Note re developments in R, that affect code in the text of

**Data Analysis and Graphics Using R – An Example-Based Approach, 2nd edn**

John Maindonald (email: john.maindonald@anu.edu.au) and John Braun

December 23, 2008

In addition to specific changes that affect the text as it stands, note the document *Draft of Changes and Additions in a Projected 3rd Edition*. This has drafts of the following changes and additions to the current text:

- A new section, likely to be included in Chapter 14, has a brief discussion of graphical user interfaces (GUIs) for use with R.
- The function `auto.arima()` in the `forecast` package (in the `forecasting` package bundle) can be used to automate model selection. A revised version of Sections 9.1 and 9.2 is included in the document *Draft of Changes and Additions in a Projected 3rd Edition*.
- A rewrite of Section 14.12 on `lattice` graphics has further details on interaction with plots. It includes also brief details on the `latticist` package that gives a simple GUI interface to `lattice`. Additionally, the `playwith` package that has extensive abilities for interaction with plots, whether created using `lattice` or by another system.

**Enhancements to the lattice package** Here are noted changes that allow simplification of code that is scattered through the text.

**Columns in parallel**

Columns can now be plotted in parallel. Code that stacks the relevant columns to create a new data frame, as on pages 55 (lines -8 to -12) and 433 (lines 6, 8 and 10), is thus redundant. The call to `xyplot()` in lines 15 through 17 on page 433 becomes:

```r
xyplot(Ontario+Quebec+BC+Alberta+Prairies+Atlantic ~ Date,
       outer=FALSE, data=jobs, ylab="Number of workers",
       scales=list(at=atdates, labels=datelabs),
       auto.key=list(columns=3, between=1))
```

The column names (Ontario, etc.) are juxtaposed in the graphics formula, with “+” as separator. Use of `outer=FALSE` causes overplotting in one panel, rather than the use of separate panels.

**Point, line and fill color settings – simpleTheme()**

Using the function `simpleTheme()`, the code in footnote 8 on page 57 (Figure 2.10B) simplifies to:

```r
xyplot(csoa ~ it | sex*agegp, data=tinting, groups=target,
       par.settings=simpleTheme(col=c("black","gray20"),
                                pch=c(1, 16)),
       auto.key=list(columns=2))
```

Similarly also pages 58 (footnote 9), 415 (lines 2 and 3) and 463 (lines -7 to -4). Lines 2 and 3 on page 415 simplify to:

```r
trellis.par.set(theme=simpleTheme(cex=0.25, lwd=2, col.line=linecols))
```
Chapter 6 – Interpretation of Regression Coefficients  Regression coefficients from the various sets of hill race data make much better sense if \( \log(\text{time}) \) is regressed on \( \log(\text{dist}) \) and \( \log(\text{climb/dist}) \). Think!

Chapter 10 – output from \texttt{mcmcsamp}(): Changes to the structure of objects created by \texttt{mcmcsamp()} affect code on pages 310 (the lower half-page), 314 (lines -16 to -10) and 343 (line 2). The code on page 310 starts with the two lines:

\begin{verbatim}
> set.seed(41)  # Use to reproduce results below
> ant111b.samp <- mcmcsamp(ant111b.lmer, n=1000)
\end{verbatim}

The code in subsequent lines is no longer valid. Instead, use the function \texttt{HPDinterval()} to extract Bayesian credible intervals, thus:

\begin{verbatim}
HPDinterval(ant111b.samp, prob=0.95)
\end{verbatim}

(If \texttt{HPDinterval()} is not in the installed version of \texttt{lme4}, attach the \texttt{coda} package.)

Chapter 14  Note the change in Sec 14.5, to the subsubsection headed Counting and identifying NAs – the use of \texttt{table}(). With the argument \texttt{exclude=NULL}, the function \texttt{table()} now tabulates missing values, both for vectors and factors. The argument \texttt{useNA}, (set to "no" or "ifany" or "always") allows finer control over the table that appears.

Additional details:  See the web page
\begin{verbatim}
\end{verbatim}