

Commentary

An Online Census Atlas for everyone

Helen Durham*, Danny Dorling** and Phil Rees†

*School of Geography, University of Leeds, Leeds LS2 9JT

Email: h.p.durham@leeds.ac.uk

**Department of Geography, University of Sheffield, Sheffield S10 5TN

†School of Geography, University of Leeds, Leeds LS2 9JT

Revised manuscript received 26 January 2006

Aims

The Online Census Atlas is a highly interactive web-based visualization and exploratory data analysis tool giving access to 88 key census variables from 1971, 1981, 1991 and 2001. It aims to deliver a simple mechanism for mapping statistics from GB censuses over three decades and to provide users with a method for mapping using a cartogram that reflects the population weight of an area.

Data resource: linking censuses through time

Central to the development of the online Atlas was the extraction of statistics that were comparable over the last four UK censuses and this process was facilitated by the use of the web-based resource, Linking Censuses through Time (LCT) (Dorling *et al.* 2001; Martin *et al.* 2002; Mitchell *et al.* 2002).

The LCT project linked 1971, 1981 and 1991 census data at a number of geographical levels via a web-based interface, with the aim of facilitating change-over-time analysis. The census data were restricted to the countries of Great Britain; Northern Ireland data were unavailable to the academic community then. The resource was completed in 2001 and built on work previously carried out by Atkins *et al.* (1993) and Dorling and Atkins (1995) in which the 1991 census data were linked to the 1981 ward geography via look-up tables. 1971 census data had also previously been aggregated to 1981 wards by Dorling (1991). As 1981 was the

midpoint of the three census periods it was chosen as the geographical base for the resource.

Access to the LCT data is via interactive menus to select options of date, geography and subject which are converted by the program into a Unix line command. This command line, which specifies which data are to be extracted from the compressed versions of the census data tables, is passed to existing C code which, in turn, reads external look-up tables to produce run-time aggregations to the requested geography. Output of the data is in CSV format.

The end result of this project was streamlined access to census data from 1971 to 1991 across a wide range of geographical areas, including 2001 geographies such as parliamentary constituencies and local authorities (Martin *et al.* 2002). However, to continue serving social scientists needing information on the change in social and spatial distribution of the population there was a need to incorporate 2001 census data. Although 2001 data are not currently available in the LCT resource, it does include 2001 geographies making it possible to extract census counts from public-domain census tables published by the statistical offices of the United Kingdom and merge the data with LCT data with relatively minor manipulation.

Overview of the Atlas

In the 1990s software developers at the German National Research Centre for Information Technology developed CommonGIS, Java-based software which provides interactive exploratory analysis of spatially

referenced data through a web interface (Cole and Hayes 2001). In 2003, this software was adapted as part of a JISC-funded project, the Collection of Historical and Contemporary Censuses (CHCC) to be used with its resources. This project was a collaborative venture between the Universities of Leeds, Manchester, Essex and Glasgow. The University of Leeds was responsible for developing online learning and teaching material to help students understand the Census Area Statistics datasets from 1971 to 2001. One of the resources developed was the Online Census Atlas.

This Atlas currently displays maps of 88 census variables at six geographical levels:

- Government Office Region (GOR).
- Local Authority (LA).
- Historical County (HC).
- County and Unitary Authority (C and UA).
- Parliamentary Constituency (PC).
- European Constituency (EC).

Users select the variable and year of interest (e.g. 1971 or 2001) or change-over-time period (e.g. 1971–1981). Two maps are produced using CommonGIS, one map being the variable displayed within generalized boundaries of the selected geography and the other is the variable displayed within a cartogram. Either map can be manipulated and displayed according to users' preferences using the CommonGIS visualization tool or the map boundaries can be downloaded in ESRI shapefile format for use in GIS or mapping packages. It was felt appropriate, when designing the Atlas, to include both a conventional map and a cartogram; they provide the same information, but the appearance of the two maps is different. To understand the philosophy behind the cartogram and how it differs from the conventional map, the two kinds of maps are now described.

Conventional maps

The familiar choropleth maps in the Online Census Atlas were produced using generalized boundaries. Great Britain is recognizable and it is relatively easy to identify the higher geography levels such as Counties and Local Authorities. A disadvantage of the conventional map is that sparsely populated rural areas may be emphasized, whereas the areas representing cities are very small, making interpretation of spatial patterns very difficult.

The generalized boundaries were produced in the School of Geography at the University of Leeds based on detailed 1981 ward boundaries. The generalization process used an algorithm developed by Dorling and reduced wards to very simple polygons with small wards represented as triangles. There were three main goals in the development of the generalized ward boundaries:

- 1 Ward boundaries needed to be in roughly the right place so that when aggregated to higher geographies, for deployment within the Atlas, the polygons were a simplified but accurate representation of the zone.
- 2 Each ward had to be represented by a simple polygon (e.g. a triangle) to keep the overall file size to a minimum, permitting faster visualization.
- 3 All wards had to be visible, although magnification would be required to view the small urban ward boundaries.

The Atlas does not allow visualization of census data at ward level, but instead uses these boundaries aggregated to the six higher level geographies defined above for display of 88 census variables. The aggregation process of wards to higher geographies was undertaken using look-up tables developed by the authors.

Cartograms

Cartograms are derived from conventional map boundaries, but each zone is proportional to a measure, such as the population. The cartograms deployed in the online Atlas were developed using spreadsheet software and labelled as Universal Data Maps or UDMs. These UDMs were converted to vector format maps for use within the Atlas. A description of the design and build process for the Atlas is given below.

The creation of the UDM involved two key but separate jobs: initial creation of the map and subsequent detailed checking, both arduous and largely manual processes. The creation of the map began with a computer-generated cartogram of parliamentary constituencies where each constituency was given equal weight. Using the circular cartogram algorithm (Dorling 1995), this resulted in a largely hexagonal tessellation of 641 circles, each circle representing a parliamentary constituency as first used in the 1997 general election. This tessellation was transferred to several large sheets of graph

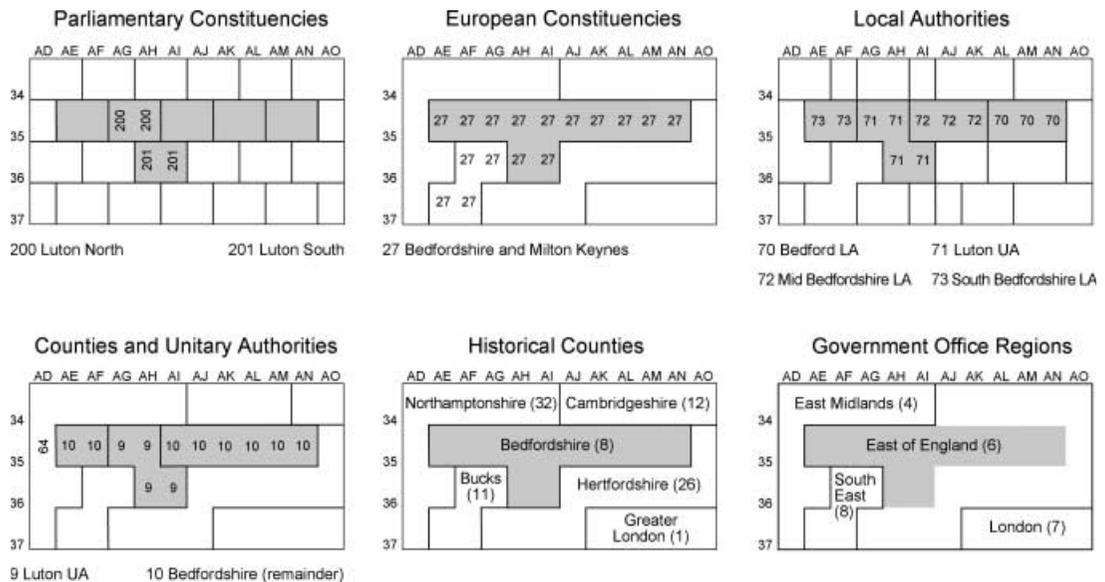


Figure 1 How the different geographical areas are made up from building bricks of half Parliamentary constituencies in the cartogram spreadsheet

paper and ‘tidied’ up by hand so that each constituency was represented by a ‘brick’ of two unit squares (topologically hexagonal in tessellation). Regional, county, health authority and former European parliamentary constituency boundaries along with other geographies, such as travel to work areas, were then drawn on this map using coloured pencils. This drawing of boundaries revealed many areas where the topology of these larger areas was broken by the particular arrangement of constituencies first generated by computer algorithm. Solving these problems would be very difficult by computer program, as the implications of rearranging the constituencies in one area were multiple and could affect parts of the map further away. There were also aesthetic considerations to bear in mind. Most simply, where constituencies are labelled as being Town East and Town West, it is intuitive that they should lie right and left of each other if possible.

Following this, the local authority geography that was in place by the time of the 2001 census was overlaid on this map. Constituency ‘bricks’ were allowed to be split East/West for when, for instance, two local authority districts largely lie within one constituency. The smallest authorities – the City, Scillies and Rutland – were maintained, despite making up only half a constituency in size. A great many further changes had to be made to the UDM

to include all local authorities and still maintain the topology of higher areal units.

These paper maps were then converted into spreadsheets using Microsoft Excel. The principle of the design was that a geographical area would be represented by the cells of a spreadsheet, with the cell area being approximately proportional to the population of the geographical area. A UDM of the Westminster Parliamentary Constituencies was the first map to be created and involved converting the hand drawn constituency map into a 56 x 52 cell matrix. Each constituency was assigned two spreadsheet cells (representing approximately 45 000 persons each) and the constituency code allocated to each cell. Higher geography UDMs were also created based on the paper map produced by Dorling, supported by look-up tables developed by the authors. The UDMs can be used for simple choropleth mapping within Microsoft Excel itself or for more demanding mapping and analysis functions the boundaries of the UDMs have been converted to MapInfo and ESRI shapefile format.

Figure 1 illustrates, for a small portion of the GB cartogram, how the different geographical areas were constructed from the half parliamentary constituency building bricks.

UDM mapping is particularly suited to the inexperienced user, especially when the choropleth

mapping process is carried out in Microsoft Excel, because no specialist mapping skills are required and users only need access to spreadsheet software rather than more expensive, specialist software. In addition, there are no copyright restrictions on the UDM boundaries. Parliamentary constituency and higher geography UDM boundaries can be downloaded from the CHCC website (<http://www.chcc.ac.uk/CAS/resources.html>) from within Unit 13: Universal Census Data Mapping, which includes a tutorial on how to produce simple choropleth maps using standard Microsoft Excel functionality.

The most obvious disadvantage of the UDM is that the final outline of the UK is distorted. For example, Scotland becomes much smaller because it has vast areas of rural, sparsely populated land. Not so obvious, until you examine the UDM areas in greater detail, is that it was not always possible to maintain the topology of the geography, i.e. some areas that should be adjacent are separated. The regular shape of the cells restricted the number of neighbours possible, meaning that only four other areas could have a common boundary and four other areas could be joined at the corner only. In addition, higher geography UDMs were built from lower geography UDMs, so in order to maintain the integrity of higher geographies topology, the topology of lower geographies sometimes had to be compromised.

When deciding whether to use cartograms or conventional maps to portray spatial distributions from the census, you need to take into consideration that the information in the census concerns people and households, rather than land, so by trying to minimize the visual bias, cartograms can be claimed to be more sensible statistically and more just socially for census mapping than conventional maps. Cartograms (and UDMs) allow the fortunes of every group of people to be shown and do not disadvantage those groups who live in the most densely populated areas (Dorling 1993, 1995).

How to use the Online Census Atlas

Two versions of the Atlas exist – one on the University of Manchester MIMAS service (<http://www.chcc.ac.uk/atlas/>) and another based at Leeds University (<http://www.ccg.leeds.ac.uk/teaching/chcc/>). The MIMAS version displays the data in both the conventional map and the UDM, but access is restricted to users with an Athens username. The Leeds Atlas is a reduced version, as it only contains

UDM maps, but is freely available to anyone as no copyright issues affect these map boundaries.

The interface to either version is identical and requires the simple selection of geography and variables from a series of pull-down boxes which produces the maps 'on the fly'. This paper describes the use of the full resource located on the MIMAS server, but the instructions will be appropriate for the resource at Leeds University.

Upon accessing the Atlas, the user is taken to the selection interface. A form allows the user to define the map that they wish to view using four stages. Each stage must be completed before the next stage will become active. Stages include the selection of the geography, the category of variable, the census year/s and the exact variable the user wishes to map. Three buttons below the selection boxes allow, in turn, the creation of the map, the resetting of the selection or the selection of the pre-defined default values. The first button, 'Create the Map' should be clicked to initiate the Java applet.

Two maps will appear as an applet within the web browser window, both showing unclassified choropleth maps of the selected variable. Figure 2 shows an example of a map pair, reproduced here in black and white. The online maps are in colour. The maps show the change in numbers of households with two or more cars available for Government Office Regions over a 30 year time span, 1971–2001. Of interest is the low growth in the Greater London Authority area, where congestion pressure and high costs of ownership have limited the gains. However, overall the percentage increases are spectacular and show how Britain became a society that is making more and more use of car transport.

The map on the left is based on conventional generalized map boundaries, whereas the map on the right is the population-weighted UDM. Although the maps may appear a little different, they both display the same information. The visualization tool has limited functionality, but will allow users to zoom in, zoom out and zoom undo, save and print a map, select and de-select objects and move the map. The power behind this simple applet is in the ability to manipulate the visualization of the data; two tabs exist to the right of each map, one headed Legend and the other Manipulate. The Legend tab defines the variable mapped and by clicking on the three symbols next to the variable description users can alter the appearance of the map by changing parameters such as fill and line colours, label fonts, line thickness and even the variable description.

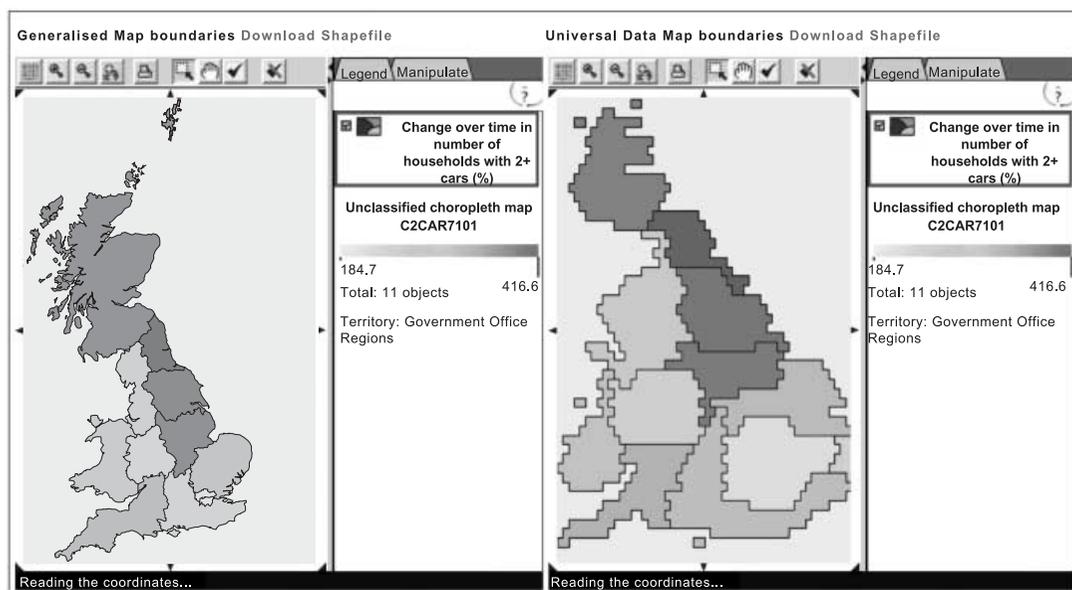


Figure 2 Example of an Online Census Atlas pair of maps: change 1971–2001 in per cent of households with access to two or more cars

The check box allows the map layer to be turned off. Users also have the ability to alter the colour of the choropleth shading by clicking on the attribute scale bar and selecting new colours to represent positive and negative values. The degree of darkness is proportional to the numeric attribute value, so higher attribute values are a darker shade than lower values.

Under the Manipulate tab, the attribute scale bar has become vertical and the data can be compared to a chosen value to more easily observe which values are above the value and which fall below it. Clicking anywhere on the attribute bar selects that point (the critical midpoint) and places the value in the Compare to box. An indicator appears on the bar at that point and a diverging colour scheme is seen above and below it, which is reflected in the shading on the map. This provides an immediate visualization of how variables in one geographical area compare to others. Checking the Dynamic map update box allows users to dynamically see the progression outward from any critical midpoint; simply drag the point indicator on the attribute bar up and down the scale to see the spatial patterns altering and trends developing. The remaining check box on this tab, Compare by click in, permits the user to select a specific geographical zone as the midpoint (turning the zone white), thereby visualizing which parts of the country have greater or lesser attribute values.

A further feature of the Atlas is the ability to download the map boundaries and data in Shapefile format. The link appears above each map and can be saved to a local PC. Users wishing to download the Generalized Map boundaries should obtain a click-use licence from HMSO.

Discussion

The Online Census Atlas provides users with a very quick and easy way to explore the spatial patterns of change in selected census topics over four censuses and three decades. The geographical units have been harmonized and comparable variables produced from each census. The Atlas adds visualization to the work of the Linking Censuses through Time project. It should be possible in future to develop this harmonization and visualization system for much smaller areas, such as output areas or their super-output aggregations. But the research effort would be considerable if the whole country and the full range of census topics were to be covered. The tools for carrying out harmonization (e.g. the ONS/GROS/NISRA Postcode Directories now available at the UKBORDERS data support unit) have improved considerably in the accuracy of the georeference attached to the postcode and the accuracy of the links to other geographies. The

ambition of the census organizations is to develop a national address base of even greater accuracy, from which harmonized units can be defined. Vital to this project will be the integration into the plan for the next census in 2011 of the re-working of previous censuses: 1991 and 2001 quite accurately and 1971 and 1981 more approximately.¹ However, prior to such a wonderful database being available, the Online Census Atlas can be used to explore which changes at higher geographies are significant and interesting before embarking on the greater effort needed at smaller spatial scales.

Acknowledgements

The research was funded by the Joint Information Systems Committee (JISC) of the Higher Education and Further Education Funding Councils in two awards to the University of Leeds: (1) the CHCC project (Collection of Historical and Contemporary Censuses and related materials), 2001–2003, and (2) the DIALOG-PLUS project (Digital Libraries in Support of Innovative Approaches to Learning and Teaching in Geography), 2003–2006, which is part of the Digital Libraries in the Classroom Programme. People who provided particular help on the projects were Linda See, Myles Gould and Simon Corne (CHCC). Andy Nelson (DIALOG-PLUS) was the main developer of the web pages. Help in using CommonGIS was provided by Manchester Computing through Jackie Carter, Manager of the CHCC project. The census data used in the project are Crown Copyright and produced by the Office for National Statistics and the General Register Office Scotland. The data were extracted from the UK Data Archive holdings (1971 and 1981 censuses) and the MIMAS service's CASWEB system (1991 and 2001 censuses). Danny Dorling is currently supported in his research by a British Academy Research Leave Fellowship.

Note

1 The census organizations worry about the threat to confidentiality posed by geographical differencing of the har-

monized areas with previously published census output areas. However, Duke-Williams and Rees (1998) showed that the risks of deriving statistics for sub-threshold areas were very small and NISRA (2005) have used database analysis to show that the risks of publishing both output area statistics and grid square statistics were minimal.

References

- Atkins D, Charlton M, Dorling D and Wymer C** 1993 Connecting the 1981 and 1991 censuses NE.RRL Research report 93/9
- Cole K and Hayes J** 2001 CASWEB 2 – An enhanced web-based interface to census area statistics: final report and demonstration ESRC/JISC 2001 Census Development Programme Fourth Workshop, Leeds, UK
- Dorling D** 1991 The visualisation of spatial social structure Unpublished PhD Department of Geography, University of Newcastle upon Tyne
- Dorling D** 1993 Map design for census mapping *Cartographic Journal* 30 167–83
- Dorling D** 1995 Visualizing changing social-structure from a census *Environment and Planning A* 27 353–78
- Dorling D and Atkins D** 1995 *Population density, change and concentration in Great Britain 1971, 1981 and 1991* Studies on medical and population subjects 58 HMSO, London
- Dorling D, Mitchell R and Martin D** 2001 Linking censuses through time – easy access to data 1971–2001: final report ESRC/JISC 2001 Census Development Programme Fourth Workshop, Leeds, UK
- Duke-Williams O and Rees P H** 1998 Can census offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure *International Journal of Geographical Information Systems* 12 579–605
- Martin D, Dorling D and Mitchell R** 2002 Linking censuses through time: problems and solutions *Area* 34 82–91
- Mitchell R, Dorling D, Martin D and Simpson L** 2002 Bringing the missing million home: correcting the 1991 small area statistics for undercount *Environment and Planning A* 34 1021–35
- NISRA** 2005 Personal communication, February Northern Ireland Statistics and Research Agency, Belfast