SparcMats and Generalized Pairs Plots

John Emerson[1], Walton Green, Avi Feller, and John Hartigan

We would like to present several graphical innovations and/or improvements. Although we hope others will find these tools interesting and useful, we are particularly enthusiastic about the power and flexibility of R for teaching and research. These graphical tools were motivated by research problems and developed in a classroom environment.

First, the proliferation of data in the computer age has made the search for methods of compressing higher dimensions onto a flat, two-dimensional graphical display more difficult and important. We illustrate a graphical, exploratory approach using a generalization of Edward Tufte's "sparklines," that can be used to represent data of continuous variables distributed in space and time. Using this and other graphical tools, we identify certain peculiarities of seven 4-year-long monthly time series of climate data on a 24 by 24 raster covering an area of South and Central America.

Second, we propose a generalized pairs plot, recognizing the fundamental importance of the roles of categorical and quantitative variables. Others have produced pairs plots for categorical variables (with each tile consisting of a small mosaic plot). We introduce a new "barcode" plot, envisioned by John Hartigan (having similarities to side-by-side boxplots and histograms), and we generalize the pairs plot to include (a) scatterplots for pairs of quantitative variables, (b) mosaic plots for pairs of categorical variables, and (c) boxplots and barcode plots for pairs consisting of one categorical and one quantitative variable. Additional options are provided.

[1] Corresponding and contributing author.