

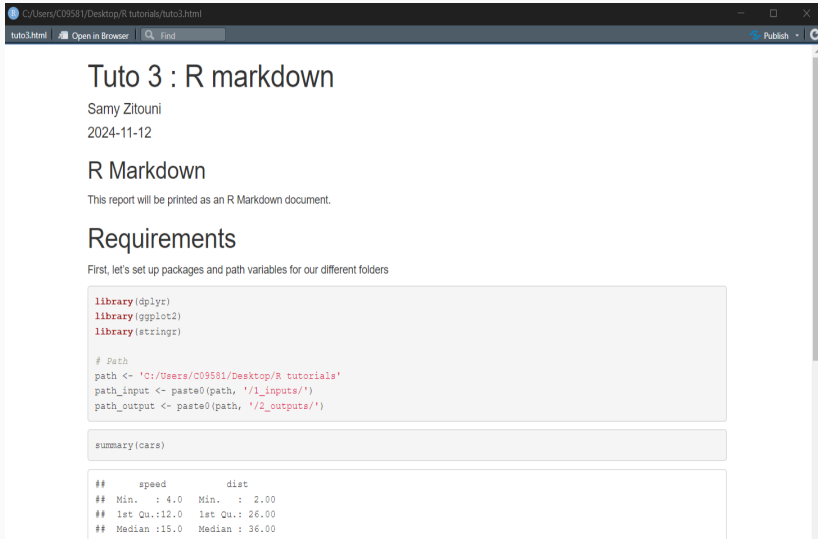
An introduction to R

R markdown basics

Samy Zitouni

October 2024

What is R Markdown ?



The screenshot shows a web browser window with the address bar displaying 'C:/Users/C09581/Desktop/R tutorials/tuto3.html'. The browser has tabs for 'tuto3.html', 'Open in Browser', and 'Find'. The document content is as follows:

Tuto 3 : R markdown

Samy Zitouni
2024-11-12

R Markdown

This report will be printed as an R Markdown document.

Requirements

First, let's set up packages and path variables for our different folders

```
library(dplyr)
library(ggplot2)
library(stringr)

# Path
path <- 'C:/Users/C09581/Desktop/R tutorials'
path_input <- paste0(path, '/1_inputs/')
path_output <- paste0(path, '/2_outputs/')

summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   : 2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##    ...      ...    ...    ...    ...
```

What is R markdown

Generates reports, in which you can include...

- Titles, text, global text formatting
- **Code** parts
- The **results** of your code, whether they are **datasets**, **plots**, regression **summaries**, etc.
- **Math** equations with \LaTeX
- It is better than commenting code, and **more practical** than exporting your results on another text editor

You can generate...

- Reports
- Presentations
- Books, widget, dashboards, etc.
- See **here** for more examples

Getting started

- Create a new rmd file through **New document** > **R Markdown**
- A popup should appear
- Fill **general information** (type of document, your information, title, date)
- This defines the **YAML metadata**, which tells how your file is built
- Click **OK**

New R Markdown

Document
Presentation
Shiny
From Template

Title:

Author:

Date:

☐ Use current date when rendering document

Default Output Format:

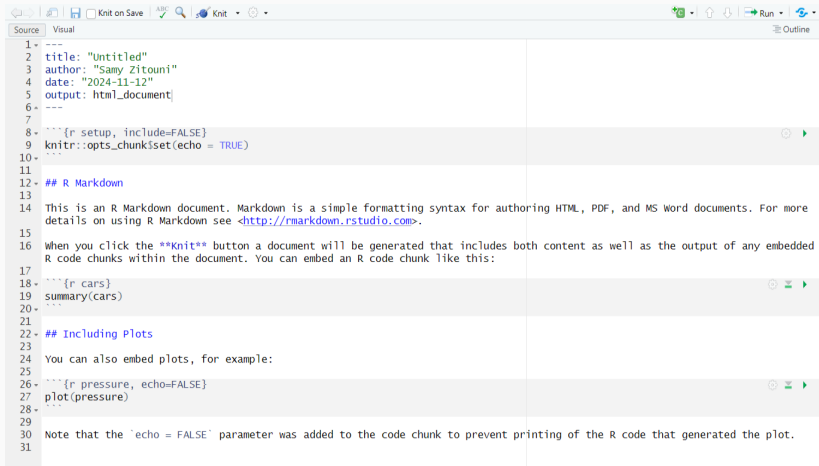
☒ HTML
Recommended format for authoring (you can switch to PDF or Word output anytime).

☐ PDF
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

☐ Word
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

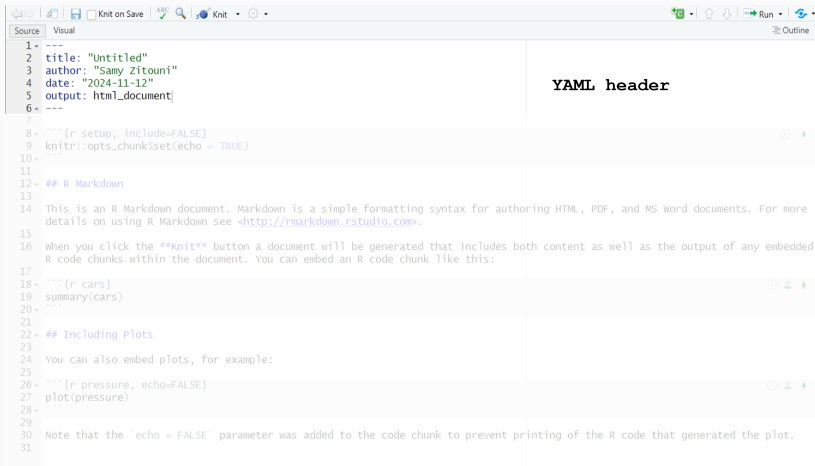
Create Empty Document OK Cancel

General structure of the code



```
1 ---
2 title: "Untitled"
3 author: "Samy Zitouni"
4 date: "2024-11-12"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more
15 details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the Knit button a document will be generated that includes both content as well as the output of any embedded
18 R code chunks within the document. You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 ```{r pressure, echo=FALSE}
29 plot(pressure)
30 ```
31
32 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

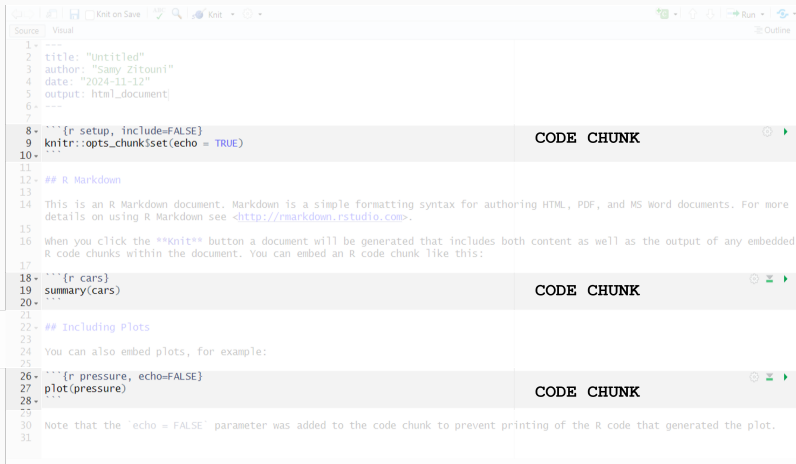
General structure of the code



```
1 ---
2 title: "Untitled"
3 author: "Samy Zitouni"
4 date: "2024-11-12"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more
15 details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the Knit button a document will be generated that includes both content as well as the output of any embedded
18 R code chunks within the document. You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 ```{r pressure, echo=FALSE}
29 plot(pressure)
30 ```
31
32 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

YAML header

General structure of the code



The screenshot shows the RStudio interface with a source editor on the left and a viewer on the right. The source editor contains an R Markdown document with the following content:

```
1 ---
2 title: "Untitled"
3 author: "Samy Zitouni"
4 date: "2024-11-12"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more
15 details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the Knit button a document will be generated that includes both content as well as the output of any embedded
18 R code chunks within the document. You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 ```{r pressure, echo=FALSE}
29 plot(pressure)
30 ```
31
32 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

The code is organized into three sections, each labeled "CODE CHUNK" on the right side of the editor:

- Chunk 1 (Lines 8-10):** Setup code for knitr, including setting the echo parameter to TRUE.
- Chunk 2 (Lines 18-20):** R code for summarizing the 'cars' dataset.
- Chunk 3 (Lines 28-30):** R code for plotting the 'pressure' dataset, with the echo parameter set to FALSE.

General structure of the code



The screenshot displays the RStudio interface with an R Markdown document. The document is titled "Untitled" and includes metadata for the author (Samy Zitouni) and date (2024-11-12). The output format is set to "html_document". The document contains three R code chunks, each followed by a text block representing its rendered output.

```
1 ---
2 title: "Untitled"
3 author: "Samy Zitouni"
4 date: "2024-11-12"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more
15 details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the Knit button a document will be generated that includes both content as well as the output of any embedded
18 R code chunks within the document. You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 ```{r pressure, echo=FALSE}
29 plot(pressure)
30 ```
31
32 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

Text part

Text part

Text part

Text part

Let's run it !

- You can run a chunk you are into : **Ctrl + Enter** or clicking on the green arrow at the top right corner of the chunk
- You can run all your code chunks : **Ctrl + Alt + R**
- **Most importantly**, you can run and preview your code : **Ctrl + Alt + K** or pressing the "Knit" button at the top

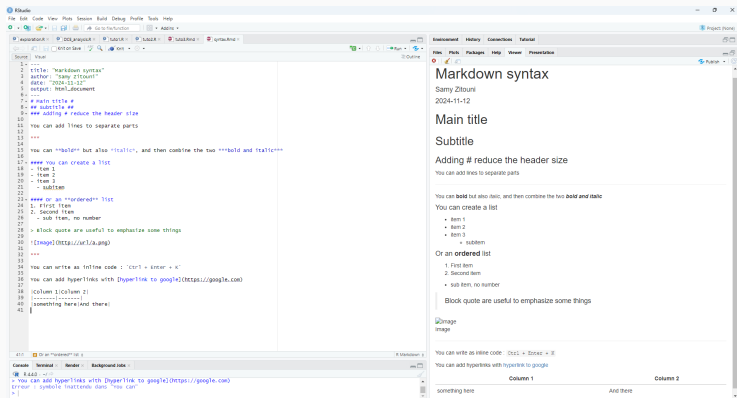
For the configuration of the header of your document

- It contains attributes which values you can set
 - Title, subtitle, author, date, the type of output (html document, word document, pdf document)
 - The yaml header is **always** between two sets of three dashes ---

The text syntax

Markdown syntax

- The text parts use the markdown syntax that is common to R Markdown, Jupyter notebooks, etc.



Code chunks

- You can write the fences or simply press add cell on the top right corner

```
`{my r chunk 1}`  
x <- 5  
x+4  
`
```

- This code chunk gives :

```
x <- 5  
x+4
```

```
## [1] 9
```

Chunk options

- Every chunk has a set of options, that can be set **TRUE** or **FALSE**
- Every option allows you to **print** or **not print** different part of the code or the output

Option	Default	Effect
eval	TRUE	Evaluate the code and include its results?
echo	TRUE	Display the code and include its results?
warning	TRUE	Display warnings?
error	FALSE	Display errors?
message	TRUE	Display messages?
results	'markup'	How to render code output? 'hide' hides output
fig.width	7	Width in inches for plots created in the chunk
fig.height	7	Height in inches for plots created in the chunk

More about inline code

- There are places in your report in which you would like to **include some results** in your text
- Inline code is typically what you want to look at
- It shows the code and you also can execute it directly by adding a ``r youcodeline``

Example

```
```${r number of rows, echo = FALSE}  
Path
path <- 'C:/Users/C09581/Desktop/R tutorials'
path_input <- paste0(path, '/1_inputs/')
path_output <- paste0(path, '/2_outputs/')

Import data
wid <- read.csv(paste0(path_input, 'wid.csv'))
a <- nrow(wid)
```
```

In this text, I am talking about a dataframe that contains ``a`` rows. I could also have written
↪ ``nrow(wid)`` rows directly

In this text, I am talking about a dataframe that contains ``r a`` rows. I could also have
↪ written ``r nrow(wid)`` rows directly

That gives

```
# Path
path <- 'C:/Users/C09581/Desktop/R tutorials'
path_input <- paste0(path, '/1_inputs/')
path_output <- paste0(path, '/2_outputs/')

# Import data
wid <- read.csv(paste0(path_input, 'wid.csv'))
a <- nrow(wid)
```

In this text, I am talking about a dataframe that contains `a` rows. I could also have written `nrow(wid)` rows directly

In this text, I am talking about a dataframe that contains 1610 rows. I could also have written 1610 rows directly

General tips

Showing dataframes in the report

- Use the "DT" library
- I don't show it here but `kable(head(data))` is neat

```
```${r my code}
 library("DT")
 datatable(mtcars, options = list(pageLength = 5))
```
```

| Show | 5 | ▼ | entries | Search: | | | | | | | |
|------------------------------|------|-----|---------|-----------------------------|------|-------|-------|----|----|------|------|
| | mpg | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb |
| Mazda RX4 | 21 | 6 | 160 | 110 | 3.9 | 2.62 | 16.46 | 0 | 1 | 4 | 4 |
| Mazda RX4 Wag | 21 | 6 | 160 | 110 | 3.9 | 2.875 | 17.02 | 0 | 1 | 4 | 4 |
| Datsun 710 | 22.8 | 4 | 108 | 93 | 3.85 | 2.32 | 18.61 | 1 | 1 | 4 | 1 |
| Hornet 4 Drive | 21.4 | 6 | 258 | 110 | 3.08 | 3.215 | 19.44 | 1 | 0 | 3 | 1 |
| Hornet Sportabout | 18.7 | 8 | 360 | 175 | 3.15 | 3.44 | 17.02 | 0 | 0 | 3 | 2 |
| Showing 1 to 5 of 32 entries | | | | Previous 1 2 3 4 5 6 7 Next | | | | | | | |

General tips

Header parameters

- In my header, I can add a parameter, let's say a name.
- Later in the code, I can access it through inline code ``r
params$name``.
- I can generate reports for each name changing one parameter only

```
---  
title: "Tuto 3 : R markdown"  
author: "Samy Zitouni"  
date: "2024-11-12"  
output: html_document  
params:  
  name: "EDGAR"  
---
```

Objective : plot the evolution of names attribution

- Download the name dataset
- **Clean** the different columns
- **Summarize** the data by **year** and and count *the total number* of birth per year
- Choose two names and **compare** the evolution of their attribution at birth
- **Style** your graphs