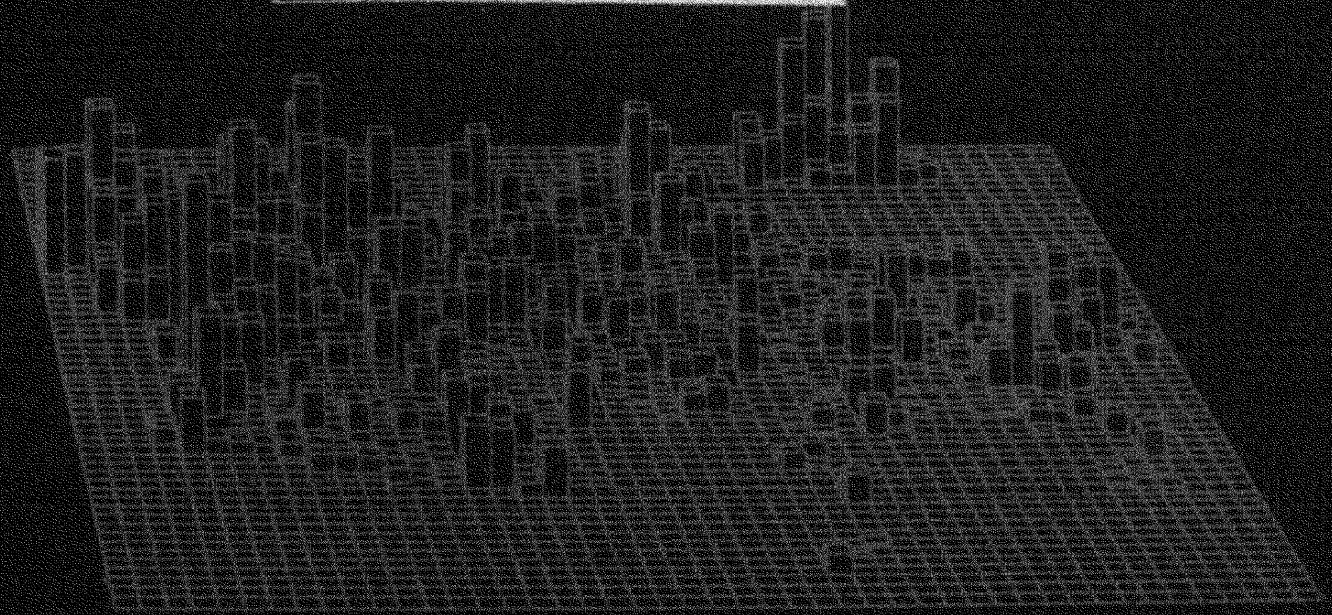
CRU

The identification of demographic types: a preliminary report on methodology

by M. Visvalingam



Census Research Unit
Department of Geography
University of Durham

Working paper 14

The Census Research Unit, Department of Geography, University of Durham, is a small group of research workers investigating aspects of the theory and use of census data. It is currently funded as a research project by the Social Science Research Council.

The diagram on the cover represents total population per 1 km grid square in the northern part of County Durham: the height of each column is proportional to the population in that square. The county is viewed from due west, Gateshead being at the extreme left margin, West Hartlepool at the far right and Bishop Auckland at the centre-right. The original surface was

CENSUS RESEARCH UNIT

DEPARTMENT OF GEOGRAPHY

UNIVERSITY OF DURHAM

WORKING PAPER 14 : THE IDENTIFICATION OF DEMOGRAPHIC TYPES : A PRELIMINARY REPORT ON METHODOLOGY

ERRATA

Page 2, line 9 for "Visvalingam, 1979" read "Visvalingam 1978"

Page 11, line 9 for "Visvalingam, 1979" read "Visvalingam 1978"

line 19 for "square" read "square j"

line 21 for "variable" read "variable i"

line 22 for "X_s^2" read "$X_{s_i}^2$"

line 23 for "$\sum_{i=1}^{N_2} X_s^2$" read "$X_{s_i}^2$"

Pages 14, 15 Interchange figures 2 and 3

Page 16, line 6 for "Visvalingam, 1979" read "Visvalingam 1978"

Page 19, line 32 for "date" read "data"

Page 29, line 10 for "Burnely" read "Burnley"

Page 39, line 23 for "Figure 7" read "Figure 10"

Page 55, line 1 for "VISVALINGAM, M (1979)" read "VISVALINGAM, M (1978)"

line 2 for "Vol. 16(1)" read "Vol. 15 (2)"

CONTENTS

UNIVERSITY OF DURHAM

DEPARTMENT OF GEOGRAPHY

CENSUS RESEARCH UNIT

WORKING PAPER No. 14

SEPTEMBER 1979

THE IDENTIFICATION OF DEMOGRAPHIC TYPES : A PRELIMINARY REPORT ON METHODOLOGY

M. VISVALINGAM

ILLUSTRATIONS

THE IDENTIFICATION OF DEMOGRAPHIC TYPES :

A PRELIMINARY REPORT ON METHODOLOGY

1. INTRODUCTION

This paper presents a simple dissection procedure for the identification of demographic types. Most automated classifications are based on some measure of similarity or dissimilarity between pairs of data units. For continuous data, these usually take the form of distance or correlation coefficients.

In the case of the Ward Library census data for the whole of Great Britain, the number of pairwise comparisons would be unmanageably large. Moreover, such measures are tailored towards the recognition of clusters of data units in property space, characterised by the properties of isolation and coherence (Jardine and Sibson, 1971); the assumption is made 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 (Sneath and Sokal, 1973), their properties of location, shape and distribution in space 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 units to be clustered a structure which may or may not correspond to a natural structure of the data.

Clusters are usually conceived as areas of high density in hyperspace, obeying gravitation-like laws (Sokal, 1974). While the delimitation of boundaries between clusters is a difficult process, dependent on linkage concepts and clustering algorithms, the boundaries are assumed to exist in transition zones characterised by a more diffuse scatter of data units.

An approach based on such premises is highly inappropriate for the identification of demographic types. Various ratio and X_s^2 representations of demographic variables indicate only one constellation of data units. The scatterplots of these representations indicate a single dense centre, a core of near-average and small populations from which data units emanate in several directions in a continuous fashion as swarms and limbs.

In the proposed classification procedure, interest is focussed on the relatively more diffuse limbs rather than on the dense central core. Thus the concepts underlying the identification of demographic types are seen as extensions pertaining to the identification and mapping of extremes in univariate distributions. The method has also been used successfully with X_s^2 representations for dichotomous variables such as demographic characteristics consisting of two categories - a category of substantive interest (A) and a remainder (B) making up the total (Visvalingam, 1979; Visvalingam and Dewdney, 1977). In such cases, both the ratio and X_s^2 representations of the two categories are, by definition, 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.

There is some measure of negative correlation in categorised data (Chayes, 1971), and the proposed method exploits this closure effect for classifying such data. Attention is focussed here on classifications based on polychotomous demographic characteristics consisting of more than two categories. The population census provides data on a large number of demographic characteristics for a large set of geographical areas. Many of these characteristics are displayed in tables or sets of data. This study uses data on household composition, which is described by counts for 48 categories of household type.

As with bivariate classification, multivariate classification is here concerned with the magnitude of deviation from average characteristics and the direction of significant departures. The classification procedure involves dissection of the measurement space, undertaken for purposes of generalisation and data reduction to facilitate description. The resultant typologies are seen as different and extreme parts of a continuum in measurement space rather than as disjoint and discrete clusters. The classification is analytic rather than synthetic. The extremities of such distributions are of immense interest, particularly because their geographical distributions and aspatial characteristics are highly distinctive. However, the boundaries which demarcate them are inevitably arbitrary and artificial, as is often the case in univariate and bivariate classifications. They are analagous to serial class limits used for histograms and for statistical maps (Evans, 1977), whereas empirical classification procedures aim to produce boundaries analogous to idiographic class limits.

Emphasis is placed on the use of basic rather than sophisticated statistical procedures, so that a clear understanding of the implications of the various procedures for the producing and assessing the classification can be maintained. The proposed method can be used to process very large volumes of data for thousands of areal units with minimal core store requirements. It can also be used quite easily with mini-computers.

This method, like all others, is sensitive to the number and definition of categories, and draws attention to deficiencies in the definition of variables at the end, even if not at the start, of the classification procedure. When necessary, a classification of areas can be produced using a few tentative categories defined on the basis of substantive interest. The resulting classification can then be used to refine and extend the range of necessary categories and typologies.

2. DATA CHARACTERISTICS

2.1. Data used, the data source and the study area

Clarke (1977), in his review of the classification of population types and regions stated that

> in spite of the fact that classification for clarification is one of the first major steps in most sciences, it has not developed very far in either demography or population geography in either establishing typologies or in regionalization

He went on to suggest that the analysis of individual demographic characteristics, consisting of several categories, was likely to yield better results than classification procedures involving several demographic 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. Clarke pointed out that even some individual characteristics, such as age or occupation, are polychotomous and involve extensive data sets which are unsatisfactorily summarised by average values, by single-figure indices or by grouping.

In the present study, a single characteristic, namely household composition, which in the census is described by 48 categories of household type, is selected for classification. The ulimate aim of the classification is to summarise the household composition data for each areal unit by codes which reflect the composition of household types. Household

composition was chosen because the data on household type (Small Area Statistics, Table 20) indicate, either directly or indirectly, the age, sex and marital status of adults and the number and ages of children in the 48 categories of household (O.P.C.S., 1976). Data on age, sex and marital status for individuals (SAS Tables 4, 6 and 7) were employed to check the distinctiveness of the resultant typologies

Another reason for the paucity of classifications in demography is the definition of populations on areal or statistical rather than on demographic or sociological criteria. The resultant population units are not demographically distinctive, often lack basic internal homogeneity and are thus not optimal for classification purposes. 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, the one-kilometre grid square. The advantages and drawbacks of grid-square data have already been discussed elsewhere (Rhind, 1975; Visvalingam and Dewdney, 1977).

The 1971 one-kilometre grid square population census data for Great Britain, supplied by the Office of Population Censuses and Surveys, forms a gigantic data set (Visvalingam and Perry, 1976). The present study is concerned with a small subset of the data, covering the same area as that considered by Visvalingam and Dewdney (1977). The area comprises the three 100 km squares whose south-west corners are defined by the grid references 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. Unsuppressed household data are available for only 8,869 squares, 53.4 per cent of all 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)* Thus the analysis excludes remarkably few households despite dealing with little more than half the inhabited squares.

---

* The population figures include the entire resident population although the classification is based on data for private households only.

2.2. Variables for producing and evaluating the classification

The classification is based on counts for household composition. Earlier attempts at classification, based on data for persons, cross-tabulated by age, sex, marital status and private or non-private households, presented problems. Such data not only violated the conditions of statistical independence - since the distributions of members of households are obviously 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 of married persons. Owing to the complex relationships between persons, a large number of categories of persons was found necessary, and the interpretation of the ecological relationships between different categories of persons within one-kilometre squares was subjective, speculative and unsatisfactory.

For these reasons, data for household composition were preferred. While the latter data set is far from ideal (see below), it does provide a more explicit indication of household structure, since various categories of adults are cross-tabulated against the numbers and ages of children. The 48 primary variables 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 parentheses. Table 2 gives the corresponding data for the study area.

The O.PC.S. definition of the terms "adult" and "child" as used in S.A.S. Table 20 of the 100% Household data requires some explanation. Only persons below the age of 15 were counted as children and all those aged 15 and above were included in the adult category. Thus it is not surprising that households with no children comprise some 64 per cent of the total. It should also be noted that the pensionable categories are based on age criteria rather than actual retirement. Pensionable persons are all males aged 65 or over and all females aged 60 or over. Consequently, non-pensionable persons are all males aged 15 to 64 and females aged 15 to 59. While the majority of adult types are clearly defined, the O.P.C.S. definition of " $\geqslant$ 2 others" is somewhat obscure This category was assumed to include at least two non-pensionable and at least one pensionable adult and represents a residual category.

TABLE 1. HOUSEHOLD TYPES : GREAT BRITAIN TOTALS

Adults (aged 15 or over) in Household	Household Type						
	No child	Children aged 0-14 in Household					
		One child		Two or more children			
		0 - 4	5 - 14	all 0 - 4	all 5 - 14	others+	Total
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 (100.00)

TABLE 2. HOUSEHOLD TYPES : TOTALS IN THE STUDY AREA

Adults (aged 15 or over) in Household	Household Type						
	No child	Children aged 0-14 in Household					
		One child		Two or more children			
		0 - 4	5 - 14	all 0 - 4	all 5 - 14	others+	Total
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)

FIGURE 1 : DERIVATION OF SECONDARY VARIABLES

<table>
<tr><td rowspan="2">Adults (aged 15 or over) in Household</td><td colspan="6">HOUSEHOLD TYPE</td></tr>
<tr><td>No child</td><td colspan="2">One child</td><td colspan="3">Two or more children</td></tr>
<tr><td></td><td></td><td>0-4</td><td>5-14</td><td>all 0-4</td><td>all 5-14</td><td>others+</td></tr>
<tr><td>One pensionable male</td><td colspan="6">VARIABLE 1</td></tr>
<tr><td>One pensionable female</td><td colspan="6">VARIABLE 2</td></tr>
<tr><td>Two or more, all pensionable</td><td colspan="6" rowspan="2">VARIABLE 3</td></tr>
<tr><td>Two or more,one not pensionable</td></tr>
<tr><td>One other male</td><td colspan="6">VARIABLE 4</td></tr>
<tr><td>One other female</td><td colspan="6">VARIABLE 5</td></tr>
<tr><td>Two or more, others*</td><td rowspan="2">VARIABLE 6</td><td colspan="2" rowspan="2">VARIABLE 7</td><td colspan="3" rowspan="2">VARIABLE 8</td></tr>
<tr><td>Two or more,none pensionable</td></tr>
</table>

* = 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)

The great majority of children are found in households with two or more non-pensionable adults, followed by households designated " $\geq$ 2 others". There are also substantial numbers of households with children under the care of one non-pensionable female.

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 these very small counts. The aggregation procedure was recognised as tentative and arbitrary, reflecting not only the investigator's interest and judgement as to which categories of household were of substantive importance but also the constraints imposed by practical considerations. Since the present analysis was exploratory and a second run through the data was desirable, it was felt that this exercise would provide an opportunity for assessing the effects of numerically or statistically inappropriate combinations of categories of data.

The choice of eight secondary variables is decidedly arbitrary but reflects other practical and technical considerations associated with mapping. On small-scale maps showing one-kilometre squares, particularly for the whole of Great Britain,it is difficult to distinguish clearly between more than seven coloured classes. Given an eight-fold categorisation of the data, the classification procedure proposed in this paper would yield nine groups. It was felt that an assessment of the spatial and aspatial characteristics of nine groups would provide a better basis for the eventual reduction (or expansion) of the number of groups. Just as too few initial categories would clearly be inadequate, the use of too many would present problems in the classification procedure, for reasons which will become obvious when the results of the analysis are discussed. Consequently, eight was chosen as a reasonable and tentative compromise.

In devising the eight-fold categorisation shown in Figure 1, several considerations were taken into account. The primary aim was to retain the demographic information on the 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 categories. Thirdly, it was recognised that the enumerated data refer to the composition of

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)

households on census night. Thus the counts, admittedly low, for households with children in the care of pensionable adults or one non-pensionable male may include temporary rather than permanent arrangements. Consequently, in the eight-fold categorisation, 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 (Figure 1) as one category and 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 variables thus defined include all private households with adults and are mutually exclusive. Since there were several subjective decisions involved in the derivation of the secondary variables, the 48 primary variables were used at a later stage to evaluate both the classification and the derivation of secondary variables. In addition, data on the age sex and marital status of individuals, derived from S.A.S. Tables 4, 6 and 7, were used to interpret and thereby further evaluate the classification. No further interpretations are made of the typologies in this paper. More exhaustive cross-tabulations with other demographic, socio-economic and housing variables will be made for the final classification at the Great Britain level.

2.3. General characteristics of household types in the study area

Table 3 shows the composition of the eight categories of household type in Great Britain and in the study area. Although the disparities between the two are quite small in proportional terms, it must be remembered that ratio values tend towards near-average figures in large populations. The differences are in fact large in terms of the numbers of people and households involved. Variable 6 (several other adults with no child) constitutes well over a quarter of all households both in Great Britain and in the study area. The study area is somewhat deficient in this type of household and also in types 4 and 5 (households with one non-pensioner adult), but has an excess of one-pensioner households (variables 1 and 2) and households with several non-pensioner adults with children (variables 7 and 8). The household composition is dominated by "several other adult" households, which form 63 per cent of the total. Pensioner households are nearly 30 per cent of all households, while households with one non-pensioner adult form only

7.5 per cent of the total.

3. METHODOLOGY

3.1. Numerical representation of the data

The data matrix consists of N x M elements, where N is the number of variables, namely eight, and M is the number of data units, in this case 8,869. This matrix of observed frequencies was standardised with respect to row and column totals, using the X_s^2 representation. (For a fuller discussion of the choice of X_s^2 as opposed to ratio variables, see Visvalingam, 1979 and Visvalingam and Dewdney, 1977). The expected frequencies were calculated in the usual way as

$$Eij = \frac{Ci.\ Rj}{N} \qquad (1)$$

where : Eij is the expected frequency for variable i in data unit j,

N is the total number of households in the study area

Ci is the total number of households of type i in the study area,

Rj is the total number of households in the grid square

For ease of computer processing, data for each grid square were processed separately and the data for each variable were standardised to X_s^2 scores as follows :

$$\sum_{i=1}^{N} X_s^2 = \frac{(Oi - Ei)^2}{Ei} \cdot \text{sgn}\ (Oi - Ei) \qquad (2)$$

The X_s^2 value for each variable was a numerical measure of the extent to which its occurrence 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_s^2 scores, where

$$X_s = \text{sgn}\ (X_s^2) \sqrt{X_s^2} \qquad (3)$$

is useful for plotting each data unit onto an eight-dimensional measurement space, the centre of which has zero deviations for all variables. The squared distance of the data unit from this centre - i.e. its deviation from average expectation - would be derived by

$$X^2 = \sum_{i=1}^{N} \left| X_s^2 \right| = \sum_{i=1}^{N} \left[\frac{(Oi - Ei)^2}{Ei} \right] \qquad (4)$$

3.2. Classification procedures

While demographic types have been recognised and studied, there is an absence of an a priori logical or philosophical theory to prescribe the number, let alone the nature of distinct and transitional demographic types, The present study is thus exploratory and is amenable to heuristic refinement.

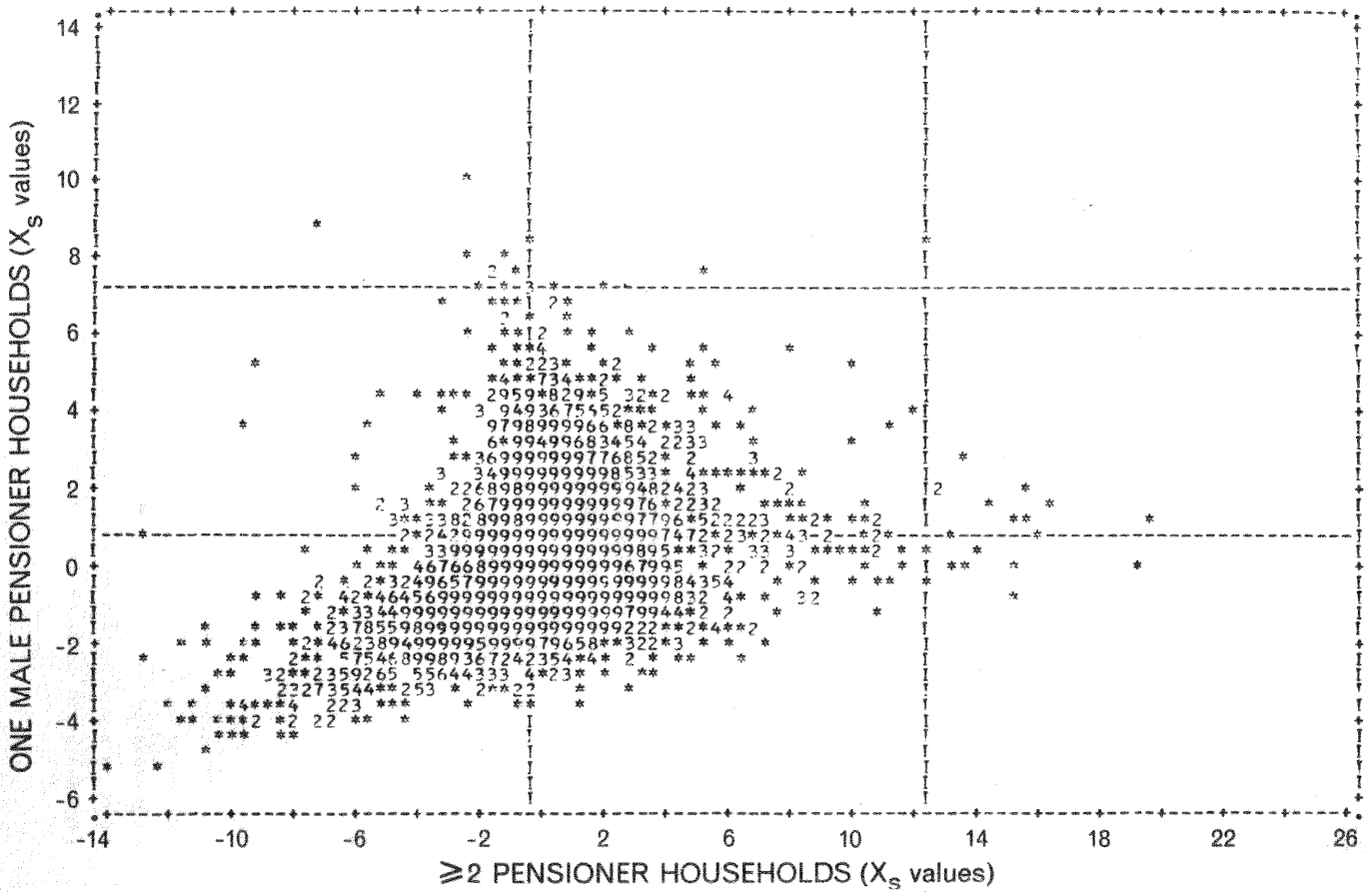
Everitt (1974) described the clustering techniques in common use and discussed the various problems associated with their use. Optimization techniques for partitioning the set of data units require excessive amounts of computer time and resources, and consequently are not recommended for use with large data sets. In addition, there are several problems associated with the assumptions regarding the shape and spacing of clusters. Mode analysis (Wishart, 1969) was also rejected for its assumption of spherical clusters or contiguous spheres and emphasis on disjoint density surfaces (for reasons, see below). The Cartet count method described by Cattell and Coulter (1966) was more appealing for its simplicity and lack of assumptions concerning the configuration of clusters in measurement space. Essentially, this method consists of partitioning the multi-dimensional space and counting the number of points in each cartet (or hypercube). While this gives some description of the distribution of data units in measurement space, the rectilinear dissection procedure has a tendency to segment groups and can result in an unmanageable number of cartets or classes; the number of cartets (k) reflecting the function $k = c^v$, where v is the number of variables or dimensions and c is the number of classes on each axis of variation. Thus, even if the eight household types were classified as being above or below average, the

dissection procedure would result in 256 cartets.

The various X_s scatterplots indicate that there is basically only one constellation of data units. Data units with average and small populations are located in or near the dense central core, while the more interesting data units with a marked excess or deficit of one or more household types are to be found in the outer, more diffuse parts of the hypersphere. This is especially noticeable in the bivariate plot of two or more pensioner households (variable 3) against one male non-pensioner households (variable 4). Note that in Figure 2 the more extreme deviations tend to have preferred directions and are elongated parallel to the two dimensions. The elongation towards the third quadrant, i.e. to the bottom left, reflects areas where both variables 3 and 4 are deficient and where there is consequently an excess of some other category of household type such as large households (variable 8). The X_s scatter-plot of variable 3 against variable 1 (one male pensioner households) exhibits similar trends (Figure 3).

Clustering procedures based solely on distance measures or density functions were considered inappropriate for classification of the observed distributions. Instead, a dissection procedure was adopted. The N-dimensional measurement space was initially separated into two major sectors, the 'average' and the 'outer' more extreme sectors. A X^2 value (equation 4) of 14.1, which conventionally corresponds to the 95% significance level at 7 degrees of freedom, was an arbitrary value chosen to demarcate the 'average' and 'outer' sectors. It must be stressed that such a procedure pools genuinely 'average' data units with others in which the numbers of households are too few to justify any statistically valid conclusions. The initial designation of an average class is a temporary expedient and relocation of some of these data units needs to be considered.

The 'outer' sector contains the more interesting departures from expectation. This was further partitioned into N sub-sectors so that each sub-sector contained data units with a marked excess of a particular household type. This 'distinctive' category is identified as possessing the maximum X_s value in the 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


ONE MALE PENSIONER HOUSEHOLDS (X_s values)
14
12
10
8
6
4
2
0
-2
-4
-6
-14
-10
-6
-2
2
6
10
14
18
22
26
≥2 PENSIONER HOUSEHOLDS (X_s values)

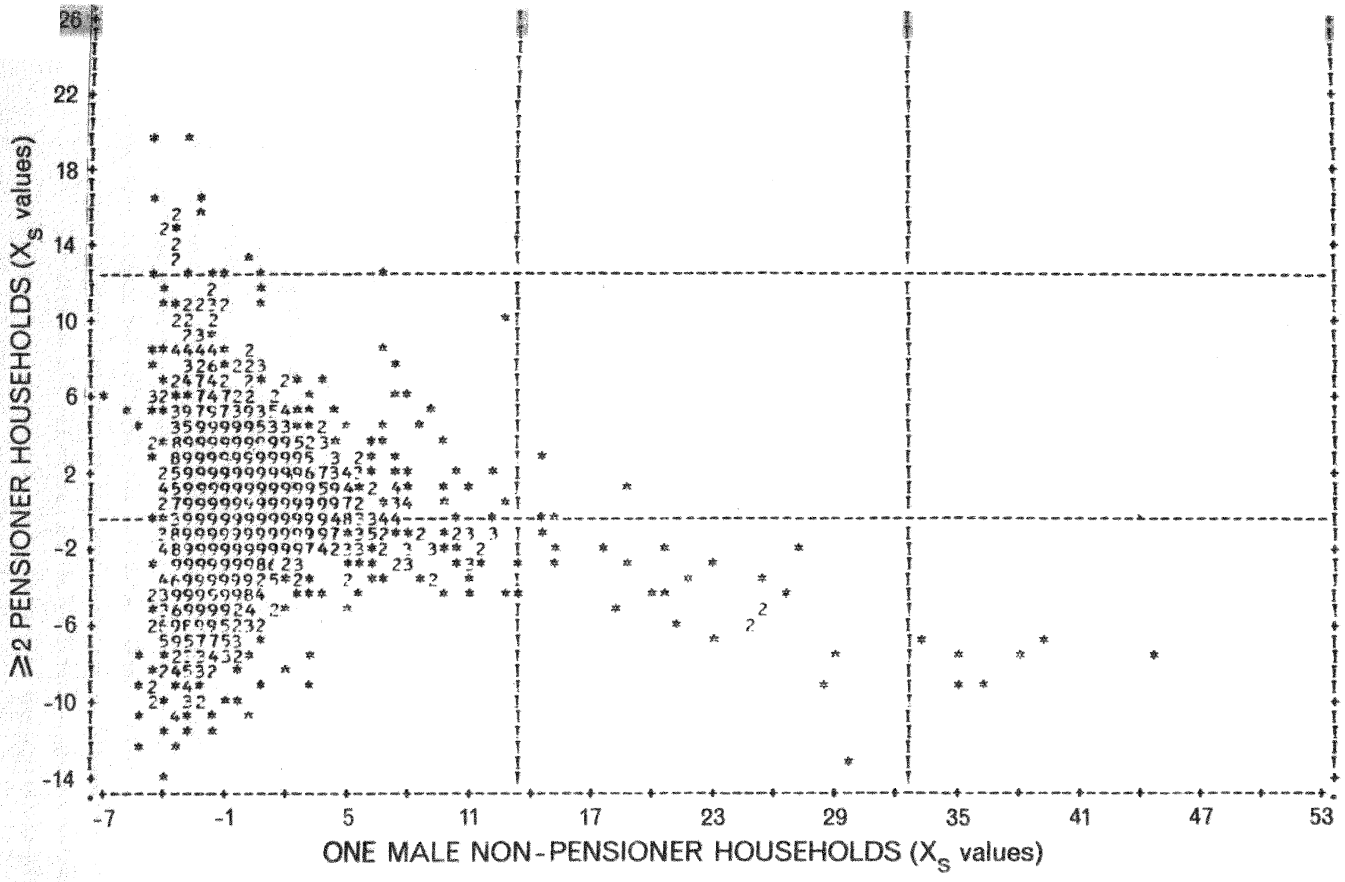


Fig. 3. Scattergram of X_s values for variables 1 and 3

allocation of data units to one of the two possible sectors in a random and unbiased fashion.

The use of the maximum X_s value exploits the closure effect in categorised data (Chayes, 1971). The proposed method is consistent with the procedure used for the classification of bivariate data (Visvalingam, 1979; C.R.U., forthcoming). In the bivariate case, the closure effect results in an inverse relationship between the two categories, both in ratio and X_s terms. 'Average' values were separated from deviant 'outer' ones by a X^2 value of 3.84, an arbitrary value which conventionally is 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 for determining the variable or 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 demographic type. The only certainty concerning the above dissection procedure is that there is a marked excess of a particular household type, the 'distinctive' category, within each data unit and consequently within the group of data units. There is no implication of the absolute predominance of the attribute. If the 'distinctive' category is positively correlated with some other category or categories, there would be a tendency towards a corresponding excess of these other categories in the group. If, on the other hand, there is a negative correlation between the 'distinctive' category and some others, there would be a deficit of the latter in the group. Table 4 gives the general correlations between the variables and the slopes of their regressions. This does not imply that the pattern of excess and deficit in group characteristics would apply to each data unit within a group.

The dissection procedure is valid so long as pairs of variables are not strongly positively correlated or off the diagonals in the first quadrant of bivariate scatterplots. The regression coefficients in Table 4 and scatterplots suggest that the only relationship likely to violate this condition is that between male and female households of one non-pensioner (variables 4 and 5). The procedure is extremely sensitive to the number, selection and definition of initial categories. Thus refinement of the classification is concerned as much with evaluating the suitability of the categorisation as with the identification and adjustment of non-homogeneous and non-unique types.

TABLE 4 : MATRIX OF CORRELATION COEFFICIENTS AND REGRESSION SLOPES BETWEEN THE EIGHT VARIABLES

		Correlation coefficients							
		V1	V2	V3	V4	V5	V6	V7	V8
Regression coefficients	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.

The eight-fold categorisation was proposed on the merit of substantive interest. Thus, despite the indications of correlation and regression coefficients, the categories were retained : and the data classified. The aim was to observe the aggregate characteristics of groups pertaining to each of these 'distinctive' categories. The uniqueness and homogeneity of these groups would be ascertained and the geographic distributions portrayed. Data on age, sex and marital status of the resident population, which were not used in the initial classification, will also be analysed to ascertain whether differences between groups persist with respect to these variables.

The classification procedure is not only simple in concept but also economical in the use of computer core and time. Since data units are processed individually, the method can cope with very large data sets, even using mini-computers. The maximum X_s and X^2 measures (equations 3 and 4) were used as indices of the direction and magnitude of deviation respectively. The distance measure was used to separate 'average' from 'outer' deviant data units and the directional measure was used to subdivide the latter. The resulting groups are as follows :

Group Name	'Distinctive' category when $X^2 > 14.1$
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.

4. AGGREGATE CHARACTERISTICS OF GROUPS

4.1. General Indications

The spatial distribution of data units within the different groups is portrayed in Figures 4,5 and 6; these exhibit contiguity and suggest a spatial sorting of household types. The aggregate frequencies of household types in each group are given in Table 5, and Table 6 gives the proportions of different household types within each group. In Table 7, the proportions of each household type found in the different groups are tabulated to give some indication of the degree and direction of sorting of each household type.

The aspatial summary statistics of the groups and their spatial distribution suggests that there are three sets of deviant groups, namely those characterised by a concentration of pensioners (B, C,D), single non-pensioner households (E,F) and several other adult households (G,H,I). Tables 6A and 7A suggest that the major differenceis between GHI and the others; variables 6, 7 and 8 (several other adults) form approximately 63 per cent of all households (Table 3) and together remain the predominant household types in all groups (Table 6A). The relative concentration of this set ranges from below 60 per cent in groups B to F to over 70 per cent in G,H and I, forming nearly three-quarters of the households in H and I. Consequently, G,H and I record the lowest proportions of pensioners and one non-pensioner households.

The remaining groups, B to F, can be further subdivided. B,C and D record relatively high proportions of pensioner households, while E and F contain markedly above-average proportions of one non-pensioner households. It must be stressed, however, that the proportions of both types of households are above average in B to F. Thus, while there seems to be a clear distinction between the aggregate characteristics of G,H and I and the rest, the distinction between B, C and D on the one hand and E and F on the other is not as sharp. D (several pensioners) and E (one other male) seem to be at the opposite ends of a continuum. D and E also exhibit distinctive spatial distributions, while date units in B,C and F do not exhibit spatial contiguity to the same extent and appear to be transitional.

TABLE 5 : FREQUENCY DISTRIBUTION OF THE EIGHT HOUSEHOLD TYPES AMONG THE NINE GROUPS

Group	Household Types								Total house-holds	Total km squares	House-hold density	% of house-holds	% of data areas
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)					
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

G, H and I (several other adults) contain 37 per cent of all households (Table 7A) but they include over 40 per cent of variables 7 and 8 (households with children). At the other extreme B,C and D contain 34 per cent of all households, but include over 42 per cent of the pensioner households. While E and F contain only 15 per cent of the households, they include nearly 32 per cent of variable 4 and 24 per cent of variable 5. There are somewhat more one-other male households in E and F with 15 per cent of all households, than in B,C and D which together include 34 per cent of the households. Note also that the male non-pensioner households in E and F are concentrated spatially within 7 per cent of the data areas, while similar numbers are distributed over 19 per cent of the data areas in B, C and D (Table 5).

There is, however, some degree of variation within each of the above three sets of groups. For example, while the aggregate proportions of all two or more non-pensioner adult households were high in G, H and I (Table 6A), Table 7A indicates that variable 6 (several adults with no child) is somewhat deficient in the GHI set. The study of individual groups is not facilitated by tabulations of row and column proportions (Table 6B and 7B), let alone by the observed frequencies (Table 5). The pattern of variation within these tables can more readily be discerned using the distribution of X_s values given in Table 8. The X_s values indicate that variables 5 and 6 are sorted to a lesser extent than the other variables. Moreover, the distribution of variable 3 (several pensioners) is not coincident with that of variables 1 and 2 (one-pensioner households). While all three are concentrated in groups B,C and D, there is an excess only of the one-pensioner households in groups E and F, which are markedly deficient in variable 3. Thus a description of each group is pertinent.

4.2. Description of groups

The distribution of the 48 household types and the composition of the nine groups can be gauged from Tables 5 to 9 and Figures 7 to 9. The structures of Tables 5 to 8 have already been discussed. Aggregate frequencies of the 48 primary household types, cross-tabulated against the groups, were processed into X_s form for presentation in Table 9. Similarly, the number of people in each five-year age group was cross-

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	18.61	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

TABLE 8 : DISTRIBUTION OF X_{s} VALUES

Groups	Household Types							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
B	50.807	9.988	9.505	0.270	-3.247	-4.691	-6.230	-19.625
C	31.547	111.045	39.897	7.776	16.574	-30.209	-38.222	-68.836
D	9.520	36.128	108.360	-28.131	-11.547	-16.614	-37.210	-60.789
E	17.518	16.799	-17.631	146.715	65.236	-17.429	-37.895	-39.089
F	-0.551	11.879	-4.967	17.770	44.957	-6.330	-13.900	-12.207
G	-17.632	-41.901	-27.304	-28.241	-20.507	62.206	18.138	-7.017
H	-18.085	-49.146	-50.145	-27.488	-27.158	20.156	71.200	28.021
I	-41.068	-86.121	-89.666	-46.524	-23.171	-8.572	48.460	155.288

tabulated against marital status and sex. The latter data are displayed as age-sex pyramids in Figure 7 . Separate age-sex pyramids for 'single' (including widowed and divorced) and married persons are also presented in Figure 8 . Using the same structure, the corresponding X_s values are plotted in Figure 9 , which highlights the variations in Figure 8 . This section of the paper is concerned with the description rather than an explanation of group characteristics, and is directed towards detecting similarities between groups.

4.2.1. Group A : the central core of average data units

Group A is a collection of 4,974 data units which lie at the core of the X_s property space, with X^2 values $\leq$ 14.1. It includes grid squares with small populations as well as larger-population areas with near-average characteristics. Thus it contains over 56 per cent of the data units but only about 17 per cent of the households (Table 5). This group has an average household density of 99 households per one-kilometre grid square. On aggregate it contains a slight excess of variables 3,6 and 7 (several-adult households with one or no children) and a deficit of other household types.

4.2.2. Group B : marked excess of one-male pensioner households (variable 1)

Group B contains 5.46 per cent of the data areas but only 2.55 per cent of all households (Table 5). The low household density (149) and the dispersed geographical distribution of Group B suggest more rural situations and Group B does appear to occur in relative abundance in rual areas of England and Wales (Fig. 4). It must be emphasised that, in both absolute and ratio terms (Tables 5 and 6B), within Group B variable 1 exceeds only variables 4 and 5. As a result of its infrequency and the low household density, Group B includes only 5.95 per cent of variable 1 (Table 7B) and even smaller proportions of the other household types. Group B has an excess of pensioner households (Table 8), especially those with no children (Table 9). It has an excess of persons over the age of 35 (Fig. 9), of married people above 50 years of age and of married persons below the age of 25. It is relatively deficient in married people between the ages of 25 and 50 and in all age groups below 20.

TABLE 9 : DISTRIBUTION OF X_S VALUES FOR THE 48 HOUSEHOLD TYPES

	Groups							
	B	C	D	E	F	G	H	I
One male pensioner								
(i)	50.80	31.54	9.48	17.51	-0.55	-17.63	-18.05	-41.08
(ii)	- 0.36	1.27	2.59	- 0.76	-0.42	- 0.63	- 0.59	- 0.96
(iii)	1.60	0.85	0.95	0.93	0.31	- 0.69	- 0.90	- 1.16
(iv)								
(v)	- 0.48	- 0.41	-0.31	- 0.04	-0.56	- 0.84	- 0.79	2.61
(vi)	- 0.23	- 0.58	-0.55	- 0.48	-0.26	4.65	- 0.37	- 0.61
One female pensioner								
(i)	10.05	111.08	36.23	16.69	11.90	-42.00	-49.04	-86.11
(ii)	- 1.05	- 0.07	1.38	0.02	- 1.23	- 1.29	- 0.58	- 1.38
(iii)	- 0.55	1.56	-1.43	3.20	0.28	0.94	- 3.02	- 1.53
(iv)	2.45	1.27	-0.87	- 0.76	- 0.42	- 0.63	- 0.59	- 0.96
(v)	- 0.95	2.12	-1.24	- 0.09	0.31	0.83	- 1.09	- 1.65
(vi)	- 0.73	- 0.81	2.15	- 0.91	- 0.86	0.27	- 0.38	- 0.43
Two or more pensioners								
(i)	6.43	38.36	97.39	-22.32	- 9.27	-26.90	-38.98	-74.32
(ii)	- 0.46	- 2.60	1.19	0.32	3.97	- 1.75	- 0.64	- 0.41
(iii)	- 0.63	- 0.08	2.02	- 0.90	- 1.75	- 2.01	- 1.35	- 1.75
(iv)	- 0.50	- 0.52	1.22	- 0.14	- 0.59	- 0.89	- 0.84	1.60
(v)	- 1.20	- 0.79	-0.59	0.21	- 0.14	- 1.17	- 1.22	1.99
(vi)	- 0.81	- 0.65	1.56	- 0.0	- 0.95	2.77	- 0.61	- 1.27
Two or more adults, one not pensionable								
(i)	7.61	18.50	55.05	- 3.17	1.96	-10.92	-31.19	-52.94
(ii)	- 1.05	1.61	4.20	0.41	- 1.11	- 1.13	- 3.07	- 0.49
(iii)	- 0.97	1.04	5.40	2.69	2.10	- 2.74	- 4.91	- 4.12
(iv)	0.50	1.20	1.17	0.66	- 1.38	- 1.24	- 1.48	0.25
(v)	- 1.48	- 1.99	4.56	0.89	0.22	- 2.19	- 2.47	0.76
(vi)	- 0.46	- 2.88	2.70	1.35	3.58	- 2.18	- 1.08	0.07

	Groups							
	B	C	D	E	F	G	H	I
One non-pensioner, male								
(i)	0.07	8.46	-26.71	150.08	17.66	-28.87	-28.30	-49.52
(ii)	- 1.14	0.22	- 4.35	1.65	0.38	- 2.27	0.24	2.22
(iii)	0.09	-0.27	- 4.91	5.53	- 0.21	1.23	1.12	0.35
(iv)	1.89	-0.30	- 1.63	0.40	0.13	0.09	0.64	1.91
(v)	0.61	-1.65	- 4.94	2.21	3.92	- 0.15	- 1.38	3.18
(vi)	0.93	-1.40	- 3.90	1.22	1.02	- 2.98	- 1.17	6.37
One non-pensioner, female								
(i)	- 1.43	20.36	- 2.37	61.22	33.75	-15.81	-26.11	-38.08
(ii)	- 1.49	2.43	- 6.76	20.86	12.66	- 7.18	- 2.94	- 3.29
(iii)	0.41	4.07	- 5.20	10.59	13.12	- 4.43	- 4.38	- 1.02
(iv)	- 2.88	-1.24	- 7.84	11.97	11.57	- 6.34	- 1.68	5.35
(v)	- 1.79	-3.46	- 7.48	4.48	14.03	- 5.52	- 5.30	14.25
(vi)	- 3.25	-2.72	-12.11	14.65	17.08	- 7.49	- 6.41	15.71
Two or more adults, some pensioners, more than one non-pensioner								
(i)	1.48	-7.61	22.75	-2.38	0.24	7.16	-19.51	-17.74
(ii)	- 0.35	-6.30	1.93	-1.03	- 3.01	0.02	1.62	3.70
(iii)	- 3.51	-8.44	11.23	-5.12	- 3.19	- 0.55	2.81	2.08
(iv)	- 0.97	-3.24	0.09	-1.11	1.17	- 1.33	- 1.22	3.91
(v)	- 4.56	-9.44	6.35	-5.80	- 1.45	- 3.47	- 5.80	11.35
(vi)	- 1.69	-9.14	1.28	-0.24	0.10	- 5.12	- 3.53	11.67
Two or more non-pensioners								
(i)	- 5.61	-29.45	-26.65	-17.77	- 6.88	63.95	29.19	- 2.31
(ii)	- 1.25	-20.38	-38.92	-15.16	- 7.12	2.00	74.84	26.83
(iii)	- 6.25	-30.85	-21.38	-36.68	-11.31	22.88	34.18	41.60
(iv)	- 4.89	-21.26	-34.48	-10.13	- 3.67	- 7.55	30.86	52.42
(v)	-15.57	-45.77	-30.29	-36.59	-12.22	3.16	9.01	101.00
(vi)	-10.86	-45.34	-45.99	-17.19	- 4.52	- 7.63	18.14	106.22

Key : (i) - no child

(ii) - 1 child aged 0-4

(iii) - 1 child aged 5-14

(iv) - 2 or more children all aged 0-4

(v) - 2 or more children all aged 5-14

(vi) - 2 or more children - at least one aged 0-4 and at least one aged 5-14

4.2.3. Group C : marked excess of one-female pensioner households (variable 2)

The average household and population composition of Group C is broadly similar to that of Group B. It has an excess of pensioner and one non-pensioner households. There is also a marked excess of households with one child and one non-pensioner female. The general pattern of excess and deficit of population constituents (Fig. 9) is quite similar to that of Group B, with an excess of older age groups and a small excess of young couples. Group C does not exhibit marked spatial contiguity, but shows some concentration in the smaller holiday resorts of North Wales and in the older industrial areas of Blackburn, Burnley, Preston, the eastern part of the Manchester conurbation and the Potteries (Fig. 4). The group includes 17 per cent of all households and 6 per cent of the data areas, the highest density of 885 households per one-kilometre square suggesting essentially urban locations.

4.2.4. Group D : marked excess of households with two or more pensioners (variable 3)

Group D is particularly noteworthy for the excess of older people and a deficit of younger adults and children (Fig. 9). Note, in Figure 7 , that the minor modes for the 20-24 age group and for pre-school children present in B and C is absent from Group D. Figure 8 suggests that this is due primarily to the progressive decrease in the numbers of married persons below the age of 50 and a marked deficit of young couples aged 20-24. Group D exhibits marked spatial contiguity and includes 8 per cent of the data areas. It extends over much of the coastal zones of North Wales and Lancashire and is also found in the higher status areas of Liverpool (Webber, 1975), St. Helens and south Manchester. A more dispersed distribution of Group D is found in the Potteries and some Rural Districts (Fig. 4). The group includes 15 per cent of households and has a household density of 611, somewhat lower than in Group C. While Group D is generally deficient in non-pensioner households, it has an excess of "other" households, both with and without children (Table 9), a point to be discussed later.

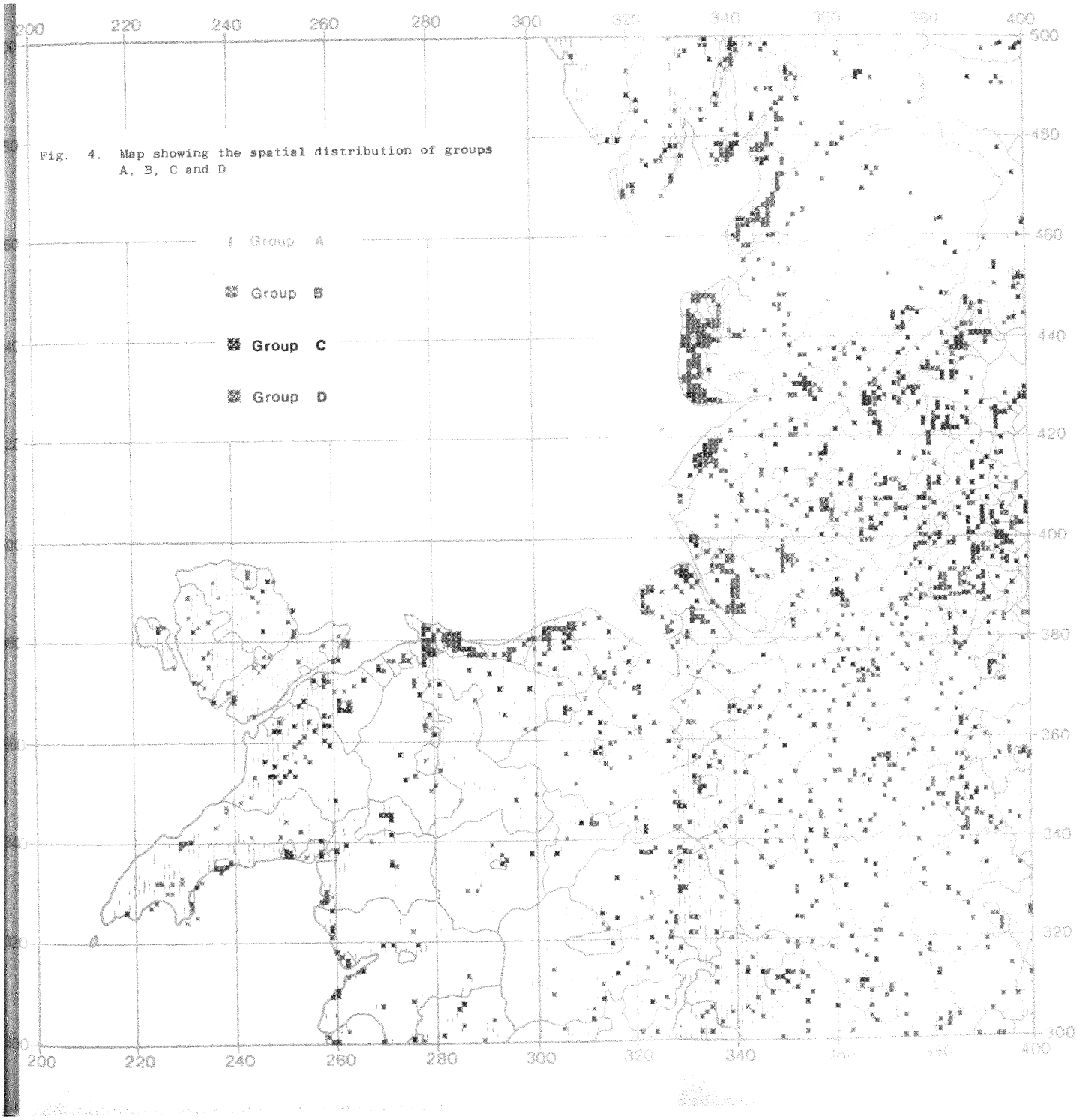


Fig. 4. Map showing the spatial distribution of groups A, B, C and D

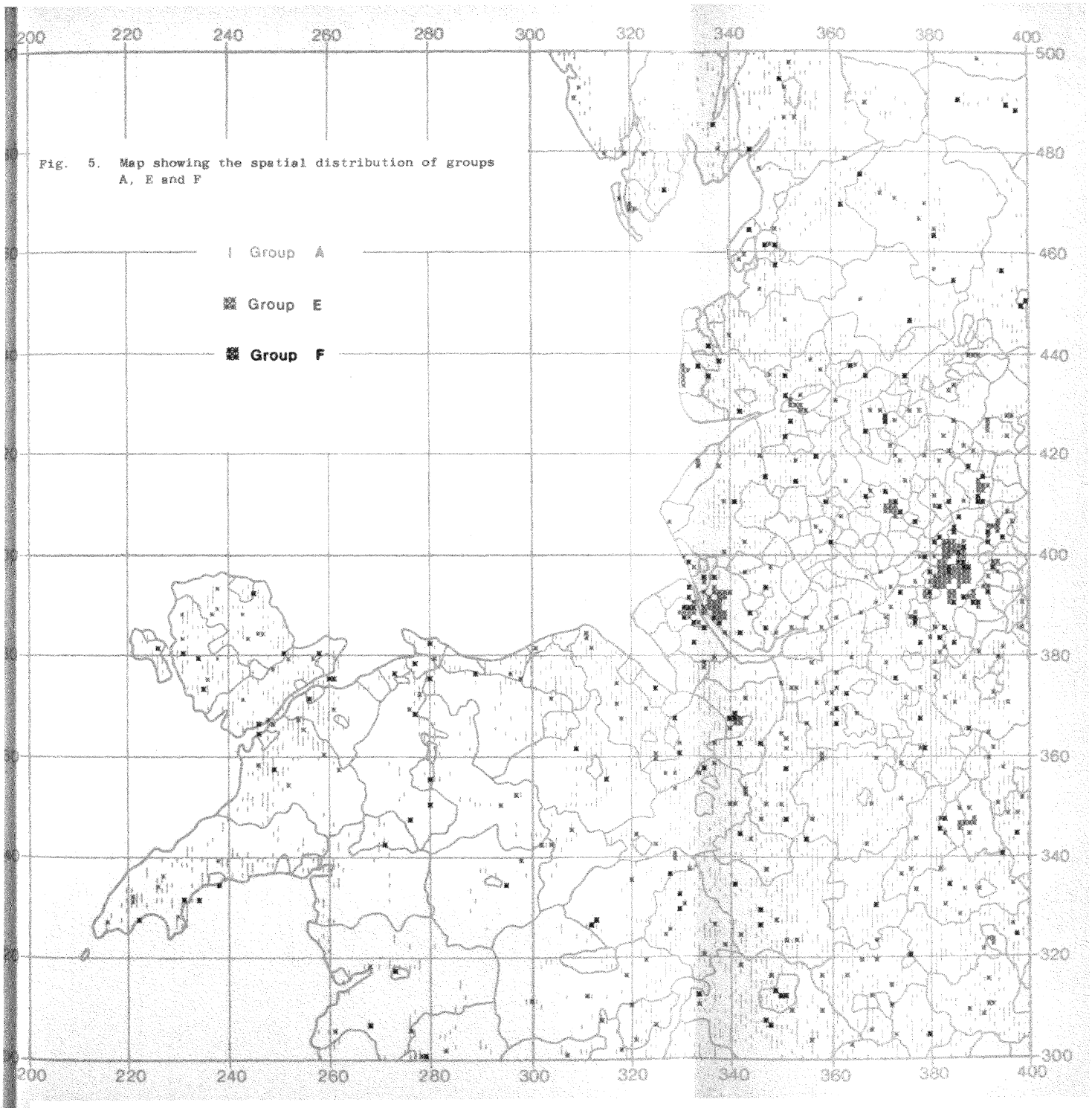


Fig. 5. Map showing the spatial distribution of groups A, E and F

Fig. 6 Map showing the spatial distribution of groups A, G, H and I

4.2.5. Group E : marked excess of one-male non-pensioner households (variable 4)

Group E is particularly notable for the preponderance of one-person households, both non-pensioner and pensioner (Table 8), male and female. With only 5 per cent of the data areas, this group contains 11.5 per cent of all households (Table 5) and 27 per cent of one male non-pensioner households (Table 7B). Average household density is high (738) and Figure 5 shows a marked concentration within inner areas of cities, such as Manchester, Liverpool, Birkenhead, Chester, Stoke, Oldham, Rochdale, Bolton, Preston and Burnely. There is also a pronounced concentration along the seafront in Blackpool and a dispersed distribution in rural areas.

There is a higher concentration of variable 5 (one female non-pensioner households) in Group E than in Group F and, while the outstanding excesses are in no-child households, this group is above average in one-parent households, especially those with one child under the care of a woman. While there is an excess of single people over 15 years in general, there is a marked excess of young adults, aged 20-24. (Figs. 7 and 9) and middle-aged single males (40-64 years). Group E is deficient in married people over 25 and in children of school age, but there is a small excess of pre-school children.

4.2.6. Group F : marked excess of one-female non-pensioner households (variable 5)

Group F shows a similar pattern, though with a different ordering of magnitude, of excess and deficit of population components to that of Group E. It has an excess of one-pensioner and non-pensioner households, of one-parent families and of pre-school children. Note, however, that it includes only 2 per cent of the data units and 3.5 per cent of the households, displaying a dispersed distribution (Fig. 5) except where it occurs in conjunction with Group E in inner city locations.

4.2.7. Group G : marked excess of childless households with several adults (varible 6)

It should be noted that variable 6, the largest single household type, is not sorted to the same extent as other household types (see X_s values in Table 8). Thus, while Group G contains 9.8 per cent of variable 6,

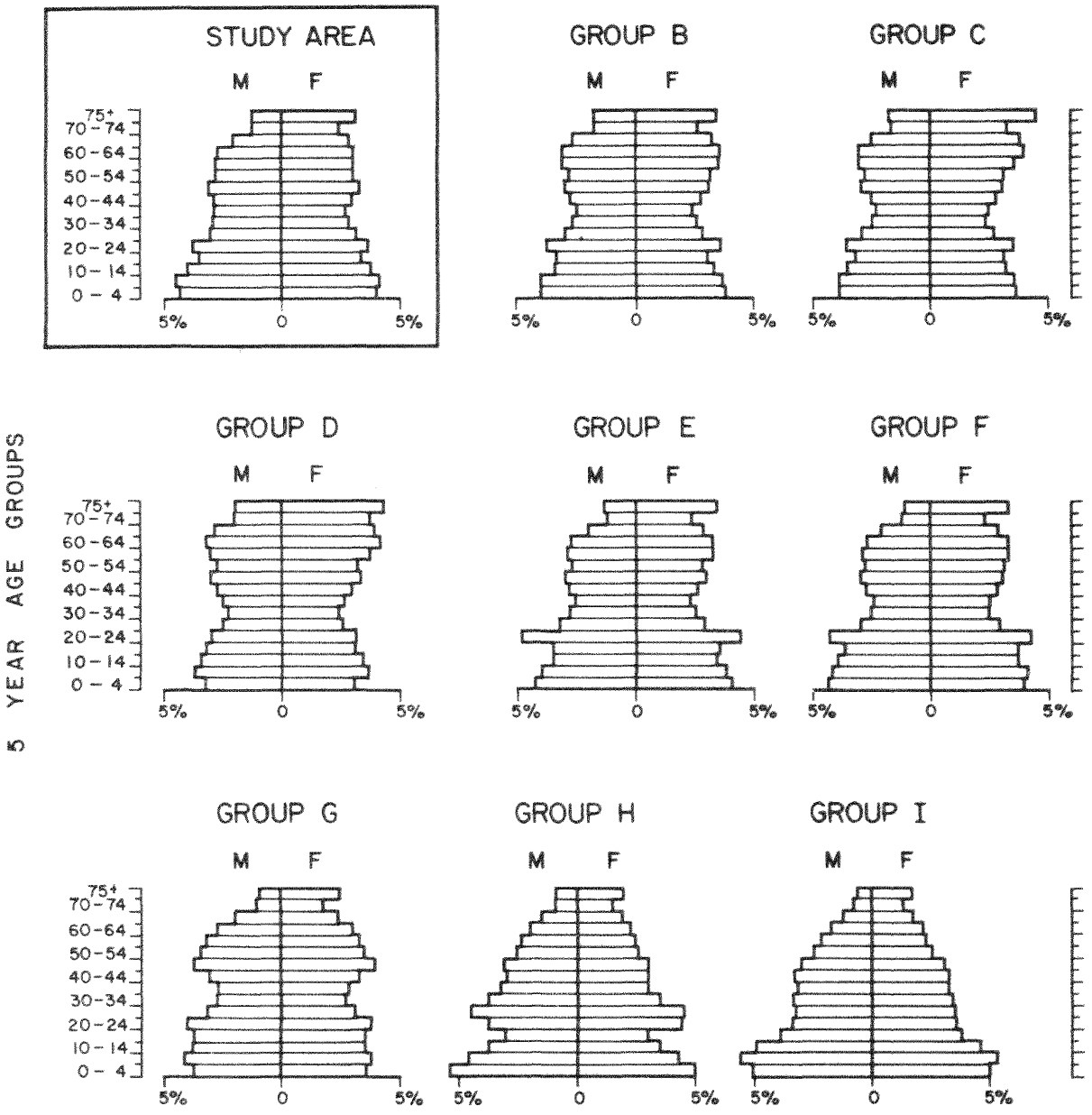


Fig. 7. Percentage distribution of population by age and sex within the study area and within each of the groups

the bulk of this household type is found in Groups I and A (Table 7B). Group G is particularly noteworthy for the excess of married people between the ages of 40 and 59 (Figs. 7 to 9), but also shows a small excess of slightly younger married couples; there is a deficit of older and very young couples. There is also an excess of single adults between the ages of 15 and 24, who are probably the non-dependent children of the middle-aged couples, while single people of all other age groups are deficient. Note the very distinctive age-sex pyramids for Group G (Fig. 7 and 8), especially that for married persons.

4.2.8. Group H : marked excess of one-child households with several adults

Group H includes nearly 7 per cent of all households and more than 4 per cent of the data units. Like Group G, it has a relatively low household density, suggesting more dispersed dwellings, possibly including areas with large gardens, and a lower degree of multiple occupancy. There is an excess of variables 6,7 and 8 and of married people between the ages of 20 and 39. There is a general lack of single adults though children, especially those under the age of 5, are numerous. However, the largest concentration of 0 to 4 - year olds is in Group I. Nevertheless, the mix of demographic components in Group H is very different from that in Group I as the age-sex pyramids in Figures 7 and 8 clearly demonstrate. There are, for example, major differences in the proportions of young married people and children.

4.2.9. Group I: marked excess of households with several adults and two or more children (variable 8)

Group I is the modal type, both in its coverage of data units and in terms of the proportion of households (Table 5), and has a fairly high average household density of 663. There is a marked excess of households with several children, whether these are under the care of one or several adults (Table 9). Group I contains the great bulk of the children in the study area and has an excess of young school-leavers and married adults aged 25-44. The modal age groups for adults are 30-44, which is somewhat older than in Group H. While Groups G and H are more striking for their excess of married people, Group I is outstanding for the concentration of children (Fig. 9). The age-

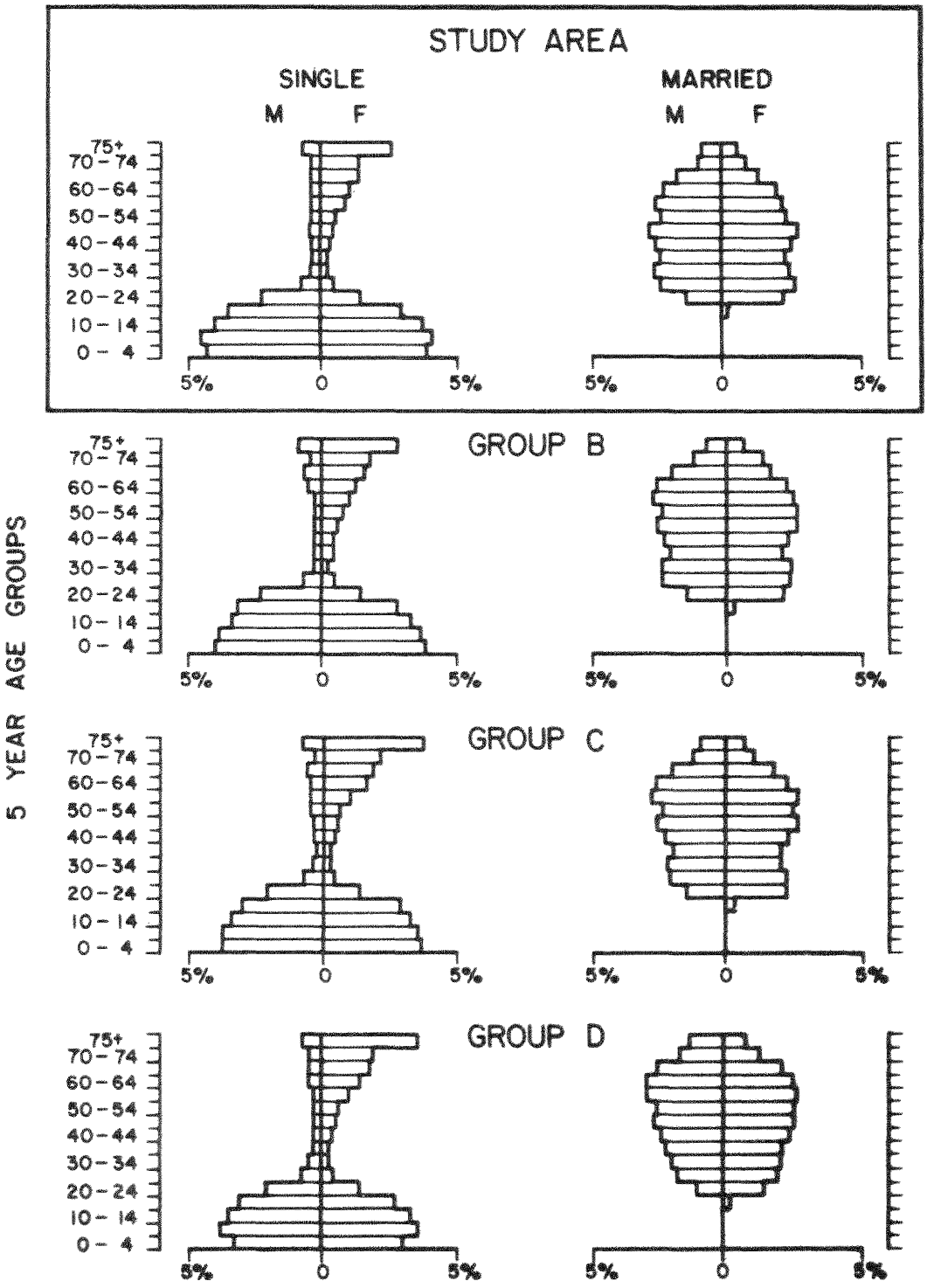


Fig. 8. Percentage distribution by age and sex of single and married people within the study area and within each group

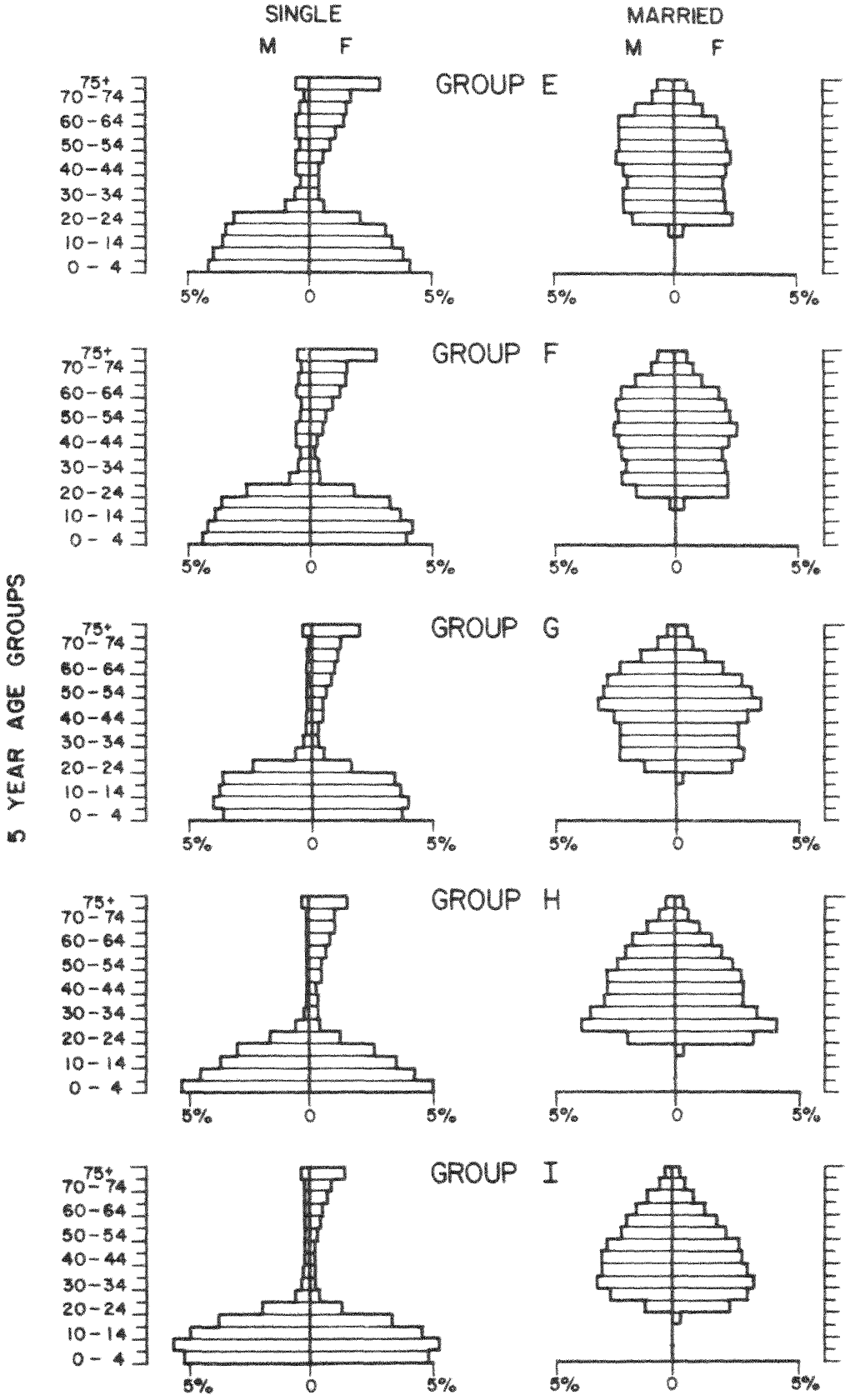

SINGLE
M F
MARRIED
M F
GROUP E
GROUP F
GROUP G
GROUP H
GROUP I
5 YEAR AGE GROUPS
75+
70 - 74
60 - 64
50 - 54
40 - 44
30 - 34
20 - 24
10 - 14
0 - 4
5%
0
5%

Fig. 8. (contd)

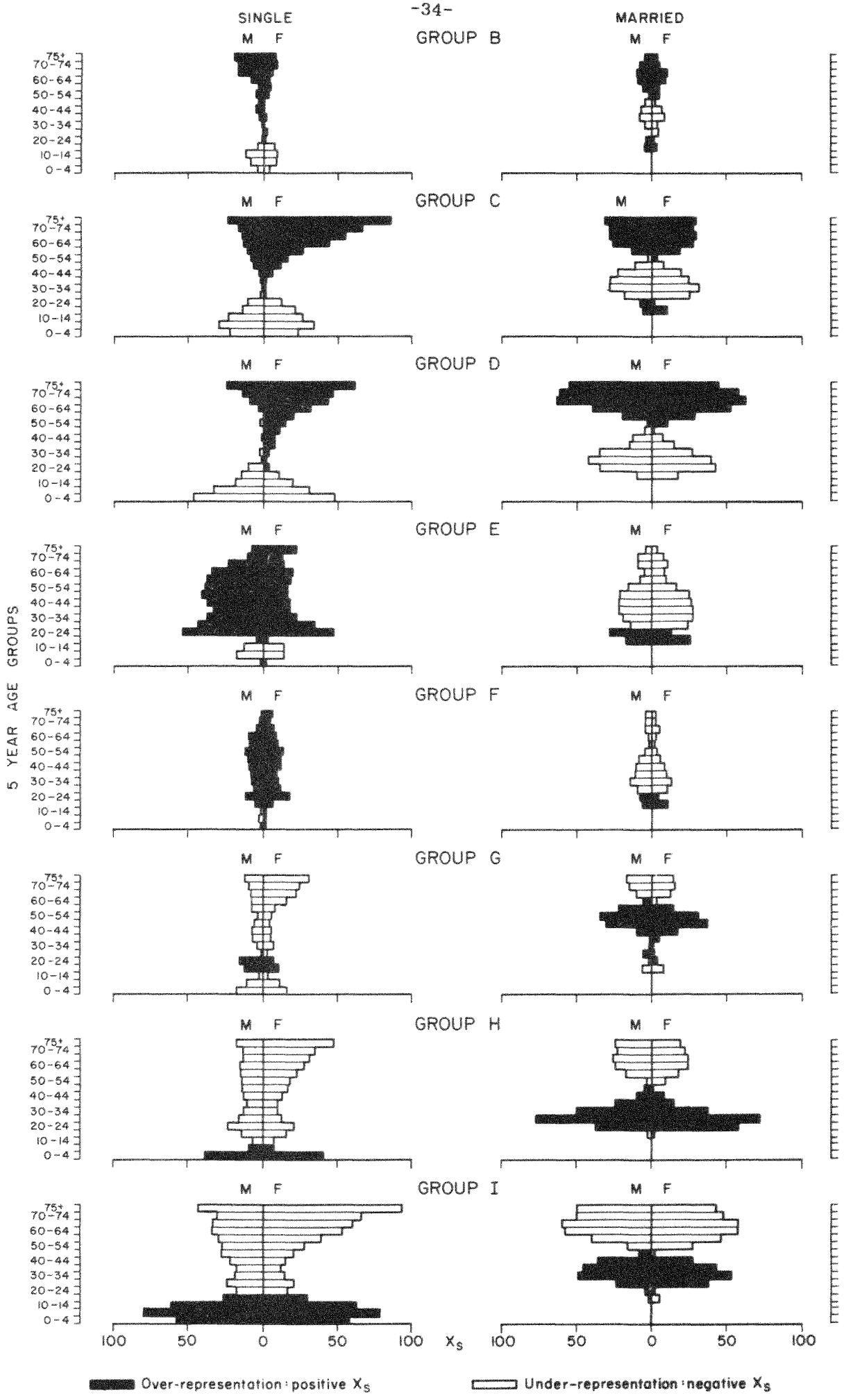


Fig. 9. Plot of X_s values showing the over- and under-representation of population components within each group

sex pyramids for Group I (Figs. 7 and 8) are highly distinctive.

The geographical distribution of Group I involves a variety of locations. Outstanding blocks occur in areas of council housing in Liverpool, the Wirral peninsula and Ellesmere Port. Other conspicuous blocks occur, for example, in Kirkby U.D., Skelmersdale and Cheadle. The concentrations of Group I in the Valley R.D. of Anglesey, in Shrewsbury and in Shifnal R.D. are probably associated with defence establishments.

4.3. General comments

Although the groups were determined on the basis of 'distinctive' categories, they are interesting for the mix of household and population components which each contains. Owing to the continuum in measurement space, the groups are by no means homogeneous; nevertheless, their aggregate characteristics show distinctive patterns. The bell-shaped age-sex pyramid of Group I is very distinctive for its broad base and marked preponderance of children (Fig. 7). In contrast, the Group D pyramid is decidely top heavy, while those of Groups E and F are noteworthy for the pronounced mode in the 20-24 age group. The suburban distribution of Group I, the inner-city concentration of Group E and the concentration of Group D in retirement and high-status areas complement each other on a broad scale.

Groups C and B are transitional in their aspatial characteristics between Groups E and D (Fig. 7). While some Group D areas seem to reflect the results of migration on retirement, Groups C and B would seem to indicate aging in situ in areas which have experienced an out-migration of younger age groups in the past. The X_s plots (Fig. 9) for Groups B and C suggest a more recent influx of very young couples. In this context, it is significant that Jones (1978), in his study of the distribution of immigrants, draws attention to the recent movement of New Commonwealth immigrants into the older industrial zones where Group C is especially strong. While Group C occurs mainly in urban locations, the distribution of Group B suggests rural areas which have experienced out-migration, especially of females.

While Group I does show some rural locations, associated, for example, with defence establishments, Groups G and H are markedly

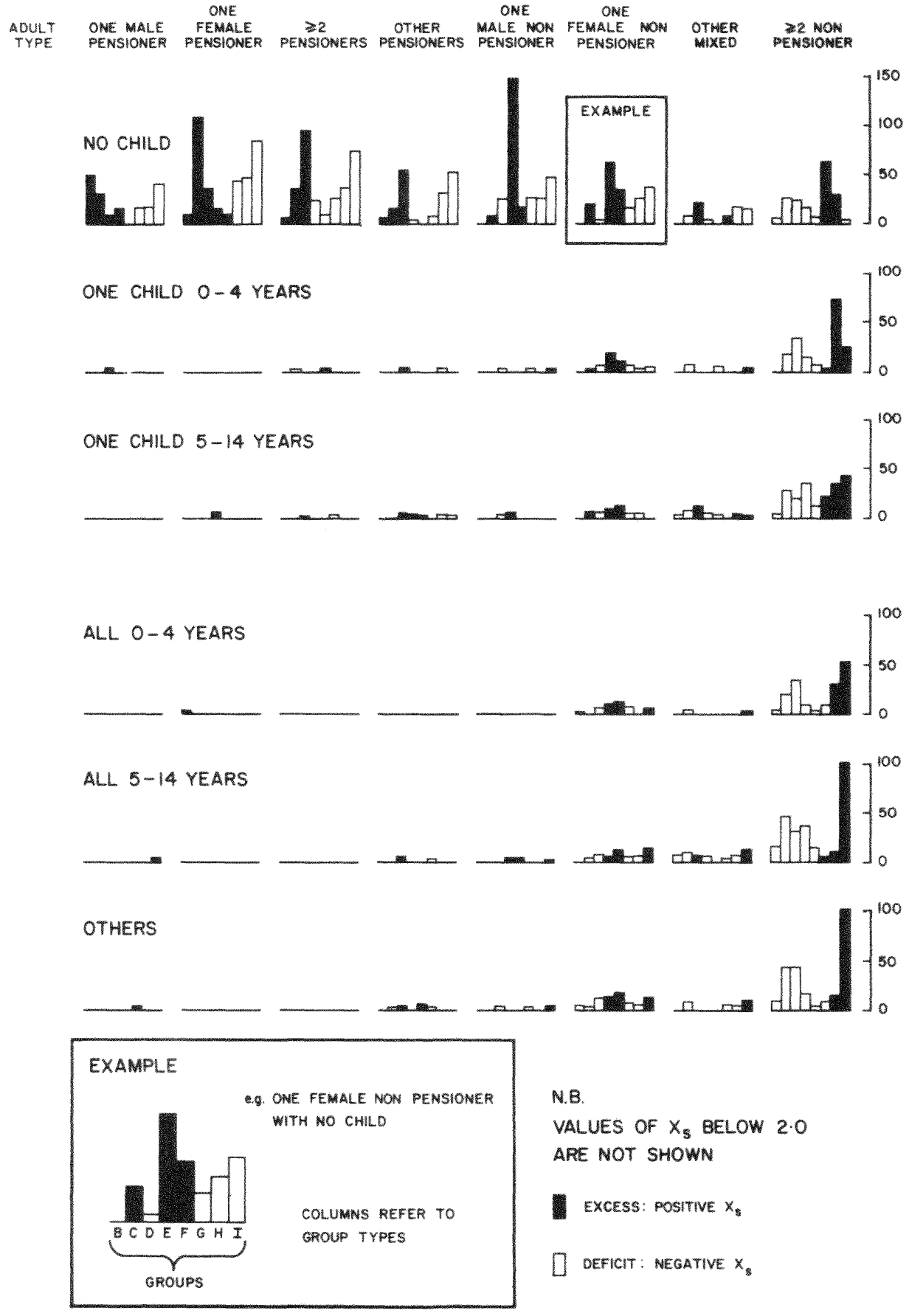


Fig. 10. Plot of X_s values showing the distribution of the 48 primary household types among the 8 deviant groups

suburban in their distribution. The age-sex pyramid for Group H (Fig. 7) shows a marked second mode on a basically bell-shaped pattern. Group G has a different composition altogether, the second mode occurring much higher up the pyramid.

The groups discussed so far each have a distinctive mix of household and population components and/or geographical distribution, but Group F appears to be a diminished version of Group E.

5. EVALUATION OF THE CLASSIFICATION

All classification procedures are sensitive to the selection, definition and number of the variables which determine the distribution of data units in measurement space. The refinement of a classification should therefore be as much concerned with the homogeneity of the descriptors as with the recognition and deletion of heterogeneous and/or non-unique groups.

5.1. Homogeneity of initial descriptors

The eight-fold categorisation was derived by aggregating subsets of the 48 primary household types (Section 2.2. and Fig. 1). In many studies, the process of derivation has involved much soul-searching in an attempt to extract, from the available data, the most appropriate indicators of themes of substantive interest. In this process, variables for statistical mapping and analysis have often been derived by aggregation. However, in most cases, little or no attempt has been made at a subsequent evaluation of the homogeneity of these summations.

Owing in part to irregularities and ambiguities in the O.P.C.S. definitions (Section 2.2.), it was recognised from the outset that the eight-fold categorisation was a tentative one and much effort was directed towards developing procedures (even if only qualitative ones) for checking it. Consequently, the X_s values in Table 9 were plotted graphically in Figure 10 in order to compare the patterns of within-descriptor and between-descriptor variations. If the allocation of the 48 primary counts to the secondary variables (descriptors) were reasonably sound, then within-descriptor primary counts should exhibit similar variations with respect to the eight groups, while the primary counts in different descriptors should exhibit dissimilar distributions between groups.

TABLE 10 : GROUP IMPORTANCE

Group	Number and percentage of units	$\sum X^2$	$\% \sum X^2$	Mean X^2	Rank
B	484 (12.43)	12,566	4.74	25.97	7
C	536 (13.76)	39,099	14.74	72.96	3
D	699 (17.95)	37,355	14.08	55.18	4
E	443 (11.37)	57,392	21.64	129.56	2
F	178 (4.57)	7,776	2.93	43.70	8
G	394 (10.12)	16,044	6.05	40.73	6
H	377 (9.68)	21,413	8.07	56.81	5
I	784 (20.13)	73,616	27.75	93.91	1
Total	3,895 (100.00)	265,261	100.00	68.10	-

A very small proportion of pensioner households include children and thus their variations cannot be over-emphasized. The difference between one male and one female pensioner households with no child lies more in the amplitude of variation than in the pattern of excess and deficit of components in the various groups. This suggests that the two categories of one pensioner households could be combined. The distributions of several pensioner and other pensioner households are fairly similar, justifying the pooling of these two counts. The aggregation of counts for one male non-pensioner households with and without children seems reasonable owing to the low proportions of such households with children. The same cannot be said for one female non-pensioner households. While childless and one-child households in this category are particularly important in Groups E,F and C, one female households with two or more children are found in excess in Groups E,F and I. On the other hand, the aggregation may be justified on the grounds that E and F remain the modal groups.

The pooling of household counts for"several non-pensionable adults" with "other mixed" households is inappropriate. While the former category is found in excess in Groups G and H, G being the modal group, the "other mixed" households have a different distribution pattern with the modal class in Group D. This partial association can also be observed in households with older children.

Figure 7 also suggests the aggregation of households with children into two classes - those with one child and those with two or more children - may not be the most useful categorisation. While the number of children can reflect fertility and household size, the O.P.C.S. classification into these two groups is not particularly useful since the "two or more" category includes a wide range of numbers of children. An alternative approach would be to aggregate counts on the basis of the age of the children, a characteristic reflecting the life-cycle stage of their parents. Indeed, the present classification, as it stands, emphasises the importance of this factor (Fig. 9). However, such a categorisation would present difficulties because of the "other children" category in the O.P.C.S tables, particularly in view of the fact that the distribution of this category (Fig. 10) is more akin to that of pre-school than that

of older children.

The definition of categories of household composition is therefore open to further discussion and refinement since, if there is a difference in the distributional pattern of component counts at the group average level, the disparity is likely to be greater at the level of the individual one-kilometre grid square. Revision of the categorisation is bound to have a significant effect on the resulting classification.

5.2. Evaluation of the groups

Classification procedures are directed towards producing classes in which members of the same class are as similar and/or members of different classes are as dissimilar as possible. Since there are several different measures of similarity between data units, there can be no single set of universally applicable rules for estimating the success of the classification, especially since data conditions and problem definitions may direct attention towards different features of the property space.

The concept of a single constellation of all data units in hyperspace forms the basis of the current classification procedure. The density of objects is greatest at the core of average values and declines progressively outwards with both ratio and X_s plots. The more extreme, and often the most interesting, departures exist in the outer, more diffusely occupied parts of the measurement space. In general, the morphology and spatial distribution of the groups seem to substantiate the proposition that the marked excess of a particular household type can be used as a measure of the tendency towards a particular mix of types. However, when the distinguishing category is highly and positively associated with some other category, the proposed method would segment a'natural'group into two separate parts. This may explain why Groups B and F were seen to be modified versions of Groups C and E respectively. Group characteristics were valuable for assessing the differences between groups, but there is also a need to evaluate the importance and homogeneity of groups.

5.2.1. Group importance

While the absence or paucity of data units in a particular sector of measurement space makes that sector uninteresting, the importance of a group cannot be evaluated as a function of the number and/or density of the data units within its catchment alone. Two groups, each containing the same number of data units, cannot be deemed equally important without reference to the location of those data units within the corresponding sectors. The frequency distribution of data units within the 'deviant' region (and sectors) is highly skewed, with the mode near the inner limits verging on the average area . Thus data units in the outer regions of the group sector are much more important than those just outside the central core and therefore just within the sector.

Consequently, group importance is here formulated in the following terms. The deviation of each data unit from average characteristics is described by the X^2 value (equation 4, section 3.1). Within the 'deviant' region, each data unit has a X^2 value in excess of 14.1. The sum of the X^2 values of all data units within the 'deviant' region forms the total deviation to be examined. This total deviation is distributed in different proportions among the eight sectors. When this is done, sectors may be ranked as in Table 10 where, taking into account both the degree of variation and the number of units involved, Groups I,E, C and D, in that order, emerge as the most interesting. Groups B and F, whose attribute structures were indistinct at the group level, are seen also to contribute least towards the total deviation. Group F, in particular, contains relatively few data units, while the mean X^2 value of Group B, which is the lowest, indicates relatively small departures.

5.2.2. Group homogeneity

Within-sector distribution of data units and deviation was studied by partitioning each group's sector into sub-areas using the second-ranked X_s value in each data unit. Data units with an excess of only one variable formed a class of their own within the group. The remaining data units were allocated to one of seven sub-classes, each corresponding to the remaining variable recording the next-highest X_s

TABLE 11 : DISTRIBUTION OF ONE-KILOMETRE SQUARES AND TOTAL VARIATION AMONG 64 CLASSES

Group B	attribute structure of variables	Squares		Deviation		Means		
		No	%	% of total	% of group	X^2	max	2nd
	1 2 3 4 5 6 7 8							
C11								
C12	+ + + - - - - -	85	17.6	0.93	19.6	29.0	14.5	4.7
C13	+ + + - - - - -	110	22.7	1.07	22.7	25.9	12.9	3.7
C14	+ - - + + - - -	57	11.8	0.70	14.8	32.7	13.2	6.7
C15	+ - - - + - - -	36	7.4	0.31	6.6	23.1	13.3	3.7
C16	+ - - - - + - -	85	17.6	0.72	15.2	22.4	15.2	1.7
C17	+ - - - - - + -	59	12.2	0.52	11.1	23.5	15.8	2.4
C18	+ - - - - - - +	52	10.7	0.48	10.1	24.4	16.5	2.5
	Total	484	100.0	4.74	100.0			
Group C								
C21	+ + + + + - - -	153	28.5	3.45*	23.37	59.7	28.6	8.0
C22	- + - - - - - -	1	0.2	0.01*	0.0	28.8	22.3	-
C23	+ + + - + - - -	218	40.7	7.01*	47.55	85.3	34.8	14.6
C24	+ + + + + - - -	54	10.1	2.08	14.08	102.0	43.6	14.6
C25	+ + + + + - - -	66	12.3	1.68	11.38	67.4	29.7	8.8
C26	- + + - - + - -	19	3.5	0.27	1.86	38.3	13.4	3.0
C27	- + - - - - + -	13	2.4	0.13	0.87	26.2	10.7	3.3
C28	- + - - - - - +	12	2.2	0.12	0.81	26.5	12.0	2.8
	Total	536	100.0	14.74	100.0			
Group D								
C31	+ + + - - - - -	118	16.9	1.46*	10.40	32.9	15.4	4.4
C32	+ + + - - - - -	289	41.3	10.06*	71.48	92.4	43.9	13.7
C33	- - + - - - - -	19	2.7	0.19	1.37	27.0	19.7	-
C34	+ - + + - - - -	48	6.9	0.46	3.23	25.2	11.2	3.5
C35	- + + - + - - -	41	5.9	0.46	3.23	29.3	13.0	3.8
C36	- - + - - + - -	111	15.9	1.18	8.38	28.2	11.7	2.1
C37	- - + - - - + -	36	5.1	0.29	2.03	21.1	10.3	1.4
C38	- - + - - - - +	37	5.3	0.44	3.11	23.5	13.0	2.5
	Total	699	100.0	14.08	100.0			
Group E								
C41	+ + + + + - - -	63	14.2	1.02*	4.73	43.1	22.8	6.7
C42	+ + + + + - - -	62	14.0	2.40*	11.10	102.8	45.0	17.8
C43	+ + + + + - - -	61	13.8	0.82	3.78	35.6	16.0	4.4
C44	- - - + - - - -	2	0.5	0.03*	0.16	46.2	43.9	-
C45	+ + - + + - - -	125	28.2	15.89*	73.45	337.2	212.8	60.0
C46	- - - + - + - -	57	12.7	0.72	3.33	33.6	20.6	3.2
C47	- - - + - - + -	41	9.3	0.41	1.90	26.5	15.5	2.9
C48	- - - + + - - +	32	7.2	0.33	1.53	27.5	14.5	3.7
	Total	443	100.0	21.64	100.0			

TABLE 11 (Cont.)

	attribute structure of variables	Squares No.	Squares %	Deviation % of total	Deviation % of group	Means X^2	Means max	Means 2nd
Group F	1 2 3 4 5 6 7 8							
C51	+ - - - + - - -	17	9.6	0.15	5.14	23.6	9.3	4.7
C52	- + - + + - - -	34	19.1	0.63	21.50	49.2	19.7	8.6
C53	- - + + + - - -	28	15.7	0.27	9.27	25.8	14.5	2.7
C54	+ + - + + - - -	46	25.8	1.34	45.87	77.5	31.3	19.1
C55	- - - - + - - -	0	0.0	0.0	0.0	-	-	-
C56	- - - - + + - -	18	10.1	0.18	6.17	26.7	12.5	2.6
C57	- - - - + + + -	12	6.7	0.11	3.86	25.0	8.6	3.4
C58	- - - - + - - +	23	12.9	0.24	8.18	27.7	12.6	3.4
	Total	178	100.0	2.93	100.0			
Group G								
C61	+ - - - - + - -	26	6.6	0.23	3.78	23.3	8.6	2.5
C62	- + - - - + - -	10	2.5	0.08	1.32	21.1	8.1	0.7
C63	- - + - - + - -	60	15.2	0.60	9.84	26.3	8.9	2.3
C64	- - - + - + - -	17	4.3	0.17	2.78	26.3	9.5	3.6
C65	- - - - + + - -	19	4.8	0.24	3.96	33.5	12.7	2.5
C66	- - - - - + - -	19	4.8	0.20	3.38	28.6	18.0	-
C67	- - - - - + + -	173	43.9	3.26*	53.97	50.1	15.3	5.1
C68	- - - - - + + +	70	17.8	1.27	20.97	48.1	11.5	4.1
	Total	394	100.0	6.05	100.0			
Group H								
C71	+ - - - - - + -	24	6.4	0.22	2.78	24.8	10.9	3.1
C72	- + - - - - + -	5	1.3	0.04	0.55	23.5	9.0	2.9
C73	+ - + - - - + -	17	4.5	0.14	1.70	21.5	7.8	2.4
C74	- - - + - - + +	20	5.3	0.18	2.22	23.8	9.6	3.0
C75	- - - - + - + +	8	2.1	0.07	0.84	22.5	9.5	1.7
C76	- - - - - + + +	139	36.9	2.76*	34.15	52.6	15.9	5.5
C77	- - - - - - + -	4	1.1	0.03	0.32	17.1	14.1	-
C78	- - - - - + + +	160	42.4	4.64*	57.45	76.9	22.8	11.5
	Total	377	100.0	8.07	100.0			
Group I								
C81	+ - - - - - - +	27	3.4	0.32	1.15	31.3	15.7	2.4
C82	- + - - - - - +	18	2.3	0.29	1.05	42.8	20.6	3.7
C83	- - + - - - - +	37	4.7	0.71	2.57	51.2	21.3	4.8
C84	- - - + - - - +	31	4.0	0.28	1.00	23.9	10.1	2.5
C85	- - - - + - + +	50	6.4	1.84*	6.64	97.8	54.7	5.6
C86	- - - - - + + +	142	18.1	2.65*	9.56	49.6	14.9	3.7
C87	- - - - - - + +	465	59.3	21.25*	76.56	121.2	54.1	11.0
C88	- - - - - - - +	14	1.8	0.41	1.46	77.0	55.9	-
	Total	784	100.0	27.75	100.0			

* ten highest proportions of total deviation

value. This produced eight classes within a sector, giving a total of 64 possible classes within the 'deviant' region. Each of these is denoted by C_{IJ}, where I refers to the variable with the highest X_s value and J to that with the next highest value.

Table 11 is a summary of the resulting distribution of data units and deviations within the 64 classes. Ten classes together contain over 73 per cent of the total deviation, indicating an irregular distribution of data units in measurement space. There are very few data units (59) which show an excess of only one household type. In general, the marked excess of one type of household reflects a tendency towards a particular composition of household types. Sometimes the dominant component is associated primarily with an excess of only one other type of household; for example, variable 8 is associated very strongly with variable 7 in Group I. In other classes and groups, there are positive associations between several types. This is the case in Group E, where one non-pensioner and pensioner households are found in excess.

Some of the groups - C, E, H and I - show a greater degree of homogeneity than others in which data units are dispersed in several directions. The modal class in Group I - C_{87}, where the two most prominent household types are the two several-adult types with children - contains 76.5 per cent of group deviation and nearly 60 per cent of the group's data units, which are located in the outermost region of the group sector; the mean X^2 value is high at 121.2. Both variables 7 and 8 have high mean X_s^2 values. Variable 7 is also found in excess in 79 per cent of the data units within the group (Table 12). C_{85}, C_{86} and C_{87} show an excess of variable 7 on average. These classes, together with C_{88}, account for more than 94 per cent of within-group deviation and 85.6 per cent of the data units, showing a tendency for the latter to be distributed in similar directions within the group sector. The anomalies in C_{81}, C_{82}, C_{83} and C_{84} are relatively unimportant, since they contribute little towards total deviation and record relatively low mean X_s^2 values for the second variables. C_{83}, where the second variable is several-pensioner households, with a mean X_s^2 in excess of 3.84 for variable 3 is an exception, the reasons for which are considered later.

Similarly, in Group E, the pattern of excess of one pensioner and one non-pensioner household types can be observed in classes C_{41}, C_{42}, C_{43} and C_{45} which, together with C_{44}, account for 93 per cent of group deviation and 70.7 per cent of the data units. Thus the aggregate characteristics of the group are found in several sub-classes which include the majority of the data units.

In Group C, classes C_{21}, C_{23}, C_{24} and C_{25} exhibit attribute structures which correspond to the aggregate characteristics of the group. Variables 1 and 3 in particular are found in excess in 74 and 80 per cent respectively of the data units (Table 12). The four classes together contain 96 per cent of the deviation and 92 per cent of the data units.

Within Group H, C_{78} and C_{76} contain the bulk of the data units, and deviations display similar patterns of excess and deficit - several other adult households with and without children are both in excess.

In Groups D and G, on the other hand, it is possible to observe two separate trends. In group D, the major trend involves an excess in all pensioner household types, namely in C_{31} and C_{32}, but these classes include only 58.2 per cent of the data units. Another 16 per cent of the data units and 8 per cent of the deviation are found in C_{36}, where the second variable is the household with several adults and no children, which has a markedly different attribute structure. Similarly, Group G, is seen to include two different trends. The bulk of the deviation is attributed to C_{67} and C_{68} - households with several adults accompanied by a child or children - which account for 62 per cent of the data units. A further 15.2 per cent of the data units are found in C63, where the second variable is households with several pensioners, which account for 10 per cent of the group deviation. The minor trends include a relatively smaller proportion of data units and the second variables show low mean X_s^2 values, as a result of which the second trend does not manifest itself in the aggregate characteristics of the group. The minor relationship between variables 3 and 6 is believed to be the product of poor definition of variables, and can be attributed to the pooling of counts for " $\geqslant$ 2 other" households, the O.P.C.S. residual category, with those for two or more non-pensioner adult households (see Sections 2.2 and 6.1.). The same factor could be responsible for the very small minor trends in C_{83}, C_{46} and C_{26}.

TABLE 12 : PERCENTAGE OF DATA UNITS WITH AN EXCESS OF EACH HOUSEHOLD TYPE WITHIN THE GROUPS

Groups	Household Type							
	1	2	3	4	5	6	7	8
B	100.0	39.9	55.2	31.4	24.6	45.5	35.1	29.1
C	74.4	100.0	79.9	47.0	56.5	18.3	13.6	7.6
D	52.1	58.4	100.0	22.3	28.9	32.6	14.9	9.7
E	44.0	39.3	41.8	100.0	53.3	40.9	25.3	28.7
F	37.6	51.1	47.2	53.4	100.0	42.1	30.9	31.5
G	21.6	9.1	25.1	12.4	14.0	100.0	65.5	37.6
H	19.4	6.4	8.8	14.1	14.6	72.7	100.0	73.7
I	11.6	4.8	10.2	12.5	23.2	47.2	79.0	100.0

The pooling of dissimilar counts in variable 5 could also be responsible for the presence of several trends in Group F. The data units in Group B are diffused in several directions in the inner parts of the group sector. The mean X^2 value for all classes is low and the class attribute structures bear little resemblance to each other.

6. DEMOGRAPHIC TYPES IN THE STUDY AREA

Although it is not the intention of this paper to discuss in detail the nature and distribution of the demographic types identified, Webber's Liverpool Social Area Study (1975) proved a valuable source of information for cross-checking the distribution of household types in and around Liverpool C.B., for assessing some of the processes responsible for their existence and for identifying other features which might help to explain variations between and within the groups identified in this paper.

The five major types or families of social areas identified in Webber's study corresponded to five major types of housing area :

1. A high-status, owner-occupied area with stable families
2. An area of subdivided housing with many young people and furnished, privately-rented accommodation in small units
3. Inner/older council estates
4. Outer/more recent council estates
5. An area of older, Victorian terraced housing, for the most part privately rented unfurnished.

Webber's statistical analysis and detailed discussion of the sub-types within each of these families of social areas indicate that the latter do not correspond - and were not intended to correspond - to demographic types. Thus it is not surprising that there is only a limited degree of correspondence between Webber's families and the demographic types identified in this working paper. Moreover, it should be remembered that the demographic types in the present study were identified with reference to average tendencies in the study area as a whole rather than to Liverpool alone.

Group E of the present study, in which one-adult, especially non-pensioner, households are over-represented, shows marked spatial contiguity in the inner-city areas of Liverpool, Manchester, Bolton, Preston and Stoke (Fig. 5). These areas have experienced a recolonisation of older housing stock with elderly residents (many of them women living alone) by various transient and/or disadvantaged groups including students, young married couples with pre-school children, and one-parent families. The distribution of Group E within 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 a result, Groups B and C are almost absent from Liverpool C.B. and from Merseyside as a whole, with the exception of Wallasey, which contains the once popular but now decaying holiday resort of New Brighton (Fig 4.). Other contiguous blocks of Group C in Preston, Blackburn, Burnley, Oldham and Stockport and on the eastern fringes of Manchester C.B. show a gross over-representation of elderly, especially one-pensioner households. These areas also show an excess of young marrried couples with pre-school children and of one-parent families. There is an under-representation of more mature families with older children.

The linear distribution of Group C, presumably along the main roads, eastwards from Manchester to such places as Ashton-under-Lyne and Oldham coincides with areas of privately-rented, unfurnished property. The high density of households (Table 5) suggests older terraced housing which had experienced an out-migration of young persons in the past but which, by 1971, had attracted numbers of young, often disadvantaged people. The inflow of younger people had not, however, been as great as that towards areas in Group E.

Group B is also found in areas of older housing, both urban and rural. Whereas in urban areas the occurrence of Group B is masked by the preponderance of Group C. (Fig. 4), it is particularly distinctive in rural districts owing to the relative paucity of Groups C and D. The low average household density of Group B also suggests

small populations. This contrast in location may point to a 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. As has already been pointed out (Sections 4.2.2. and 4.2.3), Groups B and C are otherwise very similar as regards their population components. Both groups also tend to be found in areas with an excess of property rented privately and unfurnished. It is quite likely that the older residents of such areas have not moved on retirement.

Some anomalous blocks of Group C can be found in the holiday resorts of North Wales and Lancashire, for example in Criccieth, Llandudno, Colwyn Bay, Grange, Lytham St. Annes and Southport (Fig. 4). In these areas, Group C occurs largely in districts with a marked excess of privately-rented furnished accommodation, probably reflecting the presence of numerous hotels. Here, the marked excess of female pensioners could be affected by the type of visitor present on census night but is due mainly to the age-sex composition of the resident population.

In contrast, 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. These areas also have an excess of owner-occupied property. 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 the areas of relatively high status housing identified by Webber. Such areas are of two types. The elderly age structure in owner-occupied areas such as Woolton in Liverpool C.B. or Didsbury in Manchester C.B. is in part related to the period in which housing development occurred. Much of the population has aged in situ and high house prices have been an obstacle to purchase by younger people, even those from high income groups, who have not yet benefitted from many years appreciation of their own house values. Blocks of Group D also occur in outer areas with older council estates, such as Norris Green. Webber describes the latter type of area as comprising low-density, semi-detached council houses built in the 1930's and the early post-war period. This type represents the most favoured council

estates with the highest status and the lowest level of deprivation, which has an elderly age structure and an unusually low proportion of large families.

Groups with an excess of households composed of several non-pensioners, with or without children - Groups G, H and I - show suburban locations in areas with an excess of either owner-occupied or council housing. 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 high-status family, with higher scores for middle-aged than for retired people. This is usually associated with semi-detached housing, which gives the group its lower average household density.

Group H has an excess of households with two or more non-pensioners, with and without children, but also has a marked over-representation of such households with one child. The group has a youthful age structure with an excess of pre-school children. It is interesting that areas of this type are absent from Liverpool and Manchester C.B.'s. The age structure and household composition of Group H areas suggest that they are zones of more recent housing development. Small blocks of Group H in commuter settlements in the rural districts of Cheshire suggest the presence of professional and/or higher income classes. However, the distribution of Group H coincides with areas of excess both of owner-occupied and of council dwellings in the zone extending from Orrell to Haydock and Newton-le-Willows, in Lumm, Cannock and Walsall and in the Stoke-on-Trent area.

Group I, which is particularly striking for the over-representation of households with two or more children, especially children of school age, shows a distinct pattern of spatial distribution in the Wirrall peninsula, south Manchester, Cheadle U.D., Gatley and Formby, and occurs as arcs or rings around several Municipal and County Boroughs, including Stafford, Oswestry, Chester and Liverpool. Within Liverpool C.B., the 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 benefitted from council policies. The spatial distribution of Groups H and I also confirms Webber's statement 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 present analysis is only tentative, owing to the use of heterogeneous and associated categories, the results indicate that polychotomous data for household composition can be summarised by nominal codes, each corresponding to a different mix of household types. There is some degree of variation within each group, since household types extend over more than one social environment. A finer classification may be produced by considering other related factors, such as household tenure and the mobility, ethnicity or socio-economic composition of the population. The occupational structure, for example, would discriminate between the rural and older industrial areas of out-migration (Groups B and C respectively) and would also sift out those anomalies in Group E which are related to the presence of defence establishments.

7. CONCLUSION

A very simple method has been used in this paper for the classification of areal data on the basis of household composition. It should now be tested with other polychotomous data such as those on household tenure, ethnic composition and socio-economic groups. The method exploits the closure effect in categorical data and therefore works best with relatively few categories. It is easy to comprehend and implement and requires minimal computer resources.

Since the classification procedure is sensitive to the number and characteristics of categories, methods for qualitatively ascertaining the homogeneity of the latter with respect to component counts were considered. The X_s distribution of the 48 primary counts within the eight distinctive groups (B to I) gave some indication not only of the homogeneity of derived categories but also of their interdependence (Fig. 10) and of the complex and non-linear relationships between

household types.

Data on household composition, either directly or indirectly, encode information on several demographic attributes, namely the age, sex and marital status of the adults, the size of the household and the number and age of the children. Ages of adults and children, together with the numbers of children, seem to be the most useful indicators of life-cycle stage. Together, they suggest a three-fold classification of areas into those with concentrations of 'aged', 'transient young adults' and 'settled younger families' respectively (Figures 4, 5 and 6). Marital status is another important characteristic which reinforces this classification and discriminates between areas with excesses of one-pensioner and two or more pensioner households.

Sex composition is the least discriminating feature : even if it is important at the household level, it is unlikely to manifest itself to the same extent in aggregated areal data. It appears to be important only to the extent that it reflects age composition and mobility. Very low masculinity proportions are associated with aged populations, while masculinity proportions in excess of 53 per cent were observed among 20-29 year old adults in Group E. Consequently, a reduction in the number of groups could involve the merging of male and female pensioner households (Groups B and C) and of male and female non-pensioner households (Groups E and F). In future analysis of data for the whole of Great Britain, this will be considered at the initial stage of categorisation of household types rather than at the final stage of classification.

In general, the marked over-representation of one dominant category (as measured by X_s scores) reflects a tendency towards a particular mix of household and population components. Future work on the identification of demographic types in Britain will involve the use of only six major household types. Attempts will be made to relocate some of the data units in the 'average' category and to identify and map subtypes within groups. This will involve a classification of areal units on the basis of other polychotomous data such as socio-economic groups and housing areas. Nominal codes for each of the

characteristics would then be cross-tabulated and further analysed to explore relationships between factors and to refine typologies.

ACKNOWLEDGEMENTS

The author is deeply indebted to the many people and institutions who have made this study possible. The research was funded by S.S.R.C. and the data, in computer-readable form, 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 facilities with the Department of Geography, University of Durham. These include the modern computing and graphics laboratory built up through the initiative of Dr. D.W. Rhind, photographic work by Mr. D. Hudspeth, cartographic assistance provided by Messrs A. Corner and D. Cowton and reprographic services undertaken by Messrs J. Normile and D. Ewbank. The manuscript was typed by Mrs. J. Dresser.

The author is also grateful to the other members of the Census Research Unit for their encouragement and advice, especially to Professor W.B. Fisher for the one-year post of Temporary Lecturer which made the completion of the study possible, to Professor J.I. Clarke for discussion and suggestion during the initial stages of the study, to Mr. J.C. Dewdney for collaboration, advice and much editorial work, and to Dr. I.S. Evans and Dr. D.W. Rhind for helpful comments on the paper. She also appreciates the valuable references on classification methods provided by Mr. N.J. Cox and long discussions on the study area with Mr. D. Owen.

REFERENCES

CATTELL, R.B. and COULTER, M.A. (1966) "Principles of behavioural taxonomy and the mathematical basis of the taxonomic computer program", Br. J.of Math. Statist. Psychol., 19, 237-269

CHAYES, F. (1971) Ratio Correlation, Chicago, University of Chicago Press

CLARKE, J.I. (1977) "On the classification of population types and regions", National Geographer., XII (2), 131-136.

CENSUS RESEARCH UNIT, People in Britain - a census atlas, H.M.S.O., forthcoming

EVANS, I.S. (1977) "The selection of class intervals", Transactions of the Institute of British Geographers, N.S. 2 (1), 98-124

EVERITT, B. (1974) Cluster Analysis, Heinemann Educational Books, London

JARDINE, N. and SIBSON, J. (1971) Mathematical Taxonomy, Wiley, London

JONES, P.N. (1978) "The distribution and diffusion of the coloured population in England and Wales 1961-71". Trans. Inst. Br. Geog. New Series Vol. 3 (4), 515-532

OFFICE OF POPULATION CENSUSES AND SURVEYS (1972) Census 1971, Standard Small Area Statistics (Ward Library) Explanatory Notes, H.M.S.O.

RHIND, D.W. (1975) "Mapping of the 1971 census by computer", Population Trends, 2, 9-12

SNEATH, P.H.A. and SOKAL, R.R. (1973) Numerical Taxonomy, Freeman, San Francisco.

SOKAL, R.R. (1974) "Classification : purposes, principles, progress, prospects", Science, 185 (4157), 1115-1123

VISVALINGAM, M. (1979) "The signed chi-square measure for mapping", Cartographic J., Vol.16 (1), 93-98

VISVALINGAM, M. and PERRY, B.J. (1976) Storage of the grid square based 1971 GB census data : checking procedures. Census Research Unit Working Paper No.7, Department of Geography, University of Durham

VISVALINGAM, M and DEWDNEY, J.C. (1977) The effects of the size of areal units on ratio and chi-square mapping. Census Research Unit Working Paper No.10, Department of Geography, University of Durham

WEBBER, R.J. (1975) Liverpool Social Area Study 1971, Data : Final Report, PRAG Tech. Papers TP14

WISHART, D. (1969) "Mode analysis" in Numerical Taxonomy (A.J. Cole, ed.) Academic Press, New York, 282-308.