R Markdown in Education

Andrew Duncan

***Address:*** *Inverness College UHI, 1 Inverness Campus, Inverness, IV2 5NA*

***Email:***[*andrew.duncan.ic@uhi.ac.uk*](mailto:andrew.duncan.ic@uhi.ac.uk)

# Introduction

Increasingly in recent times there has been a change of focus in teaching statistics in schools, colleges and universities. There is a movement toward using software and teaching students using data and case studies (Baumer et al. ([2014](#ref-Baumer2014)), Bray, Çetinkaya–Rundel, and Stangl ([2014](#ref-Bray2014)), Hicks and Irizarry ([2016](#ref-Hicks2016)), Wang, Rush, and Horton ([2017](#ref-Wang2017)), Wilson et al. ([2017](#ref-Wilson2016))). This change is also apparent in the introduction of a new statistics unit at SCQF level 6 by the Scottish Qualifications Authority, developed with input from the University of Strathclyde (SQA ([2015](#ref-SQA_level6))). This unit requires students to complete linear regression and a basic hypothesis test using computer software and several examples of possible software options are provided in the unit support notes.

Alongside this change in the teaching of statistics is an increasing desire in the scientific community for reproducible research. This is highlighted by the articles collated by Nature (Nature ([2017](#ref-Nature_repro_special))) and several books that have been published on the topic (for example - Stodden, Leisch, and Peng ([2014](#ref-Stodden2014))).

One of the tools that can be used to satisfy both the change in statistics teaching whilst encouraging reproducible research is the statistical software R (R Core Team ([2017](#ref-baseR))). R is an increasingly popular tool (Muenchen ([2017](#ref-Muenchen2017))) and widely taught in Universities (Baumer et al. ([2014](#ref-Baumer2014))). The integrated development environment RStudio (RStudio Team ([2016](#ref-RStudio))) can be used to help with reproducibility (Gandrud ([2015](#ref-Gandrud2015))) and personal experience has shown that it can help when introducing students to R.

One of the benefits of R is that the functionality provided by the normal "base" install can be extended by packages. These provide additional capabilities, are written by other users and like R itself are open source. One such package is the R Markdown package (Allaire et al. ([2017](#ref-rmarkdown_package))). This is one of the tools that can assist with reproducible research as it allows for the code needed for calculations and plots (and the results of these) to be integrated alongside plain text. This avoids the problem where the calculation of results is separated from the interpretation. The R Markdown package allows users to convert (render/knit[[1]](#footnote-23)) R Markdown documents into several other possible forms including HTML, Word or PDF[[2]](#footnote-24).

An R Markdown document is based on the Markdown markup language (Gruber ([2004](#ref-markdown))) which takes a plain text document and produces a HTML file. R Markdown uses the same syntax but allows users to include R code, use LaTeX syntax (LaTeX3 Team ([2017](#ref-LaTeX))) to create mathematical equations and [BibTeX](http://www.bibtex.org/) to add references.

This article will outline some of the advantages and associated challenges in using R Markdown, and its extensions, in teaching statistics and data analysis.

# Benefits of using R Markdown in the classroom

One of the major benefits of using a system like R Markdown is that a student's work is contained within a single document. It is no longer necessary to create graphs separately and import them into their write up. This applies to any subject where data analysis is presented, it need not be limited to statistics. If the data is in an Excel spreadsheet (or similar) it can be imported into R and the plots created and displayed to the students as they continue to edit their report as seen in figure 1.

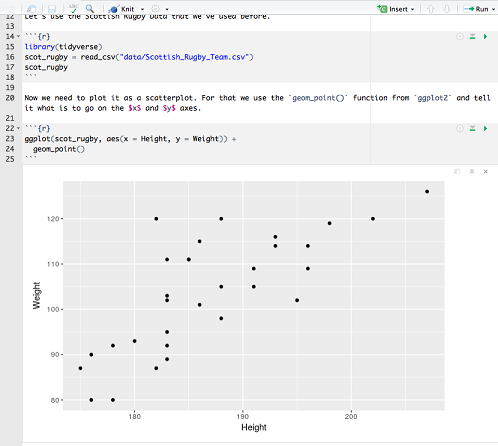


Figure 1 An example of a plot appearing beside both the code used to produce it and plain text. In this case it is taken from student notes rather than a submission.

As mentioned earlier students can easily produce HTML files from R Markdown and there are several avenues through which these might be published online for free[[3]](#footnote-28). This can help the students take ownership of the work (Bray, Çetinkaya–Rundel, and Stangl ([2014](#ref-Bray2014))). R Markdown can also be used to build HTML presentations, an entire website (an example can be found at <http://rmarkdown.rstudio.com/rmarkdown_websites.html>) or a blog (Xie ([2017c](#ref-blogdown_package)), Xie, Thomas, and Hill ([2017](#ref-blogdown_book))).

By learning R and R Markdown students are learning an increasingly popular computing language. Not only is it increasingly taught at Universities but it is on the rise as an industry requested skill for employees (Muenchen ([2017](#ref-Muenchen2017))).

# Benefits of using R Markdown as an educator

## In Assessment or Submission

Using R Markdown as a teacher/lecturer has several advantages compared to using, for example, Word and Excel, including for assessed or submitted work. If the student also submits the R Markdown document (and any associated data files if required) then, as when the student constructed it, the code used to create the results is right in the middle of the analysis. This makes marking easier and you can check whether the document submitted was produced using the corresponding R Markdown document. Everything should be reproducible (Baumer et al. ([2014](#ref-Baumer2014))).

As the R Markdown document can produce a Word or PDF document, this can be passed onto other members of staff to make second marking easier. They need not necessarily consider the R Markdown document itself.

## In Preparation of Materials

Depending on whether the reader has used LaTeX previously they may already have an opinion on the Word equation editor (or similar) and how easy it is to insert (and reference) equations with it, compared to producing and referencing equations in LaTeX. Alternatively LaTeX users may need to create, for whatever reason, a Word document, rather than a PDF. R Markdown allows the use of LaTeX equations whilst creating Word documents and for referencing we can call on the bookdown package (Xie ([2016a](#ref-bookdown_book)), Xie ([2016b](#ref-bookdown_package))).

The bookdown package extends R Markdown and as the name suggests, is designed to create books. By default the package will create an HTML book from a series of R Markdown documents. It can also create PDFs or several eBook formats, either in conjunction with the HTML (i.e. for download) or in isolation (Xie [2016a](#ref-bookdown_book)). As with a single R Markdown file, if the output format is to be a PDF then a LaTeX distribution will need to be installed too.

Aside from creating books, the bookdown package increases some of the capability of R Markdown and can be used for individual R Markdown files, this includes adding the capabilities to reference figures (produced by R or inserted as an image), tables and equations[[4]](#footnote-36). Sections or chapters can also be referenced later in the document. To get started the author of the package has produced two templates. One is a basic book template, the other contains additional files to produce a PDF in the same format as the published guide to bookdown (Xie ([2017a](#ref-bookdown_demo)), Xie ([2017b](#ref-bookdown_demo_crc)), Xie ([2016a](#ref-bookdown_book))).

One intention of bookdown is to allow public sharing with collaborative creation and editing of books (Xie ([2016a](#ref-bookdown_book))). With that in mind several books have been created recently where either the publishers have allowed the online version to be available for free (Wickham and Grolemund ([2016](#ref-r4ds_book)), Silge and Robinson ([2017](#ref-tidytext_book))). There is also an introductory statistics book, available online, with the intention that it is reproduced and tweaked/extended by other teachers/lecturers (Ismay and Kim ([2017](#ref-moderndive_book))).

Two further advantages of using R, but not specifically R Markdown, is the availability of data already nicely formatted and normally available as a package. Examples include the gapminder and fivethirtyeight packages (Bryan ([2015](#ref-gapminder_package)), Ismay and Chun ([2017](#ref-538_package))). The community of R users can provide an enormous amount of help and a wide range of free resources already exist for both staff and students.

## Examples of current use

Baumer et al. ([2014](#ref-Baumer2014)) has examples of usage and student feedback from courses at Duke University and Smith College. Student feedback was encouraging and lack of prior knowledge of markdown did not hinder students. It is also used in many other University courses including [STAT 545A and 547M](http://stat545.com/) at the University of British Columbia.

# Using R Markdown

There are many suitable guides for the R Markdown but a few examples are included here to demonstrate the capabilities.

Most simple inline formatting is available. For example:

* \_italic\_ or \*italic\* produces *italic* or *italic*
* \_\_bold\_\_ or \*\*bold\*\* produces **bold** or **bold**
* R^2^ provides superscript R2 with R~2~ providing subscript R2
* ![](path/to/image) allows you to insert an image
* [link to R Markdown website](http://rmarkdown.rstudio.com) [link to R Markdown website](http://rmarkdown.rstudio.com) produces a hyperlink.

Ordered and unordered lists can also be produced but there must be a blank line between the previous paragraph and the list.

The LaTeX syntax (LaTeX3 Team ([2017](#ref-LaTeX))) needed to create mathematical expressions is the normal single dollar signs $x^2$ producing and two dollar symbols (e.g. $$e^{i\pi} + 1 = 0$$) producing an equation.

If the equation needs to be numbered (and referenced) then the standard \begin{equation} and \end{equation} can be used with a label added using (\#eq:label). This could produce something like equation (1)

which can be referenced using \@ref(eq:label).

Although it is not necessary to include R code into an R Markdown document, inline R code can be inserted using single backticks `r 2 + 2` and multiple lines using a "chunk" wrapped in three backticks. The chunk below will produce figure 2.

```{r example-plot, fig.cap = "Example output plot"}  
library(tidyverse)  
# create subset of data  
diamonds = sample\_n(diamonds, 10000)  
# plot carat vs price  
ggplot(diamonds, aes(x = carat, y = price)) +  
 geom\_point()  
```

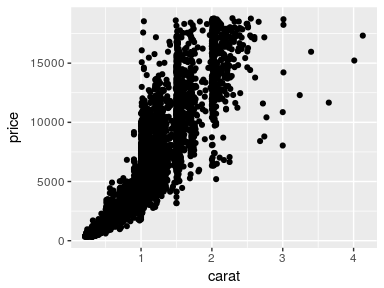


Figure 2 Example output plot

Certain elements of a document may render slightly differently in the different output formats (HTML, Word and PDF).

# Difficulties of using R and R Markdown

Using R and R Markdown is not without their difficulties too. They work best with both R and RStudio installed. Although RStudio is not strictly necessary the buttons available for render/knit and additional ease of use that RStudio provide students/staff with extra assistance. In addition to the two pieces of software every user will need access to the same packages. All of this needs to be in place before students can begin. RStudio make a server version of their software which is also open source and this may be of use to ensure that all students have the same packages installed.

There is a recommendation to use template R Markdown documents (Baumer et al. ([2014](#ref-Baumer2014))) which personal experience supports. It allows students to get started and some help can be given by including the start of an R chunk or examples of LaTeX equations.

If there is a mistake in a R Markdown document then an error will be reported when the user attempts to render/knit it. Sometimes these error messages are unclear and this is where either teacher/lecturer experience is necessary or the R community will be extremely helpful. Students can be encouraged to find the solution themselves by using questions that others have already asked on websites like [stackoverflow](https://stackoverflow.com/questions/tagged/rmarkdown).

# Conclusions

Reproducibility is seen as increasingly important in scientific research (Nature ([2017](#ref-Nature_repro_special))). R along with R Markdown can help students to gain understanding and knowledge of the importance (and advantages of) reproducibility whilst learning about tools used in modern data science (Baumer et al. ([2014](#ref-Baumer2014)), Muenchen ([2017](#ref-Muenchen2017))). With the extension of R Markdown into the bookdown package, the advantages of equation editing and referencing from LaTeX can now be leveraged in a simpler document format by both students and staff alike.

# Appendix

The R Markdown document, template and BibTeX file are available [online](https://github.com/aj2duncan/Rmd_Ed_Paper). To aide reproducibility the following setup was used to produce this article.

sessionInfo()

R version 3.2.5 (2016-04-14)  
Platform: x86\_64-pc-linux-gnu (64-bit)  
Running under: Ubuntu 14.04.5 LTS  
  
locale:  
 [1] LC\_CTYPE=en\_US.UTF-8 LC\_NUMERIC=C   
 [3] LC\_TIME=en\_US.UTF-8 LC\_COLLATE=en\_US.UTF-8   
 [5] LC\_MONETARY=en\_US.UTF-8 LC\_MESSAGES=en\_US.UTF-8   
 [7] LC\_PAPER=en\_US.UTF-8 LC\_NAME=C   
 [9] LC\_ADDRESS=C LC\_TELEPHONE=C   
[11] LC\_MEASUREMENT=en\_US.UTF-8 LC\_IDENTIFICATION=C   
  
attached base packages:  
[1] stats graphics grDevices utils datasets methods base   
  
other attached packages:  
[1] forcats\_0.2.0 stringr\_1.2.0 dplyr\_0.7.2   
[4] purrr\_0.2.3.9000 readr\_1.1.1 tidyr\_0.6.2   
[7] tibble\_1.3.3 ggplot2\_2.2.1.9000 tidyverse\_1.1.1.9000  
  
loaded via a namespace (and not attached):  
 [1] Rcpp\_0.12.12 cellranger\_1.1.0 highr\_0.6   
 [4] plyr\_1.8.4 bindr\_0.1 tools\_3.2.5   
 [7] digest\_0.6.12 lubridate\_1.6.0 jsonlite\_1.5   
[10] evaluate\_0.10.1 nlme\_3.1-127 gtable\_0.2.0   
[13] lattice\_0.20-33 pkgconfig\_2.0.1 png\_0.1-7   
[16] rlang\_0.1.1.9000 psych\_1.7.5 rstudioapi\_0.6   
[19] yaml\_2.1.14 parallel\_3.2.5 haven\_1.0.0   
[22] bindrcpp\_0.2 xml2\_1.1.1 httr\_1.2.1   
[25] knitr\_1.17 hms\_0.3 rprojroot\_1.2   
[28] grid\_3.2.5 glue\_1.1.1 R6\_2.2.2   
[31] readxl\_1.0.0 foreign\_0.8-68 rmarkdown\_1.6   
[34] bookdown\_0.5.2 modelr\_0.1.0 reshape2\_1.4.2   
[37] magrittr\_1.5 backports\_1.1.0 scales\_0.4.1.9002   
[40] htmltools\_0.3.6 rvest\_0.3.2 assertthat\_0.2.0   
[43] mnormt\_1.5-5 colorspace\_1.3-2 labeling\_0.3   
[46] stringi\_1.1.5 lazyeval\_0.2.0.9000 munsell\_0.4.3   
[49] broom\_0.4.2

# References

Allaire, JJ, Joe Cheng, Yihui Xie, Jonathan McPherson, Winston Chang, Jeff Allen, Hadley Wickham, Aron Atkins, Rob Hyndman, and Ruben Arslan. 2017. *Rmarkdown: Dynamic Documents for R*. <https://CRAN.R-project.org/package=rmarkdown>.

Baumer, Ben, Mine Çetinkaya-Rundel, Andrew Bray, Linda Loi, and Nicholas J. Horton. 2014. “R Markdown: Integrating a Reproducible Analysis Tool into Introductory Statistics.” *Technology Innovations in Statistics Education* 8 (1). <http://escholarship.org/uc/item/90b2f5xh>.

Bray, Andrew, Mine Çetinkaya–Rundel, and Dalene Stangl. 2014. “Five Concrete Reasons Your Students Should Be Learning to Analyze Data in the Reproducible Paradigm.” *CHANCE Columns,Taking a Chance in the Classroom*. <http://chance.amstat.org/2014/09/reproducible-paradigm/>.

Bryan, Jennifer. 2015. *Gapminder: Data from Gapminder*. <https://CRAN.R-project.org/package=gapminder>.

Gandrud, Christopher. 2015. *Reproducible Research with R and R Studio*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC.

Gruber, John. 2004. <https://daringfireball.net/projects/markdown/>.

Hicks, Stephanie C., and Rafael A. Irizarry. 2016. “A Guide to Teaching Data Science.” *ArXiv E-Prints*, December.

Ismay, Chester, and Jennifer Chun. 2017. *Fivethirtyeight: Data and Code Behind the Stories and Interactives at ’Fivethirtyeight’*. <https://CRAN.R-project.org/package=fivethirtyeight>.

Ismay, Chester, and Albert Y. Kim. 2017. *ModernDive: An Introduction to Statistical and Data Sciences via R*. <https://ismayc.github.io/moderndiver-book/>.

LaTeX3 Team. 2017. *The Latex Project*. <https://www.latex-project.org>.

Muenchen, Robert A. 2017. “The Popularity of Data Science Software.” <http://r4stats.com/articles/popularity/>.

Nature. 2017. “Challenges in Irreproducible Research.” *Nature Specials and Supplements Archive*. <http://www.nature.com/news/reproducibility-1.17552>.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

RStudio Team. 2016. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>.

Silge, Julia, and David Robinson. 2017. *Text Mining with R: A Tidy Approach*. O’Reilly Media. <http://tidytextmining.com/>.

SQA. 2015. *Statistics Award (SCQF Level 6) Unit Specification*. <http://www.sqa.org.uk/sqa/files_ccc/UnitStatisticsSCQF6.pdf>.

Stodden, Victoria, Friedrich Leisch, and Roger D. Peng. 2014. *Implementing Reproducible Research*. 1st ed. Boca Raton, Florida: Chapman; Hall/CRC.

Wang, Xiaofei, Cynthia Rush, and Nicholas J. Horton. 2017. “Data Visualization on Day One: Bringing Big Ideas into Intro Stats Early and Often.” *Technology Innovations in Statistics Education* 10 (1). <http://escholarship.org/uc/item/84v3774z>.

Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science*. O’Reilly Media. <http://r4ds.had.co.nz/>.

Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K. Teal. 2017. “Good Enough Practices in Scientific Computing” 13 (6). Public Library of Science: 1–20. doi:[10.1371/journal.pcbi.1005510](https://doi.org/10.1371/journal.pcbi.1005510).

Xie, Yihui. 2016a. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Boca Raton, Florida: Chapman; Hall/CRC. <https://github.com/rstudio/bookdown>.

———. 2016b. *Bookdown: Authoring Books and Technical Documents with R Markdown*. <https://github.com/rstudio/bookdown>.

———. 2016c. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <http://yihui.name/knitr/>.

———. 2017a. *A Minimal Book Example Using Bookdown*. <https://github.com/rstudio/bookdown-demo>.

———. 2017b. *A Minimal Example of Using Bookdown to Write a Book for Chapman & Hall*. <https://github.com/yihui/bookdown-crc>.

———. 2017c. *Blogdown: Create Blogs and Websites with R Markdown*. <https://github.com/rstudio/blogdown>.

Xie, Yihui, Amber Thomas, and Alison Presmanes Hill. 2017. *Blogdown: Creating Websites with R Markdown*. <https://bookdown.org/yihui/blogdown/>.

1. **knit** because the knitr package (Xie ([2016c](#ref-knitr_package))) is used in the process of creating the documents and RStudio provides users with a button to render/knit the document. [↑](#footnote-ref-23)
2. To create a PDF file users need to have a LaTeX distribution installed. [↑](#footnote-ref-24)
3. Examples include <https://rpubs.com>, <https://pages.github.com> and <https://www.netlify.com>. For students the easiest of these would be RPubs as everything can be done with no additional software and by pushing a button in RStudio. [↑](#footnote-ref-28)
4. Equation referencing works very well in PDF/HTML output but is a little fiddly in Microsoft Word. As seen in equation (1), it is perfectly possible though. [↑](#footnote-ref-36)