



## Write scientific reports in R Markdown: Meet your instructor

- PhD Candidate in Econometrics (DNB + TI+ ESE) since 2021
- Previous:
  - Bsc in Statistics (2010),
  - Bsc in Economics (2019),
  - Msc in Economics (2019),
  - Mphil in Econometrics (2022)
- R user since 2008 (but not a developer)
- Available at venesschmidt@ese.eur.nl





## Goal of this tutorial: Write scientific reports in R Markdown

- To introduce R Markdown:
  - Motivation
  - How to get started
  - Markdown writing
  - Inserting LaTeX
  - Inserting chunks with R code
  - Calling R functions along Markdown text



#### RMarkdown: what?



- RMarkdown was introduced in 2012 along the knitr package for R (and became a pagkage in 2014)
- The idea is to have code together with text, written with Markdown
- Can create reports in different formats: pdf, html, word or slides in power point, beamer, html
- Compiled from R, it can insert output from code directly in your file

#### RMarkdown: what changes from standard R code?

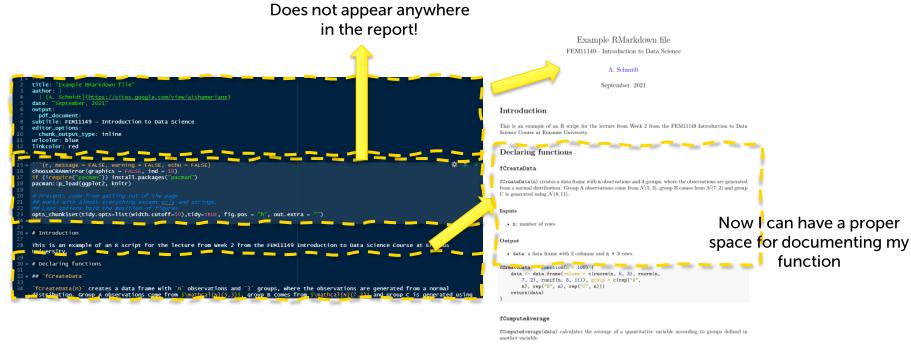
```
17 - ######################
    if (!require("ggplot2")) install.packages("ggplot2")
31 v fCreateData <- function(n = 100){
      data <- data.frame(values = c(rnorm(n, 5, 3),</pre>
                                     rnorm(n, 7, 2),
                                     runif(n, 8, 11)),
                          group = c(rep("A", n),
                                    rep("B", n),
                                    rep("c", n)))
      return(data)
    fComputeAverage <- function(data = data){</pre>
      if (!is.data.frame(data)){
        stop("Please use a data.frame object!")
      average <- tapply(data$values, data$group, mean)
```

I have some documentation for my functions, but they are displayed as comments The plot g1 will not be displayed in my \*.R code, to see the graph it you have to run the code yourself.



Figures: Standard R code

#### RMarkdown: what changes from standard R code?





#### RMarkdown: what changes from standard R code?

#### Initialization

n <- 200
dX <- fCreateData(n)

#### Main

We first check the first and last observations from dX, to see what information is there.

Table 1: First and last values of dX

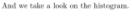
Value	Group		Variable	Group
0.2034385	A	Т	8.413659	С
6.3201182	A	i	8.877690	C
6.2419738	A	i	9.129007	C
0.0055206	A	i	10.370745	C
6.9101949	A	i	10.568385	C
8.8625207	A	i	10.458294	C

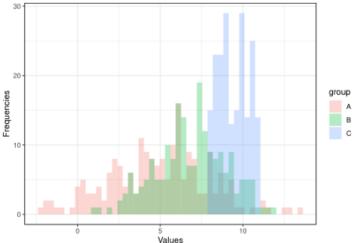
Now we take a look of the average value for each group.

Table 2: Average values per group from dX

A 4.924333 B 6.853090 C 9.430250

2







Figures: Rmarkdown output

#### RMarkdown: why?

- Improves scientific communication: you can use text to explain parts of the code to others (including your future self!)
- Makes it easier to update dynamic

#### When it might not be so helpful?

- Writing packages
- Doing heavy simulations (better run separately in a regular R file and call in the Rmarkdown document later)





#### So what are we doing, in practice?

- Instead of starting an R code file, we start an Rmarkdown file
- In it, we include code using chunks, that might or might not appear in the final document
- We can use results from the code to display tables, graphs, coding results
- We can also insert text using Markdown and equations using LaTeX

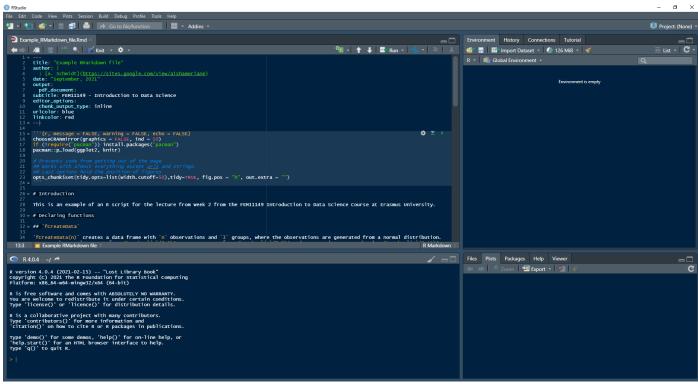
## Goal of this tutorial: Write scientific reports in R Markdown

We are here!

- To introduce R Markdown:
  - Motivation
  - How to get started
  - Markdown writing
  - Inserting LaTeX
  - Inserting chunks with R code
  - Calling R functions along Markdown text



### RMarkdown: first steps



At this point you should know how to start an Rmarkdown file based on the preparation slides.



#### RMarkdown: the anatomy of a file (1)

```
Example RMarkdown file.Rmd
                                                                                                   • Run - | • - | ≡
          🔚 👫 🔍 🖋 Knit 🕝 🌣 🕝
      title: "Example RMarkdown file"
      author:
                                                                    This is called YAML header
        [A. Schmidt](https://sites.google.com/view/aishameriane)
      date: "September, 2021'
      output:
        pdf_document:
                                                                         There are many options that can be
      subtitle: FEM11149 - Introduction to Data Science
      editor options:
        chunk output type: inline
                                                                         added here (as you can see I already
      urlcolor: blue
      linkcolor: red
                                                                               modified the 'simple' header)
  15 v ```{r, message = FALSE, warning = FALSE, echo = FALSE}
      chooseCRANmirror(graphics = FALSE, ind = 10)
     if (!require("pacman")) install.packages("pacman")
      pacman::p_load(ggplot2, knitr)
      opts_chunk$set(tidy.opts=list(width.cutoff=50),tidy=TRUE, fig.pos = "h", out.extra = "")
  24 🔺
  26 # Introduction
      This is an example of an R script for the lecture from Week 2 from the FEM11149 Introduction to Data Science Course at Erasmus University.
  30 # Declaring functions
  32 w ## `fCreateData`
      `fCreateData(n)` creates a data frame with `n` observations and `3` groups, where the observations are generated from a normal distribution.
```

You can check here for some more options for pdf file:

https://bookdown.org/yihui/rmarkdown/pdf-document.html



#### RMarkdown: the anatomy of a file (2)

```
Example RMarkdown file.Rmd
           🔚 👫 🔍 🖋 Knit 🕝 🌣 🕝
                                                                                                                  ↑ ↓ | ■ Run - | • - | =
       title: "Example RMarkdown file"
       author:
        [A. Schmidt](https://sites.google.com/view/aishameriane)
       date: "September, 2021"
       output:
        pdf_document:
       subtitle: FEM11149 - Introduction to Data Science
       editor options:
        chunk output type: inline
       urlcolor: blue
  12 linkcolor: red
  15 v ```{r, message = FALSE, warning = FALSE, echo = FALSE}
      chooseCRANmirror(graphics = FALSE, ind = 10)
                                                                                                 This is a chunk of code
      if (!require("pacman")) install.packages("pacman")
      pacman::p_load(ggplot2, knitr)
                                                                                                  Some customization is possible
      ## Last options hold the position of figures
opts_chunk$set(tidy.opts=list(width.cutoff=50),tidy=TRUE, fig.pos = "h", out.extra = "")
                                                                                                            (we will see later)
  24 🔺
  26 # Introduction
       This is an example of an R script for the lecture from Week 2 from the FEM11149 Introduction to Data Science Course at Erasmus University.
  30 # Declaring functions
  32 w ## `fCreateData`
       `fCreateData(n)` creates a data frame with `n` observations and `3` groups, where the observations are generated from a normal distribution.
```



### RMarkdown: the anatomy of a file (2)

```
Example RMarkdown file.Rmd
                                                                                                        Oc - 1 → Run - 2 - - =
      title: "Example RMarkdown file"
      author:
        [A. Schmidt](https://sites.google.com/view/aishameriane)
      date: "September, 2021"
      output:
        pdf_document:
      subtitle: FEM11149 - Introduction to Data Science
      editor options:
      chunk_output_type: inline
      urlcolor: blue
  12 linkcolor: red
  15 v ```{r, message = FALSE, warning = FALSE, echo = FALSE}
      chooseCRANmirror(graphics = FALSE, ind = 10)
  if (!require("pacman")) install.packages("pacman")
      pacman::p_load(ggplot2, knitr)
      opts_chunk$set(tidy.opts=list(width.cutoff=50),tidy=TRUE, fig.pos = "h", out.extra = "")
     # Introduction
      This is an example of an R script for the lecture from Week 2 from the FEM11149 Introduction to Data Science Course at Erasmus University.
     # Declaring functions
      ## `fCreateData`
       `fcreateData(n)` creates a data frame with `n` observations and `3` groups, where the observations are generated from a normal distribution.
```

This is plain text
Written in Markdown



#### Markdown (1)

- Markdown is intended to be an accessible syntax to write text that can be converted to html or other format types
- It is different from writing in MS Word where 'what you see is what you get' (WYSIWYG):
  - in Markdown, you have some basic syntax that needs to be interpreted in order to show its final form
- Still, it is simpler than a programming language

16

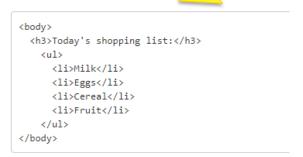
#### Markdown (2)

Say you want to make a webpage that displays this:

#### Today's shopping list:

- Milk
- Eggs
- Cereal
- Fruit

### This is the html code





This is the markdown code

### Today's shopping list:

- \* Milk
- \* Eggs
- \* Cereal
- \* Fruit

Ezafus,

Source: <a href="https://monashbioinformaticsplatform.github.io/2017-11-16-open-science-training/topics/rmarkdown.html">https://monashbioinformaticsplatform.github.io/2017-11-16-open-science-training/topics/rmarkdown.html</a>

### Markdown syntax

To delimit sections, you use # to make headers

```
# Time to learn some markdown!
## Time to learn some markdown!
### Time to learn some markdown!
#### Time to learn some markdown!
```

Time to learn some markdown!



### Markdown syntax

Plain text can be bold, italic, strikethrough

This is \*\*bold\*\* text

This is \_italic\_ text

This is ~~crossed out~~ text

This is **bold** text

This is italic text

This is <del>crossed out</del> text

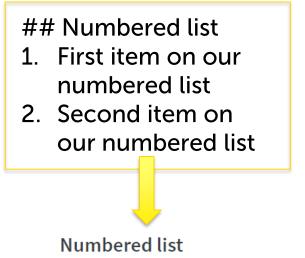


### Markdown syntax

You can make numbered or bullet lists

## Bullet list \* a bullet point - a bullet point + still a bullet point **Bullet list** · a bullet point

- · a bullet point
- · still a bullet point



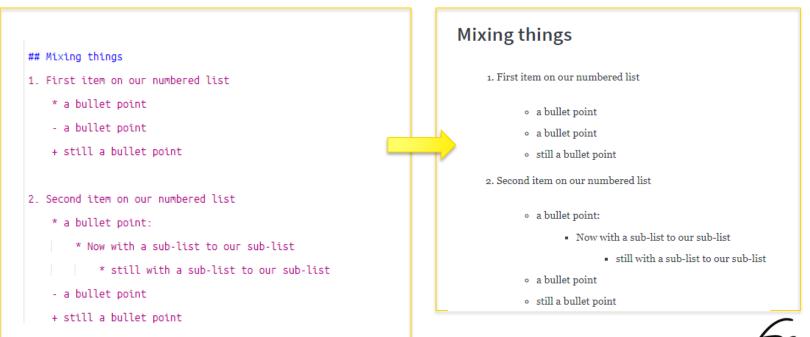
- 1. First item on our numbered list
- 2. Second item on our numbered list



20

#### Markdown syntax

You can mix elements in lists





#### Practice a bit!

Go to <a href="https://dillinger.io/">https://dillinger.io/</a>
and edit the file in there



## Goal of this tutorial: Write scientific reports in R Markdown

- To introduce R Markdown:
  - Motivation
  - How to get started
  - Markdown writing
  - Inserting LaTeX We are here!
  - Inserting chunks with R code
  - Calling R functions along Markdown text



23

### LaTeX (1)

- It is a software system intended for document preparation
- Outside RMarkdown you can use LaTeX to make whole documents (papers, thesis, books, etc)
  - In Rmarkdown we will use to write math expressions

LaTeX raw code

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

How it looks after compiling

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

24

#### LaTeX (2)

 Inside Rmarkdown, we insert LaTeX code inline using \$equation\$ or in a new line using \$\$equation\$\$

Inline LaTeX equations can be written in a pair of dollar signs using the LaTeX syntax, e.g.,  $f(k) = n \cdot p^{k} \cdot p^{k} \cdot (1-p)^{n-k}$  (actual output:  $f(k) = \binom{n}{k} p^{k} \cdot (1-p)^{n-k}$ ); math expressions of the display style can be written in a pair of double dollar signs, e.g.,  $f(k) = n \cdot p^{k} \cdot (1-p)^{n-k}$ , and the output looks like this:

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$



### LaTeX (3)



- Check <u>here</u> + <u>here</u> for some LaTeX tips;
- Be careful that not all LaTeX packages will be available, so some commands could not work on your document;
- Equation labeling and referencing is also not equal to what you do in 'plain' LaTeX
- Like any other software/language, it is the type of thing you learn by doing
  - And on Google



#### LaTeX (4)

 Try to write in a Markdown document what is written in the example for reports (link on chat)

```
## Writing what is in the instructions_report.pdf file

Equations and mathematical symbols can be easily added to a Markdown file. be written as $\frac{2}{3}$$. Other expressions are, e.g., $\hat{\lambda} = \frac{2}{\lambda}$$

(...)

An $n \times m$ matrix can be shown as

begin{equation}
n\mathbb{A}m = begin{pmatrix}
a_{11} & a_{12} & \ldots & a_{1m}\\
a_{21} & a_{22} & \ldots & a_{2m}\\
\dots & a_{11} & a_{11}
```

#### Writing what is in the instructions\_report.pdf file

Equations and mathematical symbols can be easily added to a Markdown file. A fraction of two-thirdas can be written as  $\frac{2}{3}$ . Other expressions are, e.g.,  $\hat{\lambda}=1.02$  and  $\sqrt{4}$ .

(...)

An  $n \times m$  matrix can be shown as

$$n\mathbb{A}m = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$
(1)

You can use Mathpix Snip to crop images, pdf files and other type of files and generate a LaTeX code. There are good tutorials on how to write equations in LaTeX (just be aware that in RMarkdown your possibilities with packages are a bit more limited than in a \*.tex document). Some references can be found here or in the report instructions on Canvas.

Check out this software: <a href="https://mathpix.com/">https://mathpix.com/</a> You can take 'snips' of text and it makes the LaTeX code. If you register with your Erasmus e-mail you get some extra benefits.



## Goal of this tutorial: Write scientific reports in R Markdown

- To introduce R Markdown:
  - Motivation
  - How to get started
  - Markdown writing
  - Inserting LaTeX
  - Inserting chunks with R code | We are here!
  - Calling R functions along Markdown text



#### Inserting chunks of R code into RMarkdown files (1)

- The biggest advantage of RMarkdown is to be able to use R code
- A 'chunk' is a piece of the document that will work exactly as an R file (so you cannot put text inside unless in commented lines)
- You can insert a chunk via menu or using the shortcut Ctrl+Alt+I
- There are many options for a chunk, for example size of figures that will display, hide the code (and display only results), etc

#### Inserting chunks of R code into RMarkdown files (2)





# — Inserting code

30

#### Inserting chunks of R code into RMarkdown files (3)

```
142 * # Inserting R code

143

144 * ```{r}

145  # This is the most basic R code chunk

147  x <- 1+1

148  x

149  * ```

Click here to compile all chunks that come before this one
```

Ezafus,

#### Inserting chunks of R code into RMarkdown files (4)

```
142 * # Inserting R code

143

144 * ```(r)

145  # This is the most basic R code chunk

148  X

149  * ```

Click here for some shortcuts for the chunk options
```



#### Inserting chunks of R code into RMarkdown files (5)

```
You can also position the cursor

(or select a few lines) and press

ctrl + enter
```



(Zafus)

#### Inserting chunks of R code into RMarkdown files (6)

```
142 * # Inserting R code

143

144 * ```{r}

145  # This is the most basic R code chunk

146

147  X <- 1+1

148  X

149 * ```
```

And this is how it looks When compiled

#### Inserting R code

```
# This is the most basic R code chunk

x <- 1 + 1

x
```



#### Inserting chunks of R code into RMarkdown files (7)





### Chunk options (1)

- eval = TRUE evaluate a code chunk (or FALSE to skip it)
- echo = TRUE to make the code appear in the final document (or FALSE to omit it)
- message = TRUE, warning = TRUE to allow for messages and warnings from the code
- Example: "\frac{r}{r}, eval = TRUE, echo = FALSE, warning = FALSE}
  # This is the most basic R code chunk
  x <- 1+1</pre>

 Check here for more: <u>https://bookdown.org/yihui/rmarkdown/r-code.html</u> and <u>https://yihui.org/knitr/options/</u>

## Goal of this tutorial: Write scientific reports in R Markdown

- To introduce R Markdown:
  - Motivation
  - How to get started
  - Markdown writing
  - Inserting LaTeX
  - Inserting chunks with R code
  - Calling R functions along Markdown text



#### Calling R functions inside Rmarkdown text (in line)

```
## Adding inline R code
To use R code inline, you need to type `r 3+1`. Besides doing simple math operations, you can call
objects that are already saved in the memory. Recall that the value of $x$ is `r x`.
Alternatively, you can use functions, such as `r mean(c(1,2,3,4))`. But for tables, figures and
more elaborate code, you have to use a chunk.
```



#### Calling R functions inside Rmarkdown text (in line)

```
*## Adding inline R code

To use R code inline, you need to type (r 3+1) Besides doing simple math operations, you can call objects that are already saved in the memory. Recall that the value of $x$ is (r x).

Alternatively, you can use functions, such as (r mean(c(1,2,3,4))). But for ables, figures and more elaborate code, you have to use a lunk.
```

#### Adding inline R code

To use R code inline, you need to type 4. Beside toing simple, ath operations, you can call objects that are already saved in the memory. Recan that it value of x is 2.

Alternatively, you can use functions, such as 2.5. But for tables, figures and more elaborate code, you have to use a chunk.



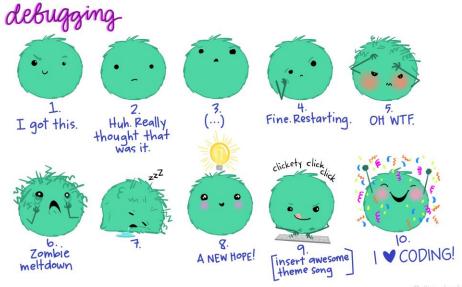
## Goal of this lecture: write scientific reports in R Markdown

- To introduce R Markdown:
  - Motivation
  - How to get started
  - Markdown writing
  - Inserting LaTeX
  - Inserting chunks with R code
  - Calling R functions along Markdown text



#### Final remarks

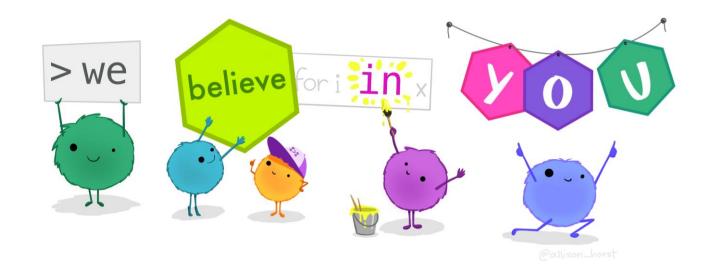
- You can find plenty of material online
- Best way to learn is starting your projects in RMarkdown
- Remember that programming can be unexpected





# — Final Remarks

#### Remember to have fun!





#### **Useful links**

- https://rmarkdown.rstudio.com/authoring\_quick\_tour.html
- https://rmarkdown.rstudio.com/docs/
- https://bookdown.org/yihui/rmarkdown/installation.html



#### Acknowledgements

 Drawings by <u>Alisson Horst</u>. Check her other awesome images here: <a href="https://github.com/allisonhorst/stats-illustrations">https://github.com/allisonhorst/stats-illustrations</a>



You can reach me at venesschmidt@ese.eur.nl

