

TOPCAT Corner Plot

Mark Taylor, University of Bristol, UK



University of
BRISTOL

m.b.taylor@bristol.ac.uk



Introduction

TOPCAT is a desktop GUI tool for working with tabular data such as source catalogues. Among other capabilities it provides a rich set of visualisation windows suitable for interactive exploration of large (many rows) and high-dimensional (many columns) datasets.

The new release v4.9 introduces a **Corner Plot** window which can present a grid of scatter-plot-like and histogram-like plots for all pair and single combinations of a supplied list of coordinates. This provides a simultaneous view of all the 2-d and 1-d projections of a high-dimensional dataset, which can be a valuable aid for identifying interesting features in complex data. This plot is also available as the command `plot2corner` from the corresponding STILTS release v3.4-9.

This type of visualisation is not novel; it has been used since the 1980s under the names “Scatter Plot Matrix”, “SPLOM”, “Pairs Plot” and “Corner Plot”, and graphics packages producing them exist for instance in Python, R and IDL. Its adoption here benefits from the performance, interactive features and ease of use provided by the existing application GUI, and gives TOPCAT users a new tool for exploring high-dimensional data.

Features

The Corner Plot shares its features with the other TOPCAT visualisation windows:

Ease of use: coordinates can be added, deleted and reordered with a few mouse clicks

Interactivity: each panel can be panned and zoomed in one or two dimensions easily using the mouse, and the other panels in the same row/column adjust to keep the axes aligned

Linked Views: graphical selections and point highlighting on one panel are immediately visible in the other panels (and in other windows)

Overplotting: multiple datasets, selections and plot types can appear in each panel

Calculations: coordinates can be existing columns or algebraic expressions based on them


Scalability: multipanel displays for millions of rows are quite responsive even on modest hardware

Flexibility: many plotting and shading options are available alongside simple scatter plots and histograms: transparency, weighted density maps, contours, KDEs, text labels, ...


Configurability: interactive controls can specify lower/upper/full panel matrix, log/flipped axes, axis labels, grid line drawing, colour map adjustments, ...

The interactive linked plots make it very easy to identify regions, populations or points that are apparent in one coordinate plane and see where they appear in the other planes.

Configuring a Plot

To set up a Corner plot in TOPCAT, click the **Corner Plot button**  in the main toolbar, and choose a table. Then select the columns or type in expressions you want to plot against each other, and adjust the results by adding or removing more quantities, or dragging coordinates up and down the list. You can pan and zoom each panel using the mouse to focus on the regions you're interested in.

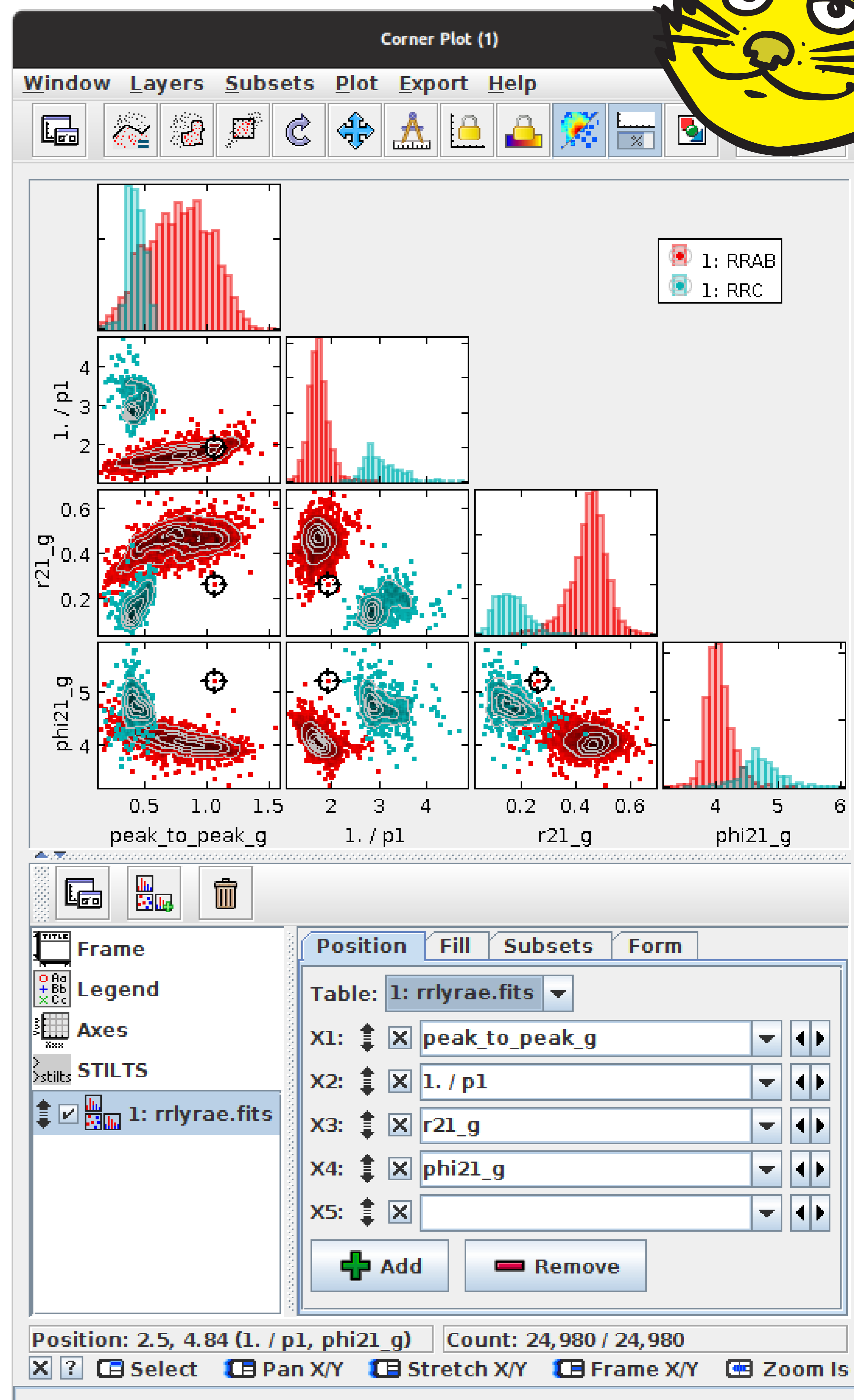
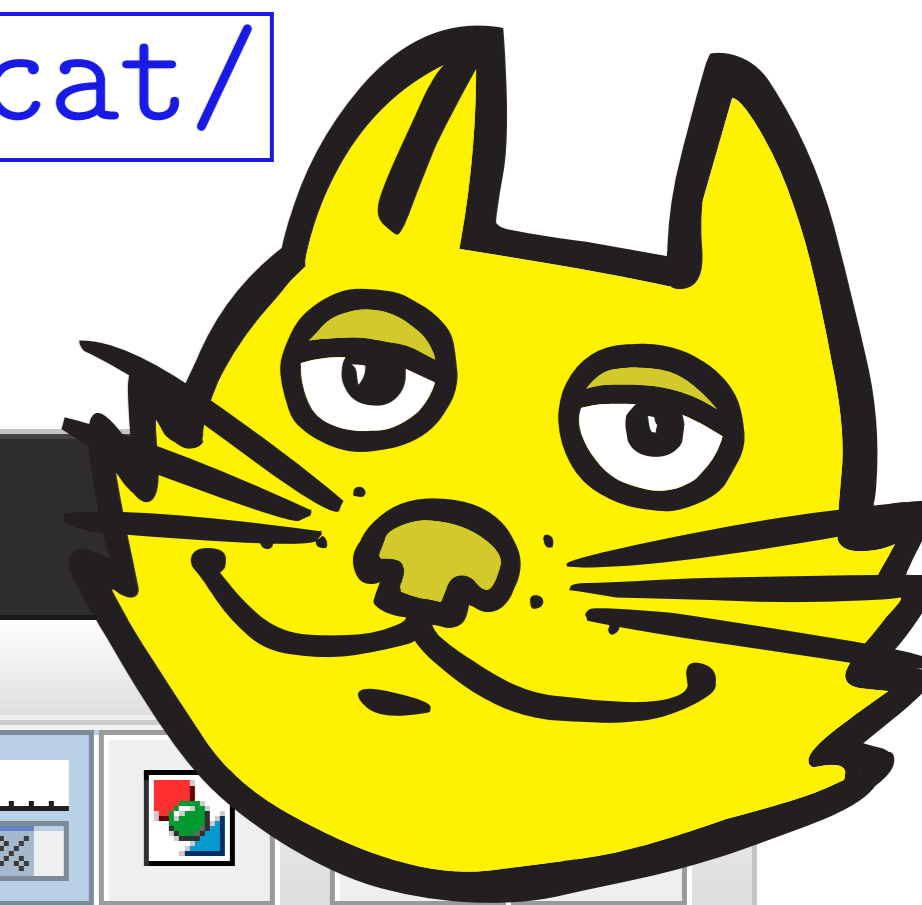
There are additional controls in the **Fill** tab to help select coordinates, including an option to use all pair differences/ratios from a selection of columns (e.g. all pairwise colours from a list of magnitudes/fluxes).

The **Float Controls button**  detaches the control panel from the plot panel, so can give more space for the plot.

Implementation

Assembling multiple plot panels in this way is not conceptually difficult, and the TOPCAT codebase already provided most of the data handling infrastructure required, as well as some support for multi-panel plots. However, implementing a plot type with multiple panels derived from a single set of input coordinates broke some of the assumptions in the existing plotting framework. This meant that significant code restructuring was required, especially for the graphical and command-line user interfaces, so implementation was quite time-consuming.

<http://www.starlink.ac.uk/topcat/>



Corner plot window showing variability relationships between two different classes of RR-Lyrae stars; data from Gaia DR1.

Future Work

The new corner plot is working quite well, but not perfectly. In particular the axis labelling developed for other plot types can become overcrowded and imperfectly aligned. Some performance improvements could be made by concurrent plotting of multiple panels and eliminating some duplicated work. Usage in the field may come up with other bugs, complaints or suggestions. If this plot type proves popular, improvements may be made in future. Other possibilities for multi-panel plots may also be investigated.

Your feedback is welcome. Ask me for a demo!