Use of *knitr* to Generate Reproducible Reports

John Maindonald

April 24, 2013

1 Setup

This document was created from the **.Rnw** file **knitDemo.Rnw**. To create the LaTeX (**.tex**) file, first ensure that *knitr* and dependencies are installed. Then place the file **knitDemo.Rnw** in your working directory, and type from the R command line:

```
library(knitr)
knit('knitDemo.Rnw')
```

Options can be set near the start of the document that will apply to the document as a whole. If there are input files, such setup code may be required for each separate input file. See Section 6 below for the setup code used here.

2 Reproducible Reports

The function `knit()` from the *knitr* package is an alternative, with somewhat different markup conventions, to Rs `Sweave()` function. As with `Sweave()` text, and code that is embedded in suitable markup, are mixed in the one file. The file is then processed to give a document that includes any desired combination of code and computer output. Output may be any or all of printed computer output, output tables and output graphs.

The present document will demonstrate how code may be embedded in LaTeX files, with markup instructions used to control what appears in the LaTeX file that is generated, and thus in the final pdf file.

A simpler alternative, which does not require the use of LaTeX is to embed code in R Markdown files. The use of R Markdown is described in a separate document.

It is hard to over-estimate the importance of the abilities provided by `knitr()` and `Sweave()` for improving the quality of the research literature and for making pre-publication scrutiny part of a process that can continue after publication. For discussion of Excel-based published work where serious mistakes in computation were later found, and that make this point with some force, see:

Paul Krugman on "The Excel Depression": http://www.nytimes.com/2013/04/19/opinion/krugman-the-excel-depression.html?_r=0

Is The Reinhart-Rogoff Result Based on a Simple Spreadsheet Error? (Matthew Yglesias): http://www.slate.com/blogs/moneybox/2013/04/16/reinhart_rogoff_coding_error_austerity_policies_founded_on_bad_coding.html

The benefits of reproducible research: a public health example (Tim Churches): https://github.com/timchurches/meta-analyses/blob/master/benefits-of-reproducible-research/benefits-of-reproducible-research.md

3 Printing and Executing Code

Here are command line calculations:

```
2*3*4*5          # * denotes 'multiply'

## [1] 120

sqrt(10)         # the square root of 10

## [1] 3.162

pi               # R knows about pi

## [1] 3.142

2*pi*6378        # Circumference of earth at equator (km)

## [1] 40074

                 # (radius at equator is 6378 km)
```

4 Plotting a Graph

Code for Figure 1 is:

```
Year <- c(1800, 1850, 1900, 1950, 2000)
Carbon <- c(8, 54, 534, 1630, 6611)
## Now plot Carbon as a function of Year
plot(Carbon ~ Year, pch=16)
```

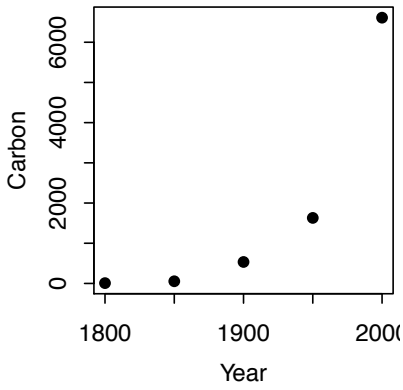

Figure 1: Plot of annual totals of carbon emissions (millions of tonnes), against year

5 Graphs Shown Side by Side

The dataset `mammals` *MASS* is an extreme example of data that cries out for transformation. Figure 2A shows the scatterplot for the raw data, while Figure 2B shows the scatterplot for the logged data.

```
library(MASS)
## Panel A
plot(brain ~ body, data=mammals, pty="s")
mtext(side=3, line=0.75, "A: Unlogged Data", adj=0)
## Panel B
plot(brain ~ body, data=mammals, log="xy", pty="s")
mtext(side=3, line=0.75, "B: Logaruthmic Scales", adj=0)
```

6 Code Chunks, with Markup

Section 1 – Setup

Setup code can conveniently be placed near the beginning of the main file (after `begin{document}`. It can also be placed at the beginning of any **.Rnw** files that are included . The chunk options can all be changed in the calls to individual code chunks. Here, the setup code was:

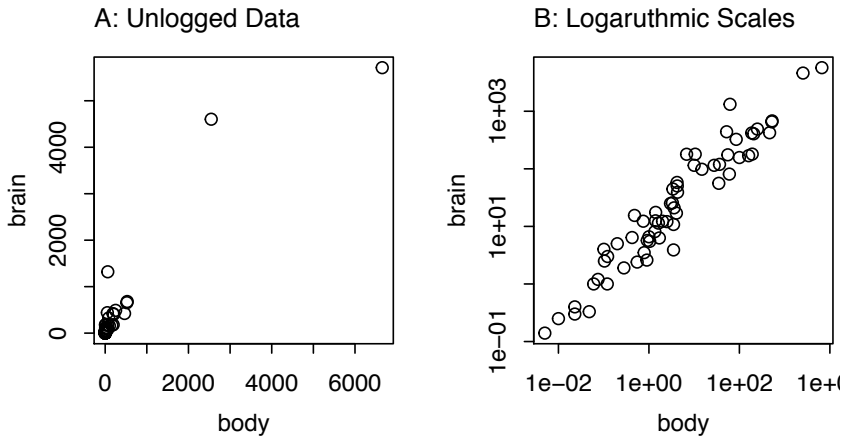

Figure 2: Brain weight (g) versus Body weight (kg), for 62 species of mammal. Panel A shows the unlogged data, while Panel B uses log scales, for both axes. Notice that the scales are labeled in the original (unlogged) units.

```
<<setup, echo=FALSE, results="hide", message=FALSE>>=
options(replace.assign=TRUE,width=90)
opts_chunk$set(fig.path='fig/demo-', cache.path='cache/manual-',
               fig.align='center', dev='pdf', tidy=FALSE,
               fig.width=2.75, fig.height=2.95,
               out.width="0.47\\texwidth", fig.show='hold',
               cache=TRUE, par=TRUE)
knit_hooks$set(par=function(before, options, envir){
if (before && options$fig.show!='none') par(mar=c(4,4,1.6,.2),
              cex=1, cex.lab=.95,cex.axis=.9,mgp=c(2,.7,0),
              tcl=-.3)
}, crop=hook_pdfcrop)
@
```

The chunk options apply to all later code chunks where there is no explicit setting for that option.

Code with markup used to load em knitr and create the LaTeX file that was then used to create this file was:

```
<<knitr-run, eval=FALSE, echo=TRUE>>=
library(knitr)
knit('knitDemo.Rnw')
@
```

Section 2 – Printing and Executing Code

The code that appears in the file, with markup added is:

```
<<cmdline, echo=TRUE, eval=TRUE, tidy=FALSE>>=
2*3*4*5          # * denotes 'multiply'
sqrt(10)         # the square root of 10
pi               # R knows about pi
2*pi*6378        # Circumference of earth at equator (km)
                 # (radius at equator is 6378 km)
@
```

The markup lines that precede and end the code have the form:

```
<<cmdline, ...>>=
@
```

Here `cmdline` is the name for this code chunk, amd `...` is replaced by chunk options. The code chunk ends with an `@` symbol.

Section 3 – Plotting a Graph

Code, with enclosing markup, is:

```
\begin{figure}
<<graph01, eval=TRUE, echo=FALSE, out.width="0.47\\textwidth">>=
Year <- c(1800, 1850, 1900, 1950, 2000)
Carbon <- c(8, 54, 534, 1630, 6611)
## Now plot Carbon as a function of Year
plot(Carbon ~ Year, pch=16)
@
\caption{Annual totals of carbon emissions (millions of tonnes),
against year}
\end{figure}
```

The option `out.width` specifies the width that the graph should occupy in the final document. Notice that it is specified as a text string. This is done so that it can be inserted as text into the LaTeX **.tex** document that is generated.

The markup code that caused the printing of the code for the graph was:

```
<<graph1-code, ref.label="graph1", eval=FALSE, echo=TRUE>>=
@
```

Notice the use of `ref.label="graph1"` to identify the code chunk that is printed.

Section 4 – Graphs Shown Side by Side

Here, the code chunk used is printed and stored under the name `mammals`, but not evaluated at this point.

```
<<mammals, eval=TRUE, echo=TRUE>>=
## Panel A
plot(brain ~ body, data=mammals, pty="s")
mtext(side=3, line=0.75, "A: Unlogged Data", adj=0)
## Panel B
plot(brain ~ body, data=mammals, log="xy", pty="s")
mtext(side=3, line=0.75, "B: Logaruthmic Scales", adj=0)
@ %
```

The code used to create the graph is then:

```
\begin{figure}
<<mammals-gph, ref.label="mammals", echo=FALSE>>=
@ %

\caption{Brain weight (g) versus Body weight (kg), for
  62 species of mammal. Panel A shows the unlogged data,
  while  Panel B uses log scales, for both
  axes. Notice that the scales are labeled
  in the original (unlogged) units.\label{fig:Animal}}
@
\end{figure}
```

Use of `ref.label="mammals"`, in the code chunk `mammals-gph`, has the effect of inserting the code at that point. As the default is `eval=TRUE`, it is evaluated and the graphs are printed.

The setup code `out.width="0.47\\texwidth"` in Section 1 allowed room for two plots side by side. As the code chunk that created Figure 2 made no change from this initial setup, the LaTeX file that is created accordingly places the plots side by side. Specify, e.g., `out.width="0.32\\texwidth"`, to allow three graphs side by side across the page.

7 Links (knitr more generally)

Elegant, flexible and fast dynamic report generation with R (Yihui Xie): `http://yihui.name/knitr/`

Slides obtained with knitr (Yihui Xie): `https://github.com/yihui/knitr-examples/blob/master/009-slides.md`

Getting started with R Markdown, knitr, and Rstudio 0.96 (Jeromy Anglim): https://jeromyanglim.blogspot.com.au/2012/05/getting-started-with-r-markdown-knitr.html

Ipython replication of Reinhart-Rogoff (Vincent Arel-Bundock): http://nbviewer.ipython.org/5409848