

Introduction to RMarkdown

Jonathan Gilligan

August 17, 2017

RMarkdown

RMarkdown is a combination of two tools: Markdown is a simple way of writing plain text that can be formatted into fancy documents. Markdown originated as a tool for easily writing blog posts and comments that include formatting, such as section headings, italic and bold-faced text, and so forth without having to learn a technical formatting language, such as HTML. Markdown is fairly generic, so it is easy to translate a single Markdown document into many output formats, such as HTML for web pages, PDF, and Microsoft Word documents.

You can learn a lot about the details of Markdown at the RStudio web site <http://rmarkdown.rstudio.com>. RStudio has a handy cheat sheet for RMarkdown that you can see by opening the Help menu, going to “Cheatsheets” and opening the Rmarkdown Cheat Sheet or the more comprehensive RMarkdown Reference.

What makes RMarkdown different from ordinary Markdown is that it allows you to combine text, formatted in Markdown, with instructions to the R statistical software to analyze data and produce graphs, tables, and other useful output.

By integrating the data analysis with the text of a document, we can easily make our research reproducible. The RMarkdown file contains all the instructions to load the data, analyze it, and generate the final report. Then if your data changes and you need to update your report, you can do so by knitting the RMarkdown document with the updated data files.

To turn an RMarkdown document into an HTML, PDF, or Microsoft Word document, you just click on the “Knit” button in RStudio. If you click on the word “Knit” on the button, RStudio will turn the RMarkdown document into the default format (see below). To knit the document into a different output format, click on the arrow just to the right of the word “Knit,” and select the output format you want.

Document header

At the top of an RMarkdown document there is a section set off by lines of three hyphens above it and below it.

Here is an example:

```
---
title: "Your document's title"
subtitle: "Your document's subtitle"
author: "Your name goes here"
date: "Aug 28, 2017"
output:
  pdf_document: default
  html_document: default
---
```

By default, when RStudio knits the document, it converts it into the format listed first under `output_format`

When the output format lists, for instance, “`pdf_format: default`”, that means that when RStudio creates a pdf document, it uses the default options. If you want to customize the output, you can specify options for RStudio to use in creating PDF, HTML, or Word output documents:

```
---
title: "Your document's title"
subtitle: "Your document's subtitle"
author: "Your name goes here"
date: "Aug 28, 2017"
output:
  pdf_document:
    toc: "true"
    number_sections: "false"
  html_document:
    self_contained: "false"
---
```

Markdown

Some of the basic elements of Markdown are:

Paragraphs

Any block of one or more lines of text, with a blank line before and a blank line after is treated as a single paragraph. To separate paragraphs, put a blank line between them:

```
This is one paragraph.
It stretches over several consecutive lines,
but it will be formatted
as a single paragraph.
```

```
This is another paragraph.
The blank line between the two
blocks of text tells Markdown
that they are separate paragraphs
```

```
    This is one paragraph. It stretches over several consecutive lines, but it will be formatted as a
    single paragraph.
```

```
    This is another paragraph. The blank line between the two blocks of text tells Markdown that
    they are separate paragraphs
```

Section headers

Any line of text that begins with one or more hash symbols (“#”) and is preceded by a blank line is treated as a section header. Top-level section headers have a single hash, and subsections, subsubsections, etc. use two, three, etc. hashes.

```
# This is a top-level section header
```

```
## This is a subsection
```

```
### This is a subsubsection
```

```
# This is another section
```

This is a top-level section header

This is a subsection

This is a subsubsection

This is another section

Formatting text

To make *italic text* and **boldface text** you surround the text with underscores or asterisks. A single underscore or asterisk means italic, two means boldface, and three means both italic and boldface:

This is *_italic text_*. This is **also italic text**. *__This is boldface__* and ***so is this***. ****This is bold italic****. This is *~~strikethrough~~*, perhaps to indicate an error.

This is *italic text*. This is *also italic*. **This is boldface** and **so is this**. ***This is bold italic***. This is ~~strikethrough~~, perhaps to indicate an error.

Lists

You can make bulleted or numbered lists easily in Markdown. Simply begin a line with an asterisk, hyphen, or plus sign. To make a sub-list, just indent the lines of the sublist by four spaces.

- * This is a list

- * This is the second item of the list.

- * This is a sub-list

- * This is another item in the sub-list

- A list item can have several paragraphs. Just indent the continuation by four additional spaces.

- Here is a third paragraph of continuation.

- * This is the main list again.

- This is a list

- This is the second item of the list.

- This is a sub-list

- This is another item in the sub-list

- A list item can have several paragraphs. Just indent the continuation by four additional spaces and do not begin it with an asterisk.

- Here is a third paragraph of continuation.

- This is the main list again.

To make numbered lists, start them with a number and a period:

1. This is a list

1. This is the second item of the list. Notice that I keep using the numeral "1", but Markdown automatically increments the numbers for the list items.

a) This is a sub-list

a) This is another item in the sub-list

A list item can have several paragraphs. Just indent the continuation by four additional spaces.

Here is a third paragraph of continuation.

i) this is a sub-sub list numbered with Roman numerals

i) this is a sub-sub list numbered with Roman numerals

i) this is a sub-sub list numbered with Roman numerals

1. This is the main list again.

1. This is a list

2. This is the second item of the list. Notice that I keep using the numeral "1", but Markdown automatically increments the numbers for the list items.

a) This is a sub-list

b) This is another item in the sub-list

A list item can have several paragraphs. Just indent the continuation by four additional spaces and do not begin it with an asterisk.

Here is a third paragraph of continuation.

i) this is a sub-sub list numbered with Roman numerals

ii) this is a sub-sub list numbered with Roman numerals

iii) this is a sub-sub list numbered with Roman numerals

3. This is the main list again.

Mathematical expressions

For simple math, you can just type stuff in RMarkdown: $1 + 2 * 3$ comes out as $1 + 2 * 3$. If you want subscripts or superscripts, you can get those by using the `~` and `^` characters, respectively: $I_{out} = \sigma T^4$ appears as $I_{out} = \sigma T^4$.

If you want fancier formatting for mathematical expressions, RMarkdown has a way to do this. However, it is complicated and I will **not** expect you to learn it for this class.

If you put an expression between dollar signs, RMarkdown interprets it differently from regular RMarkdown, and uses a format called LaTeX that is used for typesetting sophisticated mathematics. I won't try to present all of LaTeX here, but will give some common examples:

An expression between single dollar signs is interpreted as *in-line* math that appears in the middle of a line of text. LaTeX uses special terms that begin with a backslash `\` to indicate special mathematical formatting or operators. For example $\$a \times x + b\$$ appears as $a \times x + b$.



Figure 1: This is a tornado

LaTeX handles subscripts and superscripts differently from RMarkdown: You use `_{} and ^{}` with whatever you want to appear in the subscript or superscript inside the curly braces `{}`: `$I_{out} = \sigma T^4$` appears as $I_{out} = \sigma T^4$

You can include square root signs using `\sqrt{}`: `$x = \sqrt{y \times z}$` appears as $x = \sqrt{y \times z}$. You can do other roots like this: `$T = \sqrt[4]{I_{out}} / \sigma$` appears as $T = \sqrt[4]{I_{out}} / \sigma$.

To get display math, which appears on its own line, you use double dollar signs. This is useful when you want to write a mathematical expression that is much taller than a line of text: `$$\varepsilon \sigma T^4 = \frac{(1 - \alpha)}{4} \times I_{solar}$$` appears as

$$\varepsilon \sigma T^4 = \frac{(1 - \alpha)}{4} \times I_{solar}$$

If this seems too painfully complicated, don't worry. I do not expect you to learn this mathematical notation and it will not be required for this course. I only mention it in case you are interested.

Figures

You can include image files in your RMarkdown document:

```
![This is a tornado](images/tornado.jpg)
```

Using R for calculations

RMarkdown combines the “Markdown” system for formatting text with the R statistical system for performing calculations and making graphs.

We are using R for the laboratory section of this course because it allows you to easily analyze large amounts of data and produce high-quality graphs.

To enter R expressions in an RMarkdown document, we use “code blocks” and “inline code”. Code blocks are useful if we are doing a calculation, and inline code is useful if you just want to insert a number (maybe one you have calculated in a code block) into the middle of a line of text.

Code blocks begin and end with three consecutive “back-tick” characters:

Inline code appears between single back-ticks: ``r code goes here``

We express basic mathematical operations in R similarly to the way we express them in Excel and most programming languages. Addition and subtraction use the normal plus and minus signs. Multiplication uses an asterisk, division uses a slash (“/”), and powers (exponentiation) uses “^”.

R has a lot of mathematical functions you can use, such as `sqrt()`, `log()` (for the natural logarithm), `log10()` for the base-10 logarithm, and so forth.

We can assign numbers to variables using the equals sign:

```
sigma = 5.67E-8
I_solar = 1350 # watts per square meter
albedo = 0.3
I_absorbed = I_solar * (1 - albedo)
T = (I_absorbed / (4 * sigma))^0.25
```

If the last line of a code block is something with a value (e.g., the name of a variable or a mathematical expression, as opposed to an assignment with an equals sign), then R will show that value:

```
T
## [1] 254.0664
```

We can also print the value of an R expression in the middle of a line of text using inline code: `T = `r T`` will give `T = 254.0663741`.

Inline code is very useful because it lets us ask R to automatically insert a number into the text. This means that every time we knit the document, that number is generated by R. If you write a report and then realize that there was a problem with the data or the analysis, you can just fix the problem and re-knit the report using the corrected data and analysis code. You don’t need to manually go through a separate document and edit the numbers to update them with the latest results from your analysis. RMarkdown will do that for you, if you used inline code to insert the numbers into your text.

Thus, a bit of extra work at the beginning to set up the document and analysis using RMarkdown saves lots of time later on by making it trivial to update the document.

Code blocks also allow us to include graphs in our document:

Loading R Scripts

In this course, I am trying to teach you to use the basics of R. Sometimes we will need to do things for our analysis that require more complicated R programming, and then I will write R scripts that contain this code so you don’t need to write it yourself from scratch, but can just call functions that I provide.

To load a script (a file containing R code), you use the R command `source("script_file.R")`

For the exercises here, we will use two scripts, which are located in the “scripts” directory.

To load them into R so you can use them in your analysis, you would put the following code into your RMarkdown document:

```
source("scripts/format_md.R")
source("scripts/layer_diagram.R")
```

These scripts contain the code for a function, `format_md`, which formats numbers nicely for markdown documents, allowing you to control how many decimal places it shows, and optionally using scientific notation or inserting commas to separate thousands, millions, etc.

`format_md(sigma, digits = 1)` will produce 5.7×10^{-8}

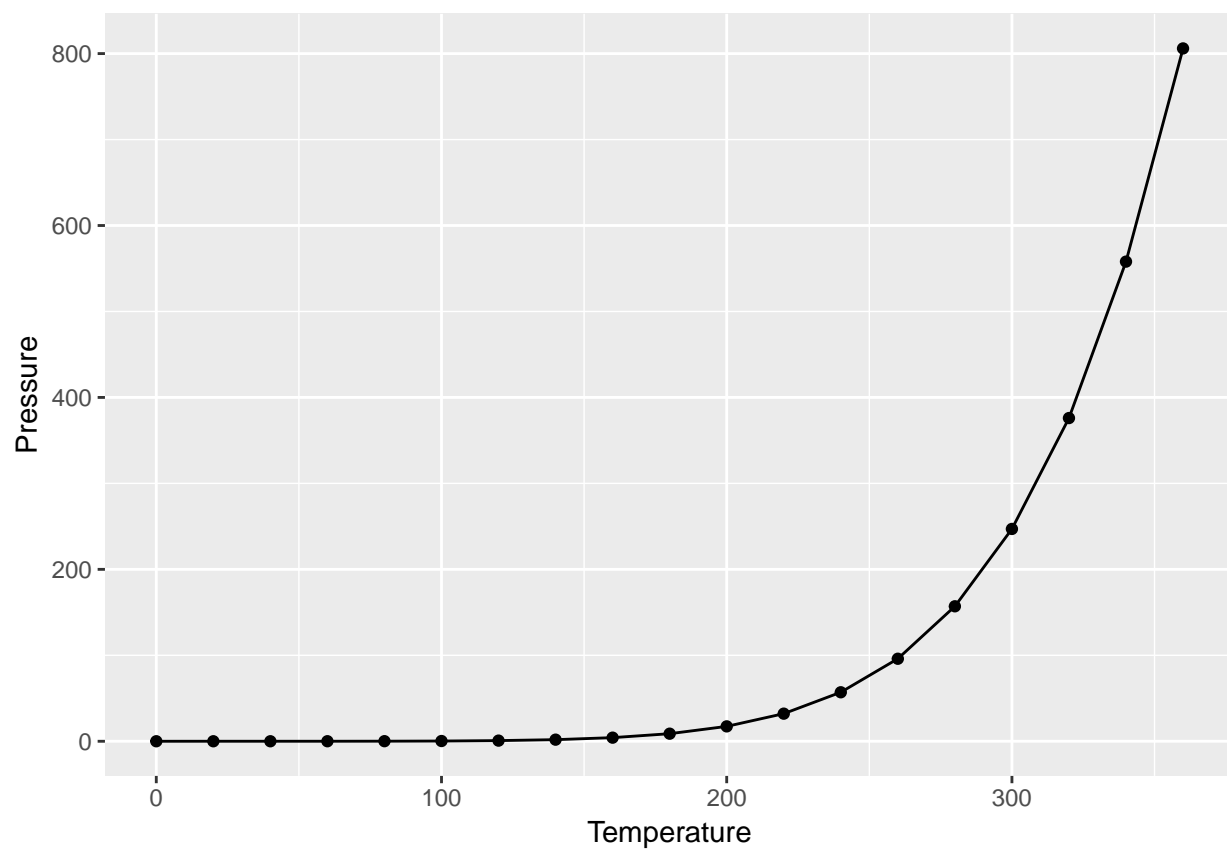


Figure 2: This is a plot of pressure versus temperature.

`format_md(sigma, digits = 1, scientific = FALSE)` will produce 0.00000006

`format_md(pi * 1E6, digits = 3, scientific = TRUE)` will produce 3.142×10^6

`format_md(pi * 1E6, digits = 3, scientific = FALSE, comma = TRUE)` will produce 3,142,000

The other script allows you to produce a layer diagram, similar to the ones in Chapter 3 of *Global Warming: Understanding the Forecast*.

`make_layer_diagram(2)`

