

Corrections and Elucidations (as of April 12, 2007) for the 2nd edition of
Data Analysis and Graphics Using R – An Example-Based Approach

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

We are indebted to Raymond Balise (Stanford University) for a number of these corrections.

Chapter 1

p.16, line -6 (Section 1.4.1)

Replace `subset(Cars93.summary, subset=1:2)`
by

`subset(Cars93.summary, subset=c(TRUE, TRUE, FALSE, FALSE, FALSE, FALSE))`

The `subset` argument requires a logical expression indicating which rows to keep. Rows may not be identified by number.

p.33, line -10 (Section 1.8)

Remove the unwanted `)` from `save.image()`

p.39, exercise 6(b) (Section 1.11)

Replace “The `log="y"` setting is automatic, after its initial use with `plot()`, for the subsequent use of `text()`.”
by
“The `log="y"` setting carries across to the subsequent `text()` commands. See Section 2.1.5 for an example.”

p.40, exercise 16 (Section 1.11)

Replace `census1889` by `census1886`, in each of the three places where it appears.

p.42, exercise 20 (Section 1.11)

Replace this exercise by:

The help page for `iris` (type `help(iris)`) gives code that converts the data in `iris3` (*datasets* package) to case-by-variable format, with column names “Sepal.Length”, “Sepal.Width”, “Petal.Length”, “Petal.Width”, and “Species”. Look up the help pages for the functions that are used, and make sure that you understand them. Then add annotation to the code that explains each step in the computation.

This should be an asterisked question.

[Note that asterisks are used to identify exercises that are challenging, or that are overly technical.]

Chapter 2

p.48, code segment in final 8 lines (ss 2.1.2)

line -7: Omit the space between `log10` and `c(1,5000*1000)`

line -2: This should read

```
plot(window(measles, start=1840, end=1882), ylim=c(0, 4600), yaxt="n")
```

[`plot` has a lower case p. Omit two redundant (albeit harmless) spaces.]

p.49, footnote 4 (ss 2.1.2)

In order not to interfere with the plotting of subsequent graphs, the code should be followed by

```
par(fig=c(0,1,0,1))
```

p.55, lines -15 to -14 (ss 2.1.5)

Replace

Labeling is done on the untransformed (`Number`) scale, with values on the logarithmic scale given in brackets:

by

Labeling is done on the transformed (`log(Number)`) scale. The scale to the right of the graph relates `log(Number)` to `Number`.

line -6

Replace `c(2,3)` by `c(3,2)`.

lines -3 to -1

Omit `see Subsection 14.3.4. Alternatively,`

Add as the final sentence in the paragraph:

Subsection 14.3.4 describes how to label the x -axis with dates, in the format Jan95, Apr95, ...

p.70, l.14 (Section 2.3.1)

Replace `6.5` by `13.2`.

p.72, l.16 (Section 2.3.4)

Replace `, discussed briefly in Subsection 2.2.1,`
by `(see Subsection 2.2.1 and Section 13.2)`"

p.77, Exercise 7 (Section 2.6)

Replace `Download and load` by `Install and attach`.

[The package must be installed, which usually involves an initial downloading step. The action of `library()` is better described as attaching, rather than loading, the relevant package.]

Chapter 4

p.106, line -3 (ss 4.1.6)

Replace `90%` by `95%`.

p.121, line 14 (Section 4.4)

Replace `using the anova() function, thus:`
by

`using the aov() function to fit the analysis of variance model, then calling the anova() function with the resulting aov object as argument, thus:`

Chapter 5

p.172 (Section 5.11)

Exercise 12 should be asterisked.

Chapter 6

p.189 (ss 6.3.3), line 3

This should read

“ Figure 6.9B does seem marginally more consistent than Figure 6.9A with the assumption of a linear relationship between predictors. ”

(i.e., interchange A with B)

p.189 (ss 6.5.1), final line

Omit the space in `odd_ books`.

p.207 (ss 6.8.1), line 3

In line 3 of Subsection 6.8.1, replace `Such error, ...`
with

`Discussion will be limited to a relatively simple ‘classical “errors in x ” model. For this model the error in $x, ...$`

p.209 (ss 6.8.1), final paragraph

Replace `Variables that are measured inaccurately ... interpreting results.`
with

Again, attention will be limited to the classical “errors in x ” model. Where one only of several variables is measured inaccurately, its coefficient may on that account not appear statistically significant, or be severely attenuated. For remaining variables (measured without error) possible scenarios include: the coefficient suggests a relationship when there is none, or the coefficient is reversed in sign. Where several variables are measured with error, there is even more room for misleading and counter-intuitive coefficient values.

p.218 (Section 6.11)

In “Additional Notes”, there is a further exercise (exercise 12) that illustrates how measurement error in one variable may lead to a spurious effect for a variable that is measured without error.

Chapter 7***p.222 (ss 7.1.1), lines -4 to -3***

There are 8 degrees of freedom (not 66) for the residual standard error. These lines should read.

```
> qtukey(p=.95, nmeans=4, df=8) * sem
[1] 20.26
```

Chapter 8***p.266 (ss 8.4.2), line3 -3 and -2***

Replace `exp(-1.282)` by `exp(-1.283)`

Replace `Bank` by `Lowerside`. Omit the final sentence; it adds nothing to what has gone before.

p.268 (ss 8.4.2), lines 11-12

Omit these lines (they repeat lines 9-10), i.e., omit

```
> ANW.glm <- glm(A ~ habitatNW + log(meters), family = quasipoisson,
+               data=moths)
```

p.282 (ss 8.7.5), line 5

Omit `(or from $\beta = 0$)`. (The parenthetical comment is superfluous).

Chapter 10*p.302, line 2*Replace `nline` by `nlme`.*p.308 (ss 10.1.3)*In line 13 (“lower” of the ...), replace `0` by `1`. In line 14, replace `1` by `2`.*p.309 (ss 10.1.3), line 5:*Following “In”, insert `hierarchical`
i.e., “In `hierarchical` multi-level ...”*p.311 (ss 10.1.3)***line 5:** Replace `0` by `3`, `1` by `2`, and `2` by `1`. Thus the sentence becomes:“ We now have three levels of variation: level `3` is house, level `2` is suburb, and level `1` is city.”**line 8:** Replace `reflected` by `identified`.*p.325 (final 2 lines) & p.326 (first 2 lines), ss 10.3.5*(First paragraph under *Plots of residuals*)

Replace these first 4 lines, i.e.

Recall that by default, fitted values adjust for all except random variation between individual vines, i.e., they account for treatment, block and plot effects. For this, set `level=2` when calculating fitted values, or the equivalent residuals. Other choices are to calculate fitted values as treatment plus block (`level=1`) or as treatment effects only ...

by

In this hierarchical model there are three levels of variation: level 1 is between blocks, level 2 is between plots, and level 3 is between vines. The function `fitted()` adjusts for all levels of random variation except between individual vines, i.e. fitted values are at level 2.

Because the older `lme()` function (`nlme`) was designed for use with hierarchical models, the `fitted` method for `lme` model objects does accept the parameter `level`, here with a choice of 0, 1 or 2. ...

Thus, the paragraph reads

“In this hierarchical model there are three levels of variation: level 1 is between blocks, level 2 is between plots, and level 3 is between vines. The function `fitted()` adjusts for all levels of random variation except between individual vines, i.e. fitted values are at level 2.

Because the older `lme()` function (`nlme`) was designed for use with hierarchical models, the `fitted` method for `lme` model objects does accept the parameter `level`, here with a choice of 0, 1 or 2. The block *effects* are differences between fitted values at level 1 and fitted values at level 0, while the plot effects are differences between fitted values at level 2 and fitted values at level 1.“

Chapter 14*p.426, line -11 (Section 14.2)*Following “The command”, insert `datafile("bostonc")`*p.433, lines 7 & 11 (ss 14.3.4)*Replace `"1Jan1990"` by `"1Jan1995"`.**Redundant spaces***p.310, mid-page*`names(ant111b.samp)`*p.463***line -13:** `trellis.par.set()`**line -7:** Omit space following `xyplot`**line -6, -5 and -4:** Omit space following `list` (4 times), i.e. `list(`, not `list (`.

[The inclusion of a space between a function name and the `(` that precedes the argument is contrary to established practice. It does no harm, however.]

Index of Terms, pp.491-500**p.494**

Replace the heading at the top of column 1, i.e.

`distribution (cont.)`

by

`examples, analyses & analysis issues (cont.)`**Index of Authors, pp.501-502**

References to pages 43 to 75 (Chapter 2) are mostly, in each case, too large by 1. Thus `Aldrich, 64` should be `Aldrich, 63`.

Add `Box & Jenkins, 299`; `Chatfield, 97`; `Hyndman et al, 299`; `McCullagh & Nelder, 283`; `Nightingale, 76`; `Ord et al, 299`; `Snijders & Bosker, 347`; `Tukey, 43`.