

TP Statistiques et Matplotlib

Objectif

L'objectif de ce TP est de prendre en main certaines fonctionnalités de la bibliothèque Matplotlib. À la fin de ce TP, vous serez capable de construire des représentations graphiques, en particulier, dans le domaine des Statistiques et vous aurez travaillé/élaboré des stratégies rencontrées lors de l'étude de données statistiques.

I. Premières manipulations (Débutants)

1. En vous aidant de l'aide mémoire, importer la bibliothèque matplotlib.
2. Placer 3 points en forme de croix rouge dont vous aurez choisi les coordonnées.
3. Créer deux listes possédant le même nombre de valeurs, une pour les valeurs en abscisses et l'autre pour les valeurs en ordonnées puis tracer la ligne polygonale passant par ces points.

Question supplémentaire

Écrire un programme où l'on crée une fonction avec l'expression de votre choix, puis, à l'aide d'une boucle, on réalise un tableau de valeurs avec une variable variant de 1 à 5 avec un pas de 0,2 (les données seront stockées dans deux listes, une pour les valeurs de la variable, une autre pour les images). Afficher enfin la ligne polygonale que l'on obtient avec les deux listes précédentes.

II. Représenter des données (Intermédiaires)

1. Dans un nouveau script, importer le fichier `département` qui contient, entre autre, les variables :
 - `numero` (une liste avec le numéro de chaque département hors Corse et DOM-TOM)
 - `population` (une liste avec la population, en milliers d'habitants, de chaque département hors Corse et DOM-TOM en 2014).
2. Afficher un histogramme composé de 25 classes des populations des départements.
3. Afficher un nuage de points (en forme de ronds verts) ayant pour abscisse le numéro des départements et pour ordonnées leur population.
4. Compléter votre graphique en ajoutant un titre, un légende pour chaque axe en utilisant : `title("titre")`, `xlabel("legende1")`, `ylabel("legende2")` avant le `show()`.
5. Par lecture graphique, quels sont les trois départements qui ont la plus grande population ? Déterminer le département de population maximale à l'aide d'une série d'instructions.

Question supplémentaire

Écrire un algorithme, puis le programme associé, qui permettrait de déterminer la valeur médiane d'une série statistique (dont les données seraient stockées dans une liste). L'idée est de créer une fonction `mediane`. L'aide mémoire est précieux ! Utiliser votre programme pour déterminer la population médiane des départements étudiés.

III. Travailler avec des séries de données (Experts)

1. Dans un nouveau script, importer le fichier `département` qui contient, entre autre, les variables :
 - `numero` et `population` (voir ci-avant) ;
 - `nombre_enfant` (une liste avec le nombre d'enfants par famille, 4 correspondant à « 4 enfants ou plus ») ;
 - `nombre_famille_SEL` (une liste du nombre de familles, en milliers, associé à la liste `nombre_enfant` en Saône-et-Loire en 2014) ;
 - `echantillon` (nombre d'enfants dans un échantillon représentatif de 200 familles françaises).

2. Afficher dans une même fenêtre (voir le document ressource « Utilisation de Matplotlib ») :
 - un diagramme en bâton → `bar(ListeX, ListeY)` ; ou un nuage de points de triangles bleus ; ayant pour abscisse le nombre d'enfants par famille et pour ordonnées les effectifs correspondants en Saône-et-Loire
 - un histogramme représentant la variable `echantillon`
3. Comparer les deux représentations précédentes et critiquer la « structure familiale » en Saône-et-Loire par rapport à l'ensemble de la France. (Attention à bien considérer le nombre total de familles à chaque fois)
4. Lorsque l'on souhaite mettre en relation trois grandeurs numériques, le graphique à bulles est un outil très utile ! Afficher le graphique à bulles donnant la population par département et une bulle plus ou moins imposante en fonction du nombre de villes de plus de 50000 habitants. Dans le fichier `departement`, la variable `commune2` permet un affichage plus adapté pour la surface des bulles et des couleurs ont été stockées dans la variable `couleur`. Compléter avec un titre et des légendes pour les axes en utilisant : `title("titre")`, `xlabel("legende1")`, `ylabel("legende2")` avant le `show()`.

Question supplémentaire

- Écrire un programme, qui contient une fonction qui permet de réaliser les effectifs cumulés croissants. Cette fonction prendra donc en paramètre la liste des effectifs et renverra la nouvelle liste souhaitée. Utiliser cette fonction pour obtenir la liste `nombre_famille_SEL_cumule` qui contient les effectifs cumulés croissants du nombre d'enfants par famille en Saône-et-Loire.
- Poursuivre ce programme pour faire en sorte de déterminer graphiquement la médiane de cette série. Il faudra afficher la ligne polygonale souhaitée mais aussi le segment attendu dont l'ordonnée des points sera calculée à l'aide de la liste des effectifs cumulés croissants. (On pourra aussi traiter les cas du premier quartile et du troisième quartile). Interpréter le résultat obtenu pour la médiane.