Editorial Introduction

As is noted in several other Case Studies in this set, human geographers, at least in the UK, have probably made more use of visualization than any other discipline within the social sciences. In part, this results from the obvious centrality of the map as a means of communicating, storing and analysing geographical patterns. In part, too, it results from proximity to visualization practice in the physical sciences through contact with physical geographers. Less academic, but none the less important has been the historically good level of access and use of computing in research and, increasingly, in application of geographical research related to how the discipline in the UK has been funded.

However, viewed from the disciplinary perspective of geography both standard statistical analysis packages and the unmodified ViSc systems are deficient in their abilities to produce and modify maps or to integrate with the now universal geographical information systems used to store and organise the data. To date, this deficiency is being made good by the development of systems for what the authors of this Case Study call Exploratory Spatial Data Analysis (ESDA). Four such systems are compared, but in making the comparison an important step was the development of a set of task-oriented evaluation criteria. These are certain to be of more general interest.

\

The Visualization of Area-based Spatial Data

Stephen Wise, Robert Haining & Paola Signoretta
Sheffield Centre for Geographic Information and Spatial Analysis
University of Sheffield
Sheffield S10 2TN

Abstract

Area-based data, such as from the decennial census of population, are an extremely important source of information for many social science disciplines. Such data have a number of characteristics which make them different from other data used in the social sciences, such as the sensitivity of analytical results to the size and arrangement of the area boundaries, and this has lead to the development of specialist software tools for their analysis and display.

This case study presents a review of the current state of the art of software for the visualization of area data. The focus is on the use of visualization methods for exploratory spatial data analysis (ESDA) which normally encompasses one or more of the following: (1) the identification of unusual data values or errors in the data, (2) the detection of patterns in the arrangement of data values, (3) hypothesis formulation from the data and (4) the assessment of models of the data.

Four packages were selected for review based on their availability within the timescale of the work, and the range of visualization tools they offer. Two elements were found in common in all four packages:

- The provision of 'standard' statistical graphs, such as histograms, boxplots and scatterplots.
- The linking of these graphs to a mapped view of the data, such that elements identified in one view were highlighted in the others.

These facilities provide an extremely flexible environment for visualising the non-spatial elements of the data - both the 'smooth'characteristics such as the distributions of variables, and the 'rough' characteristics such as outliers from the distribution.

However there was a difference in the extent to which the packages provided tools to identify the explicitly spatial characteristics of the data. In two cases, the emphasis was on the provision of flexible graphical facilities, allowing extremely quick and easy redrawing of mapped views of the data, and graphical exploration of the location on the map of cases identified from the other graphs. The other two packages provided a series of facilities specifically for spatial data, such as the drawing of boxplots for areas at defined spatial lags, spatial smoothing to identify trend and the calculation of measures of local spatial autocorrelation and spatial concentration, for the detection of clusters of high or low data values.

We consider that both approaches are valuable, and that the development of a software package which combined the strengths of both, and was easily available to the academic community would be a useful tool for those wishing to analyze area-based spatial data. . In the report, the functionality which should a package should contain is identified, drawing on the examples of best practice identified in the four packages reviewed. Such a package would need to be easy to implement on commonly used platforms, and come supplied with documentation, including an introductory guide which outlined the main facilities, and explained the user interface. Given the importance of ease of implementation

\

and documentation, it might be necessary to seek collaboration with a professional software vendor in the production of the software.

Despite the fact that graphical facilities are seen as playing such a key role in ESDA, little work has been done to assess the graphical effectiveness of visualization tools, and we identify this as one area for further work. In the report we outline one possible methodology, drawing on work from the field of statistical graphics, which could be explored further.

Work on visualization tools for area-based spatial data has taken place within the framework of existing statistical and cartographic software, rather than using existing ViSC software tools. It would be extremely interesting to see whether ViSC products could in fact be used for the visualization of area-based data, since such products are already available to the academic community and structures have been put in place to support them, and this is also suggested as a topic for future research.

1. Introduction

1.1 Background

Data relating to geographically-defined units (hereafter referred to as 'spatial data') are an extremely important source of information for many social science disciplines. The clearest example of this is the decennial census of population, which provides the basic information on the socio-economic characteristics of the population of the country. The data are collected for individuals, but because of the necessary confidentiality restrictions are reported as aggregate statistics for small areas (Enumeration Districts) and various larger, standard sets of areas (such as wards, districts and councils) which are aggregates of EDs.

The census is by no means an isolated example - data on employment, unemployment, family expenditure, health, crime, industrial productivity and agricultural productivity are all provided on a regular basis as counts or rates for geographical areas (and much of it available to researchers from the ESRC Data Archive at Essex).

Geographical location is also an important factor in many areas of social and economic policy, at both national and European level. It is recognised that social and economic conditions vary across the country, and that in some instances assistance and resources can usefully be targeted at problem 'areas'. Spatial data are not simply of interest to those working in disciplines with a strong spatial element, such as Geography and Planning, but are of relevance to researchers in Politics, Sociology, Economics, Social and Economic History and Criminology.

Areal spatial data have a number of characteristics which distinguish them from other types of data used in the social sciences, and which have necessitated the development of specialist techniques and software.

There are a very large number of ways to define a set of areal units for the collection or reporting of statistics. The difficulty is that the choice can have important consequences for the later analysis of the data. The problem is compounded by the fact that in many cases the variables being measured cannot be defined independently of the frame of measurement. This is in direct contrast to the physical sciences, where variables such as pressure and temperature can be considered to exist independently of the sampling framework. By contrast a variable such as population density only has meaning in terms of a particular physical area.

The commonest method for displaying spatial data is by the use of a map - in the case of area data, choropleth and point symbol maps are the commonest two forms. Maps are commonly used to look for 'patterns' in the data, where the word pattern normally implies regularities in the data. The purpose is often to separate out broad (or smooth) features of the data from local irregularities (or rough feautures). The problem is that the same data may be mapped many different ways, each yielding a different visual pattern. This has been the subject of a great deal of research by cartographers and geographers, and is the subject of another of the case studies in this initiative. A particular problem is that the areas will vary in their spatial extent, to the extent that physically large areas may visually dominate the map.

Areas tend to vary in their base population. Even though standard reporting areas, such as EDs and wards, are designed to have approximately equal populations, other factors such as the variation in population density and historical and logistical constraints on boundary placements, mean that base populations within one set of areal units can vary by an order of magnitude. This means that rates calculated using population as the denominator will have different degrees of reliability across the map. This leads to the possibility that some aspects of the variation in the mapped values may be due to variations in the size of the base populations from which the rates have been calculated.

Even if the purpose of the analysis is not inherently spatial, it may still be necessary to understand some of the particular characteristics of spatial data in order that analytical results will not be misinterpreted. For example, published health statistics might be analysed to look for associations between material deprivation and the incidence of ill health. The areal reporting units provide a ready-made sampling frame for examining this question, splitting the country into a series of samples which can be analysed using regression for example. However, since the samples have been drawn spatially the assumption of independence of errors may not hold, undermining, among other things the usual inferential tests of significance.

Given the widespread use of spatial data in social science disciplines where there may not be a strong tradition of expertise in spatial data handling, there is clearly a need for software tools which assists the non-specialist in the analysis of spatial data handling. In this context, visualization methods may assist in the analysis of both the spatial and non-spatial elements of area-based data in four areas:

The identification of unusual data values or errors in the data.

The detection of patterns in the arrangement of data values.

Hypothesis formulation from the data.

The assessment of models.

Areas 1-3 are often referred to as exploratory spatial data analysis (ESDA. Area 4 includes disgnostic checks on models as part of confirmatory spatial data analysis (CSDA).

Visualization methods are not intended to replace numerical methods of analysis but rather complement them. They may also assist in the development and testing of numerical models of data.

Much of the work on the visualization of area-based data has been undertaken by academic researchers in disciplines with a specialist interest in spatial data and reported in the scientific literature in these disciplines. This means that many social scientists will be unaware of the potential use of visualization in analysing area-based data. The purpose of this case study is therefore to review the state of the art in this field to help disseminate knowledge about these techniques to a wider audience. We also identify opportunities for JISC and the ESRC for further work in this area.

1.2 Aim and Structure of Report

The aims of this report are as follows:

1. To assess the state of the art in the visualization of area-based data through an examination of four recently developed pieces of software.

\

2. To suggest areas where further work is needed, and what form this might take.

To this end, we first provide a brief review of the literature in this area to provide some background. The bulk of the document then consists of a review of four pieces of software designed to assist with the visualization of area-based data. These are reviewed in terms both of the range of graphical and statistical tools provided, but also in terms of the graphical effectiveness of their visualization tools. The review was undertaken by a researcher who is skilled in social science research using spatial data, but had very little previous experience of these kinds of software packages, since we felt this would provide a more representative view of how the existing software might be received by the social science community in general. Given this focus, and the limited time available, the review is not a comprehensive assessment of all the facilities provided by these packages.

Visualization may be used for many different types of analysis, and we have decided to focus on the use of visualization for the statistical analysis of spatial data, and in particular Exploratory Spatial Data Analysis or ESDA. As part of the review, we have developed a simple data model for ESDA which suggests the range of facilities one might require in order to undertake a full programme of ESDA in the case of area-based data. By comparing the existing offerings with this theoretical 'model' we are able to suggest those areas where additional facilities might be required.

2. Software for the Visualization of Area-based Data

The development of visualization techniques for area-based spatial data has its roots in the dissatisfaction expressed by a number of authors about the poor spatial analytical tools available in Geographic Information Systems (Goodchild 1987, Burrough 1990, Goodchild, Haining and Wise 1992) Most GIS provide analytical capabilities based on the idea of Map Algebra (Tomlin 1990) in which operations on layers of data could be used to identify spatial queries such as the amount of land satisfying a set of criteria, but were unable to provide any statistical operations beyond the calculation of descriptive statistics.

A number of researchers began to explore different means by which extra analytical functionality could be added to such systems, and this work is reviewed in Haining et al (1996). One of the problems however, was that many of the standard techniques of spatial analysis (as reviewed in Haining (1990) for example) are highly numerical and rooted in the statistical tradition of model building and hypothesis testing , and would not necessarily be of interest to many GIS users. Therefore a good deal of work also focused on the potential of exploratory methods, in which the aim is to use a range of largely graphical methods in order to explore properties of data sets and suggest hypotheses, but without the need for formal model building. EDA had been developed in the context of non-spatial data, and using traditional pen and ink graphics, but it was realised that the ideas could be extended to spatial data (thus producing ESDA) and implemented using interactive graphical methods.

One of the key early pieces of software in this field was SPIDER (Haslett et al 1990) which demonstrated how highly visual exploratory methods could produce useful insights into spatial data. SPIDER later became REGARD, and was a forerunner of MANET, one of the packages we review in this report. Table 1 contains a list of all the software packages of which we are aware which provide facilities for ESDA - for a review of the field see Haining et al. (1996)

Four of these software packages (cdv, MANET, SAGE and SpaceStat) have been evaluated. All are available to the social science community to varying degrees. They were chosen because, of the packages which were available within the rather limited time available for this study, they seemed to span the range of relevant software for spatial statistical analysis in terms of statistical capability and visualization capability. The non-inclusion of other packages should not be taken to imply that they are less effective visualization packages.

Cartographic Data Visualiser (cdv) developed by Dykes (1996) from the Tcl and Tk toolkit of Ousterhout (1994) is principally a map visualization toolkit with relatively modest computational and statistical capability. MANET, developed by Unwin and Theus builds on the earlier, innovative ESDA packages of SPIDER and REGARD. MANET is highly visual but contains a number of tools, drawn from the statistics literature, with a central objective of providing the user with a suite of tools for exploring multivariate spatial data and dealing with missing values (Unwin et al 1996). SAGE, developed by Haining, Ma and Wise (1996), adds ESDA capability to the ARC/INFO GIS but also includes spatial data modelling capability so that ESDA can be applied to raw data and model residuals generated within the package. Finally SpaceStat, developed by Anselin, is principally designed for spatial modelling but has been enhanced recently to include a visualization capability through an ArcView extension (Anselin and Bao 1997).

Reference	**Description**
Anselin, L. and S. Bao. (1997)	**ArcView extension for the visualization of results from SpaceStat.**
Bailey T.C. and Gatrell A.C. (1995)	INFOMAP. PC-based mapping and Spatial Analysis package distributed on disk with textbook on spatial analysis..
Brunsdon C and M.E.Charlton (1996)	Prototype spatial analysis system written using public domain software system XLisp-Stat.
Cook, D., J.J.Majure, J.Symanzik and N.Cressie (1996)	Link between ArcView GIS and the Xgobi graphics system. Has ability to handle both vector and raster data in graphical windows linked to statistical graphs.
Dykes J. (1996)	**cdv.**
Haining R.P., Wise S.M. and Ma J.(1996)	**SAGE.**
Haslett J., Wills G. and Uniwn A.R. (1990)	**SPIDER (later called REGARD).**
MacDougall E.B. (1992)	Polygon explorer. Mac-based software for map display and analysis.
Unwin A., Hawkins G. Hofman H and Siegl B. (1996)	**MANET**
Nagel M. (1996)	System built using data analysis language ISP. Includes a range of map types (Choropleth, isoline) and spatial analysis techniques.

Table 1: Software developed for the visualization of area-based spatial data. (Packages in bold are discussed in the report).

3. Software evaluations

The assessment criteria are grouped into three categories. The first of these covers the *functionality* of the package - the range of visualization tools and their graphical effectiveness. The second assessment category covers *ease of use* whilst the third covers *implementation*. These categories are typical of those used in more formal comparisons between software packages (Wise 1991, AGOCG 1992,1993) although assessment of the *effectiveness* of graphical facilities, as opposed to simply enumerating their range, does not appear to have been undertaken before, and this is one area we identify as providing scope for further work.

3.1 Functionality

The functionality of the packages has been assessed using two sets of criteria: (1) what is the range of graphical and numerical facilities offered and (2) how effective are the graphical tools at enabling the visualization of data.

3.1.1 Range of tools

We have taken the view that the range of visualization tools provided by any software package should not be assessed against a simple list (of possible tools) but rather against an appropriate data model which identifies generic properties of spatial data and the corresponding tools needed to do the job of identifying these properties. The range of visualization tools within any package is therefore assessed against some definition of "coherence" or "complementarity". Stated differently, "does the package offer the analyst the appropriate range of tools necessary to undertake specified tasks?". A package with a small but coherent set of tools which complement one another effectively with respect to some defined analytical goal would be more valuable to the spatial data analyst than a package which may have more tools but which fails the complementarity test or for which there is no clearly defined design criterion for deciding which tools to include.

The authors of all the packages have linked their software to the requirements of ESDA. This implies that the visual tools embedded in the systems should use statistically resistant methods and employ both graphical and cartographical forms of display for the purpose of identifying data properties, suggesting hypotheses, identifying atypical values and (if modelling capability is included) assessing model fits. Visualization is an integral part of ESDA.

In the EDA literature, the data being explored is often considered to consist of two components:

DATA = SMOOTH + ROUGH.

Smooth properties of non spatial data on a single variable are characteristics such as the centre of a set of values and degree of spread around this value, while outliers would constitute the rough element. Smooth properties of two variables might be the trend line specifying the relationship between the collection of pairs of values whilst rough properties refer to residuals around the fit line. For spatial data, the data model generates four components: smooth and rough elements of attribute values (without any reference to the arrangement of these values and hence equivalent to the non spatial case) and smooth and rough properties of the first and second order arrangement characteristics of the attribute values (which loosely equate to trend and spatial autocorrelation respectively). Table 2 below expands each of these components in more detail and identifies techniques used to identify each property.

\

	Smooth		**Rough**	
	Properties	*Techniques*	*Properties*	*Techniques*
Non-Spatial	Univariate: Centre; Spread; Distribution.	Boxplot; dotplot; histogram	Outliers	Boxplot;
	Multivariate relationships	scatterplot plus resistant fit or loess curve; mosaic plot;q-q plot; multi-way dotplot matrix scatter-plot	Relationship outliers	scatterplot plus resistant fit; mosaic plot
Spatial	Trend	Median filters; Trellis displays	Spatial outliers	Moran plots with resistant fit
	Autocorrelation	Moran plot with resistant fit	Local clusters	Getis-Ord

Table 2: Classification of ESDA techniques

Taking this as a model of spatial data, then software for ESDA should provide tools which allow the analyst to explore all four components. However there are a number of additional difficulties that arise in exploring spatial data and we turn to these now.

In the case of area-based spatial data, data values are not independent of the particular set of areas used to measure them moreover there is usually no “natural” or “right” partition. This is referred to as the Modifiable Areal Unit Problem (or MAUP) which has two components:

- Scale effects - as the average size of the areal units increases, variation in data values between areas is smoothed out. One consequence of this is that statistical measures such as correlation are influenced - in this case correlation values tend to rise.
- Aggegation effects - at any given scale there are a large number of ways to subdivide the study region into areal units, and different partitions can produce different analytical results.

We suggest that a fundamental tool for ESDA is the ability to construct new spatial partitions (whenever this is possible) in order to explore the sensitivity of ESDA results to the spatial framework. This means incorporating within the software the ability to construct new regional partitions for the data from the set of smallest units stored (Wise et al 1997).

Many techniques of spatial analysis depend on the construction of the connectivity (or **W**) matrix. This matrix defines the analysts assumptions about the nature of the spatial

\

relationships between areas of the chosen regionalisation. Just as there is no “correct” regionalisation there is no “correct” connectivity matrix and we suggest that to the extent that tools are made available in the software for analysing spatial relationships and which depend on defining the spatial relationships then it should be possible to replicate the analysis for other assumptions about the matrix **W** (Haining 1990).

Whilst not unique to spatial data the problem of missing data values is a particularly severe one. The fact that values may be missing has particular significance in a data set where interest focuses on the arrangement properties of the data and where a missing value refers to a particular area, not, as in classical statistics, simply a missing value from an experimental set which contains replication. Hence, the ability of the software to handle missing cases is potentially very important.

3.1.2 Graphical effectiveness of tools

The second key area of functionality is the effectiveness of the graphical facilities in assisting and encouraging the exploration of the data and activities such as hypothesis formulation and testing. In order to develop a set of criteria for making this assessment we draw upon ideas from the literature on statistical data graphing and cartography.

Surprisingly little attention seems to have been paid to the assessment of the effectiveness of different graphical techniques used in modern visualization software. Cartographers, who have a long history of assessing the effectiveness of different designs for traditional printed maps, have only just begun to address this issue (Monmonier 1991, MacEachren and Kraak 1997) with the establishment of an ICA commission on visualization (ICA 1997). This means that we have had to define a set of guidelines ourselves, building on existing work on statistical graphing (Cleveland 1994), the design of graphical material in general (Tufte 1983, 1989) and the design of paper maps (Buttenfield and Mark 1994).

Cleveland (1994) provides a useful classification of the tasks required by an observer to interpret a statistical graph, which will be extended for the purpose of this case study to the interpretation of maps. According to Cleveland there are two main ways in which a graph displaying either quantitative or categorical data may be read:

- Table Look Up - this is the process by which values for individual observations are read and normally proceeds in a sequential fashion i.e. one observation at a time. It requires the user to do one or more of the following :
    1. Scanning - identifying values from the axis/legend.
    2. Interpolating - e.g. between tick marks on an axis.
    3. Matching - finding symbol in the key.
- Pattern perception - this is the detection of patterns and trends in data, and requires the simultaneous observation of the whole graph. It requires the user to do one or more of the following:
    1. Detection - of the meaning of the symbols used to code the values. On a graph this refers to the use of different point symbols for different categories.
    2. Assembly - visual grouping of similar symbols.
    3. Estimation - of relationship between the assembled groups. At the simplest level this means identifying that a and b are different, at the highest level it means being able to estimate the ratio between values for a and b.

The purpose of the classification is to provide a basis for evaluating how well any particular graph performs in portraying data. The six activities can also be related to many of the principles of good design which have been identified by Tufte (1983,1990) for graphics in general and Dent (1985) and Cuff and Mattson (1982) for maps. If there are differences between pattern perception on graphs and on maps they may lie with respect to the types of pattern properties that are looked for on maps and the extent to which the individual (areal) units differ in size and hence contribution to detecting overall pattern.

In a series of influential books, Tufte (1983, 1990) has developed a series of principles which should guide the development of good graphics. The key one, underlying all the others, is that a good design should show the maximum amount of data using the least amount of ink in the smallest space. An effective graph or map should focus the viewer's attention on the data being displayed, and not on the attractiveness or otherwise of the design itself. By aiming for a compact presentation, it is possible to put more data on a single page, or the same amount of information in a smaller space, both of which facilitate comparison between data values. Essentially the points Tufte makes about good design all serve to assist the reader in undertaking the six tasks defined by Cleveland (1994) - graphs should have clear annotation, with consistent and logical symbolism enabling the reader to make unambiguous interpretations of the values and patterns in the data.

Similar comments may be made about many of the general rules of map design. These are covered at length in standard texts such as Dent (1985) and Cuff and Mattson (1992) but Buttenfield and Mark (1994) provide a very useful summary. They define the main stages of map production, and the rules governing each as follows.

Generalisation - this includes the selection of an appropriate amount of material to display so that the final map is not visually cluttered. With maps, it is also generally accepted that some distortion of reality is often needed - for example roads are often simplified, and buildings moved from their true geographical location so that they lie alongside the simplified road course. It is therefore common in maps to 'lie a little to tell a better truth' (Monmonier 1991) something which appears to be less common in drawing statistical graphs but occurs for example when "jittering" a graph to reveal overlapping cases. Brassel and Weibel (1987) also identify statistical generalisation - the simplification of data to portray trends rather than details - as part of this process.

Symbolisation - this includes the selection of appropriate symbols (point symbols, line types) and colours to portray the various map elements. A series of 'rules' have developed about appropriate symbolism, some based on features of the human perception system (e.g. red is perceived as a 'stronger' colour than blue), some on natural associations (e.g. blue for water), some on convention (e.g. brown for contours on UK OS maps) and some on principles of design (e.g. the use of colour 'sequences' to portray sequence in numerical data).

Production - the final stage includes organising the geographic and non-geographic map elements (title, key, scale) on the page to achieve a balanced overall design in which the main features of the map (i.e. the data in thematic maps) is visually most prominent.

It is clear that the design rules for statistical graphs and maps have many features in common:

- The use of symbols and colours to assist with data interpretation.
- The importance of layout and clarity of the final design.
- The need for appropriate ancilliary information such as keys and legends.

Presentation	**Visualization**
Static	Dynamic
One view of data	Multiple views of data
Data unknown to viewer	Data known to viewer

\

Table 3:
Comparison between features of presentation
graphics and visualization.

However, as noted above, all these rules have been developed in the traditional world of presentation graphics, which has a number of different characteristics from the world of interactive graphic visualization software (Table 3).

This means that some of the rules developed for paper maps may no longer be appropriate. As a simple example, maps for visualization do not need a north arrow, scale bar or title (normally considered essential for presentation maps) because the analyst can be assumed to be familiar with the study area. In general the move to interactive graphical presentation will produce two types of change:

Provision of the means to alter the view of the data interactively- maps and graphs do not not need to be designed once and for all, but can be changed during viewing to reveal or hide different features. For example, gridlines are often felt too obtrusive for statistical graphs, but in a visualization context it could be useful to have the ability to turn a set of gridlines on, in order to assist with reading an individual data value.

Provision of new techniques, such as animation, fly-throughs, hotlinking and brushing.

It is clear that a good deal of work remains to be done to examine just what rules are appropriate for graphics in visualization packages, and to assess the effectiveness of the graphical tools which exist at the present. Such work will become even more urgently required if one outcome of the current AGOCG project is to try and encourage a greater use of such tools among social scientists in general.

3.2 Ease of use

Ease of use is fundamental for no matter how good a software system is, if the system is difficult to use this will discourage many from persevering with it. This assessment criterion includes the quality of the user interface and the quality of the support documentation.

The quality of the user interface refers to the organisation of the available options and the extent to which this organisation helps or hinders, in the first instance, the process of learning to use the software and subsequently, the process of continuing to use the software. Are the packages easy to come back to after a period of not using them? Windows interfaces place a barrier between user and system simply because they do not correspond to the way in which we are used to accessing tools in more general circumstances. Individual menu items are hidden and no matter how logical the structure of the menu, the user is inevitably confronted with the need to retain a mental map of how to access the required tools. This is true whether the system is essentially a cookbook compendium of tools or whether it is a system that provides the ingredients which the user must then select from and adapt in order to build their analysis.

The quality of the documentation - its comprehensiveness, how easy it is to use when needing to deal with problems - will also be assessed. Is there a tutorial guide with a trial dataset provided to help the user get started and appreciate the range of tools provided?

3.3 Implementation

The implementation criterion is a technical criterion and includes how easy a package is to install, its portability, how easily it can be extended and the kinds of data formats that it supports.

\

The installation criterion includes how easy the package is to download or install on the users system. Portability includes the range of platforms on which the software will operate. Extensibility covers how easy it is for the user to add new facilities and whether this has to be done via the source code or whether facilities to attach macros are provided. Finally the issue of data formats refers to both inputs and outputs. Does the software read in data output by other packages and write out data that can be read in by other packages - and what are the packages it has such interfaces with?

3.4 Methodology

Given the constraints on time it was not possible to make a full assessment of the four packages under all these criteria. The main focus of the assessment was to:

- Review the packages in relation to the broad criteria defined in sections 3.2 to 3.4.
- Make an initial assessment of the extent to which such software could be useful to social scientists.

Therefore the approach which was used involved:

- Identification of the facilities offered by each package from the documentation (including Web pages, as listed in the appendix to this report).
- Dowloading and installing each package (except SAGE).
- Use of a range of facilities from each package. The inital intention had been to attempt to load a set of data into each package, and carry out a simple program of ESDA involving the display of maps and graphs, and the fitting of a simple regression model. In the event, the time it took to learn even the rudiments of each system, using the various sets of manuals, help systems and tutorials, meant that all that was possible was a general review of the facilities in each of the categories listed in Table 2. The fact that this is all that could be done is testament to the fact that even highly visual systems, which have been designed to have an 'intuitive' interface, require an investment of time before they can be used effectively.

cdv	http://cs6400.mcc.ac.uk/argus/	Project Argus home page at MIDAS
	http://www.geog.le.ac.uk/argus	Project Argus page at leicester
	http://midas.ac.uk/JANUS/cdv_on_midas/session2.htm	Instructions on how to use Cdv from Midas with your own datasets and Midas.
MANET	http://www1.math.uni-augsburg.de/Manet/	Manual available on request from authors
SAGE	http://www.shef.ac.uk/~scgisa	Follow links to details of SAGE.
	ftp://hippo.shef.ac.uk/pub/uni/academic/D-H/g/sage/sagehtm/sage.htm	Download in tar file (which contains manual)
SpaceStat	http://www.rri.wvu.edu/	SpaceStat home page.
	http://www.rri.wvu.edu/spacearc.htm	To download SpaceStat extension for ArcView 3.0) . User's guide for ArcView extension in progress

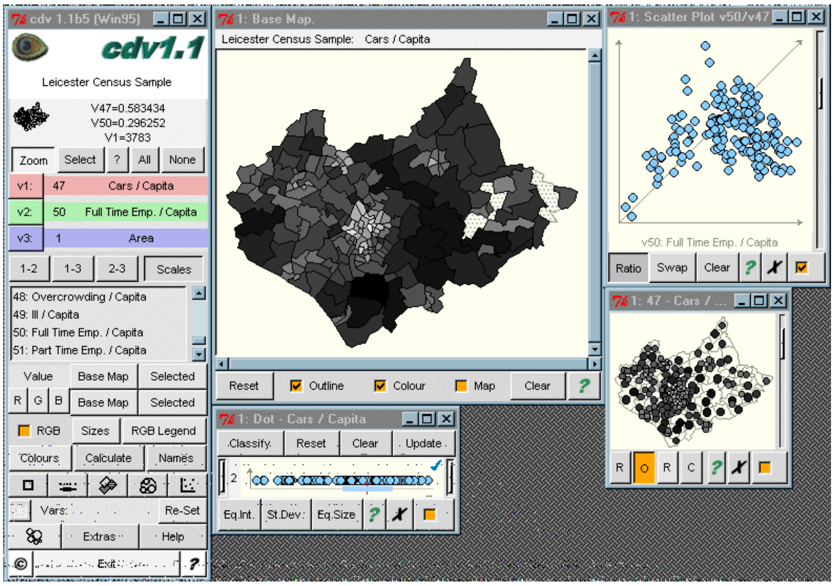
Table 4

The location of the web pages which provide further information about each of the packages described in this report.

\

4. Assessment of packages

4.1 cdv



4.1.1 Functionality

The bulk of cdv's facilities consist of graphical tools for viewing data, including a wide range of mapping options (choropleth map, point symbol map and cartogram) and some traditional statistical graphs (dotplot, histogram and scatterplot). The system will handle missing values. Most of the facilities are aimed at the exploration of non-spatial elements of the data. Exploration of spatial patterns is done by the linking of the graphs with the map, rather than by specialist spatial graphs or tools. The exception is a facility for exploring the rough properties of the spatial data, through the use of a local Moran's I coefficient. Once this tool is in operation, the selection of a polygon on the map produces a display of the value of the coefficient, and a graphical display of the relationship between the value in the selected polygon and its neighbours. We found this display difficult to interpret, and the purpose of this tool is not clear in the context of the rest of the system.

A marked feature of the package is the ease with which the graphical displays can be created and manipulated. Maps are shaded in a greyscale by default, but alternative colour schemes can be set up. In addition, multiple variables can be displayed on the map by assigning them to one of Red, Green or Blue. For those unfamilair with such maps a graphical key is provided to help interpret the resulting colours. The statistical graphs do not have labels on the axes, although the data value for any given area is automatically displayed in the main menu window whenever the area is selected with the mouse.

A particularly nice feature is the option of viewing a choropleth map and a cartogram of the same data at the same time. The key weakness of the choropleth map is that the size and arrangement of the areal units has a strong influence on the look of the map - an area which has a 'high' value for whatever is being mapped will be far more noticeable if it

\

happens to be large than if it is small. One solution to this is to produce a new set of areal units, where the size of the unit is related to its 'importance' rather than its physical extent - this is usually done by relating the size to the number of people in the area. The problem is that this distortion of the geography can make it difficult to interpret the resulting map, since areas may have to be moved away from their true geographical location to 'make way' for more populous areas.

By having a choropleth map and a cartogram on-screen some of these problems can be reduced, since it is possible to discover the true location of any area on the cartogram by clicking on it and seeing where it is located on the choropleth map. This is an excellent example of how interactive graphics can simply adopt methods developed for paper maps, but help to overcome some of their design problems.

4.1.2 Ease of Use

The initial attempt to use cdv was hindered by the fact that the version at MIDAS has no help information, although it does have a very useful tutorial guide. It was therefore possible to get started, but then became rather difficult to explore what many of the options and menus did. The version on the CD-ROM did have help information, although not the tutorial, and this was helpful for the new user.

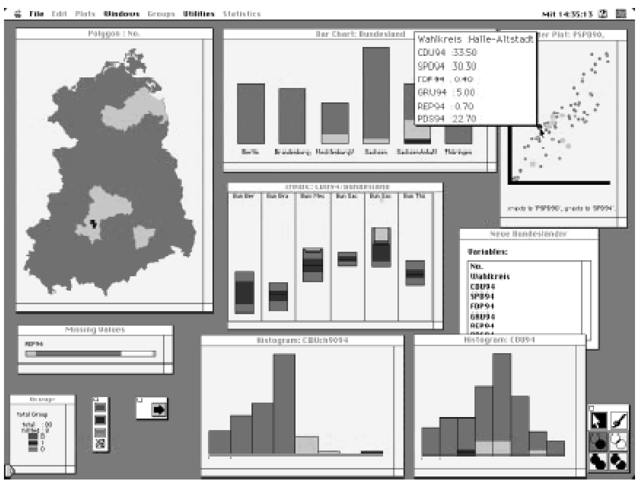
4.1.3 Implementation

There are two versions of cdv - one which may be freely downloaded from the MIDAS web site (address given in the appendix to this report) and one which was provided to us by the author (Jason Dykes) on CD-ROM. Both were installed without difficulty.

An interesting feature of cdv is that it is written in a scripting language Tcl which has a windows interface called Tk. Since the implementation inlcudes Tcl/Tk this means that an individual user could modify cdv if so desired. There was not time during this study to assess what skills would be required in order to learn Tcl/Tk and to understand enough of the structure of cdv to be able to add new facilities.

Data input is via text files, which can be fairly easily generated from a spreadsheet package. The geographical data is again supplied in text format, which can easily be generated from ARC/INFO. In order to read different files however, it is necessary to edit a parameter file to inform cdv of their existence, which is rather clumsy, and not what users would expect.

4.2 MANET



\

4.2.1 Functionality

Like cdv, MANET has been designed principally to provide facilities for the graphical exploration of data, with the additional emphasis on the development of methods for dealing imaginatively with missing values (hence the name of the package - **M**issings **A**re **N**ow **E**qually **T**reated).

In terms of our classification of ESDA functionality, MANET has excellent facilities for exploring the non-spatial elements of data. These include standard graphs, such as histograms, boxplots and scatter plots, and more innovative techniques such as trellis plots, in which a series of separate boxplots are drawn for different categories of data (e.g. data from separate years) in a tabular form (an example is shown in the central window of the screen layout above). The trellis plot is actually an example of what Tufte (1990) calls ‘small multiples’, in which repeated graphics are used to encourage comparison between different subsets of data, and would make it extremely easy to detect differences between different categories within a dataset. There are some very useful additional features including weighted and ‘spine’ versions of histograms, and the capability to automatically and continuously change the histogram bin sizes.

Exploration of the spatial elements of the data is done using the linkage between the map and the graphs. For example, by moving a selection tool across a graph, the **geography** of the data set can be explored. However, there appear to be no specialist spatial graphs or tools for detecting **spatial** properties such as trends or clusters.

All the windows are linked to one another, such that selecting an element in one (e.g. a point on a scatter plot) highlights the related elements in the others (e.g. the area on the map). The selection of elements can be done in numerous ways, selected using a special graphical menu - the default is for selection of a single element at a time. However, it is possible to set the system up so that selections are cumulative for example.

A characteristic feature of the package is that most of the displays do not show the actual values of individual data items - indeed it is very difficult to discover what the data values are at all, since there is no tabular display of the data matrix, and no tools for producing summary statistics. The focus of the software is thus entirely on the detection of patterns and trends.

4.2.2 Ease of Use

MANET is extremely easy to use, once the basic operation of the system is understood. Some of the newer features (such as drawing a map of the data) are not yet documented, so that assistance was required from the author to test these out. The manual is clear, apart from the description of the various options for selecting cases. This caused a good deal of confusion until an explanation was discovered on the Web page.

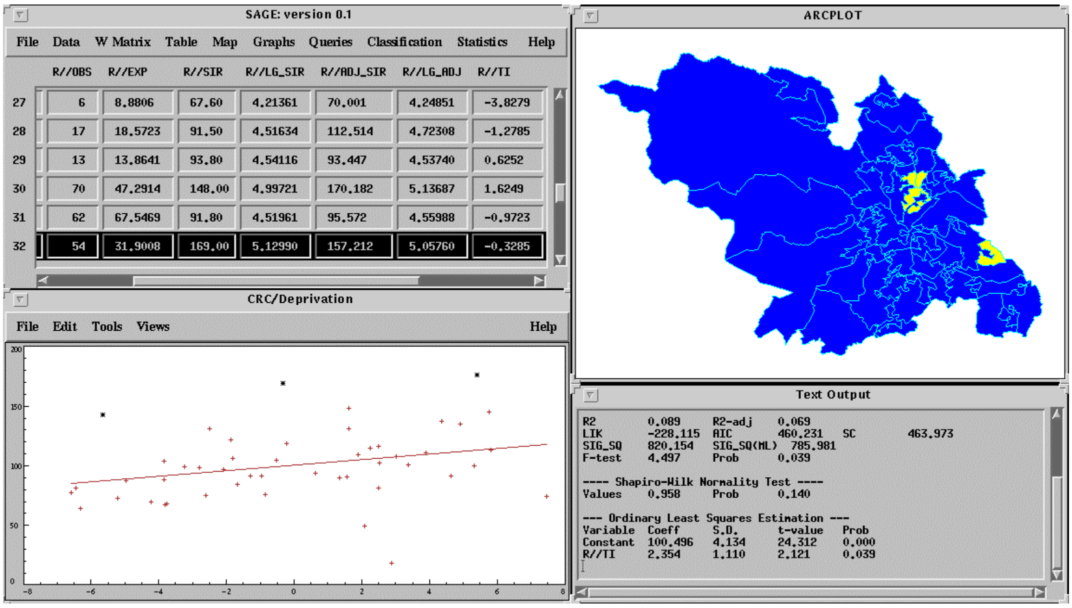
4.2.3 Implementation

MANET is available by email from the author and was downloaded and installed without difficulty.

Data can be input using a simple text file format.

\

4.3 SAGE



4.3.1 Functionality

In terms of functionality SAGE includes a balance of graphical and numerical methods and provides some tools for exploring the smooth and rough properties of both the spatial and non-spatial elements of the data. The usual range of numerical descriptive statistics (both resistant and non-resistant) can be output to a text window. Graphical tools for exploring the non-spatial elements of the data include boxplots, histograms and scatterplots. Spatial graphical tools include a Moran plot, in which the values for each area are plotted against the mean values for neighbouring areas, a type of trellis diagram based on spatial order and facilities for spatial filtering.

SAGE is unique in providing facilities for classification (beyond that needed for cartographic display) and region building. Despite the known problems of area-based data and the sensitivity of results to the areal partition, this is one area of functionality which has received very little attention in the ESDA literature. SAGE also includes facilities for constructing and modifying a wide range of weights matrices, and making adjustments for variable base populations. There is no facility for handling missing values. SAGE has a spatial modelling capability, similar to SpaceStat, so that ESDA and other tools can be used for model assessment.

All the graphical windows are linked, so that selecting an item in one automatically highlights the same item in all the other graphical windows. The statistical graphs are easy to read and contain axis labelling. Although it is possible to open multiple graph windows, tt is only possible to have one map window open, and this is drawn by the ARC/INFO GIS package. Altering the appearance of the map is done by using a series of ARC/INFO menus, and is rather long-winded and cumbersome. The default colour scheme is well suited to displaying categorical data, but not to data on rates or proportions. There is a limited spreadsheet capability that allows the user to see the non-spatial data, and to create new functions of the existing variables.

\

4.3.2 Ease of Use

SAGE is controlled using a menu interface on the main SAGE window (top left window on the diagram). The menus can sometimes be non-intuitive, the worst example being in the drawing of graphs. After selecting the graph type from the main menu, the user is prompted to enter a name for the graph - it is not clear why this is needed, and it is not possible to allow the name to default - something must be typed even if it is nonsense. After this a blank window appears! In fact this is a canvas on which one or more graphs of the desired type can be drawn - in order to draw a graph the Add option on the window menu is selected, and this finally brings up a menu where the relevant variables can be selected.

The documentation is such that it is fairly easy to learn how to use the system. However, given that some of the facilities provided are quite sophisticated, and really require the user to understand something about spatial analysis, and the use of weights matrices, it would take a beginner some time to become familiar with all the facilities.

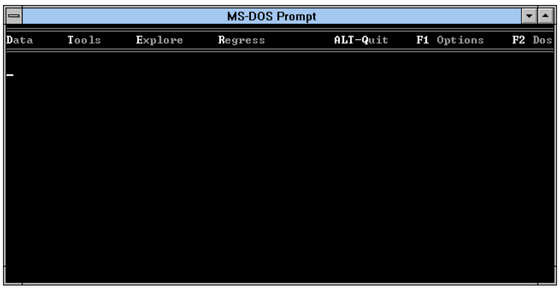
4.3.3 Implementation

SAGE was unusual in this exercise, since it did not need to be downloaded or installed, as it was written at Sheffield. However, we know from the experience of those elsewhere who have tried to dowload the software that this presents problems. SAGE is linked to a particular version of the Unix operating system on Sun workstations, and does not appear to work on any other configuration. Since there is no longer any manpower available to provide technical support for it, this unfortunately limits its availability to other researchers.

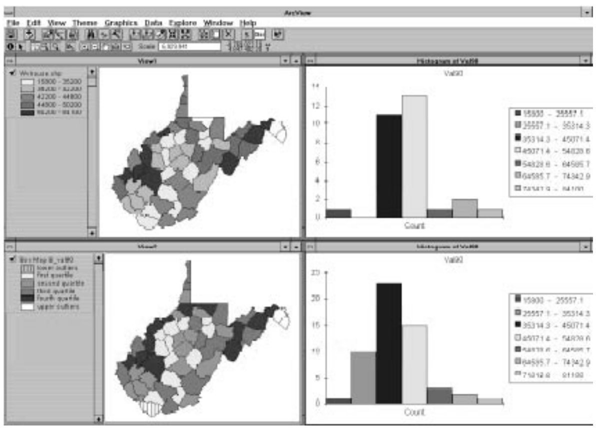
SAGE is written as a self-contained package, with no facility for additional facilities. It comes supplied with source code (in C++) but only a skilled programmer would be able to use this to modify the package significantly.

SAGE reads its data from ARC/INFO, which means it does not have any data input facilities of its own. A complication, however, is that before a dataset can be read, the user must create a View file - this is in fact another ARC/INFO facility which SAGE simply makes use of, but is rather complex for a novice, and as the first thing which must be done, rather a daunting introduction to the package.

4.4 SpaceStat



SpaceStat



ArcView extension

4.4.1 Functionality

SpaceStat differs from the other packages reviewed here in that the package itself has no graphical facilities at all. However an extension to ArcView has recently been developed, which allows the results from SpaceStat to be visualised using the facilities of ArcView.

\

The focus of SpaceStat, as the name implies, is on spatial statistics and in particular, spatial econometric modelling. The ArcView extension provides facilities for viewing the original data through a map linked to one or more statistical graphs (boxplot, histogram, scatterplot). In addition, by importing a results file from SpaceStat, it is also possible to explore the smooth and rough elements of the spatial data using lagged boxplots and local autocorrelation coefficients.

In the time available it was not possible to run an analysis in SpaceStat and export the results to ArcView. The histogram and boxplot in ArcView were fine, but there is a minor problem in that if a graph is drawn, then clicking on any of the other windows (e.g. the map) causes the graph window to be minimised onto the Windows 95 task bar.

4.4.2 Ease of Use

The menu system of SpaceStat has a very old-fashioned feel to it compared with modern Windows-based systems. However, it is actually easy to use once the basic technique of entering command parameters has been mastered.

The documentation is extremely thorough, leading the user through a series of worked examples and explaining the principles underlying the methods which are available. Given the highly specialised nature of the statistical techniques which SpaceStat has been written to implement, there is no 'quick-fix' solution to learning the package - anyone wishing to use it would simply have to devote some time to reading the manual carefully.

4.4.3 Implementation

SpaceStat comes on diskette, and installation is reasonably straightforward. One minor problem is that it requires a particular graphics device driver (ANSI.SYS) to be loaded, otherwise the menu system does not appear properly on screen. Installing ANSI.SYS is simply a matter of editing the CONFIG.SYS file - simple enough for the technically confident, but something of a deterrent for others, and not what one expects in the days of the Windows 95 Install Shield! As an extra complication, when SpaceStat was installed on a machine running Windows 95, it was not easy to find the ANSI.SYS driver and the help of our local computer service had to be enlisted! Data input is via text files, with the option of providing alternative weights matrices, either as text files or as output files from GIS packages such as ARC/INFO and IDRISI.

ESDA functionality	cdv	MANET	SAGE	SpaceStat
Non-spatial	yes	yes	yes	yes
Spatial - smooth	partial	partial	yes	yes
Spatial - rough	partial	partial	yes	yes
Sensitivity analysis				
Classification/regionalisation	no	no	yes	no
Attribute manipulation	yes	no	yes	yes
Alternative weights matrices	partial	no	yes	yes

Table 5

Summary of types of ESDA functionality in the four packages

\

5. Summary and discussion

5.1 Functionality of software

Table 5 summarise the ESDA functionality provided by the four packages.

As can be seen he packages can be divided into two groups in terms of the focus of the functionality provided.

Both cdv and MANET have an emphasis on graphical facilities, rather than numerical ones. They are both extremely flexible in the way that the appearance of graphs and maps can be altered. For example, when a point symbol map is produced in cdv, the size of the symbols used to represent the data can be altered by a simple slider bar. Not only is this an extremely simple and easy to use technique, it works quickly enough that it is possible to experiment with the effect of different sized symbols on the appearance of the map, and a suitable size can very easily be determined. This can be compared with a typical Windows mapping package, such as MapInfo, in which the same operation requires the use of a menu panel in which the size is specified numerically. The effect can only be seen by entering the new value and returning to the map, so that adjustments to the size requires a constant switching between these two windows.

MANET has a similar facility with histograms, where the number and size of the histogram bins can be altered dynamically. Again, this is superb as a tool for demonstrating the effect of changing these parameters on the histogram and would make a good teaching tool. MANET also has an excellent missing data facility and provides state of the art visualization tools such as weighted histograms, trellis displays, mosiac plots and the use of brushing and tracing.

There are situations however, where this emphasis on the graphical is more open to debate. For example, neither package appears to put labels on the axis of scatterplots, dotplots or histograms. In this sense, these graphs are unsuited to table look-up, one of Cleveland's two uses for graphical displays. It is possible to discover the value of any particular data point, in cdv at least, by clicking on that point in the graph. However, it would seem that one of the uses of graphs such as dotplots is to indicate the numerical range in the values of a variable, as well as illustrating the shape of the distribution and the presence of outliers.

This emphasis on graphical display in cdv and MANET leads these packages to provide very little in the way of numerical analysis measures. In contrast, SAGE and SpaceSat provide a much richer range of numerical methods, but with correspondingly weaker graphical facilities. SpaceStat itself has no graphics of course, but uses ArcView as a tool for exploring the data and results of analysis. In SAGE the graphical tools are integrated into the system, but ARC/INFO is used for map drawing. The contrast with cdv and MANET illustrates the difference between the traditional world of the paper map and the new one of interactive visualization. Whereas the colour symbolism of a choropleth map can be changed easily and quickly in both cdv and MANET, it is a very cumbersome process in SAGE. What is more it is impossible to get SAGE to automatically generate a greyscale which has the correct number of shades for the dataset being mapped (which happens in cdv and MANET).

All the packages provide facilities for the exploration of non-spatial elements of the data but again the two groups of packages differ in their provision of tools for exploring the spatial properties of the data - the identification of spatial trends (representing the smooth element of the data) or the presence of spatial outliers (the rough). SAGE and SpaceStat provide specialised graphs (such as the Moran scatterplot and lagged boxplot) and numerical

\

measures (such as local autocorrelation statistics) for this purpose, whereas cdv and MANET rely on the use of the map and its link to other graphs

Finally, SAGE is unique in providing facilities for the manipulation of the areal units themselves, through the use of a regionalisation facility.

5.2 Ease of Use

It was suggested at the beginning of this report that graphical tools might be intuitively easier to use for those unfamiliar with spatial data. While this may be true, it is clear that none of these software packages can be used without the investment of considerable time. The problems lie in several areas:

There is an inherent problem in any menu-based system, that the menus are rarely self-explanatory, and must be learnt. Although the use of Windows style menu interfaces has provided a uniform look to software interfaces, so that the mechanics of making selections, setting parameters, opening files etc can be learnt once and applied to several packages, the logic of what the menus mean still has to be learnt. A feature of software for spatial data (including mapping packages and GIS), which sets it apart from many other software packages is that there is a definite logic to the sequence in which menus should be used. The clearest example in the case of ESDA is that you must create a **W** matrix before some other functions can be used - SAGE does this automatically, but in SpaceStat it must be done explicitly.

Some of the visualization techniques used would require some explanation to those unfamiliar with computer graphics. For example, cdv's plotting of multivariate maps by assigning each variable to one of Red, Green or Blue is a technique which has been tried by various cartographers (it is used in Dorling's census atlas for example (Dorling 1995) but which can produce a map which is hard to interpret. There are several difficulties here:

- The idea of a single 'colour' being used to represent two variables simultaneously will certainly be a new concept to many people familiar with viewing traditional choropleth maps. In paper maps this is tackled by the careful design of a legend, and something similar is attempted in cdv, with an optional, pop-up legend.
- RGB colour space is known to be non-intuitive - for example people more familiar with colour mixing using paints would not expect Red and Green to combine to produce yellow.
- Even when the viewer has grasped the previous two concepts, the map may be hard to read. The extreme ends of the colour spectrum will be easy to interpret - vivid red clearly means one variable is at a maximum, while the other is at a minimum. However, it is far harder to assign a clear meaning to shades in the middle of the spectrum.

There were other examples in the various packages where the meaning of some of the graphs and charts was not clear. Overall, we had the strong feeling that work on these packages had concentrated on exploring what was technically possible. Now that working systems have been developed, we feel there is a need for research into just how effective these tools are for general users.

5.3 Implementation

The packages varied somewhat in the ease with which they could be implemented. Given that all four are the product of academic researchers, rather than professional software vendors, it would be unfair to be over-critical in this area, but if JISC decided that there was a case for encouraging the use of this type of software, then this is an area which would require attention.

\

6. Conclusions and Recommendations

It is clear that no single package provides the full range of tools identified here as being desirable in an ESDA package, in a truly interactive, graphical environment. From the evidence of this limited review, each of the two approaches to software development used in the packages reviewed has some merit, and could be used to develop a more comprehensive system:

- Through the extension of a free-standing package such as cdv or MANET. The challenge here would be to add further facilities such as spatial analytical tools and the ability to examine the implications of modifying the spatial framework.
- Through the development of a link with GIS software as with SAGE and SpaceStat. Here the challenge is to provide a more interactive and flexible graphical environment.

The facilities such a system should contain can be identified, drawing on the examples of best practice from each of the four packages reviewed in this report:

- The provision of a wide range of cartographical and statistical views of data in linked windows.
- A range of different tools for selecting cases in any given window. For example, in a map window, the ability to select one or more areas, all areas within a defined shape, all areas along a trajectory etc, with the possibility of dynamic update of other windows as the selection tool is moved.
- A 'spreadsheet' type display of the non-spatial data and the ability to create new variables from exisiting ones using mathematical operators.
- The interactive modification of histogram bin sizes and symbol sizes on point symbol maps.
- The ability to turn on or off the labelling of graph axes.
- The linking of multiple cartographic views of the same data. For example, linking a cartogram with a choropleth map of the same data greatly assists in the interpretation of the cartogram.
- The ability to deal with variable base populations.
- The capability to assess the sensitivity of findings to changes in the spatial framework. Facilities here might inlcude the ability to modify the weights matrix or to produce new sets of regions.
- The provision of specialist tools for exploring spatial properties of the data, in addition to the linking of a map with graphical views.
- Handling of missing values.

Within the time available for this study it was not possible to make a very thorough evaluation of the graphical effectiveness of the software packages. However we feel that this is an important aspect of visualization software, which appears to have been overlooked by the Scientific Visualization community, and which deserves closer attention. The classification proposed by Cleveland (1994) was used as the basis for our subjective evaluation, and we feel that this may form a useful starting point for the development of a more thorough methodology in this area.

Work on visualization tools for area-based spatial data has taken place within the framework of existing statistical and cartographic software, independently of ViSC software such as Iris or AVS (Earnshaw and Wiseman 1992).This may be in part due to the way in which visualization of this type of spatial data has grown out of initial concerns with the lack of spatial analysis facilities in GIS systems. Whatever the reason, it would be extremely interesting to see whether ViSC products could in fact be used for the visualization of area-based data, since they are already available to the academic community and structures have been put in place to support them and this is also suggested as a topic for future research.

7. Acknowledgements

We are indebted to the authors of the software packages described here, who have have given freely of their time and expertise to assist us in carrying out this review: Jason Dykes (cdv), Anthony Unwin (MANET) and Luc Anselin (SpaceStat). We are aware that we have had a rather limited time in which to make our evaluation, and hope that this has not caused us to misrepresent any of the systems. We would stress again that the intention of this case study is to present a review of what is the current state of the art in the visualization of area-based spatial data across the range of available software packages, rather than making detailed comparisons between them.

8. References

AGOCG (1992) *Evaluation of Visualization Systems*. Advisory Group on Computer Graphics, Loughborough University of Technology, Loughborough.

AGOCG (1993) *Image Processing Software Evaluation*. Advisory Group on Computer Graphics, Loughborough University of Technology, Loughborough.

Anselin, L. and S. Bao. (1997) "Exploratory spatial data analysis linking SpaceStat and Arc View." In M.Fischer and A.Getis (eds) *Recent Developments in Spatial Analysis: Spatial statistics, behavioural modelling and neuro-computing.* Berlin, Springer-Verlag p35-59.

Bailey T.C. and Gatrell A.C. (1995) *Interactive Spatial Data Analysis*. Longman, Cambridge..

Brassel K.E. and Weibel R. (1987) Map generalisation. In Anderson K.E. and Douglas A.V. (eds) Report on International Research and Developments in Advanced Cartographic Technology, 1984-1987, International Cartographic Association, 120-137.

Brunsdon C and M.E.Charlton (1996) Developing an exploratory spatial analysis system in XLisp-Stat . In D.Parker (ed), *Innovation in GIS 3*, 135-145. Taylor and Francis, London.

Burrough P.A. (1990) Methods of spatial analysis in GIS. *Int.J.Geographical Information Systems* 4(3),221-223.

Buttenfield B.P. and Mark D.M. (1994) Expert systems in cartographic design. in Taylor D.R.F. (ed) *Geographic Information Systems: the microcomputer and modern cartography*. Oxford, Pergamon, 129-150.

Cleveland W.S. (1994) The elements of graphing data. AT&T Bell Laboratories, Murray Hill NJ.

Cook, D., J.J.Majure, J.Symanzik and N.Cressie (1996) "Dynamics graphics in a GIS: exploring and analyzing multivariate spatial data using linked software" *Computational Statistics* Vol 11, 467-480.

Cuff D. and Mattson M.T. (1982) *Thematic maps: their design and production*. New York, Methuen.

Dent B.D. (1985) *Principles of thematic map design*. Addison-Wesley, London.

Dorling D. (1995) *New Social Atlas of Britain*. Wiley, Chichester.

Dykes J. (1996) Dynamic maps for spatial science: a unified approach to cartographic isualization. In D.Parker (ed), *Innovation in GIS 3*, 177-187. Taylor and Francis, London.

\

Earnshaw R.A. and Wiseman N. (1992) *An Introductory Guide to Scientific Visualization*. Springer-Verlag, Berlin.

Goodchild M.J., Haining R.P., Wise S.M. and 12 others (1992) Integrating GIS and spatial data analysis : problems and possibilities. *Internatinal Journal of Geographical Information Systems* 6(5), 407-423.

Goodchild, M.G. (1987) "A spatial analytical perspective on geographical information systems" *International Journal of Geographical Information Systems*, 1, p327-334.

Haining R.P., Wise S.M. and Ma J.(1996) The design of a software system for the interactive spatial statistical analysis linked to a GIS. *Computational Statistics* 11,449-466.

Haining, R.P (1990) *Spatial Data Analysis in the Social and Environmental Sciences.* Cambridge University Press.

Haslett J., Wills G. and Uniwn A.R. (1990) SPIDER - an interactive statistical tool for the analysis of spatially distributed data. *International Journal of Geographical Information Systems* 4(3),285-296.

ICA (1997) *Commission on Visualization*. International Cartographic Association. http://www.geog.psu.edu/ica/ICAvis.html.

MacDougall E.B. (1992) Exploratory Analysis, Dynamic Statistical Visualization and Geographic Information Systems. *Cartography and Geographic Information Systems* 19(4),237-246.

McEachren A.M. and Kraak M-J (1997) Guest editorial: exploratory cartographic visualization: advancing the agenda. *Computers and Geosciences* 23(4), 335-343.

Monmonier M.S. (1991) *How to Lie with Maps*. Chicago, University of Chicago Press.

Nagel M. (1996) Exploratory Analysis of Medical Mass Data. " *Computational Statistics* 11(4), 429-448.

Ousterhout J.K.(1994) Tcl and the Tk toolkit. Addison-Wesley, Reading MA.

Tomlin C.D. (1990) *Geographic Information Systems and Cartographic Modelling*. Prentice-Hall, Englewood Cliffs.

Tufte E.R. (1983) *The visual display of Quantitative Information.* Cheshire,Connecticut: Graphics Press.

Tufte E.R. (1990) *Envisioning information*. Cheshire,Connecticut: Graphics Press.

Tukey,J.W. (1977) *Exploratory Data Analysis.* Reading, Mass. Addison Wesley.

Unwin A., Hawkins G. Hofman H and Siegl B. (1996) Interactive graphics for data sets with missing values - MANET. *J.Computational and Graphical Statistics*,5,113-122.

Wise S.M. (1991) Software evaluation - lessons from the GIS evaluation. *University Computing* 13, 16-20.

Wise, S.M, R.P. Haining and J.Ma (1997) "Regionalisation tools for the exploratory spatial analysis of health data". In M.Fischer and A.Getis (Eds) *Recent Developments in Spatial Analysis: Spatial statistics, behavioural modelling and neuro-computing.* Berlin, Springer-Verlag p83-100.

\