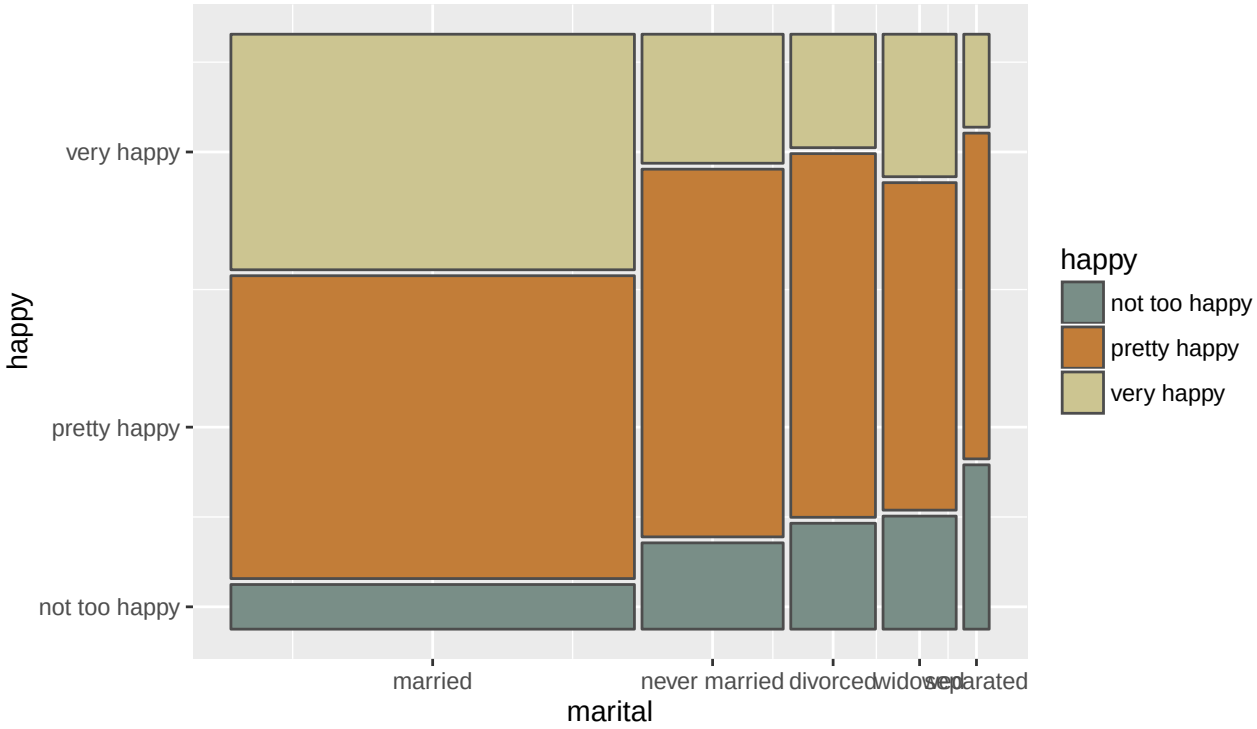
Mosaic Plot Interpretation

Haley Jeppson

March 2, 2016

A mosaic plot is a convenient graphical summary of the conditional distributions in the contingency table. The area of the graphical element is propotional to the underlying probability, so we are easily able to visualize how the joint distribuiton is composed of the product of the conditional and marginal distributions which allows us to see any association that may be occuring between the variables. Because the plot is built hierarchically, the ordering of the variables is very important.

The `productplots` package is used below to produce an example of mosaic plot for two categorical variables. In addition, below the plot is an interpretation of the plot. The data set used in the `happy` data set and the 2 variables to be considered in this example are happy (with 3 levels not too happy, pretty happy, very happy) and marital (with 5 levels married, never married, divorced, widowed, separated). For this example, all `NA`s have been removed.



In this first example of a mosaic plot we are viewing the joint distribuiton of the variables `happy` and `marital` as the product of the marginal distribtion of `marital` and the conditional distribution of `happy` conditioned on the variable `marital`, i.e. $f(happy, marital) = f(happy|marital)f(marital)$. We can consider the rows as the categories of the response variable and the columns are the categories of the explanatory variable. When the segments in the mosaic plot do not line up, there is an indication of an assocation between the variables. For example, from the plot, it appears the marital group `separated` is the least likely to have responded with a `very happy` level of happiness.