

Using Grid Graphics to produce linked micromap plots of large financial datasets

Gordon Blunt
CACI Ltd
gblunt@caci.co.uk

This paper describes how linked micromap (LM) plots, drawn using the flexibility of Grid Graphics' multiple coordinate systems, can be used to develop graphical summaries of large financial services datasets. A typical LM plot can contain sixty, seventy or more individual plot elements, which need to be arranged carefully on a page, with particular regard to alignment of the elements. The Grid Graphics package allows precise control over layouts, and so the exact placement of each plot element within the layout. Further, use of the different coordinate systems in Grid allows the flexible specification of the different plot elements – micromaps, labels, statistical graphics and text – with ease.

This flexibility comes at a price, however, in that Grid Graphics provide only low level graphical functions rather than high level functions that can produce complete plots. However, once these low level functions are understood, complex plots can be constructed with relatively little effort. This is well illustrated by the Lattice package, which uses Grid to render its plots.

LM plotting can be thought of as a method that combines the techniques of exploratory data analysis and statistical graphics, but which maintains the spatial context of the data. Many financial services datasets have pronounced geographic differences for which traditional statistical summaries – in isolation – are inadequate; on the other hand, choropleth maps on their own cannot convey the range of data to be presented. Linked micromaps are one solution to this problem, and have the following characteristics.

- They display several sequences of panels that are linked by position. These panels contain names, maps and graphical summaries of the data. The latter can be any form of statistical summary; for example histograms, time series or dot plots, among many others.
- Data are sorted by the variable(s) of interest, to improve perception between sequences of panels.
- The dataset is partitioned into these relatively small panels to allow attention to be focused on small areas at a time.
- They draw on principles from a number of disparate disciplines, including statistics (in particular exploratory data analysis), cartography and psychology to produce convenient, yet revealing, summaries of large data sets which preserve most of the important elements of the data.

Another advantage of LM plots is that they provide insights into the structure of a dataset at many levels. This is a consequence of the range of statistical graphics it is possible to use, which are able show extremes as well as measures of central tendency. LM plots are flexible enough to be used for this further analysis, and can thus be thought of as a flexible graphical data mining tool. Thus, they allow data sets with millions of observations to be summarised in a convenient – and revealing – graphical way.

LM plots have been used mainly in the fields of epidemiology, ecology and official statistics but have not, as yet, been applied to financial datasets.