Stata tip 23: Regaining control over axis ranges

Nicholas J. G. Winter
Cornell University
nw53@cornell.edu

Beginning with version 8, Stata will often widen the range of a graph axis beyond the range of the data. Convincing Stata to narrow the range can be difficult unless you understand the cause of the problem.

Using the trusty auto dataset, consider the graph produced by this command:

```
.sysuse auto, clear
(1978 Automobile Data)
.twoway scatter mpg price
```

Although price ranges from $3,291 to $15,906 in the data, the lower end of the x-axis in this graph extends to zero, leaving blank space on the left-hand side. If we do not like this space, the solution would seem to lie with the `range()` suboption of the `xscale()` option. Thus we might expect Stata to range the x-axis using the minimum and maximum of the data, given the following command:

```
.twoway scatter mpg price, xscale(range(3291 15906))
```

However, this produces the same graph: the axis still includes zero. It seems that Stata is ignoring `range()`, although it does not do that when the range is increased, rather than decreased. Consider, for example, this command, which expands the x-axis to run from 0 through 30,000:

```
.twoway scatter mpg price, xscale(range(0 30000))
```
The issue is that the range displayed for an axis depends on the interaction between two sets of options (or their defaults): those that control the axis range explicitly, and those that label the axis. The range can be expanded either by explicitly specifying a longer axis (e.g., with `xscale(range(a b))`) or by labeling values outside the range of the data.

To determine the range of an axis, Stata begins with the minimum and maximum of the data. Then it will widen (but never narrow) the axis range as instructed by `range()`. Finally, it will widen the axis if necessary to accommodate any axis labels.

By default, `twoway` labels the axes with “about” five ticks, the equivalent of specifying `xlabel(#5)`. In this case, Stata chooses four labels, one of which is zero, and then expands the x-axis accordingly. In other words, if we specify `xscale()`—but do not specify `xlabel()`—we are in effect saying to Stata “and please use the default `xlabel()` for this graph”. This default may widen the axis range.

Therefore, to get a narrower x-axis, we must specify a narrower set of axis labels. For example, to label just the minimum and maximum, we could specify

```
.scatter mpg price, xlabel(minmax)
```

![Scatter plot of miles per gallon (mpg) vs. price, with axis labels at minimum and maximum values.](image)

Of course, we could specify any other set of points to label, for example,

```
.scatter mpg price, xlabel(5000[1000]15000)
```

This issue only appeared with Stata version 8. Prior versions defaulted to labeling the minimum and the maximum of the data only. This would be equivalent to including the options `xlabel(minmax)` and `ylabel(minmax)` in Stata 8 or later.