TP1 Analyse Univariée en Python

2026-2027

https://www.umr-lastig.fr/martin-cubaud/cours/

Objectif principal : Analyser et cartographier la distribution statistique et spatiale des arbres dans Paris
Objectif secondaire : revoir les concepts d'analyse univariée, et réaliser des cartes avec Python.

0.1 Les données

0.1.1 Arbres

Les données proviennent du site opendata.Paris.fr (url)
Ils sont disponibles en plusieurs formats (KML, geoJSON, SHP, CSV…)

Les variables du jeu de données sont :

- *remarquable* : le caractère remarquable (1) ou non (0) de l'arbre; contient des valeurs NA.
- *circonferenceencm* : la circonférence en centimètres du tronc de l'arbre
- *stadedeveloppement* : stade de développement ∈ {"A" , "J" , "JA" ,"M"} ; contient des valeurs NA.
- *idbase* : identifiant unique de l'arbre.
- *arrondissement* : libellé de l'arrondissement où se trouve l'arbre, ainsi que les zones "BOIS DE BOULOGNE", "BOIS DE VINCENNES", "HAUTS-DE-SEINE", "SEINE-SAINT-DENIS", "VAL-DEMARNE"
- *idemplacement* : code de l'emplacement de l'arbre, non unique.
- *geometry* : attribut géometrique de l'arbre : simple feature de classe POINT (XY)
- *adresse* : libellé de l'adresse de l'emplacement de l'arbre e.g. "BOULEVARD SAINT GERMAIN"
- *libellefrancais* : libellé vernaculaire (=en français du quotidien) de l'espèce de l'arbre e.g. "Marronnier"
- *complementadresse* : complément de l'adresse, souvent le numéro; contient des valeurs NA
- *domanialite* : type de localisation de l'arbre ∈ {"Alignement","CIMETIERE","DAC","DASCO", "DASES","DFPE","DJS","Jardin","PERIPHERIQUE"}
- *typeemplacement* : variable constante égale à "Arbre"
- *hauteurenm* : hauteur en mètres
- *genre* : genre de l'arbre, 192 modalités, contient des valeurs NA
- *espece* : espèce de l'arbre, 537 modalités; contient des valeurs NA
- *varieteoucultivar* : variété de l'arbre, 453 modalités, contient des valeurs NA

Précision sur la hiérarchie des types d'arbres : l'ordre est Genre > Espèce > Variété (ou cultivar)

0.1.2 Contour des quartiers administratifs

Les contours des quartiers proviennent également du site opendata.Paris.fr (url) Chaque arrondissement est constitué de 4 quartiers.

Les variables disponibles sont :

- *n_sq_qu* : identifiant séquentiel du quartier, constitué de la concaténation de 75000 et du code de quartier
- *c_qu* : code du quartier , valeur entière ∈ [1; 80]
- *c_quinsee* : Numéro INSEE du quartier, valeur entière. format 751AAQQ avec AA le numéro d'arrondissement ∈ [1; 20] et QQ le numéro de quartier ∈ [1; 4]
- *l_qu* : libellé du quartier , e.g. "La Chapelle"
- *c_ar* : numéro d'arrondissement , valeur entière ∈ [1; 20]

- *n_sq_ar* : identifiant séquentiel de l'arrondissement , ∈ [750000001; 750000020], format 7500000AA, avec AA le numéro de l'arrondissement.
- *perimetre* : périmètre de la géométrie du quartier
- *surface* : aire de la géométrie du quartier
- *geometry* : attribut géometrique du quartier : simple feature de classe POLYGON

1 Distribution statistique des hauteurs des arbres

Nous allons utiliser les packages `geopandas` et `shapely`, installables avec `pip`.
Fonctions utiles :

- `gpd.read_file` pour charger les données spatiales
- Méthode `plot` pour visualiser les données
- Méthode `sjoin` pour réaliser une jointure spatiale
- Méthode `to_crs` pour projeter les données.

Vous disposez de deux jeux de données :

- Le contour des quartiers de Paris (vectoriel, polygones) ;
- L'implantation et les variables des arbres de Paris (vectoriel, ponctuel).

1.1 Charger le jeu de données des arbres, et projeter la dans un système de coordonnées géographiques planes à l'aide de la méthode.
1.2 On peut émettre l'hypothèse que les arbres remarquables sont censés être les plus grands et qu'il y a moins d'erreur dessus. Quelle est la hauteur du plus grand arbre remarquable ? Quelle est la plus grande circonférence parmi les arbres remarquables ?
1.3 Filtrer les arbres avec une taille aberrante, trop grande ou trop petite.
1.4 Tracer l'histogramme des hauteurs et des circonférences des arbres filtrés
1.5 Tracer la carte des hauteurs des arbres
1.6 A l'aide de la méthode `value_counts`, quelles sont les 6 genres (vernaculaires) les plus fréquents ?
1.7 Comparer les distributions des hauteurs pour ces 6 genres sur un même graphique.
1.8 Calculer pour chacun de ces 6 genres la moyenne et la médiane. Lequel est « généralement » le plus grand ?
1.9 Calculer l'écart-type et l'écart interquartiles des hauteurs par genre. Quel genre est le plus homogène en hauteur ?
1.10 Comparer les asymétries des genres.

2 Nombre et densité d'arbres par quartier

2.1 Calculer le nombre d'arbres par quartier

Étapes :

- Identifier les données pertinentes pour réaliser ce calcul ;
- Réaliser une jointure spatiale (laquelle ?) ;
- Calculer le nombre d'arbres par quartier ;
- Stocker le résultat dans une variable *nb_arbres* dans la couche vectorielle des quartiers.

2.2 Calculer la densité d'arbres par quartier

Étapes :

- Étapes précédentes
- Calculer la surface des quartiers (certes elle existe déjà, mais on ne sait pas comment elle a été calculée. E.g. la projection est-elle équivalente ?)

- Calculer la densité d'arbres par quartiers
- Stocker le résultat dans une variable *dens_arbres* dans la couche vectorielle des quartiers

Attribut utile :

- `area` pour calculer l'aire d'un polygone

2.3 Comparaison des variables nombre et densité

Étapes :

- Calculer la moyenne et l'écart-type du nombre et de la densité d'arbres
- Afficher les histogrammes de ces deux variables
- Calculer les indicateurs de formes (kurtosis et skewness) des distributions de ces deux variables

Commenter les résultats obtenus

3 Distribution spatiale des arbres

3.1 Cartographie du nombre d'arbres

Vous pouvez au choix :

- Réaliser une carte choroplèthe du nombre d'arbre par quartier (cartographier un nombre est déconseillé, pourquoi ?)
- Réaliser une carte avec des symboles proportionnels représentant le nombre d'arbres par quartier

  Obtenir le centre des polygones avec `geodataframe.geometry.centroid`
- Réaliser une carte de chaleur du nombre d'arbres sur une grille raster

Commenter les cartes obtenues.

3.2 Cartographie de la densité d'arbres

Etapes :

- Cartographier la densité d'arbres par quartiers

Commenter la carte obtenue

4 Export des données

Exporter les données ajoutées (nombre et densité d'arbres) dans la couche vectorielle des quartiers de Paris (format SHP ou CSV)

- Méthode : `to_file`

5 Pour aller plus loin :

A l'aide des données de voiries de Paris disponibles ici, peut-on établir un lien entre le genre des arbres et leur distance à la voirie la plus proche ?