STAT 801- R Markdown

Tyler Wiederich

Fall 2022

R Markdown is a programming language used to create reports by making it seamless to integrate R code, visualizations, and user commentary. For example, you can use R Markdown for your projects and homework in this class. The benefit from R Markdown is that you will not need to copy code/output from R and paste it into Word. Instead, you can run one program that produces a PDF, Word, or HTML file. Having your entire work completed with R Markdown also ensures reproducibility of it. R Markdown files have a .Rmd extension and these are simply plain ASCII text files that can be viewed in notepad or similar editors.

# Installation

R Markdown is installed already in RStudio. When you run a program for the first time, RStudio may prompt you to install a few new packages or update existing packages. If you use R Markdown outside of RStudio, please follow the link <https://www.rdocumentation.org/packages/rmarkdown/versions/2.16> for installation details.

# Opening an R Markdown program

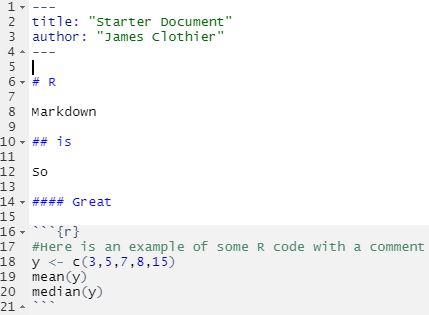
To start with a new R Markdown program, select File > New File > R Markdown… from the RStudio menu bar. To open an existing R Markdown program, select File > Open File.

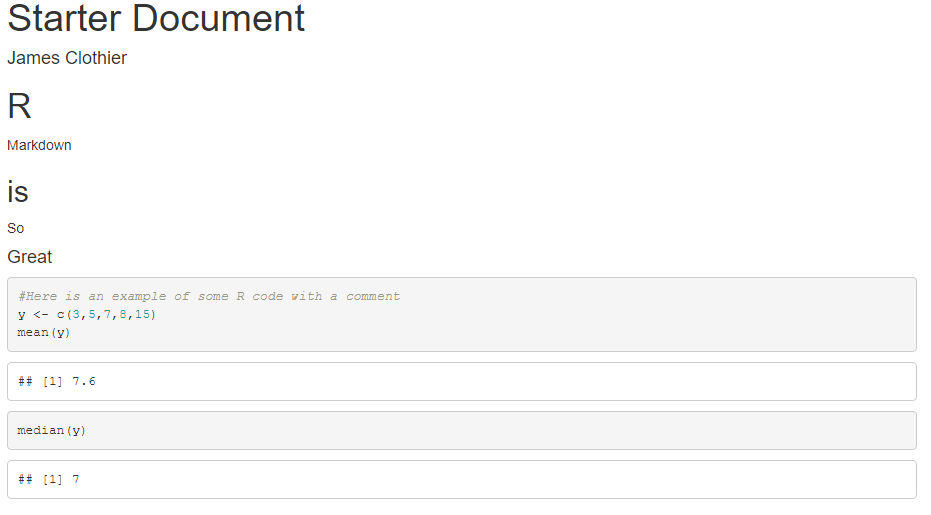
# Elements of R Markdown

There are three main elements of any R markdown file:

1. Metadata
2. Text for the report
3. R code that is run by the knitr package

We will go over each of these elements after examining a very simple “starter” program. Below is a screen capture of the corresponding R Markdown program (Starter.rmd) that creates an HTML file.





The write-up, code, and output are all merged into one document! The .rmd file contains our code and text. To run the program and “knit” together this code and text with the corresponding output, select the Knit dropdown menu.[[1]](#footnote-29) You can select “Knit to HTML”, “Knit to Word” or “Knit to PDF” to produce the final document types. When you use R Markdown to prepare documents for this class, you can submit Word or PDF files but not HTML files. The PDF file format is only available if you have LaTeX (a typesetting language and software package; <https://www.latex-project.org/get/#tex-distributions>) installed on your computer.

## Metadata

Metadata is what we put at the very beginning of the R markdown file. It is used to mention things such as title, date, and author.

## Text for the report

The text portion is whatever you want to explain in your report. There are no specific requirements for the text portion. You can use headers which use the # character (this is different from a “comment” within an R program). Headers with only one # will be the top level header. If you want to nest other headers within the top level headers you have to use more # characters.

## Code

R code is contained within *chunks* using a specific syntax. You can obtain a template structure by pressing Ctrl + Alt + I, selecting Insert > R, or by simply copying/pasting previous chunk code.

# Mathematical symbols

Mathematical symbols can be rendered using the same syntax as within LaTeX. If you do not have experience with LaTeX already, you may have used its same syntax when creating equations in Word’s equation editor. A reference for writing code for equations can be found at <http://www.stat.cmu.edu/~cshalizi/rmarkdown> under the heading “Math in R Markdown”. An introduction to LaTeX in general can be found on Chris Bilder’s website at <http://www.chrisbilder.com/stat850/sections.html> (see Section 2 and course videos on the schedule web page).

The sample mean has a designated symbol . Therefore, if you want to nicely format your result, you can use \bar{x} combined with $ symbols and write . Similarly you can create many other symbols such as , , and fractions like . If you write an equation within single $ symbol, the equation will appear inline with other text. If you want to display an equation in the center of the page, you need to enclose it in double $$ symbols. For example, the previous equation when enclosed within double $$ symbols will appear as follows.

# Text Formatting options

There are many text formatting options. A very detailed online companion on R markdown by R Studio which can be accessed at <https://rmarkdown.rstudio.com/lesson-1.html>. The R Markdown cheatsheet is at <https://rmarkdown.rstudio.com/docs/articles/rmarkdown.html>.

# Tables

Tables can be put in a markdown document using the kable() function in the knitr package. If we want to read in Lincoln\_Feb\_wind.csv (Section 2 notes) and display the first few lines of this data set as a table, we can do it as follows:

> wind <- read.csv("Lincoln\_Feb\_wind.csv")  
> kable(head(wind), caption = "A knitr table.")

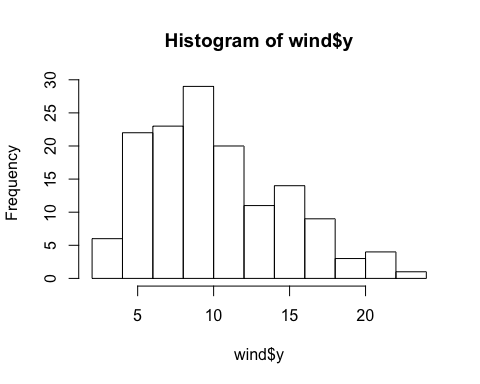
A knitr table.

| Year | Day | y |
| --- | --- | --- |
| 1 | 1 | 9.4 |
| 1 | 2 | 12.7 |
| 1 | 3 | 3.9 |
| 1 | 4 | 9.8 |
| 1 | 5 | 9.5 |
| 1 | 6 | 15.0 |

Your data file must be in the same folder as your .rmd file if you wish to use the code above.

# Embedding visualizations

Embedding visualizations in a markdown file is similar to embedding the R code. You need to embed the code generating the visualization in a chunk. For example, below is a histogram from the Section 2 notes, and we hide its corresponding R code (explained in the next section).



# Code Chunk Options

There are options that you can use within chunks. These options go in the {r} portion of the syntax. For example, suppose we want to keep a code chunk and the objects created from it but without showing either the code or any results from it. This can be done by using the option include = FALSE.

If you want to hide the code but display the output, you can use the option echo = FALSE.

[1] 3.175

[1] 3.15

A list of options is available at <https://yihui.org/knitr/options>.

Options can be set for all chunks at once by using a function. There is an odd syntax associated with it that involves embedding a function within a list so that opts\_chunk$set() is the function needed. Below is how you can use this function with a set of options recommended by Chris to use in this class.

> options(continue = " ")  
> opts\_chunk$set(prompt = TRUE, comment = NA, background = "white",  
 tidy = TRUE, tidy.opts = list(width.cutoff = 60, blank = FALSE))  
> y <- c(2.9, 3.4, 3.6, 2.8)  
> mean(y)

[1] 3.175

> median(y)

[1] 3.15

The appearance of the R output itself can be controlled using the options() function (this is from the base package). One option recommended by Chris is to remove the leading “+” from being displayed when an R command extends onto the next line.

# Inline R output

The report can be automated by using inline R code. Suppose you want to state the mean and the median calculated previously, but you do not want the normal output. You can use inline code for this specific purpose. For example, the mean is 3.175 and the median is 3.15 (see code in the .rmd file).

# Alternatives to R Markdown

There are a few alternatives to R Markdown that provide similar functionality. For example, LyX is a free software package that allows one to use a Word-like interface with LaTeX and embed R code within the document. Introductions to LyX are available on Chris Bilder’s STAT 850 website at <http://www.chrisbilder.com/stat850/sections.html> (see Section 2 and course videos on the schedule web page). In fact, all of Chris’s R notes for that course were created using the combination of LyX and R and are available for download. An introduction to including R within a LyX document is available in the knitr portion of the LyX section.

A second alternative is StatTag that was recently created by researchers at Northwestern University. This is a free plug-in for Word. Please see <https://sites.northwestern.edu/stattag> if you are interested in using it.

# Do you need to use R Markdown?

The use of R Markdown or similar alternatives is optional for students in STAT 801. We wanted to present these options to you because they can make your work easier to complete in this class. Also, trends in the creation of reports/papers that use R are going in the direction of using “dynamic document creation” tools like these. We understand though there is a start-up time cost associated with learning these tools. We recommend that you at least try these tools.

1. The knitr package is used by R Markdown for this purpose. The package was written by Yihui Xie as part of his dissertation in Statistics at Iowa State. Xie now works for Rstudio and lives in Omaha. [↑](#footnote-ref-29)