



106-2: EE4052

通識課程： 計算機程式設計 之旅

Computer Programming

Unit 04: 數據處理與繪圖功能

連 豐 力

臺大電機系

Feb 2018 - Jun 2018

問題探索與分析

問題

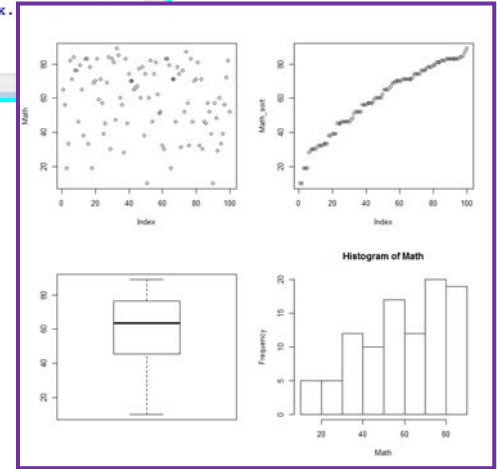
理論
(公式)

計算機
程式設計

```
R Console
> summary(data)
  Name      Chinese      English      Math
Agnes : 1      Min.   : 0.00      Min.   : 0.00      Min.   :10.00
Aiolos : 1     1st Qu.:38.75     1st Qu.:32.00     1st Qu.:45.75
Alan   : 1     Median :66.00     Median :57.00     Median :63.50
Alexis : 1     Mean   :57.98     Mean   :51.86     Mean   :59.38
Alice  : 1     3rd Qu.:78.00     3rd Qu.:71.00     3rd Qu.:76.25
Alina  : 1     Max.   :87.00     Max.   :90.00     Max.   :90.00
(Other):94
> |
```

算術平均數 $M(=\bar{X}) = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$
標準差 $S = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{X}^2}$

- `summary(Math)`
- `sort(Math)`
- `boxplot(Math)`
- `hist(Math)`



分析：數據化

分析：視覺化

問題探索與分析

計算機程式設計 - 2018S

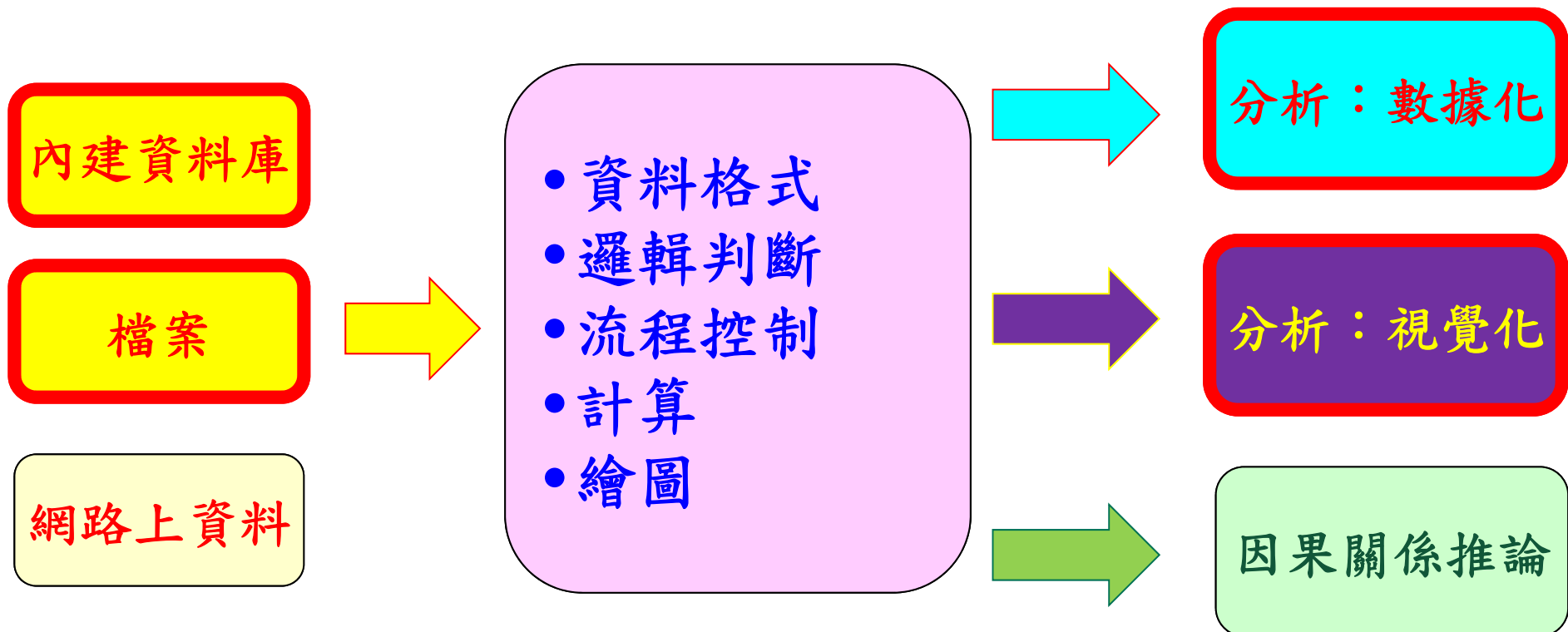
U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

輸入

程式

輸出



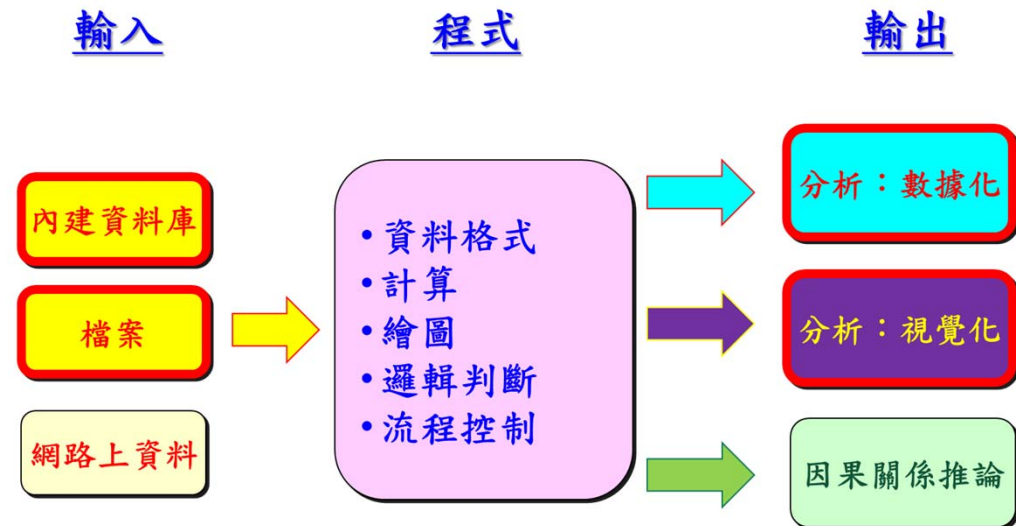
課程主題進度

計算機程式設計 - 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

- U01: 課程介紹：討論主題，作業，報告，進行方式
- U02: 主題，案例，程式，演算法，資源
- U03: 設定軟體 R 與 Rstudio
- U04: 數據處理與繪圖指令功能
- U05: 資料類別與基本運算
- U06: 邏輯判斷與流程控制
- U07: 函數：計算與排序
- U08: 多維度資料格式
- U09: 檔案資料輸入與輸出
- U10: 繪圖功能與文字
- U11: 多重繪圖與顏色
- U12: 函數：動畫與動作
- U13: 探索性資料分析
- U14: 資料間的相關性
- U15: 資料連結分析



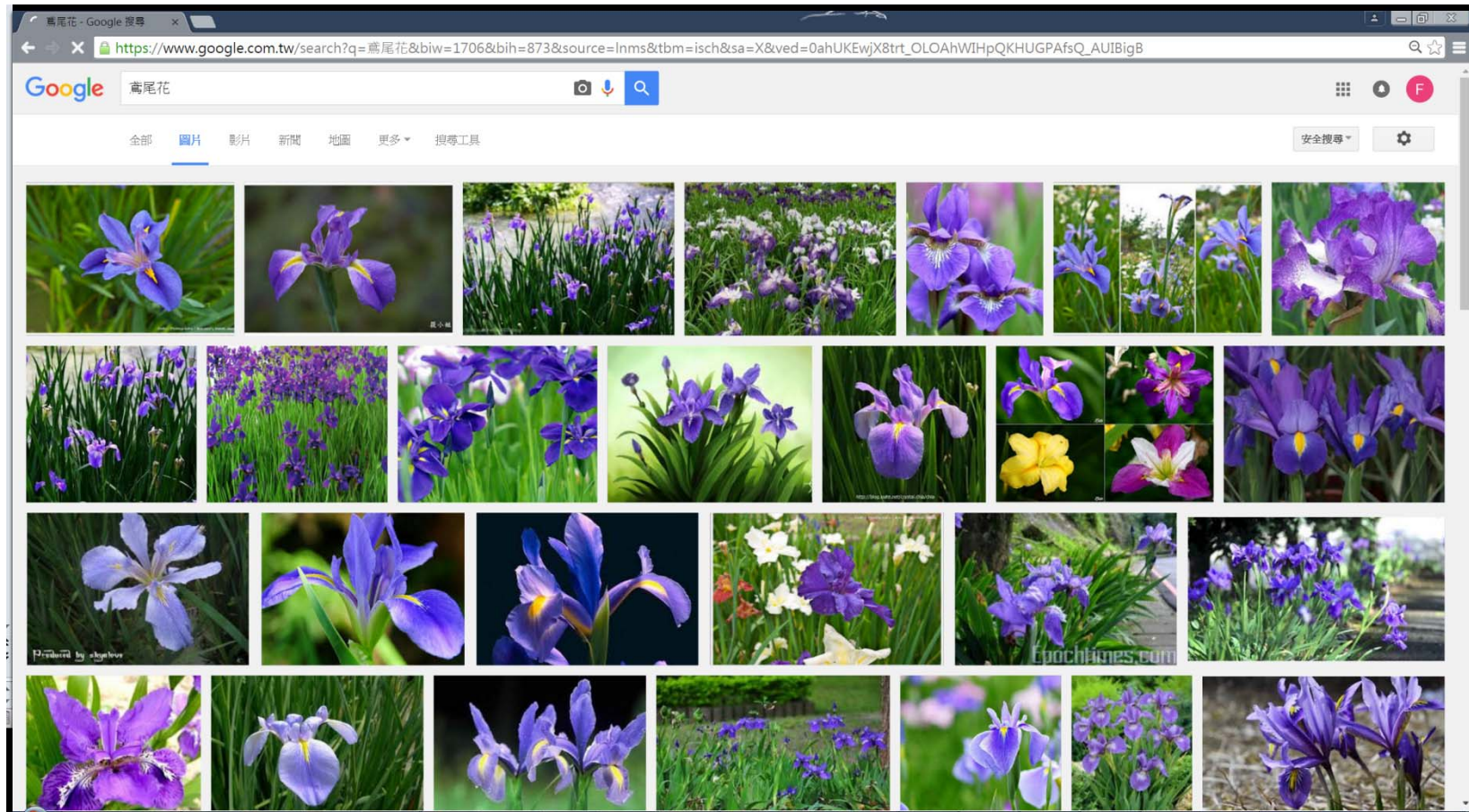
- 以 鳶尾花 - IRIS 數據為例
- 數據所在位置與數據的內容
- 初步分析數據
- 繪製圖形 -
 - 一維圖：
 - 直方圖，盒鬚圖，莖葉圖，長條記錄圖，圓餅圖，機率分布圖，經驗累積分布圖，常態機率圖
 - 多維圖：
 - 散點圖，散點直方核密度，多重分布，三維散點圖

鳶尾花 - IRIS

計算機程式設計 - 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE



鳶尾花 - IRIS

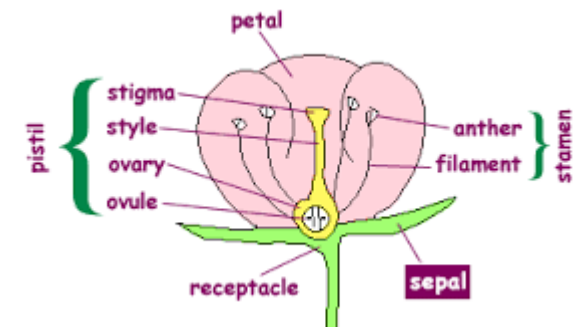
■ 鳶尾花 (iris) 資料集

- 非常著名的生物資訊資料集之一
- 取自美國加州大學歐文分校的機械學習資料庫

■ 資料的筆數為150筆，

■ 共有五個欄位：

1. 花萼長度 (Sepal Length)：計算單位是公分。
2. 花萼寬度 (Sepal Width)：計算單位是公分。
3. 花瓣長度 (Petal Length)：計算單位是公分。
4. 花瓣寬度 (Petal Width)：計算單位是公分。
5. 類別 (Class)：可分為 **Setosa**，**Versicolor** 和 **Virginica** 三個品種。



鳶尾花 - IRIS

```
Console -1
> iris
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
9 4.4 2.9 1.4 0.2 setosa
10 4.9 3.1 1.5 0.1 setosa
11 5.4 3.7 1.5 0.2 setosa
12 4.8 3.4 1.6 0.2 setosa
13 4.8 3.0 1.4 0.1 setosa
14 4.3 3.0 1.1 0.1 setosa
15 5.8 4.0 1.2 0.2 setosa
16 5.7 4.4 1.5 0.4 setosa
17 5.4 3.9 1.3 0.4 setosa
18 5.1 3.5 1.4 0.3 setosa
19 5.7 3.8 1.7 0.3 setosa
20 5.1 3.8 1.5 0.3 setosa
21 5.4 3.4 1.7 0.2 setosa
22 5.1 3.7 1.5 0.4 setosa
23 4.6 3.6 1.0 0.2 setosa
24 5.1 3.3 1.7 0.5 setosa
25 4.8 3.4 1.9 0.2 setosa
26 5.0 3.0 1.6 0.2 setosa
27 5.0 3.4 1.6 0.4 setosa
28 5.2 3.5 1.5 0.2 setosa
29 5.2 3.4 1.4 0.2 setosa
30 4.7 3.2 1.6 0.2 setosa
31 4.8 3.1 1.6 0.2 setosa
32 5.4 3.4 1.5 0.4 setosa
33 5.2 4.1 1.5 0.1 setosa
34 5.5 4.2 1.4 0.2 setosa
35 4.9 3.1 1.5 0.2 setosa
36 5.0 3.2 1.2 0.2 setosa
37 5.5 3.5 1.3 0.2 setosa
38 4.9 3.6 1.4 0.1 setosa
39 4.4 3.0 1.3 0.2 setosa
40 5.1 3.4 1.5 0.2 setosa
41 5.0 3.5 1.3 0.3 setosa
42 4.5 2.3 1.3 0.3 setosa
43 4.4 3.2 1.3 0.2 setosa
44 5.0 3.5 1.6 0.6 setosa
45 5.1 3.8 1.9 0.4 setosa
46 4.8 3.0 1.4 0.3 setosa
47 5.1 3.8 1.6 0.2 setosa
48 4.6 3.2 1.4 0.2 setosa
49 5.3 3.7 1.5 0.2 setosa
50 5.0 3.3 1.4 0.2 setosa
51 7.0 3.2 4.7 1.4 versicolor
52 6.4 3.2 4.5 1.5 versicolor
53 6.9 3.1 4.9 1.5 versicolor
54 5.5 2.3 4.0 1.3 versicolor
55 6.5 2.8 4.6 1.5 versicolor
56 5.7 2.8 4.5 1.3 versicolor
57 6.3 3.3 4.7 1.6 versicolor
58 4.9 2.4 3.3 1.0 versicolor
59 6.6 2.9 4.6 1.3 versicolor
60 5.2 2.7 3.9 1.4 versicolor
61 5.0 2.0 3.5 1.0 versicolor
```

```
-- -- -- --
91 5.5 2.6 4.4 1.2 versicolor
92 6.1 3.0 4.6 1.4 versicolor
93 5.8 2.6 4.0 1.2 versicolor
94 5.0 2.3 3.3 1.0 versicolor
95 5.6 2.7 4.2 1.3 versicolor
96 5.7 3.0 4.2 1.2 versicolor
97 5.7 2.9 4.2 1.3 versicolor
98 6.2 2.9 4.3 1.3 versicolor
99 5.1 2.5 3.0 1.1 versicolor
100 5.7 2.8 4.1 1.3 versicolor
101 6.3 3.3 6.0 2.5 virginica
102 5.8 2.7 5.1 1.9 virginica
103 7.1 3.0 5.9 2.1 virginica
104 6.3 2.9 5.6 1.8 virginica
105 6.5 3.0 5.8 2.2 virginica
106 7.6 3.0 6.6 2.1 virginica
107 4.9 2.5 4.5 1.7 virginica
108 7.3 2.9 6.3 1.8 virginica
109 6.7 2.5 5.8 1.8 virginica
110 7.2 3.6 6.1 2.5 virginica
111 6.5 3.2 5.1 2.0 virginica
112 6.4 2.7 5.3 1.9 virginica
113 6.8 3.0 5.5 2.1 virginica
114 5.7 2.5 5.0 2.0 virginica
115 5.8 2.8 5.1 2.4 virginica
116 6.4 3.2 5.3 2.3 virginica
117 6.5 3.0 5.5 1.8 virginica
118 7.7 3.8 6.7 2.2 virginica
119 7.7 2.6 6.9 2.3 virginica
120 6.0 2.2 5.0 1.5 virginica
121 6.9 3.2 5.7 2.3 virginica
122 5.6 2.8 4.9 2.0 virginica
123 7.7 2.8 6.7 2.0 virginica
124 6.3 2.7 4.9 1.8 virginica
125 6.7 3.3 5.7 2.1 virginica
126 7.2 3.2 6.0 1.8 virginica
127 6.2 2.8 4.8 1.8 virginica
128 6.1 3.0 4.9 1.8 virginica
129 6.4 2.8 5.6 2.1 virginica
130 7.2 3.0 5.8 1.6 virginica
131 7.4 2.8 6.1 1.9 virginica
132 7.9 3.8 6.4 2.0 virginica
133 6.4 2.8 5.6 2.2 virginica
134 6.3 2.8 5.1 1.5 virginica
135 6.1 2.6 5.6 1.4 virginica
136 7.7 3.0 6.1 2.3 virginica
137 6.3 3.4 5.6 2.4 virginica
138 6.4 3.1 5.5 1.8 virginica
139 6.0 3.0 4.8 1.8 virginica
140 6.9 3.1 5.4 2.1 virginica
141 6.7 3.1 5.6 2.4 virginica
142 6.9 3.1 5.1 2.3 virginica
143 5.8 2.7 5.1 1.9 virginica
144 6.8 3.2 5.9 2.3 virginica
145 6.7 3.3 5.7 2.5 virginica
146 6.7 3.0 5.2 2.3 virginica
147 6.3 2.5 5.0 1.9 virginica
148 6.5 3.0 5.2 2.0 virginica
149 6.2 3.4 5.4 2.3 virginica
150 5.9 3.0 5.1 1.8 virginica
> |
```


數據所在位置與數據的內容

- `iris[i, j]` # 某一個位置的數據
 - `iris[1, 1]`
 - `iris[1, 2]`
 - `iris[1, 3]`
 - `iris[1, 4]`
 - `iris[1, 5]`
 - `iris[2, 1]`
 - `iris[2, 2]`
 - `iris[1,]`
 - `iris[2,]`
 - `iris[3,]`
 - `iris[, 1]`
 - `iris[, 2]`
 - `iris[, 3]`

數據所在位置與數據的內容

- `iris[, j]`

某一個直行的數據

- `iris[, 1]`

- `iris[, 2]`

- `iris[, 3]`

- `iris[, 4]`

- `iris[, 5]`

- `iris$Sepal.Length`

- `iris$Sepal.Width`

- `iris$Petal.Length`

- `iris$Petal.Width`

- `iris$Species`



重新指定一個數據

計算機程式設計 – 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

- `mydata <- iris`
- `mydata`

- `mydata[i, j]`
 - `mydata[1, 1]`
 - `mydata[3,]`

初步分析數據

初步分析數據

- `mydata <- iris`
- `mydata`
- `str(mydata)` # Display the Structure

```
> str( mydata )
'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 ...
```

- `summary(mydata)` # Object Summaries

```
> summary( mydata )
  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
```

初步分析數據

- `mydata <- iris`
- `mydata`
- `head(mydata, n = 5)` # the first part of an object

```
> head( mydata, n = 5 )
  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
```

- `tail(mydata, n = 5)` # the last part of an object

```
> tail( mydata, n = 5 )
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
146           6.7           3.0           5.2           2.3 virginica
147           6.3           2.5           5.0           1.9 virginica
148           6.5           3.0           5.2           2.0 virginica
149           6.2           3.4           5.4           2.3 virginica
150           5.9           3.0           5.1           1.8 virginica
```


初步分析數據

- `mydata <- iris`
- `mydata`

- `mydata[, 1]`
- `mydata[, 2]`
- `mydata[, 3]`
- `mydata[, 4]`
- `mydata[, 5]`

the n-th column of an object

- `mydata$Sepal.Length`
- `mydata$Sepal.Width`
- `mydata$Petal.Length`
- `mydata$Petal.Width`
- `mydata$Species`

the data with the NAME

初步分析數據

- `mydata <- iris`
- `mydata`

- `mydata$Species == "setosa"` **# find the data with the NAME**
- `mydata$Species == "versicolor"`
- `mydata$Species == "virginica"`

- `mydata[mydata$Species == "setosa" ,]`
- `mydata[mydata$Species == "versicolor" ,]`
- `mydata[mydata$Species == "virginica" ,]`

- `mydata[mydata$Species == "setosa", 1]`
- `mydata[mydata$Species == "versicolor", 1]`
- `mydata[mydata$Species == "virginica", 1]`

- `mydata <- iris`
- `mydata`

- `mydata[mydata$Species == "setosa", 1:2]`
- `mydata[mydata$Species == "setosa", 1:3]`
- `mydata[mydata$Species == "setosa", 2:4]`

- `# find the subset of the data with the property`

- `subset(mydata, Species == "setosa", select = Sepal.Length)`
- `subset(mydata, Species == "setosa", select = c(Sepal.Length, Sepal.Width))`

初步分析數據

- `data1 <- iris[, 1]`
- `data1`

- `mysort <- sort(data1)` `# sort the data`

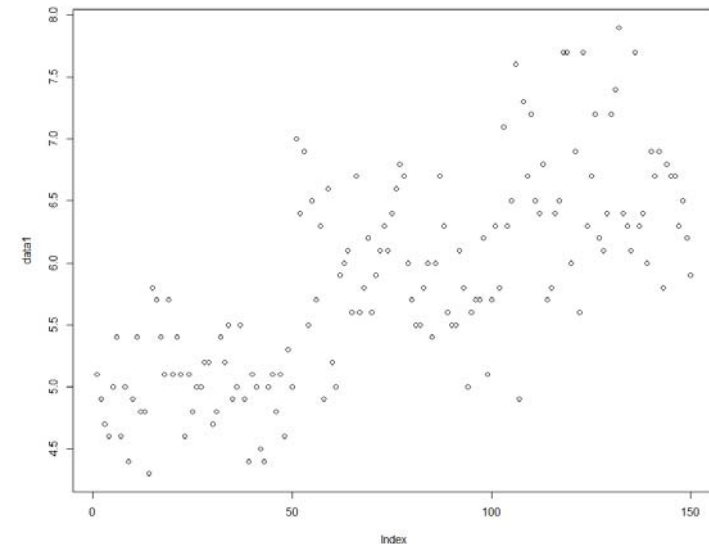
- `mysort[150*0.5]` `# the 50% data`
- `mysort[150*0.25]` `# the 25% data`
- `mysort[150*0.75]` `# the 75% data`

- `mystat <- c(min(data1), mysort[150*0.25], median(data1),
mean(data1), mysort[150*0.75], max(data1), sd(data1))`

- `summary(data1)`

繪製圖形

- `plot(data)`
 - # 個別繪製圖形
 - `data1 <- iris[, 1]`
 - `plot(data1)`
 - `data2 <- iris[, 2]`
 - `plot(data2)`
 - `data3 <- iris[, 3]`
 - `plot(data3)`
 - `data4 <- iris[, 4]`
 - `plot(data4)`
 - `data5 <- iris[, 5]`
 - `plot(data5)`



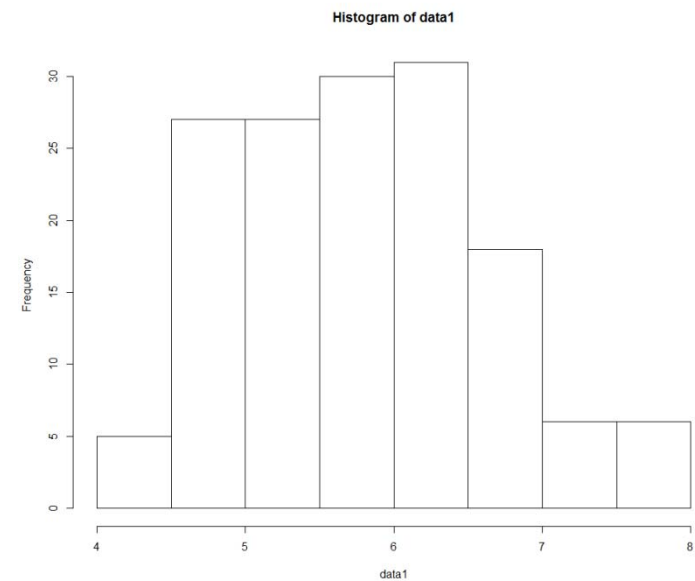
繪製圖形 – histogram 直方圖

計算機程式設計 – 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

- `hist(data)`
 - # 個別繪製圖形
- `hist(data1)`
- `hist(data2)`
- `hist(data3)`
- `hist(data4)`
- `hist(data5)`



基本數據圖示法

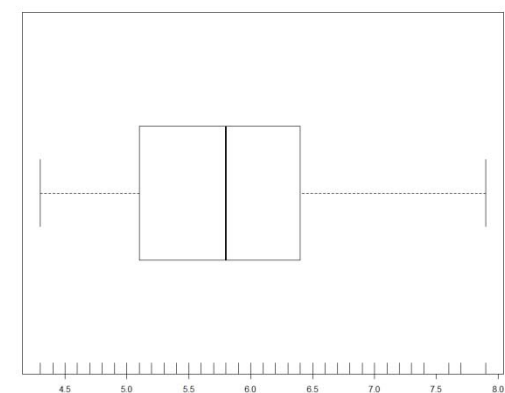
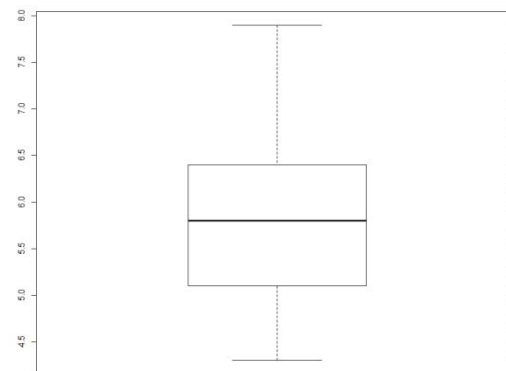
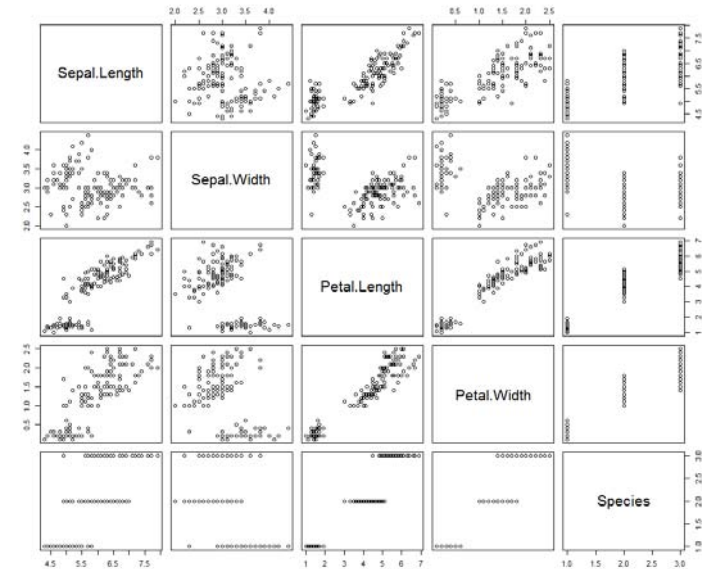
繪圖 – boxplot 盒鬚圖

計算機程式設計 – 2018S

U04: 數據處理-繪圖功能

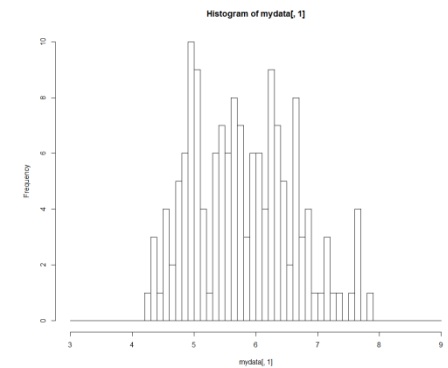
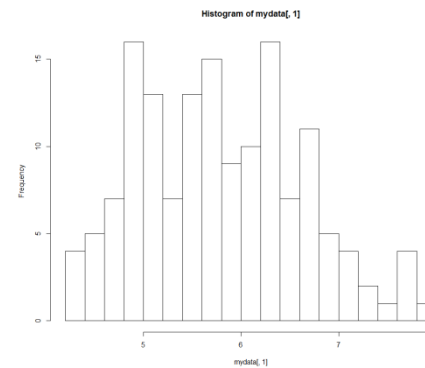
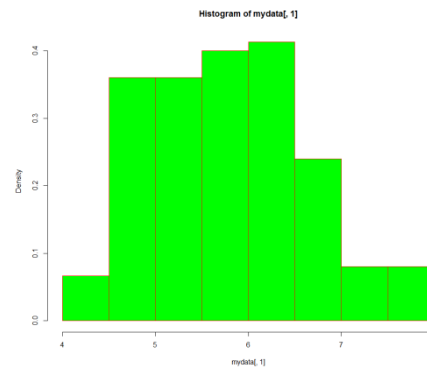
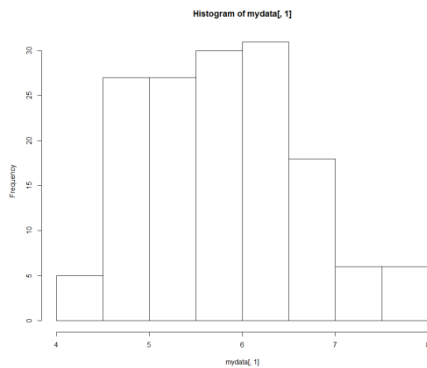
Feng-Li Lian @ NTU-EE

- `mydata <- iris`
- `plot(mydata)`
- `plot(mydata[, 1:4])`
- `boxplot(mydata[, 1])`
- `rug(mydata[, 1], side = 4)`
- `boxplot(mydata[, 1], horizontal = TRUE)`
- `rug(mydata[, 1], side = 1)`



繪圖 – histogram 直方圖

- `hist(mydata[, 1])`
- `hist(mydata[, 1], freq = TRUE)`
- `hist(mydata[, 1], freq = TRUE, breaks = "Sturges")`
- `hist(mydata[, 1], prob = TRUE, breaks = "Sturges", col = "green", border = "red")`
- `hist(mydata[, 1], freq = TRUE, breaks = 20)`
- `hist(mydata[, 1], freq = TRUE, breaks = seq(from=3, to=9, by=0.1))`



繪圖 – stem-leaf 莖葉圖

- `stem(mydata[, 1], scale = 1.0)`
- `stem(mydata[, 1], scale = 0.5)`
- `sum(mydata[, 1]) == 4.4)`
- `sum(mydata[, 1]) == 4.6)`

```
> stem( mydata[ , 1 ], scale = 1.0 )
```

```
The decimal point is 1 digit(s) to the left of the |
```

```
42 | 0
44 | 0000
46 | 000000
48 | 000000000000
50 | 00000000000000000000
52 | 00000
54 | 0000000000000000
56 | 0000000000000000
58 | 0000000000
60 | 00000000000000
62 | 00000000000000
64 | 00000000000000
66 | 0000000000
68 | 0000000
70 | 00
72 | 0000
74 | 0
76 | 00000
78 | 0
```


繪圖 – strip chart 長條記錄圖

計算機程式設計 – 2018S

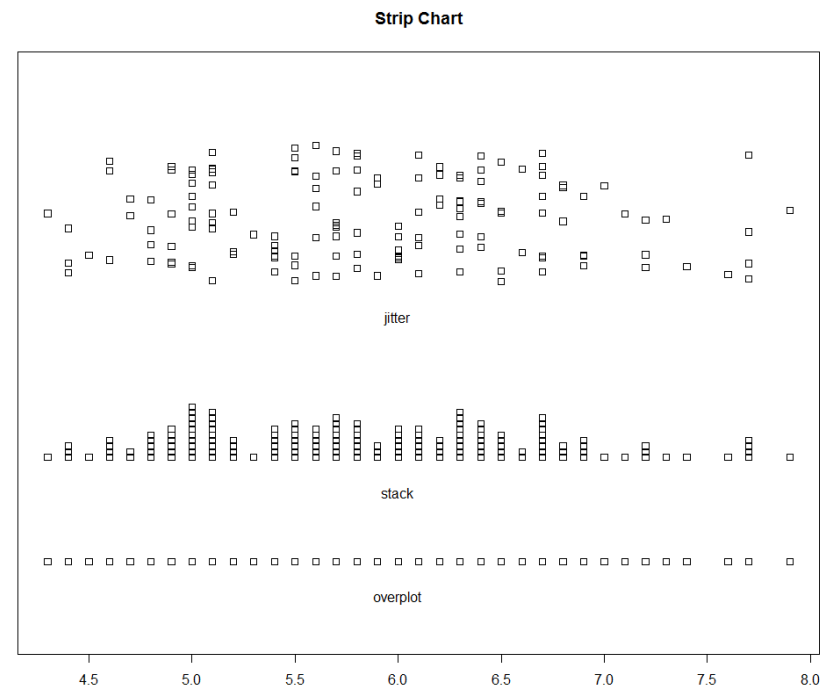
U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

- `stripchart(mydata[, 1], method = "overplot", at = 0.7)`
- `stripchart(mydata[, 1], method = "stack", add = TRUE, at = 0.85)`
- `stripchart(mydata[, 1], method = "jitter", add = TRUE, at = 1.2)`

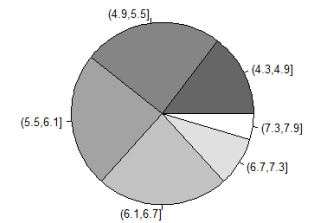
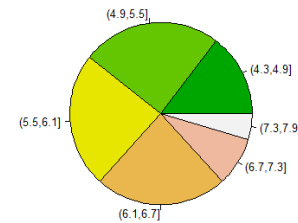
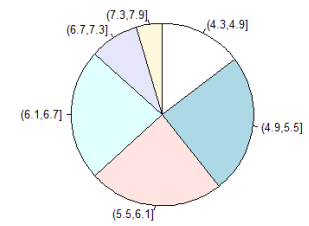
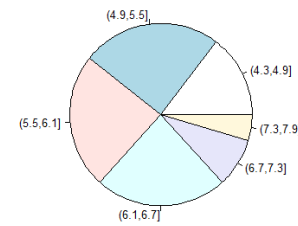
- `text(6, 0.65, "overplot")`
- `text(6, 0.8, "stack")`
- `text(6, 1.05, "jitter")`

- `title(main = "Strip Chart")`



繪圖 – pie chart 圓餅圖

- `x <- cut(mydata[, 1], breaks = 6)`
- `y <- table(x)`
-
- `pie(y)`



- `pie(y)`
- `pie(y, clockwise = TRUE)`
- `pie(y, col = terrain.colors(6))`
- `pie(y, col = gray(seq(from = 0.4, to = 1.0, length = 6)))`

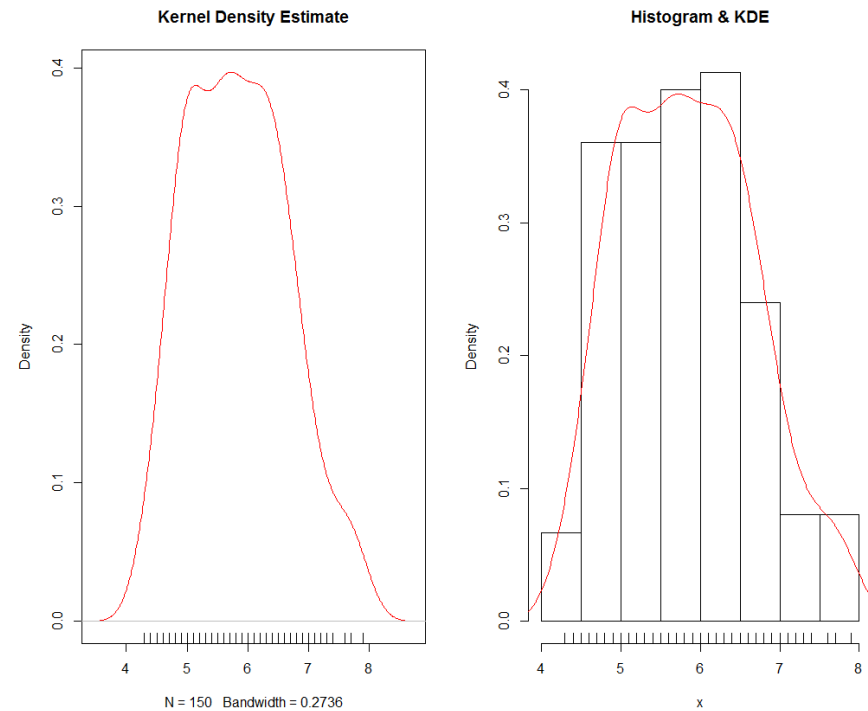
繪圖 – density 機率分布圖

計算機程式設計 – 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

- `x <- mydata[, 1]`



- `plot(density(x), col = "red", main = "Kernel Density Estimate")`
- `rug(x, side = 1)`

- `hist(x, prob = TRUE, breaks = "Sturges", main = "Histogram & KDE")`
- `lines(density(x), col = "red")`
- `rug(x, side = 1)`

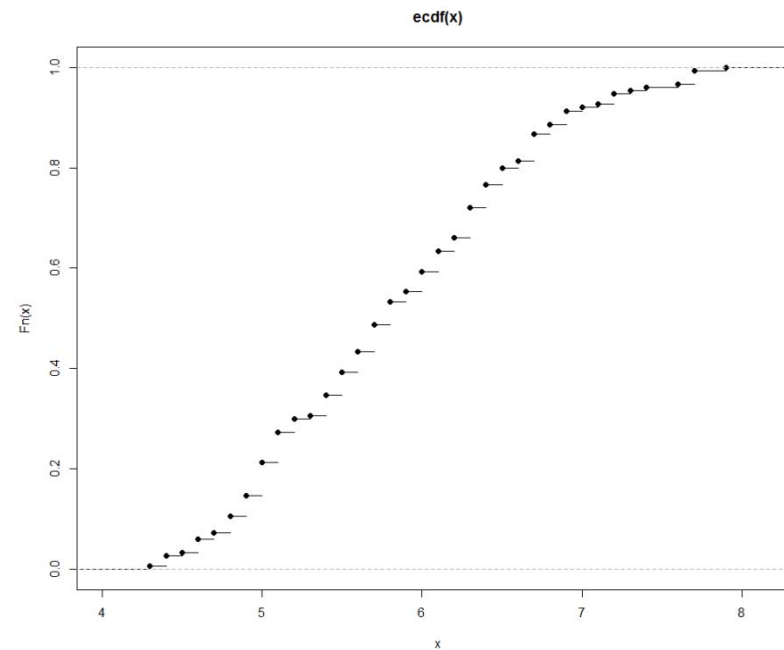
繪圖 – ECDF 經驗累積分布圖

計算機程式設計 – 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

- `x <- mydata[, 1]`
- `plot.ecdf(x)`



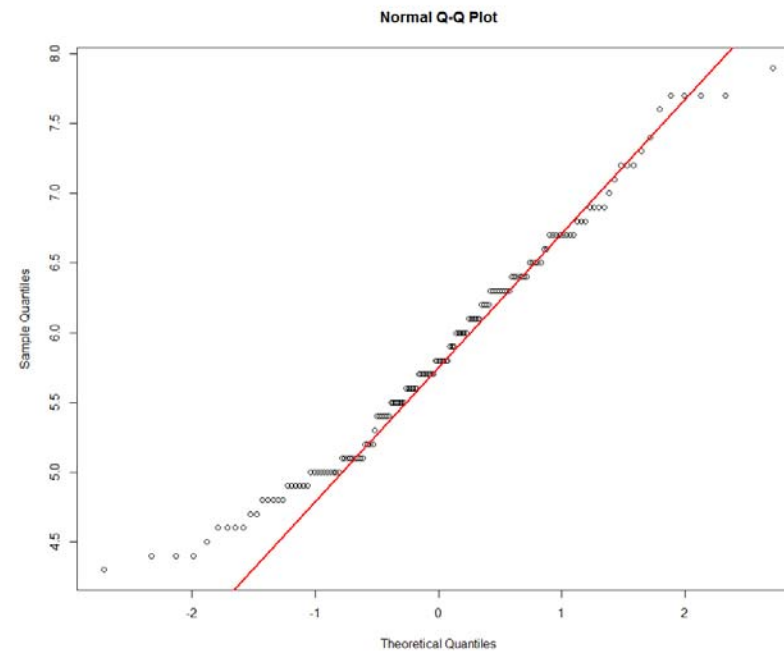
繪圖 – normal QQ 常態機率圖

計算機程式設計 – 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

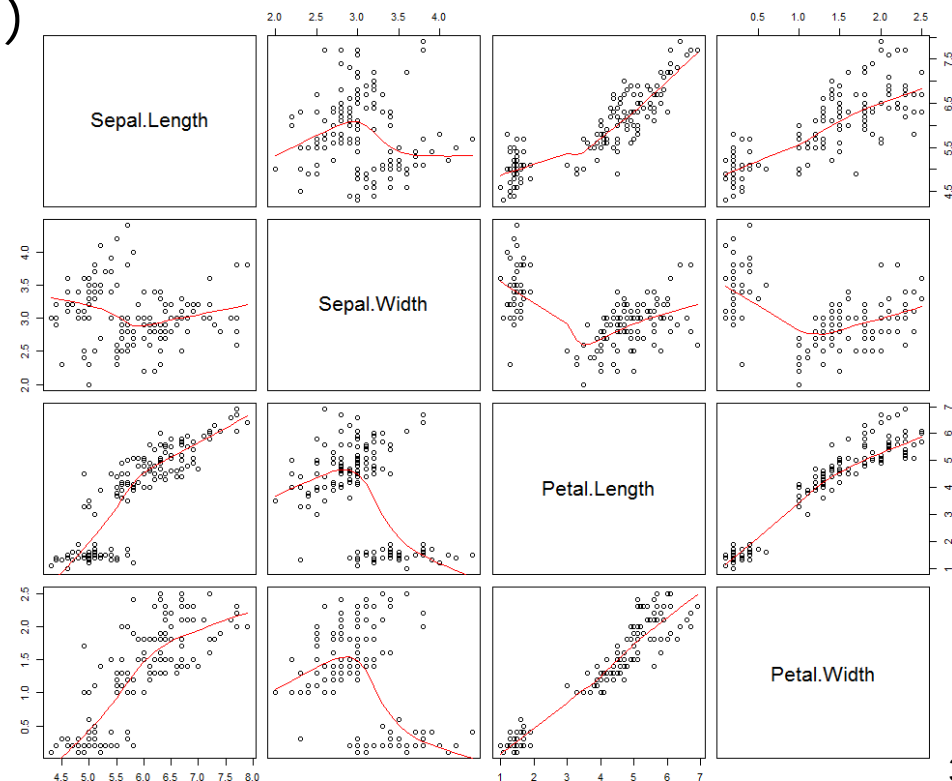
- `x <- mydata[, 1]`
- `qqnorm(x)`
- `qqline(x, col = "red", lwd = 2)`



多維數據繪圖 - 散點圖

- iris
- `x <- iris[, 1:4]`
- `plot(x)`
- `pairs(x)`
- `pairs(x, panel = panel.smooth)`

scatterplot





大綱

計算機程式設計 - 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

作業

HW02：數據處理與繪圖功能

計算機程式設計 - 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

On 3/20, 2017

- 請參考 U04 講義，以及 R code 檔案
- 請從下面資料中，自行挑選一組數據：iris, cars, women, or CO2
- 然後，請用三個以上的指令計算分析一下這組數據，
- 以及挑選三個繪製指令繪製三個圖，
- 請把從頭到尾的執行過程，編輯於 .R 檔之中，並且依序執行這些指令
- 把執行的過程，或者是產生的數據/圖形等，整理到報告檔 (pdf)
- 報告檔中，請編輯：
 - 描述進行的計算或繪圖工作名稱，
 - 所使用的的指令，
 - 產生的結果，數據 and/or 圖形
 - 解釋說明該指令的功能，產生的結果，該結果的意義，特點等

HW02：數據處理與繪圖功能

計算機程式設計 - 2018S

U04: 數據處理-繪圖功能

Feng-Li Lian @ NTU-EE

On 3/20, 2017

- 繳交下面檔案，檔案名稱：[HW02_學號_關鍵字.xxx](#)
 - R 程式檔案：[HW02_B01921001_ComputePlot.R](#)
 - 報告檔案：[HW02_B01921001_ComputePlot.pdf](#)
- 繳交方式與期限：
 - E-mail 上面兩個檔案到：ntucp2018s@gmail.com
 - E-mail 主旨：[HW02 B01921001 ComputePlot](#)
(就是，作業編號 and 您的學號 and 關鍵字)
 - 繳交期限：**3/25 (Sun), 11pm 以前**
- 學習方式：
 - 請至下面網址輸入此次的學習方式所花的時間：
 - https://docs.google.com/forms/d/e/1FAIpQLSdAZ_b-FUtvnNr_14rYQNYejMhDESy6jJ9ESh5XsjFI-DXMIw/viewform?c=0&w=1