

TP1 Analyse Univariée

M2 IGAST 2019-2020

Objectif principal : Analyser et cartographier la distribution statistique et spatiale des arbres dans Paris

Objectif secondaire : Préparer les données pour la classification

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.
- *circonferencecm* : la circonférence en centimètres du tronc de l'arbre
- *stadedeveloppement* : stade de développement $\in \{“A”, “J”, “JA”, “M”\}$; contient des valeurs NA.
- *genre* : nom latin du genre de l'arbre, 175 modalités, ; 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-DE-MARNE”
- *idemplacement* : code de l'emplacement de l'arbre, non unique.
- *geo_point_2d* : coordonnées [X,Y] de l'arbre en texte brut e.g.[48.8409288153,2.27798983391]
- *geometry* : attribut géométrique de l'arbre : simple feature de classe POINT (XY)
- *adresse* : libellé de l'adresse de l'emplacement de l'arbre
- *libellefrançais* : libellé de l'espèce de l'arbre e.g. “BOULEVARD SAINT GERMAIN”
- *complementadresse* : complément de l'adresse, souvent le numéro; contient des valeurs NA
- *domanialite* : type de localisation de l'arbre $\in \{“Alignement”, “CIMETIERE”, “DAC”, “DASCO”, “DASES”, “DFPE”, “DJS”, “Jardin”, “PERIPHERIQUE”\}$
- *typeemplacement* : variable constante égale à “Arbre”
- *hauteuren* : hauteur en mètres
- *varietoucultivar* : variété de l'arbre, 453 modalités, contient des valeurs NA
- *espece* : espèce de l'arbre, 537 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 $\in [1; 80]$
- *c_quinsee* : Numéro INSEE du quartier, valeur entière. format 751AAQQ avec AA le numéro d'arrondissement $\in [1; 20]$ et QQ le numéro de quartier $\in [1; 4]$
- *l_qu* : libellé du quartier, e.g. “La Chapelle”

- *c_ar* : numéro d'arrondissement , valeur entière $\in [1; 20]$
- *n_sq_ar* : identifiant séquentiel de l'arrondissement , $\in [750000001; 750000020]$, 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 géométrique du quartier : simple feature de classe POLYGON

0.2 Outils

Les manipulations peuvent être effectuées avec PostGIS, QGIS et MS Excel et R .

1 Distribution statistiques des arbres

Vous disposez de deux jeux de données :

- le contours des quartiers de paris (vectoriel, polygones)
- l'implantation et les variables des arbres de Paris (vectoriel, ponctuel)

1.1 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

1.2 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

1.3 Comparaison des variables

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 des distributions de ces deux variables

Commenter les résultats obtenus

2 Distribution spatiale des arbres

2.1 Carte(s) simple(s)

Réaliser une carte simple des arbres des six genres les plus représentés dans la population

Réaliser une carte simple des arbres en faisant apparaître la variable *domanialite* de façon à représenter le type d'implantation des arbres

Commenter les cartes obtenues

2.2 Cartographie du nombre d'arbres

Vous pouvez au choix :

- réaliser une carte choroplète du nombre d'arbre par quartier (cartographier un nombre est déconseillé, pourquoi ?)
- réaliser une carte de chaleur du nombre d'arbres sur une grille raster
- réaliser une carte avec des symboles proportionnels représentant le nombre d'arbres par quartier

Commenter les cartes obtenues

2.3 Cartographie de la densité d'arbres

Etapes :

- Réaliser une classification des densités en faisant varier la méthode de classification
- Cartographier la classification qui semble la plus adaptée

Commenter la carte obtenue

3 Export des données

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

4 Devoir (trop dur ?)

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