# Visualizing Multi-way Contingency Tables

David Meyer[1], Achim Zeileis[2], and Kurt Hornik[2]

[1] Department of Information Systems and Process Management
[2] Department of Statistics and Mathematics
Wirtschaftsuniversität Wien
Augasse 2–6, 1090 Vienna, Austria
{David.Meyer,Achim.Zeileis,Kurt.Hornik}@wu-wien.ac.at

## 1 Introduction

Categorical data analysis is typically based on two- or higher-dimensional contingency tables, cross-tabulating the co-occurences of levels of nominal and/or ordinal data. In order to explain these, statisticians typically look for (conditional) independence structures using common methods such as independence tests and log-linear models. One idea of visualization techniques is to use the human visual system to detect structures in the data that possibly are not obvious from solely numeric output (e.g., test statistics). Whether the task is purely exploratory or model based, techniques such as mosaic, sieve, and association plots offer good support for visualization. They have been extended over the last two decades, and implementations exist in many statistical environments.

All three graphical methods visualize aspects of (possibly higher-dimensional) contingency tables. A *mosaic plot* [**?**] is basically an area-proportional visualization of (typically, observed) frequencies, composed of tiles (corresponding to the cells) created by recursive vertical and horizontal splits of a square. Thus, the area of each tile is proportional to the corresponding cell entry *given* the dimensions of previous splits. *Sieve plots* [**?**] are similar to mosaic plots, but the area of each tile is proportional to the *expected* cell entry, and each tile is filled with a number of rectangles corresponding to the observed value. An *association plot* [**?**] visualizes the standardized deviations of observed frequencies from those expected under a certain independence hypothesis. Each cell is represented by a rectangle that has (signed) height proportional to the residual and width proportional to the square root of the expected counts, so that the area of the box is proportional to the difference in observed and expected frequencies.

Over the years, extensions to these techniques have mainly been focused on the following aspects:

- Varying the shape of mosaic plots (as well as bar plots) to yield, e.g., double-decker plots [**?**], spine plots [**?**], or spinograms.
- Using residual-based shadings to visualize log-linear models [**?**, **?**] and significance of statistical tests [**?**].
- Using pairs plots and trellis-like layouts for marginal, conditional and partial views [**?**].
- Adding direct user interaction, allowing quick exploration and modification of the visualized models [**?**, **?**].
- Providing a modular and flexible implementation to easily allow user extensions [**?**].

Current implementations of mosaic displays can be found, e.g., for SAS [**?**], `ViSta` [**?**], `MANET` [**?**], `Mondrian` [**?**], R [**?**], and S-PLUS [**?**]. Implementations of association and sieve plots can only be found in R and SAS (in the latter, these plots are available for two-way tables only). Table **??** gives an overview of the available functionality in these systems. The figures in this chapter have all been produced using the R system, using the extension packages `vcd` [**?**] and `scatterplot3d` [**?**] (Fig. **??** only), all freely available from the Comprehensive R Archive Network (`http://CRAN.R-project.org/`).

| | SAS | S-PLUS | R | ViSta | MANET/Mondrian |
|---|---|---|---|---|---|
| Basic functionality | × | × | × | × | × |
| Shape | | | × | | × |
| Residual-based shadings | × | | × | × | (×) |
| Conditional Views | × | | × | | × |
| Interaction | | | | × | × |
| Extensible Design | | | × | | |

**Table 1.** Comparison of current software environments.

This chapter will give an overview of the state of the art of mosaic, sieve, and association plots, both for exploratory visualization and model-based analysis. Exploratory techniques will include specialized displays for the bivariate case, as well as pairs plot-like displays for higher-dimensional tables. As for the model-based tools, particular emphasis will be given to methods suitable for the visualization of conditional independence tests (including permutation tests), as well as for the visualization of particular GLMs (such as log-linear models). In Sect. **??**, we start with the simple bivariate case. Section **??** explains how the use of color in residual-based shadings can support data exploration, and even promotes the methods to diagnostic and model-based tools by visualizing test statistics and residuals of independence models. In Sect. **??**, we show how the basically bivariate methods straightforwardly extend to the multivariate case by using 'flat' representations of the multi-way tables. In this section, we also introduce specialized displays for conditional independence structures. Section **??** concludes the chapter.

## 2 Two-way Tables

Throughout this section, our examples will be based on the hospital data [**?**] (see Tab. **??**).

| | **Length of stay (in years)** | | | |
|---|---|---|---|---|
| **Visit frequency** | 2–9 | 10–19 | 20+ | $\Sigma$ |
| Regular | 43 | 16 | 3 | 62 |
| Less than monthly | 6 | 11 | 10 | 27 |
| Never | 9 | 18 | 16 | 43 |
| $\Sigma$ | 58 | 45 | 29 | 132 |

**Table 2.** The hospital data.

The table relates the length of stay (in years) of 132 long-term schizophrenic patients in two London mental hospitals with the freqency of visits. The length of stay (LOS) has been categorized into 2–9 years, 10–19 years, and more than 19 years. There are also three categories for the visit frequency: regular (including patients that were allowed to go home), less than monthly, and never. Wing [**?**] concludes from this data that the longer the length of stay in hospital, the less frequent the visits, which can be seen from the column-standardized table (see Tab. **??**).

| | **Length of stay (in years)** | | |
|---|---|---|---|
| **Visit frequency** | 2–9 | 10–19 | 20+ |
| Regular | 0.74 | 0.36 | 0.10 |
| Less than monthly | 0.10 | 0.24 | 0.35 |
| Never | 0.16 | 0.40 | 0.55 |
| $\Sigma$ | 1.00 | 1.00 | 1.00 |

**Table 3.** The hospital data, corrected for the column margin.

In addition, [**?**] notes that this pattern is not significantly different in the "less than monthly" and "never" strata. From the row-standardized table (see Tab. **??**), it seems indeed that LOS were homogeneous with respect to these two visit frequency strata.

Although far from optimal, contingency tables are frequently visualized using grouped bar plots (see Fig. **??**) or even by means of 3D-bar charts (see Fig. **??**). It seems hard to detect the aforementioned pattern in these, especially in the 3D plot where the perspective view tends to distort the true proportions of the bars. In the following, we will introduce three graphical methods that are better suited for contingency tables.

| | Length of stay (in years) | | | |
|---|---|---|---|---|
| **Visit frequency** | 2–9 | 10–19 | 20+ | $\Sigma$ |
| Regular | 0.69 | 0.26 | 0.05 | 1.00 |
| Less than monthly | 0.22 | 0.41 | 0.37 | 1.00 |
| Never | 0.21 | 0.42 | 0.37 | 1.00 |

**Table 4.** The hospital data, corrected for the row margin.

### 2.1 Mosaic Displays

Mosaic displays have been introduced by [**?**, **?**] and extended, e.g., by [**?**, **?**, **?**]. They visualize the observed values of a contingency table by area-proportional tiles, arranged in a squared mosaic. The tiles are obtained by recursive partitioning splits of a square. In the following, we describe the main idea of mosaic plots, chapter XXX in this book **(Reference to other contributions)** provides more detailed information. Consider our example of the hospital data from above. Step 1 consists of splitting a square according to the marginals of one of the variables. To be consistent with the textual representation, we choose LOS with vertical splits (see Fig. **??**). The result is similar to a bar plot where not the height, but the width is adapted to visualize the counts for each level, which is also called a *spine plot* [**?**]. From this plot, we see that the number of patients decreases with the length of stay. Step 2 now is to add further splits in the other direction, i.e., horizontal splits, for the second variable. This means that each vertical bar is split according to the marginals of the second variable, *given* the first variable (see Fig. **??**). The resulting plot visualizes the contingency table where each cell has a size proportional to the corresponding table entry. We can still see the marginal distribution of LOS, and additionally, the visit frequency *given* each category of LOS. If the two variables were independent, the grid would be regular. Clearly, compared to a length of stay of 10–19 years, more patients get regular visits for stays from 2–9 years, and conversely, less patients get regular visits for stays for more than 19 years. For patients that get no visits, the pattern is reversed.

Since mosaic plots are asymmetric by construction, the choice of the variable order matters, as the first splitting variable dominates the plot. In our example, if we use 'Visit frequency' as the first splitting variable, the impression is very different compared to the previous mosaic (see Fig. **??**). In this alternative display, we see the marginal distribution of 'Visit frequency' in the rows: about half of the patients get visited regularly. This group is dominated by patients staying between 2 and 9 years. It seems apparent that the distribution of LOS is similar for monthly and never visited patients, so this two categories actually represent one homogenous group (patients visited only casually). Since the first splitting variable dominates the plot, it should be chosen as to be the explanatory variable in (hypothesized) causal relationships.

### 2.2 Sieve Plots

When we try to explain data, we assume the validity of a certain model for the generating process. In the case of two-way contingency tables, the two most common hypotheses [**?**] are

1. independence of the two variables
2. homogeneity of one variable among the strata defined by the second.

It is easy to compute the *expected* table under either of these hypotheses. Consider a 2-way contingency table with $I$ rows and $J$ columns, cell frequencies $\{n_{ij}\}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$, and row and column sums $n_{i+} = \sum_i n_{ij}$ and $n_{+j} = \sum_j n_{ij}$, respectively. For convenience, the number of observations is denoted $n = n_{++}$. Given an underlying distribution with theoretical cell probabilities $\pi_{ij}$, the null hypothesis of independence of the two categorical variables can be formulated as

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}. \tag{1}$$

| | **Length of stay (in years)** | | | |
|---|---|---|---|---|
| **Visit frequency** | 2–9 | 10–19 | 20+ | $\Sigma$ |
| Regular | 27.24 | 21.14 | 13.62 | 62 |
| Less than monthly | 11.86 | 9.20 | 5.93 | 27 |
| Never | 18.89 | 14.66 | 9.45 | 43 |
| $\Sigma$ | 58.00 | 45.00 | 29.00 | 132 |

**Table 5.** The hospital data—expected values.

Now, the expected cell frequencies in this model are simply $\hat{n}_{ij} = n_{i+}n_{+j}/n$. The expected table for our sample data is given in Tab. **??**. It could again be visualized using a mosaic plot, this time applied to the table of expected frequencies. If we cross-tabulate each tile to fill it with a number of squares equal to the corresponding number of *observed* frequencies, we get a *sieve plot* (see Fig. **??**). It implicitly compares expected and observed values since the density of the grid will increase with the deviation of the observed from the expected values. This allows the detection of general association patterns (for nominal variables) and of linear association (for ordinal variables), the latter producing tiles of either very high or very low density along one of the diagonals. In the case of our data, the density of the rectangles is marked along the secondary diagonal, indicating a negative association of the two variables. This gives evidence to the fact that for these patients, visit frequency decreases with the length of stay.

### 2.3 Association Plots

In the last section, we have seen how to compare observed and expected values of a contingency table using sieve plots. We can do this more straightforwardly by using a plot that directly visualizes the residuals. The most widely known residuals are the Pearson residuals

$$r_{ij} \quad = \quad \frac{n_{ij} - \hat{n}_{ij}}{\sqrt{\hat{n}_{ij}}}. \tag{2}$$

which are standardized raw residuals. In an association plot [**?**], each cell is represented by a rectangle that has (signed) height proportional to the corresponding Pearson residual $r_{ij}$ and width proportional to the square root of the expected counts $\sqrt{\hat{n}_{ij}}$. Thus, the area is proportional to the raw residuals $n_{ij} - \hat{n}_{ij}$. The sign is visualized by its position relative to the baseline (upward tiles for positive, and downward tiles for negative residuals). Figure **??** shows the association plot for the hospital data. Consistent with the corresponding mosaic and sieve plots, we clearly see that too many (too few) patients that stay from 2 to 9 (more than 19) years get visited regularly than would be expected under the null of independence, and that this pattern is reversed for patients visited less than monthly and that get never visited.

### 2.4 Summary

Mosaic plots are an instrument to visualize the *observed* frequencies of a contingency table by recursive conditional splits. If causal relationships between the variables are assumed, the determining variables should be used first for splitting since they will dominate the plot. Sieve plots basically visualize the table of *expected* frequencies, and in addition the deviations from the observed frequencies by the density of a grid added to each tile. They are best used to detect dependency patterns, especially for ordinal variables. Association plots directly visualize Pearson and raw residuals, i.e., standardized and non-standardized deviations of observed from expected frequencies, respectively. These plots should be used if the diagnostics of independence models is of primary concern.

## 3 Using Colors for Residual-based Shadings

As introduced in the previous section for association plots, the investigation of residuals from a posited independence model is of major interest in analyzing contingency tables. In the following, we will demonstrate how the use of colors can greatly facilitate the detection of interesting patterns. Before, we start with some general remarks on colors and color palettes.

### 3.1 A Note on Colors and Color Palettes

The plots introduced in the previous section are basically composed of tiles whose areas represent characteristics derived from the contingency tables—observed and expected frequencies in the case of mosaic and sieve plots, or residuals as visualized by association plots. When using color for these tiles, it is imperative to choose the right color palettes, derived from suitable color spaces. Apart from aesthetic considerations, wrongly chosen colors might serously affect the analysis. For example,

- Using high-chroma colors for large areas tend to produce after-image effects which can be distracting [**?**].
- Lighter colors tend to make areas look larger than darker colors, therefore using colors with unequal luminance values makes it difficult to compare area sizes [**?**].
- Color palettes derived from non-uniform color spaces may contain unbalanced colors with respect to their colorfulness or brightness. When tiles are shaded using such a palette, some of them might appear more important than others in an uncontrolled way.

Due to the three-dimensional nature of human color perception [**?**], it has been common to specify colors using the three primaries red, green, and blue (RGB colors), especially for computer devices. The appearance of the on-screen colors is affected by the device characteristics: For example, the intensity $I$ of a primary on a particular device follows the rule $I = L^\gamma$, with $L$ the value for the primary color, and $\gamma$ device-dependent (but typically close to 2.2). Therefore, a first caveat is that if colors are to appear identical on different devices, their gamma characteristics have to be taken into account.

Now, software implementations are often based on the more convenient Hue-Saturation-Value (HSV) colors (or the comparable Hue-Luminance-Saturation color scheme). Both spaces are rather similar transformations of the RGB space [**?**, **?**] and are very common implementations of colors in many computer packages [**?**]. Each color in this space is represented by three dimensions $(H, S, V)$: the hue $H$ (dominant wavelength in the spectrum, in $[0, 360]$), the saturation $S$ ('colorfulness' or 'pureness', in $[0, 100]$), and the value $V$ ('brightness', amount of gray, in $[0, 100]$).[3] These intuitive dimensions make HSV colors easyier to specifiy than, e.g, RGB colors. But HSV colors have several disadvantages. Most importantly, HSV colors are not perceptually uniform because the three HSV dimensions map only poorly to the three perceptual dimensions of the human visual system [**?**, **?**]. One important issue here is that HSV dimensions are confounded, e.g., saturation is not uniform across different hues. As an example, see Fig. **??** (left) showing a qualitative color palette (colors $(H, 100, 100)$ for varying hues $H$) in the HSV space: although

[3] In many implementations, all three dimensions are scaled to the unit interval instead of the coordinates used here.

saturation and value are fixed, the fully saturated blue is perceived much darker than the fully saturated red or green, making it difficult to judge the size of shaded areas. Furthermore, the flashy fully saturated HSV colors are hard to look at for a longer time. For similar reasons, it is equally difficult to derive acceptable diverging palettes from the HSV space, i.e., bipolar scales containing colors ranging between two very distinct colors. The upper part of Fig. **??** shows a diverging palette in the HSV space with colors ranging from a saturated red $(0, 100, 100)$ over a neutral white $(H, 0, 100)$ to a saturated blue $(240, 100, 100)$. Although the palette should be balanced with respect to colorfulness and brightness, the red colors are perceived to be more more intense and flashy than the corresponding blue colors.

The use of colors that are more 'in harmony' goes back to Munsell [**?**] who introduced a notation for balanced colors. Based on those, tools producing better palettes for specific tasks have been developed [**?**]. Other perceptually-based color spaces, especially suited for computer displays, are the CIELAB and CIELUV spaces [**?**] from which qualitative palettes for statistical graphics have been derived [**?**]. A transformation of the CIELUV space leads to the HCL (Hue-Chroma-Luminance) space. These colors are again specified by triplets $(H, C, L)$: chroma $C$ loosely corresponds to colorfulness and luminance $L$ to brightness, but in contrast to HSV colors, chroma is an *absolute* measure valid for all hues, and luminance can be varied independently from the other two dimensions. Qualitative color palettes can easily be obtained by holding constant chroma and luminance, and using varying hues. HCL colors with fixed chroma and luminance are always balanced towards the same grey and thus do not have the problem of varying saturations as the HSV colors (see Fig. **??**, right). Similarly, diverging HCL color palettes can be derived [**?**] by interpolating again between a neutral color such as $(H, 0, 90)$ and two colors with full chroma such as blue $(260, 100, 50)$ and red $(0, 100, 50)$. The resulting palette is shown in the lower part of Fig. **??**. In contrast to the HSV counterpart, matching colors (light red/blue and dark red/blue) are balanced to the same gray, and thus receive the same perceptual "weight". Note, that this varies chroma and luminance are varied simultaneously, i.e., the full chroma colors are also darker than the neutral color. Of course, it would also be possible to choose a darker neutral color (or lighter full chroma colors), however, by varying both chroma and luminace better contrasts can be achieved.

### 3.2 Highlighting and Color-based Shadings

All plots introduced in Sect. **??** are composed of empty tiles. It seems intuitive to use filled tiles to highlight information of interest. Consider again the hospital example: in Fig. **??**, we mark tiles for patients never or seldom visited to visualize their proportion in the LOS strata. For optical clarity, we set the spacing between the visit frequency tiles to 0. Clearly, the proportion of these patients increases with LOS. In fact, such a "stacked" plot can also be interpreted as spine plot with highlighting [**?**], particularly useful in analyzing

**Fig. 1.** Bar plot for the hospital data.

**Fig. 2.** 3D-bar chart for the hospital data.

**Fig. 3.** Construction of a mosaic plot for a two-way table, step 1.

**Fig. 4.** Construction of a mosaic plot for a two-way table, step 2.

**Fig. 5.** Mosaic plot for the hospital data, using 'Visit frequency' as first splitting variable.

**Fig. 6.** Sieve plot for the hospital data.

**Fig. 7.** Association plot for the hospital data.

**Fig. 8.** Qualitative color palette for the HSV (left) and HCL (right) spaces. The HSV colors are $(H, 100, 100)$ and the HCL colors $(H, 50, 70)$ for the same hues $H$.

**Fig. 9.** Diverging color palettes for the HSV space (upper part) and the HCL space (lower part), ranging from blue over a neutral color to red. The triples indicate the settings for the three dimensions: (Hue, Saturation, Value) in the upper part, and (Hue, Chroma, Luminance) in the lower part.

causal relationships in categorical data with a binary dependent variable. More variations on this theme, such as doubledecker plots, are treated in chapter XXX (**REFERENCE TO OTHER CHAPTER(S)**).

**Fig. 10.** Spine plot with highlighting for the hospital data.

Using colors, even more complementary information can be visualized, either by adding additional information, or by redundantly coding information already visualized by the 'raw' plot to support our perceptual system. First, we consider the sieve plots. The density of the grid in the raw version implicitly gives us an idea of the residuals' size, but since the plot does not include the density corresponding to zero residuals (the null model) for comparison, we cannot easily assess whether there are more, of less counts in a cell than expected under the null hypothesis. It would help if we knew the sign of the residuals. Using color, we can add this information, for example using blue for positive, red for negative, and gray for zero residuals. Figure **??** demonstrates the effect of applying such a color shading to a sieve plot for the hospital data. Color clearly emphasizes the linear association pattern in this data: Patients with a LOS from 2–9 years are associated with regular visits, whereas patients with a LOS from 10–19 and more than 19 years are associated with a visit frequency of less than monthly and no visits.

**Fig. 11.** Sieve plot with color coding of the residuals.

Mosaic plots in their initial version are monochrome displays. Friendly [**?**] introduced a residual-based shading of the tiles to additionally visualize the residuals from a given independence model fitted to the table. The idea is to use a color coding for the mosaic tiles that visualizes the sign and absolute size of each residual $r_{ij}$: Cells corresponding to small residuals ($|r_{ij}| < 2$) have no color. Cells with medium sized residuals ($2 \leq |r_{ij}| < 4$) are shaded light blue and light red for positive and negative residuals, respectively. Cells with large residuals ($|r_{ij}| \geq 4$) are shaded with a fully saturated blue and red, respectively. The heuristic for choosing the cut-offs 2 and 4 is that the Pearson residuals are asymptotically standard normal which implies that the highlighted cells are those with residuals *individually* significant at approximately the $\alpha = 0.05$ and $\alpha = 0.0001$ levels. However, the main purpose of the Friendly shading is not to visualize significance but the *pattern* of deviation from independence [**?**]. In addition to the shading of the rectangles themselves, the Friendly shading also encompasses a choice of line type and line color of

the rectangle borders with similar ideas as described above, useful when no color can be used.

In Fig. **??**, we again show the mosaic for the hospital data, this time using a Friendly-like color shading (with HCL instead of HSV colors, and no line type coding).

**Fig. 12.** Mosaic display with Friendly-like color coding of the residuals.

Clearly, the asymmetry for regular and never visited students, and the pattern inversion for lengths of stay of 2–9 and more than 19 years are emphasized using the color shading.

For association plots, residual-based shadings are redundant since all relevant information is already contained in the plot by construction. But using one of the shadings discussed above will nevertheless support the analysis process and is therefore recommended. For example, applying the Friendly shading in Fig. **??** on the one hand facilitates the discrimination between positive and negative residuals, and on the other, more importantly, allows an easier comparison of the tiles' sizes. Thus, the shading supports the detection of the pattern.

**Fig. 13.** Association plot with Friendly-like color coding of the residuals.

### 3.3 Visualizing Test Statistics

Figures **??** and **??** include the (same) $p$ value of a $\chi^2$ test of independence, frequently used to assess the significance of the hypothesis of independence (or homogeneity for stratified data) in two-way tables. The test statistic is just the sum of the squared Pearson residuals

$$X^2 \quad = \quad \sum_{i,j} r_{ij}^2 \ , \tag{3}$$

known to have a limiting $\chi^2$ distribution with $(I-1)(J-1)$ degrees of freedom under the null hypothesis. An important reason for using the unconditional limiting distribution for the $X^2$ statistic from Equ. **??** was the closed form result for the distribution. Recently, with the improving perfomance of computers, conditional inference (or permutation tests)—carried out either by

simulation or by computation of the (asymptotic) permutation distribution—have been receiving increasing attention [**?**, **?**, **?**]. For testing the independence hypothesis from Equation **??**, using a permutation test is particularly intuitive due to the permutation invariance (given row and column sums) of this problem. Consequently, all results in this paper are based on conditional inference performed by simulating the permutation distribution of test statistics of type $\lambda([r_{ij}])$.

Since the HCL space is three-dimensional and we only used two 'degrees of freedom' so far for coding information (hue for the sign and a linear combination of chroma and luminance for the size of the residuals), we can add a third information to the plot. For example, we can visualize the significance of some specified test statistic (e.g., the $\chi^2$ test statistic) using darker ('uninteresting') colors for non-significant results. These can again be derived using the interpolation procedure described in the Sect. **??** by restricting the luminance to a shorter range, ending with a "darker" value.

The heuristic for choosing the cut-off points in the Friendly shading may lead to wrong conclusions: especially in large tables, the test of independence may not be significant even though some of the residuals are "large". On the the other hand, the test might be significant even though the residuals are "small". In fact, the cut-off points are really data-dependent. Consider the case of the arthritis data [**?**], resulting from a double-blind clinical trial investigating a new treatment for rheumatoid arthritis, stratified by gender (see Tab. **??** for the female patients):

| | **Improved** | | | |
|---|---|---|---|---|
| **Treatment** | None | Some | Marked | $\Sigma$ |
| Placebo | 19 | 7 | 6 | 32 |
| Treatment | 6 | 5 | 16 | 27 |
| $\Sigma$ | 25 | 12 | 22 | 59 |

**Table 6.** The arthritis data (female patients).

Figure **??** visualizes the results for the female patients, again by means of a mosaic display. Clearly, the hypothesis of independence is rejected by the $\chi^2$ test even on a 99% level ($p = 0.0018$), but since all residuals are in the $[-1.7173, 1.8696]$ interval, the tiles remain uncolored. A solution to this issue is to use a different test statistic, for example the maximum of the absolute values of the Pearson residuals [**?**] instead of the sum of squares:

$$M \quad = \quad \max_{i,j} |r_{ij}| \ . \tag{4}$$

Given a critical value $c_\alpha$ for this test statistic, all residuals whose absolute value exceeds $c_\alpha$ violate the hypothesis of indendence at level $\alpha$ [**?**, ch. 7].

Thus, the interesting cells giving evidence for the rejection of the independence hypothesis can easily be identified. As explained above, the conditional distribution of this test statistic under the null can be obtained by simulation, sampling tables with the same row and column sums $n_{i+}$ and $n_{+j}$ using, e.g., the Patefield algorithm [**?**] and computing the maximum statistic for each of these tables. In Fig. **??**, we again visualize the arthritis data, this time using the maximum test statistic. The shading of the tiles now clearly shows that the treatment is effective: significantly more patients in the treatment group exhibit marked improvement than would be expected under independence.

### 3.4 Summary

Especially for the visualization of areas, use perceptually uniform colors and color palettes, derived from, e.g., the HCL color space. Shadings can be used to add information to the basic plots and to support the analysis. Highlighting of tiles can support the analysis of causal relationships. Residual-based shadings can be used for model diagnostics, e.g., by using the hue to code the sign of a residual, and a linear combination of chroma and luminance for the size. In addition, the significance of test statistics can be visualized through overall darker color palettes. Using the maximum instead of the $\chi^2$ test statistic to aggregate the residuals allows precise diagnostics when the null hypothesis of independence is rejected.

## 4 Selected Methods for Multi-way Tables

In Sect. **??**, we have presented basic visualization methods for two-way tables. In this section, we will show how these methods extend to multi-way tables by applying them to 'flat' representations, and we will treat specialized displays for conditional independence models.

### 4.1 Visualizing Flat-Tables

The main idea of the mosaic, sieve, and association plots presented in the previous sections is to visualize information on the tables' cells, arranged in rectangular form. For multi-way tables, mosaic plots can directly be used by simply adding further splits for each additional variable. For sieve and association plots, we apply the basic idea of mosaic plots to the table itself, i.e., simply nest the variables into rows and columns using recursive conditional splits, given the margins. The result is a 'flat' representation of the multi-way table that can be visualized in ways similar to a two-dimensional table. As an example, consider the HairEyeColor data containing two polytomous variables (hair and eye color), as well as one (artificial) dichotomous variable (gender). A 'flattened' contingency table, putting eye color in the columns and hair

color—nested in gender—in the rows, is given by Tab. **??**. The corresponding mosaic, sieve, and association plots are shown in Figs. **??**–**??**. The shadings in these plots visualize the model of hair and eye color being *conditional* independent, given gender. As we can see from the shading (many tiles are colored) and the $p$ value for the test statistic, this model fits only poorly.

| | | **Eye** | | | |
|---|---|---|---|---|---|
| **Gender** | **Hair** | Brown | Blue | Hazel | Green |
| Male | Black | 32 | 11 | 10 | 3 |
| | Brown | 38 | 50 | 25 | 15 |
| | Red | 10 | 10 | 7 | 7 |
| | Blond | 3 | 30 | 5 | 8 |
| Female | Black | 36 | 9 | 5 | 2 |
| | Brown | 81 | 34 | 29 | 14 |
| | Red | 16 | 7 | 7 | 7 |
| | Blond | 4 | 64 | 5 | 8 |

**Table 7.** The HairEyeColor data, in flat representation.

### 4.2 Displays for Conditional Independence

In addition to the generic approach to the visualization of multi-way tables outlined in the previous section, there are specialized displays designed for the visualization of conditional independence structures.

A first step in the analysis of more complex tables is to find a suitable model that fits the data. All basic plots can be combined in pairwise displays, arranged in in a matrix similar to scatterplots in a pairs plot. The diagonal cells contain the variable names, optionally with univariate statistics, whereas the off-diagonal cells contain plots whose variables are implicitly specified by the cells' position in the matrix. More formally, each cell $a_{ij}$ in such a matrix defines two variables $i$ and $j$ that can be used to specify the model visualized in that cell. Typical hypotheses are: Variables $i$ and $j$ are marginally independent; variables $i$ and $j$ are conditionally independent, given all others; variables $i$ and $j$ are jointly independent from all others, etc. Figure **??** shows a pairs display with mosaic plots visualizing mutual independence in the upper triangle, mosaics for *joint* independence in the lower triangle, and bar charts in the diagonal. In the upper part of the matrix, we can immediately detect the independence of hair and gender, and eye and gender from the non-colored cells in the corresponding mosaic plots. On the other hand, hair and eye color clearly are associated. In the lower part of the matrix, the mosaic at the intersection of hair and eye is the less colored one. Its shading represents the model of hair and eye color being jointly independent from gender, i.e., their association being homogeneous over the gender strata. It is the model that fits

best, although still rejected at the 95% level ($p = 0.0161$): possibly, because too many brown-haired male students have blue eyes, and too few of them have brown eyes, than expected under the null model?

Since mosaic displays are conditional plots by construction, we can use them for stratified data by using the conditioning variables as first splitting variables. The mosaic in Fig. **??** shows the distribution of hair and eye color for both male and female students. A larger spacing for the gender variable has been used to simplify the distinction between the two groups. However, using a single plot, the strata are not corrected for the marginal distribution of the conditioning variable(s). Especially for very unevenly distributed marginals, the strata can become very distorted, and the tiles, along with their shading, inscrutable. An alternative is to use trellis-like layouts for visualizing *partial* tables for given strata, defined by the conditioning variables. All subtables are corrected for the marginals and thus, all corresponding mosaics in the panels have the same size. In addition, trellis layouts help reducing the complexity of bigger tables. As an example, consider the well-known UCB admissions data [**?**] on applicants to graduate school at the University of California in Berkeley for the six largest departments in 1973 classified by admission and gender (see Tab. **??**). The table aggregated over all departments gives wrong evidence of gender bias, i.e., higher admission rates for male students. Figure **??** visualizes the data by means of a conditional mosaic plot. Each panel corresponds to a department, and contains a mosaic corresponding to the partial (fourfold) table of admission and gender. Accordingly, the shading visualizes the residuals from the corresponding conditional independence model (independence of gender and admission, given department), stratified by department. Clearly, the situation in department A (more women/less men accepted than would be expected under the null) causes the rejection of the hypothesis of conditional independence. The overall deviation pattern thus indicates that in contrast to the impression given by the aggregated table, male students were not advantaged, and that in department A, *female* students were even more successful than their male colleagues.

| | | **Department** | | | | | |
|---|---|---|---|---|---|---|---|
| **Gender** | **Admit** | A | B | C | D | E | F |
| Male | Admitted | 512 | 353 | 120 | 138 | 53 | 22 |
| | Rejected | 313 | 207 | 205 | 279 | 138 | 351 |
| Female | Admitted | 89 | 17 | 202 | 131 | 94 | 24 |
| | Rejected | 19 | 8 | 391 | 244 | 299 | 317 |

**Table 8.** The UCB admissions data, in flat representation.

### 4.3 Summary

Mosaic, association, and sieve plots can be used for multi-way tables by applying them to flat representations. In addition, these basic plots are leveraged by specialized frameworks for exploratory model search (pairs plots) and stratified analysis (conditional plots).

## 5 Conclusion

This chapter reviews several alternatives for the visualization of multi-way contingency tables. For two-way tables, mosaic, sieve, and association plots are suitable for the visualization of observed and expected values and the Pearson residuals, respectively. This basic methods can be enhanced by using residual-based shadings, preferably based on perceptual color palettes such as those derived from the HCL space. Residual-based shadings can be used to visualize sign and size of the residuals, as well as the significance of test statistics such as the $\chi^2$ or the maximum test statistic. The latter has the advantage of detecting residuals causing the rejection of the hypothesis of independence. The methods directly extend to the multi-way case by using 'flat' representations of the multi-way tables, and specialized displays for conditional independence such as trellis layouts of partial tables and pairs plots.

**Fig. 14.** Mosaic plot for the Arthritis data, using the $\chi^2$ test and fixed cut-off points for the shading.

**Fig. 15.** Mosaic plot for the Arthritis data, using the maximum test and data-driven cut-off points for the residuals.

**Fig. 16.** Mosaic plot for the HairEyeColor data.

**Fig. 17.** Sieve plot for the HairEyeColor data.

**Fig. 18.** Association plot for the HairEyeColor data.

**Fig. 19.** Pairs plot for the HairEyeColor data.

**Fig. 20.** Conditional mosaic plot for the UCB admissions data.