KRITTIKA SUMMER PROJECTS 2022

Gaia Data Analysis

Amrit Shankar Mishra, Arihant Tiwari, Navdha, Sudeep Maurya, Tamojeet Roychowdhury, Teja Bale, Vighnesh Nayak, and Himanshu Verma

Amrit Shankar Mishra[3], Arihant Tiwari[4], Navdha[1,2], Sudeep Maurya[5], Tamojeet Roychowdhury[1,2], Teja Bale[2], Vighnesh Nayak[2], and Himanshu Verma[1,2]

[1]Krittika - The Astronomy Club of IIT Bombay, Powai, Mumbai - 400076, India

[2]Indian Institute of Technology Bombay, Mumbai - 400076, India

[3]Indian institute of Space Science and Technology, Thiruvananthapuram, Kerala, 695547, India

[4]Indian Institute of Science Education and Research Bhopal, Bhopal, 462066, India

[5]Indian Institute of Science Education and Research Mohali, Punjab, 140306, India

PUBLISHED BY KRITTIKA: THE ASTRONOMY CLUB OF IIT BOMBAY
GITHUB.COM/KRITTIKAIITB
First Release, September 2022

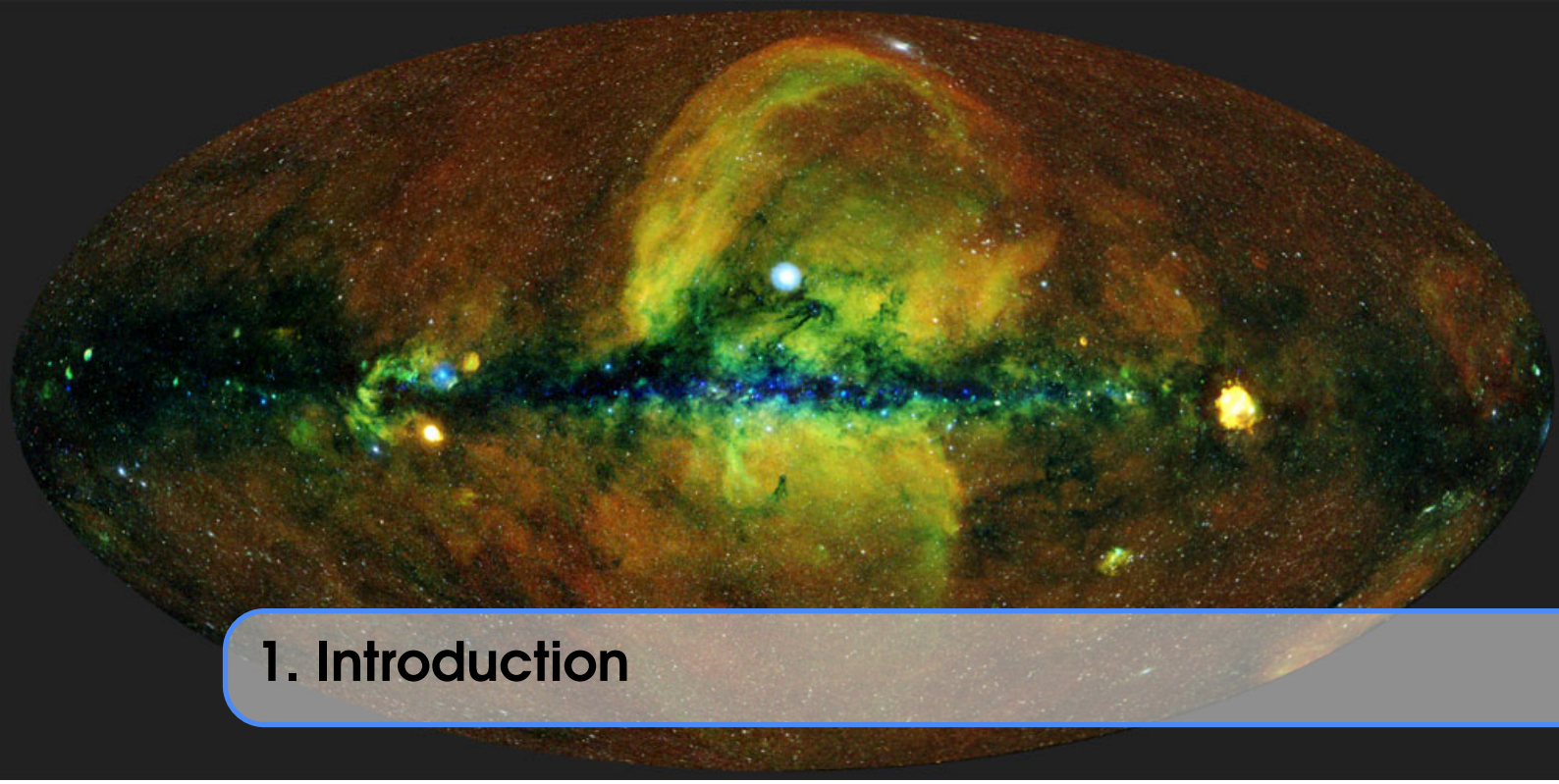
Abstract

Gaia is a space observatory of the European Space Agency (ESA) which was launched in 2013 and expected to operate until 2025. It aims to build the most accurate map of our galaxy so far, by mapping the positions, distances and proper motions of the stars with enormous precision. It currently operates around the Sun-Earth L2 Lagrangian point.

All of Gaia's data is accessible publicly via ADQL interface on Python, using the package `astroquery.Gaia`. This allows us to study the astrometric and certain photometric characteristics of over a billion stars upto magnitudes of 21 in the Gaia catalog using our computers. The Gaia Data Release 3 was released on 13 June 2022.

This project aims to use the data for specific types of stars, and then to use the trends observed to arrive at some interesting qualitative results for those chosen stars.

Contents



1. Introduction

1.1 Collecting Data

Gaia will perform its observations from a controlled Lissajous-type orbit around the L2 Lagrange point of the Sun and Earth-Moon system. During its 5-year operational lifetime, the satellite will continuously spin around its axis, with a constant speed of 60 arcsec/s. As a result, over a period of 6 hours, the two astrometric fields of view will scan across all objects located along the great circle 'perpendicular' to the spin axis. As a result of the basic angle of 106.5° separating the astrometric fields of view on the sky, objects transit the second field of view with a delay of 106.5 minutes compared to the first field.

Gaia has two telescopes with a fixed angular separation of 106.5°. Both of them rotate at the same angular velocity around an axis perpendicular to line of sight of the two telescopes. Light from both telescopes are focused onto a plane of CCDs using a mirror. The CCD array has 106 individual CCDs, and almost 1 billion pixels.

When light from the first telescope falls onto the CCD array, the first line of CCD is activated and selects the stars from the first telescope to be projected on the main CCD grid. Similarly for the second telescope, the second CCD line is activated.

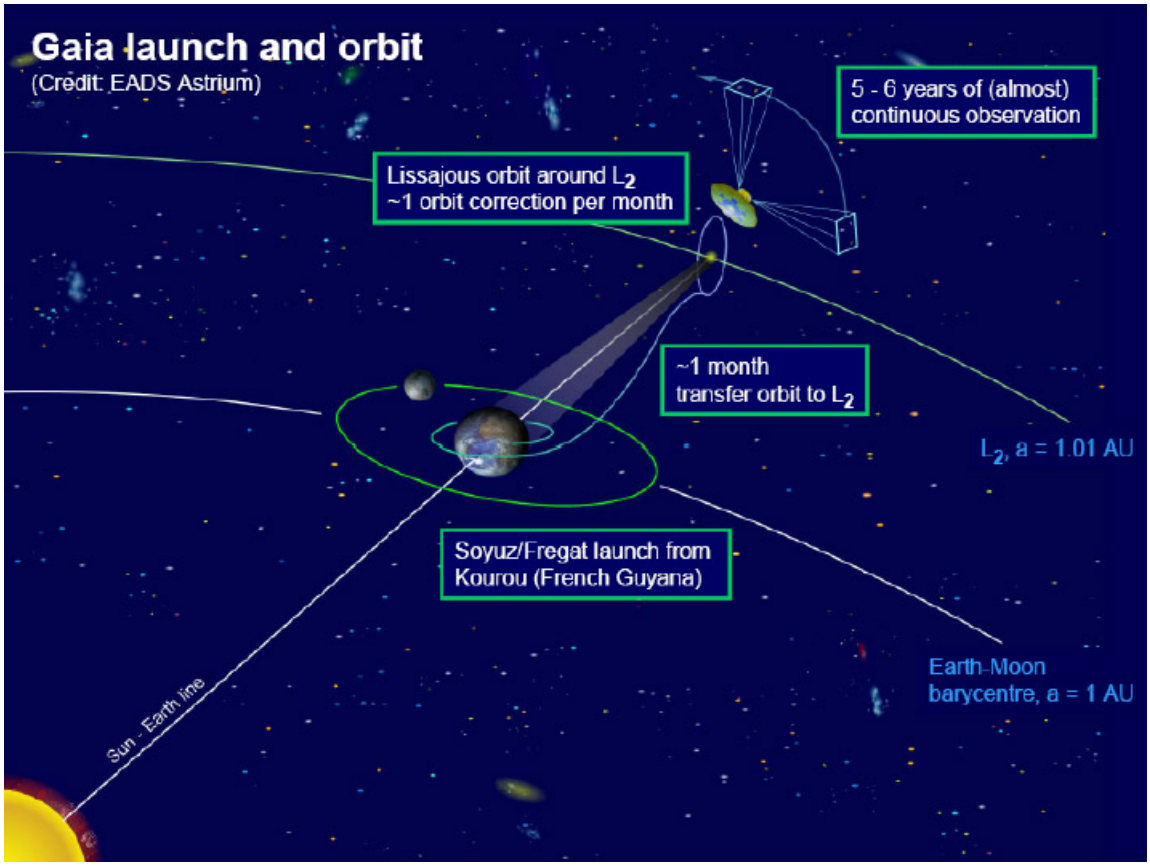


Figure 1.1: Orbit of Gaia around Sun-Earth L2

As the telescope rotates, light from each single star moves across the CCD grid. There are three main measurements made:

1. The first grid of pixels is the *astrometric field*. It tracks the position of the star. As the star moves along each CCD in the CCD grid, the on-board computer finds out its position using the CCD pixels and also finds out its brightness. As the starlight passes through the 9 columns of CCD, each column measures its position and brightness. Multiple measurements help to minimize the error, and also optimizes the amount of data needed to be sent.
2. The next two columns are meant for photometry. Light from the stars is passed through two special prisms, that directs the smaller and larger wavelength components of the star's light onto the blue and red photometers respectively. This provides information about the star's temperature (Wien's law), size and chemical composition (relative abundance of hydrogen, helium etc. using intensity of their spectral lines).
3. Light from the brighter stars is captured on the radial velocity spectrometers, which measure the stars' radial velocity using Doppler effect (using hydrogen spectral lines as the base wavelength).
4. Additionally, as Gaia moves along with Earth around the Sun, it also records each star's parallax, thus allowing us to find the distance to that star.

The collected information from the CCDs is compressed into a data packet to be stored on Gaia's onboard computer. and later transmitted back to ESA's ground stations. This process is repeated for each star as the telescopes rotate. Around 3 million stars are measured every hour. Gaia maps over a billion different stars,

upto a limiting magnitude of about 20, and would thus catalogue about 1% of our galaxy's stars. It is also expected to trace numerous asteroids and comets in our Solar System, exoplanets, brown dwarfs, supernovae and quasars - apart from numerous stars from other galaxies (Magellanic Clouds and farther in the Local Group). Each star is mapped almost 70 times over the course of 5 years before the processed data is released.

1.2 Coordinate Systems

1.2.1 Galactic coordinate system

The **galactic coordinate system** is a celestial coordinate system in spherical coordinates, with the **Sun** as its **center**, the primary direction aligned with the approximate center of the Milky Way Galaxy, and the fundamental plane parallel to an approximation of the galactic plane but offset to its north. It uses the right-handed convention, meaning that coordinates are positive toward the north and toward the east in the fundamental plane.

Galactic longitude:
Longitude (l) measures the angular distance of an object eastward along the galactic equator from the galactic center. Analogous to terrestrial longitude, galactic longitude is usually measured in degrees(°).

Galactic latitude:
Latitude (b) measures the angle of an object northward of the galactic equator (or midplane) as viewed from Earth. Analogous to terrestrial latitude, galactic latitude is usually measured in degrees(°).

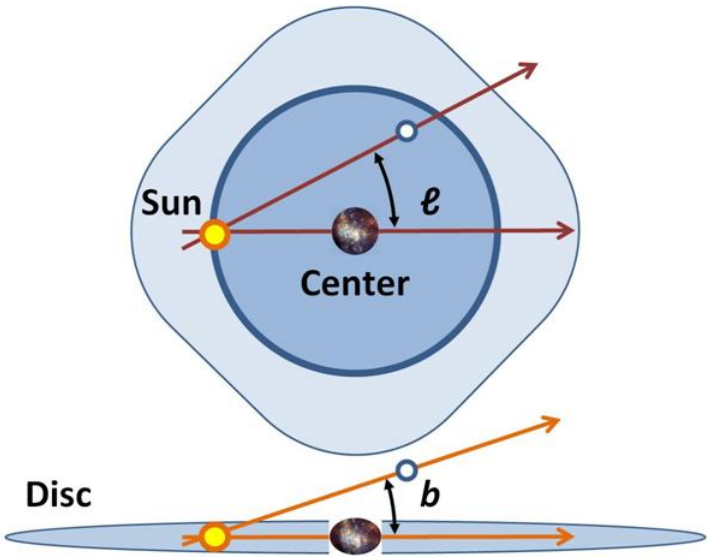


Figure 1.2: Galactic coordinate system

Here, in the from the above figure we can see that the galactic longitude is measured from the line joining the sun and galactic center and it is represented as l.

Similarly, latitude is measured above the galactic equator or midplane and it is represented by b.

1.2.2 ICRS coordinate system

- The **International Celestial Reference System (ICRS)** is the current standard celestial reference system adopted by the **International Astronomical Union (IAU).**
- Its origin is at the **barycenter** of the **Solar System**.
  Barycenter is the center of mass of two or more bodies that orbit one another and is the point about which the bodies orbit.
  Unlike galactic coordinate system (which has sun as origin), ICRS has barycenter of solar system as origin.

1.3 Observables

1.3.1 Flux

Luminosity: Luminosity is the energy released by a star in one second. It is generally denoted as L.
Flux is defined as:-

$$F = \frac{L}{4\pi r^2}$$

where r is distance of star from earth

1.3.2 Magnitude scale

Apparent magnitude

The stars' are classified on the basis of their apparent magnitude using the magnitude scale.
Apparent magnitude is defined as:-

$$m = -2.5\log_{10}\left(\frac{F}{F_0}\right)$$

here, F_0 is the reference flux of a source of our choice. Since negative of log is used in the definition, stars with higher magnitude value are fainter. The brightest object in the sky, the Sun, has an apparent magnitude of -26.74. The brightest star, Sirius, has apparent magnitude -1.46. The star Vega was historically assigned the baseline apparent magnitude of 0.00, but this was further refined later with other definitions when Vega was discovered to have slight variations in its luminosity.

Absolute magnitude

The apparent magnitude of a star is good enough to tell us about its brightness as seen from Earth, but it does not tell us anything about its intrinsic luminosity i.e. the total energy it radiates, due to the distance factor in the flux. So we use the absolute magnitude, which is the magnitude of a star as seen from a distance of 10 parsecs (1 parsec = 3.26 light years). By comparing absolute magnitudes of two different stars, we directly have a measure of their luminosities as the distance factor cancels out.

$$M = m + 5 - 5\log_{10} r$$

where r is the distance of star from earth in parsecs.

1.3.3 Using filters

Red pass filter

This filter allows higher wavelengths to pass through and blocks the lower wavelengths. The flux is available in the form of red pass apparent magnitude, denoted as G_{RP}.

Blue pass filter

This filter allows lower wavelengths to pass through and blocks the higher wavelengths. The flux is available in the form of blue pass apparent magnitude, denoted as G_{BP}.

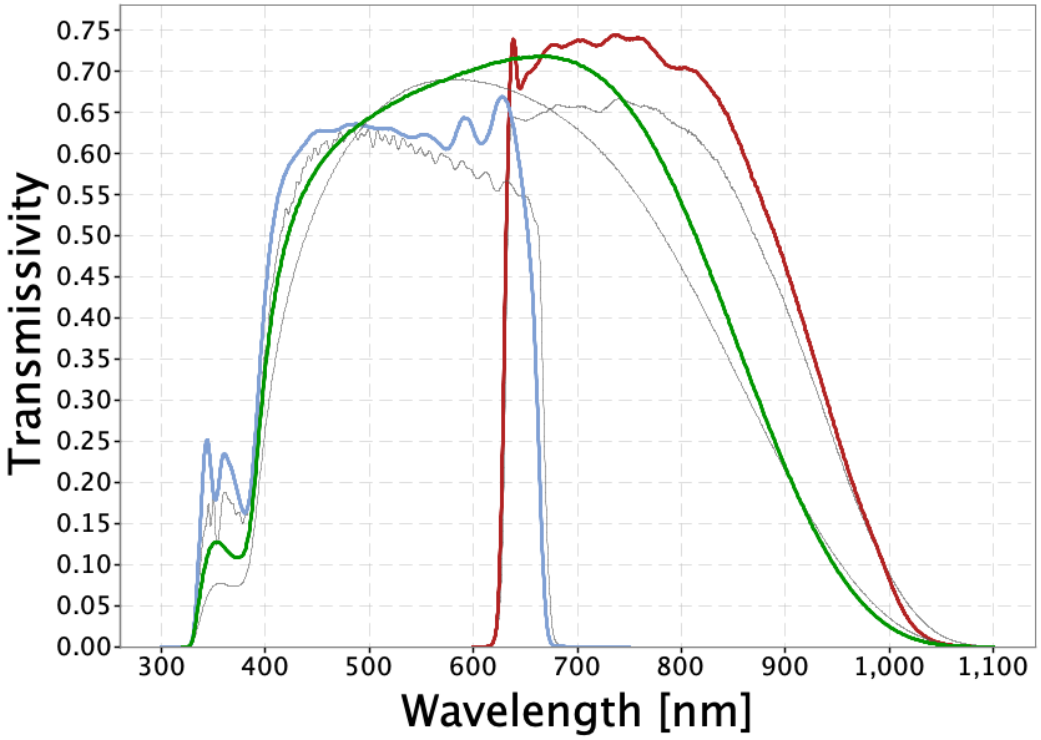


Figure 1.3: Range of Wavelengths admitted by each passband - red is for red pass filter, blue for blue pass and the green line denotes entire flux in visible range

2. Astrometry

2.0.1 Parallax

Parallax measures the angular movement of a star w.r.t. a fixed background of the sky. The angular movement is a result of the Earth (and Gaia) revolving around the sun. From different points in orbit, the line of sight to the star is slightly different resulting in parallax. It has units of angle. The unit of distance parsec is defined as the distance at which a star appears to have a parallax of 1 arcsecond (1/3600 of a degree) as seen from Earth.

Hence, parallax p of a star is directly related to its distance d as d in parsecs $= \frac{1}{p}$ where parallax is in arcseconds.

2.0.2 Proper Motion

A star also has its intrinsic relative velocity w.r.t. Earth-Sun system due to their motion around the galactic centre. This results in the star changing its position in the sky by a very small amount each year, known as proper motion. Note that the proper motion only measures the transverse component of the star's velocity, or more precisely, its angular velocity w.r.t. Earth. The transverse component of velocity is not included. Proper motion in Gaia catalogues is measured in milliarcseconds/year.

2.1 Brightest Stars.

2.1.1 Motivation and Selection Criteria

We considered the top 40,000 brightest stars in Gaia EDR3 catalog. The motivation for considering the brightest stars is that they tend to generally have higher mass, radius and temperature, which gives rise to some interesting physical phenomenon like supernovae. And also they tend to have more accurate data associated with them due to ease of observation.

The selection criteria were :

- Parallax is not null and G_{BP}, G_{RP} are not null.
- Parallax> 0 and Parallax error$< 2\%$.
- Ordered by apparent magnitude.

2.1.2 The ADQL query.

```
query="""select top 40000
    source_id,ref_epoch,ra,dec,parallax,pm,bp_rp,g_rp,
    phot_g_mean_mag
    from gaiaedr3.gaia_source
    where parallax is not null and bp_rp is not null
    and parallax>0 and parallax_over_error>50
    order by phot_g_mean_mag"""
```

2.1.3 Distribution of the stars in sky.

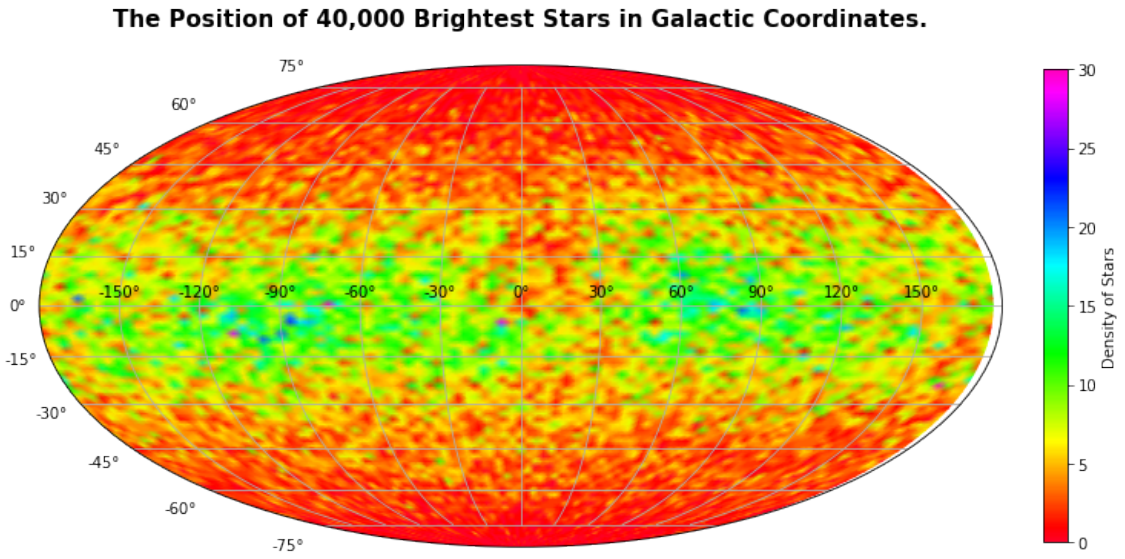
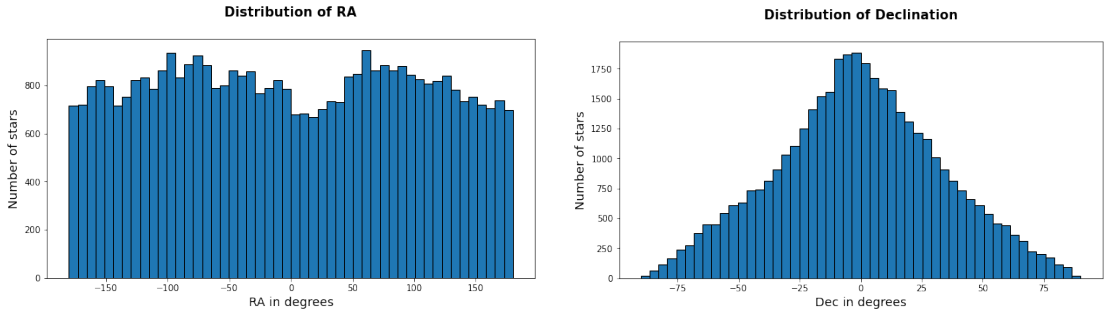


Figure 2.1: position of stars in galactic coordinates.

From the above figure we can see that most of the brightest stars are concentrated in central galactic plane of the Milky Way. This is due to their proximity to us. Since the ordering is done by apparent magnitude closer stars of milky way appear more in number in the data sample.

This fact can also be seen in the distribution of Right Ascension and Declination, wherein we can see that the stars are almost evenly distributed in RA they while are concentrated around 0° in Dec which corresponds to the galactic plane.



(a) RA distribution. (b) Dec Distribution.

Figure 2.2: Distribution of stars.

2.1.4 Distance and Proper motion distribution.

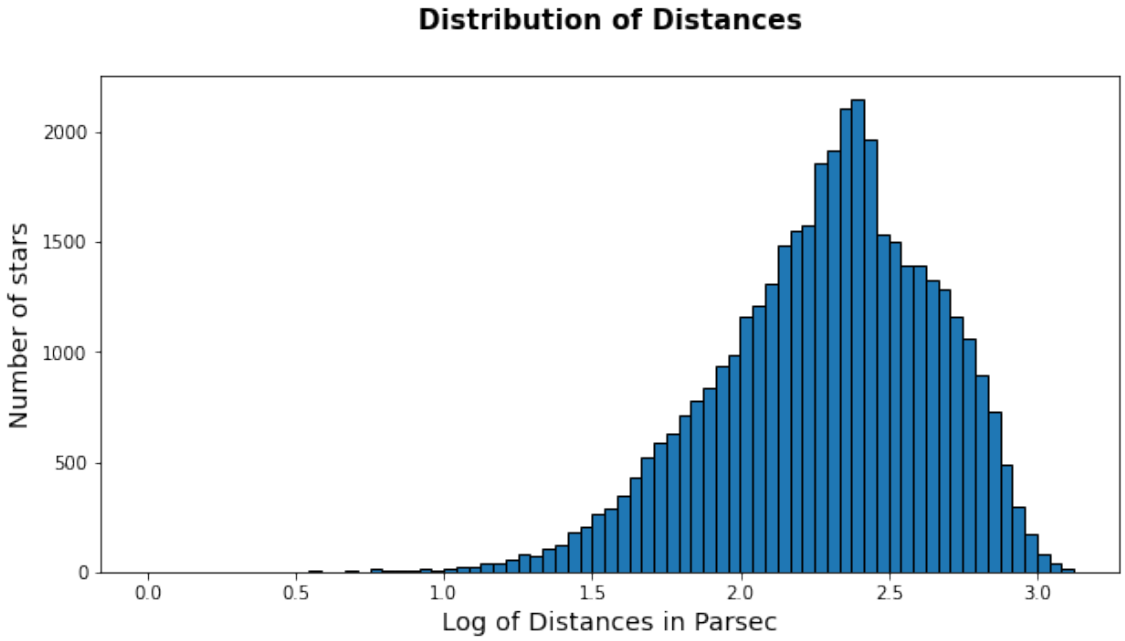


Figure 2.3: distance distribution of stars

Here we can see that at the beginning the number of stars in the sample increases with distance as the number of stars at a distance R is roughly proportional to R^2. And since brightness is inversly proportional to distance, after a maximum at 236.66 parsecs the number of stars drops of steeply.

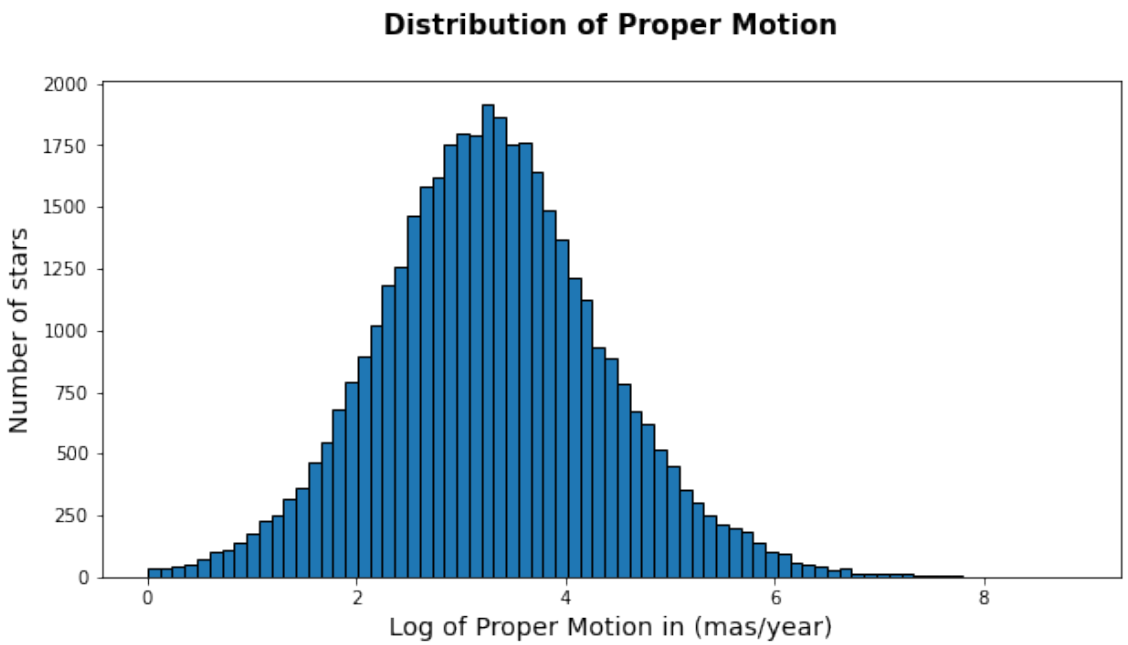


Figure 2.4: proper motion distribution of stars

The proper motion roughly follows Gaussian distribution with mean at 1.55 arcseconds/year.

2.1.5 Magnitude Distribution.

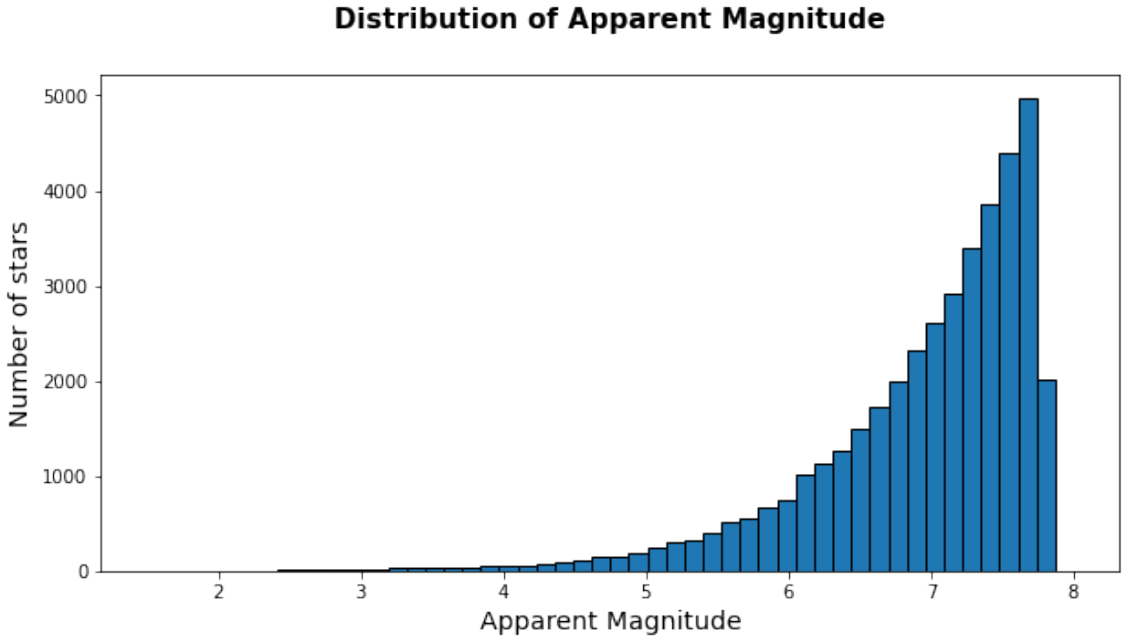
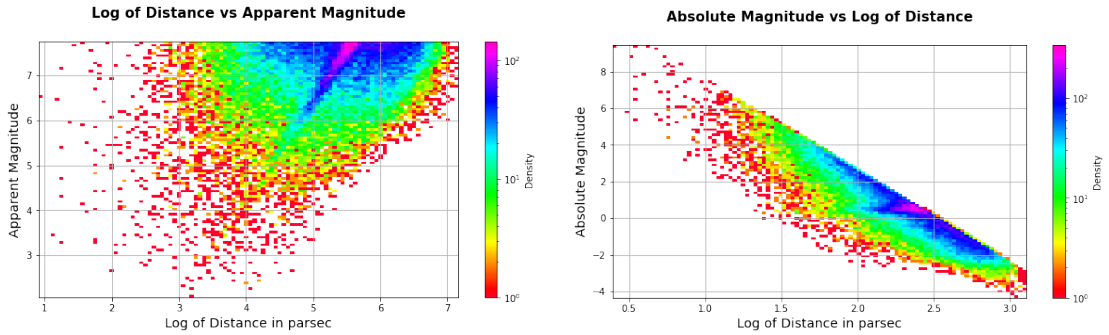


Figure 2.5: apparent magnitude distribution of stars

The apparent magnitude used here is the G-Band magnitude in GAIA database. Here we can see that the number of stars increase exponentially with apparent

magnitude as expected since stars with high brightness i.e. lower magnitude are less in number.

2.1.6 2D plots.



(a) Apparent magnitude vs Distance. (b) Absolute magnitude vs Distance.

Figure 2.6: Magnitude vs Distance.

Though apparent magnitude vs distance plot doesn't reveal much information, from absolute magnitude vs distance plot we can see that at higher distances stars with higher magnitude absolutely absent. And highest absolute magnitude at a log(distance) follows linearly decreasing trend.

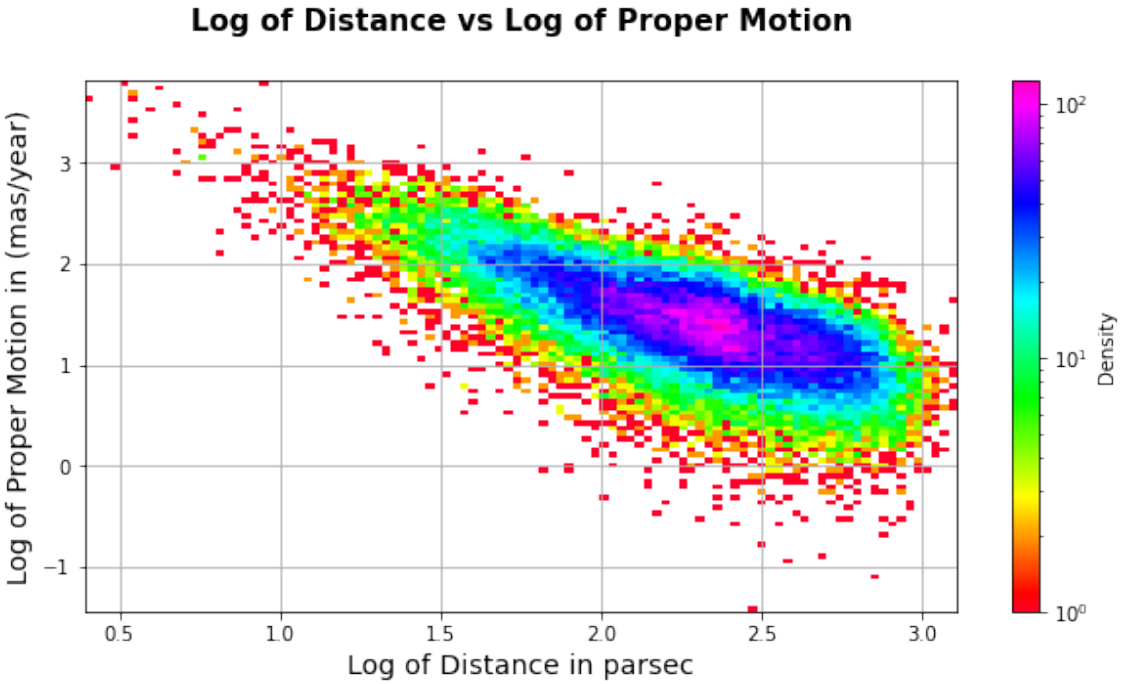
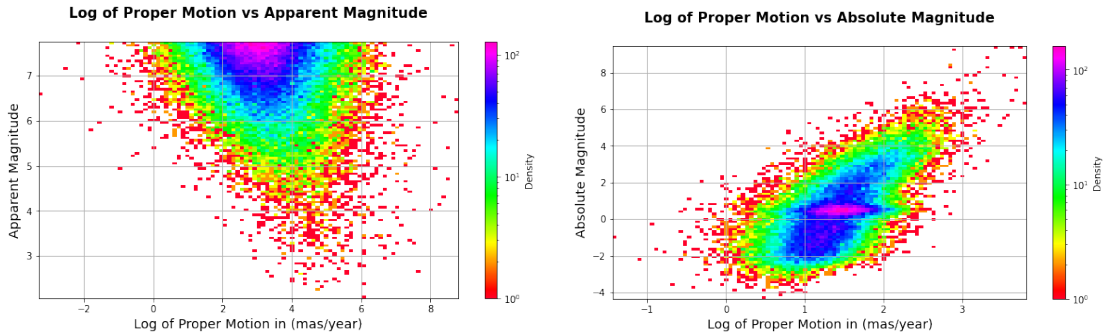


Figure 2.7: Proper Motion vs Distance



(a) Apparent magnitude vs Proper motion. (b) Absolute magnitude vs Proper motion.

Figure 2.8: Magnitude vs Proper motion.

2.2 Far away bright sources

2.2.1 ADQL Query used

The following is the ADQL query used to obtain the data:

```
query = """SELECT ra,dec, phot_g_mean_mag as mag,
phot_g_mean_mag -5*log10(r_med_geo) +5 as mg_abs,
r_med_geo as dist, pmra, pmdec,
phot_bp_mean_mag - phot_rp_mean_mag as bp_rp,  parallax
FROM gedr3dist.litewithdist
WHERE r_med_geo >300
and phot_g_mean_mag <8
and phot_g_mean_flux_over_error > 100
and abs(parallax_error/parallax) <0.01"""
```

NOTE: Since the GAVO table from which these data was taken did not have separate column for proper motion, it was found later by taking square-root of the sum of squares of pmra and pmdec.

2.2.2 Motivation and Selection Criteria

- The motivation behind this selection was that far away sources which have an apparent magnitude < 8 may include hotter main sequence stars as well as red giants.
- The gedr3dist.litewithdist table from GAVO was chosen as it had very good estimates of distance already included.
- A minimum distance of 300pc was chosen to get far away stars.
- A constraint for maximum error in apparent g-band magnitude of 1 percent was applied.
- A constraint for maximum error in parallax of 1 percent was applied to ensure minimum uncertainties in distance estimated.

2.2.3 Histograms

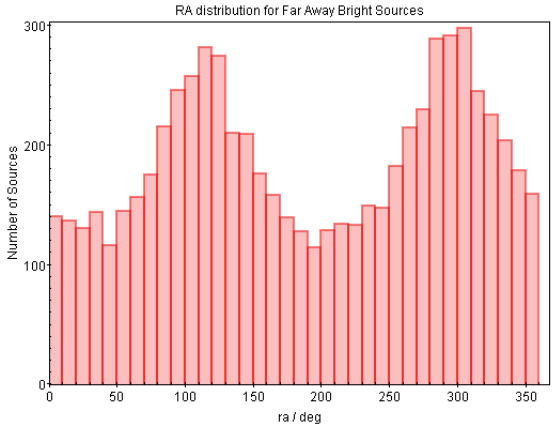


Figure 2.9: RA distribution of Far Away Bright Sources

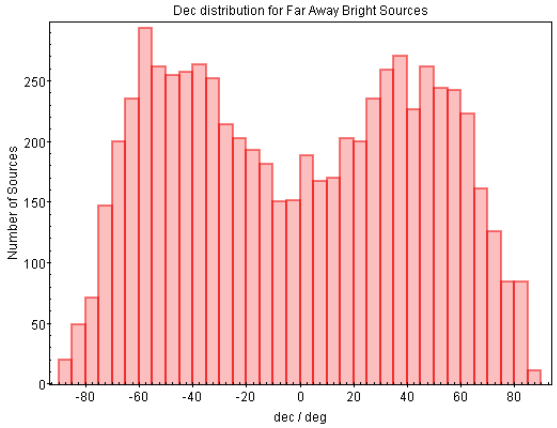


Figure 2.10: Declination distribution of Far Away Bright Sources

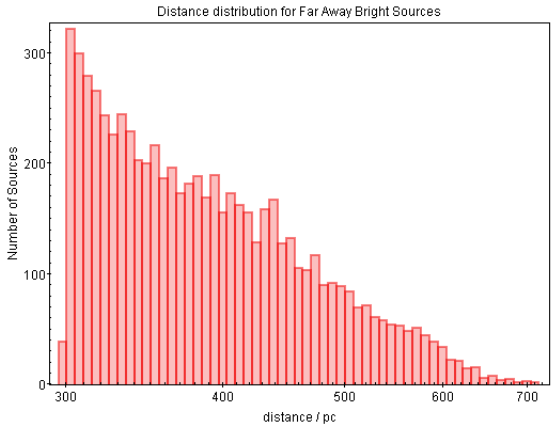


Figure 2.11: Distance distribution of Far Away Bright Sources Due to our constraint on apparent magnitude >8, the number of selected stars decreases with increase in distance as farther stars appear dimmer and have higher apparent magnitudes even for the same luminosity.

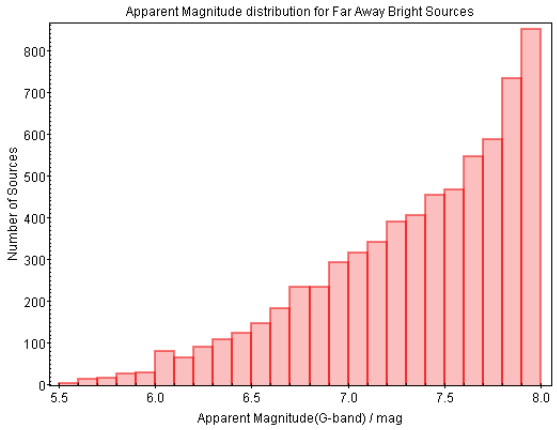


Figure 2.12: Apparent Magnitude distribution - the trend we see matches expectations as far more stars appear dimmer stars than brighter ones

2.2.4 2D plots

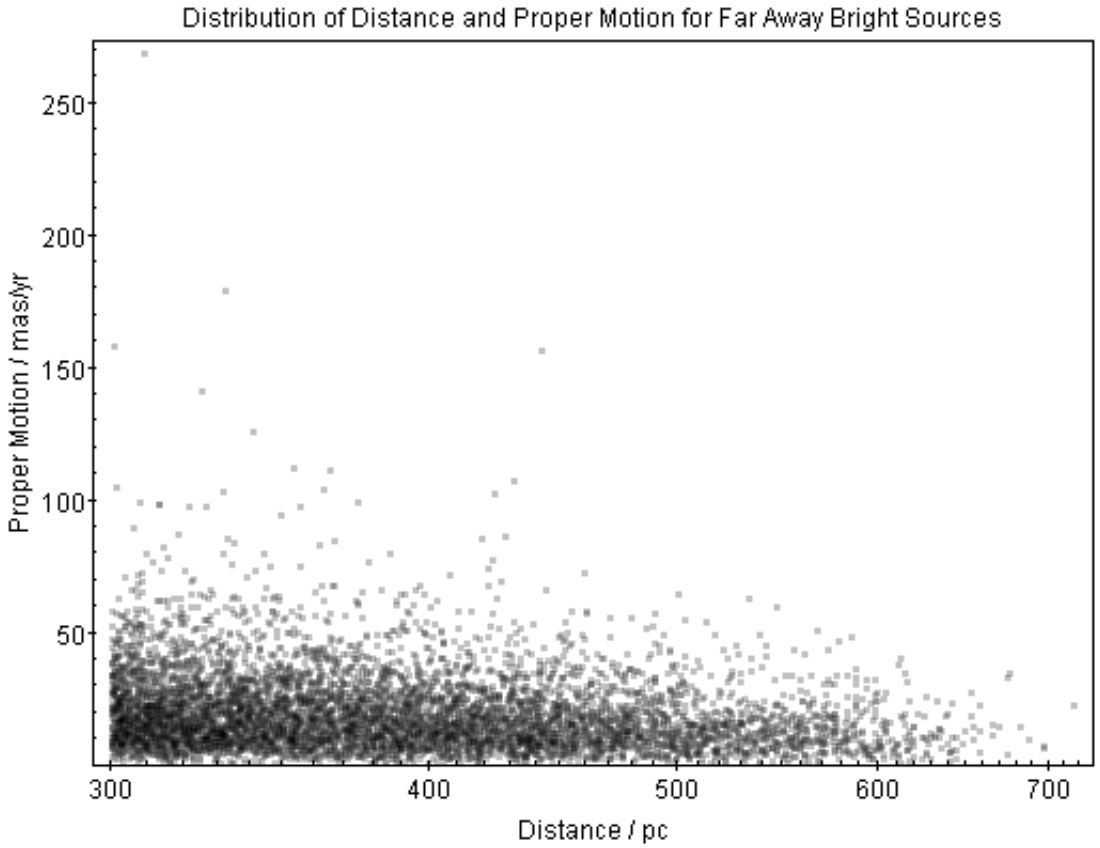


Figure 2.13: Distribution of Distance and Proper Motion of Far Away Bright Sources.Darker tones indicate higher density of sources.

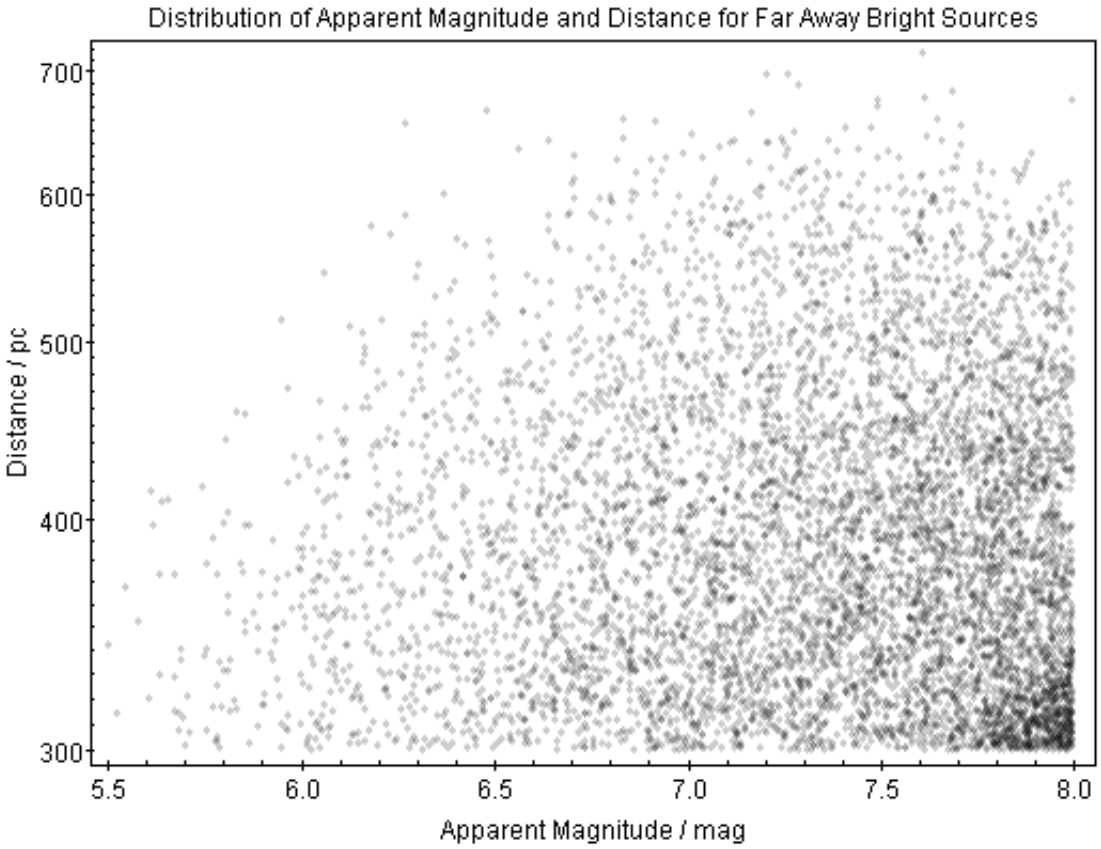


Figure 2.14: Apparent Magnitude vs Distance of Far Away Bright Sources. Darker tones indicate higher density of sources.

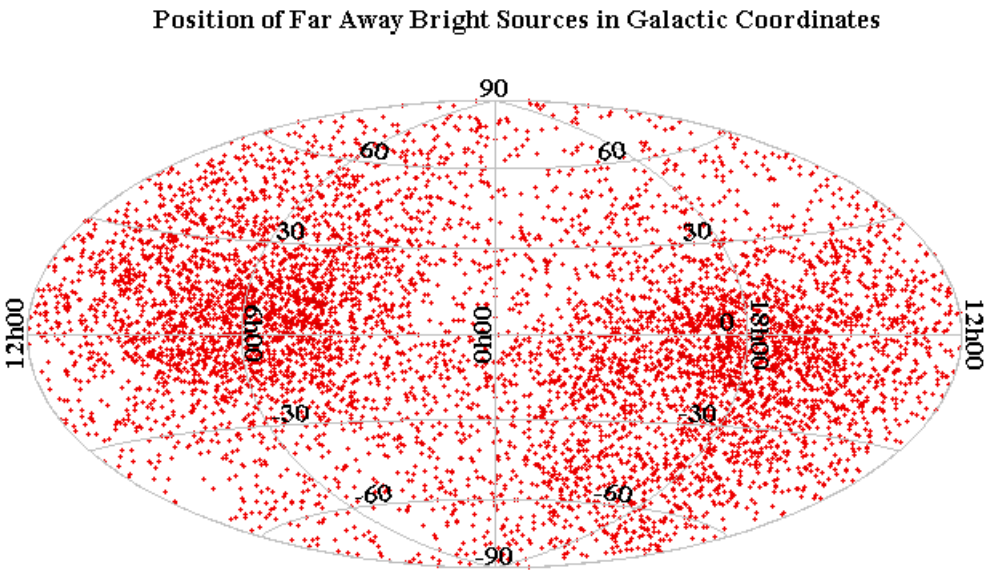


Figure 2.15: Positions of Far Away Bright Sources in Galactic Coordinates

The sky plot seems to have a low number of stars close to galactic centre but this issue vanishes if we relax the constraint on parallax error from < 1 percent to <2 percent. Thus, this pattern emerges only due to some unwanted bias in our selection criteria and not due to any trend in position of stars.

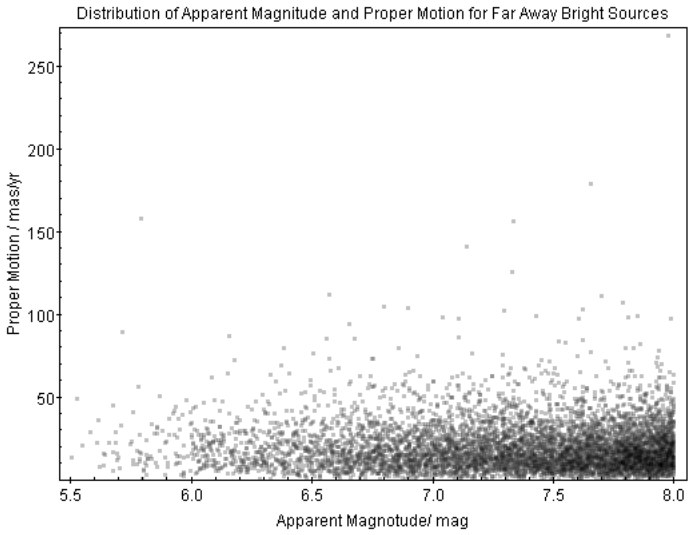


Figure 2.16: Distribution of Apparent Magnitude(G-band) and Proper Motion of Far Away Bright Sources. Darker tones indicate higher density of sources.

2.3 Omega Centauri Globular Cluster

Motivation

Globular clusters are tightly packed groups of several hundred to thousand stars which are generally old and have a common motion through space. Analyzing star cluster data provides insights into their formation as well as of the galaxy's past as a whole.

Querying the Data

The ADQL query needed to fetch stars in a certain angular window, around the coordinates of the cluster, and the parallax was kept in limits to exclude extreme outliers:

```
query = '''SELECT TOP 50000 g.ra, g.dec, g.parallax, g.phot_g_mean_mag,
g.pm, h.teff_gspphot, DISTANCE(
   POINT(201.7, -47.48),
   POINT(g.ra, g.dec)) AS ang_sep
FROM gaiadr3.gaia_source AS g, gaiadr3.astrophysical_parameters AS h
WHERE 1 = CONTAINS(
   POINT(201.7, -47.48),
   CIRCLE(ra, dec, 0.6))
AND g.source_id = h.source_id
AND phot_g_mean_mag < 20.5
AND parallax IS NOT NULL
AND parallax BETWEEN 0.18 AND 0.25
ORDER BY ang_sep ASC'''
```

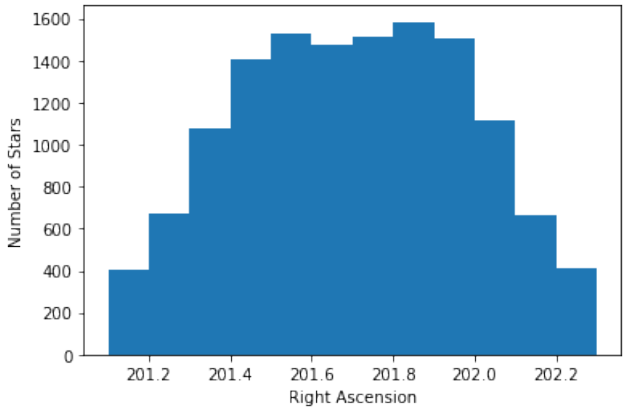


Figure 2.17: Distribution of RA

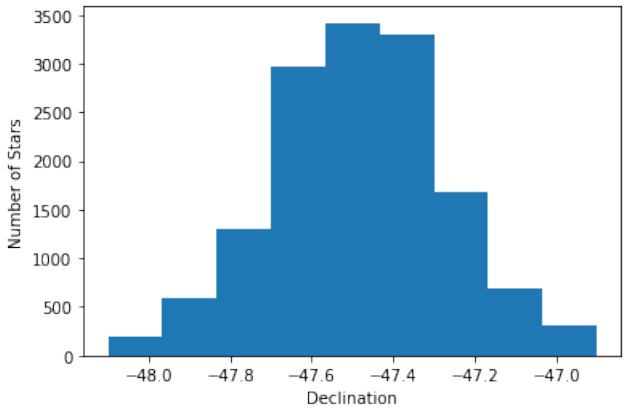


Figure 2.18: Distribution of Declination

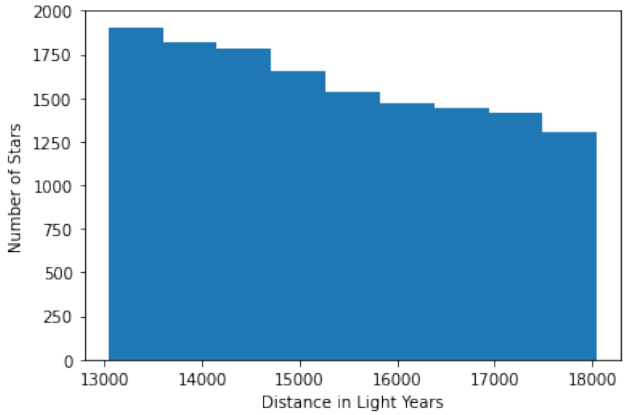


Figure 2.19: Distribution of Distances in the ADQL range

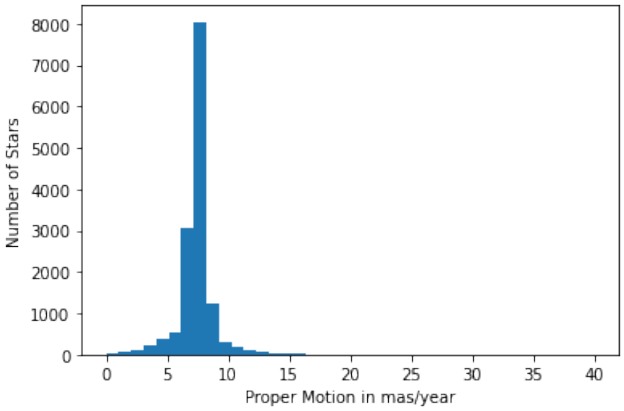


Figure 2.20: Distribution of Proper Motions

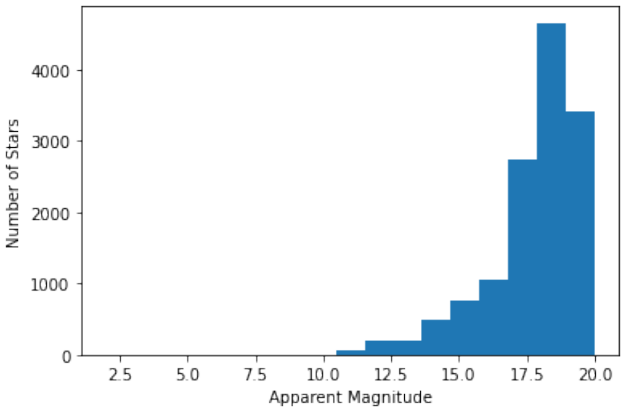


Figure 2.21: Distribution of Apparent Magnitudes

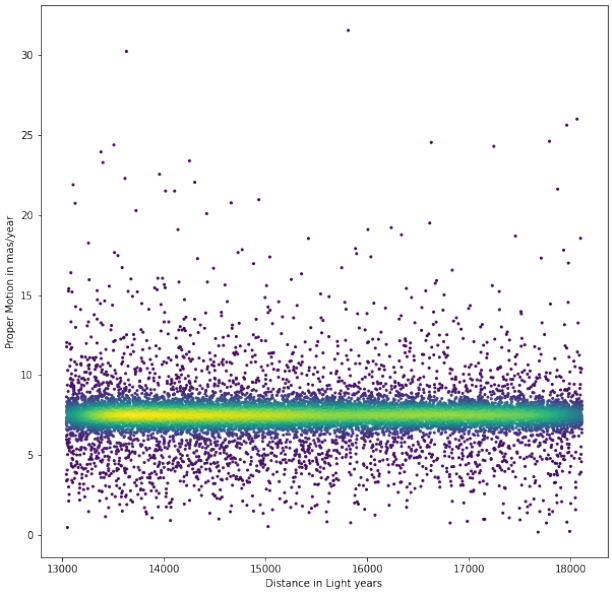


Figure 2.22: Plot of Proper Motion vs distance - most stars lie very close in proper motion, as is expected from a star cluster

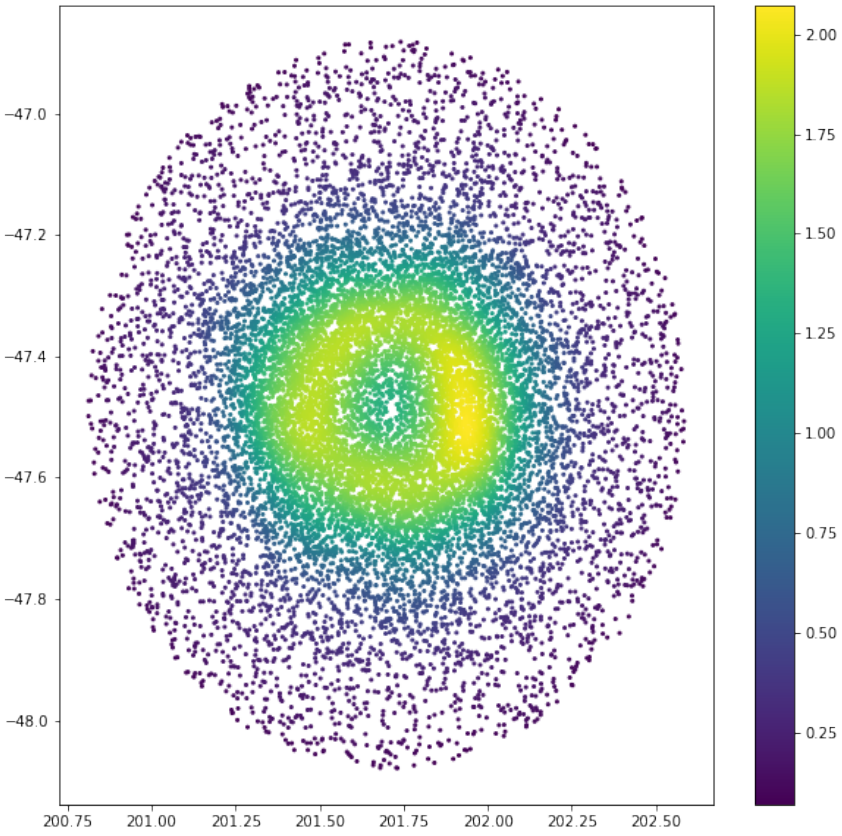


Figure 2.23: Plot of RA and Dec - for a very small window of the sky, linear plot was preferable over Mollweide plot. Numbers on the colour bar represent density of stars at that coordinate

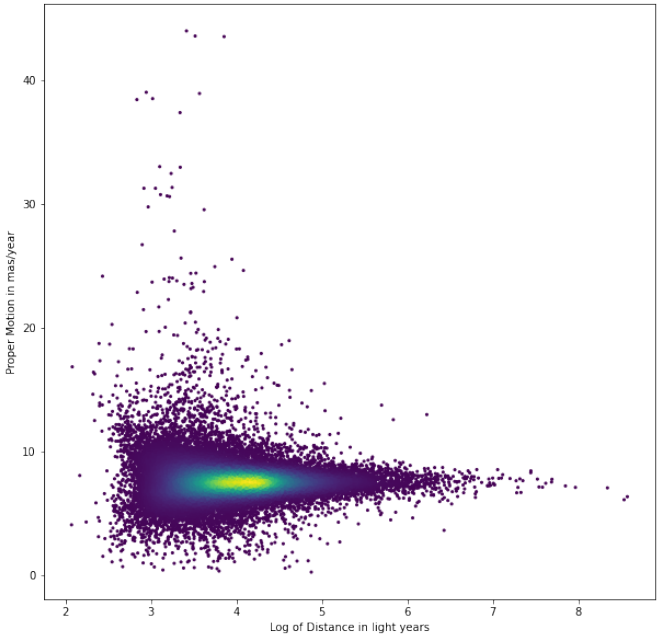


Figure 2.24: Plot of log of Distance and Proper Motion. The logarithmic plot allows to include outliers too.

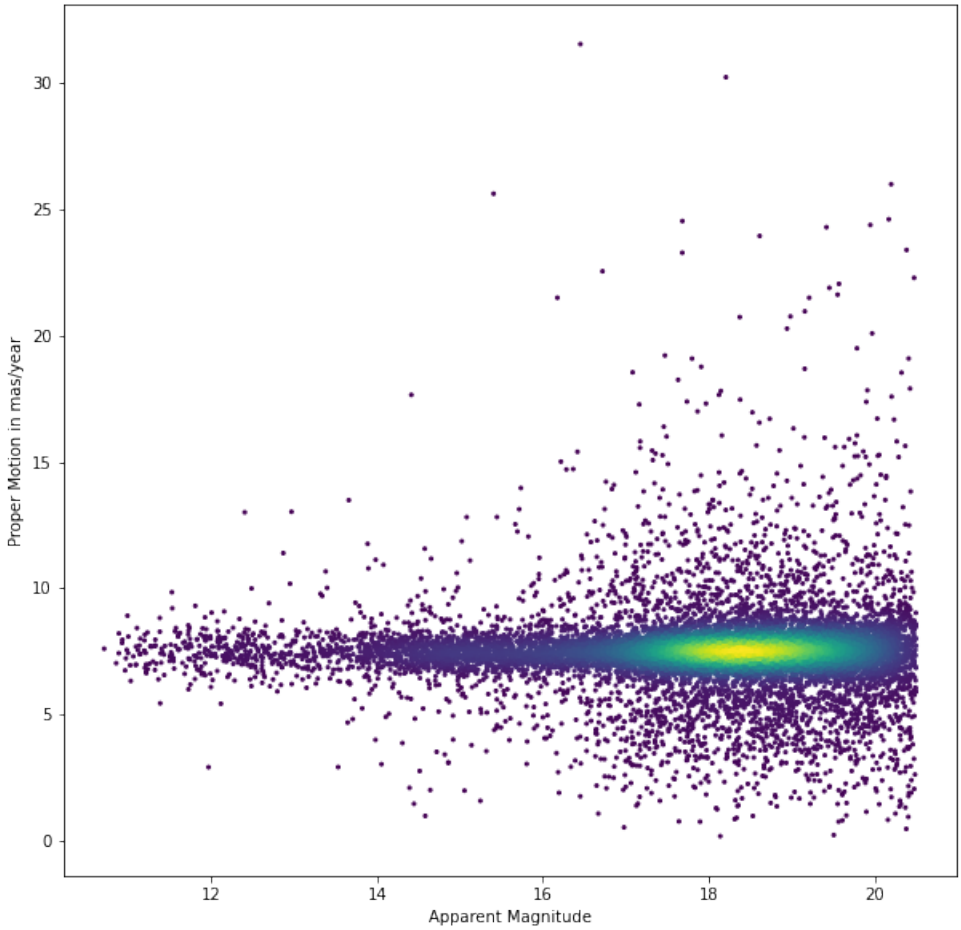


Figure 2.25: Plot of Proper Motion vs apparent magnitude

2.4 Fastest Stars

The ADQL query used

```
query = """
SELECT TOP 10000
source_id, ra, dec, parallax,phot_g_mean_mag,
radial_velocity, radial_velocity_error, pm
From gaiadr3.gaia_source as g
WHERE radial_velocity is NOT NULL
AND pm is NOT NULL
AND parallax is NOT NULL
AND phot_g_mean_mag is NOT NULL
ORDER BY  radial_velocity DESC
"""
```

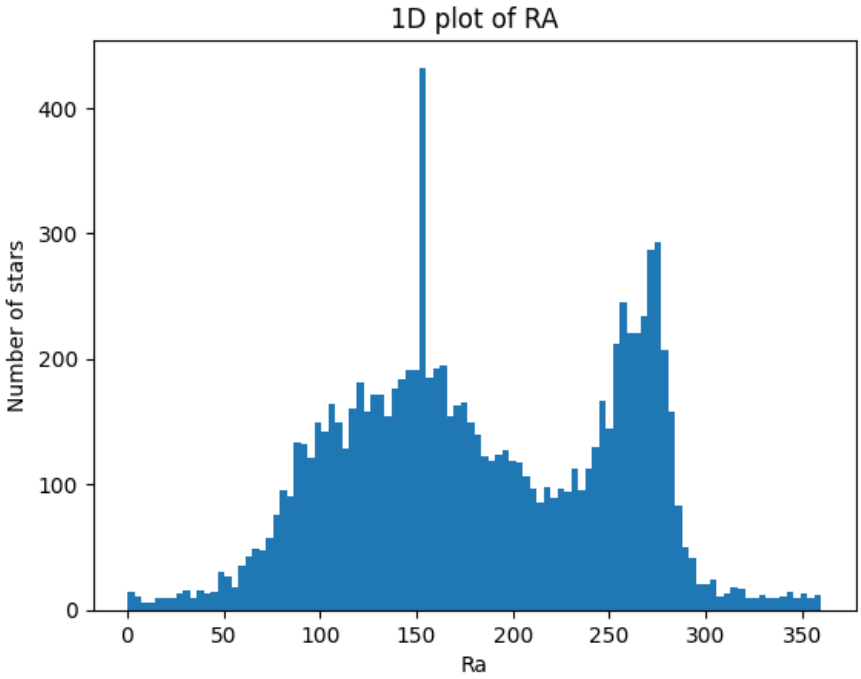


Figure 2.26

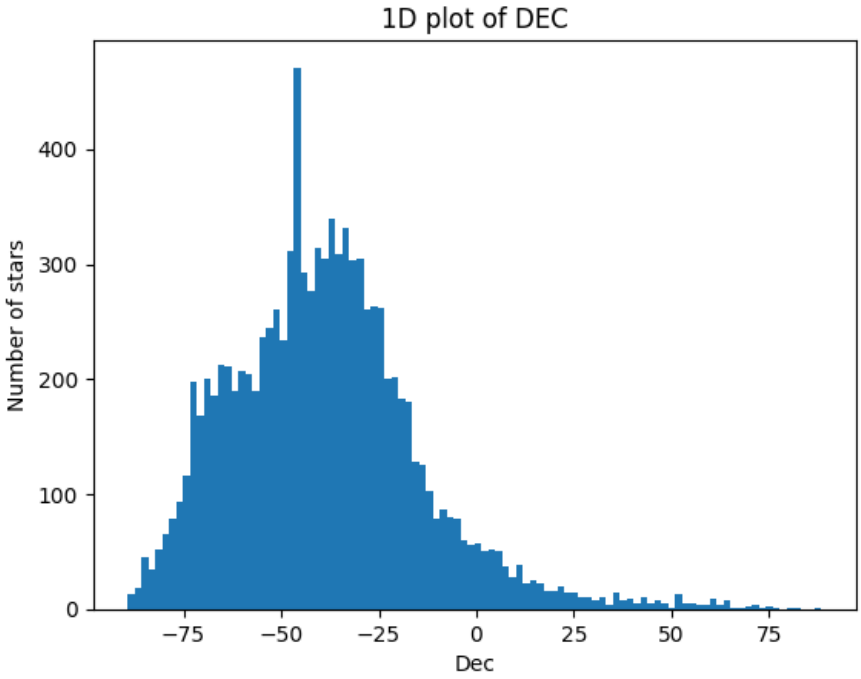


Figure 2.27

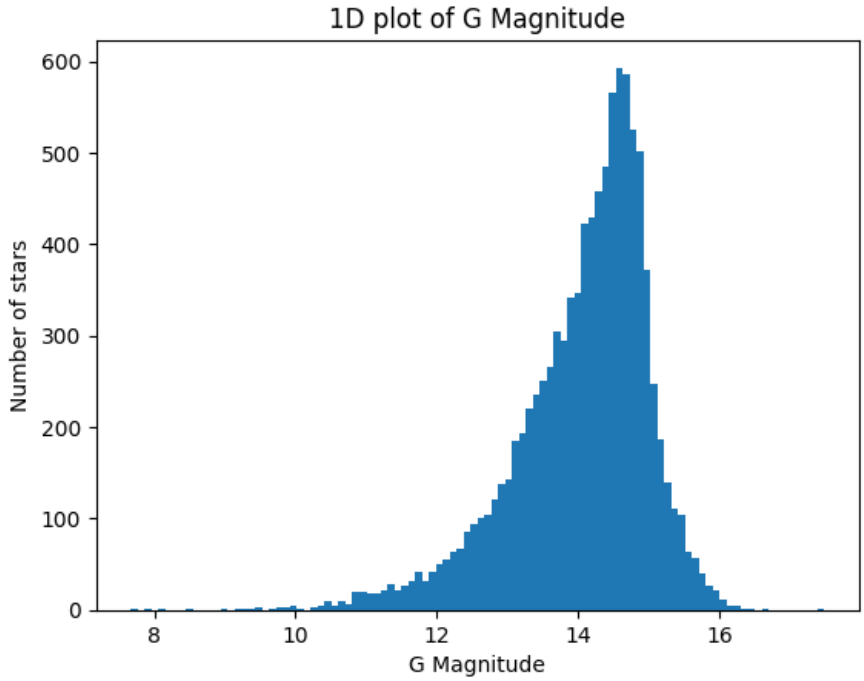


Figure 2.28

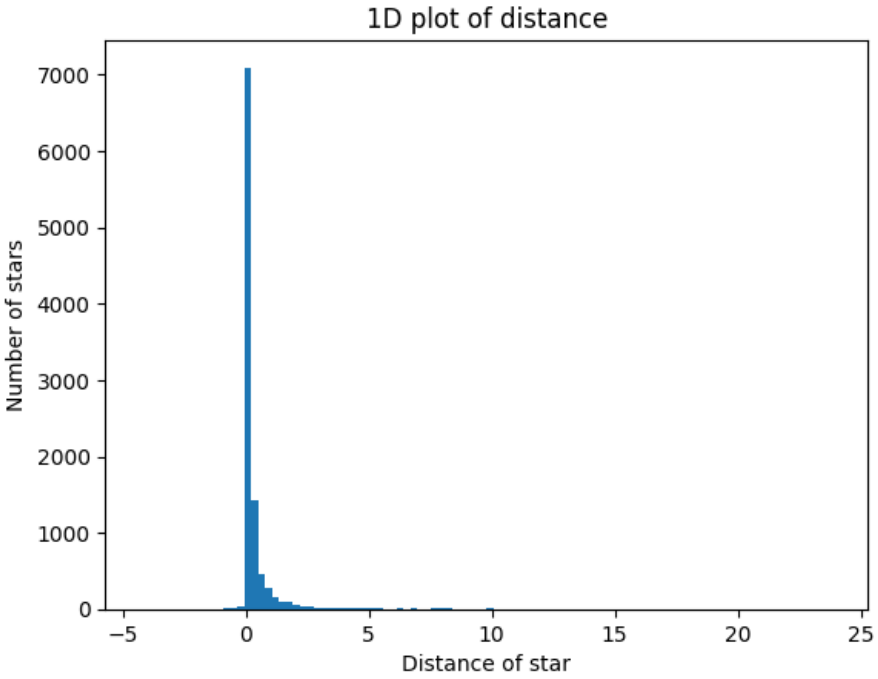


Figure 2.29

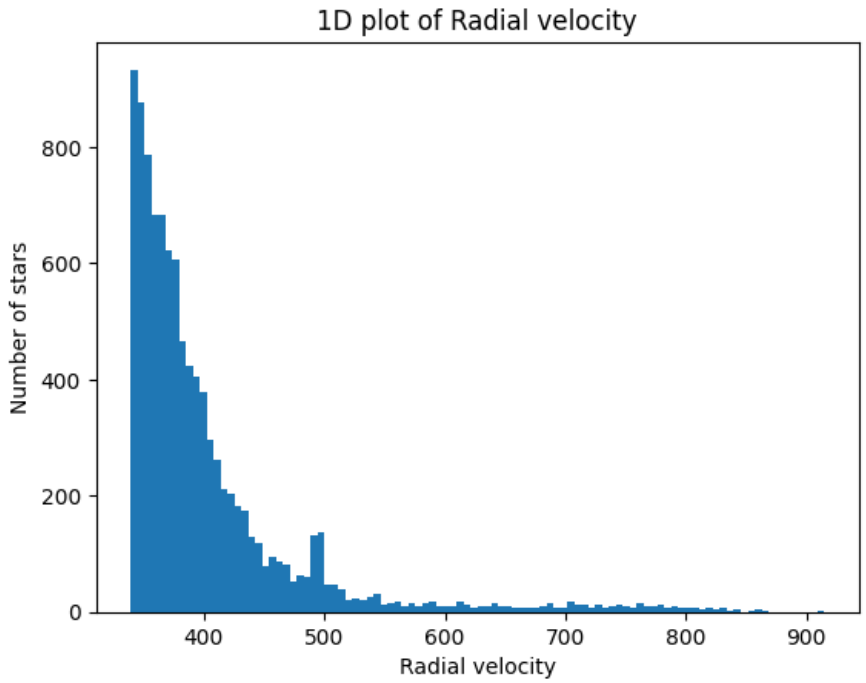


Figure 2.30

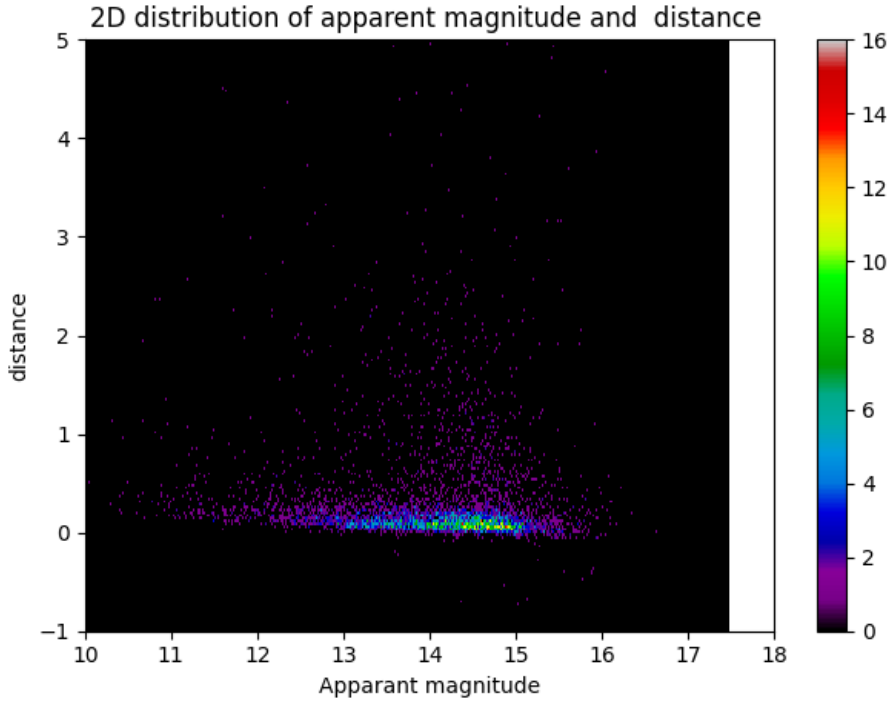


Figure 2.31

2.5 Closest Stars (within 60 parsecs)

2.5.1 Selection Criteria and Motivation

Those stars were selected which have

- Distance $<$ 60 parsecs
- Parallax error $< 1\%$
- Parallax > 0

All the stars within 60 pc of the sun and with 1 percent parallax error were selected for this study. These close stars have very precise astrometric and photometric data. This precision helps in the accurate generation of the HR diagram.

Further, the white dwarfs were selected and their distribution and proper motion was studied.

2.5.2 ADQL query

```
query = '''SELECT
          g.source_id, g.ra, g.dec, g.parallax , g.phot_g_mean_mag AS mag,
          g.phot_rp_mean_mag as magrp, g.phot_bp_mean_mag as magbp,
          g.parallax_error as par_err, d.r_med_geo as dspc_geo,
          d.r_med_photogeo as dspc_photogeo, g.pmra as pmra, g.pmdec as pmdec
          FROM gaia.edr3lite AS g, gedr3dist.main AS d
          WHERE g.source_id = d.source_id
          AND g.parallax_error/g.parallax < 0.01
          AND g.parallax > 0 AND g.parallax_error > 0
          AND d.r_med_geo < 60 AND g.phot_g_mean_mag IS NOT NULL
          AND g.phot_rp_mean_mag IS NOT NULL AND
          g.phot_bp_mean_mag IS NOT NULL'''
```

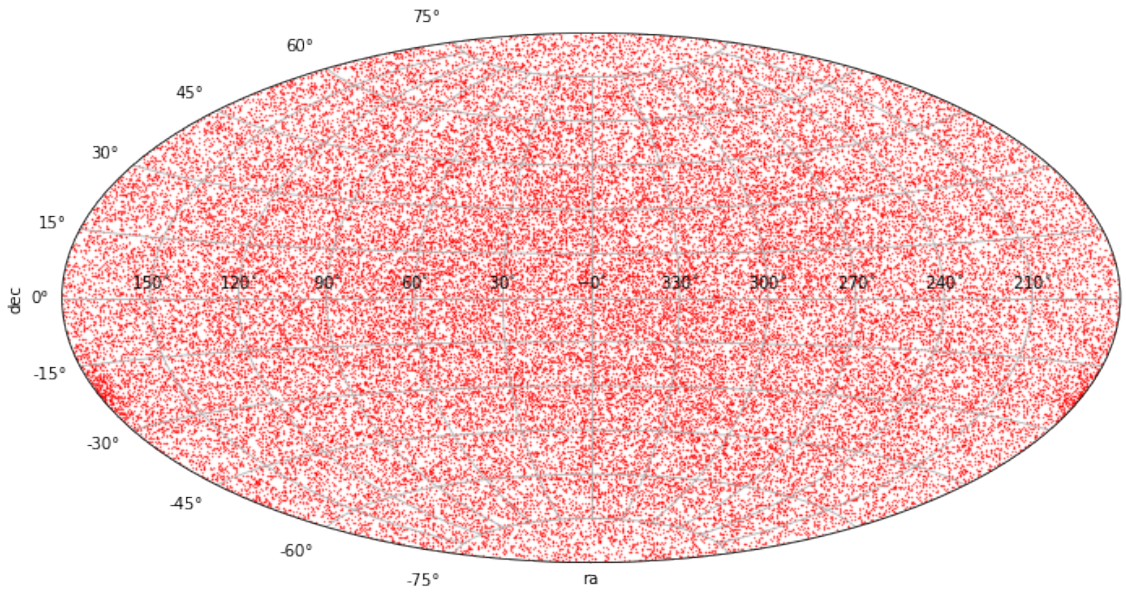


Figure 2.32: Distribution in Galactic coordinates

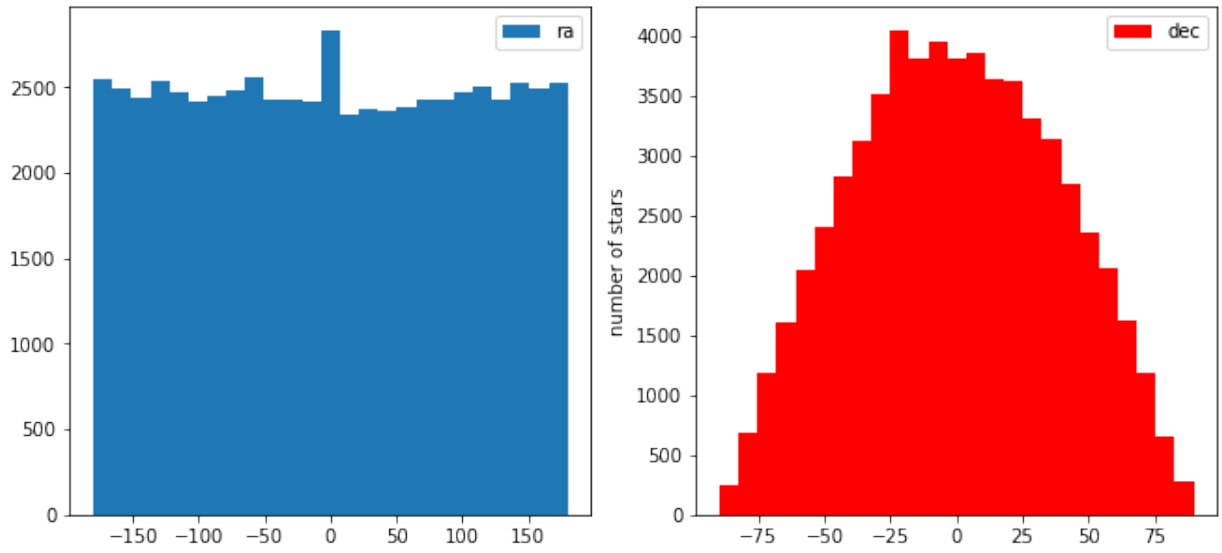


Figure 2.33: ra and dec distribution

The left plot is the histogram plot of star count vs RA and right plot is of star count vs Dec. The distribution of stars is very uniform, which points to almost no major structures within 60 pc of the sun.

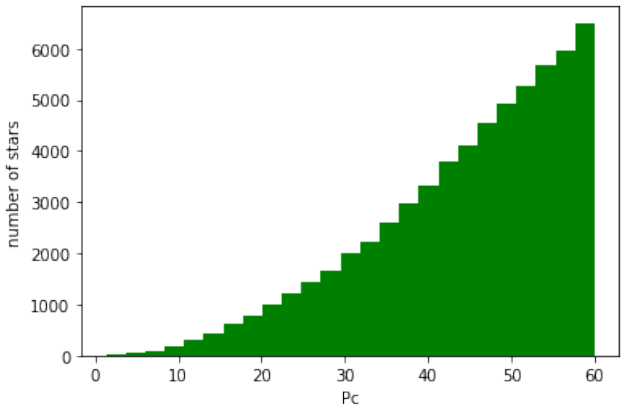


Figure 2.34: distance (in parsecs) distribution

This uniform increase in the number of stars too suggests that there are no local collection/group of stars within 60 pc of the sun.

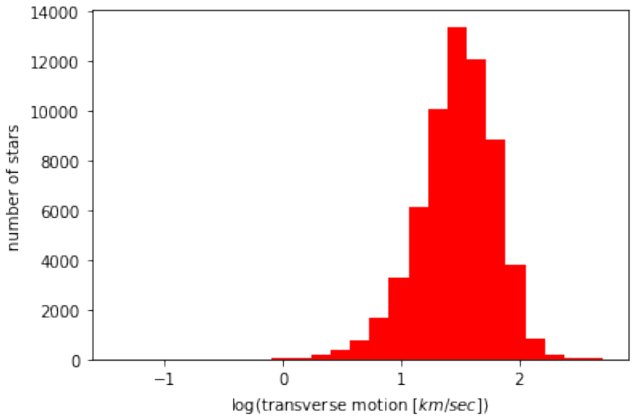


Figure 2.35: transverse motion distribution (base of log = 10)

$$T_v = 4.74\left(\frac{\mu D}{1000}\right)$$

Here, μ = proper motion (milli arc sec) and D = distance (parsecs).

The majority of stars have transverse velocity between 10 and $100\ km/sec$. The mean transverse velocity is 37.74 km/sec.

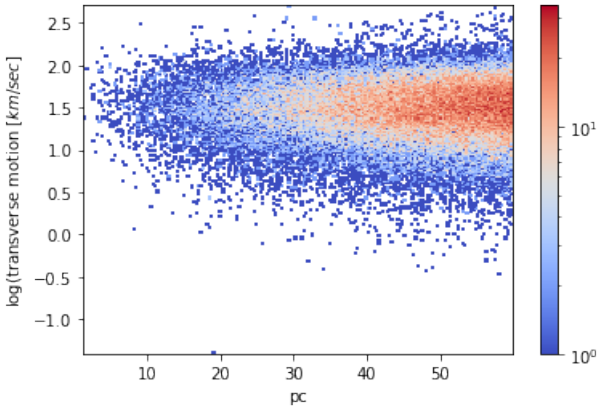


Figure 2.36: Transverse velocity vs distance (base of log = 10)

There is no affect of distance on transverse velocity. But the lower limit of velocity keeps on decreasing as distance increases.

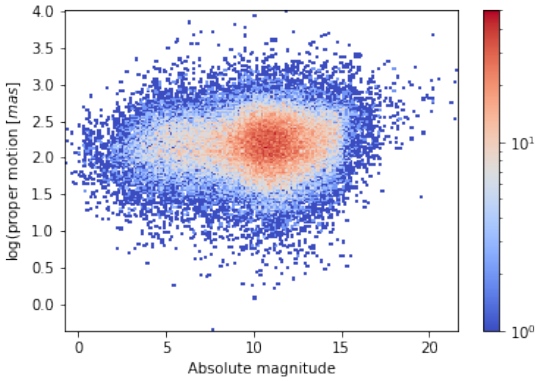


Figure 2.37: Proper motion vs absolute magnitude (base of log = 10)

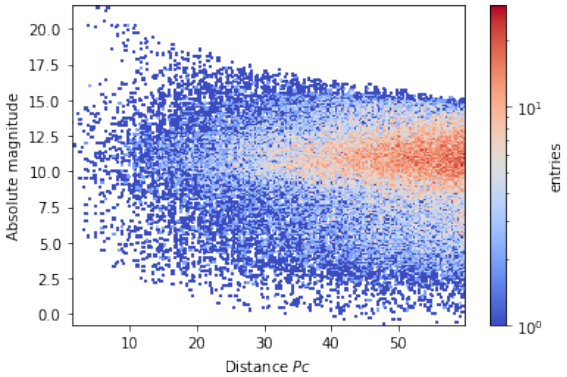


Figure 2.38: Absolute magnitude vs distance

As distance increases, number of stars with higher absolute magnitude (dim stars) decreases. At 60 pc, there are no stars with magnitude of 15 or above (there is a clear dividing line above of which, the reading abruptly is zero). Since 1% accuracy was demanded in the query, it seems that dim stars have higher uncertainty in their astrometric measurement, thus got excluded from the data.

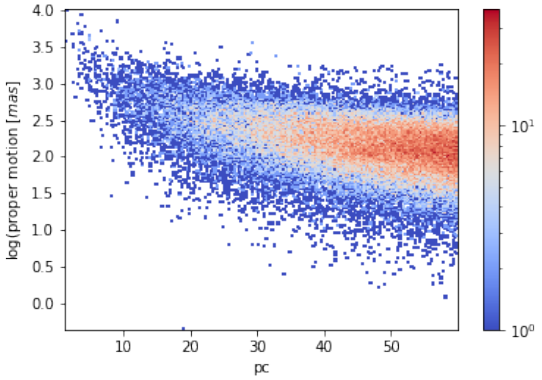


Figure 2.39: Proper motion vs distance (base of log = 10) : proper motion decreases with increasing distance as expected

2.6 Large Magellanic Cloud

The Gaia data of the LMC was chosen for analysis because like many irregular galaxies, it is rich in gas and dust, and is currently undergoing vigorous star formation activity. The LMC holds the Tarantula Nebula, the most active star-forming region in the Local Group. Added to that, it has a wide range of galactic objects and phenomena that make it known as an astronomical treasure-house, a great celestial laboratory for the study of the growth and evolution of the stars. Supernova 1987A (the nearest supernova in recent years) was in the Large Magellanic Cloud.

Astronomers discovered a new black hole inside the Large Magellanic Cloud in November 2021 using the European Southern Observatory's Very Large Telescope in Chile. It is claimed that its gravity is influenced by a nearby star, which is about five times the mass of our Sun.

The LMC also serves as a crucial midway point between the stars of our own Milky Way Galaxy (which we have some hope of studying in detail) and galaxies in the distant universe (which we can't study in detail). Astronomers use it as a milepost, measuring the distances to other galaxies relative to the LMC. Due to these features, the LMC makes for an important subject of study.

2.6.1 Query

```
query = '''select top 30000 g.source_id, g.pm, g.ra, g.dec, g.parallax,
g.phot_g_mean_mag, g.phot_bp_mean_mag, phot_rp_mean_mag,
parallax_over_error, h.teff_gspphot
from gaiadr3.gaia_source as g, gaiadr3.astrophysical_parameters as h
where g.source_id = h.source_id and
g.ra between 85 and 95 and
g.dec between -74 and -64 and
parallax > 0.01 and parallax < 0.03
order by g.parallax_over_error desc'''
```

2.6.2 Data visualization

As can be seen, the stars are distributed in a circular window around fixed RA and Dec values.

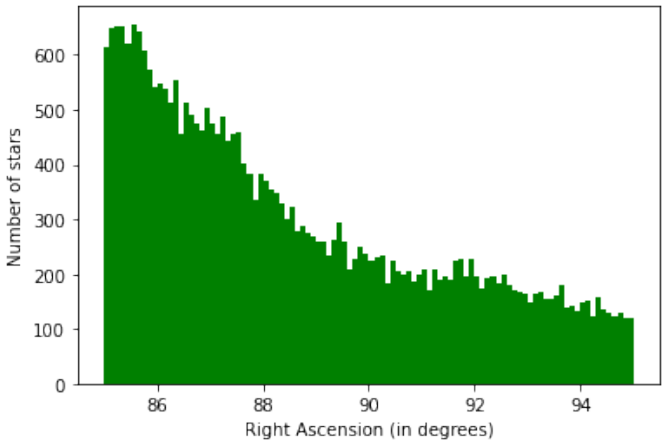


Figure 2.40: RA distribution of the stars

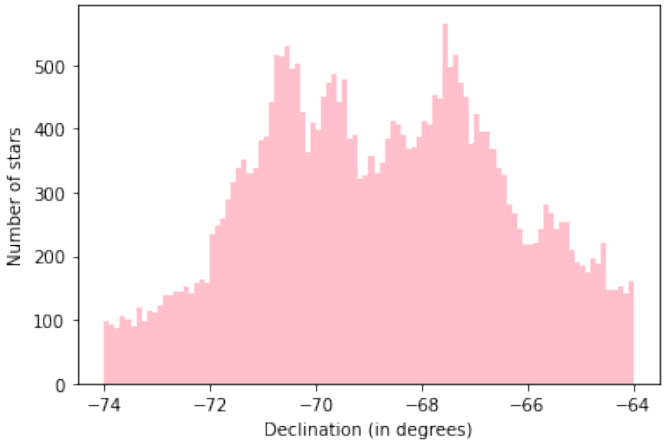


Figure 2.41: Declination distribution of the stars

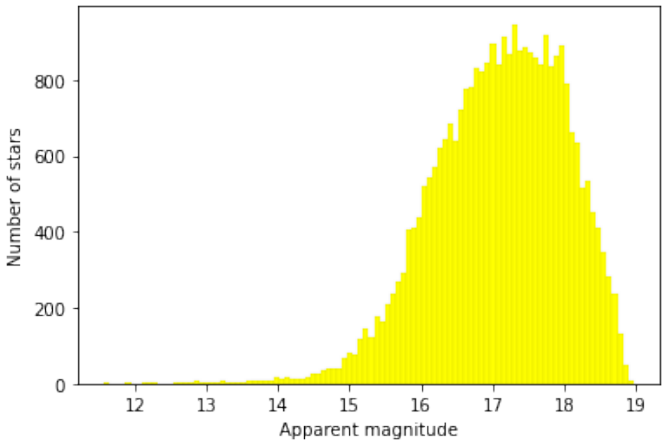


Figure 2.42: 1D distribution of apparent magnitude

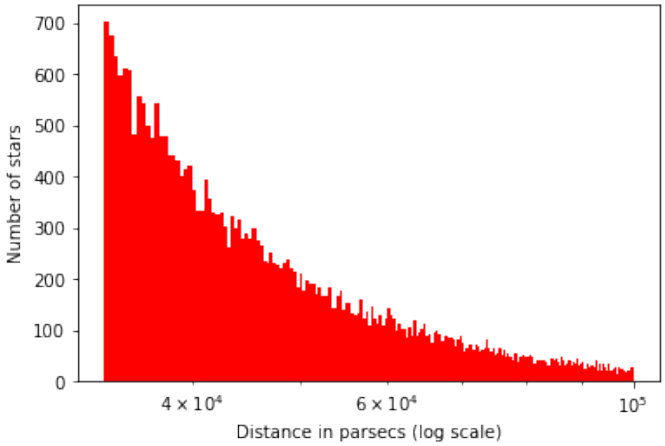


Figure 2.43: 1D distribution of distance

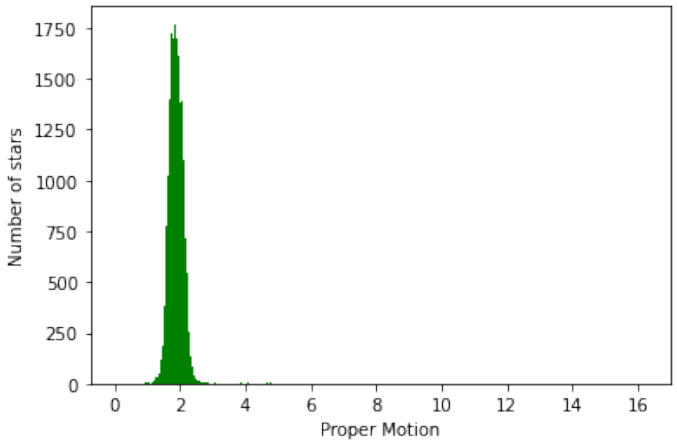


Figure 2.44: 1D distribution of Proper Motion

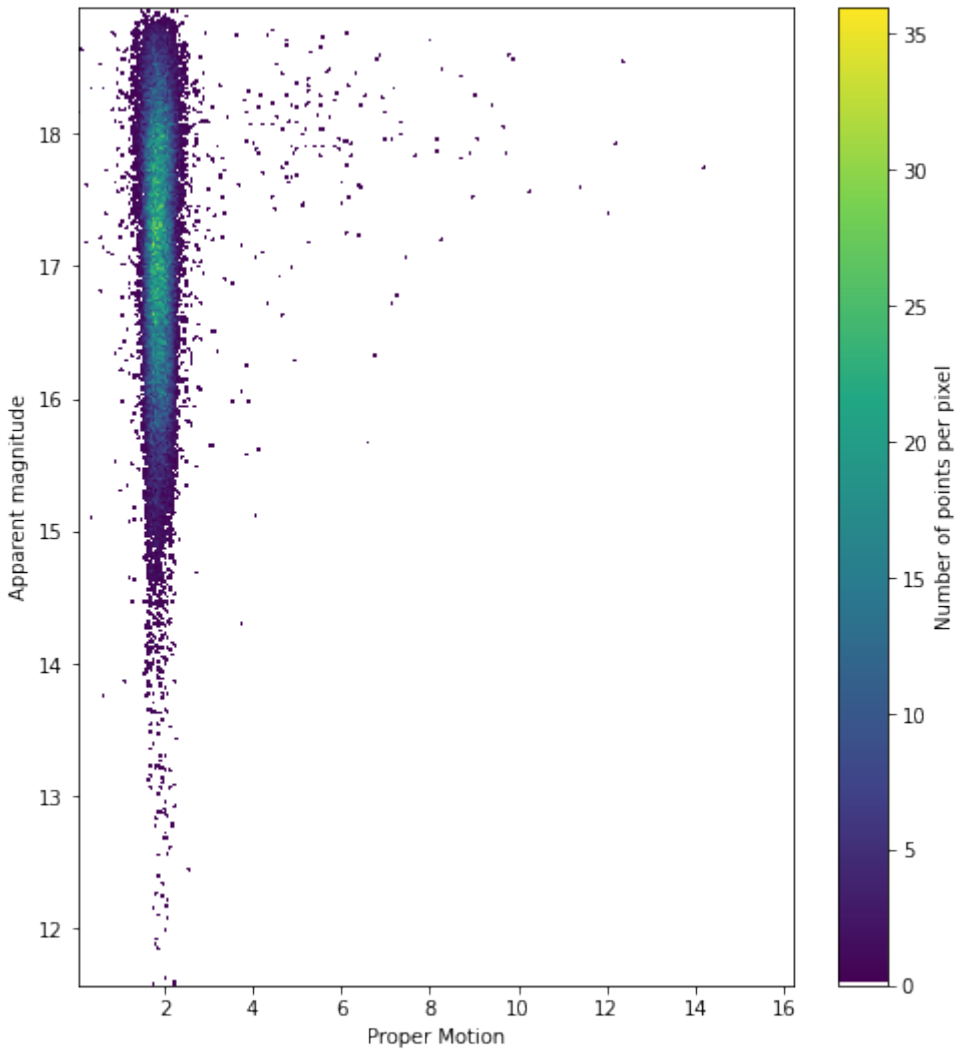


Figure 2.45: Apparent magnitude vs Proper Motion of LMC stars

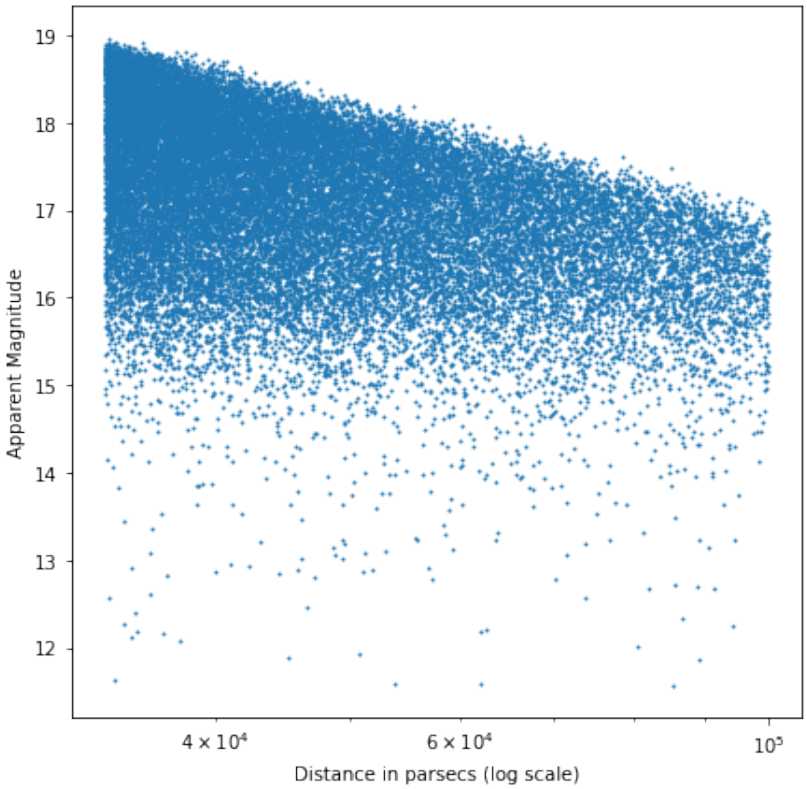


Figure 2.46: Apparent magnitude vs Distance of LMC stars

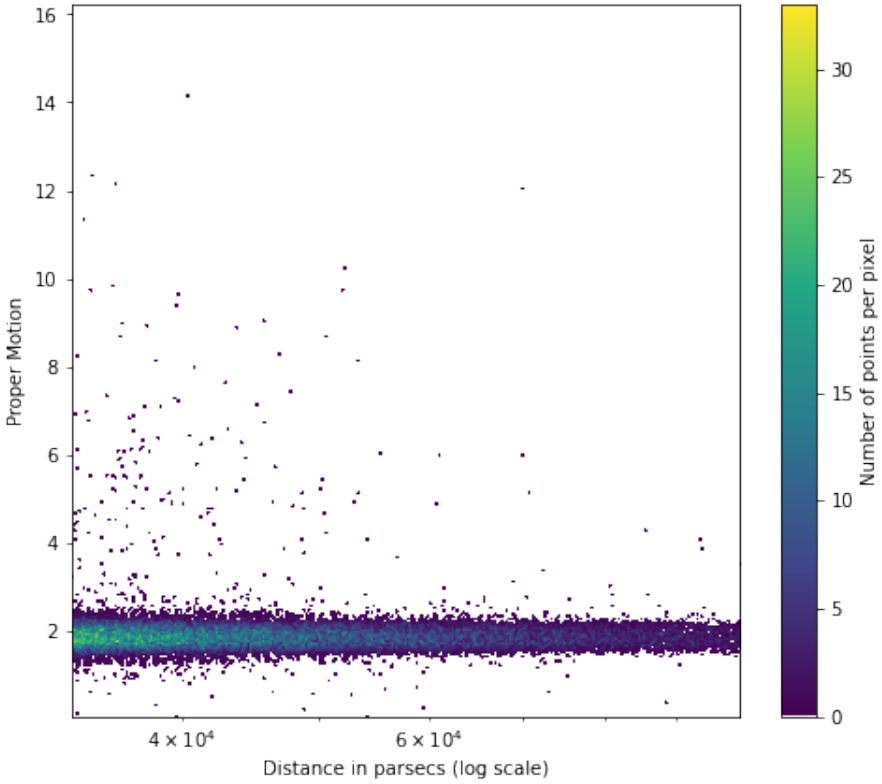
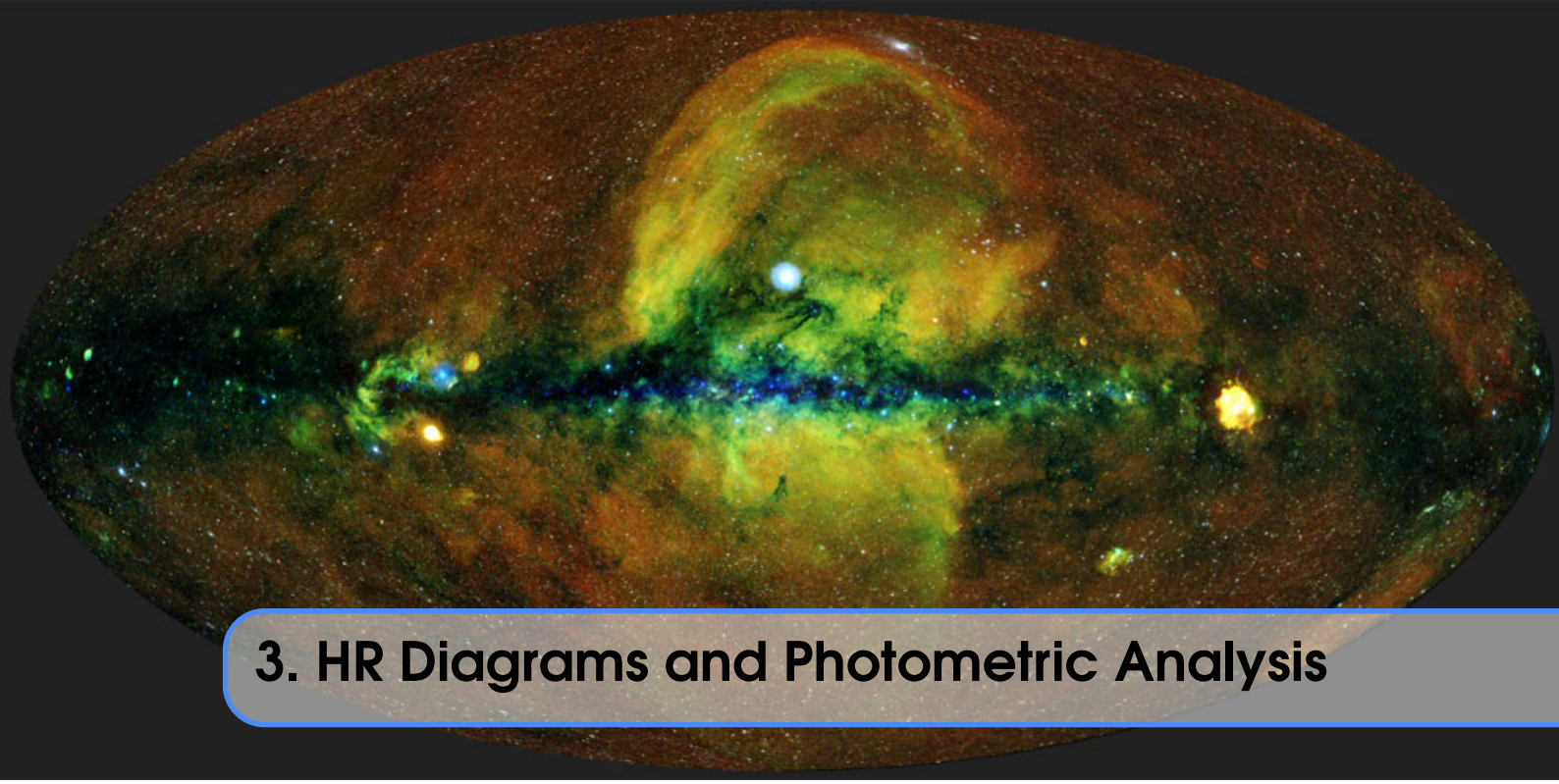


Figure 2.47: Proper Motion vs Distance of LMC stars



3. HR Diagrams and Photometric Analysis

3.1 The Physics of Stars

A star can be approximated quite well as a blackbody. For the same, we have the typical radiation curve and Planck's Law as show below

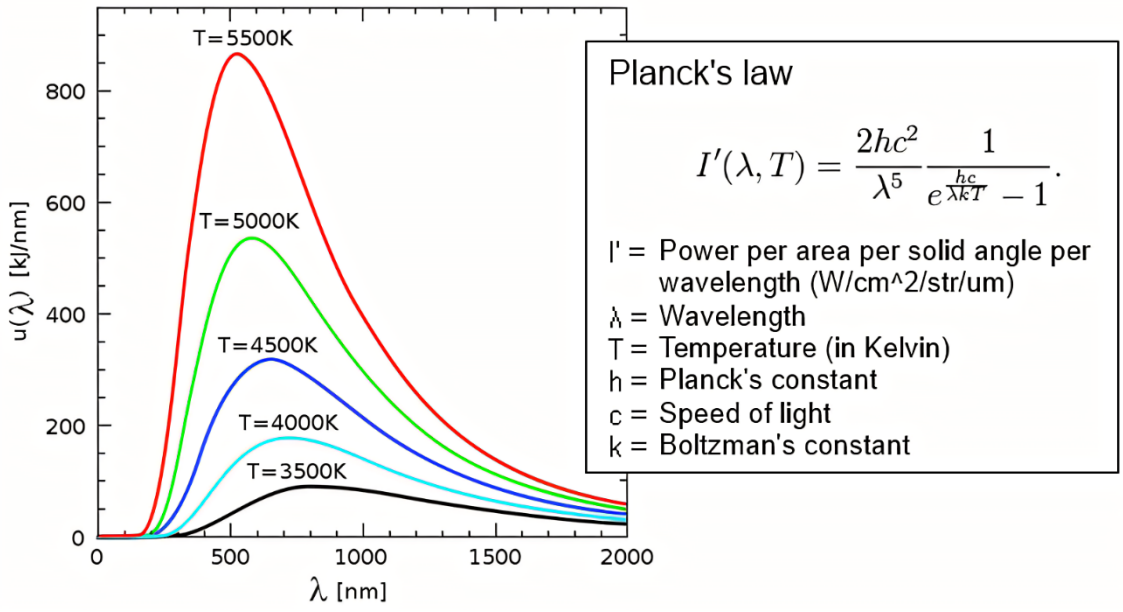


Figure 3.1: Blackbody Radiation Curve

It is easily seen that as temperature increases, the peak shifts to the left. Now consider this with respect to the blue pass and red pass filters. At a lower temperature, there is more flux from star at higher wavelengths i.e. in the red band. So G_{RP} is lower in value than G_{BP} or $G_{BP} - G_{RP}$ is more positive. The opposite happens

for higher temperature stars, i.e. $G_{BP} - G_{RP}$ is less positive or negative. So if we plot $G_{BP} - G_{RP}$ on one axis, it gives a good idea of the temperature.

Meanwhile, consider the flux-luminosity relation given by

$$F = \frac{L}{4\pi r^2}$$

and the absolute magnitude

$$M = -2.5\log_{10}\left(\frac{F}{F_0}\right) = -2.5\log_{10}(L) + c$$

where c is a fixed constant. Thus absolute magnitude serves as a measure of luminosity.

Finally, it may be noted that luminosity is related to surface temperature as

$$L = 4\pi\sigma R^2 T^4$$

where R is the radius and T the surface temperature of the star.

3.1.1 HR Diagrams

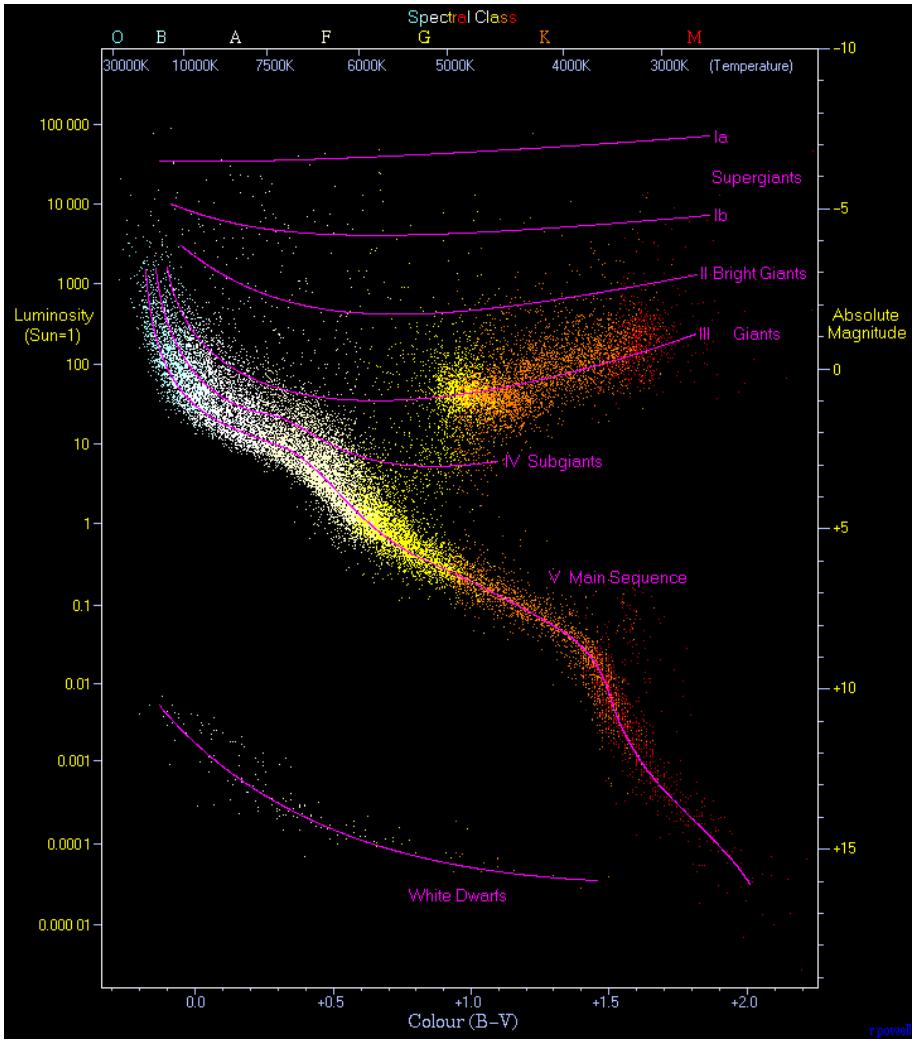


Figure 3.2: Standard HR Diagram

Normally, one would expect there to be no relation between the luminosity (absolute magnitude) and temperature ($G_{BP} - G_{RP}$) due to the radius of star being a uniformly varying value. However, the internal structure of the star (mainly its mass) dictates that only specific combinations of these are allowed, which was deduced by studying stellar data extensively. These are shown on the Hertzsprung-Russell diagrams (HR diagrams) which look something like above.

Comparing HR diagrams of Gaia star samples with these helps us to discern several interesting properties of these stars.

3.2 Brightest Stars

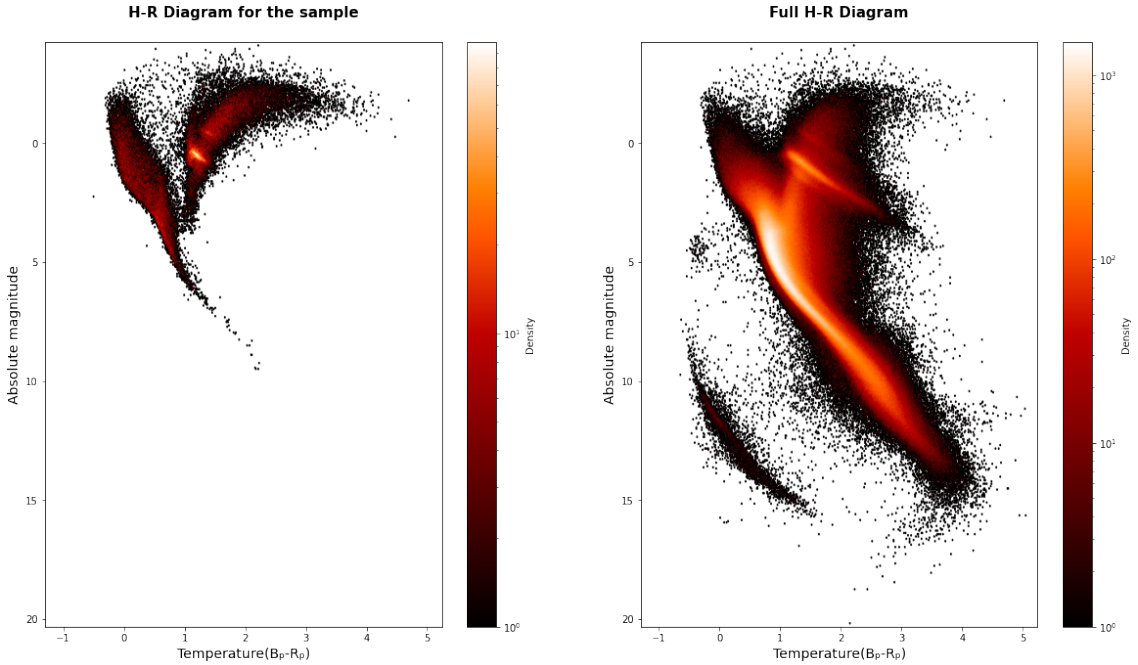


Figure 3.3: H-R diagrams.

The left H-R diagram is generated for our sample of 40,000 brightest stars. The right H-R diagram is generated from all the stars in GAIA catalog that satisfy the conditions

- Parallax error $< 2\%$.
- Parallax > 0.
- Parallax and G_{BP}, G_{RP} are not null.

which turns out to be about 30,00,000 stars.

Comparing the above two diagrams we can see that our sample contains mainly stars from Giant branch and high temperature, high mass main sequence stars. It's also interesting to see that our sample doesn't contain any white dwarfs. This is mainly due to their small size as luminosity is directly proportional to radius of star. The presence of giant stars with lower surface temperature compared to the main sequence stars in our sample can mainly be attributed to their huge size as luminosity $L = 4\pi\sigma R^2 T^4$. The main sequence stars are present mainly due to their higher surface temperatures.

3.2.1 Comparison Giant branch stars and Main Sequence Stars.

For selecting stars in giant branch we consider the area where $B_p - R_p > 0.9$ and absolute magnitude < 3.8 as shown below.

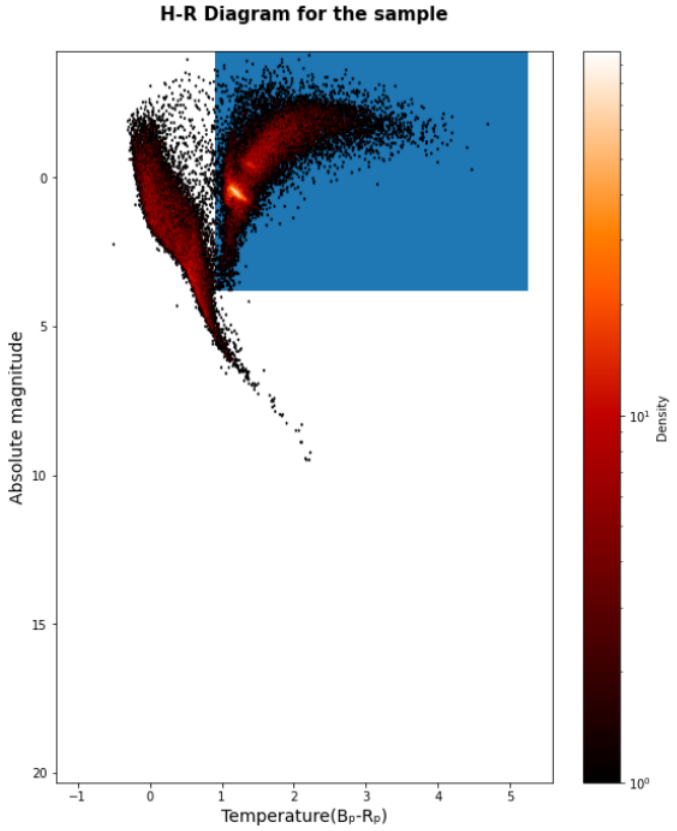
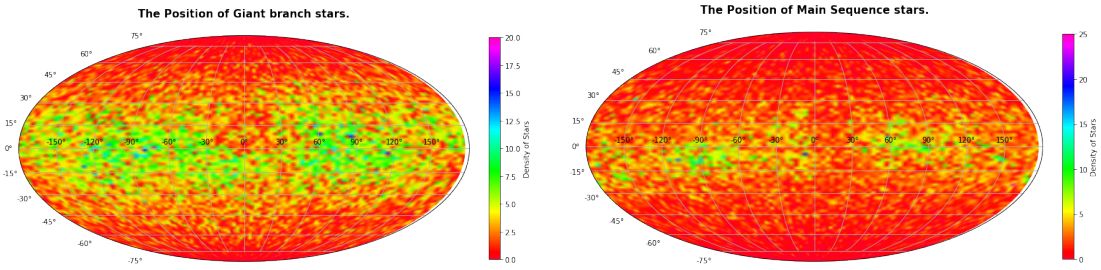


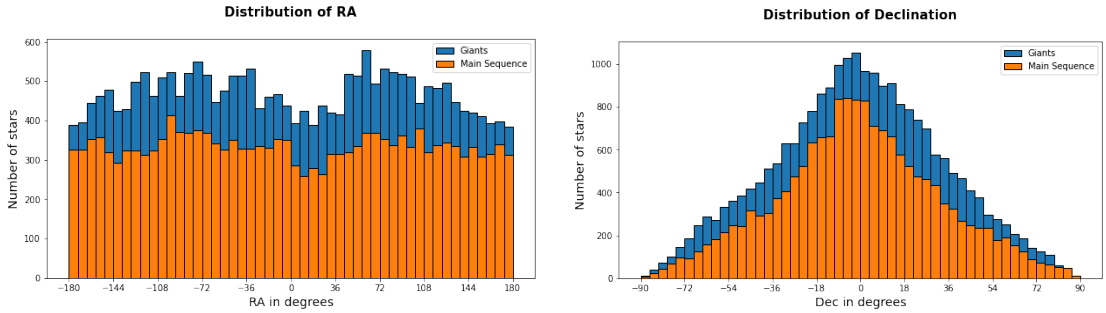
Figure 3.4: Giant branch selected from H-R diagram.

Now comparing Giant branch stars to main sequence we get following graphs.



(a) Position of Giant branch stars. (b) Position of Main sequence stars.

Figure 3.5: Position of 2 class of stars.



(a) RA distribution. (b) Dec distribution.

Figure 3.6: RA and DEC of 2 class of stars.

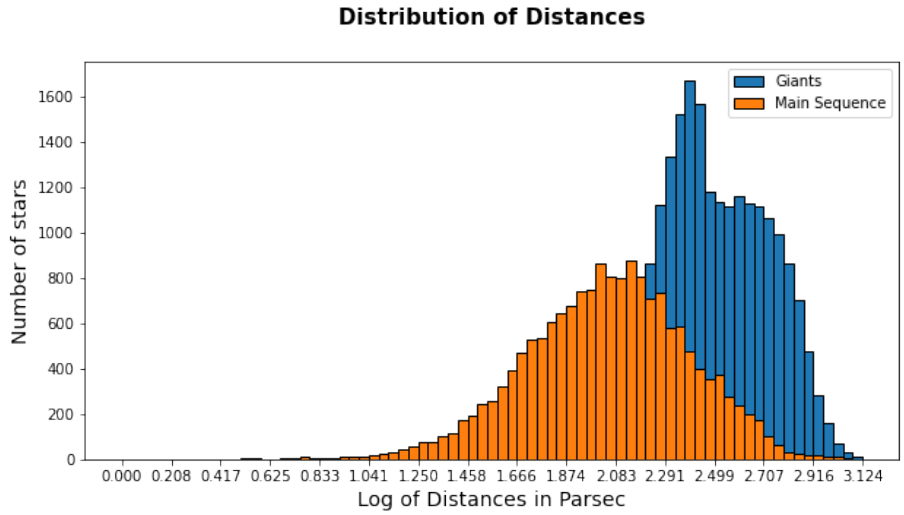


Figure 3.7: Distance distribution.

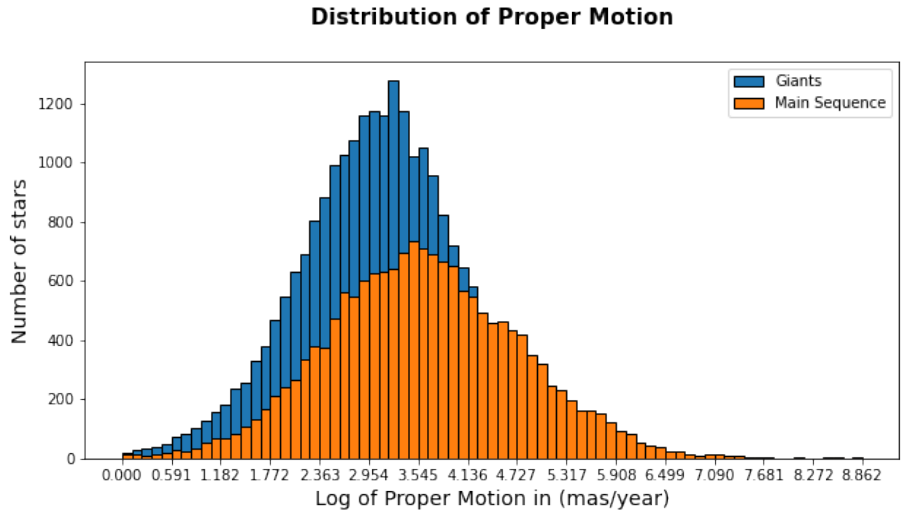


Figure 3.8: Proper motion distribution.

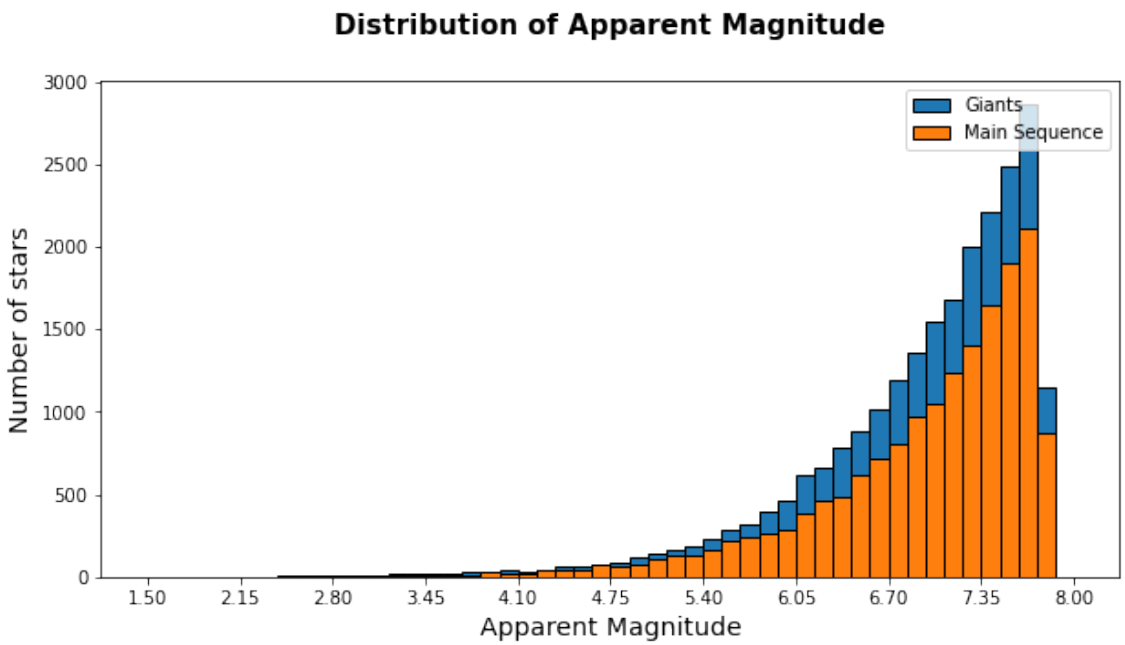


Figure 3.9: Apparent Magnitude distribution.

From above comparisons we can see that the giant branch stars and main sequence stars have almost same distribution except in the case of distances, where giant branch stars are distributed at longer distances as compared to main sequence stars on average.

3.3 Far Away Bright Sources

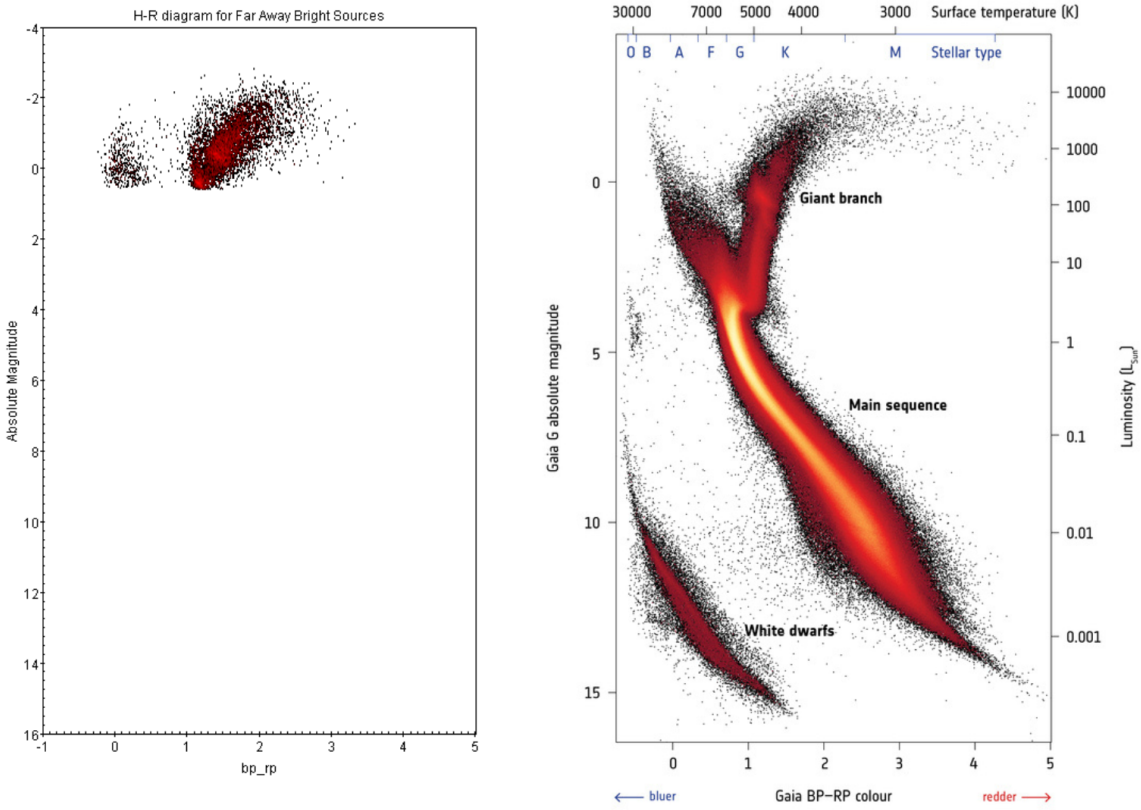


Figure 3.10: HR diagram for Far Away Bright Sources. Darker tones indicate higher density of sources. The HR diagram for the whole Gaia catalogue is given on the right for comparison.

The full HR diagram was obtained from the official site. The plot for the sample was scaled with ranges similar to the full HR plot to ease comparison, hence most of the plot is empty with stars in sample having bp-rp between -0.3 to 3.4 and absolute magnitude between -2.86 and 0.67 .

On comparing we can clearly see that most of the selected stars match the Giant branch from the full HR diagram. Some stars also correspond to hotter end of the main sequence, maybe blue giants. This correspondence was expected as only bright giants can appear bright enough to be in chosen apparent magnitude range when viewed from farther than 300 pc.

Since the pseudocolour estimates for only 176 of the 6746 selected sources were we could verify the spectral classes for the hotter main sequence stars in the sample but on comparing the two HR plots above, with the full HR one marked with the spectral classes, it seems that most of those stars are of A class with some being of B class.

We can divide the stars from giant branch from the others by taking subsets, one with the red giants for $G_{BP} - G_{RP} > 0.7$ and other for $G_{BP} - G_{RP} < 0.7$.

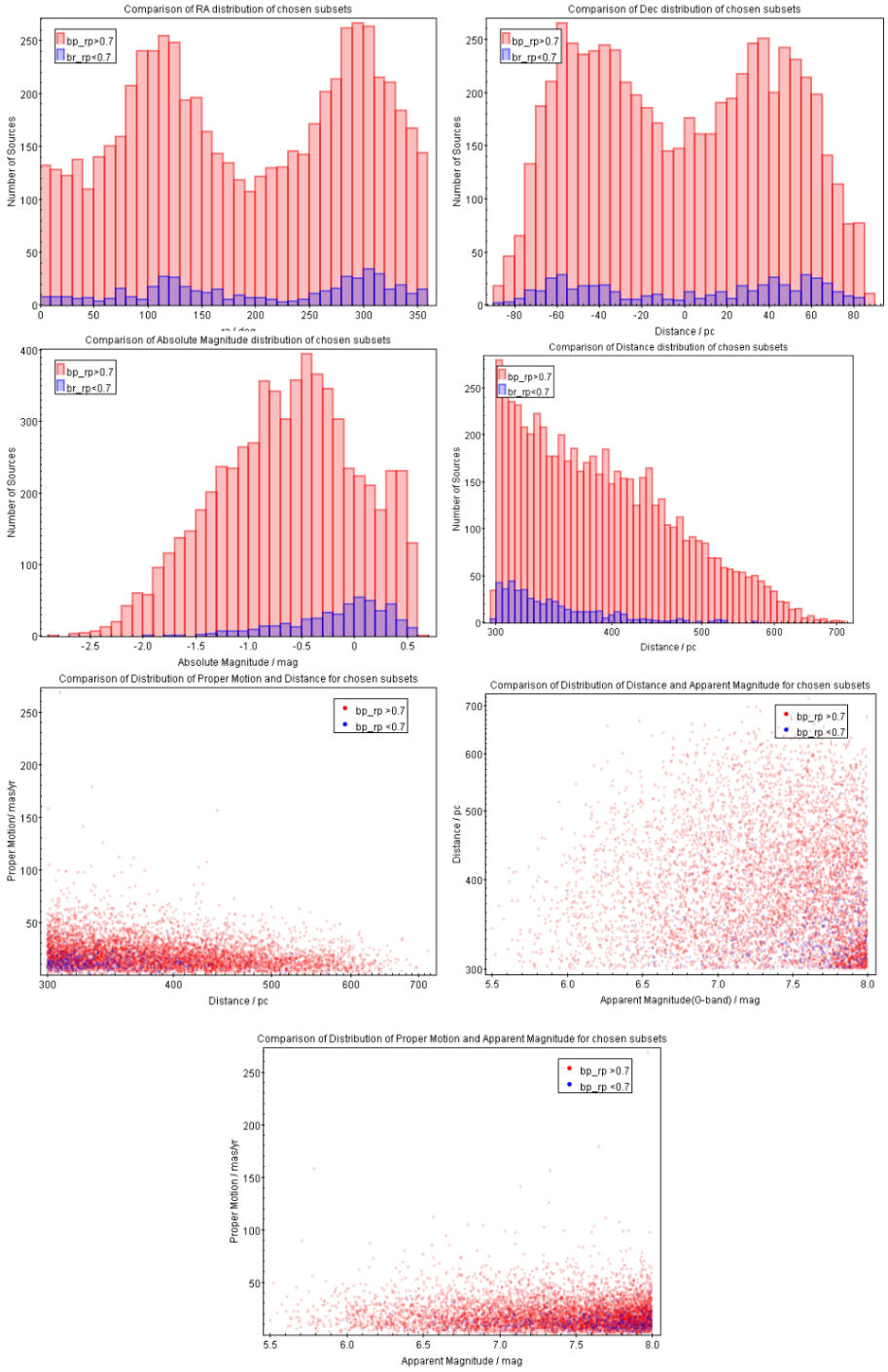


Figure 3.11: Various comparison plots for the two chosen subsets.

Due to small number of stars in sample with $G_{BP} - G_{RP} < 0.7$, no proper comparisons can be drawn.

3.4 Closest Stars

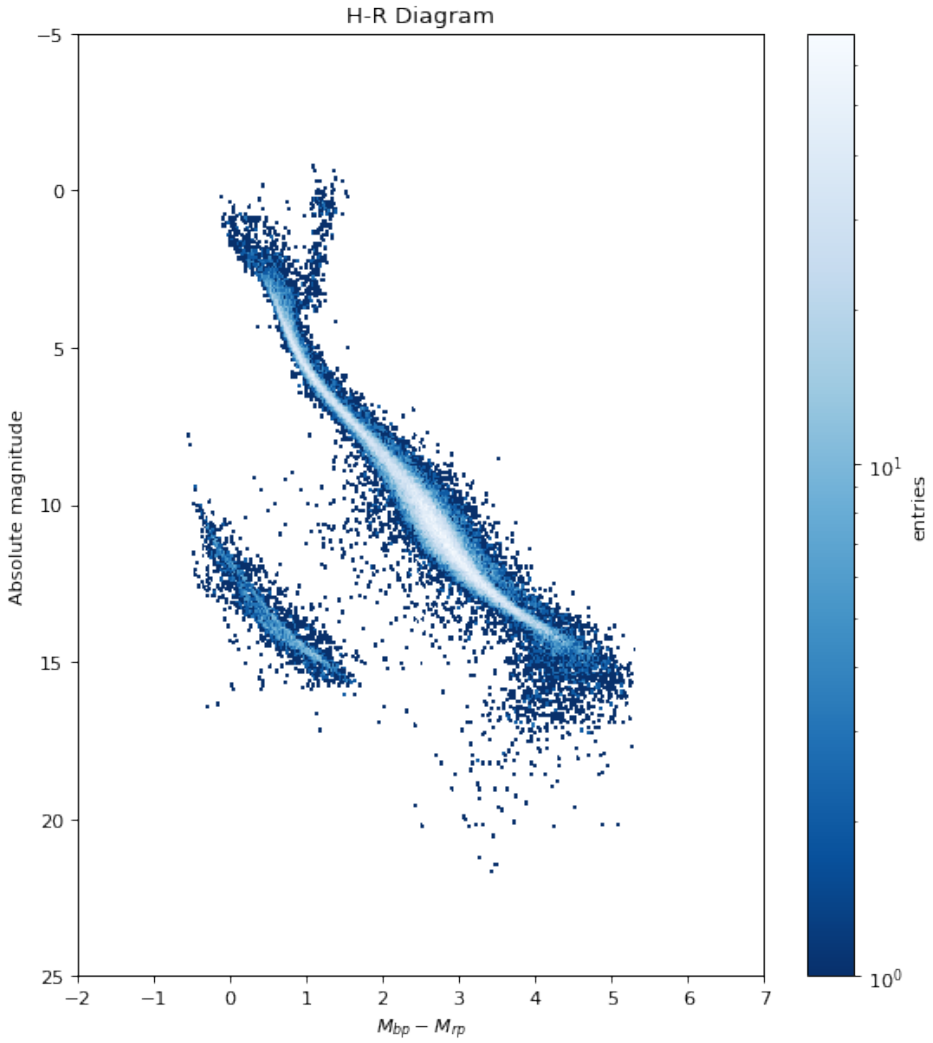


Figure 3.12: HR diagram of stars within 60 pc of the sun

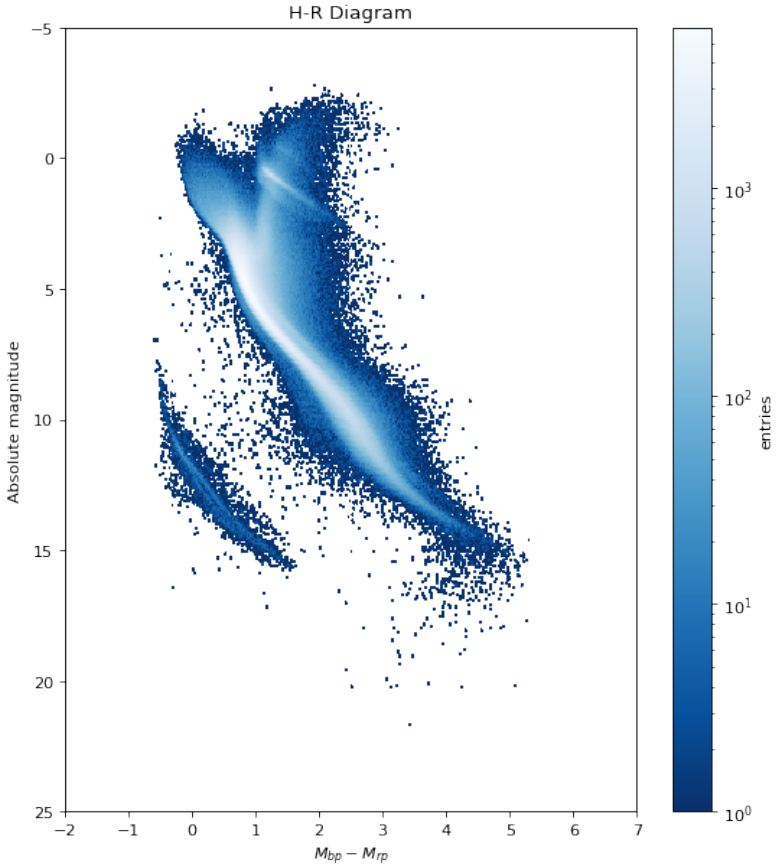


Figure 3.13: Full HR diagram

Comparing with the full HR diagram, the HR diagram of close stars has lesser proportion of high luminosity stars whereas the proportion of mid and low luminosity stars is not that stark. This points to the lack of high mass/old stars in the solar neighbourhood.

3.4.1 White dwarf stars

```
query= '''SELECT
          g.source_id, g.ra, g.dec, g.parallax , g.phot_g_mean_mag AS mag,
          g.phot_rp_mean_mag as magrp, g.phot_bp_mean_mag as magbp,
          g.parallax_error as par_err, d.r_med_geo as dspc_geo,
          d.r_med_photogeo as dspc_photogeo, g.pmra as pmra, g.pmdec as pmdec
          FROM gaia.edr3lite AS g, gedr3dist.main AS d
          WHERE g.source_id = d.source_id
          AND g.parallax_error/g.parallax < 0.01 AND g.parallax > 0
          AND g.parallax_error > 0 AND d.r_med_geo < 60
          AND ((g.phot_g_mean_mag - 5*LOG10(d.r_med_geo/10)) -
          (20/6)*((g.phot_bp_mean_mag -
          5*LOG10(d.r_med_geo/10)) - (g.phot_rp_mean_mag -
          5*LOG10(d.r_med_geo/10))) -7) > 0
          AND g.phot_g_mean_mag IS NOT NULL AND g.phot_rp_mean_mag IS NOT NULL
          AND g.phot_bp_mean_mag IS NOT NULL'''
```

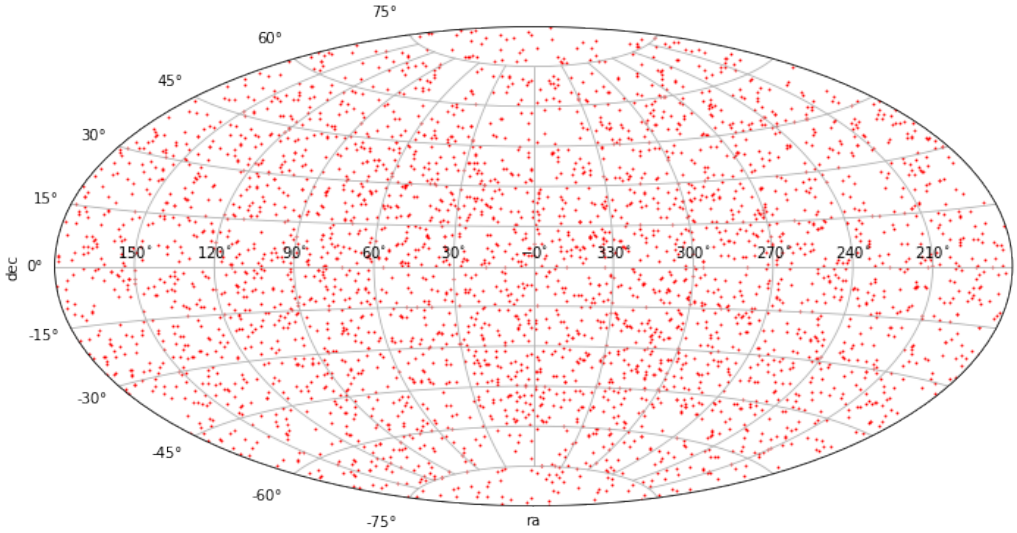


Figure 3.14: Sky map

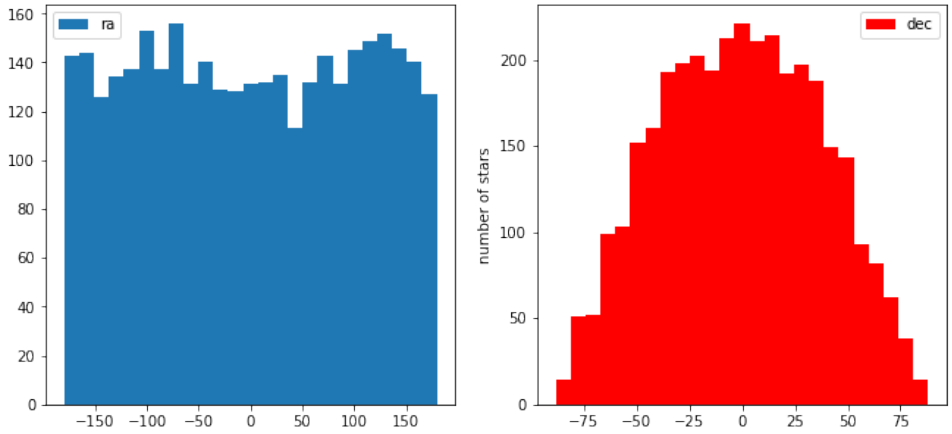


Figure 3.15: ra & dec distribution

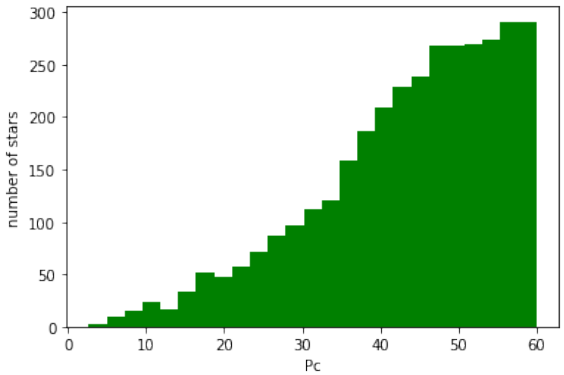


Figure 3.16: distance(pc) distribution

White dwarf stars are very evenly distributed, but after 45 pc, ther population growth decreases, due to the 1 percent parallax error condition.

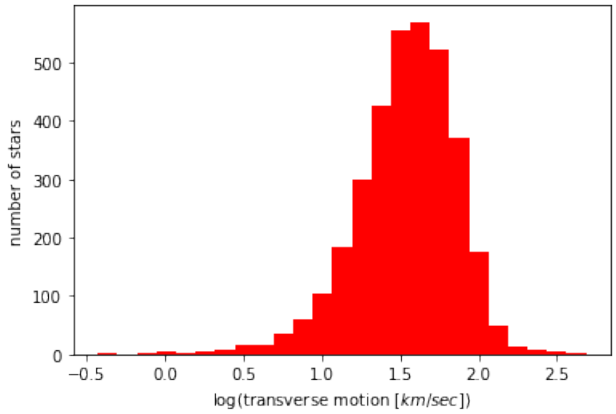


Figure 3.17: Transverse motion distribution (base of log = 10)

Again, most of the white dwarfs have transverse velocity between 10 to 100 km/sec just like main sequence stars.

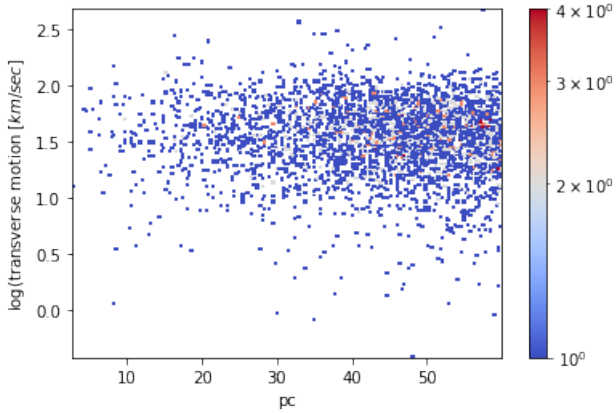


Figure 3.18: Transverse velocity vs distance (base of log = 10)

Here again, white dwarfs show same distribution as main sequence stars.

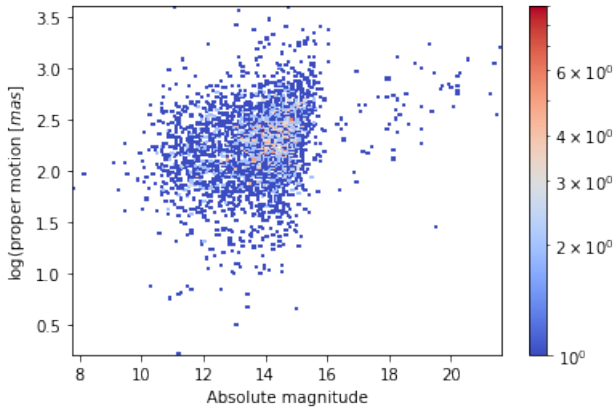


Figure 3.19: proper motion vs absolute magnitude (base of log = 10)

There is an abrupt end in the density of star population at about magnitude

of 15 which marks the end of white dwarfs, after which there are very faint red dwarfs.

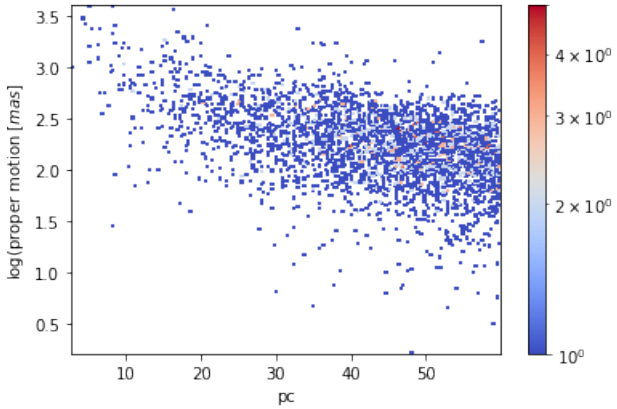


Figure 3.20: proper motion vs distance (base of log = 10)

3.4.2 Discussion

The motivation behind this exercise was to detect any difference in the distribution and transverse velocity of main sequence and white dwarf stars. This would allow us to determine if there was any condition which favoured either of the two.

3.4.3 Conclusion

The analysis of above results points out to almost no difference in the astrometric distribution of the main sequence and white dwarf stars.

3.5 Large Magellanic Cloud

In the apparent magnitude and distance plot, we saw a greater star density towards the top. Since the LMC is a large distance away, to appear in Gaia's limiting magnitude of 21, the stars' intrinsic brightness would have to be large, which agrees with the diagram. The HR diagram lacks the tail and the bend, thereby confirming the presence of stars of high luminosity. The patch on the left points to the massive high temperature, high luminosity blue main-sequence stars and the right one corresponds to the red giants and supergiants, again being one of the key features of the LMC.

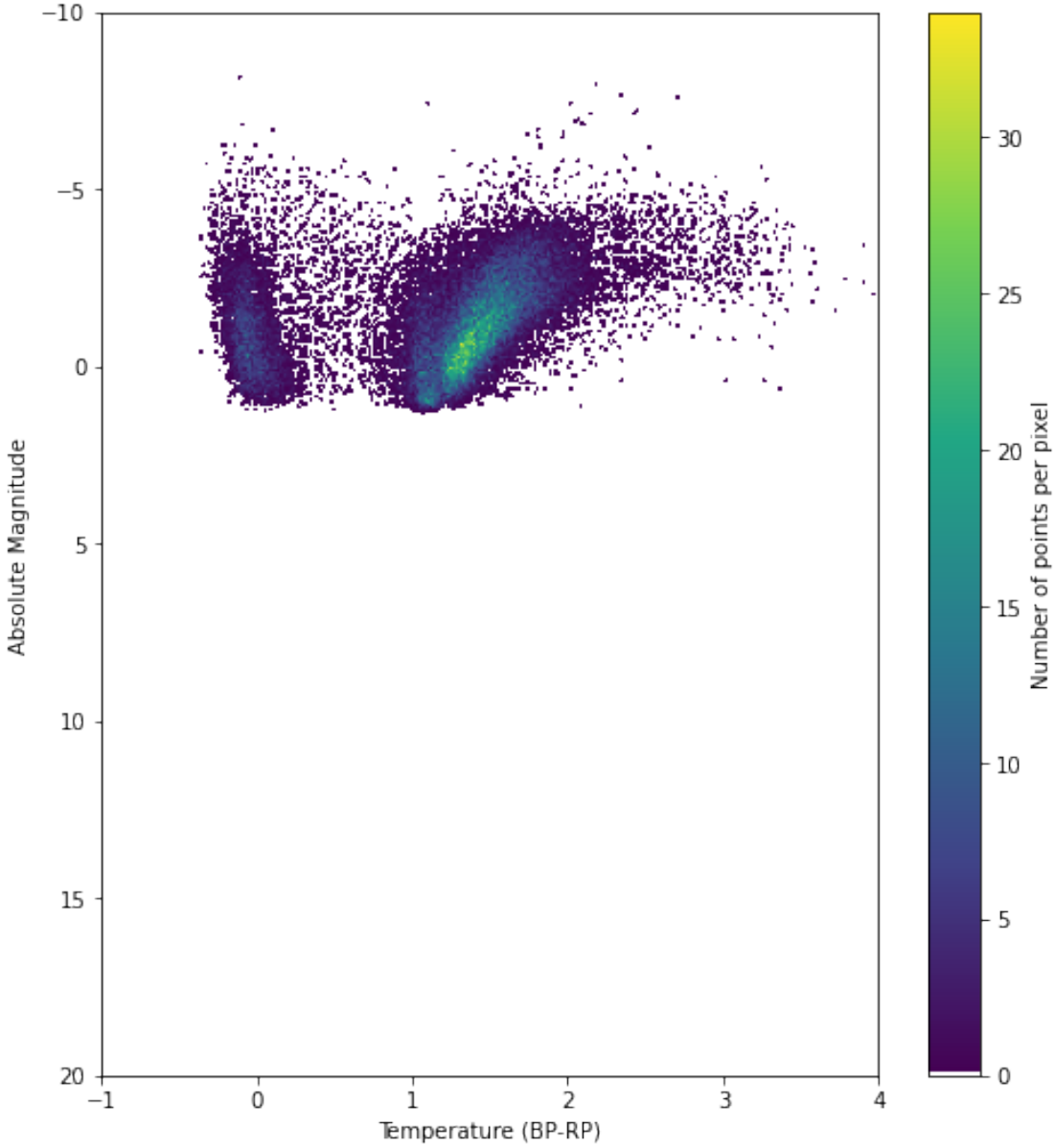


Figure 3.21: HR diagram of LMC stars

4. Star Cluster Analysis

4.1 Data Querying, Distances and HR Diagrams

Method - First we find the coordinates of the star cluster using available data on web. Then we extract all stars from Gaia whose parallax has less than 20% error using an ADQL query.

Then, to find the distance of star cluster from Earth, we plot apparent magnitude vs distance calculated as (distance in parsecs) = 1000/(parallax in milliarcseconds). The distance with greatest density of points tells us the distance of the cluster from Earth.

We next use a mask on the extracted Gaia data, which filters out only the stars that lie in the parallax range of the star cluster. For these stars we then plot the graph of absolute magnitude v/s colour $G_{BP} - G_{RP}$.

4.1.1 King Cobra Cluster (Messier 67)

We find the coordinates of M67 using previous data, then use it in ADQL query to extract them

```
query = '''select top 10000 g.source_id, g.ra, g.dec, g.parallax,
g.phot_g_mean_mag, g.phot_bp_mean_mag, phot_rp_mean_mag,
h.classprob_dsc_combmod_star + h.classprob_dsc_combmod_whitedwarf as comb
from gaiadr3.gaia_source as g, gaiadr3.astrophysical_parameters as h
where g.source_id = h.source_id and
g.ra between 132.3 and 133.3 and
g.dec between 11.3 and 12.3
and g.parallax_over_error > 5
order by g.parallax desc, comb desc'''
```

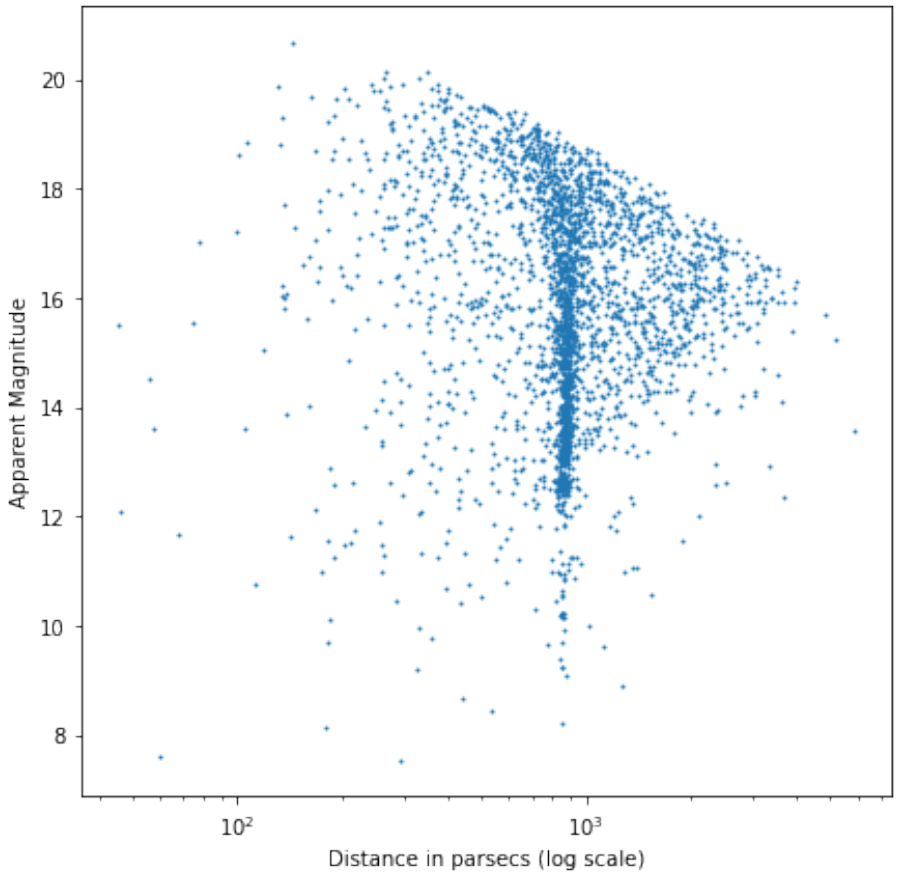


Figure 4.1: Plot of distance and apparent magnitude. The distance distribution is densest at just less than 1000 parsecs

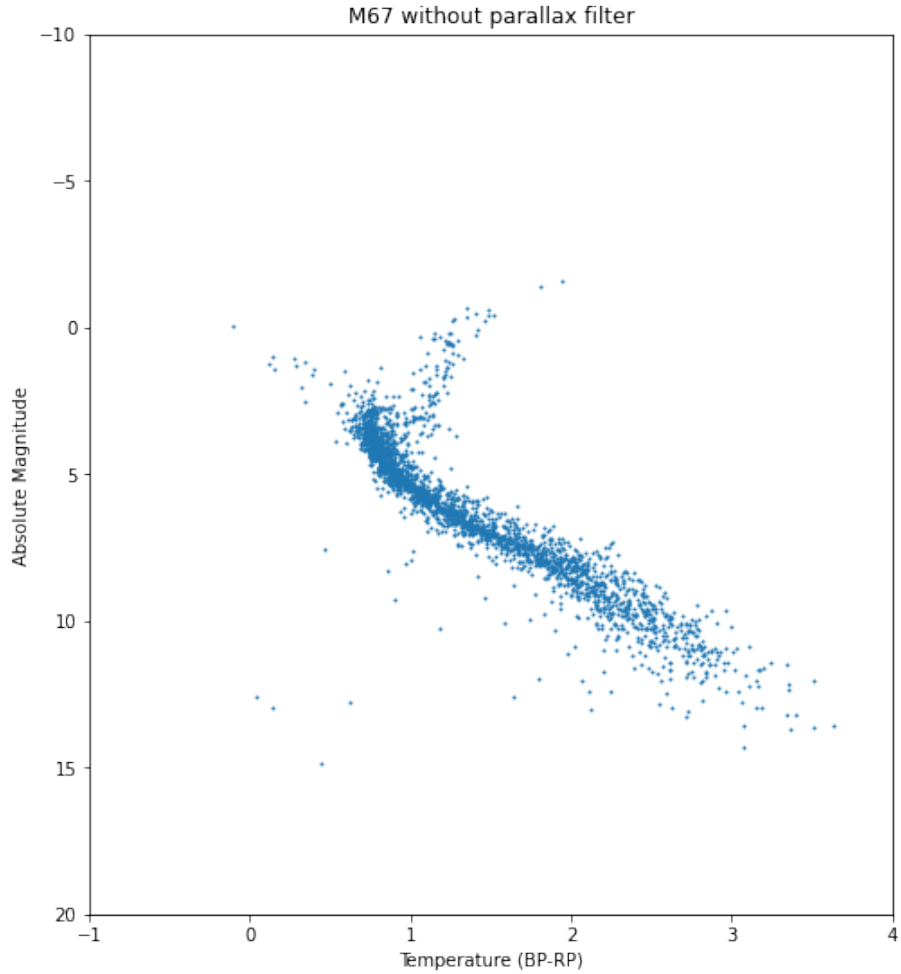


Figure 4.2: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch

.

According to the distance plot, we now choose our filter so that only stars with distance between 750 and 1000 pc, equivalent to parallax in mas > 1 snd < 1.33, are displayed.

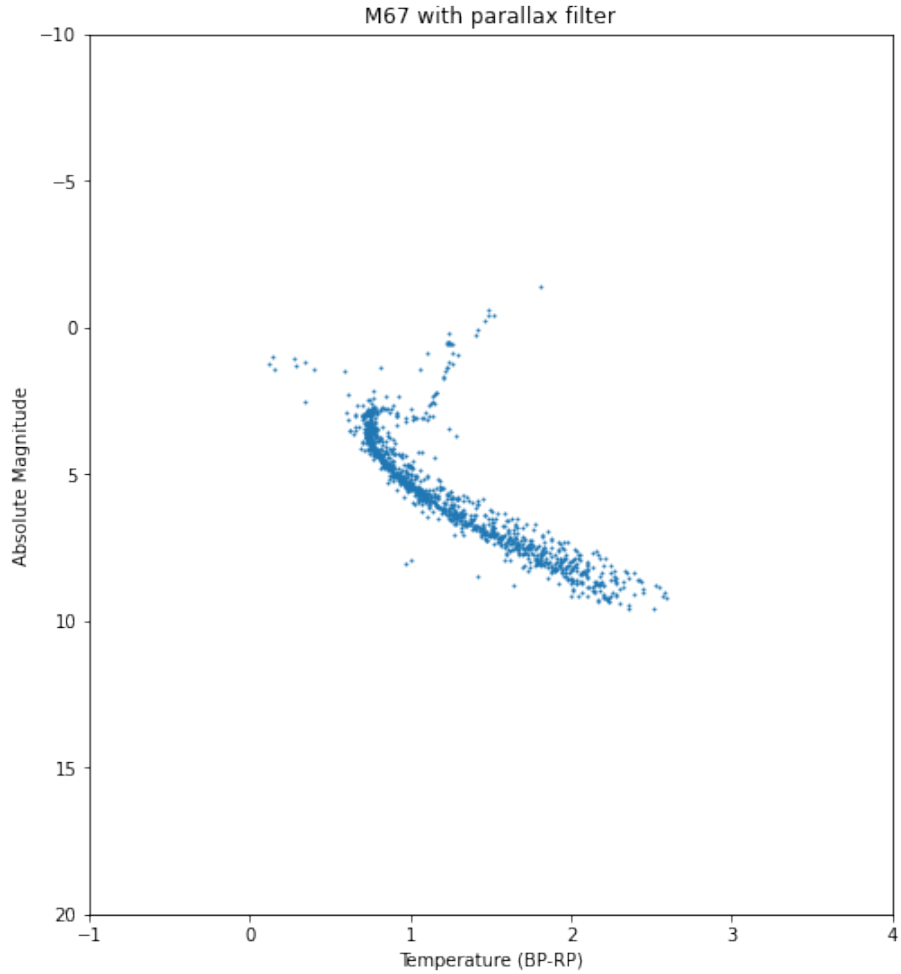


Figure 4.3: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance between 750 and 1000 pc

.

Here we get to see what happens to be the typical curve for an open star cluster:

- The main diagonally-running line are the main sequence stars, which are high in number because open star clusters are typically young.
- There is a distinct fish hook-shaped curve at the top where the stars slide into the red giant phase. Stars above that hook correspond to those stars which have moved into the red giant phase, typically low in number than the main sequence stars. This fish hook part is still noticeable in the case of M67. It is known as the **turnoff point**.
- Apart from the main sequence turning toward the red giant region, there's another fainter branch that continues on the main sequence diagonal line. This corresponds to the stars with very high temperature and luminosity, again low in number.

4.1.2 Beehive Cluster (Praesepe)

We again follow the method discussed in the start of the section, the ADQL query being exactly same except for the change in RA and Dec.

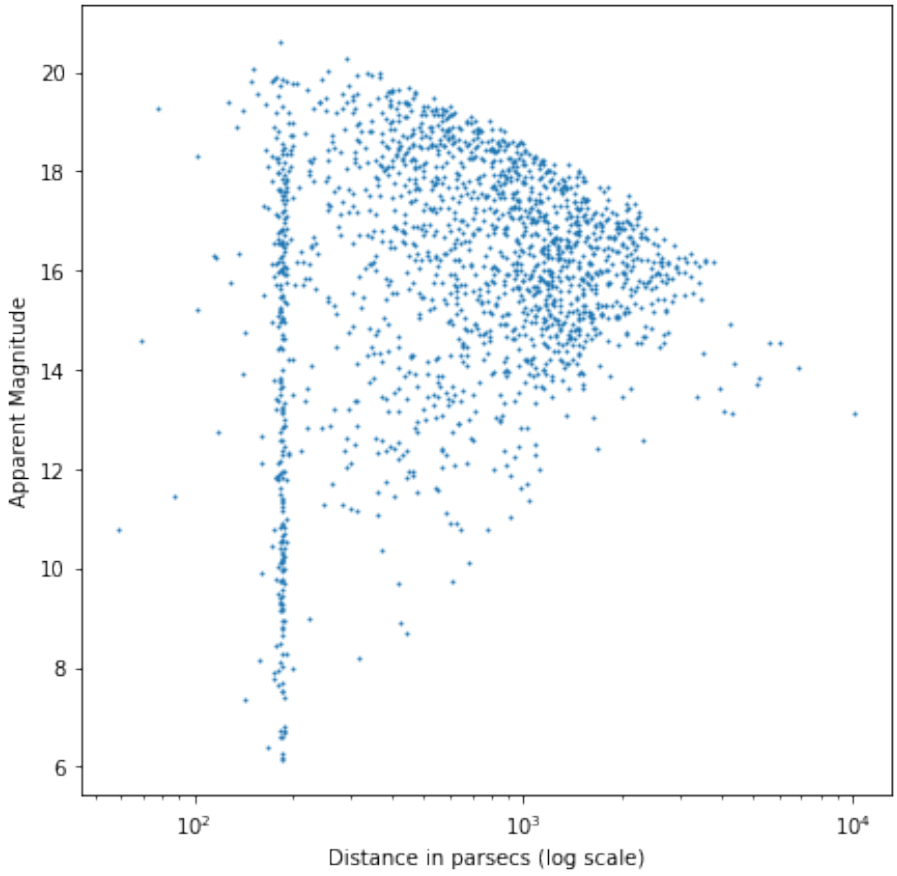


Figure 4.4: Plot of distance and apparent magnitude. The distance distribution is densest at just less than 200 parsecs

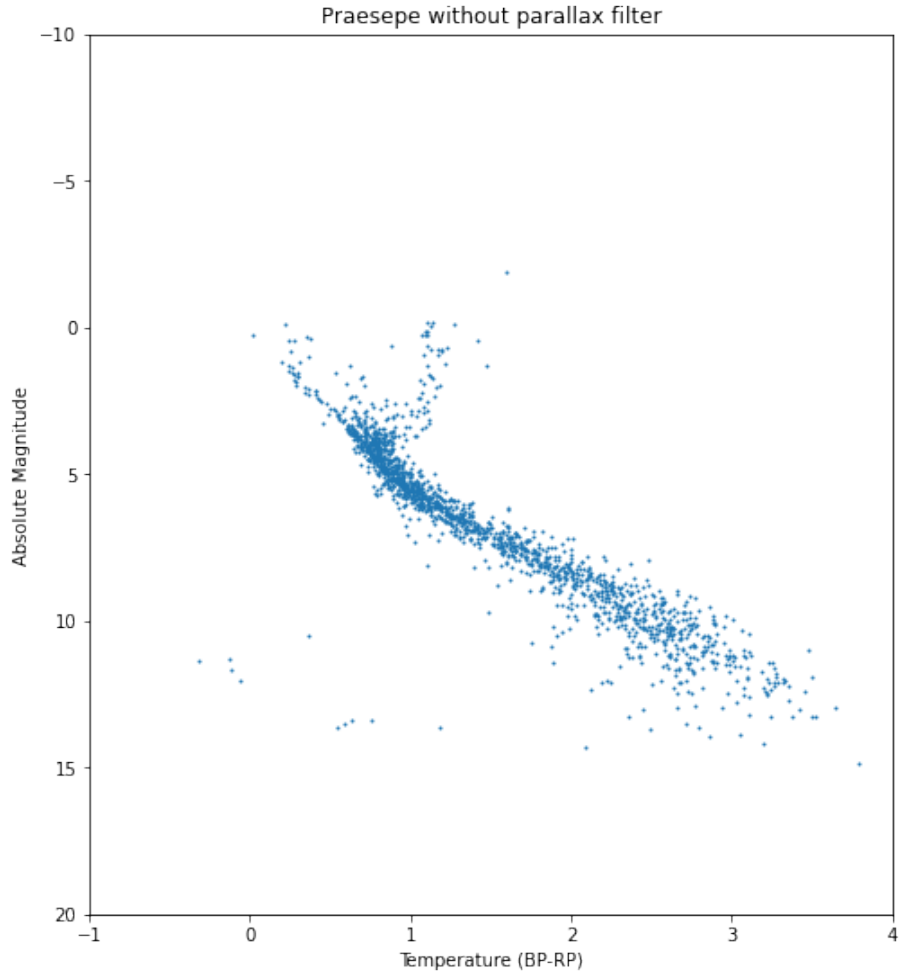


Figure 4.5: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch

.

We again choose our filter so that only stars with distance < 1000 pc, equivalent to parallax in mas > 1, are displayed.

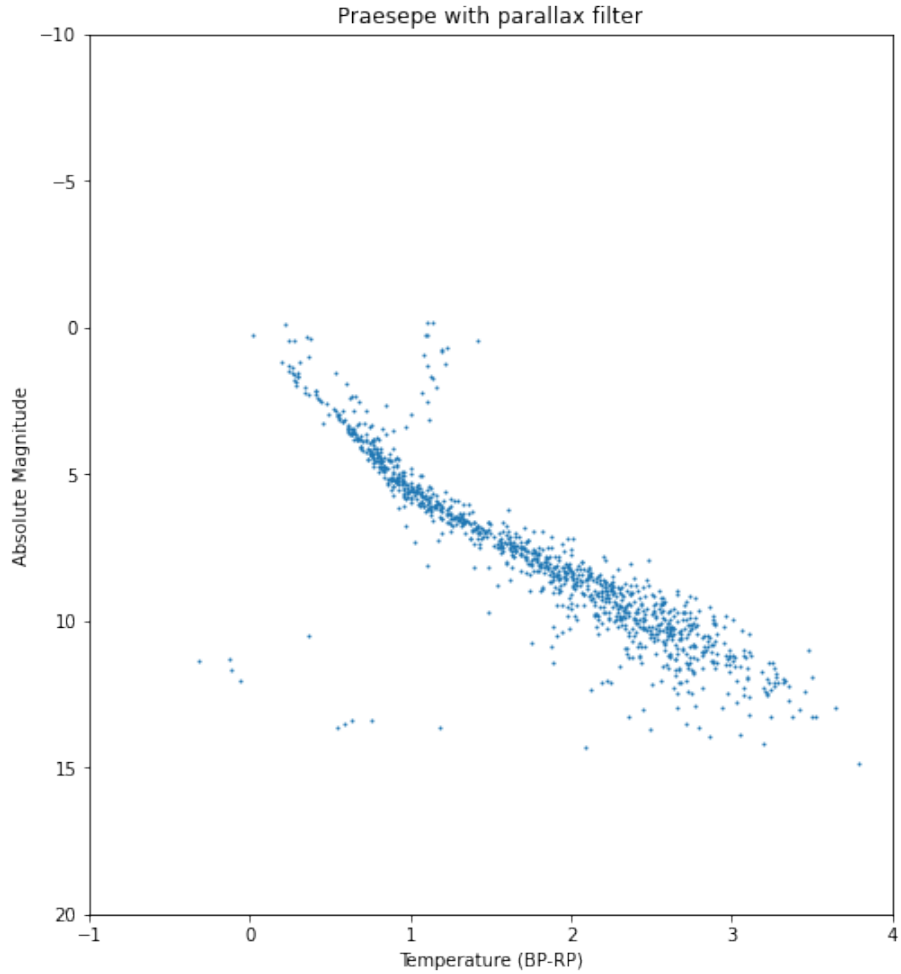


Figure 4.6: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance $<$ 1000 pc

.

A noticeable difference of this plot with the earlier plot of M67 stars is that there are fewer stars in the red giant phase, and significantly more stars in the high temperature, high luminosity region.

Another difference is the absence of cooler stars in the lower right main-sequence in M67, which are present here in the case of Beehive cluster.

Since a red giant is typically a late phase of stellar evolution, this tells us that M67 has greater population of stars in their later part of life - which means the King Cobra cluster is older than Beehive cluster.

4.1.3 Pi Puppis cluster (Collinder 135)

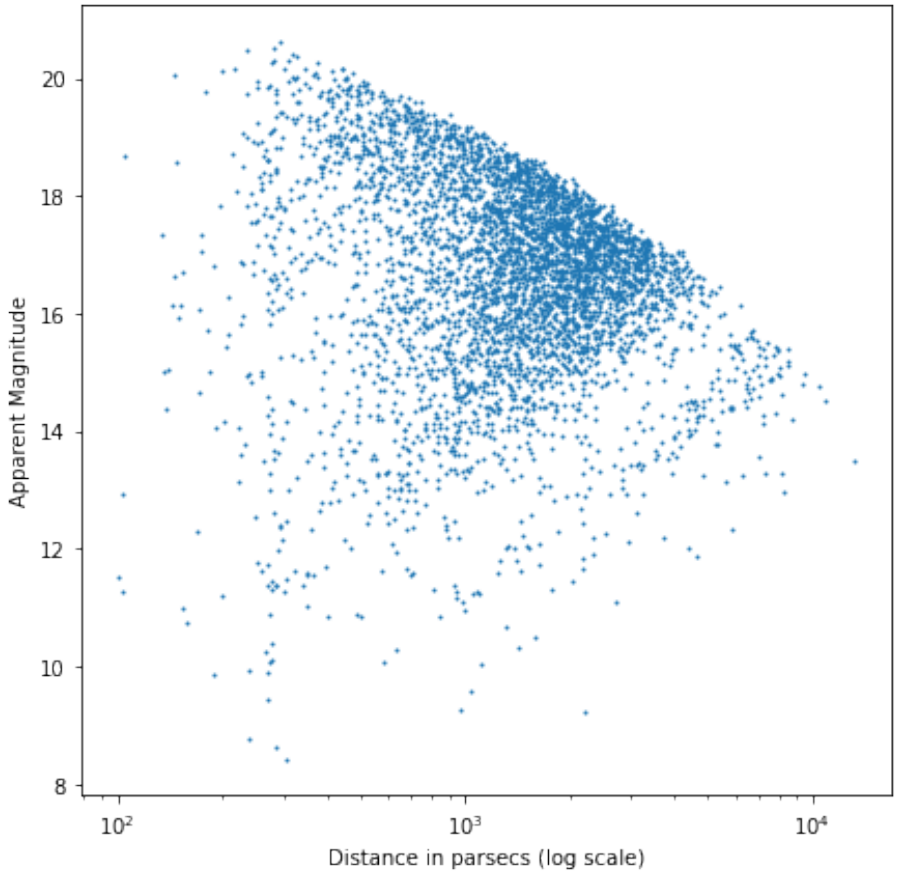


Figure 4.7: Plot of distance and apparent magnitude

.

Here we notice something different - since the star cluster is relatively faint (which is one of the reasons the name would seem unfamiliar to many), the data of the cluster's stars is mixed with data of other stars in the line of sight as well. There is a significant star density spread out over the top part of the entire graph. So in order to get a good estimate of the distance, we do the following:

1. Set the error limit of parallax to 12.5%
2. Make a histogram of distance distribution
3. Look for clearly discernible peaks in distance distribution using `scipy.interpolate` functions.

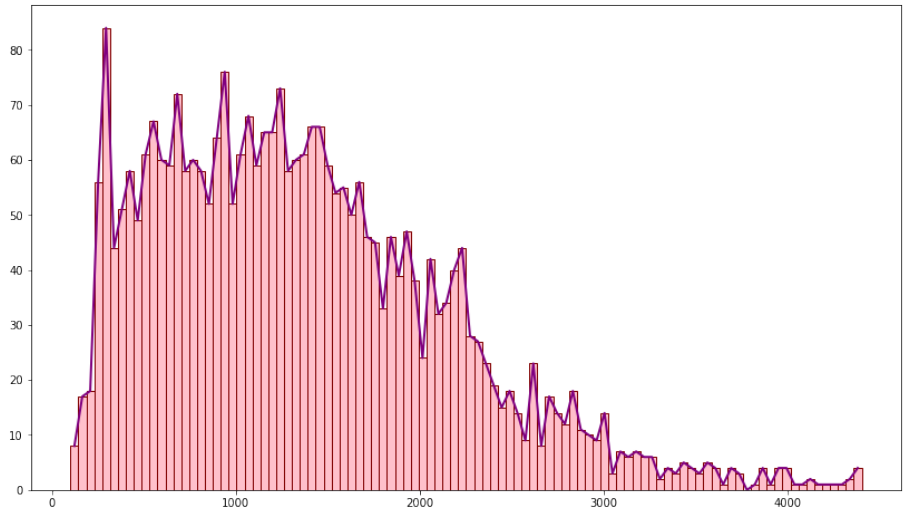


Figure 4.8: Histogram of distance distribution - distance is in parsecs

.

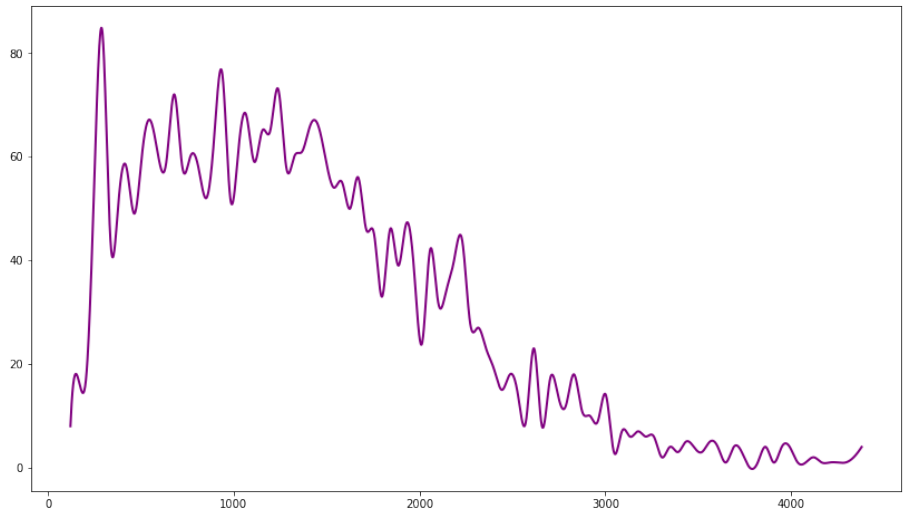


Figure 4.9: Smooth curve of distance (in pc) distribution

.

A clear peak is observable at a distance just less than 500 parsecs, after which there is a random distribution of crests and troughs until it drops at 3000-4000 pc. We accordingly choose our distance filter as < 500 pc

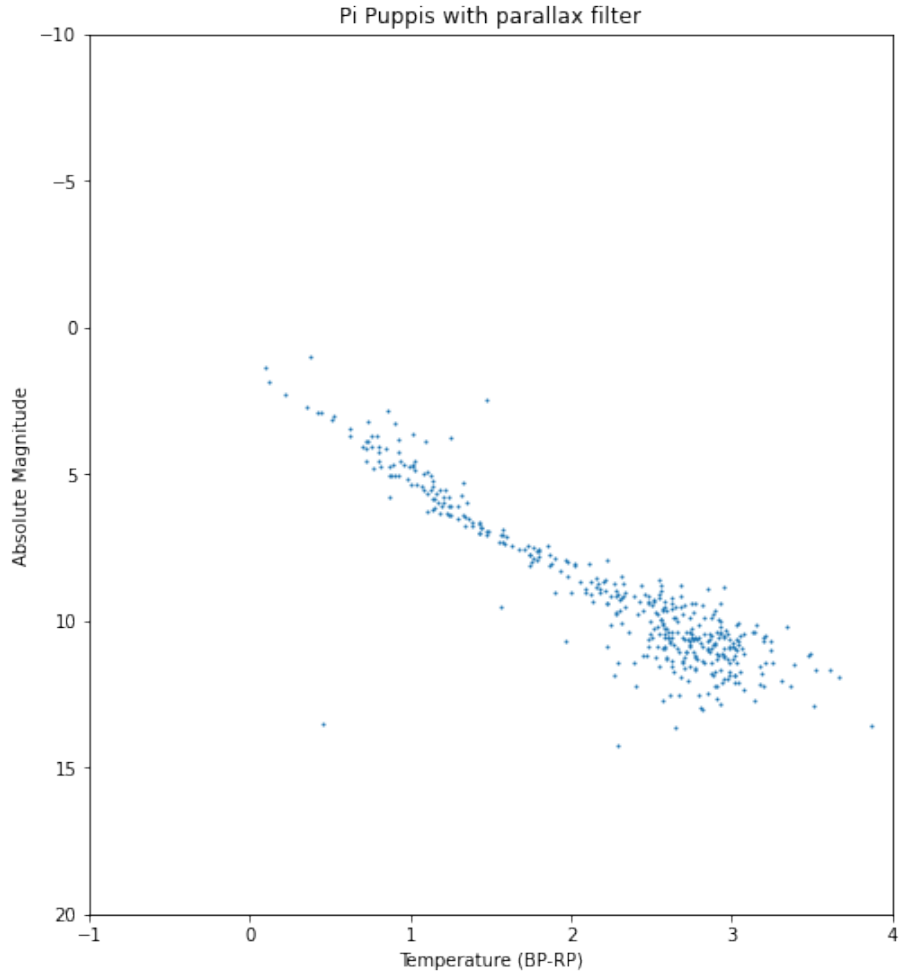


Figure 4.10: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance $<$ 500 pc

.

We have a significantly lower number of total stars in the sample, but even then it is clearly seen that most stars are in the main sequence, with barely any stars in the red giant phase. Also, a lot of stars are in the lower right part of the plot, with low temperature and low luminosity. This tells us that

- The cluster is quite young, as there are no stars in their late red giant phase
- Most stars are of low luminosity and temperature, which correspond to orange dwarfs or red dwarfs of low apparent magnitude. This also explains why only relatively fewer stars were detected in the first graph and why there was no clear distance density.

4.1.4 Butterfly Cluster (Messier 6)

In this case setting parallax error limit as $< 20\%$ led to some noisy output remaining still, so I imposed a tighter condition of error limit $< 10\%$.

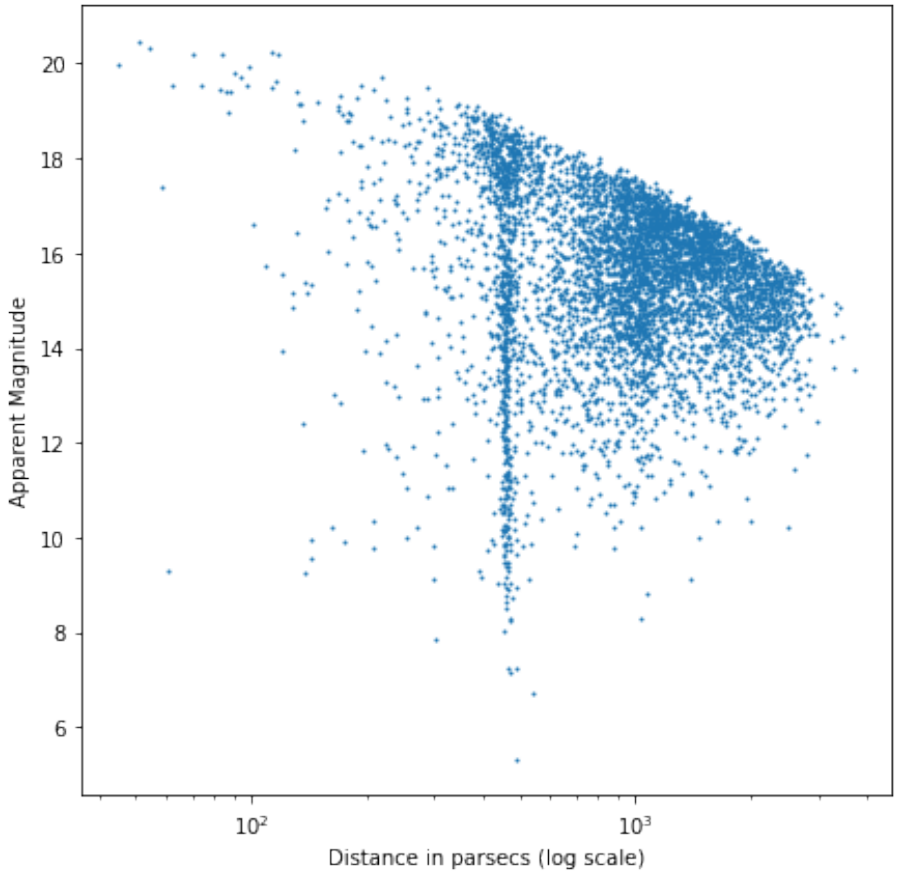


Figure 4.11: Plot of distance and apparent magnitude

.

Here we notice a clear peak at between 400 and 600 pc, but at the same time there is some density of stars of low brightness/high apparent magnitude (between 15 and 20) and at distances beyond 1000 pc, which do not belong to the Butterfly cluster.

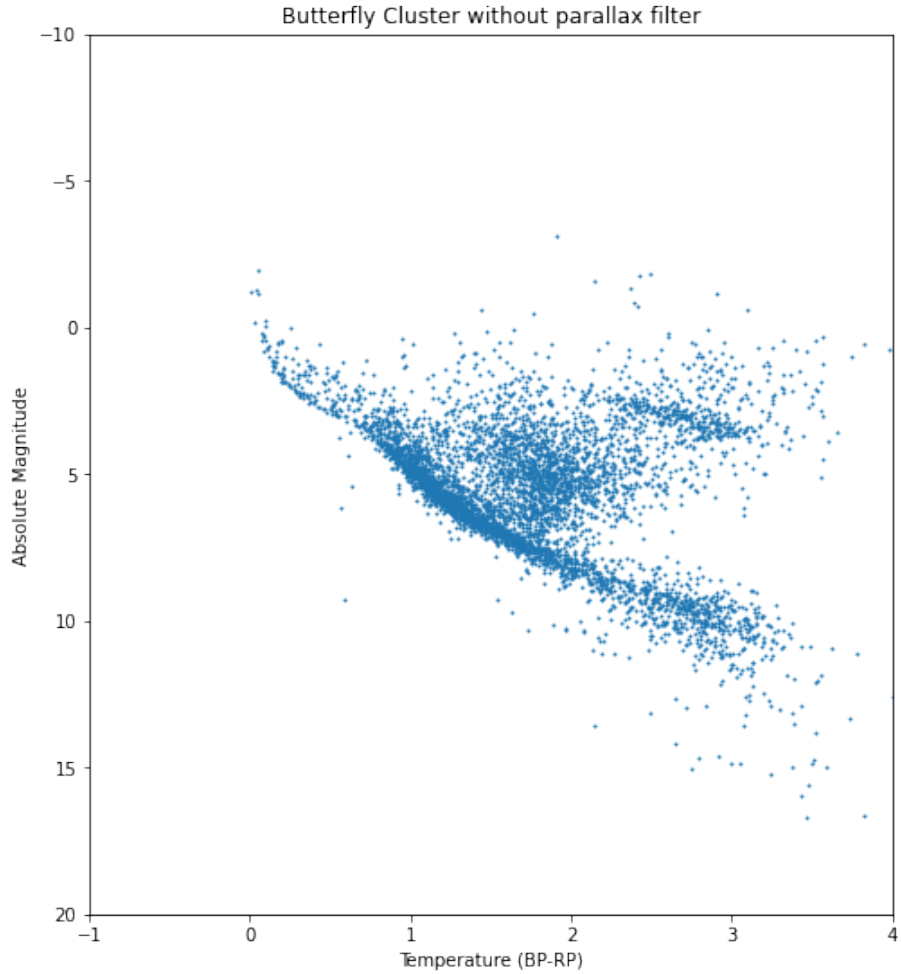


Figure 4.12: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch

.

Here we again have something interesting - instead of just the main sequence line and a paler trail going into the red giant region, we have the main sequence line, and a *significant* amount of stars in the red giant and supergiant regions too. To check if these belong to the cluster or not, we need to filter by distance.

Again, our filter is to only show stars with distance < 600 and > 400 pc, equivalent to parallax in mas > 1.67 and < 2.5, are displayed.

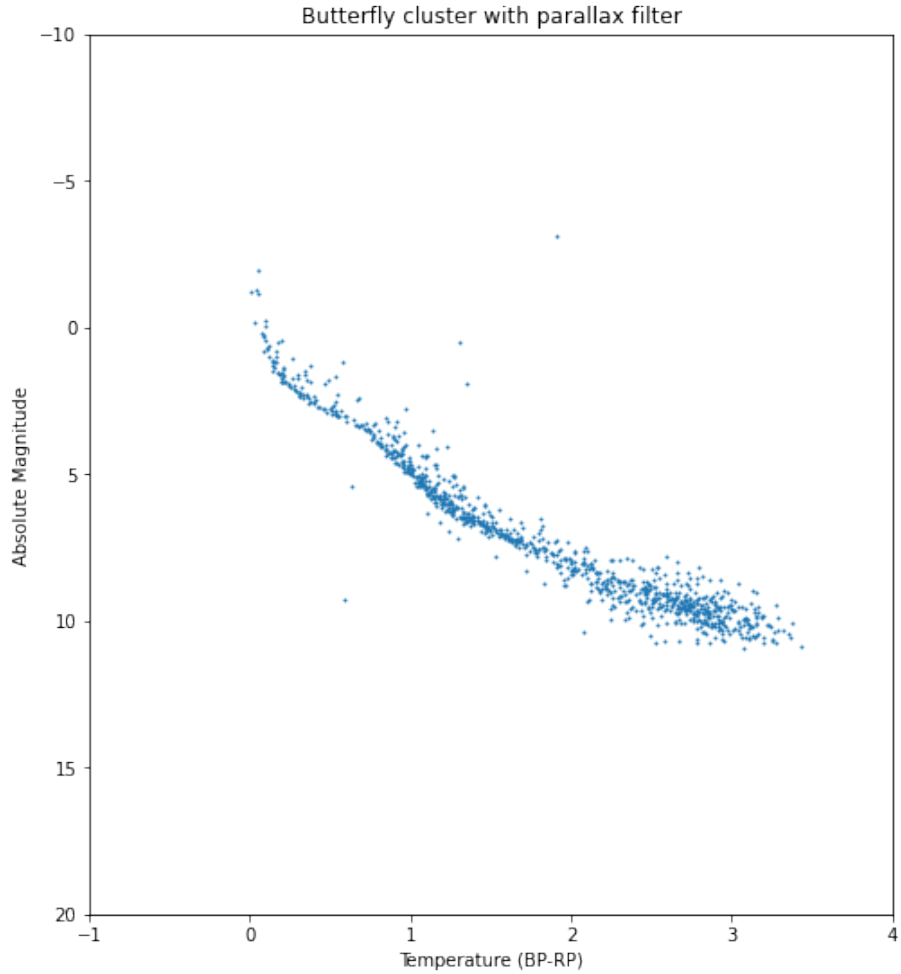


Figure 4.13: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance between 400 and 600 pc

.

With the distance filter, the red giant and supergiant patches have disappeared - which means they aren't a part of the Butterfly cluster. If we revisit the plot of distance and apparent magnitude, and compare it with the HR diagram plot, we find that

- The density of stars along the vertical line at distance somewhere less than 1000 pc corresponds to the star cluster M6 or Butterfly cluster. These stars are predominantly main sequence stars with a few in the red giant phase.
- There is another density of low brightness (apparent magnitude 15-20) stars at distances larger than 1000 pc - this hence corresponds to the red giant and supergiant patches in the first (without parallax filter) HR diagram .

4.1.5 Pleiades

Here we needed to refine our search - Pleiades has a larger angular window of about 2° while all others till now had angular diameter of less than 1°. We first impose the parallax error to be $< 10\%$. There is still a lot of background star density as in the figure below

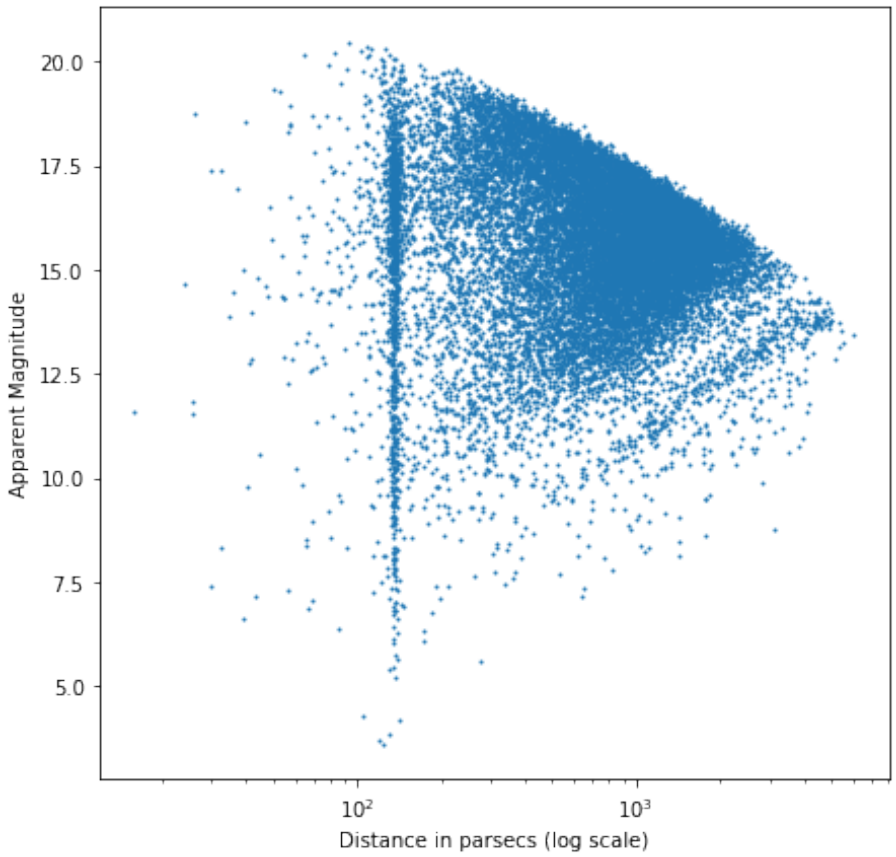


Figure 4.14: Plot of distance and apparent magnitude

.

Again, there is a high stellar density at the top of the plot, but unlike Pi Puppis cluster, we do have a clearly discernible dense vertical line somewhere between 100 and 200 pc.

Plotting the HR diagram for the full data gives

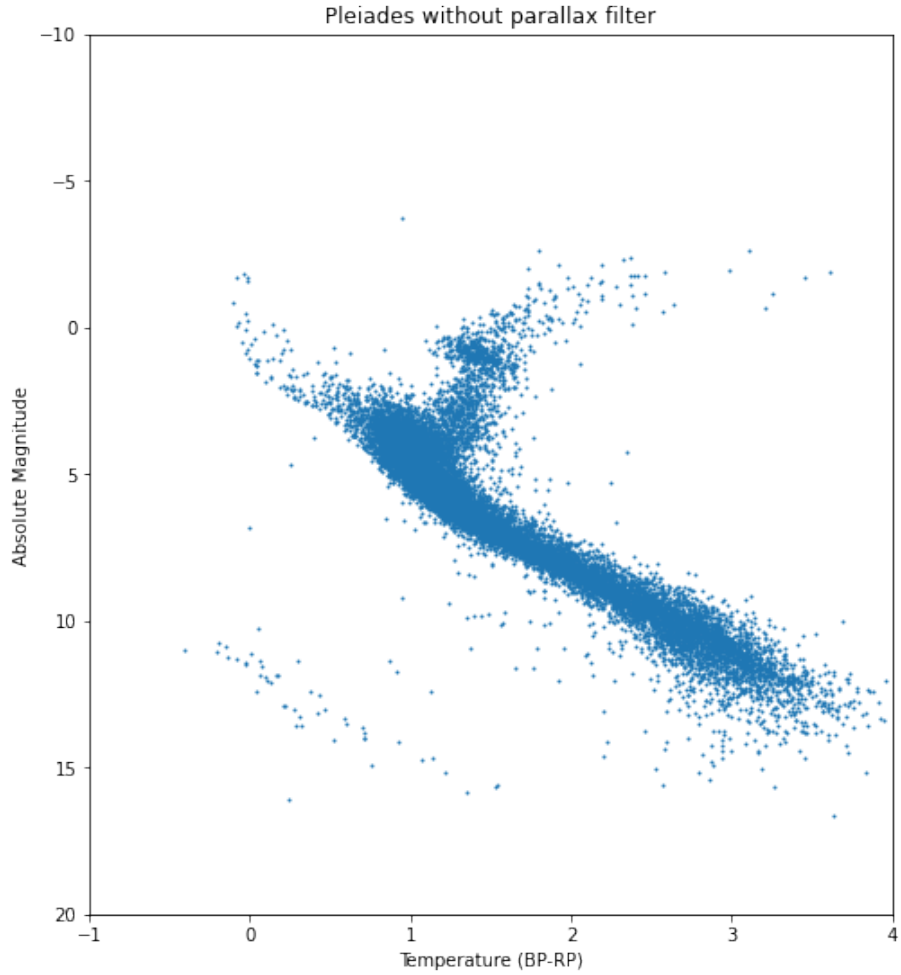


Figure 4.15: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch

.

This is the closest we have got so far to the actual HR diagram for a random sample of stars - the diagonal line is the main sequence, the branching-off on top right corresponds to the red giants and supergiants.

Additionally, we notice a sprinkling of stars in the bottom left i.e. of low luminosity but high temperature. These correspond to white dwarfs.

Applying the filter of distance < 200 pc, we get the HR diagram for Pleiades

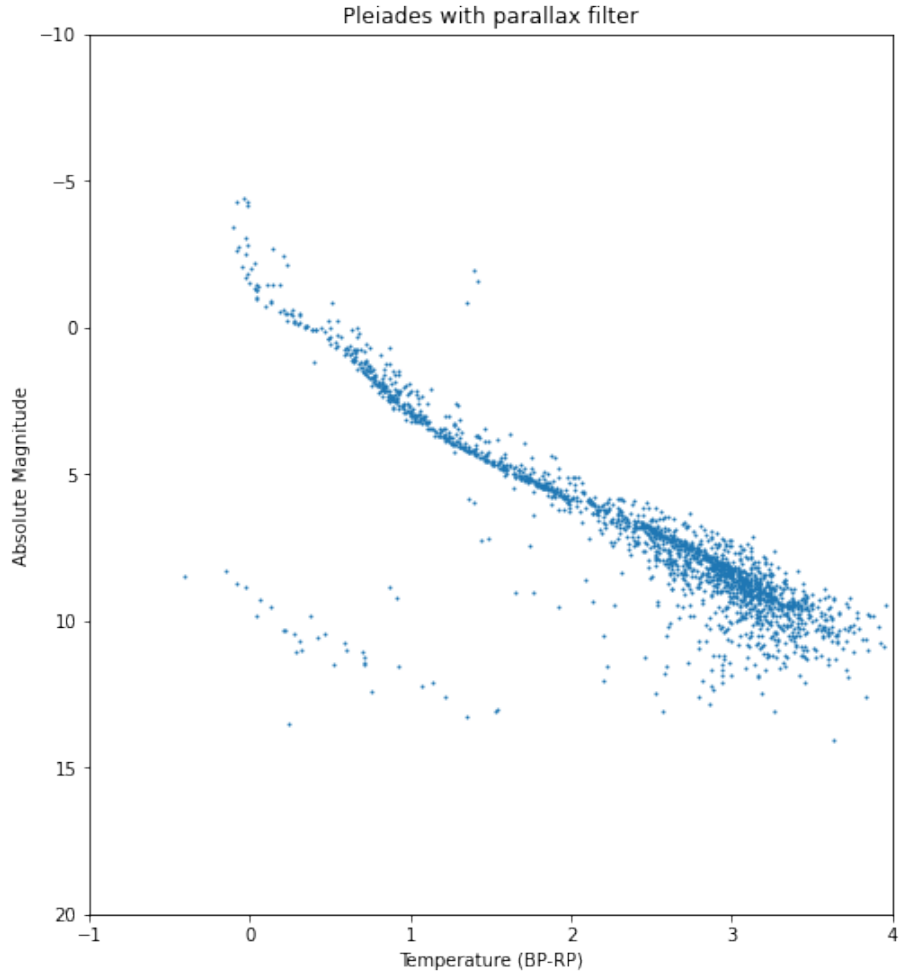


Figure 4.16: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance < 200 pc

.

Like the Pi Puppis cluster, we again have almost no stars in the red giant phase. But there are also a few differences. We can infer the following

- Absence of red giants shows higher population of younger stars, i.e. the cluster is young.
- There is a significant population in the high luminosity and temperature region as well. This agrees well with the fact that the seven brightest stars of the cluster are all blue.
- The sprinkling of white dwarfs still remains, which means these are part of Pleiades cluster.
- The red giant branch from the previous HR diagram has vanished, and the main sequence diagonal line has noticeably thinned . This means that the density of stars at distances around 1000 pc (which were removed by the 200 pc filter) were nearly all red giants or main sequence stars.

4.1.6 NGC 2264 (Christmas Tree Cluster and Cone Nebula)

This is a cluster of stars in the constellation of Monoceros. Its brightest member is an O-type blue giant; however Gaia does not capture very bright stars (whose accurate astrometric data is already available in previous catalogs) so this star was missed. Nonetheless, it was still chosen to study a star-forming nebula in more detail.

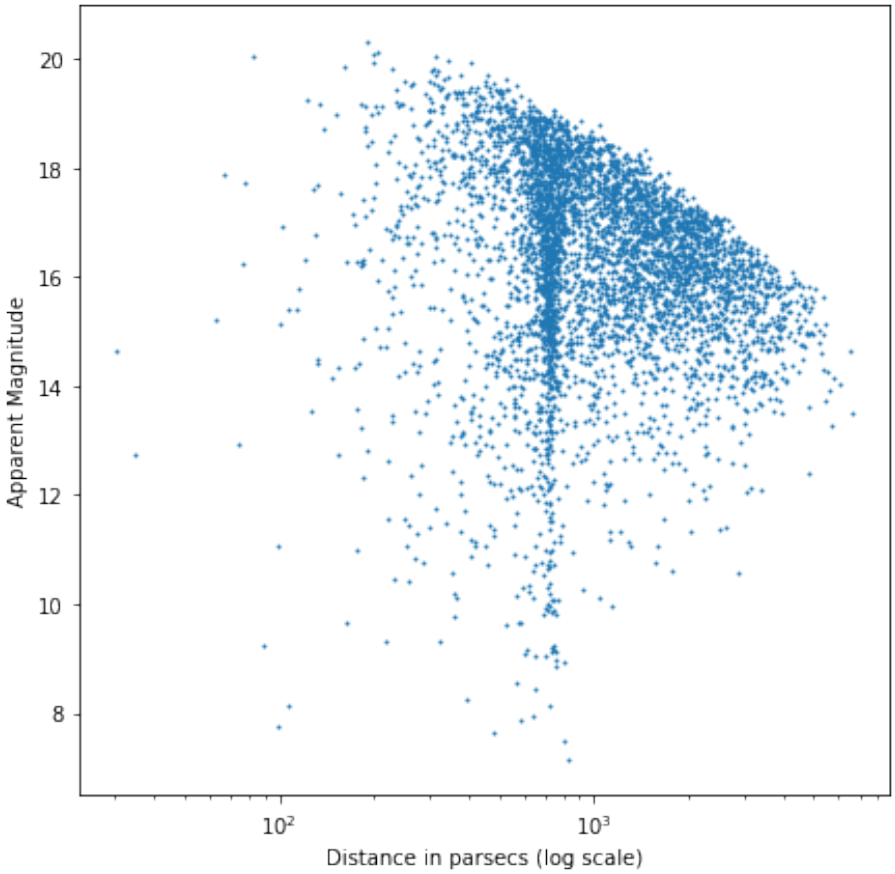


Figure 4.17: Plot of distance and apparent magnitude

.

In this case there is a clear stellar density at between 700 and 800 pc, so we set the parallax mask accordingly of parallax > 1.2 mas and < 1.5 mas.

HR diagram for the full data gives

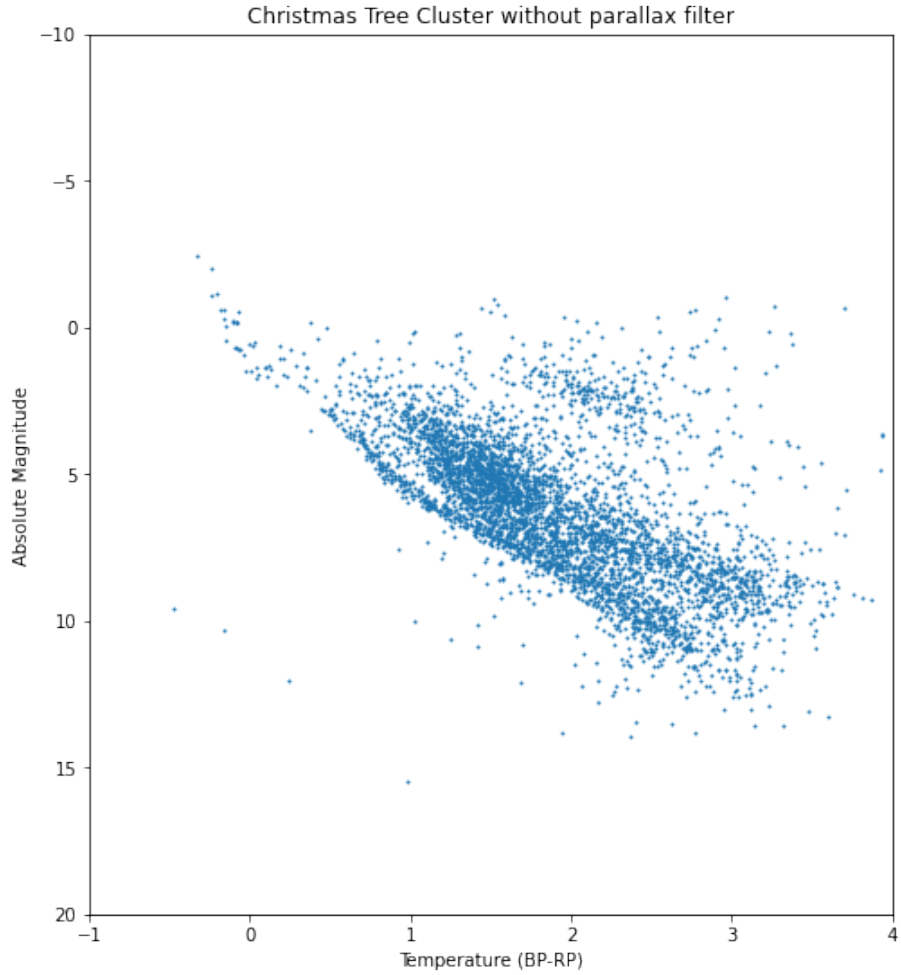


Figure 4.18: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch

.

Now we notice a few white dwarfs, more population along the main sequence and a merging of main sequence with the red giants branch.

Applying the distance/parallax filter and plotting gives

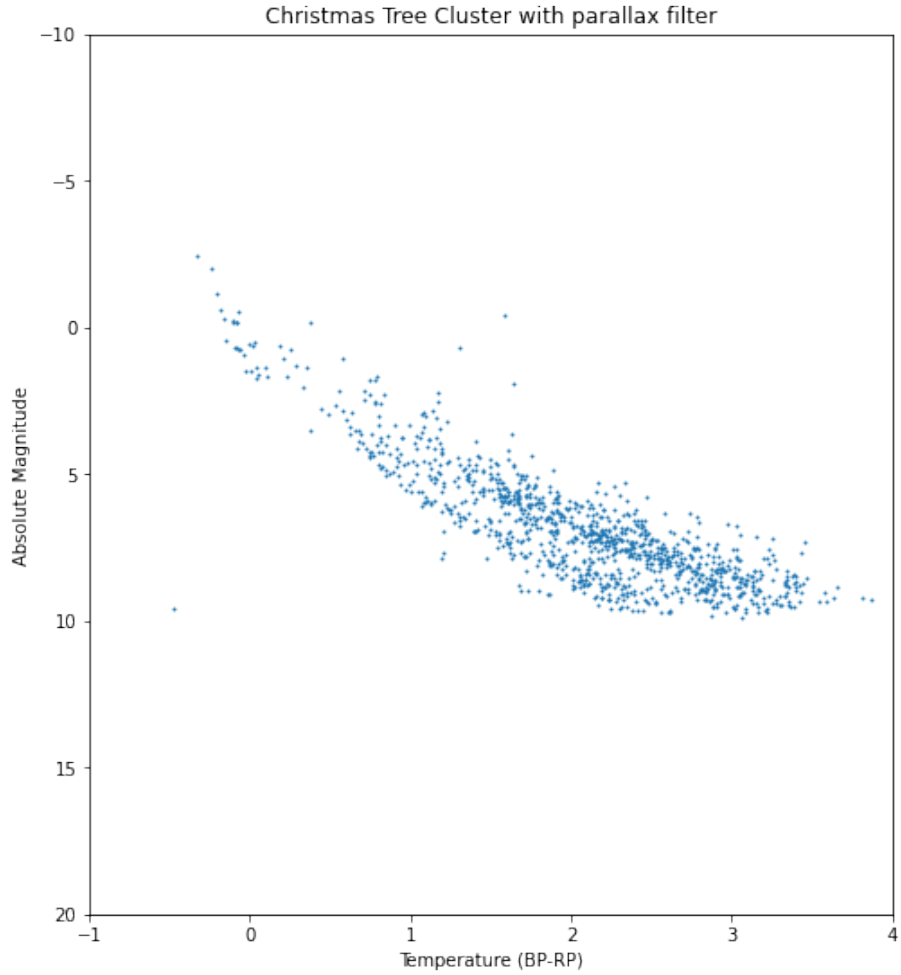


Figure 4.19: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance between 660 and 830 pc

.

Comparing the HR diagrams with and without the distance filter

- The main sequence branch is more spread out, perhaps indicative of the star-forming activity and population of young stars in the region.
- The white dwarfs have almost disappeared in the distance filter, which shows these are not part of cluster - which corresponds to the fact that the cluster has larger population of young stars (white dwarfs appear at the end of stellar life cycle).
- On similar lines as above, there are very few red giants compared to the main sequence.

4.1.7 Ptolemy Cluster (M7)

We follow the same method as used earlier to obtain the following graph. Again, the stellar background was too dense, so the parallax error was limited to $\leq 5\%$.

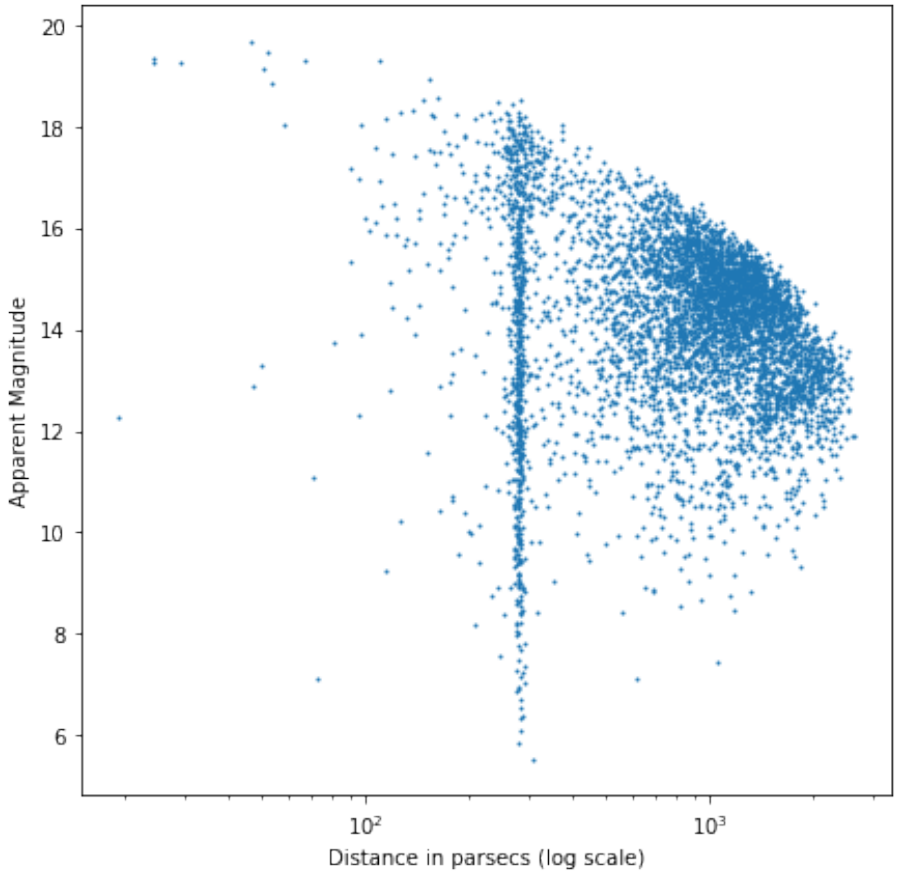


Figure 4.20: Plot of distance and apparent magnitude

.

In this case there is a clear stellar density at just under 300 pc, so we set the parallax mask accordingly.

HR diagram for the full data gives

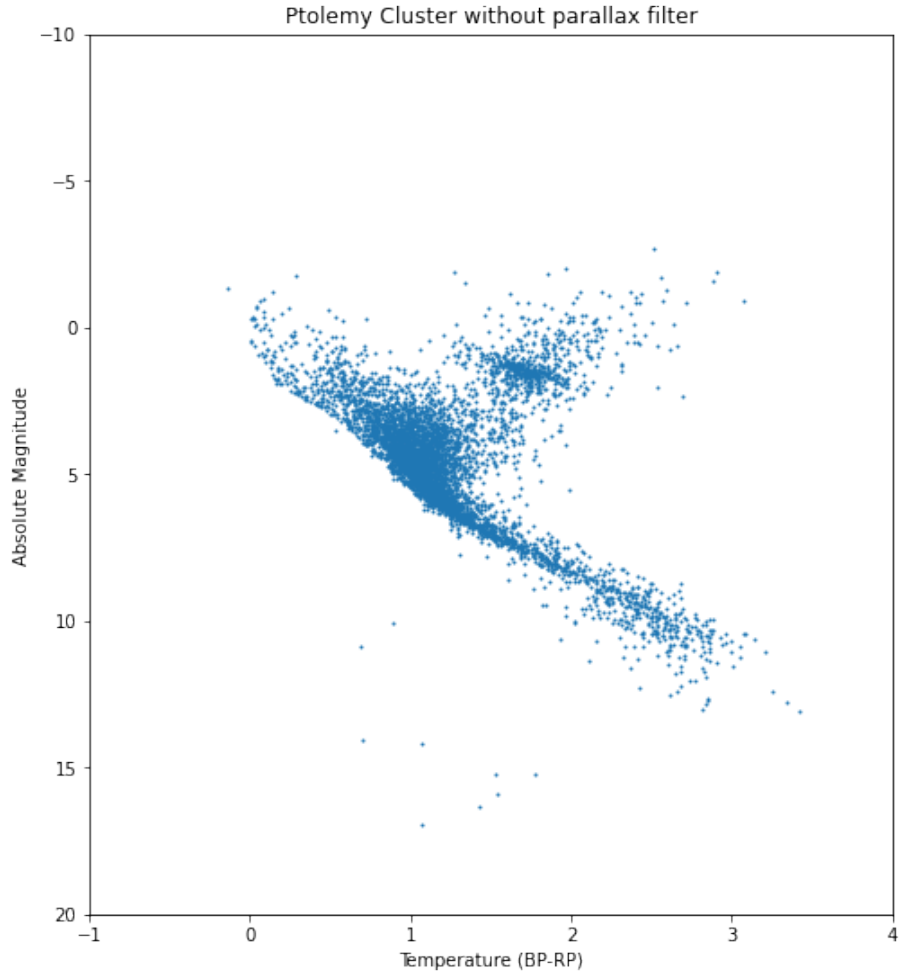


Figure 4.21: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch

.

Now we notice fewer white dwarfs than in the case of Pleiades, but there is clearly more density at the top part of main sequence (near the giant branch and hotter stars) than in the lower part (cooler, smaller stars).

Applying the distance/parallax filter and plotting gives

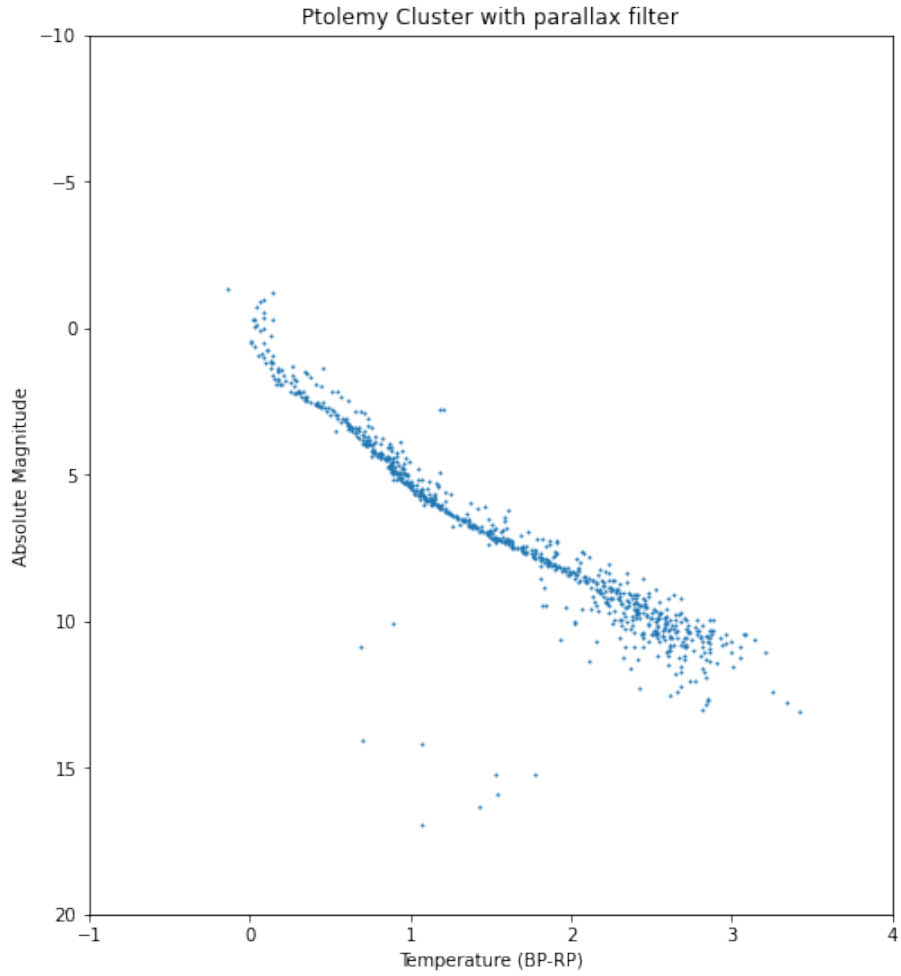


Figure 4.22: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance $<$ 300 pc

.

Comparing the HR diagrams with and without the distance filter

- The main sequence branch has nearly uniform density along the line, which means the cluster has an even population of main-sequence stars of all ages.
- There are almost no stars in red giant phase. Most of these hotter main-sequence stars and red giants thus belong to beyond 300 pc area.
- The white dwarfs remain, which means they are part of the cluster.

4.2 Spectral Analysis, Data Cleaning and Astrometric Plots

4.2.1 Dividing Stars into Classes

We now attempt to break the HR diagram for each cluster (except Pi Puppis, which has too few star samples to be studied well) into the star types. The broad classification into the giant phase, main sequence and white dwarfs was done along these lines:

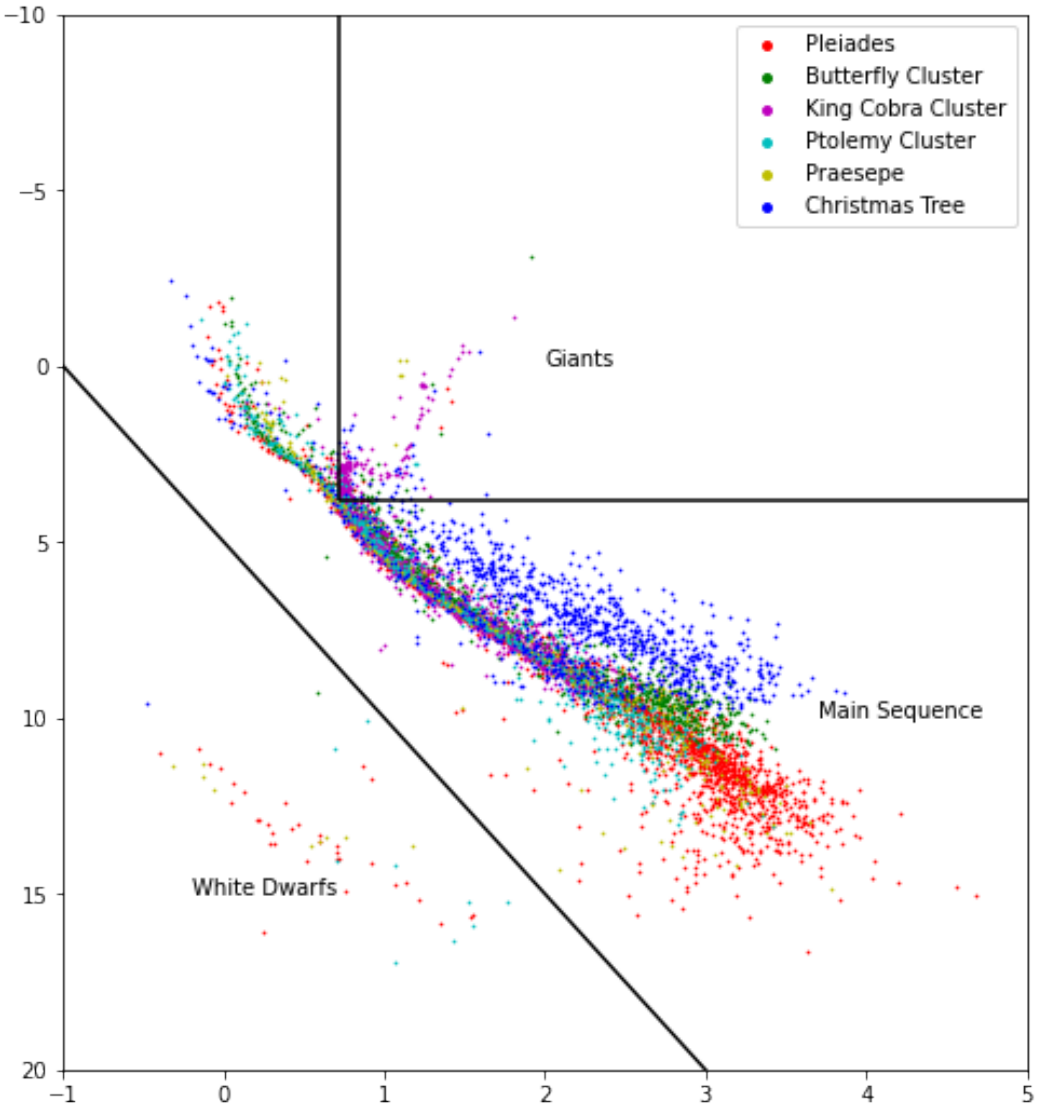


Figure 4.23: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, of our stars with the lines that were chosen to separate them - note that these are only rough boundaries, not precise borders between stages of stars' life cycle

.

We obtain from the Gaia website, the full HR diagram and the rough classification into the spectral types of main sequence (O, B, A, F, G, K, M) and split our main sequence along similar boundaries.

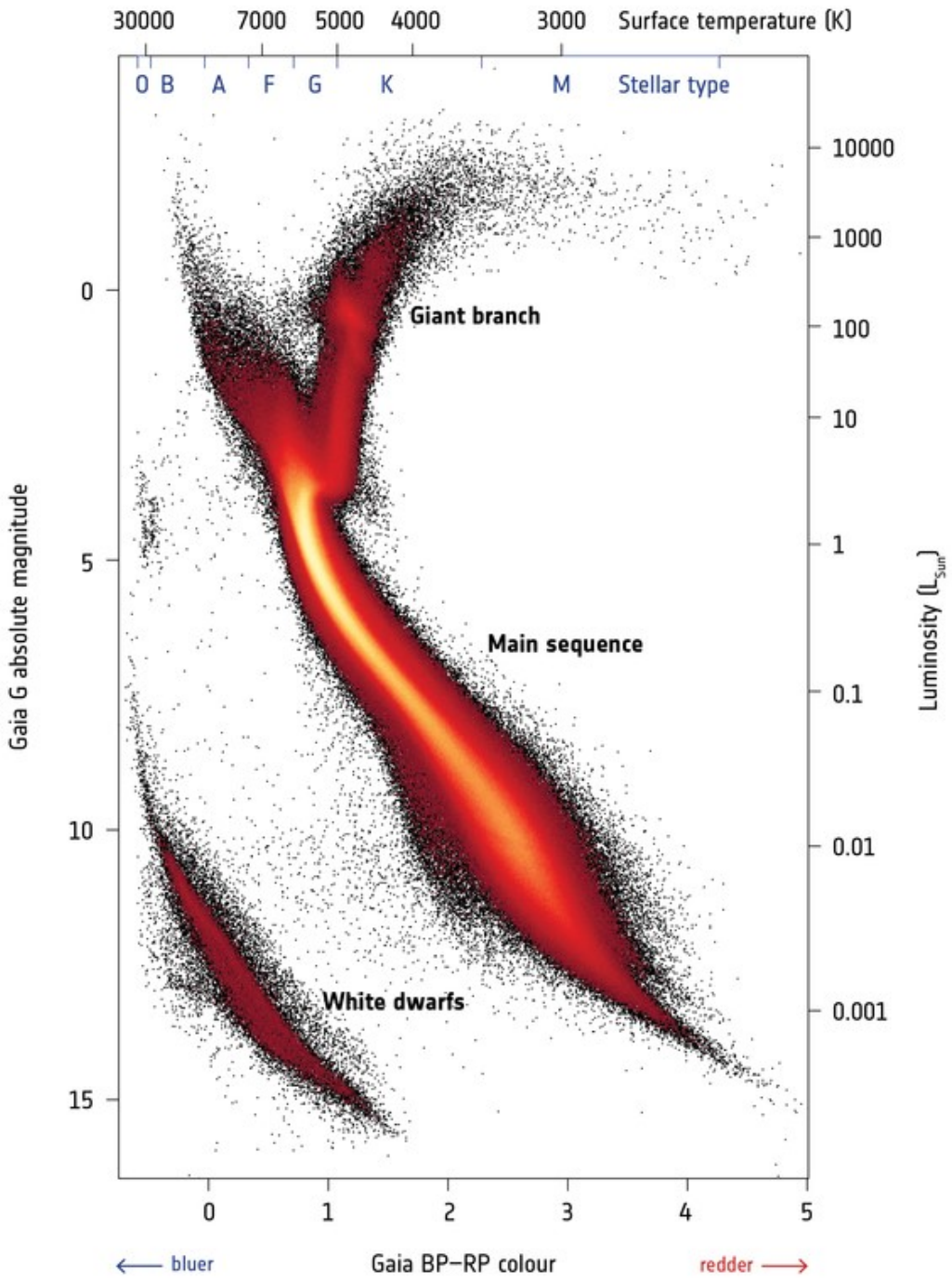


Figure 4.24: HR Diagram for Gaia stars with spectral classification

.

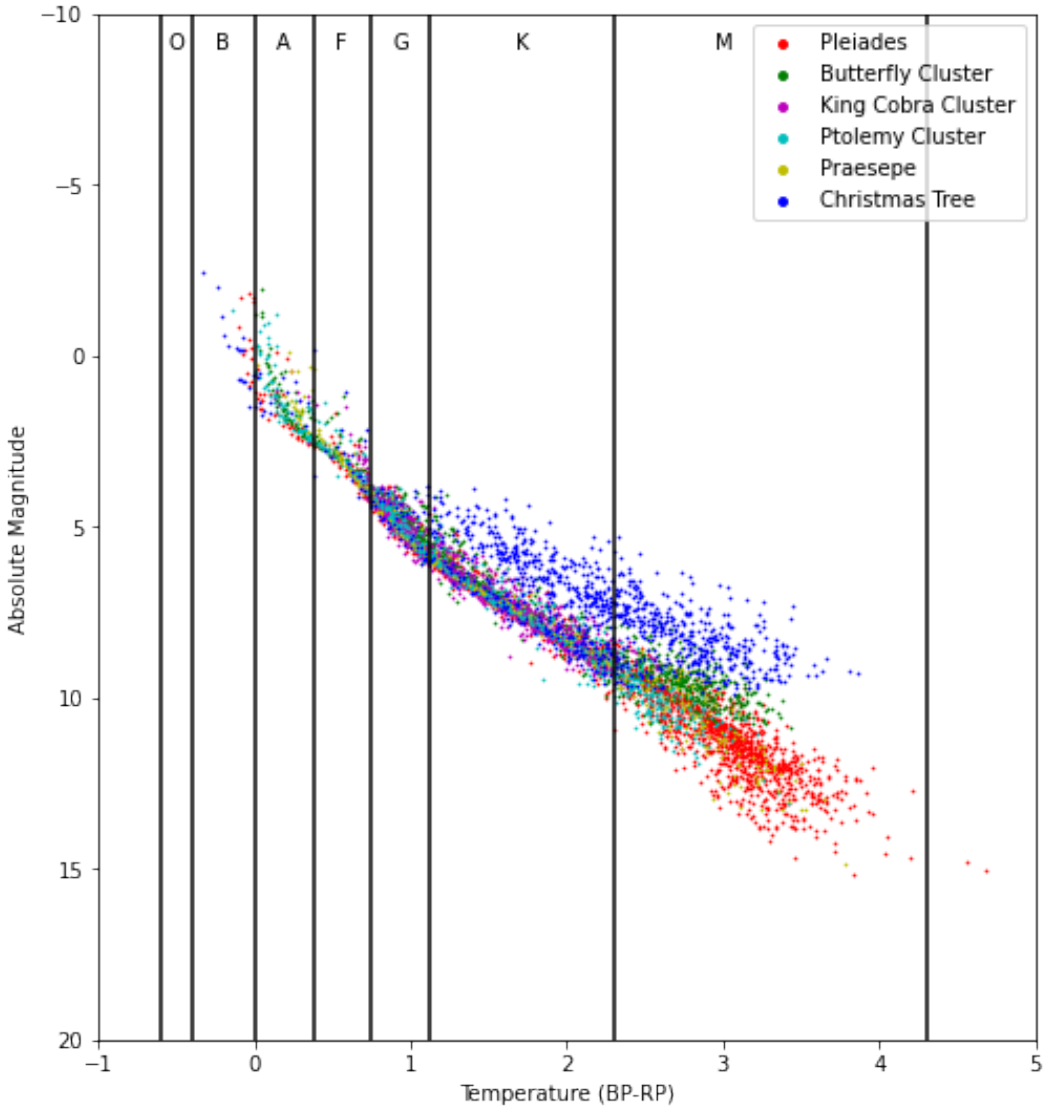


Figure 4.25: Spectral classification on our HR diagram

.

Again, the above lines are not hard borders, just a rough estimate of spectral classes.

4.2.2 Data Cleaning

One of the basic properties of a star cluster is that all its members move together in space i.e. have nearly equal value of proper motion. Based on this, we can use proper motion and distance to filter out non-members. In our analysis, we remove all stars beyond one standard deviation (from mean) along both proper motion and distance. The cleaned data can be visualized as in Fig 4.26.

We then plot the HR diagram (Fig 4.27) for the cleaned data, with the colour coding of spectral classes, as shown above. Colours used here for each class will correspond to the colours used in astrometric analysis in the next part. Red giants are marked in fuchsia and white dwarfs in green, to remove ambiguities when they are plotted together.

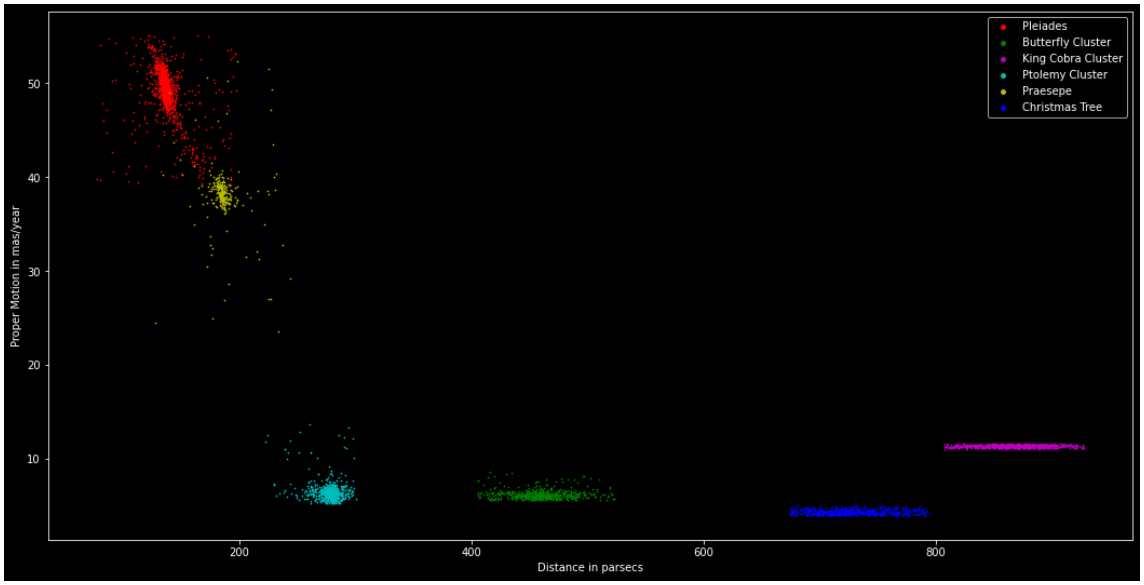


Figure 4.26: Plot of Proper Motion and Distance for each of our clusters. Outliers beyond one standard deviation from the average value are removed

.

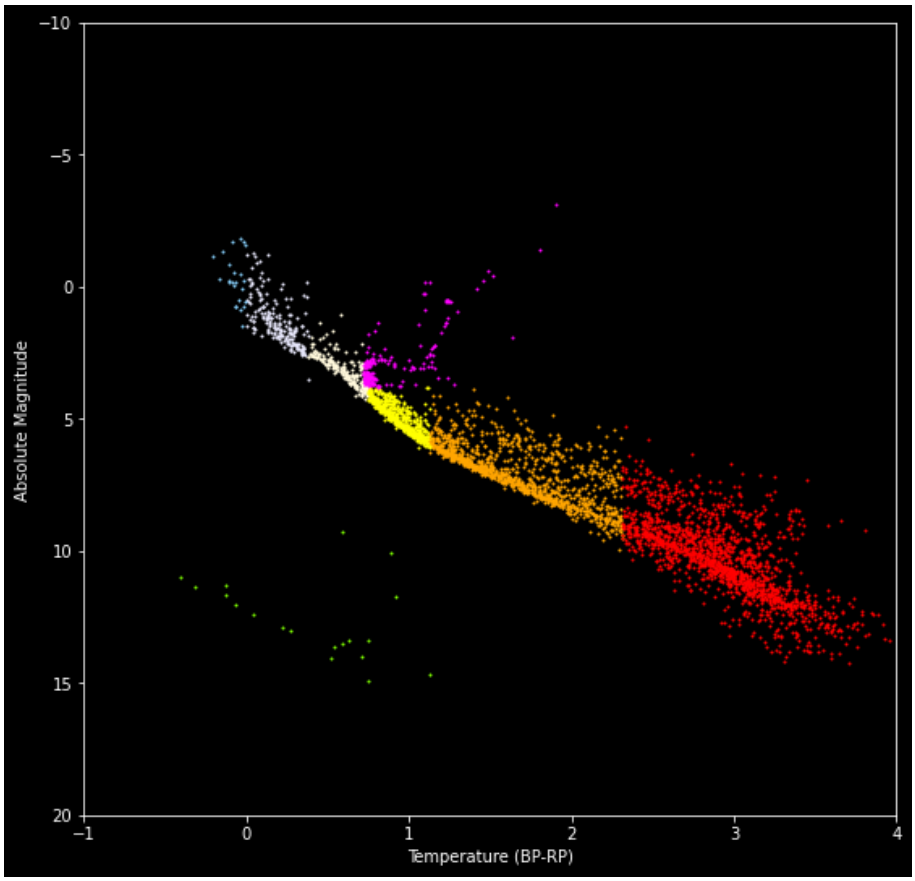


Figure 4.27: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, of stars from the cleaned data

4.2.3 Relating Kinematics to Star type

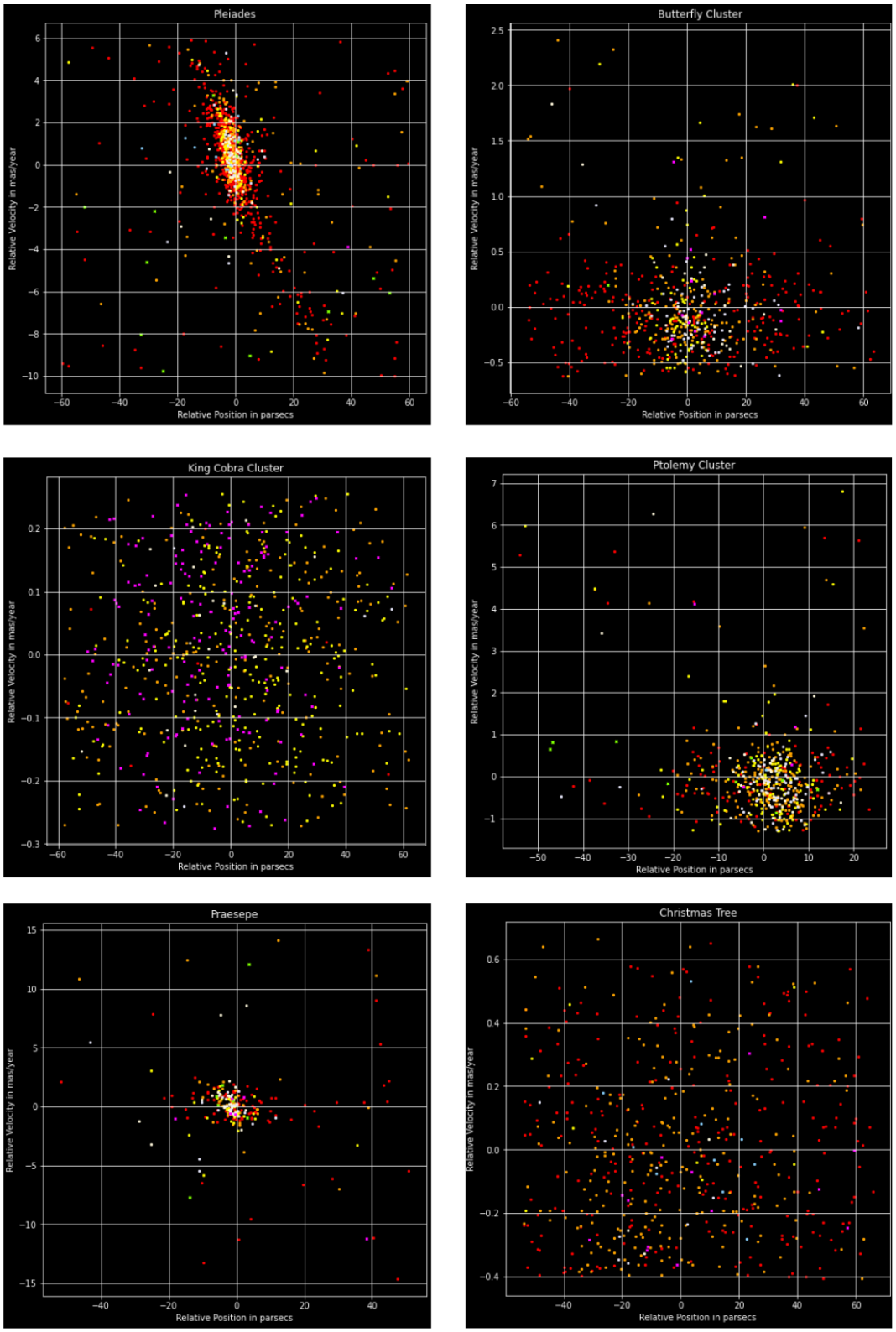


Figure 4.28: Plot of Distance and Proper Motion of all the six Open Clusters

In each of the above plots, the mean distance and mean proper motion of the cleaned data of each cluster was taken. Then plots were made for the **relative** proper motion i.e. motion w.r.t. average motion of cluster, as well as **relative** distance i.e. distance from centre of cluster.

The relative distance was again found using parallax i.e. it accounts for only the radial, line-of-sight distance. The transverse distance using the angular displacement along RA and Dec was not added as that component gave at most 2-3 parsecs, which is completely insignificant compared to the radial distance.

4.3 Observations and Modelling of Open Clusters

A standard model of open clusters is that all stars in the cluster were born from the same giant molecular cloud and have roughly the same age. They start off as a group, moving with nearly identical velocities through the galaxy, before gravitational perturbations from other objects cause them to get dispersed apart from the cluster. Given this, we would expect the following:

4.3.1 Population of Age

In old or medium-aged clusters, the most massive stars, which have the shortest lifespan, will already have evolved into their red giant phases. On the HR diagram, this corresponds to the absence of massive, luminous blue stars on top-left of main sequence line and the presence of a higher number of red giants as the cluster grows older.

In other words, if this model is correct, an open cluster cannot be expected to contain both hot massive stars and red giants in high numbers. This is corroborated by our HR diagrams, all of which either have a high number of red giants or massive blue stars, but not both together.

4.3.2 White Dwarf Deficit

Out of the six clusters studied, after data cleaning, only Pleiades (which had the highest star samples available in Gaia) had a number of white dwarfs at 14; two had between 5 and 10 and the rest had nearly 0 white dwarfs. This can be explained by our model in two ways

- Open clusters have a tendency to get dispersed over time as said above. Since a white dwarf appears at the end of a star's life cycle, it is very much possible that the star cluster has already been dispersed to a high degree by the time a star dies. However, the estimated number from this theory is still higher than the observed number.
- A white dwarf is a supernova remnant. A supernova is a huge burst of energy and matter. So any asymmetry in the distribution of matter when the outer layer of star is expelled, would result in the white dwarf acquiring a velocity in a direction to conserve momentum. This extra velocity would be sufficient for the white dwarf to shoot away from the cluster.

4.3.3 Star Formation and Structural Perturbations

This is based on the plots of relative proper motion and relative distance. We can say that members of the star cluster having low relative velocity and low relative distance, are more tightly bound together. In the plots above, the more tightly

bound cluster members hence correspond to the stars closer to the origin $(0,0)$ in the grid. Using the colour coding mentioned earlier, we observe a few trends and attempt to explain it based on our current model

1. In Pleiades, Butterfly Cluster, Ptolemy Cluster and Praesepe we observe the stars higher on main sequence (coloured in white and yellow and a few in blue) to be closer to the centre i.e. more tightly bound. In contrast, the stars lower in main sequence (mostly red and orange) are spread further away.
2. In the King Cobra Cluster, we observe a significant lack of lower main sequence red stars, but a higher number of giants (which was evident from its HR diagram as well).
3. In Christmas Tree cluster and Pleiades having a noticeable number of blue (high on main sequence) stars, we observe most blue stars concentrated around the centre.
4. White dwarfs in Pleiades, Ptolemy Cluster and Praesepe are also farther away from centre than main sequence. This can be explained on similar lines as the cause of white dwarf deficit in the preceding part.
5. Deviation along y-axis (the proper motion) is the least for King Cobra cluster and Christmas Tree cluster (about $\pm$ 0.5 or less).

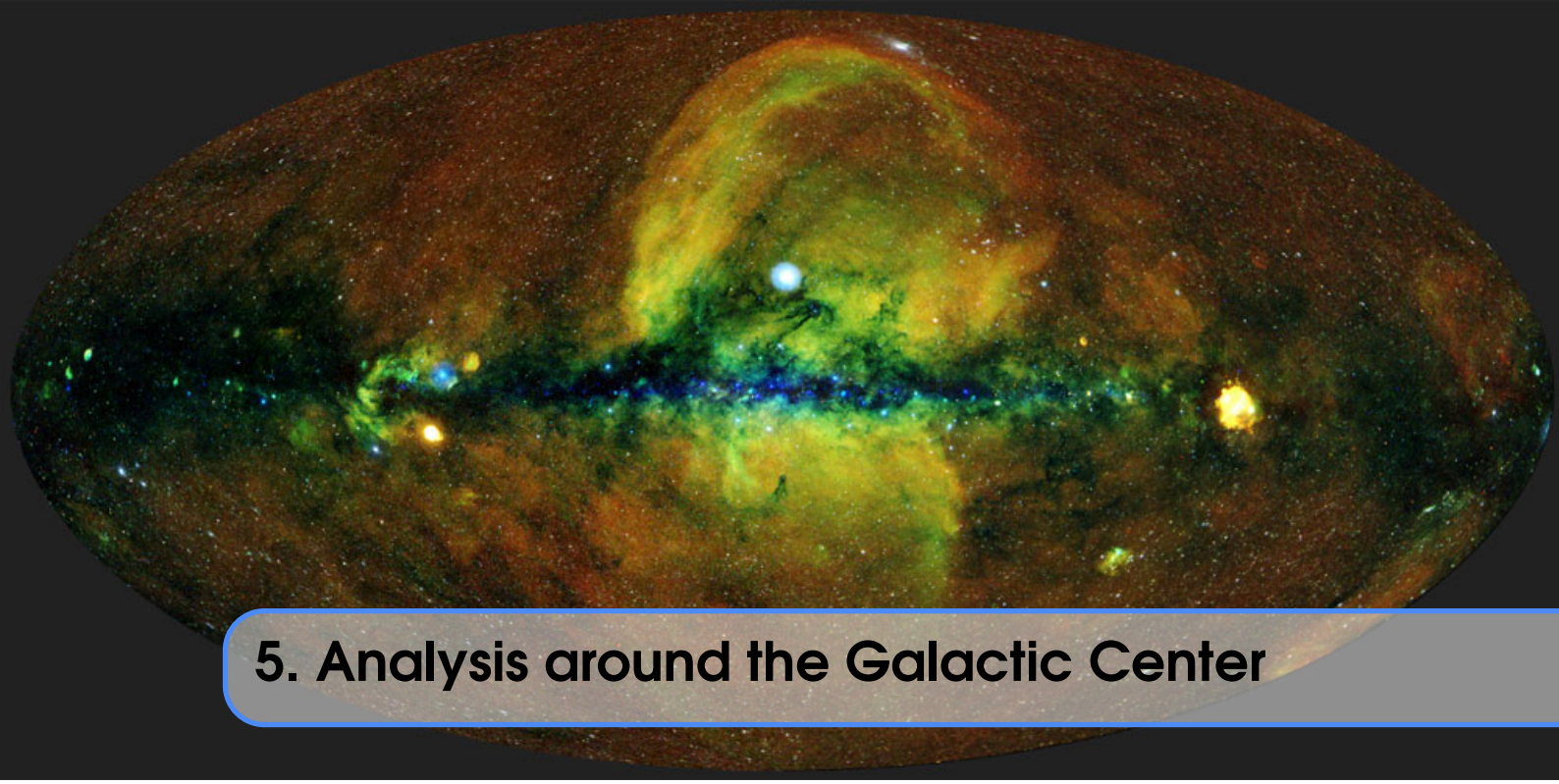
4.3.4 Explaining these with the Model

Some plausible explanations for the above observations can be formulated as follows:

1. Lower main sequence stars (red) are of lower mass than the white or yellow stars. Low mass stars in an open cluster are more susceptible to gravitational perturbations, which results in them getting farther away from the centre (average position and average proper motion) of the cluster. In Pleiades, these low mass red stars have high differences in proper motion which eventually ends up in them being ejected from the cluster. In the other three, these have nearly zero proper motion difference but are significantly far in position, which again makes them more susceptible to perturbations
2. The earliest red giants appear from the shortest-lived main sequence stars i.e. those in the high luminosity and temperature (also high mass) part of main sequence. In contrast, when these high mass stars form giants, the lower mass red main-sequence stars are still in their main sequence, as they have a much longer life-span. So age cannot be a cause for absence of red main-sequence stars.
   Here we again resort to the perturbation model as above. A high number of red giants indicates that King Cobra cluster is significantly older than any of the other five, which means it has had more time for its stars to experience perturbations from neighbouring objects. What this is expected to result in over the long run is the expulsion of low mass members of the cluster. In other words, the red main sequence stars have been around for so long that perturbations have driven them sufficiently far away from the cluster to be associated with it any more.
3. Star formation in open clusters is theorized to have occurred from the same giant molecular cloud. By the force of its own gravity, the cloud would be densest near the centre. This would facilitate the collapse of huge chunks of gas into larger stars. Hence it would lead to formation of high mass, high temperature, high luminosity blue stars near the centre of the cloud.

Along similar lines, as the molecular cloud grows thinner outwards, this would result in formation of smaller, low mass red main-sequence stars, offering another explanation of why these are removed from the centre.

4. The white dwarfs are farther away from the centre along the same reasons as those of white dwarf deficit discussed earlier. In fact, white dwarf deficit is an extreme case of them being far from the centre.
5. In case of the King Cobra cluster, its old age has resulted in most of the low-mass outliers getting ejected from the cluster, and only the tightly bound members remaining and moving with a common velocity in space. So the relative difference in proper motion is low.
   In the case of Christmas Tree cluster, the adjoining Cone Nebula is still actively forming stars. Thus it shows a higher population of main-sequence blue giants (which generally die young, hence are observable only in young clusters). So most stars, all formed around the same time, are young and hence still tightly bound to the cluster.
   In these plots, only the middle-aged star clusters have been around long enough for its low mass members to have been significantly perturbed, while at the same time remaining close enough to be labelled as part of the cluster.



5. Analysis around the Galactic Center

The **Galactic Center** is the rotational center and the barycenter of the Milky Way galaxy. Its center contains a supermassive black-hole (SMBH) of about 4 million solar masses, named **Sagittarius A***. It is a compact radio source, and is about 8 Kilo-parsecs away from Earth in the direction of the Sagittarius Constellation.

The stellar population of about 10 million in a radius of one parsec around the galactic center is dominated by red giants, substantial amount of supergiants and Wolf-Rayet stars.

In this stellar analysis I have used two approaches to extract data that represent parts of the galaxy as a function of stellar variables. The two types of datasets used are:

1. **Beam Data:** A $(1^o \times 1^o)$ square window centered at the RA and DEC coordinates of the Galactic Center is specified. All the stars, binary stars and the white dwarfs that fall into this window are extracted to form the Beam Data. This is because this data set contains all the mentioned sources in a beam with a cross-sectional area of $(1^o)^2$ from the Earth towards the Galactic Center.
2. **Sphere Data:** A sphere of radius 300 parsecs, centered at the Galactic Center is specified and all the stars, binary stars and white dwarfs falling into this region are extracted. The query parameters were determined using trigonometric formulations including the inverse of tan of the ratio of the radius of desired sphere around the center of the galaxy and the distance of the galactic center from the earth (approximated to be 8 kilo-parsecs).

5.0.1 Data Querying

The Galactic Center is located at the following RA and DEC coordinates:

$$\text{RA} = 17h\ 45m\ 40.04s$$

$$\text{DEC} = -29^{\circ}\ 00m\ 28.1s$$

Beam Data Query

```
query = '''select g.source_id, g.ra, g.ra_error, g.dec, g.dec_error,
g.parallax,g.parallax_over_error,g.bp_rp,g.pm,g.pmra,g.pmra_error,
g.pmdec,g.pmdec_error,g.phot_g_mean_mag,
g.phot_bp_mean_mag,g.phot_rp_mean_mag,g.radial_velocity,
h.classprob_dsc_combmod_star + h.classprob_dsc_combmod_whitedwarf +
h.classprob_dsc_combmod_binarystar as comb
from gaiadr3.gaia_source as g, gaiadr3.astrophysical_parameters as h
where g.source_id = h.source_id and
g.ra between 265.9157 and 266.9157 and
g.dec between -29.50728 and -28.50728
order by g.phot_g_mean_mag desc, comb desc'''
```

The above code selects a window of $1^o \times 1^o$ centered at the above mentioned Galactic Center Coordinates.

Sphere Data Query

```
query = f'''select g.source_id, g.ra, g.ra_error, g.dec, g.dec_error,
g.parallax,g.parallax_over_error,g.bp_rp,g.pm,g.pmra,g.pmra_error,
g.pmdec,g.pmdec_error,g.phot_g_mean_mag,
g.phot_bp_mean_mag,g.phot_rp_mean_mag,g.radial_velocity,
h.classprob_dsc_combmod_star + h.classprob_dsc_combmod_whitedwarf +
h.classprob_dsc_combmod_binarystar as comb
from gaiadr3.gaia_source as g, gaiadr3.astrophysical_parameters as h
where g.source_id = h.source_id and
g.ra between 264.269 and 268.564 and
g.dec between -31.155 and -26.86 and
g.parallax between 0.1205 and 0.1299
order by g.phot_g_mean_mag desc, comb desc'''
```

The above query extracts all the stars in a radius of 300 parsecs around the galactic center.

If you imagine the milky way as a disk with spiral arms emanating from the center, our sun is approximately halfway from the center to the visible edge. Our solar system lies between two prominent spiral arms, in what astronomers once thought was a mere bridge of stars, gas, and dust clouds. Our spiral arm is the Orion-Cygnus Arm, or simply, the Orion Arm or Local Arm. You sometimes still hear the names Orion Bridge or Orion Spur. Th Milky Way is a barred spiral galaxy, which means it has a central bar, so where exactly does our solar system lie?
It's sandwiched by two primary spiral arms, the Sagittarius and Perseus Arms. The artists' concepts above and below show the various spiral arms, along with the location of our sun on the Orion-Cygnus Arm.
The positional arrangement of the Galactic Bar with respect to the earth is critical for many observations that we will be making from the data. Also, the position of intermittent arms between the earth and the galactic center would be responsible for the stellar data distribution.

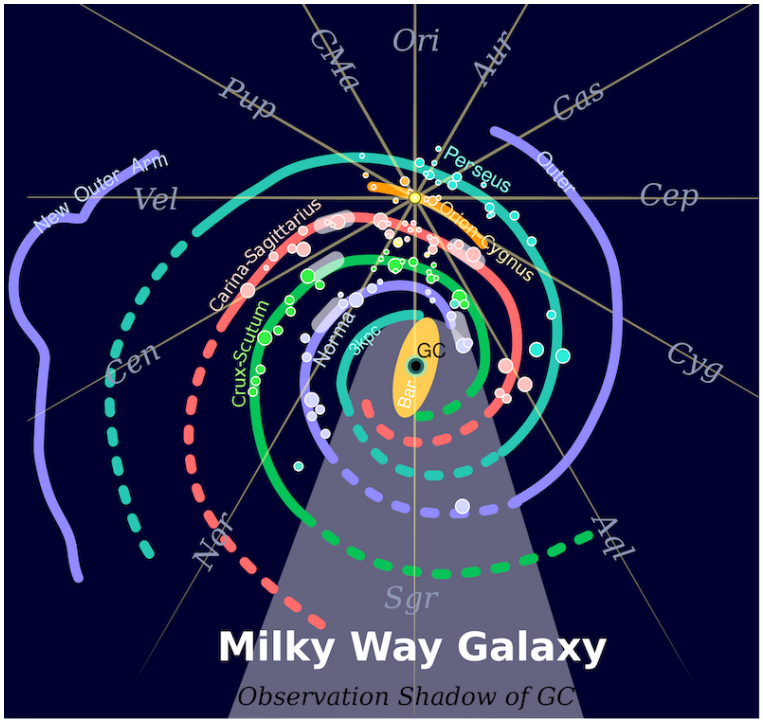


Figure 5.1: RA and DEC Distribution of Beam Data

5.1 Data Visualization

RA and DEC Distribution

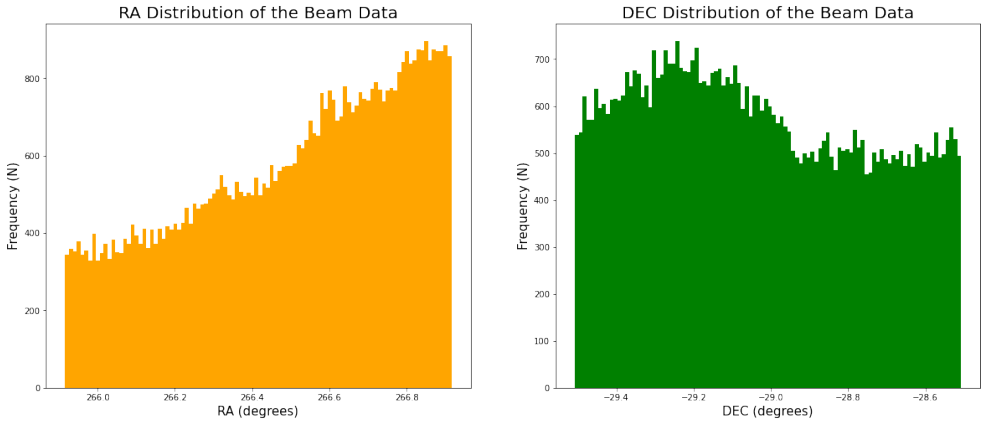


Figure 5.2: RA and DEC Distribution of Beam Data

For the Beam Data, the RA and DEC distribution are usual with the number of stars lying towards the lower left of the galactic center to be more than rest of the the window. This observation is fairly explained by the density wave model of the galactic evolution where the density of stars is pocketed due to the mathematical density wave function of Non-Linear Dynamical systems.

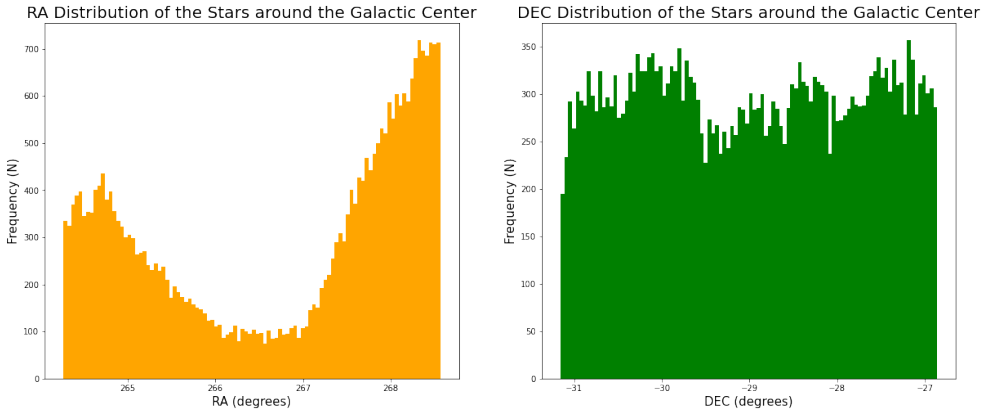


Figure 5.3: RA and DEC Distribution of Sphere Data

.

In the case of the sphere data, we see a huge lack of stars just about the galactic center. This is evident by the extremely low number of stars at the RA and DEC coordinates $(\delta 1^o$, and two sharp peaks obtained when moving away. This trend is seen in both RA and DEC distribution suggesting a structure similar to the Kuiper Belt of the solar system. The stars form a somewhat thick shell around the galactic center with fewer stars inside the shell.

> When observed from the earth, the 300 parsec radius thick shell of around the galactic center should have a brighter lower left quadrant as compared to rest of the three quadrants of the ring. This distribution can be affiliated to the position of the galactic bar with respect to the earth, which agrees with the stellar density being higher in the third quadrant of the ring.

Distance Distribution

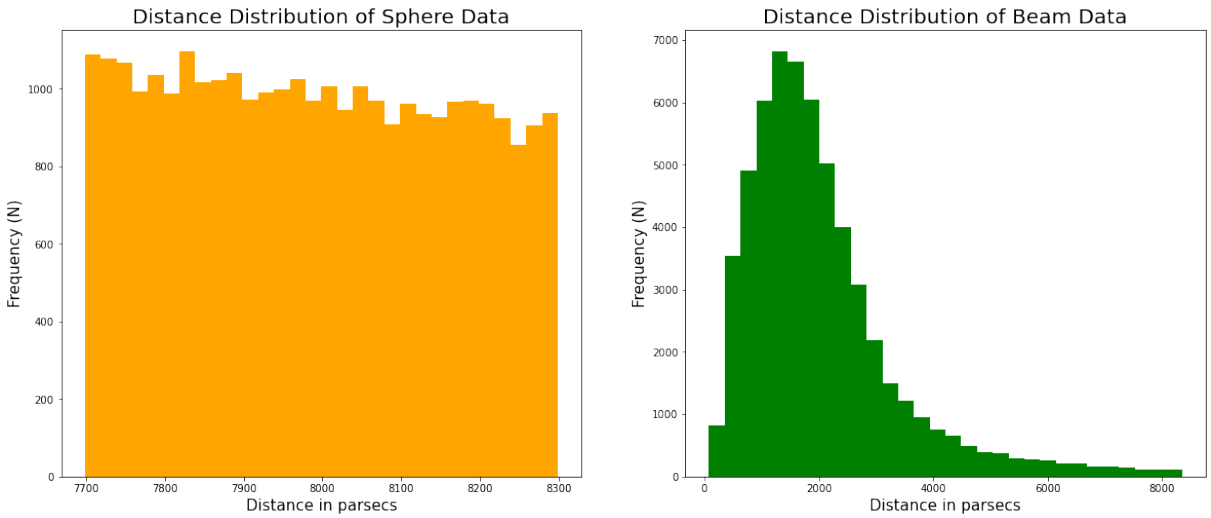


Figure 5.4: Distance Distribution of Beam and Sphere Data

The distance distribution of the stars is fairly in accordance to the accepted trend. The distances of the stars in the sphere data is fairly well distributed with slightly

more stars near the earth as compared to the other end. This can also be well explained by the fact that the galactic bar is inclined towards the earth and hence the component of the density towards earth contributes in more stars near us.
For the beam data, the number of stars is fairly low very near and very far from earth. But there is a sharp peak present at about 1800 parsecs. This is because of the intersection of our observational beam with the **Carina Sagittarius** arm, which lies around 1800 parsecs from the earth, and can be deduced to have a width of about 2000 parsecs.

Proper Motion Distribution

In the sphere data, a significant fraction of the stars have a substantial proper motion with mean of about $10\ (mas/year)$. Contrary to this, in the beam data, most of the stars have a very low proper motion, suggesting that the stars around the galactic center are moving more than the stars in the beam data. To further reinforce out claim, we made a 2D scatter plot of proper motion for both, the Beam and the sphere data. The density scatter plot of the sphere data just came out to be a zoomed in version of the central dense region of the beam data. This is in alliance with our claim that the stars in the sphere data are moving relatively fast and in harmony as compared to the beam data. In other words, we can say that the tip of the beam that intersects with out sphere data shows similar characteristics to the sphere data.

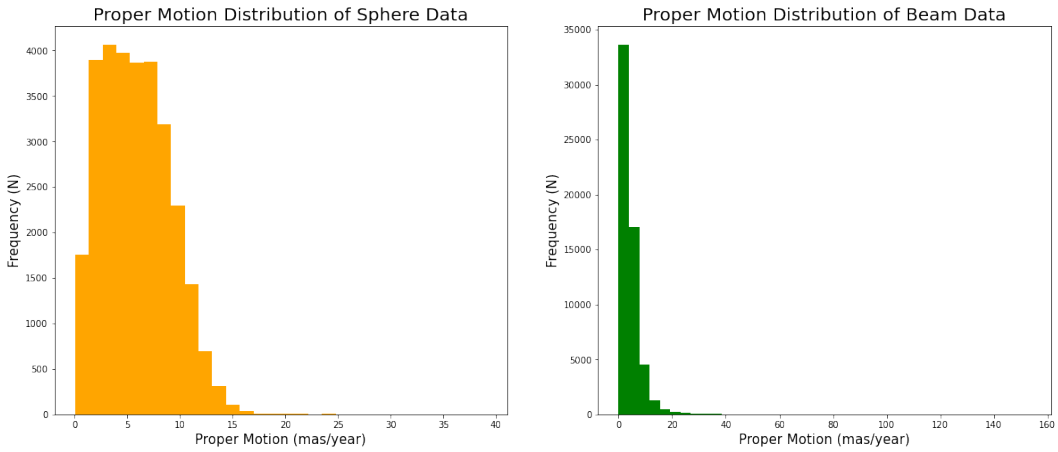


Figure 5.5: Proper Motion Distribution of Beam and Sphere Data

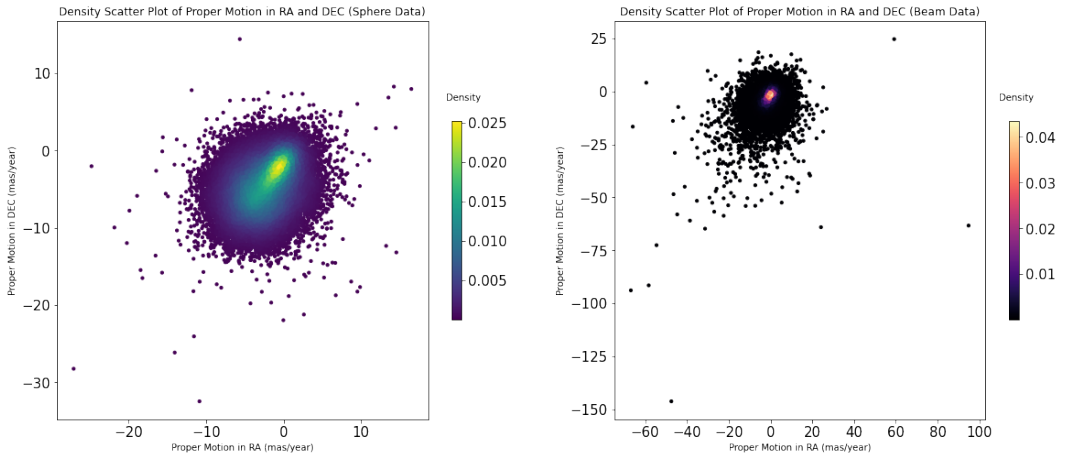


Figure 5.6: Density Scatter Plots of Proper Motion in RA and DEC for Beam and Sphere Data

Apparent Magnitude Distribution

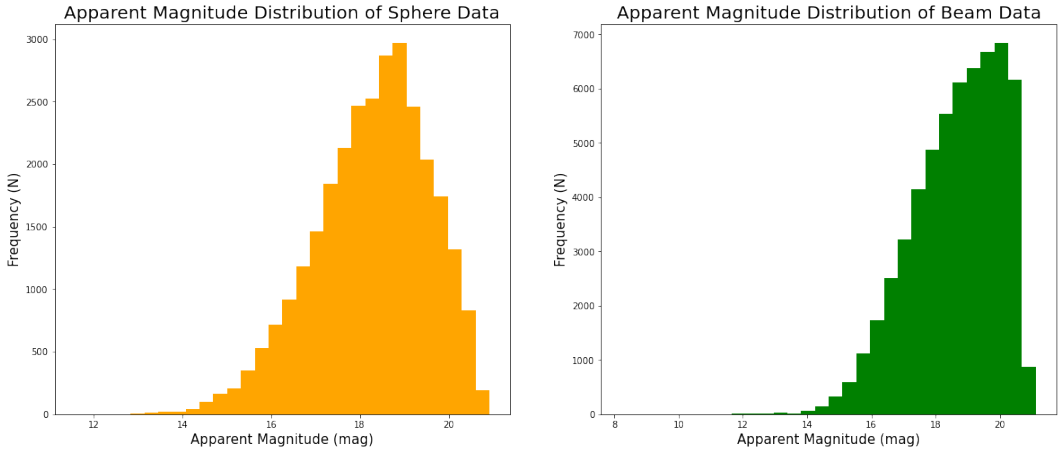


Figure 5.7: Apparent Magnitude Distribution of Beam and Sphere Data

The apparent magnitude peaks at around 18 for the sphere data, where as most of the stars in the beam data have an apparent magnitude of 20. This can be affiliated to the dust clouds present in the arms of the galaxy, which the beam data penetrates. The presence of dust leads to a lower 'apparent' magnitude as compared to the sphere data which contains little to no dust particles that diminish the magnitude of the stars. This does not mean that the absolute magnitude would also follow the same trend. The models of galactic evolution say otherwise. Since the galactic center must contains older stars, it must harbour a population of lower absolute magnitude with respect to the rest of the galaxy.

5.1.1 Data Correlation

In this section, we will find relation between two variables of the stellar population in order to find a trend and pattern. This pattern will help us understand some basic properties of the structure and the stellar population census of our galaxy.

Apparent Magnitude v/s Distance Distribution

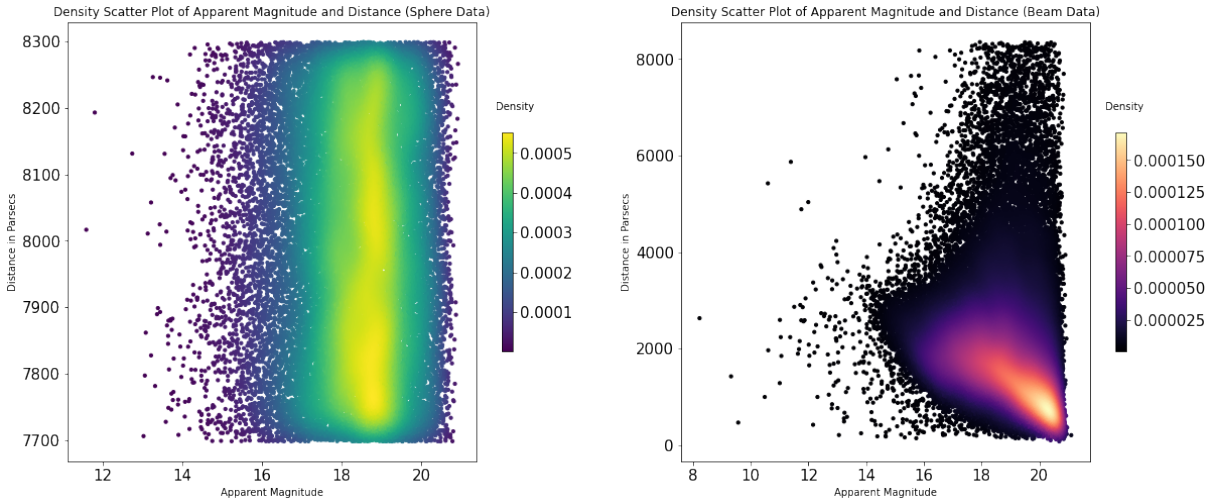


Figure 5.8: Apparent Magnitude v/s Distance Distribution of Beam and Sphere Data

- **Sphere Data:** All the stars are distributed uniformly along the distance scales but have nearly the same apparent magnitude. This displays the striking similarity between the type of stars present in the sphere and the apparent surrounding of each star in this case. The vertical density line is centered about 19 magnitude, which is in accordance to the Apparent Magnitude distribution that we made earlier.
- **Beam Data:** The stars in the Beam data are somewhat scattered as compared to the sphere. According to the graph, most of the stars are present substantially close to earth and have very low magnitude. This is in perfect accordance with our model of position of the sun in the milky way galaxy. The sub-arm contains little to no bright stars. When moving towards the galactic center i.e. away from earth, we will encounter the Carina Sagittarius arm, where the stars have a higher magnitude. This is seen in the graph as well, because when we move along the distance axis, we also see a increase in the magnitude of the stars, around the same distance as the Carina arm. The stars then become dim rapidly as we move away, because of the filtration caused by the dust clouds of the Carina arm and also the decreasing distance from the galactic center, indicating the presence of more older and dimmer stars.

Apparent Magnitude v/s Proper Motion Distribution

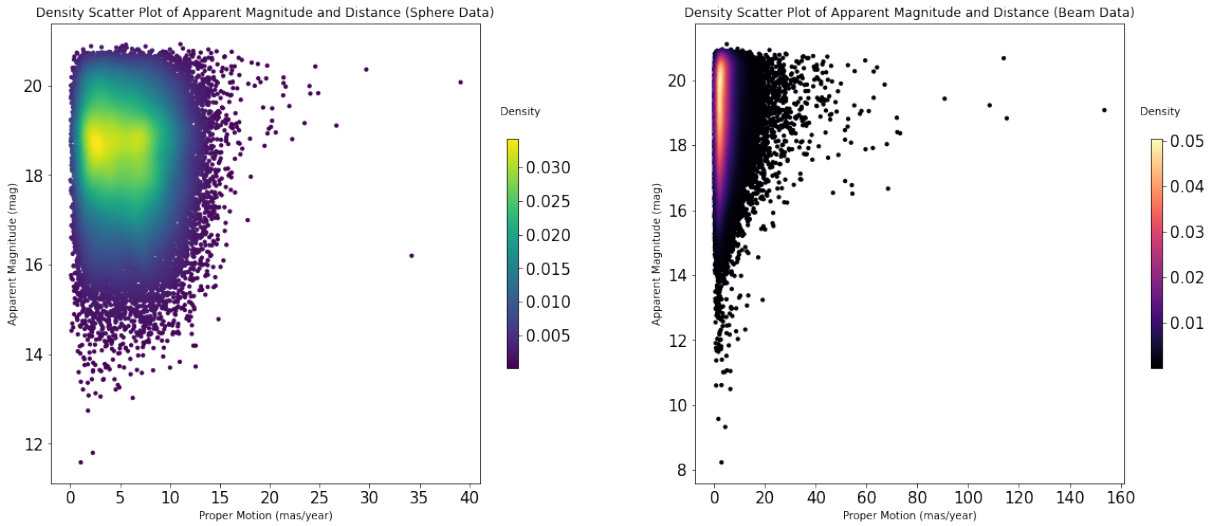


Figure 5.9: Apparent Magnitude v/s Proper Motion of Beam and Sphere Data

The plot between Proper motion and apparent magnitude just reinforces that fact declared in the previous graphs. The the similar stars are moving similarly suggesting an ordered structure with motion around the galactic center. In case of the Beam data, most of the stars have very low proper motion, with exceptions of some extremely fast stars, that lie around the dim regions of the plot. There is little to no grouping observed in this section as the stars in the beam are not bound by a single system and hence are free to move independent to each other.

Distance v/s Proper Motion Distribution

The above graphs show a relationship between the proper motion of the stars and the distance. The sphere data shows that most of the stars have a fairly uniform proper motion meaning about 8 (mas/year). We are also able to see a scatter of stars that follow a somewhat Gaussian distribution above the dense scatter of the stars. The peak of the Gaussian lies about the coordinate of the galactic center, indicating the proper motion of the stars rises rapidly when we reach near the galactic center with stars reaching a proper motion of 40 $(mas/year)$. This gives a beautiful insight into the fast moving stars around the central super massive black hole allowing us to see its impact on the stellar dynamics around the galactic center.
We also suspect the presence of some **Hyper Velocity Stars (HVSs)** in both the distributions. This is seen as the sparsely scattered stars with exceptionally high proper motion. The most basic explanation of the HVSs around the galactic center is the gravitational catapult of Binary stars where one of the star in the system absorbs a gravity assist from the blackhole and in turn throws its companion out. Such stars have a high change of leaving the galaxy and entering the Inter-Galactic Medium (IGM).

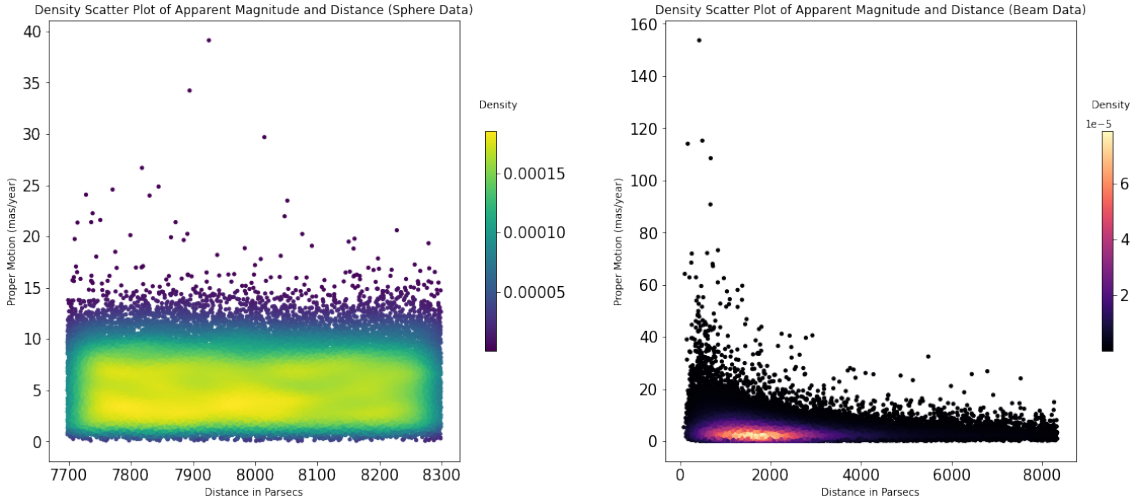


Figure 5.10: Distance v/s Proper Motion of Beam and Sphere Data

5.1.2 Color Magnitude Diagram

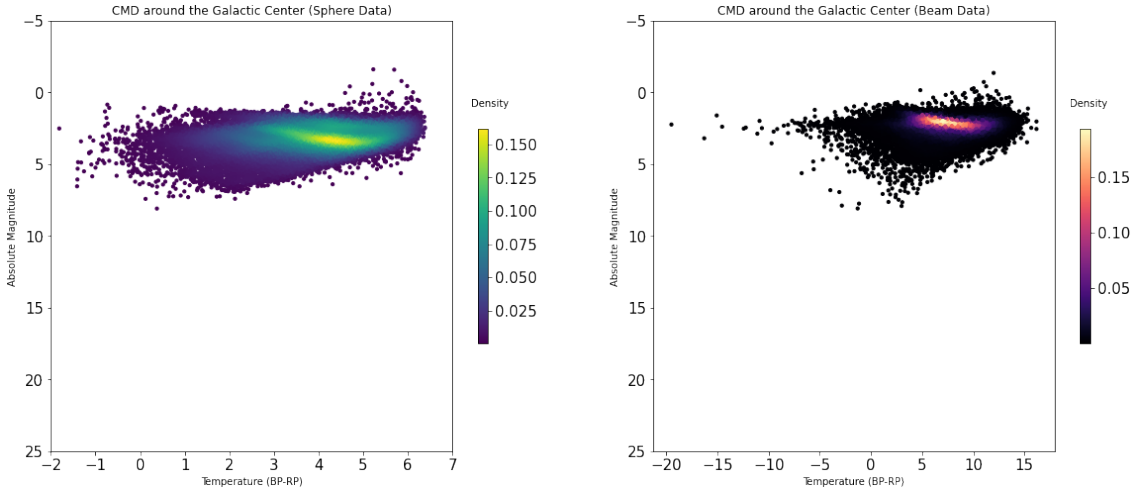


Figure 5.11: Color Magnitude Diagram for Sphere and Beam Data

The Color Magnitude Diagram for the sphere data shows a strong restriction of stars in the Giant Branch, its extension and Left region of the Main Sequence. As discussed by E.Gallego et.al. the structure of the stellar density around the SMBH is similar to a cusp, with many different types, sizes and aged stars in a symphony. But due to the stars being around the galactic center, their age is much more than the stars in the branches of the galaxy, which is evident from the fact that most of the stars present around the galactic center are restricted to the Giant Branch, with the maximum density present for the red giant stars. There is also a substantial scatter to the left of the main sequence, a place alleged for the Wolf Rayet stars. This has been discussed and been confirmed by the study of the characteristic Paschen-alpha lines as discussed by Mauerhan et. al. as they studied the presence of isolated Wolf Rayets and O Super-giants in the galactic center region.
The Beam Data is however a mix of stars as it contains all the stars ranging from

the galactic center to the slices of the arms of the milky way. Hence the stars in the region are not well distributed and even have a temperature range three times that of the stars at the galactic center. This might be because of the larger number of stars in the galactic center being giants, whereas the stars in the beam data are fairly distributed.

5.1.3 Radial Velocity Distribution

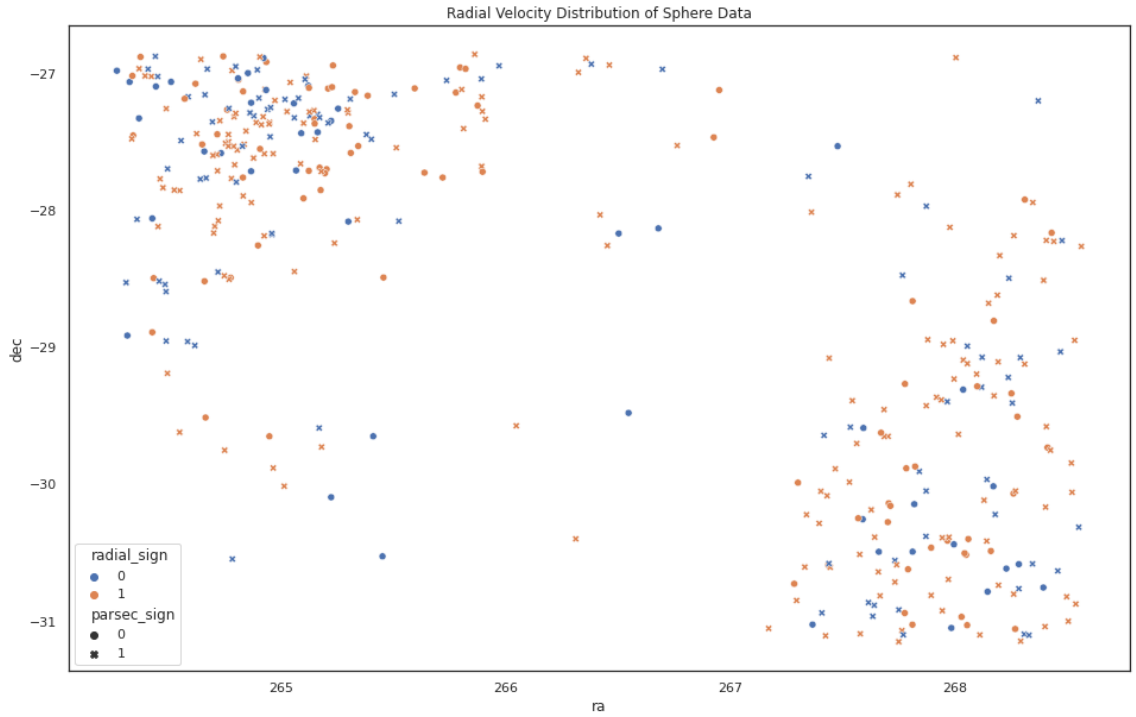


Figure 5.12: Coded Radial Velocity Distribution of the Sphere Data

The above diagram is a color and marker coded distribution of the radial velocity of the stars around the galactic center, scattered about their RA and DEC values. The number of stars in this plot is reduced drastically because of the unavailability of Radial Velocity values for most of the stars in the dataset. The distribution is color coded on the basis of the sign of the radial velocity, the stars with a **negative radial velocity are orange** where as the **stars with positive radial velocity are blue.** Similarly, on the basis of their distance from Earth, considering that the Sagittarius A* lies exactly at 8 kiloparsecs, we mark all the **stars further than 8000 parsecs with a circle** whereas the **stars with a distance less than 8000 parsecs are marked with a cross**. Hence we get the overall motion of the stars represented in a concise manner through this chart. Some confirmed motion traits that we can decipher from the plot are listed below. There is no proper grouping of stars or orderly motion, but instead a chaotic system of stars revolving in both directions.

1. **Left Cluster**
   (a) Orange Stars
      i. Circle: Further side of the SMBH and moving clockwise
      ii. Cross: Closer side of the SMBH and moving anti-clockwise
   (b) Blue Stars

   i. Circle: Further side of the SMBH and moving anti-clockwise
   ii. Cross: Closer side of the SMBH and moving clockwise

2. **Right Cluster**
   (a) Orange Stars
      i. Circle: Further side of the SMBH and moving anti-clockwise
      ii. Cross: Closer side of the SMBH and moving clockwise
   (b) Blue Stars
      i. Circle: Further side of the SMBH and moving clockwise
      ii. Cross: Closer side of the SMBH and moving anti-clockwise

6. Results and Discussion

6.1 Far Away Bright Sources

For the constraints taken for g-band apparent magnitude(< 8) and distance(> 300 pc) it was seen that most of the stars in the sample were a part of the giant branch, with a small number of stars from close to the hotter end of the main sequence(mostly B spectral class with some from A spectral class). This met the expectations as only very bright massive stars would have such low apparent magnitudes when viewed from greater than 300 parsecs away. A majority of the main sequence was absent from the HR diagram from the sample as was expected. Thus a separate study for the red giants wasn't done. Instead plots comparing the stars from sample from the giant branch with the hotter main sequence stars. No clear patterns could be drawn due to a very small number of stars in one set compared to the other.

6.2 Brightest Stars

On examining the top 40,000 brightest stars as seen from earth on the basis of apparent magnitude with parallax error $< 2\%$ we get to see that most of these stars are concentrated in the galactic plane of Milky Way, our home galaxy, due to their proximity. The distribution of number of stars with distance matched the inverse square law for brightness. The 2-D plot of Absolute magnitude and Distance shows the relation between apparent magnitude, absolute magnitude and distance. In H-R diagrams we saw that the stars in our sample are mainly Giant stars and Hotter Main Sequence stars. On separating the Giant stars and Main Sequence stars from our sample and analyzing them separately shows that the Giant stars are distributed on an average at larger distances as compared to Main Sequence stars.

6.3 Closest stars

Upon studying the stars within 60 parses of the sun it was seen that there are no major structures in this region. The stars have very even distribution in the generated sky map. The HR diagram generated shows very high proportion of main sequence stars, followed by white dwarf stars and very less proportion of high mass stars. White dwarf stars were further separated from this data on the basis of their photometric properties in order to study the difference in the sky map distribution and velocities. The graphs thus generated show no difference between the two. This points to a possibility that the main sequence and white dwarf stars have same origins.

6.4 Galactic Centre

We analysed the data is two fashions, namely the Sphere data which is the subset of stars in sphere of radius 300 parsecs around the Sagittarius A*; And the Beam data which takes a thin slice of the Milky way with all the stars from earth to Sagittarius A*. The overall conclusions about the stellar population, density and type can be understood as follows: Most of the stars are present around the SMBH (supermassive black hole) in a doughnut kind of formation, with more stars to one quadrant than the rest three. Also, there are more number of stars near to us than further in the spherical shell of stars around the SMBH; this has been attributed to the central bar in the galactic center being oriented towards us. The proper motion of the stars is much higher around the galactic center as compared to the black hole, which can be understood by the presence of dark mass around the galactic center driving the stars in a relatively faster rate as compared to the outer stars.

Most of the stars are of the same age and temperature around the galactic center, whereas the stars very close to earth are dim as we lie in a minor arm. The stars get brighter and hotter as we reach about 2 kilo-parsecs from earth, because of the beam data containing the Carina arm's stars. The distance versus the proper motion plot suggests a strong relation between the closeness of the star to the SMBH and its radial velocity, as the stars closer to the galactic center are very fast, almost following a Gaussian distribution as compared to the outer stars. Most of the stars around the galactic center lie in the Giant Branch indicating the presence of old stars, and have a comparatively less absolute magnitude than the rest of the galaxy reinforcing the age factor.

The apparent magnitude of these stars is remarkably low as well, which can be attributed to the Compton thick gas and dust clouds that circumscribe the galactic center. Finally, a radial velocity diagram helps us visualise the stellar dynamics around the SMBH in a 3 dimensional notion. Conclusively, the galactic center is a region filled with old aged dim stars that move around it in a spherical shell exhibiting density wave patterns and in both directions, clockwise and anti-clockwise, without any relation to other parameters, making it a chaotic shell of blazing red giants.

6.5 Large Magellanic Cloud

The HR diagram shows how the LMC is an active star-forming region, as well as has older red giants. Among the astrometric plots, we notice a striking feature of

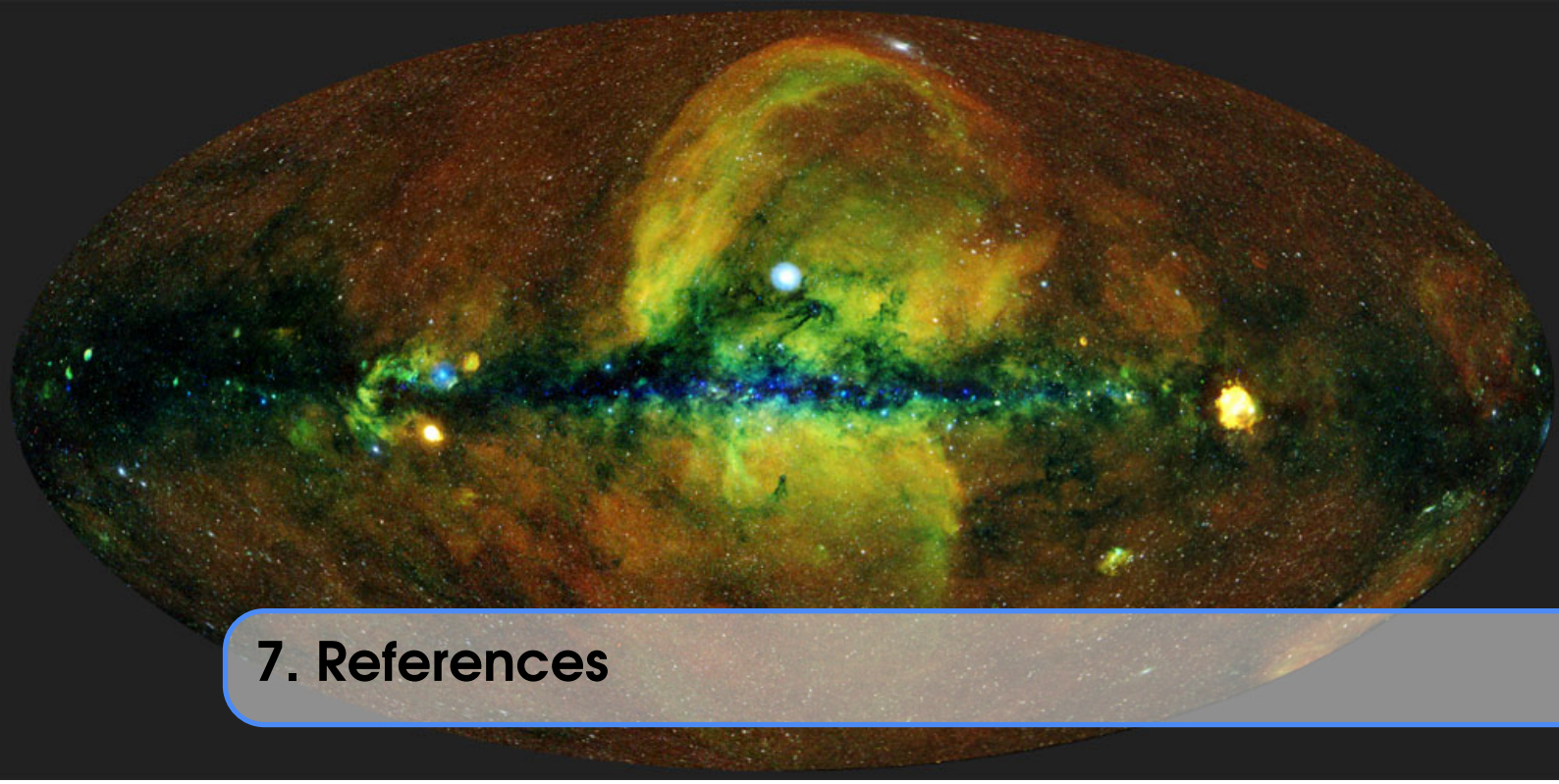
the Proper Motion plots - that all the stars are concentrated at around 2 mas/year, indicating that the LMc stars move together in the sky. We can calculate the approximate tangential velocity using the following relation it has with the mean distance and proper motion:

$$v_t = d \sin \mu$$

where μ is the proper motion of the LMC, v_t is the tangential velocity, and d is the distance in parsecs. For small proper motion (as in our case), $\sin \mu \approx \mu$. Taking d as the mean distance, and substituting the values, this comes out to be about 281 km/s.

6.6 Further Work and Possibilities

Gaia data tables contain large amounts of accurate data of an enormous number of stars. The recent Data Release 3 opens up a whole avenue of research possibilities, which will help us study the physics of stellar processes much better than was possible before. This project was to take the first steps towards more extensive analysis of stellar data, uncovering a lot more interesting phenomena and build a deeper understanding of stars, clusters, galaxies and the Universe as a whole.



7. References

- Gaia DR3 Data Model - to obtain the table names, column names and the information contained in them
- Gaia Data Access using Python - gives the procedure and syntax to extract Gaia data
- Building HR diagrams with Gaia data - the complete HR diagram for all stars in Gaia
- Stellar Classification - Wikipedia
- Temperature of Stars using Gaia Data
- Open Clusters - Wikipedia