

# Analyse d'un semis de points

*SPA1*

Marianne Guérois  
Souleymane Sidi Traoré

*Contributeur.ice.s : France, Emmanuel, Ouédraogo,  
Landry, Nicolas, Fatou, Charles, Claude, Mouftaou,  
Malika, Zacharyao, Hugues, Christine, Bénédicte*

ÉCOLE D'ÉTÉ INTERNATIONALE

## MÉTHODES ET OUTILS DES SCIENCES DES TERRITOIRES

UNE PERSPECTIVE NORD-SUD, SUD-NORD ET SUD-SUD

ÉTAPE 2 • IRSP, Ouidah (Bénin) 27 février - 10 mars 2023



## Introduction : un module d'initiation à l'analyse spatiale

1. Définition d'un semis de points
2. Résumer la distribution d'un semis de points
3. Concentration, dispersion : évaluer la forme globale d'un semis de points
4. Des points aux surfaces de densités : approche locale des concentrations via la méthode des noyaux (Kernels)

# Analyse Spatiale (AS), c'est quoi?

**L'analyse spatiale est..**

...l'étude des objets dans un espace géographique,



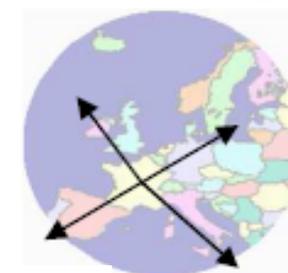
Une démarche « qui met l'accent sur l'identification de **formes ou de structures spatiales**, sur les **relations entre les objets géographiques**, sur les **processus de changement** » (Feuillet et al., 2019).

..en analysant leur distribution..



..leur voisinage...

...dans un champ spatial...



...à une résolution donnée.





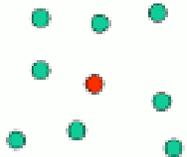
AS pour quoi faire?



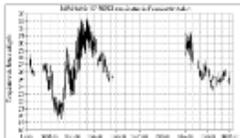
mieux connaître l'organisation des milieux dans l'espace



comprendre le fonctionnement spatial des milieux

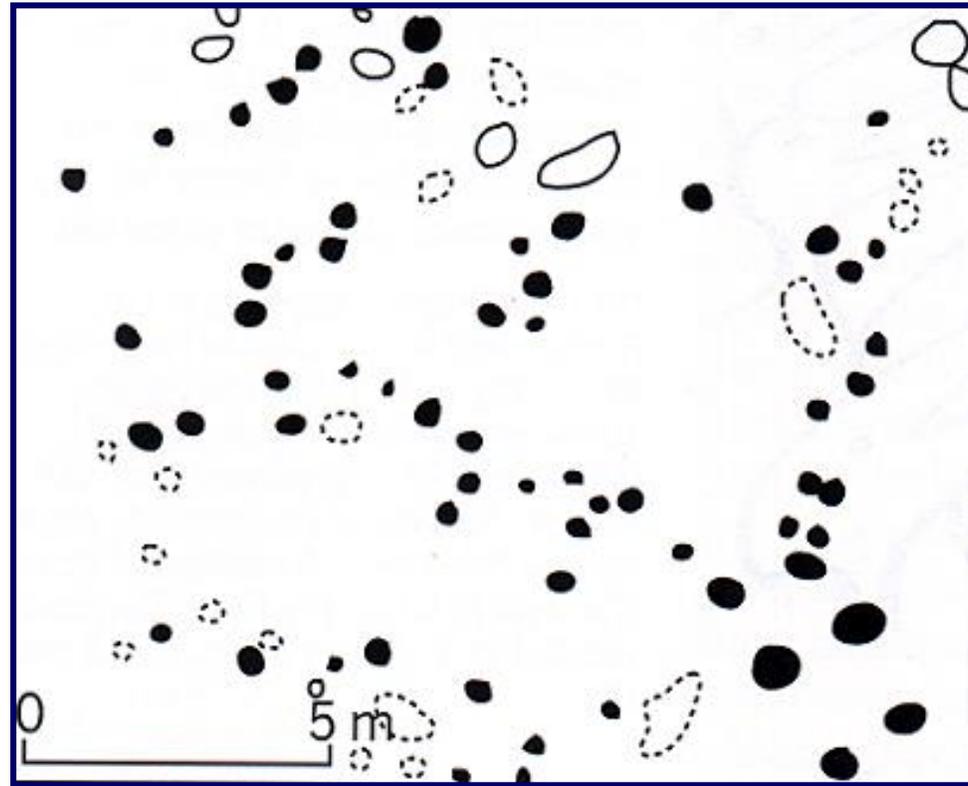


prédire des informations en un point non mesuré



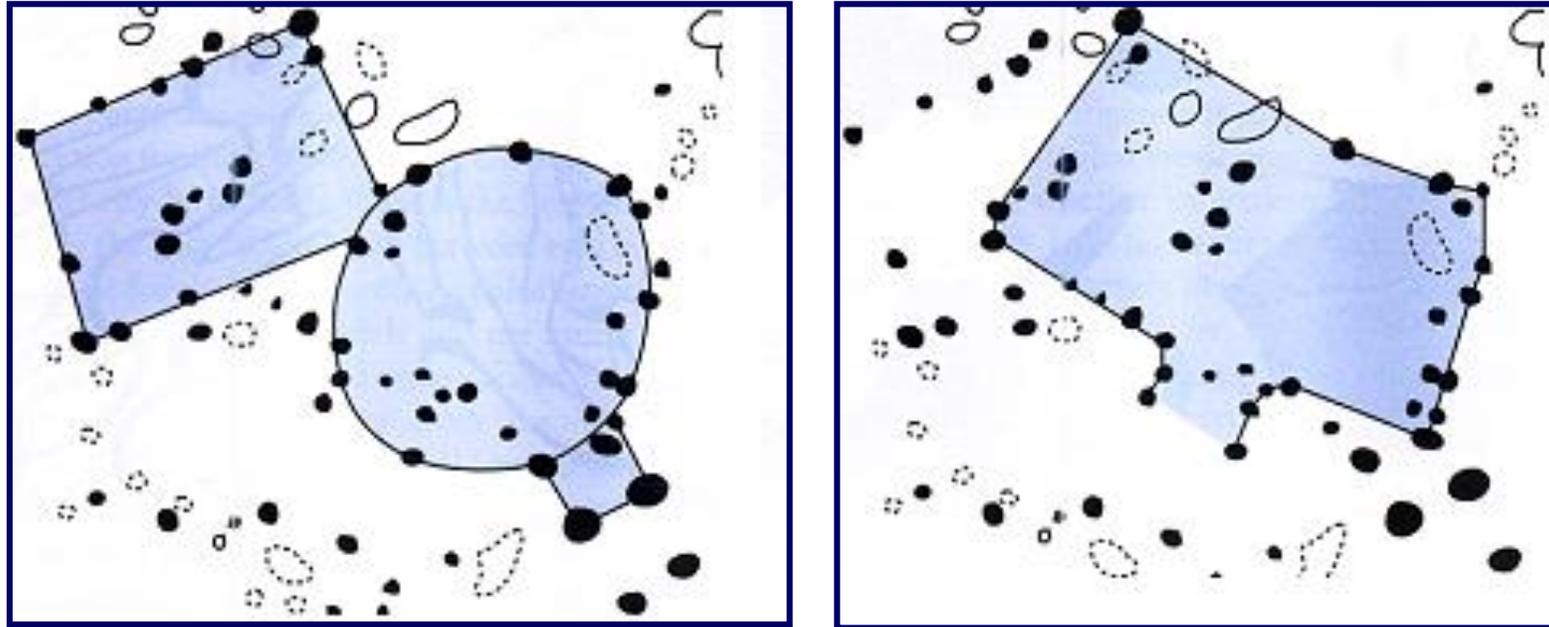
modéliser le fonctionnement spatial des milieux

A l'œil nu, quels schémas d'organisation émergent d'une distribution de lieux?



Source : Gregory R.H, and Gombrich E.H., 1973, cité par Haggett P. (2001), *Geography : a global synthesis*, Routledge, p.28.

A l'œil nu, quels schémas d'organisation émergent ?  
Comment formaliser ces observations ?



Source : Gregory R.H, and Gombrich E.H., 1973, cité par Haggett P. (2001), *Geography : a global synthesis*, Routledge, p.28.

## Objectifs du module

- **Un module d'initiation à l'AS** : examiner des méthodes qui peuvent être mises en œuvre **pour décrire un ensemble d'objets ou d'événements** distribués dans l'espace, lorsque la distribution de ces objets ou de ces événements peut être **assimilée à un semis de points**
- Répondre à des **questions** sur l'étude d'un semis de points dans le temps et dans l'espace :
  - Comment **résumer la répartition** des points ?
  - Comment **suivre dans le temps** la localisation de phénomènes ?
  - Comment identifier des **types d'organisation spatiale** (dispersées/concentrées) ?
  - Comment localiser des **zones de forte concentration locale** ?

## Compétences

- Savoir **résumer** la distribution spatiale d'un ensemble de points à l'aide d'**indicateurs élémentaires** (*points centraux, indicateurs de dispersion*)
- Connaître une des méthodes classiques d'**analyse de la forme plus ou moins concentrée ou dispersée** d'un ensemble de points (*méthode de la distance au plus proche voisin*)
- Savoir transformer une distribution de points en une **surface de densités**, pour identifier des **zones locales de concentration**, à l'aide d'une **méthode de lissage cartographique** (*KDE - Kernels Estimation Density*)

## Pré-requis et liens avec autres modules de l'EE

- **Information géographique** : localisation, systèmes de coordonnées
- **Statistiques univariées** : valeurs centrales et paramètres de dispersion d'une distribution statistique
- **Analyse spatiale** : définitions et mesures de distances

# Bases de données utilisées pour les exercices

## Africapolis, sur les agglomérations urbaines

(<https://africapolis.org/>) :

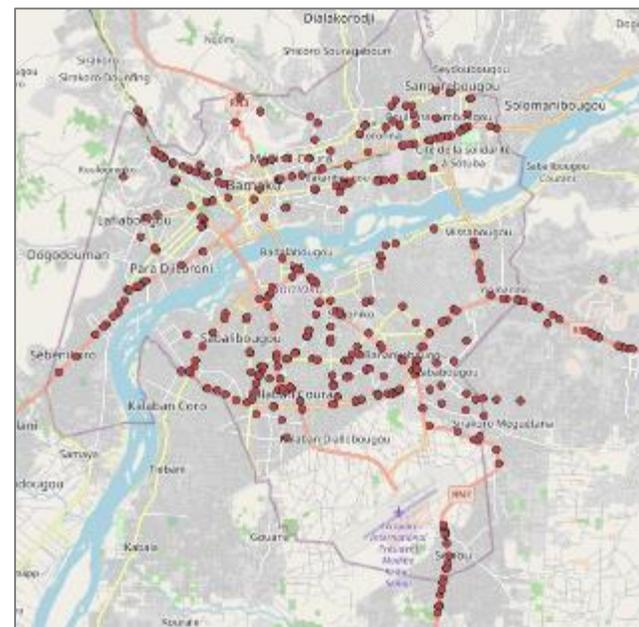
- Extraction pour 7 pays d'Afrique occidentale francophone
- Attributs: Coordonnées X,Y du centre géométrique de l'agglomération, population de 1950 à 2015



Source : Africapolis, 2020. Fond OSM.

## Stations services de Bamako :

Enquête. S.S. Traoré (février 2020)



Source : Traoré, 2020. Fond OSM.

Stations services à Bamako  
Date de l'enquête: 14 au 29 février 2020  
Auteur: SSTRACORE, sstraore@yahoo.fr

-----  
Système de ref. UTM 29 N, wgs84

-----  
Support de collecte de données

- smartphone avec application Kobbo collect

### Table d'attribut

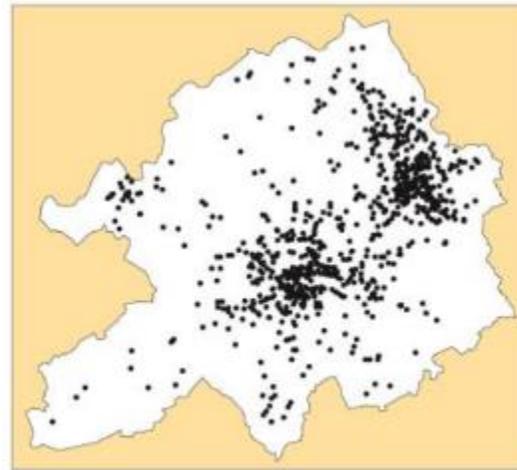
- \* Latitude / longitude: coordonnées géographiques de l'infrastructure en UTM 29 Nord, wgs 84
- \* Nom : Nom de l'infrastructure
- \* Type: type de l'infrastructure ( Station services, point de vente, dépôt de carburant)
- \* Fonctional: fonctionnalité de l'infrastructure au moment des enquêtes
- \* Rang: Nombre d'équipements publics se trouvant dans un rayon de 100m de l'infrastructure (marchés, restaurants, boutiques, écoles...)

(c) sstraore, 2020

# 1. Définition d'un semis de points

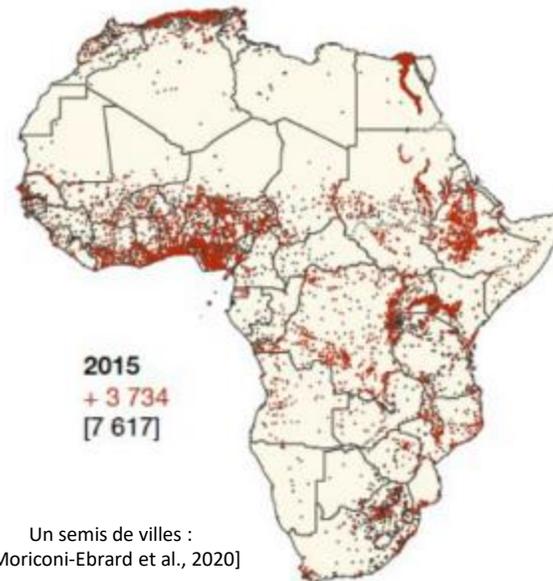
# 1. Définition d'un semis de points

**Semis de points** = ensemble de lieux pouvant être traités comme des points, à un certain degré de généralisation (habitats, équipements, arbres, événements, villes, ...)



Population de référence :  
 l'ensemble des accidents en 1997

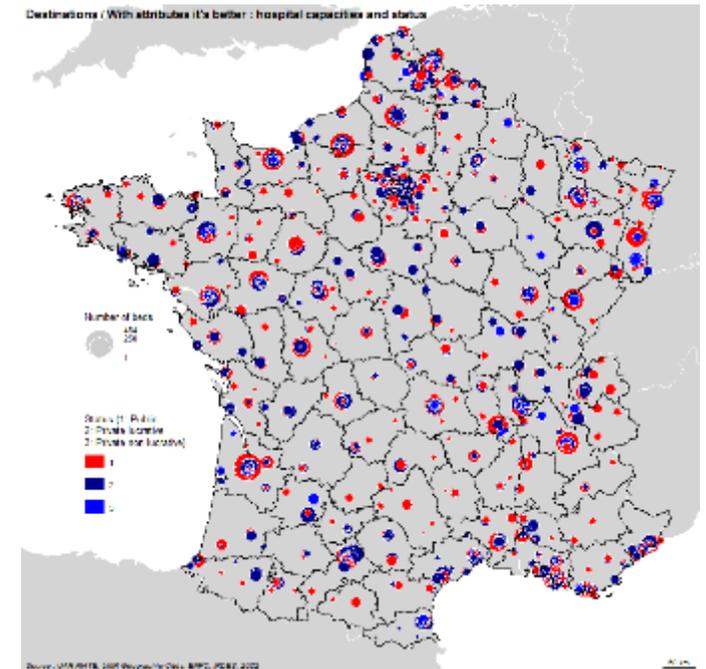
[Huguenin-Richard, 2010]



Un semis de villes :  
 [Moriconi-Ebrard et al., 2020]



[<https://tree-map.nycgovparks.org/>]



[Ysebaert, Baudet-Michel, Conti, 2022]

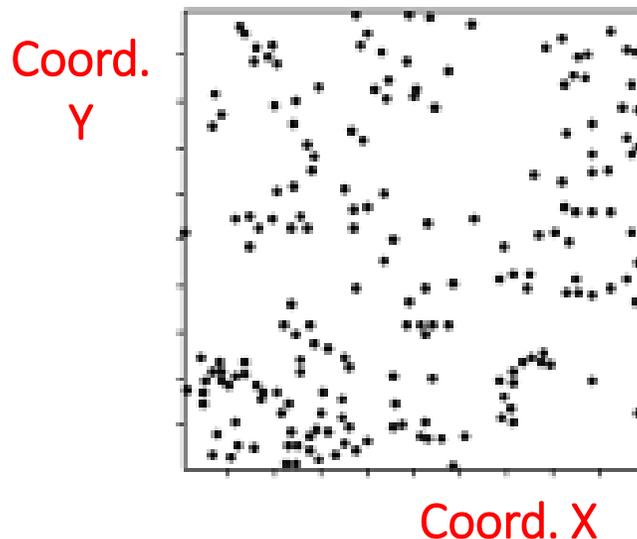
# 1. Définition d'un semis de points

Structure de l'information géographique associée :

ensemble de  $n$  points ( $1...i...n$ ) décrits par leurs coordonnées de position ( $X_i, Y_i$ ) et souvent munis d'attributs qualitatifs ou quantitatifs (par exemple taille des arbres, de la population, ...).

## Semis élémentaire

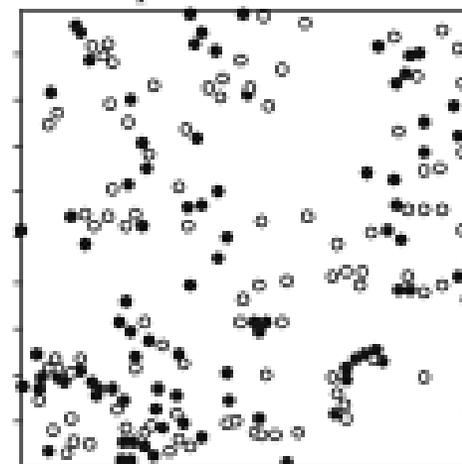
(a) Univariate point pattern



CODE	X	Y
1	250	428
2	323	427
3	435	624
4	872	125
...	...	...
100	326	453

## Semis muni d'attributs qualitatifs

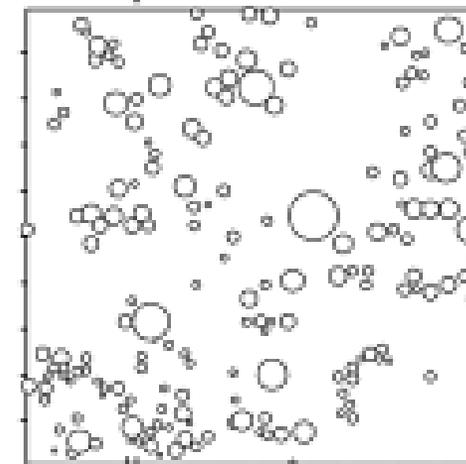
(e) Univariate point pattern w/qualitative marks



CODE	X	Y	Type
1	250	428	A
2	323	427	A
3	435	624	B
4	872	125	A
...	...	...	...
100	326	453	B

## Semis muni d'attributs quantitatifs

(f) Univariate point pattern w/quantitative marks



CODE	X	Y	Taille
1	250	428	10
2	323	427	20
3	435	624	105
4	872	125	40
...	...	...	...
100	326	453	100

# 1. Définition d'un semis de points

## Assimiler des lieux à des points : une opération parfois élémentaire... ou plus ambiguë

Les agglomérations Géopolis (**implantation surfacique**), entre Ouidah et Cotonou (Bénin)



Sources : Africapolis, 2020 ; OpenStreetMap 2023

Les agglomérations Géopolis (**implantation ponctuelle**), entre Ouidah et Cotonou (Bénin)



Quel centre pour l'agglomération de Cotonou-Abomey-Calvi ?

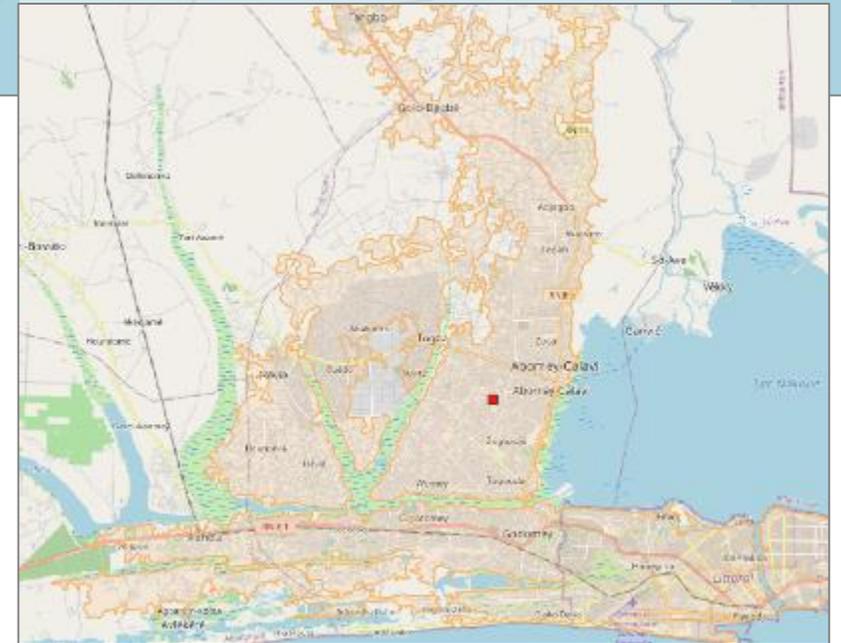
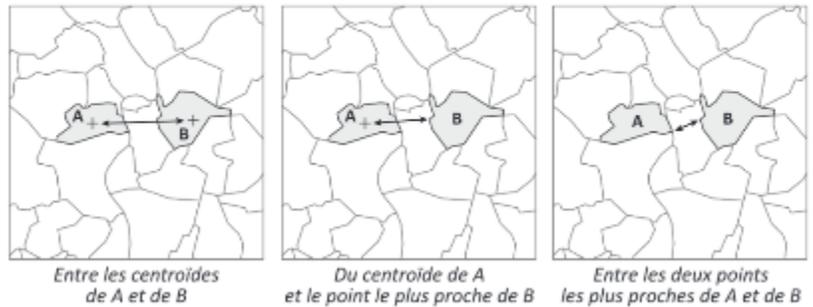


Figure 2.21 : Trois différentes façons de considérer la distance entre deux entités de géométrie surfaciques

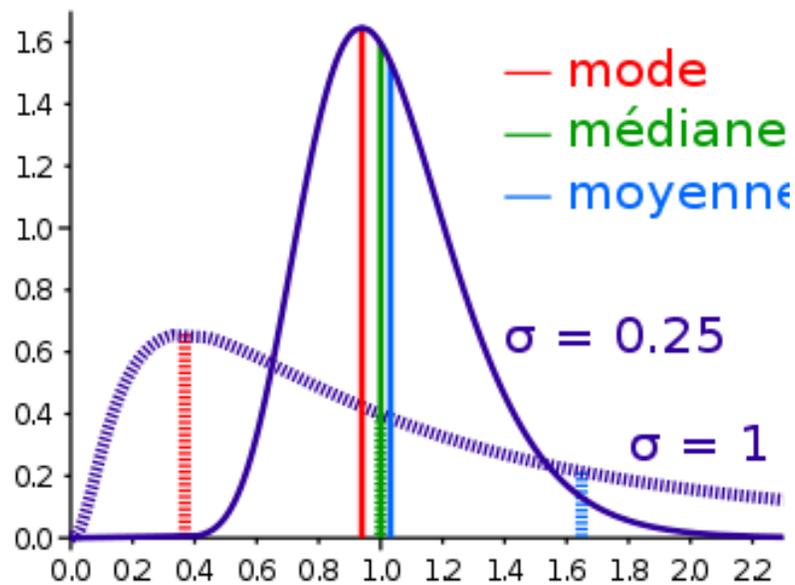


Source : Aschan-Leygonie C., Cuntly C., Davoine P.A., 2019, *Les systèmes d'information géographique*, Cours, Armand Colin, p.78

## 2. Résumer la distribution d'un semis de points

## 2. Résumer la distribution d'un semis de points

En **statistiques univariées**... résumé d'une variable quantitative continue :



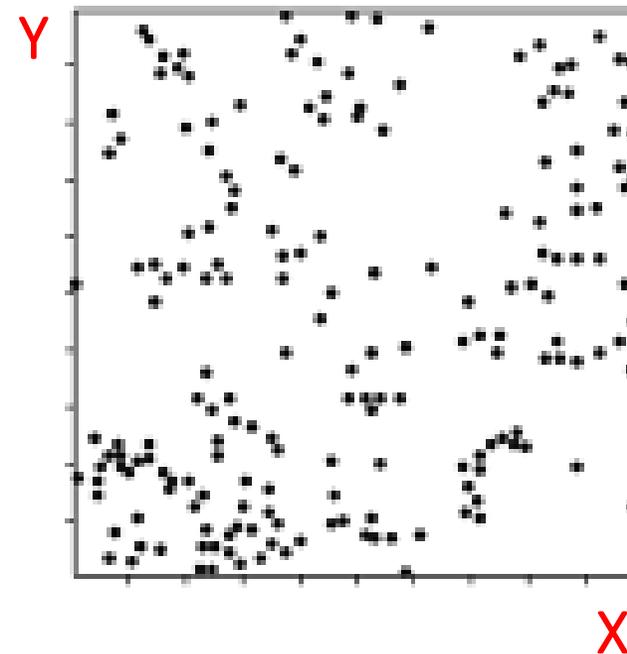
Source : Wikipedia

Et quand ces variables sont des **coordonnées spatiales** X et Y :

quelles valeurs centrales ?

quelle dispersion autour de ce centre ?

(a) Univariate point pattern



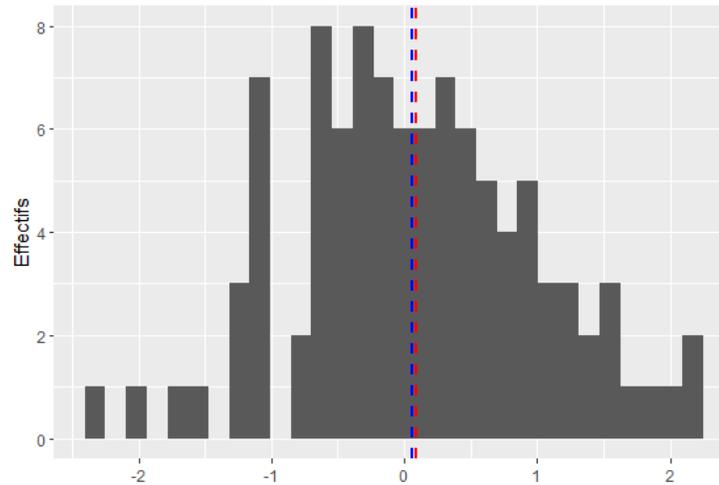
## 2. Résumer la distribution d'un semis de points

### 2.1. Points centraux (point moyen, point médian)

#### Distribution **statistique** : 1 variable (x)

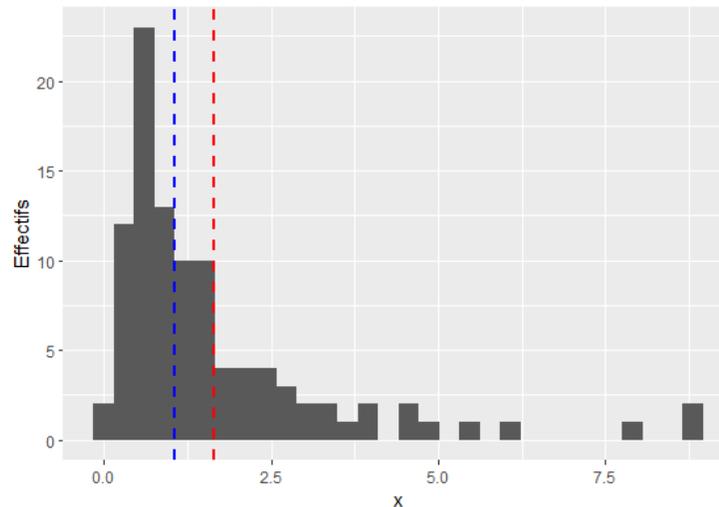
Distribution symétrique

Valeurs centrales : moyenne (rouge) et médiane (bleu)



Distribution dissymétrique

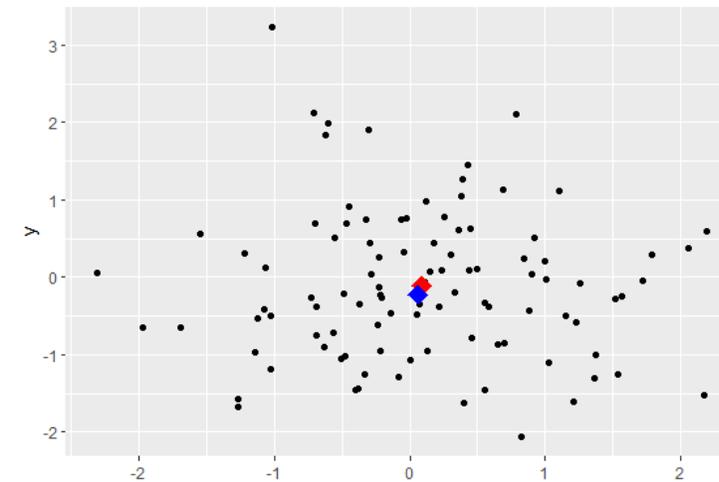
Valeurs centrales : moyenne (rouge) et médiane (bleu)



#### Distribution **spatiale** : 2 variables = coord (x,y)

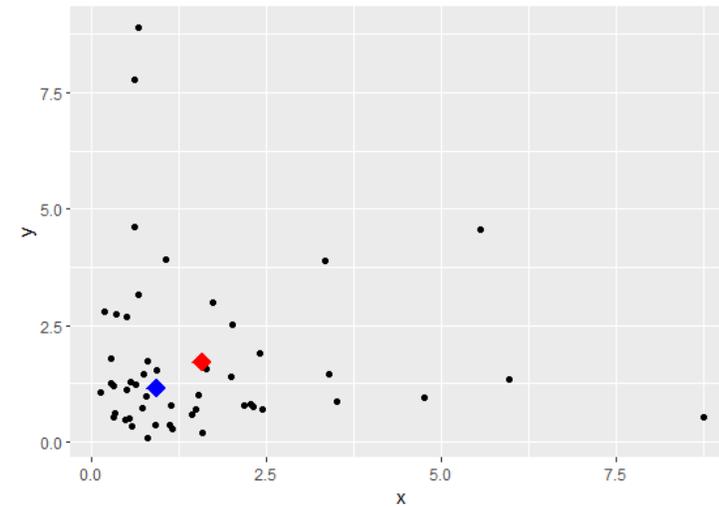
Semis de points symétrique

Points moyen (rouge) et médian (bleu)



Semis de points dissymétrique

Points moyen (rouge) et médian (bleu)



Symétrique

Dissymétrique

## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

### 2.1. Points centraux : quel est le point qui résume le mieux la position de l'ensemble des points du semis ?

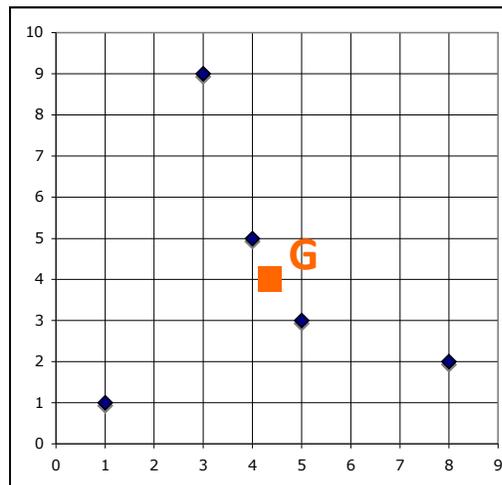
**Point moyen : simple**/pondéré

Point médian : simple/pondéré

**Centre de gravité** (noté **G**), barycentre

Pour un ensemble de **n points** (1...i...n) décrits par leurs coordonnées (**X<sub>i</sub>**, **Y<sub>i</sub>**), c'est le point qui a pour coordonnées **mX**, **mY** = moyenne des distributions des **x<sub>i</sub>** et des **y<sub>i</sub>** de l'ensemble du semis

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{Y} = \frac{\sum_{i=1}^n y_i}{n}$$



i	Coordonnées	
	X	Y
A	1	1
B	5	3
C	3	9
D	8	2
E	4	5
Total	21	20
Moyenne	4,2	4

## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

### 2.1. Points centraux : quel est le point qui résume le mieux la position de l'ensemble du semis ?

**Point moyen : simple / pondéré**

Point médian : simple / pondéré

- **Propriétés**

Comme la moyenne, c'est le point qui **minimise la somme des carrés des distances** à tous les autres points

- **Applications : utile pour...**

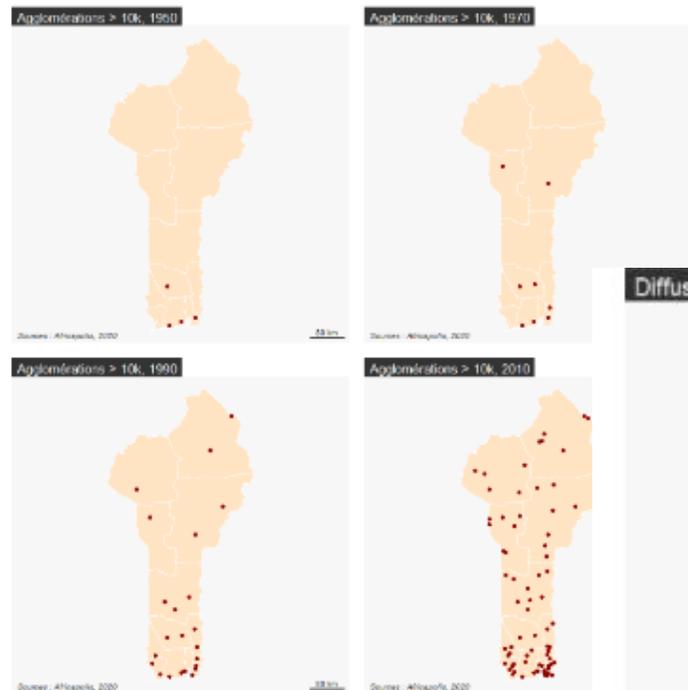
→ ... Suivre des évolutions d'une distribution à **différentes dates** (indique orientation et rythme d'une diffusion spatiale)

→ ... Comparer les distributions spatiales de **différentes catégories de points** à une même date (par ex. plusieurs essences végétales, catégories d'équipements...)

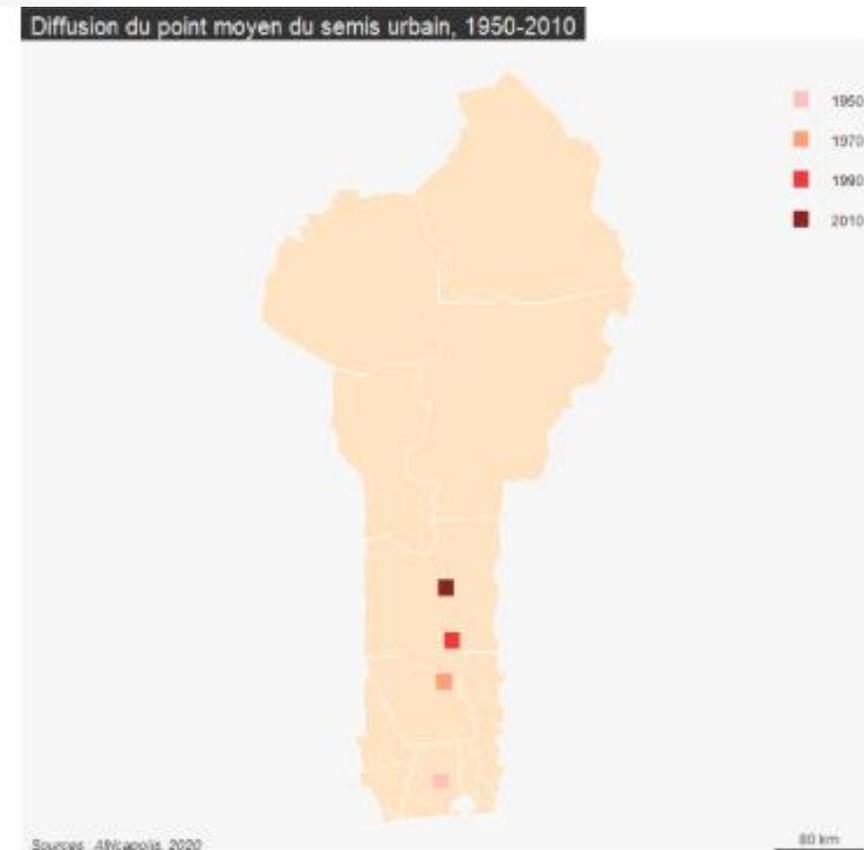
*Ex : Un analyste des infractions peut souhaiter observer si le centre moyen des cambriolages se déplace lors de l'évaluation des incidents journaliers ou nocturnes. Cela peut aider les services de police dans l'allocation des ressources.*

- **Limites :**

Comme la moyenne, **sensible aux positions extrêmes** (et aux masses extrêmes, si pondéré)



Positions du point moyen du semis d'agglomérations urbaines au Bénin (1950-2010) → cf. **Exercice 1**



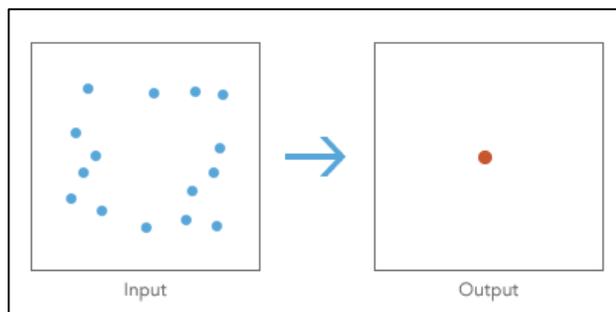
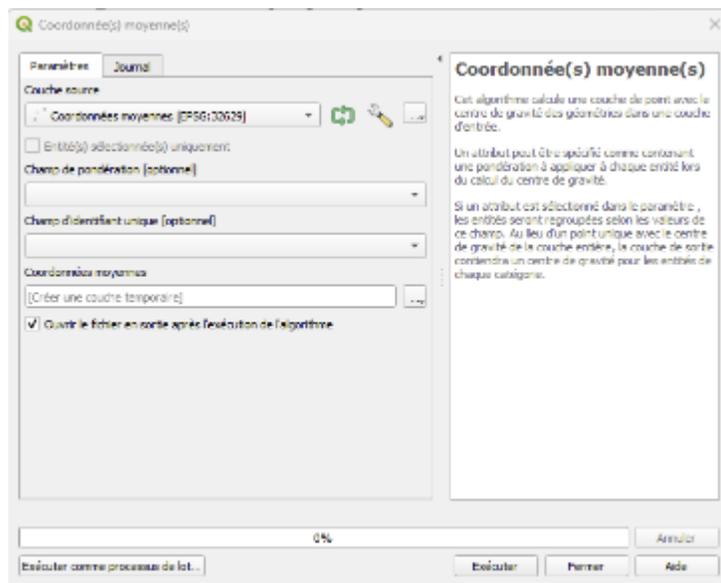
## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

Point moyen : simple/pondéré

Point médian : simple/pondéré

### Avec QGIS

Menu Vecteur – Outils d'analyse – Coordonnées moyennes



### Avec R

Fonctions de *R-base* et de *sf*

- `st_coordinates` (`data$geometry`) : pour extraire les coordonnées X, Y des points
- `mean` (X), `mean` (Y) : pour en calculer la valeur moyenne  $X_m$  et  $Y_m$
- `st_point` (`c(Xm, Ym)`) : pour transformer ces deux coordonnées en point

#### # Extraction des coordonnées des villes

```
coordsPays10 <- data.frame(Nom = Payspolis2010$Nom,  
                           X = st_coordinates(Payspolis2010$geometry)[, 1],  
                           Y = st_coordinates(Payspolis2010$geometry)[, 2],  
                           stringsAsFactors = FALSE)
```

#### # Calcul moyenne de ces coordonnées en X et Y

#### # puis transformation en point

```
ptmoy_Pays10 <- st_point(c(mean(coordsPays10[, 2]), mean(coordsPays10[, 3])))
```

## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

### 2.1. Points centraux : quel est le point qui résume le mieux la position de l'ensemble du semis ?

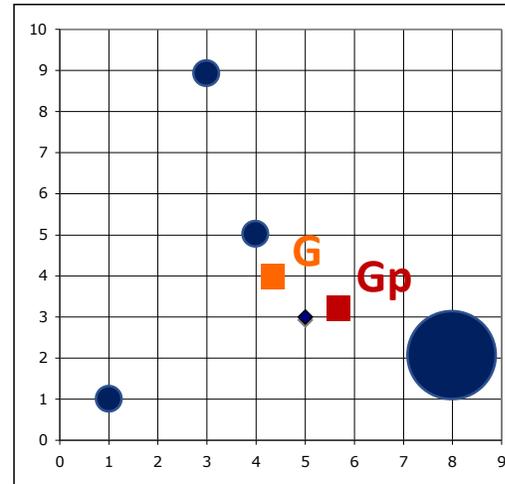
**Point moyen** : simple/**pondéré**

Point médian : simple/pondéré

**Centre de gravité pondéré** (noté **Gp**), barycentre des poids localisés (population, taille, capacité d'accueil...)

Pour un ensemble de **n points** (1...i...n) décrits par leurs coordonnées (**X<sub>i</sub>**, **Y<sub>i</sub>**) et par leur **masse m<sub>i</sub>**, c'est le point qui a pour coordonnées **mX<sub>p</sub>**, **mY<sub>p</sub>** = moyenne des distributions des **x<sub>i</sub>** et des **y<sub>i</sub>** de l'ensemble du semis pondérée par la masse

$$\bar{X}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \quad \bar{Y}_w = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

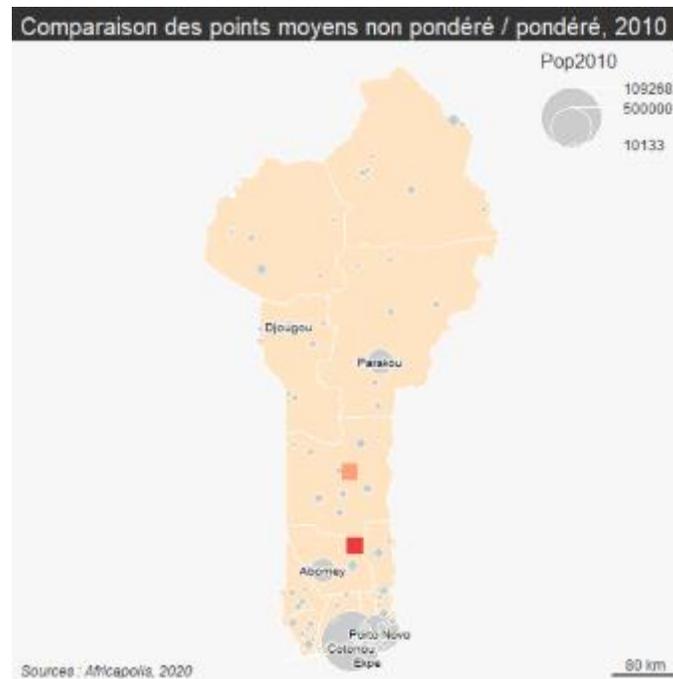


i	Coordonnées		Population	Coordonnées pondérées	
	X	Y		X pondéré	Y pondéré
A	1	1	2 000	2 000	2 000
B	5	3	1 000	5 000	3 000
C	3	9	2 000	6 000	18 000
D	8	2	8 000	64 000	16 000
E	4	5	2 000	8 000	10 000
Total	21	20	15 000	85 000	49 000
Moyenne	4,2	4		5,7	3,3

## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

**Point moyen** : simple/**pondéré**

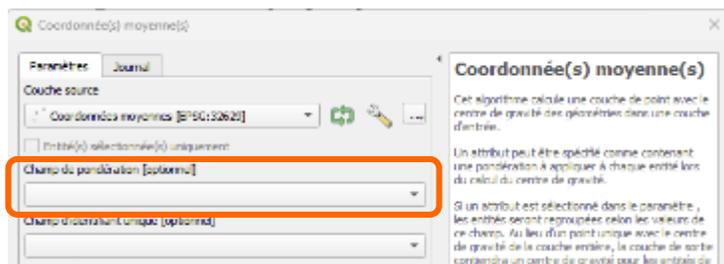
Point médian : simple/pondéré



Fonctions de R-base : `sum ( )`, `sf : st_point ( )`, `dplyr : left_join ( )`, `mutate ( )`

Menu Vecteur – Outils d’analyse – Coordonnées moyennes –  
 Champ de pondération

Avec  
 QGIS



Avec  
 R

```
# On ajoute une colonne population au tableau des coordonnées
coordsPayspop10 <- left_join(coordsPays10, Payspolis, by = "Nom", keep = FALSE) %>% select(Nom, X, Y, Pop2010)

# On calcule les coordonnées pondérées par les populations pour chaque ville
coordsPayspop10 <- coordsPayspop10 %>% mutate(Xmp = X * Pop2010, Ymp = Y * Pop2010)

# Puis on calcule les coordonnées du point moyen pondéré en divisant la somme des Xmp par la somme des populations (idem pour Ymp)
xctrp10 <- sum(coordsPayspop10$Xmp)/sum(coordsPayspop10$Pop2010)

# Enfin on crée un nouveau fichier sf pour représenter ce point moyen pondéré par les populations
ptmoyp_2010 <- st_point(c(xctrp10, yctrp10))
```

## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

### 2.1. Points centraux : quel est le point qui résume le mieux la position de l'ensemble des points du semis ?

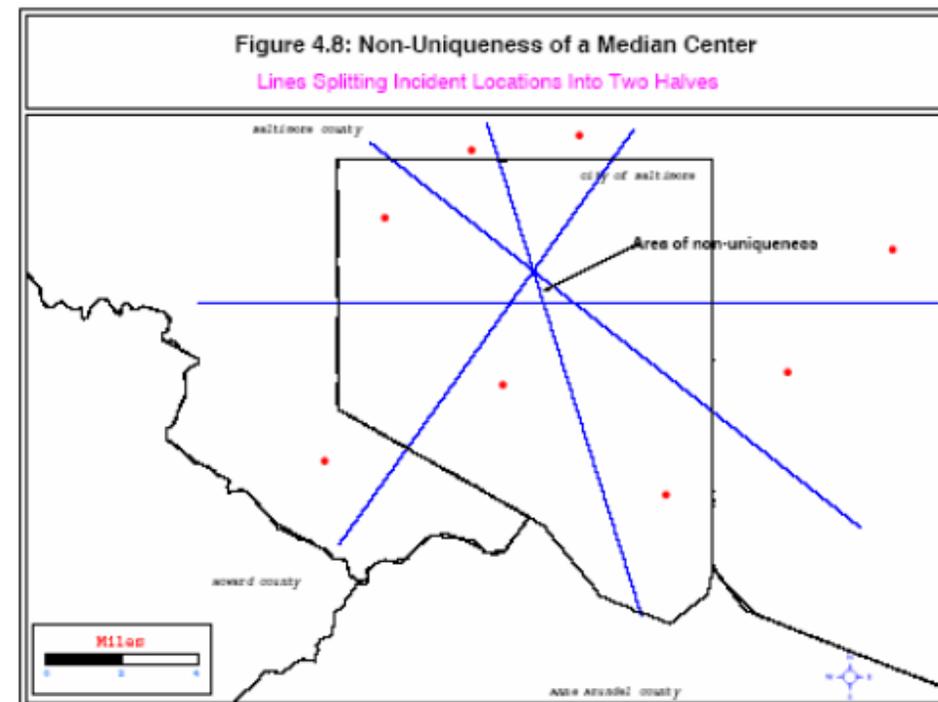
Point moyen : simple/pondéré

Point médian : simple/pondéré

Le **point médian** est une mesure de la tendance centrale robuste aux points aberrants. Il identifie l'emplacement qui **minimise les trajets** depuis et vers toutes les autres entités dans le jeu de données.

#### Méthode de détermination dans un référentiel euclidien...

La méthode utilisée pour calculer le Centre médian est une procédure itérative introduite par Kuhn et Kuenne (1962) et décrite plus en détail dans Burt et Barber (1996). A chaque étape (t) de l'algorithme, un candidat de centre médian est déterminé ( $X_t, Y_t$ ), puis affiné jusqu'à ce qu'il représente l'emplacement qui minimise la distance euclidienne  $D$  vers toutes les entités (ou toutes les entités pondérées) (i) du jeu de données.



Source: Levine et al., 2002

## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

### 2.1. Points centraux : quel est le point qui résume le mieux la position de l'ensemble des points du semis ?

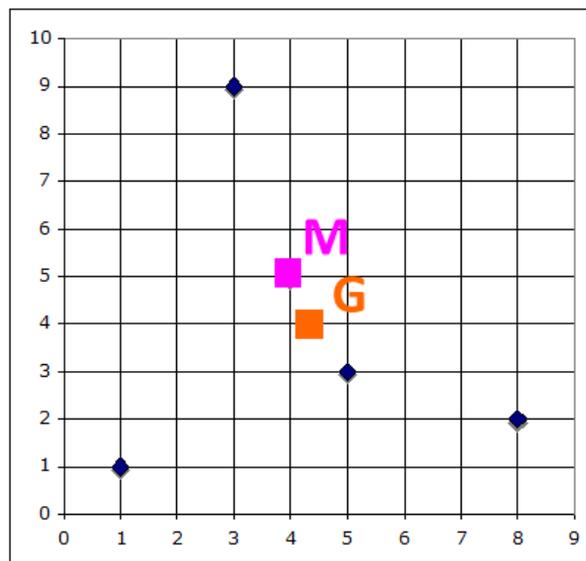
Point moyen : simple/pondéré

**Point médian** : simple/pondéré

Le **point médian** est une mesure de la tendance centrale robuste aux points aberrants. Il identifie l'emplacement qui **minimise les trajets** depuis et vers toutes les autres entités dans le jeu de données.

**Dans un référentiel rectilinéaire...**

Pour un ensemble de  $n$  points ( $1...i...n$ ) décrits par leurs coordonnées  $(X_i, Y_i)$ , c'est le point qui a pour coordonnées  $mX, mY =$  **médianes** des distributions des  $x_i$  et des  $y_i$  de l'ensemble du semis



Coordonnées		
i	X	Y
A	1	1
B	5	3
C	3	9
D	8	2
E	4	5
Total	21	20
Moyenne	4,2	4

Médiane des X

Médiane des Y

## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

### 2.1. Points centraux : quel est le point qui résume le mieux la position de l'ensemble des points du semis ?

Point moyen : simple/pondéré

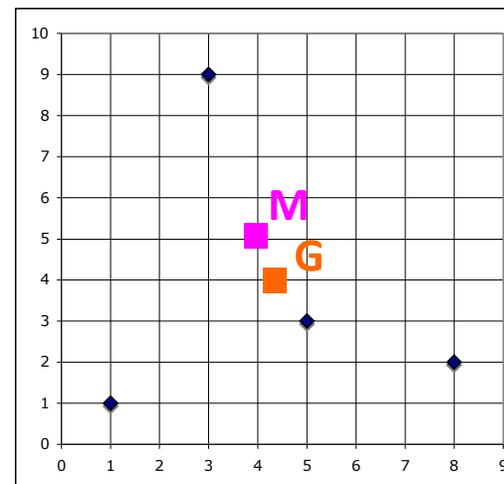
**Point médian** : simple/pondéré

Le **point médian** est une mesure de la tendance centrale robuste aux points aberrants. Il identifie l'emplacement qui minimise les trajets depuis et vers toutes les autres entités dans le jeu de données.

#### Applications possibles

Adapté lorsqu'on souhaite réaliser une **mesure de tendance centrale robuste aux points spatiaux aberrants**. Si on souhaite minimiser l'impact de phénomènes périphériques pour se focaliser davantage sur les phénomènes centraux, le centre médian est pertinent.

Il est souvent intéressant de **comparer les résultats entre centre moyen et centre médian** pour voir l'impact des entités périphériques sur le résultat et détecter la forme de la distribution spatiale (plutôt symétrique ou dissymétrique).



	Coordonnées	
i	X	Y
A	1	1
B	5	3
C	3	9
D	8	2
E	4	5
Total	21	20
Moyenne	4,2	4

## 2. Résumer la distribution d'un semis de points – 2.1. Points centraux (point moyen, point médian)

Point moyen : simple/pondéré

**Point médian** : simple/pondéré

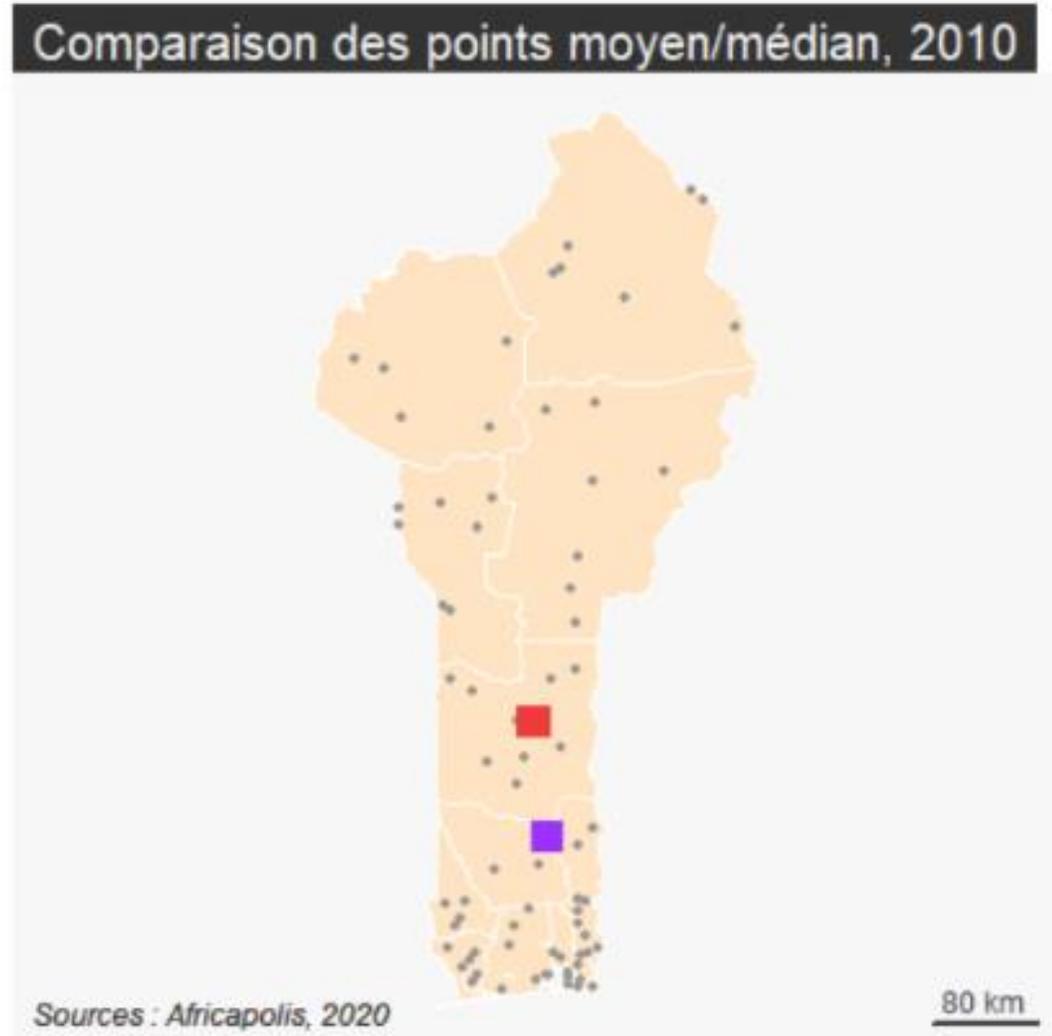
**Non disponible dans QGIS**

**Avec R**

Fonctions de *R-base* et de *sf*

- `st_coordinates` (`data$geometry`) : pour extraire les coordonnées X, Y des points
- `median` (X), `median` (Y) : pour en calculer la valeur moyenne  $X_m$  et  $Y_m$
- `st_point` (`c(Xm,Ym)`) : pour transformer ces deux coordonnées en point

```
# Les coordonnées des villes ont déjà été extraites pour le point moyen  
On repart du même fichier coordsPays10.  
# Calcul médiane de ces coordonnées en X et en Y puis transformation en point  
ptmed_Pays10 <- st_point(c(median(coordsPays10[, 2]), median(coordsPays10[, 3])))  
# pour récupérer directement la valeur de  $X_m$  et  $Y_m$ 
```

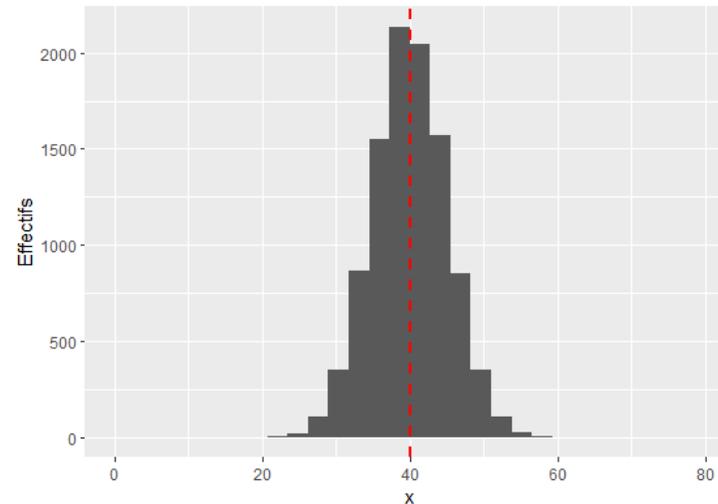


## 2. Résumer la distribution d'un semis de points

### 2.2. Dispersion autour du centre

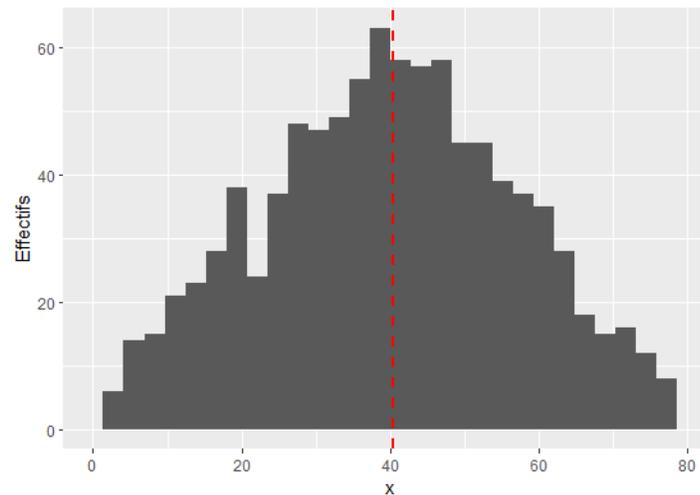
#### Distribution **statistique** : 1 variable (x)

Distribution symétrique  
Faible dispersion autour de la moyenne



**Faible**  
dispersion

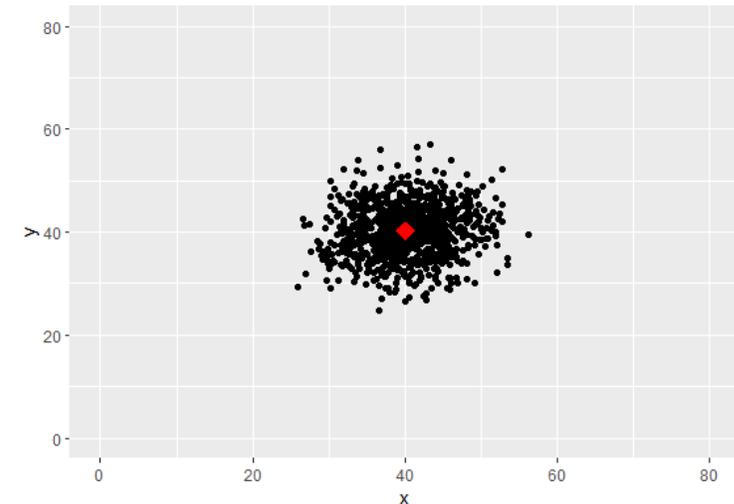
Distribution symétrique  
Forte dispersion autour de la moyenne



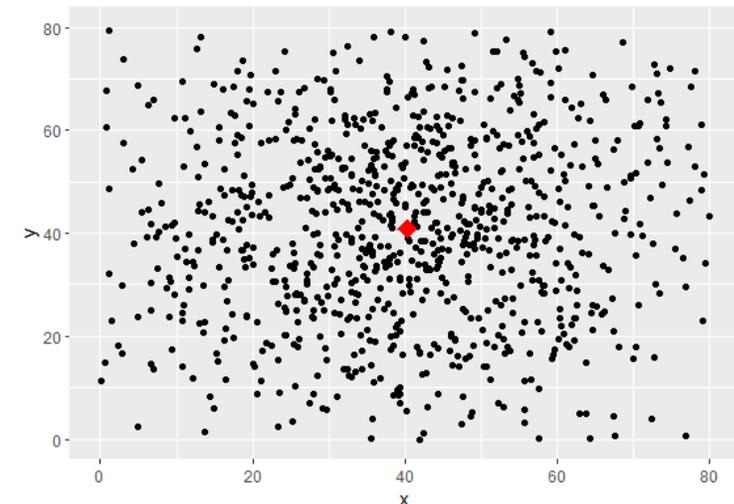
**Forte**  
dispersion

#### Distribution **spatiale** : 2 variables = coord (x,y)

Semis de points symétrique  
Faible dispersion autour du point moyen



Semis de points symétrique  
Forte dispersion autour du point moyen



## 2. Résumer la distribution d'un semis de points – 2.2. Dispersion autour du centre

Dispersion autour du point moyen : **distance-type** / ellipse de dispersion

La **distance-type** mesure la dispersion des points autour du point moyen.

Elle a pour valeur la **racine de la somme des variances** de X et de Y et résulte en une valeur unique qui est une distance.

La compacité d'un ensemble d'entités peut ainsi être représentée sur une carte en traçant un cercle de rayon égal à la valeur de distance standard.

The Standard Distance is given as:

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}} \quad (1)$$

where  $x_i$  and  $y_i$  are the coordinates for feature  $i$ ,  $\{\bar{X}, \bar{Y}\}$  represents the Mean Center for the features, and  $n$  is equal to the total number of features.

The Weighted Standard Distance extends to the following:

$$SD_w = \sqrt{\frac{\sum_{i=1}^n w_i (x_i - \bar{X}_w)^2}{\sum_{i=1}^n w_i} + \frac{\sum_{i=1}^n w_i (y_i - \bar{Y}_w)^2}{\sum_{i=1}^n w_i}} \quad (2)$$

where  $w_i$  is the weight at feature  $i$  and  $\{\bar{X}_w, \bar{Y}_w\}$  represents the weighted Mean Center.

## 2. Résumer la distribution d'un semis de points – 2.2. Dispersion autour du centre

Dispersion autour du point moyen : **distance-type** / ellipse de dispersion

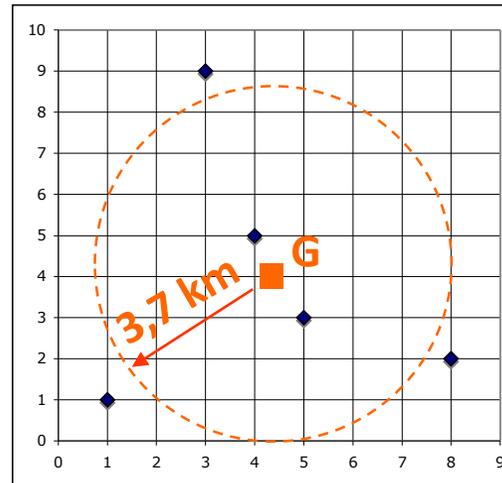
La **distance-type** mesure la dispersion des points autour du point moyen.

Elle a pour valeur la **racine de la somme des variances** de X et de Y et résulte en une valeur unique qui est une distance.

La compacité d'un ensemble d'entités peut ainsi être représentée sur une carte en traçant un cercle de rayon égal à la valeur de distance standard.

i	Coordonnées	
	X	Y
A	1	1
B	5	3
C	3	9
D	8	2
E	4	5
Moyenne	4,2	4
Variance	5,4	8

$$\text{Distance type} = \sqrt{5,4 + 8} = 3,7 \text{ km}$$



Dans l'hypothèse d'une distribution normale, 67% des points seront compris dans ce cercle

Applications possibles

Vous pouvez utiliser ces valeurs pour comparer deux distributions ou plus. Vous pouvez également comparer le même type d'entité à différentes périodes ou heures.

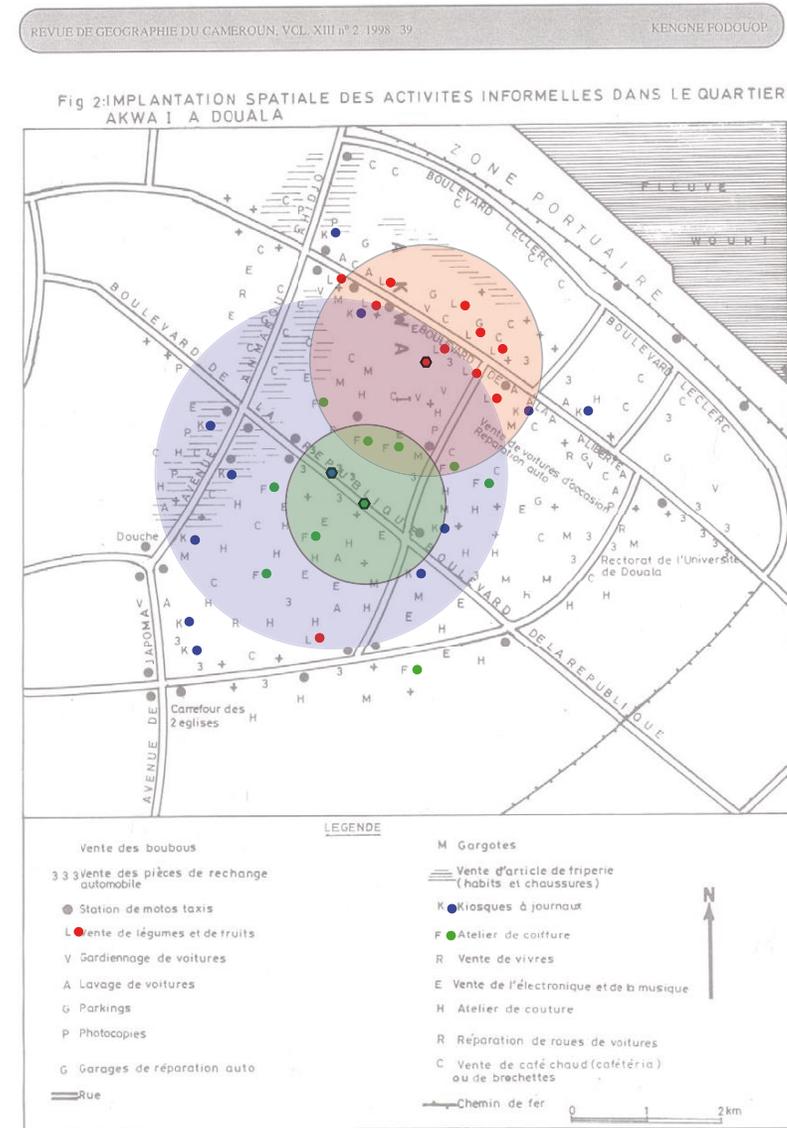
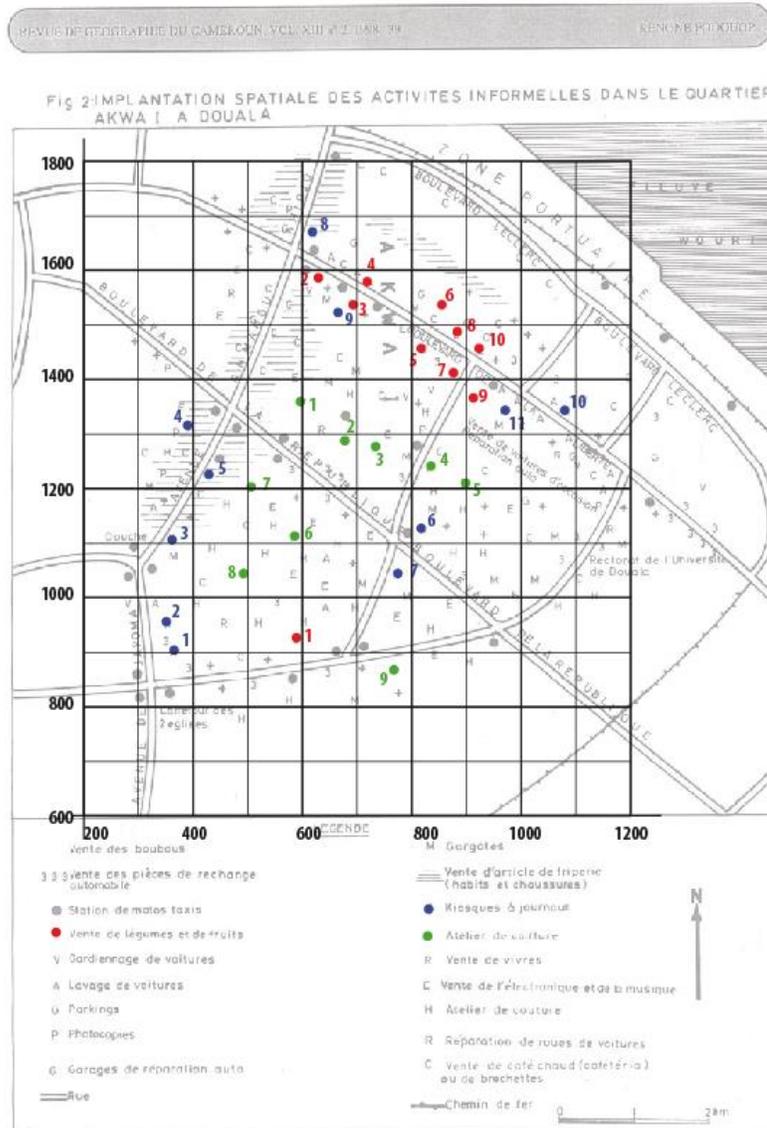
## 2. Résumer la distribution d'un semis de points – 2.2. Dispersion autour du centre

Dispersion autour du point moyen : **distance-type** / ellipse de dispersion

Clin d'œil à une autre école d'été...

La localisation de quelques activités informelles dans le quartier Akwa de Douala

Source : *Ecole d'été*  
 « Statistiques, Cartographies et Analyse spatiale » Yaoundé 2006.

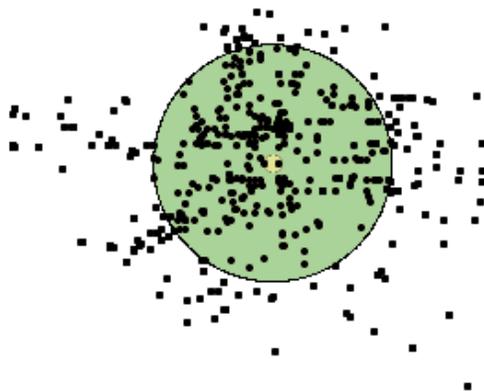
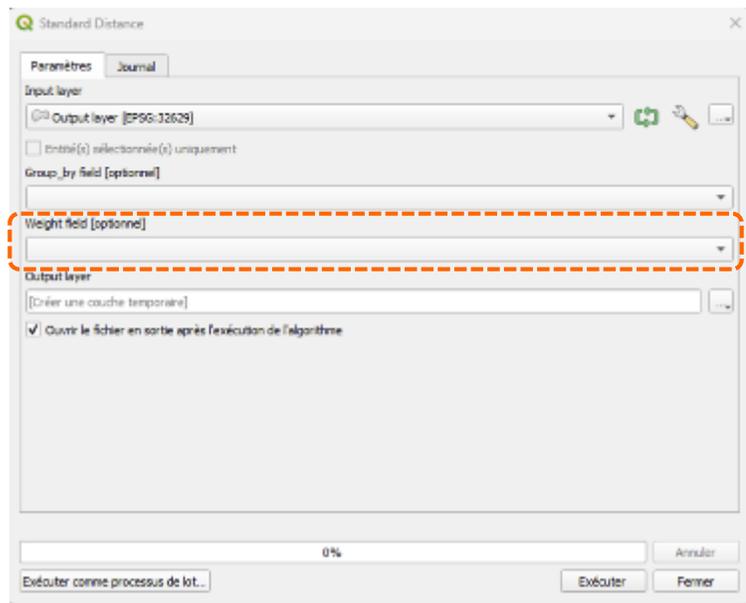


## 2. Résumer la distribution d'un semis de points – 2.2. Dispersion autour du centre

Dispersion autour du point moyen : **distance-type** / ellipse de dispersion

**Avec QGIS**

Menu traitement – boîte à outils – Distance standard

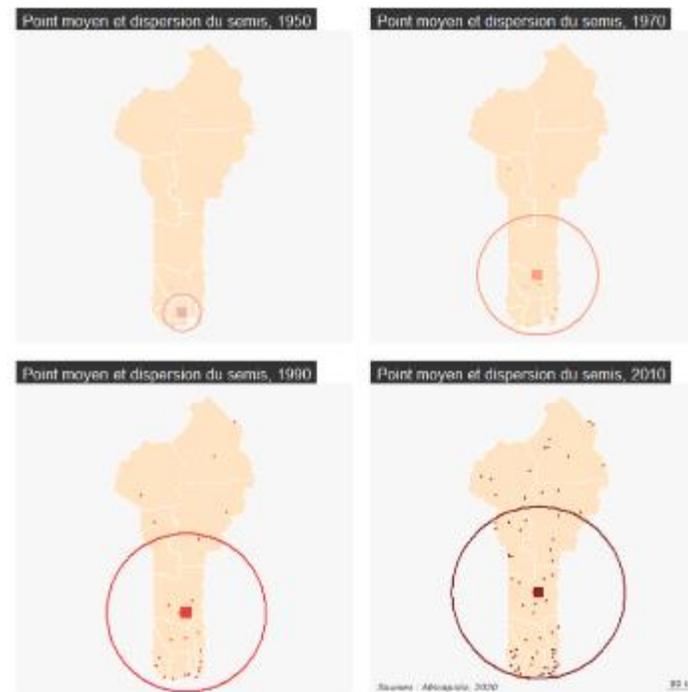


**Avec R**

Fonctions de *R-base* : `sqrt ( )`, `sum ( )`, `nrow ( )` et de *sf* : `st_buffer ( )`

```
# Calcul de la distance standard non pondérée  
sdPays50 <- sqrt(sum((coordsPays50[,2] - ptmoy50[1])^2  
+ (coordsPays50[,3] - ptmoy50[2])^2) / nrow(coordsPays50))
```

```
# Tracer le cercle (à l'aide de la fonction "zone tampon")  
sdPays50_buffer <- st_buffer(ptmoy_Pays50, sdPays50)
```



## 2. Résumer la distribution d'un semis de points – 2.2. Dispersion autour du centre

Dispersion autour du point moyen : distance-type / **ellipse de dispersion**

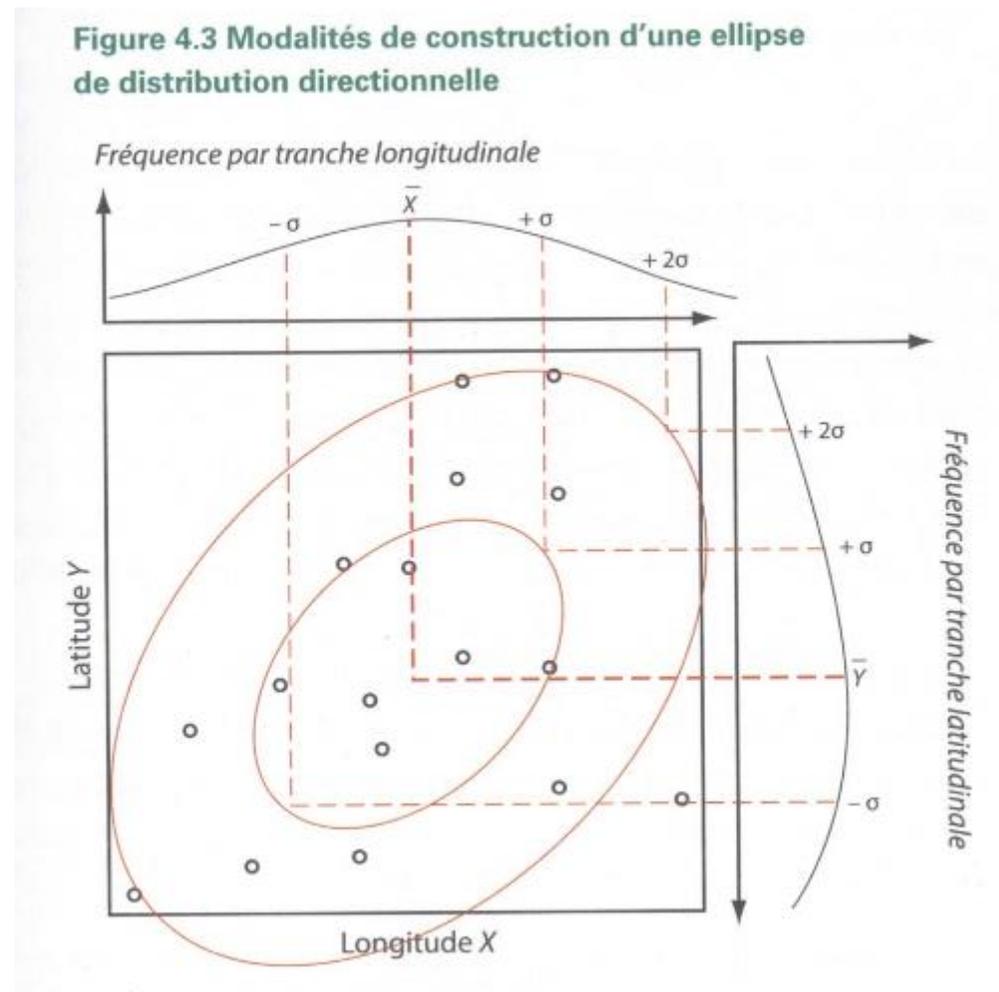
Bien qu'on puisse se faire une idée de l'orientation en dessinant les entités sur une carte, le calcul de l'ellipse de l'écart type rend la tendance évidente.

On peut **calculer l'ellipse de l'écart type** en utilisant les emplacements des points (**non pondérée**)...

... ou les emplacements **pondérés** par une valeur attributive associée aux entités.

**Démarche** pour repérer les axes qui structurent le semis :

- A partir des diagrammes de distribution de X et Y ...
- Centre de l'ellipse = point moyen
- Ecarts-types calculés : pour X, pour Y → définition des axes d'étirement
- Reconstitution de l'enveloppe (en paramétrant la longueur des axes en fonction du nombre d'écarts-types)

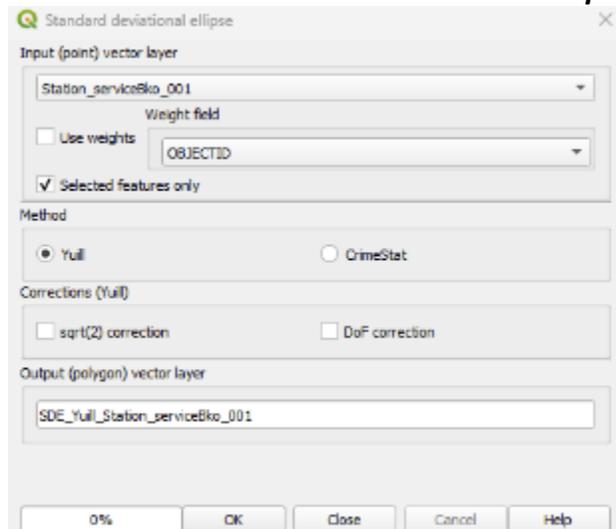


## 2. Résumer la distribution d'un semis de points – 2.2. Dispersion autour du centre

Dispersion autour du point moyen : distance-type / **ellipse de dispersion**

### Avec QGIS

#### Menu vecteur - Standard Deviational Ellipse



### Avec R

Fonctions de *car* : dataEllipse ( )

```
# Tracé de l'ellipse en 2010
```

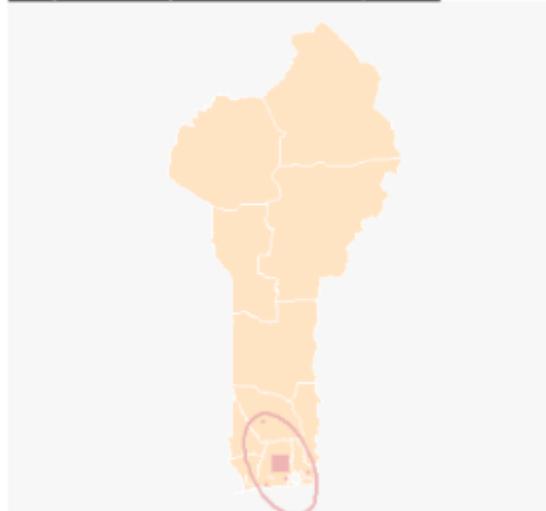
```
# D'abord transformation des coordonnées en data.frame,...
```

```
vil_pt_df_2010 <- cbind(st_drop_geometry(Payspolis2010), st_coordinates(Payspolis2010))
```

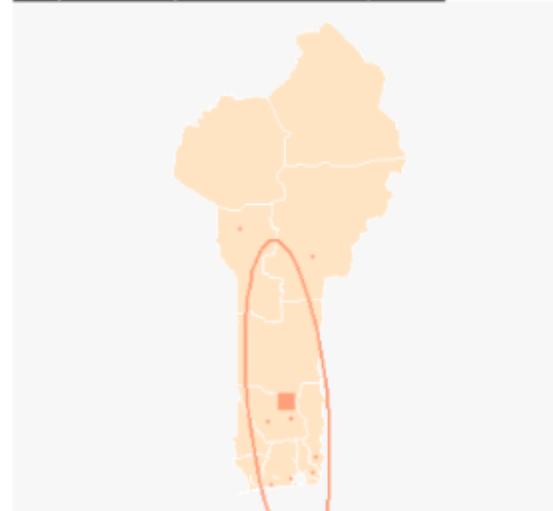
```
# pour pouvoir utiliser la fonction "dataEllipse" du package Car
```

```
dataEllipse(x = vil_pt_df_2010$X, y = vil_pt_df_2010$Y, center.pch = FALSE,  
plot.points=FALSE, col = "brown4", levels = .66, xpd=T, add = TRUE) # xpd=T permet d'écrire  
dans les marges
```

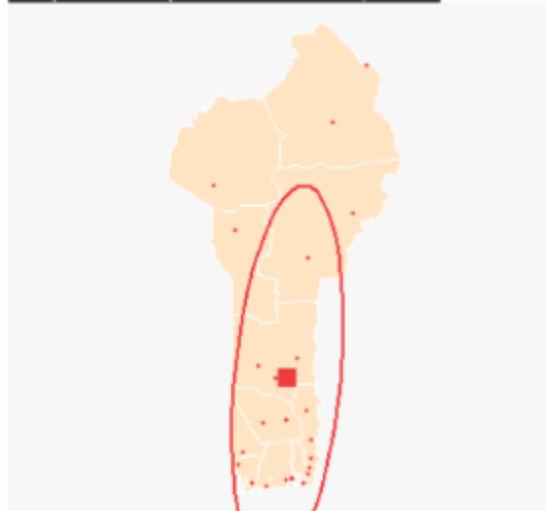
Ellipse de dispersion du semis, 1950



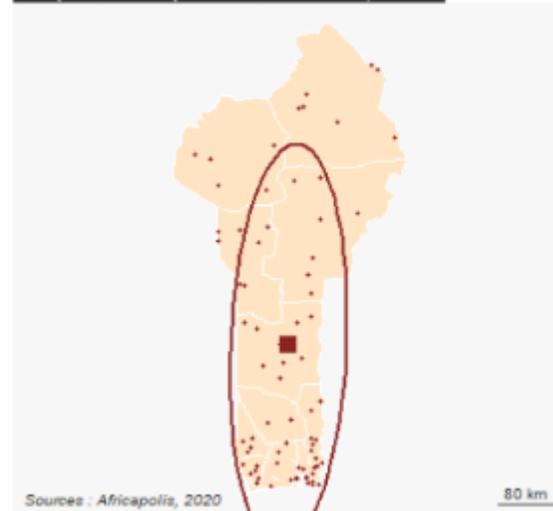
Ellipse de dispersion du semis, 1970



Ellipse de dispersion du semis, 1990



Ellipse de dispersion du semis, 2010

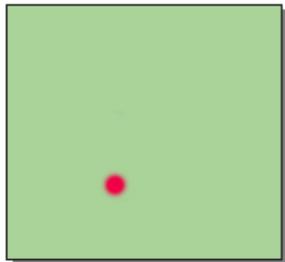
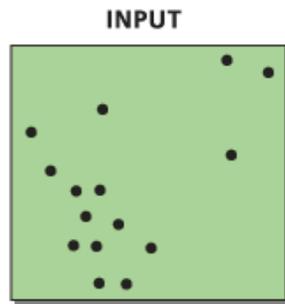


## 2. Résumer la distribution d'un semis de points

### Description de position et de dispersion dans la distribution spatiale

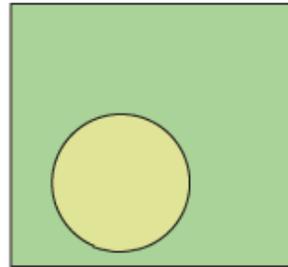
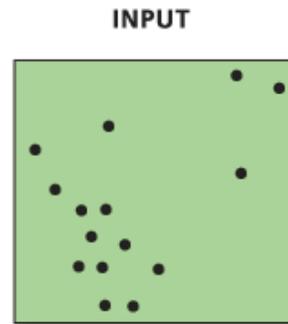
#### Entité centrale

Identifie l'entité située le plus au centre dans une classe d'entités ponctuelles, linéaires ou surfaciques.



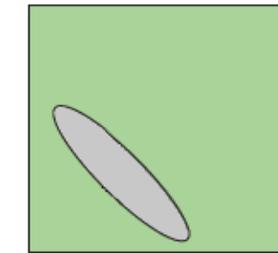
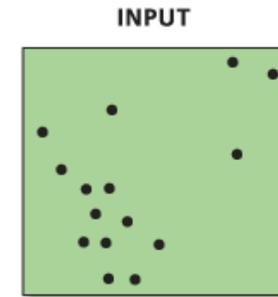
#### Standard Distance

Mesure le degré auquel les entités sont concentrées ou dispersées autour du centre moyen géométrique.



#### Directional Distribution (Standard Deviational Ellipse)

Crée des ellipses de l'écart type pour récapituler les caractéristiques spatiales d'entités géographiques : tendance centrale, dispersion et tendances directionnelles.

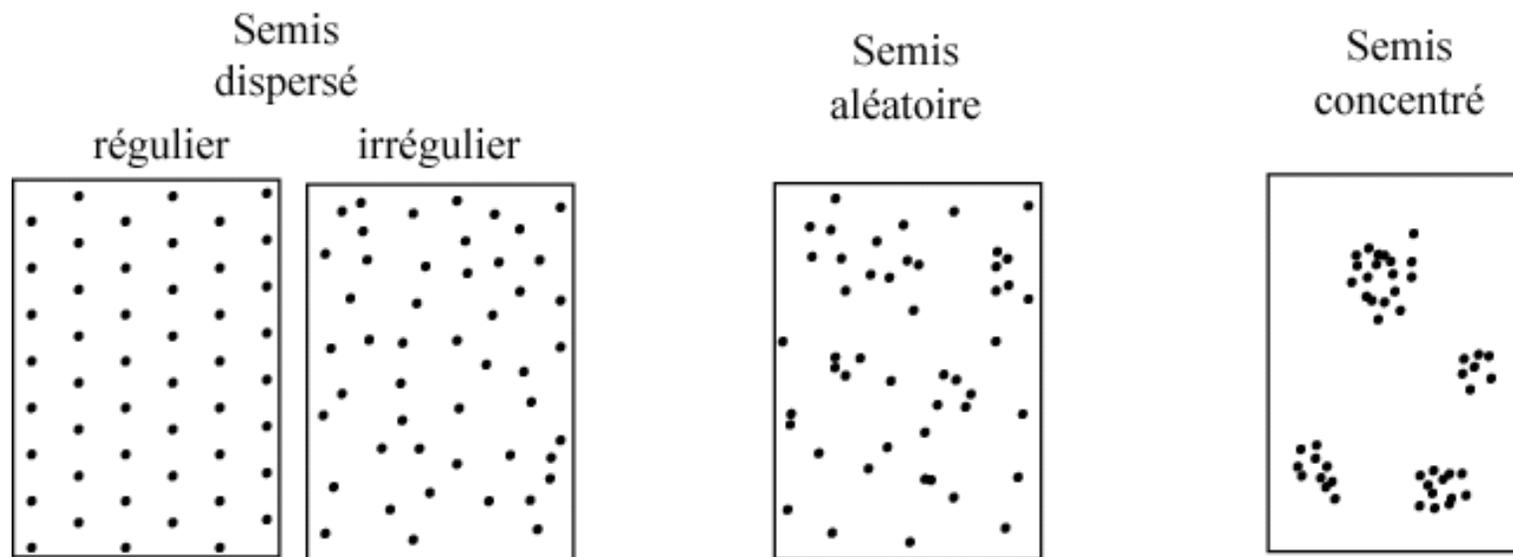


Source: ArcGIS

### **3. Concentration, dispersion: évaluer la forme globale d'un semis de points**

### 3.1. Distributions spatiales de référence

→ Trois distributions de référence



Un semis est un **ensemble de localisations**, résultant d'un processus qui a engendré des points espacés d'une certaine manière.

### 3. Concentration, dispersion : évaluer la forme d'un semis de points

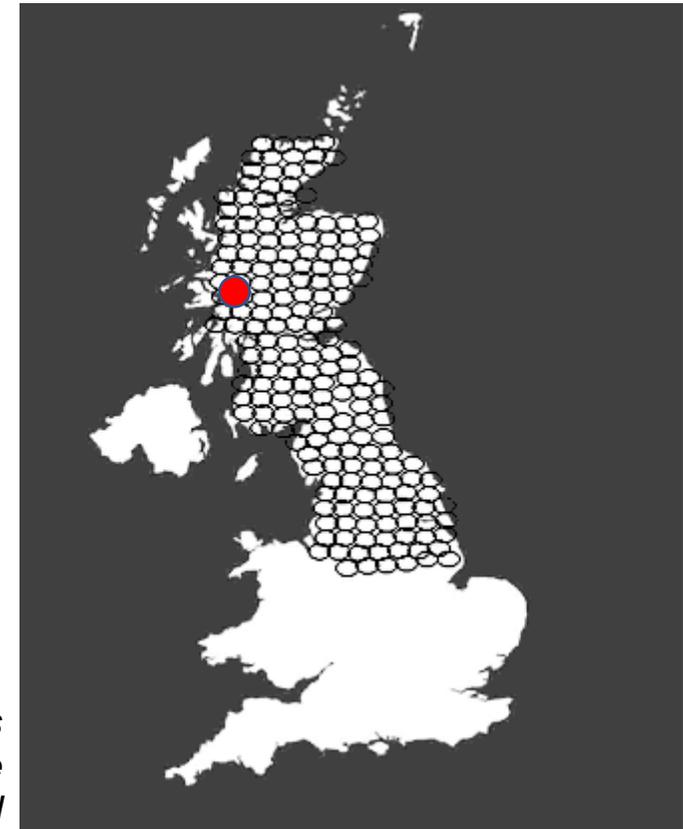
#### 3.1. Distributions spatiales de référence

→ Quel est le lien entre ces **formes** et les **processus** qui ont présidé à leur formation ?

*Stan Openshaw, la centrale de Sellafield et la leucémie infantile (1987)*



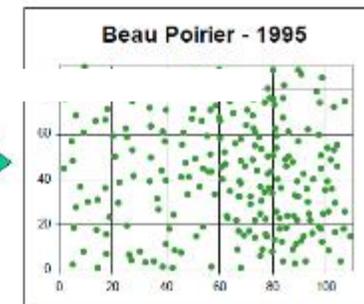
*Complexe nucléaire de Sellafield*



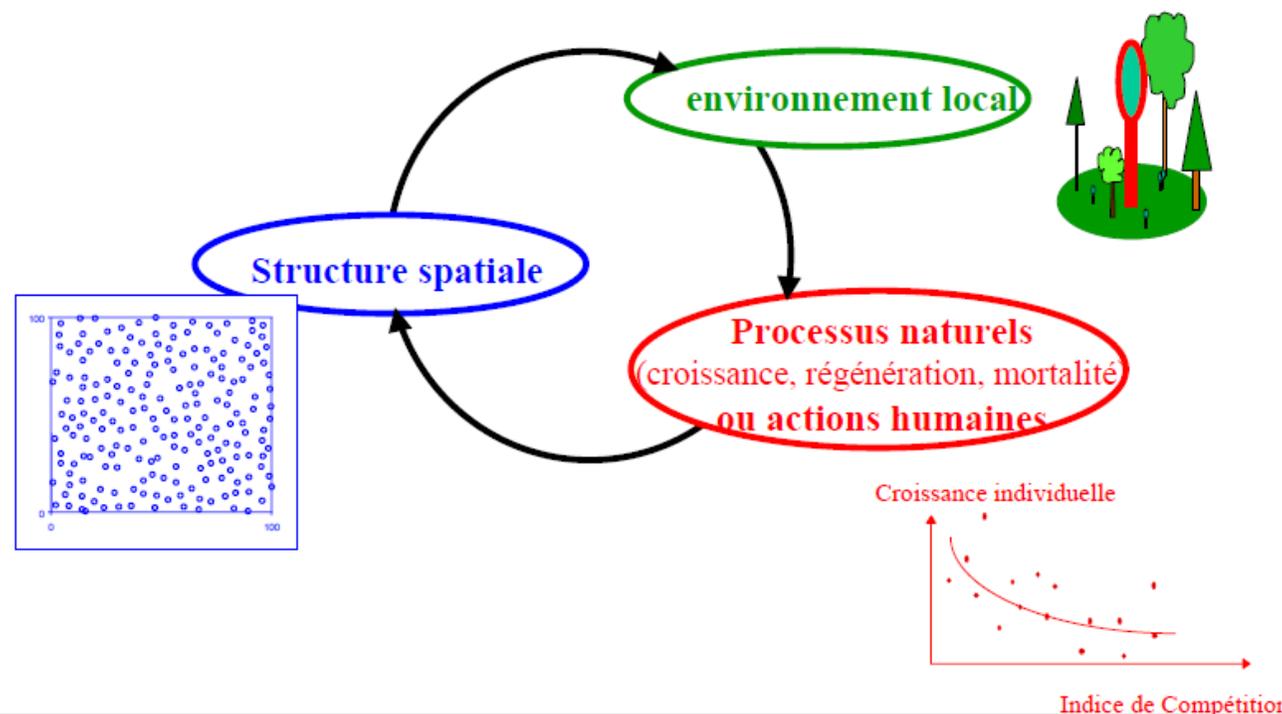
*Grille de dénombrement des cas de Leucémie autour de Sellafield*

### 3. Concentration, dispersion : évaluer la forme d'un semis de points

#### 3.1. Distributions spatiales de référence



→ Quel est le lien entre ces formes et les processus qui ont présidé à leur formation ?



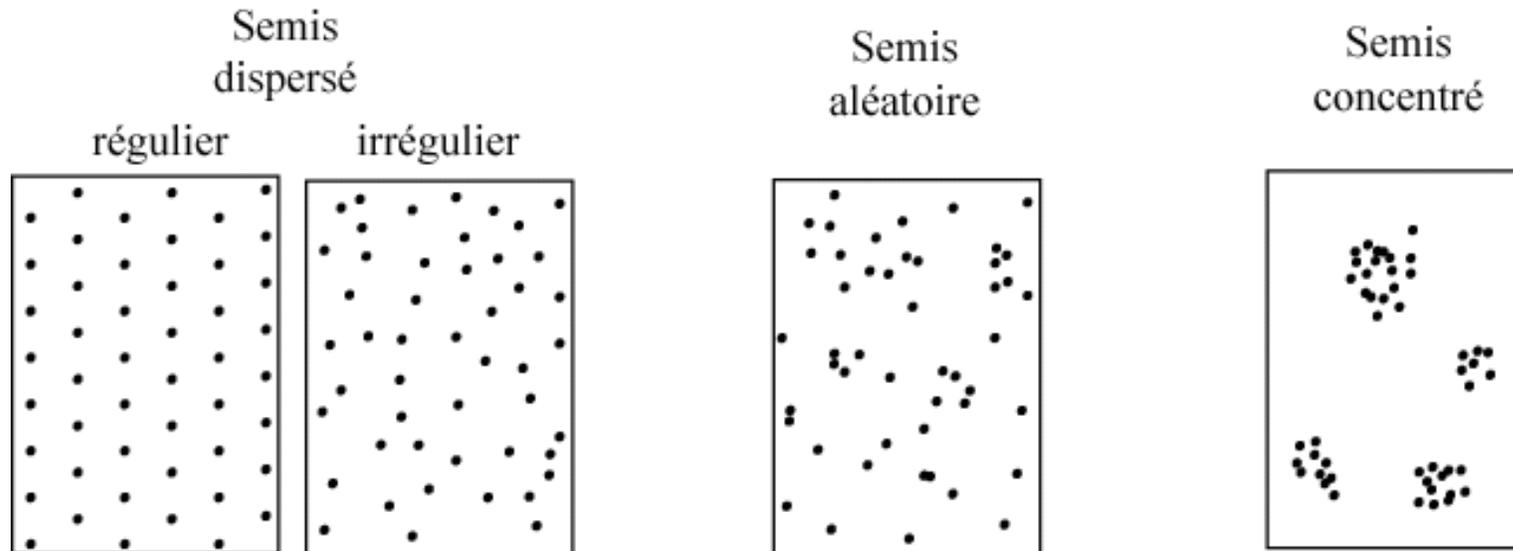
Source : F. Goreaud, 2000, Apports de l'analyse de la structure spatiale en forêt tempérée à l'étude et la modélisation des peuplements complexes, Thèse de doctorat AgroParisTech.

### 3.1. Distributions spatiales de référence

2 grandes familles de méthodes :

- Fondées sur les **densités** mesurées au sein d'un carroyage (*ex: méthode des quadrats*)
- Fondées sur les **espacements** (*ex: méthode du plus proche voisin*)

Point commun : comparaison du semis observé avec un semis aléatoire

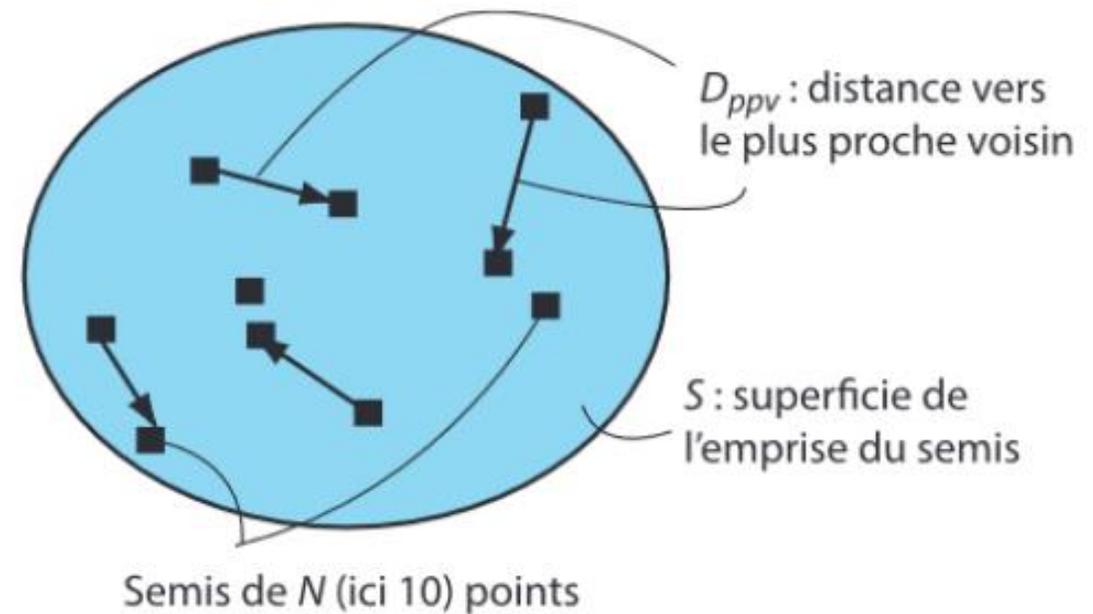


## 3.2. Méthode du plus proche voisin

**N** points distribués sur un espace de surface **S**.

**d** : densité moyenne de points par unité de surface à l'intérieur de l'espace considéré ( $d=N/S$ )

- 1) On calcule pour chaque point *i* la distance  $D_{ppv}(i)$  qui le sépare de son voisin le plus proche.
- 2) On calcule ensuite la moyenne des distances observées au plus proche voisin  $R_o$

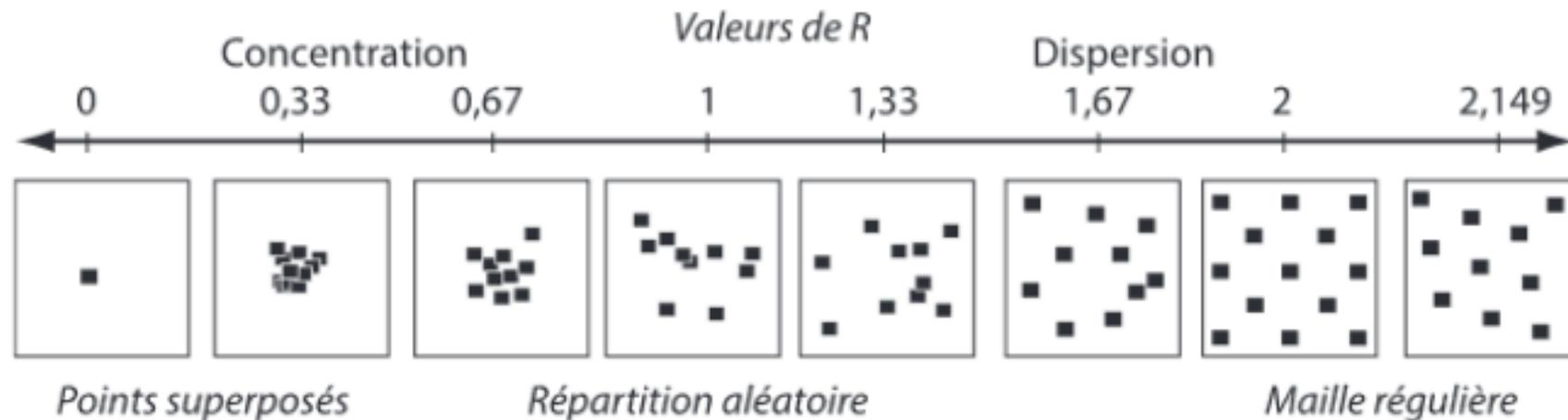
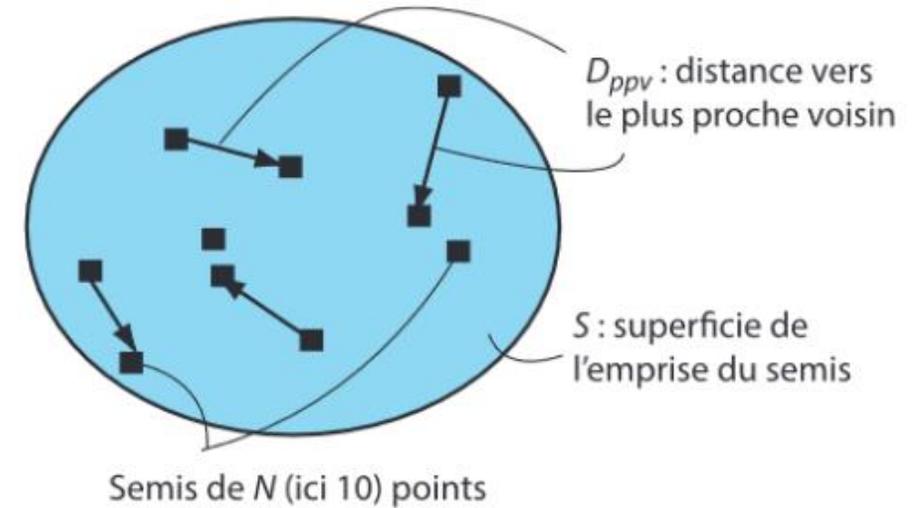


### 3.2. Méthode du plus proche voisin

3) On détermine la **distance théorique moyenne** au plus proche voisin  $R_T$  dans le cas d'une **distribution aléatoire** à l'aide de la formule :

$$R_T = 0,5/\sqrt{d}$$

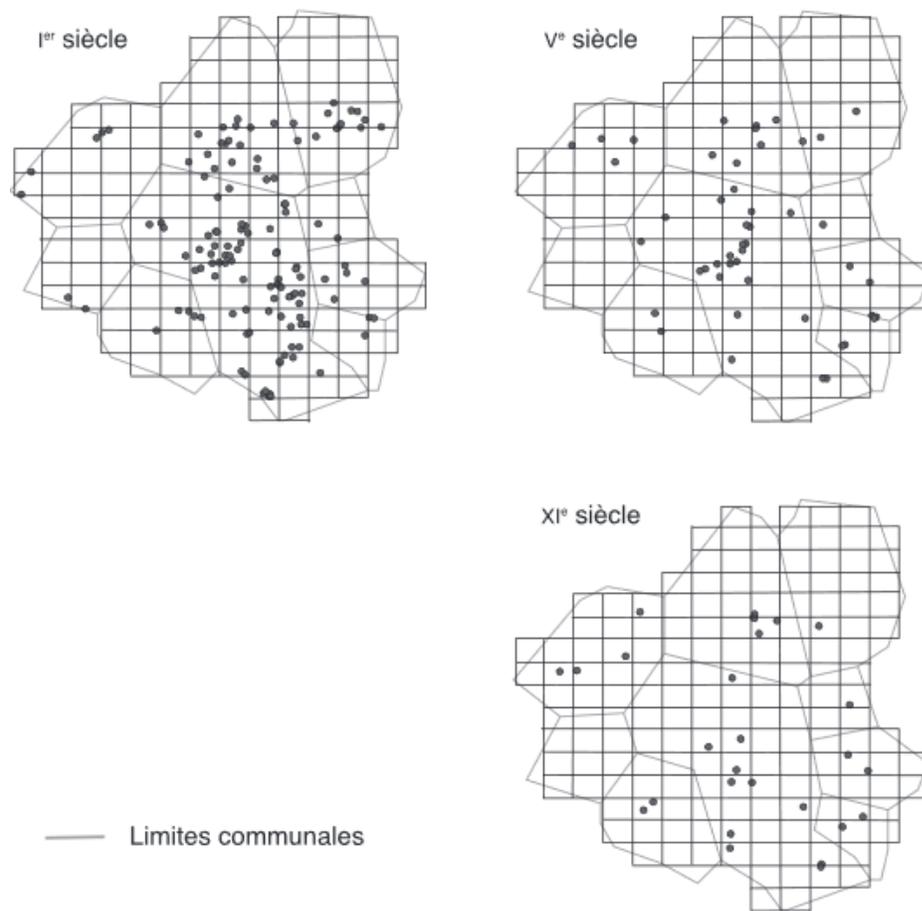
4) On calcule l'**indice de dispersion** qui est le rapport entre ces deux distances :  $R = R_O/R_T$



### 3. Concentration, dispersion : mesurer la forme d'un semis de points

## 3.2. Méthode du plus proche voisin

**FIGURE 2.10 ANALYSE DE L'ÉVOLUTION D'UN SEMIS DE POINTS : LA LOCALISATION DE L'HABITAT RURAL DANS LA VAUNAGE (GARD) AUX I<sup>ER</sup>, V<sup>E</sup> ET XI<sup>E</sup> SIÈCLES**



Source : d'après Archaeomedes, 1997.

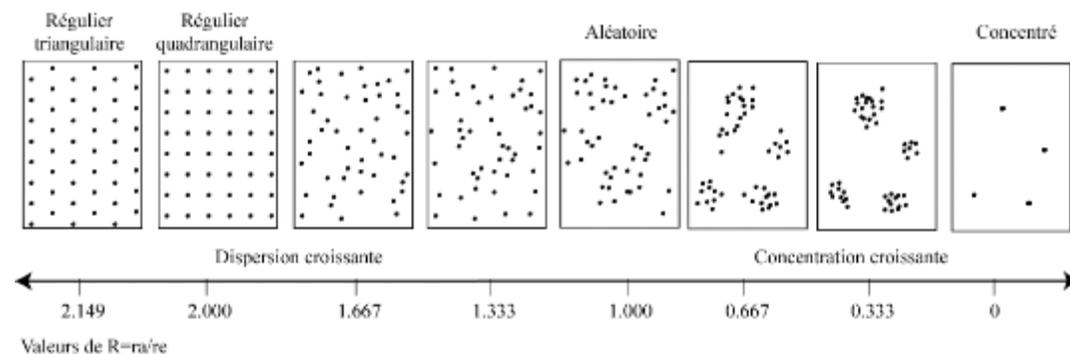
Source : Pumain, Saint-Julien, 2010 (p.91)

**Tableau 2.3b Les distributions spatiales des sites archéologiques dans la Vaunage comparées à la distribution théorique engendrée par un processus de Poisson : méthode du plus proche voisin**

	I <sup>er</sup> siècle	V <sup>e</sup> siècle	XI <sup>e</sup> siècle
m (nombre de points)	127	48	28
D = m/S	1,29	0,48	0,28
r <sub>e</sub> (espacement théorique)	0,44	0,71	0,94
r <sub>a</sub> (espacement observé)	0,28	0,51	0,67
<b>R = r<sub>a</sub>/r<sub>e</sub></b>	<b>0,63</b>	<b>0,72</b>	<b>0,71</b>
SE (r <sub>e</sub> )	0,0204	0,0539	0,0923

Source : Archeomedes, 1997.

Source : Pumain Saint-Julien, 2010 (p.95)



Valeurs de R=ra/re

## 3.2. Méthode du plus proche voisin

### Limites

#### (1) Effets de bordure :

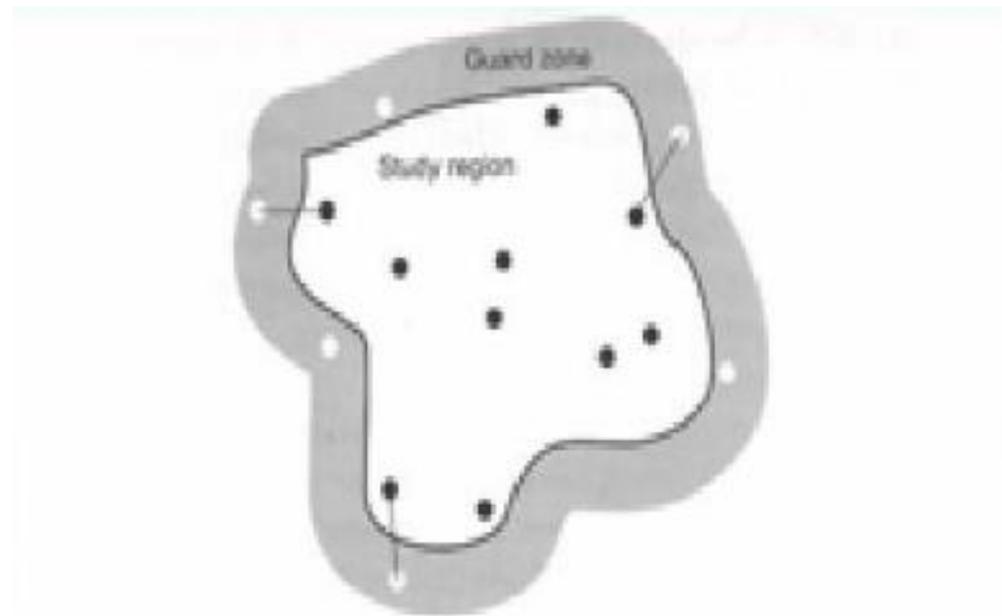


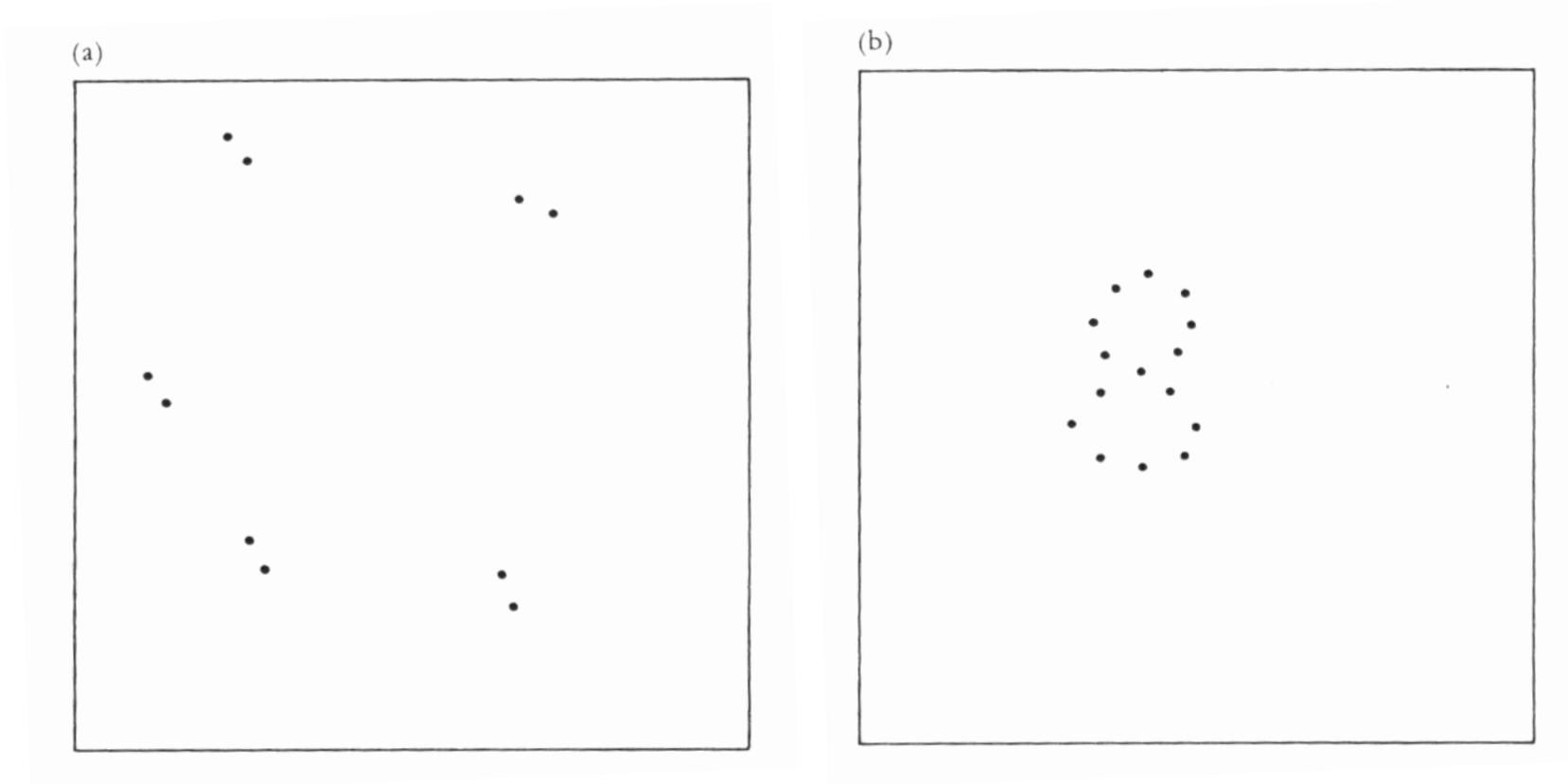
Figure 4.15 Use of a guard zone in a distance-based point pattern measure.

→ On ajoute souvent une zone « buffer », un couloir autour de la zone d'étude où les points sont relevés mais pas utilisés dans l'analyse

## 3.2. Méthode du plus proche voisin

### Limites

#### (2) Effets d'échelle :

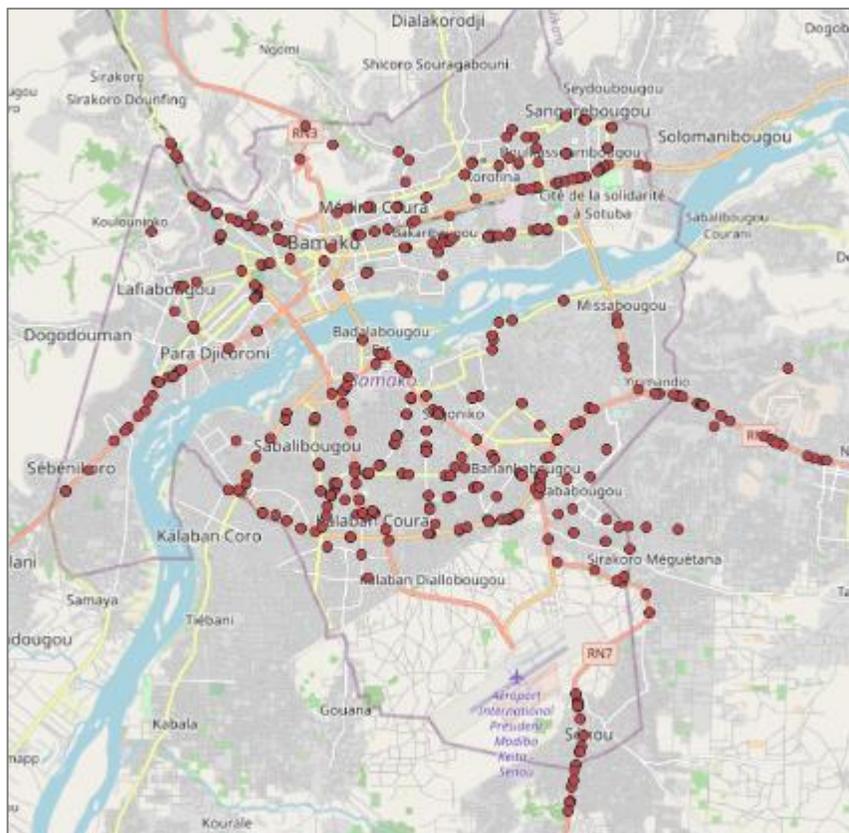


→ Possibilité d'élargir l'analyse à la distance aux 2<sup>e</sup>, 3<sup>e</sup> voisins les plus proches...

## 4. Des points aux surfaces de densités : approche locale des concentrations via la méthode des noyaux (*Kernels*)

## Des points à une surface de densité : quelles concentrations locales d'un phénomène ?

Localement, où sont les zones de plus forte concentration de stations services dans le district de Bamako ?



Sources : S.S. Traoré, 2020 ; fond OSM

### Objectifs

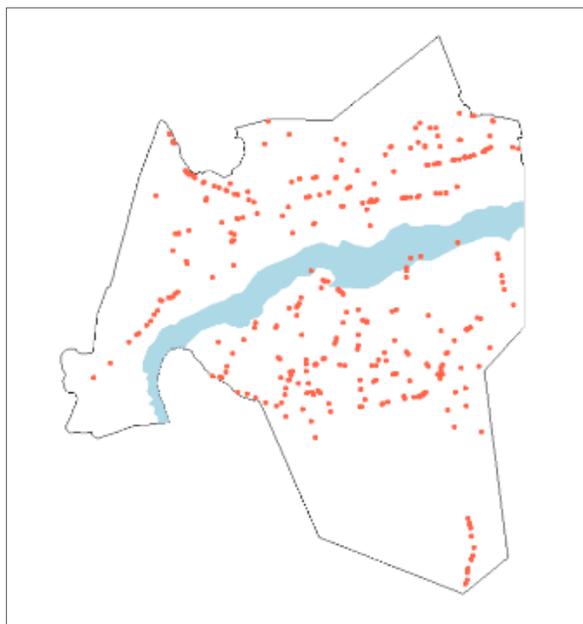
- Localiser des zones de concentration, mettre en valeur des **tendances continues** dans l'espace
- **Transformer une info ponctuelle discrète** (semis de points = présence/absence d'un phénomène) **en surface de densité** pour estimer l'intensité du phénomène étudié en tout point de l'espace

#### 4. Des points aux surfaces de densités : approche locale des concentrations via la méthode des noyaux (Kernels)

### Principe général de la méthode des noyaux (*KDE : Kernel Density Estimation*)

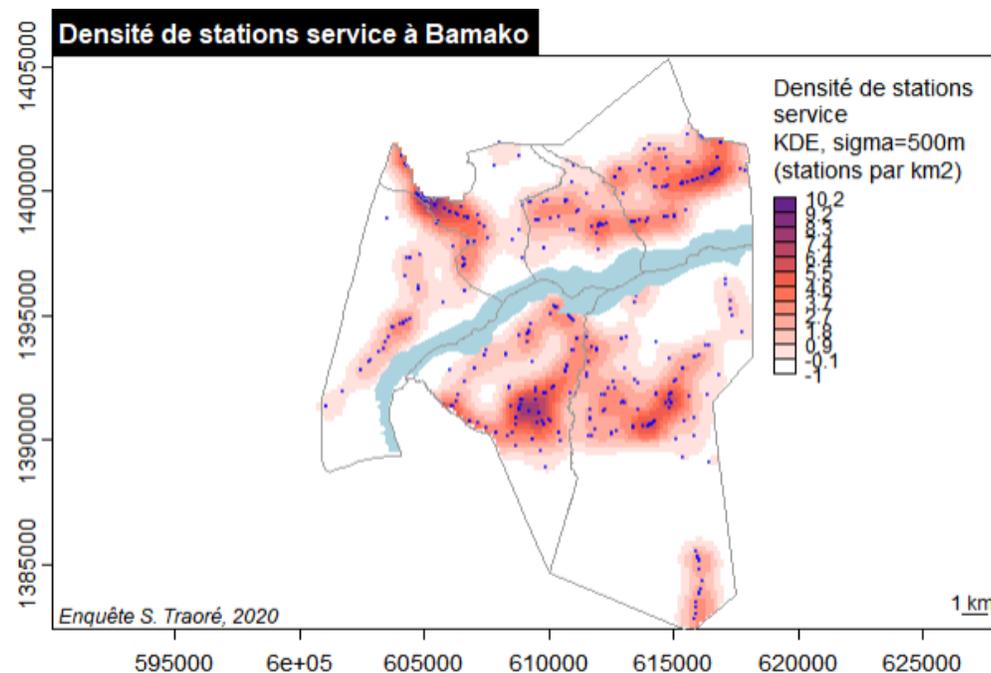
La densité d'un phénomène **en un point de l'espace** est indissociable de celle de ses **voisins** : elle peut être estimée par la densité d'observations  $i_1, i_2, \dots, i_n \dots$  qui se trouve à proximité de ce point

Carte de lissage par noyaux de densité, « *kernels* », « *heat map* »...  
= une des méthodes de **lissage** de données



Des points

...



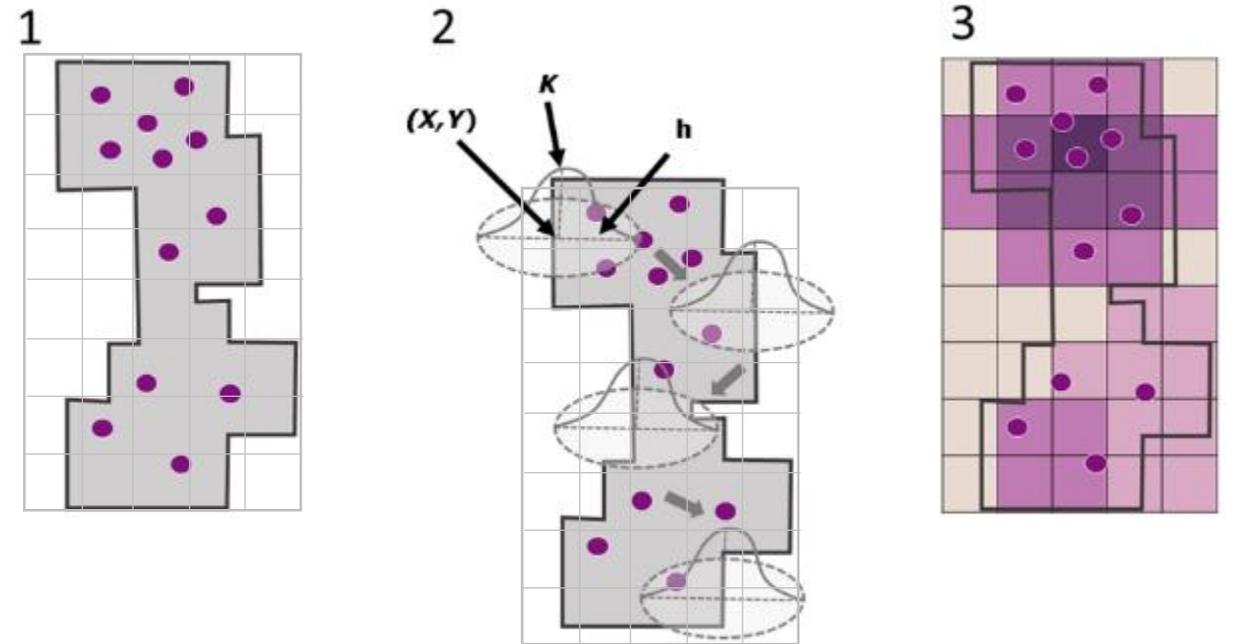
à une surface de densité

#### 4. Des points aux surfaces de densités : approche locale des concentrations via la méthode des noyaux (Kernels)

### Procédure de la méthode des noyaux (*KDE : Kernels Density Estimation*)

- Création d'une **grille d'estimation**, maillant l'espace étudié **(1)**
- Sur le centre de chaque carreau de la grille (point d'estimation), on place une « **boîte** » en forme de **cloche** au sein de laquelle on compte le **nombre de points observés dans le voisinage** du point  $j$  (2).
- Cette « boîte » est définie par deux paramètres :
  - Sa **portée** (notée  $h$  sur le schéma)
  - Sa **forme** (notée  $K$  sur le schéma).
- Par itération, on fait glisser cette boîte de d'un carreau à un autre de la grille = **fenêtre mobile qui balaie la zone étudiée** (3) pour attribuer à chaque carreau une valeur de densité d'un phénomène dans un voisinage de portée  $h$

Figure 1  
Estimation spatiale par densité de noyau



Source : Statistique Canada, Centre de démographie.

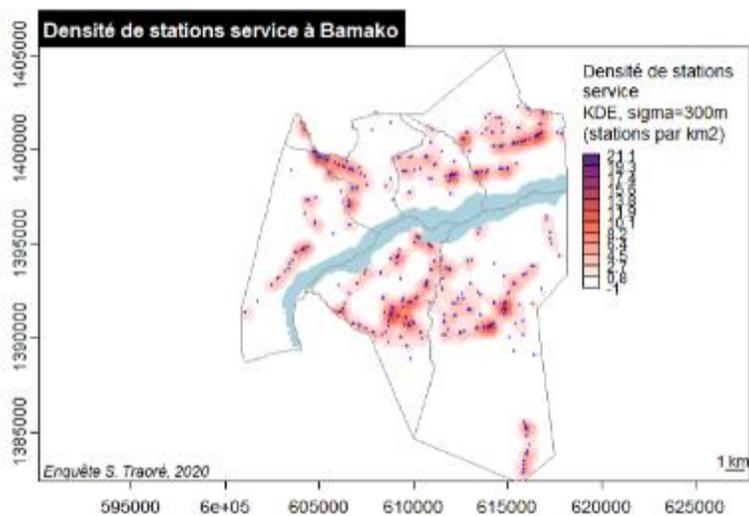
D'après <https://www150.statcan.gc.ca/n1/pub/91f0015m/91f0015m2021001-fra.htm>

## 4. Des points aux surfaces de densités : approche locale des concentrations via la méthode des noyaux (Kernels)

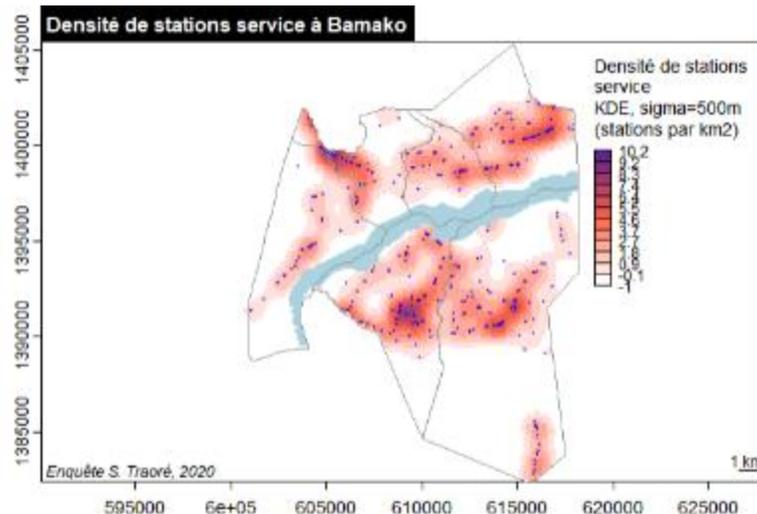
### Paramètres

- Incidence du choix de la **portée**

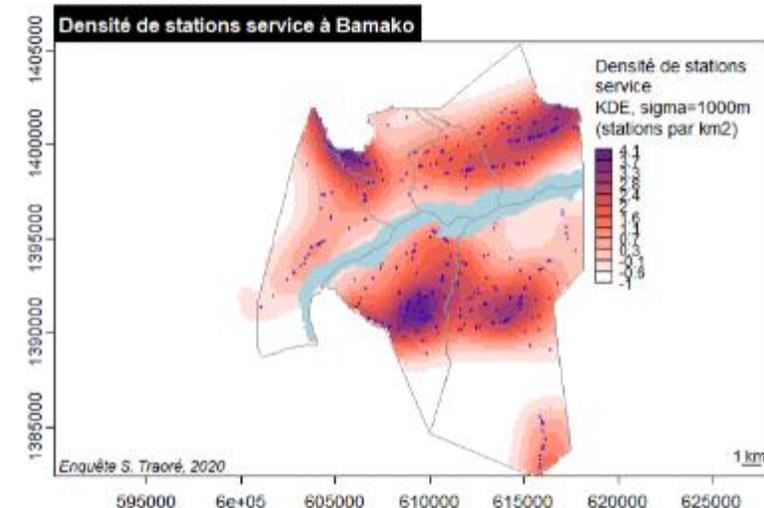
300 m



500 m



1000 m



Dans R

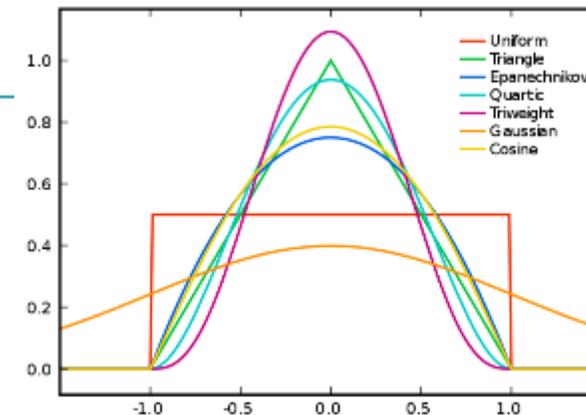
Package *spatstat* → fonction *density.ppp*, argument *sigma*

```
ds <- density.ppp(p, sigma = 1000)
```

## 4. Des points aux surfaces de densités : approche locale des concentrations

### Paramètres

- Incidence du choix de la **fonction de décroissance de la distance**

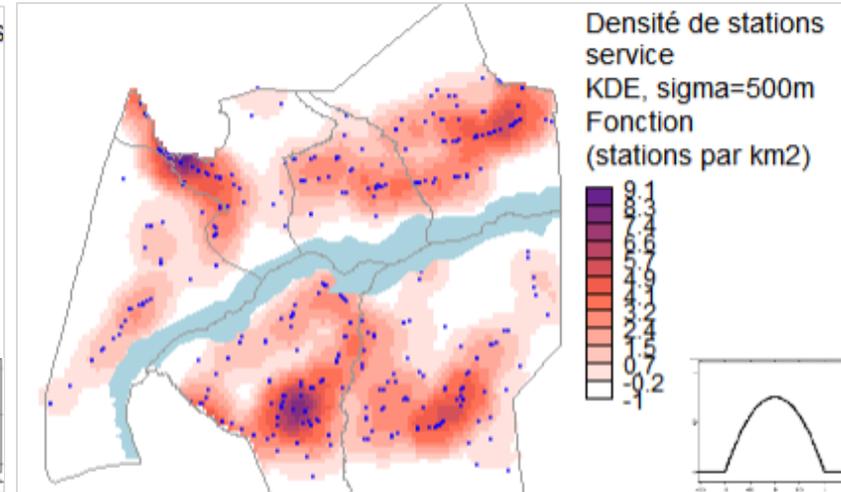
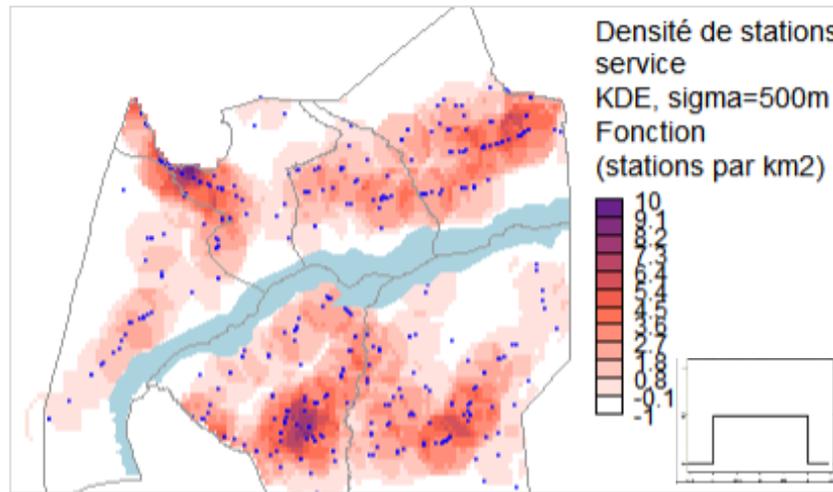
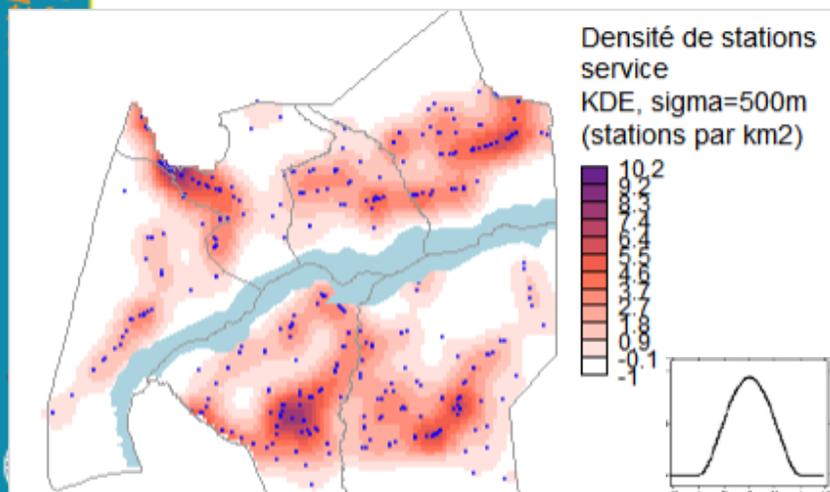


Source: wikipedia

Par défaut, la fonction **gaussienne**

... mais aussi, fonction **uniforme**

... ou **epanechnikov**



Dans R

Package *spatstat* → fonction *density.ppp*, argument *kernel*

```
ds <- density.ppp(p, kernel="disc", sigma = 500)
```

## 4. Des points aux surfaces de densités : approche locale des concentrations via la méthode des noyaux (Kernels)

**En ouverture : « When is a heat map not a heat map » ...**

Attention ! Des confusions possibles entre plusieurs méthodes de transformation de points en surfaces continues.

Le choix de ces méthodes dépend de :

→ la nature du phénomène :  
 Discret/Continu ?

→ la prise en compte ou non des attributs, au-delà de la localisation :

Qu'observe-t-on dans le voisinage des points :  
 Absence/présence d'un phénomène ?  
 Quantité de ressources ?  
 Degré de ressemblance entre les lieux ?



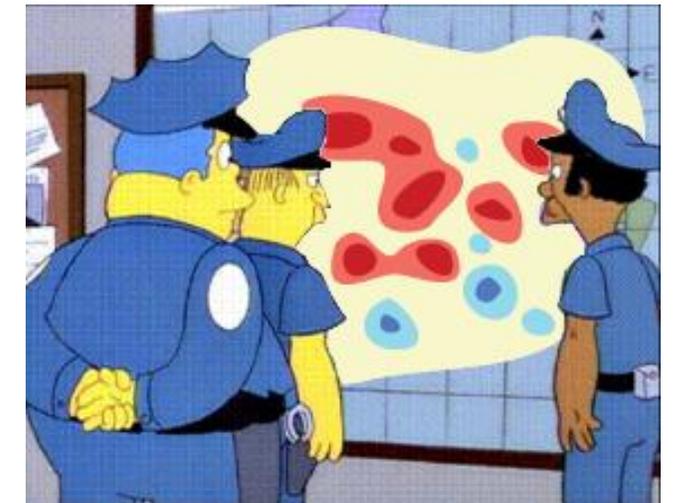
(1) Limites d'une carte par points pour identifier des formes d'organisation spatiale



(2) Au sens strict, « heat map » = lissage par noyaux sur données discrètes de présence/absence (Kernel Density Estimation)



(3) Méthodes d'interpolation : rendu visuel proche de 2) mais pertinentes pour les phénomènes continus seulement



(4) Méthodes d'autocorrélation spatiale : détection de clusters, représentation lissée possible, mais pas une « heat map » au sens strict

# Bibliographie élémentaire

**Feuillet T., Cossart E., Commenges H.** (2019). *Manuel de géographie quantitative*, Cursus, Armand Colin, 237 p.

**Grasland, C.** (2000), « Chapitre 1 : Analyse des semis de points », *Cours d'Analyse spatiale et de modélisation des phénomènes géographiques*, Université Paris VII (en ligne : <http://grasland.script.univ-paris-diderot.fr/go303/ch1/ch1.htm>)

**Pumain, D., Saint-Julien, T.** (2010). *Analyse spatiale : Les localisations*. Armand Colin. Chapitre 2 : Lieux et distribution de lieux

**Taylor P.J.** (1977). *Quantitative Methods in Geography: an introduction to Spatial Analysis*, Waveland Press, 386 p.