8. DIFFUSION BASED MAPPING TECHNIQUES: SEEING WEST MIDLANDS DEMOGRAPHY AND HEALTH IN A NEW WAY

8.1 Summary

The presentation of certain types of information on maps where shape and size is defined solely by land surface area has the potential to mislead the viewer, particularly where very different population densities exist. International and national work has demonstrated the value of density equalizing cartograms in correcting for this visual bias whilst still retaining immediate accessibility to the untrained viewer.

We discuss the development of a West Midlands density equalizing projection for local authority areas, and demonstrate variations of this method for common demographic variables and determinants of health. We provide a description of the method with the intention of encouraging local use of similar techniques, and further innovation in the geographical display of data for the West Midlands.

8.2 Introduction

Geographic Information System (GIS) techniques have become established as valuable tools in the communication of place-anchored data for both health and social care datasets. Providing sound principles are followed (Figure 8.1) large volumes of data can be concisely presented, made visually accessible, and used to identify differences and inequalities associated to place.

Figure 8.1: Relevant principles of good GIS usage

1. **Accurate raw data, including precise geolocation in standardised format**
2. **Appropriate choice of projection and mapping resources**
3. **Intelligent selection of class limits to categorise a colour representation**
4. **Selection of appropriate colour progression for continuous and discreet variable outcomes**
5. **Sufficient but not excessive groupings to achieve meaningful display whilst protecting individuals**

Health intelligence specialists in the United Kingdom have access to high quality, and freely available, raster and vector layers for use in producing GIS analysis, and maps now form a common part of many health intelligence products.

However, the complex geography of the United Kingdom, with areas of very high population density, interspersed by areas of low population rural landscapes, makes the fair presentation of spatial data difficult. As a familiar example, electoral maps of the United Kingdom and the United States often struggle to show apportionment of political preference, with rural areas that cover a large area implying one political persuasion dominates over an alternate party more favoured by the much smaller area inner city constituencies, when in fact voting numbers are much more balanced.

These problems also occur in the presentation of healthcare data. If raw numbers are presented on maps, the effect of population density is ignored, and larger areas receive undue attention. If standardised rates are used to produce choropleths (seeking to correct for variation attributed to known factors) the greater physical area of the less populated rural areas can lead the viewer to miss the relevance of the urban area rates which, although covering a much smaller area of the map, represent a far greater number of individuals experiencing that health outcomes. With either solution, comparison of rural and urban areas is difficult.

Cartograms are maps where the area of the shapes are not constrained or defined by land-surface area or exact geographical position and can be used in one form to standardise for the effect of density, whilst allowing the observer to more accurately weigh the information presented. For the production of these density-weighted cartograms, the relative size of different map areas is adjusted by a variable, allowing the shape itself to become the representation of this variable. In its simplest form, this can be an adjustment by population, allowing the map to become a representation not of surface area, but of population density. When applied well in healthcare, this has a particular value in expanding centres of dense urban population, to allow them to be fairly compared against their surrounding rural areas.

Developments in this method have been in progress for many years.[1,2] Very early examples made minimal attempt to maintain coherence to geography, or were based on their creators' subjective impression of how each piece of geography should be sized and arranged. Subsequent methods became more systematic, but relied on complex manual approaches, repeated over many iterations to produce a finished product. The advent of advanced computational techniques, and computers with sufficient processing power has greatly enhanced the production of cartograms, but there is an inherent tension between achieving a meaningful transformation whilst keeping a visually relevant display.

Dorling, one of the recent leaders in developing the cartogram method, described the ideal cartogram method, and this highlights the tension in producing a ‘good’ map; his recommendations are seen in Figure 8.2.

Dorling undertook work to attempt to meet these criteria, summarised initially in 1996.[3] It not only provides a historical overview of the development of the modern cartogram, but also developed the ‘cellular automata’ method, to produce more visually relevant displays of information. Dorling’s plain language description of this approach is to divide a map up into cells, and then iterate a process that allows ‘regions represented by too few cells … to gain cells from regions represented by too many’ until a steady state was reached.

This work remains under copyright, but the effect of this transformation on Great Britain can be seen in the paper by Gastner Newman 2004,[4] or by visiting http://www.pnas.org/content/101/20/7499/F2.expansion.html

Dorling noted the challenge of reaching steady state without losing the familiarity of the base shapes, particularly at the corners of shapes, and in the context of the coastline or map boundary. He therefore also presented an alternative method, which discarded the attempt to maintain recognisable boundaries for individual geographies, and instead use a circle of proportional area, spatially positioned in a way to remain identifiable with knowledge of the background geography. If geographies were simply replaced with a proportional circle at their centroid, there would be overlap and gaps, so a process was used to achieve disambiguation between the circles, whilst retaining relative geography, and minimal separation. Local examples can be seen in Dorling’s ‘New Social Atlas of Britain’ [5] and a dynamic example of a similar method in action for US states can be seen at Provotis,[6] an open source java implementation of many mapping techniques.

Figure 8.2: Dorling recommendations on cartogram technique

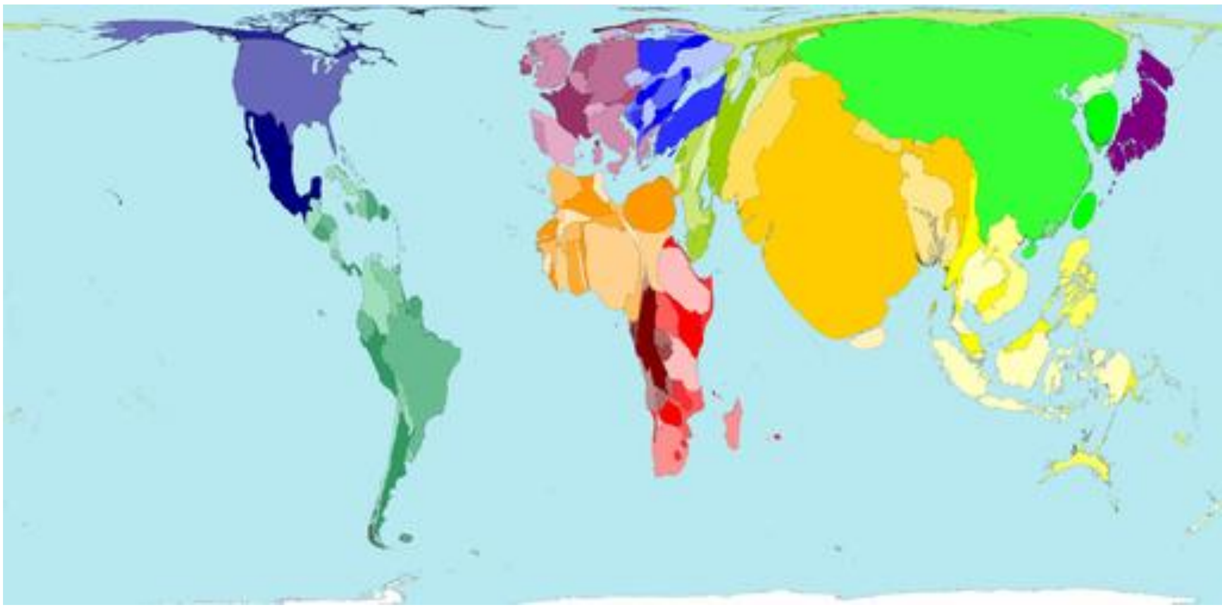
- **Be as simple and easy to understand and implement as possible**
- **Generate readable maps by minimizing the distortion of the shape of the geographical areas being mapped**
- **Preserves accuracy and maintains topological features**
- **Unambiguous**
- **Minimizes computational speed for the construction of new visualisations**
- **End result is independent of the initial projection being used**
- **End result looks aesthetically acceptable**
- **Has no overlapping regions or other more complex portrayal.**

Another key contribution to the recent rise in popularity of the cartogram is the work undertaken by Gastner and Newman.[4] Their work recognised the complexity of performing the transformation, and proposed a new mathematical theory to drive the shift. This involves the production of a 'diffusion cartogram', which uses a mathematical variation of a 'blur' on the population density to generate a flow of population from high density to low density areas. As population moves, it is allowed to take area and shape with it, whilst being 'floated' in a sea of surrounding uniform population to stop the map from equalising outwards from the margins. Gastner and Newman assert this method as a development as it allowed a more universal transformation of the map, as compared to Dorling's work which can consider the interplay of nearest neighbours only. (Other alternate solutions, including the work of Kocmoud and House,[7] have also been proposed)

Cartograms based on this method have been popularised in recent years through the 'Worldmapper' project, which offers global cartograms with geography modified by variables sourced from many different domains, including health, income, fuel, exploitation and food. Their base map, reproduced in Figure 8.3 shows a global projection modified to demonstrate population density. The Worldmapper project also produces country specific projections for population density, available through work from the University of Sheffield.[8] Closely allied to this work, and used in the production of the Worldmapper Cartograms, is the Gridded Population of the World produced by the NASA SocioEconomic Data and Applications Center (SEDAC) which provides international and national population synthetic estimates from available data sources.[9]

More recently, the Chief Medical Officer's report for England [10] used the Gastner Newman transformation based population density cartograms as the base layer for all mapping displays, as demonstrated in Figure 8.4.

Figure 8.3: Total population by country, 2002



The West Midlands has areas of very high and very low population density. Presentation of regional data therefore encounters the same problems as the national and international examples considered above. We therefore undertook work to develop a series of base geography cartograms for the West Midlands, whilst also seeking to identify a method that would be accessible and reproducible by local health intelligence teams for their own purposes, and stimulate future innovation in the presentation of West Midlands geography, demography and other health and social care outcomes.

8.3 Cartogram creation

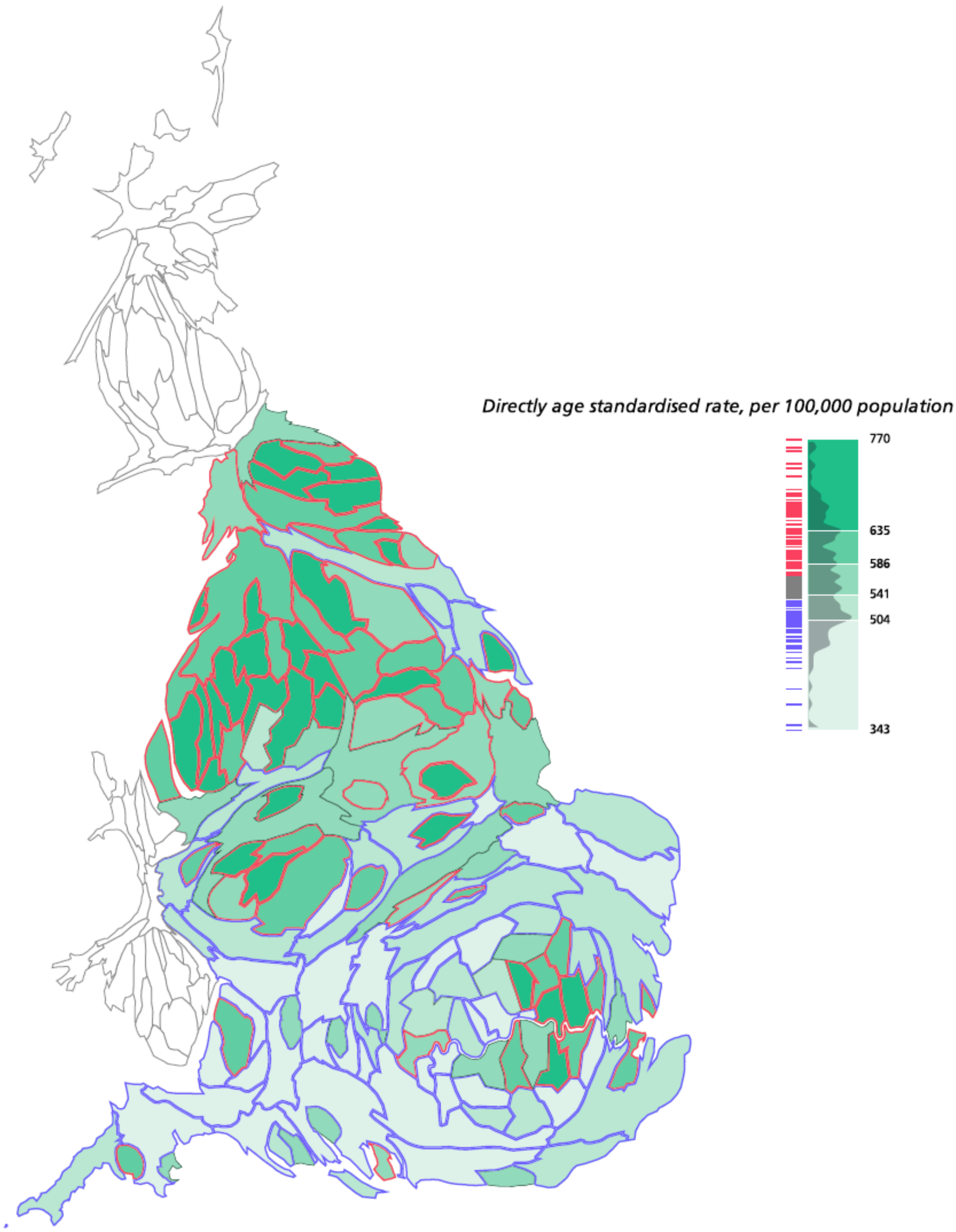
The cartograms were created using ArcMap 10 and an Environmental Systems Research Institute (ESRI) script written by Tom Gross[12] which replicates the Gastner and Newman methodology[4] and is available for free download.

Installation and utility information are detailed and proved very useful in setting up the correct layers to produce the required outputs. For example, cartograms must be generated from positive numeric attribute values and all outputs must be written to a geodatabase even though shapefiles can be used as inputs.

Production of cartograms relies on equal area based projection. The example data frame and data layers provided with the download is set up with projection GCS_WGS_1984. Experimentation with alternative projections such as, Albers conic, Lambert conformal conic, and Eckert IV produced good cartograms but their shapes were less appealing than those produced here using the default GCS_WGS_1984 projection.

The data frame and all layers were transformed to this projection. The polygon file for geographical boundaries was joined to the attribute dataset and then exported and re-imported as a new layer to the projection of the data frame.

Figure 8.4: Average annual mortality due to all causes by upper tier local authority, England, 2008-2010



Source: Death registrations and 2008 to 2010 population estimates, ONS. (Analysis by DH)

Excerpt from Report of the Chief Medical Officer [10]

The cartograms were produced using population data and additional attribute data has been displayed thematically. Data has been classified as five equal counts of areas, quintiles. This method does not account for the variation and distribution of the data itself therefore careful attention should be given to the thresholds when interpreting the data.

8.4 Alternate methods

After experimentation with stock data provided by the tool, we found the ArcMap tool relatively easy and reliable to use. However, alternate solutions do exist. For users of MapInfo, and other GIS software, it would be possible to create cartograms using the free software application, ScapeToad.[13] A shapefile can be easily imported into this standalone application and although the production of the final cartogram takes some time and the default legend needs alteration, the final output is a good cartogram. ScapeToad also produces a set of statistics by which the accuracy and appropriateness of the final cartogram can be reviewed, a feature requiring manual calculations in ArcMap. ScapeToad was used in the production of the report of the Chief Medical Officer. [10]

We also identified the plug-in ‘Cartogram Creator’ available for use with the open source Quantum GIS software, downloadable from within the QGIS software. This plug-in also performed well, generating credible cartograms in relative short time frames. However, it lacks documentation, particularly around the selection of number of iterations required to achieve appropriate transformation.

8.5 West Midlands cartograms

Figure 8.5: West Midlands local authority boundaries projected with British National grid

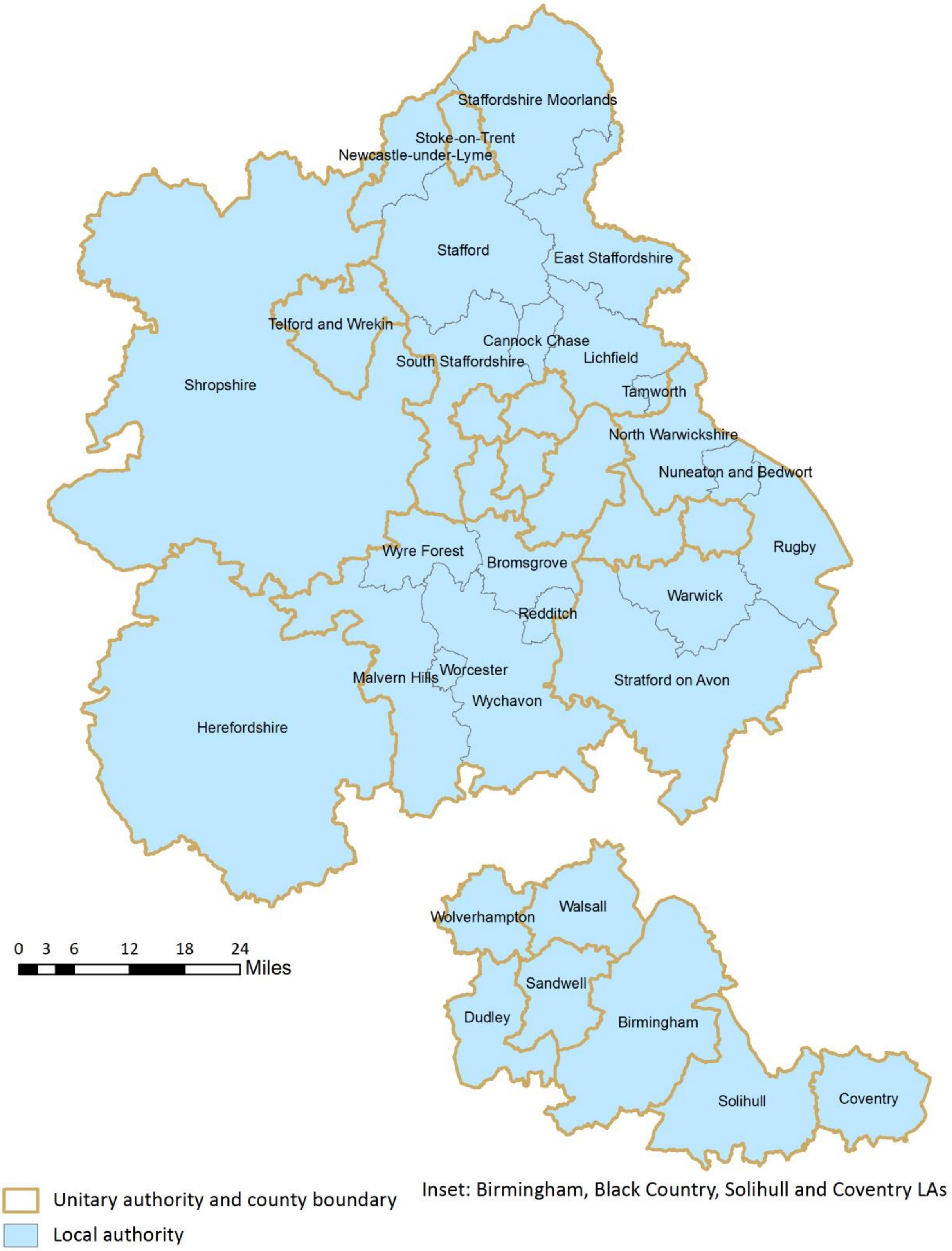
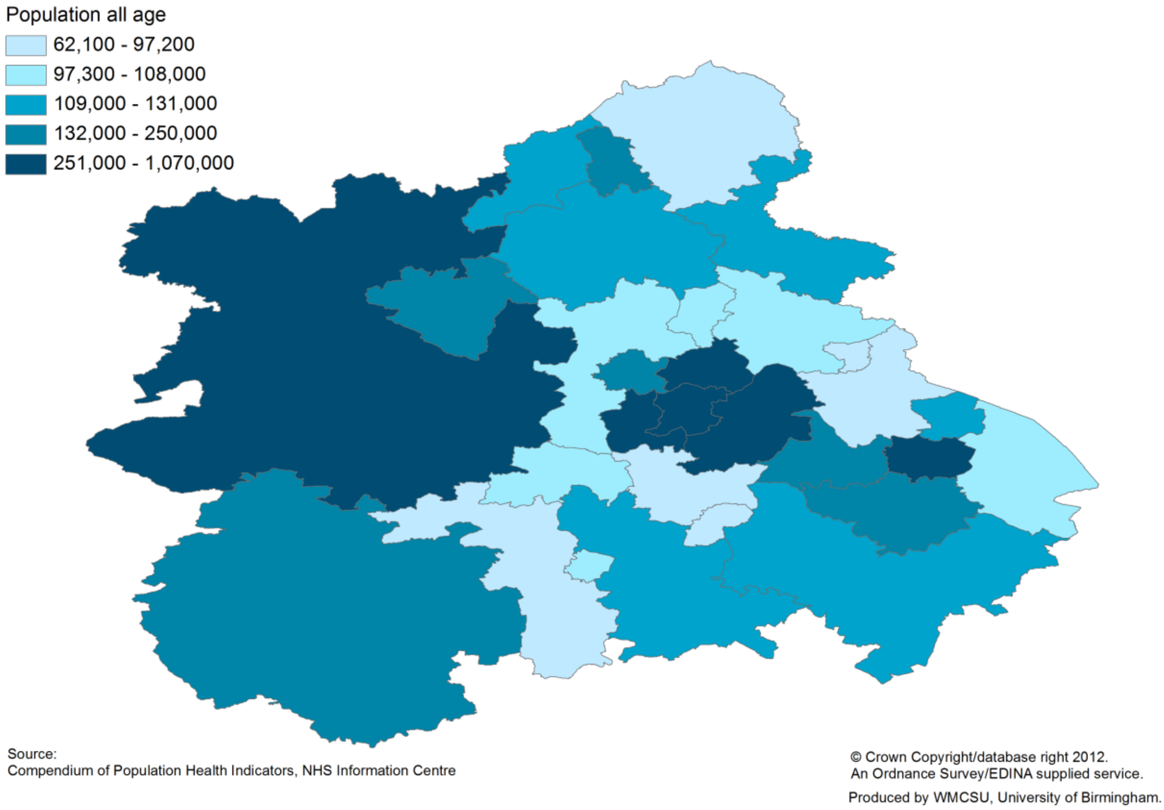


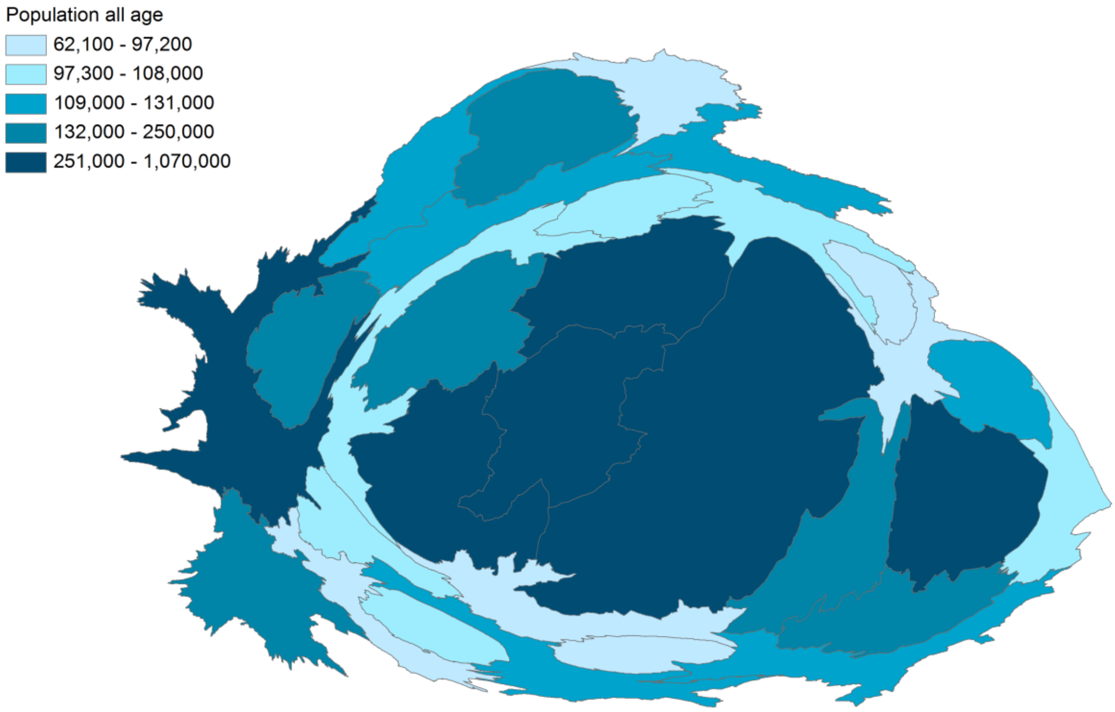
Figure 8.6: Base geography: Midyear population estimates by local authority, all ages, West Midlands 2011.



The choropleth base map of the West Midlands (Figure 8.6) demonstrates the complexity of interpreting information using land area based boundaries. The urban conurbation of Birmingham and the Black Country at the centre of the region is coloured to show a high population, as is Coventry. However, Shropshire is also shaded to represent a similar population. An observer, without additional context, has no means of identifying where the population density lies. If population count were to be replaced with a health outcome, it is easy to see that the relatively small area of the central population could have its outcomes under represented on this map.

We used this projection as the base map to generate our population cartograms.

Figure 8.7: Population cartogram sized by: Midyear population estimates by local authority, all ages, West Midlands 2011

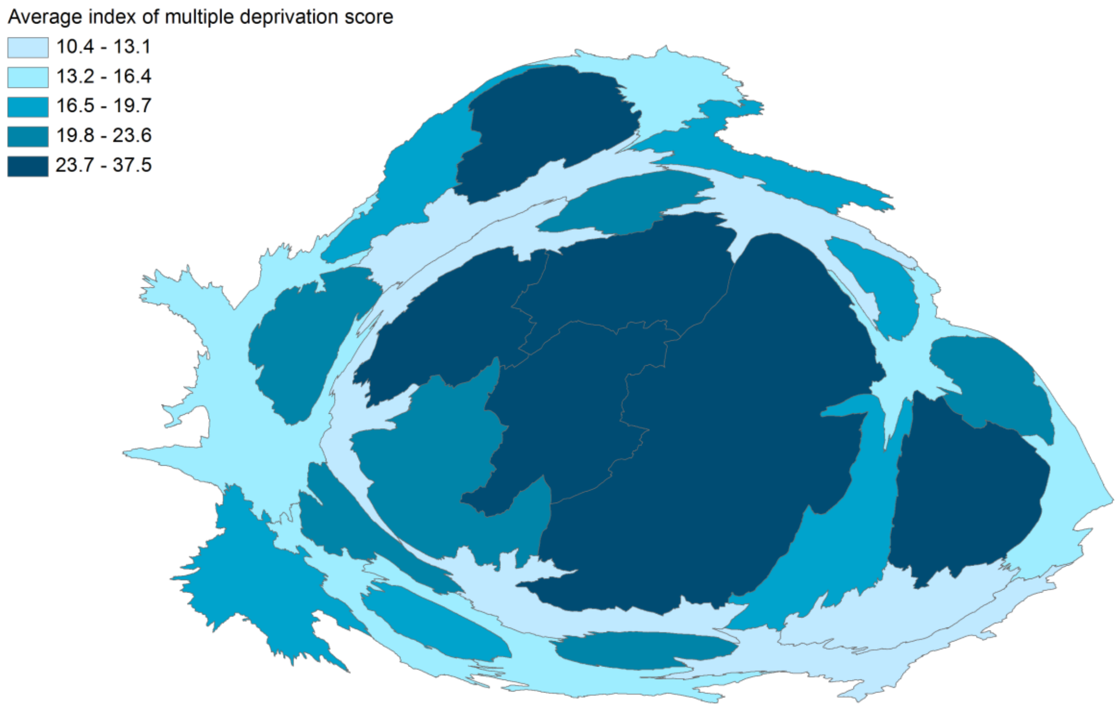


Source:
Compendium of Population Health Indicators, NHS Information Centre

Produced by WMCSU, University of Birmingham.

The population cartogram in Figure 8.7 demonstrates the base transformation applied to the West Midlands. The dense urban population has been used to expand shapes and give these areas greater significance. Shropshire has been reduced in size, but retains its relative position, and a surface area that still fairly represents its total population. The significance of the population of Coventry is now more clearly identifiable. The technique has also coped well with the other significant population centres, including Stoke, which although not in the highest population category, is now expanded to be more fairly visible.

Figure 8.8: Population cartogram sized by: Midyear population estimates by local authority, all ages, West Midlands 2011. Thematically representing: Index of multiple deprivation 2010 by local authority, West Midlands.



Source:
Compendium of Population Health Indicators, NHS Information Centre
Department for Communities and Local Government,
Indices of Deprivation 2010

Produced by WMCSU, University of Birmingham.

Using our base cartogram, we display in Figure 8.8 the IMD scores for each of the local authority regions. The significance of the high scores in the dense population regions would be visually underrepresented on a standard map, but here, the proportion of the West Midlands population exposed to high deprivation is fairly represented, and immediately obvious to the observer.

Figure 8.9: Population cartogram sized by: Midyear population estimates by local authority, all ages, West Midlands 2011. Thematically representing: Percentage population change from midyear 2011 to 2021

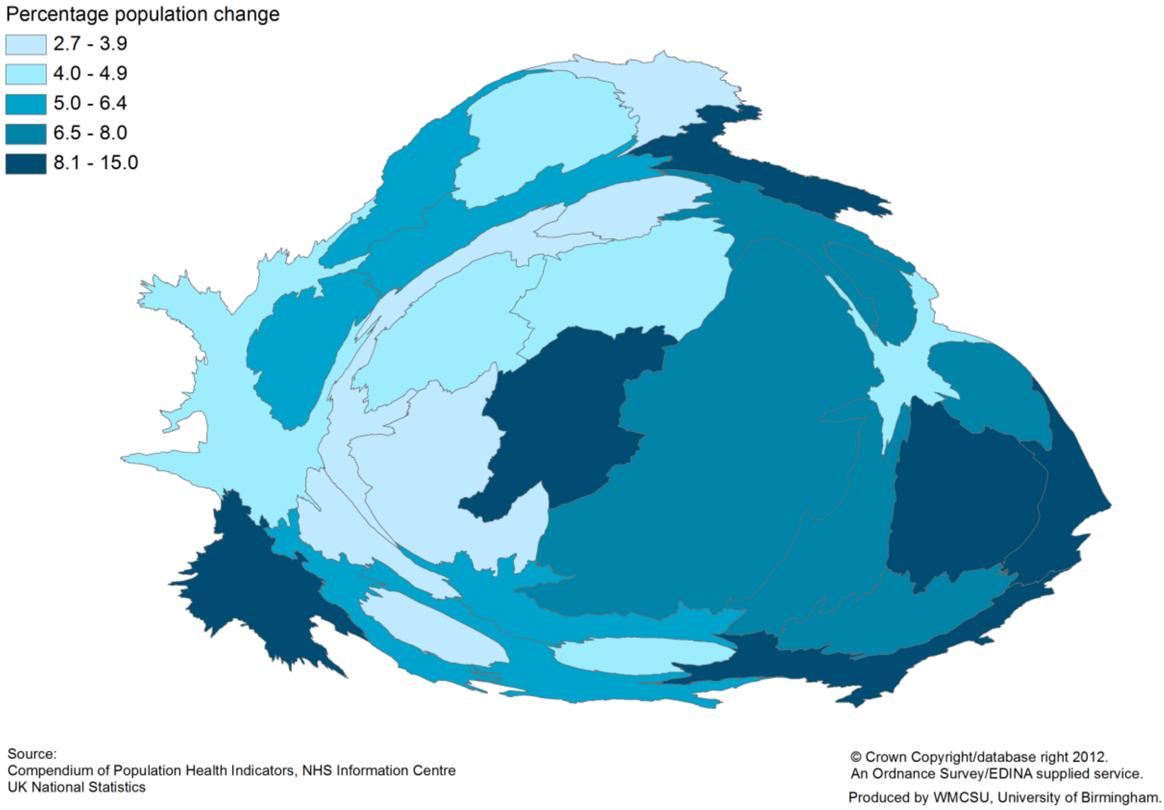
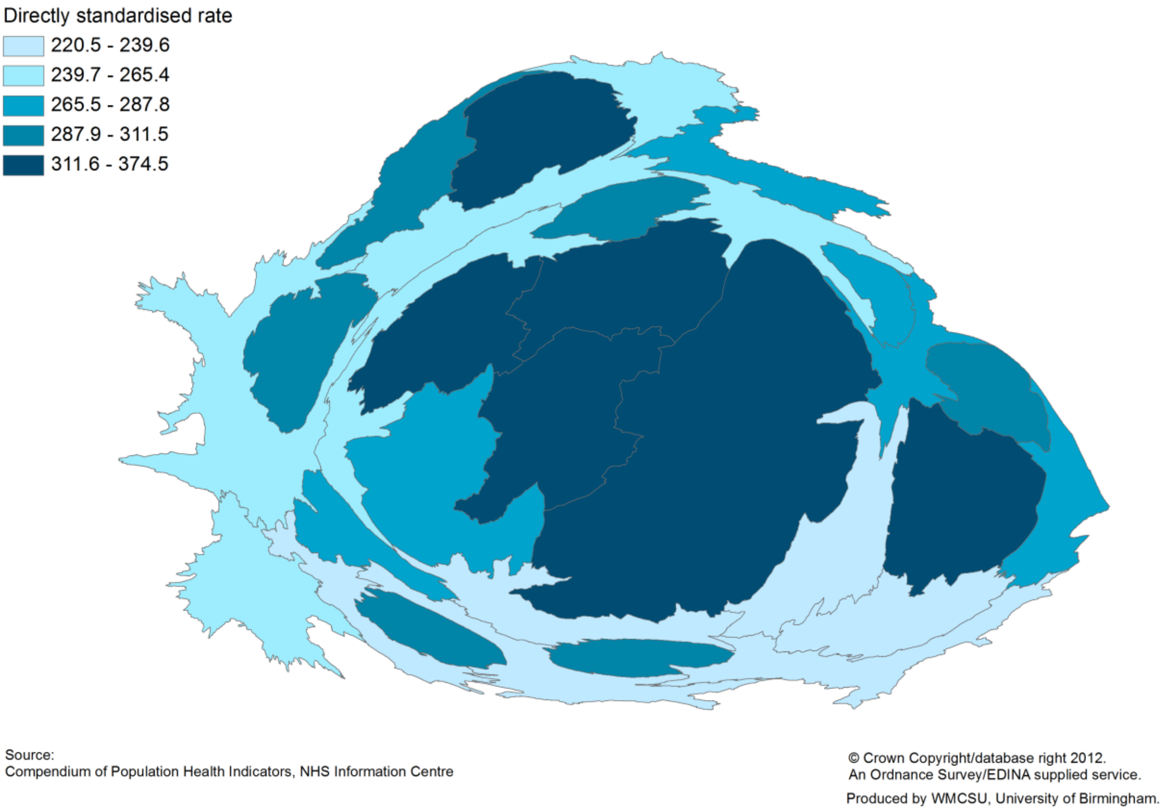


Figure 8.9 presents the forecast population change across the West Midlands between 2011 and 2021. Again, the population based cartogram allows the relative importance of this to be considered against a fair background of population density. For example, the high percentage forecast change in Herefordshire, if presented on a standard projection, could be over-represented against the equally significant changes forecast in some of the urban centres. The impact of a moderate change in a highly populous area is also more easily appreciated on this map.

Figure 8.10: Population cartogram sized by: Midyear population estimates by local authority, under 75, West Midlands 2011. Thematically representing: Mortality from all causes: directly standardised rate per 100,000 population by local authority, under 75, West Midlands 2008-10



The population based cartograms for all cause mortality in the under 75s presented in Figure 8.10, and the discreet male and female life expectancy cartograms presented in Figure 8.11 and Figure 8.12, demonstrate an important principle in the application of the transformation method. The base population used to transform the geography must be appropriate for the context being presented. In the life expectancy example, it is important that the transformation be driven by the male population only for Figure 8.11 and the female population only for Figure 8.12. Although the final transformations in these examples are almost equivalent, there are situations where this would not be the case, and the authors have noted examples of this mistake in well recognised literature.

Figure 8.11: Population cartogram sized by: Midyear population estimates by local authority, all ages, males, West Midlands 2011. Thematically representing: Male life expectancy at birth by local authority, West Midlands 2008-10.

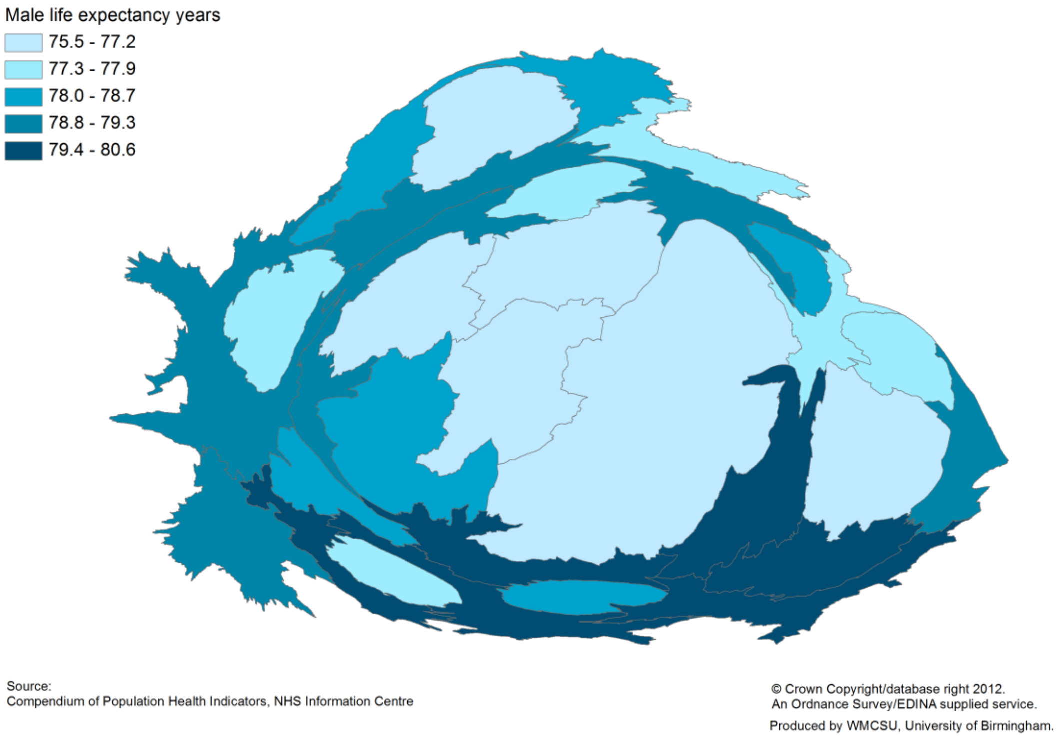


Figure 8.12: Population cartogram sized by: Midyear population estimates by local authority, all ages, females, West Midlands 2011. Thematically representing: Female life expectancy at birth by local authority, West Midlands 2008-10.

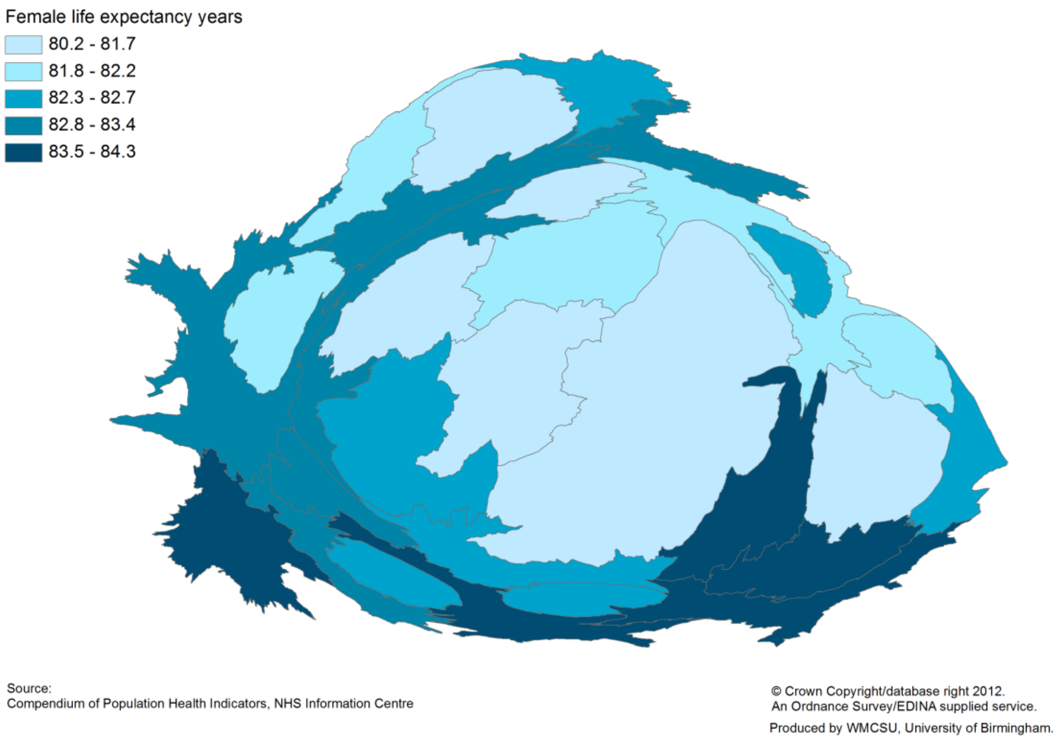
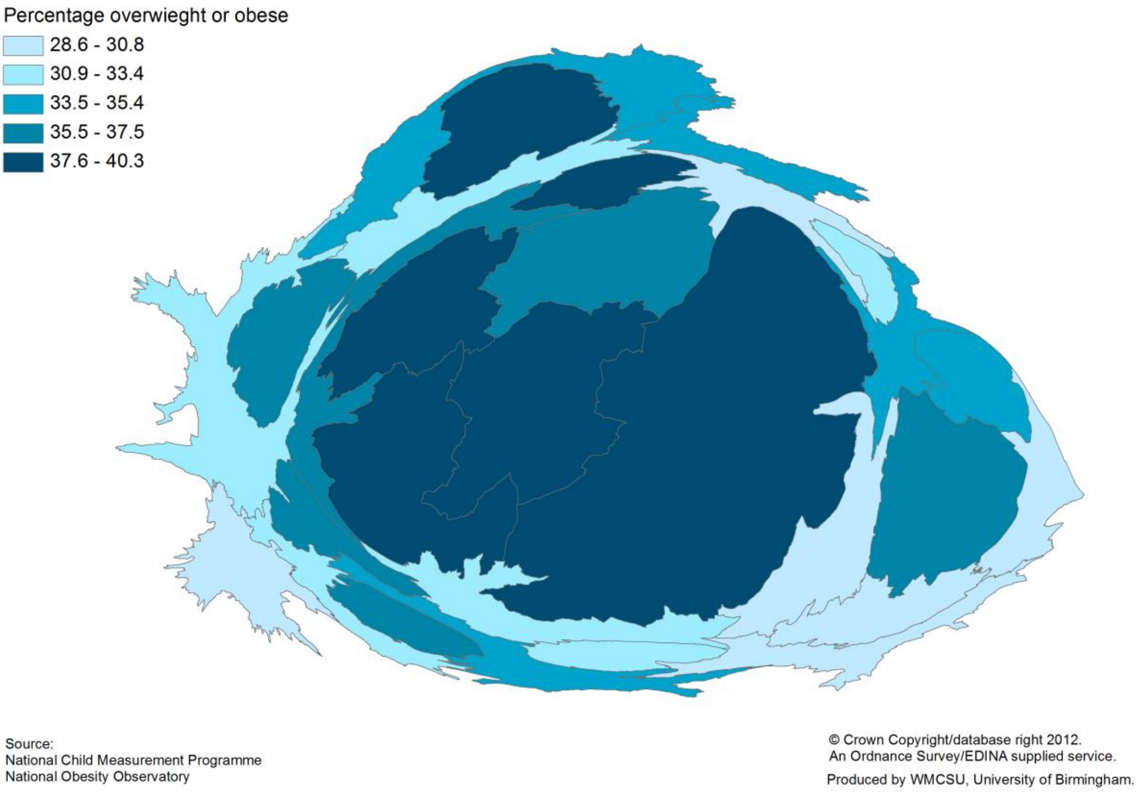


Figure 8.13: Obesity cartogram sized by: Number of children aged 10-11 overweight and obese 2011-12. Thematically representing: Percentage of children aged 10-11 overweight and obese by local authority, West Midlands, 2011-12.



The value used to drive the transformation does not necessarily need to be an overall population density function. In Figure 8.13 we used the number of overweight or obese year 6 children to make the transformation, to allow a transformation that represents the distribution of these individuals across the West Midlands, and highlighting the significant need in the urban centres.

Individual projects may demand different transformations, depending on the population group or outcome of interest.

8.6 Discussion and application

Our transformed maps allow the health outcomes of the more densely populated areas to be more fairly visualised. We established a new transformation for the West Midlands that standardises for the visual bias due to geographical area, and demonstrated the relevance of this in the display of common demographic and health variables. Appropriate use of this method should assist in regional comparisons, and has potential to also be used in smaller area geographies.

Despite the increasing popularity of value based cartograms, we found it difficult to identify a standardised method for the production of local context maps. We tested several software plug-ins, with very variable results. The approximately circular nature of the West Midlands, with a central density of population has 'squeezed' the physically peripheral local authority regions, creating greater distortion at the boundaries, and although the physical area of these shapes is representative, the ribbon effect of being spread at the boundaries has potential to visually over-moderate the significance of these regions. Sandwell's central physical location within the boundary shape file has created what the authors have locally termed the 'Mappa Mundi' effect, where a centrally positioned shape tends to be preserved, giving the impression of its co-incidental representation of the centre of the world!

We were unable to identify a viable and locally reproducible bubble cartogram method, which has potential to avoid this failure.

In the context of Dorling's recommendations for the ideal cartogram, we suggest our projections;

- ✓ Are simple and relatively easy to implement
- ✓ Generate maps that do preserve the majority of the West Midlands relationships
- – Have variable accuracy to topological features, with poor performance at the margins
- ✓ Have a single processing step, avoiding ambiguity compared to other software
- ✓ Can be performed on standard office computers
- ✘ Vary aesthetically in output based on initial projection chosen
- ✓ Have no overlapping regions

Our final solution is clearly only one part of developing more intelligent representations of geographical information for the West Midlands, but we believe it has demonstrated there is value in developing techniques for more intelligent data display.

8.7 Recommendations

In considering new mapping techniques, we make the following recommendations based on our own learning experience;

1. The use of cartograms should be considered wherever information mapping covers both densely and lightly populated regions, and where small areas physical area may mask greater need.
2. Rigorous methodology should be used in performing the transformations, particularly in the selection of projection, numerator, denominator and transformation technique.

3. The increasing use of cartograms in international and national publications requires that end users should be familiar with appraising cartograms, particularly in establishing if the presented information is a fair representation of the underlying information.

8.8 Further information

We commend to interested readers several sources of information and examples which inspired our development of alternative mapping solutions;

1. **How to Make Area Cartogram Maps in ArcGIS (Dempsey C, GIS Lounge).** Walkthrough including common pitfalls, in generating cartograms within ArcGIS.
   Accessible at: http://www.gislounge.com/how-to-make-area-cartogram-maps-in-arcgis/

2. **Scapetoad (Choros)**. As well as providing freeware javascript transformation, includes references and links to developing cartogram techniques.
   Accessible at: http://scapetoad.choros.ch/index.php

3. **Indiemaps (Johnson Z).** Blog covering variety of techniques and tools for improving data visualisation.
   Accessible at: http://indiemaps.com/blog/

4. **Improving Visualisation (Dataviz).** Commissioned by the Department of Communities and Local Government and carried out by Oxford Consultants for Social Inclusion. Provides diverse examples of data visualisation relevant to the public sector at all levels, including case studies.
   Accessible at: http://www.improving-visualisation.org/

5. **Geographer At Large (Anon).** Blog posts on different mapping techniques, particularly in the context of New York. Quote: “Geography invariably leads to revolution”
   Accessible at: http://geographer-at-large.blogspot.co.uk/2011_10_01_archive.html

6. **Gapminder World (Gapminder Foundation)**. Non profit venture promoting sustainable global development, with an interest in innovative data display.
   Accessible at: http://www.gapminder.org/

7. **MAPresso (Herzog A).** Downloadable Java applet that can be used to generate different cartogram formats, including ‘Dorling Cartograms’ on local shape files and data.
   Accessible at: http://www.mapresso.com/

8.9 References

1 Johnson Z. Early cartograms. [Internet] 2008 [Cited 24th February 2013]. Available at: http://indiemaps.com/blog/2008/12/early-cartograms/

2 Tobler W. Thirty Five Years of Computer Cartograms. 2004. *Annals of the Association of American Geographers* 94(1)58-73

3 Dorling D. *Area Cartograms: Their Use and Creation. 1996. Concepts and Techniques in Modern Geography* series no. 59, University of East Anglia: Environmental Publications.

4 Gastner M Newman M. Diffusion based method for producing density-equalizing maps. PNAS. May 18, 2004. 101 (20): 7499-7504

5 Dorling D. *A New Social Atlas of Britain.* London: John Wiley and Sons [Internet] 1995. [Cited 24th February 2013] Available at: http://sasi.group.shef.ac.uk/publications/new_social_atlas/index.html

6 Protovis: Dorling Cartograms [Internet] 2012 [Cited 24th February 2013] Available at: http://mbostock.github.com/protovis/ex/cartogram.html

7 Kocmoud C, House D. Constructing Continous Cartograms: A Constraint based approach. 1998. *8th international symposium on Spatial Data Handling,* Vol. 1:236-246

8 Hennig BD. *The Human Shape of Britain* [Internet] 2012 [Cited 24th February 2013] Available at: http://www.viewsoftheworld.net/?p=2233

9 Center for International Earth Science Information Network (CIESIN)/Columbia University, International Food Policy Research Institute (IFPRI), The World Bank, and Centro Internacional de Agricultura Tropical (CIAT). *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Density Grid.* [Internet] 2011 [Cited 24th February 2013] Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). Available at: http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-density

10 Davies SC. *Annual Report of the Chief Medical Officer, Volume One, 2011, On the State of the Public's Health.* London: Department of Health, 2012.

11 Dorling D. *Worldmapper Dataset 002: Total Population*. SASI, University of Sheffield. [Internet] 2nd May 2005 [Cited 24th February 2012]. Available at: http://sasi.group.shef.ac.uk/worldmapper/display.php?selected=2#

12 Gross, T. *Cartogram geoprocessing tool version 2*. New York: Environmental Systems Research Institute (2007, 2009). Available at: http://arcscripts.esri.com/details.asp?dbid=15638 (accessed 15/01/2013)

13 Andrieu D, Kaiser C, Ourednik A. *ScapeToad*. [Internet] 2008 [Cited 24th February 2012] Available at: http://scapetoad.choros.ch/index.php

8.10 Authors

Dr Christopher Chiswell, Specialty Registrar in Public Health, West Midlands Commissioning Support Unit, University of Birmingham

Jane Fletcher, Data analyst, West Midlands Commissioning Support Unit, University of Birmingham

Version 1.0
18 March 2013