

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*

...

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

Les points essentiels

- Mise en données et documentation d'informations scientifiques
- Définition des traitements et explicitation (documentation) des algorithmes à utiliser
- Mise en œuvre technique et scientifique (le choix des outils)
- Publication des résultats, des données et des codes



Série "Severance" S1 E2, l'équipe de Mark S. dans la salle de raffinage des données de Lumon.

openrefine.org

OpenRefine is a powerful free, open source tool for working with messy data: cleaning it from one format into another; and extending it with web services and data.

Main features

- Faceting
- Clustering
- Reconciliation
- Infinite undo/redo
- Privacy
- Wikibase

python.org

The Python™ website features a search bar and navigation links for Python, PSF, Docs, PyPI, Jobs, and Community.

Functions Defined

```
# Python is Filenumber series up to n
def fib(n):
    """ Print Fibonacci series up to n """
    a, b = 0, 1
    while a < n:
        print(a, end=' ')
        a, b = b, a+b
    print()
    fib(1000)
```

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

Get Started, **Download**, **Docs**, **Jobs**

cran.rstudio.com

The Comprehensive R Archive Network (CRAN) provides precompiled binary distributions of the base system and contributed packages. It includes sections for Download and Install R, Source Code for all Platforms, and Questions About R.

Download and Install R

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

R is part of many Linux distributions, so 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. This should 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, File of Leaves) [R-4.4.2.tar.gz](#), read [what's new](#) in the latest version.
- The CRAN directory `srcbase-prelease` contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
- Between releases, the same directory `srcbase-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](#).
- Source 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 [an frequently asked questions](#) before you send us an email.

Supporting CRAN

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

LCHS Concept Label sample csv

14550 rows

ConceptLabel	example
Fictional character	3328 Convention of Fictional character
English	1373 English (language)
Vietnam	1298 Vietnamese (language)
Craft	1223 Craft
Italy	1151 Franciacorta (Italy)
New York, N.Y.	1018 Daily News Building (New York, N.Y.)
African people	976 Kenyan (African people)
Iran	1487 Cultural Revolution (Iran)
Alewife	723 Andover Islands Reserve (Alewife)
Extinct city	719 Sumurun (Extinct city)
France	709 Fortuna (France)
Aric	691 Food Hotel Aric (Arist)
India	684 Chittagong Hills (India)
Germany	652 Einfach (Germany)
Russia	649 Zelenovskiy Forest (Russia)
Japan	649 Sakurajima (Japan)
Qing	645 Kunshan (Qing)
China	638 Kunshan (People's Republic of China)
English	626 Children's stories, British (English)
Game	621 Trivii (Game)
Mexico	571 Reserva Ecológica Cuyabá (Mexico)
Scotland	570 Harris (Scotland)
Asia	552 Fressen (Asia)
Ishio	507 Orchard Training Area (Ishio)
Islamic law	498 Game laws (Islamic law)
Tzotzil	492 Camp Walker (Tzotzil)
Computer program language	466 Computer program language
Iran	445 Rock-Roll Music (Iran)
Psi	439 Mitbrook Marsh (Psi)
B.C.	429 Featherstone (B.C.)
N.Z.	429 Puketi Rhododendron Trust Garden (N.Z.)

localhost

RStudio Launcher

Users/stephanepeouillyau/Documents/cours/Evry/2024

- Notebook
- Console
- Python 3 (ipython)
- Terminal
- Text File
- Markdown File
- Python File
- Show Contextual Help

RStudio

R version 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 "copyright" or "license" for distribution details.

R is a collaborative project with many contributors.

Type "contributors" for more information and "citation" to learn 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, Type "q" to quit R.

[Workspace loaded from ~/RData]

Open Source Components

Copy Version Ok

43 B

Langages et IDE

Langages	<i>Regular expression (Regex)</i>	Python	R
IDE	Open Refine	JupyterLab	R Studio

<https://github.com/spouyllau/Cours-Master-Histoire-Evry-2024---C05>

Survivre avec Python

Python

- L'environnement : les jupyter notebook et jupyterlab
- Aller sur <https://jupyter.org/try>
- Choisir JupyterLab : <https://jupyter.org/try-jupyter/lab/>

⚠️ **Experimental** ⚠️ several of the environments below use the **JupyterLite** experimental technology and may have some bugs, so please be patient and expect some unexpected behavior in [the JupyterLite repository](#).

JupyterLab



The latest web-based
interactive development
environment

Jupyter Notebook



The original web application
for creating and sharing
computational documents

Jupyter Widgets



Voilà

voilà

Python

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



Licence 3 | Master | Écoles d'Ingénieurs

SCIENCES SLIP

Cours
Exercices
d'application
corrigés

Patrick Fuchs, Pierre Poulain

Programmation en Python pour les sciences de la vie

2^e ÉDITION

LES + EN
LIGNE

DUNOD

Cours de Python

python.sdv.u-paris.fr

Rechercher

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

Introduction à la programmation Python pour la biologie

Patrick Fuchs et Pierre Poulain

prénom [point] nom [arobase] u-paris [point] fr

 python™

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

- Variables
- Listes
- Boucles et comparaisons
- Tests
- Fonctions

Python

Variables

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.

