Title

stata.com

dotplot — Comparative scatterplots

Syntax

Dotplot of varname, with one column per value of groupvar

```
dotplot varname [if] [in] [, options]
```

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

```
dotplot varlist [if] [in] [, options]
```

options	Description
Options	
`over(`*groupvar*`)`	display one columnar dotplot for each value of *groupvar*
`nx(#)`	horizontal dot density; default is `nx(0)`
`ny(#)`	vertical dot density; default is `ny(35)`
`incr(#)`	label every # group; default is `incr(1)`
`mean` | `median`	plot a horizontal line of pluses at the mean or median
`bounded`	use minimum and maximum as boundaries
`bar`	plot horizontal dashed lines at shoulders of each group
`nogroup`	use the actual values of *yvar*
`center`	center the dot for each column
Plot	
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than `by()` documented in [G-3] ***twoway_options***

Menu

Graphics > Distributional graphs > Distribution dotplot

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.

In the first syntax, `dotplot` produces a columnar dotplot of *varname*, with one column per value of *groupvar*. In the second syntax, `dotplot` produces a columnar dotplot for each variable in *varlist*, with one column per variable; `over(`*groupvar*`)` is not allowed. In each case, the "dots" are plotted as small circles to increase readability.

Options

Options

`over(`*groupvar*`)` identifies the variable for which `dotplot` will display one columnar dotplot for each value of *groupvar*.

`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.

`ny(#)` sets the vertical dot density (number of "bins" on the 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 `nx()` to give the most pleasing appearance.

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

[`mean` | `median`] plots a horizontal line of pluses at the mean or median of each group.

`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.

Plot

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***.

Y axis, X axis, Titles, Legend, Overall

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

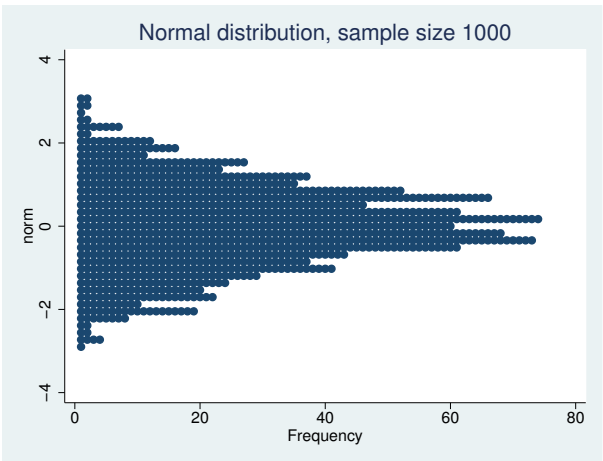
stata.com

`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.

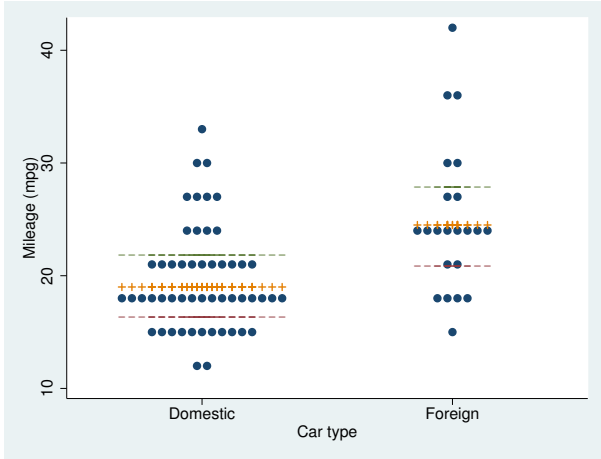
```
. set seed 123456789
. set obs 1000
. generate norm = rnormal()
. dotplot norm, title("Normal distribution, sample size 1000")
```



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.

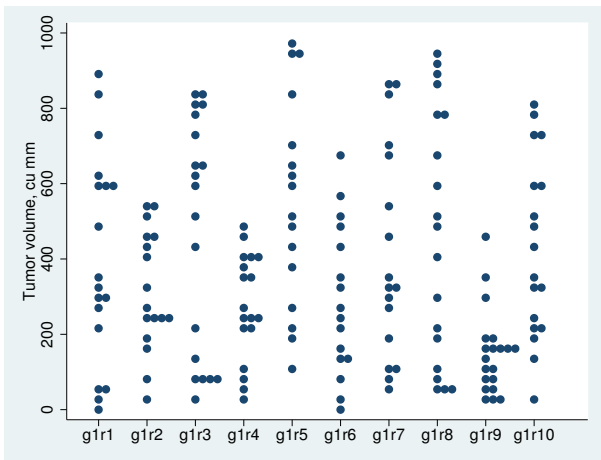
```
. use http://www.stata-press.com/data/r13/auto, clear
(1978 Automobile Data)
. dotplot mpg, over(foreign) nx(25) ny(10) center median bar
```



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.

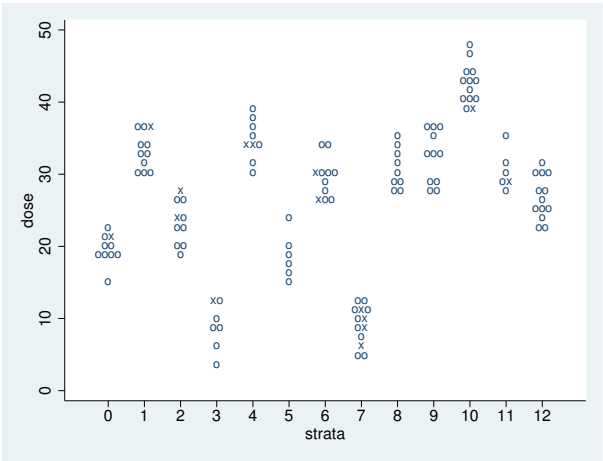
```
. use http://www.stata-press.com/data/r13/dotgr
. dotplot g1r1-g1r10, ytitle("Tumor volume, cu mm")
```



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/r13/dotdose
. label define symbol 0 "o" 1 "x"
. label values case symbol
. dotplot dose, over(strata) m(none) mlabel(case) mlabp(0) center
```

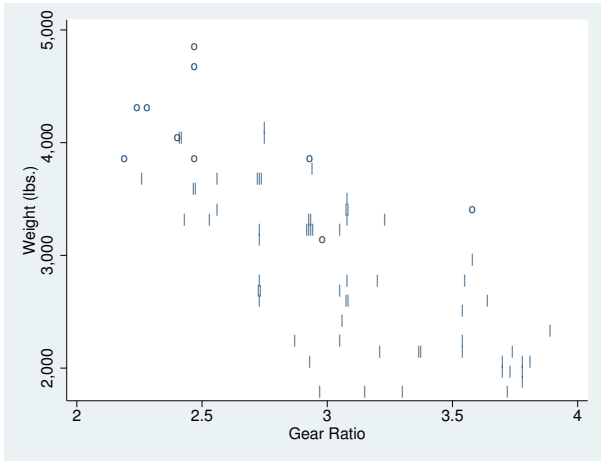


Example 5

`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/r13/auto
(1978 Automobile Data)
. generate byte hi_price = (price>10000) if price < .
. label define symbol 0 "|" 1 "o"
. label values hi_price symbol
```

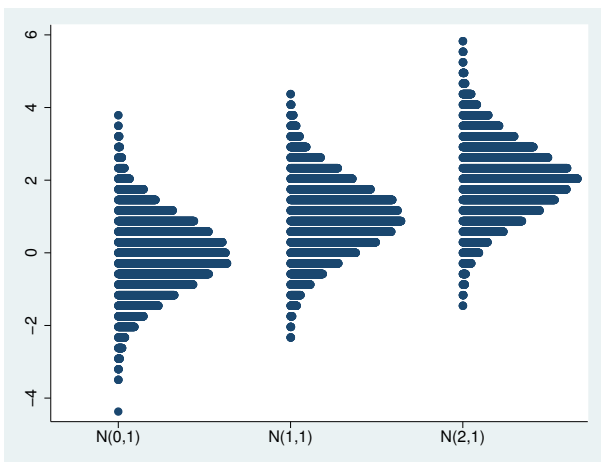
```
. dotplot weight, over(gear_ratio) m(none) mlabel(hi_price) mlabp(0) center
> xlabel(#5)
```



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
. gen norm0 = rnormal()
. gen norm1 = rnormal() + 1
. gen 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

Sasieni, P. D., and P. Royston. 1994. gr14: dotplot: Comparative scatterplots. *Stata Technical Bulletin* 19: 8–10. Reprinted in *Stata Technical Bulletin Reprints*, vol. 4, pp. 50–54. College Station, TX: Stata Press.

——. 1996. Dotplots. *Applied Statistics* 45: 219–234.