

Analyse statistique de données spatiales

3èmes Journées “Ingé+” INRA à Dijon

Thibault LAURENT

GREMAQ/CNRS, Toulouse School of Economics

22 Octobre 2010

1 Introduction

2 Cartographie

3 Analyse Exploratoire de Données Spatiales

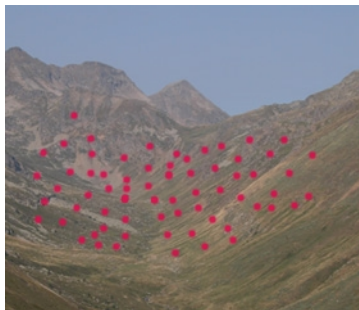
4 Autocorrélation spatiale

5 Etudes de données spatiales

6 Conclusion

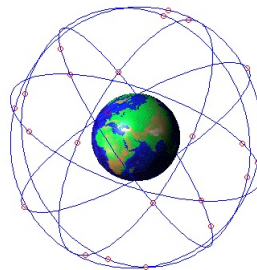
Etude de cas

Problématique : l'évolution du paysage sur une zone de pâturage laissée à l'abandon depuis 10 ans.



Mesurer l'information géographique

L'outil : le GPS



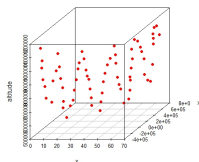
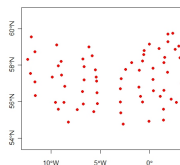
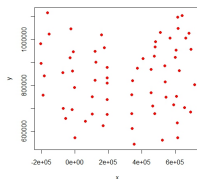
Mesurer l'information statistique

L'outil : relevés manuels, tablettes, etc effectués par le botaniste.
Obtenir à la fin un jeu de données avec n observations et p variables (quantitatives ou qualitatives) :

Individus	Variables		
	1	...	p
1	x_{11}	...	x_{1p}
2	x_{21}	...	x_{2p}
\vdots	\vdots	...	\vdots
n	x_{n1}	...	x_{np}

Quelle représentation spatiale des données ?

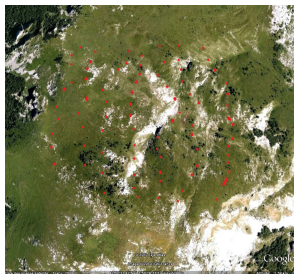
Une multitude de représentations possibles...



Aide au choix d'une représentation

Des critères possibles pour choisir :

- utiliser la norme d'un pays (dépend de la position par rapport aux pôles, géopolitique, etc)
- comment calculer la distance entre deux sites ?
- Superposition de plusieurs couches \Rightarrow choix d'un référentiel géodésique commun (ex : googlemaps).



Quelles méthodes statistiques ?

Une fois le référentiel géodésique choisi, les méthodes statistiques disponibles sont issues des domaines suivants :

- l'analyse exploratoire de données (Exploratory Data Analysis)
- l'économétrie spatiale
- la géostatistique
- les processus ponctuels

Dans cet exposé, nous traiterons essentiellement les deux premiers thèmes.

1 Introduction

2 Cartographie

- Coordinate Reference Systems (CRS)
- Geographic Information System (GIS)
- Cartographie avec R

3 Analyse Exploratoire de Données Spatiales

4 Autocorrélation spatiale

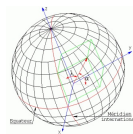
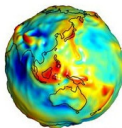
5 Etudes de données spatiales

6 Conclusion

Référentiel géodésique (datum)

Datum = système de référence permettant d'exprimer les positions au voisinage de la terre. Pour définir un datum :

- choix de la forme de la terre : sphère ou ellipsoïde (quel(s) rayon(s), quel coefficient d'applatissage, etc)
- choix des axes d'orientation (quel méridien de référence, etc)
- choix de l'origine

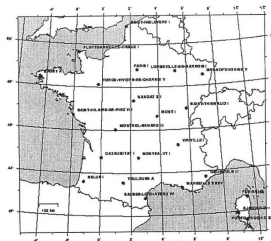
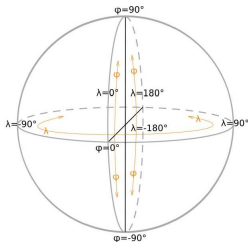


Exemples de datum

- NTF (Nouvelle Triangularisation de la France) : système utilisé en France jusqu'en décembre 2000, la plupart des cartes de l'IGN sont toujours dans ce système.
- RGF (réseau Géodésique Français) 1993 : système utilisé en France
- ED50 (European Datum) : système européen unifié.
- WGS84 (World Geodetic System) : système mondial mis au point par le Département de la Défense des Etats Unis et utilisé par le GPS.

Système de coordonnées géodésiques

- Latitude : mesure de l'angle ϕ par rapport à l'équateur
- Longitude : mesure de l'angle λ par rapport au méridien de référence

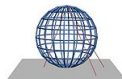


Unités de références

- degrés-minutes-secondes. Exemple : $\phi = 52^{\circ}28'00'' N$ et $\lambda = 4^{\circ}31'00'' W$
- degrés-décimaux. $\phi = 52.46667$ et $\lambda = -4.516667$

Projections

Une projection permet d'établir entre la surface de la Terre et le plan une correspondance telle que : $x = f_1(\phi, \lambda)$ et $y = f_2(\phi, \lambda)$.
Trois type de projection : cylindrique (ex : UTM associée à WGS84), conique (ex : Lambert II étendu associée à RGF) et azimutale.



Conséquence : sur un plan, le calcul des distances entre deux points se fait à l'aide de la formule de Pythagore.

Formules de passage (extrait de Wikipedia)

Formules de passage de latitude, longitude (φ, λ) aux coordonnées UTM (E, N) [modifier]

Les formules avec une précision du centimètre [modifier]

Les formules exactes sont compliquées et peu utilisables. Nous proposons des formules approchées avec une précision de l'ordre du centimètre.

Par convention, le géoïde WGS 84 décrit la terre par un [ellipsoïde de révolution](#) d'axe Nord-Sud, de rayon à l'équateur $a=6378,137$ km et d'[excentricité](#) $e=0,0818182$. On considère un point de [latitude géodésique](#) φ et longitude λ . Notons λ_0 la longitude du méridien de référence.

Les angles sont exprimés en radian. Voici des valeurs intermédiaires à calculer:

$$\nu(\varphi) = 1 / \sqrt{1 - e^2 \sin^2 \varphi}$$

$$A = (\lambda - \lambda_0) \cos \varphi$$

$$s(\varphi) = \left(1 - \frac{e^2}{4} - \frac{3e^4}{64} - \frac{5e^6}{256}\right)\varphi - \left(\frac{3e^2}{8} + \frac{3e^4}{32} + \frac{45e^6}{1024}\right)\sin 2\varphi + \left(\frac{15e^4}{256} + \frac{45e^6}{1024}\right)\sin 4\varphi - \frac{35e^6}{3072}\sin 6\varphi$$

$$T = \tan^2 \varphi, \quad C = \frac{e^2}{1 - e^2} \cos^2 \varphi, \quad k_0 = 0,9996$$

Dans l'hémisphère Nord $N_0 = 0$ et dans l'hémisphère Sud $N_0 = 10000$ km.

Voici les formules de passage donnant les coordonnées UTM E, N en kilomètres:

$$E = 500 + k_0 a \nu(\varphi) \left(A + (1 - T + C) \frac{A^3}{6} + (5 - 18T + T^2) \frac{A^5}{120} \right)$$

$$N = N_0 + k_0 a \left(s(\varphi) + \nu(\varphi) \tan \varphi \left(\frac{A^2}{2} + (5 - T + 9C + 4C^2) \frac{A^4}{24} + (61 - 58T + T^2) \frac{A^6}{720} \right) \right)$$

1 Introduction

2 Cartographie

- Coordinate Reference Systems (CRS)
- Geographic Information System (GIS)
- Cartographie avec R

3 Analyse Exploratoire de Données Spatiales

4 Autocorrélation spatiale

5 Etudes de données spatiales

6 Conclusion

Qu'est-ce qu'un SIG ?

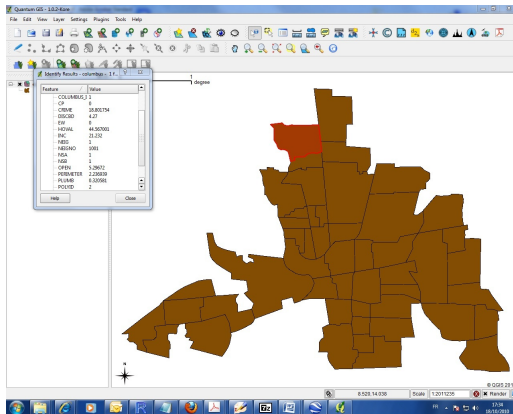
SIG (Système d'Information Géographique) : système informatique capable de gérer des données comme un gestionnaire de bases de données ordinaire, mais aussi de manipuler des objets géographiques, faire des cartes, agréger des zones, voir si une région est incluse dans une autre, si deux zones sont mitoyennes, etc.

Exemples de SIG :

- Solutions payantes : MapInfo, Arcview.
- Logiciels libres et validés PLUME : GRASS, gvSIG, Quantum GIS. Ce dernier, en plus d'avoir ses propres fonctionnalités, permet d'interfacer GRASS.

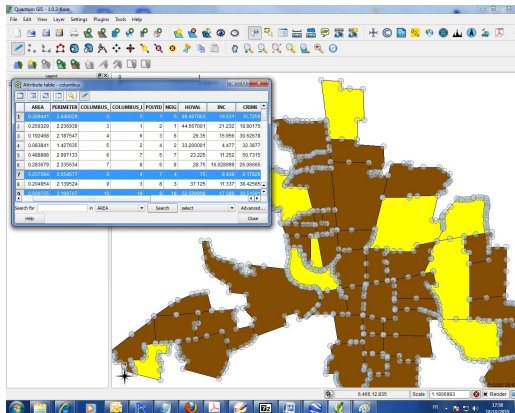
Quantum GIS (1)

Spatial Reference Identifier (SRID) : identification d'une unité spatiale



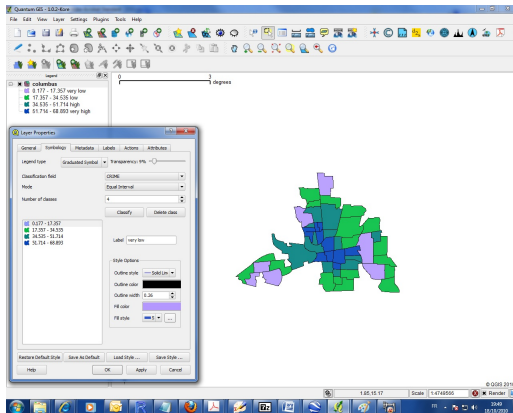
Quantum GIS (2)

Gestionnaire de base de données (possibilité d'utiliser les solutions PostgreSQL/PostGIS)



Quantum GIS (3)

Des outils statistiques de base comme le découpage d'une variable quantitative.

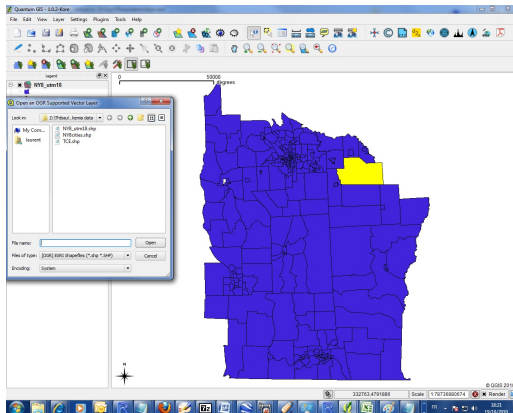


Format ESRI shapefile

C'est le format de référence d'import/export pour données géographiques (ESRI=Environmental Systems Research Institute, Inc.). Un ESRI shapefile est formé de :

- un fichier principal (.shp) qui contient toute l'information liée à la géométrie des objets décrits qui peuvent être : des points, des lignes ou des polygones.
- un fichier (.shx) qui stocke l'index de la géométrie
- un fichier dBASE (.dbf) pour les données attributaires
- des fichiers facultatifs comme un fichier sur les datums/projections (.prj).

Représentation d'un ESRI shapefile avec QGIS

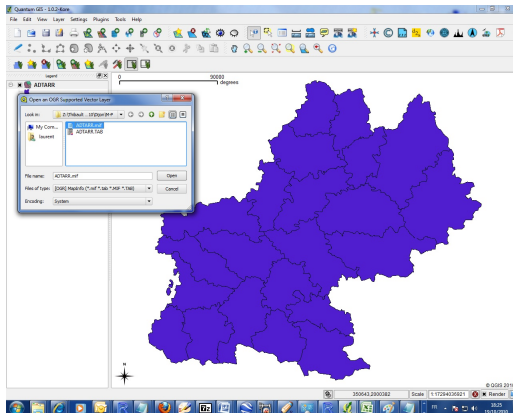


Format MIF/MID

Format propre à MAPINFO. Il contient :

- un fichier (**.mif**) avec les données graphiques + un en-tête décrivant certains paramètres (CRS, projection, index, etc.)
- un fichier (**.mid**) avec les identifiants et attributs des régions dont les contours sont donnés dans le fichier .mif

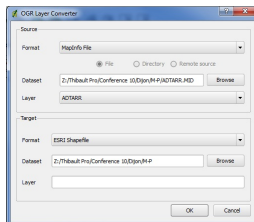
Représentation d'un MIF/MID avec QGIS



Solution GDAL/OGR (1)

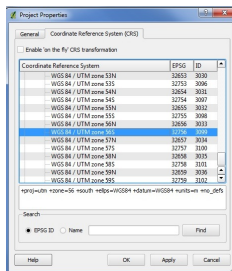
Outils intégrer dans Quantum GIS et R

- convertisseur de fichiers de données géographiques (shapefile, MIF/MID, KML, etc)
- superposition de couches de fichiers de données géographiques (.shp, etc.), de fichiers au format vectoriel (.eps, etc.) ou raster (.jpg, etc.).



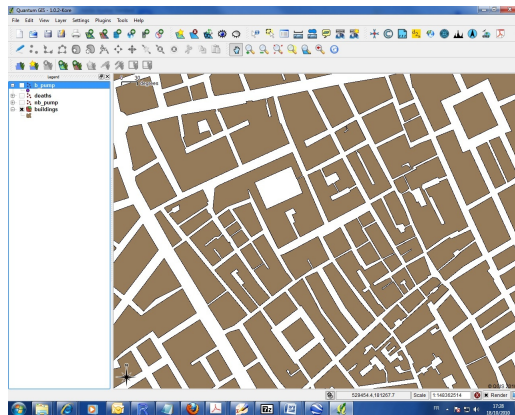
Solution GDAL/OGR (2)

- changement de référentiel géodésique (CRS) et de projection



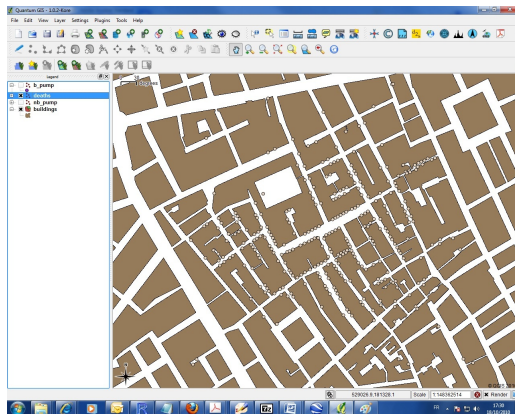
Gestion des couches (1)

Plan d'un quartier de Londres en 1850



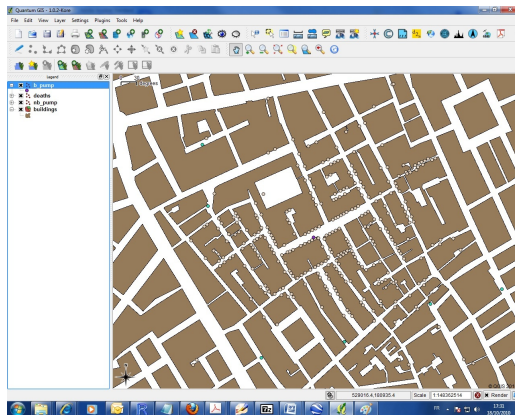
Gestion des couches (2)

Coordonnées géographiques d'individus décédés du cholera en 1850



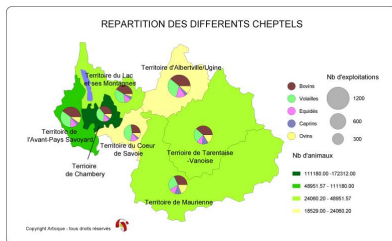
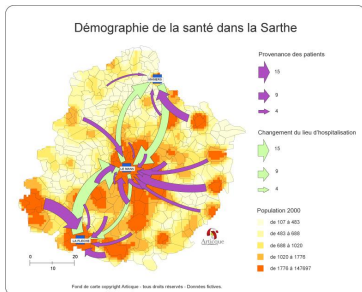
Gestion des couches (3)

Localisation des puits d'eau en 1850



Exemples de cartes réalisées avec des SIG

Exemples of maps produced by Articque
(<http://www.articque.com/>)



Limites des SIG

Les SIG n'utilisent pas d'outils d'analyse statistique sophistiqués (Moran plot, Variogramme, etc).

⇒ ces outils ont été développés par des chercheurs dans des logiciels dédiés comme GeoDa (L. Anselin) ou sur les logiciels statistiques usuels :

- Matlab (Spatial Econometric Toolbox développée par K. Pace et J. LeSage)
- R (de nombreux packages dédiés à l'analyse statistique de données spatiales)
- Stata, SAS/GIS, etc.

1 Introduction

2 Cartographie

- Coordinate Reference Systems (CRS)
- Geographic Information System (GIS)
- Cartographie avec R

3 Analyse Exploratoire de Données Spatiales

4 Autocorrélation spatiale

5 Etudes de données spatiales

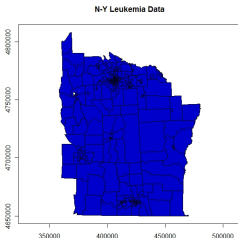
6 Conclusion

Statistique spatiale avec R

- R, logiciel libre validé PLUME :
<http://cran.r-project.org/>
- R. Biband, E. Pebesma, V. Gomez-Rubio (2008), *Applied Spatial Data Analysis with R*, Springer.
- Description des packages disponibles sur R dans
<http://cran.r-project.org/web/views/Spatial.html>
- Packages principaux pour la cartographie : `rgdal` et `maptools` (import/export de fichiers de données spatiales, convertir les CRS/projection, etc), `sp` (définition et méthodes de classes d'objet spatial propres à R) et `raster` (images de type raster).

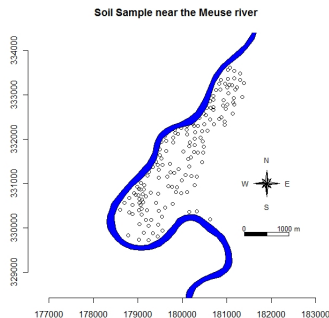
Importation et représentation d'un fichier de données spatiales

```
library(sp)
library(maptools)
xx = readShapePoly("NY8_utm18.shp", IDvar="AREAKEY",
  CRS("+proj=utm +zone=18 +datum=WGS84"))
plot(xx, main="NY data", col='blue3')
title("N-Y Leukemia Data")
```



Spatial object : SpatialPointsDataFrame

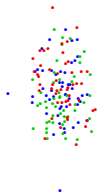
- Unité spatiale = un point $(X, Y) = \text{SpatialPoints}$
- Un jeu de données (`data.frame`) où `individu=site géographique`



Création d'un SpatialPointsDataFrame

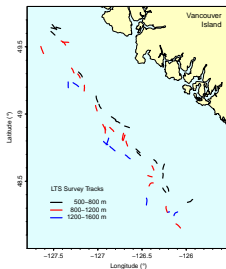
```
x = rnorm(150,0,1); y = runif(150)
sp.coord = SpatialPoints(cbind(x,y))
spdf.iris = SpatialPointsDataFrame(sp.coord,iris)
s.col=c("blue1","green3","red2")
plot(spdf.iris,pch=16,col=s.col[as.numeric(iris$Species)])
title("Variétés des Iris de Fisher")
```

Variétés des Iris de Fisher



Spatial object : SpatialLinesDataFrame

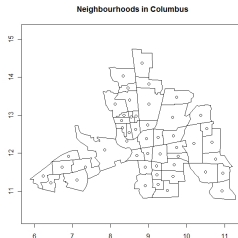
- Unité spatiale = un vecteur \overrightarrow{x} = SpatialLines
- Un jeu de données (data.frame) où individu=vecteur



Carte réalisée depuis le package PBSmapping

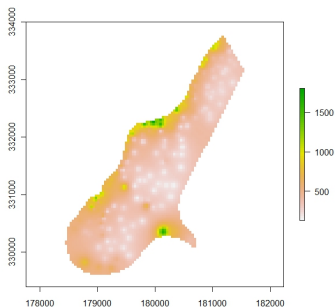
Spatial object : SpatialPolygonsDataFrame

- Unité spatiale = une zone délimitée par un contour = `SpatialPolygons`
- Un jeu de données (`data.frame`) où individu=contour
- Remarque : un contour peut être remplacé par un point en prenant par exemple le centroïde du contour.



Spatial object : SpatialPixelsDataFrame

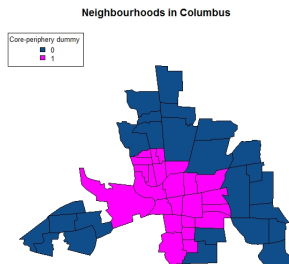
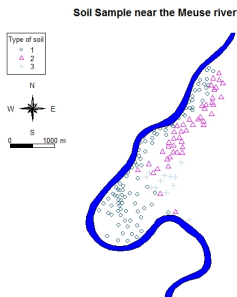
- Unité spatiale = un pixel = SpatialPixels
- Un jeu de données (data.frame) où individu=pixel



Carte réalisée avec le package raster

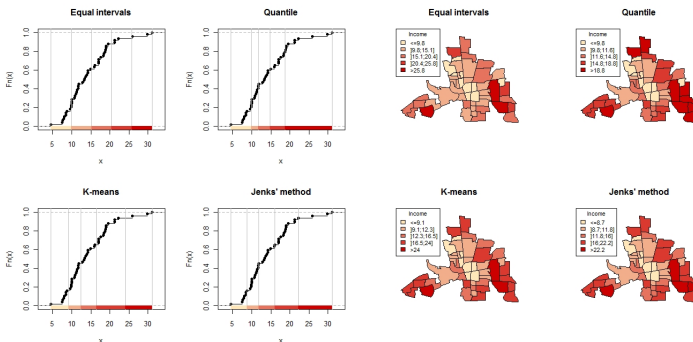
Représentation d'une variable qualitative

- options graphiques nombreuses sur R
- choix de couleurs important (package RColorBrewer)
- $R \simeq SIG$, sans interface...



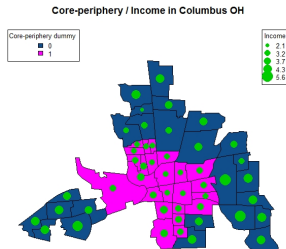
Représentation d'une variable quantitative

- discrétisation d'une variable continue en classe : bcp de méthodes statistiques dans le package `classInt` ("choix utilisateur", "intervalles égaux", "quantile", "CAH", etc).



Utilisation des Bubble

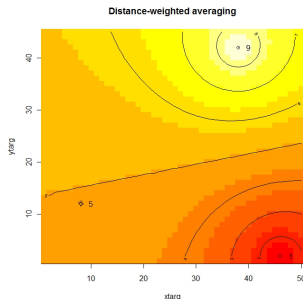
- Aire du point proportionnelle à la valeur d'une variable Z (fonction `bubble` dans `sp`).
- Superposition de deux informations provenant de 2 variables
- S. Tanimura et al. (2006), Proportional Symbol Mapping in R, *JSS*.



Aire des points dépend de la variable “revenu”.

Exemple d'interpolation spatiale (1)

- obtenir une vision spatiale globale d'une distribution (variables, résidus, coefficients de modèles GWR, etc.)
- Ex : **Distance-Weighted averaging** : $\hat{z} = \frac{z_1/d_1 + z_2/d_2 + z_3/d_3}{1/d_1 + 1/d_2 + 1/d_3}$
- Lignes de niveaux et couleurs avec les fonctions `image` et `contour`

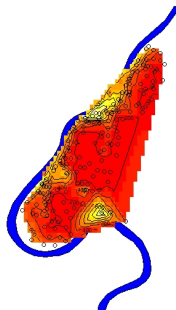


Exemple d'interpolation spatiale et représentation des lignes de niveaux

Exemple d'interpolation spatiale (2)

- Méthodes basées sur les splines cubiques dans R (package `akima`).

Akima interpolation of topsoil zinc concentration



Conclusion

Pour réaliser une “bonne” carte :

- solides connaissances multidisciplinaires : géographie, informatique, statistique, mathématiques, etc.
- ne pas oublier les classiques (légendes, échelle, source, etc)
- Un large choix de logiciels notamment dans les solutions libres

- 1 Introduction
- 2 Cartographie
- 3 Analyse Exploratoire de Données Spatiales
 - ...with GeoDa
 - ...with Matlab
 - ...with R
- 4 Autocorrélation spatiale
- 5 Etudes de données spatiales
- 6 Conclusion

Définition et objectifs de ESDA (Exploratory Spatial Data Analysis)

- Définition : combiner l'information provenant de l'analyse de données avec l'information spatiale en utilisant la cartographie, l'interactivité entre graphiques statistiques et cartes, des méthodes propres en créant par exemple des matrices de voisinage, etc
- Objectifs : donner des éléments de réponse aux problèmes classiques de l'analyse de données spatiales : détection d'autocorrélation spatiale, détection de valeurs atypiques (localement ou globalement), détection d'hétérogénéité spatiale...

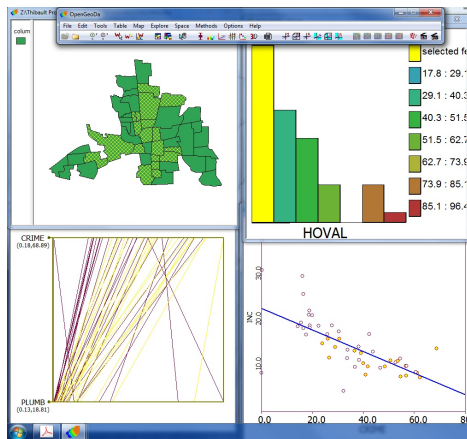
- 1 Introduction
- 2 Cartographie
- 3 Analyse Exploratoire de Données Spatiales
 - ...with GeoDa
 - ...with Matlab
 - ...with R
- 4 Autocorrélation spatiale
- 5 Etudes de données spatiales
- 6 Conclusion

Solution GeoDa (1)

Logiciel libre (interfacé) développé par Luc Anselin

- reprend les outils de base du SIG (manipulation de base de données, etc)
- couple de façon interactive une carte avec un graphique statistique (histogramme, boîte à moustache, nuage de points, etc)
- création de matrices de voisinage → représentation du diagramme de Moran
- téléchargeable sur <http://geodacenter.asu.edu/ogeoda> avec une documentation détaillée

Solution GeoDa (2)



Limites de GeoDa

- nombre limité de graphiques statistiques
- un seul format de données spatiales pris en compte (.shp)
- matrice de voisinage basée uniquement sur les distances et + proches voisins
- uniquement dédié à l'économétrie spatiale

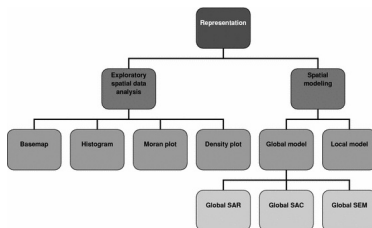
Cependant, le logiciel semble évoluer (la possibilité d'être bientôt coupler avec R?).

- 1 Introduction
- 2 Cartographie
- 3 Analyse Exploratoire de Données Spatiales
 - ...with GeoDa
 - ...with Matlab
 - ...with R
- 4 Autocorrélation spatiale
- 5 Etudes de données spatiales
- 6 Conclusion

Solution Matlab (1)

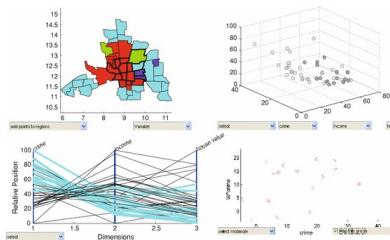
Plusieurs toolbox développées par James LeSage et Kelley Pace

- disponible sur <http://www.spatial-econometrics.com/>
- Un outil exploratoire de données spatiales décrit dans : Liu et Lesage (2009), *Arc_Mat : a Matlab-based spatial data analysis toolbox*, JGS.



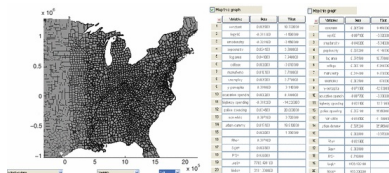
Solution Matlab (2)

Analyse interactive de données spatiales



Solution Matlab (3)

Analyse des résidus des modèles de régression spatiale



- 1 Introduction
- 2 Cartographie
- 3 Analyse Exploratoire de Données Spatiales
 - ...with GeoDa
 - ...with Matlab
 - ...with R
- 4 Autocorrélation spatiale
- 5 Etudes de données spatiales
- 6 Conclusion

Solution R

Package GeoXp : analyse exploratoire interactive de données spatiales

- package téléchargeable sur CRAN (<http://cran.r-project.org>) avec un manuel technique et une vignette (outil Sweave).
- utilise les classes d'objet définies par R. Bivand dans sp
- T.Laurent, A. Ruiz-Gazen et C. Thomas-Agnan (à paraître), "GeoXp : an R package for Exploratory Spatial Data Analysis", *JSS*.
- NB : GeoXp, également disponible sur Matlab

Utilisation de GeoXp

Arguments d'entrée d'une fonction `xxxmap` (ex : `histomap`, `barmap`, `densitymap`, etc) :

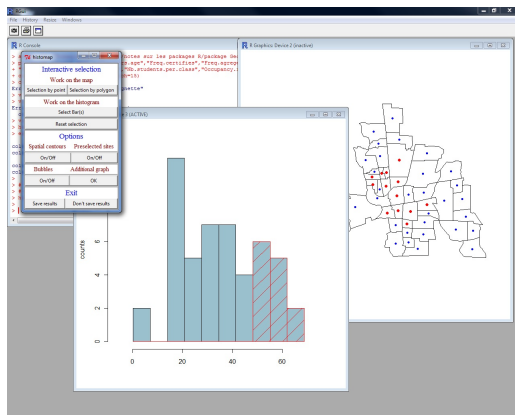
- Un objet de classe `SpatialxxxDataFrame` en entrée
- le nom d'une variable (analyse univariée) ou plusieurs (multivariée)
- un certain nombre d'options (couleurs, légendes, etc)

3 types de fonctions :

- cas univarié : `histomap(columbus,"CRIME")`
- cas bivarié : `polyboxplotmap(columbus,c("CRIME","CP"))`
- fonctions avec utilisation d'une matrice de voisinage (`moranplotmap`, `neighbourmap`, etc.)

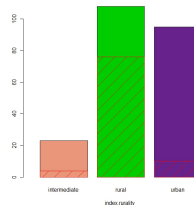
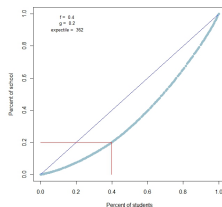
...with R

Une interface Tcl/Tk, une carte et un graphique



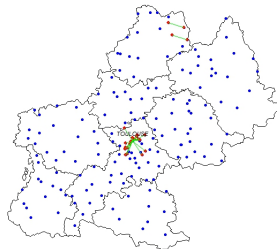
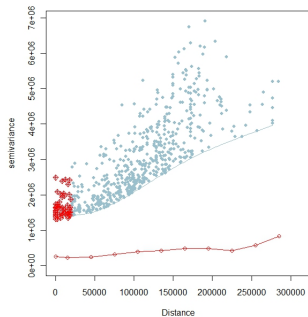
Courbe de Lorentz

```
data(mp.school)
ginimap(mp.school, "nb_student")
```



Variogramme

```
variocloudmap(mp.school, "Nb_Student_per_class")
```



Perspectives pour GeoXp

- création d'une seule interface commune à toutes les fonctions et couplage simultané de plusieurs graphiques avec la carte (comme GeoDa)
- sélection de polygones sur la carte

1 Introduction

2 Cartographie

3 Analyse Exploratoire de Données Spatiales

4 Autocorrélation spatiale

- Matrice de voisinage

- Tests statistiques

5 Etudes de données spatiales

6 Conclusion

Création de matrices de voisinage sous R (package `spdep`)

Méthodes basées sur la géométrie des unités spatiales :

- matrice de contiguïté (voisin = frontières partagées)
`ny1_nb=poly2nb(xx)`
- matrice basée sur la triangularisation de Delaunay (les centroïdes sont les points de référence)
`coords=coordinates(xx)`
`ny2_nb=tri2nb(coords)`

Méthodes basées à partir d'un calcul de distance :

- matrice basée sur les plus proches voisins
`ny3_nb=knn2nb(knearneigh(coords,4))`
- matrice basée sur une distance seuil
`ny4_nb=dnearneigh(coords,0,20000)`

Résumé d'une matrice de voisinage

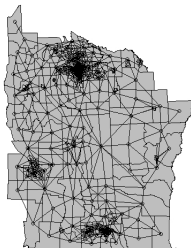
- un objet `nb` donne l'information sur qui est voisin de qui.
- pour obtenir la valeur w_{ij} des cellules, il faut créer un objet `listw` (option `style="B"` pour 0 ou 1, `style="W"` pour normaliser en ligne, etc.).
- le code suivant donne un résumé d'un objet `nb` ou `listw` (nombre moyen de voisins, résumé sur les distances géographiques entre voisins, etc)
`summary(ny2_nb, coords)`
`summary(listw2nb(ny2_nb), coords)`

L'analyse descriptive (résumé et représentation) de la matrice W peut aider l'utilisateur dans le choix de la méthode.

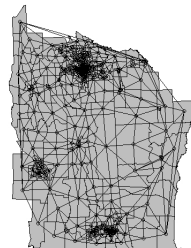
Représentation des matrices de voisinage sous R (1)

```
plot(xx,col='grey')  
plot(ny1_nb,coords,add=TRUE)
```

Matrice de contiguité

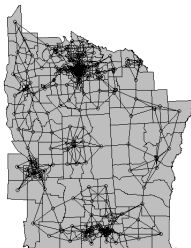


Triangularisation de Delaunay

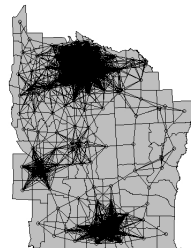


Représentation des matrices de voisinage sous R (2)

4 plus proches voisins



Distance seuil

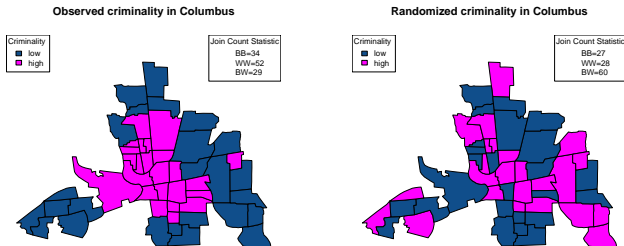


Test de "Join count"

- Mesure le degré d'autocorrélation spatiale d'une variable qualitative.

```
CRIME = columbus@data$CRIME
```

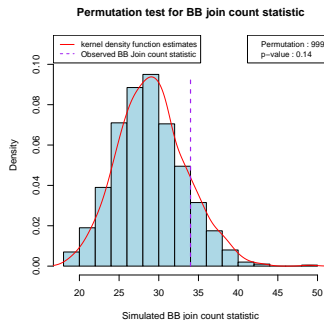
```
HICRIME = cut(CRIME, breaks = c(0, 35, 80), labels =  
c("low", "high"))
```



Test de Join count sous R

- Possibilité d'utiliser un test de Monte-Carlo et représenter graphiquement les résultats

```
joincount.mc(HICRIME, nb2listw(col.gal.nb, style="B"),  
nsim=999)
```



Résultat du test de Join count

	Joincount	Expected	Variance	z-value
low : low	34.00	29.34	18.64	1.08
high : high	52.00	26.99	17.65	5.95
high : low	29.00	58.67	26.04	-5.81

TABLE: Utilisation d'une matrice de voisinage binaire

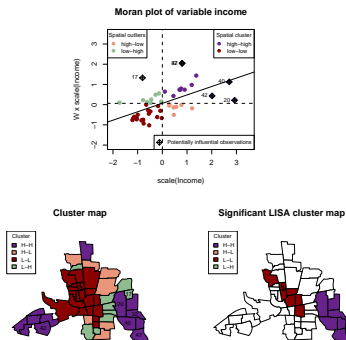
Remarque : le choix de la matrice de voisinage peut influencer les tests statistiques

Diagramme de Moran

- utile pour visualiser l'autocorrélation spatiale d'une variable quantitative
- représentation de $W \times X$ en ordonnées et X en abscisses avec la fonction `moran.plot`
- à tester sur les résidus d'un modèle OLS pour vérifier si l'utilisation d'un modèle spatial est utile

Représentation d'un diagramme de Moran

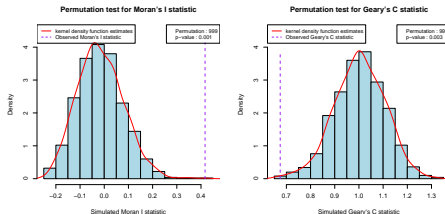
- fonction `moran.plot` de `spdep`
- représentation de deux cartes coloriées par quadrant (sur la 2ème carte, seuls les LISA - Local Indicators of Spatial Association - significatifs apparaissent)



Test de Moran sous R

- Possibilité d'utiliser un test de Monte-Carlo et représenter graphiquement les résultats

```
moran.mc(income, nb2listw(col.gal.nb, style="W"),
nsim=999)
```



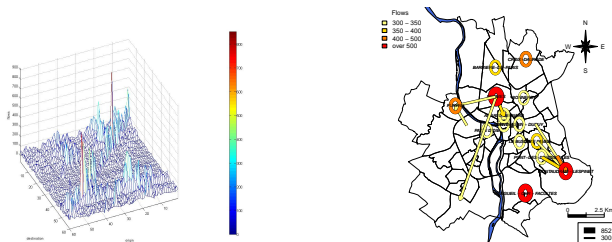
Remarque : le test de Geary (fonction `geary` ou `geary.mc`) est une alternative au test de Moran

- 1 Introduction
- 2 Cartographie
- 3 Analyse Exploratoire de Données Spatiales
- 4 Autocorrélation spatiale
- 5 Etudes de données spatiales**
 - Modèles de Flux
 - Prédictions sur modèles spatiaux
 - Réseaux sociaux
- 6 Conclusion

Variable à expliquer Y

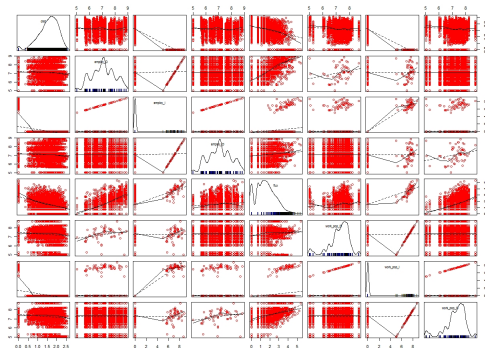
Les flux de population domicile/travail dans 60 iris de la région toulousaine

- Travail effectué avec Christine Thomas-Agnan (TSE) et Camille Rousseau (INSEE)
- graphique sous Matlab et carte sous R



Variables explicatives X

Variables observées par iris → variables à l'origine et à la destination



Modèle spatial utilisé : SAR ou Lag Model

$$Y = \rho_d W_d Y + \rho_o W_o Y + \alpha + X_d \beta_d + X_o \beta_o + \gamma g + \epsilon_{OD}$$

Pourquoi Matlab et pas R ? En 2008 :

- pas de code disponible sur R (tout comme les modèles probit et tobit, disponibles sur Matlab uniquement)
- quand $n=60$ (nombre d'iris), le modèle contient $N=3600$ observations, mais dès que n augmente, R rencontre des problèmes numériques pour calculer le Jacobien d'une matrice.
- les effets directs et indirects étaient calculés uniquement sur Matlab

En 2010, R a comblé une partie de ces lacunes en utilisant un package qui traite les calculs sur les matrices creuses...

Estimation des coefficients d'un SAR "classique"

- Les résultats d'un SAR classique sont donnés par la fonction `sar` ou `gsar` (bayésien) sur Matlab et `lagsarlm` sur R. Ils renvoient pratiquement les mêmes résultats (méthodes d'optimisation différentes sur chaque logiciel).

Variables	Coefficient	t-stat	p-level
α	-2.382467	-32.923365	0.000000
α_I	-0.940341	-2.362556	0.018149
dist	-0.108924	-6.680950	0.000000
work_pop_O	0.264135	31.209757	0.000000
work_pop_D	-0.130825	-13.944637	0.000000
work_pop_I	0.686877	9.682889	0.000000
employ_O	-0.026996	-4.347458	0.000014
employ_D	0.301888	39.984190	0.000000
employ_I	-0.183899	-4.218990	0.000025
ρ	0.636965	41.139750	0.000000
R^2	0.6476		
σ^2	0.2486		

Effets directs/indirects

- La fonction `impacts` de R renvoie l'estimation des effets directs et indirects du modèle SAR

Variables	Indirect Effect	std deviation	t-stat	p-level
employ_I	-0.3045	0.0746	-4.0808	0.0000
work_pop_D	-0.2198	0.0142	-15.4451	0.0000
dist	-0.1886	0.0179	-10.5071	0.0000
employ_O	-0.0456	0.0069	-6.5937	0.0000
work_pop_O	0.4445	0.0227	19.6251	0.0000
employ_D	0.5085	0.0239	21.2863	0.0000
work_pop_I	1.1552	0.1301	8.8783	0.0000
Variables	Direct Effect	std deviation	t-stat	p-level
employ_I	-0.1889	0.0453	-4.1735	0.0000
work_pop_D	-0.1364	0.0061	-22.2428	0.0000
dist	-0.1170	0.0096	-12.1541	0.0000
employ_O	-0.0283	0.0040	-6.9981	0.0000
work_pop_O	0.2757	0.0069	39.8998	0.0000
employ_D	0.3154	0.0056	56.0986	0.0000
work_pop_I	0.7167	0.0724	9.9029	0.0000

Variables	Total Effect	std deviation	t-stat	p-level
employ_I	-0.4934	0.1199	-4.1158	0.0000
work_pop_D	-0.3562	0.0204	-17.4916	0.0000
dist	-0.3056	0.0276	-11.0820	0.0000
employ_O	-0.0739	0.0110	-6.7429	0.0000
work_pop_O	0.7203	0.0296	24.3649	0.0000
employ_D	0.8239	0.0295	27.9190	0.0000
work_pop_I	1.8719	0.2025	9.2445	0.0000

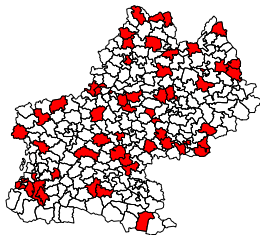
- 1 Introduction
- 2 Cartographie
- 3 Analyse Exploratoire de Données Spatiales
- 4 Autocorrélation spatiale
- 5 Etudes de données spatiales**
 - Modèles de Flux
 - **Prédictions sur modèles spatiaux**
 - Réseaux sociaux
- 6 Conclusion

Objectif

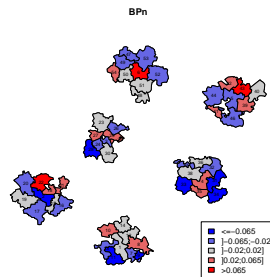
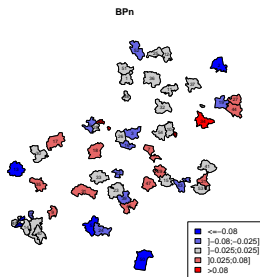
- Travail effectué avec Christine Thomas-Agnan (TSE) et Michel Goulard (INRA)
- Prédire le nombre d'habitants (variable Y) dans un canton en utilisant l'information X propre à ce canton et l'information Y et X de ses voisins.
- Modèle utilisé : SAR
- logiciel utilisé : R

Résultats sur données simulées

En rouge, les sites à prédire :



Cartographie des résidus



1 Introduction

2 Cartographie

3 Analyse Exploratoire de Données Spatiales

4 Autocorrélation spatiale

5 Etudes de données spatiales

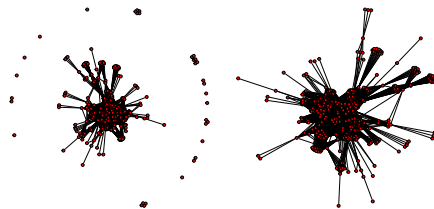
- Modèles de Flux
- Prédications sur modèles spatiaux
- Réseaux sociaux

6 Conclusion

Graphique de réseau social

Utiliser les méthodes de l'analyse statistique de données spatiales à la théorie des graphes

- Travail effectué avec Nathalie Villa-Vialaneix (IMT)
- Variable (qualitative ou quantitative) observée sur les noeuds.
- Calcul du test Join Count ou Moran avec la matrice W obtenue grâce aux arrêtes.



- 1 Introduction
- 2 Cartographie
- 3 Analyse Exploratoire de Données Spatiales
- 4 Autocorrélation spatiale
- 5 Etudes de données spatiales
- 6 Conclusion**

Conclusion

Analyse de données spatiales = vaste sujet. N'ont pas été traités, mais possible sur R :

- de nombreux modèles de régression en économétrie spatiale (GWR, SEM, etc), estimation du SAR/SEM par GMM, etc.
- les outils classiques de la géostatistique (variogramme, interpolation spatiale, etc) avec de nombreuses applications INRA
- les outils modernes des processus ponctuels

... et encore beaucoup de choses. Merci de votre attention