

# Sciences humaines, sciences sociales à l'ère numérique

L'analyse des données et l'urbanisation d'outils

Stéphane Pouyllau, ingénieur de recherche CNRS - Professeur Attaché Univ. Evry - Paris Saclay

# Méthodes, techniques, outils...

Notions abordées dans les cours à venir...

La mise en données (méthodes de modélisation des connaissances)

L'analyse des données et l'urbanisation d'outils

La pérennité de l'information

La publication Web à l'heure de l'IA générative et à la nécessité du *Low-Tech*

...

**Main features**

- Faceting**: Drill through large datasets using facets and apply operations on filtered views of your dataset.
- Clustering**: Fix inconsistencies by merging similar values thanks to powerful heuristics.
- Reconciliation**: Match your dataset to external databases via reconciliation services.
- Infinite undo/redo**
- Privacy**
- Wikibase**

**Functions Defined**

```
# Python 3: Fibonacci series up to n
>>> def fib(n):
    a, b = 0, 1
    while a < n:
        print(a, end=' ')
        a, b = b, a+b
    print()
>>> fib(1000)
0 1 1 2 3 5 8 13 21 34 55 89 144 233 377 610
987
```

Python is a programming language that lets you work quickly and integrate systems more effectively. [» Learn More](#)

**Download and Install R**

Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of the versions of R:

- Download R for Linux (Debian, Fedora/Redhat, Ubuntu)
- Download R for macOS
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

**Source Code for all Platforms**

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. They have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2024-10-31, Pile of Leaves) [R-4.4.2.tar.gz](#), read [what's new](#) in the latest version.
- The CRAN directory [src/base-prelease](#) contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
- Between releases, the same directory [src/base-prelease](#) contains snapshots of current patched and development versions. Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Alternatively, daily snapshots are [available here](#).
- Code of older versions of R is [available here](#).
- Contributed extension [packages](#).

**Questions About R**

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

**Supporting CRAN**

- CRAN operations, most importantly hosting, checking, distributing, and archiving of R add-on packages for various platforms.

**LHS Concept Label sample csv** [Permalink](#)

**14550 rows**

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

ConceptLabel	nCount	example
Fictitious character	3328	Commander Toad (Fictitious character)
England	1515	South Crofty Mine (England)
Wash.	1298	Whitemarsh Landfill (Wash.)
Calif.	1223	Clear Creek Distribution System (Calif.)
Italy	1151	Franciacorta (Italy)
New York, N.Y.	1016	Daily News Building (New York, N.Y.)
African people	976	Korane (African people)
Spain	887	Cartada Real Leonesa (Spain)
Alaska	723	Annette Islands Reserve (Alaska)
Extinct city	719	Sumuruan (Extinct city)
France	709	Forterre (France)
Ariz.	691	Fools Hollow Lake Watershed (Ariz.)
India	664	Chitrakot Falls (India)
Germany	652	Eastphalia (Germany)
Russia	649	Zaelfsovskii Forest (Russia)
Japan	649	Saku Region (Japan)
Or.	640	Brothers Fault (Or.)
China	628	Kuajishan Liyou Dujiaqu (China)
English	626	Children's stories Belzean (English)
Game	621	Ti-khi (Game)
Mexico	571	Reserva Ecológica Cuxtal (Mexico)
Scotland	570	Harris (Scotland)
Colo.	522	Fireside Mine (Colo.)
Idaho	507	Orchard Training Area (Idaho)
Islamic law	498	Game laws (Islamic law)
Tex.	492	Camp Waldemar (Tex.)
Computer program language	466	TABLOG (Computer program language)
Utah	455	Koyote Relief Mine (Utah)
Pa.	436	Millbrook Marsh (Pa.)
B.C.	429	Featherland (B.C.)
N.Z.	428	Puketi Rhododendron Trust Garden (N.Z.)

localhost

**Launcher**

- Notebook**:
  - Python 3 (ipykernel)
- Console**:
  - Python 3 (ipykernel)
- Other**:
  - Terminal
  - Text File
  - Markdown File
  - Python File
  - Show Contextual Help

**RStudio**

**R 4.4.2 (2024-10-31) -- "Pile of Leaves"**

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help,  
'?help.start()' for an HTML browser interface to help  
information, or 'q()' to quit R.

[Workspace loaded from ~/RData]

**RStudio**

**R 2024.12.0 Build 467**

Version 1.6.0 of the shiny package is required to run tutorials for RStudio.  
Please [click here](#) to update the shiny package.  
The learnr package is required to run tutorials for RStudio.

**RStudio**

"Kousa Dogwood" Release (cf37a3e5, 2024-12-11) for macOS  
Mozilla/5.0 (Macintosh; Intel Mac OS X 10\_15\_7) AppleWebKit/537.36 (KHTML, like Gecko)  
RStudio/2024.12.0+467 Chrome/126.0.6478.234 Electron/31.7.6 Safari/537.36, Quarto 1.5.57

RStudio, Quarto, and Shiny are registered trademarks of Posit Software, PBC, all rights reserved.  
See [https://www.posit.co/about/trademark/](#) for permitted uses and guidelines.

Unless you have received this program directly from Posit Software pursuant to the terms of a commercial license agreement with Posit Software, then this program is licensed to you under the terms of version 3 of the GNU Affero General Public License.

**Open Source Components**

**Copy Version** **OK**

Library  
Misc  
Movies  
Music  
Pictures  
Public  
src  
Untitled.ipynb  
X-Plane 12

# Langages et IDE



# Survivre avec Python

# Python

Python, Patrick Fuchs et Pierre Poulain, U. Paris Cité.



The screenshot shows a web browser window with the URL [python.sdv.u-paris.fr](http://python.sdv.u-paris.fr). The page has a blue header with the text 'Cours de Python'. On the left, there's a sidebar with a navigation menu:

- Cours de Python
- Avant propos
- Chapitres
  - 1. Introduction
  - 2. Variables
  - 3. Affichage
  - 4. Listes
  - 5. Boucles et comparaisons
  - 6. Tests
  - 7. Fichiers
  - 8. Dictionnaires et tuples
  - 9. Modules
  - 10. Fonctions
  - 11. Plus sur les chaînes de caractères
  - 12. Plus sur les listes
  - 13. Plus sur les fonctions
  - 14. Conteneurs
  - 15. Création de modules
  - 16. Bonnes pratiques en programmation Python
  - 17. Expressions régulières

The main content area features the Python logo (a blue and yellow snake) and the word 'python' in a large serif font. It includes a search bar labeled 'Rechercher' and a link to the PDF version of the course. A small image of the book cover is shown at the bottom left of the content area.

Version du 13/11/2024 | [PDF du cours](#)

La seconde édition de notre livre « [Programmation en Python pour les sciences de la vie](#) » est parue aux éditions Dunod en juin 2024.

Vous le trouverez dans toutes les bonnes librairies.

Patrick Fuchs & Pierre Poulain. Cours sous licence Creative Commons Attribution - Partage à l'identique (CC BY-SA)

# Python

Site ressource pour aller plus loin : [pyshs.fr](https://pyshs.fr)

A screenshot of a GitHub repository page for 'pyshs/exemples-manuel'. The repository has 11 issues and 11 pull requests. The main view shows a list of files, mostly Jupyter notebooks (ipynb) named from '0 - Prologue.ipynb' to '8 - Statistiques avancées.ipynb'. Each file entry includes a 'modified' icon, the file name, a description like 'nettoyage code', and a timestamp (e.g., 5 years ago). A commit history for the master branch is visible at the top, showing one commit by 'emilienschultz' with the message 'modif fichier' and timestamp '6bba834 · 4 years ago'.

The screenshot shows the PySHS website (<https://pyshs.fr>) and a book cover for 'Python pour les SHS' by Emilien Schultz and Matthias Bussonnier.

**PySHS Website:** The website features a header with the PySHS logo (an orange icon with two brackets) and the text 'PySHS - Python pour les Sciences Humaines et Sociales'. Below the header, there's a large title 'Python pour les Sciences Humaines et Sociales' and a brief introduction: 'Introduction à la programmation pour le traitement de données. Ce livre propose de familiariser des débutants à la programmation Python et à plus long terme favoriser le développement d'une communauté francophone des utilisateurs de Python en SHS'. Three buttons are present: 'En Savoir Plus' (orange), 'Scripts' (white), and 'Acheter' (orange).

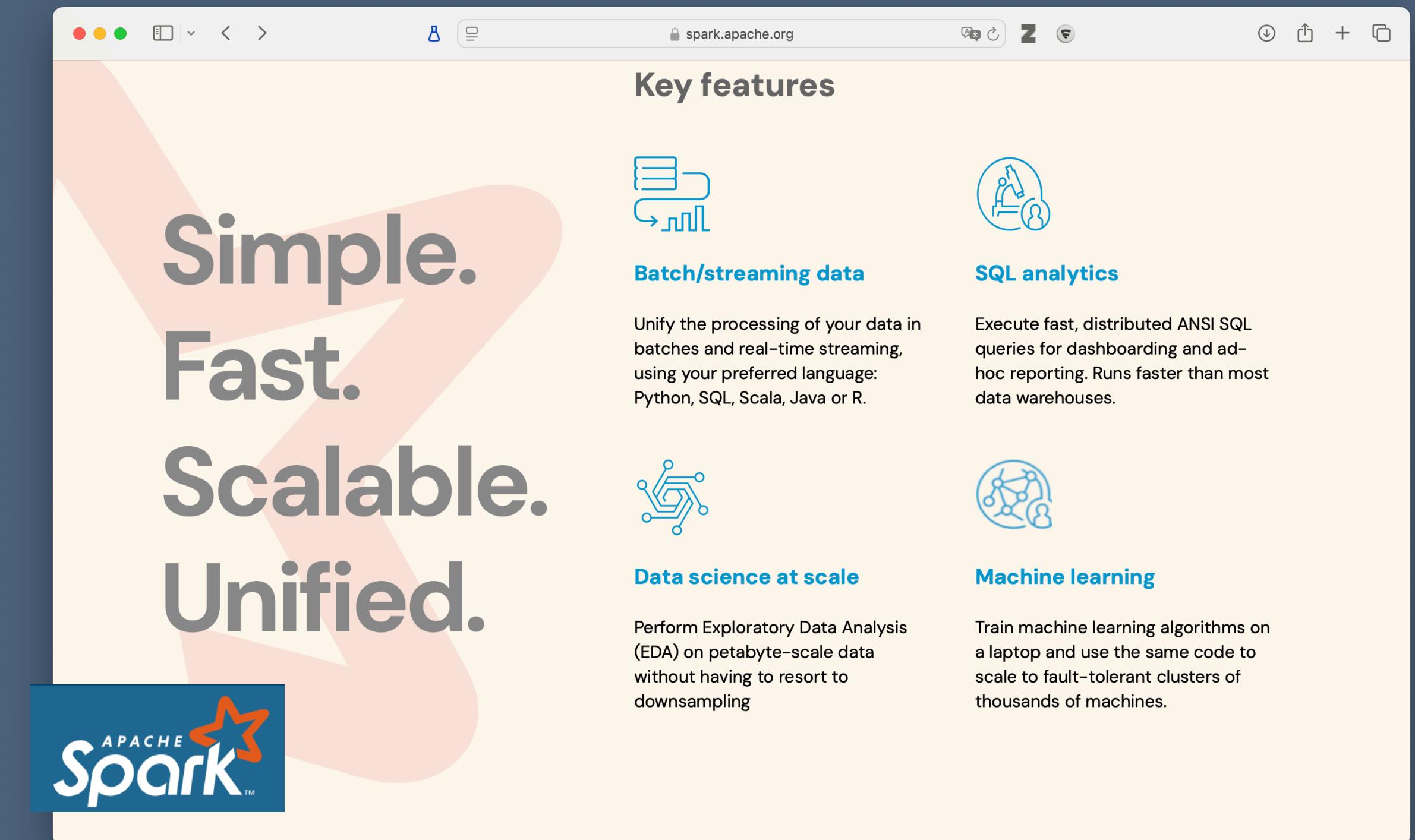
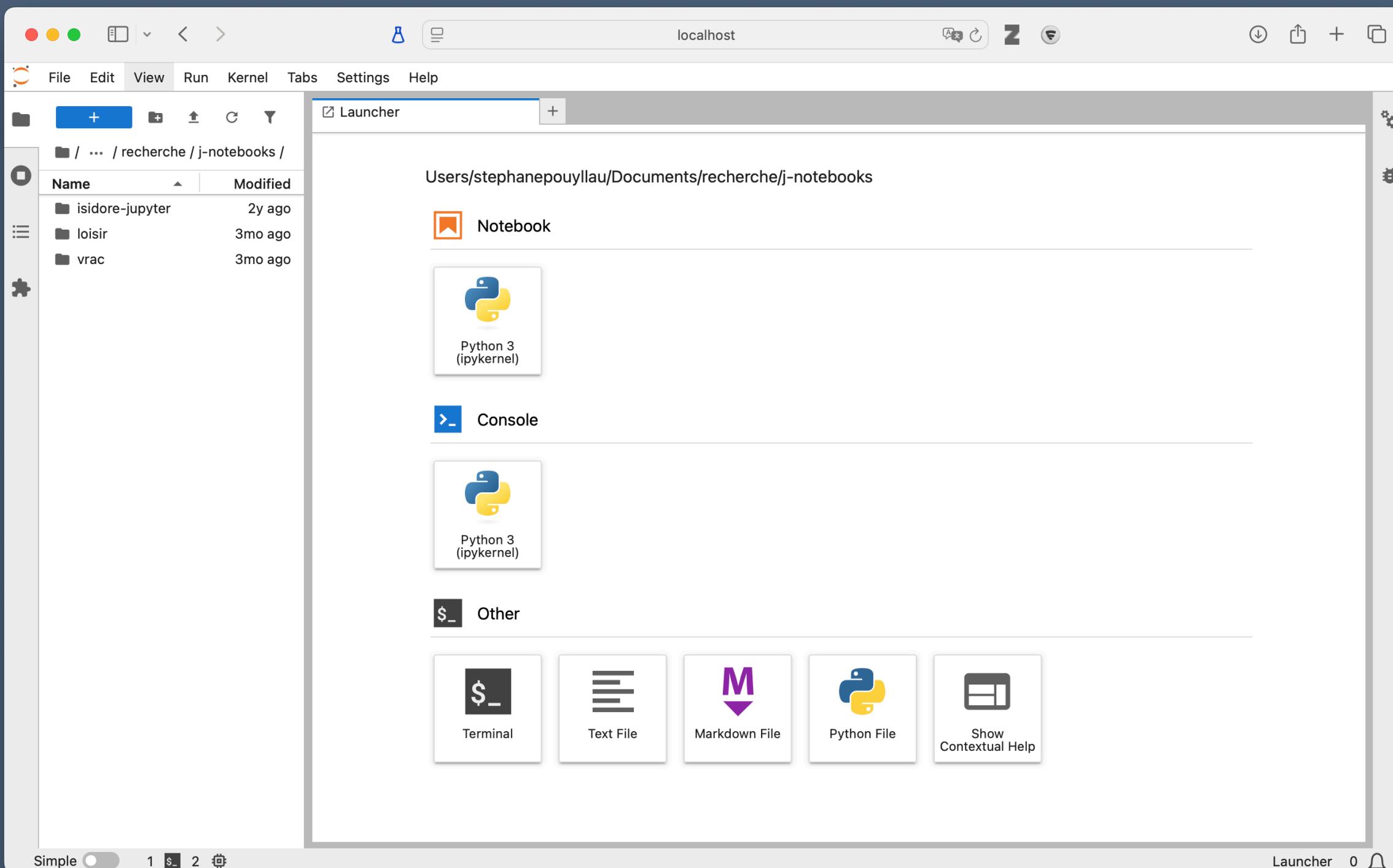
**Book Cover:** The book cover is titled 'PRATIQUE DE LA STATISTIQUE' and 'Python pour les SHS: Introduction à la programmation pour le traitement de données'. It features a black cat sitting on a keyboard, surrounded by icons of cats, laptops, and data visualization symbols. The authors' names, 'EMILIE SCHULTZ et MATTHIAS BUSSONNIER', are at the top. The bottom right corner of the book cover includes logos for 'SES' (Institut National de la Statistique) and 'PUR'.

<https://github.com/pyshs/exemples-manuel>

# Urbanisation d'outils

Résultante de l'analyse des données

- Définition : Choisir et définir, pour un processus d'analyse et de traitement de données, une chaîne d'outils, dialoguant automatiquement ou pas, afin de procéder à la conception d'un système d'information.



# Urbanisation d'outils

Résultante de l'analyse des données

- Outils « clé en main » locaux (à l'échelle individuelle) : OpenRefine, Libre Office, etc.
- Outils « en kit » (à l'échelle individuelle) : IDE + Langages (Python, R, etc.)
- Outils « clé en main » distants (collectifs) : Heurist, Grist, Sharedocs, etc. : <https://documentation.huma-num.fr/sharedocs-traitement/>
- Outils « en kit » (collectifs) : grille de calcul, IDE distant, etc. : <https://documentation.huma-num.fr/calcul-scientifique/>

# Urbanisation d'outils

Résultante de l'analyse des données

- Définir un processus face à un besoin (ou des besoins)
- Faire un état de l'art des méthodes, techniques de traitement des données ou des documents
- Faire une projection à 3 ans et à 10 ans de l'évolution scientifique et technique des objets numériques créés
- Faire une veille continue sur les projets similaires, les métiers qui sont au cœur des processus de "mise en données" (ie. Informatisation des données)

# Urbanisation d'outils

Résultante de l'analyse des données

- Exemple avec une collection de photographies anciennes "Paris ferroviaire autour de 1900"

- Processus **d'analyse et de documentation**

- **Elaboration d'une méthode** de documentation

- **Outils possibles** : MédiHAL / Dbman / etc.

