable graphics (`xyplot`, `bwplot`, `densityplot`, `levelplot`, etc.), its portability as it is written entirely in R, its maturity through years of improvement and its customizability. The last item is owing to a variety of graphical parameters, flexible use of panel functions, extensibility through the object model and its leverage of the underlying R **grid** graphics system (R Development Core Team 2008) upon which it is built. The package's author and author of the book under review is a member of the R development core team.

The nature of **lattice** may be gleaned from the following R code which creates a multipanel graphic using **lattice**. Here, `yield` and `nitro` are numeric while `Variety` and `Block` are factors. The formula specifies that for each combination of those factors a scatterplot of `yield` vs. `nitro` is to be displayed in a separate panel. Note that `type` is similar to the argument of the same name in R's classic graphics and the "key", i.e., legend, can be specified within the command itself rather than via a separate `legend` command, as would be the case in most classic graphics. Also, we see that the result of the `xyplot` function is not a graphic but an object, here assigned to `p`, that can be further manipulated. The most common operation is simply to plot the object which is done via the `print` command below.

```
R> library("lattice")
R> data(Oats, package = "MEMSS")
R> p <- xyplot(yield ~ nitro | Variety + Block, Oats,
+   type = "o", auto.key = TRUE)
R> print(p)
```

Because of its substantial power and history the package has drawn many users yet the relatively terse documentation has meant that getting up to speed usually involved scavenging sample code from the internet. One always had the feeling that the author was the sole expert in its use. This volume attempts to address that with a book length treatment of the package.

The book attempts to address both new users through Part I which is roughly the first 100 pages and more experienced users through Part II which is approximately the next 100 pages. The latter covers individual topics with chapters on graphical parameters, axes, etc. Finally, Part III, discusses extending **lattice** via panel functions and object orientation.

With 150 sample graphics throughout the book, including both black and white output as well as color plates, it has sufficient sample code to satisfy most readers yet it is not a cookbook or reference and seems to aim at conveying **lattice** concepts making use of whatever examples are needed in the process.

Part I consists of an introductory Chapter 1 which dives right into examples, followed by Chapter 2 which discusses the main concepts behind **lattice** including conditioning, shingles, aspect, layout, scales, annotation, panel functions and other concepts. The remaining chapters of Part I focus on univariate, e.g., **densityplot**, bivariate and multivariate scatterplots, e.g., **xyplot**, trivariate, e.g., **cloud** and graphical representations of tables, e.g., **dotplot**. Part I will be particularly useful for new users. For existing users it does have a number of tips that may be new and also some useful comparative examples of the effectiveness of different displays such as the various representations of the **quake** data in Chapter 5. Examples of using related packages such as **latticeExtra** (Sarkar 2007) and **hexbin** (Carr, Lewin-Koh, and Maechler 2008) will also be of interest.

Part II covers a number of specific topics such as graphical parameters (Chapter 7), axes (Chapter 8), legends (Chapter 9) and others followed by chapters focusing on data (Chapter 10), objects (Chapter 11) and interaction (Chapter 12). We note the slick single page visual in Chapter 7 organizing the multitude of otherwise bewildering graphical parameters and some unexpected applications of shingles in Chapter 10 as two sample items from this part.

Part III on extending **lattice** discusses panel functions (Chapter 13) with eye catching corrgram (Friendly 2002) and mapping examples while Chapter 14 gives examples of creating new **lattice** functions. Note that there is a package **corrgram** (Wright 2006) on CRAN but it uses classic graphics differentiating the presentation here.

The exposition is clear and while a completely linear treatment of such a complex package is probably not realistic the author was careful to make the forward references explicit when such were needed. The text has few errors although we did manage to catch a typo on page 109 where Figure 6.15 is erroneously referred to as Figure 6.4.

Not all the packages mentioned in the book are cited in the bibliography which seems an omission in light of the R **citation** function. Also the index can be difficult to use until one discovers that individual “Arguments”, “Datasets” and “Packages” are listed as second level entries under those three topics, respectively. However, once understood that organization does become convenient.

As many of the chapters are independent of each other, readers may wish to skim the book and then focus on particular sections of interest while creating their own graphics. Some additional reference to the online help may still be necessary but having read the book the user will have a better understanding of the concepts and a wealth of examples to rely on.

The book is destined to be the standard reference on **lattice** and is highly recommended to those with an interest in using **lattice** graphics in R.

References

- Becker RA, Chambers JM, Wilks AR (1988). *The New S Language*. Wadsworth.
- Carr D, Lewin-Koh N, Maechler M (2008). *hexbin: Hexagonal Binning Routines*. R package version 1.14.0, URL <http://www.bioconductor.org/>.
- Cleveland W (1985). *The Elements of Graphing Data*. Wadsworth.
- Friendly M (2002). “Corrgrams: Exploratory Displays for Correlation Matrices.” *The American Statistician*, **56**, 316–324.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Sarkar D (2007). *latticeExtra: Extra Graphical Utilities Based on lattice*. R package version 0.3-1, URL <http://CRAN.R-project.org/package=latticeExtra>.
- Wright K (2006). *corrgram: Plot a Correlogram*. R package version 0.1, URL <http://CRAN.R-project.org/package=corrgram>.

Reviewer:

Gabor Grothendieck
GKX Associates Inc.
E-mail: ggrothendieck@gmail.com