

Dupont, W.D. and Plummer, W.D. Jr. (2002), "Density Distribution Sunflower Plots," *Journal of Statistical Software*, <http://www.jstatsoft.org>

Cleveland and McGill (1984) introduced the sunflower plot as a way of graphing high-density bivariate data. We generalize this idea to convey an overall sense of the density distribution of the data while still conveying as much information as possible about the location and number of data points in small regions of the  $x$ - $y$  plane. This graphic, which we call a density distribution sunflower plot, also incorporates ideas due to Carr et al. (1987) and Huang et al. (1997).

Density distribution sunflower plots are useful for data where a conventional scatter plot is difficult to read due to overstriking of the plot symbol. The  $x$ - $y$  plane is subdivided into a lattice of regular hexagonal bins of width  $w$  specified by the user. The user also specifies the values of  $l$ ,  $d$ , and  $k$  that affect the plot as follows. Individual observations are plotted when there are less than  $l$  observations per bin as in a conventional scatter plot. Each bin with from  $l$  to  $d$  observations contains a light sunflower. Other bins contain a dark sunflower. In a light sunflower each petal represents one observation. In a dark sunflower, each petal represents  $k$  observations. (A dark sunflower with  $p$  petals represents between  $pk - k/2$  and  $pk + k/2$  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. The reader can determine the exact location of data points in low density regions, the exact number of data points in bins that contain light sunflowers, and can estimate to within  $k/2$  observations the number of data points in bins with dark sunflowers.

The use of this graphic is illustrated with data from the Framingham Heart Study. A documented Stata program, called *sunflower*, is available to draw these graphs. It can be downloaded from the Statistical Software Components archive at <http://ideas.repec.org/c/boc/bocode/s430201.html>.