



DataViz - Itération 1

on the road to mastering Matplotlib!



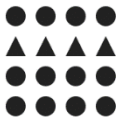
Good Figure

Objects grouped together tend to be perceived as a single figure. Tendency to simplify.



Proximity

Objects tend to be grouped together if they are close to each other.



Similarity

Objects tend to be grouped together if they are similar.



Continuation

When there is an intersection between two or more objects, people tend to perceive each object as a single uninterrupted object.



Closure

Visual connection or continuity between sets of elements which do not actually touch each other in a composition.



Symmetry

The object tend to be perceived as symmetrical shapes that form around their center.

THE GESTALT PRINCIPLES (SOURCE : [HTTPS://UXCHEAT.COM/](https://uxcheat.com/))

Objectifs du module

Savoir raconter une histoire avec vos données.

Nous sommes dans une époque où nous générons énormément de données. Un piège courant est de croire qu'une quantité importante de données facilite le traitement du sujet étudié. Dans la pratique, c'est tout le contraire. Plus nous avons d'information à traiter, plus il est nécessaire de se poser les bonnes questions :

- Quel visualisation me permet de mieux explorer/comprendre mes données ?
- Quel visualisation me permet de mieux communiquer mes conclusions ?

Vous l'aurez compris, il existe *deux* étapes à bien distinguer :

- Une étape d'**exploration** de vos données
- Une étape de **présentation** des résultats

L'objectif de ce module est donc de vous sensibiliser à ces questions clés, de vous aider à choisir la (ou les) visualisation(s) qui feront sens selon le sujet étudié.


Ce module est décomposé en quatre itérations réparties tout au long de votre cursus (trois journées pendant la période "socle" et une journée pendant la période "alternance" de la formation).

Le but de cette première itération est de vous familiariser avec l'outil de visualisation le plus utilisé dans le langage de programmation Python, à savoir la librairie Matplotlib!

Objectif de cette journée

Les objectifs de cette première journée sont les suivants:

1. Acquérir une maîtrise de la librairie matplotlib.
2. Réaliser des graphiques pour explorer un jeu de données réel.

 La manipulation de données réelles est une compétence qui va s'acquérir sur le long de votre parcours!

Démarche Pédagogique et Modalités

- Travail en autonomie
- Entraide possible
- Un notebook Jupyter est fourni pour chaque partie. Chaque notebook contient un ensemble d'exercices à réaliser avec des consignes à suivre, et parfois des questions plus ouvertes auxquelles il faudra répondre directement dans le notebook.

Compétences développées

Il y a quatre compétences DataViz en tout sur l'application Campus-Skills, seulement deux sont à valider pour cette première itération:

- **Réaliser des figures à l'aide de Pyplot:**

Je maîtrise l'interface pyplot et je sais créer des figures simples tel que des line plots, histogrammes, scatter plots...

- **Maîtriser matplotlib:**

Je sais utiliser les objets "Axes" et "Figure" de la librairie Matplotlib pour personnaliser mes figures.

Il vous suffit de terminer le premier notebook intitulé "**DataViz01-Matplotlib_Introduction.ipynb**" pour valider les compétences de cette première journée de DataViz.

Quelques ressources utiles:

- Documentation Matplotlib: <https://matplotlib.org/stable/index.html>
- Cheatsheets: <https://matplotlib.org/cheatsheets/>

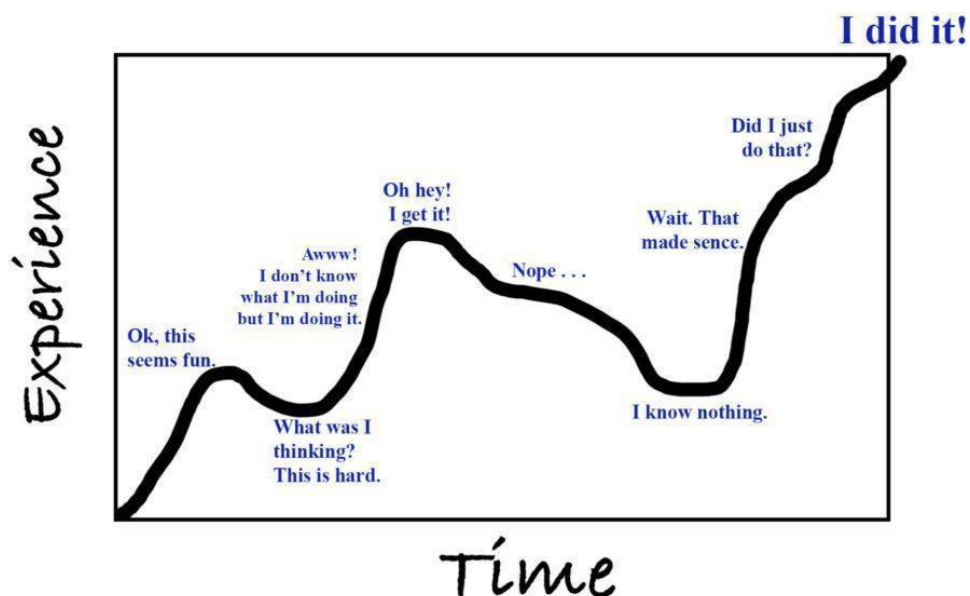
Partie 1

Introduction à matplotlib – (0.5 jour)

Cette itération a pour objectif de :

1. Comprendre la manière dont est structurée un graphique
2. Acquérir le vocabulaire liée au tracé des courbes
3. Personnaliser vos figures afin de les rendre plus intelligibles
4. Réaliser différents types de graphiques : line, histogramme, scatter.

The Learning Curve



Consignes

Téléchargez le notebook **DataViz01-Matplotlib_Introduction.ipynb** et suivre la trame proposée.

Le but de ce notebook est de vous aider à acquérir les bases de la librairie Matplotlib, en particulier à la fin de ce notebook vous saurez:

- Tracer des courbes à l'aide de l'interface pyplot
- Tracer des courbes à l'aide de l'interface artist

Partie 2

Utiliser Matplotlib avec Pandas - (0.5 jour)

Cette itération est moins guidée, le but est de vous laisser vous approprier les concepts vus dans la première partie. Vous allez aussi pouvoir approfondir votre maîtrise de la librairie Pandas en explorant son interfacement avec matplotlib qui vous permettra de tracer rapidement des courbes pendant votre phase d'exploration des données!

Contenu proposé :

- Manipuler des données réelles.
- Tracer d'autres types de graphiques: *pie-chart*, *area* et *box plots*.

Consignes

- Utiliser le jupyter notebook **DataViz01_Matplotlib-Pandas.ipynb** et suivre la trame proposée (ou pas... si vous vous sentez d'humeur créative c'est le moment de libérer l'artiste qui est en vous!)
- Si vous vous sentez déjà assez à l'aise avec matplotlib, vous pouvez en profiter pour essayer les librairies [Seaborn](#) et/ou [Plotly](#).

Références

- Visualisations proposées par RTE: <https://www.rte-france.com/fr/eco2mix/eco2mix>
- Plotly dans Pandas: <https://plotly.com/python/pandas-backend/>