Visualizing the HYG Star Database

By Paul Kassebaum

pkassebaum@gmail.com

Source of data: https://github.com/astronexus/HYG-Database/raw/master/hygdata_v3.csv

Fields in the database:

1. id: The database primary key.
2. hip: The star's ID in the Hipparcos catalog, if known.
3. hd: The star's ID in the Henry Draper catalog, if known.
4. hr: The star's ID in the Harvard Revised catalog, which is the same as its number in the Yale Bright Star Catalog.
5. gl: The star's ID in the third edition of the Gliese Catalog of Nearby Stars.
6. bf: The Bayer / Flamsteed designation, primarily from the Fifth Edition of the Yale Bright Star Catalog. This is a combination of the two designations. The Flamsteed number, if present, is given first; then a three-letter abbreviation for the Bayer Greek letter; the Bayer superscript number, if present; and finally, the three-letter constellation abbreviation. Thus Alpha Andromedae has the field value "21Alp And", and Kappa1 Sculptoris (no Flamsteed number) has "Kap1Scl".
7. ra, dec: The star's right ascension and declination, for epoch and equinox 2000.0.
8. proper: A common name for the star, such as "Barnard's Star" or "Sirius". I have taken these names primarily from the Hipparcos project's web site, which lists representative names for the 150 brightest stars and many of the 150 closest stars. I have added a few names to this list. Most of the additions are designations from catalogs mostly now forgotten (e.g., Lalande, Groombridge, and Gould ["G."]) except for certain nearby stars which are still best known by these designations.
9. dist: The star's distance in parsecs, the most common unit in astrometry. To convert parsecs to light years, multiply by 3.262. A value >= 10000000 indicates missing or dubious (e.g., negative) parallax data in Hipparcos.
10. pmra, pmdec: The star's proper motion in right ascension and declination, in milliarcseconds per year.
11. rv: The star's radial velocity in km/sec, where known.
12. mag: The star's apparent visual magnitude.
13. absmag: The star's absolute visual magnitude (its apparent magnitude from a distance of 10 parsecs).
14. spect: The star's spectral type, if known.
15. ci: The star's color index (blue magnitude - visual magnitude), where known.
16. x,y,z: The Cartesian coordinates of the star, in a system based on the equatorial coordinates as seen from Earth. +X is in the direction of the vernal equinox (at epoch 2000), +Z towards the north celestial pole, and +Y in the direction of R.A. 6 hours, declination 0 degrees.
17. vx,vy,vz: The Cartesian velocity components of the star, in the same coordinate system described immediately above. They are determined from the proper motion and the radial velocity (when known). The velocity unit is parsecs per year; these are small values (around 1 millionth of a parsec per year), but they enormously simplify calculations using parsecs as base units for celestial mapping.
18. rarad, decrad, pmrarad, prdecrad: The positions in radians, and proper motions in radians per year.
19. bayer: The Bayer designation as a distinct value
20. flam: The Flamsteed number as a distinct value
21. con: The standard constellation abbreviation
22. comp, comp_primary, base: Identifies a star in a multiple star system. comp = ID of companion star, comp_primary = ID of primary star for this component, and base = catalog ID or name for this multi-star system. Currently only used for Gliese stars.
23. lum: Star's luminosity as a multiple of Solar luminosity.
24. var: Star's standard variable star designation, when known.

25. var_min, var_max: Star's approximate magnitude range, for variables. This value is based on the Hp magnitudes for the range in the original Hipparcos catalog, adjusted to the V magnitude scale to match the "mag" field.

This visualization was inspired by this figure by u/zonination.

Set up a datastore.

```
filename = 'hygdata_v3.csv';
ds = datastore(filename);
```

Focus on the absolute visual magnitude and color index data fields.

```
ds.SelectedVariableNames = {'absmag','ci'};
```

Create a tall table of the data.

```
tbl = tall(ds);
```

Draw a two variable histogram.

```
h = histogram2(tbl.ci,tbl.absmag,'DisplayStyle','tile');
```

```
Evaluating tall expression using the Parallel Pool 'local':
- Pass 1 of 1: Completed in 10 sec
Evaluation completed in 14 sec
Evaluating tall expression using the Parallel Pool 'local':
- Pass 1 of 1: Completed in 4 sec
Evaluation completed in 5 sec
```
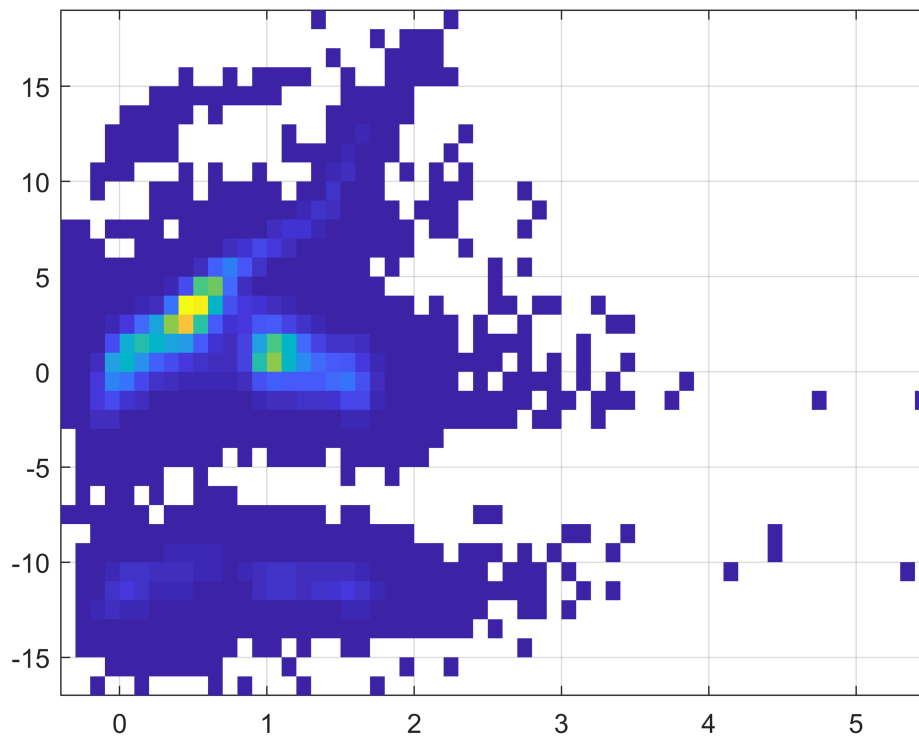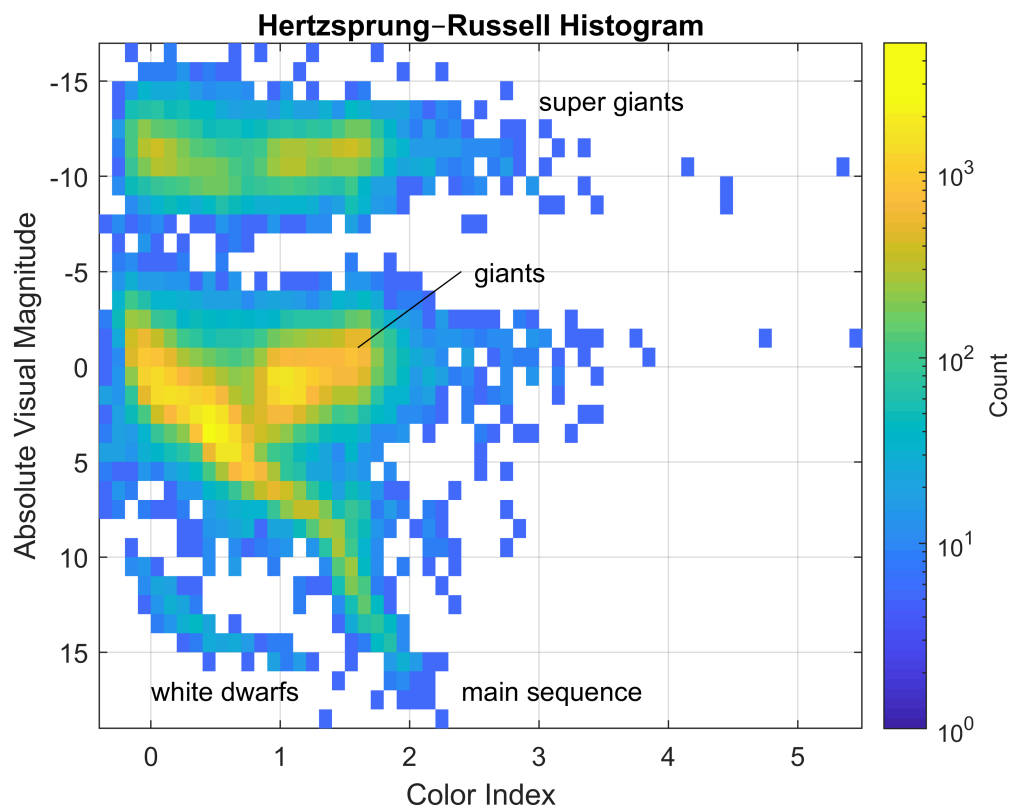
Customize the figure.

```
xlabel('Color Index');
ylabel('Absolute Visual Magnitude');
title('Hertzsprung-Russell Histogram');

ax = gca;
ax.ColorScale = 'log';
ax.YDir = 'reverse';

cb = colorbar;
cb.Label.String = 'Count';
cb.Limits = [1 5e3];

text(0, 17,'white dwarfs');
text(3,-14,'super giants');
text(2.4, 17,'main sequence');
text(2.5,-5,'giants');
line([2.4 1.6],[-5 -1],'Color','black');
```

Save the figure as a PNG.

```
print('ci_vs_absmag','-dpng');
```

Hertzsprung–Russell Histogram