



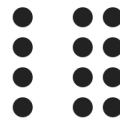
DataViz

Etape 1 : Maitriser 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 ?

Il existe donc deux étapes à bien distinguer :

- Une étape d'exploration
- Une étape de présentation de résultats

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

Ce module est décomposé en trois étapes, espacées le long de votre cursus.

Cette première étape est destinée à la maîtrise de l'outil de visualisation le plus utilisée dans le langage python : Matplotlib.


Objectif de cette journée

Pour cette étape (première journée) l'objectif est :

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

Modalités

- La journée est découpée en 2 itérations :
 1. Introduction à matplotlib
 2. Faire des plots sur un jeu de données réels
- Ce projet sera réalisé en autonomie
- Un jupyter notebook est fourni par itération. Chaque notebook contient : Une ensemble d'exercices à faire avec des consignes et/ou des questions ouvertes.

 La manipulation de données réelles est une compétence qui va s'acquérir sur le long de votre parcours. De ce fait, il suffit de finaliser la première itération pour valider les compétences de cette première journée de DataViz.

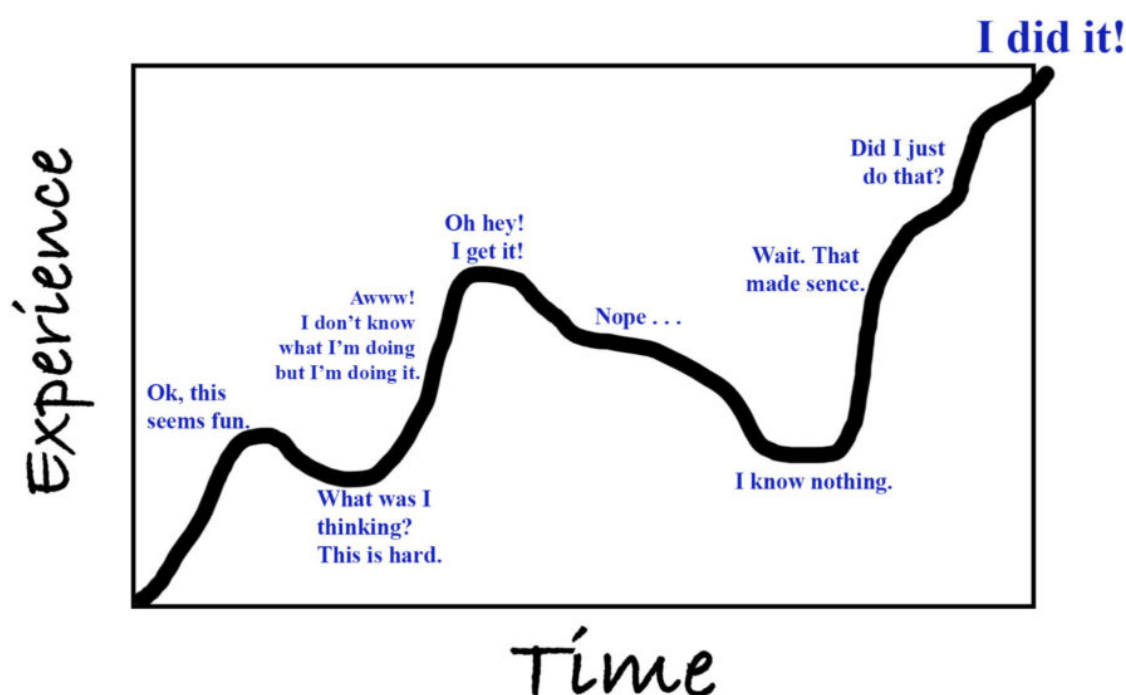
Itération 1

Introduction à matplotlib – (0.5 jour)

Cette itération a pour objectif de

1. Comprendre la structure d'un graphique
2. Acquérir le vocabulaire liée au tracée des courbes
3. *Customiser* vos graphiques afin de les rendre plus intelligibles
4. Réaliser différents types de plots : line, histogramme, scatter.

The Learning Curve



Consignes

Utiliser le jupyter notebook 'DataViz_IntroMatplotlib_Etu.ipynb', et suivre la trame proposée

Compétences à valider

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

Itération 2

Utiliser matplotlib avec pandas - (0.5 jour)

Cette itération est plutôt ouverte. Le but est de vous laisser vous approprier des concepts vus lors de la première itération, ainsi que maîtriser matplotlib et son interfaçage avec Pandas.

Contenu proposé :

- Manipuler des données réelles.
- Tracer d'autres types de courbes: pie, area, box plots.

Consignes

- Utiliser le jupyter notebook 'DataViz_RTE_Etu.ipynb', et suivre (ou pas !) la trame proposée.
- Si vous avez déjà une bonne maîtrise de matplotlib, profitez-en pour essayer plotly.

Références

R 2.1. – Visualisation proposés par RTE

<https://www.rte-france.com/fr/eco2mix/eco2mix>

R 2.2. – Plotly & pandas

<https://plotly.com/python/pandas-backend/>