Dorling, D. (1997), European micro-data availability: the special case of Britain, Chapter 6 in Geographic Information Research: Bridging the Atlantic, M.Craglia and H.Coucelis (eds), London: Taylor and Francis., 157-202.

ESF-GISDATA and NSF-NCGIA 1995 Summer Institute in Geographic Information

European Micro Spatial Data Availability: Mapping the Social Structure of Britain

First Draft - May 1995

Daniel Dorling
British Academy Fellow
Department of Geography
University of Newcastle upon Tyne
England, NE1 7RU

Tel: 41 (0) 91 222 6454
email Danny.Dorling @ UK.AC.NEWCASTLE

Confession

This paper is still unfinished. In particular, it requires a section to be written on the availability of European micro spatial data to researchers outside Britain. One excuse for this omission is that a new centre to promote the exploitation of key European datasets is being opened by the Universities of Durham and Essex in June. The final paper should include details of availability from that source and others. The paper currently relies on some black and white figures which were first printed for inclusion in a cartography bulletin (Dorling 1994). I would prefer to replace these by colour figures but am unsure if these are permitted. It could also benefit from including a short literature review — but how much space is allowed? Any comments and suggestions for changes to this draft would be very welcome!

Abstract

This paper presents the final results of a three year study sponsored by the British Academy to collect and analyse the most detailed spatial data on British human geography available, using new geographical information system techniques. The paper explores ways in which this data might be made more generally available to researchers outside Britain and discusses how the geographical information techniques developed to deal with this information could be used in a wider context.

This work was developed from the the author's PhD thesis on visualizing the social, economic and political spatial structure of British geography. The author and his supervisor explored new methods of data compression, animation, smoothing, presentation and transformation in work published originally two years ago (Dorling and Openshaw 1992). This methodological work has since been extended and is now being applied to many large data sets most of which have not before been analysed at these spatial scales (Dorling 1993).

British spatial data presents a special case for a number of reasons, all connected with the pedigree of this information. First, it tends to be extremely complex in comparison with many other countries' data. The areal-units have evolved differently in different parts of the country over time and boundary changes are common every 3 to 15 years. Britain was one of the first countries in the world to employ a Boundary Commission to change the boundaries of administrative units - which it appears to do as frequently as is possible. Comparison over time is hence problematic. Second, the British government is uniquely secretive about spatial data, having laws governing its use comparable with those of many military dictatorships. Third there exists a huge wealth of official data which dates back further, and in more detail, than that of almost any other country.

Many of the problems of spatial complexity and of the shear volume of information can be solved using Geographical Information Systems to create valuable datasets for research. The new British datasets which have been drawn together and made spatially comparable for the first time by the author include the full British Censuses of 1971, 1981 and 1991 at ward level (10,000 spatial units by 15,000 variables); selected variables from earlier censuses by the highest resolution areal unit available, which can then be compared over time with the same variables for later years; and the spatially coded results of all General Elections since 1955. The dataset also includes mortality statistics recorded by postcode in the 1980s and made comparable with local authority series dating back to the 1930s; school exam results, house sales, information on wealth and on changing patterns of migration and commuting (at the full ward to ward level in both 1981 and 1991).

Visualization techniques have been developed to analyse this information which is too vast to be studied by traditional statistical techniques. A *New Social Atlas of Britain* has been produced from these datasets which contains over two hundred pages of novel colour maps (Dorling 1995). Most of the maps use projections which distort space to highlight urban areas across the nation. The aim of this work has been to show how localities have changed over time, drawing patterns of social inclusion and exclusion across the many thousands of localities in Britain. In total over four million individual variables are presented in this atlas. Prints from the atlas will be used to illustrate the extent and value of this geographical information.

The paper will then argue that conventional quantitative statistical techniques mask the complexity of the spatial patterns in British society which computer visualization has been able to illustrate and clarify. A persistent and particularly interesting problem has been that of how to visualize a spatial pattern evolving over time. Illustrations of various solutions used in the New Social Atlas of Britain are given in this paper as well as alternative techniques and some of the lessons which are learnt through trying to visualize so much of the social data which is available for a country, nationally, across so many localities. The author is the British representative on the Cartographic Visualization, and Cartography of Time, working groups of two International Cartographic Association commissions in which a great many of these ideas have been discussed.

The lessons learnt from this work could be more widely applied to data from other countries and may challenge the assumptions that certain comparisons are not possible or that a given level of spatial resolution is acceptable. Geographical Information Systems are still rarely used to study data about social geography and when there used for this purpose elementary mistakes are often made because the operators have been trained (and the systems are geared) to deal with information about the physical geography of places rather than their human geography. This situation is counter intuitive because more money and effort is spent by government and private industry collecting spatial data about people than about any other subject. It is the market-place which has nurtured Geographical

Information Systems and yet these systems have not yet been thoughtfully applied to the problems that are of most interest to that market.

Finally the the paper will explore the extent to which the authorities in Britain could be persuaded to allow some detailed geographically coded information to be made available to researchers from other countries. Currently the legal situation is that if a researcher from, say, Eire or the United States is working in collaboration with a researcher from Britain on mapping unemployment rates across the island of Ireland, they have to travel to Britain to undertake the work - despite the instant computerised links available. This is true even thought information at the local community level, rather than at the individual or postcode level, is not commercially sensitive. International research into the problems of social geography is severely hindered by these restrictions, many of which only exist for political, military and practical reasons which have long ceased to be relevant.

Drawing a New Social Geography of Britain

In many parts of the world information about very small localities stored on computers has become available for research. This information often contains hundreds of statistics about people living in thousands of localities. This paper illustrates how a researcher can visualize these statistics to attempt to understand more of the localities of one country — Britain. An aim of this work is to show people facets of the society in which they are living which they would not otherwise recognise. The methods which are presented here could be applied to any part of the world for which the information is available, and the problems which are faced in dealing with this information are also universal.

The traditional means by which information about thousands of localities is presented is in equal-area choropleth map form. Unfortunately traditional maps distort this information by over-emphasising statistics which relate to rural localities and under-representing the characteristics of the majority of the population who live in towns and cities. This situation applies to Britain, but an even stronger case could be made in many other parts of the world where rural areas are more sparsely populated and urban areas are more densely populated.

Figure 1 shows a traditional map of the ten thousand localities in Britain for which most local statistics are available: local government wards. The figure also shows an equal population cartogram in which each of these ten thousand localities is represented by a circle (the area of which is proportional to the number of people who live there). County and region boundaries are also included in the figure so that these places can be identified. On the cartogram the statistics relating to each person are given equal visual weight. Thus the majority of space on the cartogram represents localities within the largest cities in Britain.

Cartograms can also be used effectively when only a few hundred areas are being mapped. A key advantage of using cartograms in these cases is that the cartograms can be reduced to a small size while the areas in which most people live remain visible. This means that many cartograms can be placed side by side, allowing temporal trends in geographical patterns to be compared. Figure 2 shows two cartograms of Britain based on local authority districts. The first cartogram shows the changing proportion of the population of each area born in the New Commonwealth in the 1970s, the second shows this trend in the 1980s. The first cartogram demonstrates how the mainly black immigration of that time was confined to the cities of London, the Midlands, Manchester and Yorkshire. The second cartogram reflects the mortality of the largely white older New Commonwealth born population. These

inferences cannot be made from the cartograms themselves but are possible if enough related information is examined. To be able to present the quantity of information required to draw parallels such as these, between different temporal trends, the methods of graphical presentation need to be compact as well as being fair to the population presented.

Making Area Proportional to Population

The principal of making area proportional to population can be extended from maps to diagrams of all kinds. This is often done intuitively. Figure 3 shows how the numbers of people immigrating to Britain from different countries changed each year from 1955. The vertical scale of the figure shows how many people were entering the country each year and the horizontal scale shows which year the information refers to. Thus the area of the band of colour representing each area of origin is proportional to the total number of people who immigrated from that area over the period shown. This diagram contains a great deal of information presented in a compact and just way. Unfortunately, to be reproduced here all the illustrations presented here have had to be converted into black and white, but even with this limitation it is possible to show a great deal of detail.

Traditional maps which draw places in proportion to land area can be useful, particularly when the population of interest lives in very remote areas. A good example in Britain is of people who can speak gaelic or welsh. Figure 4 presents maps and cartograms depicting their distribution in Scotland and Wales, showing also how that pattern has changed over the last decade. The map of gaelic speakers in Scotland in 1991 reflects how the proportion increased from less than one person in fifty speaking gaelic in the South East of that country to over half the population being able to speak it in the North West. However, the equivalent cartogram (in its key) shows that 94% of the population of Scotland lived in wards where less than 2% of people could speak gaelic. A similar, if less extreme pattern is seen in Wales. The cartogram of Wales shows the geography of change to be very different to that suggested by the map. Welsh speaking is not declining in most of Wales (by population) because most of the Welsh live in the Southern Valleys where the effect of teaching Welsh in schools has had much influence.

The next two figures illustrate how the principle of making the components of graphics about people proportional to the numbers of people can be extended. Figure 5 shows how many workers in each industry were in various occupational groups in 1991. This type of graph is often given as an option on computer packages but is rarely used when it would be appropriate. The graphs shown in figure 6, in contrast, cannot be created easily with standard packages. In the main graph the vertical scale shows the proportion of all workers in each industry while the horizontal scale shows the proportion of those workers who are male or female, full-time, part-time or self-employed. Thus the total area of the bars is in proportion to the size of the workforce. For each industrial group a population pyramid is also drawn, again with its area in proportion to the total number of workers in that sector. These graphs can be difficult to read and they are certainly not simple to label. However, again they contain a great deal of information and can be argued to present it fairly.

Complexity and Simplicity in Visualization

Occasionally very clear patterns are found in quite complex data. Often it is only after producing many graphics that these patterns are evident to the researcher. This is where the use of computers to

visualize social data is most advantageous. Figure 7 presents another pair of district level cartograms. The cartogram on the left shows the year in which unemployment was highest in Britain between 1979 and 1993 in each district. The North/South divide can be seen to be abrupt. The cartogram on the right shows what that highest level of unemployment was in each district. Here the urban/rural divide can be seen to be most acute. Low levels of maximum unemployment circle the districts which make up the major cities (in which over a fifth of the workforce has been unemployed at one time or another). To understand these cartograms a new geography of Britain has to be learnt. That appears to be harder for researchers who are most familiar with the traditional representation of this country. British cartographers, in particular, are often averse to these images, whereas people unfamiliar with Britain often expect to see London as the largest, rather than the smallest, place on the map!

When large quantities of data are being analysed standard types of graph often require modification. An example is the scatter-plot, in which overlapping dots are obscured. When there are many dots to plot, and many overlap, a misleading impression of the distribution can result. Figure 8 shows one method of dealing with this problem. In this figure a dot is plotted to show the relationship between the changes in house prices and unemployment in 459 districts each plotted ten times to show the situation in each of ten years. The dots are drawn with their areas proportional to the population affected. When two dots would overlap their populations are amalgamated and a single larger dot is drawn. The effect is reminiscent of the shading in old newsprint and the aim is similar, to darken certain parts of the paper more than others (in this case the parts which represent the experiences of more people). Thus, although a rough negative relationship can be discerned, it is apparent that most people in most years have experienced very little change in unemployment and have seen modestly rising housing prices. In the original version of this graphic the dots are coloured to show in which year these changes were strongest. That is not possible in black and white here.

It is possible to show the geographical distribution of average housing prices across ten thousand areas simultaneously. This is done in Figure 9 where the average prices which buyers paid for houses in each ward over the 1980s are shown at 1991 prices. The prices are calculated from the mortgage book of a major building society. Every sale is linked to a ward through the postcode of the address of the property and a weighted average of the sale prices is made with the weights allowing for inflation up to 1991. Given that in some wards the size of this sample is quite low, it is remarkable how even the pattern is. This reflects how rigid local housing markets tend to be. The huge differences between the centre and the suburbs of London, or the West and East sides of Birmingham, are immediately apparent (Figure 1 acts as a key to this figure). The pattern of low rates of unemployment in Figure 7 can be seen reflected in the high levels of average housing prices in much of the South East. Through presenting graph after map after table after figure, given enough patience and space, readers can form their own views as to what facets of society appear to be interrelated most strongly and how. This is preferable to asking them to accept the results of statistical tests which disguise the prejudices and assumptions of their authors more than do a series of picture, each concentrating on a relatively simple subject.

It is important that the form of graphics used in a social atlas varies if the reader is to remain alert. Occasionally there is merit in using "three-dimensional" charts. Figure 10 gives an example in which the almost exclusive rises in one person or in seven-plus room households in Britain are emphasised. It is when a diagram or topic appears dull that pretty graphics have their most valuable uses. Figure 11 shows an opposing case. Here the chances of people dying by two causes of death by sex and single year of age are displayed, and how those chances are changing. A very simple form of graphic can be used in this example as the contents are of greater interest to many readers and because quite a lot of

data is being portrayed which might only be confused by further embellishment.

Finally, figure 12 shows how a collage of cartograms can be used to depict a whole series of changes, in this case the results of all general election in Britain since 1955. These cartograms do not actually show who won in each constituency, but instead show which party came second. Seven thousand results are included in this one graphic. In summary the graphic shows how the Liberal party has risen from the ashes of it post-war low to be seriously contesting a majority of seats in the South of England by the end of this century. The rise of nationalism in Scotland and the exit of main stream parties from Northern Ireland are also clear messages from the figure (clear at least in its original colour form!). More importantly, by presenting this quantity of information the graphic can show to what extent these assertions are not universal. It is even possible to follow the fortunes of individual constituencies over time. A simple example is the Isle of Wight, the most Southern constituency on the cartograms. There, the second placed party has changed from Labour to Conservative to Liberal. These maps can also incorporate the effects of boundary changes. If, for instance, you examine the figure closely you can see that the number of constituency alters over time in each region.

How can the data be made more easily available?

.... still to be written .....

Conclusion

This paper has presented only twelve illustrations taken from a social atlas of Britain which contains over two hundred maps and two hundred graphs, most of which show patterns across ten thousand areas. Hopefully it has given a flavour of what can now be done given the wealth of information available and the ease with which that information can be manipulated. A home microcomputer was used to create all the maps and graphs shown here and to typeset the atlas. Apart from showing what is technically plausible this paper argues that care is needed if the maps and diagrams which social scientists produce are to represent society fairly. One guiding principle is that equal numbers of people are equally represented. Another principle is to present as much information as possible when as complex an object as society is being studied. What we need to see cannot be predetermined. Simplification conceals the complexity of reality.

References

Dorling D. and Openshaw S. (1992) Using computer animation to visualize spacetime patterns, *Environment and Planning B*, Vol.19, pp.639-650.

Dorling D. (1993) Map design for census mapping, *The Cartographic Journal*, Vol.30, No.2, pp.167-183.

Dorling D. (1994) Mapping and graphing the social structure of Britain, *Society of* Cartographers Bulletin, *Vol.28, No.1, pp.7-18.*

Dorling D. (1995) *A New Social Atlas of Britain*, London: John Wiley and Sons.

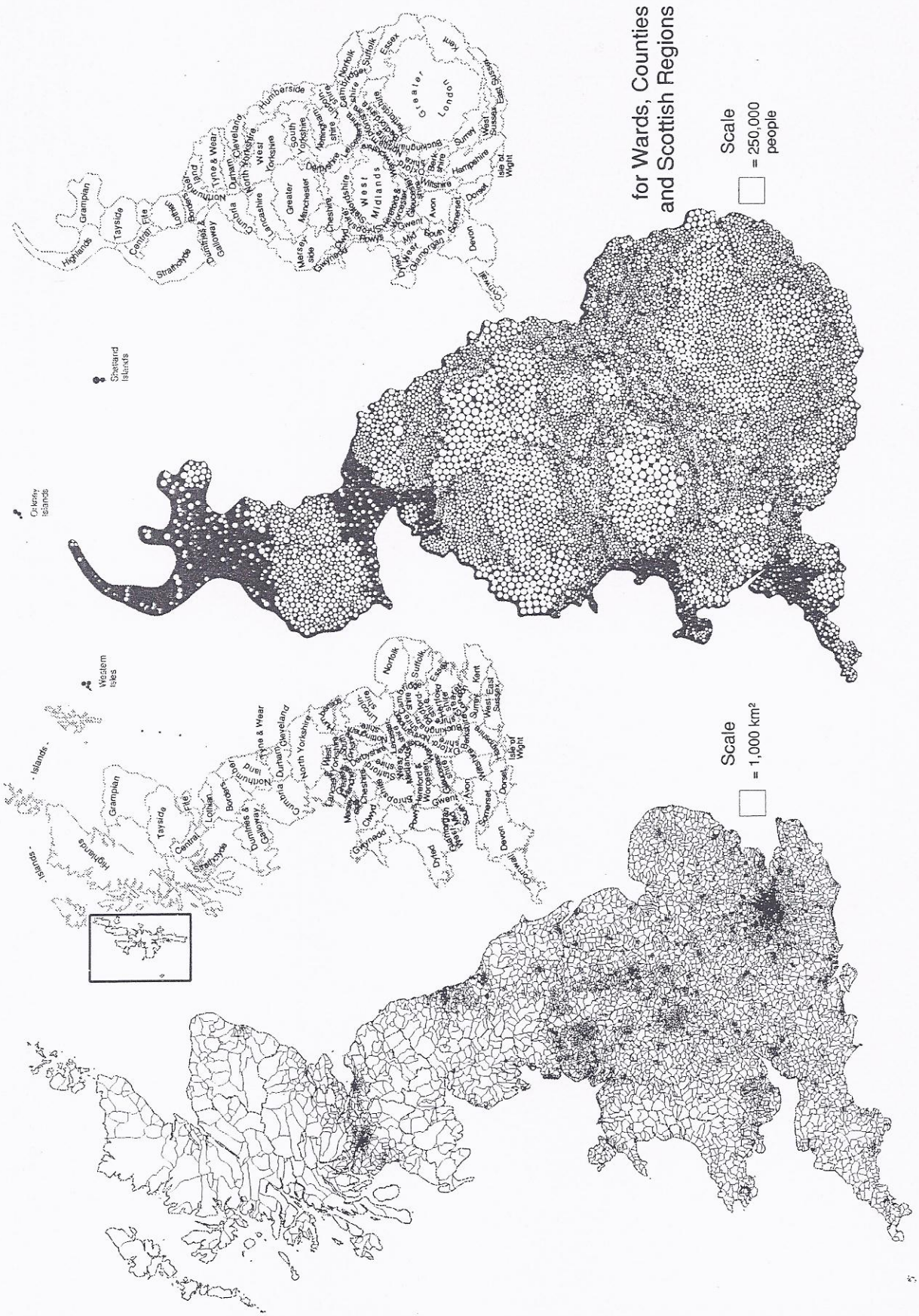

Fig 1. A traditional map and a population cartogram of Britain

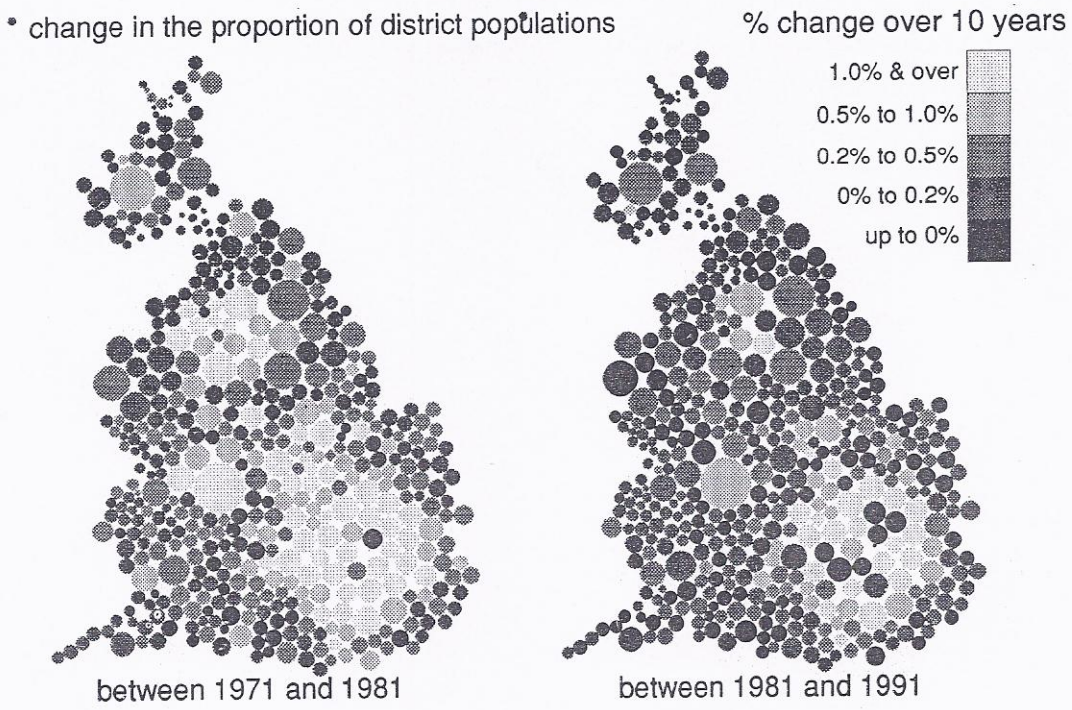

Fig 2. People resident in Britain, born in the New Commonwealth 1971-1991

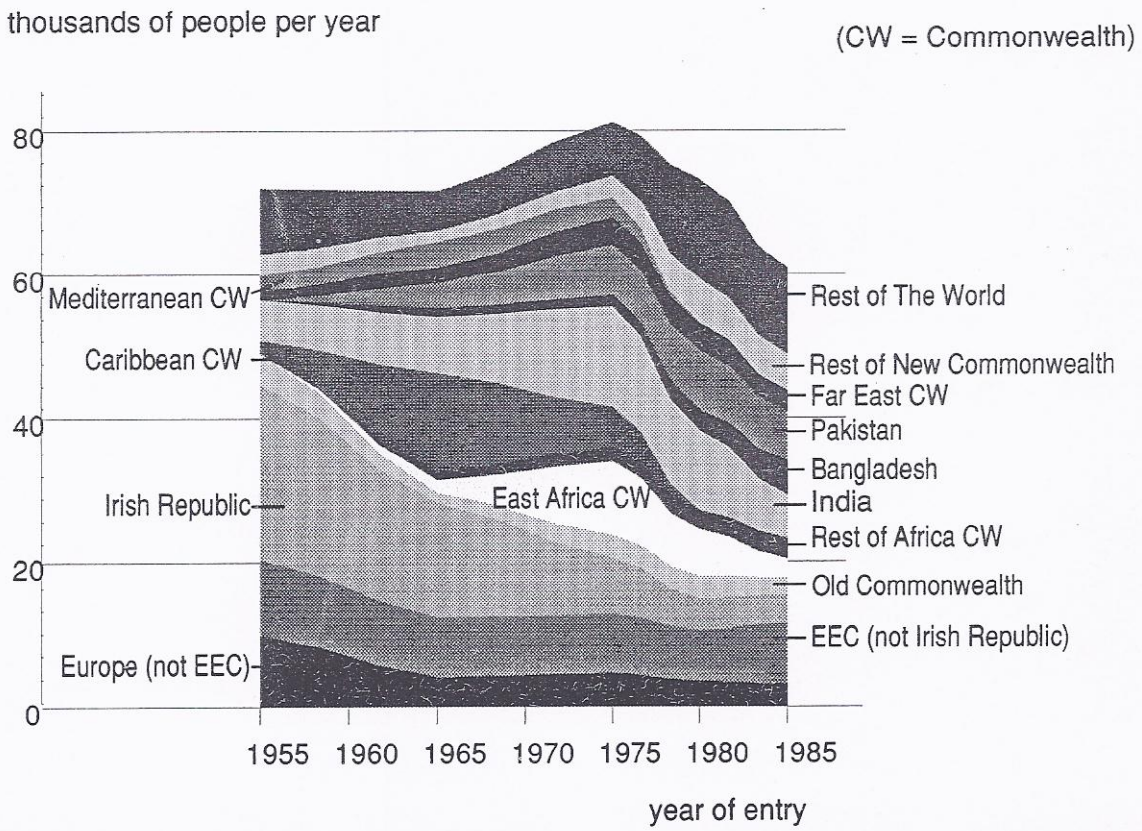

Fig 3. Residents born outside the UK by year of entry and country of birth 1989-91

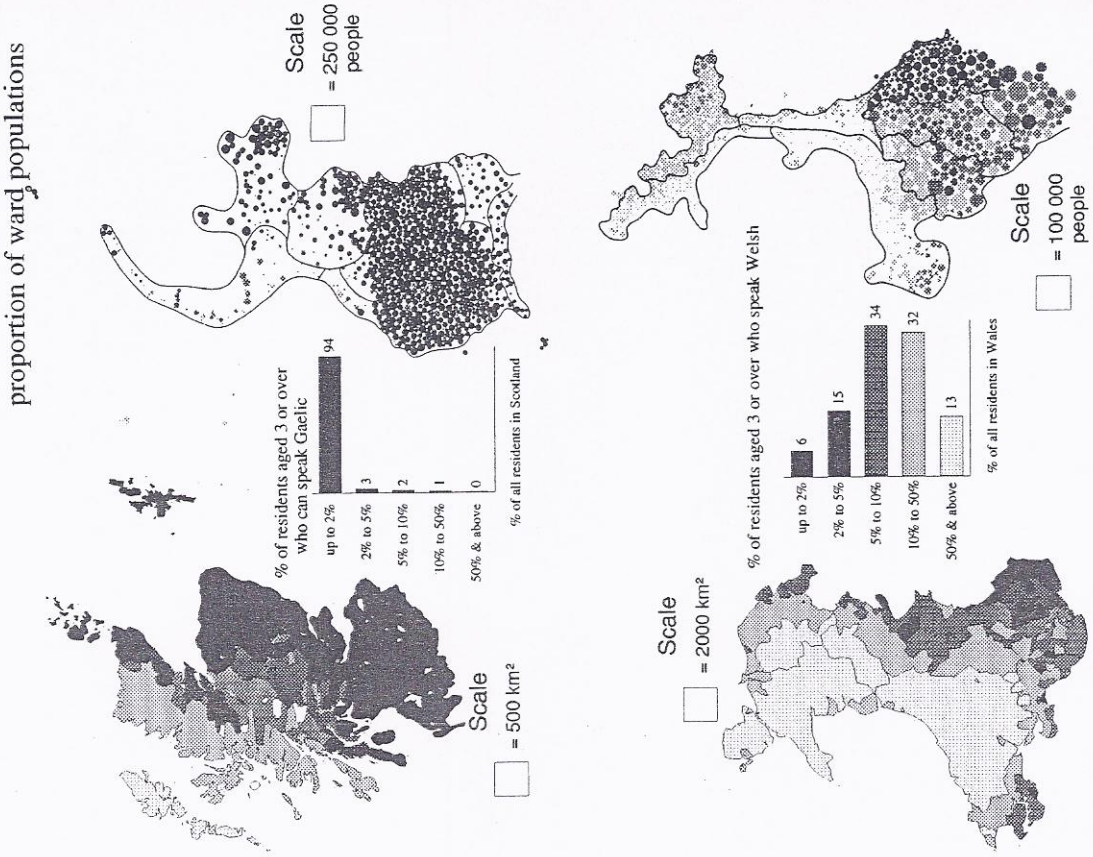

Fig 4a. Residents speaking Gaelic or Welsh 1991

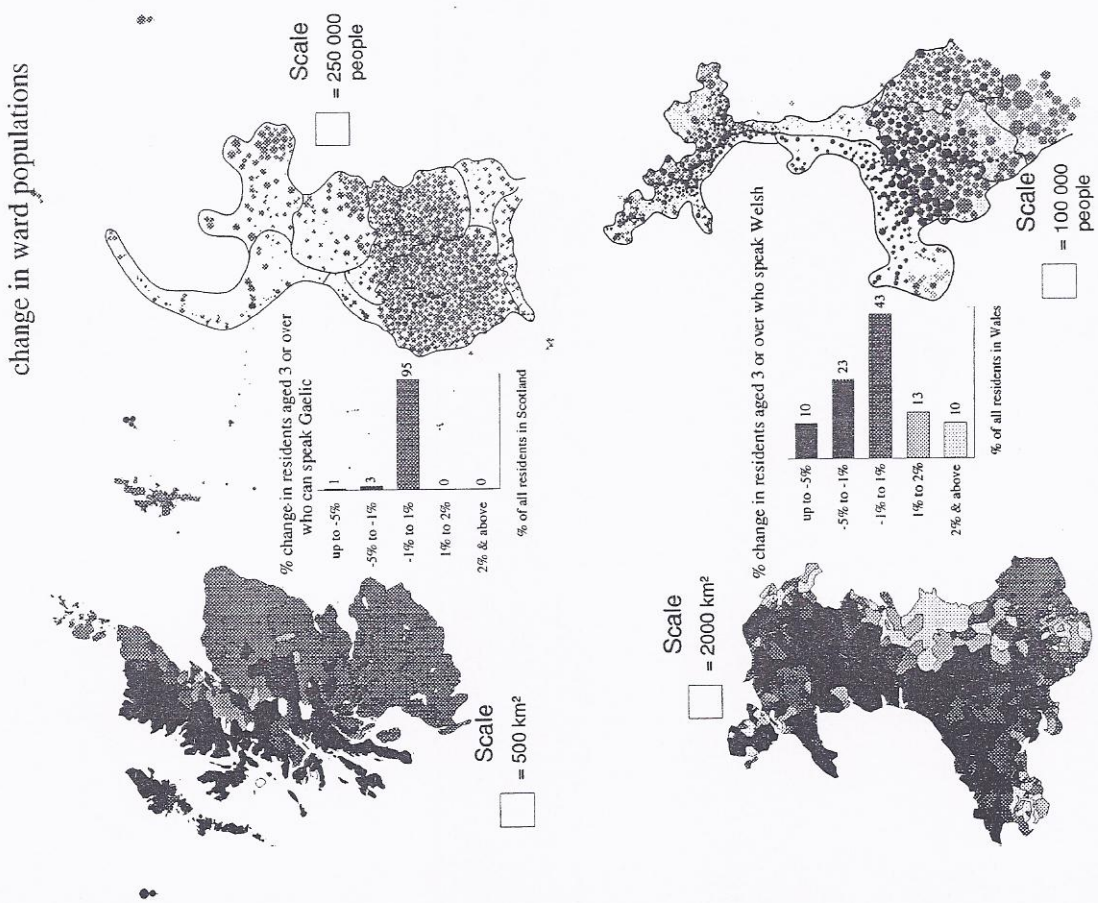

Fig 4b. Residents speaking Gaelic or Welsh 1981-1991

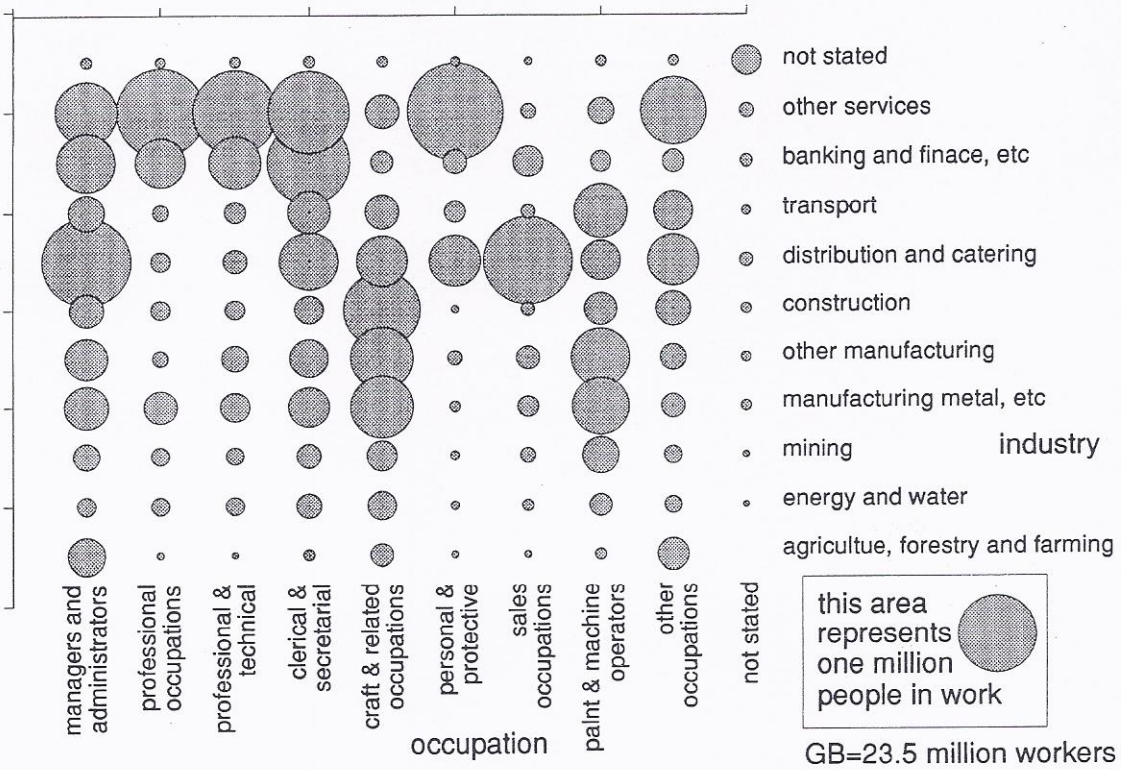

Fig 5. People in employment by industry and occupation in Britain 1991

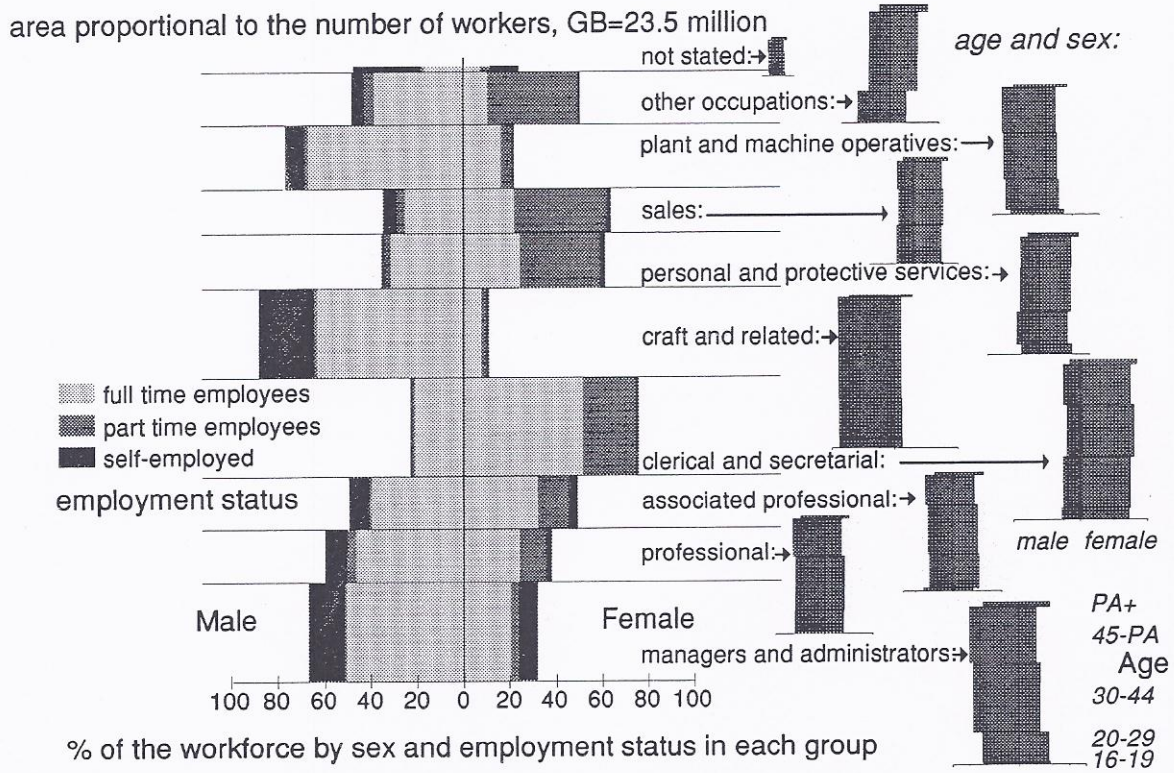

Fig 6. Residents in work in Britain by age and sex for each occupational group 1991

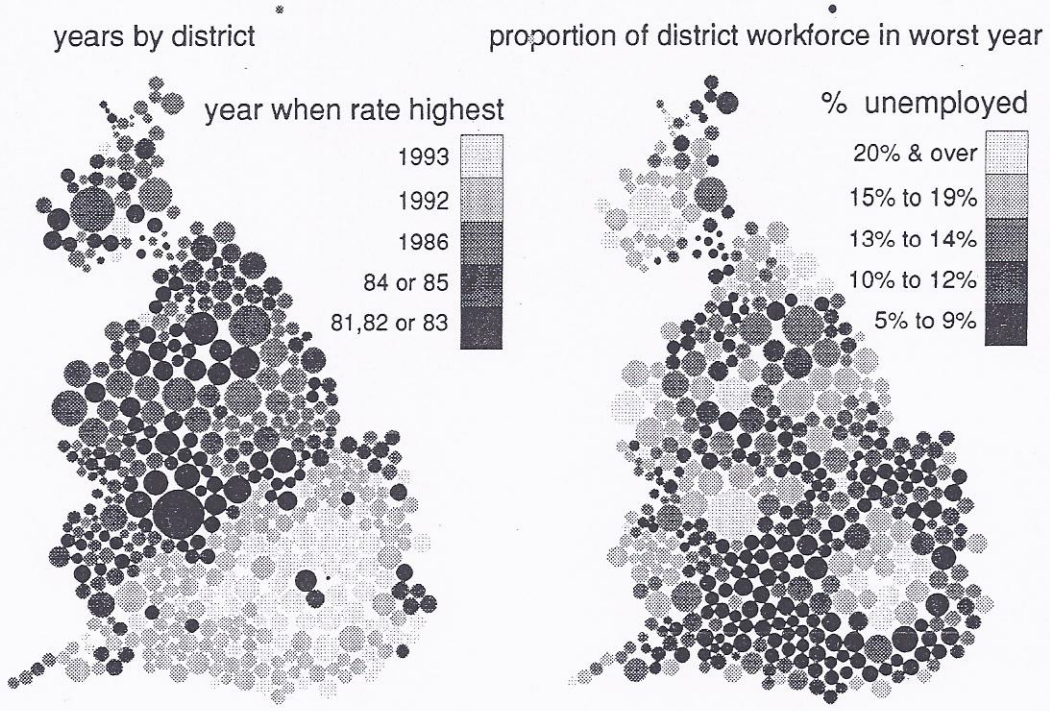

Fig 7. Year and rate of highest unemployment levels in Britain 1979-1993

Fig 8. Unemployment and house prices 1981-1991

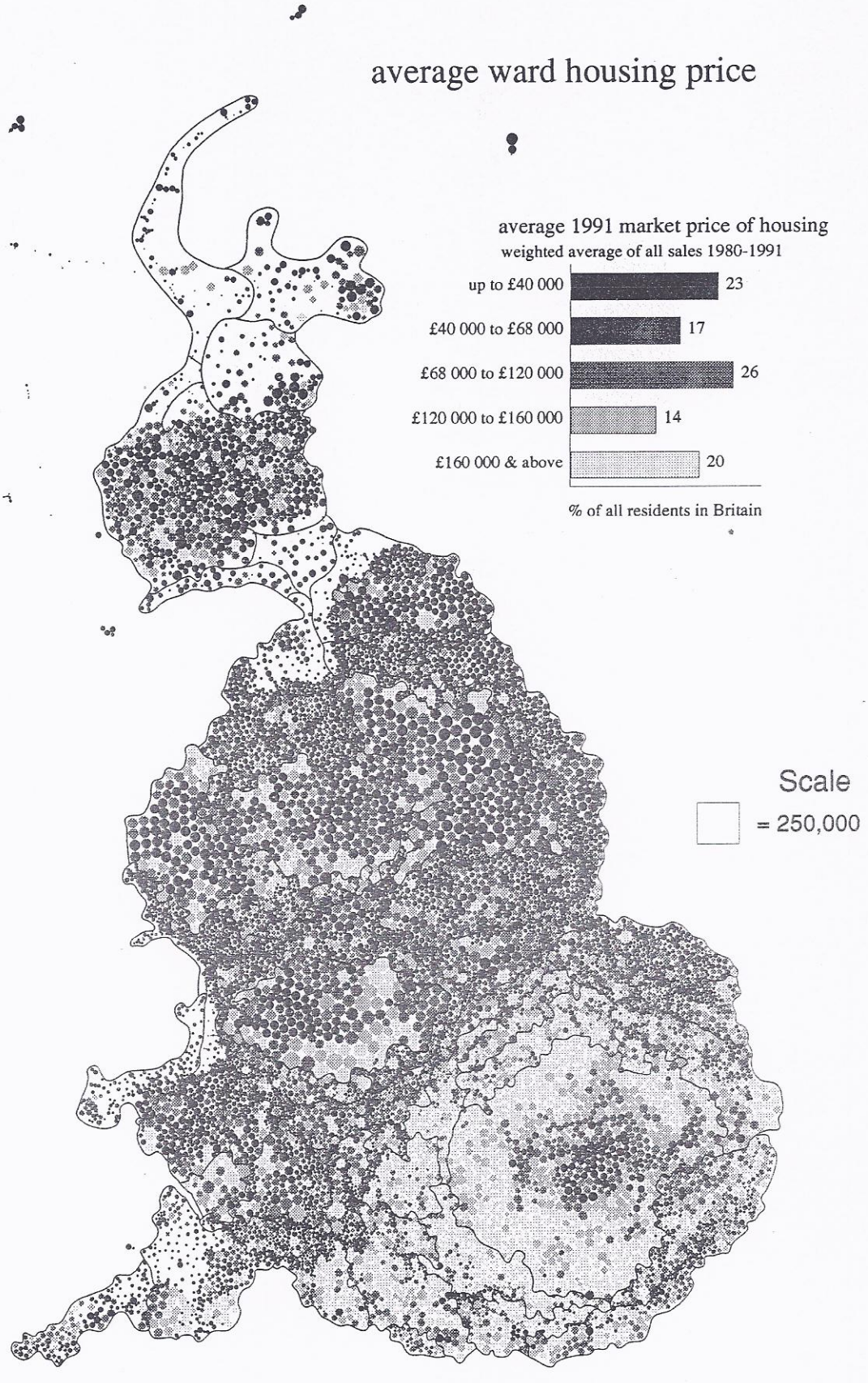

Fig 9. House prices by 1991 average ward housing price

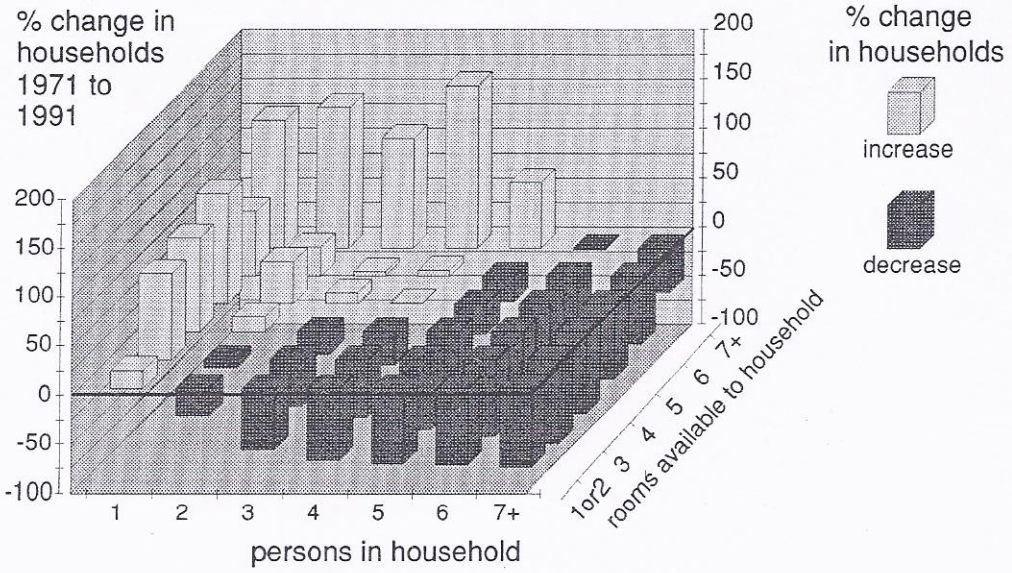

Fig 10. Persons and rooms per household in Britain 1971-1991

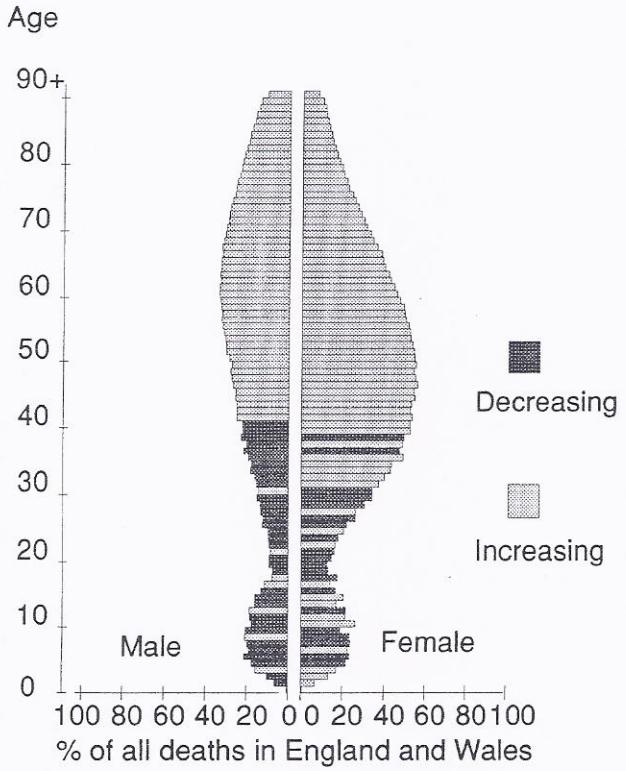

Fig 11a. Mortality from cancers between 1981 and 1989

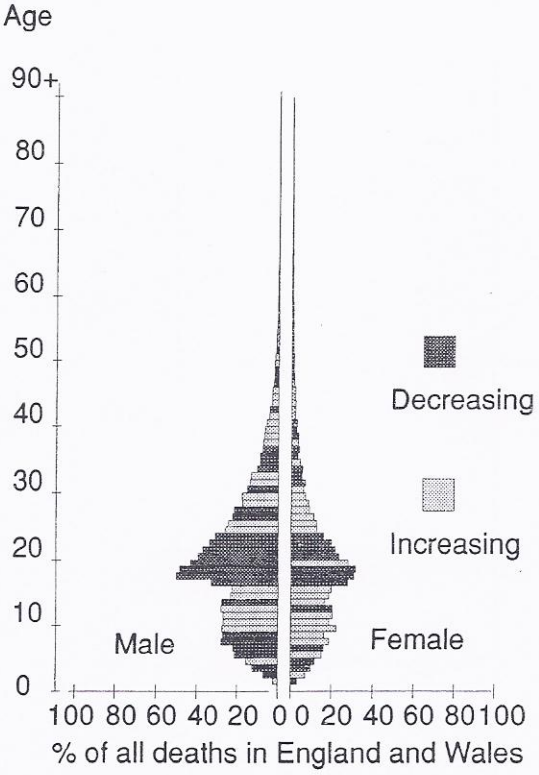

Fig 11b. Mortality from traffic accidents between 1981 and 1989

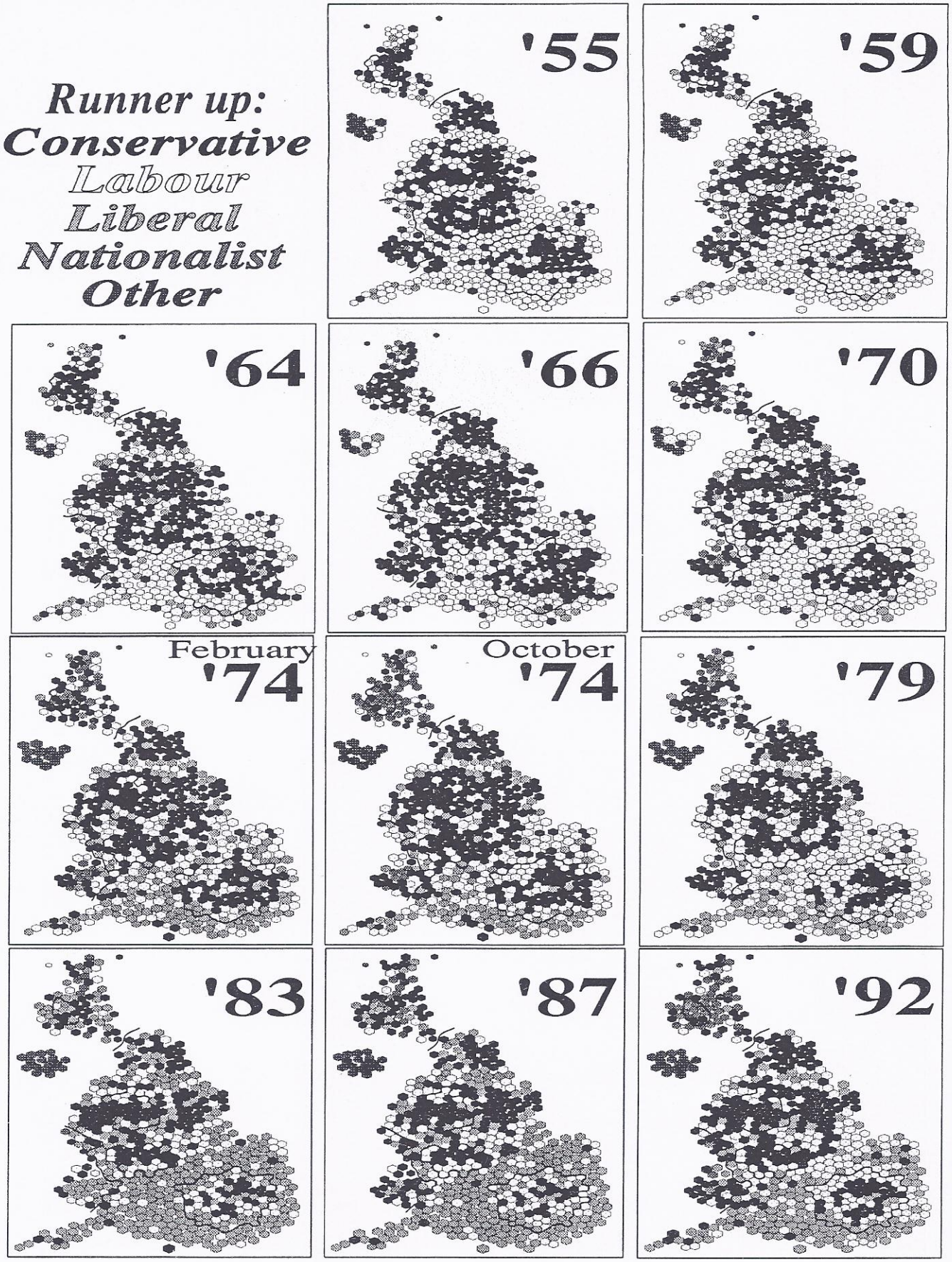

Fig 12. Party placed second in each general election in Britain, 1955 to 1992, by constituency