Using Treemaps for Variable Selection in Spatio-Temporal Visualization

Jo Wood, Aidan Slingsby, Jason Dykes

giCentre, City University London, EC1V 0HB

jwo¦a.slingsby¦jad7 @soi.city.ac.uk

http://www.gicentre.org/papers/agile08/

Introduction

Understanding the structure of large multivariate spatio-temporal datasets is challenging because of the spatial and temporal scales to be considered and the number of variable combinations (Openshaw *et al*, 1987). Individual visualization techniques can reveal structure in such data sets by summarising (Dykes and Mountain, 2003) and showing patterns involving particular variable combinations. Dynamic visual methods are useful for the exploratory analysis of data because they help the interpreter understand the structure of the dataset, may raise questions about the phenomenon represented (through apparent consistency or anomaly) and importantly, can help identify noteworthy combinations of variables for use in more detailed mapping and analysis.

We have been working with a large dataset of 90 million vehicle locations recorded by a courier fleet over an 18 month period[2]. Each location has a position, a timestamp, mode (van, large van, motorbike, large motorbike and bicycle), heading and speed. Points are recorded every ten seconds by GPS whilst any vehicle is in use. This dataset has a number of variables that can be considered at different spatial and temporal scales and it is likely that it contains interesting and complex spatial and temporal patterns relating to traffic patterns in London. We have snapped a 20% sample[3] of these GPS points to London road segments. This allows us to produce road maps of traffic volume and speed by road segment for different vehicle types by month of year, day of week or hour of day. This results in a huge number of mapping possibilities for exploratory spatio-temporal data analysis – what do we map and at which spatial and temporal scales? The complexity of the London road network is a visualization constraint. It is sufficiently dense to preclude the use of colour symbolism or other visual variables to represent data relating to particular roads when mapping all roads. We need techniques to identify interesting areas and combinations of variables for use in producing visual encodings and analysis that are potentially useful in understanding this rich data set.

Treemaps of Variable Combinations

To explore the geography of variable relationship we consider variable combinations as a hierarchy and use treemaps (Schneiderman, 1992). We use spatial and temporal ordering (Wood and Dykes, submitted) to select variables likely to be worth mapping and to try and identify patterns and raise questions during the exploration process.

Figure 1[4] is a squarified treemap (Bruls *et al*, 2000) of the ecourier data with variable combinations encoded as a hierarchy of mode, day and hour. The size of each leaf is proportional to the volume of traffic and the colour represents the average speed. According to Figure 1, vans and large motorbikes account for the greatest share of traffic, in roughly equal quantities. Large motorbikes tend to be faster. In most cases, weekdays have roughly equal quantities of traffic and have more than weekend days, but bicycles are rarely used on weekends.

---

[2] Data from eCourier. Data retrieved using the public API: [http://api.ecourier.co.uk/]

[3] Data sampled for computational convenience using consecutive GPS points (in time).

[4] Full sized colour images of all figures available online: [http://www.gicentre.org/papers/agile08/]

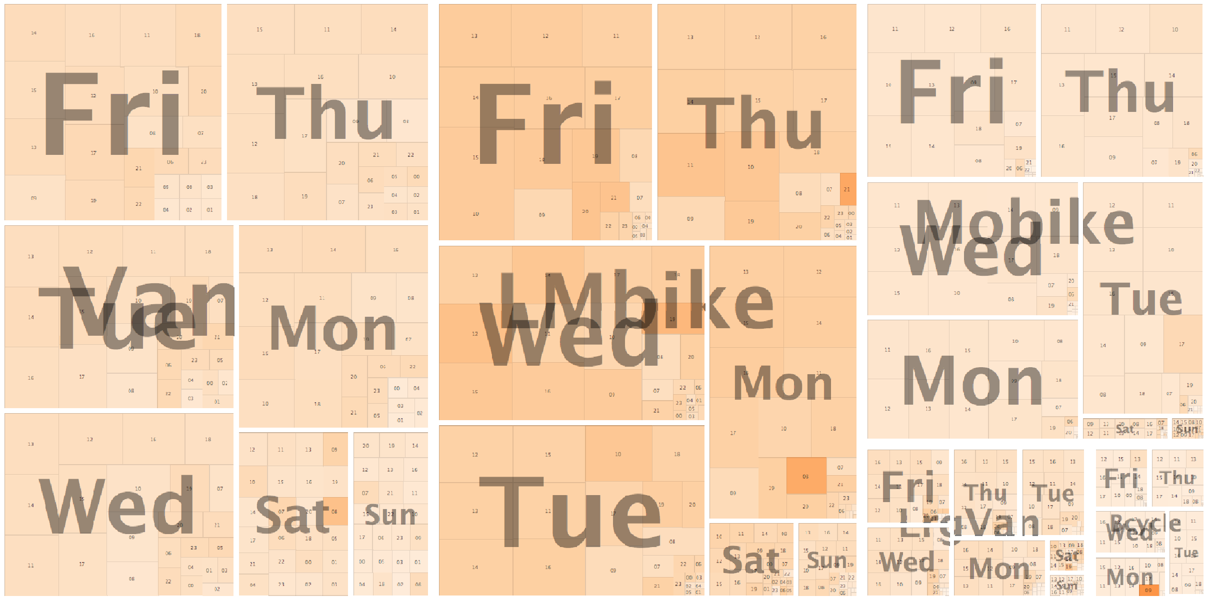

Figure 1: Treemap showing mode, day and hour of non-stationary traffic (70 million GPS points), sized by traffic volume and coloured by speed. Nodes and leaves are ordered by size from top left.

Spatial Treemaps (with Spatial and Temporal ordering)

Figure 1 shows attribute relationships between vehicular and temporal characteristics of all traffic. We constrain the area[5] (Figure 2), and use our spatial treemaps (Wood and Dykes, submitted) to show the geography of these patterns. A 1km grid has been placed across Central London in Figure 3. We use this to group points and insert it into the base of the hierarchy. Figure 3 and Figure 4 show the grid cells ordered by their locations using the spatial squarified layout. Constant cell size preserves the spatial arrangement and results in geographic small multiple treemaps in this case. Our algorithm is sufficiently robust to produce spatial treemaps of irregular areas of different sizes.

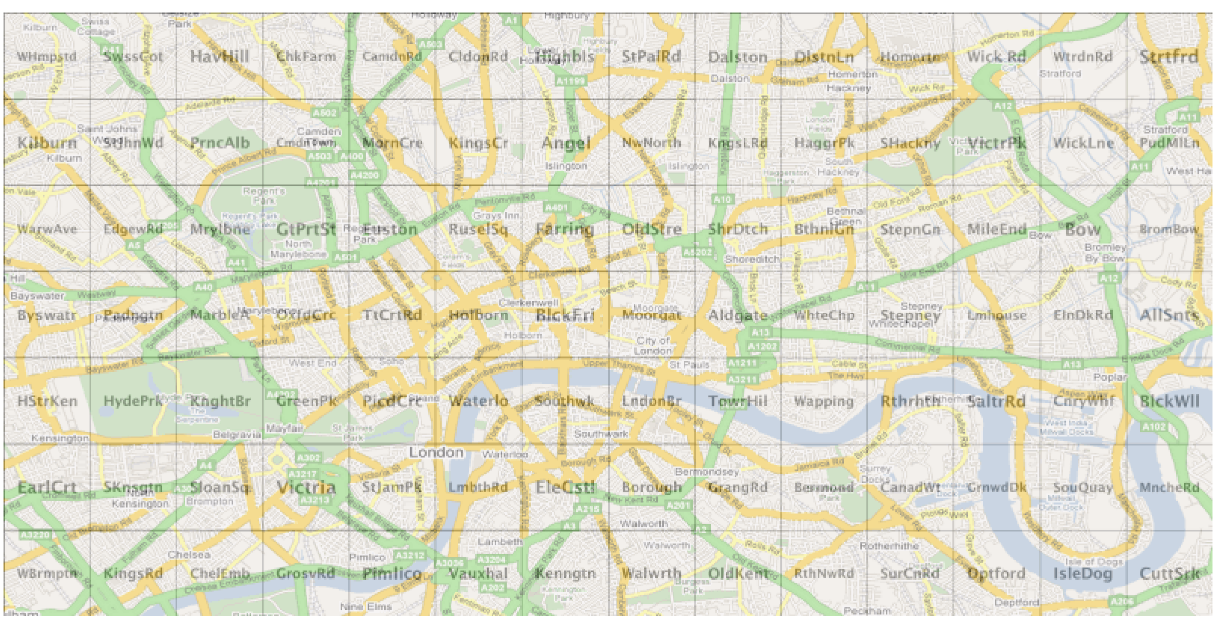

Figure 2: GPS points were allocated to 98 1km grid cells. Locations shown are approximate.

---

[5] This area of Central and East London contains 47 million points associated with non-stationary vehicles.

As reproduced here Figure 3 is too small to see detail, but we may detect generic patterns: the density of traffic is highest just north of the centre of the study area (north of the river – see Figure 2); one cell in the East has anomalous traffic density - **ElnDkRd [13,4]**; the contents of an entire cell has constantly high density of traffic compared with neighbours – **Farring [7,3]**. Accordingly, Figure 4 shows slower speeds where density of traffic is high - **ElnDkRd [13,4]** and **Farring [7,3]** have slower average speeds than surrounding cells.

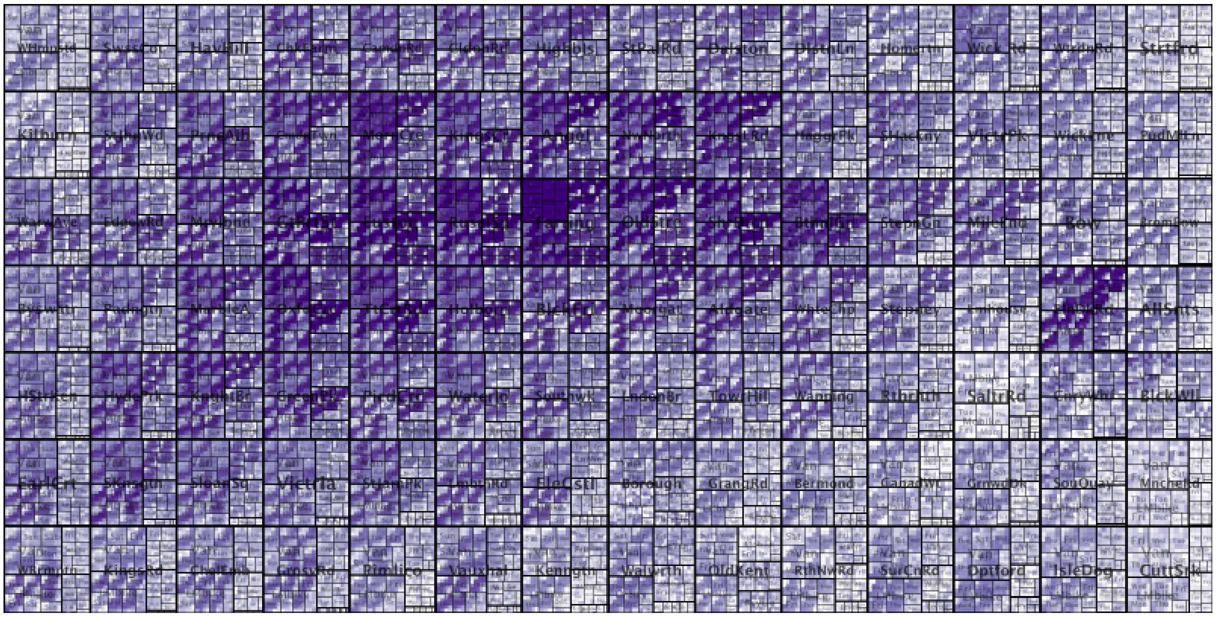

Figure 3: Spatial treemap, grouped by space, mode, day and hour, coloured by traffic density.

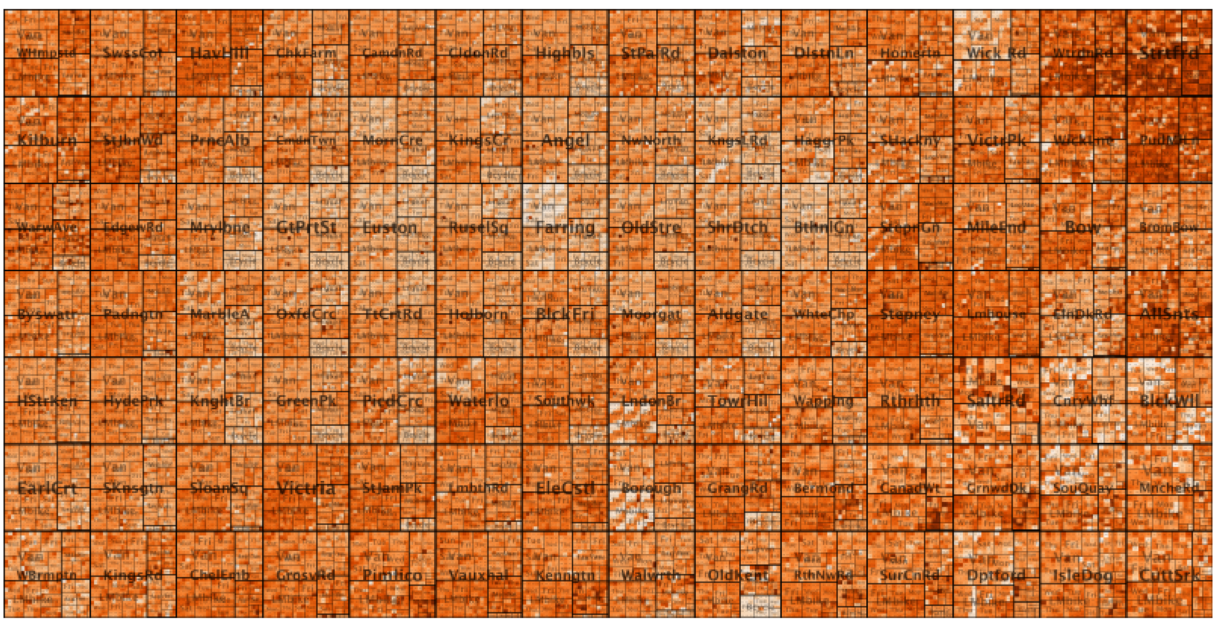

Figure 4: Spatial treemap, grouped by space, mode, day and hour, coloured by traffic speed.

Exploration Through Interaction and Linking

Our treemap software allows us to pan and zoom to see the details of the attribute treemaps within each geographic cell. Figure 5 shows an enlarged section of both treemaps for **Farring [7,3]** where van traffic is particularly dense and slow. This is consistently the case for all hours in all days of the week. We order hours chronologically from top left to bottom right, so diagonal stripes in our spatial treemaps show diurnal differences in traffic. Specifically, we might speculate that **Farring**

[7,3] is a destination for van deliveries, or that there is a van depot, or that this is an important node on a number of cross-London routes, or that there are traffic problems here.

Figure 5: Enlarged section of the treemaps for **Farring [7,3] -** Farringdon

Further exploration of such hypotheses requires consideration of the road network itself. Each cell contains somewhere in the order of 1,000 road edges enabling us to use colour symbolism or other visual variables to represent data relating to each. We are developing direct links between treemaps and interactive maps of the road network for variable selection and geovisualization.

Conclusion and Outlook

Treemaps and spatial treemaps are effective ways of summarising large datasets. Unlike network maps of London, they contain consistently sized symbols that do not overlap and to which colour symbolism can be applied. Data dense yet legible, their structure enables us to identify anomalies and repeating patterns quickly at different levels in our attribute hierarchy. For example, the occurrence and spatial persistence of the diagonal temporal patterns apparent in Figure 5 can be considered in the spatial treemaps.

Importantly, treemaps can assist in variable selection for more detailed mapping and other analysis - a key difficulty with the visual exploration of large datasets relating to complex networks. Our techniques allow us to link between treemaps and legible dynamic maps of selected combinations of time, space and attribute. We continue to develop interactive links between treemaps and maps to streamline this process.

References

Bruls, M., Huizing, K. and Wijk, J.V. (2000). "Squarified Treemaps*". http://www.win.tue.nl/~vanwijk/stm.pdf (accessed, 17th December 2007)*

Dykes J.A. and Mountain D.M. (2003). Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications*. Computational Statistics and Data Analysis, 43 (4) 581-603.*

Openshaw, S., Charlton, M.E., Wymer, C., Craft, A. (1987). "A mark 1 geographical analysis machine for the automated analysis of Point data sets". *Int. Journal of Geographical Information Systems 1 (4), pp 335–358*

Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics, 11(1), pp.92-99.*

Wood, J., Dykes, J. (submitted). From Slice and Dice to Hierarchical Cartograms: Spatial Referencing of Treemaps. *Submitted to GISRUK 2008.*