Mosaic Displays for Multi-Way Contingency Tables

Michael FRIENDLY*

Mosaic displays represent the counts in a contingency table by tiles whose size is proportional to the cell count. This graphical display for categorical data generalizes readily to multi-way tables. This article discusses extensions of the mosaic display to highlight patterns of deviations from various models for categorical data. First, we introduce the use of color and shading to represent sign and magnitude of standardized residuals from a specified model. For unordered categorical variables, we show how the perception of patterns of association can be enhanced by reordering the categories. Second, we introduce sequential mosaics of marginal subtables, together with sequential models for these tables. For a class of sequential models of joint independence, the individual mosaics provide a graphic representation of a partition of the overall likelihood ratio G^2 for complete independence in the full table into portions attributable to hypotheses about the marginal subtables.

KEY WORDS: Categorical data; Correspondence analysis; Graphical display; log-linear model; Marginal table.

1. INTRODUCTION

Statistical methods for categorical data, such as log-linear models and logistic regression, represent discrete analogs of the analysis of variance and regression methods for continuous response variables. But although graphical display techniques are common adjuncts to analysis of variance and regression, methods for plotting contingency table data are not as widely used.

Several schemes for representing contingency tables graphically are based on the fact that when the row and column variables are independent, the expected frequencies are products of the row and column totals (divided by the grand total). Then, each cell can be represented by a rectangle whose area shows the cell frequency or deviation from independence. The mosaic display, introduced by Hartigan and Kleiner (1981, 1984), represents each cell directly by a rectangle (or "tile") whose area is proportional to the cell frequency. Hartigan and Kleiner maintained that the pattern of tiles shown by the mosaic display is useful for suggesting hypotheses, making visual comparisons across portions of a frequency table, and highlighting unusually large and small counts.

Moreover, one form of the mosaic display extends quite naturally to multidimensional tables. For example, Hartigan and Kleiner (1984) presented a mosaic display of a four-way table of size $3 \times 6 \times 7 \times 12$ representing Nielson television ratings (number of viewers) broken down by television network, time of day, day of week, and weeks over a 3-month period. A graphical display of 1,512 cells requires some study, but it would be hard to imagine being able to see *any* patterns in a table.

This article extends the use of the mosaic display as a data-analytic tool in two ways. First, for a given display we can fit a baseline model of independence or partial independence and user color and shading of the tiles to reflect departures from that model. For unordered categorical variables, we show how perception of the pattern of association can be enhanced by reordering the categories to put residuals of like signs in opposite corners. A general scheme for reordering categories is based on a singular value decomposition (SVD) of residuals from independence. Second, for multi-way tables we find it useful to examine a sequence of mosaic displays of marginal subtables as successive variables are brought into the cross-classification. Although any log-linear model can be fit to the full table, a class of sequential models of joint independence provides a graphic representation of a partition of the overall likelihood ratio G^2 for complete independence in the full table into portions attributable to hypotheses about the marginal subtables.

Section 2 describes the construction of mosaic displays for two-way tables and introduces the use of color and shading and reordering of rows or columns to highlight patterns of departure from independence. Section 3 describes the extension of the display to multi-way tables and shows how fitting certain baseline log-linear models can be applied to the mosaic display. Section 4 illustrates the use of mosaic displays for several multi-way tables.

2. TWO-WAY TABLES

2.1 Notation

To establish notation, let $\mathbf{N} = \{n_{ij}\}$ be the observed frequency table of variables A and B with I rows and J columns. In what follows, an index is replaced by "+" when summed over the corresponding variable, so $n_{i+} = \sum_j n_{ij}$ gives the total frequency in row i, $n_{+j} = \sum_i n_{ij}$ gives the total frequency in column j, and $n_{++} = \sum\sum_{ij} n_{ij}$ is the grand total; for convenience, n_{++} is also symbolized by n. Estimated expected frequencies, under the hypothesis of independence, are denoted $\hat{m}_{ij} = (n_{i+}n_{+j})/n$. Finally, we express a standardized residual for cell i, j as d_{ij}. For Pearson X^2, for example, $d_{ij} = (n_{ij} - \hat{m}_{ij})/\sqrt{\hat{m}_{ij}}$, the signed root contribution for cell i, j, so that $X^2 = \sum_i \sum_j d_{ij}^2$.

2.2 Two-way Mosaics

Table 1 shows data on the relation between hair color and eye color among 592 students in a statistics course collected by Snee (1974). The Pearson X^2 for these data is 138.3 with 9 degrees of freedom, indicating substantial departure from independence. The question is how to understand the *nature* of the association between hair and eye color.

* Michael Friendly is Associate Professor of Psychology and Associate Director of the Statistical Consulting Service, York University, Toronto, Ontario, Canada M3J 1P3. The author thanks John Fox, John Sall, Howard Wainer, and the referees for helpful comments.

Table 1. Hair Color–Eye Color Data

Eye Color	Hair Color				
	Black	Brown	Red	Blond	Total
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
Total	108	286	71	127	592

For any two-way table, the expected frequencies under independence can be represented by rectangles whose widths are proportional to the total frequency in each column, n_{+j}, and whose heights are proportional to the total frequency in each row, n_{i+}; the area of each rectangle is then proportional to $\hat{m}_{ij}$. Figure 1 shows the expected frequencies for hair and eye color.

One form of the mosaic display, which suggests the name "mosaic," is similar to a divided bar chart. The width of each tile in Figure 2 is still proportional to the marginal frequency n_{+j} in each column of the table; but the height is proportional to the conditional frequency $n_{i|j}$ of each row (eye color) for a given column (hair color), so the area is proportional to cell frequency and complete independence is shown when all tiles in each row have the same height, as in Figure 1.

Several cells stand out in Figure 2. There are more blue-eyed blonds and brown-eyed black-haired people than would occur under independence, and fewer people with brown eyes and blond hair. But the rows are no longer aligned, except for the first and last rows. This makes it easier to make comparisons within hair color (columns) groups, but harder to make comparisons within eye color groups. A similar plot could be made with the first division proportional to the row totals, which would facilitate comparisons among eye color groups.

2.3 Detecting Patterns

In Hartigan and Kleiner's (1981) original version, all the tiles are unshaded and drawn in one color, so only the relative sizes of the rectangles indicate deviations from independence. We can increase the visual impact of the mosaic by using shading to reflect the size of the residual and by reordering rows and columns to make the pattern more coherent.

Color and Shading. Figure 3 extends the mosaic plot, showing the standardized deviation from independence, d_{ij}, by the color and shading of each rectangle. Cells with positive deviations are drawn black, outlined with solid lines, with shading slanted from upper left to lower right (NE to SW); negative deviations are drawn red, outlined with broken lines and shaded SE–NW. (The mosaic displays are most effective when seen in color; however, the sign information is lost if the figure is reproduced as shown here in monochrome, so we represent the sign of the deviation redundantly. The simplest solution is to add a stick-on dot with a "+" sign to cells with positive residuals, which seems to work well.) The absolute value of the deviation is portrayed by shading density.

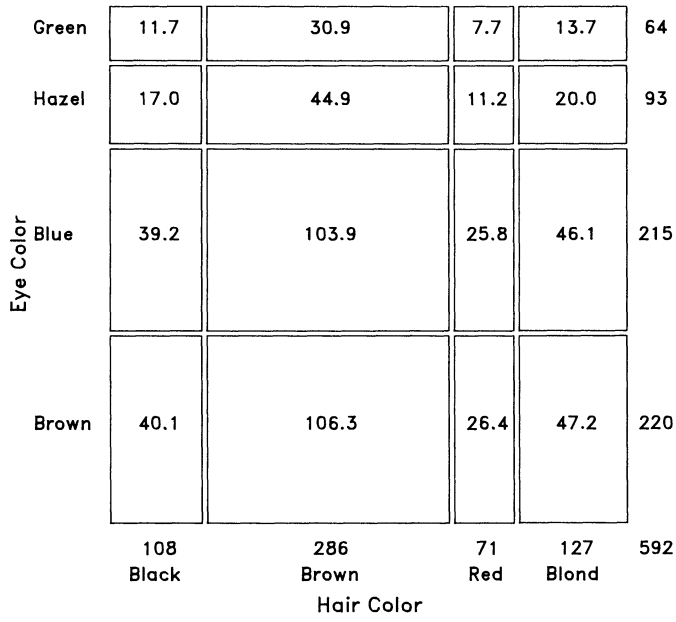

Figure 1. Expected Frequencies Under Independence. The height and width of each box is proportional to the marginal totals. Area is proportional to expected frequency, shown in each box.

Cells with absolute values less than 2 are empty and cells with $|d_{ij}| \geq 2$ are filled; those with $|d_{ij}| \geq 4$ are filled with a darker pattern. Standardized deviations are often referred to a standard Gaussian distribution; under the assumption of independence, these values roughly correspond to two-tailed probabilities $p < .05$ and $p < .0001$ that a given value of $|d_{ij}|$ exceeds 2 or 4.

Reordering Categories. When the row or column variables are unordered, we are also free to rearrange the corresponding categories in the plot to help show the nature of

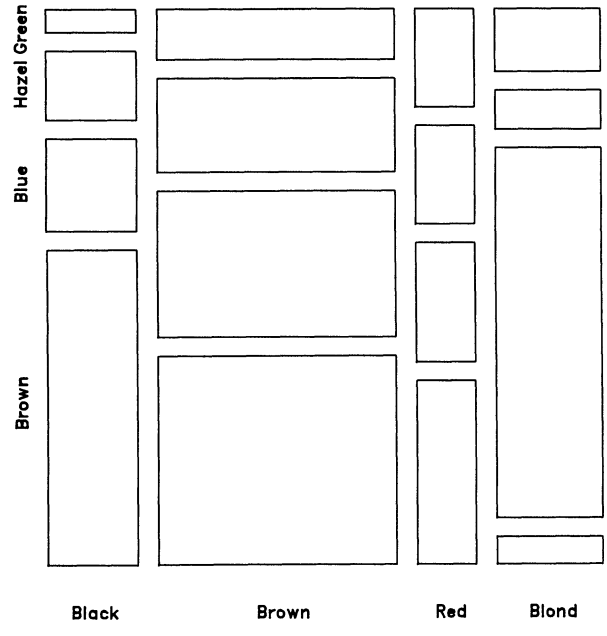

Figure 2. Condensed Column Proportion Mosaic. Each rectangle shows the conditional frequency of eye color given hair color.

association. For example, in Figure 3, the eye color categories have been permuted so that the deviations from independence have an opposite-corner pattern, with positive values running from SW to NE corners and negative values running along the opposite diagonal. With the reordering of categories coupled with size and shading of the tiles, the excess in the black-brown and blond-blue cells is even more apparent than it is in a shaded but unreordered version (not shown). We can also see clearly an overrepresentation of persons with red hair and green eyes and an underrepresentation of persons with blond hair and brown eyes and persons with black hair and blue eyes relative to independence. You could see the same things in a table of d_{ij} values if you looked hard enough, but the enhanced mosaic display makes the pattern apparent.

Note that although both hair color and eye color were treated as nominal variables, eligible for reordering, they also could be considered to be ordered along a dark–light dimension. Although the table was reordered based on the d_{ij} values, both dimensions in Figure 3 are ordered from dark to light, suggesting an explanation for the association.

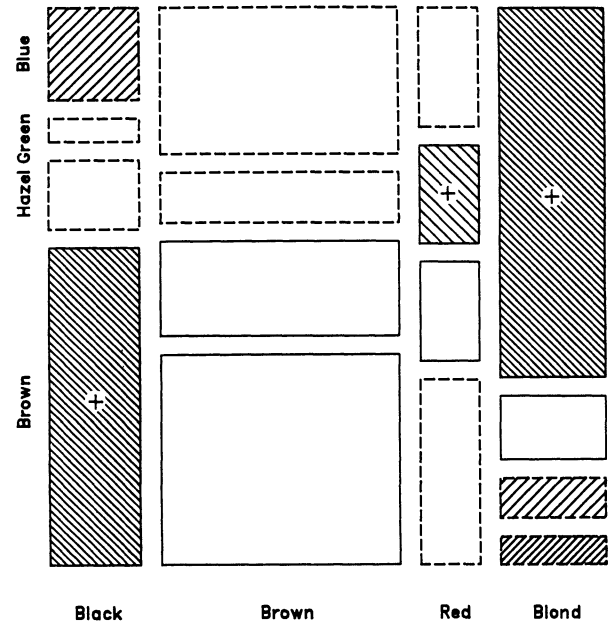

Figure 3. Condensed Mosaic, Reordered and Shaded. Deviations from independence are shown by shading. Positive deviations have solid outlines and are shaded NE-SW. Negative deviations have dashed outlines and are shaded SE-NW. The two levels of shading density correspond to standardized deviations greater than 2 and 4 in absolute value. This form of the display generalizes readily to multi-way tables.

3. MULTI-WAY TABLES

The condensed form of the mosaic plot generalizes readily to the display of multidimensional contingency tables. Imagine that each cell of the two-way table for hair and eye color is further classified by one or more additional variables—sex and ethnic group, for example. Then each rectangle in the mosaic plot can be subdivided vertically to show the proportion of males and females in that cell, and each of those portions can be subdivided horizontally to show the proportions of persons of each ethnicity in the hair–eye–sex group.

3.1 Constructing the *n*-way Mosaic

The steps in constructing the mosaic display are described here for a four-way table of variables A, B, C, and D with frequencies n_{ijkl}, $i = 1, \ldots, I$; $j = 1, \ldots, J$; $k = 1, \ldots, K$; and $l = 1, \ldots, L$. The extension to n-way tables is immediate. The construction of the mosaic is governed by an ordering of the variables, which we designate by the symbols for the table dimensions; for example, $IJKL$. The division into tiles usually alternates vertically and horizontally in this order:

1. First, the available display area is divided into vertical strips proportional to the marginal totals of variable A, so the widths are proportional to n_{i+++}.
2. Each vertical strip is then subdivided horizontally proportional to joint frequencies with the second variable, n_{ij++}. Thus each tile has a height proportional to the conditional frequency of the second variable given the first, n_{ij++}/n_{i+++}, and area proportional to n_{ij++}.
3. Next, each IJ rectangle is divided vertically proportional to n_{ijk+}, giving widths proportional to n_{ijk+}/n_{ij++}.
4. Finally, each IJK tile is divided horizontally to give areas proportional to the cell frequency n_{ijkl}.

Spacing. This procedure gives a mosaic of $IJKL$ tiles with no spacing, in which cells with small frequencies are difficult to see. Following Hartigan and Kleiner, the tiles are separated, with larger spacings at the earlier subdivisions, to help preserve the visual impact of small counts. For a four-way table with vertical splitting on dimensions I and K, the divisions of the first variable are spaced proportionally to $1/(I-1)$; divisions between levels of the third variable are spaced proportionally to $1/(IK-1)$. A more detailed description of the algorithm and a FORTRAN program for constructing n-way mosaics was given by Wang (1985). The displays shown here are implemented in SAS/IML software (SAS Institute 1989), whose combination of matrix operations, built-in functions for contingency table analysis, and graphics provide a convenient environment for graphical display for multi-way categorical data (Friendly 1991, 1992a). The program, MOSAICS.SAS (Friendly 1992b), is available for anonymous FTP transfer from UICVM.CC.UIC.EDU in the directory UICSTAT.

3.2 Fitting Log-linear Models

When three or more variables are represented in the mosaic, we can fit several different models of independence and display the residuals from that model. For a given model, we find the expected frequencies by iterative proportional fitting (Deming and Stephan 1940). We treat these models as null or baseline models, which may not fit the data particularly well. The deviations of observed frequencies from expected, displayed by shading, will often suggest terms to be added to an explanatory model that achieves a better fit.

For example, the model of complete independence, the log-linear model $[A][B][C]$ for a three-way table, puts all higher terms (and hence all association among the variables)

Table 2. Frequencies of Hair Color by Eye Color by Sex

Sex	Eye Color	Hair Color: Black	Brown	Red	Blond	Total
Male	Brown	32	38	10	3	83
	Blue	11	50	10	30	101
	Hazel	10	25	7	5	47
	Green	3	15	7	8	33
	TOTAL	56	127	34	46	264
Female	Brown	36	81	16	4	137
	Blue	9	34	7	64	114
	Hazel	5	29	7	5	46
	Green	2	14	7	8	31
	TOTAL	52	158	37	81	328

into the residuals. Another possibility is to fit the model in which variable C is jointly independent of variables A and B, the log-linear model $[AB][C]$. Residuals from this model show the extent to which variable C is related to the combinations of variables A and B, but they do not show any association between A and B. The simplest extension of joint independence to four variables is the model $[ABC][D]$.

Residuals. For computational simplicity, all examples in this article use standardized Pearson residuals. When the model being fit holds, the d_{ij} are asymptotically normal with mean 0, but their asymptotic variance, v_{ij}, is less than 1.0, with average value $\bar{v}_{ij}$ = (residual df)/(number of cells). Hence, when fitting models more complex than the model of mutual independence, the v_{ij} may be considerably less than 1, and the use of conventional Gaussian values such as ± 2 and ± 4 may be highly conservative (Agresti 1990) and fail to nominate some cells whose departure from the model should be noticed. One solution, which we simply note here, is to scale the standardized residual by its estimated standard error, giving Haberman's (1973) adjusted residuals, $r_{ij} = d_{ij}/\sqrt{v_{ij}}$, which does have an asymptotic $N(0, 1)$ distribution.

3.3 Example: Hair Color by Eye Color by Sex

Table 2 shows the breakdown of the hair color–eye color data by sex (the division by sex is contrived). Fitting the model [HairEye][Sex] allows us to see the extent to which the joint distribution of hair color and eye color is associated with sex. We might motivate such a model by asking whether the genetic information that determines hair and eye color is sex-linked. For this model, the likelihood ratio G^2 is 29.35 on 15 df ($p = .015$), indicating some lack of fit.

The three-way mosaic, shown in Figure 4, highlights two cells; males are underrepresented among persons with brown hair and brown eyes and overrepresented among persons with brown hair and blue eyes. Females in these cells have the opposite patterns, of course; however, the standardized deviates are just shy of the criterion, $|d_{ij}| \geq 2$, for shading. The d_{ij}^2 for these four cells account for 15.3 of the X^2 for the model [HairEye] [Sex]. Except for these cells, hair color and eye color appear unassociated with sex.

For comparison, the three-way mosaic for deviations from mutual independence, [Hair][Eye][Sex], is shown in Figure 5. The G^2 for this model is 179.79 on 24 df ($p < .0001$). The pattern of deviations roughly combines features seen in Figure 3 and Figure 4: the positive deviations along the SE–NW diagonal reflect the hair–eye association, and the deviations in cells for brown hair reflect the pattern seen in Figure 4. As we shall see, this result is implied by the sequence of models, [Hair][Eye] and [HairEye][Sex], whose G^2 values sum to that for the model [Hair][Eye][Sex].

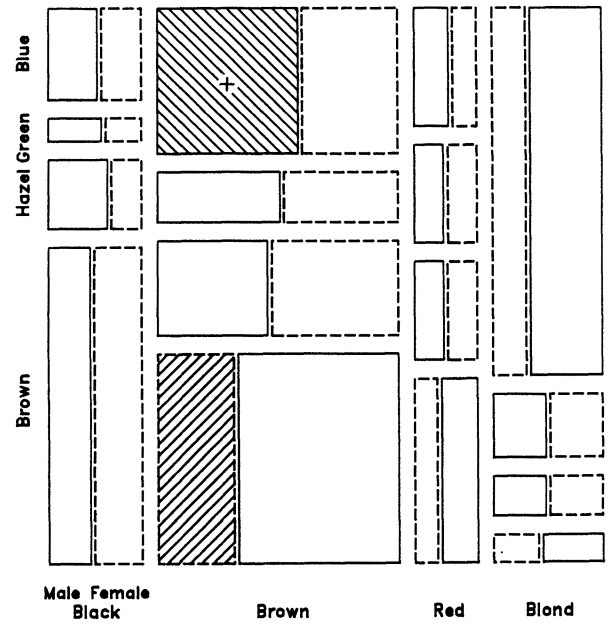

Figure 4. Mosaic Display for Hair Color, Eye Color, and Sex. The categories of sex are crossed with those of hair color, but only the first occurrence is labeled. Residuals from the model [HE][S] are shown by shading. $G^2 = 29.35$ on 15 df. The lack of fit is attributable mainly to the cells for brown hair and brown or blue eyes.

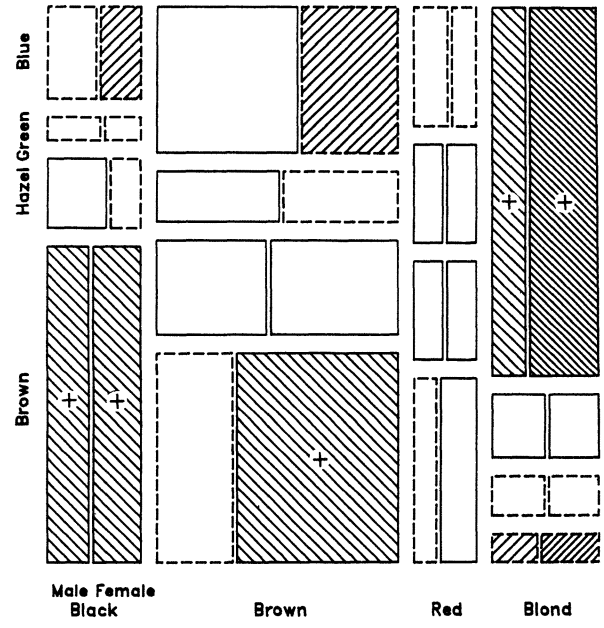

Figure 5. Mosaic Display for Hair Color, Eye Color, and Sex. This display shows residuals from the model of complete independence, [H][E][S], $G^2 = 179.79$ on 24 df.

3.4 Sequential Plots

The mosaic for a complete n-way table depicts information about the joint frequency distribution at several levels:

a. Cell frequencies are shown directly by the area of each tile.
b. The scheme for spacing allows visually combining the cells for nested marginals along both axes. For example, in Figure 5 we can visually combine across sex to see the marginal hair–eye frequencies or across eye color and sex to see the one-way frequencies of hair color in the table.
c. Approximate significance of cell contributions to lack of fit for a specified model is shown by shading.

This is a great deal of information for one plot, however. It may be more instructive, therefore, to examine the sequence of mosaics as each variable is brought into the analysis. For an n-way table, we can draw the mosaic for *each* of the marginal subtables, $\{A\}$, $\{AB\}$, $\{ABC\}$, and so on, not just the final stage as Hartigan and Kleiner did. We now consider a set of sequential models for these subtables whose residuals can be displayed in the corresponding mosaic.

3.5 Sequential Models

For the final n-way plot, we can fit any specified log-linear model or one of the baseline models described in Section 3.2. For the intermediate plots, with an arbitrary non-baseline model, we fit the reduced model in which all variables not included in the current plot are eliminated. For example, consider a four-way table in which variable D is a response and the model of interest is $[ABC][ABD][ACD][BCD]$. Then the reduced model for the two-way $\{AB\}$ plot is $[AB]$, and the reduced model for the three-way $\{ABC\}$ plot is $[ABC]$. In this case the reduced models are saturated, so all residuals are 0 and the mosaic displays simply show the marginal $\{AB\}$ and $\{ABC\}$ frequencies, which are considered fixed when variable D is the response.

For exploratory purposes, the baseline models fit some of the gross structure of the data, leaving the patterns of association that remain to be displayed in the residuals. The sequential models of joint association are of particular interest, because, following Goodman (1970, 1971), the series of hypotheses about the marginal subtables provides a partition of the hypothesis of complete independence in the full table.

Consider the hypothesis of complete independence in a three-way table. Let $H_{\{A\otimes B\}}$ denote the hypothesis that A and B are independent in the marginal subtable formed by collapsing over variable C, and let $H_{\{AB\otimes C\}}$ denote the hypothesis of joint independence of C from the AB combinations. Then Goodman's (1970, sec. 6.3) method shows that the hypothesis of complete independence, $H_{\{A\otimes B\otimes C\}}$ can be expressed as $H_{\{A\otimes B\otimes C\}} = H_{\{A\otimes B\}} \cap H_{\{AB\otimes C\}}$. When expected frequencies under each hypothesis are estimated by maximum likelihood, the likelihood ratio G^2s are additive: $G^2_{\{A\otimes B\otimes C\}} = G^2_{\{A\otimes B\}} + G^2_{\{AB\otimes C\}}$. For example, for the hair–eye data, G^2 for the model [Hair][Eye][Sex] is 179.79 on 24 df. Figure 3 and Figure 4 can be viewed as representing the partition

Model	df	G^2
[Hair] [Eye]	9	146.44
[Hair, Eye] [Sex]	15	29.35
[Hair] [Eye] [Sex]	24	179.79

This partitioning scheme extends readily to higher-way tables. For a four-way table, the hypothesis of complete independence gives rise to an analogous partitioning, $G^2_{\{A\otimes B\otimes C\otimes D\}} = G^2_{\{A\otimes B\}} + G^2_{\{AB\otimes C\}} + G^2_{\{ABC\otimes D\}}$.

This sequence of models of joint independence has another interpretation when the ordering of the variables is based on a set of ordered causal hypotheses regarding the relationships among variables (Fienberg 1980; Goodman 1973). Suppose, for example, that the causal ordering of four variables is $A \rightarrow B \rightarrow C \rightarrow D$, where the arrow means "is antecedent to." Goodman suggested that the conditional joint probabilities of B, C, and D given A can be characterized by the recursive logit models, which treat B as a response to A, treat C as a response to A and B jointly, and treat D as a response to A, B, and C. These are equivalent to the log-linear models that we fit as the sequential baseline models of joint independence, namely $[A][B]$, $[AB][C]$, and $[ABC][D]$. The combination of these models with the marginal probabilities of A gives a characterization of the joint probabilities of all four variables.

3.6 Reordering Categories

Bertin (1983, pp. 168–169) gave numerous examples of reordering the categories of nominal qualitative variables to simplify and enhance, sometimes dramatically, the perception of associations. His method is essentially one of inspection, trial and error, though he also describes a physical device called a "domino" for permuting the rows and columns of an array.

Given an ordering of the variables, the categories of the nominal variables can be reordered on the basis of sequential plots showing deviations from models of joint independence. With the variables labeled A, B, C, and D in order of division in the mosaic, the categories of variables A and B can be reordered from the $[A][B]$ plot, those of variable C can be reordered from the deviations in the $[AB][C]$ plot, and so forth. Experience shows that for small tables the orderings to diagonalize the pattern of residuals can often be determined by inspection.

A more general approach is based on the ideas of correspondence analysis (CA) (see, for example, Greenacre 1984 and Greenacre and Hastie 1987), which assigns scores to the categories so that the Pearson correlation of the optimally scaled variables is maximized. For a two-way table, the scores for the row categories, namely x_{im}, and column categories, y_{jm}, on dimension $m = 1, \ldots, M$ are derived from the SVD of Pearson residuals to account for the largest proportion of the X^2 in a small number of dimensions. This decomposition may be expressed as $d_{ij}/\sqrt{n} = \sum_{m=1}^{M} \lambda_m x_{im} y_{jm}$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$ and $M = \min(I - 1, J - 1)$. A rank-d approximation in d dimensions is obtained from the first d singular values; the proportion of X^2 accounted for by this approximation is $n \sum^d \lambda_m^2 / X^2$.

Therefore, rearranging row or column categories according to the CA scores x_{i1} or y_{j1} on the first (largest) dimension should provide an ordering for the mosaic display to best reveal the pattern of association. This ordering captures the nature of the association to the extent that $\lambda_1^2 / \sum^M \lambda_m^2$ is large. For the hair–eye data, for example, the singular values are .456 (89%), .149, and .051. The order of the scores for eye colors on the first dimension is precisely the order determined by inspection from Figure 2.

For higher-order tables, van der Heijden and de Leeuw (1985) described extensions of CA that have close links with log-linear models and with the sequential mosaic displays described here. Their method is to apply the usual correspondence analysis to a two-way "multiple table," n_{GH}, where G and H are nonoverlapping subsets of the variables A, B, . . . , which constitute the complete table. For a three-way table reshaped as $n_{(ij)k}$, they showed that the correspondence analysis solution can be interpreted in terms of residuals from the model $[AB][C]$ in a way that corresponds to the mosaic display of residuals from the same model. The row and column scores on the first dimension account for maximal association between the joint AB categories and the C categories. This suggests, therefore, that successive variables in the mosaic can be reordered in terms of their scores on the first correspondence analysis dimension from the table in which all previous variables are coded interactively.

4. EXAMPLES

Two examples are presented to illustrate the use of mosaic displays. The first example compares the results obtained with different orderings of the variables and demonstrates the use of the correspondence analysis SVD to reorder categories. The second example concentrates on the use of the mosaic display for model building. (See Friendly and Fox 1991 for the analysis of a five-way table and comparison of mosaic displays with effects plots for logit models.)

4.1 Suicide Data

We use data on suicide rates in West Germany, classified by age, sex, and method of suicide, to illustrate the effects of reordering variables and categories in analysis of multiway tables. In this section A, S, and M represent these variables. The data, from Heuer (1979, table 1), have been discussed by van der Heijden and de Leeuw (1985) and others.

Table 3. Frequencies of Suicide by Age, Sex, and Method

		Method					
Sex	Age	Poison	Gas	Hang	Drown	Gun	Jump
M	10–20	1160	335	1524	67	512	189
M	25–35	2823	883	2751	213	852	366
M	40–50	2465	625	3936	247	875	244
M	55–65	1531	201	3581	207	477	273
M	70–90	938	45	2948	212	229	268
F	10–20	921	40	212	30	25	131
F	25–35	1672	113	575	139	64	276
F	40–50	2224	91	1481	354	52	327
F	55–65	2283	45	2014	679	29	388
F	70–90	1548	29	1355	501	3	383

Table 4. Log-Linear Models for Suicide Data

Model	df	Likelihood ratio G^2	Pearson X^2
[M] [A] [S]	49	10119.6	9908.2
[M] [AS]	45	8632.0	8371.3
[A] [MS]	44	4719.0	4387.7
[S] [MA]	29	7029.2	6485.5
[MS] [AS]	40	3231.5	3030.5
[MA] [AS]	25	5541.6	5135.0
[MA] [MS]	24	1628.6	1592.4
[MA] [MS] [AS]	20	242.0	237.0

The original $2 \times 17 \times 9$ table contains 17 age groups from 10 to 90 in 5-year steps and 9 categories of suicide method. To avoid extremely small cell counts, this example uses a reduced table in which age groups are combined in 15-year intervals except for the last interval, which includes ages 70–90; the methods "toxic gas" and "cooking gas" were collapsed and the methods "knife" and "other" were deleted, giving the $2 \times 5 \times 6$ table shown in Table 3. These changes do not affect the general nature of the data.

The variables in Table 3 could be ordered in several different ways. if we regard method of suicide as a response to background variables of age and sex, then the order S, A, M is indicated. On the other hand, if no variable is singled out as a response, then other considerations may be used. A purely graphic consideration is that with three variables, the first and third will divide one dimension of the mosaic. All other things equal, it is useful to choose an order for which the product of levels, IK, is not too large, to preserve resolution in the plot. For example, the order A, S, M would have 30 divisions vertically by 2 horizontally. Because the relation between method of suicide and age, ignoring sex seems of some interest in addition to the three-way relation, we consider the variables in the order M, A, S. To illustrate the effect of changing the order of the variables, we also show the mosaic for the order S, A, M.

For a multi-way table, it is also useful to examine the fit of various log-linear models and choose an order of variables based on the association terms that appear to be important. Table 4 shows the results of all possible hierarchical log-linear models for the suicide data. It is apparent that none of these models has an acceptable fit. Given the enormous sample size (n = 48,177), even relatively small departures from any unsaturated model would appear significant, however. Nevertheless, from the differences among the G^2 and X^2 values in the sections of Table 4 containing zero, one, two, and three two-way association terms, it is clear that all three two-way associations have significant effects.

Figure 6 shows the initial mosaic for method and age in the order M, A, S, with the methods arranged as in Table 3. To show age on the horizontal axis in these figures, the first variable is placed on the vertical dimension. Some trends across age are apparent: The methods POISON, GAS, and GUN are prevalent in younger ages and decrease with age, whereas the methods HANG and DROWN show the opposite pattern.

The pattern of association between method and age can be clarified by reordering the methods according to the

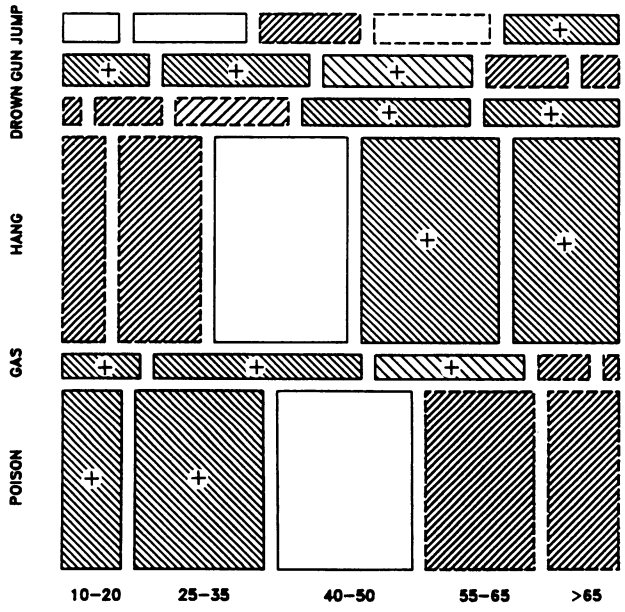

Figure 6. Mosaic Display for Method and Age. Residuals from the model [M][A] are shown by shading. The methods are ordered as in Table 3. Some trends with age can be seen, but the overall pattern of association is unclear.

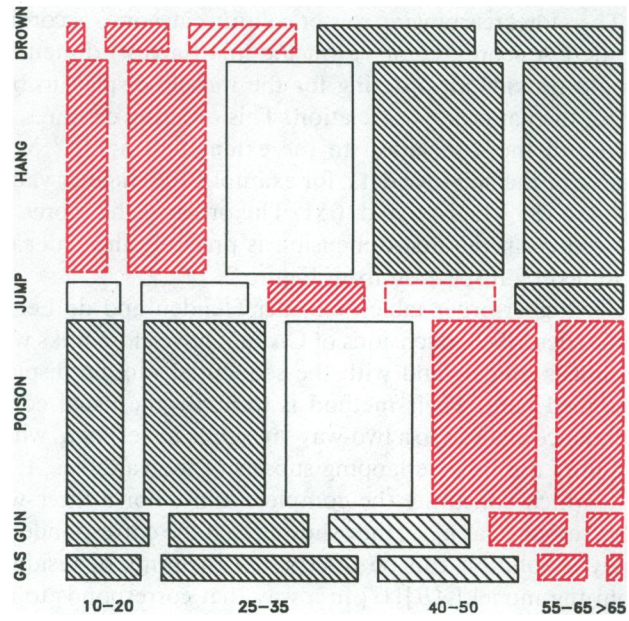

Figure 8. Mosaic Display for Method and Age, Reordered. In color figures, positive residuals are shown in black and negative residuals in red. The methods have been reordered according to their positions on dimension 1 of Figure 7. The pattern of association between method and age is now more apparent.

method scores from the SVD of the d_{ij} values. The first two dimensions of the correspondence analysis solution for the *MA* marginal table are displayed in Figure 7. The configuration is essentially one-dimensional; the first dimension accounts for 94% of the Pearson X^2.

With the methods of suicide reordered according to their positions on dimension 1 of Figure 7, the mosaic in Figure 8 is produced. The opposite corner pattern of residuals makes it easier to see the relations between methods and age described earlier. In addition, we see that the pattern of deviations for JUMP differs from an increasing or a decreasing trend with age that characterizes the other methods. This is reflected in the position of JUMP on dimension 2 of Figure 7.

The three-way mosaic showing deviations from the model [Method, Age][Sex] is shown in Figure 9, where each *MA* marginal is partitioned according to the conditional proportions of males and females. We see that the prevalent use of GAS and GUN among younger people is associated more closely with males. POISON, JUMP, and DROWN are common among females, the last two particularly among older females.

Again we note that the deviations displayed in Figure 8 and Figure 9 represent a partition of the overall G^2 for complete independence in the full three-way table. The mosaic display for deviations from the independence model,

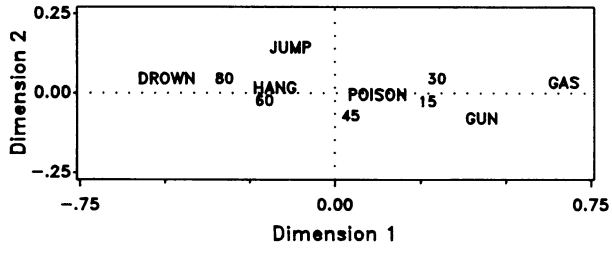

Figure 7. Two-Dimensional Correspondence Analysis Solution for the [M][A] Table. The origin represents the marginal profiles p_{i+} and p_{+j} for both the methods of suicide and age groups (which are labeled by their midpoint). Method and age points with similar positions correspond to cells with positive deviations from expected frequencies in the two-way table. The association between method and age is accounted for almost entirely by the positions on dimension 1.

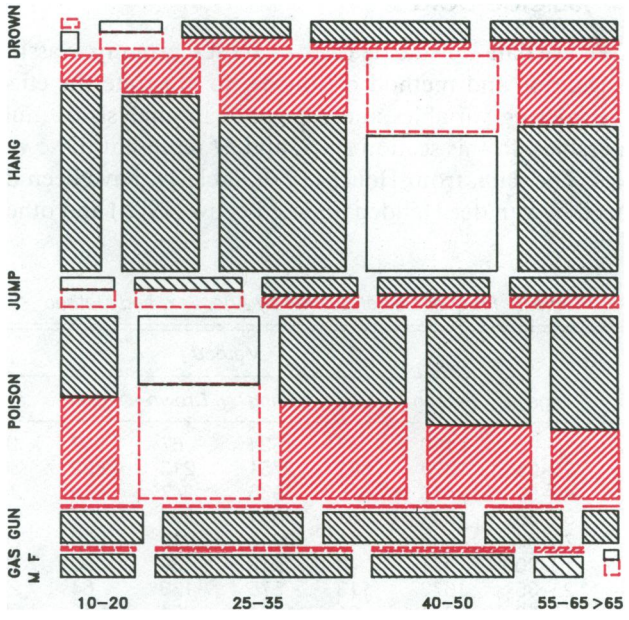

Figure 9. Mosaic Display for Method, Age, and Sex. Residuals from the model [MA][S] are shown by shading.

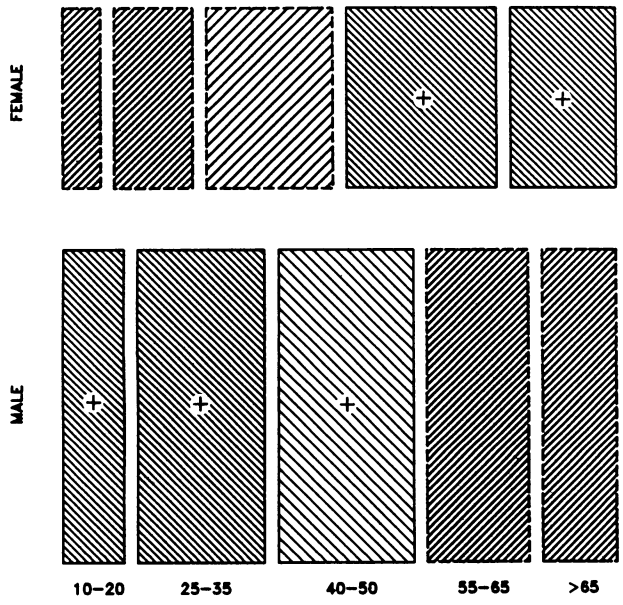

Figure 10. Mosaic Display for Sex and Age. The frequency of suicide shows opposite trends with age for males and females.

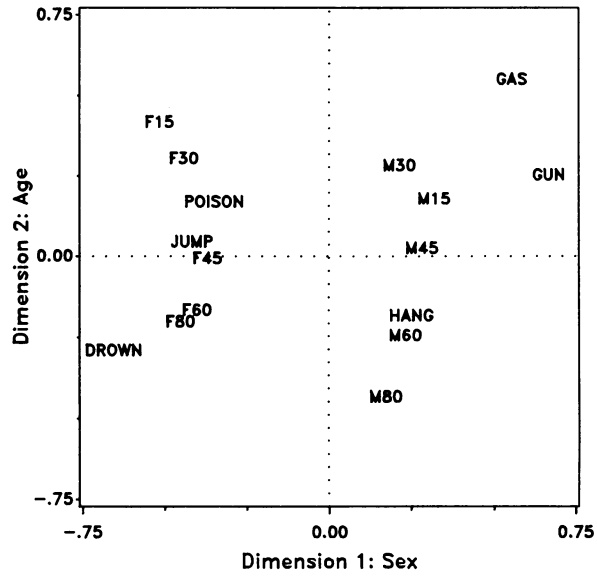

Figure 12. Two-Dimensional Correspondence Analysis Solution for the [SA][M] Multiple Table. The positions of the methods along dimension 1 provides an optimal ordering for the mosaics with the variables ordered S,A,M.

$[M][A][S]$, has the same-sized tiles as in Figure 9, of course. To save space, we do not display this mosaic.

The analogous mosaics for the order S, A, M are displayed in Figure 10 and Figure 11. Figure 10 displays the marginal relation between sex and age of persons committing suicide, ignoring method. Suicide is more common in males than in females. For males, the tendency to commit suicide decreases with age, whereas females show an opposite trend.

Figure 11 displays the breakdown of the sex–age groups by suicide method. The methods have been arranged in order of the method scores on the first dimension of the CA solution for the [Sex, Age][Method] multiple table. That solution represents a partition of the Pearson $X^2 = 8371.3$ for the model $[SA][M]$ in $M = 5$ dimensions. The first two dimensions, shown in Figure 12, account for 60.4% and 33.0% of the residuals from this model. Note that the order of the methods on dimension 1 of Figure 12 differs somewhat from the order of the methods shown in Figure 7. Comparing Figure 7 and Figure 12, it may be seen that dimension 1 in Figure 7 corresponds to projections on an axis at approximately 60 degrees to the horizontal in Figure 12.

Residuals from model [Sex, Age][Method], displayed in Figure 11, again show the prevalence of GUN and GAS among younger males and decreasing with age, whereas HANG increases with age. For females, these three methods are used less frequently than would be the case if method were independent of age and sex, whereas POISON, JUMP, and DROWN occur more often.

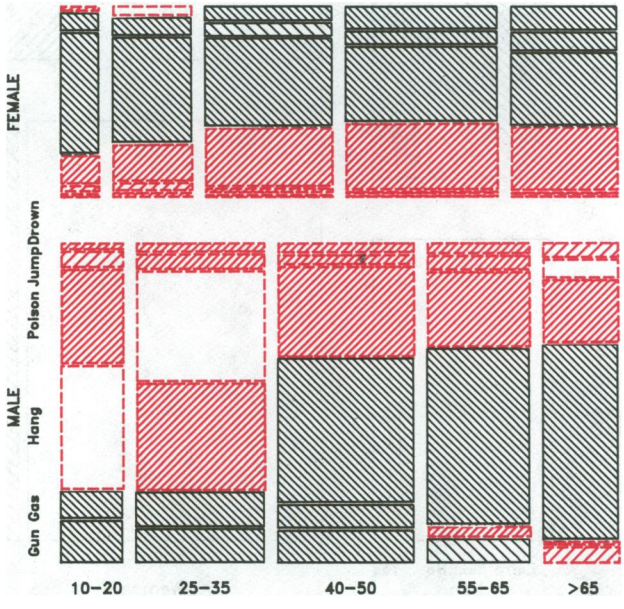

Figure 11. Mosaic Display Showing Deviations from Model [SA][M]. The methods have been reordered according to their positions on dimension 1 of the correspondence analysis solution for the [SA][M] table.

4.2 Marital Status and Premarital and Extramarital Sex

Table 5 lists the 2^4 table from a study of divorce patterns reported by Thornes and Collard (1979) and analyzed by

Table 5. Marital Status in Relation to Gender and Report of Premarital and Extramarital Sex

Gender	Premarital	Extramarital	Divorced	Still Married
Men	N	N	68	130
	N	Y	17	4
	Y	N	60	42
	Y	Y	28	11
Women	N	N	214	322
	N	Y	36	4
	Y	N	54	25
	Y	Y	17	4

Table 6. Sequential Decomposition of Likelihood Ratio G^2 for Marital Data

Model	df	G^2
[G] [P]	1	75.259
[GP] [E]	3	48.929
[GPE] [M]	7	107.956
[G] [P] [E] [M]	11	232.142

Gilbert (1981) and Agresti (1990, sec. 7.2.4). A sample of about 500 people who had petitioned for divorce and a similar number of married people were asked two questions regarding their premarital and extramarital sexual experience: (1) "Before you married your (former) husband/wife, had you ever made love with anyone else?"; (2) "During your (former) marriage (did you) have you had any affairs or brief sexual encounters with another man/woman?" The table variables are thus gender (G), reported premarital (P) and extramarital (E) sex, and current marital status (M).

In this analysis we consider the variables in the order G, P, E, and M. That is, in the first stage we treat P as a response to G and examine the [Gender][Pre] mosaic to assess whether gender has an effect on premarital sex. In the second stage we treat E as a response to G and P jointly; we examine the mosaic for [Gender, Pre][Extra] for evidence that extramarital sex is related to either gender or premarital sex. Finally, the mosaic for [Gender, Pre, Extra][Marital] is examined for evidence of the dependence of marital status on the three previous variables jointly.

Each stage results in a fitted model for the corresponding marginal table. As noted in Section 3.5, these models are equivalent to the recursive logit models whose path diagram is $G \rightarrow P \rightarrow E \rightarrow M$. The G^2 values for these models (shown in Table 6) provide a decomposition of the G^2 for the model of complete independence fit to the full table.

The [Gender][Pre] mosaic is shown in Figure 13. G^2 for the model $[G][P]$ is 75.26 on 1 df, indicating that gender and reported premarital sex are highly associated. The mosaic shows that men are much more likely to report premarital sex than are women; the sample odds ratio is 3.7. We also see that women are about twice as prevalent as men in this sample.

For the second stage, the [Gender, Pre][Extra] mosaic is shown in Figure 14. G^2 for the model $[GP][E]$ is 48.93 on 3 df, indicating that extramarital sex is not independent of gender and premarital sex jointly. From the pattern of deviations in Figure 14, we see that men and women who have reported premarital sex are far more likely to report extramarital sex than are those who have not. From the marginal totals for the $[GP][E]$ table, the conditional odds ratio of extramarital sex is 3.61 for men and 3.56 for women. The pattern of deviations in the mosaic suggests the need for a $[PE]$ term in an explanatory model for extramarital sex, but a $[GE]$ term incorporating an association between gender and extramarital sex, given premarital sex, appears unnecessary.

The mosaic for the model [Gender, Pre, Extra][Marital] for the final stage is shown in Figure 15. G^2 for this model

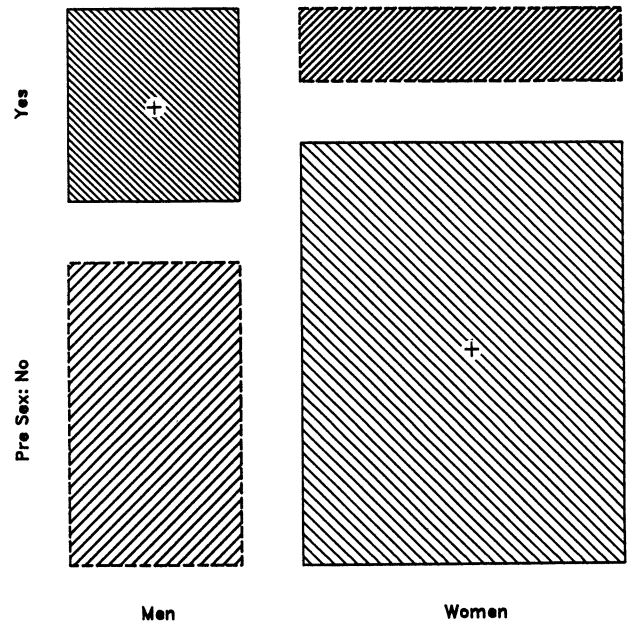

Figure 13. Mosaic Display for Gender and Premarital Sexual Experience. Residuals from the model [G][P] are shown by shading.

is 107.96 on 7 df, indicating that marital status depends strongly on gender, premarital sex, and extramarital sex jointly. The relationship displayed by the pattern of deviations in the mosaic is more complex than a single interaction. Among those reporting no premarital sex (the bottom part of Figure 15), there is a similar pattern of cell sizes and deviations for marital status in relation to gender and extramarital sex. Given that persons did not report premarital sexual experience, they are more likely to still be married if they did not report extramarital sex and more likely to be divorced if they did. Among those who do report premarital

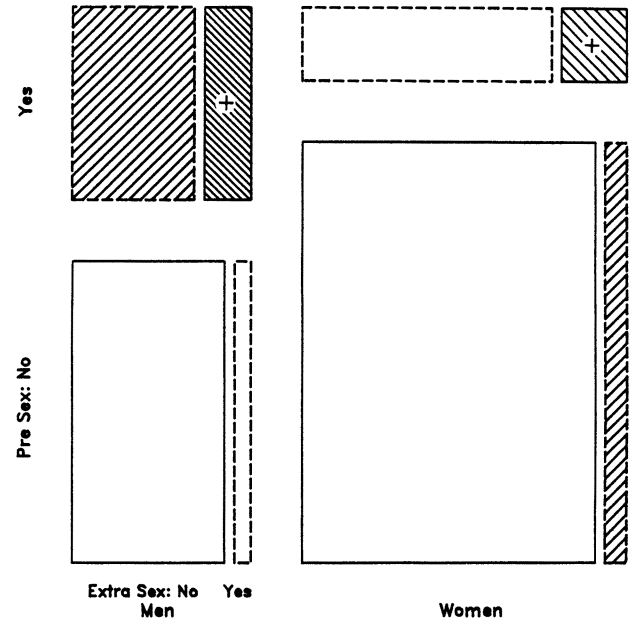

Figure 14. Mosaic Display for Gender, Premarital, and Extramarital Sexual Experience. Deviations from the model of joint independence, [GP][E] are shown.

sex (top part of Figure 15), there also a similar pattern of sign of deviations: positive for those who are divorced and negative for those who are married.

The four 2×2 blocks of the mosaic show the conditional relation of extramarital sex to marital status. Comparing these, we see that the odds ratios of divorce in relation to reported extramarital sex are considerably larger for men and women who also reported premarital sex. These observations imply the need to incorporate effects $[PM]$ and $[EM]$ of premarital and extramarital sex on marital status, and probably the interaction $[PEM]$ into an explanatory model.

Thus Figure 15 suggests the relationship between marital status and gender, premarital sex and extramarital sex can be explained better by adding the two-way associations $[PM]$ and $[EM]$ or the three-way term $[PEM]$ to the model. Because this stage considers marital status as a response to gender, premarital sex, and extramarital sex, we would normally fit the $[GPE]$ marginal table and consider the models $[GEP][PM][EM]$ or $[GPE][PEM]$ for the complete table.

The model $[GPE][PM][EM]$ does not fit particularly well, producing $G^2 = 18.16$ on 5 df ($p = .0028$). To see why, we display the residuals from this model in Figure 16. Only one cell has a standardized residual exceeding 2: There are more still-married men who reported both premarital sex and extramarital sex than the model predicts. The contribution to X^2 from this cell is $d^2 = 6.92$, which is not large enough to account for the lack of fit. Examining the signs of residuals in each of the four corner blocks in Figure 16, we see that the relationship of extramarital sex to marital status is opposite to each other in the NW and SE blocks and in the SW and NE blocks, suggesting an interaction between P and E in their effects on marital status; that is, the model $[GPE][PEM]$. This model does indeed fit quite well, $G^2 = 5.25$ with 4 df ($p = .26$).

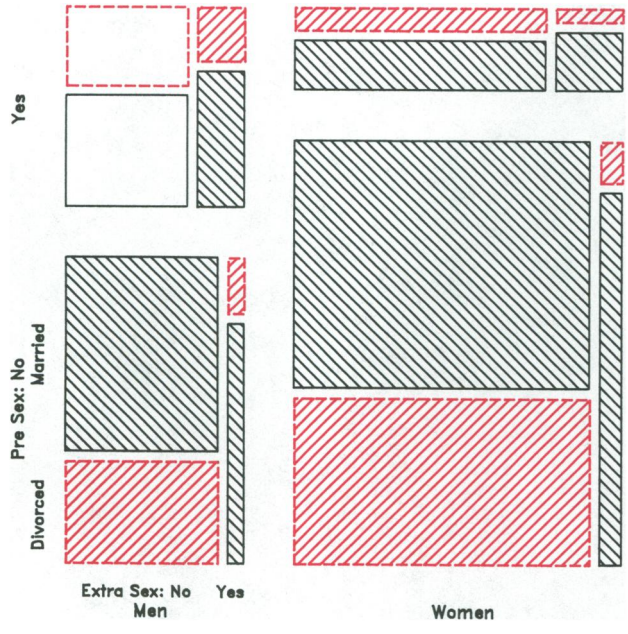

Figure 15. Four-Way Mosaic for Gender, Premarital sex, Extramarital sex, and Marital status. Deviations from the model of joint independence, [GPE][M], are shown by color and shading. The pattern of residuals suggests some terms to be included in an explanatory model.

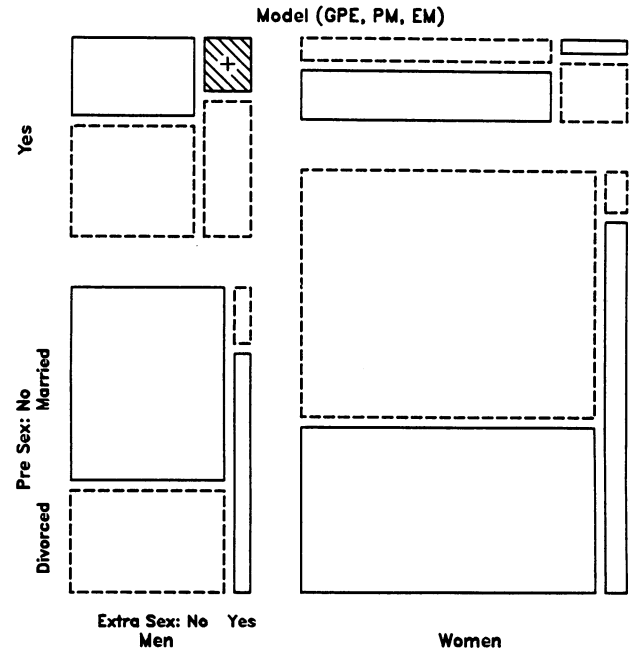

Figure 16. Four-Way Mosaic for Gender, Premarital sex, Extramarital sex, and Marital status. Shading shows residuals from the model [GPE][PM][EM]. Although only one cell has a standardized residual exceeding 2 in magnitude, the signs of residuals suggest the need for an additional term in the model.

The process of finding an acceptable model for these data clearly could be carried out numerically, by fitting all possible models, or using a method of forward or backward selection. For multi-dimensional tables with higher-order associations, the interpretation of the log-linear parameters for these associations is often difficult. The sequence of mosaic displays reveals the pattern of these associations as each variable is included. As we move from a baseline fit to an explanatory model, these associations are eliminated from the mosaic. Hence we can think of the process of finding an acceptable model as "cleaning the mosaic."

More generally, we regard the mosaic display as a natural and direct graphic adjunct to log-linear modeling for multiway contingency tables. Log-linear models can show *which* variables are associated, whereas mosaic displays reveal *how* those variables are related. The representation of the mosaic display is direct, because the size of each tile reflects cell frequency. The use of color and shading to represent sign and magnitude of residuals from a model is a natural way to portray the pattern of departure from the model and represents visually the information experienced analysts usually look for in tables of numbers.

[Received July 1991. Revised June 1992.]

REFERENCES

Agresti, A. (1990), *Categorical Data Analysis,* New York: Wiley-Interscience.

Bertin, J. (1983), *Semiology of Graphics* (trans. W. Berg), Madison, WI: University of Wisconsin Press.

Deming, W. E., and Stephan, F. F. (1940), "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known," *Annals of Mathematical Statistics,* 11, 427–444.

Fienberg, S. E. (1969), "Preliminary Graphical Analysis and Quasi-Independence for Two-Way Contingency Tables," *Applied Statistics,* 18, 153–168.
——— (1980), *The Analysis of Cross-Classified Categorical Data* (2nd ed.), Cambridge, MA: MIT Press.
Friendly, M. (1991), *SAS System for Statistical Graphics,* Cary, NC: SAS Institute.
——— (1992a), "Graphical Methods for Categorical Data," in *SAS SUGI Proceedings,* 17, 1367–1373.
——— (1992b), "User's Guide for MOSAICS: A SAS/IML Program for Mosaic Displays," Depart. of Psychology Reports, No. 206, York University.
Friendly, M., and Fox, J. (1991), "Interpreting Higher-Order Interactions in Log-Linear Analysis: A Picture is Worth 1,000 Numbers," Working Paper ISBN 1-55014-157-0, York University, Institute for Social Research.
Gilbert, G. N. (1981), *Modelling Society: An Introduction to Log-Linear Analysis for Social Researchers,* London: Allen & Unwin.
Goodman, L. A. (1970), "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classifications," *Journal of the American Statistical Association,* 65, 226–256.
——— (1971), "Partitioning of Chi-Square, Analysis of Marginal Contingency Tables, and Estimation of Expected Frequencies in Multidimensional Contingency Tables," *Journal of the American Statistical Association,* 66, 339–344.
——— (1973), "The Analysis of Multidimensional Contingency Tables When Some Variables are Posterior to Others: A Modified Path Analysis Approach," *Biometrika,* 60, 179–192.
——— (1979), "Simple Models for the Analysis of Association in Cross-Classifications Having Ordered Categories," *Journal of the American Statistical Association,* 74, 537–552.
Greenacre, M. (1984), *Theory and Applications of Correspondence Analysis,* London: Academic Press.
Greenacre, M., and Hastie, T. (1987), "The Geometric Interpretation of Correspondence Analysis," *Journal of the American Statistical Association,* 82, 437–447.
Haberman, S. J. (1973), "The Analysis of Residuals in Cross-Classified Tables," *Biometrics,* 29, 205–220.
——— (1978), *Analysis of Qualitative Data,* New York: Academic Press.
Hartigan, J. A., and Kleiner, B. (1981), "Mosaics for Contingency Tables," *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface,* ed. W. F. Eddy, New York: Springer-Verlag, pp. 268–273.
——— (1984), "A Mosaic of Television Ratings," *The American Statistician,* 38, 32–35.
Heijden, P. G. M. van der, and de Leeuw, J. (1985), "Correspondence Analysis Used Complementary to Log-Linear Analysis," *Psychometrika,* 50, 429–447.
Heuer, J. (1979), *Selbstmord bei Kinder und Jugendlichen* [Suicide by Children and Youth], Stuttgard: Ernst Klett Verlag.
SAS Institute (1989), *SAS/IML Software: Usage and Reference, Version 6, First Edition,* Cary, N.C.: SAS Institute.
Snee, R. D. (1974), "Graphical Display of Two-Way Contingency Tables," *The American Statistician,* 28, 9–12.
Thornes, B., and Collard, J., (1979), *Who Divorces?,* London: Routledge & Kegan.
Wang, C. M. (1985), "Applications and Computing of Mosaics," *Computational Statistics & Data Analysis,* 3, 89–97.