Exploratory interactive tools for spatial data analysis

Thibault LAURENT, Anne RUIZ-GAZEN and Christine THOMAS-AGNAN

GREMAQ, Toulouse School of Economics

8 juillet 2009

use R!

Plan

1 Introduction

2 GeoXp : Interactivity between statistic plots with maps

3 Innovations of GeoXp

4 Perspectives

Data Set

A data set of n observations and p variables (discrete, continuous) :

		Variables	
Individuals	1	...	p
1	x_{11}	...	x_{1p}
2	x_{21}	...	x_{2p}
⋮	⋮	...	⋮
n	x_{n1}	...	x_{np}

Data Set

A data set of *n* observations and *p* variables (discrete, continuous) :

		Variables	
Individuals	1	...	p
1	x_{11}	...	x_{1p}
2	x_{21}	...	x_{2p}
⋮	⋮	...	⋮
n	x_{n1}	...	x_{np}

Each observation is associated to a a spatial object (a point, a line or a polygon).

Definition and objectives of Spatial Exploratory Data Analysis (SEDA)

- Definition : combine information given by techniques of EDA with spatial information using mapping, interactivity between statistical plots with maps or by creating new methods using for example neighborhood matrices.

- Objectives : give qualitative response to the usual problems of spatial analysis like detection of spatial autocorrelation, detection of local or global extreme values, detection of spatial heterogeneity...

Mapping with `maptools` packages

Soil Sample near the Meuse river

Income discretised by Jenks' method
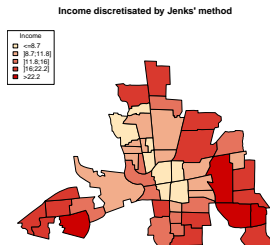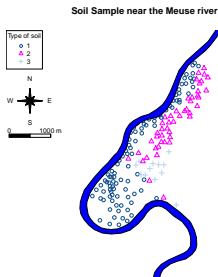
Mapping with `maptools` and `PBSmapping` packages
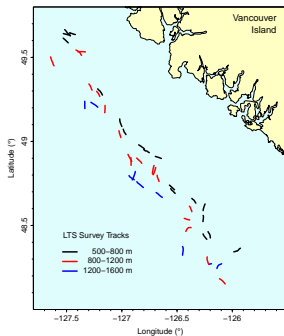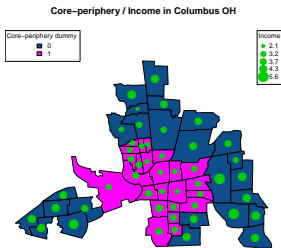
Some advices

- Which projection system ? Which reference system (very important to give units to the distances) ?

- Choice of palettes is important (see `RColorBrewer` package).

- How find class intervals for continuous numerical variables : `classInt` package ("equal interval", "quantile", "K-means", etc).

- Interpolation can be useful to have a global view of spatial distribution of : variables, residuals, coefficients of GWR model, etc (see package `akima`).

Interactivity between statistic plots with maps

Package GeoXp

- GeoXp (C. Thomas-Agnan et al.), an R package for interactive exploratory spatial data analysis : it links dynamically statistical plots like boxplot, histogram, scatterplot, etc. with a map.

- Selection of a zone on the map results in the automatic highlighting of the corresponding points on the statistical graph.

- Selection of a portion of the graph results in the automatic highlighting of the corresponding sites on the map.

EDA functions

The main statistical plots of EDA have been included in GeoXp :

- **Descriptive** functions : univariate or bivariate graphs such as histogram, barplot, scatterplot, etc.

- **Multivariate** functions : principal component analysis, cluster analysis.

Example 1 : Kernel Density function estimate (1)

Plot kernel density function estimates for the whole distribution and for the distribution of the sites selected by the user (by points or polygons) on the map (see function bkde in package KernSmooth) in red on the graphic.

densitymap(longitude,latitude,variable,...)
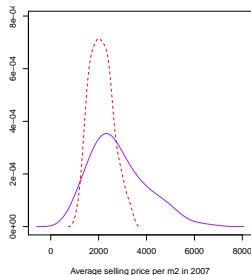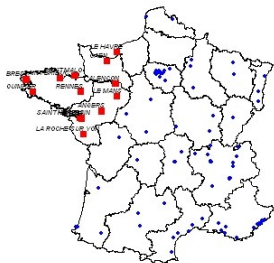
Example 1 : Kernel Density function estimate (2)

User selects an interval on the density curve by mouse clicking on the graph on the extremities of interval and selected sites are then represented in red on the map.
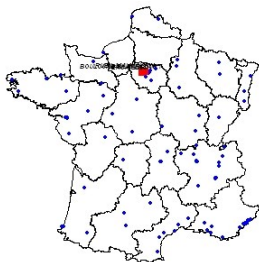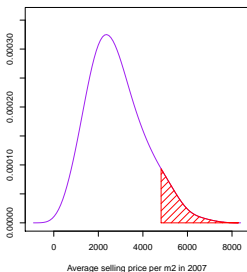
Example 2 : Principal component analysis (1)

The function pcamap draws the plots summarizing a **generalized Principal Component Analysis** (PCA) and permits an interactivity between the scatterplot of the individuals projected on a chosen principal component plane with the map.

pcamap(longitude,latitude,mat.var,...)
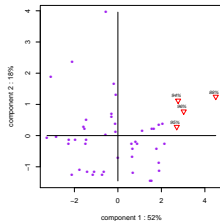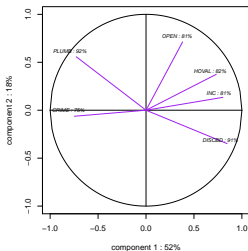
mat.var=log(1+c("HOVAL","INC","CRIME","OPEN","PLUMB","DISCBD"))

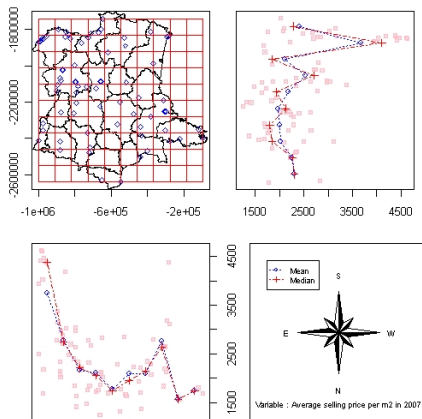# Example 2 : Principal component analysis (1)
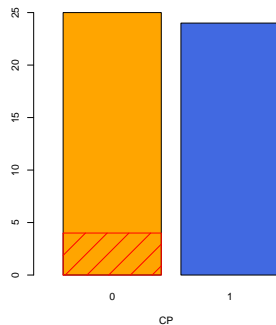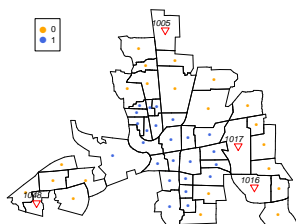
Other Examples
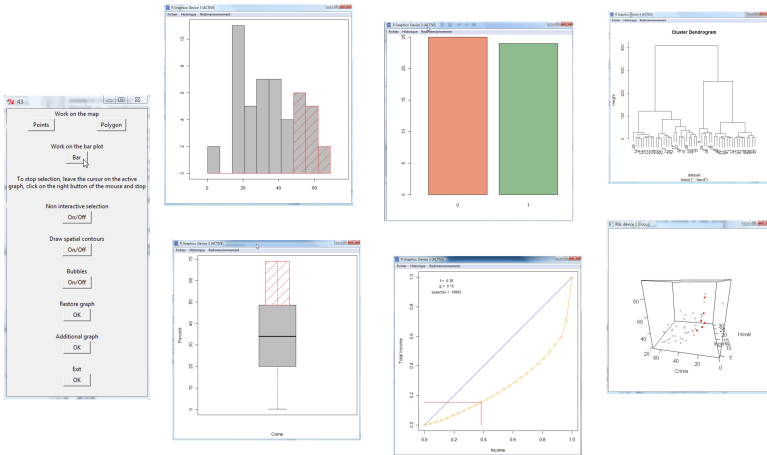
Specific statistical plots

Statistical plots specific to spatial analysis :

- **Geostatistic** functions : angle plot, drift plot.
- **Spatial Econometric** functions : Moran plot, neighbour plot

Drift map (Cressie, 1993)

`driftmap(longitude,latitude,variable,...)`

## Angle plot (Brunsdon, 1998)

- represents the absolute difference between the values of a variable at two sites s_i and s_j as a function of the angle between vector $\vec{s_i s_j}$ and the x-axis,

- possibility to draw only couples of sites whose absolute difference is larger than some α regression quantile smoothing splines.
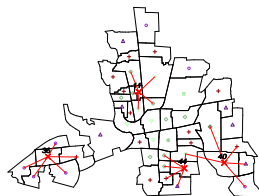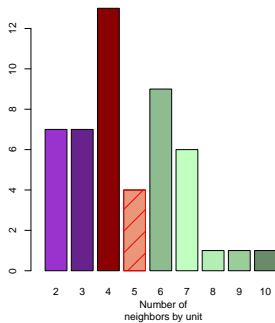
Angle plot (Brunsdon, 1998)

`angleplotmap(longitude,latitude,variable,...)`
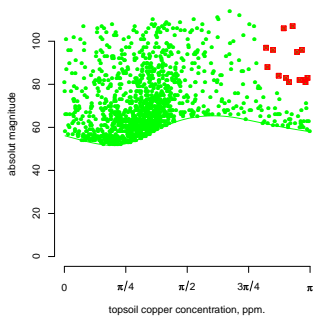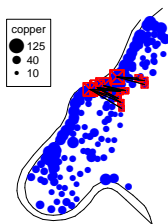
Innovations of GeoXp

1 Introduction

2 GeoXp : Interactivity between statistic plots with maps
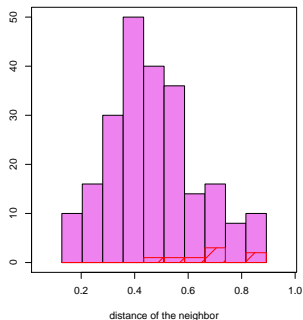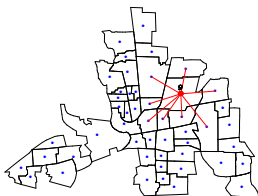
3 Innovations of GeoXp

4 Perspectives

Analysis of a neighborhood structure (1)

```
barnbmap(obj.nb, coords,...)
```
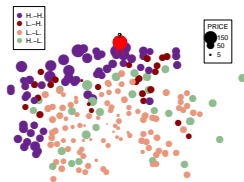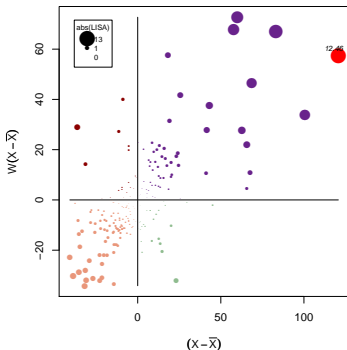
Analysis of a neighborhood structure (2)

```
histnbmap(obj.nb, coords,...)
```

Moran plot

moranplotmap(long, lat, var, mat.W, ...)
Test of spatial autocorrelation (gaussian version or permutation test)

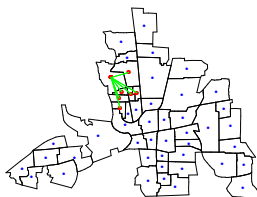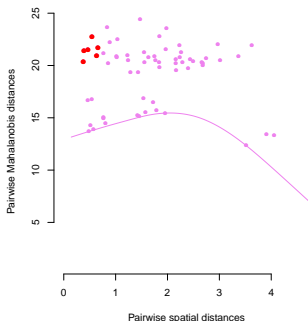$I = 0.49$ ($p - value < 0.0001$)

Pairwise map (Filmozer et al.)

"Local outliers" : observations which differ from their neighbours.
We propose to use pairwise robust Mahalanobis distances :

$$MD(x_i; x_j; C) = \{(xi - xj)'C^{-1}(xi - xj)\}^{1/2}$$

R-packages : `mvoutliers` and `rrcov`.

Pairwise map (Filmozer et al.)
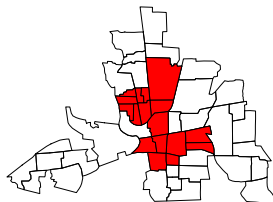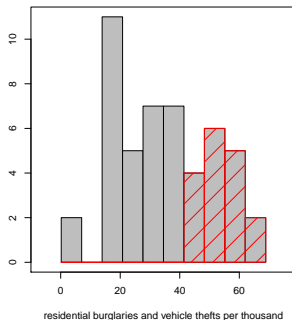
```
pairwisemap(long, lat, obj.nb, mat.var,...)
```



```
mat.var=log(1+c("HOVAL","INC","CRIME","OPEN","PLUMB","DISCBD"))
```

Perspectives

1 Introduction

2 GeoXp : Interactivity between statistic plots with maps
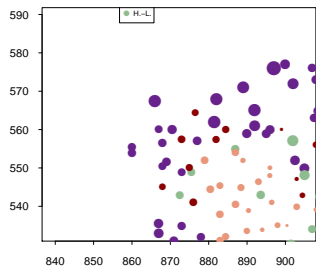
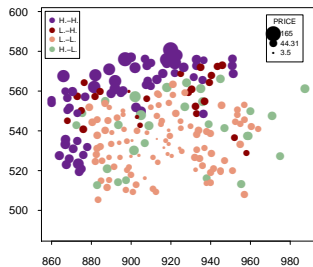**3** Innovations of GeoXp

4 Perspectives

Perspectives (1)

Use the sp norm (with the help of Roger Bivand)

residential burglaries and vehicle thefts per thousand

Perspectives (2)

Add zoom

Perspectives (3)

- A new interface (close to `Rcmdr` window)

- Add a number "unlimited" of interactive graphic window as in `GeoDa`

- Add origin-destination flow graphics

- A R vignette should appear on CRAN soon

- Thanks for attention ! (my e-mail : thibault.laurent@univ-tlse1.fr)