Description

A dotplot is a scatterplot with values grouped together vertically (“binning”, as in a histogram) and with plotted points separated horizontally. The aim is to display all the data for several variables or groups in one compact graphic.

Quick start

Dotplot of v1
    dotplot v1

Columns with separate dotplots of v1 for each level of categorical variable a
    dotplot v1, over(a)

As above, but with the dots centered in each column
    dotplot v1, over(a) center

Dotplots for v1, v2, and v3 in separate columns
    dotplot v1 v2 v3

Add a horizontal line of pluses at the mean of each variable
    dotplot v1 v2 v3, mean

Add pluses for the medians and dashed lines for the upper and lower quartiles
    dotplot v1 v2 v3, median bar

Menu

Graphics > Distributional graphs > Distribution dotplot
Dotplot of varname, with one column per value of groupvar

\texttt{dotplot \ varname \ [if] \ [in] \ [\ , \ options]} \\

Dotplot for each variable in varlist, with one column per variable

\texttt{dotplot \ varlist \ [if] \ [in] \ [\ , \ options]} \\

\textbf{Options} \\

\begin{tabular}{ll}
\textbf{Options} & \textbf{Description} \\
\hline
\texttt{over(\ varname)} & display one columnar dotplot for each value of \texttt{\ varname} \\
\texttt{nx(\#)} & horizontal dot density; default is \texttt{nx(0)} \\
\texttt{ny(\#)} & vertical dot density; default is \texttt{ny(35)} \\
\texttt{incr(\#)} & label every \# group; default is \texttt{incr(1)} \\
\texttt{mean\|median} & plot a horizontal line of pluses at the mean or median \\
\texttt{bounded} & use minimum and maximum as boundaries \\
\texttt{bar} & plot horizontal dashed lines at shoulders of each group \\
\texttt{nogroup} & use the actual values of \texttt{\ yvar} \\
\texttt{center} & center the dot for each column \\
\end{tabular}

Plot

\texttt{marker\_options} & change look of markers (color, size, etc.) \\
\texttt{marker\_label\_options} & add marker labels; change look or position \\

Y axis, X axis, Titles, Legend, Overall

\texttt{twoway\_options} & any options other than \texttt{by()} documented in \texttt{[G-3] twoway\_options}

\textbf{Options}

\texttt{over(\ varname)} identifies the variable for which \texttt{dotplot} will display one columnar dotplot for each value of \texttt{\ varname}. \texttt{over()} may not be specified in the second syntax.

\texttt{nx(\#)} sets the horizontal dot density. A larger value of \# will increase the dot density, reducing the horizontal separation between dots. This option will increase the separation between columns if two or more groups or variables are used.

\texttt{ny(\#)} sets the vertical dot density (number of “bins” on the \texttt{y axis}). A larger value of \# will result in more bins and a plot that is less spread out horizontally. \# should be determined in conjunction with \texttt{nx()} to give the most pleasing appearance.

\texttt{incr(\#)} specifies how the \texttt{x axis} is to be labeled. \texttt{incr(1)}, the default, labels all groups. \texttt{incr(2)} labels every second group.

\texttt{[mean|median]} plots a horizontal line of pluses at the mean or median of each group.

\texttt{bounded} forces the minimum and maximum of the variable to be used as boundaries of the smallest and largest bins. It should be used with one variable whose support is not the whole of the real line and whose density does not tend to zero at the ends of its support, for example, a uniform random variable or an exponential random variable.
bar plots horizontal dashed lines at the “shoulders” of each group. The shoulders are taken to be the upper and lower quartiles unless mean has been specified; here they will be the mean plus or minus the standard deviation.
	nogroup uses the actual values of yvar rather than grouping them (the default). This option may be useful if yvar takes on only a few values.

center centers the dots for each column on a hidden vertical line.

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

marker_label_options specify if and how the markers are to be labeled; see [G-3] marker_label_options. marker_label_options are not allowed if varlist is specified.

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

Remarks and examples

dotplot produces a figure that has elements of a boxplot, a histogram, and a scatterplot. Like a boxplot, it is most useful for comparing the distributions of several variables or the distribution of 1 variable in several groups. Like a histogram, the figure provides a crude estimate of the density, and, as with a scatterplot, each symbol (dot) represents 1 observation.

Example 1

dotplot may be used as an alternative to Stata’s histogram graph for displaying the distribution of one variable.
Example 2

The `over()` option lets us use `dotplot` to compare the distribution of one variable within different levels of a grouping variable. The `center`, `median`, and `bar` options create a graph that may be compared with Stata’s boxplot; see [G-2] `graph box`. The next graph illustrates this option with Stata’s automobile dataset.

```
. use http://www.stata-press.com/data/r15/auto, clear
   (1978 Automobile Data)
. dotplot mpg, over(foreign) nx(25) ny(10) center median bar
```
Example 3

The second version of *dotplot* lets us compare the distribution of several variables. In the next graph, all 10 variables contain measurements on tumor volume.

```
. use http://www.stata-press.com/data/r15/dotgr
. dotplot g1r1-g1r10, ytitle("Tumor volume, cu mm")
```

![Graph showing tumor volume measurements for 10 variables](image)
Example 4

When using the first form with the `over()` option, we can encode a third dimension in a dotplot by using a different plotting symbol for different groups. The third dimension cannot be encoded with a varlist. The example is of a hypothetical matched case–control study. The next graph shows the exposure of each individual in each matched stratum. Cases are marked by the letter ‘x’, and controls are marked by the letter ‘o’.

```
. use http://www.stata-press.com/data/r15/dotdose
. label define symbol 0 "o" 1 "x"
. label values case symbol
. dotplot dose, over(strata) m(none) mlabel(case) mlabp(0) center
```

Example 5

`dotplot` can also be used with two virtually continuous variables as an alternative to jittering the data to distinguish ties. We must use the `xlabel()` option, because otherwise `dotplot` will attempt to label too many points on the x axis. It is often useful in such instances to use a value of `nx` that is smaller than the default. That was not necessary in this example, partly because of our choice of symbols.

```
. use http://www.stata-press.com/data/r15/auto
(1978 Automobile Data)
. generate byte hi_price = (price>10000) if price < .
. label define symbol 0 "|" 1 "o"
. label values hi_price symbol
```
Example 6

The following figure is included mostly for aesthetic reasons. It also demonstrates dotplot’s ability to cope with even very large datasets. The sample size for each variable is 10,000, so it may take a long time to print.

```
. clear all
. set seed 123456789
. set obs 10000
. generate norm0 = rnormal()
. generate norm1 = rnormal() + 1
. generate norm2 = rnormal() + 2
. label variable norm0 "N(0,1)"
. label variable norm1 "N(1,1)"
. label variable norm2 "N(2,1)"
. dotplot norm0 norm1 norm2
```
 Stored results

dotplot stores the following in r():

Scalars
\( r(nx) \)  horizontal dot density
\( r(ny) \)  vertical dot density

Acknowledgments

dotplot was written by Peter Sasieni of the Wolfson Institute of Preventive Medicine, London, and Patrick Royston of the MRC Clinical Trials Unit, London, and coauthor of the Stata Press book *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*.

References
