Tutorial: Exploring Gaia data with TOPCAT and STILTS

Mark Taylor,
University of Bristol,
m.b.taylor@bristol.ac.uk
Adapted by: Ada Nebot, November 11, 2019

TOPCAT: http://www.starlink.ac.uk/topcat/ *(version 4.6-3 recommended)*

STILTS: http://www.starlink.ac.uk/stilts/ *(version 3.1-6 recommended)*

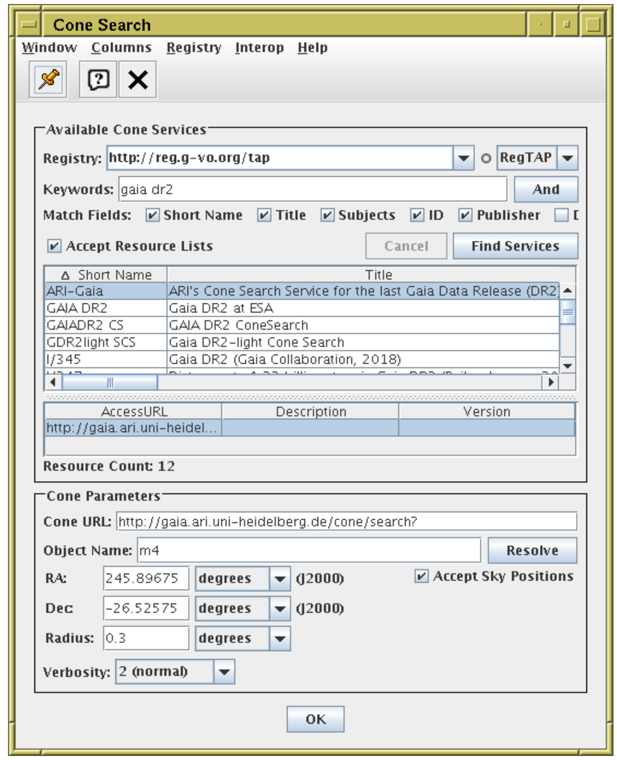
Mailing list: `topcat-user@jiscmail.ac.uk`

Contents

This tutorial uses data from Gaia DR2 [1] to lead you through some of the capabilities of TOPCAT and STILTS. For best results, you should have the manuals to hand: `http://www.starlink.ac.uk/topcat/sun253/` and `http://www.starlink.ac.uk/stilts/sun256/`.

1 Cluster identification #1: Messier 4 in proper motion space

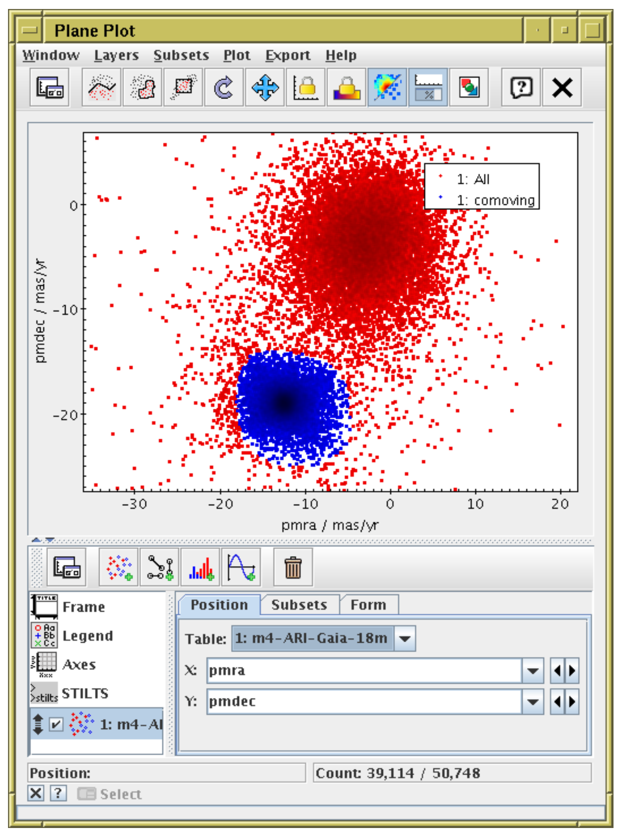
In this example we will determine the mean parallax of the stars in the globular cluster Messier 4 (M4, or NGC 6121).



1.1 Acquire Gaia data in the M4 region

1. Open the **VO|Cone Search** window (i.e. use the **Cone Search** submenu of the **VO** menu in the main topcat window)
2. Fill in **Keywords**: "`gaia dr2`", and hit **Find Services**
3. There are a few options, that should mostly give similar results. `ARI-Gaia` (probably the top one) is a good choice. Select it by clicking on it, and the partial URL of the service appears in the **Cone URL** field
4. **Object Name**: "`M4`", hit **Resolve** to fill in sky position fields
5. **Radius**: "`0.3`" (degrees)
6. Hit **OK**; new table is loaded into topcat main control window, with about 50 000 rows
7. Use the **Graphics|Sky Plot** window to see the positions on the sky

1.2 Identify comoving cluster



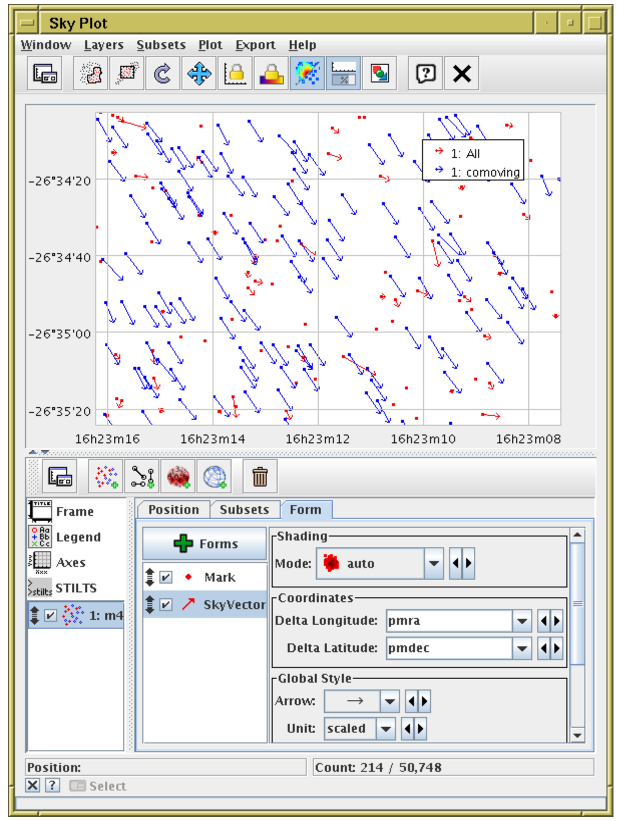
1. Plot sources in proper motion space: **Graphics|Plane Plot** window, **X**: "`pmra`", **Y**: "`pmdec`"
2. Note overdensity far from (0,0); use mouse to navigate (click little [?] button at bottom left for navigation help)
3. Graphically select this comoving cluster as new Subset: **Subsets|Draw Subset Region** button, drag mouse around cluster, hit button again
4. A **New Subset** dialogue pops up: fill in **New Subset Name**: "`comoving`", **Add New Subset**
5. Look in **Subsets** tab of plot window; turn **All** and **comoving** subsets off/on
6. Look in **Views|Row Subsets** window to see new subset

1.3 Examine cluster members

1. Go back to the **Graphics|Sky Plot** from section 1.1 (or open a new one)

2. In **Subsets** tab turn **All** and **comoving** subsets on/off

3. Plot the proper motions. In the **Form** tab, use the **Forms** menu and select the **Add SkyVector** item, with **Delta Longitude**: "`pmra`", **Delta Latitude**: "`pmdec`". The arrows will initially be much too long (units of degrees); you will have to set **Unit**: "`scaled`" (auto-scaling), then adjust to taste with the **Scale** slider. Zoom in so you can see some individual objects. All the cluster objects have similar proper motions, non-cluster ones have various directions, or none (no measured P.M.).



1.4 Determine parallax

1. Plot histogram of parallaxes:
   Graphics|Histogram Plot,
   X: "`parallax`"

2. In **Subsets** tab, make sure only subset **comoving**, and not **All**, is plotted

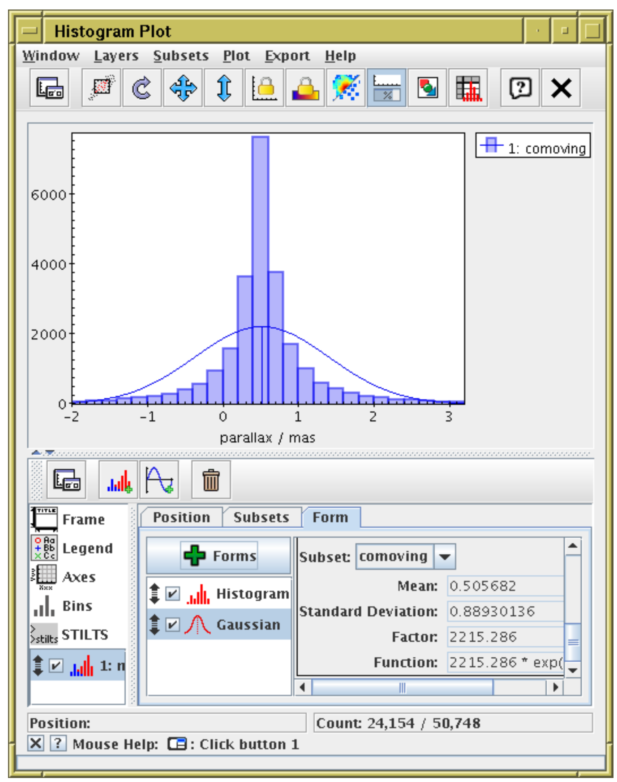
3. If necessary, rescale plot (or navigate with mouse)

4. Fit Gaussian to data: **Forms** menu in **Form** tab, **Gaussian** option.

5. Scroll to bottom of Gaussian layer description in **Form** tab, and read off parallax mean and S.D.

6. Invert to get distance to cluster (1000/parallax in mas = distance in parsec).

7. To do this without plotting, you can read off the mean and S.D. for parallax in the **Views|Column Statistics** window.



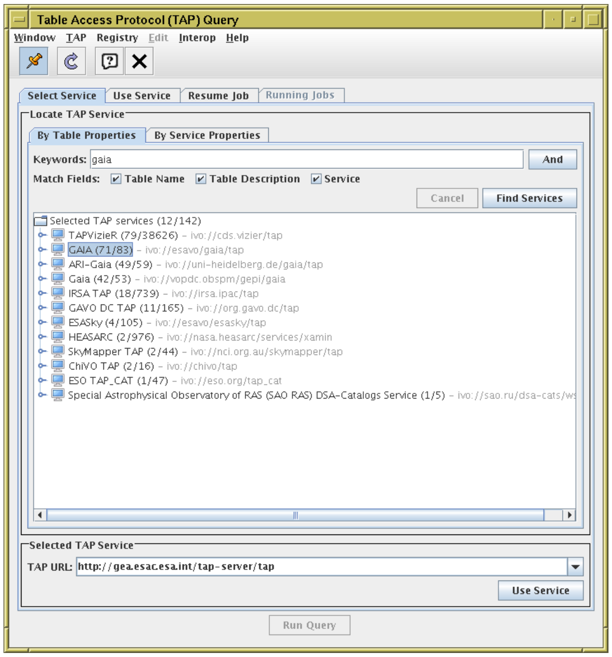
Note: careful when inverting parallaxes!
In general $r = 1/\varpi$ is *not reliable* because of errors. It's OK here because we are averaging over many measurements. Rule of thumb for single measurements: if $\varpi/\sigma_\varpi > 5$ it's probably OK. See Luri et al. 2018 [2] for full discussion.

2 Cluster identification #2: Hyades in phase space

This example locates the Hyades in 3-dimensional velocity space, using Gaia's radial velocity observations. We can't start this time by making a positional query (cone search), since the Hyades is very delocalised on the sky (because it's so close). So we need to make a more sophisticated query using TAP.

2.1 Locate Gaia TAP service

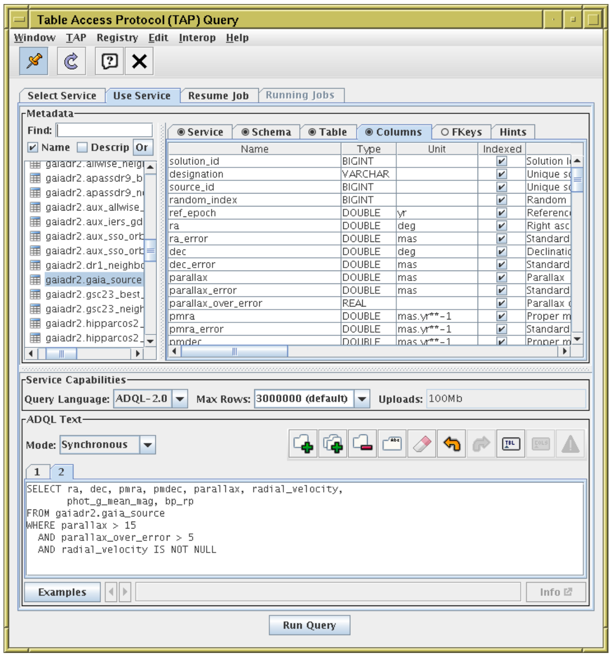
1. Open the TAP window:
   VO|Table Access Protocol (TAP) Query
2. Fill in **Keywords**: "gaia" and hit **Find Services** button
3. There are several services with Gaia data in various forms; **GAIA** (ESA) or **ARI-Gaia** (Heidelberg) are good choices. The service URL appears in the field at the bottom of the window.
4. Hit the **Use Service** button at the bottom



2.2 Explore the TAP service

Use the TAP window to explore the tables that are present and their metadata.

1. Browse the table list on the left, The tables in the `gaiadr2` schema are the ones with Gaia DR2 data.
2. Select table `gaiadr2.gaia_source` and look at **Table** and **Columns** tabs, to see information about available columns.
3. Look in the **Service** tab to see information about the service
4. Look in the **Hints** tab for a very basic ADQL cheat sheet
5. Type in some very simple ADQL:
   "`SELECT TOP 10 ra, dec FROM gaiadr2.gaia_source`"
6. Note that syntax errors (including partial or misspelt tables/columns) are highlighted in red.
7. Hit **Run Query** to run the query; if successful a new table is loaded



2.3 Acquire astrometric data

In the TAP window, execute this query:

```
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
```

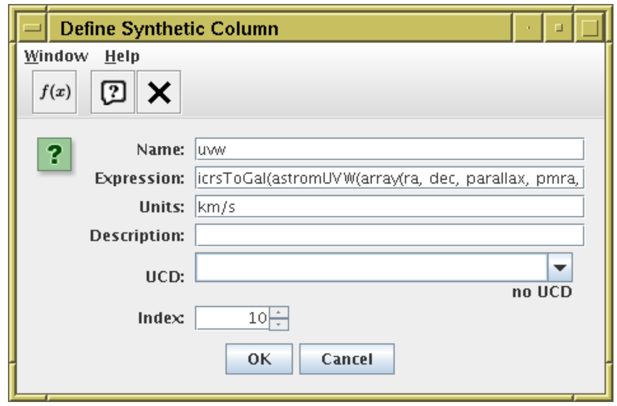
The result should contain about 26 000 rows.

The query is for all the nearby sources (nominally within $1000/15 \approx 66$ parsec) with radial velocities (only about 7 million DR2 sources have RV) and good determinations of parallax. The fact that parallax error is $\leq$20% means that it's OK to invert parallax to get distance. We are retrieving all the basic astrometric parameters and some photometry.

The Hyades should be in there somewhere.

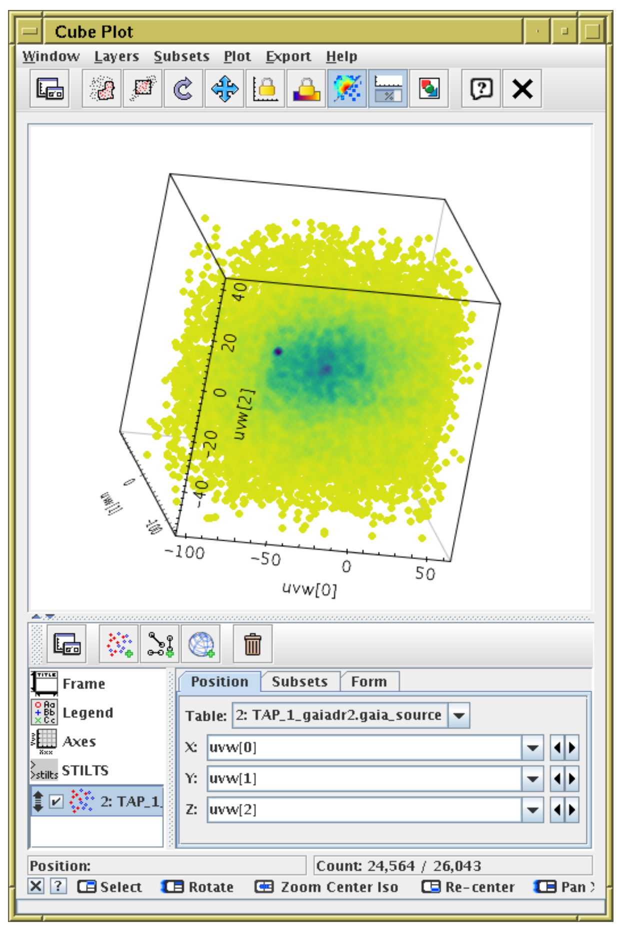
2.4 Calculate 3-d velocity components

We have the astrometric quantities measured by Gaia, and want to turn them into coordinates in 3-d phase space. TOPCAT's *expression language* can help.



1. Open the **Help|Available Functions** browser to see what functions TOPCAT provides (they are listed in the manual too).
2. Look under the **Gaia** option to see astrometry-specific items
3. We will use the `astromUVW` and maybe `icrsToGal` and `astromXYZ` functions. The **Examples** items in the function documentation are useful; for use with the `gaia_source` catalogue, you can often just cut and paste.
4. Open the **Views|Column Info** window and choose the **Columns|New Synthetic Column** menu item
5. Create a new column giving Cartesian velocity components: **Name**: "`uvw`", **Expression**: "`astromUVW(array(ra, dec, parallax, pmra, pmdec, radial_velocity))`", **Units**: "`km/s`" That gives velocities along ICRS axes. If you want it in Galactic coordinates, wrap the whole expression in the `icrsToGal(...)` function.
6. Look at the new column in the **Views|Data Window**. It is a 3-element array; you can get the array elements using expressions `uvw[0]`, `uvw[1]`, `uvw[2]`

2.5 Identify Hyades graphically in 3-d velocity space



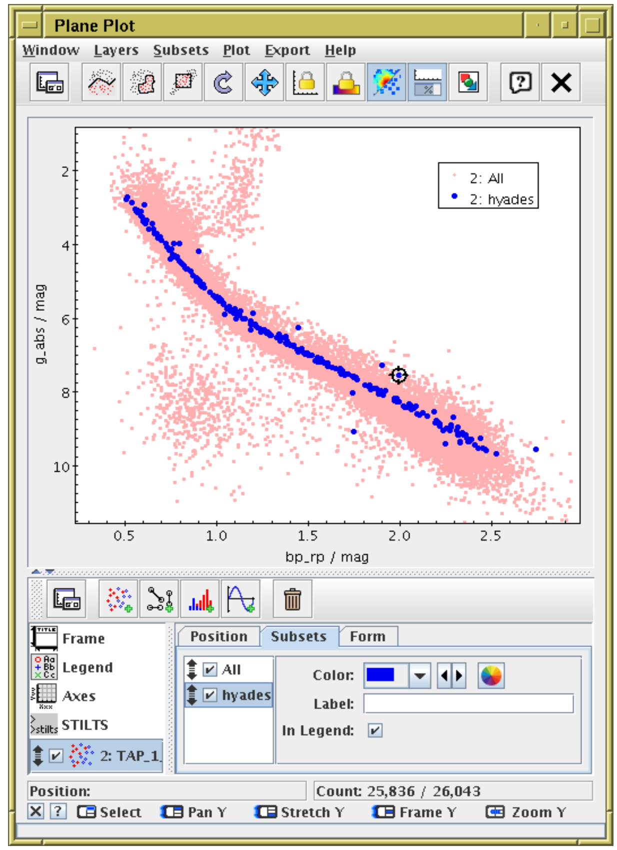
1. Plot points in 3-d space: **Graphics|Cube Plot**, **X**: "`uvw[0]`", **Y**: "`uvw[1]`", **Z**: "`uvw[2]`". Note, you can type in any expression for the plot coordinates, you don't have to just select from available columns. The `uvw` column itself doesn't appear in the selection list, since it's not a scalar.
2. Select **Mode**: "`density`" in the **Form** tab.
3. Now, navigate through the cube to find an overdensity. This takes a bit of practice, but it's fun once you work out how. Click the little ? button at bottom left for navigation help; the most useful actions are mouse wheel (2-fingered up/down drag on some trackpads) to zoom, and right click on a dense region to recenter.
4. Navigate so only the objects in the overdense region are visible inside the wireframe - these are the Hyades.
5. Use **Subsets|New Subset From Visible**, **Name**: "`hyades`", **Add Subset**.
6. Go back and plot this subset on the sky (**Graphics|Sky Plot**), or in 3-d space (**Graphics|Sphere Plot**, **Lon**: "`ra`", **Lat**: "`dec`", **Radius**: "`1000./parallax`")

2.6 Colour-magnitude diagram

1. Plot a colour-magnitude diagram: **Graphics|Plane Plot**, **X**: "`bp_rp`", **Y**: "`g_abs`". Use the **Axes** control, **Coords** tab, **Y Flip** checkbox to flip it the right way round.

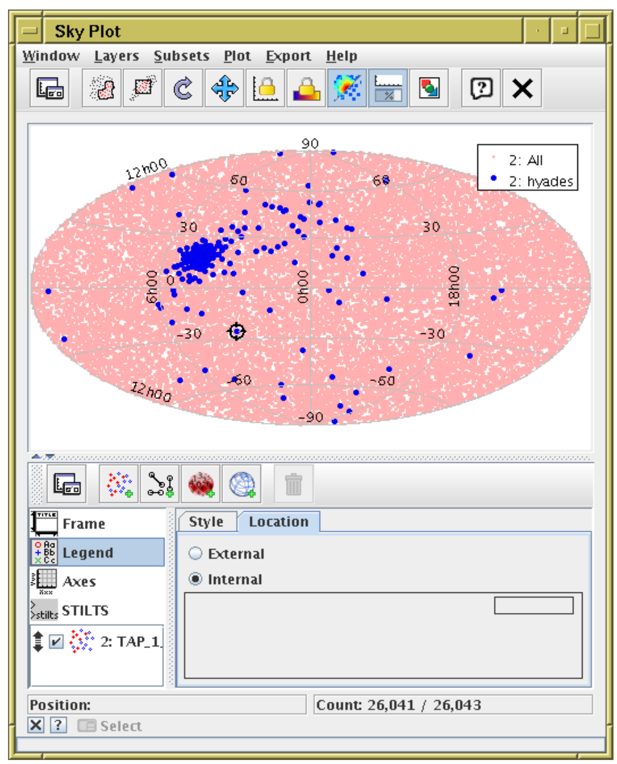
2. Hyades sit on a nice tight main sequence!

3. There are a few outliers. Click on them, see the position show up in the sky plot too. In some cases, you can see by sky position that they are non-members.



Bonus

- Can you find any other clusters in velocity space?
- Try refining the selection by localising in position space too.
- What is the mean distance to the Hyades?
- Try using Aladin and SAMP along with TOPCAT to investigate the outliers.



3 Match Gaia and HST observations

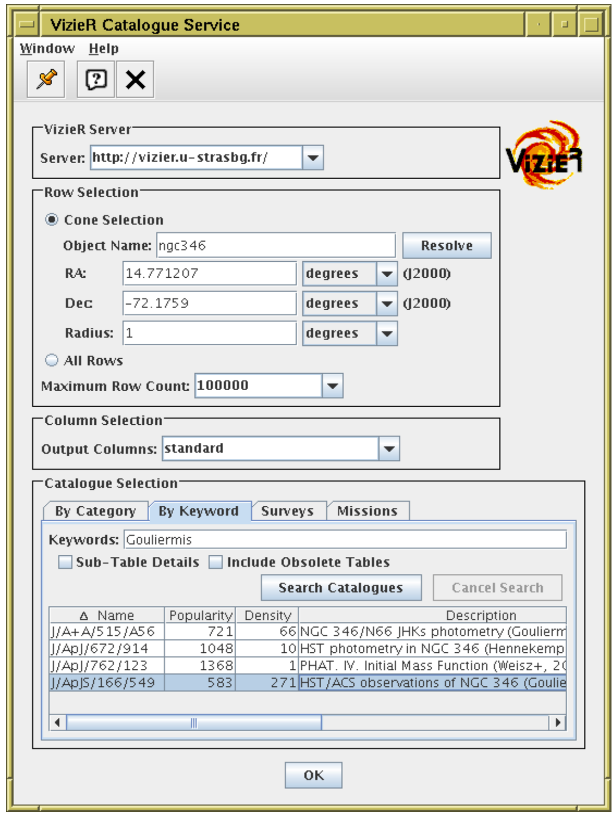
In this example we have a local catalogue from a publication by Gouliermis et al. 2006 [4], available in VizieR as J/ApJS/166/549. This contains about 100 000 sources observed by the ACS instrument on the Hubble Space Telescope at epoch ≈ J2004.6 of stars in NGC346, a cluster in the Small Magellanic Cloud. We match these positions with positions in the main Gaia catalogue at J2015.5.

3.1 Acquire HST observations

There are various ways to do this, but here we will use TOPCAT's VizieR dialogue window, which talks directly to the VizieR catalogue service.

1. Open **VO|VizieR Catalogue Service** window
2. **Object Name**: "`ngc346`", and **Resolve** to fill in **RA** and **Dec**
3. **Radius**: "`1`" (degrees)
4. **Maximum Row Count**: "`100000`" (or some large number)
5. Catalogue selection panel: **By Keyword** tab
6. Fill in **Keywords**: "`Gouliermis`"
7. Select "`J/ApJS/166/549`"
8. Hit the **OK** button at the bottom. A new table with 99 079 rows should be loaded

Alternatively, you could download the table from the VizieR web page.

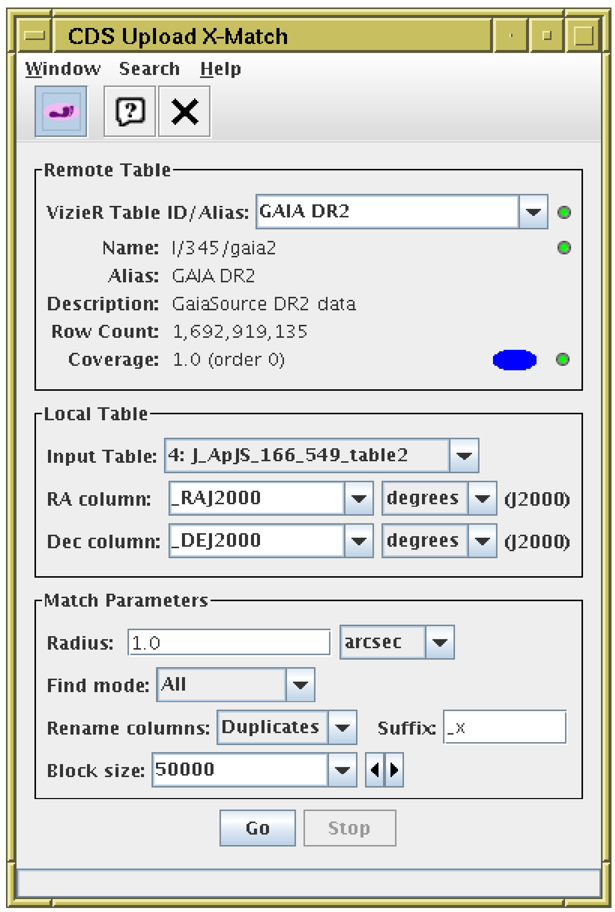


3.2 Crossmatch with Gaia

Now we want to find associations of the HST objects with sources from Gaia DR2. Use the CDS X-Match service from TOPCAT. This uploads a local table to the CDS X-Match service, where the match is made against the Gaia DR2 catalogue (or any other catalogue in VizieR). The resulting matched catalogue is then received as a new table in TOPCAT.

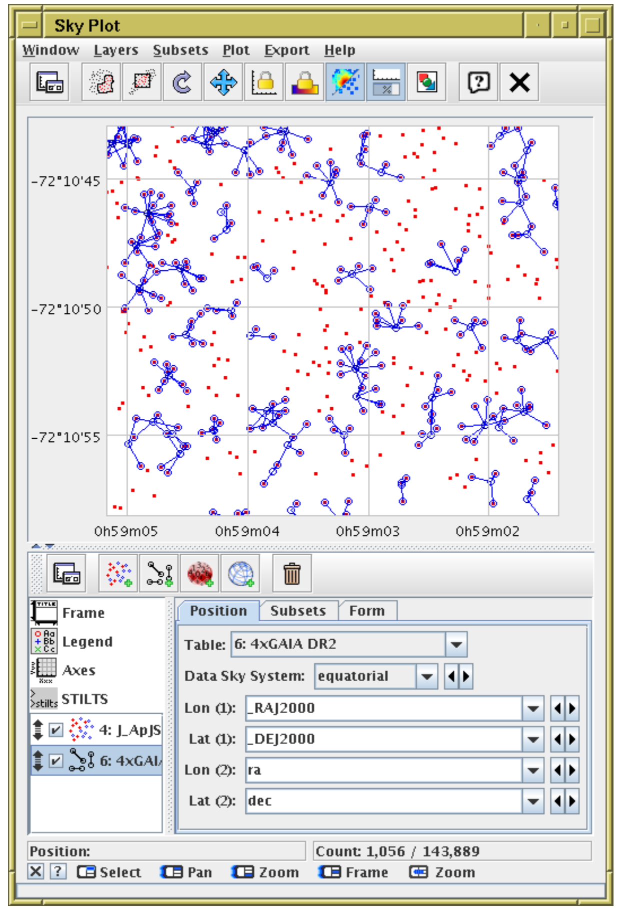
1. Open the **VO|CDS Upload X-Match** window
2. Fill in the fields:
   VizieR Table ID/Alias: "`GAIA DR2`"
   Input Table: "`J_ApJS_166_549_table2`"
   (or whatever the HST table is called)
   RA column: "`_RAJ000`", **Dec column**: "`_DEJ2000`"
   (should be filled in automatically)
   Radius: "`1`" (arcsec)
   Find Mode: "`All`"
3. Hit **Go**; within a few seconds, it should inform you that a new table has been loaded, with 22 405 rows.
4. Look at the columns of the new table (all HST followed by all Gaia) in the **Views|Column Info** window
5. View the new table in a **Graphics|Sky Plot**

Note: the match is done with Gaia coordinates rolled back (using Gaia proper motions) to J2000.0. These propagated columns are in the matched table as `ra_epoch2000`, `dec_epoch2000`.



3.3 Visualise the crossmatch

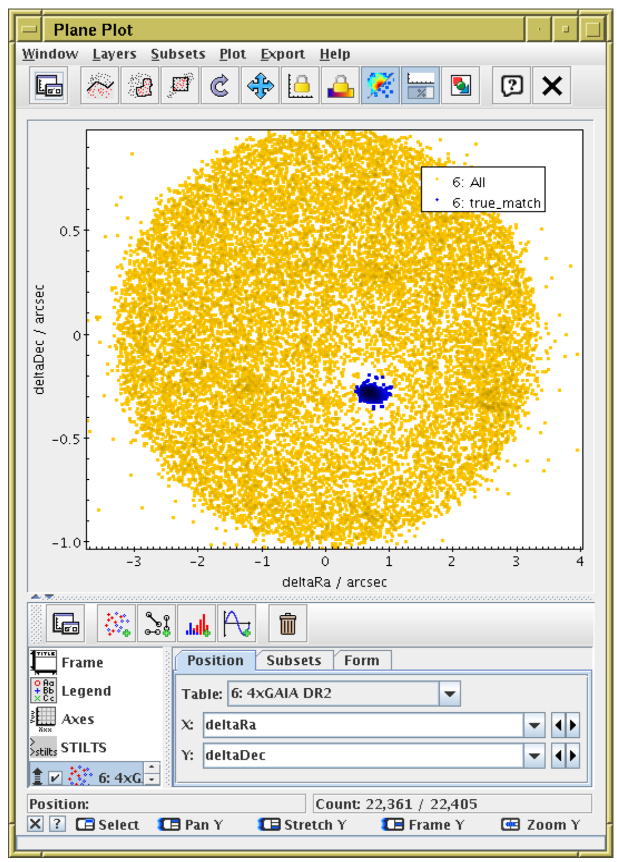
1. Open a **Graphics|Sky Plot** window
2. Plot the HST observations: **Table**: "J_ApJS_166_549_table2", **Lon**: "_RAJ2000", **Lat**: "_DEJ2000"
3. Overplot the actual matches. Add a new Pair layer: **Layers|Add Pair Control** and fill in:
   Table: "4xGAIA DR2" (or whatever the xmatch result table is called) and both sets of coordinates:
   Lon(1): "_RAJ2000", **Lat(1)**: "_DEJ2000" (HST)
   Lon(2): "ra", **Lat(2)**: "dec" (Gaia)
4. Zoom in to look at the associations. There are too many!
5. You can fiddle around with the tab to make the plot clearer, e.g. add a **Mark2** layer; change marker size, shape or colour.



Visualising the results of a crossmatch is very often a good idea, unless you're pretty sure what you're going to get. Here, you can see it was crucial to understand the results.

3.4 Investigate and identify matches

1. Add new columns giving RA/Dec discrepancies between HST and Gaia positions:
   Open **Views|Column Info** window,
   then define new columns using **Columns|New Synthetic Column**:
   Name: "deltaRa", **Expression**: "3600*(ra - _RAJ2000)", **Units**: "arcsec"
   Name: "deltaDec", **Expression**: "3600*(dec - _DEJ2000)", **Units**: "arcsec"
2. Use **Graphics|Plane Plot** window,
   plot **X**: "deltaRa", **Y**: "deltaDec"
3. Identify overdense region, select as in section 1.2, define new subset **true_match**.
4. Go back to the sky associations plot from the previous section, and use the **Subsets** tab to visualise which are the true matches.
5. Make a colour-colour diagram combining HST and Gaia photometry:
   Use **Graphics|Plane Plot** window, plot **X**: "Vmag-Imag", **Y**: "bp_rp", display **true_match** subset only.
   What are the two populations?



3.5 Alternative crossmatch: use local files

The crossmatch in section 3.2 was done by sending a local file to an external service. This is often an efficient way to do it, but there are other options. Here, we will do the same crossmatch by operating on two local files with positions covering the same sky region.

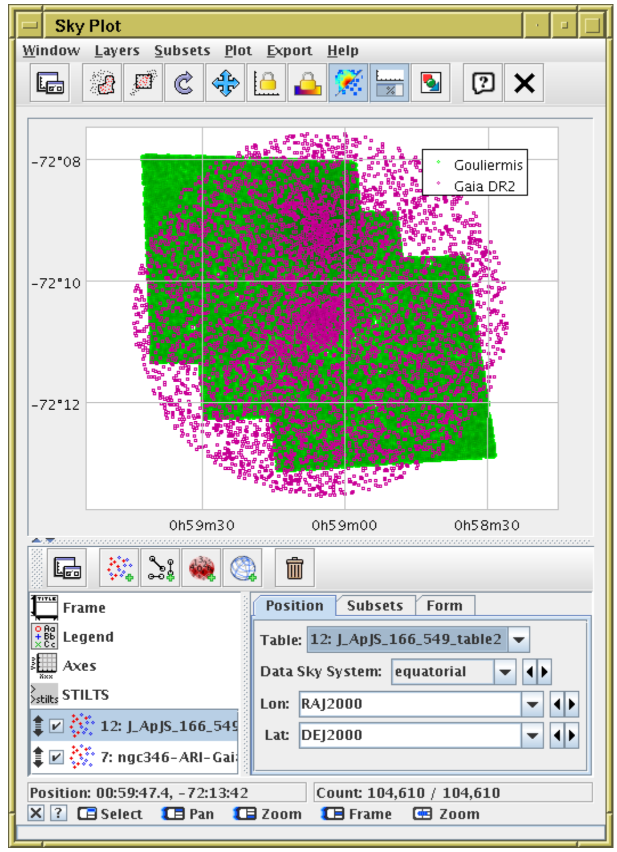
1. First, retrieve Gaia data in the region of interest. Use the **VO|Cone Search** window as in section 1.1, but this time fill in **Object Name**: "`ngc346`", **Radius**: "`0.05`" (degrees).

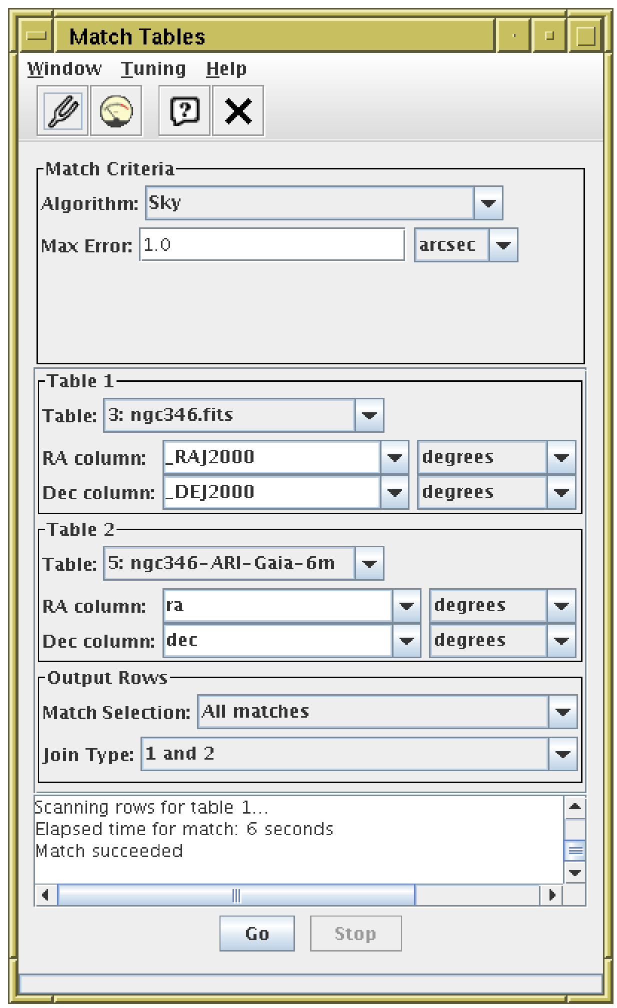
2. Plot the Gouliermis (from VizieR) and Gaia (from Cone Search) datasets on the sky: open the **Graphics|Sky Plot** window, fill in RA and Dec as **Lat** and **Lon** for one of the tables, then use the **Layers|Add Position Control** action to overplot the same thing for the other dataset.

3. Open the **Joins|Pair Match** window from the main control window. Default **Match Criteria** are OK in this case. Fill in the **Table 1** and **Table 2** details for the Gouliermis and Gaia tables. Entry **Match Selection**: "`All matches`".

4. Hit **Go** and wait a few seconds for the match to complete.

5. When complete, a popup window will tell you, and offer to **Plot Result**. If you select this option, you will see a plot like the one from section 3.3, except that the unmatched Gaia sources are also plotted (which can sometimes be useful information).





Other matching options are available in the local match windows, including identifying objects that don't match, matching internally within a table, matching between three or more tables, etc. Note that unlike most things in TOPCAT, crossmatching can take up significant amounts of memory, so matching multi-million-row tables can sometimes grind to a halt or fail.

Bonus

- Use the **Graphics|Histogram Plot** (as in section 1.4) to find the mean values of δRA, δDec for the true matches. If these are non-zero, why?

- The match done here is between Gaia positions at J2015.5 and HST positions taken at approximately J2004.6. When using the X-Match service in section 3.2, Gaia proper motions were automatically applied to predict the Gaia positions at J2000.0. Use the `epochProp` function in the expression language to do the match with the positions as Gaia proper motions predict for J2004.6. Does it make much difference?

Note that the match done here ingores positional errors. To identify the true matches errors need to be taken into account.

References

[1] Gaia Collaboration et al., "Gaia DR2: Summary of the contents and survey properties", Astronomy and Astrophysics *616*, A1 (2018), 2018A&A...616A...1G

[2] X.Luri et al., "Gaia DR2: Using Gaia parallaxes", Astronomy and Astrophysics *616*, A9 (2018), 2018A&A...616A...9L

[3] L.Lindegren et al., "Gaia DR2: The astrometric solution", Astronomy and Astrophysics *616*, A2 (2018), 2018A&A...616A...2L

[4] D.A.Gouliermis, A.E.Dolphin, W.Brandner and Th.Henning, "The Star-forming Region NGC 346 in the Small Magellanic Cloud with Hubble Space Telescope ACS Observations. I. Photometry", ApJS, 166 p.549 2006ApJS..166..549G

[5] W.-C.Jao, T.J.Henry, D.R.Gies, N.C.Hambly, "A Gap in the Lower Main Sequence Revealed by Gaia Data Release 2", ApJ Letts, 861, L11 (2018), 2018ApJ...861L..11J

Acknowledgements

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.