

Sortie de SAS

5 juin 2024

Alain Quartier-la-Tente

R et Python

- R, Python = langages et logiciels de programmation
- RStudio (Posit), Jupyter, VSCode, Anaconda = environnements de développement (IDE)
- Logiciels open-source se développant grâce à une très grande communauté de contributeurs à partir de *packages* (ensemble de fonctions) : CRAN et PyPI

NB : les logiciels évoluent vite mais il est possible de geler les versions ([renv](#) ou [pipenv](#))

- GitHub, GitLab = plateformes de partage et de développement de code fondées logiciel de gestion de versions Git

R ou Python ?

- Python est réputé pour être plus simple pour le machine learning
- R est réputé pour être plus simple pour l'analyse de données et la visualisation de données
- Beaucoup de ressources disponibles, ex : [série d'ateliers R du CSBQ](#), [Introduction à R et au tidyverse](#), [utilitR](#)
- Beaucoup d'interaction entre les deux langages ([reticulate](#) et [r2py](#)) et les langages (C++, Javascript, HTML, etc.)
Certains packages développés dans les deux logiciels
- Le choix dépend également de la formation des agents => dans les administrations généralement plus de R

Objectif de la présentation : présenter de quelques outils “classiques” (Quarto, Shiny, quelques outils de visualisation)

Quarto

- Système de publication scientifique et technique Open-Source
- Permet, à partir d'une syntaxe simple, de créer :
 - Des rapports (Word, LibreOffice, LaTeX, HTML)
 - Des présentations (PowerPoint, Beamer, Revealjs)
 - Tout un tas d'autres formats (livres, sites web, tableaux de bord, etc.)
 - Tout en intégrant du code (R, Python, Julia, Observable) recompilé à chaque modification

[DREAL Pays de la Loire](#) : Publications entièrement à partir de R

Graphiques et tableaux interactifs (1)

```
1 data <-  
2   system.file("extdata", "df_mariages_etat_dep_an.rds", package = "gouvdown") |>  
3   readRDS()  
4 DT::datatable(data,  
5   extensions = 'Buttons',  
6   options = list(  
7     pageLength = 6,  
8     dom = 'Bfrtip',  
9     buttons = c('copy', 'csv', 'excel', 'pdf', 'print')  
10  ))
```

Graphiques et tableaux interactifs (1)

Copy

CSV

Excel

PDF

Print

Search:

	amar	depmar	etamat_type	n
1	1998	01	Deux célibataires	1771
2	1998	01	Un célibataire et un divorcé	380
3	1998	01	Deux divorcés	168
4	1998	01	Autres cas	45
5	1998	02	Deux célibataires	1937
6	1998	02	Un célibataire et un divorcé	355

Showing 1 to 6 of 8,325 entries

Previous

1

2

3

4

5

...

1,388

Next

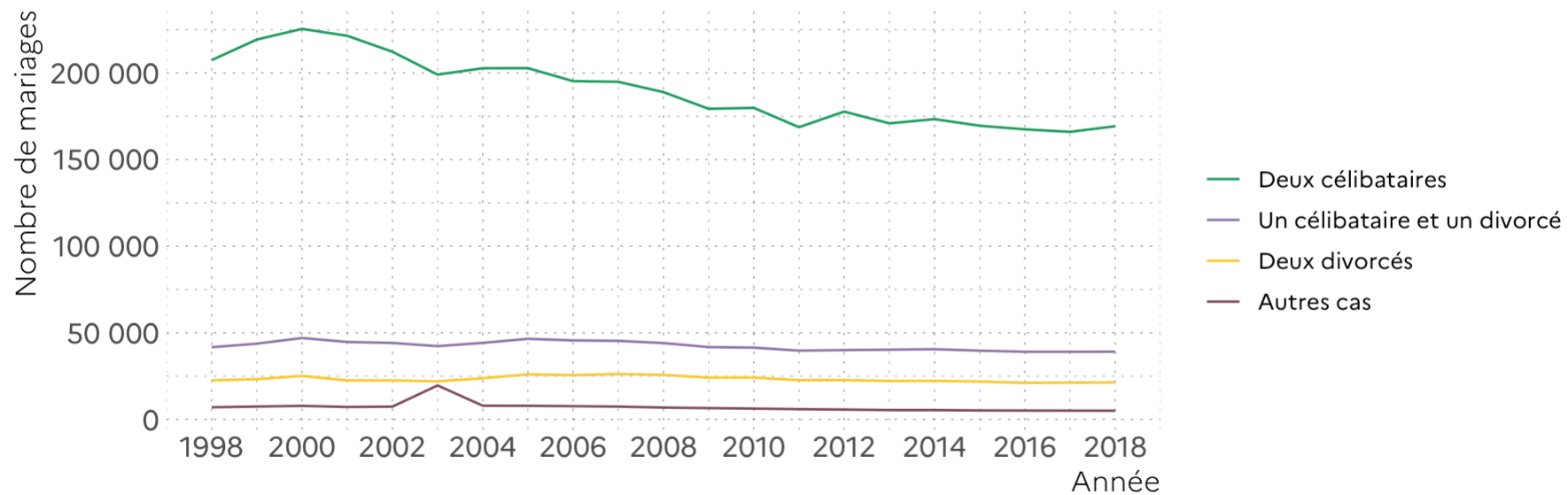
Graphiques et tableaux interactifs (2)

```
1 library(ggplot2)
2 library(gouvdown)
3 library(dplyr)
4 data <- data |>
5   group_by(amar, etamat_type) |>
6   summarise(n = sum(n))
7 p <- ggplot(data, aes(x = amar, y = n, color = etamat_type)) +
8   geom_line() +
9   scale_x_continuous(name = "Année", breaks = seq(1998, 2018, 2)) +
10  scale_y_continuous(
11    name = "Nombre de mariages",
12    breaks = seq(0, 350000, 50000),
13    labels = scales::label_number(big.mark = " ")
14  ) +
15  scale_color_gouv_discrete(name = "", palette = "pal_gouv_qual2") +
16  labs(title = "Nombre de mariages par an",
17       subtitle = "En fonction de l'état matrimonial des mariés",
18       caption = "Source : Insee, fichiers d'état civil") +
19  theme_gouv()
20 p
```

Graphiques et tableaux interactifs (2)

Nombre de mariages par an

En fonction de l'état matrimonial des mariés

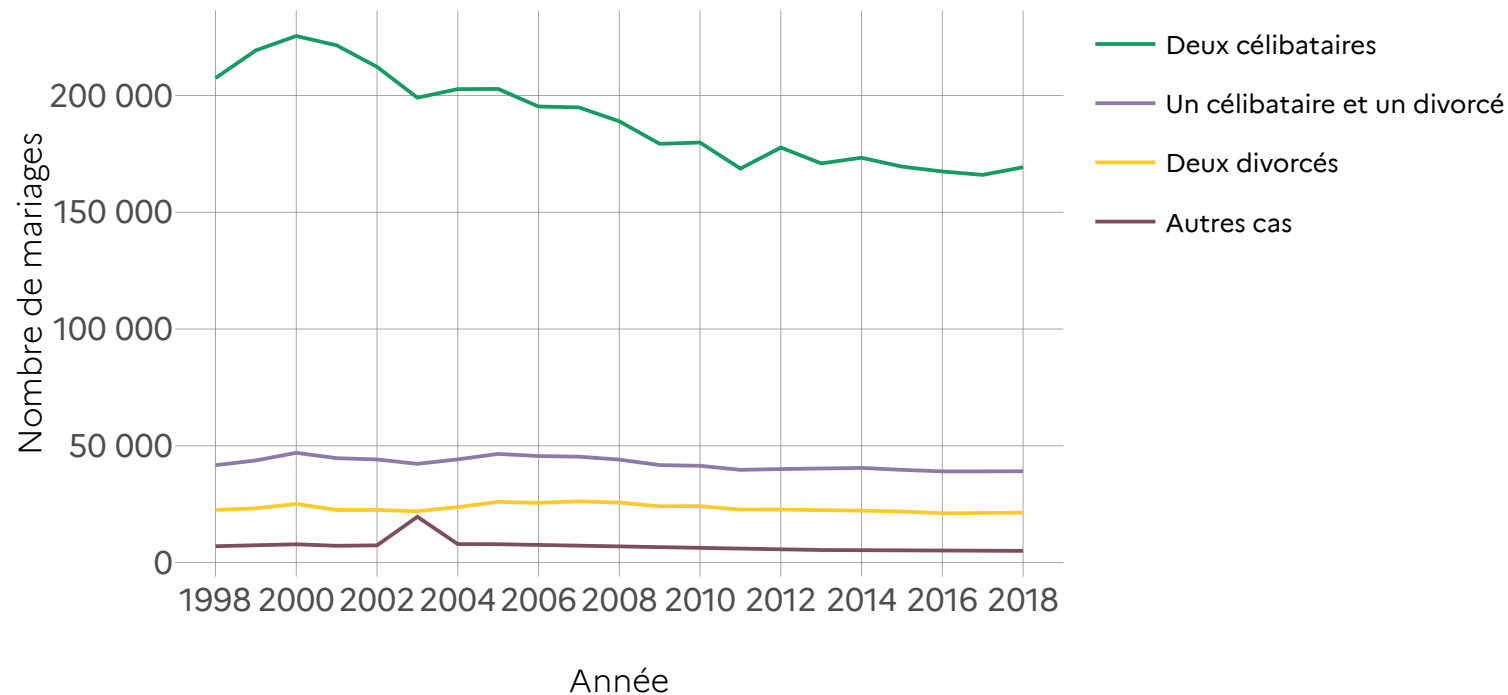


Source : Insee, fichiers d'état civil

Graphiques et tableaux interactifs (3)

```
1 plotly::ggplotly(p, height = 500, width = 800)
```

Nombre de mariages par an



Shiny

- `shiny` est un package R (maintenant également Python) qui permet de créer des applications web interactives directement depuis R sans avoir besoin de connaître le HTML ou le JavaScript
- les applications sont simples à créer mais peuvent être compliquées à déployer

Projet Sortie de SAS à l'Insee

- Objectif : abandonner SAS à horizon 2025, conversion vers un langage open source (R, Python...)
- De nombreuses ressources :
 - Formations (agents et managers)
 - Appui de la DSI
 - Communication (webinaires, ressources en ligne)
 - Support (e.g. : Tchap)
 - Parfois appel à des prestations ou création de postes
- Moins à l'Insee : Licences ChatGPT pro, GitHub Copilot... (mais nécessite de bien maîtriser R)

Recommandations :

- Pour le partage des fichiers : format parquet
- Pour le code : utilisation de logiciels de gestion de version (Git)