

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

THU THẬP VÀ TIỀN XỬ LÝ DỮ LIỆU
LAB2

CODE BOOK
IRIS PLANTS DATABASE

Họ và tên : Lưu Quang Tiến Hoàng

MSSV : 20521342

Lớp : DS103.M21

Tên bộ dữ liệu:

Iris Plants Database

Nguồn thu thập và cách thu thập:

Creator: R.A. Fisher

Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)

Date: July, 1988

Kích thước dữ liệu:**File raw data:**

Số thuộc tính: 5

Kích thước: 4.44KB

File tidy data:

Số thuộc tính: 5

Tidydata được lưu thành file: Iris.csv

Kích thước: 4.8KB

Số dòng dữ liệu: 150

Thông tin các thuộc tính:

Sepal_Length: Chiều dài đài hoa tính bằng cm

Sepal_Width: Chiều rộng đài hoa tính bằng cm

Petal_Length: Chiều dài cánh hoa tính bằng cm

Petal_Width: Chiều rộng cánh hoa tính bằng cm

Class: Phân loại:

-- Iris Setosa

-- Iris Versicolour

-- Iris Virginica

Missing Attribute Values:

Không có

Class Distribution:

33,3% cho mỗi 3 lớp.

Instruction list:

```
rm(list=ls())
```

```
mydata <- read.csv("dataset/iris.data", header = FALSE)
```

```
iris <- c("Sepal_Length", "Sepal_Width", "Petal_Length", "Petal_Width", "Class")
```

```
colnames(mydata) <- iris
```

```
write.csv(mydata,file="Iris.csv",row.names = FALSE)
```

--- HẾT ---