**BIO 4990/6990: Computational Activity 6—Unsupervised Learning**

**Principal Components**

1. How many principal components were identified in the PCA?
2. What percentage of the variance is explained by the first two principal components?
3. How well do the first two principal components separate the different castes and species (discuss each axis separately)?
4. What do the eigenvalues tell you about the importance of each principal component?
5. Is visualizing the third principal component helpful? Why or why not?
6. How are the variable loadings represented in the PCA plots, and how do you interpret them?
7. Do you see any distinct clusters in the plots? If so, describe their characteristics.

**K-means clustering**

1. Which value of k did you choose, and why?
2. How do these clusters relate to the caste and species of ants, and how do they compare to the clusters you identified by eye in question 6?

**Overall**

1. What did you learn about these data through the use of unsupervised learning?
2. Are there other unsupervised learning strategies that you think would be informative? If so, which ones?