

Lean mainstream schemes for Stata 8 graphics

Svend Juul
Department of Epidemiology and Social Medicine,
University of Aarhus, Denmark

Abstract. The new Stata 8 graphics are powerful and flexible. Now, a few months after the first release, the graphics still have some shortcomings—both in design and in the manual documenting the program—but progress is being made. The graph layout used throughout the *Graphics Reference Manual* has led some users to underestimate the potential of the program. This paper presents two schemes for a lean layout, conforming to the mainstream in scientific publishing.

Keywords: gr0002, graphs, schemes

1 Introduction

During one week, I had three calls from people who had recently acquired Stata 8. Essentially, they asked the same questions. I was not surprised as I had asked the same questions myself:

- Can I get rid of the background tinting?
- Can I make my y -axis labels horizontal?
- How do I manage to make my markers black? (Unpredictable colors were produced.)
- How do I control the aspect ratio?

Stata 7 graphics were good for analysis but insufficient for producing publishable graphs. With the new Stata 8 graphics, I can do almost anything I did with SigmaPlot[®], the nicest graphics program I have used so far. Stata's graphics have the advantage of being integrated with a powerful statistical analysis program.

The above questions illustrate, however, a demand for easy access to a mainstream graph layout leaner than the layout presented in the *Graphics Reference Manual* and the variations offered by the current official schemes. Part of the reason for the questions was a difficult-to-access *Graphics Reference Manual*.

In the now-classic book, *The Elements of Graphing Data*, Cleveland (1994) advocates and demonstrates a lean graph layout. The information should stand out clearly, and superfluous elements and distracting decoration should be avoided. Except for a demonstration of the use of colors in some special instances, every graph in the book is kept black and white.

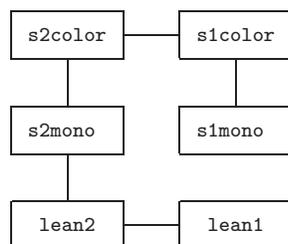
Cleveland's recommendations are generally accepted; they represent the mainstream in scientific graphs. Most scientific journals use a lean black and white look, and even the few journals that add a tint, e.g., *British Medical Journal* (BMJ), prefer to receive black-and-white originals.

This paper presents two new schemes, `lean1` and `lean2`, attempting to hit a lean mainstream graph style.

2 What is a scheme?

The default graph style is determined by a scheme, which is a set of instructions about line width, marker appearance, background appearance, etc. Thus, the default appearance is controlled by instructions outside the program, and you can easily change the overall appearance of a graph—if there is an appropriate scheme. The scheme does not restrict your possibilities; if you want a blue marker, you can have it by specifying the `mcolor(blue)` plot option, regardless of the scheme applied.

The “mother” of all current schemes is `s2color`; the other schemes are descendants of `s2color`. Including the schemes presented here, the family tree is



Generating your personal scheme may be complicated. The scheme syntax is different from the `graph` command syntax; documentation will probably be available by the time this is in print. To illustrate the syntax, the first few lines of `lean2.scheme` are shown below:

```

// lean2.scheme. Modifications to s2mono.scheme
// version 1.1 17apr2003
// Author Svend Juul

#include s2mono

// Background not tinted
color background white

// y-axis labels horizontal
anglestyle vertical_tick horizontal

// Axis titles: a bit larger and a bit more distance
gsize axis_title medium
margin axis_title small

```

The *Graphics Reference Manual* (p. 470) suggests that you edit a copy of `s2color.scheme` to create your personal scheme. However, `s2color.scheme` has some 1,100 lines, and you will get lost. The method illustrated above is more feasible.

3 The lean schemes

Here, I present two schemes, `lean1` and `lean2`. My main decisions when developing the schemes were as follows:

- No background tinting. Background tinting in by-graph subtitle fields and matrix plot title fields has been removed, as well.
- Horizontal labels at the y -axis. This is mainstream style, and it is more legible than the vertical orientation in Stata's current official schemes.
- All marker symbols are black and are not governed by the automatic allocation of shifting colors. Especially for hollow symbols and other thin-line symbols, this is an obvious advantage; a thin-line symbol in a light color is too faint.
- With all marker symbols black, hollow and thin line symbols can be given higher priority on the marker style list (p1–p15).
- Outlines for bars, boxplots, and pies are black, giving a more distinct appearance.
- With black bar outlines and color not an issue concerning marker appearance, I chose lighter default fill colors. Stata currently uses the same color style list (p1–p15) for markers and bar fills, although the visual effect of color is very different for small and large objects.
- The legend is moved to the lower right corner (4 o'clock) with no legend outline. This placement facilitates the reading of both the x -axis information and the legend itself.
- The axis title font size and the margin between the axis label and the title are slightly larger.
- Grid lines are thin black lines. `s2mono`'s 'woolly' grid lines at least have the problem that they do not reproduce well.
- Confidence intervals for fitted lines are thin black lines for the same reason as the grid lines.
- Matrix plots have small hollow circles as the default marker. This gives the best resolution when there are more than a few observations.

Figures 1–3 let you compare the impact of the suggested schemes `lean1` and `lean2` with the current official schemes `s1mono` and `s2mono`.

```
. sysuse auto, clear
(1978 Automobile Data)

. twoway
  (scatter mpg weight if foreign==0)
  (scatter mpg weight if foreign==1),
  title("Scatterplot. Scheme ...")
  legend(label(1 "Domestic") label(2 "Foreign"))
```

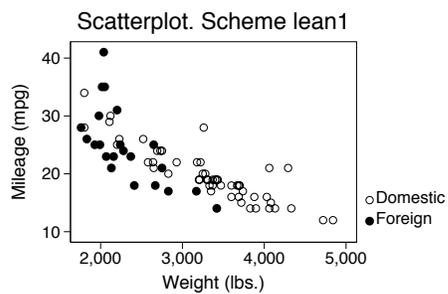
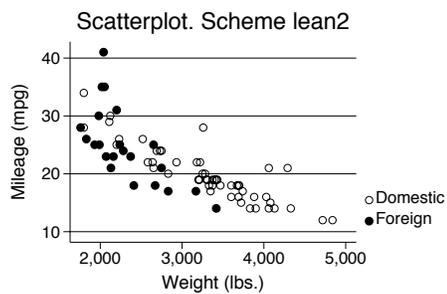
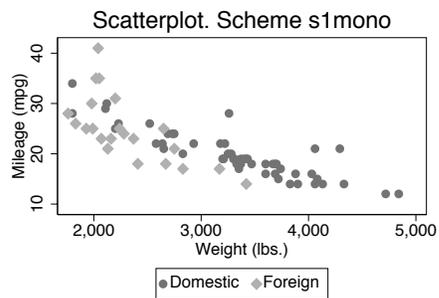
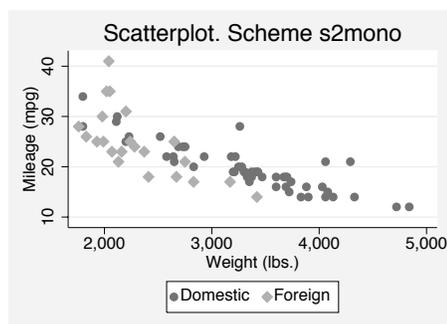


Figure 1: The same scatterplot displayed with four different schemes. The lean schemes have distinct markers, and the legend placement improves the readability of the legend and the x -axis title and labels. Horizontal y -axis labels read better than vertical labels. However, legend placement and size affect the plot aspect ratio.

```
. sysuse citytemp, clear
(City Temperature Data)
. graph bar (mean) tempjuly tempjan, over(region)
    title("Bar graph. Scheme ...")
    ytitle("Degrees Fahrenheit")
    legend( label(1 "July") label(2 "January"))
```

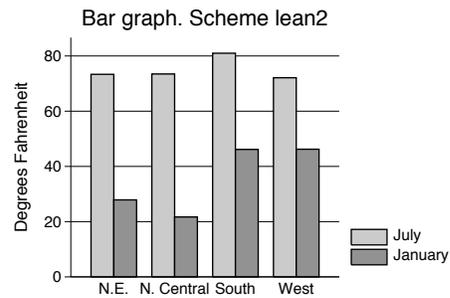
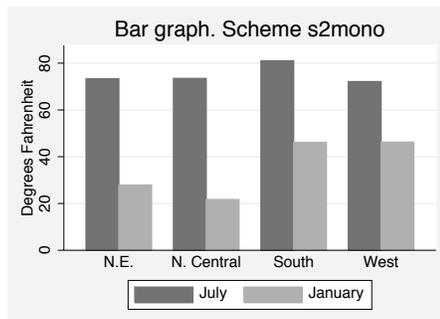


Figure 2: The same bar graph displayed with two different schemes. The appearance of bars is more distinct in the lean scheme display.

```
. sysuse auto, clear
(1978 Automobile Data)
. graph matrix price mpg weight,
    title("Matrix scatterplot. Scheme ...")
```

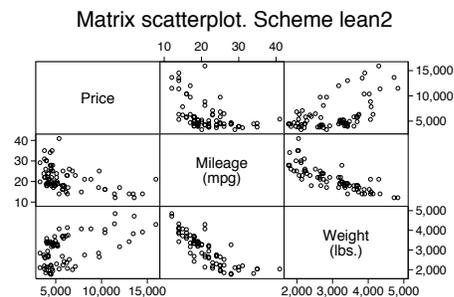
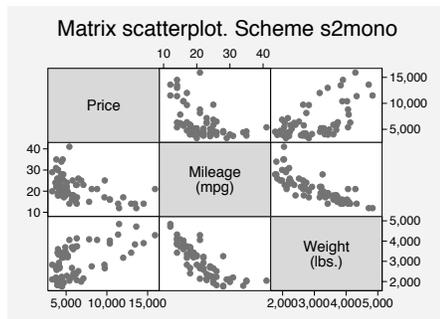


Figure 3: The same matrix scatterplot displayed with two different schemes. Small hollow markers are well suited for matrix scatterplots.

The lean schemes are in better accordance with the idea of “clear vision”, making the data stand out and avoiding superfluity (Cleveland 1994). I also believe that editors of scientific journals will accept a lean design more readily than the current Stata designs. At least, I like it much better myself.

4 The aspect ratio problem

The observant reader might wonder why the aspect ratio (the y/x ratio) varies between the official and the lean schemes; in Figure 1, the aspect ratio seems highest in the lean schemes. In Stata terminology, the opposite is the case; the official schemes have an aspect ratio of 0.73 versus 0.67 for the lean schemes. Why?

Figure 4 shows the anatomy of a Stata graph. In Stata terminology, the aspect ratio refers to the graph area, and you can determine the size of the graph area by the `xsize()` and `ysize()` options. However, the visually important area is the plot area, and the relevant aspect ratio is the ratio between axis lengths, as described by Cleveland (1994). I will refer to this as the plot aspect ratio.

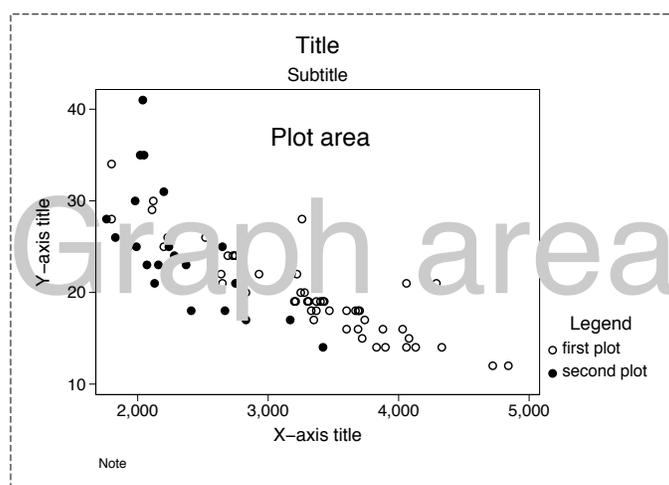


Figure 4: The anatomy of a graph. The user can control the graph area size directly and the plot area size and aspect ratio only indirectly.

Unfortunately, Stata does not allow you to determine axis lengths directly. Instead, axis lengths are determined indirectly by letting the plot area have what happens to be left over, once the title, legend, etc., have taken their share of the graph area. The reason why the lean schemes have the highest plot aspect ratio in Figure 1 is the placement of the legend. To compensate for this, I increased the default x -axis size, but at a cost: In graphs without legends, as in Figure 3, the plot aspect ratio is affected in an undesirable way.

The opportunity to determine axis lengths directly would be a major empowerment of the users. As it is now, the plot aspect ratio changes if I make a legend text shorter or longer. In the matrix scatterplot example, Figure 3, I would prefer square plot areas but that can only be obtained by trial and error.

5 Conclusion

While working to understand and use the new Stata 8 graphics, I found the graph layouts presented in the manual and made available through the current official schemes less than optimal. I especially missed a lean layout, conforming to mainstream scientific journal style. I also strongly missed the opportunity to determine axis lengths, and hence, the plot aspect ratio, directly. The first edition of the *Graphics Reference Manual* turned out to be difficult to use due to a confusing organization, even to an experienced user.

However, the new Stata 8 graphics is a very flexible graphing system with great potential, and the new dialog system makes it easy for any user to create graphs. The scheme principle allows Stata and advanced users to make a variety of layouts available to users, as this paper demonstrates.

I have been told that the issues of allocation of colors to markers and bar fills and the opportunity to give the user direct control over axis lengths are being seriously considered.

6 Acknowledgment

Vince Wiggins of Stata Corporation has been extremely helpful giving suggestions, answering numerous questions, and correcting mistakes.

7 References

Cleveland, W. S. 1994. *The Elements of Graphing Data*. Summit, NJ: Hobart Press.

About the Author

Svend Juul is an associate professor in the Department of Epidemiology and Social Medicine, University of Aarhus, Denmark. He teaches epidemiology to medical students and computer skills to Ph.D. students. His main research area is epidemiological studies of fertility problems.