

Chapter 6

BASIC DATA ANALYSIS (1)

資料處理

在資料處理的實作時

一般我們都是輸入一組或多組資料，再經過程式處理之後

輸出成檔案

目錄查詢

getwd(), setwd() 使用方式

getwd() // 抓到工作目錄位置

setwd(path) // 設定工作目錄位置

```
1 setwd("C:/")
2 getwd()
3
4 setwd("F:/Course/1062/data mining/code/R/ch6")
5 getwd()
```



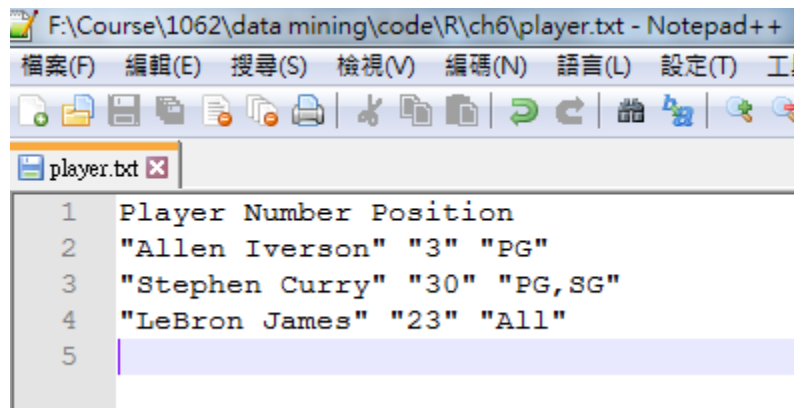
The screenshot shows an R console window with two tabs: 'Console' and 'Terminal'. The 'Terminal' tab is active, showing the current directory path 'F:/Course/1062/data mining/code/R/ch6/'. Below the path, the following commands and their outputs are displayed:

```
> setwd("C:/")
> getwd()
[1] "C:/"
>
> setwd("F:/Course/1062/data mining/code/R/ch6")
> getwd()
[1] "F:/Course/1062/data mining/code/R/ch6"
> |
```

資料輸入

read.table (txt) 使用方式

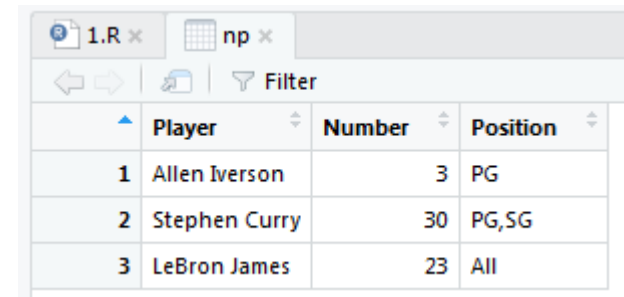
read.table(資料路徑 ,
 header = T/F,
 stringsAsFactors = T/F)



A screenshot of the Notepad++ application window. The title bar shows the file path: F:\Course\1062\data mining\code\R\ch6\player.txt - Notepad++. The menu bar includes 檔案(F), 編輯(E), 搜尋(S), 檢視(V), 編碼(N), 語言(L), 設定(T), and 工具(T). The toolbar contains icons for file operations and editing. The editor shows a file named player.txt with the following content:

```
1 Player Number Position
2 "Allen Iverson" "3" "PG"
3 "Stephen Curry" "30" "PG,SG"
4 "LeBron James" "23" "All"
5
```

```
1 fp <- "player.txt"
2
3 np <- read.table(fp, header=TRUE,
4                 stringsAsFactors = FALSE)
5
6 View(np)
```



A screenshot of the RStudio interface. The top pane shows the R script with the same code as the previous block. The bottom pane shows the environment with two objects: 1.R and np. The np object is selected, and its data is displayed in a table view. The table has three columns: Player, Number, and Position. The data is as follows:

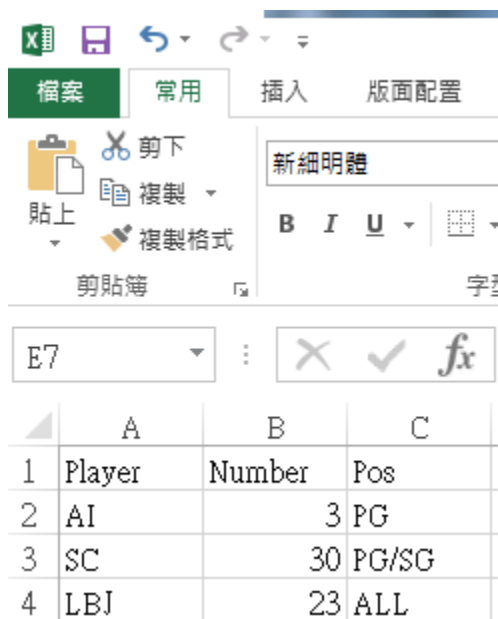
	Player	Number	Position
1	Allen Iverson	3	PG
2	Stephen Curry	30	PG,SG
3	LeBron James	23	All

資料輸入

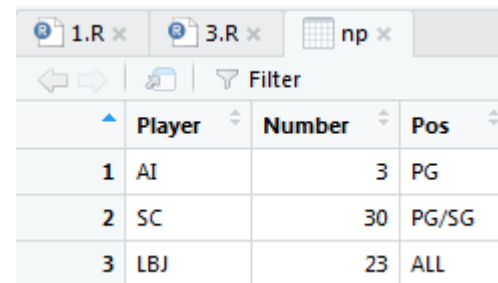
read.table (csv) 使用方式

read.table(資料路徑 ,
 header = T/F,
 stringsAsFactors = T/F
 sep =“,”)

```
1 fp <- "player.csv"
2
3 np <- read.table(fp, header=TRUE,
4                 stringsAsFactors = FALSE,
5                 sep=",")
6
7 View(np)
```



	A	B	C
1	Player	Number	Pos
2	AI	3	PG
3	SC	30	PG/SG
4	LBJ	23	ALL



	Player	Number	Pos
1	AI	3	PG
2	SC	30	PG/SG
3	LBJ	23	ALL

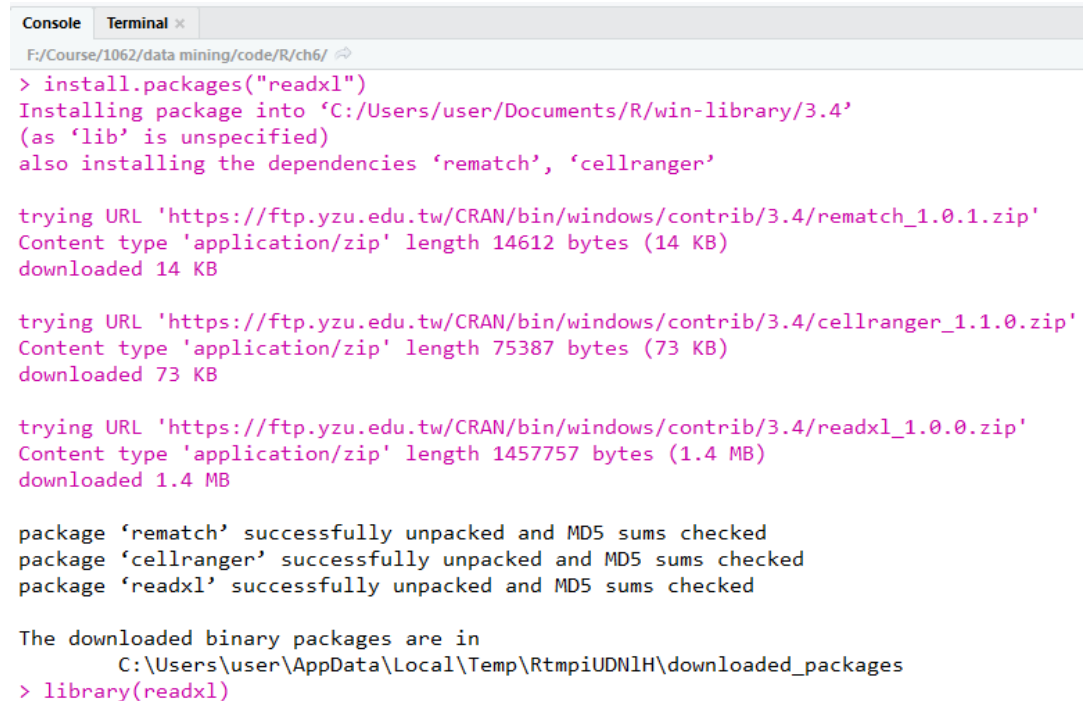
資料輸入

read.excel (xlsx) 使用方式

我們需要先安裝套件才能打開
Excel 檔案

```
install.packages("readxl")  
library(readxl)
```

```
1 install.packages("readxl")  
2 library(readxl)
```



```
Console Terminal x  
F:/Course/1062/data mining/code/R/ch6/ ↵  
> install.packages("readxl")  
Installing package into 'C:/Users/user/Documents/R/win-library/3.4'  
(as 'lib' is unspecified)  
also installing the dependencies 'rematch', 'cellranger'  
  
trying URL 'https://ftp.yzu.edu.tw/CRAN/bin/windows/contrib/3.4/rematch_1.0.1.zip'  
Content type 'application/zip' length 14612 bytes (14 KB)  
downloaded 14 KB  
  
trying URL 'https://ftp.yzu.edu.tw/CRAN/bin/windows/contrib/3.4/cellranger_1.1.0.zip'  
Content type 'application/zip' length 75387 bytes (73 KB)  
downloaded 73 KB  
  
trying URL 'https://ftp.yzu.edu.tw/CRAN/bin/windows/contrib/3.4/readxl_1.0.0.zip'  
Content type 'application/zip' length 1457757 bytes (1.4 MB)  
downloaded 1.4 MB  
  
package 'rematch' successfully unpacked and MD5 sums checked  
package 'cellranger' successfully unpacked and MD5 sums checked  
package 'readxl' successfully unpacked and MD5 sums checked  
  
The downloaded binary packages are in  
C:/Users/user/AppData/Local/Temp/RtmpiUDN1H/downloaded_packages  
> library(readxl)
```

資料輸入

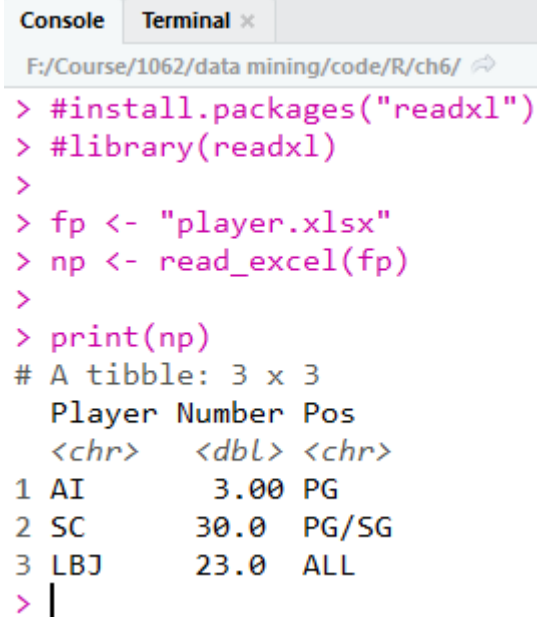
read.excel (xlsx) 使用方式

read.excel(資料路徑)



	A	B	C
1	Player	Number	Pos
2	AI	3	PG
3	SC	30	PG/SG
4	LBJ	23	ALL

```
1 #install.packages("readxl")
2 #library(readxl)
3
4 fp <- "player.xlsx"
5 np <- read_excel(fp)
6
7 print(np)
```



```
Console Terminal x
F:/Course/1062/data mining/code/R/ch6/

> #install.packages("readxl")
> #library(readxl)
>
> fp <- "player.xlsx"
> np <- read_excel(fp)
>
> print(np)
# A tibble: 3 x 3
  Player Number Pos
  <chr>    <dbl> <chr>
1 AI         3.00 PG
2 SC        30.0 PG/SG
3 LBJ        23.0 ALL
> |
```

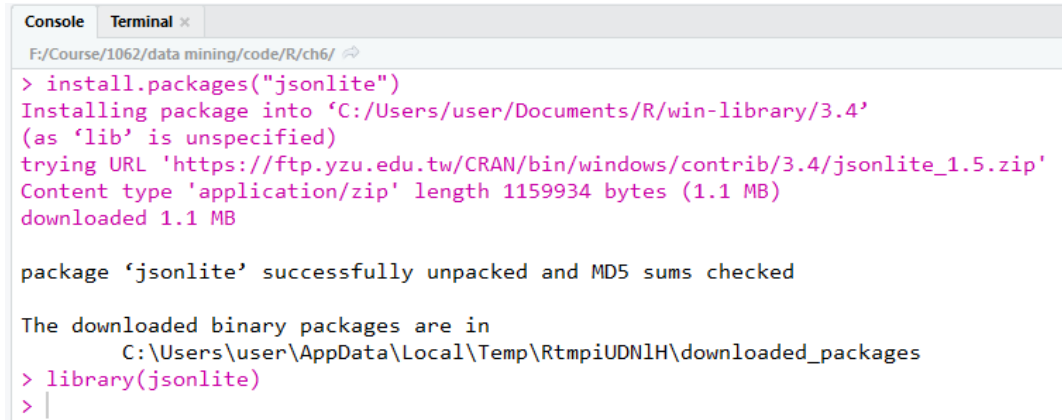
資料輸入

fromJSON 使用方式

我們需要先安裝套件才能打開
JSON(javascript object notation)

```
install.packages("jsonlite")  
library(jsonlite)
```

```
1 install.packages("jsonlite")  
2 library(jsonlite)  
3
```

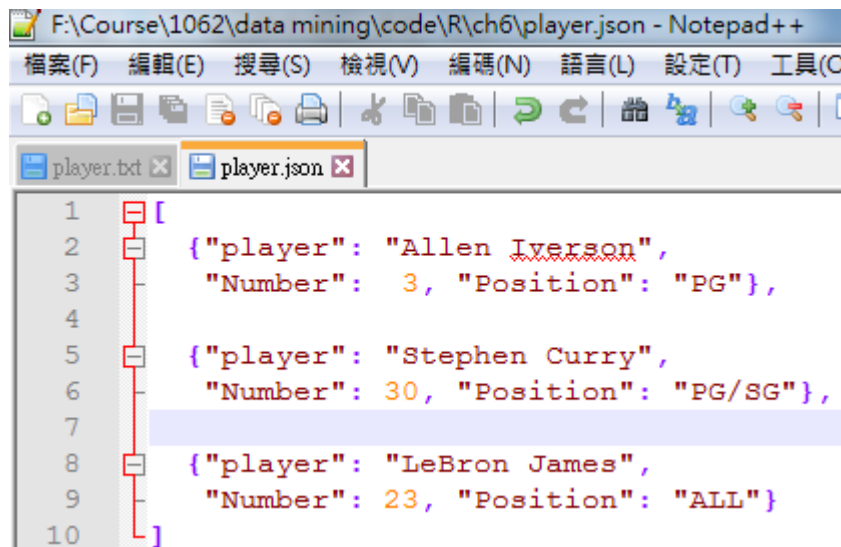


```
Console Terminal x  
F:/Course/1062/data mining/code/R/ch6/ ↗  
> install.packages("jsonlite")  
Installing package into 'C:/Users/user/Documents/R/win-library/3.4'  
(as 'lib' is unspecified)  
trying URL 'https://ftp.yzu.edu.tw/CRAN/bin/windows/contrib/3.4/jsonlite_1.5.zip'  
Content type 'application/zip' length 1159934 bytes (1.1 MB)  
downloaded 1.1 MB  
  
package 'jsonlite' successfully unpacked and MD5 sums checked  
  
The downloaded binary packages are in  
C:\Users\user\AppData\Local\Temp\RtmpiUDNlH\downloaded_packages  
> library(jsonlite)  
> |
```


資料輸入

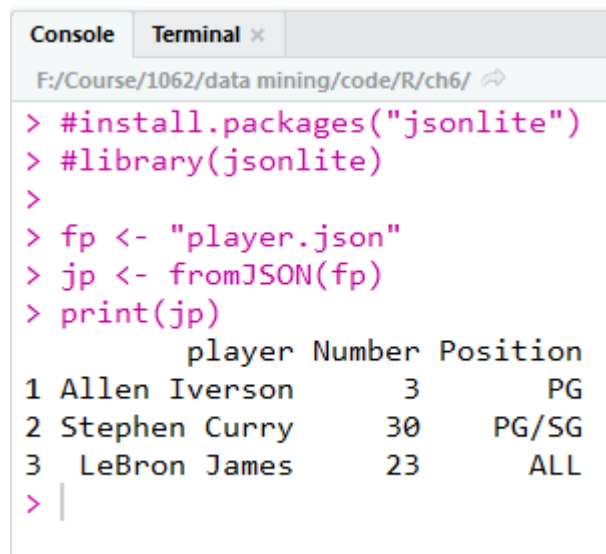
fromJSON 使用方式

fromJSON(資料路徑)



```
1 [{"player": "Allen Iverson",
2   "Number": 3, "Position": "PG"},
3
4   {"player": "Stephen Curry",
5     "Number": 30, "Position": "PG/SG"},
6
7   {"player": "LeBron James",
8     "Number": 23, "Position": "ALL"}
9 ]
10 ]
```

```
1 #install.packages("jsonlite")
2 #library(jsonlite)
3
4 fp <- "player.json"
5 jp <- fromJSON(fp)
6 print(jp)
7 |
```



```
Console Terminal x
F:/Course/1062/data mining/code/R/ch6/

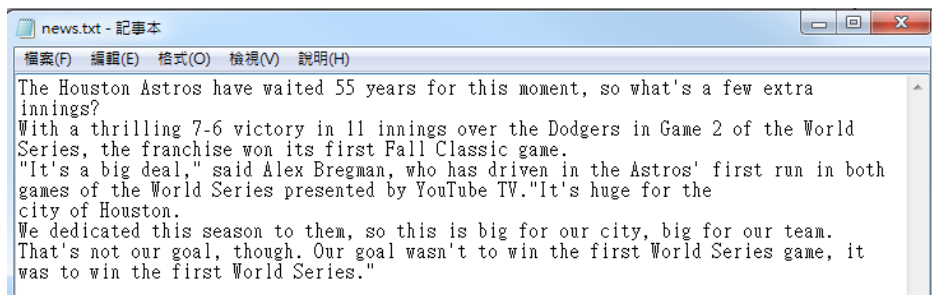
> #install.packages("jsonlite")
> #library(jsonlite)
>
> fp <- "player.json"
> jp <- fromJSON(fp)
> print(jp)
      player Number Position
1 Allen Iverson      3      PG
2 Stephen Curry    30 PG/SG
3  LeBron James    23     ALL
> |
```

資料輸入

readLines 使用方式

不是所有檔案都是有經過整理的
所以我們使用 readLines 來一行
一行讀取

readLines(資料路徑)



```
1 fp <- "news.txt"
2
3 np <- readLines(fp)
4
5 print(np)
```

```
> fp <- "news.txt"
```

```
>
```

```
> np <- readLines(fp)
```

```
>
```

```
> print(np)
```

```
[1] "The Houston Astros have waited 55 years for
```

```
[2] "With a thrilling 7-6 victory in 11 innings  
first Fall Classic game."
```

```
[3] "\"It's a big deal,\" said Alex Bregman, who  
presented by YouTube TV.\"It's huge for the cit
```

```
[4] "We dedicated this season to them, so this i
```

```
[5] "That's not our goal, though. Our goal wasn'  
ries.\""
```

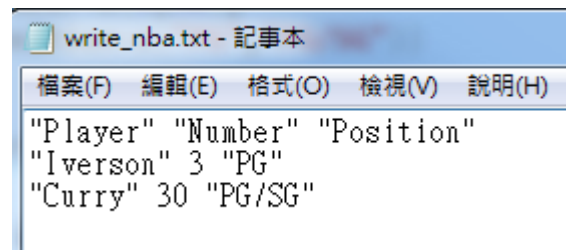
```
> |
```

資料輸出

write.table (txt) 使用方式

write.table(變數 ,
 資料路徑 ,
 rows.name = T/F)

```
1 player <- data.frame(Player=c("Iverson", "Curry"),  
2                       Number=c(3, 30),  
3                       Position=c("PG", "PG/SG"))  
4 fp <- "write_nba.txt"  
5  
6 write.table(player, fp, row.name=FALSE)
```



write_nba.txt - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

"Player" "Number" "Position"
"Iverson" 3 "PG"
"Curry" 30 "PG/SG"

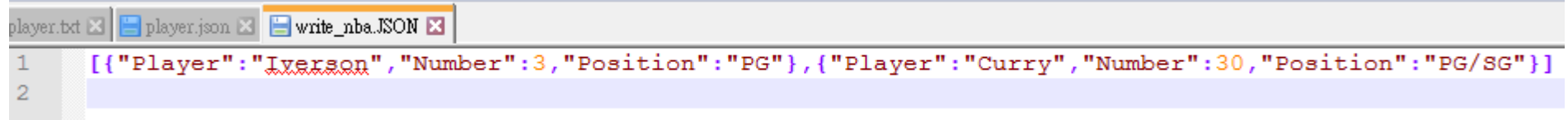
資料輸出

toJSON (JSON) 使用方式

toJSON(變數)

writeLine(變數 ,con= 路徑)

```
1 player <- data.frame(Player=c("Iverson","Curry"),
2                       Number=c(3,30),
3                       Position=c("PG","PG/SG"))
4 fp <- "write_nba.json"
5
6 a<-toJSON(player)
7 writeLines(a, con=fp)|
```



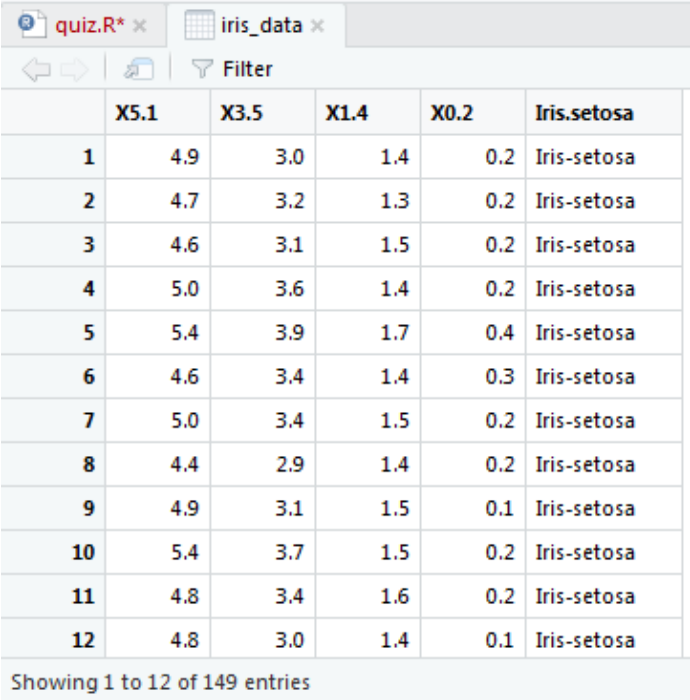
```
1 [{"Player": "Iverson", "Number": 3, "Position": "PG"}, {"Player": "Curry", "Number": 30, "Position": "PG/SG"}]
2
```

隨堂練習 1-1

1. 讀取 UCI IRIS data

(<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>)

2. 轉成 data.frame



	X5.1	X3.5	X1.4	X0.2	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
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa
10	5.4	3.7	1.5	0.2	Iris-setosa
11	4.8	3.4	1.6	0.2	Iris-setosa
12	4.8	3.0	1.4	0.1	Iris-setosa

Showing 1 to 12 of 149 entries

隨堂練習 1-2

1. 將四個欄位名稱 (X5.1, X3.5, X1.4, X0.2) 換成如下
sepal length, sepal width, petal length, and petal width
2. 將最後一個欄位設成 class

	sepal l	sepal w	petal l	petal w	class
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
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa
10	5.4	3.7	1.5	0.2	Iris-setosa
11	4.8	3.4	1.6	0.2	Iris-setosa

Showing 1 to 12 of 149 entries

Any Questions !?