c. Discussion of results from the experiments that you run in part 3b above. Which clustering version gives the best results? What seems to be the best number of clusters for each method? --- 3 points

When we look at the three clustering methods, we think that the K-Means method gives us the most insight into the question at hand – “Which of the mushrooms are actually in the unkown class”. Althought K-NN is very accurate and is able to predict and classify the test set well, we are unable to train this on ‘unknown’ so therefore unable to answer the question. Using K-means, we could dive much further to do investigation of the characteristics that stand out within each cluster, whether these characteristics seem to belong to poisonous, edible, or if they seem to stand out as a class of their own. If they do not seem to be part of either ‘poisonous’ or ‘edible’ then we can assume that this cluster may be part of the ‘unknown’ class.

Looking at the data set on a whole, we may think of ways to dive further into the data, to determine the unknown class (given the time and scope of this project these ways would not be feasible) – or even to verify our conclusions from the clustering in K-means. We could possibly take our ‘predicted’ unknown’ cluster observations and label these as ‘unknown’ in the original data set – then use K-NN to classify three classes and see how the model does.

We conclude that the best number of clusters for K-Means seems to be 7 clusters on the 70/30 train/test split. For Hierarchical clustering, we would use 7 clusters as well, this time starting with all variables and cutting the algorithm off when it reaches 7 clusters. For K-NN we see that the accuracy is high for mostly all the number of nearest neighbors. For accuracy and performance with the algorithm, we would like to use more neighbors, so we will say k=5 works the best. K=5 is a number of neighbors that is frequently used as a baseline in this methodology and seems to give us 99.5% accuracy on the 70/30 train/test set.