

Họ tên: Nguyễn Hữu Hiệu  
MSSV: 20520506

## IRIS

### 1. Code book

Thông tin	Nội dung
Tên bộ dữ liệu	Iris Plant Database
Nguồn thu thập	Fisher, R.A. "The use of multiple measurements in taxonomic problems" (1936)
Số thuộc tính	5
Số mẫu dữ liệu	150
Thông tin các thuộc tính	1. sepal length in cm (float) 2. sepal width in cm (float) 3. petal length in cm (float) 4. petal width in cm (float) 5. class: (int) -- Iris Setosa -- Iris Versicolour -- Iris Virginica
Thông tin tác giả	R.A.Fisher

### 2. Raw data

Bao gồm **bezdekIris.data**, **Index**, **iris.data** và **iris.names**.

### 3. Tidy data

Tidy data được lưu lại thành file **iris.csv**

### 4. Instruction list

```
import pandas as pd

# Insert the directory path in here
path = 'iris.data'

col_name = ["sepal_length", "sepal_width", "petal_length", "petal_width"]

# Add column names into dataset
df = pd.read_csv(path, names=col_name)

print('Task finished!')

df.to_csv('iris.csv', index=False)
```