Extending Mosaic Displays: Marginal, Conditional, and Partial Views of Categorical Data*

Michael Friendly
York University

Abstract

This paper first illustrates the use of mosaic displays for the analysis of multiway contingency tables. We then introduce several extensions of mosaic displays designed to integrate graphical methods for categorical data with those used for quantitative data.

The scatterplot matrix shows all pairwise (bivariate marginal) views of a set of variables in a coherent display. One analog for categorical data is a matrix of mosaic displays showing some aspect of the bivariate relation between all pairs of variables. The simplest case shows the bivariate marginal relation for each pair of variables.

Another case shows the conditional relation between each pair, with all other variables partialled out. For quantitative data this represents (a) a visualization of the conditional independence relations studied by graphical models. and (b) a generalization of partial residual plots.

The conditioning plot, or *coplot* shows a collection of partial views of several quantitative variables, conditioned by the values of one or more other variables. A direct analog of the coplot for categorical data is an array of mosaic plots of the dependence among two or more variables, stratified by the values of one or more *given* variables. Each such panel then shows the *partial* associations among the foreground variables; the collection of such plots show how these associations change as the given variables vary.
Key words: categorical data, conditional independence, coplots, correspondence analysis, graphical models, log-linear models, scatterplot matrix, mosaic matrix

1 Introduction

Graphical methods for quantitative data and categorical data are often viewed as quite distinct, although the underlying linear models are close analogs. Here we develop some graphic connections and parallels between these methods, through extensions of the mosaic display, a general method for visualizing n-way contingency tables.

We first describe the design goals and visualization principles for the mosaic display, and illustrates its use for the analysis of several multiway contingency tables (Section 2). Second, we introduce several extensions of mosaic displays designed to integrate graphical methods for categorical data with those used for quantitative data. The mosaic matrix (Section 3) is an analog of the scatterplot matrix, showing all pairwise, bivariate views of an n-way table. A conditional mosaic matrix (Section 4), and its' analog for continuous data shows the conditional relation for each pair, with all other variables

*Paper presented at the Workshop on "Data Visualization in Statistics", July 6-10, 1998, held at Drew University. This work is supported by Grant 8150 from the National Sciences and Engineering Research Council of Canada.

controlled. Partial views (Section 5) are provided by mosaic displays stratified by one or more *given* variables, and are categorical analogs of coplot displays (Cleveland, 1993).

One essential difference between quantitative data and categorical data lies in the nature of the natural visual representation (Friendly, 1995, 1997). For quantitative data, magnitude can be represented by length (in a bar chart) or by position along a scale (dotplots, scatterplots). When the data are categorical, design principles of perception, detection, and comparison (Friendly, 1999) suggest that frequencies are most usefully represented as areas. In spite of the fact that (in magnitude estimation tasks) judgments of area are known to be less accurate than those of length (e.g., Cleveland and McGill (1984)), there are two fundamental reasons why area is a preferred visual representation for count data: (a) multiplicative relations of probabilities and expected frequencies translate readily into height and width of rectangles, whose area then depicts a cell value; (b) a concrete, physical model for categorical data (Friendly, 1995) based on count $\sim$ area yields a surprising range of correct, but novel interpretations for statistical principles (maximum likelihood), estimation techniques (iterative proportional fitting, Newton-Raphson) and phenomena (power, why components of likelihood-ratio G^2 can be negative).

One final introductory point: the graphics shown here are, of necessity, static graphs, designed to show both the data and some model-based analysis. Their ultimate use will, I believe, be most productive as interactive graphics, tightly coupled with the model-building methods themselves. One needs to design good widgets first, however, before learning how to employ them most effectively. A simple interactive web applet may be found at `http://www.math.yorku.ca/SCS/Online/mosaics/`, which also contains pointers to other implementations, both static and dynamic.

2 Mosaic displays

The mosaic display (Friendly, 1992a, 1994, 1997, 1999, Hartigan and Kleiner, 1981, 1984) is a graphical method for visualizing an n-way contingency table and for building models to account for the associations among its variables. The frequencies in a contingency table are portrayed as a collection of rectangular "tiles" whose areas are proportional to the cell frequencies; the areas are colored and shaded to portray the residuals from a specified log-linear model. Whereas goodness-of-fit statistics provide an overall summary of how well a model fits the data, the mosaic display reveals the pattern of lack of fit, and helps suggest an alternative model that may fit better.

The construction of the mosaic is easily understood as a straightforward application of conditional probabilities. For a two-way table, with cell frequencies n_{ij}, and cell probabilities $p_{ij} = n_{ij}/n_{++}$, a unit square is first divided into rectangles whose width is proportional to the observed marginal frequencies n_{i+}, and hence to the marginal probabilities $p_i = n_{i+}/n_{++}$. Each such rectangle is then subdivided horizontally in proportion to the conditional probabilities of the second variable given the first, $p_{j|i} = n_{ij}/n_{i+}$. Hence the area of each tile is proportional to the observed cell frequency and probability,

$$p_{ij} = p_i \times p_{j|i} = \left(\frac{n_{i+}}{n_{++}}\right) \times \left(\frac{n_{ij}}{n_{i+}}\right) \tag{1}$$

The order of conditioning matters, of course. In static graphs, placing explanatory variable(s) first shows how the response(s) depend on them. In interactive, multi-windowed systems such as ViSta (Young, 1994) it is easy to provide both views of a two-way table or allow the order of variables to be chosen interactively.

For example, Table 1 shows data on the relation between hair color and eye color among 592

Table 1: Hair-color eye-color data

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

subjects (students in a statistics course) collected by Snee (1974). The Pearson χ^2 for these data is 138.3 with 9 df, indicating substantial departure from independence.

The basic two-way mosaic for these data, shown in the left panel of Figure 1, is then similar to a divided bar chart. If hair color and eye color were independent, $p_{ij} = p_i \times p_j$, and then the tiles in each row would all align. This is shown in the right panel of Figure 1, which displays the expected frequencies, $m_{ij} = n_{i+}n_{+j}/n_{++}$, under independence.

Figure 1: Basic mosaic display for hair-color and eye color data. The area of each rectangle is proportional to the cell frequency. Left: observed frequencies; right: expected frequencies under independence.

2.1 Design goals and visualization principles

One important design goal for visualization methods for categorical data is to serve various needs in the analysis of contingency tables (Friendly, 1999):

- ***reconnaissance***—a preliminary examination, or an overview of a possibly complex terrain;
- ***exploration***—help detect patterns or unusual circumstances, or to suggest hypotheses;
- ***model building & diagnosis***—critique a fitted model as a reasonable statistical summary.

Enhancements to the basic mosaic designed to meet these needs are described below.

2.1.1 Enhanced mosaics

The enhanced mosaic display (Friendly, 1992a, 1994) achieves greater visual impact by using color and shading to reflect the size of the residual from independence and by reordering rows and columns to make the pattern of association more coherent. The resulting display serves exploratory goals (by showing the pattern of observed frequencies in the full table, or any marginal subtable), and model building goals (by displaying the residuals from a given log-linear model).

Figure 2 gives the extended mosaic plot, showing the standardized (Pearson) residual from independence, $d_{ij} = (n_{ij} - m_{ij})/\sqrt{m_{ij}}$ by the color and shading of each rectangle: cells with positive residuals are outlined with solid lines and filled with slanted lines; negative residuals are outlined with broken lines and filled with grayscale. The absolute value of the residual is portrayed by shading density: cells with absolute values less than 2 are empty; cells with $|d_{ij}| \geq 2$ are filled; those with $|d_{ij}| \geq 4$ are filled with a darker pattern. (In color, it is visually more effective to use two opposing hues, blue and red, say, with varying lightness to portray magnitude.) 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. For exploratory purposes, we do not usually make adjustments (e.g., Bonferroni) for multiple tests because the goal is to display the pattern of residuals in the table as a whole.

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 association. For example, in Figure 2, the eye color categories have been permuted so that the residuals from independence have an opposite-corner pattern, with positive values running from bottom-left to top-right corners, negative values along the opposite diagonal.

Coupled with size and shading of the tiles, the excess in the black-brown and blond-blue cells, together with the under-representation of brown-haired blonds and people with black hair and blue eyes is now quite apparent. Though the table was reordered based on the d_{ij} values, both dimensions in Figure 2 are ordered from dark to light, suggesting an explanation for the association. A general method (Friendly, 1994) is to sort the categories by their scores on the largest dimension in a (correspondence analysis) singular value decomposition of residuals. This reordering of categories illustrates the principle of ***effect-ordering*** for data displays (see Friendly (1999))—sort the data by the effects to be observed, here, the *structure* of association. We now observe that one cell (hazel-eyed redheads) departs from the opposite-corner pattern, suggesting that this combination differs in some way from the light-dark association.

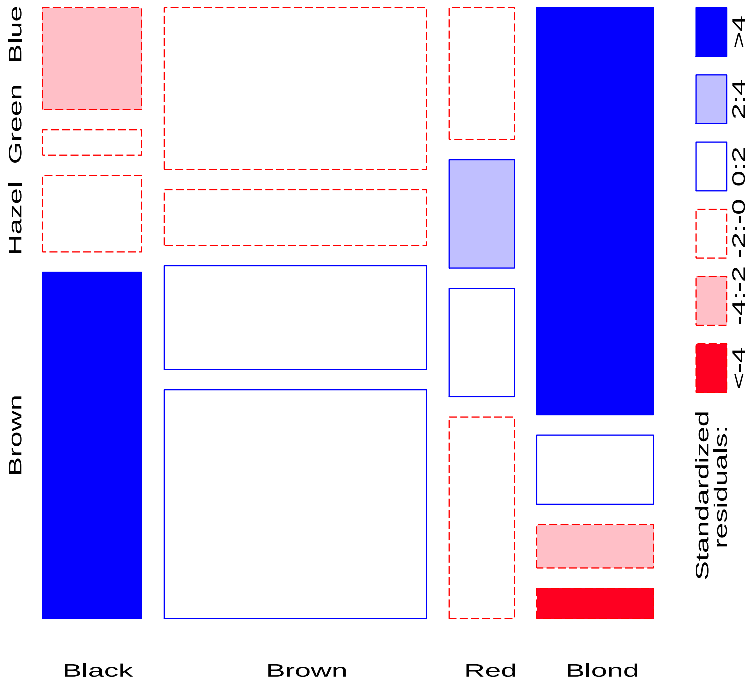


Figure 2: Extended mosaic, reordered and shaded. The two levels of shading density correspond to standardized residuals greater than or equal to 2 and 4 in absolute value.

2.1.2 n-way tables

Another design goal is that graphical methods extend naturally to three-way and higher-way tables, in much the same way that graphical methods for quantitative data do. For an n-way table, with variables $A, B, C, \ldots$, the construction of the mosaic generalizes recursively to

$$p_{ijk\ell\cdots} = \underbrace{\overbrace{p_i \times p_{j|i}}^{\{AB\}} \times p_{k|ij}}_{\{ABC\}} \times p_{\ell|ijk} \times \cdots \tag{2}$$

The braces in Eqn. (2) are meant to suggest that the first two terms provide a mosaic for the marginal frequencies of variables A and B, the first three terms give a mosaic for the $\{ABC\}$ marginal table, and so forth, up to the display of the full n-way table.

For example, 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 ethnicity, for example. Then each rectangle can be subdivided horizontally to show the proportion of males and females in that cell, and each of those horizontal portions can be subdivided vertically to show the proportions of people of each ethnicity in the hair-eye-sex group.

Figure 3 shows the mosaic for the three-way table, with hair and eye color groups divided according to the proportions of Males and Females: We see that there is no systematic association between sex and the combinations of hair and eye color—except among blue-eyed blonds, where there are an overabundance of females. (Do they have more fun?)

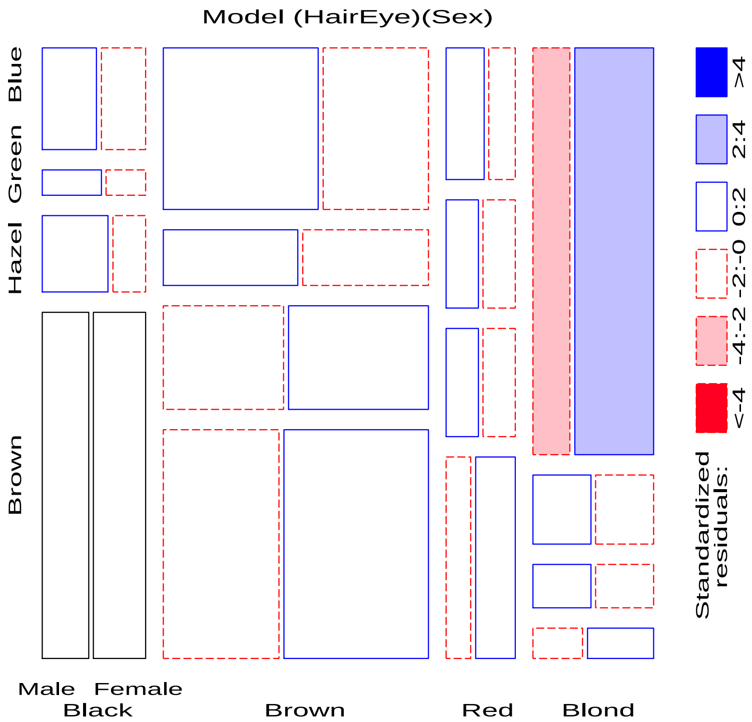


Figure 3: Three-way mosaic display for hair color, eye color, and sex. The categories of sex are crossed with those of hair color, but only the first occurrence is labeled. Residuals from the model of joint independence, $[HE]\,[S]$ are shown by shading. The only lack of fit is an overabundance of females among blue-eyed blonds.

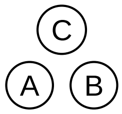
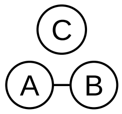
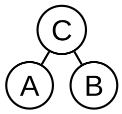
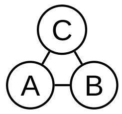
2.2 Fitting models

When three or more variables are represented in the mosaic, we can fit different models and display the residuals from each. We treat these as null or baseline models, which may not fit the data particularly well. The deviations of observed frequencies from expected ones, displayed by shading, will often suggest terms to be added to an explanatory model that achieves a better fit.

For a three-way table, with variables A, B and C, some of the possible model types are described below and summarized in Table 2; permutation of the variable letters give other model instances. We use $[\,]$ notation to list the high-order terms in a hierarchical log-linear model; these correspond to the margins of the table which are fitted exactly. Any other associations present in the data will appear in the pattern of residuals. Here, $A \perp B$ is read, "A is independent of B", and π_{ijk} refers to theoretical probabilities. Table 2 also depicts the relations among variables as an association graph, where associated variables are connected by an edge.

H_1**: Mutual independence.** The model of mutual independence, $A \perp B \perp C$, asserts that all joint probabilities π_{ijk} are products of the one-way marginal probabilities: $\pi_{ijk} = \pi_{i++}\,\pi_{+j+}\,\pi_{++k}$. This corresponds to the log-linear model $[A]\,[B]\,[C]$. Fitting this model leaves all higher terms, and hence *all* association among the variables, in the residuals, which are displayed by shading in the mosaic.

Table 2: Fitted margins, model symbols and interpretations for some hypotheses for a three-way table

Hypothesis	Fitted margins	Model symbol	Independence Interpretation	Association graph
H_1	$n_{i++}, n_{+j+}, n_{++k}$	[A][B][C]	$A \perp B \perp C$	
H_2	n_{ij+}, n_{++k}	[AB][C]	$A, B \perp C$	
H_3	n_{i+k}, n_{+jk}	[AC][BC]	$A \perp B \mid C$	
H_4	$n_{ij+}, n_{i+k}, n_{+jk}$	[AB][AC][BC]	-	

H_2**: Joint independence.** The model in which variable C is jointly independent of variables A and B, ($A, B \perp C$), has $\pi_{ijk} = \pi_{ij+}\, \pi_{++k}$, and corresponds to 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, since that association is fitted exactly.

H_3**: Conditional independence.** Two variables, say A and B, are conditionally independent given the third (C) if A and B are independent when we control for C, symbolized as $A \perp B \mid C$. This means that conditional probabilities, $\pi_{ij|k}$, obey $\pi_{ij|k} = \pi_{i+|k}\, \pi_{+j|k}$. The corresponding log-linear models is denoted $[AC]\,[BC]$. When this model is fit, the mosaic shows the conditional associations between variables A and B, controlling for C, but does not show the associations between A and C, or B and C.

H_4**: No three-way interaction.** For this model, no pair is marginally or conditionally independent, so there is no independence interpretation. However, the partial association between any two variables is the same at each level of the third variable. The corresponding log-linear model formula is $[AB]\,[AC]\,[BC]$, indicating that all two-way margins are fit exactly and so are not shown in the residuals. Only a possible three-way association appears in the mosaic.

For example, with the data from Table 1 broken down by sex, fitting the joint-independence model [HairEye][Sex] allows us to see the extent to which the joint distribution of hair-color and eye-color is associated with sex. For this model, the likelihood-ratio G^2 is 19.86 on 15 *df* ($p = .178$), indicating an acceptable overall fit. The three-way mosaic for this model was shown in Figure 3. Any other model fit to this table will have the same tiles in the mosaic since the areas depend on the observed frequencies; the residuals, and hence the shading of the tiles will differ.

2.3 Sequential plots and models

The mosaic display is constructed in stages, with the variables listed in a given order. At each stage, the procedure fits a (sub)model to the marginal subtable defined by summing over all variables not yet entered. For example for a three-way table, $\{ABC\}$, the marginal subtables $\{A\}$ and $\{AB\}$ are calculated in the process of constructing the three-way mosaic. The $\{A\}$ marginal table can be fit to a model where the categories of variable A are equiprobable (or some other discrete distribution); the independence model can be fit to the $\{AB\}$ subtable, and so forth. The series of plots can give greater insight into the relationships among all the variables than a single plot alone.

Moreover, the series of mosaic plots fitting submodels of joint independence to the marginal subtables have the special property that they can be viewed as partitioning the hypothesis of mutual independence in the full table (Friendly, 1994, Goodman, 1970). For example, for the hair-eye data, the mosaic displays for the [Hair] [Eye] marginal table (Figure 2) and the [HairEye] [Sex] (Figure 3) table can be viewed as representing the partition

Model	df	G^2
[Hair] [Eye]	9	146.44
[Hair, Eye] [Sex]	15	19.86
[Hair] [Eye] [Sex]	24	155.20

This partitioning scheme for sequential models of joint independence extends directly to higher-way tables. The MOSAICS program Friendly (1992b) implements a variety of schemes for fitting a sequential series of submodels, including mutual independence, joint independence, conditional independence, partial independence and Markov chain models. Two examples illustrate the visual comparison of models and sequential displays.

2.4 Example: Berkeley admissions

Bickel et al. (1975) analyzed data (Freedman et al., 1978, p. 14) on applicants to graduate school at Berkeley for the six largest departments in 1973 classified by admission and gender. At issue was whether the data show evidence of sex bias in admission practices. The aggregate data, across departments, are shown in Figure 4. Of the 4526 applicants, 2691 (59.5%) were male; among males, 1198 (44.5%) were admitted, while among females 557 (30.4%) were admitted. The residuals show a strong association between gender and admission ($G^2(1) = 93.45$); the sample odds ratio, Odds (Admit|Male) / (Admit|Female) is 1.84, indicating that males were almost twice as likely to be admitted. Is this evidence for gender bias?

To collapse over departments, we must assume that men and women apply in roughly the same proportions to all departments. Figure 5 shows three-way mosaics for two models. Treating gender and department as explanatory variables and admission as the response, the model [Dept Gender] [Admit] asserts that admission is independent of department and gender. This baseline model fits poorly ($G^2(11) = 877.06$) as shown in the left panel of Figure 5. The residuals suggest that more men are accepted in departments A and B, while more women are accepted in departments E and F than would be the case if admission depended on neither department nor gender.

The right panel shows the same observed frequencies, but fits the model of conditional independence, [Admit Dept] [Gender Dept], for which admission is independent of gender, given department. This model also fits poorly ($G^2(6) = 21.74$), but the residuals in the mosaic suggest that the lack of fit is due primarily to department A, where a greater proportion of women are admitted than men. We

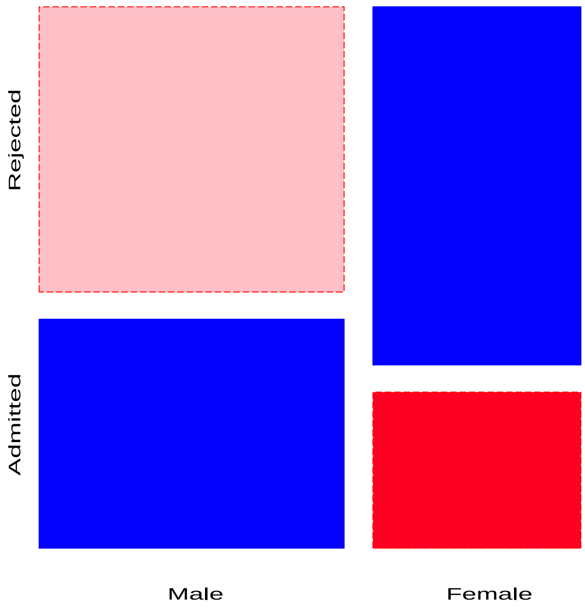


Figure 4: Mosaic display for Berkeley admissions: Evidence for sex bias?

return to these paradoxical findings—strong evidence of gender bias in favor of men in the marginal view (Figure 4), but none (except for department A) in the conditional view—in Section 3.3.

2.5 Example: Survival on the *Titanic*

There have been few marine disasters resulting in the staggering loss of life which occurred in the sinking of the *Titanic* on April 15, 1912 and (perhaps as a result) few that are so widely known by the public. It is surprising, therefore, that neither the exact death toll from this disaster nor the distributions of death among the passengers and crew are universally agreed. Dawson (1995, Table 2) presents the cross-classification of 2201 passengers and crew on the *Titanic* by Age, Gender, Class (1st, 2nd, 3rd, Crew) shown in Table 3 and describes his efforts to reconcile various historical sources. Let us see what we can learn from this data set.

Examining the series of mosaics for the variables ordered Class, Gender, Age, Survival will show the relationships among the background variables and how these are related to survival. The letters C, G, A, S respectively are used to refer to these variables below.

Figure 6 shows the two-way and three-way plots among the background variables. The two-way mosaic shows that the proportion of males decreases with increasing economic class, and that the crew was almost entirely male. The three-way plot shows the distribution of adults and children among the Class-Gender groups. The residuals display the fit of a model in which Age is jointly independent of the the Class-Gender categories. Note that there were no children among the crew, and the overall proportion of children was quite small (about 5 %). Among the passengers, the proportion of children is smallest in first class, largest in third class. The only large positive residuals correspond to a greater number of children among the 3rd class passengers, perhaps representing families traveling or immigrating together.

Two four-way mosaics are shown in Figure 7. The first fits the model $[CGA][S]$ which asserts that survival is independent of Class, Gender and Age jointly. This is the minimal null model when the

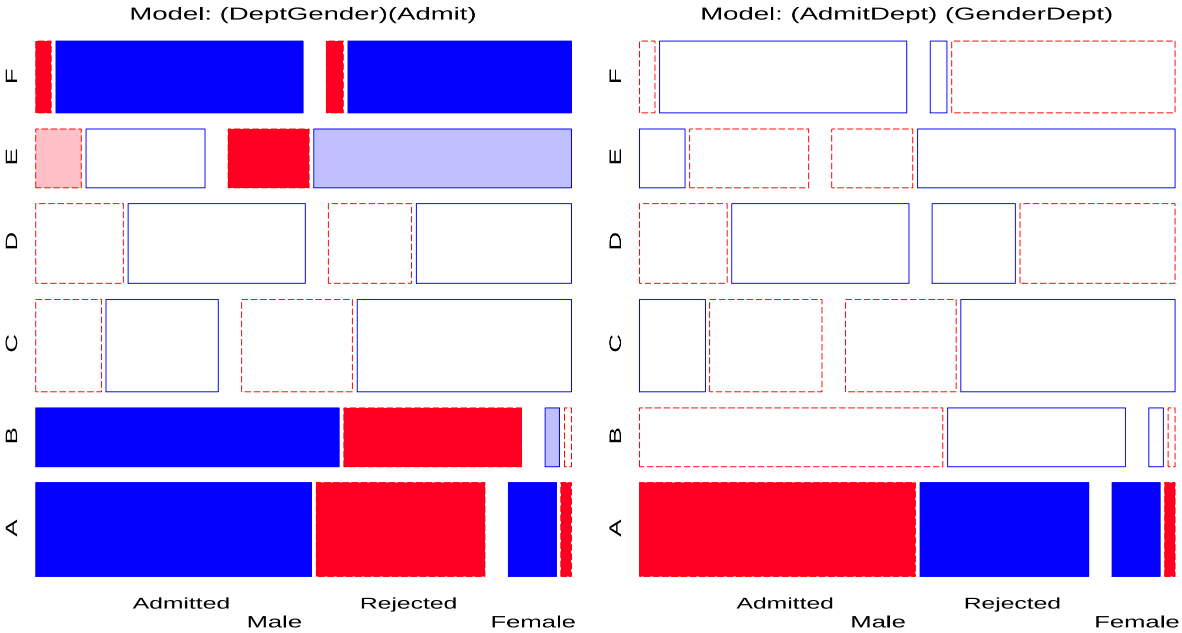


Figure 5: Three-way mosaic plots for Berkeley data. Left: Joint independence; right: Conditional independence

first three variables are explanatory. It is clear that greater proportions of women survived than men in all classes, but with greater proportions of women surviving in the upper two classes. Among males, the proportion who survived also increases with economic class. However, this model fits very poorly ($G^2(15) = 671.96$), and we may try to fit a more adequate model by adding associations between survival and the explanatory variables.

Adding a main effect of each of Class, Gender and Age on Survival amounts to fitting the model $[CGA][CS][GS][AS]$. That is, each of the three variables is associated with survival, but have independent, additive effects. The mosaic for this model is shown in Figure 7(b). The fit of this model is much improved ($\Delta G^2(5) = 559.4$), but still does not represent an adequate fit ($G^2(10) = 112.56$). There are obviously interactions among Class, Gender and Age on their impact on survival, some of which we have already noted.

Noting the rubric of "women and children first", we next fit the model $[CGA][CS][GAS]$ in which Age and Gender interact in their influence on survival (Figure 8(a)). Adding the association of Age and Gender with survival has improved the model slightly, however the fit is still not good ($G^2(9) = 94.54$). If we add the interaction of Class and Gender to this (the model $[CGA][CGS][GAS]$), the likelihood-ratio chi-square is reduced substantially ($G^2(6) = 37.26$), but the lack of fit is still significant.

Finally, we try a model in which Class interacts with both Age and Gender to give the model $[CGA][CGS][CAS]$, whose residuals are shown in Figure 8(b). The likelihood-ratio chi-square is now 1.69 with 4 df, a very good fit, indeed.

The import of these figures is clear. Regardless of Age and Gender, lower economic status was associated with increased mortality; the differences due to Class were moderated, however, by both Age and Gender. Although women on the *Titanic* were more likely overall to survive than men, the interaction of Class and Gender shows that women in 3rd class did not have a significant advantage,

Table 3: Survival on the Titanic

			Class			
Gender	**Age**	**Survived**	1st	2nd	3rd	Crew
Male	Adult	Died	118	154	387	670
Female			4	13	89	3
Male	Child		0	0	35	0
Female			0	0	17	0
Male	Adult	Survived	57	14	75	192
Female			140	80	76	20
Male	Child		5	11	13	0
Female			1	13	14	0

while men in 1st class did compared to men in other classes. The interaction of Class and Age is explained by the observation that while no children in 1st or 2nd class died, nearly two-thirds in 3rd class died; for adults, mortality increases progressively as economic class declines. Hence, although the phrase "women and children first" is mellifluous and appeals to a sense of Edwardian chivalry a more adequate description might be "women and children (according to class), then 1st class men".

3 Mosaic matrices for categorical data

One reason for the wide usefulness of graphs of quantitative data has been the development of effective, general techniques for dealing with high-dimensional datasets. The scatterplot matrix shows all pairwise (marginal) views of a set of variables in a coherent display, whose design goal is to show the interdependence among the collection of variables as a whole, and which allows detection of patterns which could not readily be discerned from a series of separate graphs. In effect, a multivariate data set in p dimensions (variables) is shown as a collection of $p(p-1)$ two-dimensional scatterplots, each of which is the projection of the cloud of points on two of the variable axes. For multivariate normal data, this gives a visualization of the covariance matrix. These ideas can be readily extended to categorical data.

A multiway contingency table of p categorical variables, $A, B, C, \ldots$, also contains the interdependence among the collection of variables as a whole. The saturated log-linear model, $[ABC\ldots]$ fits this interdependence perfectly, but is often too complex to describe or understand. By summing the table over all variables except two, A and B, say, we obtain a two-variable (marginal) table, showing the bivariate relationship between A and B, which is also a projection of the p-variable relation into the space of two (categorical) variables. If we do this for all $p(p-1)$ unordered pairs of categorical variables and display each two-variable table as a mosaic, we have a categorical analog of the scatterplot matrix, called a ***mosaic matrix***. Like the scatterplot matrix, the mosaic matrix can accommodate any number of variables in principle, but in practice is limited by the resolution of our display to three or four variables.

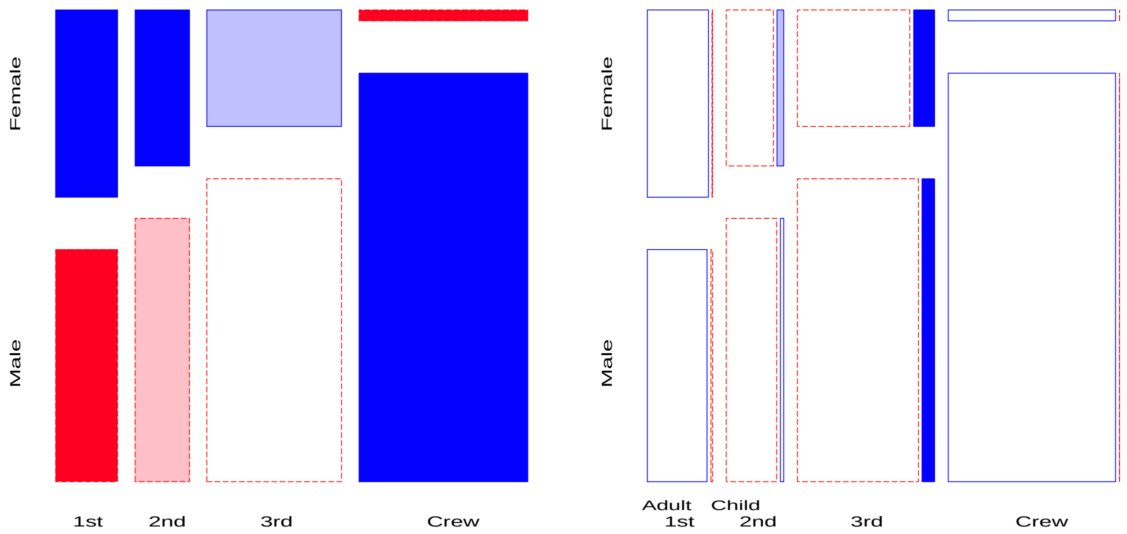


Figure 6: Titanic data, explanatory variables. (a) Class and Gender; (b) Class, Gender, Age. The levels of Age (Adult, Child) repeat for each Class.

3.1 MCA and the Burt matrix

The mosaic matrix has another interpretation as a direct visualization of the so-called "Burt matrix" which forms the basis of multiple correspondence analysis (MCA). A p-way, $J_1 \times J_2 \times \cdots \times J_p$ contingency table of $K = \prod J_i$ cells can be represented in a vector of frequencies $\boldsymbol{n} = (n_1, \ldots, n_K)^\mathsf{T}$ and a $K \times p$ matrix $\boldsymbol{X}$ whose i^{th} column gives the factor levels for variable i in each cell of the table. Let $\boldsymbol{Z}_i$ be the $K \times J_i$ indicator (design) matrix corresponding to $\boldsymbol{x}_i$, so that $\boldsymbol{Z}_i(k, \ell) = 1 \iff x_{ki} = \ell$, and let $\boldsymbol{Z}$ be the $K \times \sum^p J_i$ partitioned matrix $[\boldsymbol{Z}_1 \,|\, \boldsymbol{Z}_2 \,|\, \ldots \,|\, \boldsymbol{Z}_p]$.

Then the Burt matrix is the symmetric partitioned matrix

$$\boldsymbol{B} = \boldsymbol{Z}^\mathsf{T} \mathrm{diag}(\boldsymbol{n}) \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{N}_{[1]} & \boldsymbol{N}_{[12]} & \cdots \\ \boldsymbol{N}_{[21]} & \boldsymbol{N}_{[2]} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where each diagonal block, $\boldsymbol{N}_{[i]}$, is a diagonal matrix of the one-way marginal frequencies of variable i and each off-diagonal block $\boldsymbol{N}_{[ij]} = \boldsymbol{Z}_i^\mathsf{T} \mathrm{diag}(\boldsymbol{n}) \boldsymbol{Z}_j$ is the two-way marginal contingency table for variables i and j, with its transpose in $\boldsymbol{N}_{[ji]}$. Classical MCA (see, e.g., Greenacre (1984)) can be defined as an ordinary correspondence analysis (a singular value decomposition) of the matrix $\boldsymbol{B}$ which produces scores for the categories of all variables so that the greatest proportion of the pairwise associations in all off-diagonal blocks is accounted for in a small number of dimensions. The mosaic matrix of these two-way margins thus provides a visual representation of the Burt matrix, and the total amount of shading in all the individual mosaics portrays the total pairwise associations decomposed by MCA. (The representation would be complete if the one-way margins where drawn in the diagonal cells.)

To what extent do these displays portray the total association between variables? That question is easy to answer for the mosaic matrix: For an n-way contingency table, the total associa-

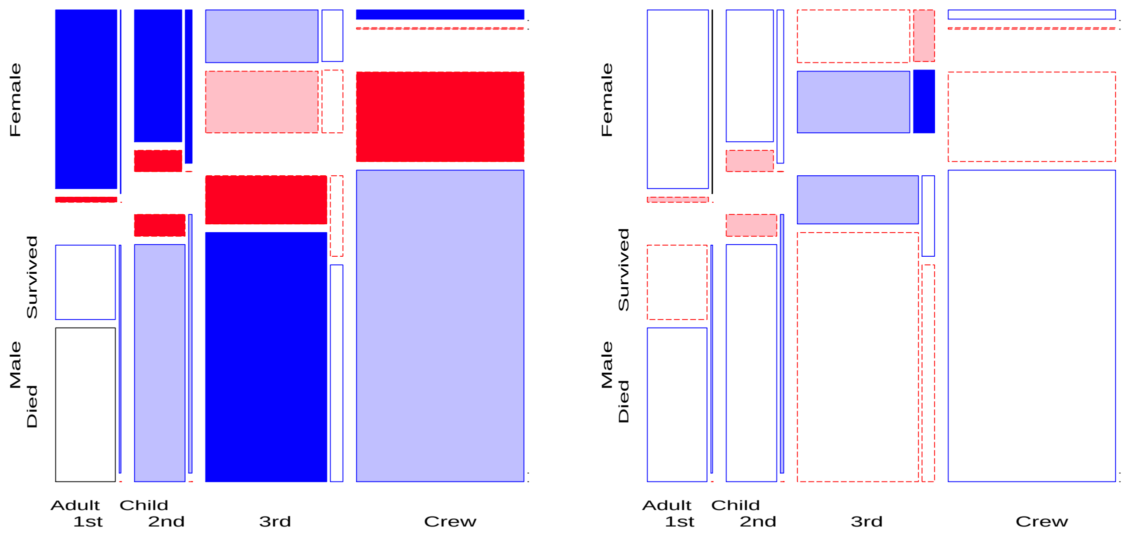


Figure 7: Titanic data, Class, Gender, Age, and Survival: (a) Joint independence; (b) Main effects of Age, Gender and Class on Survival.

tion among all variables is just the lack-of-fit statistic (Pearson χ^2 or likelihood-ratio G^2) $\chi^2_{[1]}$ for the model of mutual independence, $[A][B][C]\ldots$. The bivariate, marginal mosaic matrix shows (by residual shading) all pairwise associations, corresponding to the all-two-way loglinear model $[AB][AC][AD]\ldots[BC][BD]\ldots[CD]\ldots$, and so, the association *not* displayed is just the lack-of-fit statistic, $\chi^2_{[2]}$, for this model. A pseudo R^2 measure of the extent to which the bivariate mosaic captures all associations is then $1 - \chi^2_{[1]}/\chi^2_{[2]}$.

For MCA, the answer is more complicated, because the eigenvalues of the Burt matrix depend on the one-way marginal frequencies (which are not relevant to questions of association) as well as the two-way cross-tables. We simply note here that (a) classical MCA underestimates the degree of association captured in a given number of dimensions, and (b) the method of joint correspondence analysis Greenacre (1988, 1997) provides a χ^2decomposition commensurable with that of the mosaic matrix.

3.2 Example: Survival on the *Titanic*

Figure 9 shows the mosaic matrix for the bivariate relations in the *Titanic* data. The bottom row and the rightmost column show the associations between each of the background variables and Survival collapsing over other variables. There are strong associations of all three variables, but particularly for Gender (females more likely to have survived overall) and for Class ("1st" most likely to have survived overall). Other off-diagonal panels show the pairwise associations among the background variables. The panel in row 3, column 1 (numbered from the upper left corner, as in a table) is the bivariate relation between Class and Gender, shown earlier in Figure 6. The panels in row 2 show that very few children sailed on the *Titanic*, and that most were in 3rd class, and female.

The mosaic matrix in Figure 9 may be compared with the the results of an MCA analysis of the *Titanic* data. Figure 10 shows the 2-dimensional solution (using the typical chi-squared distance

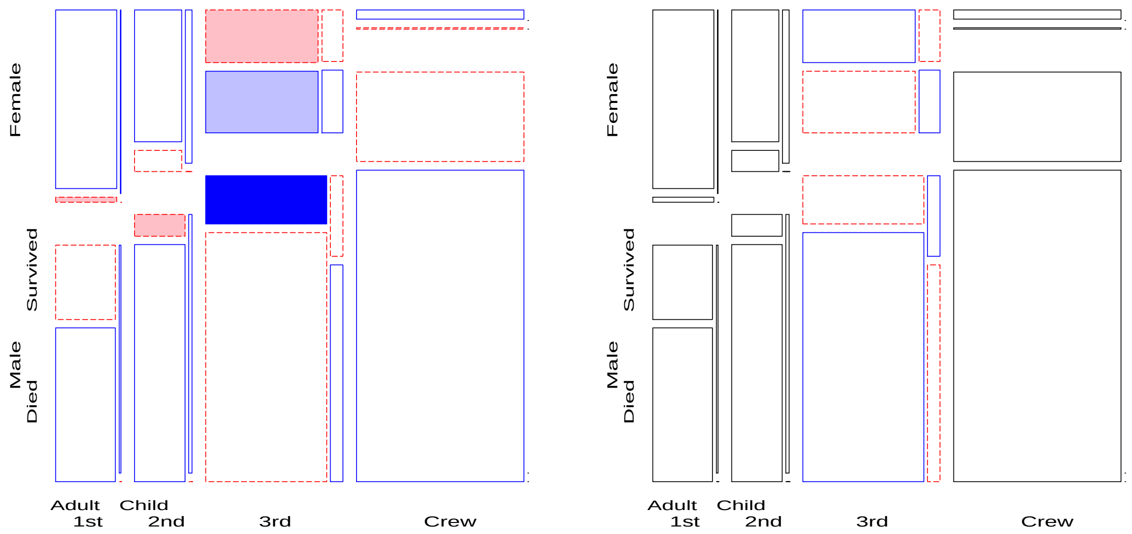


Figure 8: Titanic data, interaction models. (a) Model $[CGA][CS][GAS]$: Age*Gender on Survival; (b) Model $[CGA][CGS][CAS]$: Age*Gender + Class*Gender on Survival

scaling). The positions of the category points for all factors accounts for 50% of the total association ($\chi^2(81) = 15533.4$), representing all pairwise interactions among the four factors. The points for each factor have the property that the sum of coordinates on each dimension, weighted inversely by the marginal proportions, equals zero, so that high frequency categories (e.g., Adult) are close to the origin. The first dimension is perfectly aligned with the Gender factor, and also strongly aligned with Survival. The second dimension pertains mainly to Class and Age effects. Considering those points which differ from the origin most similarly (in distance and direction) to the point for Survived, gives the interpretation that survival was associated with being female or upper class or (to a lesser degree) being a child.

The mosaic matrix, although more complex, captures all of the pairwise associations; the all two-way model in turn reflects 91% of all association (Table 4). The MCA plot, however, shows only 50% in two dimensions. (A third dimension would account for an additional 17% here.) Most importantly, the pairwise associations are shown explicitly in the mosaic matrix, while they must be inferred from the positions of category points in the MCA plot.

Table 4: Lack-of-fit statistics for some models for the *Titanic* data

Model	Terms	df	G^2	R^2
Mutual independence	[C][G][A][S]	25	1243.663	
Survival $\perp$ Class, Gender, Age	[CGA][S]	15	671.962	0.460
All two-way	[CG][CA][CS][GA][GS][AS]	13	116.588	0.906
All thre-way	[CGA][CGS][CAS][GAS]	3	0.001	0.999

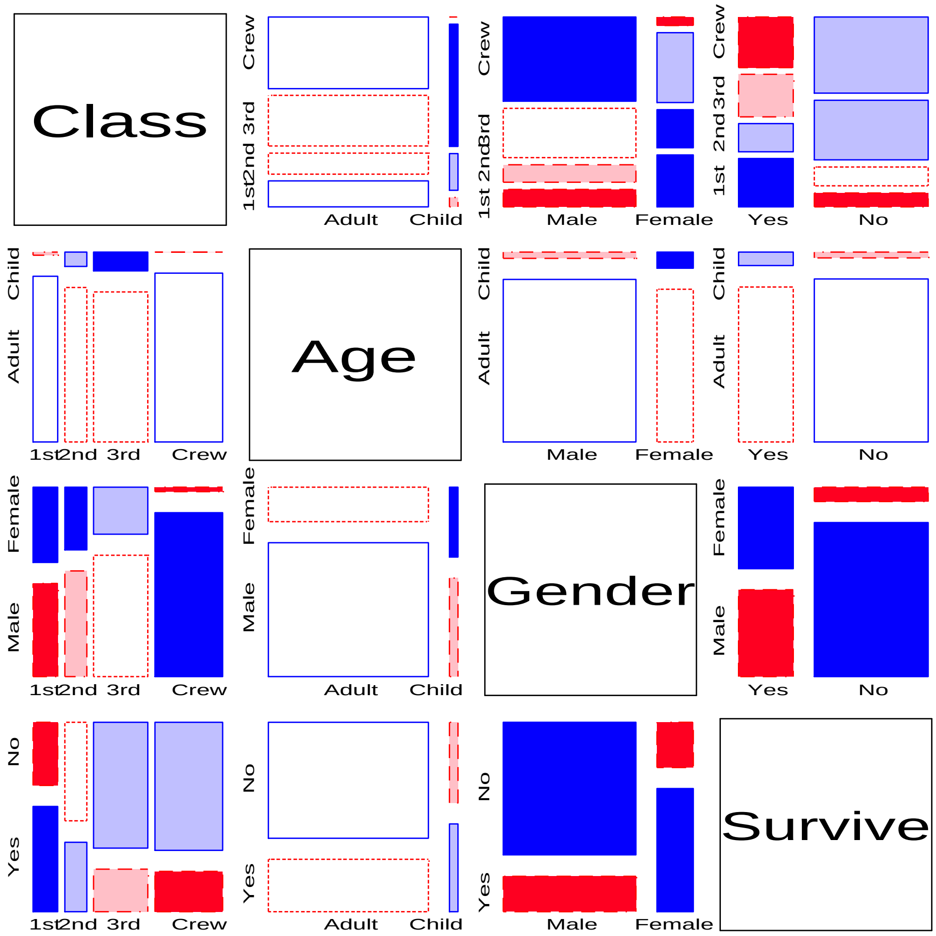


Figure 9: Mosaic matrix of *Titanic* data. Each panel shows the marginal relation, fitting an independence model between the row and column variable, collapsed over other variable(s).

3.3 Example: Berkeley admissions

Figure 11 shows the pairwise marginal relations among the variables Admit, Gender and Department in the Berkeley data which were examined earlier (Figure 4 and Figure 5). The panel in row 2, column 1 shows that Admission and Gender are strongly associated marginally, as we saw in Figure 4, and overall, males are more often admitted. The diagonally-opposite panel (row 1, column 2) shows the same relation, splitting first by gender.(Note that this is different than just the transpose or interchange of horizontal and vertical dimensions as in the scatterplot matrix, because the mosaic display splits the total frequency first by the horizontal variable and then (conditionally) by the vertical variable. The areas of all corresponding tiles are the same in each diagonally opposite pair, however, as are the residuals shown by color and shading.)

The panels in the third column (and third row) illuminate the explanation for the paradoxical result (see Figure 5) that, within all but department A, the likelihood of admission is nearly equal for men and women, yet, overall, there appears to be a bias in favor of admitting men (see Figure 4) The (1,3) and (3, 1) panels shows the marginal relation between Admission and Department; departments A and B have the greatest overall admission rate, departments E and F the least. The (2, 3) panel shows that

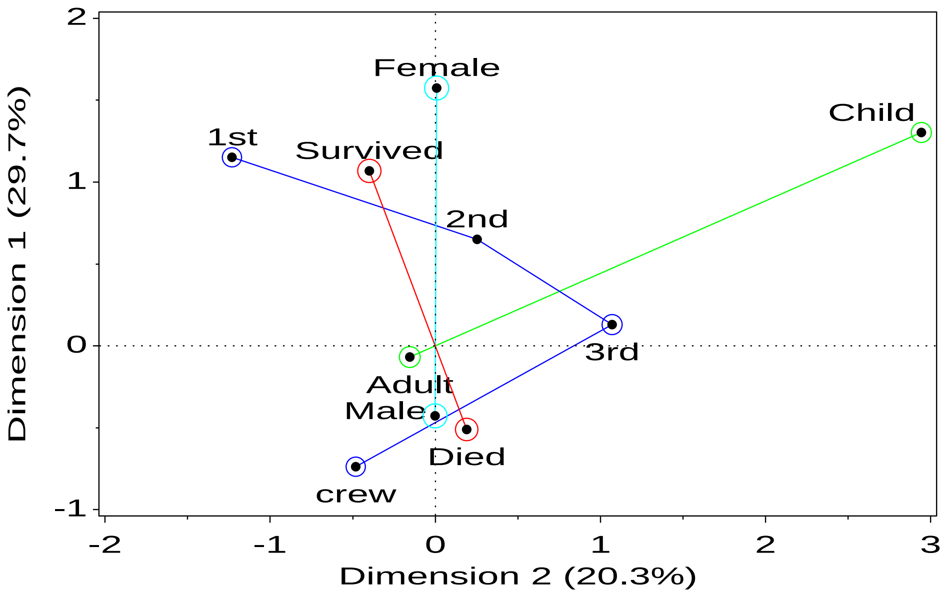


Figure 10: Titanic data: MCA analysis

men apply in much greater numbers to departments A and B, while women apply in greater numbers to the departments with the lowest overall rate of admission. (This explanation ignores the possibility of structural bias, that is, differential resources allocated to departments to which women predominantly apply.)

4 Conditional views of categorical and quantitative data

Several further extensions are now possible. First, we need not show the marginal relation between each pair of variables in the mosaic matrix. For example, Figure 12 shows the pairwise *conditional* relations among these variables. All panels show the same observed frequencies by the areas of the tiles, but each fits a model of conditional independence between the row and column variable, with the remaining variable controlled. Thus, the shading in the (1,2) and (2,1) panels show the fit of the model [Admit,Dept] [Gender, Dept], which asserts that Admission and Gender are independent, given (controlling for) department. Except for Department A, this model fits quite well, again indicating lack of gender bias. The (1,3) and (3,1) panels show the relation between admission and department controlling for gender, highlighting the differential admission rates across departments.

Second, the analogous conditional matrix plot for quantitative variables is of some interest itself. For each pair of variables, X_i, X_j, we plot $\widetilde{X_i} = X_i - \widehat{X_i}|$others against $\widetilde{X_j} = X_j - \widehat{X_j}|$others, where "others" is the complementary set excluding X_i, X_j; that is, each plot shows the residuals from regressions of X_i and X_j on all other variables. For multivariate normal data, Whittaker (1990) shows that X_i, X_j are conditionally independent of the others if and only if the corresponding element of the inverse covariance matrix $\Sigma^{-1} = \{\sigma^{ij}\}$ is zero,

$$\begin{aligned} \rho_{ij|\text{ others}} = 0 \quad &\Longleftrightarrow \quad \sigma^{ij} = 0 \\ &\Longleftrightarrow \quad X_i \perp X_j | \text{ others} \end{aligned} \tag{3}$$

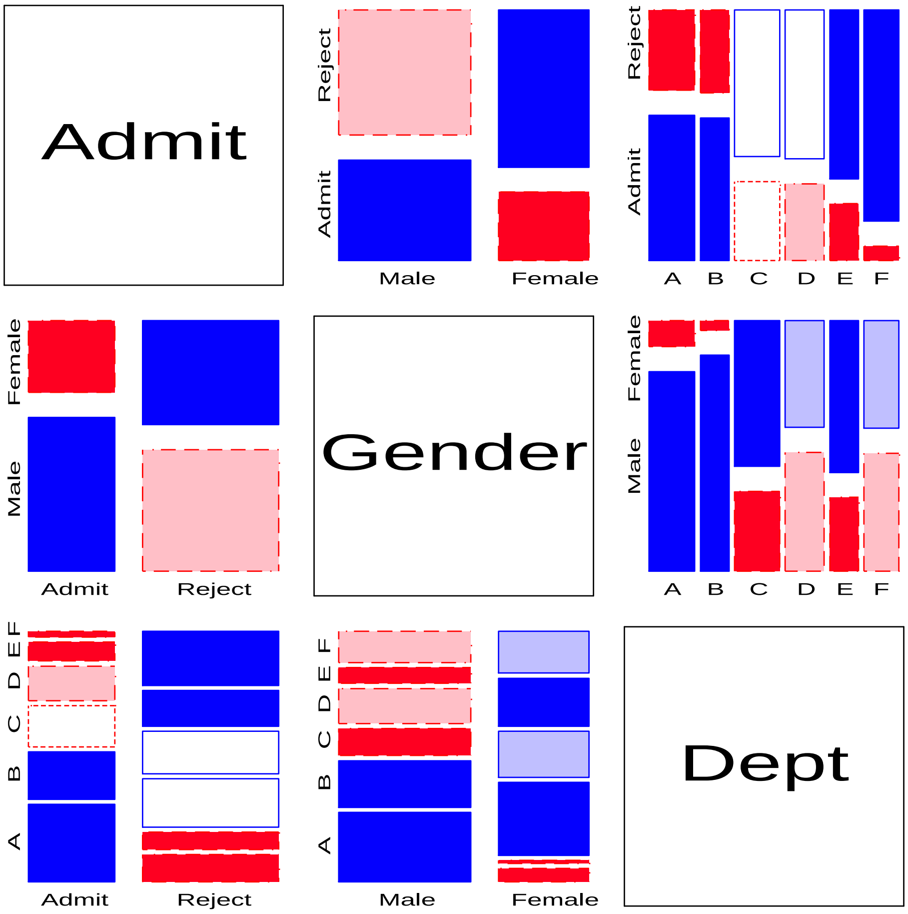


Figure 11: Mosaic matrix of Berkeley admissions. Each panel shows the marginal relation, fitting an independence model.

Zero partial correlation plays the same role in (undirected) graphical models for quantitative variables as two-way terms in graphical log-linear models. Hence, the conditional scatterplot matrix for quantitative variables provides a visualization of the pairwise partial correlations (or inverse covariance matrix) among all variables and of the conditional independence relations studied in Gaussian graphical models. Moreover, when one variable, Y, is a response, the panels in the row for Y are just the partial regression (added variable) plots. The other rows treat each variable in turn as a response, giving a multiway generalization of partial regression plots.

For example, Figure 13 shows a conditional scatterplot matrix of the well-known Iris data (Anderson, 1935), wherein each panel depicts the partial correlation between row and column variable given the remaining two variables. In the analogous scatterplot matrix of marginal relations (too familiar to most readers to show here) all pairs of variables are positively correlated and the three iris species are widely separated. The conditional plot tells a different and simpler story, however. When other variables are controlled, pairs consisting of the same flower component (petal or sepal) or the same measurement (length or width) are positively correlated, while cross component-measure pairs (e.g., petal width, sepal length) are negatively associated. There are also no apparent differences in

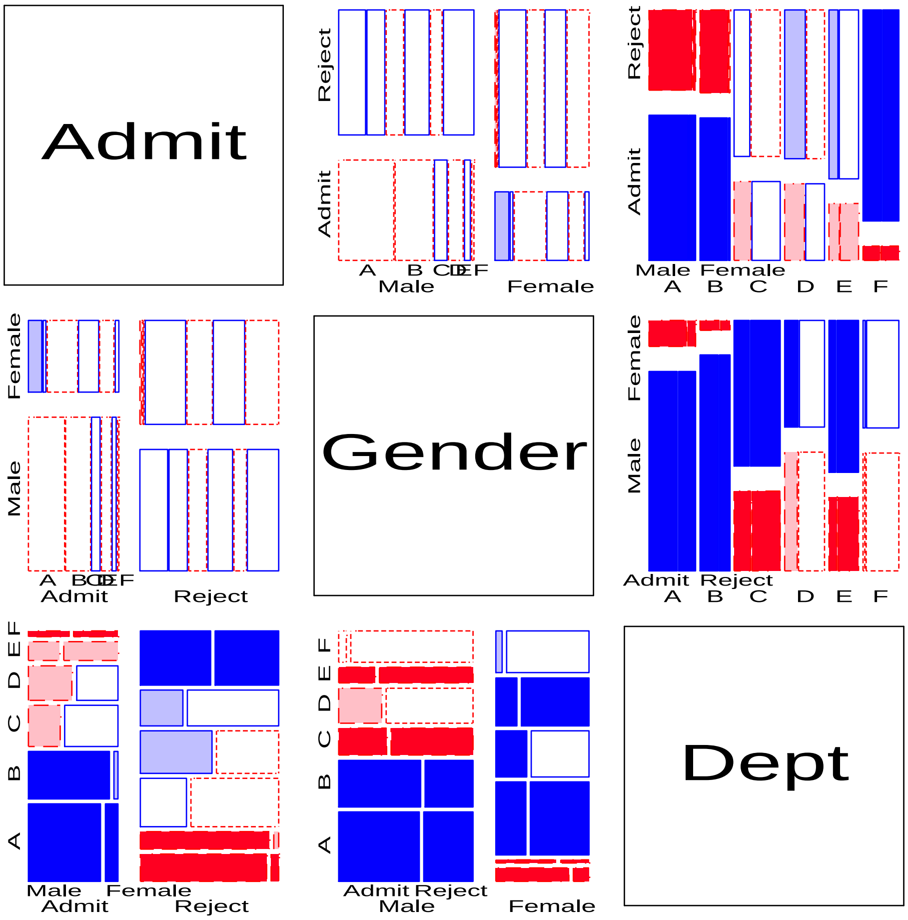


Figure 12: Conditional mosaic matrix of Berkeley admissions. Each panel shows the conditional relation, fitting a model of conditional independence between the row and column variable, controlling for other variable(s).

means among species, although more detailed modeling (Whittaker, 1990, Example 11.5.1) suggests that more complex models may be appropriate.

Hence, for the Iris data, no pair of variables is conditionally independent. Figure 14 shows a form of the independence graph (with line thickness proportional to the magnitude of partial correlation and line style indicating direction), summarizing the partial correlations shown explicitly in Figure 13. In the marginal plots, the large differences among species means imply that the 0-order correlations are poor summaries of the bivariate relations. The conditional plots in Figure 13 indicate that the species effects have been removed by partialling other variables, so that the partial correlations are not confounded by species differences.

Third, the framework of the scatterplot matrix can now be used as a general method for displaying marginal or conditional relations among a mixture of quantitative and categorical variables. For marginal plots, pairs of quantitative variables are shown as a scatterplot, while pairs of categorical variables are shown as a mosaic display. Pairs consisting of one quantitative and one categorical variable can be shown as a set of boxplots for each level of the categorical variable. For conditional plots, we

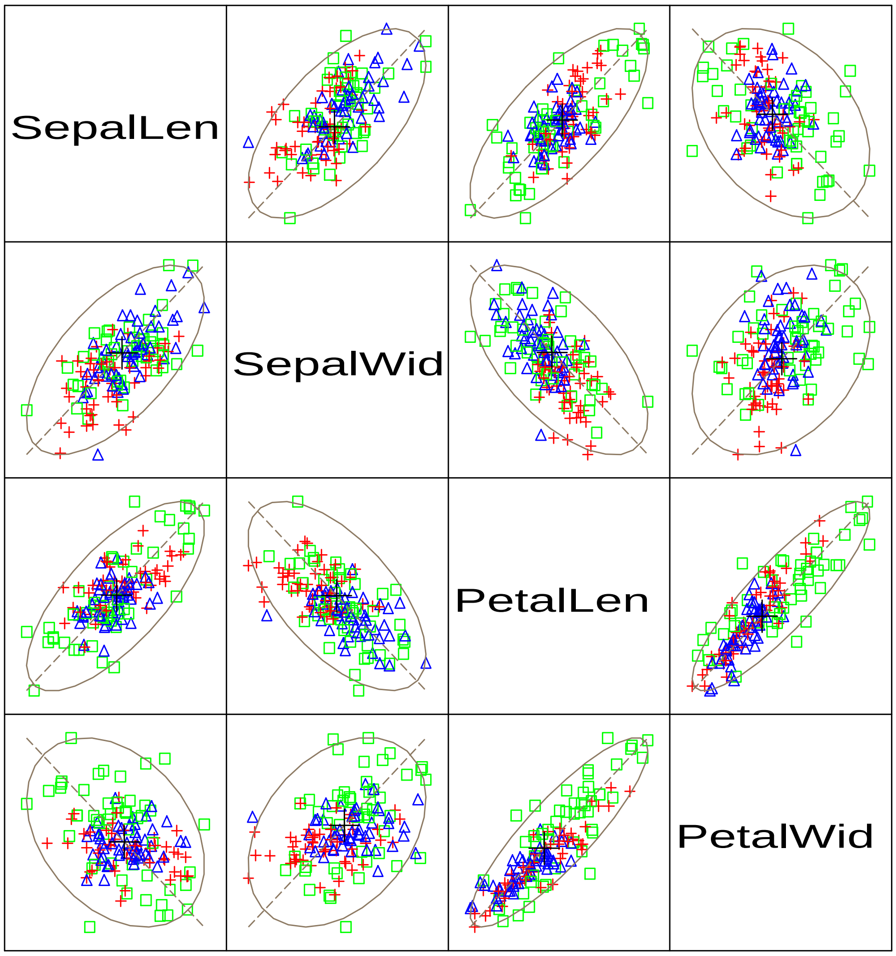


Figure 13: Conditional scatterplot matrix for Iris data. Plot symbols indicate species: *Setosa* ($\triangle$), *Versicolor* ($+$), *Virginica* ($\square$).

can fit a pair of generalized linear models, predicting the row and column variables from the others,

$$\begin{aligned} g(\mu_i) &= x^{\mathsf{T}}_{\text{others}}\beta \\ g(\mu_j) &= x^{\mathsf{T}}_{\text{others}}\beta \end{aligned}$$

with an identity link for quantitative variables, and log link for discrete variables. The mixed conditional plot then shows the residuals as in the marginal views. The details of this extension are a topic for future research.

5 Partial views: Coplots for categorical data

Conditional relations among variables may also be visualized by stratifying the data on the given variables, rather than by partialling out. For quantitative variables, a visually effective device is the ***coplot*** (or Trellis) display (Cleveland, 1993).

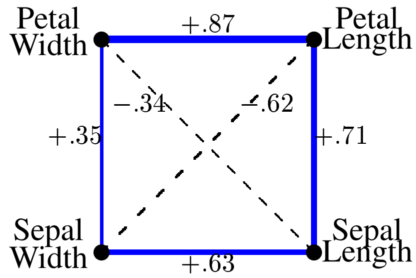


Figure 14: Independence graph for Iris Data. Numbers along each edge give the partial correlation, controlling for other variables.

One analog of the coplot for categorical data is an array of plots of the dependence among two or more variables, stratified by the values of one or more *given* variables. Each such panel then shows the *partial* associations among the foreground variables; the collection of such plots show how these change as the given variables vary.

For categorical data, models of independence fit to the strata separately have the useful property that they decompose a model of conditional independence fit to the whole table. Consider, for example, the model of conditional independence, $A \perp B \mid C$ for a three-way table. This model asserts that A and B are independent within *each* level of C. Denote the hypothesis that A and B are independent at level $C(k)$ by $A \perp B \mid C(k)$. Then one can show (Anderson, 1991) that

$$G^2_{A \perp B \mid C} = \sum_{k}^{K} G^2_{A \perp B \mid C(k)} \tag{4}$$

That is, the overall G^2 for the conditional independence model with $(I-1)(J-1)K$ degrees of freedom is the sum of the values for the ordinary association between A and B over the levels of C (each with $(I-1)(J-1)$ degrees of freedom). Thus, (a) the overall G^2 may be decomposed into portions attributable to the AB association in the layers of C, and (b) the collection of mosaic displays for the dependence of A and B for each of the levels of C provides a natural visualization of this decomposition.

These partial mosaics have the additional useful property that they adjust automatically for differing marginal frequencies across the strata, because the area of each partial mosaic is the same. This facilitates controlled comparison, allowing us to focus attention on the association of the foreground variables.

Figure 15 and Figure 16 show two further examples, using the mosaic display to show the partial relations [Admit][Dept] given Gender, and [Admit][Gender] given Dept, respectively. Figure 16 shows the same results displayed in Figure 5: no association between Admission and Gender, except in Dept. A, where females are relatively more likely to gain admission. But one can also see how the proportion admitted decreases regularly from Dept. A to F and how the proportion of females changes across departments. The breakdown of the overall G^2 from Eqn. (4) is given in Table 5.

Figure 15 shows that there is a very strong association between Admission and Department—different rates of admission, but also shows two things not seen in other displays: First, the *pattern* of

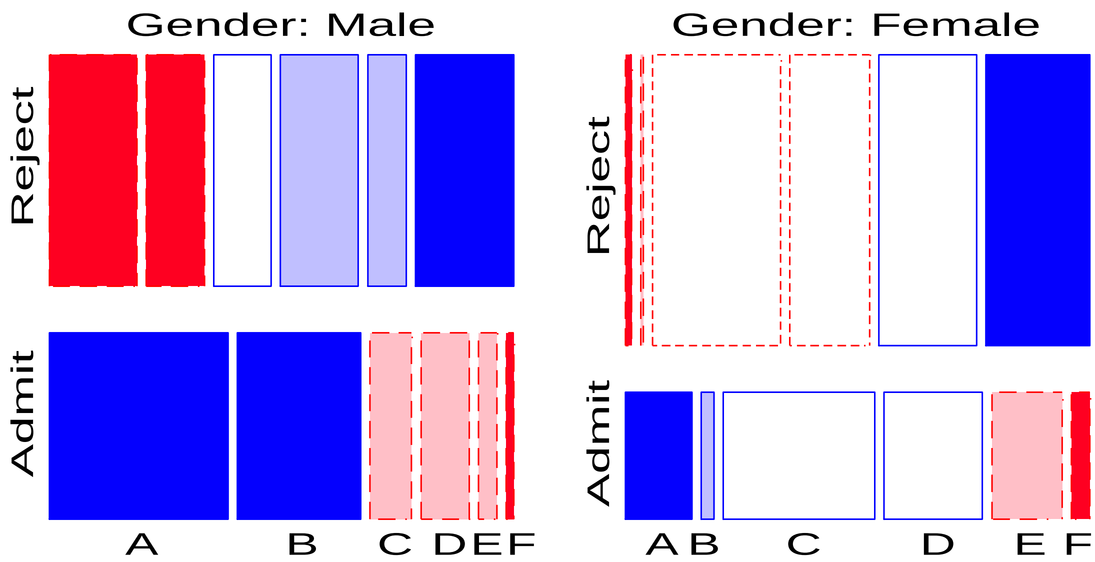


Figure 15: Mosaic coplot of Berkeley admissions, given Gender. Each panel shows the partial relation, fitting a model of independence between Admission and Department.

Table 5: Partial tests of independence of Gender and Admission, by Department

Dept	df	G^2	p
A	1	19.054	0.000
B	1	0.259	0.611
C	1	0.751	0.386
D	1	0.298	0.585
E	1	0.990	0.320
F	1	0.384	0.536
Total	6	21.735	0.001

association is qualitatively similar for both men and women; second the association is quantitatively stronger for men than women—larger differences in admission rates across departments.

5.1 Summary

Taken together, mosaic matrices and mosaic coplots extend the use of the mosaic display in simple, but powerful ways, and provide useful techniques for the graphical display of categorical and quantitative data within a common framework.

References

Anderson, E. The irises of the Gaspé. peninsula. *Bulletin of the American Iris Society*, 35:2–5, 1935.

Anderson, E. B. *Statistical Analysis of Categorical Data*. Springer-Verlag, Berlin, 1991.

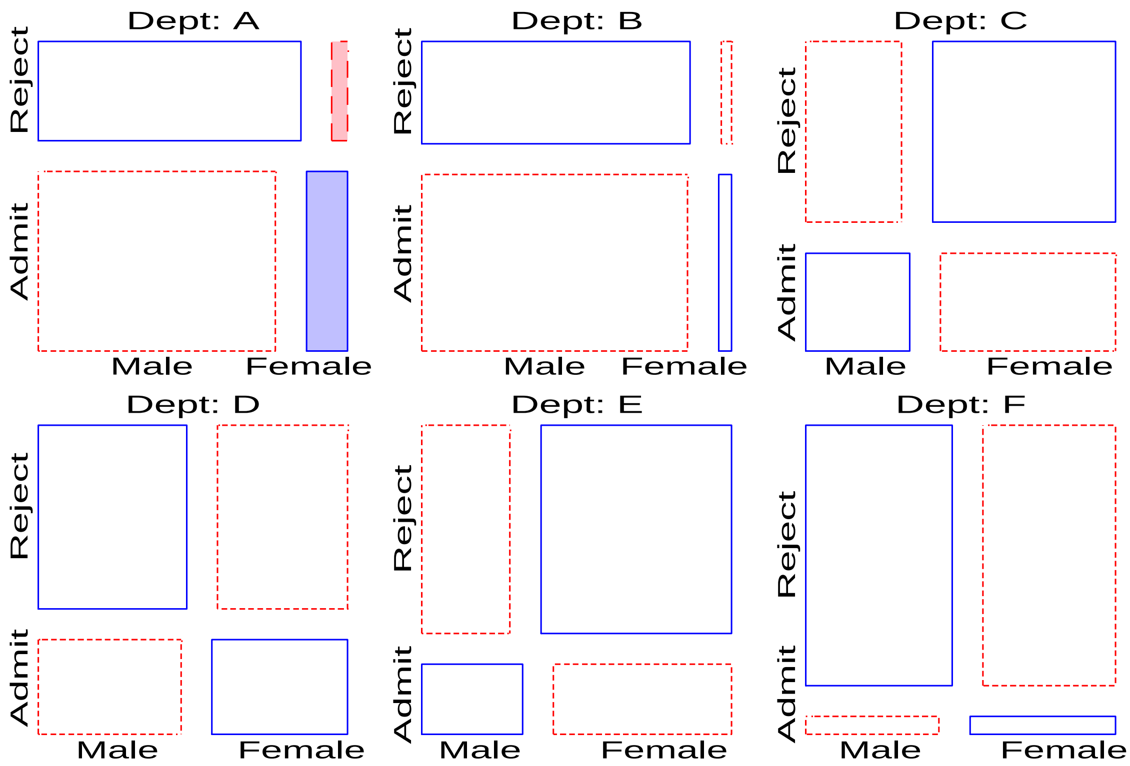


Figure 16: Mosaic coplot of Berkeley admissions, given Department. Each panel shows the partial relation, fitting a model of independence model between Admission and Gender.

Bickel, P. J., Hammel, J. W., and O'Connell, J. W. Sex bias in graduate admissions: Data from Berkeley. *Science*, 187:398–403, 1975.

Cleveland, W. S. *Visualizing Data*. Hobart Press, Summit, NJ, 1993.

Cleveland, W. S. and McGill, R. Graphical perception: Theory, experimentation and application to the development of graphical methods. *Journal of the American Statistical Association*, 79:531–554, 1984.

Dawson, R. J. M. The "unusual episode" data revisited. *Journal of Statistics Education*, 3(3), 1995.

Freedman, D., Pisani, R., and Purves, R. *Statistics*. Norton, New York, 1978.

Friendly, M. Mosaic displays for loglinear models. In *ASA, Proceedings of the Statistical Graphics Section*, pages 61–68, Alexandria, VA, 1992a.

Friendly, M. User's guide for MOSAICS. Technical Report 206, York University, Psychology Dept, 1992b. `http://www.math.yorku.ca/SCS/mosaics.html`.

Friendly, M. Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89:190–200, 1994.

Friendly, M. Conceptual and visual models for categorical data. *The American Statistician*, 49:153–160, 1995.

Friendly, M. Conceptual models for visualizing contingency table data. In Greenacre, M. and Blasius, J., editors, *Visualization of Categorical Data*, chapter 2, pages 17–35. Academic Press, San Diego, CA, 1997.

Friendly, M. Visualizing categorical data. In Sirken, M., Herrmann, D., Schechter, S., Schwarz, N., Tanur, J., and Tourangeau, R., editors, *Cognition and Survey Research*, chapter 20, pages ??–?? John Wiley and Sons, New York, 1999. In press.

Goodman, L. A. The multivariate analysis of qualitative data: Interactions among multiple classifications. *Journal of the American Statistical Association*, 65:226–256, 1970.

Greenacre, M. *Theory and Applications of Correspondence Analysis*. Academic Press, London, 1984.

Greenacre, M. Correspondence analysis of multivariate categorical data by weighted least squares. *Biometrika*, 75:457–467, 1988.

Greenacre, M. Diagnostics for joint displays in correspondence analysis. In Blasius, J. and Greenacre, M., editors, *Visualization of Categorical Data*, pages 221–238. Academic Press, 1997.

Hartigan, J. A. and Kleiner, B. Mosaics for contingency tables. In Eddy, W. F., editor, *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, pages 268–273. Springer-Verlag, New York, NY, 1981.

Hartigan, J. A. and Kleiner, B. A mosaic of television ratings. *The American Statistician*, 38:32–35, 1984.

Snee, R. D. Graphical display of two-way contingency tables. *The American Statistician*, 28:9–12, 1974.

Whittaker, J. *Graphical Models in Applied Multivariate Statistics*. John Wiley and Sons, 1990.

Young, F. W. ViSta: The visual statistics system. Technical Report RM 94-1, L.L. Thurstone Psychometric Laboratory, UNC, 1994.