Mosaic Plots and Their Variants

Heike Hofmann

13.1 Definition and Construction .............................................................. 619
13.2 Interpreting Mosaic Plots ................................................................. 622
  Probabilities in Mosaic Plots ............................................................... 622
  Visualizing Interaction Effects ............................................................. 624
13.3 Variants ............................................................................................ 627
  Doubledecker Plots .............................................................................. 627
  Fluctuation Diagrams ......................................................................... 628
  Others .................................................................................................. 632
13.4 Related Work and Generalization .................................................... 635
  Treemaps ........................................................................................... 635
  Trellis Plots ....................................................................................... 636
  Pivot Tables ....................................................................................... 638
13.5 Implementations ................................................................................ 640
In this chapter we consider mosaic plots, which were introduced by Hartigan and Kleiner (1981) as a way of visualizing contingency tables. Named “mosaic plots” due to their resemblance to the art form, they consist of groups of rectangles that represent the cells in a contingency table. Both the sizes and the positions of the rectangles are relevant to mosaic plot interpretation, making them one of the more advanced plots around. With a little practice they can become an invaluable tool in the representation and exploration of multivariate categorical data.

In this chapter we will be discussing ways of constructing and interpreting mosaic plots, including their connection to loglinear models (Hofmann, 2001; Theus and Lauer, 1999; Friendly, 1992). In Sect. 13.1 we will be discussing ways of constructing mosaic plots (Hofmann, 2003). Mosaic plots have the huge advantage of preserving all of the information in multivariate contingency tables while simultaneously providing an overview of it. As the mosaic plot follows the hierarchy its corresponding contingency table exactly, the order of the variables in the table is important. Selecting the “right,” or at least “good,” ordering is commonly found to be one of the main difficulties first-time users experience with mosaic plots. We will discuss the effects of changes in the order and provide recommendations about how to obtain “good plots.”

Multivariate categorical modeling is usually done with loglinear models. It can be shown (Hofmann, 2001; Theus and Lauer, 1999; Friendly, 1992) that mosaic plots have excellent mathematical properties which enable the strengths of interaction effects to be assessed visually and provide tools for checking residuals and modeling assumptions. We will discuss the relationship between mosaic plots and loglinear models in Sect. 13.2.

Close relatives of the mosaic plot, such as fluctuation diagrams and doubledecker plots (Hofmann et al., 2000), have also been found to be very useful in practice. We are therefore going to have a look at those and other important variants of mosaic plots too in Sect. 13.3. All of these variants are essentially simplifications of the default mosaic construction. While some information is lost in the process, these plots place additional emphasis on a specific aspect of the data.

Shneiderman (1992) and trellis plots (Becker et al., 1994) are generalizations of two different aspects of mosaic plots. While trellis plots and mosaic plots share the same structure, trellis plots are more flexible since numbers do not necessarily have to be displayed as rectangles. Treemaps, on the other hand, always use rectangles, but are able to deal with more general partitions than mosaic plots. These generalizations do not come without losses, though. We will compare mosaic plots to these other displays in Sect. 13.4, and comment on the strengths and weaknesses of each. Software implementations of mosaic plots are becoming more frequent. An implementation in R (Gentleman and Ihaka, 1995) was created by Emerson (1998). Mosaic plots in JMP (John Sall, 1989) have some limited interactive features. Fully interactive mosaic plots (Hofmann, 2000) are implemented, e.g., in MANET (Unwin, Hawkins, Hofmann and Siegl, 1997), Mondrian (Theus, 2002) and KLIMT (Urbanek, 2002).