

Research Article

Interactive maps for visual data exploration*

GENNADY L. ANDRIENKO and NATALIA V. ANDRIENKO

GMD—German National Research Centre for Information Technology, Schloss
Birlinghoven, Sankt-Augustin, D-53754 Germany

E-mail: gennady.andrienko@gmd.de

Abstract. Descartes (formerly called IRIS) is a software system designed to support visual exploration of *spatially referenced data*, e.g. demographic, economical, or cultural information about geographical objects or locations such as countries, districts, or cities. Descartes offers two integrated services: **automated presentation** of data on maps, and facilities to **interactively manipulate** these maps. Automated mapping is enabled by incorporating generic knowledge on map design into the system. Descartes selects suitable presentation methods according to characteristics of the variables to be analysed and relationships among those variables—if more than one were selected simultaneously. The cartographic knowledge of Descartes allows non-cartographers to receive proper presentations of their data, and the automation of map construction helps the users to save valuable time that can better be used for data analysis and problem-solving.

Exploratory data analysis requires highly interactive, dynamic data displays. We strive to develop various interactive techniques for map manipulation that could enhance the expressiveness of maps and thus promote data exploration. We are convinced that a technique can be made especially productive if it is directed towards a particular presentation method: it can utilise peculiarities of this method and support those analytical operations that best fit to the method.

1. Introduction

The notion of exploratory data analysis developed in statistics has over the past decade spread to cartography, see DiBiase (1990) for one of the first discussions of the potential of EDA for cartography. Cartographers have recognised the demand in a new software allowing specialists in various disciplines (i.e. not professional map designers) to generate thematic maps and use them as tools facilitating ‘visual thinking’ about spatially referenced data (Kraak 1998).

Our system Descartes has been designed as such an environment for exploratory analysis of spatially referenced data. In this role it offers the user two complementary instruments: *automated, knowledge-based cartographic visualisation* and *interactive manipulation* of the generated maps and supplementary data displays. Interested readers are invited to run Descartes in the WWW at the URL <http://allanon.gmd.de/and/java/iris/>. The user interface of the system is implemented in Java 1.0.2, and therefore expected to run within any Java-enabling WWW browser.

*Web supplement located at: <http://allanon.gmd.de/and/IcaVisApplet>

The interface part communicates with the server, which performs data management and visualisation design. The server runs under UNIX. Descartes will not work on a computer connected to the Internet through a firewall that disables communication between the client and the server.

The present paper is mostly devoted to the interactive techniques enabling the use of maps as aids in 'visual thinking'. The next section contains a survey of the literature devoted to interactive, dynamically changing data displays. The principles of map design in Descartes are briefly described in §3. A more detailed consideration can be found in (Andrienko and Andrienko 1997, 1998a).

Our research is map-centred: we strive to advance traditional data presentation principles from thematic cartography by adding interactivity and dynamics. From here on, we use the words 'dynamics' and 'dynamic' in the sense that a display changes in real time in response to user's actions rather than to denote animated presentations of time-series data. The primary goal of this paper is to introduce the variety of techniques for map manipulation that we have developed thus far. As there is no solid theory yet about how to create and validate such techniques, in the course of our work we roughly adhered to the following strategy:

1. Address each cartographic presentation method individually.
2. Try to compensate for potential weaknesses of the target presentation method.
3. Try to enhance the advantages of the target presentation method with regard to the analysis task it is best suited for.
4. Exploit and affect specific visual properties of the target method.

Following the brief presentation of the interface for map viewing and manipulation in §4, the interactive manipulation techniques of Descartes are described in §5. Section 6 contains some concluding remarks concerning the introduced techniques.

2. Interactive manipulation of graphic displays in exploratory data analysis (literature survey)

A graphical presentation intended for use in exploratory data analysis should not be static. Recognising this, Bertin (1967) even proposes some methods of building transformable 'paper' presentations. It is therefore quite natural that researchers in statistical graphics, where the concept of the exploratory data analysis emerged (Tukey 1977), rather early started to exploit the potential of computers for creating interactive, transformable, and later dynamic (i.e. with real time motion) data displays. An overview of interactive techniques proposed in statistics can be seen in Cleveland and McGill (1988).

Buja *et al.* (1996) suggest a taxonomy of tools for manipulating graphical displays of data. They classify the tools into three fundamental categories: *focussing* individual views, *linking* multiple views, and *arranging* many views. These categories approximately correspond to three principal tasks the authors distinguish in data exploration: finding Gestalt, posing queries, and making comparisons.

Focussing techniques include selecting subsets and variables (projections) for viewing and various manipulations of the layout of information on the screen: choosing an aspect ratio, zooming and panning, 3-D rotations, etc. Tools of this kind are present in all contemporary systems for graphical data analysis.

As stated in Buja *et al.* (1991), 'a consequence of focussing is that each view will only convey partial information about the data'. This can be compensated for by displaying multiple views that 'need to be linked so that the information contained

in individual views can be integrated into a coherent image of the data as a whole'. The method of linking depends on whether the views are displayed in sequence over time or in parallel, simultaneously. In the first case, linking is provided by smooth animation. The most popular method for linking parallel views is identical marking of corresponding parts of multiple displays, e.g. with the same colour or some other form of highlighting. Usually highlighting is applied to objects interactively selected by the user in one of the displays. This method is a generalisation of the 'scatterplot brushing' technique first implemented by Newton (1978) and later elaborated in several directions: colour encoding of distance from the moveable focus point (McDonald 1982), simultaneous painting of several subsets using multiple colours and mixing of colours (Carr *et al.* 1986), various ways of selection and combination of selection operations (a synopsis is given in Wills 1996), and application to various types of data displays such as time-series plots, tables, raster images (Buja *et al.* 1991), bar charts (Hurley 1993, Wills 1995, Dawkes *et al.* 1996, Hummel 1996, Unwin *et al.* 1996), trees (Buja *et al.* 1996), mosaic plots (Unwin *et al.* 1996), etc. Our particular interest is the application of brushing to spatially referenced data. The relevant works will be considered separately.

The purpose of arrangement of multiple views is to facilitate comparisons. Usually a system for graphical analysis offers some built-in method(s) of arrangement such as the scatterplot matrix rather than any interactive tools for re-arrangement besides ordinary window operations.

Developers of interactive manipulation techniques in statistical graphics have paid significant attention to spatially referenced data. Already in 1982 McDonald had the idea of connecting a map to a scatterplot by painting regions in the map in the same colours as the corresponding points in the scatterplot. Buja *et al.* (1991) consider data referring to geographical locations. They propose to represent locations by a scatterplot of the latitudes and longitudes and link it with other graphics representing thematic information. The same approach is described in MacDougall (1992) for point- and raster-related data. The author also proposes a prototype interface with a map display of geographical objects specified as polygons. Instead of incorporating map displays into a system for interactive graphical analysis, Symanzik *et al.* (1996) achieve linking between maps and statistical graphics by coupling two independent programs: a GIS (ArcViewTM) displaying maps and a dynamic graphics system XGobi with a variety of statistical graphics (see also Cook *et al.* 1997). An analyst can brush points in either ArcView or XGobi, and the corresponding points are highlighted using the same colour or glyph in another application.

In all the brushing-based approaches, the role of the map is merely to carry the information about spatial location. All thematic information is usually represented and analysed by means of non-cartographic displays. It is reasonable to expect that involvement of the potential of thematic cartography would enhance the analytical capabilities of the tools proposed for exploratory analysis of spatially referenced data.

Cartographers have long adhered to the view of the map as primarily a medium of communication (MacEachren 1995). Recently, influenced by ideas of scientific visualisation and exploratory data analysis, they started to pay more attention to the role of the map as a tool to support visual thinking and decision making (MacEachren 1994b, MacEachren and Kraak 1997, Kraak 1998). To play this role effectively, the map needs two principal additions: interaction and dynamics. Current efforts of cartographers on developing map interactivity and dynamics are reflected

in a special issue of the journal *Computers and Geosciences* entitled 'Exploratory Cartographic Visualisation' (1997, 23).

It should be noted that the interactive manipulation techniques considered by cartographers are mostly based on the linking and brushing principles borrowed from the statistical graphics (Monmonier 1989, Dykes 1997). Dykes also implements some additional techniques such as 'dynamic re-expression', that is, replacement of the presentation method used in the map with another method. For example, a choropleth map may be substituted for a map with graduated symbols that presents the same data. Still, very few steps have been made up to now in another direction: advancement of presentation methods developed in thematic cartography, i.e. *making traditional methods interactive and dynamic*. We could find only two relevant works. Egbert and Slocum (1992) on the basis of an earlier work of Yamahira *et al.* (1985) propose an exploratory tool specifically designed for classed choropleth maps. This tool allows the user to change the classification interactively, and the map is repainted to reflect the changes. So, the tool affects the parameters of the presentation method. It is interesting that another manipulation technique for choropleth maps is proposed by researchers in statistics (Unwin and Hofmann 1998). The technique allows transformations of the function mapping from numeric values to shades. Initially the function is linear. Making it non-linear and varying parameters, the user can visually emphasise differences within selected subintervals of the value range.

Development of traditional cartographic presentation methods towards interactivity and dynamics is of primary concern in our system Descartes (though we do not deny the usefulness of linking maps to other types of displays). In general, Descartes is designed as an environment for exploratory cartographic analysis. This implies more than just interactivity of maps and multiple dynamically linked views on data. Kraak (1998) notes that, unlike the traditional cartography where a professional map designer creates a map to communicate spatial information to an audience, exploration usually involves non-cartographers creating maps for their private use. In doing this, the analysts should be able to rely on cartographic expertise provided by the software or some other means. Descartes answers this requirement: it is a knowledge-based system that incorporates cartographic expertise and on this basis performs automated map design.

3. Visualisation design in Descartes

Descartes greatly facilitates the process of map generation: a user only selects one or more data variables (table columns) to visualise, and the system immediately responds with a list of map presentations that properly correspond to characteristics of the variable(s) and, in the case of several variables, to relationships among them. After selecting a map in this list, a window is created displaying it. The user can now view and study the map and compare it with other ones.

To depict data on maps, Descartes applies presentation methods as developed in cartography. Each presentation method is based on the use of one or more of the *visual variables* introduced by Bertin (1967). These variables, according to their perceptual properties, have different *levels of organisation*: associative, selective, ordered, and quantitative. Presentation techniques for given data should be selected so that each data variable is represented by a visual variable with the corresponding level of organisation.

The cartographic presentation methods used in Descartes can be divided into those based on painting contours (applied only for area objects) and those based on

the use of signs. Signs may be either simple or structured. Painting methods exploit the visual variables *colour* and *value* (degree of darkness), the former in cases of nominal data, and the latter for ordered and numeric variables. Some clarification is necessary concerning the use of the visual variable *value* for numeric variables.

According to Bertin, *value* is unsuitable for representing quantitative data because it enables only ordered but not quantitative perception. Nevertheless, encoding numeric data by *value* (e.g. in choropleth maps) has become traditional in cartography. Such usage of this visual variable is productive for analysing spatial distributions and revealing spatial trends and patterns. Therefore Descartes adheres to this cartographic tradition but always complements a choropleth map with another presentation of the same data enabling quantitative perception: it is the general strategy of the system to supply the user with multiple presentations of the same data.

The rules for selecting presentation methods depending on data characteristics and relationships are summarised in table 1.

When several data variables are selected for visualisation, and relationships between them are heterogeneous, or some of them are incomparable, the system divides the variables into groups with homogeneous relationships between their elements; the variables belonging to the same group should be comparable. In particular, a group may contain only one element. Descartes will select presentation methods or visual variables for the groups and then try to combine them. Some possible combinations are listed in table 2.

According to Bertin (1967) and Tufte (1983), relationships between two or more data variables can be efficiently examined by comparing several maps, each presenting spatial distribution of one of the variables. For this reason we have added the presentation by multiple maps to the arsenal of combination techniques employed in Descartes. In this presentation each variable is represented by the area painting method (choropleth map), which favours perception of the distribution as a single image. Multiple maps are united in one panel. The system supports their simultaneous resizing and zooming so that they always have the same size and show the same territory. This simplifies their joint analysis.

The strategy of Descartes is to apply all presentation methods that are suitable for user-selected data (note that most of the rows in table 1 contain several variants of presentation). All alternative maps are presented to the user. The objective is not only to give the user the freedom to choose the most convenient map. As a rule, such multiple presentations of the same data are *complementary*. Each presentation method suits certain types of analysis tasks better than others. For example, pie charts are good for estimating proportions but worse for assessing differences between values. In the course of data exploration analysts perform different kinds of tasks, and there hardly exists one 'optimal' presentation method (MacEachren 1994a). To the contrary, the availability of several complementary presentations of the same data can greatly support the variety of activities comprising data analysis, especially exploratory analysis.

4. The interface to view and manipulate maps

Each map built by Descartes appears in a separate window containing the map itself, the legend, tools for zooming and handling geographical layer, and, what is most important, controls affecting visual properties of the data presentation (a map in Descartes may not only contain presentation of attribute data and spatial objects they refer to, but also other geographical features, e.g. rivers or forest areas, form-

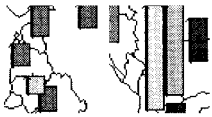




Table 1. Selection of visualisation methods for individual data variables and groups of related variables (*N* stands for number of variables).

| <i>N</i> | Type(s) of variable(s) | Relationship (where appropriate) | Visual variables | Presentation methods | Examples |
|----------|-----------------------------|----------------------------------|------------------|---|----------|
| =1 | nominal (including logical) | | colour or shape | area colouring, coloured signs, shape signs | |
| =1 | ordinal | | value | area shading, shaded signs | |
| =1 | numeric | | value or size | area shading, standalone bars | |
| >1 | logical | | colour | structured signs varying by presence or absence of coloured sectors | |
| >1 | numeric | comparable | size | parallel bars | |
| >1 | numeric | included in a common total | size | pies, standalone segmented bars, parallel bars | |
| >1 | numeric | ordered inclusion (a>b>...) | size | nested squares, parallel bars | |

ing map layers, or themes, in GIS terminology). In many cases, additional non-geographical presentations of the same data set are provided. Some of them are parts of manipulation devices, like the dot plots in the upper left section of the window in figure 1. For all screenshots given in the paper there are Java applets that show corresponding maps and enable all interactive operations. The applets are available at the URLs <http://allanon.gmd.de/and/IcaVisApplet/> and <http://ais.gmd.de/descartes/IcaVisApplet/>. The maps have been produced and stored by Descartes. The code supporting map viewing and manipulation is an extract from Descartes. This code does not communicate with the Descartes server and therefore is not disabled by firewalls.

The map is sensitive to mouse movement. When the cursor is positioned on some geographical object for which associated data exist, this object is highlighted in the map, and its name and the corresponding values are shown in a special section of

Table 2. Selection of combination techniques according to contents of variable groups and relationships among them.

| Group descriptions | Combination techniques | Examples |
|--|---|---|
| One nominal or ordinal variable + one numeric variable | Coloured or shaded standalone bars |  |
| One nominal variable + a group of numeric variables | Area colouring with superimposed diagrams |  |
| Two groups of comparable numeric variables | Two-sided bar diagrams (2 diverging groups of parallel bars) |  |
| Several groups of numeric variables with inclusion in a common total in each group | Parallel segmented bars if the groups are comparable, radial segmented bars otherwise |  |
| Several incomparable numeric variables | Radial bars |  |

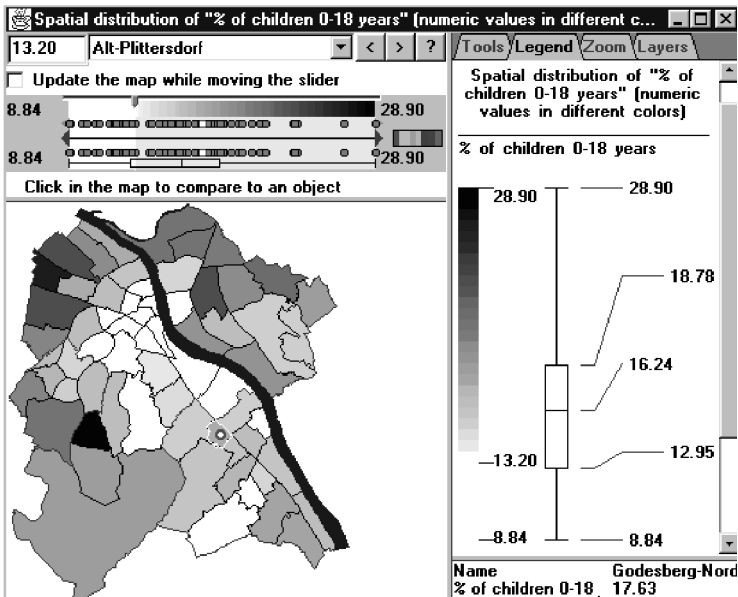


Figure 1. An example map window with dynamic manipulation tools.

the map window. If the map window contains a dot plot or a scatterplot as an additional non-geographical presentation of the data, the dot representing this object is also highlighted in the graphic. Linking between the map and the additional graphic works also in the opposite direction: when the cursor points at some dot in the graphic, both this dot and the corresponding object in the map are highlighted.

Furthermore, the map may be linked with other currently visible maps. This means that the highlighted object is also highlighted in the other maps, irrespective of their current scales, presented variables, and visualisation methods applied. So, one can observe that we propose our specialised, map-centred interactive manipulation techniques as a complement rather than a substitute to the widely recognised focussing and linking principles.

Manipulation techniques designed for different methods of presenting numeric information exploit similar interactive widgets that are organised in a construct further referred to as **slider unit**, see figure 2.

A slider unit represents the value range of a numeric variable presented in a map. It contains one or more sliders that can be moved using the mouse. Position(s) of the slider(s) affect visual properties of the map in a way that depends on the visualisation method applied. The dot plots show the distribution of values of the variable relative to its minimum and the maximum values. Each dot corresponds to one geographical object. The box-and-whiskers plot (Tukey 1977) offers a generalised presentation of the value distribution.

The small triangles at both ends of the slider can be used to limit the value interval to be visually examined. Like sliders, these triangles (further called **delimiters**) can be moved using the mouse. The use of delimiters is described in the next section together with the other techniques dynamically affecting the appearance of maps.

5. How to make use of the maps: exploration by map manipulation

Our approach is to design **specialised** manipulation tools addressing particular presentation methods. In so doing, we take into account (1) what data features and relationships the target method potentially helps to uncover; (2) what graphical expressive means the method employs. The technique being designed is intended to reinforce visibility of these features and to facilitate their disclosure. It operates with the same graphic means and, possibly, adds redundant visual variables to increase the legibility and expressiveness of the display.

5.1. *Dynamic choropleth maps*

In this subsection we describe a technique designed for *unclassed choropleth maps* representing numeric data. The ‘choropleth map’ visualisation method consists in encoding values of some attribute referring to units of territory division by colours or shades. ‘Unclassed’ means that values of a numeric attribute are converted into degrees of darkness directly, without previous classification: the degrees of darkness are (roughly) proportional to the numbers they represent.

A merit of a choropleth map is that it is capable of producing an integral image



Figure 2. An instance of slider unit.

of the spatial distribution of data, and thus enables the highest, overall level of map reading (Bertin 1967). Since the whole distribution can be grasped as one image, it is possible to compare two or more distributions presented by choropleth maps. By such comparison one can reveal relationships between several variables: relatedness will manifest itself in similar spatial patterns.

‘Dynamic visual comparison’ is an interactive tool intended to enhance inherent capabilities of the given presentation method. With this tool some number N within the value range of the shown variable is interactively selected, and the map is immediately redrawn using a *diverging*, or *double-ended*, colour scheme (Brewer 1994): values higher than N are encoded by shades of one colour (hue), and those lower than N are shown by shades of another colour. The greater the difference is between some value and N , the darker is the shade used to represent it. The values exactly equal to N are encoded by white. When the user changes N , the map is dynamically repainted.

Descartes offers a number of ways to control the reference value N (see the upper-left section of the window in figure 1):

- (a) moving the slider;
- (b) clicking on an object in the map;
- (c) selecting an object by name in the choice widget. In both cases (b and c), the value associated with the selected object becomes the reference value for the comparison;
- (d) automatically locating the object with the previous or with the next value by pressing the buttons “<” and “>”;
- (e) entering an exact number in the edit field.

The sequence of screenshots shown in figure 3 demonstrates how a map dynamically changes as the slider moves from left to right (the reference value N grows). Shades of brown are used for values higher than N , and shades of cyan for those below N .

The map of this example shows the variation of mean household sizes through the districts of Bonn. (The source data were kindly provided by the Statistical department of the Bonn city administration). In the first map (A), the reader can observe a certain spatial trend: the values on the outskirts are higher than in the centre. Applying the visual comparison tool, one can investigate the spatial variation more comprehensively.

In the following maps (B)–(D), the cyan areas representing low values are very well distinguished from the brown areas encoding higher values. Map (B) shows that the lowest values are located in the historical centre and in some districts on the

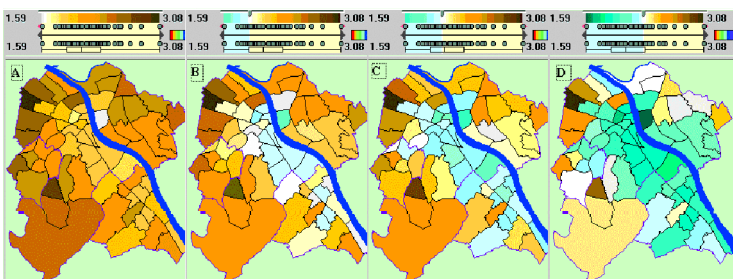


Figure 3. Several states of a map in the process of ‘dynamic visual comparison’.

south-east. Moving the slider farther to the right, one can observe how the cyan area spreads to the south-east and north-west. At a certain moment the extension in this direction stops, and the cyan area starts spreading to the west and the east.

So, in the process of visual comparison we have found several interesting spatial patterns: centre-periphery, north-west–south-east stretching, and an ‘arc’ formed by the north and west outskirts. It would be much more difficult to reveal this information only from one static picture like (A). This phenomenon can be explained as follows.

In an unclassed choropleth map different numbers are encoded by (ideally) different shades. This, to some extent, inhibits visual grouping and division into regions since, according to Bertin (1967), *value* is a selective visual variable but not associative. Dynamic comparison adds the variable *hue* to the expressive means used in the map. This variable is not only selective but also associative, and therefore encourages visual grouping of objects: neighbouring objects painted in the same colour tone tend, despite differences in shades, to be associated into a single figure. In this way, dynamic comparison favours revealing spatial patterns. It obviously helps in the task that is considered to be the most important application of the choropleth map presentation method, that is, looking for spatial patterns in value distributions.

A disadvantage of the unclassed choropleth map method is the difficulties humans meet in distinguishing close shades (Dobson 1973). It is especially hard to figure out the right order if compared objects are spatially disjoint and have surroundings with different degrees of darkness. With the dynamic visual comparison technique these difficulties are easy to overcome: to compare values for two objects, it is sufficient to click on one of them and observe in what hue the other is repainted.

Diverging colour schemes are used in cartography for emphasising progressions outward from a critical midpoint of data (Brewer 1994). In exploratory data analysis, however, such critical points are often previously unknown. The ‘dynamic comparison’ technique can help to empirically reveal them, i.e. to find such reference values that lead to meaningful spatial patterns.

The technique described can be applied as well in a joint study of several choropleth maps presenting different variables. Descartes provides this opportunity whenever a user selects two or more numeric variables to be visualised. One of the presentations the user receives in such cases is a multi-map presentation: several choropleth maps are shown in one window that contains controls for manipulating all the maps. The user can handle each map independently of the others or all maps in parallel.

5.2. *Dynamic bars*

Another method to depict values of a numeric variable is to encode them by proportional sizes of graphical symbols, e.g. by heights of bars: the greater the value, the higher is the bar. Unlike the choropleth map method, this method does not favour forming integral images. Instead, it is more beneficial for estimating absolute values and differences in values. According to our strategy, the interactive manipulation tool designed for this method should also work in this direction.

It is obvious that the values of a variable for two geographical objects can be more easily compared if the user does not have to mentally subtract the height of one of the bars from the height of the other bar. The system can automatically do this for the user: after s/he clicks on an object in a bar map, the system redraws the

map so that corresponding bar heights for all objects become proportional to the differences between the associated values and the value for the selected object. The orientations of bars (upward or downward) show the signs of the differences. For better perception, orientation is reinforced by colour, i.e. two different hues are used to represent positive and negative differences.

To control the reference value for visual comparison on a bar map, the same facilities are available as for dynamic choropleth maps (cf. the map windows in figures 1 and 4). However, the graphical elements being manipulated are different: colours and shades of area painting in one case and bar heights in the other.

5.3. Dynamic focussing on a value subrange of a numeric variable

To match numeric values with bar heights, as in bar maps, or degrees of darkness, as in choropleth maps, Descartes builds a *linear* encoding function: $f(x) = (v_{max} - v_{min}) \times (x - x_{min}) / (x_{max} - x_{min})$, where x is a value to be encoded, x_{max} and x_{min} are the highest and lowest values of the represented data variable in the given sample, and v_{max} and v_{min} are the adopted highest and lowest values of the visual variable (*size* or *value*). If the sample contains both positive and negative values, $x_{min} = 0$, and $x_{max} = \max(-s_{min}, s_{max})$, where s_{min} is the minimum (negative) value for the sample, s_{max} is the maximum value.

Sometimes the mapped value set contains, relative to the bulk of it, a few

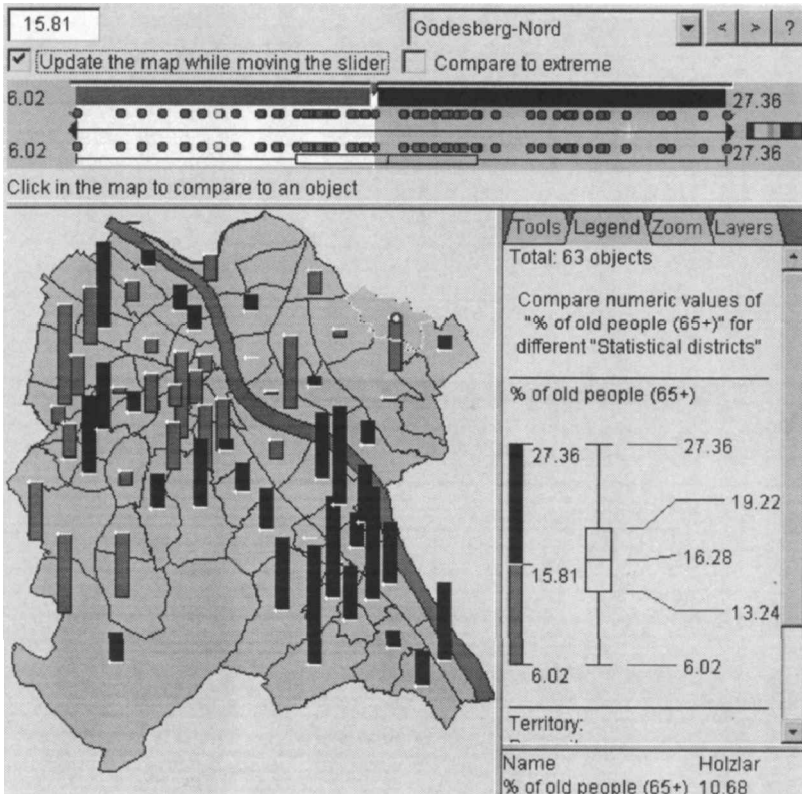


Figure 4. Visual comparison of percentages of old people in districts of Bonn with that for a selected district (Godesberg-Nord) on a bar map.

extremely high or extremely low values ('outliers'). In this case, the generated map will not be distinctive or expressive enough with regard to the mainstream values. Look, for example, at the left map in figure 5 (the source data have been taken from *The World Fact Book '95* published by CIA). Albania has a very high birth rate when compared to all other European countries. The dot plots in the slider unit clearly illustrate the distance from Albania's maximum value of 21.70 to all other values, which are all within the interval from 10.56 to 15.93. As the maximum degree of darkness is assigned to 21.5, the maximum of the remaining values, which is 15.93, is encoded by the degree approximately equal to 48 % of the total maximum. As an effect, countries with rather different values (e.g. Switzerland 12.04, France 13.00, Ireland 14.04) are painted in very similar shades. The same problem arises with a bar map: all the bars except that for Albania appear to be of almost the same height.

To handle such data, Descartes allows the user to cut the represented interval, i.e. remove outliers, and subsequently redefine the encoding function. Referring to our example, if Albania is excluded from the presentation, the full scale of darkness will be matched with a significantly shorter interval. The value 15.93 will be encoded by 100 % darkness, and differences among the mainstream values will become apparent in the map, as is clearly seen in figure 5 (right).

Outlier removal (and, in general, focussing on any subinterval of the whole value range for studying more in detail) is done interactively by using *delimiter widgets* as introduced in §4. In a slider unit (see upper sections of the map windows in figure 5), the line between two dot plots represents the whole value range of the visualised variable. The vertical edges of the delimiters (triangles) mark the boundaries of the currently depicted subinterval. Initially, the delimiters are located at the ends of the line. Moving the delimiters along the line, the user may select a subinterval for further study. The upper dot plot in the slider unit also contains a pair of delimiters (smaller and lighter triangles). They can be used for more exact adjustment of the

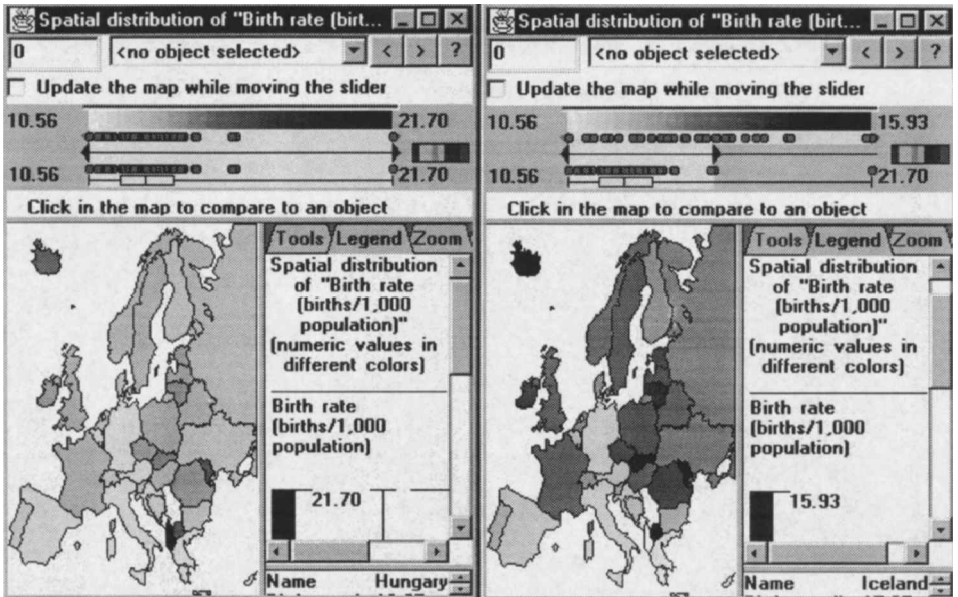


Figure 5. The effect of dynamic focussing.

boundaries of the selected interval. After any movement of a delimiter the encoding function is redefined, and the map redrawn. Geographical objects with values lying outside the focussed interval are marked in the map without painting or drawing signs.

The appearance of the slider unit also changes. The lower dot plot and the box-and-whiskers plot represent, as before, the whole data sample. Light background painting indicates which part of the range is currently presented on the map. The upper dot plot only shows the focussed subinterval stretched to the whole width of the plot area. This zooming makes it easier to distinguish dots representing close values. The dynamic linking between the dot plots and the map is preserved.

The technique of interactive removal of outliers (in application to scatterplots) was proposed by Cleveland (1993). Removing objects standing far apart of the rest allows the scale of the scatterplot to be enlarged and makes the mainstream objects better distinguished.

The outlier problem can also be solved by applying non-linear encoding functions, e.g. logarithmic ones. This expedient is frequently used in statistical graphics (see, for example, Carr *et al.* 1986). Unwin and Hofmann (1998) propose a tool for interactive non-linear transformations of mapping from numbers to shades in a choropleth map. In our opinion, however, such transformations are undesirable for maps, at least those intended for use in exploratory analysis. Unlike scatterplots and other statistical graphics containing axes and labelled ticks, a map itself does not provide any information on how numbers are encoded by shades or sizes. This information can only be retrieved from the legend. At the same time exploratory analysis greatly depends on the immediate impression (MacEachren 1995, pp. 70–71) that comes before the map reader starts learning the legend or applying any knowledge schemata. Besides, in a scatterplot one usually looks for functional dependencies between two variables. Transformation of axes does not impede noticing of such dependencies. In a map one looks for spatial patterns made by neighbouring signs (symbols or painted areas) with similar visual appearance. With a non-linear encoding function, similar appearance may not necessarily reflect closeness of values, and vice versa, close values may be encoded rather differently. So, visual grouping of geographical objects (that usually happens on the pre-attentive stage of image processing by a human) may be inconsistent with actual relationships. Therefore, maps intended for visual exploration should allow intuitive interpretation (see also Tufte 1983, p. 153) that is hardly supported by non-linear encoding functions.

One can observe some similarity between the dynamic focussing technique and the well-known 'Dynamic query' technique (Ahlberg *et al.* 1992). The real similarity, however, mainly refers to the interface rather than function. In 'Dynamic query', the goal is to allow the user to conveniently specify constraints and to see which objects satisfy them. Accordingly, the display is changed by removing the objects not satisfying the query. In our case, the goal is to present values lying within some user-selected interval with maximum possible expressiveness. Therefore, the main effect consists in a transformation of the encoding function. Objects with associated values lying outside the focus interval are actually not removed from the view, rather their values are not visually expressed.

5.4. Choropleth maps with dynamic classification

The dynamic classification technique is designed for classed choropleth maps. Unlike the unclassified choropleth maps considered earlier, values of a numeric variable

are not directly encoded by proportional degrees of darkness. Instead, the value range of the variable is first divided into two or more intervals, and a specific colour is assigned to each interval. Objects with values fitting in the same interval will thus have identical colour in the map. Due to the classification, which is a kind of generalisation, a part of the initial information concerning differences in values is lost. However, this method is very suitable for grouping and observing spatial distributions of groups of objects.

A choropleth map with a fixed classification, which is typical for ‘paper’ cartography, inevitably imposes a subjective view on the data. In Descartes, the classification is mobile since the user can interactively move the boundaries of the intervals, easily change the number of intervals, and select various kinds of automatic division such as equal frequency or equal length intervals. All these operations result in immediate repainting of the map according to the new classification. The ease of applying these operations and the immediate feedback greatly encourage users to search and find interesting, geographically meaningful groupings of objects.

Figure 6 shows the interface of the dynamic classification technique. The main manipulation widget is the slider unit. Unlike in visual comparison, it may contain several sliders. Positions of the sliders reflect the values of class boundaries. When the user moves any of the sliders, the corresponding boundary changes. A mouse click between two sliders splits the corresponding interval into two parts, and a new class appears. When a slider closely approaches one of its neighbouring sliders, it is merged with the neighbour, and the number of classes decreases by one.

Besides the slider unit, a number of other widgets are available for interactive classification. They allow the user to manually enter the values of class boundaries, to see results of automatically dividing the interval into 2 to 9 equal length or equal frequency intervals, or to select different variants of colouring.

The dynamic classification tool also allows users to visually isolate a particular class in order to observe its spatial distribution. For this purpose, the user simply clicks on the corresponding colour on the colour band in the slider unit. In response,

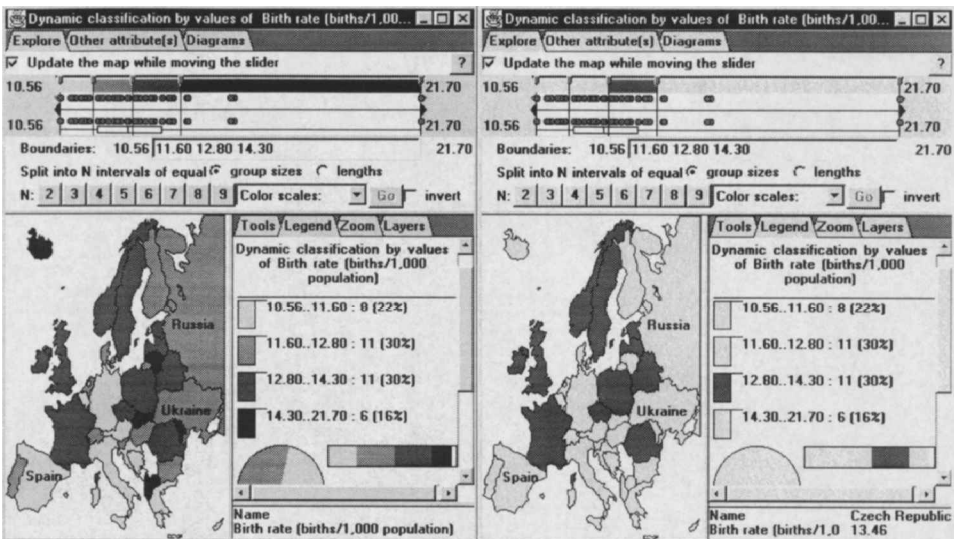


Figure 6. A classed choropleth map and the tool to manipulate classes.

only the objects belonging to this class will be coloured, whereas the other objects will be shown in neutral grey (see the right map window in figure 6). Another click on the same segment will result in colouring objects from the original class and its two adjacent classes. One more click makes the whole map coloured again. This type of visual isolation also occurs during slider movement: only the two classes affected by the corresponding boundary change are visualised.

In what we have described so far, our dynamic classification tool can be regarded as an upgrade of the work of Egbert and Slocum (1992) by utilising recent progress in graphical hardware, software, and user interface design. The essential novel feature is the opportunity to investigate relationships of the base variable (the one used for classing) with other variables. In Descartes, the user can select one or more variables to study, and the system will then calculate the statistics of value distribution of these variables for each class. Statistical results are presented to the user in the form of averaged 'portraits' of the classes and by 'box and whiskers' plots (see figure 7). All statistical results are recalculated when the classes change.

In the example of figure 7, some relationship between birth rates and death rates can be observed across the countries of Europe. The countries are classified according to birth rates. The supplementary 'class statistics' display (figure 7) represents the statistics of distribution of the death rates across the classes. It can be noticed that death rates are lower in the countries either with low or with high birth rates. The dynamic classification tool allows the user to investigate whether this trend preserves when the classification changes.

A more detailed consideration of the dynamic classification technique as well as an example of data exploration with its use can be found in Andrienko and Andrienko (1998b).

Similar to the dynamic classification that is based on a single numeric variable, we have developed a tool supporting cross-classification with two base variables. The map in figure 8 demonstrates this technique by an example: a classification of European countries on the basis of the variables 'National product per capita, \$', and 'Part (%) of old people (65 and more years) in total population'.

Analogous to the previous variant of classification, the value ranges of both variables are divided into intervals. This potentially defines $M \times N$ classes, where M and N are the numbers of intervals for the first and for the second variable, respectively. To assign colours to classes, first a colour scale is separately selected for each variable. In our example, a scale from yellow to green was selected for the

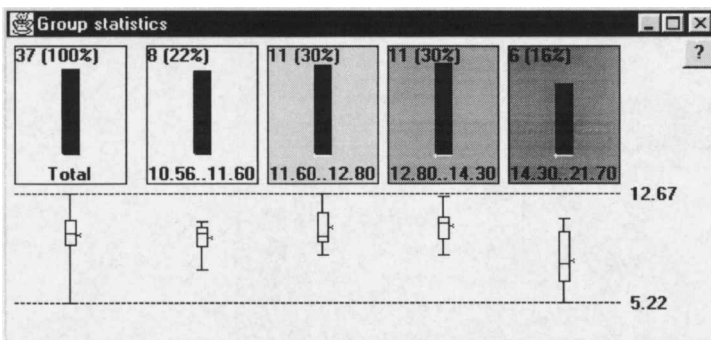


Figure 7. A display of class statistics.

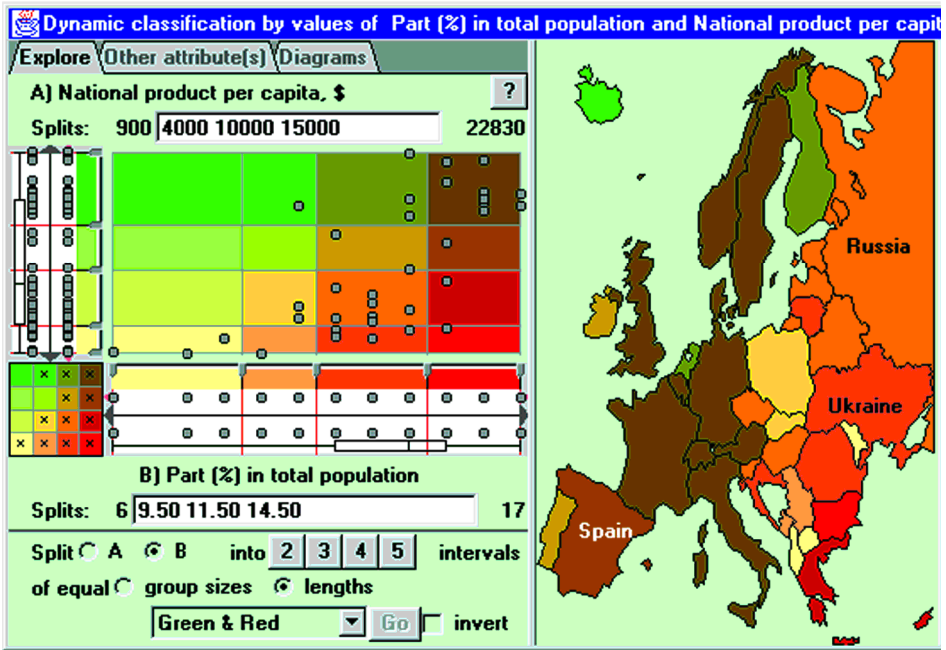


Figure 8. Cross-classification on the basis of two numeric variables.

first variable, and a scale from yellow to red for the second one. Now, the system generates a colour matrix, like the one seen in figure 8 (left), by mixing colours from the two scales. A geographical object in the map is painted according to the colour of the cell of its associated values. The distribution of colours across the map can manifest interesting relationships between variables. Presence of a relationship is signified by a clear division of the territory into regions formed by similarly painted neighbouring objects.

The investigation of relationships is also supported by a scatterplot **linked** with the map (see the left side of figure 8). This linking and the coloured background of the scatterplot allows the user to establish a correspondence between objects in the scatterplot and those in the map. One may argue that the scatterplot alone is sufficient to find relationships. However, relationships revealed using a map have a different quality than those revealed in scatterplots. In particular, a map manifests relationships inherently based on the spatial distribution, while a scatterplot can show a kind of functional dependency. In the extreme, the map can visually form cohesive regions while the corresponding scatterplot only shows a chaotic diffusion of dots. In any case, a scatterplot does not show any spatial information, e.g. whether objects with close values are also geographically close. For example, the map in figure 8 demonstrates an apparent spatial pattern with brown west and red east that cannot be derived from the scatterplot.

All operations provided for one-dimensional classification are also supported in the two-dimensional case. In addition, the scatterplot can be used as an interactive device to manipulate the map. Clicking in a coloured cell results in visual isolation of objects belonging to the corresponding class: only these are coloured in the map, the others are shown in neutral grey. It is also possible to visually isolate classes

belonging to one row or one column of the matrix. The dynamic focussing technique described in the previous section works also for classed choropleth maps. As the focussed interval of one of the base variables is narrowed, the scatterplot is zoomed.

In general, grouping is very important in exploratory data analysis. However, to be really useful, it should necessarily be dynamic. In particular, any *static* division of the value range of a numeric variable into steps is counter-productive for exploration: according to Bertin (1967), ‘determining the steps is, in fact, the goal of the graphic operation, not its means’.

5.5. Dynamic isolation of qualitative values

This tool is designed for maps that encode values of qualitative variables by colours or shapes. It allows the user to temporarily switch off the depiction of some values. Corresponding colours or shapes are removed from the map. This helps to concentrate the user’s attention on the remaining values and more easily observe their distribution. As the user changes the selection of values, those changes are instantaneously reflected in the display, thus helping the user to concentrate her/his attention on another group of values. The work of the dynamic isolation tool is illustrated by figure 9.

Dynamic isolation is designed for presentations based on the use of the associative visual variables *colour* and *shape*. It is especially useful for analysing presentations based on *shape*, which is, according to Bertin (1967), only associative but not selective. When all values are shown in a map, it is difficult to concentrate on some of them and disregard the others. Dynamic isolation effectively compensates for this lack of selectivity. After masking some of the values, the associative potential of shape or colour promotes seeing the geographical distribution of the remaining values as a single image.

The work of dynamic isolation in Descartes resembles that of ‘Dynamic query’

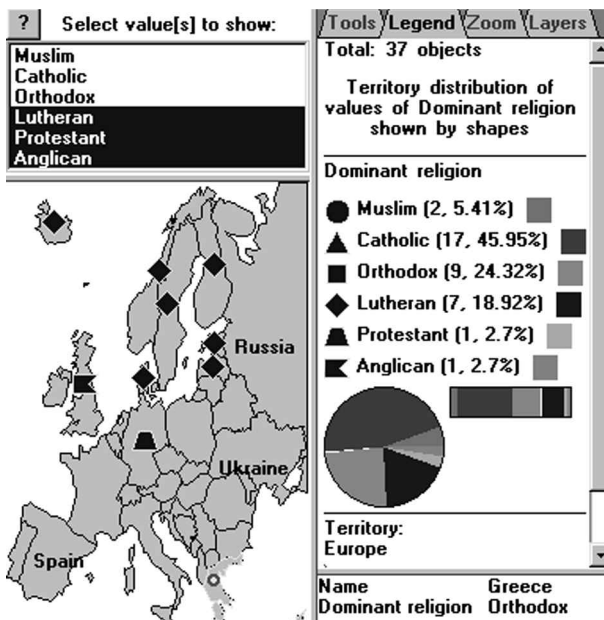


Figure 9. A map presenting a result of dynamic isolation.

(Ahlberg *et al.* 1992): removal of some symbols from the display in response to specifying constraints on values of a variable. However, like in the case of dynamic focussing on value subintervals of a numeric variable, the goal of our tool is different: it is intended primarily for studying spatial distributions rather than search for objects satisfying some constraints.

6. Conclusions

Descartes fits in the paradigm of cartographic visualisation defined by MacEachren (1994b) and further elaborated by MacEachren and Kraak (1997). This paradigm includes three key elements: private map use (an individual generates a map for his or her own needs), direction towards revealing unknowns rather than presenting known facts, and high human-map interaction. Accordingly, in developing Descartes we address two objectives:

- allow people in various disciplines (not professional cartographers) to easily present their data by maps (*private map use*).
- support *high interaction* of the analysts with the maps to facilitate *revealing unknowns*.

To achieve the first objective, Descartes incorporates cartographic expertise and on this basis automates map generation with the application of most widely known presentation methods from thematic cartography. The second goal is approached by advancing these presentation methods towards interactivity and dynamics. This approach is intended to complement the popular way of adding interactivity to a map by linking it with other types of data displays. Usually these non-cartographic displays are given the leading role in the analysis of attribute information while the map is used mainly to show the spatial location of objects. In our research we tried to bring the potential of traditional cartographic means of visualisation to exploratory data analysis.

The exploratory tools described here were developed according to our strategy of designing specialised techniques based on particular features of each presentation method, and directly affecting its inherent symbol system. With these tools we try to create more favourable conditions for data analysis by compensating for weaknesses and enhancing potential capabilities of the target presentation methods.

Map authors have always been struggling to increase map expressiveness. For this purpose, such expedients as various colour schemes, classification strategies, non-linear encoding functions, etc. were developed. However, in static presentations the use of these practical aids often results in partial loss of information, distorted perception, or/and imposing some pre-interpretation of source data. Such 'side effects' are counter-productive in exploratory data analysis. Dynamic, interactive displays allow users to achieve necessary expressiveness without significant sacrifices, and may neutralise undesirable consequences. For example, flexible midpoints or class boundaries allow users to reject any pre-set view on data, and to empirically find a view illustrating important, non-trivial properties of the spatial distribution.

Our work on enhancing map interactivity goes on. At this stage, only a part of the cartographic presentation methods implemented in Descartes have been made as flexible as we think is possible and desirable. We hope that the paper has demonstrated the usefulness of such a development.

References

- AHLBERG, C., WILLIAMSON, C., and SCHNEIDERMAN, B., 1992, Dynamic queries for information exploration: an implementation and evaluation. In *Proceedings ACM CHI '92* (ACM Press), pp. 619–626.
- ANDRIENKO, G. L., and ANDRIENKO, N. V., 1997, Intelligent cartographic visualisation for supporting data exploration in the IRIS system. *Programming and Computer Software*, **23**, 268–282.
- ANDRIENKO, G. L., and ANDRIENKO, N. V., 1998a, Intelligent visualisation and dynamic manipulation: Two complementary instruments to support data exploration with GIS. In *Proceedings Advanced Visual Interfaces '98* (New York: ACM Press), 66–75.
- ANDRIENKO, G. L., and ANDRIENKO, N. V., 1998b, Dynamic categorization for visual exploration of spatial information. *Programming and Computer Software*, **24**, 108–115.
- BERTIN, J., 1967–1983, *Semiology of Graphics. Diagrams, Networks, Maps* (Madison: The University of Wisconsin Press).
- BREWER, C. A., 1994, Colour use guidelines for mapping and visualisation. In *Visualisation in Modern Cartography* (New York: Elsevier Science Inc.), pp. 123–147.
- BUJA, A., McDONALD, J. A., MICHALAK, J., and STUETZLE, W., 1991, Interactive data visualisation using focusing and linking. In *Proceedings IEEE Visualization '91* (Washington: IEEE Computer Society Press), pp. 156–163.
- BUJA, A., COOK, D., and SWAYNE, D. F., 1996, Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics*, **5**, 78–99.
- CARR, D. B., NICHOLSON, W. L., LITTLEFIELD, R. J., and HALL, D. L., 1986, Interactive color display methods for multivariate data. In *Statistical Image Processing and Graphics* (New York: Marcel Dekker), pp. 215–250.
- CLEVELAND, W. S., and MCGILL, M. E., 1988, *Dynamic Graphics for Statistics* (Belmont: Wadsworth and Brooks).
- CLEVELAND, W. S., 1993, *The Elements of Graphing Data* (Murray Hill: AT&T Bell Laboratories).
- COOK, D., SYMANZIK, J., MAJURE, J. J., and CRESSIE, N., 1997, Dynamic graphics in a GIS: more examples using linked software. *Computers and Geosciences*, **23**, 371–385.
- DAWKES, H., TWEDDIE, L., and SPENCE, B., 1996, VICKI—The visualisation construction kit. In *Proceedings of Advanced Visual Interface '96* (New York: ACM Press).
- DI BIASE, D., 1990, Visualization in the earth sciences. *Earth and Mineral Sciences, Bulletin of the College of Earth and Mineral Sciences, Penn State University*, **59**, 13–18.
- DOBSON, M. W., 1973, Choropleth maps without class intervals: A comment. *Geographical Analysis*, **5**, 358–360.
- DYKES, J. A., 1997, Exploring spatial data representation with dynamic graphics. *Computers and Geosciences*, **23**, 345–370.
- EGBERT, S. L., and SLOCUM, T. A., 1992, EXPLOREMAP: an exploration system for choropleth maps. *Annals of the Association of American Geographers*, **82**, 275–288.
- HUMMEL, J., 1996, Linked bar charts: analysing categorical data graphically. *Computational Statistics*, **11**, 23–33.
- HURLEY, C., 1993, The plot-data interface in statistical graphics. *Journal of Computational and Graphical Statistics*, **2**, 365–379.
- KRAAK, M.-J., 1998, The cartographic visualization process: from presentation to exploration. *The Cartographic Journal*, **35**, 11–15.
- MAC EACHREN, A. M., 1994a, *Some Truth with Maps: A Primer on Symbolization and Design* (Washington: Association of American Geographers).
- MAC EACHREN, A. M., 1994b, Visualization in modern cartography: setting the agenda. In *Visualisation in Modern Cartography* (New York: Elsevier Science Inc.), pp. 1–12.
- MAC EACHREN, A. M., 1995, *How Maps Work: Representation, Visualization, and Design* (New York: The Guilford Press).
- MAC EACHREN, A. M., and KRAAK, M.-J., 1997, Exploratory cartographic visualization: advancing the agenda. *Computers and Geosciences*, **23**, 335–344.
- MCDONALD, J. A., 1982, *Interactive Graphics for Data Analysis*. Project Orion, **11** (Stanford: Stanford University).

- MACDOUGALL, E. B., 1992, Exploratory analysis, dynamic statistical visualization, and geographic information systems. *Cartography and Geographic Information Systems*, **19**, 237–246.
- MONMONIER, M., 1989, Geographic brushing: enhancing exploratory analysis of the scatterplot matrix. *Cartographycal Analysis*, **21**, 81–84.
- NEWTON, C. M., 1978, Graphics: from alpha to omega in data analysis. In *Graphical Representation of Multivariate Data* (New York: Academic Press), pp. 59–92.
- SYMANZIK, J., MAJURE, J., and COOK, D., 1996, Dynamic graphics in a GIS: a bidirectional link between ArcView 2.0 and XGobi. *Computing Science and Statistics*, **27**, 299–303.
- TUFTE, E. R., 1983, *The Visual Display of Quantitative Information* (Cheshire: Graphics Press).
- TUKEY, J. W., 1977, *Exploratory Data Analysis* (Reading: Addison-Wesley).
- UNWIN, A., HAWKINS, G., HOFMANN, H., and SIEGL, B., 1996, Interactive graphics for data sets with missing values—MANET. *Journal of Computational and Graphical Statistics*, **5**, 113–122.
- UNWIN, A., and HOFMANN, H., 1998, New interactive graphic tools for exploratory analysis of spatial data. In *Innovations in GIS*, **5** (London: Taylor and Francis), pp. 46–55.
- WILLS, G., 1995, Visual exploration of large structured data sets. In *New Techniques and Trends in Statistics* (Amsterdam: IOS Press), pp. 237–246.
- WILLS, G., 1996, Selection: 524,288 ways to say this is interesting. In *Proceedings IEEE Information Visualization '96* (Washington: IEEE Computer Society Press), 54–60.
- YAMAHIRA, T., KASAHARA, Y., and TSURUTANI, T., 1985, How map designers can represent their ideas in Thematic maps. *The Visual Computer*, **1**, 174–184.