Graphical displays

Dan Carr

Volume 2, pp 933–960

in

Encyclopedia of Environmetrics
(ISBN 0471 899976)

Edited by

Abdel H. El-Shaarawi and Walter W. Piegorsch

Graphical displays

The purpose of environmental visualization via graphical displays is to facilitate scientific and public understanding of environmental status, trends, and processes. Such understanding is incremental owing to:

- the complexity of the environment;
- the difficulty of parsimoniously conceptualizing this complexity;
- the logistic and political impediments to collecting adequate, representative data;
- and the limits of human perception and cognition for understanding multivariate summaries.

This entry can only hint at the range of challenges faced in attempts to characterize the environment and communicate using graphical summaries. It can only touch on the background knowledge that helps experienced readers to assess and interpret environmental graphics. However, the background is important since environmental visualization goes far beyond routine production and superficial interpretation of environmental graphics.

After a brief introduction, the focus herein is on graphical design principles and graphical templates for representing environmental summaries. The quality of environmental graphics depends on many factors: conceptualization, data collection, modeling, summarization, and, as emphasized here, sound graphical design.

Our understanding of quantitative graphical design continues to evolve. Since one design principle is to use familiar templates, a tension arises between using familiar templates and introducing new templates that offer richer, more focused or accurately communicated content. This entry includes some new templates and uses design principles to motivate their inclusion. The sequence of static templates presented is far from complete. Further, a printed publication has difficulty in doing justice to interactive graphics and no such attempt is made here. This entry attempts to help the reader fill in omissions by providing references into the literature.

Another omission is a discussion of integrating environmental graphics into extended documents. Interested readers can refer to Stone et al. [99] and Wahlström et al. [109] as examples of excellence. The reader may also want to adapt methods and graphics illustrated in the award-winning *Atlas of United States Mortality* [Pickle et al. 89] to environment applications.

Environmental Complexity

The environment is complex. Spatial and temporal processes at different scales influence environmental status and trends. Large external influences include solar radiation and lunar gravity inducing tides. Tiny ocean plankton are crucial to the food chain (*see* **Community food webs**). Mankind is busy rearranging the molecular composition of the earth's land surface, ocean, and atmosphere. Computer models that simulate the earth processes should, at some scale, account for changes potentially induced by mankind's actions, including those motivated by simulations. Our environment hosts the full range of processes to visualize: self-balancing processes, processes sensitive to the flap of a butterfly's wings, and processes influenced by political decisions.

Concepts and Indicators

The development of concepts to characterize environmental status and trends is itself an ongoing process. The seemingly simple status question, 'how many lakes are there in the nation?', depends on the definition of a **lake**. What width, depth, duration and other properties must a body of water have to be a lake? Obtaining consensus can be difficult. The concept of a lake is relatively simple compared with many other environmental concepts. For example, the notion of ecoregions [4–6, 86, 87] is generally accepted as useful but much more difficult to characterize than lakes. Which multivariate descriptors and thresholds should be involved? The process of relating existing ecoregion definitions to other variables continues [29].

The variables available typically start out as field measurements. Environmental scientists are inclined to retain some measurements, such as pH, as they stand. However, scientists often transform the field measurements to produce more useful variables. The percentage of time that a body of water has dissolved oxygen values below a critical threshold may be more instructive than the dissolved oxygen values by themselves. A big challenge is to convert a collection of field measurements into indicators, such as an indicator of biological integrity. (A discussion of indicators appears in [63].) Indicators need to be meaningful across different environmental habitats and their composition often needs to vary from region to region. Efforts continue toward the development of concepts, indicators, and indices that summarize multivariate relationships. Often the goal is to provide a broad univariate characterization of environmental health or safety for humans. Can a single index of **air quality** capture the human health implications of varying air toxins (such as lead, **ozone**, and volatile organic compounds)? The well-known US consumer price index was a compromise that took on a life of its own as an index. Environmental scientists continue to seek analogous simplifications whose utility as communication devices outweighs the dangers of their simplification.

Data, Estimates, Comparison and Interpretation

The sources of environmental data include field samples, survey samples (*see* **Sampling, environmental**), satellite imagery (*see* **Remote sensing**), administrative records and computer simulation. To be useful such data must be transformed into estimates that are scientifically meaningful. Each type of data comes with its own set of issues to address in the process of producing estimates that are worthy of evaluation. Common issues include calibrating instruments, scaling variables, estimating variables as surrogates for the desired variables of interest, adjusting for covariates, assessing representative coverage of the population of interest, and validating simulated or indirect estimates.

Graphics can be no better than the estimates presented. The reader should be concerned whenever quantitative graphics fail to show confidence bounds for estimates. The lack of confidence bounds is often a warning that estimates have not been assessed with respect to accuracy (bias) and precision (variability).

In many cases the available data are inadequate to address the question of interest. In such cases the presentation of tables and graphs derived from unrepresentative or marginally related data promotes the illusion of serious scientific monitoring and assessment. Appropriate interpretation of graphics depends upon understanding the meta-data, the data about the underlying data and resulting estimates. The meta-data provide important information about data quality. The concern about accuracy is not limited to statistical estimates but extends to spatial databases [61].

The heart of graphics is comparison. Quality graphics help the reader to make meaningful comparisons. Consider Figure 1, which shows **times series** of CO_2 production per capita for energy used in OECD (Organization for Economic Cooperation and Development) nations. First note that confidence bounds are not present. This complicates making comparisons. When comparing estimates without confidence bounds, the reader should immediately wonder if the estimates are worthy of comparison. In the absence of confidence bounds, the reader is tempted to make assumptions that may not be justified. A first plausible assumption is that a nation's estimates are comparable over time. (A comparison of 1995 and 1997 OECD compendium values shows that some nations continue to refine their historical estimates.) A much more questionable assumption in studying Figure 1 is whether estimates from different nations are comparable. The methodology that nations employ to obtain estimates can vary greatly, especially in situations involving Third World nations. The reader should interpret the ranking of nations (by 1995 values) in Figure 1 cautiously, not only because some values are very close. While the OECD works toward harmonization of estimates from member nations, the goal is often not in hand. The reader should be aware that nations, like people, are motivated to put themselves in a good light. In discussing maps, Wood [118] says that every map serves an agenda. The same is true for published tables and graphs. Scientific comparison is difficult in the presence of differing degrees of 'good light' estimates.

One class of methods for making estimates look better involves transformations that change the reader's perspective. Showing values per capita in Figure 1 is more favorable to the US than showing the total CO_2 production values. While the top panel in Figure 1 is taller to accommodate a bigger range of values, the US per capita values are roughly on the same scale as other OECD nations. Since the US has a high gross domestic product, an even more favorable view shows CO_2 production relative to the gross domestic product. This suggests that US energy production is more efficient. The agenda influences the choice of transformation.

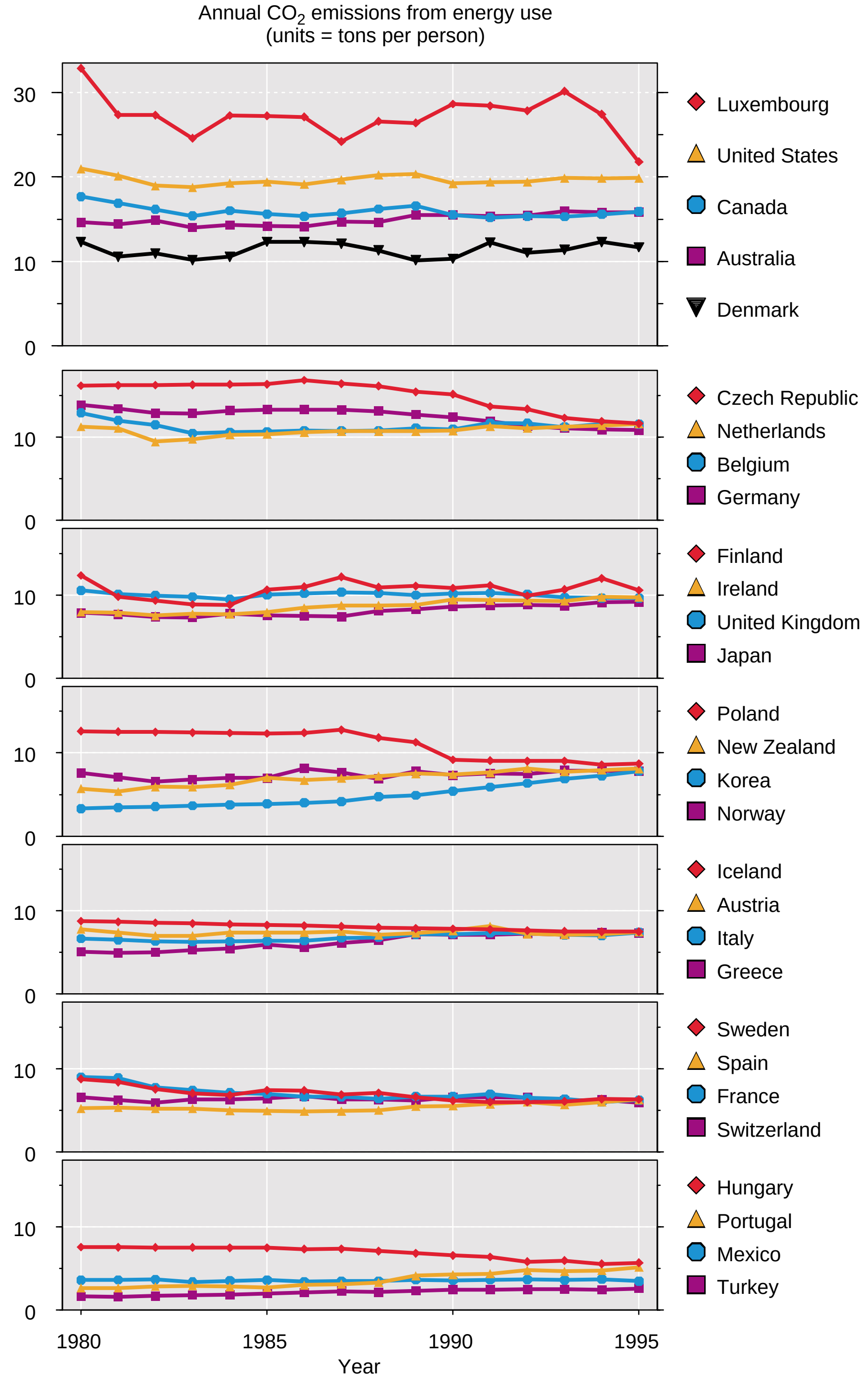

Figure 1 Times series sorted and grouped into panels. Values and ranking are subject to question

Comparability issues are not limited to differences among nations, but arise whenever researchers employ different methods. Historically, the US **Environmental Protection Agency**'s (EPA) STORET database accumulated statistics on hundreds of thousands **water quality** samples each year. Even after the EPA made efforts to provide the data in the same units, comparability issues remained due to different sample collection, chemical analysis, and recording procedures. Those interested in US water quality are inclined to focus attention on the subset of water quality data from the US Geological Survey (USGS), because of the consistent methodology and high standards. Integrating environmental information from multiple sources that use different methods is problematic (*see* **Meta-analysis**). Consequently many data are never used beyond the original study even when made readily available in public databases.

Changes over time produce comparability problems. Political entities and boundaries change. How should the unification of West and East Germany be handled in Figure 1? Measurement and calculation methods tend to improve over time. Researchers are not inclined to make statistical adjustments so that analyses can span estimates based on older data and methods. A common attitude is that the new estimates are exciting and so much better than previous estimates that they should serve as benchmarks for the future assessment of trends. This postpones the estimation of short-term trends.

High-quality, representative environmental data are often difficult and expensive to collect. For example, atmospheric scientists want a detailed snapshot of the whole earth's atmosphere. The logistics and expense of such a massive simultaneous collection effort represent a major barrier. Statistical researchers have developed a representative spatial sampling methodology that could produce estimates with uncertainty bounds for a host of variables [97, 98]. Olsen et al. [83] provide an overview of its use in national monitoring. Studies [65, 88] have used the methodology for various regions within the US, but the methodology is not employed at a national scale in the US, presumably due to the expense. Also, it is naive to be unaware of corporate and political interests opposed to the collection of environmental data, while privacy considerations make it difficult to access the collected data even as advances in databases and web technology would seem to improve access. The lack of available, high-quality, representative data is a common problem in environmental visualization.

Figure 2, a multiple panel bar plot of EPA's toxic release inventory estimates, raises numerous comparison and interpretation issues [18]. Note first the absence of confidence bounds. The estimates derive from federally mandated company self-reports. Without substantial efforts to obtain external validation measurements, the reader is not in a position to access the accuracy of the self-reported estimates. Assuming comparability across states, the reader infers that the state of Louisiana has a surface water problem, but how much better or worse might it really be? EPA publications make it clear that the survey does not include all release sources nor all kinds of toxic releases. This suggests that the real totals are higher.

Another interpretation problem with Figure 2 is that the unit of measurement, total pounds, is an index comprised of many different things. This index is not well suited for purposes of communicating risk and the personal implications of being part of a toxin-distributing society. If the index decreases 2% from one year to the next, is that good? What does a percentage decrease in pounds mean to the US in terms of genetic mutations (*see* **Mutagenesis, environmental**), **morbidity and mortality** in plant, animal and human populations? The public may infer that fewer pounds are better than more pounds, but that can be wrong due to the changing mixture of toxins involved. Even if the mixture keeps the same proportions, continuing accumulations of toxins may lead to increased risk. Interpretation of Figure 2 is problematic, even though the multiple panel bar chart with panel sizing to make comparable scales is an excellent template.

Research in environmental visualization should develop indices to help people to understand the risks involved. An index of toxic pounds per person makes the index more personal. While such an index may draw attention to the implications of living in the society, it does not translate the index into quantitative implications. The task of understanding and communicating risk is challenging. For very simplistic scenarios it is possible to do some calculations. In terms of one effect, mortality, scientists could use LD_{50}s for mice (the dose that is lethal to half the exposed mice; *see* **Biological assay**) to calculate how many mice could be killed by direct exposure assuming additive effects. This scenario ignores such things as exposure pathways, transformations of the toxins

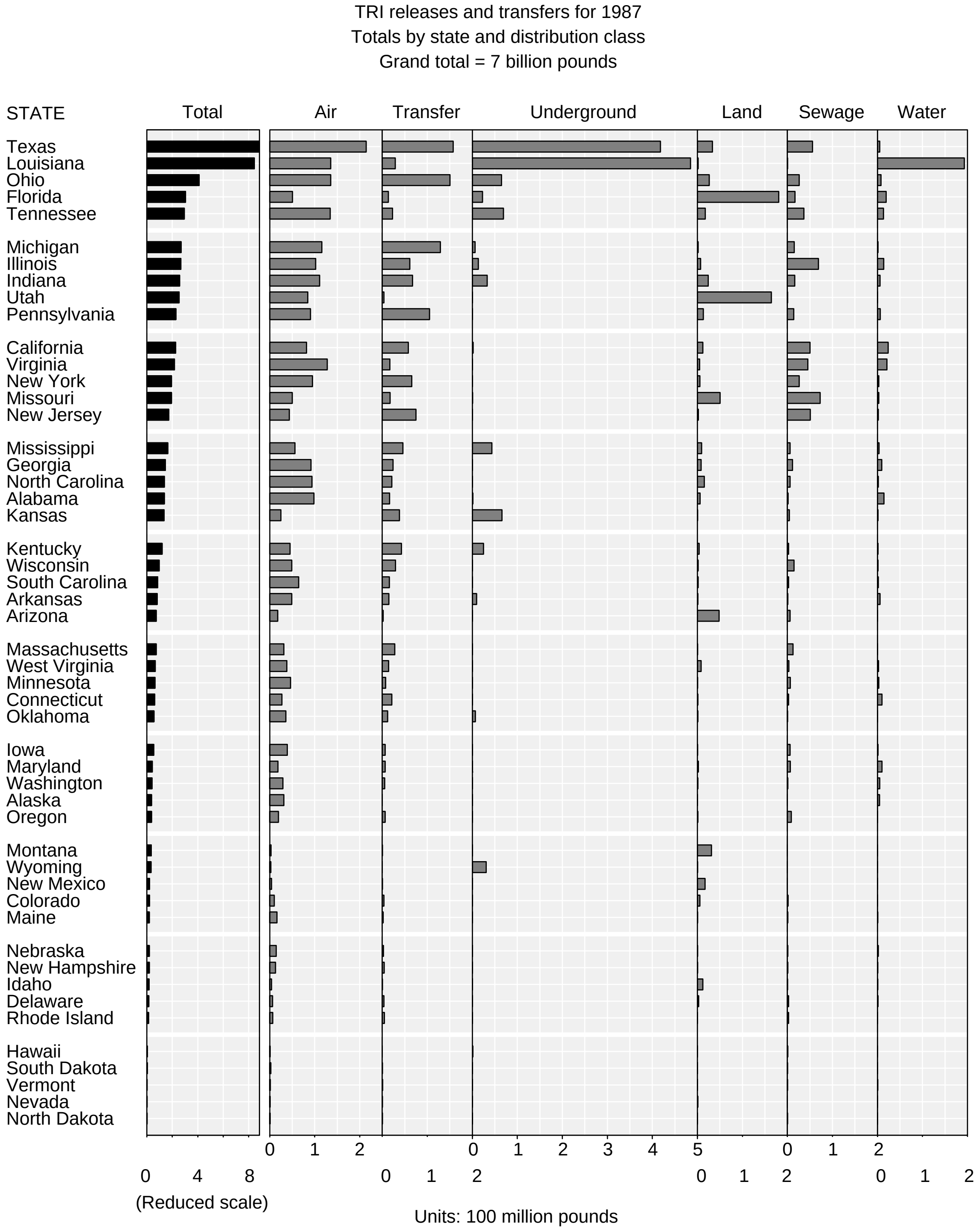

Figure 2 A multiple panel bar plot with perceptual grouping. Panel widths vary to make bar lengths comparable. Estimates are self-reported. Interpretation in terms of toxicity is also problematical

along the pathways, differing within-species susceptibilities, and toxic interactions. Species differences in susceptibility complicate scaling results to other species where the LD_{50}s are not known.

Knowledge about actual effects of exposure to multiple toxins is limited, in part due to the overwhelming combinations to be studied. It is interesting to note that the majority of USGS water samples containing one pesticide actually contain more than one. At the same time, the EPA safety standards address pesticides individually, as if the presence of multiple toxins does not change response thresholds (*see* **Joint action models**).

A mortality index does not incorporate outcomes such as mutations, cancer incidence (*see* **Carcinogenesis, environmental**), or the loss of genetic variety that goes with species extinction. Any index that begins to communicate risk is likely to be controversial. Much is known about mapping hazards [82] but developing understandable, scientifically and politically acceptable indices remains a challenge.

The quality of graphics depends heavily upon the quality of data summaries or estimates being represented. There are many ways of producing estimates for summaries. Designed sampling studies provide one source of estimates. Models operating on the data that happen to be available provide a more common source of estimates. If the available data are not adequately representative of the underlying phenomena, then model estimates, however sophisticated in terms of handling spatial and temporal correlations, can miss the mark. Often direct data are not available for regions of interest, so analysts use models developed in other locations or circumstances and covariates for the regions of interest to produce estimates. The available covariates may not be adequate for adapting models to different situations. Privacy issues and the expense of collecting observational data have motivated the development of many simulation models. Sometimes researchers validate simulation results against observational data and sometimes they do not or can not. With some understanding of estimate limitations and interpretation problems in hand, the next step is to describe graphics templates developed to communication estimate descriptions and summaries.

Templates for Environmental Graphics

Environmental complexity motivates the use of multivariate graphics templates. Univariate and bivariate graphics provide starting points, as building blocks for more multivariate graphics.

Univariate Guidelines

Cleveland and McGill [39] discuss human perceptual accuracy of extraction and indicate preferred methods for univariate encoding. Their research subjects judged relative magnitudes of graphically encoded variables. Their results ranked the graphical encoding methods into three classes described here as best, good, and poor.

The two best encoding methods represent variables using position along a common scale as shown in Figure 3 and position along identical nonaligned scales. That humans do well in judging the position of a point relative to scale should come as no surprise. Marr [77] notes the 'quintessential fact of human vision – that it tells us about shape and space and spatial arrangement'. Locating the position of objects is a fundamental visual task. Map makers have long used position along a scale as the fundamental encoding for spatial coordinates. MacEachren's [74] review of the perception literature attests to the power and primacy of positional encoding.

Length, angle, and orientation are good encodings. Figure 4 shows that transforming line segments into a standard position converts the task of judging length into a task of judging the position of one endpoint against a scale. While this is not necessarily what people do, the example suggests that judging line length is more complicated than judging position.

Figure 5 shows angle encoding. Rotation of the angles puts them in a position for comparison against equivalent angular scales shown in gray. The transformation suggests that while angle comparisons work

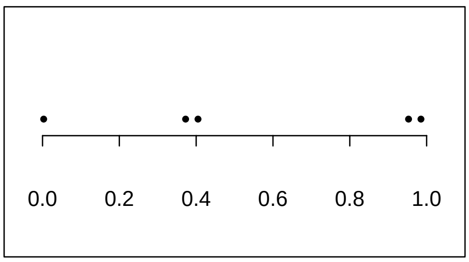

Figure 3 The best continuous univariate encoding – position along a scale. Reproduced from the *Encyclopedia of Biostatistics*, Vol. 4, pp. 2864–2886, by permission of John Wiley & Sons, Ltd. © 1998

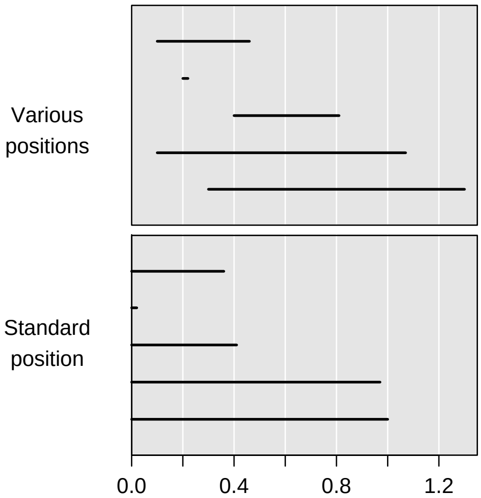

Figure 4 A good continuous univariate encoding – line length. Reproduced from the *Encyclopedia of Biostatistics*, Vol. 4, pp. 2864–2886, by permission of John Wiley & Sons, Ltd. © 1998

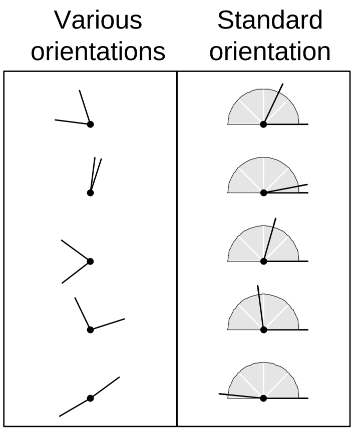

Figure 5 A good continuous univariate encoding – angle. Reproduced from the *Encyclopedia of Biostatistics*, Vol. 4, pp. 2864–2886, by permission of John Wiley & Sons, Ltd. © 1998

reasonably well, they are more complicated than direct comparison against angle scales.

Area, volume, point density, and color saturation are poor encodings. The reader familiar with experimental results involving Steven's law will not be surprised about poor results for the area and volume encodings. Steven's law states that the perceived magnitude of a stimulus follows a power law:

$$p(x) = ax^b \tag{1}$$

where x is the magnitude of the true stimulus (i.e. length, area, volume), and where the constants a and b depend on the type of stimulus. Based on values cited in Baird and Noma [7], Table 1 provides ranges of the characteristic exponents b for length, area, and volume. That is to say, people's perception of length tends to be directly proportional to object length. However, we tend to judge area and volume nonlinearly. Consider comparing areas, one of 4 square units and the other of 1 square unit. With an exponent of 0.75, the ratio of perceived magnitudes is not 4 to 1, but 2.8 to 1. We underjudge the large areas relative to small areas. If everyone had the same exponent, graphical encoding could adjust for systematic human bias. However, the range of values for b in Table 1 indicates substantial variability from person to person. Providing a set of reference symbols in a legend helps people calibrate to the intended interpretation, but the best strategy is to use better encodings whenever possible.

Weber's law, a fundamental law in human perception, also has important ramifications in terms of accurate human decoding. A simple example gives the basic notion of the law. The probability of detecting that a 1.01 in. line is longer than a 1 in. line is about the same as the probability of detecting that a 1.01 ft line is longer than a 1 ft line. In absolute terms 0.01 in. is much smaller than 0.01 ft. The use of a finer resolution scale allows more accurate judgments on an absolute scale. A common application is to put tic marks on a ruler to help us make more accurate assessments. The graphical equivalent [36] is to use grid lines to provide a finer resolution scale for more precise comparisons. In interactive graphics, zooming in provides a finer scale. Computer–human interface implementations often provide sliders that allow the reader to change the reference scale to make more accurate judgments (see, for example, [1], [50] and [92]).

Table 1

Encoding	Exponent range
Length	(0.9, 1.1)
Area	(0.6, 0.9)
Volume	(0.5, 0.8)

We render most graphics on a plane. We could show values of a continuous variable as points on a line, but there are good reasons for not doing so. For categorical variables using bar chart and pie charts to show percentages is common and dot plots could be used. From a perceptual accuracy and labeling convenience viewpoint, bar charts and dot plots are preferable. Decoding bar charts and dot plots involves judging position along a common scale while pie chart decoding compares angles at different positions. The bar chart vs. pie chart controversy is old. The merit of pie charts is that the reader assumes that percentages add to 100. Bar charts and dot plots can handle this with a footnote if the context does not make it obvious. The labeling alone for all of these forms demands a planar or higher-dimensional representation.

Bivariate Guidelines and Examples

Tufte [102] notes that it took over 5000 years for mankind to generalize from the early use of clay tablet maps to representing other kinds of point pairs with a scatterplot. It is an excellent representation since the two orthogonal axes allow two coordinates to be encoded independently as position along a common scale. While the popular press in the US still considers the scatterplot too complicated for the general public, in the sciences the scatterplot is the standard for representing continuous bivariate data. Common bivariate activities include assessing univariate distributions, comparing univariate distributions, and looking for functional relationships.

We humans do not assess data density accurately even when points are plotted on a line or on a plane. Overplotting just makes things worse. Consequently, it is advantageous to compute data density and show it as directly as possible. Figure 6 illustrates the construction of a kernel density estimate (*see* **Meteorological extremes**) based on a sample of five univariate values. The locations of the white triangles relative to the x-axis indicate the magnitudes of observed values. The basic idea is that each observed value is a surrogate for values in a neighborhood. We then associate a relative likelihood with a neighborhood about the value. Figure 6 shows the five likelihoods (or kernels) as bell-shaped curves, one in each of the upper panels. For each location where we want to estimate the data density (the x locations of white lines in Figure 6) we simply average the five likelihoods at that location, one for each observed value. The white lines indicate the locations of the density estimates. In the bottom panel each white line is the average height of all the white lines directly above it. (When panels show no white lines directly above, they contribute zero to the average.) The construction is straightforward.

Scott [93] provides the theory behind density estimation along with many graphical examples for univariate and higher-dimensional density estimates. For a valid density estimate, the kernel needs to integrate to 1. The hard part is in deciding how wide to make the kernel. Scott describes methods for making this decision.

In the environmental sciences, cumulative distribution plots and quantile plots are commonly used to describe the distributions of populations. The two types of plots are essentially the same, being transposes of each other. For quantile plots the x-axis shows cumulative probabilities and the y-axis shows sorted observed values. Figure 7 is a quantile plot. The construction plots cumulative probabilities against the sorted values from a random sample. While integrating an estimated density function up to a value approximates a cumulative probability, there are two more common approaches to calculating cumulative probabilities: the order-statistics approach and the empirical approach. The order-statistics approach here follows Cleveland [35] and uses the expression $(i - 0.5)/n$ for $i = 1, \ldots, n$ to calculate the cumulative probabilities, where n is the sample size. While often used, the empirical approach yields probabilities that imply future values will never be more extreme values than those already observed. Here the guessed probability is $0.5/n$ that a smaller value could be observed and $0.5/n$ that a larger value could be observed. This is usually a minor detail for large samples. (The construction can be adapted if the observations are not equally representative of the population of interest. See Cook et al. [42] for a definition of a spatial cumulative distribution function and an application to a crown **defoliation** index.) To finish the construction, Cleveland interpolates between the point pairs. This produces a piecewise linear curve.

Interpretation is straightforward. For any probability covered on the x-axis it is possible to determine a quantile. To obtain the 0.5 **quantile** (or estimated median) go straight up from 0.5 on the x-axis to

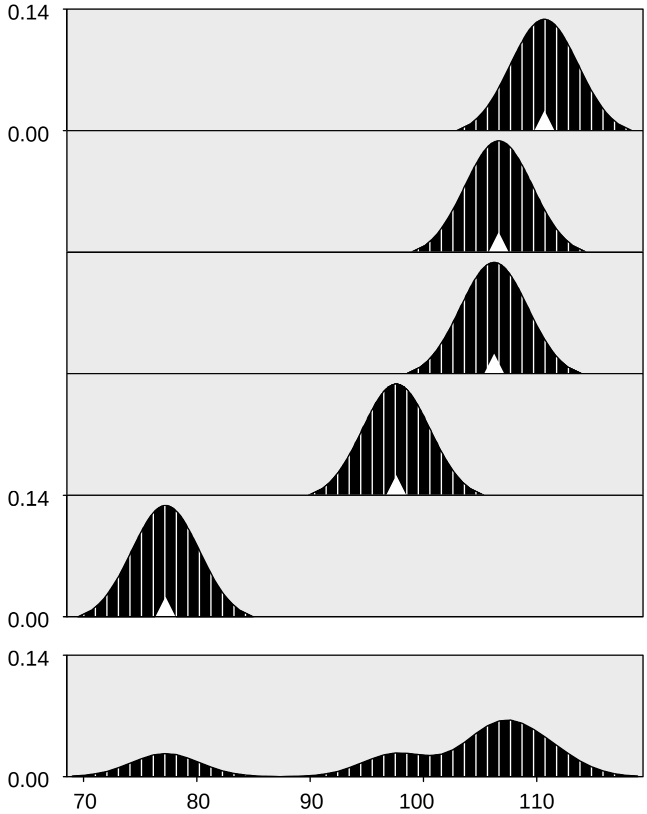

Figure 6 The construction of a kernel density estimate

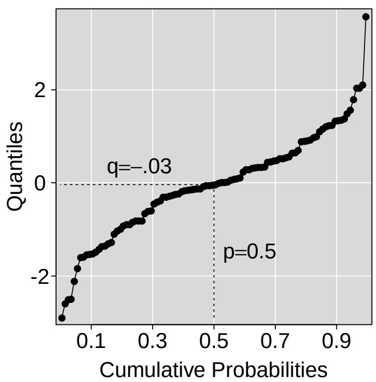

Figure 7 A quantile plot. The coordinates for the *y*-axis are typically the sorted observations. The coordinates for the *x*-axis are the corresponding cumulative probabilities. The quantile corresponding to the cumulative probability of 0.5 is also known as the median. Quantiles for other cumulative probabilities can be found graphically or by linear interpolation

the curve and then straight across to the *y*-axis and read the value. Starting with 0.25 and 0.75 yields corresponding quantiles also known as the 1st and 3rd quartiles, or 25th and 75th percentiles, respectively. Similarly one can go from quantiles to cumulative probabilities. Since scientists use such plots to describe collections of data in a database as well as environmental populations, the hardest interpretation task is often to decide if inference about an environmental population is appropriate.

Quantile or cumulative distribution plots are useful for characterizing environmental populations. For example, quantile plots can indicate the fraction of lakes in a defined region that have eutrophication values below a given threshold. Quantile plots are helpful on maps and provide a frame of reference for observing change over time. For maps, Carr and Olsen [23] highlight selected cumulative probability and quantile pairs using a parallel coordinate approach to save space. They note the importance of

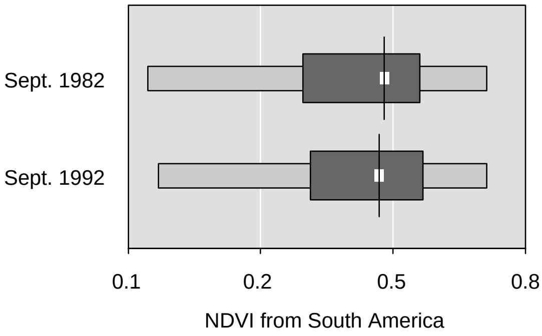

Figure 8 A variation on boxplots. The median: a long horizontal line. First and third quartiles: ends of thicker boxes. Adjacent values: ends of thinner boxes. Outliers: open circles (none present). Test intervals for different medians: white lines inside boxes. NDVI: normalized difference vegetation index

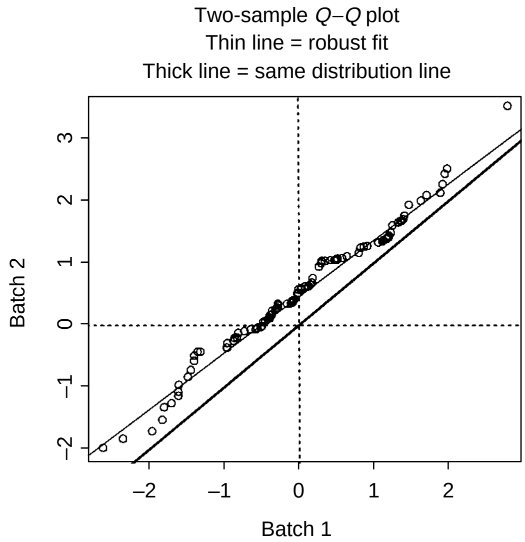

Figure 9 A two-sample Q–Q plot. A good straight line fit suggests similar distributional shapes. Given similar shapes, the slope shows the ratio of scale parameters such as standard deviations. Given a slope of one, the intercept shows the difference of location parameters such as the two means. Reproduced from the *Encyclopedia of Biostatistics*, Vol. 4, pp. 2864–2886, by permission of John Wiley & Sons, Ltd. © 1998

considering which distribution to show. For example, in mortality rate mapping, **geographic information system** (GIS) defaults will typically base the legend on the number of regions, and not other variables associated with regions. Showing the fraction of people living in regions with human mortality rates below given values is much more to the point than showing the fraction of regions with mortality rates below given values. Thoughtful selection of the distribution can lead to more meaningful quantile plots.

The **boxplot** is a distribution caricature that has achieved wide acceptance. Although it is used to represent individual distributions, the common use is to compare distributions. Figure 8 provides an example of a set of boxplots. The features shown include the median, quartiles, adjacent values, and outliers. Cleveland [35] describes the determination of adjacent values and outliers. Variations [57, 78] may show extrema rather than adjacent values and outliers. The design variation in Figure 8 uses a white line [19] to provide comparison intervals for the medians. If two comparison intervals do not overlap, then the medians are significantly different.

Q–Q plots provide the preferred graphic to make detailed continuous distribution comparisons [35]. For theoretical distributions, the cumulative distribution function $F(\cdot)$ provides the correspondence between the probability and quantile pairs via $p = F(q)$. In simple cases the quantile function $Q(\cdot)$ is the inverse of $F(\cdot)$ and $Q(p) = q$. Familiar pq pairs from the standard normal distribution are (0.5, 0) and (0.975, 1.96). Comparison of two distributions, denoted 1 and 2, proceeds by plotting quantile pairs $[Q_1(p), Q_2(p)]$ over a range of probabilities, such as from 0.05 to 0.95 in steps of 0.05. For two distributions of observed data, the calculations described for the quantile plots (above) are appropriate. Figure 9 shows a Q–Q plot for two batches of data. The x-axis shows quantiles from Batch 1 and the y-axis shows quantiles from Batch 2. Sometimes statisticians choose $Q_1(p)$ to be from a theoretical family of distributions, such as the normal family, to see if parametric modeling is reasonable using the family of distributions.

A strong merit of Q–Q plots is that in simple cases they have a nice interpretation. If points fall on a straight line, then the distributions have the same shape (basically, the same moments higher than two). This is the case in Figure 9 since the robust fit thin line matches the quantiles quite well.

The slope and intercept of the approximating straight line tell us about the discrepancies in the second moment (scale parameter) and first moment (location). The slope of the thin line tells us about the ratio of the scale estimates (for example, standard deviations). The thick line in the figure is the reference line for identical distributions. The lines are not quite parallel in Figure 9 so standard deviations are not quite the same. Graphical fitting of the scale ratio and location difference can start by guessing the ratio and dividing this into the y-axis quantiles until the lines are parallel. When the lines are parallel, the vertical distance between the two lines gives the difference in location (or means) given the Batch 2 rescaling. In Figure 9 the lines are nearly parallel so a reasonable guess is that the distributions differ in location by about 0.5.

$Q-Q$ plots avoid the visually deceptive procedure of superimposing two cumulative distribution functions or two survival curves. As Figure 10 suggests, humans are really poor at judging the distance between curves. Our visual processing assesses the closest differences between curves rather than the correct vertical distances [36]. Adding grid lines can help, but it is often better to plot the difference explicitly or make comparisons using $Q-Q$ plots.

It is straightforward to associate quantiles from three or more distributions based on the same cumulative probability. Jones and Cook [68] have generalized $Q-Q$ plots to higher dimensions and application of this is worth considering.

Before-and-after comparisons are common in science. The general idea is to control for the variation in experimental units by studying the change in experimental unit values. This differs from $Q-Q$ plots in that the study unit is the basis for pairing rather than cumulative probabilities. Figure 11 shows a paired comparison plot for two low-resolution satellite images of the same region. The traditional reference line for equality is a 45° line through the origin. Figure 11 also shows a mean and difference plot as proposed by John Tukey and described in Cleveland [35]. The x-axis shows the mean of the paired values and the y-axis shows the difference. This transformation rotates the plot so the reference line for identical values is a horizontal line at zero. Making transformations to simplify the visual reference is an important graphical design principle. Statisticians often study the variability in distributions of data. They find that the variability of differences often increases as the mean increases. To assess the variability of differences, they sometimes plot the square root of absolute difference vs. the mean and then fit a smooth line. Cleveland calls this a spread–location

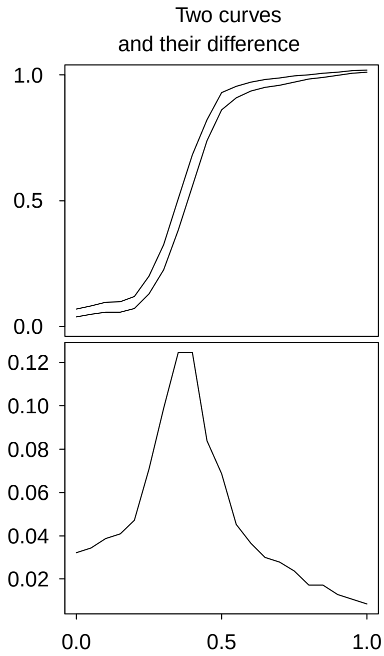

Figure 10 Explicit difference of two curves. Humans tend to see closest differences between curves, not differences in the y direction. Reproduced from the *Encyclopedia of Biostatistics*, Vol. 4, pp. 2864–2886, by permission of John Wiley & Sons, Ltd. © 1998

plot. This is one way to use the scatterplot in studying functional relationships. The general topic is discussed below.

Before proceeding to functional relationships it may be helpful to comment on the extension of $Q-Q$ plots to multiple distributions. There are two basic approaches. The first shows all paired comparisons using a scatterplot matrix, and the second establishes a common reference distribution and makes all comparisons against the common reference distribution. An example of the latter appears in Figure 15 below.

Functional Relationships and Smoothing

When y is considered a function of x, common practice is to enhance scatterplots of (x, y) pairs by adding a smooth curve. To avoid the considerable human variability in sketching an eyeballed fit, the standard procedure is to model data using a computational procedure that others can replicate.

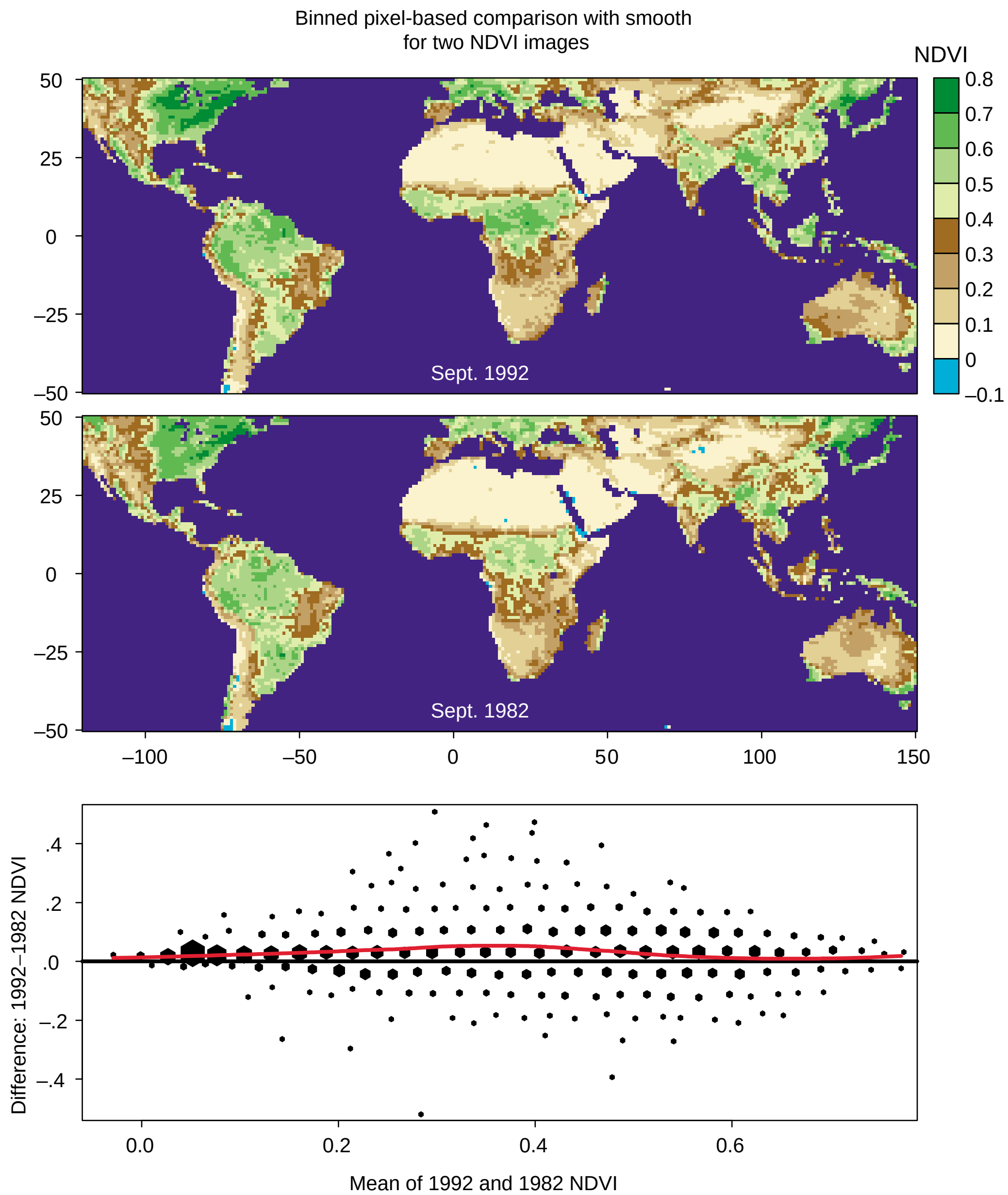

Figure 11 Paired comparison of grid cells using differences and averages. Hexagon binning and symbols avoid overplotting. Interpretation issues include equal angle cells rather than equal area cells and mixed grid cell composition. NDVI: mean normalized difference vegetation index

Figure 12 shows a scatterplot with a smooth line generated using LOESS (LOcal regrESSion) (see [35], or the entry on **nonparametric regression model** for more details). LOESS smooths the data using weighted local regression. That is, the regression uses data local to x_0 to predict a value at x_0. Points closest to x_0 receive the greatest weight. The use of many local regressions produces a set of pairs (x, y) that comprise the smooth curve. Each regression in the smooth shown in Figure 12 used a linear model in x

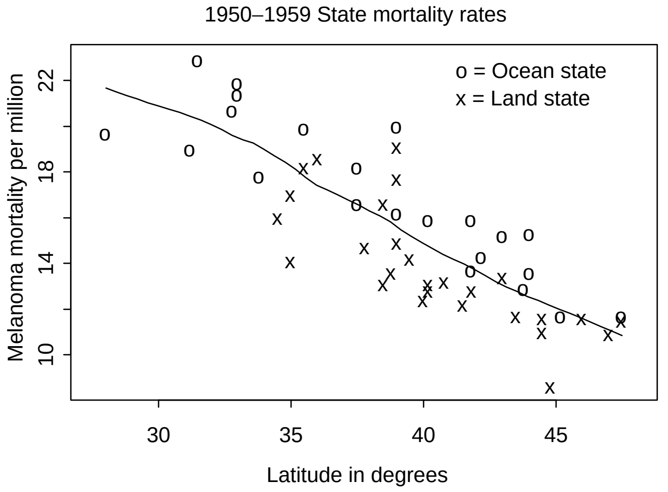

Figure 12 A smooth for scatterplots. An explicit smooth suggests the same functional relationship to different people. Reproduced from the *Encyclopedia of Biostatistics*, Vol. 4, pp. 2864–2886, by permission of John Wiley & Sons, Ltd. © 1998

and included the closest 60% of the observations to the prediction point x_0. Those with the data [52] and the algorithm can reproduce the smooth. The smooth in Figure 12 draws further attention to the distinction between ocean and land states and additional modeling is appropriate. A first step might be to smooth the ocean and land states separately.

Smoothing is an extremely important visual enhancement technique. It helps us to see the structure through the noise. The decomposition of data into smooth and residual parts is fundamental in statistical modeling. Hastie and Tibshirani [64] provide a good introduction to smoothing methods. Their description includes **generalized additive models** that cover more situations than LOESS.

Numerous smoothers are available. Historically, many researchers used cubic splines as smoothers. Cubic splines have a continuous second derivative and that is sufficient to make curves appear smooth to humans. The elegant mathematical formulation behind splines increased their popularity in segments of the statistical community (*see* **Splines in nonparametric regression**). However, there is no a priori best smoother. New methods, such as the **wavelet** smoothing in Bruce and Gao [15], keep appearing in statistical software. Different smoothers have different merits. Recently developed wavelets smoothers are better than many smoothers (but not necessarily all smoothers) at tracking discontinuities in the functional form. The older local median smoothers still do well at handling discontinuities.

Smoothers typically have some form of smoothing parameter that needs to be estimated or specified by the user. With computational power at hand, **cross-validation** methods have become increasingly popular as a community standard. This reduces the judgment burdens on the analyst, but of course does not guarantee a match between an empirical curve and a hypothesized true but unknown underlying curve. Hastie and Tibshirani [64] discuss cross-validation for moderate-sized applications. Golub and von Matt [60] discuss generalized cross-validation for large-scale problems.

Multivariate Visualization

Environmental visualization is inherently multivariate. Environmental scientists are interested in the relationships between many attributes and the attributes have space–time coordinates. The purpose of multivariate graphics is to show multivariate patterns and to facilitate comparisons (*see* **Multivariate data visualization**). As in low dimensions, the patterns concern population distributions or models with at least one dependent variable. After converting attributes and space–time coordinates to images for evaluation, human comparisons typically fall into three categories: comparison of external images with each other; comparisons of external images with external visual references; and comparison of external images with the analyst's internal references. These internal references include scientific knowledge, statistical expectations, and process models. The visualization investigation process often seeks to convert internal references into external visual references subject to further manipulation. With external images and references available, the next step often involves transformation to simpler forms in terms of our perceptual–cognitive processing abilities.

Multivariate graphics must deal not only with the noise that obscures patterns but also with the challenge of conveying the structure in large datasets with relationships that are much higher than two-dimensional (2-D). Advances in remote sensing provide difficult challenges for visualization. Imagine trying to view 30 m resolution land cover of the continental US [108]. A back of the envelope calculation suggests this will take over 7000 workstation images each with 1024×1280 pixels. This only addresses intensities for one spectral band. Researchers have used hyperspectral **image analysis** to partition pixels into the constituents of their mixtures, so pixel description is bound to become increasingly detailed. Modeling is becoming increasingly important to reduce the information to structure that is suitable for human visualization and understanding.

Databases providing geospatial frameworks for modeling continue to evolve. Frameworks for analysis include the US National Hydrography Dataset (NHD), a comprehensive set of digital spatial data that contains information about surface water features, digital elevation, groundwater flow and age (*see* **Groundwater monitoring**), soils, climate, and land cover. For example, the NHD at 1/10 000 resolution is gradually being upgraded to 1/24 000 resolution. The inclusion of smaller streams and their connections will impact the study of contaminant transport and fate using such computer modeling programs as SPARROW (SPAtially Referenced Regression On Watershed attributes) [91]. Urban planners will be able to take more streams into account. Frameworks influence visualization in many ways.

The spatially detailed presentation of estimates can be controversial. Obtaining good small area estimates is often problematic, but because interest increases as observations get closer to home, there is pressure to produce local estimates. As an interesting small area example, EPA staff modeled 1990 long-term cumulative concentrations of 148 hazardous air pollutants (HAPs) for the 60 803 US census tracts in the 48 contiguous states [100]. Public officials suggested that the EPA should not release the estimates because the underlying 1990 data were old and limited. The public could be unduly concerned and decisions to move to apparently safer places could be misguided. Ostensibly, the decision not to distribute the estimates centered on estimate quality and the difficulties in communicating this quality.

Graphical Design Principles

The above description begins to demonstrate the enormity of the multivariate visualization challenge. At the same time, Kosslyn [70] warns that 'The spirit is willing but the mind is weak.' We should approach the challenge prepared to do battle. As Cleveland [35] says, 'tools matter'. Our tools are design principles and templates. Some of our tools include:

- distributional caricatures such as boxplots to help us deal with large datasets;
- map caricatures that let us show small multiples;
- modeling to reduce noise and complexity;
- layering and separating to manage the information flow;
- partitioning and sorting to promote and simplify comparisons;
- linking to peek into higher dimensions.

The basic formats for comparison graphics include juxtaposition, superposition, or the direct display differences. The art of multivariate graphics is to select the methods and enhancements that work best in view of the phenomenon's complexity view and in view of human perceptual and cognitive limitations.

This entry cannot begin to cover all the representation tools and principles. A few pointers to the literature may help the reader explore several facets of multivariate graphics. MacEachren [73] provides a readily accessible primer on symbolization and design. The classic covering a wide variety of visual symbols and signs is Bertin [11]. Grinstein and Levkowitz [62] cover perceptual issues in visualization. Kosslyn [70] provides a gentle introduction to the application of human perception and cognition in graph design, along with excellent references into the literature. MacEachren [74] gives an extended treatment on how maps work.

Foley et al. [53] provide an extensive overview of computer graphics methods. The methods are most immediately relevant to low-dimensional visualization. Wegman and Carr [112] cover selected computer graphics methods and address issues in perception and connections to statistical graphics.

Gnanadesikan [59] covers many of the basics in multivariate statistics, and numerous texts have followed. The multivariate analysis literature deals with important methods such as **clustering**, **classification**, **factor analysis**, **discriminant analysis**, and dimension reduction that are not described here.

Early work in multivariate statistical graphics provides a continuing source of ideas. Fienberg [51] provides an early review. Barnett [8] contains a stimulating collection of papers. The work of John Tukey (see Cleveland [37]) had a profound influence on statistical graphics and is a third resource worth revisiting.

Cleveland's recent books [35, 36] capture much of his protracted efforts to guide scientists toward superior statistical graphics. Cleveland and McGill [38] provide an early survey on dynamic multivariate graphics that foreshadows the visualization revolution in computer science. Tufte [102–104] puts principles to work and draws attention to works of elegance and beauty that appear on the printed page.

Much literature is available on the use of color. A good starting point is Brewer [12] and Levkowitz [71]. Humans are very sensitive to a dark-to-light scale that is referred to in the literature by terms such as value, lightness, or brightness. This is an ordered scale and very important in visual interpretation. Friedhoff and Benzon [55] describe three visual processing channels, especially a high-resolution dark-to-light channel. Humans get their shape information and many depth cues (linear perspective, interposition, shadow, and detail perspective) through this dark-to-light channel. Tufte [103] and others warn that when rainbow colors represent an ordered variable, lightness jumps create unintended edges and patterns that can be confusing. Brewer [13] cites numerous papers opposed to the use of the spectral ordering, but found that when using spectral color to represent a few ordered classes the approach did well in usability studies after reducing the brightness of yellow to be more consistent with the neighboring spectral colors.

In addition to brightness, the literature also describes two other color dimensions, i.e. saturation and hue, along with many other trivariate descriptions of color. (There are many other related descriptions.) A saturation scale goes from an achromatic color such as medium gray, to a saturated color such as vivid red. This scale is also ordered so it can represent an ordered variable. However, saturation supports fewer distinctions than a dark-to-light scale. The hue dimension can be thought of as a circle that includes points between the colors red, yellow, green, cyan, blue and magenta. Hue is not an ordered scale and is good for distinguishing six or fewer categorical variables. Wilkinson [116] cites literature indicating that humans perceive hue and brightness as integral dimensions, so we should not use them to encode two variables. Additional color choice considerations apply to people with impaired color vision. Brewer et al. [14] and Olson and Brewer [85] provide guidance.

As more work is done in computing environments, issues around the computer–human interface become increasingly important. Card et al. [16] edited a book of readings that gathers many important concepts.

The computing revolution has increased access to and usage of visualization methodology by all disciplines. For example, people routinely get maps from the internet to help in their travels. However, the progress in quantitative graphics has been slow in terms of common application. The simple, elegant dots plots promoted by Cleveland in the mid-1980s are hard to find in publications. This is due partly to the limited options available in highly used spreadsheet graphics. The presence of Wilkinson's [116] book *The Grammar of Graphics* and the corresponding JAVA implementation suggests that a graphics revolution is about to take place. Environmental scientists may soon find it easy to produce graphics that follow some of the templates illustrated here and in other documents generated by those with special resources.

The evolving literature on human perception and cognition provides the foundation for graphical design principles. In terms of quantitative graphics, the grand design goal may be stated as to reduce the cognitive effort required to make appropriate comparisons and decisions. Since most people can work with four items of information, this entry elaborates this goal into four broad categories of quantitative design principles:

- use encodings that have high perceptual accuracy of extraction;
- provide context for appropriate interpretation;
- strive for simple appearance;
- involve the reader.

The organizing categories contain some conflicting guidelines. For example, a long list of caveats may provide the context for appropriate interpretation but conflict with simple appearance and reader involvement. Balancing among the guidelines remains something of an art form. The communication objectives influence the balance.

Communication Objectives

Multivariate graphics can have many different communication objectives. Four common objectives are to provide an overview, to tell a story, to suggest hypotheses, and to criticize a model. In providing an overview, coverage is important. Hiding details is often crucial to achieve clarity in the coverage shown. Similarly, in telling a story the predetermined message must shine through. Tufte [104] is an important resource on the topic of visual explanations. Scientists often fail to tell simple stories because they are reluctant to suppress caveats and a host of details that qualify the basic results. Interactive web graphics [30] can alleviate the archival side of this problem by showing the basic graphics and providing ready access to meta-data, supplemental documents and gigabyte-sized databases. It still takes careful design, however, to lure readers to the details.

This entry leans toward graphics discovery objectives that include suggesting hypotheses and criticizing models. For discovery, it is often crucial to see through the known and miscellaneous sources of variation. In the context of mortality mapping, Tukey [105] said, 'the unadjusted plot should not be made'. Today mortality maps begin to control for known variation by being sex- and race-specific. The maps control for age either by limiting the age range or by statistical adjustments. Typically there are also known risk factors that warrant further adjustments. After such controls and adjustments, inverse variance weighted smoothing can be used in the attempt to bring out the central structure in the remaining noise.

In terms of discovery, balanced visual emphasis of the variables helps the data to speak. Of course a happenstance emphasis of some variables over others occasionally leads to insight, but even then the careful analyst will move toward the symmetrical position of trying all permutations of the variables.

Functions of Two or More Variables

Multivariate visualization involves showing densities, functional relationships, and maps that have more than two coordinates. The density of bivariate points constitutes a third coordinate. The basic idea is that bivariate kernel density estimation is similar to that for univariate estimation: average local likelihoods. The basic difference in the local neighborhood is bivariate. The result is a surface $z = f(x, y)$. Estimating functional relationships of the form $z = f(x, y)$ also follows the pattern established with one less variable. The domain (x, y) is not limited to spatial coordinates, and statistics sometimes consider attributes divorced from spatial indices. Stepping up one dimension higher leads to the study of hypersurfaces. Scott [93] provides graphics showing contours of hypersurfaces. Methods for modeling surfaces are similar whether or not the domain consists of spatial coordinates. However, there are some important issues to address, such as spatial correlation. The interested reader can refer to Cressie [43] or the entries on **spatial covariance** or **kriging**.

Maps

Maps are an important part of environmental visualization. Spatial indices provide a basis for computers and people to organize and access information. Cartographers have developed map projections that are useful for many different purposes [96]. For environmental visualization there are usually good reasons to use equal area projections. Olsen et al. [84] discuss the application of equal area global grids to sampling. The US National Center for Health Statistics uses an Albers equal area projection and the US EPA uses a Lambert equal area projection. NASA (the US National Aeronautics and Space Administration) bases its storage of satellite information and level-three satellite products on latitude and longitude. This has the merit of being familiar but equal angle grids make the North Pole as wide as the equator. This is not directly suitable for polar or global modeling.

Rather, the information must be regridded with the attendant information losses.

Representing attributes on maps is the subject of numerous books; see, for example, Dent [46], Monmonier [81], MacEachren [74] and Slocum [95]. One common representation is the choropleth map (*see* **Landscape pattern metrics**). Regions on a map are colored and the color indicates the region's membership in a class. The classes may be based on a categorical variable or on breaking a continuous variable into class intervals. Discriminating and keeping track of many different colors is not easy, so general guidance is to limit the number of classes to six or fewer.

Authors such as Dent [46] describe various limitations of choropleth maps. The maps are not particularly informative when the variable represented is highly correlated with a region's area. Standard guidance is to represent rates whose denominators adjust for variables related to area. For example, showing pesticide application per unit area or deaths per 100 000 people is reasonable.

Another common difficulty occurs when region boundaries have little to do with the spatial structure of the variable. The spatial variation of the variable within a region may be considerable. The single estimate for a region may be a ratio with large denominator and have a small **standard error**. The calculated uncertainty for the region may provide few clues about the spatial variability. The problem of obtaining different values at different geographical scales is known as the modifiable areal unit problem [54, 117] (*see* **Sample support**).

Another problem with choropleth maps is the difficulty in representing the facets of estimate quality. Typical choropleth maps discard confidence bounds for the estimates and other indicators of estimate quality. Some things can be done, however. MacEachren et al. [76] use light and dark stripes to mark regions whose estimates have low reliability. MacEachren [73] discusses other representations for uncertainty.

Cartograms provide a controversial approach for representing spatially indexed estimates. Cartograms distort the spatial relationships to provide equal representation based on a variable such as human population. Dorling [48] provides numerous examples. The approach is readily applicable to populations of birds, mammals, lakes and so on. The distorted spatial relationships make it difficult for people to associate other information they have stored mentally based on spatial landmarks. Dykes [49] juxtaposes traditional maps with cartograms to ameliorate the problem. Since many people get used to particular map views, cartograms are not likely to see widespread use, but some people find them helpful to see patterns.

Cartographic representations are available for point features, linear features, and surfaces [73, 79]. Rather than represent local values, a common mapping approach estimates surfaces of the form $z = f(x, y)$. The surface can be represented as contours on a map. Given a surface value, z_0, a contour line consists of pairs (x, y) that satisfy the equation $z_0 = f(x, y)$. A typical contour plot shows approximate contour lines for several values of z. Labeled contour lines do not have much visual impact, however. Several methods can improve this. An easy approach is to communicate contour values by contour line thickness. Another option is to fill the regions between contour lines with color. (The colors should be ordered.) Interestingly, there are no confidence bounds for contour lines. Consequently Carr et al. [28] estimate values on a hexagon grid, define class intervals based on the distribution of estimated values, choose colors for the hexagons based on the intervals, and call the result a hexagon mosaic map. If the data and modeling provide justification, then confidence bounds can be calculated for the estimated values. Of course the same can be done for square grids. However, hexagons have merits over squares of the same area [17, 93].

Many surface representations are available such as color-draped perspective wireframes and full rendered color surfaces with highlights. See, for example, Cleveland [35]. Wegman and Luo [113] note that specular reflection highlights local density anomalies. Tufte [104] shows that the pairing of contour and surface plots can aid understanding. The surface tends to provide a good overall impression (except for what is hidden) while the contours help to locate features, such as local extrema, on the plane. Wireframes and translucent surfaces can be superimposed on a map. Sometimes researchers stack three distinct views, a tipped map, contour plot, and surface, so they appear to be aligned in three dimensions. Perspective views appeal to many people and have also been used to show local values, such as the height of three-dimensional bars in animated flyovers. However, perspective foreshortening complicates accurate decoding of values and comparisons.

Researchers have used various extensions. Tufte [103] uses small multiples very effectively to show changes over time. Researchers often use animation to show temporal change. This can reveal rapid change, but it is hard to remember the old views needed to make comparisons over longer intervals. There is little reason to believe that people can visualize the difference between surfaces any better than they can the differences between curves. Consequently, showing the explicit differences is useful in both small multiples and in animation. Carr [20] reports on an early release of satellite data where images of sea surface levels were not properly registered. The previously undetected problem became immediately evident when animating the *difference* between consecutive images.

More complicated extensions include the simultaneous display of two surfaces. This is possible using translucence. Another approach is to have surface height encode one variable and surface color the other.

Of special interest is the fast nonparametric shift histogram technique of Scott and Whittaker [94] for estimating surfaces. This can incorporate sampling weights and handle three or four variables in addition to the spatial coordinates. For instance, it can show a smoothed surface conditioned by other variables.

There are many ways of representing estimates and spatial indices. This entry calls attention to three additional static approaches: plotting glyphs, linking plots and maps, and juxtaposed maps.

Glyphs

Glyphs provide one way to represent multivariate observations. Estimated values control the glyph parameters. We can think of bar charts, pie charts and boxplots as glyphs that represent local distributions. We can also use tiny scatterplots with smooths as glyphs. Thus, glyphs can show local functional relationships. A circle is a simple commonly used glyph. We do not decode circle area very accurately. (Using a legend with a few reference values helps the reader to judge glyph values more accurately.) Carr [21] notes that for multivariate glyphs it is hard to assess the multivariate distance between glyphs so that geometrical pattern-finding breaks down. Carr et al. [29] suggest several variations for using simple ray angle glyphs. For large datasets and maps, Carr [17] and Carr et al. [26, 28] use hexagonal binning to provide symbol congestion control. The ray glyph provides a summary for each hexagonal region and with only one symbol per hexagon, overplotting is not a problem. When rays represent estimated values with confidence intervals, the authors represent the confidence intervals using arcs. Small reference wheels at the base of the ray provide an unobtrusive angular scale for comparison. The angle can be judged accurately in the context of comparison against a scale. Carr et al. [26] develop an angular boxplot glyph. Carr [17] uses a bivariate ray glyph to show two dependent variables. A ray pointing to the right encodes one variable (for small values the ray points down and for large values it points up) and a ray pointing to the left encodes the other. Chambers et al. [32] describe many other such graphical representations including a closely related metroglyph that represents both wind direction and speed. See also [2].

Many glyphs have a long history. Some glyphs, such as Chernoff faces [33, 34], have attracted considerable attention. Takacs [101] provides information about human face recognition that suggests some glyph enhancements. Some, such as the trees of Kleiner and Hartigan [69], have seen little use. Star glyphs, a polar variant of a parallel coordinate plot, occasionally appear. What survives in the long term remains to be seen.

With numerous methods available for producing stereoscopic graphics, stereo rays [22] provide yet another way to show two variables in addition to two spatial coordinates. Few researchers have seriously tackled the visualization of six-dimensional data. A notable exception is Bayly et al. [9]. They successfully used colored stereo ellipsoids to evaluate problems in improving an electrostatic potential model. Their article includes color side-by-side stereo figures. In a long sequence of efforts Bayly (personal communication) failed to obtain insight using a wide variety of encodings. Their eventual success suggests trying a variant of their glyph that devotes two of the coordinates to representing a position on a map. The image of small translucent or wireframe dirigibles comes to mind.

As indicated above, symbol congestion limits the number of glyphs that can be placed on maps. Puckett [91] partly addresses this issue by placing the pie charts (glyphs) around the map and drawing lines from each pie chart to the spatial location. This may suffice for local booking but complicates local spatial comparisons of the distributions. Cuffney et al. [44] provide an interesting glyph composed of six rectangles indicating the status of metal, nonpesticide agricultural intensity (NPAI), fish, invertebrates and algae at monitoring stations. The rectangles of red, yellow, green and white represent the impairment status of severe, moderate, unimpaired, and no data, respectively. The height of six lines could have shown the original continuous values and color could still indicate the class membership. Carr and Olsen [24] show 159 variables using line height. It is possible to represent many variables as a glyph with a sound encoding position along a scale. However, variable labeling and symbol congestion challenges remain.

Linked Plots

Linking points across plots provides a way to connect variables that are represented in different plots. Linking provides a weaker binding of the multivariate observations than glyphs. Linking methods include linking by lines, colors, names, pointers, and spatial linking by juxtaposition. The following discussion emphasizes line linking and color linking.

Diaconis and Friedman [47] discuss M and N plots that link points in different plots with lines. For example, they represent four-dimensional data using two two-dimensional scatterplots. There is nothing that prevents one plot from being a map with spatial coordinates. The first plot represents the two coordinates and the second plot represents the remaining two coordinates. A line between a bivariate point in one plot and a bivariate point in the second plot indicates that bivariate points really represent one four-coordinate point. Their general description includes linking across multiple plots of varying dimensionality. For example a four-dimensional representation might link a one-dimensional plot to a two-dimensional plot to a one-dimensional plot.

Parallel coordinate plots are the only variation of M and N plots that have caught on. The parallel coordinate plot for p dimensions is a sequence of p univariate plots. The representation connects p coordinates with $p-1$ line segments. An early example appears in Bertin [11]. Inselberg [66] and Wegman [110] introduce the mathematical and statistical aspects of parallel coordinate plots. They and Inselberg and Dimsdale [67] describe the point–line duality and other mathematical relationships that provide a basis for extended interpretation. For example, Inselberg has used the representation to find the closest distance between two lines in four dimensions. Interpretation of some patterns requires significant background. Other patterns are easy. For example, Wegman notes that one can readily assess the correlation between adjacent variables. Many crossing segments between adjacent axes indicate a high negative correlation and many parallel segments indicate a high positive correlation. Wegman and Luo [114] also use parallel coordinates for high-dimensional clustering. Even for two variables, parallel coordinates are not very good at communicating a detailed functional relationship. Parallel coordinates seem particularly well suited for showing **time series** or multispectral intensities. In both cases the data units are the same.

Increasingly, exploratory analyses use interactive color brushing [10] to highlight elements represented both in statistical plots and maps. Common plots used in this fashion are scatterplots and parallel coordinate plots. Cook et al. [42] provide a good entry point into the domain of dynamic graphics and GIS. This work builds on the software Xgobi, which supports higher-dimensional exploration using grand tour [3] and projection pursuit [41]. The grand tour or projection pursuit combined with scatterplot or parallel coordinate plots can show evolving linear combinations of variables [111]. The projection pursuit algorithms progressively modify the linear combinations to bring out different features in the plot. Furnas and Buja [58] indicate that we can learn about structure dimensionality through projection and sectioning.

Figure 13 introduces a relatively new template that Carr and Pierson [25] and Carr et al. [27] called linked micromap (LM) plots. This extends the idea of linking statistical plots and maps as promoted by Monmonier [80]. The basic template is composed of three parallel sequences: small generalized maps, region labels, and statistical panels. This design implements the idea of small multiples and parallelism recommended by Tufte. The design partitions the units of study into small perceptual groups and highlights different study units in different panels to encourage selective focus. In Figure 13, adapted from Carr et al. [28], the study units are level 2 Omernik ecoregions. Finding all parts of disjoint ecoregions is easy due to the selective focus and color link. The boxplots summarize the spatial variation in the half million grid cells that partition the US. Each grid cell has estimates produced

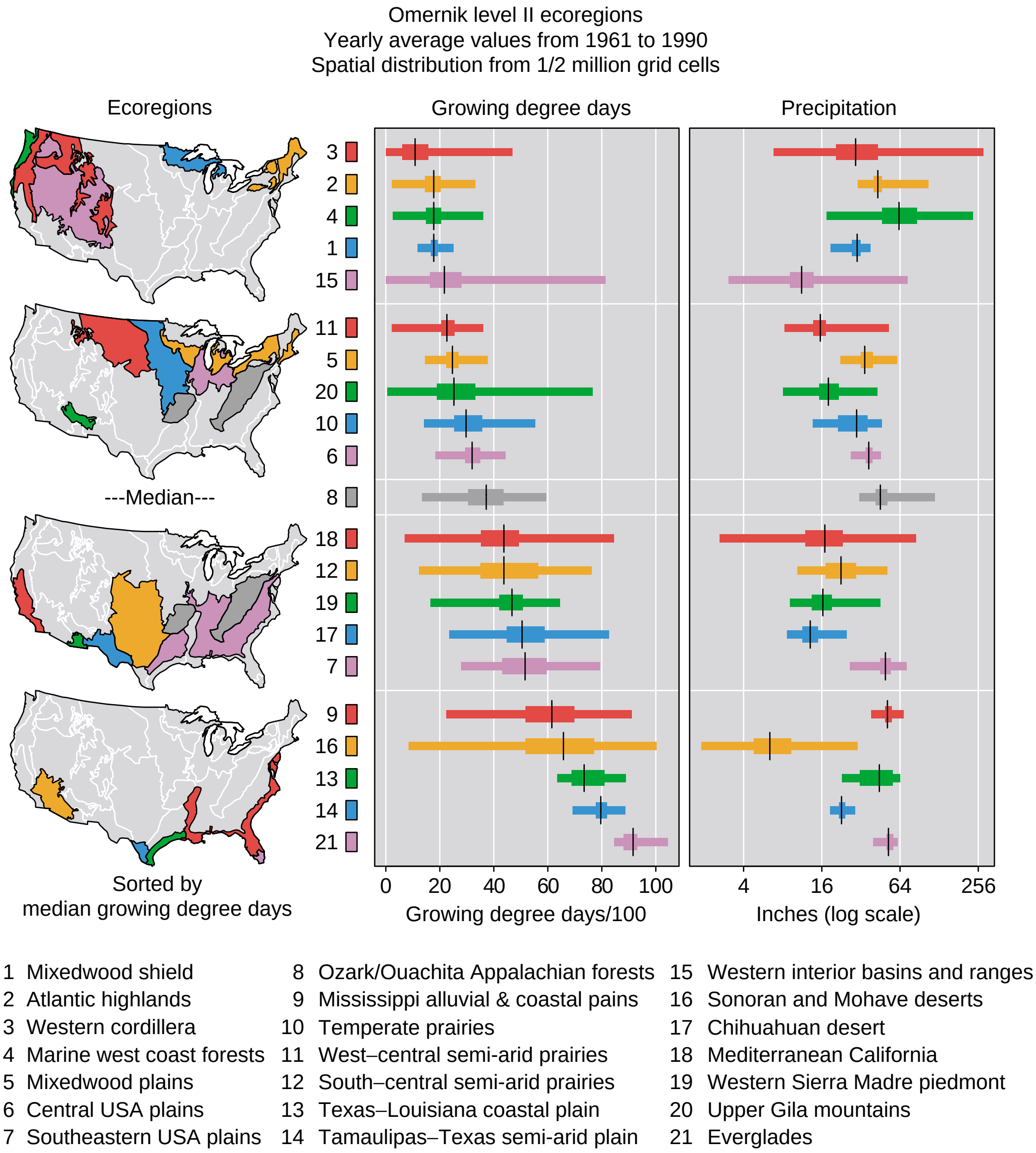

Figure 13 An LM plot with boxplots showing spatial variation. The ecoregions may not be familiar but the horizontal color linking makes it easy to find them, even the ones that are disjoint

by Daly et al. [45] giving the 30-year average precipitation and average number of growing degree days. The distribution of cell values within ecoregions shows considerable variability. If the two variables are closely related to the concept of ecoregion and influence ecoregion definition, then the expectation is that variability will rapidly decrease when considering level-three and higher ecoregions that provide a much finer partition of the US. Note the linking of statistical estimates and spatial coordinates using cyclic colors. There is a learning curve for understanding the color when reading a sequence of micromaps. The highlighted regions in the sequence of panels can reveal more detailed patterns than typical classed choropleth maps, especially in examples with many panels.

There are many variations on LM plots. Carr et al. [27] show the use of dot plots with confidence bounds in the background. Carr et al. [29] use line heights to represent percentages for the 159 land classes defined by Loveland et al. [72]. The boxplots in Carr et al. [31] reveal the variation in mortality rates for local health service areas that contrast with the seemingly stable state estimates. Layouts showing values for counties are now available for several states. Statistical panels can show time series and even bivariate boxplots (*see* **Boxplot, bivariate**). The micromaps can show sites and river segments. Interactive extensions can involve zooming into LM plots to show progressively revealed detail.

Conditioned Plots

The simplest form of conditioned plots partition the estimates (or data) into sets based on classes of conditioning variables. The different sets appear in different juxtaposed panels. The visualization task is then to study how the distributions or function relationships shown in the panels vary across the conditioning panels. This approach is very similar to the nested plots of Tukey and Tukey [107]. An early exposition on conditioned plots (or coplots) appears in Cleveland et al. [40]. Conditioned plots are typically two-dimensional plots, but they can be three-dimensional wireframe plots or other higher-dimensional plots. People readily understand one- and two-way conditioning and the corresponding layout of panels. Thus, conditioned plots prove a reasonable way to study relationships involving three to five dimensions.

Conditioned views do not have to partition the data strictly to produce different panels. Cleveland et al. [40] introduced the notion of shingles that allows the same observations to appear in more than one panel. This is helpful when smoothing a scatterplot because it increases the number of points in the plots and addresses poor smoothing at the panel edges.

Carr et al. [31] developed the coplot idea in the context of maps and call the resulting template conditioned choropleth (CC) maps. Figure 14 provides an example. The data for Figure 14 include many different estimates that describe each equal-area hexagon. These estimates include the number of species found for birds, mammals, insects and reptiles. Additional estimates include the average elevation, the number of different land classes and so on. For Figure 14 the variable chosen for study is the number of bird species. The conditioning variables chosen are number of different land classes and the number of different mammal species. None of the variables is categorical. Consequently, the analyst chooses a transformation that uses class intervals to convert the variables into ordered classes. The bar with five colors at the top shows class boundaries and class colors that partition the hexagons based on the number of birds. The chosen boundaries put 20% of the hexagons in each class. In terms of the number of land cover classes, the four numbers at the bottom of the plot define boundaries that partition the hexagons into three roughly equal classes, each with about one-third of the hexagons. Similarly, the numbers on the left partition the hexagons based on the number of mammal species. The panel is a 3×3 layout based on the classes of the conditioning variable. In a single panel only those hexagons appear that satisfy the row and column conditioning constraints. Each hexagon only appears once in the full plot.

The visual task in Figure 14 is to compare the distributions. The little box in each panel shows the mean for the panel. The box background color classifies the mean. The diagonal color pattern indicates the condition–variable interaction. Figure 15 shows more detailed comparisons using Q–Q plots. These plots compare the distribution of values in each panel with the composite across all panels. This is superior for comparing distributions. However, Figure 14 keeps the spatial context and allows the analyst to generate hypotheses about the spatial patterns observed.

Carr et al. [31] show a layout for an interactive version where the analyst controls the class boundaries with partitioning sliders. This kind of interactivity encourages the reader's involvement, which was one of the four guiding principles. Sophisticated researchers may prefer to move directly to modeling and the study of structure and **residuals**. Exploration using a few variables and a few classes and sliders is limiting. There are recursive partitioning models that have built trees from over two million potential explanatory variables. However, tools like CC maps are easy to understand and can draw people towards more sophisticated modeling.

Laying out multiway panels in rows and columns across many pages is often a good start. However, there are many facets to graphical design and general purpose algorithms have not yet captured all the

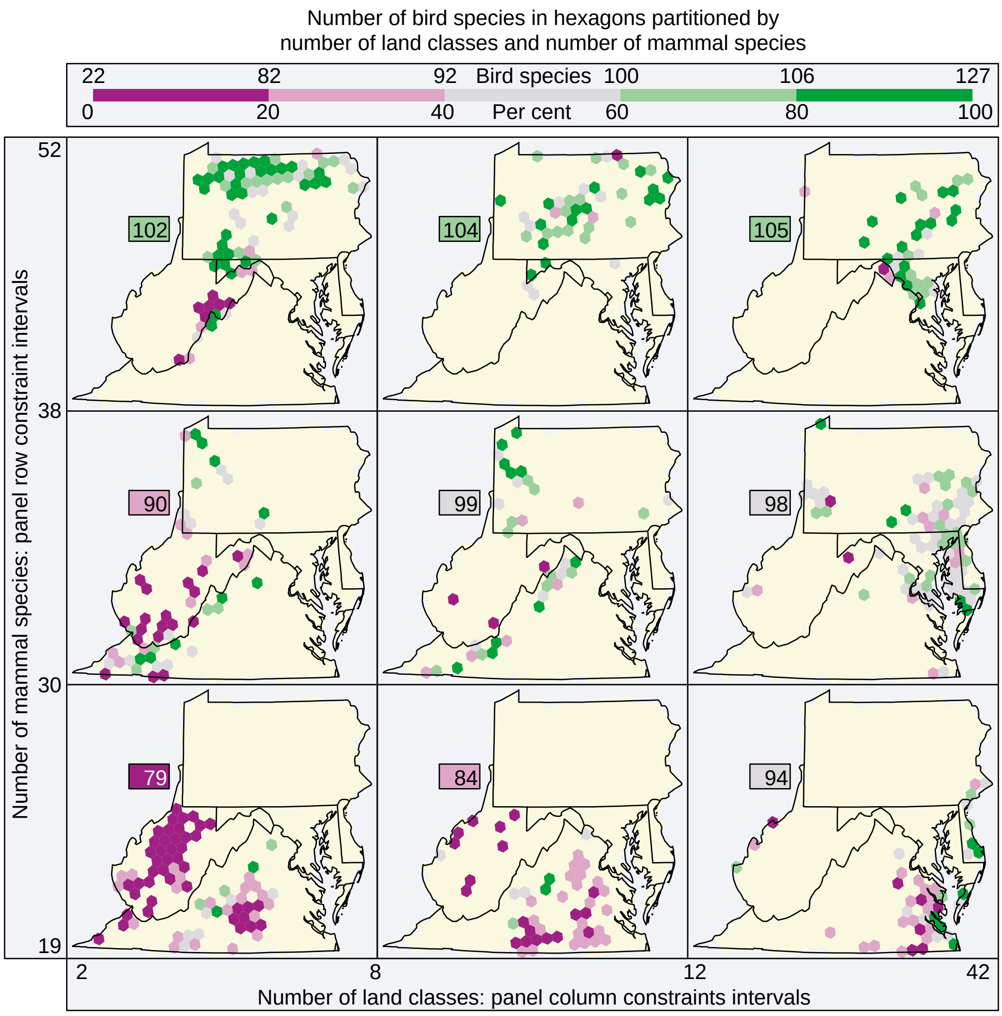

Figure 14 A conditional chloropleth (CC) map. The units are equal area hexagons covering the Mid-Atlantic Region. The plot omits fractional hexagons with less than one-half of the area inside the region. Values in the tabs are the average number of bird species present based on the hexagons in each panel. Conditioning on associated or causal variables reduces variation. Analysts may then see other spatial patterns

current graphical design expertise. For example, Carr and Olsen [24] have found the spanning tree traversal described by Friedman and Rafsky [56] very useful for multivariate sorting. Methods for simplifying visual appearance remain applicable. These include the key strategies of perceptual grouping of information, sorting, presenting the information in layers, the removal of redundant information and the purposeful use of white space [24, 70]. The graphics make it easy to apply thoughtful sorting, but the analyst still has to do the thinking.

The difficulty of seeing patterns across many levels of conditioning factors and pages needs to be recognized. Humans are not good at integrating low-dimensional relationships into higher-dimensional or overview patterns. When the information appears across pages, the limits of our short-term memories compound the difficulty. When across-panel insights occur, they are likely to be based on panels juxtaposed closely in space or time. Careful attention to the choice of layout is often the key to obtaining multivariate insights.

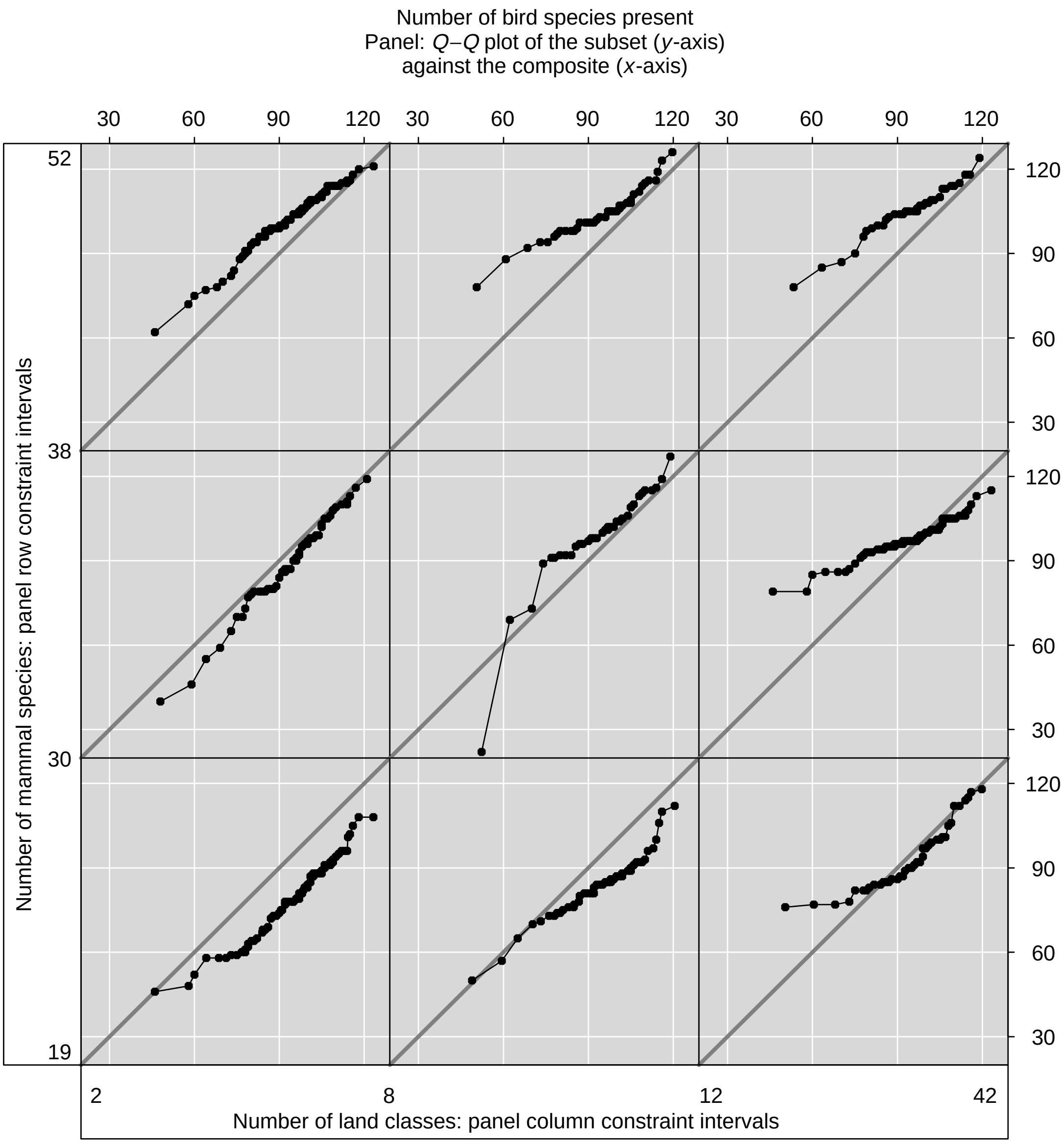

Figure 15 Two-way conditioned *Q–Q* plots. *Q–Q* plots reveal distributional differences of subsets, not just changes in the mean. Comparison with pooled data automatically implies some similarity. However, a single distribution provides a convenient framework for comparison

Closing Remarks

The tools for environment visualization continue to advance. The 1997 special issue of *Computers and Geosciences*, on exploratory cartographic visualization, edited by MacEachren and Kraak [75], contains articles on a variety of topics such as the representation of uncertainty, encoding for characterizing landscapes, and dynamic methods. This entry does not individually list all the instructive articles in that collection and certainly does not begin to capture all the literature that is available. Undoubtedly the presentation here is slanted toward the literature the author knows best. The tools will continue to evolve along with both our understanding of the environment and ourselves. The tools help us to think about the available data.

Sometimes we need to think about the data that we are not seeing. Sometimes there are barriers to scientific inquiry that those who see beyond the data should address.

Increasingly the size, detail, and complexity of environmental datasets will overwhelm our visualization capacity. In thinking about the future, Tukey [106] coined the term cognostics (diagnostics interpreted by a computer rather than a human). The idea was to compute features of merit and have a computer rank plots by their potential interest to humans. This is frightening; for example algorithms can miss little details like the hole in the **ozone** layer. However, humans miss a lot because our looking is not automated and not optimized for understanding.

Some may think environmental visualization is easy. It is not. Rather, it is a huge intellectual challenge that spans developing concepts, collecting data thoughtfully, modeling, grappling with complexity, and dealing with our own limits.

Acknowledgments

Portions of this work were supported by NSF Grants 9819945 and 9983461 and EPA Cooperative Agreement CR825564-01-2. This entry has not be reviewed by the agencies, does not necessarily reflect agency views and no endorsement should be inferred.

References

[1] Ahlberg, C. & Schneiderman, B. (1994). Visual information seeking: tight coupling of dynamic query filters with starfield displays, *Proceedings of the ACM CHI International Conference on Human Factors in Computing (Chi'94)*, Boston, pp. 303–317.

[2] Anderson, E. (1960). A semi-graphical method for the analysis of complex problems, *Technometrics* **2**, 287–292.

[3] Asimov, D. (1985). The grand tour: a tool for viewing multidimensional data, *SIAM Journal of Scientific and Statistical Computing* **6**, 128–143.

[4] Bailey, R.G. (1995). *Description of the Ecoregions of the United States*, USDA Forest Service, Washington, Miscellaneous Publications No. 1391 (review).

[5] Bailey, R.G. (1995). *Ecosystem Geography*, Springer-Verlag, New York.

[6] Bailey, R.G. (1998). *Ecoregions Map of North America: Explanatory Note*, USDA Forest Service, Washington, Miscellaneous Publications No. 1548.

[7] Baird, J.C. & Nome, E. (1978). *Fundamentals of Scaling and Psychophysics*, Wiley, New York.

[8] Barnett, V., ed. (1981). *Interpreting Multivariate Data*, Wiley, New York.

[9] Bayly, C.I., Cieplak, P., Cornell, W.D. & Kollman, P.A. (1993). A well-behaved electostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *Journal of Physical Chemistry* **97**, 10 269–10 280.

[10] Becker, R.A. & Cleveland, W.S. (1987). Brushing scatterplots, *Technometrics* **29**, 127–142.

[11] Bertin, J. (1967, 1983). *Semiologie Graphique*, Gathier-Villars, Paris. *Semiology of Graphics*, Translated by W.J. Berg, The University of Wisconsin Press, Madison.

[12] Brewer, C.A. (1994). Color use guidelines for mapping and visualization, in *Visualization in Modern Cartography*, A.M. MacEachren & D.R.F. Taylor, eds, Pergamon/Elsevier, Oxford, pp. 123–147.

[13] Brewer, C.A. (1997). Spectral schemes: controversial color use on maps, *Cartography and Geographic Information Systems* **24**, 203–220.

[14] Brewer, C.A., MacEachren, A.M., Pickle, L.W. & Herrmann, D. (1997). Mapping mortality, evaluation color schemes for choropleth maps, *The Annals of the Association of American Geographers* **87**, 411–438.

[15] Bruce, A. & Gao, H. (1996). *Applied Wavelet Analysis with S-PLUS*, Springer-Verlag, New York.

[16] Card, K.J., Mackinlay, D. & Schneiderman, B., eds (1999). *Information Visualization Using Vision to Think*, Morgan Kaufmann, San Francisco.

[17] Carr, D.B. (1991). Looking at large data sets using binned data plots, in *Computing and Graphics in Statistics*, A. Buja & P. Tukey, eds, Springer-Verlag, New York, pp. 7–39.

[18] Carr, D.B. (1994). Converting Tables to Plots, Technical Report No. 101, Center for Computational Statistics, George Mason University, Fairfax.

[19] Carr, D.B. (1994). A colorful variation on boxplots, *Statistical Computing and Graphics Newsletter* **5**, 19–23.

[20] Carr, D.B. (1996). Perspectives on the analysis of massive data sets, *Computing Science and Statistics, Proceedings of the Twenty-seventh Symposium on the Interface*, Interface Foundation of North America, Fairfax, pp. 410–419.

[21] Carr, D.B. (1998). Multivariate graphics, in *Encyclopedia of Biostatistics*, Vol. 4, P. Armitage & T. Colton, eds, Wiley, New York, 2864–2886.

[22] Carr, D.B. & Nicholson, W.L. (1988). EXPLOR4: a program for exploring four-dimensional data, in *Dynamic Graphics for Statistics*, W.S. Cleveland & M.E. McGill, eds, Wadsworth, Belmont, pp. 309–329.

[23] Carr, D.B. & Olsen, A.R. (1995). Parallel coordinate plots for representing distribution summaries in map legends, *Proceedings of the Seventeenth International Cartography Association Conference, Tenth General Assembly of the ICA*, pp. 733–742.

[24] Carr, D.B. & Olsen, A.R. (1996). Simplifying visual appearance by sorting: an example using 159 AVHRR

classes, *Statistical Computing and Graphics Newsletter* **7**, 10–16.

[25] Carr, D.B. & Pierson, S.M. (1996). Emphasizing statistical summaries and showing spatial context with micromaps, *Statistical Computing and Graphics Newsletter* **7**, 16–23.

[26] Carr, D.B., Littlefield, R.J., Nicholson, W.L. & Littlefield, J.S. (1987). Scatterplot matrix techniques for large N, *Journal of the American Statistical Association* **82**, 424–436.

[27] Carr, D.B., Olsen, A.R., Courbois, J.P., Pierson, S.M. & Carr, D.A. (1998). Linked micromap plots: named and described, *Statistical Computing and Graphics Newsletter* **9**, 24–32.

[28] Carr, D.B., Olsen, A.R. & White, D. (1992). Hexagon mosaic maps for display of univariate and bivariate geographical data, *Cartography and Geographic Information Systems* **19**, 228–236, 271.

[29] Carr, D.B., Olsen, A.R., Pierson, S.M. & Courbois, J.P. (2000). Using linked micromap plots to characterize Omernik ecoregions, *Data Mining and Knowledge Discovery* **4**, 43–67.

[30] Carr, D.B., Valliant, R. & Rope, D. (1996). Plot interpretation and information webs: a time-series example from the bureau of labor statistics, *Statistical Computing and Graphics Newsletter* **7**, 19–26.

[31] Carr, D.B., Wallin, J.F. & Carr, D.A. (2000). Two new templates for epidemiology applications: linked micromap plots and conditioned choropleth maps, *Statistics In Medicine* **19**, 2621–2538.

[32] Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. (1983). *Graphical Methods for Data Analysis*, Wadsworth and Brooks/Cole, Pacific Grove.

[33] Chernoff, H. (1973). Using faces to represent points in K-dimensional space graphically, *Journal of the American Statistical Association* **68**, 361–368.

[34] Chernoff, H. & Haseeb, R.M. (1975). Effect on classification error or random permutation of features in representing multivariate data by faces, *Journal of the American Statistical Association* **70**, 548–554.

[35] Cleveland, W.S. (1993). *Visualizing Data*, Hobart Press, Summit.

[36] Cleveland, W.S. (1994). *The Elements of Graphing Data*, Hobart Press, Summit.

[37] Cleveland, W.S. ed. (1988). *The Collect Works of John W. Tukey:* Vol. V, *Graphics 1965–1985*, Wadsworth and Brooks Cole, Pacific Grove.

[38] Cleveland, W.S. & McGill, M.E., eds (1988). *Dynamic Graphics for Statistics*, Chapman & Hall, New York.

[39] Cleveland, W.S. & McGill, R. (1984). Graphical perception: theory, experimentation, and application to the development of graphics methods, *Journal of the American Statistical Association* **79**, 531–554.

[40] Cleveland, W.S., Grosse, E. & Shyu, W.M. (1992). Local regression models, in *Statistical Models*, J.M. Chambers & T.J. Hastie, eds, Wadsworth and Brooks Cole, Pacific Grove.

[41] Cook, D., Buja, A., Cabrera, J. & Hurley, C. (1995). Grand tour and projection pursuit, *Journal of Computational and Statistical Graphics* **4**, 155–171.

[42] Cook, D., Symanzik, J., Majure, J. & Cressie, N. (1997). Dynamic graphics in a GIS: more examples using linked software, *Computers and Geosciences* **23**, 371–385.

[43] Cressie, N.A.C. (1993). *Statistics for Spatial Data*, Wiley, New York.

[44] Cuffney, T.F., Meador, M.R., Porter, S.D. & Gurtz, M.E. (1997). Distribution of fish, benthic invertebrate, and algal communities in relation to physical and chemical conditions, Yakima River Basin, Washington, 1990, *US Geological Survey Water Resources Investigation Report 96, 4280*, US Geological Survey, Denver.

[45] Daly, C., Neilson, R.P. & Phillips, D.L. (1994). A statistical–topographic model for mapping climatological precipitation over mountainous terrain, *Journal of Applied Meteorology* **33**, 140–158.

[46] Dent, B.D. (1990). *Cartography – Thematic Map Design*, Brown, Dubuque.

[47] Diaconis, P. & Friedman, J.H. (1980). M and N plots, *Pub-2495*, Stanford Linear Accelerator Center, Stanford University, Stanford.

[48] Dorling, D. (1995). *A New Social Atlas of Britain*, John Wiley, Chichester.

[49] Dykes, J.A. (1997). Exploring spatial data representation with dynamic graphics, *Computers and Geosciences* **23**, 345–371.

[50] Eick, S.G., Steffen, J. & Sumner, E. (1992). Seesoft – a tool for visualization software, *IEEE Transaction on Software Engineering* **18**, 957–968.

[51] Fienberg, S.E. (1979). Graphical methods in statistics, *The American Statistician* **33**, 165–178.

[52] Fisher, L.D. & van Belle, G. (1993). *Biostatistics: A Methodology for The Health Sciences*, Wiley, New York.

[53] Foley, J.D., van Dam, A., Feiner, W.K. & Hughes, J.F. (1990). *Computer Graphics: Principles and Practice*, Addison-Wesley, Reading.

[54] Fotheringham, A.S. & Wong, D.W.S. (1991). The modifiable a real unit problem in multivariate statistical analysis, *Environment and Planning, Series A* **23**, 1025–1044.

[55] Friedhoff, R.M. & Benzon, W. (1991). *The Second Computer Revolution, Visualization,* W.H. Freeman, New York.

[56] Friedman, J.H. & Rafsky, L.C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests, *The Annals of Statistics* **7**, 697–717.

[57] Frigge, M., Hoaglin, D.C. & Iglewicz, B. (1989). Some implementations of the boxplot, *The American Statistician* **43**, 50–54.

[58] Furnas, G.W. & Buja, A. (1994). Prosection views: dimensional inference through sections and projections, *Journal of Computational and Graphical Statistics* **3**, 323–353.

[59] Gnanadesikan, R. (1977). *Methods of Statistical Data Analysis of Multivariate Observations*, Wiley, New York.

[60] Golub, G.H. & von Matt, U. (1997). Generalized cross-validation for large-scale problems, *Journal of Computational and Graphical Statistics* **6**, 1–34.

[61] Goodchild, M. & Gopal, S., eds (1989). *Accuracy of Spatial Databases*, Taylor & Francis, New York.

[62] Grinstein, G. & Levkowitz, H., eds (1995). *Perceptual Issues in Visualization*, Springer-Verlag, New York.

[63] Hammond, A., Adriaanse, A., Rodenburg, E., Byran, D. & Woodward, R. (1995). *Environmental Indicators: A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development*, World Resources Institute, Washington.

[64] Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized Additive Models*, Chapman & Hall, New York.

[65] Herlihy, A.T., Larsen, D.P., Paulsen, S.G., Urquhart, N.S. & Rosenbaum, B.J. (2000). Designing a spatially balanced, randomized site selection process for regional stream surveys: the EMAP Mid-Atlantic pilot study, *Environmental Monitoring and Assessment* **63**, 95–113.

[66] Inselberg, A. (1985). The plane with parallel coordinates, *The Visual Computer* **1**, 69–96.

[67] Inselberg, A. & Dimsdale, B. (1994). Multidimensional lines II: proximity and applications, *SIAM Journal of Applied Mathematics* **54**, 578–596.

[68] Jones, P.G. & Cook, D. (1995). Multivariate Q-Q plots based on quantile contours, *Computing Science and Statistics* **27**, 269–278.

[69] Kleiner, B. & Hartigan, J.A. (1981). Representing points in many dimensions by trees and castles, *Journal of the American Statistical Association* **76**, 260–276.

[70] Kosslyn, S.M. (1994). *Elements of Graph Design*, W.H. Freeman, New York.

[71] Levkowitz, H. (1997). *Color Theory & Modeling for Computer Graphics, Visualization & Multimedia*, Kluwer, Dordrecht.

[72] Loveland, T.R., Merchant, J.W., Reed, B.C., Brown, J.F., Ohlen, D.O., Olson, P. & Hutchinson, J. (1995). Seasonal land cover regions of the United States, *The Annals of the Association of American Geographers* **85**, 339–355.

[73] MacEachren, A.M. (1994). *Some Truth with Maps: A Primer on Symbolization & Design*, Association of American Cartographers, Washington.

[74] MacEachren, A.M. (1995). *How Maps Work*, Guilford Press, New York.

[75] MacEachren, A.M., Kraak, M.J., eds (1997). *Computers & Geosciences* **23**(4).

[76] MacEachren, A.M., Brewer, C.A. & Pickle, L.W. (1998). Visualizing georeferenced data: representing reliability of health statistics, *Environment and Planning, Series A* **30**, 1547–1561.

[77] Marr, D. (1985). Vision: the philosophy and the approach, in *Issues in Cognitive Modeling*, M. Aitkenhead & M.M. Slack, eds, Lawrence Erlbaum, London, pp. 103–126.

[78] McGill, R., Tukey, J.W. & Larsen, W.A. (1978). Variation of boxplots, *The American Statistician* **32**, 12–16.

[79] Mitas, L., Brown, W.M. & Mitasova, H. (1997). Role of dynamic cartography in simulating of landscape processes based on multivariate fields, *Computers & Geosciences* **23**, 437–446.

[80] Monmonier, M. (1988). Geographical representations in statistical graphics: a conceptual framework, *1988 Proceedings of the Section on Statistical Graphics*, American Statistical Association, Alexandria, pp. 1–10.

[81] Monmonier, M. (1993). *Mapping It Out*, University of Chicago Press, Chicago.

[82] Monmonier, M. (1998). *Cartographies of Danger – Mapping Hazards in America*, University of Chicago Press, Chicago.

[83] Olsen, A.R., Sedransk, J., Edwards, D., Gotway, C.A. & Liggett, W., et al. (1999). Statistical issues for monitoring ecological and natural resources in the United States, *Environmental Monitoring and Assessment* **54**, 1–45.

[84] Olsen, A.R., Stevens, D.L. Jr & White, D. (1998). Application of global grids in environmental sampling, *Computing Science and Statistics* **30**, 279–284.

[85] Olson, J.M. & Brewer, C.A. (1997). An evaluation of color selections to accommodate map users with color vision impairments, *The Annals of the Association of American Geographers* **87**, 103–134.

[86] Omernik, J.M. (1987). Ecoregions of the conterminous United States, *The Annals of the Association of American Geographers* **77**, 18–25.

[87] Omernik, J.M. (1995). Ecoregions: a spatial framework for environmental management, in *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*, W.S. Davis & T.P. Simon, eds, Lewis, Boca Raton, pp. 49–62.

[88] Peterson, S.A., Larsen, D.P., Paulsen, S.G. & Urquhart, N.S. (1998). Regional lake trophic patterns in the northeastern United States: three approaches, *Environmental Management* **22**, 789–801.

[89] Pickle, L.W., Mingle, M., Jones, G.K. & White, A.A. (1997). *Atlas of United States Mortality*, National Center for Health Statistics, Hyattsville.

[90] Preston, S.D., Smith, R.A., Schwartz, G.E., Alexander, R.B. & Brakebill, J.W. (1998). Spatially referenced regression modeling of nutrient loading in the Chesapeake Bay watershed, *Proceedings of the First Federal Interagency Hydrologic Modeling Conference; Bridging the Gap Between Technology and Implementation of Surface Water Quality and Quality Models in the Next Century*, Meeting: First Federal Interagency Hydrologic Modeling Conference, Las Vegas, pp. 1.143–1.150.

[91] Puckett, L.J. (1999). Nonpoint and point sources of nitrogen in major watersheds of the United States US Geological Survey Water-Resource, Investigation Report 94-4001, Reston.

[92] Rao, R. & Card, S.K. (1994). The table lens: merging graphical and symbolic representation in an interactive focus + context visualization for tabular information, *Proceedings of the ACM Chi International Conference on Human Factors in Computing*, pp. 318–322.

[93] Scott, D.W. (1992). *Multivariate Density Estimation; Theory, Practice and Visualization*, Wiley, New York.

[94] Scott, D.W. & Whittaker, G. (1996). Multivariate applications of the ash in regression, *Communications in Statistics–Theory and Methods* **25**, 2521–2530.

[95] Slocum, T. (1998). *Thematic Cartography and Visualization*, Prentice-Hall, New York.

[96] Snyder, J.P. (1982). *Map Projections Used by the US Geological Survey*, US Government Printing Office, Washington.

[97] Stevens, D.L. Jr (1997). Variable density grid-based sampling designs for continuous spatial populations, *Environmetrics* **8**, 167–195.

[98] Stevens, D.L. Jr & Olsen, A.R. (1999). Spatially restricted surveys over time for aquatic resources, *Journal of Agricultural, Biological, and Environmental Statistics* **4**, 415–428.

[99] Stone, D., Lijelun, L., Reiersen, L., Nilson, A., et al. (1997). Arctic pollution issues: a state of the Arctic environmental report, *AMAP*, Box 8100 Dep. N-0032 Oslo, Norway.

[100] Symanzik, J., Carr, D.B., Axelrad, D.A., Wang, J., Wong, D. & Woodruff, T.J. (1999). Interactive tables and maps – a glance at EPA's cumulative exposure project web page, *1999 Proceedings of the Section on Statistical Graphics*, American Statistical Association, Alexandria.

[101] Takacs, B. (1996). Perception and recognition of human faces, Ph.D. Thesis, George Mason University, Fairfax.

[102] Tufte, E.R. (1983). *The Visual Display of Quantitative Information*, Graphics Press, Cheshire.

[103] Tufte, E.R. (1990). *Envisioning Information*, Graphics Press, Cheshire.

[104] Tufte, E.R. (1997). *Visual Explanations*, Graphics Press, Cheshire.

[105] Tukey, J.W. (1979). Statistical mapping: what should not be plotted, *Proceedings of the 1976 Workshop on Automated Cartography*, Department of Health, Education and Welfare Publication No. (PHS) 79-1254, pp. 18–26.

[106] Tukey, J.W. (1983). Another look at the future, *Computer Science and Statistics, Proceedings of the Fourteenth Symposium on the Interface*, Vol. 14, K.W. Heiner, R.S. Sacher & J.W. Wilkinson, eds, Springer-Verlag, New York, pp. 2–8.

[107] Tukey, J.W. & Tukey, P.A. (1983). Some graphics for studying four-dimensional data, *Computer Science and Statistics, Proceedings of the Fourteenth Symposium on the Interface*, Vol. 14, K.W. Heiner, R.S. Sacher & J.W. Wilkinson, eds, Springer-Verlag, New York, pp. 60–66.

[108] Vogelmann, J.E. & Wickham, J.D. (2000). *Implementation Strategy for Production of National Land-Cover Data (NLCD) from the Landsat 7 Thematic Mapper Satellite*, EPA 600/R-00/051 Office of Research and Development, US Environmental Protection Agency, Washington.

[109] Wahlström, E., Hallanaro, E. & Manninen, S., eds (1996). *The Future of the Finnish Environment*, Edita, Helsinki.

[110] Wegman, E.J. (1990). Hyperdimensional analysis using parallel coordinates, *Journal of the American Statistical Association* **85**, 664–675.

[111] Wegman, E.J. (1991). The grand tour in K dimensions, *Computer Science and Statistics, Proceedings of the Twenty-second Symposium on the Interface*, pp. 127–136.

[112] Wegman, E.J. & Carr, D.B. (1993). Statistical graphics and visualization, in *Handbook of Statistics, Computational Statistics*, Vol. 9, C.R. Rao, ed., North-Holland, New York, pp. 857–958.

[113] Wegman, E.J. & Luo, Q. (1994). *Visualizing Densities*, Technical Report No. 100, Center for Computational Statistics, George Mason University, Fairfax.

[114] Wegman, E.J. & Luo, Q. (1997). High-dimensional clustering using parallel coordinates and the grand tour, *Computing Science and Statistics* **28**, 361–368.

[115] Wiken, E.B. (1986). Terrestrial ecozones of Canada, *Environment Canada, Ottawa, Ontario, Canada, Ecological Land Classification Series No. 19*.

[116] Wilkinson, L. (1999). *The Grammar of Graphics*, Springer-Verlag, New York.

[117] Wong, D.W.S. & Amrhein, C.G. (1996). Research on the MAUP: old wine in a new bottle or real breakthrough?, *Geographical Systems* **3**, pp. 73–76.

[118] Wood, D.W. (1992). *The Power of Maps*, Guilford Press, New York.

(*See also* **Influence diagrams**; **Nelder plots**; **Risk perception**; **Statistical computing in environmental science**)

DAN CARR