Several polychotomous characteristics, when expressed in chi-score form, exhibit distributions in property space which deviate from the inherent assumptions of the more commonly used methods of cluster analysis. This paper suggests a novel approach to classification which exploits the observed and inferred characteristics of categorised data. The concepts and methods are applied to data on household composition to extract a demographic typology from the one-kilometre grid square population census data for 1971.

The Signed Chi-score Measure for the Classification and Mapping of Polychotomous Data

M. Visvalingam

Geography Department, University of Hull

INTRODUCTION

Visvalingam[12] introduced the signed chi-square measure as an alternative to ratios and numerical differences for ordering data sets. It is a compromise measure which simultaneously considers the magnitude of the absolute and relative deviations from expectation. It was used in preference of ratios for the ordering and classification of dichotomous variables in 'People in Britain—a census atlas'.[3,4,5] Since the signed chi-square values for the two categories are inversely correlated, the spatial distribution of unemployment for example, is the inverse of that of employment, just as that of masculinity is the inverse of the distribution of femininity.

This paper considers the use of the signed chi-score (X_s) measure for the classification and mapping of polychotomous data. Individual characteristics such as age, occupation, deaths by various causes, unemployment in different sectors and others are polychotomous and involve extensive data sets, which hitherto are unsatisfactorily summarised by average values, single indices or by grouping. Clarke[6] suggested that the analysis of individual characteristics, consisting of several categories, was likely to yield better results than classification procedures involving several characteristics. Although the latter may be closely correlated their distribution patterns are far from identical, owing to the changing relationships between these factors and the increasing mobility and concentration of people.

Polychotomous data have been analysed in a variety of ways. The nature of individual characteristics have been ascertained by evaluating the statistical relationships between categories of interest, for example via correlation matrices,[28] Duncan and Duncans' indices of segregation and dissimilarity[7] and other indices such as the Gini co-efficient and the probability share index.[8] These generalised measures have been complemented by separate maps of selected categories to assess their spatial relationships. Even the more recent atlases have tended to portray categories of one characteristic, such as socio-economic grouping, individually rather than simultaneously.[3,9] Separate maps produce a complex visual correlation problem for map users, especially since the indication of choropleth maps depend on the class interval used.[10] In this context even overlays are not entirely satisfactory. Cartographic correlation[11] has been suggested as a means of quantitatively comparing two spatial distributions but this has not been widely practised for several reasons. Such comparisons are tedious if several categories are employed. Moreover, the method assumes a continuous variation in geographic space, which is unrealistic when used with demographic and social data. Empirical evidence[12,13] has repeatedly established the existence of discrete zones and sectors of residential differentiation.

Composite maps of various types[3,14] have been used to study the spatial relationships between categories. Morgan's[14] composite map of Exeter was based on the concept of over-representation; a particular household type was considered to be over-represented in an enumeration district if the percentage resident in the district lay within the upper quartile of the percentile distribution. Primary symbolism associated with each of the household types were overlayed to produce composite symbols in an effective map of household composition. Holtermann[15] had a similar approach for investigating the spatial coincidence of various forms of deprivation. The Durham Census Research Unit[3] used three primary colours with signed chi-square values to produce a composite map of age structure. While the latter was much more informative than maps of individual age divisions, it was evident that such an approach was not entirely satisfactory owing to practical and other reasons (see section on classification procedures).

In addition to the above approaches, a variety of multivariate techniques—including principal component analysis, factor analysis and cluster analysis—have been applied to demographic, socio-economic and other data to extract from them the principal dimensions of residential differentiation. Although these methods have been widely applied by academics and planners[17, 18] there is growing concern over the relevance of such techniques[19, 20, 42] (see section on classification procedures). A major practical problem associated with many cluster analysis procedures is their excessive demands on computer main store and time. Hence many of these procedures are only applicable to relatively small data sets.[18, 21]

This paper presents a very quick and simple method for classification of polychotomous data. The method is easily implemented, even on mini-computers. Since, the process of classification is easy to understand, errors in judgement can easily be evaluated with the help of preliminary results. The scope of the paper is best outlined within the framework of some crucial stages in classification. These involve the selection of:

1. the number of categories or dimensions (N) needed to adequately portray a characteristic;
2. the ordering and scaling systems to be used for projecting entities onto the N-dimensional measurement space;
3. the procedures to be adopted for establishing boundaries in measurement space, *i.e.* the choice of the classification scheme; and
4. the procedures to be used for evaluating the classification scheme and resulting typologies.

For illustrative purposes, this paper uses the one-kilometre grid square data on household composition provided by the population census (S.A.S. Table 20 of the 100 *per cent* Household data).[22] This characteristic is described by 48 categories of household types. For several reasons it is often necessary to reduce the dimensions of study. Many of the 48 categories contained only small numbers and proportions of households. Thus for results to be statistically meaningful it was necessary to aggregate or combine categories. In many studies the process of aggregation has involved much soul searching in an attempt to extract from the available data, the most appropriate indicators of themes of substantive interest. However in most cases little or no attempt has been made at a subsequent evaluation of the homogeneity or suitability of these summations. Thus stages 1 and 4 above are closely associated and these form the subject of a separate paper. In this paper the 48 categories of household type were reduced to eight for reasons given below.

This paper discusses the rationale for using the signed chi-score value for projecting categorised data onto measurement space. It also discusses at some length the dissection procedure used for partitioning the measurement space, *i.e.* the paper focuses on stages 2 and 3 above. The immediate aim of the exercise is to classify individual areal units into groups in which the composition of household types is recognisably different.

ORDERING AND SCALING SYSTEMS FOR PROJECTING DATA ONTO MEASUREMENT SPACE

Absolute numbers, numerical differences, ratios and signed chi-square measures constitute different ordering schemes and therefore produce different sets of relationships in measurement space. This is especially so when there is a wide range in sample sizes. In such cases the numbers of children and retired will tend to be positively correlated, reflecting their dependence on total population. When such data are standardised with respect to total population, percentage children will tend to be negatively associated with percentage retired partly because of a discrepancy in their spatial distribution but also because this effect is exaggerated by the closure effect in categorised data.[23] While absolute numbers involve no standardisation what-so-ever, the more commonly used ratio measures only include a partial standardisation with respect to the base population. Ratio measures based on small samples are unreliable[24, 25] and their reliability increases with sample size. The signed chi-score measure is analogous to a standard score for ratios[1] and automatically includes some measure of population weighting in its derivation. Ratios and other conventional measures are frequently subjected to various normalising transformations[26] but these are often empirically chosen and the rationale for the choice of appropriate transformations has not been adequately expounded. In this paper a square root transformation is employed for reasons given below.

The data matrix consists of $N \times M$ elements, where N is the number of categories of household types and M is the number of spatial units. This matrix of observed frequencies was standardised with respect to row and column totals, using the procedure for deriving the standard chi-square statistic. The expected frequencies can be calculated in the conventional way as

$$E_{ij} = \frac{C_i \cdot R_j}{T} \tag{1}$$

where, E_{ij} is the expected frequency of category i in spatial unit j,

T is the total number of households in the study area,

C_i is the total number of households of type i,

R_j is the total number of households in spatial unit j.

However it must be stressed that the key issue in the use of this technique is the formulation of expectation.[1] Where there is no adequate theory for projecting some *a priori* expectation, values for expectation can be derived mechanically from contingency tables or related to some probability distribution. Here expectation is based entirely on the available data and (C_i/T) refers to the proportion of households of type i in the study area. There is no reason why it should not be the proportion of households of type i in the nation or some other circumventing area, especially where comparative studies are envisaged.

For ease of computer processing, data for each unit j were processed separately and the data for each category i were standardised to $X^2_{S_i}$ scores as follows:

$$X^2_{S_i} = \frac{(O_i - E_i)^2}{E_i} \cdot \mathrm{sgn}(O_i - E_i) \tag{2}$$

The X^2_S value for each element of the matrix is a measure of the extent to which a category in a particular area deviated from expectation, the excess or deficiency being indicated by a positive or negative sign respectively. The square root transformation (X_S) of the X^2_S scores, where

$$X_S = \mathrm{sgn}(X^2_S)\sqrt{X^2_S} = \frac{O_i - E_i}{\sqrt{E_i}} \tag{3}$$

is useful for plotting each data unit onto an N-dimensional space, the centre of which has zero deviations for all categories. For then, the squared Euclidean distance of the data unit j from this centre, *i.e.* its deviation from average expectation in N-dimensional space would be derived by:

$$X_j^2 = \sum_{i=1}^{N} |X_S^2| = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} \qquad (4)$$

The standard X^2 statistic would then be the sum of the X_j values, *i.e.*

$$X^2 = \sum_{j=1}^{M} X_j^2$$

CLASSIFICATION PROCEDURES

Classification involves the demarcation of boundaries in measurement space. Even for univariate distributions there are several systems in existence for the delimitation of class intervals.[10] Everitt[27] reviewed the several systems for classifying data in N-dimensional space. These may be divided into serial systems, which involve a variety of partitioning and dissection techniques, and idiographic or synthetic systems which involve search algorithms to identify clusters of data units in measurement space.

Automated synthetic clustering techniques are very demanding of computer resources since they are based on some measure of similarity or dissimilarity between pairs of spatial data units. For continuous data these usually take the form of correlation and distance coefficients. The number of pairwise comparisons can prove unmanageably large for many studies using readily available published data sets for a large number of geographical areas. This requires some pre-processing of the data (for example see Webber[21]), and involve the aggregation or sampling of spatial units and/or the reduction of the dimensions of study, using methods such as principal component analysis. Reservations have been expressed concerning the use of principal component analysis[18] but as these are related more to the choice and interpretation of the dimensions of study, they are not discussed here. The end products of cluster analysis can sometimes be difficult to evaluate, especially when the method produces irregular boundaries in measurement space.

Webber,[29] for example, used an iterative relocation procedure based upon minimum error sum of squares. Owing to computing restrictions a sample of 4000 enumeration districts were grouped progressively until 5 broad families were produced and the remaining 116 000 enumeration districts were allocated to clusters to which they were most similar. Evans[28] reports the criticisms advanced by Openshaw and Gillard,[19] who were concerned that there were no tests on the repeatability of clusters given that a different sample of clusters would produce somewhat different clusters. They showed that quite different classifications may be produced from the same variables in a study for a particular conurbation. Another relevant comment was that the classification itself had not been demonstrated to be significant in terms of each cluster being clearly separated from neighbouring clusters.

Grove and Roberts[20] applied a variety of clustering methods, using 1951 data, initially to reproduce the semi-subjective classification of 157 towns in England and Wales by Moser and Scott[30] and then to classify the towns using the 1971 data. They too concluded that different clustering methods, adopting either explicitly or implicitly a different definition of a cluster, produce different classification schemes for the towns. While Mode Analysis was able to reproduce the board outlines of the Moser and Scott scheme, there was very little agreement at a more detailed level. Brindley and Raine express that 'it is after all the aim of the method to differentiate areas however similar they are in fact' since the output is always a set of clusters. Elgie[31] complains that the standardising procedures suppress critical information about the variation in attributes for it is not easy to deduce whether the socio-economic dimension, for example, extends from very rich to very poor or from somewhat rich to poor.

Despite variations in cluster definitions there are some common themes. Clustering techniques are tailored towards the recognition of clusters of units in property space, characterised by the properties of isolation and coherence.[32] The implicit assumption is that distinct clusters do exist.

Even when clusters are defined as sets of data units in hyperspace, exhibiting neither random nor regular distribution patterns and meeting one or more of the various criteria imposed by a particular cluster definition,[16] their properties of location, shape and distribution are generally formulated in terms amenable to statistical processing, involving concepts of central tendency and dispersion. The various algorithms developed for clustering impose on the data to be clustered a structure which may or may not correspond to the natural structure of the data. Mode Analysis[33] is rejected here for its assumption of spherical clusters and contiguous spheres and emphasis on disjoint density surfaces.

Clusters are usually conceived as areas of high density in hyperspace obeying gravitation-like laws according to Sokal.[34] While the delimitation of boundaries between clusters is a difficult process, dependent upon linkage concepts and clustering algorithms, the boundaries are assumed to exist in transition zones, almost always characterised by a more diffuse scatter of data units.

An approach based on such premises is highly inappropriate for several demographic and socio-economic characteristics. Ratio and X_S scatterplots of several pairs of categories indicate only one constellation of data units at the one-kilometre grid square level. This consists of a single dense core from which data units emanate in several directions as swarms and limbs in a continuous fashion.

In this context, the serial systems of dissection were more appealing for their simplicity and lack of assumptions concerning the configuration of clusters in measurement space. The Cartet Count method of Cattell and Coulter[35] essentially consists of partitioning a multi-dimensional space and counting the number of data units in each cartet or hypercube. While this gives a description of the distribution of data units in measurement space, the rectilinear dissection procedure has a tendency to segment natural groups and can result in an unmanageable

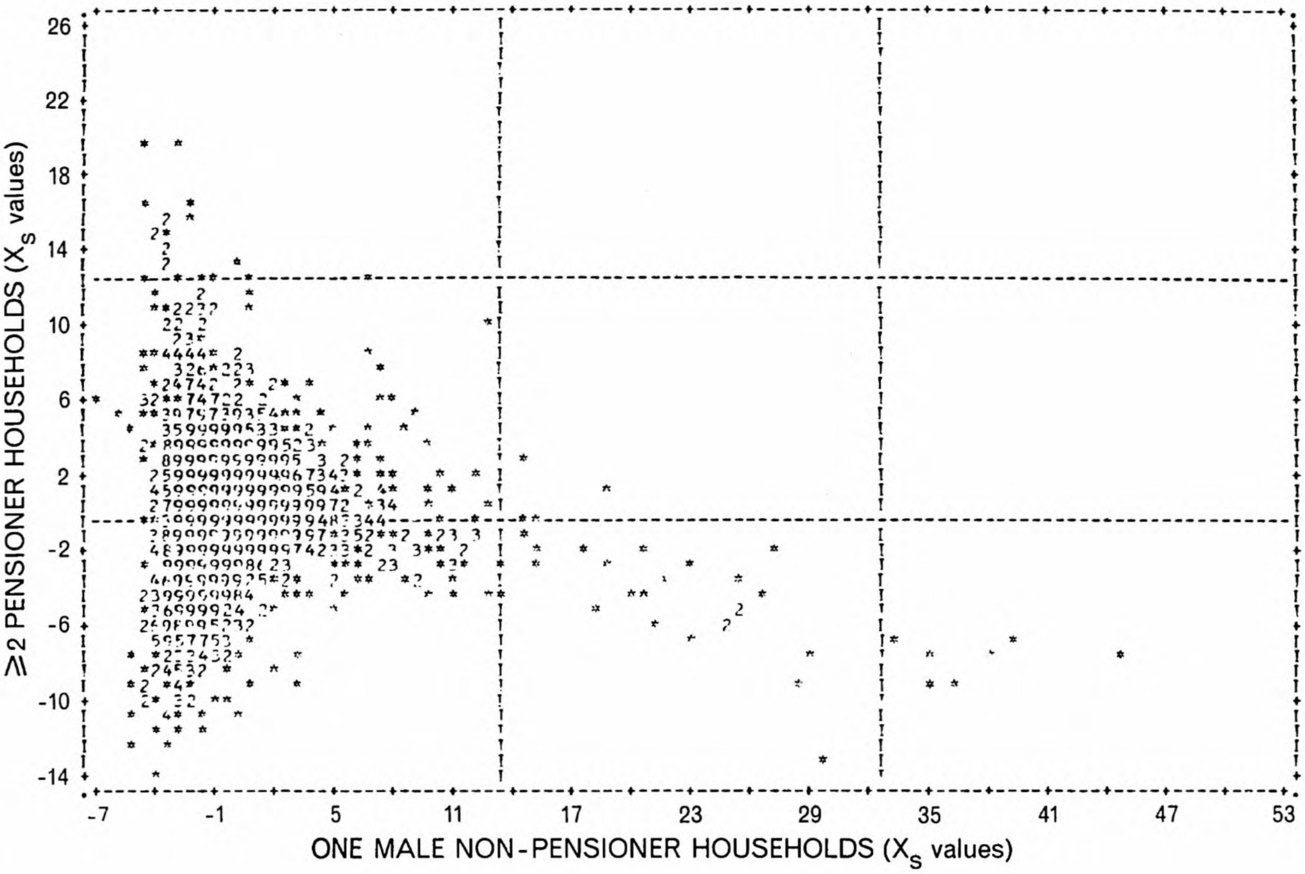

Figure 1. Scattergram of X_S values for variables 3 and 4.

number of cartets or classes. The results can be very confusing since the links between cartets are difficult to establish.

The technique adopted by Holtermann[15] and Morgan[14] for producing composite representations effectively employ divides that are parallel to the axes in measurement space. However the quantile system adopted for locating the divide does not consider the distribution of data areas on each dimension. Thus regardless of the spread of a category, there would always be a proportion of areas in excess just as there would always be an exclusion of areas outside the divide even when these include significant proportions of a category; there is the possibility that some areas with the same concentration of a category may become included while others are excluded if this approach was applied to a very large number of areas. Again the classification is not robust, for the position of the divide is not stationary but hinges on the particular set of data and this does not facilitate any comparison between case studies. For other reservations on the quantile system see Visvalingam and Dewdney.[36] For these reasons the Durham University Census Research Unit[3] selected divides based on X_S^2 cut-off values for the composite map of age structure, using three age categories. The results of such classification schemes can be very confusing where several categories are involved since this results in an exponential increase in the number of classes. Thus attention was directed to observed and inferred characteristics of categorised data in the quest for an alternative procedure.

It was observed that in X_S scatterplots data units with average concentrations are located in or near the dense central core, while the interesting data units with a marked excess or deficit of one or more categories are found in the outer more diffuse parts of the hyperspace. In the proposed classification procedure interest is focused on the relatively diffuse limbs rather than the dense central core and the concepts underlying the classification of polychotomous data are seen as extentions of considerations pertaining to the identification and mapping of extremes in univariate distributions. In *Figure 1*, an X_S plot of two or more pensioner households against one male non-pensioner households, the more extreme deviations tend to have preferred directions and are parallel to the two axes. The elongation towards the third quadrant, *i.e.* the bottom left reflects areas where both categories are deficient and where there is consequently an excess of some other category or categories such as large households. Chayes[23] has demonstrated that there is some measure of negative correlation in categorised data and that this effect is progressively more obvious with a reduction in the number of exhaustive categories. The proposed method exploits this closure effect for classifying polychotomous data. As with the classification of bivariate data[1] the classification of polychotomous data is concerned with the magnitude of deviation from average characteristics and the direction of the significant departures. The classification takes the form of a dissection of the measurement space, undertaken for purposes of generalisation and data reduction to facilitate cognitive description. The resultant typologies are seen as different and extreme parts of a continuum in measurement space rather than as disjoint and discrete clusters. These extremities are of immense interest since their geographic distributions and aspatial characteristics are highly distinctive. However, the boundaries which demarcate them are inevitably arbitrary and artificial, especially near the dense central core, as is often the case in univariate and bivariate classifications.

The N-dimensional measurement space is initially separated into two major sector, the 'average' and the 'outer' more extreme sectors. An arbitrary X_j^2 value (see Equation 4), for example that which corresponds to the 95 *per cent* significance level at $N-1$ degrees of freedom, can be used to separate the two sectors. The latter contains the statistically more interesting departures from expectation and is further partitioned into N sub-sectors so that each sub-sector contained data units with a marked excess of a particular household type. The distinctive category is identified as possessing the maximum X_S value in that geographic unit. Although there is a possibility of a tie between two variables, the inexact representation of floating point numbers in the computer causes the unbiased and random allocation of data units into sectors.

The proposed method is consistent with the method used for the classification of bivariate data.[1] In the bivariate case the closure effect results in a correlation of -1.0 between the two categories both in ratio and X_S terms. 'Average' values were separated from 'outer' ones by an X_j^2 value of 3.84, an arbitrary value which is conventionally associated with the 95 *per cent* significance level for one degree of freedom. The sign of the X_S value (directional measure) was a convenient means of determining the category with the maximum X_S value and was used for allocating data units into one of the two extreme or 'outer' sectors.

The collection of data units within each sector forms a separate group or type. The only certainty concerning the above dissection procedure is that there is a marked excess relative to expectation of a particular category within each data unit and group of data units. There is no implication of the absolute predominance of the attribute. If the distinctive category is positively correlated with others, there would be a tendency towards a corresponding excess of these others in the group. If on the other hand there is a strong negative correlation between the distinctive category and others, there would be deficit of the latter in the group. This does not imply that the pattern of excess and deficit in group characteristics would apply to each data unit within the group. Such inferences cannot be based on weak correlation coefficients since the observed distributions tend to deviate from the assumptions of the general linear model, and may suggest the existence of complex sets of relationships.

The dissection procedure is valid so long as pairs of categories are not strongly positively correlated or off the diagonals in the first quadrant of bivariate scatterplots. However this is not a severe restriction on the use of the method since strong positive correlation suggests information redundancy and spatial associations. In this event it is quite appropriate to aggregate the two categories into a compound category. The procedure is sensitive to the number, selection and definition of initial categories. Thus refinement of the classification is not only concerned with the recognition and adjustment of heterogeneous and similar types but also with evaluating the character of the categorisation.

CLASSIFICATION FOR PURPOSES OF IDENTIFICATION OF DEMOGRAPHIC TYPES

One of the problems confronting the identification of demographic types results from the definition of populations on areal or statistical rather than on demographic or sociological criteria.[6] The resultant population units often lack internal homogeneity and are not demographically distinctive. While all areal data suffer from this deficiency the problems associated with aggregation become more serious the larger the areal unit involved. The present study makes use of data at a relatively fine level of resolution, namely the 1971 one-kilometre grid square population census data supplied by the Office of Population Censuses and Surveys. This forms a gigantic data set and the present example uses a small subset of the data, covering the same area as that considered by Visvalingam and Dewdney.[36] The area comprises the three one-hundred kilometre squares whose south west corners are 300 400, 200 300 and 300 300 respectively It includes 16 612 inhabited one kilometre squares with a total population of 8 497 690 living in 2 834 396 private households. Owing to confidentiality restrictions household data are suppressed when there are less than 8 households in a kilometre square. Unsuppressed household data are available for 8869 squares, 53.4 *per cent* of inhabited squares, but these contain 8 399 868 people (98.8 *per cent* of the population) and 2 834 080 households (99.99 *per cent* of households). Thus the analysis excludes remarkably few households despite dealing with little more than half the inhabited squares. The classification is based on counts for household composition given in S.A.S. Table 20. For several reasons this data was used in preference to data for persons since the latter not only violated the requirements of statistical independence (the distribution of members of the household are related) but also introduced problems related to sample size. For example, the numbers of single persons above the age of 15 tended to be disproportionately small compared with those for married persons. Owing to the complex spatial relationships between types of people, a large number of categories of persons was found necessary and the interpretation of the ecological relationships between different categories proved subjective, speculative and unsatisfactory.

Taeuber and Taeuber[37] and Poole and Boal[8] consider the household to be the fundamental residential and decision making unit. Data on household composition provide a more explicit indication of the relationship between persons since various categories of adults are cross-tabulated against categories of the numbers and ages of children. The data indirectly provide information on household size, stage in family life cycle and fertility. The 48 primary categories and their Great Britain totals are shown in Table 1, where the proportions of the total number of households found in each cell are given in parenthesis. Table 2 gives the corresponding data for the study area.

Many of the 48 categories contained only small numbers and proportions of households. For results to be statistically meaningful, particularly at the one kilometre grid square level, it was necessary to aggregate some of the very small counts. The aggregation

Adults (aged 15 or over) in Household	HOUSEHOLD TYPE					
		One child		Two or more children		
	No child	0–4	5–14	all 0–4	all 5–14	others †
One pensionable male	VARIABLE 1					
One pensionable female	VARIABLE 2					
Two or more, all pensionable						
Two or more, one not pensionable	VARIABLE 3					
One other male	VARIABLE 4					
One other female	VARIABLE 5					
Two or more, others*						
Two or more, none pensionable	VARIABLE 6	VARIABLE 7			VARIABLE 8	

* = households with two or more adults, some of pensionable age and more than one not of pensionable age (O.P.C.S. definition)

† = two or more children, one or more aged 0–4 and one or more aged 5–14 (O.P.C.S. definition)

Figure 2. Derivation of secondary variables.

adopted in this paper is tentative and arbitrary; it reflects not only the investigator's interest and judgment as to which categories of household type were of substantive importance but also the constraints imposed by practical considerations. For illustrative purposes a small number of categories was desirable. Some consideration was paid to the categorisations adopted by Morgan[14] and Abu-Lughod and Foley.[38] These researchers were concerned with household segregation at different stages of the family life cycle. Since family life cycle is just one of the factors leading to demographic differences, and as there is still an absence of an *a priori* logical or philosophical theory for predicting the number, let alone the nature, of distinct and transitional demographic types, (especially at the present scale and extent of study), the investigation was of necessity empirical and exploratory.

Several considerations were taken into account in devising the eight-fold categorisation shown in *Figure 2.* The primary aim was to retain the demographic data on age, sex and marital status of adults together with some discriminating information on children. Secondly, the aggregation procedure had to consider likely similarities and dissimilarities in the spatial distribution of primary categories. Thirdly, it was recognised that the enumerated data refer to the composition of households on census night and is a *de facto* rather than a *de jure* tabulation. Thus the counts of households with children with one pensioner adult or one non-pensioner make may include temporary as well as permanent arrangements. Consequently, emphasis was placed on the types of adults and the eight O.P.C.S. classes were reduced to six by aggregating rows three and four as one category and rows seven and eight as another. Since the bulk of the children are in this last category, this several adult type was subdivided on the basis of the numbers of children present.

The eight categories thus defined include all private households with adults and are mutually exclusive. Table 3 shows the composition of the eight categories in Great Britain and in the study area. Table 4 gives the general correlations between the categories and the slope of their regressions. The correlation and regression coefficients and bivariate scatterplots suggest that the relationship between male and female households of one non-pensioner (variables 4 and 5) is the only one which is likely to violate the constraints on the use of the procedure. Since the present investigation is not only exploratory but also illustrative, this exercise provides an opportunity for assessing the effects of numerically or statistically inappropriate combinations of categories. Thus the eight-fold categorisation was retained, especially since it was proposed on the merit of substantive interest. The resulting groups are as follows:

Group name	*Distinctive category when*
A	none; average class
B	one male pensioner
C	one female pensioner
D	several pensioners
E	one male non-pensioner
F	one female non-pensioner
G	several adults; no child
H	several adults; one child
I	several adults; two or more children

The spatial distribution of groups A, B, C and D is portrayed in *Figure 3.* The aggregate frequencies of household types in each group are given in Table 5. Table 6 gives the proportions of different household types within each group. In Table 7, the proportions of each household type found in different groups are tabulated to give some indications of the degree and direction of concentration of each household type. The age-sex pyramids for the various groups are also provided in *Figure 4.* For further illustration and a detailed commentary see Visvalingam (1979).[39]

The spatial and aspatial characteristics of the groups suggest that the irregularities present in the data are not the product of random or chance variations alone. Group A, which contains over 56 *per cent* of the data units and 17 *per cent* of households, is an average group which requires further analysis. Consequently it is omitted in the present discussion. Groups B and C comprise a broadly similar mix of household and population constituents. Neither of them exhibit marked spatial contiguity and are found in areas of

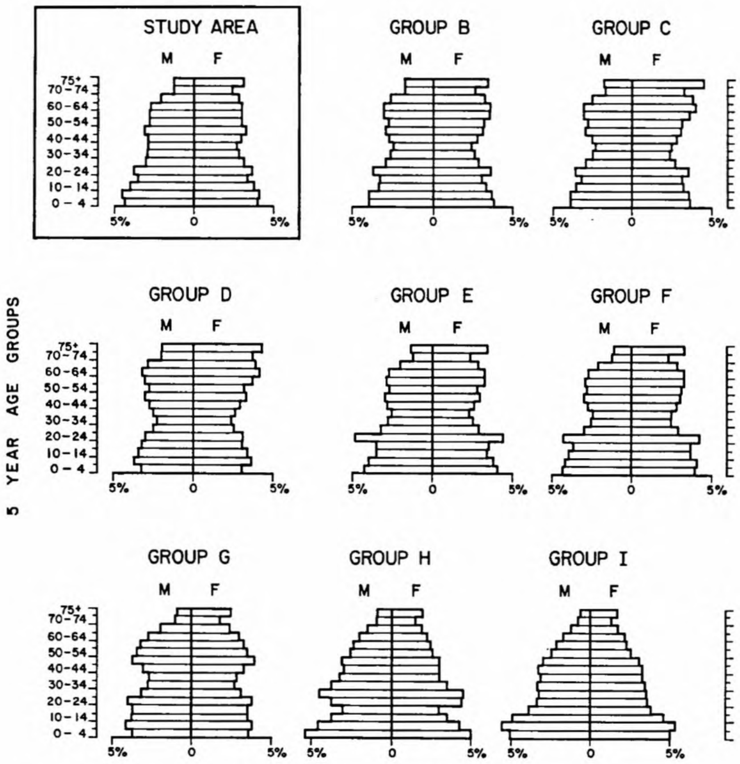

Figure 4. Percentage distribution of population by age and sex within the study area and within each of the groups.

older housing, both urban and rural, in areas with an excess of property rented privately and unfurnished. Whereas in urban areas the occurrence of Group B is masked by the preponderance of Group C, it is particularly distinctive in the rural districts of England and Wales, owing to the relative paucity of Groups C and D. The low household density (149) and the dispersed geographical distribution of Group B suggest small populations. Group C on the other hand shows some concentration in the older industrial areas of Blackburn, Burnley, Preston and the eastern part of the Manchester conurbation and in the Potteries. This group records the highest household density (885) suggesting essentially urban locations. This contrast in location, coupled with the marked over-representation of females in Group C, may point to persistent migration of females from Group B to Group C areas; employment opportunities for women in agricultural areas being inferior to those in textile manufacturing districts. Both Groups B and C had experienced an out-migration of young persons in the past. However, the age-sex pyramids for these groups suggest that by 1971 these areas were attracting numbers of young people. Group C in particular has an excess of married persons under the age of 25 and one non-pensioner households, including those with one child and one non-pensioner female. These figures suggest that young and disadvantaged people are colonising these areas with a sessile old population. It is quite likely that the older residents in Groups B and C have tended not to move on retirement owing to social and economic factors unlike the older residents in Group D.

Group D is found along large stretches of the North Wales and Lancashire coasts which have experienced an in-migration of retired people, both single and married. Distinct blocks of Group D are also found in Newcastle-under-Lyme, Manchester C.B. (around Didsbury), St Helens and Northwich. Within Liverpool C.B. Group D is found mainly in areas of relatively high status housing identified by Webber.[21] These areas have an elderly age structure and unusually low proportions of large families and younger persons in general.

Group E is particularly notable for the preponderance of one-person households, both non-pensioner and pensioner, male and female. Average household density is high (738) and its spatial distribution[39] shows a marked concentration within inner areas of cities such as Manchester, Liverpool, Birkenhead, Chester, Stoke, Oldham, Rochdale, Bolton, Preston and Burnley. There is also a pronounced concentration of this group along the sea front and a dispersed distribution in rural areas. While the outstanding excess is in no-child households, this group is also above average in one-parent households, especially those with one child under the care of a woman. There is also a small excess of pre-school children and an under-representation of married persons of pensionable age.

The process of recolonisation of older housing stock with elderly residents by various transient and/or disadvantaged groups, which was evident in Groups B and C, is more advanced both in intensity and extent in Group E. The distribution of Group E in Liverpool and Manchester suggests that this process has occurred not only in the three-storey villa areas of rooming houses and in areas of older terraced housing but also in the inner, older council estates as suggested by Webber. As a result Group B and C are almost absent from Liverpool C.B. and from Merseyside as a whole, with the exception of Wallasey, which contains a once popular but now decaying holiday resort of New Brighton.

Group F is the least interesting of the eight deviant groups. The correlation and regression coefficients presented in Table 4 suggest a moderately strong association between male and female non-pensioner households. Further statistical analysis of the aggregate frequencies of household types in each group (Table 5) indicates that the concentration of variable 5 in Group E is more significant than its concentration in Group F.

Groups with an excess of several non-pensioner households, with or without children—*i.e.* Groups G, H and I—show suburban locations.[39] There is a marked paucity of these types in the rural parts of Wales, Cheshire, Lancashire and Yorkshire. Data units belonging to Group G exhibit contiguity in an area north-west of Manchester from Worsley to Bury and in the urban areas of St Helens, Crewe, Wolverhampton and a zone from Wolverhampton and Kidsgrove to Stoke-on-Trent. The only contiguous block of Group G within Liverpool C.B. occurs in Webber's[21] high status family, with higher scores for middle-aged than retired people. This is usually associated with semi-detached housing, which gives the group its lower average household density. Indeed Group G is particularly noteworthy for the excess of married persons between the ages of 40 and 59 and of single adults between the ages of 15 and 24, who are probably the non-dependent children of the middle-aged couples.

The age structure and household composition of

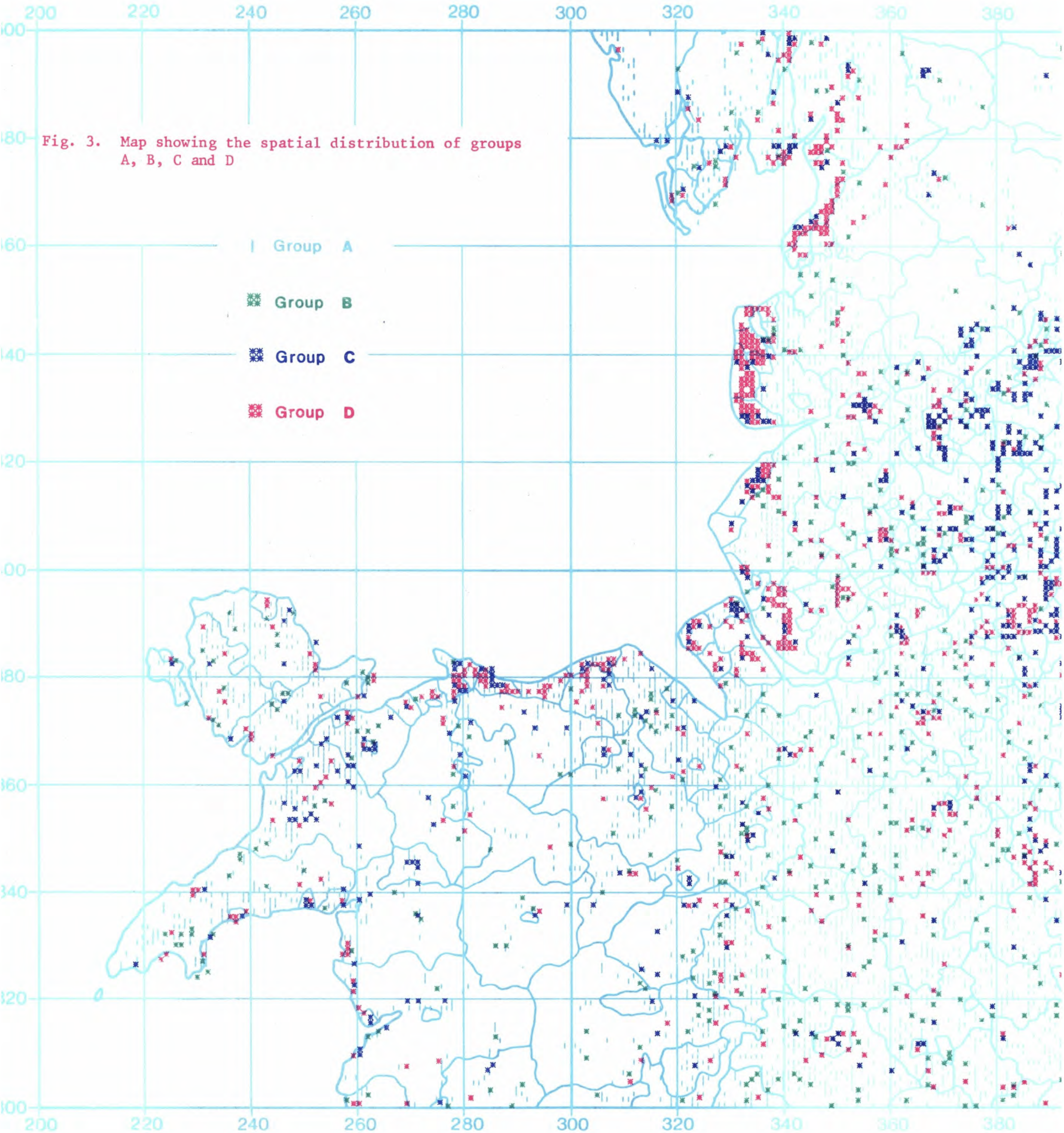

Fig. 3. Map showing the spatial distribution of groups A, B, C and D

TABLE 1

Household types: Great Britain totals

Adults (aged 15 or over) in household	Household Type						
	No child	Children aged 0–14 in Household					Total
		One child		Two or more children			
		0–4	5–14	all 0–4	all 5–14	others†	
One pensionable male	351 772 (1.93)	19 (n)	377 (n)	3 (n)	75 (n)	11 (n)	352 257 (1.94)
One pensionable female	1 781 280 (9.79)	346 (n)	3 716 (0.02)	53 (n)	731 (n)	184 (n)	1 786 310 (9.82)
Two or more, all pensionable	1 591 641 (8.75)	540 (n)	3 917 (0.02)	77 (n)	816 (n)	239 (n)	1 597 230 (8.78)
Two or more, one not pensionable	1 471 785 (8.09)	7 960 (0.04)	28 679 (0.16)	1 780 (0.01)	11 522 (0.06)	5 470 (0.03)	1 527 196 (8.39)
One other male	625 274 (3.44)	3 602 (0.02)	13 051 (0.07)	1 240 (0.01)	12 429 (0.07)	5 176 (0.03)	660 772 (3.63)
One other female	555 780 (3.05)	30 613 (0.17)	57 641 (0.32)	18 320 (0.10)	67 302 (0.37)	55 848 (0.31)	785 404 (4.32)
Two or more, others*	675 826 (3.71)	43 070 (0.24)	110 539 (0.61)	17 381 (0.10)	81 702 (0.35)	56 615 (0.31)	985 133 (5.41)
Two or more, none pensionable	4 674 672 (25.69)	919 122 (5.05)	1 342 388 (7.38)	596 538 (3.28)	1 540 715 (8.47)	1 427 441 (7.85)	10 500 876 (57.71)
Total	11 728 030 (64.46)	1 005 272 (5.52)	1 560 308 (8.58)	635 392 (3.49)	1 715 471 (9.43)	18 195 457 (8.42)	18 195 457 (100.0)

TABLE 2

Household types: totals in the study area

Adults (aged 15 or over) in Household	Household Type						
	No child	One child		Children aged 0–14 in household	Two or more children		Total
		0–4	5–14	all 0–4	all 5–14	others†	
One pensionable male	56 754 (2.00)	5 (n)	72 (n)	0 (n)	0 (n)	2 (n)	56 842 (2.01)
One pensionable female	294 264 (10.38)	43 (n)	594 (0.02)	5 (n)	98 (n)	21 (n)	295 025 (10.41)
Two or more, all pensionable	244 683 (8.63)	62 (n)	566 (0.02)	10 (n)	123 (n)	26 (n)	245 470 (8.66)
Two or more, one not pensionable	231 103 (8.15)	1 428 (0.05)	4 781 (0.17)	263 (0.01)	1 896 (0.07)	911 (0.03)	240 382 (8.48)
One other male	92 252 (3.26)	608 (0.02)	2 289 (0.08)	275 (0.01)	2 102 (0.07)	922 (0.03)	98 448 (3.47)
One other female	79 793 (2.82)	4 390 (0.15)	9 024 (0.32)	2 708 (0.10)	10 413 (0.37)	8 825 (0.31)	115 153 (4.06)
Two or more, others*	105 927 (3.74)	7 339 (0.26)	18 076 (0.64)	2 628 (0.09)	12 732 (0.45)	9 141 (0.32)	155 843 (5.50)
Two or more, none pensionable	703 036 (24.81)	146 353 (5.16)	213 017 (7.52)	92 834 (3.28)	238 970 (8.43)	232 707 (8.21)	1 626 917 (57.41)
Total	1 807 812 (63.79)	160 228 (5.65)	248 419 (8.77)	98 723 (3.48)	266 343 (9.40)	252 555 (8.91)	2 834 080 (100.00)

n. = negligible *i.e.* less than 0.005 *per cent* of all households.
* = households with two or more adults, some of pensionable age and more than one not of pensionable age (O.P.C.S. definition).
† = two or more children, one or more aged 0–4 and one or more aged 5–14 (O.P.C.S. definition).

TABLE 3

Summary statistics for the eight secondary variables

Variable	Great Britain		Study Area	
	No.	(%)	No.	(%)
1. One male pensioner	352 257	(1.94)	56 842	(2.01)
2. One female pensioner	1 786 310	(9.82)	295 025	(10.41)
3. Several pensioners	3 124 426	(17.17)	485 852	(17.14)
	5 262 993	(28.93)	837 719	(29.56)
4. One other male	660 772	(3.63)	98 448	(3.47)
5. One other female	785 504	(4.32)	115 153	(4.06)
	1 446 276	(7.95)	213 601	(7.53)
Several other adults with:				
6. no child	5 350 498	(29.41)	808 963	(28.54)
7. one child	2 415 119	(13.27)	384 785	(13.58)
8. two or more children	3 720 392	(20.45)	589 012	(20.78)
	11 486 009	(63.13)	1 782 760	(62.90)
Total	18 195 457	(100.00)	2 834 080	(100.00)

TABLE 4

Matrix of correlation coefficients and regression slopes between the eight variables

Regression coefficients	Correlation coefficients							
	V1	V2	V3	V4	V5	V6	V7	V8
V1		0.43	0.30	0.22	0.16	−0.28	−0.38	−0.43
V2	0.28		0.55	0.18	0.28	−0.45	−0.62	−0.67
V3	0.19	0.54		−0.06	−0.01	−0.31	−0.61	−0.67
V4	0.14	0.18	−0.06		0.58	−0.20	−0.35	−0.31
V5	0.15	0.40	−0.01	0.85		−0.28	−0.39	−0.29
V6	−0.28	−0.69	−0.50	−0.32	−0.31		0.21	−0.05
V7	−0.33	−0.82	−0.61	−0.48	−0.36	0.18		0.52
V8	−0.25	−0.60	−0.61	−0.29	−0.18	−0.03	−0.35	

Note: The values in this table provide a useful summary of the nature of the bivariate distributions. Since the latter violate the assumptions of the general linear model, these figures should not be used for inferential purposes

Group H areas suggest that they are zones of more recent housing development. While Group H has a marked over-representation of several adult with one child households, it also records an excess of other adult households without dependent children or with relatively larger families. It is interesting that areas of this type are absent from Liverpool and Manchester. Small groups of Group H occur in commuter settlements in the rural districts of North Cheshire.[39] The relatively low household density in these areas suggest more dispersed dwellings and a lower degree of multiple occupancy. Group H also occurs in areas with an excess of council dwellings in the zone extending from Orrell to Haydock and Newton-le-Willows, in Lumm, Cannock, Walsall and in the Stoke-on-Trent area.

Group I is the modal type, both in its coverage of data units and in terms of the proportions of households and has a fairly high average household density of 663. The modal age group for married persons is 30 to 44, which is somewhat older than in Group H. While Groups G and H are striking for their excess of married people, Group I is outstanding for the concentration of children. The age–sex pyramid shows a very marked bell-shaped distribution. The geographical distribution of Group I involves a variety of locations. The concentration of Group I in the Valley R.D. of Anglesey, in Shrewsbury and in Shifnal R.D. are probably associated with a defence establishment. Group I shows a distinctive pattern of spatial distribution in the Wirral peninsula, south Manchester, Cheadle U.D., Gatley and Formby and occurs as arcs and rings around several municipal and county boroughs including Stafford, Oswestry, Chester and Liverpool. Within Liverpool C.B. this group overlaps with council estates built since 1966. The demographic structure in these areas reflects council policy in the allocation of public housing. One parent families, especially those with two or more children have also benefited from Council policies. The spatial distribution of Groups H and I also confirm Webber's observation that larger families tend to be allocated houses in nearer council estates while younger couples with one child are allocated to more distant estates, for example in the zone from Orrell to Haydock and Newton-le-Willows.

Although the groups were identified on the bais of distinctive categories, they are interesting for the varying mix of household and population components. These variations, although relatively small in ratio terms, suggest different demographic environments and societal processes in operation. The results confirm the Shevsky–Bell theory[40] that the stage in family life cycle is an important factor associated with, even if not responsible for, residential differentiation. while a discussion of this topic is outside the scope of this paper the methodology proposed here promises to be of value to the study of residential differentiation.

Returning to the demographic typology, there is some degree of variation present within each group. A finer classification could be produced in several ways. One approach would examine the distribution in measurement space more closely. Using this approach groups could be subdivided on the basis of the relationship of the distinctive category to other categories. Alternatively, groups could be segmented by employing a series of X_j^2 values (equation 4) which would subdivide the outer sector into a series of 'concentric' zones, thereby separating out areas where the processes of segregation have resulted in progressively more marked concentrations of particular types of households. A completely different approach would consider other related factors such as tenure, and the mobility, ethnicity or socio-economic composition of the population to discern their effects on demographic variability.

CONCLUSION

This paper dwelt on two associated themes. It considered the use of the signed chi-score measure for projecting polychotomous data onto measurement space. Secondly it discussed the rationale for the adoption of yet another approach to classification. The proposed dissection procedure has the merits of simplicity and minimal use of computer resources. This extends its applicability to very large volumes of data and avoids the need for sampling. Moreover, values for expectation can be chosen to facilitate either comparability between different case studies or to determine the degree of sorting within each study area separately. The method is consonant with the observed distribution of data in measurement space and is directed towards detecting the directions of variation rather than clusters *per se*. It draws attention to extreme variations in diffuse areas of the property space rather than to dense areas. In this respect it resembles the conventional approaches for classifying univariate distributions. It is also the user's responsibility to distinguish between average and deviant groups through a selection of X_j^2 values.

Population census data on household composition was used to identify demographic types and to illustrate the concepts and methods involved. However, the procedures have been tested successfully with other polychotomous data including data on tenure characteristics and on the birthplace of recent immigrants into the United Kingdom. The method promises to be a simpler and more explicit alternative to factor analytic techniques for the study of spatial and social differentiation.

The immediate aims of classification are the definition of spatial sub-systems and the compilation of a detailed catalogue of group tendencies for purposes of cognitive description and comparison. It is appreciated that data generalisation and reduction and an analysis of ecological patterns cannot in themselves provide

TABLE 5

Frequency Distribution of the eight household types among the nine groups

Group	Household Types												
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total house-holds	Total km squares	House-hold density	% of house-holds	% of data areas
A	9 132	47 415	89 076	13 909	16 849	147 832	68 253	101 527	493 993	4 974	99	17.43	56.08
B	3 380	8 377	13 427	2 520	2 756	19 923	9 180	12 593	72 156	484	149	2.55	5.46
C	12 595	74 078	92 728	17 482	21 582	124 331	54 725	77 004	474 525	536	885	16.74	6.04
D	9 453	52 109	102 595	11 418	15 843	116 186	49 061	70 704	427 369	699	611	15.08	7.88
E	7 978	37 144	51 890	26 998	20 808	88 025	36 417	57 778	327 038	443	738	11.54	4.99
F	1 960	11 506	16 316	4 479	6 871	27 180	11 822	18 814	98 948	178	559	3.49	2.01
G	3 281	16 766	32 777	5 240	7 083	79 119	33 331	44 690	222 287	394	564	7.84	4.44
H	2 829	13 543	24 683	4 594	5 604	61 265	38 532	46 802	197 852	377	525	6.98	4.25
I	6 234	34 087	62 360	11 808	17 757	145 102	83 464	159 100	519 912	784	663	22.22	8.84
Total	56 842	295 025	485 852	98 448	115 153	808 963	384 785	589 012	2 834 080	8 869	320	100.00	100.00

TABLE 6

Percentage distribution of household types within each group

A. Percentage distribution of pensioner, one non-pensioner and several-other-adult household types within each group

Group	Household Type: Pensioner (1, 2, 3)	One non-pensioner (4, 5)	Several other adults (6, 7, 8)	Total
A	28.63	6.23	64.28	100
B	34.90	7.31	57.78	100
C	37.81	8.23	53.96	100
D	38.41	6.38	55.21	100
E	29.67	14.62	55.73	100
F	30.10	11.47	58.93	100
G	23.77	5.55	70.68	100
H	20.76	5.15	74.11	100
I	19.75	5.69	74.56	100
Total	29.56	7.53	62.90	100

B Percentage distribution of the eight household types within each group

Group	Household Type (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A	1.02	9.60	18.03	2.82	3.41	29.92	13.81	20.55
B	4.68	11.61	3.49	3.49	3.82	27.61	12.72	17.45
C	2.65	15.61	19.54	3.68	4.55	26.20	11.53	16.23
D	2.21	12.19	24.01	2.67	3.71	27.19	11.48	16.54
E	2.44	11.36	15.87	8.26	6.36	26.72	11.14	17.87
F	1.98	11.63	16.49	4.53	6.94	27.47	11.95	19.01
G	1.48	7.54	14.75	2.36	3.19	35.59	14.99	20.10
H	1.43	6.85	12.48	2.32	2.83	30.97	19.48	23.66
I	1.20	6.56	11.99	2.27	3.42	27.91	16.05	30.60
Total	2.01	10.41	17.14	3.47	4.06	28.54	13.58	20.78

TABLE 7

Percentage distribution of household types among groups

A. Percentage distribution of household types among groups with an excess of pensioner, one non-pensioner and several-other-adult types

Groups	Household Types (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
A	16.07	16.07	18.33	14.13	14.63	18.27	17.74	17.24	17.43
B, C, D	44.734	45.611	42.967	31.916	34.893	32.194	29.358	27.215	34.37
E, F	17.483	16.490	14.038	31.974	24.037	14.241	12.536	13.003	15.03
G, H, I	21.716	21.827	24.661	21.983	26.438	35.290	40.367	42.544	37.04
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

B. Percentage distribution of household types among individual groups

Groups	Household Types (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Total
A	16.07	16.07	18.33	14.13	14.63	18.27	17.74	17.24	17.43
B	5.95	2.84	2.76	2.56	2.39	2.46	2.39	2.14	2.55
C	22.16	25.11	19.09	17.76	18.74	15.37	14.22	13.07	16.74
D	16.63	17.66	21.12	11.60	13.76	14.36	12.75	12.00	15.08
E	14.04	12.59	10.68	27.42	18.07	10.88	9.46	9.81	11.54
F	3.45	3.90	3.36	4.55	5.97	3.36	3.07	3.19	3.49
G	5.77	5.68	6.75	5.32	6.15	9.78	8.66	7.59	7.84
H	4.98	4.59	5.08	4.67	4.87	7.57	10.01	7.95	6.98
I	10.97	11.54	12.84	11.99	15.42	17.94	21.69	27.01	22.22
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

functional explanations. However, the output from such studies provide an insight into the complex associations between attributes within sub-systems. Classifications are a bonus to practical applications for they provide some basis for stratified sampling,[41] locational studies, area selection studies for specific surveys, regionalisation procedures and for identifying 'target' groups for service provision.[17] Social maps are of interest not only to academics but also to planners, politicians and the commercial sector. Brindley and Raine[18] discuss how social maps at different scales could be used at various levels of planning. Areal studies are also a valuable input to behavioural and other studies whether these are for assessing voting behaviour or for purposes of market research. Finally, the method of classification proposed in this paper can aid sampling from large data sets for input to multivariate cluster analysis.

Acknowledgements

The author is deeply indebted to the many people and institutions who have made this study possible. The research was funded by the Social Science Research Council and the data were provided by the Office of Population Censuses and Surveys. The Durham and NUMAC computer units provided indispensable computing facilities. The writer is particularly grateful for the excellent technical help provided by the Department of Geography of the University of Durham, where the research was started: these include photographic work by D. Hudspeth, cartographic assistance provided by Messrs A. Corner and D. Cowton and reprographic services undertaken by Messrs J. Normile and D. Ewbank. Thanks are also due to other members of the Census Research Unit of the University of Durham for their encouragement, advice and comments on earlier drafts of the paper, to Dr N. J. Cox for some valuable references on classification methods and to Miss Janet Smith for typing the manuscript.

References

1. M. Visvalingam, The signed chi-square measure for mapping, *Cartographic Journal*, **15,** 2 (1978), 93–98.
2. M. Visvalingham, Chi-square as an alternative to ratios for statistical mapping and analysis, Census Research Unit *Working Paper no. 8*, Dept. of Geography, University of Durham, 1976, 21.
3. Census Research Unit, *People in Britain—a census atlas*, H.M.S.O., 1980.
4. Office of Population Censuses and Surveys, People in Britain: an educational wallchart, 1979.
5. D. W. Rhind, I. S. Evans and M. Visvalingam, Making a national atlas of population by computer, *Cartographic Journal*, in press.
6. J. I. Clarke, On the classification of population types and regions, *National Geographer*, **XII,** 2 (1977), 131–6.
7. O. D. Duncan and B. Duncan, Occupational stratification and residential distribution, *Am. J. Sociology*, **60** (1955), 493–503.
8. M. A. Poole and F. W. Boal, Religious and residential segregation in Belfast in mid-1969: a multi-level analysis, in B. D. Clark and M. B. Gleave (eds.), *Social Patterns in Cities*, Inst. Br. Geog. Spec. Publ. No. 5, (1973), 1–40.
9. P. J. Taylor, A. M. Kirby, K. J. Harrop and G. Gudgin, *Atlas of Tyne and Wear*, Univ. of Newcastle-upon-Tyne, Dept. of Geography, Research Paper 11, 1976.
10. I. S. Evans, The selection of class intervals, *Trans. Inst. Br. Geog.* New Series **2,** 1 (1977), 98–124.
11. A. H. Robinson and R. A. Bryson, A method for describing quantitatively the correspondence of geographical distributions, in G. J. Demko *et al* (eds.) *Population geography: a reader*, McGraw-Hill, New York, 1970, 99–113.
12. R. J. Johnston, *Urban residential patterns: an introductory review*, Bell & Sons, London, 1971.
13. B. D. Clark and M. B. Gleave, *Social patterns in cities*, Inst. of Br. Geog. Spec. Publ. No. 5, 1973.
14. B. S. Morgan, The bases of family status segregation: a case study in Exeter, *Trans. Inst. Br. Geog.*, New Series, **1,** 1 (1976), 83–107.
15. S. Holtermann, Census indicators of urban deprivation, WN-6, ECUR Division, Department of Environment, 1975.
16. P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*, Freeman, San Francisco, 1973.
17. D. T. Priestnall, Implications of social area analysis of S. Hampshire, South Hampshire Plan Technical Unit, *Working Paper 17*, 1972.
18. T. S. Brindley and J. W. Raine, Social area analysis and planning research, *Urban Studies*, **16,** 3 (1979), 273–89.
19. S. Openshaw and A. Gillard, forthcoming.
20. D. M. Grove and K. A. Roberts, Principal component and cluster analysis of 185 large towns in England and Wales, *Urban Studies*, **17,** 1 (1980), 77–82.
21. R. J. Webber, Liverpool Social Area Study, 1971 Data: Final Report, *PRAG Tech. Papers TP14*, 1975.
22. Office of Population Censuses and Surveys, *Census 1971, Standard Small Area Statistics (Ward Library) Explanatory Notes*, H.M.S.O.
23. F. Chayes, *Ratio correlation*, Chicago, University of Chicago Press, 1971.
24. M. Choynowski, Maps based on probabilities, *J. Am. Statis. Assoc.* **54** (1959), 385–8.
25. J. C. Dewdney and D. W. Rhind (eds.) *People in Durham—a census atlas*, Census Research Unit, Dept. of Geography, Univ. of Durham, 1975, 34–35.
26. I. S. Evans, J. Catterall and D. W. Rhind, Specific transformations are necessary, Census Research Unit *Working Paper No. 2*, Dept. of Geography, Univ. of Durham, 1975.
27. B. Everitt, *Cluster Analysis*, Heinemann, London, 1974.
28. I. S. Evans, Census Data, in N. Wrigley and R. J. Bennett (eds.), *Quantitative Geography in Britain: retrospect and prospect*, Routledge & Kegan Paul, London, forthcoming.
29. R. J. Webber, Census enumeration districts: a socio-economic classification, OPCS Occasional Paper 14, 1979.
30. C. A. Moser and W. Scott, *British towns: a statistical study of their societal and economic differences*, Oliver & Boyd, Edinburgh, 1961.
31. R. A. Elgie, The segregation of socio-economic groups in urban areas: a comment, *Urban Studies*, **16,** 2 (1979), 191–5.
32. N. Jardine and J. Sibson, *Mathematical taxonomy*, Wiley, London, 1971.
33. D. Wishart, Mode analysis, in A. J. Cole (ed.) *Numerical taxonomy*, Academic Press, New York, 1969, 282–308.
34. R. R. Sokal, Classification: purposes, principles, progress, prospects, *Science*, **185** (4157), 1974, 1115–23.
35. R. B. Cattell and M. A. Coulter, Principles of behavioural taxonomy and the mathematical basis of the taxonomic computer program, *Br. J. Math. Statist. Psychol*, **19** (1966), 237–69.
36. M. Visvalingam and J. C. Dewdney, The effects of the size of the areal units on ratio and chi-square mapping, Census Research Unit *Working Paper No. 10*, Univ. of Durham, 1977.
37. K. E. Taeuber and A. F. Taeuber, *Negroes in Cities*, Chicago, 1965.
38. J. Abu-Lughod and M. M. Foley, Consumer differences, in N. N. Foote, J. Abu-Lughod, M. M. Foley and L. Winnick (eds.), *Housing choices and constraints*, New York, 1960, 95–113.
39. M. Visvalingam, The identification of demographic types: a preliminary report on methodology, Census Research Unit *Working Paper No. 14*, Univ. of Durham, 1979.
40. E. Shevsky and W. Bell, *Social Area Analysis*, Stanford sociological series no. 1, Greenwood Press, Connecticut, 1955.
41. D. T. Herbert and D. J. Evans, Urban sub-areas as sampling frameworks for social survey, *Town Planning Review*, **45,** 1974, 171–88.
42. D. T. Herbert and R. J. Johnston, *Social areas in cities: spatial processes and form*, John Wiley, London, 1976.