

# Using linked micromap plots to produce graphical summaries of large financial datasets

Gordon Blunt  
CACI Ltd  
[gblunt@caci.co.uk](mailto:gblunt@caci.co.uk)

This paper describes how linked micromap (LM) plots can be used to develop graphical summaries of large financial services datasets. In this context, 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 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; for example, using them for high level overviews can guide further drill down. This is a consequence of the range of statistical graphics that can be used, which can 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.