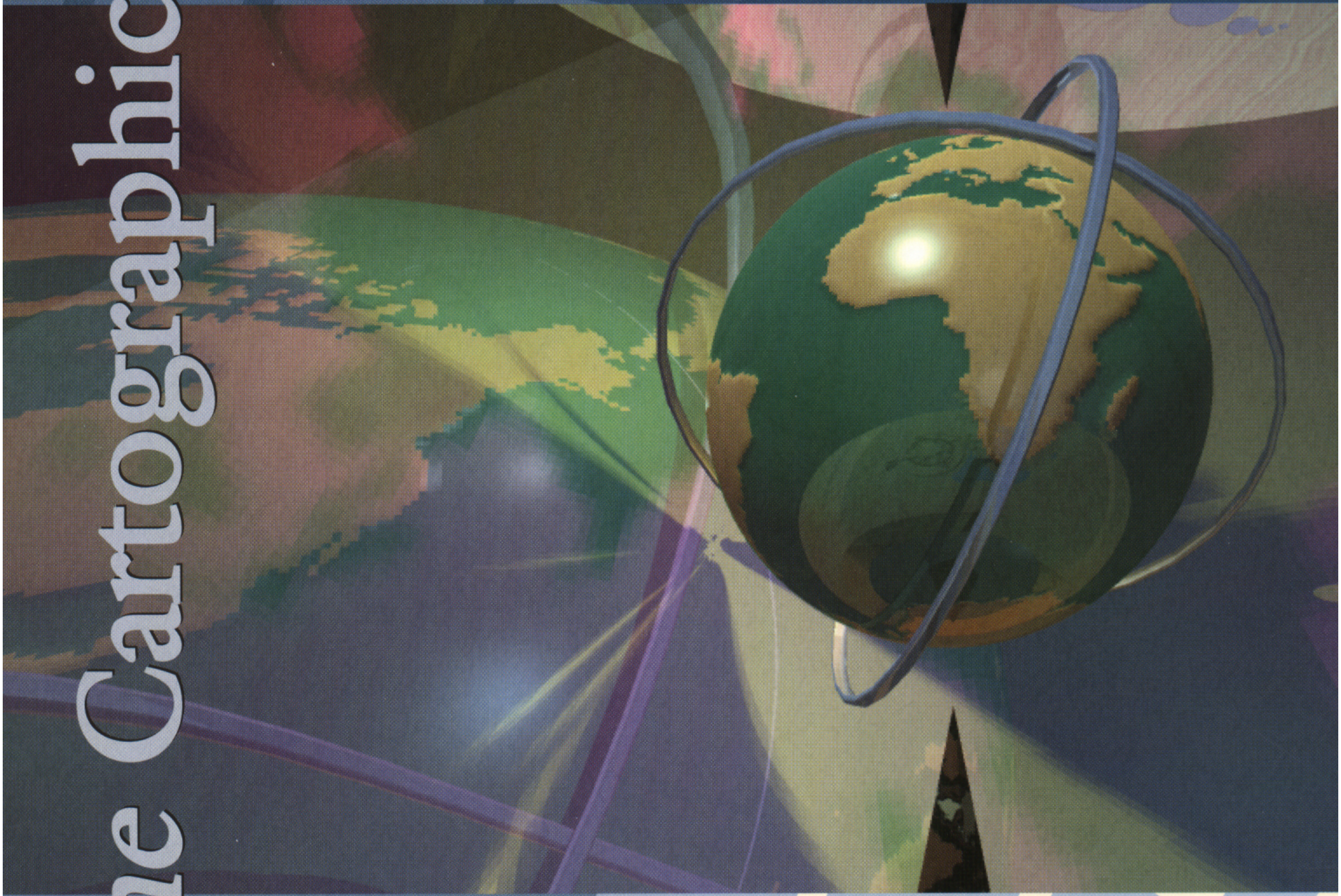
The Cartographic Journal

ISSN 0008-7041

The British Cartographic Society

Registered Charity No. 240034

The World of Mapping



THE CARTOGRAPHIC JOURNAL

Volume 37, Number 1 **June 2000**

Contents

The Cartographic Journal *Vol. 37 No. 1 pp. 51–64 June 2000*

Two Software Packages for Mapping Geographically Aggregated Data on the Apple Power Macintosh

Philippe Waniez

Institut Français pour le Développement, IRD, École Normale Supérieure, Paris, France.
Email. philgeo@club-internet.fr
Web page. http://perso.club-internet.fr/philgeo

In the world of the physical sciences geographers often develop specialized software packages for their scientific work that are directly applicable to their own domain and related domains of research. Philcarto and Philexplo for the Apple Power Macintosh are two such examples of this kind of software package. They are designed specifically to map spatially aggregated data for censuses (population and flats, agriculture, etc.) or specialized surveys (education, health, etc.) in which the 'national space' is divided up into spatial units (regions, departments, municipalities, etc.). Philcarto is an 'interactive cartomatic' program. Using two or three mouses clicks the user redefines the characteristics of the map which he wishes to obtain (the mode of representation, type of 'discretization', colours scales, symbols sizes, final annotations of the map, etc.). The value of this software does not lie with its methodological novelty, but with the ease of its use and the immediacy of displaying a map often comprising several thousand spatial units. Philexplo, by contrast, is a more innovative 'exploratory cartographic' program, the origins of which are based upon difficulties experienced in analyzing spatially aggregated data using 'traditional' statistical methods. The program uses linked windows to display various plots of the data, including a map window. Using various tools, all the displays can be dynamically explored: the plots can be transformed according to various actions of the user (selection of spatial units, 'colourization', etc.) and these transformations are displayed in all the other windows, in all the other plots.

GEOGRAPHICAL RESEARCH AND SOFTWARE DEVELOPMENT

Commercial software: function versus the end-users

With the increasing commercialization of GIS (Geographical Information Systems) software and the continuing drive for profits, the geographer/cartographer/researcher consumer now seems to have to accept, one way or another, what is on 'offer' in the marketplace, with limited options to negotiate the price and choice of computer type and operating system. Alternatives for these particular end-users often lies either with resorting to high levels of investment for the purchase of specialized computer equipment and a lot of intellectual effort to read thick manual(s), or to resort to the use of software packages which are often poorly adapted to their requirements and carry far too much functionality, the result of the 'encyclopaedic methodology' of the major software producers.

Towards the software requirements of geographers

Fortunately, the world of academic scientific research is still not completely dependent on financial interests (in France, at least). The freedom which researchers still enjoy allows them to develop original software packages, working in a computer environment where they are a 'member of a minority party' (e.g. Apple Macintosh), well adapted to specific practices in the scientific world. Naturally, however, it is difficult for a small group of researchers to set up integrated systems designed to satisfy large number of users. Nevertheless, researchers can develop specialized applications of little commercial interest, but with high scientific interest, with 'applicability' in their own domain and in the related domains of research.

Are geographers qualified to develop software?

In the world of the physical sciences it is common to come across researchers conceptualizing and developing software packages necessary for their research. But, in the field of the human sciences (at least in France) such an activity now seems more rare. This difference can be explained by the literary character of a major part of French geography. But, in spite of the real existence of some French poles of excellence formerly regrouped partially under the banner of the Public Interest Group Network for Study of Changes in Locations and the Spatial Units (GIP RECLUS), the development of new software packages adapted to the newer computers remains relatively marginal and rarely goes beyond the stage of the 'prototype', use of which remains strictly with the competence of the programmer! Unfortunately, when such an effort turns out to be impossible, one often observes a net recession in the use of techniques that have formerly been widely used, especially when they have

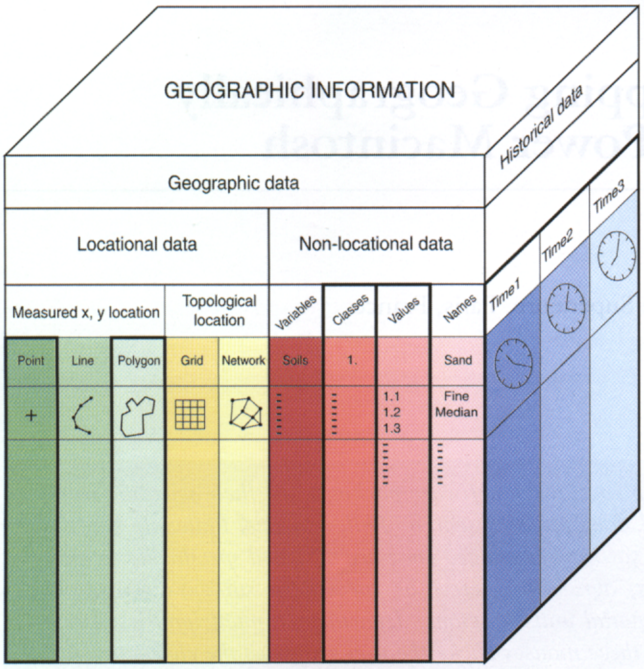


Figure 1. Three conceptual components of geographic based information system (after J. Dangermond)

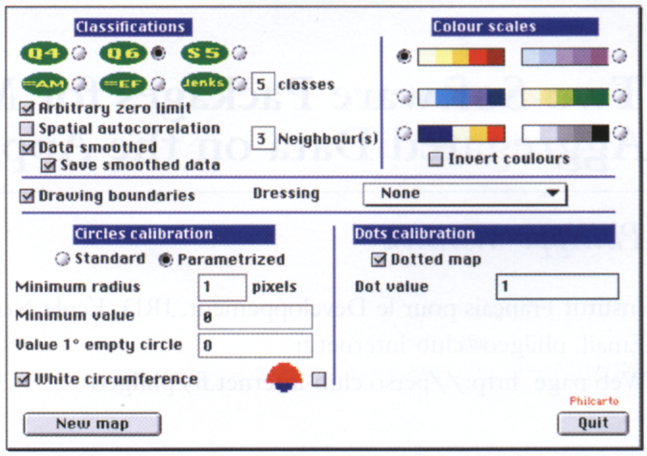


Figure 3. Philcarto: the 'Options' window

largely disappeared from the main systems available (for example, triangular diagrams).

Ten years experience in developing cartographic software

About ten years ago, for reasons of personal dissatisfaction and because the author wished to keep complete control of the computer programming involved in his geographic research it led to him developing several software packages for the thematic cartography of aggregated statistical data on the basis of administrative spatial units. The first of them, Cabral 1500, developed within the GIP RECLUS at the beginning of 1990's, was very successful with some one thousand copies being used in Europe and in South America. This permitted a fertile dialogue in the context of French co-operation with the countries of the South (mainly Brazil, Bolivia, Peru and the Equator, but also in Senegal and Algeria). However, because of the rapid rate of development of powerful microcomputers, the creation of two new software packages subsequently turned out to be

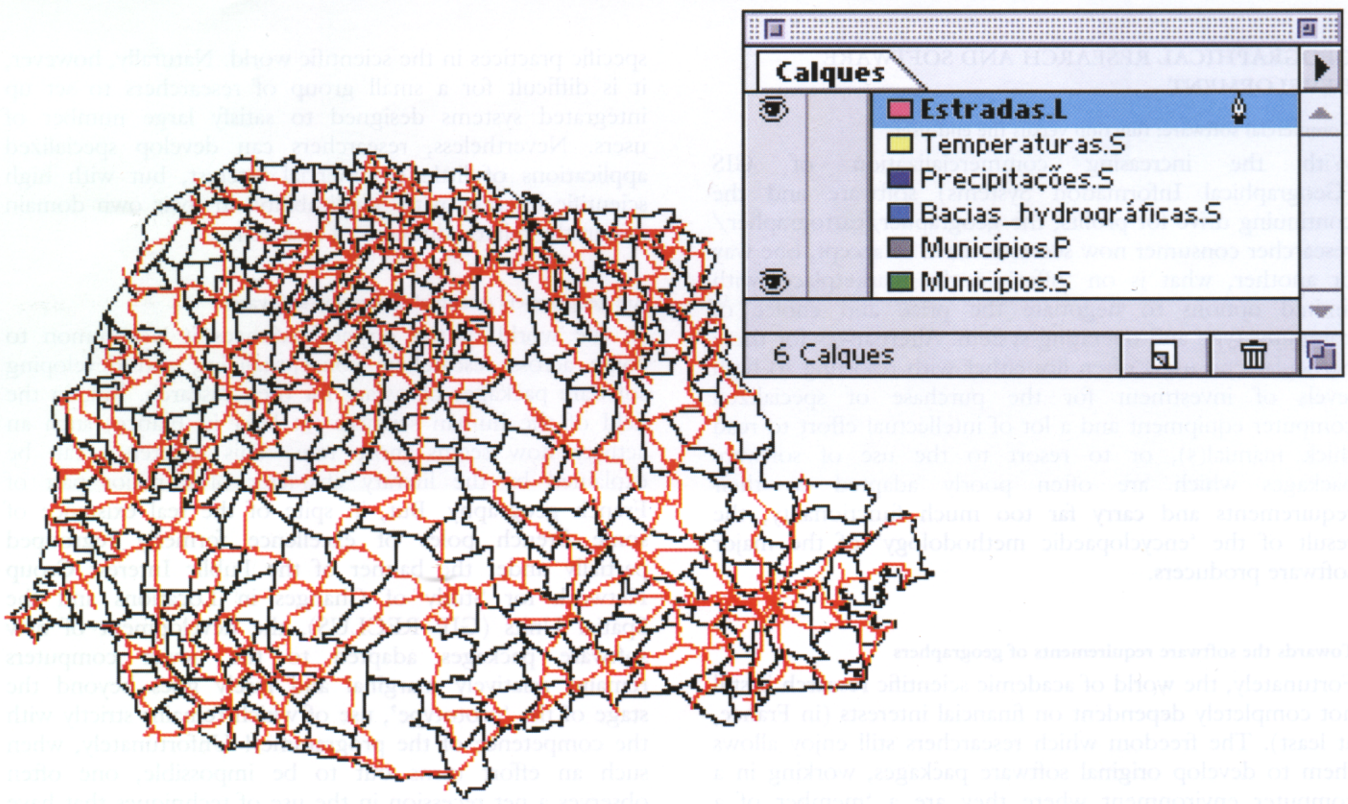


Figure 2. The base map of the Paraná State (Brazil) showing the communes and roads on different layers in Adobe Illustrator

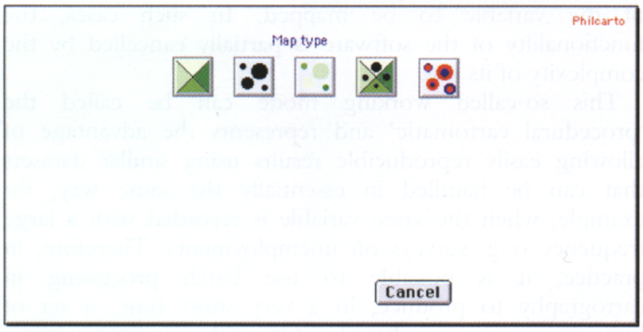


Figure 4. Philcarto: the 'Map' window

necessary: Philcarto and Philexplo for Apple Macintosh and Power Macintosh (downloadable on Internet — with several utilities and numerous base maps — at the following address: http://perso.club-internet.fr/philgeo).

THE CARTOGRAPHY OF GEOGRAPHICALLY AGREGATED DATA

Spatial information matrix

Before describing these two software packages it is necessary to clarify which types of data they can handle. This data corresponds mainly to that published, all over the world, by the national institutes of statistics undertaking various censuses (e.g. population, housing, agriculture, etc.) or specialized surveys (education, health, etc.). In these data files, the national area is often divided into regions, departments and municipalities (with a number of levels and names which can differ from one country to another one). Every element of this division of space is refered to as a 'spatial unit'. For every investigated subject, the user acquires a matrix (**T**) where the rows are the spatial units; in the columns appear the categories of objects (the classes of surface or farms, for example). The cell formed by the

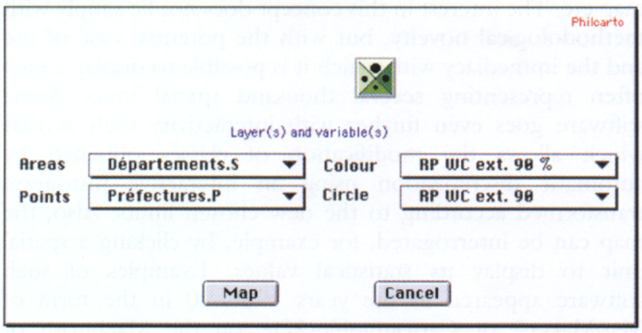


Figure 5. Philcarto: the 'Map' window. The local menus are used to choose the basic elements of the map

crossing of a row and a column (t_{ij}) contains the number of objects characterized with the category **j**, found in the spatial unit **i**. These data are spatially aggregated so the user never has access to the individual information (because censuses and surveys are undertaken in a statutory frame which has to guarantee the confidentiality of answers).

Data, maps and spatial variability

In relation to the scale of cartographic representation, the spatial units are symbolized either by surfaces or by points, according to the legibility of the final document: for example, a city can be represented by a surface at the scale 1:50 000; but because the spatial limits are not visible at the scale 1:10 000 000 it takes the shape of a point.

To show the spatial variability of data, two types of representation are available. Proportional symbols placed inside every surface or at every point representing spatial unity are used for a quantity (e.g. the number of farms). It is the variability of the surface of these symbols that allows the eye of the reader to perceive the spatial variation of

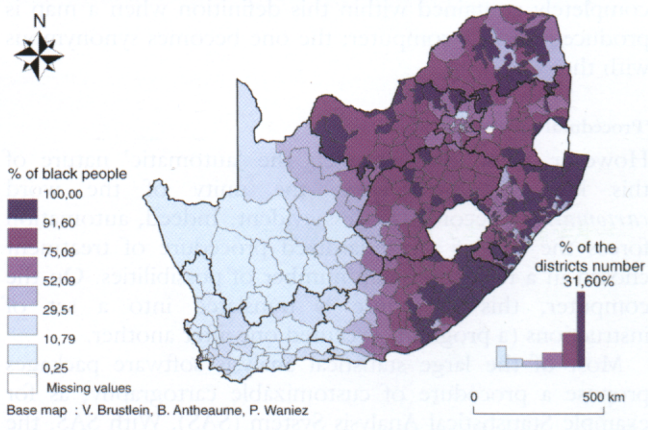


Figure 6. South Africa (districts). Percentage of black people in population, 1991

these quantities. These symbols can be geometric or non-geometric. In general one prefers the use of simple geometric symbols (circle or square) so as not to disrupt the visual perception of the reader by specific details contained in the selected symbol.

When the values are relative (e.g. ratios, percentages, indexes), calculated from the initial (**T**) matrix, the spatial variation is expressed by the intensity of a range in colours or in grey levels. As the eye is not capable of perceiving thousands of colours, one rarely chooses ranges exceeding ten different tints. The transformation of a statistically continuous series (for example, percentages varying between nought and one hundred per cent) to a number limited by colours (five or six), is called discretization. It is possible to do this by using different techniques of classification although no one method is ideal because they all cause quite a loss of variance.

Geographical information and geographically agregated data

The domain of software application presented here is limited to certain elements of geographical information as defined by Dangermond (1990) in Figure 1; locational data in the form of 'points' and 'polygons'; non-locational data as 'classes' and 'values'. Because space is considered as invariant in this software, it is not possible to consider the temporal dimension for the locational data (sometimes an insoluble problem when the spatial division or unit is transformed over time, as it is in the case of a young country such as Brazil). On the other hand, the mapping of rate or indexes of variation across time presents no difficulty when the spatial division does not change or when it is possible to elaborate historical data in a specific division of space; for example by calculating statistical frequencies again from individual data, assuming that they still exist and are accessible.

CARTOMATIC VERSUS EXPLORATORY CARTOGRAPHY

The word *cartomatique*, constructed by Brunet (1987) from the words 'cartography' and 'automatic' concerns the use of mathematical and graphical procedures designed to translate, on a base map, the spatial variation of a statistical variable. The cartography of spatially aggregated data is completely contained within this definition when a map is produced using a computer; the one becomes synonymous with the other one.

'Procédural cartomatic'

However, when one considers the 'automatic' nature of this kind of cartography, the unity of the word *cartomatique* becomes far less evident. Indeed, automation forces the use of a standardized procedure of treatment chosen in a relatively small number of possibilities. On the computer, this procedure is translated into a set of instructions (a program) executed one after another.

Most of the large statistical analysis software packages propose a procedure of customizable cartography, as for example Statististical Analysis System (SAS). With SAS, the creation of a map using coloured surfaces (a choropleth map) is extremely simple:

Proc GMAP data=myStat map=myMap;
ID theCode;
Choro myVar / levels=5;

This set of insructions can be read in the following way:

- To create a map with statistical data contained in the file **myStat** and the base map recorded in the file **myMap**;
- The identification code of the spatial units is **theCode**, and it is present in both the file **MyStat** and the file **myMap**;
- The choropleth map is represented by the variable **myVar** (which must be present in the file **myStat**); the statistical range of **myVar** has to be divided in five classes (from 0 to 20, 20 to 40, 80 to 100).

Naturally, such a procedure can be preceded by more elaborate statistical results, recorded in myVar, to be mapped and to produce a display of the corresponding map on the selected output display unit (e.g. screen, printer etc.). If the map is not satisfactory, the program must be corrected, then executed again, until one obtains the desired result. The user-interfaces used by the large statistical analysis software packages usually facilitates the input of the program text using shortcuts or menus. Modification of the program may be limited to changing some parameters but may also require a completely new conception and a complete rewriting of the source code, requiring the cartographer to have programming capability. When the cartographer is not able to master the programming language built-in to the statistical software, a programmer can help to prepare screens allowing the input of the required parameters to produce a better map. Such screens can be reduced even to the input of just the name of the variable to be mapped. In such cases, the functionalilty of the software is partially cancelled by the complexity of its use.

This so-called working mode can be called the 'procedural cartomatic' and represents the advantage of allowing easily reproducible results using similar datasets that can be handled in essentially the same way; for example, when the same variable is recorded with a large frequency (e.g. surveys on unemployment). Therefore, in practice, it is possible to use batch processing in cartography to produce, in a very short time, a set of homogeneous maps based upon either existing maps. However, beyond the simplicity of this, this process brings numerous uncertainties about the quality of the maps. The most important is the 'blind' character: anybody who takes time to examine the statistical distributions will wonder about the pertinence of the chosen discretization mode, of the selected number of classes, of the choice of colours scales, etc. In some cases, nobody knows how the maps have been derived (i.e. what is contained in the procedure) the programmer having changed jobs!

Although it is still widespread, the 'procedural cartomatic' is considered to be like a dinosaur proceeding from the prehistory of microcomputing, characterized by the omnipotence of the programmers and the quasi/no-interaction environment that exists between the software program and the user. Although effectively condemned, this method of treatment survives by the force of usage and the difficulty announcing new methods.

'Interactive cartomatic'

The 'interactive cartomatic' software adapts the user-interface by facilitating the submission of successive cartographic treatments. With two or three mouse clicks, the user redefines the characteristics of the map which he wishes to obtain: mode of representation, type of discretization, colours scales, symbols sizes, and the final annotation of the map etc. The interest in this concept does not lie simply with methodological novelty, but with the potential ease of use and the immediacy with which it is possible to display a map often representing several thousand spatial units. Some software goes even further with interactivity such as that which allows the modification of classes obtained by automatic discretization using an interactive histogram transformed according to the new chosen limits. Also, the map can be interrogated, for example, by clicking a spatial unit to display its statistical values. Examples of such software appeared in the years 1985–90 in the form of MapMaker® or Cartographie-2D® on the Macintosh or Atlas Pro® and MapInfo® on PCs, with different degrees of success, some of which has disappeared over time, whilst others examples have been transformed into GIS through the integration of database query functions.

Often interactive cartomatic software can be irritating to use because of the need for a large number of mouses clicks and because of the long and somewhat vague navigation of the interface before being able to obtain a map. Such a situation results from the commercial necessity of always having to offer more options. Philcarto, described here, can be categorized as interactive cartomatic software but without this disadvantage. Developed in an academic scientific environment, Philcarto is convenient for both students and researchers in the human sciences who need to illustrate their text documents with maps by creating them in the same Macintosh environment.

'Exploratory' cartography

Exploratory cartographic data analysis must be considered from an opposing point of view to procedral approaches. This statement is justified by considering the difficulty encountered when trying to analyze certain types of data using traditional statistical methods, a difficulty which is not specific to geographic data because it is characteristic of all the human sciences and certain domains of biology.

'It is difficult to analyze geographic data from traditional statistical models for reasons liking to spatial interdependences, and to complexity of spatial structures' [Unwin, 1994].

Attempts by statisticians to adapt classic statistical methods (regression, factorial analysis, etc.) to spatial data have often led to really complex solutions and ones that were difficult to put into practice in the real world. A number of researchers, from both biology and the social sciences, have to use more intuitive, 'less hard' , and less binding methods. There is already around thirty years of experience. J.W. Tukey, the father of the Exploratory Data Analysis (EDA) proposed a set of techniques for analysis to encourage the researcher to directly interrogate data which was not only based on the statistical parameters which summarize them (average, variance, etc.). He insisted on the necessity of reacting to the appearance of exceptional or missing values, to breaks in distributions, to groupings, etc. (Tukey, 1977). Most of the techniques proposed by Tukey are based on graphical representations allowing one to observe the data from different perspectives so as to trap significant information, allowing one to improve the knowledge of the object studied. Of these rules arises a real 'school' of statistical data analysis based on the observation of graphic representations. The book **Visualizing Data** (Cleveland, 1993) represents a pedagogical summary indispensable to anyone who wants use EDA.

At the time of Tukey, EDA remained a difficult approach to utilize. However, due to graphic interfaces for operating systems, the EDA found another dimension: interactivity. Today, the user can not only observe the different configurations of data on graphic displays, but he can also act on them, by selecting them, colouring them in, masking them, etc. And, it is not a coincidence that EDA's first public domain software appeared on the Apple Macintosh Computer. In a book published in 1991, the author drew up a comparative study of the various EDA software sold on the market for Macintosh (Waniez, 1991). It was also an appropriate occasion to present the principles necessary for a good investigation. But, unfortunately on this occasion none of the software available had any cartography capability, very regrettable for anyone who wanted to analyze geographical data. To find good examples of what should be the right software for the Exploratory Cartography of Data (ECD) it was necessary to turn to prototypes developed by researchers such as Polygon Explorer, Regard, Manet on Macintosh or Exploremap on PC.

Besides a certain variety of statistical graphs (the box and whiskers plot, histograms, bivariate graphs, rotating graphs) associated with different functions of selection, these softwares have a Map window which answers, as do the other graphic windows, to the actions of the user. Therefore the ECD allows one to analyze subtle spatial configurations. It therefore becomes possible to link a dynamic map and graphs for the researcher to explore the data. Philexplo was inspired by A. Unwin in Trinity College of Dublin. It differs, however, in several ways. First of all, several types of graphs are different; for example, Philexplo contains a dynamic triangular diagram. Secondly, an there is an important difference in Philexplo's completely operational character; a non-specialist researcher without expertise in the techniques of geographic information can easily and quickly analyze the data based on some thousand spatial units. Philexplo has already served in true geographic research (not only methodological); for example to analyze the changes in Brazilian agriculture (David, 1997; Waniez, 1999).

COMMON FEATURES OF PHILCARTO AND PHILEXPLO

Mastering the development and the programming of these two software applications confers on them a number of common characteristics from the point of view of their conception. It is due to the fact that they were conceived and developed at the same time.

Strong links with the geographical research and academic studies

Due to the strong relation within the geographer/ cartographer network, the development of Philcarto and Philexplo has benefited from the experience of professional users. What matters here, is less the technique than the contribution to the field of geographical analysis. Philcarto and Philexplo follow closely the rules of graphic semiology defined by (Bertin, 1977). Options allow one to modify the characteristics of the graphic elements. But, the freedom for the user is limited by the rules (it is the case, for example, that the colours scales are not directly modifiable in the program). To say that it is impossible to produce bad maps using Philcarto or Philexplo would be an exaggeration, but the software was conceived so as to grant to the user a 'restricted liberty'; contrary to the sales argument for commercial software that says 'let's make everything and anything'! What makes Philcarto and Philexplo simple software useful to both researchers and students does not lie in the technique but in the obtaining good cartographic documents quickly.

A STRONG SOLIDARITY WITH ADOBE ILLUSTRATOR

The base maps used by Philcarto and Philexplo are recorded in postscript format readable by Adobe Illustrator.

Why is Adobe Illustrator preferred?

Adobe Illustrator, originally conceived for the Apple Macintosh (now version 8), provided the means to integrate the most innovative concepts of vector computer graphics: postscript language, layers editable independently

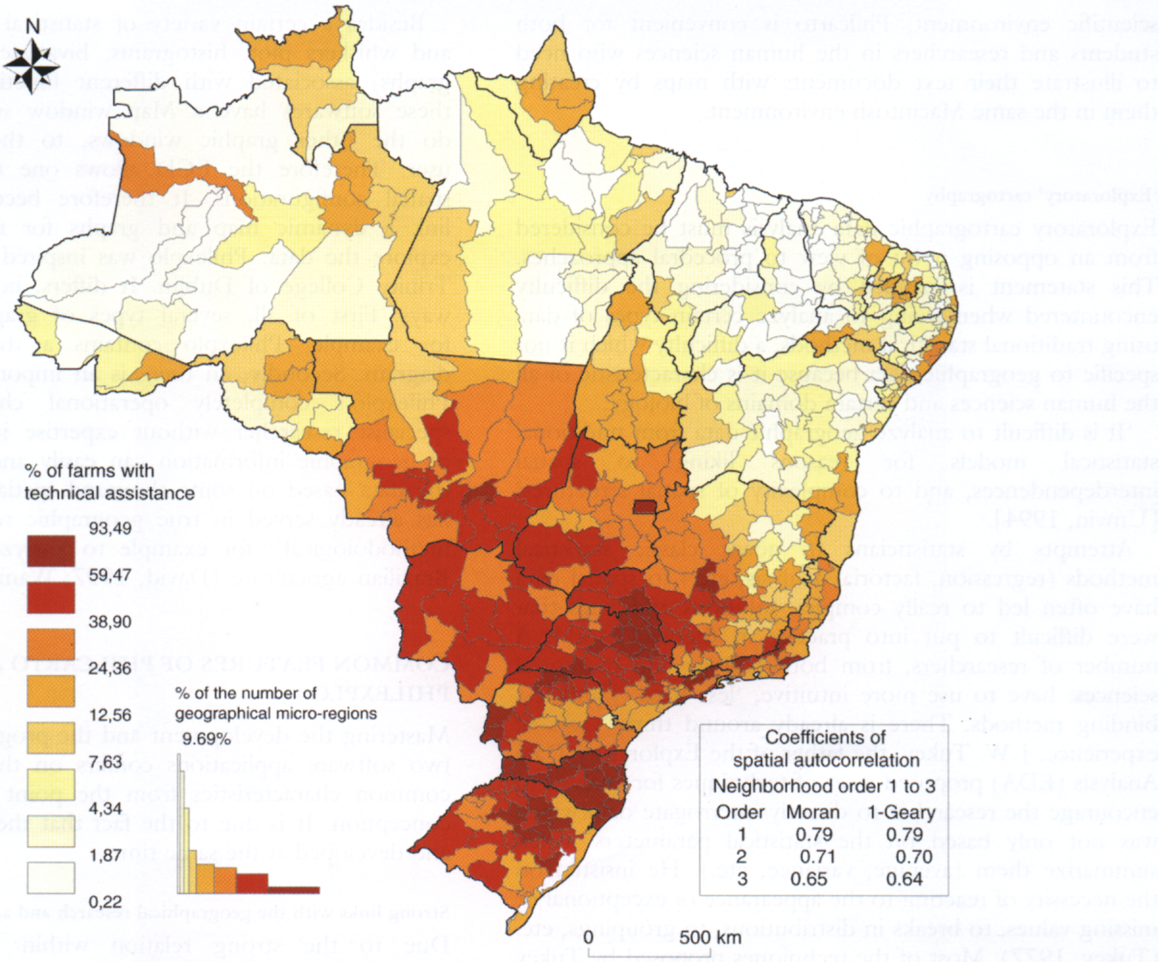


Figure 7. Brazil (geographical micro-regions). Percentage of farms which use technical assistance, 1996

of each other, operations on vector objects (grouping, intersecting, etc.), universal format PDF etc. Every new version of Illustrator provided new functions without a cost to the final user. Illustrator remains at the foreground of software for vector computer graphics on the Apple Macintosh, and since version 7 on the PC. Due to its widespread distribution, many graphic designers are able to create base maps from scanned documents, georeferenced or not, depending on the initial document. A utility supplied with Philcarto allows one to convert files of numerical coordinates (in the format Identifier, X, Y) in postscript format, ready for import into Illustrator: and every imported file becomes a layer.

The map of the Paraná's municipalities (municípios) in Figure 2 shows how the base map is recorded by Illustrator. This file contains six different layers. The first layer, Municípios.S, represents the limits of the municipalities ; the suffix '.S' states that the spatial units appearing on this layer are surfaces. The second layer, Municípios.P, contains the location of the administrative centers of the municipalities which are represented by points (.P). The other layers contain additional maps as the limits of the hydrographic fields, the surfaces among isohyets, the surfaces between them isotherms, and roads. In this figure, only layers Municípios.S and Estradas. L are visible. On every layer, the identification of the spatial units (points, lines and polygons) must be recorded in a note which can be edited in a specific Illustrator window.

The integration of Philcarto with Illustrator goes beyond just the input base map. Output maps are recorded in postscript format with the same size as the input map. If the base map is ready to be published, maps produced by Philcarto are ready too. Maps created with Philcarto can be easily be added to with additional cartographic layers (rivers, roads, scale, etc.) using a simple 'copy/paste' function.

The integration of Philexplo with Illustrator is different for reasons of the speed of display, because in Philexplo, the cartographic window has to be updated after any action of the user. The Illustrator base map file is transformed using a conversion utility which writes a binary file rapidly readable by Philexplo. For output, the window map is recorded in the PICT format, native on Macintosh, which can be imported by Illustrator, but with different geometrical characteristics from the base map (contrary to Philcarto). The preservation of the geometrical characteristics of the base map remains possible by the importing into Philcarto the attributes of the spatial units calculated by Philexplo (completely transparent for the user).

FOR POWER MACINTOSH ONLY, IN LIGHT CONFIGURATION

Philcarto and Philexplo works both on Macintosh and Power Macintosh, under MAC OS 7.5 to 9.0. System

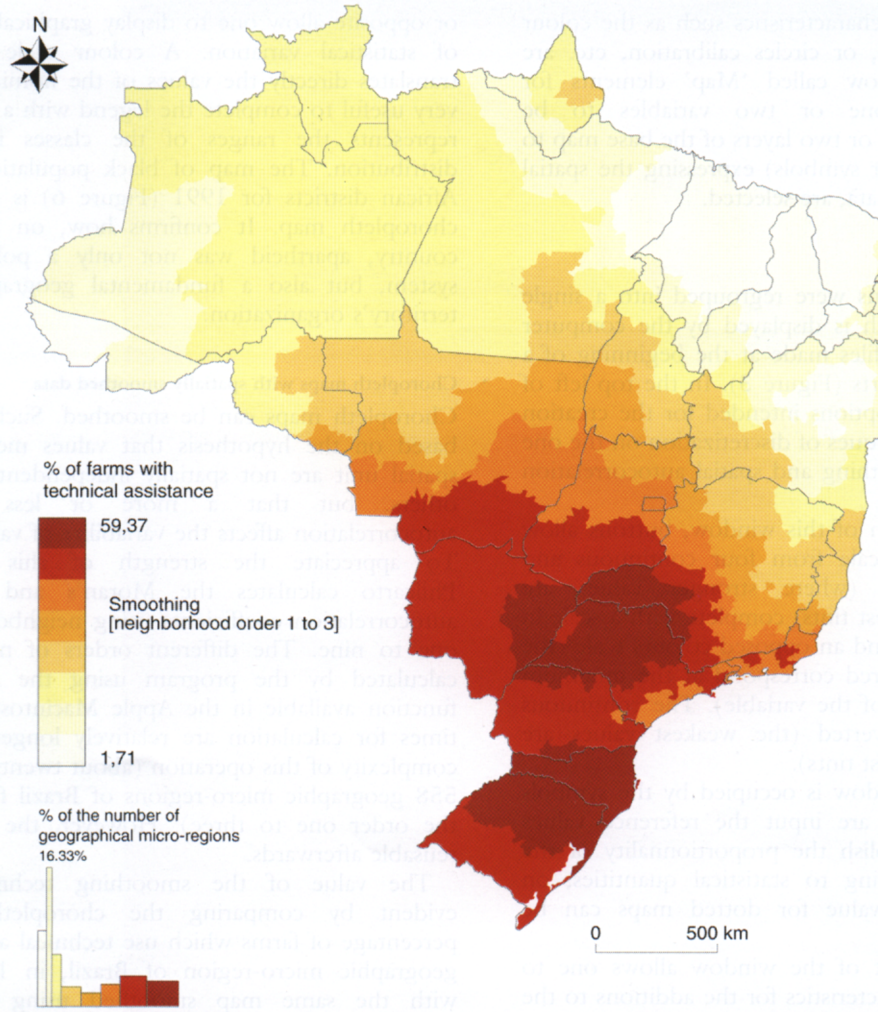


Figure 8. Brazil (geographical micro-regions). Percentage of farms which use technical assistance, 1996. Smoothed data using the neighbourhood orders 1–3

8 is recommended as well as Power Macintosh. Memory requirements necessary for Philcarto and Philexplo are about 20 megabytes, allowing a comfortable use on iMac or iBook, provided with 64 megabytes of RAM.

Why develop new cartographic software for the Apple Macintosh?

The crisis suffered by the Apple Computer company during 1997–1998 ended with the launch on the market of the Power Macintosh G3, then the G4, and the iMac and iBook provided with Mac OS8, a newer version of the Apple's innovative operating system. Regrettably, the designers and the editors of their software, shaken by the repeated 'alarm bells' concerning the possible disappearance of the company, were led into a reorientation of their strategies of development in favour of other operating systems. It led them to focus on other computer platforms, and sometimes even to abandon MAC OS versions of their products. The software offerings for statistical cartography under the MAC OS became very impoverished. However, the Apple computer platform continues to occupy an important place in the field of desktop publishing and, to a lesser extent, in research laboratories in the social sciences. Important scientific and technical efforts have accompanied the creation of Philcarto and Philexplo to take advantage of the the extended capabilities of the new computers produced by Apple, and to maintain the ease of use for cartography.

PHILCARTO, AN 'INTERACTIVE PROCEDURAL' CARTOMATIC SOFTWARE FOR MACINTOSH

There is a large variety of methods for the cartographic representation of economic and social data. Philcarto proposes only a limited choice of such methods. Some are very common as, for example, coloured surfaces or proportional circles; the others are less widely used such as smoothing or links. The limited options provided by Philcarto correspond to the main requirements of the social sciences, regional development and town planning.

Use of Philcarto requires opening, at the beginning of a session, an Illustrator file containing the base map, and a statistical datafile in the format of 'text with tabulations'. The use of only two window allows the end-user to completely define the map to be displayed. By means of the window called 'Options', characteristics such as the colour scale, discretization mode, or circles calibration, etc. are chosen. Using the window called 'Map' elements for composing the map: one or two variables to be represented, as well as one or two layers of the basc map to carry attributes (colours or symbols) expressing the spatial variation of the statistical data, are selected.

The 'Options' window

All of the Philcarto options were regrouped into a single window, the first of which is displayed by the computer after the selection of the files made at the beginning of a session. It includes five parts (Figure 3). In the top left of the window appears the options intended for the creation of choropleth maps: techniques of discretization on the one hand, parameters of smoothing and spatial autocorrelation index on the other.

In the top right portion of this window, buttons allow one to select a colour scale from four continuous and increasing colour scales (where stronger values are represented with the darkest tints) complete with a scale in increasing levels of grey, and an opposed colours scale (the darkest tints of blue and red correspond at the minimum and the maximum value of the variable). The continuous colours scale can be inverted (the weakest values are represented with the darkest tints).

The bottom of the window is occupied by the symbols parameters. On the left, are input the reference values which allow one to establish the proportionnality of the surfaces of circles according to statistical quantities; on the right the reference value for dotted maps can be edited.

Finally, the central part of the window allows one to select two important characteristics for the additions to the map: to draw the limits of the spatial units, and to stack onto the map a supplementary layer (other administrative divisions or other map elements as rivers, roads, etc.).

The 'Map' window and the different map types

When the end-user has clicked the button 'map' of the 'Options' window then it appears. The contents, extremely simple, iconize the various methods of cartographic representation allowed by Philcarto (Figure 4). A click on one of these icons displays local menus intended to allow one to select the constituent elements of the map (Figure 5). For example, to draw a coloured proportional circles map, it is necessary to choose, using the left local menu, the layer .P which contains the localization of the circle's centers; on the right, two other local menus are intended to choose the variable to be discretized (coloured) and a second variable used to calculate the diameters of the circles. Here are the available map types listed from the left to the right of the 'Map' window.

Choropleth maps

Maps with coloured surfaces (or with grey levels) are scientifically known as choropleth maps and allow one to map numerical, continuous or discrete variables, with the exception of quantities or actual values. Various methods of discretization allow one to divide the data into classes continuous variable. Different colours scales, warm, cold or opposite allow one to display graphically the expression of statistical variation. A colour scale named 'parrot' translates directly the values of the nominal variable. It is very useful to complete the legend with a histogram which represents the ranges of the classes in the statistical distribution. The map of black population in the South African districts for 1991 (Figure 6) is an example of a choropleth map. It confirms how, on the scale of the country, apartheid was not only a political and social system, but also a fundamental geographic fact of the territory's organization.

Choropleth maps with spatially smoothed data

Choropleth maps can be smoothed. Such an operation is based on the hypothesis that values measured on every spatial unit are not spatially independent of some of the others, but that a more or less strong spatial autocorrelation affects the variability of values in the space. To appreciate the strength of this autocorrelation, Philcarto calculates the Moran's and Geary's spatial autocorrelation coefficients using neighbours of the order one to nine. The different orders of neighborhood are calculated by the program using the region dilatation function available in the Apple Macintosh's toolbox. The times for calculation are relatively longer because of the complexity of this operation (about twenty minutes for the 558 geographic micro-regions of Brazil for neighbours of the order one to three). However, the files created are reusable afterwards.

The value of the smoothing technique is visually evident by comparing the choropleth map of the percentage of farms which use technical assistance in every geographic micro-region of Brazil, in 1996 (Figure 7), with the same map smoothed using three successive neighbours (Figure 8) whose values are balanced (1 for the first neighbour, 1/2 for the second and 1/3 for the third). The depth of smoothing is justified by a relatively strong spatial autocorrelation in the successive orders of the neighborhood measured by the coefficients of Moran and Geary. The choropleth map suggests the existence of a spatial trend from two centres, the one situated in the borders of the States of São Paulo and Minas Gerais, the other one, more in the South, in the States of Santa Catarina and Rio Grande do Sul. The big variability of values (from 0.22 per cent to 93.49 per cent) provides an image of this trend parasited by a big wealth of detail. The smoothed map erases the rough (what has the effect of tightening values between 1.71 per cent and 59.37 per cent) and shows in a very suggestive image the gradient of modernization in the pioneers fronts in Mato Grosso, Goiás and Minas Gerais, and even beyond towards Bahia and Tocantins. The Nordeste region presents a marked depression, more particularly in the sertão, limited by more developed littoral fringe.

Proportional circle maps

The use of this type of representation is more delicate than it appears to be at first sight because a grading of circles is necessary in most cases, and several attempts are often necessary to obtain a satisfactory result. Part of the cartographic problem lies with the density of the spatial units leading to overlapping which it is necessary to manage in the best way possible: the smallest circles are drawn in front of the bigger circles and a white circumference allows one to distinguish the small circles localized on the big. This type of map is really only efficient when the number of spatial units is not too large (some hundreds at most), and when they are distributed in space in relatively homogeneous way. An example of this map type, very classic in population geography, is shown by the population of Vietnam in 1989 (Figure 9).

Coloured proportional circle maps

Proportional circle maps can be used to produce different variations without removing the limitations. Maps with coloured proportional circles allow one to simultaneously represent quantities (circles) and continuous or discreet numerical values (colours of circles). For example, the proportion of homes equipped with outside toilets in the French departments in 1991 (Figure 10) is a good indicator of the degree of outdatedness of the flats (but also certain cultural features). It is in the North of the country that this situation is the most obvious with more than 13 per cent of flats without internal toilets. The West does not appear to be very well off either, but certain tourist departments (Vendee, Loire-Atlantique, Finistère) are better equipped than the others. On the other hand, the central part of the Parisian Basin, the East and the Mediterranean coastline seem better equipped.

Proportional circles on choropleth maps

Proportional circle maps on choropleths maps allow one to also simultaneously represent quantities (circles) and continuous or discreet numerical values (colours of areas). This construction is even more sensitive to the dimension of the spatial units and to their distribution in space. It is necessary to avoid, on the one hand overlapping circles, and on the other hand ones that completely cover the coloured surfaces below them (it is also possible to draw only the circumferences of circles, but the reading of the map turns out to be more difficult). In most cases, the intervention of the graphic designer for the location of the circles on the map is needed. The figure showing hispanic population in the United States in 1996 (Figure 11) underlines the importance of California and Texas. But, it is in New Mexico, doubtless for historical reasons, that the proportion of hispanics is the highest, but with far fewer persons than in California or in Texas. The map shows other important locations, although secondary, in Florida (Cuban immigration), and in the industrial areas of the Illinois, the State of New York and New-Jersey.

Faced semicircles maps and rosettes maps

Faced semicircles maps and rosette maps (two superimposed circles) allow one to compare two quantities relating to two categories of the same population, either to the same category at two historical moments in time. The perception of the difference imposes on circles of a large diameter, that limits their number in a drastic way. The map of agricultural and industrial employment in the italian regions in 1998 (Figure 12) is a model example of this map type: the number of regions is small (nineteen) and

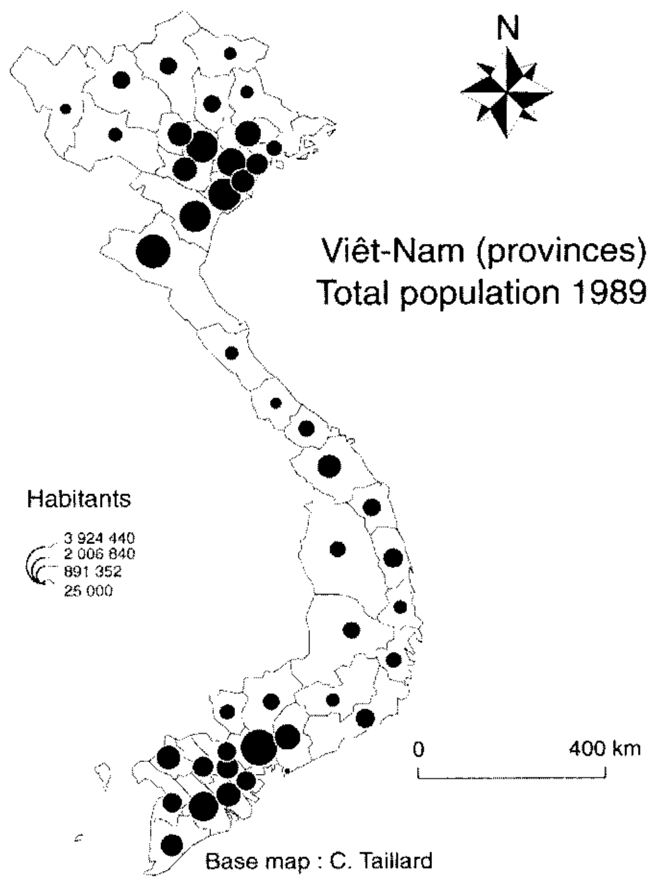


Figure 9. Vietnam (provinces). Total population, 1989

differences within the same region are strong, even though industrial employment everywhere is more important than agricultural. In Lombardia, in the North of the country, the ratio of industry/agriculture amounts to 14.5 while in Sicily, it reaches only 1.5. The map shows, without ambiguity, the decrease of this ratio from the North of the country to the South.

Dotted maps

When the button 'Dotted Map' of the Options window is selected, the icon corresponding to this type of map replaces that of the choropleth map in the left part of the Map window; a click on this icon starts the display of a dotted map. Dotted maps represent quantities, as a proportional circles map. Every point is located in a unpredictable way on the surface of the spatial unit to which it belongs. The density effect is often interesting, especially when the division of the base map is rather fine to visualize the location of the phenomenon. The use of this technique is more complex than the proportional circles because it is necessary to calibrate the elementary point to obtain a total number of points is not too big to avoid too voluminous graphic files in output. When there are too many points, the visual phenomenon of saturation occurs: certain spatial units are too small to receive all the points calculated by the program. It is not very annoying from a semiological point of view because visible saturation indicates a big population, but this unfortunately increases the size of the output file too.

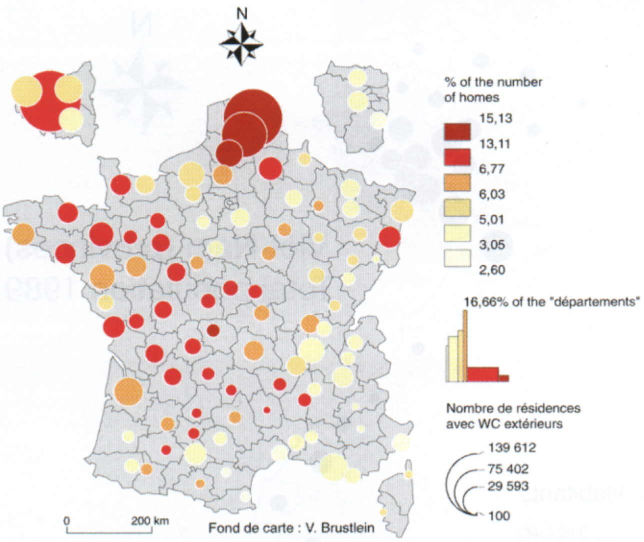


Figure 10. France (départements). Homes equipped with outside toilet, 1991

The Dotted map of the population of Bolivia in 1992 (Figure 13) demonstrates the value of this technique. The 'Andin Arc' appears neatly. Oriented NW–SE, then N–S, it regroups the main part of the country's population. Regions characterized by the highest densities of population form the part in the NW–SE direction which connects Cochamamba in La Paz. Outside this arc, the population is mainly located in Santa Cruz of the Sierra region which knows a demographic boom due to the agricultural development of the region.

Links maps

Links maps, so called 'sea urchin maps', allow one to show not only streams in a network, but relations between two points. Lines connecting points can be drawn in colour according to values of the variable. This map type is extremely efficient when one needs to express a polarized spatial organization. In this case, the links look like the spines of the sea urchin, which allows one to distinguish the boundary areas of influence. It is this that shows the areas of influence of the two main cities of the North of the State of Paraná in Brazil, Londrina and Maringa (Figure 14).

PHILEXPLO, AN EXPLORATORY CARTOGRAPHIC SOFTWARE FOR MACINTOSH

Principles of interactive cartography

The methodology for exploratory data analysis consists of looking at data from different viewing angles by adopting the mind of a detective who examines the scene of a crime: to keep an open mind (that is to avoid jumping to conclusions), to look for clues in the data (thereby avoiding placing *a priori* data in models of analysis which are often badly adapted to their characteristics). The exploratory approach becomes integrated into a process of research similar to a 'radiography' of data using multiple images which also show the abnormalities or extreme cases all too often rejected by more traditional methods which treat them as aberrant. Data exploration requires the extensive use of graphical representations, and the map can be considered as one of the many possible images of information to be analyzed. Philexplo's originality, in relation to the other software for exploratory data analysis, such as DataDesk for example, lies first of all with the presence of a window called 'Map' in addition to the different graphic windows. When graphic representations are displayed on the screen, the study of the distributions of the variable and the search for their possible correlation requires action on the spatial units (selection, colourization or concealation), which, when created in one or other one of the window's representations, are echoed on the others. These actions are realized using tools adapted to the characteristics of every graphic's window: slicing on histograms, selection with the lasso on the points, selection of neighbour spatial units on the map.

The principle of Philexplo's functioning is its dynamic nature, which makes it difficult to illustrate the process of investigation. The example below limits itself to a description of the elements of the interface specifically designed to facilitate investigation. Copies of the screens allow the reader to observe the functional wealth of the software. The user is always placed in front of a virtual desktop displayed on the screen of the computer on which are placed the different graphical representations of the statistical data. Philexplo uses the graphic interface of the Macintosh but with some modifications: several windows are used to display numerical parameters and graphic or cartographic representations; various cursors are used to interact with every type of graph; sounds are used to confirm the execution of an action; and there is a toolbar allowing one to choose the variables to be analyzed and the types of graphs to be displayed as well as allowing one to manage selections resulting from successive interactions.

The windows

Philexplo's desktop includes four components. The window containing the map occupies the left half of the screen. At the beginning of a session the program displays the base map formed by polygons bounding the spatial units (Figure 15). The right half of the screen is reserved for the visualization of the statistical graphs. It includes one or

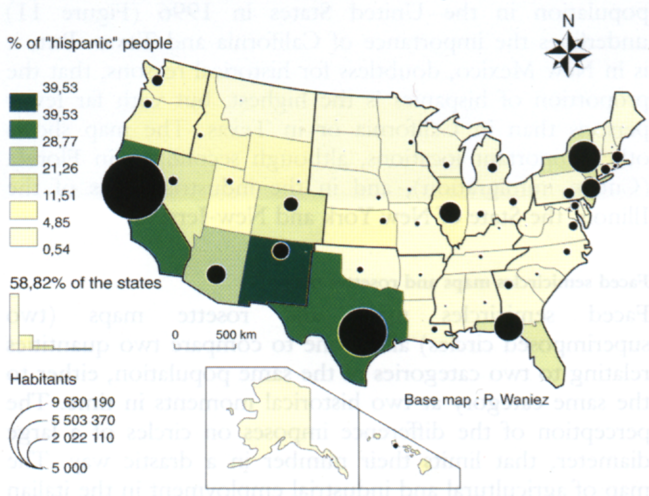


Figure 11. USA (states). Percentage of Hispanic people in the population, 1996

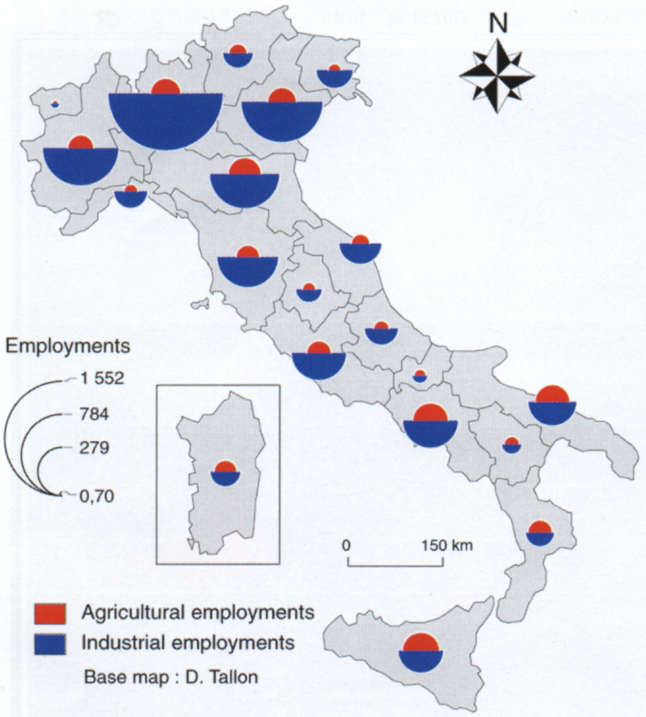


Figure 12. Italy (regions). Agricultural and industrial employment, 1998

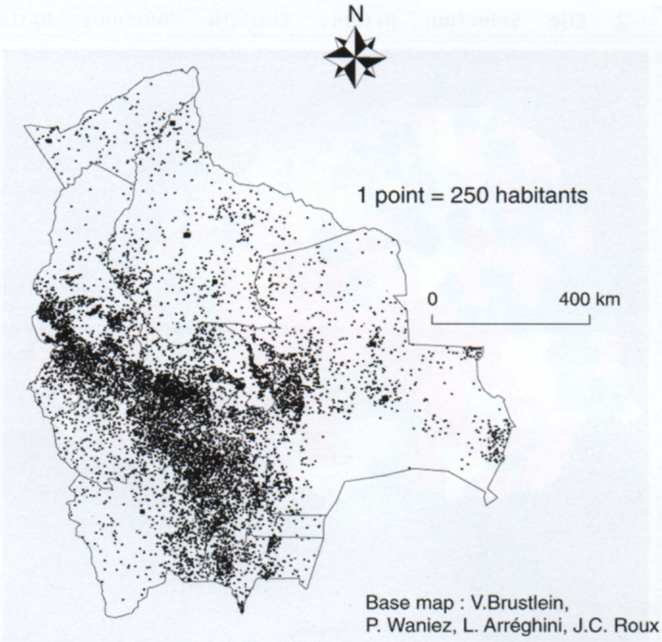


Figure 13. Bolivia (counties). Total population, 1992

more windows according to the number of graphs to be shown, and the number of variables selected. The lower third left part of the screen receives windows of different types, according to the necessities at the time: the window is intended to help identify the spatial units, windows are used to select options (colours scales, histograms number of classes, etc.). Finally, the lower right hand two thirds of the screen is occupied by the numerical values of the statistical parameters chosen by the user to evaluate the characteristics of the spatial units selections. Every column is identified with a colour corresponding to the selections. There are 6 colours (plus the black, in fact the white in the map's window that corresponds to the absence of selection).

The cursors

The interactions between the user and the desktop become established through cursors reacting to the clicks of the mouse (or to combinations of the mouse + keys on the keyboard). As the possible actions are different according to each of the windows, the shape of the cursors was chosen so as to remind the user of the current action:

- a click on the map with the finger allows one to select a spatial unity (or to deselect if it is already selected);
- the target is the equivalent of a finger when acting on graphs where spatial units are displayed using points. When the point is located in the centre of the target, a click allows one to select it (or to deselect if it is already selected);
- the lasso is intended to operate the selection of a group of spatial units shown in a window, polygons in a map window, points in a bivariate graph or triangular diagram window;
- the cross realizes selections of spatial units trapped using a circle, of which the centre is the location of the click on the button of the mouse;
- the knife is useful to slice histograms and box and whiskers plots;
- magnifying glasses allow, by a click inside the map's window, to zoom in or out the map with an increase 2× or 4×.

When the cursor is moved on the screen, it takes on a shape by default, dependent on the contents of the window where it is located: a finger on the map, a knife on histograms or box and whisker plots, a target on bivariate graphs and triangular diagrams, and the symbol 'No way' somewhere else. The presence of the cursor on a window automatically activates this window, and is visualized by a change of the window's background colour (of the black in the grey and the inversion for the subsequently deactivated windows).

The activation of a window with a particular content provides access to a family of particular interactions in this type of window (slicing using the knife on a

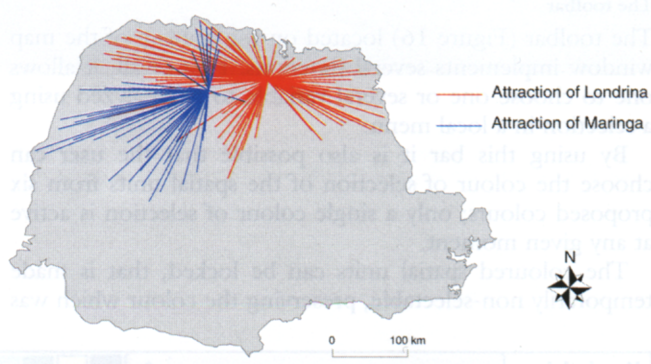


Figure 14. Paraná State, Brazil (communes). Links map:attraction of Londrina and Maringa

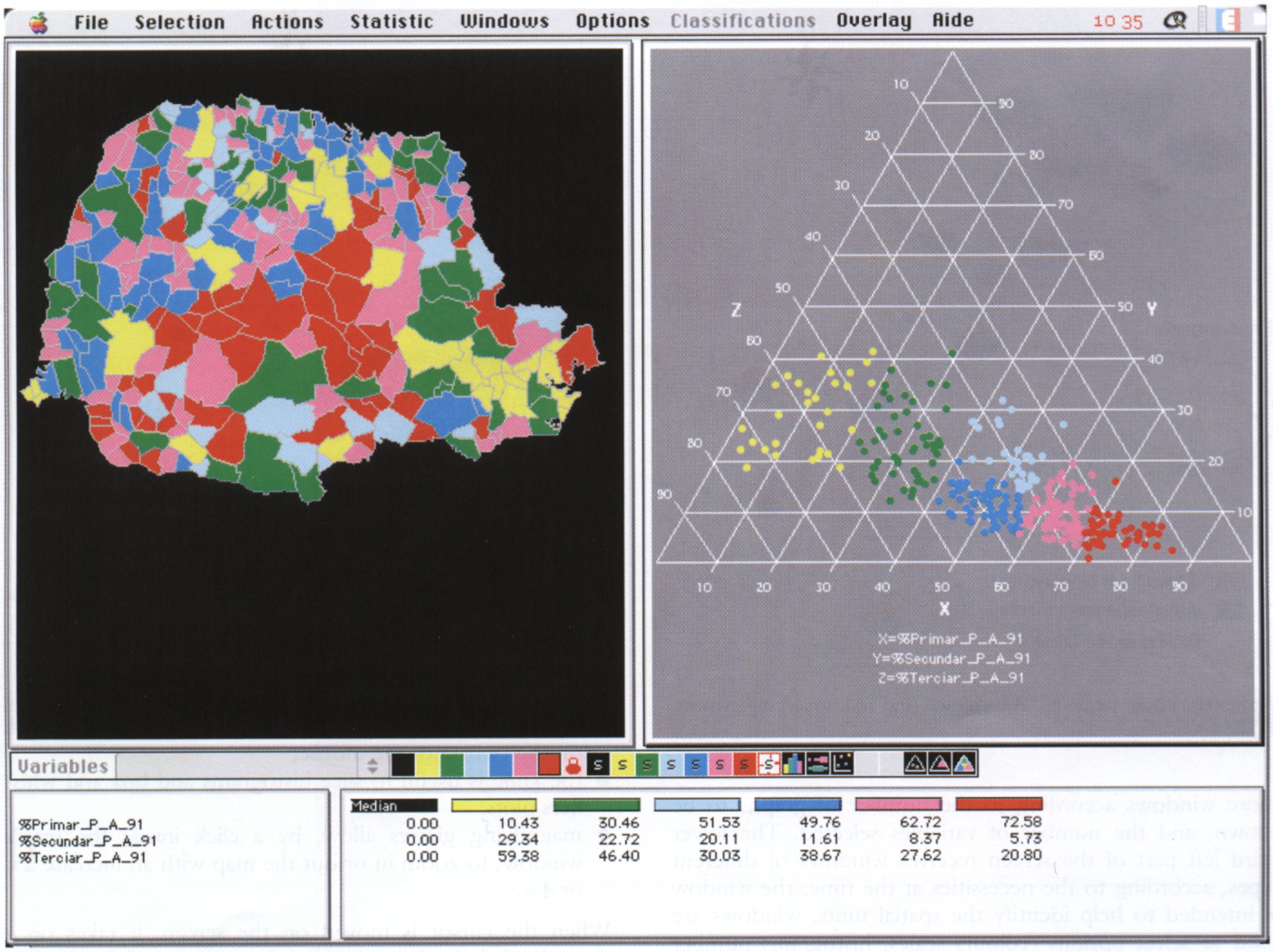


Figure 15. The desktop of Philexplo

histogram window, lasso selection in the lasso on a bivariate graph window etc.). Automatic activation is part of the standard interface of Macintosh which utilizes a mouse click in the window or on its bar to turn it active. However, this activation mode provokes too many disruptions in the interaction process through multiple mouses clicks and should be avoided to improve an investigation.

The toolbar

The toolbar (Figure 16) located on the bottom of the map window implements several functions. First of all, it allows one to choose one or several variable to be analyzed using a selection in a local menu.

By using this bar it is also possible that the user can choose the colour of selection of the spatial units from six proposed colours; only a single colour of selection is active at any given moment.

The coloured spatial units can be locked, that is made temporarily non-selectable, preserving the colour which was attributed to them before the activation of the bolt (the not-selected/not-coloured area cannot be locked).

The coloured icons set in the toolbar's centre which contains the letter 's' are selectors. They are useful to hide the spatial units which do not have the same colours as one or several active selectors: a click on a selector removes from the display (map or graph) all the spatial units which are not the colour of the selector. Several selectors can be activated successively, and their action is cumulative. The groups of spatial units can be easily isolated, and it is even possible to flash these spatial units (by repeated mouses clicks on a selector) to better perceive the position of the groups in relation to the others. The last icon containing the letter 's' cancels all the selectors (let us note that if all the selectors are active, it means having no active selector).

Finally, the choice of graphs is made by means of the fourth part of the toolbar. The icons which comprise it appear in a dynamic way, according to the number of variables held for the analysis: with one variable or more, histograms and box and whiskers plot are available; two

Figure 16. The toolbar of Philexplo

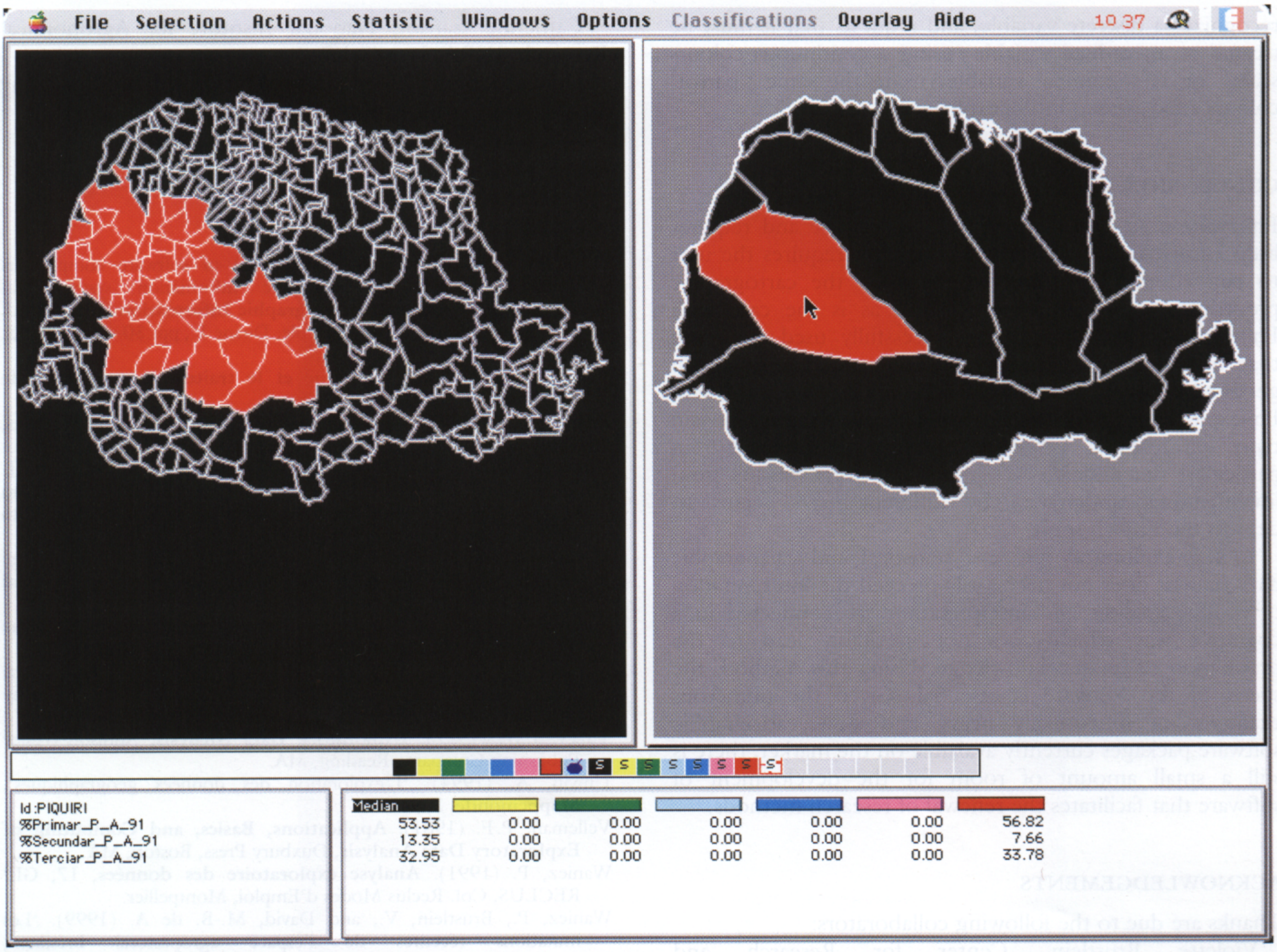


Figure 17. Philexplo: the selection of spatial units using the cross-map selection

variables add a bivariate graph, and three or more add a triangular diagram. When the graph type is chosen, additional icons available for this graph are displayed, for example drawing of regresion lines on bivariate graphs (using the full distribution and for each colour effectively used in the selections), resize and automatic partition of the triangular diagrams. To arrange the desktop in a better way, Philexplo shows only the graphs of the same type; at any one time the user can change a graph by a single click on the icon of the type of desired graph; the speed of Philexplo ensures fluidity without a clash with the display.

The menus

The menus complete the interface by providing access to functions less frequently used than those that compose the toolbar. The 'Selection' menu authorizes the inversion of the current selection. It allows a search for the neighbours of the selected spatial units: these neighborhoods can take another colour allowing one to discover the possible effects of spatial autocorrelation. These functions can be used by entering commands from the keyboard.

The 'Statistic' menu has two main functions. On the one hand, it allows one to choose the statistical parameter values displayed in the lower right window for each colour; it is the most common descriptive parameter (average, median, etc.). On the other hand, this menu gives access to a non-hierarchical classification which aggregates the spatial units not selected around seeds composed by the selected spatial units. The classification process is iterative, but quick, and stops when the spatial units do not move from a class to another.

The 'Overlay' menu makes it possible to select spatial units according to their inclusion (partial or total) in spatial units of another map (Figure 17). For example, one can select the municipalities of Paraná localized inside one or several hydrographic basins: two maps are displayed simultaneously on the desktop: a click in a hydrographic basin map (on the right) automatically selects the municipalities included (left map) It is also possible to manage existing selections, and to indicate their future in the final selection, that is if they must be added or excluded from the new selection. To do this, the icons of the 'Overlay' menu are displayed in a dynamic way, according to the existence or not of a current selection.

The 'File' menu allows one to record maps and graphs in the PICT format (a Macintosh toolbox format), but also to select a numerical format, what allows import into Philcarto to create other types of map such as, for example, a coloured proportional circles map with the colours of selection generated using Philexplo. Philcarto automatically recognizes a discrete variable, and requests that it must be treated as an ordinal variable (using a continuous colours scale), or as a nominal variable (using the same ' parrot' colours scale as used in Philexplo).

CONCLUSION

Philexplo is an easy software package to use and requires little knowledge of statistics. However, it requires the user to pay attention because he can trust the cartographic procedures programmed in advance as is the case with Philcarto. Philcarto has been successfully used for more than one year by a number of researchers and teachers in geography and cartography. This is not yet the case with Philexplo because its principles of use can confuse certain types of research based on Descartes's principles: in particular, the difficulty of getting the same results from investigations undertaken by different users seems to depress the French spirit!

In the exploratory process, statistical and cartographic data analysis does not necessarily preceed the interpretation of results; analysis and interpretation are connected in a cognitive way which does not inevitably lead to the production of unique knowledge. Using this method, the speed of the software is essential. Given the numerous examples of functionally heavy and slow cartographic software packages currently avalaible on the market, there is still a small amount of room for the development of software that facilitates the renewal of research methods.

ACKNOWLEDGEMENTS

Thanks are due to the following collaborators:

Violette Brustlein, Center for Research and Documentation on Latin America (CREDAL-CNRS), Paris, France.

Catherine Cansot, National Institute for Agronomical Research, INRA, Ivry/Seine, France.

Dora Rodrigues Hees, Brazilian Institute for Geography and Statistics (IBGE), Rio de Janeiro, Brazil.

And the users of Philcarto and Philexplo.

REFERENCES

Dangermond, J. (1990). 'A classification of software components commonly used in geographic information systems' in **Introductory readings in Geographic Information Systems**, ed. by Peuquet, Donna J., and Marble, Duane F., pp. 30–51, Taylor & Francis, London.

Bertin, J. (1977). **La graphique et le traitement graphique de l'information**, Flammarion, Paris.

Brunet, R. (1987). **La carte Mode d'Emploi**, Fayard/Reclus, Paris.

Cleveland, W. S. (1993). **Visualizing data,** Hobart Press, Summit, NJ.

David, M.-B. de A., Waniez, P., and Brustlein, V. (1997). 'Situaão social e demográfica dos beneficiários da Reforma Agraria', **Revista Brasileira de Estudos de População**, 14, 1/2, 169–84.

Hartwig, F., and Dearing, B. (1982). 'Exploratory Data Analysis', Sage University Paper, **Quantitative Applications in the Social Sciences Series**, 16, Sage Publications, Beverly Hills, CA.

Haslett, J. (1992). 'Spatial data analysis — challenges', **The Statistician**, 41, 271–84.

MacDougall, E. B. (1992). 'Exploratory analysis, dynamic statistical visualization and geographic information systems', **Cartography and Geographic Information Systems**, 19, 4, 237–46.

Tukey, J. W. (1977). **Exploratory Data Analysis,** Addison-Wesley Publishing Company, Reading, MA.

Unwin, A. (1994). 'L'exploration des données géographiques', **Mappemonde**, 3, 94.

Velleman, P. F. (1981). **Applications, Basics, and Computing of Exploratory Data Analysis**, Duxbury Press, Boston, MA.

Waniez, P. (1991). **Analyse exploratoire des données**, 17, GIP RECLUS, Col. Reclus Modes d'Emploi, Montpellier.

Waniez, P., Brustlein, V., and David, M.-B. de A. (1999). 'Les mutations récentes de l'espace agropastoral brésilien', **Mappemonde**, mars 1999, 26–35.

Waniez, P. (1999). **La cartographie des données économiques et sociales**, L'Harmattan, Paris.