# Tutorial: Exploring Gaia data with TOPCAT and STILTS, part II

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

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

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

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

## Contents

**1 Local Herzsprung-Russell Diagram** **2**
- 1.1 Acquire data from TAP service . . . 2
- 1.2 Plot HRD . . . 2
- 1.3 Exclude astrometrically suspect sources . . . 3
- 1.4 Exclude photometrically suspect sources . . . 3
- 1.5 Explore the HRD . . . 3

**2 Local Herzsprung-Russell Diagram using STILTS** **4**
- 2.1 Get STILTS running . . . 4
- 2.2 Acquire data from TAP service . . . 4
- 2.3 Manipulate the downloaded table using `tpipe` . . . 4
- 2.4 Plot the diagram . . . 5
- 2.5 Use TOPCAT to refine the plot command . . . 5

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

# Advanced usage of Topcat

# 1 Local Herzsprung-Russell Diagram

In this example we will use a TAP query to download all the nearby Gaia sources with good astrometry and photometry, and calculate their absolute magnitudes to construct an HR diagram, performing a couple of cleaning operations to improve the data. This procedure loosely follows Appendix C of the Gaia DR2 astrometry paper Lindegren et al. 2018 [3].

## 1.1 Acquire data from TAP service

1. Open the TAP window **VO|Table Access Protocol (TAP) Query**

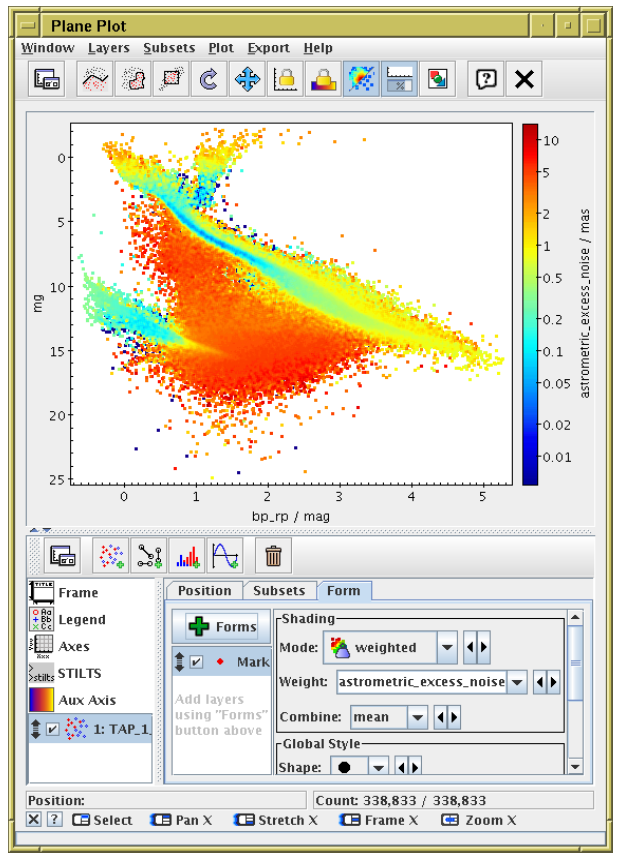
2. Select one of the Gaia services (probably the ESA one) and **Use Service**

3. Choose **Mode**: "`Asynchronous`" (just above the ADQL text entry panel). This query may take a minute or two, so a synchronous query might time out (with the unhelpful result, probably, "*TAP response is not a VOTable*").

4. Execute the following query:

```
SELECT ra, dec, parallax, phot_g_mean_mag, bp_rp,
       astrometric_excess_noise,
       phot_bp_rp_excess_factor
FROM gaiadr2.gaia_source
WHERE parallax > 10
  AND parallax_over_error > 10
  AND phot_bp_mean_flux_over_error > 10
  AND phot_rp_mean_flux_over_error > 10
```

   You should get a table with 338 833 sources; they are nominally within 100 pc, and have $\varpi$, BP and RP with small errors. In particular, the parallax error is small enough that $\varpi^{-1}$ is a reasonable estimate of distance.




## 1.2 Plot HRD

1. Add a new column calculating absolute G magnitude, using parallax: **Columns|New Synthetic Column** in **Column Info** window:
   **Name**: "g_abs",
   **Expression**: "`phot_g_mean_mag + 5*log10(parallax/100)`",
   **Units**: "`mag`"

2. Make a **Graphics|Plane Plot**, with **X**: "`bp_rp`" ($BP - RP$ colour), **Y**: "`g_abs`" (absolute G magnitude). Use the **Axes** control, **Coords** tab, **Y Flip** checkbox to flip it the right way round. Structure visible, but lots of interlopers.

3. Play around with different **Shading Modes** in **Form** tab.

4. Colour points using the other columns using modes **Aux**, **Weighted**. What's the difference between the two?

### 1.3 Exclude astrometrically suspect sources

1. Try different weighting/aux columns to see which one can be used to exclude points in the unwanted region between the main sequence and white dwarf branch.

2. Experiment with the colour map settings ( **Aux Axis** control) to find a suitable threshold.

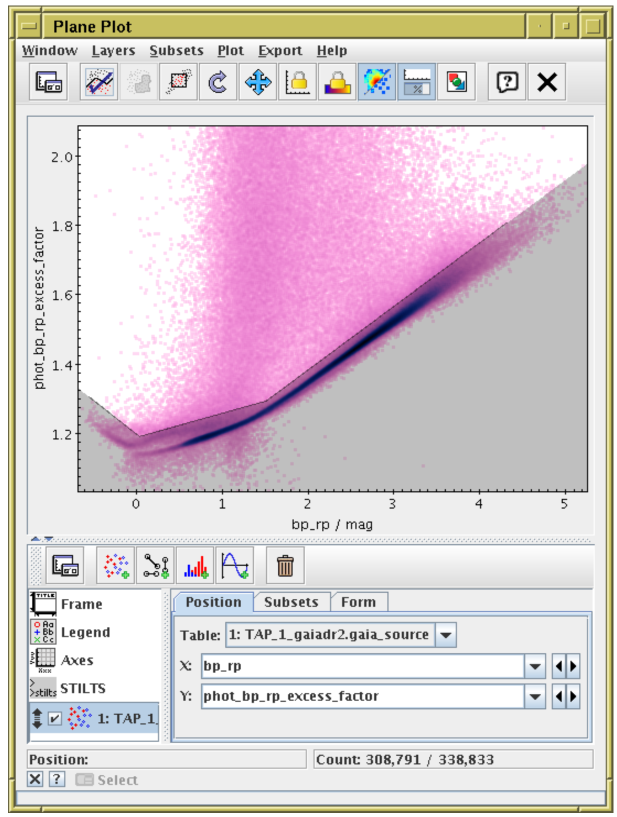
3. Create a subset using an algebraic expression that excludes the spurious points: go to the **Views|Row Subsets** window and use the **Subsets|New Subset** action: **Subset Name**: "`astrom_ok`", **Expression**: "`astrometric_excess_noise < 1`"

4. Go back to the plot, and make sure only the `astrom_ok` subset is plotted (**Subsets** tab). Now there are fewer spurious sources.

The `astrometric_excess_noise` column characterises the goodness of fit of the astrometric solution to the observations. But you didn't need to know that to improve the data like this.

### 1.4 Exclude photometrically suspect sources

1. Open a new **Graphics|Plane Plot**, with **X**: "`bp_rp`" ($BP - RP$ colour), **Y**: "`phot_bp_rp_excess_factor`".




The quantity on the Y axis is some measure of photometric reliability. High values are bad, but how high is colour-dependent. We will define a region in this space to exclude the unusually high values. This time we will interactively draw a polygon rather than a blob.

2. Click the **Subsets|Draw Subset Polygon** action

3. Select **Point inclusion mode**: "`BELOW`" in the popup

4. Click on a few points above the overdense region until the shaded area roughly covers it.

5. When you're done, click on the button again.

6. Fill in the **Subset Name** field in the popup (e.g. `photom_ok`) and hit **OK**

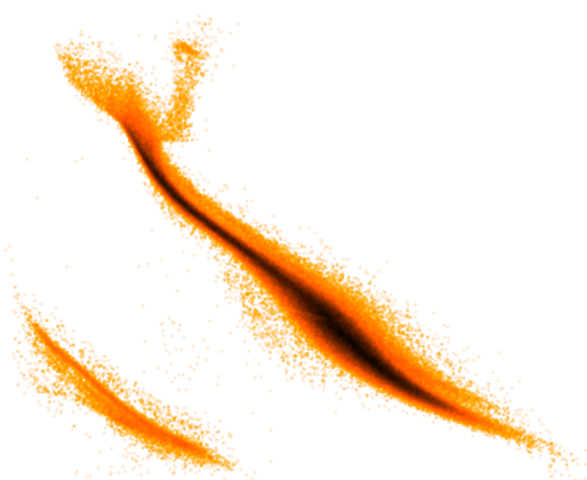
7. Go to the **Views|Row Subsets** window, where you can see the new subset alongside `astrom_ok`.

8. Create a new subset combining the two: **Subsets|New Subset** action: **Subset Name**: "`ok`" **Expression**: "`astrom_ok && photom_ok`"

9. Go back to the plot, and make sure only the `ok` subset is plotted (**Subsets** tab). Now there are fewer spurious sources.

Leave this session open, we'll need some of it for the next section.

### 1.5 Explore the HRD



The photometry and astrometry in Gaia DR2 is so good that plotting a Herzsprung-Russell Diagram by following the steps above gives a lot of astrophysical information. Identify the different populations with the help of the zoom and shading options in TOPCAT: main sequence, giant branches, the double-stranded white dwarfs sequence representing the split between helium and hydrogen burning, and if you look closely a notch in the main sequence near absolute G magnitude of 10 (Jao et al 2018 [5]).

# 2 Local Herzsprung-Russell Diagram using STILTS

In this section we will reproduce the cleaned HRD from the previous section, but this time from the command line, though we will use some of the results of the earlier GUI activity. Using TOPCAT interactively to work out how to write STILTS scripts for later batch use like this is a common pattern of work.

## 2.1 Get STILTS running

1. Unlike TOPCAT, you can't bluff your way through STILTS by pointing'n'clicking; you'll need some documentation. Find the manual here:

   http://www.starlink.ac.uk/stilts/sun256/

   or just google for "stilts" and go to the **Documentation** section.

2. Make sure you have it installed. If you have topcat, you can run "`java -jar topcat-full.jar -stilts ...`". More convenient (on Un*x), download the `stilts` script into the same directory as `topcat-full.jar` (or `stilts.jar`). We will write just "`stilts`" from now on.

3. Run a command:
   ```
   stilts calc expression="1+2"
   ```
   should print out "3"

4. Find the documentation for the `calc` command in the manual (Appendix B). Look at the **Usage** and **Examples** subsections.

5. Get command-line help by running "`stilts calc help`" and "`stilts calc help=expression`".

## 2.2 Acquire data from TAP service

1. We will use the `tapquery` command; find the entry in the manual.

2. Go back to the TOPCAT TAP window to find the TAP URL at the bottom of the **Select Service** tab (value of `tapurl` parameter)

3. Use the ADQL from step 4 of section 1.1 (value of `adql` parameter).

4. Run the query and output the result to a local FITS file `hrd.fits`:

```
stilts tapquery \
  sync=false \
  tapurl=http://gea.esac.esa.int/tap-server/tap \
  adql="SELECT ra, dec, parallax, phot_g_mean_mag, \
               phot_g_mean_mag, bp_rp \
               astrometric_excess_noise, \
               phot_bp_rp_excess_factor \
        FROM gaiadr2.gaia_source \
        WHERE parallax > 10 \
          AND parallax_over_error > 10 \
          AND phot_bp_mean_flux_over_error > 10 \
          AND phot_rp_mean_flux_over_error > 10" \
  out=hrd.fits
```

## 2.3 Manipulate the downloaded table using `tpipe`

We will perform some processing on the downloaded table using the `tpipe` command. This reads an input table, optionally performs operations on it, and writes it to output, either to a file (in the same or different format) or some other destination. It works like a Unix pipeline.

The operations are defined by adding *filters* such as "`select`" (retain only some rows) or "`addcol`" (add a new column). Filters are specified by adding using zero or more "`cmd=`" parameters on the command line, and documented in Section 6.1: "Processing Filters" of the STILTS manual.

The default output mode is write to a file, but other options such as "`meta`" (display column metadata), "`count`" (count rows) and "`stats`" (calculate mean, st.dev etc) are also available. Select non-default output modes using the "`omode=`" parameter on the command line, documented in Section 6.4: "Output Modes" of the manual.

1. Count the rows in the query result table:

```
stilts tpipe in=hrd.fits omode=count
```

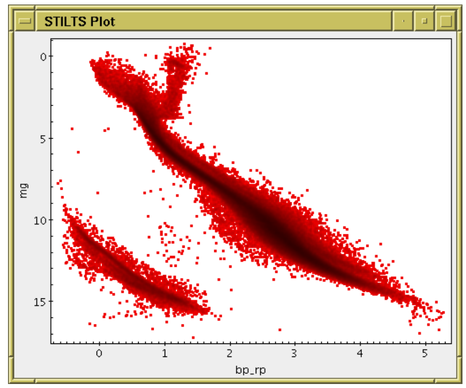
2. See what columns the query result table has:

```
stilts tpipe in=hrd.fits omode=meta
```

3. Prepare an expression for a new column giving absolute magnitude (see item 1 from section 1.2)
4. Prepare expressions for rows corresponding to the astrometry and photometry selections from sections 1.3 and 1.4; go back to TOPCAT's **Subsets** window, and see the expressions used to restrict the sources in the final HRD.
5. Put items 3 and 4 together to produce a plot-ready table (note use of quotes):

```
stilts tpipe in=hrd.fits \
             cmd=’addcol g_abs "phot_g_mean_mag + 5*log10(parallax/100)"’ \
             cmd=’select "astrometric_excess_noise < 1"’ \
             cmd=’select "phot_bp_rp_excess_factor < polyLine(bp_rp, -0.56,1.307, 0.03,1.192, \
                                                          1.51,1.295, 4.31,1.808)"’ \
             out=hrd_clean.fits
```

6. Count the rows in the output file.

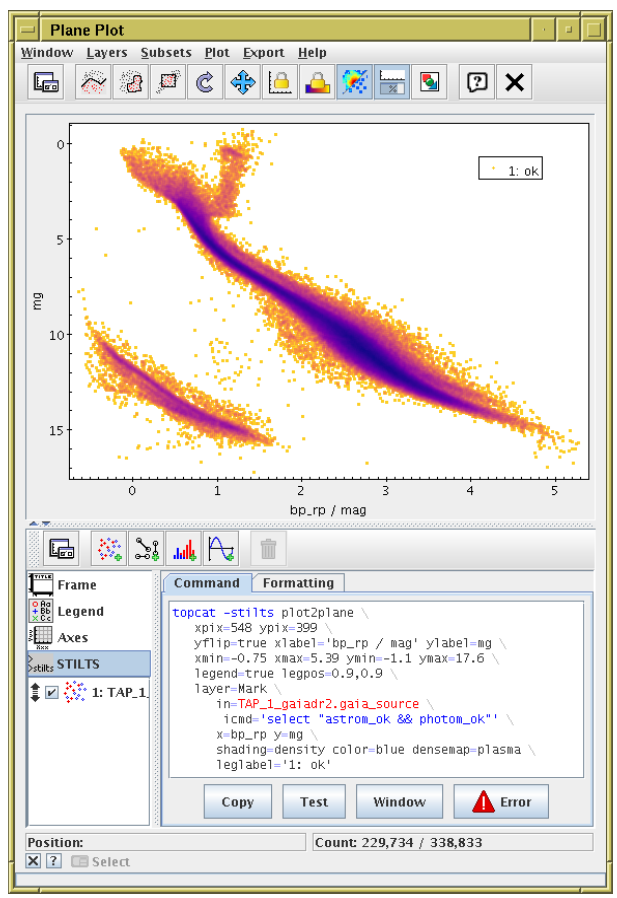



## 2.4 Plot the diagram

1. Plot the HRD on the screen using the 2d plotting command `plot2plane`:

```
stilts plot2plane layer=mark in=hrd_clean.fits \
                  x=bp_rp y=g_abs yflip=true
```

2. By default, a "live" plot shows up on the screen: use the mouse to zoom and drag it around like in TOPCAT.
3. Output it to a graphics file instead, by adding the "`out`" parameter, e.g. "`out=hrd.png`" or "`out=hrd.pdf`".




## 2.5 Use TOPCAT to refine the plot command

There are *lots* more options to the plot commands.

1. In the cleaned HRD plot from the end of section 1, select the **STILTS** control at the lower left. This is the command that will reproduce the plot from the command line. It probably won't work from the command line exactly as written, because some tables and subsets only exist in TOPCAT (expected problems are shown in colour). But we have now generated the post-processed file using `tpipe`.
2. Hit the **Window** button near the bottom to break out the command display into a separate window. You can see *all* the options (including defaults) with the **Formatting|Include Defaults** menu item.
3. Now adjust plot options in the TOPCAT GUI and see how the command changes.
4. Get the plot looking like you want in TOPCAT, then copy and paste the details into your STILTS `plot2plane` command line.
5. Experiment with changing the appearance by modifying the script.

## Bonus

Go back and reproduce all the other exercises using STILTS!

## References

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

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

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

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

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

## Acknowledgements

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