



Végéталіsons la ville de Paris

SOMMAIRE

1. CONTEXTE DU PROJET
2. PRÉSENTATION GÉNÉRALE DU JEU DE DONNÉES
3. DÉMARCHE MÉTHODOLOGIQUE D'ANALYSE
4. SYNTHÈSE DE L'ANALYSE DES DONNÉES



1. CONTEXTE DU PROJET

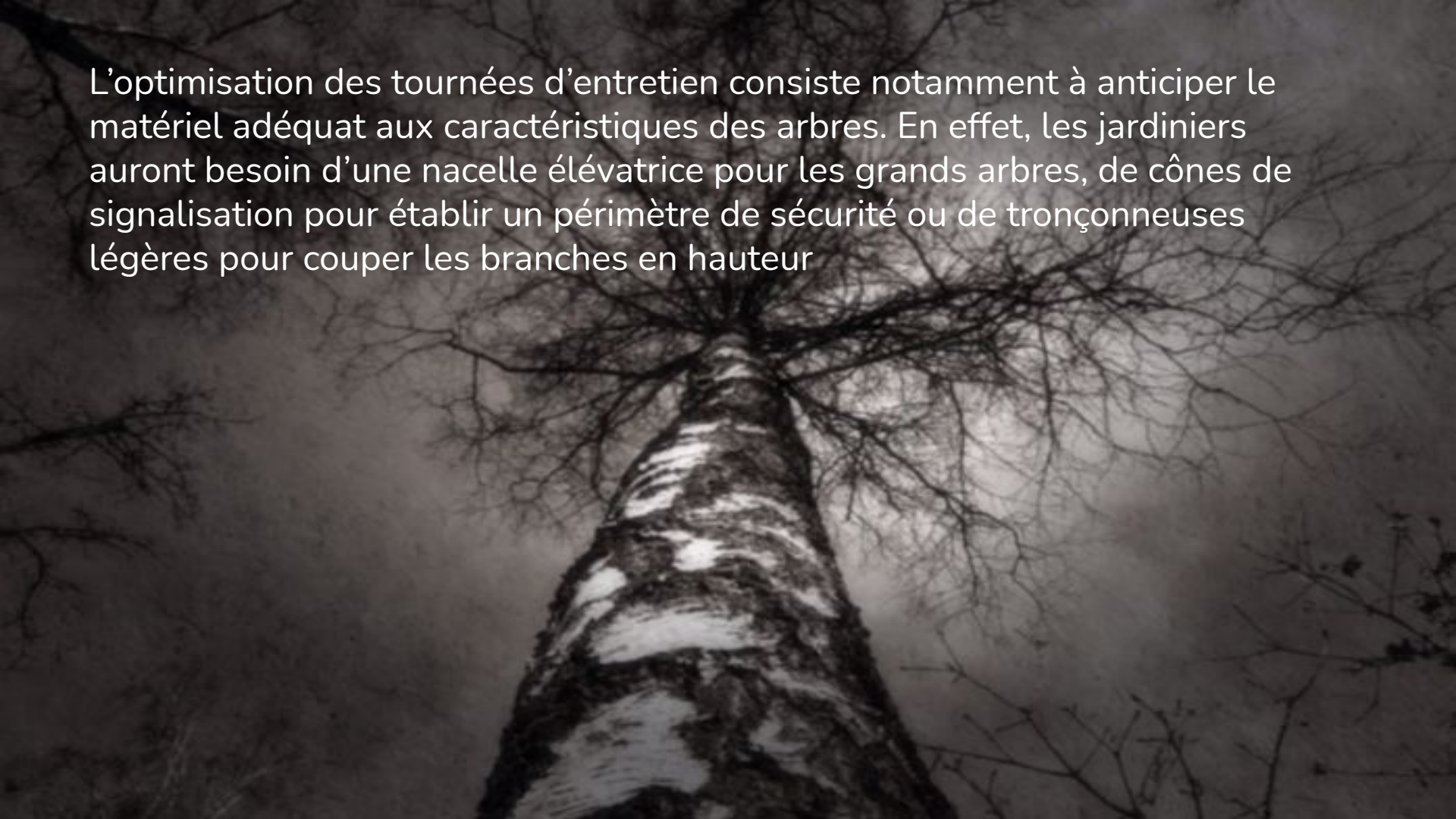
Le programme “Végétons la ville” de Paris doit servir à optimiser les tournées d’entretien des arbres de la ville. Pour cela, **une analyse exploratoire** sera faite sur le jeu de données des arbres de la ville (fichier p2-arbres-fr.csv).





Aujourd'hui, les agents jardinier souhaite entretenir les arbres du **20e arrondissement**, et précisément ceux **du cimetière**.
Mais ils ont un problème, ils doivent préparer le matériel adéquat et planifier la durée de cette tournée. Comment doivent ils faire pour optimiser cette tournée ?

L'optimisation des tournées d'entretien consiste notamment à anticiper le matériel adéquat aux caractéristiques des arbres. En effet, les jardiniers auront besoin d'une nacelle élévatrice pour les grands arbres, de cônes de signalisation pour établir un périmètre de sécurité ou de tronçonneuses légères pour couper les branches en hauteur






Les agents doivent optimiser les tournées en regroupant les arbres par âge. En effet, les jeunes arbres nécessitent plus de soins et de manière plus fréquente pour l'arrosage et l'élagage et le fait de tous les regrouper dans une même tournée.

Certaines espèces ou genre
d'arbres tels que le cèdre sont
sensibles aux froid et
nécessitent des gaines de
protection





L'organisation du travail d'entretien suivant la domanialité (alignement, cimetière ou jardin) n'est pas la même. C'est une donnée à prendre en compte dans l'optimisation des tournées.



L'entretien d'arbre remarquable tel
que l'orme de Saint Gervais
nécessite des soins particuliers
contre certaines maladies au vue de
son âge avancé.

2. PRÉSENTATION GÉNÉRALE DU JEU DE DONNÉES



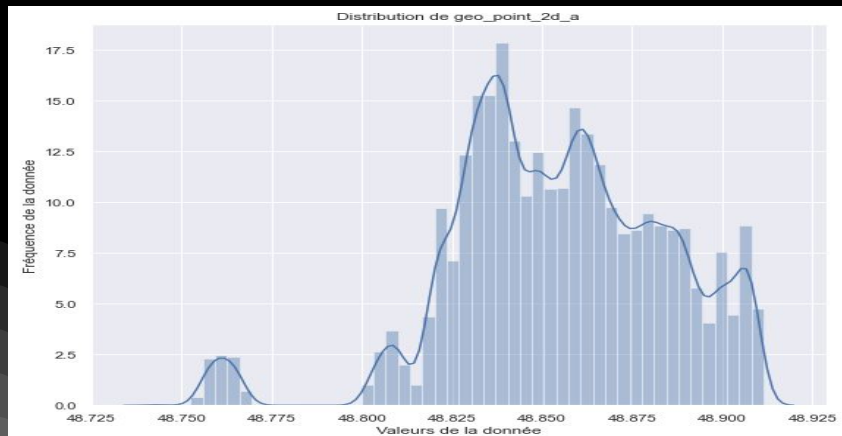
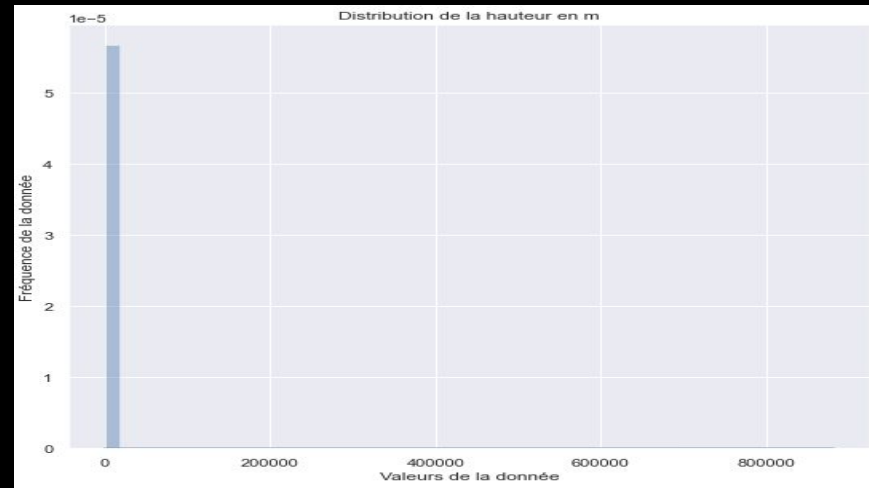
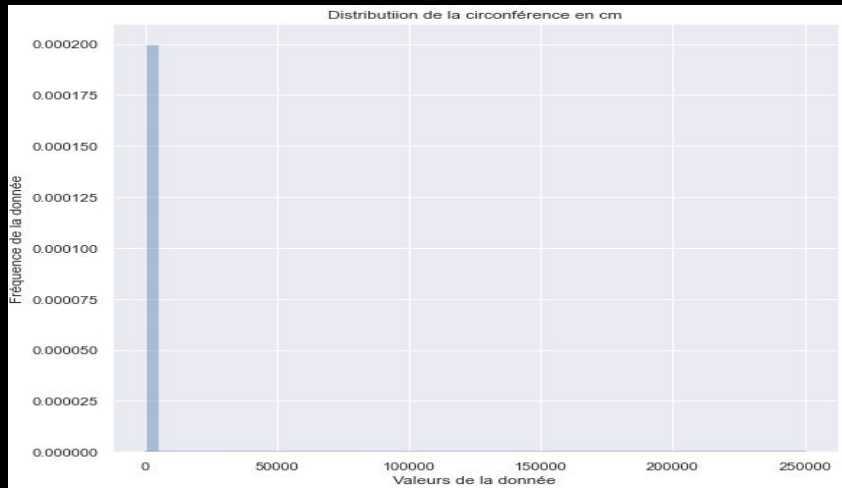
- Le jeu de données des arbres de la ville (fichier p2-arbres-fr.csv) contient 17 colonnes et 200137 lignes.
- Le fichier contient des informations sur la localisation géographique de chaque arbre ainsi que ces caractéristiques telles que la hauteur, la circonférence, le genre, l'espèce, le stade de développement, la domanialité et la donnée 'remarquable'.
- Les colonnes 'complement_adresse', 'numero', 'libelle_francais', 'genre', 'espece', 'variete', 'stade_developpement', 'remarquable' contiennent des données manquantes NaN qu'il faut traiter pour notre analyse.

- Indicateurs statistiques du jeu de données avant leur traitement

acement	libelle_francais	genre	espece	variete	circonference_cm	hauteur_m	stade_developpement	remarquable	geo_point_2d_a	geo_point_2d_b
200137	198640	200121	198385	36777	200137.000000	200137.000000	132932	137039.000000	200137.000000	200137.000000
69040	192	175	539	436	NaN	NaN	4	NaN	NaN	NaN
101001	Platane	Platanus	x hispanica	Baumannii'	NaN	NaN	A	NaN	NaN	NaN
1324	42508	42591	36409	4538	NaN	NaN	64438	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	83.380479	13.110509	NaN	0.001343	48.854491	2.348208
NaN	NaN	NaN	NaN	NaN	673.190213	1971.217387	NaN	0.036618	0.030234	0.051220
NaN	NaN	NaN	NaN	NaN	0.000000	0.000000	NaN	0.000000	48.742290	2.210241
NaN	NaN	NaN	NaN	NaN	30.000000	5.000000	NaN	0.000000	48.835021	2.307530
NaN	NaN	NaN	NaN	NaN	70.000000	8.000000	NaN	0.000000	48.854162	2.351095
NaN	NaN	NaN	NaN	NaN	115.000000	12.000000	NaN	0.000000	48.876447	2.386838
NaN	NaN	NaN	NaN	NaN	250255.000000	881818.000000	NaN	1.000000	48.911485	2.469759

- On observe des valeurs aberrantes pour la **circonférence** et la **hauteur** avec des valeurs de 250255 cm et 881818m.

- Distribution des variables quantitatives



- On observe une asymétrie vers la droite des distributions de la circonférence et de la hauteur avec des valeurs aberrantes au-delà de 66 m. Cela est mis en évidence l'indicateur du **skewness empirique** qui est très important et égale à 447.294. De même pour **le kurtosis empirique** qui est important et qui reflète l'influence des valeurs aberrantes sur ces 2 variables quantitatives.
- Pour les données '**geo_point _2d_a**' et '**geo_point 2d b**', on observe une asymétrie de la distribution mais non significative. Elles sont plus aplaties que la distribution normale car le kurtosis empirique est négatif -0.01 et ne présentent pas de valeurs aberrantes ou atypiques.


NETTOYAGE ET TRAITEMENT DES DONNÉES

Les données vont être
traitées et certaines colonnes
(`complement_adresse`,
`numero` et `variete`)
supprimées.



3. DÉMARCHE MÉTHODOLOGIQUE D'ANALYSE



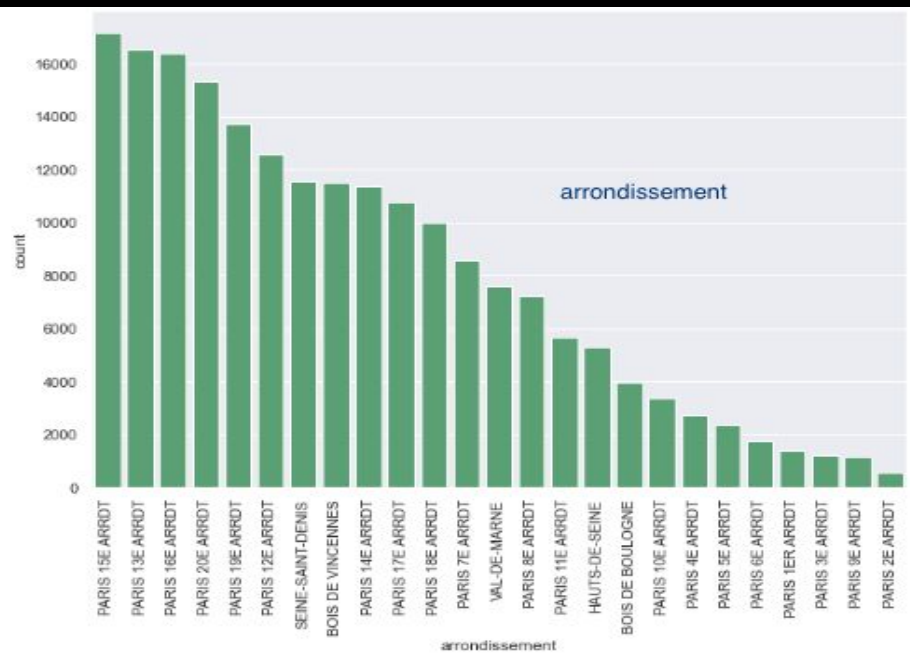
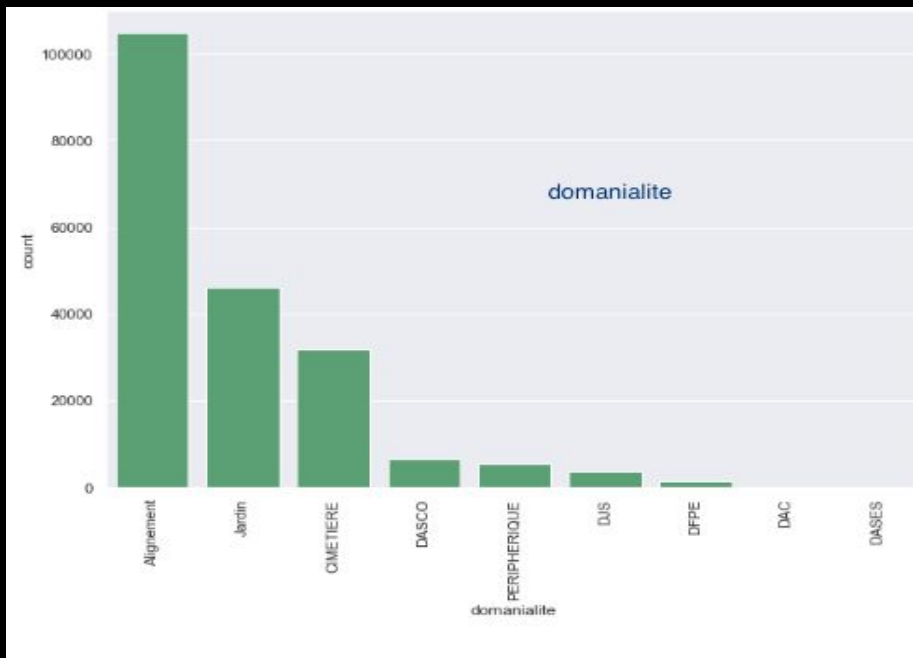
The background is a dark blue gradient with abstract, glowing elements. It features a faint, stylized financial candlestick chart in the upper right, a red line graph with a peak and trough in the center, and a white line graph with multiple peaks and valleys in the lower left. Binary code (0s and 1s) is scattered throughout, particularly on the right side. Horizontal red and blue light streaks are visible at the bottom.

Nous allons réaliser une analyse exploratoire des données concernant les arbres de Paris afin de mettre en évidence la répartition géographique des hauteurs, des circonférences, de l'âge, du genre d'arbre, de la remarquabilité et de la domanialité.

Nous allons chercher les éventuelles corrélations entre certaines données.

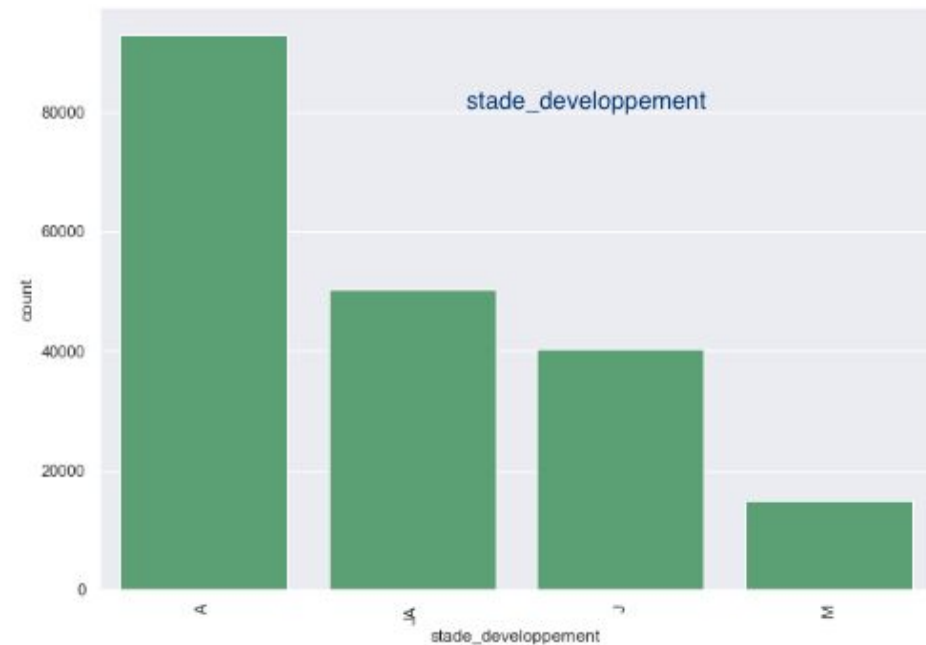
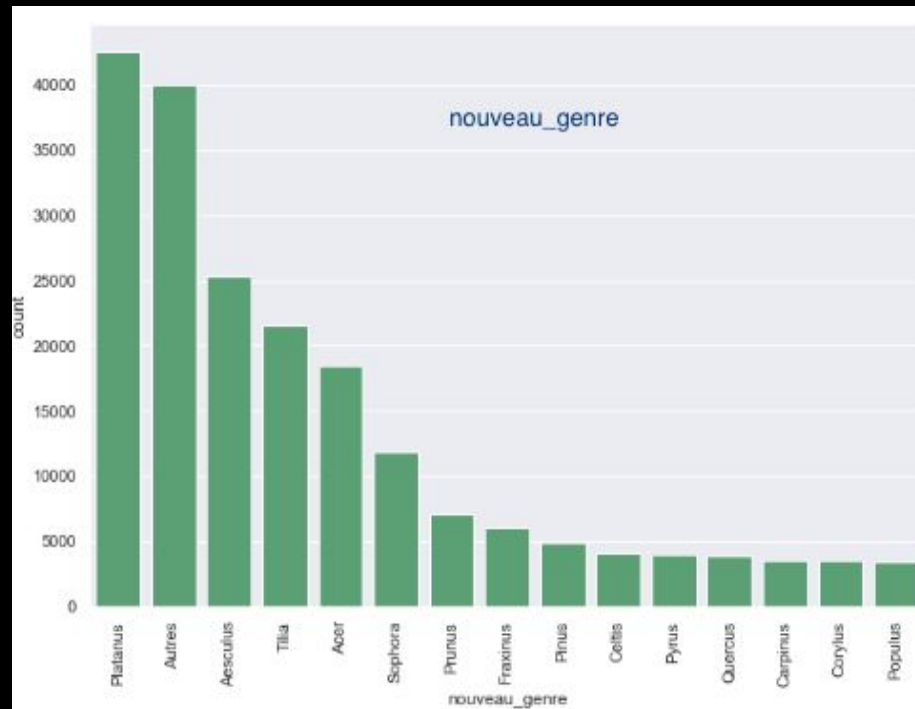
Pour l'analyse exploratoire, les données qualitatives domanialité, arrondissement, genre et stade_developpement vont être étudiées.

Distribution des variables qualitatives



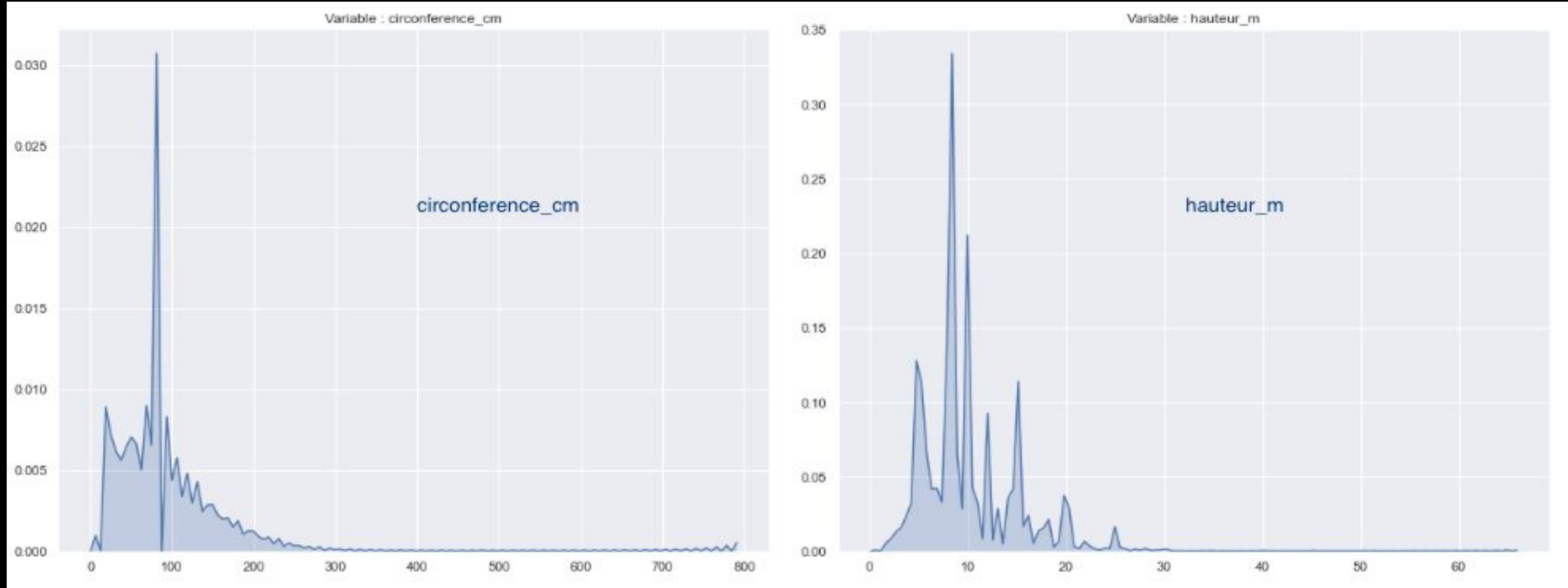
La domanialité nous montre que la majorité des arbres se trouve dans les alignements, les jardins puis les cimetières.

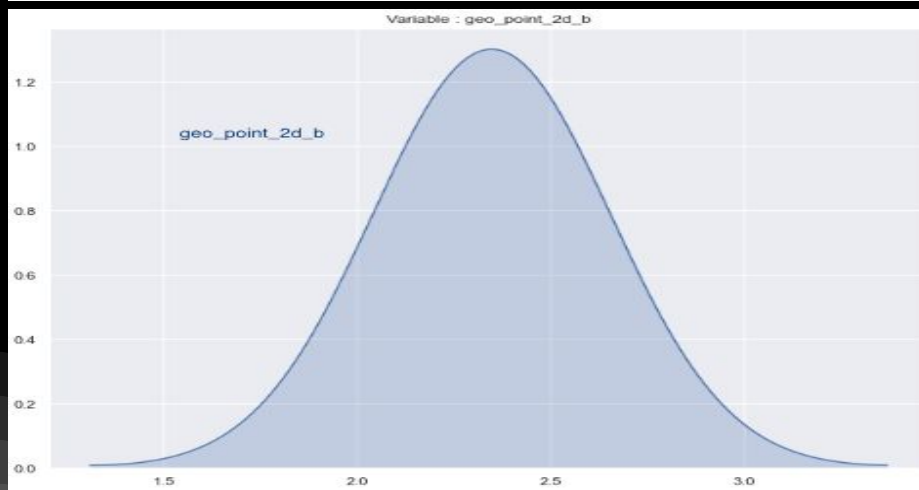
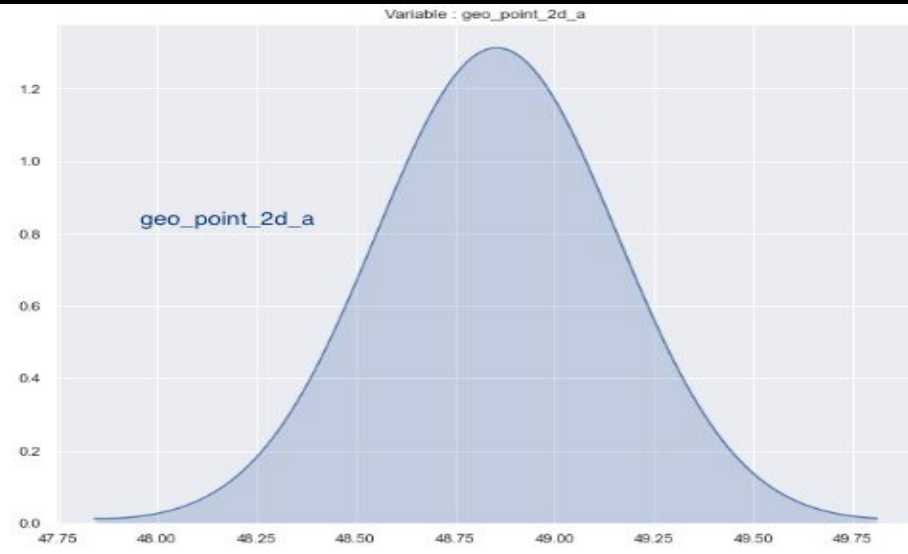
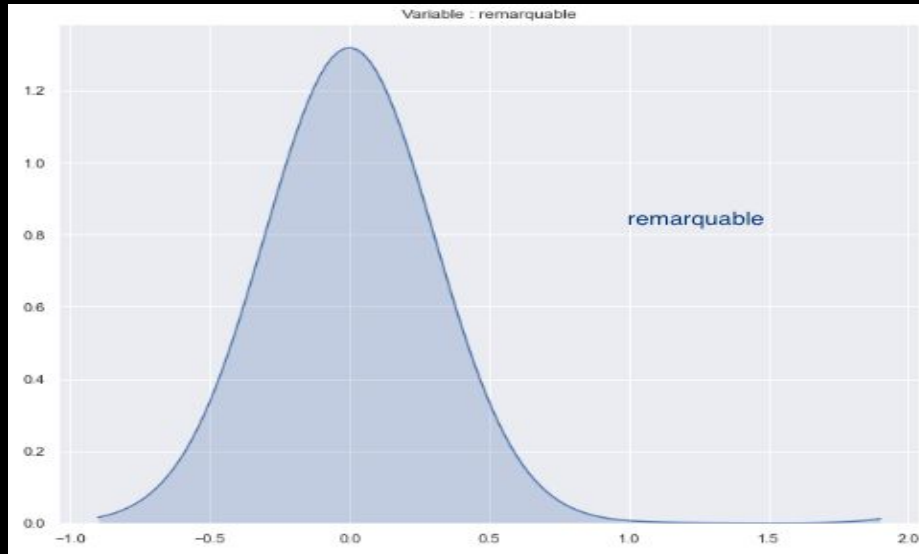
Les arrondissements 15, 13, 16, 20, 19 et 20 concentrent une grande majorité des arbres.



Les genres d'arbre prédominants sont le Platanus, Aesculus, Tilia et Acer. Quant au stade de développement, les arbres adultes sont en majorité.

Distribution des variables quantitatives hauteur_m < 66m,
circonference_cm < 800cm, remarquable, geo_point_2d_a et geo_point_2d_b





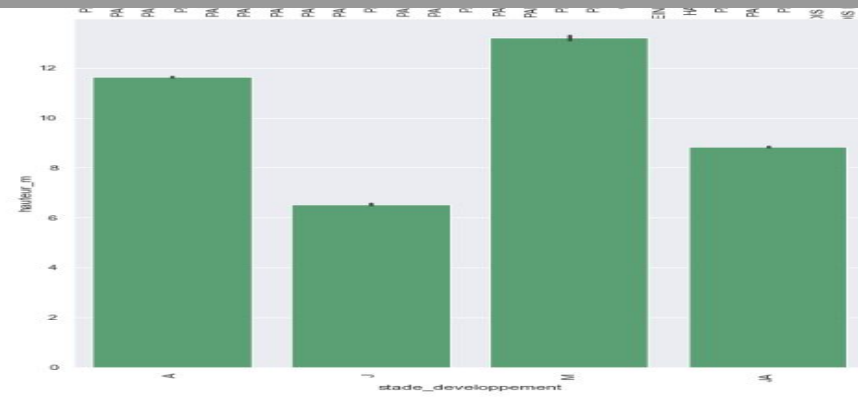
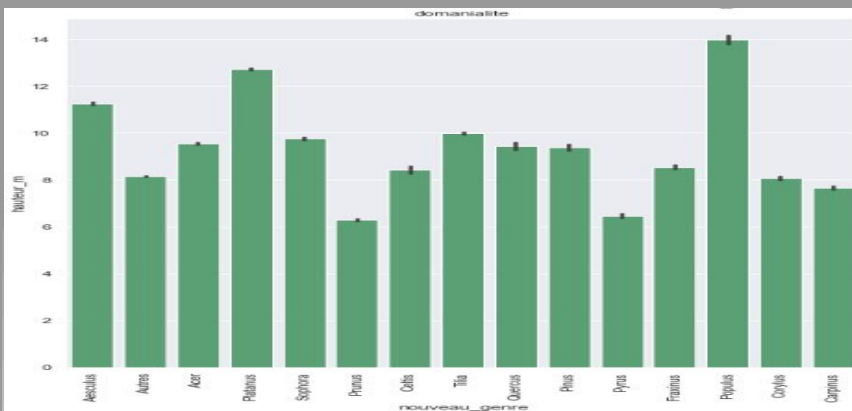
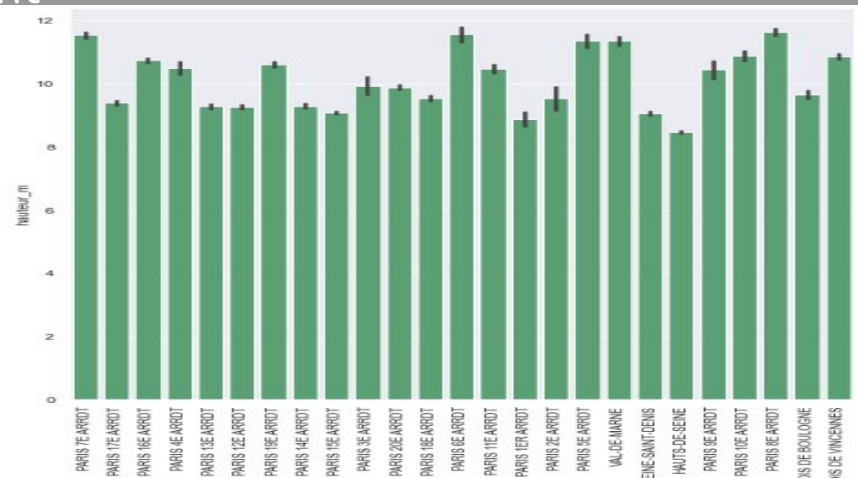
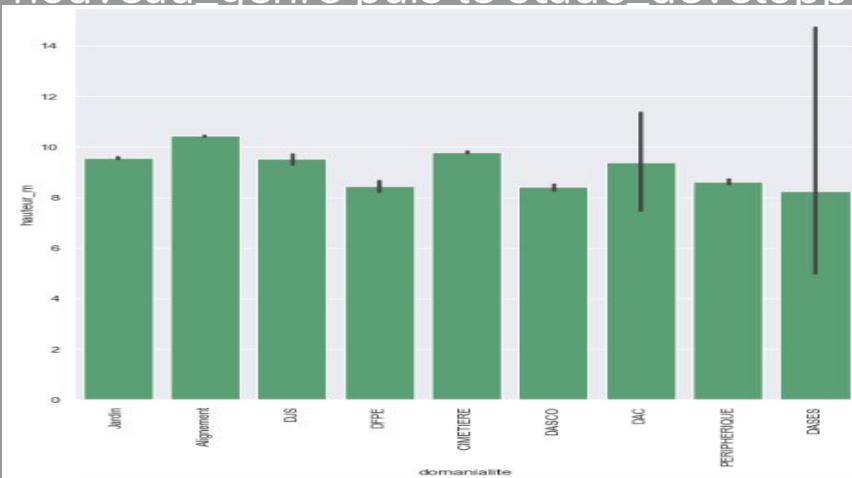
Les données `geo_point_2d_a` et `geo_point_2d_b` ont la même distribution de Gauss ce qui signifie que 68.6% des arbres se situent dans une zone de largeur $2 \times \text{écart-type}$. De même pour la donnée `remarquable` qui est centrée en 0 .

Cherchons les éventuelles corrélations entre la hauteur et les données circonférence, remarquable, geo_point_2d_a et geo_point_2d_b. Le coefficient de corrélation de Pearson est calculé entre chaque donnée.

	circonference_cm	hauteur_m	remarquable	geo_point_2d_a	geo_point_2d_b
circonference_cm	1.000000	0.741702	0.110078	-0.004692	0.006237
hauteur_m	0.741702	1.000000	0.060189	0.005444	0.000848
remarquable	0.110078	0.060189	1.000000	0.002577	-0.009693
geo_point_2d_a	-0.004692	0.005444	0.002577	1.000000	-0.010499
geo_point_2d_b	0.006237	0.000848	-0.009693	-0.010499	1.000000

On constate une forte corrélation entre la hauteur et la circonférence avec un coefficient de corrélation de 0.7417

Sur ces graphiques, on observe la répartition des hauteurs suivant la domanialité et suivant les arrondissements. Et la corrélation entre la hauteur et le nouveau_genre puis le stade_developpement



La connaissance de la répartition géographique des hauteurs et circonférences permettent d'optimiser les tournées d'entretien.



4. SYNTHÈSE DE L'ANALYSE DE DONNÉES



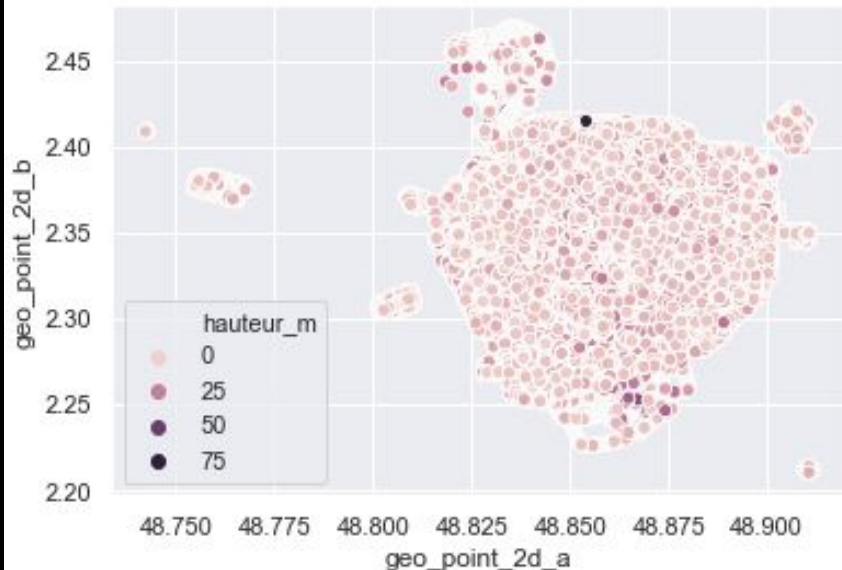
Dans le cadre du projet **Végé**talisons la ville, l'optimisation des tournées d'entretien des arbres passe par la connaissance de la **répartition géographique** des différentes hauteurs et circonférences mais aussi du genre d'arbre, de l'âge, de la domanialité et de son caractère remarquable.

En analysant la répartition géographique des différentes hauteurs, on observe que les $\frac{2}{3}$ des arbres sont réunis dans les arrondissements 15, 13, 16, 20, 19 et 12 ainsi que le Bois de Vincennes et la Seine-Saint-Denis.



- Point géographique 2d (latitude,longitude):

```
1 plot = sns.scatterplot(data=data_df, x="ge
```



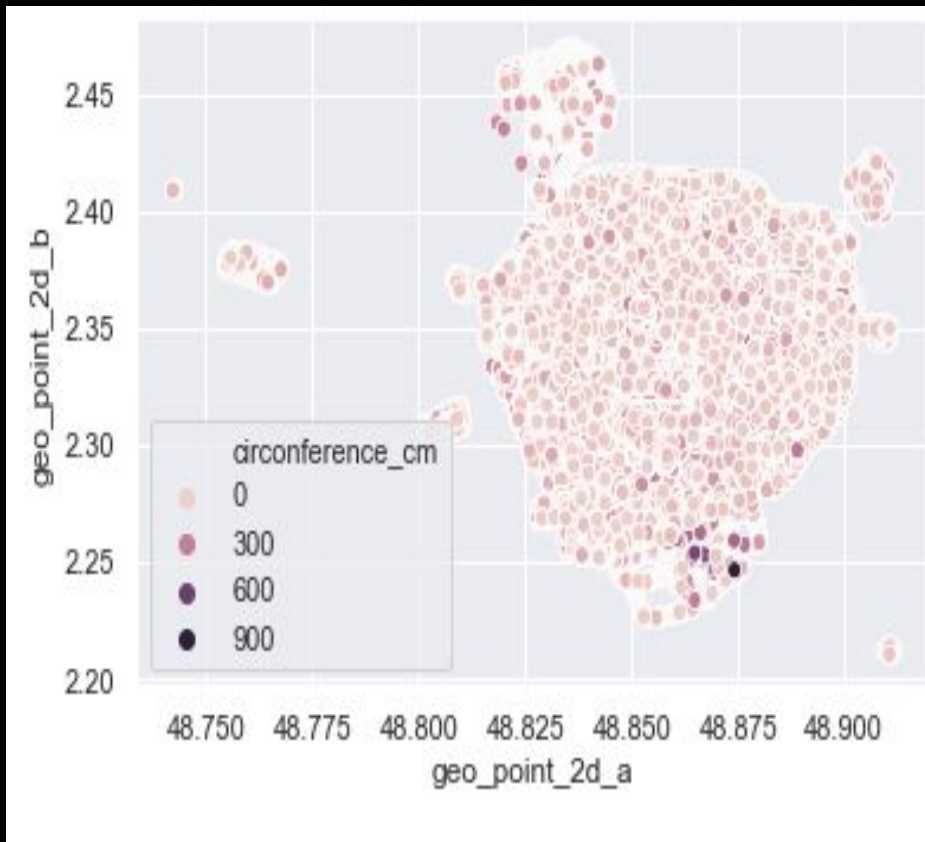
La répartition des différentes tranches de hauteur selon la latitude et la longitude des arbres :

- tranche [0 , 25 [
- tranche [25 , 50 [
- tranche [50 , 75 [
- tranche [75 , ... [

Une grande majorité des arbres ont une hauteur inférieure à 30 m et sont localisés entre les latitudes 48.825 et 48.9 et entre les longitudes 2.25 et 2.4.



Synthèse sur la circonférence des arbres

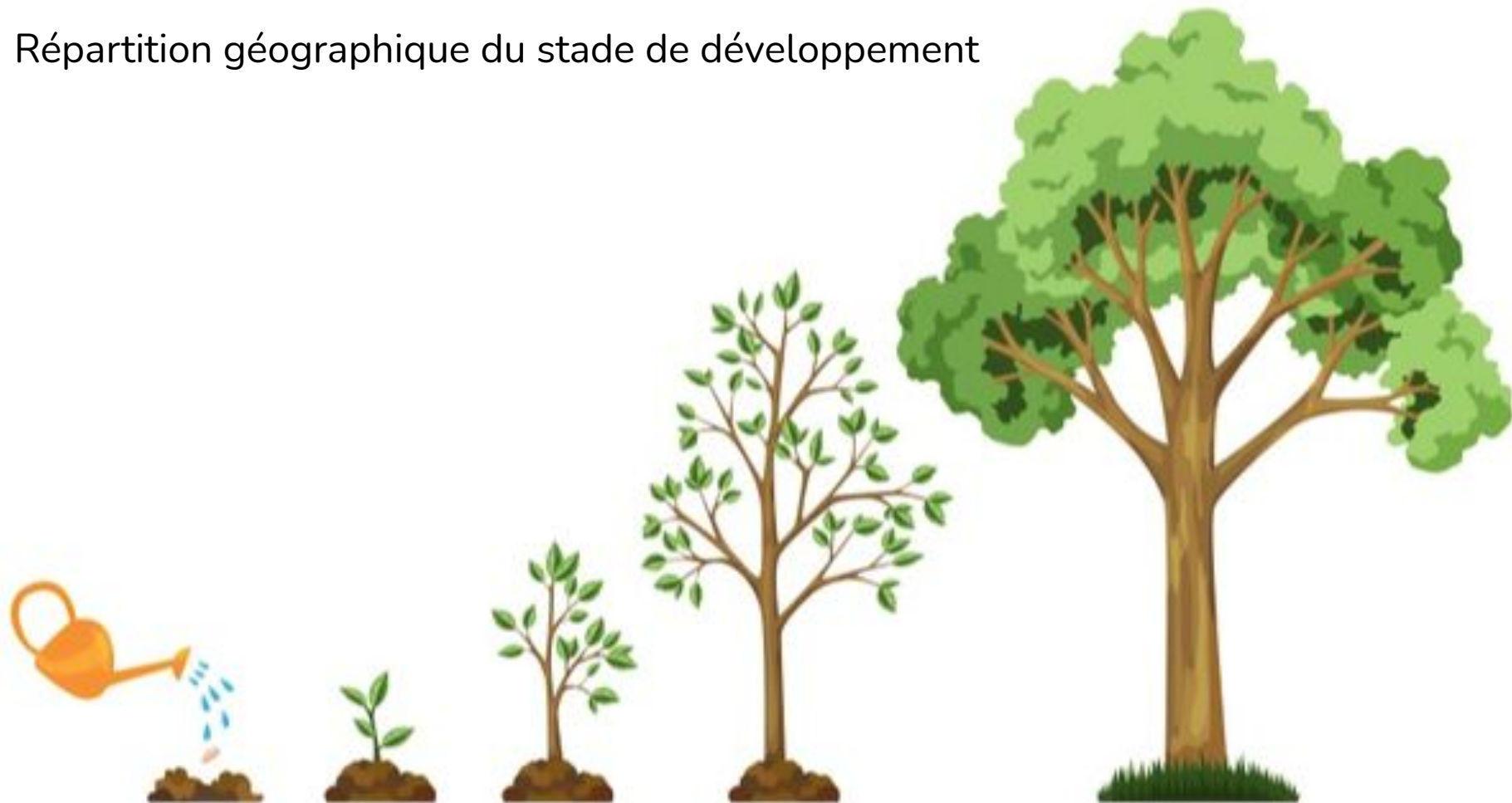


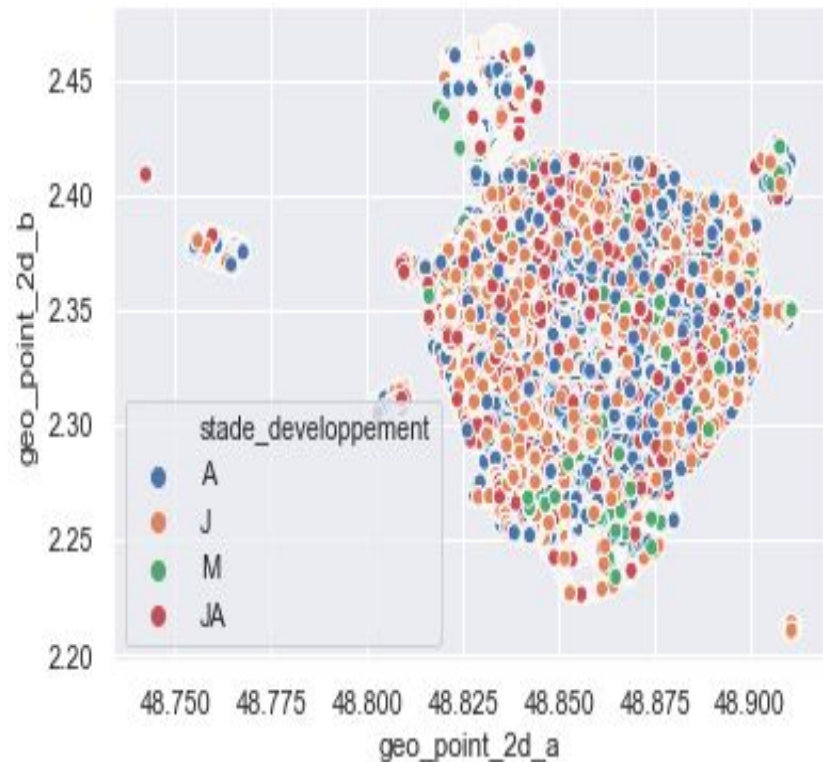
La répartition des différentes tranches de circonférence selon la latitude et la longitude des arbres :

- tranche [0 , 300 [
- tranche [300 , 600 [
- tranche [600 , 900 [
- tranche [900 , ... [

Une grande majorité des arbres ont une circonférence inférieure à 300 cm et sont localisés entre les latitudes 48.825 et 48.9 et entre les longitudes 2.25 et 2.4.

Répartition géographique du stade de développement



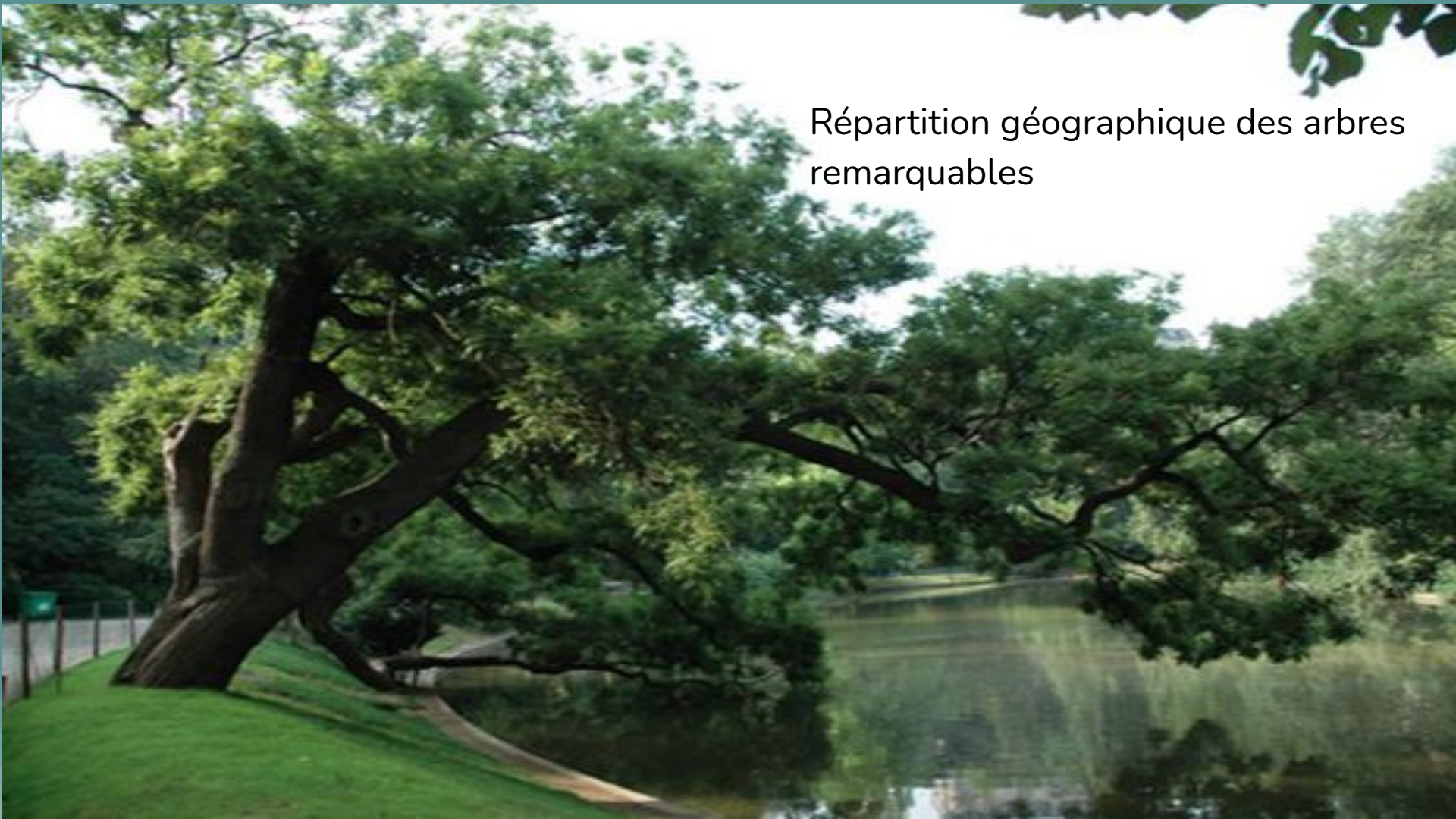


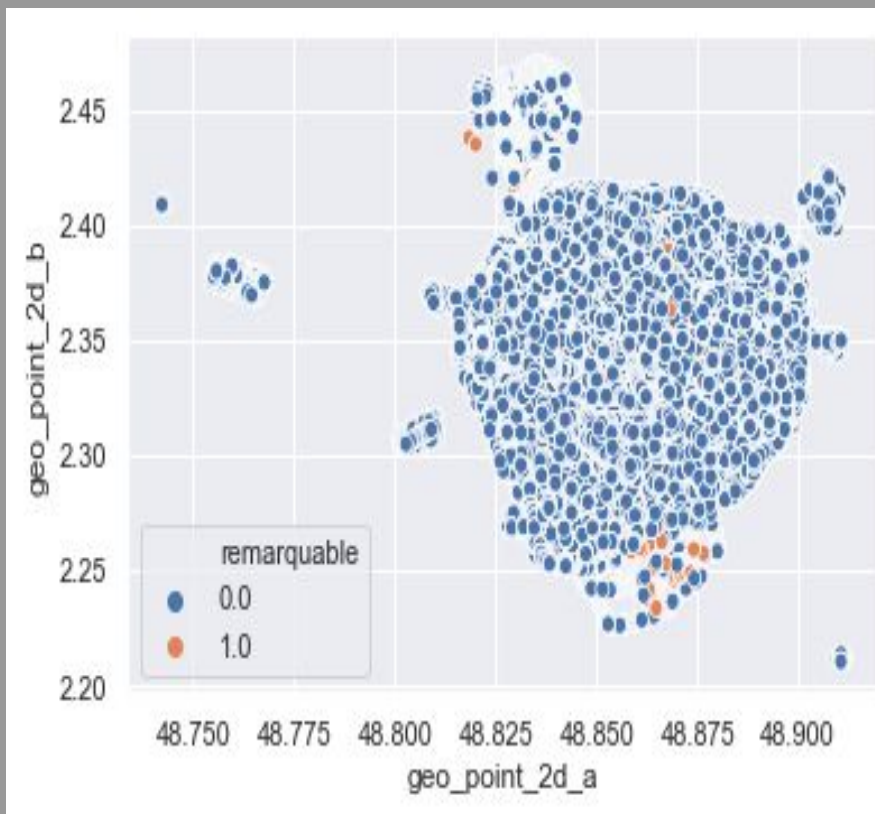
La répartition des différentes valeurs du stade de développement selon la latitude et la longitude :

- A (Adulte)
- J (Jeune)
- M (Mature)
- JA (Jeune Adulte)

Il montre visuellement qu'il y a une majorité de jeunes arbres (J) uniformément répartis dans Paris, puis vient les arbres adultes (A).

Répartition géographique des arbres remarquables



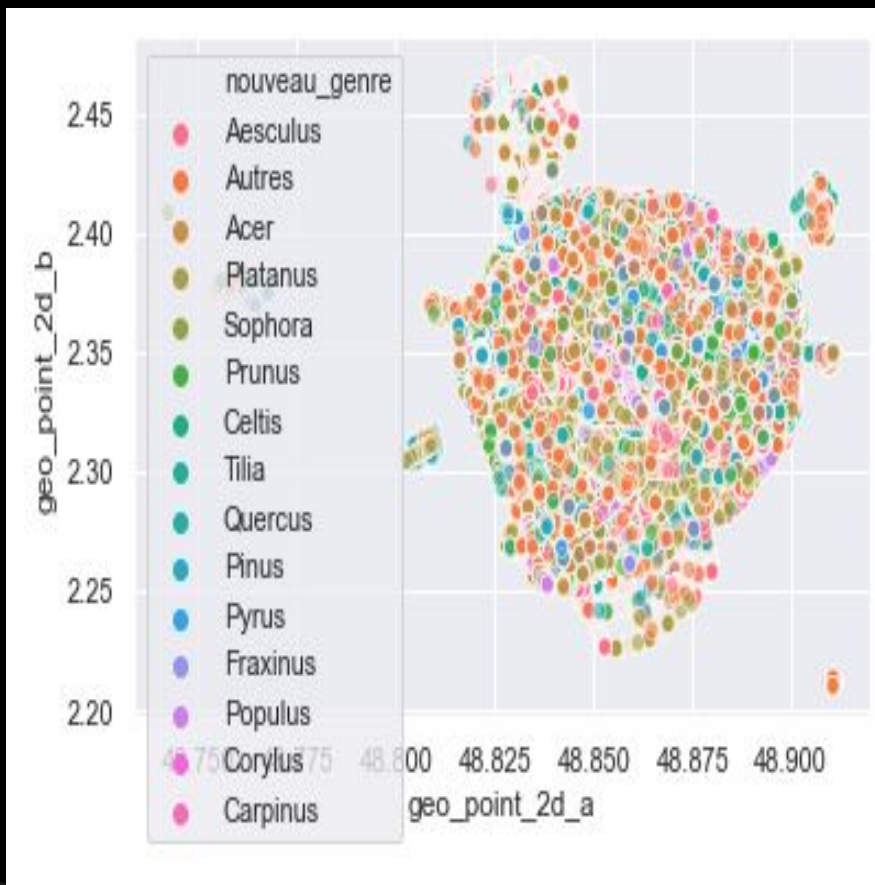


Ce graphique nous montre que la majorité des arbres ne sont pas remarquables (0.0).

On peut voir que beaucoup d'arbres remarquables sont localisés à un même endroit. Ces arbres nécessitent un traitement particulier étant donné leurs grands âges ou leur grande taille.

Synthèse des genres d'arbres

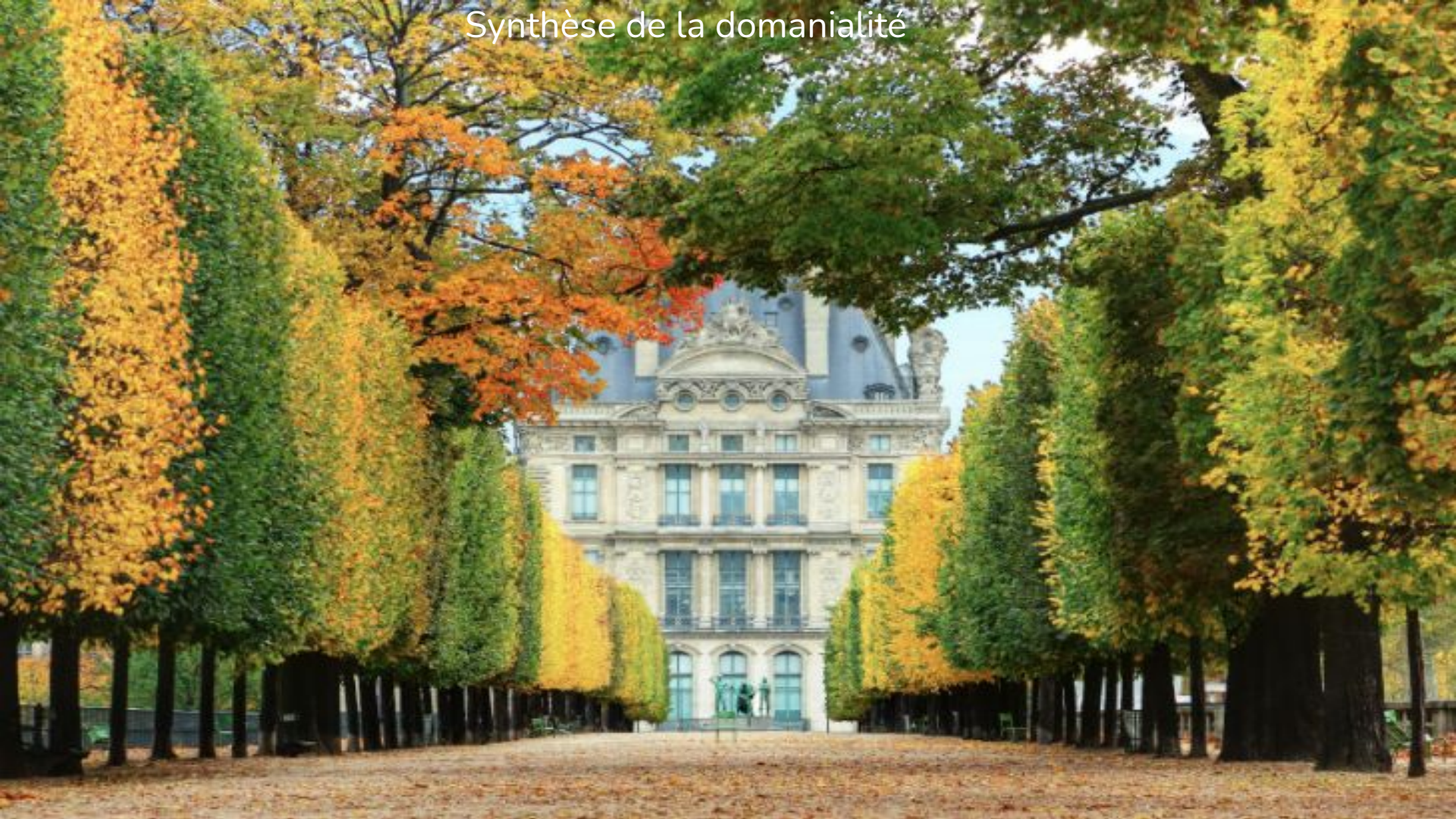


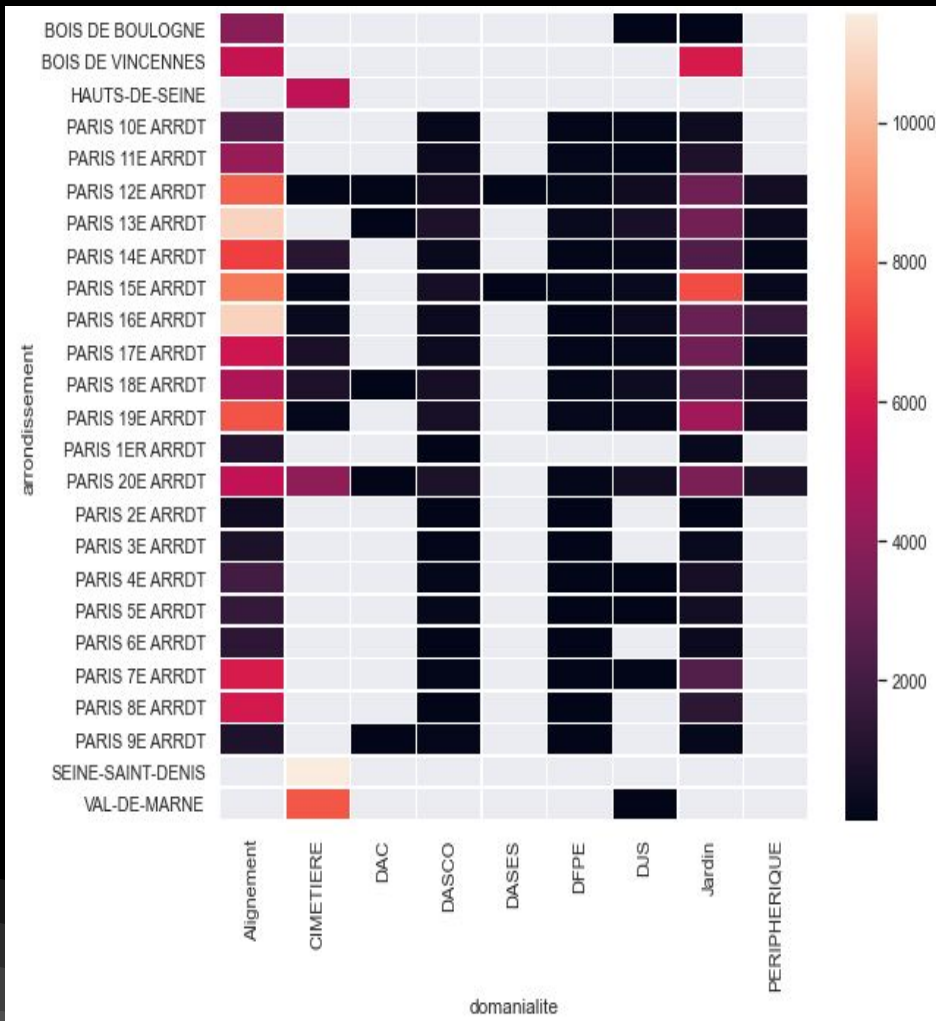


Ce scatterplot nous indique la répartition des différents genres d'arbre selon la latitude et la longitude dans Paris :

- Platanus
- Aesculus
- Tilia
- Sophora
- Acer
- Prunus
- Celtis
- Quercus

Synthèse de la domanialité





Les différentes domanialités regroupant le plus d'arbres sont :

- Alignement
- Jardin
- Cimetière

Les arbres se trouvant dans les alignements peuvent être proches des routes ou des trottoirs donc la mise en place d'un périmètre de sécurité est nécessaire.

Grâce à notre analyse exploratoire, l'agent jardinier sait que dans les cimetières du 20^e arrondissement, il trouvera principalement des platanes, des tilleuls, le genre Acer et Aesculus.

Les arbres ne sont pas remarquables à cet endroit et sont majoritairement adultes avec quelques jeunes arbres. La hauteur des arbres ne dépasse pas 10 m et leur circonférence inférieures à 300 cm.

