sunflower — Density-distribution sunflower plots

Description

sunflower draws density-distribution sunflower plots (Plummer and Dupont 2003). Dark sunflowers, light sunflowers, and marker symbols represent high-, medium-, and low-density regions of the data, respectively. These plots are useful for displaying bivariate data whose density is too great for conventional scatterplots to be effective.

Quick start

Density-distribution sunflower plot showing the relationship between x and y

    sunflower y x

Set the center of the reference bin to $x = 5$ and $y = 8$

    sunflower y x, xcenter(5) ycenter(8)

Same as above, but specify the width of the hexagonal bins to be 1.5

    sunflower y x, xcenter(5) ycenter(8) binwidth(1.5)

Same as above, but set the minimum number of observations needed for a bin to be represented by a light sunflower to 4

    sunflower y x, xcenter(5) ycenter(8) binwidth(1.5) light(4)

Use the s1color scheme

    sunflower y x, scheme(s1color)

Specify that only the petals are shown and the hexagons are omitted

    sunflower y x, flowersonly

Suppress display of the table

    sunflower y x, notable

Menu

Graphics > Smoothing and densities > Density-distribution sunflower plot

Syntax

> sunflower *yvar xvar* [*if*] [*in*] [*weight*] [, *options*]

options	Description
[Main]	
nograph	do not show graph
notable	do not show summary table; implied when by() is specified
marker_options	affect rendition of markers drawn at the plotted points
Bins/Petals	
binwidth(#)	width of the hexagonal bins
binar(#)	aspect ratio of the hexagonal bins
bin_options	affect rendition of hexagonal bins
light(#)	minimum observations for a light sunflower; default is light(3)
dark(#)	minimum observations for a dark sunflower; default is dark(13)
xcenter(#)	x-coordinate of the reference bin
ycenter(#)	y-coordinate of the reference bin
petalweight(#)	observations in a dark sunflower petal
petallength(#)	length of sunflower petal as a percentage
petal_options	affect rendition of sunflower petals
flowersonly	show petals only; do not render bins
nosinglepetal	suppress single petals
Add plots	
addplot(*plot*)	add other plots to generated graph
Y axis, X axis, Titles, Legend, Overall, By	
twoway_options	any options documented in [G-3] *twoway_options*

bin_options	Description
[l|d]bstyle(*areastyle*)	overall look of hexagonal bins
[l|d]bcolor(*colorstyle*)	outline and fill color
[l|d]bfcolor(*colorstyle*)	fill color
[l|d]blstyle(*linestyle*)	overall look of outline
[l|d]blcolor(*colorstyle*)	outline color
[l|d]blwidth(*linewidthstyle*)	thickness of outline

petal_options	Description
[l|d]flstyle(*linestyle*)	overall style of sunflower petals
[l|d]flcolor(*colorstyle*)	color of sunflower petals
[l|d]flwidth(*linewidthstyle*)	thickness of sunflower petals

All options are *rightmost*; see [G-4] **Concept: repeated options**.

fweights are allowed; see **[U] 11.1.6 weight**.

Options

Main

nograph prevents the graph from being generated.

notable prevents the summary table from being displayed. This option is implied when the by() option
is specified.

marker_options affect the rendition of markers drawn at the plotted points, including their shape, size,
color, and outline; see [G-3] *marker_options*.

Bins/Petals

binwidth(#) specifies the horizontal width of the hexagonal bins in the same units as *xvar*. By default,

$$binwidth = \text{max}(\text{rbw}, \text{nbw})$$

where

$$\text{rbw} = \text{range of } xvar/40$$

$$\text{nbw} = \text{range of } xvar/\text{max}(1,\text{nb})$$

and

$$\text{nb} = \text{int}(\text{min}(\text{sqrt}(n),10*\text{log10}(n)))$$

where

$$n = \text{the number of observations in the dataset}$$

binar(#) specifies the aspect ratio for the hexagonal bins. The height of the bins is given by

$$binheight = binwidth \times \# \times 2/\sqrt{3}$$

where *binheight* and *binwidth* are specified in the units of *yvar* and *xvar*, respectively. The default is
binar(*r*), where *r* results in the rendering of regular hexagons.

bin_options affect how the hexagonal bins are rendered.

lbstyle(*areastyle*) and dbstyle(*areastyle*) specify the look of the light and dark hexagonal bins,
respectively. The options listed below allow you to change each attribute, but lbstyle() and
dbstyle() provide the starting points. See [G-4] *areastyle* for a list of available area styles.

lbcolor(*colorstyle*) and dbcolor(*colorstyle*) specify one color to be used both to outline the shape
and to fill the interior of the light and dark hexagonal bins, respectively. See [G-4] *colorstyle* for a
list of color choices.

lbfcolor(*colorstyle*) and dbfcolor(*colorstyle*) specify the color to be used to fill the interior of
the light and dark hexagonal bins, respectively. See [G-4] *colorstyle* for a list of color choices.

lblstyle(*linestyle*) and dblstyle(*linestyle*) specify the overall style of the line used to outline the
area, which includes its pattern (solid, dashed, etc.), thickness, and color. The other options listed
below allow you to change the line's attributes, but lblstyle() and dblstyle() are the starting
points. See [G-4] *linestyle* for a list of choices.

lblcolor(*colorstyle*) and dblcolor(*colorstyle*) specify the color to be used to outline the light and
dark hexagonal bins, respectively. See [G-4] *colorstyle* for a list of color choices.

lblwidth(*linewidthstyle*) and dblwidth(*linewidthstyle*) specify the thickness of the line to be used to outline the light and dark hexagonal bins, respectively. See [G-4] *linewidthstyle* for a list of choices.

light(*#*) specifies the minimum number of observations needed for a bin to be represented by a light sunflower. The default is light(3).

dark(*#*) specifies the minimum number of observations needed for a bin to be represented by a dark sunflower. The default is dark(13).

xcenter(*#*) and ycenter(*#*) specify the center of the reference bin. The default values are the median values of *xvar* and *yvar*, respectively. The centers of the other bins are implicitly defined by the location of the reference bin together with the common bin width and height.

petalweight(*#*) specifies the number of observations represented by each petal of a dark sunflower. The default value is chosen so that the maximum number of petals on a dark sunflower is 14.

petallength(*#*) specifies the length of petals in the sunflowers. The value specified is interpreted as a percentage of half the bin width. The default is 100%.

petal_options affect how the sunflower petals are rendered.

lflstyle(*linestyle*) and dflstyle(*linestyle*) specify the overall style of the light and dark sunflower petals, respectively.

lflcolor(*colorstyle*) and dflcolor(*colorstyle*) specify the color of the light and dark sunflower petals, respectively.

lflwidth(*linewidthstyle*) and dflwidth(*linewidthstyle*) specify the width of the light and dark sunflower petals, respectively.

flowersonly suppresses rendering of the bins. This option is equivalent to specifying lbcolor(none) and dbcolor(none).

nosinglepetal suppresses flowers from being drawn in light bins that contain only 1 observation and dark bins that contain as many observations as the petal weight (see the petalweight() option).

___ Add plots ___

addplot(*plot*) provides a way to add other plots to the generated graph; see [G-3] *addplot_option*.

___ Y axis, X axis, Titles, Legend, Overall, By ___

twoway_options are any of the options documented in [G-3] *twoway_options*. These include options for titling the graph (see [G-3] *title_options*), options for saving the graph to disk (see [G-3] *saving_option*), and the by() option (see [G-3] *by_option*).

Remarks and examples

A sunflower is several line segments of equal length, called petals, that radiate from a central point. There are two varieties of sunflowers: light and dark. Each petal of a light sunflower represents 1 observation. Each petal of a dark sunflower represents several observations. Dark and light sunflowers represent high- and medium-density regions of the data, and marker symbols represent individual observations in low-density regions.

sunflower divides the plane defined by the variables *yvar* and *xvar* into contiguous hexagonal bins. The number of observations contained within a bin determines how the bin will be represented.

- When there are fewer than light(#) observations in a bin, each point is plotted using the usual marker symbols in a scatterplot.

- Bins with at least light(#) but fewer than dark(#) observations are represented by a light sunflower. Each petal of a light sunflower represents one observation in the bin.

- Bins with at least dark(#) observations are represented by a dark sunflower. Each petal of a dark sunflower represents multiple observations.

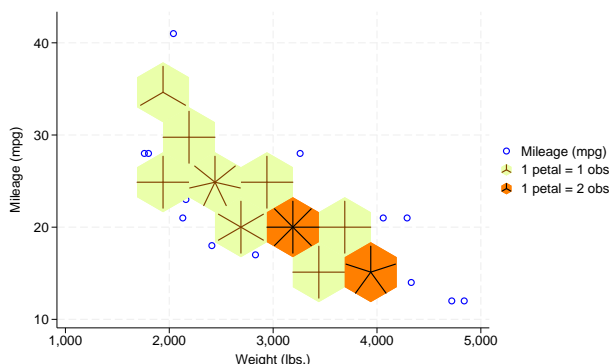See Dupont (2009, 87–92) for a discussion of sunflower plots and how to create them using Stata.

▷ Example 1

Using the auto dataset, we want to examine the relationship between weight and mpg. To do that, we type

```
. use https://www.stata-press.com/data/r19/auto
(1978 automobile data)
. sunflower mpg weight, binwid(500) petalw(2) dark(8) scheme(stcolor)
Bin width          =        500
Bin height         =    6.50422
Bin aspect ratio   =  .0112656
Max obs in a bin   =         15
Light              =          3
Dark               =          8
X-center           =       3190
Y-center           =         20
Petal weight       =          2
```

Flower type	Petal weight	No. of petals	No. of flowers	Estimated obs	Actual obs
none				13	13
light	1	3	1	3	3
light	1	4	5	20	20
light	1	6	1	6	6
light	1	7	1	7	7
dark	2	5	1	10	10
dark	2	8	1	16	15
				75	74

The two darkly shaded sunflowers immediately catch our eyes, indicating a group of 10 cars that are heavy (nearly 4,000 pounds) and fuel inefficient and a group 16 of cars that get about 20 miles per gallon and weight in the neighborhood of 3,000 pounds. The lighter sunflowers each indicate groups of three to seven cars that share similar weight and fuel economy characteristics. To obtain the number of cars in each group, we count the number of petals in each flower and consult the graph legend to see how many observations each petal represents.

◁

Acknowledgments

References

Cleveland, W. S., and R. McGill. 1984. The many faces of a scatterplot. *Journal of the American Statistical Association* 79: 807–822. https://doi.org/10.2307/2288711.

Dupont, W. D. 2009. *Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data*. 2nd ed. Cambridge: Cambridge University Press.

Dupont, W. D., and W. D. Plummer, Jr. 2005. Using density-distribution sunflower plots to explore bivariate relationships in dense data. *Stata Journal* 5: 371–384.

Huang, C., J. A. McDonald, and W. Stuetzle. 1997. Variable resolution bivariate plots. *Journal of Computational and Graphical Statistics* 6: 383–396. https://doi.org/10.1080/10618600.1997.10474749.

Levy, D. E. 1999. *50 Years of Discovery: Medical Milestones from the National Heart, Lung, and Blood Institute's Framingham Heart Study*. Hoboken, NJ: Center for Bio-Medical Communication.

Plummer, W. D., Jr., and W. D. Dupont. 2003. Density distribution sunflower plots. *Journal of Statistical Software* 8: art. 3. https://doi.org/10.18637/jss.v008.i03.

Steichen, T. J., and N. J. Cox. 1999. flower: Stata module to draw sunflower plots. Statistical Software Components S393001, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s393001.html.

For suggested citations, see the FAQ on citing Stata documentation.