1 Introduction

A popular approach to investigation of geographical patterns in a spatially referenced data set is the concept of *local statistics* (Getis and Ord, 1992; Unwin and Unwin, 1998), perhaps arising from Openshaw’s stated dissatisfaction with ‘whole map statistics’ – (Openshaw *et al.*, 1987). A point location $u$ in the study area is selected, and some statistical technique is applied only to the data within some radius $h$ centred around $u$. Applying this procedure to a number of locations spanning the study area gives a view of spatial variability in the distribution of the data values. For the ‘geographically weighted’ (GW-) approaches, $h$ is either fixed for all $u$, or chosen to equal the distance from each $u$ to its $k$th nearest neighbour - see for example Brunsdon *et al.* (1996). Mapping the results for each location for some given $h$ is an effective way of showing local changes in distribution of one or more attribute variables. However, a number of spatial processes operate at several scales simultaneously. While static mapping is useful for exploring spatial variation for a fixed $k$ or $h$ (Brunsdon *et al.*, 2002), visualising variability of the spatial patterns with $h$, is arguably too complex a task for a static map. Interactive graphics provide opportunities for exploring complex structured data sets (Thomas and Cook, 2005). The intention here is to provide an interactive tool for varying $h$ through a process of visualization to help gain insight into spatial data and characterise spatial processes.

Here we propose a series of graphics and interactions that meet this aim and present an interactive visualisation tool that allows analysts to investigate the properties of spatial data at a range of scales, through a number of views.
2. Context: Challenges for Multivariable Spatial Analysis - Guerry’s Moral Statistics of France

We focus on a particular multivariate geographic data set – that collated and graphically represented by André-Michel Guerry in his ‘Essai sur la statistique morale de la France’ (Whitt and Reinking, 2002). He identified geographic outliers and some regional trends and used such visual inspections to hypothesize about relationships between variables. Friendly (2006; in press) uses statistical and graphical methods to revisit the data set. Regression analysis shows that some of Guerry’s postulated associations do not hold and that others omitted by Guerry exist. Friendly (2006; in press) uses multivariate graphics and conditioned choropleths (Carr et al., 2005) to augment Guerry’s univariate maps.

3. Geographically Weighted Graphics

Graphical representations of geographically weighted summary statistics (Brunsdon et al, 2002) are the basis of our geographically weighted interactive graphics, or ‘geowigs’

3.1 Summary Statistics

A number of summary statistics are proposed in Brunsdon et al (2002) – for example, a geographically weighted mean taken at a point \( \mathbf{u} \) with a bandwidth \( h \) is defined by

\[
M(\mathbf{u}, h) = \frac{\sum w_i(\mathbf{u}, h) x_i}{\sum w_i(\mathbf{u}, h)} \tag{1}
\]

where \( w_i(\mathbf{u}, h) \) is the weight associated with the observed variable \( x \) at location \( i \). If location \( i \) is represented by the vector \( \mathbf{u}_i \) then

\[
w_i(\mathbf{u}, h) = \left( 1 - \frac{||\mathbf{u}_i - \mathbf{u}||^2}{h^2} \right) \tag{2}
\]

is one possible weighting scheme. In a similar manner other geographically weighted statistics – such as a geographically weighted median, or other geographically weighted quantiles, can be defined – see Brunsdon et al (2002) for further details. Note that we often choose \( \mathbf{u} \) to be the centroid of the \( i \)th department here. In this case, we use the notation \( M(i, h) \) to refer to the gw-mean at the centroid of department \( i \) with bandwidth \( h \).

3.2 Graphic Types

The Guerry data set contains six key quantitative variables for each of the 86 departments of France in 1830. Here we use the index \( i \) to refer to the departments, which are mapped by Friendly and Guerry and shaded according to rank (Figure 1).

Guerry regarded high values are being indicative of moral character and so in accordance with Guerry and Friendly’s maps low values are represented by darker shades to reveal ‘la France obscure and la France éclairée’.
Figure 1. Unclassified choropleth maps of Guerry’s six key variables.

Spatial variation in any single variable can be depicted at a range of scales by displaying geographically weighted means. In Figure 2 we show maps of $gw$-means for variable 1.

Figure 2. Geographically weighted maps for variable 1.
Weighting maps (Figure 2, top) show the relative contributions of local departments \( w_{ij} \) – ie the weighting applied to department \( j \) in the calculation of \( M(i,h) \) for a single unit with five increasing values of \( h \). Using consistent symbolism in gw-maps (Figure 2, bottom) shows the effects of increasing \( h \) on the local weighted value of a single statistic.

The univariate gw-maps reveal some trends at particular scales, but it can be useful to compare local variation at a range of scales. We achieve this by generating gw-boxplots at a range of scales from gw-percentiles (Figure 3). The lighter boxplots represent the national figures (consistent across all values of \( h \)). Darker gw-boxplots show local variation for any combination of \( u \) and \( h \), which tend towards the national values as \( h \) increases. The smaller darker circle shows the local value \( z(u) \), the larger lighter circle represents the gw-mean – \( M(u,h) \).

![Figure 3. gw-boxplots for a single department – variable 1 shown at five different scales.](image)

In Figure 3, Creuse \((i=23)\), is evidently a local high at \( h=0.025 \) – as suggested clearly by the map and identified by Guerry, though not an outlier. Also note that ‘local’ variation at scale \( h=0.050 \) in ‘crime against persons’ exceeds that observed in the national data set – a trend that is more subtle than the identification of a local high, the regional analysis of Friendly or the national ‘obscure / éclairée’ classification.

A scalogram uses the x-axis to represent \( h \) and the y-axis for \( M(i,h) \). Lines linking \( M(u,h) \) for all \( h \) for every \( u \) enable us to consider variation in multiple zones at a range of scales for any variable (Figure 4).
The three views in Figure 4 show the mapped data (left), the *scalogram* for Creuse (center), which is an outlier in the original data and the complete *scalogram* for all 86 departments (right). The vertical lines show values of $h$ for which $M(i,h)$ has been calculated. A number of other effects are also identifiable.

### 3.3 Interactions – Software Implementation

Our demonstrator software uses maps, *gw-boxplots* and *scalograms*. A series of interactions and alternative symbolism that enable users to rapidly move between these and other views to interrogate the spatial structure of data sets. Clicking the maps cycles through the variables available (Figure 1) the computed values of $h$ (Figure 2) and the four spatial views (Figure 5).

The four spatial views available are: the data view – choropleth map showing original values (see Figure 1); the *gw-map* – *gw-mean* values for a particular $h$ (see Figure 2); the *gw-residual map* - shows local effects of the geographic weighting at a particular scale for all zones; the *weighting maps* – shades relate directly to the effect of each department.
on the value of $M(u,h)$ (see Figure 2). The gw-effect map uses a diverging scheme as advocated by Harrower and Brewer (2002). When reproduced in greyscale darker shades represent greater variation between original values and local means.

All of the views are linked so that any interaction results in updates to all views and selecting a location on the map causes the gw-boxplots to be centred on that location (Figure 6).

The gw-boxplots are dynamically updated as the map is brushed, supporting rapid comparison and exploration. The departments shown here are the 'outliers' identified in Friendly's paper.

Our software shows gw-boxplots for all selected variables and so these interactions occur for multiple variables concurrently. Figure 7 shows the effects of interactively varying the mapped variable the scale ($h$), the department of interest and the map view whilst undertaking exploratory analysis.

The examples in Figure 7 map variables 1 and 2 respectively and show the effects of highlighting departments 23 (Creuse) and 43 (Haute Loire). Each shows a choropleth of the original values, and then pairs of multivariate gw-boxplots and weighting maps for $h=0.025$ (top) and $h=0.100$ (bottom).

We also display the scalogram at all times and this is also updated dynamically to correspond with any changes in variable or scale (Figure 8). The links are graphically and geographically weighted so that whenever an item in a view is selected through interaction symbolism is changed to emphasize other local items according to their degree of locality (Dykes and Mountain, 2003).
Figure 7. gw-boxplots for two departments for six variables.

Figure 8 shows maps and scalograms for two different combinations of department and variable: $v=1$, $i=23$ (top) and $v=2$, $i=43$ (bottom). In each case the first scalogram is shaded according to the original statistical value recorded for each department. The second scalogram uses geographically weighted shading in which the opacity of the line is varied such that departments that are closest to that which has been brushed with the cursor are visually emphasized. The currently selected value of $h$ is shown by emboldening the appropriate horizontal line and through a weighting map focused on the brushed spatial unit.
Figure 8. *scalograms* with statistical and geographically weighted shading ($z$ and $w_i$).

4. Summary and Conclusions

Here we use geographically weighted interactive graphics, or *geowigs*, to hypothesize about the geographic variation in the data set. We introduce the *gw-boxplot*, the *scalogram*, and geographical highlighting, symbolism and interactions. These approaches build upon existing methods and technologies (Dykes, 1998; Brunsdon, 1998; Brunsdon et al., 2006) and are implemented in demonstrator software that permits further analysis. Our consideration of the geographic and scale-based variation in the Guerry data responds to Friendly’s invitation to rise to Guerry’s visualization challenge. Whilst our focus is predominantly on the six key quantitative variables used in Friendly’s work (Friendly 2006; Friendly in press) our techniques are extensible to consider higher numbers of variables. These methods address the need for graphical displays of multivariate local variation that consider of ‘neighbourhoods’ in a flexible manner (Unwin and Unwin, 1998).

Acknowledgments

Michael Friendly’s help in providing the Guerry data set is gratefully acknowledged.

References


