

Hyades Clustering

Charlie Gels, Samantha Hsu, Yash Mehra

Dataset Description

The dataset we have used is the Hipparcos dataset of 2719 stars. The star selections are based on the set of stars that lie between parallax 20 and 25 mas. The dataset consists in a tabular data form with 9 columns.

HIP	Vmag	RA	DE	Plx	pmRA	pmDE	e_Plx	B-V
-----	------	----	----	-----	------	------	-------	-----

1. HIP = Hipparcos star number
2. Vmag = Visual band magnitude. This is an inverted logarithmic measure of brightness
3. RA = Right Ascension (degrees), positional coordinate in the sky equivalent to longitude on the Earth
4. DE = Declination (degrees), positional coordinate in the sky equivalent to latitude on the Earth
5. Plx = Parallax angle (mas = milliarcseconds). $1000/\text{Plx}$ gives the distance in parsecs (pc)
6. pmRA = Proper motion in RA (mas/yr). RA component of the motion of the star across the sky
7. pmDE = Proper motion in DE (mas/yr). DE component of the motion of the star across the sky
8. e_Plx = Measurement error in Plx (mas)
9. B-V = Color index of star (mag)

Description of the Hyades Star Cluster

The Hyades star cluster is an open star cluster in the Hipparcos stars. We assume that it consists of about 100 stars all moving in the same direction, according to Encyclopedia.com, and are positioned around RA=67 degrees and DE=16 degrees and have mean parallax angles of 22 mas.

The analytical question(s) studied

1. Can we accurately plot the Hipparcos Dataset on an Hertzs-Russell Diagram?

We managed to successfully plot the Hipparcos Dataset on a H-R Diagram, shown below. Because our dataset wasn't classified, we manually classified and plotted the stars by spectrum.

2. How accurately can we find the Hyades star cluster from the Hipparcos dataset using different clustering techniques?

Using Sci-Kit learn we ran 3 different clustering techniques.

1. Kmeans
2. Spectral
3. DBSCAN

For each of the clustering algorithms we ran different k values in order to find the optimal k that would provide the closest accuracy to with the Hyades dataset.

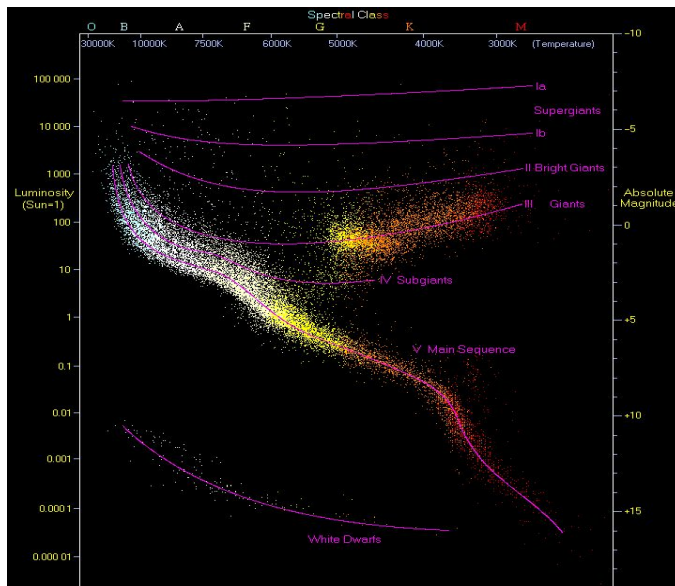
Below you will see several visualizations for each of these questions and clustering algorithms with a report on their accuracy with the Hyades cluster and finally a report on the best k-values seen and the best clustering algorithm.

Results of the use of the methods visualized where possible

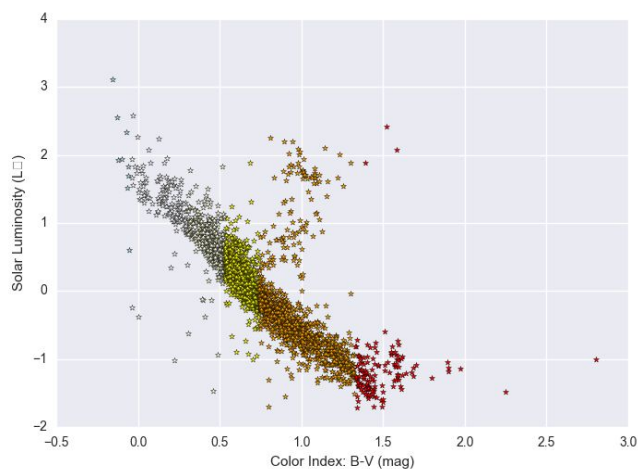
1. Hipparcos Dataset

We manually classified the Hipparcos cluster by spectrum as shown below:

And plotted the classified Hipparcos:

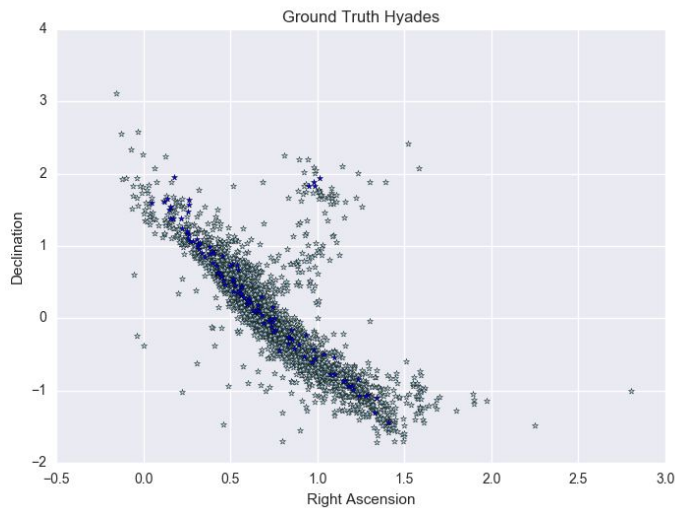


Here is the actual (ground truth) HR diagram for this particular dataset. The X-axis and Y-axis are the same as our visualization below as well as our incorporation of the spectral classes. We added a classification for all spectral classes that fell into a certain range as well. As you can see below our H-R diagram color classification matches quite well with the ground truth figure.



2. Finding the Hyades cluster from the Hipparcos

Using multiple clustering algorithms, we clustered the Hyades star cluster among the Hipparcos dataset and computed the accuracy against the ground truth:



The blue star in the figure to the left represent the data values for the Hyades Cluster and the remainder are stars in the Hipparcos Data set that wasn't included.

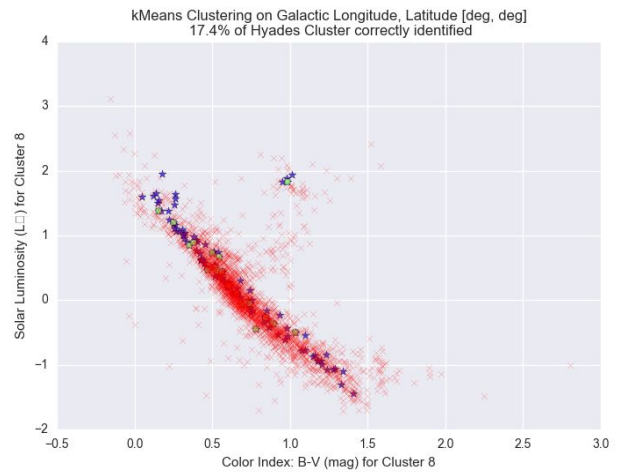
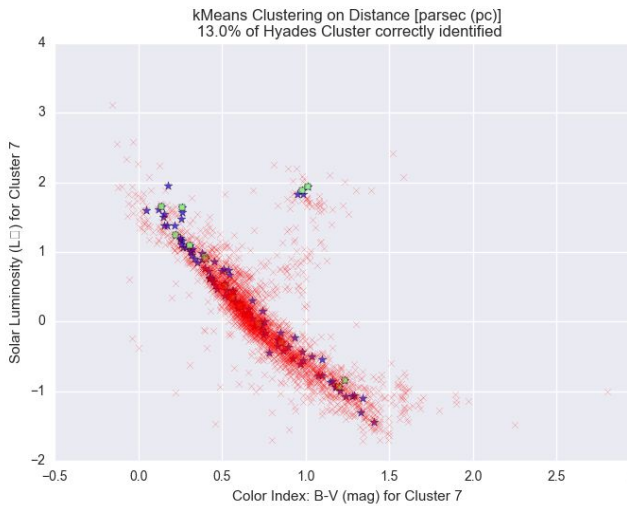
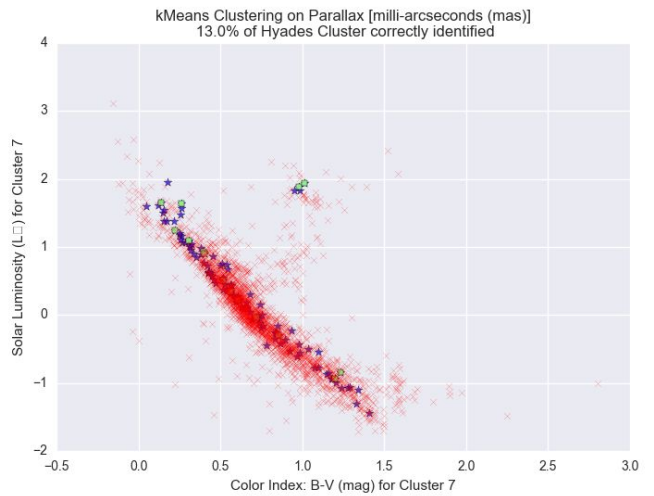
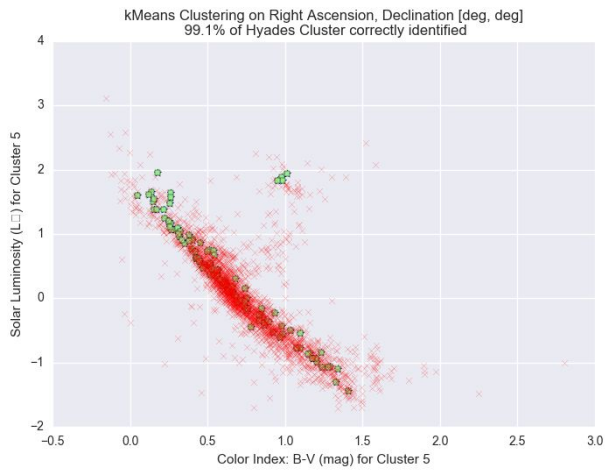
We ran a study for each criteria to run clustering on. We wanted to know which inputs to the clustering algorithms would yield the best results. The table below shows the criteria we ran each algorithm on, the best k-value and the accuracy.

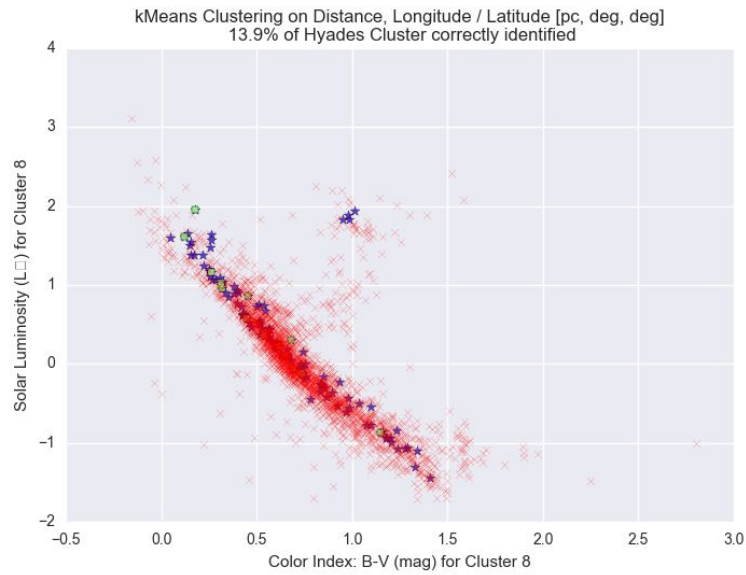
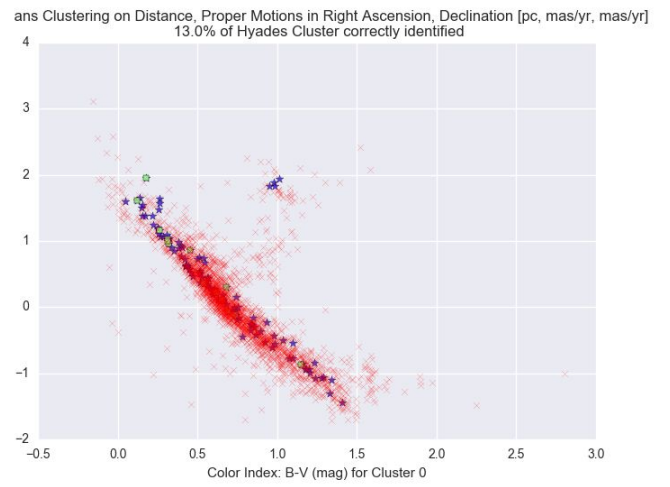
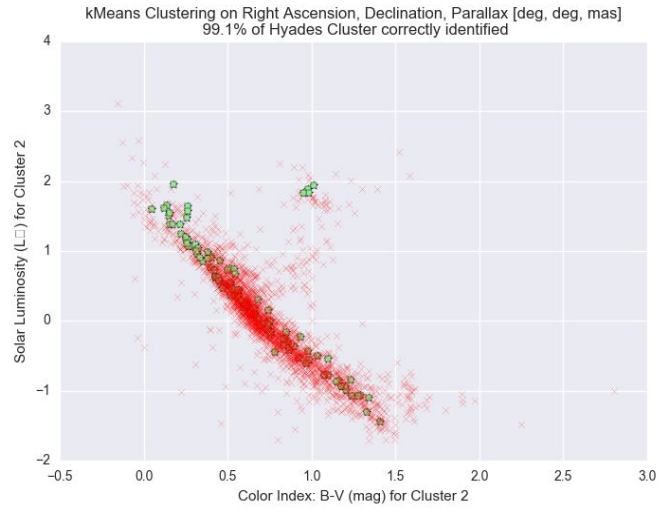
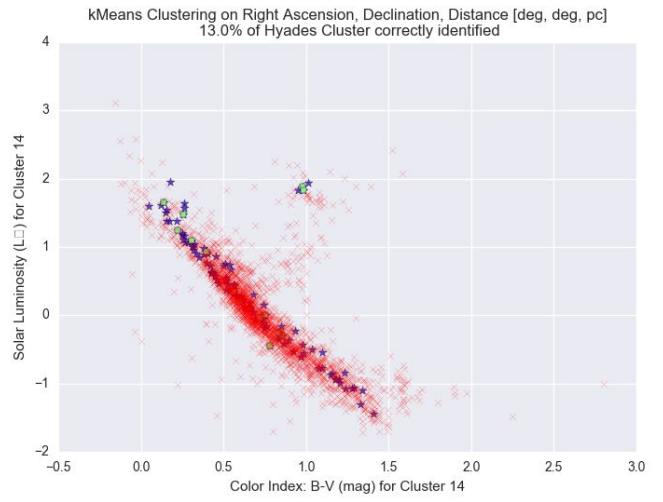
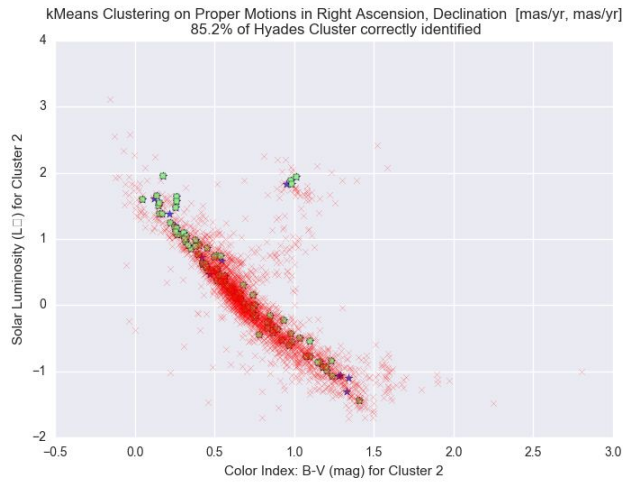
KMEANS

No.	Study	K-Value	Accuracy
0	Ra, Dec	11	0.9913
1	Parallax	14	0.1478
2	Distance	15	0.1391
3	Longitude / Latitude	12	0.1739
4	Proper Motions	11	0.8869
5	Ra, Dec, Distance	14	0.1304
6	Ra, Dec, Parallax	11	0.9913

7	Distance, Longitude, Latitude	14	0.1304
8	Distance, Proper Motions	14	0.1304

Graphs for each study





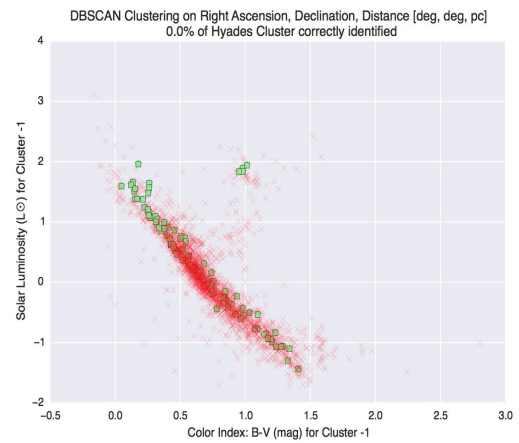
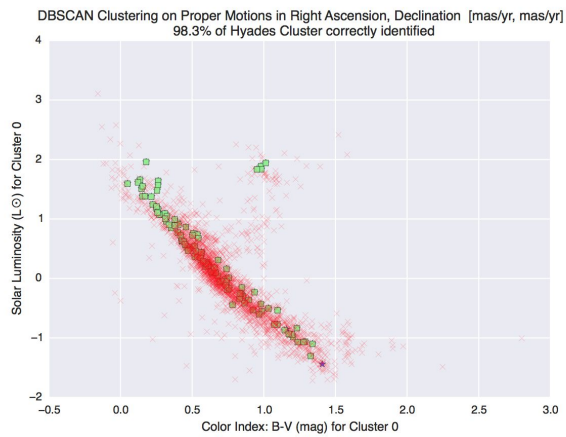
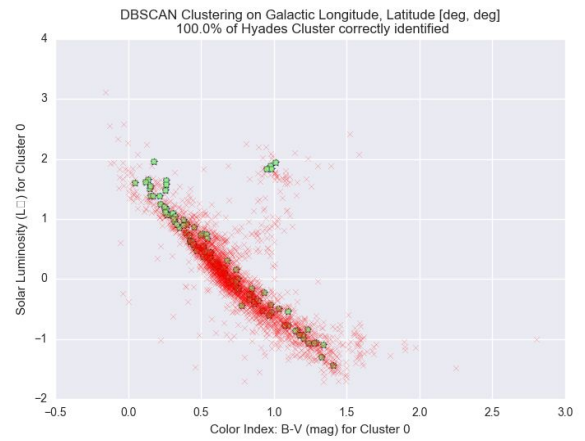
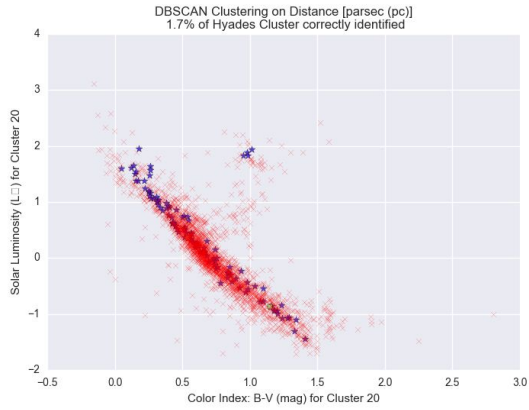
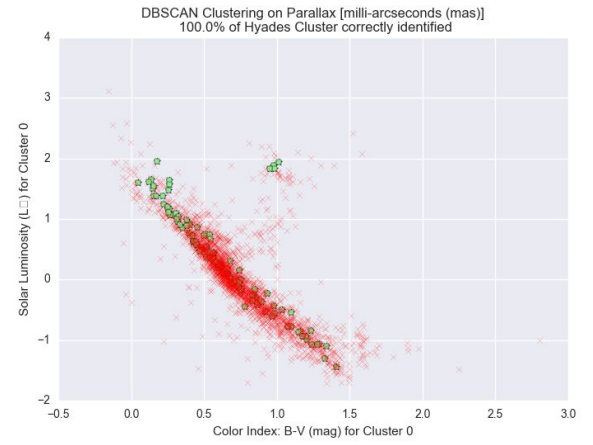
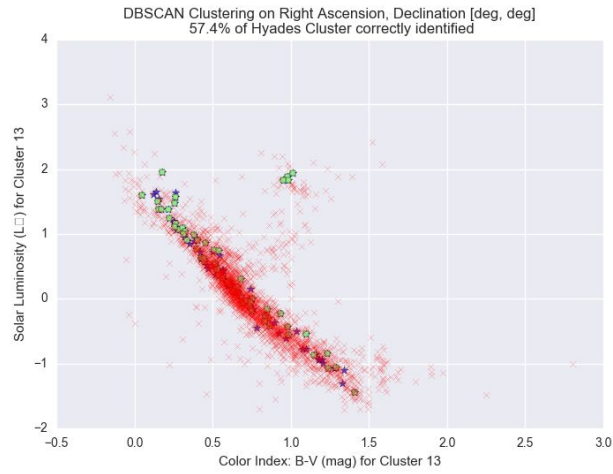
Conclusion Kmeans

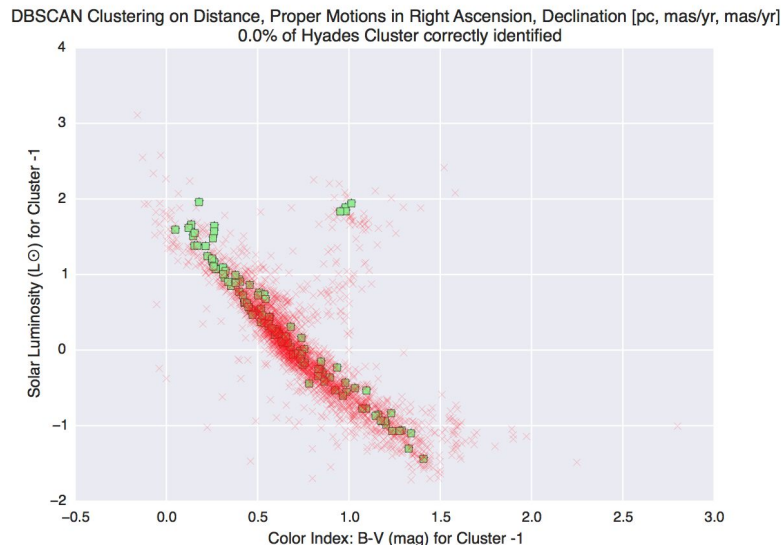
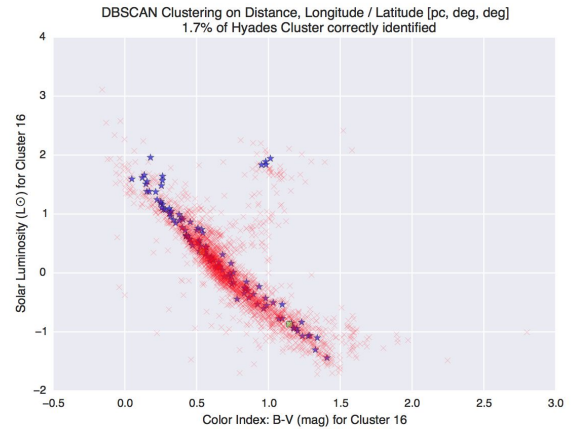
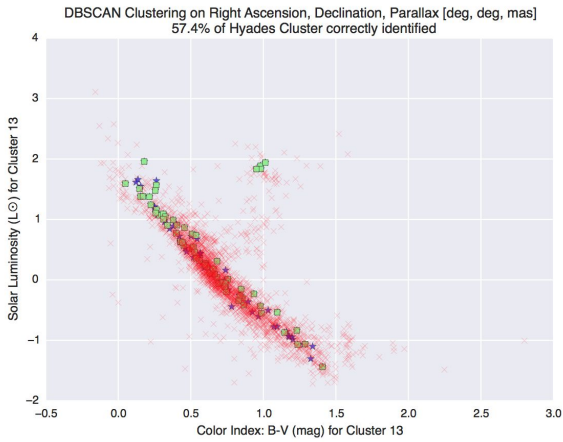
From our study above you can see there are a few studies that performed exceptionally well and a few that did not. We see that studies 0, 4, 6 were by far the most accurate with 0.9913, 0.8869 and 0.9913 respectively. Notably, KMeans clustering by Right Ascension and Declination (Study 0) and clustering by Right Ascension, Declination, and Parallax(Study 6) managed to identify 91% of hyades members. The reason for the accuracy being so high is that these studies were clustering along right ascension and declination where the Hyades is tightly clustered near right ascension +67, declination +16 degrees. Furthermore, we see similar results when adding parallax to the clustering matrix because the Hyades cluster has a mean parallax of 22 milliarcseconds. Study 4 clusters on the proper motions in both right ascension and declination; its accuracy is expected as well since it makes sense that a cluster of stars collectively moves in a similar direction in space.

DBSCAN

No.	Study	Epsilon	Accuracy
0	Ra, Dec	1.179	0.574
1	Parallax	1.179	1.00
2	Distance	1.179	0.017
3	Longitude / Latitude	1.179	1.00
4	Proper Motions	1.179	0.983
5	Ra, Dec, Distance	1.179	0.00
6	Ra, Dec, Parallax	1.179	0.574
7	Distance, Longitude, Latitude	1.179	0.017
8	Distance, Proper Motions	1.179	0.00

Graphs for each study



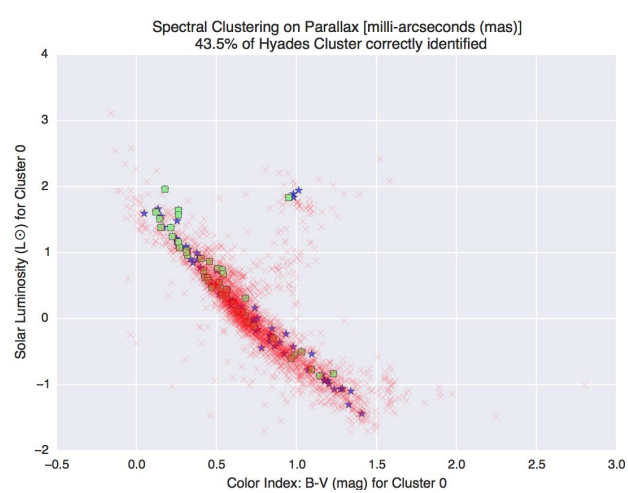
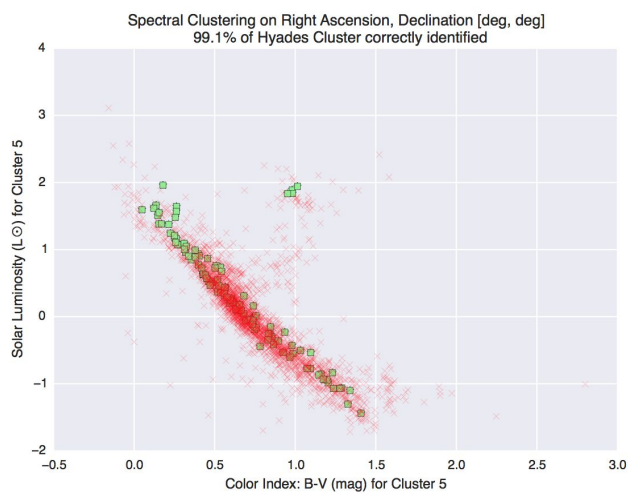


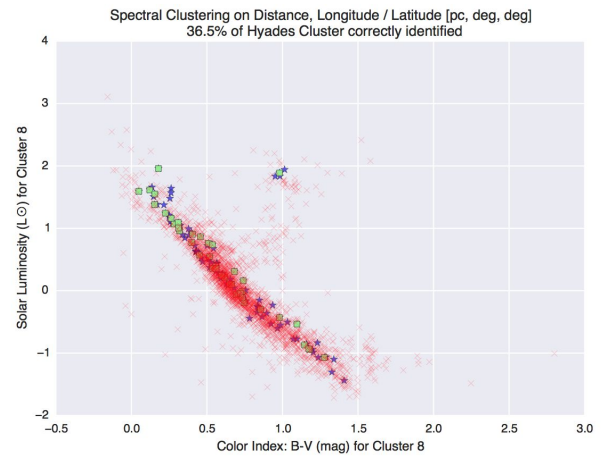
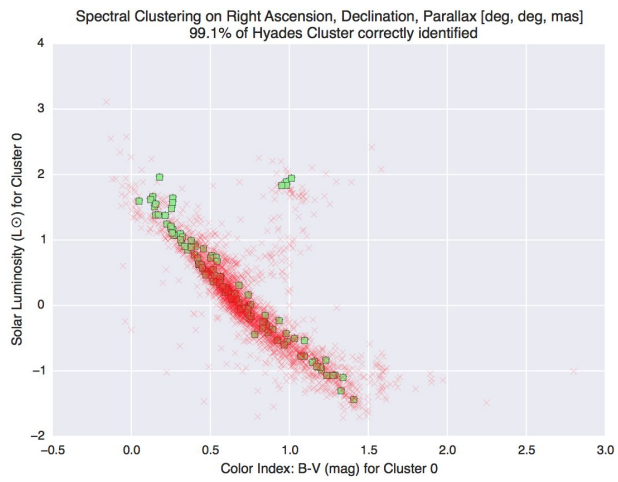
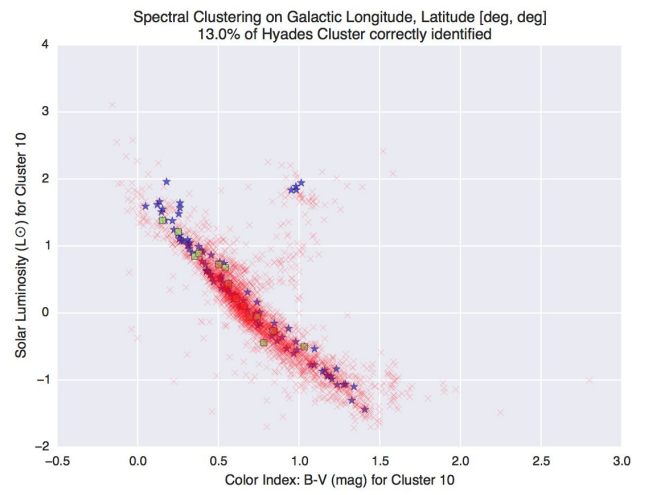
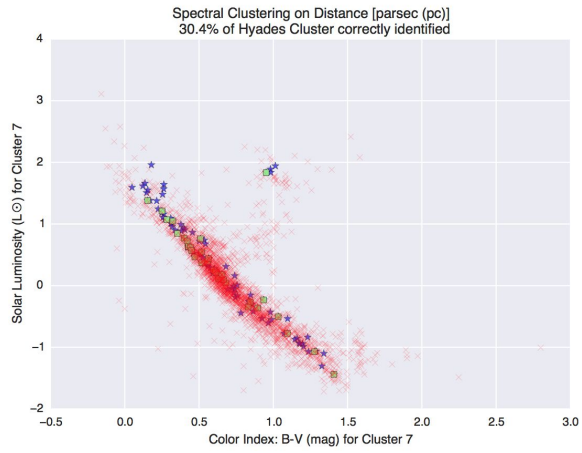
Conclusions on DBSCAN

From our study above you can see the studies were split between performed exceptionally well versus not. Here we see that studies 1, 3, 4 were by far the most accurate with 1.00, 1.00 and 0.983 respectively. There wasn't as much variety when we checked for differences in epsilon. But we got some great results when checking for accuracy. Additionally, it is critical to note the false positives found with DBSCAN. Even though we classified 100% of the Hyades cluster using various techniques, our model predicted many stars to belong in the same clusters as the Hyades, yet these stars are not considered Hyades members.

SPECTRAL

No.	Study	K-Value	Accuracy
0	Ra, Dec	8	0.9913
1	Parallax	10	0.3014
2	Distance	10	0.1391
3	Longitude / Latitude	14	0.9913
4	Proper Motions	--	--
5	Ra, Dec, Distance	--	--
6	Ra, Dec, Parallax	8	0.365
7	Distance, Longitude, Latitude	13	0.435
8	Distance, Proper Motions	--	--





Conclusion from Spectral

From our study above you can see that there were also a few studies that performed exceptionally well while a few did not. Here we see that studies 0 and 3 were by far the most accurate with 99.13 and 99.13 accuracy respectively. There wasn't as much variety when we checked for differences in the number of clusters, k . Again, we saw that Right Ascension and Declination was the best way to identify the Hyades cluster.