

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

Use'R Toulouse 2019

Sébastien Rochette



This presentation on Github: 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**

Dompteur de
dragons données,
formatrice logiciel R



**Colin
Fay**

Data scientist
et R hacker



**Sébastien
Rochette**

Modélisateur,
Formateur R, Joueur
de cartographies



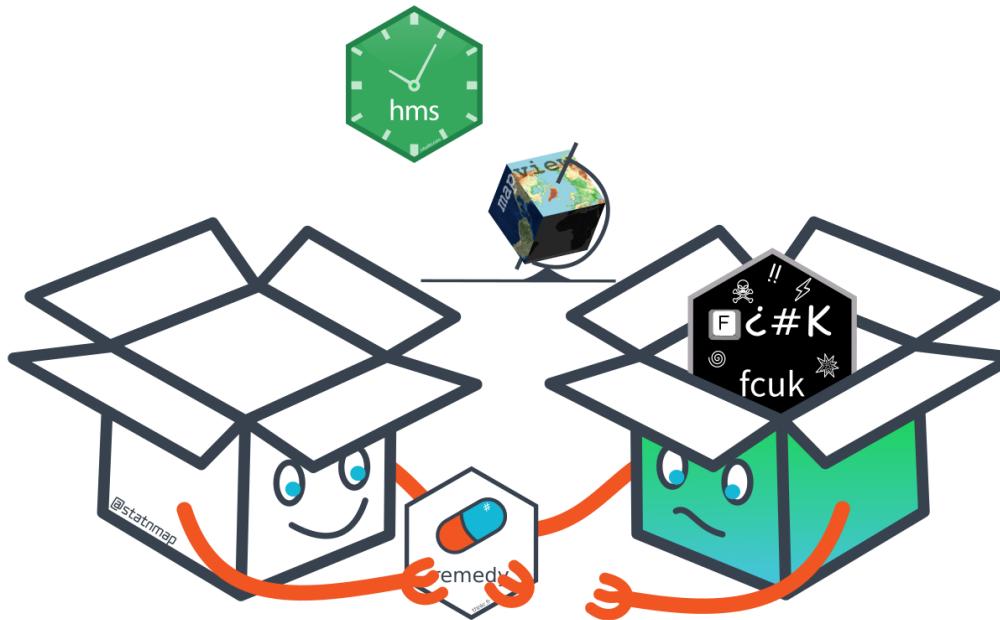
**Cervan
Girard**

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

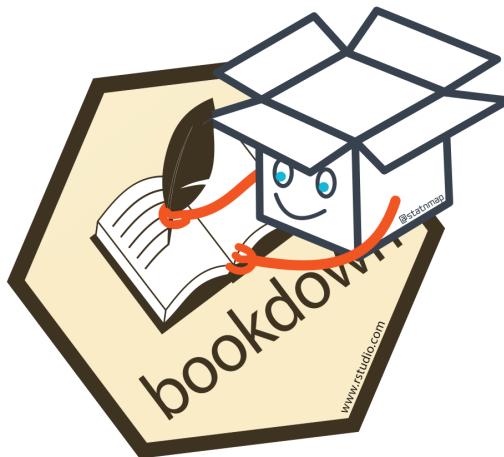


Why is documentation the last thing you do?

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

New Project

Create New Project

Back

Directory name:
userproject

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

Create a git repository

Use packrat with this project

Open in new session



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

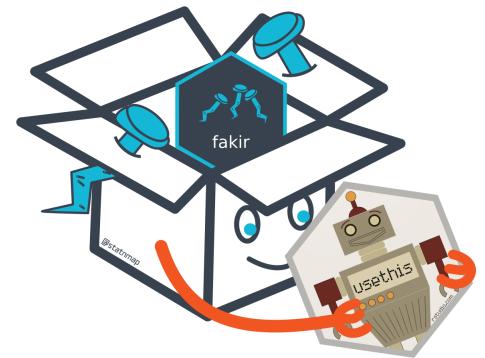
Let's try a workflow for data analysis

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}`)

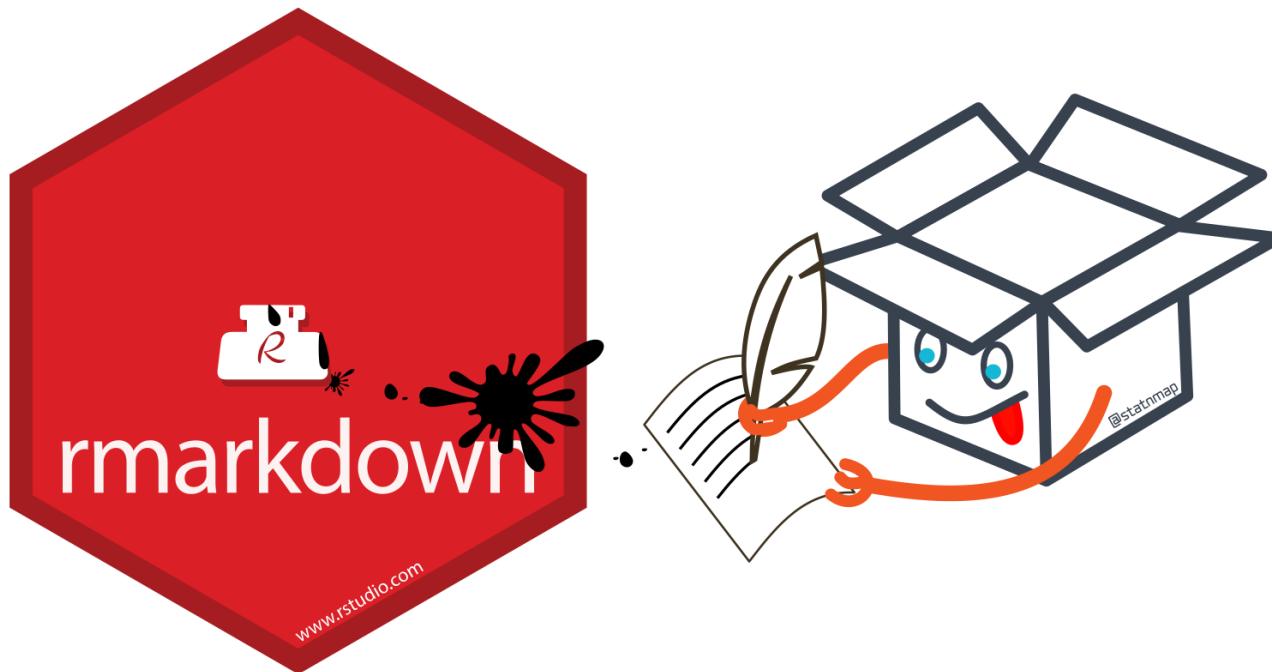
Files				Plots	Packages	Help	Viewer	
				New Folder	Delete	Rename	More	C
↻ Data > github > statnmap > userproject > data-raw  ...								
		Name	Size	Modified				
<input type="checkbox"/>	 ..							
<input type="checkbox"/>	 DATASET.R		601 B	Jun 10				
<input type="checkbox"/>	 clients.csv		73.1 KB	Jun 10				



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
Insert ▾ Run ▾ R Markdown ▾

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 ## Table by department
33 - Create a function to filter on a particular 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 ▾
```



...do what you said



```
DATASET.R * my-analysis.Rmd *
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
```

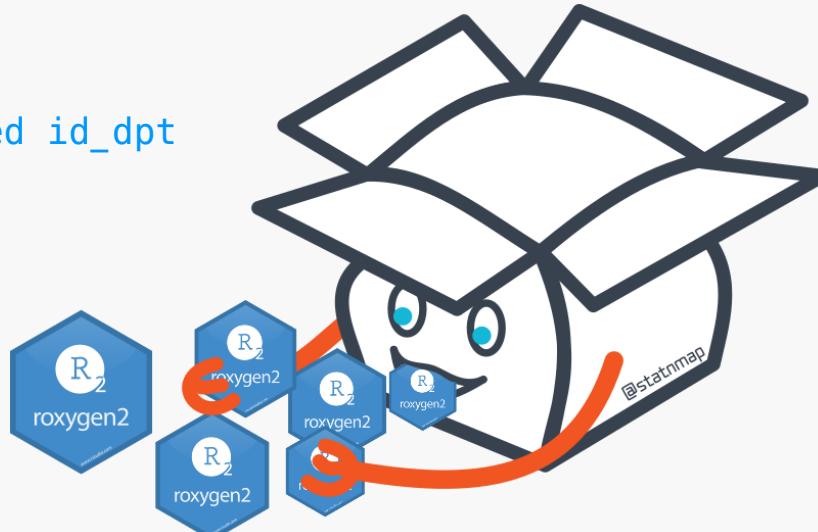


This is not
an hex !

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 - Write a test for function `filter_by_dpt`  
43  
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
```

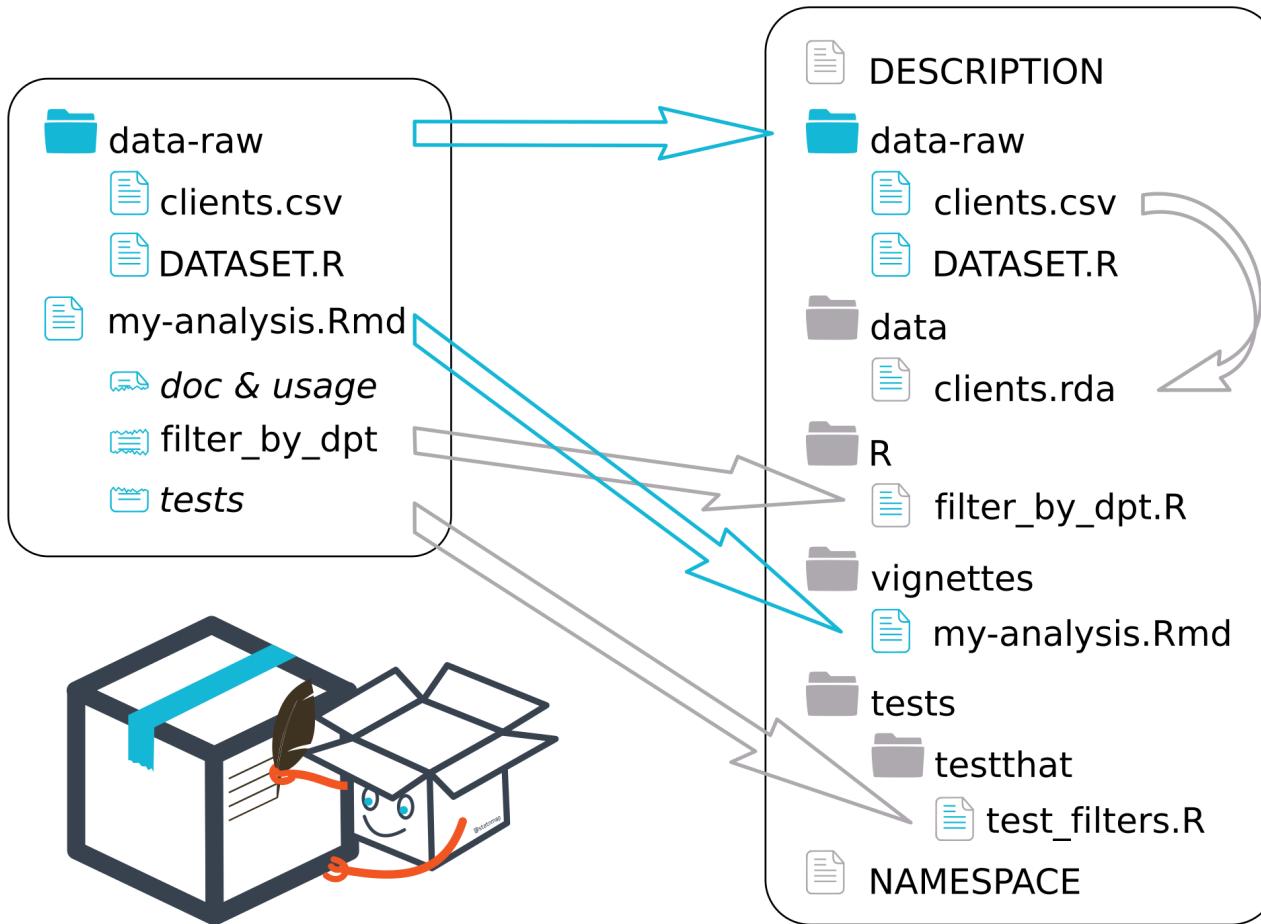


Where are we ?



We are (almost) in a package !

Clean your Rmd to see it



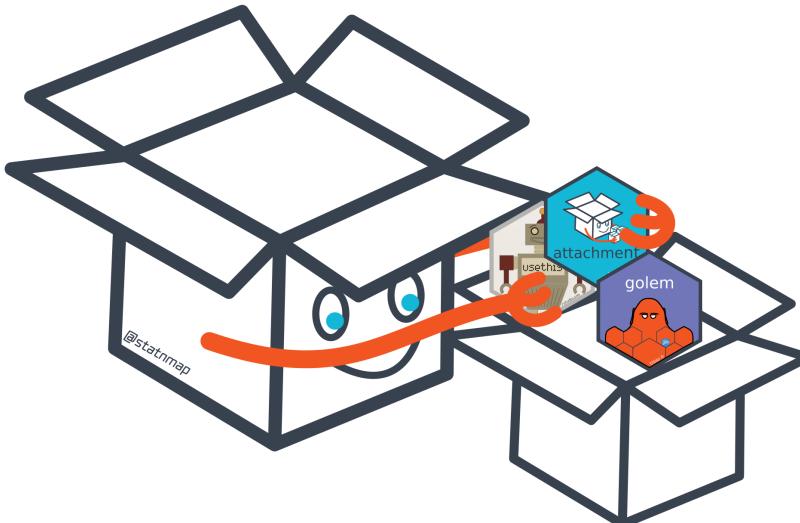
What's next?

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 a user guide for the "userproject" package. At the top, there's a navigation bar with a logo, the package name "userproject 0.0.0.9000", and links for "Home", "Reference", "Articles", and "Show/Hide all code". Below the title "userproject", the goal of the package is described: "The goal of userproject is to propose an example package with 'Rmd-first' method for UseR 2019 conference." A numbered list of steps follows:

- 1. Directly create new "R package using devtools"**
 - `usethis::create_package("myproject")` if forgotten
 - 2. Create "dev_history.R" at the root
 - Use it to store all your usethis:: and co...
 - 3. Fill in DESCRIPTION file
 - 4. Create a small dataset
 - Store in "inst/example-data" or in "data"
 - 5. "Rmd first": Create a vignette
 - `usethis::use_vignette("aa-exploration")`
 - 6. Build your functions along with vignette**
 - Document parameters
 - Add example of use
 - Store in R/
 - Run `attachment::att_to_description()`
 - Add unit tests
 - 7. Check your package regularly
 - 8. Use git and continuous integration
 - 9. Show your work to your colleagues and customers

userproject is a **THINKR** package. Learn more at rtask.thinkr.fr.

Developed by Sébastien Rochette. Site built by [pkgdown](#).

'Rmd first' method for every project

Documentation matters

Document for you, document for developpers

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](#)

'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

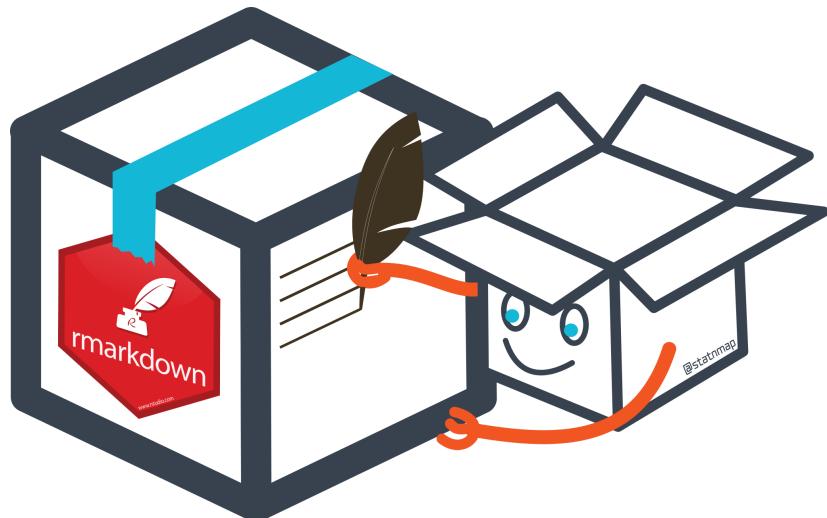
THINK PACKAGE !

- Publish your analysis

THANK YOU for your attention

See more:

- rtask.thinkr.fr



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