

# **Outils de l'analyse exploratoire des données géo-référencées**

# Plan

- ▶ **Introduction**
- ▶ **Outils cartographiques**
  - **Décrire et visualiser les distributions spatiales**
  - **Visualiser les observations extrêmes : outliers**
- ▶ **Statistiques d'autocorrélation spatiale**
  - **Autocorrélation spatiale globale**
    - ▶ **Principes généraux**
    - ▶ **Statistique Join Count : variables qualitatives**
    - ▶ **Statistique I de Moran : variables quantitatives**
    - ▶ **Statistique c de Geary : variables quantitatives**
    - ▶ **Diagramme de Moran**
  - **Autocorrélation spatiale locale**
    - ▶ **Principes : Indicateurs locaux d'association spatiale**
    - ▶ **Statistiques locales de Moran (Anselin, 1995)**
    - ▶ **Statistiques de Getis et Ord (1992, 1995)**
- ▶ **Le cas particulier des taux ou proportions**
- ▶ **Quelques problèmes méthodologiques**

# Introduction

- ▶ **Qu'est-ce que Analyse exploratoire de données géo-référencées ?**
- ▶ **Analyse exploratoire des données +**
  - **Décrire et visualiser les distributions spatiales**
    - ▶ Enrichir l'information cartographique à l'aide de graphiques statistiques
  - **Identifier les observations atypiques : points extrêmes ou « outliers » spatiaux**
    - ▶ Box-plot, Box-map
  - **Mettre en évidence des schémas d'association spatiale**
    - ▶ concentrations spatiales
    - ▶ visualiser les mesures et la significativité de l'autocorrélation spatiale globale et locale
  - **Détecter les schémas d'hétérogénéité spatiale et suggérer des régimes spatiaux**
    - ▶ non-stationnarité spatiale – hétérogénéité spatiale
    - ▶ autocorrélation spatiale locale

# **Outils cartographiques (1)**

## **Décrire et visualiser les distributions spatiales**

# Carte plane ou choropleth

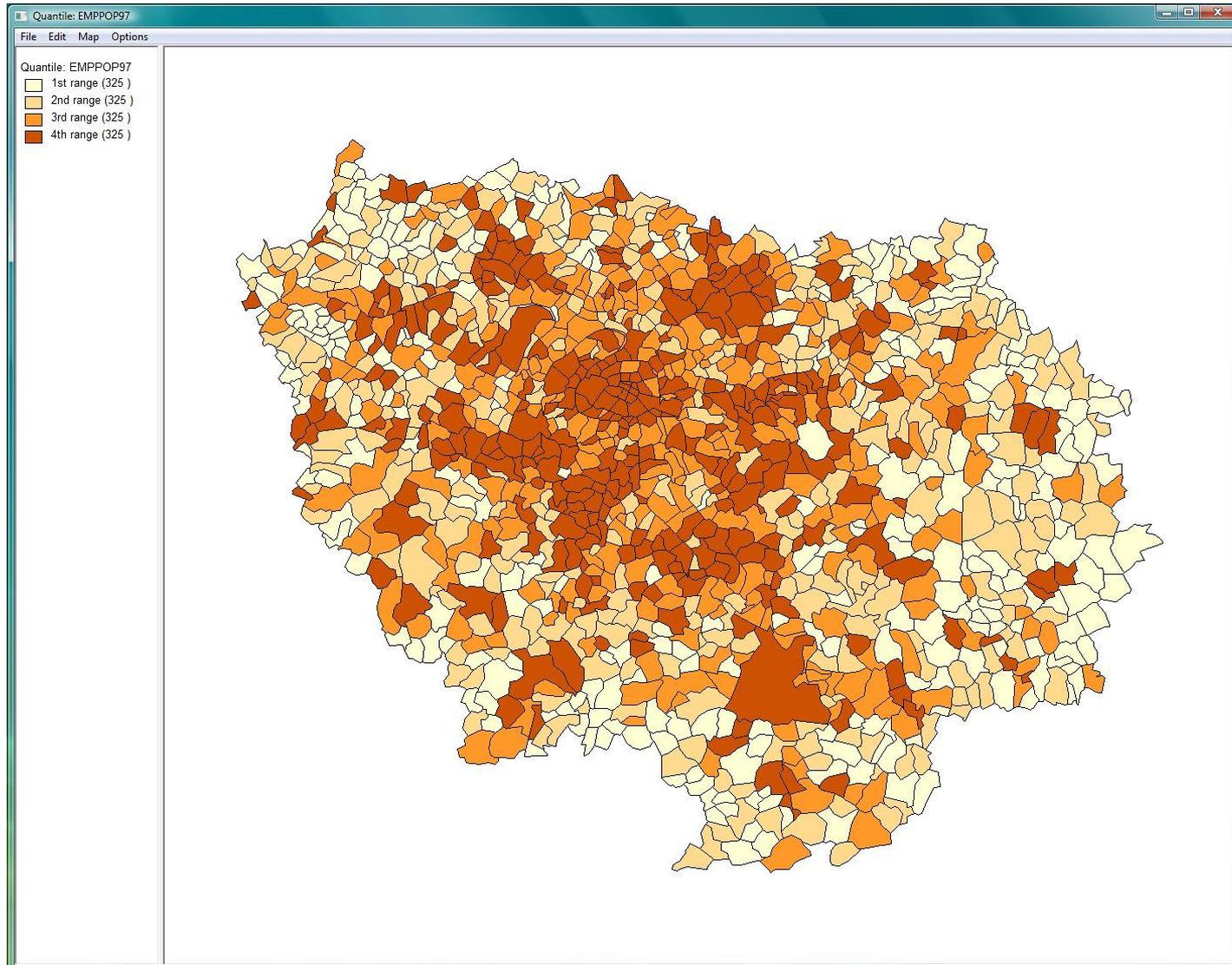
- ▶ Histogramme standard lié à une représentation cartographique
- ▶ Histogramme cartographique
  - valeurs pour chaque unité spatiale discrète
  - Cartes dans GeoDa :
    - ▶ quantiles : choix du nombre de classes (par défaut 4)
      - médiane = 2 classes : 50% des obs. dans chaque classe
      - quartiles = 4 classes : 25% des obs. dans chaque classe
      - quintiles = 5 classes : 20% des obs. dans chaque classe
    - ⇒ chaque classes a le même nombre d'obs.
    - ▶ pourcentiles : 6 classes : <1%, 1-10%, 10-50%, 50-90%, 90-99%, >99%
    - ▶ Intervalles égaux : choix du nombre d'intervalles (par défaut 4)
    - ▶ Variable qualitative ou quantitative discrète (nombre de valeurs distinctes max. 10)
  - Carte de dispersion :
    - ▶ la classe k regroupe les observations se situant à  $\bar{x} \pm k\hat{\sigma}$
    - ▶ le nombre d'observations dans la classe dépend de la distribution des données
- ▶ Problèmes pratiques
  - choix du type de graphique et/ou du nombre d'intervalles
    - ▶ quantiles, intervalles égaux, dispersion
  - illusions d'optique : choix des couleurs
    - important pour la perception des schémas
    - les grandes surfaces semblent plus importantes

# Choix des couleurs à <http://www.colorbrewer.org>

The image shows the ColorBrewer.org website interface. At the top, it displays the logo 'dg QG' and the title '5-class diverging RYB'. The main map shows a 5-class diverging RYB color palette applied to a map of the United States, with colors ranging from red to blue. The interface is divided into several sections:

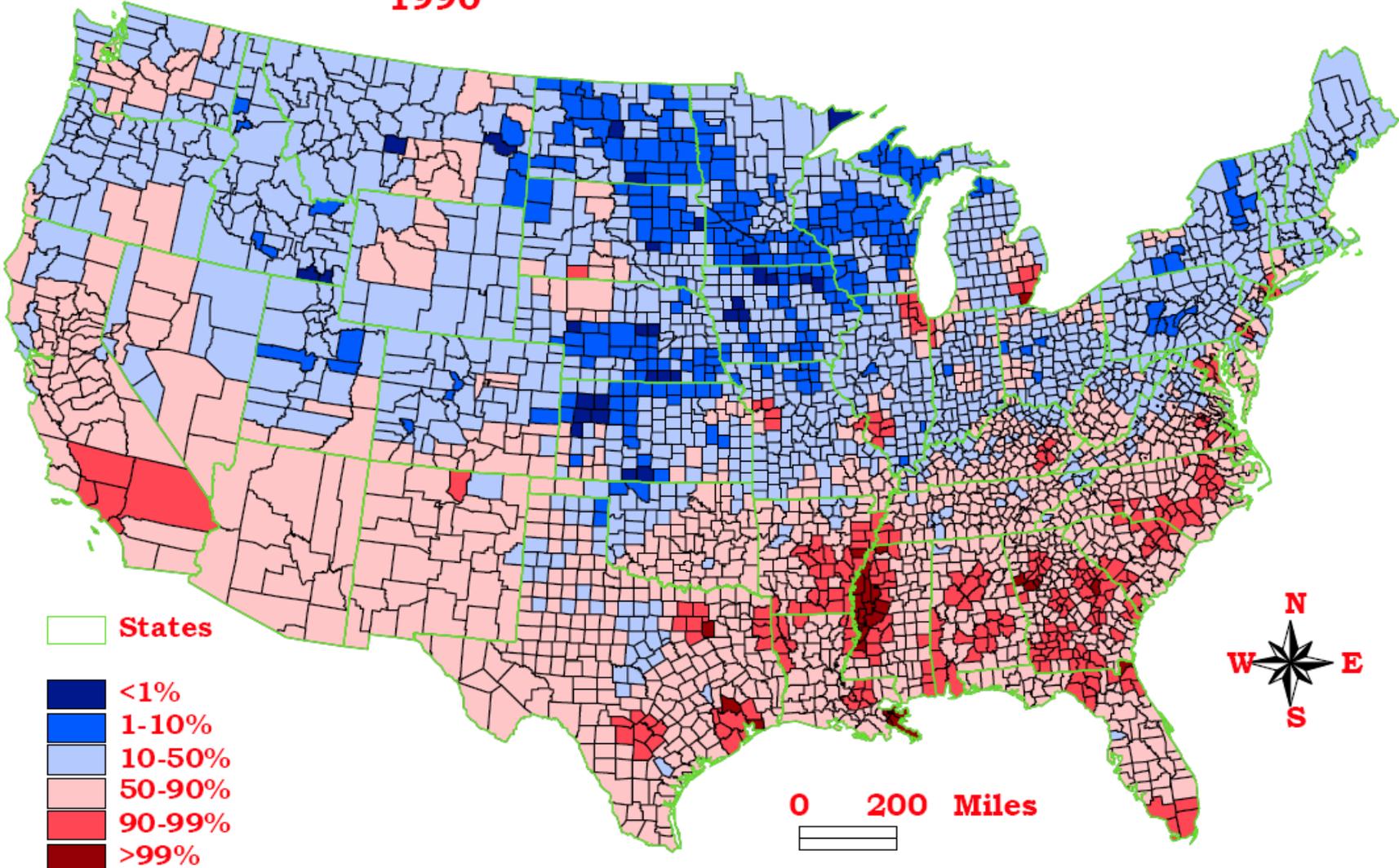
- Step 1:** 'number of classes' set to 5, with a 'learn more' link.
- Step 2:** 'legend type' set to 'diverging', with options for 'sequential' and 'qualitative'.
- Step 3:** 'mini legends' showing various color schemes.
- Color Selection:** 'cmyk', 'rgb', 'hex', 'Lab', and 'AV3' tabs, with a color selection tool.
- Map Settings:** 'map zoom', 'map borders' (off), 'city symbols' (off), 'road network' (on), and 'background color' (white/black).
- Export:** 'ArcView 3.x' and 'export' buttons.

# Choropleth du ratio emploi sur population en Ile de France en 1997 : quartiles



# Choropleth des homicides aux Etats-Unis en 1990

**Percentile Map  
Homicide Rate (SRS)  
1990**

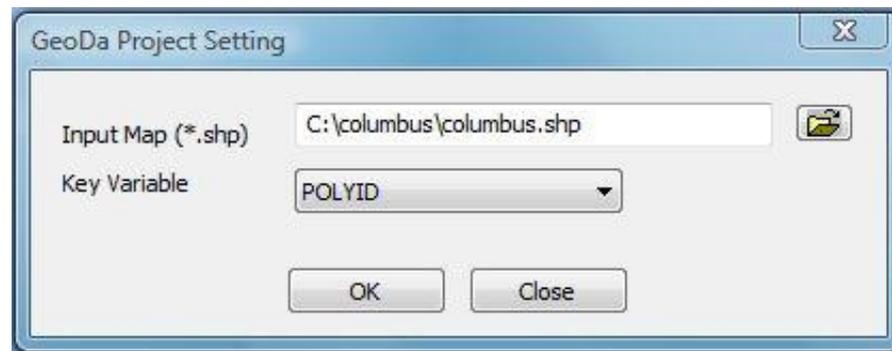


# Utiliser GeoDa 1.0 (Anselin, 2003)

- ▶ Lancer OpenGeoDa 1.0.0.7 (10 sept 2008)



1. Ouverture d'un projet : menu FILE – Open Project : nom du fichier.shp  
fichier contenant les variables pour les unités spatiales géo-référencées  
valeurs + objets spatiaux (polygones ou points) extension .shp (SIG)
  2. Choix de la variable codant les unités spatiales : Key Variable
  3. OK
- ▶ exemple : columbus.shp / Key Variable POLYID/ OK



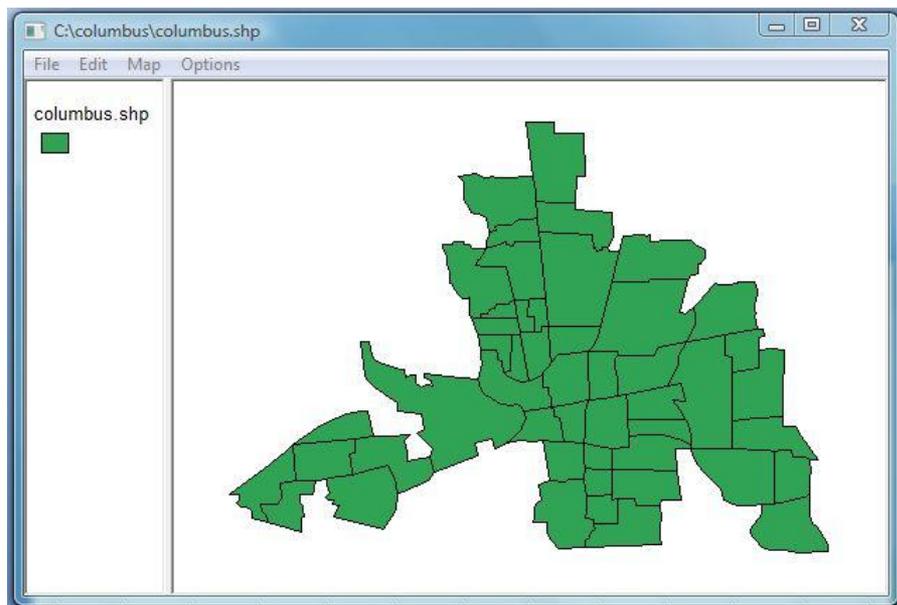
# Représentations graphiques dans GeoDa



Feuille  
de calcul

Nouvelle  
carte

Dupliquer  
la carte

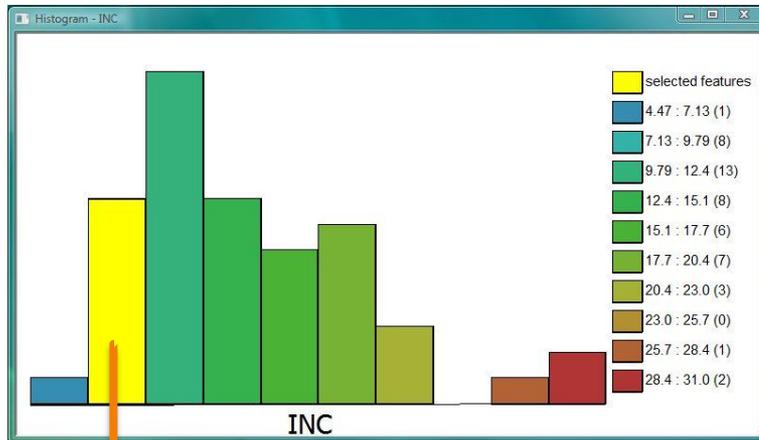


	AREA	PERIMETER	COLUMBUS_1	COLUMBUS_1	POLYID	NEIG	HOVAL	INC	CRIME	OPEN	PLUMB	DISCBD
1	0.302441	2.440630	2	5	1	5	80.467000	19.531000	15.726000	2.850750	0.217155	5.030000
2	0.259329	2.236940	3	1	2	1	44.567000	21.232000	18.801800	5.296720	0.320581	4.270000
3	0.192468	2.187550	4	6	3	6	26.350000	15.956000	30.626800	4.534650	0.374404	3.890000
4	0.083841	1.427640	5	2	4	2	33.200000	4.477000	32.387800	0.394427	1.186940	3.700000
5	0.488888	2.997130	6	7	5	7	23.225000	11.252000	50.731500	0.405664	0.624596	2.830000
6	0.283079	2.335630	7	8	6	8	28.750000	16.029000	26.066700	0.563075	0.254130	3.780000
7	0.257084	2.554580	8	4	7	4	75.000000	8.438000	0.178269	0.000000	2.402400	2.740000
8	0.204954	2.139520	9	3	8	3	37.125000	11.337000	38.425900	3.483480	2.739730	2.890000
9	0.500755	3.169710	10	18	9	18	52.600000	17.586000	30.515900	0.527488	0.890736	3.170000
10	0.246689	2.087240	11	10	10	10	96.400000	13.598000	34.008000	1.548350	0.557724	4.330000
11	0.041012	0.919488	12	38	11	38	19.700000	7.467000	62.275400	0.000000	1.479920	1.900000
12	0.035769	0.902125	13	37	12	37	19.900000	10.048000	56.705700	3.157890	2.635050	1.910000
13	0.034377	0.936590	14	39	13	39	41.700000	9.549000	46.716100	0.000000	6.328420	2.090000
14	0.060884	1.128420	15	40	14	40	42.900000	9.963000	57.066100	0.477104	5.110960	1.830000
15	0.106653	1.437610	16	9	15	9	18.000000	9.873000	48.585500	0.174325	1.311470	1.700000
16	0.093154	1.340060	17	36	16	36	18.800000	7.625000	54.838700	0.533737	4.687500	1.100000
17	0.102087	1.382360	18	11	17	11	41.750000	9.798000	36.868800	0.448232	1.619750	4.470000
18	0.055494	1.183350	19	42	18	42	60.000000	13.185000	43.962500	24.998100	13.849300	1.580000
19	0.061342	1.249250	20	41	19	41	30.600000	11.618000	54.522000	0.111111	2.622960	1.530000
20	0.444629	3.174600	21	17	20	17	81.267000	31.707000	0.223797	5.318610	0.167224	3.570000
21	0.699298	5.077490	22	43	21	43	19.975000	10.655000	40.074100	1.643760	1.599580	1.410000
22	0.192891	1.992720	23	19	22	19	30.450000	11.709000	33.705000	4.539750	1.785710	2.450000
23	0.247120	2.147530	24	12	23	12	47.733000	21.155000	20.048500	0.532632	0.216763	4.780000
24	0.192226	2.240390	25	35	24	35	53.200000	14.236000	38.297900	0.628220	18.811100	0.420000
25	0.171680	1.666490	26	32	25	32	17.900000	8.461000	61.299200	0.000000	6.528850	0.830000
26	0.107298	1.406820	27	20	26	20	20.300000	8.085000	40.969700	1.238290	2.534280	1.500000
27	0.137802	1.780750	28	21	27	21	34.100000	10.822000	52.794400	19.368100	1.483520	2.260000
28	0.174773	1.637150	29	31	28	31	22.850000	7.856000	56.919800	0.509305	3.001070	1.410000
29	0.085972	1.312160	30	33	29	33	32.500000	8.681000	60.750400	0.000000	2.645050	0.810000
30	0.104355	1.524930	31	34	30	34	22.500000	13.906000	68.892000	1.638780	15.600600	0.370000
31	0.117409	1.716050	32	45	31	45	31.800000	16.940000	17.677200	3.936440	0.853890	3.780000
32	0.185580	2.108950	33	13	32	13	40.300000	18.942000	19.145600	2.221020	0.255102	4.760000
33	0.087472	1.507970	34	22	33	22	23.600000	9.918000	41.968200	0.000000	1.023890	2.280000
34	0.226594	2.519130	35	44	34	44	28.450000	14.948000	23.974000	3.029900	0.386803	3.060000
35	0.175453	1.974940	36	23	35	23	27.000000	12.814000	39.175100	4.220400	0.633675	2.370000
36	0.178130	1.790060	37	46	36	46	36.300000	18.739000	14.305600	6.773330	0.332349	4.230000
37	0.121154	1.402250	38	30	37	30	43.300000	17.017000	42.445100	4.839270	1.203330	1.080000
38	0.053881	0.934509	39	24	38	24	22.700000	11.107000	53.710900	0.000000	0.800000	1.580000
39	0.174823	2.335400	40	47	39	47	39.600000	18.477000	19.109000	0.000000	0.314663	5.530000
40	0.302908	2.285490	41	16	40	16	61.950000	29.833000	16.241300	6.451310	0.132743	4.400000
41	0.137024	1.525100	42	14	41	14	42.100000	22.207000	18.905100	0.293317	0.247036	5.330000
42	0.266541	2.176540	43	49	42	49	44.333000	25.873000	16.491900	1.792990	0.134439	3.870000
43	0.060241	0.967793	44	29	43	29	25.700000	13.380000	36.663600	0.000000	0.589226	1.950000
44	0.173337	1.868040	45	25	44	25	33.500000	16.961000	25.962300	1.463990	0.329761	2.670000
45	0.256431	2.193040	46	28	45	28	27.733000	14.135000	29.028900	1.006120	2.391200	2.130000
46	0.124728	1.841030	47	48	46	48	76.100000	18.324000	16.530500	9.683950	0.424628	5.270000
47	0.245249	2.079990	48	15	47	15	42.500000	18.950000	27.822900	0.000000	0.268817	5.570000
48	0.069762	1.102030	49	27	48	27	26.800000	11.813000	26.645300	4.884390	1.034810	2.330000
49	0.205964	2.199170	50	26	49	26	35.800000	18.796000	22.541500	0.259826	0.901442	3.030000

# Histogramme standard



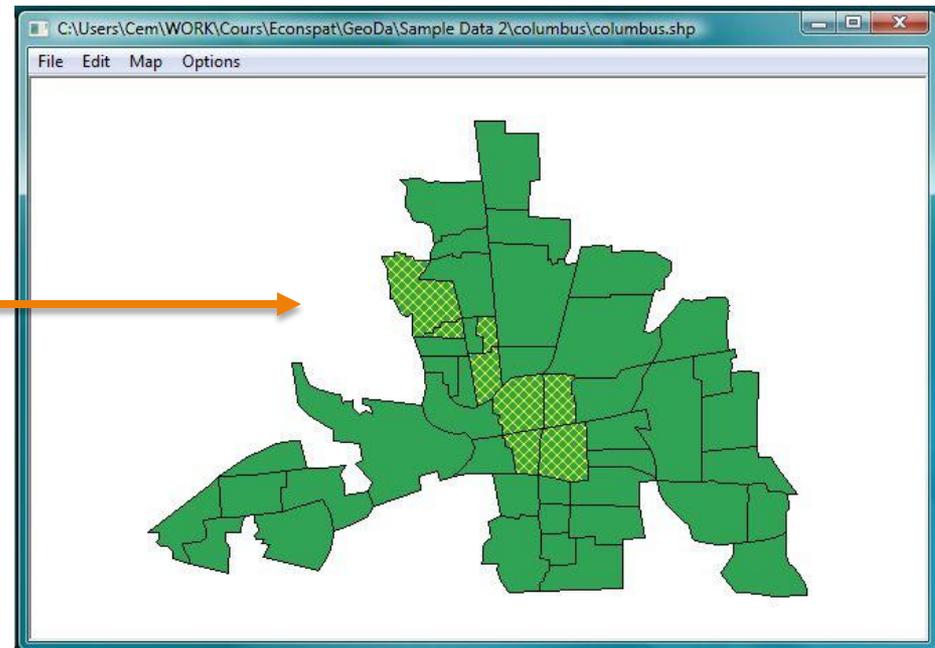
Histogramme standard



Sélection liée

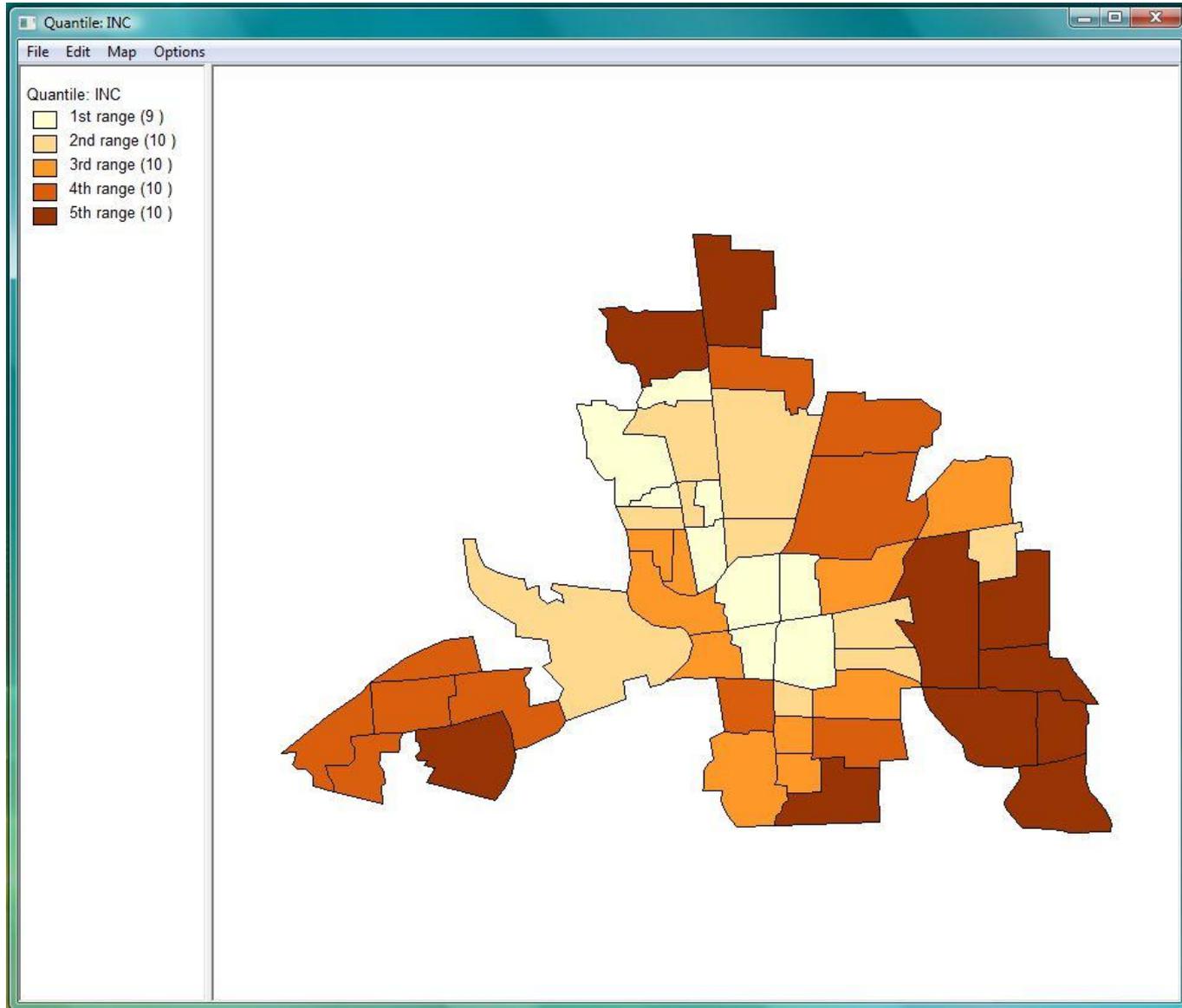
Localisation des quartiers pour lesquels le revenu annuel moyen des ménages appartient à la 2<sup>ème</sup> classe [7.13; 9.79[

INC : Revenu annuel moyen des ménages en milliers de dollars en 1980 dans les quartiers de la ville de Columbus Ohio (USA) (Anselin, 1988)



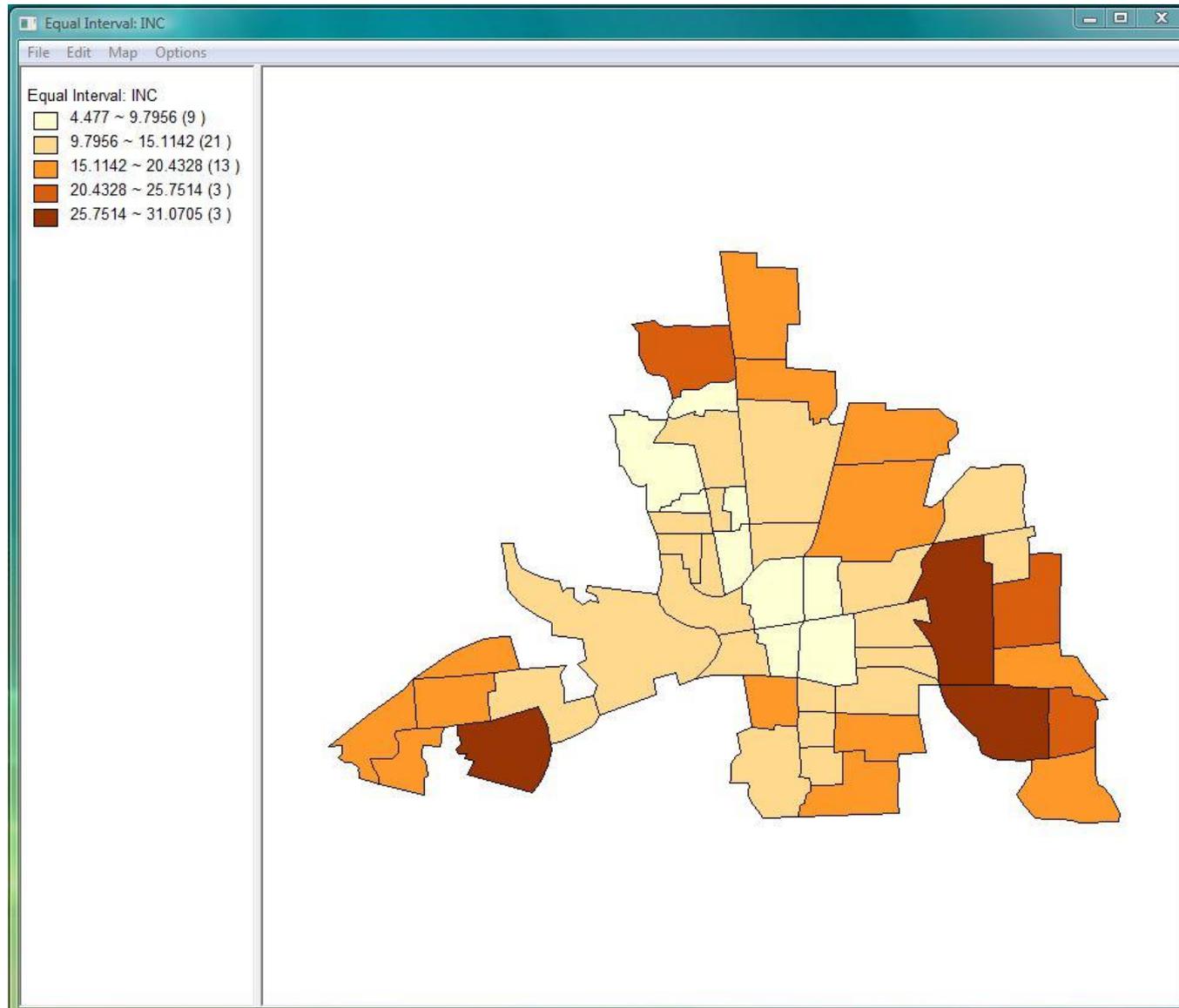
# Carte des quintiles (GeoDa)

## revenu des ménages à Columbus (Ohio, USA)



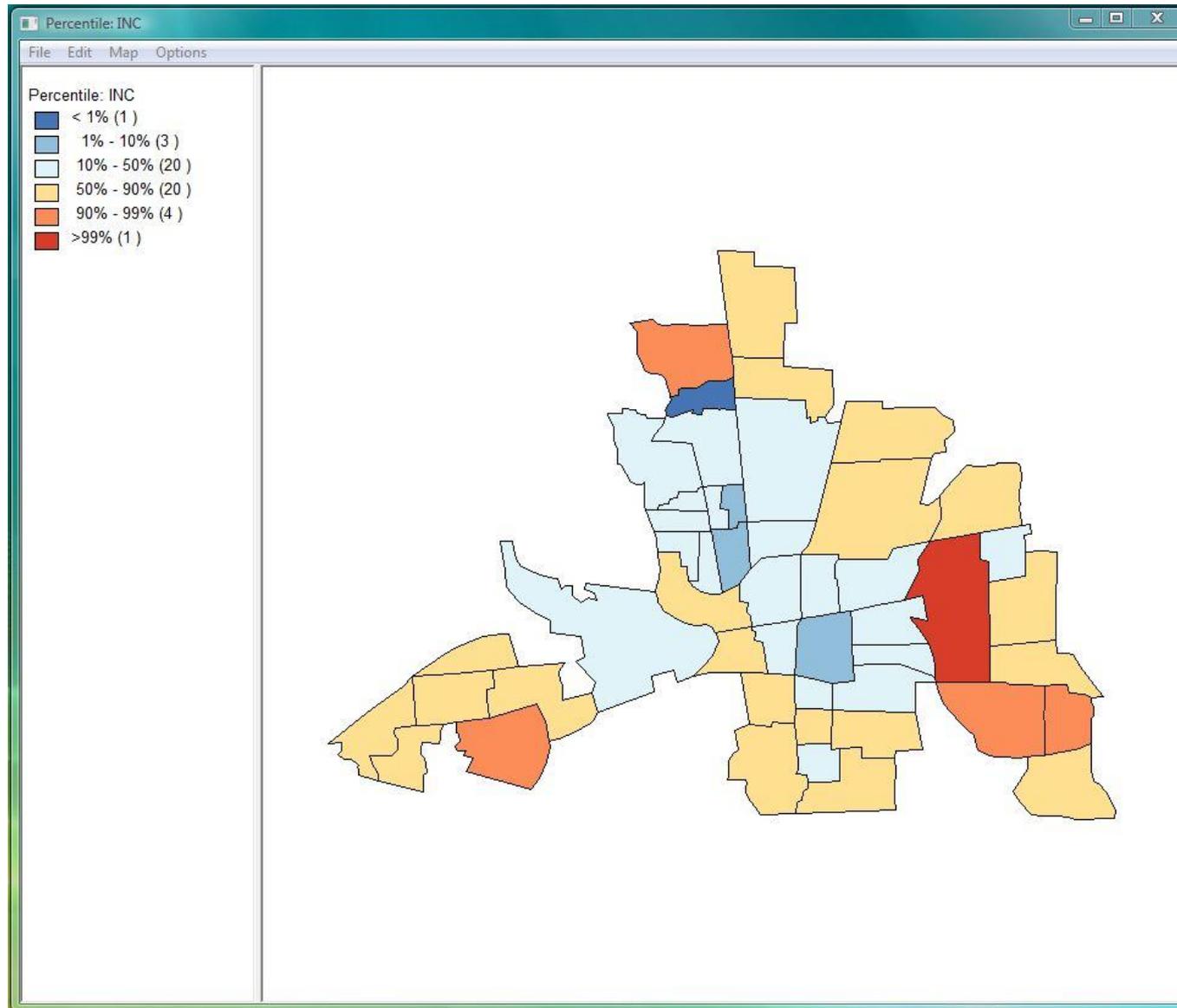
# Carte à intervalles égaux

## Revenu des ménages à Columbus (Ohio, USA)

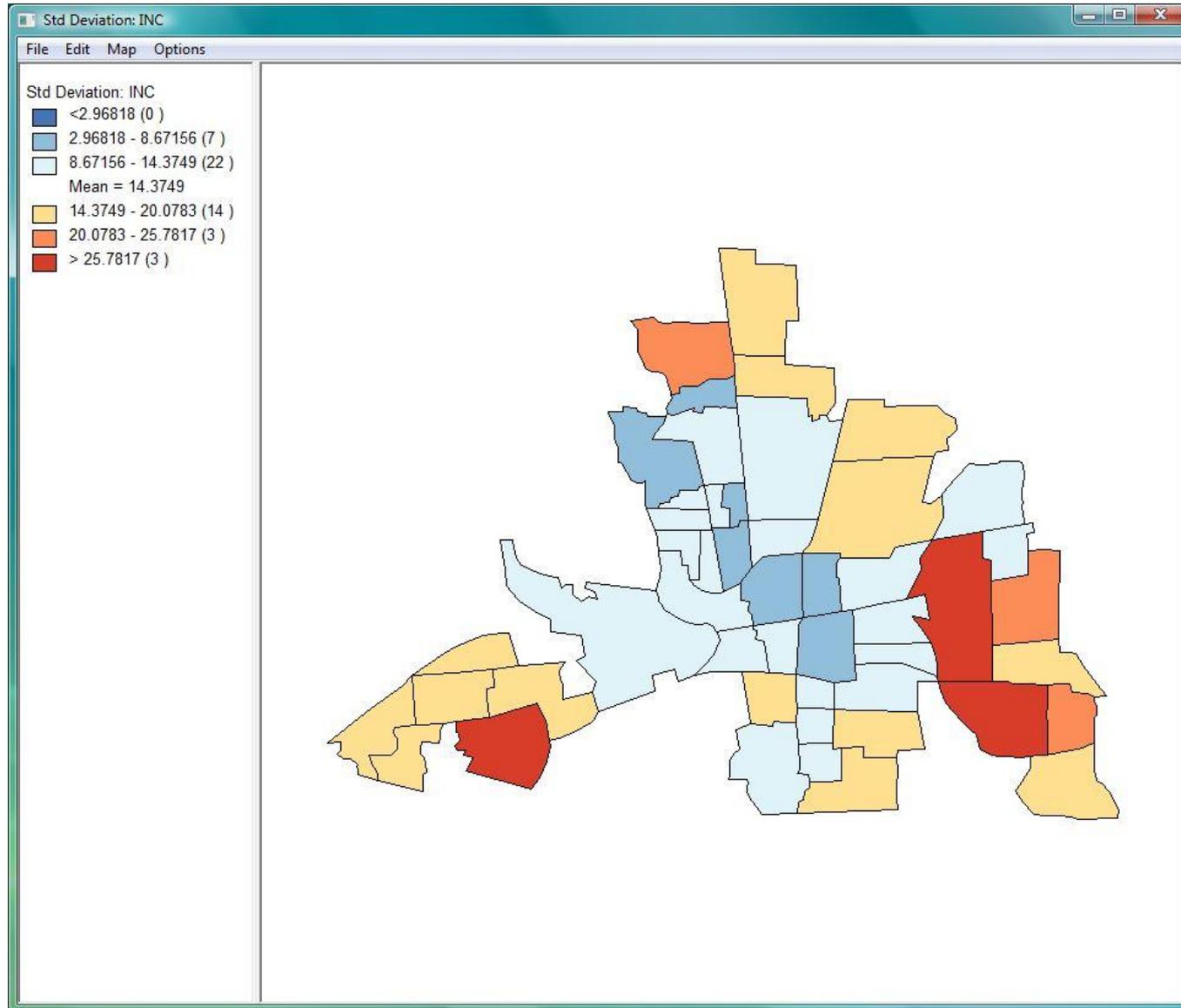


# Carte des pourcentiles

## Revenu des ménages à Columbus (Ohio, USA)



# Carte de dispersion (GeoDa) revenu des ménages à Columbus (Ohio, USA)



# Cartogrammes

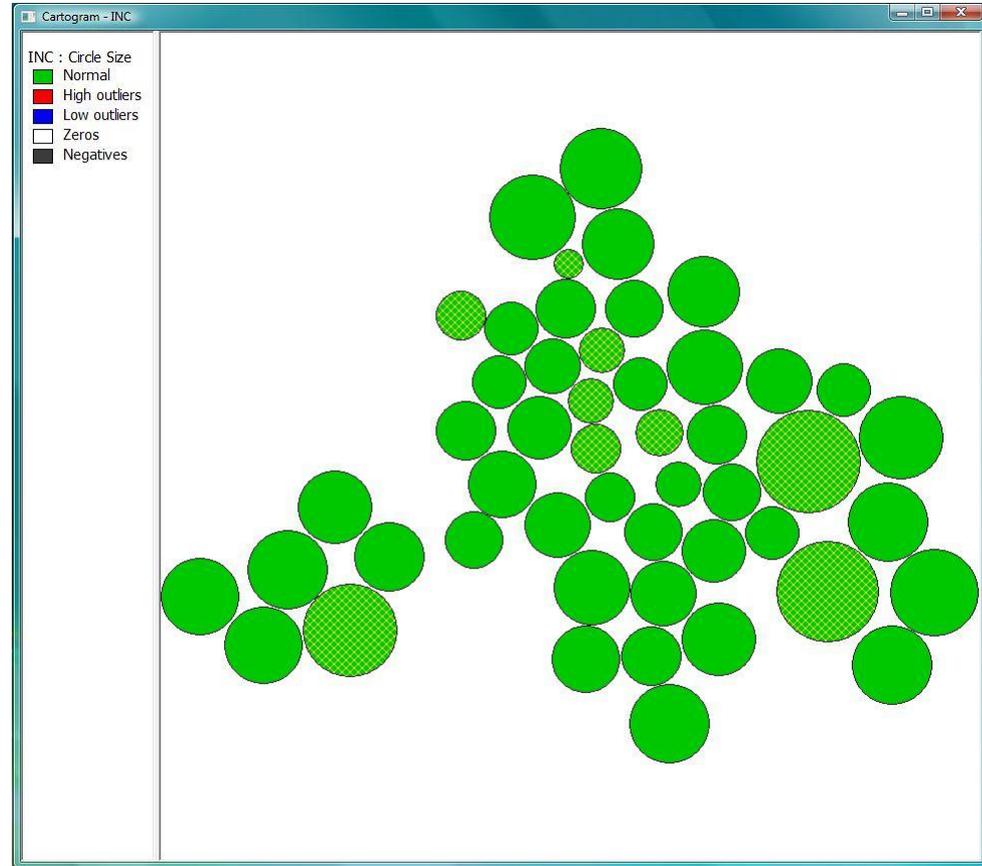
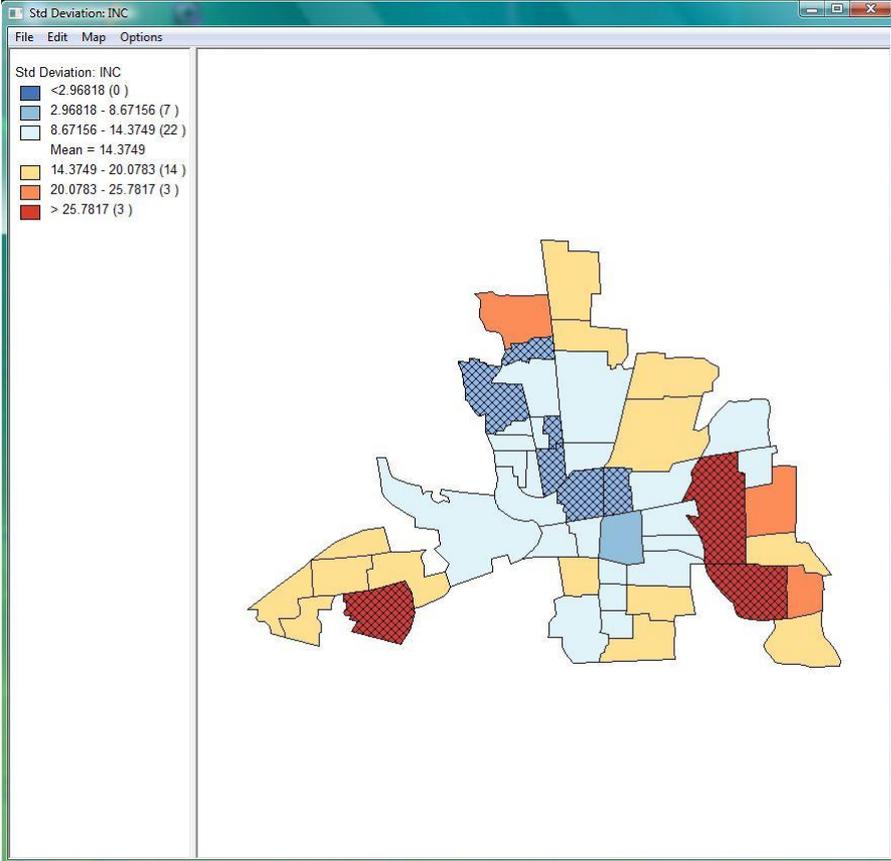
- ▶ Les aires plus grandes apparaissent plus importantes
- ▶ On peut superposer des symboles sur les aires
  
- ▶ Cartogrammes
  - On représente la taille plutôt que l'aire
  - La topologie est respectée
  - Cartogrammes contiguës : la contiguïté est respectée
  - Cartogrammes circulaires : les zones sont transformées en cercle



Cartogramme

# Cartogramme circulaire (GeoDa)

## Revenu des ménages à Columbus (Ohio, USA)



# **Outils cartographiques (2)**

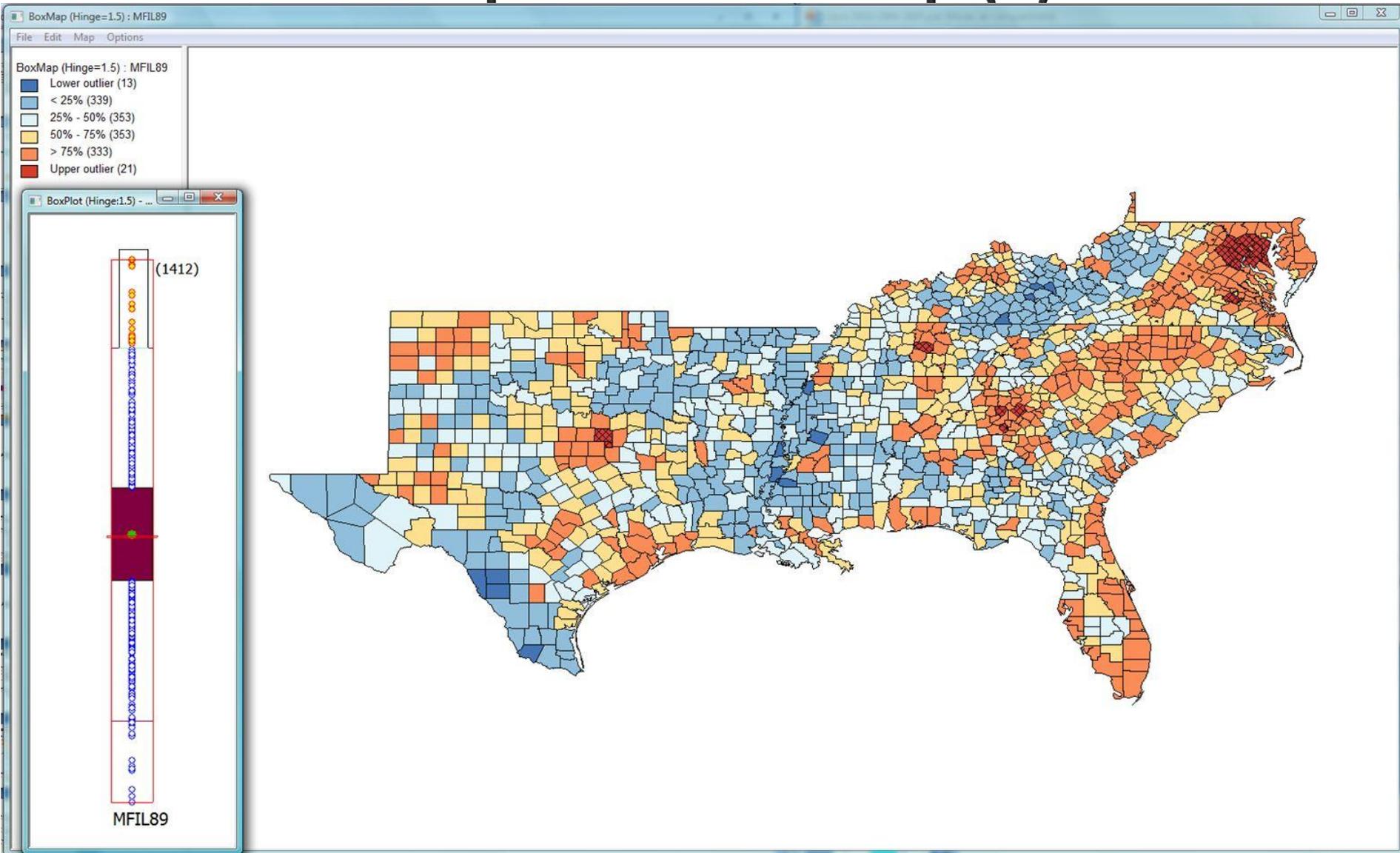
## **Visualiser les observations extrêmes : outliers**

# Définitions

## ▶ Plusieurs définitions

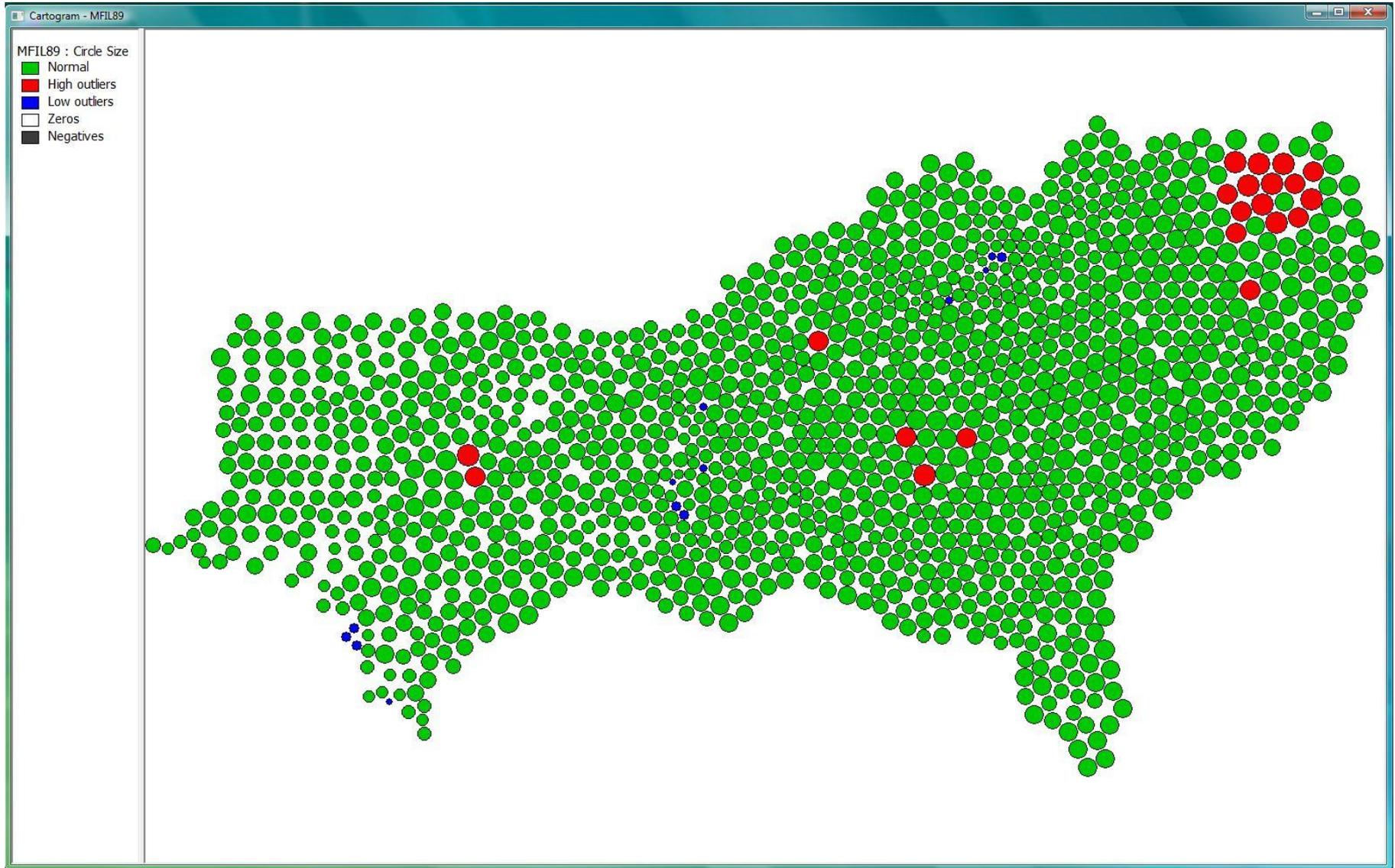
- ▶ Une observation est dite extrême si elle est :
  - supérieure à  $Q3 + 1,5 * (Q3 - Q1)$
  - inférieure à  $Q1 - 1,5 * (Q3 - Q1)$
  - $Q1$  et  $Q3$  sont respectivement les 1er et 3ème quartiles de la distribution,  $Q3 - Q1$  est l'écart interquartile
  - On utilise parfois 3 au lieu de 1,5 (hinge)
  - Visualisation à l'aide de la boîte à moustaches (box-plot) et de cartes (box-map) ou cartogrammes où ces observations sont particularisées
- ▶ Une observation est dite extrême si elle est :
  - supérieure à  $\bar{x} + 2\hat{\sigma}$
  - inférieure à  $\bar{x} - 2\hat{\sigma}$
  - Visualisation à l'aide de cartes de dispersion

# Box-plot et Box-map (1)



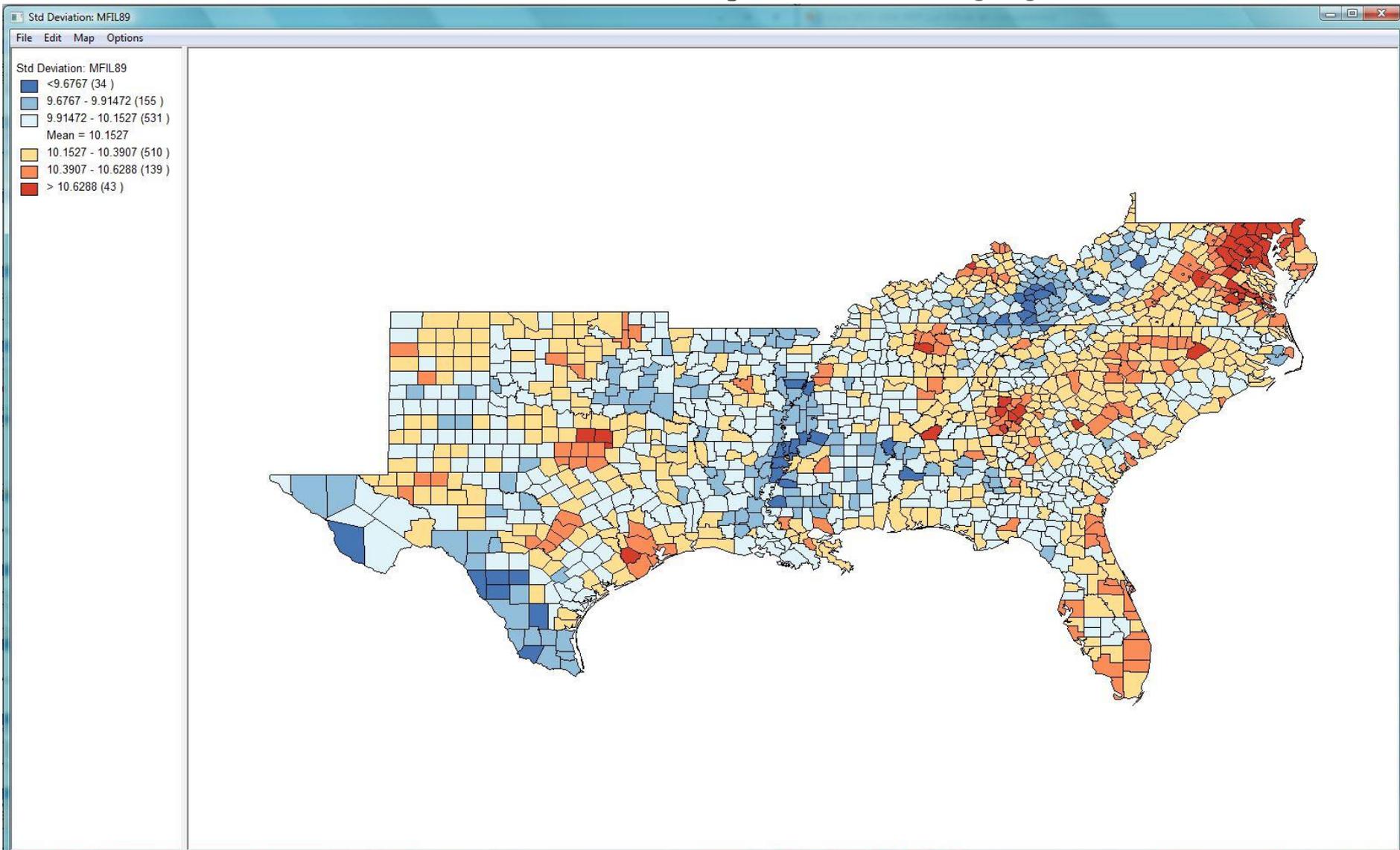
**Log du revenu médian des ménages dans les comtés du Sud-Est des Etats-Unis (1989)  
(1412 comtés)**

# Cartogramme (1)



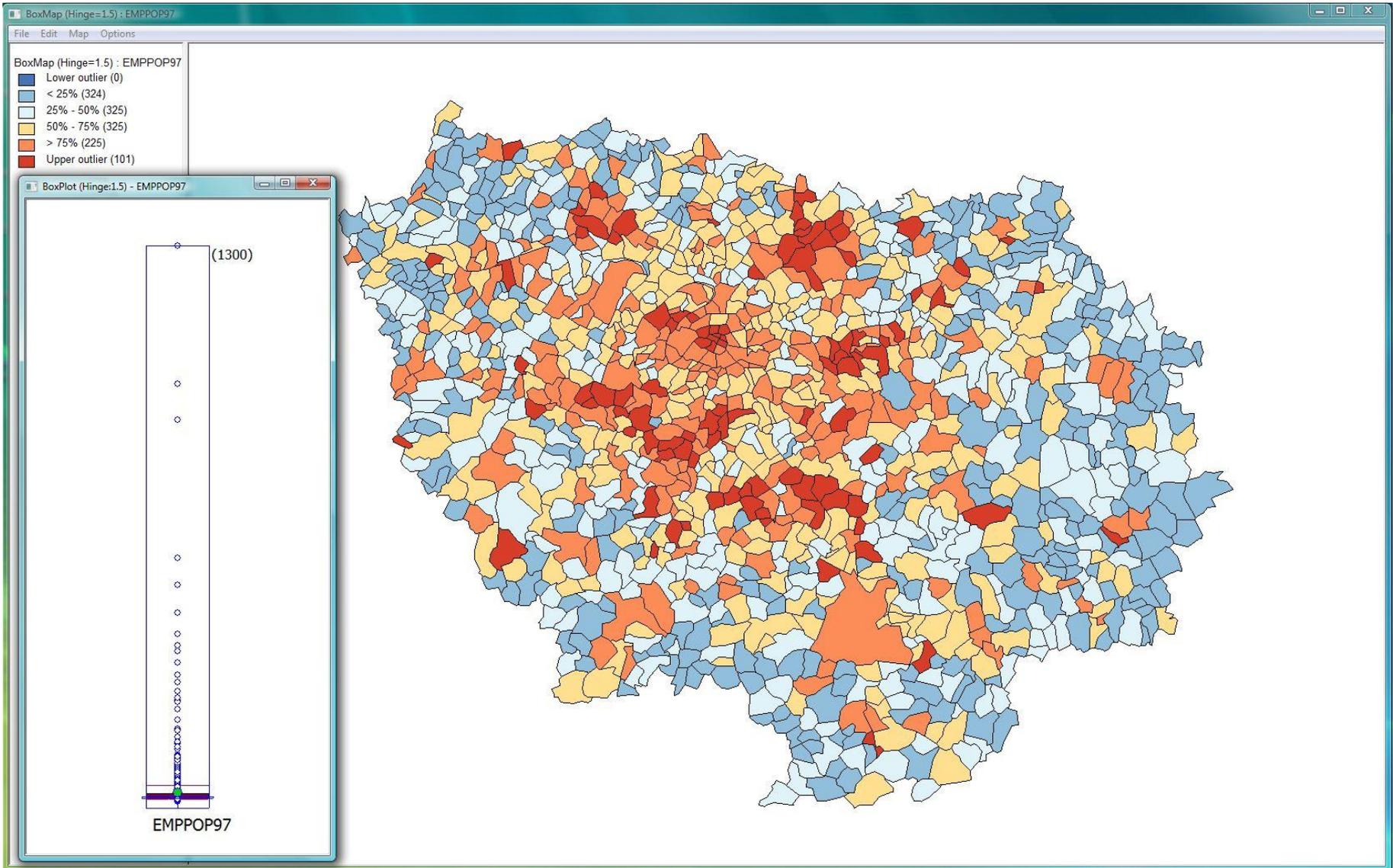
**Log du revenu médian des ménages dans les comtés du Sud-Est des Etats-Unis (1989)  
(1412 comtés)**

# Carte de dispersion (1)



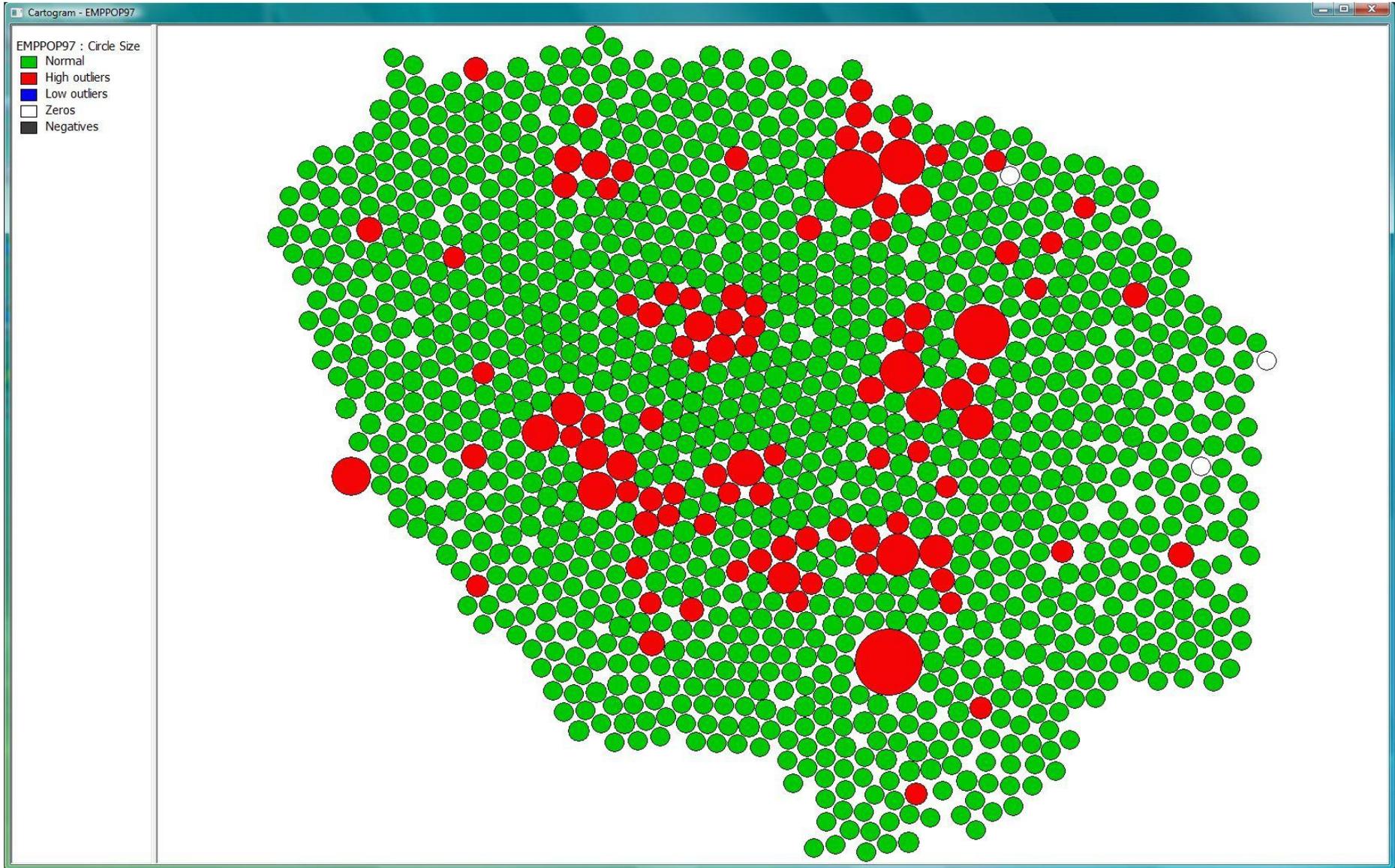
**Log du revenu médian des ménages dans les comtés du Sud-Est des Etats-Unis (1989)  
(1412 comtés)**

# Box-plot et Box-map (2)



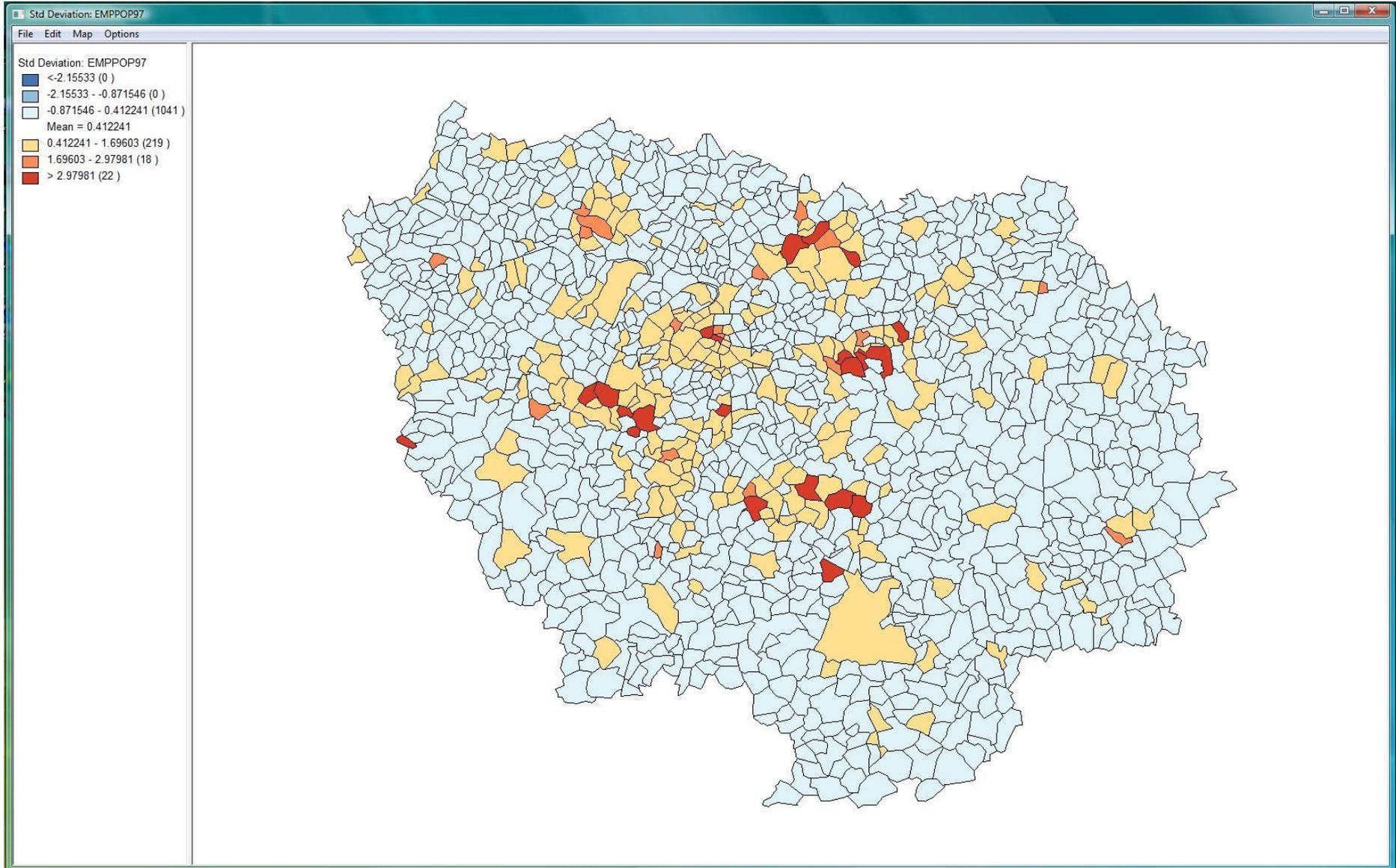
Ratio emploi sur population en Ile de France (1997)

# Cartogramme (2)



Ratio emploi sur population en Ile de France (1997)

# Carte de dispersion (2)



Ratio emploi sur population en Ile de France (1997)