

**QUANT
ITATIVE**

GEO **RICHARD
HARRIS**
**GRA
PHY**
THE BASICS

 **SAGE**

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne



Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/1 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Robert Rojek
Editorial assistant: Matt Oldfield
Production editor: Katherine Haw
Copyeditor: Kate Campbell
Proofreader: David Hemsley
Indexer: Cathryn Pritchard
Marketing manager: Sally Ransom
Cover design: Stephanie Guyaz
Typeset by: C&M Digitals (P) Ltd, Chennai, India
Printed and bound in Great Britain by Bell and
Bain Ltd, Glasgow

© Richard Harris 2016

First published 2016

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

Library of Congress Control Number: 2015958574

British Library Cataloguing in Publication data

A catalogue record for this book is available from
the British Library

ISBN 978-1-4462-9653-0
ISBN 978-1-4462-9654-7 (pbk)

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the PREPS grading system. We undertake an annual audit to monitor our sustainability.

8

MAPPING AND GIS

8.1 INTRODUCTION

The previous chapter looked at ways of presenting data but it neglected something important to geography's cartographic traditions: the mapping of geographic data. This chapter makes amends for that omission, concentrating on some of the principles behind Geographic Information Systems (GIS) – software for the storage, analysis and visualisation of geographic information.

Figure 8.1 sets the scene. It is a stylised map of London's boroughs, showing the ethnic composition of their resident population according to the 2011 Census. It would be possible to present these data in a table, and it would be better to do so if the exact population numbers were required. However, the map reveals something that the table would not: the ethnic geography of London – its geographical patterning. Despite reports about the declining number of people in London who regard themselves as White British, this group remains dominant in the capital. There are clusters of people of Asian Commonwealth heritage (Bangladeshi, Indian or Pakistani), especially in the west, northwest and east of London. Black groups (Black African, Black Caribbean or Black other) are found especially, but not exclusively, to the southeast of the centre. Most of the boroughs contain an ethnically diverse population, although that is less true of some at the outer boundary of London.

The map is a simplification of a complex reality. It is a cartographic product – a pictorial abstraction – but it is still linked to data that are the outcome of social, economic and demographic processes that are actually happening on the ground. Knowing where something occurs can be a first step towards understanding why it occurs or helping to manage its impact. GIS are technologies that help manage, make sense of and draw meaning from geographic data.

8.2 VIEWS OF THE WORLD

Figure 8.2 gives two different maps of the 2001 Census population of Northern Ireland. The first is a raster view; the second is vector. The differences between the

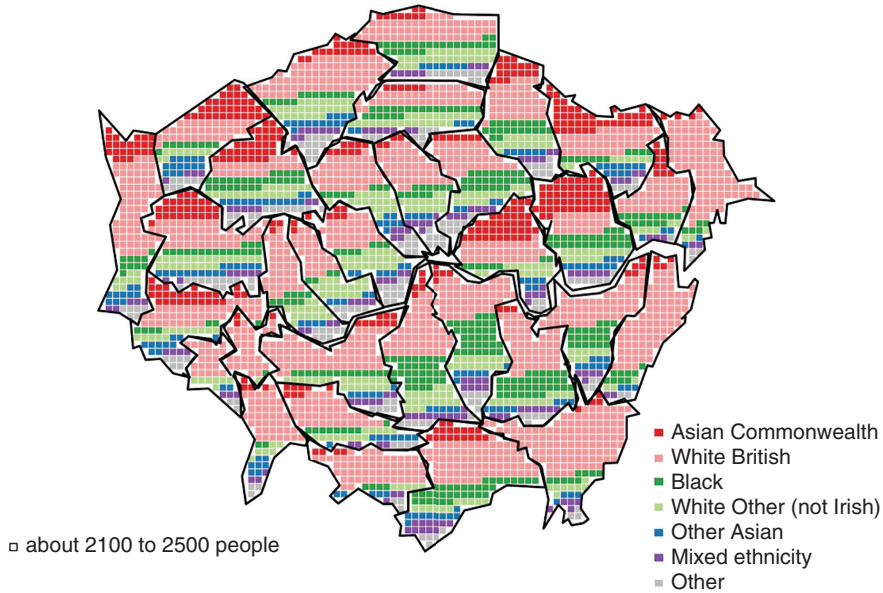


FIGURE 8.1 A map showing the ethnic composition of London's boroughs in 2011 (source of data: UK Census)

maps are more apparent in Figure 8.3, which zooms into the area in and around Belfast city centre. The blocky, grid structure of the raster map contrasts with the polygonal structure of the vector map.

Raster and vector are the primary data models used in GIS. These are the sorts of desktop software used to store, query, manipulate, analyse and display geographic data. Geographic data, also known as spatial data, can be mapped because they include a georeference identifying where an observation (a measurement) has been taken – a grid coordinate, or a longitude and latitude, for example. Geographic data combine what has been measured (the attribute data) with where it has been measured (the location or georeference). There are a number of well-known commercial GIS, and also free software including the popular QGIS package, available for Windows, Mac and Linux.¹ The content of this chapter was produced in R, the open source, statistical computing and graphical software mentioned at the end of the previous chapter. Its various spatial analytical libraries provide a powerful open source tool for geographical analysis and mapping.² Chris Brunsdon's and Lex Comber's book *An Introduction to R for Spatial Analysis & Mapping* (2015) offers what its title suggests; Bivand et al.'s *Applied Spatial Data Analysis with R* (2013) is a little more advanced.

¹<http://qgis.org/en/site/>

²<https://cran.r-project.org/web/views/Spatial.html>

MAPPING AND GIS

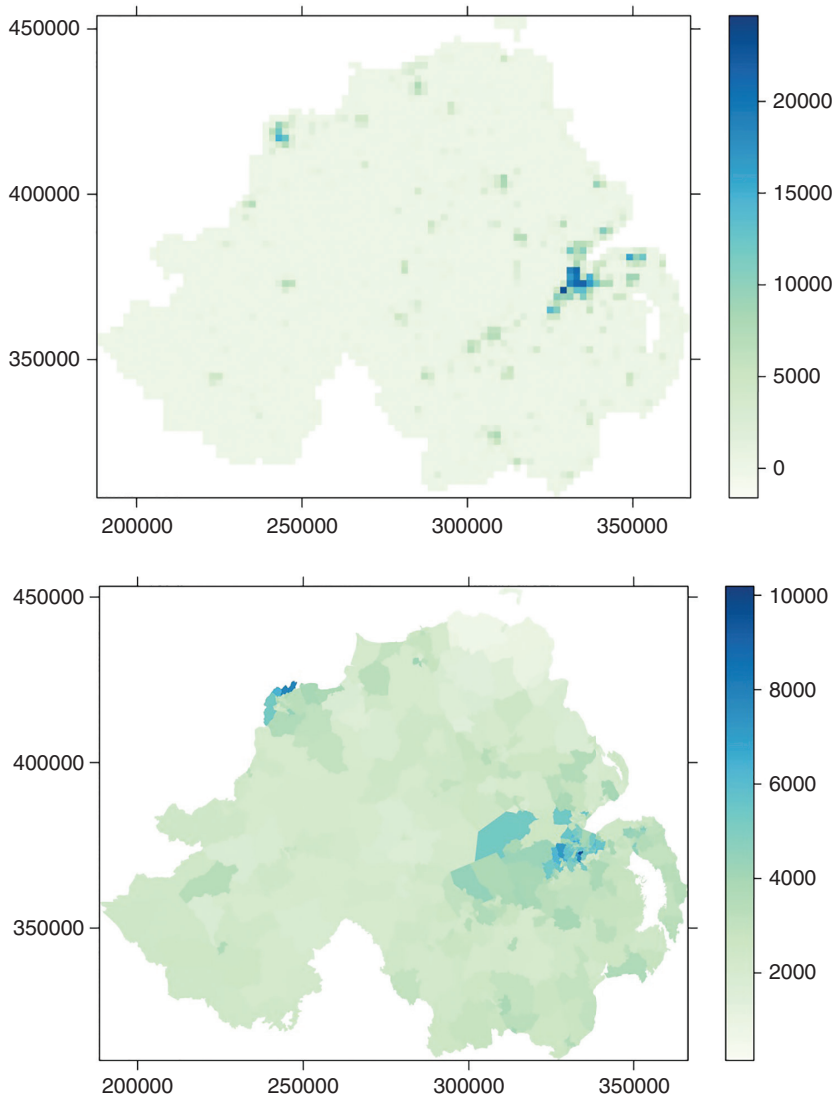


FIGURE 8.2 A raster representation of the Census population of Northern Ireland in 2001 (top). A vector representation (bottom) (source of data: <http://www.nisra.gov.uk/geography/home.htm>)

8.3 RASTER DATA

A raster is a grid structure. At its simplest, it is a series of values, one for each of the grid squares, also known as cells. The Northern Ireland map is a 73 row by 90 column grid, the product of which is 6570 cells and therefore 6570 values.

QUANTITATIVE GEOGRAPHY

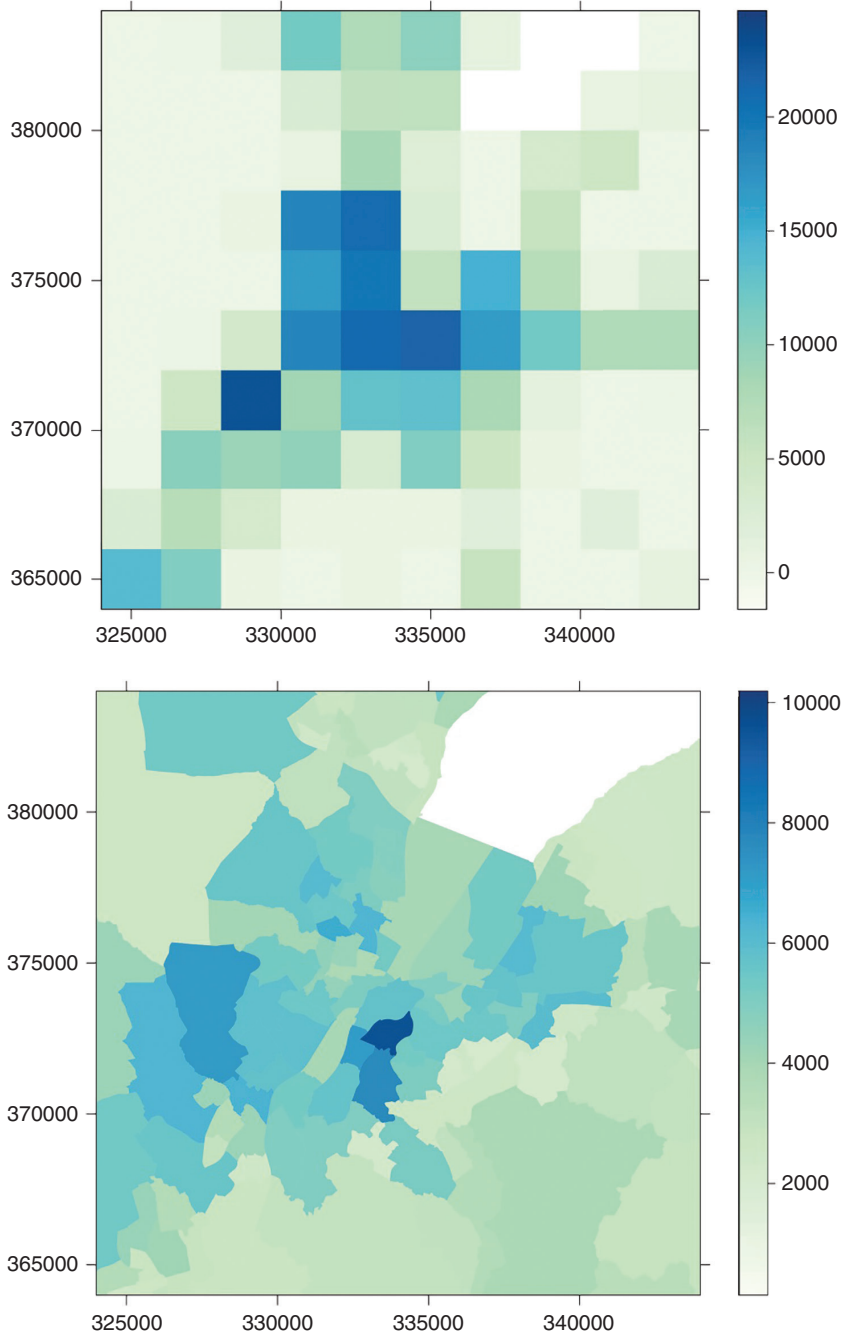


FIGURE 8.3 Showing the areas in and around the city of Belfast: in raster format (top); in vector format (bottom)

Not all of those cells record a population count, because they are offshore and in the sea. They could be assigned zero population, but it is better to say they contain no measurement at all. They will then appear unshaded in the maps and not distort any statistical calculation. There is an average of 444 persons per grid cell on the land, but that average drops to 257 if the offshore cells are taken as zero, giving a misleading indication of the population density. Therefore, in addition to the measured values, an NA (not applicable) value is required to indicate an empty cell.³

To map the data, the locational information is required. The geography of the grid is defined by assigning a geographical coordinate to one corner and combining that with knowledge of the number of rows (*nrow*), the number of columns (*ncol*), and the length/width of each grid cell. In addition, it may be necessary to include a coordinate reference system (CRS), especially for a raster covering a large area of the Earth's curved surface. The CRS will identify the ellipsoid used to model the shape of the Earth, the datum defining the origin ((0,0) point) of the coordinate system, and the projection required to flatten the globe and draw it on a map.

The area of each grid cell is the cell size and defines the resolution of the grid. The raster shown in Figures 8.2 and 8.3 is of size 2 km × 2 km, or 4 km². Decreasing the cell size increases the resolution of a grid, decreasing the 'blockiness' of the raster and the appearance of pixelation. However, it increases the size of the file and therefore the computational demands. Halving the cell size halves the length and also the width of each cell, quadrupling the number of cells. If each value requires 2–8 bytes of memory, then high-resolution rasters covering a large study area will quickly become voluminous. Figure 8.4 is a digital elevation model (DEM) showing the height of the Rocky Mountains from just north of the Canadian–US border (top left of the image), down to New Mexico. The heights are represented by a grid of 8.64 million cells, with a file size of about 17.3 MB. It can also be downloaded as a grid of over 138 million cells and a file size of 277 MB.⁴ In either case it is only a small extract from the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010).

The raster values can be stored in an *nrow* by *ncol* matrix. However, where there are many zero or, alternatively, many NA values, the matrix becomes sparse of other numbers. There may then be more efficient ways of storing the data, including methods of data compression. Some methods of compression are lossless, in that

³The same arguments could be extended to Lough Neagh to the west of Belfast and to Lower and Upper Lough Erne near the western borders with the Republic of Ireland, indeed to any 'empty' land including the various mountain ranges. How we classify the data will affect the analysis that follows on from that classification; uncertainty in the classification will create uncertainty in the results.

⁴From <http://earthexplorer.usgs.gov/>

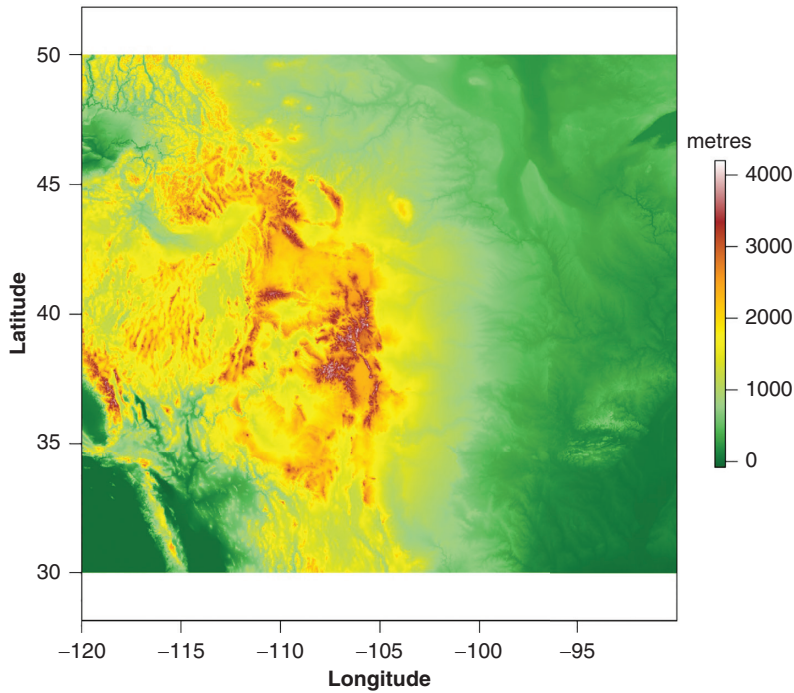


FIGURE 8.4 Digital elevation model showing the height of the Rocky Mountains through the United States (source: GMTED2010)

they reduce the file size but the process can be reversed to recover the original values. Others are lossy, in that they simplify but also degrade the image quality. The image file format, TIFF, may be used to store raster images, with or without compression.⁵ A GeoTIFF allows a CRS to be embedded in the TIFF.

8.4 EXAMPLES OF RASTER ANALYSIS

Cell aggregation

Figure 8.5 shows some typical raster processes. The first is a simple aggregation of the data – in this case, the 2 km × 2 km grid cells of Northern Ireland’s population are grouped into larger, 20 km × 20 km squares. The length of the larger squares is ten times the original grid cells, giving 10 × 10 grid cells per new square. Summing together the values for those 100 cells gives the total population count for the larger cell. Other functions could be applied during the aggregation, to find the minimum, maximum, mean or median values, for example.

⁵JPEG files usually are smaller, but employ lossy compression.

Cell reclassification

Figure 8.5b illustrates an even simpler raster operation: a reclassification of the original grid values. Suppose all cells with a population count of 400 or greater are defined as built-up areas/urban settlements. The raster can be then reclassified such that $z < 400 \rightarrow 0$ and $z \geq 400 \rightarrow 1$, where z is the original raster value.

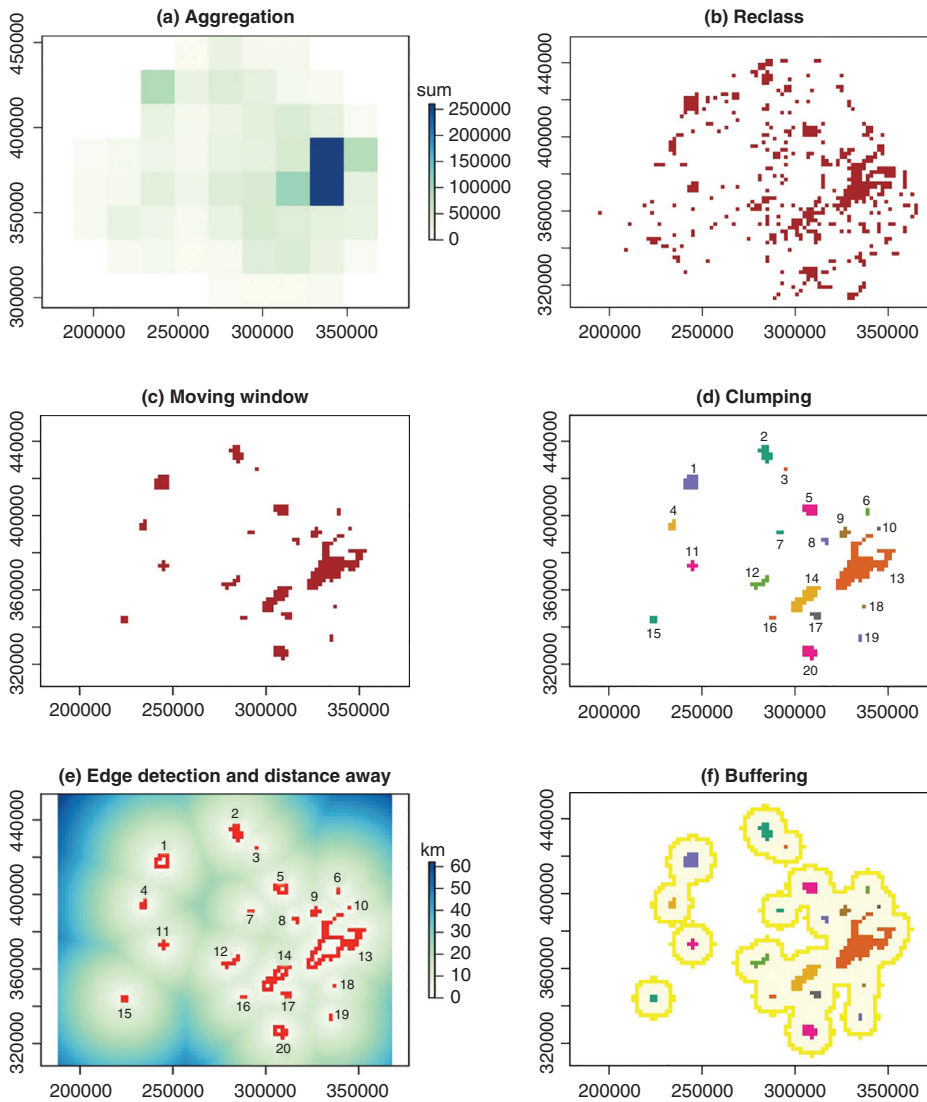


FIGURE 8.5 Examples of raster methods of analysis

The data are now categorical: the values of 0 and 1 have no quantitative meaning other than to distinguish one category (low population density/rural) from the other (higher population density/urban). Sometimes categorical data are described as qualitative because the numbers are arbitrary and simply a substitute for words (0 = rural, 1 = urban). Personally, I try and avoid it: the data are not qualitative in the sense that many in the social sciences and humanities would mean it, where focus groups, discourse analysis, ethnography, participant observation and other non-quantitative approaches are adopted.

Moving windows

Figure 8.5c illustrates a moving window operation, also known as a focal operation because it focuses (centres) on each grid cell in turn. Figure 8.6 elaborates upon the process. In this example, a 3 × 3 moving window is centred upon one of the grid cells. The number at the centre is then substituted in a new grid with the median of the nine values within the window of the old. The process is repeated, centring the window on each of the other cells in turn. The result is one of spatial smoothing – it removes the speckle in the map. Increasing the size of the window increases the amount of smoothing. Functions other than the median may be employed. Methods similar to moving windows are used in creating geographically weighted statistics (see Chapter 11). A potential problem is one of edge effects. Along any of the sides, the window extends beyond the study region, with a maximum of six cells contained within it and four at the corners.

Clumping

Figure 8.5d shows the results of a clumping operation, detecting patches of connected cells. Connections can be defined in two ways: the rook's or queen's case. By analogy with chess, where the rook can move forwards, backwards, from side to side, but not diagonally, the rook's case connects only cells that meet along a

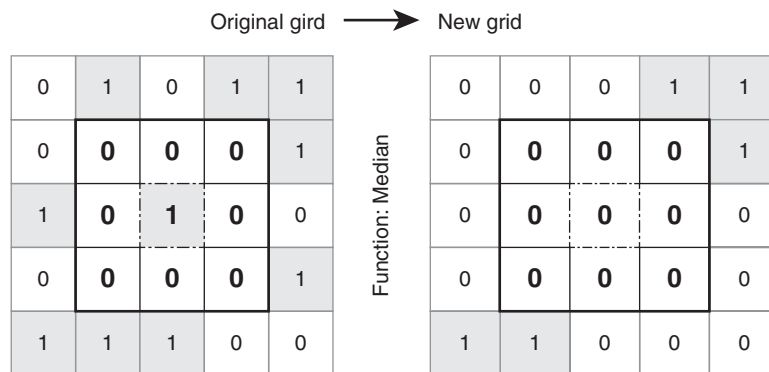


FIGURE 8.6 Illustrating a moving window operation

shared boundary (the shared side of two cells). The queen can move diagonally, so the queen's case also connects cells that meet at a corner.

Edge detection, distance calculations and buffering

Figure 8.5e arises from a process of edge detection. Finding an edge is a moving window operation: an edge exists where not all the values are the same and therefore both the range and standard deviation are greater than zero. Also shown is the distance of cells to the nearest clump of higher population density. From those distances, a 10 km buffer region is drawn around the settlements in Figure 8.5f.

Map algebra

Map algebra is another common raster operation. If two grids are geographically aligned and have the same resolution then the cell values for their corresponding locations can be added, subtracted, multiplied and divided in a straightforward way. Figure 8.7 provides an example.

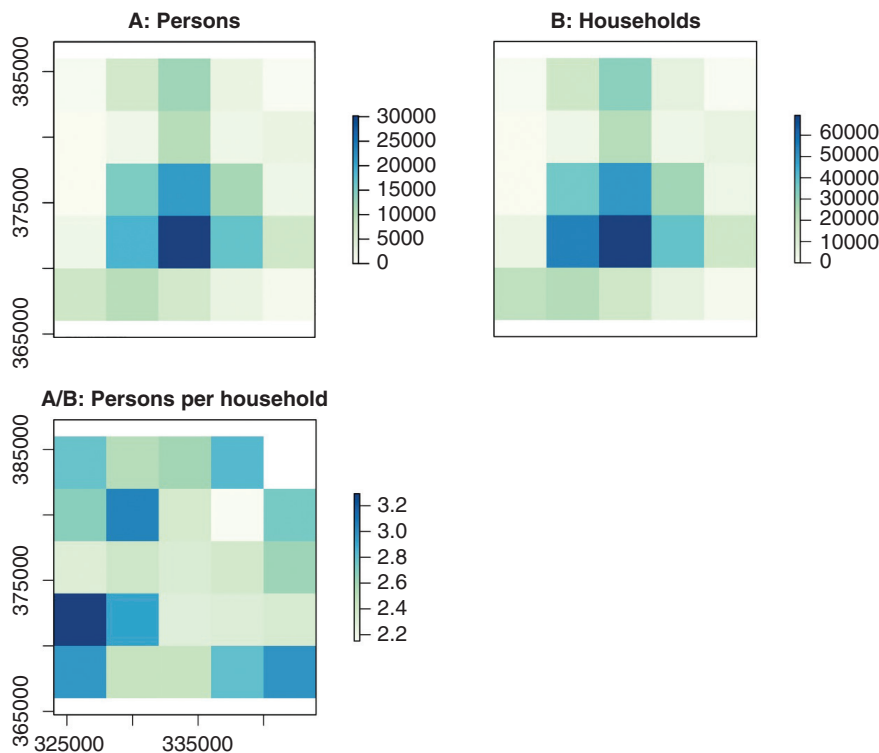


FIGURE 8.7 An example of map algebra. In this case, the number of persons per household is calculated by dividing the raster count of population by the raster count of households

Route analysis

Figure 8.8 show the result of a more sophisticated raster procedure, determining the least-cost route from A to B, where the cost increases with the height difference in traversing from one cell to the next. In other words, it seeks out the flattest route, balancing distance with the total ascent and descent.

Together, these and other raster processes provide a toolkit for location analysis, risk assessment, terrain analysis, measuring and modelling land use and handling remotely sensed data from aerial sensor and satellites, amongst other applications.

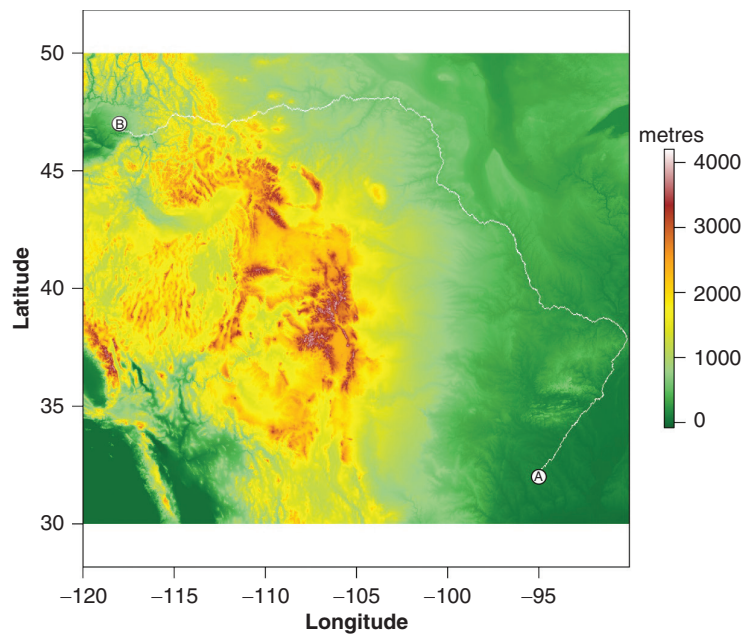


FIGURE 8.8 The results of a least-cost procedure, finding a flat route from A to B on the map

8.5 VECTOR DATA

The vector data model is more complex than raster's grid structure, although it reduces to a simple idea: locating a point on a map. Vector data come in three formats: lists of points, lists of lines and lists of polygons. A point is defined by a geographical coordinate, (x, y) . A straight line has a start point, (x_1, y_1) , and an endpoint, (x_2, y_2) . A series of lines, plotting the course of a river for example, can be represented by a 'polyline', joining multiple points together: $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$. Closed shapes (i.e. polygons) are produced when the start and the end points are the same: $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n), (x_1, y_1)\}$.

Points, lines and polygons represent geographical features on a map. For example, the location of a shopping mall, the path of a railway or the boundary of a city. In addition to this geographic information, it is common to have some information about the features' attributes – their name, their length, the number of people who live there, and so forth. This is usually stored separately from the geographic information, with a shared ID to link them back together. This link allows features to be searched for by attribute and to be highlighted on the map; reciprocally, it allows for features to be selected on the map and the information about them to be displayed in the attribute table.

Efforts are made to avoid redundant information. Imagine there are two neighbourhoods and that they share a boundary along one side. The perimeter of neighbourhood A could be recorded as one list of point coordinates; the perimeter of neighbourhood B as another. However, the two separate lists will duplicate the information for the point coordinates that lie along their shared border. A more efficient system is to create a series of tables. First, a table listing all the points for all the features in the study region, with each row identifying a point and its coordinates (without duplication). Second, a table of line segments, with each row identifying where in the points table the geographical coordinates of the lines' start and end points can be found. Third, a table of polygons, identifying which line segments form the polygon. These tables cascade downwards: the polygons data link to the lines data that link to the points data. In addition, topological information may be recorded such as which polygons lie on either side of a line. Topological information is used to detect and to ratify geographical errors in the data (e.g. unclosed polygons). It can also speed up some analytical operations such as deleting the boundary between neighbouring polygons to merge them.

As with raster data, a coordinate reference system may be required to map the data correctly. Both R and QGIS utilise EPSG codes from an online data set and registry that can be queried at <http://www.epsg.org>. Searching on code 3857, for example, provides information about the coordinate system used by Google Maps.

Figure 8.9 shows three layers of vector data where the points (representing trees) are overlaid upon the lines (representing paths and highways) that, in turn, are overlaid upon the polygons (representing buildings). The region is the area around the Brandenburg Gate, Berlin. The data are from OpenStreetMap, which is 'an initiative to create and provide free geographic data, such as street maps, to anyone.'⁶ You can download the data or contribute to the map at <http://www.openstreetmap.org>.

Vector data typically are available in shapefile format (.shp). A shapefile is actually a collection of files that are stored together and have the same name but different extensions: for example, mymap.shp, mymap.shx, mymap.dbf, mymap.prj.

⁶http://wiki.osmfoundation.org/wiki/Main_Page

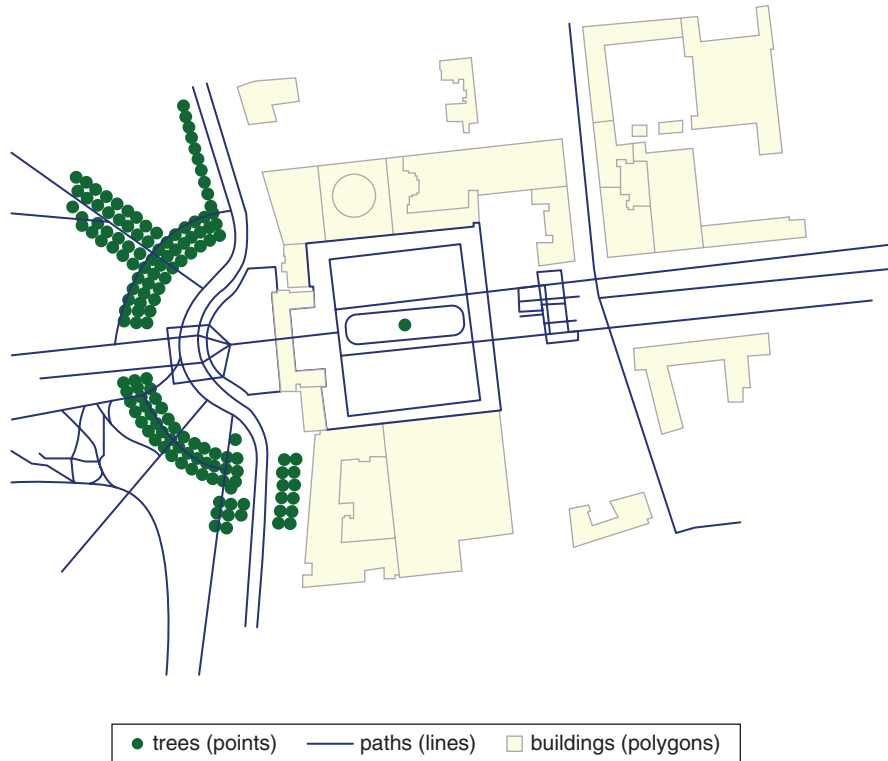


FIGURE 8.9 Three layers of vector data showing the area around the Brandenburg Gate (source of mapping data: OpenStreetMap)

The first two of these are for the locational information and provide the geography for the map; the third is for the attribute data. The fourth, which is optional, contains the CRS.

8.6 EXAMPLE OF VECTOR ANALYSIS

Vector data are well suited for geometric operations, a number of which are shown in Figure 8.10. Within England, schools do not have fixed catchment areas, but their *de facto* catchments can be modelled by mapping the locations of pupils who go to a school. These are what the points in Figure 8.10a represent. The school is represented by the + symbol, and the task is to define the core catchment area around it. One possibility would be to draw a circle around the school, extending the catchment out to a given radius. However, to do so assumes that the pupils are spread out evenly around the school, which, looking at the map, is not actually the case.

The bounding box and convex hull

The bounding box is the minimum-enclosing rectangle that contains all the points within the study region. It is shown as a dotted line in Figure 8.10a. It exaggerates the spatial extent of the school's catchment by including a lot of empty space where no pupils attending the school live. Shown as a solid line, the convex hull does better. A convex hull defines the smallest polygon that contains all the points, where all the interior angles of the polygon are less than 180 degrees. Although smaller than the bounding box, it still includes spatial outliers – the pupils who live at greatest distance from the school and remote from other pupils. A better model of the school's core catchment would exclude those outliers. One way to do so is to use a distance calculation: the distance of each pupil from the school, excluding those who are furthest away. However, this is equivalent to drawing a circle around the school and omitting those beyond a threshold distance. Again, it wrongly assumes an isotropic plain where the pupils are equally distributed in all directions around the school.

Voronoi polygons and Delaunay triangles

What matters is not just how far pupils are from the school but also how far they are from other pupils attending the same school.⁷ Voronoi, also known as Thiessen polygons, and their close relation, Delaunay triangles, are methods of space-filling from point data. All the space within the boundary of a Voronoi polygon is closer to the point around which the polygon has been drawn than it is to any other of the points. In this way, the polygon defines a point's area of influence. Delaunay triangles can be used to model three-dimensional data, to create terrain models from irregularly spaced point data, for example.⁸

Figure 8.10b illustrates Voronoi polygons (and also Delaunay triangles) created for the whole of the study region but with only an area around the school shown, to allow it to be seen more clearly. The points are the pupil locations, the solid black lines the boundaries of the polygons, the dotted lines the triangles. Looking at the image, it may be observed that the further a pupil is from their nearest neighbours, the larger the polygon around them. Therefore, the spatial outliers are those for which the Voronoi polygons are largest.

Choosing subsets by attribute

By storing the area of each Voronoi polygon in an attribute table, the largest can be searched for and omitted. Doing so leaves a subset of the polygons, representing the part of the study region where the points and therefore the pupils are closest together. This is shown in Figure 8.10c.

⁷Harris et al. (2016) use an algorithm that includes both.

⁸They create TINs: triangulated irregular networks.

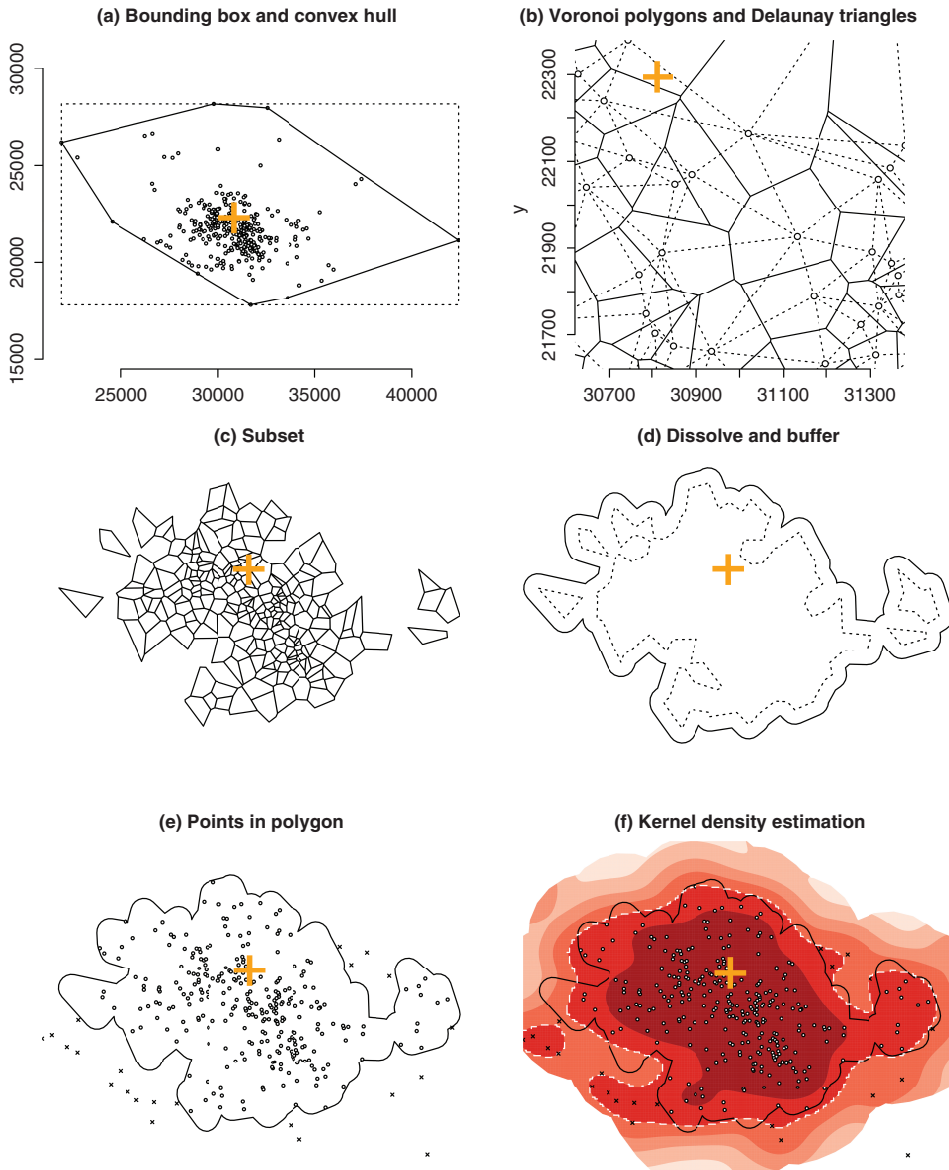


FIGURE 8.10 Examples of vector analysis, used to model the core catchment area of a school (see text for details)

Dissolving and buffering

The dissolve operation, shown in Figure 8.10d, removes the boundaries between polygons. Having done so, the catchment area consists of five disconnected parts – the

large central area and four ‘islands’ around it. These can be joined into a single whole by using a buffering process.

Points in polygon

A common vector procedure identifies which points fall within which polygons. This is an overlay operation and can be used to assign the attributes of the polygons to the points they contain (e.g. each point can be given the ID or name of the polygon that contains it). In Figure 8.10e the process is used to identify which of the pupils are within the modelled catchment area and which are not. The method can be extended to consider whether the social and demographic attributes of pupils who are in the catchment area and who go to the school are similar to those of the pupils who are in the catchment area but do not go to the school. This approach can be used to look at processes of social sorting in the admission processes to schools: do some schools disproportionately attract the highest-attaining pupils? Is there evidence that different ethnic groups are attending different schools from one another?⁹

Kernel density estimation

Figure 8.10f confirms that, in general, the modelled catchment area corresponds to the areas with the highest density of pupils, as indicated by the shading. Kernel density estimation (KDE) was used as the basis for the shading. This procedure is similar to that used to generate density estimates of how a variable is distributed (see Chapter 7), except that the distribution is now in a two-dimensional geographical space and employs a moving-window type method to identify spatial clusters of points. The KDE is a method of rasterising the vector data: its input is a list of vector points; its output is a grid of density values. The KDE could itself be used to model the catchment area – for example, as the area shown within the white dotted line.¹⁰

Unions, intersects and clipping

Finally, Figure 8.11 illustrates an instance of two overlapping polygons representing two school catchment areas, for school A and for school B. The total area served by the schools is found by a union operation, combining their catchments. Also shown is the area where the schools most obviously compete for pupils, which is where their catchment areas intersect. A related operation (not shown) is to clip one vector layer to the spatial extent of another. We might, for example, have a polyline file showing all the interstate highways in the United States, and a

⁹See, for example, Harris and Johnston (2008).

¹⁰KDEs are used in conjunction with a space-filling method to remove gaps in the modelled catchment area by Singleton et al. (2010).

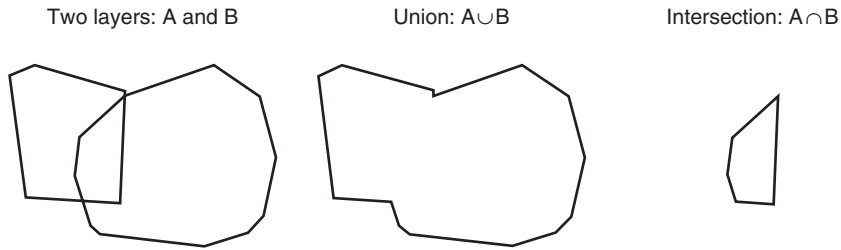


FIGURE 8.11 The union and intersection of two vector layers

polygon file giving the boundary of Florida. Clipping the first to the second leaves a map of the highways traversing the Sunshine State.

8.7 VECTOR OR RASTER?

To reiterate, vector and raster are the two main data models used in GIS. In general, vector data are used to represent discrete objects: those that have a clear boundary, route or point location. They include Census neighbourhoods. However, the boundaries can be somewhat forced – designed for administrative convenience rather than to reflect the ‘natural’ boundaries of communities or other population distributions. Diversity within the areas can be concealed. Rasters are often used to represent more continuous distributions that don’t stop at clear boundaries – for example, the height of an undulating surface. The analysis of raster data can be faster because of its simple grid structure, but, as noted, the file sizes can be large (although not necessarily greater than vector). Vector data is usually easier on the eye. The lack of pixelation makes for more aesthetically pleasing maps.

To some extent, the formats are interchangeable: raster can be vectorised; vectors can be rasterised. Raster cells can be used to represent points, lines or polygons. Elevation and other surfaces can be represented as contour lines on a map, and grid structures can be replicated as a series of polygons (as squares, known as vector grids). There is an old adage, ‘vector is better, but raster is faster’, but the truth of that statement depends on the context: ultimately the choice will be governed by how the data are collected and the application for which they will be used.

A third data model is a network structure showing the connections and interactions between people or places. A spatial weights matrix can be represented in this way, as a connectivity graph (see Chapter 11).

8.8 HOW TO LIE WITH MAPS

In a series of books that includes *How to Lie with Maps* (1996), *Air Apparent: How Meteorologists Learned to Map, Predict, and Dramatize Weather* (1999), *Bushmanders and*

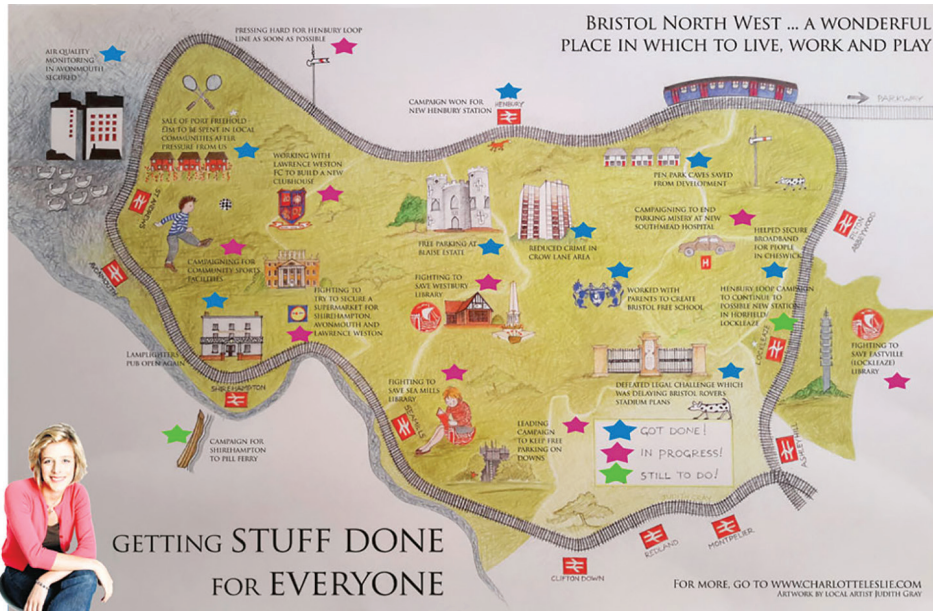


FIGURE 8.12 A map designed to support political campaigning (source : <http://charlotteleslie.com/getting-stuff-done-for-everyone-our-amazing-map-of-bristol-north-west/>)

Bullwinkles: How Politicians Manipulate Electronic Maps and Census Data to Win Elections (2001), and *No Dig, No Fly, No Go: How Maps Restrict and Control* (2010), Mark Monmonier explores the persuasive power of maps and how they are used and abused. Similarly, in *The Power of Maps* (1993) and *Rethinking the Power of Maps* (2010), Denis Wood shows how maps do not merely present things as they are but are instruments of communication, persuasion and power that reflect their authors' interests and points of view.

What we see in a map is a function of how the map is designed, and for that there are choices: how to represent real-world features, which to select, which to omit, whether to group them in some way, which colours to use, and so forth. Figure 8.12 is a clear example of a map being used for political purposes. It is from the website of the Member of Parliament for Bristol North West, near where I live. I have no objection to the map; it is a creative way of presenting the MP's involvement in local issues. However, it is certainly not impartial. I imagine voters for other political parties would present the map in very different ways.

Map classes

Figure 8.13 is more prosaic but makes clear how different choices affect the appearance of a map and therefore how it is read. In England, 80 per cent of the population classified itself as White British from the categories available in the 2011 Census. That

leaves one-fifth from other groups that include Black Africans, Black Caribbeans, Bangladeshis, Indians, Pakistanis, Chinese, other White groups and those of mixed ethnicity. Pick a place at random, though, and you are unlikely to find that 20 per cent of its population is from a group other than the White British. The reason is that groups are not evenly spread across England but concentrated in particular urban locations (although less concentrated than they have been in the recent past).

Where are these locations? Using Figure 8.13a as a guide, you'd probably conclude that they are mainly in London. It is a choropleth map where each area – each local authority – is shaded according to the percentage of its resident population that is not White British. Specifically, it is shaded according to which of four groups its percentage falls into. Those groups, which are the map classes shown in the legend, split the data into four equally spaced intervals.

Whilst the map classes are equally spaced, the same cannot be said of the data. Those are unevenly spread between the lowest and highest values. There are far more observations in the bottom category (2.4 to less than 22.62 per cent, for which there are 261) than there are in the top category (63.05 to 83.27 per cent, 11 observations). Consequently, most of the map has the lightest shade and little has the darkest. In contrast, Figure 8.13b shades the map so that approximately the same amount of the map is filled by each colour. This has the effect of greatly expanding the top category, which now ranges from 11.49 to 83.27 per cent and contains more observations than any other (133). This happens because the local authorities containing the highest percentages of groups not White British are also those of smallest area (but highest population density), requiring more of them to be grouped together to fill one quarter of the map.

Figure 8.13c is a quantile classification – specifically, a quartile classification, given there are four groups. It aims to have the same number of observations in each group. The not White British groups do not appear as dispersed as they did in Figure 8.12b but still more so than in Figure 8.12a, giving greater emphasis to the Midlands and the North West of England.

In Figure 8.13d I have chosen a manual classification, which I have set at the 50th, 75th, 90th and 100th percentiles of the distribution. Like the equal interval classification, this has the effect of emphasising the places where the percentages of the population not White British are highest. The distribution of the percentage not White British variable is shown in the strip charts of Figure 8.14, which also show how the various map classes dissect the data. An alternative classification, found in some GIS packages, is what is called a natural breaks classification. What it tries to do is choose thresholds that reflect the shape of the distribution. For the current data, it produces breaks at 11.75, 26.08 and 47.83 per cent, as well as at the minimum and maximum values of 2.40 and 83.27: five breaks in total, to give four map classes.¹¹ If we change the number of map classes, then the breaks will change too. That will further affect the appearance of the map.

¹¹The number of break points is always the number of map classes plus one.

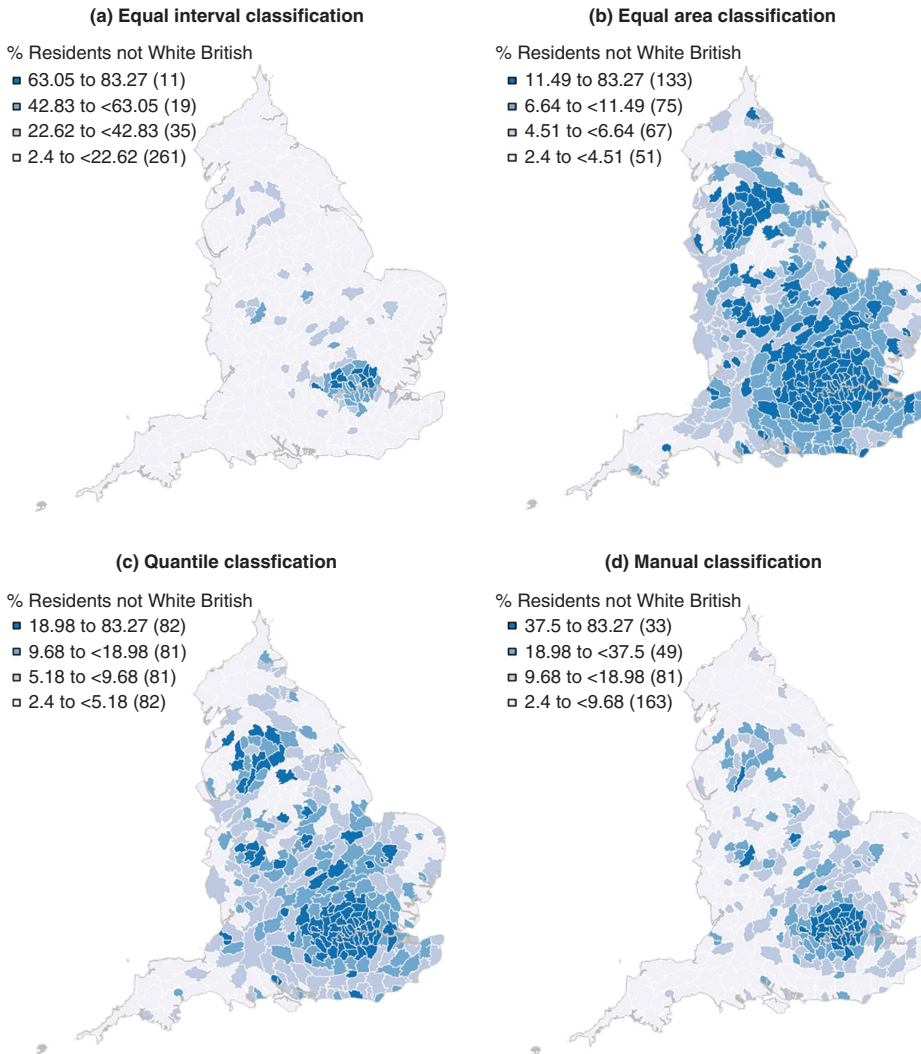


FIGURE 8.13 Changing the map classes changes our understanding of how concentrated or dispersed the non-White British groups were in England according to the 2011 Census data

Colour palettes

The reason for choosing four map classes in Figure 8.13 is that the map is shaded sequentially from the lowest to the highest values and the difference between the shading becomes hard to distinguish once there are more than four or five categories, especially if the maps are printed out in black and white (grey scale). The maps are designed to highlight more of something; that is, a greater percentage of

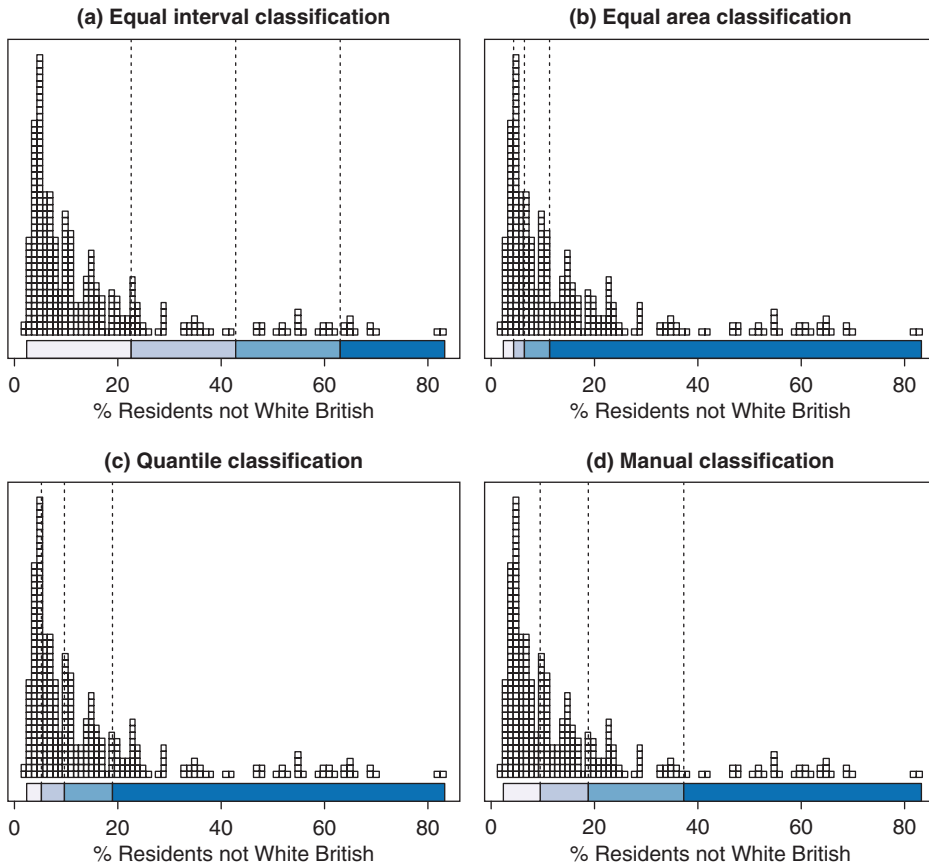


FIGURE 8.14 Showing the map classes used in Figure 8.13 and how they relate to the distribution of the variable percentage of residents not White British

the non-White British. The colour palette is one of the sequential schemes listed at <http://colorbrewer2.org> and is regarded as colour-blind safe.

Figure 8.15 adopts a different approach. Firstly, the local authorities have been ranked from the one with the least percentage of non-White British residents (rank 1) to the one with the greatest (rank 326). Second, the map has been shaded using a continuous colour ramp as opposed to discrete map classes. Finally, a diverging colour palette has been used, emphasising both the lowest and highest ranks.

A third type of colour palette is for ‘qualitative’ (i.e. categorical) data. This is used when there is no quantitative ordering to the data and the colour scheme should not imply that some values are numerically greater than others. The map at the beginning of this chapter (Figure 8.1) had a qualitative palette.

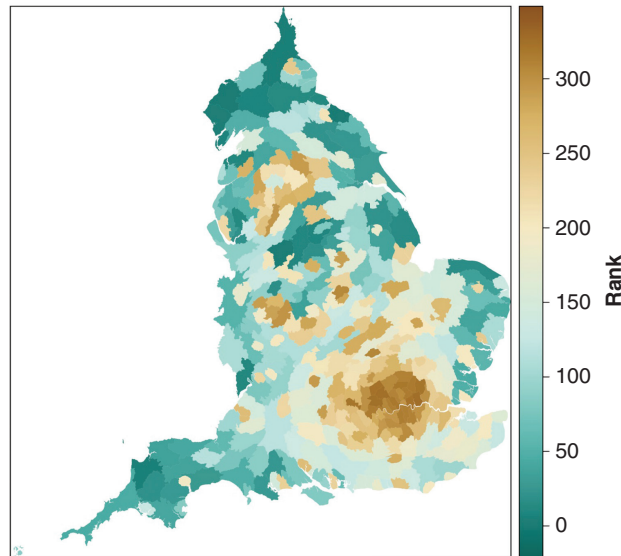


FIGURE 8.15 Showing the local authorities ranked by the percentage of their population that is not White British (highest rank = highest percentage). A colour ramp based on a diverging colour palette has been applied

Cartograms

It is time for a confession. The maps shown in Figures 8.13 and 8.15 are geographically dishonest in that they exaggerate the size of London and other urban areas. Taking Figure 8.13d as an example, it should really look like Figure 8.16a if the map is to accurately reflect the size (area) and the distances between places. However, that is not the only goal. The map is intended to communicate which places have the greatest percentages of a population that is not White British. The problem is that those places are also the smallest, despite the fact that a lot of people live within them. Their lack of geographical size makes them hard to see on the map.

A popular solution to this problem is the use of cartograms.¹² These resize the places in accordance with something other than their physical area. For example, Figure 8.16b has them sized according to the number of people who live within them. The result is to make London and other urban locations much more prominent. Another approach is to expand the areas according to the count of the variable of interest. In Figure 8.16c it is the number of people not White British. This is not simply a duplication of the shading: the colour indicates the *percentage* of the

¹²The ones shown here were produced using the Java version of ScapeToad: <http://scapetoad.choros.ch>

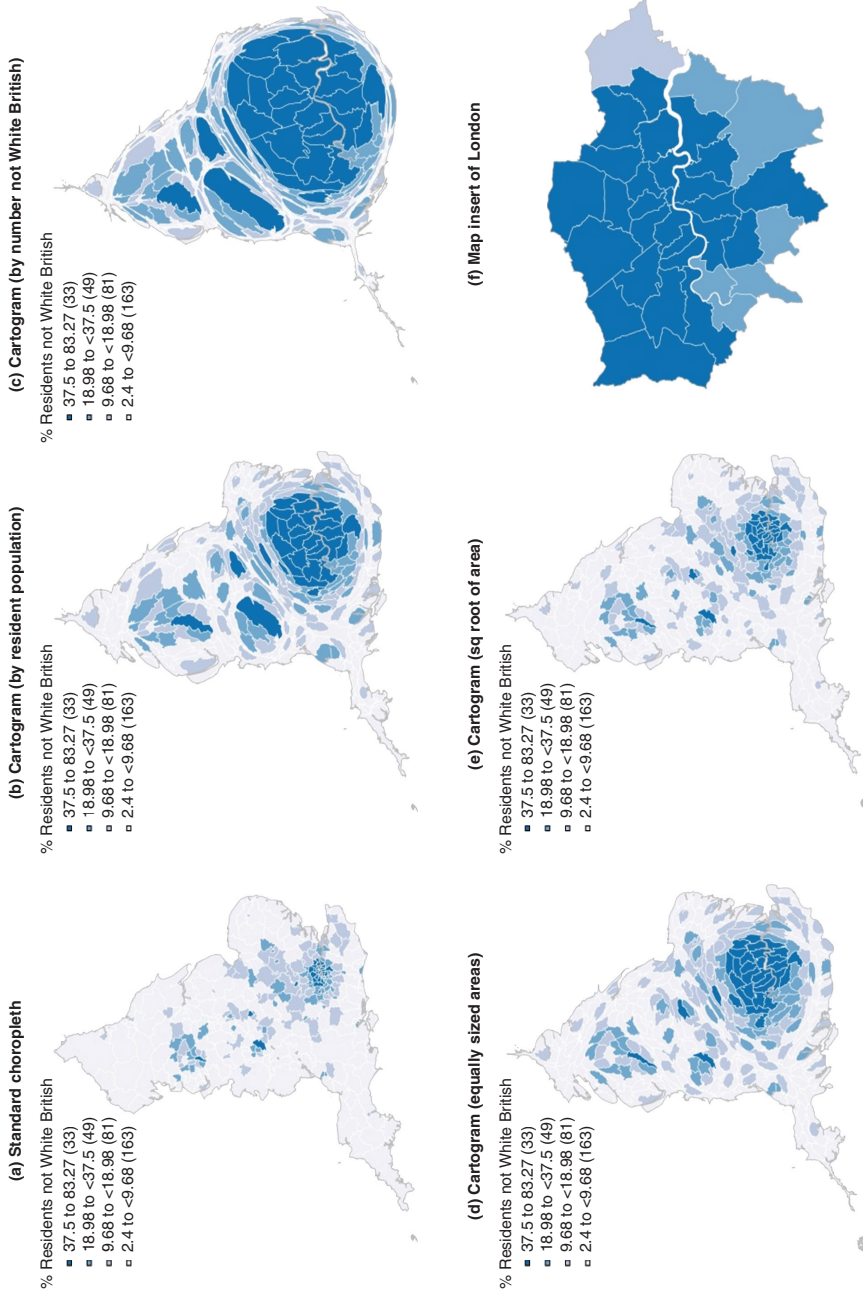


FIGURE 8.16 Showing the use of cartograms as well as a map insert to map the Census data

population that is not White British; the size of the areas indicates the *number*. The resulting map gives even greater emphasis to where those groups are located.

Cartograms can be used to create unconventional, sometimes provocative and at times unsettling views of the world – see, for example, Benjamin Hennig’s website at <http://www.viewsoftheworld.net>. They can be particularly effective when the aim is to draw attention to how unevenly something is distributed. For example, there is an interesting cartogram showing how many scientific research papers each country produced, numbers that are strongly biased towards the Northern hemisphere and against the global South, including Africa.¹³ This says something of interest about the post-colonial nature of academic knowledge production.

Nevertheless, there is a problem: as the geography of the map becomes more contorted, it loses geographical detail and it becomes harder to work out exactly which place is which. Compromise solutions are shown in Figures 8.16d and 8.16e. The first gives each local authority equal space on the map – they are all approximately the same size (but not the same shape, although that is a possibility too¹⁴). The second resizes the places according to the square root of their actual areas. This was the one used for Figures 8.13 and 8.15. It has the effect of enlarging the smaller places relative to the largest, but with less geographical distortion. A more traditional solution is to use a map insert. In many respects, I prefer it.

8.9 CONCLUSION

There is much more that could be written about Geographical Information Systems and also about geographical information science, the second phrase pointing away from the software and towards the methods, principles and theories that underpin GIS-based analyses. There are plenty of textbooks on the subject, and a very useful resource is the online edition of *Geospatial Analysis*.¹⁵ Some of the ideas will be explored further in Chapters 10 and 11.

Although they are geographical by name, GIS have not always been wholly welcomed, especially in human geography. An early criticism of GIS was that they are tools of the state, offering a very partial view of the world; that they act to objectify the world without giving voice or, rather, appearance to the plurality of often contested meanings that infuse the places around us. There is truth in this accusation. Nevertheless, although official maps of cities and states may be, for better or for worse, the products of government agencies, the same tools, and mapping more

¹³<https://theconversation.com/its-time-to-redraw-the-worlds-very-unequal-knowledge-map-44206>

¹⁴See, for example, <https://richardbrath.wordpress.com/2015/10/15/equal-area-cartograms-and-multivariate-labels/>

¹⁵<http://www.spatialanalysisonline.com>

generally, can be used to create counter-views and to assign different meanings to the spaces found on a map. This is the spirit of community mapping, participatory GIS and some aspects of counter-tourism.¹⁶ The geographer, James Cheshire, and the designer, Oliver Uberti, have harnessed a wide range of data about London, providing what their book describes as ‘100 maps and graphics that will change how you view the city’ (Cheshire and Uberti, 2012). Katherine Harmon’s book, *You Are Here* (2003), focuses on maps that venture beyond the boundaries of geography or convention. Maps have the power to persuade but also to challenge. The featured graphics in the journal *Environment Planning A* are free to access from its website, and many are cartographic.¹⁷ They include ‘The human planet’ (Hennig, 2013), ‘The virtual Bible belt’ (Zook and Graham, 2010), ‘Digital divide: the geography of Internet access’ (Graham et al., 2012), ‘Tweets by different ethnic groups in Greater London’ (Adnan and Longley, 2013), and ‘Lives on the line: mapping life expectancy along the London Tube network’ (Cheshire, 2012). They are well worth a look.

¹⁶See, for example, <http://www.participatorymethods.org/method/participatory-geographical-information-systems-pgis>, <http://www.ppgis.net> and the videos at <https://vimeo.com/channels/pgis>

¹⁷http://epn.sagepub.com/site/Virtual_Pages/Featured_Graphics_1.xhtml