Residual-based Shadings for Visualizing (Conditional) Independence

Achim Zeileis, David Meyer, and Kurt Hornik

Visualizing independence by association and mosaic plots.

Enhancements: diverging color palette based on HCL color space combined with visualization of the result of a significance test.

Extensions to visualizing conditional independence.

Key Words: Association plots; Conditional inference; Contingency tables; HCL colors; HSV colors; Mosaic plots.

1. INTRODUCTION

independence in 2-way tables

association and mosaic displays

colors and color spaces because the three perceptual dimensions (hue, lightness, saturation) are not properly mapped to the three dimensions of the color space and hence are confounded (Brewer 1999)

somewhere: mention Augsburg stuff, visualization of log-linear models via mosaics (Theus and Lauer 1999; Hofmann 2003, 2001).

Association plots as residual plots for log-linear models (Meyer, Zeileis, and Hornik 2003), especially in coplot or trellis layout.

2. INDEPENDENCE IN 2-WAY TABLES

In this section, the basic tools for testing and visualizing independence in 2-way tables are briefly reviewed. For illustration, a data set about treatment and improvement of patients with rheumatoid arthritis from Koch and Edwards (1988) is used. The data set is also discussed in Friendly (2000) and the subset of the 59 female patients from the study is given in Table 1.

2.1 Tests

To fix notations, we consider a 2-way contingency table with 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. Given an underlying distribution with theoretical cell probabilities π_{ij}, the null hypothesis of independence of the two categorical variables can be formulated as

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

The expected cell frequencies in this model are $\hat{n}_{ij} = n_{i+}n_{+j}/n_{++}$. The probably best known and most used measure of discrepancy between observed and expected values are the Pearson residuals

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

Table 1: Treatment and improvement among 59 patients with rheumatoid arthritis.

		Improved		
		None	Some	Marked
Treatment	Placebo	19	7	6
	Treated	6	5	16

The most convenient way to aggregate the $I \times J$ residuals to one test statistic is their squared sum

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

because this is known to have an unconditional limiting χ^2 distribution with $(I-1)(J-1)$ degrees of freedom under the null hypothesis. This is the well-known χ^2 test which is typically introduced at the very beginning of the chapter about independence in 2-way tables in statistics textbooks (see e.g., Agresti 2002).

But the sum of squares is not the only plausible way of capturing deviations from 0 in the residuals. There are many other conceivable functionals $\lambda(\cdot)$ which lead to reasonable test statistics $\lambda([r_{ij}])$: one which is particularly suitable for identifying the cells which cause the dependence (if any) is the maximum of the absolute values

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

Given a critical value c_α for this test statistic, all residuals whose absolute value exceeds c_α violate the hypothesis of indendence at significance level α (Mazanec and Strasser 2000, ch. 7). Thus, the interesting cells causing the dependence can easily be identified.

Furthermore, an important reason for using the unconditional limiting distribution for the X^2 statistic from (3) 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 (Pesarin 2001; Strasser and Weber 1999). For testing the independence hypothesis from (1), 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 $\lambda([r_{ij}])$.

Other measures of discrepancy (e.g., deviance residuals) could, of course, also be used instead of the the Pearson residuals $[r_{ij}]$. But as the ideas discussed here extend straightforwardly to that situation, we do not go into detail about this.

For the arthritis data from Table 1, both tests indicate a clearly significant dependence of improvement on treatment: the sum-of-squares statistic from (3) is $X^2 = 11.296$ with a p value of $p = 0.0032$, and the maximum statistic from (4) is $M = 1.87$ with $p = 0.0096$. Both p values have been computed from a sample of size 5000 from the permutation distribution under independence generated via sampling tables with the same row and column sums n_{i+} and n_{+j} using the Patefield (1981) algorithm and computing the maximum statistic for each of these tables.

2.2 Visualizations

Two well-established visualization techniques for independence in 2-way tables are mosaic plots and association plots. Both are suitable to bring out departures of an observed table $[n_{ij}]$ from the expected table $[\hat{n}_{ij}]$ in a graphical way. The latter focuses on the visualization of the Pearson residuals r_{ij} (under independence) while the former primarily displays the observed frequencies n_{ij}.

Mosaic plots (Hartigan and Kleiner 1981) can be seen as an extension of grouped bar charts where width and height of the bars show the relative frequencies of the two variables: a mosaic plot simply consists of a collection of tiles with areas proportional to the observed cell frequencies as shown in the left panel of Figure 1. A rectangle corresponding to 100 percent of the observations is first split horizontally with respect to the treatment frequencies and then vertically with respect to the conditional improvement frequencies. This shows that there have been more placebo than treated patients with no improvement and vice versa for marked improvement. This strategy of splitting with respect to conditional frequencies given all previous variables can also directly be used for visualizing multi-way tables (see Hofmann 2003, for an overview of how to construct and read mosaic displays).

Association plots (Cohen 1980) visualize the table of Pearson residuals: 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 association plot for the arthritis data is shown in the right panel of Figure 1 which leads to the same interpretation as the mosaic plot: there are more placebo patients with no improvement and fewer with marked improvement than expected under independence—vice versa for the treated patiens.

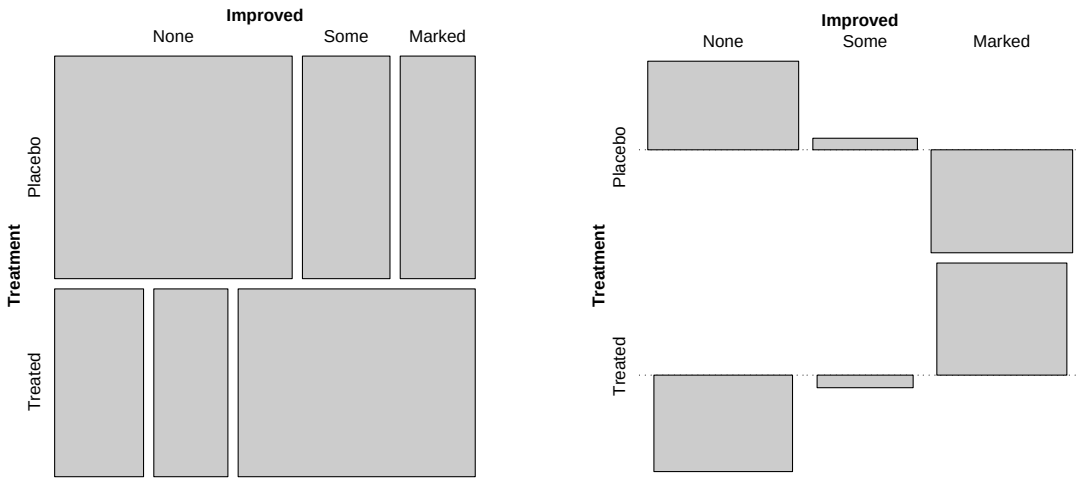

Figure 1: Classic mosaic and association plot for the arthritis data.

3. RESIDUAL-BASED SHADINGS

Colors are commonly used to enhance mosaic and association plots. To integrate a visualization of the residuals $[r_{ij}]$ into the mosaic display—which in its 'raw' version only visualizes the observed frequencies $[n_{ij}]$—Friendly (1994) suggested a residual-based shading for the mosaic tiles that can also be applied to the rectangles in association plots (Meyer et al. 2003). In this section, we first briefly review the Friendly (1994) shading, before we suggest different colors and a combination of visualization and significance testing to extend these residual-based shadings.

3.1 Friendly Shading

The extensions of Friendly (1994) to mosaic plots provide a substantial improvement of the original mosaic plots enhancing them from a plot for contingency tables to a visualization technique for log-linear models—and thus also for independence problems including 2-way tables as the simplest case.

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 ($|r_{ij}| < 2$) are shaded white. 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. Mosaic plots enhanced by this shading can thus also bring out departures from independence (or other log-linear models in multi-way tables) graphically and visualize patterns of dependence. The heuristic for choosing the cut offs 2 and 4 is that the Pearson residuals are approximately 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 shading is not to visualize significance but the *pattern* of deviation from independence (Friendly 2000, p. 109).

In addition to the shading of the rectangles themselves, the Friendly shading also encompasses a choice of line type and line color of the borders of the rectangles with similar ideas as described above. As both mosaic and association plots are area-proportional visualization techniques, we focus on area shadings and always use solid black borders throughout this paper, but the extensions suggested in the following could also be applied to control line type and color.

3.2 Colors

The way the (light) blue and red colors are chosen differs somewhat between various implementations of Friendly mosaic plots: In his original SAS implementation (see Friendly 2000), Michael Friendly uses colors from a palette based on HLS (Hue–Luminance–Saturation) color space. The implementation in the R system for statistical computing and graphics (R Development Core Team 2005) employs colors from HSV (Hue–Saturation–Value) space. The latter is a very common implementation of colors in many computer packages and makes the generation of the Friendly shading very simple.

The HSV space originally looks like a cone with black at its peak (zero value) and full color wheels for different saturations at the other end, centered around a white (full value). Type and amount of color are controlled by hue and saturation, respectively. Typically, polar coordinates (h, s, v) rescaled to the unit interval are used in this space, giving it the appearance of a cylinder. For generating colors in the Friendly shading, the following strategy is used: The hue h codes the sign of the residuals—$h = 0$ (red hue) is used for negative residuals, $h = 2/3$ (blue hue) for positive residuals. The absolute size of the residuals is then coded by the saturation s which is set to 0, 0.5 and 1, respectively, for small/medium/large residuals. The value is always fixed at $v = 1$. This is also depicted in the upper panel of Figure 2 which shows two slides from the HSV cylinder side by side. The plot shows the saturation/value plane for the given hues $h = 0$ and $h = 2/3$. The significant color palette shows the colors used for the Friendly shading when the residuals are increasing from left to right. The set of non-significant colors will be explained in Section 3.3.

Although this HSV-based shading is already very useful for enhancing mosaic and association plots and although HSV is a very commonly available implementation of color spaces, HSV colors in general and the Friendly shading in particular have a number of 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 (Brewer 1999; Ihaka 2003). Consequently, the HSV dimensions are confounded, e.g., saturation is not uniform across different hues. A fully saturated blue $(2/3, 1, 1)$ is perceived to be much darker than a fully saturated red $(0, 1, 1)$ or green $(1/3, 1, 1)$. This makes it more difficult for the human eye to judge the size of shaded areas and can therefore lead to color-caused optical illusions when used in statistical graphs (Cleveland and McGill 1983). Furthermore, flashy fully saturated HSV colors are good for drawing attention to a plot, but hard to look at for a longer time (Ihaka 2003) which makes HSV-shaded graphics harder to interpret. Finally, white is employed as the neutral color for small residuals in the Friendly

shading, but typically grey is found to convey neutrality or un-interestingness much better than white (Brewer 1999).

Alternative ways to choose colors have been available for a long time, but have been only slowly adopted for implementations of colors in computer packages in general and for shading in statistical graphs in particular. The idea of using perceptually based colors that are 'in harmony' go back to Munsell (1905) who introduced a color notation for balanced colors. Based on similar principles, Cynthia Brewer and co-workers suggested different types of palettes (qualitative/sequential/diverging) and provided the online tool **ColorBrewer.org** (Harrower and Brewer 2003) for selecting an appropriate palette for a specific problem. Furthermore, the Commission Internationale de l'Éclairage (CIE) introduced the two perceptually based color spaces CIELAB and CIELUV where the latter is typically preferred for emissive color technologies such as computer displays. Ihaka (2003) discusses how CIELUV colors can be used for choosing qualitative palettes for statistical graphics such as barplots. By taking polar coordinates in CIELUV space, it is called HCL (Hue–Chroma–Luminance) space and qualitative palettes can easily be chosen by using a range of hues for fixed values of chroma and luminance. Such colors are always balanced towards the same grey and thus do not have the problem of varying saturations as the HSV colors.

Here, we discuss how similar ideas can be used for deriving diverging HCL palettes that provide a suitable translation of the ideas from the Friendly shading to perceptually uniform HCL colors. The HCL space looks like a distorted double cone with black (zero luminance) at one end and white at the other (full luminance). In its middle, there is a full color wheel for different values of chroma (that controls the colorfulness). Unfortunately, the HCL space is not as regular as the HSV space and consequently cannot be standardized in the same way as the HSV space. Its dimensions are usually given by a hue ranging in $[0, 360]$ degrees and chroma and luminance ranging in $[0, 100]$ percent. But not all combinations (h, c, l) yield valid HCL colors and the admissable combinations of c and l vary across different hues h. But for the problem of constructing a diverging palette, this problem can easily be overcome as we just need two different hues (a 'negative' and a 'positive' hue) and hence we can choose two hues that correspond to similar shapes in the chroma/luminance plane. The lower panel of Figure 2 shows two such planes side by side for the hues $h = 0$ and $h = 260$. To obtain a sequence of colors with the same properties as the Friendly shading, the palette starts at a fully saturated red $(0, 100, 50)$, goes via a neutral color, ends at a fully saturated blue $(260, 100, 50)$, and uses linear interpolation in between. Instead of using white $(0, 0, 100)$ as the neutral color, a light grey $(0, 0, 90)$ is employed as motivated above.

This diverging palette (see the 'significant' colors in Figure 2) uses both chroma—i.e., the colorfulness—and luminance—i.e., the amount of grey—to code the absolute size of the quantity visualized—i.e., the residuals r_{ij}—when applied to the independence problem. By changing the neutral color or by changing the maximum chroma, respectively, this can be changed to using only chroma or luminance for this purpose, but using both seems to be a very effective way of visualization.

3.3 Significance

The shading scheme of Friendly (1994) was suggested to visualize the pattern of dependence in contingency tables, as discussed above, but the presence (or absence) of colors in a plot also always conveys an impression of interestingness (or un-interstingness, respectively). That is, viewers might be tempted to interpret the absence of color in a plot as a clue that there is no significant departure from independence. Or vice versa, colored cells would convey the impression that there is significant dependence. Currently, both is not true as can be seen in the left panel of Figure 3 which shows the mosaic display for the arthritis data with Friendly shading. Although there is significant dependence, no residual exceeds an absolute value of 2 and hence no cell is colored. Of course, it can be argued that the shading was not designed for this purpose and that different cut offs than 2 and 4 should be used here. But then again, it would be nice if these cut offs could be chosen automatically in a data-driven fashion which is what we do below.

The Friendly shading can be interpreted to be a visualization of the maximum statistic M from (4) which always employs the critical values c_α 2 and 4. But the problem is that it is not

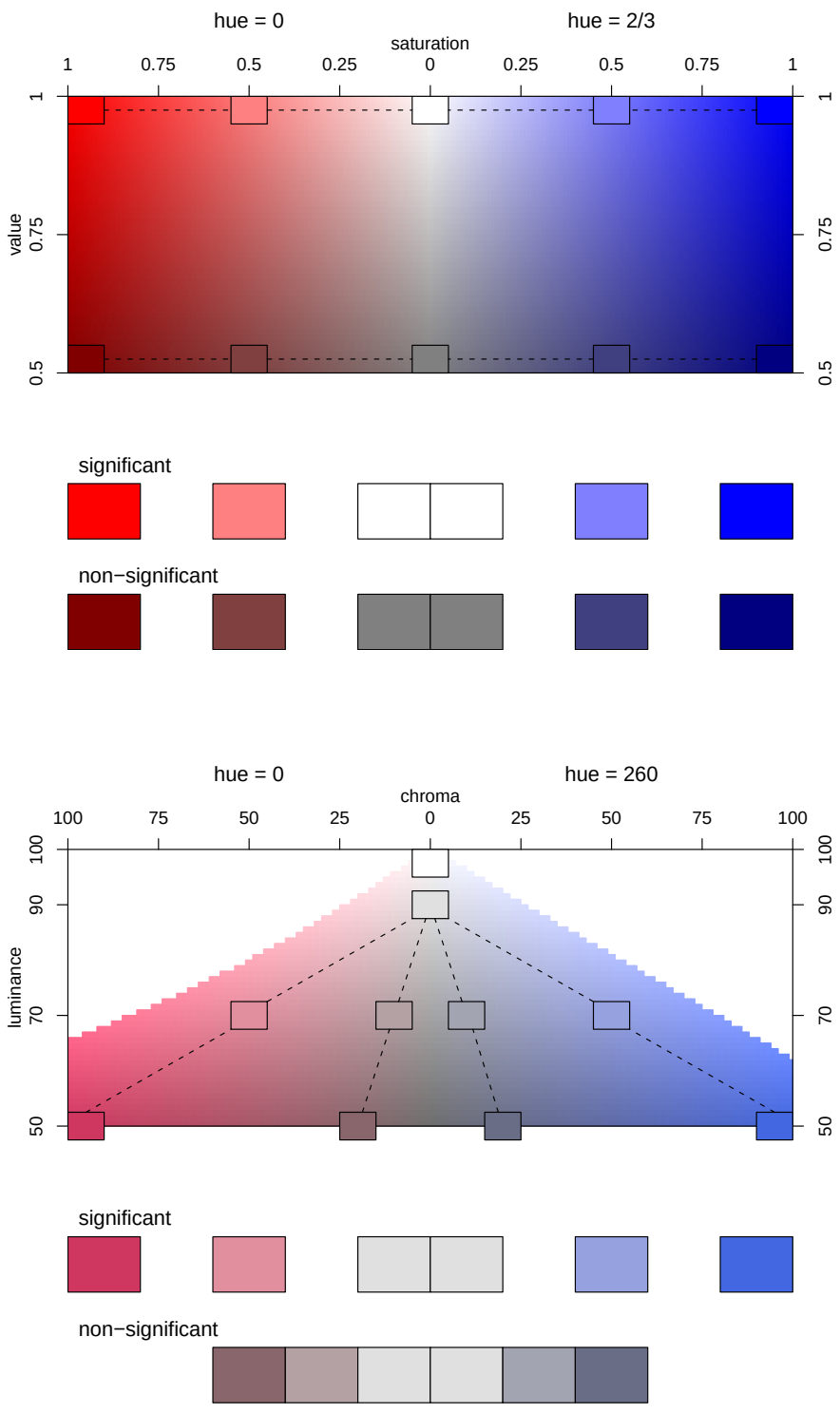

Figure 2: Extended shading in HSV (upper) and HCL space (lower).

clear to which significance levels α these critical values correspond because the distribution of M depends on the underlying contingency table. The natural solution to this problem is to compute the critical values from the distribution of M in a data-driven way (i.e., for the table visualized) and use these instead of the hard-coded values 2 and 4. In the right panel of Figure 3 this is done for the arthritis data by employing the critical values 1.24 at level $\alpha = 0.1$ and 1.64 at level $\alpha = 0.01$ (and using the diverging HCL palette) derived from the permutation distribution of M for the arthritis data as described in Section 2.1. By using these cut offs, the presence of color in the plot is equivalent to significance (of the maximum statistic M) at level $\alpha = 0.1$ and $\alpha = 0.01$, respectively, and exactly the cells which 'cause' the dependence are highlighted. For the arthritis data, these are in particular the cells in the last column that signal that there are significantly more treated patients and fewer placebo patients with marked improvement than would be expected under independence.

The significance levels $\alpha = 0.1$ and $\alpha = 0.01$ are chosen because this leads to displays where fully colored cells are clearly significant ($p < 0.01$), cells without color are clearly non-significant ($p > 0.1$), and cells in between can be considered to be weakly significant ($0.01 \leq p \leq 0.1$). Of course, users could choose any other set of significance levels they feel comfortable with, e.g., only a single cut off at $\alpha = 0.05$ or three cut offs at 0.1, 0.05 and 0.01 etc. Another option could be to use a continuous shading where the p value corresponding to a cell controls the interpolation between the neutral and the full color. However, this typically results in too much color in the plot which in turn tends to conceal the important cells and over-emphasize the unimportant ones. Hence, a discrete shading with few colors is much easier to interpret.

This maximum shading is already very flexible and combines visualization and inference. However, it can only be applied when employing the maximum statistic because it is the only aggregation functional $\lambda(\cdot)$ where a single large residual $|r_{ij}|$ exceeding its critical value is equivalent to a significant value of the whole test statistic $\lambda([r_{ij}])$.[1] Typically, applying the maximum statistic is feasible and also appropriate for exploratory analysis, but it would be desirable to also have a residual-based shading that can incorporate visualization of significance when the sum-of-squares statistic X^2 (or any other functional $\lambda(\cdot)$) is used. For the reasons discussed above, it is not possible to achieve this by shading individual cells differently but can only be realized by using different colors for the whole table. As outlined before, colorfulness is intuitively matched with interestingness, therefore a rather natural idea is to use the fully colored palette only when the corresponding test is significant and to use a less colorful palette if not. For the HCL scheme, the amount of color can conveniently be controlled by varying the maximum chroma value used. For the full colors, the maximum chroma was set to 100 as shown in Figure 2 and is reduced to 20 for the non-significant palette. This palette still codes the absolute size of the residuals by luminance (i.e., the amount of grey), uses the same neutral grey for small 'un-interesting' residuals, codes positive and negative residuals by different hues, but gives less emphasis to the pattern by making the plot less colorful. A similar effect can be obtained in the HSV space if the value is reduced from 1 to 0.5 for non-significant palettes (see Figure 2). But as for the significant colors, the dimensions used for creating this palette are confounded and hence the HCL scheme is clearly preferable.

4. EXTENSION TO CONDITIONAL INDEPENDENCE

To introduce the new residual-based shadings without too much overhead in Section 3, we have only considered the independence problem in 2-way tables. But, in fact, none of the ideas described are really confined to the 2-dimensional case as will be outlined in this section.

Mosaic displays have been emphasized in the literature to be an excellent means of visualization for log-linear models (Friendly 1999; Theus and Lauer 1999). Problems of interest include complete, joint or conditional independence for all of which the tables of expected values and residuals

[1] This can easily be seen from the following consideration: To obtain a level α test, where any residual $|r_{ij}|$ exceeding the critical value c_α gives a significant test result, the equation $\alpha = P(\text{any } |r_{ij}| > c_\alpha)$ must hold. This is equivalent to $1 - \alpha = P(\text{all } |r_{ij}| \leq c_\alpha) = P(\max_{i,j} |r_{ij}| \leq c_\alpha)$.

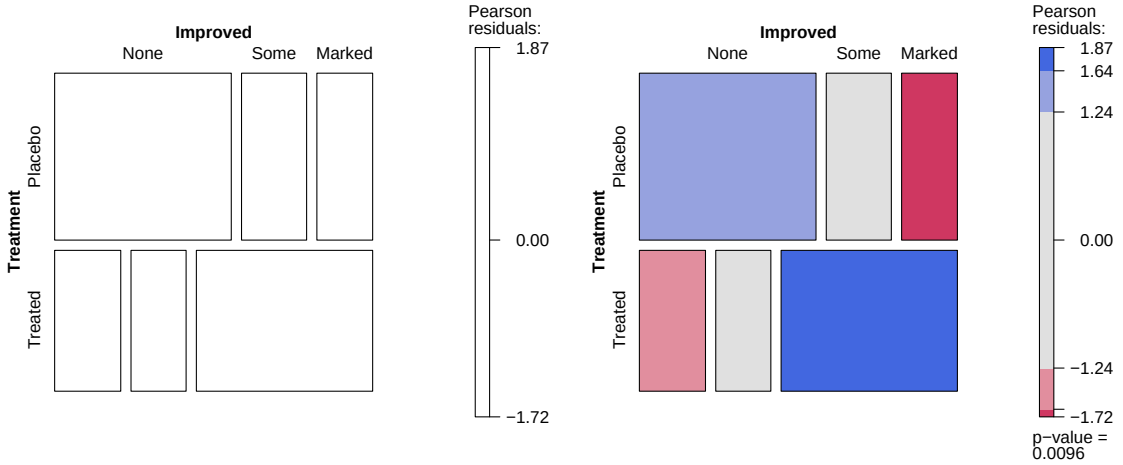

Figure 3: Mosaic plot for the arthritis data with Friendly shading (left) and maximum shading (right).

(again Pearson or deviance) can be computed. Friendly (1994, 1999) shows that his residual-based shading scheme can directly be applied to these more complex independence models. For inference, the most commonly used aggregation functional for the residuals is again the sum of squares yielding the associated Pearson or likelihood ratio statistic, respectively (Agresti 2002). Hence, our residual-based shading with different HCL palettes for significant and non-significant results can directly be applied to these models. For the maximum shading with data-driven cut offs, even more flexibility is possible than in the 2-way case when the structure of the independence problem is exploited which will be discussed below for conditional independence

Association plots are not commonly used for contingency tables with more than two margins, although there is nothing in the definition that would prevent application in higher dimensions. However, as argued for the mosaic plots by Friendly (1999) and Theus and Lauer (1999), it becomes increasingly important to choose a good layout as the number of variables grows.

To show how the structure of the independence problem can be exploited for choosing appropriate critical values in the maximum shading and for selecting a suitable layout of mosaic or association plots, we consider the conditional independence problem. To keep the notational overhead low again, only the 3-way case is spelt out, but the same ideas extend fairly straightforwardly to conditional independence in higher dimensions. In a 3-way table $[n_{ijk}]$ with underlying theoretical probability distribution $[p_{ijk}]$ the conditional independence problem can be formulated as

$$H_0 : \pi_{ij|k} = \pi_{i+|k}\pi_{+j|k}. \tag{5}$$

where $\pi_{ij|k}$ are the conditional probabilities given the stratum k with $k = 1, \ldots, K$. Under the assumption of conditional independence, we can again compute expected frequencies $[\hat{n}_{ijk}]$ and the corresponding residuals $[r_{ijk}]$. To test the conditional independence hypothesis, usually the sum-of-squares statistic is used

$$\sum_{i,j,k} r_{ijk}^2 = \sum_k X_k^2, \tag{6}$$

which is simply the sum of the individual sum-of-squares statistics X_k^2 in each stratum k.

Alternatively, a maximum statistic similar to that from Equation (4) can be constructed

$$\max_{i,j,k} |r_{ijk}| \quad = \quad \max_{k} M_k. \tag{7}$$

As in the 2-way case, this allows for identification of the cells that are responsible for the deviation from conditional independence (if any). If it is not so much of interest in which *cell* but only in which *stratum* the deviation occurs, then it would be natural to use

$$\max_{k} \sum_{i,j} r_{ijk}^2 \quad = \quad \max_{k} X_k^2. \tag{8}$$

Given a critical value for this statistic, all strata k whose associated sum-of-squares statistics X_k^2 exceed the critical value are in conflict with the hypothesis of conditional independence.

All these statistics are of type

$$\lambda_{\text{agg}}(\lambda_{\text{indep}}(r_{ijk})), \tag{9}$$

where λ_{indep} is a functional for assessing independence in stratum k and λ_{agg} is a functional for aggregating over the $k = 1, \ldots, K$ strata. If the maximum is used for the latter, then identification of the strata causing the independence is possible. If additionally λ_{indep} is the maximum, the corresponding cells can also be identified. Hence, the double maximum statistic from Equation (7) is the only functional allowing for detection of both the strata and the cells which caused the dependence.

But the main purpose of the formulation of the different test statistics is not so much inference but their applicability to diagnostic plots via residual-based shadings. As already discussed, it is possible for all aggregation functionals to simply use either the significant or the non-significant shading for all cells in the contingency table—this strategy would have to be used for the sum-of-squares statistic from Equation (6). If λ_{agg} is the maximum as in Equation (8), then the significant palette would only be used in those strata in conflict with the hypothesis of conditional independence whereas the non-significant palette would be used for the remaining strata. Finally, if both λ_{agg} and λ_{indep} are the maximum, then the same strategy as in Section 3 can be pursued, i.e., only the significant palette is used but with data-driven cut offs derived from the distribution of the double maximum statistic from Equation (7).

To arrange the shaded rectangles of the association or the mosaic plot, respectively, the most intuitive approach is to use the same conditioning in the display that was also used for conditioning in the model. For this situation, Friendly (1999) discusses a grouping similar to coplots (conditioning plots, see Cleveland 1993) that lead to trellis graphics (Becker, Cleveland, and Shyu 1996). Thus, a natural visualization of such an independence model would be a trellis-like coplot where each stratum k could be visualized by an association or mosaic display. This has also the advantage that only the conditional independence problem but not the conditioning distribution over $k = 1, \ldots, K$ is visualized which could obscure departures from conditional independence if the number of observations in each stratum n_{++k} are very different.

For illustration, the famous admissions data of the University of California at Berkley (UCB) from Bickel, Hammel, and O'Connell (1975) are employed. Testing the conditional independence of admission and gender given the department leads to the well-known result that there is no gender discrimination in five of six departments, but that women are more likely to be admitted than men in the first department. This departure from conditional independence is picked up by all three statistics formulated above: the sum-of-squares statistic from Equation (6) is $\sum_k X_k^2 = 19.938$ ($p = 0.0034$), the maximum sum-of-squares statistic from Equation (8) is $\max_k X_k^2 = 17.248$ ($p = 0.0006$), and the double maximum statistic from Equation (7) is $\max_k M_k = 3.134$ ($p = 0.0004$). The p values are again computed by drawing 5000 samples from the corresponding permutation distribution. The latter two test statistics both indicate the first department as the source of departure and the maximum statistic additionally identifies the deviations in the female applicants group. This result is also visualized by the conditional association plot in a trellis layout based on the double maximum shading in Figure 4. This can be interpreted as a diagnostic residual plot

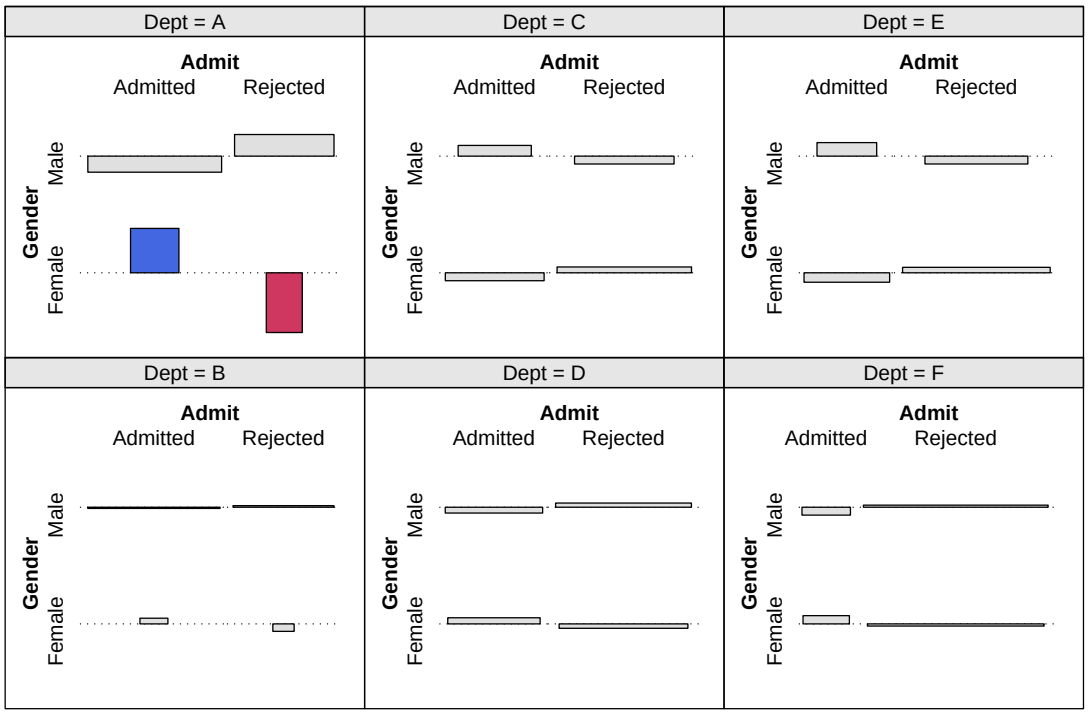

Figure 4: Conditional association plot for UCB admissions with double maximum shading.

for the associated log-linear model because it shows that there are significant departures from the model in department A but not in departments B to F where all residuals are very small.

5. CONCLUSIONS

Various strategies for constructing residual-based shadings for visualizing (conditional) independence in contingency tables via mosaic and association plots are discussed. The shading of Friendly (1994) is extended in two directions: the use of perceptually uniform HCL colors and the combination of visualization and significance testing. To achieve the former, a general guideline for constructing diverging palettes in HCL space is introduced. The advantages of using this HCL shading scheme instead of an HSV scheme are that the colors from this perceptually based color space are device independent and provide uniform saturations over different hues. Furthermore, the colorfulness in this shading can be controlled independently from the other two dimensions.

To combine visualization and significance testing, two approaches are presented: The first approach always uses a fully colored palette but relies on data-driven cut offs such that the presence of color is equivalent to significance of the associated maximum test. The second approach uses pre-defined cut offs (such as 2 and 4) but codes the result of the significance test by using full colors only for significant results and the same type of palette but with a reduced amount of color for non-significant results.

Both results can not only be applied to the simple independence problem in 2-way contingency tables, but also to arbitrary inpendence models fitted via log-linear models in higher dimensions. In addition, it might be possible to exploit the structure of a given independence problem to achieve better visualizations which is illustrated for the conditional independence problem.

All associated significance tests are carried out using a conditional inference approach instead of relying on unconditional asymptotic results.

COMPUTATIONAL DETAILS

The results in this paper were obtained using R 2.1.1 (R Development Core Team 2005, http://www.R-project.org/) and the packages **vcd** 0.9-1 (Meyer, Zeileis, and Hornik 2005), **MASS** 7.2-16 (Venables and Ripley 2002), **grid** 2.1.1 (Murrell 2002) and **colorspace** 0.9 (Ihaka 2004).

REFERENCES

Agresti, A. (2002), *Categorical Data Analysis*, Hoboken, New Jersey: John Wiley & Sons, 2nd ed.

Becker, R. A., Cleveland, W. S., and Shyu, M.-J. (1996), "The Visual Design and Control of Trellis Display," *Journal of Computational and Graphical Statistics*, 5, 123–155.

Bickel, P., Hammel, E., and O'Connell, J. (1975), "Sex Bias in Graduate Admissions: Data from Berkeley," *Science*, 187.

Brewer, C. A. (1999), "Color Use Guidelines for Data Representation," in *Proceedings of the Section on Statistical Graphics, American Statistical Association*, Alexandria, VA, pp. 55–60.

Cleveland, W. S. (1993), *Visualizing Data*, Summit, New Jersey: Hobart Press.

Cleveland, W. S. and McGill, R. (1983), "A Color-Caused Optical Illusion on a Statistical Graph," *The American Statistician*, 37, 101–105.

Cohen, A. (1980), "On the Graphical Display of the Significant Components in a Two-Way Contingency Table," *Communications in Statistics—Theory and Methods*, A9, 1025–1041.

Friendly, M. (1994), "Mosaic Displays for Multi-Way Contingency Tables," *Journal of the American Statistical Association*, 89, 190–200.

— (1999), "Extending Mosaic Displays: Marginal, Conditional, and Partial Views of Categorical Data," *Journal of Computational and Graphical Statistics*, 8, 373–395.

— (2000), *Visualizing Categorical Data*, Carey, NC: SAS Insitute.

Harrower, M. A. and Brewer, C. A. (2003), "**ColorBrewer.org**: An Online Tool for Selecting Color Schemes for Maps," *The Cartographic Journal*, 40, 27–37.

Hartigan, J. A. and Kleiner, B. (1981), "Mosaics for Contingency Tables," in *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, ed. Eddy, W., New York: Springer, pp. 268–273.

Hofmann, H. (2001), "Generalized Odds Ratios for Visual Modelling," *Journal of Computational and Graphical Statistics*, 10, 1–13.

— (2003), "Constructing and Reading Mosaicplots," *Computational Statistics & Data Analysis*, 43, 565–580.

Ihaka, R. (2003), "Colour for Presentation Graphics," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria*, eds. Hornik, K., Leisch, F., and Zeileis, A., ISSN 1609-395X.

— (2004), **colorspace***: Colorspace Manipulation*, R package version 0.9.

Koch, G. and Edwards, S. (1988), "Clinical Efficiency Trials with Categorical Data," in *Biopharmaceutical Statistics for Drug Development*, ed. Peace, K. E., New York: Marcel Dekker, pp. 403–451.

Mazanec, J. A. and Strasser, H. (2000), *A Nonparametric Approach to Perceptions-based Market Segmentation: Foundations*, Berlin: Springer.

Meyer, D., Zeileis, A., and Hornik, K. (2003), "Visualizing Independence Using Extended Association Plots," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria*, eds. Hornik, K., Leisch, F., and Zeileis, A., ISSN 1609-395X.

— (2005), **vcd***: Visualizing Categorical Data*, R package version 0.9-1.

Munsell, A. H. (1905), *A Color Notation*, Boston, Massachusetts: Munsell Color Company.

Murrell, P. (2002), "The **grid** Graphics Package," *R News*, 2, 14–19.

Patefield, W. M. (1981), "An Efficient Method of Generating $R \times C$ Tables with Given Row and Column Totals," *Applied Statistics*, 30, 91–97, Algorithm AS 159.

Pesarin, F. (2001), *Multivariate Permutation Tests*, Chichester: John Wiley & Sons.

R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-00-3.

Strasser, H. and Weber, C. (1999), "On the Asymptotic Theory of Permutation Statistics," *Mathematical Methods of Statistics*, 8, 220–250.

Theus, M. and Lauer, S. R. W. (1999), "Visualizing Loglinear Models," *Journal of Computational and Graphical Statistics*, 8, 396–412.

Venables, W. N. and Ripley, B. D. (2002), *Modern Applied Statistics with S*, New York: Springer, 4th ed., ISBN 0-387-95457-0.

A. REPLICATION OF RESULTS IN R

```
## random seed
rseed <- 1071

## load packages and data
library("vcd")
art <- xtabs(~ Treatment + Improved, data = Arthritis, subset = Sex == "Female")
ucb <- aperm(UCBAdmissions)

## Figure 1
mosaic(art)
assoc(art)

## Figure 3
mosaic(art, gp = shading_Friendly(lty = 1, eps = NULL))
set.seed(rseed)
mosaic(art, gp = shading_max)

## Figure 4
set.seed(rseed)
cotabplot(~ Admit + Gender | Dept, data = ucb, panel = cotab_coindep,
          n = 5000, type = "assoc")

## Inference
## X^2 (3)
set.seed(rseed)
coindep_test(art, n = 5000)
## M (4)
set.seed(rseed)
coindep_test(art, n = 5000, indepfun = function(x) sum(x^2))

## sum X^2 (6)
set.seed(rseed)
coindep_test(ucb, "Dept", n = 5000, indepfun = function(x) sum(x^2), aggfun = sum)
## max M (7)
set.seed(rseed)
coindep_test(ucb, "Dept", n = 5000)
## max X^2 (8)
set.seed(rseed)
coindep_test(ucb, "Dept", n = 5000, indepfun = function(x) sum(x^2))
```