

# COURS MAUP

## MODIFIABLE AREAL UNIT PROBLEM

FLORIAN BAYER - M2 IGAST



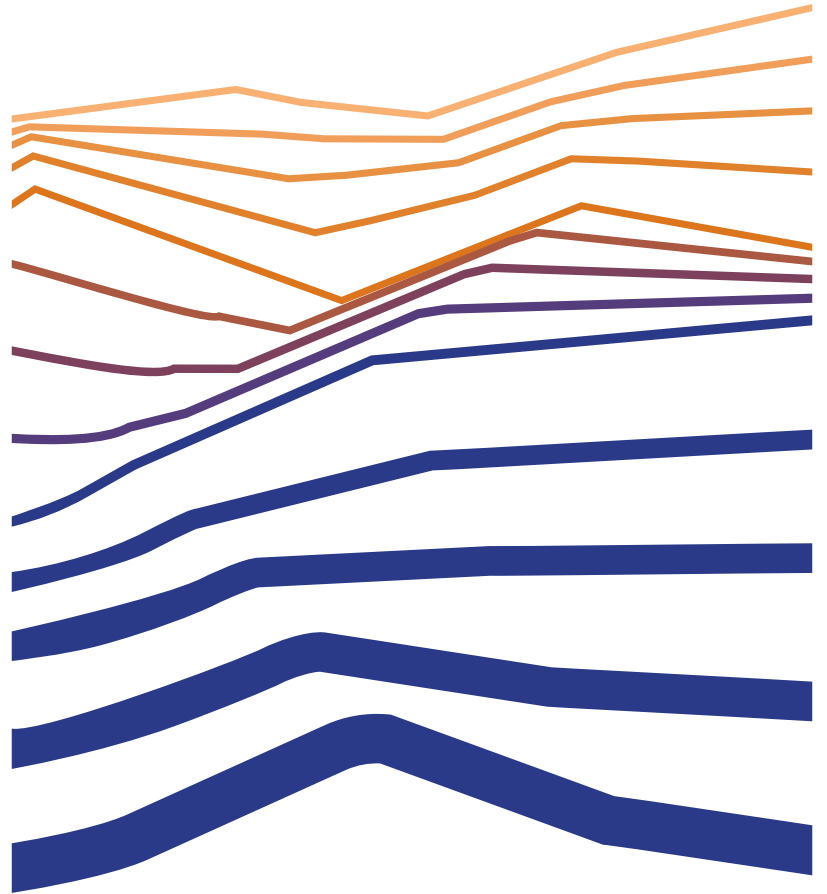
# © OBJECTIFS DU TD

Ce cours a pour objectif de vous sensibiliser au Modifiable Areal Unit Problem (MAUP) et à ses implications pour l'analyse spatiale.

Il vous permettra de comprendre comment le choix des unités spatiales peut influencer les résultats de l'analyse et d'apprendre à identifier et à atténuer les effets du **MAUP**.

# 1 - LE MAUP

## COMPRENDRE LE PROBLÈME



## 1.1 - DÉFINITION DU MAUP

Le **MAUP (Modifiable Areal Unit Problem)** est un biais statistique qui peut influencer les résultats en analyse spatiale.

Il survient lorsque des mesures ponctuelles de phénomènes spatiaux sont agrégées en unités spatiales (comme des régions ou des maillages abstraits) pour calculer des **valeurs résumées** telles que des taux ou des proportions.

Ces valeurs sont influencées à la fois par la **forme** et par l'**échelle** de l'unité d'agrégation.

## 1.1 - DÉFINITION DU MAUP

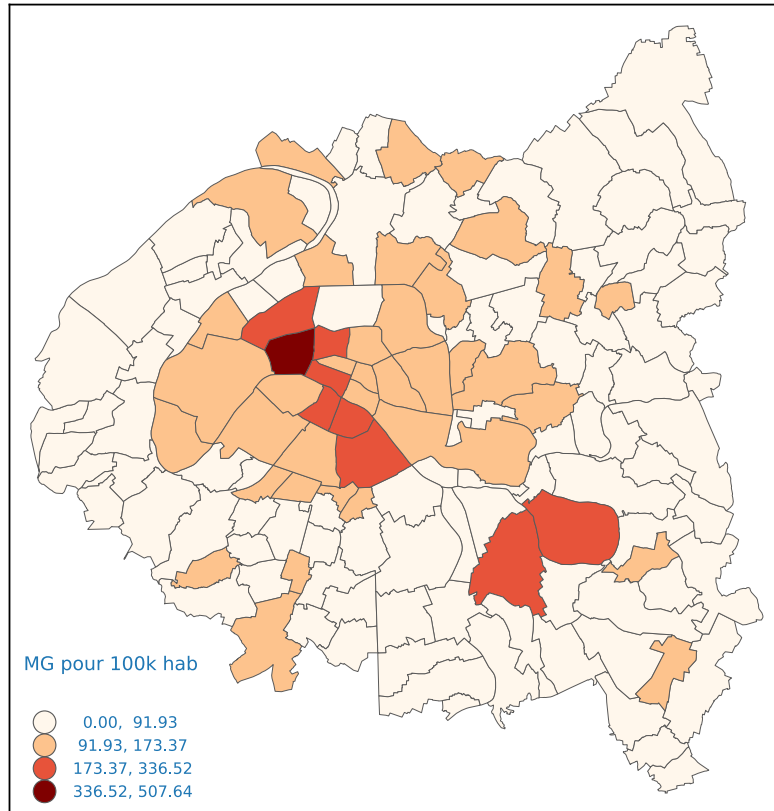
Les conclusions que vous tirez de l'analyse de données spatiales peuvent varier en fonction de la **taille**, de la **forme** et de la **configuration** des unités géographiques que vous utilisez.

Ce problème peut conduire à des **interprétations erronées** des données et à des conclusions inexactes.

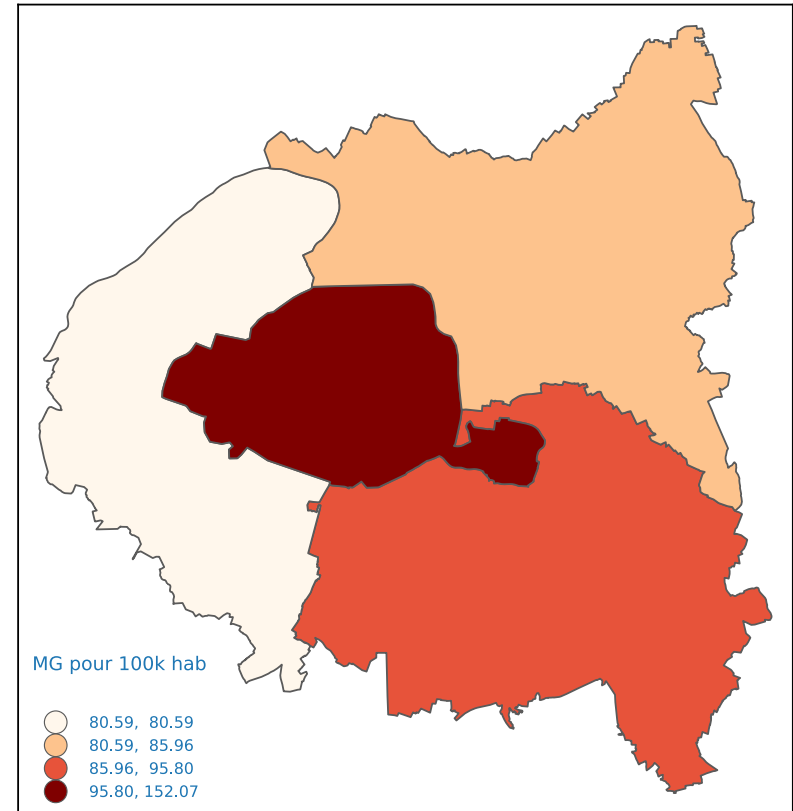
**Les résultats de vos études dépendront du découpage spatial choisi !**

# 1.1 - EXEMPLE : DENSITÉ DE MÉDECINS GÉNÉRALISTES

Niveau Communal



Niveau Départemental



**Niveau communal** : Forte variabilité spatiale de la densité entre communes. **Niveau départemental** : Lissage des variations par agrégation. **Constat** : Distribution spatiale et variabilité changent avec l'échelle.

# 1.1 - CONSÉQUENCES DU MAUP

## BIAIS DANS LES INTERPRÉTATIONS

Résultats variant selon le mode d'agrégation géographique

## BIAIS DANS LES ÉTUDES

Corrélation réelle masquée

---

Association artificielle créée

## GÉNÉRALISATION LIMITÉE

Résultats non généralisables à d'autres zones avec limites différentes

## IMPACT SUR LES DÉCISIONS

Décisions biaisées basées sur données agrégées

## 1.2 - LES DEUX MANIFESTATIONS DU MAUP

Le MAUP survient lorsque les données sont agrégées géographiquement.

Il se manifeste par **deux effets principaux** :

1. **L'effet de zonage** → forme des unités
2. **L'effet d'échelle** → taille des unités



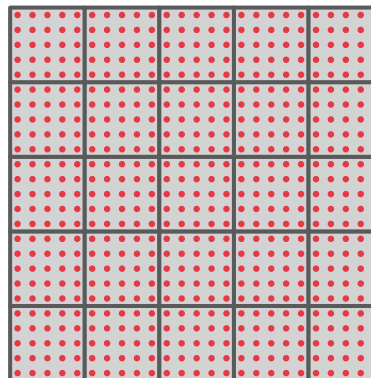
## 1.2 - EFFET DE ZONAGE

Il s'agit de la variation des statistiques causée par le regroupement des données en **configurations différentes mais à la même échelle**.

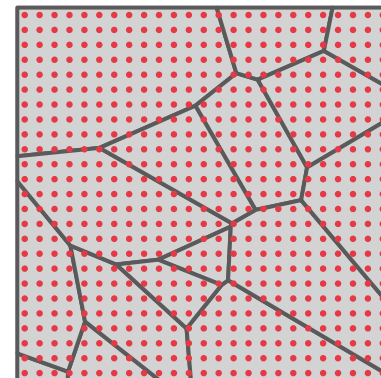
Exemple : deux maillages différents pour la même zone avec une population répartie régulièrement dans l'espace. Les valeurs agrégées varient selon le découpage choisi.

**Même échelle, formes différentes = résultats différents !**

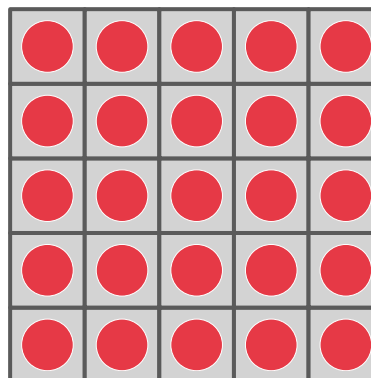
Grille régulière + points



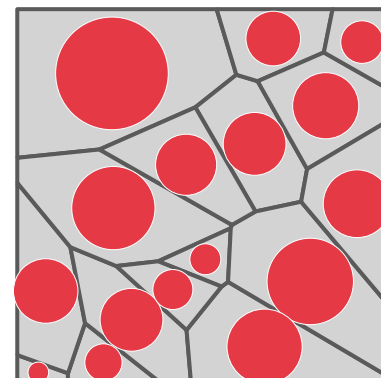
Maillage administratif + points



Agrégation grille



Agrégation administrative



# ✖ 1.2 - EFFET D'ÉCHELLE

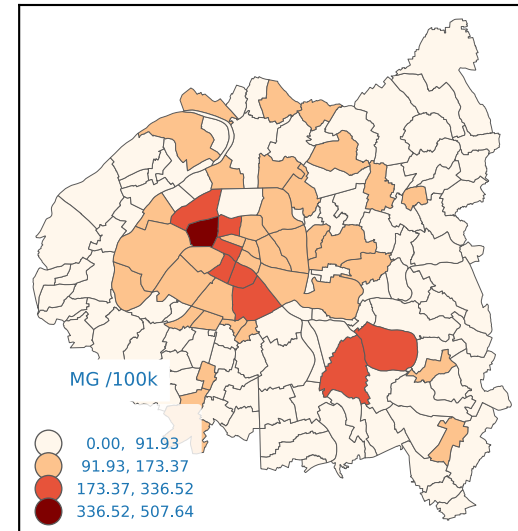
L'effet d'échelle provoque une variation des résultats entre **niveaux d'agrégation différents**.

Exemple de la densité de médecins à Paris PC :

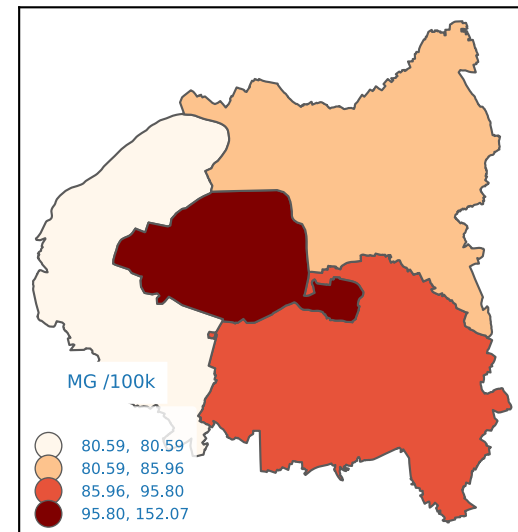
- Niveau communal : forte variabilité.
- Niveau départemental : lissage des variations.

**Échelles différentes = résultats différents !**

Communal (n=143)



Départemental (n=8)

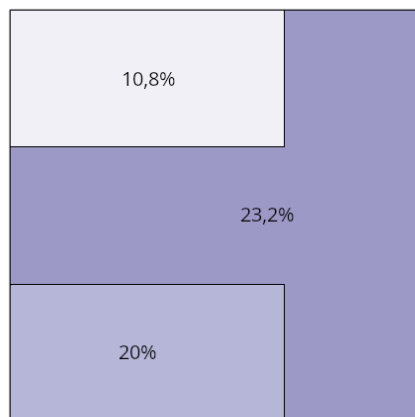


# 1.2 - EXEMPLE DES DEUX EFFETS

Population générale

500	100	700
200	400	100
300	600	200

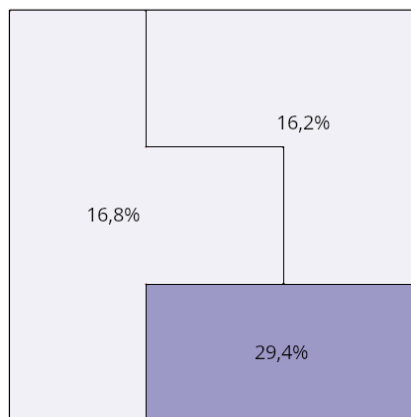
Effet de zonage - configuration 1



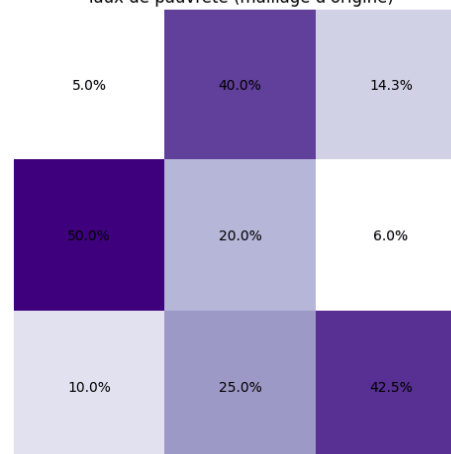
Population pauvre

25	40	100
100	80	6
30	150	85

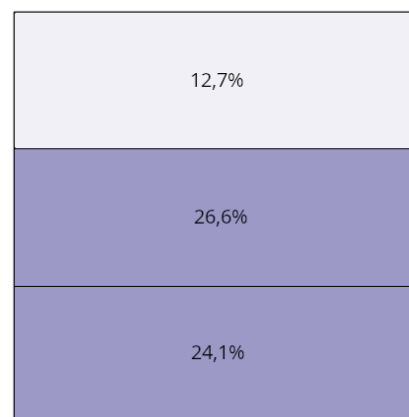
Effet de zonage - configuration 2



Taux de pauvreté (maillage d'origine)



Effet d'échelle - 3 lignes d'agrégat



# 1.3 - EXEMPLE SYNTHÉTIQUE

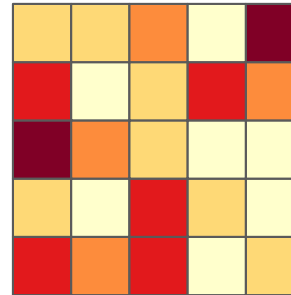
**500 observations** distribuées aléatoirement **Deux variables corrélées** (R 0.70)

**Trois maillages à comparer :**

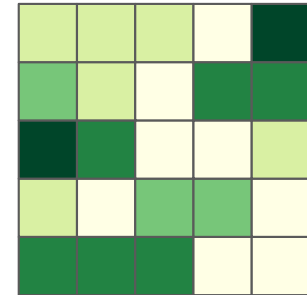
- Grille 10km (25 zones)
- Grille 20km (9 zones)
- Maillage administratif (6 zones)

→ Mêmes données, agrégations différentes

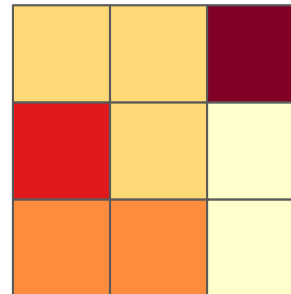
Grille 10km - Variable 1



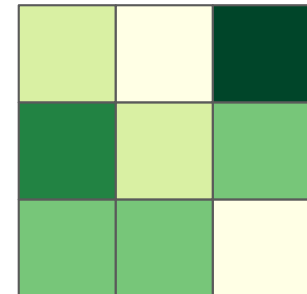
Grille 10km - Variable 2



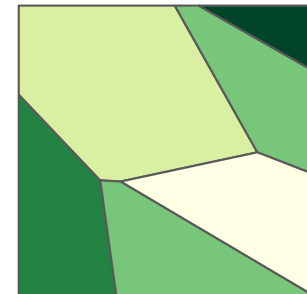
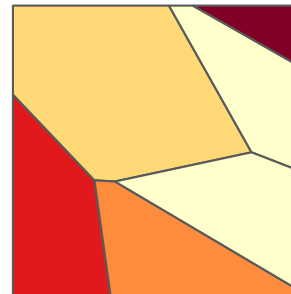
Grille 20km - Variable 1



Grille 20km - Variable 2



Maillage administratif - Variable 1Maillage administratif - Variable 2

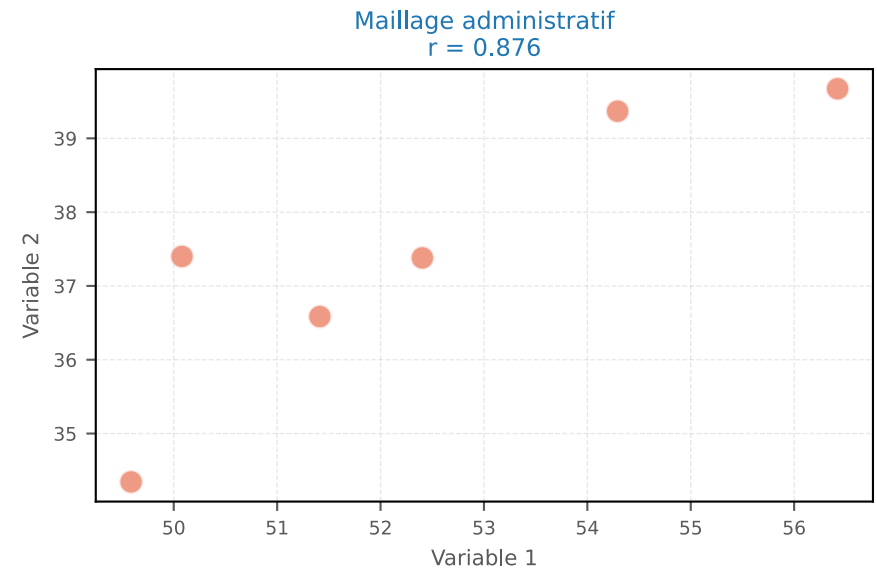
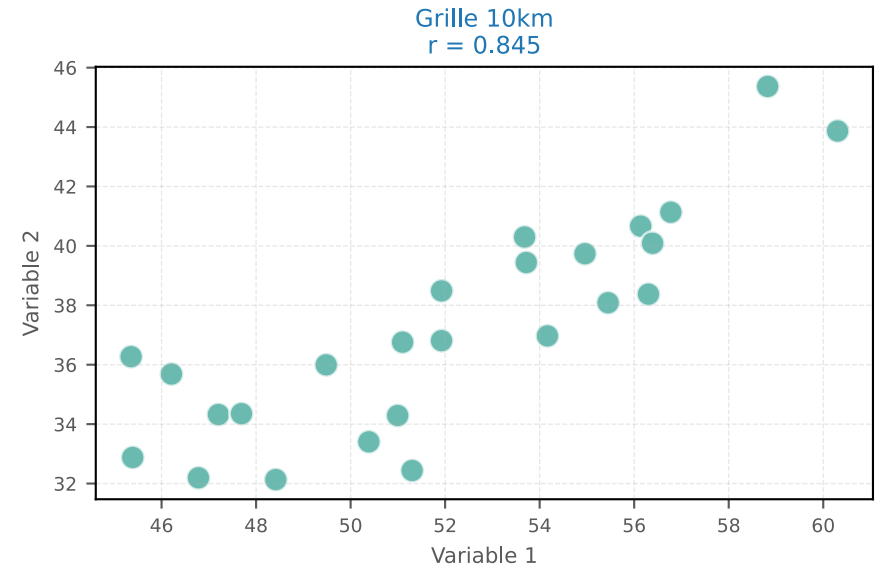


# 1.3 - IMPACT DU ZONAGE

Même ensemble de données,  
**zonages différents :**

- Grille 10km :  $r = 0.845$
- Maillage administratif :  $r = 0.876$

**La forme du zonage modifie la  
corrélation observée !**



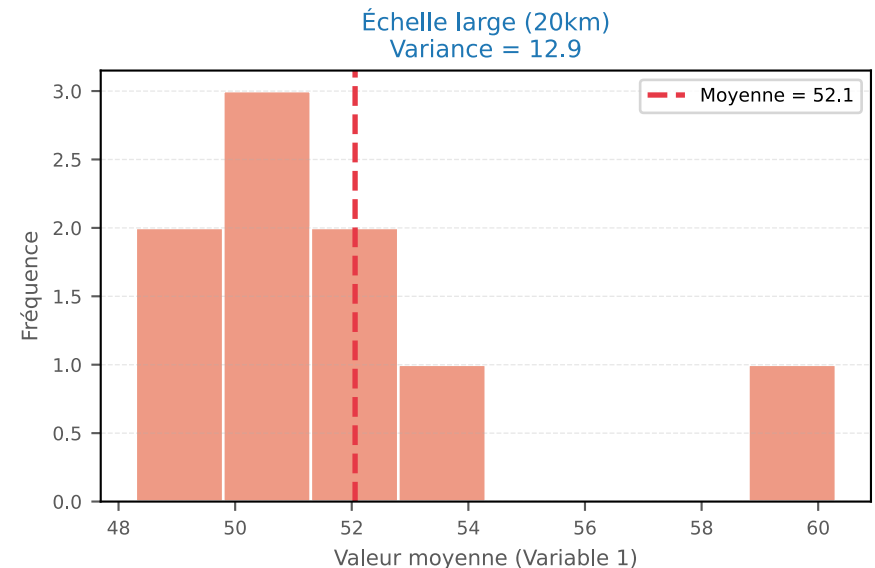
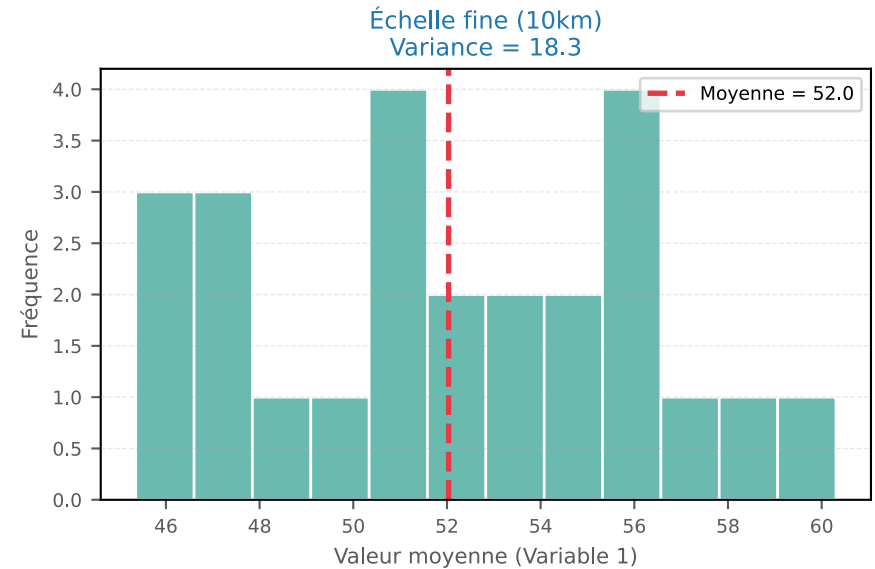
# ✖ 1.3 - IMPACT DE L'ÉCHELLE

Même ensemble de données, même forme (grille), **échelles différentes** :

- Échelle fine (10km) : 25 zones, variance = **18.3**
- Échelle large (20km) : 9 zones, variance = **12.9**

**L'agrégation à une échelle plus large réduit la variance !**

→ Effet de lissage spatial : les valeurs extrêmes sont moyennées



# 1.4 - PARADOXE DE SIMPSON

	Employés	Chômeurs	Total	Chômeurs %
<b>Zone A</b>				
Jeunes	81	9	90	10%
Vieux	9	1	10	10%
<b>Total</b>	90	10	100	<b>10%</b>
<b>Zone B</b>				
Jeunes	40	10	50	20%
Vieux	40	10	50	20%
<b>Total</b>	80	20	100	<b>20%</b>
<b>Zone A+B</b>				
Jeunes	121	19	140	<b>13.6%</b>
Vieux	49	11	60	<b>18.3%</b>
<b>Total</b>	170	30	200	<b>15%</b>

## Données de départ :

- Zone A : 10% de chômage (tous âges)
- Zone B : 20% de chômage (tous âges)

## Après agrégation (A+B) :

- Jeunes : 13,6% ( $\neq$  15%)
- Vieux : 18,3% ( $\neq$  15%)

**Paradoxe** : L'agrégation crée des différences qui n'existaient pas dans les données initiales !

## 2.3.1 - OBSERVATIONS

- Les zones A et B ont chacune une population totale de 100 personnes.
- Dans la zone A, 10% de jeunes et 10% de vieux sont au chômage.
- Dans la zone B, 20% de jeunes et 20% de vieux sont au chômage.
- **La part des chômeurs est donc la même dans les deux zones.**

Si l'on combine les zones A et B, on pourrait s'attendre à ce que le taux de chômage global soit de **15%** pour les deux groupes.

Mais ce n'est pas le cas !



## 2.3.2 - LE BIAIS OBSERVÉ

Résultats observés en combinant A+B :

- Le taux de chômage combiné pour les jeunes est de **13,6%** (au lieu de 15%)
- Le taux de chômage combiné pour les vieux est de **18,3%** (au lieu de 15%)

Ce résultat inattendu est dû à la différence de la taille des populations de chaque groupe dans les zones A et B :

- La zone A a une population **jeunes beaucoup plus importante** que la population âgée
- La zone B a des populations jeunes et âgées **égales**

En combinant les zones, le taux de chômage global est davantage influencé par la zone A, qui a une population jeune plus importante.

## ! 2.3.2 - CONSÉQUENCE DU BIAIS

**Cela crée un biais qui donne l'impression que le taux de chômage est plus faible pour les jeunes que pour les vieux**

Le choix de l'unité spatiale d'agrégation (zone A, zone B ou A+B) a un impact direct sur les statistiques calculées et peut conduire à des conclusions erronées.

Les statistiques agrégées peuvent raconter une histoire complètement différente de la réalité.

## 2.4 - LE GERRYMANDERING

Le **Gerrymandering** est une pratique qui consiste à :

- manipuler les frontières des circonscriptions électorales
- afin de favoriser un parti politique ou un groupe particulier.



## 2.4 - TECHNIQUES DE MANIPULATION

### CONCENTRATION (PACKING)

Concentrer les électeurs d'un parti adverse

---

Réduire leur influence dans les autres districts

---

Sacrifier quelques districts pour gagner ailleurs

### DISPERSION (CRACKING)

Disperser les électeurs d'un parti adverse

---

Empêcher d'obtenir une majorité dans chaque district

---

Diluer leur poids électoral

### FORMES BIZARRES

Districts de formes non compactes

---

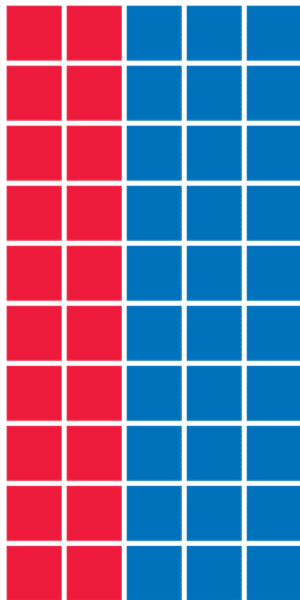
Regrouper les électeurs favorables

---

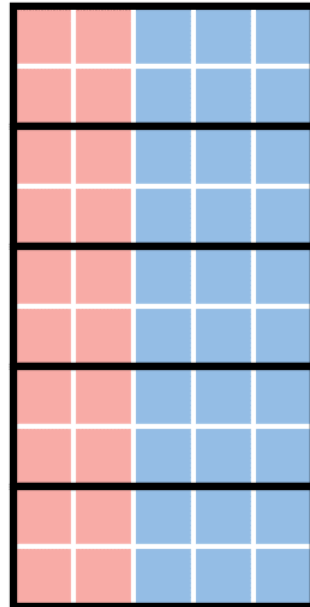
Créer des circonscriptions géométriquement étranges

## 2.4.1 - DANS LE PRINCIPE

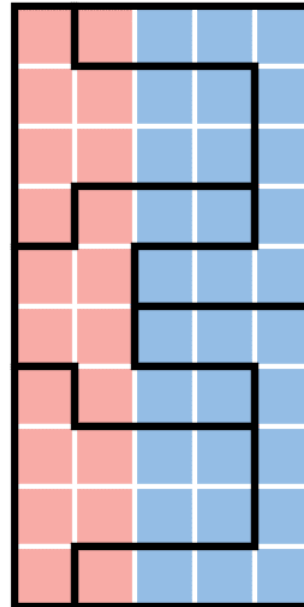
### HOW TO STEAL AN ELECTION



50 PRECINCTS  
**60% BLUE**  
**40% RED**



5 DISTRICTS  
5 BLUE  
0 RED  
**BLUE WINS**



5 DISTRICTS  
3 RED  
2 BLUE  
**RED WINS**

### Effet de zonage appliqué à l'électoral

Même distribution d'électeurs :

- Bleus : 60%
- Rouges : 40%

Résultats selon le découpage :

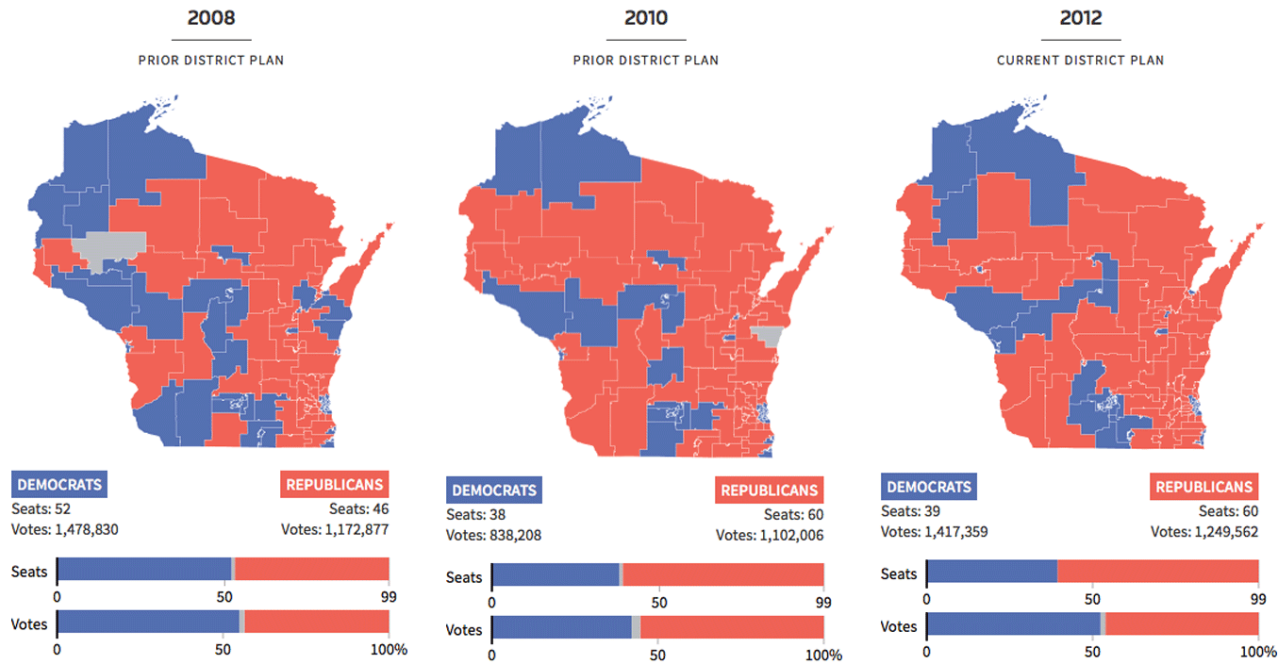
- Les rouges peuvent gagner la majorité malgré leur minorité

## 🌐 2.4.2 - DANS LA RÉALITÉ : WISCONSIN

Observez :

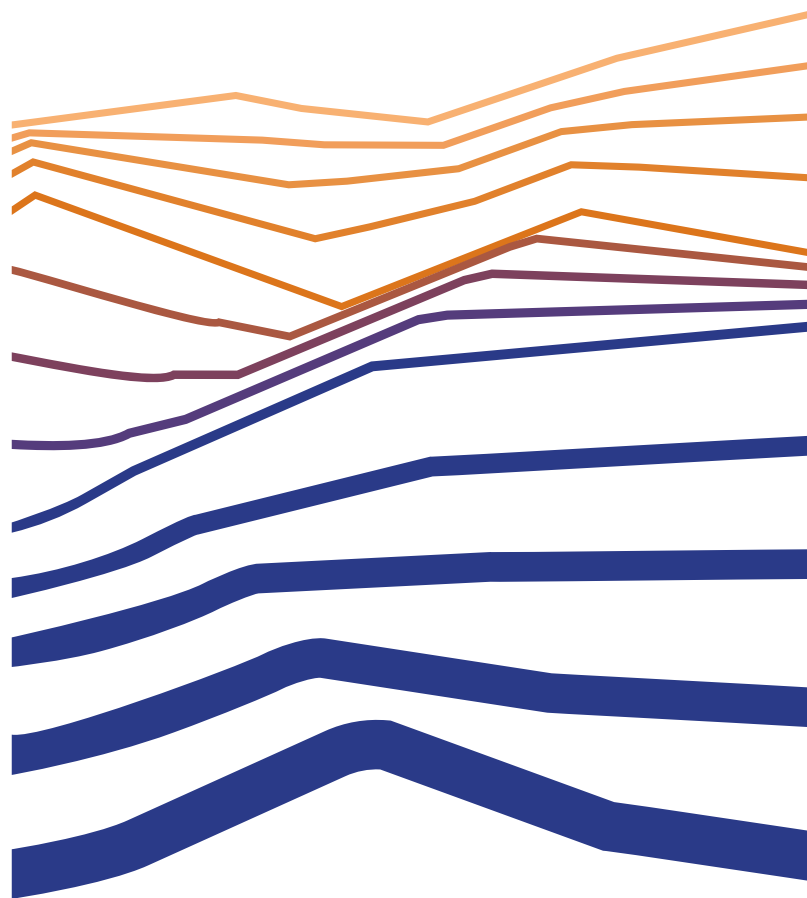
- La forme des circonscriptions
- Les résultats électoraux vs votes

La manipulation des frontières spatiales peut complètement changer les résultats !



# 3 - SOLUTIONS

POUR LIMITER LE  
MAUP



## 3 - SOLUTIONS POUR LIMITER LE MAUP

Le MAUP présente un défi en analyse spatiale. Plusieurs approches permettent de minimiser son influence et d'obtenir des résultats plus robustes.

### Cinq approches principales :

1. Sensibilisation au problème
2. Utilisation de données à échelle fine
3. Analyse de sensibilité
4. Techniques de régression adaptées
5. Lissage spatial et géographique



## 3.1 - SENSIBILISATION AU PROBLÈME

### OBJECTIF

Faire prendre conscience du problème

---

Première étape pour minimiser l'impact du MAUP

### PRINCIPE

Comprendre que le choix des unités spatiales influence les résultats

---

Reconnaître l'impact de la forme et de la taille des zones

---

Éviter les interprétations naïves des données agrégées

### MISE EN PRATIQUE

Former les analystes aux enjeux du MAUP

---

Documenter les choix de zonage dans les rapports

---

Questionner systématiquement la robustesse des résultats

## 3.2 - DONNÉES À ÉCHELLE FINE

### PRINCIPE

Utiliser des données à l'échelle la plus fine possible

---

Minimiser l'impact de l'agrégation

### AVANTAGES

Conservation de la précision spatiale

---

Évite l'agrégation excessive

---

Permet de choisir le niveau d'agrégation optimal

### LIMITES

Puissance statistique insuffisante avec petits maillages

---

Problèmes de confidentialité des données individuelles

---

Coût d'acquisition et de traitement des données fines

## 3.3 - ANALYSE DE SENSIBILITÉ

### PRINCIPE

Évaluer l'impact du choix des unités spatiales sur les résultats

---

Approche importante pour tester la robustesse

### MÉTHODE

Effectuer des analyses avec différentes configurations d'unités spatiales

---

Comparer les résultats pour identifier les zones de sensibilité

---

Utiliser le rapport de variance pour étudier l'effet de la configuration spatiale

### RÉSULTAT

Quantification de la stabilité des résultats

---

Identification des paramètres sensibles au MAUP

---

Aide à la décision pour le choix du zonage

## 3.4 - TECHNIQUES DE RÉGRESSION

### PRINCIPE

Réduire les biais du MAUP dans les analyses de régression

---

Approches adaptées aux données spatiales

### RÉGRESSION SPATIALE PONDÉRÉE

Prend en compte l'autocorrélation spatiale

---

Pondère les observations selon leur localisation

---

Réduit les biais liés à l'agrégation spatiale

### MODÈLES HIÉRARCHIQUES BAYÉSIENS

Combinent données agrégées et données individuelles

---

Modélisent les relations entre variables à différents niveaux d'agrégation

---

Tiennent compte de la structure hiérarchique des données spatiales

---

Permettent l'inférence écologique robuste

## 3.5 - LISSAGE GÉOGRAPHIQUE - PRINCIPE

### DÉFINITION

Lissage spatial appliqué aux cartes

---

Représentations cartographiques simplifiées

### MÉTHODE

Valeur observée remplacée par moyenne pondérée

---

Calcul sur le voisinage dans un rayon défini

### MATRICE DE VOISINAGE

Matrice de voisinage d'ordre  $n$

---

Pondération inverse de l'ordre

---

Fonction de pondération = compromis biais-variance

# ☒ EXEMPLE THÉORIQUE : MATRICE DE VOISINAGE

Ordres de voisinage (cellule centrale)

10	15	20	25	18
12	18	25	22	15
15	22	30	24	18
13	20	26	20	16
11	16	21	17	14

Cellule cible  
Voisins ordre 1  
Ordre 2+

## Application de la pondération

Cellule centrale : 30  
Voisins (ordre 1) : [22, 25, 24, 20, 26, 22, 24, 18]  
Moyenne des voisins : 22.6

Pondération 50%-50% :  
 $V_{\text{lissée}} = 0.5 \times 30 + 0.5 \times 22.6$   
 $V_{\text{lissée}} = 26.3$

Pondération 75%-25% :  
 $V_{\text{lissée}} = 0.75 \times 30 + 0.25 \times 22.6$   
 $V_{\text{lissée}} = 28.2$

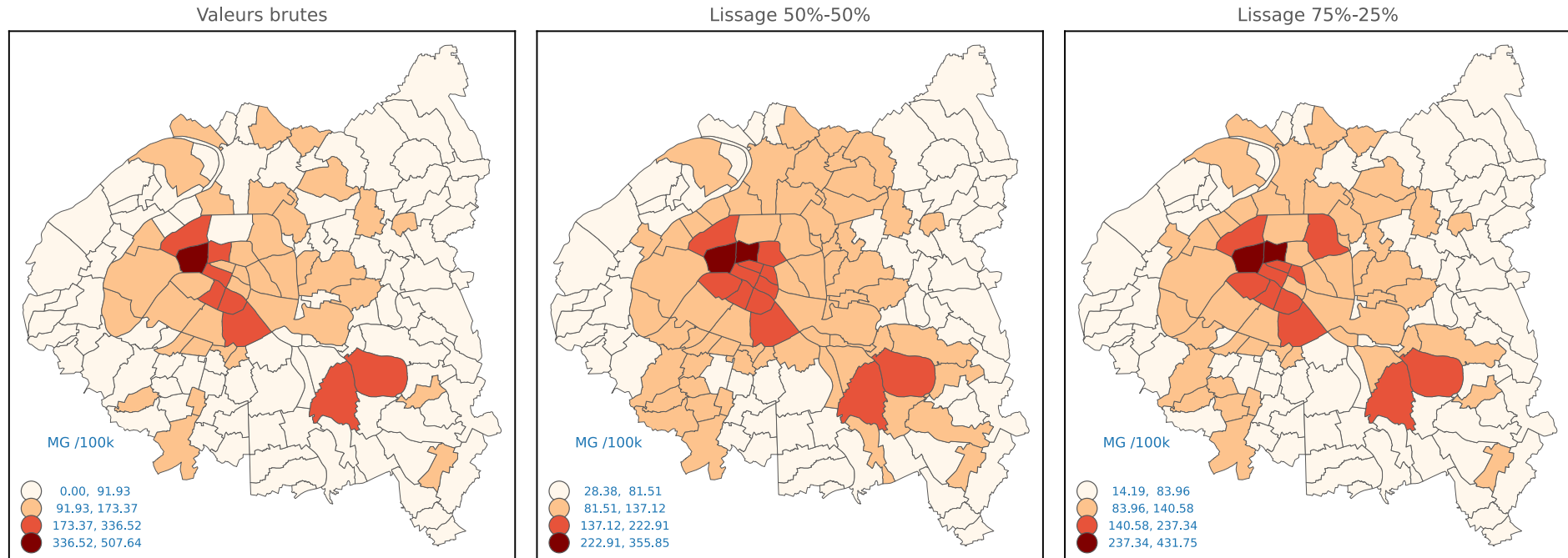
Impact : Plus le poids de la cellule centrale est élevé, moins le lissage est fort.

**Ordre de voisinage** : Les cellules adjacentes (partageant un côté ou un coin) sont des voisins d'ordre 1.

**Pondération** : Le paramètre clé contrôle l'équilibre entre la valeur propre et la moyenne des voisins.

**Plus le poids est élevé pour la cellule centrale, moins le lissage est prononcé.**

# ❧ EXEMPLE : LISSAGE AVEC MATRICE DE VOISINAGE



**Matrice de voisinage Queen** : Contiguïté par côtés et coins.

**50%-50%** : Lissage équilibré. **75%-25%** : Lissage plus conservateur (préserve davantage les valeurs locales).

**Impact** : Plus le poids de la cellule centrale est élevé, moins le lissage est prononcé. Le choix de la pondération dépend de l'objectif analytique.

## └┐ 3.6 - LE LISSAGE SPATIAL

Le **lissage spatial** est une méthode d'estimation non paramétrique de la fonction d'intensité de données ponctuelles.

Il permet de révéler des structures spatiales sous-jacentes en filtrant l'information.

**S'affranchit des découpages administratifs arbitraires !**



# 3.6 - PARAMÈTRES DU LISSAGE SPATIAL

Il s'agit d'une modélisation locale qui repose sur le choix de **deux paramètres clés** :

## 1. LE KERNEL

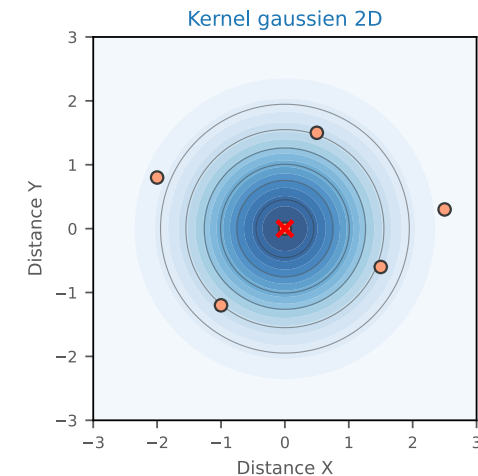
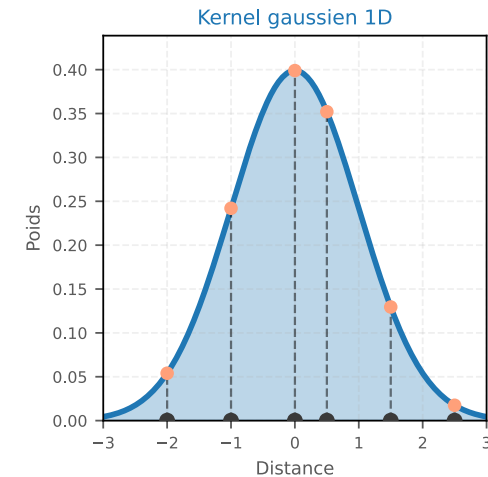
Le kernel décrit la façon dont le voisinage est appréhendé (type de fonction à l'intérieur de la fenêtre) : gaussien, uniforme, triangulaire, etc.

## 2. LA BANDWIDTH

La bandwidth quantifie la « taille » du voisinage.

**Le choix de la bandwidth est un arbitrage** entre la précision spatiale de l'analyse et sa qualité statistique (compromis biais-variance).

On peut utiliser un **histogramme de Moran** afin de choisir la distance juste **avant la chute la plus importante** de l'autocorrélation spatiale (avant la plus faible structure spatiale).

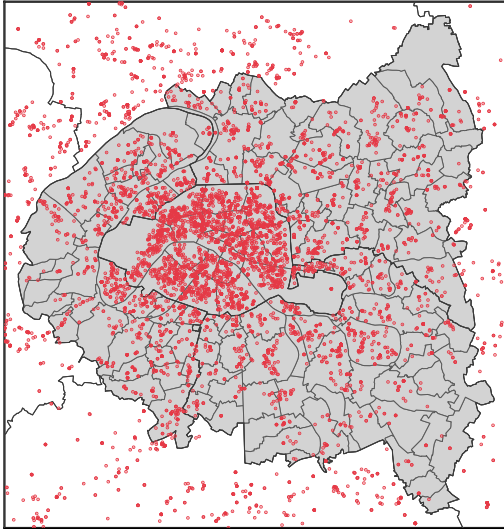




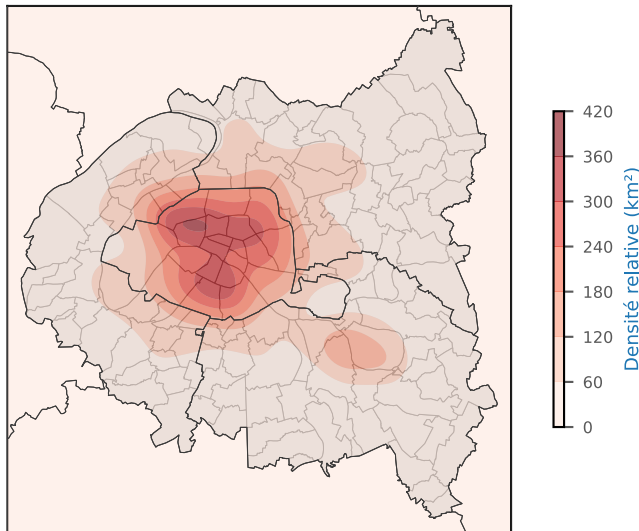
# EXEMPLE DE LISSAGE SPATIAL

## DONNÉES PONCTUELLES

Médecins généralistes géocodés



Lissage spatial (KDE)



Les médecins généralistes sont géocodés à l'adresse exacte.

## APRÈS LISSAGE (KDE)

On applique un Kernel Density Estimator (KDE) :

- Fonction gaussienne
- Bandwidth 200m
- Révèle la structure spatiale sans découpage arbitraire

## ⇄ 3.7 - LISSAGE SPATIAL VS GÉOGRAPHIQUE

### LISSAGE SPATIAL

S'affranchit des découpages administratifs

---

Révèle les structures spatiales continues

---

Estimation non paramétrique de l'intensité

### LISSAGE GÉO

Pondération par les voisins

---

Appliqué aux cartes avec zonage existant

---

Matrice de voisinage d'ordre n

### CHOIX DE LA BANDWIDTH

Histogramme de Moran pour choisir la bandwidth

---

Prendre la valeur juste avant la plus grande chute d'autocorrélation spatiale

---

Compromis biais-variance

### LIMITES

Atténuent les ruptures mais peuvent masquer des phénomènes réels

---

Nécessitent une expertise pour paramétrer correctement

---

Dépendance au zonage administratif initial

# CONCLUSION

## LE MAUP EN PRATIQUE



# CONCLUSION - LE MAUP EN PRATIQUE

## NATURE DU PROBLÈME

Présent dans toute analyse spatiale

---

Forme et taille des unités modifient les résultats

---

Peut conduire à des interprétations erronées

## IMPACTS OBSERVÉS

Modifie statistiques descriptives

---

Affecte analyses de corrélation

---

Impact sur inférences statistiques

## APPROCHES PRATIQUES

Stratégies de minimisation nécessaires

---

Analyse de sensibilité recommandée

---

Documentation des choix méthodologiques

# ✓ CONCLUSION - RECOMMANDATIONS PRATIQUES

## APPROCHE CONTEXTUELLE

Solution dépend du contexte

---

Adapter selon la question de  
recherche

---

Tenir compte des données  
disponibles

## TESTS DE ROBUSTESSE

Combiner plusieurs méthodes

---

Tester différentes échelles

---

Analyser la sensibilité des résultats

## DOCUMENTATION

Documenter les choix  
méthodologiques

---

Rapporter l'impact potentiel du  
MAUP

---

Assurer la reproductibilité de  
l'analyse

## BIBLIOGRAPHIE - OUVRAGES DE RÉFÉRENCE

**Openshaw, S. (1984).** *The Modifiable Areal Unit Problem*. Geo Books, Norwich.

Ouvrage fondateur qui explore en profondeur le MAUP, ses implications et ses effets sur les analyses spatiales.

---

**Fotheringham, A. S., & Wong, D. W. S. (1991).** "The Modifiable Areal Unit Problem in Multivariate Statistical Analysis". *Environment and Planning A*, 23(7), 1025-1044.

Article clé qui étudie le MAUP dans le contexte d'analyses multivariées, avec des exemples de l'impact du MAUP.

# BIBLIOGRAPHIE - ANALYSES SPATIALES

**Bailey, T. C., & Gatrell, A. C. (1995).** *Interactive Spatial Data Analysis*. Longman, Essex.

Référence pour les techniques de lissage et les statistiques spatiales appliquées.

---

**Silverman, B. W. (1986).** *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Ouvrage de base sur l'estimation de densité, qui décrit les méthodes de lissage par noyau (KDE).

---

**Waller, L. A., & Gotway, C. A. (2004).** *Applied Spatial Statistics for Public Health Data*. Wiley, Hoboken.

Un manuel détaillant les approches statistiques spatiales, y compris le lissage et la correction des effets de bord.



# BIBLIOGRAPHIE - TECHNIQUES DE RÉGRESSION

**Anselin, L. (1988).** *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.

Ouvrage de référence pour les méthodes de régression spatiale.

---

**Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014).** *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC.

Exploration des modèles hiérarchiques bayésiens et de leurs applications en analyses spatiales.

---

**Dark, S. J., & Bram, D. (2007).** "The Modifiable Areal Unit Problem (MAUP) in Physical Geography". *Progress in Physical Geography*, 31(5), 471-479.

Article de synthèse qui explique les effets du MAUP en géographie physique.