Geographically Weighted Visualization: Interactive Graphics for Scale-Varying Exploratory Analysis

Jason Dykes and Chris Brunsdon

Abstract — We introduce a series of geographically weighted (GW) interactive graphics, or *geowigs*, and use them to explore spatial relationships at a range of scales. We visually encode information about geographic and statistical proximity and variation in novel ways through *gw-choropleth maps*, multivariate *gw-boxplots*, *gw-shading* and *scalograms*. The new graphic types reveal information about GW statistics at several scales concurrently. We impement these views in prototype software containing dynamic links and GW interactions that encourage exploration and refine them to consider directional geographies. An informal evaluation uses interactive GW techniques to consider Guerry's dataset of 'moral statistics', casting doubt on correlations originally proposed through visual analysis, revealing new local anomalies and suggesting multivariate geographic relationships. Few attempts at visually synthesising geography with multivariate statistical values at multiple scales have been reported. The *geowigs* proposed here provide informative representations of multivariate local variation, particularly when combined with interactions that coordinate views and result in *gw-shading*. We argue that they are widely applicable to area and point-based geographic data and provide a set of methods to support visual analysis using GW statistics through which the effects of geography can be explored at multiple scales.

Index terms — Geographical weighting, exploratory data analysis, scale, multivariate, directional, interaction, coordinated views

1 Introduction

Definitions of geovisualization and information visualization suggest mutual exclusivity – yet in practice much geovisualization is dependent on and informed by aspatial graphics and many views that combine elements of geography and statistical spaces provide useful insights. The value added in combining these elements to help explain geographical processes is that exploring relationships between measured phenomena that are 'near' in statistical space and geographical space may provide insights that cannot be achieved with maps or statistical graphics alone. The definition of 'nearness' is fundamental to such approaches: How near do two houses have to be before the selling price of one influences the selling price of the other? How near does an area with a high crime rate have to be from a house to influence its selling price? Is the relationship a simple monotone or might there be more than one 'critical distance' at which interactions occur? And do these relationships vary across space and with direction? These kinds of questions consider geographical interaction between variables but also imply that an investigation of scale must be considered.

A popular approach to the investigation of geographical patterns is the concept of *local statistics* [11,18] perhaps arising from Openshaw's stated dissatisfaction with 'whole map statistics' [15]. A point location **u** in a study area is selected, and some statistical technique applied to the data weighted by proximity to **u**. Applying this procedure to a number of locations spanning the study area gives an indication of the spatial variability in the distribution of the data values. Such approaches have incorporated not only formal inferential methods, such as the *g*-statistic [11], or *geographically weighted regression* (GWR) [1], but also descriptive tools, such as *geographically weighted summary statistics* [2]. The latter approach allows us to consider spatial patterns of attributes through localized descriptive statistics and is suitable for considering the geography of a statistical data set in the context of exploratory enquiry.

For the 'geographically weighted' approaches, the rate at which weighting reduces with distance is controlled by a parameter h: typically h is either a fixed distance for all **u**, or chosen to equal the distance from each **u** to its kth nearest neighbour [1]. In the former case the scale of localisation is fixed for all **u** in physical space, and in the latter it is fixed for all **u** in areas of a given density. Mapping the results for each location for some given h or k is an effective way of showing local changes in the distribution of some measured attribute, or changes in the relationships between several of these. However, a number of spatial processes operate at several scales simultaneously. Consider for example house prices which may exhibit patterns within streets, between parts of a town and between national regions, or topographic features whereby a local peak may be part of a larger valley when measured at a wider scale [20].

It can be helpful to consider the effect of varying either k or h in order to identify all scales at which patterns operate, particularly when spatial data are analysed in an exploratory situation. Fotheringham *et al.*[8] describe the process of so doing as involving a 'spatial microscope' whose focal length may be varied to allow the identification of patterns at different scales. For example, Foody [7] uses GWR to explore the relationship between species richness and three explanatory environmental variables in sub-Saharan Africa, and considers how spatial patterns in the geographically varying regression parameters change in association with h in order to investigate the underlying geographical scale of the relationships between these variables. Scale-space analysis [13] provides a multi-scale framework for analysing data recorded in a regular rectangular array in this manner and is used for feature recognition. It may, along with other pixel-based approaches, be applied to data recorded in the kinds of irregular zones that are typical of many geographic data sets [17].

While static mapping may be useful for investigating spatial variation for a fixed k or h [2], visualizing variability of the spatial patterns with h, is arguably too complex a task for conventional cartography, particularly where multiple variables are considered concurrently. Interactive graphics provide opportunities for exploring complex structured data sets and generating knowledge from them [16]. For example, interactive graphical techniques have been developed to show the effects of scale on local surface derivatives [21]. Here we extend these ideas by developing complimentary graphical techniques, with coordinated behaviours and interactions through which these can be accessed. These new methods allow us to explore variations in h and are applicable to multivariate data sets

- *Dr. Jason Dykes is based at the giCentre, Department of Information Science, City University, London EC1V OHB, Email: jad7@soi.city.ac.uk*
- *Prof. Chris Brunsdon is based at the Department of Geography, University of Leicester, Leicester LE1 7RH, Email: cb179@leicester.ac.uk*

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 27 October 2007.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

measured using irregular zones. They are suitable for investigating the properties of spatial data at a range of scales. They include standard choropleth maps, maps of geographic weightings for any h (*weighting maps*) maps of geographically weighted means (*gw-mean maps*) and local variations from the mean (*gw-residual maps*), a localised (GW) version of the box-and-whisker plot that we term a *gw-boxplot* and a *scalogram* – a plot showing the variation of localised summary statistics as the value of h changes. Our prototype software contains interactions and visual encodings that are both geographic and geographically weighted. In combination these views and coordinated interactions are designed to help analysts gain insight into spatial patterns through which geographic processes may be characterised and understood.

2 Context : Challenges for Multivariable Spatial Analysis – Guerry's Moral Statistics of France

By way of example 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*' [19]. Guerry meticulously collected and importantly *related* data on a number of themes for the departments of France to analyse social issues in the early 19th century. He considered geography predominantly by visually inspecting his univariate choropleth maps and identified geographic outliers and some regional trends. Visual inspections were used to hypothesize about relationships between variables. Friendly [9] uses statistical and graphical methods to revisit Guerry's data set. Regression analysis shows that some of Guerry's postulated associations do not hold and that others omitted by Guerry exist. Geography is considered in a hierarchical manner by comparing data for departments in each of the five regions of France and suggestions are made about relationships between these regional aggregations [10]. Multivariate graphics and conditioned choropleths [4] are used to augment Guerry's univariate maps.

3 Geographically Weighted Graphics

The geographically weighted interactive graphics, or '*geowigs*', developed here are graphical representations of geographically weighted summary statistics.

3.1 Summary Statistics

A number of geographically weighted summary statistics can be calculated at locations in a spatial data set [2]. The most fundamental of these is the geographically weighted mean. This is simply a moving spatial window mean smoother. If we have a number of observations with values x_i at points $\mathbf{u}_i$, then the geographically weighted mean at any point $\mathbf{u}$ is

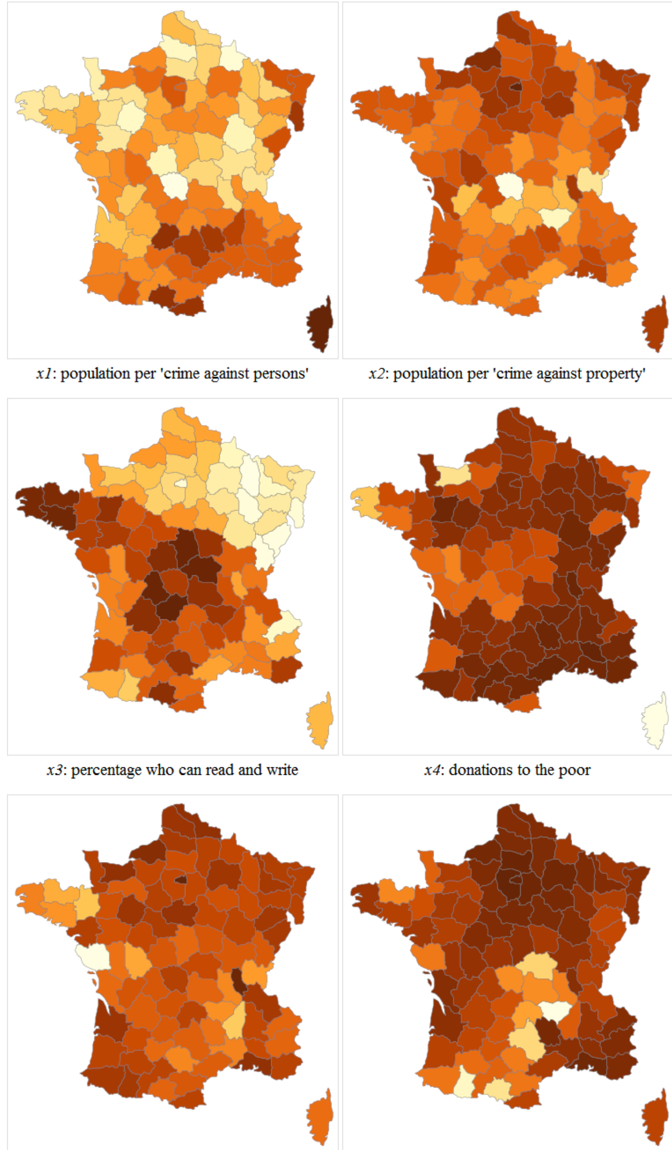
$$M(\mathbf{u},h) = \frac{\sum x_i w_i(\mathbf{u})}{\sum w_i(\mathbf{u})} \qquad (1)$$

where $w_i(\mathbf{u})$ is the weight applied to the observation at location $\mathbf{u}_i$ when computing the geographically weighted mean at location $\mathbf{u}$. This weight is typically a monotone decreasing function of the distance between $\mathbf{u}$ and $\mathbf{u}_i$ with h a parameter controlling the rate at which the decay occurs. h is referred to as the *bandwidth* of the locally weighted statistic – it effectively controls the size of the moving window. Here, we use a Gaussian decay function:

$$w_i(\mathbf{u}) = \exp\left(-\frac{|\mathbf{u}-\mathbf{u}_i|^2}{2h}\right) \qquad (2)$$

This has a maximum value of one when $\mathbf{u} = \mathbf{u}_i$. We can simplify the formula for $M(\mathbf{u},h)$ if we replace the standard weights by a set of weights that sum to one, so that

$$W_i(\mathbf{u}) = \frac{w_i(\mathbf{u})}{\sum w_i(\mathbf{u})} \text{ and } M(\mathbf{u},h) = \sum x_i W_i(\mathbf{u}) \qquad (3)$$



x1: population per 'crime against persons' *x2*: population per 'crime against property'

x3: percentage who can read and write *x4*: donations to the poor

x5: population per illegitimate birth *x6*: population per suicide

Fig. 1. Unclassified choropleth maps of Guerry's six key variables.

This last expression demonstrates that a geographically weighted mean is the mean of a mass-point distribution with a discrete set of (value, probability) pairs $L = \{(x_i, W_i)\}$. The other locally weighted summary statistics of importance here are the distributional quantiles associated with the construction of boxplots. They are computed in terms of the quantiles of the mass-point distribution L. If the (x_i, W_i) pairs are ordered in increasing values of x_i then the cumulative distribution of L makes a series of discrete 'jumps' at each value of x_i – jumping from zero to W_1 at x_1, then to $W_1 + W_2$ at x_2 and so on. Thus each point x_i is the $(W_1 + W_2 + \ldots + W_i)$th quantile of the distribution and other specific quantiles (such as the median or quartiles) can be derived by linear interpolation.

These statistics form the basis of our *geowigs*. In the case of area-based data such as that collected and analysed by Guerry we may select the centroids of our zones as the positions at which $\mathbf{u}$ is calculated. We use the notation $M(\mathbf{u}_i,h)$ to refer to the *gw-mean* at the centroid of department with index i at a 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. These are shaded according to rank in Friendly and Guerry's maps. We symbolise the ratios using unclassified choropleths (Figure 1). The colour schemes used throughout are continuous and based upon Brewer's sequential and diverging schemes [12]. Guerry regarded high values are being indicative of moral character and so in accordance with Guerry and Friendly's maps low data values are represented by darker shades to reveal '*la France obscure and la France éclairée*'. We use the YlOrBr scheme [12] for maps of original statistical values and GW means. Spatial variation in any single variable can be depicted at a range of scales by displaying GW means. Choropleth maps for variable 1 are shown in Figure 2 with five values of h increasing from left to right. $M(\mathbf{u},h)$ is precomputed for selected values of h in our implementation. The *weighting maps* (top) show the relative contributions of local departments w_{ij} in the calculation of $M(\mathbf{u},h)$ for a single unit using the YlGn scheme [12] – in which zones making no contribution to the statistic are light shaded yellow. Symbolism that is consistent across the resultant *gw-mean maps* (middle) shows the absolute effects of increasing h on the local weighted value of a single statistic. Symbolism that varies between *gw-mean maps* (bottom) shows the relative effects of varying h. Each map shows the local statistic for 'population per crime against persons' with 'nearness' being considered more widely successively from left to right. The value computed for any location $\mathbf{u}_i$ is dependent upon the geography considered in calculating the statistic. In this case the values are calculated for the 86 department centroids.

The univariate *gw-mean maps* reveal some regional 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, as discussed in section 3.1. In Figure 3 the light grey boxplots represent the national figures (not weighted and consistent across all values of h). The green *gw-boxplots* show local variation for any combination of $\mathbf{u}$ and h, which tend towards the national values as h increases (from left to right). The smaller grey circles represents the local value $x(\mathbf{u})$, the larger green circle represents the GW mean – $M(\mathbf{u},h)$.

The geographically weighted box and whisker plots visually encode the GW median and GW percentiles. A series of these allows multivariate GW relationships to be considered at a variety of scales. The *gw-boxplots* contrast with the maps in that they are projected in a statistical space with the y-axis representing statistical values and the x-axis ordering a series of alternative *gw-boxplots* by variable (x) or scale (h).

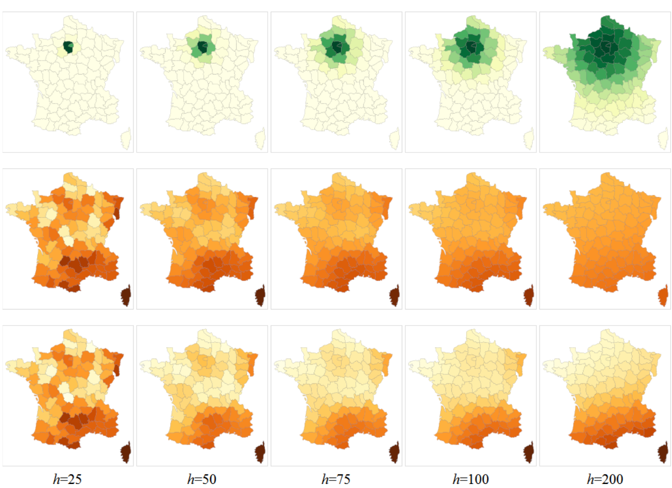


Fig. 2. Geographically weighted maps for variable 1 – 'population per crime against persons', with h increasing from left to right. From top to bottom: *weighting maps* for one department; *gw-mean maps* using a common colour scheme (for considering inter-map variation); *gw-mean maps* using a unique colour scheme for each value of h (for considering intra-map variation). The maps on the right hand side of the figure give a good indication of the range of shades used in the YlGn (top right) and YlOrBr (bottom right) colour schemes [12].

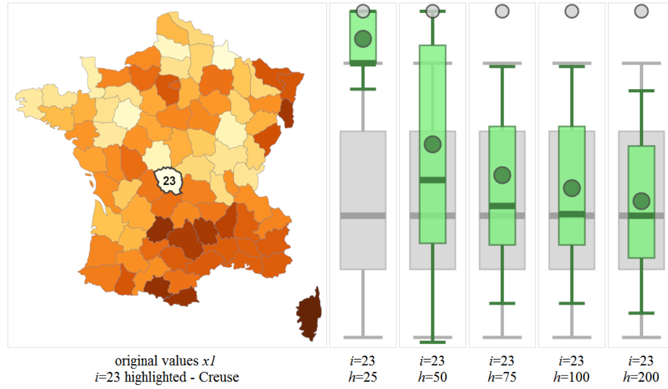


Fig. 3. *gw-boxplots* for a single department – variable 1 'population per crime against persons' at five different scales (h=25 to h=200).

Figure 3 shows *gw-boxplots* for i=23, Creuse, at five scales. The department is evidently a local maximum at h=25, as suggested clearly by the light zone on the map and as identified by Guerry, though it is not considered a statistical outlier in the boxplot. This maximum is not persistent across scale however. Locations to the north have higher GW mean values at a scale of h=50 and several departments to the east have higher GW means at h=100 as shown in the *gw–mean maps* of Figure 2. Not unrelated is that fact that 'local' variability in 'crime against persons' around Creuse at scale h=50 is greater than that observed in the national data set. These trends and findings are more subtle and more spatial than those detected in previous analyses, including Guerry's identification of a local high, the regional analysis of Friendly and the national '*obscure / éclairée*' trend. The way that local maxima depend upon the definition of 'nearness' and that a process operating at a local scale (between h=25 and h=50) focused on Creuse is resulting in more variation than is evident in the wider national data set may be of some interest and is indicative of the kind of knowledge that might be elicited from *geowigs*.

The *scalogram* is designed to help explore such variations and the scales at which they occur. Whilst the *gw-mean maps* are projected in geographical space, the *scalogram* fills an abstract space with orthogonal geographic and statistical dimensions. The x-axis is used to represent h and the y-axis for $M(\mathbf{u},h)$. Lines linking $M(\mathbf{u},h)$ for all h for every $\mathbf{u}_\mathbf{i}$ enable us to consider variation in multiple zones at a range of scales for any variable. The three views in Figure 4 show the original data mapped as a choropleth (left), the scalogram for i=23 – Creuse, which is the maximum value in the original data (center), and the complete scalogram for all 86 departments (right). The vertical lines in the *scalograms* show values of h for which $M(\mathbf{u},h)$ has been computed. The path of the single line in the central *scalogram* of Figure 4 corresponds with the GW means in Figure 3 (the green circles). Flatter sections of the lines in the full *scalogram* (right) denote departments where varying scale or the definition nearness has little effect on the GW value.

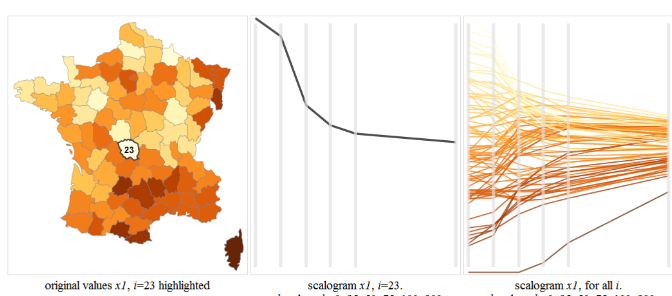
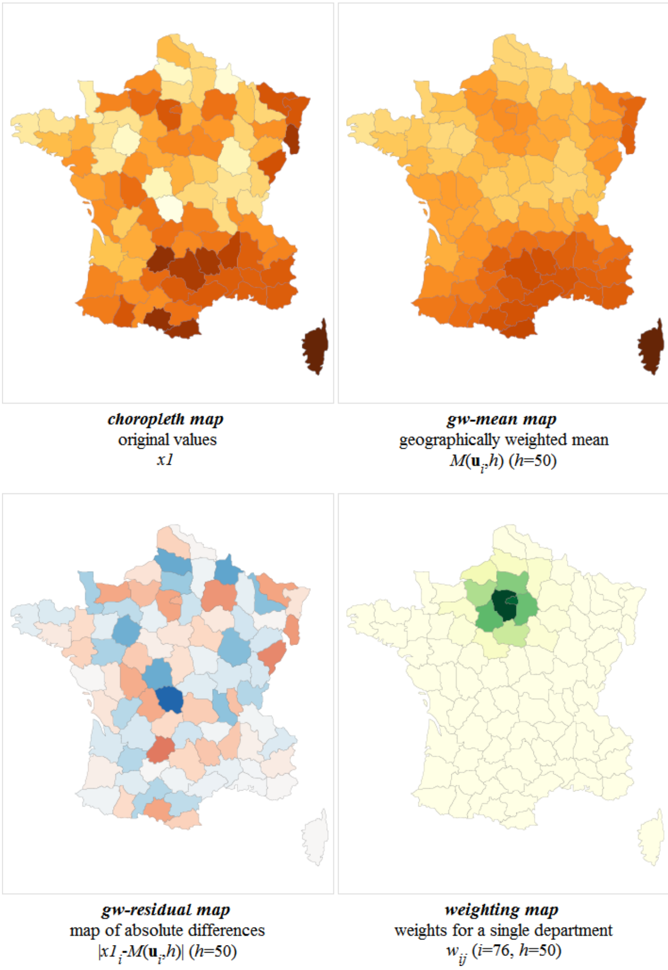


Fig. 4. *scalograms* for variable 1 – 'population per crime against persons'. A single department (centre) and all departments (right).



5. Map views – choropleth map of original values, *gw-mean map*, *gw-residual map* and *weighting map*.

The way in which the GW mean centred on Creuse decreases as our definition of 'nearness' increases in scale is shown by the falling profile in Figure 4 (centre). The reducing dominance of the location as a 'high' as we expand our definition of 'nearness' is reflected in the profile dipping beneath those of other departments in Figure 4 (right). The bottom line in Figure 4 (right) is Corsica, and the lack of any local effect until h=100 relates to the geographic isolation of the island. Steeper line sections within a scalogram draw attention to local variability at a particular scale and show the scale at which this occurs. These are characteristic of a number of departments. Some departmental scalogram profiles contain local maxima or minima identifying a scale at which local variation is particularly atypical.

3.3 Interactions – Software Implementation

Our demonstrator software contains three linked views: a map, *gw-boxplots* and a *scalogram*. A series of interactions supports rapid navigation between alternative visual encodings within these graphic types to interrogate the spatial structure of data sets. Clicking the maps cycles through the variables available (Figure 1), the precomputed values of h (Figure 2) and the four spatial encodings (Figure 5) with related updates in all other views.

The four spatial views available are: the *choropleth map* – sequential shades relate to the original values (see Figure 1); the *gw-mean map* – colours show the GW mean values for a particular h using constant (comparable) or map-specific sequential shading schemes (see Figure 2); the *gw-residual* map – showing local effects of the geographic weighting at a particular scale for all zones with a diverging scheme (see Figure 5) – we use the RYB scheme [12] where red is positive, with values higher than the locality denoting a local high and blue is negative, denoting a local low ; the *weighting map* – a sequential scheme shows the relative weightings of all departments in localities based upon a particular source department – the visual emphasis relates directly to the contribution of each department to M($\mathbf{u}$,h) (see Figure 2).

All of the views are coordinated so that any interaction that changes x (the mapped variable) or h (the scale used in weightings) results in appropriate updates to all views. Links between the *scalogram* and map are dynamically updated as the map is clicked; selecting a location on the map updates the *gw-boxplots* and *weighting map* so that they are centred on the relevant zone (Figure 6). *gw-boxplots* for five departments are shown in Figure 6 to highlight the spatial differences in scale effects for a single variable. The departments selected are the 'outliers' identified by Friendly.

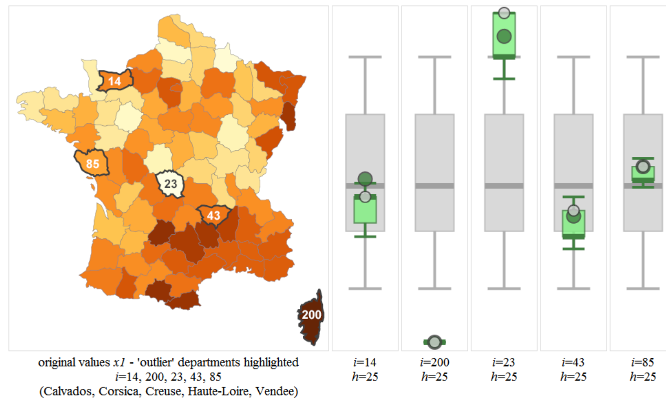


Fig. 6. Choropleth and *gw-boxplots* for one variable at a single scale – five different departments, those selected identified as 'outliers' [10].

Our software shows *gw-boxplots* for all variables in the data set simultaneously and so the interactions described occur for multiple variables concurrently. Figure 7 shows the effects of interactively changing the mapped variable (x), the scale (h) and the zone of interest (i) whilst undertaking exploratory analysis.

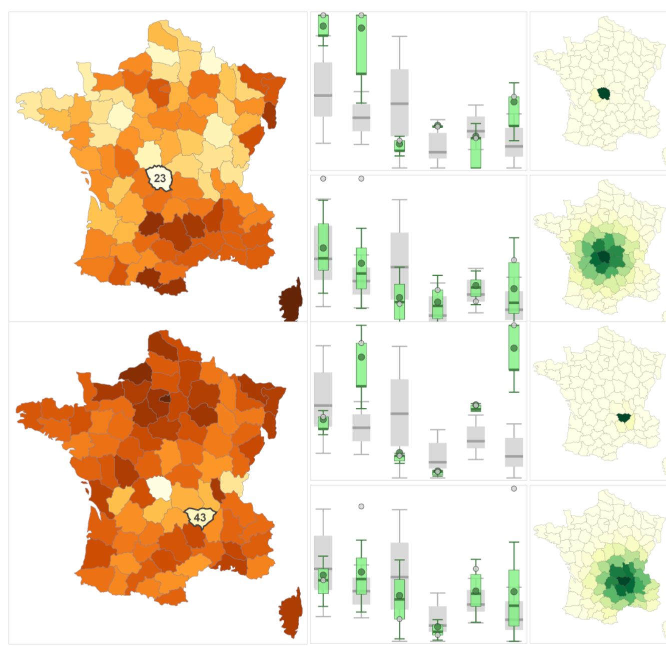


Fig. 7. Maps for two variables and *gw-boxplots* for all six variables for two departments at two scales. These two examples map variables 1 and 2 respectively and show the effects of highlighting departments 23 (Creuse) and 43 (Haute Loire). Each shows a choropleth of the original values, and then pairs of multivariate *gw-boxplots* and *weighting maps* for h=25 (top) and h=100 (bottom).

Shading used in the map is reflected in the *scalogram* and so when *weighting maps* are displayed, colour encodings in the scalogram emphasize items according to their weighted contribution to the locality [6]. Figure 8 illustrates with maps and *scalograms* at h=50 for two variables: *x1* (top set) and *x2* (bottom set). In each case the first scalogram is shaded according to the original statistical values of the selected variable (x) recorded for each department (i). The second scalogram uses *gw-shading* – GW highlighting in which the colour of each line is varied such that departments closest to that selected on the map are visually emphasized. The currently selected values of h and i are shown by emboldening the appropriate vertical line in the *scalogram* and through a *weighting map* focused on the brushed spatial unit. In Figure 9, the graphics use the format shown in Figure 8 and focus on i=69, Rhône. The peak in the darkest curve (i=69) indicates a scale effect that may be of interest.

These visual encodings constitute new graphic types for geographic enquiry. When combined, the dynamic features of these geographically weighted interactive graphics or '*geowigs*' provide the basis for exploring local variations in the effects of scale on a series of geographic variables. This configuration supports rapid comparison and exploration. Computing the geographically weighted statistics and subsequent use of visual mappings provides a spatial perspective on the analysis of geographic data. In terms of visual information seeking we are filtering by geography in two ways – by location (**u**) and nearness (h) and provide graphical details on demand (*gw-boxplots* and *scalograms*). The geographic nature of our enquiry requires the Gestalt of the graphical overview and so maps are provided concurrently. Dynamic brushing helps relate graphical overview with graphical detail as data are filtered.

4 Geographically Weighted Analysis

The graphic types introduced here show that the identification of maxima, minima and notable 'outliers' is scale dependent when considering geographic data and draw attention to these dependencies. Our preliminary analysis allows us to suggest patterns at a range of scales. Individual departments identified as 'outliers' in Guerry's data set may not be atypical for all variables at all scales (Figure 7). The shapes of the scalograms help us relate local and national variations in single variables. For example, when comparing the full scalograms in Figure 8, the wider range of values of $M(\mathbf{u},h)$ at h=200 for *x1* than *x2* suggest that trends measured at a national scale are more dominant in case of the former variable ('crime against persons') than the latter ('crime against property').

Equally, we can identify particular scales at which local effects occur. For example, in Figure 9, Rhône (i=69) has a low value for *x2* at low values of h, but is part of a wider locality with values of 'crime against property' above the national average where h=50 and h=75. The peak in the curve suggests that, relative to other departments, Rhône is a local high for 'crime against property', but only when considered at certain scales. These scales do not correspond to either the highest resolution available (the departments for which the data were originally collected) or a formal regional scale (the regions into which departments are aggregated in the administrative hierarchy [10]). The peak suggests that there is more intra-regional variation than inter-regional variation in this variable at this location, enabling us to detect a scale at which the geography of 'crime against property' may operate here. It is important to note that a more traditional approach to identifying visual patterns would prescribe a scale at which to view patterns, most likely settling for either departments or regions, rather than exploring a range of scales. In this case, a pattern has occurred at a scale not coincident with typical 'official' reporting scales for statistical mapping. Another difference with the more 'traditional' hierarchical approach to scale is that aggregation is spatially abrupt – neighbouring departments may be related, but this information is lost if a neighbour pair consists of two departments in different regions. The 'moving window' approach, where larger scale neighbourhoods move continuously with **u**, overcomes this problem.

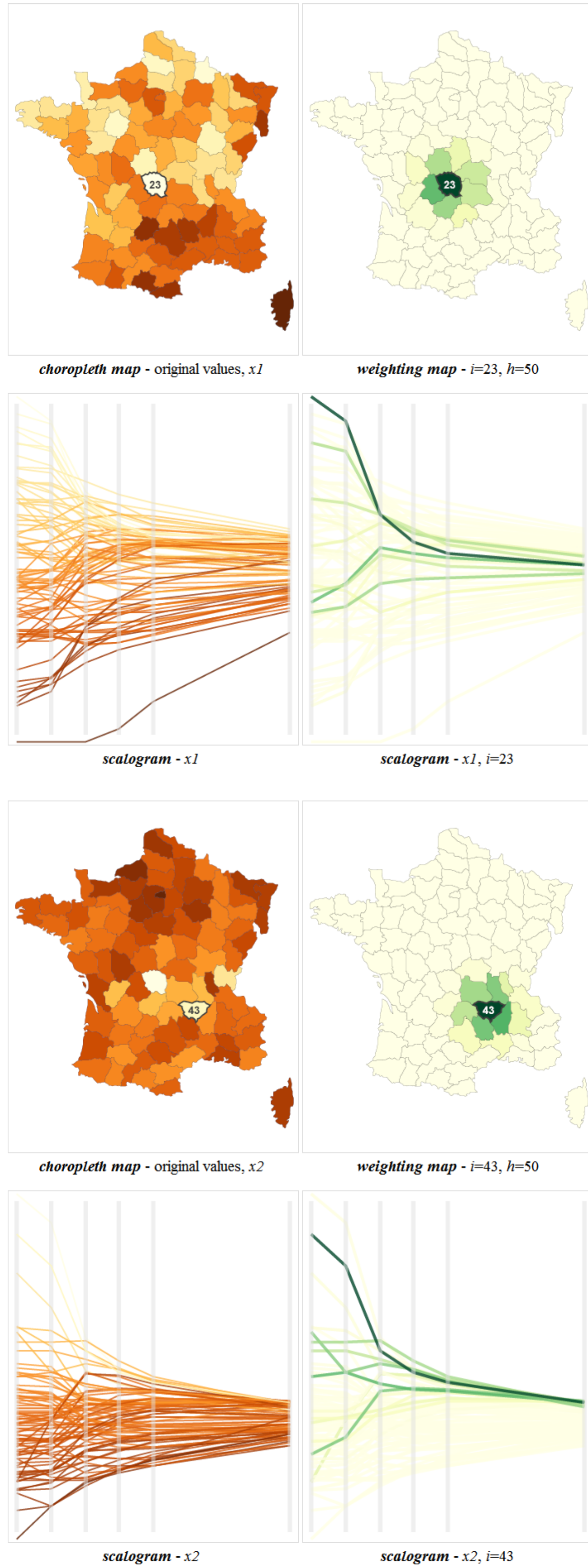


Fig. 8. *scalograms* with statistical and geographically weighted shading (x_i and w_{ij}). i=23, *x1* (top) and i=43, *x2* (bottom). The pairs of figures show maps and *scalograms* for two variables: *x1* (top quartet) and *x2* (bottom quartet).

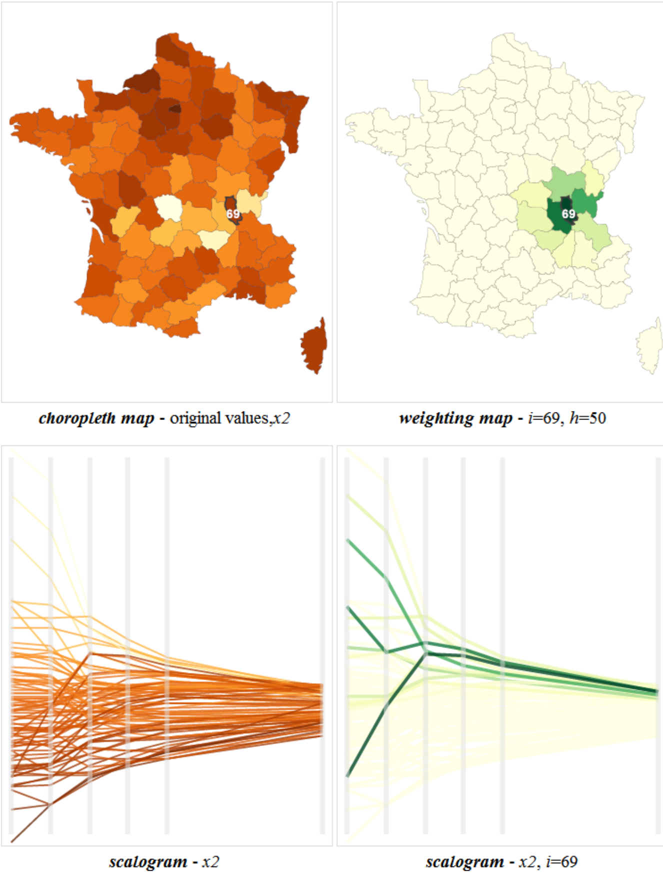


Fig. 9. *scalograms* with statistical and geographically weighted shading (x_i and w_{ij}). i=69, $x2$. The graphics use the format shown in Figure 8 and focus on i=69, Rhône. The peak in the darkest curve of the scalograms (i=69) indicate a scale effect that may be of interest.

Another feature of the interactive nature of the *geowigs* presented here is the ability to 'strum' the set of scalogram curves – here 'strumming' means running the cursor quickly up or down the curves while brushing. Highlighting each curve in quick succession gives an indication of how unusual the shape of the scalogram curve is. Not only has this approach highlighted an effect at an unexpected scale, in the case of Rhône, it has also demonstrated that this is localised to one particular area. Processes may work at different scales in different places.

Other geographic patterns are detectable in the GW views. Loire-Inferiure is notable as being unlike its immediate locality (which has nationally high values) in variable 5 (illegitimate births) despite having an individual value close the national statistical average; Lozere and Rhône are local highs and local lows in variables 6 ('suicide') and 2 ('crime against property') respectively and Loiret is spatially invariant when variable 3 ('literacy') is considered. These are not the kinds of relationships that have been identified in previous analyses of the data set and are detectable through the multi-scale geovisualization techniques introduced here.

The serial display of boxplots enables local comparison between multiple variables at a range of scales. For example, the degree of variance (i.e. GW interquartile range) for the variables 'population per crime against property' (*x1*) and 'population per crime against the person' (*x2*) is relatively large when centred on Rhône at low values of h, but less significant for both variables at nearby Saone-et-Loire. Probing the map and considering the *gw-boxplots* indicates that a third variable relating to literacy ('percent read and write' – *x3*) maintains a more constant GW interquartile range in the vicinity of Rhône then either *x1* or *x2*.

5 Directed Geographic Weighting

Weighting has been applied isotropically thus far, with weighting functions depending only on the distance between a pair of points, and not on their angular separation. However direction, as well as distance, plays a key role in the study of many geographical phenomena. In the domain of physical geography there are many examples, such as wind speed, direction of water flows and orientation of faultlines. Examples in human geography include population migration flows and the direction of commuting patterns. When such phenomena and processes are considered, similarity between a pair of places may not just depend on their distance apart, but also on the direction from one place to the other. For example, if this direction were coincident with population flow, one might expect two places to have more social characteristics in common than otherwise.

We extend the idea of geographical weighting to incorporate direction as well as distance by pre-multiplying the expression for w_i by the expression

$$\exp\left(-\lambda\cos\left(\theta_i-\varphi\right)\right) \tag{4}$$

Here, θ_i is the angle of the path between locations $\mathbf{u}$ and $\mathbf{u}_i$ and φ is the principal direction of the weighting (angles are measured counter-clockwise from the east). For a fixed distance, $\mathbf{u}_i$ values along this direction from $\mathbf{u}$ will gain the highest weighting. Those in the opposite direction will receive the lowest weighting. The parameter λ controls the relative sharpness of the directional effect. Setting λ to zero removes any directional bias in the weighting, while increasing it results in an expansion of the ratio between the highest and lowest weight values.

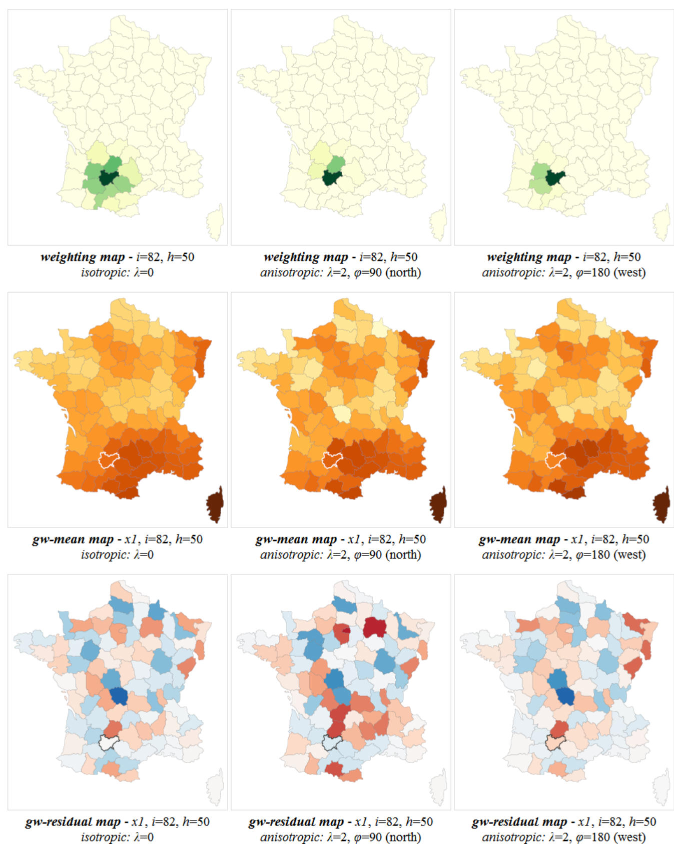


Fig. 10. *weighting maps (top)*, *gw-mean maps* (middle) and *gw-residual maps* (bottom) using three different directional weightings: isotropic, λ=0 (left); φ=90, λ=2 – 'north' (centre), φ=180, λ=2 – 'west' (right). The *weighting maps* focus on i=82 Tarn et Garonne, which is highlighted in the other views and can be seen to vary in terms of the relationship between original value and local mean with direction.

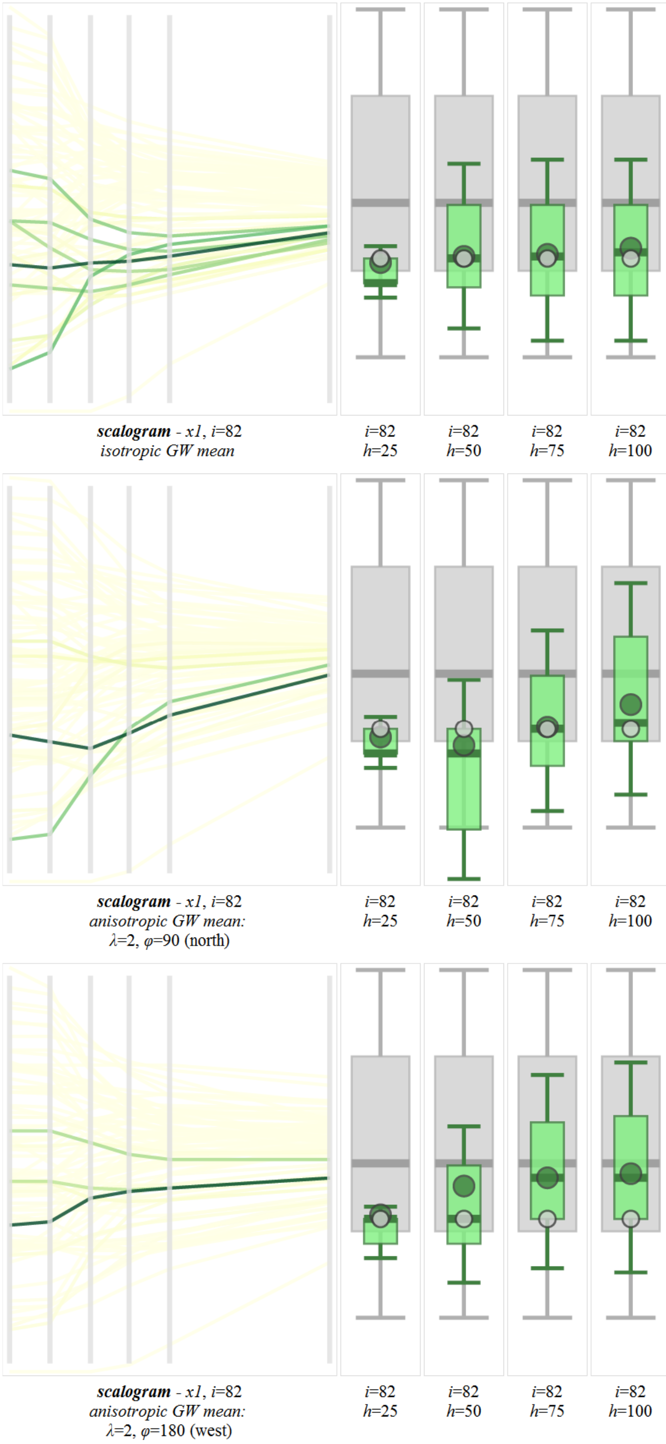


Fig. 11. *gw-shaded scalograms* and multi-scale *gw-boxplots* for *i*=82, Tarn et Garonne using three different directional weightings: isotropic, λ=0 (top); φ=90, λ=2 – 'north' (middle), φ=180, λ=2 – 'west' (bottom). The *gw-shading* in the *scalograms* relates to *h*=50 in all cases.

To reduce computation time, we only consider φ values on so-called 'clock points' – that is angles in the set $\{0,\pm\pi/6, \pm\pi/3, \pm\pi/2,\pm2\pi/3,\pm5\pi/6,\pi\}$. Statistics computed using this approach will be referred to as *directed* GW statistics. Angles are measured counter-clockwise from east and referred to in degrees, as is usual in the geographical literature. We can use this directional extension to geographical weighting with our *geowigs* – to explore the data as before, but investigating effects associated with specific directions as well as isotropic patterns.

For example, department 82, Tarn et Garonne, shows interesting variation when the directionless GW mean is compared with *directed* GW means at 90^o (north) and 180^o (west) at multiple scales. Whilst Tarn et Garonne has a low value of variable *x1* ('crime against persons') when compared to the national distribution at a scale of *h*=50, the directionless GW mean indicates a value close to the local median and the department is thus locally typical (Figure 10). If we consider the directed statistics however, more spatial structure is revealed. With λ = 2 and φ = 90 (*directed* GW mean with bias towards the north) Tarn et Garonne is identified as a local directional low due to the high value of the northern neighbour – department 46, Lot. With λ = 2 and φ = 180 (*directed* GW mean with bias towards the west) Tarn et Garonne is identified as a local directional high due to the lower values of neighbours to the west, particularly the immediate neighbours – Lot et Garonne (*i*=47) and Gers (*i*=32). There is evidently directional structure in the local statistics associated with Tarn et Garonne, which exhibits variously typical, low and high values of 'population per crime against person' depending upon the uniform, northern and western directional biases used in calculating the local statistic. This explains the colour profile of outlined department 82, Tarn et Garonne, in the *residual maps* of Figure 10.

These differences are not consistent across scales. The local statistics for Tarn et Garonne are relatively independent of scale from *h*=50 upwards when the directionless local statistic is considered, but when a northern bias is applied to the statistics (λ=2 and φ=90) the considerable variation in values to the north results in a profile showing that Tarn et Garonne varies from local low, to local average to local high through scales from *h*=50 to *h*=100. There is less scale dependency in the western biased GW means (λ=2 and φ=180) as there is less variation in values with scale to the west of the department. Here, Tarn et Garonne is regarded as a local high across scales from *h*=50 upwards (at *h*=25 neighbouring zones make little contribution to the local statistic). These relationships are depicted in the *gw-shaded scalograms* and multi-scale *gw-boxplots* shown in Figure 11.

A single measured value in space can thus be considered low, high and typical in relation to its neighbours and these characteristics can be variously scale dependent and scale invariant when direction is considered. These differences are important because as we have seen, geographic phenomena are often directional. This *directional* analysis uses *geowigs* to demonstrate some of the complexity associated with geographic relationships – spatial processes are neither scale invariant nor isotropic. When we consider a value recorded in space its relationships with 'near' neighbours depend upon definitions of locality and directional emphases. Such differences are difficult to detect in standard choropleth maps and the GW statistics can help with our exploratory spatial data analysis. Interactive visualization of the type described here in our *geowigs* allows us to explore various scale-based definitions of locality and nearness and to investigate the effects of directional bias upon relationships between attributes recorded in geographic spaces.

6 Summary, Discussion and Conclusions

Geographically weighted interactive graphics are introduced as a means of hypothesizing about the spatial variation in a geographic data set. We introduce *gw-maps*, *gw-boxplots*, the *scalogram*, and *gw-shading* in conjunction with appropriate geographic interactions. These approaches build upon existing methods and technologies [1,2,3,5,6] 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. The *geowig* techniques provide methods that may be used to compliment the analysis of Geurry, Friendly and others in exploring the spatial structure of area based data at a range of scales and address an identified need for graphical displays of multivariate local variation that consider 'neighbourhoods' in a flexible manner [18]. Whilst our focus is predominantly on the six key quantitative variables used in Friendly's work [9,10] the techniques are extensible to higher numbers of variables.

There is considerable potential for extending and exploring the techniques outlined here and using them further. For example, it may be helpful to benchmark *geowigs*, by considering their performance when applied to a test bed of geographical data sets – investigating the change in outcome when different sizes and shapes of geographical regions are used. Doing so may help explore the dependence of the GW mean on the geometry of nearby regions at low values of h and relate this effect to those associated with the aggregation of continuous geographic phenomena into discrete irregular units for enumeration. The locally weighted statistics are computed on centroids associated with irregular units in our implementation. The statistics could, however be centred on any points, such as those comprising a regular grid. This might be useful for speeding up certain visualisation routines, particularly when large numbers of spatial units are involved, and in such situations it would not be necessary to compute a local statistic for each unit. Comparing *geowigs* generated from discrete units with those produced from more regular continuous representations of phenomena in geographic space (perhaps using centroids re-allocation techniques [14]) would enable us to explore a number of issues relating to the scale effects associated with alternative models of geographic information. Additionally, our focus here has been predominantly on highlighting multi-scale effects for individual variables, however, one could extend the ideas of Brunsdon *et. al.* [2] by measuring local multivariate patterns and mapping a GW correlation coefficient. This would enable the effect of local bivariate association to be explored at different scales, using a *scalogram* based upon GW correlation instead of the GW mean. Each of these techniques could be used with non-Euclidian distances, such as estimated travel time – perhaps a more realistic indicator of potential human interaction.

In commenting on Guerry's original maps, Friendly [10] states that "at the very least, this work testified to the importance of detailed data, sensibly presented, to inform the debate on the relations of crime and education." The techniques presented here are intended, at the very least, to draw attention to geographically weighted graphics and interactions in visualization and to inform debate on the possibilities for using interactive graphics to explore and reveal spatial structure at multiple scales in multivariate and anisotropic geographic data. By adapting ideas from scale-space analysis to irregular, multivariate data we have developed techniques for geographers and policy analysts to explore regional variability in relationships between social variables. These techniques provide instruments for helping uncover processes operating in specific localities and at particular scales and can draw attention to some of the subtleties of spatial information as it encodes geography – such as the effects of using zones of irregular shape and of dealing with scale using an aggregated hierarchical approach. We hope that they may improve understanding of our models and our geography and support informed decision-making.

Acknowledgments

Michael Friendly's efforts in documenting Guerry's work and help in providing the Guerry data set are gratefully acknowledged. The constructive critical comments of four anonymous referees were also greatly appreciated. Their contributions have helped us considerably in refining our work and communicating our ideas and their implications. The time and effort that these colleagues have invested in engaging with our ideas has helped us to develop our work considerably.

References

[1] Brunsdon, C., Fotheringham, A.S. and Charlton, M.. (1996) Geographically weighted regression: A method for exploring spatial nonstationarity. Geographical Analysis, 28:281–289.

[2] Brunsdon, C., Fotheringham, A.S. and Charlton, M., (2002) Geographically weighted summary statistics – a framework for localised exploratory data analysis. Computers, Environment and Urban Systems, 26(6):501–524.

[3] Brunsdon, C. (1998). Exploratory Spatial Data Analysis and Local Indicators of Spatial Association with XLISP-STAT. The Statistician 47, 471-484.

[4] Carr, D., White, D., and MacEachren, A. (2005). Conditioned choropleth maps and hypothesis generation. Annals of the Association of American Geographers 95, 32-53.

[5] Dykes, J.A. (1998). Cartographic Visualization: Exploratory Spatial Data Analysis with Local Indicators of Spatial Association Using TcI/Tk and cdv. The Statistician 47, 485-497.

[6] 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, 581-603.

[7] Foody, G.M., (2004) Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-saharan endemic avifauna. Global Ecology and Biogeography, 13:315–320.

[8] Fotheringham, A.S., Brunsdon, C. and Charlton, M., (2002) Geographically Weighted Regression: the Analysis of spatially varying relationships. Wiley, Chichester.

[9] Friendly, M. (2006). André-Michel Guerry's Moral Statistics of France: Challenges for Multivariate Spatial Analysis. Available (10/07/07) http://euclid.psych.yorku.ca/SCS/Gallery/guerry/

[10] Friendly, M. (in press). A. M. Guerry's Moral Statistics of France: Challenges for Multivariable Spatial Analysis. Statistical Science. Available (10/07/07) http://www.imstat.org/sts/future_papers.html

[11] Getis, A. and Ord, J.K., (1992) The analysis of spatial association by use of distance statistics. Geographical Analysis, 24:189–206.

[12] Harrower, M., and Brewer, C.A. (2003). ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. The Cartographic Journal 40, 27-37(11).

[13] Lindeberg, T., (1994). Scale-space theory: A basic tool for analyzing structures as different scales. Journal of Applied Statistics 21, 225-270.

[14] Martin, D. (1989) Mapping population data from zone centroid locations. Transactions of the Institute of British Geographers NS, 14, 90-97.

[15] Openshaw, S. and Charlton, M. and Wymer, C. and Craft, A.W. (1987). A mark I geographical analysis machine for the automated analysis of point data sets. International Journal of Geographical Information Systems 1, 335-358.

[16] Thomas, J.J., and Cook, K.A., eds. (2005). Illuminating the Path: The Research and Development Agenda for Visual Analytics. National Visualization and Analytics Center.

[17] Tobler, W., (1995) The Resel-Based GIS. International Journal of Geographical Information Systems 9, 95-100.

[18] Unwin, A., and Unwin, D. (1998). Exploratory Spatial Data Analysis with Local Statistics. The Statistician 47, 415-421.

[19] Whitt, H.P., and Reinking, V.W. (2002). *English Translation of* Guerry, A.-M. (1833) Essai Sur La Statistique Morale de la France, Paris: Crochard. Edwin Mellen Press, Lewiston, N.Y.

[20] Wood, J. (1996). Scale-based characterisation of digital elevation models. In *Innovations in GIS 3* (Parker, D., ed, pp. 163–75. Taylor & Francis, London.

[21] Wood, J. (2002). Visualizing the structure and scale dependency of landscapes. In Virtual reality in geography (Fisher, P.F., and Unwin, D.J., eds.), pp. 163–74. Taylor & Francis, London.