I continued my work from last quarter, by first fixing the search criteria for parallax error, which changed the length of the data table from 25,163 rows to 39,853 rows. I then proceeded to transform the proper motions and heliocentric radial velocities into galactocentric coordinates. I was having trouble manipulating matrices in python, so I ended up doing a series of variable transformations by hand. I understand that this most likely leaves a lot of room for errors in the transformations, so I will be double checking my work in the following quarter.

After transforming the velocity coordinates I made several 3D scatter plots, as well as kernel density estimate plots of U, V, and W. Some structure of the data was revealed in these plots, but it was not quite comparable to the plots in Bovy and Hogg's 2009 paper, *The Velocity Distribution of Nearby Stars from Hipparcos Data. I. The Significance of the Moving Groups*. This lack of structure may be due to possible errors in my coordinate transformations.

I then proceeded to apply the k-means clustering algorithm to the data, which initially provided no insight as the data points were too densely packed to identify anything other than uniform clusters. Gwen suggested that my subsamples (cut down to a few thousand rows or so out of fear of more being unmanageable) may be ordered in a way that could lead to the data being so uniform, so I implemented random selection to the subsample, which then provided less uniform plots and clusters.

In the following quarter, I will be rechecking my coordinate transformations, as well as possibly looking into using data from the Gaia DR2. I also will implement Gaussian mixture models for density estimation, to see if this provides more insightful clustering than k-means.