

# Reproducible workflows

To fully trust the results of the work we perform, we must be able to repeat it. The most credible and trustworthy work is reproducible.

The basis of keeping data science reproducible has to do with managing small tasks, like keeping the code executed in the right order, and starting from scratch when rerunning all your code to produce figures or tables.

**Humans** are good at creative tasks like writing, and designing a study, or working on a paper, or a report. **Computers** are much better at automating repetitive tasks, so we should let them do that type of work.

This step will focus on this way of working - getting the computer to do the small tasks which are difficult to remember, and leaving the person to do the writing. We call this process a '**reproducible workflow**'.

## What is R Markdown?

- R Markdown (.Rmd) is an authoring format that enables easy creation of documents, presentations, and reports from R. It uses the R packages `rmarkdown` and `knitr` to make it all work. All coding exercises that you complete for this course and throughout the microcredential need to be in R Markdown.
- It combines the core syntax of **markdown** (an easy-to-write plain text format) **with embedded R code chunks** that are run so their output can be included in the final document.
- R Markdown documents are fully reproducible (they can be automatically regenerated whenever underlying R code or data changes).
- An R Markdown file is a plain text file that has the extension `.Rmd`

## Tutorials with Quang

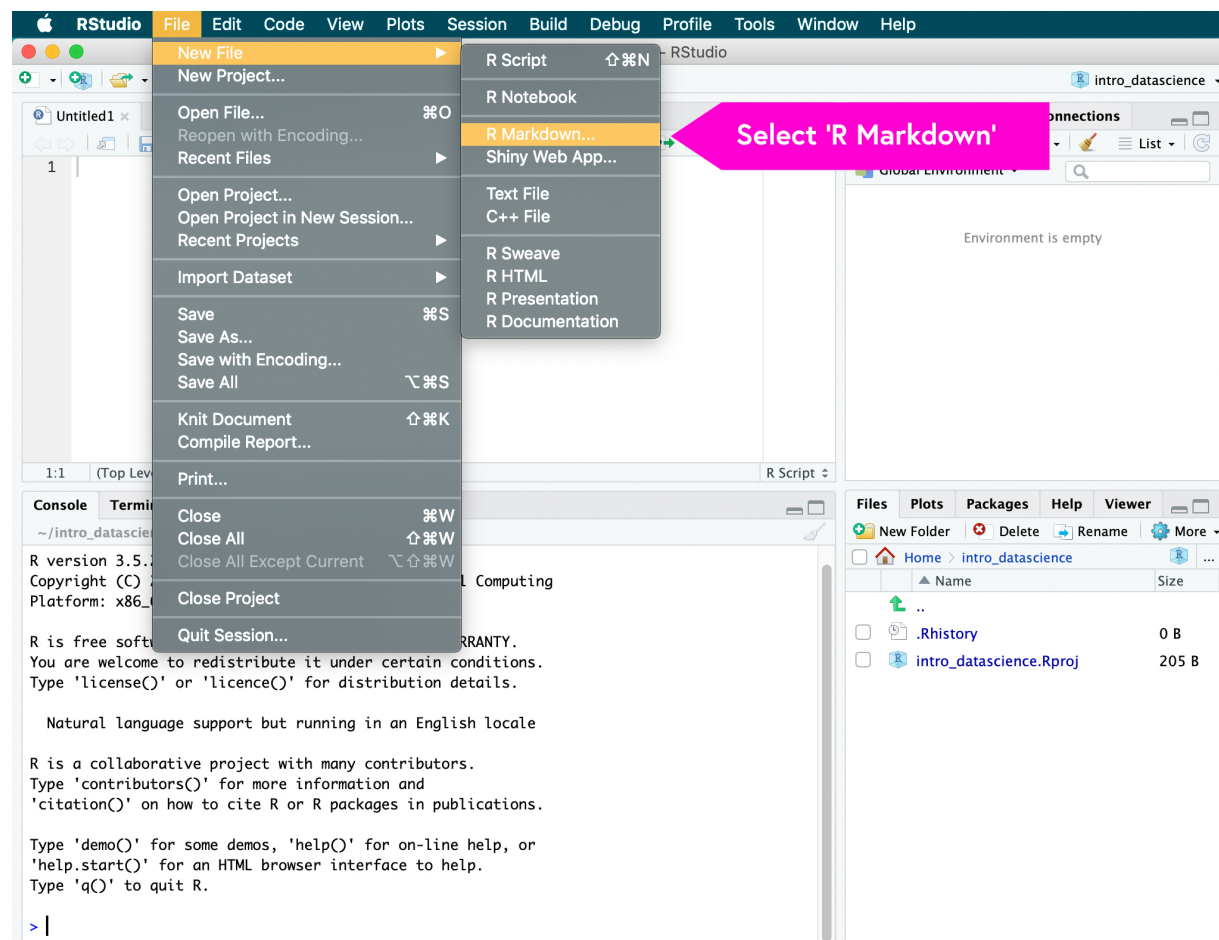
Consider watching the **first** video in a series of tutorials by Quang related to the coding needed for the exercises on this step, and then watch the rest in this series on YouTube ([https://www.youtube.com/playlist?list=PLmUx1KJpXh5\\_BY-bezPHpJJfW-cUdLjL6](https://www.youtube.com/playlist?list=PLmUx1KJpXh5_BY-bezPHpJJfW-cUdLjL6)) for further guidance. We hope you find them useful.

*Watch on YouTube to adjust the video size to a format that best suits you.*

embed ([https://www.youtube.com/watch?v=MhPp\\_82Fc7c](https://www.youtube.com/watch?v=MhPp_82Fc7c))

## Create your own R Markdown

If you haven't already, open RStudio on your computer and then create your own R Markdown file by selecting **File -> New File -> R Markdown**.



View a larger version of the image of RStudio. (<https://ugc.futurelearn.com/uploads/assets/6e/b5/6eb5951a-bb7b-4455-a2c8-9396fd9b4168.jpg>)

## Code chunks

Chunks of R code surrounded by ````` (three backticks) and `{r}` specifies that this is R code. The R code chunk below loads the `tidyverse` (a collection of R packages designed for data science), stores a sample of the `diamonds` data (which comes from the `tidyverse` and is now loaded into your R session) as an object named `smaller`, then prints the sample data:

```
```{r}
library(tidyverse)
smaller <- filter(diamonds, carat <= 2.5)
smaller
```
```

## Headings

`#` is a top-level heading/section, `##` is a second-level heading/subsection, `###` is a third-level heading/subsubsection, etc.

## Lists

Use `'` or `*` or `+` at the beginning of several lines to create a bullet or unordered list. To create a numbered or ordered list, use `'1.'` and `'2.'` at the beginning of each line. Create sub-lists by indenting with **four** spaces, followed by the same syntax. For example:

```
- Item A
  - Item a
  - Item aa
- Item B
```

## Links

Links can be included in your document by inserting the URL. If you want to hyperlink a text, surround the text with `'[]'` followed by `'()'` around the link, for example, `[RStudio cheatsheets](https://www.rstudio.com/resources/cheatsheets/)` will produce the following hyperlink: RStudio cheatsheets (<https://www.rstudio.com/resources/cheatsheets/>).

## Images

A simple way to include images is

`![Image of postcard of Melbourne.](https://ugc.futurelearn.com/uploads/assets/2a/22/2a22717f-0a39-49ad-8438-f5ceea0e3c18.jpg "Image of postcard which embeds the image from the web into your document:`



## Attribution

Remember, when you use someone else's work, you need to check its copyright, permissions or licensing to determine if you can use it, and then appropriately attribute. For example, the image of the unused postcard is in the Public Domain (<https://creativecommons.org/publicdomain/mark/1.0/>) and made available by the National Museum of Australia (<https://www.nma.gov.au/>). The image could be attributed in the following format:

`'Collins Street, Looking East, Melbourne' [Image] National Museum of Australia Retrieved from http://collectionsearch.nma.gov.au/object/720`

# Compiling, or ‘making the output’

To compile your document, select ‘**Knit**’. If you have a big document, build it up in pieces. You can run just one code chunk at a time, several chunks together, or even one line of code by itself. The ‘**Run**’ button has a menu of options for running the coding in pieces.

## Reporting styles

With R Markdown, you can create different kinds of reports by changing the output YAML (Yet another markup language) tag or by selecting the ‘**Knit**’ drop down in RStudio.

There are many extension reporting types available using external packages. For example, you can create interactive tutorials with the `learnr` (<https://rstudio.github.io/learnr/>) package or interactive dashboards with the `flexdashboard` (<https://rmarkdown.rstudio.com/flexdashboard/>) package. You can even write whole books (using the `bookdown` (<https://bookdown.org/>) package ) or use it to create a personal website (using the `blogdown` (<https://bookdown.org/yihui/blogdown/>) package), or even use other computer languages (<http://rmarkdown.rstudio.com/>) in the same document. These all follow the same principles as regular R Markdown, so once you understand the basics, you can easily try out other styles!

## Give it a go!

Continue to develop your skills in RStudio by making your way through this exercise - creating your first R Markdown document that **reads** in a dataset distributed by fivethirtyeight (<https://github.com/fivethirtyeight/data>), **makes** a summary table of the data, **generates** a few plots, and then **compiles** the R Markdown document into a HTML file.

- **Step 1:** Open your course project open in RStudio.
- **Step 2:** In RStudio, select **File --> New File --> R Markdown** .
- **Step 3:** Give your report a memorable title, but don't forget to include your name and today's date in the YAML header.
- **Step 4:** Include one code chunk that loads the `tidyverse` and reads in the data. For example, if you were interested in reading in the data set for the 538 article about the Avengers (<https://fivethirtyeight.com/features/avengers-death-comics-age-of-ultron/>) your code chunk should contain the following code:

```
library(tidyverse)
data_url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/avengers/avengers.csv"
avengers <- read_csv(data_url)
```

- **Step 5:** Include one chunk for making a summary table of the data. You might like to look up the functions `summary` or `glimpse` .
- **Step 6:** Include one chunk to plot the data. You might like to look up some base R functions like `hist()` or `plot()` (built-in R functions). For a challenge, you might like to try the `visdat` package we installed previously or `ggplot2` (which will have been loaded into your R session after loading the `tidyverse` ). For example, the code chunk below uses the `ggplot()` function to create a box plot of the number of appearances for each Avenger by the Avenger's sex:

```
ggplot(avengers, aes(x = Gender, y = Appearances)) +
  geom_boxplot()
```

- **Step 7:** Knit the `.Rmd` file you have made into a HTML document. Open your HTML in a web browser to view your work.

## Tell us how you went

Within the **Comments**, consider sharing with other learners your experience of creating your first R Markdown document.

- **How did you go? Were you able to read in the data, make a summary table and generate plots?**
- **As part of the exercise, you were asked to compile the document into a HTML file. Using RStudio, how could you have created a Word document instead of HTML?**

Also consider reading and commenting on contributions made by other learners or following learners with similar interests to you.