

# Week 03 Worksheet - part 1

A first gentle introduction

INSERT YOUR NAME HERE

## Welcome to RStudio!

This worksheet will introduce you to the RStudio interface and the concept of literate programming using R.

## The RStudio Interface (Lab Tutor Tour)

When you open RStudio, you'll see four main panes:

1. **Source Editor** (top-left): This is where you write and edit your code and text.
2. **Console** (bottom-left): This is where R commands are executed and where you see the output.
3. **Environment/History** (top-right): This shows the objects in your workspace and your command history.
4. **Files/Plots/Packages/Help** (bottom-right): This multi-purpose pane shows your files, plots, installed packages, and help documentation.

## This is a .qmd file

It is markdown - a text only language that can be edited anywhere, even on your phone, because it uses commonly occurring symbols to do all the formatting. If you happen to know html, it's pretty similar. It's simple (once you get the idea), very small in terms of file-size, easily accessed, and versatile.

## Understanding Literate Programming

Literate programming is a paradigm that combines explanatory text with executable code in a single document. This approach, pioneered by Donald Knuth in 1984, aims to make programming more accessible, understandable, and maintainable. When you learn about Open Science practices next week with Stacey, she will probably talk about sharing of data, replicability and other big issues. One way in which we can make Science more accessible and open is by using techniques like this.

### Key Concepts

1. **Integration of Code and Documentation:** In literate programming, the code is interspersed with narrative text that explains the purpose and functionality of the code.
2. **Human-Oriented:** The primary focus is on making the program understandable to humans, rather than just computers.
3. **Executable Documents:** The resulting document can be both read as a coherent explanation and executed as a functional program.

### Benefits in Data Science and Research

Literate programming is particularly valuable in data science and research for several reasons:

1. **Reproducibility:** By combining code, results, and explanations, others can easily reproduce and verify your work.
2. **Clear Communication:** It allows you to explain your thought process, methodology, and interpretation of results alongside the code that generates them.
3. **Error Checking:** The close proximity of code and explanation makes it easier to spot inconsistencies or errors.
4. **version Control:** Changes in both code and narrative can be tracked together, providing a comprehensive history of the project's evolution.

### Example in R

Here's a simple example of literate programming in R using a Quarto document:

In this analysis, we'll explore the relationship between a car's horsepower and its fuel efficiency using the `mtcars` dataset (this is a commonly used dataset that comes installed in R).

## “The Results of My Amazing Experiment”

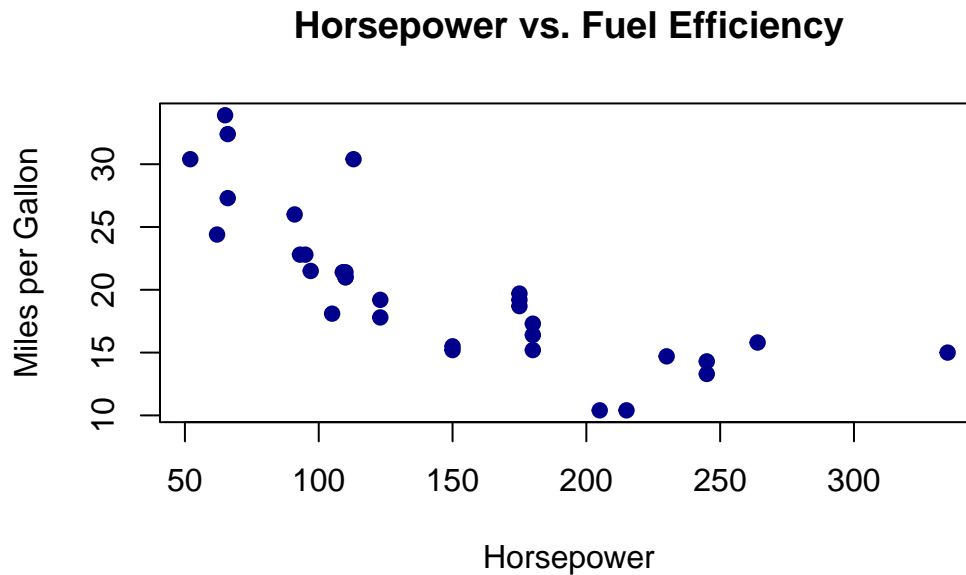
Dear Reader, be amazed at my wonderfulness! Below, I am going to walk you through the analysis. If you are reading this as a .qmd file, you will see all the code. But after I render it to html or pdf, you'll only see what I want you to see, such as just the pretty pictures, without the code! Or I can fold up the code here too!

Another aspect of literate programming, other than narrating the larger document, you can also add comments in the code, to draw an interested reader to specialist aspects of what's happening there, or just to signal what the code is doing. When you try to work with someone else's code, comments like the ones below (but better) are really helpful!

If you are reading along in the .qmd file, press the green arrow to the right of the code cell to execute it. And bingo! R does it's work and performs the calculations.

```
# Load the mtcars dataset
data(mtcars)

# Create a scatter plot
plot(mtcars$hp, mtcars$mpg,
     main="Horsepower vs. Fuel Efficiency", # Clearly these are the most important things!
     xlab="Horsepower",
     ylab="Miles per Gallon", # Can we change this to Kilometers per Litre for other countries?
     pch=19,
     col="darkblue") # The best colour for car-based statistics
```



## The Visual Editor

RStudio's visual editor provides a user-friendly interface for creating Quarto documents. It allows you to write text and code in a “What You See Is What You Get” (WYSIWYG) environment.

To switch to the visual editor, click the “Visual” button at the top-left of the Source Editor.

## Writing Narrative Text

In the visual editor, you can simply type your text as you would in any word processor. You can use the formatting toolbar at the top to:

- Apply **bold** or *italic* formatting
- Create bullet or numbered lists
- Add headings
- Insert links or images

Try writing a short paragraph about why you're studying Psychology below:

[Your text here]

## Adding Code Cells

To add a code cell:

1. Click the “+C” button in the toolbar or use the keyboard shortcut Cmd+Option+I (Mac) or Ctrl+Alt+I (Windows/Linux)
2. You’ll see a new code cell appear
3. Type your R code inside this cell

Let’s try a simple calculation. Add a code cell and type the following:

```
2 + 2
```

```
[1] 4
```

## Running Code

To run the code in a cell:

Click the “Run” button (green play icon) at the top-right of the cell, or Use the keyboard shortcut: Cmd+Enter (Mac) or Ctrl+Enter (Windows/Linux)

The output will appear directly below the code cell. Rendering Your Document To create the final document:

Click the “Render” button at the top of the editor Choose your desired output format (HTML or PDF) RStudio will process your document and display the result

## Exercise

Load the data Gordon has been using today. [https://github.com/DrDeception/Week0203/blob/3276cb8fa56a466a3e25761523cea475ed4ee4bd/materials/data/Y1W3\\_data.csv](https://github.com/DrDeception/Week0203/blob/3276cb8fa56a466a3e25761523cea475ed4ee4bd/materials/data/Y1W3_data.csv)

```
# Welcome to R! This is a comment. R doesn't run anything on a line after a '#'.

# Before we can read our data, we need a special tool called 'readr'.
# Think of this like buying a book from a bookstore.
# We only need to do this once, just like you only need to buy a book once.
install.packages("readr")
```

- readr [2.1.5]

```
# Installing packages -----
```

Successfully installed 1 package in 6.5 milliseconds.

```
# In R, we do this by 'loading' the library. We'll need to do this each time we start a new l
```

```
library(readr) # This is like opening the book 'readr' to use its contents
```

```
url <- "https://raw.githubusercontent.com/DrDeception/Week0203/3276cb8fa56a466a3e25761523cea"
```

```
# This is like using the skills we learned from the book to read a document in the library.
```

Rows: 109 Columns: 27

Delimiter: ", "

```
dbl (13): PokeNumber, LoginCount, EyeContact, Psy4me, DrWho, Maths, E_TIPI,...
```

i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
# This is like skimming the first page of the document.
```

```
# A tibble: 6 x 27
```

6

```

5           5 charmeleon https://r~           61 Do you~ How oft~ Why did~ 10'00"
6           6 charizard  https://r~           58 How of~ What ty~ Describ~ 18'00"
# i 19 more variables: `A-Level` <chr>, Coin <chr>, EyeContact <dbl>,
#   Psy4me <dbl>, DrWho <dbl>, Maths <dbl>, InsectApocalypse <chr>,
#   DogCatBoth <chr>, LarkorOwl <chr>, E_TIPI <dbl>, A_TIPI <dbl>,
#   C_TIPI <dbl>, ES_TIPI <dbl>, O_TIPI <dbl>, MBTI <chr>, Image <chr>,
#   Commentary <chr>, Row <dbl>, Column <dbl>

# Now, let's get a summary of our data.
# This is like reading the table of contents and index to get an overview.
print(summary(data))

```

```

PokeNumber   PokeName      PokeImage      LoginCount
Min.   : 1   Length:109    Length:109     Min.   : 1.00
1st Qu.: 28   Class :character Class :character 1st Qu.:10.00
Median : 55   Mode  :character Mode  :character Median :15.00
Mean    : 55                                     Mean   :19.43
3rd Qu.: 82                                     3rd Qu.:25.00
Max.    :109                                     Max.   :61.00
                                                NA's   :12

Q_Scale      Q_choice      Q_option      CompTime
Length:109   Length:109   Length:109   Length:109
Class :character Class :character Class :character Class1:hms
Mode  :character Mode  :character Mode  :character Class2:difftime
                                                Mode   :numeric

```

```

A-Level      Coin          EyeContact      Psy4me
Length:109   Length:109   Min.   : 1.000 Min.   :1.000
Class :character Class :character 1st Qu.: 5.000 1st Qu.:4.000
Mode  :character Mode  :character Median : 7.000 Median :4.000
                                                Mean   : 6.314 Mean   :4.114
                                                3rd Qu.: 8.000 3rd Qu.:5.000
                                                Max.   :10.000 Max.   :5.000
                                                NA's   :39     NA's   :39

```

```

DrWho      Maths      InsectApocalypse DogCatBoth
Min.   : 0.000 Min.   :1.000 Length:109 Length:109
1st Qu.: 1.000 1st Qu.:2.250 Class :character Class :character
Median : 1.500 Median :3.000 Mode  :character Mode  :character
Mean    : 3.471 Mean    :2.986
3rd Qu.: 5.000 3rd Qu.:4.000

```

Max. :10.000	Max. :5.000		
NA's :39	NA's :39		
LarkorOwl	E_TIPI	A_TIPI	C_TIPI
Length:109	Min. :2.000	Min. :2.000	Min. :2.000
Class :character	1st Qu.:3.000	1st Qu.:3.500	1st Qu.:3.000
Mode :character	Median :3.000	Median :4.000	Median :3.500
	Mean :3.179	Mean :3.814	Mean :3.221
	3rd Qu.:3.500	3rd Qu.:4.500	3rd Qu.:3.500
	Max. :4.500	Max. :5.500	Max. :5.000
	NA's :39	NA's :39	NA's :39
ES_TIPI	O_TIPI	MBTI	Image
Min. :2.000	Min. :2.000	Length:109	Length:109
1st Qu.:2.500	1st Qu.:4.000	Class :character	Class :character
Median :3.000	Median :4.500	Mode :character	Mode :character
Mean :2.993	Mean :4.243		
3rd Qu.:3.500	3rd Qu.:5.000		
Max. :5.000	Max. :5.500		
NA's :39	NA's :39		
Commentary	Row	Column	
Length:109	Min. : 1.000	Min. : 1.000	
Class :character	1st Qu.: 4.000	1st Qu.: 3.000	
Mode :character	Median : 5.500	Median : 5.000	
	Mean : 5.357	Mean : 5.386	
	3rd Qu.: 7.000	3rd Qu.: 7.000	
	Max. :10.000	Max. :10.000	
	NA's :39	NA's :39	

```
# We can check how big our dataset is.
# This is like counting the pages and chapters in our document.
cat("Our dataset has", dim(data)[1], "rows (like pages) and", dim(data)[2], "columns (like chapters)")
```

Our dataset has 109 rows (like pages) and 27 columns (like chapters)

```
# Finally, let's see what information (columns) we have in our dataset.
# This is like looking at the headings in our document.
cat("Our dataset contains information about:\n")
```

Our dataset contains information about:



```
print(colnames(data))
```

[1]	"PokeNumber"	"PokeName"	"PokeImage"	"LoginCount"
[5]	"Q_Scale"	"Q_choice"	"Q_option"	"CompTime"
[9]	"A-Level"	"Coin"	"EyeContact"	"Psy4me"
[13]	"DrWho"	"Maths"	"InsectApocalypse"	"DogCatBoth"
[17]	"LarkorOwl"	"E_TIPI"	"A_TIPI"	"C_TIPI"
[21]	"ES_TIPI"	"O_TIPI"	"MBTI"	"Image"
[25]	"Commentary"	"Row"	"Column"	

```
# Congratulations! You've just used R to bring in data from the internet and take a first look
# It's like you've gone to a digital library, found a document, and skimmed through its contents
```

## Normally...

That code chunk would only be like this:

```
install.packages("readr")
```

The following package(s) will be installed:

- readr [2.1.5]

```
These packages will be installed into "~/Week0203/renv/library/macos/R-4.4/x86_64-apple-darwin20.0.0/"
```

```
# Installing packages -----
- Installing readr ... OK [linked from cache]
Successfully installed 1 package in 5.3 milliseconds.
```

```
library(readr)
```

```
data <- read_csv("https://raw.githubusercontent.com/DrDeception/Week0203/3276cb8fa56a466a3e2d")
```

Rows: 109 Columns: 27

```
-- Column specification -----
```

Delimiter: ", "

```
chr (13): PokeName, PokeImage, Q_Scale, Q_choice, Q_option, A-Level, Coin, ...
```

```
dbl (13): PokeNumber, LoginCount, EyeContact, Psy4me, DrWho, Maths, E_TIPI,...
```

```
time (1): CompTime
```

i Use ``spec()`` to retrieve the full column specification for this data.

i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
summary(data) # Even this isn't necessary if you know what the data includes already!
```

PokeNumber	PokeName	PokeImage	LoginCount
Min. : 1	Length:109	Length:109	Min. : 1.00
1st Qu.: 28	Class :character	Class :character	1st Qu.:10.00
Median : 55	Mode :character	Mode :character	Median :15.00
Mean : 55			Mean :19.43
3rd Qu.: 82			3rd Qu.:25.00
Max. :109			Max. :61.00
			NA's :12

Q_Scale	Q_choice	Q_option	CompTime
Length:109	Length:109	Length:109	Length:109
Class :character	Class :character	Class :character	Class1:hms
Mode :character	Mode :character	Mode :character	Class2:difftime
			Mode :numeric

A-Level	Coin	EyeContact	Psy4me
Length:109	Length:109	Min. : 1.000	Min. :1.000
Class :character	Class :character	1st Qu.: 5.000	1st Qu.:4.000
Mode :character	Mode :character	Median : 7.000	Median :4.000
		Mean : 6.314	Mean :4.114
		3rd Qu.: 8.000	3rd Qu.:5.000
		Max. :10.000	Max. :5.000
		NA's :39	NA's :39

DrWho	Maths	InsectApocalypse	DogCatBoth
Min. : 0.000	Min. :1.000	Length:109	Length:109
1st Qu.: 1.000	1st Qu.:2.250	Class :character	Class :character
Median : 1.500	Median :3.000	Mode :character	Mode :character
Mean : 3.471	Mean :2.986		
3rd Qu.: 5.000	3rd Qu.:4.000		
Max. :10.000	Max. :5.000		
NA's :39	NA's :39		

LarkorOwl	E_TIPI	A_TIPI	C_TIPI
Length:109	Min. :2.000	Min. :2.000	Min. :2.000
Class :character	1st Qu.:3.000	1st Qu.:3.500	1st Qu.:3.000
Mode :character	Median :3.000	Median :4.000	Median :3.500
	Mean :3.179	Mean :3.814	Mean :3.221
	3rd Qu.:3.500	3rd Qu.:4.500	3rd Qu.:3.500
	Max. :4.500	Max. :5.500	Max. :5.000
	NA's :39	NA's :39	NA's :39

ES_TIPI	O_TIPI	MBTI	Image
Min. :2.000	Min. :2.000	Length:109	Length:109
1st Qu.:2.500	1st Qu.:4.000	Class :character	Class :character
Median :3.000	Median :4.500	Mode :character	Mode :character
Mean :2.993	Mean :4.243		
3rd Qu.:3.500	3rd Qu.:5.000		
Max. :5.000	Max. :5.500		
NA's :39	NA's :39		

Commentary	Row	Column
Length:109	Min. : 1.000	Min. : 1.000
Class :character	1st Qu.: 4.000	1st Qu.: 3.000
Mode :character	Median : 5.500	Median : 5.000
	Mean : 5.357	Mean : 5.386
	3rd Qu.: 7.000	3rd Qu.: 7.000
	Max. :10.000	Max. :10.000
	NA's :39	NA's :39

```
dim(data) # or this
```

```
[1] 109 27
```

```
colnames(data) # or this.
```

```
[1] "PokeNumber"      "PokeName"        "PokeImage"       "LoginCount"
[5] "Q_Scale"         "Q_choice"        "Q_option"        "CompTime"
[9] "A-Level"         "Coin"            "EyeContact"      "Psy4me"
[13] "DrWho"           "Maths"           "InsectApocalypse" "DogCatBoth"
[17] "LarkorOwl"       "E_TIPI"          "A_TIPI"          "C_TIPI"
[21] "ES_TIPI"         "O_TIPI"          "MBTI"            "Image"
[25] "Commentary"      "Row"             "Column"
```

## Data Viewer

If you look in the Environment panel to the right, you will see `data` and if you click on the little spreadsheet icon to the right, you can look at it like a spreadsheet!

## My First Plot

```
# COPY SOME CODE IN HERE from the slides from this morning - choose one of the early, simpler
```

## Run the code cell

Render your document to html (website) or pdf (document) and view the results!

## Conclusion

You've now learned the basics of:

### Note

- Navigating the RStudio interface Using the visual editor for Quarto documents
- Writing narrative text and adding formatted elements
- Inserting and running code cells to do sums and create graphs
- Rendering your document

Keep practicing these skills as you continue your journey with R and data analysis!