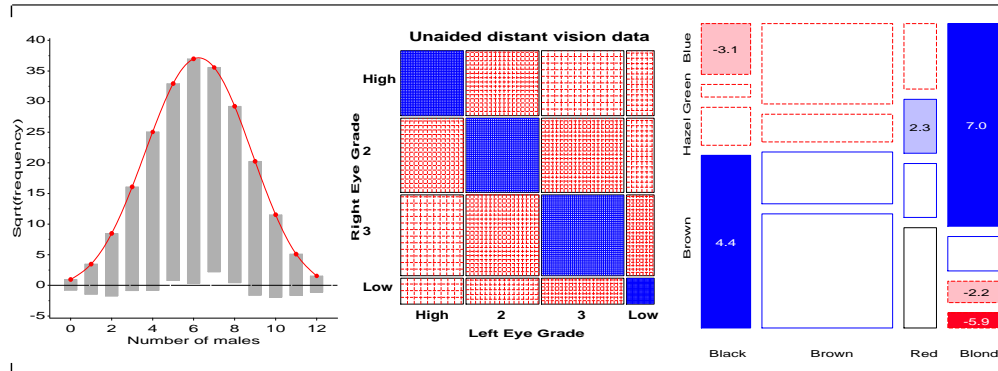# Visualizing Categorical Data with SAS and R

Michael Friendly
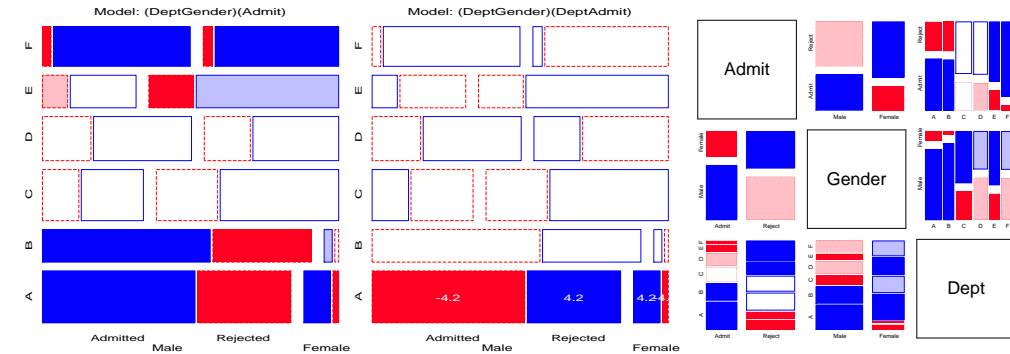
York University

SCS Short Course, 2016
Web notes: `datavis.ca/courses/VCD/`

Part 3: Mosaic displays and loglinear models

Topics:

- Mosaic displays
- loglinear models for n-way tables
- Visualizing loglinear models: SAS & R
- Models for square and structured tables
- Larger tables

---

Mosaic displays: Basic ideas

Hartigan and Kleiner (1981), Friendly (1994, 1999)

- Area-proportional display of frequencies in an n-way table
- Tiles (cells): recursive splits of a unit square—
  - V1: width $\sim$ marginal frequencies, n_{i++}
  - V2: height $\sim$ relative frequencies $|$ V1, n_{ij+}/n_{i++}
  - V3: width $\sim$ relative frequencies $|$ (V1, V2), n_{ijk}/n_{ij+}
  - $\cdots$
- $\Rightarrow$ area $\sim$ cell frequency, n_{ijk}

UCB Admissions: Observed frequencies
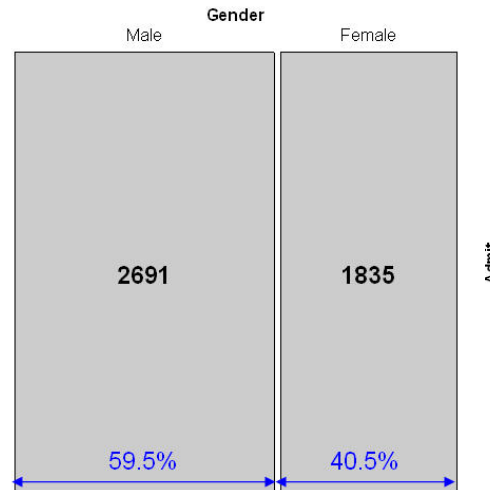
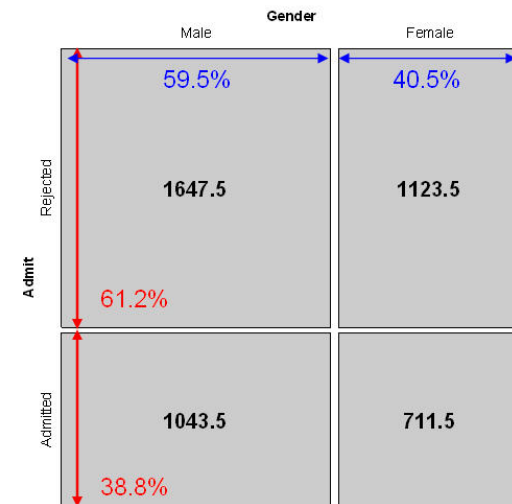# Mosaic displays: Basic ideas

- Independence: Two-way table
- Expected frequencies:

$$\widehat{m}_{ij} = \frac{n_{i+}n_{+j}}{n_{++}} = n_{++}\,\text{row \%col \%}$$

- $\Rightarrow$ rows & columns align when variables are independent
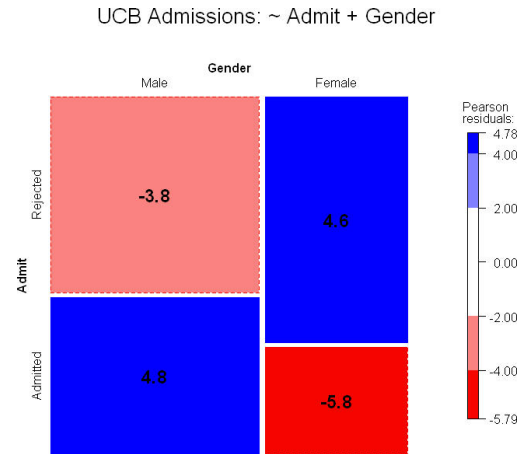

Independence: Expected frequencies

Mosaic displays: Residuals & shading

- Pearson residuals:

$$d_{ij} = \frac{n_{ij} - \widehat{m}_{ij}}{\sqrt{\widehat{m}_{ij}}}$$

- Pearson $\chi^2 = \Sigma\Sigma d_{ij}^2 = \Sigma\Sigma \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$
- Other residuals: deviance (LR), Freeman-Tukey (FT), adjusted (ADJ), ...
- Shading:
  - Sign: $-$ negative in red; $+$ positive in blue
  - Magnitude: intensity of shading: $|d_{ij}| > 0, 2, 4, \dots$
- $\Rightarrow$ Independence: rows align, or cells are empty!

UCB Admissions: ~ Admit + Gender

Loglinear models: Perspectives I

Loglinear approach

Loglinear models were first developed as an analog of classical ANOVA models, where *multiplicative* relations (under independence) are re-expressed in *additive* form as models for log(frequency).

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B \equiv [A][B] \equiv\sim A + B$$

- This expresses the model of independence for a two-way table (no A*B association)
- The notations $[A][B] \equiv\sim A + B$ are shorthands

Loglinear models: Perspectives II

GLM approach

More generally, loglinear models are also generalized linear models (GLMs) for log(frequency), with a Poisson distribution for the cell counts.

$$\log \mathbf{m} = \mathbf{X}\boldsymbol{\beta}$$

- This looks just like the general linear ANOVA, regression model, but for log frequency
- This approach allows quantitative predictors and special ways of treating ordinal factors

Loglinear models: Perspectives III

Logit models

When one table variable is a binary response, a logit model for that response is equivalent to a loglinearmodel (as discussed in Part 4).

$$\log(m_{1jk}/m_{2jk}) = \alpha + \beta_j^B + \beta_k^C \equiv [AB][AC][BC]$$

- $\log(m_{1jk}/m_{2jk})$ represents the log odds of response category 1 vs. 2
- The model formula includes only terms for the effects on A of variables B and C
- The equivalent loglinearmodel is [AB] [AC] [BC]
- The logit model assumes [BC] association, and [AB] $\to \beta_j^B$, [AC] $\to \beta_k^C$

Loglinear models: Overview

Two-way tables: Loglinear approach

For two discrete variables, A and B, suppose a multinomial sample of total size n over the IJ cells of a two-way $I \times J$ contingency table, with cell frequencies n_{ij}, and cell probabilities $\pi_{ij} = n_{ij}/n$.

- The table variables are statistically independent when the cell (joint) probability equals the product of the marginal probabilities, $\Pr(A = i \, \& \, B = j) = \Pr(A = i) \times \Pr(B = j)$, or,

$$\pi_{ij} = \pi_{i+}\pi_{+j} \ .$$

- An equivalent model in terms of expected frequencies, $m_{ij} = n\pi_{ij}$ is

$$m_{ij} = (1/n) \, m_{i+} \, m_{+j} \ .$$

- This multiplicative model can be expressed in additive form as a model for $\log m_{ij}$,

$$\log m_{ij} = -\log n + \log m_{i+} + \log m_{+j} \ . \tag{1}$$

Loglinear models: Overview

Independence model

By anology with ANOVA models, the independence model (1) can be expressed as

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B \ , \tag{2}$$

- μ is the grand mean of $\log m_{ij}$
- the parameters λ_i^A and λ_j^B express the marginal frequencies of variables A and B — "main effects"
- typically defined so that $\sum_i \lambda_i^A = \sum_j \lambda_j^B = 0$ as in ANOVA

Loglinear models: Overview

Saturated model

Dependence between the table variables is expressed by adding association parameters, λ_{ij}^{AB}, giving the *saturated model*,

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \equiv [AB] \equiv \sim A * B \ . \tag{3}$$

- The saturated model fits the table perfectly ($\widehat{m}_{ij} = n_{ij}$): there are as many parameters as cell frequencies. Residual df = 0.
- A global test for association tests $H_0 : \boldsymbol{\lambda}_{ij}^{AB} = \mathbf{0}$.
- If reject H_0, which $\lambda_{ij}^{AB} \neq 0$?
- For ordinal variables, the λ_{ij}^{AB} may be structured more simply, giving tests for ordinal association.

Example: Independence

Generate a table of Education by Party preference, strictly independent

```
educ <- c(50, 100, 50)  # row marginal frequencies
names(educ) <- c("Low", "Med", "High")

party <- c(20, 50, 30)  # col marginal frequencies
names(party) <- c("NDP", "Liberal", "Cons")

table <- outer(educ, party) / sum(party)   # row x col / n
names(dimnames(table)) <- c("Education", "Party")
table

##            Party
## Education NDP Liberal Cons
##      Low   10      25   15
##      Med   20      50   30
##      High  10      25   15
```

Example: Independence

All row (and column) proportions are the same:

```
prop.table(table,1)

##          Party
## Education NDP Liberal Cons
##      Low  0.2     0.5  0.3
##      Med  0.2     0.5  0.3
##      High 0.2     0.5  0.3
```

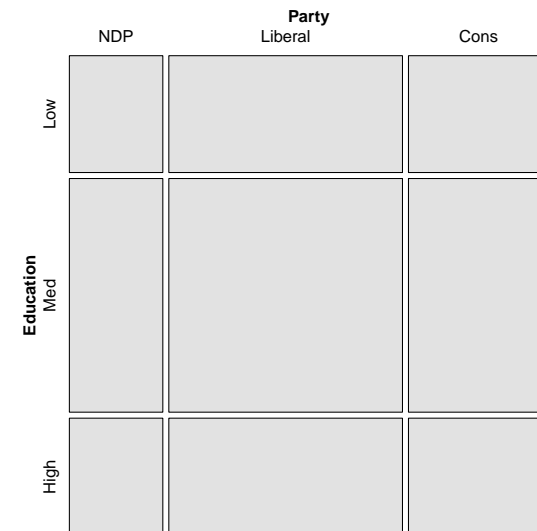All statistics are 0:

```
vcd::assocstats(table)

##                    X^2 df P(> X^2)
## Likelihood Ratio    0  4        1
## Pearson             0  4        1
##
## Phi-Coefficient   : 0
## Contingency Coeff.: 0
## Cramer's V        : 0
```

Mosaic plot shows equal row and column proportions:

```
library(vcd)
mosaic(table, shade=TRUE, legend=FALSE)
```

Two-way tables: GLM approach

In the GLM approach, the vector of cell frequencies, $\mathbf{n} = \{n_{ij}\}$ is specified to have a Poisson distribution with means $\mathbf{m} = \{m_{ij}\}$ given by

$$\log \mathbf{m} = \mathbf{X}\boldsymbol{\beta}$$

- $\mathbf{X}$ is a known design (model) matrix, expressing the table factors
- $\boldsymbol{\beta}$ is a column vector containing the unknown λ parameters.
- This is the same as the familiar matrix formulation of ANOVA/regression, except that
    - The response, $\log \mathbf{m}$ makes multiplicative relations additive
    - The distribution is taken as Poisson rather than Gaussian (normal)

Example: 2 x 2 table

For a 2×2 table, the saturated model (3) with the usual zero-sum constraints can be represented as

$$\log \begin{pmatrix} m_{11} \\ m_{12} \\ m_{21} \\ m_{22} \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{pmatrix} \mu \\ \lambda_1^A \\ \lambda_1^B \\ \lambda_{11}^{AB} \end{pmatrix}$$

- only the linearly independent parameters are represented. $\lambda_2^A = -\lambda_1^A$, because $\lambda_1^A + \lambda_2^A = 0$, and so forth.
- association is represented by the parameter λ_{11}^{AB}
- can show that $\lambda_{11}^{AB} = \frac{1}{4}\log(\theta)$ (log odds ratio)
- Advantages of the GLM formulation: easier to express models with ordinal or quantitative variables, special terms, etc. Can also allow for *over-dispersion*.

Assessing goodness of fit

Goodness of fit of a specified model may be tested by the likelihood ratio G^2,

$$G^2 = 2 \sum_i n_i \, \log \left(\frac{n_i}{\widehat{m}_i} \right) \quad , \tag{4}$$

or the Pearson X^2,

$$X^2 = \sum_i \frac{(n_i - \widehat{m}_i)^2}{\widehat{m}_i} \quad , \tag{5}$$

with degrees of freedom df = # cells - # estimated parameters.

- E.g., for the model of independence, $[A][B]$, df = $IJ - [(I-1) - (J-1)] = (I-1)(J-1)$
- The terms summed in (4) and (5) are the squared *cell residuals*
- Other measures of balance goodness of fit against parsimony, e.g., *Akaike's Information Criterion* (smaller is better)

$$AIC = G^2 - 2df \text{ or } AIC = G^2 + 2 \text{ \# parameters}$$

Three-way tables

Saturated model

For a 3-way table, of size $I \times J \times K$ for variables A, B, C, the saturated loglinear model includes associations between all pairs of variables, as well as a 3-way association term, λ_{ijk}^{ABC}

$$\begin{aligned} \log m_{ijk} = \mu &+ \lambda_i^A + \lambda_j^B + \lambda_k^C \\ &+ \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \quad . \end{aligned} \tag{6}$$

- One-way terms $(\lambda_i^A, \lambda_j^B, \lambda_k^C)$: differences in the *marginal frequencies* of the table variables.
- Two-way terms $(\lambda_{ij}^{AB}, \lambda_{ik}^{AC}, \lambda_{jk}^{BC})$ pertain to the *partial association* for each pair of variables, *controlling* for the remaining variable.
- The three-way term, λ_{ijk}^{ABC} allows the partial association between any pair of variables to vary over the categories of the third variable.
- Fits perfectly, but doesn't *explain* anything, so we hope for a simpler model!

Three-way tables: Reduced models

Reduced models

- Loglinearmodels are usually hierarchical: a high-order term, such as $\lambda_{ijk}^{ABC} \rightarrow$ *all* low-order relatives are automatically included.
- Thus, a short-hand notation for a loglinear model lists only the high-order terms,
- i.e., the saturated model (6) $\equiv [ABC]$, and implies all two-way and one-way terms
- The usual goal is to fit the *smallest* model (fewest high-order terms) that is sufficient to explain/describe the observed frequencies.
- This is similar to ANOVA/regression models with all possible interactions

Three-way tables: Reduced models

Reduced models

- For a 3-way table there are a variety of models between the mutual independence model, $[A][B][C]$, and the saturated model, $[ABC]$
- Each such model has an independence interpretation: $A \perp B$ means an hypothesis that A is independent of B.

Table: Log-linear Models for Three-Way Tables

Model	Model symbol	Interpretation
Mutual independence	$[A][B][C]$	$A \perp B \perp C$
Joint independence	$[AB][C]$	$(A\,B) \perp C$
Conditional independence	$[AC][BC]$	$(A \perp B) \mid C$
All two-way associations	$[AB][AC][BC]$	homogeneous assoc.
Saturated model	$[ABC]$	interaction

Three-way tables: Model types

- **Joint independence**: $(AB) \perp C$, allows A*B association, but asserts no A*C and B*C associations

$$[AB][C] \equiv \log m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB}$$

- **Conditional independence**: $A \perp B$, controlling for C

$$[AC][BC] \equiv \log m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

- **Homogeneous association**: All two-way, but each two-way is the *same* over the other factor

$$[AB][AC][BC] \equiv \log m_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

Goodness of fit tests

As noted earlier, overall goodness of fit of a specified model may be tested by the likelihood ratio G^2, or the Pearson X^2,

$$G^2 = 2 \sum_i n_i \log \left(\frac{n_i}{\widehat{m}_i} \right) \qquad X^2 = \sum_i \frac{(n_i - \widehat{m}_i)^2}{\widehat{m}_i} ,$$

with residual degrees of freedom $\nu = \#$ cells $- \#$ estimated parameters.

- These measure the lack of fit of a given model— a large value $\mapsto$ a poor model
- Both are distributed as $\chi^2(\nu)$ (in large samples: all $\widehat{m}_i > 5$)
- $\mathcal{E}(\chi^2) = \nu$, so G^2/ν (or X^2/ν) measures lack of fit per degree of freedom (overdispersion)
- But: how to compare or test competing models?

Nested models and ANOVA-type tests

Nested models

Two models, M_1 and M_2 are nested when one (say, M_2) is a special case of the other

- Model M_2 (with ν_2 df) fits a subset of the parameters of M_1 (with ν_1 df)
- M_2 is more restrictive — cannot fit better than M_1: $G^2(M_2) \geq G^2(M_1)$
- The least restrictive model is the saturated model [ABC...] with $G^2 = 0$ and $\nu = 0$

Therefore, we can test the difference in G^2 as a specific test of the added restrictions in M_2 compared to M_1. This test has has a χ^2 distribution with df = $\nu_2 - \nu_1$.

$$\begin{aligned} \Delta G^2 \equiv G^2(M_2 \mid M_1) &= G^2(M_2) - G^2(M_1) \\ &= 2 \sum n_i \log(\widehat{m}_{i1}/\widehat{m}_{i2}) \end{aligned} \qquad (7)$$

Example: Berkeley admissions data

For the UC Berkeley data, with table variables [A]dmit, [D]ept and [G]ender the following models form a nested chain

$$[A][D][G] \subset [A][DG] \subset [AD][AG][DG] \subset [ADG]$$

Table: Hierarchical G^2 tests for loglinear models fit to the UC Berkeley data

Type	LLM terms	G^2	df	$\Delta(G^2)$	$\Delta(df)$	$\Pr(> \Delta(G^2))$
Mutual ind	[A][D][G]	2097.67	16			
Joint	[A][DG]	877.06	11	1220.62	5	0.0000
All 2-way	[AD][AG][DG]	20.20	5	1128.70	5	0.0000
Saturated	[ADG]	0.0	0	20.20	5	0.0011

- Only testing decrease in G^2 from one model to the next
- Here, each model is significantly better than the previous
- Joint vs. all 2-way: does Admit depend on Dept and/or Gender?
- Absolute fit of the all 2-way model is not terrible. Investigate further!

Fitting loglinear models: SAS

SAS

- PROC CATMOD

```
1 %include catdata(berkeley);
2 proc catmod order=data data=berkeley;
3    format dept dept. admit admit.;
4    weight freq;                    /* data in freq. form */
5    model dept*gender*admit=_response_ ;
6    loglin admit|dept|gender @2   / title='Model (AD,AG,DG)'; run;
7    loglin admit|dept dept|gender / title='Model (AD,DG)';  run;
```

- PROC GENMOD

```
1 proc genmod data=berkeley;
2    class dept gender admit;
3    model freq = dept|gender dept|admit / dist=poisson;
4 run;
```

- mosaic macro usually fits loglin models internally and displays results
- You can also use PROC GENMOD for a more general model, and display the result with the mosaic macro.

Fitting loglinear models in R

loglm() - data in contingency table form (MASS package)

```
1 data(UCBAdmissions)
2    ## conditional independence (AD, DG) in Berkeley data
3 mod.1 <- loglm(~ (Admit + Gender) * Dept, data=UCBAdmissions)
4    ## all two-way model (AD, DG, AG)
5 mod.2 <- loglm(~ (Admit + Gender + Dept)^2, data=UCBAdmissions)
```

glm() - data in frequency form

```
1 berkeley <- as.data.frame(UCBAdmissions)
2 mod.3 <- glm(Freq ~ (Admit + Gender) * Dept, data=berkeley,
3                    family='poisson')
```

- loglm() simpler for nominal variables
- glm() allows a wider class of models and quantitative predictors (covariates)
- gnm() fits models for structured association and generalized *non-linear* models
- vcdExtra package provides visualizations for all.

Example: Berkeley admission data

Fit the model of mutual independence using loglm()

```
data("UCBAdmissions")
library(MASS)
berk.loglm0 <- loglm(~ Dept + Gender + Admit, data=UCBAdmissions)
berk.loglm0

## Call:
## loglm(formula = ~Dept + Gender + Admit, data = UCBAdmissions)
##
## Statistics:
##                    X^2 df P(> X^2)
## Likelihood Ratio 2097.7 16        0
## Pearson          2000.3 16        0
```

Example: Berkeley admission data

Fit other models with loglm()

```
# conditional independence [AD] [AG]
berk.loglm1 <- loglm(~ Admit * (Dept + Gender), data=UCBAdmissions)

# joint independence [A] [DG]
berk.loglm2 <- loglm(~ Admit + (Dept * Gender), data=UCBAdmissions)
berk.loglm2

## Call:
## loglm(formula = ~Admit + (Dept * Gender), data = UCBAdmissions)
##
## Statistics:
##                    X^2 df P(> X^2)
## Likelihood Ratio 877.06 11        0
## Pearson          797.70 11        0

# all two-way model [AD] [AG] [DG]
berk.loglm3 <-loglm(~(Admit+Dept+Gender)^2, data=UCBAdmissions)
```

Example: Berkeley admission data

Compare nested models with `anova()`

```
anova(berk.loglm0, berk.loglm2, berk.loglm3, test="Chisq")

## LR tests for hierarchical log-linear models
##
## Model 1:
##  ~Dept + Gender + Admit
## Model 2:
##  ~Admit + (Dept * Gender)
## Model 3:
##  ~(Admit + Dept + Gender)^2
##
##            Deviance df Delta(Dev) Delta(df) P(> Delta(Dev))
## Model 1    2097.671 16
## Model 2     877.056 11   1220.615         5         0.00000
## Model 3      20.204  5    856.852         6         0.00000
## Saturated     0.000  0     20.204         5         0.00114
```

Example: Berkeley admission data

`LRStats()` in vcdExtra gives one line summaries of a collection of models

```
LRstats(berk.loglm0, berk.loglm1, berk.loglm2, berk.loglm3)

## Likelihood summary table:
##             AIC  BIC LR Chisq Df Pr(>Chisq)
## berk.loglm0 2273 2282     2098 16     <2e-16 ***
## berk.loglm1 1336 1352     1149 10     <2e-16 ***
## berk.loglm2 1062 1077      877 11     <2e-16 ***
## berk.loglm3  217  240       20  5     0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- AIC and BIC are GOF measures adjusted for model parsimony
- Not not significance tests, but smaller is better
- Also apply to non-nested models

$$AIC = G^2 + 2 \times \# \text{ parameters}$$
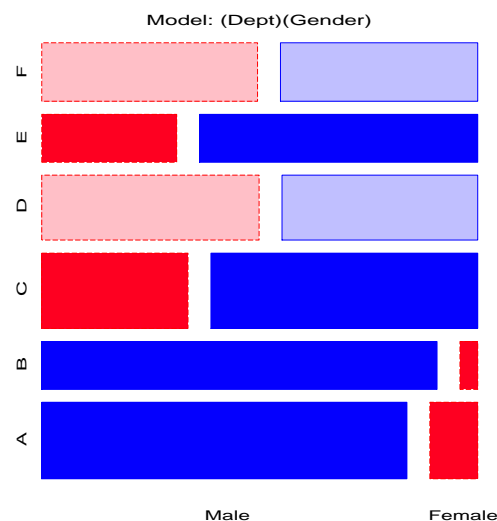$$BIC = G^2 + 2\log(n) \times \# \text{ parameters}$$

Mosaic displays: Predictor variables

Berkeley data: Departments $\times$ Gender (ignoring Admit):

- Did departments differ in the total number of applicants?
- Did men and women apply differentially to departments?

Model: (Dept)(Gender)

- Model [Dept] [Gender]: $G^2_{(5)} = 1220.6$.
- *Note*: Departments ordered A–F by overall rate of admission.
- Men more likely to apply to departments A,B; women more likely in depts C–F

Mosaic displays for multiway tables

- Generalizes to n-way tables: divide cells recursively
- Can fit *any* log-linear model (e.g., 2-way, 3-way, ...),
  - For a 3-way table: [A][B][C], [AB][C], [AB][AC], ..., [ABC]
- Each mosaics shows:
  - **DATA** (size of tiles)
  - (some) **marginal** frequencies (spacing $\rightarrow$ visual grouping)
  - **RESIDUALS** (shading) — what associations have been omitted?
- Visual fitting:
  - Pattern of lack-of-fit (residuals) $\rightarrow$ "better" model— smaller residuals
  - "cleaning the mosaic" $\rightarrow$ "better" model— empty cells
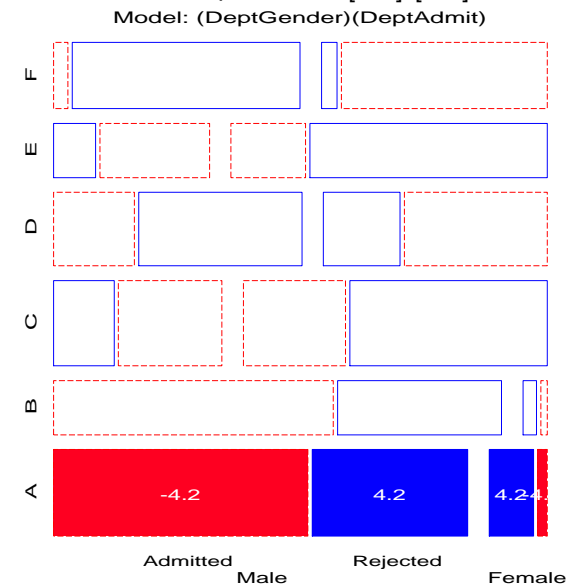 - best done interactively!

Joint independence, [DG][A] (null model, Admit as response) $[G^2_{(11)} = 877.1]$:

Model: (DeptGender)(Admit)
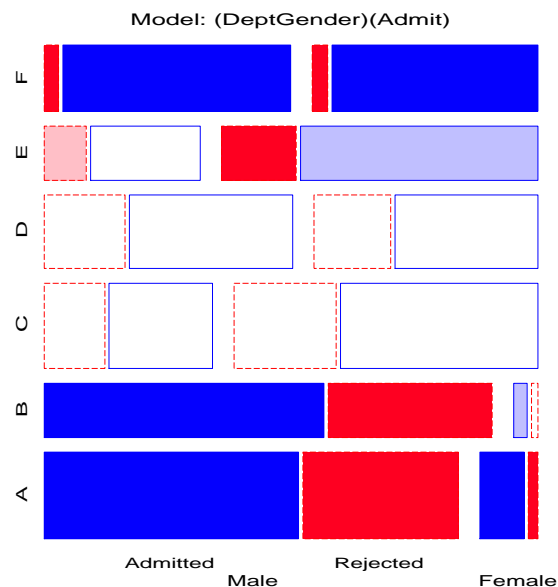
Mosaic displays for multiway tables

Conditional independence, [AD] [DG]:

Model: (DeptGender)(DeptAdmit)

- E.g., Add [Admit Dept] association $\rightarrow$ Conditional independence:
  - Fits poorly: ($G^2_{(6)} = 21.74$)
  - But, only in Department A!
- GLM approach allows fitting a special term for Dept. A
- Note: These displays use *standardized residuals*: better statistical properties.

Other variations: Double decker plots

- Visualize dependence of one categorical (typically binary) variable on predictors
- Formally: mosaic plots with vertical splits for all predictor dimensions, highlighting the response by shading
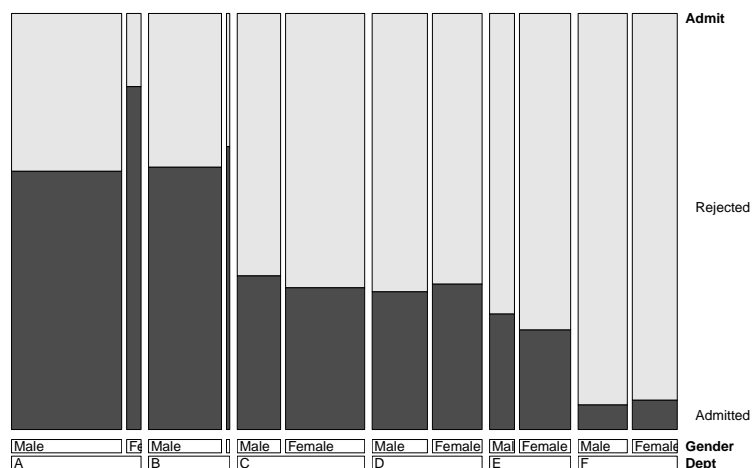
Sequential plots and models

- Mosaic for an *n*-way table $\rightarrow$ hierarchical decomposition of association
- Joint cell probabilities are decomposed as

$$p_{ijk\ell\cdots} = \underbrace{\overbrace{p_i \times p_{j|i}}^{\{v_1 v_2\}} \times p_{k|ij}}_{\{v_1 v_2 v_3\}} \times p_{\ell|ijk} \times \cdots \times p_{n|ijk\cdots}$$

  - First 2 terms $\rightarrow$ mosaic for v_1 and v_2
  - First 3 terms $\rightarrow$ mosaic for v_1, v_2 and v_3
  - $\cdots$
- Roughly analogous to sequential fitting in regression: X_1, $X_2|X_1$, $X_3|X_1 X_2$, $\cdots$
- The order of variables matters for interpretation

Sequential plots and models

Sequential models of *joint independence* $\rightarrow$ additive decomposition of the total association, $G^2_{[v_1][v_2]\ldots[v_p]}$ (mutual independence),
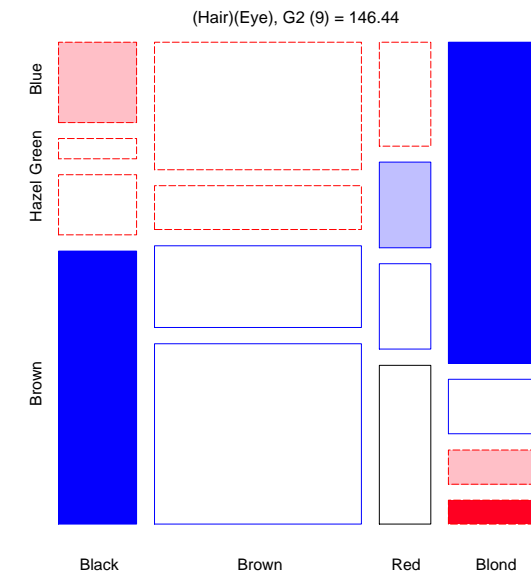
$$G^2_{[v_1][v_2]\ldots[v_p]} = G^2_{[v_1][v_2]} + G^2_{[v_1 v_2][v_3]} + G^2_{[v_1 v_2 v_3][v_4]} + \cdots + G^2_{[v_1\ldots v_{p-1}][v_p]}$$

e.g., for Hair Eye color data

Model	Model symbol	df	G^2
Marginal	[Hair] [Eye]	9	146.44
Joint	[Hair, Eye] [Sex]	15	19.86
Mutual	[Hair] [Eye] [Sex]	24	166.30

Sequential plots and models: Example
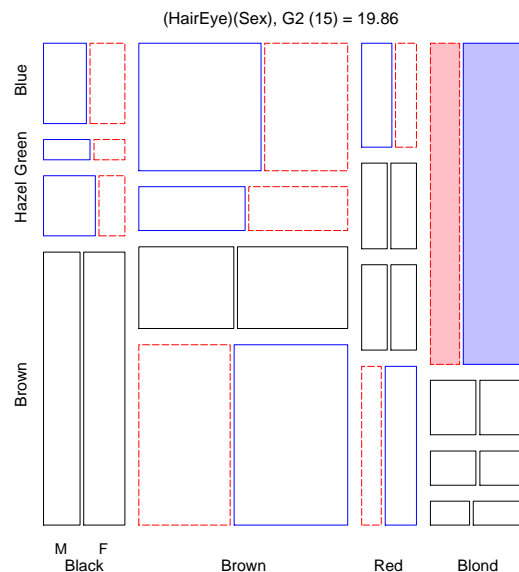
- Hair color x Eye color marginal table (ignoring Sex)

(Hair)(Eye), G2 (9) = 146.44

Sequential plots and models: Example

- 3-way table, Joint Independence Model [Hair Eye] [Sex]

(HairEye)(Sex), G2 (15) = 19.86

Sequential plots and models: Example

- 3-way table, Mutual Independence Model [Hair] [Eye] [Sex]

(Hair)(Eye)(Sex), G2 (24) = 166.30

Sequential plots and models: Example

Marginal
(Hair)(Eye), G2 (9) = 146.44

Joint
(HairEye)(Sex), G2 (15) = 19.86

Total
(Hair)(Eye)(Sex), G2 (24) = 166.30

[Hair] [Eye]
$G^2_{(9)} = 146.44$

[Hair Eye] [Sex]
$G^2_{(15)} = 19.86$

[Hair] [Eye] [Sex]
$G^2_{(24)} = 166.30$
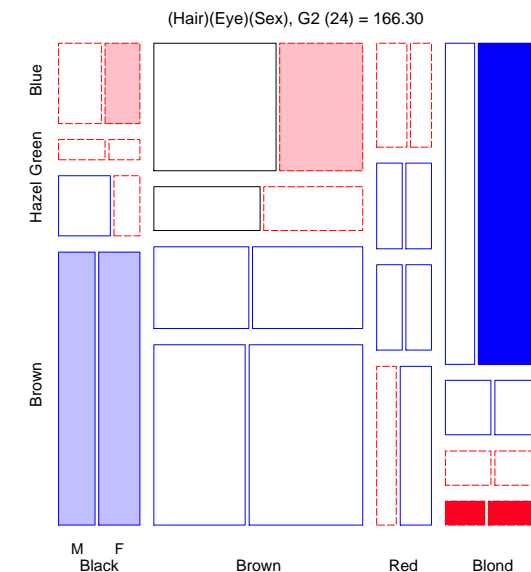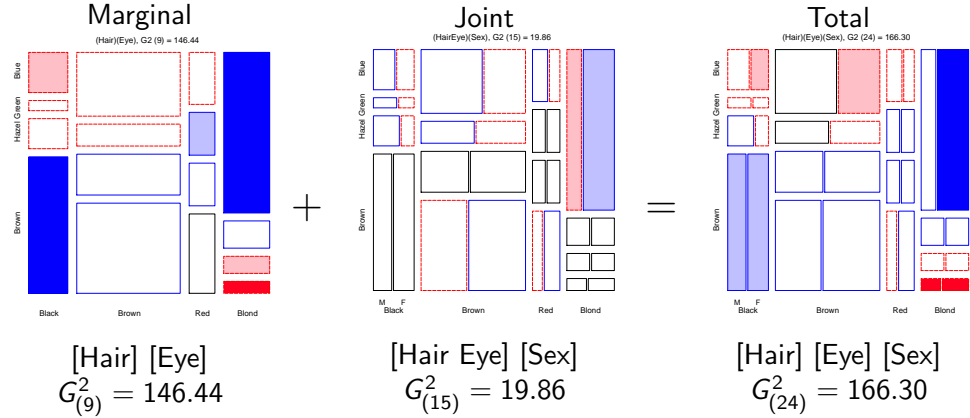
Mosaic matrices

- Analog of *scatterplot matrix* for categorical data (Friendly, 1999)
  - Shows all $p(p-1)$ pairwise views in a coherent display
  - Each pairwise mosaic shows bivariate (marginal) relation
  - Fit: marginal independence
  - Residuals: show marginal associations
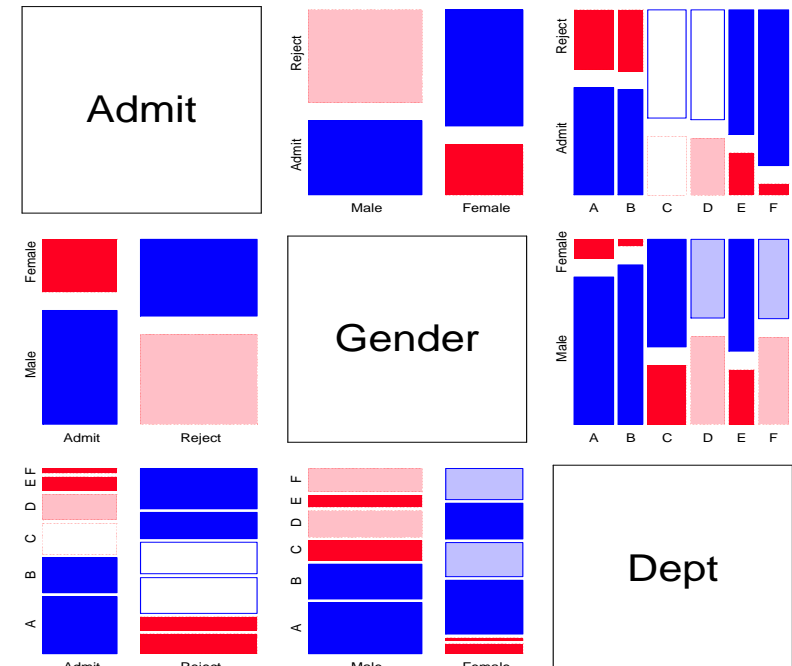 - Direct visualization of the "Burt" matrix analyzed in MCA for p categorical variables
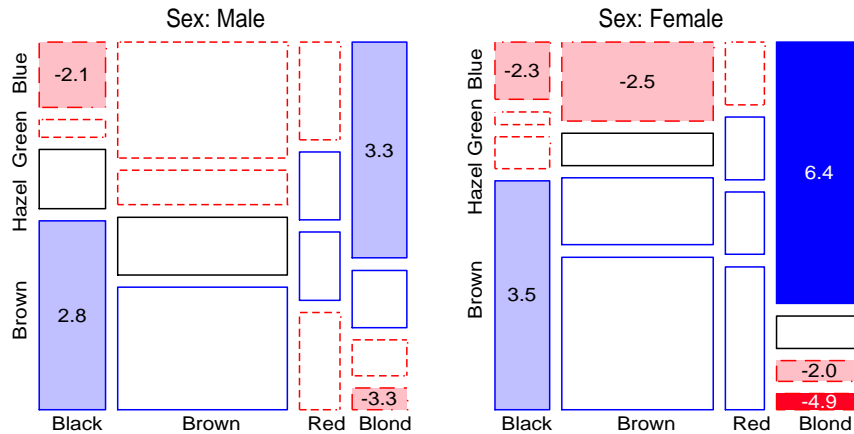
Hair, Eye, Sex data:
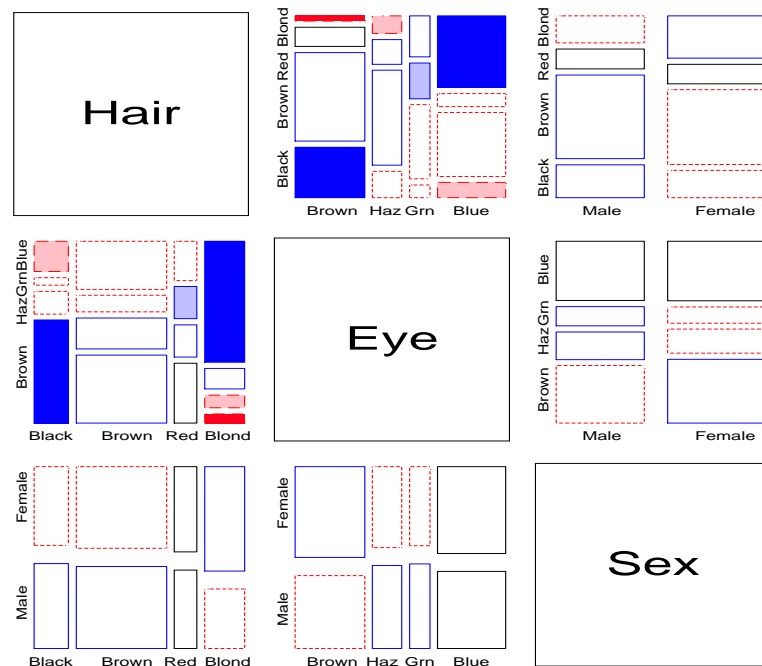
Berkeley data:

Partial association, Partial mosaics

- **Stratified analysis:**
  - How does the association between two (or more) variables vary over levels of other variables?
  - Mosaic plots for the main variables show *partial association* at each level of the other variables.
  - E.g., Hair color, Eye color *BY* Sex ↔ `TABLES sex * hair * eye;`

Sex: Male

Sex: Female

Partial association, Partial mosaics

Stratified analysis: conditional decomposition of G^2

- Fit models of partial (conditional) independence, $A \perp B \mid C_k$ at each level of (controlling for) C.
- $\Rightarrow$ partial G^2s add to the overall G^2 for conditional independence, $A \perp B \mid C$

$$G^2_{A \perp B \mid C} = \sum_k G^2_{A \perp B \mid C(k)}$$

Table: Partial and Overall conditional tests, $Hair \perp Eye \mid Sex$

Model	df	G^2	p-value
$[Hair][Eye] \mid$ Male	9	44.445	0.000
$[Hair][Eye] \mid$ Female	9	112.233	0.000
$[Hair][Eye] \mid$ Sex	18	156.668	0.000

Software for Mosaic Displays: Web applet
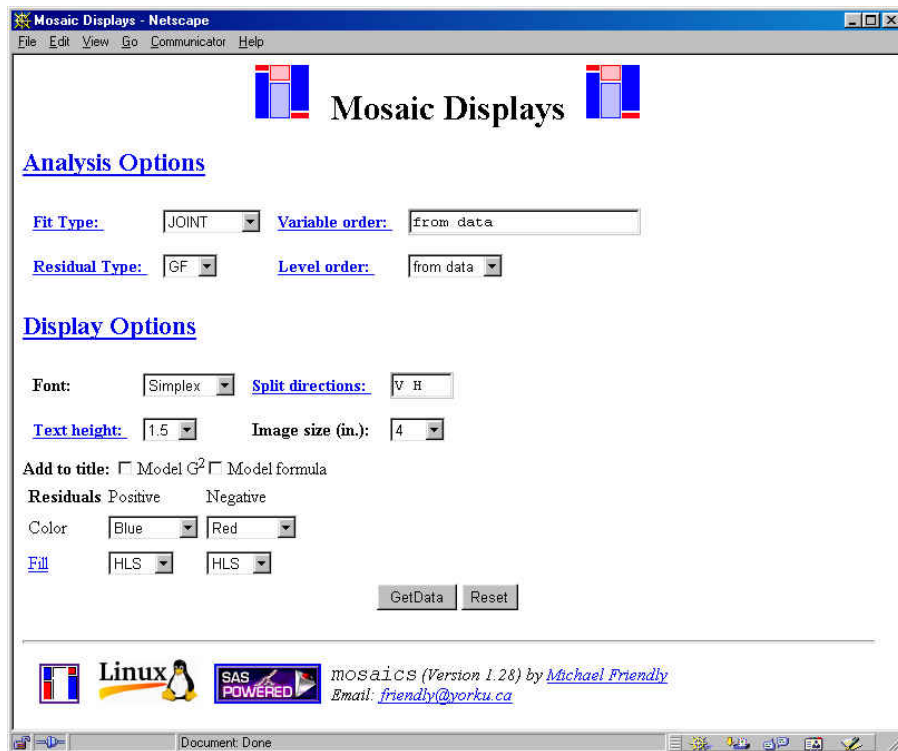
Demonstration web applet

Go to: `http://datavis.ca/online/mosaics/`

- Runs the *current* version of `mosaics.sas` via a cgi script (perl)
- Can:
  - run *sample* data,
  - *upload* a data file,
  - *enter* data in a form.
- Choose model *fitting* and *display* options (not all supported).
- Provides (limited) interaction with the mosaics via javascript

Software for Mosaic Displays: SAS

Mosaic Displays - Netscape
File Edit View Go Communicator Help

Mosaic Displays

Analysis Options

Fit Type: JOINT   Variable order: from data
Residual Type: GF   Level order: from data

Display Options

Font: Simplex   Split directions: V H
Text height: 1.5   Image size (in.):   4
Add to title: ☐ Model G² ☐ Model formula
Residuals Positive   Negative
Color   Blue   Red
Fill   HLS   HLS

GetData   Reset

Linux   SAS POWERED   *mosaics* (Version 1.28) by *Michael Friendly*
Email: *friendly@yorku.ca*

Document: Done

Software for Mosaic Displays: SAS

SAS software & documentation

http://datavis.ca/mosaics/mosaics.pdf - User Guide
http://datavis.ca/books/vcd/macros.html - Software

- **Examples**: Many in *VCD* and on web site
- **SAS/IML modules**: `mosaics.sas`— Most flexible
    - Enter frequency table directly in SAS/IML, or read from a SAS dataset.
    - Select, collapse, reorder, re-label table levels using SAS/IML statements
    - Specify structural 0s, fit specialized models (e.g., quasi-independence)
    - Interface to models fit using PROC `GENMOD`

Software for Mosaic Displays: SAS

- **Macro interface**: `mosaic` macro, `table` macro, `mosmat` macro
- `mosaic` **macro**— Easiest to use
    - Direct input from a SAS dataset
    - No knowledge of SAS/IML required
    - Reorder table variables; collapse, reorder table levels with `table` macro
    - Convenient interface to *partial mosaics* (BY=)
- `table` **macro**
    - Create frequency table from raw data
    - Collapse, reorder table categories
    - Re-code table categories using SAS formats, e.g., 1='Male' 2='Female'
- `mosmat` **macro**
    - Mosaic matrices— analog of scatterplot matrix (Friendly, 1999)

---

mosaic macro example: Berkeley data

berkeley.sas

```
1  title 'Berkeley Admissions data';
2  proc format;
3      value admit 1="Admitted" 0="Rejected"        ;
4      value dept  1="A" 2="B" 3="C" 4="D" 5="E" 6="F";
5          value $sex  'M'='Male'   'F'='Female';
6  data berkeley;
7      do dept = 1 to 6;
8          do gender = 'M', 'F';
9              do admit = 1, 0;
10                 input freq @@;
11                 output;
12     end; end; end;
13 /* -- Male --  - Female- */
14 /* Admit  Rej   Admit Rej */
15 datalines;
16     512   313      89    19   /* Dept A */
17     353   207      17     8   /*      B */
18     120   205     202   391   /*      C */
19     138   279     131   244   /*      D */
20      53   138      94   299   /*      E */
21      22   351      24   317   /*      F */
22 ;
```

Data set berkeley:

dept	gender	admit	freq
1	M	1	512
1	M	0	313
1	F	1	89
1	F	0	19
2	M	1	353
2	M	0	207
2	F	1	17
2	F	0	8
3	M	1	120
3	M	0	205
3	F	1	202
3	F	0	391
4	M	1	138
4	M	0	279
4	F	1	131
4	F	0	244
5	M	1	53
5	M	0	138
5	F	1	94
5	F	0	299
6	M	1	22
6	M	0	351
6	F	1	24
6	F	0	317

mosaic macro example: Berkeley data

mosaic9m.sas

```
1  goptions hsize=7in vsize=7in;
2  %include catdata(berkeley);
3
4  *-- apply character formats to numeric table variables;
5  %table(data=berkeley,
6      var=Admit Gender Dept,
7      weight=freq,
8      char=Y, format=admit admit. gender $sex. dept dept.,
9      order=data, out=berkeley);
10
11 %mosaic(data=berkeley,
12     vorder=Dept Gender Admit,  /* reorder variables */
13     plots=2:3,                 /* which plots?      */
14     fittype=joint,             /* fit joint indep.  */
15     split=H V V, htext=3);     /* options           */
```
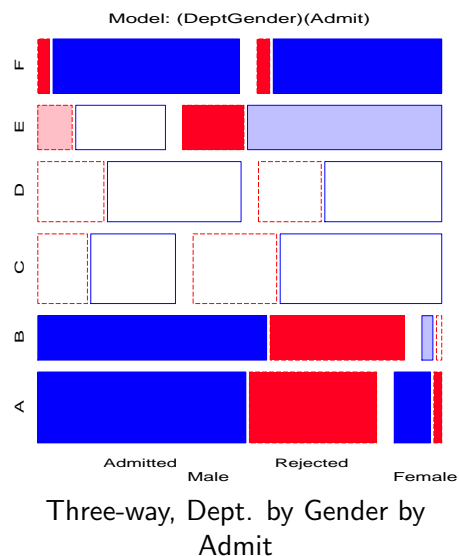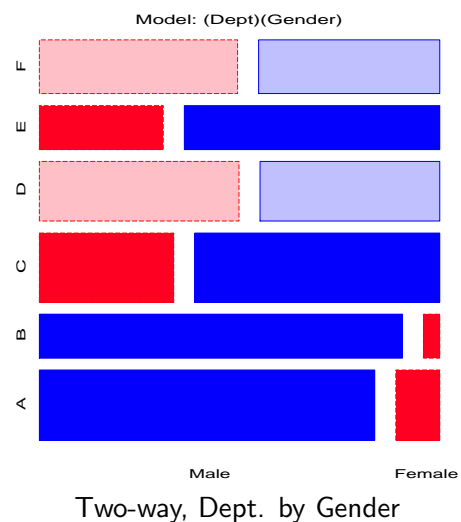
NB: The `fittype=` argument allows various types of sequential models: joint, conditional, etc.

mosaic macro example: Berkeley data

Two-way, Dept. by Gender
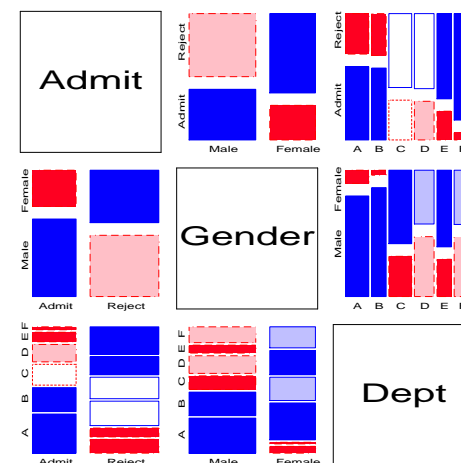
Three-way, Dept. by Gender by Admit

mosmat macro: Mosaic matrices

mosmat9m.sas

```
1  %include catdata(berkeley);
2  %mosmat(data=berkeley,
3      vorder=Admit Gender Dept, sort=no);
```
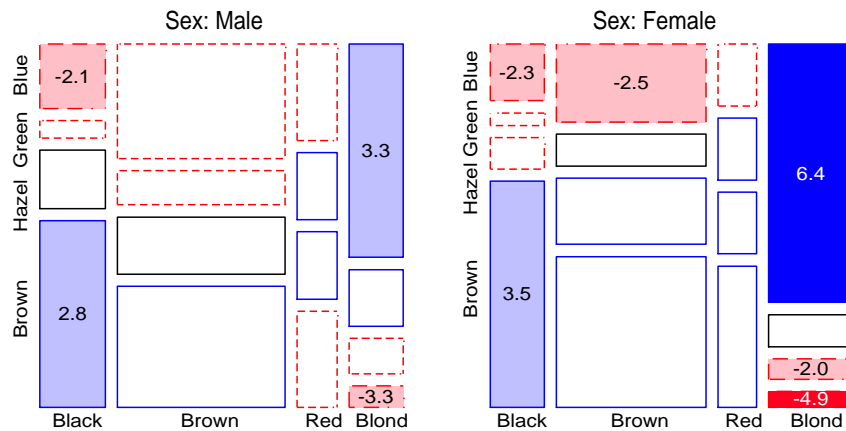
Partial mosaics

```
                          ┌──── mospart3.sas ────┐
1  %include catdata(hairdat3s);
2  %gdispla(OFF);
3  %mosaic(data=haireye,
4      vorder=Hair Eye Sex, by=Sex,
5      htext=2, cellfill=dev);
6  %gdispla(ON);
7  %panels(rows=1, cols=2);    /* make 2 figs -> 1 */
```


Using the vcd package in R

```
>library(vcd)          # load the vcd package & friends
>
>data(HairEyeColor)
>structable(Eye ~ Hair + Sex, data=HairEyeColor)
```

	Eye	Brown	Blue	Hazel	Green
Hair	Sex				
Black	Male	32	11	10	3
	Female	36	9	5	2
Brown	Male	53	50	25	15
	Female	66	34	29	14
Red	Male	10	10	7	7
	Female	16	7	7	7
Blond	Male	3	30	5	8
	Female	4	64	5	8

- The structable() function →'flat' representation of an *n*-way table, similar to mosaic displays
- Formula interface: Col factors ∼ row factors

Using the vcd package in R

- The loglm() function fits a loglinear model, returns a loglm object
  - Fit the 3-way mutual independence model: Hair + Eye + Sex ≡ [Hair] [Eye] [Sex]
  - Printing the object gives a brief model summary (badness of fit)

```
>## Independence model of hair and eye color and sex.
>mod.1 <- loglm(~Hair+Eye+Sex, data=HairEyeColor)
>mod.1
```
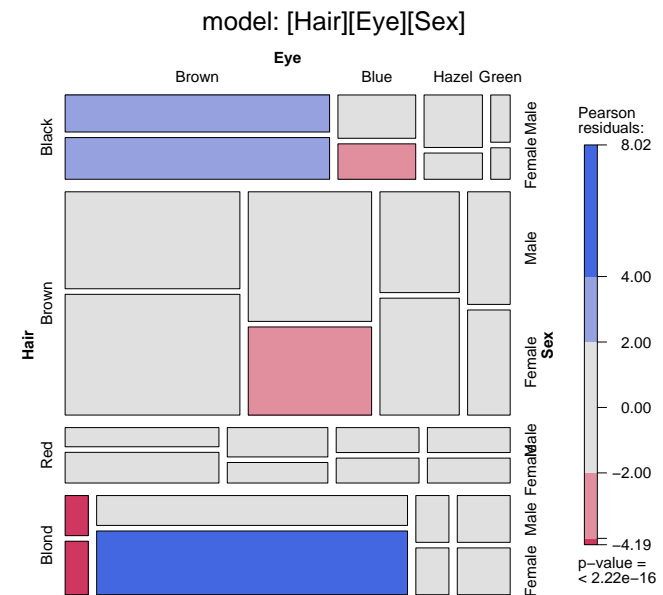
```
Call:
loglm(formula = ~Hair + Eye + Sex, data = HairEyeColor)

Statistics:
                    X^2 df P(> X^2)
Likelihood Ratio 166.3001 24        0
Pearson          164.9247 24        0
```

- The mosaic() function plots the object.
- the vcdExtra package extends mosaic() to glm() models.

```
>mosaic(mod.1, main="model: [Hair][Eye][Sex]")
```

vcd: Other models

```
>## Joint independence model.
>mod.2 <- loglm(~Hair*Eye+Sex, data=HairEyeColor)
>mod.2
```

```
Call:
loglm(formula = ~Hair * Eye + Sex, data = HairEyeColor)

Statistics:
                  X^2 df  P(> X^2)
Likelihood Ratio 19.85656 15 0.1775045
Pearson          19.56712 15 0.1891745
```

```
>## Conditional independence model:  Hair*Eye + Sex*Eye
>mod.3 <- loglm(~(Hair+Sex)*Eye, data=HairEyeColor)
>mod.3
```
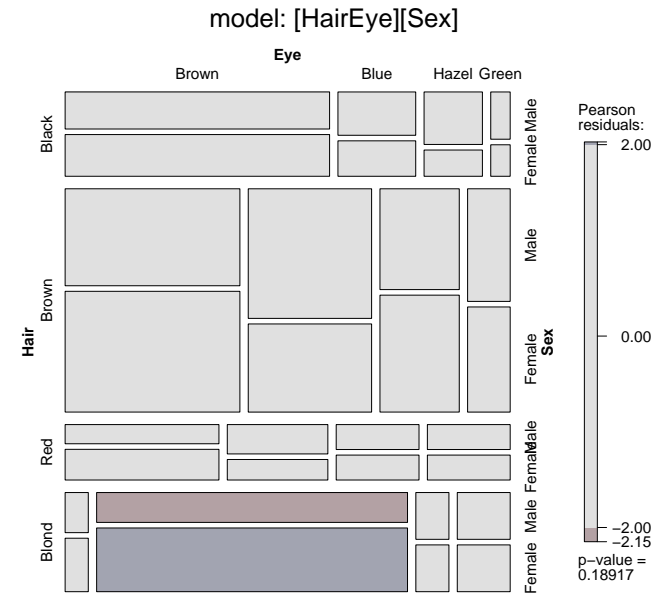
```
Call:
loglm(formula = ~(Hair + Sex) * Eye, data = HairEyeColor)

Statistics:
                  X^2 df  P(> X^2)
Likelihood Ratio 18.32715 12 0.1061122
Pearson          18.04110 12 0.1144483
```
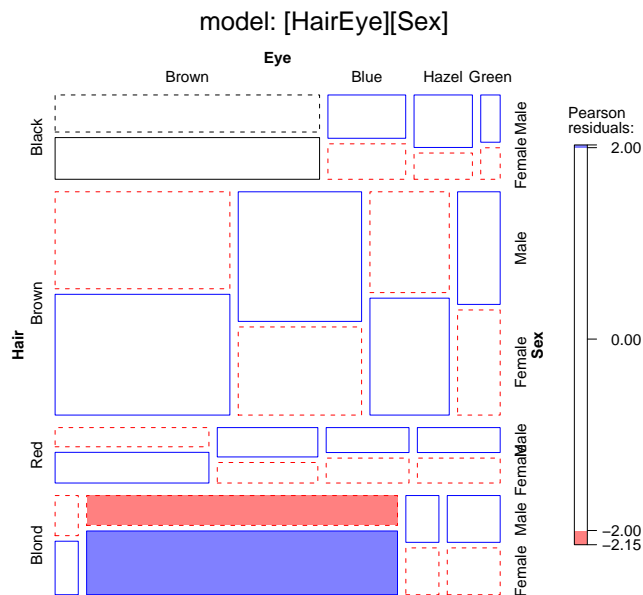
```
>mosaic(mod.2,  main="model: [HairEye][Sex]")
```


model: [HairEye][Sex]

```
>mosaic(mod.2,  main="model: [HairEye][Sex]", gp=shading_Friendly)
```


model: [HairEye][Sex]

Testing differences between models

- For nested models, $M_1 \subset M_2$ (M_1 nested within, a special case of M_2), the difference in LR G^2, $\Delta = G^2(M_1) - G^2(M_2)$ is a specific test of the difference between them. Here, $\Delta \sim \chi^2$ with $df = df_1 - df_2$.
- R functions are object-oriented: they do different things for different types of objects.

```
>anova(mod.1, mod.2)
```

```
LR tests for hierarchical log-linear models

Model 1:
 ~Hair + Eye + Sex
Model 2:
 ~Hair * Eye + Sex
```

	Deviance	df	Delta(Dev)	Delta(df)	P(> Delta(Dev))
Model 1	166.30014	24			
Model 2	19.85656	15	146.44358	9	0.0000
Saturated	0.00000	0	19.85656	15	0.1775

More structured tables

Ordered categories

Tables with ordered categories may allow more parsimonious tests of association

- Can represent λ_{ij}^{AB} by a small number of parameters
- $\rightarrow$ more focused and *more powerful* tests of lack of independence (recall: CMH tests)
- Allow one to "explain" the pattern of association in a compact way.

Square tables

For square $I \times I$ tables, where row and column variables have the same categories:

- Can ignore diagonal cells, where association is expected and test remaining association (*quasi-independence*)
- Can test whether association is *symmetric* around the diagonal cells.
- Can test substantively important hypotheses (e.g., mobility tables)

All of these require the GLM approach for model fitting

Ordered categories I

- **Ordinal scores**
  - In many cases it may be reasonable to assign numeric scores, $\{a_i\}$ to an ordinal row variable and/or numeric scores, $\{b_i\}$ to an ordinal column variable.
  - Typically, scores are equally spaced and sum to zero, $\{a_i\} = i - (I+1)/2$, e.g., $\{a_i\} = \{-1, 0, 1\}$ for I=3.
- **Linear-by-Linear (Uniform) Association**: When *both* variables are ordinal, the simplest model posits that any association is *linear* in both variables.

$$\lambda_{ij}^{AB} = \gamma\, a_i b_j$$

  - Only adds one additional parameter to the independence model ($\gamma = 0$).
  - It is similar to CMH test for linear association
  - For integer scores, the local log odds ratios for *any* contiguous 2×2 table are all equal, $\log \theta_{ij} = \gamma$
  - This is a model of *uniform association* — simple interpretation!

Ordered categories II

For a two way table, there are 4 possibilities, depending on which variables are ordinal, and assigned scores:

B→ A↓	Nominal	Col scores b_j, j=1,…J
Nominal	General association df: (I-1)(J-1) parm: λ_{ij}^{AB}	Row effects df: I-1 parm: $\alpha_i\, b_j$
Row scores a_i, i=1, … I	Col effects df: J-1 parm: $a_i\, \beta_j$	Uniform association df: 1 parm: $\gamma\, a_i\, b_j$

Ordered categories III

- **Row Effects and Column Effects**: When only one variable is assigned scores, we have the *row effects model* or the *column effects model*.
  - E.g., in the row effects model, the row variable (A) is treated as nominal, while the column variable (B) is assigned ordered scores $\{b_j\}$.

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \alpha_i b_j$$

  where the row parameters, α_i, are defined so they sum to zero.
  - This model has $(I-1)$ more parameters than the independence model.
  - A Row Effects + Column Effects model allows both variables to be ordered, but not necessarily with linear scores.
- **Fitting models for ordinal variables**
  - Create *numeric* variables for category scores
  - PROC `GENMOD`: Use as quantitative variables in `MODEL` statement, but *not* listed as `CLASS` variables
  - R: Create numeric variables with `as.numeric(factor)`

Ordered categories: RC models

- **RC(1) model**: Generalizes the uniform association, R, C and R+C models by relaxing the assumption of specified order and spacing.

$$RC(1) : \log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \phi\mu_i\nu_j$$

  - The row parameters (μ_i) and column parameters (ν_j) are estimated from the data.
  - ϕ is the measure of association, similar to γ in the uniform association model
- **RC(2) ... RC(M) models**: Allow two (or more) log-multiplicative association terms; e.g.:
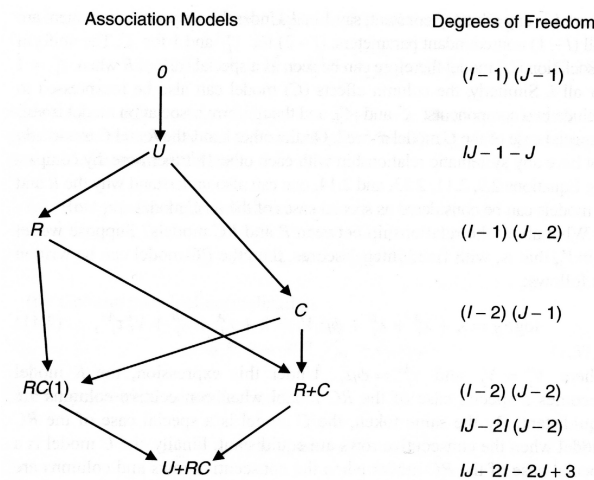
$$RC(2) : \log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \phi_1\mu_{i1}\nu_{j1} + \phi_2\mu_{i2}\nu_{j2}$$

Related to CA, but provide hypothesis tests, std. errors, etc.
- **Fitting RC models**
  - SAS: no implementation
  - R: Fit with `gnm(Freq ~ R + C + Mult(R, C))`

Relations among models

- Structured models: different ways to account for association
- Ordered by: df (# of parameters)
- Arrows show nested models (compare directly: $\Delta\chi^2$)
- All can be compared using AIC (or BIC)

Example: Mental impairment and parents' SES

- Srole et al. (1978) Data on mental health status of ~1600 young NYC residents in relation to parents' SES.
  - Mental health: Well, mild symptoms, moderate symptoms, Impaired
  - SES: 1 (High) − 6 (Low)

Mental health	Parents' SES					
	High	2	3	4	5	Low
1: Well	64	57	57	72	36	21
2: Mild	94	94	105	141	97	71
3: Moderate	58	54	65	77	54	54
4: Impaired	46	40	60	94	78	71

Before fitting models, it is often useful to explore the relation amongs the row/column categories. Correspondence analysis is a good idea!


Mental impairment and SES
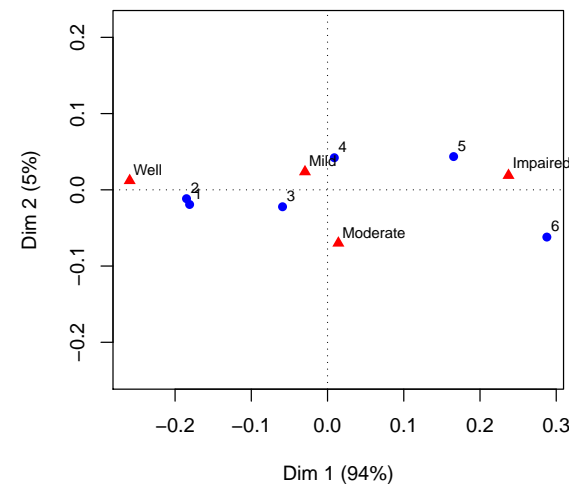
- Essentially 1D
- Both variables are ordered
- High SES goes with better mental health status
- Can we treat either or both as equally-spaced?
- GLM approach allows testing/comparing hypotheses vs. eye-balling
- Parameter estimates quantify effects.

Visual assessment of various loglin/GLM models: mosaic displays

Mental Impairment and SES: Independence
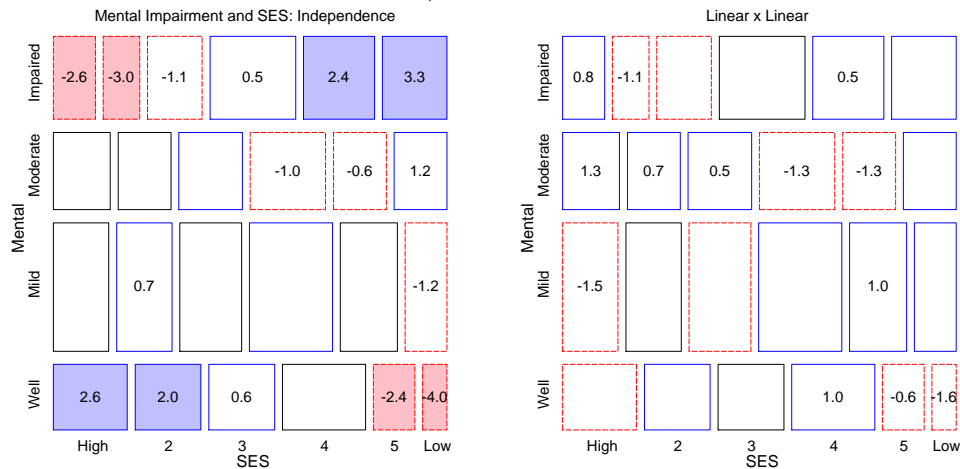
Linear x Linear

- Residuals from the independence model show an opposite-corner pattern. This is consistent with both:
  - Linear × linear model: equi-spaced scores for both Mental and SES
  - Row effects model: equi-spaced scores for SES, ordered scores for Mental

Statistical assesment:

Table: Mental health data: Goodness-of-fit statistics for ordinal loglinear models

Model	G^2	df	$\Pr(> G^2)$	AIC	AIC-best
Independence	47.418	15	0.00003	65.418	35.523
Col effects (SES)	6.829	10	0.74145	34.829	4.934
Row effects (mental)	6.281	12	0.90127	30.281	0.386
Lin × Lin	9.895	14	0.76981	29.895	0.000

- Both the Row Effects and Linear × linear models are significantly better than the Independence model
- AIC indicates a slight preference for the Linear × linear model
- In the Linear × linear model, the estimate of the coefficient of $a_i b_j$ is $\hat{\gamma} = 0.0907 = \widehat{\log \theta}$, so $\hat{\theta} = \exp(0.0907) = 1.095$.
- $\mapsto$ each step down the SES scale increases the odds of being classified one step *poorer* in mental health by 9.5%.
- Compare with purely exploratory (CA) interpretation: mental health increases with SES

Fitting these models with PROC ~~GENMOD~~:

mentgen2.sas

```
1  %include catdata(mental);
2  data mental;
3     set mental;
4     m_lin = mental;      *-- copy m_lin and s_lin for;
5     s_lin = ses;         *-- use non-CLASS variables;
6
7  title 'Independence model';
8  proc genmod data=mental;
9     class mental ses;
10    model count = mental ses / dist=poisson obstats residuals;
11    format mental mental. ses ses.;
12    ods output obstats=obstats;
13 %mosaic(data=obstats, vorder=Mental SES, resid=stresdev,
14 title=Mental Impairment and SES: Independence, split=H V);
```

Row Effects model:

mentgen2.sas

```
16 proc genmod data=mental;
17    class mental ses;
18    model count = mental ses mental*s_lin / dist=poisson obstats;
19    ...
```

Linear × linear model:

mentgen2.sas

```
21 proc genmod data=mental;
22    class mental ses;
23    model count = mental ses m_lin*s_lin / dist=poisson obstats;
```

Fitting these models with glm() in R (see: mental-glm.R for plots)

```r
library(vcdExtra)
data(Mental)
# Integer scores for rows/cols
Cscore <- as.numeric(Mental$ses)
Rscore <- as.numeric(Mental$mental)

indep <- glm(Freq ~ mental+ses, family = poisson, data=Mental)

# column effects model (ses)
coleff <- glm(Freq ~ mental + ses + Rscore:ses,
              family = poisson, data = Mental)

# row effects model (mental)
roweff <- glm(Freq ~ mental + ses + mental:Cscore,
              family = poisson, data = Mental)

# linear x linear association
linlin <- glm(Freq ~ mental + ses + Rscore:Cscore,
              family = poisson, data = Mental)

# compare models
AIC(indep, coleff, roweff, linlin)
```
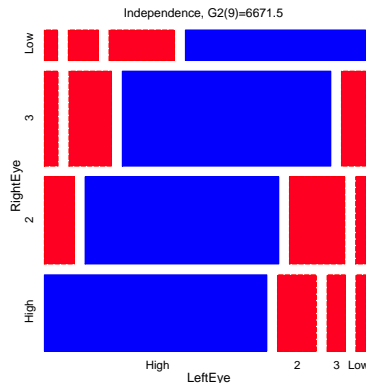
Square tables

- Tables where two (or more) variables have the same category levels:
    - Employment categories of related persons (mobility tables)
    - Multiple measurements over time (panel studies; longitudinal data)
    - Repeated measures on the same individuals under different conditions
    - Related/repeated measures are rarely independent, but may have simpler forms than general association
- E.g., vision data: Left and right eye acuity grade for 7477 women

Independence, G2(9)=6671.5

Square tables: Quasi-Independence

- Related/repeated measures are rarely independent— most observations often fall on diagonal cells.
- Quasi-independence ignores diagonals: tests independence in remaining cells ($\lambda_{ij} = 0$ for $i \neq j$).
- The model dedicates one parameter (δ_i) to each diagonal cell, fitting them exactly,

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \delta_i\, I(i = j)$$

where $I(\bullet)$ is the indicator function.
- This model may be fit as a GLM by including indicator variables for each diagonal cell: fitted exactly

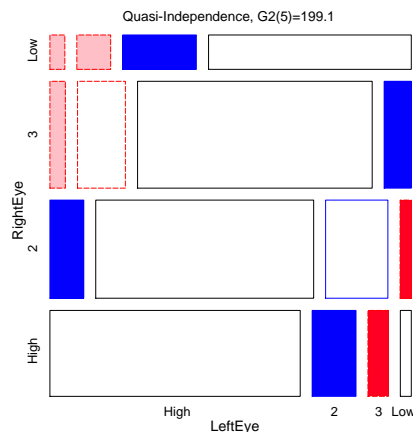| diag	4 rows		4 cols	
1	0	0	0
0	2	0	0
0	0	3	0
0	0	0	4

- Using PROC GENMOD

```
··· mosaic10g.sas
1 title 'Quasi-independence model (women)';
2 proc genmod data=women;
3     class RightEye LeftEye diag;
4     model Count = LeftEye RightEye diag /
5         dist=poisson link=log obstats residuals;
6     ods output obstats=obstats;
7 %mosaic(data=obstats, vorder=RightEye LeftEye, ...);
```

Mosaic:

Quasi-Independence, G2(5)=199.1

Square tables: Symmetry

- Tests whether the table is symmetric around the diagonal, i.e., $m_{ij} = m_{ji}$
- As a loglinear model, symmetry is

$$\log m_{ij} = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \ ,$$

subject to the conditions $\lambda_i^A = \lambda_j^B$ and $\lambda_{ij}^{AB} = \lambda_{ji}^{AB}$.
- This model may be fit as a GLM by including indicator variables with equal values for symmetric cells, and indicators for the diagonal cells (fit exactly)

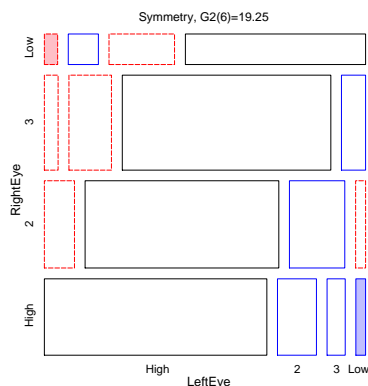| symmetry	4 rows		4 cols)	
1	12	13	14
12	2	23	24
13	23	3	34
14	24	34	4

- Using PROC GENMOD

··· mosaic10g.sas

```
1 proc genmod data=women;
2 class symmetry;
3 model Count = symmetry /
4 dist=poisson link=log obstats residuals;
5    ods output obstats=obstats;
6 %mosaic(data=obstats, vorder=RightEye LeftEye, ...);
```

Mosaic:

Symmetry, G2(6)=19.25

- **Quasi-Symmetry**
  - Symmetry is often too restrictive: $\mapsto$ equal marginal frequencies ($\lambda_i^A = \lambda_i^B$)
  - PROC GENMOD: Use the usual marginal effect parameters + symmetry:

··· mosaic10g.sas

```
1 proc genmod data=women;
2 class LeftEye RightEye symmetry;
3 model Count = LeftEye RightEye symmetry /
4 dist=poisson link=log obstats residuals;
5    ods output obstats=obstats;
```

Quasi-Symmetry, G2(3)=7.27
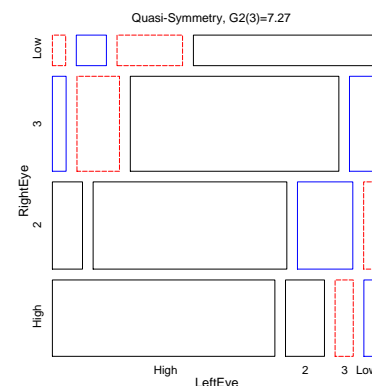
Comparing models

Table: Summary of models fit to vision data

Model	G^2	df	$Pr(> G^2)$	AIC	AIC - min(AIC)
Independence	6671.51	9	0.00000	6685.51	6656.23
Linear*Linear	1818.87	8	0.00000	1834.87	1805.59
Row+Column Effects	1710.30	4	0.00000	1734.30	1705.02
Quasi-Independence	199.11	5	0.00000	221.11	191.83
Symmetry	19.25	6	0.00376	39.25	9.97
Quasi-Symmetry	7.27	3	0.06375	33.27	3.99
Ordinal Quasi-Symmetry	7.28	5	0.20061	29.28	0.00

- Only the quasi-symmetry models provide an acceptable fit: When vision is unequal, association is symmetric!
- The ordinal quasi-symmetry model is most parsimonious
- AIC is your friend for model comparisons

Using the gnm package in R

- `Diag()` and `Symm()`: structured associations for square tables
- `Topo()`: more general structured associations
- `mosaic.glm()` in vcdExtra

```
library(vcdExtra)
library(gnm)
women <- subset(VisualAcuity, gender=="female", select=-gender)

indep <- glm(Freq ~ right + left, data = women, family=poisson)
mosaic(indep, residuals_type="rstandard", gp=shading_Friendly,
       main="Vision data: Independence (women)"  )

quasi.indep <- glm(Freq ~ right + left + Diag(right, left),
       data = women, family = poisson)

symmetry <- glm(Freq ~ Symm(right, left),
       data = women, family = poisson)

quasi.symm <- glm(Freq ~ right + left + Symm(right, left),
       data = women, family = poisson)

# model comparisons: for *nested* models
anova(indep, quasi.indep, quasi.symm, test="Chisq")
anova(symmetry, quasi.symm, test="Chisq")
```

Survival on the *Titanic*

Survival on the *Titanic*: 2201 passengers, classified by Class, Gender, Age, survived. Data from:

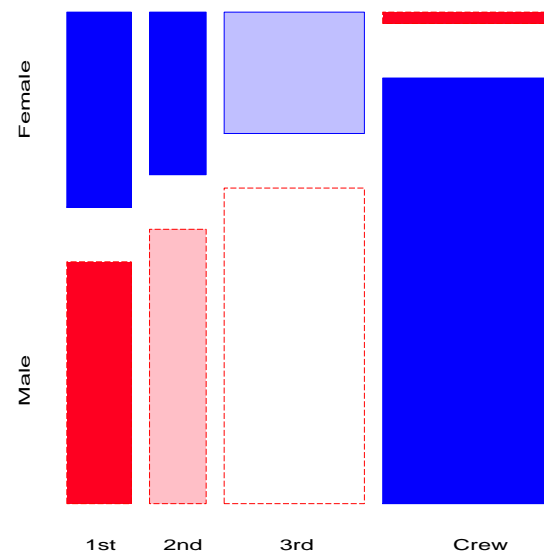- Mersey (1912), *Report on the loss of the "Titanic" S.S.*
- Dawson (1995)

			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

Order of variables in mosaics: Class, Gender, Age, Survival
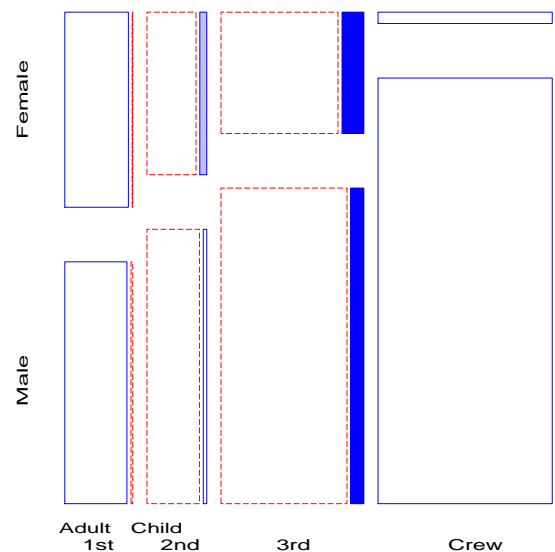
Survival on the *Titanic*: Background variables

Class × Gender:
- % males decreases with increasing economic class,
- crew almost entirely male

Sequential mosaics: understand associations among background variables
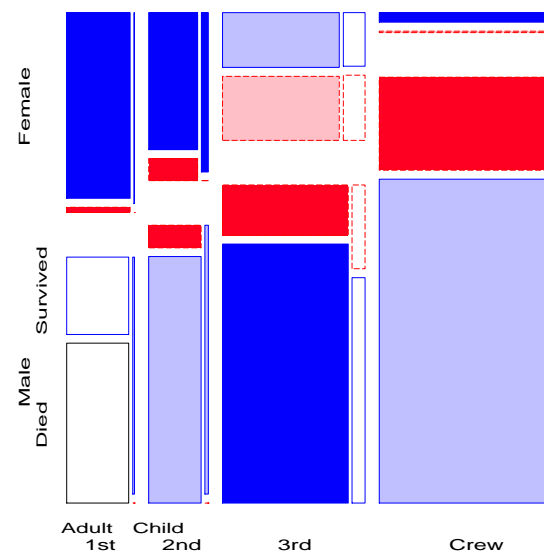
Survival on the *Titanic*: Background variables

3 way: {Class, Gender} ⊥ Age ?

- Overall proportion of children quite small (about 5 %).
- % children smallest in 1st class, largest in 3rd class.
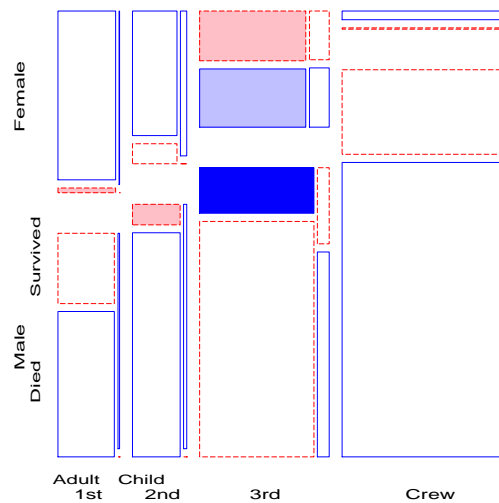- Residuals: greater number of children in 3rd class (families?)

Survival on the *Titanic*: 4 way table

4 way: {Class, Gender, Age} ⊥ Survival ?

- Joint independence: [CGA][S]
- Minimal null model when C, G, A are explanatory
- More women survived, but greater % in 1st & 2nd
- Among men, % survived increases with class.
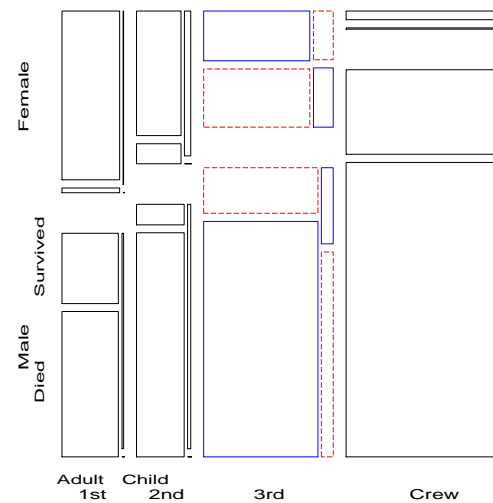- Fits poorly [$G^2_{(15)} = 671.96$] ⇒ Add *S*-assoc terms

Survival on the *Titanic*: Better models

women and children first $\longrightarrow$

- model [*CGA*][**CS**][**GAS**] (Age and Gender affect survival, independent of Class)
- Model improved slightly, but still not good ($G^2_{(9)} = 94.54$).

Survival on the *Titanic*: Better models

Class interacts with Age & Gender on survival:

- Model [*CGA*][**CGS**][**CAS**]
- $G^2_{(4)}$ now 1.69, a very good fit.
- Perhaps too good? (Overfitting?) $\rightarrow$ check AIC!

Titanic Conclusions

Mosaic displays allow a detailed explanation:

- Regardless of Age and Gender, lower economic status $\longrightarrow$ increased mortality.
- Differences due to Class were moderated by both Age and Gender.
- Women more likely *overall* to survive than men, but:
    - Class $\times$ Gender: women in 3rd class *did not* have a significant advantage
    - men in 1st class *did*, compared to men in other classes.
- Class $\times$ Age:
    - no children in 1st or 2nd class died, but
    - nearly two-thirds of children in 3rd class died.
    - For adults, mortality $\uparrow$ as economic class $\downarrow$.
- Summary statement:
  "**women and children (according to class), then 1st class men**".

Summary: Part 3

- **Mosaic displays**
    - Recursive splits of unit square $\rightarrow$ area $\sim$ observed frequency
    - Fit *any* loglinear model $\rightarrow$ shade tiles by residuals
    - $\Rightarrow$ see *departure* of the data from the model
    - SAS: `mosaic` macro, `mosmat` macro; R: `mosaic()`
- **Loglinear models**
    - Loglinear approach: analog of ANOVA for $\log(m_{ijk\cdots})$
    - GLM approach: linear model for $\log(\mathbf{m}) = \mathbf{X}\beta \sim$ Poisson()
    - SAS: PROC `CATMOD`, PROC `GENMOD`; R: `loglm()`, `glm()`
    - Visualize: `mosaic`, `mosmat` macro; R: `mosaic()`
    - Complex tables: sequential plots, partial plots are useful
- **Structured tables**
    - Ordered factors: models using ordinal scores $\rightarrow$ simpler, more powerful
    - Square tables: Test more specific hypotheses about pattern of association
    - SAS: PROC `GENMOD`; R: `glm()`, `gnm()`

References I

Bangdiwala, S. I. Using SAS software graphical procedures for the observer agreement chart. *Proceedings of the SAS User's Group International Conference*, 12:1083–1088, 1987.

Bowker, A. H. Bowker's test for symmetry. *Journal of the American Statistical Association*, 43:572–574, 1948.

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

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. Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Journal of Computational and Graphical Statistics*, 8(3): 373–395, 1999.

References II

Friendly, M. Multidimensional arrays in SAS/IML. In *Proceedings of the SAS User's Group International Conference*, volume 25, pp. 1420–1427. SAS Institute, 2000.

Friendly, M. Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*, 56(4):316–324, 2002.

Friendly, M. and Kwan, E. Effect ordering for data displays. *Computational Statistics and Data Analysis*, 43(4):509–539, 2003.

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*, pp. 268–273. Springer-Verlag, New York, NY, 1981.

Hoaglin, D. C. and Tukey, J. W. Checking the shape of discrete distributions. In Hoaglin, D. C., Mosteller, F., and Tukey, J. W., editors, *Exploring Data Tables, Trends and Shapes*, chapter 9. John Wiley and Sons, New York, 1985.

Koch, G. and Edwards, S. Clinical efficiency trials with categorical data. In Peace, K. E., editor, *Biopharmaceutical Statistics for Drug Development*, pp. 403–451. Marcel Dekker, New York, 1988.

References III

Landis, J. R. and Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174., 1977.

Mersey, L. Report on the loss of the "Titanic" (S. S.). Parliamentary command paper 6352, 1912.

Mosteller, F. and Wallace, D. L. *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. Springer-Verlag, New York, NY, 1984.

Ord, J. K. Graphical methods for a class of discrete distributions. *Journal of the Royal Statistical Society, Series A*, 130:232–238, 1967.

Srole, L., Langner, T. S., Michael, S. T., Kirkpatrick, P., Opler, M. K., and Rennie, T. A. C. *Mental Health in the Metropolis: The Midtown Manhattan Study*. NYU Press, New York, 1978.

Tufte, E. R. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.

Tukey, J. W. Some graphic and semigraphic displays. In Bancroft, T. A., editor, *Statistical Papers in Honor of George W. Snedecor*, pp. 292–316. Iowa State University Press, Ames, IA, 1972.

References IV

Tukey, J. W. *Exploratory Data Analysis*. Addison Wesley, Reading, MA, 1977.

van der Heijden, P. G. M. and de Leeuw, J. Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, 50:429–447, 1985.