

Data mining with Topcat and ADQL

Creating a target list

Ingrid Pelisoli
Research workshop on evolved stars
04.09.2019

Overview



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



Tool for Operations on Catalogues And Tables

Does what you want with tables

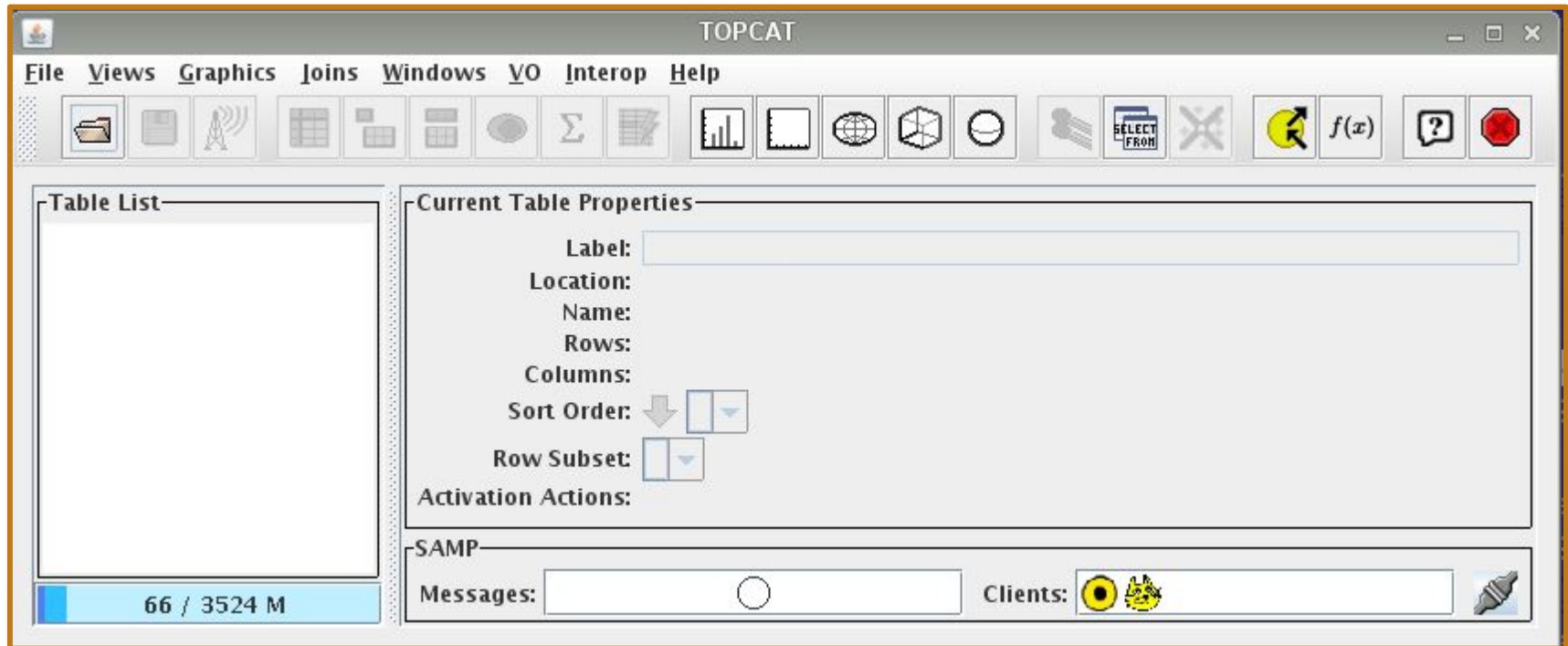
- Website: <http://www.star.bristol.ac.uk/~mbt/topcat/>
- Manual: <http://www.starlink.ac.uk/topcat/sun253/>
- Why TOPCAT?
 - Easy to use
 - Easy to learn
 - Easy to investigate data — good for exploratory analysis
 - Simple things obvious, complicated things documented
 - Easy to install and run
 -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



TOPCAT

- 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
 - Trigger some event when a row is selected
 - Talk to other astro tools (SAMP)

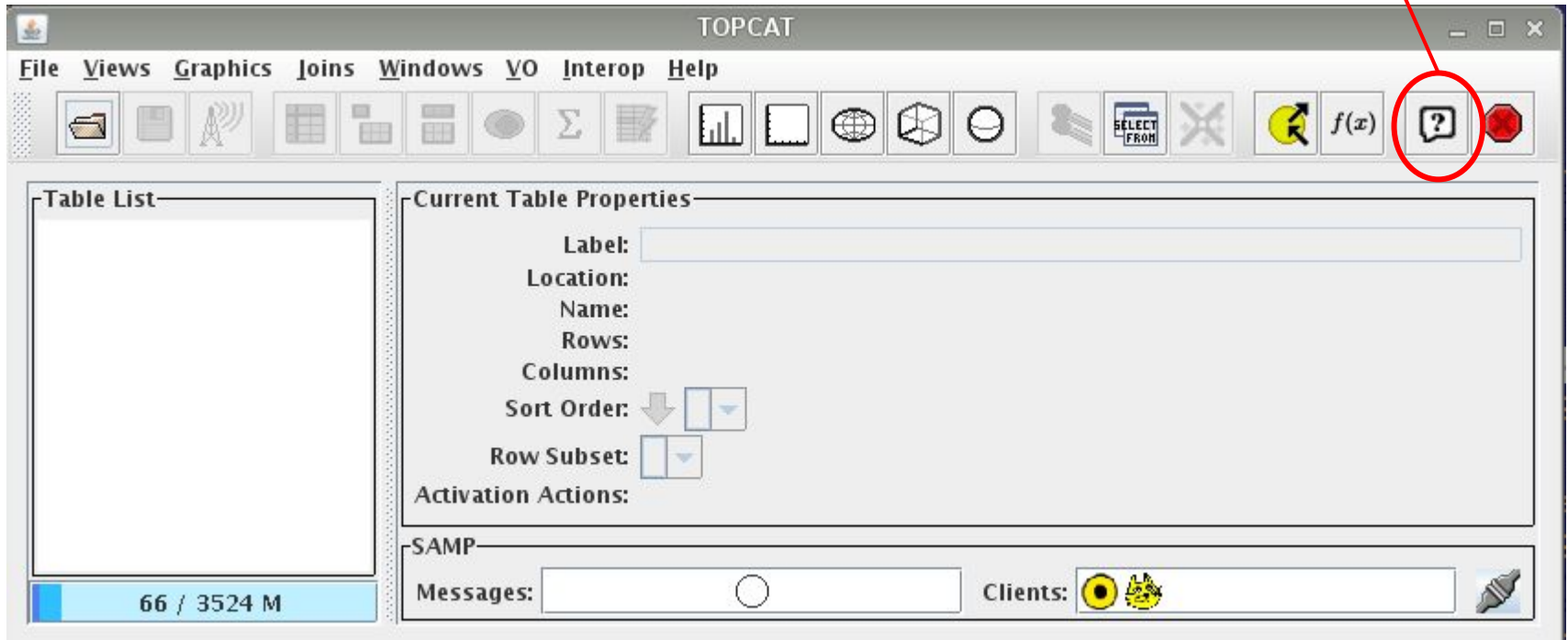
TOPCAT – start window



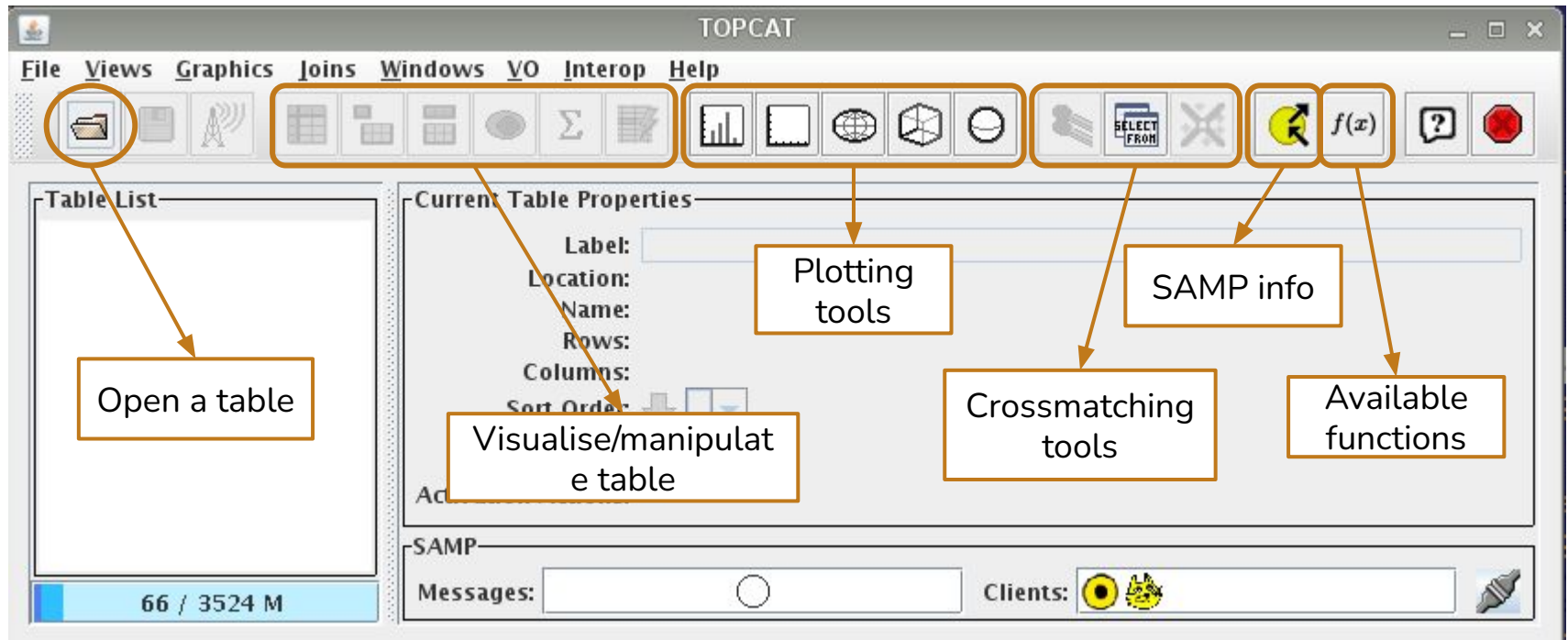
TOPCAT – start window



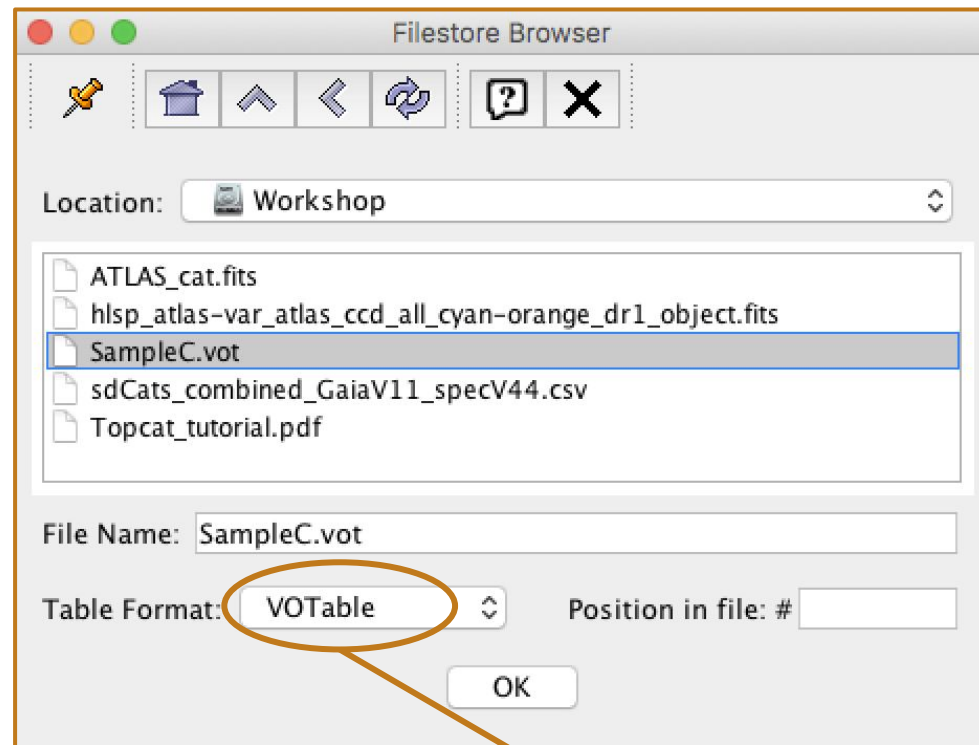
Most important button!



TOPCAT – start window



TOPCAT – open a table



It's necessary to set the correct table format

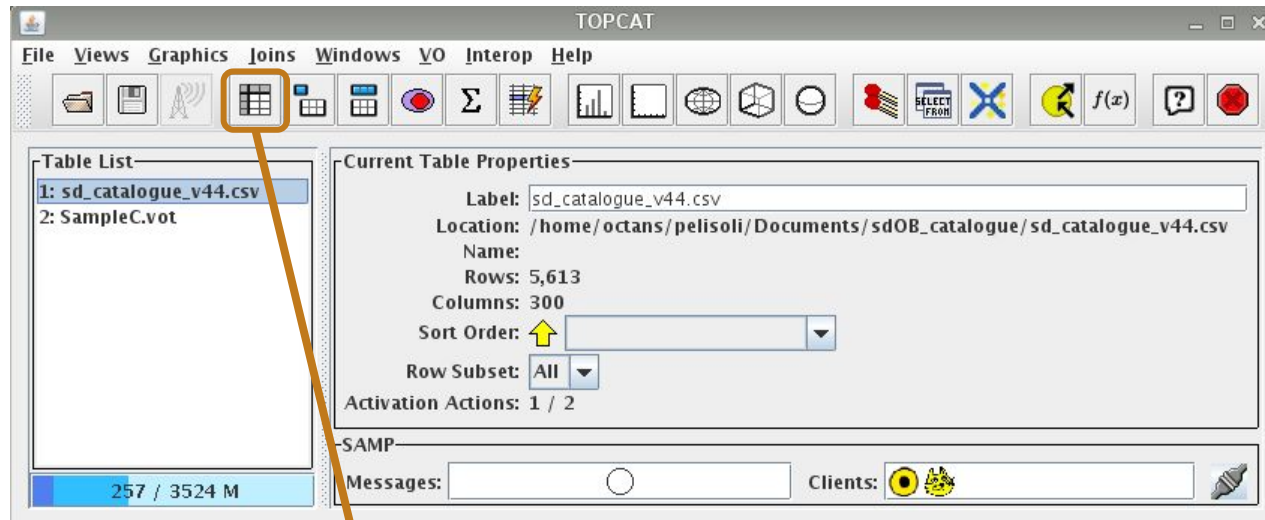
TOPCAT – tables

The screenshot displays the TOPCAT software interface. At the top, there is a menu bar with options: File, Views, Graphics, Joins, Windows, VO, Interop, and Help. Below the menu bar is a toolbar containing various icons for file operations, data visualization, and analysis. The main workspace is divided into two primary panels, both highlighted with orange borders:

- Table List:** A list of tables with the following entries:
 - 1: sd_catalogue_v44.csv
 - 2: SampleC.vot
- Current Table Properties:** A detailed view of the selected table with the following information:
 - Label: sd_catalogue_v44.csv
 - Location: /home/octans/pelisoli/Documents/sdOB_catalogue/sd_catalogue_v44.csv
 - Name:
 - Rows: 5,613
 - Columns: 300
 - Sort Order: [Dropdown menu]
 - Row Subset: All [Dropdown menu]
 - Activation Actions: 1 / 2

At the bottom of the interface, there is a status bar showing memory usage as 257 / 3524 M. On the right side of the status bar, there is a 'SAMP' section with 'Messages:' and 'Clients:' fields, each containing a small circular icon.

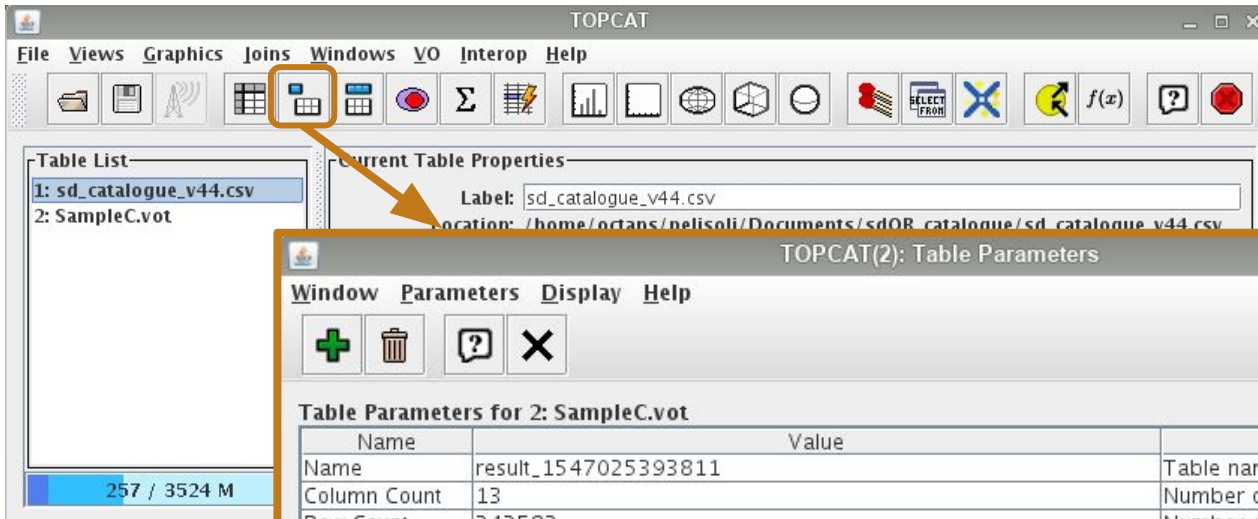
TOPCAT – browse a table



The screenshot shows the 'TOPCAT(2): Table Browser' window. The table is titled 'Table Browser for 2: SampleC.vot' and contains 13 rows of data. The columns are: source_id, ra, dec, parallax, pmra, pmdec, phot_g_me..., phot_bp_m..., phot_rp_m..., bp_rp, teff_val, radius_val, and radial_velocity.

	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



The screenshot shows the 'TOPCAT(2): Table Parameters' dialog box. The dialog has a title bar with 'TOPCAT(2): Table Parameters' and standard window controls. Below the title bar is a menu bar with 'Window', 'Parameters', 'Display', and 'Help'. There are four icons: a green plus sign, a trash can, a question mark, and a close button. The main content area is titled 'Table Parameters for 2: SampleC.vot' and contains a table with the following data:

Name	Value	Description
Name	result_1547025393811	Table name
Column Count	13	Number of columns
Row Count	242582	Number of rows
QUERY_STATUS	OK	
PROVIDER	ARI (Astronomisches Rechen Institut - Heidelberg, Germany)	ARI's TAP access to the Gaia Archive.
QUERY	SELECT source_id, ra, dec, g.parallax, pmra, pmdec, g.phot_g_...	

Below the table, there are several input fields for the 'QUERY' parameter:

Name: QUERY
Class: String
Shape:
Units:
Description:
UCD:
Utype:
Value: SELECT source_id, ra, dec, g.parallax, pmra, pmdec, g.phot_g_mean_mag, phot_bp_mean_mag, radius_val, radial_velocity FROM gaiadr2.gaia source AS g JOIN TAP_UPLOAD.t1.USING (source_id)

TOPCAT – column metadata

TOPCAT

File Views Graphics Joins Windows VO Interop Help

Table List

- 1: sd_catalogue_v44.csv
- 2: SampleC.vot

Current Table Properties

Label: sd_catalogue_v44.csv
 Location: /home/octans/pelisoli/Documents/sdOB_catalogue/sd_catalogue_v44.csv
 Name:
 Rows: 5,613
 Columns: 300
 Sort Order: ↑
 Row Subset: All
 Activation Actions: 1, 2

SAMP

Messages: Clients:

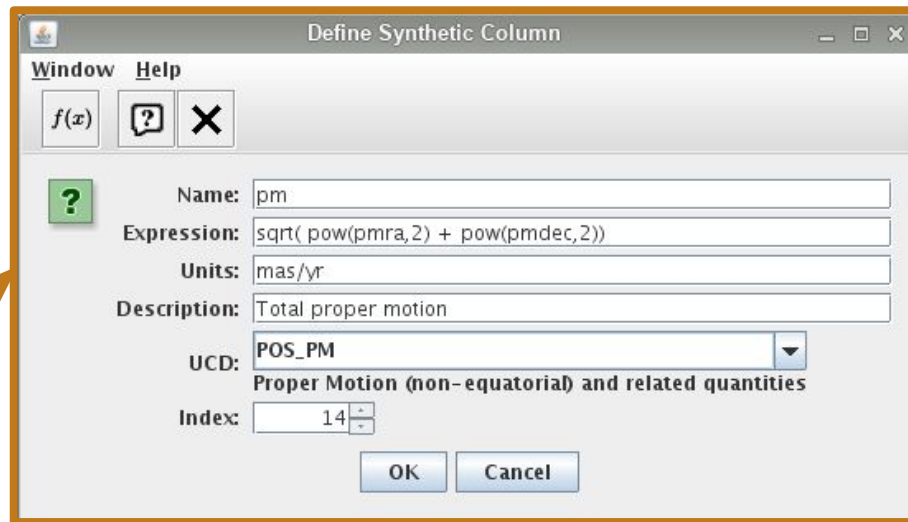
TOPCAT(2): Table Columns

Window Columns Display Help

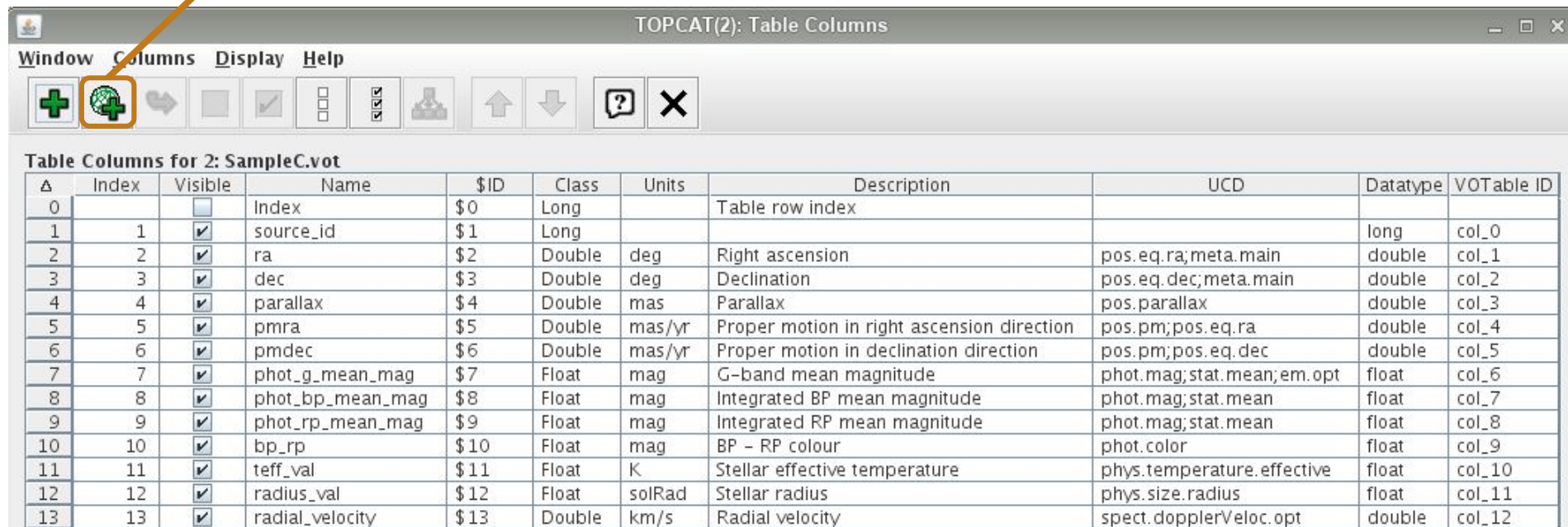
Table Columns for 2: SampleC.vot

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

TOPCAT – create new column



The dialog box is titled "Define Synthetic Column". It has a menu bar with "Window" and "Help". Below the menu bar are three buttons: a button with $f(x)$, a help button with a question mark, and a close button with an X. The main area contains several input fields: "Name" with the value "pm", "Expression" with the formula $\text{sqrt}(\text{pow}(\text{pmra},2) + \text{pow}(\text{pmdec},2))$, "Units" with "mas/yr", "Description" with "Total proper motion", "UCD" with a dropdown menu showing "POS_PM" and a tooltip "Proper Motion (non-equatorial) and related quantities", and "Index" with a spinner box set to "14". At the bottom are "OK" and "Cancel" buttons.



The main window is titled "TOPCAT(2): Table Columns". It has a menu bar with "Window", "Columns", "Display", and "Help". Below the menu bar is a toolbar with various icons, including a green plus sign in a circle. The main area displays a table with the following columns: Δ , Index, Visible, Name, \$ID, Class, Units, Description, UCD, Datatype, and VOTable ID. The table contains 14 rows of data.

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

TOPCAT – create subsets

TOPCAT(1): Row Subsets

Window Subsets Display Interop Help

ID	Name	Size	Fraction	Expression
.1	All	5613	100%	

Define Row Subset

Window Help

Simple expression

Subset Name: bright

Expression: `G_Gaia < 10`

OK Cancel

Define Row Subset

Window Help

Use of functions

Subset Name: fast

Expression: `abs(RV) > 400`

OK Cancel

Define Row Subset

Window Help

Missing values

Subset Name: noRVdata

Expression: `NULL_RV`

OK Cancel

TOPCAT

File Views Graphics Joins Windows VO Interop Help

Table List

- 1: sd_catalogue_v44.csv
- 2: SampleC.vot

Current Table Properties

Label: sd_catalogue_v44.csv

Location: /home/octans/pelisoli/Documents/sdOB_catalogue/s

Name:

Rows: 5,613

Columns: 300

Sort Order: ↑

Row Subset: All

Activation Actions: 1 / 2

SAMP

Messages: ○ Clients: ☀ 🐛

257 / 3524 M

TOPCAT – create column based on subset

Define Synthetic Column

Window Help

f(x) ? X If statement

? Name: Observe?

Expression: (fast && bright) ? "yes" : "no"

Units:

Description:

UCD: no UCD

Index: 301

OK Cancel

TOPCAT(2): Table Columns

Window Columns Display Help

+ + ? X

Table Columns for 2: SampleC.vot

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

TOPCAT – Visualisation tools

The image displays the TOPCAT software interface. On the left, the 'Table List' shows two tables: '1: sd_catalogue_v44.csv' and '2: SampleC.vot'. The 'Current Table Properties' panel shows details for 'sd_catalogue_v44.csv', including its location, name, rows (5,613), and columns (300). A toolbar on the left contains various icons, with the histogram icon highlighted by a red box and an arrow pointing to the main plot area.

The main plot area, titled 'Histogram Plot', shows a histogram of 'C_Gaia' values. The x-axis ranges from 8 to 21, and the y-axis ranges from 0 to 600. The histogram bars are colored red. A legend in the top right corner indicates three series: '1: All' (red), '1: bright' (green), and '1: fast' (blue). The histogram shows a distribution peaking around 16.5.

Below the plot, the 'Position' and 'Subsets' panels are visible. The 'Position' panel shows the table '1: sd_catalogue_v44.csv' and the x-axis 'C_Gaia'. The 'Subsets' panel shows a list of subsets, with '1: sd_cat' selected. A red box highlights the 'Add function' button in the 'Subsets' panel, with an arrow pointing to a text box that says 'Add function (e.g. Gaussian)'. Another red box highlights the 'Add another histogram' button in the 'Subsets' panel, with an arrow pointing to a text box that says 'Add another histogram'.

The status bar at the bottom shows the current position and count: 'Position: X ? Select' and 'Count: 5,596 / 5,613'.

TOPCAT – Visualisation tools

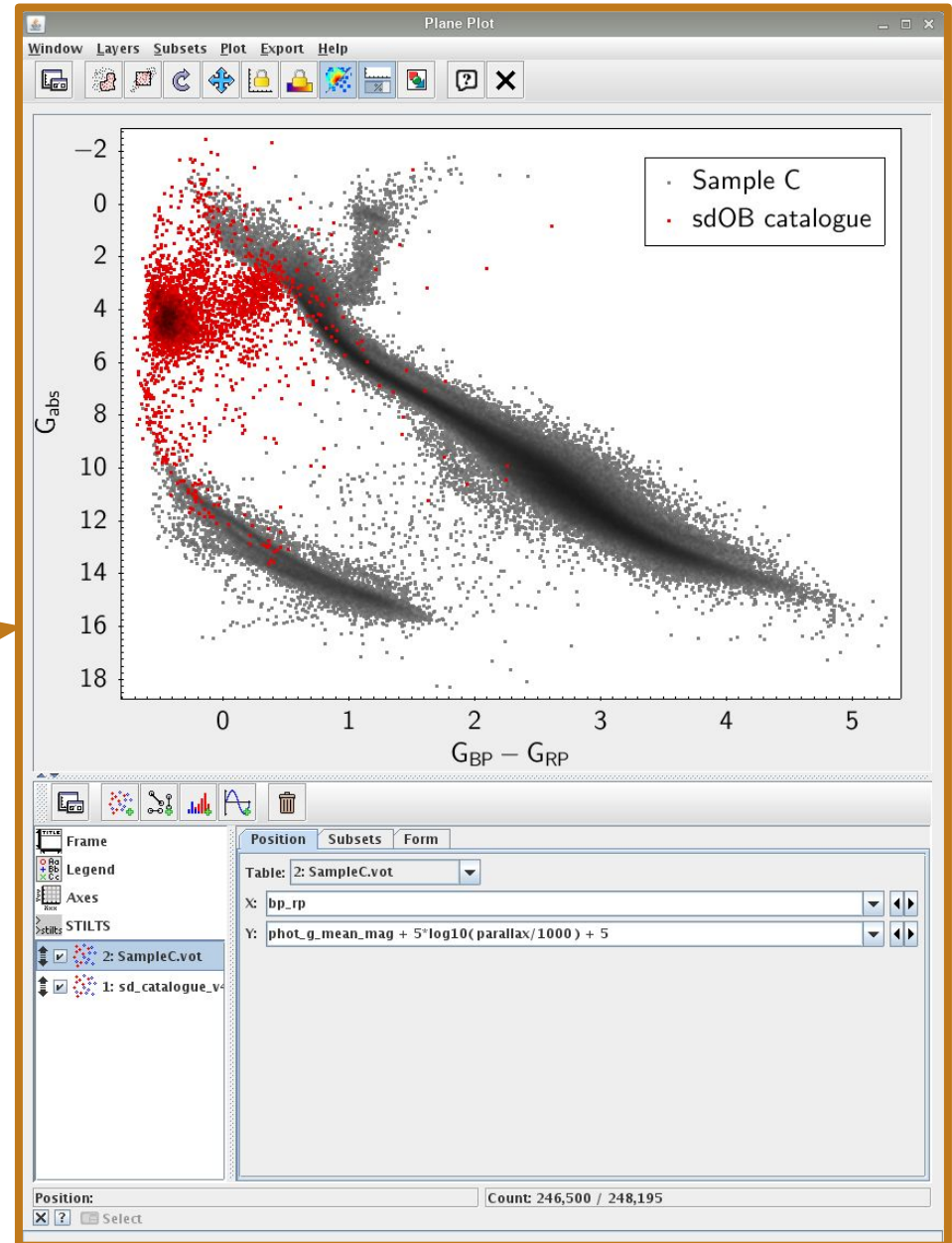
The screenshot shows the TOPCAT software interface. The 'Current Table Properties' panel displays the following information:

- Label: sd_catalogue_v44.csv
- Location: /home/octan/pelissoli/Documents/sdOB_catalogue/sd_catalogue_v44.csv
- Name:
- Rows: 5,613
- Columns: 300
- Sort Order:
- Row Subset: All
- Activation Actions: 1 / 2

The 'Table List' panel shows two tables:

- 1: sd_catalogue_v44.csv
- 2: SampleC.vot

The interface also includes a menu bar (File, Views, Graphics, Joins, Windows, VO, Interop, Help) and a toolbar with various icons. A callout box labeled 'Plane plot' with an arrow points to the 'Plane Plot' window shown in the adjacent image.



TOPCAT – Visualisation tools

TOPCAT

File Views Graphics Joins Windows VO Interop Help

Current Table Properties

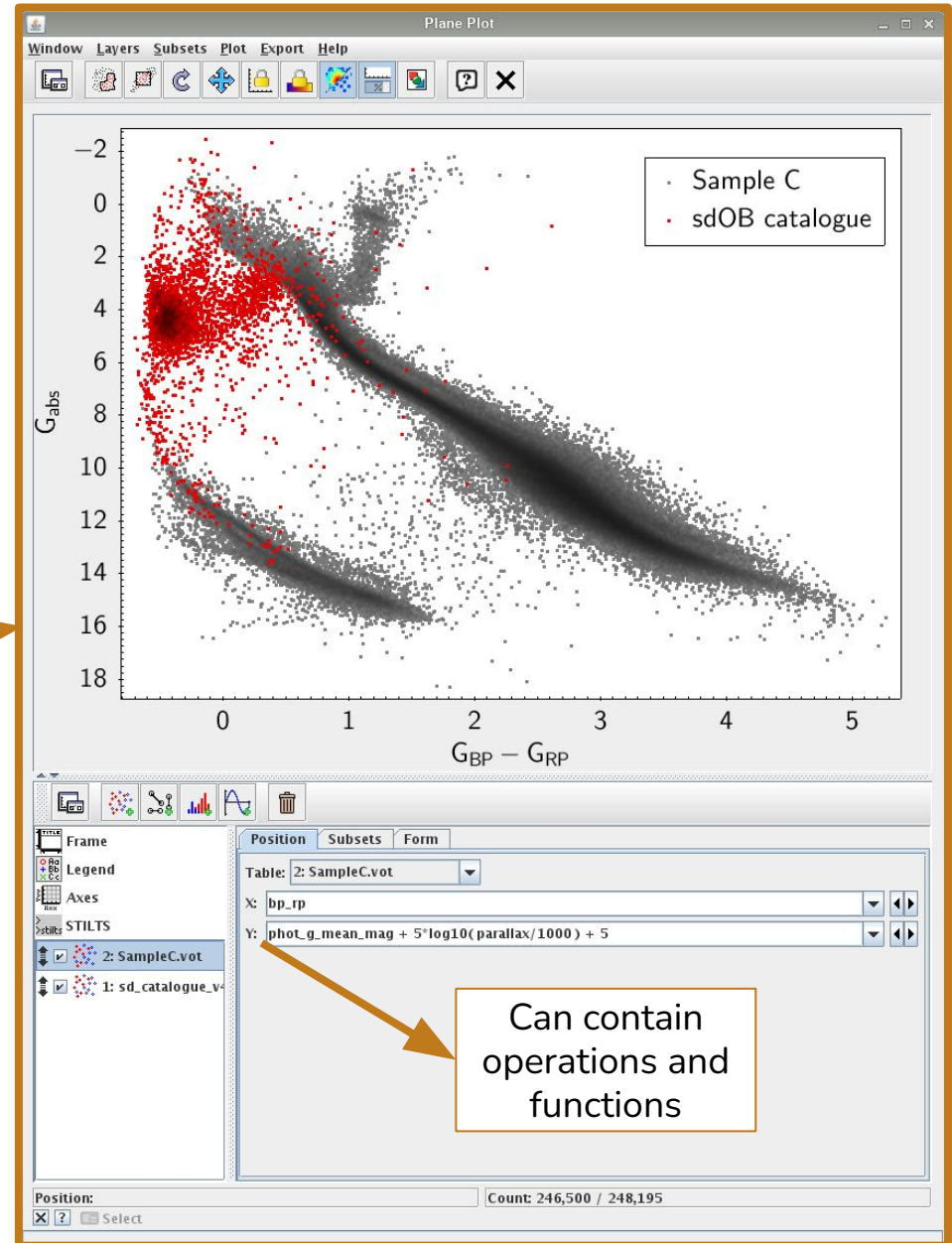
Label: sd_catalogue_v44.csv
Location: /home/octan/pelissoli/Documents/sdOB_catalogue/sd_catalogue_v44.csv
Name:
Rows: 5,613
Columns: 300
Sort Order:
Row Subset: All
Activation Actions: 1 / 2

Table List

1: sd_catalogue_v44.csv
2: SampleC.vot

257 / 3524 M

Plane plot



TOPCAT – Visualisation tools

TOPCAT

File Views Graphics Joins Windows VO Interop Help

Current Table Properties

Label: sd_catalogue_v44.csv
Location: /home/octan/pelissoli/Documents/sdOB_catalogue/sd_catalogue_v44.csv
Name:
Rows: 5,613
Columns: 300
Sort Order:
Row Subset: All
Activation Actions: 1 / 2

Messages:

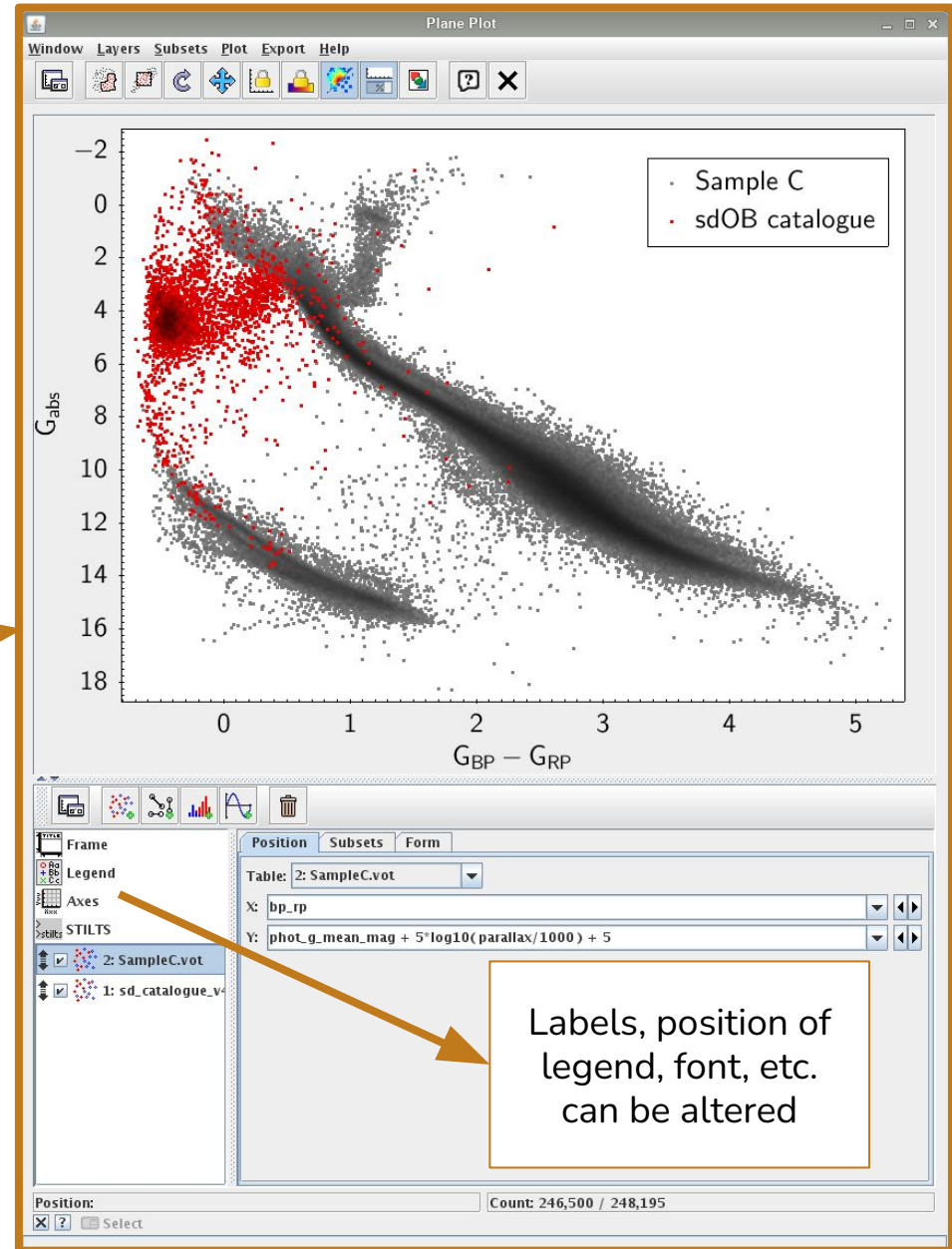
SAMP

Table List

- 1: sd_catalogue_v44.csv
- 2: SampleC.vot

257 / 3524 M

Plane plot



Labels, position of legend, font, etc. can be altered

TOPCAT – Visualisation tools

TOPCAT

File Views Graphics Joins Windows VO Interop Help

Current Table Properties

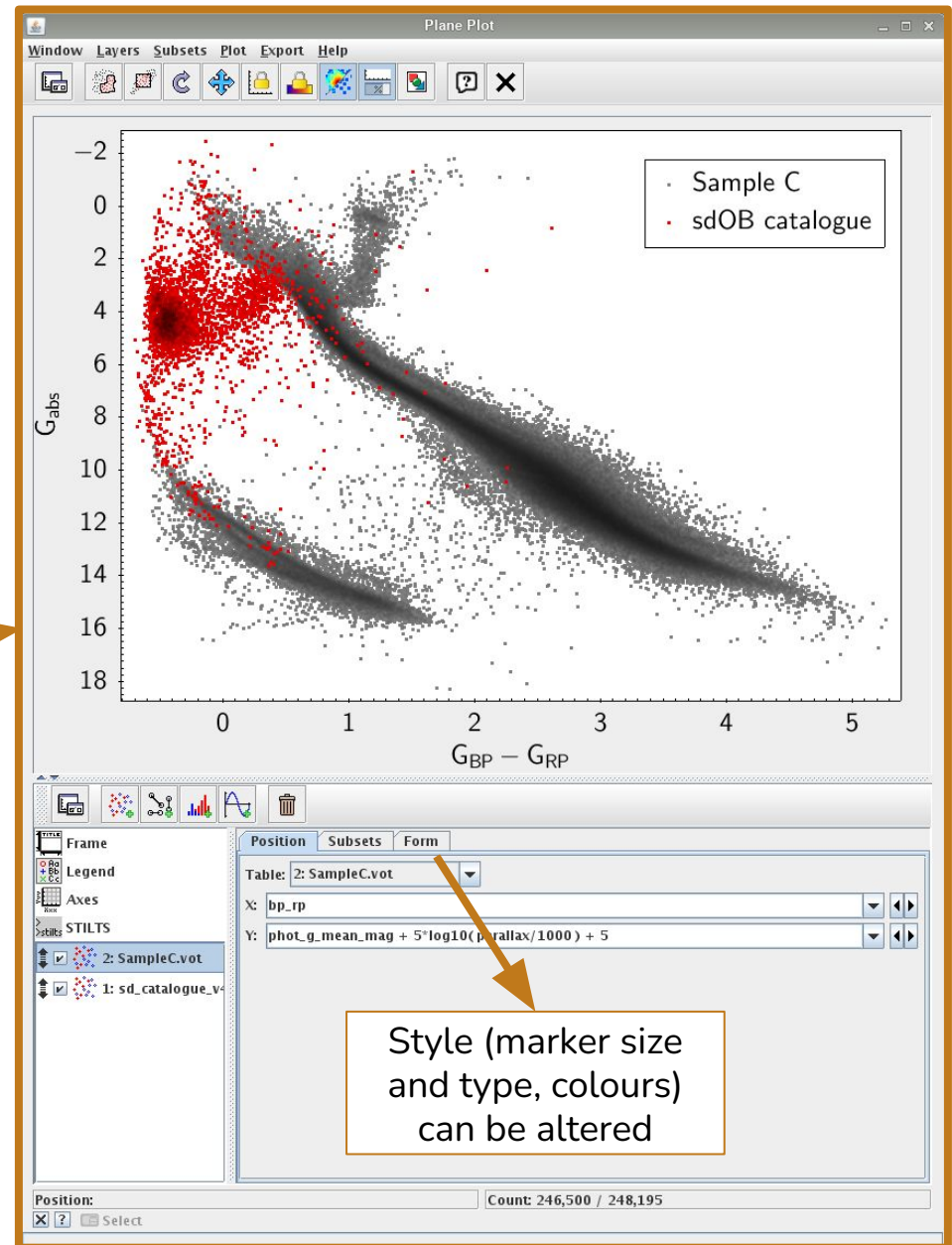
Label: sd_catalogue_v44.csv
Location: /home/octan/pelissoli/Documents/sdOB_catalogue/sd_catalogue_v44.csv
Name:
Rows: 5,613
Columns: 300
Sort Order:
Row Subset: All
Activation Actions: 1 / 2

Table List

1: sd_catalogue_v44.csv
2: SampleC.vot

257 / 3524 M

Plane plot



TOPCAT – Visualisation tools

TOPCAT

File Views Graphics Joins Windows VO Interop Help

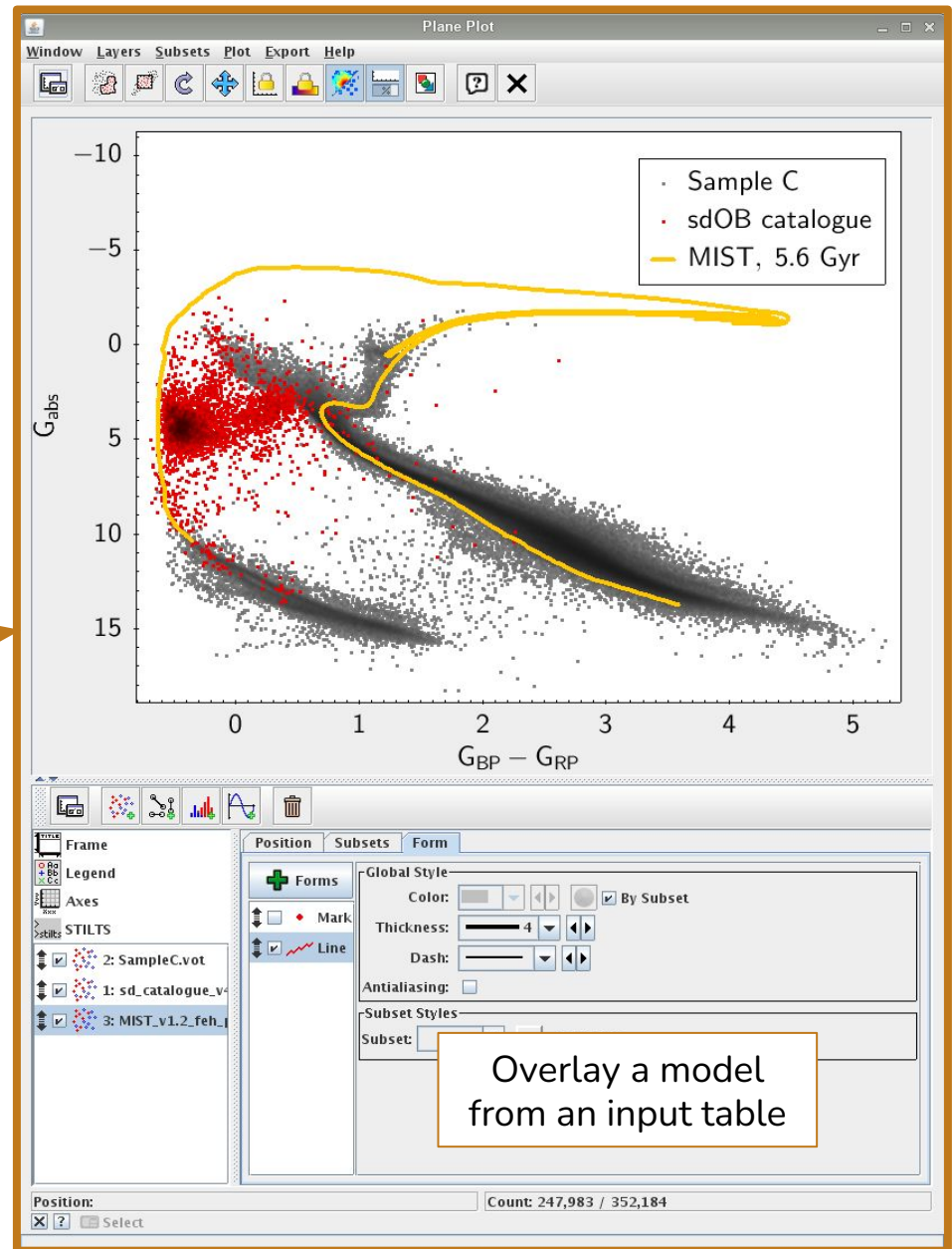
Current Table Properties

Label: sd_catalogue_v44.csv
Location: /home/octan/pelissoli/Documents/sdOB_catalogue/sd_catalogue_v44.csv
Name: sd_catalogue_v44.csv
Rows: 5,613
Columns: 300
Sort Order:
Row Subset: All
Activation Actions: 1 / 2

Table List

1: sd_catalogue_v44.csv
2: SampleC.vot

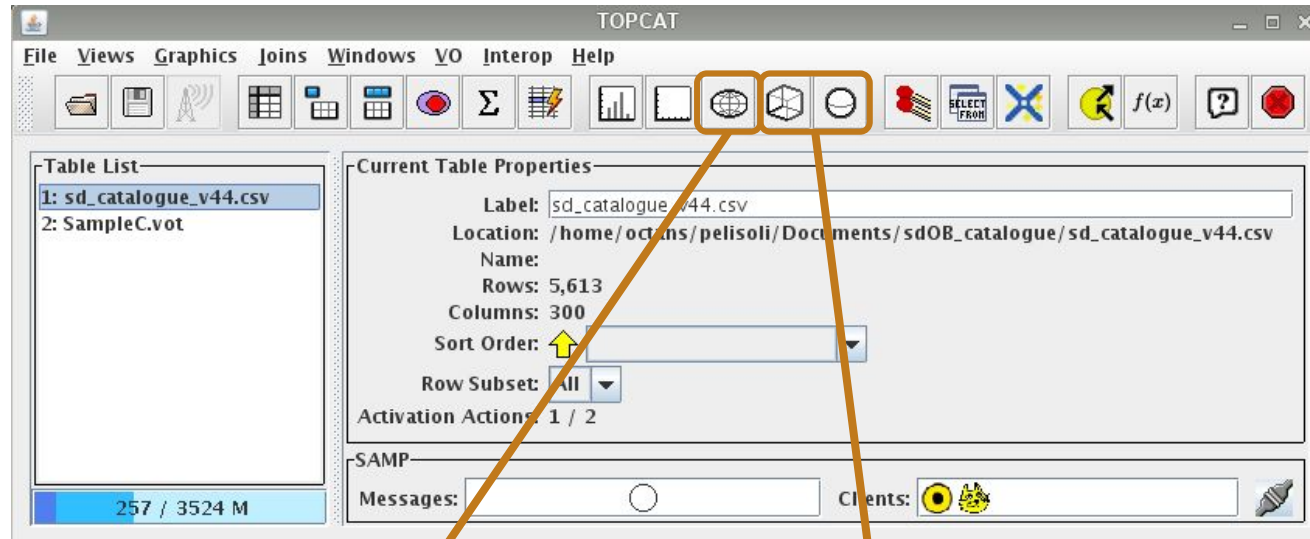
257 / 3524 M



Plane plot

Overlay a model from an input table

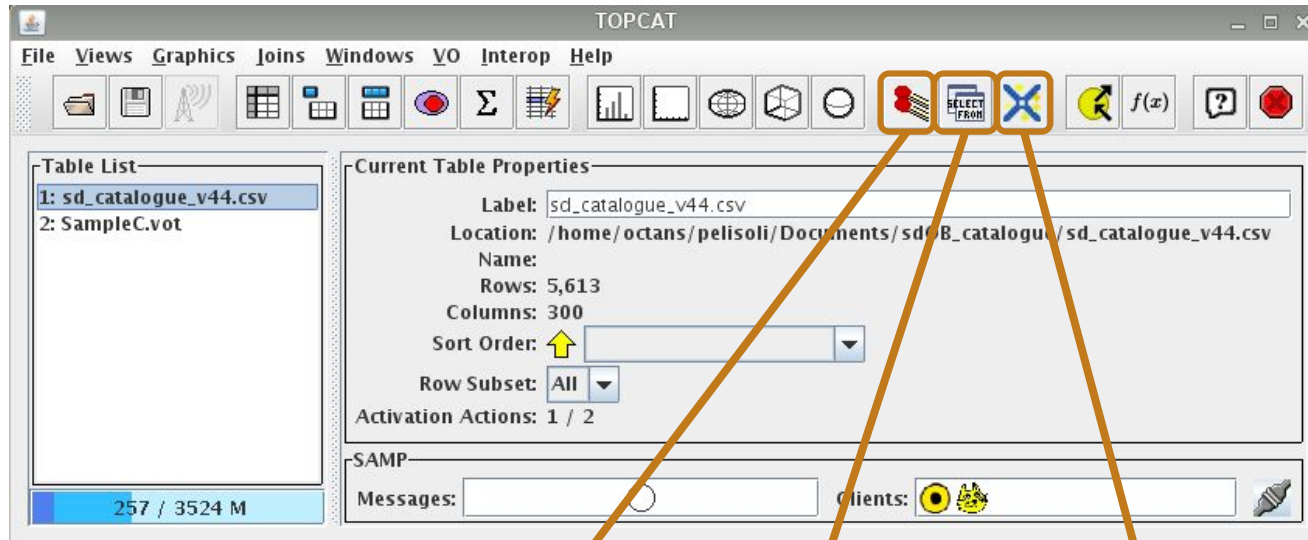
TOPCAT – Visualisation tools



Sky plotting

3D plotting

TOPCAT – Crossmatching



Match two local tables

Query using ADQL

Match local table to survey (e.g. SDSS, Gaia...)

ADQL queries



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

Very basic summary of a query:

```
SELECT [TOP (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)]
```


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 gaiadr2.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(POWER(pmra,2)+POWER(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 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 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 gaiadr2.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 gaiadr2.gaia_source AS b USING(source_id)
```

ADQL queries – Geometries



- Useful for searching a radius around given coordinates
- E.g.: get Gaia DR2 proper motions for stars with *unknown* source_id (3" search)

ADQL queries – Geometries

- Useful for searching a radius around given coordinates
- E.g.: get Gaia DR2 proper motions for stars with *unknown* source_id (3" search)

```
SELECT b.source_id, a.NAME_SDCAT, b.pmra, b.pmdec
FROM TAP_UPLOAD.t10 AS a
JOIN gaiadr2.gaia_source AS b ON 1=CONTAINS (
    POINT('ICRS', a.RAJ2000, a.DEJ2000),
    CIRCLE('ICRS', b.ra, b.dec, 3./3600.))
```








- Note: same thing could be done with a TOPCAT crossmatch, but that is not always the case (e.g. if a table is not listed for crossmatching).

Exercise: Pleiades


From the tutorial at

<http://andromeda.star.bris.ac.uk/topcat/tutorial/>

(Credit: Niall Deacon, Hawaii)

- Open the VizieR load dialog () (click on "VO" at the top bar menu)
- Search for all the objects within 3 degrees of the Pleiades in the Tycho-2 catalogue:
 - Check **Cone Selection** button
 - Object name Pleiades, **Resolve**
 - Radius 3 degrees
 - Catalogue Selection **Surveys** tab
 - Click on row Tycho-2 (Name column is ordered alphabetically)
 - Click **OK**
 - Loads 2 tables (2 tables in VizieR under that heading) - pick the one with most rows
- Visualise proper motions:
 - Open a scatter plot window 
 - X = pmRA, Y = pmDE
 - Zoom in to find a cluster with non-zero motion
 - Draw a blob round it to create a new subset (click ; drag out the cluster region, click  again)
- Draw colour-magnitude diagram:
 - Open a different scatter plot window 
 - X = VTmag - BTmag, Y = VTmag, flip Y
 - See where the new cluster subset you identified sit in colour-magnitude space (main sequence?).
- Save the cluster identification:
 - Go to the Subsets window 
 - Select the row corresponding to the cluster subset
 - Create a new boolean table corresponding to this subset by clicking the **To Column**  toolbar button
 - Save the table.

Creating our photometric target list

- 
- Now that you have familiarised yourself with TOPCAT, we can create a list of targets for photometry!
 - We want to observe hot subdwarf stars with suspected variability.
 - We are going to use a table containing 40,000+ hot subdwarf and candidates:

http://www.astro.physik.uni-potsdam.de/~pelisoli/AstroWorkshop/sdCats_combined_GaiaV11_specV44.csv

- To identify candidate variables, we will use the ATLAS catalogue:

<https://archive.stsci.edu/prepds/atlas-var/>

(download the “Object Table”)

* This table is 7GB in size! If the download takes too long, download instead:

http://www.astro.physik.uni-potsdam.de/~pelisoli/AstroWorkshop/ATLAS_cat.fits

Creating our photometric target list

- **Step 1:** import both tables to TOPCAT.
- **Step 2:** select only relevant columns from the ATLAS table.
 - There are 197 (!) columns in this table – they describe many parameters in the variability search algorithm run by ATLAS.
 - Using the column metadata shortcut, deselect all columns, then select only:
 - ATO_ID
 - ra and dec (we need those to do a crossmatch)
 - fp_period
 - fp_fitrms
 - fp_fitchi
 - CLASS (this is the type of variation ATLAS identified)

We are interested in short period binaries. These parameters describe the fitted period, root-mean-square, and chi-square of the short-period algorithm in ATLAS.

Creating our photometric target list

- **Step 3: cross-match both tables**



Match Tables

Window Tuning Help

Match Criteria

Algorithm: Sky

Max Error: 5 arcsec

Table 1

Table: 16: sdCats_combined_GaiaV11_specV44

RA column: RAJ2000 degrees

Dec column: DEJ2000 degrees

Table 2

Table: 17: hlsp_atlas-var_atlas_ccd_all_cyan-orange_dr1_obj...

RA column: ra degrees

Dec column: dec degrees

Output Rows

Match Selection: Best match, symmetric

Join Type: 1 and 2

Scanning rows for table 2...
Eliminating multiple row references...
Elapsed time for match: 14 seconds
Match succeeded

Go Stop

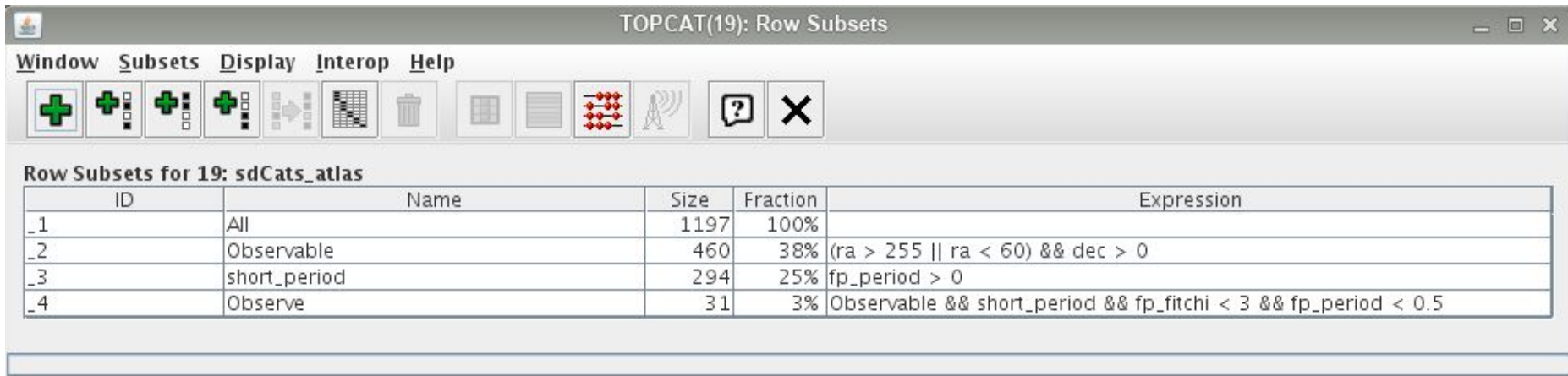
Creating our photometric target list



- **Step 4:** create a subset with objects worth observing for our science case, taking into account the time and site constraints (for next week in Ondřejov).
 - You can use staralt: <http://catserver.ing.iac.es/staralt/index.php>
(Ondřejov location: 14.781°E 49.915°N 500m, UT-offset +2)
 - We want objects that do have a short period determined.
 - Preferably objects whose period can be covered in one night.

Creating our photometric target list

- **Step 4:** create a subset with objects worth observing for our science case, taking into account the time and site constraints (for next week in Ondrejov).
 - We want objects that do have a short period determined.
 - Preferably objects whose period can be covered in one night.



TOPCAT(19): Row Subsets

Window Subsets Display Interop Help

ID	Name	Size	Fraction	Expression
_1	All	1197	100%	
_2	Observable	460	38%	(ra > 255 ra < 60) && dec > 0
_3	short_period	294	25%	fp_period > 0
_4	Observe	31	3%	Observable && short_period && fp_fitchi < 3 && fp_period < 0.5

Creating our photometric target list

TOPCAT(19): Table Browser

Window Subsets Help

Table Browser for 19: sdCats_atlas

	NAME_SDCAT	RAJ2000	DEJ2000	SPEC_SDCAT	phot_g_me...	ATO_ID	fp_period	fp_fitrms	fp_fitchi	CLASS
1073		331.33941	35.55149		15.80169	J331.3394+35.5514	0.08587	0.03249	1.20774	NSINE
943		299.98457	50.61615		17.09661	J299.9845+50.6161	0.09119	0.07073	1.26616	dubious
997		307.21987	6.16752		14.81179	J307.2199+06.1675	0.09422	0.02025	1.56105	NSINE
971		303.32853	42.42771		15.61423	J303.3285+42.4276	0.09526	0.03289	2.42345	dubious
754		267.15569	9.16338		16.71258	J267.1556+09.1633	0.09646	0.07756	1.39432	SINE
758		267.90184	14.73861		16.94822	J267.9018+14.7386	0.09951	0.07587	1.65196	dubious
257		30.59676	51.89702		15.11298	J030.5967+51.8970	0.1008	0.02152	1.25241	NSINE
1102		343.38397	47.69991		16.41222	J343.3839+47.6999	0.1055	0.03974	0.8814	NSINE
1026		316.00593	34.61008		17.47401	J316.0059+34.6100	0.11855	0.08564	1.9227	dubious
1105		344.33419	49.65927		17.56062	J344.3341+49.6592	0.12966	0.08458	1.2062	dubious
189	HS2035+0418	309.50381	4.48565	sdB	14.77305	J309.5037+04.4856	0.13103	0.0249	1.46974	dubious
164	KeplerJ184307+425918	280.77823	42.98835	sdB+WD	15.58791	J280.7782+42.9883	0.13726	0.0395	2.02105	dubious
1024		315.11791	59.65741		16.3627	J315.1179+59.6574	0.13772	0.05055	1.28395	NSINE
1129	SDSSJ012458.96+475640.9	21.24568	47.94472	sd	16.92145	J021.2457+47.9447	0.14013	0.07384	1.13479	CBF
792		274.57916	6.89912		17.27594	J274.5791+06.8991	0.14707	0.10659	1.93357	NSINE
226		0.63041	42.88611		14.33737	J000.6304+42.8861	0.15578	0.02251	1.6455	SINE
1077		333.07139	52.02175		17.46384	J333.0713+52.0217	0.16019	0.10316	1.74243	dubious
212	PG2259+134	345.44094	13.64374	sdB	14.51706	J345.4409+13.6437	0.16346	0.02577	1.8531	NSINE
817		280.39493	38.99883		15.85566	J280.3949+38.9988	0.1655	0.04998	2.95899	SINE
1184	SDSSJ192059.78+372220.0	290.24908	37.37222	sdB+dM	15.77123	J290.2490+37.3722	0.16896	0.03841	1.64564	SINE
215	FBS2304+440	346.62686	44.31354	sdB	14.30496	J346.6269+44.3135	0.17589	0.03356	2.88013	CBF
1115		352.34433	32.23316		16.92967	J352.3443+32.2331	0.17644	0.07152	1.09889	NSINE
219	Pn2311-18	351.71858	12.50608	sdB	14.3078	J351.7186+12.5060	0.21191	0.02353	2.7015	IRR
975		304.26974	53.71505		16.33385	J304.2697+53.7150	0.21286	0.04614	1.39473	SINE
1074		331.66585	32.72679		16.962	J331.6658+32.7267	0.22041	0.06343	1.14331	NSINE
229		4.23059	51.23049		16.35795	J004.2305+51.2304	0.27096	0.04629	1.17926	NSINE
842		284.85281	7.85064		15.80786	J284.8528+07.8506	0.29756	0.05785	2.25908	SINE
245		18.47086	50.08699		14.97042	J018.4708+50.0870	0.31029	0.0239	1.37276	NSINE
211	GALEXJ22392+1819	339.80672	18.3295	sdB	14.07094	J339.8067+18.3294	0.36676	0.01961	1.88989	PULSE
1018		312.41318	30.08182		13.5045	J312.4131+30.0818	0.42977	0.02234	2.12267	SINE
925		296.70749	39.99371		14.39431	J296.7074+39.9936	0.45116	0.02622	2.38232	SINE

Creating our photometric target list



- To determine the best targets, you can also inspect the light curves and perform a period search.
- At <http://www.astro.physik.uni-potsdam.de/~pelisoli/lightcurves/ATLAS/dat/> you can find a Jupyter notebook containing instructions, as well as the code, to perform a Lomb-Scargle periodogram and phase-fold the data.

The data is available in this same directory.

Exercise – ADQL queries in TOPCAT

- Draw the Gaia DR2 HR diagram (absolute magnitude $M_G = G - 5 \log(d[\text{pc}]) + 5$ as a function of colour $G_{BP} - G_{RP}$) for 100.000 stars closer than 100 pc.
 - Which variables do you need to select?
 - From which table?
 - How to limit this for 100.000 stars?
 - How to limit this to $d < 100$ pc?
- Inspect this diagram. Is there something odd with it? Why?

Exercise – ADQL queries in TOPCAT

- Retrieve the variables `parallax_over_error`, `phot_bp_mean_flux_over_error`, `phot_rp_mean_flux_over_error`, `phot_bp_rp_excess_factor`, `astrometric_chi2_al`, `astrometric_n_good_obs_al`, and `astrometric_excess_noise` for the stars in the table resulting from your previous query. Hint: use `JOIN USING`.
- Create a subset with objects showing `parallax_over_error < 5`.

How does the HR-diagram look like with only these objects?

- Now you know how important quality control parameters are!
- Use the following conditions to further improve your HR-diagram:

$\text{parallax_over_error} > 10$

$\text{astrometric_excess_noise} < 1.0$

$\text{phot_bp_mean_flux_over_error} > 10$

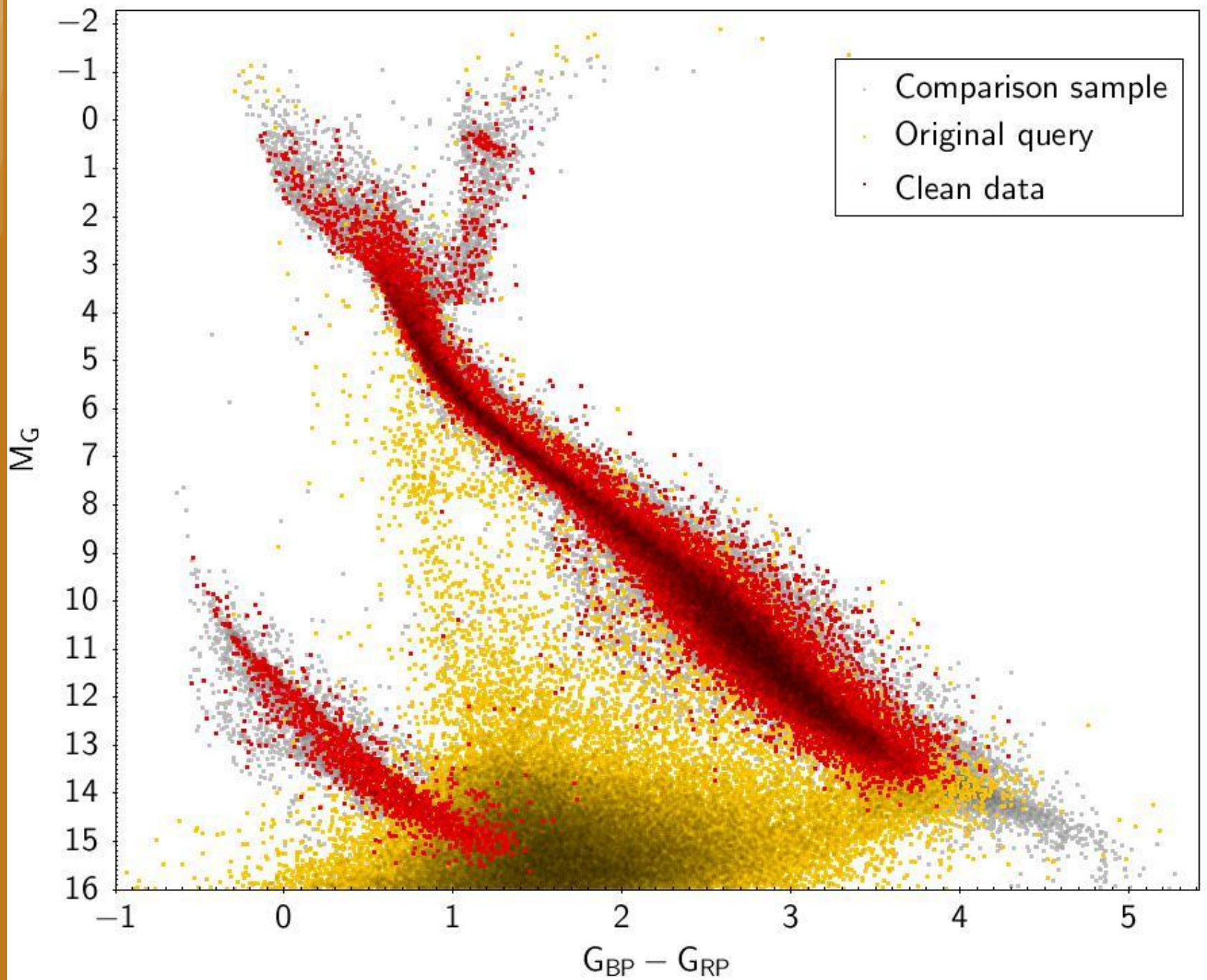
$\text{phot_rp_mean_flux_over_error} > 10$

$\text{phot_bp_rp_excess_factor} < 1.3 + 0.06 * \text{power}(\text{bp_rp}, 2)$

$\text{phot_bp_rp_excess_factor} > 1.0 + 0.015 * \text{power}(\text{bp_rp}, 2)$

$\text{astrometric_chi2_al} / (\text{astrometric_n_good_obs_al} - 5)$
 $< 1.44 * \max(1, \exp(-0.4 * (\text{phot_g_mean_mag} - 19.5)))$

Check out this paper: <https://arxiv.org/abs/1804.09366> if you want to understand more about where all of these parameters come from.



Creating our spectroscopic target list



- Now that you have familiarised yourself with TOPCAT, ADQL, and some Gaia DR2 parameters, it is time to create our list of targets for spectroscopy!
- We want to observe bright candidate hot subdwarf stars identified in *Gaia*. Spectra can confirm (or rule out) their nature.
 - Step 1:** identify the position of these stars in the HR-diagram.
 - Step 2:** define a colour-cut.
 - Step 3:** do a query in *Gaia* recovering stars within your colour cut, also using quality control parameters.
 - Step 4:** validate your query. Does the result make sense?
 - Step 5:** observational constraints (brightness, RA and DEC).

Creating our target list - STEP 1



- Plot the HR-diagram for the comparison sample

Sample C:

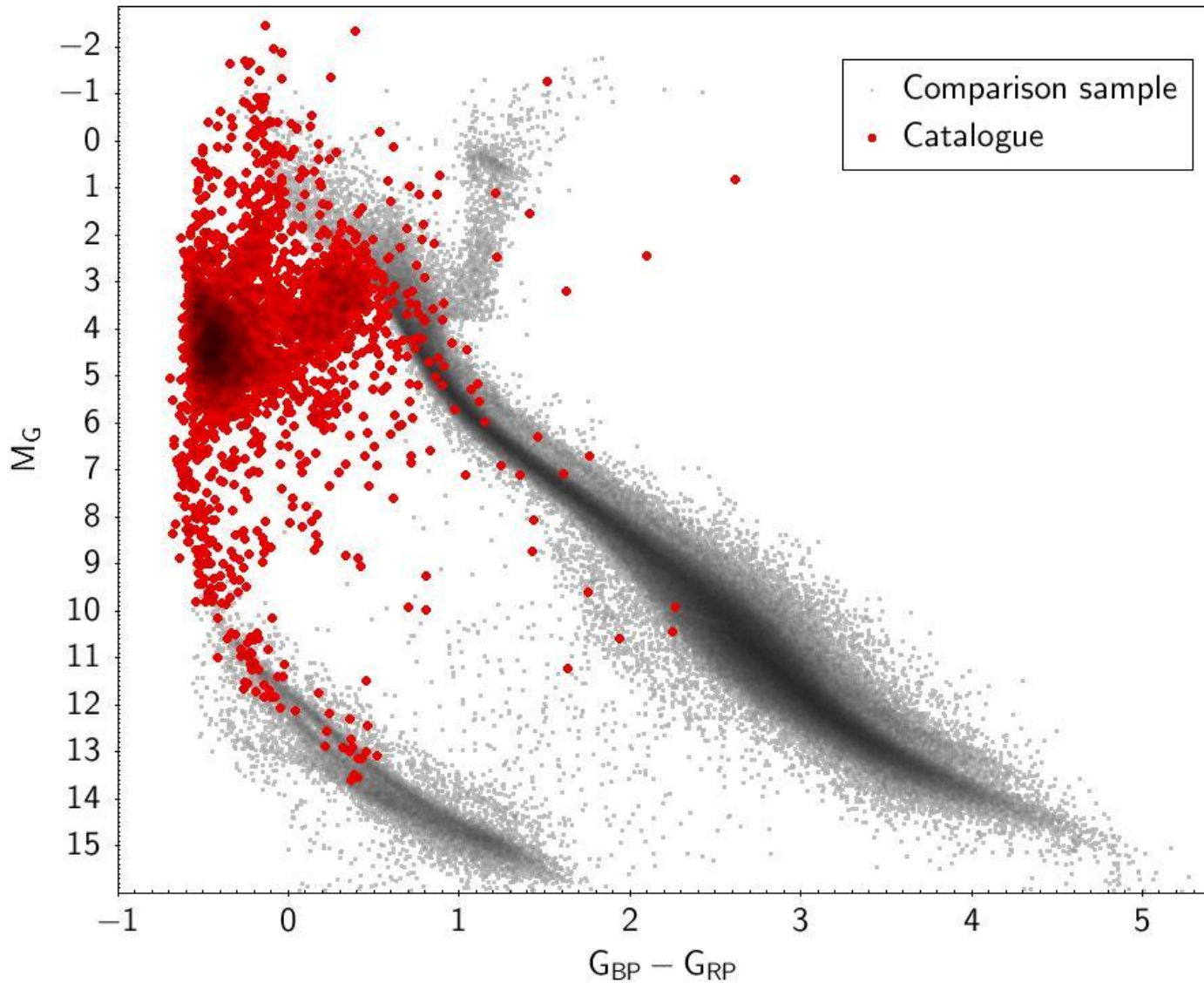
<http://www.astro.physik.uni-potsdam.de/~pelisoli/AstroWorkshop/SampleC.vot>

- Overplot the known hot subdwarfs from Prof. Geier's catalogue

Catalogue:

www.astro.physik.uni-potsdam.de/~pelisoli/AstroWorkshop/sd_catalogue_v44.csv

Creating our target list - STEP 1

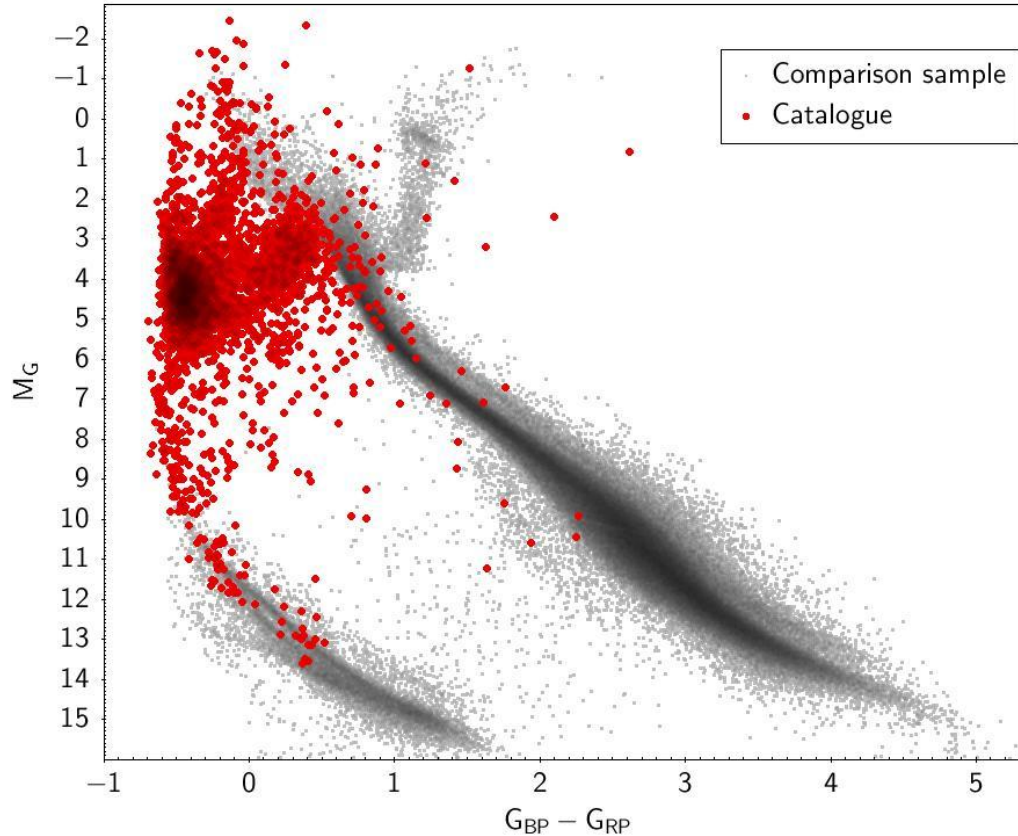


Creating our target list - STEP 2

- Define a colour-cut. Where do these stars concentrate?

$$?? < M_G < ??$$

$$?? < G_{BP} - G_{RP} < ??$$

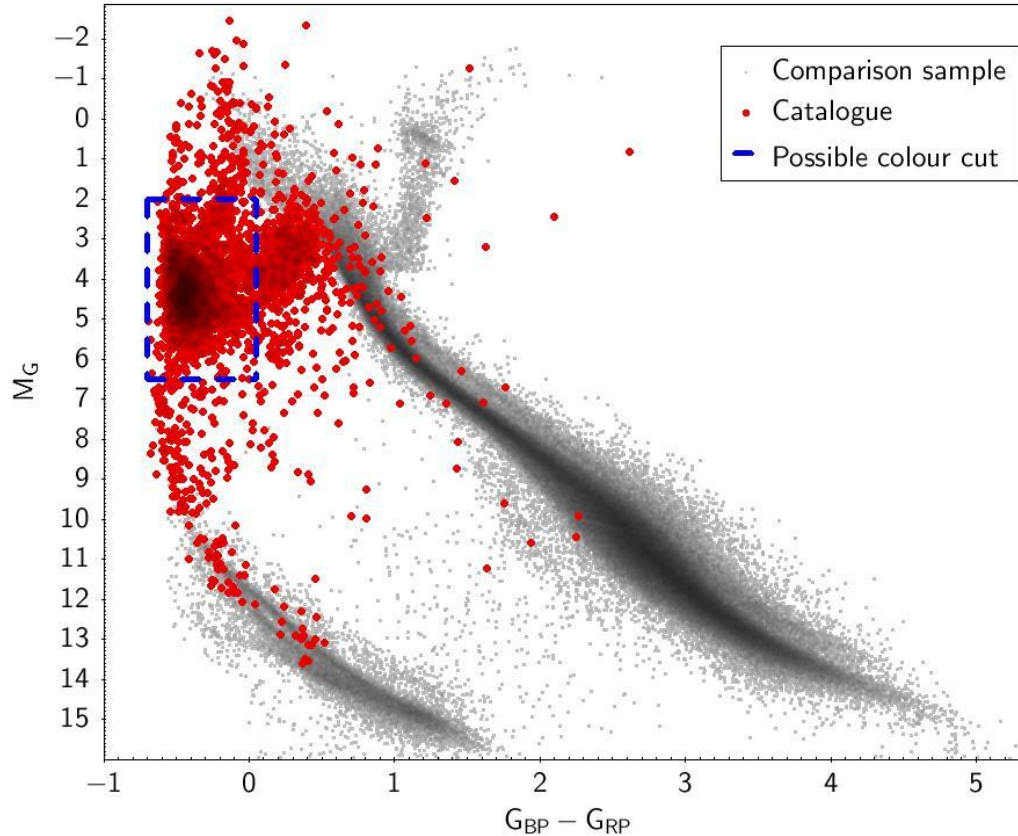


Creating our target list - STEP 2

- Define a colour-cut. Where do these stars concentrate?

$$?? < M_G < ??$$

$$?? < G_{BP} - G_{RP} < ??$$

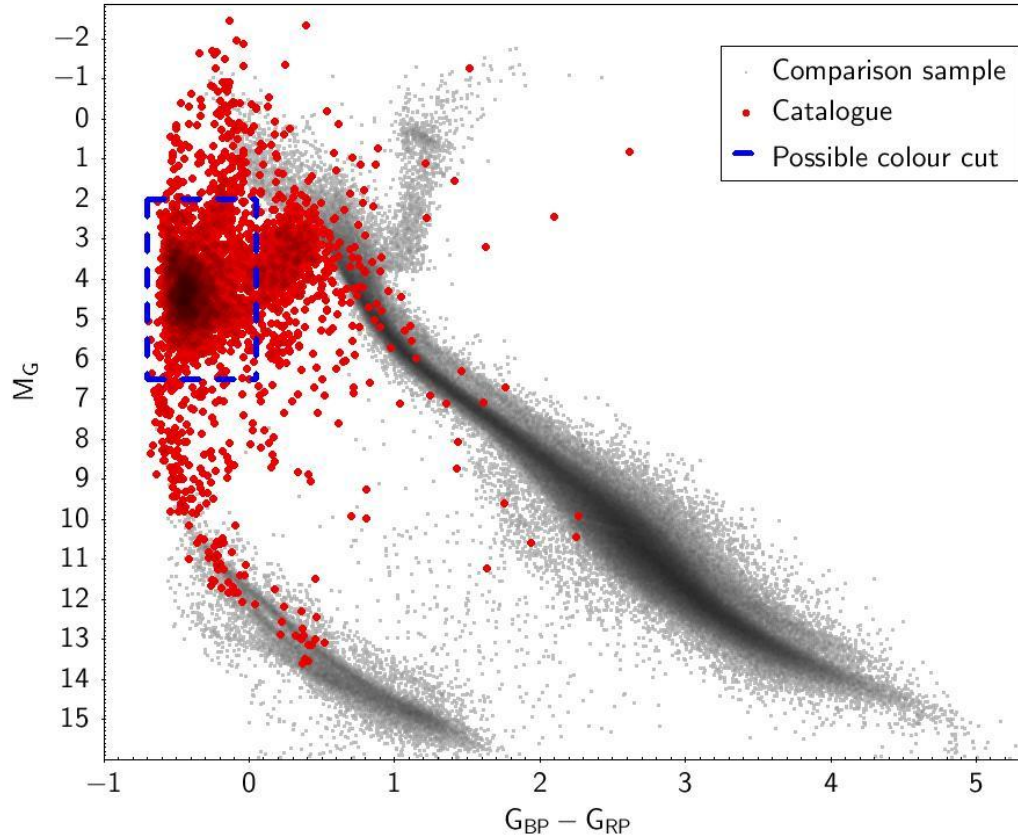


Creating our target list - STEP 2

- Define a colour-cut. Where do these stars concentrate?

$$2 < M_G < 6.5$$

$$-0.7 < G_{BP} - G_{RP} < 0.05$$



Creating our target list - STEP 3



- Write an ADQL query in *Gaia* recovering stars within your colour cut, also using quality control parameters:

```
parallax_over_error > 5
astrometric_excess_noise < 1.0
phot_bp_mean_flux_over_error > 10
phot_rp_mean_flux_over_error > 10
phot_bp_rp_excess_factor < 1.3+0.06*power(bp_rp,2)
phot_bp_rp_excess_factor > 1.0+0.015*power(bp_rp,2)
astrometric_chi2_al/(astrometric_n_good_obs_al-5)
  < 1.44*max(1, exp(-0.4*(phot_g_mean_mag-19.5)) )
```

Creating our target list - STEP 3



- Write an ADQL query in *Gaia* recovering stars within your colour cut, also using quality control parameters.

```
SELECT source_id, ra, dec, parallax, phot_g_mean_mag, bp_rp
FROM gaiadr2.gaia_source
WHERE parallax_over_error > 5
AND phot_bp_mean_flux_over_error>10
AND phot_rp_mean_flux_over_error>10
AND phot_bp_rp_excess_factor < 1.3+0.06*power(phot_bp_mean_mag-phot_rp_mean_mag,2)
AND phot_bp_rp_excess_factor > 1.0+0.015*power(phot_bp_mean_mag-phot_rp_mean_mag,2)
AND ( astrometric_chi2_al/(astrometric_n_good_obs_al-5)<1.44
      OR
      astrometric_chi2_al/(astrometric_n_good_obs_al-5)<1.44*exp(-0.4*(phot_g_mean_mag-19.5)) )
AND bp_rp > -0.7 AND bp_rp < 0.05
AND 5+5*log10(parallax/1000)+phot_g_mean_mag < 6.5 and
5+5*log10(parallax/1000)+phot_g_mean_mag > 2.0
```

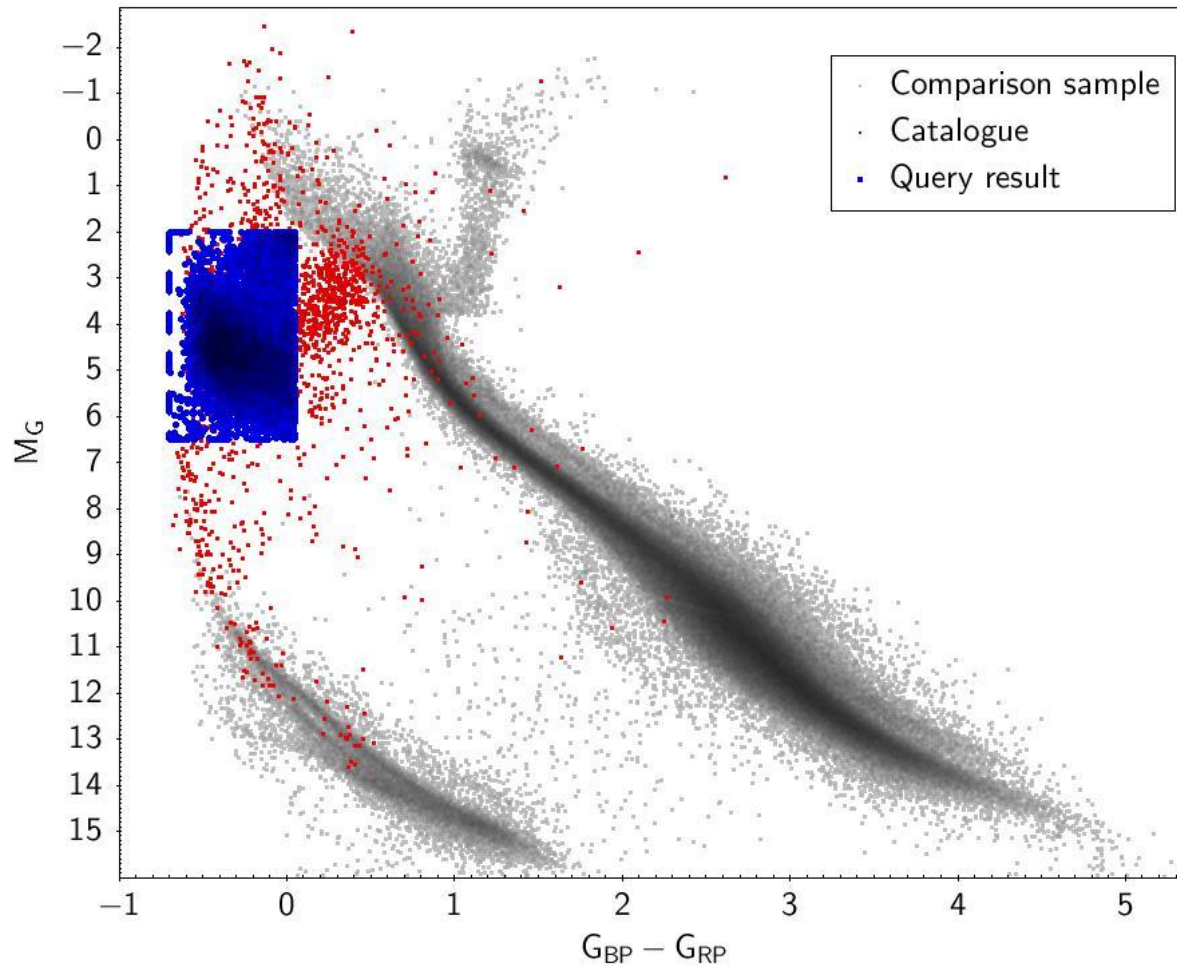
Creating our target list - STEP 4



- Overplot the result of your query on the HR-diagram. Is everything where it is supposed to be?

Creating our target list - STEP 4

- Overplot the result of your query on the HR-diagram. Is everything where it is supposed to be?



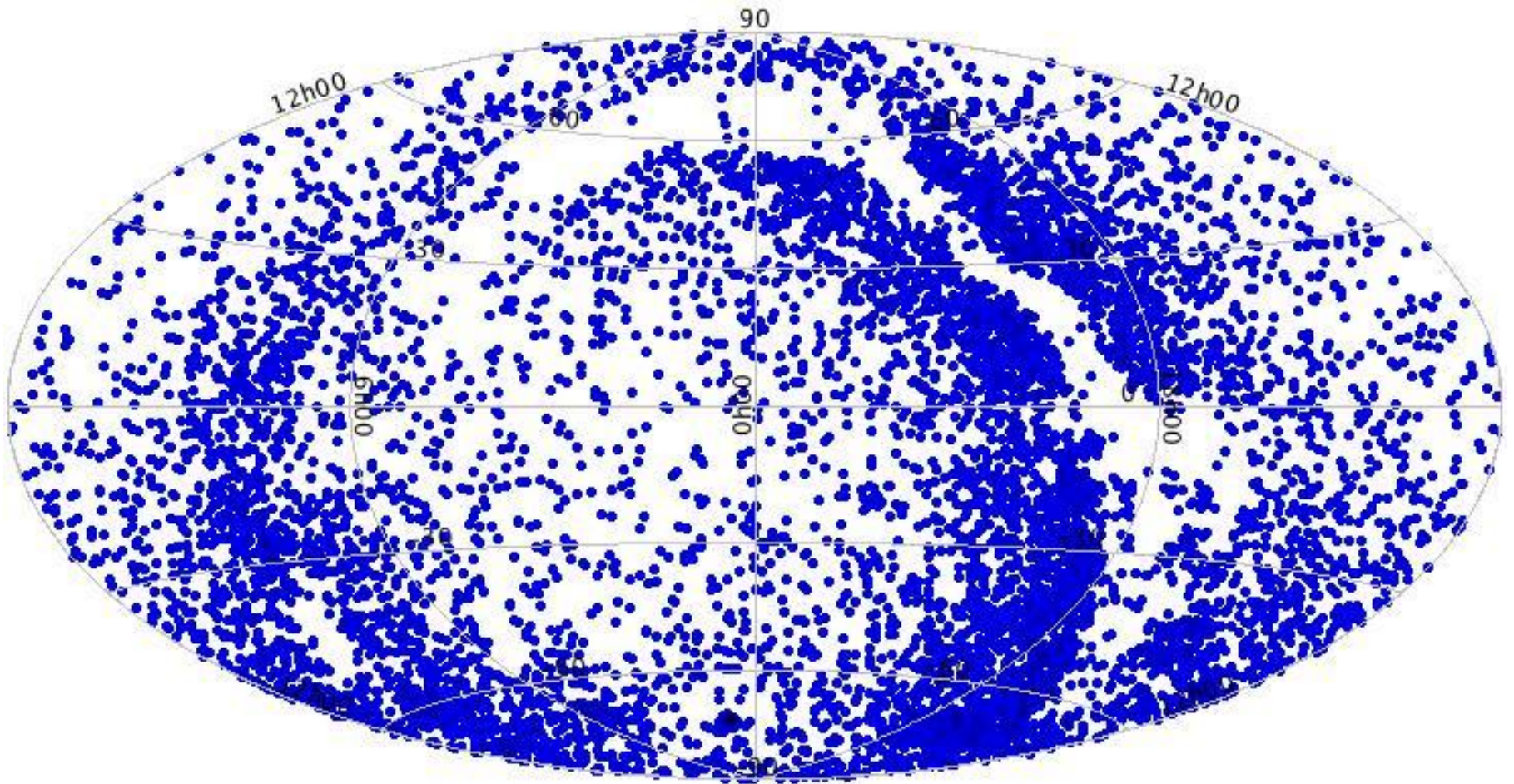
Creating our target list - STEP 4



- Make a sky-plot of the objects in your query. Anything weird?

Creating our target list - STEP 4

- Make a sky-plot of the objects in your query. Anything weird?



Creating our spectroscopic target list



- Congratulations! You have done some proper science.

However, you were too slow... "someone" has already published a catalogue of candidate hot subdwarfs in *Gaia*:

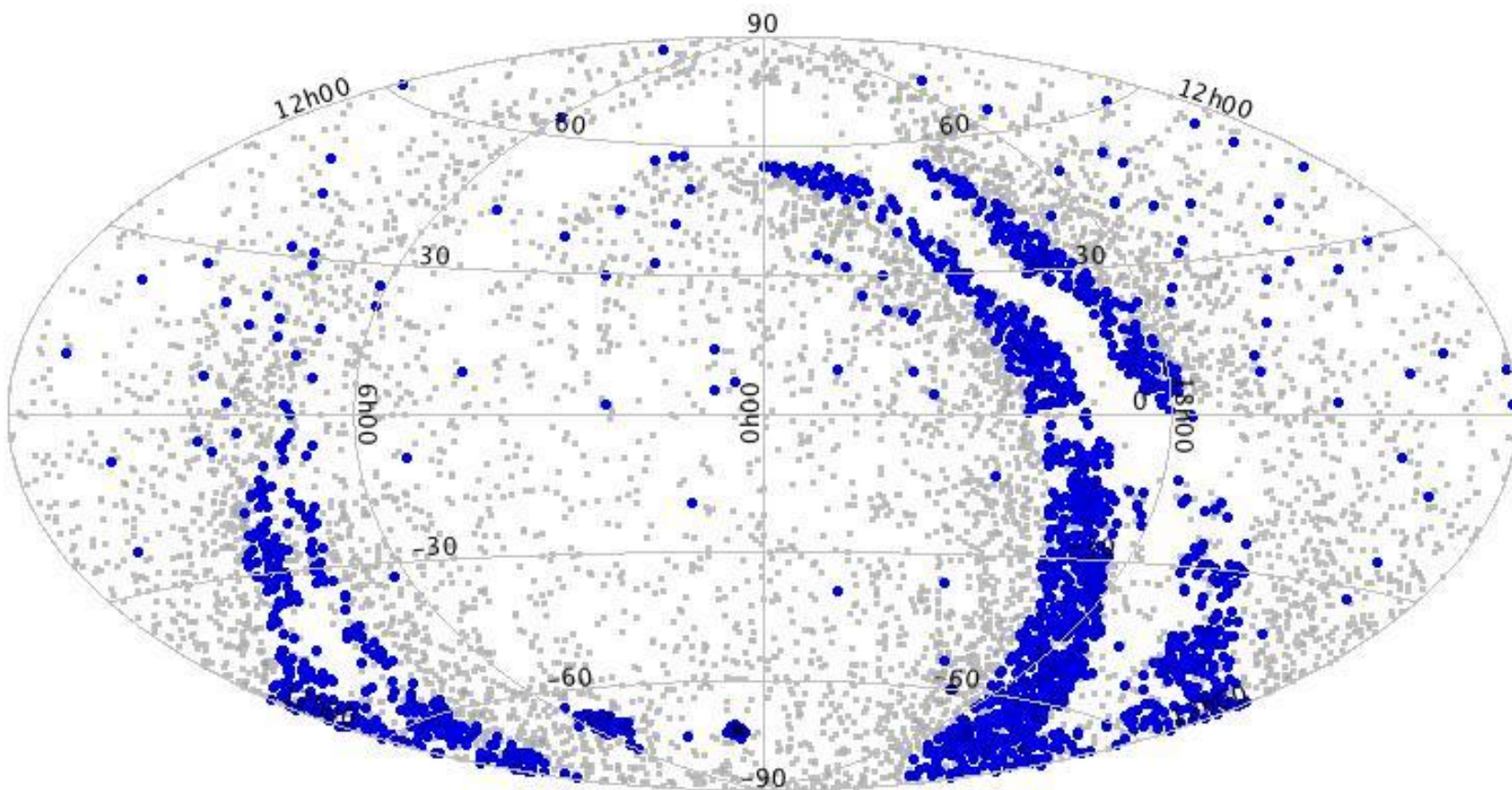
Geier et al. 2019:

[The population of hot subdwarf stars studied with Gaia. II. The Gaia DR2 catalogue of hot subluminous stars](#)

http://www.astro.physik.uni-potsdam.de/~pelisoli/AstroWorkshop/sd_catalogue_gaia_v11.csv


There are things taken into account in the published catalogue that we did not discuss here (e.g. problems in crowded areas), so, from here on, we will use the published catalogue.

Crowded areas are problematic



In blue = in our query, but not in the final catalogue. Essentially the disc and the Magellanic clouds! These regions need stricter quality control cuts.

Creating our spectroscopic target list

- 
- Open the catalogue in TOPCAT.
 - What is the brightness constraint for our telescope?
 - What is the declination constraint given our location?
 - What is the constraint in right ascension for this time of year?
 - You can use [staralt](#) again.

Our locations is approximately: 12.97°E 52.41°N 32m
UT+2

Creating our spectroscopic target list

- Open the catalogue in TOPCAT.
- What is the brightness constraint for our telescope?
- What is the declination constraint given our location?
- What is the constraint in right ascension for this time of year?

NAME_SDCAT	RAJ2000_HMS	DEJ2000_HMS	SPEC_SIMBAD	SPEC_SDCAT	G_GAIA	BP-RP_GAIA
	00:54:35.22	19:11:18.32	A1Vn		6.1981	-1.8391
	00:09:20.15	79:42:52.44	A7IV		6.6357	-3.1425
	20:10:45.14	20:29:12.74	B8V		7.4272	-0.0909
	22:02:56.66	44:39:00.53			7.579	-0.252
	21:49:48.89	34:55:00.57	A0		7.9073	0.1132
	01:53:19.25	43:23:21.98	A2		8.3128	0.202
	19:27:09.59	16:26:27.48	A2		9.013	0.1901
	19:03:01.82	42:32:46.14	A2		9.1905	0.2682
	23:01:16.37	44:29:48.04	A3		9.228	0.2118
	19:24:19.22	31:55:35.58	A0		9.2762	0.0005
	21:09:47.38	20:12:29.18	B5		9.4754	-0.1932
	19:36:22.02	19:38:21.50	A0		9.5522	0.1759
	20:04:08.90	16:59:57.27	G5		9.6318	0.379
FB179	21:59:41.98	26:25:57.40	sdO6	sdO	9.6508	-0.4614
	20:46:17.78	28:52:47.31	A0		9.7432	0.2434
	20:35:52.47	31:01:34.93	A0		9.7728	0.1405
	19:56:33.68	29:13:26.66	A0		9.8662	0.2232
	20:10:56.20	22:37:18.64	A0		9.8978	0.1707
	21:04:55.58	46:32:31.16	B9V		9.9396	0.1792
BD+37442	01:58:33.43	38:34:23.85	sdOHe	sdO	9.9485	-0.3984