Title

stata.com

stem — Stem-and-leaf displays

Description Quick start Menu Syntax
Options Remarks and examples Stored results References
Also see

Description

`stem` displays stem-and-leaf plots.

Quick start

Stem-and-leaf plot for `v1`

```
stem v1
```

As above, with 2 lines per interval of 10

```
stem v1, lines(2)
```

Specify that each stem has a width of 5 (equivalent to above)

```
stem v1, width(5)
```

Stem-and-leaf plot for `v2` with an interval of 100

```
stem v2, digits(2)
```

As above, with 4 lines per interval of 100

```
stem v2, digits(2) lines(4)
```

Display `v2` rounded to the nearest hundred

```
stem v2, round(100)
```

Do not display empty stems

```
stem v2, prune
```

Menu

Statistics > Summaries, tables, and tests > Distributional plots and tests > Stem-and-leaf display

Syntax

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

options	Description
Main	
prune	do not print stems that have no leaves
round(#)	round data to this value; default is round(1)
truncate(#)	truncate data to this value
digits(#)	digits per leaf; default is digits(1)
lines(#)	number of stems per interval of $10^{\texttt{digits}}$
width(#)	stem width; equal to $10^{\texttt{digits}}/\texttt{width}$

by is allowed; see [D] **by**.

Options

Main

prune prevents printing any stems that have no leaves.

round(#) rounds the data to this value and displays the plot in these units. If round() is not specified, noninteger data will be rounded automatically.

truncate(#) truncates the data to this value and displays the plot in these units.

digits(#) sets the number of digits per leaf. The default is 1.

lines(#) sets the number of stems per every data interval of $10^{\texttt{digits}}$. The value of lines() must divide $10^{\texttt{digits}}$; that is, if digits(1) is specified, then lines() must divide 10. If digits(2) is specified, then lines() must divide 100, etc. Only one of lines() or width() may be specified. If neither is specified, an appropriate value will be set automatically.

width(#) sets the width of a stem. lines() is equal to $10^{\texttt{digits}}/\texttt{width}$, and this option is merely an alternative way of setting lines(). The value of width() must divide $10^{\texttt{digits}}$. Only one of width() or lines() may be specified. If neither is specified, an appropriate value will be set automatically.

Note: If lines() or width() is not specified, digits() may be decreased in some circumstances to make a better-looking plot. If lines() or width() is set, the user-specified value of digits() will not be altered.

Remarks and examples

stata.com

▷ Example 1

Stem-and-leaf displays are a compact way to present considerable information about a batch of data. For instance, using our automobile data (described in [U] **1.2.2 Example datasets**):

```
. use http://www.stata-press.com/data/r14/auto
(1978 Automobile Data)
. stem mpg
Stem-and-leaf plot for mpg (Mileage (mpg))
  1t | 22
  1f | 44444455
  1s | 66667777
  1. | 88888888899999999
  2* | 00011111
  2t | 22222333
  2f | 444455555
  2s | 666
  2. | 8889
  3* | 001
  3t |
  3f | 455
  3s |
  3. |
  4* | 1
```

The stem-and-leaf display provides a way to list our data. The expression to the left of the vertical bar is called the stem; the digits to the right are called the leaves. All the stems that begin with the same digit and the corresponding leaves, written beside each other, reconstruct an observation of the data. Thus, if we look at the four stems that begin with the digit 1 and their corresponding leaves, we see that we have two cars rated at 12 mpg, 6 cars at 14, 2 at 15, and so on. The car with the highest mileage rating in our data is rated at 41 mpg.

The above plot is a five-line plot with `lines()` equal to 5 (five lines per interval of 10) and `width()` equal to 2 (two leaves per stem).

Instead, we could specify `lines(2)`:

```
. stem mpg, lines(2)
Stem-and-leaf plot for mpg (Mileage (mpg))
  1* | 22444444
  1. | 556666777788888888899999999
  2* | 00011111222223334444
  2. | 555556668889
  3* | 0014
  3. | 55
  4* | 1
```

`stem mpg, width(5)` would produce the same plot as above.

The stem-and-leaf display provides a crude histogram of our data, one not so pretty as that produced by `histogram` (see [R] **histogram**), but one that is nonetheless informative.

◁

▷ Example 2

The miles per gallon rating fits easily into a stem-and-leaf display because, in our data, it has two digits. However, `stem` does not require two digits.

```
. stem price, lines(1) digits(3)
Stem-and-leaf plot for price (Price)
   3*** | 291,299,667,748,798,799,829,895,955,984,995
   4*** | 010,060,082,099,172,181,187,195,296,389,424,425,453,482,499, ... (26)
   5*** | 079,104,172,189,222,379,397,705,719,788,798,799,886,899
   6*** | 165,229,295,303,342,486,850
   7*** | 140,827
   8*** | 129,814
   9*** | 690,735
  10*** | 371,372
  11*** | 385,497,995
  12*** | 990
  13*** | 466,594
  14*** | 500
  15*** | 906
```

The `(26)` at the right of the second stem shows that there were 26 leaves on this stem—too many to display on one line.

We can make a more compact stem-and-leaf plot by rounding. To display `stem` in units of 100, we could type

```
. stem price, round(100)
Stem-and-leaf plot for price (Price)
price rounded to nearest multiple of 100
plot in units of 100
   3* | 33778889
   4* | 000011122223444555556677778 99
   5* | 11222447788899
   6* | 2233359
   7* | 18
   8* | 18
   9* | 77
  10* | 44
  11* | 45
  12* | 0
  13* | 056
  14* | 5
  15* | 9
```

`price`, in our data, has four or five digits. `stem` presented the display in terms of units of 100, so a car that cost $3,291 was treated for display purposes as $3,300.

◁

❑ Technical note

Stem-and-leaf diagrams have been used in Japanese railway timetables, as shown in Tufte (1990, 46–47).

❑

Stored results

`stem` stores the following in `r()`:

Scalars
- `r(width)` width of a stem
- `r(digits)` number of digits per leaf; default is 1

Macros
- `r(round)` number specified in `round()`
- `r(truncate)` number specified in `truncate()`

References

Cox, N. J. 2007. Speaking Stata: Turning over a new leaf. *Stata Journal* 7: 413–433.

Emerson, J. D., and D. C. Hoaglin. 1983. Stem-and-leaf displays. In *Understanding Robust and Exploratory Data Analysis*, ed. D. C. Hoaglin, C. F. Mosteller, and J. W. Tukey, 7–32. New York: Wiley.

Tufte, E. R. 1990. *Envisioning Information*. Cheshire, CT: Graphics Press.

Tukey, J. W. 1972. Some graphic and semigraphic displays. In *Statistical Papers in Honor of George W. Snedecor*, ed. T. A. Bancroft and S. A. Brown, 293–316. Ames, IA: Iowa State University Press.

——. 1977. *Exploratory Data Analysis*. Reading, MA: Addison–Wesley.

Also see

[R] **histogram** — Histograms for continuous and categorical variables

[R] **lv** — Letter-value displays