# Data mining with TOPCAT and ADQL

## Creating a target list

Max Pritzkuleit
Research workshop on evolved stars
27.08.2024

# Overview

- Topcat
  - Basic overview
  - Table visualisation/manipulation
  - Visualisation tools
  - Crossmatching
- ADQL
  - Basic commands and hands-on exercise
- Exercise: cross-match with ATLAS – creating our target list for photometry
- Creating our target list for spectroscopy
  - Defining the region of interest
  - ADQL query
  - Observational constraints

# TOPCAT

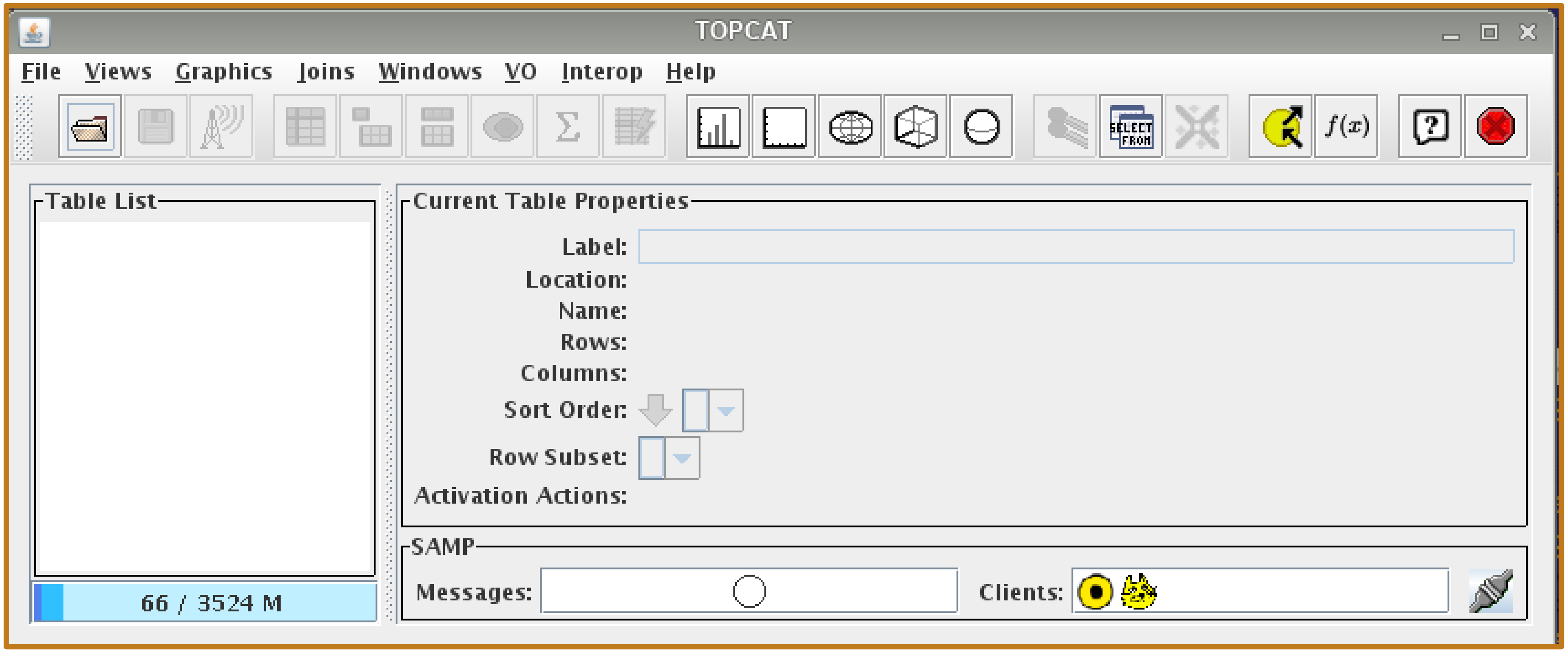
**Tool for OPerations on Catalogues And Tables**

***Does what you want with tables***

- Website: http://www.star.bristol.ac.uk/~mbt/topcat/
- Manual: http://www.starlink.ac.uk/topcat/sun253/
- Why TOPCAT?
  - Easy to use
  - Easy to learn
  - Easy to investigate data — good for exploratory analysis
  - Simple things obvious, complicated things documented
  - Easy to install and run
  - Fast
  - Copes with large data sets

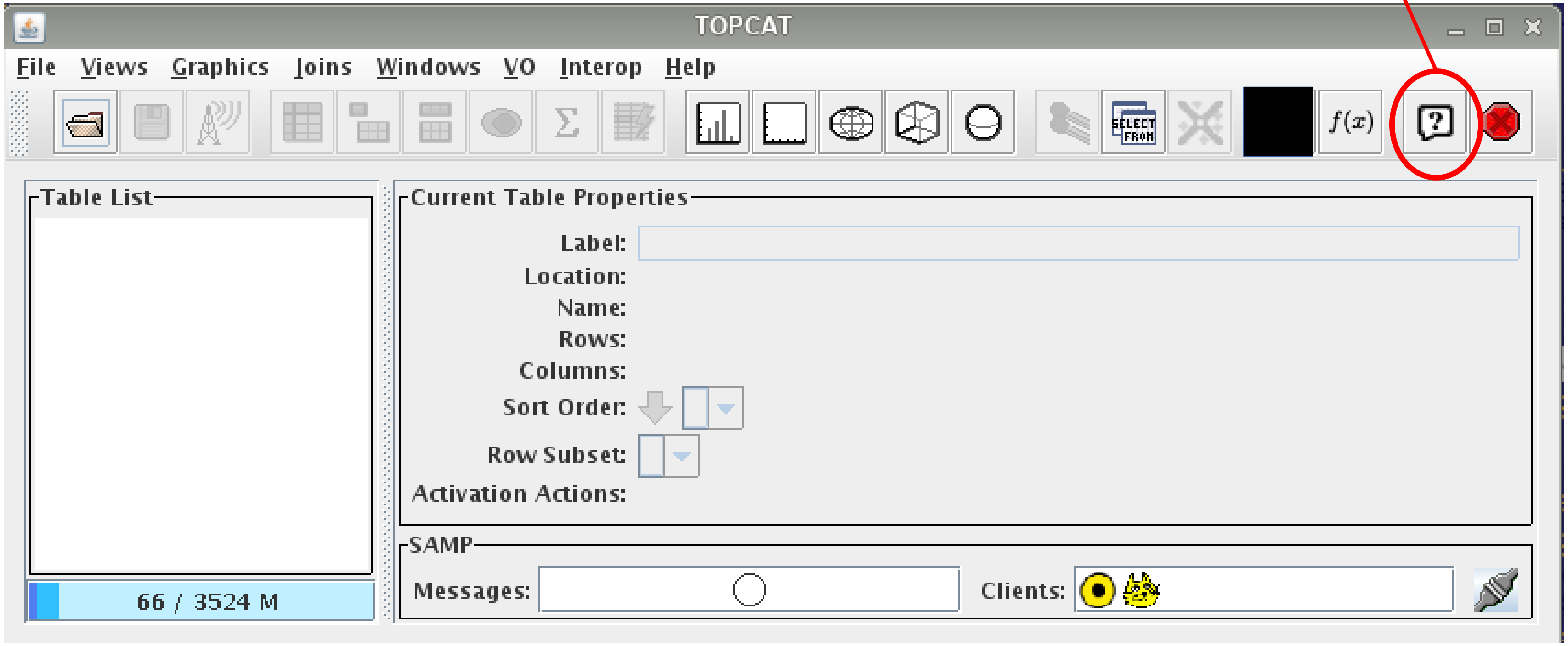
- What can we do with TOPCAT?
  - Read/write tables in multiple formats
  - View/edit data
  - View/edit metadata
  - Plot data
  - Crossmatch — efficient and very flexible
  - (Simple) Calculations
  - Access Virtual Observatory (VO) services
- What can't we do with TOPCAT?
  - Images and spectra (tables only!)
  - Do astronomy for you

# TOPCAT – start window

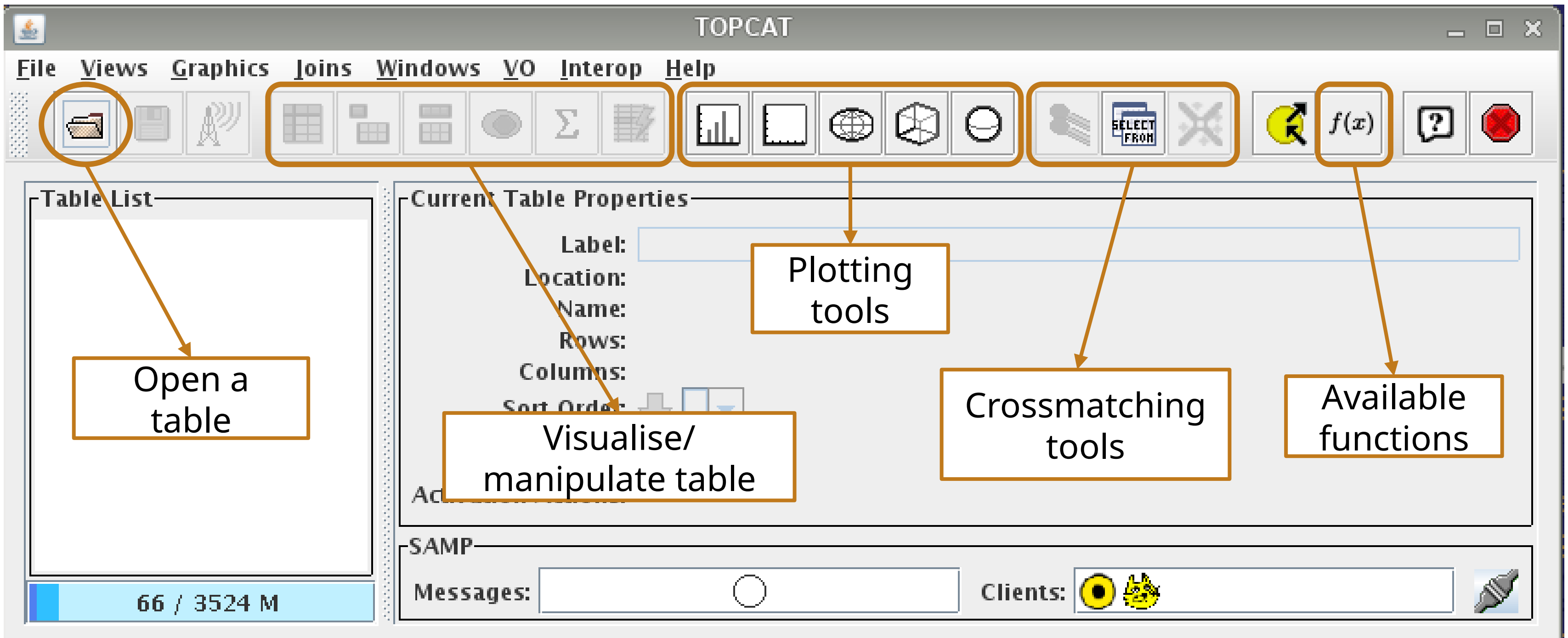


# TOPCAT – start window

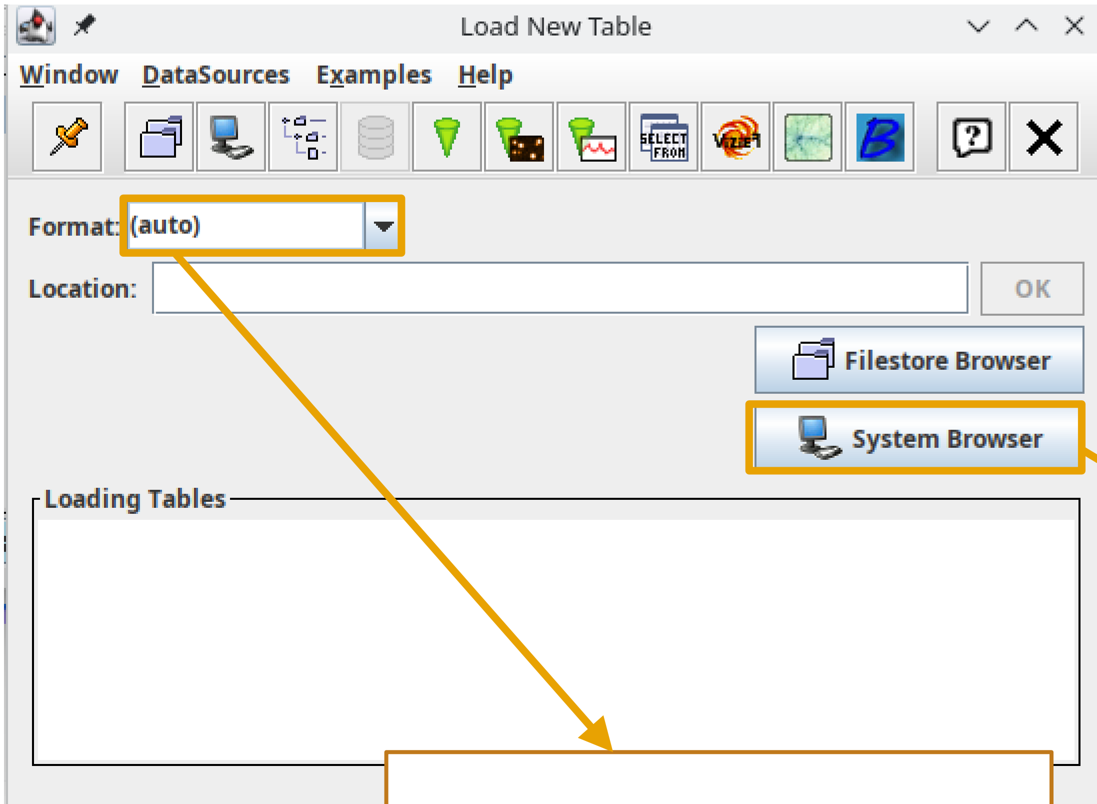
Most important button!



# TOPCAT – start window



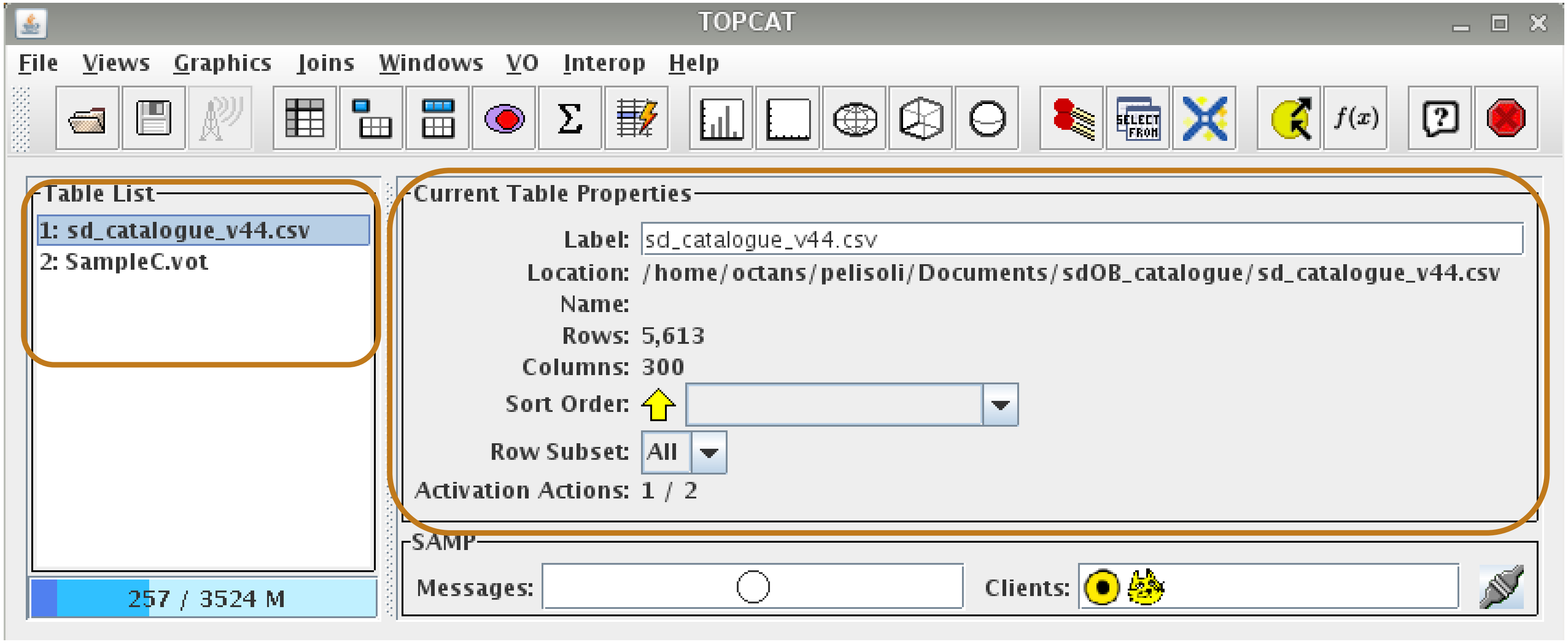
# TOPCAT – open a table



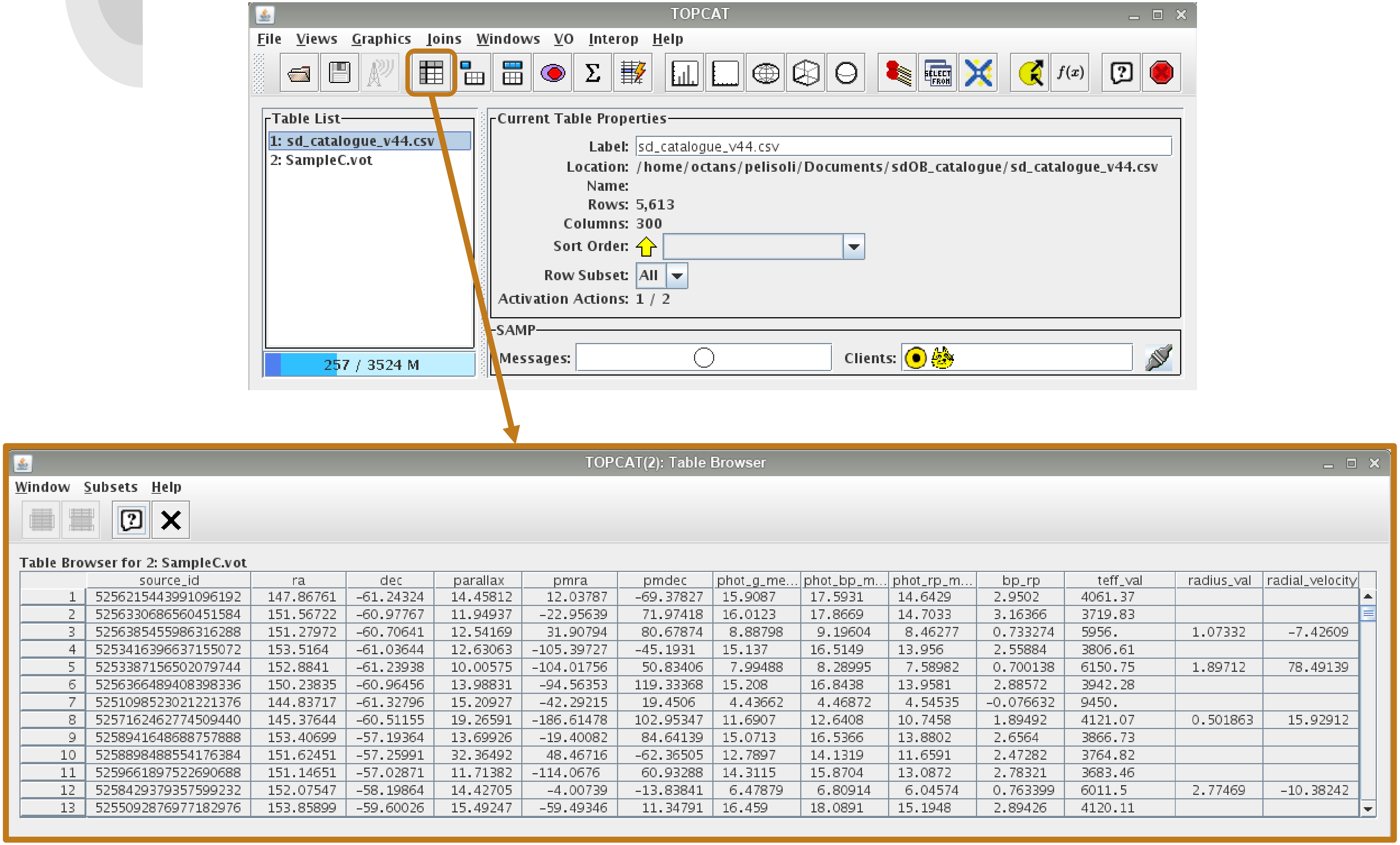

Select path where table is located, e.g. your share folder

You can keep “auto” here but it might fail sometimes and then you have to set the correct format of your table

# TOPCAT – tables

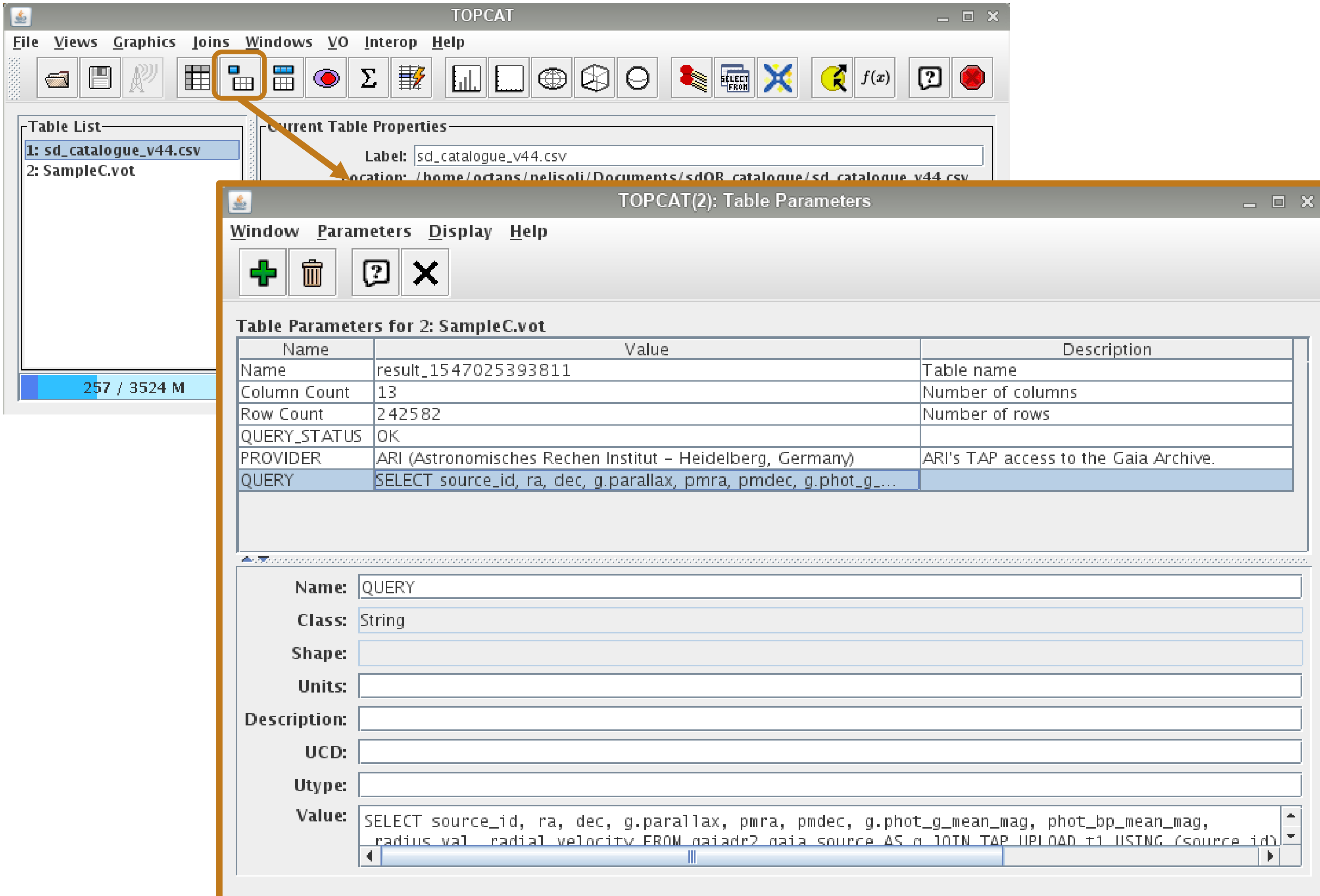


# TOPCAT – browse table entries

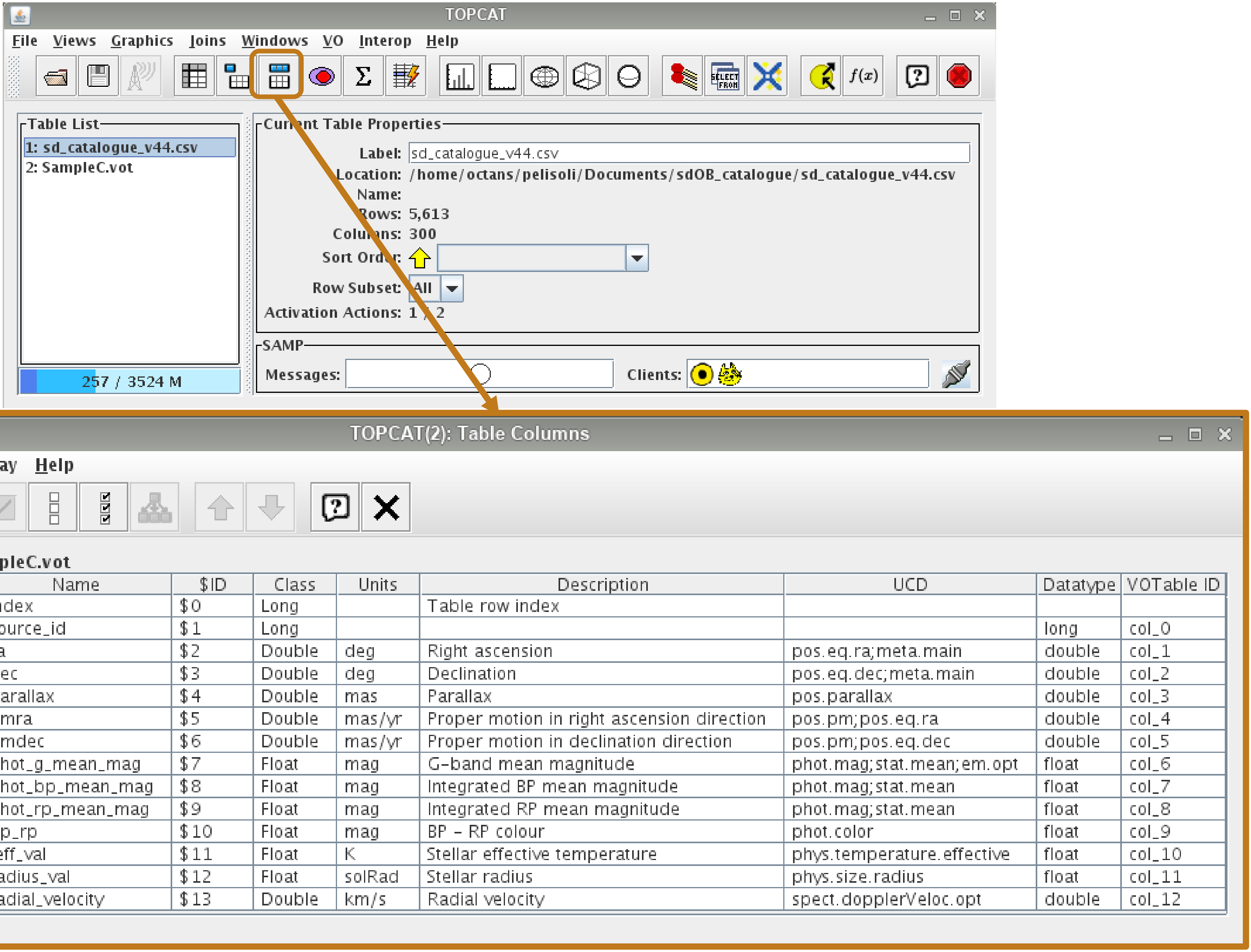



| | source_id | ra | dec | parallax | pmra | pmdec | phot_g_me... | phot_bp_m... | phot_rp_m... | bp_rp | teff_val | radius_val | radial_velocity |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5256215443991096192 | 147.86761 | -61.24324 | 14.45812 | 12.03787 | -69.37827 | 15.9087 | 17.5931 | 14.6429 | 2.9502 | 4061.37 | | |
| 2 | 5256330686560451584 | 151.56722 | -60.97767 | 11.94937 | -22.95639 | 71.97418 | 16.0123 | 17.8669 | 14.7033 | 3.16366 | 3719.83 | | |
| 3 | 5256385455986316288 | 151.27972 | -60.70641 | 12.54169 | 31.90794 | 80.67874 | 8.88798 | 9.19604 | 8.46277 | 0.733274 | 5956. | 1.07332 | -7.42609 |
| 4 | 5253416396637155072 | 153.5164 | -61.03644 | 12.63063 | -105.39727 | -45.1931 | 15.137 | 16.5149 | 13.956 | 2.55884 | 3806.61 | | |
| 5 | 5253387156502079744 | 152.8841 | -61.23938 | 10.00575 | -104.01756 | 50.83406 | 7.99488 | 8.28995 | 7.58982 | 0.700138 | 6150.75 | 1.89712 | 78.49139 |
| 6 | 5256366489408398336 | 150.23835 | -60.96456 | 13.98831 | -94.56353 | 119.33368 | 15.208 | 16.8438 | 13.9581 | 2.88572 | 3942.28 | | |
| 7 | 5251098523021221376 | 144.83717 | -61.32796 | 15.20927 | -42.29215 | 19.4506 | 4.43662 | 4.46872 | 4.54535 | -0.076632 | 9450. | | |
| 8 | 5257162462774509440 | 145.37644 | -60.51155 | 19.26591 | -186.61478 | 102.95347 | 11.6907 | 12.6408 | 10.7458 | 1.89492 | 4121.07 | 0.501863 | 15.92912 |
| 9 | 5258941648688757888 | 153.40699 | -57.19364 | 13.69926 | -19.40082 | 84.64139 | 15.0713 | 16.5366 | 13.8802 | 2.6564 | 3866.73 | | |
| 10 | 5258898488554176384 | 151.62451 | -57.25991 | 32.36492 | 48.46716 | -62.36505 | 12.7897 | 14.1319 | 11.6591 | 2.47282 | 3764.82 | | |
| 11 | 5259661897522690688 | 151.14651 | -57.02871 | 11.71382 | -114.0676 | 60.93288 | 14.3115 | 15.8704 | 13.0872 | 2.78321 | 3683.46 | | |
| 12 | 5258429379357599232 | 152.07547 | -58.19864 | 14.42705 | -4.00739 | -13.83841 | 6.47879 | 6.80914 | 6.04574 | 0.763399 | 6011.5 | 2.77469 | -10.38242 |
| 13 | 5255092876977182976 | 153.85899 | -59.60026 | 15.49247 | -59.49346 | 11.34791 | 16.459 | 18.0891 | 15.1948 | 2.89426 | 4120.11 | | |

# TOPCAT – table metadata

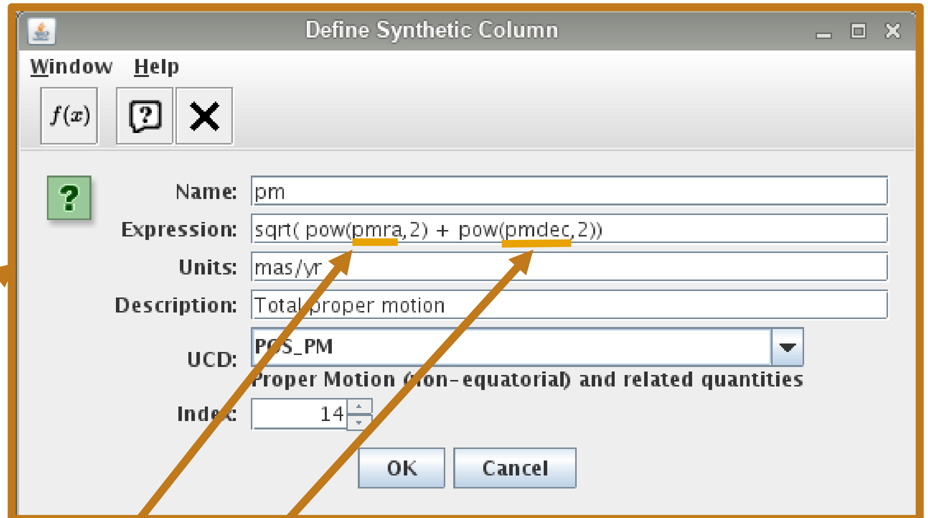


# TOPCAT – column metadata



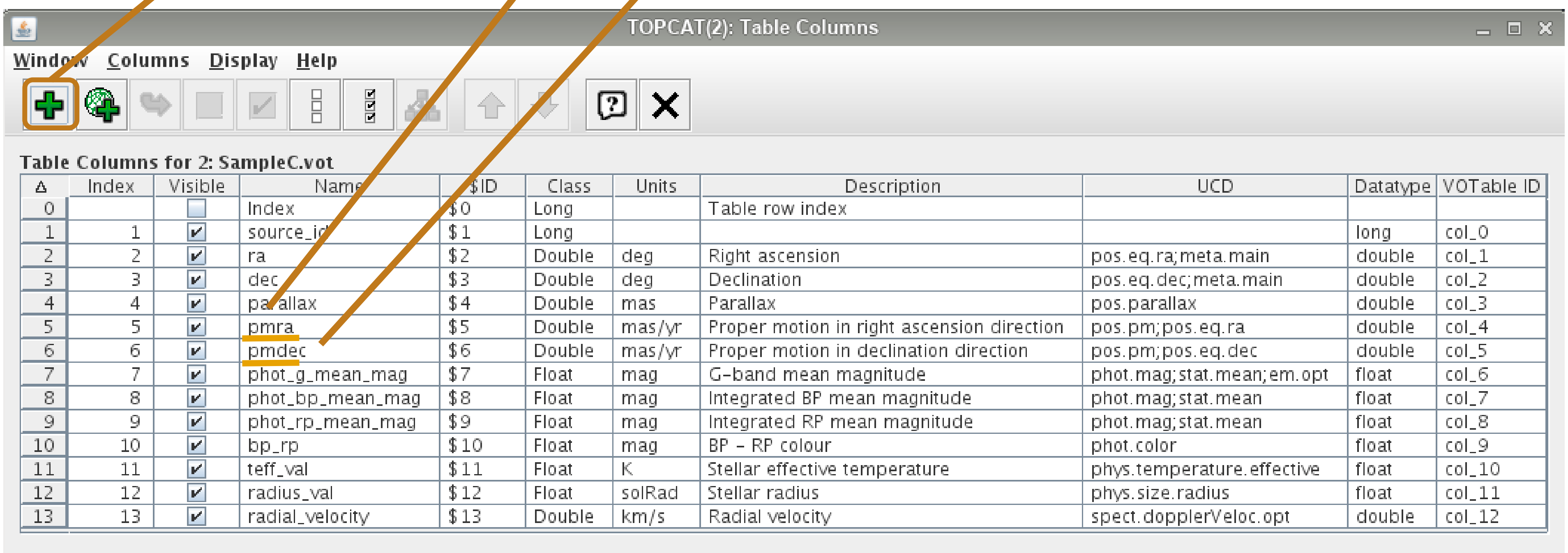

| Δ | Index | Visible | Name | $ID | Class | Units | Description | UCD | Datatype | VOTable ID |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | ☐ | Index | $0 | Long | | Table row index | | | |
| 1 | 1 | ☑ | source_id | $1 | Long | | | | long | col_0 |
| 2 | 2 | ☑ | ra | $2 | Double | deg | Right ascension | pos.eq.ra;meta.main | double | col_1 |
| 3 | 3 | ☑ | dec | $3 | Double | deg | Declination | pos.eq.dec;meta.main | double | col_2 |
| 4 | 4 | ☑ | parallax | $4 | Double | mas | Parallax | pos.parallax | double | col_3 |
| 5 | 5 | ☑ | pmra | $5 | Double | mas/yr | Proper motion in right ascension direction | pos.pm;pos.eq.ra | double | col_4 |
| 6 | 6 | ☑ | pmdec | $6 | Double | mas/yr | Proper motion in declination direction | pos.pm;pos.eq.dec | double | col_5 |
| 7 | 7 | ☑ | phot_g_mean_mag | $7 | Float | mag | G-band mean magnitude | phot.mag;stat.mean;em.opt | float | col_6 |
| 8 | 8 | ☑ | phot_bp_mean_mag | $8 | Float | mag | Integrated BP mean magnitude | phot.mag;stat.mean | float | col_7 |
| 9 | 9 | ☑ | phot_rp_mean_mag | $9 | Float | mag | Integrated RP mean magnitude | phot.mag;stat.mean | float | col_8 |
| 10 | 10 | ☑ | bp_rp | $10 | Float | mag | BP - RP colour | phot.color | float | col_9 |
| 11 | 11 | ☑ | teff_val | $11 | Float | K | Stellar effective temperature | phys.temperature.effective | float | col_10 |
| 12 | 12 | ☑ | radius_val | $12 | Float | solRad | Stellar radius | phys.size.radius | float | col_11 |
| 13 | 13 | ☑ | radial_velocity | $13 | Double | km/s | Radial velocity | spect.dopplerVeloc.opt | double | col_12 |

# TOPCAT – create new column




Column names become variables

Column names should not contain mathematical operators and spaces!

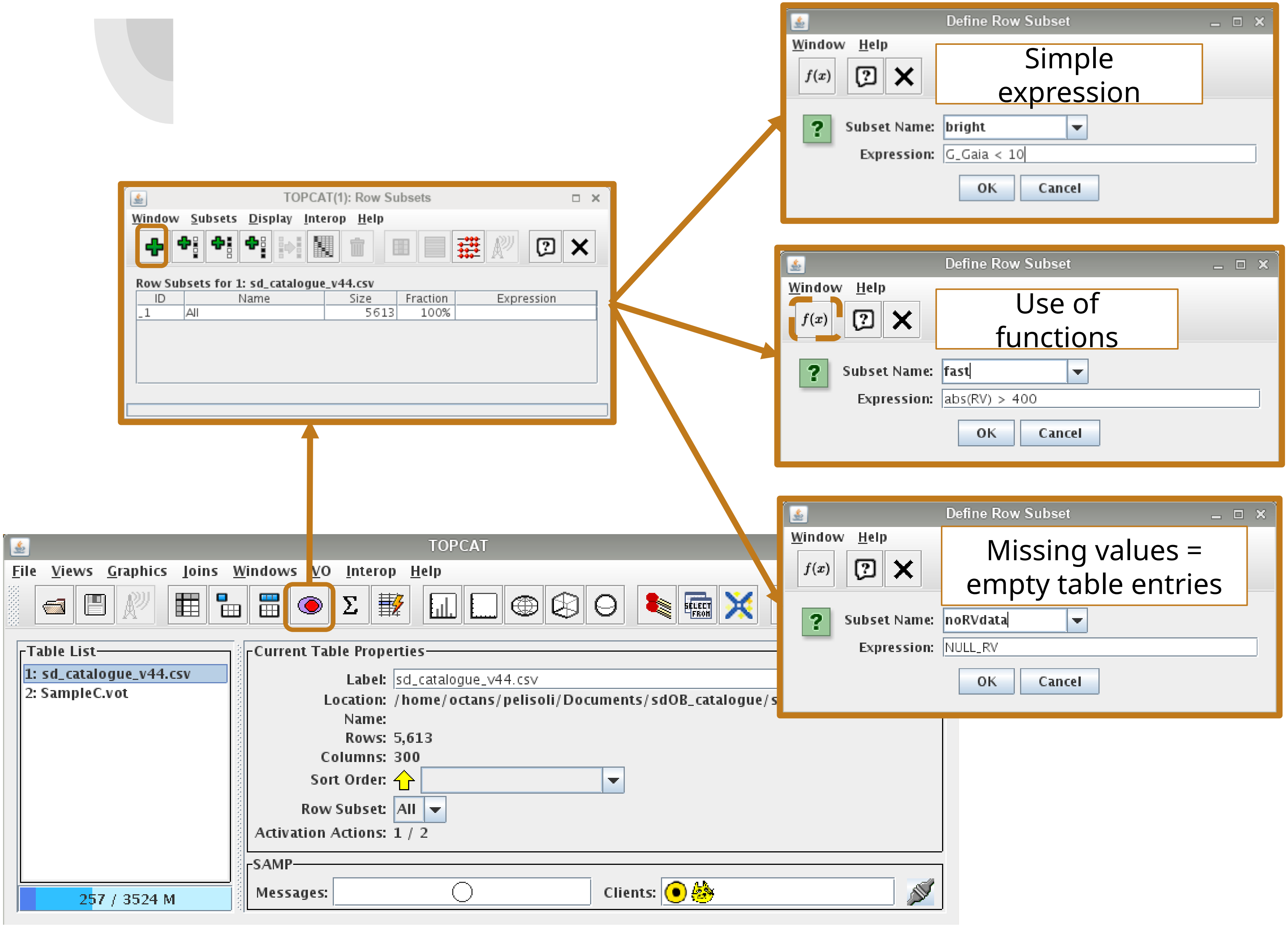



| Δ | Index | Visible | Name | $ID | Class | Units | Description | UCD | Datatype | VOTable ID |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | ☐ | Index | $0 | Long | | Table row index | | | |
| 1 | 1 | ☑ | source_id | $1 | Long | | | | long | col_0 |
| 2 | 2 | ☑ | ra | $2 | Double | deg | Right ascension | pos.eq.ra;meta.main | double | col_1 |
| 3 | 3 | ☑ | dec | $3 | Double | deg | Declination | pos.eq.dec;meta.main | double | col_2 |
| 4 | 4 | ☑ | parallax | $4 | Double | mas | Parallax | pos.parallax | double | col_3 |
| 5 | 5 | ☑ | pmra | $5 | Double | mas/yr | Proper motion in right ascension direction | pos.pm;pos.eq.ra | double | col_4 |
| 6 | 6 | ☑ | pmdec | $6 | Double | mas/yr | Proper motion in declination direction | pos.pm;pos.eq.dec | double | col_5 |
| 7 | 7 | ☑ | phot_g_mean_mag | $7 | Float | mag | G-band mean magnitude | phot.mag;stat.mean;em.opt | float | col_6 |
| 8 | 8 | ☑ | phot_bp_mean_mag | $8 | Float | mag | Integrated BP mean magnitude | phot.mag;stat.mean | float | col_7 |
| 9 | 9 | ☑ | phot_rp_mean_mag | $9 | Float | mag | Integrated RP mean magnitude | phot.mag;stat.mean | float | col_8 |
| 10 | 10 | ☑ | bp_rp | $10 | Float | mag | BP - RP colour | phot.color | float | col_9 |
| 11 | 11 | ☑ | teff_val | $11 | Float | K | Stellar effective temperature | phys.temperature.effective | float | col_10 |
| 12 | 12 | ☑ | radius_val | $12 | Float | solRad | Stellar radius | phys.size.radius | float | col_11 |
| 13 | 13 | ☑ | radial_velocity | $13 | Double | km/s | Radial velocity | spect.dopplerVeloc.opt | double | col_12 |

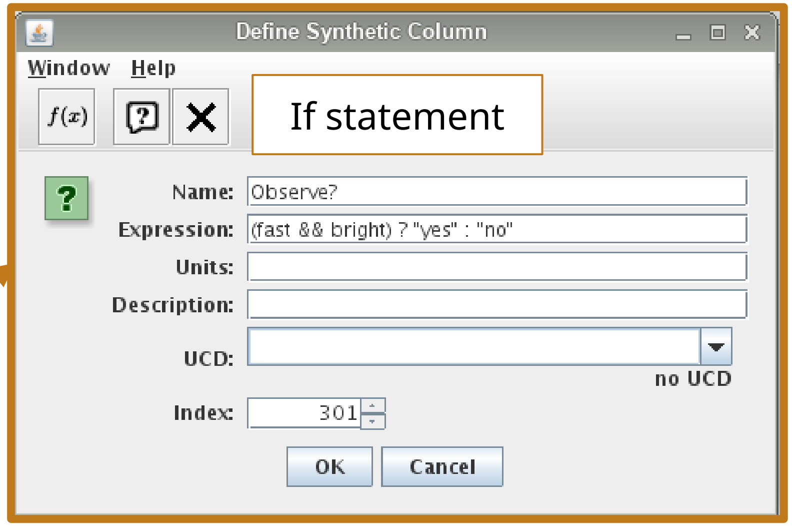
# TOPCAT – most important math operators

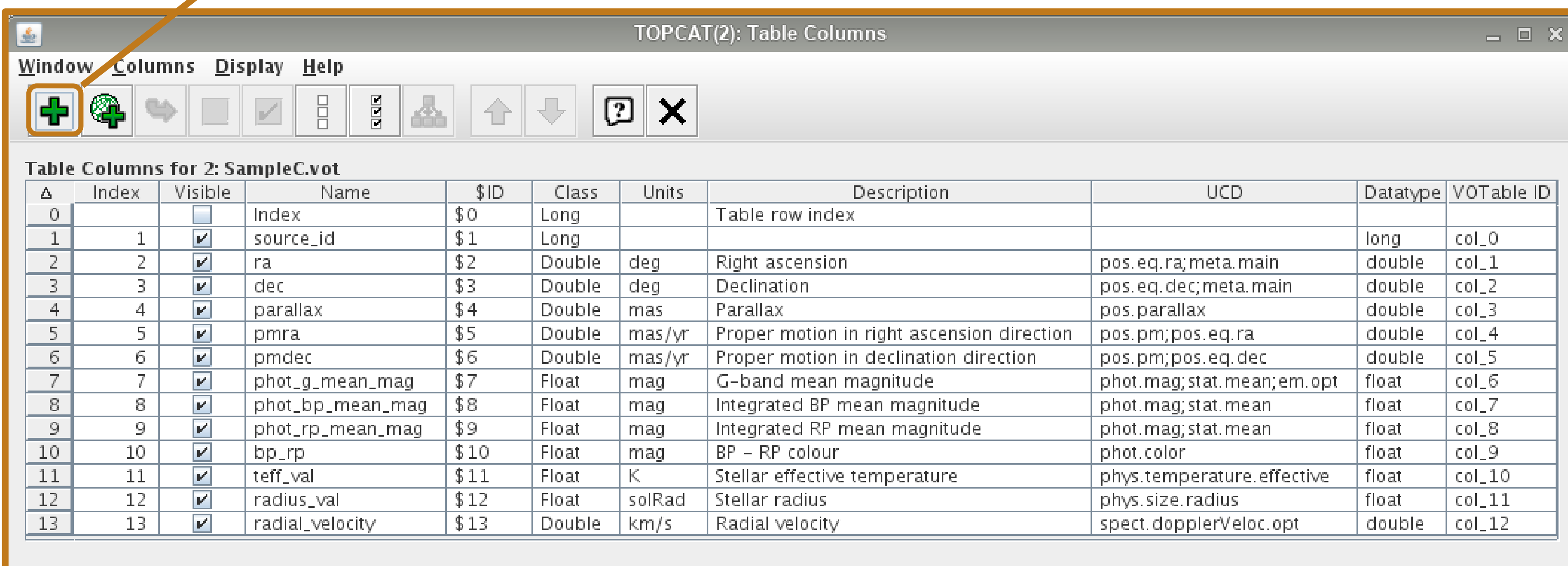
| Real | Expression in TOPCAT |
| --- | --- |
| + | + |
| - | - |
| * | * |
| / | / |
| log(5) | log10(5) |
| $\sqrt{5}$ | sqrt(5) |
| $10^5$ | pow(10,5) |

# TOPCAT – create subsets



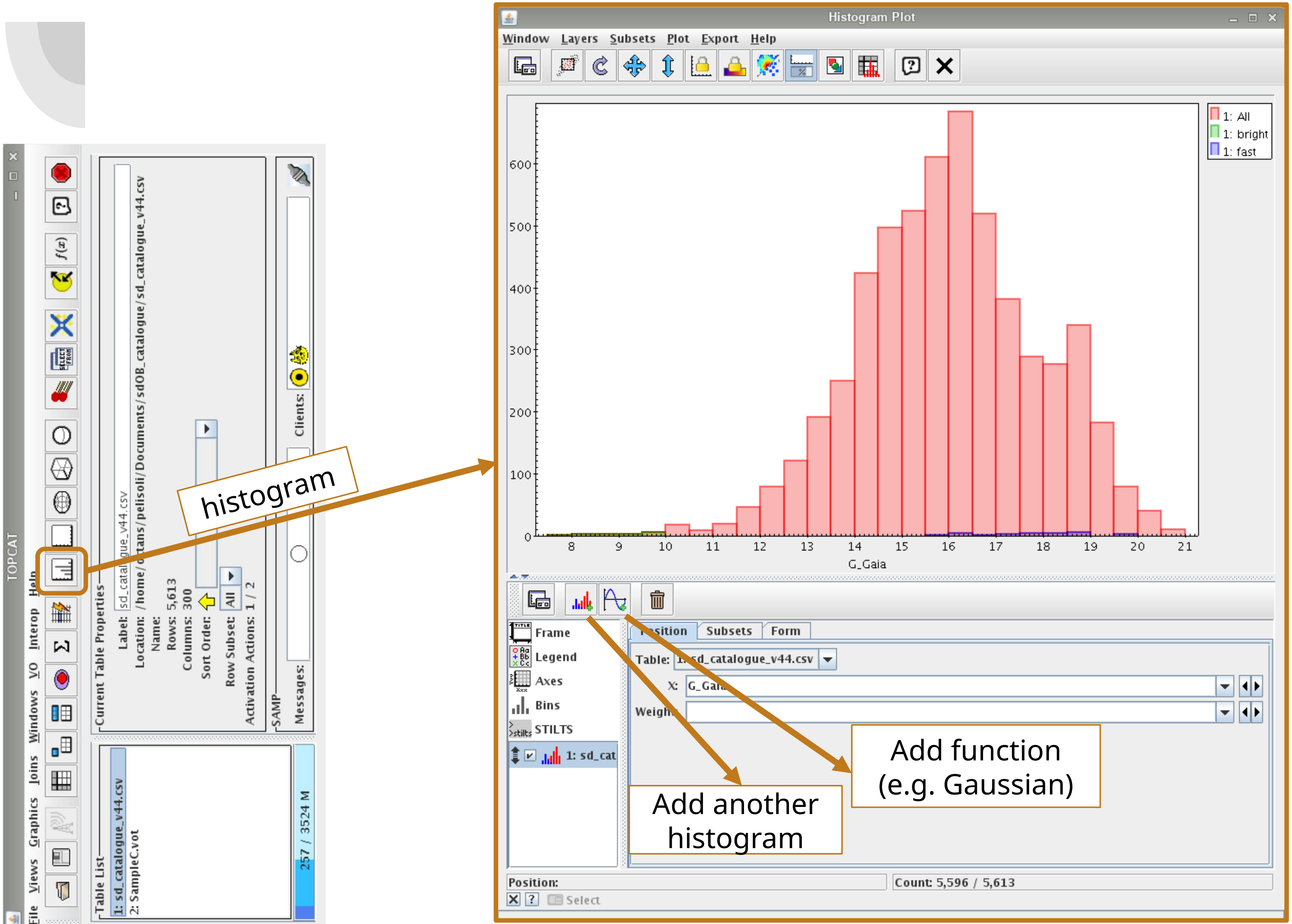
# TOPCAT – create column based on subset



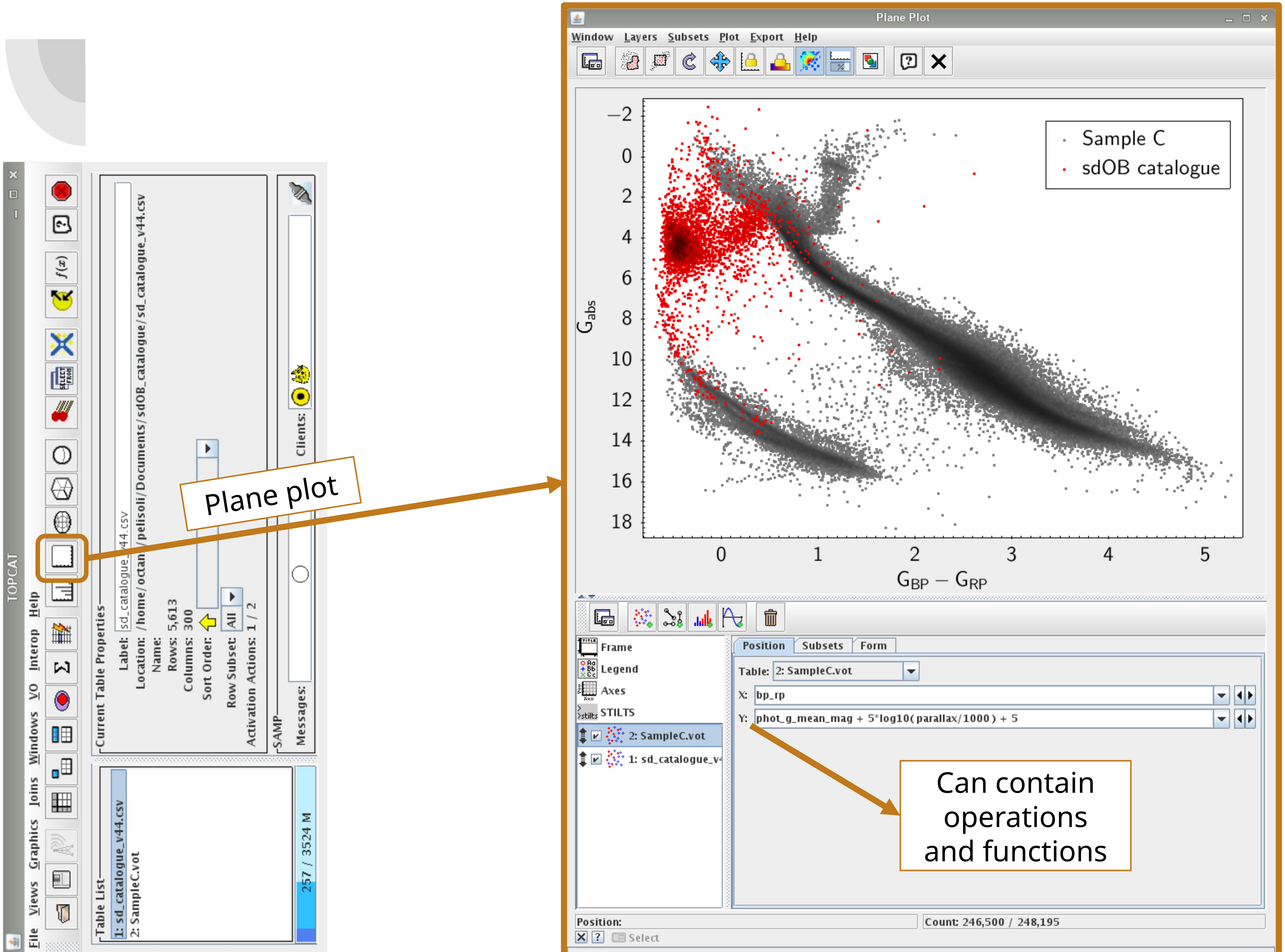




| Δ | Index | Visible | Name | $ID | Class | Units | Description | UCD | Datatype | VOTable ID |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | ☐ | Index | $0 | Long | | Table row index | | | |
| 1 | 1 | ☑ | source_id | $1 | Long | | | | long | col_0 |
| 2 | 2 | ☑ | ra | $2 | Double | deg | Right ascension | pos.eq.ra;meta.main | double | col_1 |
| 3 | 3 | ☑ | dec | $3 | Double | deg | Declination | pos.eq.dec;meta.main | double | col_2 |
| 4 | 4 | ☑ | parallax | $4 | Double | mas | Parallax | pos.parallax | double | col_3 |
| 5 | 5 | ☑ | pmra | $5 | Double | mas/yr | Proper motion in right ascension direction | pos.pm;pos.eq.ra | double | col_4 |
| 6 | 6 | ☑ | pmdec | $6 | Double | mas/yr | Proper motion in declination direction | pos.pm;pos.eq.dec | double | col_5 |
| 7 | 7 | ☑ | phot_g_mean_mag | $7 | Float | mag | G-band mean magnitude | phot.mag;stat.mean;em.opt | float | col_6 |
| 8 | 8 | ☑ | phot_bp_mean_mag | $8 | Float | mag | Integrated BP mean magnitude | phot.mag;stat.mean | float | col_7 |
| 9 | 9 | ☑ | phot_rp_mean_mag | $9 | Float | mag | Integrated RP mean magnitude | phot.mag;stat.mean | float | col_8 |
| 10 | 10 | ☑ | bp_rp | $10 | Float | mag | BP - RP colour | phot.color | float | col_9 |
| 11 | 11 | ☑ | teff_val | $11 | Float | K | Stellar effective temperature | phys.temperature.effective | float | col_10 |
| 12 | 12 | ☑ | radius_val | $12 | Float | solRad | Stellar radius | phys.size.radius | float | col_11 |
| 13 | 13 | ☑ | radial_velocity | $13 | Double | km/s | Radial velocity | spect.dopplerVeloc.opt | double | col_12 |

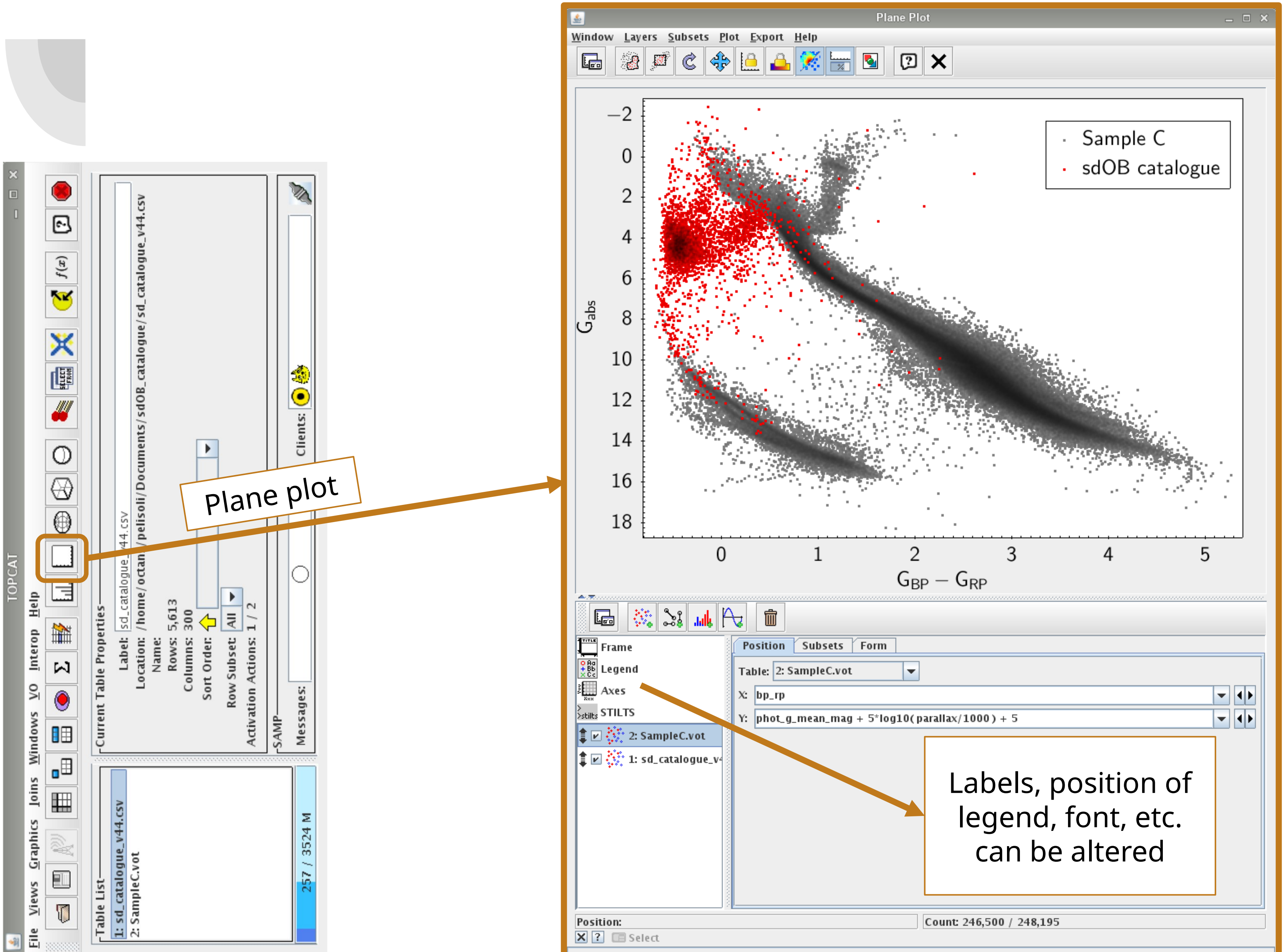
# TOPCAT – Visualisation tools



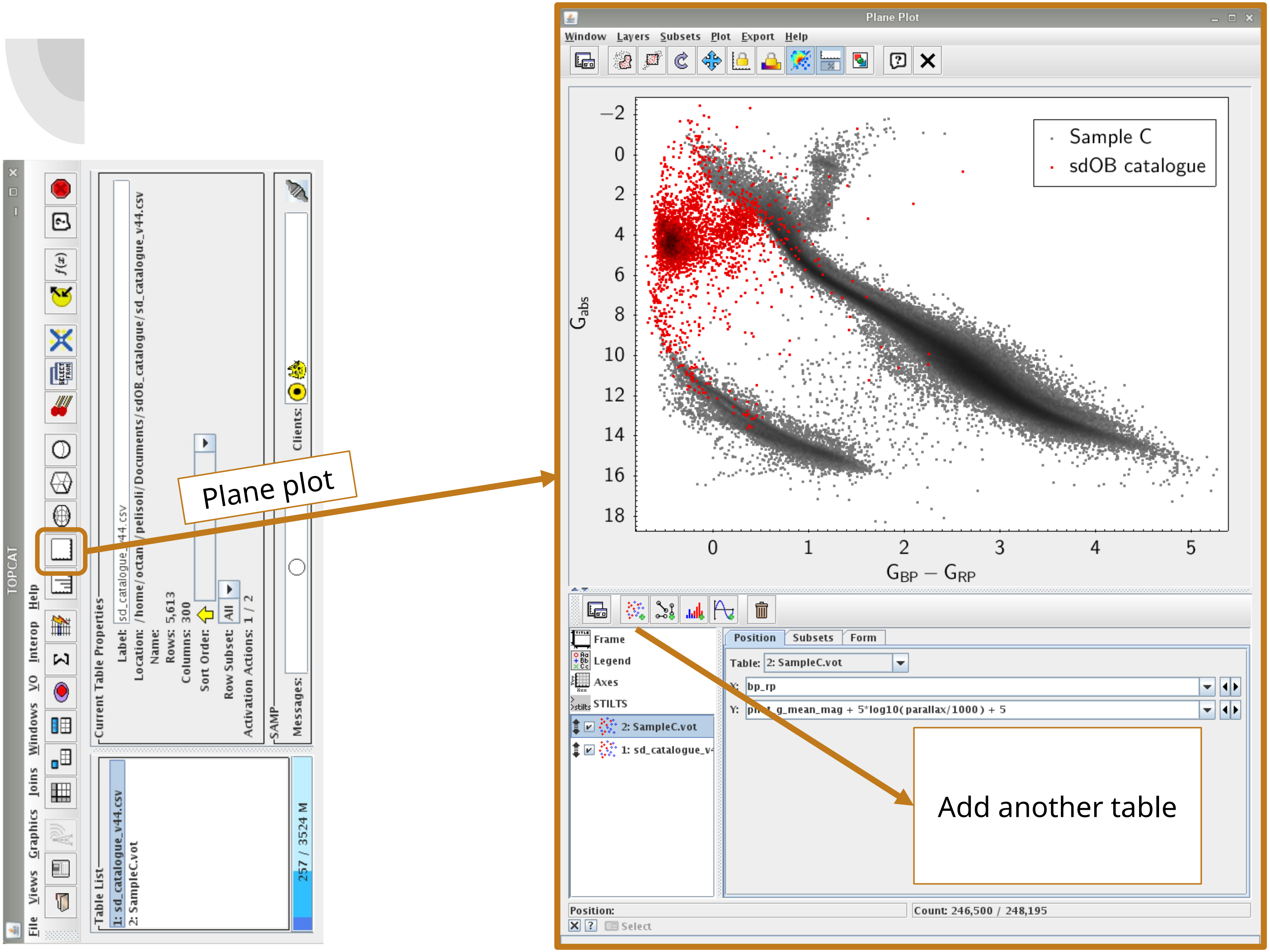
# TOPCAT – Visualisation tools



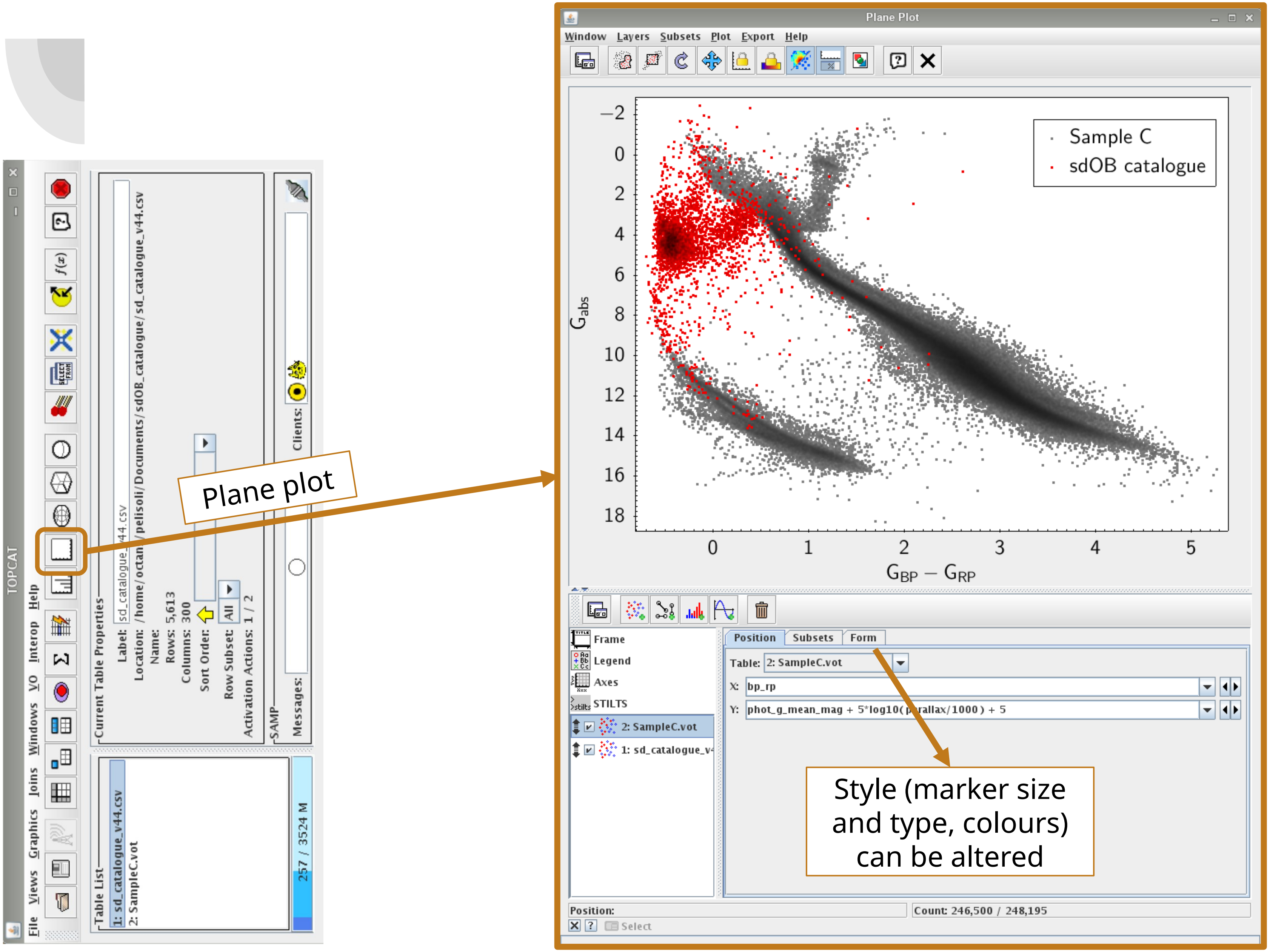
# TOPCAT – Visualisation tools



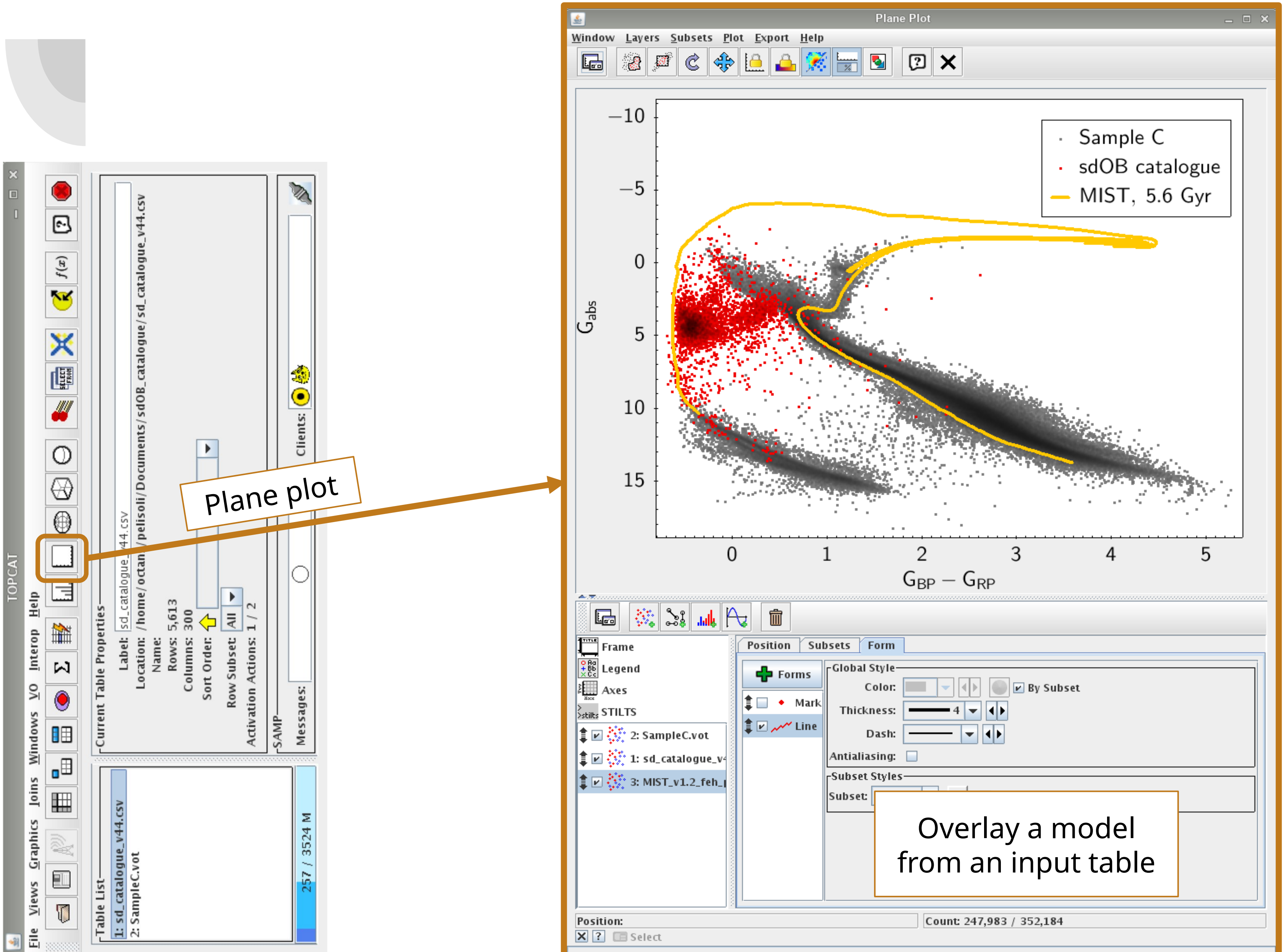
# TOPCAT – Visualisation tools



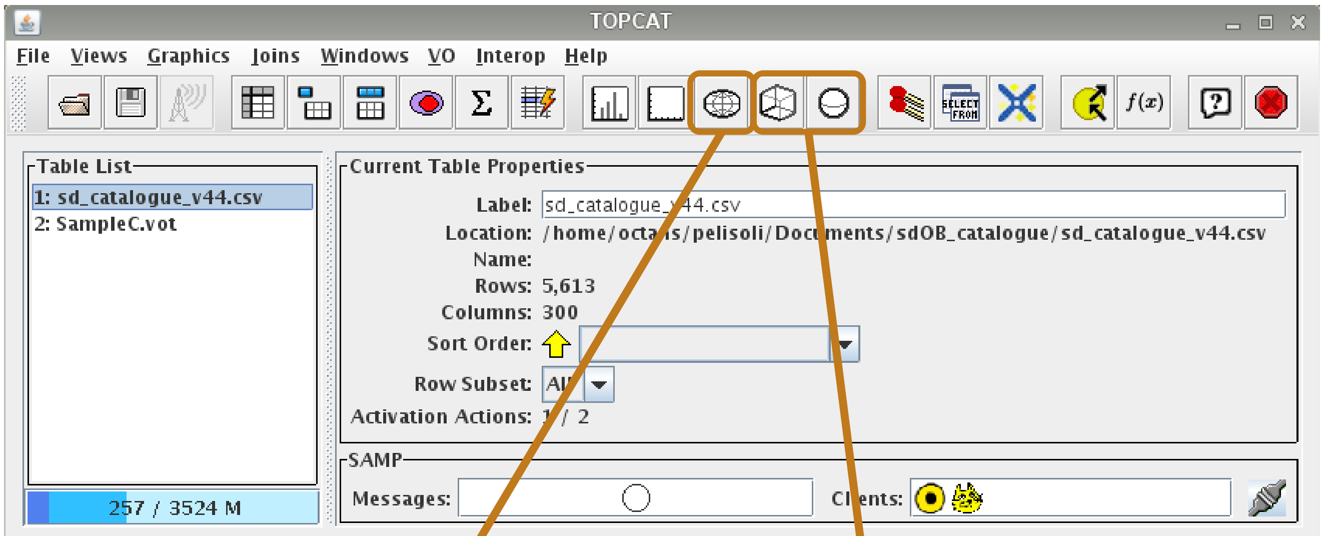
# TOPCAT – Visualisation tools



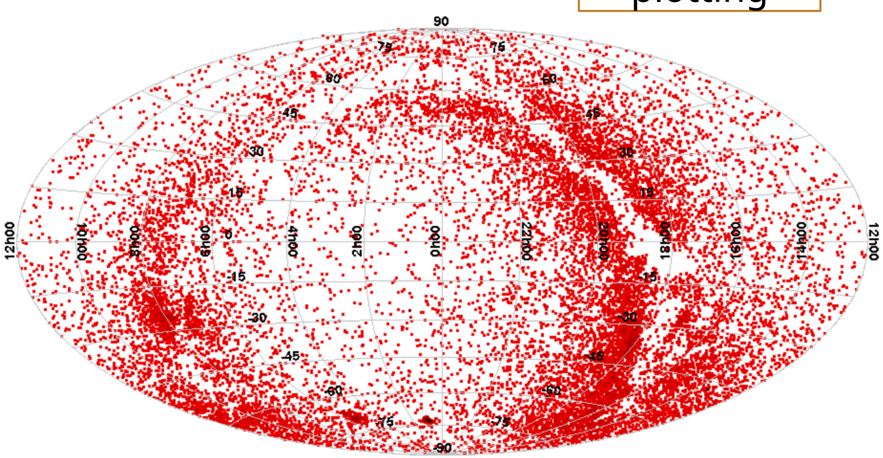
# TOPCAT – Visualisation tools



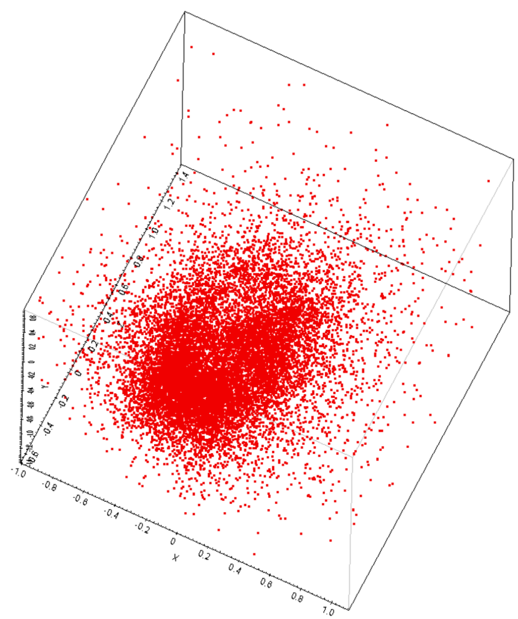
# TOPCAT – Visualisation tools



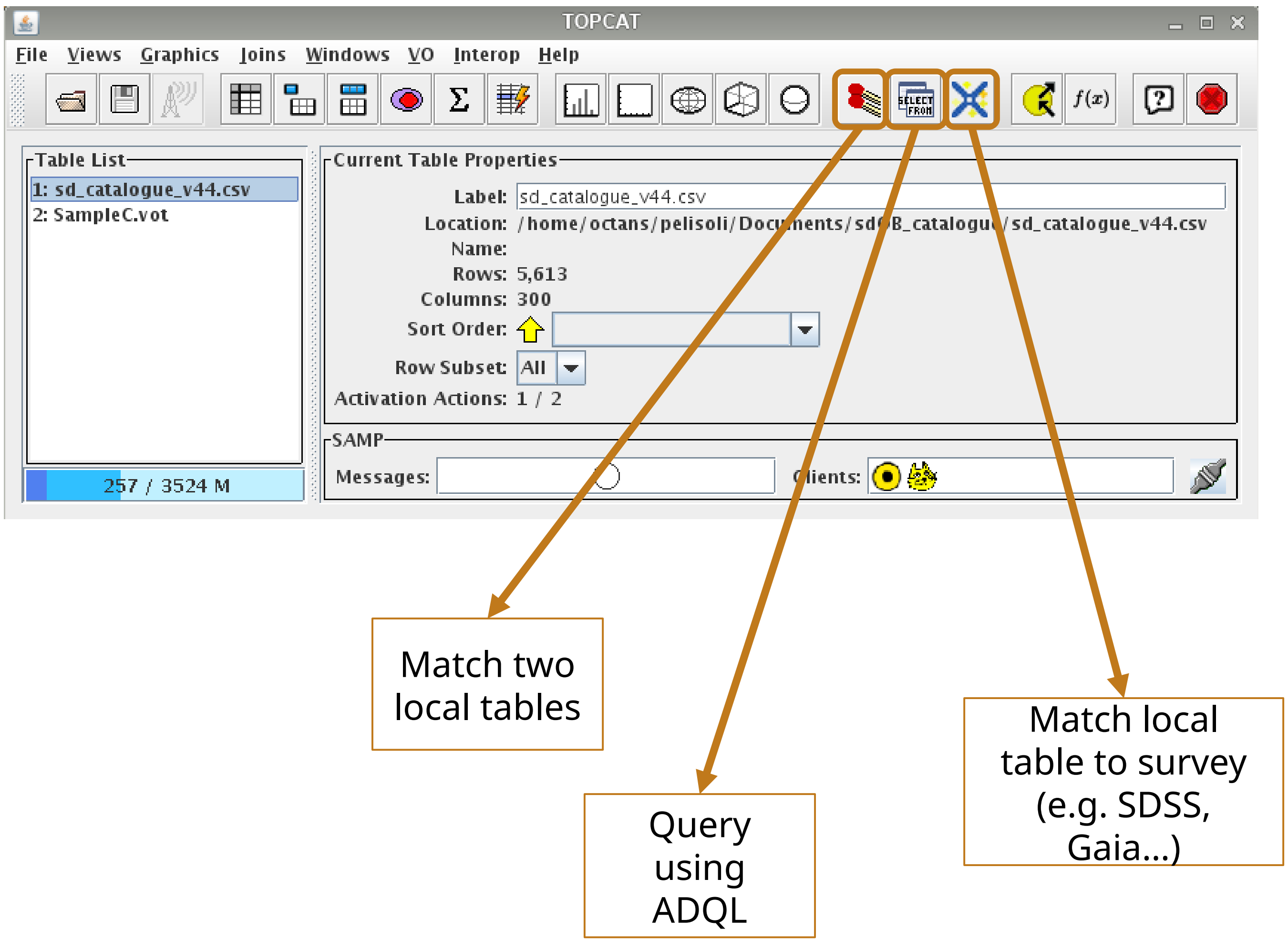

Sky plotting



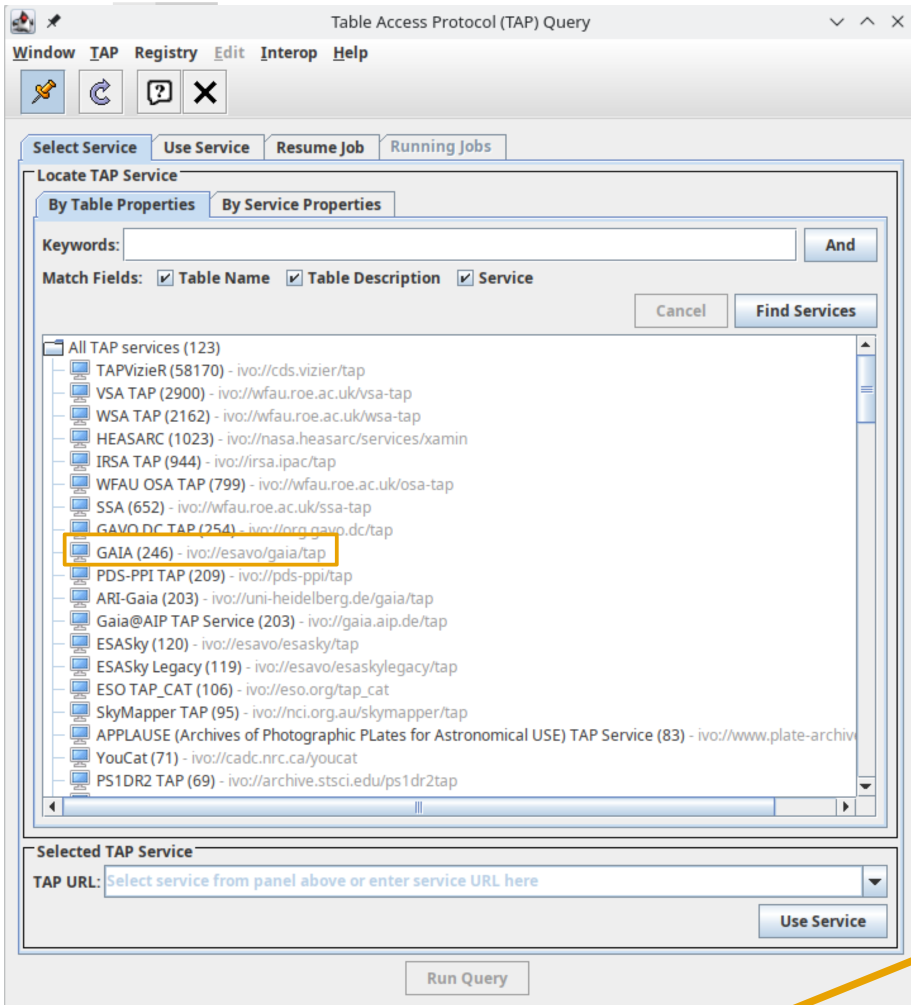
3D plotting

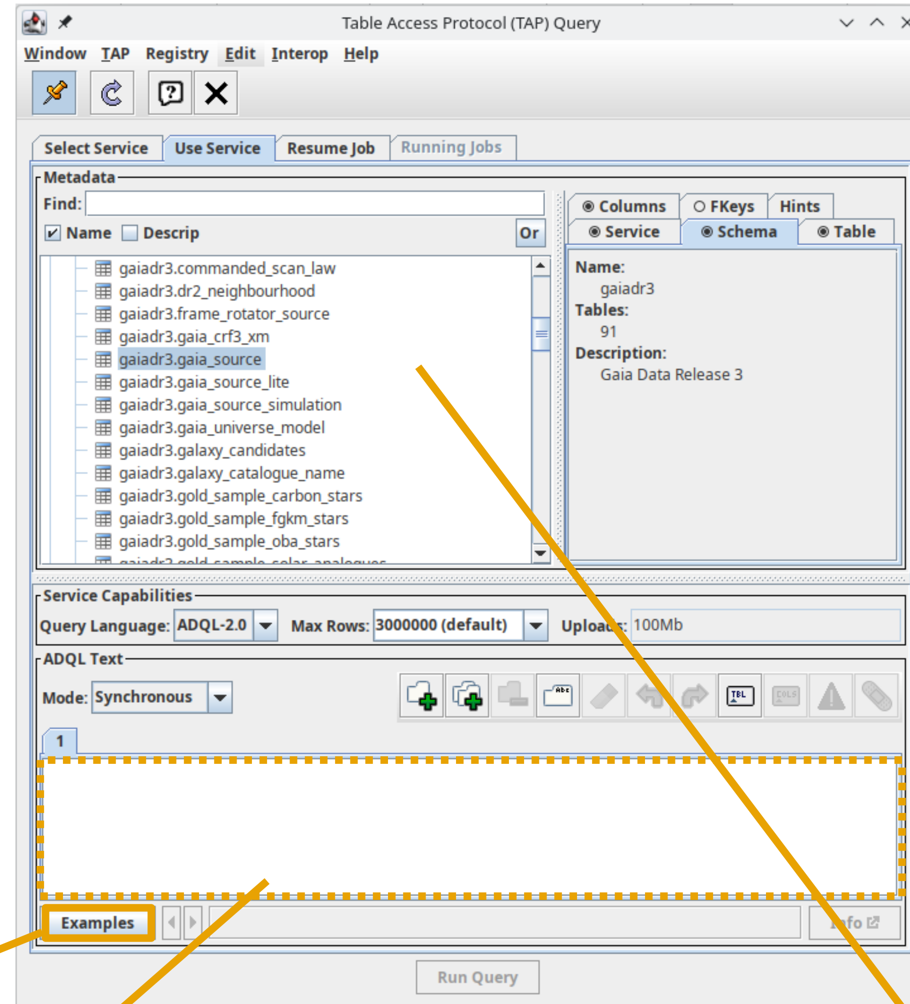


# TOPCAT – Crossmatching



# ADQL queries



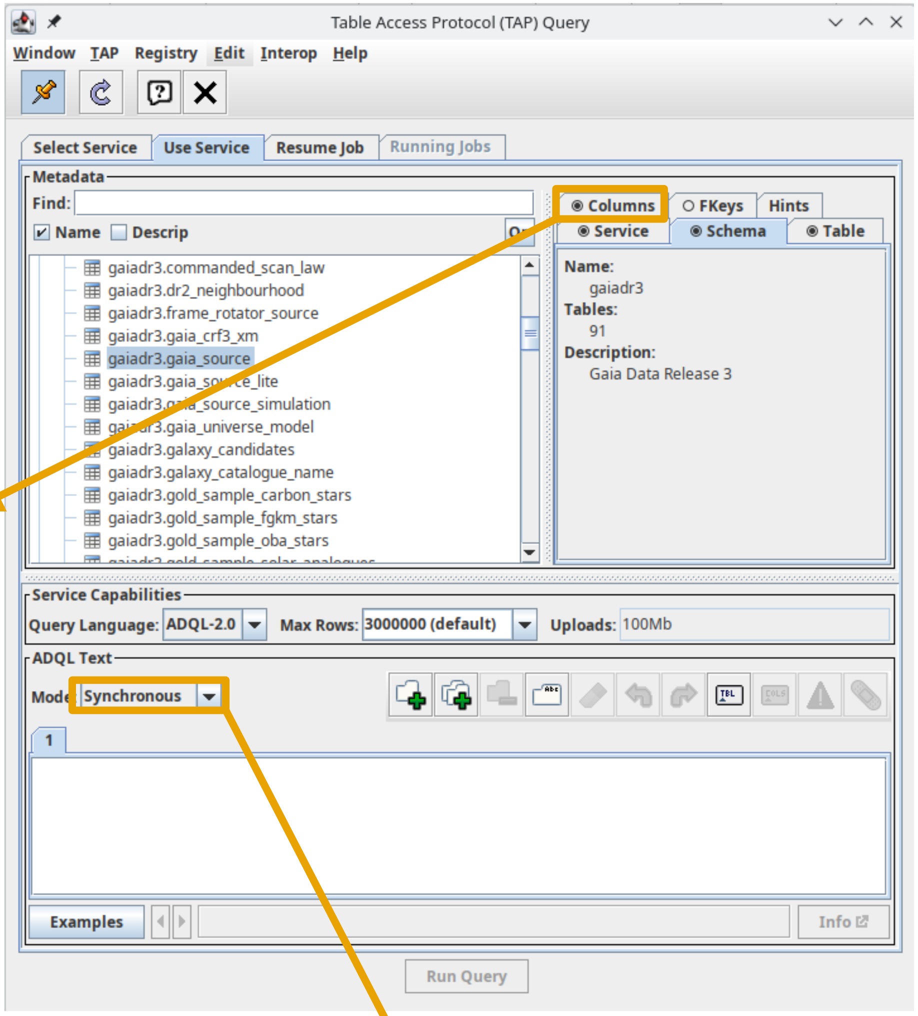




Provides simple queries that can be adapted

Here goes query

Spectroscopy group will use this catalogue

# ADQL queries




Select this to see names of columns

Set to Asynchronus if query times out

# ADQL queries

- ADQL = Astronomical Data Query Language
- Useful tutorial http://docs.g-vo.org/adql-gaia/html/
- A dialect of SQL

Column names

Very basic summary of a query:

```
SELECT [TOP¹ (number of rows)] [source table index].(variables you need)
FROM (table you're querying) [AS (table index)]
[WHERE (condition 1) AND (condition 2) OR (condition 3)]
[ORDER BY (variable)]
```

[1] Instead of TOP a * can be used to select everything that fulfills criteria or until maximum number of entries is reached

# ADQL queries – SELECT: ORDER BY

- Useful to select brightest, fastest, etc. from a table
- E.g.: 50 brightest stars in Gaia DR2

- E.g.: 20 highest proper motion stars in Tycho

# ADQL queries – SELECT: ORDER BY

- Useful to select brightest, fastest, etc. from a table
- E.g.: 50 brightest stars in Gaia DR2

```
SELECT TOP 50 source_id, phot_g_mean_mag, parallax, bp_rp
FROM gaiadr3.gaia_source
ORDER BY phot_g_mean_mag
```

- E.g.: 20 highest proper motion stars in Tycho

```
SELECT TOP 20 source_id, parallax, phot_g_mean_mag,
              SQRT(POW(pmra,2)+POW(pmdec,2)) AS pm
FROM gaiadr1.tgas_source
ORDER BY pm DESC
```

# ADQL queries – SELECT: WHERE clause

- WHERE introduces a logical expression, in a similar way to other languages, with operators AND and OR.
- E.g.: stars brighter than 12, closer than 50 pc.

# ADQL queries – SELECT: WHERE clause

- WHERE introduces a logical expression, in a similar way to other languages, with operators AND and OR.
- E.g.: stars brighter than 12, closer than 50 pc.

```
SELECT source_id, phot_g_mean_mag, parallax, bp_rp
FROM gaiadr3.gaia_source
WHERE phot_g_mean_mag < 12.0 AND parallax > 20.0
```

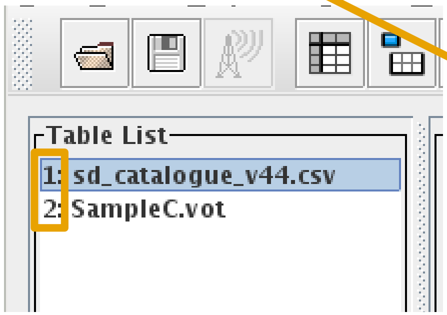
# ADQL queries – SELECT: JOIN USING

- For joining two tables with a same column
- E.g.: get Gaia DR2 proper motions for stars with known `source_id`

# ADQL queries – SELECT: JOIN USING

- For joining two tables with a same column
- E.g.: get Gaia DR2 proper motions for stars with known `source_id`

```
SELECT source_id, a.phot_g_mean_mag, a.parallax,
       a.bp_rp, b.pmra, b.pmdec
FROM TAP_UPLOAD.t6 AS a
JOIN gaiadr3.gaia_source AS b USING(source_id)
```




You can find this number in your table list

# ADQL queries – Exercise: variable sources in Gaia

Select variable sources in Gaia and crossmatch the result with the catalogue of known hot subluminous stars.

- Variability index: $$V_G = \frac{\sigma_G}{\langle G \rangle} \times \sqrt{n_{\mathrm{obs,G}}}$$
- phot_g_mean_flux_error/phot_g_mean_flux*sqrt(phot_g_n_obs)

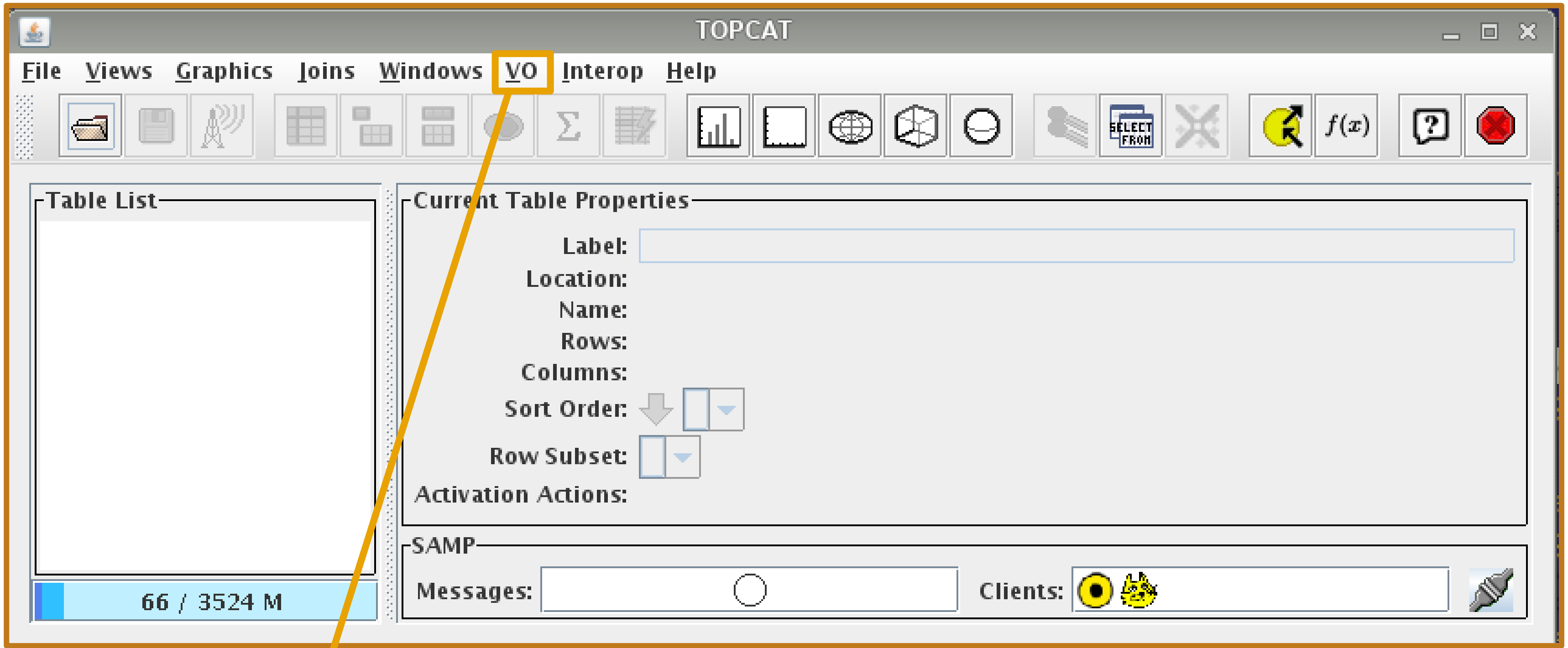
1. Select all stars that fulfill the following conditions:

```
Parallax >= 0.1
parallax_error/parallax<=0.3
phot_g_mean_mag<16
bp_rp<0
Varindex>0.1
```

Should result in 109 objects

# ADQL queries – Excersise: variable sources in Gaia

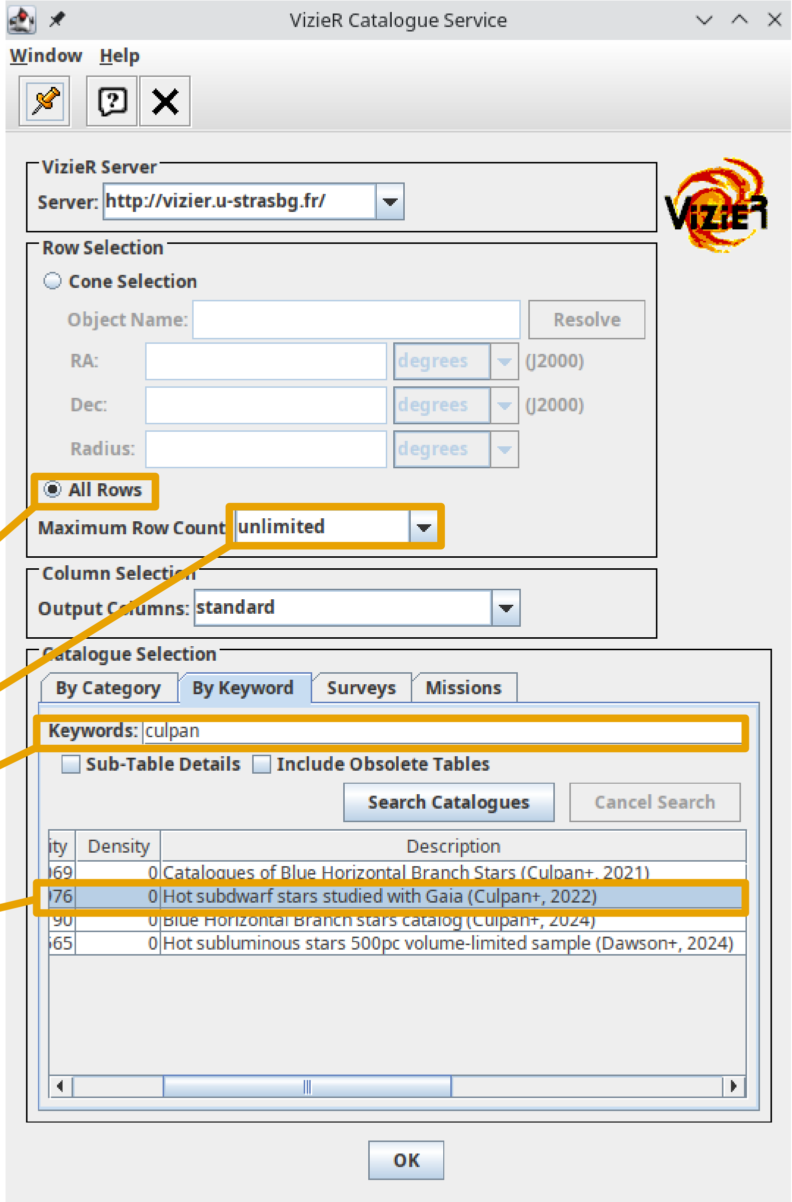
2. Download the catalogue of hot subluminous stars and crossmatch this with the table from the query




Select VizieR catalogue service

# ADQL queries – Excersise: variable sources in Gaia

This will give you two catalogues: _knownhsd and _hotsd. Use _knownhsd for the crossmatch

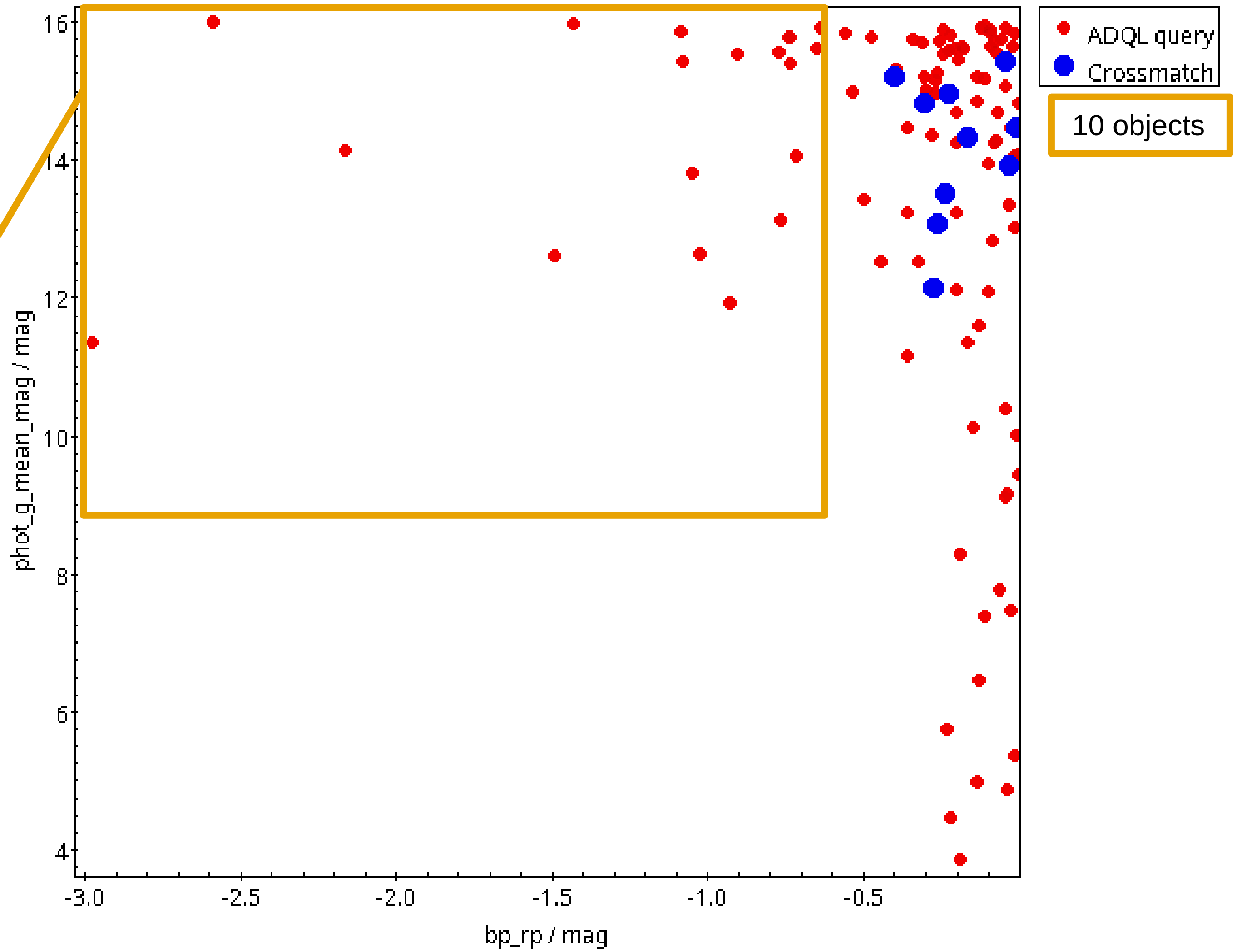



Select these settings

# ADQL queries – Excersise: variable sources in Gaia

3. Plot a Colour Magnitude Diagram (CMD) of the stars in the ADQL query table and the targets from the crossmatch. Use the columns bp_rp and phot_g_mean_mag

# ADQL queries – Excersise: variable sources in Gaia




10 objects

Are these reasonable?

# Now create the target lists for your projects

# Spectroscopic selection

Selection criteria

Parallax>=0.1
Absolute magnitude<=3.7
bp_rp<=0.05
phot_g_mean_mag<=11
Tangential velocity>=100

Quality cuts

phot_bp_rp_excess_factor>=1.0+0.015*bp_rp$^2$
visibility_periods_used>=8
phot_g_mean_flux_over_error>=50
phot_bp_mean_flux_over_error>=20
phot_rp_mean_flux_over_error>=20
4.74*pm/(parallax+parallax_error)>=30
Ruwe<=1.4