

The "Rmd first" method: when projects start with documentation

Use'R Toulouse 2019

Sébastien Rochette



This presentation on Github: [statnmap/prez](https://github.com/statnmap/prez)

Sébastien

Data Scientist, R expert.

- <https://rtask.thinkr.fr>
- <https://github.com/ThinkR-open>
- https://twitter.com/thinkr_fr



**Vincent
Guyader**

Codeur Fou,
formateur et expert
logiciel R



**Diane
Beldame**

Dompteuse de
dragons données,
formatrice logiciel R



**Colin
Fay**

Data scientist
et R hacker



**Sébastien
Rochette**

Modélisateur,
Formateur R, Joueur
de cartographies



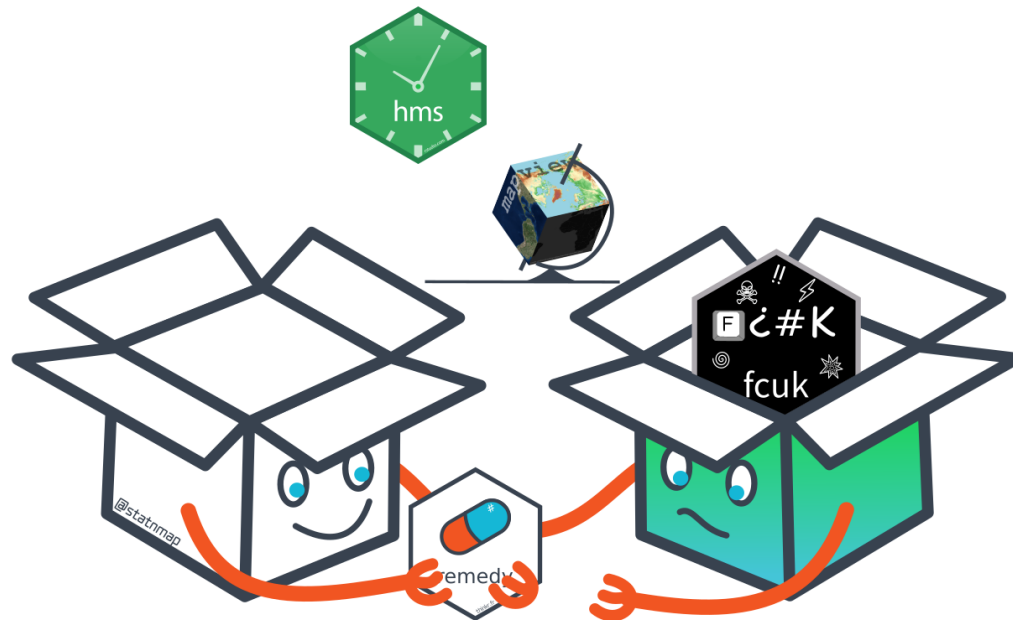
**Cervan
Girard**

Le nouveau

Why do you need documentation ?

Prevent code sickness

- You take back your project 6 months later
- You share your scripts with your young padawan
- A colleague has to run your analysis when you are on a sick leave
- Your client/boss asks for an update on your work
- You want to share your work with the world



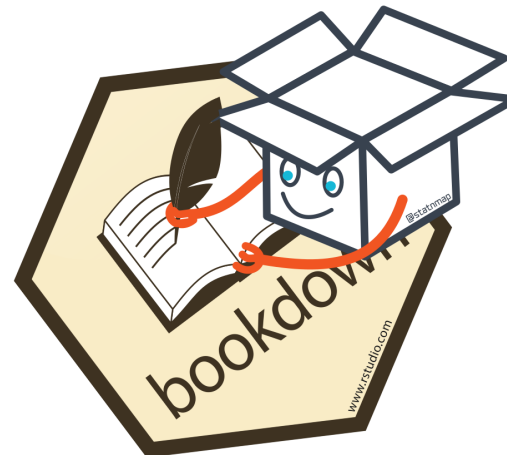
Why is documentation the last thing you do?



Why is documentation the last thing you do?

Start with the documentation !

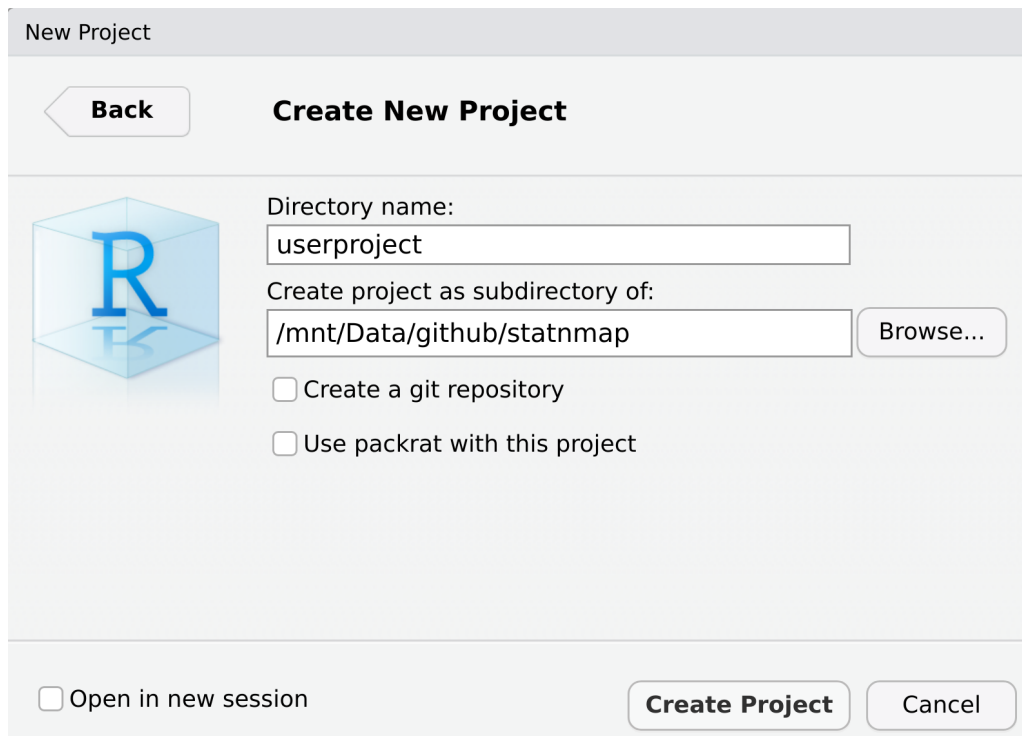
Rmd first !



Let's try a workflow for data analysis

Always work in a project (one analysis = one project)

1. New *baby* project

A screenshot of the 'New Project' dialog box in RStudio. The window title is 'New Project'. It has a 'Back' button on the left. The main heading is 'Create New Project'. On the left is the R logo. The 'Directory name:' field contains 'userproject'. The 'Create project as subdirectory of:' field contains '/mnt/Data/github/statnmap' and has a 'Browse...' button to its right. There are two checkboxes: 'Create a git repository' and 'Use packrat with this project', both of which are unchecked. At the bottom left is an unchecked checkbox 'Open in new session'. At the bottom right are 'Create Project' and 'Cancel' buttons.

New Project

Back Create New Project

Directory name:
userproject

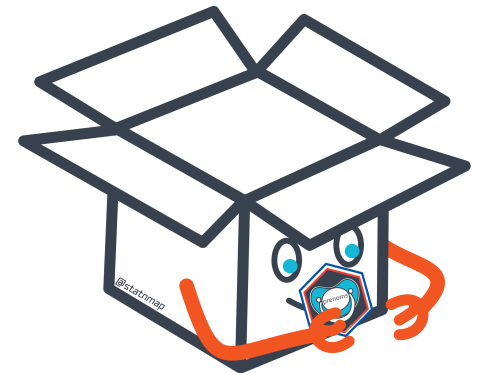
Create project as subdirectory of:
/mnt/Data/github/statnmap Browse...

Create a git repository

Use packrat with this project

Open in new session

Create Project Cancel



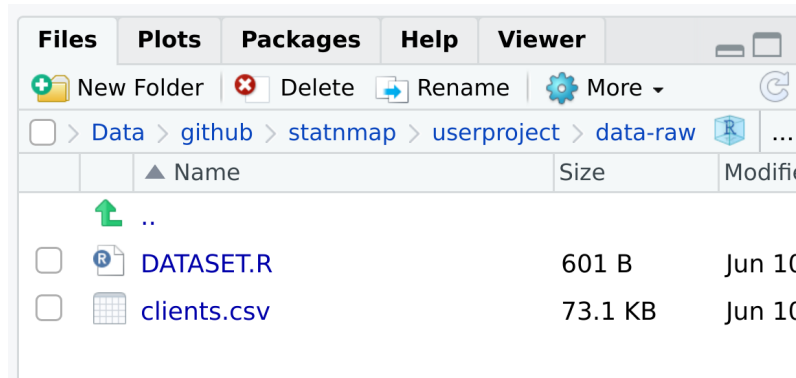
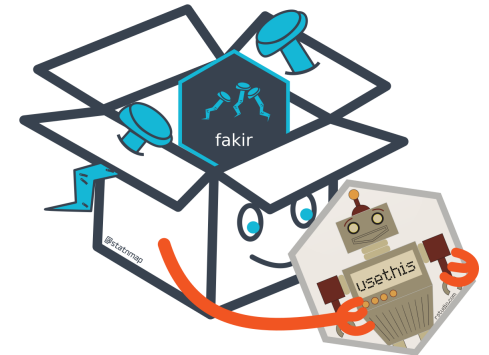
⚠ These steps are to illustrate our approach, this is not exactly the workflow we will recommend. Stay focus until the end! ⚠

Let's try a workflow for data analysis

Always work in a project (one analysis = one project)

2. Add some data in a folder called 'data-raw'

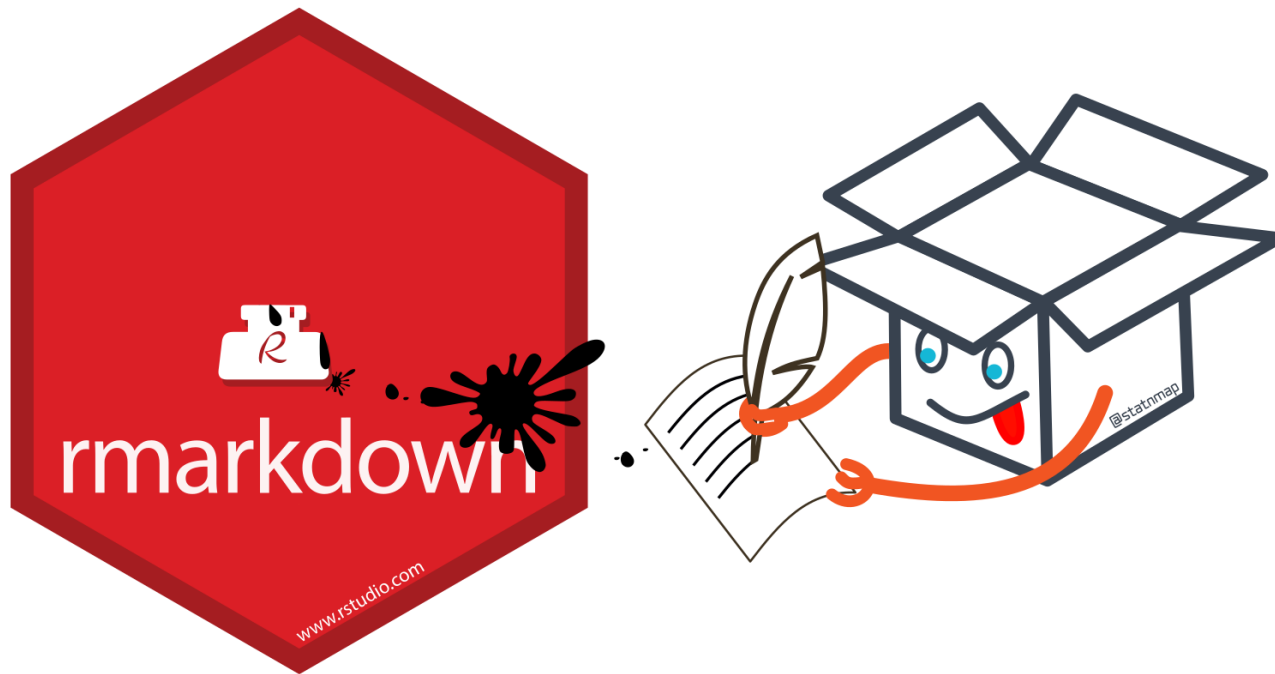
- use `usethis::use_data_raw()`
 - Add your dataset to analyse in "data-raw"
 - e.g. client database (issued from `{fakir}`)



Let's try a workflow for data analysis

Always work in a project (one analysis = one project)

3. Create a Rmd file at the root of the project

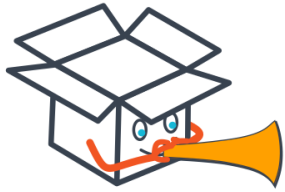


Say what you will do...

...do what you said

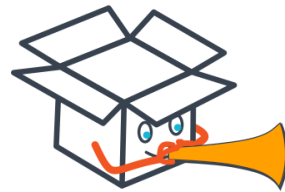


Say what you will do...



```
DATASET.R x my-analysis.Rmd x
Knit
Insert Run

24 ▾ ## Read a client database
25
26 Dataset has been built using package {fakir} available on Github with
   `remotes::install_github("ThinkR-open/fakir")`. It is saved as a CSV
   file in this project.
27
28 ▾ ```{r, message=FALSE}
29 dataset_path <- here("data-raw/clients.csv")
30 clients <- read_csv(dataset_path)
31 ```
32
33 ▾ ## Table by department
34
35 - Create a function to filter on a particular department
36
37 ▾ ```{r}
38 filter_by_dpt <- function(x, dpt) {
39     filter(x, id_dpt == dpt)
40 }
41 # Examples
42 filter_by_dpt(clients, dpt = 11)
43 ```
23:1 Load packages R Markdown
```



...do what you said



```
DATASET.R x my-analysis.Rmd x
Knit
Insert Run
24 ## Read a client database
25
26 Dataset has been built using package {fakir} available on Github with
   `remotes::install_github("ThinkR-open/fakir")`. It is saved as a CSV
   file in this project.
27
28 ```{r, message=FALSE}
29 dataset_path <- here("data-raw/clients.csv")
30 clients <- read_csv(dataset_path)
31 ```
32
33 ## Table by department
34
35 - Create a function to filter on a particular department
36
37 ```{r}
38 filter_by_dpt <- function(x, dpt) {
39   filter(x, id_dpt == dpt)
40 }
41 # Examples
42 filter_by_dpt(clients, dpt = 11)
43 ```
23:1 Load packages R Markdown
```



This is not an hex !



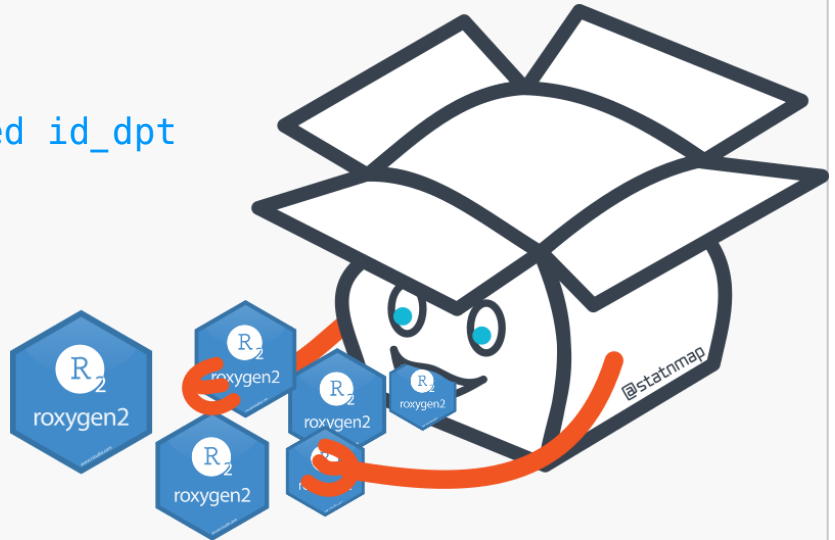
Document your function

- Document the function using roxygen skeleton
- **You already have an example in the Rmd !**

```

#' Filter by department
#'
#' @param x dataframe with column named id_dpt
#' @param dpt department number
#'
#' @return a filtered dataframe
#'
#' @examples
#' \dontrun{
#'   filter_by_dpt(clients, dpt = 11)
#' }
filter_by_dpt <- function(x, dpt) {
  filter(x, id_dpt == dpt)
}

```



Create some tests on this function

- Use the example to build some tests

```
42  
43 - Write a test for function `filter_by_dpt`  
44  
45 ```{r}  
46 test <- filter_by_dpt(clients, dpt = 11)  
47 testthat::expect_is(test, "data.frame")  
48 testthat::expect_equal(nrow(test), 10)  
49 ```  
50
```

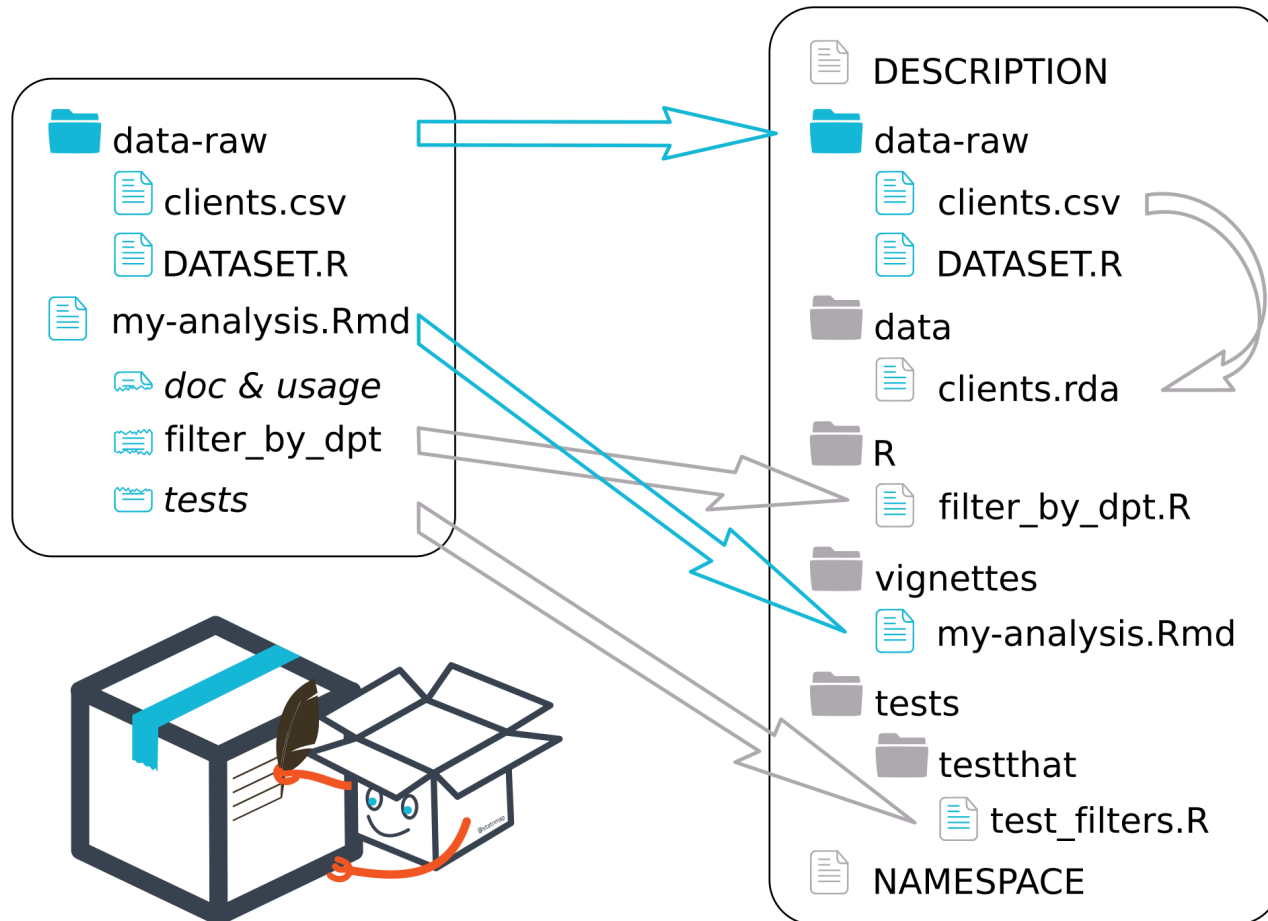


Where are we ?



We are (almost) in a package!

Clean your Rmd to see it



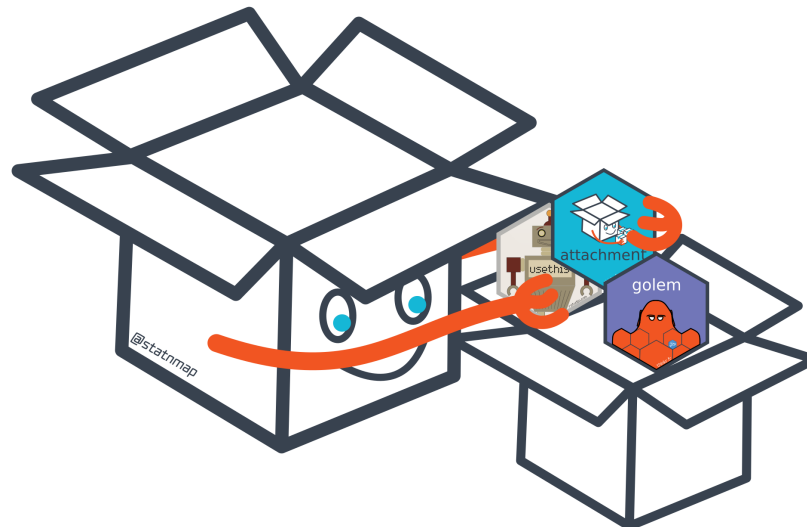
What's next ?

Package

- `usethis::create_package("myproject")`

Document

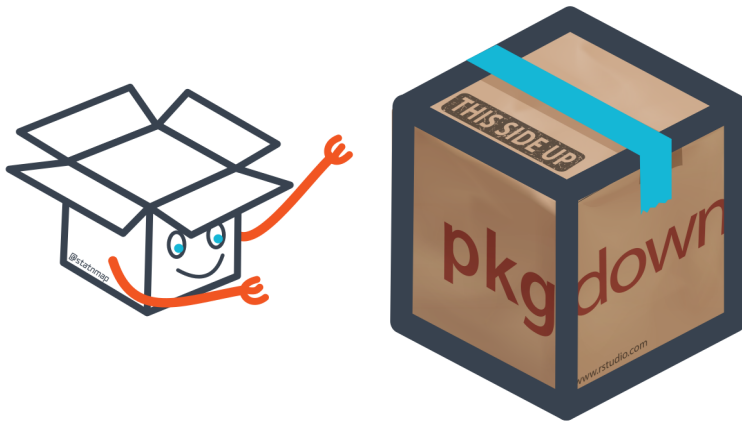
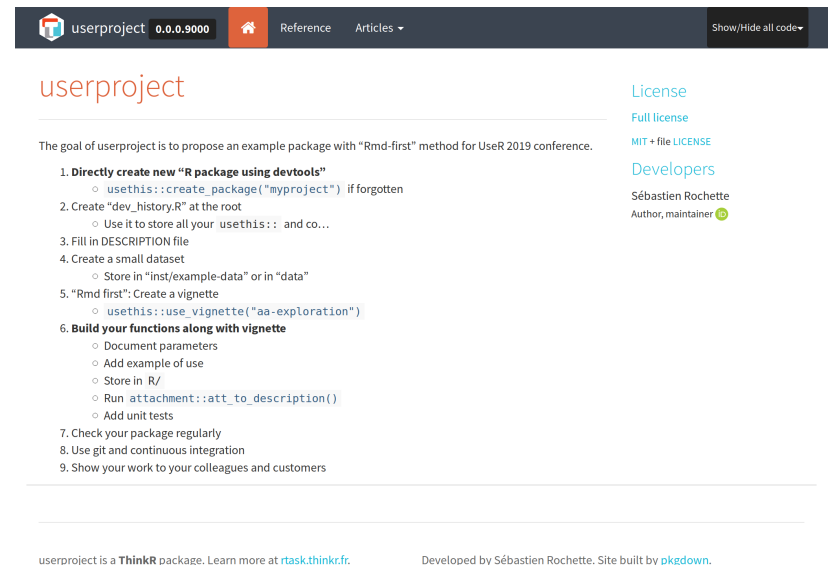
- Create "dev_history.R" to document your development steps
 - see [{golem} for examples](#)
- `attachment::att_to_description()`
 - Document and deal with dependencies



Publish for your colleagues and customers

Share the userguide

- Use `{pkgdown}` to publish
- Customize your Rmd templates
 - [experimental] `chameleon::build_pkgdown()`
 - [experimental] `chameleon::open_pkgdown_function()`

The screenshot shows the GitHub repository page for 'userproject' (version 0.0.0.9000). The page includes a navigation bar with 'Reference' and 'Articles' tabs, and a 'Show/Hide all code' button. The main content area features the repository name 'userproject' and a list of instructions for creating an R package using devtools. The instructions are numbered 1 through 9. On the right side, there are links for 'License', 'Full license', 'MIT + file LICENSE', and 'Developers', along with the author's name 'Sébastien Rochette' and a GitHub profile icon.

'Rmd first' method for every project

Documentation matters

Document for you, document for developers
Document for customers, document for your boss

Start with Rmd

- Start with a Rmd as a sandbox
- Document your functions with reproducible examples
- Create your tests while you code



This presentation on Github: [statnmap/prez](https://github.com/statnmap/prez)

'Rmd first' method for every project

Documentation matters

Document for you, document for developers
Document for customers, document for your boss

Start with Rmd

- Start with a Rmd as a sandbox
- Document your functions with reproducible examples
- Create your tests while you code

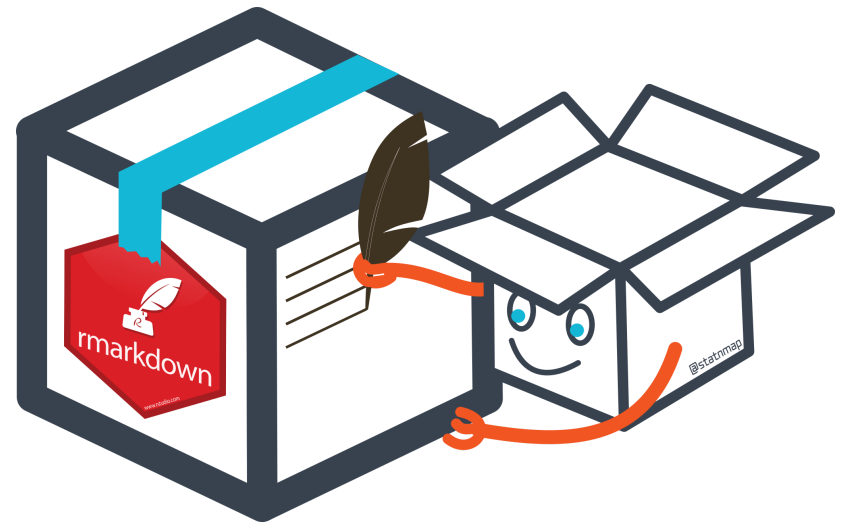
THINK **PACKAGE** !

- Publish your analysis

THANK YOU for your attention

See more:

- rtask.thinkr.fr



This presentation on Github: [statnmap/prez](https://github.com/statnmap/prez)