



R in a Nutshell, A Desktop Quick Reference

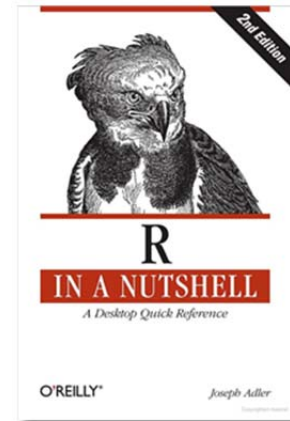
by Joseph Adler. O'Reilly 2011
ISBN: 978-0-596-80170-0

Three months ago I switched from a Dell with Windows and Linux to a MAC with a Windows virtual machine. I had been using wgnuplot for 13 years for most of my graphing but was determined to find an alternative native to the MAC. I hadn't had much luck until I came across a posting about R on crystallography BB. I downloaded a copy and started playing with it. I was lost. I started looking at the online documentation but I needed something more. I went online and found this O'Reilly book.

This is the first O'Reilly book I've read cover-to-cover and the first programming book in more than a decade. R is a programming language for graphics and statistics. It provides for interactive graphing and analysis. It is much more powerful than wgnuplot but with that comes more structure and perhaps pitfalls. It is maintained by CRAN and freely available and covered by the GNU General Public License. Adler provides many examples of data sets that can be downloaded from the Internet for analysis.

The author begins with the basics: how to get R, how to install R, how to drive the user interface and how get packages that extend the capabilities beyond the basics. In the second part Adler covers the R language with chapters on syntax, objects, symbols and environments, and functions. This part concludes with a chapter on the basics of object oriented programming as applied to R and making R high performance.

The next section of the book deals with data and data structures, and graphing. The default graphics are good but there is an implementation of Trellis graphics, called



lattice, that really makes some interesting plots. Lattice is not loaded by default so you'll want to load that package.

Part III covers the analysis of data, statistical tests and regression. This was a fun section because I feel that I really learned something about the statistics. I am reading the *The Black Swan* by N. N. Taleb and am also unlearning things about statistics but that is another review. Also covered are classification methods, survival methods and times series data. I was disappointed to find no Fourier transform in R since the very thing I do with time series data is an FT to analyze the components. The last chapter of Part III provides an example of the use of R in a bioinformatics study.

The remainder of the book, not quite 100 pages, is the R reference manual. If you are using R, you want this book. If you are doing statistics, there are certainly better references but this is a good primer.

I should point out that in the two years since this review first appeared I have built gnuplot on OSX. The reason is that I have many older gnuplot files and sometime have to replot them on a modern device. Unfortunately they all have to be hand edited since many commands have changed or became deprecated over the last 15 years.

Book review by Joseph Ferrara, Ph.D.