**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)
```

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

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

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

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>nograph</strong></td>
</tr>
<tr>
<td><strong>notable</strong></td>
</tr>
</tbody>
</table>

#### Bins/Petals

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>binwidth(#)</strong></td>
</tr>
<tr>
<td><strong>binar(#)</strong></td>
</tr>
<tr>
<td><strong>light(#)</strong></td>
</tr>
<tr>
<td><strong>dark(#)</strong></td>
</tr>
<tr>
<td><strong>xcenter(#)</strong></td>
</tr>
<tr>
<td><strong>ycenter(#)</strong></td>
</tr>
<tr>
<td><strong>petalweight(#)</strong></td>
</tr>
<tr>
<td><strong>petallength(#)</strong></td>
</tr>
<tr>
<td><strong>floweronly</strong></td>
</tr>
<tr>
<td><strong>nosinglepetal</strong></td>
</tr>
</tbody>
</table>

#### Add plots

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>addplot(plot)</strong></td>
</tr>
</tbody>
</table>

### twoway_options

- any options documented in [G-3] *twoway_options*  

#### bin_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>l d bstyle(areastyle)</strong></td>
</tr>
<tr>
<td><strong>l d bcolor(colorstyle)</strong></td>
</tr>
<tr>
<td><strong>l d bcolor(colorstyle)</strong></td>
</tr>
<tr>
<td><strong>l d blstyle(linestyle)</strong></td>
</tr>
<tr>
<td><strong>l d bicolor(colorstyle)</strong></td>
</tr>
<tr>
<td><strong>l d blwidth(linewidthstyle)</strong></td>
</tr>
</tbody>
</table>

#### petal_options

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>l d fli style(linestyle)</strong></td>
</tr>
<tr>
<td><strong>l d flicolor(colorstyle)</strong></td>
</tr>
<tr>
<td><strong>l d flwidth(linewidthstyle)</strong></td>
</tr>
</tbody>
</table>

All options are *rightmost*; see [G-4] *Concept: repeated options*.  
*weight* are allowed; see [U] 11.1.6 weight.
Options

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.

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

\[
binwidth = \max(rb, nb) 
\]

where

\[
rb = \text{range of } xvar/40
\]

\[
lb = \text{range of } xvar/\max(1, nb)
\]

and

\[
lb = \lfloor \min(\text{sqrt}(n), 10 \times \log10(n)) \rfloor
\]

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(area) and dbstyle(area) 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] area for a list of available area styles.

lbcolor(color) and dbcolor(color) 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] color for a list of color choices.

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

lblstyle(linestyle) and dblblstyle(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 dblblstyle() are the starting points. See [G-4] linestyle for a list of choices.

lblcolor(color) and dblblcolor(color) specify the color to be used to outline the light and dark hexagonal bins, respectively. See [G-4] color for a list of color choices.
**sunflower — Density-distribution sunflower plots**

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).

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

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/r17/auto
(1978 automobile data)
. sunflower mpg weight, binwid(500) petalw(2) dark(8) scheme(s2color)
```

---

<table>
<thead>
<tr>
<th>Flower type</th>
<th>Petal weight</th>
<th>No. of petals</th>
<th>No. of flowers</th>
<th>Estimated obs</th>
<th>Actual obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>light</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>light</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>light</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>dark</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>dark</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>dark</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>16</td>
<td>15</td>
</tr>
</tbody>
</table>

76 74
The three darkly shaded sunflowers immediately catch our eyes, indicating a group of eight cars that are heavy (nearly 4,000 pounds) and fuel inefficient and two groups of cars that get about 20 miles per gallon and weight in the neighborhood of 3,000 pounds, one with 10 cars and one with 8 cars. The lighter sunflowers with seven petals each indicate groups of seven cars that share similar weight and fuel economy characteristics. To obtain the number of cars in each group, we counted the number of petals in each flower and consulted the graph legend to see how many observations each petal represents.

Acknowledgments

We thank William D. Dupont and W. Dale Plummer Jr., both of the Department of Biostatistics at Vanderbilt University, who are the authors of the original sunflower command, for their assistance in producing this version.

References


