

106-1 生物統計學二 實習課

R : Introduction

周芷好

2017.09.14

大綱

- 實習課說明
- Introduction to R
 - 簡介與安裝
 - 基本操作

實習課說明

課程助教

- 周芷妤 d04849010@ntu.edu.tw
- 徐麗茵
- 陳怡婷

實習課內容

- 時間：星期四 第八、九節(15:30 - 17:20)
- 地點：公衛大樓214電腦教室
- 課程內容：R 上機操作
- 實習課教材會放在ceiba上

作業繳交規定

- **正課**作業

- 繳交日期：依照老師規定繳交
- 繳交時間、方式：第六節(13:20)，以紙本繳交給助教
- 作業會於繳交後一週後發還，並於實習課時檢討
- 遲交作業：
 - 於規定繳交時間至檢討作業之間繳交：原始分數× 0.8
 - 檢討作業後繳交：原始分數× 0.6
 - 11/6(一)(期中考週的星期一)17:00後，不得再繳交期中考前之作業
 - 1/8(一)(期末考週的星期一)17:00後，不得再繳交期末考前之作業
- 嚴禁抄襲作業
 - 若發現抄襲，將取其中一人分數，除上抄襲人數作為該次作業成績

作業繳交規定

- 實習課作業

- 繳交日期、時間：請盡量於當週實習課下課前上傳ceiba作業區繳交

- * 最晚上傳期限為：當週禮拜六中午12:00

- 遲交作業將不計該次作業分數

- 每次將挑一位同學的作業，徵得同意後作為解答提供同學們參考

- 嚴禁抄襲作業

- 若發現抄襲，將取其中一人分數，除上抄襲人數作為該次作業成績

發問管道

- ceiba 討論區
- E-mail
正好 d04849010@ntu.edu.tw

Introduction to R

簡介與安裝
基本操作

R 下載

- 官網：<https://www.r-project.org/>



1. [Home]

Download
CRAN

Here !!



R Project

About R

Logo

Contributors

What's New?

Reporting Bugs

Development Site

Conferences

Search

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

2.

Taiwan

<https://ftp.yzu.edu.tw/CRAN/>

<http://ftp.yzu.edu.tw/CRAN/>

<http://cran.csie.ntu.edu.tw/>

Department of Computer Science and Engineering, Yuan Ze University

Department of Computer Science and Engineering, Yuan Ze University

National Taiwan University, Taipei

- R version 3.4.2 (Short Summer) prerelease versions** will appear starting Monday 2017-09-18.
Final release is scheduled for Thursday 2017-09-28.
- The R Journal Volume 9/1** is available.
- R version 3.4.1 (Single Candle)** has been released on Friday 2017-06-30.
- R version 3.3.3 (Another Canoe)** has been released on Monday 2017-03-06.

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

3.

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.



4.

[base](#)

[contrib](#)

[old contrib](#)

[Rtools](#)

Binaries for base distribution (managed by Duncan Murdoch). This is what you want to [install R for the first time](#).

Binaries of contributed CRAN packages (for R >= 2.11.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.

Binaries of contributed CRAN packages for outdated versions of R (for R < 2.11.x; managed by Uwe Ligges).

Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself.



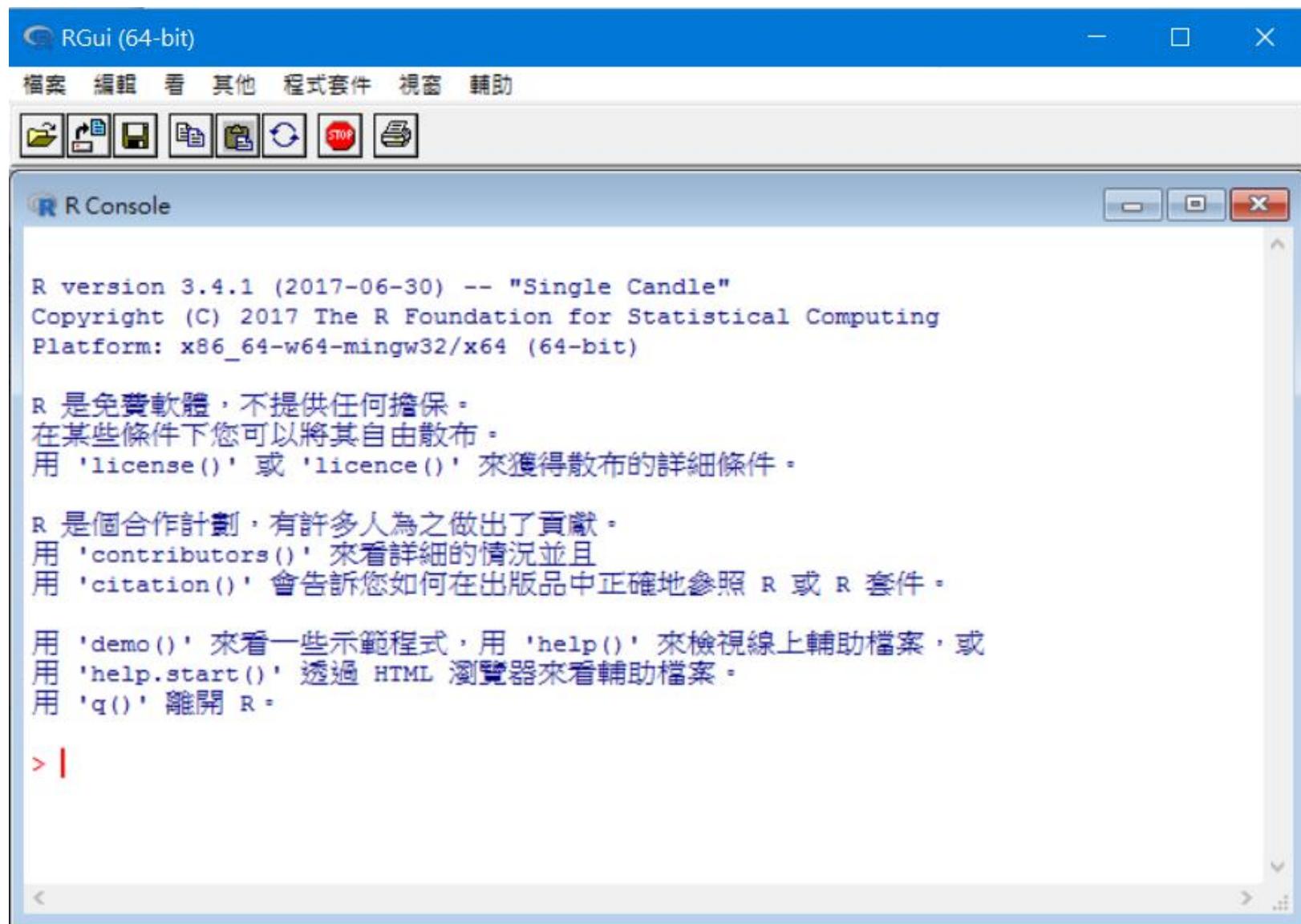
5.

[Download R 3.4.1 for Windows](#) (75 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

R 的介面



R 的基本操作

- 「**>**」表示程式待命中，可進行運算
(輸入完成後，利用Enter鍵即可執行該指令)
- 「**+**」表示指令尚未完成
(需要繼續輸入或是利用Esc鍵取消)
- **#** 表示後面指令為註解，不需執行
- 避免使用R內建指令命名 (ex: **matrix, var, c, ...**)
- 須注意大小寫有區別

R 的基本操作

```
★ > # 定義物件:  
> a <- 3  
>  
★ > # 四則運算: (注意先乘除後加減)  
> a+2          #加法  
[1] 5  
> a-2          #減法  
[1] 1  
> a*2          #乘法  
[1] 6  
> a/2          #除法  
[1] 1.5  
> a^2          #次方 or a**2  
[1] 9  
> sqrt(a)      #開根號 or a^0.5  
[1] 1.732051  
> exp(a)       #e^a  
[1] 20.08554  
>  
★ > # 產生序列:  
> 1:3          #產生從1到3的整數  
[1] 1 2 3  
> rep(1,3)      #把1重複3次  
[1] 1 1 1  
> seq(1,3,0.5)  #從1到3每隔0.5產生一個數字  
[1] 1.0 1.5 2.0 2.5 3.0  
>  
★ > # 產生矩陣及向量  
> v <- c(1:6)    #產生向量  
> matrix(1:6,3,2) #產生3x2的矩陣，預設為從第一行開始排，排完後再到第二行  
[1,] [2]  
[1,] 1 4  
[2,] 2 5  
[3,] 3 6
```

敘述性統計

```
> # 使用R內建的iris資料檔
> iris          #直接呼叫資料
   Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
1          5.1       3.5         1.4       0.2    setosa
2          4.9       3.0         1.4       0.2    setosa
3          4.7       3.2         1.3       0.2    setosa
4          4.6       3.1         1.5       0.2    setosa
5          5.0       3.6         1.4       0.2    setosa
6          5.4       3.9         1.7       0.4    setosa
7          4.6       3.4         1.4       0.3    setosa
8          5.0       3.4         1.5       0.2    setosa

> str(iris)      #觀察資料結構
'data.frame': 150 obs. of 5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
> names(iris)    #呼叫變項名稱
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
> summary(iris)  #描述資料
   Sepal.Length   Sepal.Width   Petal.Length   Petal.Width   Species
Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa   :50
1st Qu.:5.100  1st Qu.:2.800  1st Qu.:1.600  1st Qu.:0.300  versicolor:50
Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
```

敘述性統計

```
> #總和  
> sum(iris$Sepal.Length)  
[1] 876.5  
> #平均值  
> mean(iris$Sepal.Length)  
[1] 5.843333  
> #中位數  
> median(iris$Sepal.Length)  
[1] 5.8  
> #變異數  
> var(iris$Sepal.Length)  
[1] 0.6856935  
> #標準差  
> sd(iris$Sepal.Length)  
[1] 0.8280661  
> #共變異數  
> cov(iris$Sepal.Length,iris$Sepal.Width)  
[1] -0.042434  
> #相關係數  
> cor(iris$Sepal.Length,iris$Sepal.Width)  
[1] -0.1175698  
  
> #指定資料  
> attach(iris)  
> var(Sepal.Length)  
[1] 0.6856935  
> detach(iris)    #取消指定  
> var(Sepal.Length)  
Error in is.data.frame(x) : 找不到物件 'Sepal.Length'
```

繪圖

```

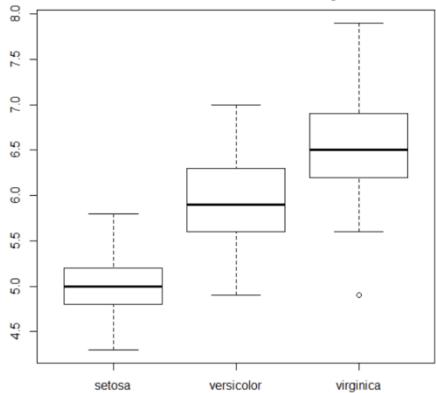
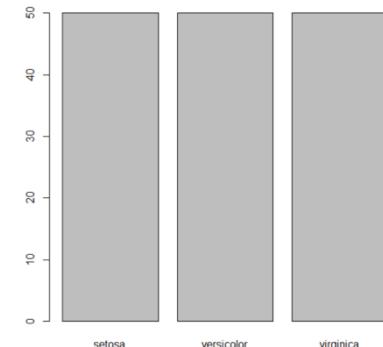
> attach(iris)
> table(Species)
Species
  setosa versicolor virginica
      50      50      50

```

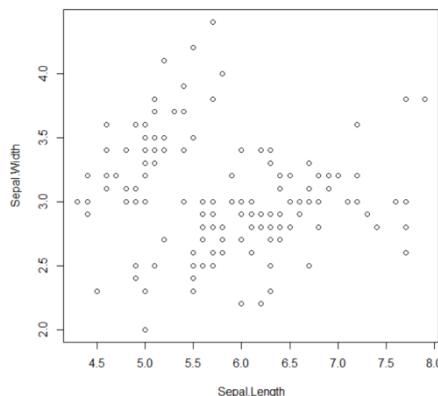
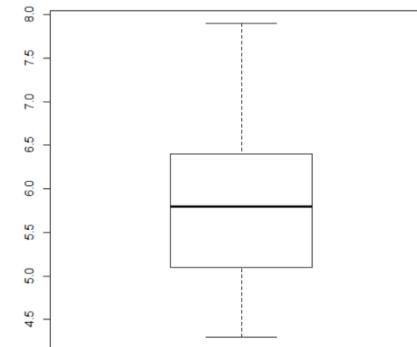
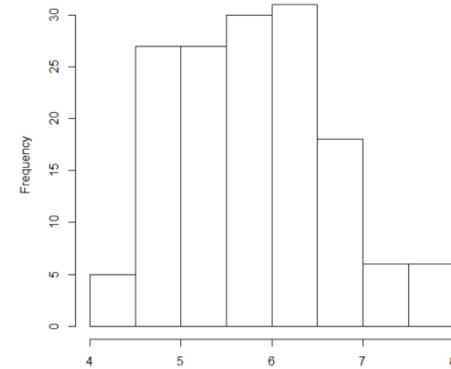
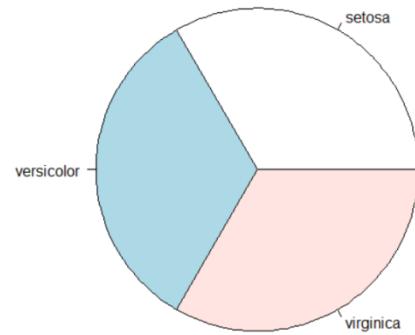
```

> barplot(table(Species))
> pie(table(Species))
> boxplot(Sepal.Length)
> boxplot(Sepal.Length~Species)
> hist(Sepal.Length)
> plot(Sepal.Length, Sepal.Width)

```



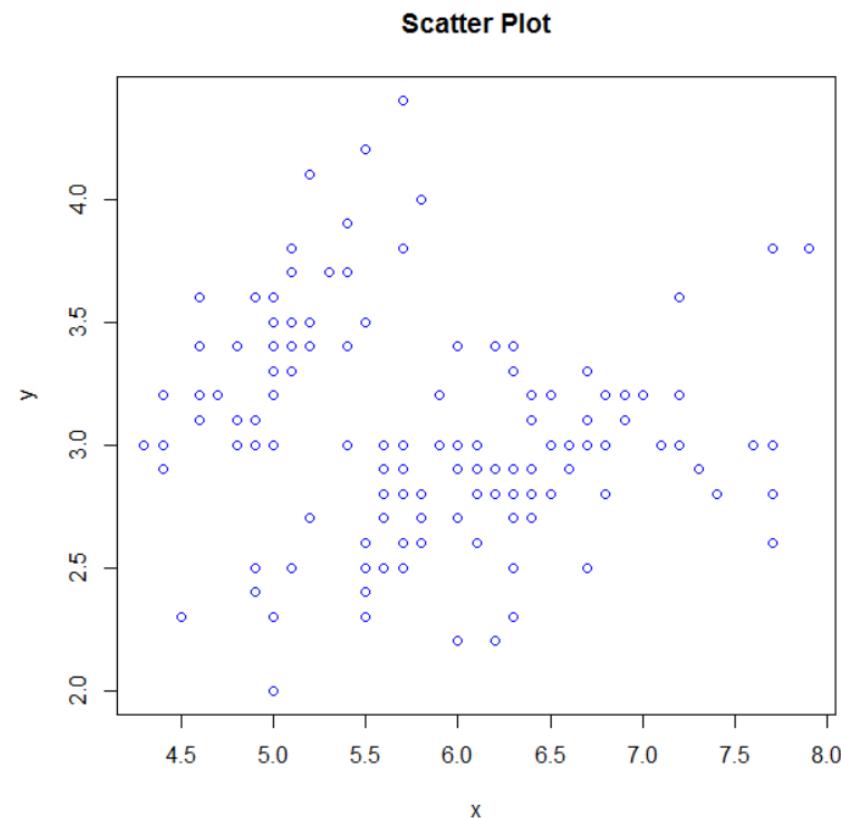
#長條圖
#圓餅圖
#盒鬚圖
#盒鬚圖 (依Species分組)
#直方圖
#散佈圖: plot(X, Y)



繪圖相關指令

```
> plot(Sepal.Length, Sepal.Width, main="Scatter Plot", xlab="x", ylab="y", col="blue")
```

語法	說明	備註
main	標題	(main="title")
xlab	X軸名稱	(xlab="my label")
ylab	Y軸名稱	
col	顏色	(col="red", "blue", ...)



資料匯入

- 點選方式

```
> data1 <- read.table(file.choose(), sep=",", header=T)
```

- 輸入路徑

```
> data1 <- read.table("C:\\MyData1.csv", sep=",", header=T)
```

- 顯示前幾筆資料

```
> head(data1)
```

語法	說明	備註
file.choose()	點選檔案來源	也可以直接輸入路徑 "C:\\MyData1.csv"
sep	資料分隔方式	空白 “ ” 逗號 “ , ” TAB鍵 “ \\t ”
header	是否將第一列視為變數名稱	T or F

資料匯出

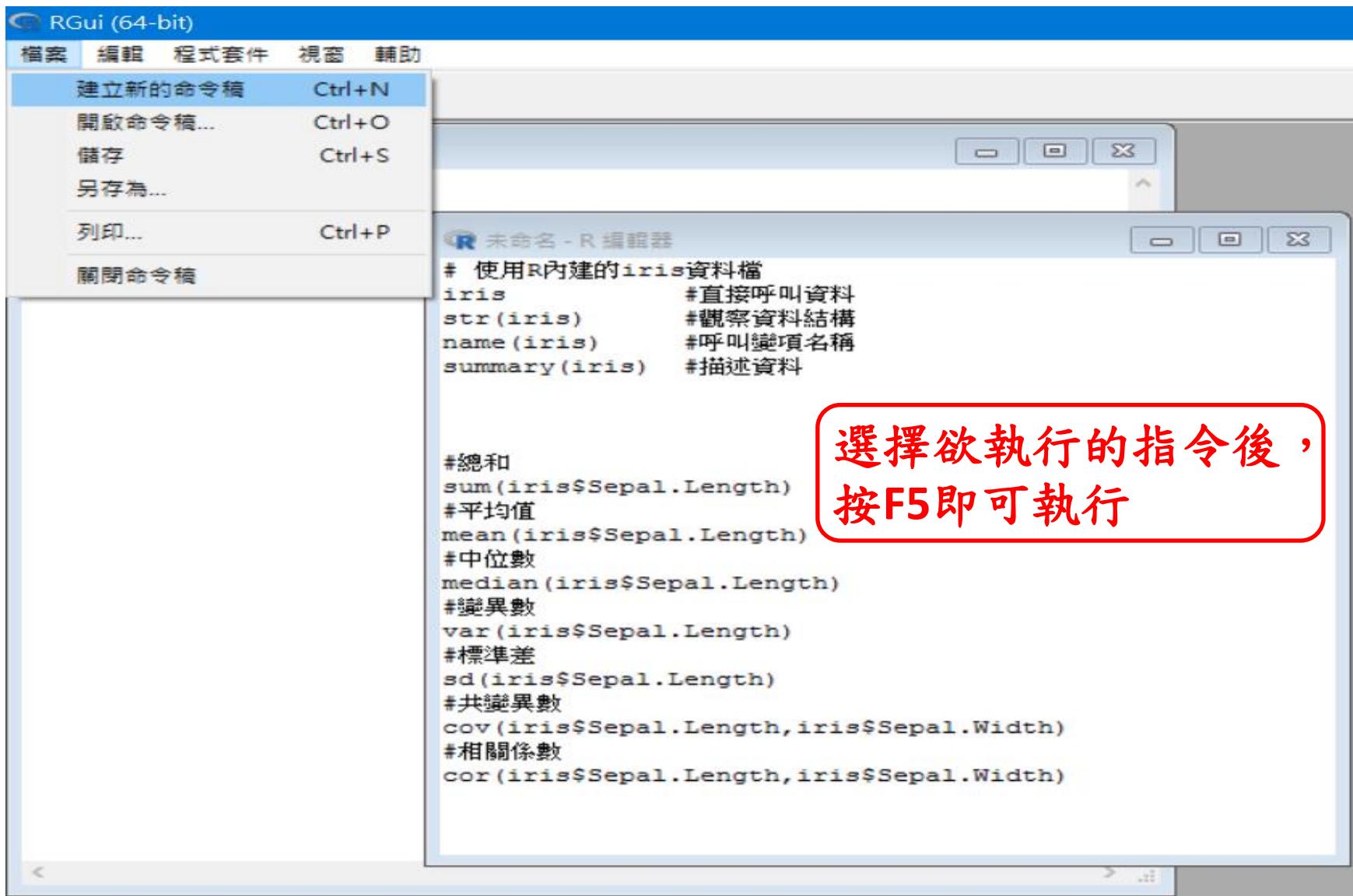
```
> write.table(data1, "C:\\mydata.csv", sep=",", row.names=F)
```

語法	說明	備註
data1	欲匯出的資料名稱	
"C:\\mydata.csv"	匯出的檔案名稱及路徑	
sep	資料分隔方式	空白 “ ” 逗號 “,” TAB鍵 “\t”
row.names	是否要儲存橫列名稱	T or F



```
> data1
   id sex birthday weight height McDonlond home department sport
1   1   M 11Aug1960     63  168.0       1     1          6     1
2   2   M 29Mar1962     60  180.0       1     1          6     1
3   3   F 25Feb1966     54  165.0       1     1          1     1
4   4   M 08Nov1966     78  168.0       1     1          5     1
5   5   M 12Dec1967     70  168.5       0     1          1     0
6   6   F 12Jun1971     52  152.0       1     1          1     1
```

Script(命令稿)



查詢指令

- 已知指令名稱，想查詢詳細用法時
 - ✓ `help(barplot)`
 - ✓ `?barplot`
- 只知道方法名稱，但不知道指令時
 - ✓ `help.search("bar chart")`
 - ✓ `??"bar chart"`
 - ✓ `google`

課堂練習

- 請先將資料檔 “MyData1.csv” 汇入R中
- 請計算 weight 的 3 個敘述統計量
- 請畫出 weight (X 軸) 及 height (Y 軸) 的scatter plot