CO-405

INTERACTIVE METHODS FOR SPATIO-STATISTICAL DISTRIBUTION DISCRETIZATION

JOSSELIN D., PRACHE B.
UMR ESPACE, AVIGNON, FRANCE

1. THE DISTOGRAM

The Distogram (Josselin, 2001) has been developed to provide an interactive method to discretize distributions (Cauvin, 2008). It is a peculiar histogram (figure 1) which aims at dynamically analysing spatial distributions of studied data and their statistical distribution. As a "real" histogram (whose class root can change), the surface of its classes are proportional to the number of individuals, so it doesn't require any Y axis. It includes different classical automatic methods to separate individuals in classes: classes of equal widths, quantile repartitions, discretization based on several centres and widths (e.g. standard-deviation) and now, methods based on variance analysis. Also, users can move the classes limits with the mouse, tune the number of classes to see what happens on the map, separate or aggregate on-fly some contiguous classes, or process some distribution transformations to focus more accurately on some parts of the data. Spatial autocorrelation indexes (LISA, Anselin, 1990) are also processed in real time to help for discretizing (figure 2). Thanks to its high interactivity (LipStat, Tierney, 1990) and the multiple representations interacting with each others, the Distogram can be used to build spatial partitioning. The objective for the user is to aggregate the data in such a way the classes are the more homogeneous possible and each classes limit is located at an inflexion point in the distribution of the values. Thus, the Distogram is useful for finding spatial discontinuities. Some tests has been made previously, showing the majority of the users converge to close spatial partitioning (Josselin, 2003) using automatic manual delineation (figure 2).

The Distogram main functionalities are (the D rule): Double view (maping vs statistics), Dynamic Distributions of Data, Spatial Discretization, Discontinuities and Delineation. It is possible to interactively: change of the number of classes, split a class by fixing a new limit, merge two classes, move class limits, select data, focus on selection, cross histogram, assign colors, transform data, have any M-to-N possible dynamic links and views.

2. OBJECTIVES

The global objective of the Distogram (Josselin, 2001) was to build pertinent statistical and spatial partitioning using dynamic links between maps and plots in a Exploratory Spatial Data Analysis (Andrienko & Andrienko, 2005).

The objective of this poster is to minimize the variance in the classes and to maximize the variance between classes, using different methods :

- Jenks method
- F test of variance
- RSQ (R-squared): coefficient of determination
- Local Indices of Spatial Autocorrelation

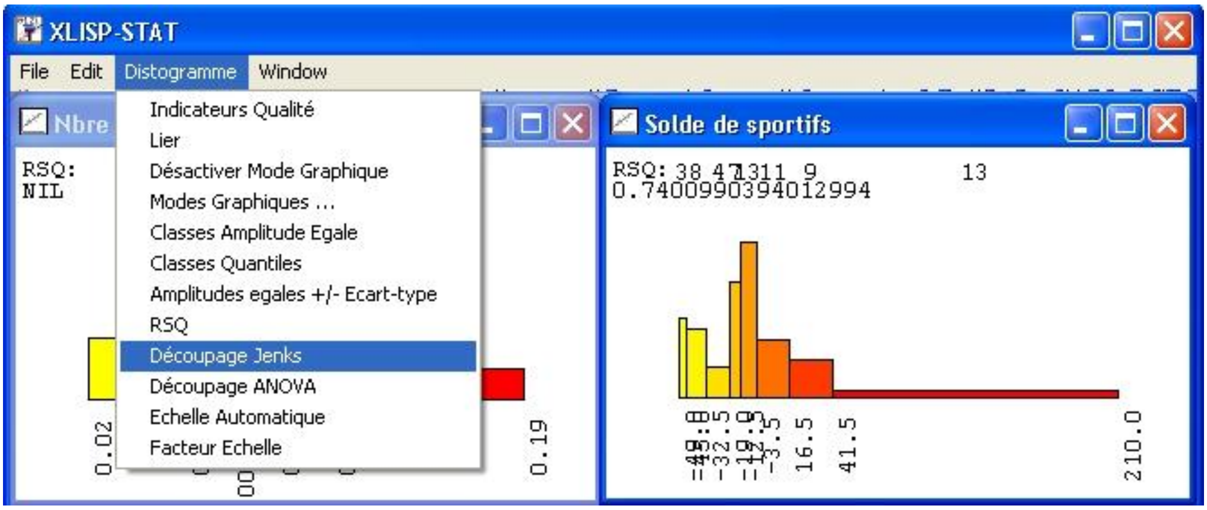

Figure 1: The Distogram: an histogram including different discretization methods

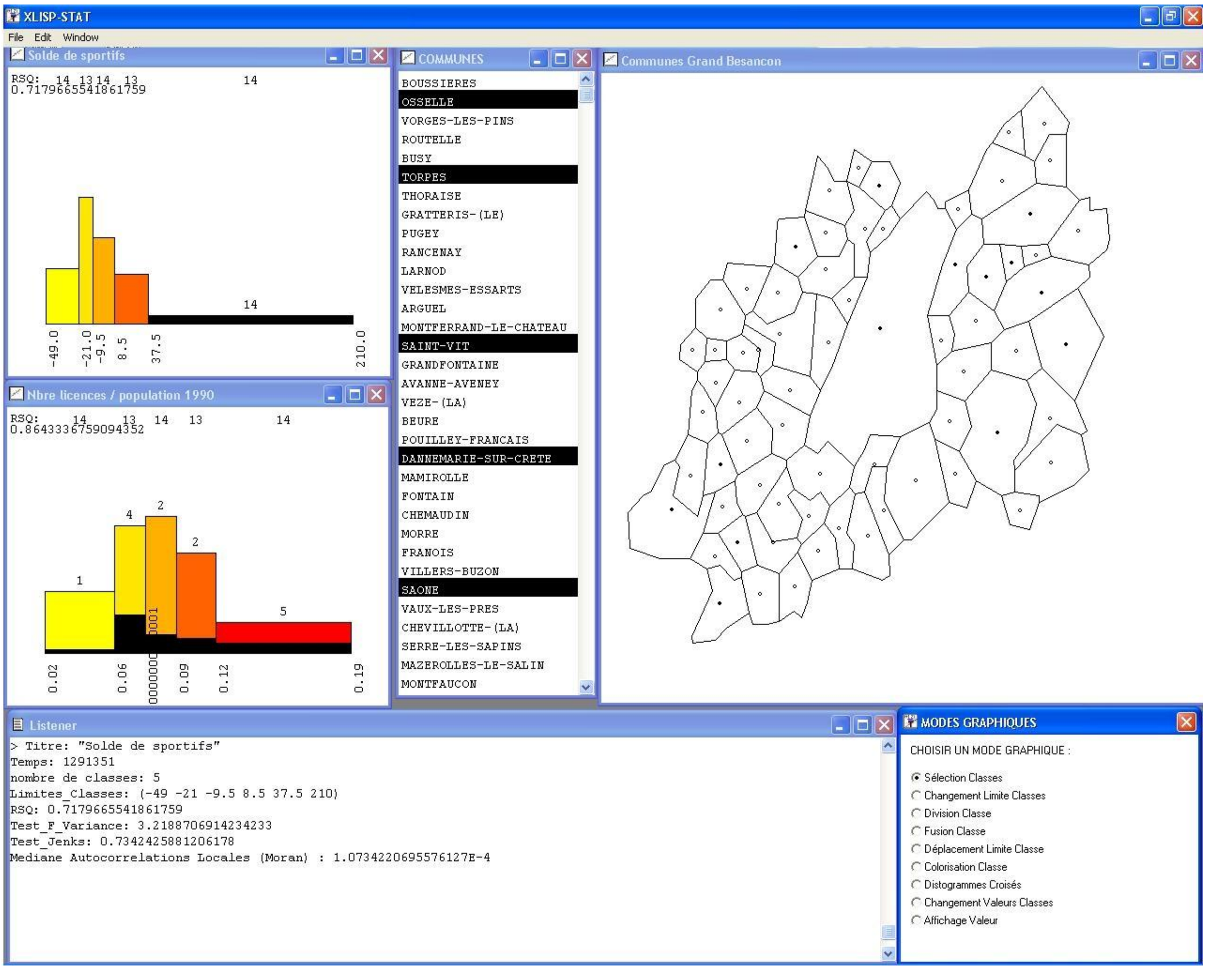

Figure 2: Maps and graphs are dynamically linked and interact

3. VARIANCE ANALYSIS

We recently improved the discretizing function by including current variance estimations. The goal is to maximize inter-group variance and minimize intra-group variance. What is original in our approach is that the variance criteria are permanently calculated (Jenks indice, 1971, F-Variance), each time the user changes the distribution. The R-Square (RSQ: intra-class variance / total variance) is processed in a separate graph (figure 3) which plots the RSQ (Y) and the number of possible classes (X). This graph can be used to find the best compromise between the number of classes and the variance analysis results. It interacts with the statistical distribution through a many-to-many dynamic link: it is possible to click on one of its points and to activate the discretization changes in the appropriate distribution (with the relevant number of classes and the corresponding RSQ). It helps a lot the user to make reliable discretization and spatial partitioning due to interactivity.

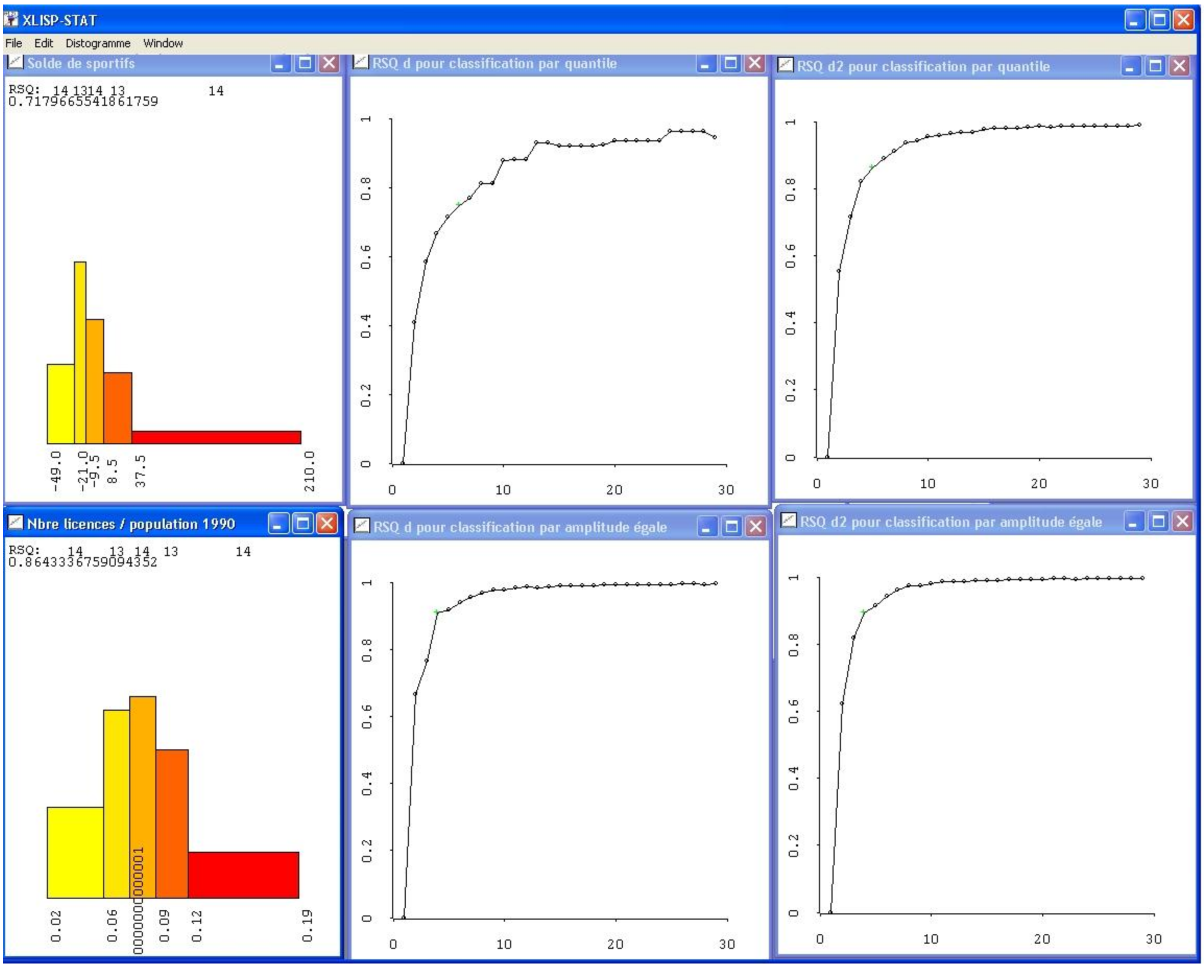

Figure 3: RSQuare values depending on the number of classes and the type of discretization method

The figure 3 depicts a case of two variables with their respective distograms d and d2 about a few French municipalities attractivity and penetration rate for sport, Besançon, North-East of France). By clicking on one of the points, the corresponding distribution will be depicted according to the number of classes and the type of discretization. Also the map colorization is updated.

CONCLUSION

The Distogram allows to jointly analyse statistical and spatial distributions. It is possible to build discriminant discretization that maximize spatial autocorrelation and minimize variance in the classes and the agregates.The variance analysis and the RSQ help in finding the best compromise between the number of classes and the quality of the discretization. Working in an Exploratory Spatial Data Analysis framework brings a lot to the users, because it allows to dynamically test different distribution discretizations with specific statistical parameters. This makes the user analatical process closer to the data and the phenomena they describe, in a systemic approach.

REFERENCES

- Andrienko Natalia and Gennady, 2005, Exploratory Analysis of Spatial and Temporal Data. A Systematic Approach, Springer-Verlag.

- Anselin Luc, 1995, Local indicators of spatial association – LISA, Geographical Analysis, 27, pp. 93-115.

- Cauvin Colette, Francisco Escobar, Aziz Serradj, 2007, Cartographie thématique 2 des transformations incontournables, Traité Information Géographique et Aménagement du Terroitoire, Hermès lavoisier.

- Jenks George F., 1967, The Data Model Concept in Statistical Mapping, International Yearbook of Cartography 7: 186-190.

- Josselin Didier, 2003, L'analyse des discontinuités spatiales avec le Distogramme, Contexte théorique, présentation, évaluation, vol. 13, n°1/2003, Revue Internationale de Géomatique, 29-48, Hermès, Lavoisier, Paris.

- Josselin Didier, 2005, Interactive Geographical Information System using LISPSTAT : prototypes and applications. Journal of Statistical Software. Volume 13, Issue 6, 20 pages, http://www.jstatsoft.org/v13/i06

- Tierney Luke, 1990, Lisp-Stat, an object oriented environment for statistical computing and dynamic graphics, Wiley-Interscience Publication, John Wiley and Sons.