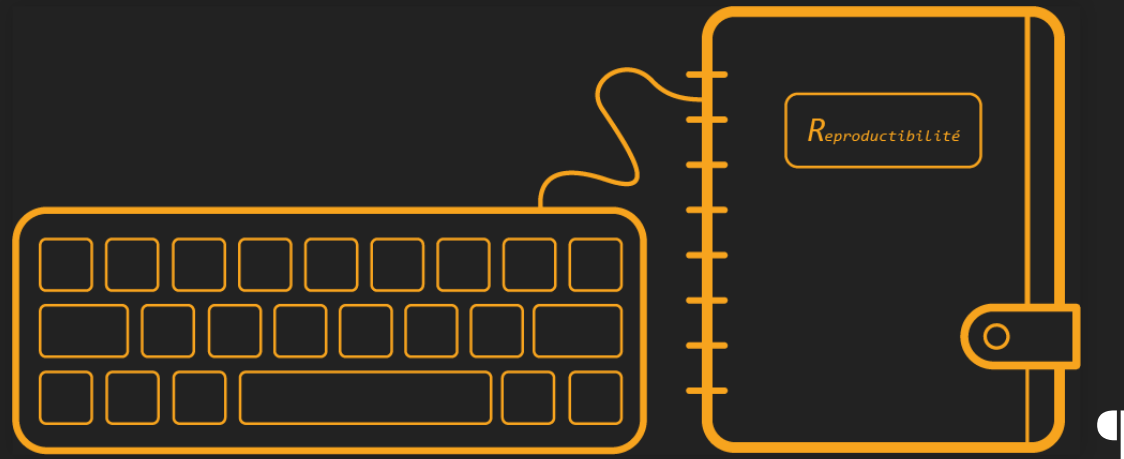


LE NOTEBOOK et la programmation lettrée

Documenter ses traitements



Journées annuelles du réseau Mate-SHS, 17 mai 2022



Hugues Pecout (CIST, CNRS) - Timothée Giraud (RIATE, CNRS)
Sébastien Rey (IDEES, Université Rouen-Normandie)

Bonjour à tou·te·s

Ingénieurs CNRS en sciences de l'information géographique

Timothée



UAR RIATE

Hugues



FR CIST

- Le langage R comme outil principal
- Production régulière de notebook R
- Responsables du projet Rzine

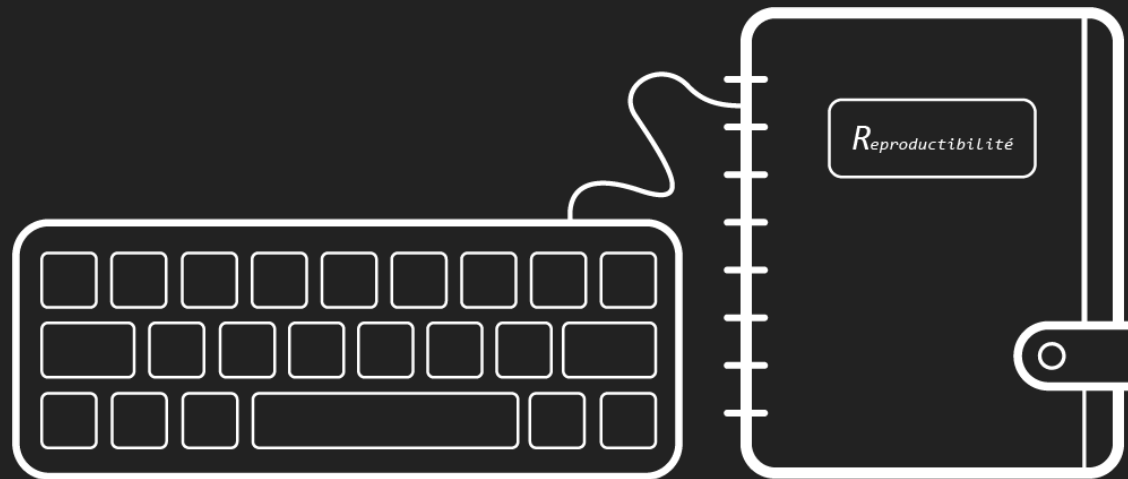
Sébastien



Univ. Rouen-Normandie
UMR IDEES

- Ingénieur de recherche
- Géomatique/Informatique
- Coord. du GT Notebook
- Membre du projet Rzine

Le notebook



Définition

Interface de programmation qui permet de combiner des sections en langage naturel et des sections en langage informatique

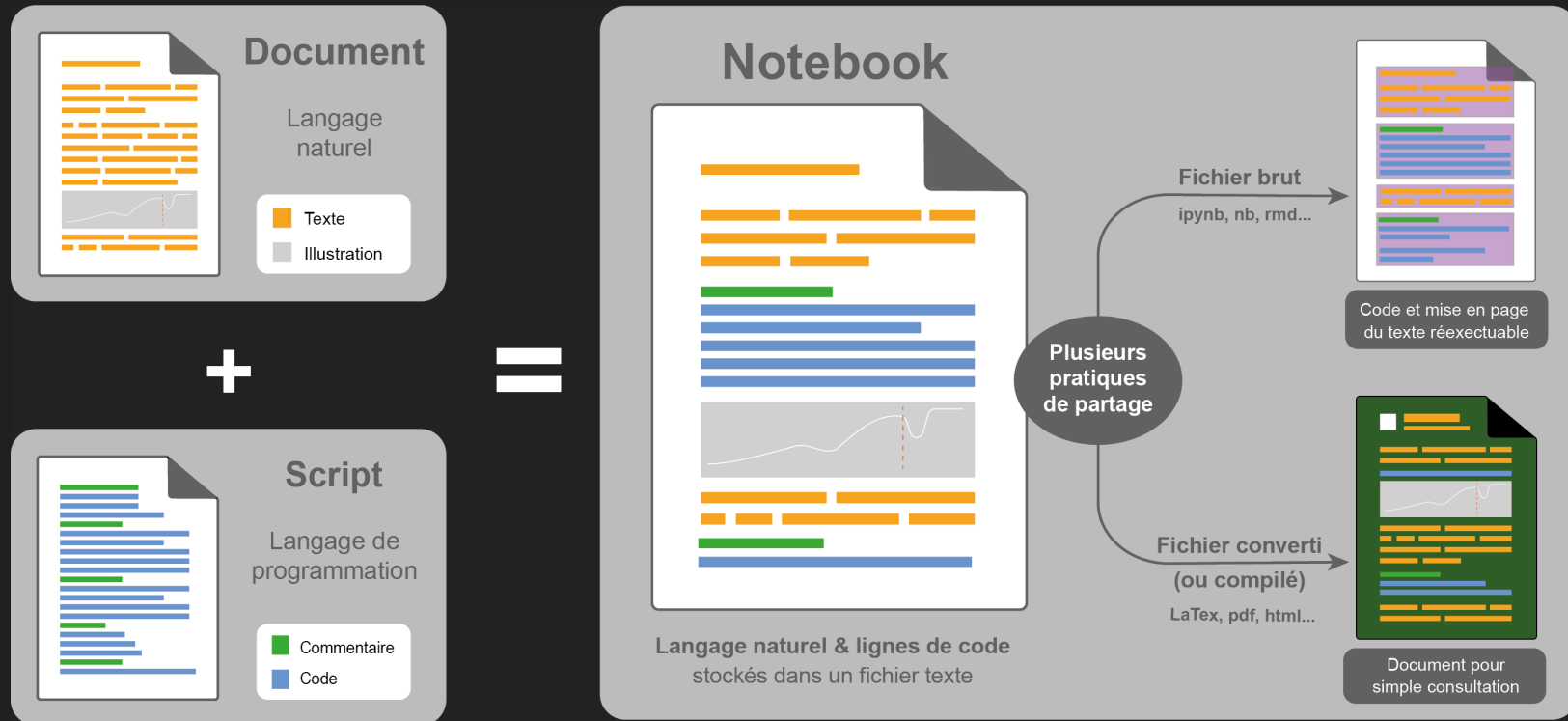
De nombreuses dénominations existent :

- *Article exécutable*
- *Document computationnel*
- *Document électronique interactif*
- *Bloc-note*
- *Cahiers de programmation*
- *Cahier électronique d'analyse*
- *Calepin électronique*
- *Carnet de code*
- *Manifeste algorithmique*
- **Notebook !**
- ...

Un notebook permet d'appliquer le paradigme de programmation lettrée, qui facilite le partage, le travail collaboratif et la reproductibilité



Une logique... des pratiques

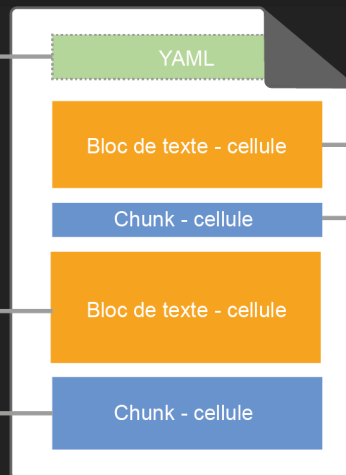


La mise en forme du texte se fait à l'aide d'un langage de balisage (Markdown, html...)

Anatomie d'un notebook

Simple fichier texte

En-tête comportant des métadonnées
Ne concerne pas tous les notebooks



Langage naturel balisé en Markdown

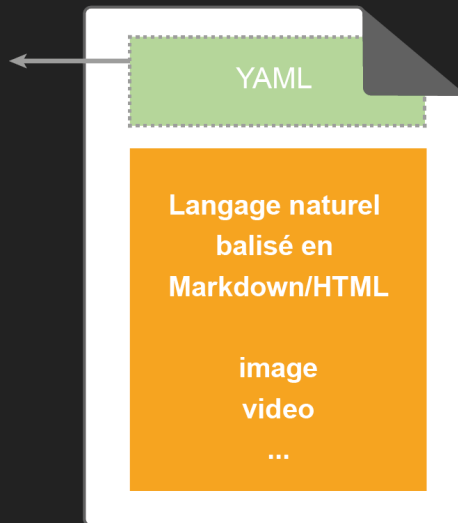
Langage de programmation

Mis en page à l'exécution de la cellule
ou à la compilation du document

Comporte des options de visibilité et
d'exécution du code et des résultats

Notebook sans code ?

Modèle de mise en page
Feuille de style (.css)
Fichier Bibliographique (.bib)
...

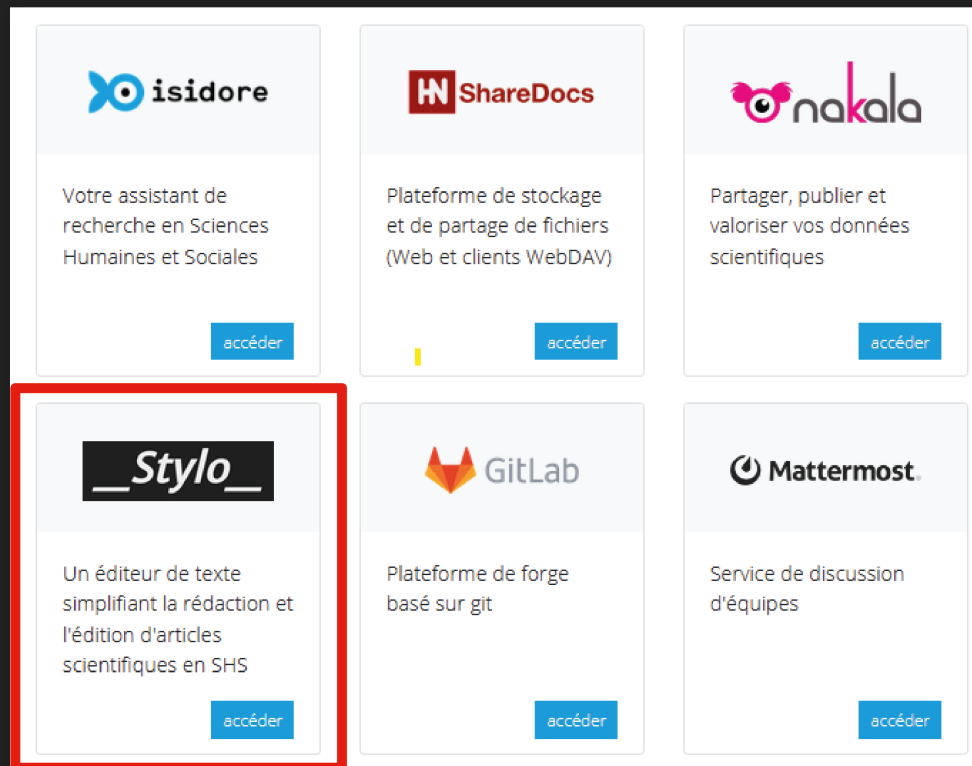


**Livre, Thèse
Lettre, Diaporama
Article scientifique
Support de cours
site web statique
curriculum vitae...**



Stylo d'Human-Num

Développé par le [CRC-EN](#) avec le support [Érudit](#) et de la [TGIR Huma-Num](#)



Sinon, il existe beaucoup d'éditeurs libres ([CodiMD...](#))

Le Markdown (Gruber, 2004)

Fichier Markdown (.md)

```
# Mon premier Markdown



## Journées annuelles 2022 :satisfied:

-> Consultez [**le programme**] (https://ja-mate2022.sciencesconf.org/resource/page/id/1!)

### A. Interventions prévues


1. *Le notebook et la programmation lettrée*
  - Notebook ?
  - Historique
  - Dernière génération
  - Notebook + Git
  - Exemple Quarto

3. *Le projet Rzine*
  - Le *projet Rzine*
  - Le [*site web*] (rzine.fr)
```

-> Visualiser sur [HackMD](#)

Texte mis en forme (sortie html)

Mon premier Markdown



Journées annuelles 2022 😊

-> Consultez [le programme !](#)

A. Interventions prévues

1. *Le notebook et la programmation lettrée*
 - Notebook ?
 - Historique
 - Dernière génération
 - Notebook + Git
 - Exemple Quarto
2. *Le projet Rzine*
 - [Le projet Rzine](#)
 - [Le site web](#)
 - [La collection](#)

B. Intervenants

Nom	Prénom
Giraud	Timothée
Pecout	Hugues
Rey	Sébastien

Le texte brut balisé en Markdown reste lisible pour l'humain, se stocke dans un simple fichier texte et de nombreux outils libres permettent de l'éditer.

Exemple de Notebook

3.7

3.8



Le Notebook Rmarkdown

Fichier source (.rmd)

```
---
title: "Documenter mes traitements"
output: html_document
---



## Introduction

Ce document présente une chaîne de traitement statistique appliquée à des données particulières...


### Les données
Tableau de données utilisé :

```{r chunk1, eval=TRUE, echo=FALSE}
DT::datatable(cars, options = list(pageLength = 5))
```

### Résumé statistique
```

Fichier en sortie (.html)

Documenter mes traitements



Introduction

Ce document présente une **chaîne de traitement statistique** appliquée à des données particulières...

Les données

Tableau de données utilisé :

Show entries Search:

| | speed | dist |
|---|-------|------|
| 1 | 4 | 2 |
| 2 | 4 | 10 |
| 3 | 7 | 4 |
| 4 | 7 | 22 |
| 5 | 8 | 16 |

Showing 1 to 5 of 50 entries Previous **1** 2 3 4 5 ... 10 Next

Résumé statistique

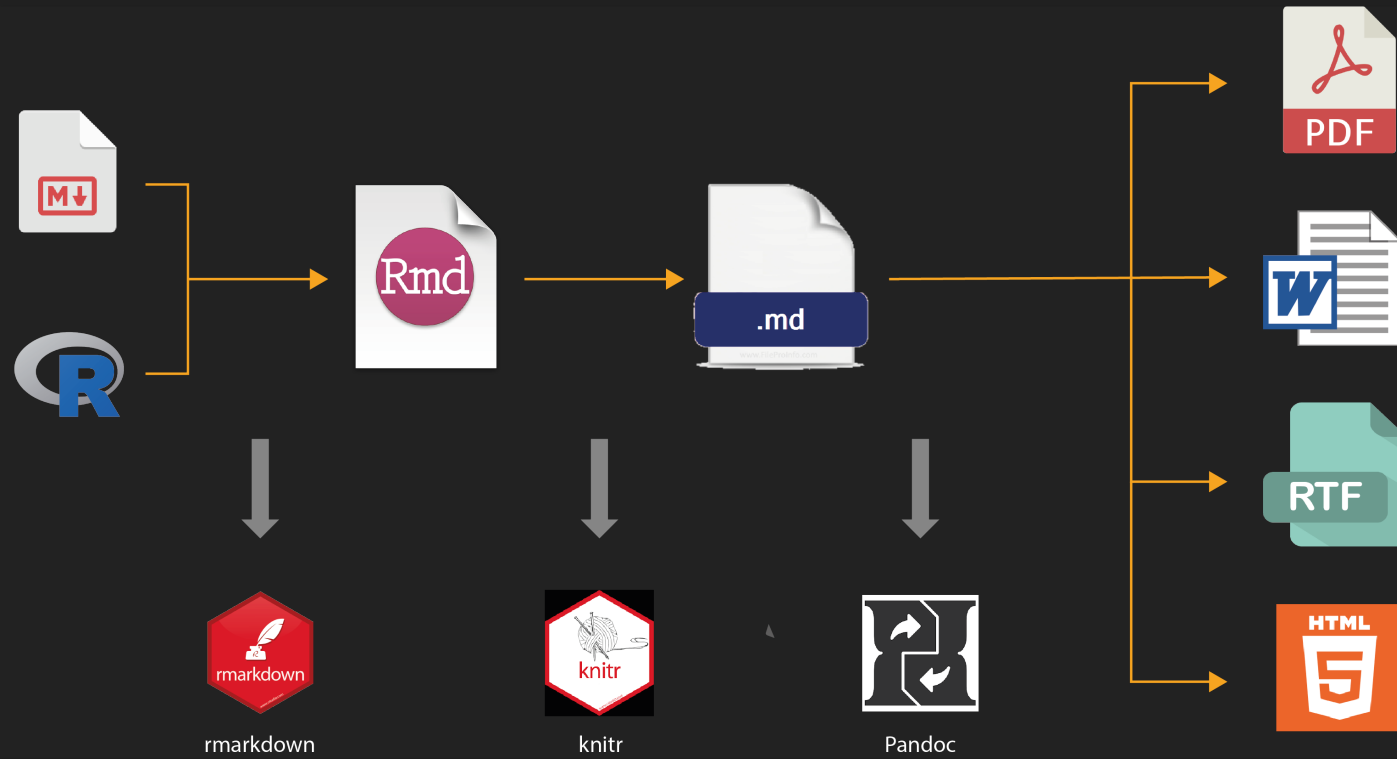
```
summary(cars)

##      speed      dist
## Min.   : 4.0   Min.   : 2.00
## 1st Qu.:12.0   1st Qu.: 26.00
## Median :15.0   Median : 36.00
## Mean   :15.4   Mean   : 42.98
## 3rd Qu.:19.0   3rd Qu.: 56.00
## Max.   :25.0   Max.   :120.00
```

Visualisation

Visualisons les données, mais je ne vous montre pas mon code !

« Kniter » un .rmd




```
1 ---
2 title: "Documenter mes traitements"
3 output: html_document
4 ---
5
6 
8 ## Introduction
9
10 Ce document présente une chaîne de traitement statistique appliquée à des
11 données particulières...
12 ### Les données
13 Tableau de données utilisé :
14
15 ```{r chunk1, eval=TRUE, echo=FALSE}
16 DT::datatable(cars, options = list(pageLength = 5))
17 ```
18
19 ### Résumé statistique
20
21 ```{r chunk2, eval=TRUE, echo=TRUE}
22 summary(cars)
23 ```
24
25 ## Visualisation
26
27 Visualisons les données, mais je ne vous montre pas mon code !
28
29 ```{r plot, echo=FALSE, fig.height=4, fig.width=10}
30 plot(cars,
31      pch = 20,
32      cex = 3,
33      col= "red")
34 ```
35
36 ## Info session
37
38 ```{r , echo=FALSE, fig.height=8, fig.width=10}
39 sessionInfo()
40 ```
41
42 29:51 Chunk 3: plot z R Markdown z
```

Environment History Connections Git

Files Plots Packages Help Viewer

Documenter mes traitements



Introduction

Ce document présente une **chaîne de traitement statistique** appliquée à des données particulières...

Les données

Tableau de données utilisé :

Show 5 entries Search:

| | speed | dist |
|---|-------|------|
| 1 | 4 | 2 |
| 2 | 4 | 10 |
| 3 | 7 | 4 |
| 4 | 7 | 22 |
| 5 | 8 | 16 |


Showing 1 to 5 of 50 entries Previous 1 2 3 4 5 ... 10 Next

Résumé statistique

```
summary(cars)
##      speed      dist
##  Min.   : 4.0   Min.   : 2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Visualisation

Visualisons les données, mais je ne vous montre pas mon code !



```

1 ---
2 title: "Documenter mes traitements"
3 output: html_document
4 ---
5
6 
8 ## Introduction
9
10 Ce document présente une chaîne de traitement statistique appliquée à des
11 données particulières...
12
13 ### Les données
14 Tableau de données utilisé :
15
16 {r chunk1, eval=TRUE, echo=FALSE}
17 DT::datatable(cars, options = list(pageLength = 5))
18
19 ### Résumé statistique
20
21 {r chunk2, eval=TRUE, echo=TRUE}
22 summary(cars)
23
24
25 ## Visualisation
26
27 Visualisons les données, mais je ne vous montre pas mon code !
28
29 {r plot, echo=FALSE, fig.height=4, fig.width=10}
30 plot(cars,
31      pch = 20,
32      cex = 3,
33      col= "red")
34
35
36 ## Info session
37
38 {r , echo=FALSE, fig.height=8, fig.width=10}
39 sessionInfo()
40
41

```

Environment History Connections Git

Files Plots Packages Help Viewer

Publish

Résumé statistique

```
summary(cars)
```

| ## | speed | dist |
|-------------|-------|----------------|
| ## Min. | : 4.0 | Min. : 2.00 |
| ## 1st Qu.: | 12.0 | 1st Qu.: 26.00 |
| ## Median : | 15.0 | Median : 36.00 |
| ## Mean : | 15.4 | Mean : 42.98 |
| ## 3rd Qu.: | 19.0 | 3rd Qu.: 56.00 |
| ## Max. : | 25.0 | Max. : 120.00 |

Visualisation

Visualisons les données, mais je ne vous montre pas mon code !

Info session

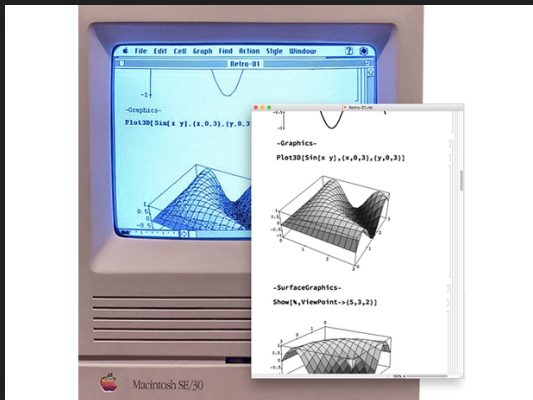
```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 18.04.6 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/openblas/Libblas.so.3
## LAPACK: /usr/lib/x86_64-linux-gnu/libopenblas-r0.2.20.so
##
## locale:
## [1] LC_CTYPE=fr_FR.UTF-8 LC_NUMERIC=C
## [3] LC_TIME=fr_FR.UTF-8 LC_COLLATE=fr_FR.UTF-8
## [5] LC_MONETARY=fr_FR.UTF-8 LC_MESSAGES=fr_FR.UTF-8
## [7] LC_PAPER=fr_FR.UTF-8 LC_NAME=C
## [9] LC_ADDRESS=C LC_TELEPHONE=C
## [11] LC_MEASUREMENT=fr_FR.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
##
## loaded via a namespace (and not attached):
## [1] digest_0.6.29 R6_2.5.1 jsonlite_1.7.2 magrittr_2.0.1
## [5] evaluate_0.14 highr_0.9 rlang_0.4.12 stringi_1.7.6
## [9] jquerylib_0.1.4 DT_0.15 bslib_0.3.1 rmarkdown_2.11.2
## [13] tools_4.0.2 stringr_1.4.0 htmlwidgets_1.5.4 crosstalk_1.2.0
## [17] xfun_0.29 yaml_2.2.1 fastmap_1.1.0 compiler_4.0.2
## [21] htmltools_0.5.2 knitr_1.37 sass_0.4.0
```



Les prémices

Ancêtres du notebook :

- 1970 : Scratchpad (Axiom)
- 1987 : MathCad
- 1988 : Mathematica
- 1989 : Maple
- 2011 : IPython



Émergence de la recherche reproductible

En 1976, Jon Claerbout publie l'ouvrage : *Fundamentals of Geophysical Data Processing*

Cette 1ère version, non-reproductible, sera le début d'un long travail de recherche sur la reproductibilité de ses travaux... Via des notebooks.

Émergence du Literate Programming

Le concept de programmation lettrée pensé et mis en oeuvre (1977-78) par Donald Knuth est lié à l'histoire des notebooks.

Programmation lettrée

La **programmation lettrée** est une approche de la programmation qui se veut différente du paradigme de programmation structurée.

Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.

Donald Knuth, [LP Website citations page](#)

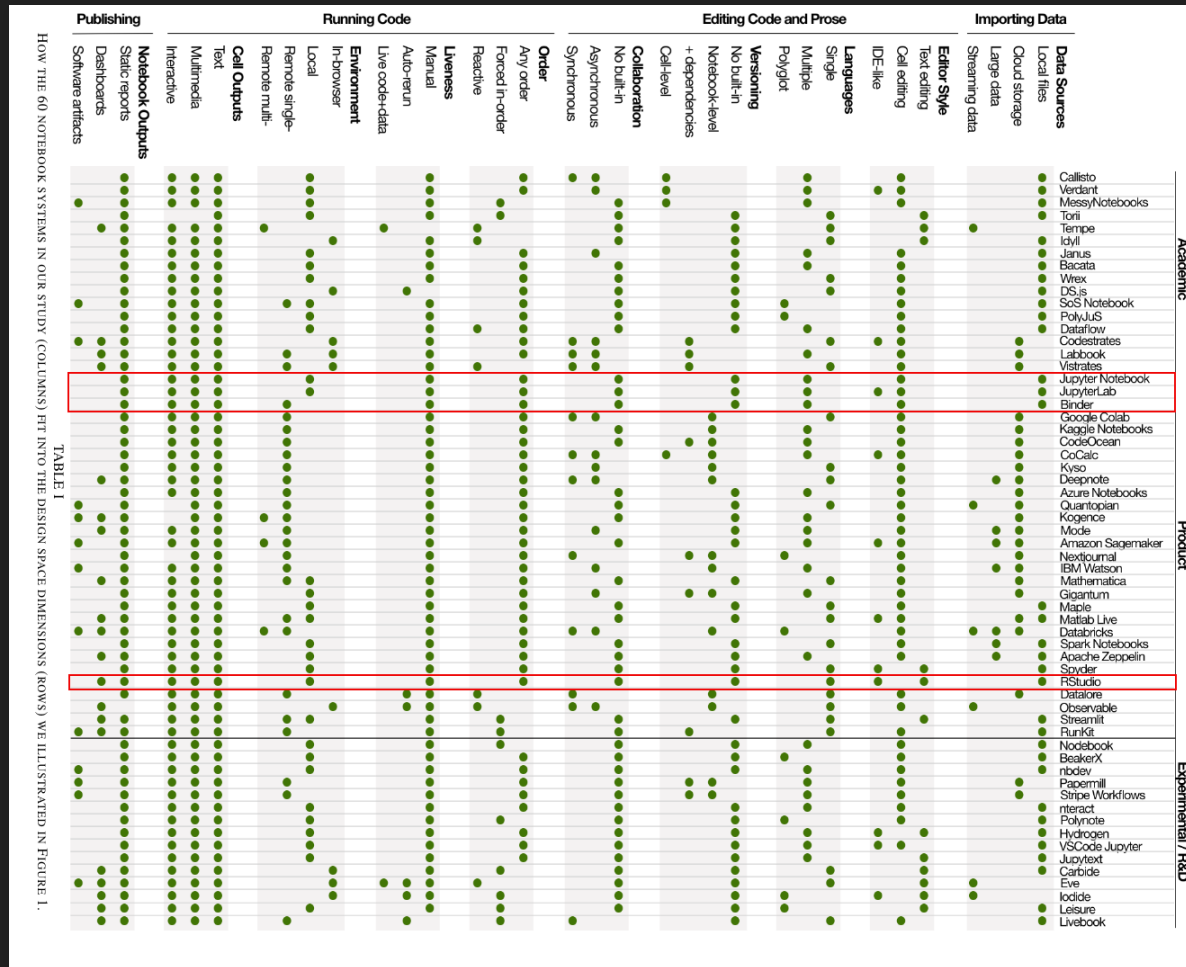
The practitioner of literate programming (...) strives for a program that is comprehensible because its concepts have been introduced in an order that is best for human understanding (...).

Donald Knuth, [LP Website citations page](#)

Le paradigme de la programmation lettrée, tel qu'il a été conçu par **Knuth**, s'éloigne dans son écriture de l'ordonnancement imposé par l'ordinateur, et à la place **autorise les développeurs à utiliser un ordre imposé par la logique et le fil de leur pensée.**



Multiplication des NB et des plateformes




4.3

S. Lau, I. Drosos, J. M. Markel and P. J. Guo, “The Design Space of Computational Notebooks: An Analysis of 60 Systems in Academia and Industry,” 2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2020

4.4


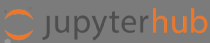


Dernière génération

 **Carnet Jupyter** 2014

«Extension» du projet Ipython
Open Source
Plusieurs langages

Notebook réactif
Code réexécutable


IP[y]:  



 **Rmarkdown & co** 2015

 Studio comme acteur principal
Open Source
Plusieurs langages

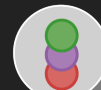
Nombreux formats
Programmation lettrée


        

 **Notebook Observable** 2016

 De Mike Bostock, créateur de  .js
Opensource - Javascript

Notebook réactif **Représentation**
Code réexécutable **graphique interactive**

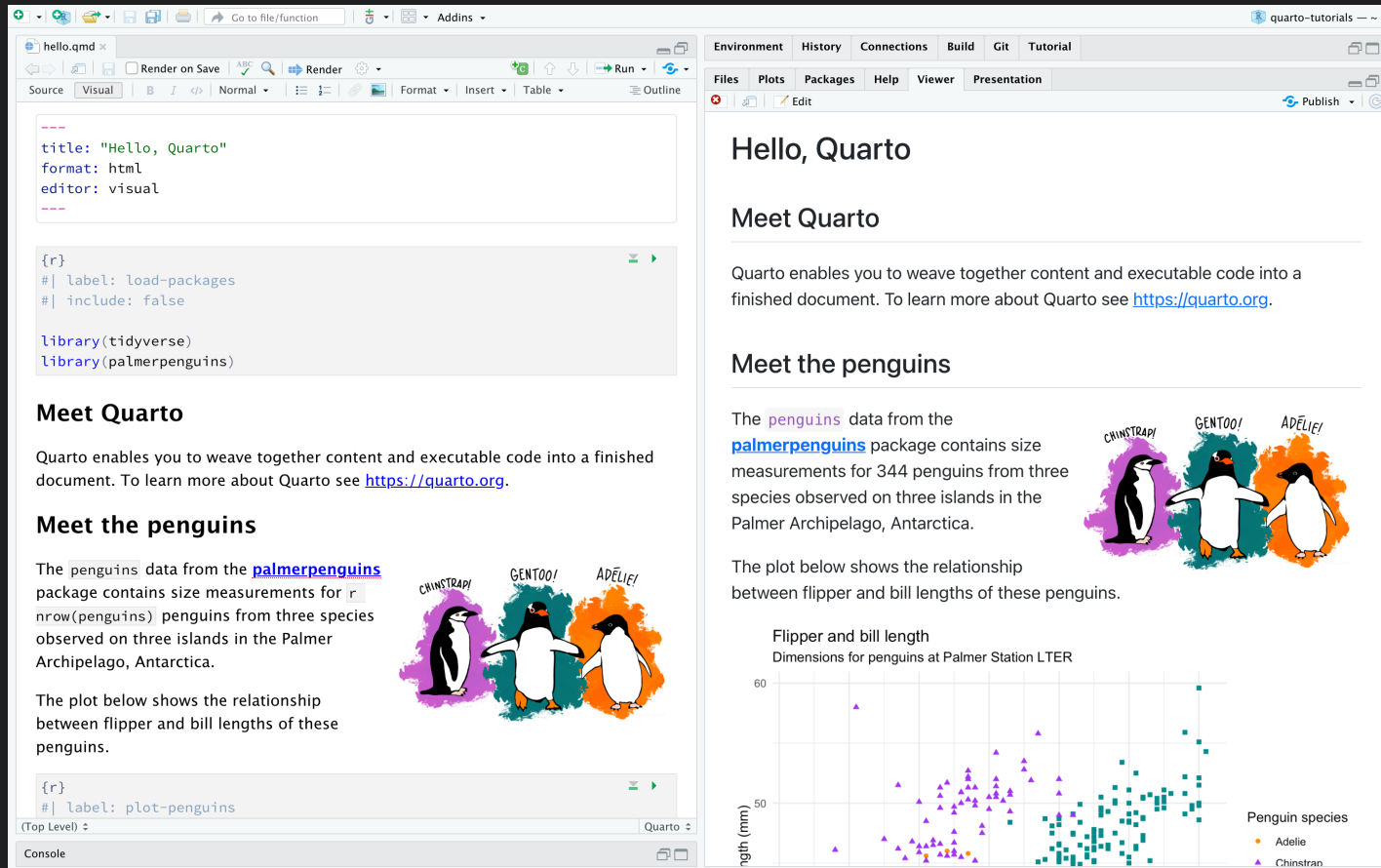
 **Notebook Pluto** 2020

 Open Source
Plusieurs langages

Notebook réactif
Code réexécutable



2022, l'année quarto®



The screenshot displays the Quarto editor interface. On the left, the source code is visible, including a YAML header and R code for loading packages and rendering a plot. The right pane shows the rendered document, which includes a title, a section header, a paragraph of text, a sub-section header, another paragraph, a small illustration of penguins, and a scatter plot titled 'Flipper and bill length'.


```
---  
title: "Hello, Quarto"  
format: html  
editor: visual  
---  
  
{r}  
#| label: load-packages  
#| include: false  
  
library(tidyverse)  
library(palmerpenguins)
```

Meet Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

Meet the penguins

The `penguins` data from the `palmerpenguins` package contains size measurements for `r` `nrow(penguins)` penguins from three species observed on three islands in the Palmer Archipelago, Antarctica.



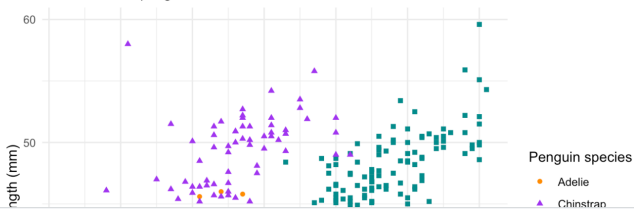
The plot below shows the relationship between flipper and bill lengths of these penguins.

```
{r}  
#| label: plot-penguins
```

(Top Level) | Quarto | Console

Flipper and bill length

Dimensions for penguins at Palmer Station LTER



Bill length (mm)

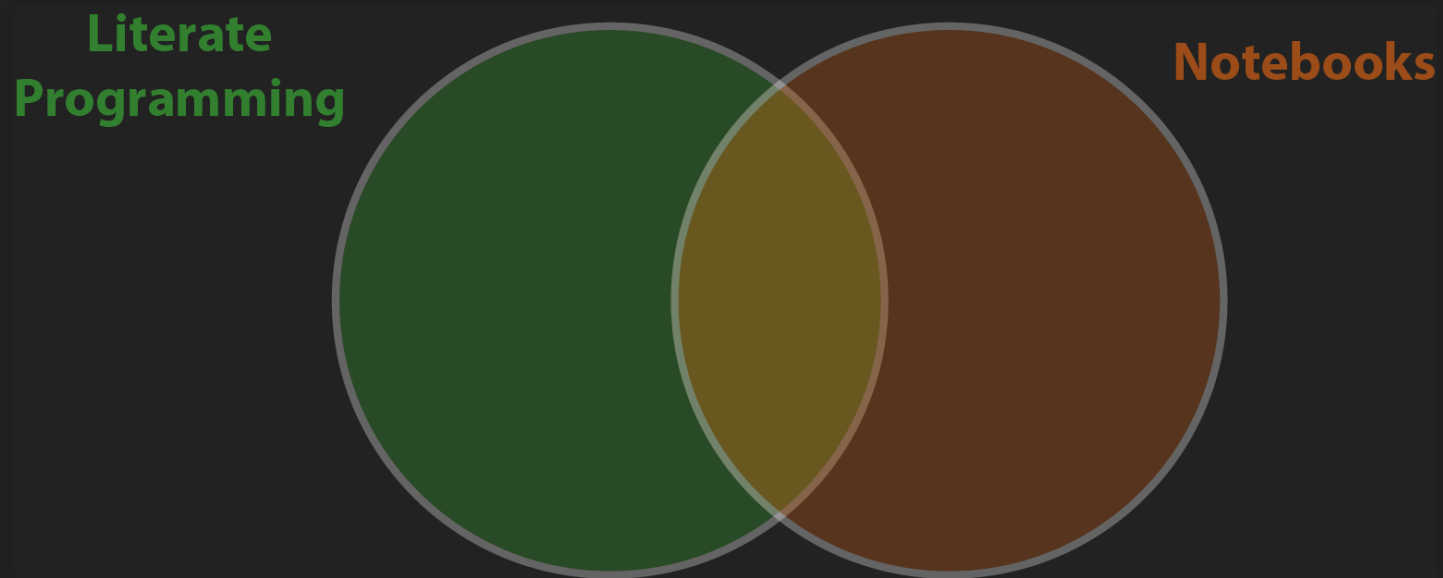
Penguin species

- Adelle
- Chinstran

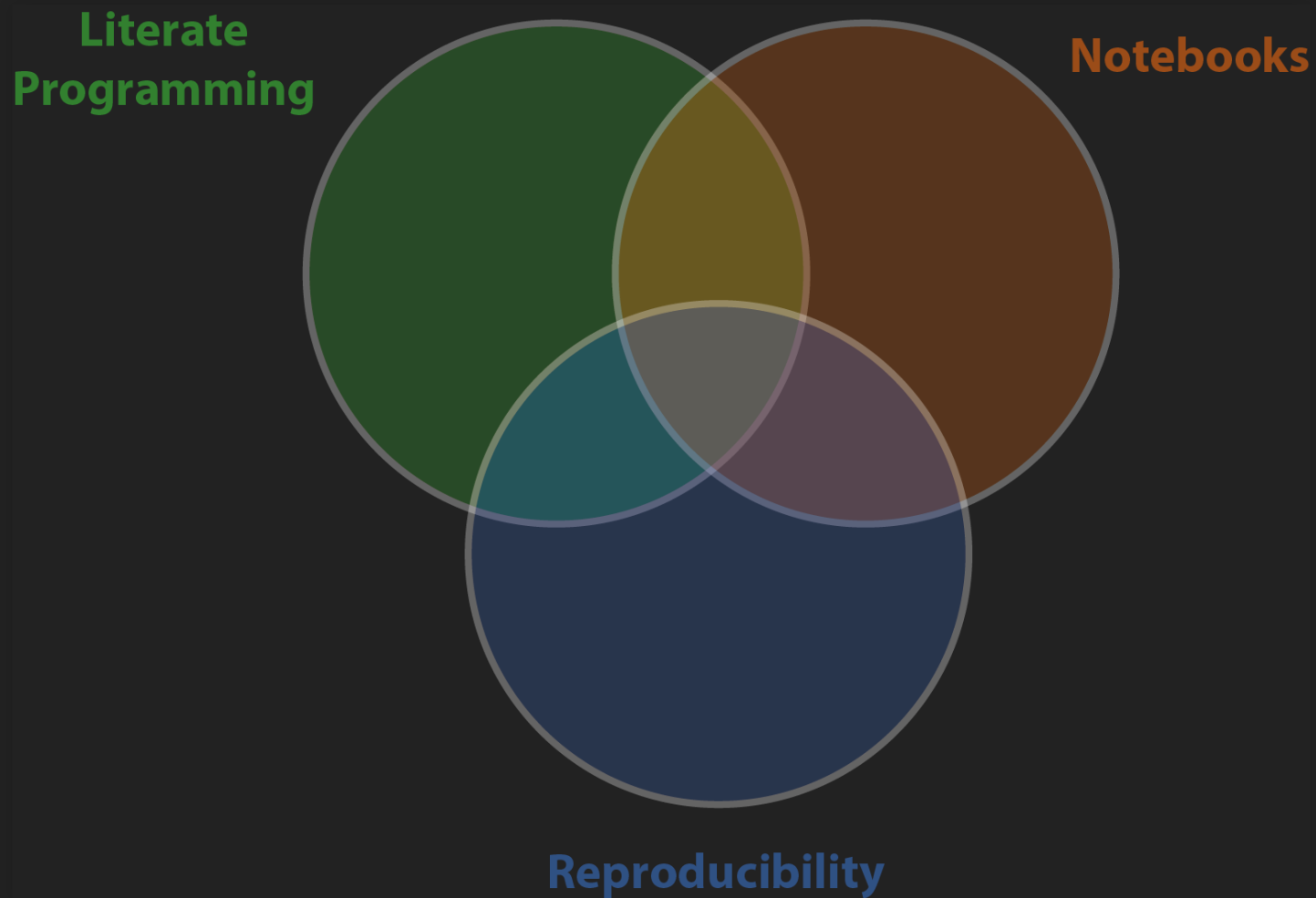
Quarto permet l'utilisation de R, Python, Julia, Javascript (OJS)...

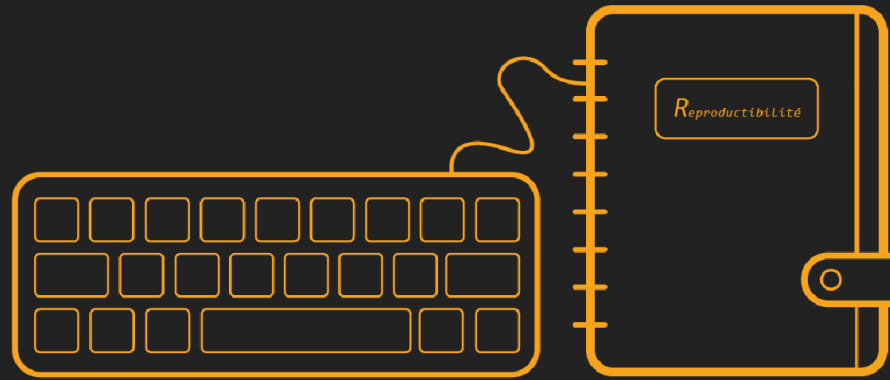


Programmation lettrée & Notebook



LP, Notebook & Reproductibilité





+



Git, GitHub, GitLab

Git est un système de gestion de version distribué pour :

- tracker les changement dans les fichiers texte
- gérer l'historique du code source
- partager le code dans des dépôts distants

GitHub et **GitLab** hébergent des dépôts distants et des services pour gérer des projets (issue tracker, collaboration, hébergement web, CI/CD).



Notebook + Git

- Les sources des notebooks sont des fichiers texte
 - Suivi des modifications (*commit*, branches, etc.)
- Certains ont besoin d'être compilé
 - *Continuous Integration*
- Certains peuvent être déployés sur le web
 - *Continuous Deployment*

Exemple notebook Quarto



[Cliquez-ici !](#)

```
---  
title: "CONCLUSION"  
output: html_document  
---
```

```
## Pour Conclure ...
```

Le *****notebook*****, un format efficace pour la ****recherche reproductible****

1. un format de publication légitime ?

```
```{r}  
cat("Un format de publication émergent !")
```
```

2. un objet pour de nombreux usages

```
```{r}  
print(c("Livre, turorial, manuel,
 cours, communication,
 exercice, application,
 site internet..."))
```
```



Conclusion

Un format de publication
scientifique légitime

- [Rzine](#), pour la pratique de R en SHS
- [Computo](#) de la SFdS
- [Programming Historian](#)
- ...

Nombreux templates,
multiples usages

- *Article*
- *Manuel*
- *Rapports*
- *Livre*
- *Tutoriel*
- *Cours - Exercice*
- *Blog - Site web*
- *curriculum vitae*
- *Diaporama*
- ...

Ceci est un notebook...

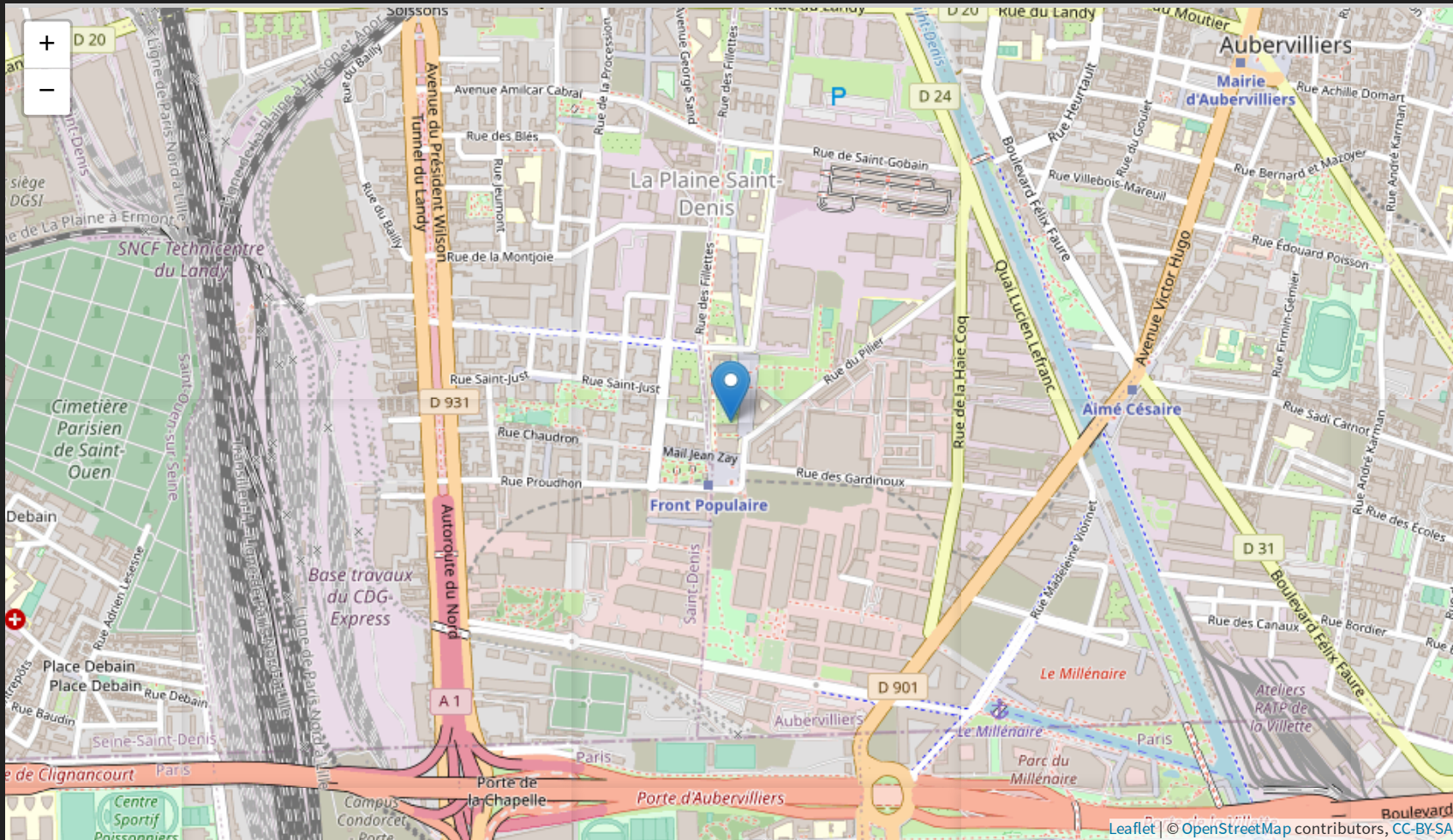
```
# Chargement des librairies
library(sf)
library(leaflet)

# Création d'un point (localisation centre des colloques)
ptsfg <- st_point(c(2.36652, 48.9071886))
ptsfc <- st_sfc(ptsfg, crs = (4326))
ptdf <- data.frame(id= 1, name = "Vous êtes ici !")
Here <- st_sf(ptdf, geometry = ptsfc)

# Affichage du point - carte interactive
mamap <- leaflet(Here, width = 950, height = 550) %>%
  addTiles() %>%
  addMarkers(data = Here, popup = "❤️mate-shs❤️")
```



Vous êtes ici !



Merci de votre attention

Diapositives libres (CC BY-SA 4.0)

Slides : https://huguespecout.github.io/notebook_mateshs/#/

Code source : https://github.com/HuguesPecout/notebook_mateshs

Contact

hugues.pecout@cnrs.fr

timothee.giraud@cnrs.fr

sebastien.rey-coyrehourcq@univ-rouen.fr (GT Notebook)

 [@rzine_shs](https://twitter.com/rzine_shs)

Info session

```
## R version 4.2.2 Patched (2022-11-10 r83330)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 22.04.1 LTS
##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.10.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0
##
## locale:
## [1] LC_CTYPE=fr_FR.UTF-8 LC_NUMERIC=C LC_TIME=fr_FR.UTF-8
## [4] LC_COLLATE=fr_FR.UTF-8 LC_MONETARY=fr_FR.UTF-8 LC_MESSAGES=fr_FR.UTF-8
## [7] LC_PAPER=fr_FR.UTF-8 LC_NAME=C LC_ADDRESS=C
## [10] LC_TELEPHONE=C LC_MEASUREMENT=fr_FR.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
##
## other attached packages:
## [1] leaflet_2.1.1 sf_1.0-7
##
## loaded via a namespace (and not attached):
## [1] revealjs_0.9 tidyselect_1.1.2 xfun_0.32 bslib_0.3.1 purrr_0.3.4
## [6] vctrs_0.4.1 generics_0.1.2 htmltools_0.5.2 yaml_2.3.5 utf8_1.2.2
```



