Recall that almost **all**of our real world models are supervised. That means we are going to point our model at at dataset.

In the exercises to follow we will use a **famous**data set called the Iris Dataset

The data we will use is a very simple flower database known as the Iris dataset.

The Iris dataset is composed of 50 samples from three different species of Iris. Each sample contains four features. (What's a feature again?) The length and width of the sepals, the length and width of the pedals.

We have 150 observations of the iris flower specifying some measurements: sepal length, sepal width, petal length and petal width together with its subtype: Iris setosa, Iris versicolor, Iris virginica.

This data is stored in the .data member, which is a (n\_samples, n\_features) array.

The class of each observation is stored in the .target attribute of the dataset. This is an integer 1D array of length n\_samples.

To use this dataset with the SciKit-Learn, we transform each 8x8 image into a vector of length 64.