

Data is for Good

Végétalisons la ville de Paris



Stéphanie ROULLAND

SOMMAIRE



Contexte



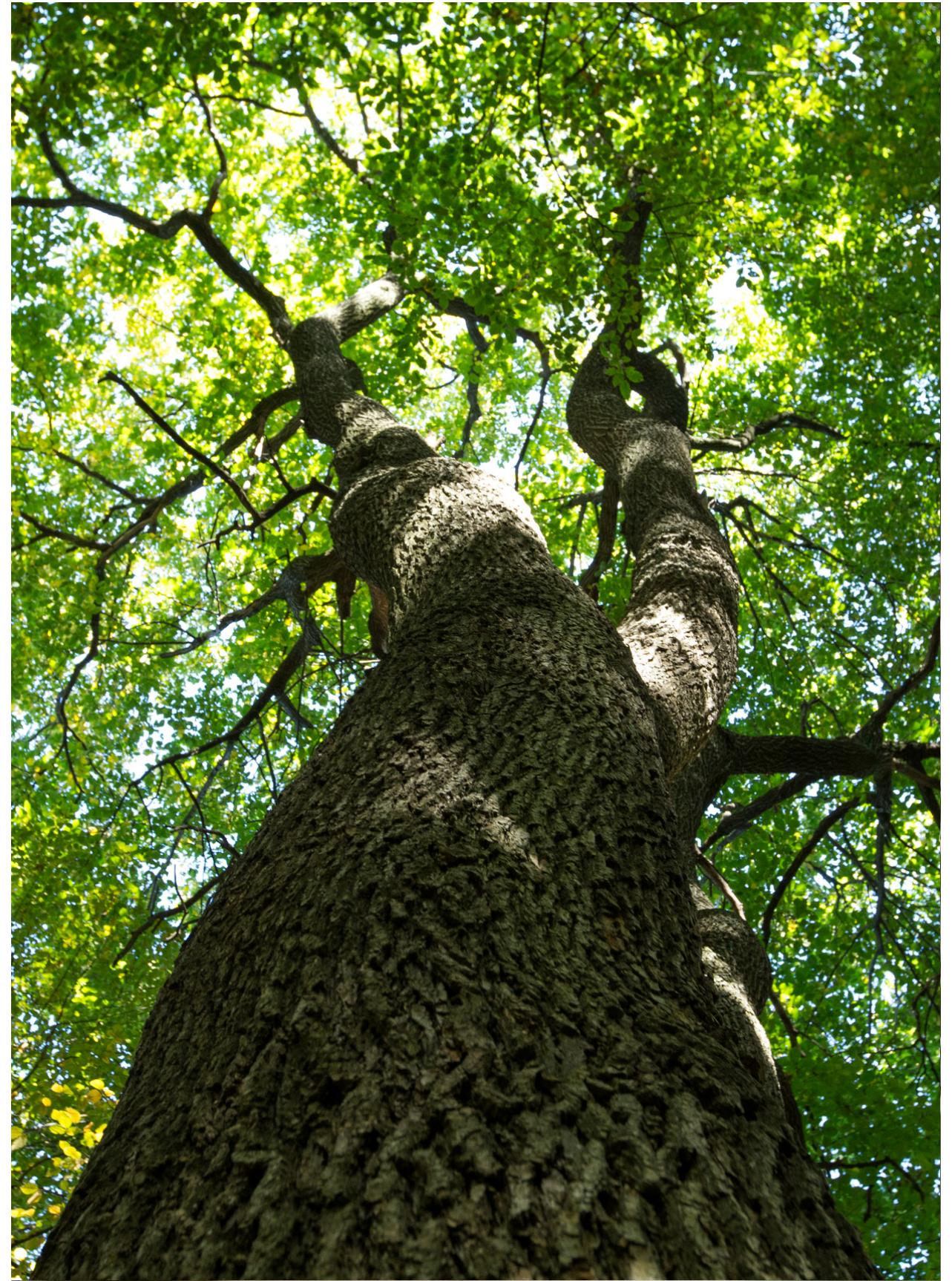
Présentation du jeu de données



Analyse du jeu de données



Synthèse de l'analyse





CONTEXTE

Végétalisons la ville

Contribuer à l'optimisation
des tournées pour
l'entretien des arbres de la
ville de Paris

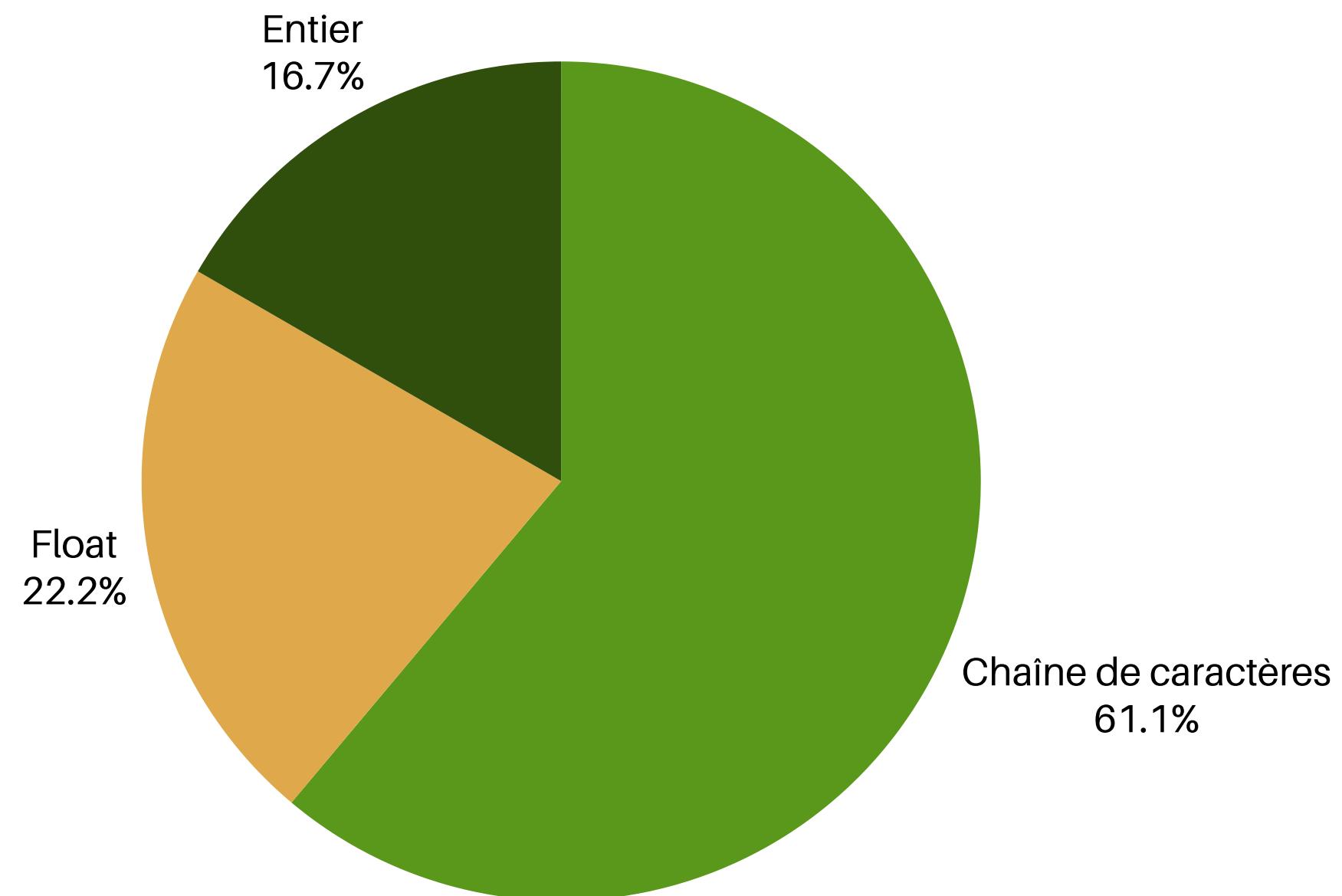




PRÉSENTATION DU JEU DE DONNÉES

18 VARIABLES DANS LE JEU DE DONNÉES

200 137 arbres



Variables quantitatives

7

Variables qualitatives

11

7 VARIABLES QUANTITATIVES

	% de valeurs nulles	Nombre de valeurs distinctes
id	0 %	200 137
numero	100 %	0
circonference_cm	0 %	531
hauteur_m	0 %	143
remarquable	32 %	2
geo_point_2d_a	0 %	200 107
geo_point_2d_b	0 %	200 114

11 VARIABLES QUALITATIVES

	% de valeurs nulles	Nombre de valeurs distinctes		% de valeurs nulles	Nombre de valeurs distinctes
type_emplacement	0 %	1	espece	1 %	539
domanialite	0 %	9	variete	82 %	436
arrondissement	0 %	25	stade_developpement	34 %	4
complement_adresse	85 %	3 795			
lieu	0 %	6 921			
id_emplacement	0%	69 040			
libelle_français	1 %	192			
genre	0 %	175			



ANALYSE DU JEU DE DONNÉES

VARIABLE ID

VARIABLE QUANTITATIVE

Identifiant unique de chaque arbre

→ Aucun doublon dans les id

Colonne non analysée

VARIABLE TYPE_EMPLACEMENT

VARIABLE QUALITATIVE

Type d'emplacement
→ Une valeur unique : “Arbre”

Colonne non analysée

VARIABLE DOMANIALITE

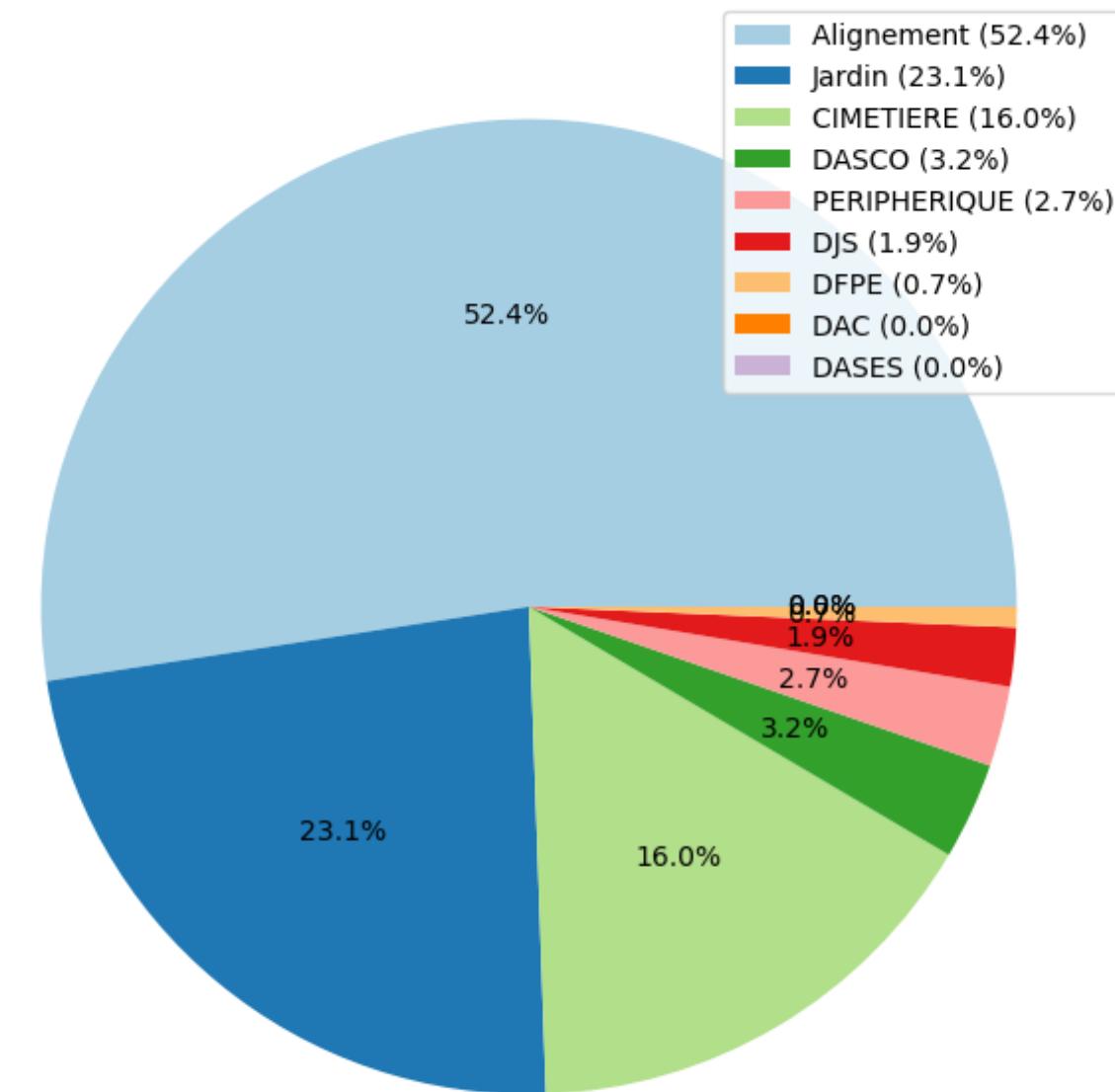
VARIABLE QUALITATIVE

Remplacement des
valeurs nulles par
“Jardin”

Domaine

→ 9 valeurs distinctes et une valeur nulle

Répartition des arbres par domanialité

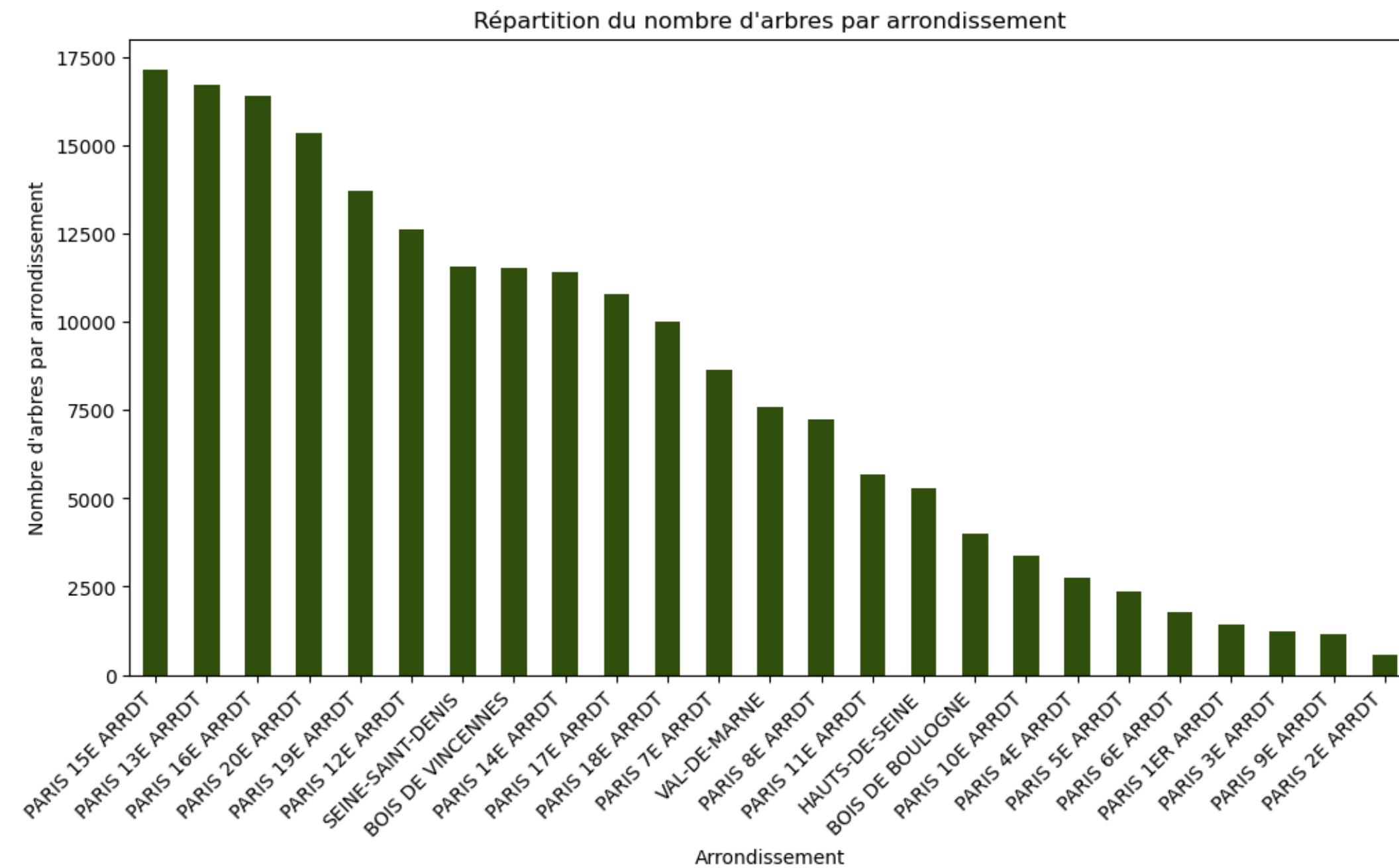


VARIABLE ARRONDISSEMENT

VARIABLE QUALITATIVE

Arrondissement de Paris

→ 25 valeurs distinctes



VARIABLE

COMPLEMENT_ADDRESS

VARIABLE QUALITATIVE

Complément d'adresse
→ 85% de valeurs nulles

Colonne non analysée

VARIABLE NUMERO

VARIABLE QUANTITATIVE

Numéro
→ 100% de valeurs nulles

Colonne non analysée

VARIABLE LIEU

VARIABLE QUALITATIVE

Adresse

→ 6 921 valeurs distinctes

Colonne non analysée

VARIABLE

ID_EMPLACEMENT

VARIABLE QUALITATIVE

Identifiant de l'emplacement
→ 69 040 valeurs distinctes

Colonne non analysée

VARIABLE LIBELLE_FRANCAIS

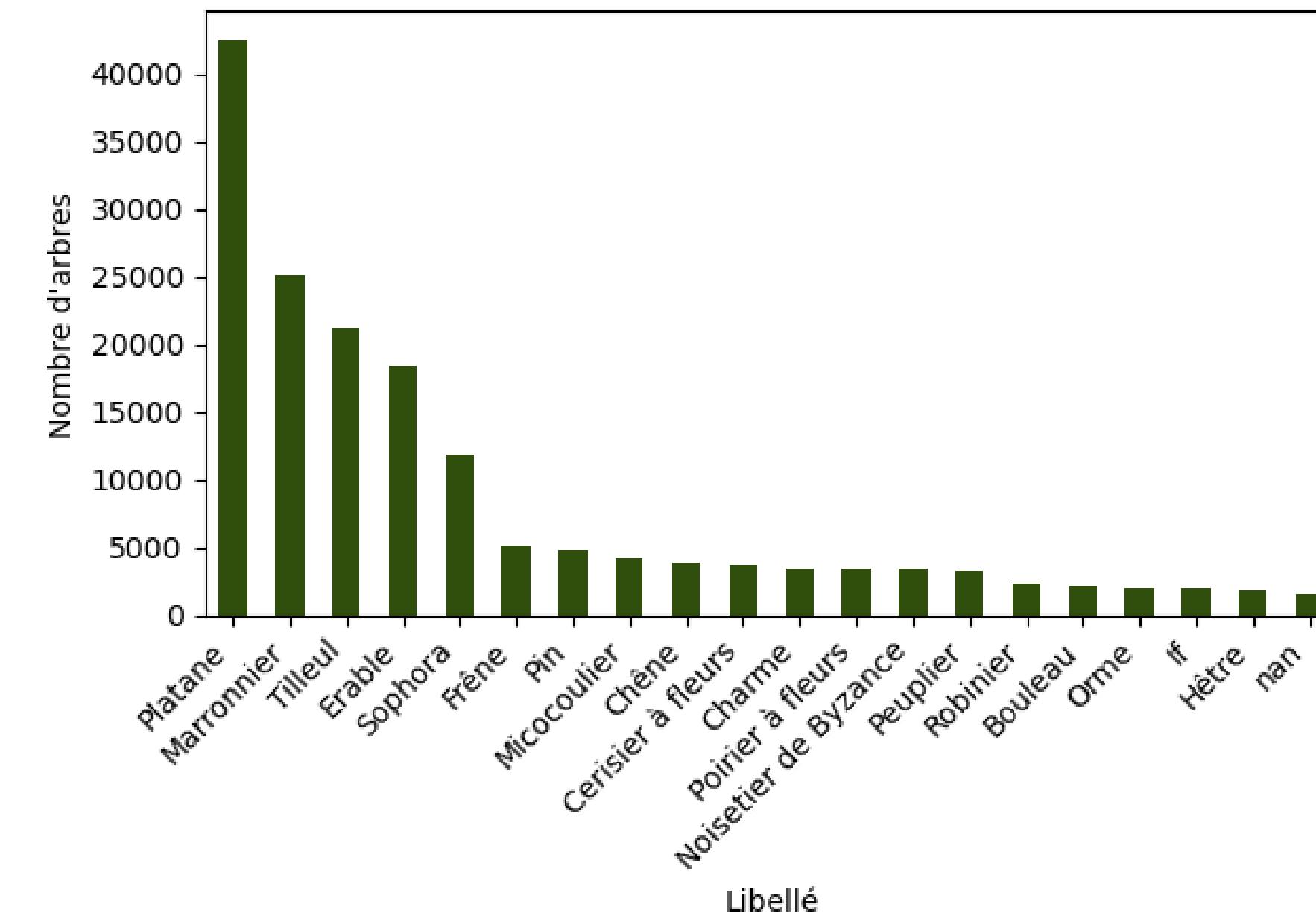
VARIABLE QUALITATIVE

Création d'une valeur
“Inconnu” pour
remplacer les valeurs
nulles

Nom français de l'arbre

→ 192 valeurs distinctes

TOP 20 des arbres par libellé

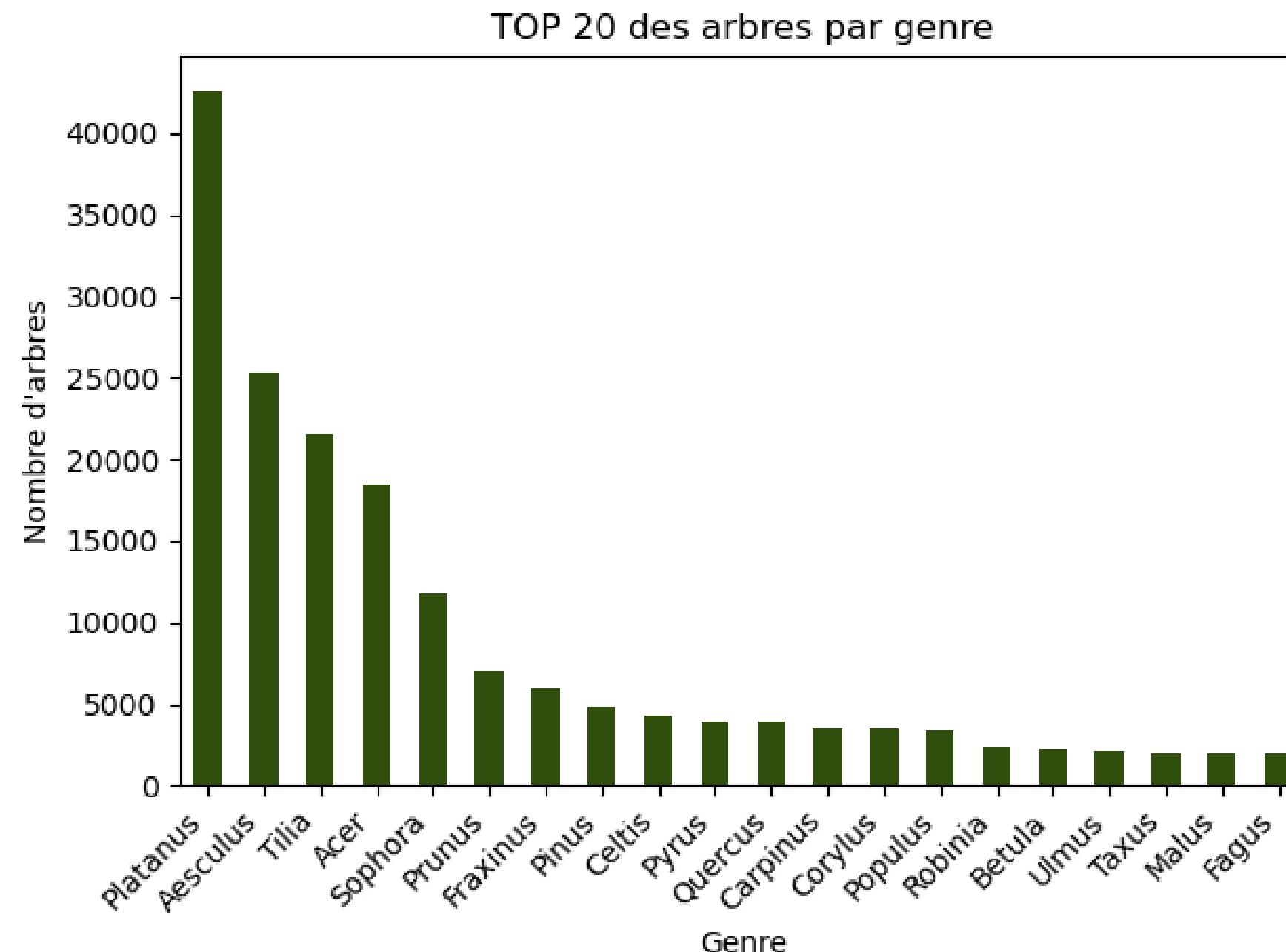


VARIABLE GENRE

VARIABLE QUALITATIVE

Genre de l'arbre

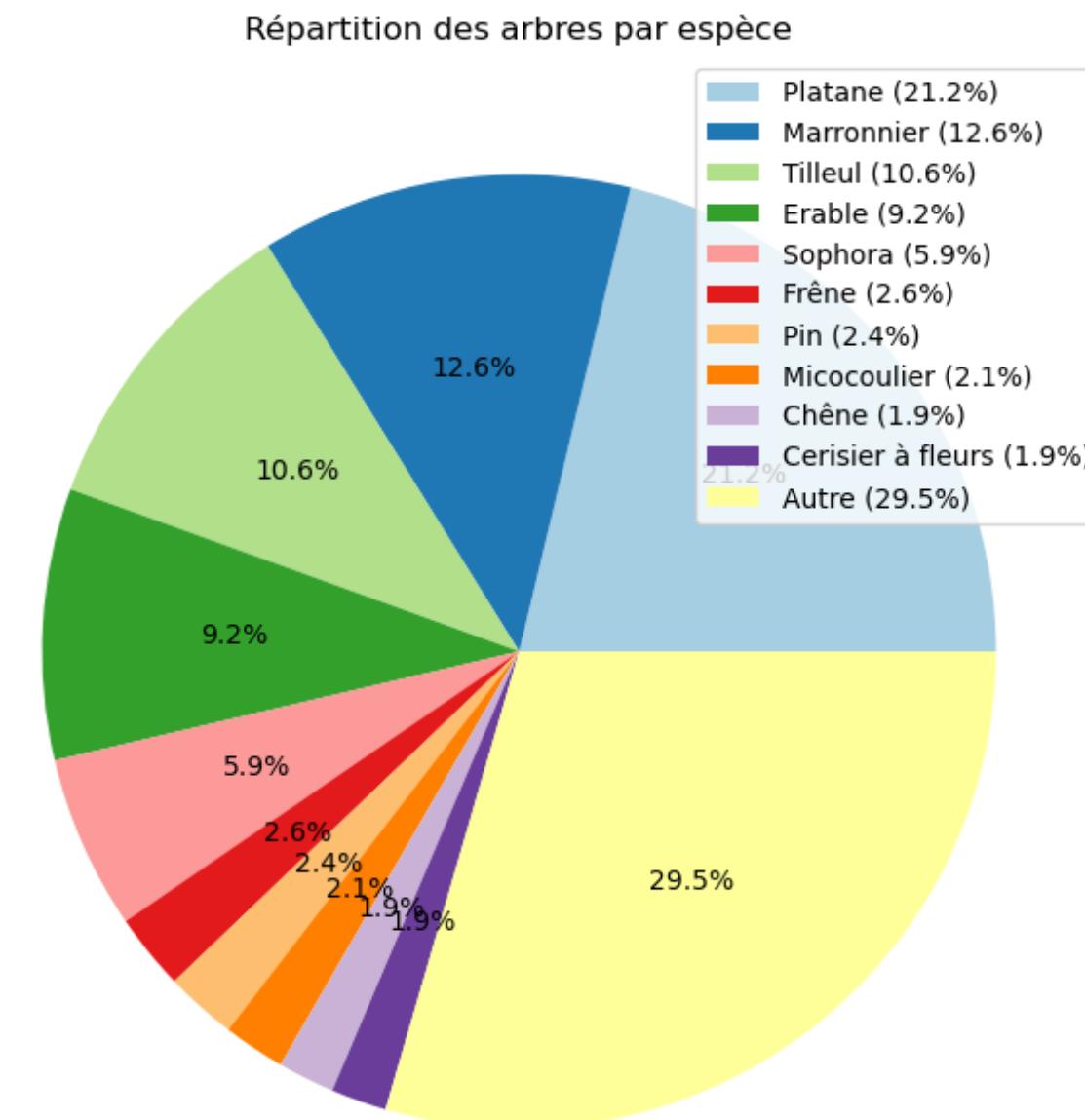
→ 175 valeurs distinctes



VARIABLE ESPECE

VARIABLE QUALITATIVE

Espèce de l'arbre
→ 539 valeurs distinctes

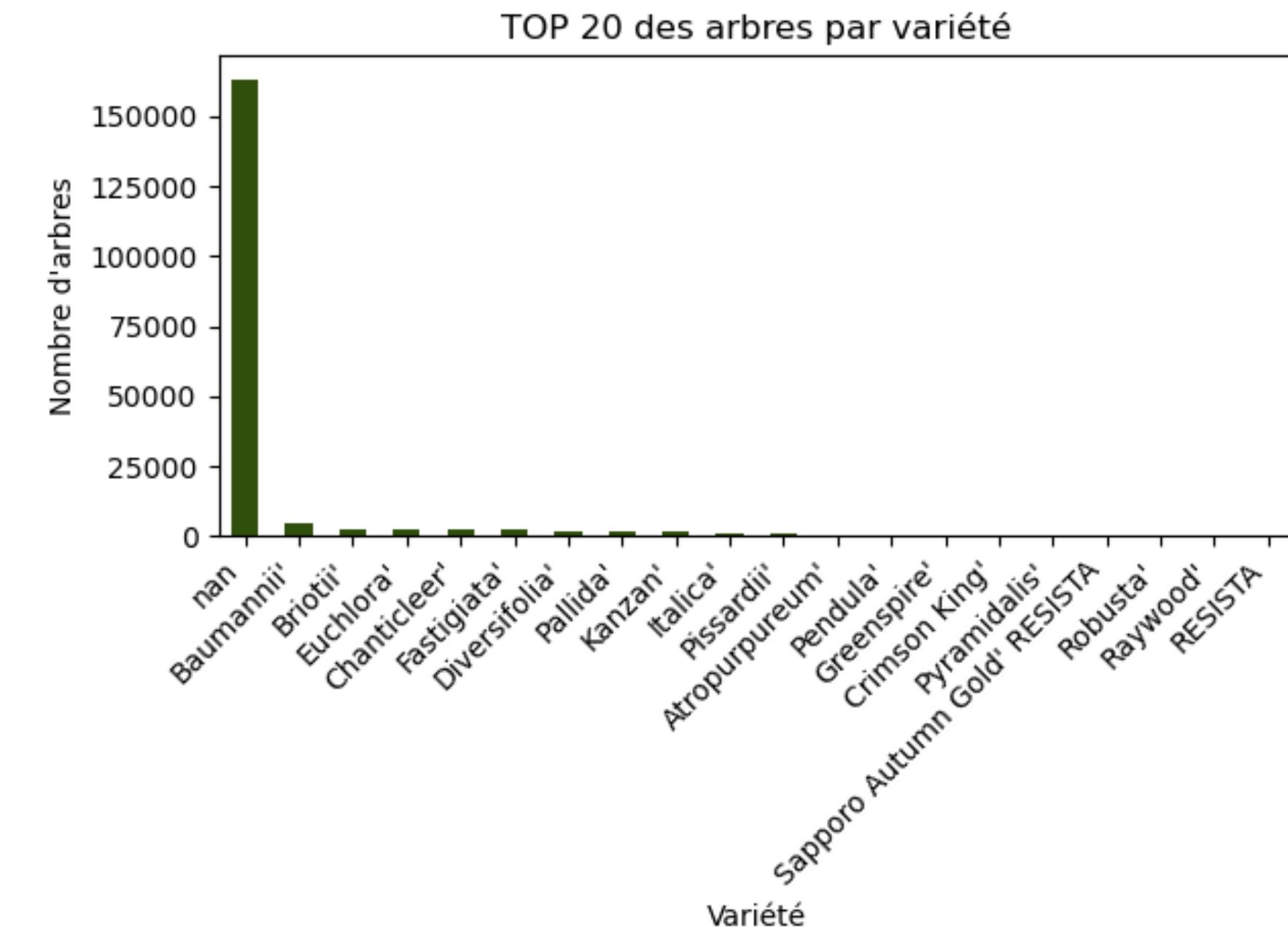


VARIABLE VARIETE

VARIABLE QUALITATIVE

Variété de l'arbre

→ 436 valeurs distinctes



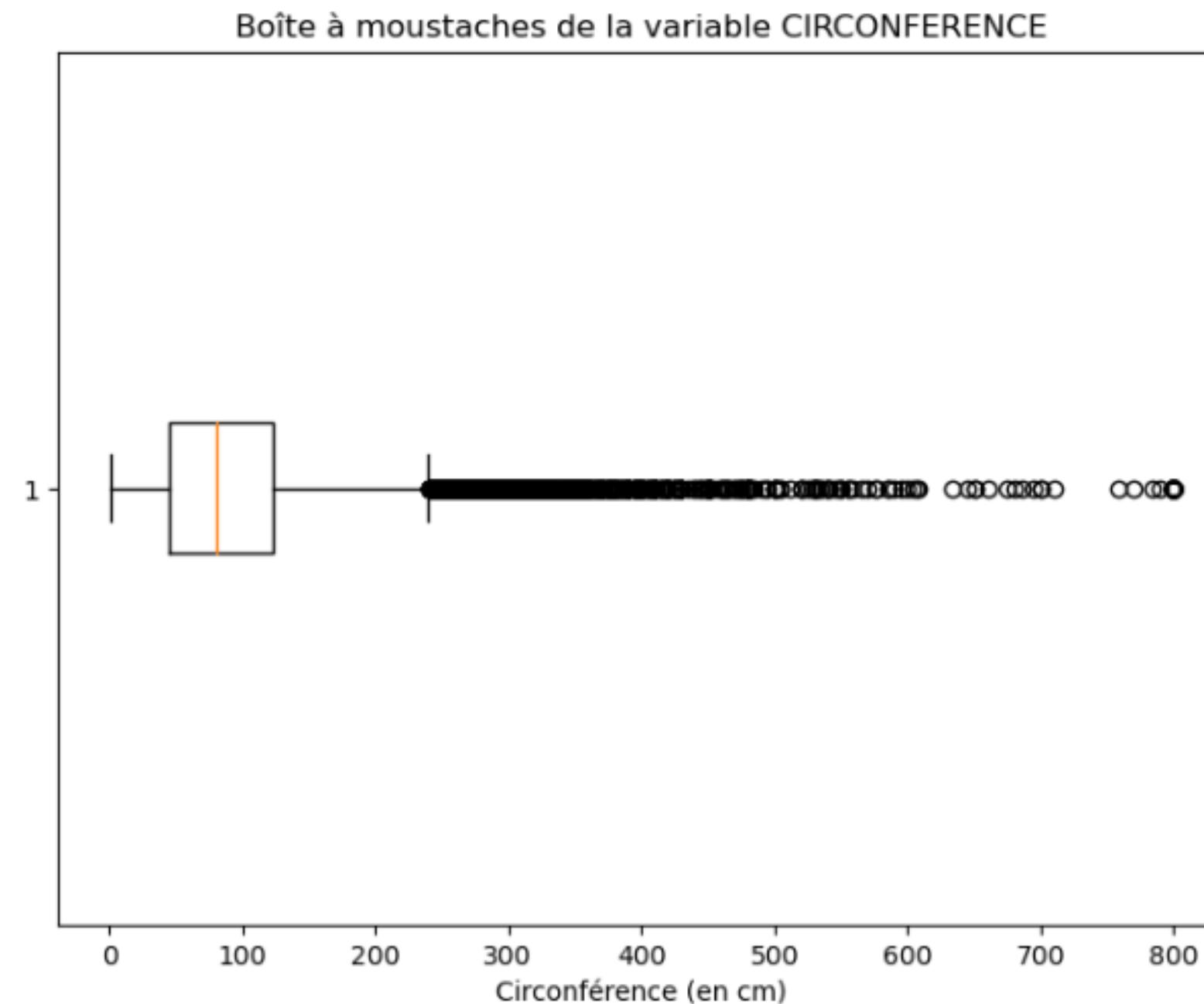
VARIABLE CIRCONFERENCE_CM

VARIABLE QUANTITATIVE

Remplacement des
valeurs supérieures
à 800 cm

Circonférence en cm

→ Valeurs comprises entre 0 et 250 255 cm



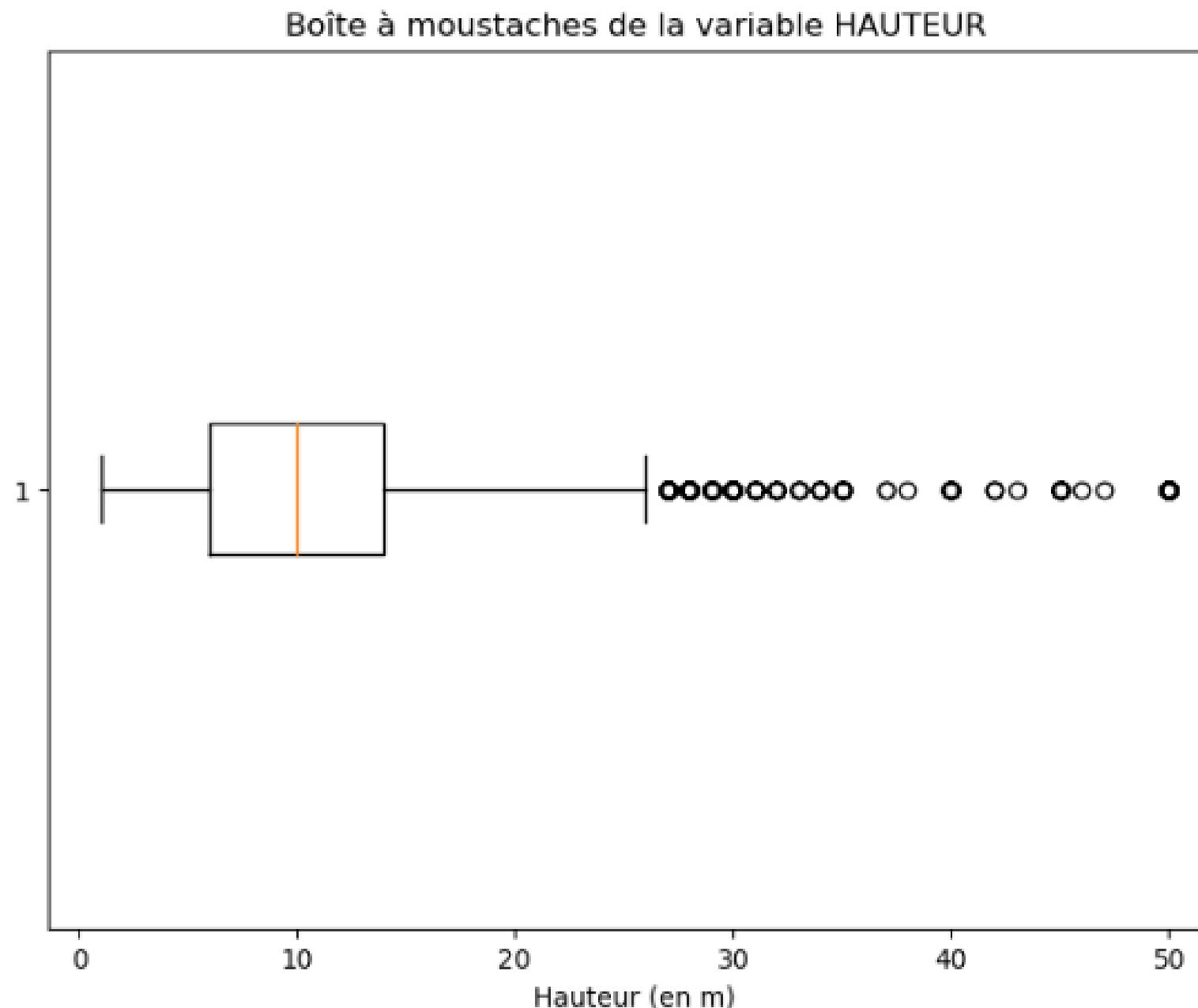
VARIABLE HAUTEUR_M

VARIABLE QUANTITATIVE

Remplacement des
valeurs supérieures
à 50 m

Hauteur en m

→ Valeurs comprises entre 0 et 881 818 m



VARIABLE STADE_DEVELOPPEMENT

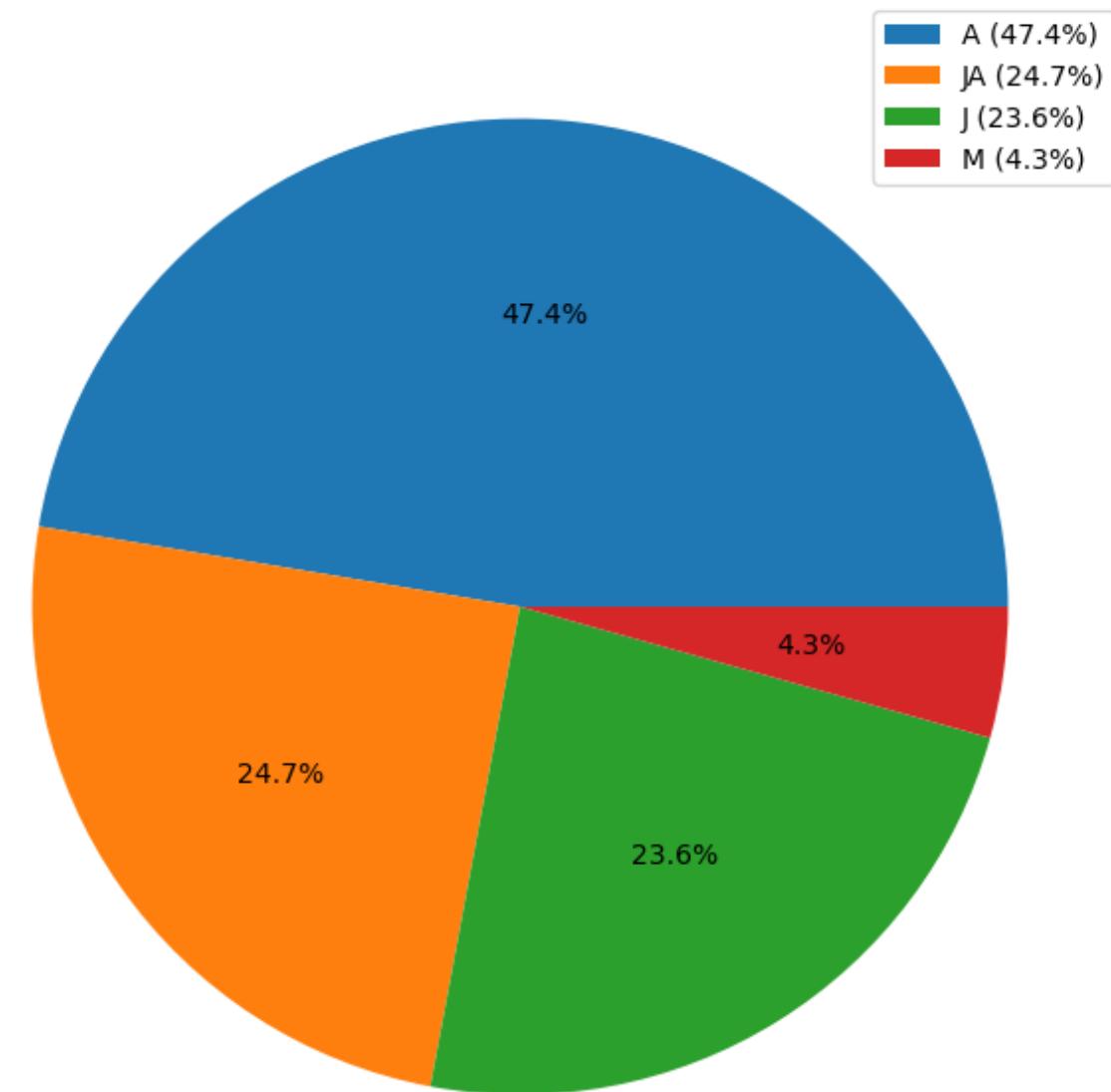
VARIABLE QUALITATIVE

Remplacement des
valeurs manquantes
en utilisant un
algorithme KNN

Stade de développement

→ 4 valeurs distinctes : A / JA / J / M

Répartition des arbres par stade de développement



VARIABLE REMARQUABLE

VARIABLE QUANTITATIVE

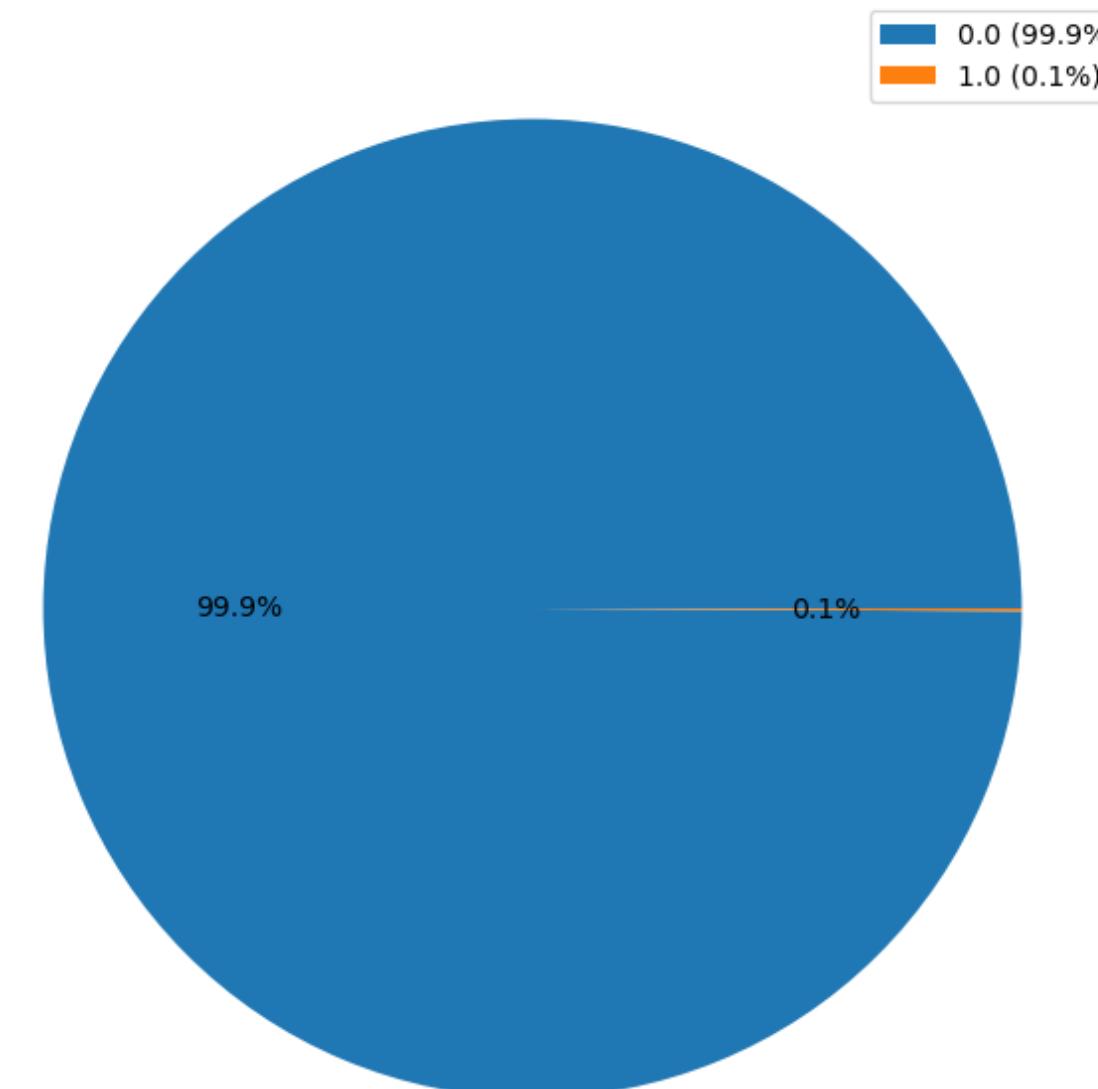
188 arbres
remarquables à Paris
d'après le site
opendata

Remplacement par 0
des valeurs nulles

Arbre remarquable ou non

→ 2 valeurs distinctes : 0 (non) et 1 (oui)

Répartition des arbres remarquables et non remarquables

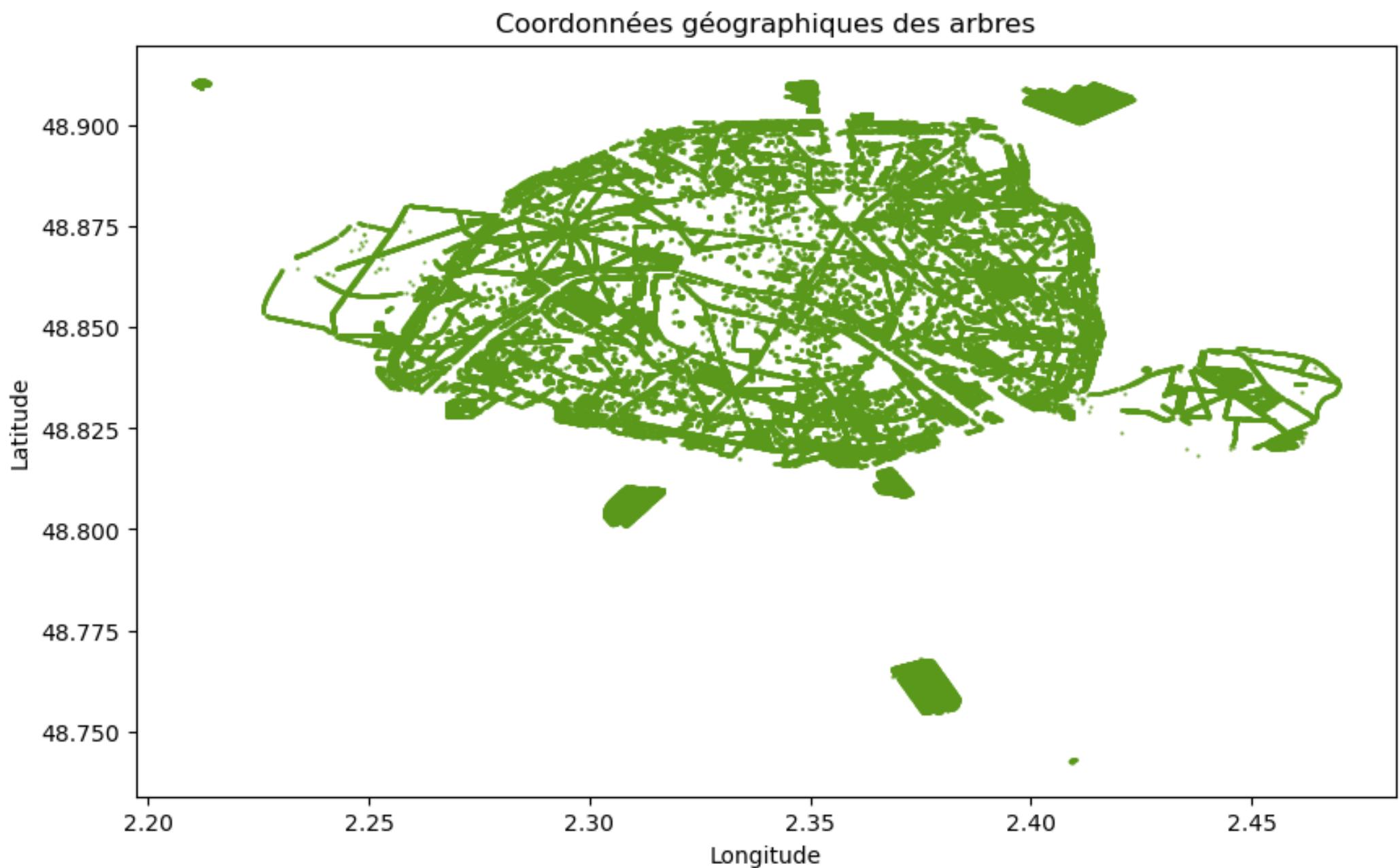


VARIABLES

GEO_POINT_2D_A ET GEO_POINT_2D_B

VARIABLE QUANTITATIVE

Latitude et longitude





SYNTHÈSE DE L'ANALYSE

SEGMENTATIONS GÉOGRAPHIQUES

1/6

Utilisation de la latitude et la longitude pour segmenter les arbres en différentes zones géographiques.

**REGROUPEMENT DES ARBRES
PROCHES POUR MINIMISER LES
DÉPLACEMENTS**



OPTIMISATION DES ITINÉRAIRES

2/6



Segmentation des arbres par zones et identification des arbres prioritaires

+

Utilisation algorithme du voyageur du commerce pour déterminer l'itinéraire le plus court et le plus efficace



**RÉDUCTION DU TEMPS DE
DÉPLACEMENT ET DE LA
CONSOMMATION DE CARBURANT**



ANALYSE PAR ARRONDISSEMENT

3/6



Identification des arrondissements ayant le plus grand nombre d'arbres ou ceux nécessitant le plus d'entretien



**ALLOCATION DES RESSOURCES
D'ENTRETIEN SELON LES BESOINS DE
CHAQUE ARRONDISSEMENT**



PRISE EN COMPTE DES CONTRAINTES EXTERNES

4/6



Intégration de données telles que la météo,
les événements spéciaux ou les travaux en
cours dans la ville



**AJUSTEMENT ET OPTIMISATION
DES TOURNÉES D'ENTRETIEN**



PLANIFICATION SAISONNIÈRE

5/6

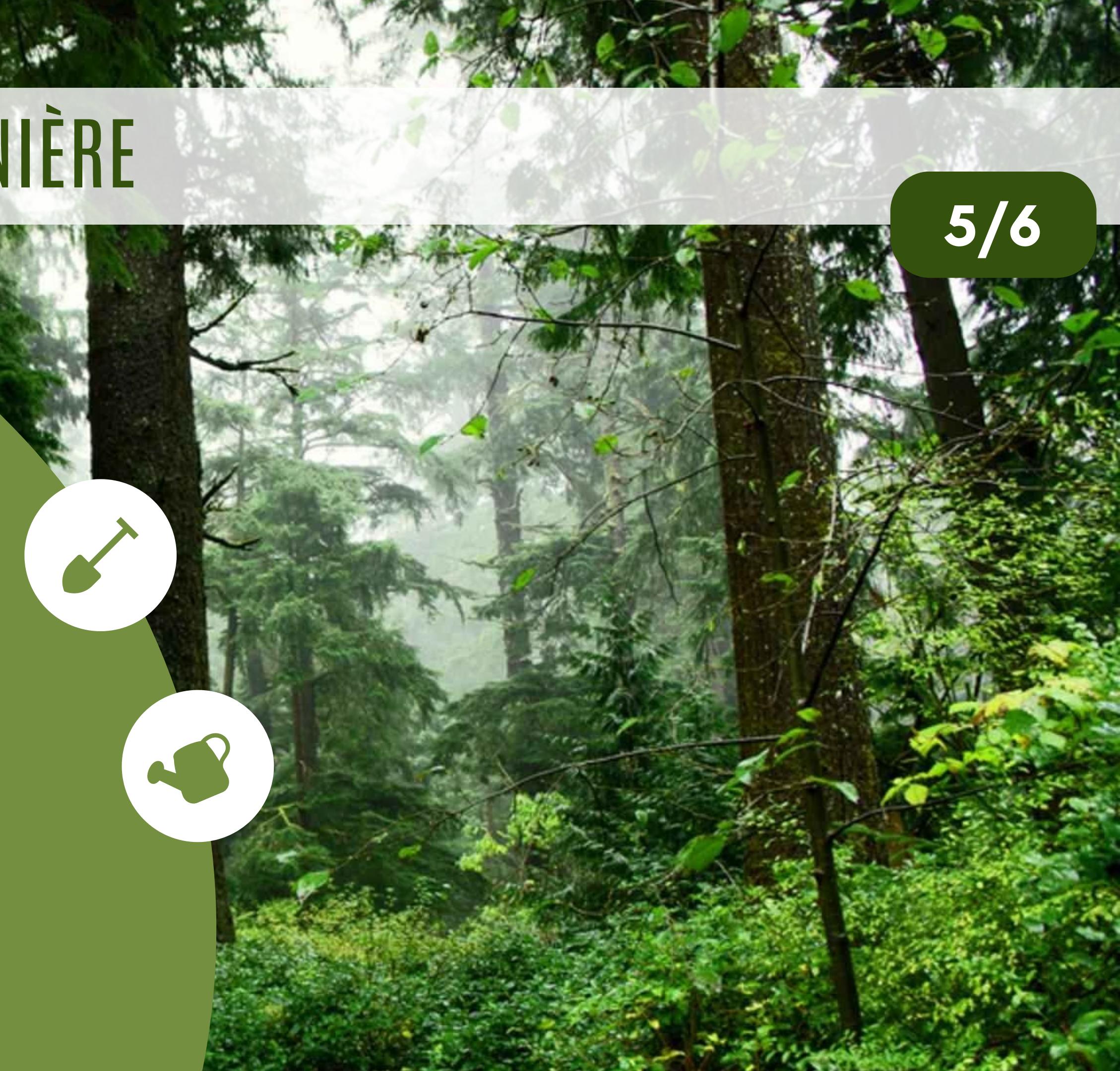


Vérification des soins spécifiques
à certaines espèces

=> Taille en fonction de la saison



**CRÉATION DE TOURNÉES
SAISONNIÈRES OPTIMISÉES**



PRIORISATION DES ESPÈCES SENSIBLES OU EN DANGER

6/6

Priorisation des espèces plus sensibles aux conditions environnementales urbaines ou classées comme en danger

**CONTRIBUTION DE LA
CONSERVATION ET DE LEUR SANTÉ
À LONG TERME**



Merci



Des questions ?

Stéphanie ROULLAND