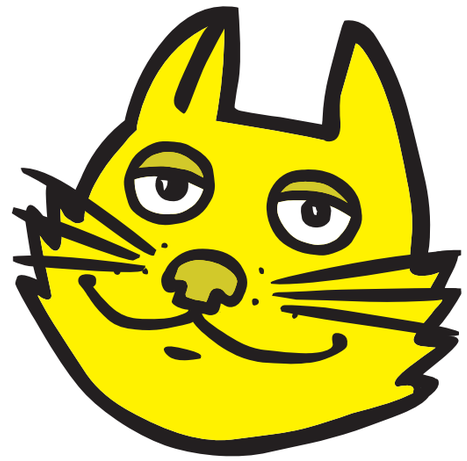
Introduction to TOPCAT

Mark Taylor (University of Bristol)

Shristi *Astronomy from Archival Data*
Online seminar

22 August 2020



$Id: topcat.tex,v 1.5 2020/08/21 13:31:00 mbt Exp $

Outline

TOPCAT

- What is it?
- What can it do?

Demo

- Pleiades in Gaia DR2 and 2MASS
- Hyades in 3D phase space using TAP

Q+A

Hands on

- Gaia DR2

Overview

TOPCAT = Tool for OPerations on Catalogues And Tables

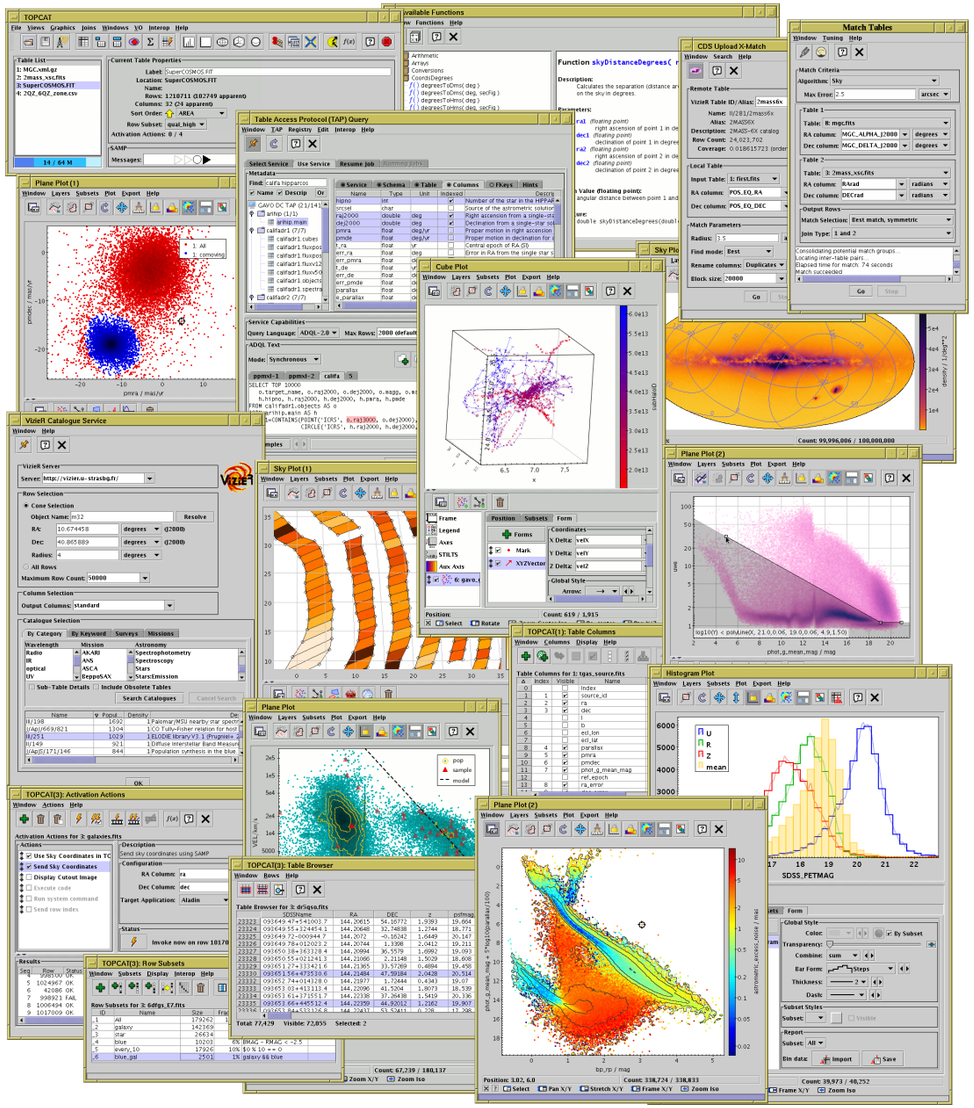
"Does what you want with tables"

Suitable for:

- Interactive exploration
- Quick look at unfamilar data
- In-depth analysis

Overall aim:

- Makes table manipulation easy,
  so users can concentrate on doing science



Characteristics

Aims:

- User-friendly
  - ▷ Easy to install and run (pure Java — one download file, no library issues)
  - ▷ Easy to get started
  - ▷ Simple things fairly obvious
  - ▷ Complicated things at least well-documented

    ... this does get harder as more functionality is added
- High Performance
  - ▷ Most things are fast
  - ▷ Handles fairly large tables: millions of rows, hundreds of columns easily (can be much more)
  - ▷ ... even on modest hardware
- Does the things that astronomers need
  - ▷ Development is led by community input (mailing list, personal emails, tutorials, feature requests, bug reports...)
  - ▷ Feedback always welcome!

Capabilities

It can do:

- Read/write tables in various formats (FITS, VOTable, CSV, ...)
- View data
- View metadata
- Calculations and simple statistics (expression language)
- Visualisation (many options, interactive)
- Make/combine/display row selections in various ways (linked views)
- Crossmatching (many options)
- Access external data services (VO and others)
- Talk to other astro tools (SAMP)
- Trigger some event when a row is selected

It can't do:

- Images, spectra (it's only for tables)
- Scripting (but see STILTS)
- *Very* large tables (but see STILTS)
- Every ASCII-like format known to man
- Write your papers for you

Table Data and Metadata

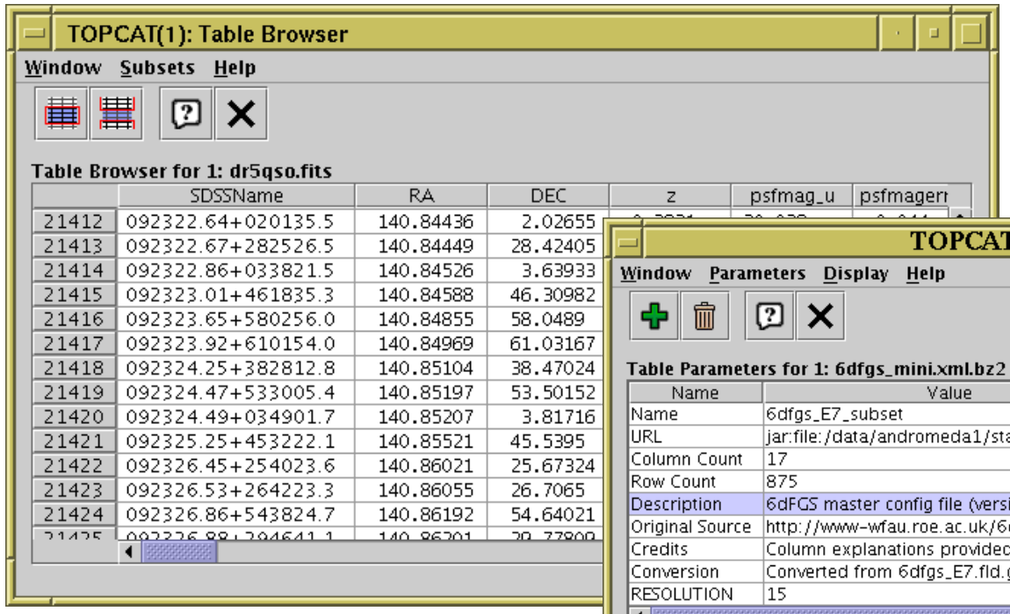


Table data view

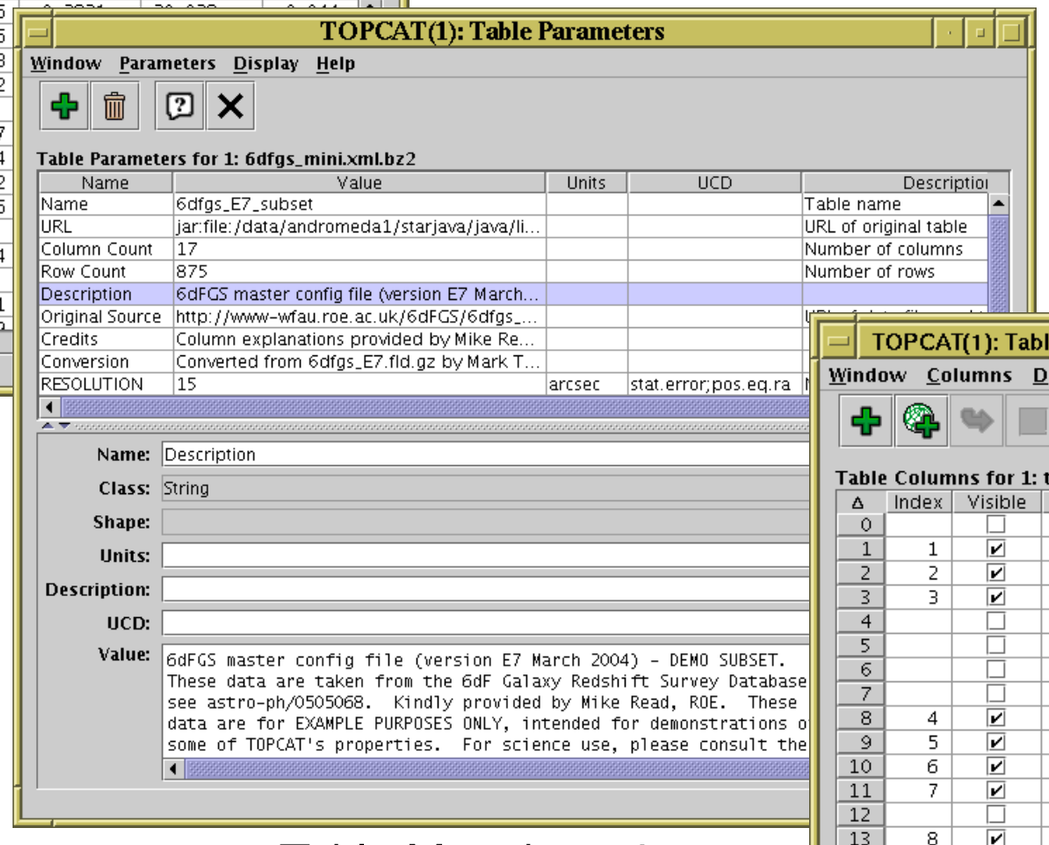
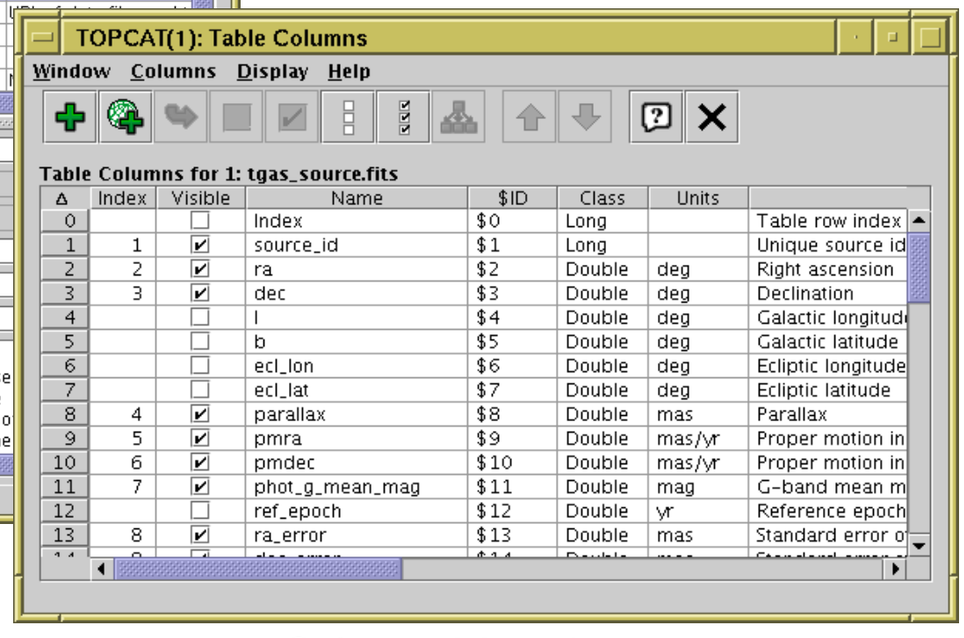


Table Metadata view



Column Metadata view

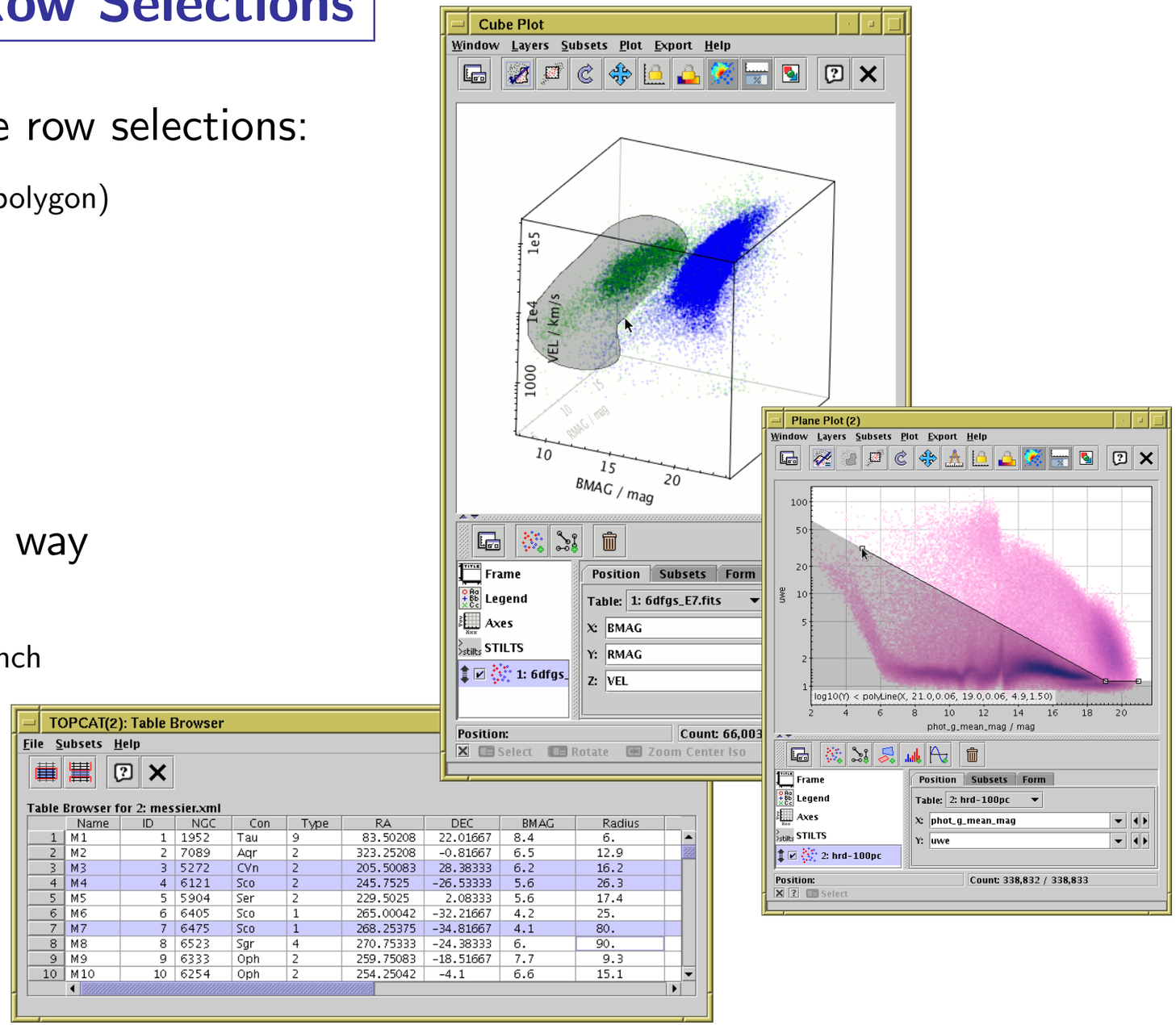
Row Selections

Different ways to make single or multiple row selections:

- Select points graphically from a plot (freehand or polygon)
- Select rows from the table view
- Use an algebraic expression
- Combine existing subsets
- Receive from an external application (SAMP)

Linked views mean a selection made one way is visible in other ways

- Perform crossmatch only on items in red giant branch
- Where on the sky is this colour cut?
- Spot outliers
- Identify objects on ds9 image display



TOPCAT(2): Table Browser

File Subsets Help

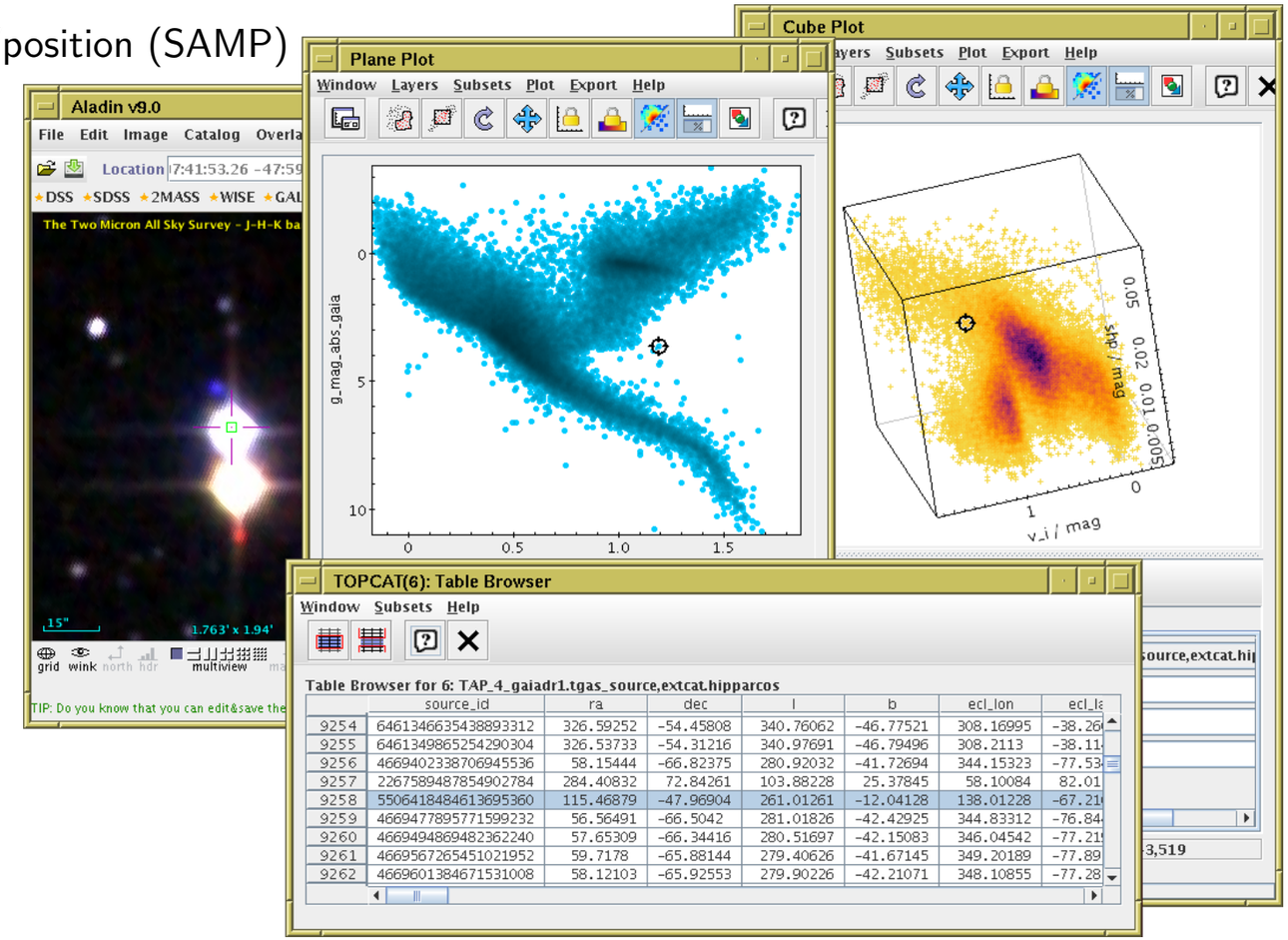
Table Browser for 2: messier.xml

	Name	ID	NGC	Con	Type	RA	DEC	BMAG	Radius
1	M1	1	1952	Tau	9	83.50208	22.01667	8.4	6.
2	M2	2	7089	Aqr	2	323.25208	-0.81667	6.5	12.9
3	M3	3	5272	CVn	2	205.50083	28.38333	6.2	16.2
4	M4	4	6121	Sco	2	245.7525	-26.53333	5.6	26.3
5	M5	5	5904	Ser	2	229.5025	2.08333	5.6	17.4
6	M6	6	6405	Sco	1	265.00042	-32.21667	4.2	25.
7	M7	7	6475	Sco	1	268.25375	-34.81667	4.1	80.
8	M8	8	6523	Sgr	4	270.75333	-24.38333	6.	90.
9	M9	9	6333	Oph	2	259.75083	-18.51667	7.7	9.3
10	M10	10	6254	Oph	2	254.25042	-4.1	6.6	15.1

Row Highlighting

Row selection is coordinated between linked views:

- Click on row in table browser or plot
- Same row is highlighted in other plots & table browser
- Can configure external tools to highlight same object/position (SAMP)



Calculations

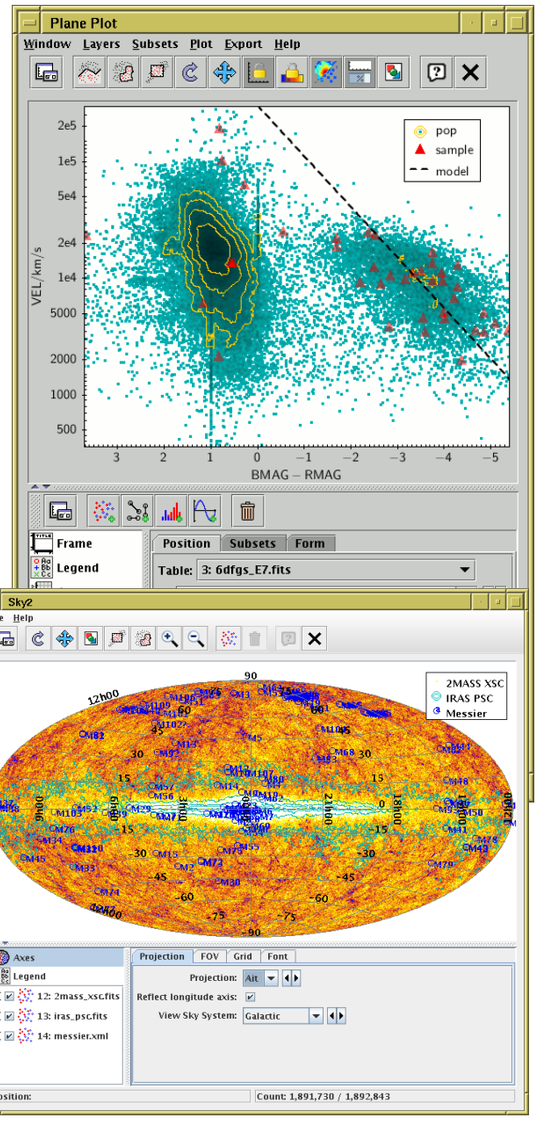
- Expression language used for creating columns, defining selections, making plots etc:
  - Straightforward arithmetic syntax (C-like)
  - Use column names as variables
  - Standard arithmetic operators (`+`, `-`, `/`, `*`)
  - Conditional expressions (`q?a:b`)
  - Standard mathematical functions (`abs`, `max`, `round`, `sin`, `cos`, `pow`, ...)
  - Sky coordinates (degrees, sexagesimal, sky distances)
  - Astrometry (epoch propagation with/without errors, ...)
  - Cosmological distances (redshift, luminosity dist, lookback time, ...)
  - Fluxes (Johnson AB Magnitudes, Jansky)
  - Time conversions (ISO8601, MJD, Julian, Besselian)
  - ... and more (and it's extensible)
- Examples:
  - `mag_u - mag_g`
  - `janskyToAb(flux)`
  - `skyDistanceDegrees(ra, dec, 14.1, -72.9) < 1.2`

Visualisation

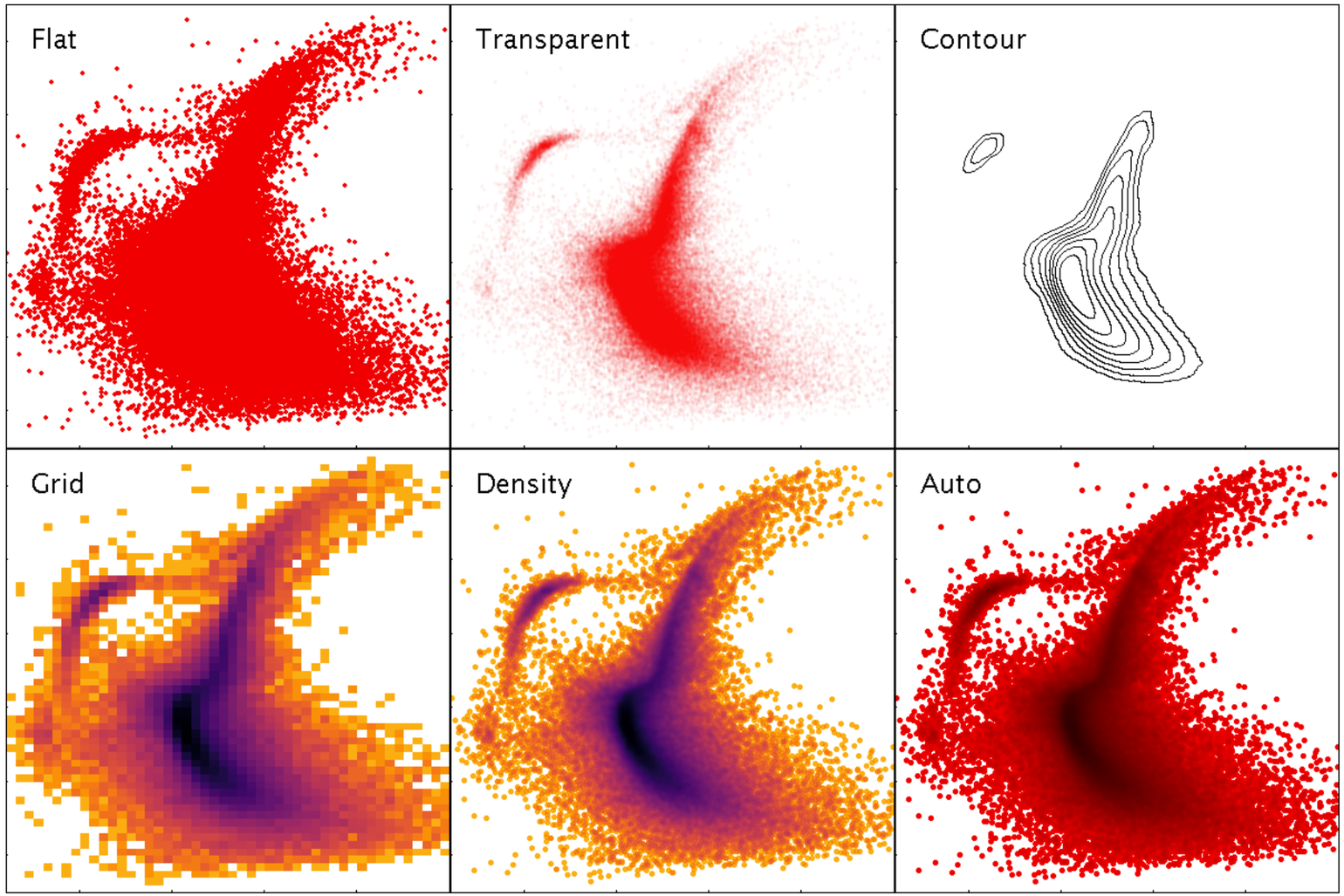
Very good for interactive exploration of large (or small) datasets:

- Many plot types!
  - 2d/3d scatter plots, histograms, HEALPix, density maps, error bars/ellipses, vectors, lines, quantiles, text labels, contours, KDEs, analytic functions, spectrograms, ...
- Many options!
  - Colour, colour maps, shading mode, weighting, marker shape/size line style, sky projection, sky system, coordinate grid, axis labelling, smoothing, binning, ...
- Highly responsive
  - Interactive changes to options update plot immediately
- Special attention to large data sets
  - Plot arbitrarily large datasets in fixed memory
  - Represent very dense plots in comprehensible ways
  - Many options for high-dimensional visualisation
- Publication-quality output?
  - Export to PDF, EPS, PNG, SVG (coming soon), ...
  - Optional LaTeX annotation
  - Script output (STILTS) for reproducibility
  - ... but not quite as good as Matplotlib/IDL/R

Visualisation: Plot Types



Visualisation: Dense plots



Different options for shading scatter-plot data.

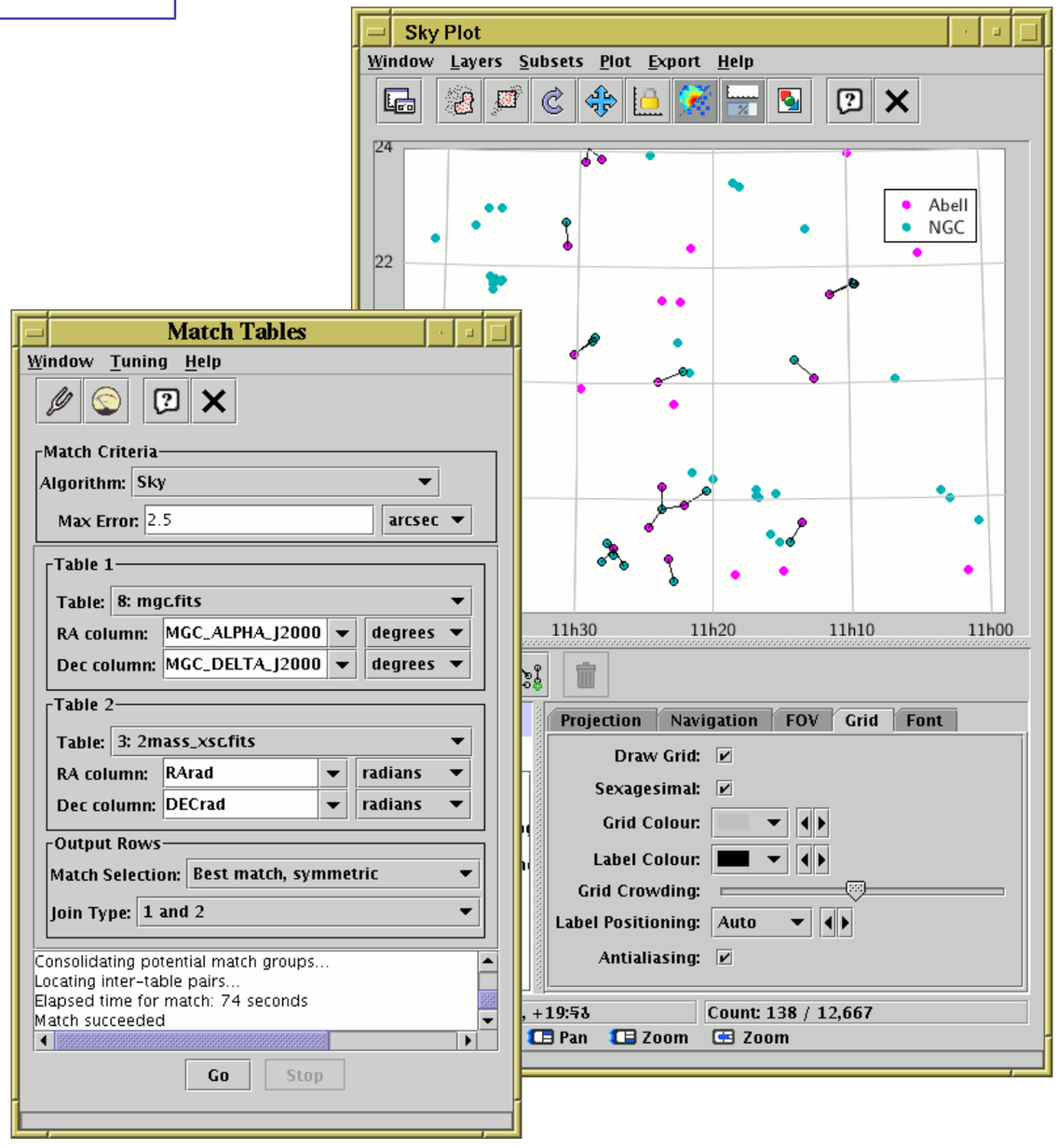
Crossmatching

Internal

- Both/all files loaded into TOPCAT
- Works well up to ∼1 million rows each
- Pretty fast (≤couple of minutes)
- Very flexible (sky, Cartesian, exact, 3D, ellipses, errors, combinations...)

External

- One or both tables too big to download
- Several options, with different pros and cons:
  - ▷ CDS X-Match
    (any VizieR table, sky match, fast, easy)
  - ▷ Multiple cone search
    (many tables available, sky match, slow)
  - ▷ TAP
    (few tables available, flexible, tricky)



Virtual Observatory

What is the Virtual Observatory (VO)?

- *"All astro archives in your computer"*
- A set of protocols that allows software clients to talk to external data services in a uniform way
- In most cases you (the software user) don't need to understand the details, but it's under the hood making data access work

External data access from TOPCAT:

- Cone Search: positional query of remote catalogue
- Table Access Protocol (TAP): SQL-like queries against remote databases
- Simple Image Access/Simple Spectral Access: positional query of image/spectrum archives
- CDS services: Simbad, VizieR cone/all-sky, X-Match, Hips2fits
- Registry: service discovery
- SAMP: communication with other desktop/web applications

STILTS

STIL Tool Set (STIL = Starlink Tables Infrastructure Library)

- Has pretty much the same capabilities as TOPCAT
- but works from the command line (also JyStilts from Jython)

TOPCAT 	**STILTS** 
GUI	Command line
Interactive	Scriptable
Easy to use	Reproducible
Good for data exploration	Good for batch/programmed use
Exploratory phase	Production phase
few$\times 10^6$ rows	Unlimited size (for most things)

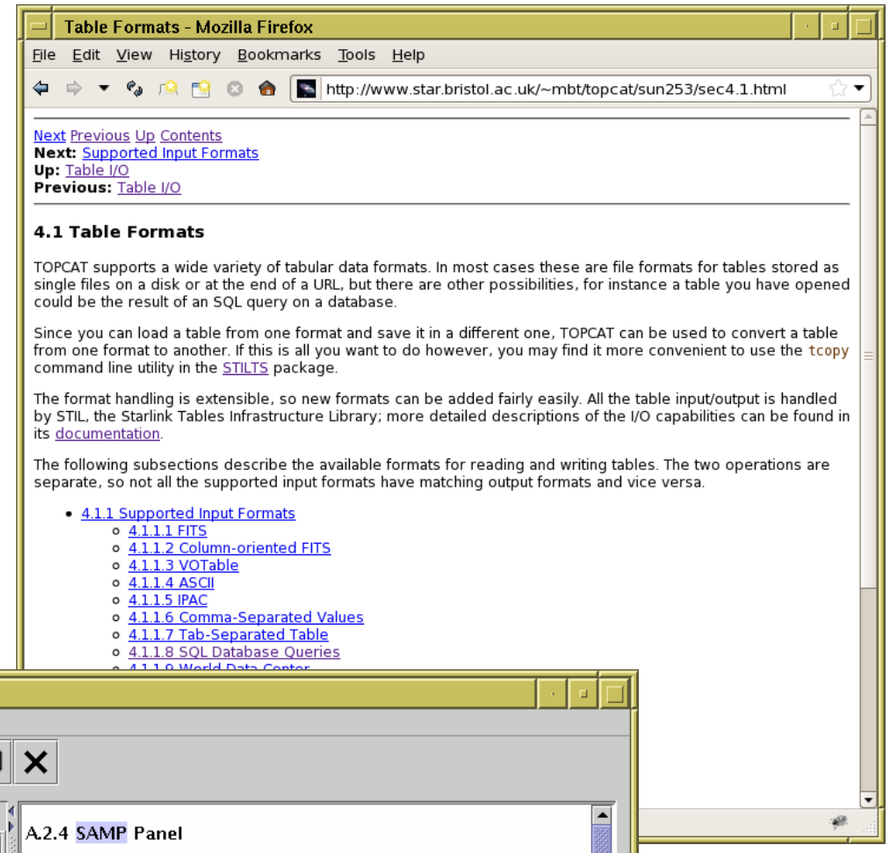
Typical usage:

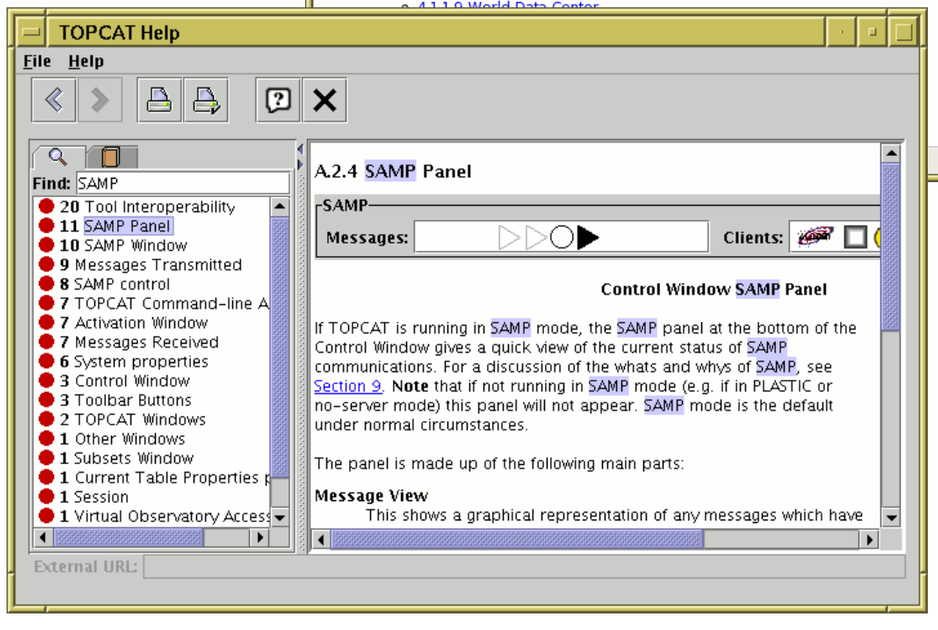
- start off with TOPCAT
- maybe move on to STILTS for more specialised requirements
- TOPCAT >stilts STILTS control helps constructing plot commands

Further Information

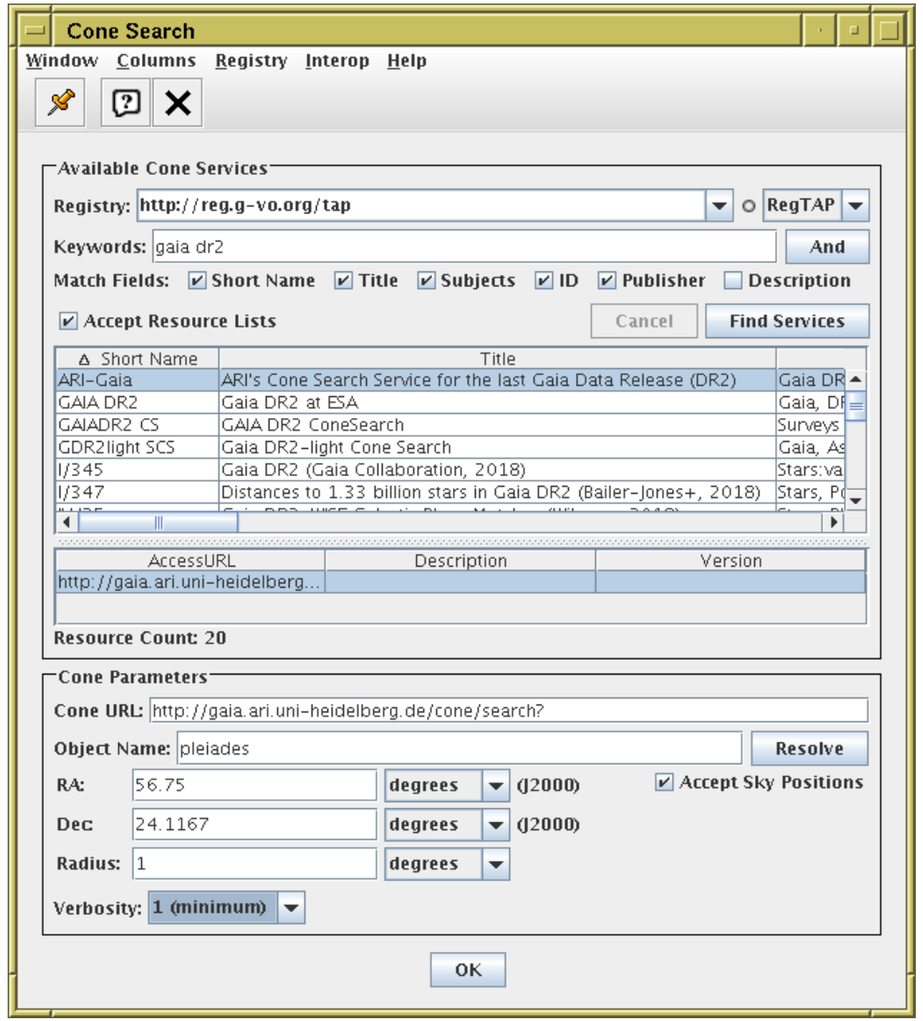
- There are things I haven't mentioned!
- Full tutorial and reference documentation:
  - ▷ HTML/PDF manual on web page

    `http://www.starlink.ac.uk/topcat/` (or Google it)
  - ▷ **Help for Window** button [?] on every window
  - ▷ Help browser includes search tool
  - ▷ More options in Help Menu (including **Help for Window in Browser** item [?])
  - ▷ Or print out the 500-page manual
- Support by email:
  - ▷ on list: topcat-user@jiscmail.ac.uk
  - ▷ in person: m.b.taylor@bristol.ac.uk
  - ▷ All feedback and questions (even dumb questions) welcome!

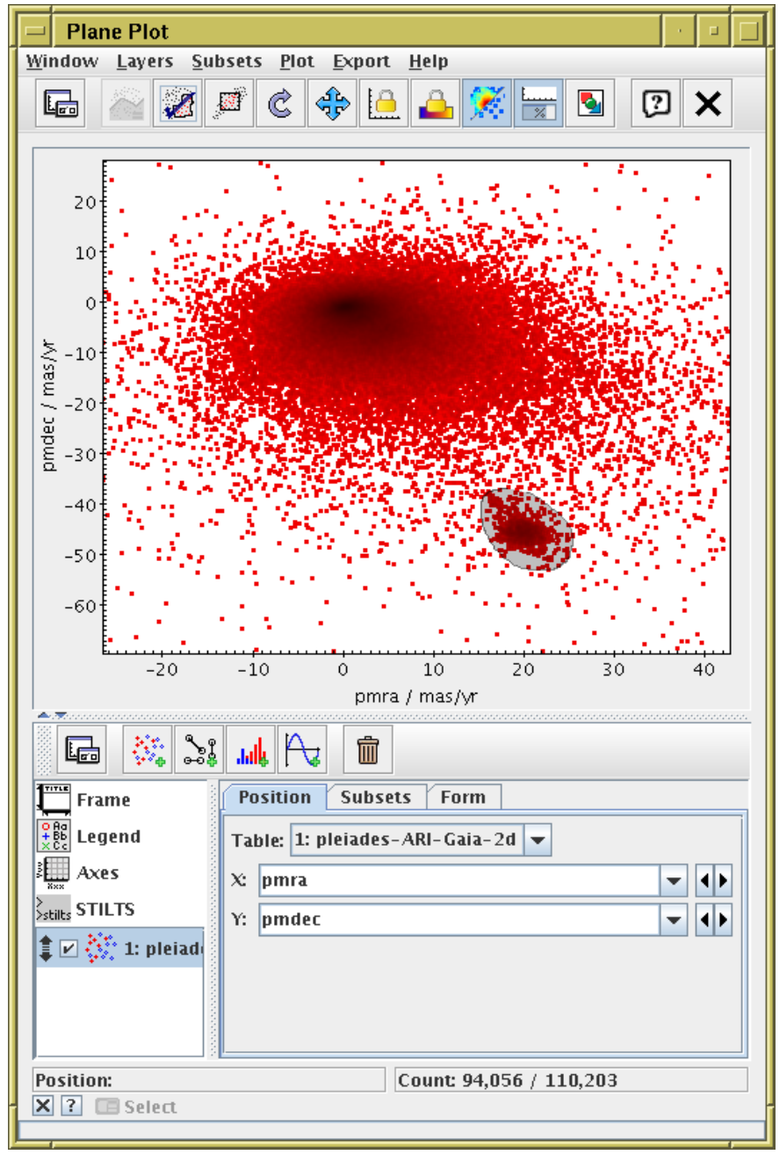




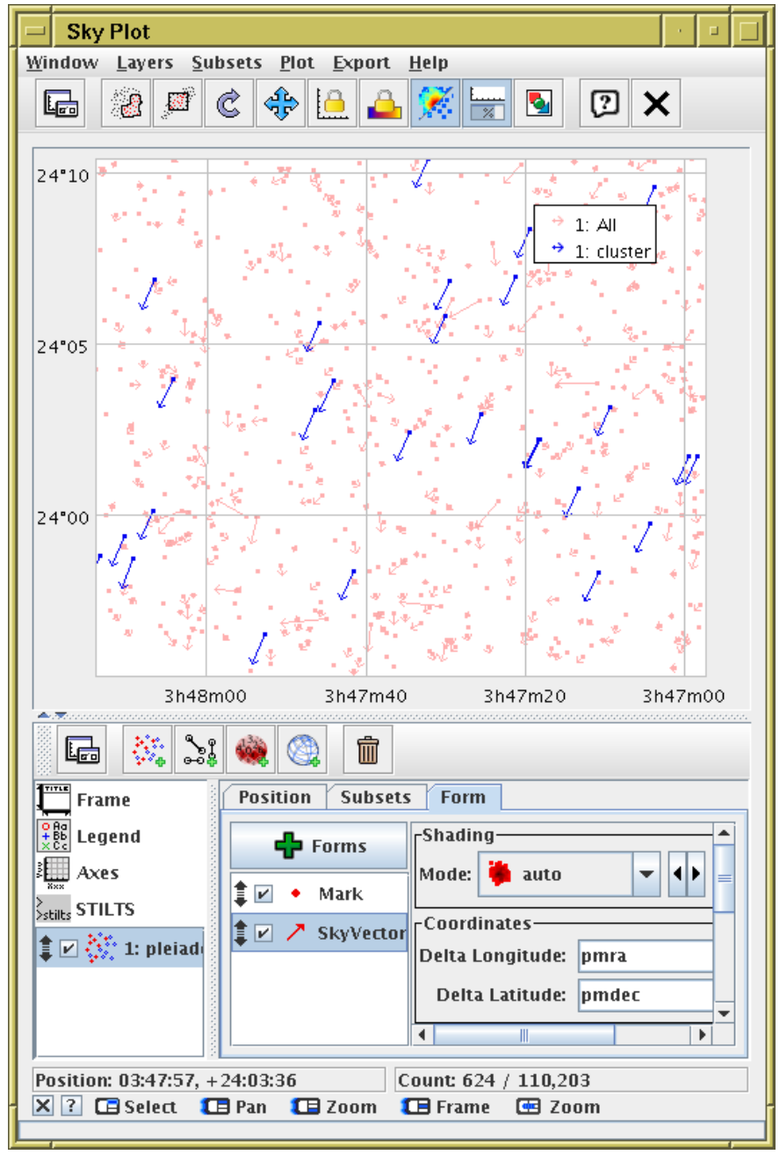
Demo 1: Pleiades in Gaia and 2MASS



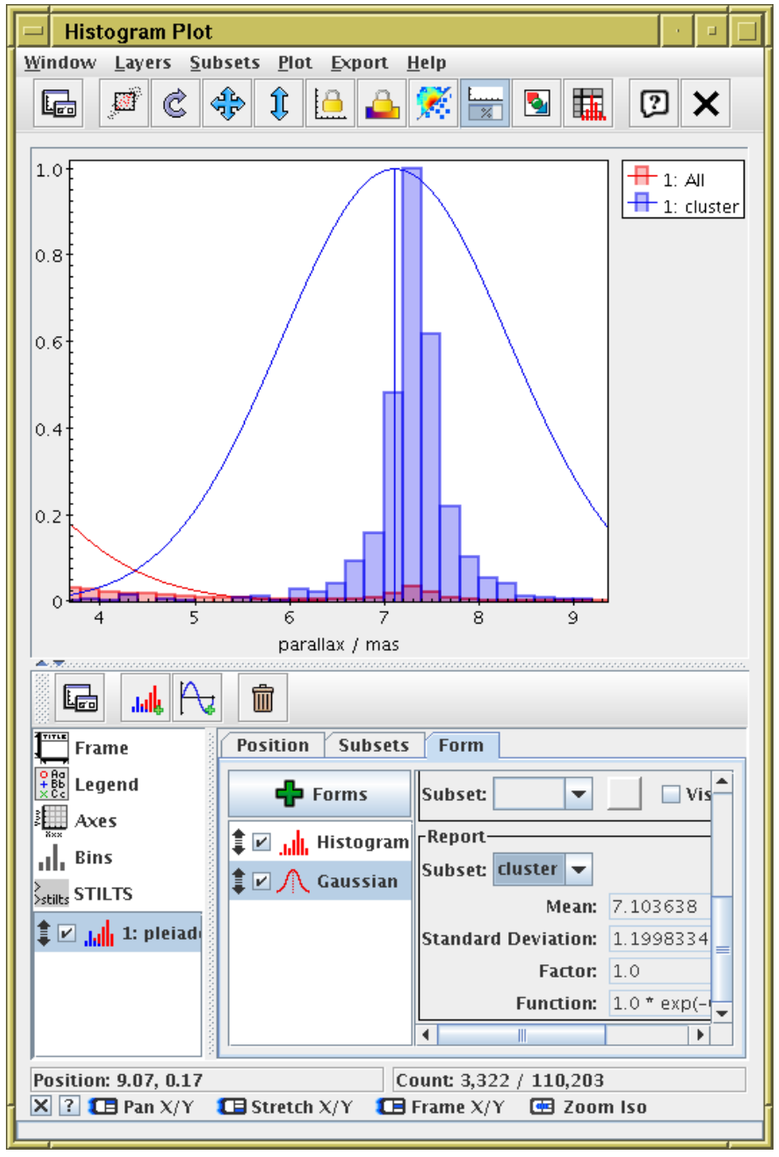
Acquire Gaia DR2 data in the region of the Pleiades using the Cone Search window



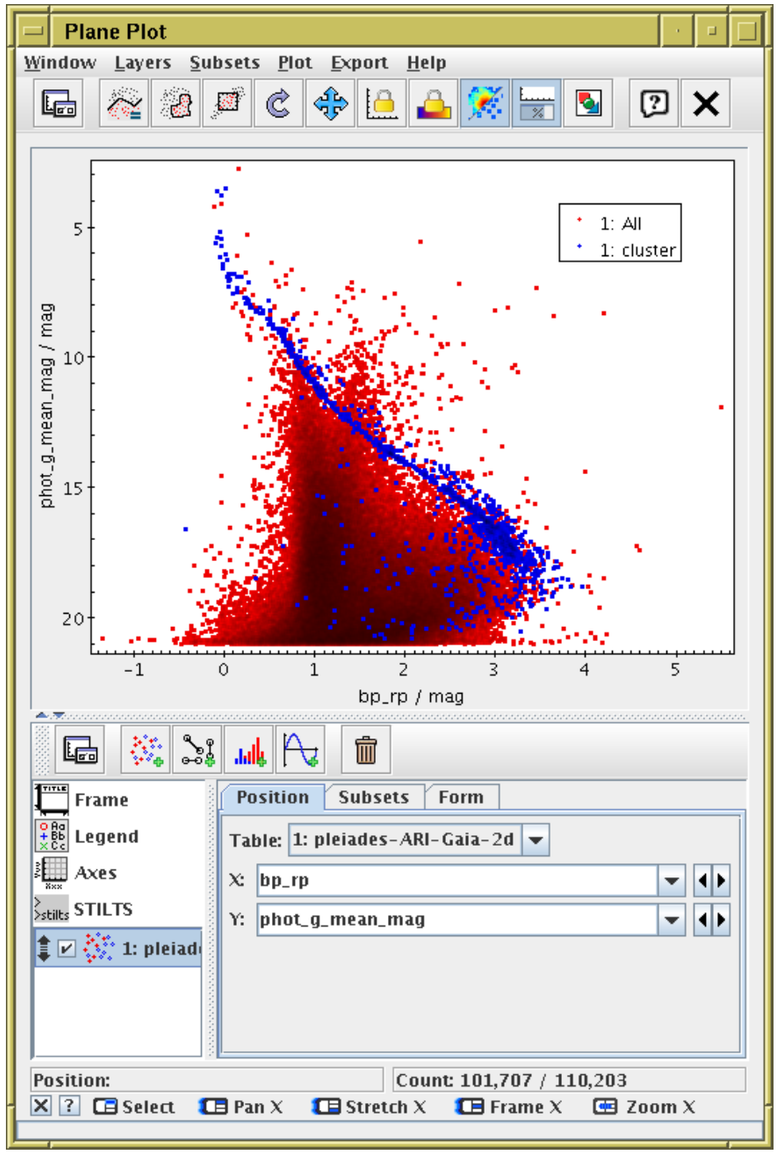
Plot the points in proper motion space,
and select the comoving sub-population graphically to create a new subset



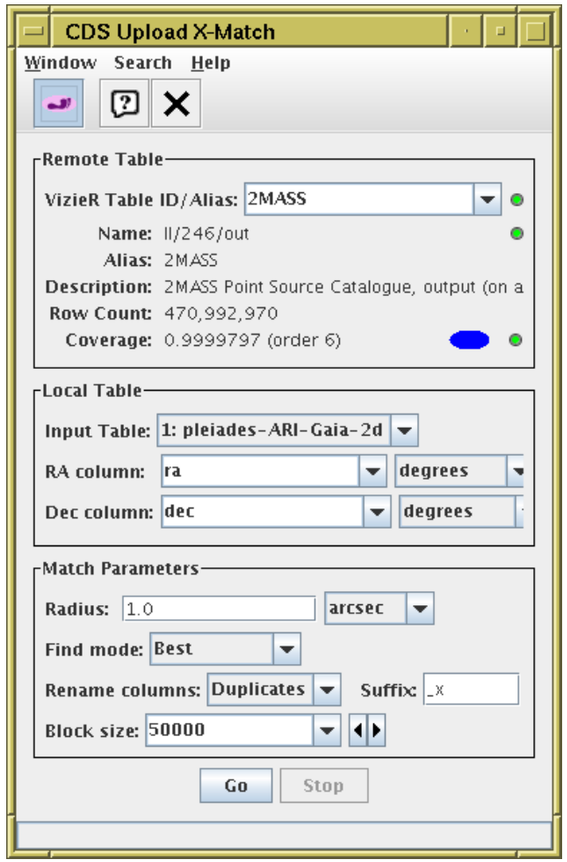
Plot the background and cluster objects on the sky with their proper motion vectors



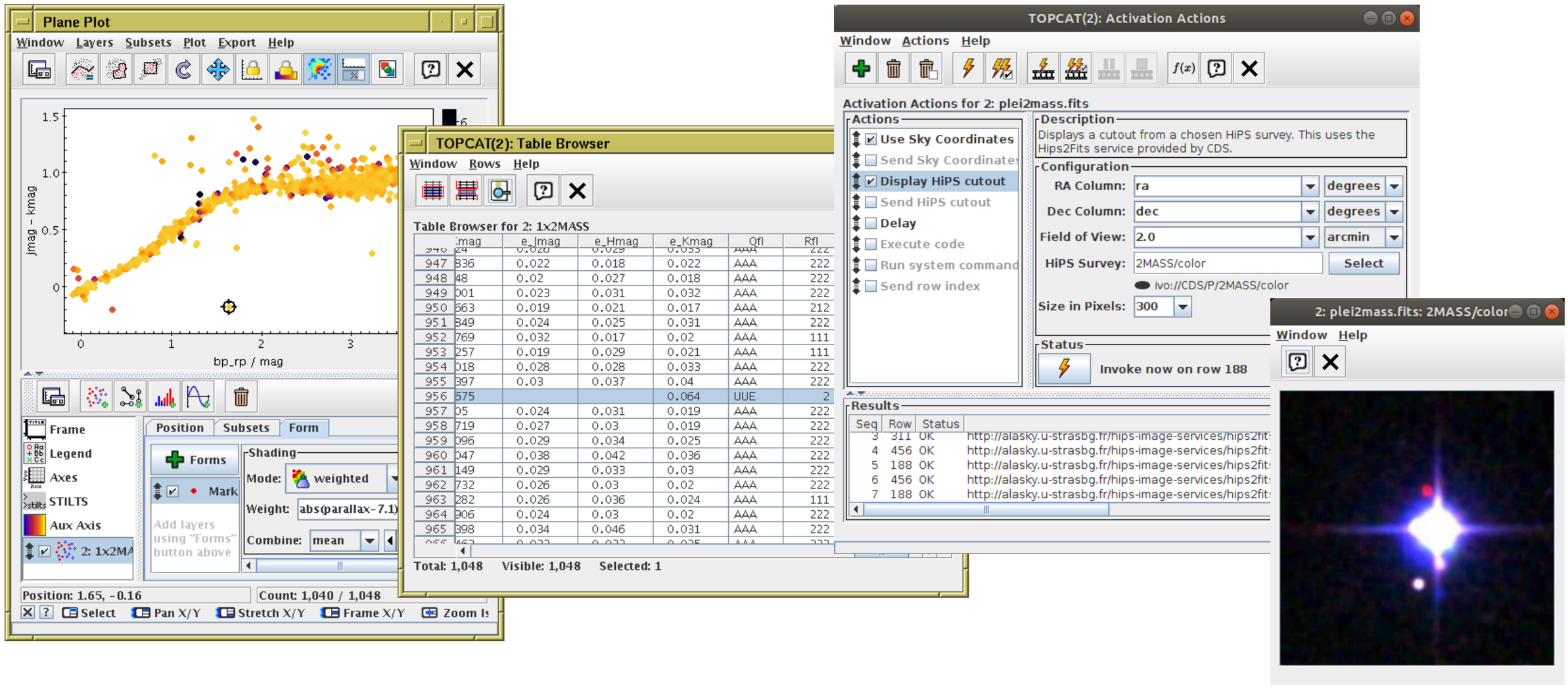
Plot a histogram of the background and cluster objects.
Fit a Gaussian to calculate mean parallax hence distance of cluster objects: $1000/7.1\,\mathrm{mas} \simeq 140\,\mathrm{pc}$.



Plot colour-magnitude diagram using Gaia photometry for cluster and background objects

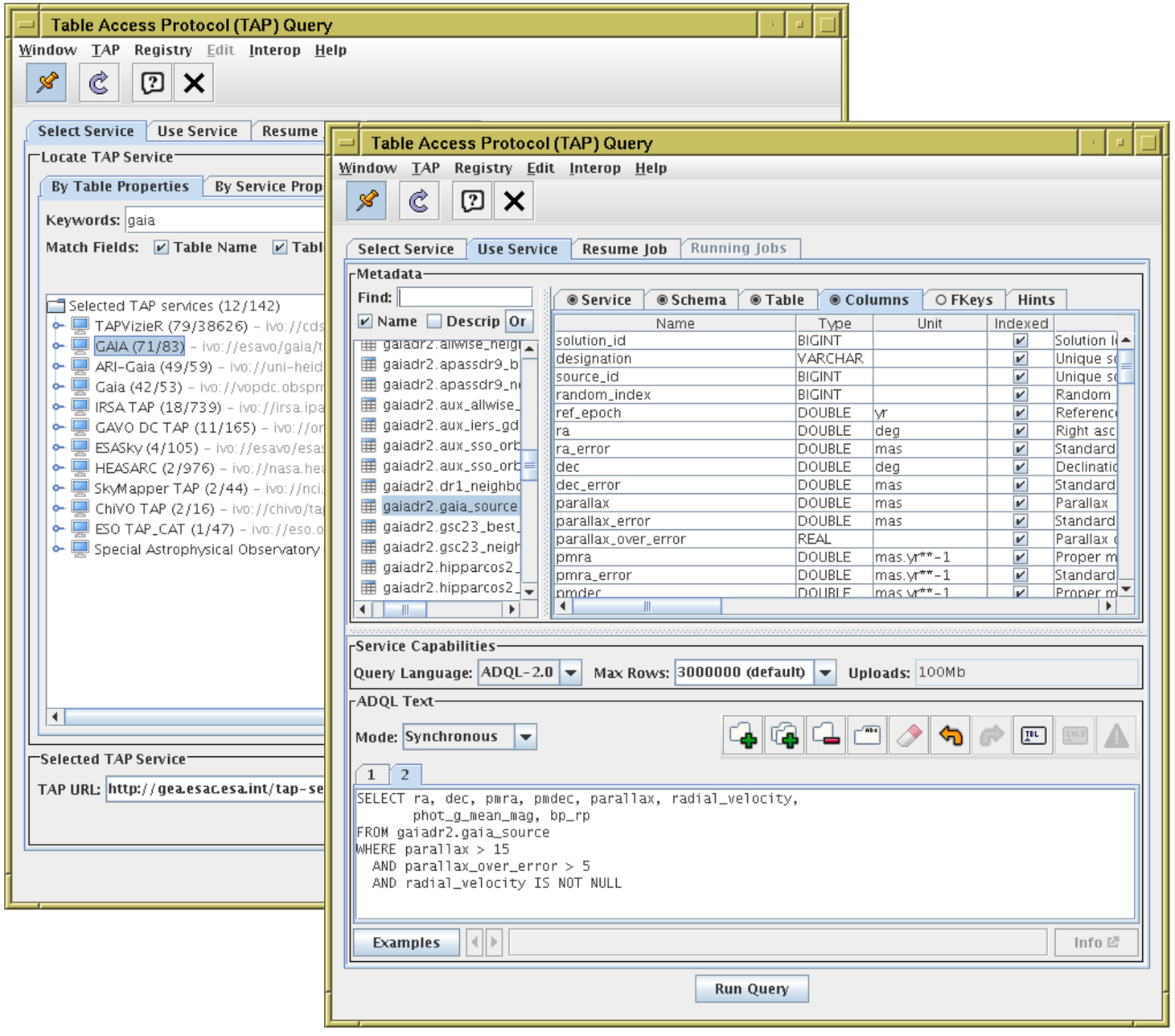


Find 2MASS associations for cluster objects using CDS X-Match window



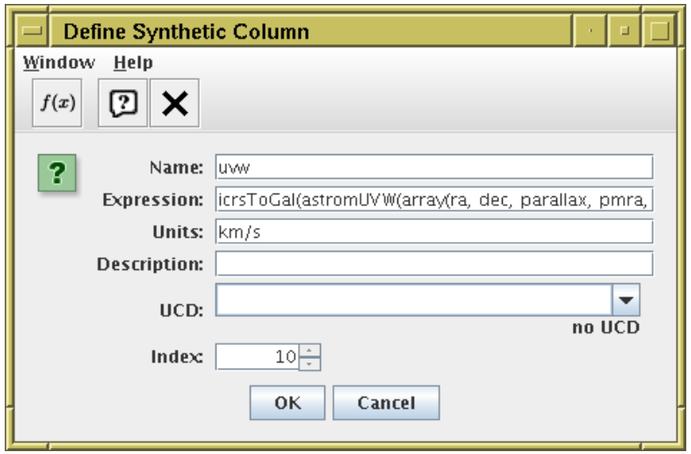
Plot colour–colour diagram using Gaia and 2MASS photometry (Gaia `bp_rp` *vs.* 2MASS `Jmag - Kmag`) and investigate outliers.
Colour points using distance from mean cluster parallax; darker ones are PM interlopers.
Set up Activation Action **Display HiPS Cutout**; use survey `2MASS/color`.
Click on point of interest for linked view: in table display and 2MASS imagery in image window.

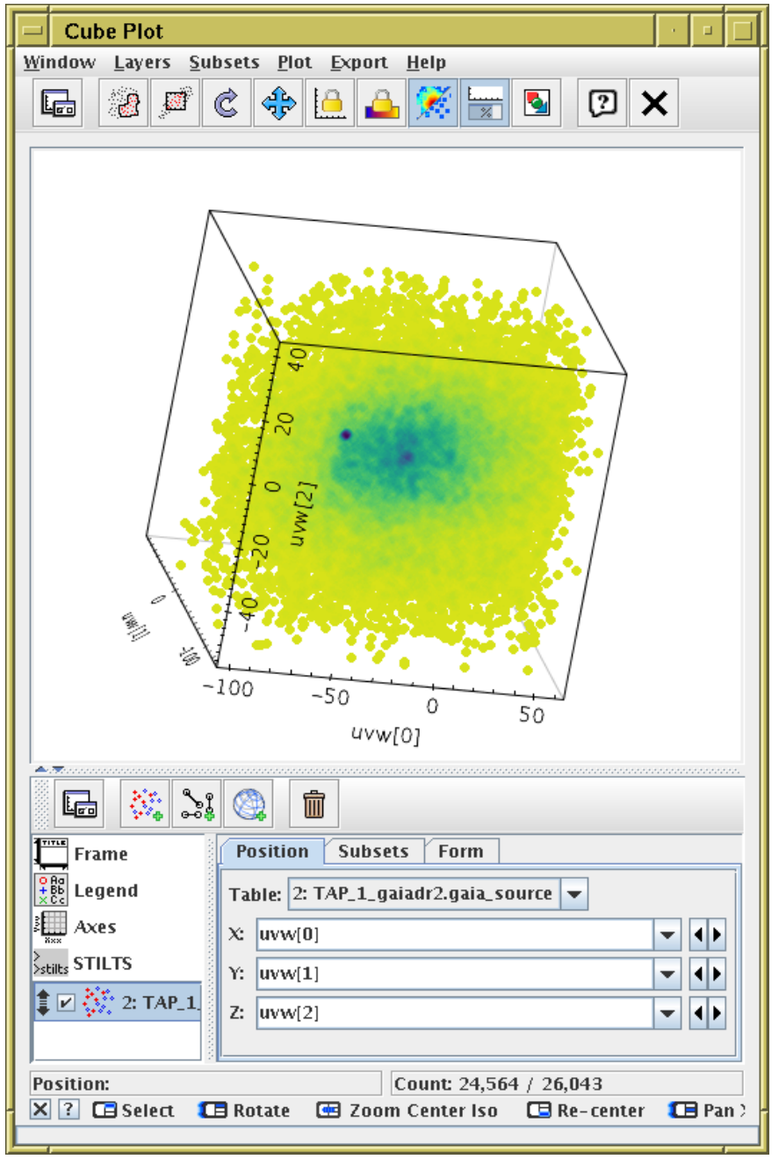
Demo 2: Hyades in 3D using TAP



Use ESA Gaia TAP service
Investigate metadata
Load position and velocity information
for nearby sources:

```
SELECT ra, dec, pmra, pmdec, parallax,
       radial_velocity, bp_rp,
       phot_g_mean_mag + 5*log10(parallax/100) as g_abs
FROM gaiadr2.gaia_source
WHERE parallax > 15
  AND parallax_over_error > 5
  AND radial_velocity IS NOT NULL
```





Define a new column `uvw` giving *Cartesian* velocity components:

```
astromUVW(array(ra, dec, parallax, pmra, pmdec, radial_velocity))
```

Plot in 3D velocity space
Navigate to overdense region - it's the Hyades.
Make a subset and plot it on the sky or a colour-magnitude diagram.

Do It Yourself!

Work through the tutorial at

https://github.com/mbtaylor/tctuto/releases/download/asterics-vo-school-4/tctuto.pdf

- The first part is what I've demonstrated today
- Later parts contain other examples and functionality
- Do whichever parts you like!

Full documentation available online (or in topcat):

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

http://www.starlink.ac.uk/stilts/