

Data Visualisation avec Python

Formation proposée par le groupe MALIA de la SFdS

Baptiste Gregorutti (LumenAi), Charlotte Laclau (UDL)

June 10, 2021

Association à but non lucratif fondée en 1997, reconnue d'utilité publique

Missions

- promouvoir l'utilisation de la statistique et sa compréhension et favoriser ses développements méthodologiques
- lieu privilégié de rencontres, d'échanges et de réflexion des acteurs
- majeurs du domaine provenant des administrations, des entreprises ou des établissements d'enseignement ou de recherche
- Groupes Spécialisés: Machine Learning et Intelligence Artificielle
- **Adhérez !**

Objectifs

Se familiariser avec les librairies de visualisation en python, notamment

- matplotlib : librairie incontournable
- seaborn : plus de fonctionnalités
- plotly : visualisation interactive

Organisation

1. Quelques rappels en Python
2. Panorama des librairies de data visualisation
3. TP1 : matplotlib, seaborn et pandas
4. TP2 : plotly

Quelques rappels en Python

Python impose l'indentation

```
1 for elt in [1, 2, 3]:  
2     print(elt)
```

→ Exécution OK

```
1 for elt in [1, 2, 3]:  
2 print(elt)
```

→ IndentationError: unexpected indent

Quelques rappels en Python

Programmation orientée objet

Python est un langage orienté objet dans lequel des classes définissent une structure de données ou un type :

```
1 # obj est un objet de classe Myclass
2 obj = Myclass()
3
4 # one_method est une methode de la classe Myclass
5 obj.one_method()
6
7 # one_attribute est un attribut de la classe Myclass
8 obj.one_attribute
```

Points importants :

- Une classe doit toujours être définie avec une lettre capitale, même si Python est très permissif
- Le '.' désigne l'accès à des méthodes et des attributs de classes. Grande différence avec R.

Quelques rappels en Python

Importation de librairies

```
1 # Importe le sous-package pyplot de matplotlib
2 import matplotlib.pyplot
3 fig = matplotlib.pyplot.figure()
4
5 # Idem avec un alias
6 import matplotlib.pyplot as plt
7 fig = plt.figure()
8
9 # Importe la fonction figure de pyplot
10 from matplotlib.pyplot import figure
11 fig = figure()
12
13 # Importe tous les elements de pyplot
14 from matplotlib.pyplot import *
15 fig = figure()
```

Quelques rappels en Python

PROSCRIRE l'utilisation de 'import *'

Exemple : la fonction `sqrt` existe dans plusieurs libraries :

- Dans `math` : calcule la racine carré pour un scalaire
- Dans `numpy` : calcule la racine carré pour un scalaire ou pour chaque élément d'un tableau

```
1 from numpy import *
2 from math import *
3 print(sqrt([1, 2, 3]))
```

→ `TypeError: must be real number, not list`

Panorama des librairies: périmètre de la formation

Matplotlib

- Librairie *basique* de visualisation avec Python
- Peut s'utiliser dans des scripts python, shells Python et IPython , Jupyter notebook, serveurs d'application web.

Seaborn

- Librairie basée sur Matplotlib
- Graphiques statistiques plus attrayants et informatifs

Pandas

- Librairie basée sur Matplotlib
- Adapté pour les objets pandas.DataFrame (se rapproche de R)

Plotly

- Librairie de visualisation dédié à la création de graphe interactif et de cartes
- Disponible pour Python, R et Javascript
- Permet d'afficher des graphiques dans des notebooks Jupyter; sauvegarder des graphiques au format HTML; intégration dans des applications web (Dash).

Si vous avancez très très vite !

- **Dash** : développement d'une application web (création de tableau de bord)
- **Dtale** : client web, basée sur Plotly, pour visualiser des données tabulaires (notamment de type `pandas.DataFrame`)

Panorama des librairies: en dehors du périmètre

ggplot

- Implémentation python de ggplot2 (librarie R)
- Chevauchement de fonctionnalités avec ggplot2, mais pas nécessairement un mimétisme

Bokeh

- Très proche de plotly (également basé sur du js)
- Ne gère pas la visualisation 3D

Altair

- S'appuie sur le langage Vega (format json)

Folium

Visualiser des cartes, basé sur OpenStreepMap

- 9h30: TP1 - Matplotlib, seaborn et pandas
- 10h30: Pause
- 10h45: TP2 - Plotly
- 12h15: Temps d'échange

Des questions ?