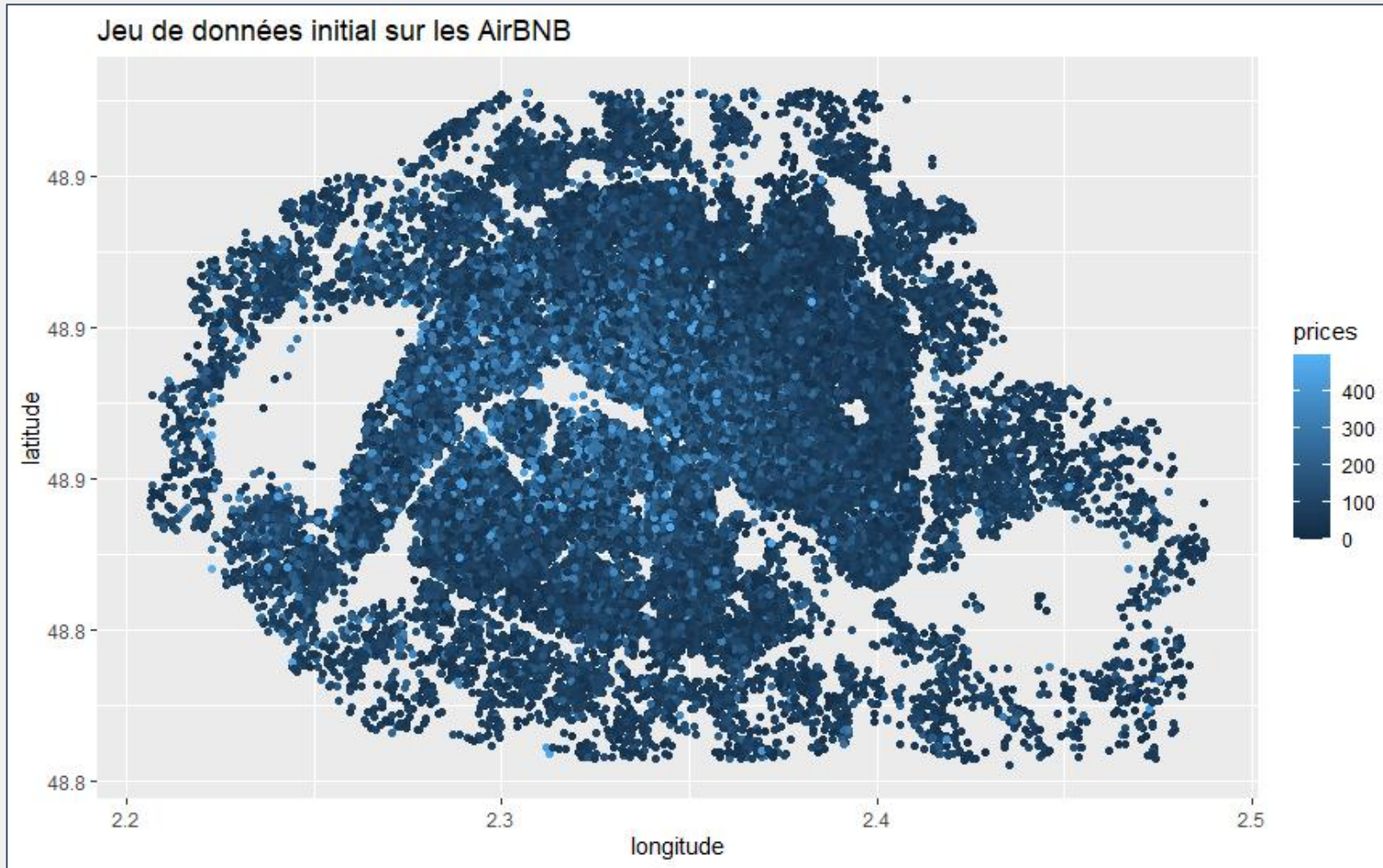


# Etude de la répartition spatiale et de la distribution des prix des locations AirBNB dans Paris pour définir des quartiers touristiques

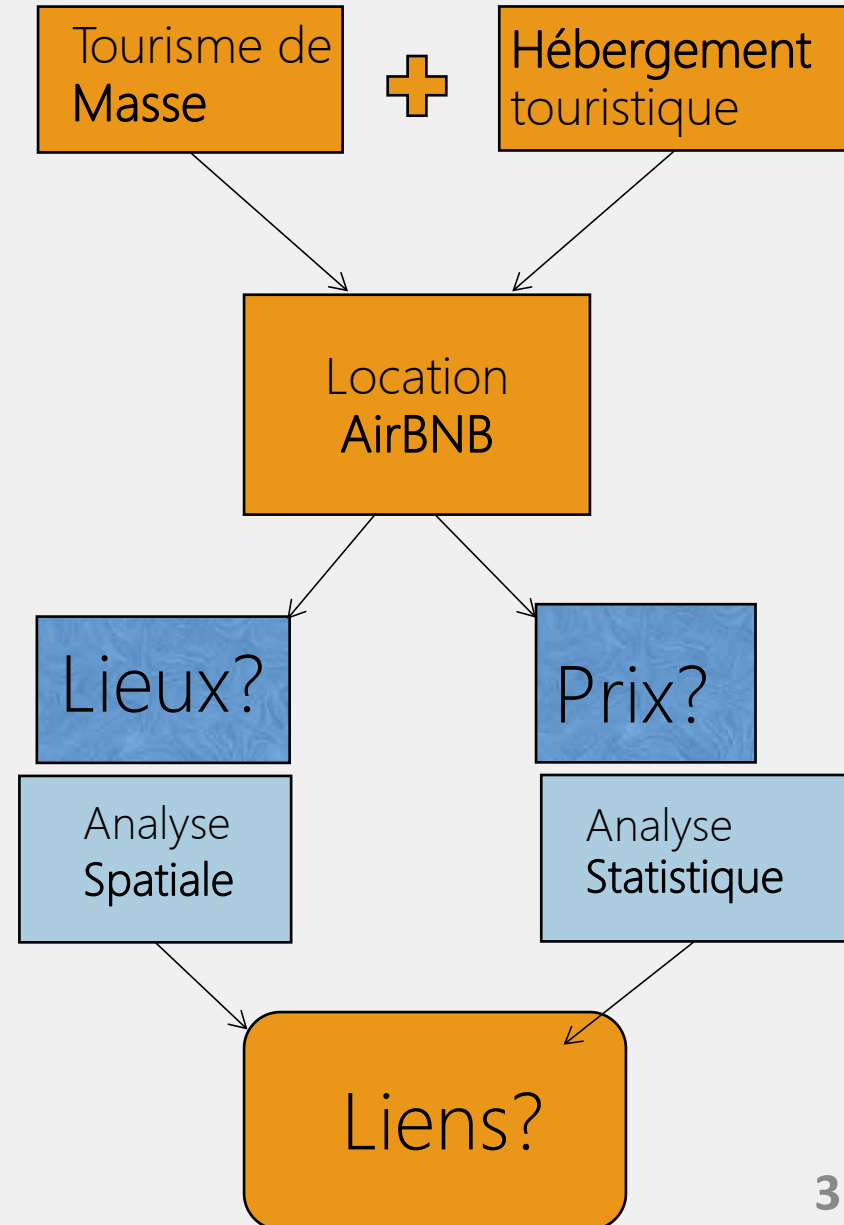
Mise en place de méthodes statistique de détermination de quartiers à fort potentiel d'attraction touristique

# Introduction

# CONTEXTE



**Fig.1** : Nuage de points des prix/nuit des locations AirBNB à Paris (en euros)



# INTRODUCTION

- Comment mettre en place des méthodes d'analyse statistique et spatiale permettant de déterminer au mieux les zones géographiques à potentielle touristique à Paris ?
- Comment mettre en place un système de classification spatiale pour déterminer des clusters d'AirBNB intra-arrondissement ?

# Sommaire

01

Exploration et  
visualisation  
des données

02

Analyse de la  
distribution des prix

03

Analyse de la  
répartition  
spatiale

04

Débuts de  
classification  
intra-arrondissements

01

# Exploration des données

# DATA :

- Locations AirBNB (*InsideAirBNB*)
- Arrondissements et quartiers parisiens(*Open Data ville de Paris*)
- Station de métro et de RER dans Paris (*Data Idf : Mobilités*)
- Sites touristiques en Ile de France (*Data Idf : Sites touristiques*)

## ► Données utilisées

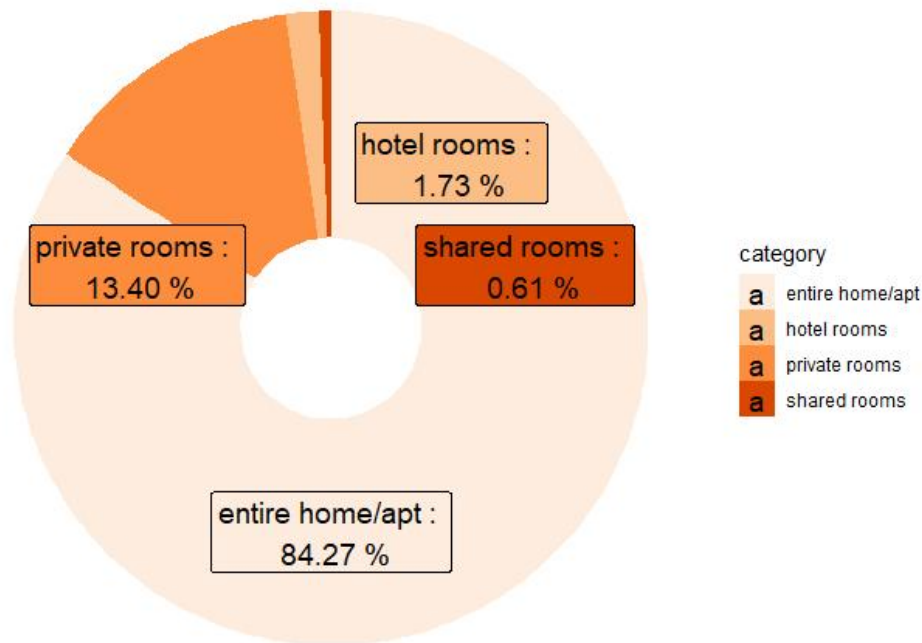
---

- Prix, type et localisation des locations AirBNB (61 365)
- Données spatiales [*polygones*] :
  - Arrondissements (20)
  - Quartiers (80)
- Données spatiales [*points*] :
  - Gares RER
  - Stations de métro
  - Monuments et sites touristiques

Zone d'étude : Paris

## ► Données AirBNB

Répartition des locations AirBNB à Paris selon le type de chambre en pourcentage



**Fig.2** : Diagramme en disque de la répartition des types de locations AirBNB à Paris (en %)

Emprise : Paris

Nombre : 61 365 entités

Objectifs : Analyser et évaluer les systèmes énergétiques en Europe.

- Type de chambre : 4 type de chambre répartis selon :
  - ~84% de logements entiers
  - ~13% de chambres privées
  - < 2% de chambres d'hôtel
  - < 1% de chambres partagées

Avec des distribution de prix différentes.



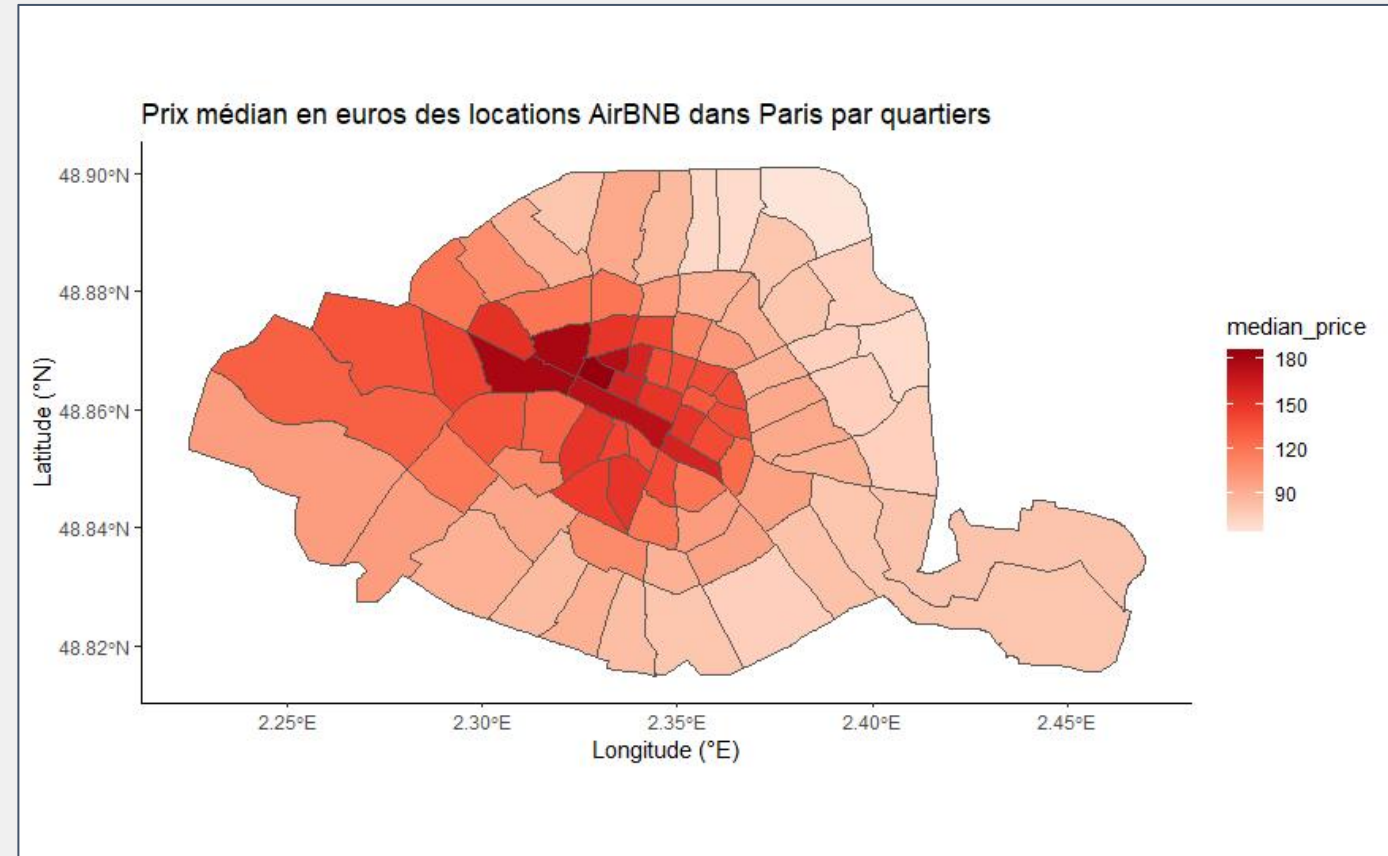
# 02 Analyse des prix

## OBJECTIFS

1. Détection de valeurs de prix de locations aberrantes (Outliers)
2. Histogramme et Distribution des prix de l'échantillon.
3. Définir des arrondissements attractifs grâce aux prix
4. Définir des quartiers attractifs grâce aux prix
- 5. Première ébauche de classification**

# Hypothèses

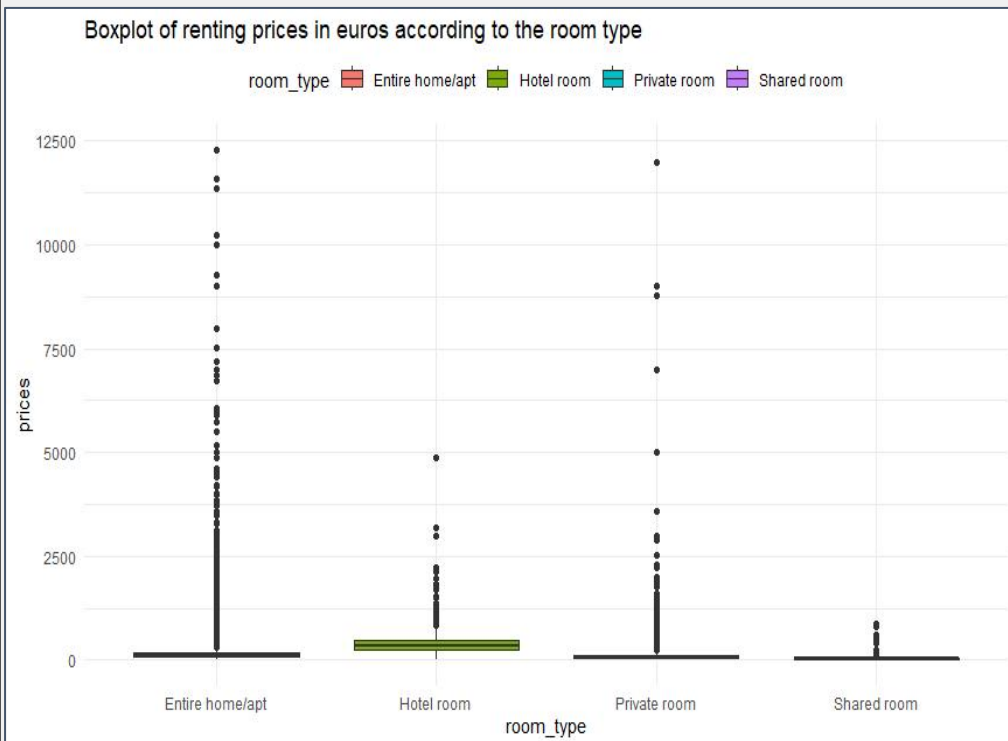
1. Les prix par zones reflètent une certaine attractivité touristique
2. La Distribution des prix varient selon l'arrondissement.
3. On peut définir des arrondissements attractifs grâce à des mesures de la dispersion des prix.



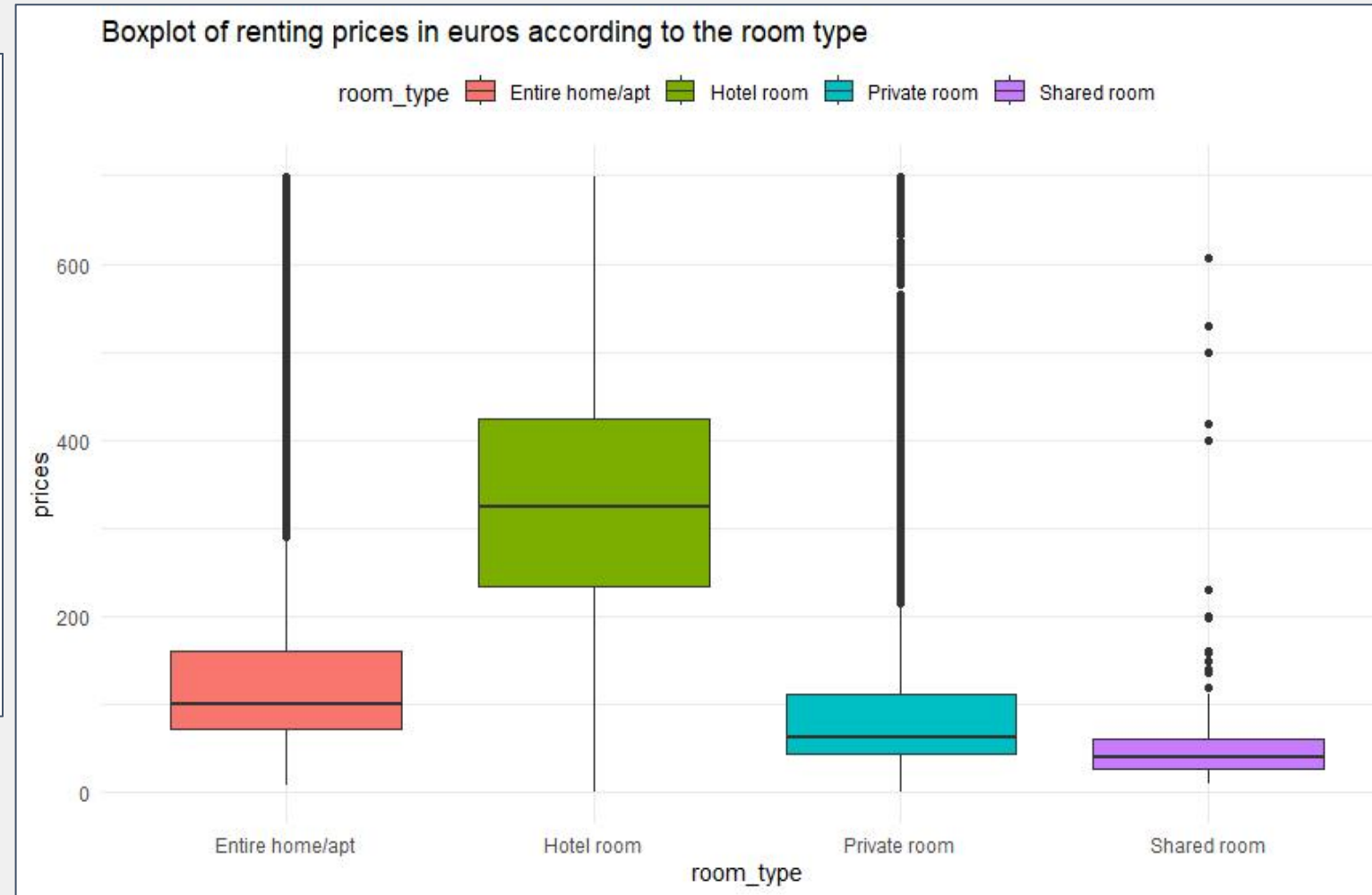
**Fig.3 :** Carte des prix médian en euros des AirBNB par quartiers parisiens

# Outliers

=> Prix / nuit > 500 euros + Hôtel et chambres partagées



**Fig.4 :** Histogramme des prix en euros des AirBNB par type de chambre



**Fig.5 :** Histogramme des prix médian en euros des AirBNB par type de chambre

## Distribution

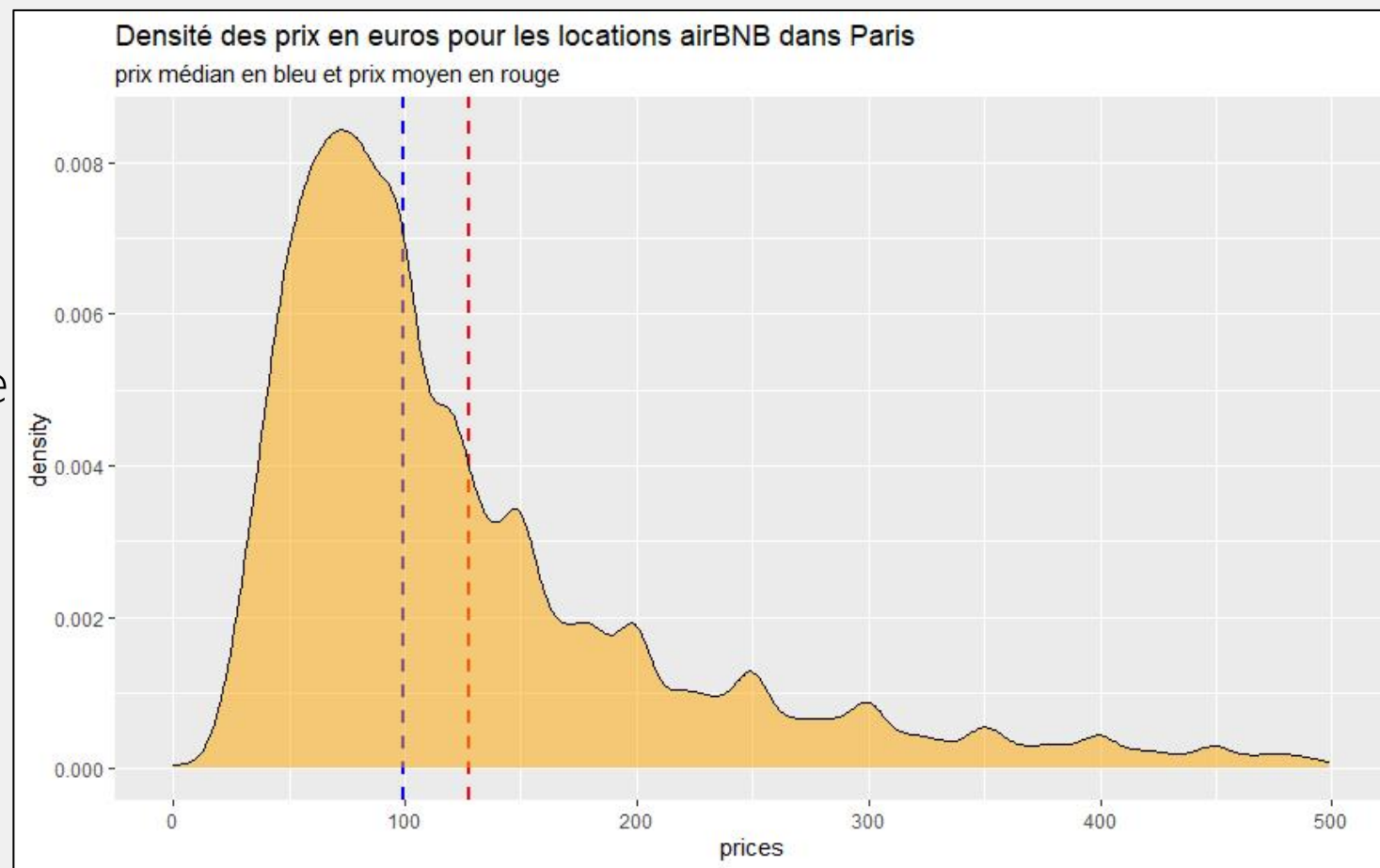
# ANALYSE DES PRIX

3 mesures de tendance centrale :

- mode principal : «pic» avant la médiane
- prix médian (ligne bleue) = 99 €
- prix moyen (ligne rouge) = 127 €

Distribution tirée par les prix élevés :

- Moyenne > Médiane



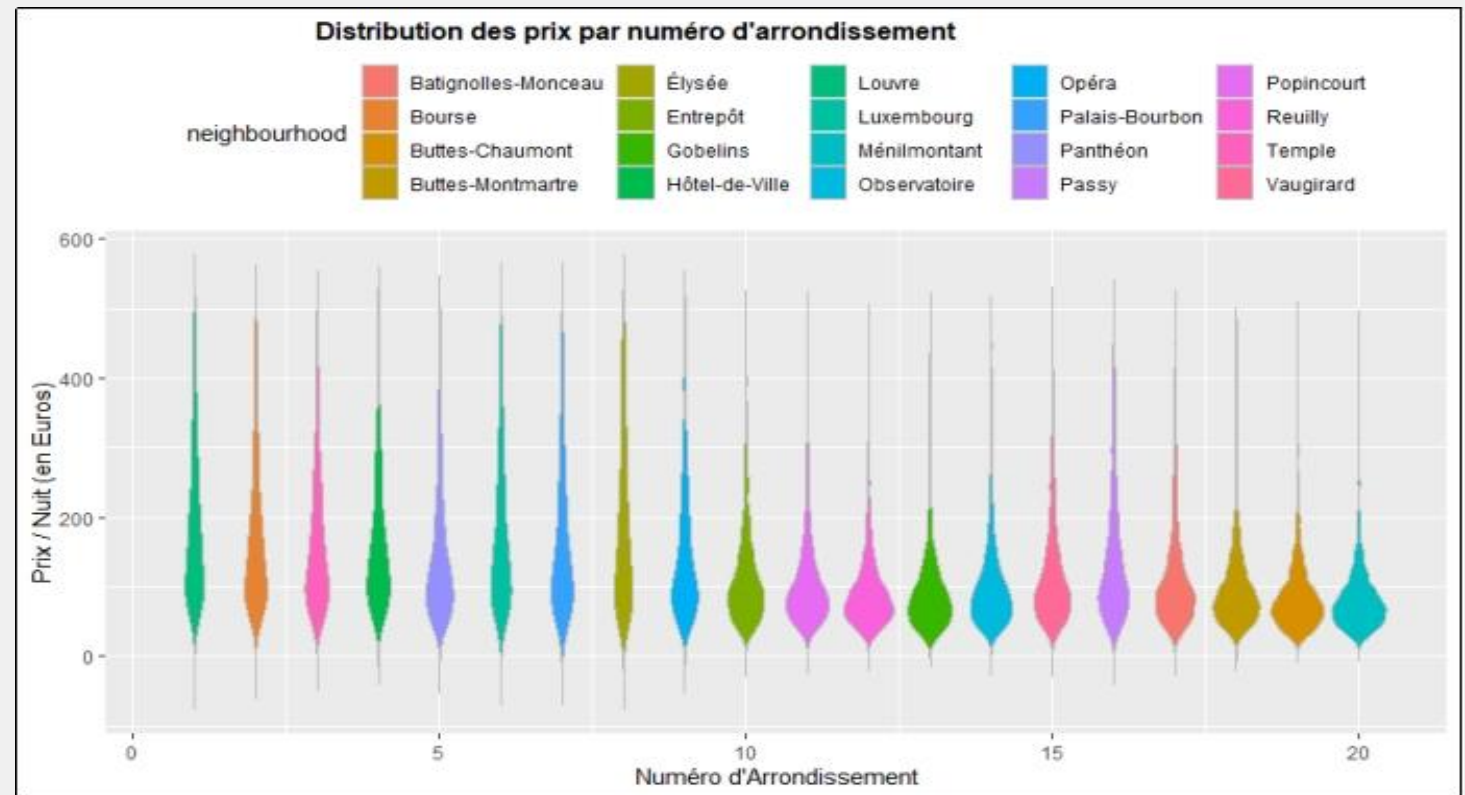
**Fig.6** : Distribution des prix/nuit (en €) des AirBNB dans Paris

## Distribution par arrondissement

# ANALYSE DES PRIX

On fait l'intersection entre la couche des arrondissements et celle des locations AirBNB :

- Visuellement on observe plusieurs profil de prix en fonction de l'arrondissement :
  - Les profils «mince» et allongé avec des prix plus élevé.
  - Les profils «large» et aplatie avec des prix moins élevé
  - Les profils «intermédiaires» situé à la croisée des deux profils précédent

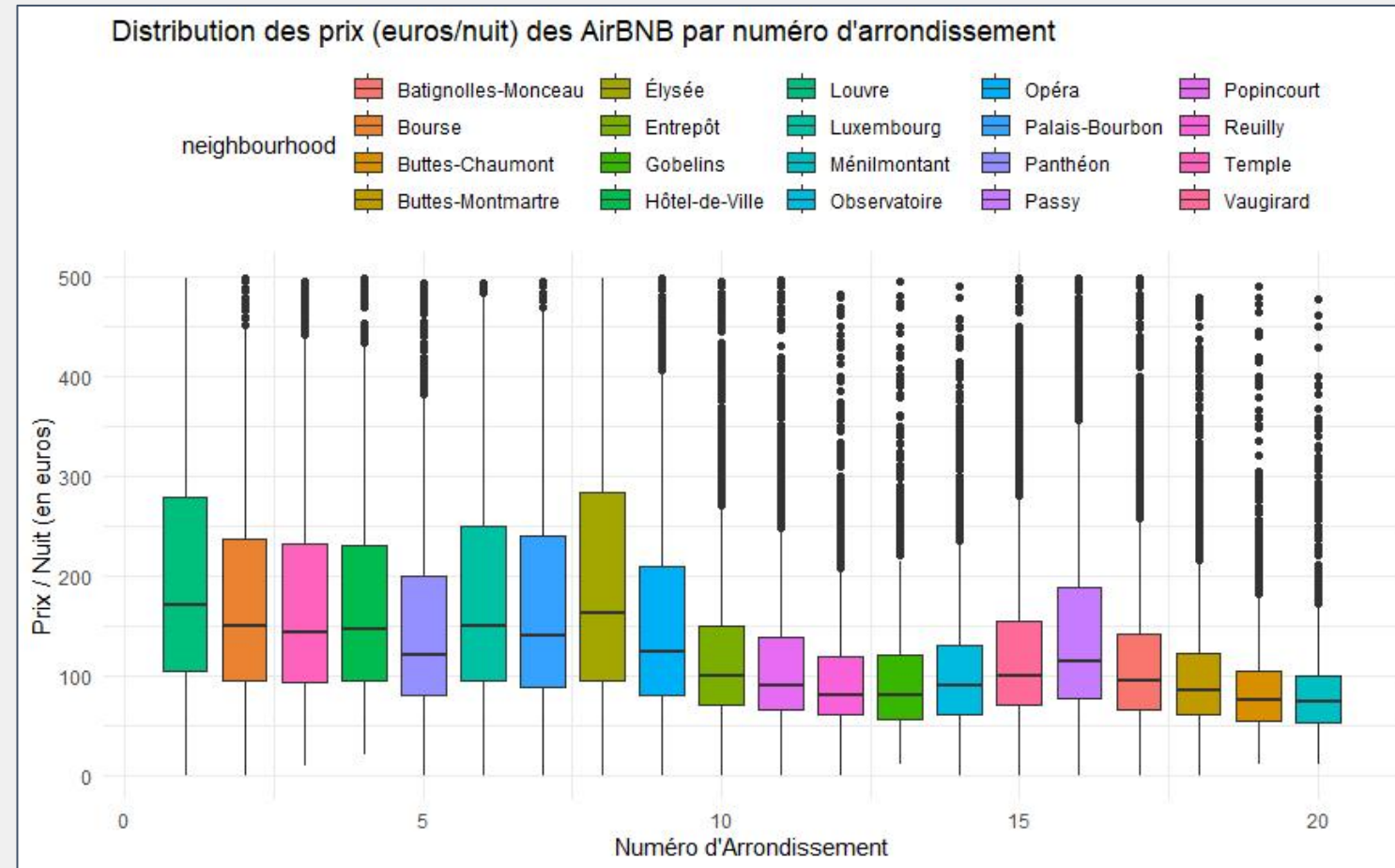


**Fig.7 :** Violin plot des prix/nuit en euros des locations AirBNB à Paris (dont les prix sont inférieurs à 500 euros)

## Distribution par arrondissement

# ANALYSE DES PRIX

- Visuellement on observe plusieurs profil de prix en fonction de l'arrondissement :
- Les profils «mince» et allongé avec des prix médian plus élevé et des écarts inter-quartiles fort.
- Les profils «large» et aplatie avec des prix médian plus faible et écart inter-quartiles faible
- Les profils «intermédiaires» situé à la croisée des deux profils précédent.



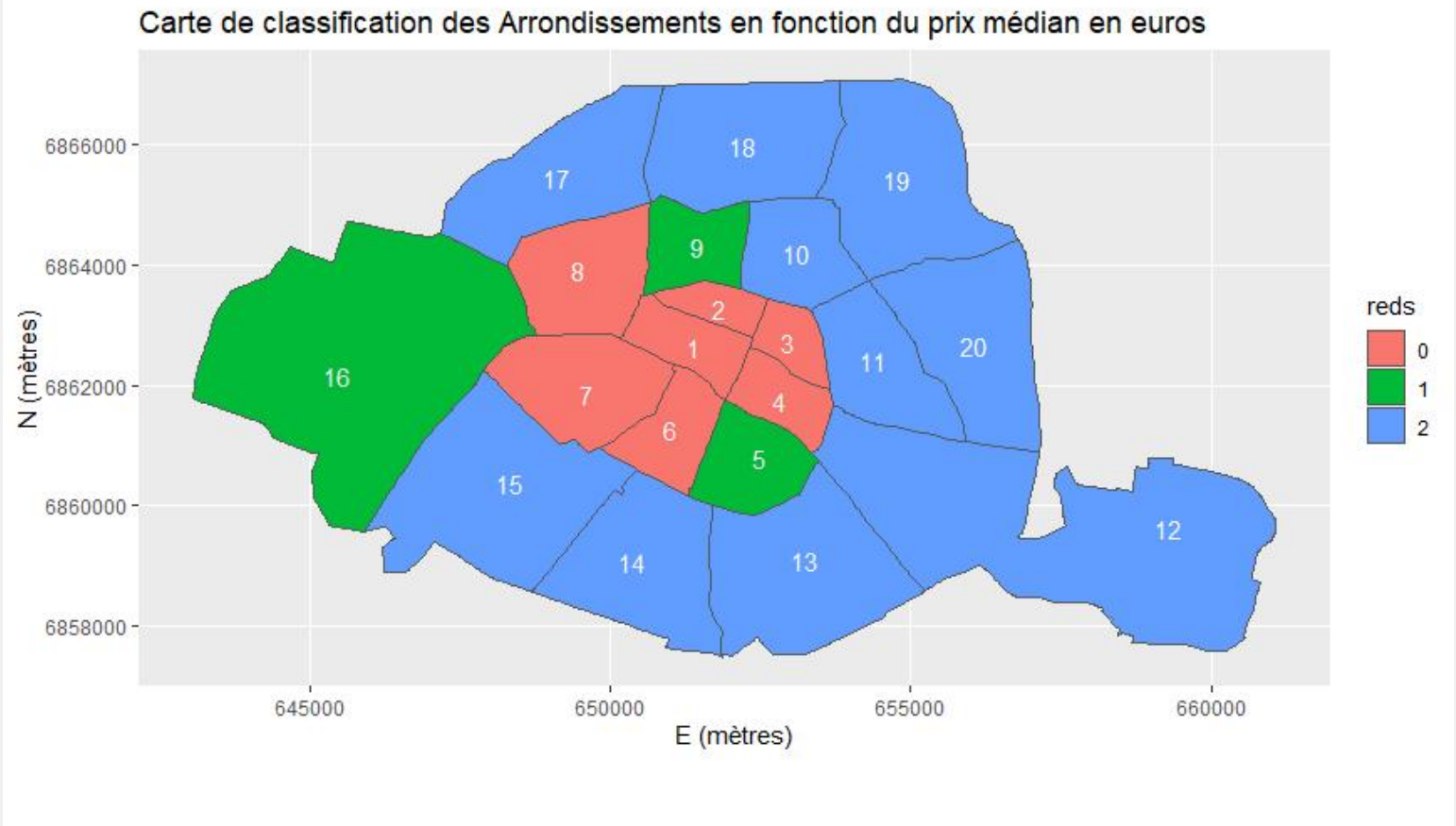
**Fig.8:** Boxplot des prix/nuit en euros des locations AirBNB à Paris (dont les prix sont inférieurs à 500 euros)



# Première Classification

## CLASSIFICATION

- *Classe : 0*; Arr. attractif
  - => prix médian > 130 euros
- *Classe : 2*; Arr. abordable
  - => prix médian < 110 euros
- *Classe : 1*; Arr. intermédiaire
  - => à la fois attractif et abordable



**Fig.9:** Carte des classes d'arrondissement à Paris en fonction des prix médians en euros

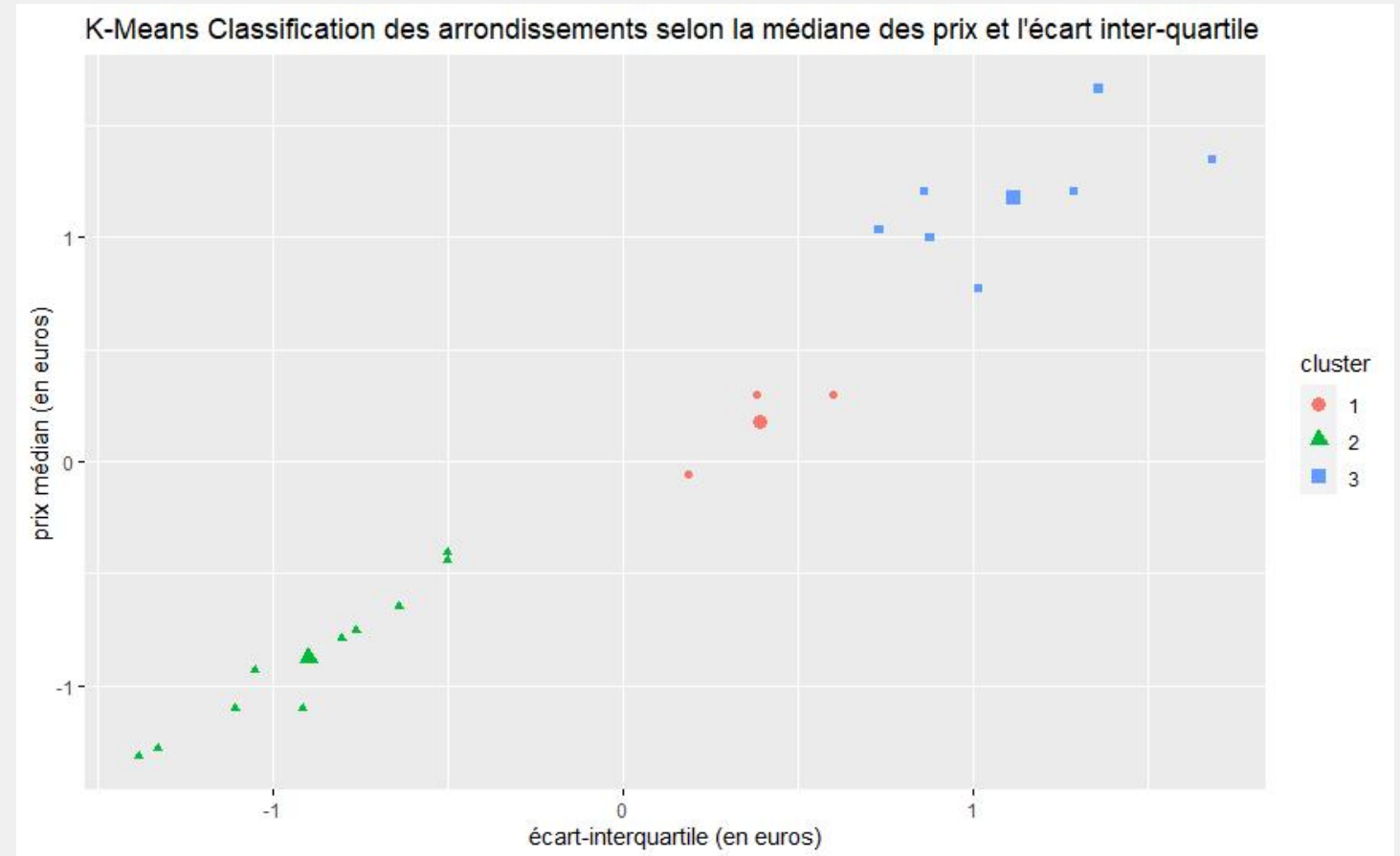


# Première Classification

## K-MEANS

- *Classe : 3*; Arr. attractif = {1,2,3,4,6,7,8}
- *Classe : 2*; Arr. abordable = {11,12,13,14,15,17,18,19,20}
- *Classe : 1*; Arr. intermédiaire = {5,9,15}

En utilisant un algorithme d'apprentissage supervisée K-Means ( $n = 3$ ) sur le prix médian et l'écart interquartile des arrondissement



**Fig.10:** Classification par K-Means des arrondissements à Paris en fonction des prix médians et des écart inter-quartile (normalisés)

# 03 Analyse de la répartition spatiale

## Hypothèses

- **On va désormais considérer la répartition spatiale des Airbnb**
- **On considère qu'il y a une corrélation spatiale entre la densité de gares de transport et de sites touristiques avec la densité de AirBNB dans Paris**
- **Cela nous permettras de définir des quartiers accessibles (transport) et attractifs (monuments)**

# Méthode

**1**

Pour chaque type de point  
Kernel Density Estimation

**4**

Comparaison avec une distribution de  
points aléatoire

**2**

Corrélation spatiale entre la  
densité de AirBNB et celle points  
d'intérêts

**5**

Validation de l'hypothèse

**3**

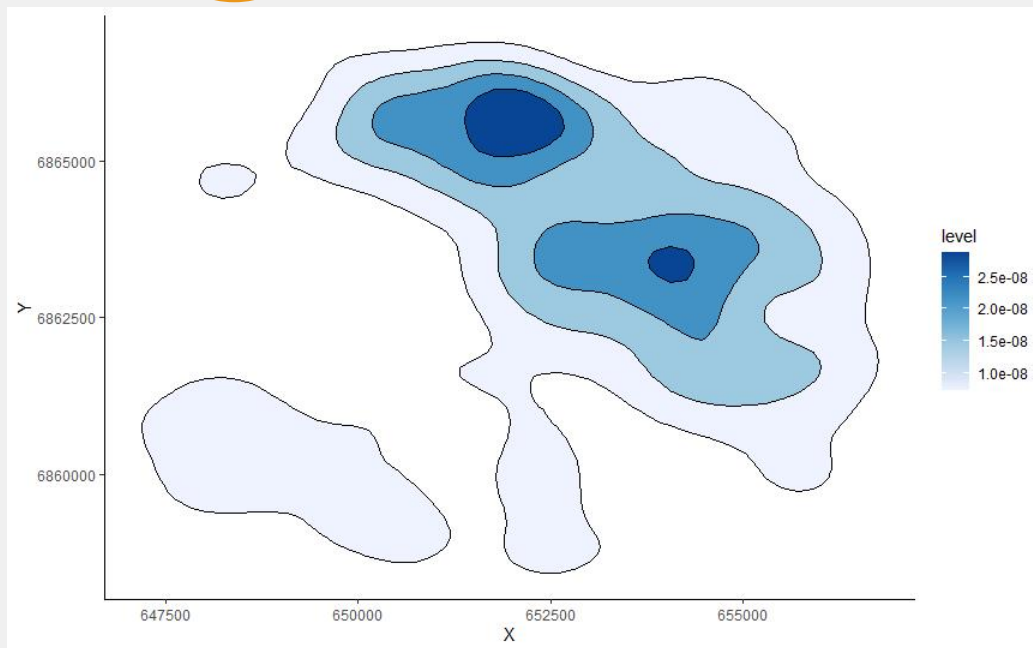
Bootstrap pour obtenir une incertitude  
sur ces corrélations

# Méthode

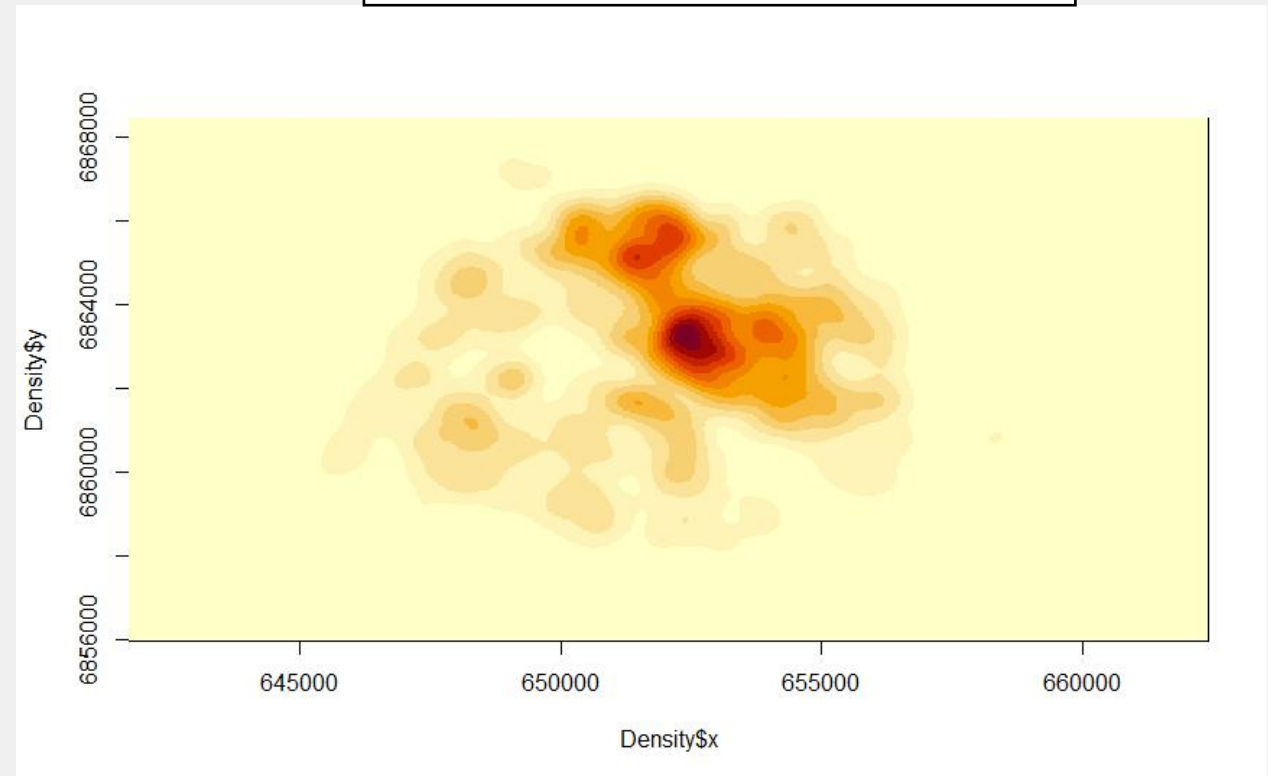
$h = 1050$  mètres ;  $n = 500$

1

Pour chaque type de point  
Kernel Density Estimation



**Fig.12:** Kernel Density Estimation des locations AirBNB à Paris

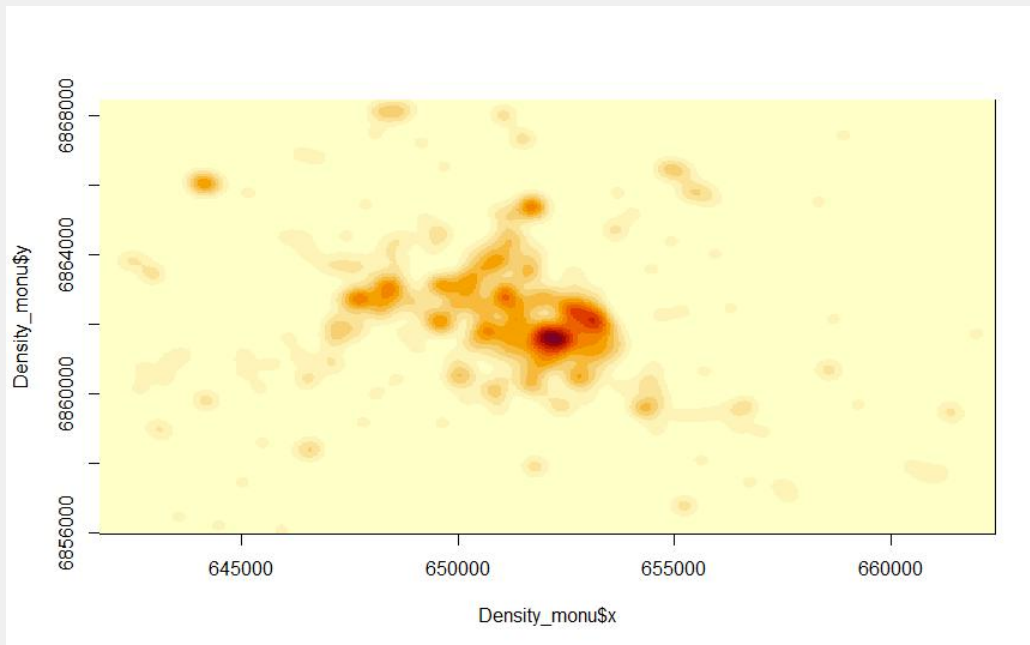


**Fig.11:** Kernel Density Estimation( $h = 1050, n = 500$ ) des locations AirBNB à Paris

# Méthode

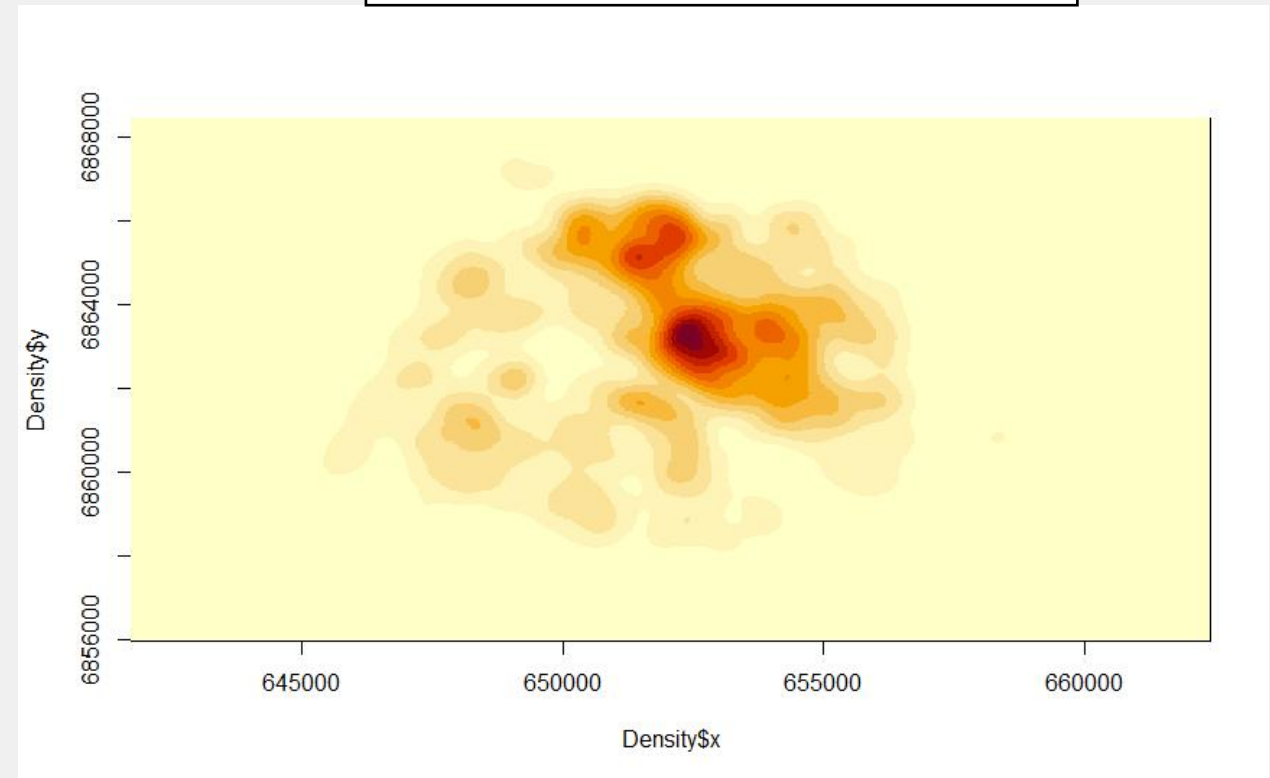
1

Pour chaque type de point  
Kernel Density Estimation



**Fig.13:** Kernel Density Estimation des sites touristiques à Paris

$h = 1050$  mètres ;  $n = 500$

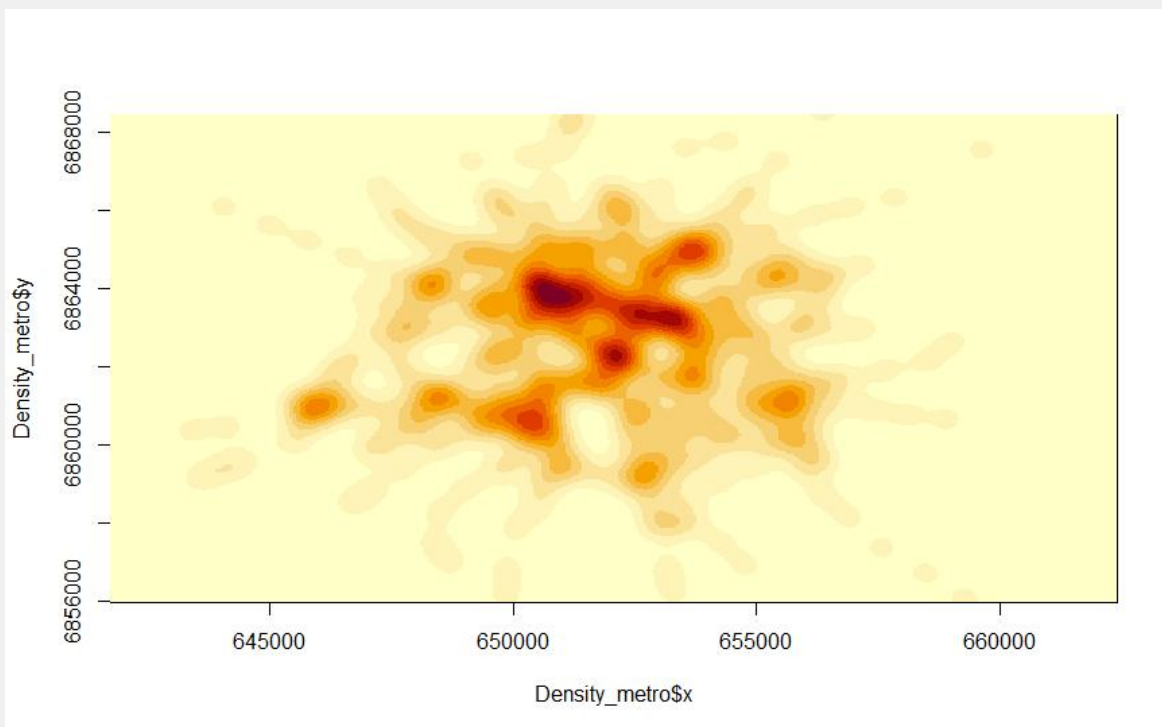


**Fig.11:** Kernel Density Estimation( $h = 1050, n = 500$ ) des locations Airbnb à Paris

# Méthode

1

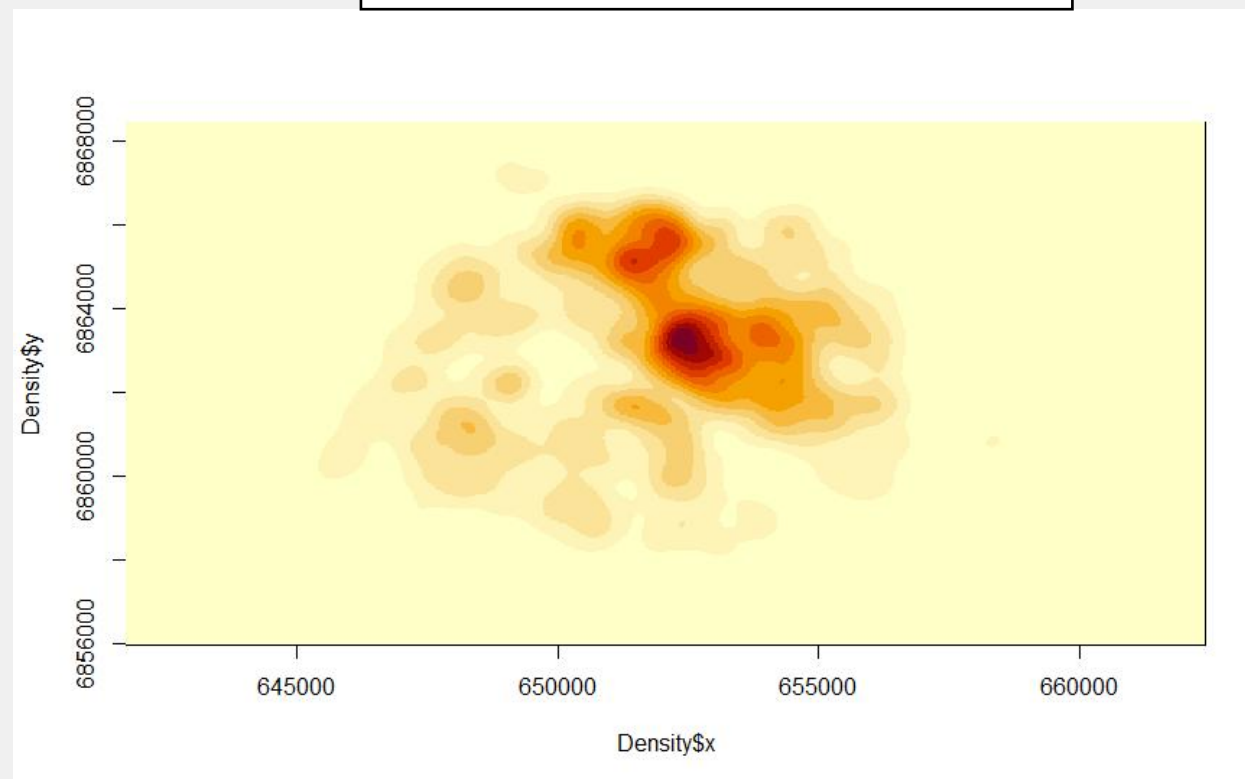
Pour chaque type de point  
Kernel Density Estimation



**Fig.14:** Kernel Density Estimation ( $h = 1300, n = 500$ )  
des stations de métro à Paris

Correlation AirBNB - métro  
[Totale] =  $(75.5 \pm 4)\%$

$h = 1050$  mètres ;  $n = 500$



**Fig.11:** Kernel Density Estimation( $h = 1050, n = 500$ ) des locations  
AirBNB à Paris

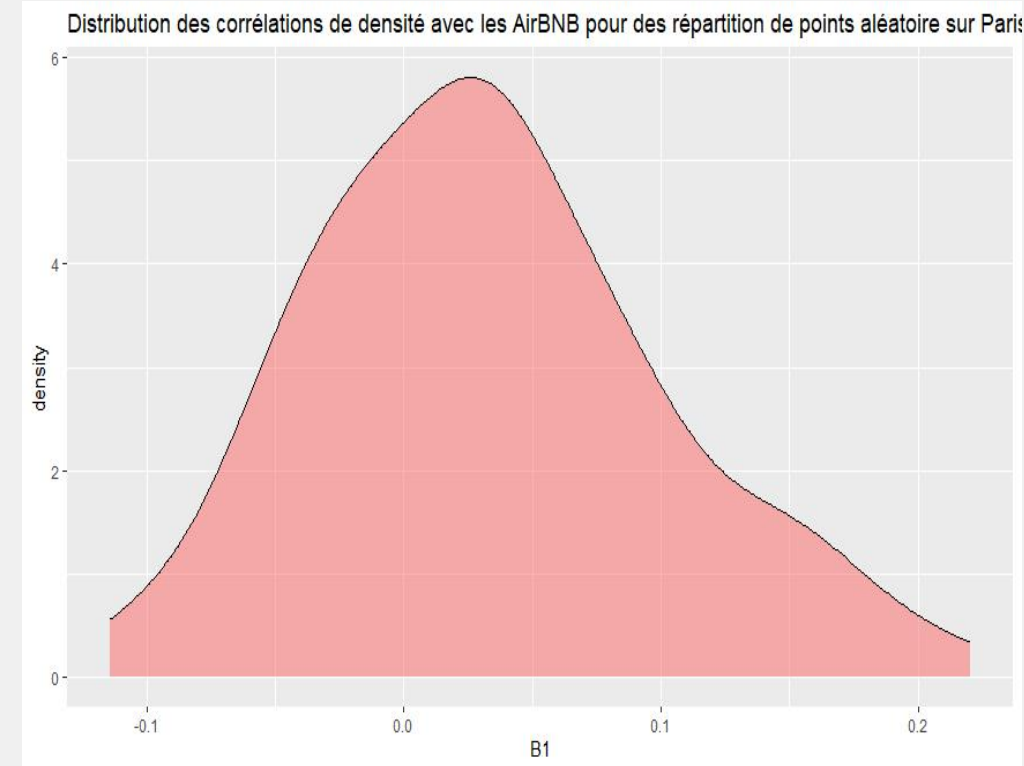
# Résultats

2

Corrélation spatiale réelle entre la densité de AirBNB et celle points d'intérêts

	Corrélation spatiale	Incertitude Bootstrap	Sup à 95% aléatoire
AirBNB - Métro	(75.5 +/- 4)%	0,04	Yes
AirBNB - Monu	(41.7 +/- 3)%	0,03	Yes
AirBNB - RER	(37.1 +/- 3)%	0,03	Yes

**Tab.1:** Résultats de la corrélation spatiale entre les densité de points



**Fig.16:** Distribution des corrélations spatiales de densité avec AirBNB pour des répartitions aléatoire



04

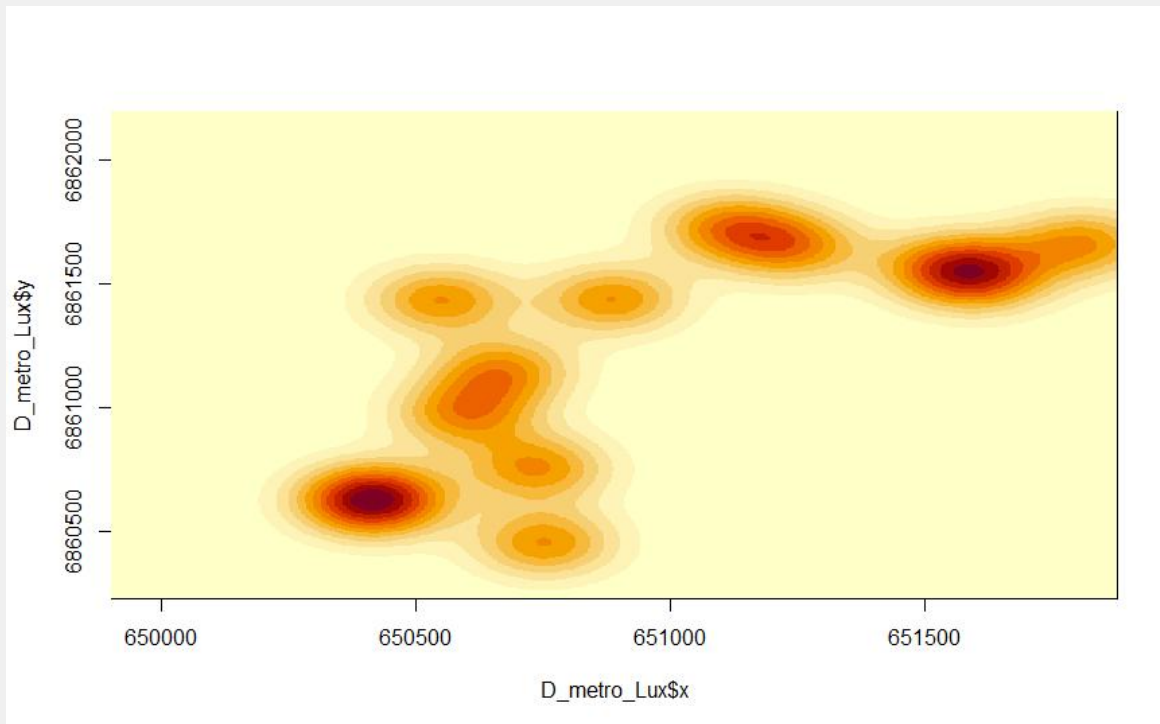
# Début de Classif intra-arrondissement

# Densité intra-arrondissement

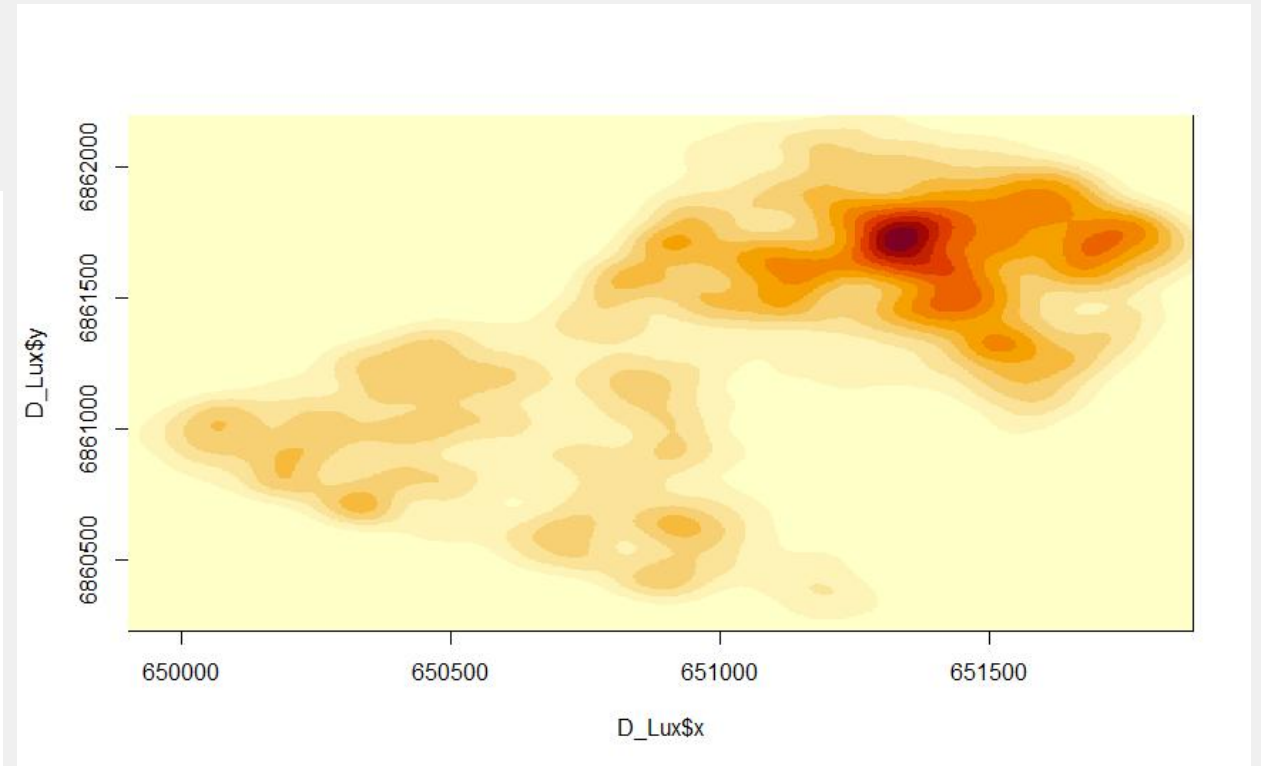
Correlation AirBNB - métro  
[Luxembourg] =  $(75.5 \pm 4)\%$

1

Pour chaque type de point  
Kernel Density Estimation



**Fig.17:** Kernel Density Estimation ( $h = 400, n = 500$ ) des stations de métro à Paris dans l'arrondissement du Luxembourg (6)

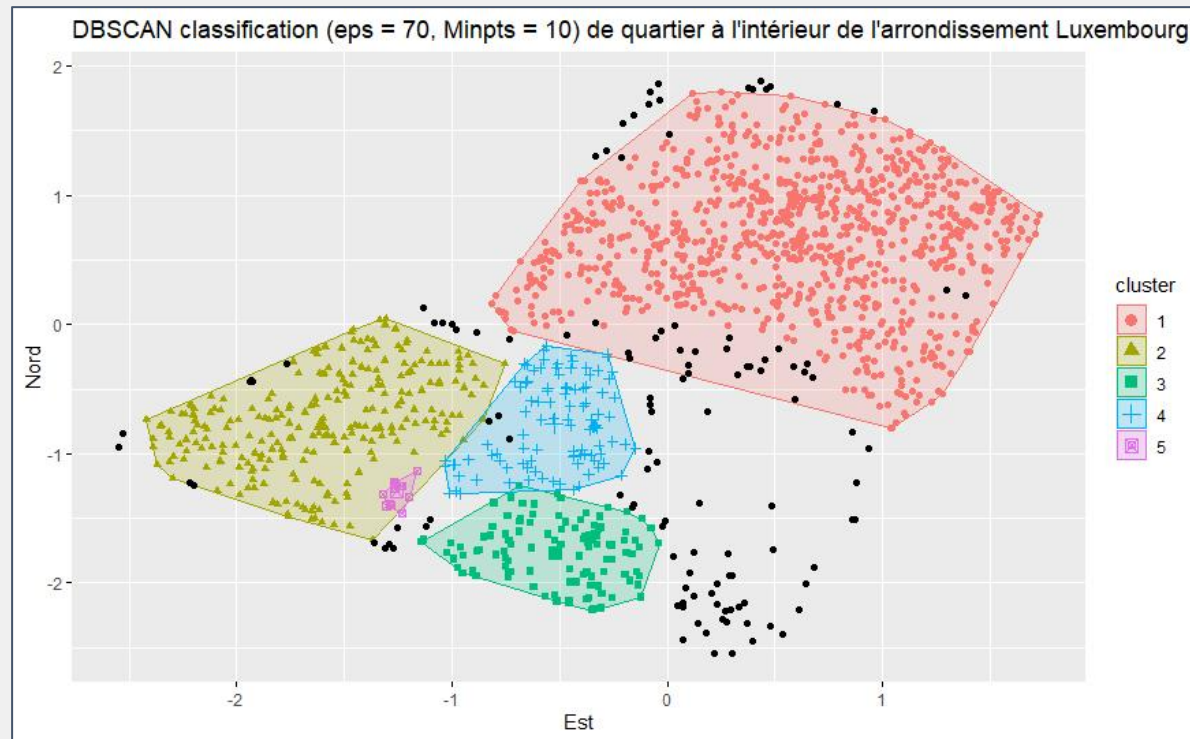


**Fig.15:** Kernel Density Estimation ( $h = 200, n = 500$ ) des locations AirBNB dans l'arrondissement du Luxembourg (6)

# DBSCAN Intra-Arr

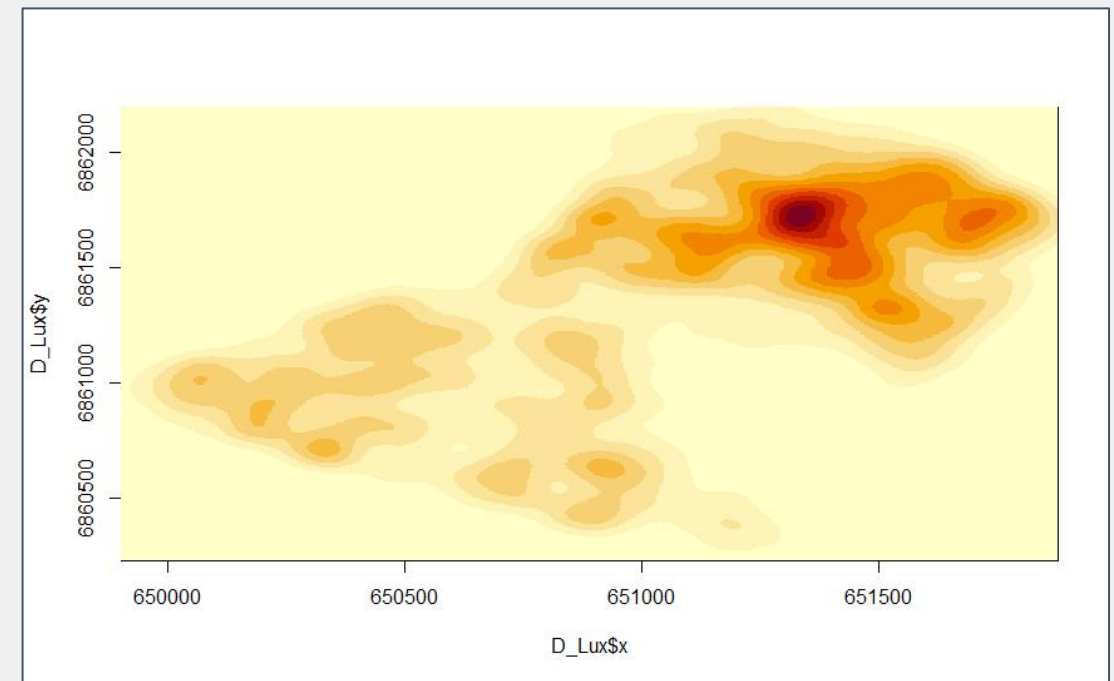
1

Par Arrondissement on applique un DBSCAN



**Fig.18:** Classification DBSCAN (eps= 70, Minpts= 10) dans l'arrondissement du Luxembourg

$h = 200$  mètres ;  $n = 500$



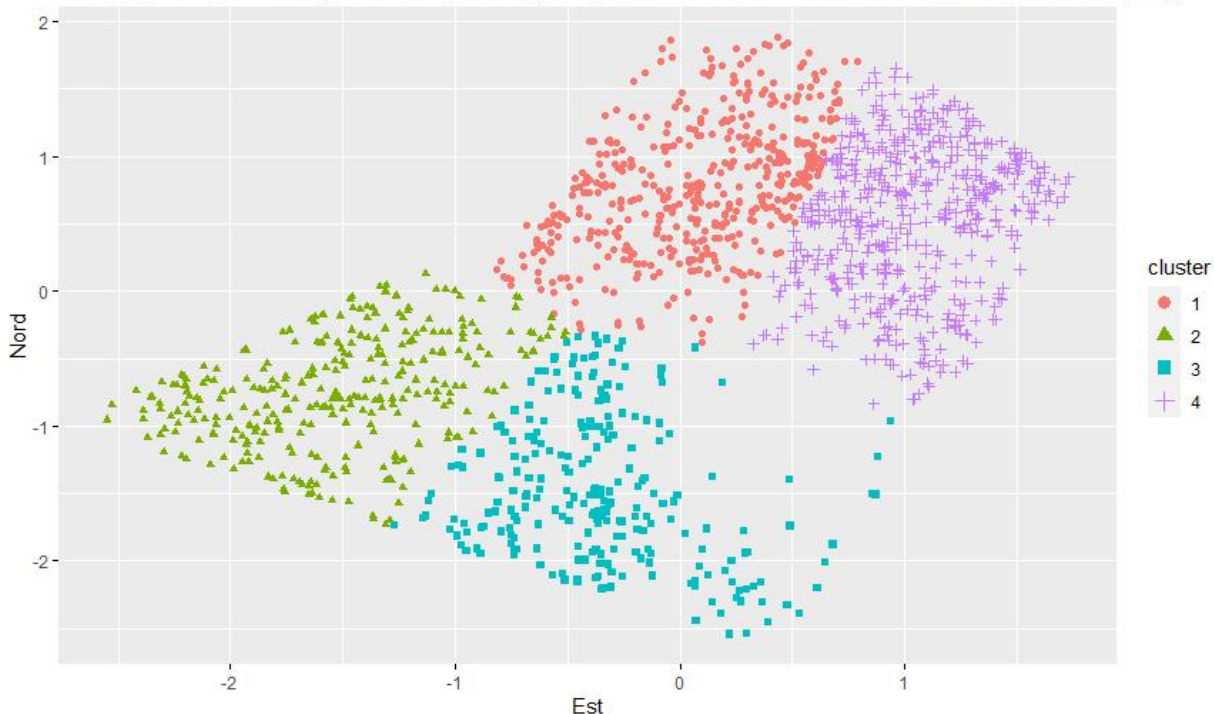
**Fig.15:** Kernel Density Estimation( $h = 200, n = 500$ ) des locations Airbnb dans l'arrondissement du Luxembourg (6)

## DBSCAN Intra-Arr

2

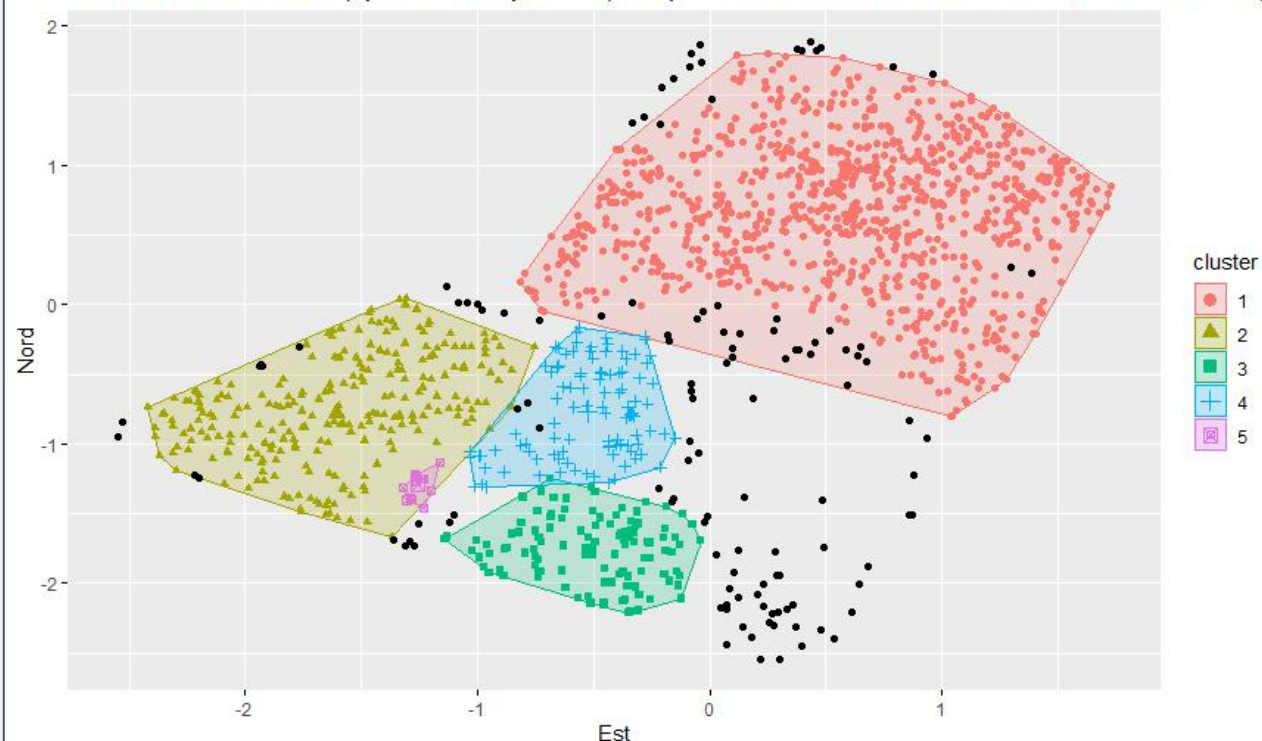
Par Arrondissement on applique un K-Means pour retrouver des quartiers.

K-Means classification (n = 4, Minpts = 10) de quartier à l'intérieur de l'arrondissement Luxembourg



**Fig.19:** Classification K-Means ( $k = 4$ ) dans l'arrondissement du Luxembourg

DBSCAN classification (eps = 70, Minpts = 10) de quartier à l'intérieur de l'arrondissement Luxembourg



# CONCLUSION

- Dernière partie à refaire pour tous les arrondissement !
- Comparaison quartiers de Densité et DBSCAN/KMeans
- Pondération par les prix
- Faire l'analyse des prix pour les quartiers + ACP
- Mieux définir les zones attractives et accessibles via les corrélations spatiales de densité dans les arrondissements => indices d'influences
- Exploiter les mesures du nuage de point dans chaque arrondissement.
- Indice d'autocorrélation spatiale : Indice de Moran/ Indice LISA

Merci pour votre  
**Attention !**

Avez vous des  
questions ?