

Using density-distribution sunflower plots to explore bivariate relationships in dense data

William D. Dupont
william.dupont@vanderbilt.edu

W. Dale Plummer, Jr.
dale.plummer@vanderbilt.edu

Department of Biostatistics
Vanderbilt University School of Medicine

Abstract. Density-distribution sunflower plots are used to display high-density bivariate data. They are useful for data where a conventional scatterplot is difficult to read due to overstriking of the plot symbol. The x - y plane is subdivided into a lattice of small, regular, hexagonal bins. These bins are divided into low-, medium-, and high-density groups. In low-density bins, the individual observations are plotted as in a conventional scatterplot. Medium- and high-density bins contain light and dark sunflowers, respectively. In a light sunflower, each petal represents one observation. In a dark sunflower, each petal represents a specific number of observations. The user can control the sizes and colors of the sunflowers. By selecting appropriate colors and sizes for the light and dark sunflowers, plots can be obtained that give both the overall sense of the data-density distribution, as well as the number of data points in any given region.

Sunflower plots are also contrasted with contour plots of bivariate kernel-density estimates. The appearance of these plots is markedly affected by the choice of smoothing parameters and the spacing of points at which the probability density function is evaluated. Sunflower plots can be helpful in guiding the selection of these parameters and in distinguishing between chance and systematic variation in the distribution of bivariate data.

Keywords: gr0016, scatterplot, sunflower plot, bivariate data, density plot, probability density function, graphical statistics