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'''
```

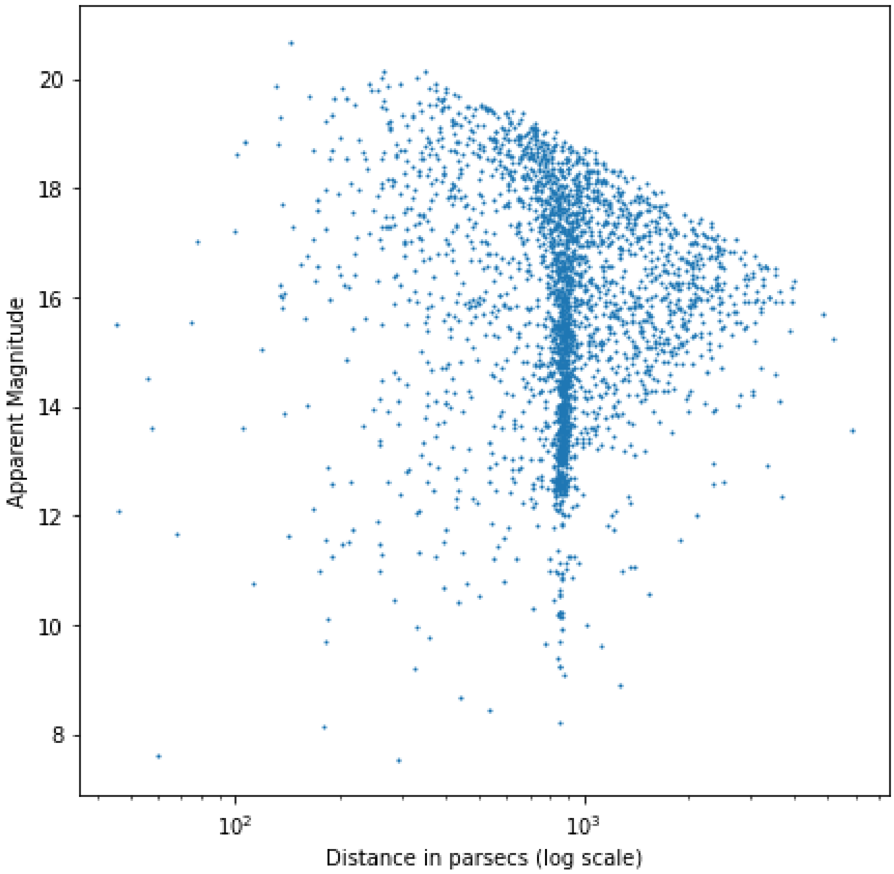

Figure 4.1: Plot of distance and apparent magnitude. The distance distribution is densest at just less than 1000 parsecs

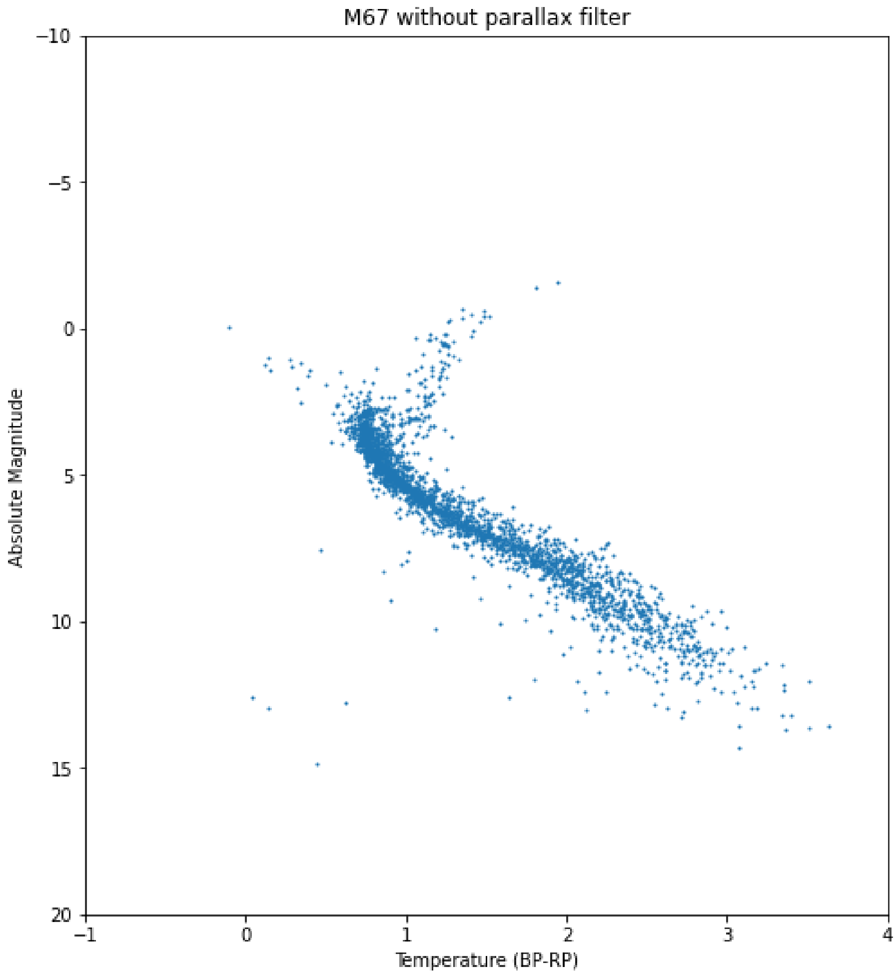

Figure 4.2: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch

.

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.

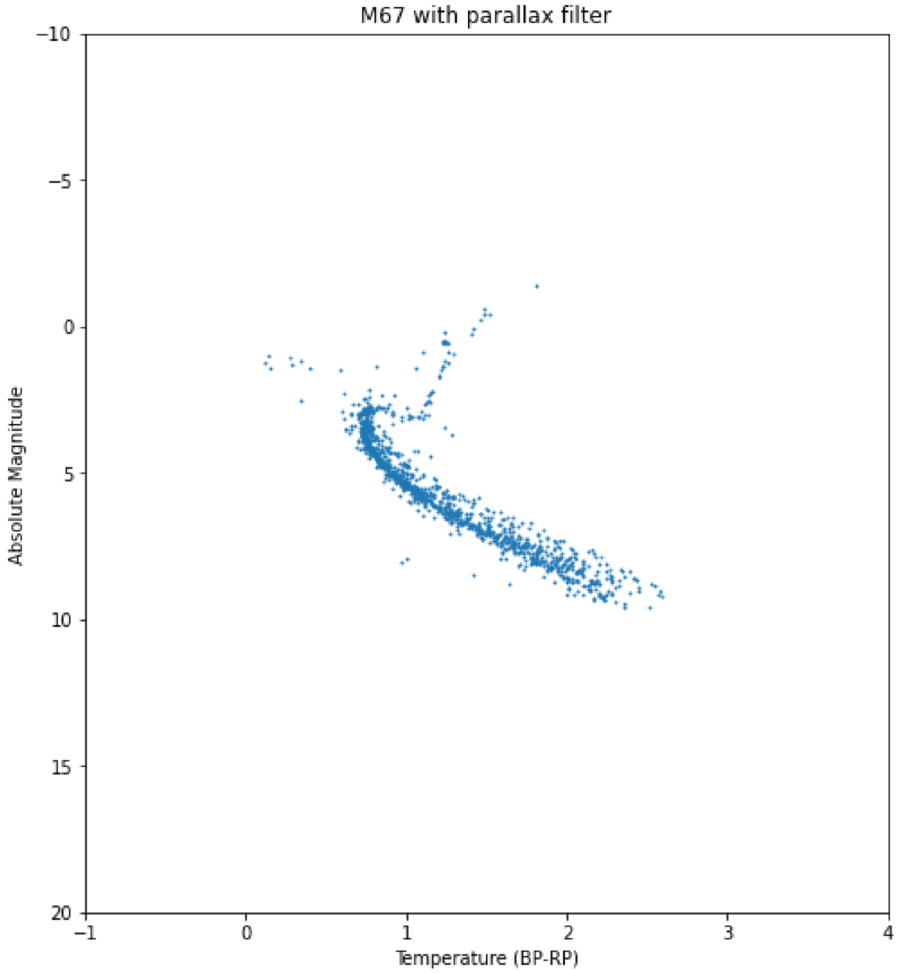

Figure 4.3: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance between 750 and 1000 pc

.

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.

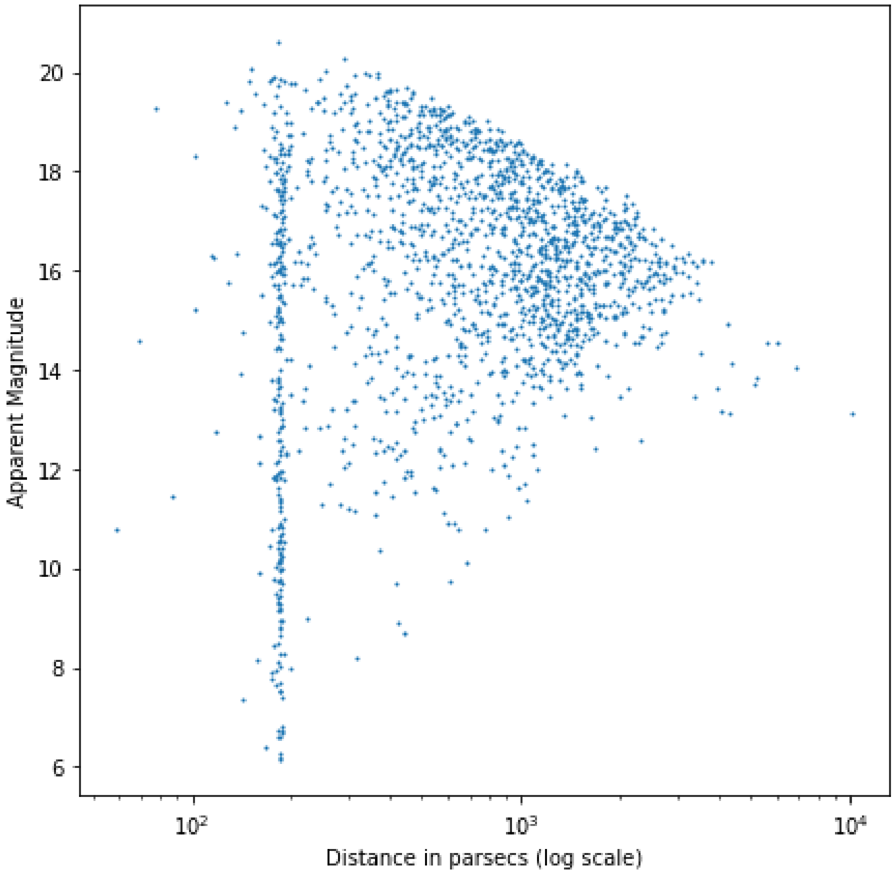

Figure 4.4: Plot of distance and apparent magnitude. The distance distribution is densest at just less than 200 parsecs

.

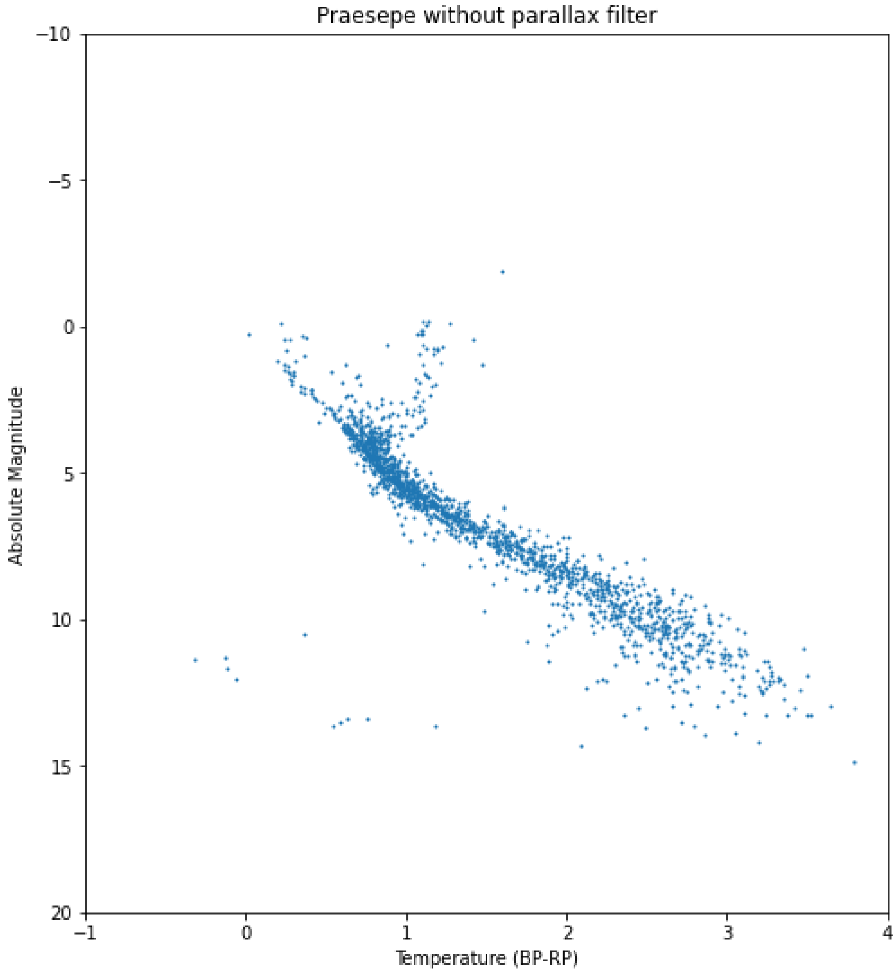

Figure 4.5: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch
.

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

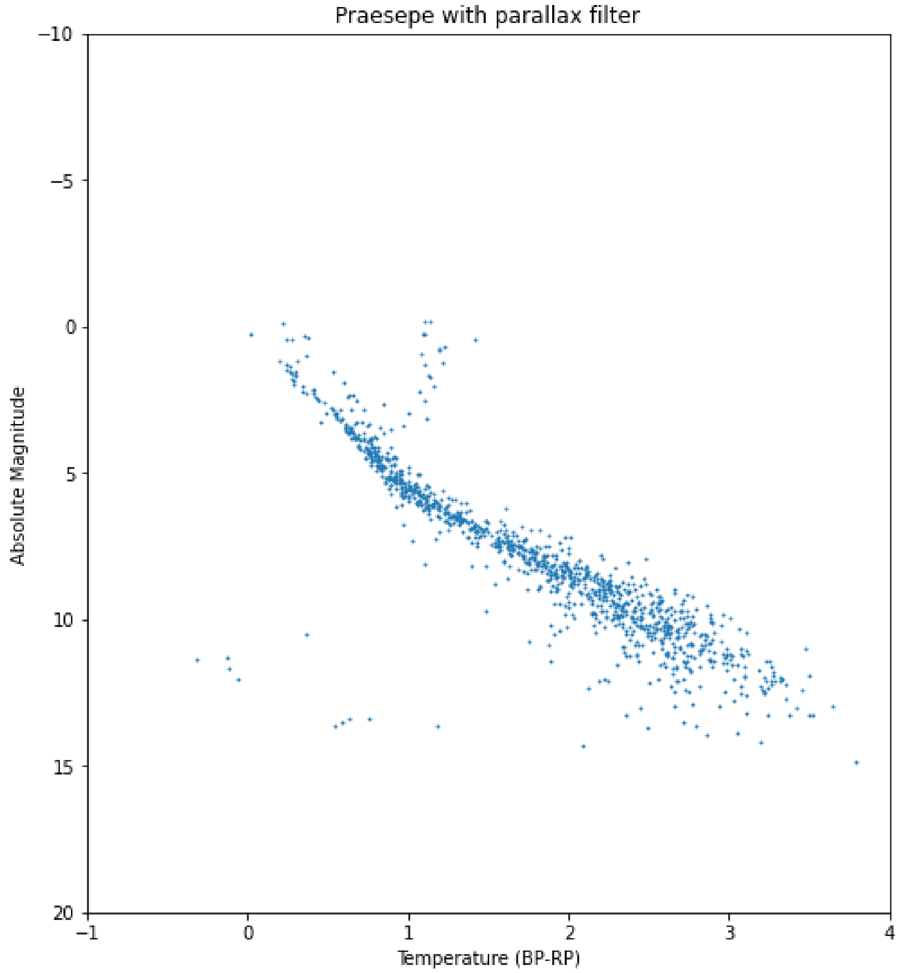

Figure 4.6: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance $<$ 1000 pc

.

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)

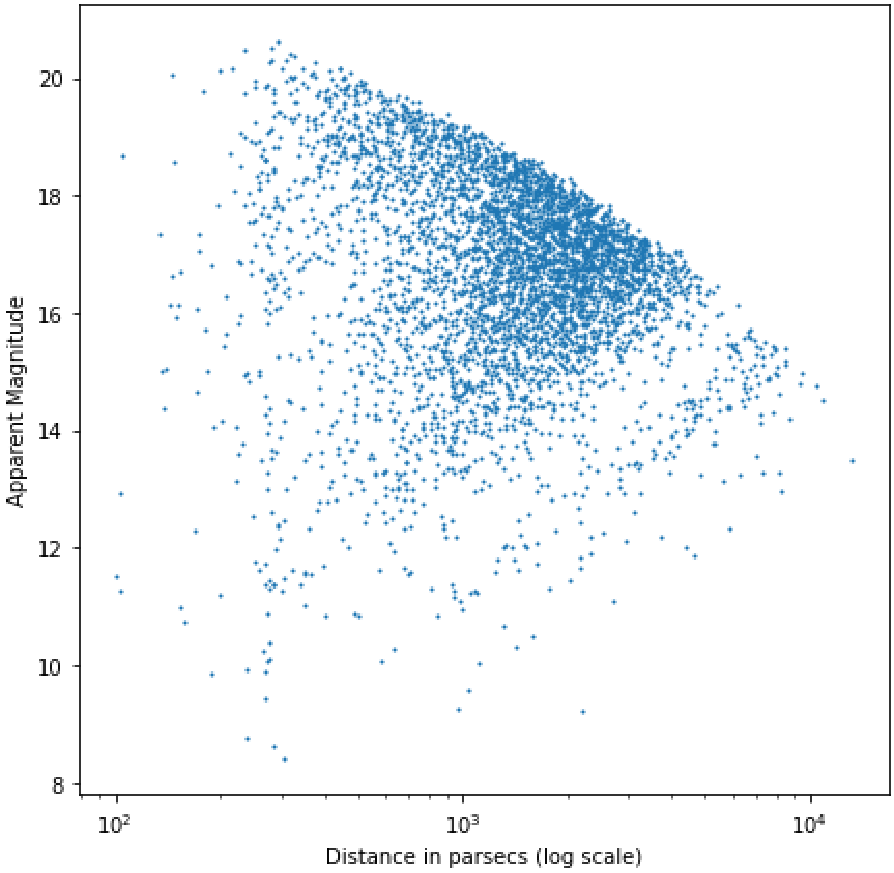

Figure 4.7: Plot of distance and apparent magnitude

.

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.

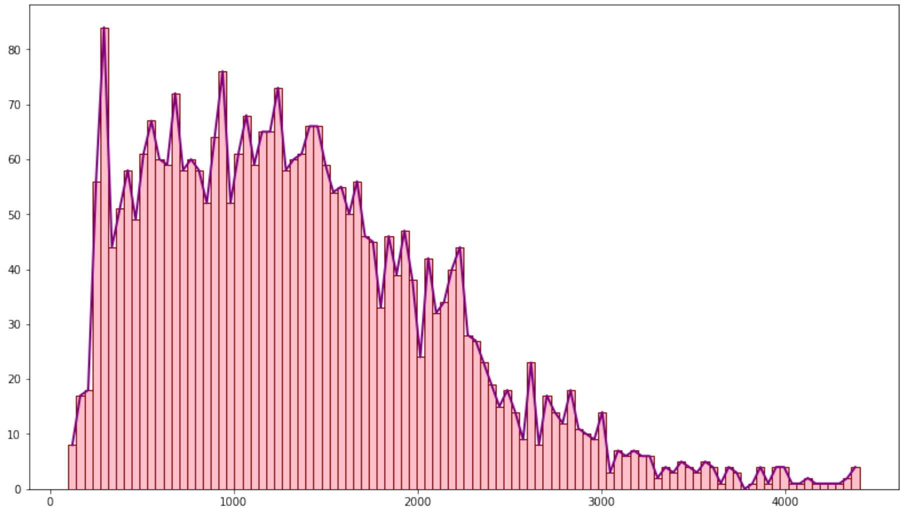

Figure 4.8: Histogram of distance distribution - distance is in parsecs

.

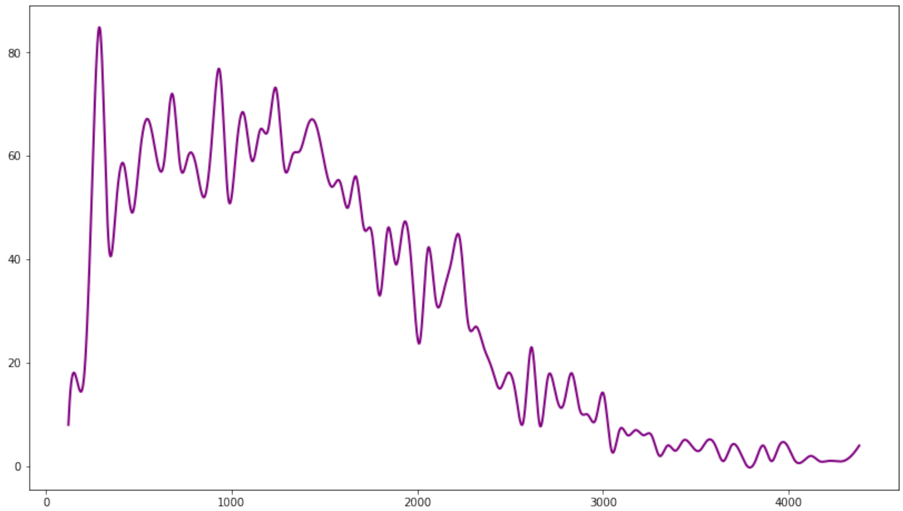

Figure 4.9: Smooth curve of distance (in pc) distribution

.

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

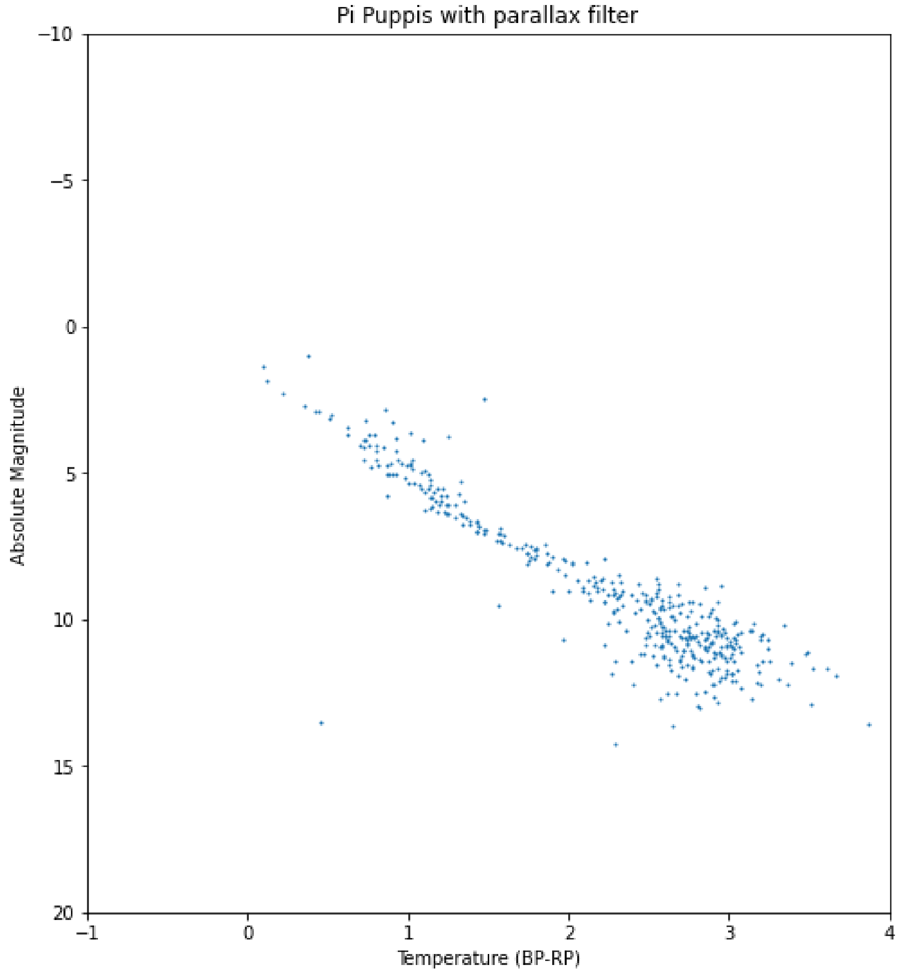

Figure 4.10: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance < 500 pc

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 $< 10\%$.

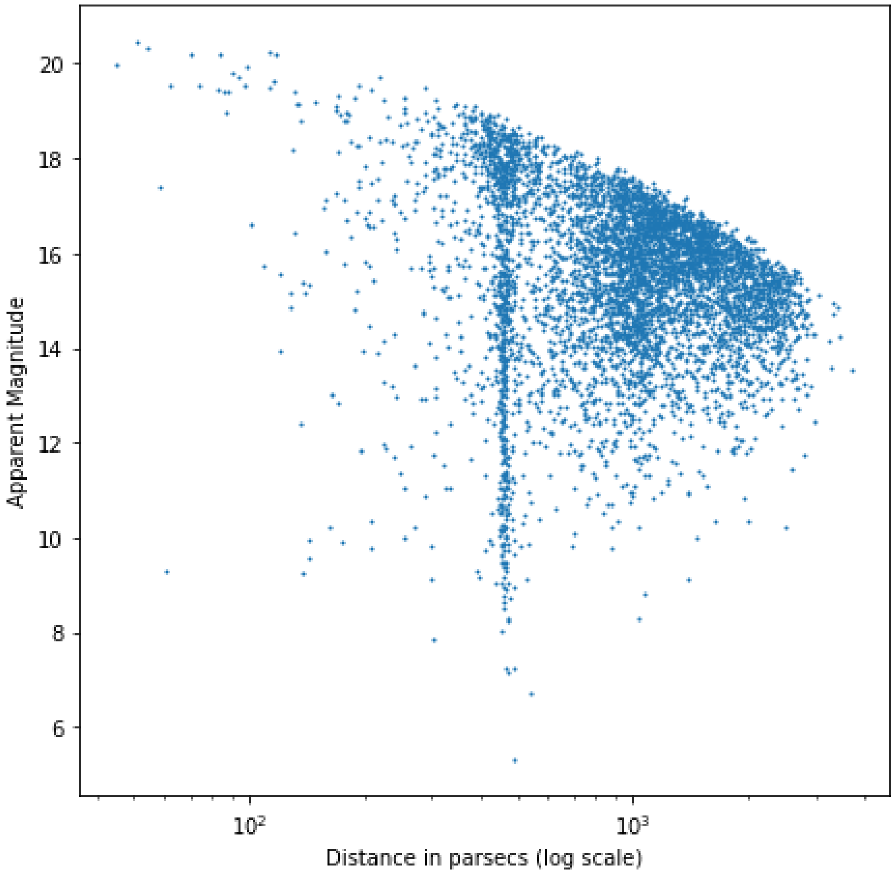

Figure 4.11: Plot of distance and apparent magnitude

.

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.

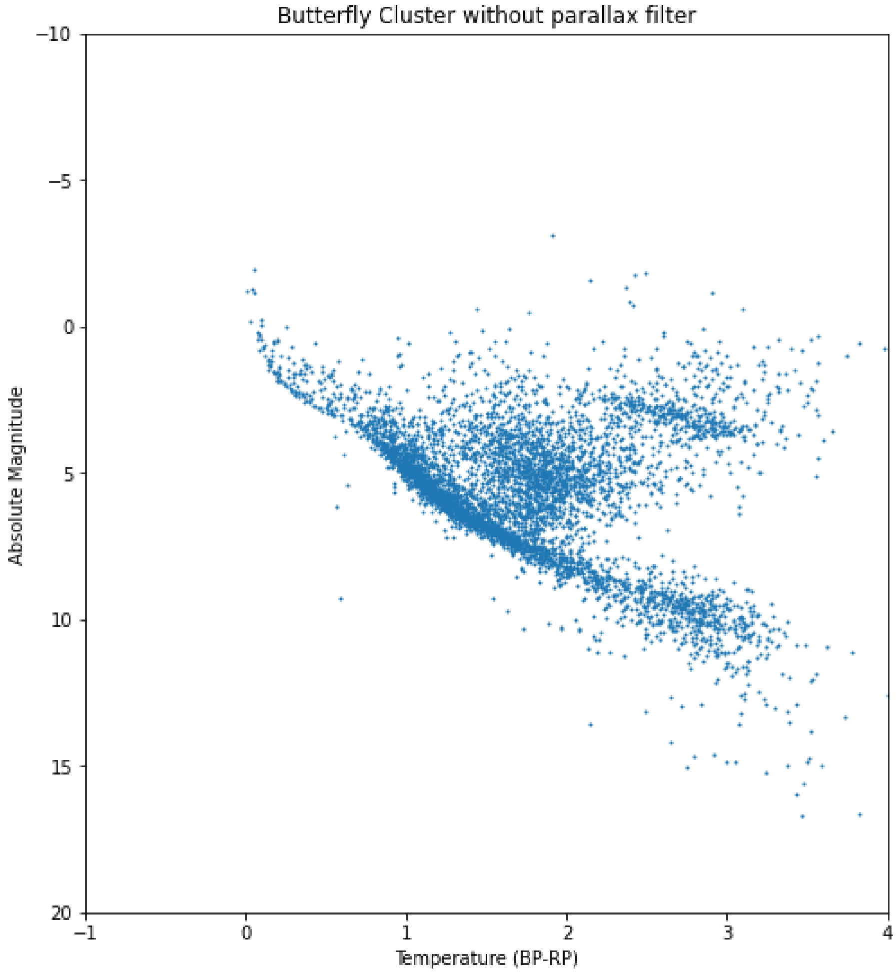

Figure 4.12: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch
.

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.

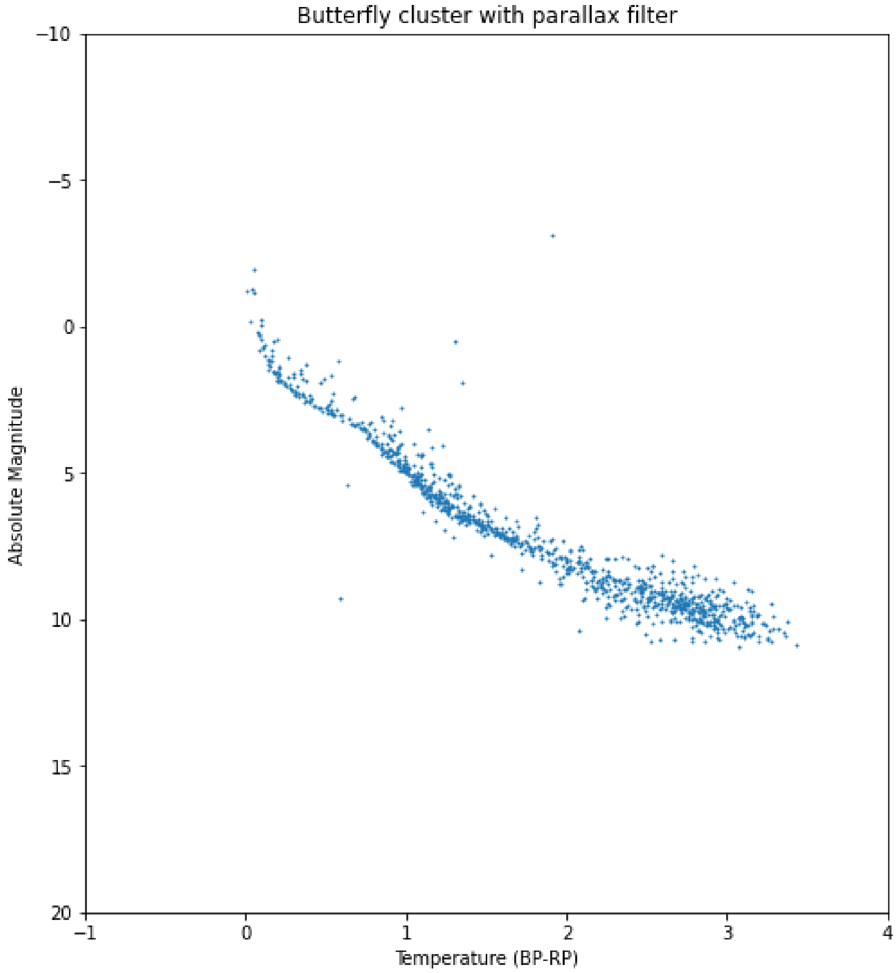

Figure 4.13: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance between 400 and 600 pc

.

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

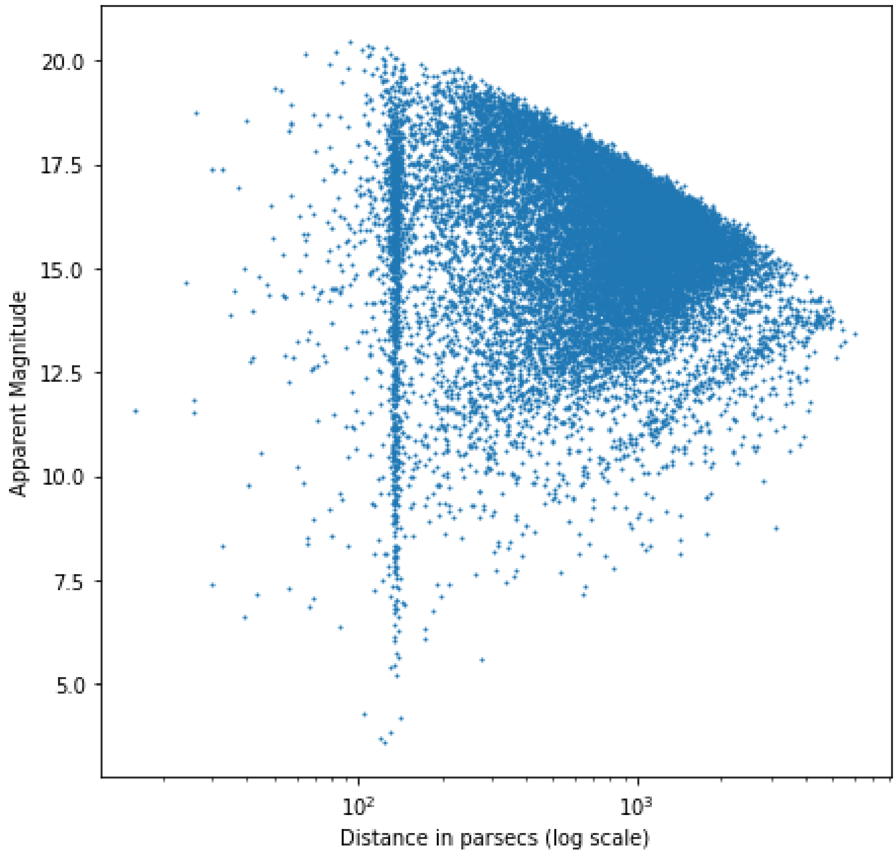

Figure 4.14: Plot of distance and apparent magnitude

.

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

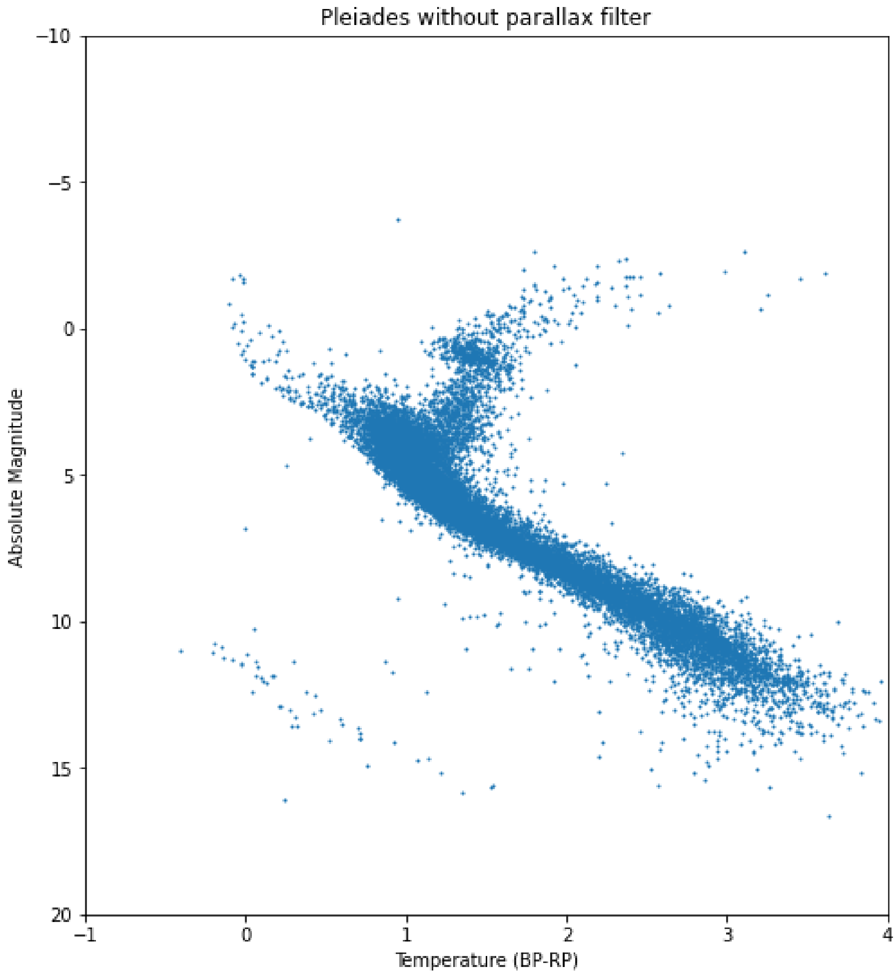

Figure 4.15: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch
.

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

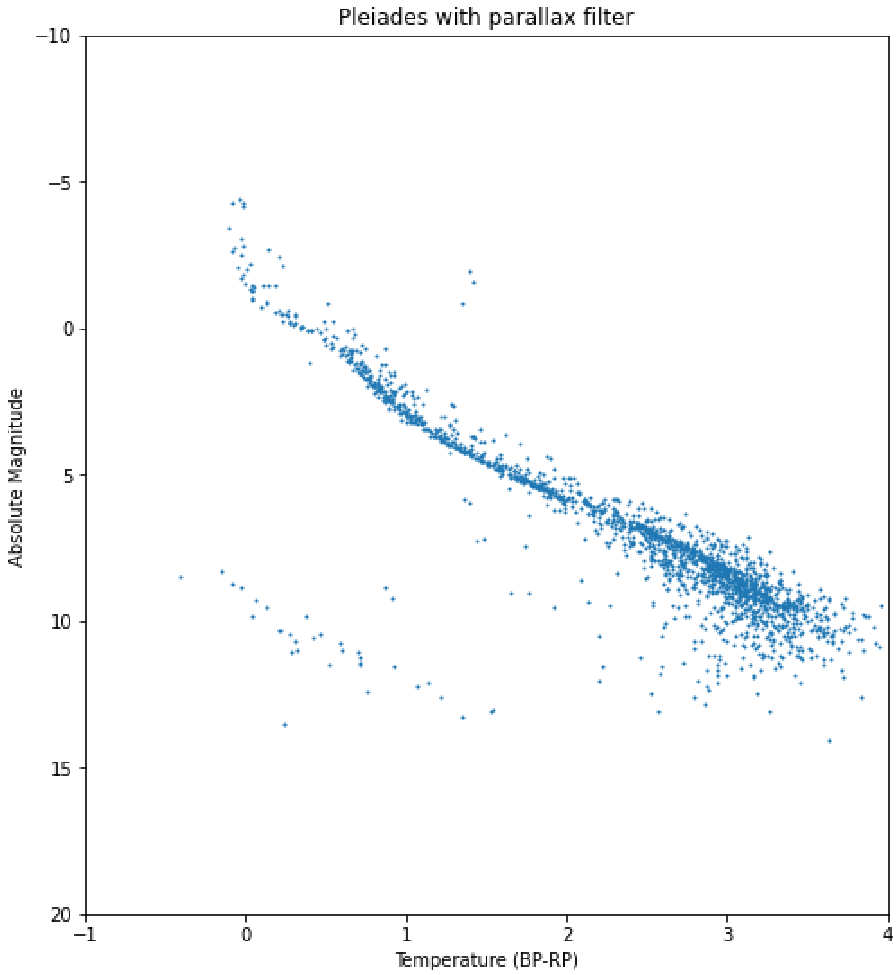

Figure 4.16: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance $<$ 200 pc

.

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.

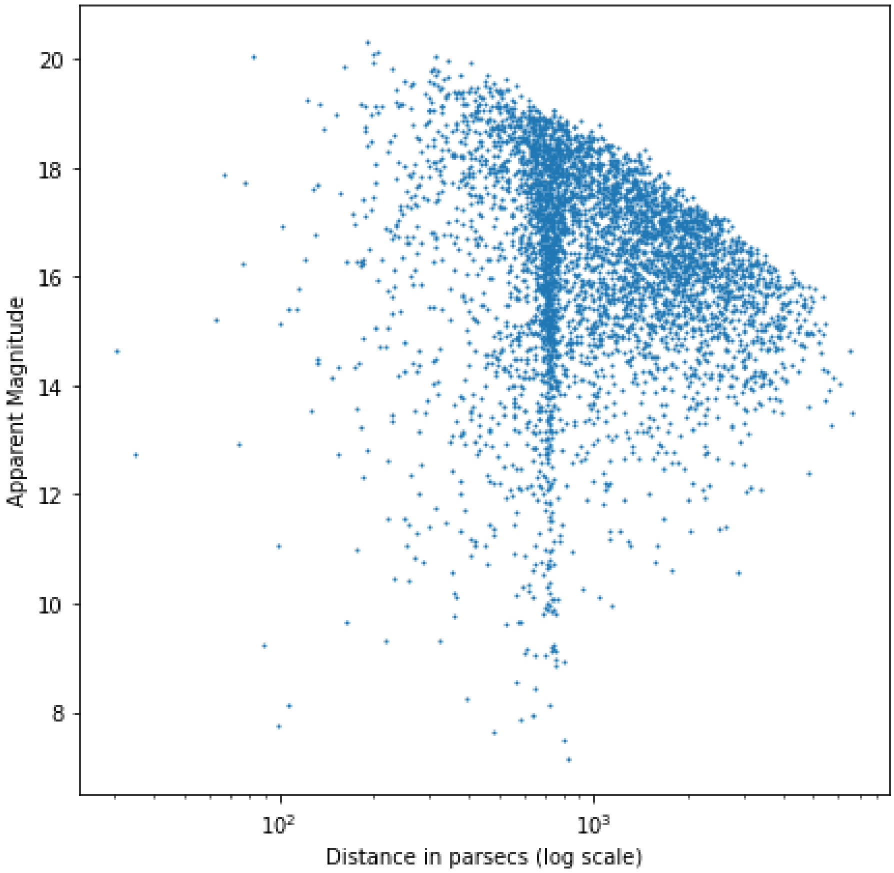

Figure 4.17: Plot of distance and apparent magnitude

.

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

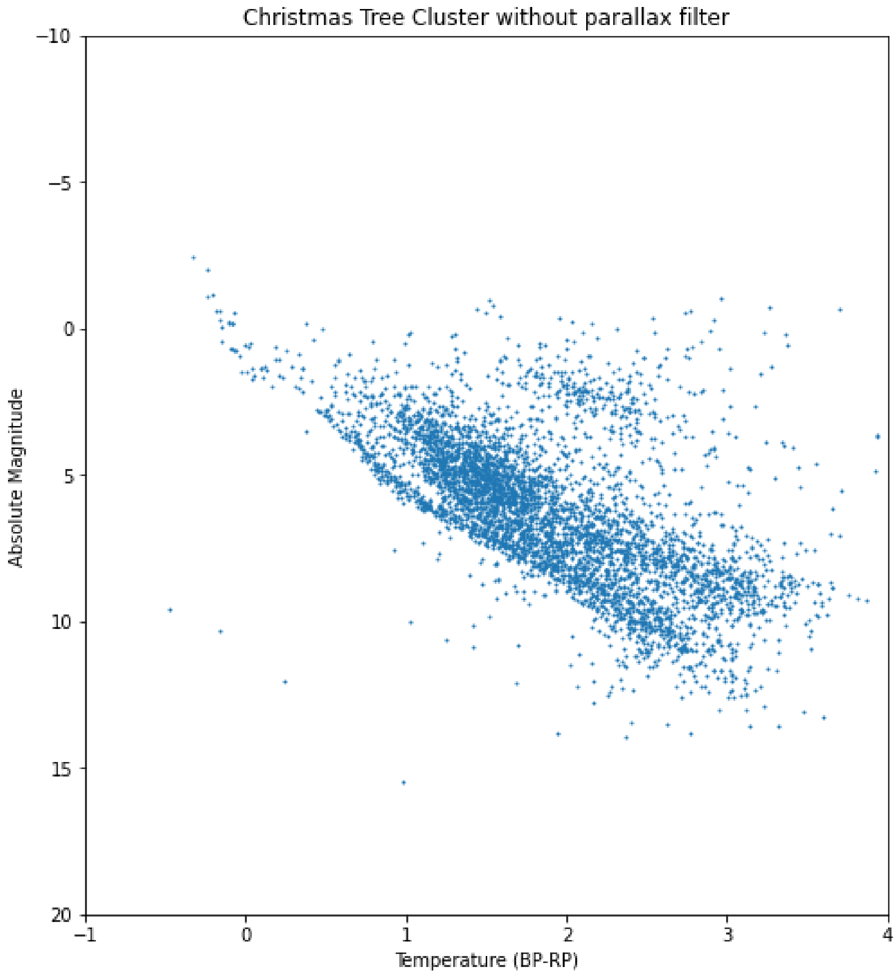

Figure 4.18: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch
.

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

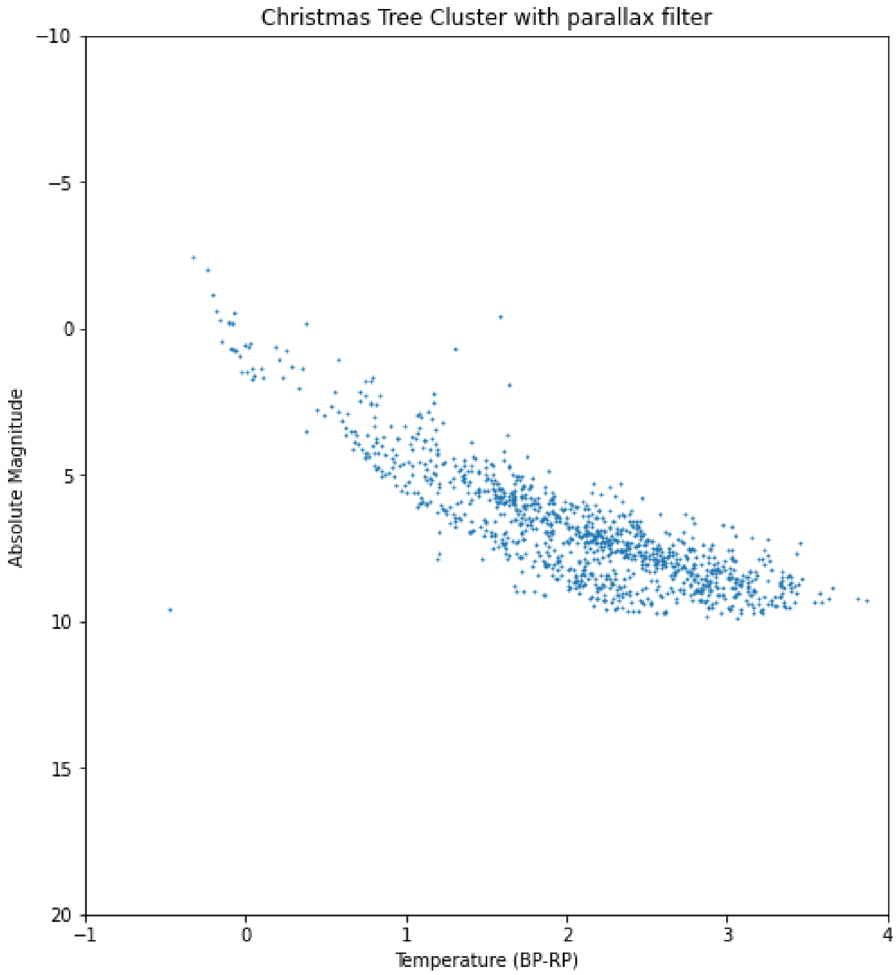

Figure 4.19: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance between 660 and 830 pc

.

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 $\leq 5\%$.

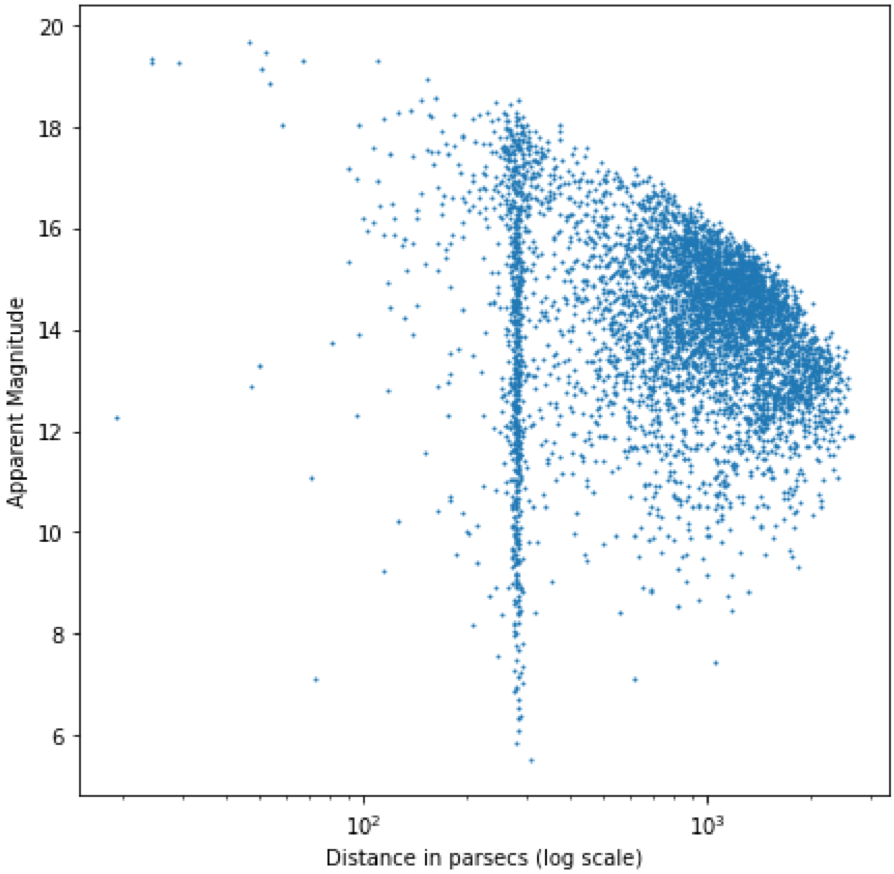

Figure 4.20: Plot of distance and apparent magnitude

.

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

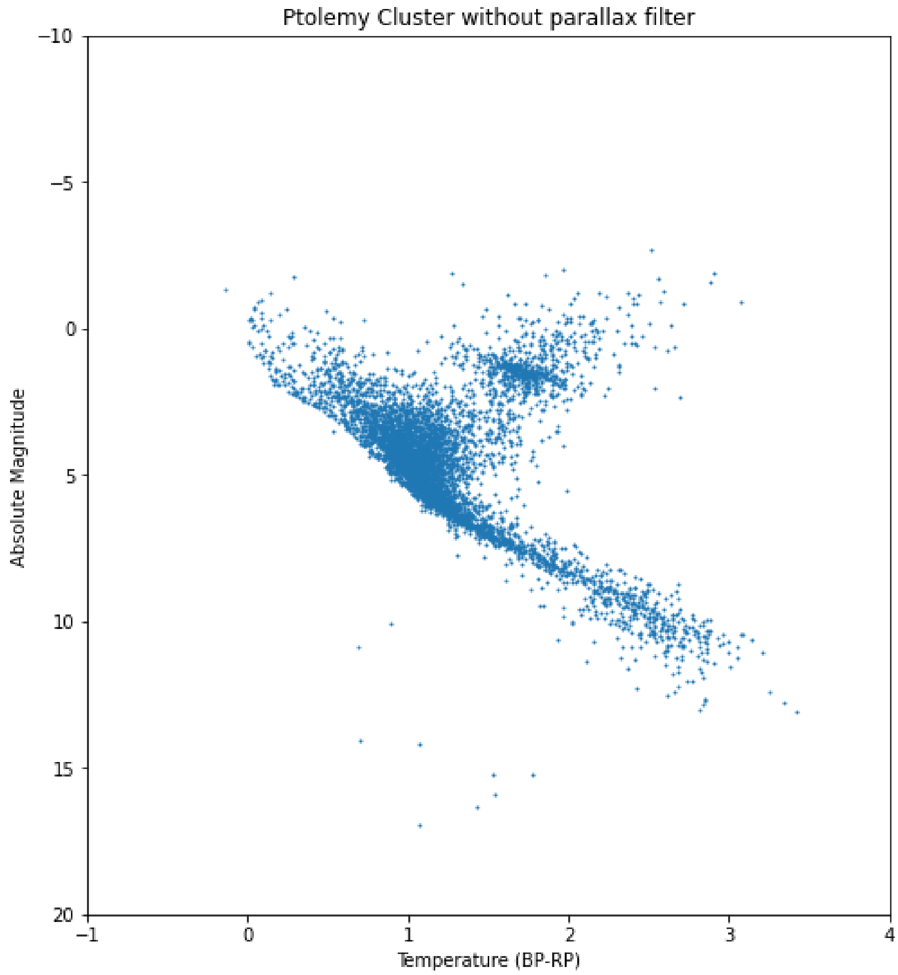

Figure 4.21: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for all stars in that sky patch

.

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

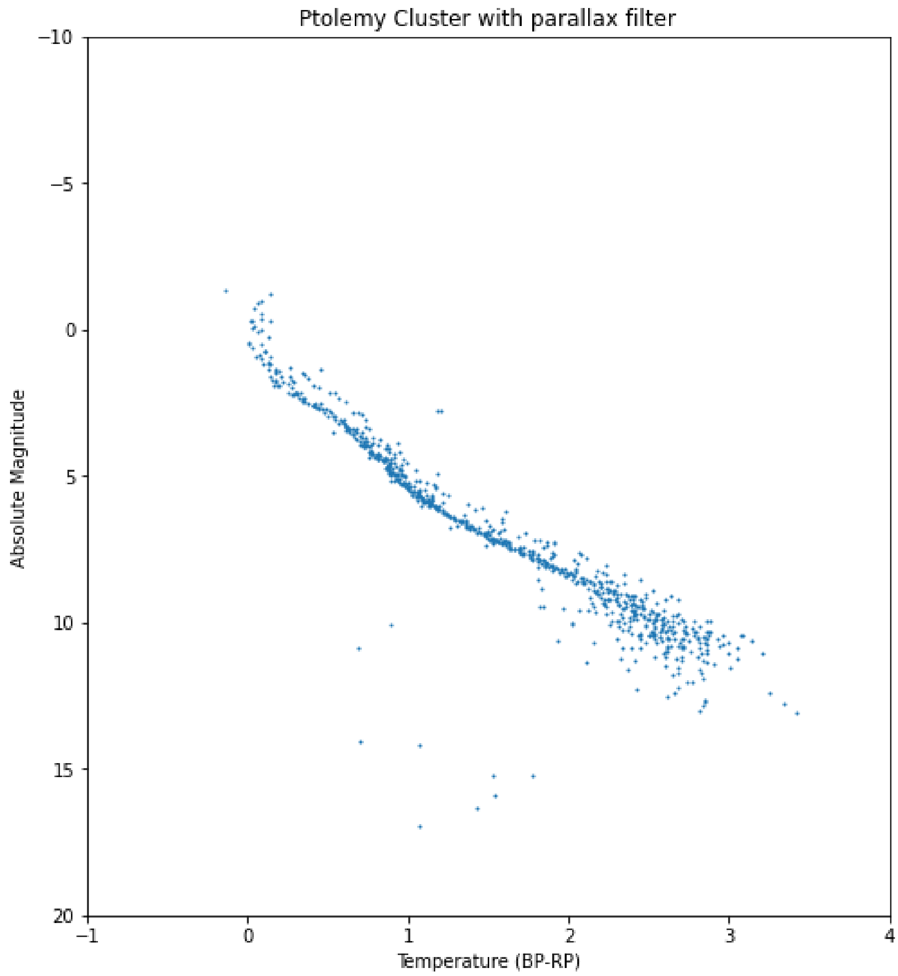

Figure 4.22: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, for only those stars in that sky patch at a distance $<$ 300 pc

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:

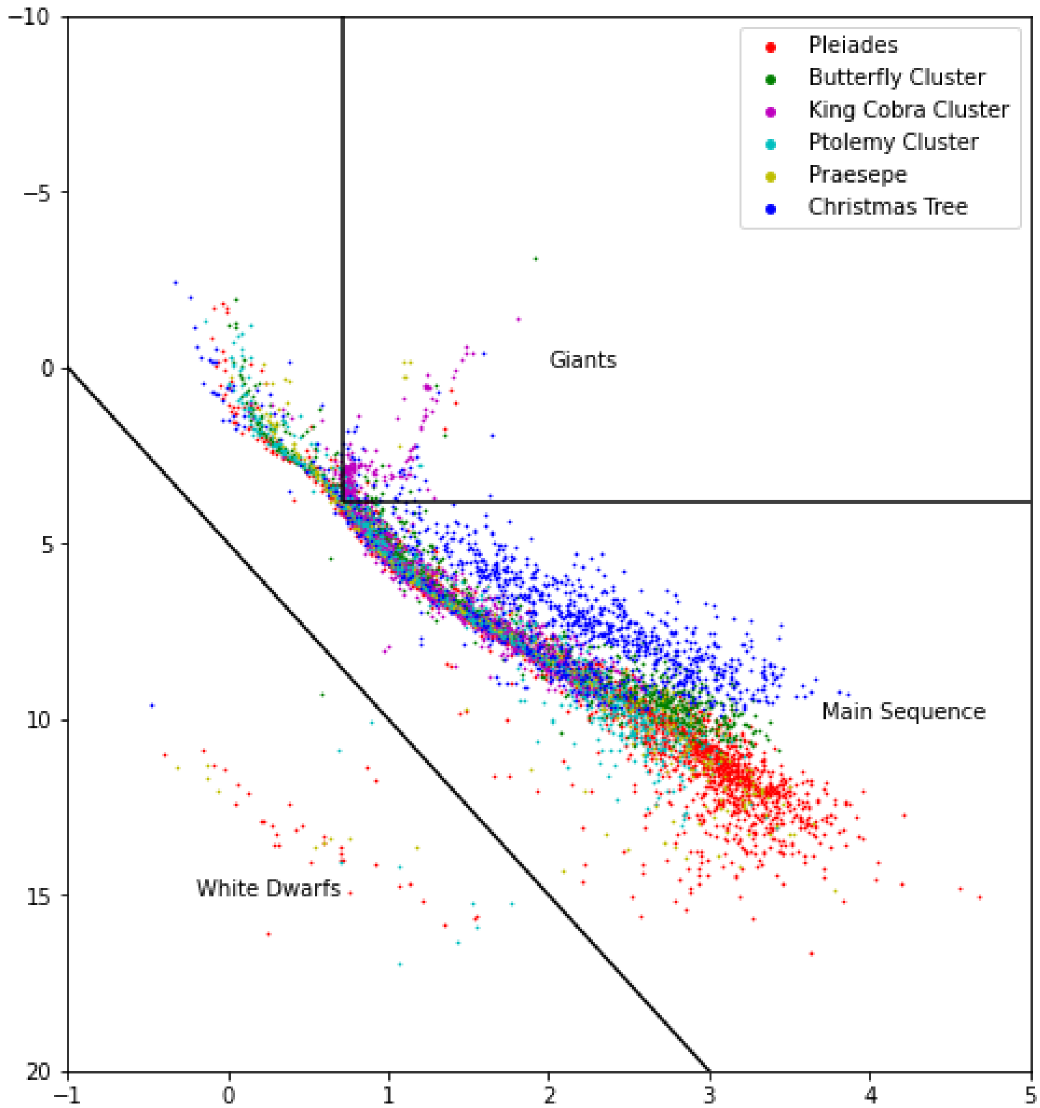

Figure 4.23: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, of our stars with the lines that were chosen to separate them - note that these are only rough boundaries, not precise borders between stages of stars' life cycle

.

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.

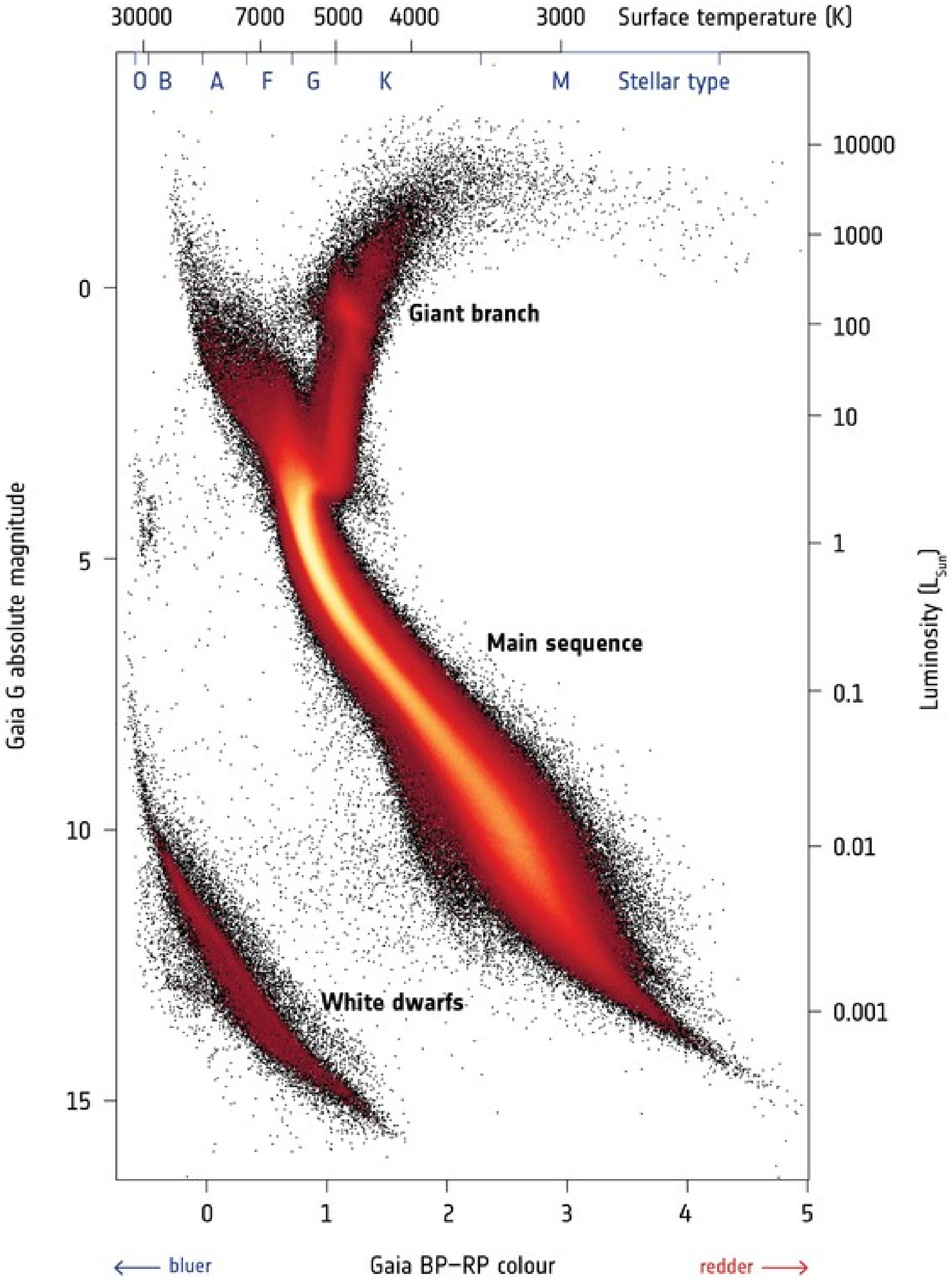

Figure 4.24: HR Diagram for Gaia stars with spectral classification

.

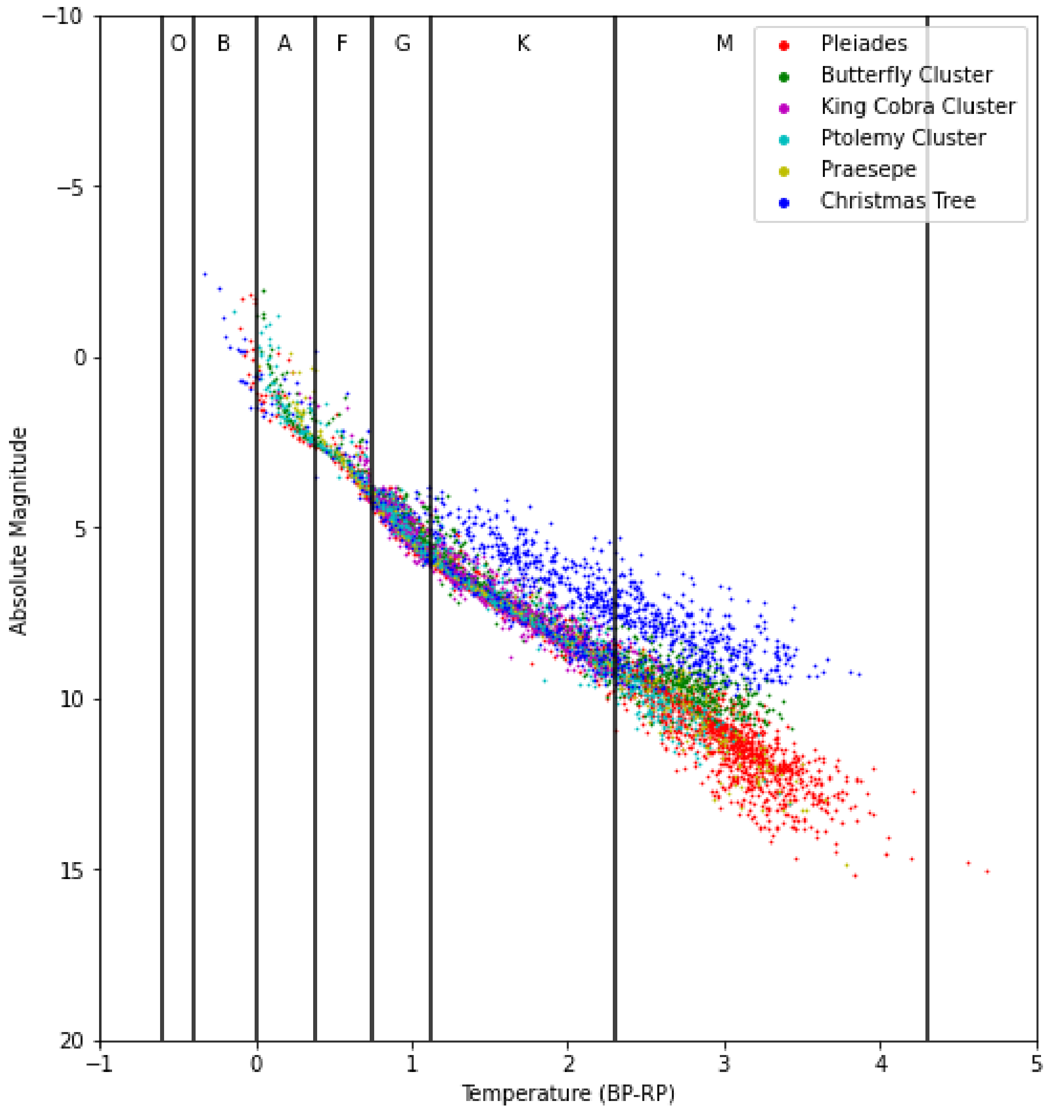

Figure 4.25: Spectral classification on our HR diagram

.

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.

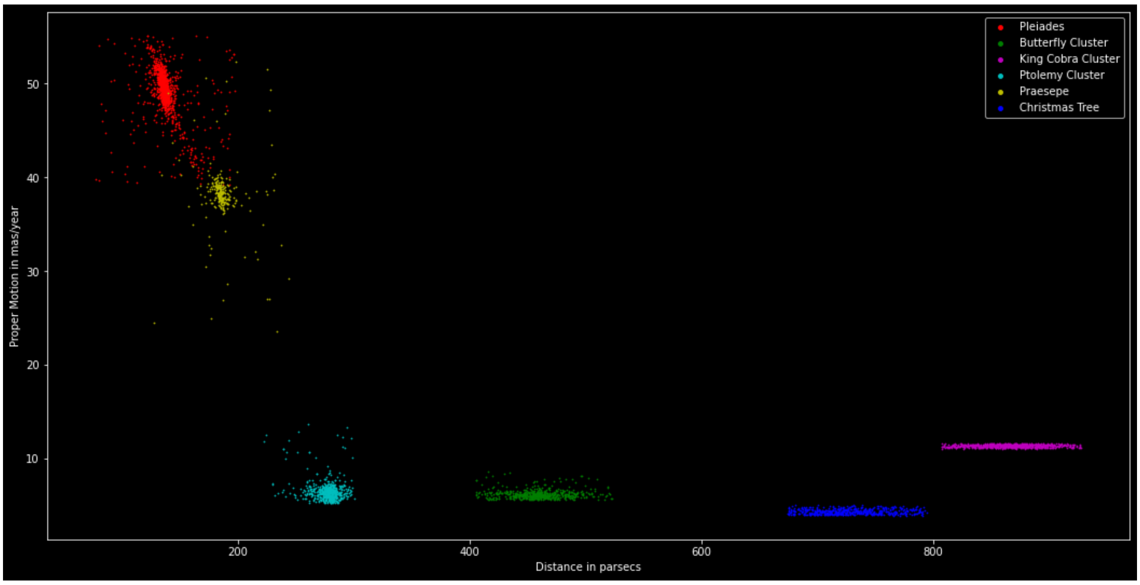

Figure 4.26: Plot of Proper Motion and Distance for each of our clusters. Outliers beyond one standard deviation from the average value are removed

.

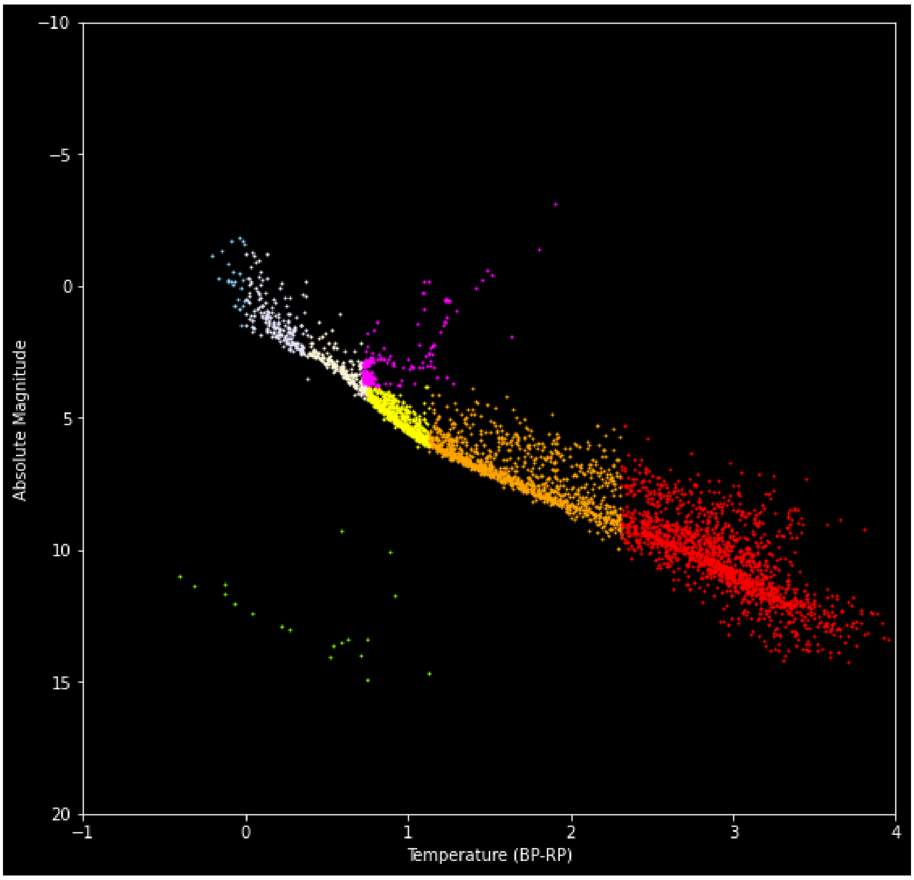

Figure 4.27: Plot of $G_{BP} - G_{RP}$ and absolute magnitude, of stars from the cleaned data

4.2.3 Relating Kinematics to Star type

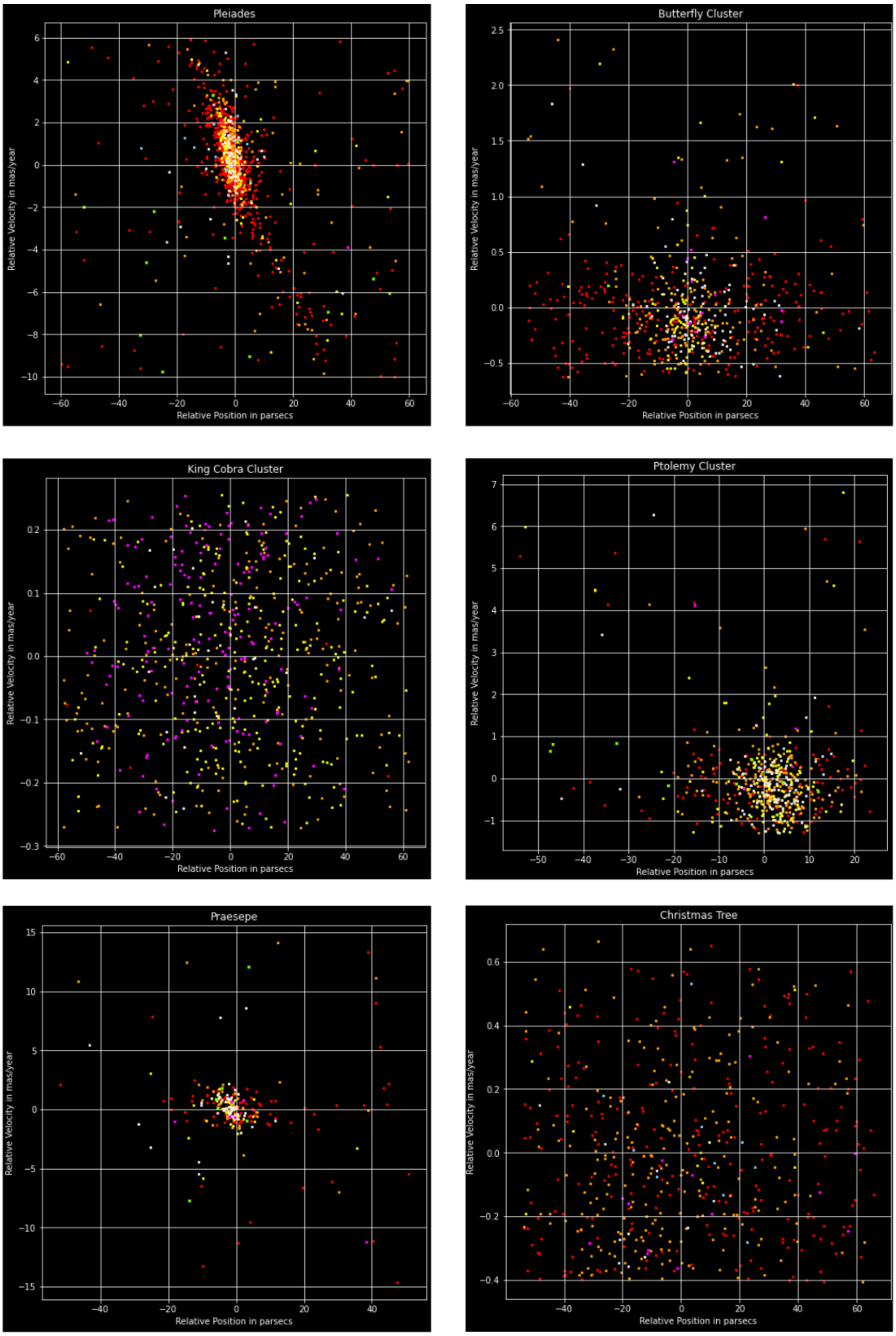

Figure 4.28: Plot of Distance and Proper Motion of all the six Open Clusters

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.