The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the web page

http://www.stata-journal.com

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index®
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch®)

**Copyright Statement:** The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible web sites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata and Mata are registered trademarks of StataCorp LP.
Stata tip 76: Separating seasonal time series

Nicholas J. Cox
Department of Geography
Durham University
Durham, UK
n.j.cox@durham.ac.uk

Many researchers in various sciences deal with seasonally varying time series. The part rhythmic, part random character of much seasonal variation poses several graphical challenges for them. People usually want to see both the broad pattern and the fine structure of trends, seasonality, and any other components of variation. The very common practice of using just one plot versus date typically yields a saw-tooth or roller-coaster pattern as the seasons repeat. That method is often good for showing broad trends, but not so good for showing the details of seasonality. I reviewed several alternative graphical methods in a Speaking Stata column (Cox 2006). Here is yet another method, which is widely used in economics. Examples of this method can be found in Hylleberg (1986, 1992), Ghysels and Osborn (2001), and Franses and Paap (2004).

The main idea is remarkably simple: plot separate traces for each part of the year. Thus, for each series, there would be 2 traces for half-yearly data, 4 traces for quarterly data, 12 traces for monthly data, and so on. The idea seems unlikely to work well for finer subdivisions of the year, because there would be too many traces to compare. However, quarterly and monthly series in particular are so common in many fields that the idea deserves some exploration.

One of the examples in Franses and Paap (2004) concerns variations in an index of food and tobacco production for the United States for 1947–2000. I downloaded the data from http://people.few.eur.nl/paap/pbook.htm (this URL evidently supersedes those specified by Franses and Paap [2004, 12]) and named it ftp. For what follows, year and quarter variables are required, as well as a variable holding quarterly dates.

```
    . egen year = seq(), from(1947) to(2000) block(4)
    . egen quarter = seq(), to(4)
    . gen date = yq(year, quarter)
    . format date %tq
    . tsset date
    . gen growth = D1.ftp/ftp
```

Although a line plot is clearly possible, a scatterplot with marker labels is often worth trying first (figure 1). See an earlier tip by Cox (2005) for more examples.

© 2009 StataCorp LP
. scatter growth year, ms(none) mla(quarter) mlabpos(0)

Figure 1. Year-on-year growth by quarter for food and tobacco production in the United States: separate series

Immediately, we see some intriguing features in the data. There seems to be a discontinuity in the early 1960s, which may reflect some change in the basis of calculating the index, rather than a structural shift in the economy or the climate. Note also that the style and the magnitude of seasonality change: look in detail at traces for quarters 1 and 4. No legend is needed for the graph, because the marker labels are self-explanatory. Compare this graph with the corresponding line plot given by Franses and Paap (2004, 15).

In contrast, only some of the same features are evident in more standard graphs. The traditional all-in-one line plot (figure 2) puts seasonality in context but is useless for studying detailed changes in its nature.
Figure 2. Quarterly food and tobacco production in the United States

The apparent discontinuity in the early 1960s is, however, clear in a plot of growth rate versus date (figure 3).

Figure 3. Year-on-year growth by quarter for food and tobacco production in the United States: combined series
An example with monthly data will push harder at the limits of this device. Grubb and Mason (2001) examined monthly data on air passengers in the United Kingdom for 1947–1999. The data can be found at http://people.bath.ac.uk/mascc/Grubb.TS; also see Chatfield (2004, 289–290). We will look at seasonality as expressed in monthly shares of annual totals (figure 4). The graph clearly shows how seasonality is steadily becoming more subdued.

```
. egen total = total(passengers), by(year)
. gen percent = 100 * passengers / total
. gen symbol = substr("123456789OND", month, 1)
. scatter percent year, ms(none) mla(symbol) mlabpos(0) mlabsize(*.8) xtitle(""
> ytitle(\% in each month) yla(5(5)15)
```

Figure 4. Monthly shares of UK air passengers, 1947–1999 (digits 1–9 indicate January–September; O, N, and D indicate October–December)

Because some users will undoubtedly want line plots, how is that to be done? The `separate` command is useful here: see Cox (2005), [p] separate, or the online help. Once we have separate variables, they can be used with the `line` command (figure 5).
You may think that the graph needs more work on the line patterns (and thus the legend), although perhaps now the scatterplot with marker labels seems a better possibility.

If graphs with 12 monthly traces seem too busy, one trick worth exploring is subdividing the year into two, three, or four parts and using separate panels in a `by()` option. Then each panel would have only six, four, or three traces.

**References**


