# dataMining02-data_exploration-iris

October 28, 2019

## 1 From UCI Machine Learning Repository

### 1.1 Iris Dataset

This data file does not have a header with column names.

- Look at the ".names" text file in the Data Folder, read (visually) the column names and store them in a list
- the url is 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
- read the file with `read_csv` using also the `names` parameter
- show the head of the file, just for a quick inspection

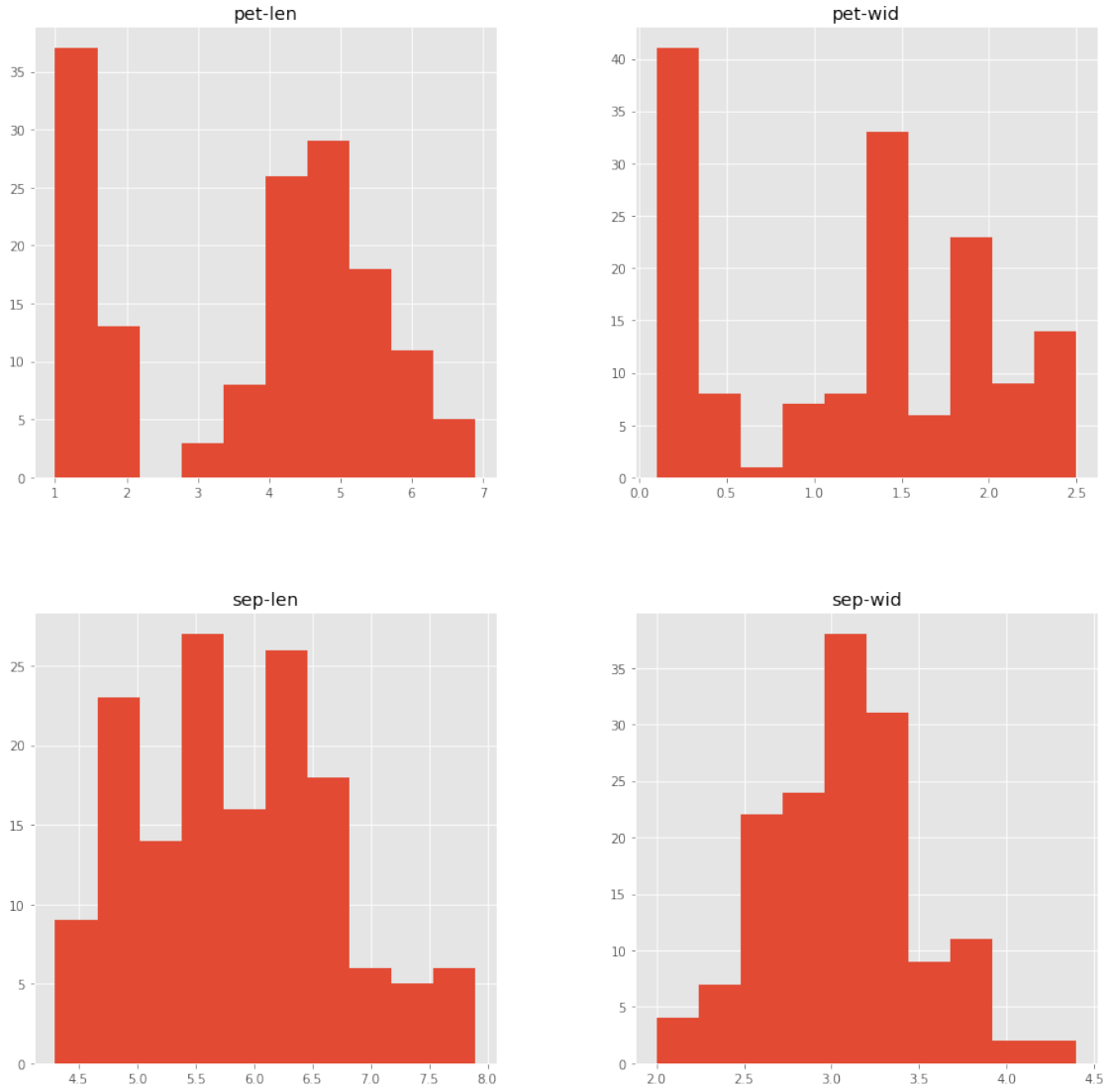
Notice: Iris is also available directly with the datasets package, together with some other *toy* datasets used as examples

Out[9]:

| | sep-len | sep-wid | pet-len | pet-wid | class |
|---|---|---|---|---|---|
| 0 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| 1 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
| 2 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 3 | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| 4 | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa |

## 2 Print histogram of numeric values

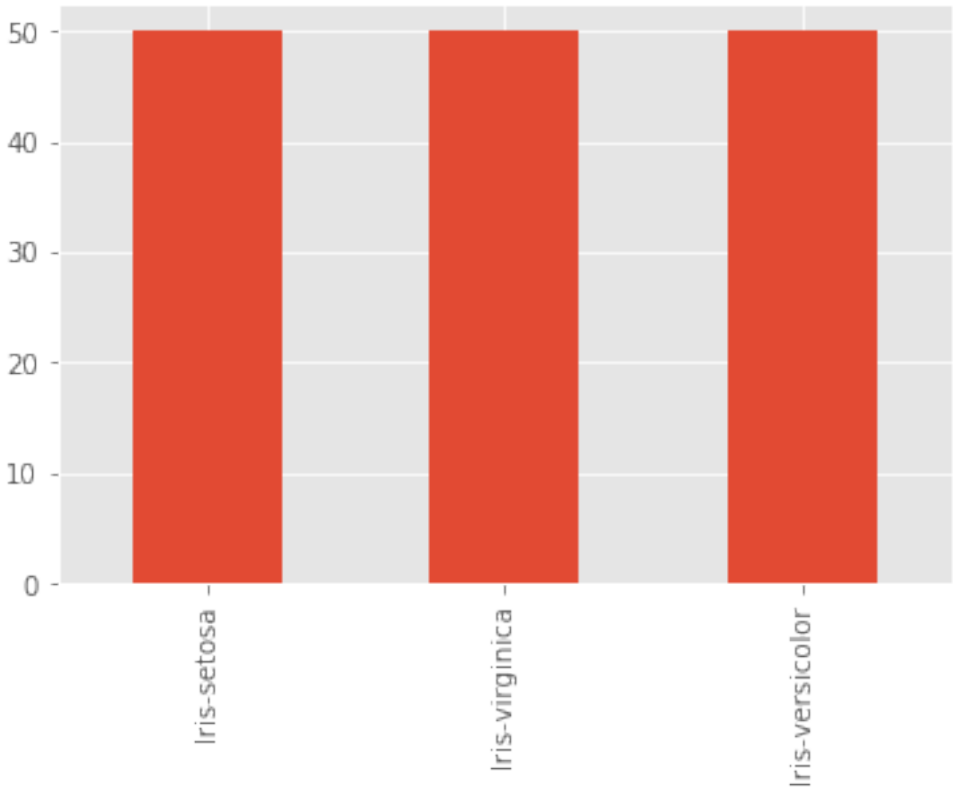
Use the `hist` method again




### 2.0.1 Print histogram of frequencies for the class value

Use the `value_count` method on `class` column, then plot the result with `kind = 'bar'`

Out[11]: `<matplotlib.axes._subplots.AxesSubplot at 0x228c9c15da0>`


50
40
30
20
10
0
Iris-setosa
Iris-virginica
Iris-versicolor