Journal of Maps, 2010, 135-146

Treemap Cartography for showing Spatial and Temporal Traffic Patterns

AIDAN SLINGSBY, JO WOOD and JASON DYKES

giCentre, Department of Information Science, City University London, Northampton Square, London, EC1V 0HB, UK; a.slingsby@city.ac.uk.

Abstract

Depicting spatial and temporal aspects of traffic flows of different types is challenging. We use a treemap-based technique that is able to show multiple aspects of large quantities of spatial and temporal traffic data simultaneously. Treemaps present multivariate data as a hierarchy of rectangles that are nested within each other. Each level of the hierarchy is used to carry information about one variable, with rectangle size, arrangement and colour being potential information-carrying 'channels' for reflecting properties of the data.

We show information about the vehicles operated by eCourier (UK) Ltd. by location, vehicle type, day of the week and hour of the day. Our two maps use colour to show the volumes and speeds of traffic at each of 82,320 combinations of location, vehicle type, day of week and hour of day, concurrently. Crucially, we use a regular grid to represent location, give all the treemap nodes a constant size and order them spatially. This cartographic representation allows multiple aspects of large traffic datasets to be viewed concurrently, such that spatial and temporal patterns can be identified.

(Received 30th April 2009; Revised 1st December 2009; Accepted 7th January 2010)

ISSN 1744-5647
doi:10.4113/jom.2009.1071

1. Introduction

The depiction of spatial and temporal aspects of traffic flow on conventional maps is challenging. Most approaches use multiple maps and graphics to show spatial change through time, as small-multiple, interactive or animated maps. Our treemap-based approach (Slingsby et al., 2008; 2009) aims to represent these aspects as single, rich and data-dense graphics, in such a way that spatial and temporal trends can be identified visually. This paper shows *(a)* how treemaps can be used in a spatial context to produce novel multivariate maps, *(b)* describes the cartographic decisions made in producing an example showing traffic volumes and flow and *(c)* reflects upon these. The movements of a courier company's (http://www.ecourier.co.uk/) fleet of vehicles is used as a case study.

2. Methods

The data were obtained from eCourier's public API (http://api.ecourier.co.uk/) which gives access to over 300 million GPS points representing the whereabouts of the company's fleet of vehicles. We obtained a sample of 42.2 million GPS points representing a year's worth of courier vehicle positions within a 98 km^2 area of London (Figure 1). Each GPS point has a location, timestamp, vehicle type, and speed. eCourier categorise their vehicles into five types: vans ('van'), large vans ('lvan'), motorbikes ('mbike'), large motorbikes ('lmbike') and bicycles ('pbike' – pushbike). GPS points were recorded at regular time intervals (10sec); for this reason we could use the number of recorded points as an estimate of traffic density.

Our challenge was to display this large dataset in such a way that spatial and temporal patterns can be detected in space and time, within and between vehicle types. We used methods based on treemaps because of their ability to display multiple variables concurrently.

Treemaps were originally developed for displaying hierarchical data (Shneiderman, 1991), but they have also had success in visualising large multivariate datasets (e.g. Kolatch and Weinstein, 2001; Vliegen et al., 2006; Slingsby et al., 2008; 2009) in a data-dense and space-efficient manner. They recursively subdivide space for each variable, resulting in a set of tessellating and non-occluding rectangles, each of which has the three potential information-carrying properties of *size*, *colour* and *arrangement*.

Figure 2 summarises traffic volume (rectangle size) and average speed including time resting (rectangle colour) by vehicle type, day of the week and hour of the day using

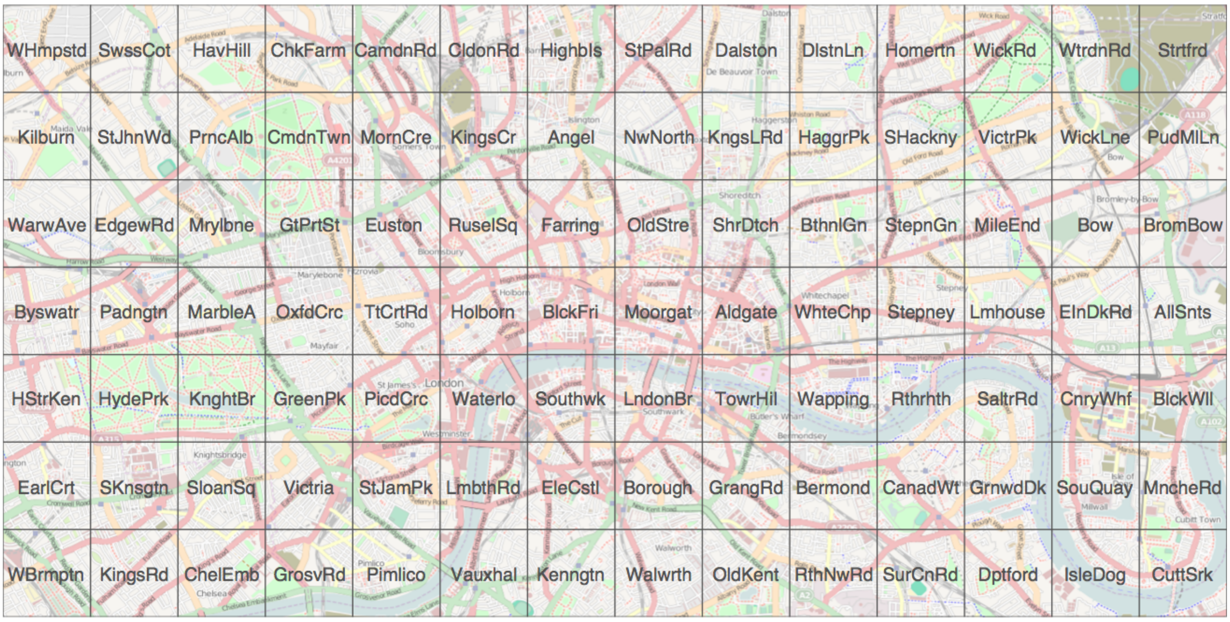

Figure 1. The 98 km^2 case study area, divided into 1×1 km grid cells. *Source: http://www.openstreetmap.com/.*

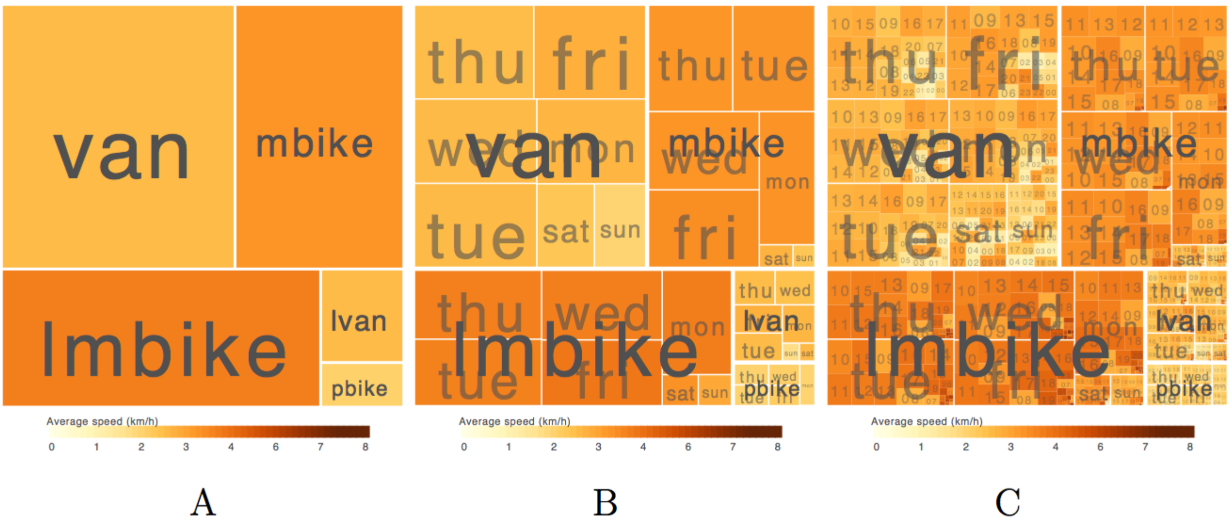

Figure 2. Treemaps that show traffic volume (area size) and average speed (colour; km/h) of traffic by three categorical variables (day, hour and type), organised into the hierarchy type-day-hour. **A:** First level of the hierarchy. **B:** First and second levels. **C:** All three levels of the hierarchy. *Colour scheme: ColorBrewer 'Yl-Or-Br' from http://www.colorbrewer.org/.*

'squarified' treemaps (Bruls et al., 1999) in which rectangles are set to be as square as possible to make relative size comparison easier. Squarified treemaps usually arrange rectangles in decreasing order of size (within their containing rectangle) from the top right to bottom left to improve their square aspect. Figure 2A shows that vans comprise about a third of all traffic and that large motorbikes ('lmbike') are faster, on average. In Figure 2B, each vehicle type has been subdivided into days of the week. There is more traffic on weekdays, proportionally more vans on weekend than other modes of transport and no bicycle ('pbike') use at weekends. Figure 2C also includes the hour of the day.

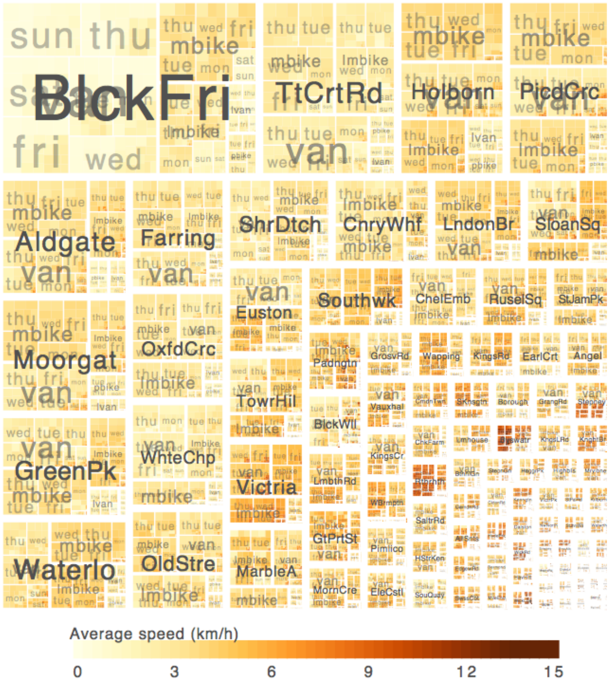

Figure 3. As the treemaps in Figure 2, but with a spatial variable (grid square) added to the base of the hierarchy.

Space can be added to these treemap representations by inserting a spatial variable into the hierarchy. Figure 3 shows this, in which GPS points have been aggregated into the 1km^2 grid squares identified in Figure 1. Blackfriars ('BlckFri') has the most traffic, vans dominate in Blackfriars and motorbikes dominate in and around Marble Arch ('MarbleA').

Although comparisons can be made between categorical values – e.g. the comparison of Blackfriars and Marble Arch – it is difficult to detect broader spatial patterns. By arranging the cell squares geographically – as in Figure 4 – the treemap is transformed into a cartogram (Wood and Dykes, 2008) and spatial patterns are discernible; for example, there is more traffic in the centre of the study area; Canary Wharf ('CnryWhf' in the SE) has significantly more traffic than its surroundings; and peripheral grid squares tend to have higher average speeds.

There are two characteristics of Figure 4 that make it difficult to interpret. Firstly, the small nodes associated with low traffic volumes are difficult to resolve – this is problematic for identifying patterns relating to low volumes of traffic. Secondly, sizing geographically-arranged areas by anything other than their geographical area will result in (often significant) displacement. This is characteristic of all cartograms, but those

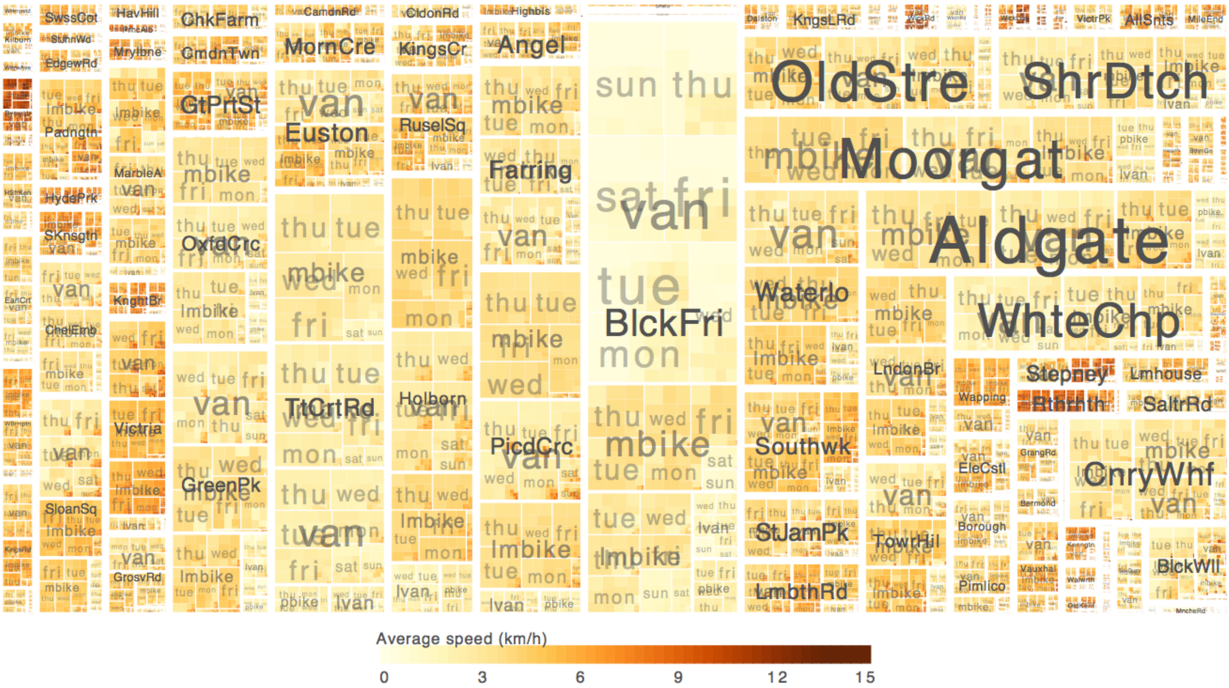

Figure 4. As Figure 3, but the rectangles corresponding to grid squares are arranged geographically.

Figure 5. As Figure 4, but with rectangle sizes fixed for each hierarchical level (size no longer relates to traffic volume). This is one of the components of the map.

that maintain area contiguity – such as Gastner Cartograms (Gastner and Newman, 2004) – are usually easier to interpret because a recognisable frame of reference is often maintained. We address these problems in Figure 5 by *fixing the size* of the rectangles at each hierarchical level with the dual effect of overwhelmingly *improving the consistency of layout* and giving each rectangle *equal prominence* regardless of the associated traffic volume. Note that consistency of grid-square position is achieved because the geographical area of the grid squares are equal to each other, each contains an identical number of rectangles and the aspect ratio of the geographical extent of the study area is used (2:1). The loss of size as in information-carrying property has the result of only being able to show either traffic volume (purple) or average speed (orange), but with the benefit of increased readability.

The importance of the arrangement of rectangles to detect spatial patterns also extends to temporal patterns (Slingsby et al., 2009). Figure 6 is an enlarged portion (coloured by traffic volume) that shows how different and appropriate arrangements are used at each hierarchical level. Arrangements used are *slice-and-dice* (Shneiderman, 1991), *ordered squarified* (Wood and Dykes, 2008) and *spatial* (Wood and Dykes, 2008).

- *Level 1:* grid squares (labelled with an abbreviated area name) with spatial arrangement (i.e. 'MarbleA' is south of 'Mrylbne').
- *Level 2:* transport type with slice-and-dice arrangement (sliced into two horizonal strips: upper for 4-wheeled traffic; lower for 2-wheeled traffic below).
- *Level 3:* transport mode with slice-and-dice arrangment (diced from left to right from large to small).
- *Level 4:* day-of-week with slice-and-dice arrangement (sliced into vertical strips from Sunday on the left to Saturday on the right).
- *Level 5:* hour-of-day with slice-and-dice arrangement (diced from top to bottom with midnight at the top to 2300 at the bottom). (Levels 4 and 5 together represent 2D temporal ordering).

The figure shows that most of the traffic is during working hours, but that this varies spatially and by vehicle type. Vans appear to be in use at all times of the day and week, large motorbikes ('lmbike') appear to start fairly abruptly at 0700 but gradually decrease in use in the evening compared to other vehicle types, large vans ('lvan') are more in use in the SE of the 4 km^2 area and the same appears to be true for bicycles.

With this arrangement in mind, spatial and temporal patterns can be identified in the map even though the fine detail may be unresolvable. For example, in the upper portion of the map (traffic volume; purple):

- Bicycles (bottom right of each grid square) are only used in Central London

Figure 6. Excerpt from Figure 5 (coloured by traffic volume rather than average speed, showing the consistency of layout. Numbers refer to hour of the day. *Colour scheme: ColorBrewer 'Rd-Pu' from* http://www.colorbrewer.org/*; grey indicates that there were no vehicles.*

- Large motorbikes (bottom left) are more active in the western half of the study area and locally around CnryWhf (east) along with motorbikes (bottom middle) and vans (top right).

- Around PicdCrc (centre) and CnryWhf, motorbikes are used more than vans, but only during office hours.
- KingsCr has very high early morning usage.

In the lower portion of the map (average speed; orange; also reproduced in Figure 5):

- Speeds are slower in the centre of the study area, but large motorbike show higher speeds in the early morning.
- High average speeds are found in the northeast and west.
- Higher speeds are found where there is lower traffic use, but there is also high variation in speed, some of which is due to smaller sample sizes.

This hierarchy of variables (grid-type-day-hour) and layout is not suitable for all tasks. Although the consistent layout allows, for example, van traffic to be identified in each grid square, their comparison across space is more difficult than if they were grouped spatially. Similarly, comparing van traffic in grid cells between the eastern and western edges of the map is more difficult than if they were adjacent to each other. Reconfiguring the variable hierarchy and layout helps address different research questions (Slingsby et al., 2009). Other graphics such as traffic frequency by binned average speed might also be useful. Note also that we chose to increase stationary traffic in our computation of average speed, which is why the values are so low. However, a very high proportion of the traffic in Blackfriars is stationary and occurs throughout the night. This suggests that some of the GPS units record vehicle positions when vehicles are not in use, something worth investigating further. Depicting alternative statistical summaries of speed such as median, standard deviation and coefficient of variation can also assist in interpretation (Slingsby et al., 2008).

Figure 6 shows the two aspect ratios associated with the non-square number of vehicle types. In Figure 7A, the least utilised vehicle type ('pbike') has been dropped, resulting in a more pleasing symmetrical arrangement in which the temporal information for each vehicle type is of the same shape. Whilst comparison may be easier in the less data dense Figure 7A, the asymmetry in Figure 7B makes, for example, the top left portion ('lvan') more distinguishable then the bottom left ('lmbike').

In Figure 8 we attempt to show traffic volume by size using the contiguity-preserving property of the Gastner cartogram algorithm. This is an improvement over Figure 4 because contiguity is preserved, but the shape-distortion does reduce the consistency of layout present in the fixed-node treemaps, reducing readability, particularly at the

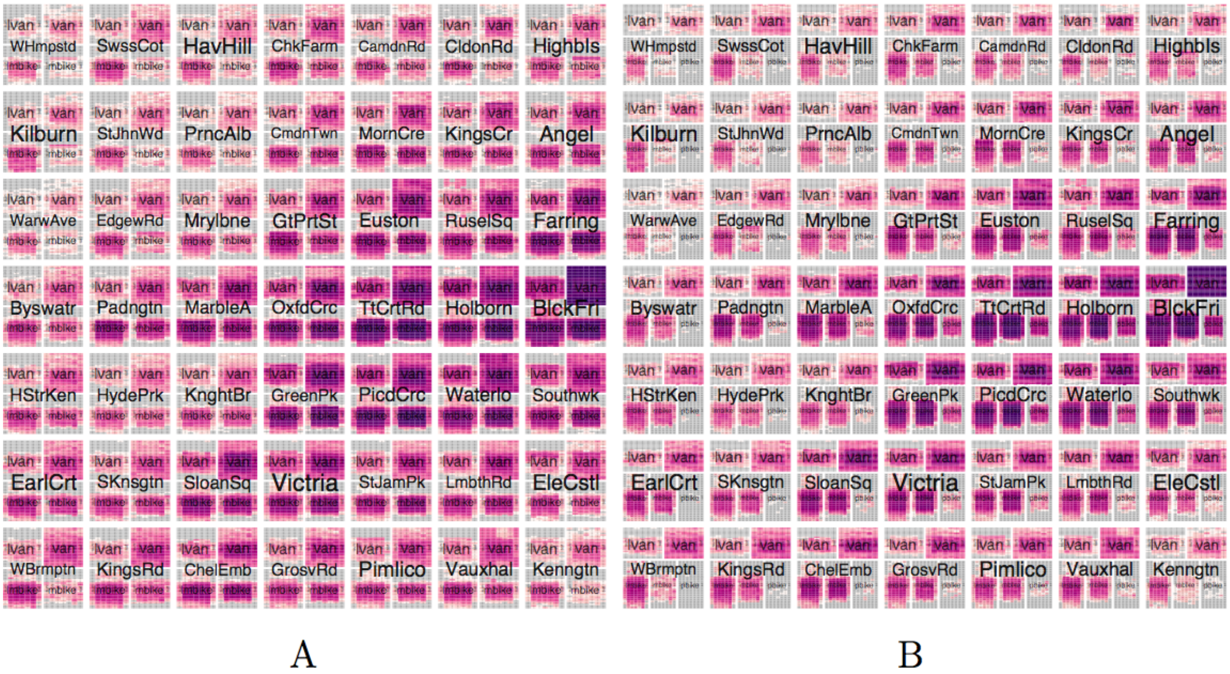

Figure 7. Fixed-size treemaps coloured by traffic volume. **A:** Four vehicle types. **B:** Five vehicle types.

edges. However, there are a number of possible benefits. One of these is that size draws more attention to the huge differences that exist in traffic density than logarithmically-scaled colour. Another is that it is easier to relate both aspects of traffic to individual nodes – but only where traffic volume is high enough for nodes to be resolvable. A useful effect of this is that attention is drawn towards nodes with greater sample sizes (more traffic) and away from those with lower sample sizes. Many of the high average speeds in Figure 3 represent 'noise' associated with low sample sizes. The Gastner cartogram lessens the visual impact of this noise by reducing the prominence of average speeds based on small sizes. It shows that the high average speeds for isolated hours in CanryWhf are based on relatively high sample sizes, an observation that is not as easily identifiable when using separate images.

3. Conclusion

We conclude that some treemap techniques have a useful role to play in cartography where multivariate data are the subject of the map, particularly where the data have spatial and temporal components. The availability of slice-and-dice, ordered squarified and spatial treemap layout algorithms and the ability to apply these separately to different levels of a variable hierarchy of variables, is a powerful means to produce rich and data-dense maps.

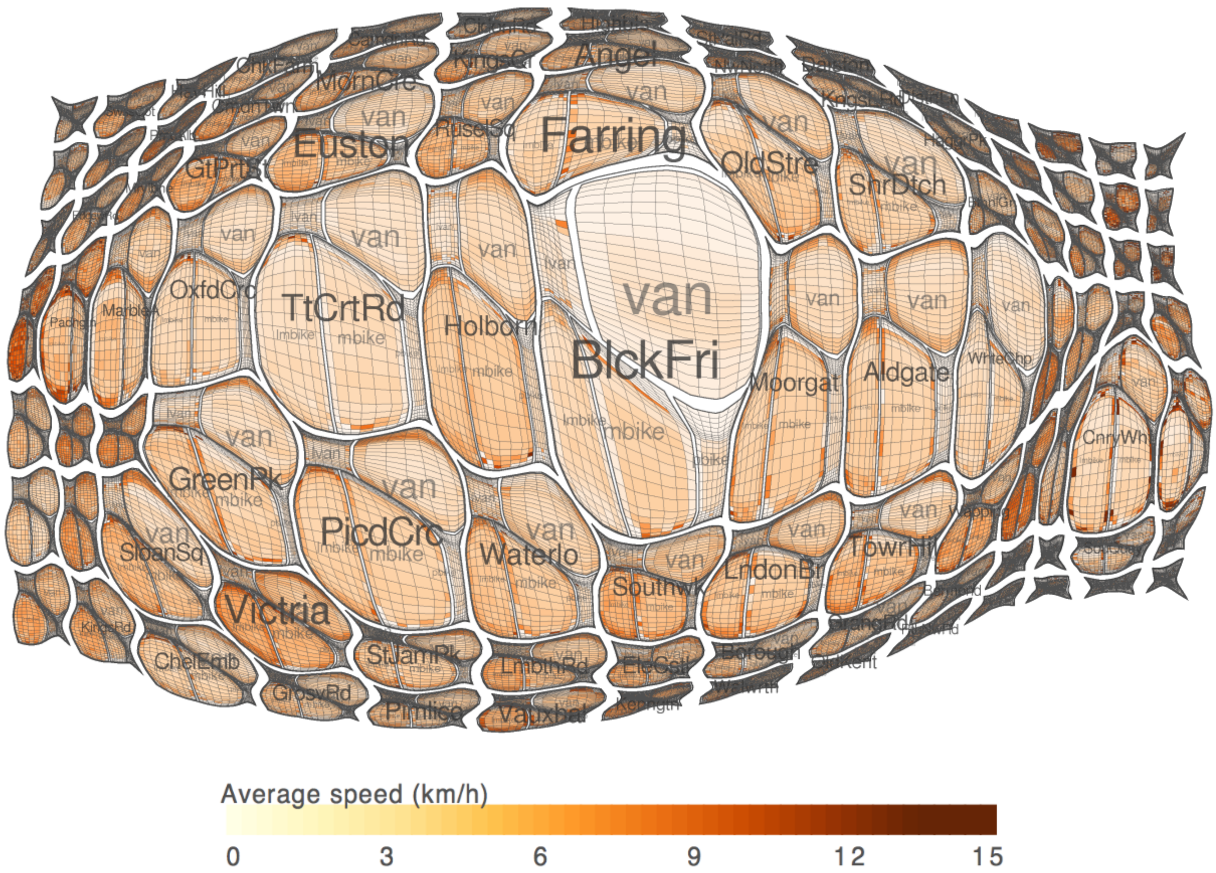

Figure 8. Gastner cartogram of the fixed-size treemap shown in Figure 5, coloured by traffic volume. The fixed-size rectangles have been distorted so that their areas correspond to traffic density.

As illustrated in the previous section, the use of size, colour and arrangement in treemaps draws attention to different aspects of the dataset, often resulting in tradeoffs, the evaluation of which is part of cartographic design. Of the examples presented in the previous section, we chose the two treemaps with fixed-size rectangles – one for traffic volume and one for average speed – for our map, which we consider to be the most appropriate for revealing spatial and temporal patterns in the traffic dataset. Although the potential information-carrying property of size has been lost, the resulting maps possess a logical consistency that makes the temporal components of the data more interpretable.

Although not discussed here, interactive techniques such as zooming, panning, switching hierarchies, switching hierarchy depths, brushing and multiple coordinated views enhance the potential for finding patterns in the data and investigating hypotheses that may arise (Slingsby et al., 2008; 2009). An interactive map at http://gicentre.org/treemaptraffic/ demonstrates some of these techniques, illustrating their potential.

Software

The data were retrieved from eCourier's public api (http://api.ecourier.co.uk/) using PHP, the coordinates were projected to the GB National Grid (OSGB) using PHPCoord (http://www.jstott.me.uk/phpcoord/) and then loaded into a PostgreSQL database http://www.postgresql.org/. Standard SQL was used to aggregate the data as used for this work.

The treemap layouts were produced by TreeMappa (http://www.treemappa.com/). This software takes a CSV file, the format of which is described on the website and a number of parameters that control the treemap appearance. The software supports a variety of treemap layout algorithms and will export the resulting treemap as a raster graphic (PNG), vector graphic (SVG), Shapefile (SHP) or an ASCII-based coordinate list.

High-quality PDF outputs can be produced by printing SVGs in most web browsers to PDFs. To obtain a high degree of control over the output, we used Processing (http://www.processing.org/) to render the maps from the ASCII coordinate lists exported from TreeMappa and then using Processing's PDF libraries to export into PDFs.

The Gastner cartogram of the treemap (Figure 8) was produced from TreeMappa's Shapefile output using ScapeToad (http://chorogram.choros.ch/) and rendered using Processing.

The interactive map at http://gicentre.org/treemaptraffic/ was produced using Processing.

References

BRULS, M., HUIZING, K. and VAN WIJK, J. (1999) Squarified Treemaps, In Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization, Press, pp. 33–42.

GASTNER, M. T. and NEWMAN, M. E. J. (2004) Diffusion-based method for producing density-equalizing maps, Proceedings of the National Academy of Sciences of USA, 101, 7499–7504.
URL http://dx.doi.org/10.1073/pnas.0400280101

KOLATCH, E. and WEINSTEIN, B. (2001) CatTrees: Dynamic Visualization of Categorical Data Using Treemaps. Available from: http://www.cs.umd.edu/class/spring2001/cmsc838b/Project/Kolatch_Weinstein/, [Last accessed: 7 January, 2010].

SHNEIDERMAN, B. (1991) Tree visualization with Tree-maps: A 2-d space-filling approach, ACM Transactions on Graphics, 11, 92–99.
URL http://dx.doi.org/10.1145/102377.115768

SLINGSBY, A., DYKES, J. and WOOD, J. (2008) Using treemaps for variable selection in spatio-temporal visualisation, Information Visualization, 7, 210–224.
URL http://dx.doi.org/10.1057/palgrave.ivs.9500185

SLINGSBY, A., DYKES, J. and WOOD, J. (2009) Configuring Hierarchical Questions to Address Research Questions, IEEE Transactions on Visualization and Computer Graphics, 15, 977–984.
URL http://dx.doi.org/10.1109/TVCG.2009.128

VLIEGEN, R., VAN WIJK, J. and VAN DER LINDEN, E. (2006) Visualizing Business Data with Generalized Treemaps, IEEE Transactions on Visualization and Computer Graphics, 12, 789–796.
URL http://dx.doi.org/10.1109/TVCG.2006.200

WOOD, J. and DYKES, J. (2008) Spatially Ordered Treemaps, Visualization and Computer Graphics, IEEE Transactions on Visualization and Computer Graphics, 14, 1348–1355.
URL http://dx.doi.org/10.1109/TVCG.2008.165