

Principles of Computer Science II

Working with Pandas

Ioannis Chatzigiannakis

Sapienza University of Rome

Lecture 11



Introduction to CSV

- ▶ CSV : Comma separated values
- ▶ Other separators are supported
- ▶ Module CSV provides useful functions to handle structured text files

```
5.1,3.5,1.4,0.2,setosa
4.9,3.0,1.4,0.2,setosa
4.7,3.2,1.3,0.2,setosa
4.6,3.1,1.5,0.2,setosa
5.0,3.6,1.4,0.2,setosa
5.4,3.9,1.7,0.4,setosa
```



Import File

Download file:

<http://aima.cs.berkeley.edu/data/iris.csv>

Read dataset in CSV format

```
import pandas as pd
dataset = pd.read_csv('iris.csv', header='infer')
```

Recall the manual page for details on the different parameters used:

https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_csv.html



Iris Flower Dataset

- ▶ Study of 150 Iris Plants
- ▶ 3 different types: Setosa, Versicolour, Virginica
- ▶ Each plant sampled for
 1. Sepal Length
 2. Sepal Width
 3. Petal Length
 4. Petal Width



Exercise – Scatter plots

- ▶ Create Scatter plots for all iris species:
 1. Sepal Length vs Sepal Width
 2. Sepal Length vs Petal Length
 3. Sepal Length vs Petal Width
 4. Sepal Width vs Petal Length
 5. Sepal Width vs Petal Width
 6. Petal Length vs Petal Width



Exercise – Histograms

- ▶ For each iris species create Histograms:
 1. Sepal Length
 2. Sepal Width
 3. Petal Length
 4. Petal Width



Exercise – 3D Plots

- ▶ For each iris species create 3D plots:
 1. Sepal Length vs Sepal Width vs Petal Length
 2. Sepal Length vs Petal Length vs Petal Width
 3. Sepal Width vs Petal Length vs Petal Width

