

Reproducible Papers with RMarkdown

Michael Loranty, mloranty@colgate.edu

Heather Kropp, hkropp@colgate.edu

2019-01-18

Abstract

I really like using R (R Core Team 2015) for science because of tools like RStudio (RStudio Team 2015) and RMarkdown (RMarkdown Team 2015). This document is a quick demonstration of writing an academic paper in RMarkdown. There's a lot of other resources available on the web but hopefully you'll find this document useful as an example.

Introduction

Writing reports and academic papers is a ton of work but a large amount of that work can be spent doing monotonous tasks such as:

- Updating figures and tables as we refine our analysis
- Editing our analysis and, in turn, editing our paper's text
- Managing bibliography sections and in-text citations/references

These monotonous tasks are also highly error-prone. With RMarkdown, we can close the loop, so to speak, between our analysis and our manuscript because the manuscript can become the analysis.

As an alternative to Microsoft Word, RMarkdown provides some advantages:

- Free to use
- Uses text so we can:
- Use version control for
 - Tracking changes
 - Collaborating
- Edit it with our favorite and most powerful text editors
- Use the command line to for automation

The rest of this document will show how we get some of the features we need such as:

- Attractive typesetting for mathematics
- Figures, tables, and captions
- In-text citations
- Bibliographies

Methods

Our analysis will be pretty simple. We'll use the `diamonds` dataset from the `ggplot2` (Wickham 2009) package and run a simple linear model. At the top of this document, we started with a code chunk with `echo=FALSE` set as a chunk option so that we can load the `ggplot2` package and `diamonds` dataset without outputting anything to the screen.

For our analysis, we'll create a really great plot which really shows the relationship between price and carat and shows how we include plots in our document. Then we'll run a linear model of the form $y = mx + b$ on

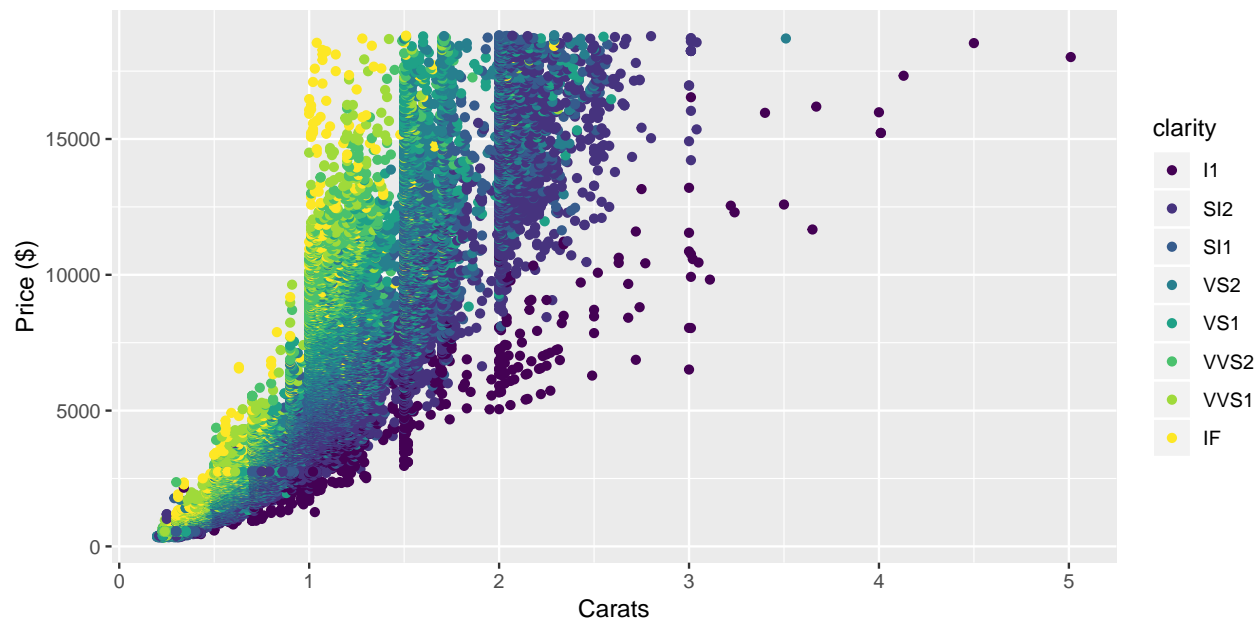


Figure 1: Figure 1: The relationship between price and carat for the diamonds dataset.

the relationship between price and carat and shows how we include tables in our document. We can also put some more advanced math in our paper and it will be beautifully typeset:

$$\sum_{i=1}^N \log(i) + \frac{\omega}{x}$$

We can also use R itself to generate bibliographic entries for the packages we use so we can give proper credit when we use other peoples' packages in our analysis. Here we cite the `ggplot2` package:

```
> citation('ggplot2')
```

To cite `ggplot2` in publications, please use:

H. Wickham. `ggplot2`: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

A BibTeX entry for LaTeX users is

```
@Book{,
  author = {Hadley Wickham},
  title = {ggplot2: Elegant Graphics for Data Analysis},
  publisher = {Springer-Verlag New York},
  year = {2009},
  isbn = {978-0-387-98140-6},
  url = {http://ggplot2.org},
}
```

And then we just place that in our `.bibtex` file. # Results

But the model was even better:

Table 1: Table 1: This is a broomed linear model summary table.

term	estimate	std.error	statistic	p.value
(Intercept)	-2256.36	13.06	-172.83	0
carat	7756.43	14.07	551.41	0

We were delighted to find that the slope parameter was 7756.43.

Discussion

This was just a quick demonstration of a reproducible paper that combined text, analysis, figures, tables, and citations into multiple output formats (HTML, PDF). Hopefully you found it useful.

A lot of people are using RMarkdown these days so there are tons of resources online but here are a few choice ones specifically about making papers:

- http://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html
- <http://svmiller.com/blog/2016/02/svm-r-markdown-manuscript/>
- <http://www.petrkeil.com/?p=2401>

References

- R Core Team. 2015. “R: A Language and Environment for Statistical Computing.” <http://www.r-project.org>.
- RMarkdown Team. 2015. *Rmarkdown: R Markdown Document Conversion, R Package*. Boston, MA: RStudio, Inc. <http://rmarkdown.rstudio.com/>.
- RStudio Team. 2015. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>.
- Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <http://ggplot2.org>.