THE 1981 CENSUS SMALL AREA STATISTICS FOR
WEST YORKSHIRE WARDS: HOW TO ACCESS,
EXTRACT AND ANALYSE.

Mark Birkin and Philip Rees

School of Geography
University of Leeds
LEEDS LS2 9JT
UK.

July 1982

CONTENTS

ACKNOWLEDGEMENTS

ABSTRACT

The Small Area Statistics (SAS) from the 1981 Census are now available for analysis.   This paper describes a system for accessing, extracting and analysing a subset of the SAS data: the 100% tables for West Yorkshire wards.   The data files, extraction and analysis programs described should help researchers prepare for the availability towards the end of 1982 of enumeration district statistics at University regional computing centres as processed by the major package SASPAC.

Since the data are of interest in themselves the paper concludes with a detailed analysis of the ethnic geography of Leeds metropolitan district in April 1981.

LIST OF TABLES

LIST OF FIGURES

1.   INTRODUCTION

    1.1   The 1981 Census and its publication

        On the 5th April, 1981 a census of the population of England and
    Wales was carried out by the Office of Population Censuses and Surveys.
    Such censuses are administered every ten years in the United Kingdom.
    The 1981 Census provides a comprehensive picture of the state of the
    British population in terms of the characteristics of individuals, of
    the households they are members of and of the housing spaces they
    occupy.

        The results of the 1981 Census are currently in the course of
    publication.   Preliminary reports (OPCS, 1981a, 1981b) on the counts
    of people in administrative areas have been published, and selected
    tabulations (OPCS, 1982a) for those areas have been produced in the
    form of OPCS County Monitors.   A few County Reports (OPCS, 1982b)
    are now available and many more are in the pipeline.   Within the
    counties Small Area Statistics (SAS) are being published for enumera-
    tion districts, electoral wards and local government districts in a
    standard set of tables.   Publication consists of production on
    printed paper, on microfilm or on magnetic tape.   All this is
    happening at amazing and unprecedented speed within 18 months of
    Census day.

    1.2   Plans for use of the Census 1981 SAS by the higher
            education community

        Far reaching and ambitious steps have been taken to make the
    Census 1981 SAS available to the academic community.   The Social
    Science Research Council (SSRC) has purchased a national set which
    is to be accessed either via the SSRC Data Archive at the University
    of Essex or via five regional computing centres at the Universities
    of London, Bristol, Manchester, Newcastle and Edinburgh.   The SSRC
    has also bought software to make possible the easy use of this vast
    data set in the form of a package called SASPAC prepared at the
    University of Durham and at the Program Library Unit, University of
    Edinburgh under the direction of Professor David Rhind.   Towards the
    end of 1982 the SAS data should be mounted on special disk packs
    at the regional computing centres and accessible via SASPAC from all
    Universities via the regional computing networks.

1.3 Local use at the University of Leeds

In order to obtain experience with the SAS prior to and in preparation for the coming online of the 'national' system, a set of ward level SAS was purchased for the School of Geography through the good offices of Professor Alan Wilson and Martin Clarke using funds available from an SSRC Grant. These ward data have been stored on diskspace assigned to the Geography program and data library (GEO6LIB) for general use by members of the School (staff, postgraduates, undergraduates).

1.4 Aims of this paper

This paper is designed to serve as a guide to the SAS data for West Yorkshire and as an introduction to the national SAS.

The general structure of the Census 1981 SAS data is described in section two of the paper and the key documents that need to be consulted are referenced.

The particular structure of the 1981 Census SAS for West Yorkshire wards as held on the Leeds University computer is outlined in the third section of the paper.

The fourth section of the paper presents and gives instructions for a general data extraction and tabulation program (TABULATE) written by the authors. It can be used to produce tabulations for all wards of a district on a particular topic.

The fifth section details a more specific data extraction and further analysis 'system' as an example of the kinds of program users may wish to write themselves.

Finally, some preliminary results are presented for the spatial patterning of ethnic groups in the city of Leeds to provide a flavour of what can now be done.

2. THE CENSUS 1981 SAS DATA

Here reference is made to the key documents that need to be read prior to use of the data contained in the Census 1981 SAS.

2.1 Small areas

The geography of the 1981 Census in England and Wales is described in detail in Denham (1980). Very briefly the hierarchy of small areas areas and the larger areas to which they sum is as follows:

```
                    enumeration districts
                   /                     \
          electoral wards              civil parishes
         (as of April 5, 1981)        (rural areas)
         /                  \
  Parliamentary          County Districts & London Boroughs
  constituencies         (Metropolitan & non-metropolitan)
                                    |
                                    v
                         Counties (Metropolitan & non-
                         metropolitan)
                                    |
                                    v
                         Standard regions
```
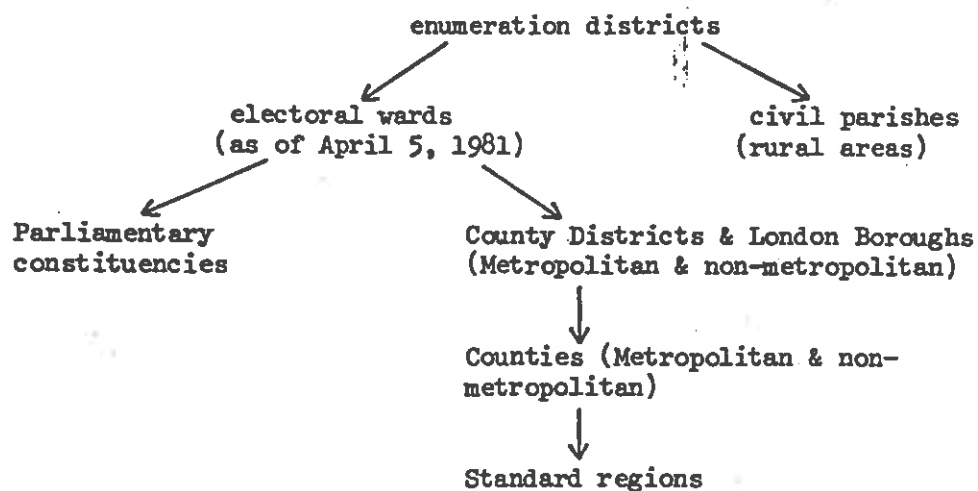
Enumeration districts (EDs) are territorial units covered on Census day by one enumerator. The average ED in England and Wales in 1971 contained circa 450 persons. The boundaries of EDs change substantially from Census to Census, though currently OPCS proposes to put together statistics for amalgamations of EDs roughly equivalent in Census 1971 and Census 1981 (Morgan and Denham, 1982; OPCS, 1981c).

Wards are defined each year in urban areas as the territorial basis of local elections with one or more Councillors being elected from each ward. Ward boundaries change marginally from year to year and can change radically between censuses as a result of local government or parliamentary constituency reorganisation. Those used in Census 1981 are those current for the May 1981 local elections. In rural areas the equivalent units are civil parishes, whose boundaries are less subject to change. The average population size of wards and civil parishes in 1971 was circa 3,000 but in urban areas they average over 20,000 persons. The Metropolitan District of Leeds at the time of the 1981 Census contained 33 wards with average population of 21,360.

2.2 The topics covered

For each small area a standard set of tables are produced in three sections

"1. Population and household tables on a 100 per cent basis, available for areas throughout Great Britain ...

2. Population tables on a 10 per cent basis available throughout Great Britain ...

3. Tables available for areas in England or these tables and the Welsh language table available for areas in Wales only ... or tables available for areas in Scotland only ..."

(OPCS, 1980, p.11). Details of the table layout can be derived from:

1. OPCS (1980) 1981 Census: Small area statistics. Prospectus. OPCS Monitor CEN 80/8. OPCS, London.

2. OPCS (1982c) Census 1981 Small Area Statistics. Tape specifications. 100% data. OPCS, Titchfield, Hants.

3. OPCS (1982b) Census 1981. County Reports. Part 1. HMSO, London.

The general topics covered are:

Demographic characteristics

Economic characteristics

Housing and amenities

Household composition

Socioeconomic and place of birth characteristics.

A full discussion of the population bases, population and household definitions, the definition and classification of communal establishments, the definitions of topic variables and the Census areas is provided in OPCS (1981b).

3.   CENSUS DATA FOR WEST YORKSHIRE WARDS AT THE UNIVERSITY OF LEEDS

3.1   West Yorkshire wards

Figure 1 shows how West Yorkshire Metropolitan County is divided into 5 metropolitan districts and electoral wards.   Table 1 lists and names the wards of Bradford, Calderdale, Kirklees, Leeds and Wakefield Metropolitan districts.

3.2   The computer files for West Yorkshire wards

For each Metropolitan district some four data files have been placed on the Geography program and data library permanent disk on the Amdahl V7 computer at the University of Leeds.   Table 2 lists the names of the twenty files.

The files with type CDATA contain the records supplied by OPCS. There are some 210 records of 80 characters per ward containing circa 3075 data items or counts.   Currently only 100% count data are available in the CDATA files.   Sample or 10% count data will be added when they become available.

Type BOUNDARY files contain the raw coordinate data for ward boundaries suitable for input to the GIMMS mapping system (Waugh, 1978) or for users to employ with other mapping systems.   The files of type GIMMS contain the same boundary information already processed by the GIMMS package and ready for further use with that system.

The MEMO files contain brief documentation on each district's set of files.   An example is shown in Figure 2.

3.3   Accessing the census data

There are two routes to accessing the Census data, depending on username status.

(1)   For undergraduates (with usernames GEO2xx or GEO3xx)

You are automatically linked to the GEO6LIB library disk when you logon, with read only access.

Files behave as if they are on your 'D' disk.

eg.   To obtain a listing on the screen.
        TYPE LEEDS81 MEMO D

This will inform you about the data available.

LEEDS

BRADFORD

WAKEFIELD

KIRKLEES

CALDERDALE
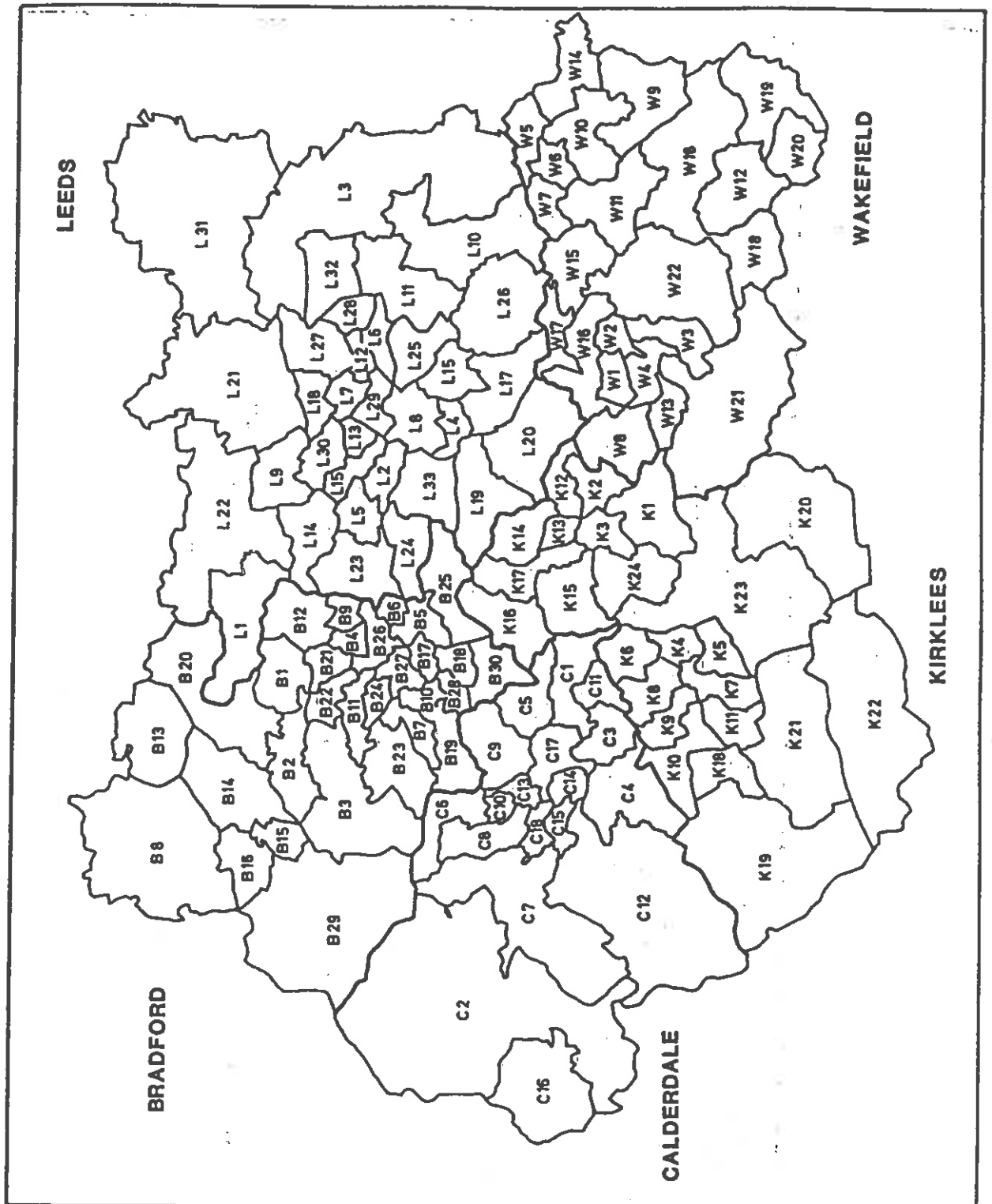
Table 1.   Ward names in West Yorkshire, Census 1981

BRADFORD		KIRKLEES		LEEDS	
B1	Baildon	K1	Thornhill	L1	Aireborough
B2	Bingley	K2	Dewsbury East	L2	Armley
B3	Bingley Rural	K3	Dewsbury West	L3	Barwick & Kippax
B4	Bolton	K4	Dalton	L4	Beeston
B5	Bowling	K5	Almondbury	L5	Bramley
B6	Bradford Moor	K6	Deighton	L6	Burmantofts
B7	Clayton	K7	Newsome & Central	L7	Chapel Allerton
B8	Craven	K8	Birkby & Fartown	L8	City & Holbeck
B9	Eccleshill	K9	Lindley	L9	Cookridge
B10	Great Horton	K10	Paddock	L10	Garforth &
B11	Heaton	K11	Crosland Moor		Swillington
B12	Idle	K12	Batley East	L11	Halton
B13	Ilkley	K13	Batley West	L12	Harehills
B14	Keighley North	K14	Batley North	L13	Headingley
B15	Keighley South	K15	Hightown	L14	Horsforth
B16	Keighley West	K16	Cleckheaton	L15	Hunslet
B17	Little Horton	K17	Heckmondwike	L16	Kirkstall
B18	Odsal	K18	Colne Valley East	L17	Middleton
B19	Queensbury	K19	Colne Valley West	L18	Moortown
B20	Rombalds	K20	Denby Dale	L19	Morley North
B21	Shipley East	K21	Holme Valley North	L20	Morley South
B22	Shipley West	K22	Holme Valley South	L21	North
B23	Thornton	K23	Kirkburton	L22	Otley & Wharfedale
B24	Toller	K24	Mirfield	L23	Pudsey North
B25	Tong			L24	Pudsey South
B26	Undercliffe			L25	Richmond Hill
B27	University	WAKEFIELD		L26	Rothwell
B28	Wibsey			L27	Roundhay
B29	Worth Valley	W1	Wakefield No. 1	L28	Seacroft
B30	Wyke	W2	Wakefield No. 2	L29	University
		W3	Wakefield No. 3	L30	Weetwood
		W4	Wakefield No. 4	L31	Wetherby
CALDERDALE		W5	Castleford No. 5	L32	Whinmoor
		W6	Castleford No. 6	L33	Wortley
C1	Brighouse	W7	Castleford No. 7		
C2	Calder Valley	W8	Osset No. 8		
C3	Elland	W9	Pontefract No. 9		
C4	Greetland &	W10	Pontefract No. 10		
	Stainland	W11	Featherstone No. 11		
C5	Hipperholm &	W12	Hemsworth Urban No. 12		
	Lightcliffe	W13	Horbury No. 13		
C6	Illingworth	W14	Knottingley No. 14		
C7	Luddenden Foot	W15	Normanton No. 15		
C8	Mixenden	W16	Stanley No. 16		
C9	Northowram & Shelf	W17	Stanley No. 17		
C10	Ovenden	W18	Hemsworth Rural No. 18		
C11	Rastrick	W19	Hemsworth Rural No. 19		
C12	Ryburn	W20	Hemsworth Rural No. 20		
C13	St. John's	W21	Wakefield Rural No. 21		
C14	Skircoat	W22	Wakefield Rural No. 22		
C15	Sowerby Bridge				
C16	Todmorden				
C17	Town				
C18	Warley				

Table 2.   A list of Census 1981 data and associated files held on
the Geography program and census data library disk

filename	file type	filename	filetype
For Bradford M.D.		**For Leeds M.D.**	
BRADFD81	CDATA	LEEDS81	CDATA
BRADFD81	BOUNDARY	LEEDS81	BOUNDARY
BRADFD81	GIMMS	LEEDS81	GIMMS
BRADFD81	MEMO	LEEDS81	MEMO
For Calderdale M.D.		**For Wakefield M.D.**	
CALDER81	CDATA	WAKEFD81	CDATA
CALDER81	BOUNDARY	WAKEFD81	BOUNDARY
CALDER81	GIMMS	WAKEFD81	GIMMS
CALDER81	MEMO	WAKEFD81	MEMO
For Kirklees M.D.			
KIRKLS81	CDATA		
KIRKLS81	BOUNDARY		
KIRKLS81	GIMMS		
KIRKLS81	MEMO		

General Notes

1.   CDATA type files contain the raw census data by ward.

2.   BOUNDARY type files contain the boundary coordinates and associated
     information for use with the GIMMS package.

3.   GIMMS type files contain the BOUNDARY data in processed form. Do
     not attempt to list.

4.   MEMO type files provide brief documentation on each district set
     of files.

5.   The CDATA files contain the following numbers of 80 character records:

BRADFD81:	6300
CALDER81:	3780
KIRKLS81:	5040
LEEDS81:	7140
WAKEFD81:	4620

Leeds 1981 Census Data: Memo

(1) OPCS file specifications: there are 3 copies of this document in the Geography Map Library (1 copy for reference, 1 copy for loan, 1 copy for reference in 41 University Road).

(2) Printouts of the raw data are also held in the Geography Map Library.

(3) Maps of the ward boundaries as of Census 1981 (April 5th/6th) are also held in the Geography Map Library.

(4) Details of software that can be used to access the data can be found in:-

Birkin, M.H. and Rees, P.H. (1982) The 1981 small area statistics for West Yorkshire wards: how to access, extract and analyze. Working Paper, School of Geography, forthcoming.

(5) The following files are held on the b disk of the Geography program and data library (GEO6LIB):

LEEDS81 CDATA	Census 1981 100% population and households, small area statistics for wards.
LEEDS81 BOUNDARY	Raw boundary coordinate data for Leeds wards at Census 1981.
LEEDS81 GIMMS	Processed boundary coordinate data on Leeds wards for input to the GIMMS mapping system.
LEEDS81 MEMO	Description of the Census 1981 data on Leeds held in the Geography program and data library.

(6) Similar sets of files for Bradford, Kirklees, Calderdale and Wakefield should shortly be available in the Geography program and data library.

Figure 2.   A listing of a sample memo file

(2)   For postgraduates and staff (with usernames GEO5xxx or GEO6xxx)

You must obtain a work disk using the FILES command and then issue the following instructions

LINK GEO6LIB 191 195 RR

ACCESS 195 B/A

You are then linked to the B disk of GEO6LIB.

Having obtained access to the library files, you may wish to printout your own copies for manual extraction of data. However, it would be better to use a program to extract from the census data files those counts of direct use or the statistics of those wards of interest.

Section 4 describes a program for producing ward by count tabulations and data sets that can be employed by the user to extract information from the CDATA files. This program called TABULATE is also stored on the GEO6LIB permanent disk. Figure 3 provides a picture of the activity involved in abstracting a data set from the library. The user prepares a data file of instructions to the TABULATE program and issues LINK and ACCESS commands to connect to the GEO6LIB B disk. The issuing of a TABULATE command then initiates the reading, printing and writing to disk of the census variables that the user has requested.

GEOGRAPHY PROGRAM AND CENSUS DATA LIBRARY DISK

USER'S WORK DISK

FILE

EXAMPLE DATA

BRADFORD

CALDERDALE

KIRKLEES

LEEDS

WAKEFIELD

FILES OF CENSUS 1981 DATA FOR WARDS

ACCESSED

TERMINAL

USER ISSUES LINK, ACCESS AND TABULATE COMMANDS

ACCESSED

ACCESSED

ACCESSED

TABULATE TEXT

TABULATE EXEC

TABULATE MODULE

PROGRAM FILE

COMMANDS FILE

MODULE FILE

FILE

ARRAY FOR ANALYSIS

TABLES OF STATISTICS FOR LEEDS WARDS

PRINTER

Figure 3.   An example of a census data extraction system

4.   A GENERAL PROGRAM FOR THE EXTRACTION AND TABULATION OF CENSUS DATA

4.1   Introduction

In this section and the next we illustrate the relative ease with which it is possible to extract information from the census data files and how the analysis might then proceed.   We begin by looking at the format of the data file itself, and then introduce our simple but general purpose program for extraction of the data.   This may be copied and adapted by a user familiar with FORTRAN77 - sub-section 4.3 will hopefully be of assistance to such users.   Alternatively the program may be used as a package with the instructions provided in subsection 4.4.

4.2   The format of the data

Figure 4 illustrates the structure of the data files for a typical ward.   We see from the figure that the data for the ward are headed by three label records.   The first of these locates the ward within the United Kingdom as a whole, the second gives the name of the ward and the third the name of the metropolitan district within which the ward is located.   The fourth record is another label relating to the SAS themselves, telling us that the first section - containing the 100% population and household data - will follow immediately.   The four label records are followed by 174 records of data, and then another label identifies the end of Section 1 and the beginning of Section 3 of the SAS.   After a further 31 records of data come the header labels for the next ward and the data for that ward in the same format (and so on).   Thus there are a total of 210 records for each ward.

Each of the records containing data is characterised by the appearance of the number 5 in the first column followed by three spaces. There then follow 15 integer numbers in fields of five characters each, making a total of 79 characters in each record.   Note that sequential numbers will only be separated (by a space) if the second is less than 5 digits long (ie. a number less than 10000).   The only data records which do not share this format are the final records in the two distinct SAS sections.   Record number 178 has only two data items (14 characters) and number 210 has three items (19 characters).

Figure 4.   The SAS data for Aireborough ward, Leeds, Census 1981

```
5   4104 4436 3314  894  556  183   221  153 3496 2526  860 2127 1560  437  207
5     59  161  120 3399 1709 1087   312 1740 1214 4937 2619 1652 1054  598  304
5     61  243 1189  325  864 2570   591 197916598 2472 6400 3252 646611802 1792
5   5306 1316 3388 8996 1525 4916  1022 1533 2123   44  136  189 1754  616  212
5    224   92   88   47   11   30    13   13 6796  680 1102 1936 3078  255   35
5    123   48   49 1985  277  957   467  204  731  358   11  342   20 3825   10
5     11 1079 27252410011636 5228 640812464 5995 6469 1441  695  695          746
5    746           1656  852  852        804  804      2030 1025 1025      1005 1005
5          375  187  187        188  188      1472  792  784    0  680  661   19
5   1468  744  549  195  724  367   357 1672  803  248  555  869  149  720 2204
5   1119  173  946 1085  104  981  1725  866  105  761  859  104  755 1455  730
5     84  646  725   78  647 1341   678   62  616  663   80  583 1206  645   70
5    575  641   84  557 1313  613    51  562  700  127  573 1176  953   52  501
5    623  186  437 1105  537   89   448  648  277  371  982  386   66  320  596
5    339  257  741  264   72  192   477  332  145  394  104   43   61  290  239
5     51  184   43   21   22  141   125   16 9028 5458 1319 1456  326   69 1396
5   1330   25   31    5    5  639   535   47   47    0    2  940  918   11    7
5      3    1 2123  903  455  579   148   34 3930 2172  781  792  162      2318598
5  11485 3236 3048  647 14211799  6543 2367 2297  502   90  684  440  129   78
5     32    5 6467 3387  375  243  1700  949 1402  965 1712 1110 1278  120 3285
5   1533 1752  234  197   37  921   426  495  936  379  557 1076  504  572  118
5     27   91 1534 1402  225  173   199  173  210  169  157  164  165  156  156
5    136  133  156  141  130  140   145  202  376    2    0    1    1    0    4
5      5   14    8   46   15   42    34   83   50   85   87  101  354  339  156
5    127   62   70   42   51   26    31   20   26   27   22   11    8    7    3
5      3    1 1174  910   67   44   137  102  167  115  131  126  143  105  127
5     99  123  117  134   95  145   107  120   66   14    5    9    7   25   11
5     16   11   15    9   13    9    13    6   10    0    5    8 5502 2760 2742
5    280  124  156  263  131  132   330  165  165  271  126  145  298  149  149
5    313  169  144  296  156  140   329  163  166  330  164  166  388  201  187
5    405  211  194  369  188  181   432  220  212  421  209  212  401  196  205
5    376  180 18018598 8077 6408 1972  497 9721 6469 1569 1683 4612 2339  750
5   1556   25 2273 1096 1124   53  5384 2715 2353  236  126 2669 2383  109  177
5   4493 2489 2254  122  113 2004  1713   83  208 4109 1334 1043   58  233 2775
5   1277  253 1245 9026 6915 6180   337  396 2111  297  330 1484 1000  849  712
5    123   14  151   62   54   35  2729 2435 2258   94   83  294  114   44  136
5   2669 2374 2208   73   93  295    64   43  188 2628 1257 1002   47  208 1371
5     57  189 112524101206611951  527  623 3440  785  452 2201 2511 2166 1932
5    207   27  345  160   90   95  9752 8926 8617  147  162  826  370   67  389
5   7652 7064 6783  102  179  588   154   64  370 4186 2505 2179   71  255 1641
5    101  231 1349    4    4   17    11   41   28    3    3    0   11    8    3
5     27   23    4   33    7  112    80  254  144    7    2    5   73   37   36
5    127   60   67   25   25    0     0   24   24    0    0    0    3    0    0
5      3    0    0    0    0    0     0   25   25    0    0   24   24    0
5      0 9027 1349  517  167 2396  2218 1016  626  581  157 9014 6356 2002  220
5     25   58  286   67 1346  602   512  128    0    0   98    6  515  322  120
5     17    2    4   22   28  167    75   80    2    0    0    9    1 2392 1727
5    509   45    5   18   81    7  2217 1906  254    7    8   18   21    3 1016
5    775  210    1    5    6   17     2  626  510  104    2    3    4    3    0
5    579  350  155   17    2    6    31   18  157   91   61    0    0    1    3
5      1   13    3    2    0    4     1    0    0    2    0 2252 2246 1145  774
5    165    0    3  153    6    6  1442 4060 2792 1318 1442 1186  219    5    5
5     10   10    7 4058 3272  683    10   14   31   47    1 2786 1627  904   97
5      2    9  145    2 1313  669   430  109    1    0  100    4    0    2    6
5      5  752 1911   65  232  745     2    2 1899    2    2    3  226  453
5   2006  268 1994    2    3    3   455 3484  528 3476    2    4    0  631  392
5     48  390    1    0    0  198   489  236  489    0    0  115 3169  168
5     42  252 3001 6944 1399 3809   396  510   10    0   30  500 1333  236  921
5    208 3163    2    3  167    0     0   42    0    0  250    0    0 2996    2
5      3 6930    5    6 1398    0     1 3801    4    3  396    0    0    0    0
5      0    0    0    0    0    0     0  591  114   30  171  477 1090  268  616
5    107 3064  117  101  611  523   575  330  573  234  160    4    5    7   34
5     30   27   17   36 3012   20    13  114    0    0  100    0    0  601    4
5      2  497    9    6  570    3     2  325    2    2  572    1    1  233    1
5      0   19    1    0    4   14     0    0    0    0 2002   83   80  534  507
5    273  238  231   56 5335 4109  1319  183  400  231  100   20 5276   30   21
5   4055   27   20 1286   15   12   175    5    1   20   20   14    6 2956 2579
5   1040  149
4 S32598 SAS SECTION 3  ( ENGLAND AND WALES ONLY TABLES )
5   9488 9455  387 9000                 33            13    1 9015 8989  363
5   8567   45    7    7   26   26     0   13    0  473  466   24  433
5            7                  0  145609 104544361  178   25   59   23    4    7
5     25  315  125  164   26    0   630  111  510    9    0 1993   82 1904    7
5      0 4876   22 4848    6    0   952    0  950    2    0  189    0  187    2
5      0   29    1    9    3    0     0            24077 49923452  101   25 1860
5    244 1557   34   25 2908  105  2787   16    0 1618   13 1599    6    0 1848
5      1 1845    2    0  596    0   595    1    0  142    0  142    0    0   43
5      0   42    1    0   22    6     6    0    2    0    0    0   11    0   11
5      0    0  106    0  106    0     0 3633   94 3492   23   24 5264  270 4958
5     36    0 9014  364 8567   59    24    1    1    5    7   14 8949  363 8534
5     52    0    2    0    2    0     0   10    0   10    0    0   19    0    1
5      0   18   15    0    4    7     4    0    0    0    0    0   10    0    7
5      0    3    0    0    0    0     0   10    0   10    0    0    6    0    0
5      0    0    0    1    0    6  9013  363 8567   58   25 6296    2 6287    7
5      0   60   10   50    0    0  2001  160 1794   28   19  220  187   29    4
5      0   26    0   24    2    0    57    2   53    2    0  286    2  275    1
5      0   67    0   55    6    6    15    8    2    2    0    2    0    1
5     57   13   57    0    0    0  112409923529  570 1427   13 3991   69 4502
5    111 5236  148 4348  144 4025  85234452196  249 1393   12 3888   52 4401
5     60 5193   51 4317   52 4004    22  176   93   83    2    0   24    1   46
5      6    6   17   12   36    3    23  207   90  109   15    0   44   11   20
5     23   10   47    3   17    6    11  271  142  129   17    1   35    5   35
5     22   27   33   16   39   12    29 9028 8812   62   57   97 6295  904 1998
5   1227 1570  473   90   25   62    19   24    3   10    3    3    0   58
5    204 1216 2186 1565 1060 629734156    0   79  345 2550 6221 4978 383918012
5      0    0    0    7   24   17    12    0   60  274    0    0    0   50   47
5     36    0  141      6297 6276     8    9    4    7   10 71801217965   16
5     27    4   14   21   14   60    60    0    0    0    0    0    0  142  142
5      0    0    0    0    0    0     3   64   22  362    0    2    0   12 1378
5   2839   16   29
```

Figure 4. Continued

The meaning of each item of data is defined by its position in the file. Thus the first data field of the first record is assigned the identifying number 1 (or V1 for 'variable number 1'). We see from Figure 4 that for Aireborough ward the value of V1 = 24704. Identifying numbers are likewise assigned to the following data items in sequence. Thus the contents of the second field on the first data record are identified as number 2, the first field of the second data record would contain variable number 16 (since there are a total of 15 variables on the first record) and the sixth field of the twelfth data record would be defined as V171 (and so forth). Notice, however, that the simple structure of the data is somewhat disturbed by the presence of the second SAS label on record 179 (see Figure 4).

The next thing we need to know are the meanings of the identifying numbers themselves. These can be found in the 'Coding Book' (OPCS, 1982c), an illustrative sheet of which is shown in Figure 5. Thus, for example, we see that total residents in a ward may be found in location 320: that is, the number may be found in the fifth field of the twenty-second data record. The next number will give the total male population of the ward and so on. As an example the twenty-second to twenty-sixth data records for Aireborough ward in Leeds have been reproduced in full in Figure 6. The fifth field of the twenty-second record contains the number 25020-- this, then, is the total number of residents in Aireborough ward at census day. Reading across the record we see that 12053 of Aireborough's residents were males and 12967 females, 23827 were born in England and so on. By reading across records we can replace the general identifiers of Figure 5 with the actual values for Aireborough ward.

Finally in this section it is important to recognise the significance of those fields in the data which contain no information at all. These blank fields relate to quantities that have not been measured in the census but are given a location to maintain the pattern of the identifying numbers. A blank location is therefore to be distinguished from a field containing a zero, which indicates that an attribute has been counted but does not occur in the ward in question.

4 All residents		
Country of birth	Males	Fmles
1	TOTAL	321	322
2	England	324	325
3	Scotland	327	328
4	Wales	330	331
5	Rest of UK	333	334
6	Irish Rep	336	337
7	Old Comm'th	339	340
8	New Comm'th	342	343
9	East Africa	345	346
10	Africa Rem	348	349
11	Caribbean	351	352
12	India	354	355
13	Bangladesh	357	358
14	Far East	360	361
15	Mediterr	363	364
16	Remainder	366	367
17	Pakistan	369	370
18	Other E.C.	372	373
19	Other Europe	375	376
20	Rest of World	378	379

	TOTAL PERSONS
1	320
2	323
3	326
4	329
5	332
6	335
7	338
8	341
9	344
10	347
11	350
12	353
13	356
14	359
15	362
16	365
17	368
18	371
19	374
20	377

Figure 5.   Country of birth page from the coding book for the Census 1981 SAS (OPCS, 1981)

```
5        0    0    0    725020120531296723821114951232 6  366  175  191  110   54
5       56   92   47   45  145   57   88   56   23   33  152   82   70   16    ë
5       11   16    9    7   12    6    6   52   29   23    8    8    0   29   14
5       15   15    9    6    4    2    2   10    5    5  136   48   88   61   36
5       25   71   31   4019520 9296 2825 647110224 3672 655211882 7125 1815 5310
                                                                                ꞓ
```

Figure 6.   Selected records from the SAS for Aireborough ward, Leeds,
             Census 1981

4.3   Main features of the program

We have seen in the last section how it is possible to fill out the tables of the coding book with their actual values for particular wards.   It is this task which our general purpose extraction and tabulation program performs.

The text of the FORTRAN77 program TABULATE is given in full in Figure 7.   The program reads in the number of wards in the district in question and their names, and then the names and code numbers of the variables to be read off.   In the latter part of the program the variables are written back in a tabular form as a wards times variables array.

Following the initial comments and parameter assignments, the key part of the program in terms of accessing the census data occurs on lines 47-74.   The first 'IF' block (lines 49-53) checks to see if the data item is in Section 1 or Section 3 of the SAS.   From its identifying number the program then computes the record number (NREC) and the field of that record in which the data item is to be found (lines 54-62).   Once NREC has been determined an external ASSEMBLER subroutine ('READ') is called.   This subroutine sets up the required data file with direct access, extracts record number NREC and transfers it to a temporary file on Channel 4.   Data can then be read off as required for the tabulation routine (lines 78-99).

This somewhat circuitous procedure is necessary to obtain direct access to the 'read only' files of the GEO6LIB disk.   In normal circumstances direct access requires both read and write facilities.

In addition there is a facility for variables to be printed out in matrix form for further analysis (lines 100-106).

```
@PROCESS  SC (READ,GTARG)                                              TAB00010
C::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::  TAB00020
C       A FORTRAN77 PROGRAM TO EXTRACT AND TABULATE VARIABLES FROM   : TAB00030
C          THE 1981 CENSUS SMALL AREA STATISTICS.                   : TAB00040
C     WRITTEN BY M.H.BIRKIN, SCHOOL OF GEOGRAPHY, UNIVERSITY OF LEEDS. : TAB00050
C                      MAY, 1982.                                    : TAB00060
C::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::  TAB00070
C       INPUT CHANNEL IS 1.                                           TAB00080
C       OUTPUT CHANNELS ARE 2 FOR TABLES AND                          TAB00090
C       3 FOR 'NAKED' (?!) DATA.                                      TAB00100
C                                                                     TAB00110
        CHARACTER*80 TITLE, TITLE1, TITLE2                           TAB00120
        CHARACTER*10 ZLABEL(40), VNAME1(10), VNAME2(10)              TAB00130
        CHARACTER*11 DOT(11), COLON(11)                              TAB00140
        CHARACTER *80 BUFF                                           TAB00150
        CHARACTER *8 ARGS(3), BLANK8                                 TAB00160
        INTEGER VARB(10)                                             TAB00170
        REAL DUMMY(15), VAL(10,40)                                   TAB00180
        BLANK8='        '                                            TAB00190
        READ(5,326)ARGS(1),ARGS(2), ARGS(3)                         TAB00200
        IF(ARGS(1).EQ.BLANK8) THEN                                   TAB00210
          WRITE(2,323)                                               TAB00220
          STOP                                                       TAB00230
        ELSE                                                         TAB00240
          CONTINUE                                                   TAB00250
        ENDIF                                                        TAB00260
        DO 5000, K=1,11                                              TAB00270
          DOT(K)='...........'                                       TAB00280
 5000     COLON(K)=':::::::::::'                                     TAB00290
        READ(1,100) TITLE1                                           TAB00300
        READ(1,100) TITLE2                                           TAB00310
        READ(1,*) NZONE                                              TAB00320
        READ(1,102) (ZLABEL(K),K=1,NZONE)                            TAB00330
 1050   READ(1,*) NVARBS                                             TAB00340
        IF (NVARBS.EQ.0) STOP                                        TAB00350
        READ(1,*) NARRAY                                             TAB00360
        READ(1,106) TITLE                                            TAB00370
        DO 1000, L=1,NVARBS                                          TAB00380
          READ(1,104) VNAME1(L), VNAME2(L), VARB(L)                 TAB00390
 1000   CONTINUE                                                     TAB00400
  326   FORMAT(A8)                                                   TAB00410
  323   FORMAT('NO FILENAME GIVEN FOR CENSUS DATA FILE. CAN''T CONTINUE')TAB00420
  100   FORMAT(A80)                                                  TAB00430
  102   FORMAT(7A10)                                                 TAB00440
  106   FORMAT(A80)                                                  TAB00450
  104   FORMAT(2A10,I4)                                              TAB00460
        DO 1001, IWARD=1,NZONE                                       TAB00470
          DO 1001, N=1,NVARBS                                        TAB00480
            IF ((VARB(N).GT.2597).AND.(IWARD.EQ.1)) THEN             TAB00490
              KOUNT=VARB(N)+28                                       TAB00500
            ELSE                                                     TAB00510
              KOUNT=VARB(N)                                          TAB00520
            ENDIF                                                    TAB00530
          I=INT(KOUNT/15)+5                                          TAB00540
          J=KOUNT+75-(I*15)                                          TAB00550
            IF (J.EQ.0) THEN                                         TAB00560
              J=15                                                   TAB00570
              I=I-1                                                  TAB00580
            ELSE                                                     TAB00590
              CONTINUE                                               TAB00600
            ENDIF                                                    TAB00610
          NREC=((IWARD-1)*210)+I                                     TAB00620
          CALL READ(ARGS,NREC,BUFF,80,'F',NGOT,IFAIL)               TAB00630
```

Figure 7. A listing of the TABULATE program

```
          IF (NGOT.NE.80.OR.IFAIL.NE.0) THEN                          TAB00640
            WRITE(2,321)NGOT,IFAIL                                    TAB00650
            STOP                                                      TAB00660
          ELSE                                                        TAB00670
            REWIND 4                                                  TAB00680
            WRITE(4,322)BUFF                                          TAB00690
            REWIND 4                                                  TAB00700
            READ(4,400)(DUMMY(K),K=1,15)                             TAB00710
          ENDIF                                                       TAB00720
          VAL(N,IWARD)=DUMMY(J)                                       TAB00730
 1001     CONTINUE                                                    TAB00740
  321     FORMAT(' BAD DATA FROM CDATA FILE, NGOT= ',I5,'IFAIL= ',I5) TAB00750
  322     FORMAT(A80)                                                 TAB00760
  400     FORMAT(4X,15F5.0)                                           TAB00770
C================================================================TAB00780
C       TABULATION ROUTINE FOLLOWS.                                   TAB00790
C                                                                     TAB00800
          NV=NVARBS+1                                                 TAB00810
          WRITE(2,310) TITLE1, TITLE2                                 TAB00820
  310     FORMAT(1H1,A80/A80)                                         TAB00830
          WRITE(2,311)                                                TAB00840
  311     FORMAT(1H0,'    ')                                          TAB00850
          WRITE(2,312) TITLE                                          TAB00860
  312     FORMAT(1H0,A80)                                             TAB00870
          WRITE(2,313)  (DOT(K),K=1,NV)                               TAB00880
  313     FORMAT(1H ,'...',11A11)                                     TAB00890
          WRITE(2,314)  (VNAME1(M),M=1,NVARBS)                        TAB00900
          WRITE(2,314)  (VNAME2(N),N=1,NVARBS)                        TAB00910
  314     FORMAT(1H ,':',12X,':',10(A10,':'))                        TAB00920
          WRITE(2,315)  (COLON(K),K=1,NV)                             TAB00930
  315     FORMAT(1H ,':::',11A11)                                     TAB00940
          DO 2001, J=1,NZONE                                          TAB00950
            WRITE(2,320) ZLABEL(J), (VAL(I,J),I=1,NVARBS)             TAB00960
  320       FORMAT(1H ,': ',A10,' :',10(F8.0,'  :'))                 TAB00970
 2001     CONTINUE                                                    TAB00980
          WRITE(2,315)  (COLON(K),K=1,N)                              TAB00990
C================================================================TAB01000
C       IF NARRAY=1 HAS BEEN SPECIFIED, NAKED VARIABLES ARE           TAB01010
C         WRITTEN TO CHANNEL 3 FOR FURTHER PROCESSING.                TAB01020
          IF (NARRAY.NE.1) GOTO 1050                                  TAB01030
          DO 1800, N=1,NVARBS                                         TAB01040
 1800     WRITE(3,201)  (VAL(N,IWARD),IWARD=1,NZONE)                 TAB01050
  201     FORMAT (7(8F7.0/))                                          TAB01060
C       RETURN TO READ NEXT SET OF VARIABLES.                         TAB01070
C                                                                     TAB01080
          GO TO 1050                                                  TAB01090
          END                                                         TAB01100
```

4.4   Instructions for using the program

4.4.1   The instructions file

Prepare, using the editor, a file that contains the instructions to the TABULATE program for running the job.   The instructions consist of titles, labels and parameters (for example, the number of zones) and the names and code numbers of the variables to be extracted.   Figure 8 lays out the instructions for preparing this file, which should be of filetype DATA.   A simple example of such a control file is shown in Figure 9.

4.4.2   Accessing the census data

Follow the instructions given in section 3.3 to link your username to the Geography program and census data library and to access the permanent disk from your workdisk.

4.4.3   The TABULATE command for running the program

A TABULATE exec has been set up which enables you to run the program by issuing a command of the general form:

TABULATE filename1 filename2 disklabel

The TABULATE command is followed by three arguments:

(1)   'Filename1' is the name of the file on which you have prepared instructions to the program.   It should be of filetype DATA and be present on your workdisk.   The filename is used to name the file containing the tabulations which has the filetype RESULTS and the file containing the array of raw numbers with filetype ARRAY.

(2)   'Filename2' is the name of the GEO6LIB file that contains the census data you wish to access.   The filenames are listed in Table 2.

(3)   'Disklabel' is an alphabetic character which defines the location of the GEO6LIB disk and depends on your status as a user.   For undergraduate users the letter D should be specified;   for postgraduates and staff the letter B.

Record	Variable	Format	Meaning
1	titlel	a80	Two title records for the results
2	title2	a80	file (up to 80 characters per record)
3	nzone	free	Number of wards in the district of interest
4 TO N	zlabel	7al0	Ward names for nzone wards. Each ward name occupies 10 characters; 7 ward names per record
N+1	nvarbs	free	Number of variables in the first table. nvarbs must be less than or equal to 10.
N+2	narray	free	Parameter governing the production of a raw numbers array. narray = 0 − no array narray = 1 − array produced
N+3	title	a80	Label of up to 80 characters for the first table
N+4	vnamel	al0	Labell for variable selected ($\leq$10 characters)
	vname2	al0	Label2 for variable selected ($\leq$10 characters)
	varb	i4	Code number of variable (as in 'Coding Book', OPCS, 1982c)
	.	.	
	.	.	
	repeat for nvarbs variables		
	.	.	
N+(m+3)	vnamel	al0	As above
	vname2	al0	
	varb	i4	
N+(m+4)	Either	record set N+1 to N+(m+3) is repeated for a new table	
	Or	a zero is entered on a separate record to signal the end of the analysis.	

Figure 8. Instructions for preparing the data file for input to the TABULATE program

```
TABULATION OF SELECTED VARIABLES FROM 1981 CENSUS SMALL AREA
 STATISTICS : DATA FOR 33 WARDS IN LEEDS.
33
  AIREBORO    ARMLEY    BARWICK    BEESTON    BRAMLEY    B'TOFTSCHAPEL ALL
       CITY COOKRIDGE  GARFORTH      HALTON HAREHILLSHEADINGLEY HORSFORTH
    HUNSLET KIRKSTALL MIDDLETON  MOORTOWN MORLEY N. MORLEY S.      NORTH
      OTLEY PUDSEY N. PUDSEY S.RICHMOND H  ROTHWELL  ROUNDHAY  SEACROFT
UNIVERSITY  WEETWOOD  WETHERBY  WHINMOOR    WORTLEY
  6
  1
TABLE 1: TOTAL RESIDENTS BY COUNTRY OF BIRTH - BRITAIN AND IRELAND.
   TOTAL     PERSONS  320
             ENGLAND  323
             SCOTLAND 326
             WALES    329
   REST        UK     332
   IRISH    REPUBLIC  335
 0
```

Figure 9.   A sample instruction file for the TABULATE program

The command used with the sample instructions file shown in Figure 9 was

TABULATE ETHTAB LEEDS81 B

The RESULTS and ARRAY files may be printed in the usual way.

4.4.4 The TABPRINT command for running the program

An alternative TABPRINT exec with exactly the same arguments may be used if you wish to print out the tabulations directly (when there is insufficient room on your workdisk). First of all, spool the printer to an appropriate location, then issue the command:

TABPRINT filename1 filename2 disklabel

In this case no RESULTS file is produced.

4.4.5 Output from the program

The TABULATE program produces tables of census statistics in ward times variable format with user input names for the wards and labels for the variables. Figure 10 shows the table produced by running the program with the ETHTAB instructions of Figure 9. The raw numbers from this table are also written to a separate ARRAY file for further analysis, and Figure 11 shows a listing of the contents of the raw numbers file associated with the ETHTAB example.

TABULATION OF SELECTED VARIABLES FROM 1981 CENSUS SMALL AREA
STATISTICS : DATA FOR 33 WARDS IN LEEDS.

TABLE 1: TOTAL RESIDENTS BY COUNTRY OF BIRTH - BRITAIN AND IRELAND.

	TOTAL PERSONS	ENGLAND	SCOTLAND	WALES	REST UK	IRISH REPUBLIC
AIREBORO	25020.	23821.	366.	110.	92.	145.
ARMLEY	22887.	21137.	466.	91.	116.	276.
BARWICK	21821.	21013.	306.	85.	52.	76.
BEESTON	17646.	16259.	264.	56.	93.	300.
BRAMLEY	21664.	20746.	268.	71.	90.	127.
B'TOFTS	21916.	20436.	286.	57.	138.	398.
CHAPEL ALL	23663.	17822.	370.	104.	186.	473.
CITY	21517.	19207.	533.	88.	146.	437.
COOKRIDGE	21026.	19503.	430.	145.	120.	156.
GARFORTH	24006.	23110.	295.	107.	70.	92.
HALTON	19570.	18755.	218.	81.	64.	143.
HAREHILLS	22775.	17980.	439.	101.	163.	745.
HEADINGLEY	16114.	12713.	353.	125.	137.	389.
HORSFORTH	21577.	20477.	347.	96.	71.	139.
HUNSLET	16009.	14711.	315.	49.	127.	132.
KIRKSTALL	19659.	17758.	366.	75.	163.	389.
MIDDLETON	19757.	19028.	250.	56.	67.	116.
MOORTOWN	20113.	18050.	324.	127.	93.	208.
MORLEY N.	21424.	20754.	196.	69.	51.	80.
MORLEY S.	22678.	21986.	214.	67.	49.	73.
NORTH	20277.	18623.	393.	145.	108.	157.
OTLEY	23166.	22099.	346.	143.	65.	119.
PUDSEY N.	23023.	21945.	257.	96.	58.	86.
PUDSEY S.	22227.	21330.	223.	84.	65.	121.
RICHMOND H	22711.	21418.	290.	72.	150.	334.
ROTHWELL	21271.	20507.	216.	86.	61.	84.
ROUNDHAY	20304.	18042.	446.	172.	132.	295.
SEACROFT	20992.	19916.	289.	79.	146.	223.
UNIVERSITY	17706.	14392.	478.	99.	207.	491.
WEETWOOD	18250.	16669.	406.	134.	95.	265.
WETHERBY	23396.	21802.	607.	298.	103.	146.
WHINMOOR	19279.	18214.	311.	71.	102.	141.
WORTLEY	23263.	21982.	327.	76.	147.	206.

Figure 10.   A sample tabulation produced by the TABULATE program

25020.	22887.	21821.	17646.	21664.	21916.	23663.	21517.
21026.	24006.	19570.	22775.	16114.	21577.	16009.	19659.
19757.	20113.	21424.	22678.	20277.	23166.	23023.	22227.
22711.	21271.	20304.	20992.	17706.	18250.	23396.	19279.
23263.							
23821.	21137.	21013.	16259.	20746.	20436.	17822.	19207.
19503.	23110.	18755.	17980.	12713.	20477.	14711.	17758.
19028.	18050.	20754.	21986.	18623.	22099.	21945.	21330.
21418.	20507.	18042.	19916.	14392.	16669.	21802.	18214.
21982.							
366.	466.	306.	264.	268.	286.	370.	533.
430.	295.	218.	439.	353.	347.	315.	366.
250.	324.	196.	214.	393.	346.	257.	223.
290.	216.	446.	289.	478.	406.	607.	311.
327.							
110.	91.	85.	56.	71.	57.	104.	88.
145.	107.	81.	101.	125.	96.	49.	75.
56.	127.	69.	67.	145.	143.	96.	84.
72.	86.	172.	79.	99.	134.	298.	71.
76.							
92.	116.	52.	93.	90.	138.	186.	146.
120.	70.	64.	163.	137.	71.	127.	163.
67.	93.	51.	49.	108.	65.	58.	65.
150.	61.	132.	146.	207.	95.	103.	102.
147.							
145.	276.	76.	300.	127.	398.	473.	437.
156.	92.	143.	745.	389.	139.	132.	389.
116.	208.	80.	73.	157.	119.	86.	121.
334.	84.	295.	223.	491.	265.	146.	141.
206.							

Figure 11.   A listing of the raw data file associated with the sample
tabulation

5. TECHNIQUES IN THE EXTRACTION AND FURTHER PROCESSING OF SPECIFIC
    DATA SETS

In this section we narrow the focus of our attention to one
particular table of the SAS.   We have chosen to investigate spatial
patterns of place-of-birth groups of the resident population, the
variables for which were listed in Figure 5.

In general terms there are three operations which we would like
to perform on the data.   Firstly, we would like to undertake a
statistical analysis.   Secondly, we should like to map the variables
so derived (or possibly in some variant of their crude form).
Finally, it is necessary to make various tabulations so that the
information is not lost in the cartographic exercise.   It will be
appreciated that the aim of mapping work in this context is better
seen in terms of 'clarification through simplification', rather than
an attempt to represent the whole of the data in a map form which
could then be used for autonomous statistical 'quarrying'.   The
overall process is summarised as Figure 12 around which the following
sections are built.

5.1   Extraction

Our need for data from the census file in this analysis is
specific to the extent that only one output file is required, to be
used as an input to IOFD FORTRAN, a program written at Leeds (Rees,
Brookes and Sellers, 1981) to compute the indices of dissimilarity
between social groups in cities.   The extraction program used is
reproduced in full as Figure 13.

We use essentially the same direct access read procedure as
for the TABULATE program with a temporary file on channel 4.   The
first data item is number 320 which, as we saw earlier, comes five
fields into the twenty-second data record - the twenty-sixth in all -
hence the determination of NREC in line 16.   The five records of
data containing the ethnic variables are all read into the temporary
file (lines 18-29).   Since the variable codes follow simple and
well-defined patterns (see Figure 5), the 'DO' loops can be nested
straightforwardly to extract the data from the temporary file
(lines 32-33).
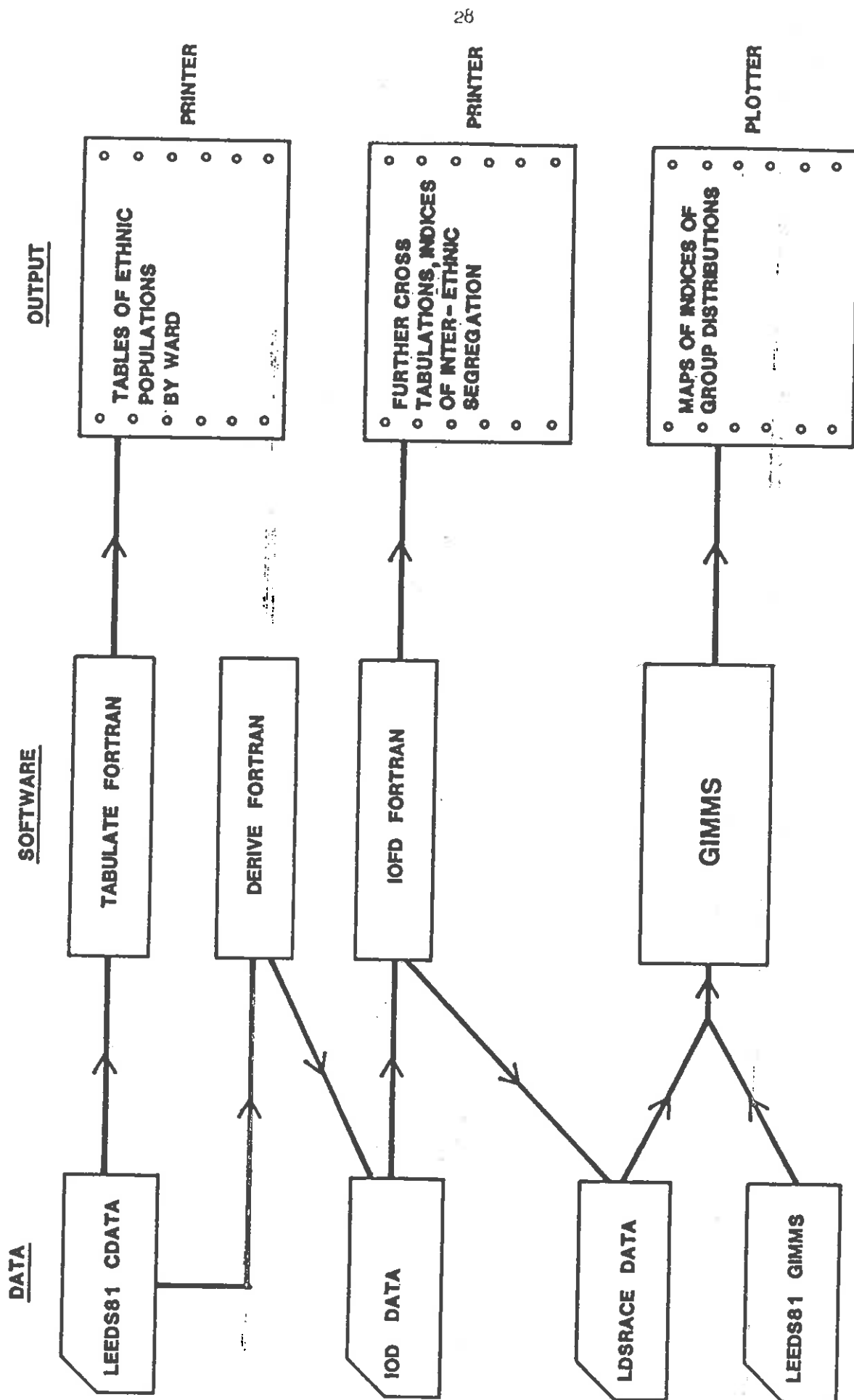
Figure 12. A system of extracting and analysing ethnic variables

```
@PROCESS SC (READ,GTARG)                                         DER00010
C***********************************************************     DER00020
C    A FORTRAN PROGRAMME TO EXTRACT VARIABLES FROM    *          DER00030
C  TABLE 4 OF THE UNPROCESSED CENSUS DATA FILE, AND   *          DER00040
C  TO ARRANGE THE DATA FOR INPUT TO IOFD.             *          DER00050
C***********************************************************     DER00060
      REAL*8 ETHNIC(33,20,3), COLTOT(20), IX(33)                 DER00070
      CHARACTER *6 GROUP(20)                                     DER00080
      CHARACTER *10 WARD(33)                                     DER00090
      CHARACTER*80 BUFF                                          DER00100
      CHARACTER*8 ARGS(3)                                        DER00110
      ARGS(1)='LEEDS81 '                                         DER00120
      ARGS(2)='CDATA   '                                         DER00130
      ARGS(3)='B       '                                         DER00145
      DO 999, I=1,33                                             DER00150
       NREC=((I-1)*210)+26                                       DER00160
       MREC=NREC+4                                               DER00170
       DO 998, NUM=NREC,MREC                                     DER00180
        CALL READ(ARGS,NUM,BUFF,80,'F',NGOT,IFAIL)               DER00190
         IF ((NGOT.NE.80).OR.(IFAIL.NE.0)) THEN                  DER00200
          WRITE(5,321)NGOT, IFAIL                                DER00210
321       FORMAT('BAD DATA FROM CDATA FILE'/                     DER00220
  1            'NGOT= ',I5/                                      DER00230
  2            'IFAIL=',I5)                                      DER00240
          STOP                                                   DER00250
         ELSE                                                    DER00260
          WRITE(4,322)BUFF                                       DER00270
322       FORMAT(A80)                                            DER00280
         ENDIF                                                   DER00290
998     CONTINUE                                                 DER00300
       REWIND 4                                                  DER00310
       READ(4,101) ((ETHNIC(I,E,X),X=1,3),E=1,20)                DER00320
101    FORMAT (24X,11F5.0/(4X,15F5.0))                           DER00330
999   REWIND 4                                                   DER00340
C------------------------------------------------------------C   DER00350
C   BEGIN ROUTINE TO PRINT DATA TO CHANNEL 2, 'IOD DATA'         DER00360
C------------------------------------------------------------C   DER00370
      WRITE(2,401)                                               DER00380
401 FORMAT('LEEDS CENSUS DATA, 1981 - ETHNIC GROUPS'/            DER00390
  1      'DATA FOR 33 WARDS AND 18 GROUPS'/'18'/'33')            DER00400
      READ(1,501) (WARD(N),N=1,33)                               DER00410
      READ(1,500) (GROUP(M),M=1,20)                              DER00420
500 FORMAT(10A7)                                                 DER00430
501 FORMAT(6A11)                                                 DER00440
      DO 600, I=1,20                                             DER00450
600 COLTOT(I)=0.0                                                DER00460
      DO 1001, J=1,20                                            DER00470
       DO 1000, I=1,33                                           DER00480
1000    COLTOT(J)=COLTOT(J)+ETHNIC(I,J,1)                        DER00490
1001   CONTINUE                                                  DER00500
      DO 1005, K=2,20                                            DER00510
       IF (K.EQ.8) GO TO 1005                                    DER00520
       WRITE(2,410) GROUP(K),COLTOT(K)                           DER00530
410   FORMAT(A10/F7.0)                                           DER00540
1005  CONTINUE                                                   DER00550
      DO 1010, L=1,33                                            DER00560
       WRITE(2,420) WARD(L), ETHNIC(L,1,1)                       DER00570
420   FORMAT(A10/F6.0)                                           DER00580
1010  CONTINUE                                                   DER00590
      WRITE(2,430)                                               DER00600
430 FORMAT('1')                                                  DER00610
      DO 1050, II=1,33                                           DER00620
       WRITE(2,440) (ETHNIC(II,IJ,1),IJ=2,7)                     DER00630
       WRITE(2,441) (ETHNIC(II,JJ,1),JJ=9,20)                    DER00640
440   FORMAT(6F7.0)                                              DER00650
441   FORMAT(6F7.0/6F7.0)                                        DER00660
1050  CONTINUE                                                   DER00670
      STOP                                                       DER00068
      END                                                        DER00069
```

Figure 13.   A listing of the DERIVE program

In the second half of the program the data are sorted into exactly the right form for input into the next program in the chain. Some tables of 'primary' information may also be constructed from the raw data file using the TABULATE program described in the last section. The use of this program allows us to exert tighter control over the presentation of this basic data than is possible with IOFD.

5.2 Statistical analysis

Once the data file has been prepared as described above it is run with the IOFD program. This produces the following sets of statistics:

- Tabular output of zone by group populations. (This information can be used to check that the data has been input correctly but as noted above is not suitable for final presentation).

- Tables of indices of dissimilarity (see next section) between pairs of social groups and between each social group and the whole population.

- Tables of indices of dissimilarity between pairs of city zones and between each city zone and the whole region.

- For each group a table across zones of the following characteristics:

    (a) the proportion of a zone's population belonging to the said group
    (b) the proportion of the total group population locating in the said zone
    (c) the location quotient - an index of the relative concentration of a group's population in particular zones

These tables are particularly important because the information they contain is set to the file IDSRACE DATA and then used in the GIMMS package to produce cartographic representations of the three features. An example of one of these tables is shown in Table 3.

— For each zone a table across groups of the following:

(a) the proportion of the zone's population of particular ethnic origins

(b) the proportion of different group populations locating in the zone

(c) the location quotients for each group in that zone.

In fact what we have, therefore, are the same variables as in the group tables but in a transposed form, to highlight the characteristics of the spatial units rather than the structure of the 'sectoral' units (racial groups). An example of such a table is given in Table 4.

5.3 Mapping

The GIMMS (Geographic Information Management and Mapping System) package is a relatively new implementation at the University of Leeds and so a few comments regarding its salient features may be useful. The major source of information is the Reference Manual (Waugh, 1978). In addition a 'HELP' file on the Amdahl elucidates some of the (many!) grey areas in the manual.

Two types of information are needed for GIMMS — locational information detailing the spatial boundaries of the region under consideration, and non-locational information giving values for different variables to be mapped for the various zones. Any project is likely to have three iterative and ongoing phases according to Waugh (1978, p. 7).

(1) Locational data banks are set up. In our case this means setting up a BOUNDARY data file for Leeds wards, and from this a processed GIMMS file (see Section 3). Processed GIMMS files exist for all the West Yorkshire metropolitan districts and are available on the Geography Library (GEO6LIB) disk. Thus for users interested in analysing West Yorkshire census data this stage of the operation has been effectively short-circuited. For users interested in setting up further locational data files, instructions for digitising area boundaries using the GTCO digitiser and Apple II microcomputer with links to the Amdahl mainframe are given in Birkin (1982).

Table 4. <u>Country of birth statistics for persons resident in Headingley ward, Leeds, Census 1981</u>

ZONE13   : HEADINGLEY

GROUP	PERCENT IN GROUP OF ZONAL POPULATION	PERCENT IN ZONE OF GROUP POPULATION	LOCATION QUOTIENT
1 ENGLISH	78.89	1.98	0.86
2 SCOTS	2.19	3.15	1.36
3 WELSH	0.78	3.77	1.63
4 REST OF UK	0.85	3.88	1.68
5 IRISH	2.41	5.14	2.22
6 OLD COMM	0.22	2.98	1.29
7 EAST AFRIC	1.44	12.76	5.52
8 REST AFRIC	0.63	13.00	5.62
9 CARIBBEAN	1.01	4.17	1.80
10 INDIA	3.34	11.61	5.02
11 BANGLADESH	0.05	1.25	0.54
12 FAR EAST	1.94	21.01	9.08
13 MEDIT	0.10	2.86	1.24
14 REST NEWCW	0.10	8.84	3.82
15 PAKISTAN	2.57	12.44	5.38
16 REST EEC	0.86	4.18	1.81
17 REST EUR	0.59	3.09	1.34
18 REST WORLD	2.04	8.14	3.52
TOTAL	100.00	2.31	1.00

(2)   Set up non-locational data banks.   This operation has basically
two elements.   Firstly, the data to be stored must be obtained.   In
the above example the file LDSRACE DATA is created from the 'group
tables' of the IOFD run.   Secondly, this data must be stored in a
processed form by an initial GIMMS run to create a file called here
LDSRACE DATMAN.   Details of this relatively simple operation are
given in Chapter 5 of the Reference Manual.

(3)   The two sets of data banks are brought together and maps produced
using the application modules described in Chapter 5 of the manual.
An example is given in Figure 14 to demonstrate the simplicity of the
procedure.

To begin with, an instructions file of type 'GIMMS' is set up –
let us call this EXAMPLE GIMMS.   To get access to the GIMMS routines
the command SETUP PACKAGES is issued.   It is also necessary to link
to the GEO6LIB disk to use the locational data files stored there.
Having done this the appropriate locational data file, in this case for
Leeds, is linked to input channel 10.   The non-locational data file
will normally be on the user's own disk (filemode = a) – this is linked
to channel 11.   To process the file the command GIMMS EXAMPLE C1039 is
issued.   This produces two files – the file EXAMPLE LISTING traces the
steps in the execution procedure;   the EXAMPLE PLOT file contains the
codes for plotting the map produced by the run.   If everything has
gone smoothly this may be set to the Calcomp 1039 using the PLOT EXAMPLE
command.

For the ethnic groups project a total of 54 maps were produced
using three measures for each of eighteen groups.   Percentage of zonal
population in a group was mapped using a simple grid shade; the propor-
tion of group population in a zone using proportional squares;   and
location quotients using a combination of grid and bar shading.   It
is hoped that these maps are illustrative of the power of the GIMMS
system, although they are in no sense exhaustive of its potential.

```
TYPE EXAMPLE GIMMS

*PLOTPARM GINO, 'M.H.BIRKIN - SCHOOL OF GEOGRAPHY'
*SYSPARM DIAGS=3
*TEXTPARM ALPHABET=13
*PLOTPROG
  *NEWMAP 10 9.5 FRAME
  *GIMMSFILE 10, FACTOR=0.001, PLOT=0.5,0.5, MAP=550,900
    *DATMAN 33 6 6
    *DATA FILE 11 LEEDSDAT
    *DEFINE IRISH=13
    *SET 5 (IRISH)
    *END
  *LEGEND 0.1 9.5 0.4 2 COMPACT
  *LEVELS 6
  *INTERVALS 0.0 0.25 0.5 1.0 2.0 4.0 8.0 5
  *SYMBOLISM AREA GRID 0.24 0.20 0.16 0.12
                       0.08 0.04
  *TEXT X=0.1,Y=9.0,SIZE=0.3,
    TEXT='@ZPOP. BORN IN REP. OF IRELAND'
  *MAP 5
  *END
*STOP


END OF FILE
R; T=0.01/0.07 17:11:52


 .


SETUP PACKAGES
R; T=0.03/0.06 16:58:48

LINK GEO6LIB 191 195 RR
R; T=0.01/0.01 16:58:58

ACCESS 195 B/A
'195' REPLACES ' B (195) '
B (195) R/O
R; T=0.01/0.01 16:59:04

FI 10 DISK LEEDS81 GIMMS B4
R; T=0.01/0.01 16:59:18

FI 11 DISK LDSRACE DATA A4
R; T=0.01/0.01 16:59:28

GIMMS EXAMPLE  C1039
GIMMS 3.0      VM/CMS Version 1.1        Apr 1982      Leeds University.
YOUR PLOTFILE IS - EXAMPLE PLOT
EXECUTION BEGINS...

R; T=1.19/1.41 16:59:44

PLOT EXAMPLE
PRT FILE 6407 TO  GUP      COPY 001  NOHOLD
- YOUR PRINTER IS NOW SPOOLED TO SYSTEM !
R; T=0.04/0.12 16:59:53
```
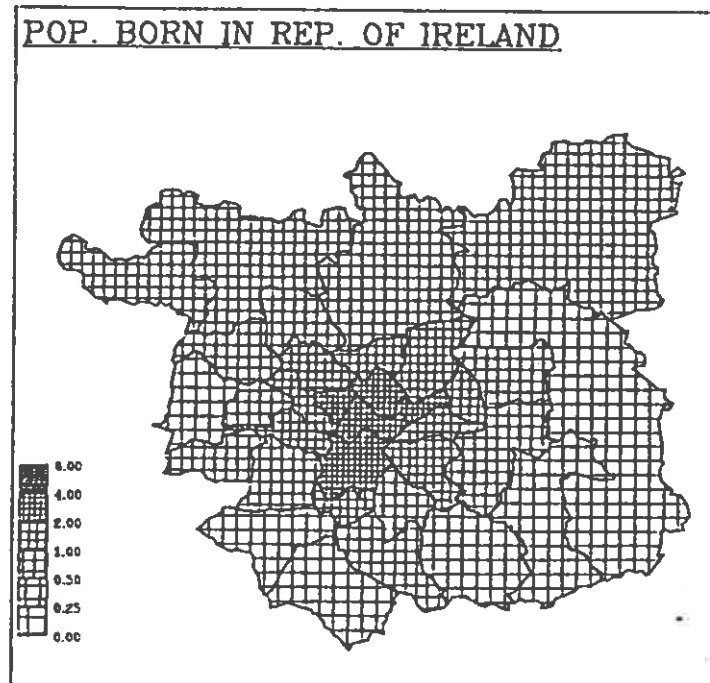

Figure 14.   A specimen run using the GIMMS package

6. THE ETHNIC GEOGRAPHY OF LEEDS, 1981

6.1 Introduction

In this section of the paper, the ethnic geography of Leeds at Census 1981 is described as revealed by the Table 4 SAS statistics for persons by country of birth. There are a number of drawbacks (see section 6.6) affecting the interpretation and use of these statistics as indicators of ethnicity, and there are additional data in Table 37 of the SAS (OPCS, 1982c) and Tables 10, 11, 32 of the County statistics (OPCS, 1982b) which should be employed. Here we assume a rough equivalence between birth in a country and membership of that national ethnic group. However, the Table 4 data are sufficiently interesting in their own right to justify a careful description of the patterns of residential location, and thus to provide the reader of this technical manual filled with computing jargon with his geographic reward.

6.2 An ordering and classification of the ethnic groups

The first question we pose is 'how segregated spatially is each ethnic group?'. We can answer this question by comparing the spatial distribution of the group with that of the whole population, using the most frequently used index — the index of dissimilarity. The index of dissimilarity is defined as:

$$D_{ef} = 100 \times 0.5 \times \sum_{i=1}^{N} \left| \frac{P_e^i}{P_e^*} - \frac{P_f^i}{P_f^*} \right|$$

where D_{ef} is the index of dissimilarity between group e and group f of the population; P_e^i and P_f^i are the populations of group e and f respectively in zone i; P_e^* and P_f^* are the populations of groups e and f in all zones, of which there are N; and $|a|$ indicates the absolute value function where $|a| = a$ and $|-a| = a$. The index has a range of 0 (no segregation) to 100 (maximum segregation).

The first column of Table 5 gives the index of dissimilarity between each of 18 ethnic groups and the whole population. The country of birth groups have been arranged in ascending order of the index. Comparison of the ethnic group with the whole population has the drawback than that we are comparing the group with itself, to a certain

Table 5.   Key indicators of ethnic residential segregation, Leeds, 1981

Country of birth group	Index of dissimilarity versus			Percentage of Leeds population (4)	No. in group (5) -
	Whole population (1)	Rest of population (2)	English (3)		
England	2.1	26.9	C.0	92.18	642,202
FAMILY I					
Scotland	12.3	12.5	14.3	1.61	11,193
Old Commonwealth	15.7	15.8	17.2	0.17	1,173
EED	15.7	15.8	17.4	0.47	3,305
Wales	16.0	16.1	17.3	0.48	3,314
Rest of the UK	17.7	17.8	19.3	0.51	3,530
FAMILY II					
NC Mediterranean	21.7	21.7	23.2	0.08	562
Irish Republic	28.2	28.5	3C.1	1.09	7,563
Rest of Europe	33.5	33.7	35.4	0.44	3,069
Rest of World	34.4	34.6	36.3	0.58	4,044
NC Far East	38.8	38.9	40.8	0.21	1,493
FAMILY III					
NC East Africa	43.0	43.1	45.0	0.26	1,816
NC India	43.7	44.0	45.6	0.66	4,627
NC Rest of New Comm.	44.8	44.9	45.4	0.03	177
NC Rest of Africa	45.1	45.1	46.9	0.11	779
FAMILY IV					
NC Caribbean	61.1	61.4	62.9	0.56	3,889
Pakistan	62.7	63.0	64.6	0.48	3,337
NC Bangladesh	75.6	75.7	77.1	0.09	641

Source: Output of the IOFD program using Country of birth statistics for
           Leeds wards, Census 1981.

Column (2): the index of residential segregation
Column (3): the index of spatial assimilation

NC = New Commonwealth
EEC = European Economic Community

All statistics refer to the usually resident population.   The total
usually resident population of Leeds at Census 1981 was 696,714.

extent.   It is better to subtract the group itself from the whole
population and make a comparison between the group and the rest of
the population.   This is done in column (2) of Table 5.   The index
is conventionally labelled 'the index of residential segregation'.
All indices are raised, though only that for the English changes
significantly because the English born make up 92 per cent of the
Leeds population.   An even more critical comparison is between each
group and the host population (English born) which is given in the
third column of the table.   We will call this last statistic 'the
index of spatial assimilation'.   The ordering of the non-English
born groups remains constant over these three indices of dissimilarity.

The ethnic groups fall fairly simply into four classes which we
have termed 'families of ethnic groups'.

The groups in Family I have indices of spatial assimilation
between 14 and 20.   Members of the groups are either born in other
parts of the United Kingdom or in the three dominions (Australia,
Canada and New Zealand) or in the other countries of the European
Community bar Eire.   The groups in Family II exhibit spatial assimi-
lation indices between 23 and 41, and comprise a mixed group of
European (apart from the EEC bar Eire) and American countries together with
the colony of Hong Kong.   Family III groups show indices of spatial
assimilation between 45 and 47 and comprise four groups of New
Commonwealth countries.   The fourth family have indices above 60 and
include Caribbean born, Pakistan born and Bangladesh born Leeds
residents.

6.3   The spatial patterns

6.3.1   Proportional distribution

The index of dissimilarity gives a one dimensional picture only
of the residential pattern of an ethnic group.   Very different spatial
patterns can be associated with the same index number.   One way to map
spatial distributions of populations is to represent the number in a
zone by a symbol proportional to the number.   Because of the wide dis-
parity in group size (columns 4 and 5 of Table 5), it is difficult to
produce comparable maps using the same symbol scale.   So the propor-
tional numbers (P_e^i/P_e^*) were mapped instead, in effect reducing all maps
to the same symbol scale.

Figures 16 through 18 show the resulting maps*, arranged in
ascending order of the index of spatial assimilation and grouped roughly
in families.   The Family I maps (Figure 16) show fairly even distribu-
tions, although they are clearly more spatially uneven than the English
born map.   The Family II maps (Figure 17) show a degree more uneveness.
All show a tendency to concentration in the more central wards.   The
Irish born show a wide spread over the central wards, the Rest of
Europe born (mainly Poles) show a concentration in the Chapeltown (the
ward for traditional settlement by poorer immigrant groups) and Moortown
(the ward in which these groups spread as their circumstances improved).
The concentration of Far East born in Headingley and University wards
undoubtedly reflects their association as faculty or students with the
educational institutions located there (Leeds University, Leeds
Polytechnic, various Colleges of Higher Education, a major teaching
hospital).

The Family III groups (Figures 17 and 18) show increased spatial
concentration in particular wards, although members of the groups are
found in virtually all wards.   Note that the spatial patterns of those
born in East Africa and in India are very close, reflecting the close
ethnic ties between the groups since most of the former are displaced
Indian-origin families from Kenya and Uganda.   The final three maps
for Family IV groups show the most intense concentration.   The
Bangladesh-born are the only group to show total absence from a sub-
stantial number of wards (14 out of 33).   In fact, the only other
group to show such absence is the tiny 'Born in the Rest of the New
Commonwealth' group (5 out of 33).

6.3.2   Percentage distributions

A more frequently used method is the shading map of the percentages
that a group makes up of the zonal populations.   Figures 19 through 21
show the choropleth maps of zonal percentages for the 18 country-of-birth
groups.   The zonal percentages are defined as:

$$q_c^i = 100 \left(\frac{p_c^i}{p_*^i} \right)$$

Three different scales with six shading categories are used: one for the
English, one for the larger or more concentrated groups (born in Scotland,

*Figure 15 provides the ward names and boundaries associated with these
maps together with an indication of the continuously built up area.

Figure 15.

Built-up Areas and Electoral Ward Boundaries,
Census 1981, Leeds Metropolitan District

Central Business
District

Built-up Areas

ENGLAND

SCOTLAND

OLD COMMONWEALTH

REST OF EEC

1·0
0·5
0·25
0·125

WALES

REST OF UK
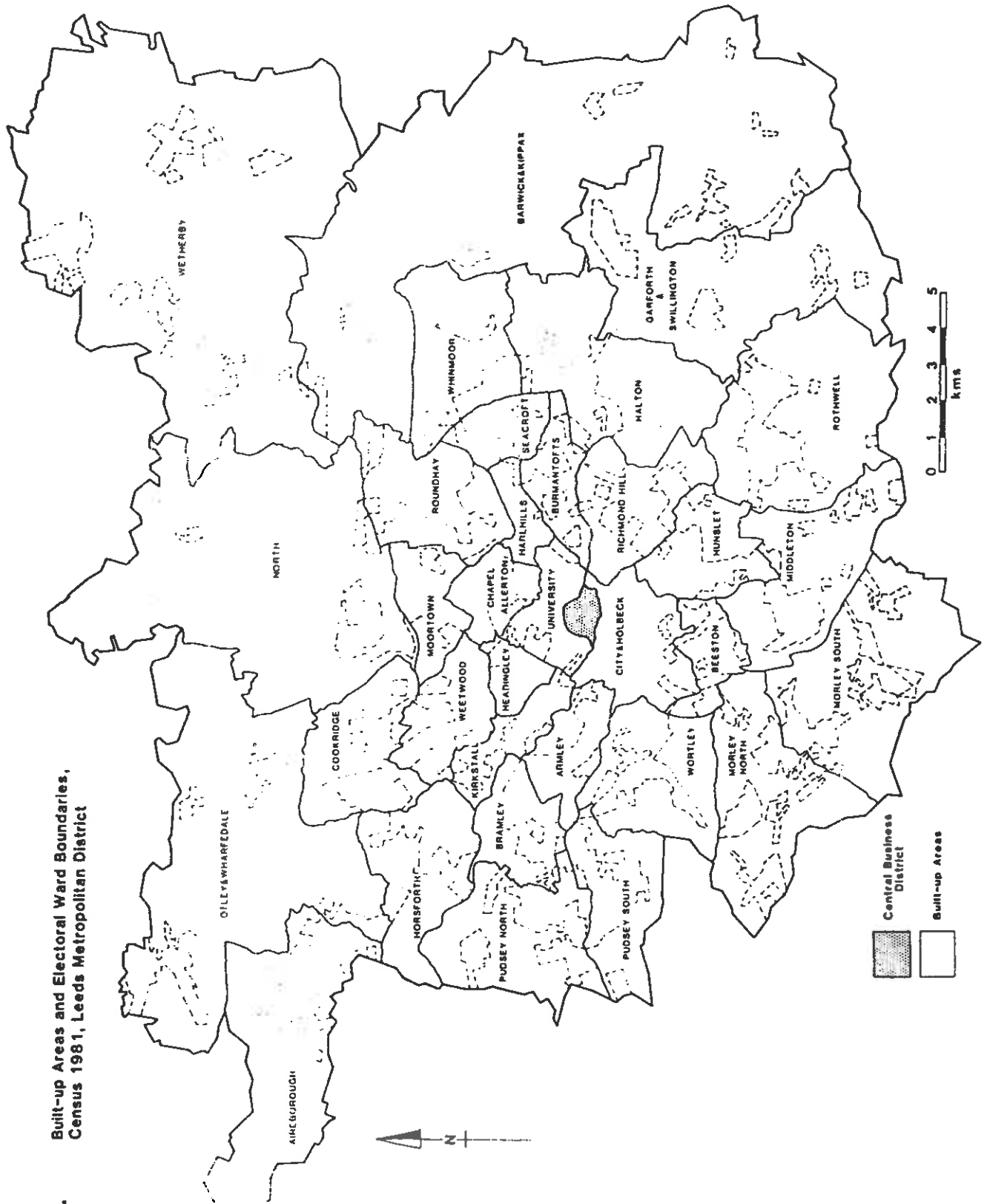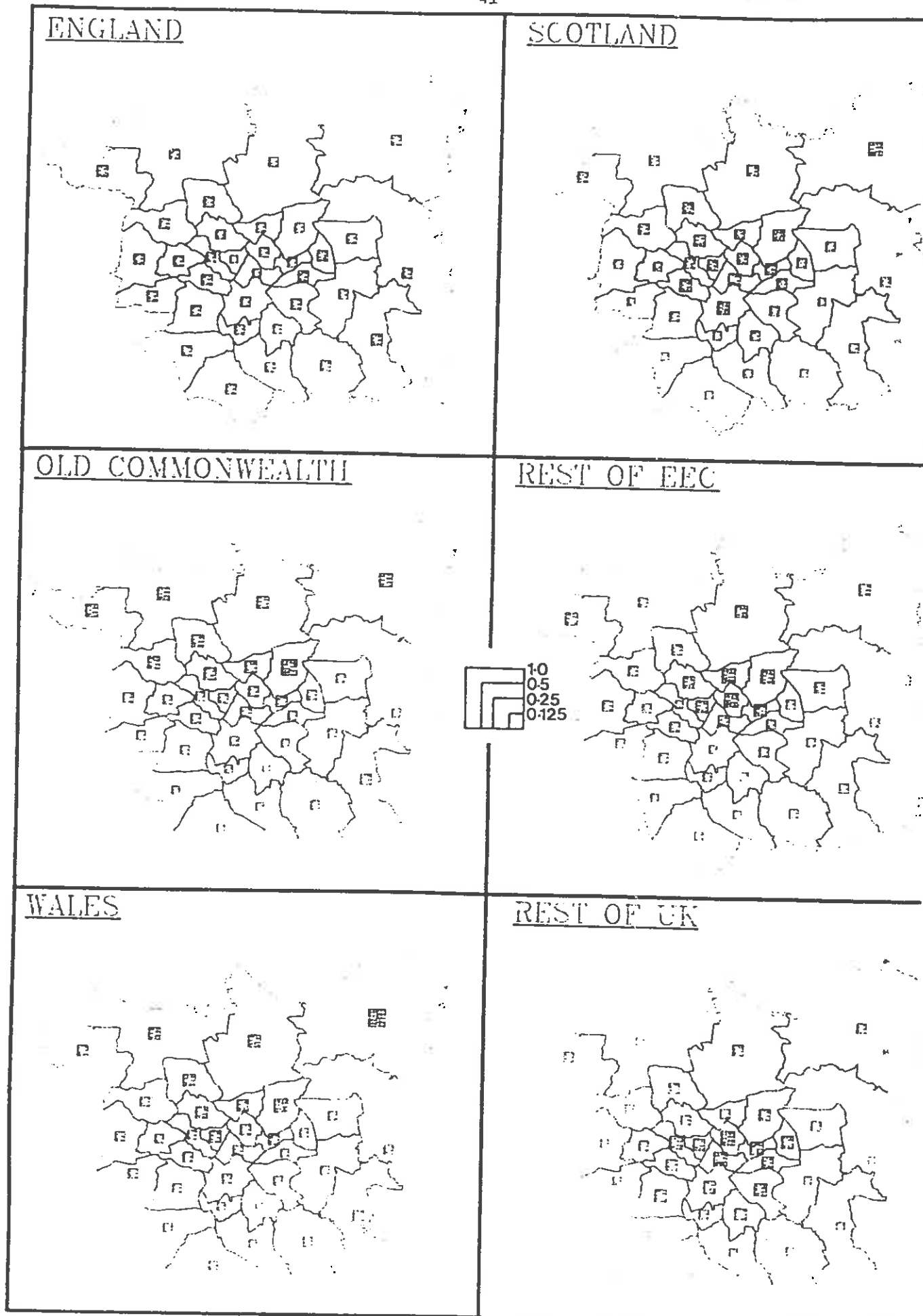
Figure 1c. The proportional distribution of ethnic groups in Leeds, 1981:
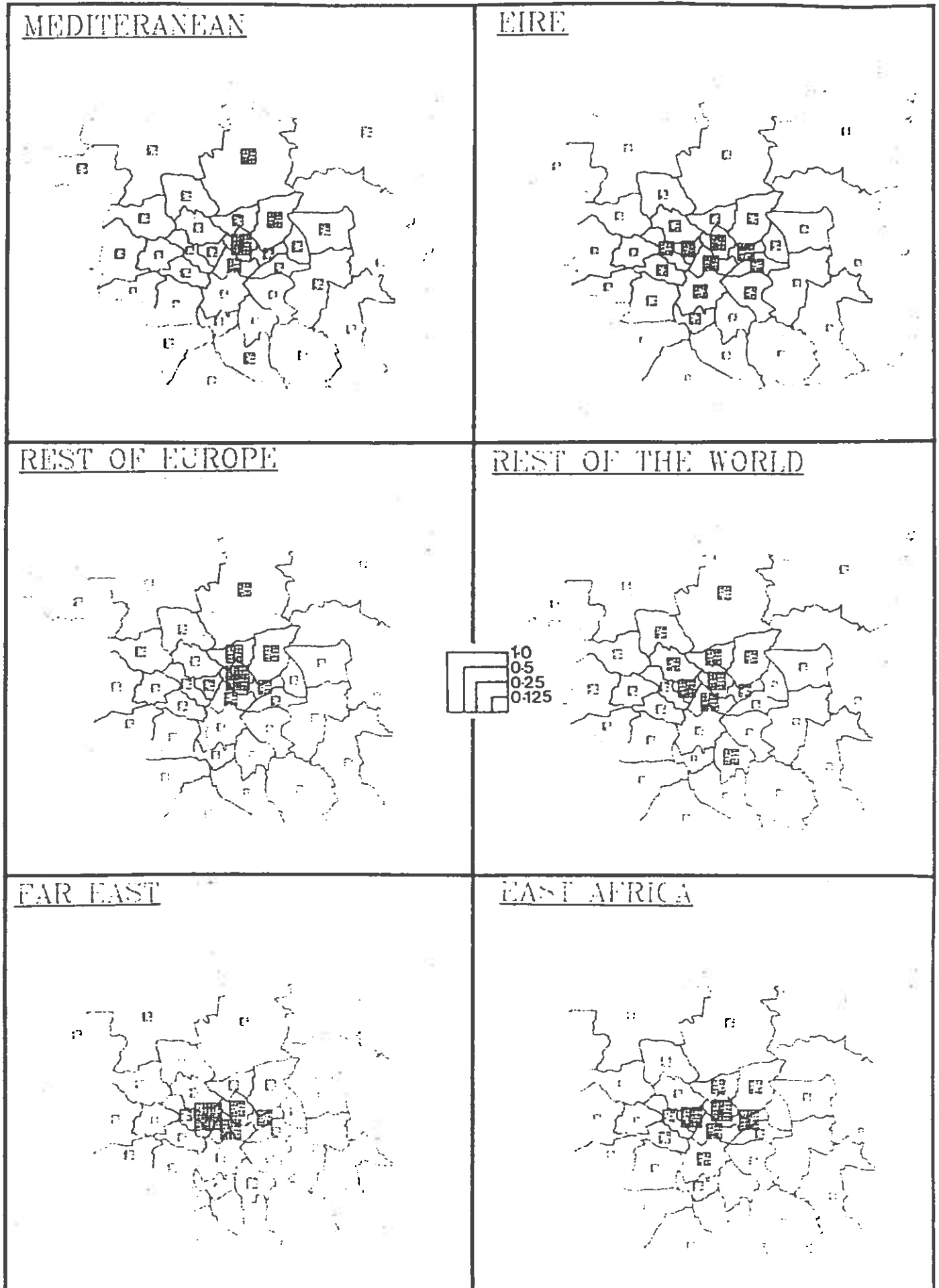
Figure 17. The proportional distribution of ethnic groups in Leeds, 1981:
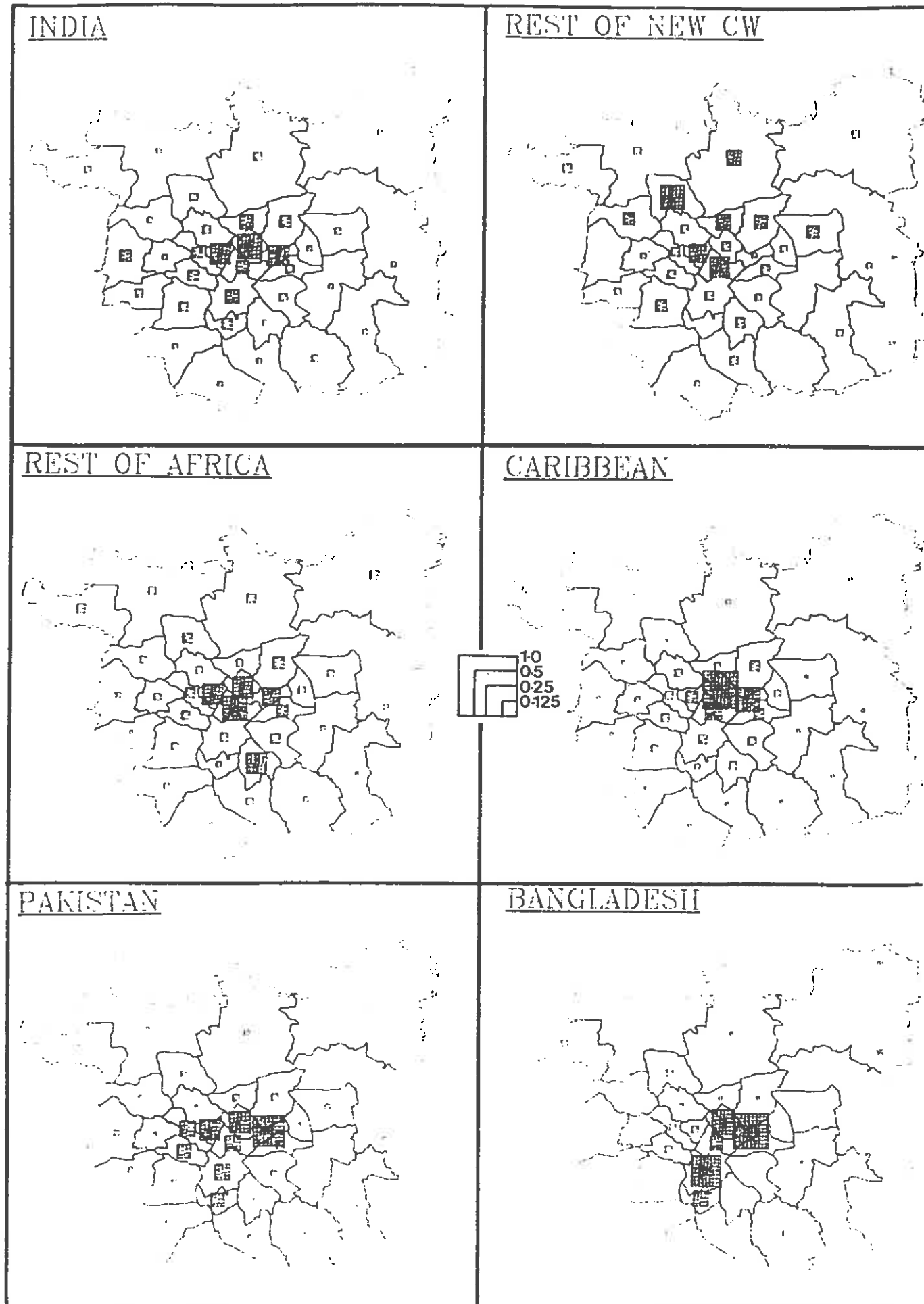
Figure 18. The proportional distribution of ethnic groups in Leeds, 1981:

ENGLAND
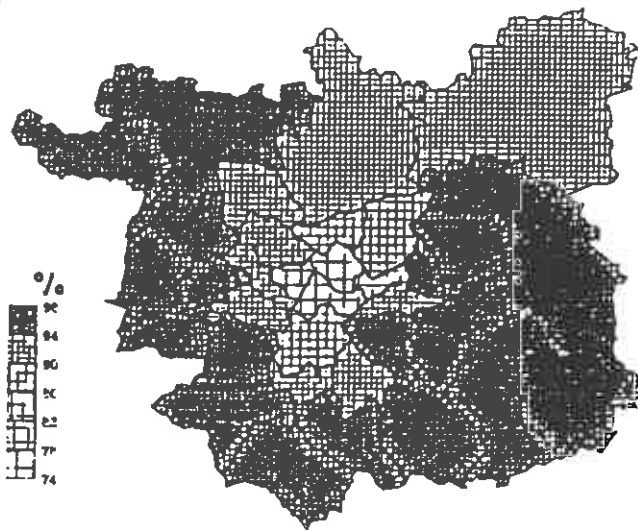
SCOTLAND
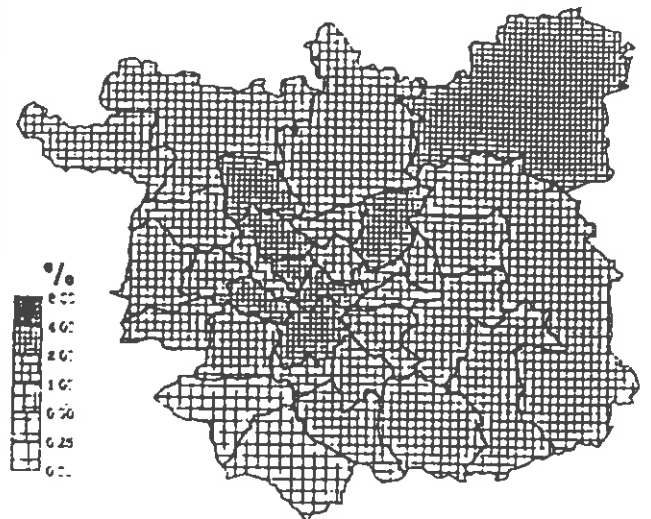
OLD COMMONWEALTH
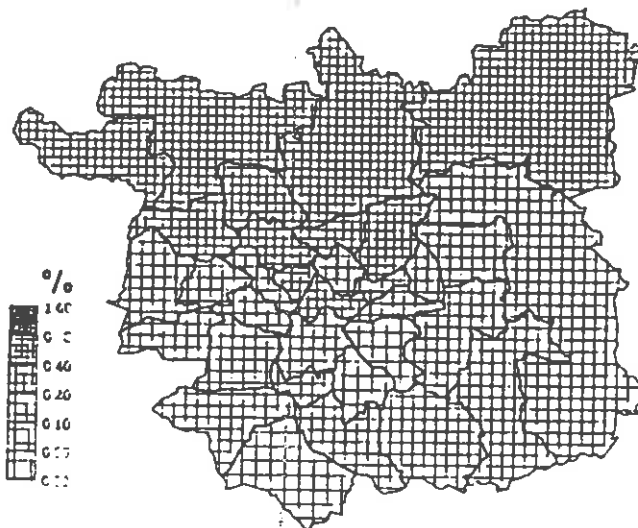
REST OF EEC
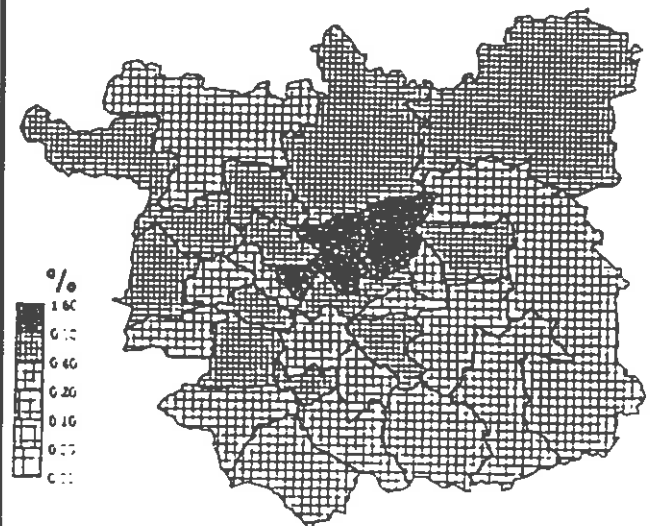
WALES
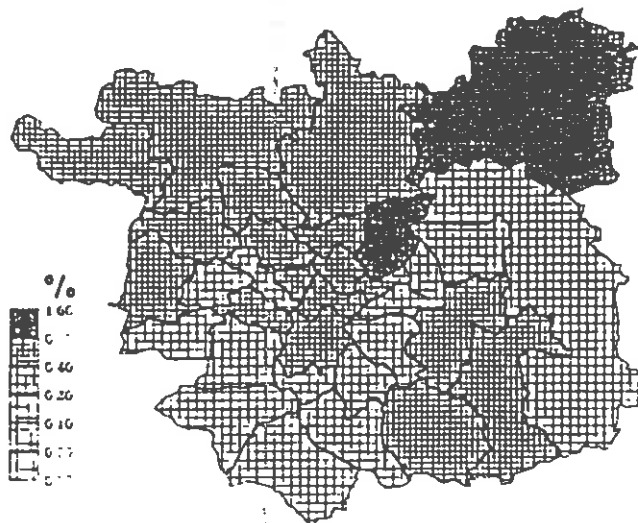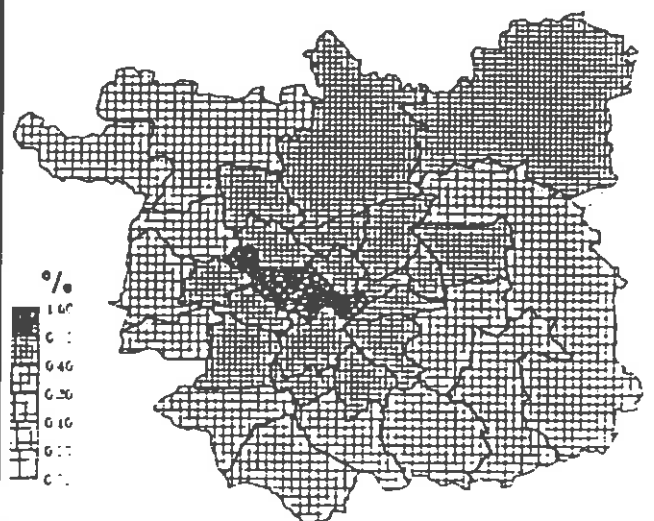
REST OF UK

Figure 19.   Tenal percentage distribution of ethnic groups in Leeds, 1981.
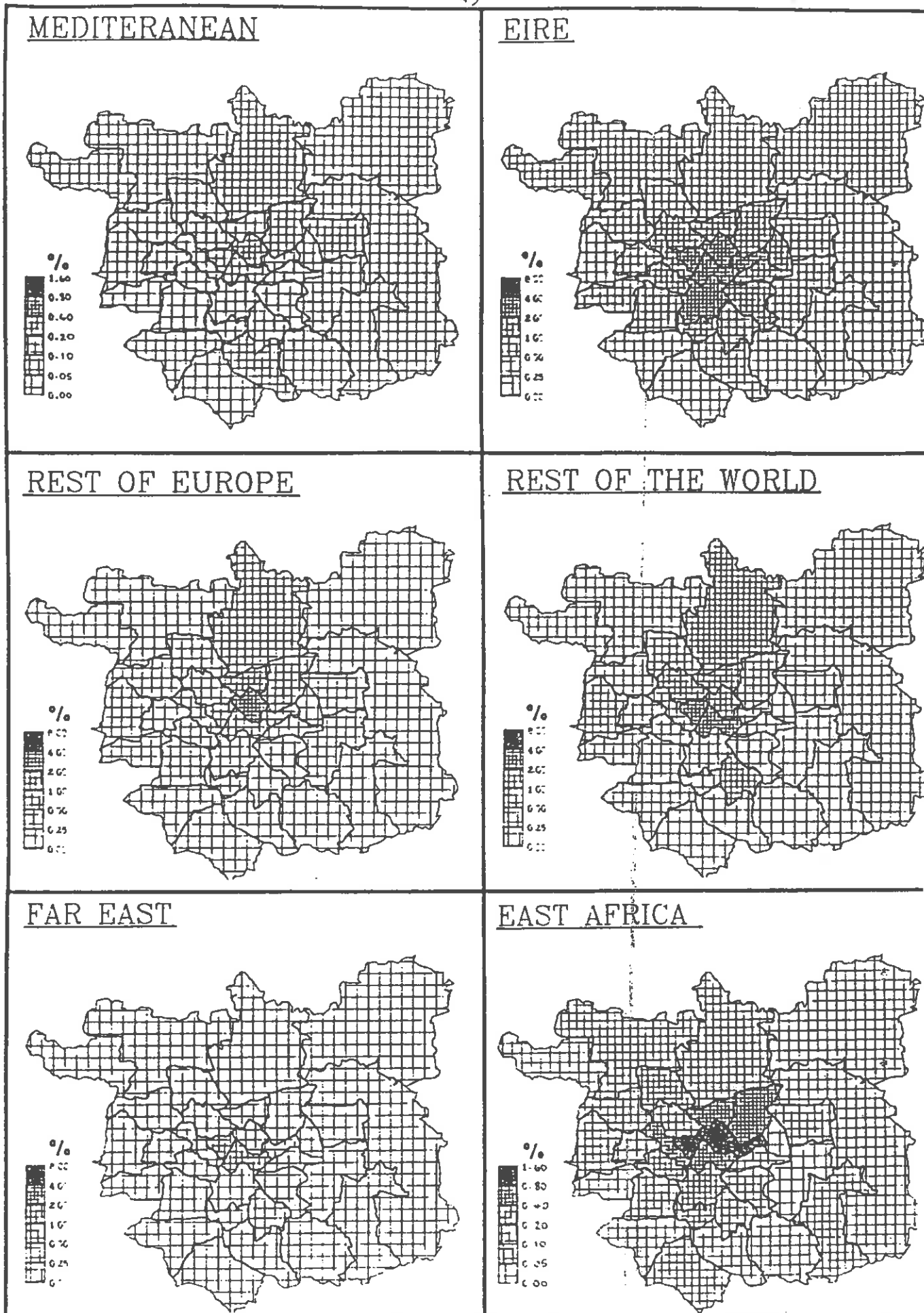
Figure 20. Zonal percentage distribution of ethnic groups in Leeds, 1981:
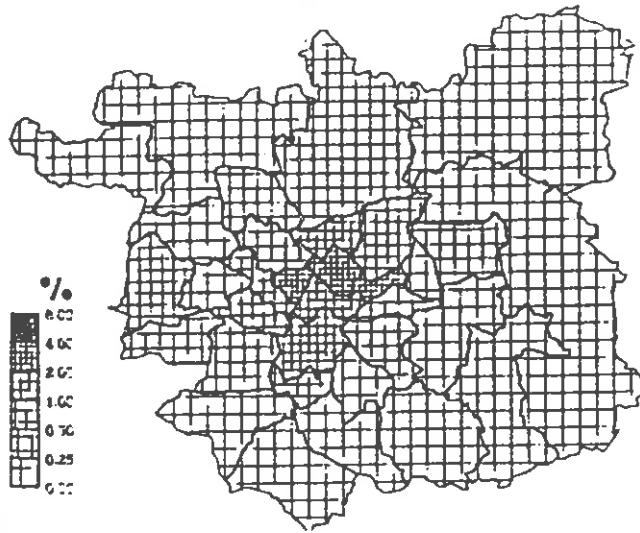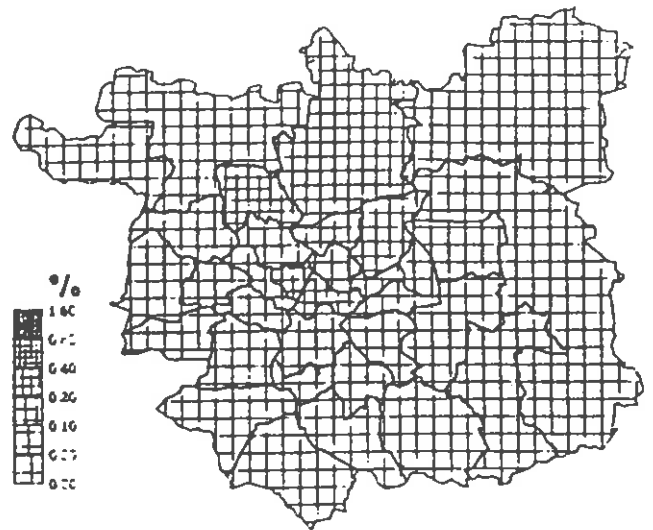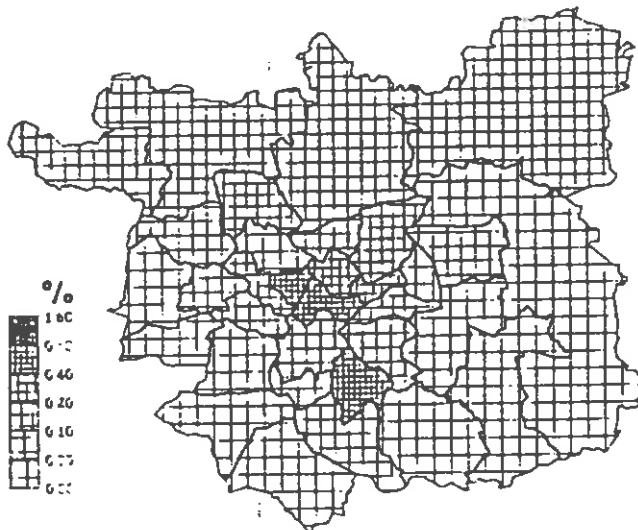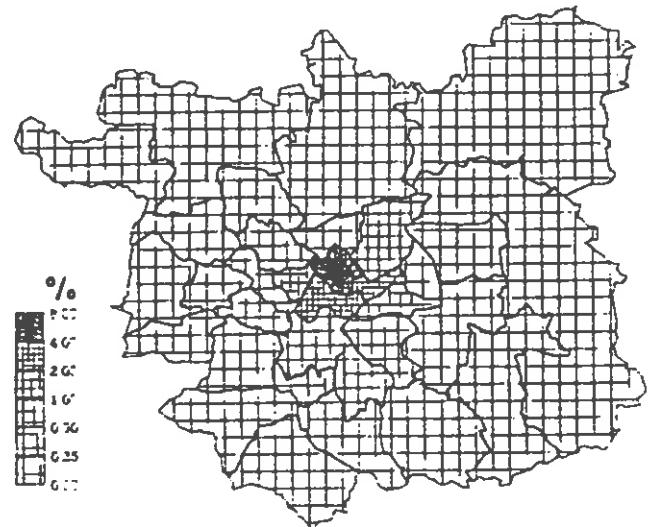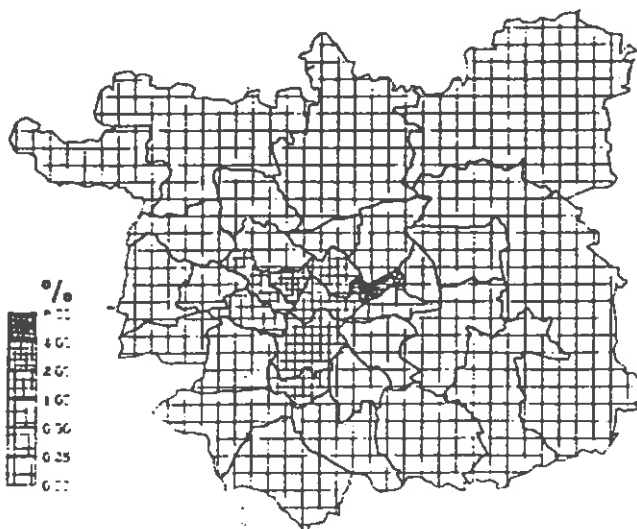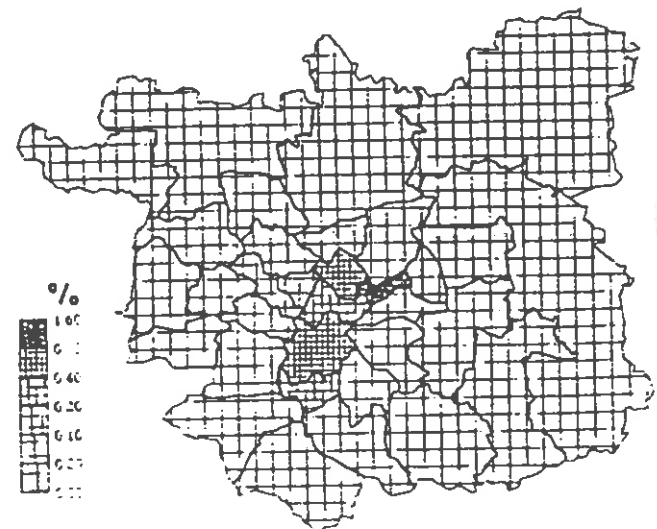
Figure 21. Zonal percentage distributions of ethnic groups in Leeds, 1981:
... of Africa born (Fig. III): Caribbean

Eire, Rest of Europe, Rest of the World, Far East, India, Caribbean and
Pakistan) and one for the smaller or less concentrated groups.

6.3.3 Location quotient maps

Another aspect of the concentration in a ward or absence from a ward of
group is the degree to which it is more concentrated in that ward than
is the population as a whole.   This can be measured by the location
quotient

$$LQ_e^i = \frac{P_e^i/P_e^*}{P_*^i/P_*^*}$$

where P_*^i is the total population of zone i and P_*^* is the total population
of all zones.   A location quotient for a ward above 1 indicates that a
group is more concentrated there than is the whole population.

Figures 22 through 24 depict these degrees of concentration.
Compare, for example, the Welsh born map with that for persons born
in Bangladesh.   The Welsh born are most highly concentrated in
Wetherby, in the other north suburban wards but also in Headingley and
University (the most affluent parts of the city).   However, in com-
parison with the Bangladesh born, the Welsh group are not highly
concentrated.   The Bangladesh born show location quotients of 10.8,
8.5, 4.6 and 2.97 for Harehills, City and Holbeck, Chapel Allerton and
Beeston wards (perhaps the least affluent parts of the city) res-
pectively whereas the Welsh born location quotient for Wetherby is
only 2.68.

Clearly, a great deal more could be said about the spatial
distribution of each individual group.   The maps, together with a good
local knowledge, are good means of generating hypotheses about the
residential location of the groups concerned: the testing of those
hypotheses requires careful work with other census-derived indicators
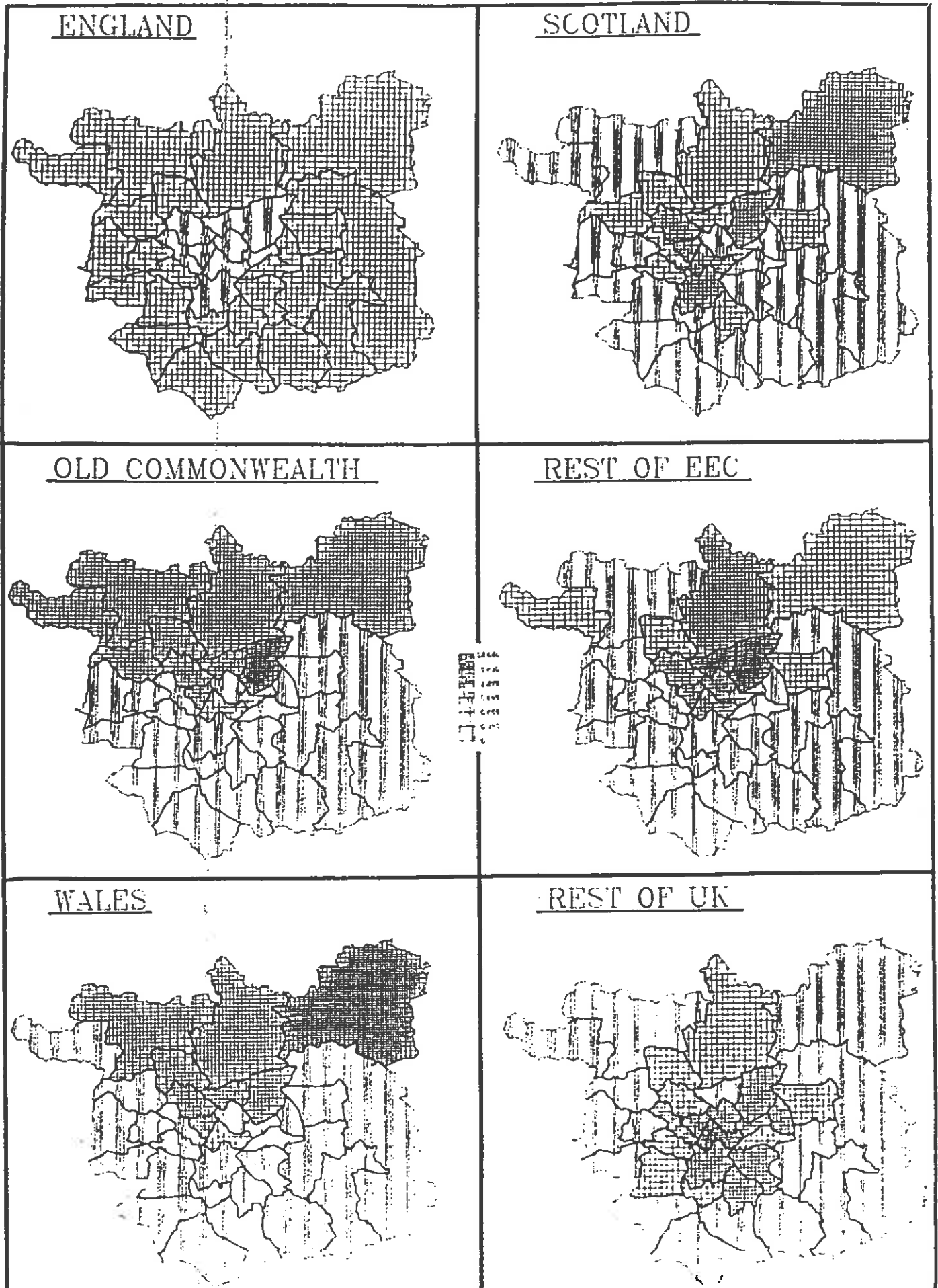and additional variables.

Figure 22.   Degrees of concentration (location quotients) for ethnic groups in Leeds, 1981:   1) England born; Scotland to Rest of UK
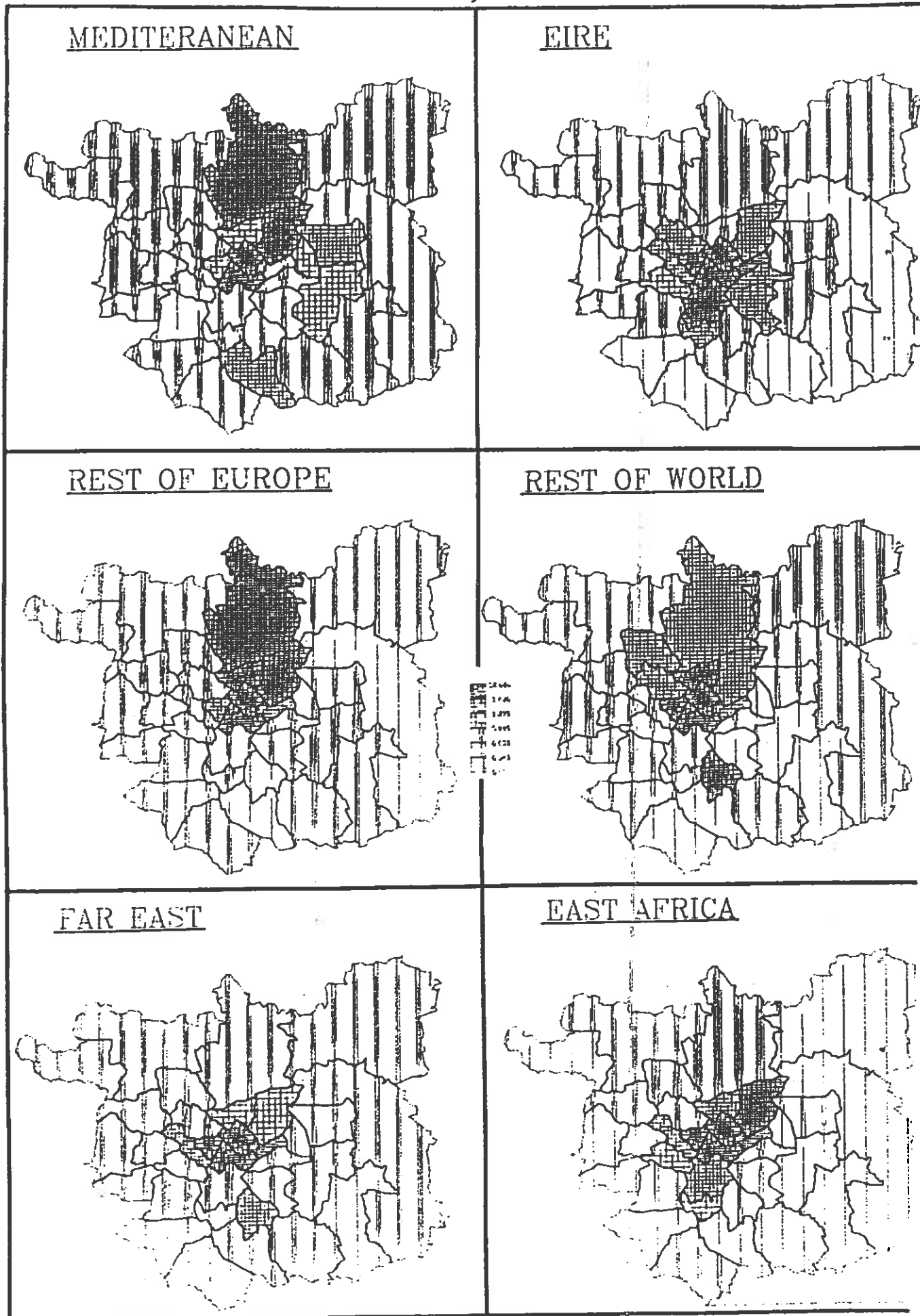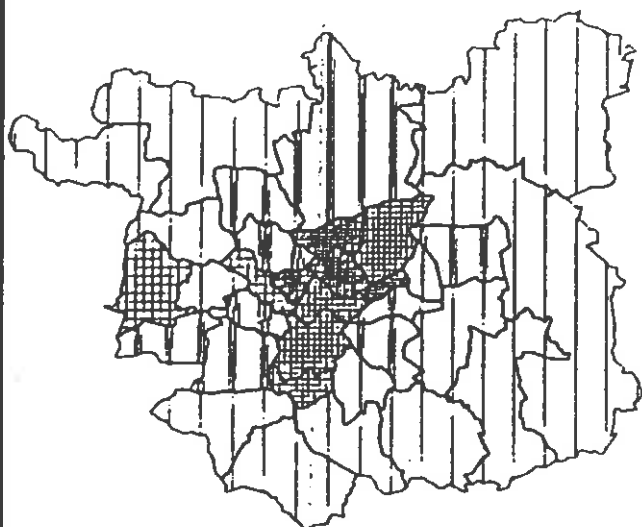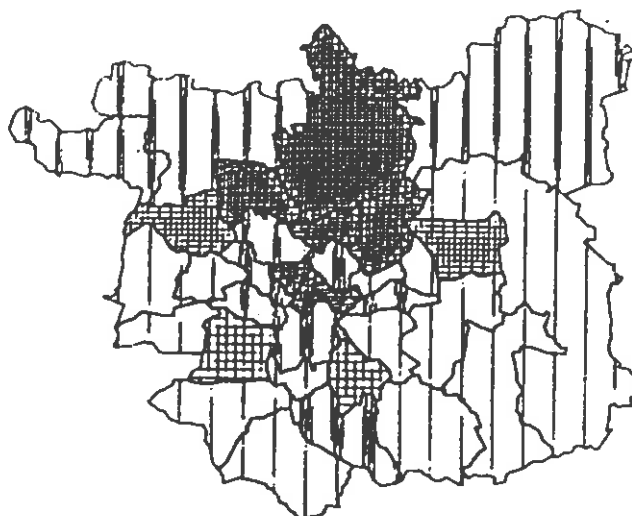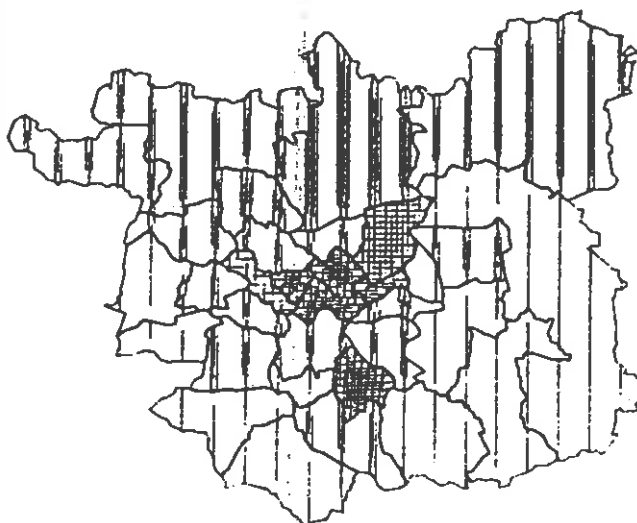
Figure 23. Degrees of concentration (location quotients) of ethnic groups in Leeds, 1981: (2) Mediterranean to Far East born

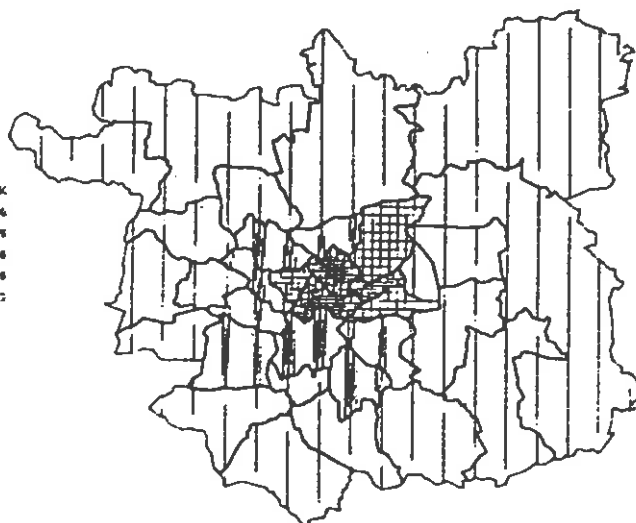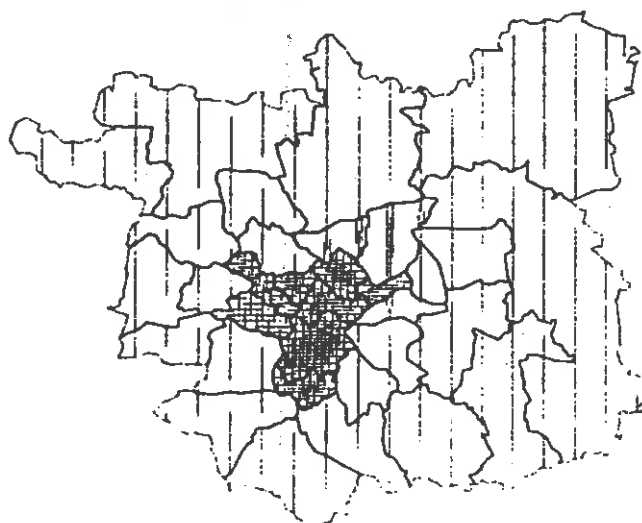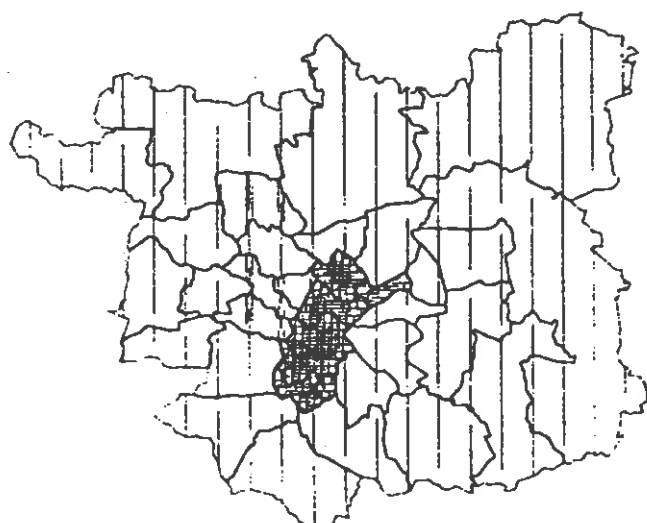Figure 24. Degrees of concentration (location quotients) of ethnic groups
in Leeds, 1981: (3) India to Rest of Africa born (Family III);
Caribbean to Bangladesh born (Family IV)

6.4   Interrelationships between the ethnic group distributions

We have already commented in passing on the sameness or contrast between the maps of particular pairs of ethnic groups.   Those maps can be compared in one respect by computing indices of dissimilarity between ethnic groups.   If we look at the interrelationships between the 17 country of birth groups (besides the English born) we can compute $(17 \times (17 - 1))/2$ or 136 pairwise indices of dissimilarity.

These are arranged in blocks in Table 6 in which groups of each family are compared with each family of groups.   The 'British' groups of Family I are very similar in their distribution (Table 6.1) particularly the first four members of the group, whose indices of dissimilarity are lower with fellow group members than with the English born.   Those born in the Rest of the UK (mainly Northern Ireland) are a little more segregated from other Family I groups.   The two lowest indices for this group are with the Scots born (14) and the Irish born (15)!

Other intra-family dissimilarity indices are higher than those of Family I except for the index value (19) comparing those Born in the New Commonwealth, Mediterranean and those born in the Rest of Europe, and that (value = 12) comparing those born in the New Commonwealth, East Africa and those born in India.   Comparisons between families of ethnic groups (Table 6.11) tend on average to follow the ordering of the groups by degree of spatial assimilation, though there are many exceptions in detail that show that groups born in different geographical areas but linked by cultural tradition or family ties have lower indices than the general ordering would suggest.

6.5   A composite picture of the ethnic geography of Leeds, 1981

So far we have examined the ethnic geography of Leeds from a group point of view.   The IOFD program provides, as we have shown, statistics, such as a matrix of inter-ward dissimilarity indices, which are the mirror image of the group statistics we have already discussed.   What are the types of ethnic areas that characterise the city of Leeds? What are the corresponding statements that can be made from a zonal (ward) point of view?

Table 6.   Indices of dissimilarity between pairs of ethnic groups

6.1   Family I vs Family I

BORN IN:	Scotland	Old Common-wealth	Rest of EEC	Wales	Rest of UK
Scotland	O	12	12	13	14
Old Commonwealth		O	10	8	23
Rest of EEC			O	12	17
Wales				O	23
Rest of UK					O

6.2   Family I vs Family II

BORN IN:	NC Medi-terranean	Irish Republic	Rest of Europe	Rest of World	NC for East
Scotland	21	23	29	28	35
Old Commonwealth	19	31	27	27	39
Rest of EEC	16	25	22	22	34
Wales	21	31	29	27	39
Rest of UK	22	15	28	26	30

6.3   Family I vs Family III

BORN IN:	NC East Africa	NC India	Rest of NC	NC Rest of Africa
Scotland	37	39	40	41
Old Commonwealth	39	42	40	46
Rest of EEC	34	37	37	40
Wales	41	43	41	46
Rest of UK	32	36	39	35

6.4   Family I vs Family IV

BORN IN	NC Caribbean	Pakistan	NC Bangladesh
Scotland	59	57	72
Old Commonwealth	61	63	76
Rest of EEC	56	59	73
Wales	61	64	77
Rest of UK	53	52	68

6.5 Family II vs Family II

BORN IN:	NC Mediterranean	Irish Republic	Rest of Europe	Rest of World	NC Far East
NC Mediterranean	0	30	19	25	35
Irish Republic		0	31	29	28
Rest of Europe			0	23	34
Rest of World				0	25
NC Far East					0

6.6 Family II vs Family III

BORN IN:	NC East Africa	NC India	Rest of NC	NC Rest of Africa
NC Mediterranean	33	37	38	37
Irish Republic	24	28	43	32
Rest of Europe	28	25	41	37
Rest of World	29	31	30	24
NC Far East	28	31	36	20

6.7 Family II vs Family IV

BORN IN:	NC Caribbean	Pakistan	NC Bangladesh
NC Mediterranean	52	55	67
Irish Republic	46	40	59
Rest of Europe	43	52	61
Rest of World	49	50	69
NC Far East	47	44	68

6.8 Family III vs Family III

BORN IN	NC East Africa	NC India	Rest of NC	NC Rest of Africa
NC East Africa	0	12	42	27
NC India		0	48	32
Rest of NC			0	39
Rest of Africa				0

6.9 Family III vs Family IV

BORN IN:	NC Caribbean	Pakistan	NC Bangladesh
NC East Africa	38	25	51
NC India	35	29	48
Rest of NC	59	61	80
NC Rest of Africa	42	41	64

6.10 <u>Family IV vs Family IV</u>

	NC Caribbean	Pakistan	NC Bangladesh
NC Caribbean	0	41	53
Pakistan		0	34
NC Bangladesh			0

6.11 <u>Average inter-group indices of dissimilarity</u>

Families of ethnic groups	Family I	Family II	Family III	Family IV
English born	17	33	37	68
I 'British'	14	27	39	63
II 'European'		28	32	53
III 'Indian & African'			33	48
IV 'West Indian & Pakistan				43

Computed from indices in Table 6.1-6.10

We could carry out a sophisticated cluster and region building analysis to answer these questions, but, for present purposes a simpler analysis suffices. First, we list in Table 7 some of the key ward characteristics. The first two columns provide a measure of the ethnic diversity of a ward by recording the number of groups with location quotients above 1 and above 2. The University ward records a location quotient above 1 for 17 out of 18 groups - only the English born are underpresented there, and 11 of those 17 groups have location quotients above 2. Neighbouring Headingley ward shows similar characteristics, again associated with having a University-Polytechnic-Colleges of Higher Education linked resident population. Almost as ethnically diverse are the wards of Harehills and Chapel Allerton. At the other end of the spectrum are the intensely 'English' wards of Barwick and Kippax, Bramley, Garforth and Swillington, Morley North, Morley South, Pudsey South and Rothwell in which the only over-represented group is the English born.

If we look at the groups (columns (3) to (5) in Table 7) with the highest percentages in a ward (apart from the English born who are in a 75% or over majority in every ward), then the list is dominated by the numerically largest groups of Scots and Irish born, displaced from first place only in Chapel Allerton by the Caribbean born, in Harehills by the Pakistani born and in Headingley by the Indian born. The picture of groups with the highest concentration in particular wards (columns (6) to (8) in Table 7) is much more diverse and the first column includes 17 of the 18 groups.

A convenient summary of the ethnic character of the different parts of the city is provided if we compute the tripartite percentage distribution of each ward's population into:

(A) English born

(B) Born in Family I or II (Scotland, Old Commonwealth, Rest of EEC, Wales, Rest of UK, NC Mediterranean, Eire, Rest of Europe, Rest of World, NC Far East)

(C) Born in Family III or IV (NC East Africa, NC India, Rest of NC, NC Rest of Africa, NC Caribbean, Pakistan, NC Bangladesh)

Table 7.   Selected ethnic characteristics of Leeds wards, Census 1981

Ward	No. of groups with LQs above >1 (1)	>2 (2)	Group with highest % — First (3)	Second (4)	Third (5)	Groups with highest LQ >1 — Highest LQ (6)	Next highest LQ (7)	Third highest LQ (8)
1. Aireborough	3	0	Scots	Irish	Rest of EEC	Old Comm.	Rest of EEC	English
2. Armley	3	0	Scots	Irish	Pakistani	Pakistani	Scots	Irish
3. Barwick & Kippax	1	0	Scots	Welsh	Rest of EEC	English	-	-
4. Beeston	7	2	Irish	Scots	Pakistani	Bangladesh	Pakistani	Irish
5. Bramley	1	0	Scots	Irish	Rest of UK	English	-	-
6. Burmantofts	4	0	Irish	Scots	Rest of UK	Irish	Rest of UK	Caribbean
7. Chapel Allerton	13	10	Caribbean	Indian	Rest of Europe	Caribbean	Rest of Europe	Indian
8. City & Holbeck	7	2	Scots	Irish	Indian	Bangladeshi	Pakistani	Irish
9. Cookridge	8	1	Scots	Rest of World	Irish	Rest of NC	Welsh	Old Comm.
10. Garforth & Swillington	1	0	Scots	Welsh	Irish	English	-	-
11. Halton	2	0	Scots	Irish	Welsh	Mediterranean	English	-
12. Harehills	14	7	Pakistani	Irish	Indian	Bangladeshi	Pakistani	Caribbean
13. Headingley	16	8	Indian	Pakistani	Irish	NC Far East	NC Rest Africa	NC East Africa
14. Horsforth	3	0	Scots	Irish	Welsh	Rest of NC	Old Comm.	English
15. Hunslet	6	2	Scots	Rest of World	Irish	NC Rest Africa	Rest of World	Rest of UK
16. Kirkstall	7	1	Irish	Scots	Rest of UK	Pakistani	Irish	NC East Africa
17. Middleton	2	0	Scots	Irish	Rest of UK	Mediterranean	English	Irish
18. Moortown	10	4	Scots	Rest of Europe	Indian	Rest of Europe	Rest of NC	Rest of World
19. Morley North	1	0	Scots	Irish	Welsh	English	-	-
20. Morley South	1	0	Scots	Irish	Welsh	English	-	-
21. North	9	2	Scots	Rest of World	Rest of Europe	Rest of NC	Mediterranean	Rest of Europe
22. Otley & Wharfedale	3	0	Scots	Welsh	Irish	Old Comm.	Welsh	English
23. Pudsey North	2	0	Scots	Indian	Rest of World	Indian	English	-
24. Pudsey South	1	0	Scots	Irish	Welsh	English	-	-
25. Richmond Hill	3	0	Irish	Scots	Rest of UK	Irish	Rest of UK	English
26. Rothwell	1	0	Scots	Welsh	Irish	English	-	-
27. Roundhay	14	4	Scots	Irish	Rest of World	Mediterranean	Old Comm.	Rest of Europe
28. Seacroft	3	0	Scots	Irish	Rest of UK	Rest of UK	Mediterranean	English
29. University	17	11	Irish	Scots	Caribbean	NC Rest Africa	Rest of NC	Caribbean
30. Weetwood	7	0	Scots	Irish	Rest of World	Rest of World	Old Comm.	Welsh
31. Wetherby	5	1	Scots	Welsh	Irish	Welsh	Scots	Old Comm.
32. Whinmoor	5	0	Scots	Irish	Rest of EEC	Rest of NC	Mediterranean	Rest of EEC
33. Wortley	3	0	Scots	Irish	Rest of UK	Rest of UK	Rest of NC	English

and plot the position of the wards on triangular coordinate paper
(Figure 25). Three clusters can be distinguished on the scatter plot:

Cluster   I - 'native areas' with more than 94% of their
                population born in England, less than 5%
                born in ethnic families I or II and less
                than 2% born in ethnic families III or IV;

Cluster II - 'moderately ethnic areas' with between 89
                and 94% of their population born in England,
                Between 5 and 10% of their population born
                ethnic families I or II and up to 5% born
                in ethnic families III or IV;

Cluster III - 'intensely ehtnic areas' with between 75%
                and 82% of their population born in England,
                between 9 and 13% born in ethnic families I
                and II and between 6 and 15% born in ethnic
                families III and IV.

The inset map shows the spatial location of these ward clusters.
Cluster I wards cover the outer suburban and small town ring around
the periphery of Leeds metropolitan district. Cluster II wards
occupy parts of the inner city, suburbs of the city immediately
beyond the inner city plus the outer suburbs on the north-northeastern
fringe. Cluster III wards occupy the inner city north of the Central
Business district.

6.6   Caveats and further analysis

Our analysis of the ethnic geography of Leeds in 1981 has shown
how rich a source of information the decennial census is. The findings have
served to underline the value of easy access to that data on disk,
the value of writing specific software designed for the subject of
interest, and the value of using a well written general package.

However, there is scope for improvement of the analysis in
several ways.

Figure 22. A scatterplot of Leeds wards in 'ethnic space'.

PERCENTAGE BORN IN ETHNIC FAMILIES I & II

PERCENTAGE BORN IN ETHNIC FAMILIES III & IV

PERCENTAGE BORN IN ENGLAND

CLUSTER I "NATIVE AREAS"

CLUSTER II "MODERATELY ETHNIC AREAS"

CLUSTER III "INTENSELY ETHNIC AREAS"

CLUSTER I
CLUSTER II
CLUSTER III

FAMILIES OF ETHNIC GROUPS

I   SCOTLAND, OLD COMMONWEALTH, REST OF EEC, WALES, REST OF UK

II MEDITERRANEAN, EIRE, REST OF EUROPE, REST OF WORLD, FAR EAST

III EAST AFRICA, INDIA, REST OF NEW COMMONWEALTH, REST OF AFRICA

IV CARIBBEAN, PAKISTAN, BANGLADESH

WARDS

AIREBOROUGH
ARMLEY
BARWICK & KIPPAX
BEESTON
BRAMLEY
BURMANTOFTS
CHAPEL ALLERTON
CITY & HOLBECK
COOKRIDGE
GARFORTH & SWILLINGTON
HALTON
HAREHILLS
HEADINGLEY
HORSFORTH
HUNSLET
KIRKSTALL
MIDDLETON
MOORTOWN
MORLEY NORTH
MORLEY SOUTH
NORTH
OTLEY & WHARFEDALE
PUDSEY NORTH
PUDSEY SOUTH
RICHMOND HILL
ROTHWELL
ROUNDHAY
SEACROFT
UNIVERSITY
WEETWOOD
WETHERBY
WHINMOOR
WORTLEY

6.6.1 The definition of ethnicity

The country of birth place statistics undoubtedly underestimate the size of the non-English ethnic groups. Further SAS tables, County report tables and tables from the General Household Survey and Labour Force Survey can be used to improve estimates of the distribution of the population among ethnic groups. What is needed is a model for estimating from available statistics the numbers in more sharply defined ethnic groups. Such a model would have to deal with the conceptual difficulties of mixed birth for individuals and of mixed ethnicity for households.

6.6.2 The definition of spatial segregation

The index of dissimilarity which we have used to summarise the spatial pattern or interrelationships of groups needs improvement. Further indices need to be used. Lieberson (1981) has suggested re-use of an index that reflects the asymmetry of segregation between a larger and a smaller group. It would also be interesting to use our knowledge of the spatial arrangement of wards to compute a set of spatial indices such as the mean inter-group and intra-group distances: this would provide a very interpretable measure of spatial separation.

6.6.3 The scale problem

And, of course, there is the scale problem. If we measured the patterns at enumeration district scale or in terms of clusters of enumeration districts the patterns identified at ward scale are likely to be much sharper. Levels of segregation are likely to be higher (cf. Poole and Boal, 1973; Rees, 1979, p. 299-302). Inter-ethnic distance measures are also likely to be affected, but perhaps to a lesser extent.

6.6.4 Evolution over time

In order to understand the processes that have produced the spatial patterns of 1981 an analysis of their evolution over time is necessary. Severe problems of change in ward and enumeration district definition need to be overcome. It will also be necessary to build up a picture of the demographic changes experienced by the various ethnic groups. The work of King (1974, 1976) will be an invaluable guide here. It should

also be possible to follow and improve on King's work in estimating the socioeconomic characteristics of ethnic groups, and the housing they live in.

6.6.5   Models of ethnic residential location

Given understanding of the spatial and social patterns and their evolution over time, it should then be possible to test rigorously a set of hypotheses concerning the determinants of ethnic residential patterns, either through cross-sectional analysis or through simulation modelling.   A major improvement on these previous studies would be the use of a much wider variety of ethnic spatial histories and the use of a more general microsimulation methodology.

6.6.6   Further issues

The analysis should conclude with a careful evaluation of the costs and benefits of the various residential patterns to group members and to society as a whole (cf. Boal, 1976), and an assessment of the influence of public policy in housing and in employment on ethnic residential location.

7.   CONCLUSIONS

We have in this paper described an information system.   An information system consists of data, a storage device, a description, an accessing system and an analysis system.   Traditionally, data have been stored in printed books in libraries, described in the contents and introductory pages of those books; the accessing system has been the way in which people may join libraries and borrow books and the analysis system has been reading and making notes on the information contained in the books.

Using data stored in a library is an elementary activity learnt early in life.   We have attempted to make the information system for 'the 1981 Census Small Area Statistics for West Yorkshire Wards' almost as easy to use.   The data are numbers contained in machine readable records stored on a computer disk (GEC6LIB's B disk); the data are described in a simple coding book (OPCS, 1982c).   Just as you have to gain permission to use a library so you have to acquire permission to use the computer (your username), and the accessing system is the easy-to-use TABULATE program.   The analysis system we have left, as does the library, in the hands of the reader of the data, but have shown how simple spatial statistics may be computed using the IOFD program and how some of those statistics may be mapped using the GIMMS package.

We hope that the information system we have designed minimises the routine tasks in getting access to part of the 1981 Census SAS so as much time as possible can be spent in the task of thought and innovation in its analysis.   We trust that in the near future the same will apply to the national information system for accessing enumeration district data at regional computing centres.

REFERENCES

OPCS = Office of Population Censuses and Surveys.

Birkin, M.H. (1982) Digitizing maps on the GTCO digitizer and Apple II system for use with the GIMMS package. Computer Manual, forthcoming, School of Geography, University of Leeds.

Boal, F.W. (1976) Ethnic residential segregation. In Herbert, D.T. and Johnston, R.J. (eds.) *Social areas in cities, Volume 1. Spatial processes and form.* Wiley, London.

Denham, C.J. (1980) The geography of the Census 1971 and 1981. *Population Trends*, 19, 6-12.

King, J.R. (1974) Immigrant fertility trends and population growth in Leeds. *Environment and Planning*, 6, 509-46.

King, J.R. (1976) Immigrants in Leeds: an investigation into their socioeconomic characteristics, spatial distribution, fertility trends and population growth. PhD dissertation, School of Geography, University of Leeds.

Lieberson, S. (1981) An asymmetrical approach to segregation. In Read, C., Robinson, V. and Smith, S. (ed.) *Ethnic segregation in cities.* Croom Helm, London.

Morgan, C. and Denham, C.J. (1982) Census small area statistics (SAS): measuring change and spatial variation. *Population Trends*, 28, 12-17.

OPCS (1980) *Census 1981. OPCS Monitor. CEN 80/8.* OPCS, London.

OPCS (1981a) *Census 1981. Preliminary report. England and Wales.* HMSO, London. CEN 81 PR(1).

OPCS (1981b) *Census 1981. Preliminary report for towns, urban and rural population. England and Wales.* HMSO, London. CEN 81 PR(2).

OPCS (1981c) *Census 1981. OPCS Monitor. CEN 81/85.* OPCS, London.

OPCS (1982a) *Census 1981. OPCS County Monitors.* OPCS, London. CEN 81 CM1-46.

OPCS (1982b) *Census 1981. County Reports. Part 1.* HMSO, London.

OPCS (1982c) Census 1981 Small Area Statistics. Tape specifications. 100% data. OPCS, Titchfield, Hants.

OPCS (1982d) *Census 1981. Definitions. Great Britain. CEN 81 DEF.* HMSO, London.

Poole, M.A. and Boal, F.W. (1973) Religious residential segregation in Belfast in mid-1969: a multi-level analysis. In Clark, B.D. and Gleave, M.B. (ed.) *Social patterns in cities.* IBG Special Publication No. 5.

Rees, P.H. (1979) *Residential patterns in American cities: 1960.* Research Paper 189, Department of Geography, University of Chicago, Chicago.

Rees, P.H., Brookes, S.M. and Sellers, P.J. (1981) A manual for using *IOD*, an index of dissimilarity program. Second edition. Computer Manual 5, School of Geography, University of Leeds.

Waugh, T.C. (1978) *GIMMS.* Second Edition, Inter-University Research Council Series, Report No. 30. Program Library Unit, University of Edinburgh.