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://github.com/ThinkR-open
- https://twitter.com/thinkr_fr

Vincent Guyader
Codeur Fou, formateur et expert logiciel R

Diane Beldame
Dompteur de 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

⚠ 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
## Read a client database

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.

```r
```
```
```

## Table by department

- Create a function to filter on a particular department

```r
```
```r
filter_by_dpt <- function(x, dpt) {
  filter(x, id_dpt == dpt)
}
```

# Examples

```r
```
```
```r
```
```
```r
filter_by_dpt(clients, dpt = 11)
```
```
```r
```
...do what you said

```r
## Read a client database

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.

```r
\`
\`
{r, message=FALSE}
dataset_path <- here("data-raw/clients.csv")
clients <- read_csv(dataset_path)
```

## Table by department

- Create a function to filter on a particular department

```r
\`
\`
{r}
filter_by_dpt <- function(x, dpt) {
  filter(x, id_dpt == dpt)
}
# Examples
filter_by_dpt(clients, dpt = 11)
```

This is not an hex!
Document your function

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

```r
#' 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

```r
# Write a test for function `filter_by_dpt`

```test
```
```
```
```
```
```
```
```
```
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()`
'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
'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