



Semaine 1 – Données, Python et notebooks

Outils informatiques appliqués aux géosciences II

Sébastien Biass 

sebastien.biasse@unige.ch

1st floor, office 103

9 octobre 2026



<https://e5k.github.io/outils-info-2/>



CONFLIT DISCRIMINATION VIOLENCE HARCÈLEMENT

Un espace est à votre disposition pour vous aider et vous guider vers les ressources de l'UNIGE:

PERMANENCE POUR LES ÉTUDIANT-ES

Mardi | 9h–12h et 14h–17h

Point Vie de campus

Uni Dufour | 1^{er} étage



UNIGE.CH/HELP



Pourquoi programmer?



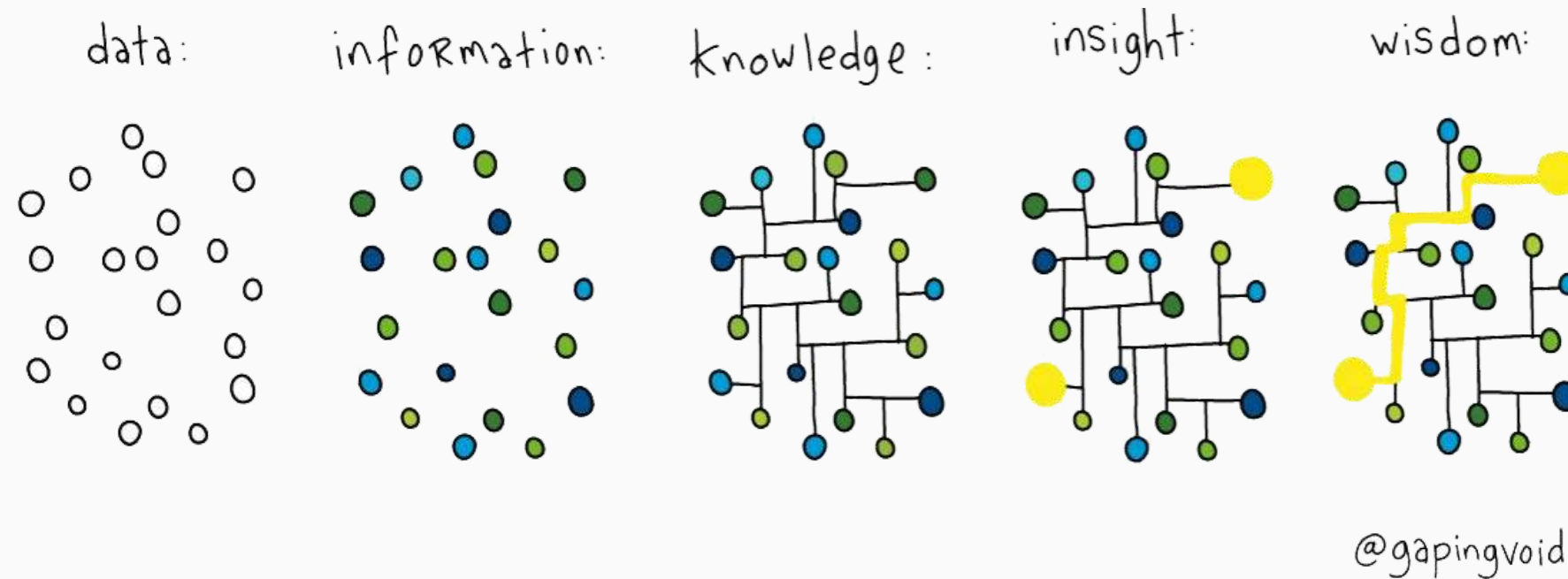
Le déluge de données

Infrastructure	Ordre de grandeur	Contenu
PANGAEA	~447 000 jeux	Géologie marine, géochimie, paléoclimat
Landsat	>13,8 Po	Images de la Terre depuis 1972
EarthScope	>215 To ajoutés en 2024	Données sismologiques et géophysiques
ERA5	~5 Po natifs	Reconstruction horaire du climat

Le problème n'est plus de **trouver des données** mais de savoir **y accéder**, **les valider** et **les exploiter**.



Des données à la décision



1. **Données**: observations brutes
2. **Information**: données organisées et contextualisées
3. **Connaissance**: relations ou mécanismes interprétés
4. **Sagesse**: prise de décision

Recherche scientifique

- **Approche déductive:** l'hypothèse précède les données. Une théorie ou un mécanisme conduit à une prédiction qui est ensuite confrontée aux observations.
- **Approche inductive:** les données précèdent l'hypothèse. L'exploration des observations permet d'identifier des régularités à partir desquelles une explication générale est proposée.
- **Approche abductive:** observations et hypothèses évoluent ensemble. Un résultat surprenant conduit à proposer l'explication la plus plausible, puis à rechercher de nouvelles observations permettant de l'évaluer.

Une analyse reproductible

Entre une observation brute et une conclusion, il faut:

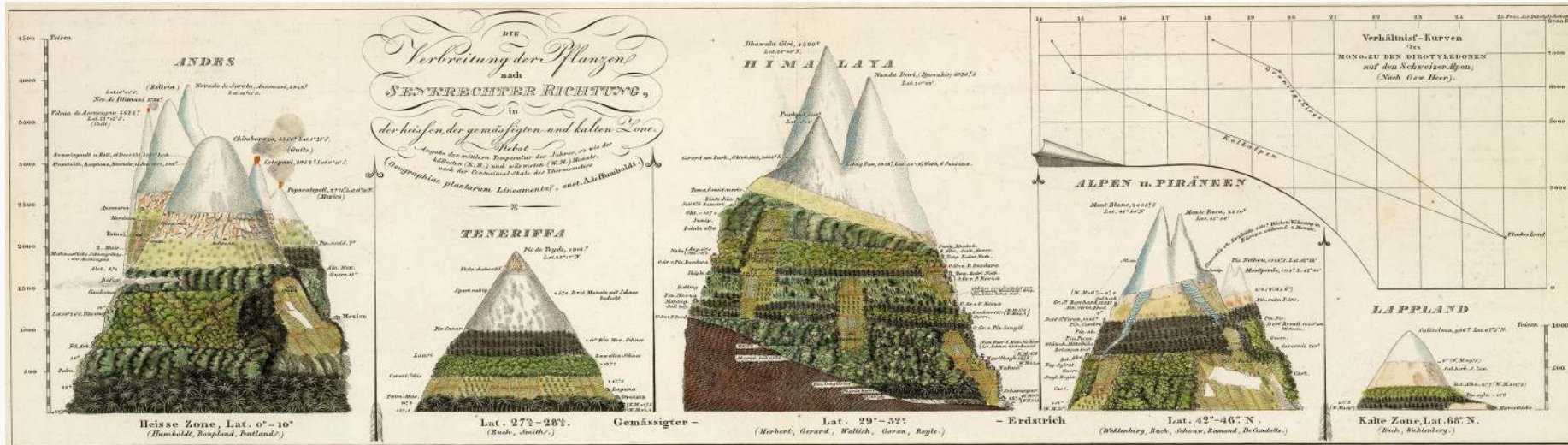
- **Traiter les données**: sélectionner les observations utiles, contrôler qualité et unités;
- **Analyser les données**: analyses statistiques et numériques;
- **Valider les données**: vérifier la robustesse, assurer reproductibilité et transparence;
- **Communiquer les résultats**: visualisation, documenter l'interprétation;



Un bon code conserve une **trace explicite** de ces choix et permet de répéter l'analyse lorsque les données changent.

Visualisation de données

Avant...



Zones de végétations en fonction de la latitude et de l'altitude (Berghaus 1838)

Maintenant... The perfect storm



Pourquoi Python?





Python et son écosystème

Le langage

- Libre et multiplateforme
- Syntaxe relativement lisible
- Notebooks interactifs
- Richesse de bibliothèques scientifiques





Python et son écosystème

Le langage

- Libre et multiplateforme
- Syntaxe relativement lisible
- Notebooks interactifs
- Richesse de bibliothèques scientifiques

Les bibliothèques (= Libraries)

- **pandas**: données tabulaires
- **NumPy** et **SciPy**: calcul et statistiques
- **Matplotlib** et **seaborn**: figures
- **statsmodels**: inference statistique
- **xarray**, **GeoPandas**, **rasterio**: géodonnées



Python comme une langue

Français	Python	Exemple
Nom	objet ou variable	<code>volcans</code> , <code>df</code>
Verbe	fonction ou méthode	<code>mean()</code> , <code>sort_values()</code>
Complément	variable ciblée	<code>"Elevation (m)"</code>
Adverbe	argument	<code>ascending=False</code>
Phrase	instruction	<code>df.sort_values(...)</code>

```
1 df.sort_values("Elevation (m)", ascending=False)
```

⚠ Python contrôle la grammaire du code. Il ne contrôle pas la **pertinence scientifique** de la phrase.



Administration

Objectif du cours

Développer une autonomie scientifique

- Formuler une **question de recherche** et avoir les outils pour y répondre
- Choisir les données pertinentes et les préparer
- Produire un résultat vérifiable
- Interpréter ses limites

Développer une pensée statistique

- Décrire et visualiser les données (stats **uni** et **bivariées**)
- Tester des hypothèses
- Inférer les propriétés d'une **population** à partir d'un **échantillon**

Informations pratiques

Cours

- Vendredi, 13h15–15h00
- Salle 0013, Sciences III
- Introduction + exercice guidé
- Assistante: [Ana Gonzalez](#)
- Autonomie augmente au fil du semestre
- Exercices → avancer l'argument scientifique

Support

- Site du cours comme manuel
- Commentaires avec [Hypothesis](#)
- Questions et retours bienvenus



Plan du semestre

Semaines	Données	Objectif
1	Installation et premiers exemples	Environnement de travail
2	Volcans	Interroger des données
3–5	Tsunamis	Décrire et visualiser un échantillon
6–8	Pétrophysique	Comparer et quantifier des relations
9–11	Recherche personnelle	Projet et communication

Travail accompli ce semestre = **pré-requis** pour **Analyse spatiale** (printemps) et **travail de bachelor**



Structure d'une séance

Introduction

- Contexte scientifique, données, théorie et nouvelles fonctions

Exercise

- Exploration initiale + question de recherche
- Développement de code autonome

Fin du cours

- Réflexion sur limites / prochaines étapes
- Identification de questions supplémentaires
- Publication des réponses + notebook corrigé

Resources

Livres

- Petrelli M (2021) Introduction to Python in Earth Science Data Analysis: From Descriptive Statistics to Machine Learning (**Petrelli 2021**)
- Trauth MH (2022) Python Recipes for Earth Sciences (**Trauth 2022**)

Cheat sheets

- PDF résumant les fonctions les plus utilisées pour chaque bibliothèque

Documentation

- Documentation simplifiée pour chaque bibliothèque



Examen

Format

Par groupe de 2, identifier une question de recherche, trouver un jeu de données, réaliser une analyse, écrire un **rapport** sous forme d'un notebook et **présentation** à la classe.



Examen

Format

Par groupe de 2, identifier une question de recherche, trouver un jeu de données, réaliser une analyse, écrire un **rapport** sous forme d'un notebook et **présentation** à la classe.

Evaluation

Notebook

- Question et données
- Méthode reproductible
- Résultats quantifiés
- Discussion des limites

Présentation

- Même raisonnement scientifique
- Figures lisibles
- Conclusion défendable
- Semaine 10 → 11 décembre

Usage des LLM

Un LLM peut faire gagner du temps si vous pouvez:

- Lire et expliquer le code proposé
- Identifier les fonctions et leurs arguments
- Vérifier que les variables correspondent aux données
- Tester si le résultat répond à la question



Accepter une réponse sans la comprendre empêche d'évaluer sa **validité scientifique**.

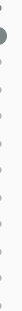
Programme de cette semaine





Plan de la séance

1. Installer ou vérifier **Miniconda** et l'environnement `outils-info`
2. Configurer **VS Code** et ses extensions
3. Créer un premier notebook **Jupyter**
4. Alternner cellules Python et **Markdown**
 - Prenez un peu de temps pour étudier le guide **Markdown**
5. Vérifier que le notebook se réexécute dans l'ordre



Les quatre composantes

1. Miniconda

Gère des environnements Python isolés.

2. Environnement `outils-info`

Contient Python et les bibliothèques du cours.

3. Jupyter

Exécute et documente l'analyse dans un notebook.

4. Visual Studio Code

Éditeur utilisé pour organiser les fichiers.



Jupyter dans VS Code

Annotations: I, II, III, IV, V (green); A, B, C, D (purple); ENV (red); VI (green).

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

print("✓ Félicitations ! Votre environnement Python et Jupyter sont correctement configurés.")

# Petit test de visualisation (données fictives de rapports géochimiques)
contextes = ['MORB', 'OIB', 'Arc']
ba_nb_ratio = [20, 30, 500]

plt.bar(contextes, ba_nb_ratio, color=['blue', 'orange', 'red'])
plt.ylabel('Rapport Ba/Nb')
plt.title('Test de visualisation - Outils Info Géosciences')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```

Terminal Output:

```
✓ Félicitations ! Votre environnement Python et Jupyter sont correctement configurés.
```

Jupyter Variables:

Name	Type	Size	Value
ba_nb_ratio	list	3	[20, 30, 500]
contextes	list	3	['MORB', 'OIB', 'Arc']

Deux types de cellules

Code

```
1 altitude = 4200  
2 print(alitude)
```

→ Produit immédiatement un texte, un tableau ou une figure.

Markdown

```
1 ## Résultats  
2  
3 Le volcan culmine à **4200 m**.
```

→ Documente la question, la méthode et l'interprétation.

Utiliser correctement un notebook

- ✓ Organiser les cellules dans un ordre lisible, du début à la conclusion
- ✓ Exécuter une cellule après chaque modification
- ✓ Expliquer en Markdown pourquoi chaque analyse est réalisée
- ✓ Conserver les sorties nécessaires à l'interprétation
- ✓ Avant de partager: **Restart Kernel**, puis **Run All**

! Un notebook fiable doit produire le même résultat lorsqu'une autre personne l'exécute du début à la fin.

Le kernel conserve la mémoire

Le kernel garde en mémoire les variables et fonctions déjà exécutées.

```
1 altitude = 4200
```



Modifier cette cellule sans la réexécuter ne modifie pas la valeur mémorisée!



Règle d'or

Après une modification importante:

1. Redémarrer le kernel pour effacer la mémoire
2. Ré-exécuter toutes les cellules dans l'ordre.

Markdown comme carnet scientifique



Documentez votre réflexion, pas juste le code!

```
1 # Question
2 La hauteur des volcans dépend-elle du contexte tectonique ?
3
4 ## Méthode
5 Nous comparons les distributions d'altitude.
6
7 ## Conclusion
8 ...
```

- **Markdown** permet aussi d'insérer des **liens**, des images, des tableaux, du **code** et des équations comme $PV = nRT$.
- Le notebook doit pouvoir être compris par une personne qui ne connaît pas encore les données.

Raccourcis essentiels

Action	macOS	Windows / Linux
Exécuter et avancer	Shift + Entrée	Shift + Entrée
Exécuter et rester	Cmd + Entrée	Ctrl + Entrée
Palette de commande	Cmd + Shift + P	Ctrl + Shift + P
Sauvegarder	Cmd + S	Ctrl + S

A vous!

1. (Re-) lisez l'**introduction** à un moment ou un autre
2. **Suivez méthodiquement** le **guide d'installation**
3. Lisez l'introduction à **Jupyter** et **Markdown**

Tip

- `Cmd + Shift + P` | `Ctrl + Shift + P`: Open web browser
- Entrez https://e5k.github.io/outils-info-2/course/semaine_01_intro.html
- Cliquez sur le tab **Browser** et bougez-le sur la droite de l'écran pour avoir instructions et notebook sur un même écran

References

Berghaus HKW (1838) Physikalischer Atlas: Oder Sammlung von Karten, Auf Denen Die Hauptsächlichsten Erscheinungen Der Anorganischen Und Organischen Natur Nach Ihrer Geographischen Verbreitung Und Vertheilung Bildlich Dargestellt Sind. Justus Perthes, Gotha

Petrelli M (2021) **Introduction to Python in Earth Science Data Analysis: From Descriptive Statistics to Machine Learning**. Springer International Publishing

Trauth MH (2022) **Python Recipes for Earth Sciences**. Springer International Publishing, Cham