

Lesson 20 (Clustering Iris) Most of the commands need to answer the questions in this lesson can be found in the Jupyter notebook `K-Means.ipynb`. The commands below are useful for labeling each species of iris with an integer.

```
df['species'] = df.species.astype('category')
df['species code'] = df.species.cat.codes
df1.plot.scatter(x='asymmetry',y='grove',c=df1.clusters.values,s=marker_size,alpha=marker_alpha)
```

- (a) Create a data frame of the iris data set. Use Sci-Kit Learn's K-Means algorithm to cluster the sepal length and sepal width measurements into 3 clusters. Add a column to the iris data frame containing the cluster label (0, 1 or 2) generated by K-Means clustering.
 - (b) Plot two scatter plots of sepal length vs sepal width. Color the first scatter plot by species and the second by the cluster label determined by K-Means clustering. Does K-Means do a reasonable job of clustering the iris data? Do the K-Means clusters correspond to iris species? Explain.
-