Data is for Good - aidons Paris à devenir une smart-city !

Ville de Paris - programme "Végétalisons la ville"

Sommaire

VILLE DE PARIS

PARTIE 1 - PRÉSENTATION GÉNÉRALE DU JEU DE DONNÉES

VILLE DE PARIS

1.1. Caractéristiques générales

	id	type_emplacement	domanialite	arrondissement	complement_addresse	numero	lieu	id_emplacement	libelle_francais	genre	espece	variete	circonference_cm	hauteur_m	stade_developpement	remarquable	geo_point_2d_a	geo_point_2d_b
0	99874	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	19	Marronnier	Aesculus	hippocastanum	NaN	20	5	NaN	0.0	48.857620	2.320962
1	99875	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	20	If	Taxus	baccata	NaN	65	8	A	NaN	48.857656	2.321031
2	99876	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	21	If	Taxus	baccata	NaN	90	10	A	NaN	48.857705	2.321061
3	99877	Arbre	Jardin	PARIS 7E ARRDT	NaN	NaN	MAIRIE DU 7E 116 RUE DE GRENELLE PARIS 7E	22	Erable	Acer	negundo	NaN	60	8	A	NaN	48.857722	2.321006
4	99878	Arbre	Jardin	PARIS 17E ARRDT	NaN	NaN	PARC CLICHY-BATIGNOLLES-MARTIN LUTHER KING	000G0037	Arbre à miel	Tetradium	daniellii	NaN	38	0	NaN	NaN	48.890435	2.315289
5	99879	Arbre	Jardin	PARIS 17E ARRDT	NaN	NaN	PARC CLICHY-BATIGNOLLES-MARTIN LUTHER KING	000G0036	Arbre à miel	Tetradium	daniellii	NaN	38	0	NaN	NaN	48.890470	2.315228
6	99880	Arbre	Jardin	PARIS 17E ARRDT	NaN	NaN	PARC CLICHY-BATIGNOLLES-MARTIN LUTHER KING	000G0035	Arbre à miel	Tetradium	daniellii	NaN	37	0	NaN	NaN	48.890504	2.315168
7	99881	Arbre	Jardin	PARIS 16E ARRDT	NaN	NaN	SQUARE ALEXANDRE ET RENE PARODI / 1 PLACE DE L...	35	Platane	Platanus	x hispanica	NaN	260	17	NaN	NaN	48.876722	2.280222
8	99882	Arbre	Jardin	PARIS 16E ARRDT	NaN	NaN	JARDIN DE L AVENUE FOCH / 10 AVENUE FOCH	802008	Sophora	Sophora	japonica	NaN	145	14	A	0.0	48.871990	2.275814
9	99883	Arbre	Jardin	PARIS 16E ARRDT	NaN	NaN	JARDIN DE L AVENUE FOCH / 10 AVENUE FOCH	802009	Sophora	Sophora	japonica	NaN	135	10	A	0.0	48.872046	2.275752
10	99884	Arbre	Jardin	PARIS 16E ARRDT	NaN	NaN	JARDIN DE L AVENUE FOCH / 10 AVENUE FOCH	802007	Prunus n. sp.	Prunus	n. sp.	NaN	15	3	J	0.0	48.871948	2.275867
11	99885	Arbre	Jardin	PARIS 16E ARRDT	NaN	NaN	JARDIN DU RANELAGH	20001	Hêtre	Fagus	sylvatica	Atropunicea'	30	0	NaN	NaN	48.858222	2.269287
12	99887	Arbre	Jardin	PARIS 16E ARRDT	NaN	NaN	JARDIN DU RANELAGH	20003	Micocoulier	Celtis	occidentalis	NaN	205	0	NaN	NaN	48.858212	2.268794
13	99888	Arbre	Jardin	PARIS 16E ARRDT	NaN	NaN	JARDIN DU RANELAGH	20004	Tilleul	Tilia	tomentosa	NaN	155	0	NaN	NaN	48.858139	2.268634
14	99889	Arbre	Jardin	PARIS 16E ARRDT	NaN	NaN	JARDIN DU RANELAGH	20007	Chêne	Quercus	robur	NaN	25	0	NaN	NaN	48.858227	2.268489

↕ 200 137 lignes

↔ 18 colonnes

VILLE DE PARIS

1.1. Caractéristiques générales

	Description colonne	Groupe colonne
id	identifiant de chaque arbre sur la forme d'un ...	identification
type_emplacement	type de végétal	biométrie
domanialite	type d'espace public où se situe l'arbre	localisation
arrondissement	arrondissement de Paris	localisation
complement_addresse	complément d'adresse	localisation
numero	numéro de l'adresse	localisation
lieu	adresse de l'arbre	localisation
id_emplacement	identifiant de l'emplacement	identification
libelle_francais	nom de l'espèce en langage commun	biométrie
genre	genre de l'arbre	biométrie
espece	nom de l'espèce en latin	biométrie
variete	variété de l'espèce	biométrie
circonference_cm	circonférence en centimètres	biométrie
hauteur_m	hauteur en mètres	biométrie
stade_developpement	stade de développement, reflète l'âge	biométrie
remarquable	précise si oui ou non l'arbre est remarquable	biométrie
geo_point_2d_a	latitude	localisation
geo_point_2d_b	longitude	localisation

Différents types d'information :

- Biométrie
- Localisation
- Identification

VILLE DE PARIS

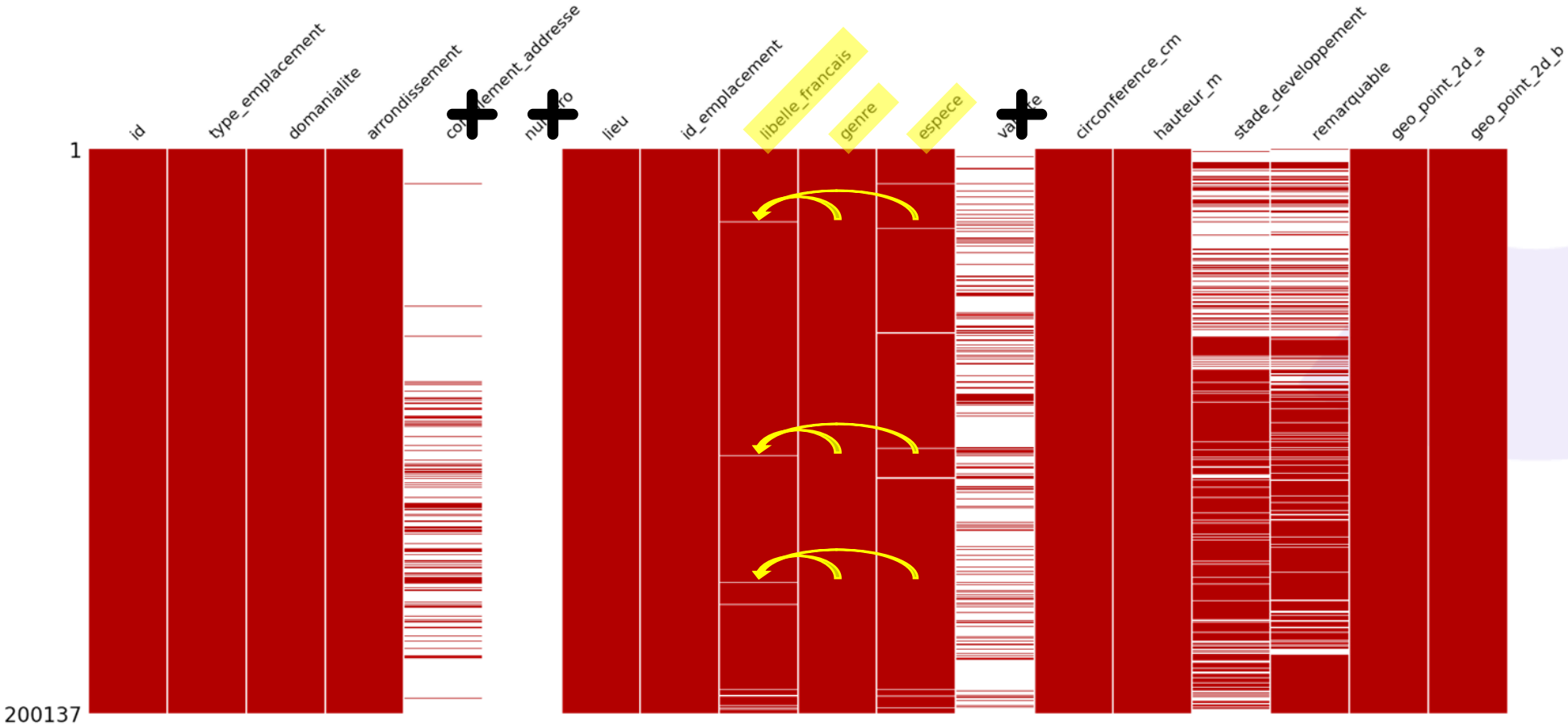
1.2. Des valeurs manquantes ?

	Description colonne	Groupe colonne	Nb NaN	% NaN
id	identifiant de chaque arbre sur la forme d'un ...	identification	0	0 %
type_emplacement	type de végétal	biométrie	0	0 %
domanialite	type d'espace public où se situe l'arbre	localisation	1	0 %
arrondissement	arrondissement de Paris	localisation	0	0 %
complement_addresse	complément d'adresse	localisation	169235	85 %
numero	numéro de l'adresse	localisation	200137	100 %
lieu	adresse de l'arbre	localisation	0	0 %
id_emplacement	identifiant de l'emplacement	identification	0	0 %
libelle_francais	nom de l'espèce en langage commun	biométrie	1497	1 %
genre	genre de l'arbre	biométrie	16	0 %
espece	nom de l'espèce en latin	biométrie	1752	1 %
variete	variété de l'espèce	biométrie	163360	82 %
circonference_cm	circonférence en centimètres	biométrie	0	0 %
hauteur_m	hauteur en mètres	biométrie	0	0 %
stade_developpement	stade de développement, reflète l'âge	biométrie	67205	34 %
remarquable	précise si oui ou non l'arbre est remarquable	biométrie	63098	32 %
geo_point_2d_a	latitude	localisation	0	0 %
geo_point_2d_b	longitude	localisation	0	0 %

Les prendre en compte ?

VILLE DE PARIS

1.2. Des valeurs manquantes ?



1.3. Mesures statistiques sur les données brutes

Colonnes numériques :

	id	numero	circonference_cm	hauteur_m	remarquable	geo_point_2d_a	geo_point_2d_b
count	2.001370e+05	0.0	200137.000000	200137.000000	137039.000000	200137.000000	200137.000000
mean	3.872027e+05	NaN	83.380479	13.110509	0.001343	48.854491	2.348208
std	5.456032e+05	NaN	673.190213	1971.217387	0.036618	0.030234	0.051220
min	9.987400e+04	NaN	0.000000	0.000000	0.000000	48.742290	2.210241
25%	1.559270e+05	NaN	30.000000	5.000000	0.000000	48.835021	2.307530
50%	2.210780e+05	NaN	70.000000	8.000000	0.000000	48.854162	2.351095
75%	2.741020e+05	NaN	115.000000	12.000000	0.000000	48.876447	2.386838
max	2.024745e+06	NaN	250255.000000	881818.000000	1.000000	48.911485	2.469759

Moyenne > Médiane

Valeurs aberrantes

VILLE DE PARIS

1.3. Mesures statistiques sur les données brutes

Colonnes non numériques :

	type_emplacement	domanialite	arrondissement	complement_addresse	lieu	id_emplacement	libelle_francais	genre	espece	variete	stade_developpement
count	200137	200136	200137	30902	200137	200137	198640	200121	198385	36777	132932
unique	1	9	25	3795	6921	69040	192	175	539	436	4
top	Arbre	Alignement	PARIS 15E ARRDT	SN°	PARC FLORAL DE PARIS / ROUTE DE LA PYRAMIDE	101001	Platane	Platanus	x hispanica	Baumannii'	A
freq	200137	104949	17151	557	2995	1324	42508	42591	36409	4538	64438

Ne conserver que les plus représentatives ?

VILLE DE PARIS

PARTIE 2 - DÉMARCHE MÉTHODOLOGIQUE D'ANALYSE DE DONNÉES

2.1. Exploration - variables d'identification

Colonne id : des arbres en doublon ?

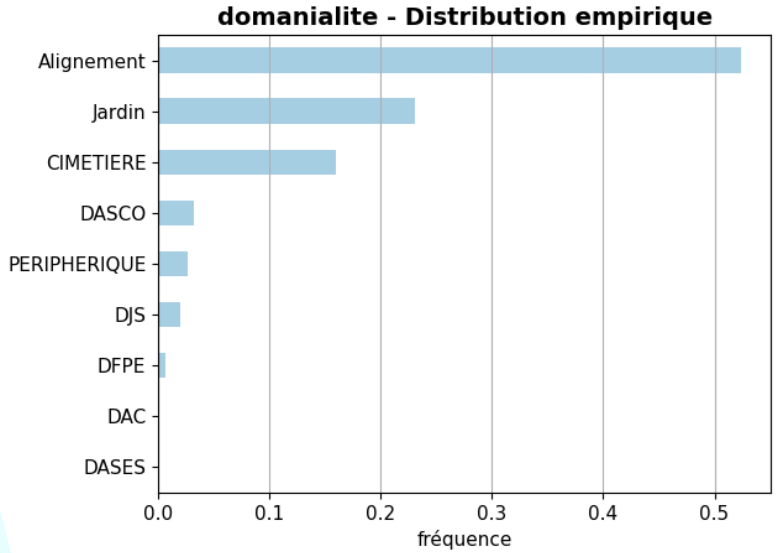
Colonne id_emplacement : des problèmes de formats ?

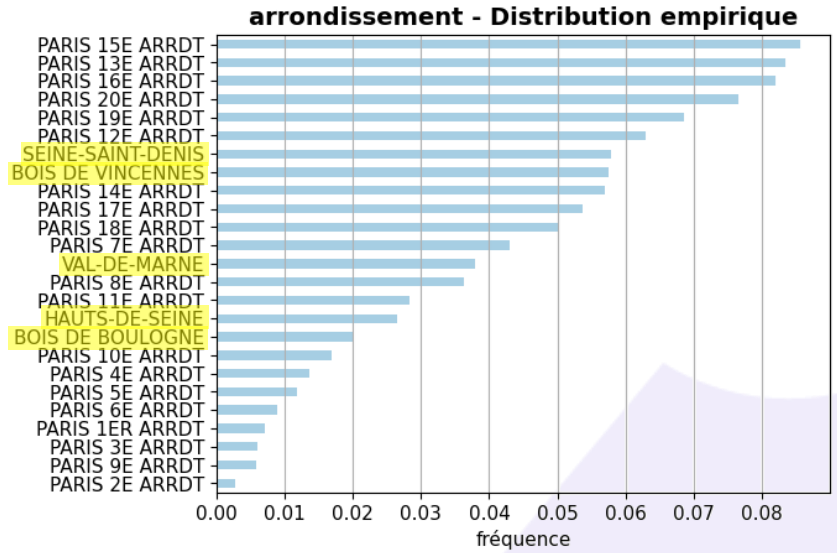
	Nb caracteres id_emp	Nb occurrence
0	6	92166
1	12	31928
2	7	27354
3	8	18300
4	2	10920
5	5	7187
6	3	5241
7	9	3802
8	1	2894
9	4	316
10	10	21
11	15	8

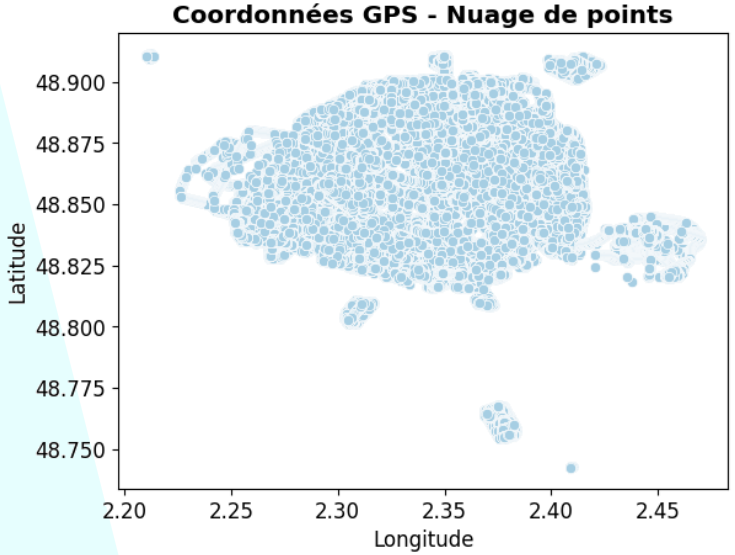
→ ✗ Ne sera pas considérée

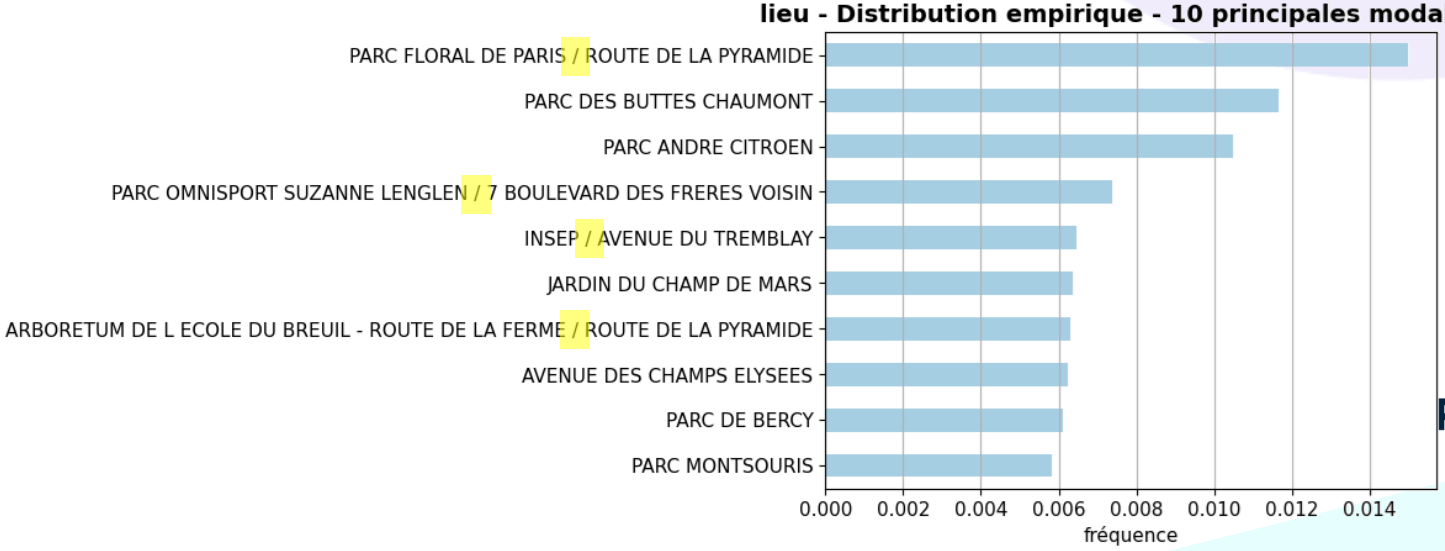
VILLE DE PARIS

2.1. Exploration - variables de localisation



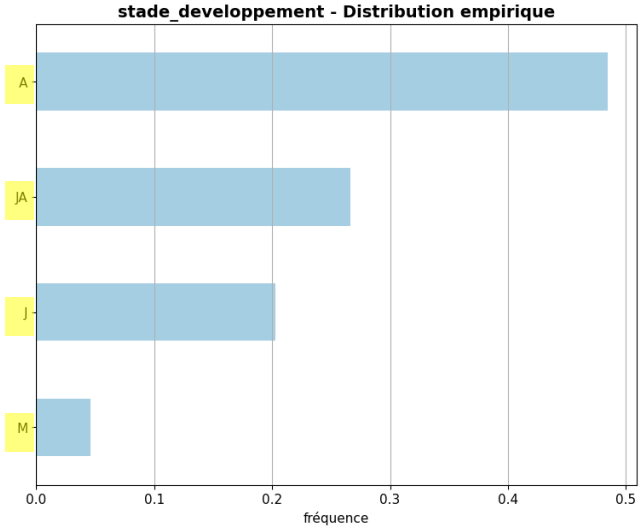




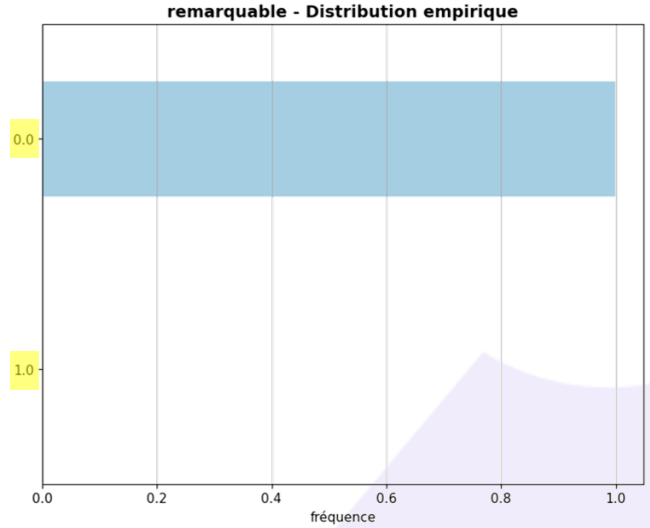


VILLE DE PARIS

2.1. Exploration - variables biométriques



renommer



renommer

VILLE DE PARIS

2.1. Exploration - variables biométriques

libelle_francais

genre

espece

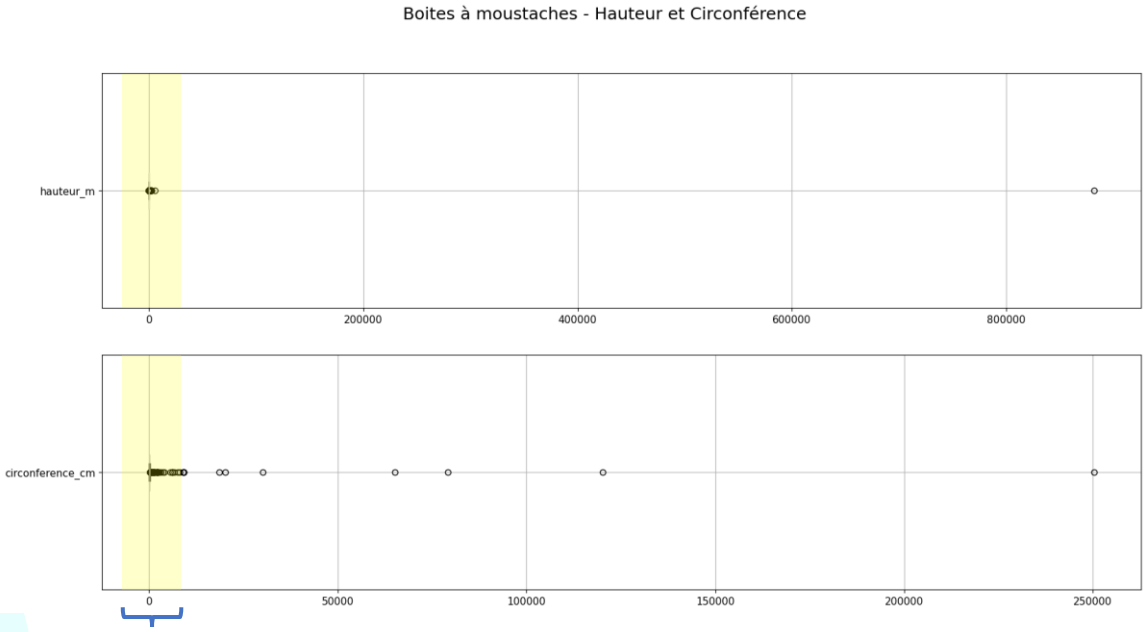
Quelle représentativité minimale pour une bonne observation ?

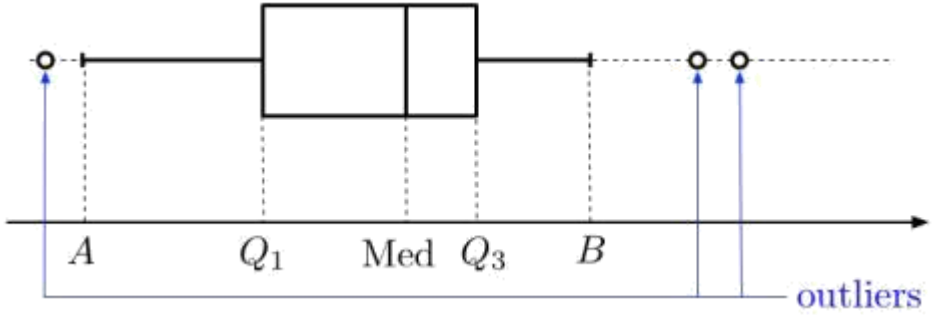
	libelle_francais	lib %	lib % cumul	genre	gen %	gen % cumul	espece	esp %	esp % cumul
0	--TOP--	---	---	--TOP--	---	---	--TOP--	---	---
1	Platane	21.4 %	21 %	Platanus	21.28 %	21 %	x hispanica	18.35 %	18 %
2	Marronnier	12.69 %	34 %	Aesculus	12.66 %	34 %	hippocastanum	10.1 %	28 %
3	Tilleul	10.73 %	45 %	Tilia	10.77 %	45 %	japonica	5.96 %	34 %
4	Erable	9.26 %	54 %	Acer	9.23 %	54 %	n. sp.	4.57 %	39 %
5	Sophora	5.94 %	60 %	Sophora	5.91 %	60 %	tomentosa	4.52 %	43 %
6	Frêne	2.6 %	63 %	Prunus	3.52 %	63 %	pseudoplatanus	3.75 %	47 %
7	Pin	2.44 %	65 %	Fraxinus	3.01 %	66 %	platanoides	3.17 %	50 %
8	Micocoulier	2.11 %	67 %	Pinus	2.43 %	69 %	nigra	2.49 %	53 %
9	Chêne	1.95 %	69 %	Celtis	2.14 %	71 %	x europaea	2.37 %	55 %
10	Cerisier à fleurs	1.9 %	71 %	Pyrus	1.96 %	73 %	x carnea	2.23 %	58 %
11	Charme	1.75 %	73 %	Quercus	1.94 %	75 %	australis	2.08 %	60 %
12	Poirier à fleurs	1.72 %	74 %	Carpinus	1.75 %	77 %	cordata	1.94 %	62 %
13	Noisetier de Byzance	1.7 %	76 %	Corylus	1.73 %	78 %	excelsior	1.92 %	63 %
14	Peuplier	1.67 %	78 %	Populus	1.67 %	80 %	occidentalis	1.75 %	65 %
15	Robinier	1.16 %	79 %	Robinia	1.16 %	81 %	betulus	1.75 %	67 %
16	Bouleau	1.13 %	80 %	Betula	1.13 %	82 %	colurna	1.7 %	69 %
17	Orme	1.04 %	81 %	Ulmus	1.07 %	83 %	calleryana	1.58 %	70 %
18	--BOTTOM--	---	---	--BOTTOM--	---	---	--BOTTOM--	---	---
19	Jujubier	0.0 %	100 %	Euscaphis	0.0 %	100 %	lusitanica subsp.azorica	0.0 %	100 %
20	Maackie	0.0 %	100 %	Phyllanthus	0.0 %	100 %	oliveri	0.0 %	100 %
21	Garrya	0.0 %	100 %	Washingtonia	0.0 %	100 %	delavayi subsp. potaninii	0.0 %	100 %

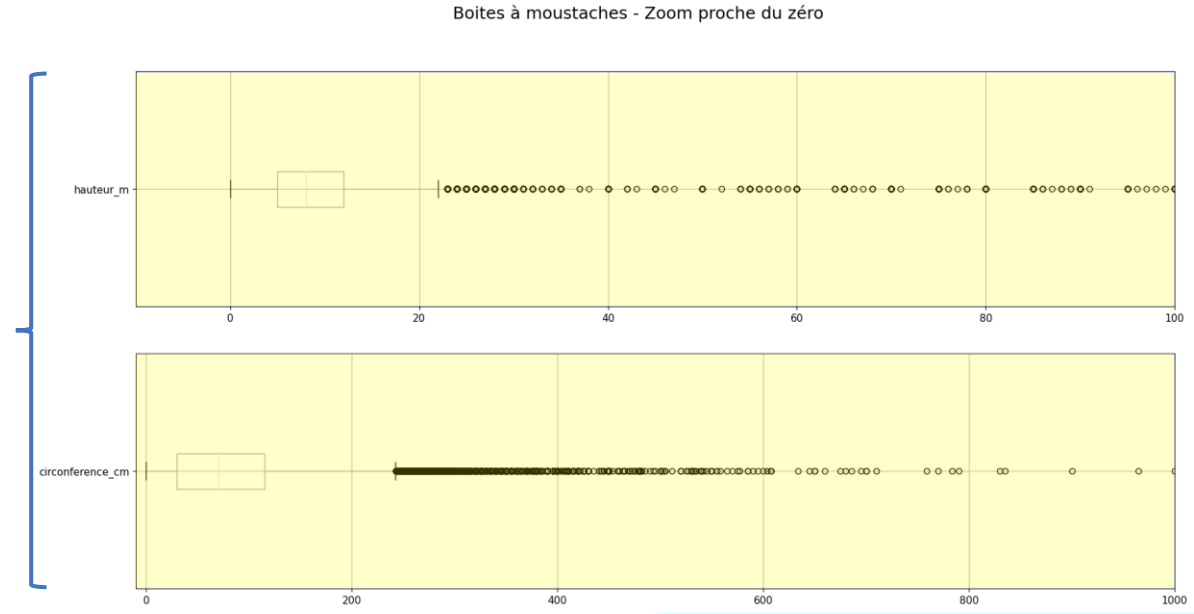
VILLE DE PARIS

2.1. Exploration - variables biométriques

Hauteur_m et circonference_cm

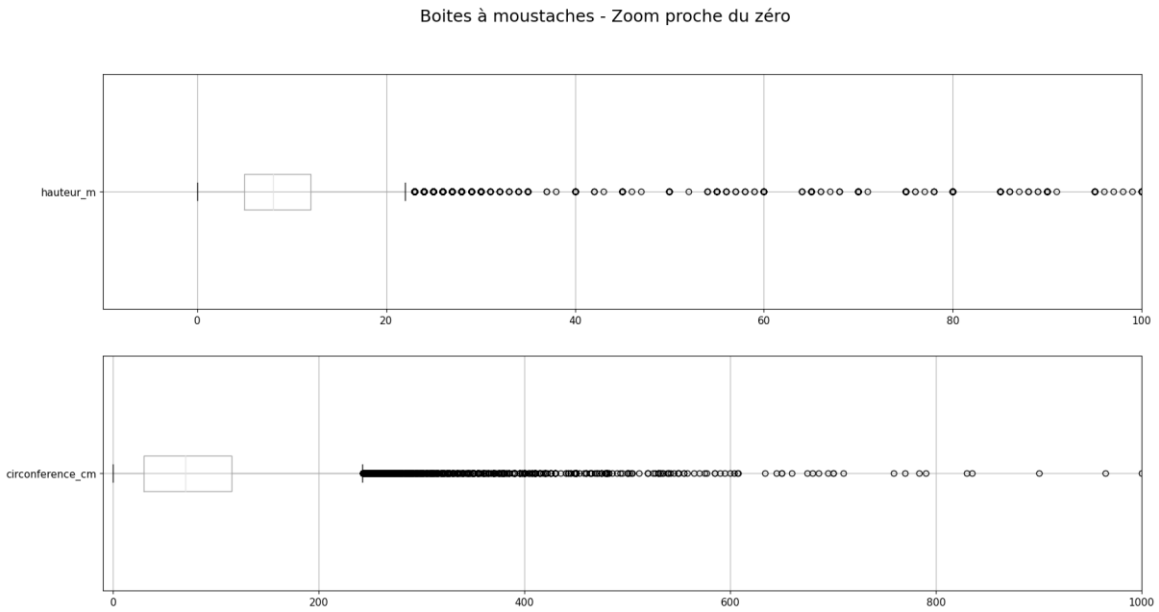






2.1. Exploration - variables biométriques

Hauteur_m et circonference_cm



Cas du zéro …

- Aberrant
- Rappel : % valeurs manquantes = 0%

➡ zéros = valeurs manquantes

- Mais : $N_{\text{zéros hauteur\_m}} > N_{\text{zéros circonference\_cm}}$…

➡ hypothèse : problèmes arrondi

VILLE DE PARIS

2.1. Exploration - variables biométriques

Hauteur_m et circonference_cm

Choix pour nettoyage :

- Hauteur_cm
- Zéros :
  - circonference_cm = 0 ➡ circonference_cm valeur manquante
  - Hauteur_cm = 0 ⬇
    - circonference_cm = 0 ➡ Hauteur_cm valeur manquante
    - circonference_cm > 0 ➡ Hauteur_cm laissée à 0
- Puis, traiter valeurs hautes avec critère Tukey

VILLE DE PARIS

2.2. Nettoyage – suppressions

Type_emplacement
Numero
Complement_adresse
Id_emplacement
variete

2.2. Nettoyage – renommage

Stade_développement : J, JA, A, M → 1-Jeune, 2-Jeune Adulte, 3-Adulte, 4-Mature

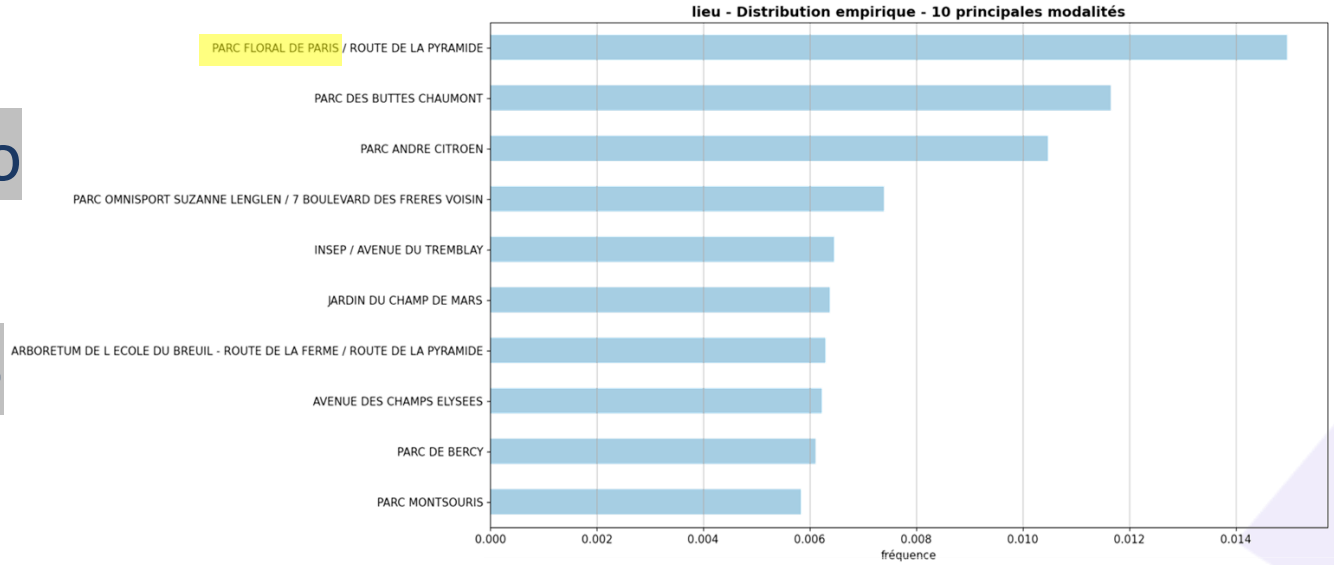
Remarquable : 0.0, 1.0 → non, oui

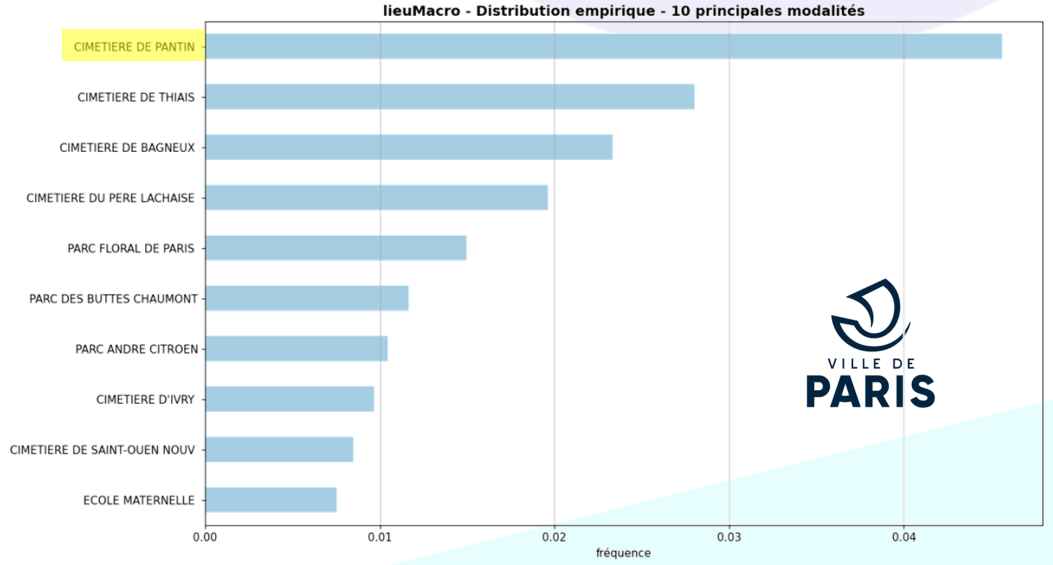
VILLE DE PARIS

2.2. Nettoyage – lieu

Lieu

- LieuMacro
- LieuMicro





2.2. Nettoyage 🖌 – libelle_francais, genre, espece

Imputer libelle_francais grâce à genre et espece

a) Créer libelle_francais_GUESS

b) Filtrer genre et espece : valeurs pour lesquelles libelle_francais = valeur manquante

c) Pour chacune d'elles, regarder les différentes valeurs possibles de libelle_francais

d) Hypothèse/Choix : **si** une seule et unique valeur de libelle_francais

➡ libelle_francais_GUESS reçois celle-ci pour les arbres concernés

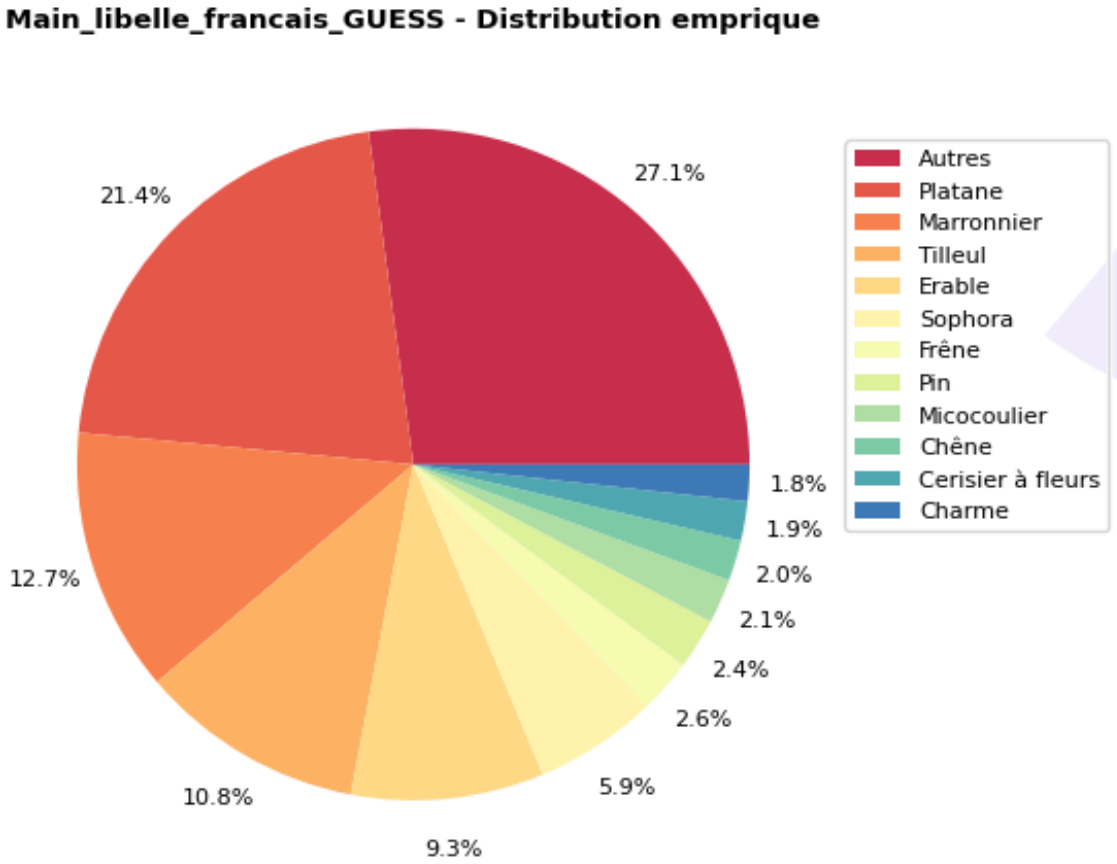
2053 valeurs manquantes ➡ 823 pour libelle_francais_GUESS après nettoyage

VILLE DE PARIS

2.2. Nettoyage – libelle_francais, genre, espece

Créer Main_libelle_francais_GUESS, Main_genre, Main_espece

a) Seuil de représentativité = 1,75%
b) Sous de seuil, valeurs remplacées par « Autres »
c) Plus facile pour visualiser :

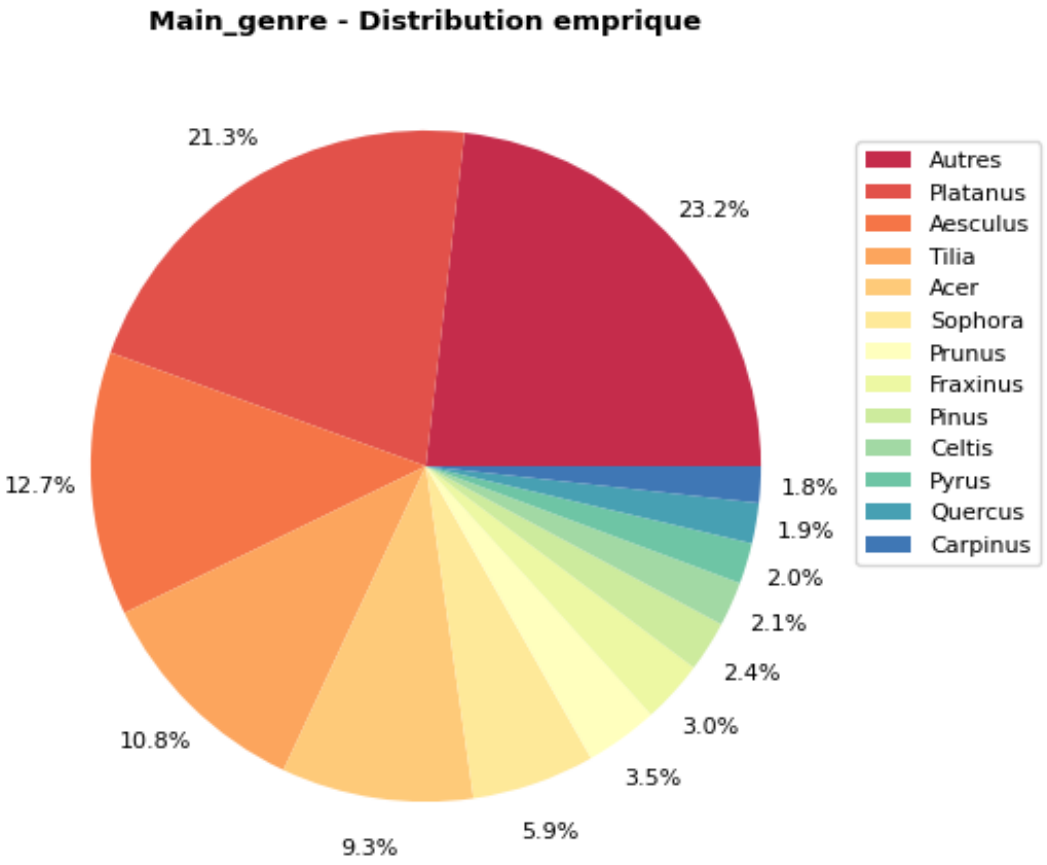


VILLE DE PARIS

2.2. Nettoyage – libelle_francais, genre, espece

Créer Main_libelle_francais_GUESS, Main_genre, Main_espece

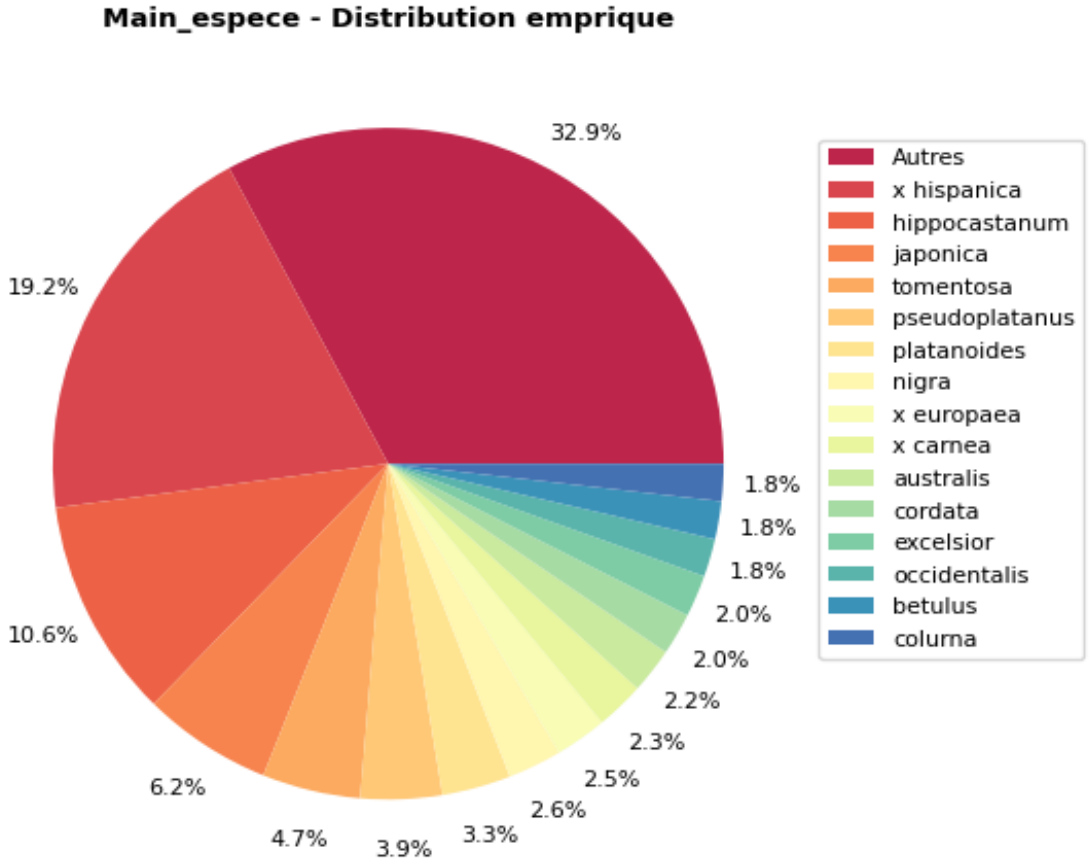
a) Seuil de représentativité = 1,75%
b) Sous de seuil, valeurs remplacées par « Autres »
c) Plus facile pour visualiser :



2.2. Nettoyage 🧹 – libelle_francais, genre, espece

Créer Main_libelle_francais_GUESS, Main_genre, Main_espece

a) Seuil de représentativité = 1,75%
b) Sous de seuil, valeurs remplacées par « Autres »
c) Plus facile pour visualiser :



VILLE DE PARIS

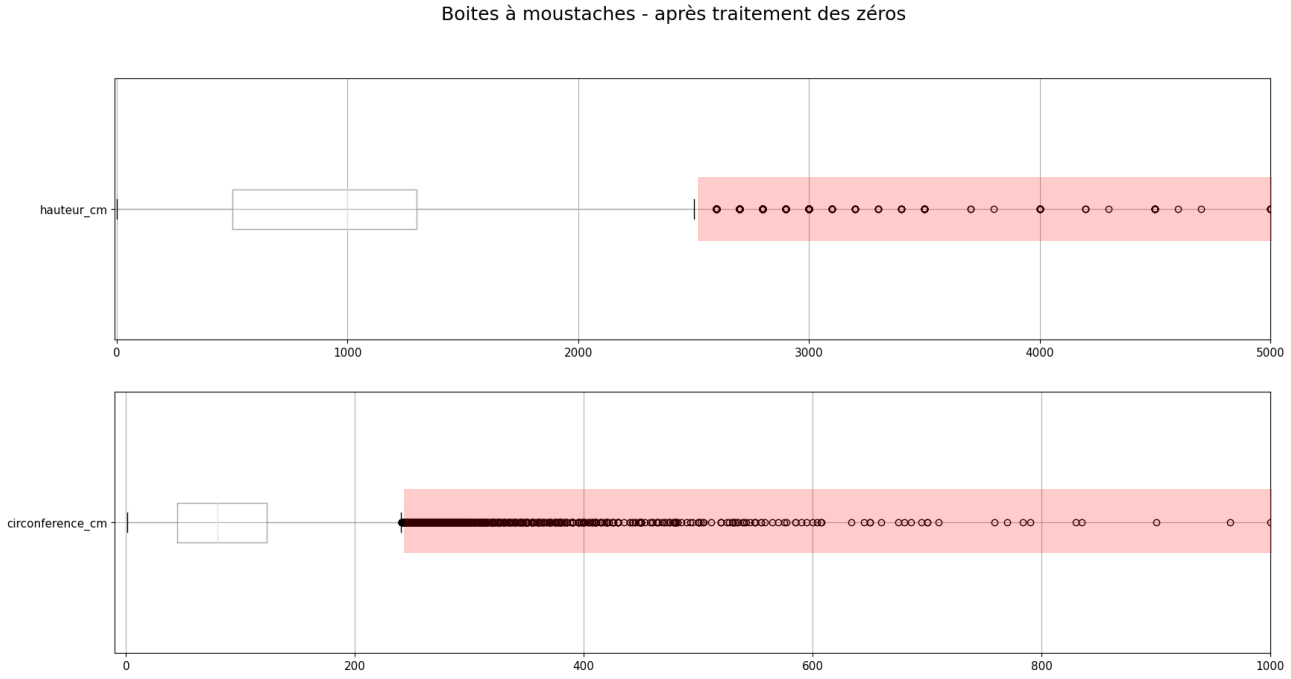
2.2. Nettoyage – circonference, hauteur

a) Créer hauteur_cm

b) Traiter les zéros

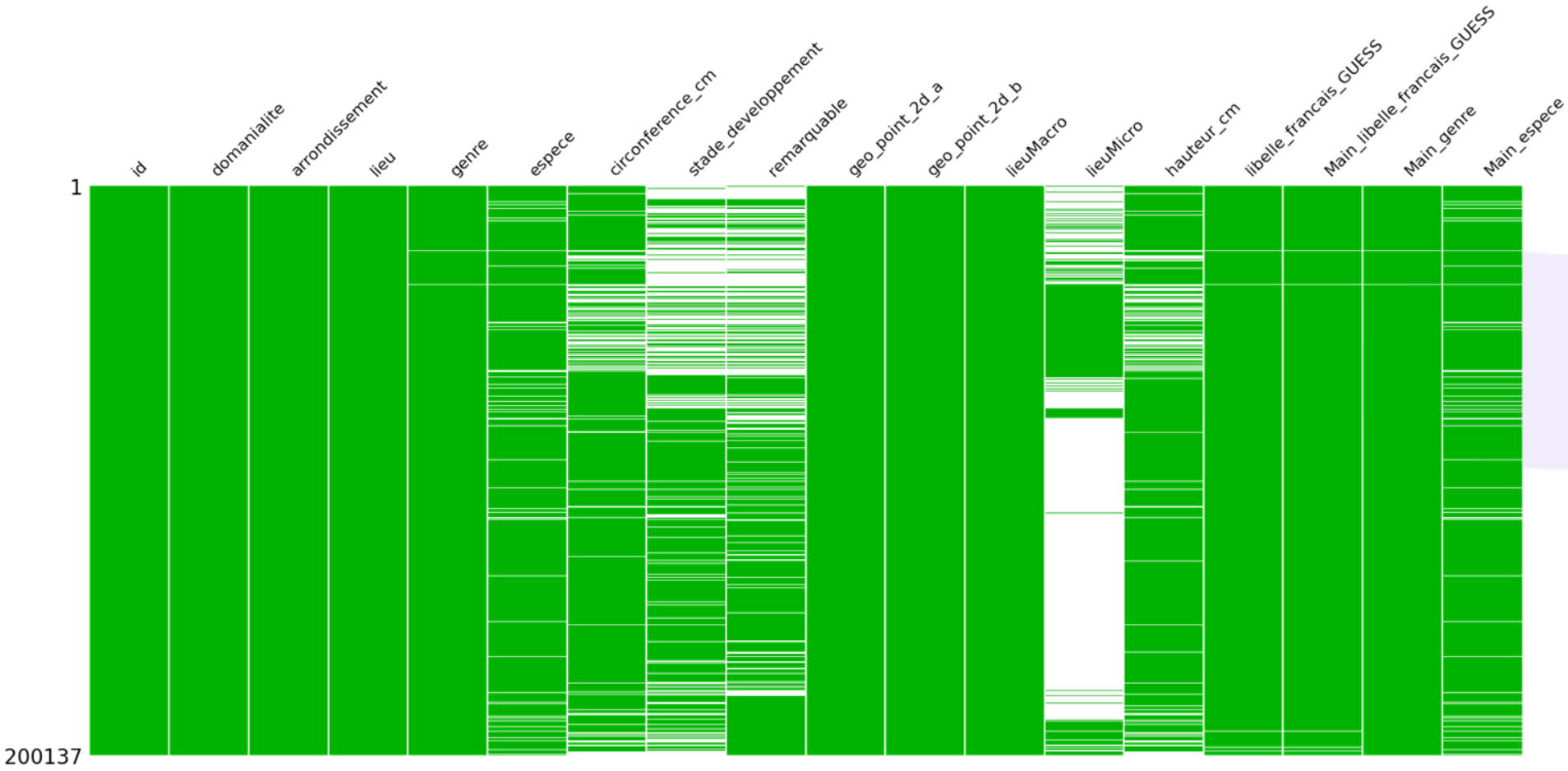
- circonference_cm = 0 ➔ circonference_cm valeur manquante
- Hauteur_cm = 0
  - circonference_cm = 0 ➔ Hauteur_cm valeur manquante
  - circonference_cm > 0 ➔ Hauteur_cm laissée à 0

c) Traiter les valeurs aberrantes hautes :



VILLE DE PARIS

2.2. Nettoyage



VILLE DE PARIS

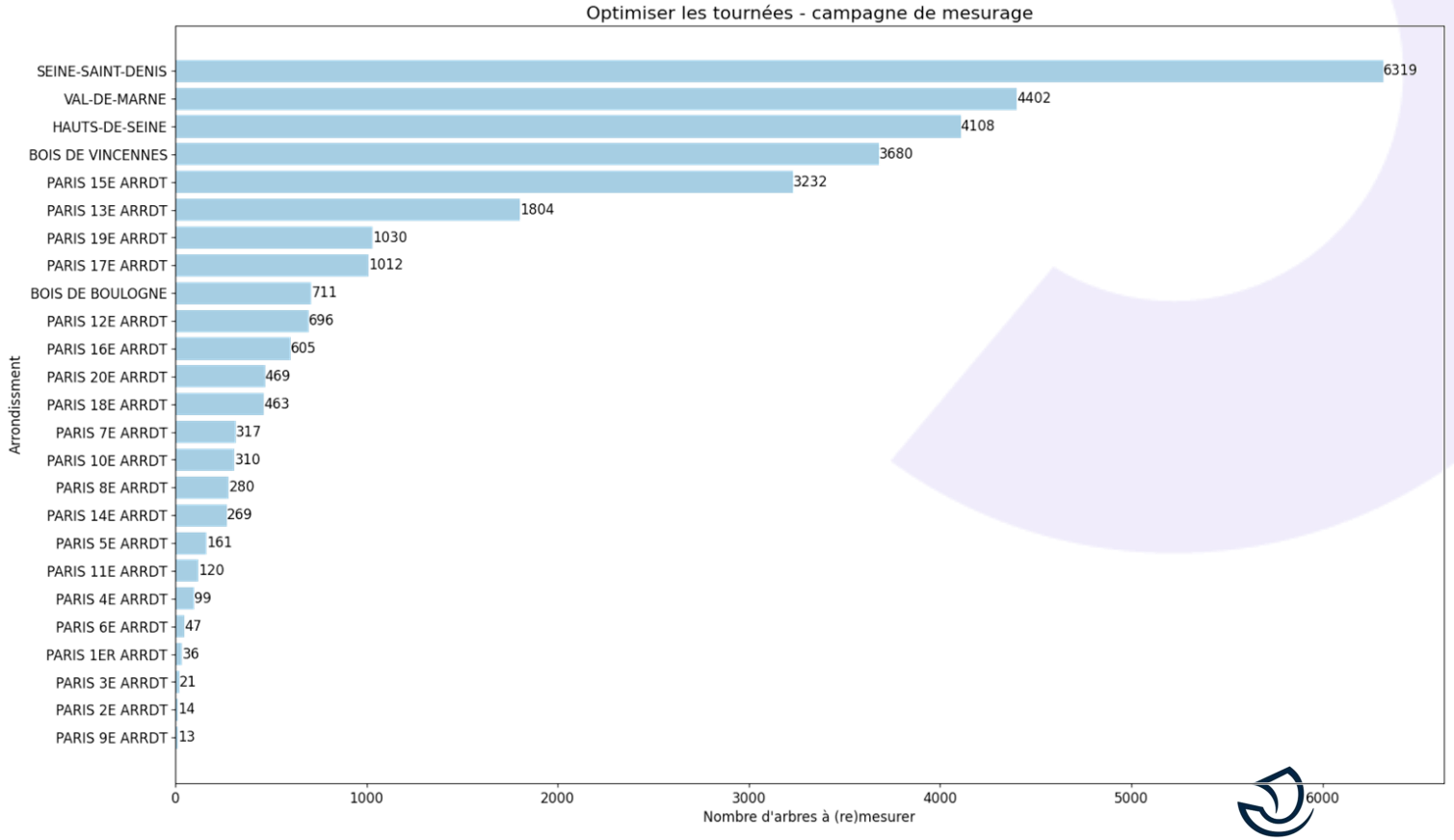
PARTIE 3 - SYNTHÈSE DE L'ANALYSE DE DONNÉES

3.1) Optimiser les tournées - 📏 mesurer les arbres suspects

- Dimensions conditionnent entretien (ex : élagage).

- Données plus exhaustives permettent plus de précisions :
  - Nombre **réel** d'arbres à entretenir par secteur
  - Dimensionnement des **équipes**
  - **Planification** / anticipation

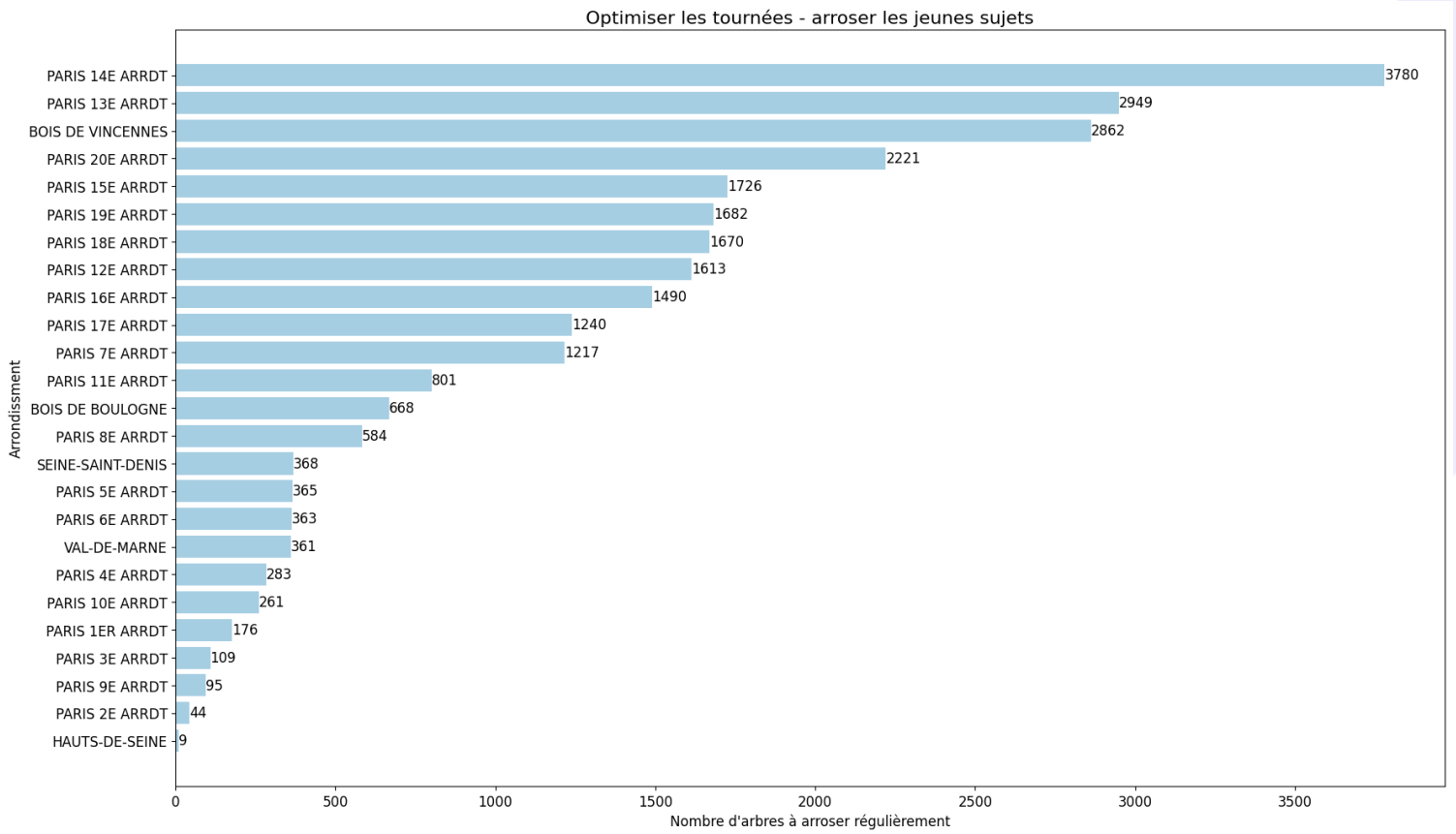
- **Campagne de mesurage** sur les arbres dont les dimensions ont été écartées lors du nettoyage :



VILLE DE PARIS

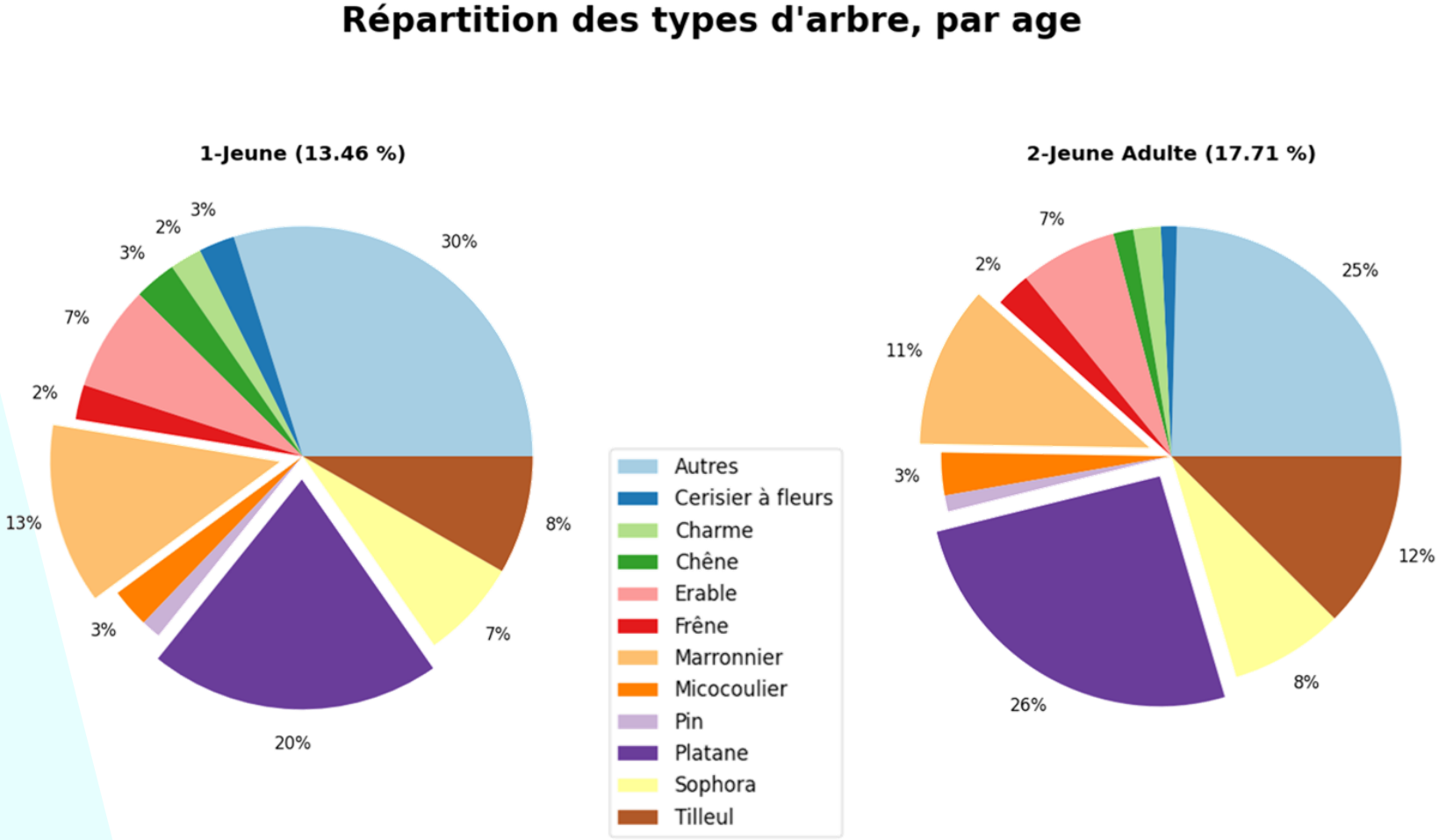
3.2) Optimiser les tournées - arroser le jeunes arbres

- Jeunes sujets ont besoin :
  - Arrosage **régulier**
  - Pendant **2-3 ans**
- Objet d'un entretien **spécifique**
- Sur le même principe, utiliser jeu de données pour optimiser, dimensionner et planifier



VILLE DE PARIS

3.3) Optimiser les tournées - savoir quel arbre planter



Comment évolue la répartition et la diversité des essences ?

Les projets d'espaces verts neufs ou de renouvellement ont un impact.

➡ exemple avec le Platane ou le Marronnier dont la représentativité diminue au cours des générations

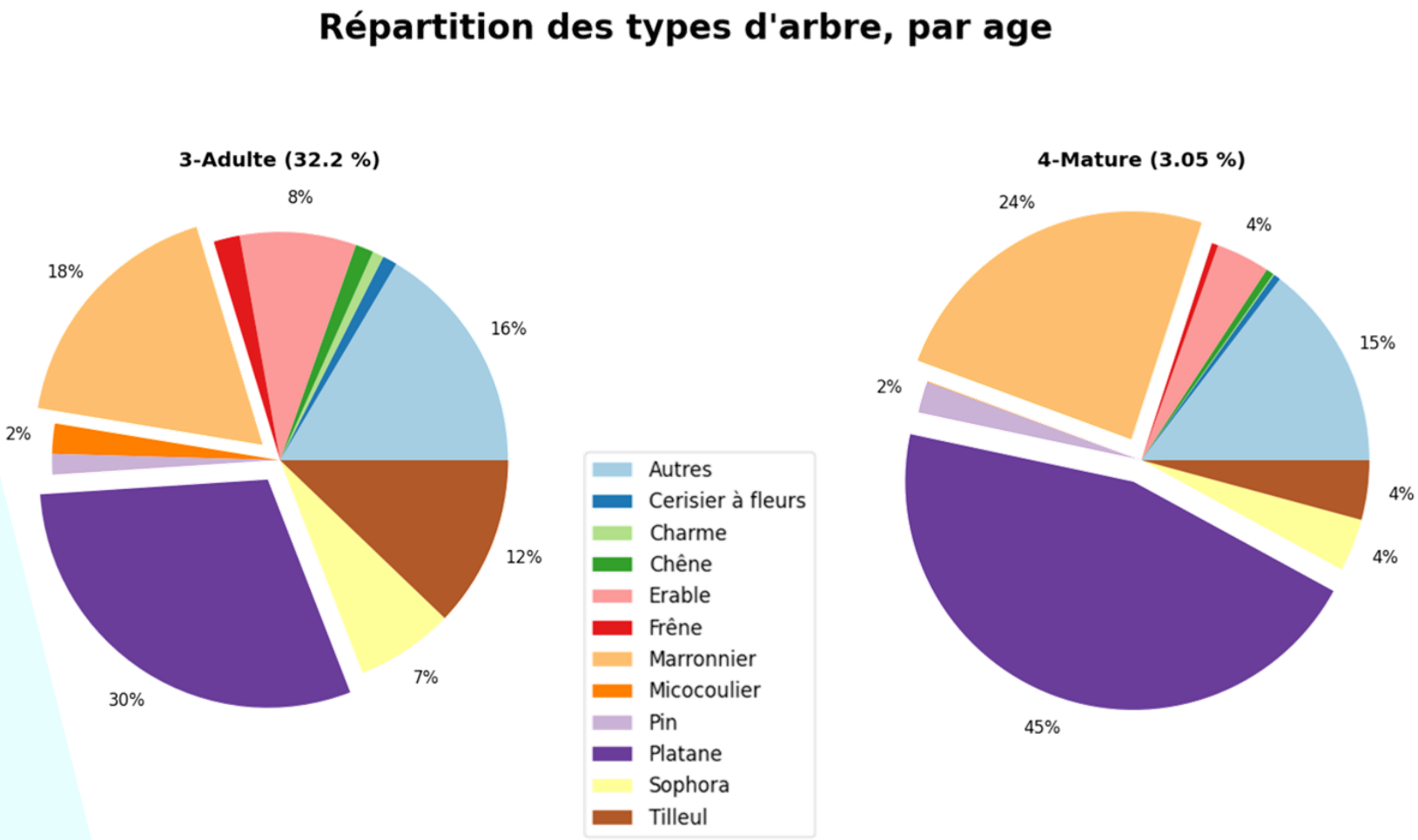
Le jeu de données permet :

- Pleine conscience du phénomène
- Futurs arbitrages

VILLE DE PARIS

Voir aussi slide suivante ➡

3.3) Optimiser les tournées - savoir quel arbre planter



Comment évolue la répartition et la diversité des essences ?

Les projets d'espaces verts neufs ou de renouvellement ont un impact.

➡ exemple avec le Platane ou le Marronnier dont la représentativité diminue au cours des générations

Le jeu de données permet :

- Pleine conscience du phénomène
- Futurs arbitrages

VILLE DE PARIS

Merci

VILLE DE PARIS