This paper describes the conceptual and practical problems encountered and solved in producing a multi-colour atlas of population characteristics in Great Britain. The atlas itself is in A4 format; it consists of some thirty-four maps of Great Britain in four colours and the same number of regional maps, together with descriptive text. All maps were plotted on a laser plotter with a resolution of 127 microns.

The paper describes how mapping of ratios, such as percentages, was found to be highly misleading and describes the novel probability mapping solution adopted, based on the signed chi-square statistic. In addition, the rationale for selecting the class intervals and for selecting colour schemes is described.

Making a National Atlas of Population by Computer*

D. W. Rhind, I. S. Evans and M. Visvalingam

Census Research Unit,** Department of Geography, University of Durham

INTRODUCTION

The decennial Census of Population is possibly the most widely used single data set in Britain; it plays a fundamental rôle in the analysis and monitoring of changes in the distribution of population, of the characteristics of households and of such socio-economic variables as journey-to-work patterns and migration. It is also needed for standardisation of many other data sets, such as mortality statistics. Despite this importance, there has been no routine mapping of British census data. This paper describes the design and production of an atlas based on census data for 1971 and intended for use by the 'intelligent layman'. Though all censuses are important, that of 1971 was, in some respects, unprecedented. More detailed both in terms of geography and of questions asked than any of its precursors, it produced the first major data set in the UK to be available for regularly shaped areas, in addition to the customary and frequently changing irregularly shaped and sized enumeration districts, wards and local authority areas. All of the small area data were available in computer-compatible form. Detailed descriptions of the data have been published (*e.g.* OPCS 1978; Rhind, Evans and Dewdney, 1977) but data characteristics are summarised here since, as in any major mapping project, they condition the mapping response.

DATA CHARACTERISTICS

The data set obtained by the CRU from OPCS covered approximately 40 magnetic tapes, and a sizeable proportion of the early work consisted of converting these to the format of the local computer and checking the integrity of the results. An experimental atlas for the 80 km by 65 km area of County Durham was published in 1975 (Dewdney and Rhind, 1975) but the last of the national data were not made available until September 1976. The first multi-colour national maps from these were plotted within a few weeks and published ten weeks after receipt of the data in the December 1976 *Geographical Magazine* (Rhind, Visvalingam, Perry and Evans, 1976).

All of these census data pertained to, and were referenced only by, 1 km square areas of the British National Grid; all 18 million households in the country were allocated an approximate grid reference at the survey stage and the grid square data were produced by OPCS sorting the responses into these areas. In principle, OPCS had the capability to sort the data into squares of any resolution although confidentiality and accuracy of grid referencing considerations have shown that 1 km squares were probably the smallest areas suitable for covering the whole country.

Data for regular and consistently-sized areas such as grid squares have two disadvantages by comparison with the more usual data related to irregular zones: they are unfamiliar and, if interpreted strictly, are unrelated to functional entities in the real world. Thus they dissect apartment blocks, roads and homogeneous areas. By contrast,

* This paper was originally given at the Ninth Internal Cartographic Association Conference in Maryland, USA in July 1978. It was delayed pending the printing of illustrative materials and was modified while one author (D. W. R.) was on sabbatical leave.

** The Census Research Unit (CRU) has been funded by the Social Science Research Council to work on statistical and cartographic aspects of the grid square census data. The necessary data were kindly provided by the Office of Population Censuses and Surveys (OPCS). At the time of writing, the CRU comprised the three authors of this paper and Professor J. I. Clarke, Mr J. C. Dewdney and Mrs J. Coulter.

their advantages are many. They are easy to map, since the sizes of symbols can be chosen beforehand to avoid overlap. All areas (except around the coast and inland lakes) contain the same area of land, so counts are automatically density measures and absolute numbers can be mapped directly. The areas are unchanged over time and hence facilitate the examination of trends through time. In giving constant rather than the more usual varying spatial resolution, they are directly analogous to the week, year or decade which are used for the analysis of variation over time, in preference to non-constant units such as the length of a King's reign or of a presidency. Finally, the areas are simple to aggregate into larger areas which may either be square or be areas of irregular shape, approximating, for example, to administrative areas.

For each of the 1 km grid squares, some 1571 items of information—such as the number of married females in that area between 45 and 49 years old—are (in theory) available. Thus the basic data set is a matrix of 152 440 grid squares by 1571 items of information, but this is incomplete because many items are suppressed in squares with low populations, so as to preserve confidentiality. Inevitably, such a sizeable data set, originally nearly one billion bytes* in size, had to be compressed before it was usable (Visvalingam and Rhind, 1976). In addition, the fixed size of data units inevitably meant that the population frequency distribution was highly skewed—most 1 km squares contained few people while a small number contained up to 24 000 people, the mean value being just over 300 people. Many of the raw variables were stored in various different forms—as counts or as ratios (such as 'per 1000 resident population')—and, because of the way they were defined, were related to each other in sometimes complex fashions. Inter-relating different variables was complicated by the data being stored as four records per area, in two pairs, each pair on separate magnetic tapes.

Such complications, however, were not as significant as two effects resulting from confidentiality considerations. The first is the suppression of data where there are few people or households: in the population data record, for instance, only 3 out of 471 items of information were provided unless the population was greater than 24 people. Only some 68 000 1 km squares contained all population data as a result of this suppression. In addition, all data values given for unsuppressed areas had a value of -1, Ø or 1 randomly added to them, except in those circumstances where the result would have become negative. As has been demonstrated elsewhere (Rhind, Evans and Dewdney, 1977) this can have the effect of giving highly spurious and misleading ratio values which, in certain cases, may become infinitely large. Checks were applied to the data to ensure this had not occurred and impossible values were corrected. Furthermore, some of the data were not based upon 100 *per cent* enumeration but upon a 10 *per cent* sample; they were thus of lower reliability and available only for 54 000 squares.

MAPPING CENSUS DATA

In the past, mapping of the UK census data has been carried out either by hand (*e.g.* Hunt, 1968) or by the relatively crude computer lineprinter (*e.g.* Rosing and Wood, 1971).

* A 'byte' is IBM computer parlance for eight binary digits (bits). The standard storage unit on most IBM (and many other) computers consists of four bytes.

More sophisticated computer output devices have been used for limited areas such as counties (*e.g.* Coates (1974), Dewdney and Rhind (1975), Shepherd *et al.* (1974)), or at a very generalised level of geographical resolution (*e.g.* Department of the Environment, 1977). Coarse resolution, normally based upon administrative areas, inevitably results in misleading mapping except in the unusual situation where the mapped variable changes in value precisely at the boundary of the data zones such as counties or states: the result is a pernicious form of filtering which is difficult to interpret because of the varying size, shape and orientation of the areal units for which the data are expressed. Visually, choropleth maps of such areas are dominated by the larger sub-divisions, which often contain a very small proportion of the population.

While comparatively expensive in machine time, using data for the 1 km grid squares homogenises smoothing of the data. Thus the problems listed above and the need to produce compact, high-quality output in atlas form within an acceptable time scale (see below) precipitated the following major cartographic consequences.

(i) The problem of selecting which of the 1571 variables (or their permutations) to map.

(ii) The selection of the areas and the resolution of the data to be mapped.

(iii) The need to produce symbolism which did not obscure unpopulated areas, did not overlap and which bore obvious relationship to the areal units mapped.

(iv) The problem of representing the comparative unreliability of data in small-population areas and particularly for variables based upon 10 *per cent* sample data.

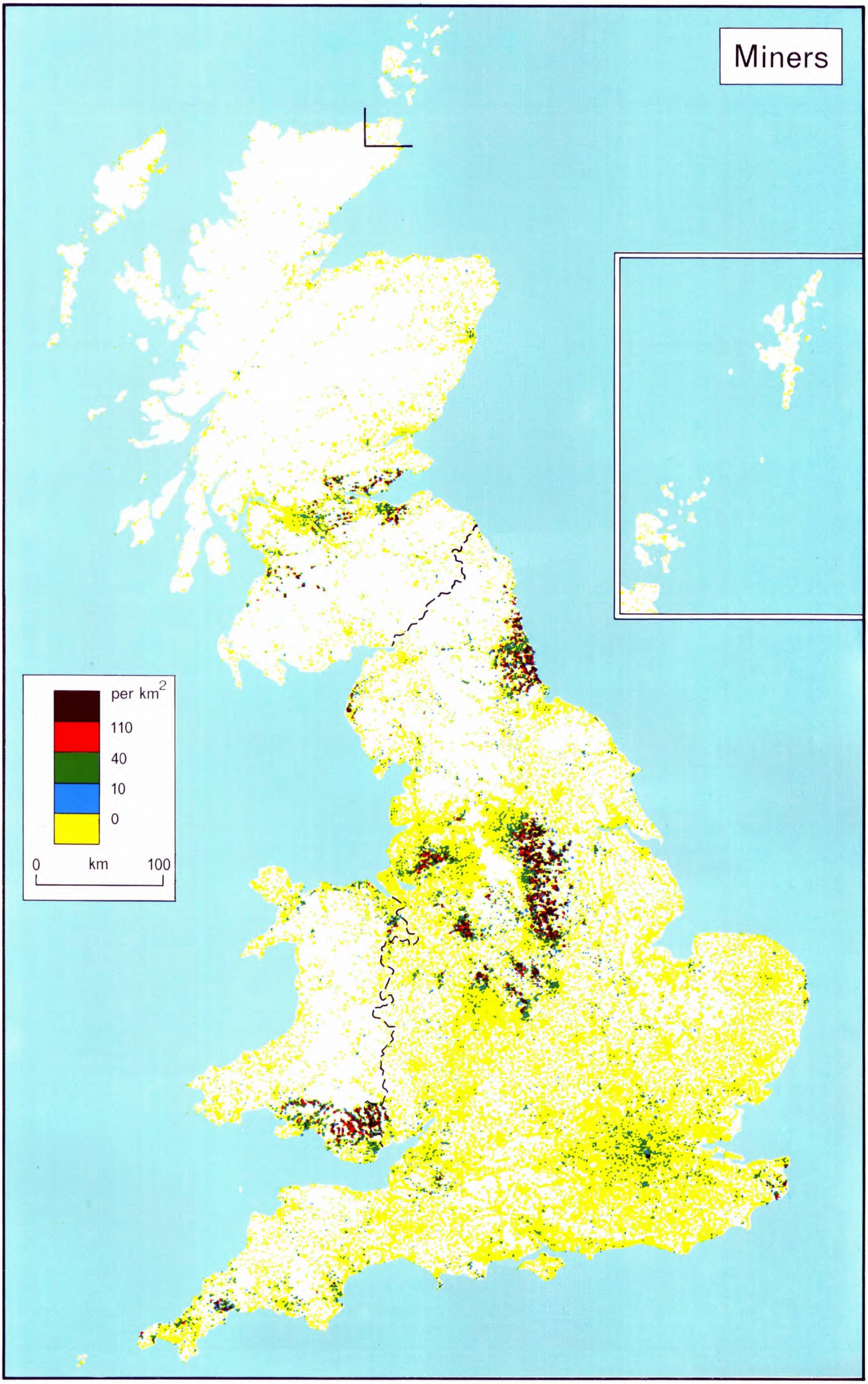
(v) The selection of the class intervals for the data.

(vi) The overall layout of the atlas maps and accompanying text, including page format.

(vii) The selection of the colour schemes to be used.

(viii) The selection of the technology for mapping up to 150 000 zones on each A4 size sheet of paper, involving considerations of computational efficiency, availability, reliability, resolution, repeatability, cost, final reproduction and related matters.

The end result of decisions made after consideration of the above is an atlas, titled 'People in Britain—a Census Atlas' being published by Her Majesty's Stationary Office and authored jointly by CRU and OPCS. It contains 34 maps of all of Great Britain (see Appendix 1) at approximately 1 : 4 million scale, and four maps for each of seven conurbation-centred regions at 1 million scale. Almost all of the maps are reproduced in three process colours plus black and all the maps are described in accompanying text. Appendices in the atlas give the detailed statistical background. Selections from these maps and a map of population density originally at 1 : 1 million scale were used by OPCS to produce a wall chart for schools, illustrating the work and benefits of the population census. The larger scale was produced by slightly different means (such as the plotting of class, rather than colour, separations): its production is not discussed further. The rest of this paper describes the procedures utilised to cope with the topics listed above.


Miners
per km^2
110
40
10
0
0
km
100

MAKING THE NATIONAL ATLAS

(i) *Selecting variables for mapping*

The traditional method of selecting census variables is to utilise those which have been used in the past (thus permitting historical comparisons), those which have broad substantive interest (such as the numerical dominance of old people in the community), or those which have specific policy implications (such as households without basic amenities). The first stage of the CRU selection procedure followed this practice and 102 variables, almost all of which were percentages, were derived by selection and aggregation from the raw census data and stored as one record per 1 km grid square. In addition, the corresponding 121 numerators and denominators used to produce these ratios were also stored; these played a vital role in later mapping.

All of these variables were mapped in one way or another but the final selection of variables for the atlas was based upon three criteria:

—that the variables mapped did not possess high positive or very high negative intercorrelations;

—that the variables were of substantial general interest;

—that the visible, mapped patterns were significantly different from those in other maps.

In addition, two special items—a multivariate map showing the distribution of ages of the population in each 1 km square and a block diagram of the density of population—were introduced because of their illustration of alternative ways to view the data.

The correlation analysis, described in Evans (1979), was a central part of the statistical analysis in the project and involved producing Pearson product-moment correlation statistics for ratio values, weighted by population and previously transformed to minimise skewness in the univariate frequency distributions. As a consequence of this, large numbers of variables could be eliminated from consideration; any visible differences between a map of the distribution of people and another of the distribution of households in Great Britain, for example, would be due solely to the class intervals chosen—the two variables are correlated as +0.997.

(ii) *Selecting the areas to be mapped, the level of resolution and the format*

Several considerations impinged on the choice of which areas to map and the levels of resolution at which the data were to be mapped (since the data are for square areas, aggregation to larger units is very simple). The general desire to show as much geographical detail as possible in order to overcome the filtering effects of larger areas had to be tempered by the format of the atlas, the available technology and the perception aspects: planning to make maps of the whole country at 1 km resolution would have been futile if this level of detail could not be printed and interpreted. Early experiments suggested such resolution could be achieved and perceived, hence all of the maps in the atlas except for one (a 10 km square map of population for comparison of the resolution levels) were based upon data for the 1 km areas. It must be emphasised, however, that variations in quality of plotting on film, of proofing and of final reproduction (particularly of registration), meant that the appropriateness of this decision could only be vindicated after the final production of the atlas. The concepts and methods of production of this atlas pushed the then available technology to its limits.

It was inevitable that a national atlas should include a substantial number of maps of the whole country, as covered by the data; no data were available for Northern Ireland, the Isle of Man and the Channel Islands, so these do not appear except on location maps. As produced, such national maps permit the comparison of regional differences with local variations; however, because of their scale and detail, no base other than a coastline and one separate map with place names could be provided. As a result of this difficulty of local identification, and because some variables are of urban rather than national interest, additional maps at larger scale were included in the atlas. The seven regions which these maps cover are conurbation-centred and thus include 85 *per cent* of Great Britain's population, even though they cover only about half of the land area.

Considerable disagreement was found over the best format. Ultimately the standard A4 size (210×297 mm) was selected—for convenience, and because Great Britain (with minor modifications) fitted neatly on a page without photographic reduction. Because of the type of plotter used (see below) the actual scales available for maps without photographic processing were approximately 1 to 8 million, 4 million, 2.6 million, 2 million, *etc.* It was considered vital to avoid photographic reduction of the computer output for two reasons: it would lead to extra stages in map production, and hence extra costs, and it could well lead to filling-in of the extremely intricate detail. Therefore, reproduction material was prepared directly on the computer plotter as wrong-reading positives, with the exception of the black-plate material (base-map and key).

(iii) *Selecting the symbolism*

Choice of symbolism to represent values for each grid square was one of the simpler decisions: it had to be contained within the square, to avert any overlaps; it had to be present for all mapped values so that missing data (from data suppression or unpopulated areas) could be left blank; it had to relate closely to the data areas; and it had to be capable of discrimination in at least four and possibly five states when the size of a 1 km grid square on paper was only 0.25 mm. Such considerations ensured that variable size and density squares (the size indicating the base population and the density or colour the mapped ratio value), or even monochrome density maps would be unclear unless resolution was sacrificed by aggregation of data areas. The obvious solution was coloured squares of fixed size.

In practice, the symbolism is not exactly of this form. It is compiled by building a matrix of dots produced on film by a laser beam; thus a symbol on the national maps is produced by exposing four dots, two across, two down, each dot being nominally offset from the others by five thousandths of an inch (127 μm). However, each dot is circular in shape and the laser light distribution across the dot is Gaussian. Overlap occurs along the rows and down the columns of dots because their effective diameter (where the laser light is at least at half intensity) is some 180 μm to fill in corners of the squares. In addition, the resulting symbol usually has 'soft edges'. These limitations have not been found to be serious ones except where over-exposure occurs or where the laser beam has not been sharply focussed; the former situation results in infilling of small, enclosed areas on the original positive.

(iv) *Representation of varying reliability*

In mapping any data for small areas, particularly those which are obtained on a sample basis, considerable attention must be given to variations in the likely validity of the data. The larger is the base population (the denominator in a ratio such as percentage of working-age population unemployed), the more reliable is the resulting ratio, yet the less likely is it to fall in a non-average class. All of the 102 variables described above were originally plotted as monochrome ratio maps on film and careful examination of these showed that, for most, the dominating areas of very high and very low values were those in which very small populations occurred. The choice of class intervals to accommodate such extremes, based often on only a few people, was such as to mask very substantial and important variations elsewhere in the country; the resulting maps clearly emphasised rural areas at the expense of urban ones.

Several possible solutions for this dilemma were investigated. The first, and most obvious, was to aggregate the grid squares to give larger and thus more stable populations. The difficulty here is that the population frequency distribution is autocorrelated and highly skewed, so that very considerable aggregation had to be carried out before the problem abated. Even with 5×5 km squares, many contained only two people (Visvalingam and Dewdney, 1977). In addition, aggregation of data—some of which were previously suppressed by OPCS—meant that any larger area would be characterised only by the unsuppressed areas within it. An extension of this argument would be to omit all squares with populations below a specified threshold, mapping only the remainder. While this would have been statistically satisfactory, different thresholds would have been necessary for different variables—complicating the layman's interpretation of the maps—and the maps would have been extremely 'spotty'.

Yet another alternative was to map absolute numbers in the particular category, such as 'numbers of one parent families' and leave any interpretation to the user. To know whether the presence of 100 Irish in a given area was unusual or merely reflected a large number of people of all nationalities, the user would have to compare the map of numbers of Irish with a corresponding one of the total population in each grid square. Such maps emphasise urban areas in comparison with rural ones, since the low populations in the latter ensure that all rural squares fall in the lower classes.

After considerable experiment, two types of map were included in the atlas. Seven national maps and one reference map of total population for each of the regions are absolute number (or density) maps. The variables thus shown are those, such as numbers of farmers and of miners, whose spatial patterns vary greatly from that of total population. All other maps are related to the national average figure and are based upon mapping a signed chi-square value for that variable in that square (Visvalingam, 1976; Visvalingam and Dewdney, 1977), using the tape of numerators and denominators mentioned in (i) above.

The signed chi-square statistic is a compromise measure, taking into account both the absolute deviation (in number or persons, households, *etc.*) from the national average, and the relative deviation (effectively the proportionate difference from expectation). It is computed as

$$X^2 = \Sigma[(O-E)^2/E]$$

where O is the observed number of individuals, E is the expected number based upon the national average, and X^2 is the chi-square value, summed over two classes. It is then given a negative sign if the expected number (derived by multiplying the national average by the base population of the area) is greater than the observed number.

The effect of this computation is that a high-population square will obtain a larger chi-square statistic than will a low-population square which has the same percentage difference from the national average. In other words, a weighting is applied such that small population squares must be very unusual indeed before they are given a high chi-square value. *Figure 1* illustrates the way in which variations in base population and in the proportion of interest interact in calculation of the chi-square statistic and, subsequently, in assignation of map class and colour (see section (vi)).

It is important to note that, like all mapping schemes, this one contains certain arbitrary elements. The conventional squaring of (O–E), for example, homogenises the variance for binomial variables only. Our justification for the method is that it certainly steers a middle course between emphasising high or low population areas, and it does not lead to undue loss of information. That the results are more readily interpretable than those of either ratio or absolute-number based maps is, of course, a further benefit.

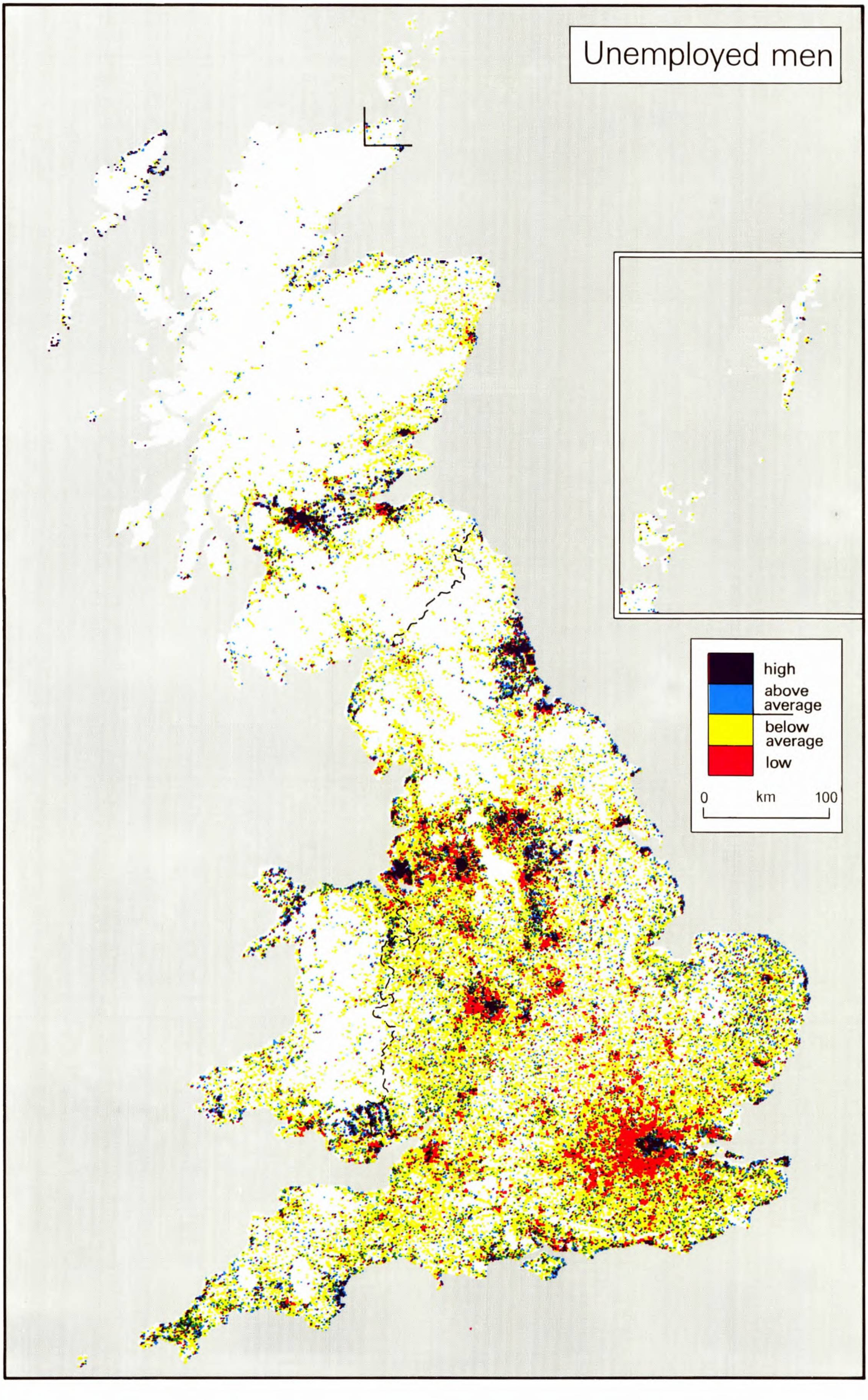
(v) *Choosing the class intervals*

Because of the extremely small size of each data area on the national maps, the limited grey tone capability of the particular map-making technology adopted (see (viii) below) and the lithographic means of final reproduction, the desirable objective of obviating class intervals (Tobler, 1973) was not feasible in these atlas maps. On this basis, various class interval schemes were considered, as reviewed by Evans (1977).

The class interval schemes adopted inevitably differ between those used for the 'absolute number' or density maps and that for the signed chi-square maps. Frequency distributions of absolute numbers were always positively skewed: hence geometric progressions of class width with four or five classes were used. (An arithmetic basis would have placed very few squares indeed in the upper classes.) The class limits were chosen after consultation of histograms for each mapped variable, and rounded off. The keys for all absolute number maps were labelled with the upper inclusive class limits used.

Maps based on the signed chi-square statistic are rather different since the measurement scale is centred on the national average figure, the two extremes away from this being those of most substantive interest in any particular map. Thus choosing the value of zero—the national average—as the mid-point on the signed chi-square scale and opting for two classes above and two below the national average enabled these maps to be viewed in two ways. On a simple dichotomy basis, all areas above the national average and all those below could be distinguished in a fashion identical to that on any ratio map. More usefully, perhaps, the notably divergent areas could be observed with ease and those near the average ignored. To be complete, the solution required only two more decisions: the cut-off points for the extreme classes, and a description of the meaning of these classes suitable for the non-statistician.

After much discussion, the obvious possibility of mapping the n most extreme grid squares (where n might



Unemployed men
high
above average
below average
low
0
km
100

be, say, 500) as being in each extreme class was abandoned. To map a fixed number in each extreme category is to ignore the differing variability in the different mapped variables: some have low standard deviations, others have enormous ones. The solution adopted was to use the —3.84 and 3.84 values as the inner limits of the two extreme classes. These values happen to be the tabulated 95 *per cent* significance levels for chi-square with one degree of freedom, but with spatial data of this non-binomial nature we do not wish to press an interpretation based on statistical significance. An appendix to the atlas gives a fuller description of this rationale and tabulates the number of squares in each class for each map.

Perhaps the most intransigent problem was that of the labels for the chi-square map keys. Since labelling these with the —3.84, 0 and 3.84 values would have communicated nothing without accompanying details of the underlying units of measurement, the use of qualitative labels was unavoidable. Descriptions for the four classes which were statistically appropriate, such as 'great surplus/surplus/deficit/great deficit', were usually unsatisfactory either from the view point of lay understanding or, as in this case, for the possible emotional impact of such terms. Finally, the simple terms 'very high/high//low/very low' were adopted as being understandable and relatively value-free; strictly, the accurate use of these is based upon the understanding that 'very high' is above the national average and involves a reasonable number of cases.

(vi) *The selection of colour schemes*

It has already been emphasised that the minuteness of the 1 km grid squares on the 1 : 4 million scale national maps conditioned the overall design of the atlas and ensured that the use of colour was essential. In practice, the restriction was even more severe: because of the size of each data area in the national maps, no screening of a resolution suitable for the anticipated printing could be used and thus the colours used had to be solid ones. Printing costs limited the number of colours to four and those used were the three process colours and black.

The regional maps, at 1 : 1 million scale, presented different design problems, notably of what to do with squares which overlapped coastlines: ultimately these were left untouched since they give visual evidence of the spatial averaging involved. In these maps, however, the squares were large enough (1 mm) to consider screening and experiments were successful; solid colours were used in the final maps because they were believed to give more obvious distinctions between classes, as well as compatability with the national maps.

The colour schemes which could be produced from the permutations of these colours are obviously limited. Table 1 gives those finally selected. In the absolute number maps, the overall concept was to match increasing density of colour (related partially to the number of over-prints) to increasing mapped value and to ensure that any mis-interpretations (*e.g.* green for blue) were, where possible, only between adjacent classes. In the signed chi-square maps, the two extremes were emphasised and paler colours as closely related as possible to these were used for the innermost classes.

(vii) *Layout of maps and text*

Given the use of A4 paper, the scale of the national maps in the atlas is automatically constrained to 1 : 4 million;

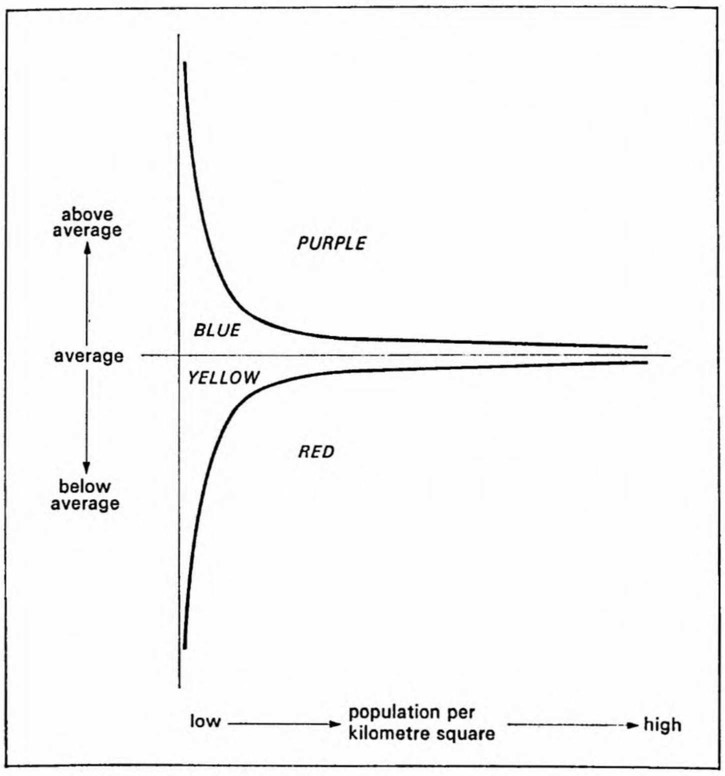


Figure 1.

spreading the maps across two pages was not a viable proposition since areas of considerable interest would have lain in the gutter. Thus the national maps are arranged 'portrait' on each page, each with a facing text. The Shetland Islands were moved by computer program to an inset, chosen sufficiently far south that any misunderstanding of their position was believed impossible—the extreme parts of the mainland and part of Orkneys appear on both main map and inset. The key for each map deliberately differs in position (as well as form), depending on whether it is an absolute number one or a chi-square map. A pale tint was added to the sea to differentiate it from unpopulated or suppressed land squares.

The regional maps are 'landscape', stacked two to a page and with the four maps for each region facing each other and separated from the previous region by a key map of place names, administrative boundaries, *etc.*, and descriptive text. Because of constraints of space, the two keys for all of the regional maps had to be placed in an introductory section. For each region (see Appendix 1) there are three chi-square maps and one map of total population: the usual purple/blue/yellow/red colour scheme for the former served to distinguish them from the yellow, blue, green, orange and brown of the latter.

(viii) *The map-making technology*

The decisions and *desiderata* described above made very onerous demands upon the available technology. In short, this had to be capable of processing very substantial amounts of data and creating maps within the resources available in the university computer centre. The plotting devices had to be capable of extremely high resolution and repeatability (to ensure correct registration of colour separations), at high speed and at low cost. Since the data were originally on a grid and since the symbols chosen were square in shape, it was logical to use a raster rather than a vector plotter for the colour separations.

In the event, not all those ideal conditions were obtained. The processing was carried out on the Northumbrian

ORGANISATION OF THE LASER MAP PROGRAM SUITE

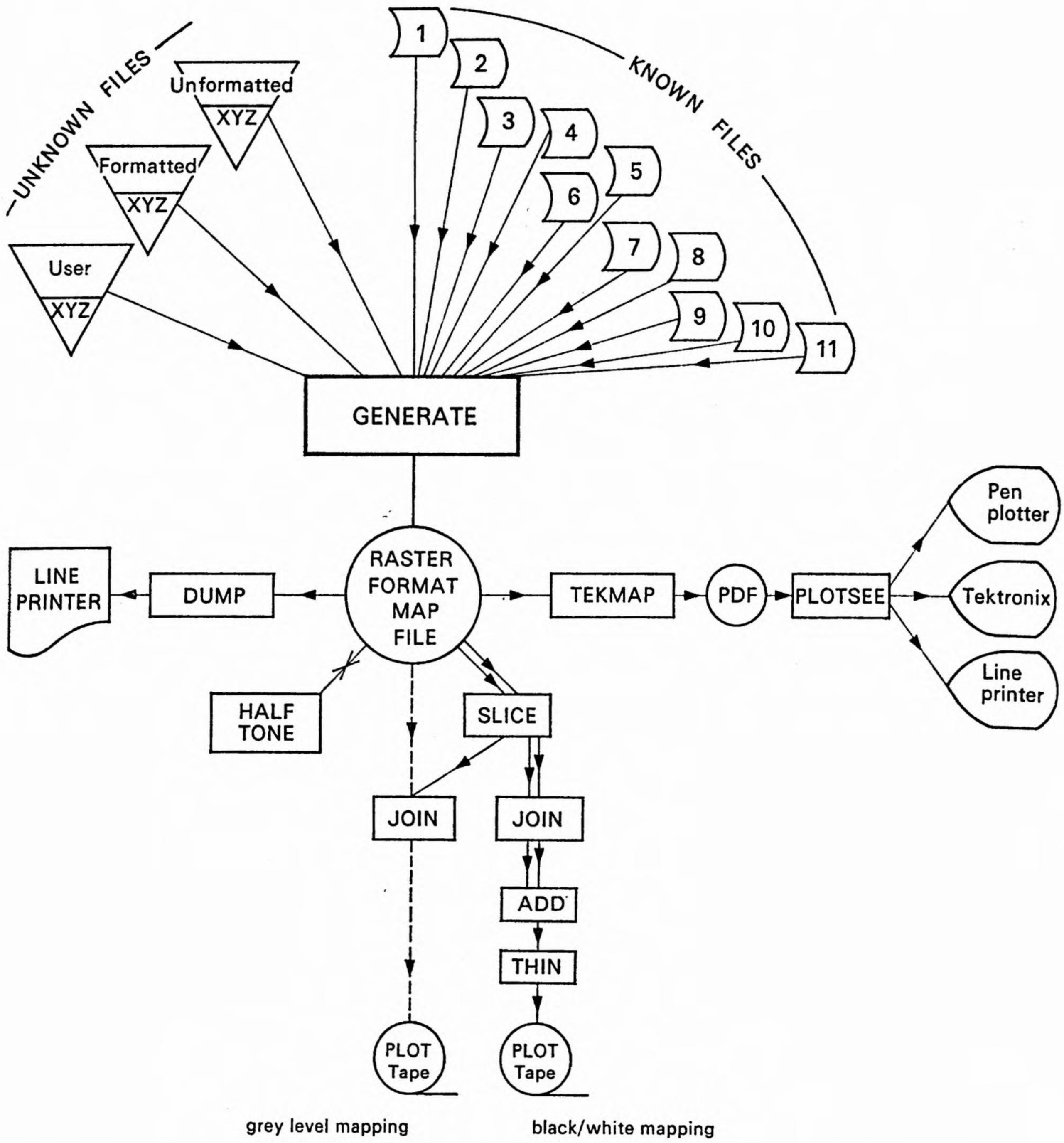
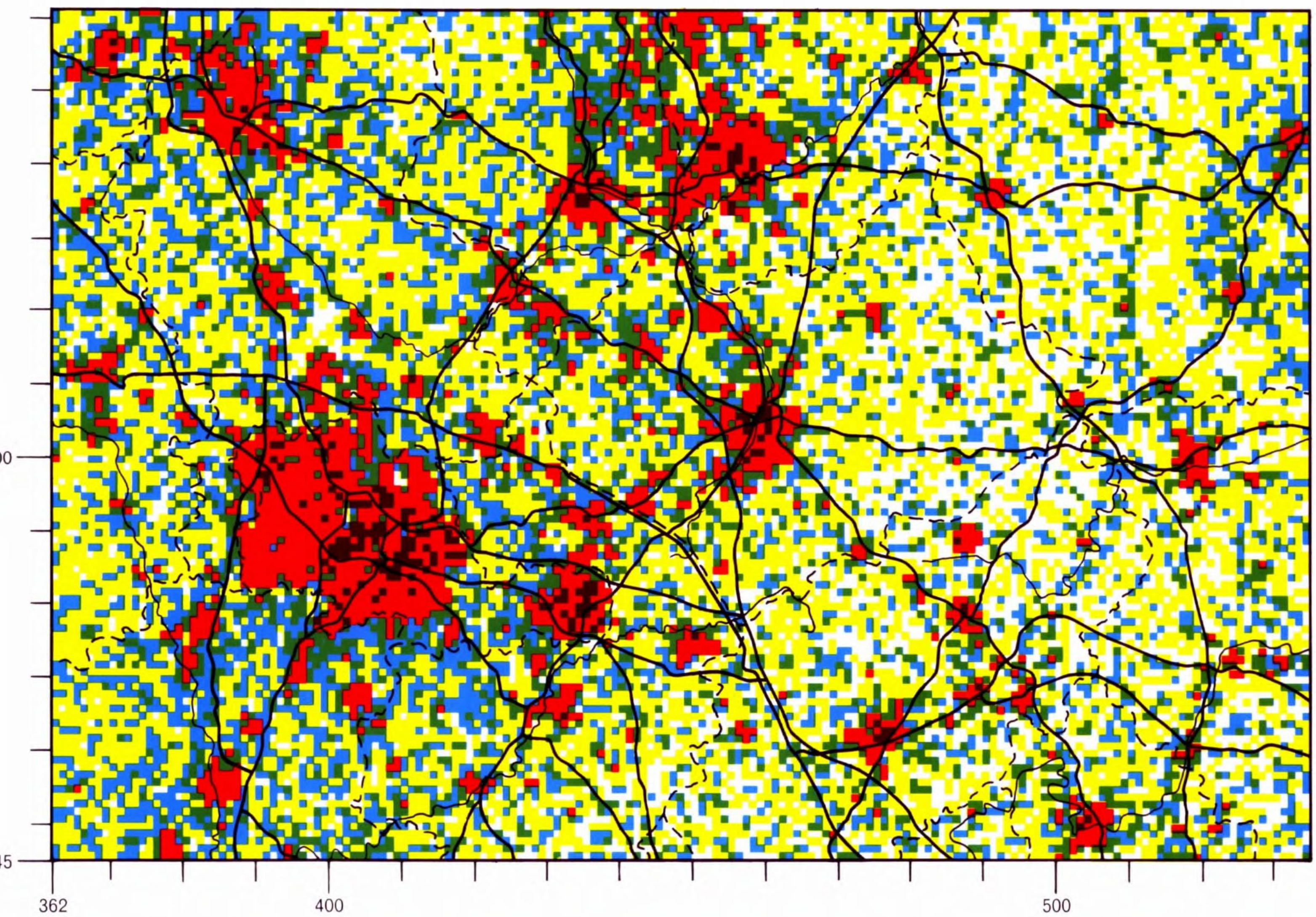


Figure 2.

Universities Multi-Access Computer System (NUMAC), shared by the Universities of Durham and Newcastle; the actual processing was carried out on an IBM 370/168 computer running under the Michigan Terminal system (MTS). The overall structure of the LASERMAP suite of computer programs used to produce the maps is set out in *Figure 2*. Programs in this suite may be run either in batch mode or from a terminal; both were used in practice.

All input data were passed through the GENERATE program which created a raster array of classified map values, typically 1300 by 700 in size to cover the whole country at 1 km resolution. GENERATE has the ability to recognise all raw census data files and previously derived files such as those of ratios and counts, or a standard-format inverted file ('known files'). For most users, no knowledge of where or how such data are stored is necessary since the program calls the appropriate disks and files and reads the data as appropriate. The program also has the ability to read most other point- or grid-based data files although this necessitates supplying it with additional information on their characteristics. The parameters for each map (such as minimum and maximum X and Y co-ordinates, minimum and maximum mapped values, the data measurement scale, and so on) are either supplied by the user to GENERATE, calculated by the program or defaulted. These are stored in a map header. All subsequent programs—to slice out values to be printed in a particular colour (SLICE), to shift the Shetlands to an inset (ADD), to map on a Tektronix 4014 cathode ray tube, pen plotter or line printer for a check plot (PLOTSEE), to assemble maps economically on the 40 inch wide film used in the plotter (JOIN), and so on—required virtually no supplied parameters because each program first read this header and checked the feasibility of the operations

Population density



Born outside the U. K.

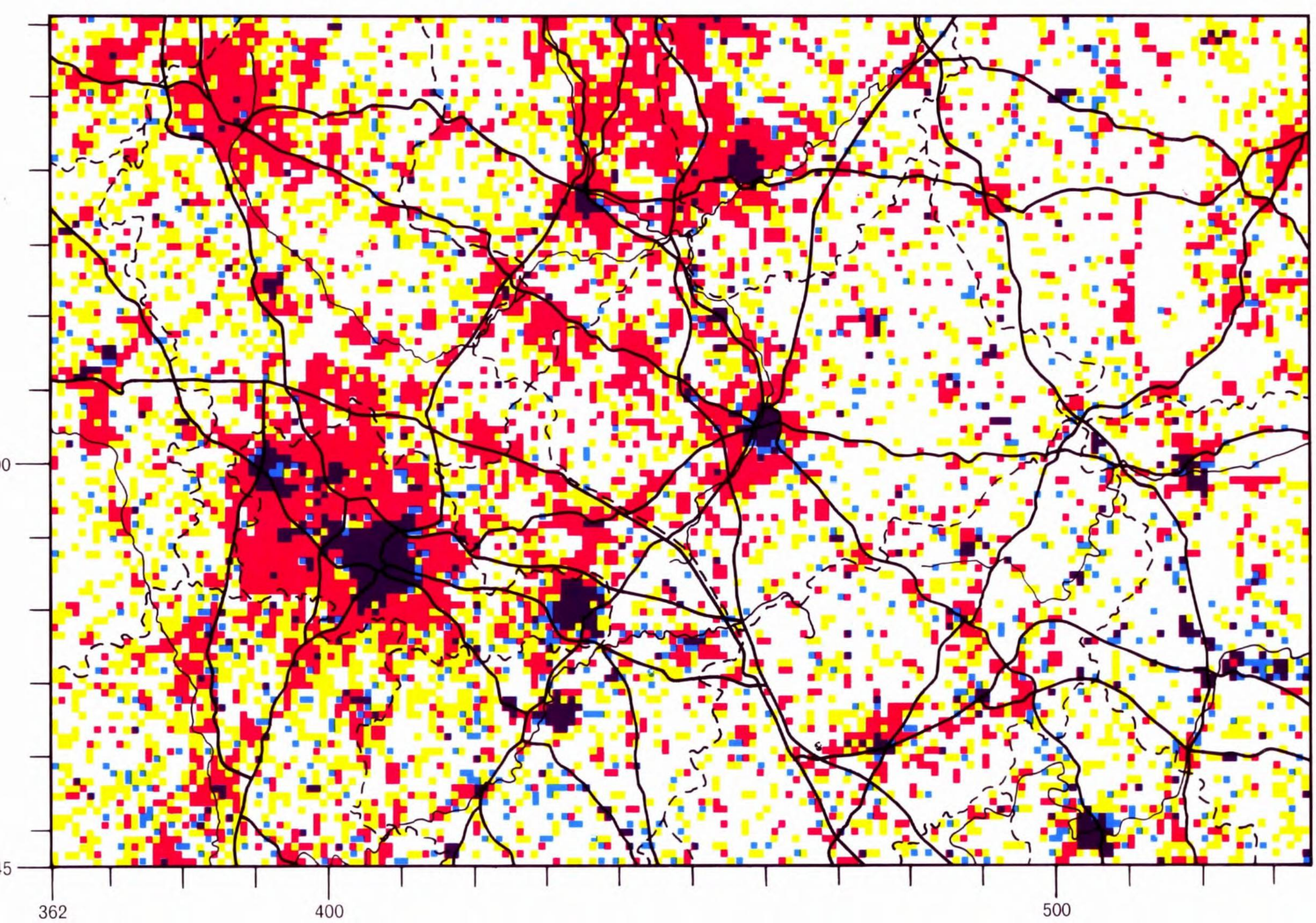


TABLE 1

Map colour schemes

	Absolute number maps 4 class	5 class	Signed chi-square maps Class		
High		Brown (red and blue and yellow)	5		
	Brown (red and blue and yellow)	Orange (red and yellow)	4	Purple (red and blue)	Very high
	Orange (red and yellow)	Green (blue and yellow)	3	Blue	High
	Blue	Blue	2	Yellow	Low
Low	Yellow	Yellow	1	Red	Very high

Note: 'no data' areas are shown as white.

requested: any attempt, for example, to ADD together maps of different variables would have resulted in an error message.

None of the programs required substantial machine resources. Since all the input (X, Y, Z) data were previously sorted from north to south, all the programs could operate in a 'push broom fashion', never requiring more than one 1 km deep row of data in core at any one time. All of the programs were written in IBM FORTRAN IV, compiled under the H level optimising compiler, but made use of system-specific input/output routines to speed data transfers by a factor of ten times over the standard FORTRAN unformatted input/output. The cost of using GENERATE to map the whole country by such means varied according to the location and blocking of the data but was, in some instances, as little as £2 at internal rates, based on about 20 seconds of central processor and 5 to 15 minutes of elapsed time. Other operations necessary to make the final maps approximately doubled this computing cost. It should be noted that such costs do not take into account the amortisation of software or commercial charging for computer time—in NUMAC, full commercial charges were three times the internal rate at the time of writing.

The plotting was carried out on a bureau laser plotter manufactured for and run by Seiscom Delta Ltd. This plotter, described in Rhind (1974), plots on a raster basis with a basic step size of 5 thousandths of an inch (*c.* 0.127 mm) direct on stable base film 40 inches wide. The impossibility of varying this width, difficulties of communicating over 500 km via magnetic tape and post, and cost-consciousness dictated that as many maps as possible be stacked across the film and the maximum number (80 separations) be arranged on each magnetic tape. The plotter was capable of producing output in up to 16 grey levels, each raster element being separately controllable using a four-bit intensity code, or dual intensity (black/white only) using a single-bit intensity code. Though initial ratio maps were produced in up to thirteen grey levels and automated half-toning was accomplished (program HALF-TONE), the extreme difficulty of reproducing these subtle variations accurately at such high resolution led to its abandonment; all final maps were produced as dual intensity colour separations. At the cost of increased computer time to thin the data from 4 to 1 bit representation (program THIN), this halved the plotting costs. Towards the end of the time these maps were plotted (late 1976 to late 1977), the cost of plotting 40 inch wide film was £20 per foot length for full intensity plotting and £10 per foot for dual intensity work. Five separations of Great Britain at A4 size, each containing between 50 000 and 150 000 zones (between 200 000 and 600 000 dots) could be stacked across the film. The usual procedure was to plot all the red separations for five maps in one strip of film, all the yellow directly below and all the blue below those. Thus the total cost paid for plotting each map was £6. Following this stage, the sheets of film were returned to the CRU and were proofed by a local commercial firm using the Chromaline process; proofing each map cost between £8 and £10.

Several important conclusions arise from this. The first is that the weakest link in the initial chain of reproduction was certainly the transfer of data to an external plotting agency. Even though they were normally extremely helpful, changes in their personnel, difficulties with the mail (several tapes went astray for periods of up to two weeks), and minor but fatal mistakes made in generating the maps which were not detected in test plots on the cathode ray tube ensured that much frustration and delay were caused. These effects, compounded by use of an overloaded university computer and overworked staff, ensured that the active map preparation spread over a period of months. Certain of the difficulties arose from the vital need to save cash by stacking maps across the film and using other plotter-specific features.

Yet the positive benefit of such an operation is its extremely low cost. By using an £80 000 plotter on a bureau basis, because the staff were carrying out the mapping of their own volition, because of the nature of the computer financing methods in British universities and because of the efficiency of the programs and careful job planning, the total cost of producing and proofing the reproduction material for seventy totally original four-colour maps, each based upon very large volumes of data, was about £1500. It should be noted that even minor changes to the method of operation might have greatly increased this figure: using the other laser plotter then available in Britain for this task (to which it is ill-suited) would have raised the plotting cost at that time by a factor of three or four.

THE ADMINISTRATION OF THE ATLAS PROJECT

The original OPCS planning envisaged that the grid square data would be available before the end of 1973. In the event, the last of these data were provided in their raw form in September 1976; within three months, all the data had been converted into IBM form, arranged as a data base and the first maps had been produced. Thereafter, over one hundred grey-tone ratio maps were produced, examined and subsequently abandoned (see (iv), above). The search for a more satisfactory mapping method—culminating in the use of the signed chi-square measure—plus very considerable amounts of statistical analysis of the data and experiments on class intervals, followed by plotting and colour proofing, ensured that the last of the reproduction material was not produced until early 1978. Manual cartography, such as making masks and producing the black plate reproduction material, together with writing of the text accompanying each map and the more general chapters, took a further few months; the reproduction material and almost the entirety of the text were handed to OPCS in mid-May 1978. A series of unfortunate internal factors, especially priority conflicts, caused delay until August 1979 before Her Majesty's Stationary Office (HMSO) found a printer for the atlas. Though less delayed internally, the production of the wall map involved even more organisations: the CRU produced the film positives, OPCS negotiated a contract with the Central Office of Information (COI) who designed the layout and wrote the printing specification; HMSO then found a printer who sub-contracted the plate making. The plate maker called in a small cartographic firm to make the sea mask and to implement various parts of COI's design. It will be readily appreciated that such long chains of command readily lead to delays and errors, however well-intentioned are the individual parties. The delays in the project, then, were substantially outside the control of the CRU and were such as to diminish greatly the practical utility of the end products although they do not detract from the cartographic innovations made.

CONCLUSIONS

The making of a census atlas by computer may now be an extremely rapid matter: some line printer atlases of such data have been produced in a matter of two or three months but many of these are simple in concept and execution and crude in appearance; several have been heavily criticised by reviewers. The atlas described in this paper took much longer to produce—something akin to eighteen months from the time when the last of the data were available to the time the reproduction material was sent for proofing. Thus its production is not remarkable for its rapidity, largely because of the factors described in the preceding section. The end result, however, is possibly the most detailed computer-produced atlas ever to be published for a country the size of Great Britain. Despite this detail, inter-regional and large scale variations may readily be observed; because of the detail, the gross filtering effects resulting from use of large data areas in other atlases are absent.

Its gestation covered a period of experiment into procedures for checking and cleaning data, of carrying out univariate and multivariate statistical analyses and relating these to the mapping, of testing the signed chi-square mapping technique, of implementing the results of class interval theory and of persevering with and refining the mapping programs—quite apart from production of the reproduction material. The result is an innovative atlas which makes use of detailed small-area data to portray the real world rather than administrative abstractions of it, yet also attempts to tackle the severe problems stemming from small-area data and confidentiality restrictions. The results described in this paper have application beyond grid square data and the techniques have other applications—as a short term benefit, wall maps have been produced; in the longer term, the same procedures will facilitate comparison between 1971 and 1981 data (if grid square data are produced from the next census). Most fundamentally, we believe our experience casts considerable doubt on the validity of many other choropleth-type maps based upon small and varying populations (Visvalingam and Dewdney, 1977).

If conclusions of some general value may be drawn, the project has also posed other questions for further investigation. Do a reasonable number of map users understand the chi-square mapping procedures? (indeed, does this matter so long as they receive a faithful impression of spatial variation?). Can mapping at such small scales—as claimed—really be used and, if so, for what specific applications? Would any organisation be prepared to pay the full economic costs of such mapping? How valuable is mapping of data which are eight years old (though it is at least four years until the next data are available)? Are the procedures capable of being implemented on locally run mini-computers? It is obvious to the authors that cartographers are poised at the point where changes in mapping technology, data availability and some mapping concepts have combined to produce an uncertain but potentially very different future. In any case, we have shown that automated thematic cartography can surpass the quality of manual cartography, and reveal spatial distributions in new ways.

ACKNOWLEDGEMENTS

Apart from the generous support of the Social Science Research Council in providing a grant to the CRU and that of the Office of Population Censuses and Surveys in providing the grid square data, we specifically wish to acknowledge the assistance of the following individuals who have helped us in various ways: Sybil Brown, Jane Coulter, Joan Dresser, Catherine Hakim, Barbara Perry, Judith Rattenbury, Kay Stanness, Susan Wileman, Gerald Boston, Arthur Blacklock, John I Clarke, Arthur Corner, Chris Denham, John C. Dewdney, Peter Ferrar, Derek Hudspeth, John Steele.

The final manuscript of the paper was compiled and typed while one author (D. W. R.) was on sabbatical leave in the Australian National University and the Canberra College of Advanced Education.

BIBLIOGRAPHY

1. B. Coates, *Census Atlas of South Yorkshire*, Department of Geography, University of Sheffield (1974).
2. J. C. Dewdney and D. W. Rhind (eds), *People in Durham: a census atlas*, Census Research Unit, Department of Geography, University of Durham (1975).
3. Department of Environment, *Atlas of the Environment*, DOE, London (1977).
4. I. S. Evans, 'The selection of class intervals', *Trans., Inst. Brit. Geogr.*, N.S. **2**, (1) (1977), 98–124.
5. I. S. Evans, 'Relationships between G.B. census variables at the 1 km aggregate level' in N. J. Wrigley (ed.), *Statistical applications in the spatial sciences*, Pion, London (1979).
6. A. J. Hunt (ed.), 'Population maps of the British Isles, 1961', *Trans., Inst. Brit. Geogr.* No. 43 (1968).
7. O.P.C.S. *1971 Census grid square statistics explanatory notes* (1977), 35 pp.
8. D. W. Rhind, 'High speed maps by laser beam', *Geographical Magazine* **46**, (1974), 393–4.
9. D. W. Rhind, M. Visvalingam, B. Perry and I. S. Evans, 'People mapped by laser beam', *Geographical Magazine* **48**, (1976), 148–54.
10. D. W. Rhind, I. S. Evans and J. C. Dewdney, 'The derivation of new variables from population census data', *Census Research Unit, Working Paper*, No. 9, Department of Geography, University of Durham (1977).

11. K. E. Rosing and P. A. Wood, *Character of a conurbation. A computer atlas of Birmingham and the Black Country*. London (1971).
12. J. Shephard, J. Westway and T. Lee, *Social atlas of London*. Oxford (1974).
13. W. R. Tobler, 'Choropleth maps without class intervals?', *Geographical Analysis* 5, (1973), 262–5.
14. M. Visvalingam and D. W. Rhind, 'Compaction of the 1971 UK census data', *Computer Applications* **3**, (1976), 499–511.
15. M. Visvalingam, 'Chi-square as an alternative to ratios for statistical mapping and analysis', *Census Research Unit, Working Paper*, No. 8, Department of Geography, University of Durham (1976), 21 pp.
16. M. Visvalingam and J. C. Dewdney, 'Size of areal units and the performance of ratio and chi-square measures', *Census Research Unit, Working Paper*, No. 10, Department of Geography, University of Durham (1977), 31 pp.

APPENDIX 1 *People in Britain: A Census Atlas*

LIST OF CONTENTS

Colour maps with this article are from "People in Britain—a Census Atlas", published by HMSO, 1980.

Reproduced by permission of the Controller, HMSO, Crown Copyright Reserved.

Map 1. The distribution of miners in 1971, as an absolute number map; based upon 10% sample data. Scale 1: 4 million.

Map 2. The distribution of unemployed men in 1971, as a signed chi-squared map. Scale 1: 4 million.

Map 3 (top) 'Population density' in and around Birmingham. Scale 1: 1 million. A 'place-names' map is provided in the atlas. Colours are as in map 1 except that an extra class (white) is included for unpopulated areas and densities are higher—the top class is more than 6000 people per km^2.

Map 4 (bottom) 'born outside the UK' in and around Birmingham. A signed chi-squared map, using the same key as map 2.