

Graphical User Interface for Modifying Structables and their Mosaic Plots

Richard M. Heiberger and Erich Neuwirth
Temple University University of Vienna

A structable object is a representation in R of a k -dimensional contingency table. The structable object has two attributes:

split_vertical that carries information on assignment of the factors to row or columns, and

dnames that carries information on sequencing of the factors.

The printed display of a structable as a *flat* table in two dimensions shows the row and column assignment but is unable to illustrate the joint sequencing of the horizontal and vertical splits.

The default plot of a structable is as a mosaic plot with recursive splits of the factors in the specified sequence and according to the orientation given in **split_vertical**. Each split is along the vertical or horizontal direction associated with the column or row assignment of its factor.

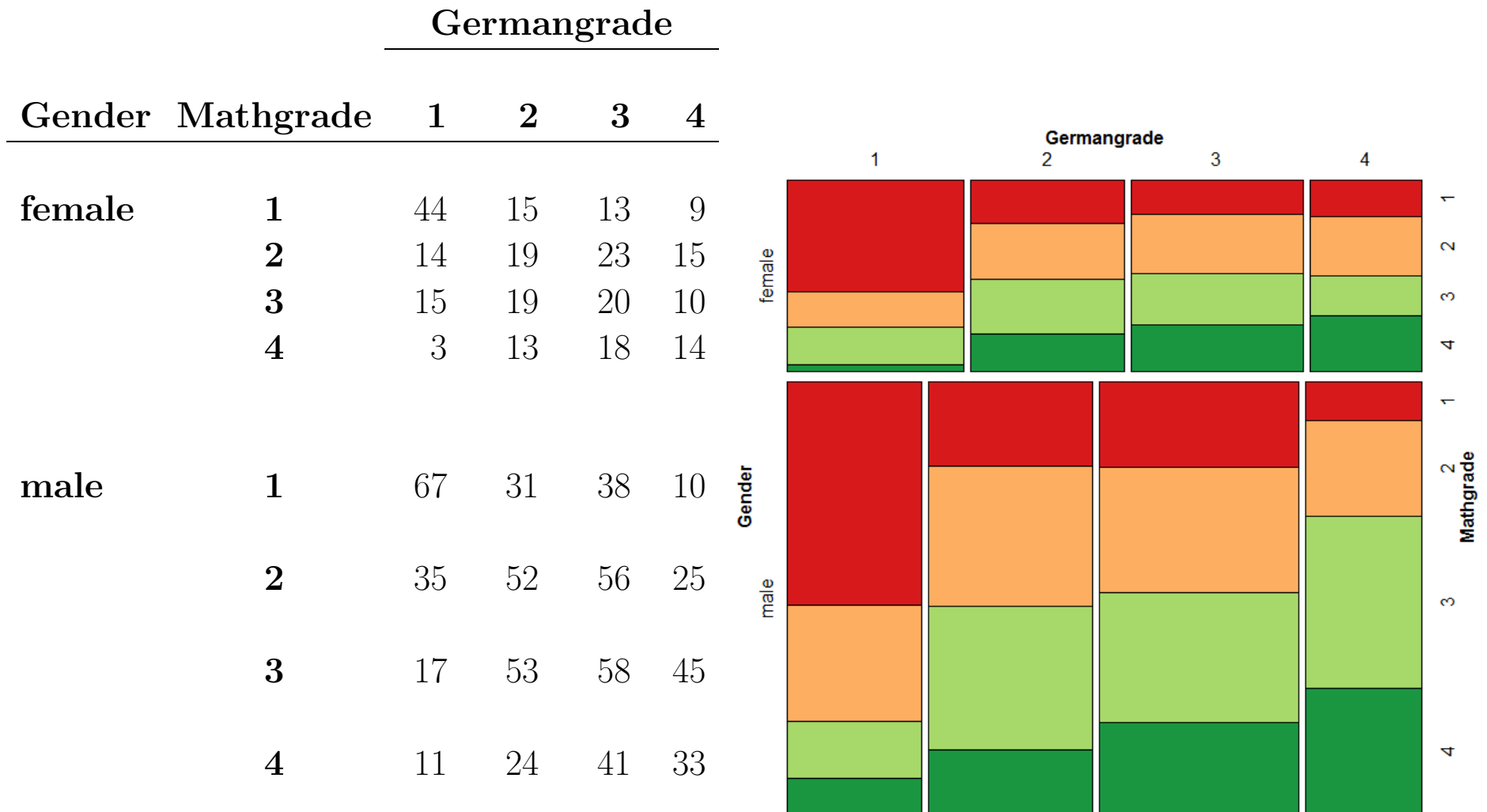


Figure 1: In the left panel we show a three-factor contingency table displayed as a `structable`. In the right panel we show a mosaic plot of the same data. The mosaic uses the default configuration, conditioning first on the outer rows showing Gender, then on the columns showing Germangrade nested within Gender, and finally on the inner rows showing Mathgrade nested within Germangrade nested within Gender.

1 Example

StudentData and Pivot Table by RExcel

The StudentData (Neuwirth, 2011) contains measurements on 1126 Austrian undergraduates collected over a ten year period. We will look at three of the factors here.

Gender: Student's Gender: **m** for man and **w** for woman.

Mathgrade: Discrete values (1, 2, 3, 4) with 1 as the best grade.

Germangrade: Discrete values (1, 2, 3, 4) with 1 as the best grade.

We can construct a contingency table for these factors as a pivot table in Excel and then export it directly to a structable in R (Figure 2), or we can construct the structable directly in R (Figure 3). Either way we have the same structable, called GGM here.

The image shows an Excel PivotTable with the following data:

Count of Gender	Column Labels	1	2	3	4	Grand Total
female		76	66	74	48	
1		44	15	13	9	
2		14	19	23	15	
3		15	19	20	10	
4		3	13	18	14	
male		130	160	193	113	
1		67	31	38	10	
2		35	52	56	25	
3		17	53	58	45	
4		11	24	41	33	
Grand Total		206	226	267	161	

A context menu is open over the PivotTable, with the 'Put Pivottable' option highlighted. A 'Put table in R' dialog box is also open, showing the table name 'GGM' and buttons for 'Get from R', 'Get from Cell', 'OK', and 'Cancel'.

Figure 2: Pivot table in Excel that was constructed from the original data (one observation per student) and then exported to R using the RExcel Put Pivottable context menu item.

```
> GGM <- structable(~ Gender + Germangrade + Mathgrade, StudentData)
> GGM
```

		Germangrade	1	2	3	4
Gender	Mathgrade					
female	1		44	15	13	9
	2		14	19	23	15
	3		15	19	20	10
	4		3	13	18	14
male	1		67	31	38	10
	2		35	52	56	25
	3		17	53	58	45
	4		11	24	41	33

```
>
```

Figure 3: Construction of structable in R with the `structable` function on the dataframe containing the original data (one observation per student).

```
> mosaicPermDialog(GGM)
[1] "mosaic(GGM) "
```

```
>
```

a. R command line
(all platforms)

b. Rcmdr menu from Rcmdr window
(all platforms)

c. Rcmdr menu from Excel
(Windows only)

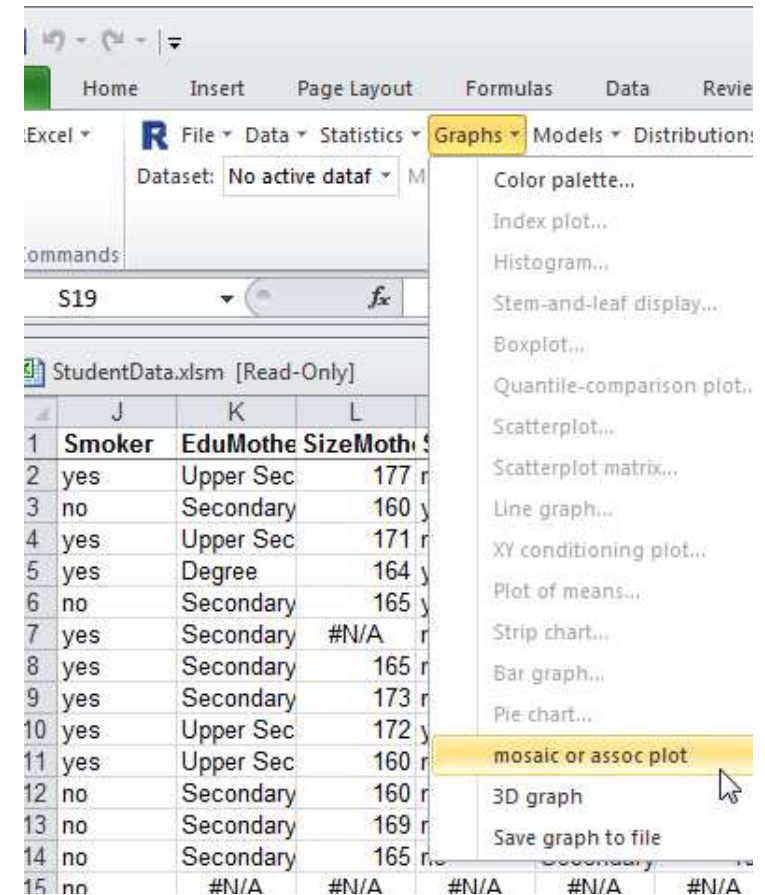
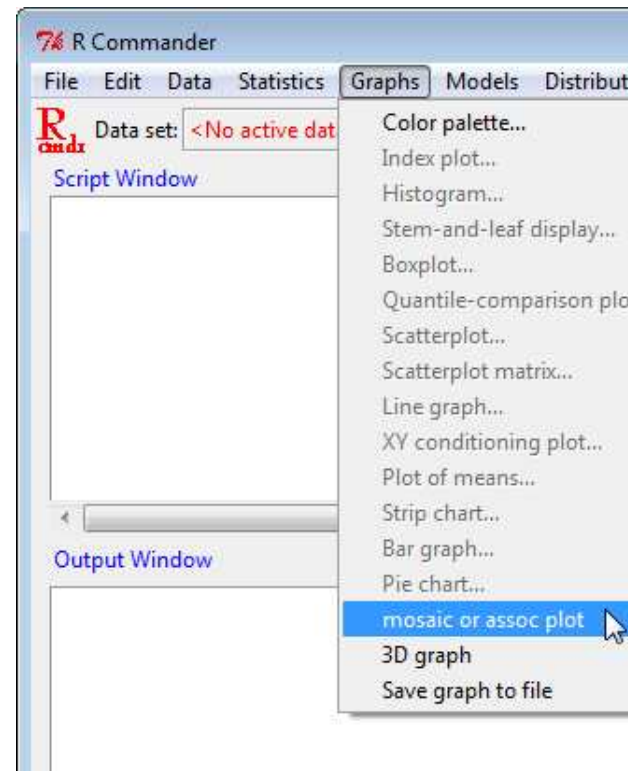


Figure 4: Three equivalent ways to start the mosaic dialog shown in Figure 5.

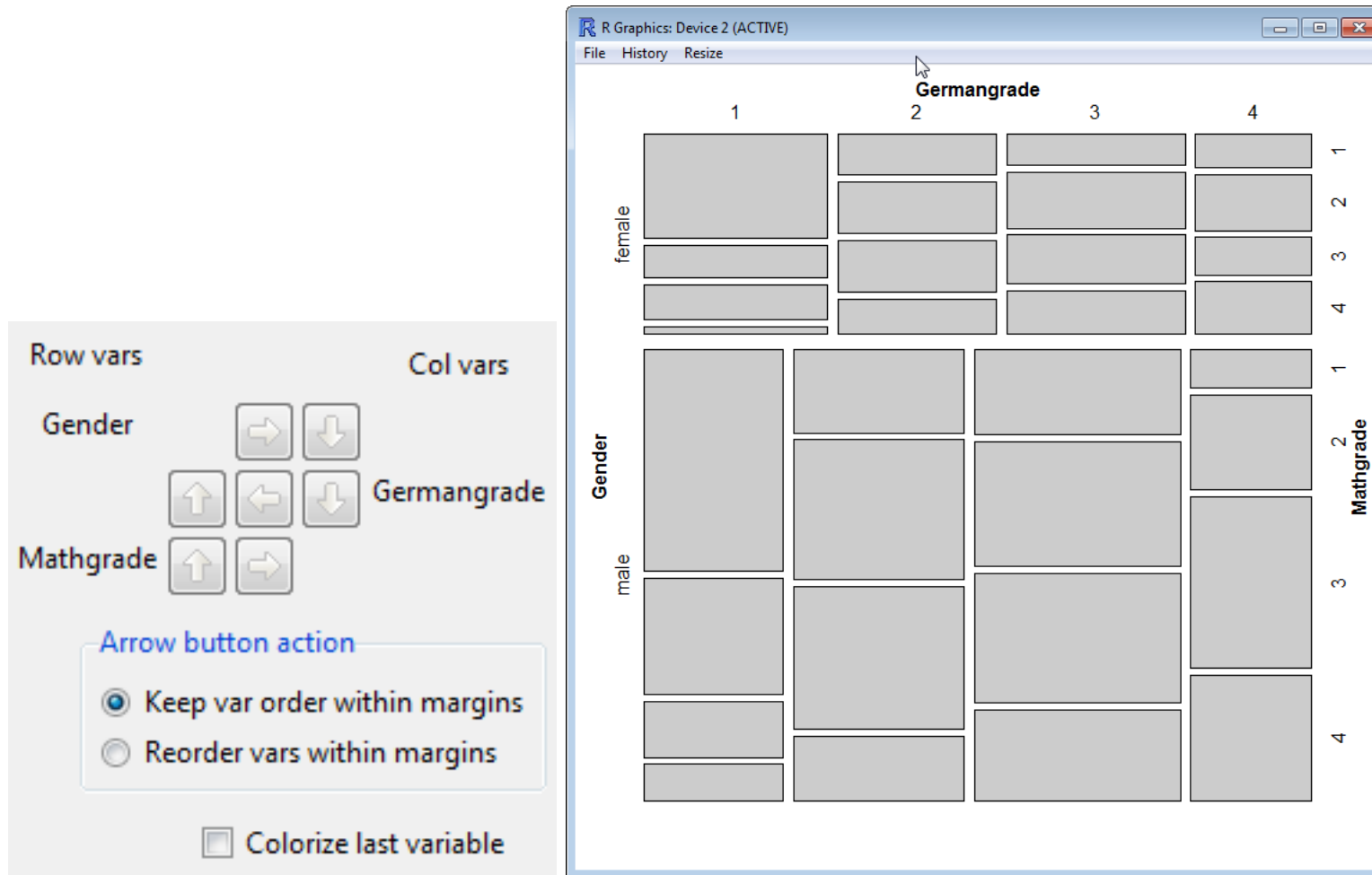


Figure 5: Initial dialog box and mosaic display. The labeled rows of buttons in the dialog box show the sequence of splits illustrated in the mosaic plot. The columns of labels in the dialog box show the assignment of factors to the rows and columns of the structable and mosaic plot. The arrow buttons indicate the specification of split sequence and row/column assignment that are possible from the current display.

The printed representation of a structable displays only the separate derived sequencing of the horizontal and vertical factors. As a consequence, multiple structables—and their associated mosaic plots—can yield the same printed flat table.

We have developed a graphical user interface, and corresponding R functions, that simplify the specification of the alternate sequencing of splits—hence associated mosaic plots—consistent with a printed flat table. The user interface also permits rearrangement of the flat table, either by reassigning factors to rows or columns or by changing the order of the factors within the same row or column assignment. The primary R function is an **aperm** (array permutation) method designed for structables.

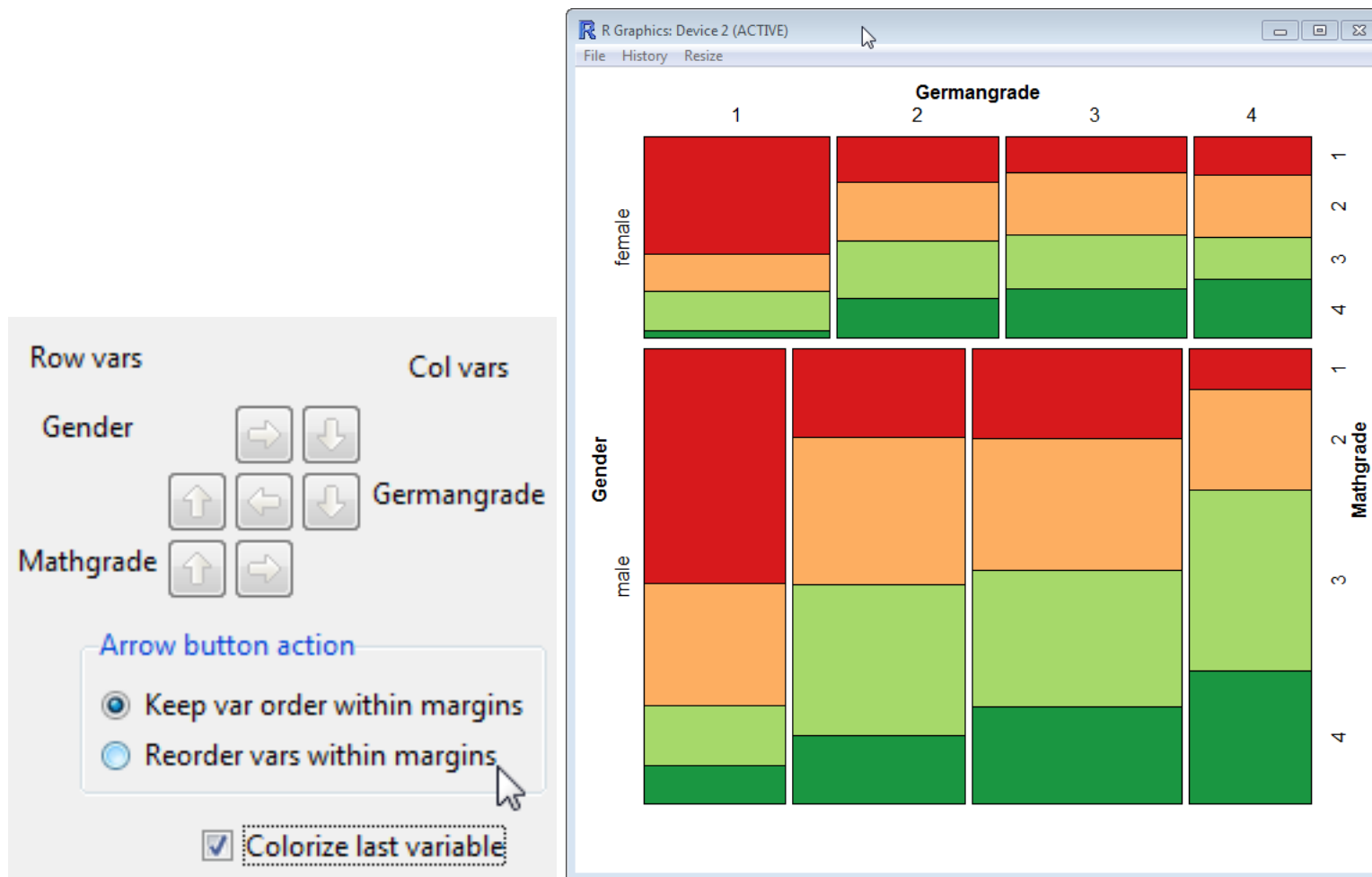


Figure 6: For this and the remaining figures in the paper, we clicked the Colorize last variable checkbox. The coloring matches the last split, Mathgrade here.

We click the German Down arrow to get Figure 7.

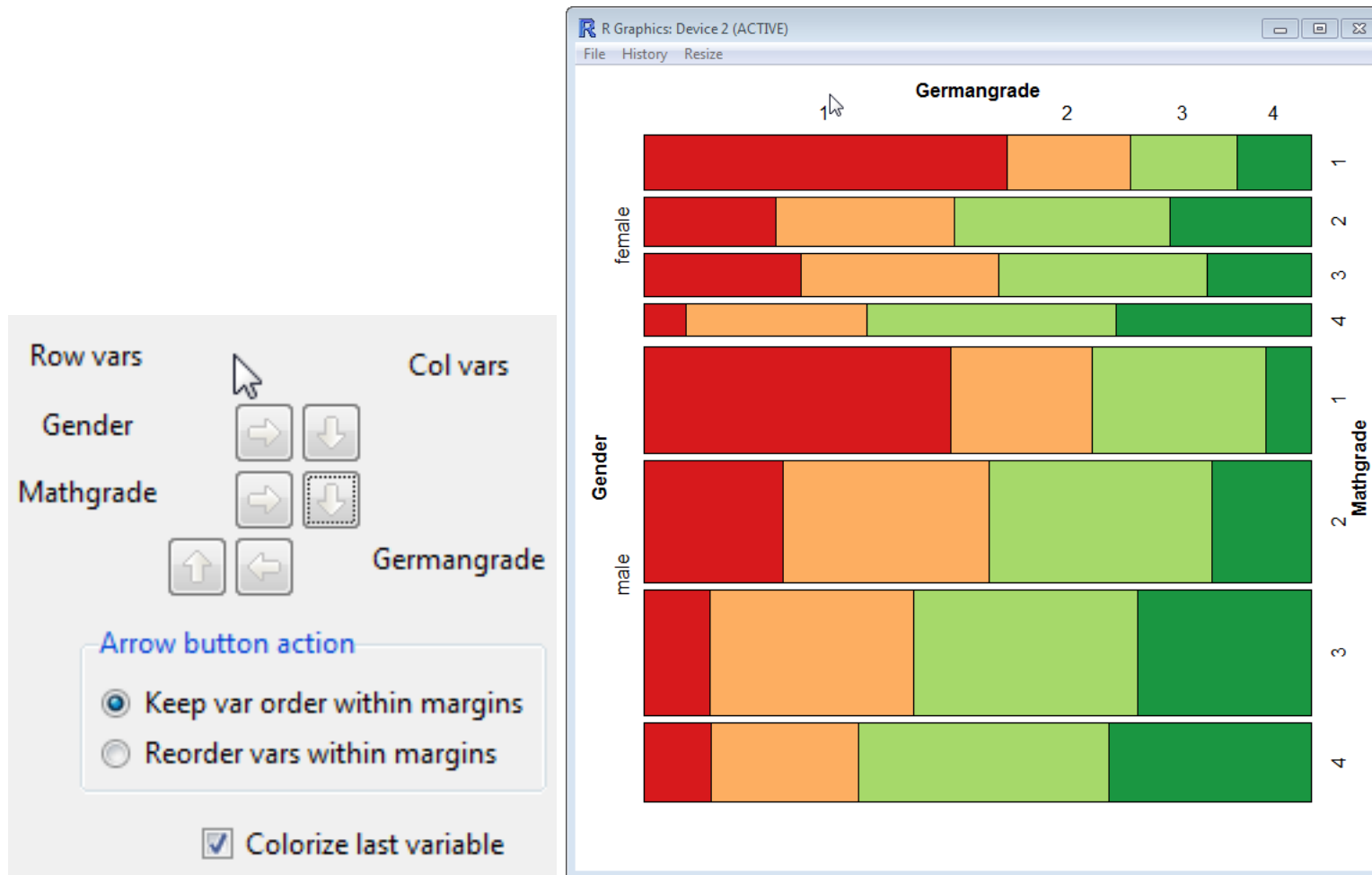


Figure 7: Gender is still the first split. In this figure Mathgrade is the second split. The coloring matches the last split, Germangrade here.

We click the Gender Down arrow to get Figure 8.

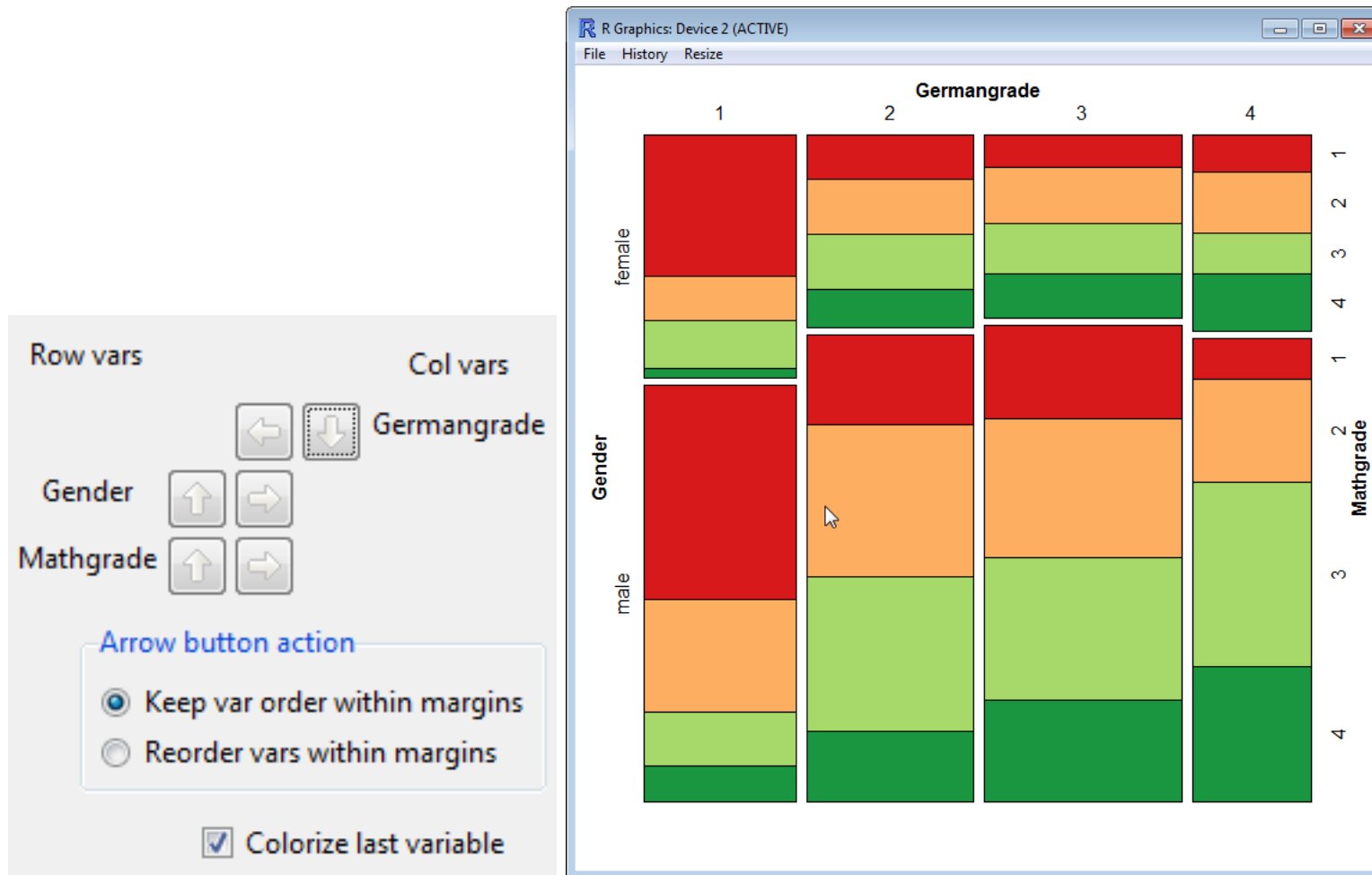


Figure 8: The structable for this figure (Figure 8) is identical to the structable for Figures 6 and 7. The structable is displayed in Figure 11. In all three of these figures, the areas of the rectangles corresponding to each count in the structable are proportional to the count. The interpretation as conditional probabilities of this plot (distribution of Gender by MathGrade conditional on GermanGrade) doesn't make much sense, where the similar interpretation of the other two plots as distributions of the two grades conditional on Gender does make sense.

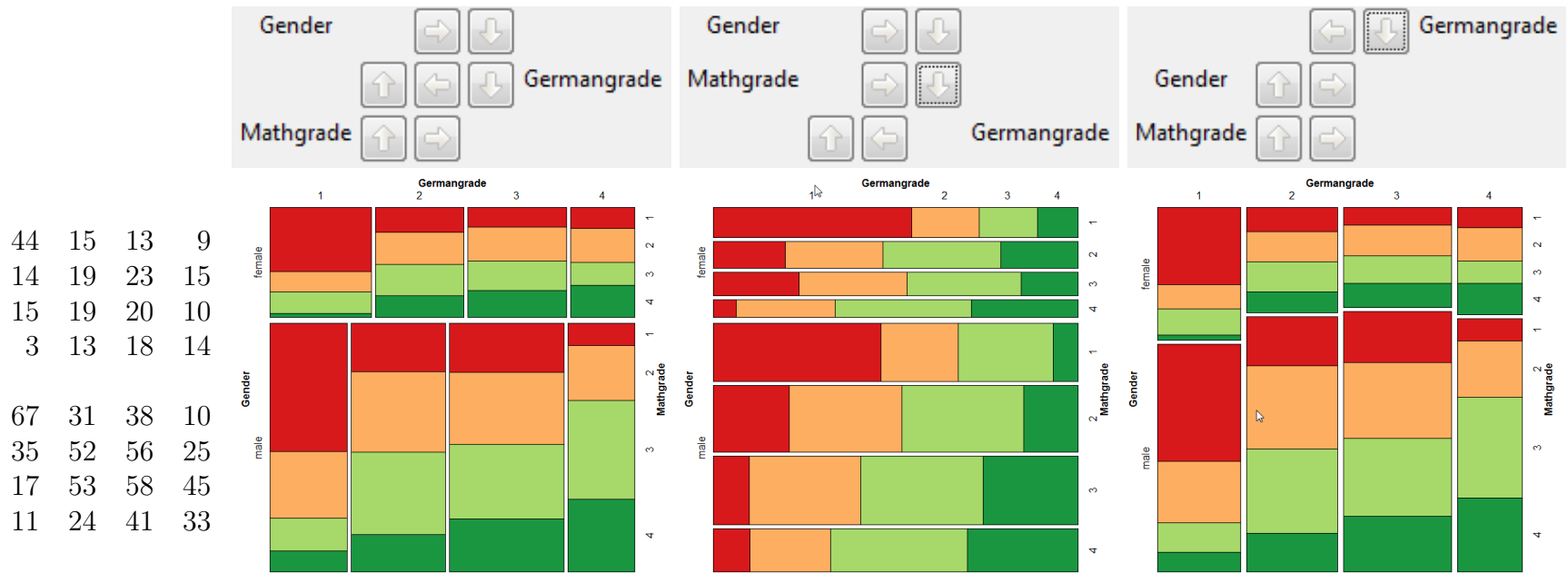


Figure 9: All three mosaic plots that correspond to the same flat display of their structable. The conditioning of the splits differs. The relative position of each cell of the three-way table is the same.

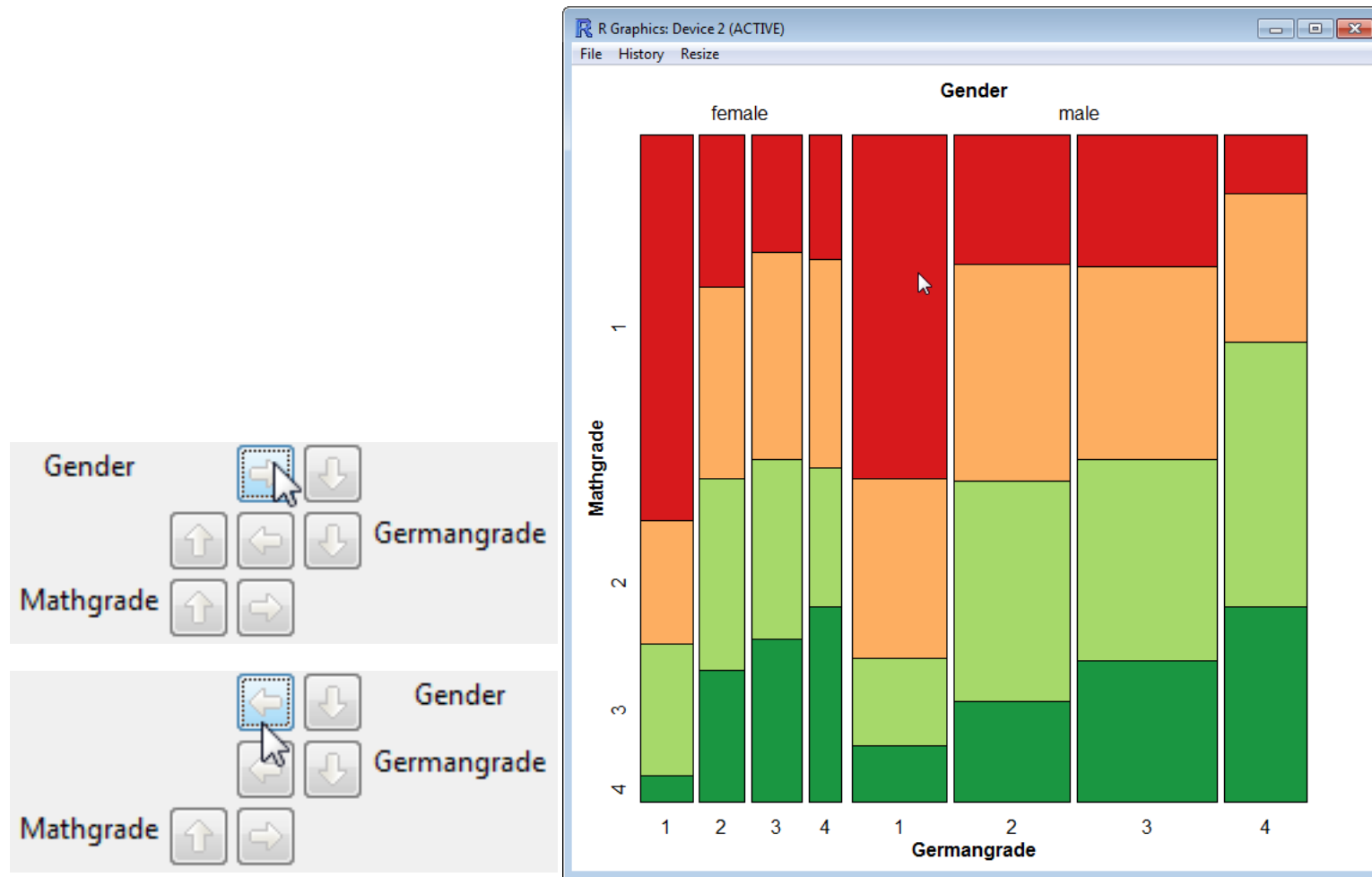


Figure 10: In this figure, we moved Gender from a row factor to a column factor. Now the flat structable has a different appearance.

The dialog box on the left is titled 'Row vars', 'Col vars', and 'Active vars'. It shows 'Gender' and 'Mathgrade' in the row vars, and 'Germangrade' in the col vars. All three are checked in the 'Active vars' column. Below this, there are arrow buttons for moving variables between the lists. The 'Arrow button action' section has two radio buttons: 'Keep var order within margins' (selected) and 'Reorder vars within margins'. There is also a checkbox for 'Colorize last variable'. The 'Graphics type' section has two radio buttons: 'mosaic plot' (selected) and 'assoc plot'. The 'Returned object' section has two radio buttons: 'return plot command' and 'return structable object' (selected). A mouse cursor is pointing at the 'return structable object' option.

The output window on the right shows the following R code and its output:

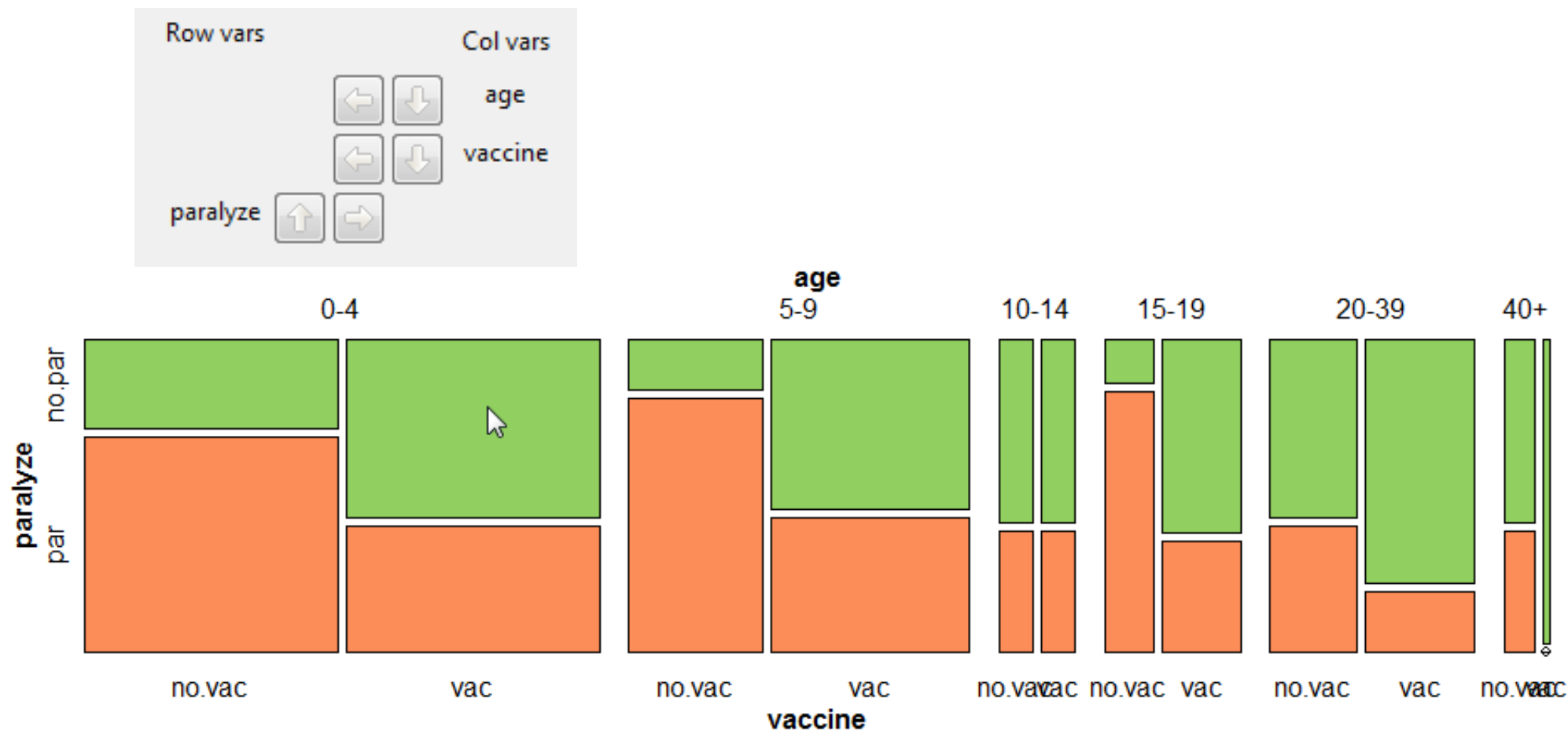
```
> mosaic(GGM)
> mosaic(GGM, highlighting=3, highlighting_fill=br
> GGM
```

		Germangrade			
		1	2	3	4
Gender	Mathgrade				
female	1	44	15	13	9
	2	14	19	23	15
	3	15	19	20	10
	4	3	13	18	14
male	1	67	31	38	10
	2	35	52	56	25
	3	17	53	58	45
	4	11	24	41	33

Figure 11: The dialog box can return either the structable object or the function call that generates the structable object.

Salk Vaccine

(Chin et al, 1961), also in (Agresti, 1990) and in (Heiberger and Holland, 2004), discuss 174 polio cases classified by age of subject, whether or not the subject received the Salk polio vaccine, and whether the subject was ultimately paralyzed by polio. We wish to learn if symptom status (paralysis or not) is independent of vaccination status after controlling for age.



	0-4		5-9		10-14		15-19		20-39		40+	
	no	vac	no	vac	no	vac	no	vac	no	vac	no	vac
no.par	10	20	3	15	3	3	1	7	7	12	3	1
par	24	14	15	12	2	2	6	4	5	3	2	0

Figure 12: Mantel–Haenszel–Cochran test for the Salk polio example. It is easy to see from the mosaic plot that the upper right box in each age group is taller than the upper left box in its own age group. That is, the proportion of cases without paralysis in the vaccinated treatment is higher for all age groups.

References

Erich Neuwirth (2008), Student data collected in classes, 1998–2008, Distributed on CRAN as part of the `RthroughExcelWorkbooksInstaller` package (Heiberger and Neuwirth, 2011).

Richard M. Heiberger (`rmh@temple.edu`) and Erich Neuwirth (`erich.neuwirth@univie.ac.at`) (2009). *R Through Excel: Introductory and Advanced Statistics, Data Analysis, and Graphics with a Spreadsheet Interface and Spreadsheet Tools*, Springer, *Use R* series, ISBN: 9787-1-4419-0051-7.

Richard M. Heiberger (`rmh@temple.edu`) and Erich Neuwirth (`erich.neuwirth@univie.ac.at`) (2011).

`RthroughExcelWorkbooksInstaller`: Excel Workbooks supporting Statistics courses using *R through Excel*. R package version 1.2-6.

<http://CRAN.R-project.org/package=RthroughExcelWorkbooksInstaller>

Erich Neuwirth (`erich.neuwirth@univie.ac.at`) (2011).

`RExcelInstaller`: Integration of R and Excel, (use R in Excel, read/write XLS files). R package version 3.2-2. <http://rcom.univie.ac.at>

David Meyer, Achim Zeileis, and Kurt Hornik (2011). `vcd`: Visualizing Categorical Data. R package version 1.2-11.

David Meyer, Achim Zeileis, and Kurt Hornik (2006). The Strucplot Framework: Visualizing Multi-Way Contingency Tables with `vcd`. *Journal of Statistical Software*, 17(3), 1–48. <http://www.jstatsoft.org/v17/i03/>

Chin, T. W., Hall, E., Gravelle, C., and Speers, J. (1961). “The influence of Salk vaccination on the epidemic pattern and spread of the virus in the community”, *American Journal of Hygiene*, 73:67–94.

Agresti, A. (1990). *Categorical Data Analysis*. Wiley.

Heiberger, R. M. and Holland, B. (2004). *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-Plus, R, and SAS*. Springer. <http://springeronline.com/0-387-40270-5>.

Table Classes

R has 3 classes for tables: **table**, **fable**, and **structable**.

In **table**, all factors have a joint order and there is no horizontal-vertical orientation.

In **fable**, factors are divided into two sets, horizontal and vertical, and only ordered within these sets; there is no joint order.

In **structable**, factors are ordered (as in **table**) and also assigned to either horizontal or vertical (as in **fable**). The separate ordering of the horizontal and vertical factors is derived from the joint ordering.

structable inherits from **fable**, and not **table**. Since printing a **structable** and printing an **fable** produces typographically equivalent results, this seems to make sense. But the print method hides the joint order. The joint order is important for mosaic plots.

table and **structable** have mosaic plots as their default plot method. **fable** does not have a plot method of its own. Plotting an **fable** will produce a not very useful scatterplot. In the object hierarchy of R, **structable** inherits from **fable**, and not from **table**. With respect to the default plotting method, this is a somewhat unfortunate class hierarchy.