

SIX NUANCES D'A-GREY-GATION

L'influence du choix des unités spatiales d'agrégation sur la lecture et l'interprétation des résultats cartographiés



Jean-Philippe Antoni¹, Émilie Lerond¹, Stéphane Moisy² et Olivier Klein³

¹ Laboratoire ThéMA, UMR 6049, Dijon (F)

² OTE Ingénierie, Strasbourg (F)

³ Luxembourg Institute of Socio-Economic Research, Belval (L)

13^e Rencontres de Théo Quant – Besançon – 17-19 mai 2017



1. INTRODUCTION

1.1. Contexte | 1.2. Problématique | 1.3. Hypothèses et objectif

2. METHODOLOGIE

2.1. Construction des unités d'agrégation | 2.2. Évaluation

3. RÉSULTATS

3.1. Cartographie | 3.2. Enquête | 3.3. Indicateurs

4. CONCLUSION ET PERSPECTIVES

1.1. Contexte

Une nouvelle génération de données

Massives

Désagrégées

Multidimensionnelles

1.1. Contexte

Une nouvelle génération de données

Massives

Désagrégées

Multidimensionnelles



Une richesse

Spatiale

Thématique

Temporelle

1.1. Contexte

Une nouvelle génération de données

Massives

Désagrégées

Multidimensionnelles



Une richesse

Spatiale

Thématique

Temporelle



Une nouvelle difficulté

Comment visualiser des milliers d'objets/individus en mouvement de manière à ce que l'information reste lisible, mais que l'on puisse également en dégager une structure spatialement interprétable ?

1.1. Contexte

Une nouvelle génération de données

Massives

Désagrégées

Multidimensionnelles



Une richesse

Spatiale

Thématique

Temporelle



Une nouvelle difficulté

Comment visualiser des milliers d'objets/individus en mouvement de manière à ce que l'information reste lisible, mais que l'on puisse également en dégager une structure spatialement interprétable ?



Interpolation

Heat maps (carte de densité)

Agrégation

1.1. Contexte

Une nouvelle génération de données

Massives
Désagrégées
Multidimensionnelles



Une richesse

Spatiale
Thématique
Temporelle



Une nouvelle difficulté

Comment visualiser des milliers d'objets/individus en mouvement de manière à ce que l'information reste lisible, mais que l'on puisse également en dégager une structure spatialement interprétable ?



Interpolation
Heat maps (carte de densité)
Agrégation

Un paradoxe

L'agrégation spatiale reste nécessaire pour visualiser des données dont l'un des principaux intérêts repose au départ sur leur caractère désagrégé

Illustration du paradoxe

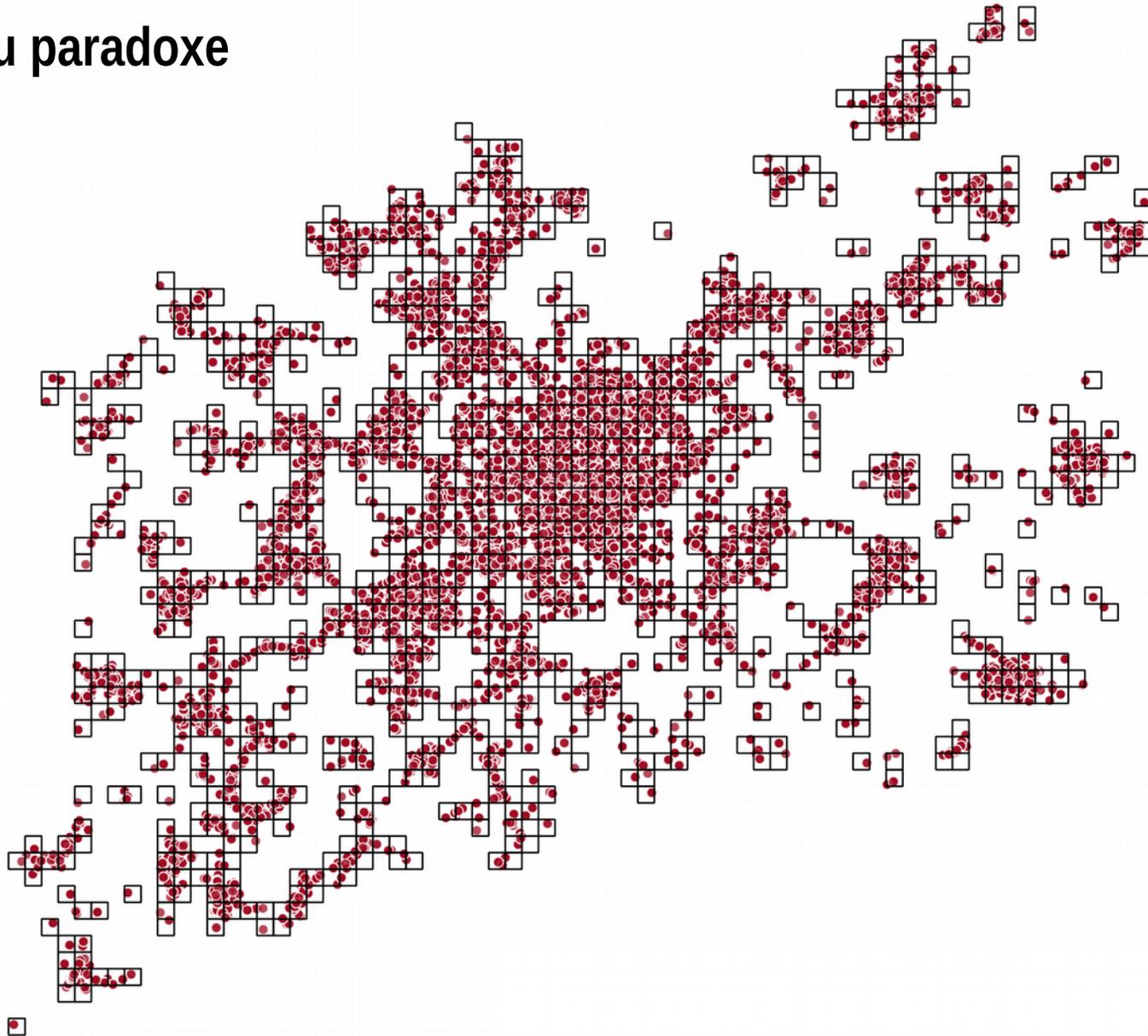
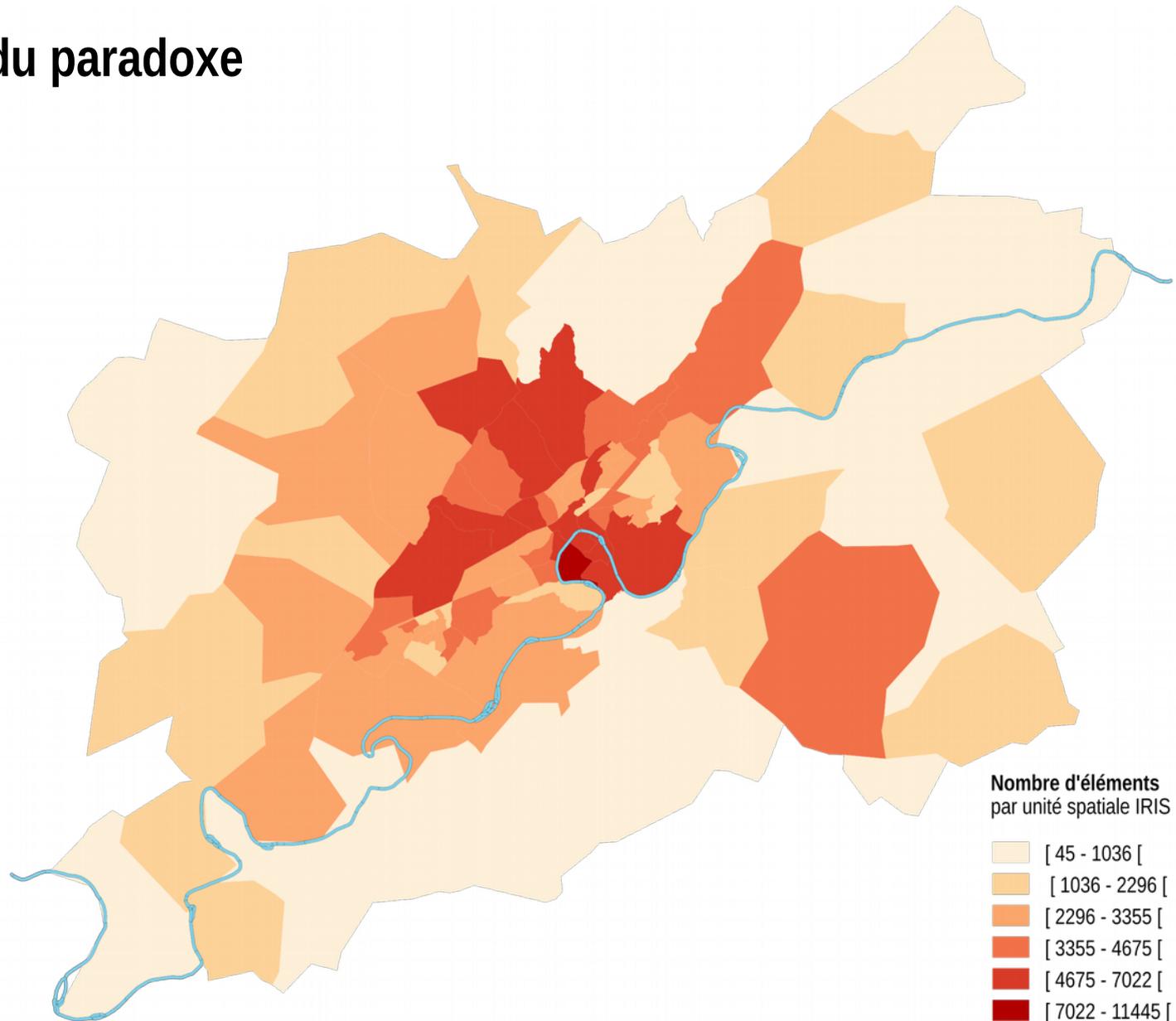


Illustration du paradoxe



1.2. Problématique

L'agrégation est nécessaire. Elle s'appuie sur un double choix

1. La **taille** des unités spatiales au sein desquelles les données seront regroupées, ce qui revient à déterminer la résolution des résultats (problématique du carroyage) (cf. Delahaie, 1987 ; Cauvin, 2008)



Variation de taille

2. La **forme** de ces unités spatiales, selon une logique qui relève du MAUP (cf. Openshaw, 1983 ; Epson, 2006)



Variation de forme

Un degré de liberté supplémentaire

Les données désagrégées ne sont contraintes par aucune structure au départ

1.2. Hypothèses et objectif

Hypothèses

1. Pour un jeu de données identique, la forme des unités spatiales d'agrégation influence les résultats quantifiés et cartographiés
2. Cette même forme influence la lecture et l'interprétation des résultats et peut conduire à des conclusions fallacieuses ou contradictoires

1.2. Hypothèses et objectif

Hypothèses

1. Pour un jeu de données identique, la forme des unités spatiales d'agrégation influence les résultats quantifiés et cartographiés
2. Cette même forme influence la lecture et l'interprétation des résultats et peut conduire à des conclusions fallacieuses ou contradictoires

Objectifs

1. **Vérifier** ces deux hypothèses à partir d'un exemple concret
2. Mieux **comprendre** les enjeux de la représentation de données individuelles désagrégées
3. Envisager une méthodologie pour **généraliser** la démarche à n'importe quel cas d'étude

1. INTRODUCTION

1.1. Contexte | 1.2. Problématique | 1.3. Hypothèses et objectif

2. METHODOLOGIE

2.1. Construction des unités d'agrégation | 2.2. Évaluation

3. RÉSULTATS

3.1. Cartographie | 3.2. Enquête | 3.3. Indicateurs

4. CONCLUSION ET PERSPECTIVES

2.1. Construction des unités d'agrégation

Terrain d'étude et données : sans importance

Trois contraintes

1. Couverture de l'ensemble de la zone d'un seul tenant et sans enclave
2. Nombre d'unités spatiales identiques (autant que possible)
3. Taille des unités spatiales identiques (à assouplir pour D, E et F)

2.1. Construction des unités d'agrégation

Terrain d'étude et données : sans importance

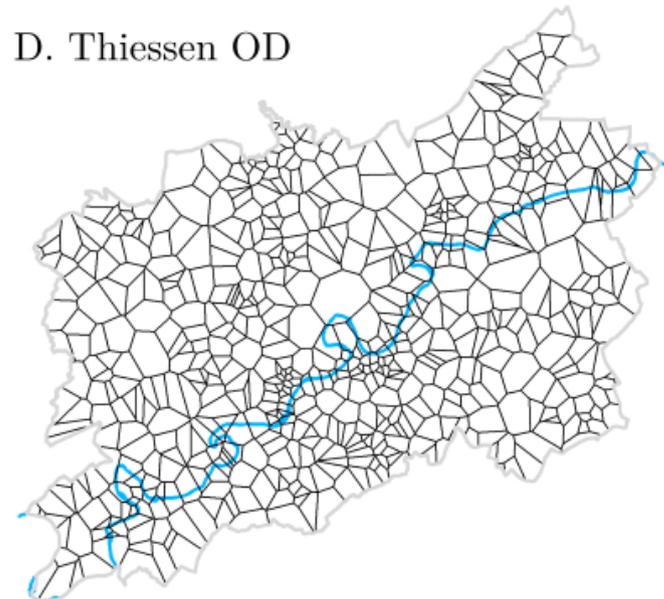
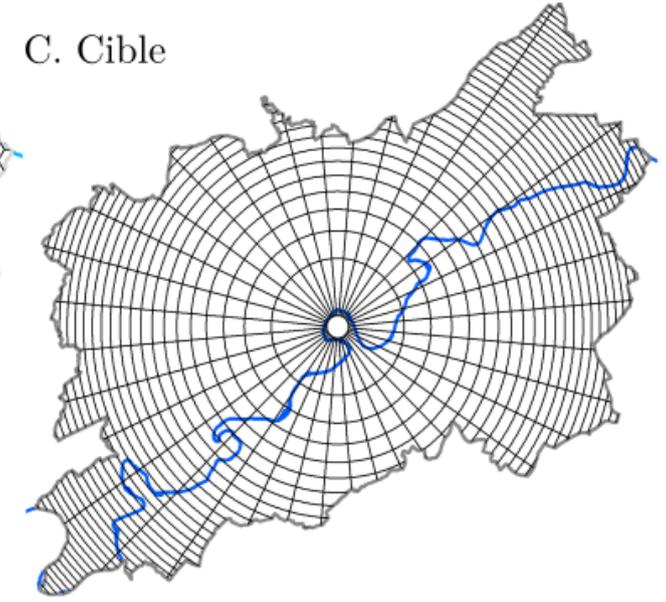
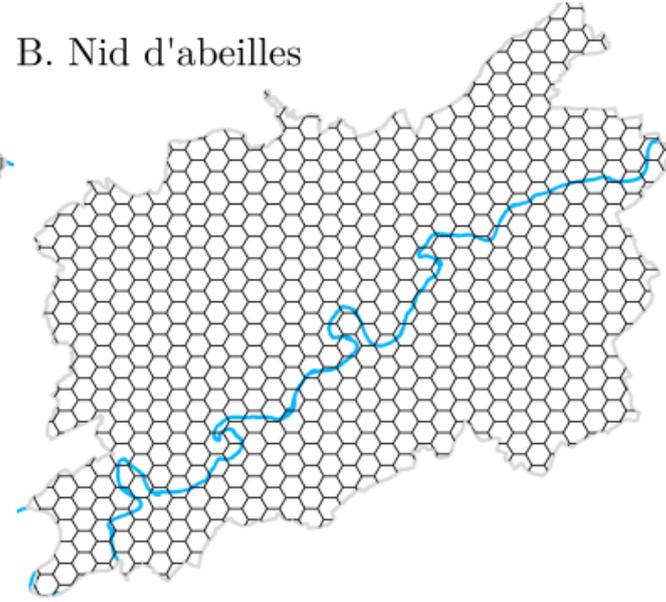
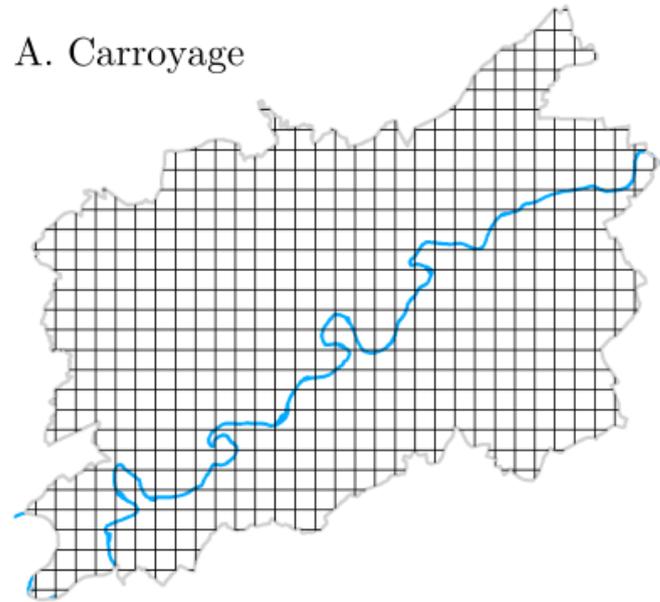
Trois contraintes

1. Couverture de l'ensemble de la zone d'un seul tenant et sans enclave
2. Nombre d'unités spatiales identiques (autant que possible)
3. Taille des unités spatiales identiques (à assouplir pour D, E et F)

Six ensembles

A. Carroyage	628 cellules	$S=806\ 192\ \text{m}^2$
B. Nid d'abeilles	624 cellules	$S=806\ 192\ \text{m}^2$
C. Cible	655 cellules	$S=806\ 154\ \text{m}^2$
D. Pol. Thiessen (centroïdes)	626 cellules	$18\ 310 < S < 4\ 652\ 990\ \text{m}^2$
E. Pol. Thiessen (aléatoire 1)	626 cellules	$57\ 950 < S < 3\ 233\ 065\ \text{m}^2$
F. Pol. Thiessen (aléatoire 2)	626 cellules	$38\ 645 < S < 2\ 822\ 297\ \text{m}^2$

2. MÉTHODOLOGIE



2.2. Évaluation

Deux indicateurs avancés

(cf. Jenks and Caspall, 1971 ; Smith, 1986 ; Armstrong et al., 2003)

Goodness of variance fit

Compare la variance totale du jeu de données
à la variance interne des classes d'intervalle
 $0 < GVF < 1$

$$GVF = 1 - \frac{\sum_{j=1}^k \sum_{i=1}^{N_j} (z_{ij} - \bar{z}_j)^2}{\sum_{i=1}^{N_j} (z_i - \bar{z}_j)^2}$$

Overview accuracy index

Idem mais en incluant la surface des unités spatiales
Comparées à la surface totale de l'aire d'étude
 $0 < OAI < 1$

$$OAI = 1 - \frac{\sum_{j=1}^k \sum_{i=1}^{N_j} |z_{ij} - \bar{z}_j| S_{ij}}{\sum_{i=1}^{N_j} |z_i - \bar{z}_j| s_i}$$

2.2. Évaluation

Deux indicateurs avancés

(cf. Jenks and Caspall, 1971 ; Smith, 1986 ; Armstrong et al., 2003)

Goodness of variance fit

Compare la variance totale du jeu de données
à la variance interne des classes d'intervalle
 $0 < GVF < 1$

$$GVF = 1 - \frac{\sum_{j=1}^k \sum_{i=1}^{N_j} (z_{ij} - \bar{z}_j)^2}{\sum_{i=1}^{N_j} (z_i - \bar{z}_j)^2}$$

Overview accuracy index

Idem mais en incluant la surface des unités spatiales
Comparées à la surface totale de l'aire d'étude
 $0 < OAI < 1$

$$OAI = 1 - \frac{\sum_{j=1}^k \sum_{i=1}^{N_j} |z_{ij} - \bar{z}_j| S_{ij}}{\sum_{i=1}^{N_j} |z_i - \bar{z}_j| s_i}$$

Un « détournement »

GVF et OAI ne sont pas utilisés pour tester différentes méthodes de discrétisation sur un même jeu de données, mais une même méthode de discrétisation (Jenks) sur différents jeux de données

1. INTRODUCTION

1.1. Contexte | 1.2. Problématique | 1.3. Hypothèses et objectif

2. METHODOLOGIE

2.1. Construction des unités d'agrégation | 2.2. Évaluation

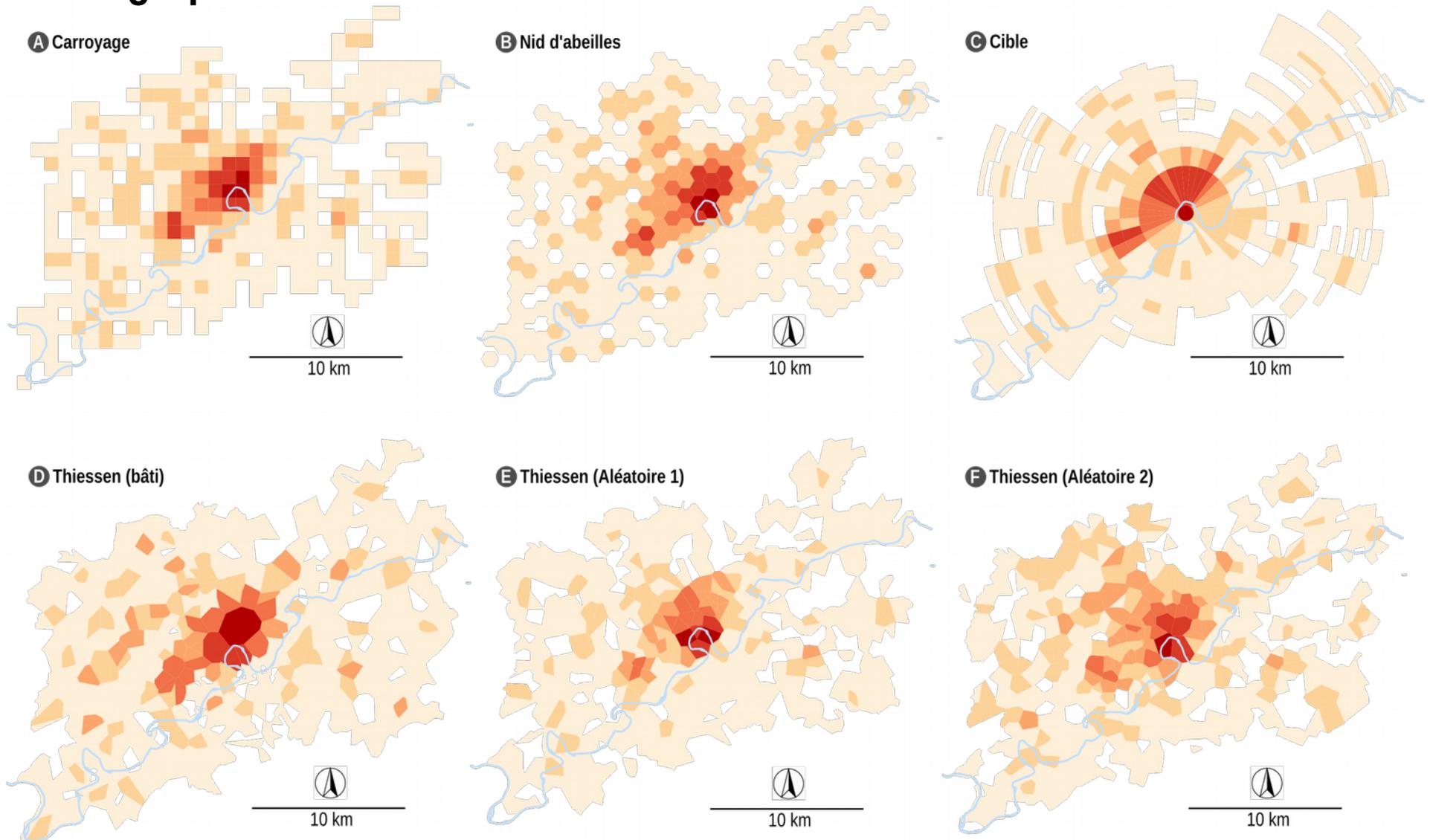
3. RÉSULTATS

3.1. Cartographie | 3.2. Enquête | 3.3. Indicateurs

4. CONCLUSION ET PERSPECTIVES

3. RÉSULTATS

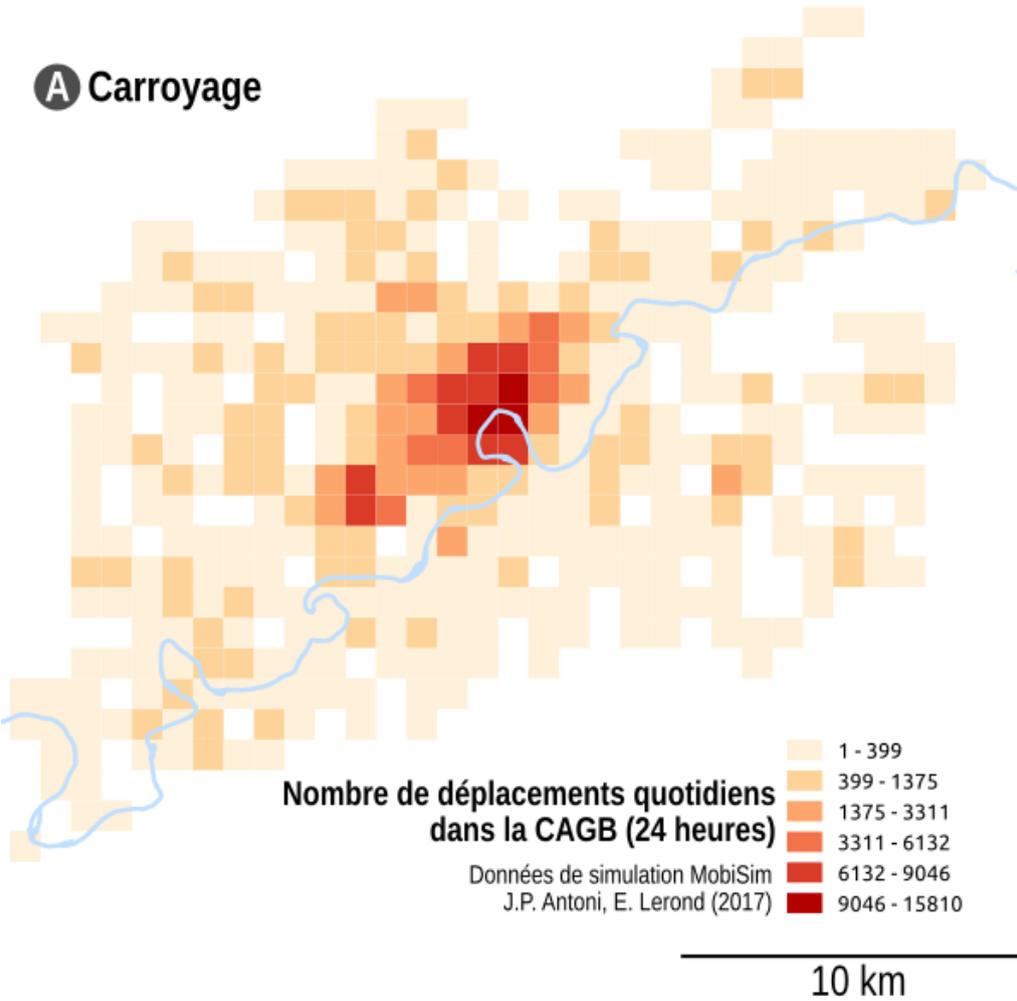
3.1. Cartographie



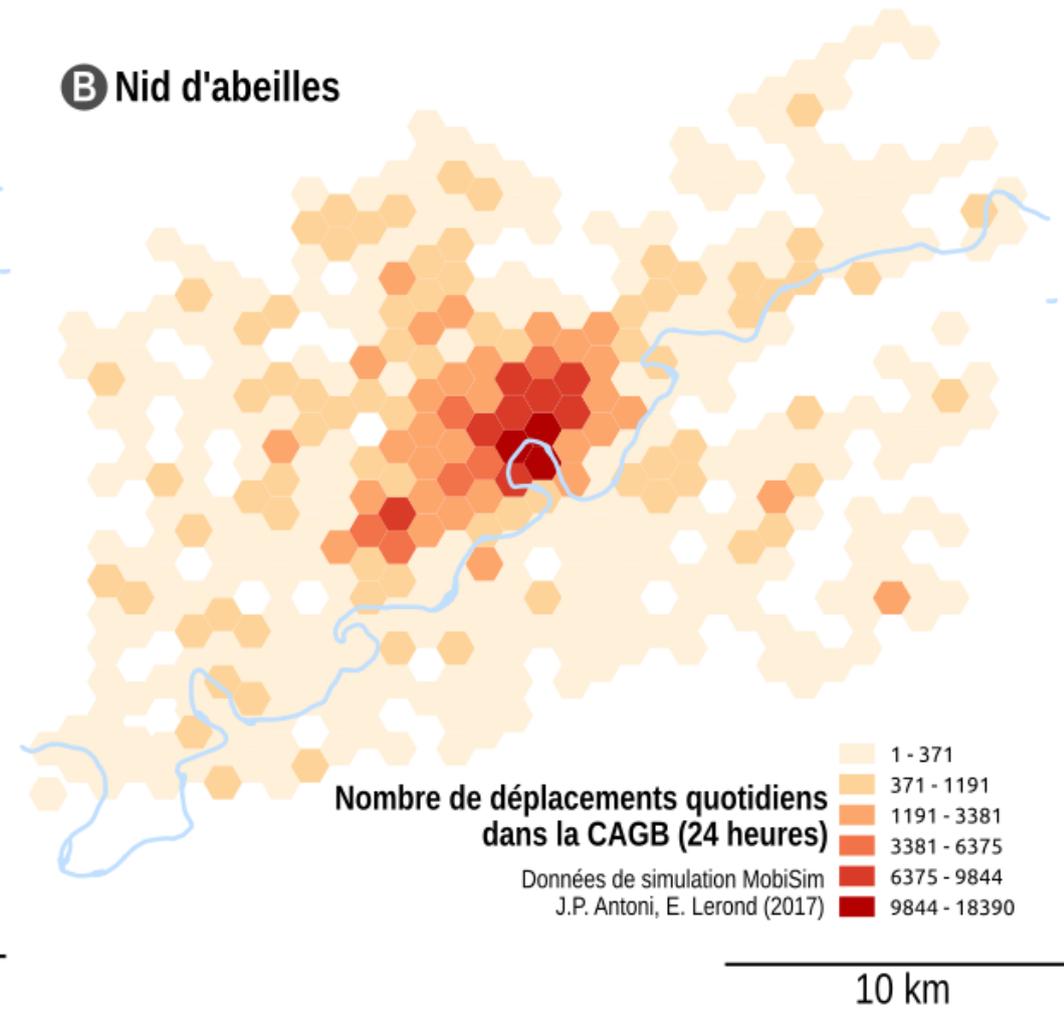
3. RÉSULTATS

3.1. Cartographie

A Carroyage



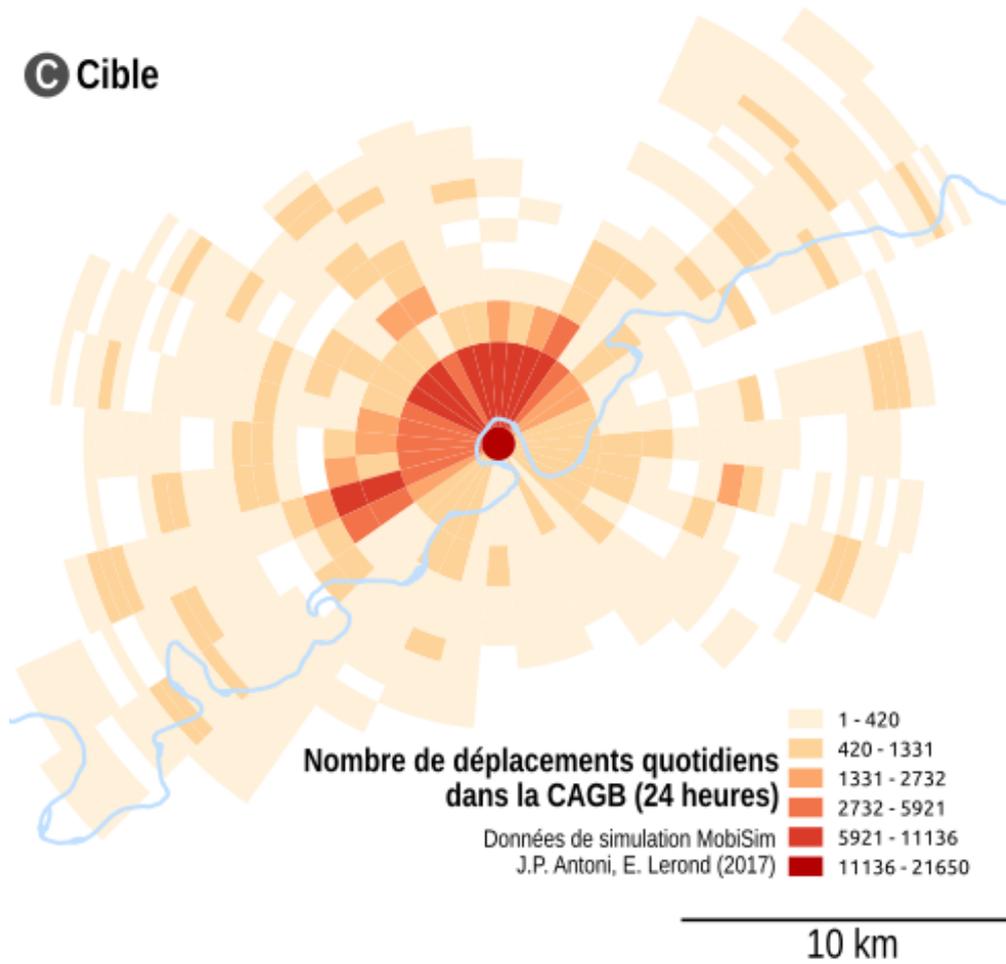
B Nid d'abeilles



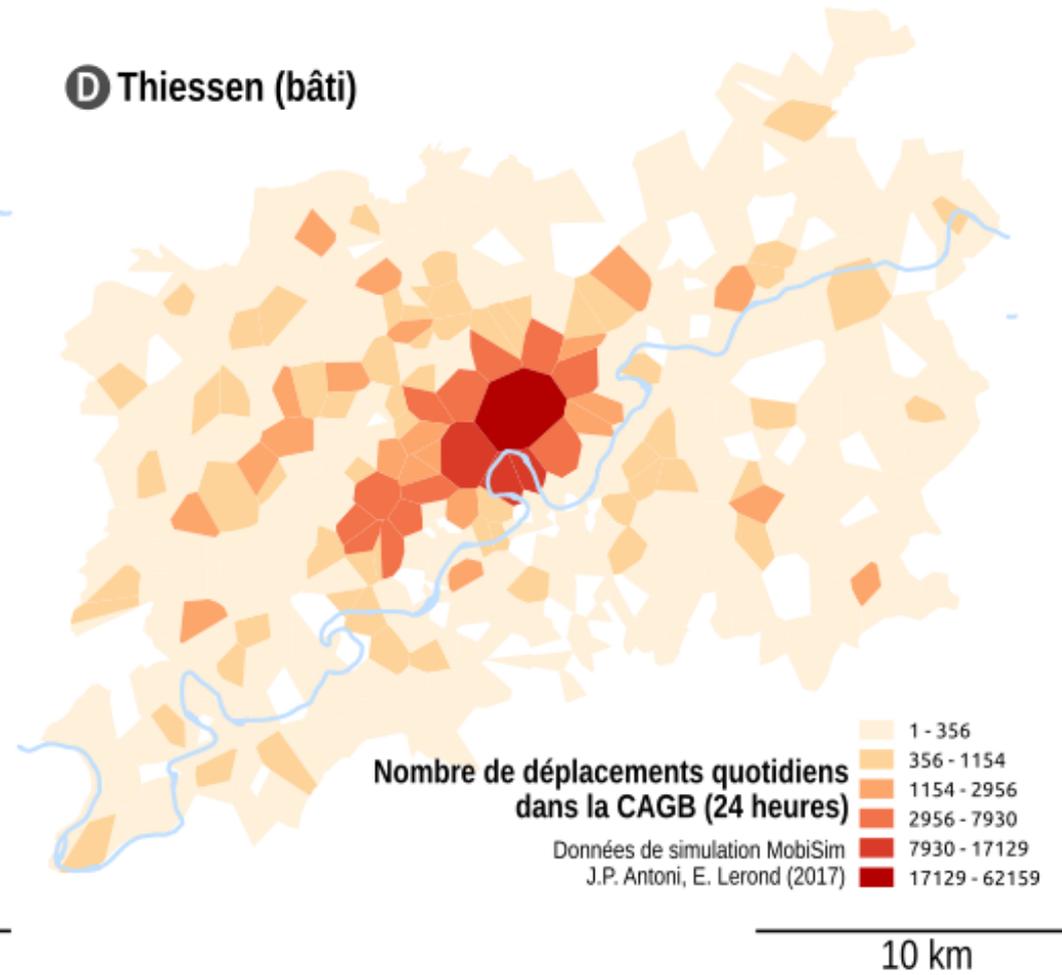
3. RÉSULTATS

3.1. Cartographie

C Cible



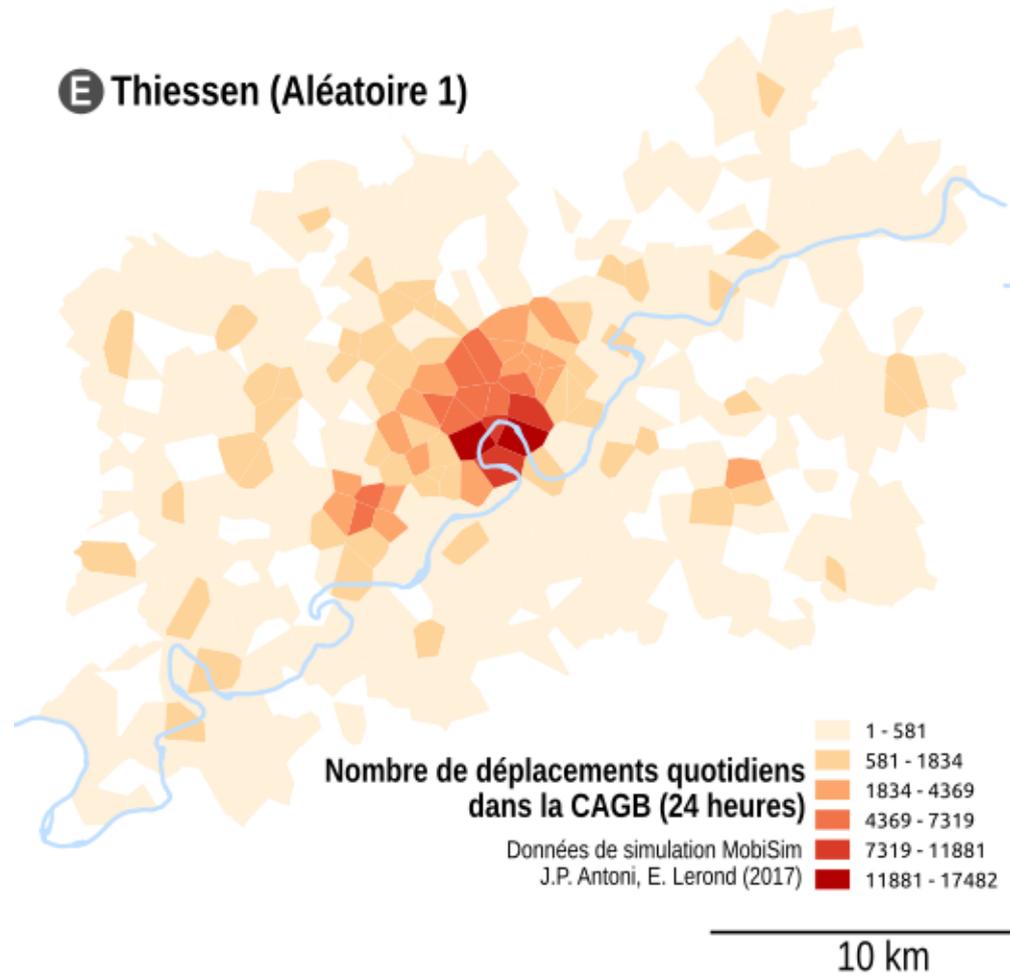
D Thiessen (bâti)



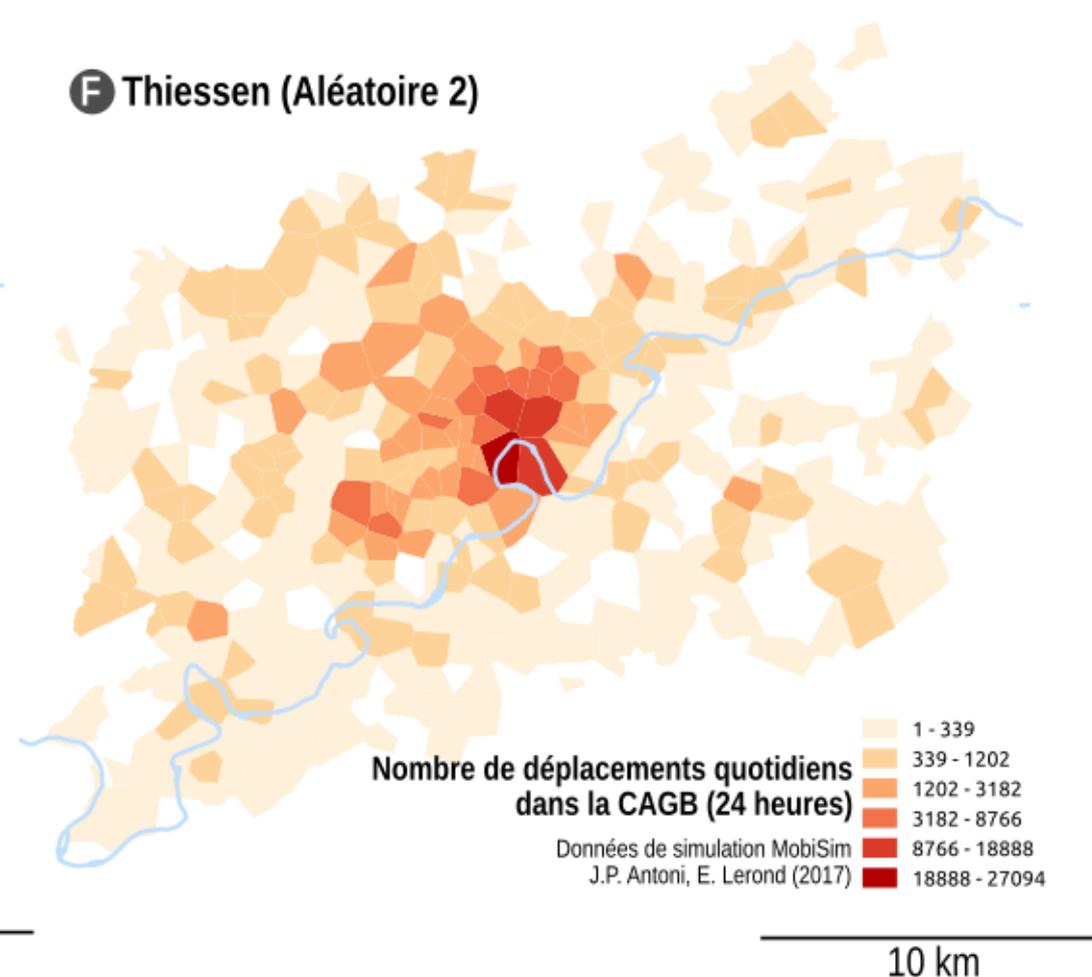
3. RÉSULTATS

3.1. Cartographie

E Thiessen (Aléatoire 1)



F Thiessen (Aléatoire 2)

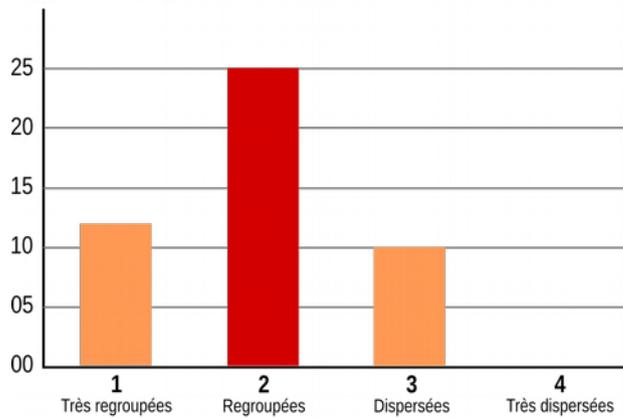


3.2. Enquête

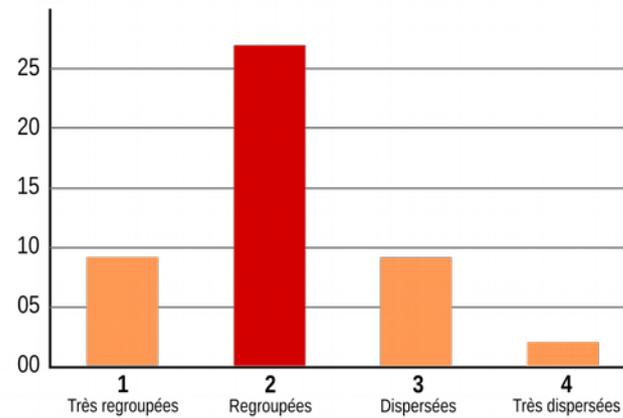
Question 1

Les données représentées sur chaque carte sont
1. Très regroupées 2. Regroupées 3. Dispersées 4. Très dispersées

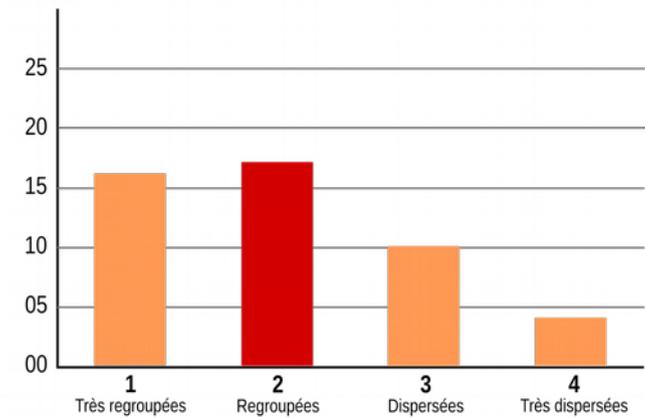
A Carroyage



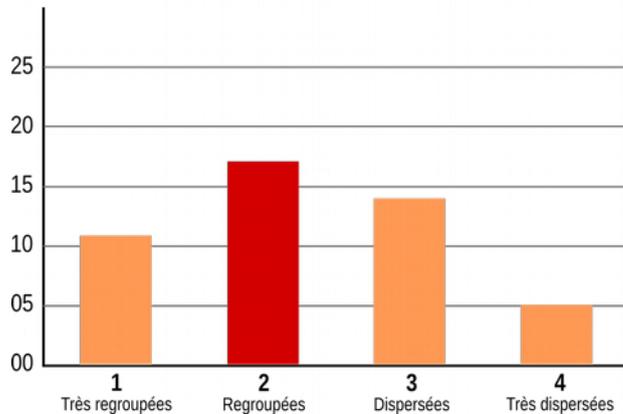
B Nid d'abeilles



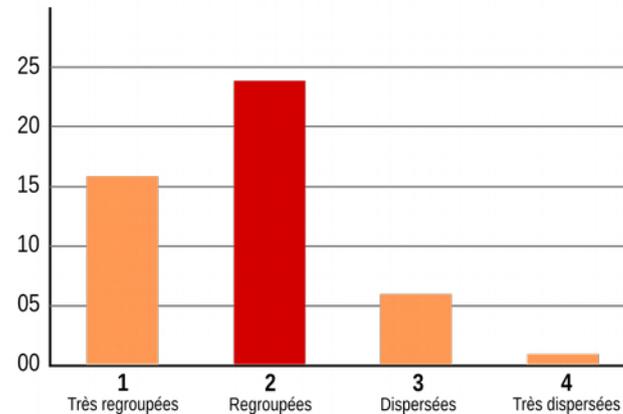
C Cible



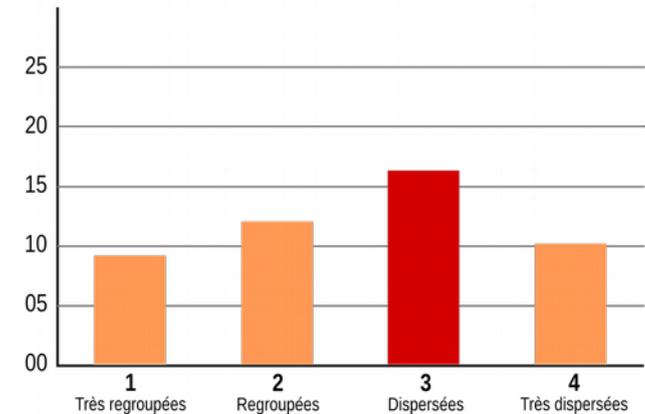
D Thiessen (bâti)



E Thiessen (Aléatoire 1)



F Thiessen (Aléatoire 2)

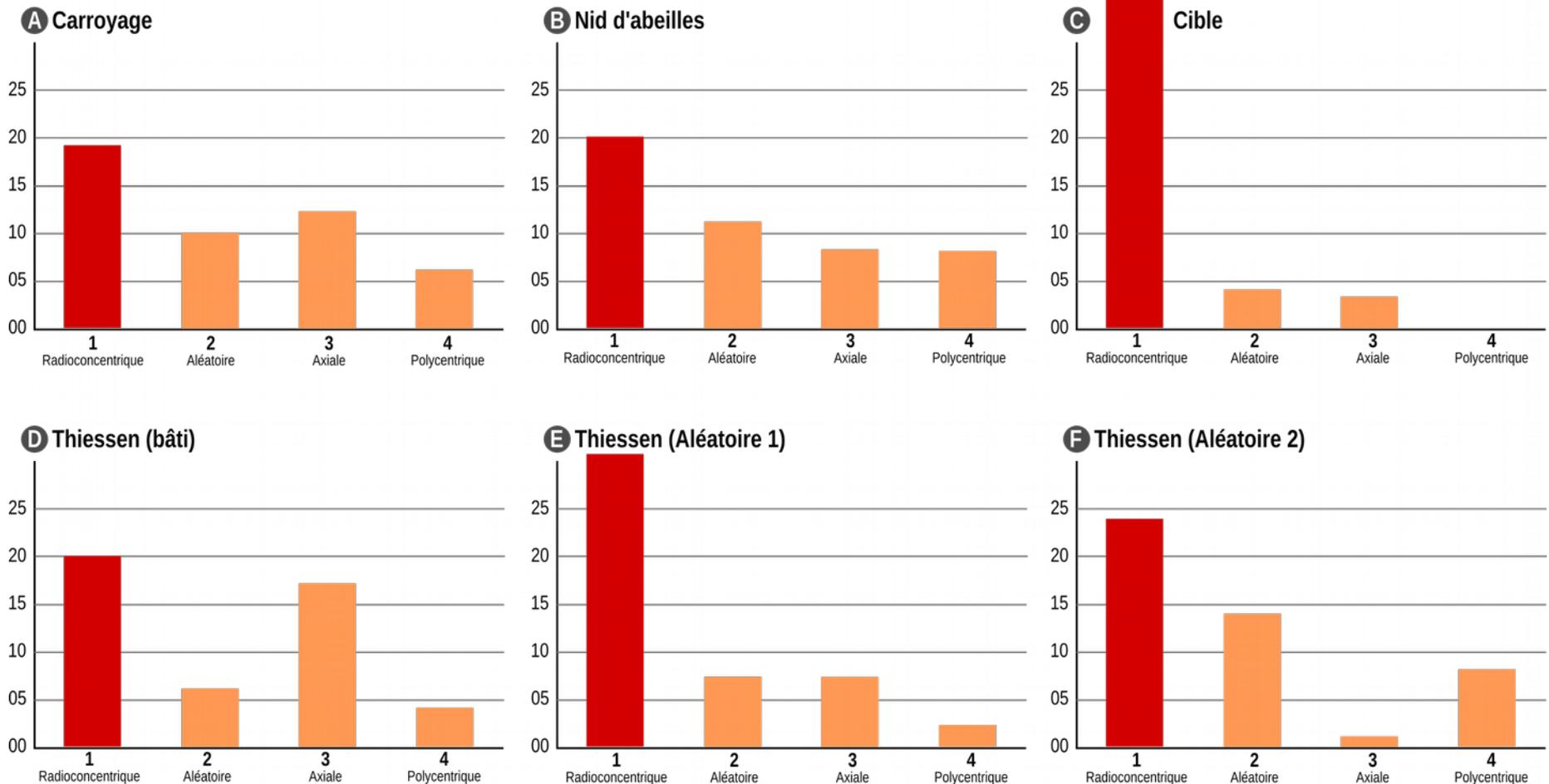


3.2. Enquête

Question 2

Les données représentées sur chaque carte sont réparties selon une logique

1. Radioconcentrique 2. Aléatoire 3. Axiale 4. Polycentrique

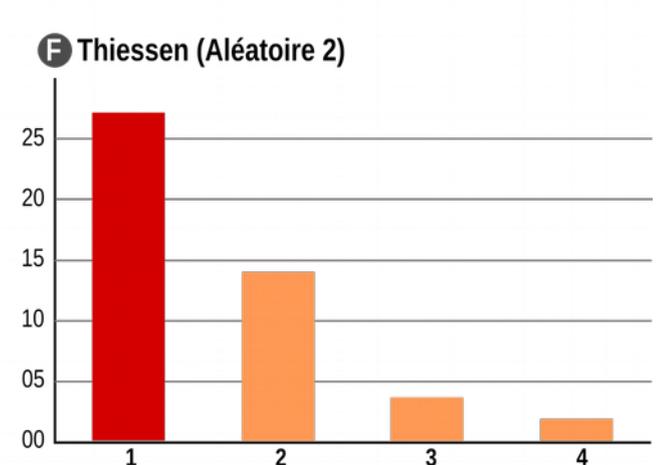
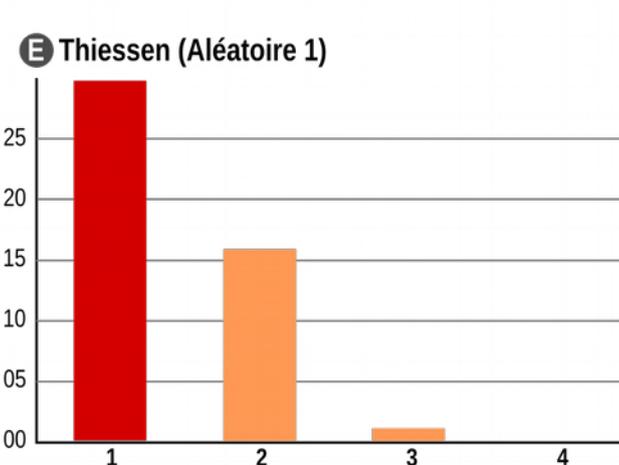
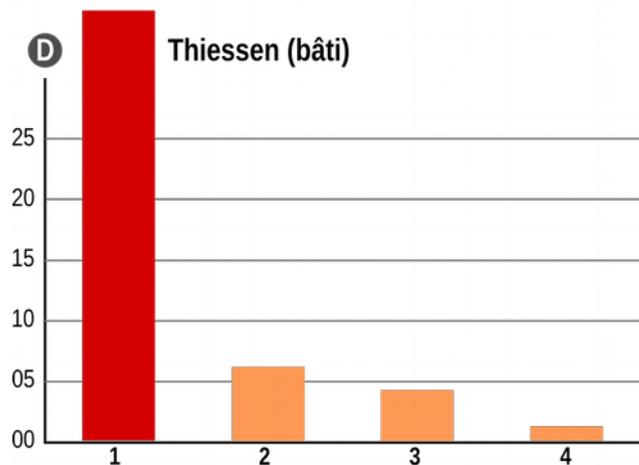
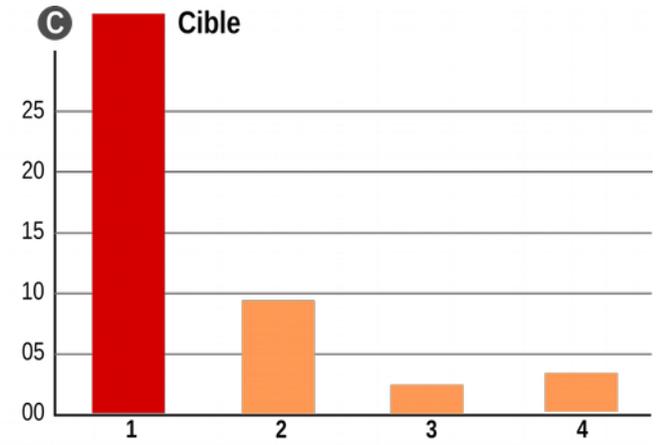
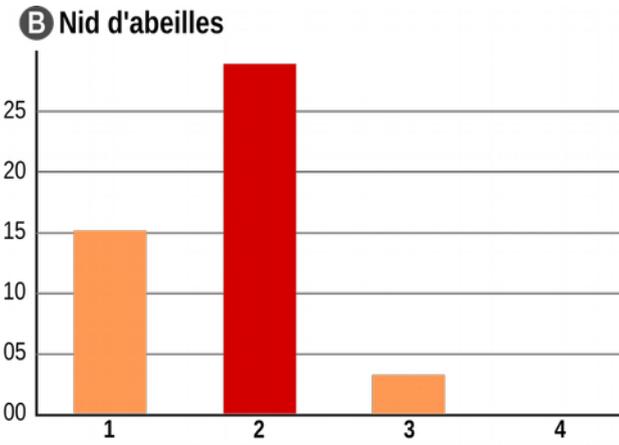
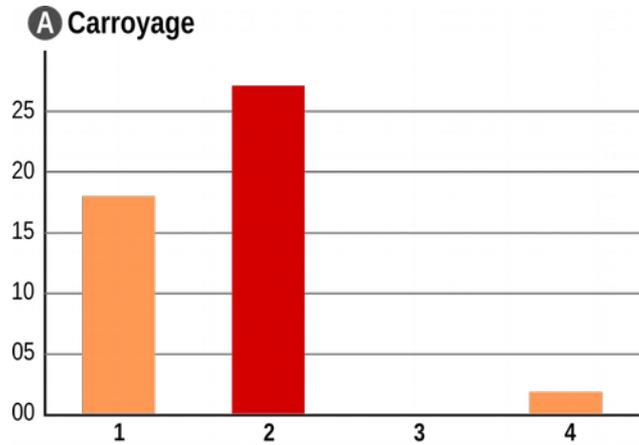


3.2. Enquête

Question 3

Sur chaque carte, combien de centres pouvez-vous identifier ?

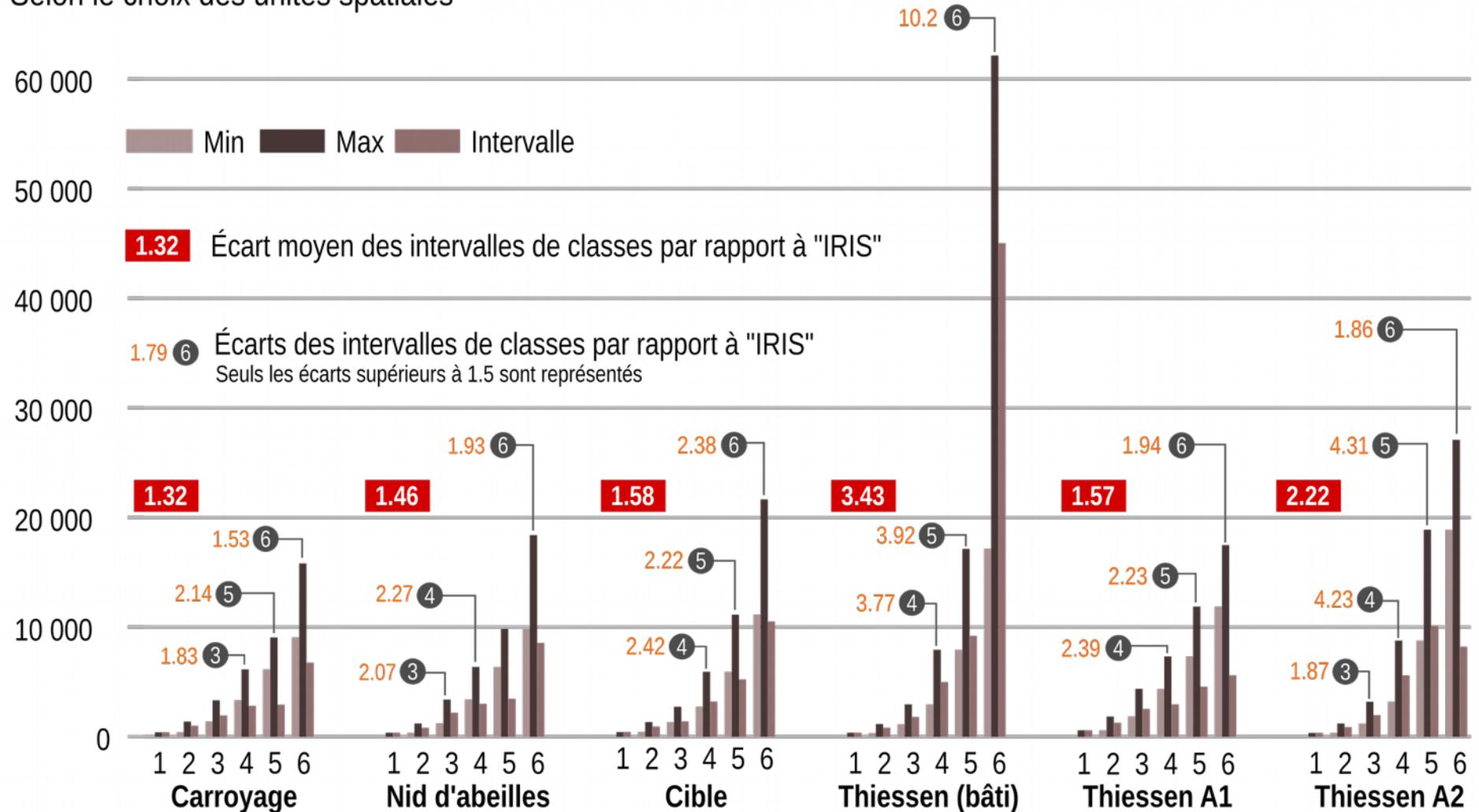
1. 1 2. 2 3. 3 4. 4



3.3. Indicateurs

Nombre de déplacements par classes

Selon le choix des unités spatiales



3.3. Indicateurs

	GVF	TAI	OAI	
IRIS	0.965	0.790	0.769	(Le plus mauvais)
A. Carroyage	0.979	0.805	0.805	
B. Nid d'abeilles	0.974	0.809	0.809	
C. Cible	0.975	0.762	0.762	
D. Thiessen (bâti)	0.991	0.793	0.877	(Le meilleur)
E. Thiessen (A1)	0.969	0.765	0.757	
F. Thiessen (A2)	0.972	0.694	0.726	

3.3. Indicateurs

	GVF	TAI	OAI	
IRIS	0.965	0.790	0.769	(Le plus mauvais)
A. Carroyage	0.979	0.805	0.805	
B. Nid d'abeilles	0.974	0.809	0.809	
C. Cible	0.975	0.762	0.762	
D. Thiessen (bâti)	0.991	0.793	0.877	(Le meilleur)
E. Thiessen (A1)	0.969	0.765	0.757	
F. Thiessen (A2)	0.972	0.694	0.726	

Interprétation

La logique géographique (fondée sur le bâti) se révèle statistiquement plus efficace que la logique géométrique, systématique ou aléatoire

1. INTRODUCTION

1.1. Contexte | 1.2. Problématique | 1.3. Hypothèses et objectif

2. METHODOLOGIE

2.1. Construction des unités d'agrégation | 2.2. Évaluation

3. RÉSULTATS

3.1. Cartographie | 3.2. Enquête | 3.3. Indicateurs

4. CONCLUSION ET PERSPECTIVES

Des hypothèses en voie de confirmation

La structure des unités spatiales d'agrégation :

A. influence les résultats quantifiés et cartographiés

B. influence l'interprétation des résultats et peut conduire à des conclusions contradictoires

Des hypothèses en voie de confirmation

La structure des unités spatiales d'agrégation :

A. influence les résultats quantifiés et cartographiés

B. influence l'interprétation des résultats et peut conduire à des conclusions contradictoires

Un test trop partiel

Un nombre quasi infini de tailles et de formes d'unités spatiales

Une pré-enquête fondée sur un échantillon trop faible (n=47)

Un profil d'enquêtés qui mériterait d'être approfondi et intégré à l'analyse

Des hypothèses en voie de confirmation

La structure des unités spatiales d'agrégation :

A. influence les résultats quantifiés et cartographiés

B. influence l'interprétation des résultats et peut conduire à des conclusions contradictoires

Un test trop partiel

Un nombre quasi infini de tailles et de formes d'unités spatiales

Une pré-enquête fondée sur un échantillon trop faible (n=47)

Un profil d'enquêtés qui mériterait d'être approfondi et intégré à l'analyse

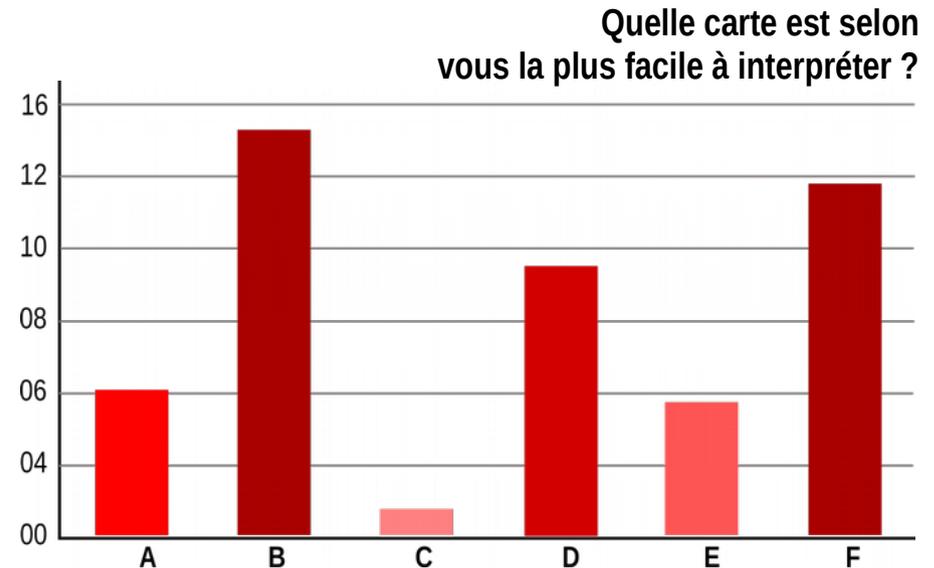
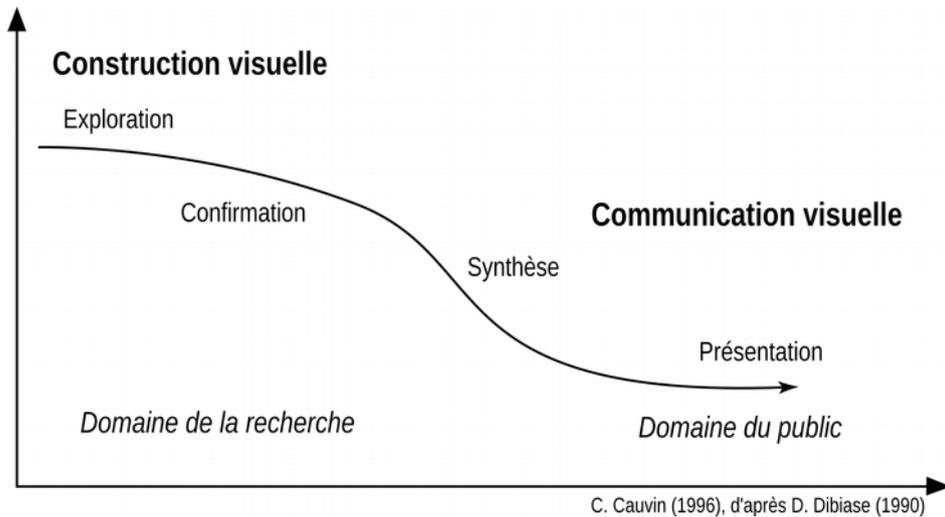
Un optimum difficile à atteindre

La meilleure solution consiste-t-elle à atteindre un optimum ?

Quel optimum choisir entre l'optimum statistique et l'optimum cognitif ?

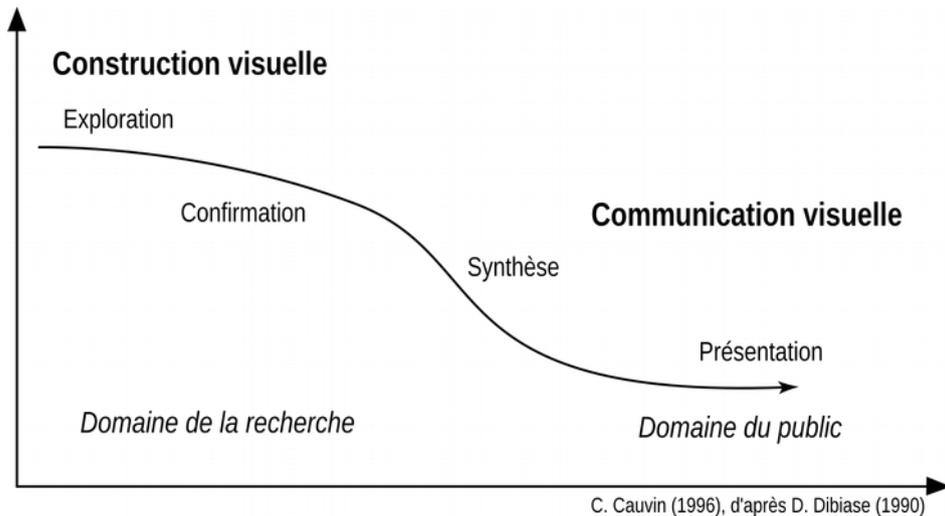
Perspectives immédiates

1. Étudier la convergence entre optimums statistique et cognitif (cf. Dibiase, 1990)
2. Multiplier les exemples en associant le découpage à un sens géographique
3. Se méfier des « effets de structure »

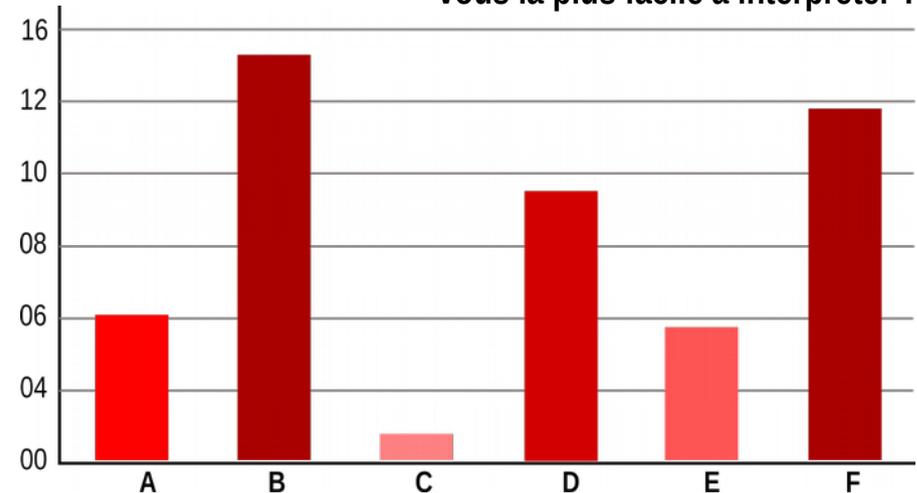


Perspectives immédiates

1. Étudier la convergence entre optimums statistique et cognitif (cf. Dibiase, 1990)
2. Multiplier les exemples en associant le découpage à un sens géographique
3. Se méfier des « effets de structure »



Quelle carte est selon vous la plus facile à interpréter ?



Perspectives générales

1. Intégrer ces résultats dans la visualisation des données de simulation (MobiSim)
2. Intégrer la question des formes d'agrégation aux débats sur l'exploitation des Big Data

MERCI POUR VOTRE ATTENTION