information. If a user has set an upper limit on the price they are prepared to pay, they would almost certainly wish to know about any house which failed only that limit and by a very small amount. Movement of a limit to enclose a black house will cause it to turn green, confirming that it now satisfies all limits.

The black encoding is extremely valuable in those circumstances wherein there are no houses that satisfy the stated limits, a situation depicted in Figure 3.59. In this example we see that by spending a little more, one house becomes available, whereas by accepting a house with three bedrooms a number of houses become available. Astonishingly, it is rare to find such guidance in online house sales.

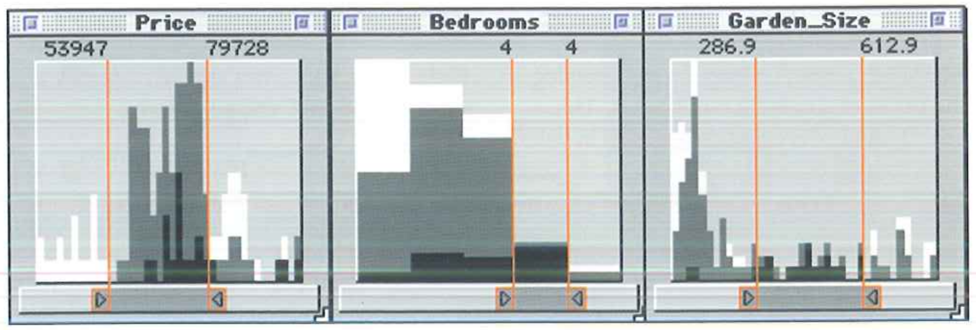

FIGURE 3.59 Even if no houses satisfy all attribute limits, black houses, which fail only one limit, provide guidance as to the effect of relaxing limits

Potential applications of the Attribute Explorer technique will no doubt occur to the reader, especially after experiencing, for example, the frustration of trying to arrange flights on specific dates at convenient times or trying to decide which mobile telephone or washing machine to buy. Combination with other techniques could prove useful, such as providing an Attribute Explorer histogram along each axis of a parallel coordinate plot to enable better insight into the distribution of attribute values and, by brushing, the correlation between them. The Explorer is, of course, not restricted to numerical data: it can handle a combination of numerical, categorical and ordinal data. Concurrently with the invention of the Attribute Explorer, Eick (1994) proposed the technique of Data Visualization Sliders. He drew attention to the fact that, as well as providing a mechanism for the interactive specification of limits to an attribute, 'the effectiveness of sliders may be increased by using the space inside . . . as an interactive colour scale, a bar plot for discrete data, [or] a density plot for continuous data'. The effectiveness of the Attribute Explorer has been studied (Li and North, 2003) by comparing it with a Dynamic queries interface (see Chapter 5, Figure 5.23) in the context of a number of tasks. It was observed that the 'brushing histograms', as the Explorer was called, tended to offer advantages with the more complex tasks. It was commented that brushing histograms could be extended in various ways, for example to be zoomable and to possess granularity controls.

While originally invented to support the selection of one object from many on the basis of its attributes, the Attribute Explorer also offers considerable potential to support exploration with the goal of acquiring **insight** into data. An example is provided in its use as an investigative tool by the Swedish Defence Research Agency and illustrated briefly in Figure 3.60. (A more extensive

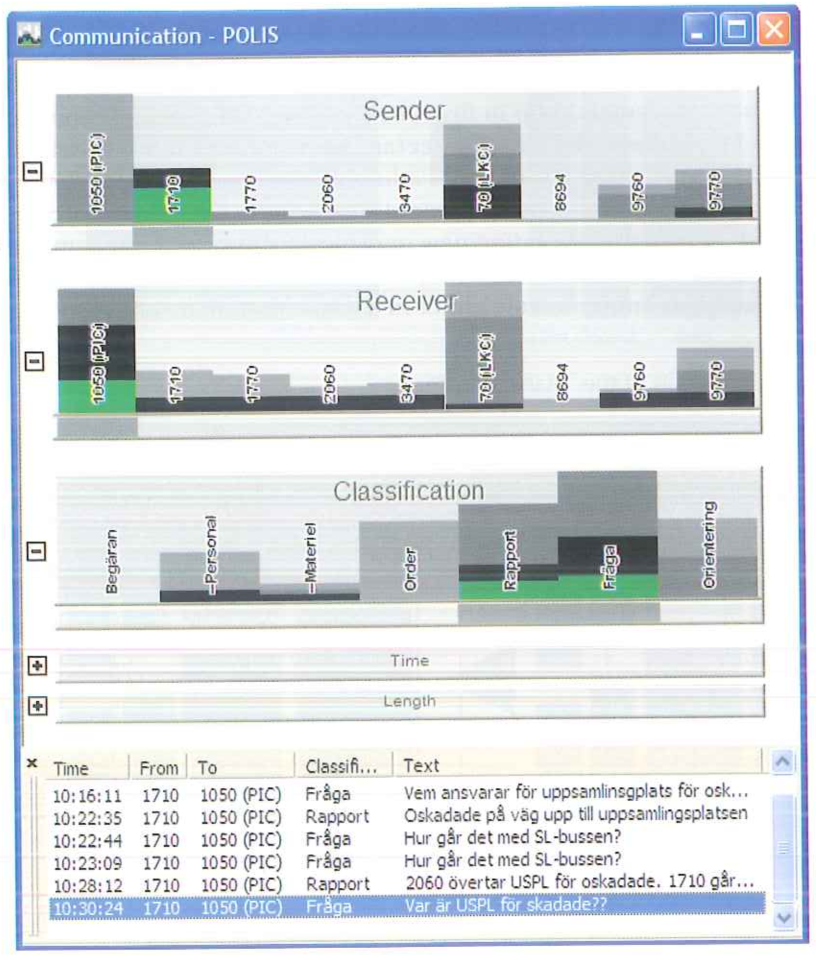

FIGURE 3.60 An Attribute Explorer representation of three dimensions of communication data captured during an emergency services exercise, supporting interactive exploration by an analyst

account is the subject of one of the case studies in Chapter 6.) The figure relates to communication data accumulated during a practice scenario of a train derailment on the Stockholm subway (Morin and Albinsson, 2005) involving three data dimensions. A single sender (1710) has been identified and a single receiver (1050), the on-site commander, as well as two categories of classification (reports and questions). The green 'hits' indicate that a large part of the total communication initiated by 1710 is indeed reports and questions directed to 1050. These 'hits' are also shown in the list below the Attribute Explorer.

Mosaic plots

The idea behind mosaic plots (Friendly, 1994) can be demonstrated by an example related to the Titanic disaster of April 1912, when 1,731 of the 2,201 passengers and crew were lost (Dawson, 1995). Table 3.1 shows the raw numerical data, involving four attributes: gender, survival, class and adult/child. Whereas it is difficult without careful examination to obtain much insight from the table, a so-called mosaic plot (Friendly, 1992, 1994, 2000) can profitably be generated as follows. We start with a rectangle whose area is proportional to the number of passengers and crew (Figure 3.61(a)). We then break down that area according to class of travel (Figure 3.61(b)), again representing number by area. Next we break down the resulting rectangles according to gender (Figure 3.61(c)), from which we can already begin to gain immediate insight into, for example, the male/female ratio in first-, second- and third-class accommodation. Finally, we break down the existing rectangles according to two attributes: survival (green) or otherwise (black), and adult/child (Figure 3.61(d)). The resulting mosaic plot is capable of providing rapid insight into the nature of the Titanic disaster and, of course, generating new questions – Who were the females travelling third class? Why did the vast majority of first-class females survive? Why, proportionately, did more female children survive than male children?

TABLE 3.1 Details of the Titanic disaster

Survived	Age	Gender	Class			
			1st	2nd	3rd	Crew
No	Adult	Male	118	154	387	670
Yes			57	14	75	192
No	Child		0	0	35	0
Yes			5	11	13	0
No	Adult	Female	4	13	89	3
Yes			140	80	76	20
No	Child		0	0	17	0
Yes			1	13	14	0

The order in which the original rectangle is broken down is, of course, open to choice, and the reader may wish to experiment with different orders to gain some feeling for the potential offered (see Exercises). Many other examples, and a fuller discussion, are provided by Friendly (2000) whose website is well worth visiting in this respect.

Icons

Object visibility is a property of representation techniques in which a single object is so portrayed that a number of its attributes can easily be assimilated, qualitatively if not quantitatively. Two examples are discussed here. One was originally invented to characterize the properties of geological samples, whereas the other was evaluated in the context of house selection.

Professor Chernoff, a statistician at Stanford University, observed that human beings are very sensitive to a wide range of human facial characteristics and suggested that facial features – such as eye size and the length of a nose – are not only numerous (he identified 18) but could, in a cartoon face, take on a sufficiently large number of 'values' to offer a useful encoding mechanism (Figure 3.62). He applied this idea (Chernoff, 1973) to the study of geological

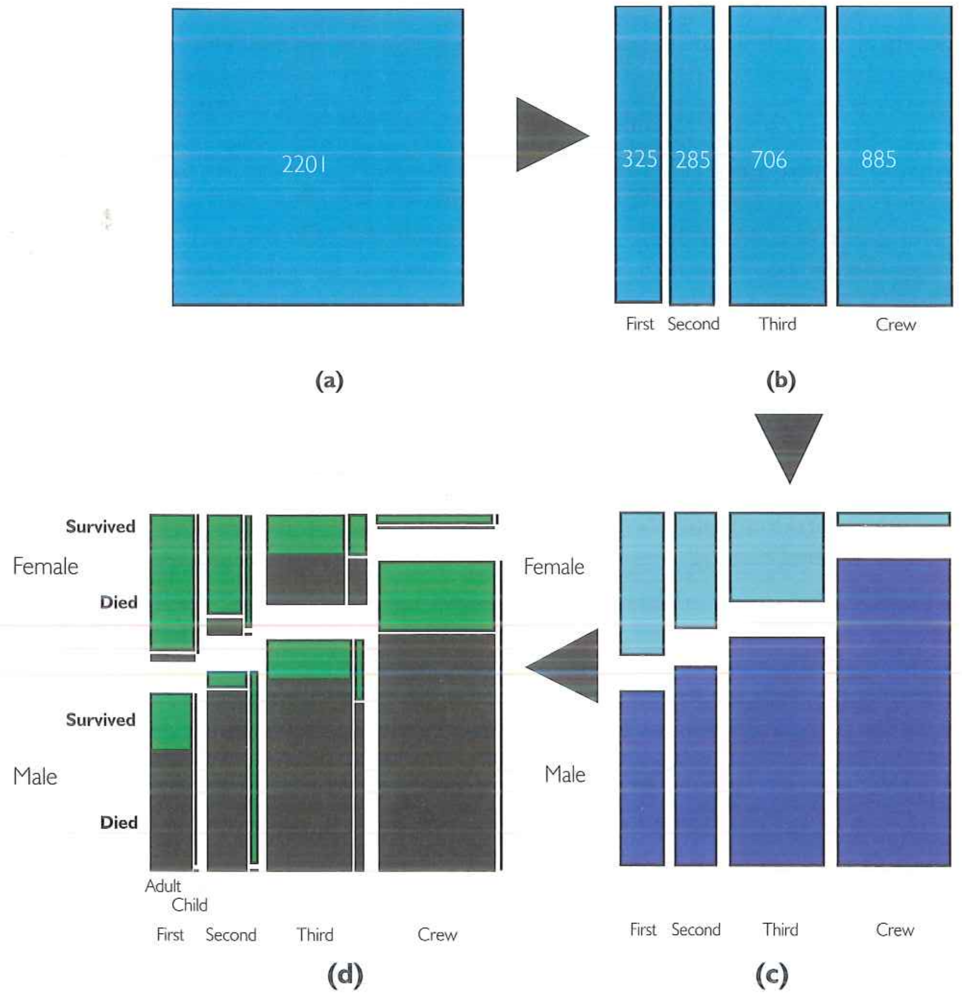

FIGURE 3.61 Steps in the creation of a mosaic plot representing the Titanic disaster

FIGURE 3.62 Chernoff Faces allow attribute values to be encoded in the features of cartoon faces

samples, each characterized by 18 attributes such as salt content and water content, and found that the display of so-called Chernoff Faces facilitated the identification of interesting groups of samples. If this use of computer-generated cartoon faces should appear frivolous it should be remarked that accountants, not usually associated with frivolity, have nevertheless explored the use of Chernoff Faces to display accountancy data (Stock and Watson, 1984). Since one would expect some facial features to be more informative than others (Morris *et al.*, 1999), it is useful to note that a study by De Soete (1986) established the relative value of various facial features. Ware (2004, page 253) comments that we may have special neural hardware to deal with faces.

In the context of house purchase the multidimensional icons shown in Figure 3.63 attempted to portray attributes more directly, for example representing a houseboat by a shape suggestive of a boat and a garage by a rectangular box containing a simple representation of a car. A total of eight house attributes were represented in this way. A question that immediately arises, of course, is whether such multidimensional icons offer any advantage over an equivalent textual description, shown in Figure 3.64. To investigate this issue, controlled experiments were undertaken (Spence and Parr, 1991). Subjects were given a number of tasks to complete, of which the following is representative:

> *You can spend up to £200,000 on accommodation. Locate the best you can with regard to the number of bedrooms and the size of garden, but it must have central heating.*

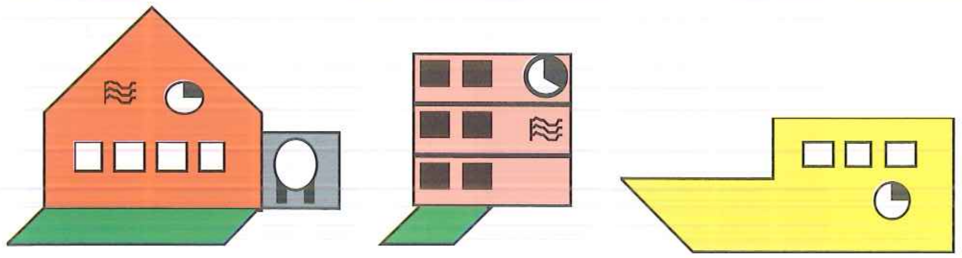

FIGURE 3.63 Multidimensional icons representing eight attributes of a dwelling

house	flat	houseboat
£400,000	£300,000	£200,000
garage	no garage	no garage
central heating	central heating	no central heating
four bedrooms	two bedrooms	three bedrooms
good repair	poor repair	good repair
large garden	small garden	no garden
Victoria 15 mins	Victoria 20 mins	Victoria 15 mins

FIGURE 3.64 Textual descriptions of the dwellings represented by the multidimensional icons shown in Figure 3.63

The subjects were presented with two conditions. Details of 56 dwellings were presented in either iconic (see Figure 3.63) or textual (see Figure 3.64) form, the layout in both cases being seven rows of eight dwellings (Figure 3.65). The 56 dwellings were so designed that there was a unique solution to all the tasks. Overall it was found that the time taken to identify the appropriate house using icons was about half what it was when using textual description. Observation of the subjects while they carried out the tasks strongly suggested that, if the display had been an interactive one rather than printed on paper, some means of tagging individual dwellings – either to 'discard' them or 'include' them – would have simplified the task. With the paper display, Post-it stickers were employed for tagging.

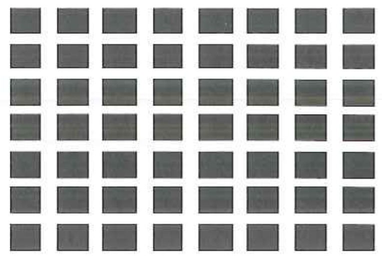

FIGURE 3.65 Layout employed in the experiment carried out to compare the relative merits of iconic and textual descriptions of dwellings

We have discussed two examples of metaphorical icons: in one there is a direct relation between icon and object (the house icon) while in the other there is no direct relation between facial features and the geological attributes they represent. Siirtola (2005) has observed that experiments support the idea that a data-related **'glyph'** (icon) is more favoured by users and, with respect to the task of information acquisition, leads to about 13 per cent more accuracy. Direct metaphorical icons find wide application (see, for example, Miller and Stasko, 2001).

7

Object and attribute visibility

We return briefly to the concept of visibility, discussed earlier in the context of different encoding techniques (Teoh and Ma, 2005).[4]

We can say that the property of object visibility is such that each object is represented as a single and coherent visual entity such as a point. Object visibility is desirable when a user is interested in knowing an object's attribute values in many different dimensions and how different objects relate to one another. Examples include Chernoff Faces, multidimensional icons and star plots (Figure 3.66). We do not include static parallel coordinate plots in this set because an object is then represented by a collection of points complemented by a polyline joining those points and it is extremely difficult either to discern a single object's attribute values or to see how objects relate to each other. The same comment applies to the Attribute Explorer.

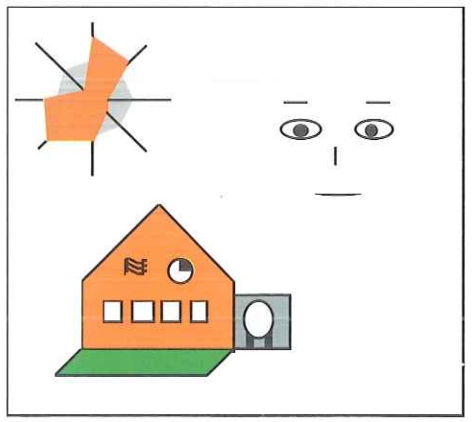

FIGURE 3.66 Representations of multi-attribute objects supportive of object visibility

[4] We use the term 'visibility' in place of 'coherence' used by Teoh and Ma.

A similar form of description applies to attribute visibility, also known as dimension visibility. In a representation with attribute visibility the distribution of objects' attribute values in each dimension is clear. Attribute visibility is desirable when the user wants a clear picture of how objects are distributed in an attribute dimension – for example when there are clusters present. Attribute visibility is associated with the Attribute Explorer and parallel coordinate plots (Figure 3.67), but not with Chernoff Faces, multidimensional icons and star Plots. Teoh and Ma (2005) also introduced the concepts of *object correlation* and *attribute (dimension) correlation*. An example of the latter is provided by a parallel coordinate plot in which points on adjacent axes are connected by lines which clearly indicate a trade-off or correlation. Object correlation may be facilitated by representations such as star plots.

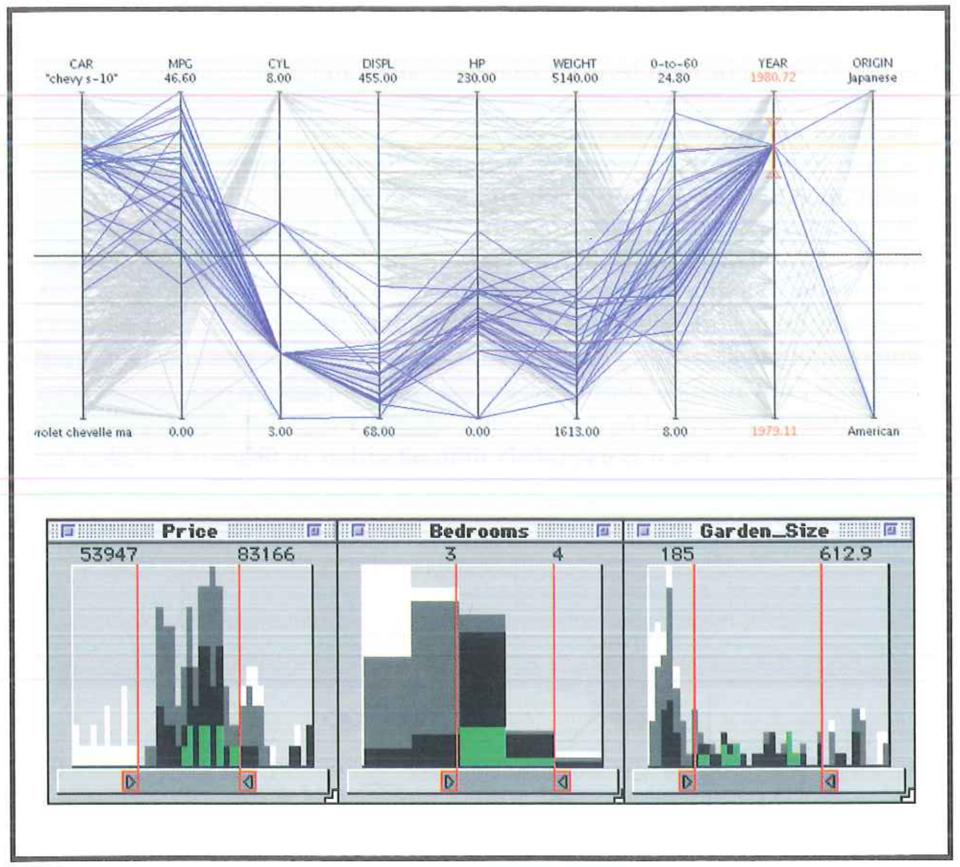

FIGURE 3.67
Representations of multi-attribute objects supportive of attribute visibility

Although the emergence of visibility and correlation as useful characterizations of a representation is quite recent and remains to be validated, it already appears subjectively to be useful and, in addition, seems to have some relation to the 'popping out' property discussed earlier.

3.2 The encoding of relation

In our earlier discussion of data it was established that two types of data exist: *values* and *relations*. We have examined many ways of encoding values; we now ask, 'What do we mean by a relation?' and 'How do we represent it?'.

The dictionary definition says:

> **relation** (n): a logical or natural association between two or more things; relevance of one to another; connection.

A relation between two or more things can be represented in many different ways. A simple straight line or lines can be used to show that John Smith married Mary Robinson (Figure 3.68) or that John borrows money from the Stingy Bank to purchase a 1930 Bentley (Figure 3.69). The relation could be mathematical, of the form y = f(x), and, in turn, could be represented by a node and directed link diagram (Figure 3.70). As shown earlier in Chapter 2, the relation between a car and its price can be represented by a common colour (Figure 3.71). Figure 3.72 summarizes aspects of the warfare in Anglo-Saxon England between 550 AD and 700 AD far more effectively (for me) than an equivalent area of text – the warlike nature of the West Saxons immediately stands out (Arnold, 1997), as does the fact that the Britons continually took quite a beating. Colour and line thickness provide additional scope for encoding aspects of a relation.

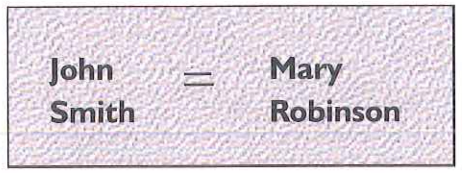

FIGURE 3.68
A simple symbol indicates the relationship of marriage

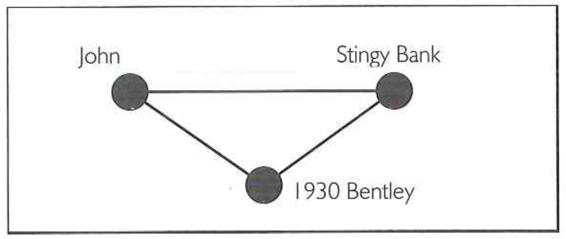

FIGURE 3.69
Lines indicate relationship

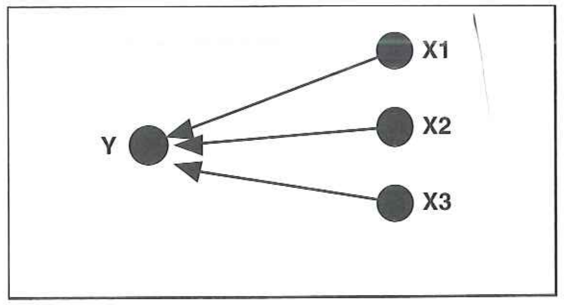

FIGURE 3.70
Arrows indicate unique unilateral functional relations