Remarks on:

The Structplot Framework: Visualizing Multi-way Contingency Tables with vcd

by D. Meyer, A. Zeileis and K. Hornyk

The paper describes the Structplot Framework, which is the core of the vcd-package in R. This package offers many functions to plot and modify various forms of graphical representations of multivariate categorical data.

The paper is well written and technically sound. Only a few remarks to improve it:

1. The comparison of the different implementations of mosaic plots is nice, but a bit short. For a journal on statistical software it would be desirable to have a couple of examples how the plots/implementations look like, and what the specific advantages and disadvantages are. Furthermore, are there no differences in the implementation of mosaic plots in MANET and Mondrian?

2. The coding of empty cells in the structplot framework is done via boxes of zero area additionally annotated with a "o". This obviously follows the conventions in MANET, except for the fact that MANET (as most other implementations) do not subdivide empty cells any further – which would not really make any sense since the division is done arbitrarily.
   The coding of empty cells should explicitly be explained. Furthermore, the splitting of empty cells should be optionally, and the default should not split the empty cells!

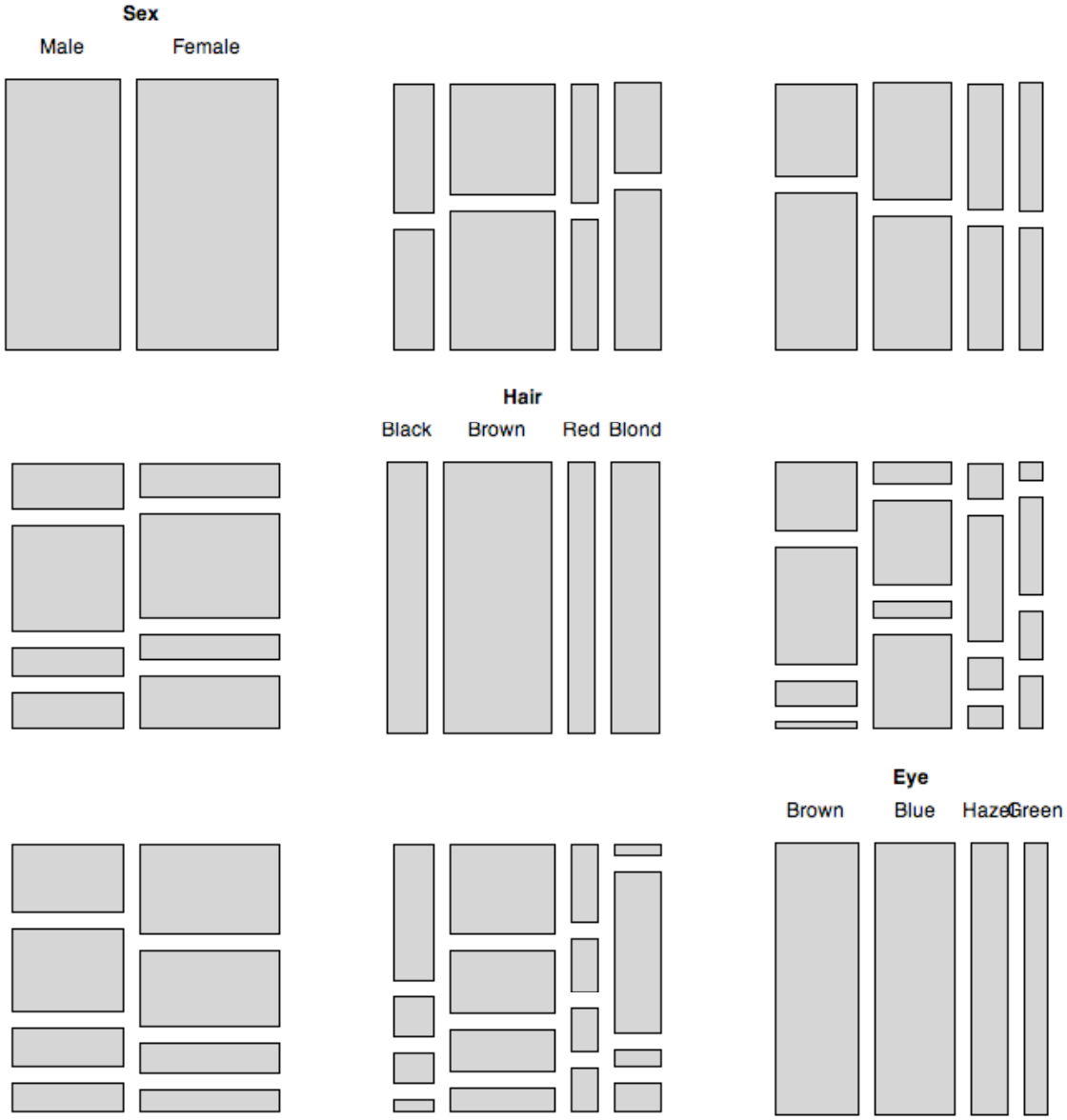
3. On page 9 it is stated that the `doubledecker()` function is just a wrapper for the `mosaic()` function. It would be nice for the reader to actually see how the wrapper (in its structure) looks like.

4. The "sieve plot" on page 9 has gaps, which is not faithful to the original definition by Riewyl & Schupach. Of course, the gaps can be set to anything the user wants to see, but the default should match the original definition – unless there is a good reason for not doing so, which should be mentioned in the paper.

5. Aren't Figure 11 and 14 identical except for a different aspect ratio and a slightly larger gap in the first split? The choice of the example for Figures 13-15 is unfortunate since the groups for "male" and "female" are almost of same size. All-in-all section 2.2 could be a bit clearer in what the options do, why one would use the one or the other.

6. The `pairs()` plot on page 18 has some drawbacks. In a traditional paris-scatterplot, each observation is aligned along two marginal dimensions. In the plot in Figure 16, the groups for eye-color align in the third column, but the sexes don't in the first row. This is due to the conditional nature of mosaic plots.
   A more consistent pairs plot will use the inverse mosaic plot in the lower_panel (which is actually the default!) and probably the matching one-dimensional mosaic plot in the diagonal.

Example:



(not really solving the problem, but a bit more consistent)
At any rate, the association plots in the lower panel are a bit irritating, as for a comparison with the raw data in the upper panel one must transpose the entries.

7. The possibility to modify the panels through their viewports is quite nice. Are the viewports and names global, and not associated with a specific plot? If so, this technique will be pretty error-prone when multiple plots will be used.

8. Section 3.3 does not really fit into the description of the package/framework. I would recommend to leave this section out and put it in a separate, smaller article, e.g. for the R-Newsletter, or the Statistical Graphics and Statistical Computing Newsletter.

9. The footnote on page 42 gives a hint to the `spine()` function to plot spine plots and spinograms. What are spinegrams and how are they defined? Either leave the footnote out, or give a decent example of what the `spine()` function is able to do, especially how a spinegram would look like.

10. When using the package/framework, it turned out to be quite slow. This is obviously due to the extensive use of viewports within Grid. Would it be necessary to spend a section on performance issues for more complex plots than the examples in the current paper?!

Summary:

The paper should definitely be published. Nonetheless, most of the above remarks need to be addressed by the authors in order to make this paper ready for publication.

(I actually didn't look at the code, as the authors should know what they do when programming in R!)