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

Gordon Blunt, CACI Ltd
gblunt@caci.co.uk
useR! 2006, Vienna

Agenda

Linked micromaps (LM plots)
- what they are
- background from statistical graphics

Use of Grid Graphics
- enable one layout template to plot many maps
- multiple coordinate systems cope easily with different data sets
- Grid Graphics' *grid.layout()*
  - more flexible than traditional graphics' *par()*, *split.screen()* or *layout()*

Conclusions & further work

N.B. This talk is not an overview of Grid Graphics
- much better done by Paul Murrell
  R Graphics (2005), Chapman & Hall

Linked micromaps

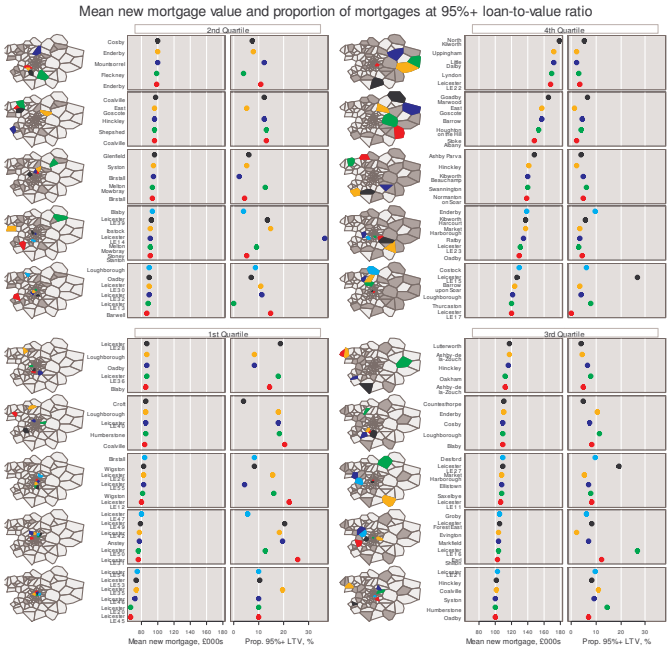
Method for plotting statistical data
- while maintaining the spatial context

Presented by Carr et al. (1998)
- following earlier work in related areas
- Carr and Pierson (1996)
  - others shown in references

Dan Carr, George Mason University
- seems to have been involved in most published LM plotting work
- used to display data sets from
  - official statistics
  - epidemiology
  - ecology

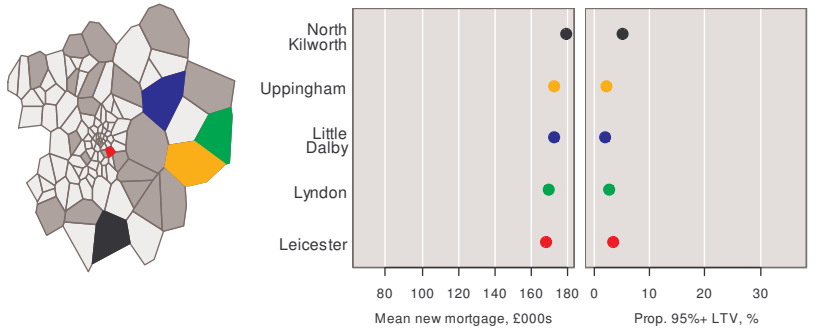
Linked micromaps

Elements from
- Cleveland (1985, 1993)
  - dot plots
    - can easily extend to other types of statistical graphic
  - ordering
  - good statistical graphics and layout
    - especially common scales, position
- Tufte (1990, 1997, 2001)
  - small multiples
- Kosslyn (1994)
  - psychology of graphical perception
    - colour
    - number of elements
    - linking these unambiguously



LM plots do not work well on screen – the resolution is too coarse, hard copy prints are usually needed.

Top right part of plot



Axes added to plot for this example, they are usually shown only on bottom most plots.

Key features of linked micromaps

Data sorted
- but the map caricatures don't show data

Each column of plots to same scale

Simplified boundaries
- Dirichlet tesselation for internal boundaries
- simplify external boundaries too

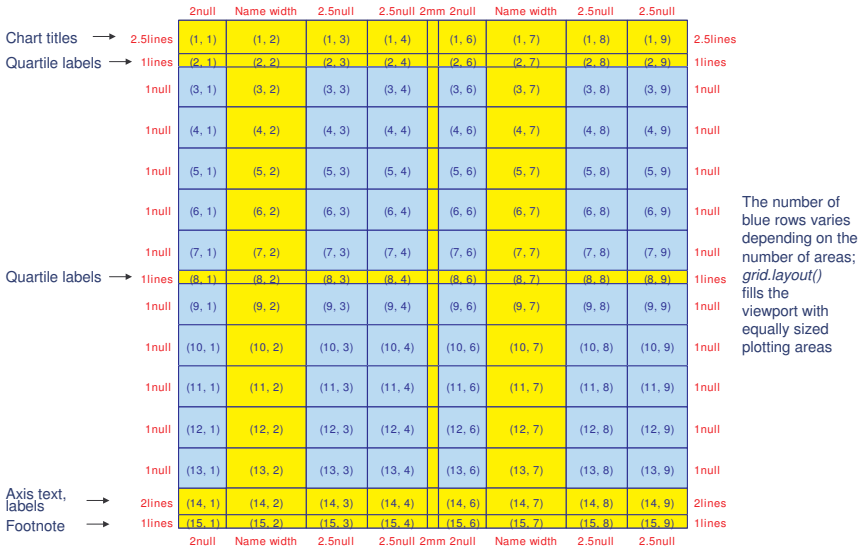
Legend

Small multiples

Colour

Distort the map if necessary
- it spatially indexes the data, it is not accurate geography
- e.g. for UK, we typically squash it north – south, move some of the islands nearer the mainland

grid.layout(), *viewports* and Grid's *coordinate systems* essential



The number of blue rows varies depending on the number of areas; *grid.layout()* fills the viewport with equally sized plotting areas

Adapted from Paul Murrell's 'grid.show.layout()' function.

Coordinate systems

Many coordinate systems are available in Grid

We used the following to produce LM plots

- *native* (locations relative to *x* and *y* scales of the current viewport)
- *npc* (normalised parent coordinates)
- *strwidth* (string width – for the area names in columns 2 and 7)
- *lines* (locations specified relative to current font size)
- *mm* (millimetres)
- *char* (locations specified relative to multiples of current font size)

Also, and specific to grid layouts

- *null* (allows plot elements to be equally sized – by Grid – to fit around the areas where the size needs to be fixed)

There are other coordinate systems too!

Other Grid features

Not essential, but useful in our implementation

A plot drawn in a Grid *viewport* fills the viewport

- the axes are outside the viewport by default
- unless we use *plotViewport()*
  - sets up a central plotting region within a viewport
- LM plots require only one set of axes at the bottom of each column of plots

We've found it easier to explain LM plots to clients if each plot is separated – slightly – from every other element

- have used the *width* and *height* arguments to shrink each plot element by a consistent amount
- in conjunction with *npc* and *native* coordinates Grid graphics does this regardless of x and y limits

Can place text precisely in viewports

- or across viewports, if the *clip* argument is *off*

Conclusions

LM plots add insight to our discussions with our clients

We have adapted Carr et al.'s (1998) methods for displaying financial datasets; the following are crucial

- variables are shown by distance along a linear scale
- they are measured from a common axis
- the data need to be sorted
- number of points in each plot element needs to be relatively small
- the plot elements are arranged in small multiples
- white gridlines should be used to add guides but not clutter

If the quartile shadings indicate little, or no, spatial pattern:

- we might conclude that LM plots are unnecessary
- so the data could probably be better represented by a simpler method

Further work

Different types of statistical graphic

have also used (slightly modified) box percentile plots

see Frank Harrell's *bpplot()*, based on Esty and Banfield (2003)

used by Dan Carr too …

who has also used

time series

rates of change

however, the dot chart seems ideal for LM plots

easily understood by my (mostly) non-statistical clients

Our functions work for up to 144 areas

is it possible to plot more?

I suspect not on A3 paper

this shouldn't be a limitation