

# A reproducible journal article with RMarkdown: A building science template with examples for beginners based on the Elsevier journal format

Paul Raftery<sup>\*,a</sup>, Dana Miller<sup>a</sup>, Next contributor<sup>1</sup>

<sup>a</sup>*Center for the Built Environment, UC Berkeley, 390 Wurster Hall, Berkeley, CA, 94720, USA*

<sup>b</sup>*Another organization, and their address*

## Abstract

Enter the text for your abstract here. All data and analysis code in this template is publicly available at <https://github.com/CenterForTheBuiltEnvironment/rmd-example> and can be freely adapted and reused. Suggestions or pull requests to improve this template or create additional templates are welcome.

Keywords:

Keyword 1; Keyword 2; Keyword 3; Keyword 4; Keyword 5; Keyword 6 (max)

---

\*Corresponding Author

*Email address:* `p.raftery@berkeley.edu` (Paul Raftery)

**Highlights:**

- Add your 3 -5 highlights
- As bulletpoints here
- Making sure each is less than 85 characters in length

**Graphical Abstract**

L<sup>A</sup>T<sub>E</sub>X

## 1. Introduction

The purpose of this (very much work-in-progress) document is to provide a complete R markdown template for an Elsevier journal submission (based on the rrticles repository Allaire et al. (2017)), along with useful examples and packages to improve usability for folks who are just starting out with this workflow. The eventual intent is to capture minimal examples of the common things that authors need to do when writing papers in R markdown; provide examples of useful packages, workflows, and tools; and provide solutions to common issues that folks encounter.

You may wish to have short sub-sections for:

- Terminology
- Objectives

## 2. Methods (this is a ‘level 1 heading’)

### 2.1. Markdown (this is a ‘level 2 heading’)

This RMarkdown document is a special type of interactive document that can contain both code chunks (in R, python, or other languages), and text written in a format called Markdown. Here are some examples of how to format text with Markdown, and a [link](#) to an RMarkdown guide.

#### 2.1.1. Level 3 heading

Here’s how to **bold** or *italicize* a piece of text. This is how you do a bullet point list:

- First bullet
- Second bullet
  - A sub-bullet
  - Another sub-bullet

Or an ordered option:

1. Item 1
2. Item 2
  - Item 2a
  - Item 2b

## 2.2. Tables

Here is an example of a table created from the .csv file in the `SupplementaryMaterial/Data` folder using the `gt` package from RStudio. You can read more about `gt` [here](#), and there's helpful examples to further customize tables (eg with color, footnotes, significant figures, re-labeling columns, and more) [here](#).

Example irises			
Data on five randomly selected irises			
iris_class	sepal_length	petal_length	petal_length_level
Iris-virginica	6.4	5.5	High
Iris-virginica	6.9	5.4	High
Iris-setosa	4.4	1.3	VeryLow
Iris-virginica	7.7	6.7	VeryHigh
Iris-versicolor	5.0	3.3	Medium

## 2.3. Cross-referencing

This is how you refer to a figure in your text: Figure [1](#). Simply reference the title of the code chunk, and ensure that the code chunk includes a figure caption.

## 2.4. Calculations in text

The holy grail of markdown - doing all of your calculations in the same file so you never need to worry about updating them after someone<sup>1</sup> asks you to make changes. . . . again! It's as easy as pi: 3.14. Incidentally, you can selectively override the 'global' options set at the beginning, to say for example, show more decimals: 3.1416.

This is an example of outputting the result of a calculation that you perform within a code chunk in the document somewhere prior to the location where you first refer to it: 36.

## 2.5. Citations

Citing other literature is remarkably easy, just like this Coakley, Raftery, & Keane ([2014](#)). This citation key references the tag associated with an entry in Bibliography.bib (a BibTeX file). I've found it easiest to use Zotero to manage my library of references and to generate the BibTeX file, though any software that creates a

---

<sup>1</sup>Often I'm the someone, sorry CBE grad students. Also, look, it's an example of a footnote!

valid BibTeX file should work fine. Zotero allows you to create a ‘Collection’ (or folder) that gathers together all of the references used for a particular document. When combined with the Better BibTeX plugin, that collection can be exported to a BibTeX file that is continually updated as you add or edit references in that Collection. Better BibTeX also puts the citation key - the text after the ‘@’ symbol in the .Rmd file - on the upper right of each entry, which is convenient for adding citations.

There’s not much else involved in citing, as the references list gets built and formatted automatically based on the selected style. The only other issue I’ve had to look around to solve was figuring out how to combine multiple citations, which is easy when you know how. (Coakley et al., 2014; Zhai et al., 2015)

The instructions for installing BetterBibtex are [here](#).

### *2.5.1. Citing software*

It is also good practice to include citations not such for literature like journal articles and reports, but also the software used in the analysis. Many software packages have accompanying publications that can be cited, and even if they don’t it is important to attribute which version and which software you used. Citing software helps to provide credit for the creators and maintainors of the software and demonstrate that people (like you!) are using the software, which is especially important for open-source tools supported through public research funding.

You can use the **greatful** package to quickly generate a list of the packages for software used in an R project using the `cite_packages` function from **greatful**. There is an additional function called `get_citations` that generates the Bibtex-formatted citations for each package. It’s not elegant, but you can use the first to list packages used, then generate the `pkg-refs.bib` file (saved to `/Paper/SupplementaryMaterial` by default in this template), copy the content into your overall `Bibliography.bib` file for the whole document, and then cite the packages in your document as you would another source, eg by using `@base` for an in-text citation for core R software before other packages are installed (also known as ‘base R’).

- base R - R Core Team (2018)
- ggpmisc - Aphalo (2016)
- grateful - Rodriguez-Sanchez (2018)
- gt - Iannone, Cheng, & Schloerke (2019)
- here - Müller (2017)
- knitr - Xie (2018)
- tidyverse - Wickham (2017)

- rticles - Allaire et al. (2017)
- bookdown - Xie (2016)
- rmarkdown - Allaire et al. (2018)

## 2.6. Equations and math

Here’s a basic example inline  $example_{\text{subscript}} = \frac{D}{R}$ , or you display it on a whole line if needed. Google latex math cheat sheets for more information.

$$\sum_{i=1}^n x_i^2$$

Here is another equation:

$$CD_{\text{rated}} = \frac{4 * Q}{\pi * D^2} = 2.0 \text{ m/s}$$

## 2.7. Other packages

There are lots of packages that are useful for markdown docs and customizing plots. We encourage you to search for these whenever you encounter a new thing you need to do and to propose an addition to this repository accordingly. Some examples to start: `ggExtra`, `gridExtra`, `RColorBrewer`, `ggrepel`...

## 2.8. Writing style

This is a little off topic for an Rmd example but a convenient place to remind our grad students about writing style. In almost all cases, active voice is better than passive voice. Several psychological studies show that the active voice is more easily understood by readers, and that information is more accurately reported by authors when writing in active voice. For example, research Klenbort & Anisfeld (1974) has shown that the “active [voice] offers a neutral structure for conveying information”. Authorship guides for highly regarded journals often indicate a preference for the active voice instead of passive:

- Nature: “Nature journals like authors to write in the active voice (‘we performed the experiment...’) as experience has shown that readers find concepts and results to be conveyed more clearly if written directly.” (“Nature - How to write a paper,” n.d.)

- Science: “Use active voice when suitable, particularly when necessary for correct syntax (e.g., ‘To address this possibility, we constructed a lZap library . . .,’ not ‘To address this possibility, a lZap library was constructed. . .’).” Ruben, 2012, & Am (2012)

And, on top of all that, you also end up with less text if you write in active voice, saving space for useful information and making it easier for your readers to understand.

#### *2.8.1. A fun way to spot passive voice:*

If you can add the words ‘by zombies’ (“A scary-easy way to help you find passive voice!” 2014) to the end of the sentence and the sentence still makes logical sense, then the sentence is in passive voice. You can also switch on the grammar settings in Microsoft Word’s spelling and grammar checker and it will show up that way.

How to fix it?

Change:

“These measurements are not quantitatively reported in the paper” (... by zombies)

To

“The paper does not quantitatively report these measurements”

Or even better, it’s really the authors doing the reporting as the paper is an inanimate object. . .

“We do not quantitatively report these measurements”.

Change:

“Six different table and partition configurations were tested” (... by zombies)

To

“We tested six different table and partition configurations.”

### **3. Results**

#### *3.1. Scatter plot example*

Figure 2 shows petal widths by iris type.

### 3.2. Violin plot example

Figure 3 shows petal widths by iris type.

## 4. Discussion

## 5. Limitations

It is important to include a discussion of limitations in any paper. Some limitations of this template include:

- Whether or not figures are included inline with text appears to depend of the version of LaTeX the document is compiled with. Compiling on a computer where TinyTeX is installed (eg in the Binder associated with the repository for this project on GitHub) will render the figures inline, whereas compiling the same document on a Windows computer with MikTeX installed does not
- Does not include instructions on how to align figures

## 6. Conclusion

## 7. Acknowledgements

Don't forget to acknowledge the funder(s) with associated grant numbers if required. The same goes for folks who significantly assisted you with this paper but that are not authors. Eg "Agency (grant number 12345) supported this work, with cost share provided by the Center for the Built Environment. We thank Person1 and Person2 for assistance with data collection"

## 8. Declaration of interest

Describe any relevant interests of the authors, particularly if there is a link to the research that is relatively uncommon and could be perceived as a conflict of interest. Otherwise : All authors declare no conflict of interest.



## References

A scary-easy way to help you find passive voice! (2014). *A scary-easy way to help you find passive voice!* / *Grammarly Blog*. <https://www.grammarly.com/blog/a-scary-easy-way-to-help-you-find-passive-voice/>.

Allaire, J., R Foundation, Wickham, H., Journal of Statistical Software, Xie, Y., Vaidyanathan, R., ... Yu, M. (2017). *Rticles: Article formats for r markdown*. Retrieved from <https://CRAN.R-project.org/package=rticles>

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., ... Chang, W. (2018). *Rmarkdown: Dynamic documents for r*. Retrieved from <https://CRAN.R-project.org/package=rmarkdown>

Aphalo, P. J. (2016). *Learn r ...as you learnt your mother tongue*. Leanpub. Retrieved from <https://leanpub.com/learnr>

Coakley, D., Raftery, P., & Keane, M. (2014). A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, 37(Supplement C), 123–141. <https://doi.org/10.1016/j.rser.2014.05.007>

Iannone, R., Cheng, J., & Schloerke, B. (2019). *Gt: Easily create presentation-ready display tables*. Retrieved from <https://github.com/rstudio/gt>

Klenbort, I., & Anisfeld, M. (1974). Markedness and perspective in the interpretation of the active and passive voice. *Quarterly Journal of Experimental Psychology*, 26(2), 189–195. <https://doi.org/10.1080/14640747408400404>

Müller, K. (2017). *Here: A simpler way to find your files*. Retrieved from <https://CRAN.R-project.org/package=here>

Nature - How to write a paper. (n.d.). [https://www.nature.com/authors/author\\_resources/how\\_write.html](https://www.nature.com/authors/author_resources/how_write.html).

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Rodriguez-Sanchez, F. (2018). *Grateful: Facilitate citation of r packages*. Retrieved from <https://github.com/Pakillo/grateful>

Ruben, A., 2012, & Am, S. (2012). How to Write Like a Scientist. *Science / AAAS*. <https://www.sciencemag.org/careers/2012/03/write-scientist>.

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida:

Chapman; Hall/CRC. Retrieved from <https://github.com/rstudio/bookdown>

Xie, Y. (2018). *Knitr: A general-purpose package for dynamic report generation in r*. Retrieved from <https://yihui.name/knitr/>

Zhai, Y., Zhang, Y., Zhang, H., Pasut, W., Arens, E., & Meng, Q. (2015). Human comfort and perceived air quality in warm and humid environments with ceiling fans. *Building and Environment*, 90, 178–185. <https://doi.org/10.1016/j.buildenv.2015.04.003>

**List of Figures**

1	Correlation. Source: XKCD, <a href="http://xkcd.com/552">xkcd.com/552</a> . . . . .	12
2	Petal and sepal lengths from iris dataset . . . . .	13
3	Iris petal widths by species . . . . .	14

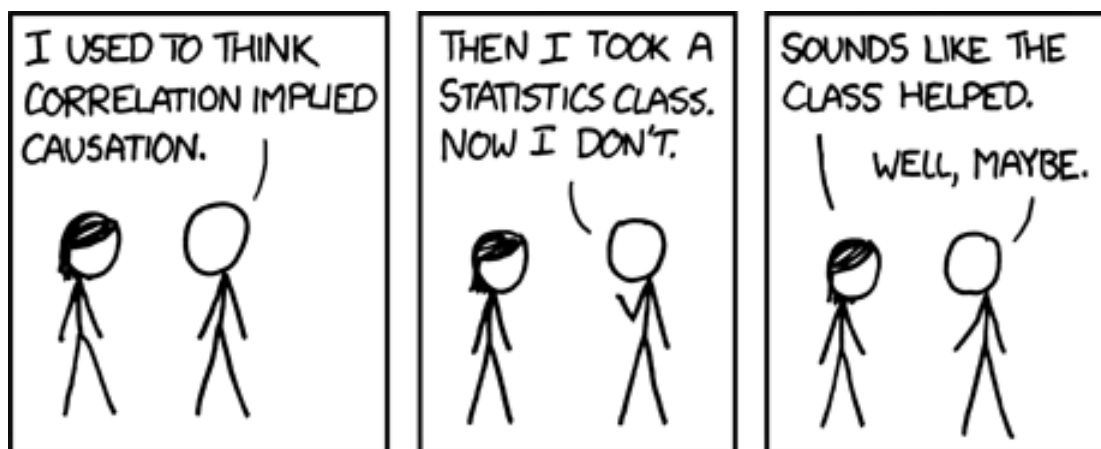


Figure 1: Correlation. Source: XKCD, [xkcd.com/552](http://xkcd.com/552)

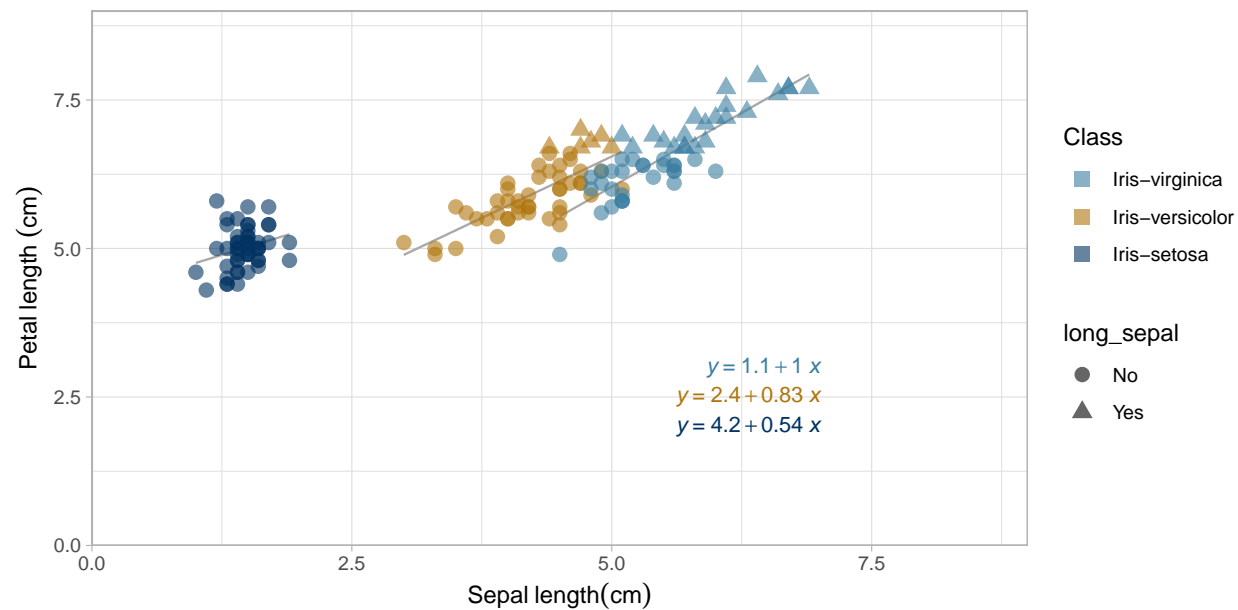


Figure 2: Petal and sepal lengths from iris dataset



Figure 3: Iris petal widths by species