Hyperdimensional Data Analysis Using Parallel Coordinates

EDWARD J. WEGMAN*

This article presents the basic results of using the parallel coordinate representation as a high-dimensional data analysis tool. Several alternatives are reviewed. The basic algorithm for parallel coordinates is laid out and a discussion of its properties as a projective transformation is given. Several duality results are discussed along with their interpretations as data analysis tools. Permutations of the parallel coordinate axes are discussed, and some examples are given. Some extensions of the parallel coordinate idea are given. The article closes with a discussion of implementation and some of my experiences.

The classic scatter diagram is a fundamental tool in the construction of a model for data. It allows the eye to detect such structures in data as linear or nonlinear features, clustering, outliers and the like. Unfortunately, scatter diagrams do not generalize readily beyond three dimensions. For this reason, the problem of visually representing multivariate data is a difficult, largely unsolved one. The principal difficulty, of course, is the fact that whereas a data vector may be arbitrarily high dimensional, say n, Cartesian scatterplots may only be done easily in two dimensions and, with computer graphics and more effort, three dimensions.

I propose as a multivariate data analysis tool the following representation. Instead of using a scheme that tries to preserve the orthogonality of the n-dimensional coordinate axes, draw them as parallel. A vector $(x_1, x_2, \ldots, x_n)$ is created by plotting x_1 on axis 1, x_2 on axis 2, and so on through x_n on axis n. These points are joined by a broken line. Figure 1 illustrates two points (one solid, one dashed), plotted in parallel coordinate representation, that agree in the fourth coordinate. The principal advantage of this plotting device is clear. Each vector $(x_1, x_2, \ldots, x_n)$ is represented in a planar diagram, so each vector component has essentially the same representation.

My parallel coordinates proposal has its roots in a number of sources. Griffen (1958) considered a two-dimensional parallel coordinate device as a method for graphically computing Kendall's tau correlation coefficient. Hartigan (1975) described his "profiles algorithm" as "histograms on each variable connected between variables by identifying cases" (p. 29). Although he did not recommend drawing all profiles, a profile diagram with all profiles plotted is, in fact, a parallel coordinate plot. There is, however, far more mathematical structure, particularly high-dimensional structure, to the parallel coordinate diagram than Hartigan exploited. Inselberg (1985) originated the parallel coordinate representation as a device for computational geometry. His 1985 paper is the culmination of a series of technical reports dating from 1981. Finally, Diaconis and Friedman (1983) discussed the so-called M and N plots. Their special case of a 1-and-1 plot is a parallel coordinate plot in two dimensions. Indeed, the 1-and-1 plot is sometimes called a before-and-after plot and has a much older history. Also related is the Andrews (1972) plot, which can be viewed as a Fourier-series interpolation of the points on the parallel coordinate axes. This interpolation preserves some least squares properties because of Parseval's relationship but does not enjoy the statistical interpretations available for parallel coordinates because of projective geometry dualities.

The fundamental theme of this article is the highly structured mathematical nature of the transformation from Cartesian coordinates to parallel coordinates, which maps mathematical objects into mathematical objects. Thus although the use of parallel coordinates as a data analysis tool may, at first glance, appear unlikely to be successful, because it is highly structured interpretations may be given and intuition developed. Of course, Cartesian coordinate representations have a long history; consequently, there has been much development of intuition about the appearance of structures represented by them. Similar intuition for parallel coordinate representations must, therefore, be developed. One should suspend reservations about parallel coordinates due to the lack of developed intuition in anticipation that this article and future work will develop the needed intuition.

Section 1 is a discussion of other multivariate data representations. In Section 2, I discuss some of the basic facts about parallel coordinate geometry. Parallel coordinates are closely linked to ideas in projective geometry, so these connections are delineated in Section 3. Some statistical interpretations are offered in Section 4 and an illustrative example is given in Section 5. Section 6 focuses on some variants of parallel coordinates. Appendix A provides a rather nice result on minimal permutations of parallel coordinate axes and Appendix B a discussion of the connection between parallel coordinates and star diagrams.

1. OTHER MULTIVARIATE DATA REPRESENTATIONS

Alternative static multidimensional representations have been proposed by several authors, including Chernoff's (1973) faces, Fienberg's (1979) star diagrams, and Cleveland and McGill's (1984a,b), Carr, Nicholson, Littlefield, and Hall's (1986), and others' scatterplot matrices. The Chernoff faces and star diagrams are what might be

* Edward J. Wegman is Director and Professor for the Center for Computational Statistics, George Mason University, Fairfax, VA 22030. This research was supported by Air Force Office of Scientific Research Grant AFOSR-87-0179, Army Research Office Contract DAAL03-87-K-0087, and National Science Foundation Grant DMS-8701931.

Journal of the American Statistical Association
September 1990, Vol. 85, No. 411, Theory and Methods

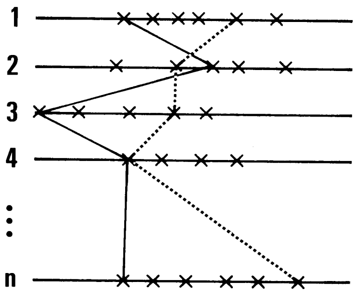

Figure 1. Parallel Coordinate Representation of Two n-Dimensional Points.

thought of generically as icon-based representations. Scatterplot matrices are representations of projections of data into grouped two-dimensional plots. Much of Carr et al.'s work focused on a mixture of projection- and icon-based representations. An important alternative technique based on the use of motion is the computer-based kinematic display, yielding the illusion of three-dimensional scatter diagrams. This technique was pioneered by Friedman and Tukey (1973) under the name PRIM-9 and is now available in a variety of commercial software packages. Coupled with easy data manipulation, the kinematic display techniques spawned the exploitation of such methods as projection pursuit (Friedman and Tukey 1974) and the grand tour (Asimov 1985). Clearly, projection-based techniques are highly successful and lead to important insights concerning data. Nonetheless, one must be cautious in making inferences about high-dimensional data structures based on projection methods alone.

I inject this cautionary note because for hyperdimensional geometry, normal two- and three-dimensional geometric intuition fails very rapidly with increase in dimension. An easy illustration involves the general intersection of two two-planes in Euclidian four-space. The normal three-dimensional intuition is that the general intersection of two planes is a line. One can construct four-space analytically, however, as the Cartesian product of two orthogonal two-planes. It is immediate from that observation that the general intersection of two two-planes in four-space is a point. The following two examples illustrate the potential pitfalls of projection-based methods. See Kendall (1961) for a general treatment of n-dimensional geometry.

Example 1.1 Diagonals in Hyperspace

Consider the general ray diagonal in n-dimensional space, that is, the vector passing through $(d_1, d_2, \ldots, d_n)$ and originating at $(0, 0, \ldots, 0)$. Here $d_j = \pm 1$. Choose any diagonal and fix it. For purposes of discussion, consider the ray diagonal through $(1, 1, \ldots, 1)$. It is easy to see for $n = 2$ that the angle θ_2 between the ray diagonal and any of the coordinate axes is characterized by $\cos \theta_2 = 1/\sqrt{2}$. A simple inductive argument leads to $\cos \theta_n = 1/\sqrt{n}$ in the general case. For $n \to \infty$, $\cos \theta_n \to 0$ so that $\theta_n \to \pi/2$. Thus the ray diagonals are nearly orthogonal to the coordinate axes for reasonably large n. A simple computation shows that there are 2^n ray diagonals or 2^{n-1} diagonals in n-dimensional space. From a data analytic perspective, this computation implies that data structures lying near a diagonal in hyperspace will be mapped nearly into the origin in every lower-dimensional projection onto the original coordinates, but a similar data structure located near a coordinate axis will not be. Thus the intuition gained from a scatter diagram in two or three dimensions is highly dependent on the coordinate axis system chosen. It is relatively easy to miss real data structures by simply examining lower-dimensional projections.

Example 1.2 Hypervolume of a Thin Shell

The area of a circle of radius r is c_2r^2, the volume of a sphere of radius r is c_3r^3, and in general, the hypervolume (content) of a hypersphere of radius r is c_nr^n. In general, then, the hypervolume of a thin shell is $c_n[r^n - (r - \varepsilon)^n]$, and the hypervolume of the thin shell relative to the hypervolume of the n sphere is $1 - (1 - \varepsilon/r)^n$. Since $1 - \varepsilon/r < 1$, the relative hypervolume converges to 1 as $n \to \infty$. Loosely speaking, most of the hypervolume is close to the $(n - 1)$-dimensional hypersurface of the n sphere.

Consider then a probability measure that is uniform in the volume of an n-dimensional hypersphere. Consider further a random sample of observations drawn at random according to this measure. If n-dimensional scatter diagrams could be visualized, most of the observations would lie close to the $(n - 1)$ hypersurface of the n sphere. If on the other hand, we project observations onto a two-plane, we get a scatter diagram with a circular cross-section. The most intense concentration of observations would be near the center; that is, the distribution would appear unimodal. Indeed, it is easy to show that the marginal density would be unimodal. [In the case of two dimensions, the density on the one-plane would be $f(x) = \pi^{-1}(1 - x^2)^{1/2}$.] Thus, curiously, in the n-dimensional scatter diagram, most of the observations would lie near the boundaries of the n sphere, but in the two-dimensional projection, most of the observations would lie near the center. I contend that the two-dimensional projection may convey the wrong intuition.

These examples show that exploratory data analysis based on projection techniques including two- and three-dimensional scatter diagrams is potentially misleading. It would be highly desirable to have a simultaneous representation of all coordinates of a data vector, especially if the representation treated all components in a similar manner. The standard Cartesian coordinate representation fails because of the requirement for orthogonal coordinate axes. In a three-dimensional world, it is difficult to represent more than three orthogonal coordinate axes. Thus we are motivated to give up the orthogonality requirement and replace the standard Cartesian axes with a set of n parallel axes.

2. PARALLEL COORDINATE GEOMETRY

The parallel coordinate representation enjoys some elegant duality properties with the usual Cartesian orthogonal coordinate representation. Consider a line $\mathcal{L}$ in the Cartesian coordinate plane given by $\mathcal{L}$: $y = mx + b$, and

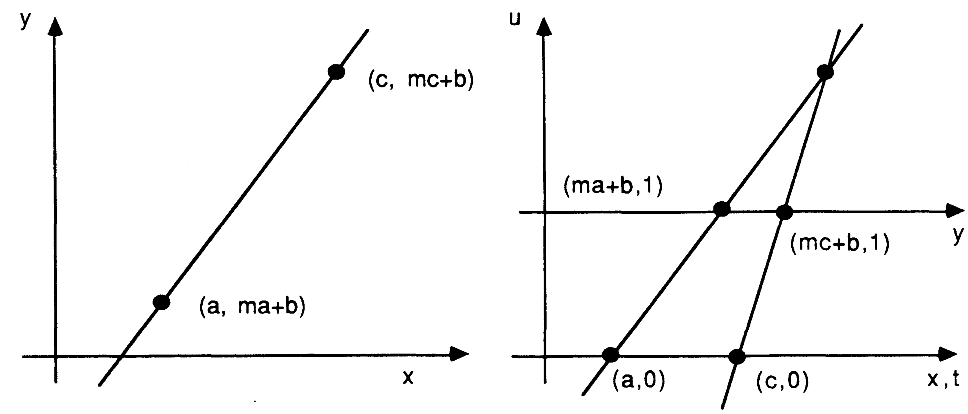

Figure 2. Cartesian and Parallel Coordinate Plots of Two Points. The tu Cartesian coordinate system is superimposed on the xy parallel coordinate system.

consider two points lying on that line, say $(a, ma + b)$ and $(c, mc + b)$. For simplicity of computation, consider the xy Cartesian axes mapped into the xy parallel axes as described in Figure 2. Superimpose Cartesian coordinate axes tu on the xy parallel axes so that the y parallel axis has the equation $u = 1$. The point $(a, ma + b)$ in the xy Cartesian system maps into the line joining $(a, 0)$ to $(ma + b, 1)$ in the tu coordinate axes. Similarly, $(c, mc + b)$ maps into the line joining $(c, 0)$ to $(mc + b, 1)$. It is a straightforward computation to show that these two lines intersect at a point (in the tu plane) given by $\bar{\mathcal{L}}$: $[b(1 - m)^{-1}, (1 - m)^{-1}]$. Notice that this point in the parallel coordinate plot depends only on m and b, the parameters of the original line in the Cartesian plot. Thus $\bar{\mathcal{L}}$ is the dual of $\mathcal{L}$, giving the interesting duality that points in Cartesian coordinates map into lines in parallel coordinates and lines in Cartesian coordinates map into points in parallel coordinates.

For $0 < (1 - m)^{-1} < 1$, m is negative and the intersection occurs between the parallel coordinate axes. For $m = -1$, the intersection is exactly midway. A ready statistical interpretation can be given. For highly negatively correlated pairs, the dual line segments in parallel coordinates tend to cross near a single point between the two parallel coordinate axes. The scale of one of the variables may be transformed in such a way that the intersection occurs midway between the two parallel coordinate axes, in which case the slope of the linear relationship is -1.

In the case of $(1 - m)^{-1} < 0$ or $(1 - m)^{-1} > 1$, m is positive and the intersection occurs external to the region between the two parallel axes. In the special case of $m = 1$, this formulation breaks down. It is clear, however, that the point pairs are $(a, a + b)$ and $(c, c + b)$. The dual lines to these points are the lines in parallel coordinate space with slope b^{-1} and intercepts $-ab^{-1}$ and $-cb^{-1}$, respectively. Thus the duals of these lines in parallel coordinate space are parallel lines with slope b^{-1}. We thus append the ideal points to the parallel coordinate plane to obtain a projective plane. The ideal points may be thought of as extra points added to the ordinary plane and, intuitively, as the points where parallel lines intersect. There are thus as many ideal points as there are slopes. Consequently, these parallel lines intersect at the ideal point in direction b^{-1}. One model for the projective plane is a hemisphere with diametrically opposed equatorial points identified.

In the statistical setting, the following interpretation can be made. For highly positively correlated data, lines tend not to intersect between the parallel coordinate axes. By suitable linear rescaling of one of the variables, the lines may be made approximately parallel in direction with slope b^{-1}. In this case the slope of the linear relationship between the rescaled variables is 1. See Figure 3 for an illustration of a sequence of correlations ranging from large positive to large negative.

3. NATURAL HOMOGENEOUS COORDINATES AND CONICS

The point–line, line–point duality seen in the transformation from Cartesian to parallel coordinates extends to conic sections. A more complete discussion of projective

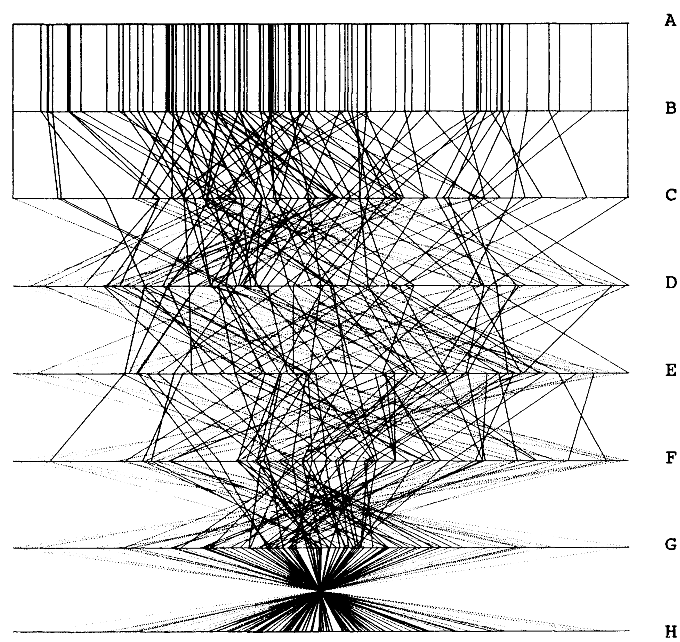

Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8,$ and -1.

transformations of conics was given by Dimsdale (1984). Inselberg (1985) generalized the notion of conics in parallel coordinates to what he calls *h* stars. A general reference for projective geometry and natural homogeneous coordinates can be found in Fishback (1962). For the discussion of conics, however, consider both the xy plane and the tu plane to be augmented by suitable ideal points so that both may be regarded as projective planes. The representation of points in parallel coordinates is thus a transformation from one projective plane to another. Computation is simplified by an analytic representation. The usual coordinate pair, (x, y), is not, however, sufficient to represent ideal points. I plan to represent points in the projective plane by triples, (x, y, z). Consider two distinct parallel lines having the equations $ax + by + cz = 0$ and $ax + by + c'z = 0$. Simultaneous solution yields $(c - c')z = 0$, so $z = 0$ describes ideal points. The representation of points in the projective plane is by triples, (x, y, z), which are called natural homogeneous coordinates. If $z = 1$, the resulting equation is $ax + by + c = 0$, and so $(x, y, 1)$ is the natural homogeneous-coordinate representation of a point (x, y) in Cartesian coordinates lying on $ax + by + c = 0$. Notice that if (px, py, p) is any multiple of $(x, y, 1)$ on $ax + by + c = 0$, then

$$apx + bpy + cp = p(ax + by + c) = p \cdot 0 = 0.$$

Thus the triple (px, py, p) represents equally well the Cartesian point (x, y) lying on $ax + by + c = 0$, so the representation in natural homogeneous coordinates is not unique. If p is not 1 or 0, however, we can simply rescale the natural homogeneous triple to have a 1 for the z component and thus read off the Cartesian coordinates directly. If the z component is 0, we know immediately that we have an ideal point.

Notice that we could equally well consider the triples (a, b, c) as natural homogeneous coordinates of a line. Thus triples can represent either points or lines reiterating the fundamental duality between points and lines in the projective plane. Recall now that the line $\mathcal{L}$: $y = mx + b$ is mapped into the point $\bar{\mathcal{L}}$: $[b(1 - m)^{-1}, (1 - m)^{-1}]$ in parallel coordinates. In natural homogeneous coordinates, $\mathcal{L}$ is represented by the triple $(m, -1, b)$ and the point $\bar{\mathcal{L}}$ by the triple $[b(1 - m)^{-1}, (1 - m)^{-1}, 1]$ or, equivalently, by $(b, 1, 1 - m)$. The latter yields the appropriate ideal point when $m = 1$. A straightforward computation shows that for

$$A = \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & -1 \\ 1 & 0 & 0 \end{bmatrix},$$

$t = xA$ or $(b, 1, 1 - m) = (m, -1, b)A$. Thus the transformation from lines in orthogonal coordinates to points in parallel coordinates is a particularly simple projective transformation with the rather nice computational property of having only addition and subtraction.

Similarly, a point $(x_1, x_2, 1)$ expressed in natural homogeneous coordinates maps into the line represented by $(1, x_1 - x_2, -x_1)$ in natural homogeneous coordinates. Another straightforward computation shows that the linear transformation given by $t = xB$ or $(1, x_1 - x_2, -x_1) = (x_1, x_2, 1)B$, where

$$B = \begin{bmatrix} 0 & 1 & -1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

describes the projective transformation of points in Cartesian coordinates to lines in parallel coordinates. Because these are nonsingular linear transformations, hence projective transformations, it follows from the elementary theory of projective geometry that conics are mapped into conics. This is straightforward to see because an elementary quadratic form in the original space—say, xCx', where x' denotes x transpose—represents the general conic. Clearly then, since $t = xB$, B nonsingular, we have $x = tB^{-1}$; so $tB^{-1}C(B^{-1})'t'$ is a quadratic form in the image space. An instructive computation involves computing the image of an ellipse $ax^2 + by^2 - cz^2 = 0$ with $a, b, c > 0$. The image in the parallel coordinate space is $ct^2 - b(u + v)^2 = av^2$, a general hyperbolic form.

The quadratic form does not describe a locus of points, but rather the natural homogeneous coordinates of a locus of lines, a line conic. The notion of a line conic is, perhaps, a strange one. By this I mean a locus of lines the natural homogeneous coordinates of which satisfy the equation for a conic. These may be more easily related to the usual notion of a conic by realizing that the envelope of this line conic is a point conic. In this computation, the point conic in the original Cartesian coordinate plane is an ellipse and the image in the parallel coordinate plane is, as we have just seen, a line hyperbola with a point hyperbola as envelope, as illustrated in Figure 4 with a parallel coordinate plot of a five-dimensional hypersphere. As it turns out, this has an important statistical interpretation.

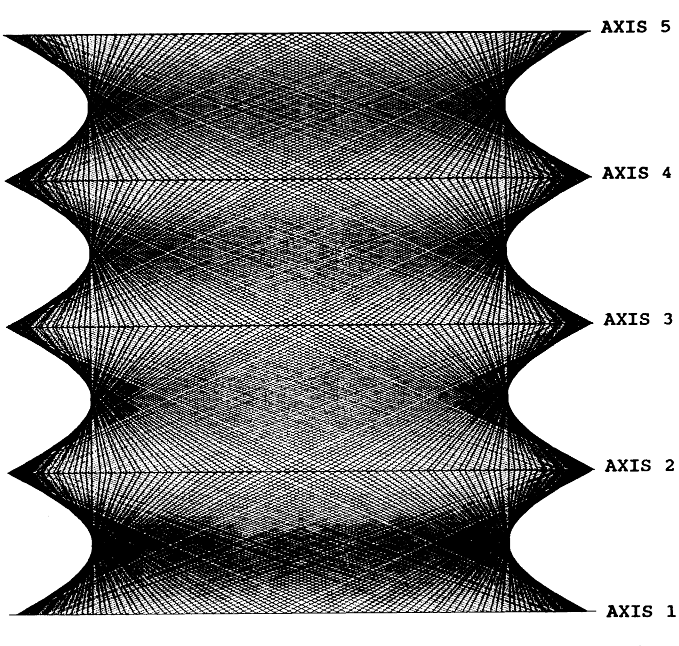

Figure 4. Parallel Coordinate Plot of a Five-Dimensional Hypersphere. Every axis pair is correspondingly the parallel coordinate plot of a circle. Notice the hyperbolic envelopes.

I mentioned the duality between points and lines and conics and conics. There are two other nice dualities. Rotations in Cartesian coordinates become translations in parallel coordinates and vice versa. Perhaps more interesting from a statistical point of view, points of inflection in Cartesian space become cusps in parallel coordinate space and vice versa. Thus the relatively hard-to-detect inflection point property of a function becomes the notably more easy-to-detect cusp in the parallel coordinate representation. Inselberg (1985) discussed these properties in detail. It is well worth noting that the natural homogeneous coordinate representation is a standard device in computer graphics because a nonlinear transformation (rotation in two-space) can be represented as a 3×3 linear transformation of natural homogeneous coordinates. See, for example, Plastock and Kalley (1986).

4. STATISTICAL INTERPRETATIONS

Since ellipses map into hyperbolas, they provide an easy template for diagnosing uncorrelated data pairs. With a completely uncorrelated data set, we would expect the two-dimensional scatter diagram to fill substantially a circumscribing circle. As illustrated in Figure 4, the parallel coordinate plot would approximate a figure with a hyperbolic envelope. As the correlation approaches -1, the hyperbolic envelope would deepen, so in the limit, we would have a pencil of lines, what I call the crossover effect. As the correlation approaches $+1$, the hyperbolic envelope would widen, with fewer and fewer crossovers, so that in the limit we would have parallel lines. Thus correlation structure can easily be diagnosed from the parallel coordinate plot. As noted earlier, Griffen (1958) used this as a graphical device for computing Kendall's tau.

Griffen, in fact, attributed the graphical device to Holmes (1928), which predates Kendall's discussion. Griffen demonstrated that the computational formula for computing Kendall's tau by means of Holmes's graphical method is $r = 1 - [4X/n(n - 1)]$, where X is the number of intersections resulting from connecting the two rankings of each member by lines, one ranking having been put in natural order. Although the original formulation was framed in terms of ranks for the x and y axes, it is clear that the number of crossings is invariant to any monotone increasing transformation of either x or y, the ranks being one such transformation. Because of this scale invariance, one would expect rank-based statistics to have an intimate relationship to parallel coordinates.

It is clear that if there is a perfect positive linear relationship with no crossings, then $X = 0$ and $r = 1$. Similarly, if there is a perfect negative linear relationship, the bottom pair of coordinates in Figure 3 is appropriate and we have a pencil of lines. (A pencil of lines is a set of lines that are coincident at a single point.) Since every line meets every other line, the number of intersections is $\binom{n}{2}$, so

$$r = 1 - [4\tbinom{n}{2}/n(n - 1)] = -1.$$

Linear relationships are comparatively easy to diagnose by using parallel coordinates, particularly negative linear relationships, since the eye seems to note the crossover effect quickly. Moreover, linear relationships exhibited by several sets of adjacent pairs of parallel coordinate axes may be interpreted as several sets of collinearities. Two sets of collinearities, in turn, may be interpreted as points lying in a two-dimensional plane, with d sets of collinearities being interpreted as points lying in a d-dimensional hyperplane. Thus detecting linear structure is important in understanding data structure, particularly if one is interested in fitting multiple linear regression models. A linear rescaling of one or more of the axes is sometimes helpful because it guides the eye in looking for approximately parallel line segments. Of course, nonlinear relationships will not respond to simple linear rescaling. By suitable nonlinear transformations, it should be possible, however, to transform to linearity. Knowing the nonlinear transformation that yields linearity in the data gives a fundamental model building tool.

Clustering is easily diagnosed by using the parallel coordinate representation. See, for example, Figure 5 (a) and (b), illustrating separation in both x and y and in only the first coordinate. Indeed, the individual parallel coordinate axes represent one-dimensional projections of the data. Thus separation on any one axis represents a view of the data that allows the detection of clustering. Because of the connectedness of the multidimensional parallel coordinate diagram, it is usually easy to see whether this clustering propagates through other dimensions. My experience indicates that clustering may occur not in any single dimension but in combinations. Figure 5(c) indicates the appearance of three clusters in both Cartesian plots and parallel coordinate plots. In neither projection do these clusters separate, but they do separate in their joint relationship. A more extensive example of this is illustrated by the five-dimensional automobile example in Figure 10.

So far I have developed intuition for pairwise parallel coordinate relationships. The idea, however, is to stack these diagrams and represent all n dimensions simultaneously. Recall that Figure 4 is a parallel coordinate plot of a five-dimensional hypersphere. A five-dimensional ellipsoid would have a similar general shape, but with hyperbolas of different depths. Figure 6 illustrates some data structures one might see in a five-dimensional data set. First, the plots along any given axis represent dot diagrams, hence convey graphically the one-dimensional marginal distributions. In this illustration, the first axis is meant to have an approximately normal distribution shape and axis two has the shape of the negative of a chi-square. Figure 6 illustrates a number of instances of linear (both negative and positive), nonlinear, and clustering situations. Indeed, it is clear that there is a three-dimensional cluster along coordinates 3, 4, and 5. Note also that the left cluster in Figure 6 consists of a series of parallel line segments between axes 4 and 5 and between axes 3 and 4, indicating that these points are collinear in the 4–5 and 3–4 planes. Since two general lines determine a plane, these points are coplanar in the 3–4–5 three-dimensional space. Moreover these same line segments exhibit the crossover effect between the 2 and 3 axes, and they are also collinear there. Thus because three general lines de-

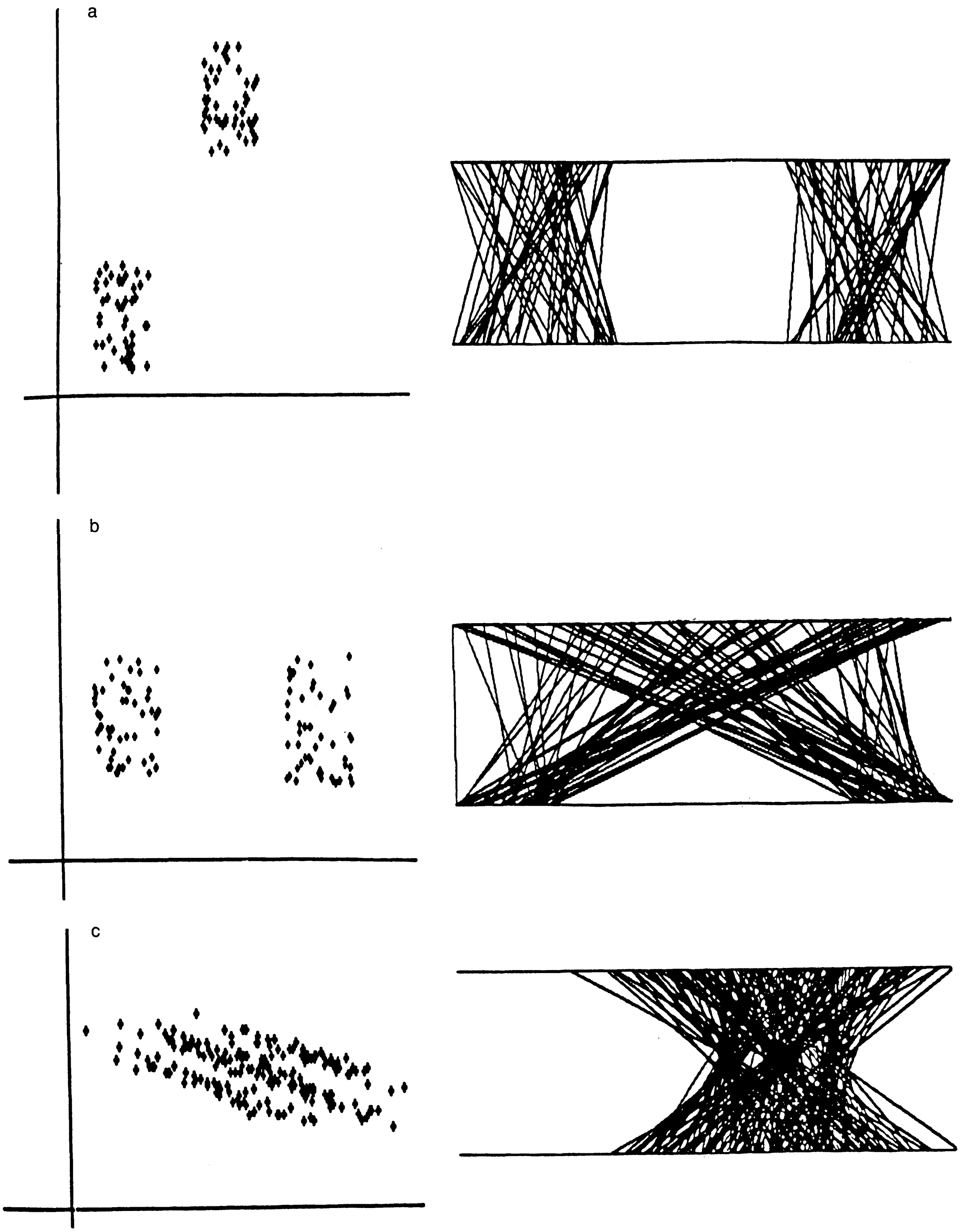

Figure 5. (a) Clustering That Is Separated in Both x and y, Represented in Parallel Coordinates. (b) Clustering That Is Separated in the x Coordinate but not the y Coordinate. (c) Clustering That Is Separated in Neither Projection. Since the first coordinate of the crossover point for negatively correlated variables is dependent on the intercept, we can separate overlapping clusters.

termine a three-dimensional hyperplane, this left cluster must lie in a three-dimensional hyperplane in five-dimensional space.

Consider also the appearance of a mode in parallel coordinates. The mode is, intuitively speaking, the location of the most intense concentration of probability. Hence in a sampling situation, it will be the location of the most intense concentration of observations. Since observations are represented by broken line segments, the mode in parallel coordinates will be represented by the most intense bundle of broken line paths in the parallel coordinate diagram. Roughly speaking, we should look for the most intense flow through the diagram. In Figure 6, such a flow begins near the center of coordinate axis 1 and finishes on the left side of axis 5.

Figure 6 thus illustrates some data analysis features of the parallel coordinate representation, including the ability to diagnose one-dimensional features (marginal densities), two-dimensional features (correlations and nonlinear structures), three-dimensional features (cluster-

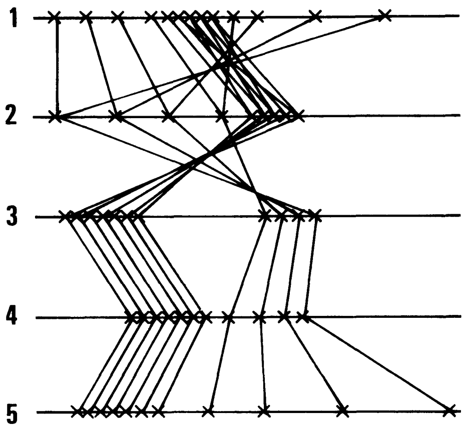

Figure 6. A Five-Dimensional Scatter Diagram in Parallel Coordinates Illustrating Marginal Densities, Correlations, Three-Dimensional Clustering and Hyperplanes, and a Five-Dimensional Mode.

ing and hyperplanes), and a five-dimensional feature (the mode). Notice that in Figure 6, the parallel coordinate axes were ordered from 1 through 5. This allowed an easy pairwise comparison of 1 with 2, 2 with 3, and so on. The pairwise comparison of 1 with 3, 2 with 5, and so on was not easily done, however, since these axes were not adjacent. A natural question arises about the number of permutations required so that in some permutation every axis is adjacent to every other axis. Although there are $n!$ permutations, many of these duplicate adjacencies. Actually, far fewer permutations are required—to be precise $(n + 1)/2$ for an n-dimensional data set. Here $(\cdot)$ is the greatest integer function. The details of this result are given in Appendix A. This result is used in Section 5, where I consider a real data set that will illustrate some additional capabilities of parallel coordinates.

5. AN AUTOMOBILE DATA EXAMPLE

We consider data on 74 1979 model-year automobiles; in particular, we consider a five-dimensional set of data consisting of measured variables—price, miles per gallon, gear ratio, weight, and cubic-inch displacement. Based on the discussion in Appendix A, for $n = 5$, $[n + 1]/2 = 3$ presentations are needed to present all pairwise permutations (see Figs. 7–9). In Figure 7, perhaps the most striking feature is the crossover effect evident in the relationship between gear ratio and weight. This suggests a negative correlation, which is reasonable because a heavy car would tend to have a large engine providing considerable torque, thus requiring a lower gear ratio. Conversely, a light car would tend to have a small engine providing small amounts of torque, thus requiring a higher gear ratio.

Consider as well the relationship between weight and cubic-inch displacement. In this diagram there is a considerable amount of approximate parallelism (relatively few crossings), suggesting positive correlation. This is a graphic representation of a fact most people are prepared to believe: that big cars tend to have big engines. Quite striking, however, is the negative slope going from low weight to moderate cubic-inch displacement. This is clearly

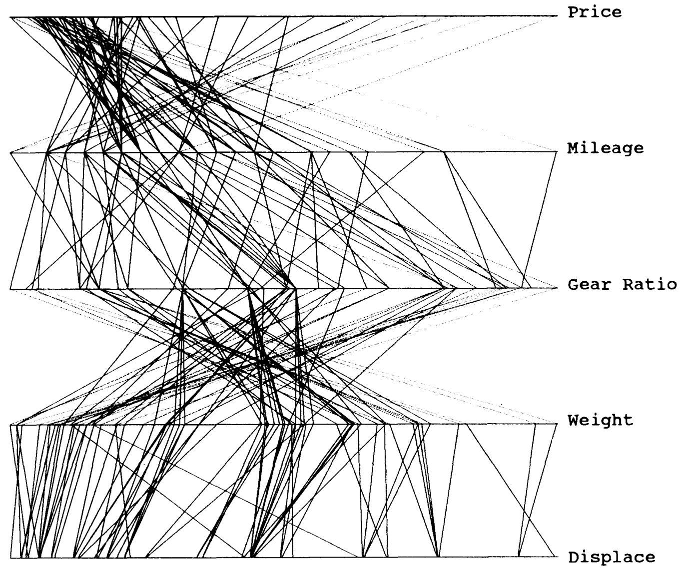

Figure 7. A Parallel Coordinate Plot in Five Dimensions of Automobile Data. Note the negative correlation between gear ratio and weight.

an outlier that is unusual in neither variable but, rather, in their joint relationship. The same observation is highlighted in Figure 8.

The relationship between miles per gallon and price is also worthy of comment. The left side shows an approximate hyperbolic boundary and the right side clearly illustrates the crossover effect. This suggests that for inexpensive cars or poor mileage cars, there is relatively little correlation. Costly cars, however, almost always get relatively poor mileage whereas good gas mileage cars are almost always relatively inexpensive.

In Figure 8, the relationship between gear ratio and miles per gallon is instructive. This diagram suggests two classes. Notice that there are a number of observations represented by line segments tilted slightly to the right of vertical (high positive slope) and a somewhat larger number with a negative slope of about -1. Within each of

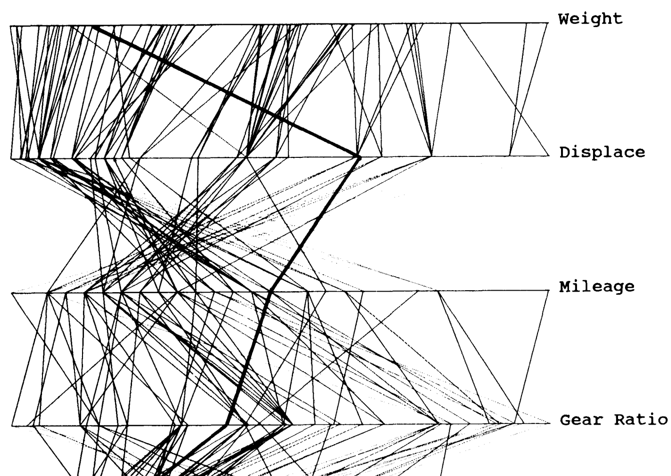

Figure 8. The Second Permutation of the Five-Dimensional Presentation of the Automobile Data. There are two classes of linear relationships between gear ratio and miles per gallon.

these two classes, there is approximate parallelism, suggesting that the relationship between gear ratios and miles per gallon is approximately linear. This is a believable conjecture because low gears = big engines = poor mileage but high gears = small engines = good mileage. It is intriguing, however, that there seem to be two distinct classes of automobiles, each exhibiting a linear relationship but with different linear relationships within each class.

Indeed in Figure 9, the third permutation, this separation into two classes is highlighted in a truly five-dimensional sense. Figure 10 represents a subset of the data in Figure 9. In particular, Figure 10(a) describes a class of vehicles that have relatively poor gas mileage, are relatively heavy, are relatively inexpensive, have relatively large engines, and have relatively low gear ratios. Figure 10(b) highlights a class of vehicles that have relatively good gas mileage, are relatively light weight, are relatively inexpensive, have relatively small engines, and have relatively high gear ratios. In 1979, these two characterizations describe, respectively, domestic automobiles and imported automobiles.

6. GRAPHICAL EXTENSIONS OF PARALLEL COORDINATE PLOTS

The basic parallel coordinate idea suggests some additional plotting devices. I call these, respectively, the parallel coordinate density plots and color histograms. These are extensions of the basic idea of parallel coordinates, but they are structured to exploit additional features to convey certain information more easily.

6.1 Parallel Coordinate Density Plots

Although the basic parallel coordinate plot is a useful device in itself, like the conventional scatter diagram, it suffers from heavy overplotting with large data sets. To get around this problem, the parallel coordinate density plot is computed as follows. The algorithm is based on Scott's (1985) notion of average shifted histogram (ASH), adapted to the parallel coordinate context. As with an ordinary two-dimensional histogram, appropriate rectangular bins are selected. A potential difficulty arises because a line segment representing a point may appear in two or more bins in the same horizontal slice. Obviously, if we have k n-dimensional observations, we would like to form a histogram based on k entries. Since the line segment could appear, however, in two or more bins in a horizontal slice, the count for any given horizontal slice is at least k and may be bigger. Moreover, every horizontal slice may not have the same count. To get around this, convert line segments to points by intersecting each line segment with a horizontal line passing through the middle of the bin. This gives an exact count of k for each horizontal slice. Construct an ASH for each horizontal slice

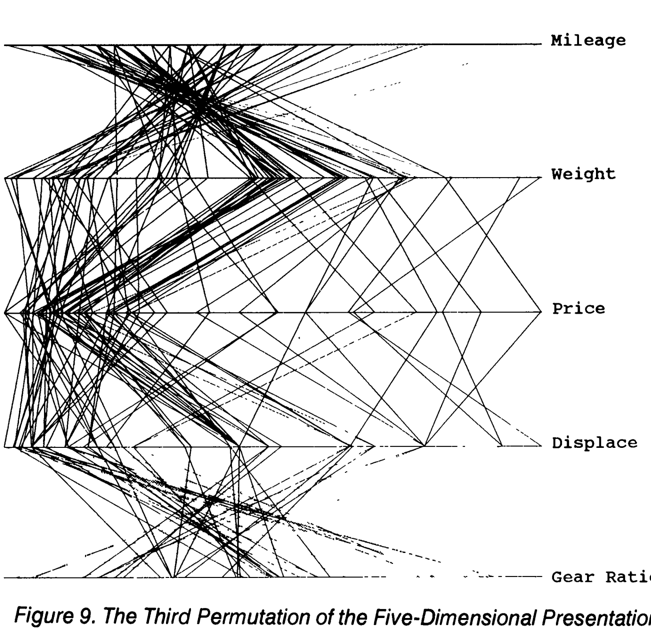

Figure 9. The Third Permutation of the Five-Dimensional Presentation of the Automobile Data. This presentation illustrates two five-dimensional clusters that are individually highlighted in Figure 10.

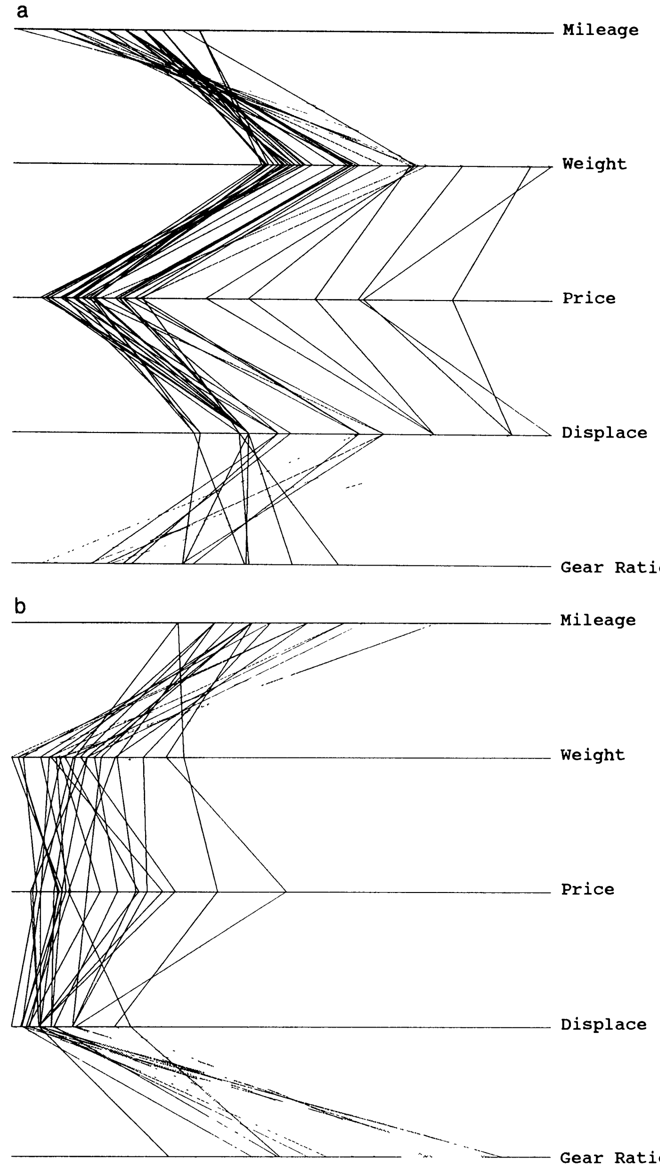

Figure 10. (a) Parallel Coordinate Plot of Cars Characterized as American Automobiles. (b) Parallel Coordinate Plot of Cars Characterized as Imported Automobiles.

(typically averaging five histograms to form the ASH). I have used contours to represent the two-dimensional density, although gray-scale shading could be used in a display with sufficient bit-plane memory. Parallel coordinate density plots have the advantage of being graphical representations of data sets that are simultaneously high dimensional and very large.

Examples of parallel coordinate density plots are given in Figures 11 and 12. These are particularly interesting because they are, respectively, the parallel coordinate density plots of a solid four-dimensional sphere and a four-dimensional sphere with a hole of radius .5. An ordinary three-dimensional sphere with a hole is a difficult challenge from an exploratory data analysis perspective because the spherical symmetry will disguise the hole in all projections. Slicing the three-sphere with a two-plane will reveal a hole. Such a device will not work with a four-sphere having a hole, however, because any three-dimensional projection of a four-sphere will be a solid three-sphere. Thus slicing with a two-plane will not reveal any hole. The differences show up readily in a parallel coordinate density plot. The construction of parallel coordinate density plots is discussed further in Miller and Wegman (1990).

6.2 Color Histograms

For an n-dimensional data set, there are n parallel axes. A vertical section through the diagram corresponds to an observation. The idea is to code the magnitude of an observation along a given axis by a color bin, the colors being chosen to form a color gradient, typically 5–15 colors. The diagram is drawn by choosing an axis, say x_k, and sorting the observations in ascending order. Along this axis, blocks of color are arranged according to the color gradient, with the width of the block being proportional to the number of observations falling into the color bin. The observations on the other axes are arranged in the order corresponding to the x_k axis and color coded according to

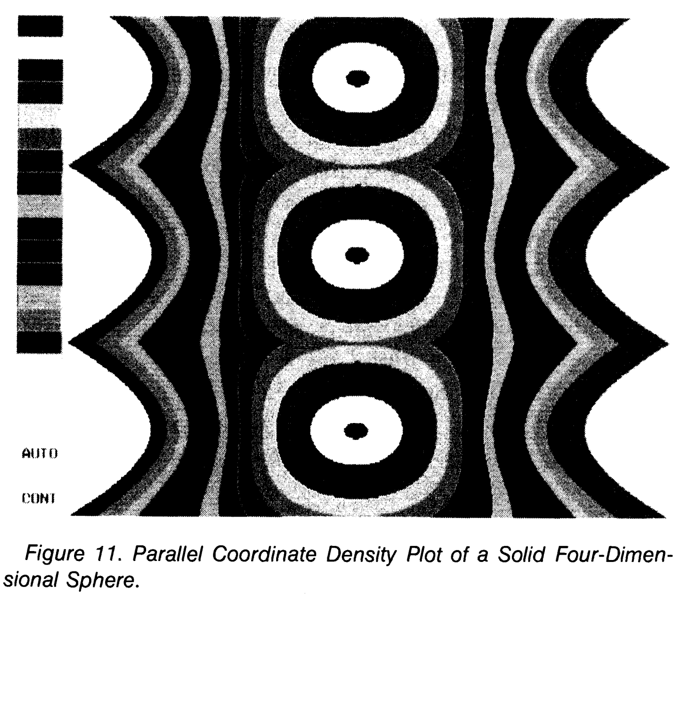

Figure 11. Parallel Coordinate Density Plot of a Solid Four-Dimensional Sphere.

their magnitude. Of course, if the same color gradient shows up on, say, the x_m axis as on the x_k, then we know x_k is positively "correlated" with x_m. If the color gradient is reversed, we know the "correlation" is negative. We used the term "correlation" advisedly, since in fact, if the color gradient is the same but the color block sizes are different, the relationship is nonlinear. Of course, if the x_k is unrelated to x_m. An example of a color histogram is given in Figure 12b. Figure 12b is based on 10 measurements of the automobile data plotted in Figures 7–10. The color histogram clearly reveals the nonlinear positive association among the variables engine displacement, turning circle, length, weight, and trunk space and the nonlinear negative association of these with gear ratio and mileage. There is also a slight positive association with price.

7. IMPLEMENTATIONS

My parallel coordinate data analysis software has been implemented in two forms. One is a PASCAL program operating on the IBM RT under the AIX operating system. This code allows for up to four simultaneous windows and offers simultaneous display of parallel coordinates and scatter diagrams. It offers highlighting, zooming, and other similar features and allows the possibility of nonlinear rescaling of each axis. It incorporates the axis permutations described in Appendix A as well as parallel coordinate density plots and color histograms. The second implementation is under development for MS-DOS machines and includes similar features. In addition, it has a mouse-driven painting capability and can do real-time rotation of three-dimensional scatterplots. Both programs use EGA graphics standards, with the second also using VGA or Hercules monochrome standards.

I regard the parallel coordinate representation as complementary to scatterplots. A major advantage of the parallel coordinate representation over the scatterplot matrix is the linkage provided by connecting points on the axes. This linkage is difficult to duplicate in the scatterplot matrix. Because of the projective line–point duality, the structures seen in a scatterplot can also be seen in a parallel coordinate plot. Moreover, Cleveland and McGill's (1984b) work suggests that it is easier and more accurate to compare observations on a common scale. The parallel coordinate plot and the derivatives of it, de facto, have a common scale and so, for example, a sense of the variability and central tendency among the variables is easier to grasp visually in parallel coordinates when compared with the scatterplot matrix. On the other hand, one might interpret all of the ink generated by the lines as a significant disadvantage of the parallel coordinate plot. My experience with this is mixed. Certainly, for large data sets on hard copy, this is a problem. On the other hand with traditional scatterplots viewed on an interactive graphics screen, particularly a high-resolution screen, I have often found that individual points can get lost because they are simply not bright enough. That does not happen in a parallel coordinate plot. If many points are plotted by using

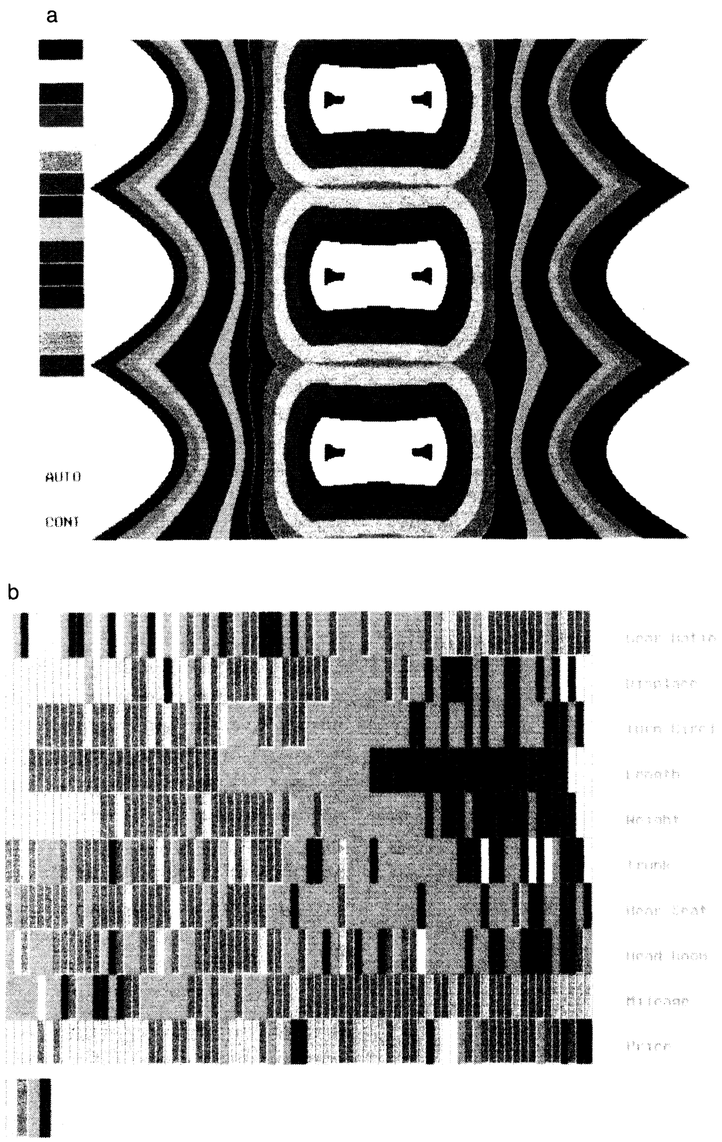

Figure 12. (a) Parallel Coordinate Density Plot of a Four-Dimensional Sphere With a Hole of Radius .5. (b) Color Histogram Representing Ten-Dimensional Automobile Data.

parallel coordinates on a monochrome screen, however, it is hard to distinguish them. I have gotten around this problem by plotting distinct points in different colors. In an EGA or VGA implementation, this means 16 colors. This is surprisingly effective in separating points. In one experiment, I plotted 5,000 five-dimensional random vectors using 16 colors. In spite of total overplotting, I was still able to see structure. In data sets of somewhat smaller scale, I have implemented a scintillation technique. With this technique, when there is overplotting, the screen view scintillates between the colors representing the overplotted points. The speed of scintillation is proportional to the number of points overplotted, and by carefully tracing colors, one can follow an individual point through the entire diagram. I have found painting to be an extraordinarily effective technique in parallel coordinates, using a painting scheme that paints not only all lines within a given rectangular area but also all lines lying between two slope constraints. This is very effective in separating clusters. I also use invisible paint to eliminate observation points from the data set temporarily. This is a natural way of doing a subset selection.

This latter implementation is known as *Mason Hypergraphics* (see Wegman and Bolorforoush 1989) and contains, in addition, the capability for rotating scatterplots, stereographic scatterplots, scatterplot matrices, star diagrams, and grand tour. I expect it to be commercially available in the near future. The RT implementation is available on request, but with very limited documentation.

APPENDIX A: PERMUTATION OF THE AXES FOR PAIRWISE COMPARISONS

A construction for determining the permutations is represented in Figure A. A graph is drawn with vertices representing coordinate axes, labeled clockwise 1 to n. Edges represent adjacencies, so vertex 1 connected to vertex 2 by an edge means axis 1 is placed adjacent to axis 2. Constructing a minimal set of permutations that completes the graph is equivalent to finding a minimal set of orderings of the axes so that every possible adjacency is present. Figure A(a) illustrates the basic zigzag pattern used in the construction, creating an order in Figure A(a) of 1 2 7 3 6 4 5. For n even, this general sequence can be written as $1, 2, n, 3, n - 1, 4, n - 2, \ldots, (n + 2)/2$, and for n odd, as $1, 2, n, 3, n - 1, 4, n - 2, \ldots, (n + 3)/2$.

An even simpler formulation is

$$n_{k+1} = [n_k + (-1)^{k+1}k] \bmod n, \qquad k = 1, 2, \ldots, n - 1, \tag{A.1}$$

with $n_1 = 1$. Here it is understood that $0 \bmod n = n \bmod n = n$. This zigzag pattern can be recursively applied to complete the graph. That is, if we let $n_k^{(1)} = n_k$, we may define

$$n_k^{(j+1)} = (n_k^{(j)} + 1) \bmod n, \qquad j = 1, 2, \ldots, [(n - 1)/2], \tag{A.2}$$

where $[\cdot]$ is the greatest integer function. For n even, it follows that this construction generates each edge in one and only one permutation. Thus $n/2$ is the minimal number of permutations needed to assure that every edge appears in the graph or, equivalently, that every adjacency occurs in the parallel coordinate representation. For n odd, the result is not exactly the same. There is no duplication of adjacencies for $j < [(n - 1)/2]$. However, $j < [(n - 1)/2]$ will not provide a complete graph. The case of $j = [(n - 1)/2]$ in Equation (A.2) will complete the graph, but it will also create some redundancies. Nevertheless, it is clear that $[(n + 1)/2]$ permutations are the minimal number needed to complete the graph and thus provide every adjacency in the parallel coordinate representation. Thus the minimal number of permutations of the n parallel coordinate axes needed to insure adjacency of every pair of axes is $[(n + 1)/2]$. These permutations may be constructed by using Equations (A.1) and (A.2).

APPENDIX B: STAR DIAGRAMS AND PARALLEL COORDINATES

The star diagrams discussed in Fienberg (1979) can be extended in a natural way by computations similar to those in Section 2. The basic idea is to consider an n-dimensional plot as follows. Draw n radial lines from a common center point at equal angles ($\theta = 2\pi/n$) and label each according to one of the variables, x_i. Mark each axis proportionally to the size of the variable, x_i, and join markings on adjacent axes by a straight line segment. This is, of course, similar to the parallel coordinate paradigm in that points map into line segments, except that for star diagrams the coordinate axes intersect. Although not commonly done, it is possible to plot all data points on the same

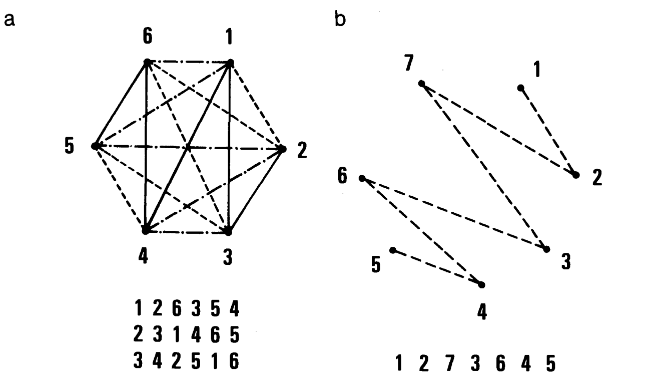

Figure A. (a) Complete Graph Representing the Three Permutations Necessary for All Adjacencies in a Six-Dimensional Parallel Coordinate Plot. (b) Basic Construction for Determining Permutations of Parallel Coordinate Axes Shown Here for a Seven-Dimensional Plot.

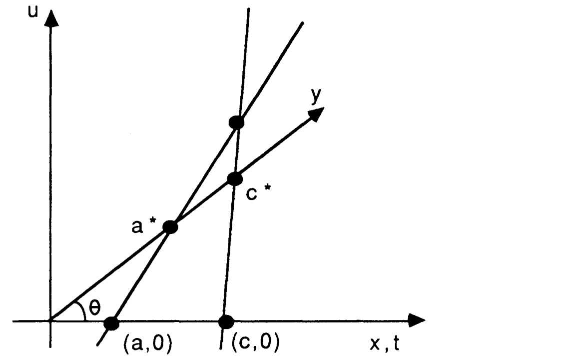

Figure B. Cartesian Coordinate Diagram and Sector of Star Plot. The tu Cartesian coordinate system is superimposed on the xy sector of the star plot. Here, a is the point [(ma + b)cos(θ), (ma + b)sin(θ)] and c* is the point [(mc + b)cos(θ), (mc + b)sin(θ)].*

diagram. One sector of such a diagram is illustrated in Figure B. If points $(a, ma + b)$ and $(x, mc + b)$ lie on a straight line, $\mathcal{L}$, in Cartesian coordinates, we have already seen that the parallel coordinate dual of $\mathcal{L}$, $\bar{\mathcal{L}}$, is a point that depends only on the parameters m and b determining $\mathcal{L}$. A natural question is whether or not such a corresponding phenomenon happens in the star diagram. It is not difficult to see that the point $(a, ma + b)$ is mapped into the line segment joining $(a, 0)$ to $[(ma + b)\cos\theta, (ma + b)\sin\theta]$, where the latter points are given in the tu coordinate system. Similarly, $(c, mc + b)$ is mapped into the line segment joining $(c, 0)$ to $[(mc + b)\cos\theta, (mc + b)\sin\theta]$. The line segments joining these points have the following respective equations in the tu coordinate system:

$$u = [(ma + b)\sin\theta(t - a)]/[(ma + b)\cos\theta - a]$$

and

$$u = [(mc + b)\sin\theta(t - c)]/[(mc + b)\cos\theta - c].$$

Simultaneous solution of these equations does not yield solutions independent of a and c, hence the locus of intersection points is not degenerate. In fact, since points on the y axis are projectively related (but not by a central perspectivity) to points on the x axis (since they are linear transforms, $x \to mx + b$) and the two axes are not coincident, by elementary projective gemetry the locus of line segments is a conic (see Fishback 1962, p. 142). If the points are related by a central perspectivity, then the conic is degenerate. This is true in the special case shown earlier for parallel coordinates. Notice that in the special case of $b = 0$, the lines given above become

$$u = [m \sin\theta(t - a)]/[m \cos\theta - 1]$$

and

$$u = [m \sin\theta(t - c)]/[m \cos\theta - 1].$$

Thus the two line segments have a common slope independent of a or c but an intercept that is dependent on a and c. Thus the lines do have a common intersection at an ideal point whose direction depends only on m and θ. The star diagrams have mathematical roots in projective geometry in common with parallel coordinates, but because of the lack of parallelism, they do not share the line–point duality properties that yield useful statistical interpretations.

[*Received October 1988. Revised November 1989.*]

REFERENCES

Andrews, D. F. (1972), "Plots of High Dimensional Data," *Biometrics*, 28, 125–136.

Asimov, D. (1985), "The Grand Tour: A Tool for Viewing Multidimensional Data," *SIAM Journal of Scientific and Statistical Computing,* 6, 128–143.

Carr, D. B., Nicholson, W. L., Littlefield, R., and Hall, D. L. (1986), "Interactive Color Display Methods for Multivariate Data," in *Statistical Image Processing and Graphics,* eds. E. Wegman, and D. DePriest, New York: Marcel Dekker, pp. 215–250.

Chernoff, H. (1973), "Using Faces to Represent Points in k-Dimensional Space," *Journal of the American Statistical Association,* 68, 361–368.

Cleveland, W. S., and McGill, R. (1984a), "The Many Faces of the Scatterplot," *Journal of the American Statistical Association,* 79, 807–822.

——— (1984b), "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods," *Journal of the American Statistical Association,* 79, 531–554.

Diaconis, P., and Friedman, J. (1983), "M and N Plots," in *Recent Advances in Statistics,* New York: Academic Press, pp. 425–447.

Dimsdale, B. (1984), "Conic Transformations and Projectivities," Technical Report 6320-2753, IBM Los Angeles Scientific Center, Santa Monica, CA.

Fienberg, S. (1979), "Graphical Methods in Statistics," *The American Statistician,* 33, 165–178.

Fishback, W. T. (1962), *Projective and Euclidian Geometry,* New York: John Wiley.

Friedman, J., and Tukey, J. W. (1973), *PRIM-9* [Film], Stanford, CA: Stanford Linear Accelerator Center, Bin 88 Productions.

——— (1974), "A Projection Pursuit Algorithm for Exploratory Data Analysis," *IEEE Transactions in Computing,* C-23, 881–889.

Griffen, H. D. (1958), "Graphic Computation of tau as a Coefficient of Disarray," *Journal of the American Statistical Association,* 53, 441–447.

Hartigan, J. A. (1975), *Clustering Algorithms,* New York: John Wiley.

Holmes, S. D. (1928), "Appendix B: A Graphical Method for Estimating R for Small Groups," in *Educational Psychology,* ed. P. Sandiford, New York: Longmans, Green, pp. 391–394.

Inselberg, A. (1985), "The Plane With Parallel Coordinates," *The Visual Computer,* 1, 69–91.

Kendall, M. G. (1961), *A Course in the Geometry of* n*-Dimensions,* London: Charles Griffen.

Miller, J. J., and Wegman, E. J. (1990), "Construction of Line Densities for Parallel Coordinate Plots," to appear in *Computational Statistics and Graphics,* eds. A. Buja and P. Tukey, New York: Springer-Verlag.

Plastock, R. A., and Kalley, G. (1986), *Computer Graphics,* New York: McGraw-Hill.

Scott, D. W. (1985), "Average Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions," *Annals of Statistics,* 13, 1024–1040.

Wegman, E. J., and Bolorforoush, M. (1989), *Mason Hypergraphics,* Fairfax Station, VA: Professional Statisticians Forum.