Reproducible Analytical Science using R and Python

John Minter

2018-07-02

Contents

Chapter 1

Prologue

I spent a 36 year career in the Analytical Sciences Division of a large corporation. My main areas of focus were electron microscopy, image analysis, and electron microanalysis using X-ray energy-dispersive spectroscopy (EDS). I quickly learned that embracing automation and scripted analysis gave me more time to think about the meaning of the results I obtained and eliminated error-prone repetitive data processing.

By the end of my career, I had embraced Donald Knuth's idea of **Literate Programming**, where the code to run the data analysis is embedded as **code chunks** in the report, automatically generating the results. Like many scientists who work in analytical laboratories, I found myself repeatedly writing the same code. By that time I had embraced both **python** and **R**, and had learned the benefit of collecting code into functions and packages maintained under version control and tested with unit tests.

> If you write the same code more than twice, incorporate it into a function and keep it in a package under version control.

I also discovered that the **python** and **R-stats** communities were welcoming and had already solved many of the problems I needed to solve. There were packages that did much of what I needed to do and I could make my non-proprietary code available as a way of giving something back. My management supported these activities because most of the tools that I used were Open Source and we were always under severe budgetary constraints.

My new workflow was both more enjoyable and more intellectually stimulating. As our headcount went down, I found like-minded individuals in the larger community helped me to continue to learn and grow. This benefitted all of our employers without revealing proprietary information.

In this book I will show how to use these tools to make a more robust and reproducible data analysis.

Chapter 2

Introduction to R

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter 2. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (Xie, 2018) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015).

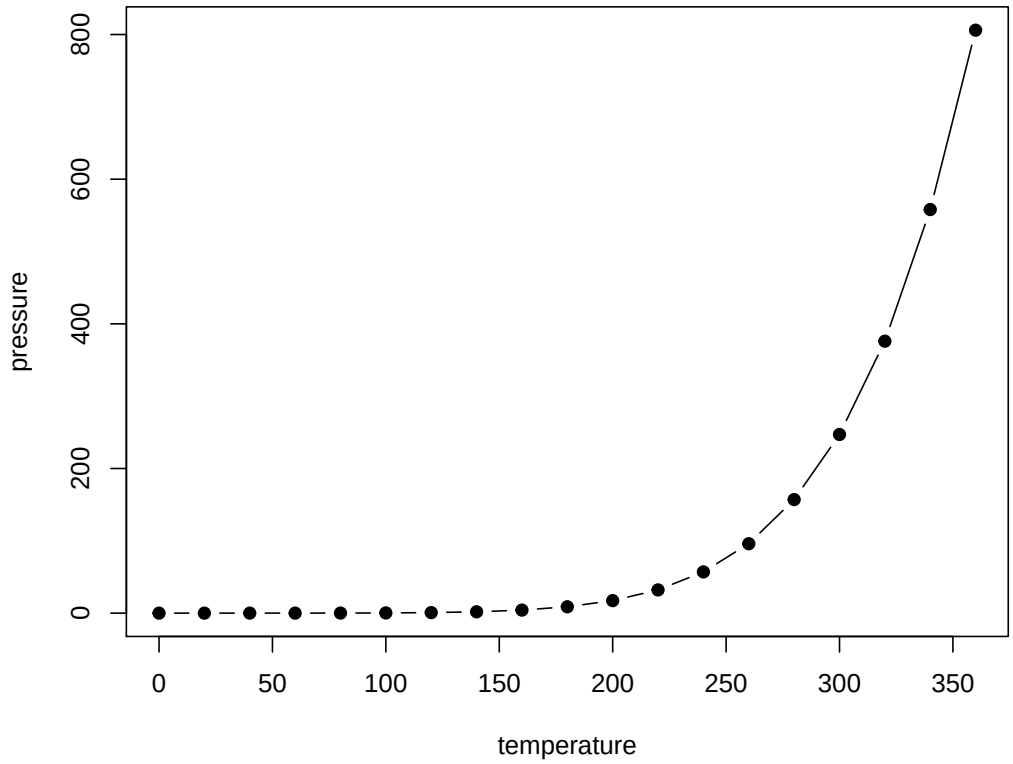

Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Chapter 3

Introduction to Python

We describe our methods in this chapter.

Chapter 4

Image Analysis with Python and Jython

Image analysis using scikit-image and Jython scripting with Fiji/ImageJ

Chapter 5

Image Analysis Data Reduction with R

Use R to do data analysis and reduction of feature vectors measured with Python and ImageJ/Fiji.

5.1 Example one

5.2 Example two

Chapter 6

Other Analytical Data Reduction with R

Use R to do data analysis and reduction non microscopy data.

6.1 Example one

6.2 Example two

Chapter 7

Final Words

We have finished a nice book.

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2018). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.7.10.