Gaia 2016 DR1 workshop. Hands-on session

Contributors: ESDC and Gaia SOC. Coordinated by Alcione Mora

AMF-020. 2016-10-30

+ Gaia Archive:
http://gea.esac.esa.int/archive/

+ Sign-in with Cosmos account (top right corner)
Self-registration is needed before the workshop, NON-DPAC MEMBERS ONLY.

+ Topcat download and installation:
http://www.star.bris.ac.uk/~mbt/topcat/#install

+ ADQL cookbook
https://gaia.ac.uk/science/gaia-data-release-1/adql-cookbook

+ Part 1: follow transcribed tutorials in the Gaia Archive. Original location: Help ➔ Tutorials

+ Part 2: Reproduce Gaia DR1 Brown et al. 2016 plots, with some extras
http://www.aanda.org/articles/aa/pdf/forth/aa29512-16.pdf

+ Part 3: Compute SDSS to Gaia proper motions. Following Sergey Koposov during NYC Gaia Sprint

+ Part 4: Python access to the Gaia Archive
Advanced tutorial. Needs Python 3.5, Jupyter Notebook, Numpy and Matplotlib python modules

Tutorial 1: white dwarfs exploration

Location: Gaia Archive ➔Help ➔ Tutorials ➔White dwarfs exploration
Adapted from original tutorial by Jesús Salgado. ESDC

This is a tutorial is focused on possible scientific exploration exercise using the Gaia Archive.

We are going to explore white dwarfs observed by Gaia. First, we have a typical known white dwarfs catalogue at Vizier:

1 Go to Vizier to:
IR photometry of 2MASS/Spitzer white dwarfs
http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=J/ApJ/657/1013&-out.max=50&-out.form=HTML%20Table&-out.add=_r&-out.add=_RAJ,_DEJ&-sort=_r&-oc.form=sexa

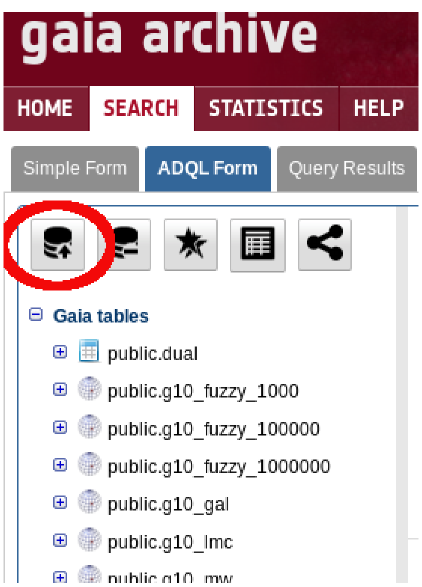
2 Download the catalogue by:
Select in Preferences: max=unlimited, output format = VOTable.
Click on Submit.
A file called vizier_votable.vot will be created.

3 Open in a different tab the Gaia Archive:
http://archives.esac.esa.int/gaia/

4 Sign in Gaia Archive.

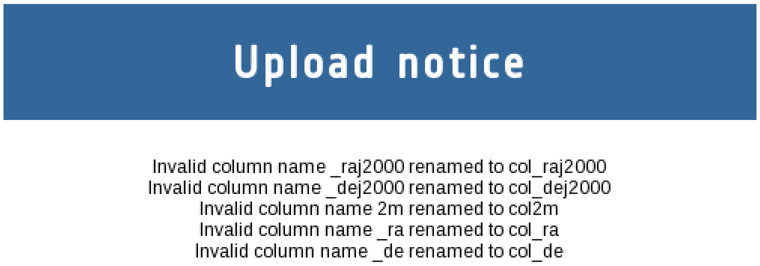
5 Click on SEARCH Tab and, inside it, the ADQL Form.

6 Click on the upload table button

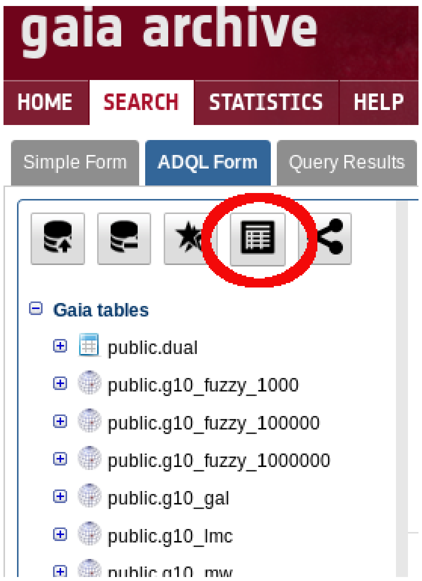


and select the vizier_votable.vot VOTable.

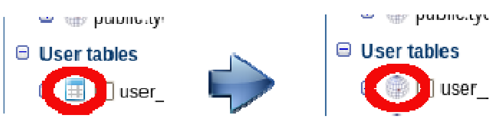
7 Use as table name "dwarfs" and click on the Upload button. As some of the column names of the downloaded VOTable are not compatible with TAP/ADQL, the system will automatically rename them showing the next notice:



8 Select the new table dwarfs (under your schema) and click on the edit table button



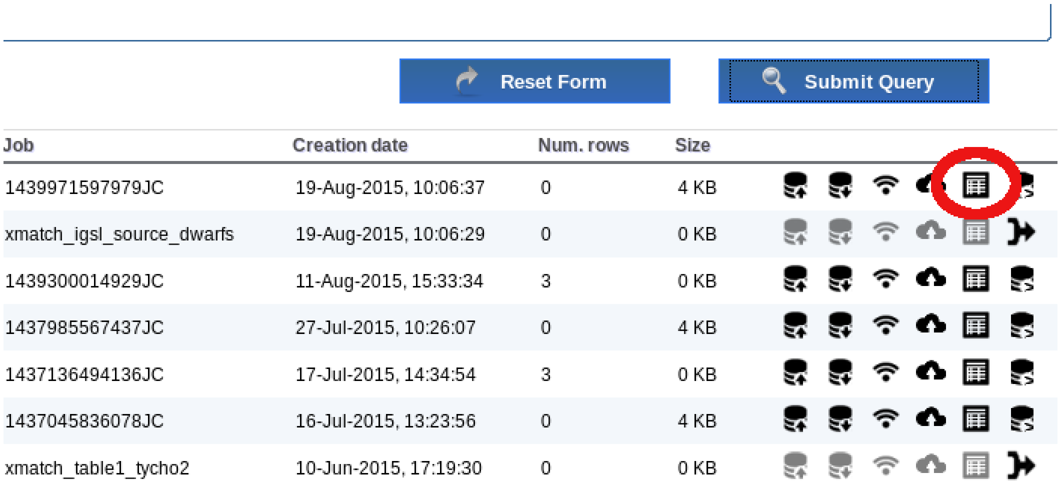
9 For column col_raj2000 select the flag 'Ra'. And for column col_dej2000 select the flag 'Dec'. Then click 'UPDATE'. The table icon in the tree will change to a "Positional indexed table"



10 Inspect the table content inside Gaia Archive. Type in the form at the top of the page:

select top 100 * from user_<your_user_name>.dwarfs

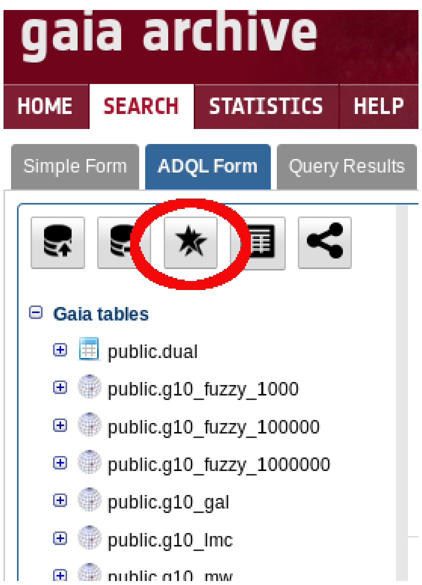
where <your_user_name> is your own username, and click 'Submit Query'. A new job will start to be exectued and, when finished, the table result could be inspected by clicking on the "Display top 2000 results" button. Alternatively, clicking on the table name, the option "show first 20 rows" also can be used without the need of creating a job)



Job	Creation date	Num. rows	Size
1439971597979JC	19-Aug-2015, 10:06:37	0	4 KB
xmatch_igsl_source_dwarfs	19-Aug-2015, 10:06:29	0	0 KB
1439300014929JC	11-Aug-2015, 15:33:34	3	0 KB
1437985567437JC	27-Jul-2015, 10:26:07	0	4 KB
1437136494136JC	17-Jul-2015, 14:34:54	3	0 KB
1437045836078JC	16-Jul-2015, 13:23:56	0	4 KB
xmatch_table1_tycho2	10-Jun-2015, 17:19:30	0	0 KB

Now, we want to obtain metadata of Gaia catalogues for these sources. In order to do that, counterparts should be found by the execution of a crossmatch operation. A positional crossmatch (identification by position) will be executed. More complex algorithms will be offered in future Gaia Archive versions.

11 Click on the crossmatch button

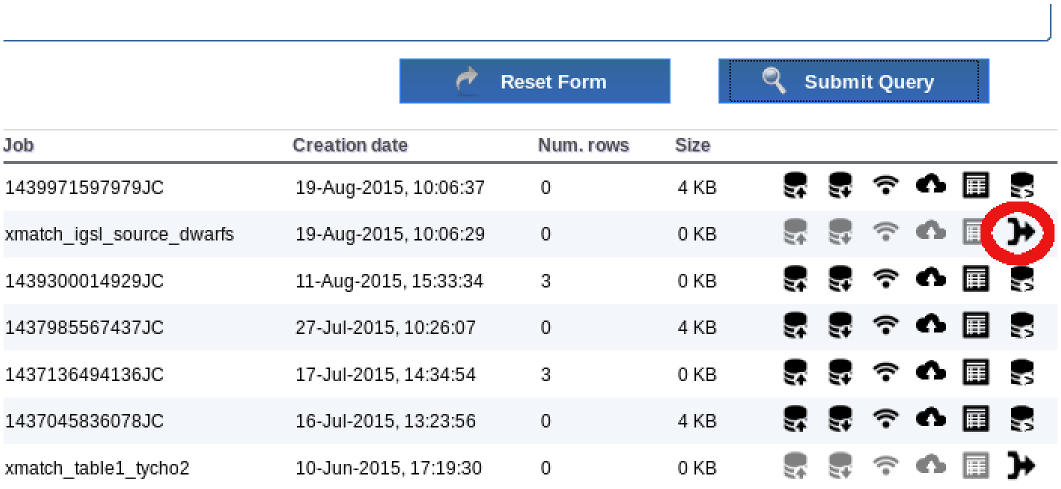


12 Select as Table A: public.igsl_source and as Table B: user_<your_user_name>.dwarfs with a radius of 1 arcsecond

Note: IGSL is a combination catalogue from other external catalogues. It has the size of the expected future Gaia catalogue and a synthetic photometry on band G (Gaia). The calculation of this photometry could fail for peculiar objects. See more info of IGSL at: http://www.cosmos.esa.int/web/gaia/iow_20131008

13 Execute the crossmatch. A new job will start. At the end of the execution, a new join table (called xmatch_igsl_source_dwarfs by default) will be created between IGSL and "dwarfs" catalogues.

14 When finished, click on the "Show join query"



Job	Creation date	Num. rows	Size
1439971597979JC	19-Aug-2015, 10:06:37	0	4 KB
xmatch_igsl_source_dwarfs	19-Aug-2015, 10:06:29	0	0 KB
1439300014929JC	11-Aug-2015, 15:33:34	3	0 KB
1437985567437JC	27-Jul-2015, 10:26:07	0	4 KB
1437136494136JC	17-Jul-2015, 14:34:54	3	0 KB
1437045836078JC	16-Jul-2015, 13:23:56	0	4 KB
xmatch_table1_tycho2	10-Jun-2015, 17:19:30	0	0 KB

This query is an example on how to contain all the metadata of the two catalogues. Reduce the content of the metadata by replacing the SELECT part of the ADQL sentence as follows

SELECT a."source_id", b."dwarfs_oid", b."name", a."ra", a."dec", b."col_raj2000", b."col_dej2000", a."mag_g", b."f_hmagc", b."f_jmagc", b."f_kmagc", b."hmag2", b."hmagc", b."jmag2", b."jmagc", b."kmagc", b."ksmag2"

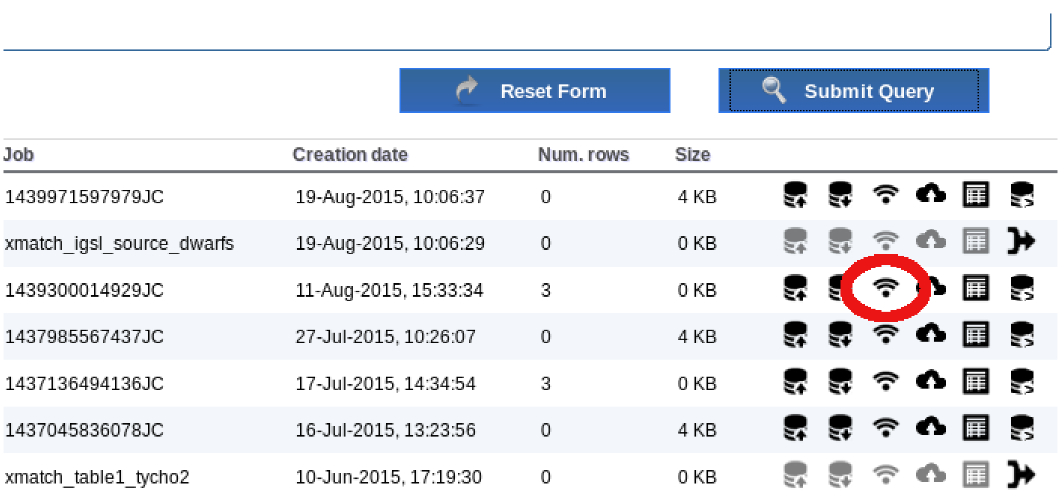
Note: FROM and WHERE conditions must be preserved
This query will contain the ids of the source in both catalogues (to explore possible duplications), the name of the dwarf stars and magnitudes from both catalogues.

Click on "Submit Query" to launch a new job.

15 Open the VO application Topcat:
http://www.star.bris.ac.uk/~mbt/topcat/topcat-lite.jnlp

16 Send result of the latest created table to Topcat through SAMP.

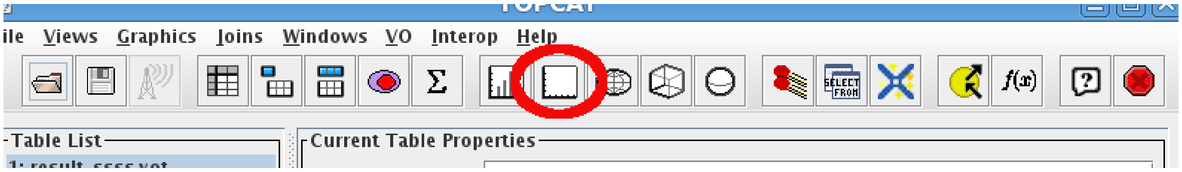
Click on SAMP button



Job	Creation date	Num. rows	Size
1439971597979JC	19-Aug-2015, 10:06:37	0	4 KB
xmatch_igsl_source_dwarfs	19-Aug-2015, 10:06:29	0	0 KB
1439300014929JC	11-Aug-2015, 15:33:34	3	0 KB
1437985567437JC	27-Jul-2015, 10:26:07	0	4 KB
1437136494136JC	17-Jul-2015, 14:34:54	3	0 KB
1437045836078JC	16-Jul-2015, 13:23:56	0	4 KB
xmatch_table1_tycho2	10-Jun-2015, 17:19:30	0	0 KB

During this process, some security authorization should be granted. Also, your LDAP credentials will be asked by Topcat as an extra security measure.

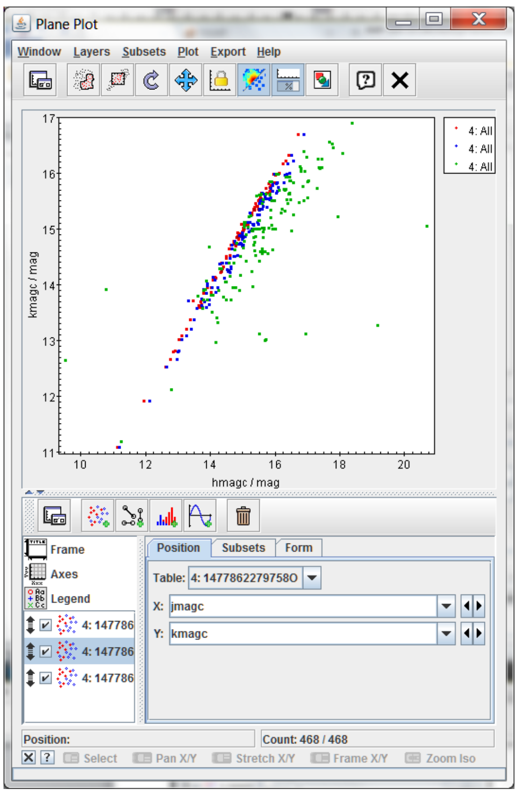
17 Click on the "Plane Plotting Window" button



18 Add, by clicking on the "Add a new positional control to the stack", to create three plots: (select also the table on the widget)

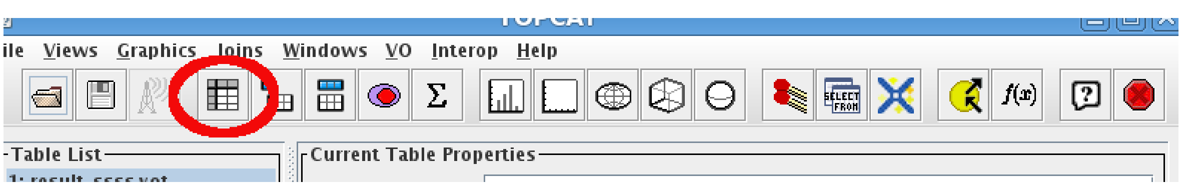
kmagc versus hmagc
kmagc versus jmagc
mag_g versus jmagc

The result should be as follows

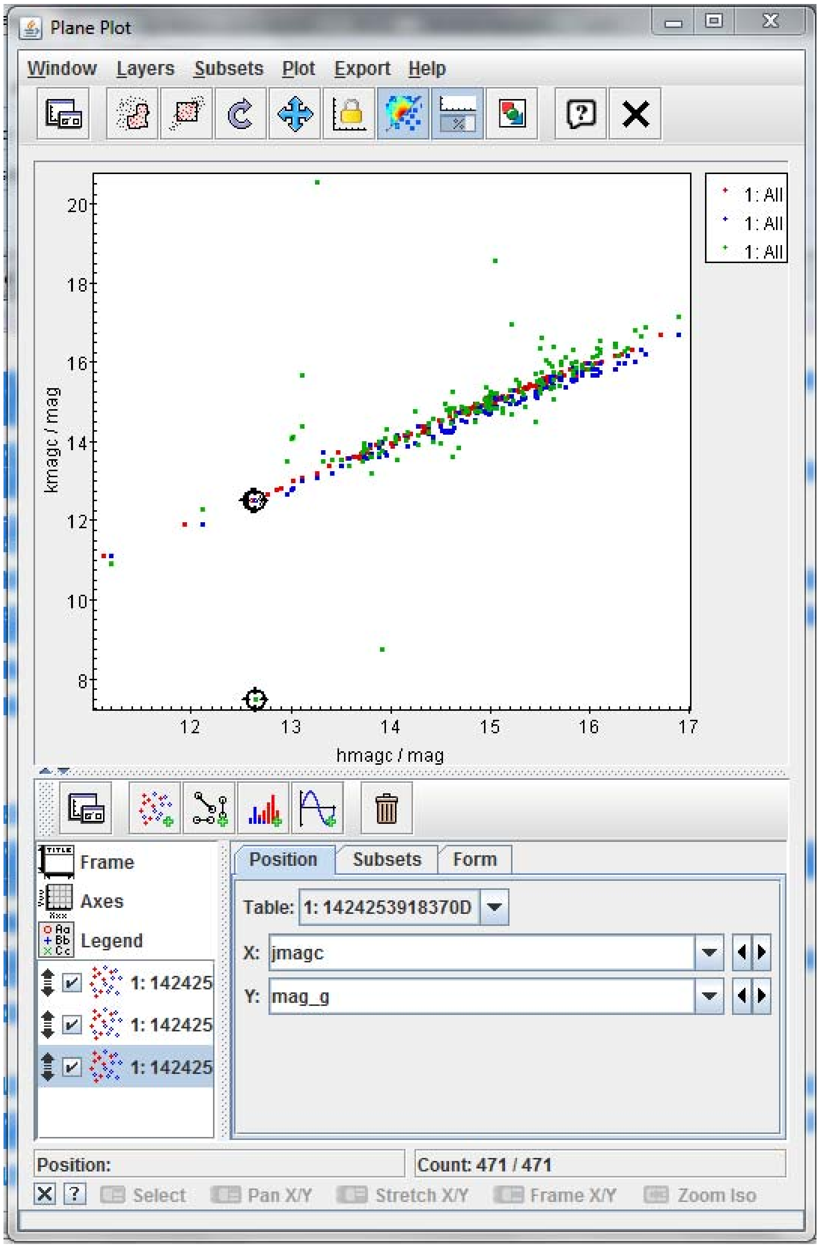


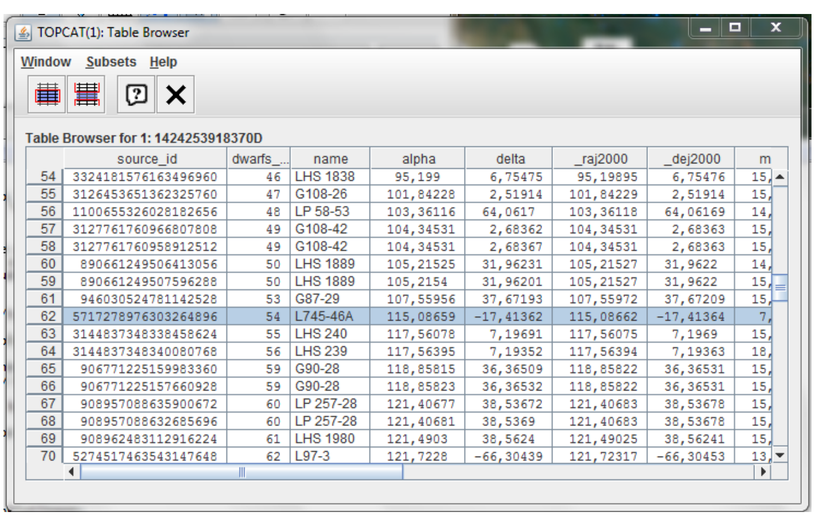
Most of the points are located in a clear line except some few up and two of them clearly below it. That could imply that they emits on the G band more than the expected (as the G band is synthetic for IGSL, this is not fully clear) or that the crossmatch is not fully correct for these sources.

Click on the "Display table cell data" button on the main Topcat window.



19 Click on the Plan Plot, one at once, on the two strange sources. Once you click on one source, the two plots will be synchronized





TOPCAT(1): Table Browser

Window Subsets Help

Table Browser for 1: 1424253918370D

	source_id	dwarfs_...	name	alpha	delta	_raj2000	_dej2000	m
54	3324181576163496960	46	LHS 1838	95,199	6,75475	95,19895	6,75476	15,
55	3126453651362325760	47	G108-26	101,84228	2,51914	101,84229	2,51914	15,
56	1100655326028182656	48	LP 58-53	103,36116	64,0617	103,36118	64,06169	14,
57	3127761760966807808	49	G108-42	104,34531	2,68362	104,34531	2,68363	15,
58	3127761760958912512	49	G108-42	104,34531	2,68367	104,34531	2,68363	15,
60	890661249506413056	50	LHS 1889	105,21525	31,96231	105,21527	31,9622	14,
59	890661249507596288	50	LHS 1889	105,2154	31,96201	105,21527	31,9622	15,
61	946030524781142528	53	G87-29	107,55956	37,67193	107,55972	37,67209	15,
62	5717278976303264896	54	L745-46A	115,08659	-17,41362	115,08662	-17,41364	7,
63	3144837348338458624	55	LHS 240	117,56078	7,19691	117,56075	7,1969	15,
64	3144837348340080768	56	LHS 239	117,56395	7,19352	117,56394	7,19363	18,
65	906771225159983360	59	G90-28	118,85815	36,36509	118,85822	36,36531	15,
66	906771225157660928	59	G90-28	118,85823	36,36532	118,85822	36,36531	15,
67	908957088635900672	60	LP 257-28	121,40677	38,53672	121,40683	38,53678	15,
68	908957088632685696	60	LP 257-28	121,40681	38,5369	121,40683	38,53678	15,
69	908962483112916224	61	LHS 1980	121,4903	38,5624	121,49025	38,56241	15,
70	5274517463543147648	62	L97-3	121,7228	-66,30439	121,72317	-66,30453	13,

20 Checking the name column in the Table browser, the two sources are:

LTT 4816
L745-46A

The first object has been identified as a pulsating white dwarf and the second as a simple white dwarf.

The analysis of the result could suggest a failure on the calculation of the synthetic G magnitude for these objects or some peculiarity on the emission.

http://simbad.u-strasbg.fr/simbad/sim-id?Ident=LTT%204816

Tutorial 2: cluster analysis

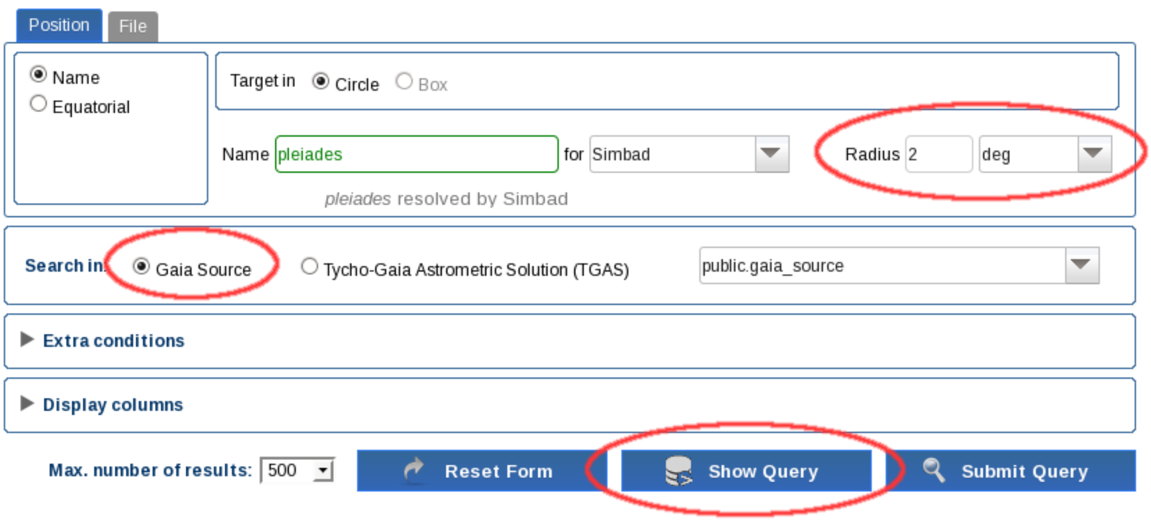
Location: Gaia Archive →Help → Tutorials →Cluster analysis
Adapted from original tutorial by Raúl Gutiérrez. ESDC

This is a tutorial is focused on possible scientific exploration exercise using the Gaia Archive.

We are going to explore a known cluster as the Pleiades (M45) using Gaia data. First, we are going to retrieve all the available data in the region of interest:

1 As we are going to use the private storage area, we have to sign in the archive (not needed if this is done after first tutorial).

2 Go to Search -> Simple Form. In positional search, enter "pleiades" in the field "Name". Once it is resolved, select a 2 degrees search radius and make sure "Gaia source" is selected. Click on "Show Query" button.



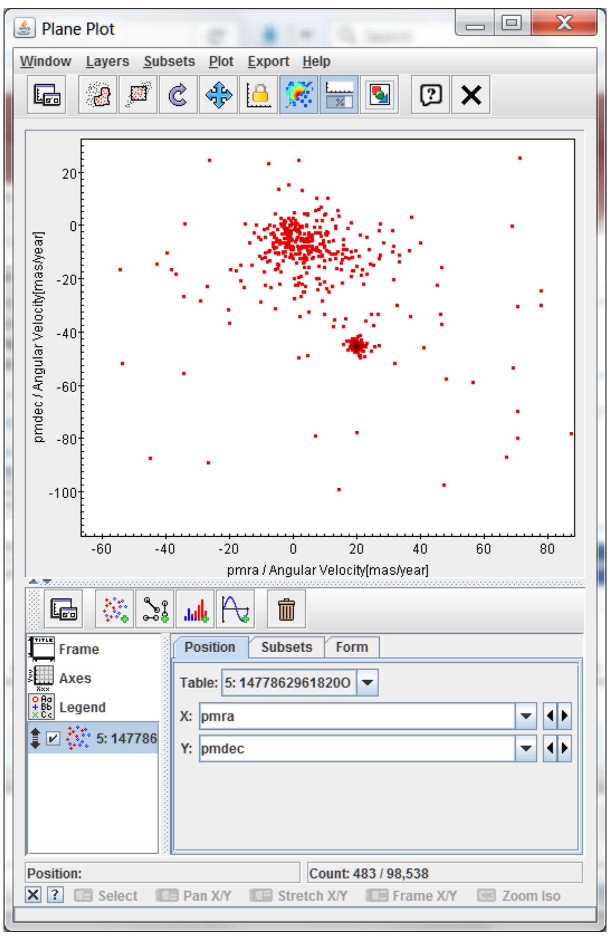
3 Enter "m45" in the field Job name. Edit the ADQL query and remove the TOP 500 restriction. The query should be:

```
SELECT *
FROM gaiadr1.gaia_source
WHERE
CONTAINS(POINT('ICRS',gaiadr1.gaia_source.ra,gaiadr1.gaia_source.dec),CIRCLE('ICRS',56.75,24.1167,2))=1
```

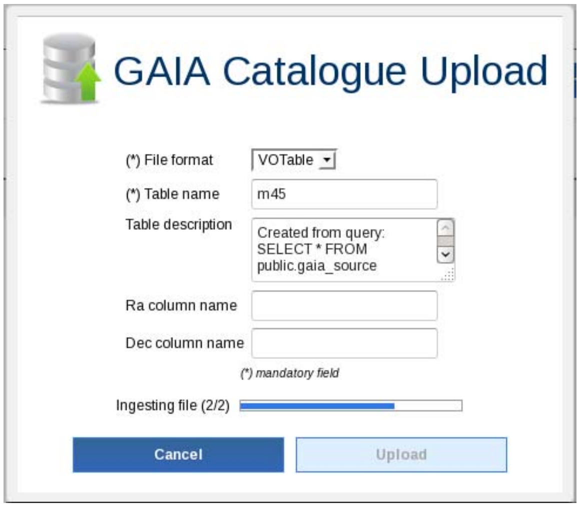
Execute the query. Around 1e5 results are found (the query could take some time).

4 The number of results are small enough to be represented by a local application. Open Topcat and click Send to SAMP button. As the job is private, Topcat could ask for your credentials (as it could have been registered in previous tutorial).

Once the data have been loaded, you could show the results in the sphere, or create a proper motion plane plotting window to identify the cluster (cuts modified to pmra in [-60,80], pmdec in [-120,30] mas/yr).



5 Go to the archive and filter data by quality. For that, upload the table to your local environment clicking in the upload table icon on the job row).

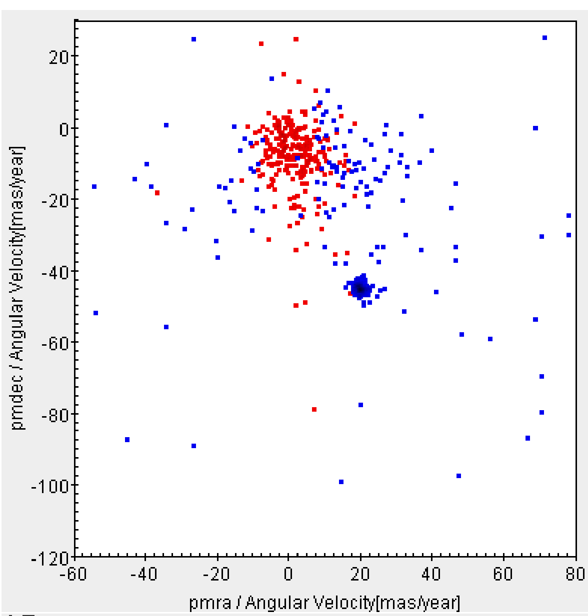


Hint: In this upload operation, no heavy network traffic occurs between the user machine and the server. Job results are stored in the server, so data to be ingested in the user area from job results do not leave the server in the upload process.

Enter "m45PmFilter" as job name and perform the next query (do not forget to replace <username> with your user name):

SELECT * FROM user_<username>.m45
WHERE abs(pmra_error/pmra)<0.10
AND abs(pmdec_error/pmdec)<0.10
AND pmra IS NOT NULL AND abs(pmra)>0
AND pmdec IS NOT NULL AND abs(pmdec)>0;

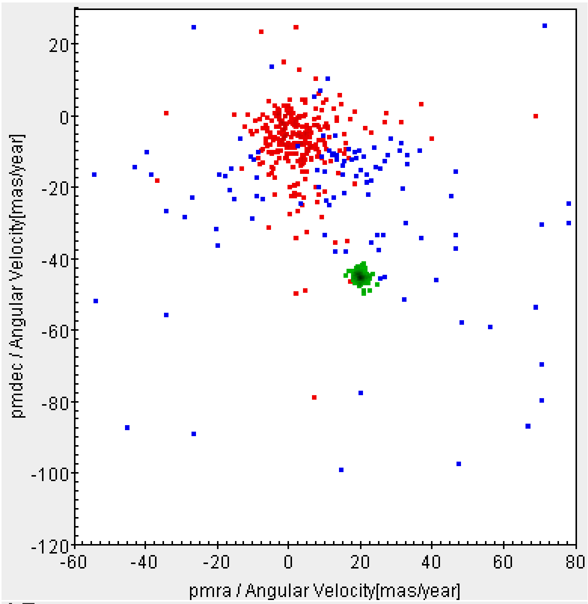
Execute the query. You can send the results to Topcat via SAMP and overplot the new results in the same panel.



6 Create the table m45pmfilter from m45PmFilter job results. Now we are going to take the candidate objects to be in the cluster. Based on the proper motion plot, we execute the next filter:

SELECT * FROM user_<username>.m45pmfilter
WHERE pmra BETWEEN 15 AND 25
AND pmdec BETWEEN -55 AND -40;

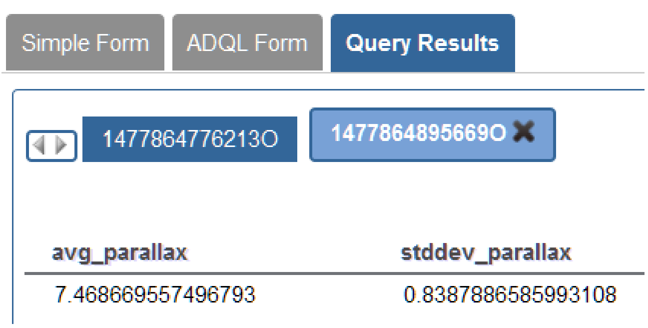
Name the job as m45cluster and execute the job. You could send the results to Topcat and plot over the previous proper motion plots.



7 Using the ADQL interface we can perform analysis queries on the results. Create the table m45cluster from m45cluster job results. Name the job as m45clusterParallaxAvg. Execute the next query:

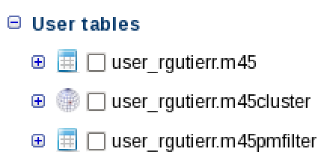
SELECT avg(parallax) as avg_parallax, stddev(parallax) as stddev_parallax
FROM user_<username>.m45cluster

Execute the query and show the results using the corresponding button from the job results.



8 Now, we want to add information from other catalogues. To do so, we will make use of the crossmatch functionality of the archive. We need to identify which columns of our table have the geometrical information (ra and dec in our case). Select m45cluster table by checking its check box and click on the Edit table button. Find column ra and set the flag Ra. Find the column dec and set the flag Dec. Click on Update button.

This action will create a positional index on Ra and Dec, and the table will be identified as a geometrical table. This allows the use of the crossmatch functionality.



9 Click on crossmatch button. Select user_<username>.m45cluster for TableA and gaiadr1.tmass_original_valid for TableB. Click on Execute.

This will create a new job of crossmatch type and a new table called xmatch_m45cluster_tmass_original_valid. This table is a join table between m45cluster and tmass_original_valid tables. A helper function is available in the crossmatch job to create a join query between the two tables.

Hint: Positional crossmatch is performed for the time being. All counterparts falling in the search radius are considered matches. Distance is provided for further filtering.

10 To create a table with all the information from the two tables, click in the Show join query button of the crossmatch job. A query like this is automatically loaded in the ADQL query panel:

SELECT c."dist", a."dec", a."m45cluster_oid", a."ra", a."astrometric_chi2_ac", a."astrometric_chi2_al", a."astrometric_delta_q", a."astrometric_excess_noise", a."astrometric_excess_noise_sig", a."astrometric_go_f", a."astrometric_n_obs_ac", a."astrometric_n_obs_al", a."astrometric_n_outliers_ac", a."astrometric_n_outliers_al", a."astrometric_params_solved", a."astrometric_primary_flag", a."astrometric_priors_used", a."astrometric_rank_defect", a."astrometric_relegation_factor", a."astrometric_weight_ac", a."astrometric_weight_al", a."dec_error", a."dec_parallax_corr", a."dec_pmdec_corr", a."dec_pmra_corr", a."dec_pmradial_corr", a."m45_oid", a."m45pmfilter_oid", a."matched_observations", a."parallax", a."parallax_error", a."parallax_pmdec_corr", a."parallax_pmra_corr", a."parallax_pmradial_corr", a."phot_bp_mean_flux", a."phot_bp_mean_flux_error", a."phot_bp_mean_mag", a."phot_bp_n_obs", a."phot_g_mean_flux",

a."phot_g_mean_flux_error", a."phot_g_mean_mag", a."phot_g_n_obs", a."phot_rp_mean_flux", a."phot_rp_mean_flux_error", a."phot_rp_mean_mag", a."phot_rp_n_obs", a."phot_variable_flag", a."pmdec", a."pmdec_error", a."pmdec_pmradial_corr", a."pmra", a."pmradial", a."pmradial_error", a."pmra_error", a."pmra_pmdec_corr", a."pmra_pmradial_corr", a."ra_dec_corr", a."radial_velocity", a."radial_velocity_constancy_probability", a."radial_velocity_error", a."ra_error", a."random_index", a."ra_parallax_corr", a."ra_pmdec_corr", a."ra_pmra_corr", a."ra_pmradial_corr", a."ref_epoch", a."scan_direction_mean_k1", a."scan_direction_mean_k2", a."scan_direction_mean_k3", a."scan_direction_mean_k4", a."scan_direction_strength_k1", a."scan_direction_strength_k2", a."scan_direction_strength_k3", a."scan_direction_strength_k4", a."solution_id", a."source_id", b."dec", b."ra", b."designation", b."err_ang", b."err_maj", b."err_min", b."ext_key", b."h_m", b."h_msigcom", b."j_date", b."j_m", b."j_msigcom", b."k_m", b."k_msigcom", b."ph_qual", b."tmass_oid"
FROM user_<username>.m45cluster AS a, public.tmass_original_valid AS b, user_<username>.xmatch_m45cluster_tmass_original_valid AS c
WHERE (c.m45cluster_m45cluster_oid = a.m45cluster_oid AND c.tmass_original_valid_tmass_oid = b.tmass_oid)

Hint: All the columns of the joined tables are shown for convenience. This way, the user has only to remove columns to get the desired output.

Name the job as xmatch and execute it. Show the results and verify that both Gaia and 2Mass data are present.

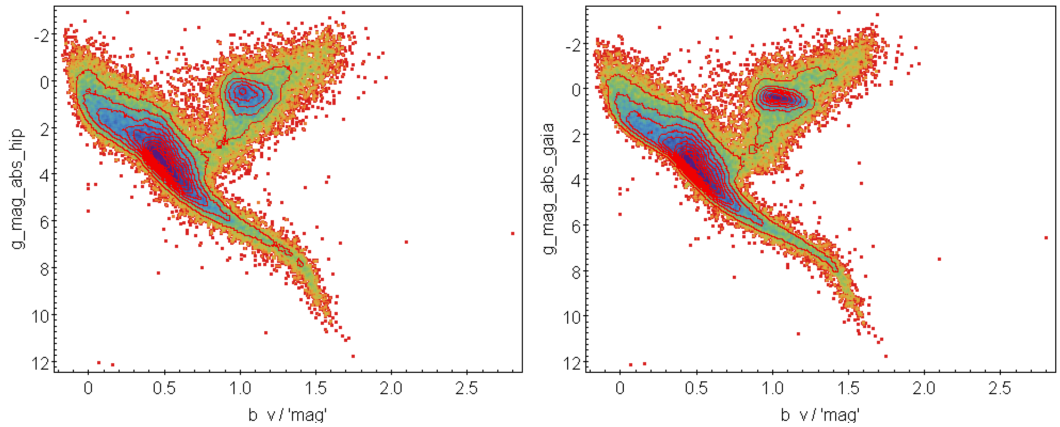
11. Now, we are going to use the sharing functionality of the archive to share these results with some colleagues. First, create a new table cluster_2mass from the previous job (a warning of automatic column renaming may appear). Go to SHARE > Groups and create a group called cluster. The new group appears in the tree. Select the group and click on Edit. Use the User to include field, include your colleague’s user name and click on add. Repeat this to add any other collaborator. Then click on Update. The group has been updated with the new members.

Return to the ADQL search page, check the cluster_2mass table and click on Share button. Select group cluster, click Share and then Update. You will see that a little Share icon is added to the table icon. From this moment, the users in cluster group will be notified and they will have access to this table.

Hint: Size of shared tables is only accounted for the quota of the table owner.

Exercise 1: Hipparcos and Gaia DR1 B-V HR diagrams for well-behaved Hipparcos stars. Figure 3a, 3b
Adapted from figures and queries in Gaia DR1 paper, Brown et al. (2016)

Objective: Construct the Hipparcos absolute magnitude vs B-V HR diagram for good quality objects



Conditions:
- Parallax_tgas / Error_parallax_tgas >= 5
- Parallax_hipparcos / Error_parallax_hipparcos >= 5
- 0 < error_B-V_hipparcos <= 0.05
- error_G_mag <= 0.05

Hints:
+ TGAS data (including G mag) are in table gaiadr1.tgas_source
+ Hipparcos data are in table public.hipparcos_newreduction
+ Logarithms in ADQL: log (natural), log10 (decimal)
+ error_mag ≈ 2.5 / log(10) * flux_error / flux
+ absolute magnitude = apparent_magnitude + 5 * log10(parallax) – 10
+ Join both tables using the Hipparcos number (same column in both tables). Join example:

```
select top 10 * from gaiadr1.tgas_source
inner join public.hipparcos_newreduction
on gaiadr1.tgas_source.hip = public.hipparcos_newreduction.hip
```

+ A positional cross-match would produce different results (no proper motion propagation)
+ Use 10 contours log scale. Smoothness factor 10
+ Topcat output might be slightly different than above (depends on window size and other options)

Solution:

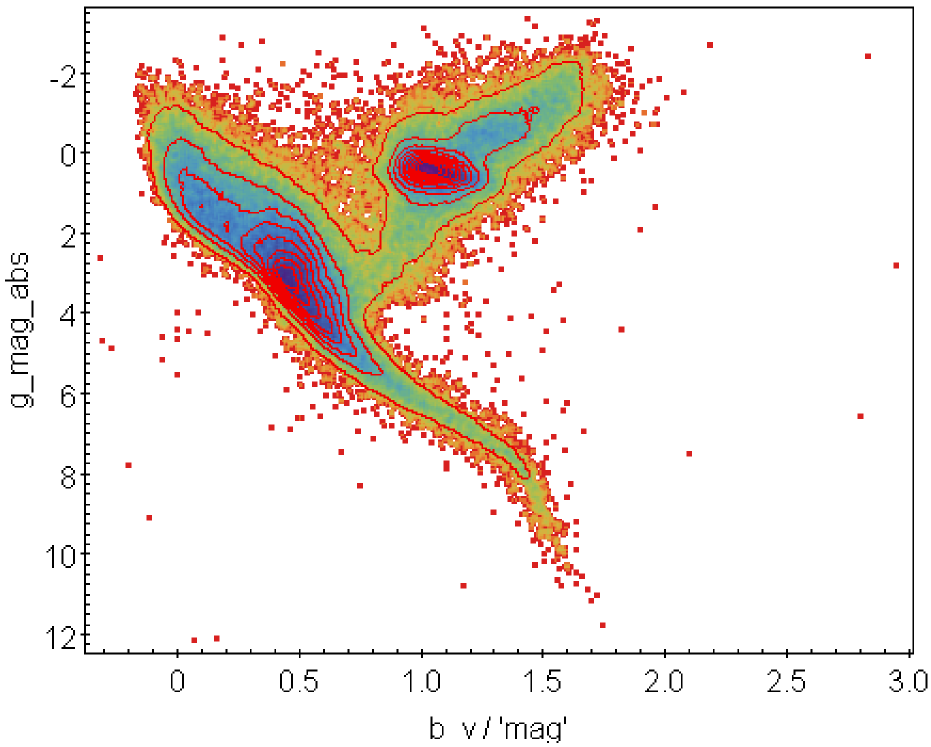
a) ADQL query (43546 rows):

```
select gaia.source_id, gaia.hip,
        gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs_gaia,
        gaia.phot_g_mean_mag+5*log10(hip.plx)-10 as g_mag_abs_hip,
        hip.b_v
from gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
        on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
        hip.plx/hip.e_plx >= 5 and
        hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
        2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05
```

b) Export to Topcat (via SAMP) and create a scatter plot, invert y-axis, use a density map option with the SRON colour palette and add 10 contours in log-scale, smoothing factor 10

Exercise 2: Gaia DR1 B-V HR diagram. Figure 3c
Adapted from figures and queries in Gaia DR1 paper, Brown et al. (2016)

Objective: Construct the Gaia absolute magnitude vs B-V HR diagram for good quality objects, not necessarily good in Hipparcos as well



Conditions:
- Parallax_tgas / Error_parallax_tgas >= 5
- 0 < error_B-V_hipparcos <= 0.05
- error_G_mag <= 0.05

Hints:
+ TGAS data (including G mag) are in table gaiadr1.tgas_source
+ Hipparcos data are in table public.hipparcos_newreduction
+ Logarithms in ADQL: log (natural), log10 (decimal)
+ error_mag ≈ 2.5 / log(10) * flux_error / flux. Example calculation:
select top 10 2.5 / log(10) * phot_g_mean_flux_error / phot_g_mean_flux as relative_flux_error from gaiadr1.tgas_source
+ absolute magnitude = apparent_magnitude + 5 * log10(parallax) - 10
+ Join both tables using the Hipparcos number (same column in both tables)
+ A positional cross-match would produce different results (no proper motion propagation)
+ Use 10 contours log scale. Smoothness factor 10
+ Topcat output might be slightly different than above (depends on window size and other options)

Solution:

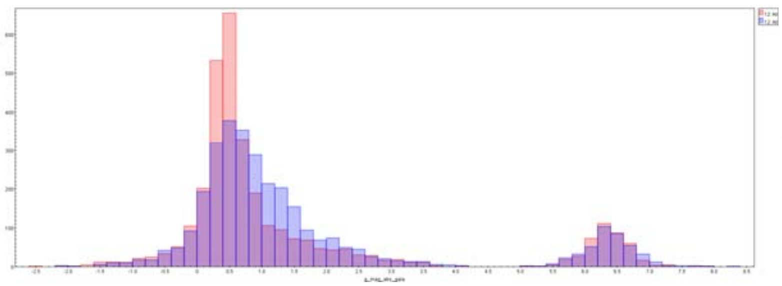
a) ADQL query (74771 rows):

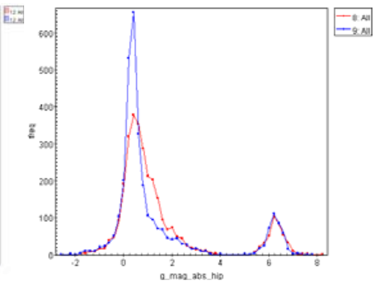
```
select gaia.source_id, gaia.hip,
gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs,
	hip.b_v
from gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
	on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
	hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
	2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05
```

b) Export to Topcat (via SAMP) and create a scatter plot, invert y-axis, use a density map option with the SRON colour palette and add 10 contours in log-scale, smoothing factor 10

Exercise 3: Red clump giants (and some red dwarfs) absolute magnitudes. Figure 4. Individual points
Adapted from figures and queries in Gaia DR1 paper, Brown et al. (2016)

Objective: See how defined is the red clump of giants in terms of absolute magnitude, the narrower the better. For that, the absolute magnitudes in a small range of colours will be computed and compared for TGAS and Hipparcos





Conditions:
- Parallax_tgas / Error_parallax_tgas >= 5
- Parallax_hipparcos / Error_parallax_hipparcos >= 5
- 0 < error_B-V_hipparcos <= 0.05
- 1.0 =< B-V =< 1.1
- error_G_mag <= 0.05

Hints:
+ TGAS data (including G mag) are in table gaiadr1.tgas_source
+ Hipparcos data are in table public.hipparcos_newreduction
+ Logarithms in ADQL: log (natural), log10 (decimal)
+ error_mag ≈ 2.5 / log(10) * flux_error / flux
+ absolute magnitude = apparent_magnitude + 5 * log10(parallax) - 10
+ Topcat output might be slightly different than above (depends on window size and other options)

Extra: Compute the histograms directly in the Gaia archive.
+ Computed directly in the archive ➔ no bulk download ➔ <u>very important for Gaia DR2!</u>
+ Hint: aggregate the data from the first query in bins of 0.2 mag using an integral index variable of value floor(g_mag * 5) and subsequently divide by 5.
+ Hint: the subquery needs a name
+ Example histogram of TGAS G magnitudes with bin size of 0.25 mag (57 rows):

```
select index / 4 as g_mag, count(*) as n from (
        select floor(phot_g_mean_mag * 4) as index from gaiadr1.tgas_source) as subquery
group by index
```

Solution:
a) ADQL query (3174 rows):

```
select gaia.source_id, gaia.hip,
	gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs_gaia,
	gaia.phot_g_mean_mag+5*log10(hip.plx)-10 as g_mag_abs_hip
from gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
	on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
	hip.plx/hip.e_plx >= 5 and
	hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
	hip.b_v >= 1.0 and hip.b_v <= 1.1 and
	2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05
```

b) Export to Topcat via SAMP. Generate histogram plot. Bin size = 0.2 mag. Superimpose Gaia and Hipparcos absolute magnitudes

Extra: Compute the histograms directly in the Gaia archive.

c1) Histogram Hipparcos with 0.2 mag bin size
ADQL query (47 rows):

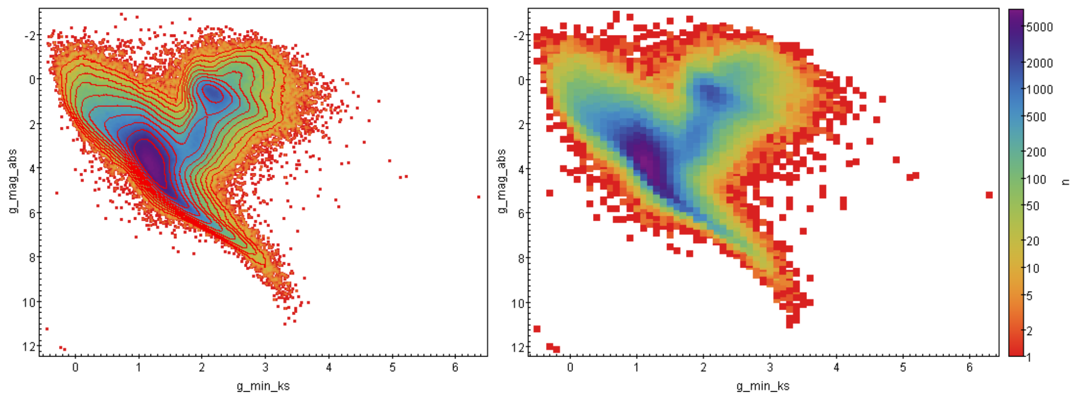
```
select g_mag_abs_hip_index / 5. as g_mag_abs_hip, count(g_mag_abs_hip_index) as freq from (select
	floor((gaia.phot_g_mean_mag+5*log10(hip.plx)-10) * 5) as g_mag_abs_hip_index
from gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
	on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
	hip.plx/hip.e_plx >= 5 and
	hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
	hip.b_v >= 1.0 and hip.b_v <= 1.1 and
	2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05) query
group by g_mag_abs_hip_index
order by g_mag_abs_hip
```

c2) Histogram Gaia with 0.2 mag bin size
ADQL query (46 rows):

d) Export to Topcat via SAMP. Generate 2D scatter plot and superimpose both curves

Exercise 4: Gaia DR1 G-Ks HR diagram. Figure 5
Adapted from figures and queries in Gaia DR1 paper, Brown et al. (2016)

Objective: Construct the Gaia-Hipparcos absolute magnitude vs G-Ks HR diagram for good quality objects



Conditions:
- Parallax_tgas / Error_parallax_tgas >= 5
- error_G_mag <= 0.05 (redundant condition, can be omitted)
- error_G-Ks <= 0.05
- 2MASS photometry quality is best: 'AAA'

Hints:
+ TGAS data (including G mag) are in table gaiadr1.tgas_source
+ 2MASS data are in table gaiadr1. gaiadr1.tmass_original_valid
+ Join using Gaia-2MASS cross-match best neighbours are in table gaiadr1.tmass_best_neighbour
+ A positional cross-match would produce different results (no proper motion propagation)
+ Logarithms in ADQL: log (natural), log10 (decimal)
+ error_mag ≈ 2.5 / log(10) * flux_error / flux
+ error_colour = sqrt(error_mag_1 ^ 2 + error_mag_2 ^ 2)
+ absolute magnitude = apparent_magnitude + 5 * log10(parallax) – 10
+ 2MASS – cross-match key: tmas_oid
+ Make sure to use single quotes for the 'AAA' condition and not other characters (e.g. accents).
+ Use 10 contours log scale. Smoothness factor 10
+ Topcat output might be slightly different than above (depends on window size and other options)

Extra: Compute the HR diagram as a 2D histogram directly in the Gaia archive.
+ Computed directly in the archive ➔ no bulk download ➔ <u>very important for Gaia DR2!</u>
+ Hint: aggregate the data from the first query in bins of 0.1 mag in colour and absolute magnitude using an integral index variable of value floor(g_mag * 10) and subsequently divide by 10. Two subquery levels are needed.
+ Hint: the subquery needs a name
+ Example histogram of TGAS G magnitudes with bin size of 0.25 mag (57 rows):

```
select index / 4 as g_mag, count(*) as n from (
        select floor(phot_g_mean_mag * 4) as index from gaiadr1.tgas_source) as subquery
group by index
```

Solution:

a) ADQL query (1004207 rows):

```
select gaia.source_id,
	gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs,
	gaia.phot_g_mean_mag-tmass.ks_m as g_min_ks
from gaiadr1.tgas_source as gaia
inner join gaiadr1.tmass_best_neighbour as xmatch
	on gaia.source_id = xmatch.source_id
inner join gaiadr1.tmass_original_valid as tmass
	on tmass.tmass_oid = xmatch.tmass_oid
where gaia.parallax/gaia.parallax_error >= 5 and
	 ph_qual = 'AAA' and
	sqrt( power(2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux,2) ) <= 0.05 and
	sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux,2)
	+ power(tmass.ks_msigcom,2)) <= 0.05
```

b) Export to Topcat (via SAMP) and create a scatter plot, invert y-axis, use a density map option with the SRON colour palette and add 10 contours in log-scale, smoothing factor 10

Extra: Compute the HR diagram as a 2D histogram directly in the Gaia archive (5550 rows).

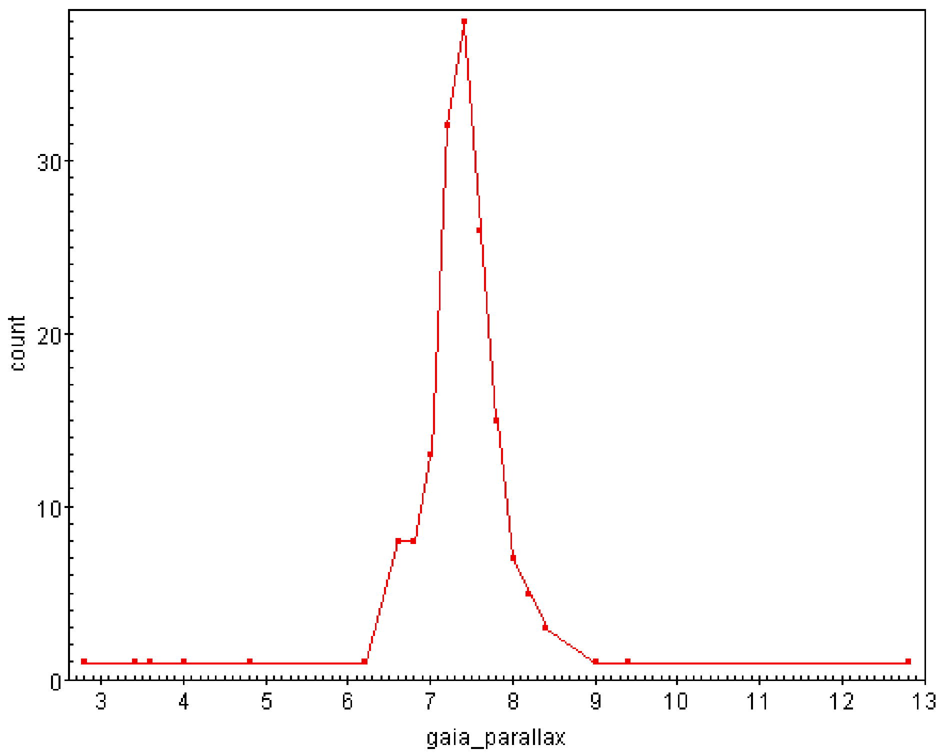
c) ADQL query (3078 rows):

```
select
	g_min_ks_index / 10 as g_min_ks,
	g_mag_abs_index / 10 as g_mag_abs,
	count(*) as n
from (
	select gaia.source_id,
		floor((gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10) * 10) as g_mag_abs_index,
		floor((gaia.phot_g_mean_mag-tmass.ks_m) * 10) as g_min_ks_index
	from gaiadr1.tgas_source as gaia
	inner join gaiadr1.tmass_best_neighbour as xmatch
		on gaia.source_id = xmatch.source_id
	inner join gaiadr1.tmass_original_valid as tmass
		on tmass.tmass_oid = xmatch.tmass_oid
	where gaia.parallax/gaia.parallax_error >= 5 and
		ph_qual = 'AAA' and
		sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux,2)  )  <=  0.05
and
		sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux,2)
			+ power(tmass.ks_msigcom,2)) <= 0.05
	)as subquery
group by g_min_ks_index, g_mag_abs_index
```

d) Export to Topcat (via SAMP) and create a scatter plot, invert y-axis, use axiliary axis for shading, select the column n and apply a logarithmic scale. Use a square symbol and increase its size.

Exercise 5: Pleiades cluster parallax. Fig. 9 (modified as 2D scatter plot)
Adapted from figures and queries in Gaia DR1 paper, Brown et al. (2016)

Objective: Create a histogram of parallaxes for the Pleiades cluster



Conditions:
- Stars within a circle centred on RA = 56.75, DEC = 24.12, radius = 5 deg
- Proper motion selection centre: (pmra, pmdec) = (20.5, - 45.5) mas/yr
- Proper motion maximum difference (modulus) +- 6 mas/yr

Hints:
+ Computed directly in the archive ➔ no bulk download ➔ <u>very important for Gaia DR2!</u>
+ Use a subquery to get all parallaxes and aggregate them in bins of 0.2 mas size (see Exercise 3)
+ Compute the modulus of the proper motion difference vector, it has to be smaller than 6 mas/yr
+ Use ADQL contains, point and circle functions to construct a cone search as in following example:
select gaia.source_id, ra, dec, pmra, pmdec
from gaiadr1.tgas_source as gaia
where contains(point('ICRS',gaia.ra,gaia.dec),circle('ICRS',30,60,0.1)) = 1
+ Make sure to use the same label (e.g. 'ICRS') and single quotes for both the point and circle
+ Use a parallax index = floor (parallax * 5), and then group on it and divide by 5 to recover the parallax

Solution:

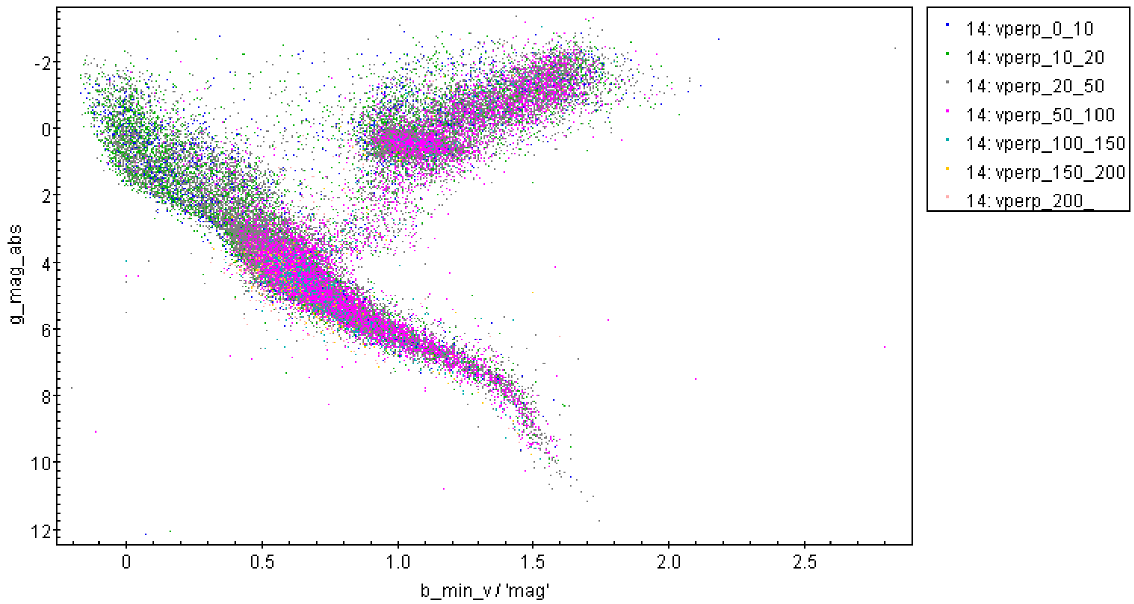
a) ADQL query (19 rows):

```
select gaia_parallax_index / 5. as gaia_parallax, count(gaia_parallax_index) as n from (
        Select floor(gaia.parallax * 5) as gaia_parallax_index
        from gaiadr1.tgas_source as gaia
        where contains(point('ICRS',gaia.ra,gaia.dec),circle('ICRS',56.75,24.12,5)) = 1
        and sqrt(power(gaia.pmra-20.5,2)+power(gaia.pmdec+45.5,2)) < 6.0) query
group by gaia_parallax_index
order by gaia_parallax
```

b) Export to Topcat via SAMP. Generate 2D scatter plot

Exercise 6: HR diagram with transverse velocity information. Fig. 6
Adapted from figures and queries in Gaia DR1 paper, Brown et al. (2016)

Objective: Create an HR diagram separating different groups of stars based on the transverse velocity



Conditions:
- Three queries are needed: Hipparcos stars, Tycho stars with good B-V colour and Tycho stars with bad B-V but good URAT1 (APASS) B-V
- Parallax_tgas / Error_parallax_tgas >= 5
- error_G_mag <= 0.05
- error (B-V) < 0.05
- parallax >= 10.0 mas OR proper_motion > 200 mas/yr OR G <= 7.5

Hints:
+ Any TGAS star has either a Hipparcos or a Tycho identifier, but not both
+ TGAS data (including G mag) are in table gaiadr1.tgas_source
+ Hipparcos data are in table public.hipparcos_newreduction
+ Tycho2 data are in table public.tycho2
+ URAT1 data are in table gaiadr1.urat1_original_valid
+ Gaia-URAT1 cross-match best neighbours are in table gaiadr1.urat1_best_neighbour
+ A positional cross-match would produce different results (no proper motion propagation)
+ Logarithms in ADQL: log (natural), log10 (decimal)
+ error_mag ≈ 2.5 / log(10) * flux_error / flux
+ error_colour = sqrt(error_mag_1 ^ 2 + error_mag_2 ^ 2)
+ absolute magnitude = apparent_magnitude + 5 * log10(parallax) – 10
+ transverse velocity (km/s) = proper motion (mas/yr) / parallax (mas) * 4.74047
+ Export the results from the three queries to Topcat, concatenate, create 7 subsets on transverse velocity (0-10, 10-20, 20-50, 50-100, 100-150, 150-200, 200+ km/s) and create scatter plot showing each subset differently.
+ Topcat output might be slightly different than above (depends on window size and other options)

Solution:

a1) Hipparcos stars

ADQL query (30009 rows):

```
select gaia.source_id,
	gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs,
	hip.b_v as b_min_v,
	sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2))/gaia.parallax*4.74047 as vperp
from gaiadr1.tgas_source as gaia
inner join public.hipparcos_newreduction as hip
	on gaia.hip = hip.hip
where gaia.parallax/gaia.parallax_error >= 5 and
	hip.e_b_v > 0.0 and hip.e_b_v <= 0.05 and
	2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
	 (gaia.parallax >= 10.0 or
	sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
	gaia.phot_g_mean_mag <= 7.5)
```

a2) Tycho stars

ADQL query (8983 rows)

```
select gaia.source_id,
	gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs,
	0.85*(tycho2.bt_mag-tycho2.vt_mag) as b_min_v,
	sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2))/gaia.parallax*4.74047 as vperp
from gaiadr1.tgas_source as gaia
inner join public.tycho2 as tycho2
	on gaia.tycho2_id = tycho2.id
where gaia.parallax/gaia.parallax_error >= 5 and
	sqrt(power(tycho2.e_bt_mag,2) + power(tycho2.e_vt_mag,2)) <= 0.05 and
	2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
	 (gaia.parallax >= 10.0 or
	sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
	gaia.phot_g_mean_mag <= 7.5)
```

a3) APASS stars with bad Tycho colour

ADQL query (2144 rows, 2923 rows if Tycho stars are not removed)

```
select gaia.source_id,
	gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10 as g_mag_abs,
	 (urat.b_mag-urat.v_mag) as b_min_v,
	sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2))/gaia.parallax*4.74047 as vperp
from gaiadr1.tgas_source as gaia
inner join public.tycho2 as tycho2
	on gaia.tycho2_id = tycho2.id
inner join gaiadr1.urat1_best_neighbour as uratxmatch
	on gaia.source_id = uratxmatch.source_id
inner join gaiadr1.urat1_original_valid as urat
	on uratxmatch.urat1_oid = urat.urat1_oid
where gaia.parallax/gaia.parallax_error >= 5 and
	sqrt(power(tycho2.e_bt_mag,2) + power(tycho2.e_vt_mag,2)) > 0.05 and
	sqrt(power(urat.b_mag_error,2) + power(urat.v_mag_error,2)) <= 0.05 and
	2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux <= 0.05 and
	 (gaia.parallax >= 10.0 or
```

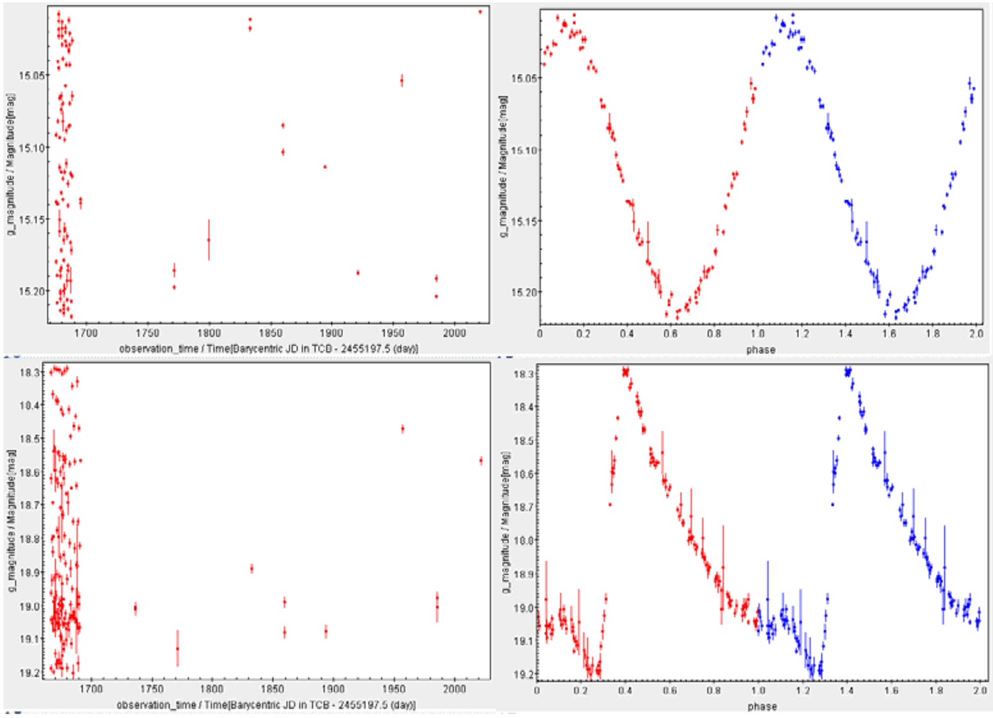
```
sqrt(power(gaia.pmra,2)+power(gaia.pmdec,2)) >= 200 or
gaia.phot_g_mean_mag <= 7.5)
```

b) Export the results from the three queries to Topcat, concatenate, create 7 subsets on transverse velocity (0-10, 10-20, 20-50, 50-100, 100-150, 150-200, 200+ km/s) and create scatter plot showing each subset differently.

Exercise 7: Cepheid and RR Lyrae phase-folded light curves

Adapted from figures and queries in Gaia DR1 paper, Brown et al. (2016)

Objective: Construct the light curves, time and phase folded, for a Cepheid and RR Lyrae in Gaia DR1



Conditions:

- Cepheid source_id: 4660526118774786304
- RR Lyrae source_id: 5284240582308398080

Hints:

+ Use G magnitudes
+ Fold the phase using the period (inverse of frequency)
+ The light curve is in table gaiadr1.phot_variable_time_series_gfov
+ The variable star period is in table gaiadr1.variable_summary
+ Join both tables using the source_id column
+ Phase = mod(time, period) / period = mod(time, 1 / frequency) * frequency
+ error_mag ≈ 2.5 / log(10) * flux_error / flux
+ Export data to Topcat and create scatter plots with error bars. For the phase folded diagrams, represent G vs phase and G vs phase + 1 in the same plot
+ Topcat output might be slightly different than above (depends on window size and other options)

Solution:

a1) Cepheid light curve
Cepheid source_id: 4660526118774786304
ADQL query (100 rows)

```
select
	curves.observation_time,
	mod(curves.observation_time, 1 / vars.phot_variable_fundam_freq1)
		* vars.phot_variable_fundam_freq1 as phase,
	curves.g_magnitude,
	2.5/log(10)* curves.g_flux_error/ curves.g_flux as g_magnitude_error
from gaiadr1.phot_variable_time_series_gfov as curves
inner join gaiadr1.variable_summary as vars
	on vars.source_id = curves.source_id
where vars.source_id = 4660526118774786304
```

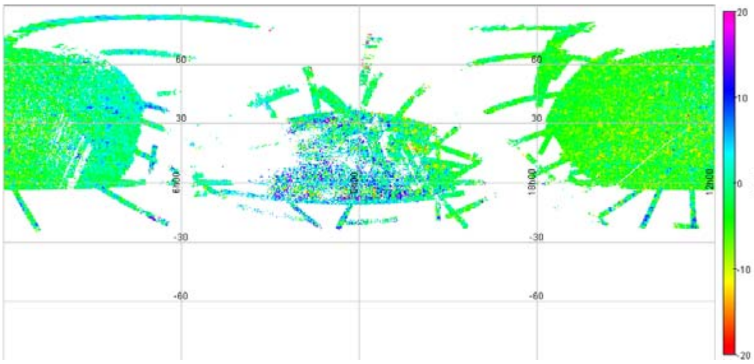
a2) RR Lyrae light curve
Source_id = 5284240582308398080
ADQL query (144 rows)

```
select
	curves.observation_time,
	mod(curves.observation_time, 1 / vars.phot_variable_fundam_freq1)
		* vars.phot_variable_fundam_freq1 as phase,
	curves.g_magnitude,
	2.5/log(10)* curves.g_flux_error/ curves.g_flux as g_magnitude_error
from gaiadr1.phot_variable_time_series_gfov as curves
inner join gaiadr1.variable_summary as vars
	on vars.source_id = curves.source_id
where vars.source_id = 5284240582308398080
```

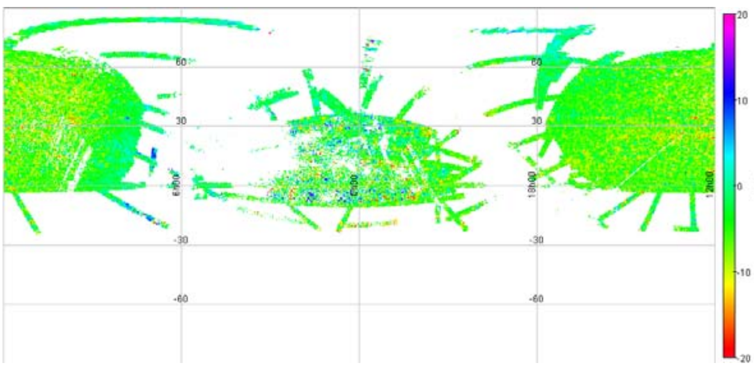
b) Export the results to Topcat via SAMP, create a scatter plot, flip y-axis and represent G vs time and phase. For the phase folded diagrams, overplot G vs phase + 1 to display two consecutive periods.

Exercise 8: SDSS proper motions (inspired by Sergey Koposov, NYC Gaia Sprint)

Objective: estimate proper motions for a subset of SDSS stars and compute average in Healpix map



PM RA



PM DEC

Conditions:
- 18 < SDSS r < 19
- 0.2 < SDSS g-r < 0.4
- Use Healpix 7 resolution
- Compute average proper motions for Healpix tiles with 5 or more stars.

Hints:
+ Gaia positions are in table gaiadr1.gaia_source
+ SDSS data are in table gaiadr1. sdssdr9_original_valid
+ Join using Gaia-SDSS cross-match best neighbours are in table gaiadr1. sdssdr9_best_neighbour
+ Gaia epoch MJD: 57023
+ proper motion RA = Delta RA * cos DEC / Delta time
+ Healpix level 7 = source_id / 35184372088832 (integral division, see source_id data model description)
+ Export data to Topcat via SAMP. Create sky plot and colour according to proper motion within +- 20 mas/yr
+ Topcat output might be slightly different than above (depends on window size and other options)

Solution:

a) ADQL query (945520rows): Gaia - SDSS cross-match for proper motion computation. Big table (332 MB), Takes around 12 min to complete query.

```
select * from gaiadr1.gaia_source
inner join gaiadr1.sdss_dr9_best_neighbour
        on gaiadr1.gaia_source.source_id = gaiadr1.sdss_dr9_best_neighbour.source_id
inner join gaiadr1.sdssdr9_original_valid as sdss
        on gaiadr1.sdss_dr9_best_neighbour.sdssdr9_oid = sdss.sdssdr9_oid
where 18 < sdss.r_mag and sdss.r_mag < 19
        and 0.2 < sdss.g_mag - sdss.r_mag
        and sdss.g_mag - sdss.r_mag < 0.4
```

b) Save table in user space, e.g. user_amora.sdss_proper_motions

c) ADQL query (57543 rows): Proper motion computation and Healpix histogram creation

```
select
        count(*) as n, avg(pmra) as pmra, avg(pmdec) as pmdec, avg(ra) as ra, avg(dec) as dec,
        avg(l) as l, avg(b) as b, healpix7 from (
        select
                ra, dec, l, b,
                (57023 - mjd) / 365.25 as years,
                (ra  - ra_2)  / ((57023 - mjd) / 365.25) * 3600000 * cos(radians(dec)) as pmra,
                (dec - dec_2) / ((57023 - mjd) / 365.25) * 3600000 as pmdec,
               source_id / 35184372088832 as healpix7
        from user_amora.sdss_proper_motions) as subquery
group by healpix7
having count(*) >= 5
```

d) Export data to Topcat via SAMP. Create sky plot. Use car projection. Add auxiliary colour axis according to proper motion in RA and DEC. Fix scale in interval [-20, 20] mas/yr

Accessing the Gaia Archive with Python

Tutorial by Enrique del Pozo. Gaia SOC

1. Pre-Requisites

If you want to follow the Python hands-on session you need at least the following software installed in your computer for running the code examples:

- Python 3.5.+
- Jupyter Notebook
- Numpy and Matplotlib python modules

There is a great manager for Python modules named Anaconda that contains all the needed pre-requisites to follow this hands-on session. See: https://www.continuum.io/downloads

2. Lets PyDance!

a) Getting the hands-on material

+ Go to ftp://ftp.sciops.esa.int/pub/epozo/GACS-Python/GACS-Workshop.ipynb and download the Jupyter notebook for this session.
+ If you do not want to use the Jupyter Notebook, the python code for replicating the samples can be found in ftp://ftp.sciops.esa.int/pub/epozo/GACS-Python/GACS-Workshop.py
+ For those that cannot install python , you can follow the tutorial in: ftp://ftp.sciops.esa.int/pub/epozo/GACS-Python/GACS-Workshop.html

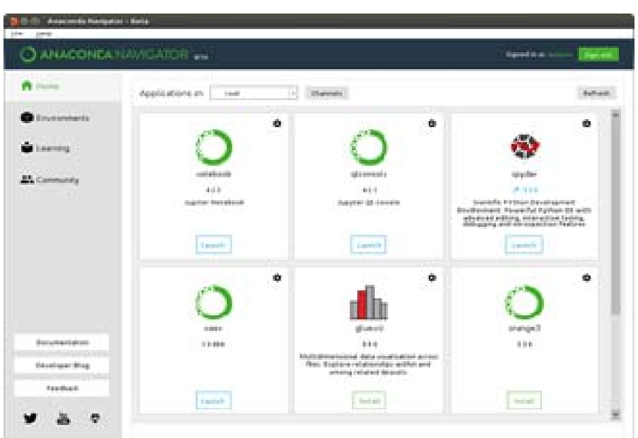
b) Launch the Jupyter Notebook

- From command line (OSX and Linux users):

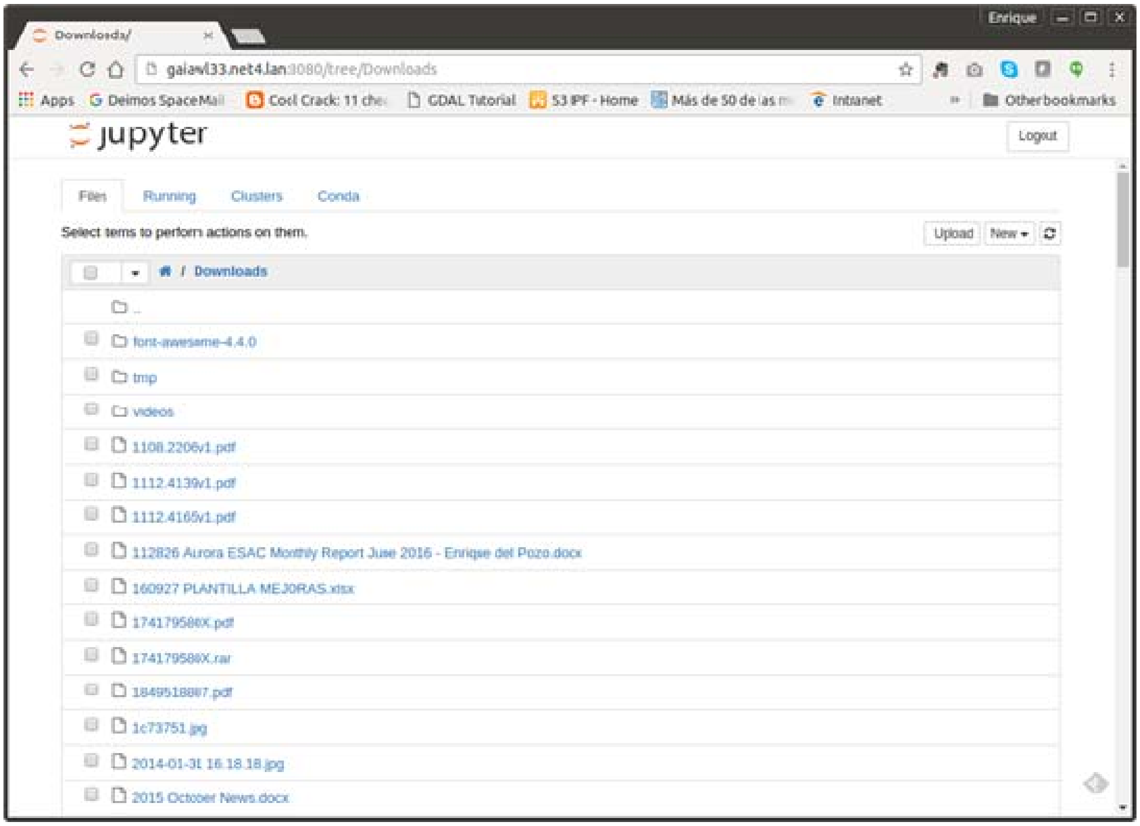
    Open a terminal and run $> jupyter-notebook

- Using Anaconda Navigator:

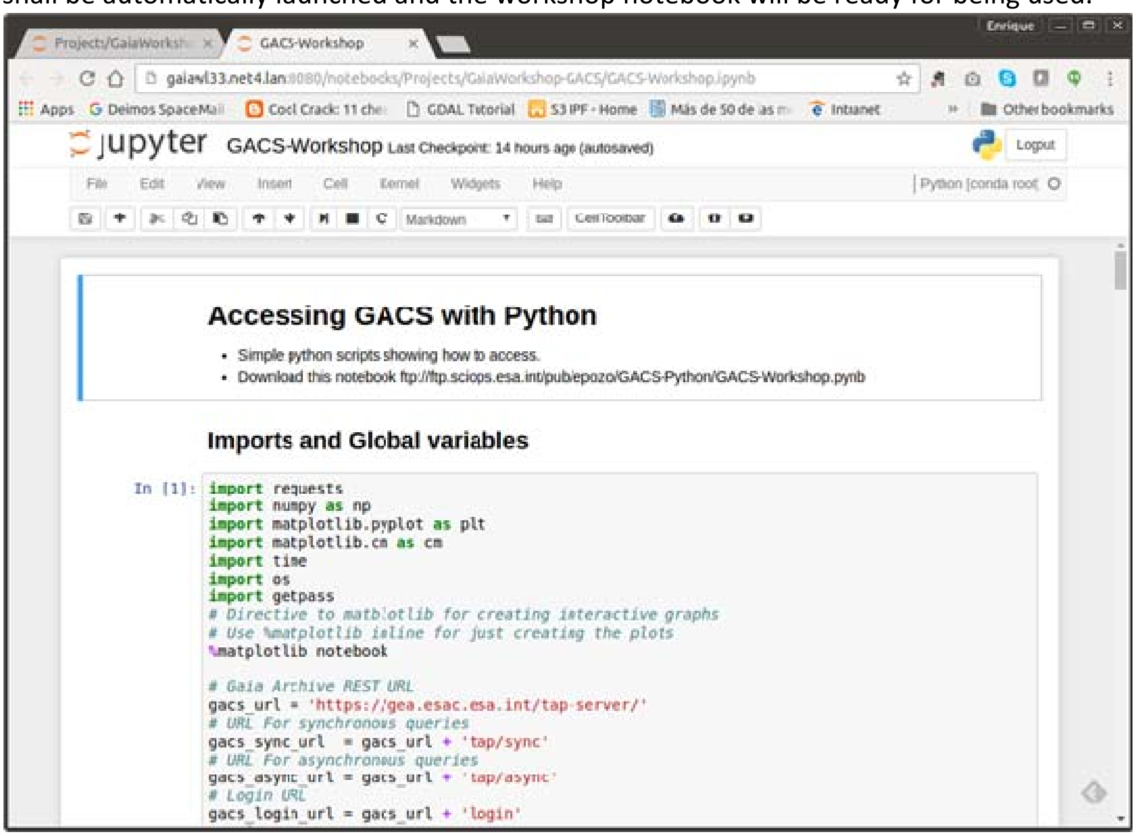
    Launch the Anaconda Navigator and click on “Launch” button within the “Notebook” widget.



In the moment the Jupyter Notebook is launched a new tab will be opened in your default browser.



In the File tree search for the downloaded GACS-Workshop.pyn file … and click on it, the python Kernel shall be automatically launched and the workshop notebook will be ready for being used.



Jupyter Notebook usage tips

- Each cell can be executed using “Ctrl + Enter” or “Command + Enter”
- Alternatively “Shift + Enter” can be used to executed a cell and automatically pass to the next cell in the notebook.
- Cells can be mark as “Code” or “Markdown” in case you want to enter text information.
- The charts that will be created are interactive , remember to click on the shutdown icon of the chart if you want to create a different one.