Title:	**Visualising Past Geographies: The use of animated cartograms to represent long-run demographic change in Britain**
Authors:	Humphrey Southall and Ben White
Address:	Department of Geography, Queen Mary & Westfield College, Mile End Road, London E1 4NS
Telephone:	0171-975-5413
Fax:	0181-981-6276
Email:	H.R.Southall@qmw.ac.uk

Abstract:

Some social processes are directly experienced, but the effects of demographic change are often slow and imperceptible. Further, in a country such as the UK there is much geographical variation and many of the extremes are found among urban populations concentrated into small areas which barely figure on conventional maps. Cartograms -- maps in which areas are made proportional to some other variable such as population -- help solve this second problem while animation deals with the first. The paper presents early results of research based on combining a large historical GIS for Britain, constructed at QMW and containing both a large volume of census and vital registration data from 1851 onwards and the CHANGING boundaries of the various reporting units, and an algorithm developed by Daniel Dorling (Bristol) for the automatic computation of cartograms. Each district is represented by a circle whose changing radius shows population growth or decline; processes contributing to that growth, such as net migration, are shown by changing shading. The animated cartograms we create cannot be conventionally published but can be distributed on CD or viewed over the World-Wide Web.

Visualising Past Geographies:
The use of animated cartograms to represent long-run demographic change in Britain

A week is a long time in politics, and the 'medium term' can mean a year or so to an economist, but a decade is a relatively short time to the demographer, concerned with change over the generations. Unsurprisingly, politicians, economic policy makers and the general public tend to neglect the consequences of demographic change until it hits them in the face. This periodically happens, and three good examples are the consequences of an aging population for pensions and health care provision; the consequences of out-migration from the great cities for hospital provision, and hence hospital closures; and out-migration from Labour's heartlands in the north for the party's share of the vote. These examples are relatively well known, but other more obscure shifts may be just as important, such as the aging of the workforce in declining regions and its consequences for productivity and adaptability.

This paper reports on early results from a major research project based at Queen Mary and Westfield College which aims not just to reconstruct and analyse these changes but to make them comprehensible to a wide audience through visualisation and animation techniques. Our current funding is varied and problematic, but our goals are clear: firstly, to assemble a comprehensive statistical database for the UK drawing on census and vital registration data for the last two centuries; secondly, to link this to a Geographical Information System (GIS) mapping the **changing** boundaries of the principal reporting units used, down to parish level; and thirdly to use this very large data structure as the basis for both novel research and a comprehensive atlas of the country's changing population. Funding permitting, this atlas will appear in March 2001 to mark the bi-centenary of the first census, and we already have the support of the Office of National Statistics (ONS) for this.

We are not short of raw data. The UK has had a comprehensive system of demographic reporting since the mid-nineteenth century with two main elements. Firstly, a population census has been carried out every ten years since 1801, with the exception of 1941; and since 1841 it has recorded basic information on age, sex, occupation and birthplace for every individual inhabitant. Secondly, since 1837 the civil registration system has recorded births, marriages and deaths, the last with cause. However, most of this information is decidedly raw: demographic data is available in machine-readable form from the Office of National Statistics, via the Essex Data Archive, only from 1971; for earlier periods, it takes the form of page after page of tabular statistics in the printed reports of the census, and the *Annual Reports* and *Decennial Supplements* of the Registrar General. A major part of our work has therefore been to input and otherwise assemble statistics from these sources.

The data available to us is limited partly by data entry resources but also by the information that was gathered and tabulated. Firstly, 19th and early 20th century censuses recorded much more limited information for each individual and household than those for the post-WWII period. Secondly, where modern demographers have access to enumeration district-level Small Area Statistics and individual deaths recorded with postcodes, we are limited to the coarser networks of reporting units used in the printed reports. Thirdly, use of some variables in long-run studies is problematic due to changing categories; for example, the 1851 census tabulated occupations using a system in which the most fundamental division was into workers with animal products, workers with vegetable products and with mineral products, while notions of industrial structure and social status were blended into a single occupational classification in all censuses up to 1911. Similarly, cause of death classifications have been drastically altered over time.

These problems mean that we are currently concentrating on strictly demographic variables which entirely avoid problems of classification, and can therefore be gathered for all censuses and inter-censal periods from 1851 onwards. The first is the age- and sex- structure of the population: all census reports tabulate this using five year age bands, so we know for example the number of people in each geographical unit who were, say, males aged 30 to 34. The second is the number of deaths over an inter-censal period similarly tabulated in terms of age and sex. Taken together, these statistics permit computation of two key indicators of demographic process. Firstly, life expectancy at various ages, which involves a complex calculation starting from crude death rates for each age band. Secondly, age- and sex-specific net migration rates, estimated by comparing the size of a given population cohort as recorded by one census with the size of the same cohort in the previous census; once we adjust for mortality in the inter-censal period, any differences in size should be due to in- or out-migration by individuals in the relevant age- and sex- group.

There is rather greater variation in how mortality statistics were recorded, so the example which follows is based purely on census data recording the age and sex structure of the population. Between 1851 and 1911, we have data drawn from every census report for England and Wales recording this for each of the c.620 Registration Districts, units based closely on Poor Law Unions and in general combining towns with their rural hinterlands. For 1921 and 1931, we have data for each of the c.1,800 local government districts (County and Municipal Boroughs, Urban and Rural Districts); we also have funding and permission from ONS to add data for the same units in 1951 and 1961, while Small Area Statistics are available from the Data Archive and MIDAS for 1971, 1981 and 1991.

Unfortunately, while the classification of age/sex structure is consistent over time, the reporting units are not. Firstly, the numbers of units given in the previous paragraph are approximate because in every inter-censal period some units were created through sub-division or eliminated by mergers, while a far larger number of units had their boundaries substantially revised. Much of our work in GIS construction has involved carefully recording all these changes, enabling us to generate an accurate base map for any date. Secondly, there have been two complete changes in the system of reporting units: in 1911, when the Registrar General switched from Registration Districts to local government districts; and in 1974, when the Redcliffe Maud reforms transformed the geography of local government.

Figure 1 shows the sequence of operations involved in creating an animated visualisation of changing population structure, and the first stage is the creation of a new dataset for a constant set of units. For now, we are using the 635 Registration Districts used by the 1911 census, and for many Districts constructing data for earlier and later dates is unproblematic: if the boundary was not changed, data from censuses from 1851 to 1901 can be used unaltered, while for 1921 and 1931 we need simply to aggregate together the Rural District and the one or more Urban Districts into which the Registration District was divided. However, for many other Districts this procedure is insufficient and our current method is as follows: first, use the GIS to overlay the boundaries of reporting units for the date in question onto the 1911 Registration District boundaries, and for each reporting unit compute the percentage of its area which was in each 1911 District; then use these percentages to re-assign numbers in each age/sex band to the 1911 Districts.

This procedure is obviously unsatisfactory in areas where major boundary changes occurred. For example, Barrow-in-Furness RD (Lancashire) was carved out of Ulverston RD on July 1st 1876, and as it was a geographically small District with a large population while much of the remainder of Ulverston was moorland our current method grossly underestimates its population for the censuses 1851-71. However, there is large scope for improving our re-assignments by extending the underlying GIS and database. Firstly, a much more detailed parish-level GIS is under construction and we are beginning to acquire the corresponding parish-level population counts from the census reports, permitting us to relax the crude assumption of constant population density over each District. Secondly, the census reports for 1851 to 1881 also tabulate age/sex structure for c.2000 Registration sub-Districts, which we are in the process of mapping, while from 1891 onwards they include tabulations for the developing network of local government districts. Our present crude method of re-assignment suffices for the development of visualisation methods.

Once we have demographic data for a constant set of geographical units we could represent change by simply mapping the data onto a sequence of conventional maps of 1911 Districts. However, such a map would give a very misleading impression because the population of England and Wales is highly urbanised, and was even more highly urbanised in the late 19th and early 20th centuries: most of the population and most of the associated demographic process was concentrated into geographically small highly urbanised districts, barely identifiable on the conventional map; conversely, the map is dominated by large districts in areas of sparse population. The standard cartographic solution to this problem is to replace the conventional map by an area cartogram, in which the size of each unit is in proportion to its population, not its geographical extent. Manually constructed cartograms have been used for many years, particularly in the presentation of electoral data, but are very time consuming to construct. Unfortunately, no entirely satisfactory computational method has yet been developed for constructing true area cartograms, but our collaborator Daniel Dorling has developed software which creates cartograms based on proportional circles. Note that animated cartograms permit, in effect, the mapping over time of not one variable but two: the first determines the changing shade or colour of the circle, the other its changing size.

The next stage in the creation of our animated cartograms is therefore based around Dr. Dorling's software whose functionality we have extended in three ways (NB the tasks carried out by this program are enclosed by the dashed box in figure 1). Firstly, we embedded the code for cartogram generation in a loop which covers a whole sequence of different dates. Secondly, we have given it the ability to extract data directly from our Oracle database by embedding SQL commands within the original 'c' code. Thirdly, where the original program creates an input file for the Arc/Info GIS package, which is then used to create the actual graphic, our program directly outputs a sequence of Postscript files, one for each date. These changes were made partly to automate the construction of a sequence of cartograms but also to accelerate a computationally intensive process: the original program as embedded within Arc/Info took c. 45 minutes per cartogram while the extended program takes c. 15 minutes, mainly because we no longer use Arc/Info to retrieve data and output graphics.

Further time savings can only come from revising the actual computation of the cartograms. The algorithm takes as input, firstly, two variables taken from the database which define the 'size' of each unit and the fill pattern — for example, population and the sex ratio — and secondly a text file which defines the 'geography' to be used. This file is created by an Arc Macro Language (AML) program run within the GIS and record, for each unit, the coordinates of its centre (or 'centroid'), the other units with which it shares a boundary and, for each such neighbouring unit, the length of the shared boundary. The algorithm then proceeds as follows. Firstly, a conventional proportional circle map is created, the circle for each unit being located on the centroid; this inevitably contains large overlaps and gaps between circles which must be eliminated from the final cartogram. Then an iterative process moves the circles so as to reduce overlaps and gaps to create a set of neatly nested circles, while seeking to keep units which were adjacent in the original map as close together as possible; the length of their shared boundaries determines the priority given to this. This process requires roughly one-and-a-half as many iterations as there are units in the map, or c.1000 iterations with 1911 Registration Districts; however the final result, as illustrated in figure 2, accurately maps population sizes, avoids overlaps and preserves the broad shape of England and Wales.

The final stage takes us from a series of Postscript files, one for each individual date, to a single file containing an animation. The method partly depends on the specific final format, the main options being Microsoft's AVI, Apple's Quicktime and the non-proprietary MPEG format; all are bit-map based formats which achieve a high level of compression by using similarities between successive frames as well as redundancies within them. For development work so far, we have been using a very awkward series of conversion, using Adobe Illustrator to convert Postscript files into GIFs; Paintshop Pro to convert from GIFs to DIBs; and Video for Windows to convert a sequence of DIB files into an AVI 'movie'. This leads to degraded images as well as being time consuming, and we are hoping to use a specialist national service at Manchester to convert directly from PostScript to MPEG, which will provide better cross-platform support than AVI.

We are only six months into a two year project, although it has a number of other goals, and so far our main emphasis has been on simply generating some kind of animated cartogram. Future refinements will be in three areas. Firstly, we need to accelerate the software and one obvious possibility is to reduce the number of iterations needed to create all but the first frame by using the centroids of the previous frame rather than real-world centroids to generate an initial proportional circle map; this requires a substantial rewrite of Dr. Dorling's code. Secondly, we aim to explore how best to create an effective visual display. For example: how many interpolated frames for non-census dates are needed to create a smooth animation? should we make use of Postscript's ability to precisely specify colours within a continuous range to directly represent the underlying continuous variable, rather than using a limited set of classes to select the fill for each unit? how best can we indicate the date, given that a flickering set of digits is hard to read. Lastly, as already noted, we intend to plot more meaningful variables which are more directly a measure of demographic process, covering mortality and migration.

The nature of our work makes inclusion of examples impossible on the printed page. Figure 2 is based on a single frame, for 1911, taken from a sequence showing the sex ratio among young adults; high ratios of males are apparent for the South Wales and Yorkshire coalfields, growing rapidly in this period, and for garrison towns in the south-east; in the animation, a relationship between a high proportion of males and rapid growth is apparent. However, examples of full animations can be viewed on our Web site:

`http://www.qmw.ac.uk/~gbhgis/examples.html`

One word of warning is necessary: we include one or two animated GIFs which will work in a standard Web browser, but the other formats we are using require various plug-ins; our site includes links to sources from which these can be downloaded. Note that to minimise file sizes and therefore download times the animated images are kept as small as possible; unchanging titles and legends are provided by separate GIFs, and explanatory text as HTML.

Acknowledgments:

Construction of the GIS and database has been funded to date by the Aurelius Trust, the Economic and Social Research Council, the Joint Information Systems Committee, the Leverhulme Trust, the Nuffield Foundation, the Pilgrim Trust, the Population Investigation Committee and the Wellcome Trust. Our visualisation work as presented here is funded by the Joint Information Systems Committee under their JISC Technology Application Programme (JTAP), through a project titled 'Authoring methods for low-cost electronic atlases of change and the past'. Construction of the GIS has been led by Ian Gregory, and he was responsible for creating the datasets for standardised units. The core of our program for generating animated cartograms was supplied by Daniel Dorling of Bristol University. Parts of our two main demographic datasets, for age structure and age-specific mortality, were supplied by Eilidh Garrett (ESRC Cambridge Group), David Gatley (Staffordshire University), Graham Mooney (Centre for Metropolitan History) and Robert Woods (Liverpool University). Eilidh Garrett and Graham Mooney have also assisted us in the development of methods for computing, respectively, age-specific net migration and life expectancy.

Figure 1:

Stages in Creating an Animated Cartogram

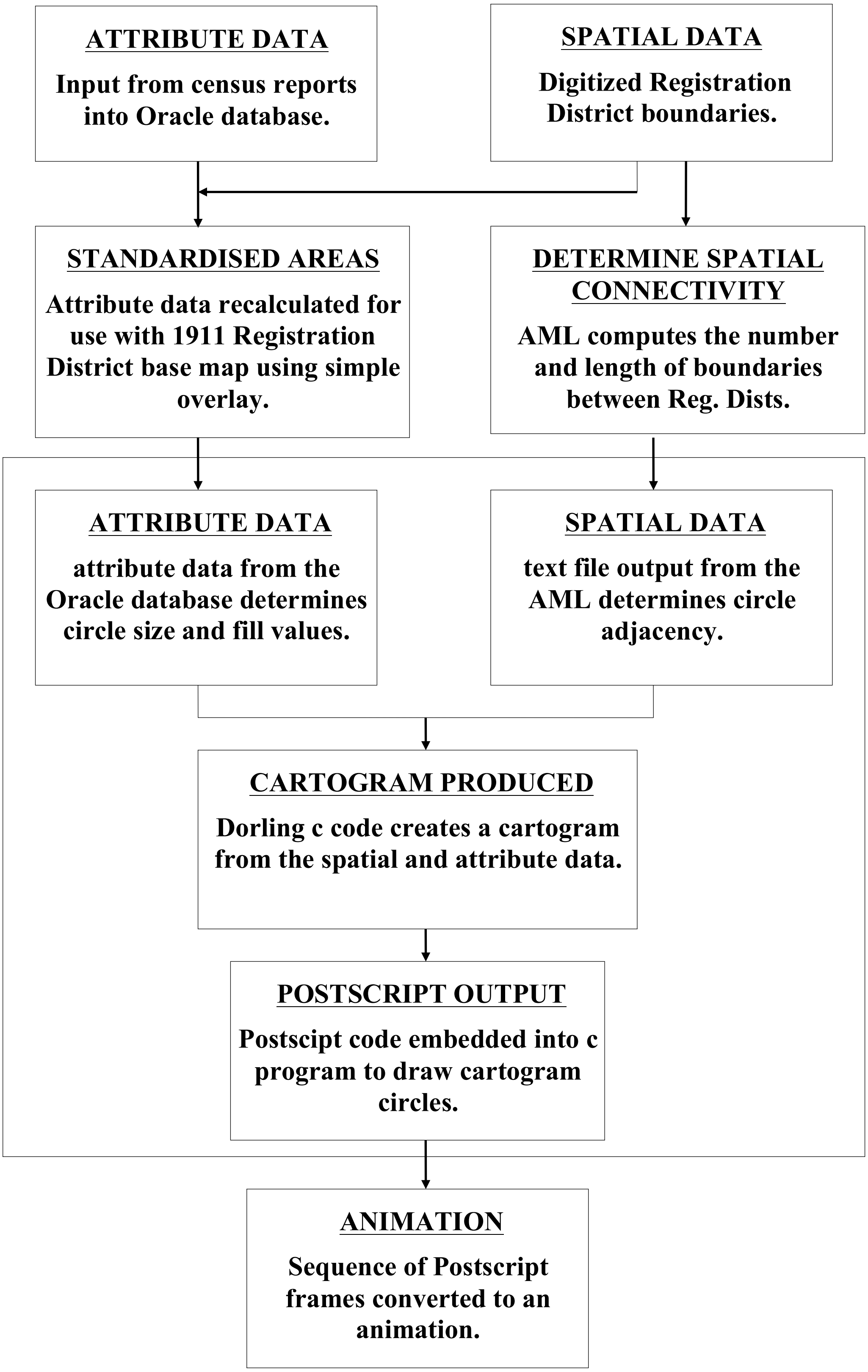

Figure 2:

Males as a Proportion of All Young Adults in 1911