Case Studies of Visualization in the Social Sciences:
An Introduction

David Unwin & Peter Fisher

Departments of Geography

Birkbeck College
7-15 Gresse St
London
W1P 1PY

University of Leciester
University Road
Leicester
LE1 7RH

Email: d.unwin@geog.bbk.ac.uk
pff1@le.ac.uk

1. Background

In the last few years Scientific Visualisation (ViSc) (as defined in the landmark report to the US National Science Foundation; McCormick *et al*., 1987) has become one of the standard investigative tools of all the physical and natural sciences (Hall, 1994). Visualisation in the sense used here is not a set of techniques to communicate the known as is done with traditional graphics and maps, but a means, a mechanism, for exploring data sets and structures, to generate ideas and explore alternatives. This has been defined as **'exploring data and information graphically, as a means of gaining understanding and insight into the data'** (Earnshaw and Wiseman, 1992: Brodlie *et al*., 1993), and is often referred to as *visual data analysis*.

Many factors, both practical and theoretical, have driven this change from traditional static displays or presentations to visualization:

- Developments in sensor technology and automated data capture have provided data at rates faster than can be easily converted into knowledge. These streams, or firehoses, of data are being added to the *warehouses* full of old data (McCormick *et al*., 1987). Such developments are appropriate to much social science data at, for example, the Essex Data Archive such as the Census and Opinion Polls.
- Some of the most exciting discoveries in science have been associated with non-linear dynamics where apparently simple equations like the finite difference form of the logistic $X_{t+1} = aX_t (1 - X_t)$ conceal enormously complex, but real-world like behaviour that can only be appreciated when displayed graphically (Lorenz, 1993).
- As science has progressed to produce ever more complex simulation models, so it has become necessary to use visualisation as the only practicable way to assimilate all the model outputs.
- Improvements in computing mean that we now have networks of very fast workstation computers available whose primary output is to high resolution, colour screens. In looking at these effects of technology it is easy to lose sight

of what is fast becoming commonplace. An obvious example is resolution and colour. In 1980 the Census Research Unit at Durham University published an Atlas of the 1971 UK Census of Population (Census Research Unit, 1980) which used the then new laser technology to produce, at great difficulty and expense, maps with individual colour symbolism for each and every kilometre grid square over Britain. At the time, these were the most detailed population maps at this scale and resolution ever produced, but using modern hardware similar displays are relatively easy to create at even higher spatial resolution using a full gamut of colours (Martin and Bracken, 1991).

This newer technology means that we can create entirely new forms of display which extend the number of available graphic variables beyond Bertin's original seven (plan, size, shape, value, orientation, hue, texture). These new visual variables include projection (Dorling, 1992, 1994), animation (Tobler, 1970; Moellering, H., 1976) and sound (Fisher, 1994). A major problem confronting the use of these new graphic variables is that, unlike conventional maps we know almost nothing about good design using them (Monmonier, 1991; Krygier, 1996). This leads to a potential for contrasting styles of use (Table 1).

Table 1
Comparison of Traditional and Computer Displays

TRADITIONAL GRAPHICS and MAPS	**COMPUTER DISPLAYS**
Use symbolism	Often aim for realism (VR)
Are selective	Try to use as much data as possible
Are end products	Are aids to understanding
Demonstrate the known	Detect the unknown
Intended for many viewers	Used by one person
Permanent	Temporary
Used many times	Used once
Restricted dimensions (x,y)	Multidimensional

ViSc is clearly in the tradition of exploratory data analysis in statistics (Tukey, 1977) in as much as it emphasises the use of graphics in the development of ideas, not, as in traditional graphics, in their presentation. Indeed, ViSc often turns the traditional process on its head by developing ideas graphically and then presenting them using non-graphic means.

McCormick *et al*. (1987) talk about visualisation as a method of analysis which 'transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations' (p3) and 'a tool for both interpreting image data fed into a computer, and for generating images for complex multidimensional data sets'.

2. Concerns

Against the background of widespread use of visualization in the natural sciences, there is perceived to be a relative lack of use of this approach in the social sciences and humanities, although outstanding examples of use can be shown. Interestingly, much of the current philosophical debate within social sciences would indicate that the approach of ViSc would be very welcome. Visualization involves close engagement with data rather than models of the data, while collection of the data necessarily involves abstraction of some kind that abstraction can be at many different levels.

Having identifed this potential use within Social Science, the JISC funded Advisory Group on Computer Graphics called a meeting at Burleigh Court Loughborough on 8/9th May 1997 which attempted to:

- Survey current work in social science making use of graphical computing and visualisation techniques;
- To evaluate the potential of the available technology to enhance teaching and research in social science;
- To explore the pictorial; data requirements of the social sciences;
- To make recommendations to AGOCG on the infrastructure necessary to support these developments.

Arising from that meeting a series of actions were suggested. One of these was to do with education and awareness, and a clear need for some readily accessible case studies of modern graphical computing in use in social sciences was identified. It was decided to address this need by a call for proposals for two Reviews and a number of Case Studies of how visualization is used in the Social Sciencees. These have been jointly funded by AGOCG and ESRC. This volume presents the Case Studies arising from that initiative.

The Objectives of the case studies are to:

- introduce the unfamiliar aspects of vsualization from two directions, by posing the problem of making use of visualisation
- give a readily available source of technical information on the software and hardware systems used.

3. Why are social science data different?

Data in the physical sciences result usually from controlled experiments or as model outputs. Investigators has choice on frequency, sampling, mode of measurement and domain covered. A good example is an atmospheric GCM of the type used to investigate global warming (REF). The output is a regular grid of values of , say, temperature, and the visualisation problem is straightforward. A number of fundamental differences are evident in the Social Sciences:

- In most work in ViSc in the physical and natural sciences, it is important to realise that the domains that are sampled are almost always assumed to be continuous. Very commonly in social science applications, although the domain is continuous it may be that our measurements are an enumeration from a discrete region which is a subset of this space. This is typical of census data and provides the social sciences with a major analytical difficulty referred to as the modifiable areal unit problem (MAUP).
- Typically, ViSC software systems also insist on regularly distributed point sampling of these continuous domains. Such data are common in many applications in natural and physical science, for example as the outputs from simulation models or from devices of the type used in medical imaging. Their use also implies that the data models used in ViSc are very simple, at least by compared with social science information and the standards of, for example, Geographical Information Systems. Census data are, of course, nothing like

this, though the use of interpolation and/or density estimation enables ‘surface models’ to be created (Martin, 1989).

- In ViSc all independent variables have the same potential importance. Using a ViSc, the domain may well refer to an abstract space given by two other variables about which the investigator has little *a priori* information, so that location within the domain is not a paramount concern, although patterning within it most certainly is. A major problem in using ViSc techniques in applications where the domain is real geography is to find ways by which the spatial scientist can be provided with locational clues to say where he or she is in the real world.
- Fundamental to much social science information is its strong and important location in time and space component. This is in common with some previous work in ViSc and particuarly in animation, but much remain to be done, and the methods used to be evaluated for the social sciences.
- Much social sciences information is qualitative, being based on questionnaire survey. Here measurement is an attempt to record subjective opinions, using traditional Lickert scales (such as Agree, Neutral, Disagree), binary answers and free text. Analytical devices for these are varied, but research in ViSc has not addressed any aspect of how these should be visualised.
- Social information is usually multivariate and necessarily so. This is because variables are often surrogates for concepts that have no single unique or readily agreed operational definition. An example would be social deprivation or class.
- Just as social concepts are hard to define, so too are the entities or objects of study. This leads to results being highly conditional on aggregation used and scale of analysis. Ecological fallacy is a well known if not understood.
- For many years the mixed mode of social science data has given problems in statistical analysis. Within one analytical framework data may be in nominal, ordinal, interval and ratio. This also presents visualisation problems.

4. The Case Studies

This report is made up of nine case studies. To assist the reader these are arranged into subject groups.

- The first of these is enumerated census information. It is noteworthy that none of these authros are using standard packages for their work. Three case studies are inclued:

- Steven Wise and others **review four different software packages which are all designed for the visualization of such data with variable statistical functionality. The packages are public domain or semi-public, but they find that they are of variable usability with differing support.**
- Jason Dykes and David Unwin **present one of the programs reviewed by Wise and others.**
- **Finally,** Chris Brunsdon and others **present some of their work on manipulating, analysing and visualizing multidimensional census data.**

- The second group of two papers examine the instance of chronological and categorical data:
  - Brian Francis and John Pritchard **present a new view of a traditional graphic device (although one which is little used).**
  - Humphrey Southall and Ben White **examine visualization of life-history trajectories.**
- Virtual Reality is a fast developing area of computer science and a number of social scientists are exploring its use and implicaitons:
  - Paul Carey and others **explain how the data sets required for Virtual environments can be developed quickly and efficiently using modern digital photogrammetric workstations.**
  - Andy Smith and others **explore developments of Virtual Cities both real and imagined.**
- Finally, an major features of modern computing, indeed the modern world, is the development of the World Wide Web, which (as well as highly publicised uses) has a potential for serious applications:
  - Ralph Schroeder and Ray Lee **examine the use of the Web (and Virtual Reality) in training the administers of questionnaires.**
  - Steve Carver and co-workers **review the use of the Web for conducting questionnaires.**

Acknowledgement

All contributors to these case studies wish to express their appreciation to ESRC and AGOCG for funding. More especially, we wish to thank Anne Mumford who, on behalf of AGOCG, facilitated and oversaw the initiative, and Sue Cunningham, who mounted the collection on the Web and oversaw printing.

References

Brodlie, K.W., Carpenter, L.A., Earnshaw, R.A., Gallop, J.R., Hubbold, R.J., Mumford, A.M., Osland, C., and Quarendon, P. (1993) *Scientific visualization techniques and applications*. Berlin : Springer-Verlag.

Census Research Unit (1980) *People in Britain - a census atlas*. London : HMSO.

Dorling, D.(1992) Visualizing people in time and space. *Environment and Planning B: Planning and Design* 19, 613-637.

Dorling, D. (1994) Cartograms for visualizing human geography. pp.85- 102 Hearnshaw, H.H. and Unwin, D.J. (eds) *Visualization in geographical information systems*. Chichester : John Wiley & Sons.

Earnshaw, R.A. and Wiseman, N. (eds) (1992). *An introductory guide to scientific visualization*.,Berlin : Springer-Verlag.

Fisher, P.F. (1994) Animation and sound for the visualization of uncertain spatial information. pp.181-185 Hearnshaw, H.H. and Unwin, D.J. (eds) *Visualization in geographical information systems*. Chichester : John Wiley & Sons.

Hall, S.S. (1994) *Mapping the next millennium: the discovery of new geographies*. New York : Random House.

Krygier, J.B.(1996) Sound variables, sound maps and cartographic visualisation. In MacEachren, A.M. and Taylor, D.R.F. *Progress in cartographic visualization*. London : Pergamon, London.

Lorenz, E.N.(1993) *The essence of chaos*. London: UCL Press.

McCormick, B., DeFanti, T.A. and Brown, M.D. (1987) *Visualisation in scientific computing*. *ACM, SIGGRAPH Computer Graphics* 1987, vol 21(6), special issue

Martin, D. and Bracken, I. (1991) Techniques for modelling population-related raster databases. *Environment and Planning A*. 23, 1069-1075.

Moellering, H. (1976) The potential uses of a computer animated film in the analysis of geographical patterns of traffic crashes. *Accident Analysis and Prevention*. 8, 215-227.

Monmonier, M. (1991) *How to lie with maps*. Chicago : University of Chicago Press.

Tobler, W.R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234-240.

Tukey, J.W. (1977) *Exploratory data analysis*. Reading : Addison-Wesley.