

# R Markdown

04.05.2022, Data Science (SpSe 2022): T9

**Prof. Dr. Claudius Gräbner-Radkowsch**

**Europa-University Flensburg, Department of Pluralist Economics**

[www.claudius-graebner.com](http://www.claudius-graebner.com) | [@ClaudiusGraebner](https://twitter.com/ClaudiusGraebner) | [claudius@claudius-graebner.com](mailto:claudius@claudius-graebner.com)

# Prologue:

# Prologue

## Feedback and exercises

- XX of you filled out the feedback survey. Main take-aways:
  - TBA
- What were the main problems with the exercises?

# Goals for today

- I. Understand what R Markdown is about
- II. Write your first R Markdown document
- III. Render R Markdown documents into html and PDF format
- IV. Become aware of specific challenges for project management when using R Markdown

# What is R Markdown?

# What is R Markdown?

- R Markdown is a document format that allows you to write documents containing code of two languages:
  - R code to perform statistical analysis → we know this (almost 😊)
  - Markdown code to create formatted text using a plain text editor
- Markdown is readable in its source and, if rendered, allows for formatting, such as **bold** or *italic* fonts, tables, headings...

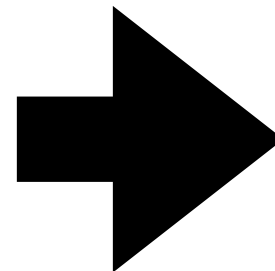
```
# Formatting in markdown
This is the source code. It is ugly,
but it is readable.

> Note: You can edit it on every editor,
and still can format your text.

You can have italics, you can have bold
texts, and much more!

## Tables

| Column 1 | Column 2 |
|:-----:|:-----:|
| Cell 1   | Cell 2   |
| Cell 3   | Cell 4   |
```



## Formatting in markdown

This is the source code. It is ugly, but it is readable.

**Note:** You can edit it on every editor, and still  
can format your text.

You can have *italics*, you can have **bold** texts, and much more!

## Tables

Column 1	Column 2
Cell 1	Cell 2
Cell 3	Cell 4

# What is R Markdown?

- R Markdown is a document format that allows you to write document comprising two languages:
  - R code to perform statistical analysis → we know this (almost 😊)
  - Markdown code to create formatted text using a plain text editor
- Markdown is readable in its source and, if rendered, allows for formatting, such as **bold** or *italic* fonts, tables, headings...
- We will not cover the basics of markdown → quite boring in a group
  - Please do the **interactive Markdown tutorial** on the course webpage
- Together, R and Markdown allows you to write formatted texts and conducts statistical analysis within one file
  - Perfect to make research accessible and reproducible

# What is R Markdown

The header -  
contains meta  
information

Markdown

Chunk  
options

R Chunks

```
1 ---
2 title: 'GDP and development'
3 author: Claudius
4 date: '2022-04-06'
5 output:
6   html_document:
7     theme: readable
8     highlight: tango
9     toc: true
10    toc_depth: 2
11    number_sections: true
12 ---
13
14 # Packages used
15
16 ```{r}
17 library(DataScienceExercises)
18 library(ggplot2)
19 ```
20
21 # GDP and development indicators
22
23 While there are convincing critiques of GDP as a measure of well-being, there
24 is also a clear relationship between GDP and socio-economic wellbeing indicators,
25 such as life expectancy:
26
27 ```{r, include=FALSE}
28 gdp_data <- DataScienceExercises::gdplifexp2007
29 head(gdp_data, 3)
30 ```
31
32 ```{r, echo=FALSE, warning=FALSE}
33 plot_preview <- ggplot2::ggplot(
34   data = gdp_data,
35   mapping = ggplot2::aes(
36     x = gdpPercap,
37     y = lifeExp,
38     size = pop,
39     fill = continent
40   )
41 ) +
42   ggplot2::geom_point(
43     shape=21, color="black", alpha=0.5
44   ) +
45   labs(
46     title = "Life expectancy and income per capita",
47     caption = "Note: size of bubbles represents population. Data: Gapminder",
48     x = "GDP per capita (int. Dollar)",
49     y = "Life expectancy in years"
50   ) +
51   ggplot2::scale_x_continuous(
52     labels = scales::number_format(scale = 0.001, suffix = "k")
53   ) +
54   ggplot2::scale_size_continuous(
55     guide = "none",
56     range = c(0.1, 24)
57   ) +
58   scale_fill_brewer(
59     palette = "Dark2"
60   ) +
61   ggplot2::theme_bw() +
62   theme(
63     legend.position = "bottom",
64     legend.title = ggplot2::element_blank(),
65     panel.border = ggplot2::element_blank(),
66     axis.line = ggplot2::element_line(colour = "grey"),
67     axis.ticks = ggplot2::element_line(colour = "grey")
68   )
69   plot_preview
70 ```
71
72 This relationship seems to be, however, heterogeneous across countries.
```

## GDP and development

Claudius

2022-04-06

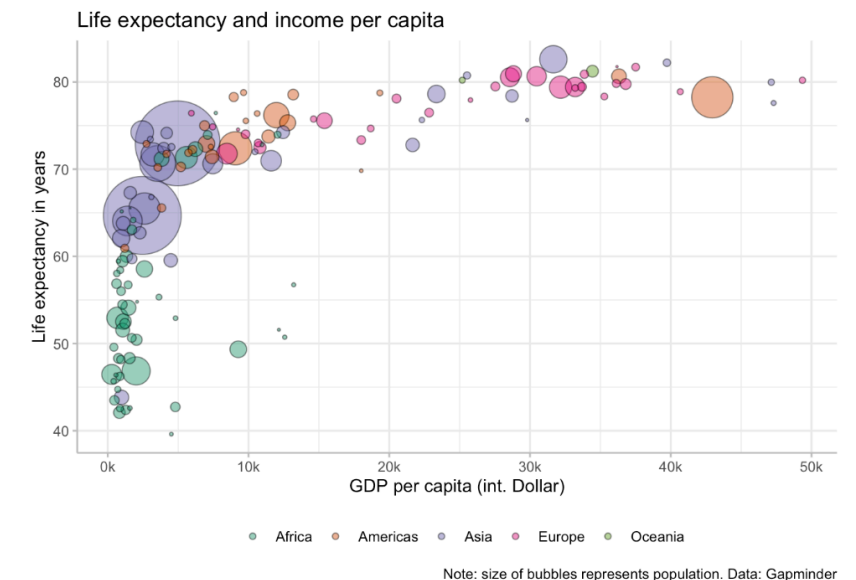
- 1 Packages used
- 2 GDP and development indicators
- 3 Trends of divergence

### 1 Packages used

```
library(DataScienceExercises)
library(ggplot2)
```

### 2 GDP and development indicators

While there are convincing critiques of GDP as a measure of well-being, there is also a clear relationship between GDP and socio-economic wellbeing indicators, such as life expectancy:



This relationship seems to be, however, heterogeneous across countries.

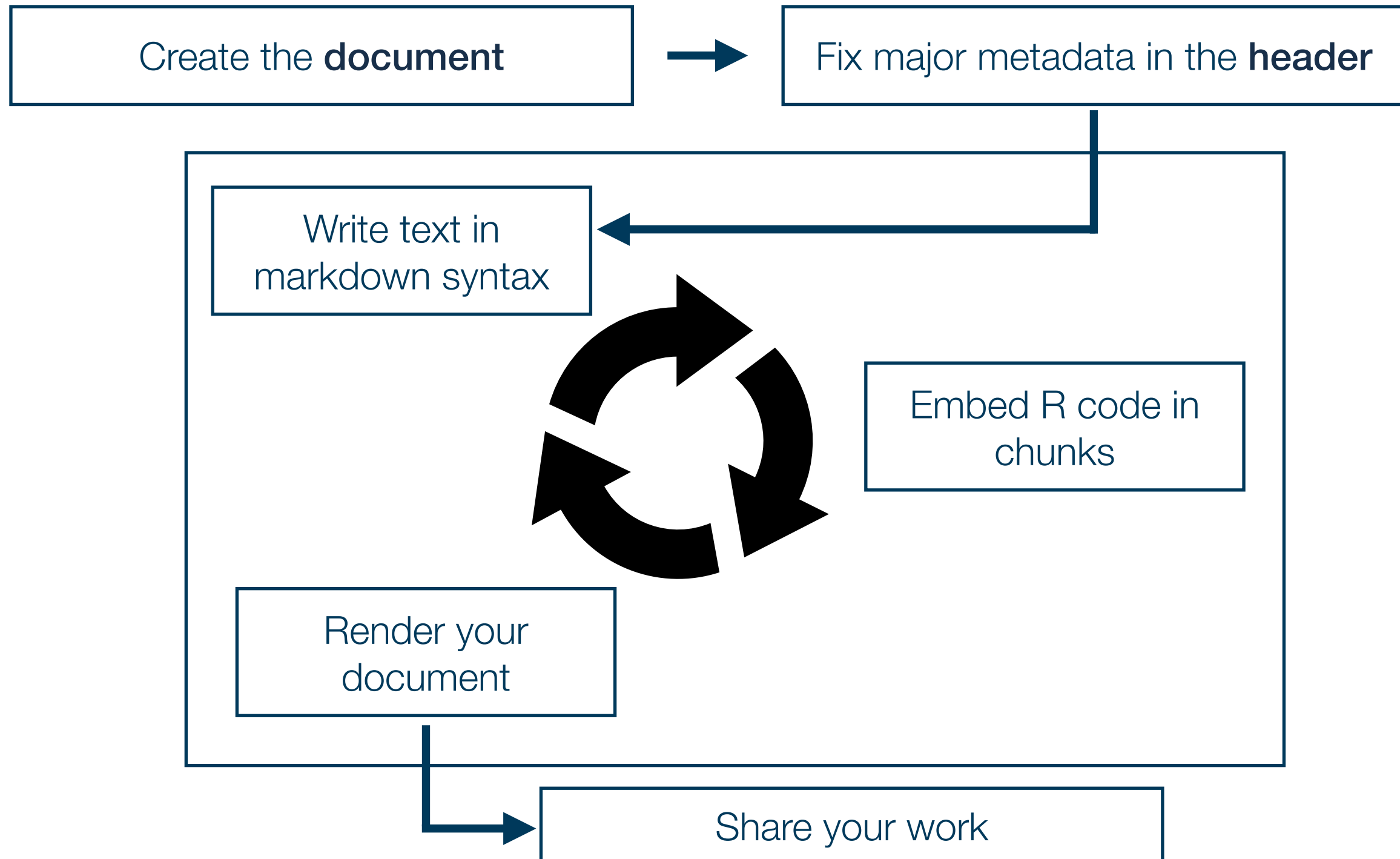


# What is R Markdown?

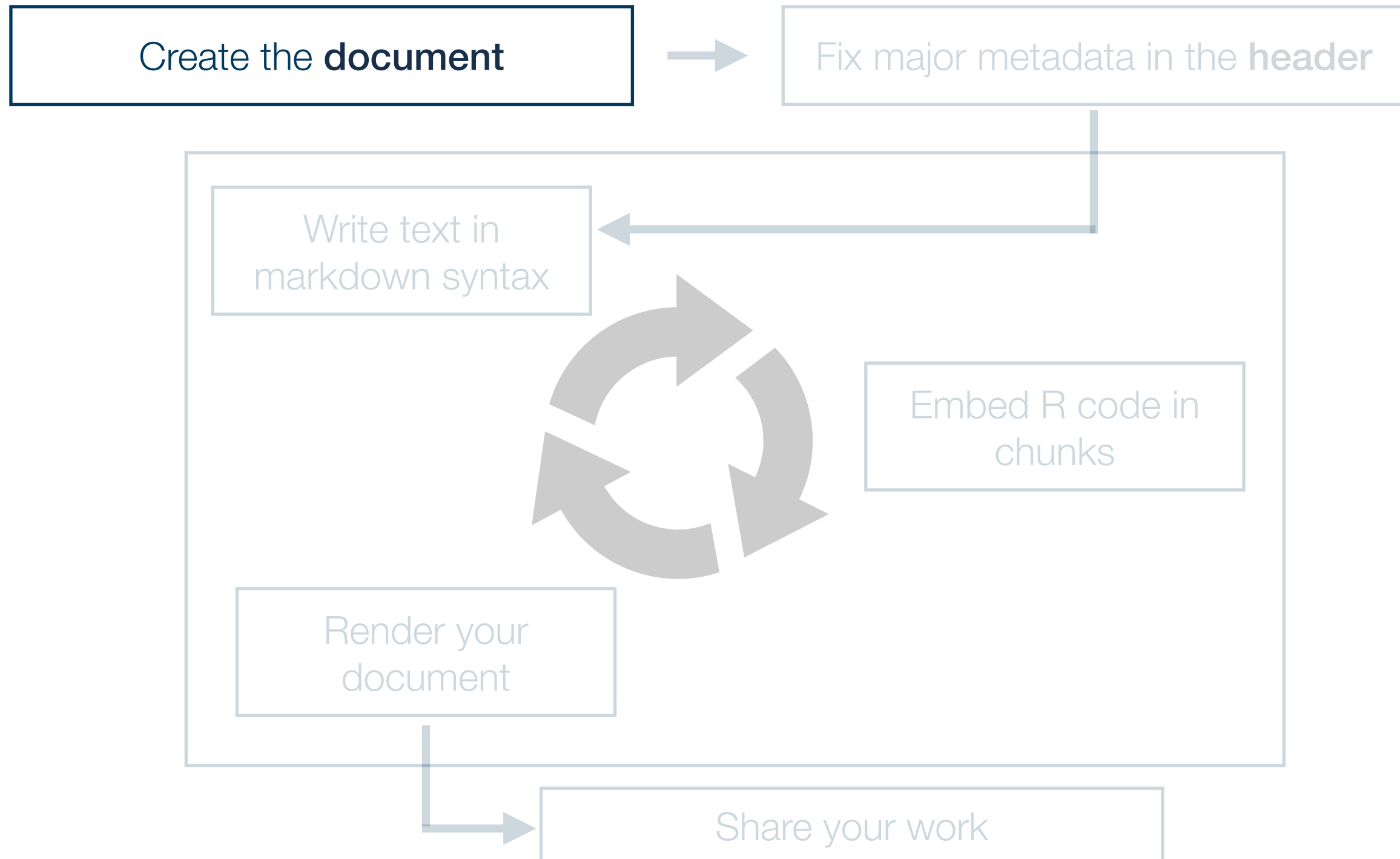
- R Markdown documents can be rendered to very different formats
  - HTML, PDF, Word, ...
- There are also many extensions available...
  - `bookdown` to write books
  - `blogdown` to create websites
  - `learnr` to create exercises
  - And many more...
- Basic **syntax the same** for all applications → this will be the focus here
  - Now go through the single steps required to get a R Markdown working

# R Markdown Step by Step

# R Markdown step-by-step



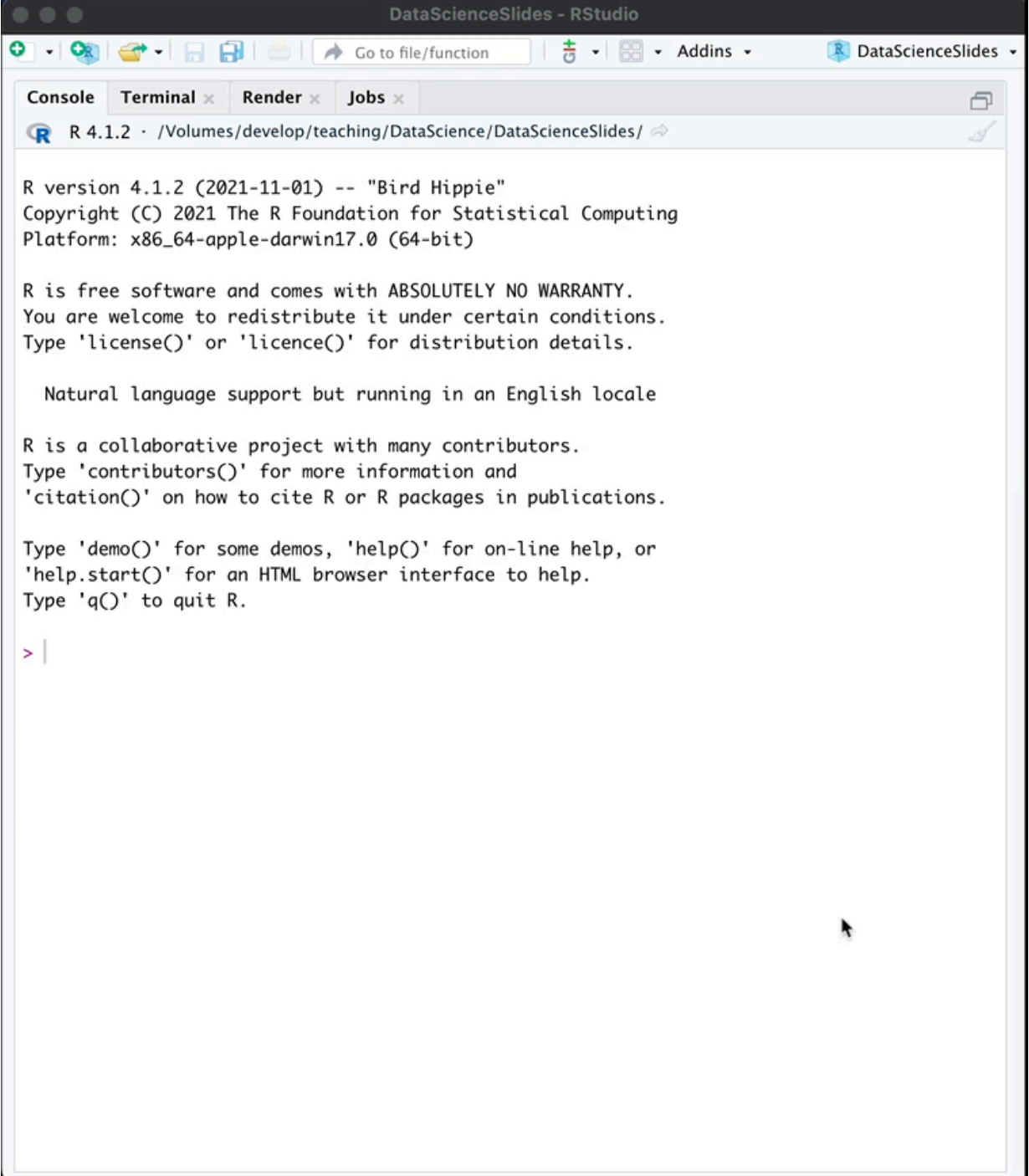
# R Markdown step-by-step



# R Markdown step-by-step

## 1. Creating the document

- First create a new Rmd document, and choose, if adequate, a template
- There are plenty of templates distributed as packages
  - For learning purposes its always best to start with a blank document
- After creating the document its best to save it immediately
  - Either in the subdirectory R, or in a separate top-level directory **markdown**



The screenshot shows the RStudio interface with the console window open. The title bar reads "DataScienceSlides - RStudio". The console shows the R startup message for version 4.1.2 (2021-11-01) -- "Bird Hippie". The message includes copyright information for The R Foundation for Statistical Computing and the platform "x86\_64-apple-darwin17.0 (64-bit)". It also states that R is free software with absolutely no warranty, and provides instructions on how to get help or quit R. The prompt ">|" is visible at the bottom of the console.

```
R version 4.1.2 (2021-11-01) -- "Bird Hippie"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

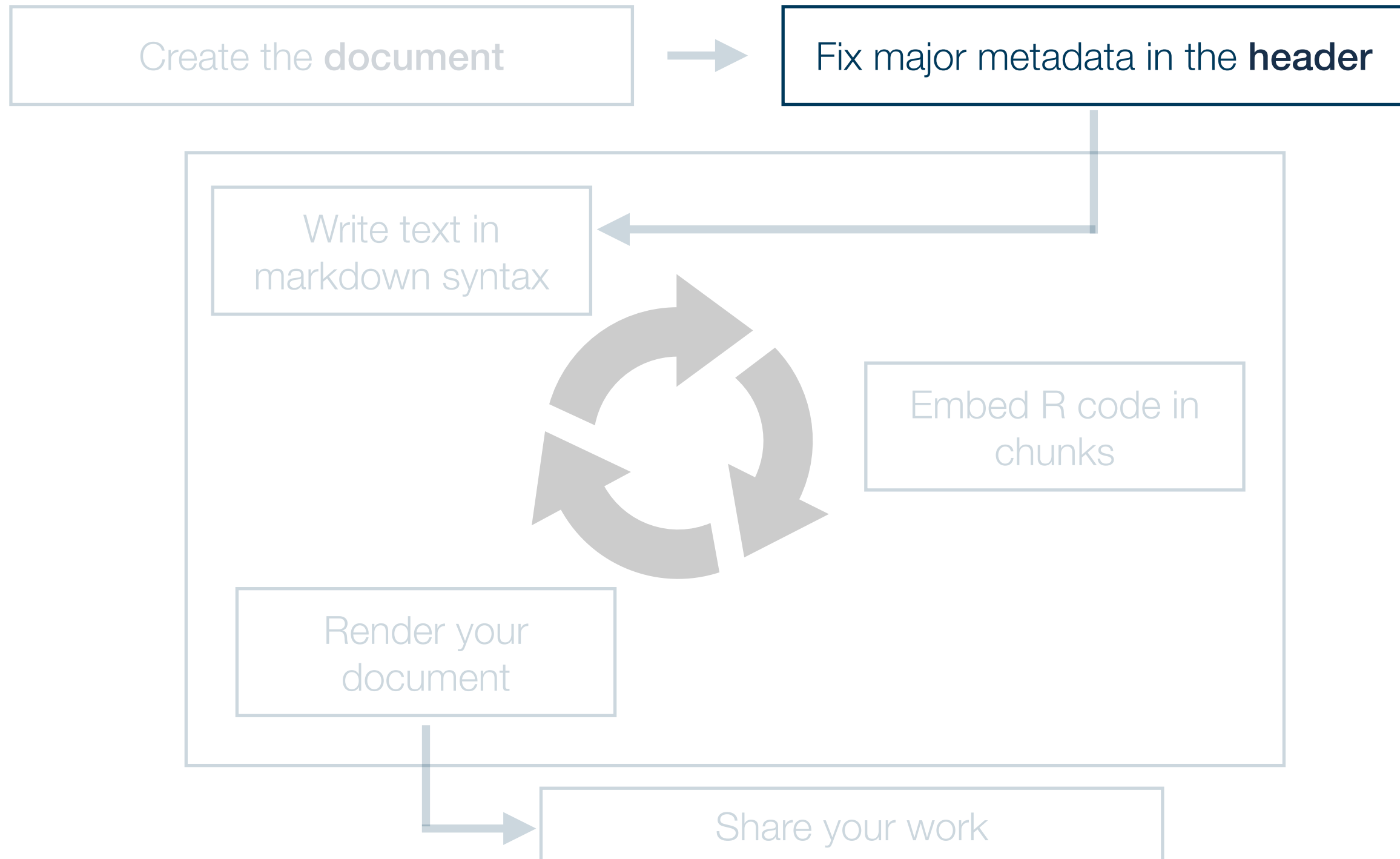
Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

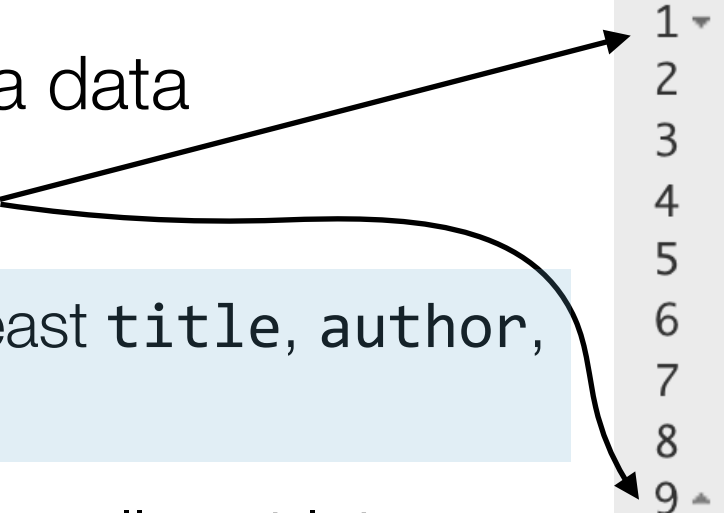
# R Markdown step-by-step



# R Markdown step-by-step

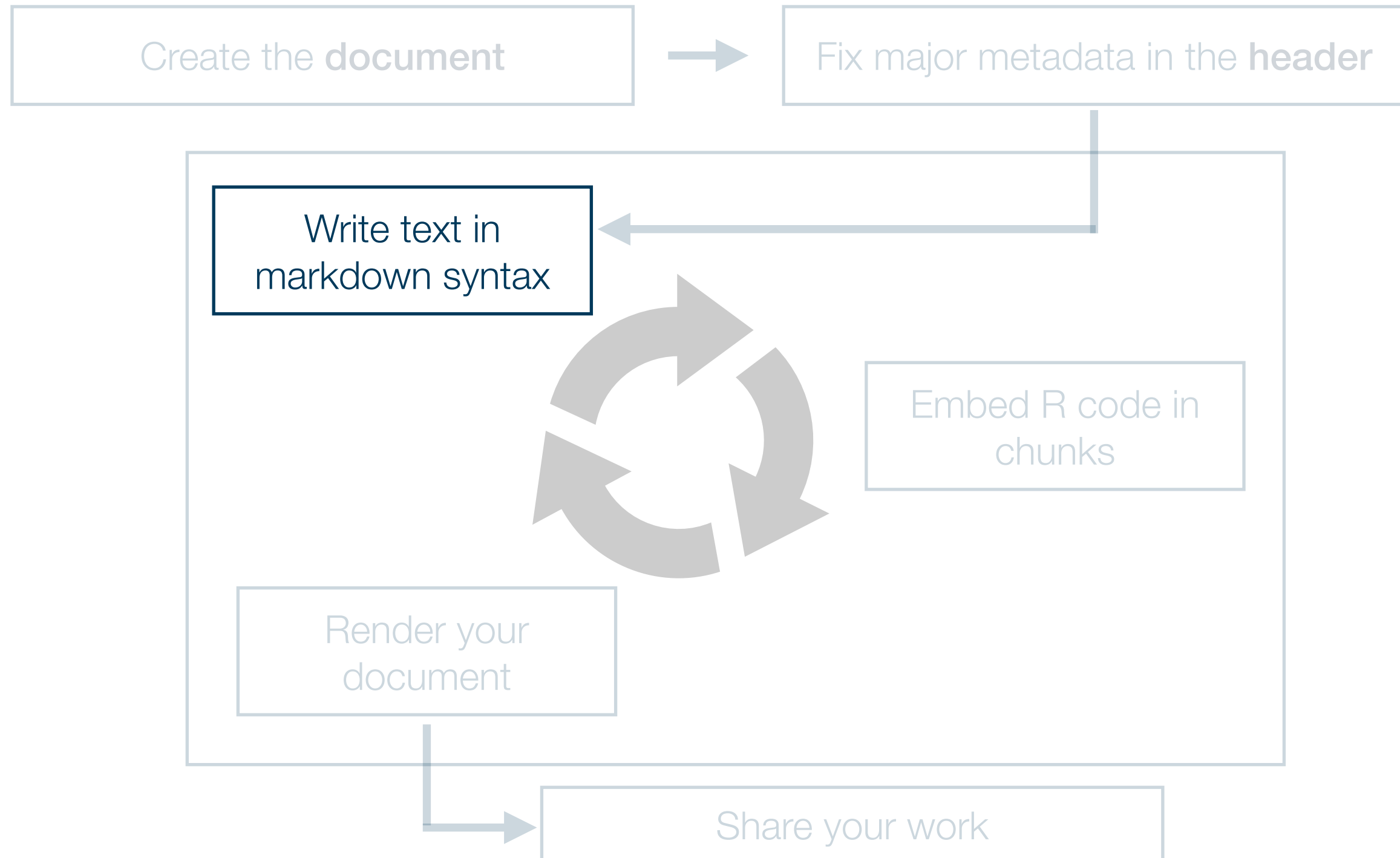
## 2. Specify the header

- The header contains meta data
  - It starts and ends with `---`
  - Usually you should set at least `title`, `author`, and `date`
- The output arguments are usually set later
  - Determines the function used by the markdown package to render output (`rmarkdown::html_document()`, `rmarkdown::pdf_document()`,...)
  - The more specific comments translate into arguments of `rmarkdown::*_document()`
- The headers is written in YAML -
  - There is a nice overview over the major keywords in the Markdown Coodbook (see further readings)



```
1 ---
2 title: "The title of my document"
3 author: "Claudius"
4 date: "3/23/2022"
5 output:
6   html_document:
7     toc: true
8     toc_float: true
9 ---
```

# R Markdown step-by-step





# R Markdown step-by-step

## 3. Write the main text

- Just write the text as you would do in any normal text editor
  - To format the text, follow the Markdown syntax
  - This syntax is best learned by example → do the interactive Markdown tutorial on the course homepage

01

### Introduction

Each lesson introduces a single Markdown concept with an example. When you see a red pulsing circle in the example, select to examine it for details.

After studying the example, try a few practice exercises with your new knowledge. Skip to any lesson at any time via the navigation controls. Experiment and have fun!

This tutorial is open source – [help us improve it!](#)

BEGIN LESSON →

WHAT IS MARKDOWN?

Intro ■  
Emphasis ■  
Paragraphs ■  
Headings ■  
Blockquotes ■  
Lists ■  
Links ■  
Images ■  
Code ■  
Nested Lists ■  
The End ■

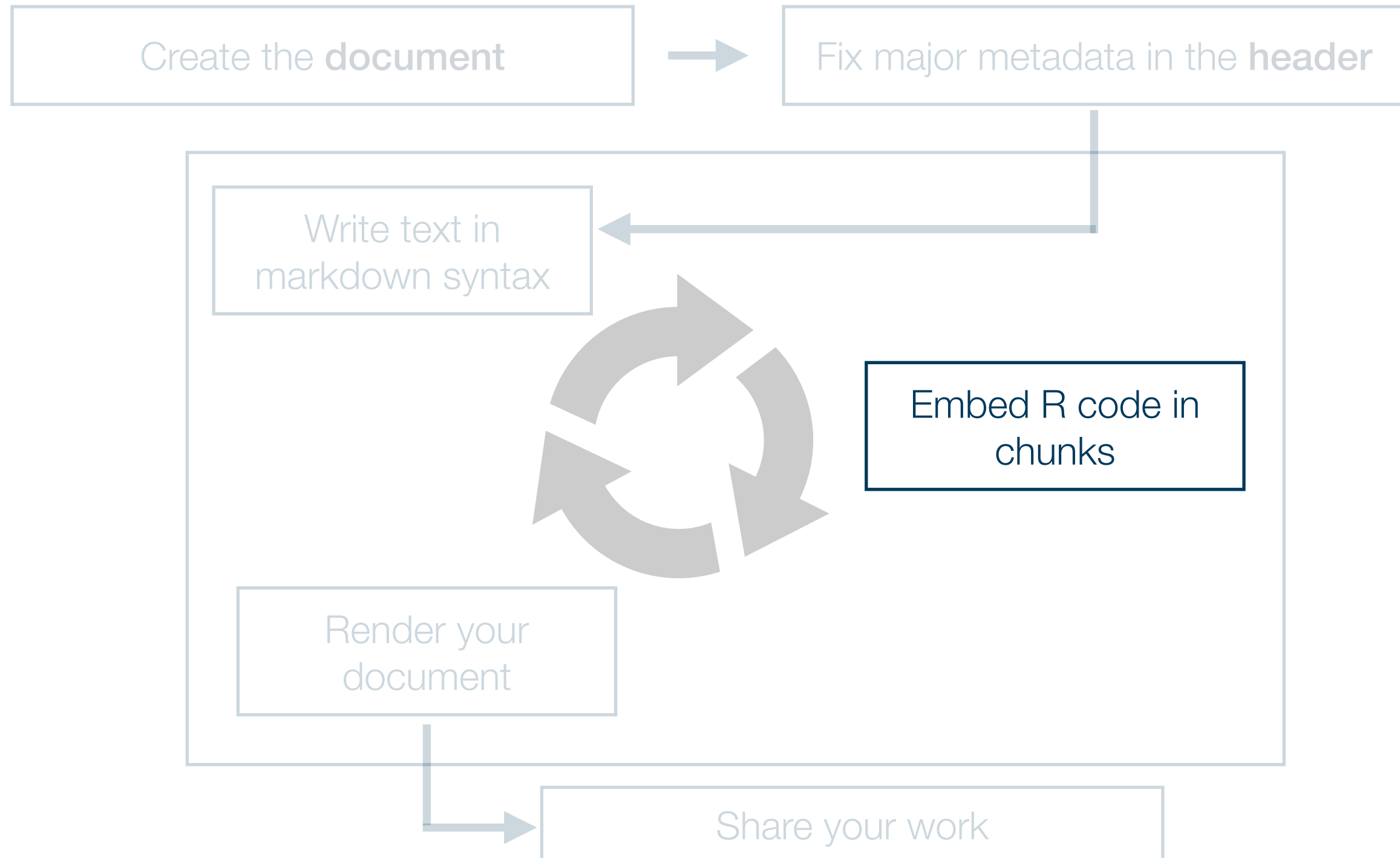
The screenshot shows the RStudio interface with a presentation titled "DataScienceSlides - RStudio". The main editor window displays an R Markdown document with the following content:

```
1 ---  
2 title: "The title of my document"  
3 author: "Claudius"  
4 date: "3/23/2022"  
5 output:  
6   html_document:  
7     toc: true  
8     toc_float: true  
9 ---  
10  
11
```

The bottom panel of RStudio shows the "Console" tab with the following output:

```
R 4.1.2 · /Volumes/develop/teaching/DataScience/DataScienceSlides/  
  
R version 4.1.2 (2021-11-01) -- "Bird Hippie"  
Copyright (C) 2021 The R Foundation for Statistical Computing
```

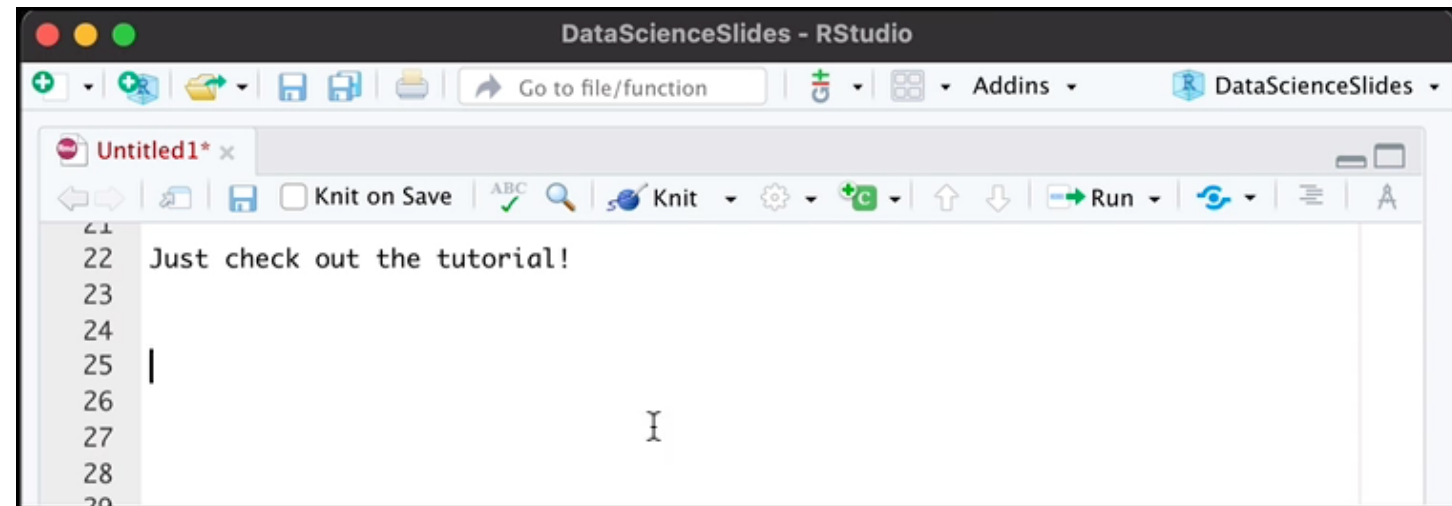
# R Markdown step-by-step



# R Markdown step-by-step

## 4. Embed R code into your document

- R code is written within **chunks**
  - Shortcut on Mac: `⌘ ⌥ i`
- Each chunk starts with a line ````{R}` and ends with `````



- Within the chunk you write R code just as you already know it
- You can refer to variables defined in previous chunks
  - You could in principle refer to all objects defined but you should not → would cause problems when rendering the file
- To execute the chunk or all previous chunks you might use the buttons:

# R Markdown step-by-step

## 4. Embed R code into your document

- How the R code gets shown and executed in the final document is controlled via the chunk options
- They can be added to the first line of the chunk:

```
```${r name, echo=FALSE}  
```
```

This is a chunk with `echo=TRUE` :

```
2 + 2
```

```
## [1] 4
```

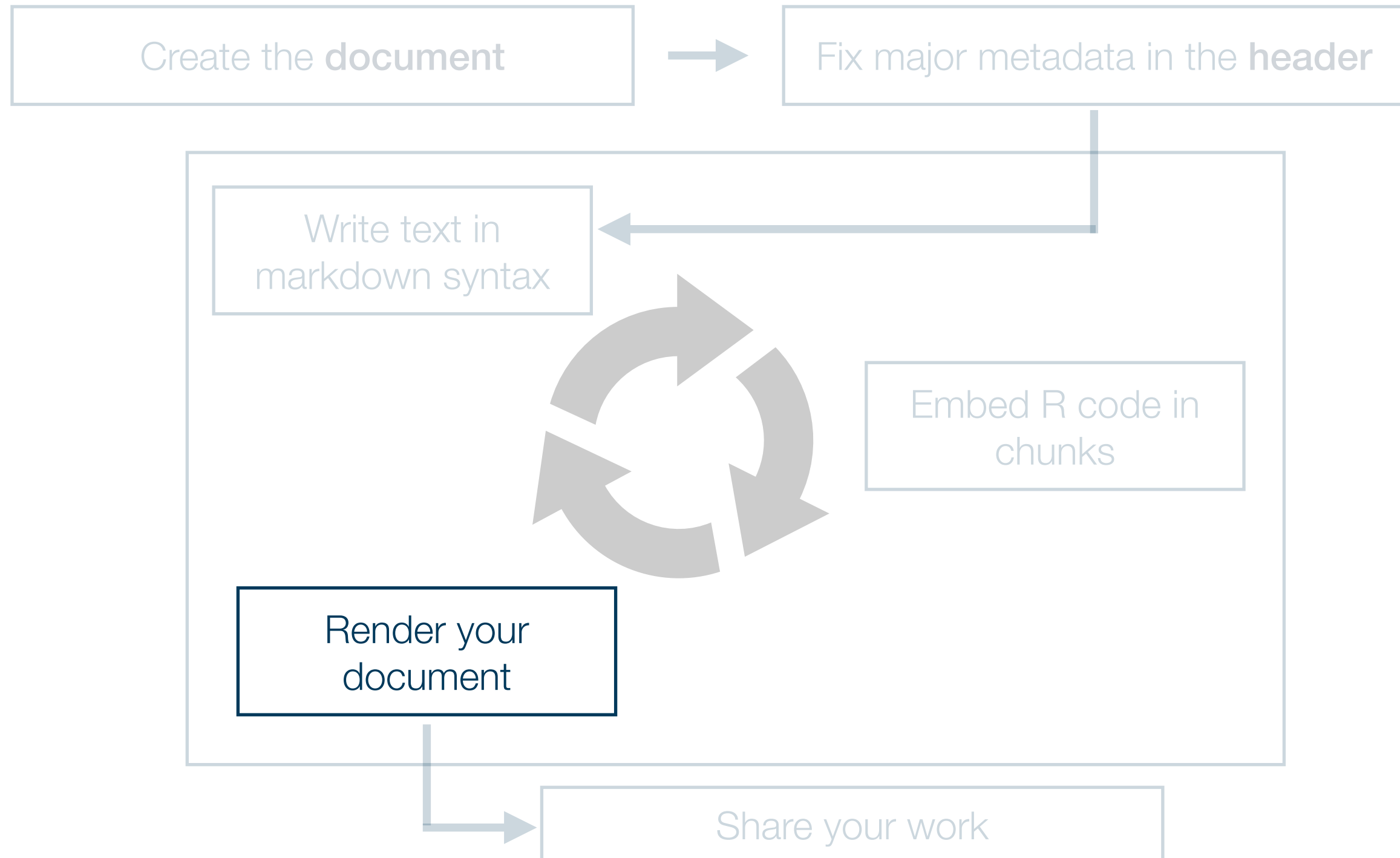
This is a chunk with `echo=FALSE` :

```
## [1] 4
```

- You can set default options for chunk options in the beginning of the document:
- A full list of all chunk options can be found here: <https://yihui.org/knitr/options/>

```
```${r setup, include=FALSE}  
knitr::opts_chunk$set(  
  echo = TRUE,  
  message = FALSE,  
  warning = FALSE  
)  
```
```

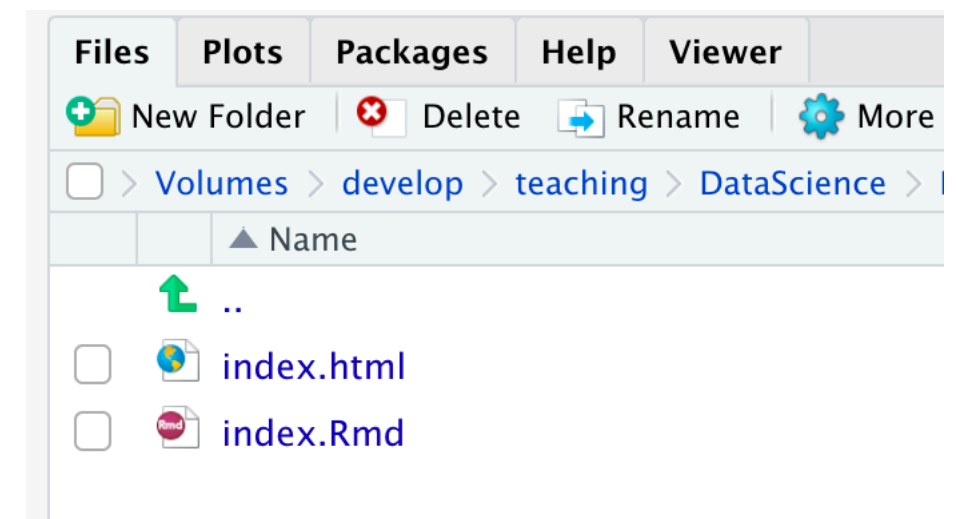
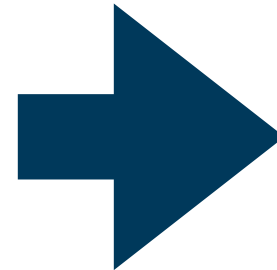
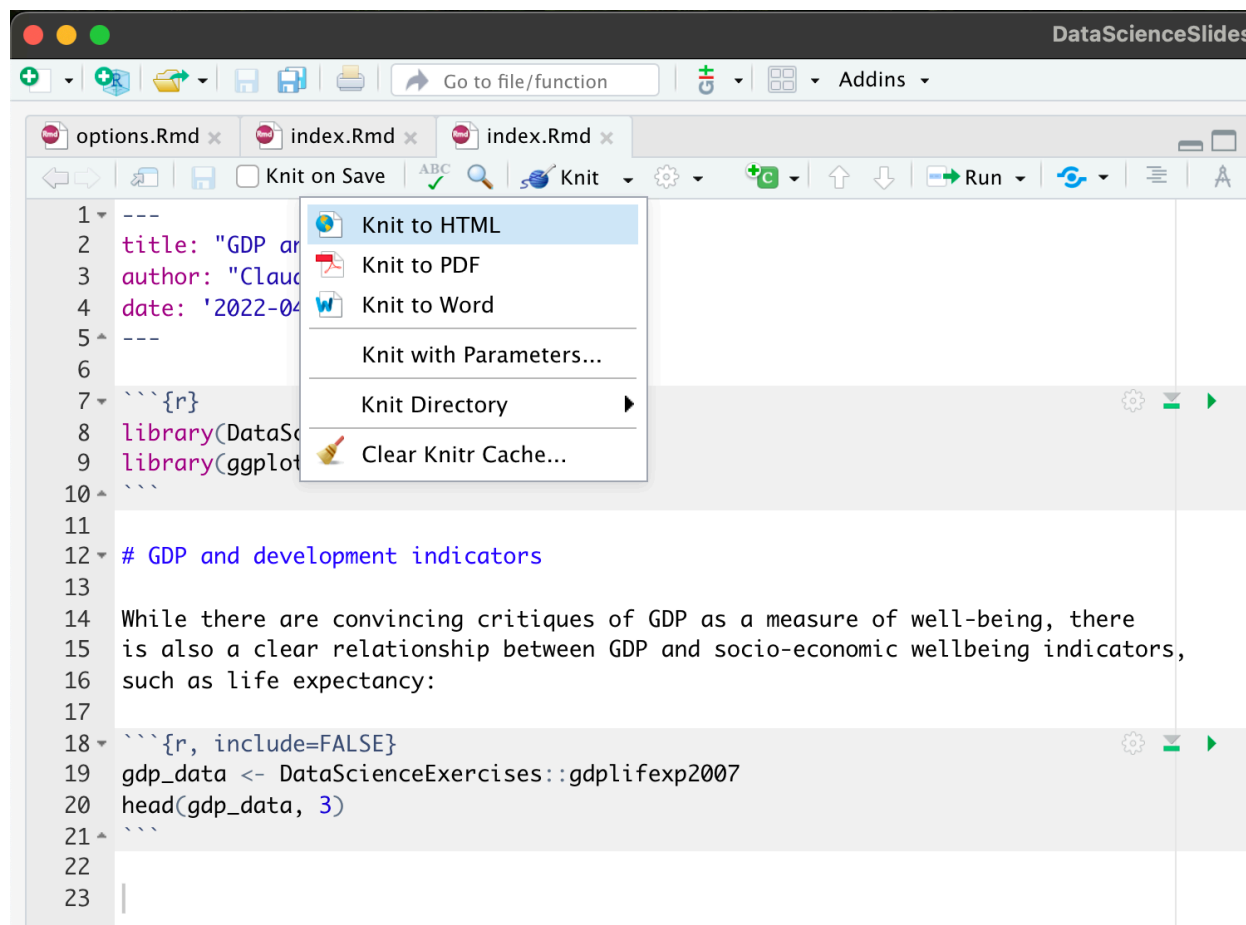
# R Markdown step-by-step



# R Markdown step-by-step

## 4. Render the documents

- This should in fact be done regularly during step 3
  - Otherwise its hard to identify the source of an error



- You can also render the document via a function directly
- You can adjust the output via the output options in the header

# R Markdown step-by-step

## 4. Render the documents - examples for output options

The indented options below only apply to documents that are rendered into html format using `rmarkdown::html_document()`

General property sets (fonts, paragraphs, etc.)

How to color the code in chunks

Specification of the table of contents

Activate the numbering of sections

```
1 ---
2 title: 'GDP and development'
3 author: Claudius
4 date: '2022-04-06'
5 output:
6   html_document:
7     theme: readable
8     highlight: tango
9     toc: true
10    toc_depth: 2
11    number_sections: true
12 ---
```

**For a complete overview over output options and possible themes you should check the internet for the desired output format!**

# R Markdown step-by-step

## 4. Render the documents - implications of output options

```
1 ---
2 title: 'GDP and development'
3 author: Claudius
4 date: '2022-04-06'
5 ---
6
```



### GDP and development

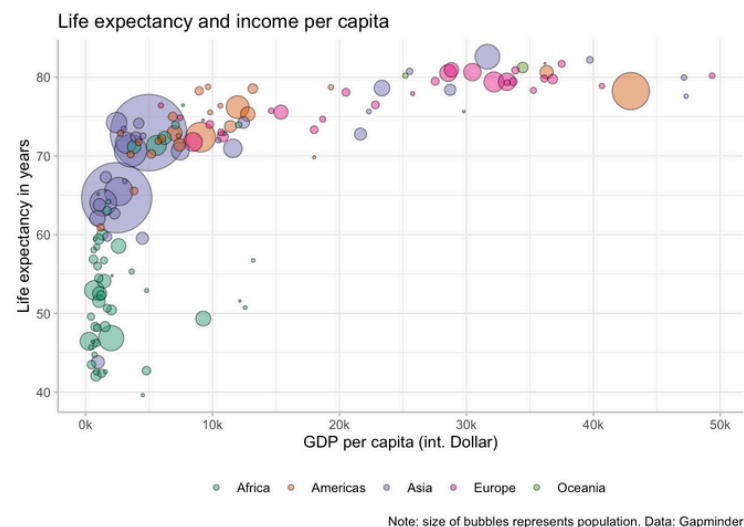
Claudius  
2022-04-06

#### Packages used

```
library(DataScienceExercises)
library(ggplot2)
```

### GDP and development indicators

While there are convincing critiques of GDP as a measure of well-being, there is also a clear relationship between GDP and socio-economic wellbeing indicators, such as life expectancy:



This relationship seems to be, however, heterogeneous across countries.

```
1 ---
2 title: 'GDP and development'
3 author: Claudius
4 date: '2022-04-06'
5 output:
6   html_document:
7     theme: readable
8     highlight: tango
9     toc: true
10    toc_depth: 2
11    number_sections: true
12 ---
```



### GDP and development

Claudius  
2022-04-06

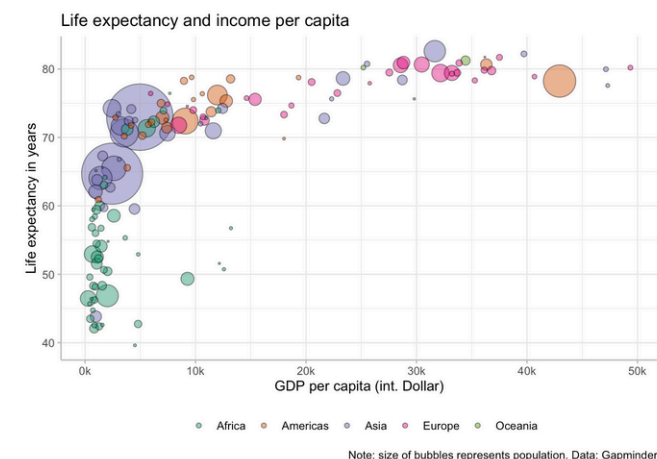
- 1 Packages used
- 2 GDP and development indicators
- 3 Trends of divergence

#### 1 Packages used

```
library(DataScienceExercises)
library(ggplot2)
```

### 2 GDP and development indicators

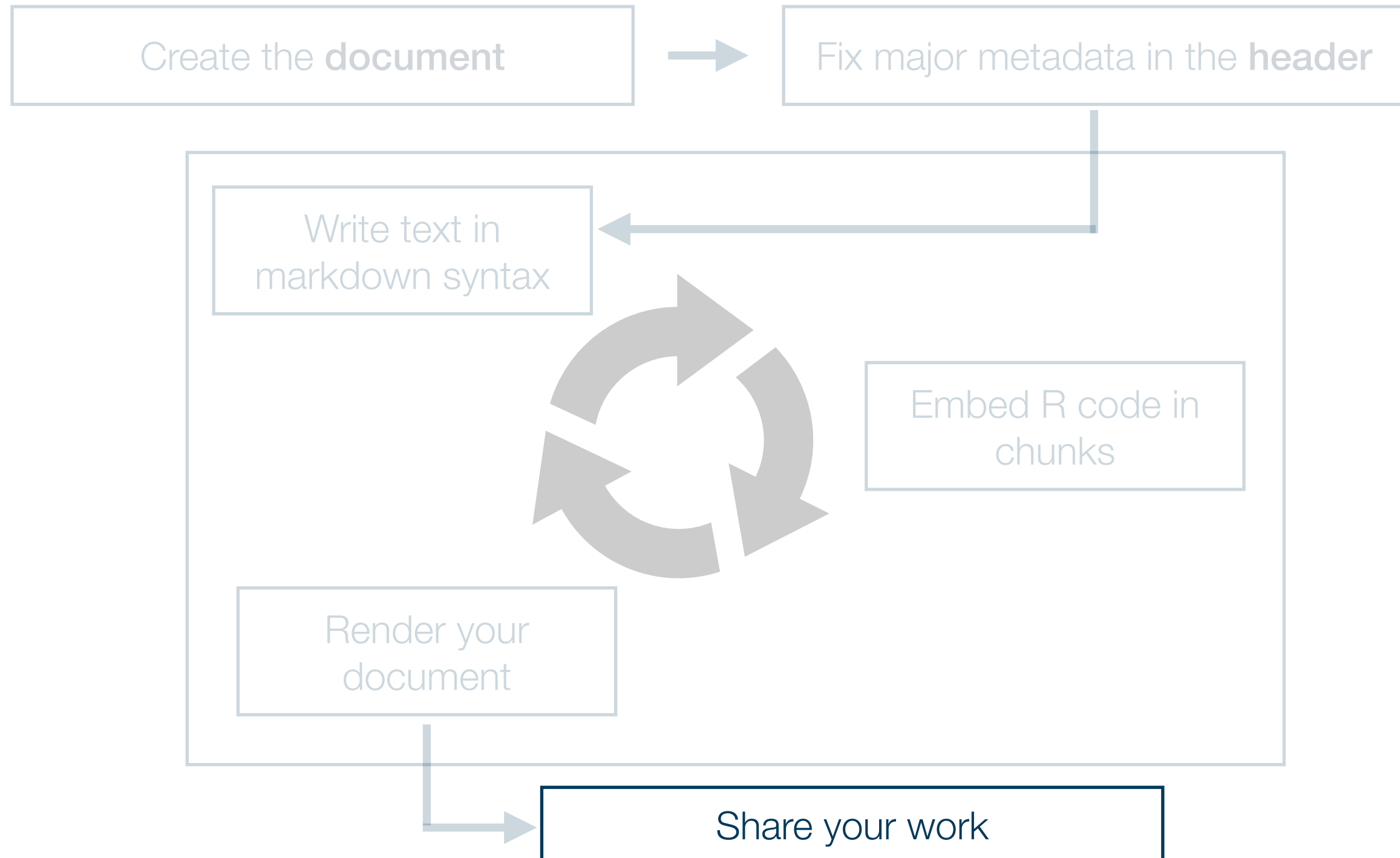
While there are convincing critiques of GDP as a measure of well-being, there is also a clear relationship between GDP and socio-economic wellbeing indicators, such as life expectancy:



This relationship seems to be, however, heterogeneous across countries.



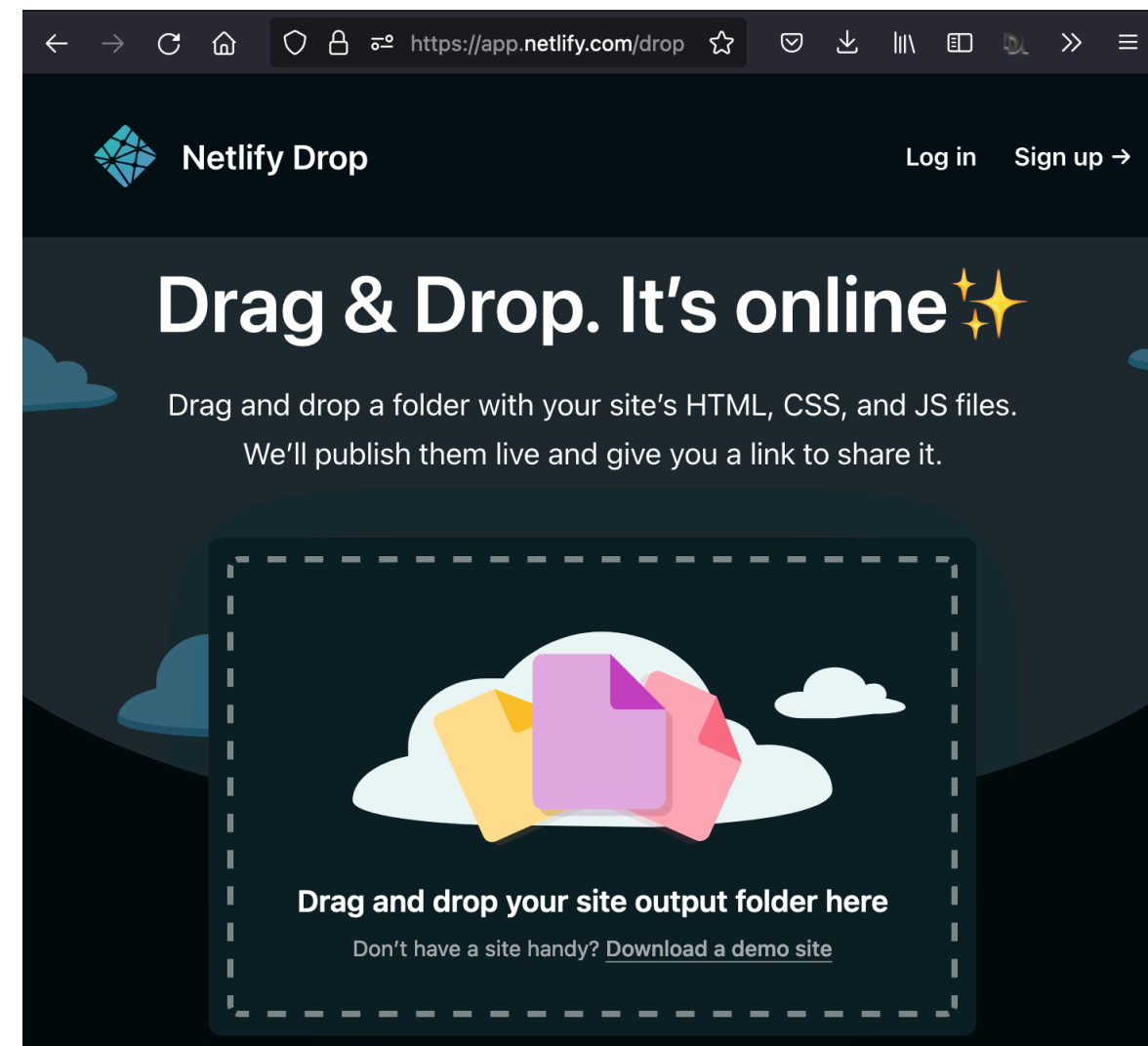
# R Markdown step-by-step



# R Markdown step-by-step

## 5. Share your work

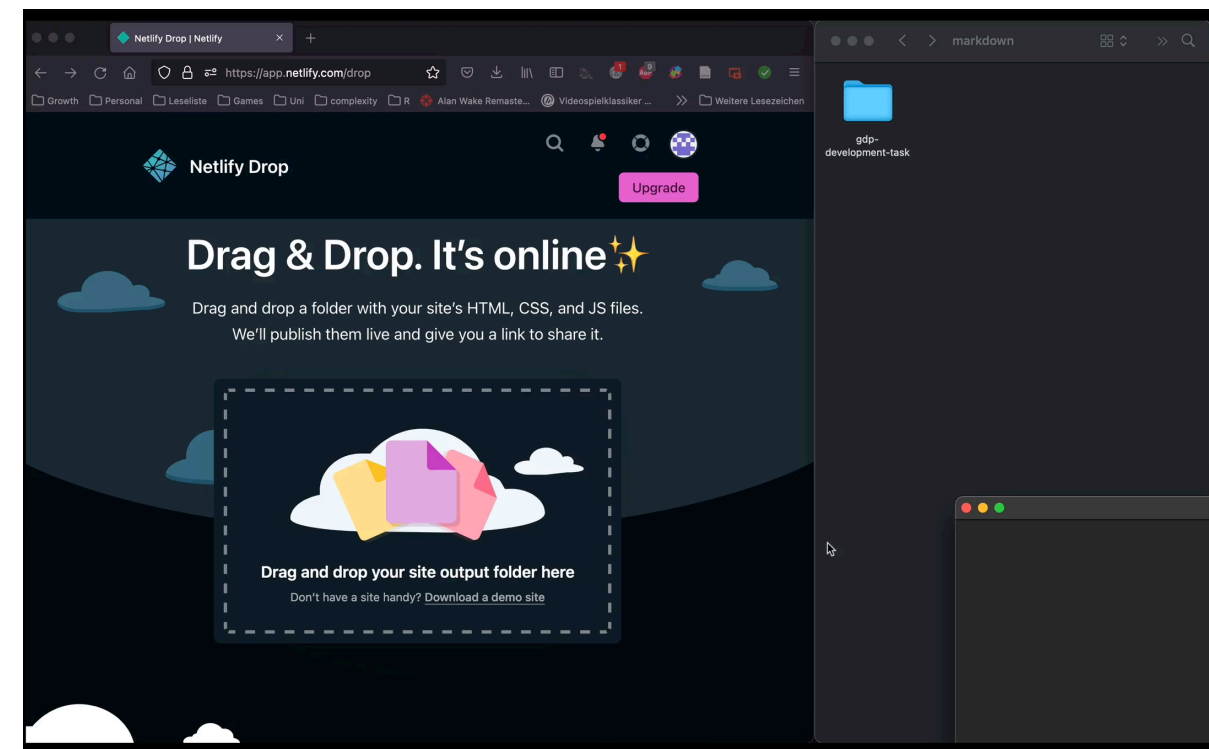
- Many different possibilities → check the further readings
  - Key question is often about the format
- Here we consider a very straightforward solution for html-documents
  - Netlify Drop: <https://app.netlify.com/drop>
- Prerequisites:
  - You create an html document
  - All relevant output data is in one folder
  - You produce an output `index.html`
- Then its a nice way to distribute your document quickly



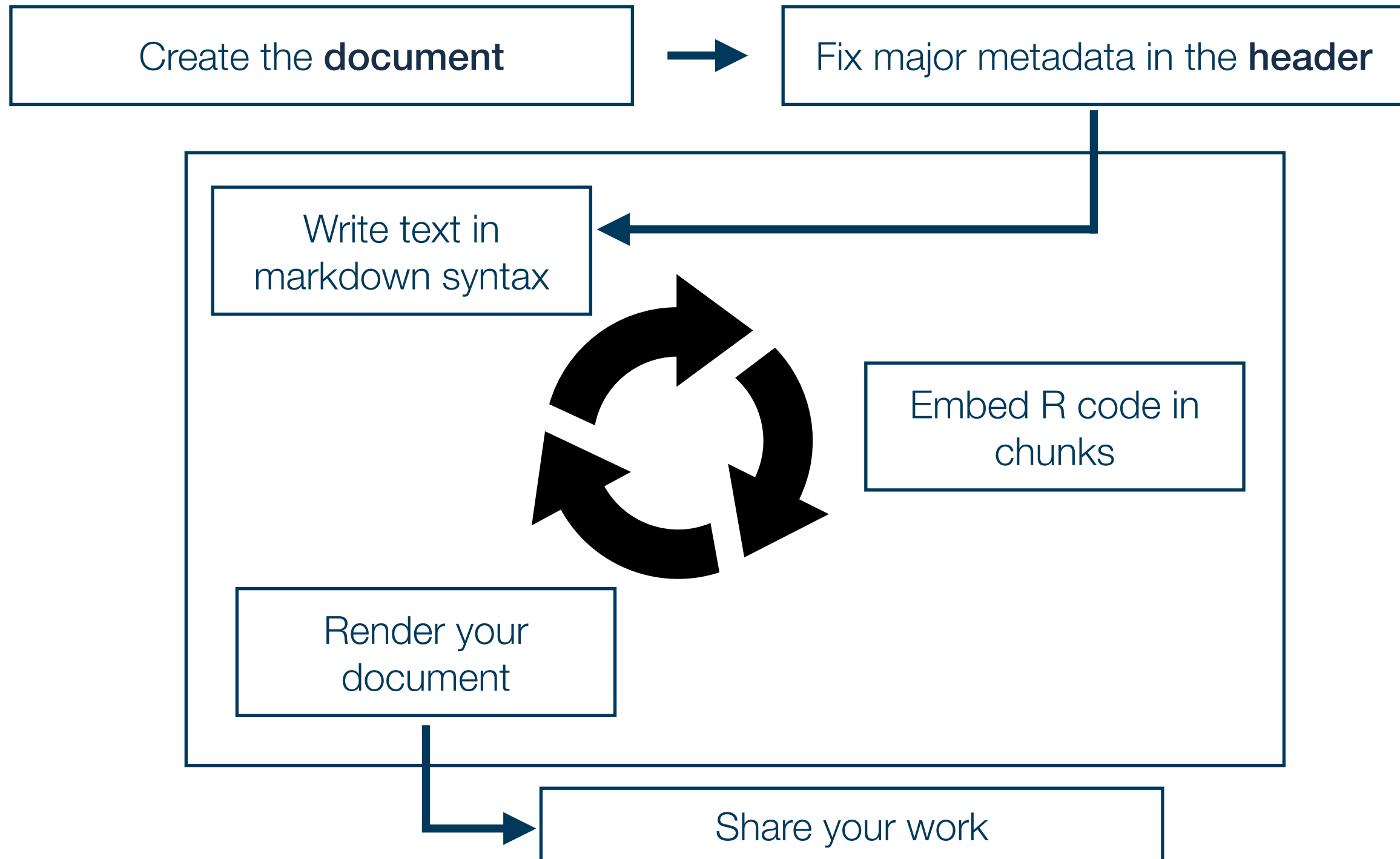
# R Markdown step-by-step

## 5. Share your work

- Many different possibilities → check the further readings
  - Key question is often about the format
- Here we consider a very straightforward solution for html-documents
  - Netlify Drop: <https://app.netlify.com/drop>
- Prerequisites:
  - You create an html document
  - All relevant output data is in one folder
  - You produce an output `index.html`
- Then its a nice way to distribute your document quickly
  - If you are registered you can also choose a custom URL and much more



# R Markdown step-by-step



# Practice!

- Write and render a document!
- The text should include...
  - ...a heading of level 1 and 2
  - ...a text body
  - ...a plot made with ggplot2
- Also add meta data on the **author**, **title**, and **date**
- Render the document into **html** and **PDF**
- Deploy the document via Netlify Drop and post the link on Moodle

# Some final remarks on project organisation

- The use of the **here** package is especially important when writing R markdown documents
  - When rendering an Rmd document, your computer is **not** using your current working directory
  - Rather, the **working directory** is set to the **location of the Rmd file**
  - This means that you cannot copy-paste code from R scripts that contains relative paths, except you use the **here**-package
- Its usually a good idea to put Rmd-files either into the folder R or create a separate top level directory markdown

# Avoid common mistakes

# Avoiding common markdown mistakes

- There are some very common mistakes
  - These screw up your document considerably and make it painful to read...
  - ...but are actually very easy to avoid
- Thus, after completing a markdown document, always look at the rendered version
  - Check whether any of the problems below still exists and eliminate them when necessary

## Task:

Download **DesasterMarkdown.pdf** from the course homepage and collect the aspects that bother you the most!



# The black list of markdown turpitudes

| Problem | Solution |
|---------|----------|
|         |          |
|         |          |
|         |          |
|         |          |
|         |          |
|         |          |

- Upgrade: make tables pretty with `knitr::kable()` and `kableExtra`
- See the overview over all chunk options here: <https://yihui.org/knitr/options/>

# Summary & outlook

# Summary and outlook

- R Markdown is a document format that allows you to write documents that contain...
  - R code to perform statistical analysis
  - Markdown code to create formatted text using a plain text editor
- This allows you to do and describe your statistical analysis within one consistent document → makes research fully transparent and reproducible
- We covered the main steps of working with R Markdown:
  - (1) create the documents, (2) set meta-data in the header, (3) write text, (4) embed R code, (5) render the document, and (6) share the final result
  - To learn markdown, just do the interactive tutorial
- There are a few mistakes that might easily screw up your document, but are just as easy to avoid

# Summary and outlook

- We have now covered all the fundamentals of data preparation
- This was the most important part of the lecture since these are the tools you *always* need
- Now we will turn to some more advanced programming techniques and statistical applications: theory and modelling

## Tasks until next week:

1. Fill in the **quick feedback survey** on Moodle
2. Read the **mandatory readings** posted on the course page
3. Complete the **interactive Markdown tutorial** linked on the course page
4. Do the **exercises** provided on the course page and **discuss problems** and difficulties via the Moodle forum