106-2: EE4052

通識課程：計算機程式設計之旅

Computer Programming

# Unit 12: 資料間的相關性

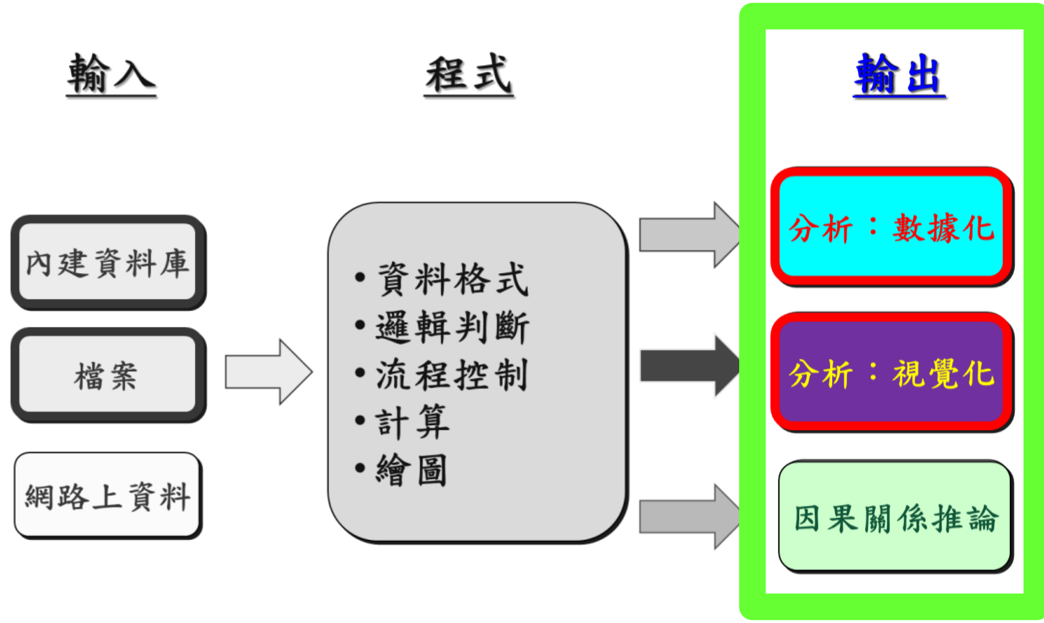
連　豐　力

臺 大 電 機 系

Feb 2018 - Jun 2018

# 課程主題進度

- **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: 資料間的相關性**
- **U13: 探索性資料分析**
- **U14: 資料連結分析**
- **U15: 影像與動畫**



# 大綱

- 資料間的線性關係
- lm: Linear Model
- nhanes2, cars, iris 的線性回歸模型
- 資料間的相關性
- 多維關係繪圖

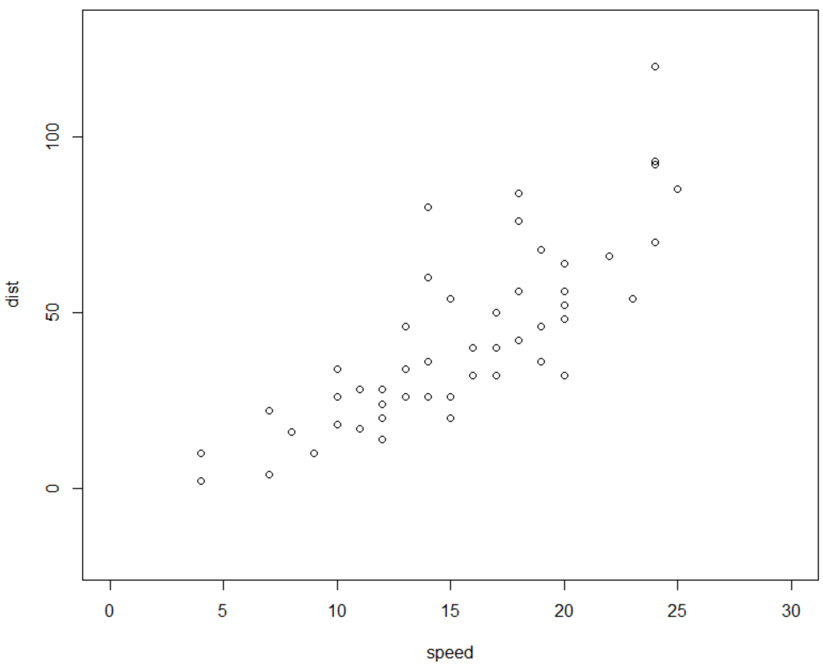
# 大綱

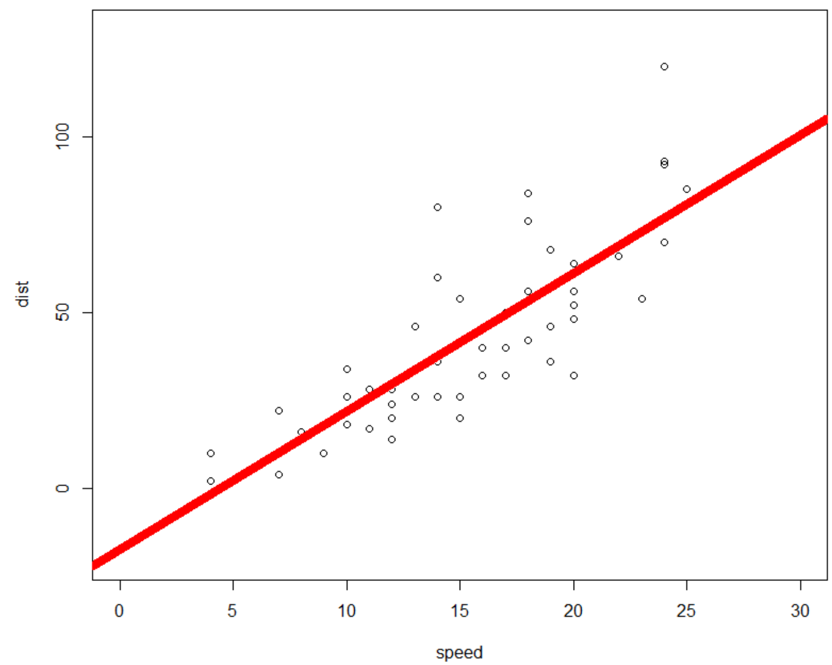
## 資料間的線性關係

# 資料庫：cars



- cars
- plot( cars, xlim = c(0, 30 ), ylim = c(-20, 130 ) )

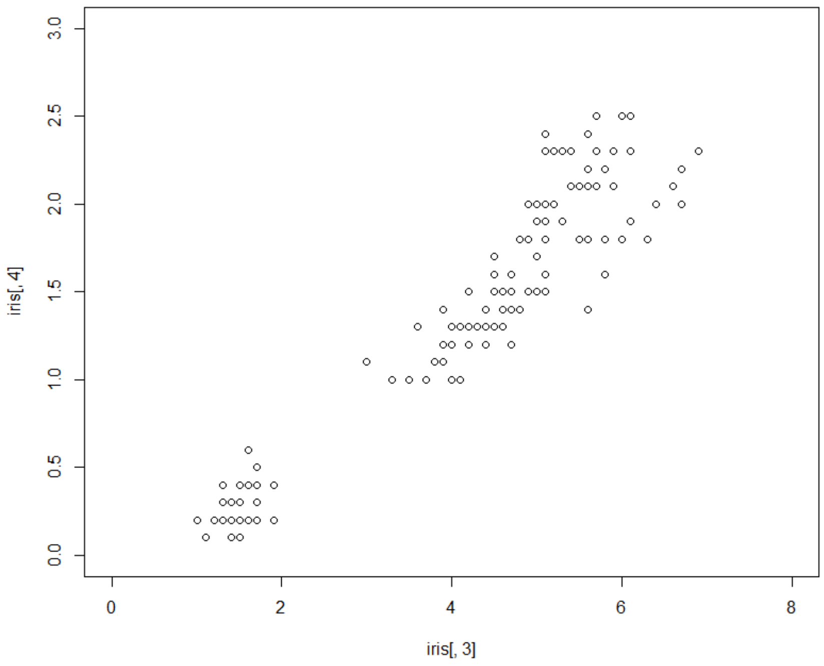


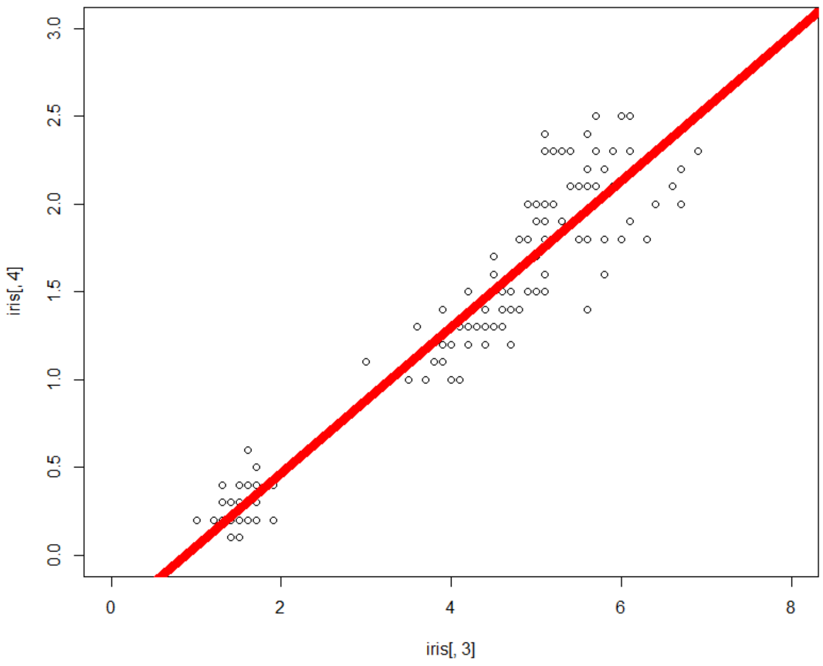




- abline( a = -17.579, b = 3.932, col = "red", lwd = 8 )

# 資料庫： iris

- iris
- plot( iris[ , 3 ], iris[ , 4 ], xlim = c( 0, 8 ), ylim = c( 0, 3 ) )




- abline( a = -0.3631, b = 0.4158, col = "red", lwd = 8 )

# 大綱

## lm: Linear Model

## Least Squares Approximation

# Least Squares Approximation

- 參考資料：http://www.ms.uky.edu/~ma138/Spring15/Curve_fitting.pdf

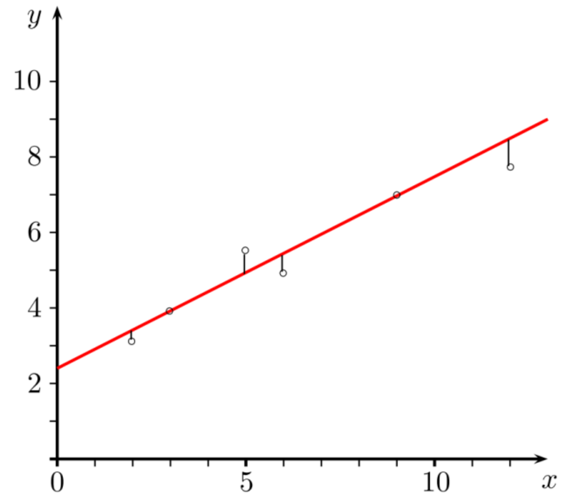



FIGURE 1: Fitting a straight line to data by the method of least squares

$$y = ax + b$$

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \\ \quad\vdots \\ ax_n + b = y_n \end{cases} \quad \leftrightsquigarrow \quad \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\delta_1 = (ax_1 + b) - y_1, \quad \delta_2 = (ax_2 + b) - y_2, \quad \ldots, \quad \delta_n = (ax_n + b) - y_n.$$

$\sqrt{\delta_1^2 + \delta_2^2 + \ldots + \delta_n^2}$ is as small as possible.

$$\widehat{a} = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad \widehat{b} = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - \widehat{a}\sum_{i=1}^{n} x_i\right),$$

$$y = \widehat{a}\, x + \widehat{b}$$

# Least Squares Approximation

- 參考資料：http://www.ms.uky.edu/~ma138/Spring15/Curve_fitting.pdf

| $t$ (sec) | 0.5 | 1.1 | 1.5 | 2.1 | 2.3 |
|---|---|---|---|---|---|
| $T$ (°C) | 32.0 | 33.0 | 34.2 | 35.1 | 35.7 |

$T = at + b,$

$$\begin{cases} 0.5\,a + b &= 32.0 \\ 1.1\,a + b &= 33.0 \\ 1.5\,a + b &= 34.2 \\ 2.1\,a + b &= 35.1 \\ 2.3\,a + b &= 35.7 \end{cases} \quad \leftrightsquigarrow \quad \begin{bmatrix} 0.5 & 1 \\ 1.1 & 1 \\ 1.5 & 1 \\ 2.1 & 1 \\ 2.3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 32.0 \\ 33.0 \\ 34.2 \\ 35.1 \\ 35.7 \end{bmatrix}.$$

$$A^T A = \begin{bmatrix} 0.5 & 1.1 & 1.5 & 2.1 & 2.3 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 1 \\ 1.1 & 1 \\ 1.5 & 1 \\ 2.1 & 1 \\ 2.3 & 1 \end{bmatrix} = \begin{bmatrix} 13.41 & 7.5 \\ 7.5 & 5 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 0.5 & 1.1 & 1.5 & 2.1 & 2.3 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 32.0 \\ 33.0 \\ 34.2 \\ 35.1 \\ 35.7 \end{bmatrix} = \begin{bmatrix} 259.42 \\ 170 \end{bmatrix}$$

$$\begin{bmatrix} 13.41 & 7.5 \\ 7.5 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 259.42 \\ 170 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 13.41 & 7.5 & 259.42 \\ 7.5 & 5 & 170 \end{array}\right] \text{ is equivalent to } \left[\begin{array}{cc|c} 1 & 0 & 2.0463 \\ 0 & 1 & 30.93 \end{array}\right]$$

$\widehat{a} = 2.0463$ and $\widehat{b} = 30.93.$

$T(t) = 2.0463\,t + 30.93$

| year | 1980 | 1985 | 1990 | 1995 |
|---|---|---|---|---|
| population | 227 | 237 | 249 | 262 |

$P(t) = at + b.$

$$\begin{bmatrix} 0 & 1 \\ 5 & 1 \\ 10 & 1 \\ 15 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 227 \\ 237 \\ 249 \\ 262 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 0 & 5 & 10 & 15 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 5 & 1 \\ 10 & 1 \\ 15 & 1 \end{bmatrix} = \begin{bmatrix} 350 & 30 \\ 30 & 4 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 0 & 5 & 10 & 15 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 227 \\ 237 \\ 249 \\ 262 \end{bmatrix} = \begin{bmatrix} 7605 \\ 975 \end{bmatrix}$$

$$\begin{bmatrix} 350 & 30 \\ 30 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 7605 \\ 975 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 350 & 30 & 7605 \\ 30 & 4 & 975 \end{array}\right] \text{ is equivalent to } \left[\begin{array}{cc|c} 1 & 0 & 117/50 \\ 0 & 1 & 1131/5 \end{array}\right]$$

$P(t) = 117/50 \cdot t + 1131/5.$

# 大綱

## 三個資料庫

## nhanes2, cars, iris

# 資料庫

- install.packages( "mice" )　　# 安裝 mice 軟體套件
- library( mice )　　# 載入 mice 軟體套件
- data( nhanes2 )
- nrow( nhanes2 )　　# nhanes2 資料集的橫列數
- ncol( nhanes2 )　　# nhanes2 資料集的直行數
- summary( nhanes2 )　　# nhanes2 資料集的概括資訊
- head( nhanes2 )

```
> head( nhanes2 )
    age  bmi  hyp chl
1 20-39   NA <NA>  NA
2 40-59 22.7   no 187
3 20-39   NA   no 187
4 60-99   NA <NA>  NA
5 20-39 20.4   no 113
6 60-99   NA <NA> 184
```

```
> summary( nhanes2 )
    age          bmi             hyp           chl       
 20-39:12   Min.   :20.40   no  :13   Min.   :113.0  
 40-59: 7   1st Qu.:22.65   yes : 4   1st Qu.:185.0  
 60-99: 6   Median :26.75   NA's: 8   Median :187.0  
            Mean   :26.56             Mean   :191.4  
            3rd Qu.:28.93             3rd Qu.:212.0  
            Max.   :35.30             Max.   :284.0  
            NA's   :9                 NA's   :10     
```

# 線性回歸模型預測數值

- data0 <- nhanes2 # 針對第2, 4組數據
- subNA <- which( is.na( nhanes2[ , 4 ] ) == TRUE | is.na( nhanes2[ , 2 ] ) == TRUE )
- dataOK <- nhanes2[ -subNA, ]
- dataOK
- dataNA <- nhanes2[ subNA, ]
- dataNA

```
> lm_chl_bmi

Call:
lm(formula = chl ~ bmi, data = dataOK)

Coefficients:
(Intercept)          bmi
     87.130        3.963
```

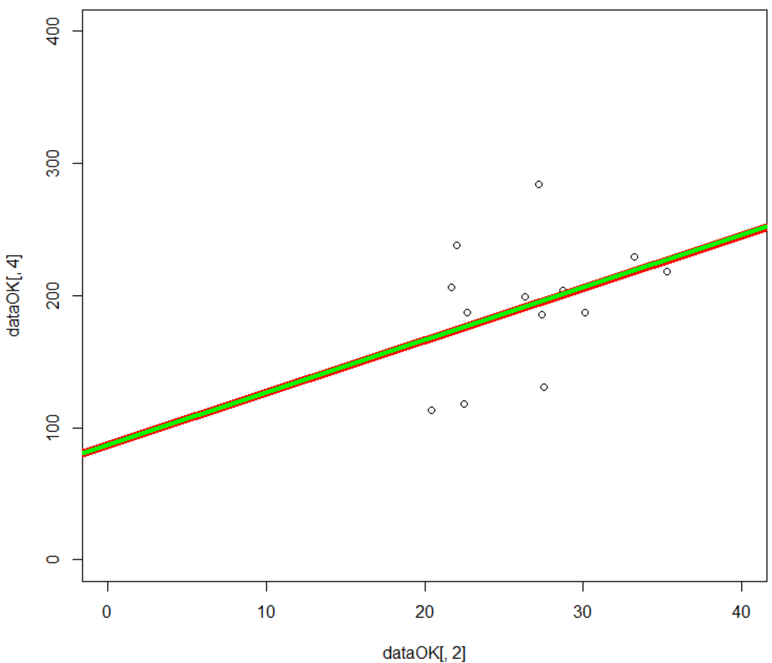
```
chl = 3.963 * bmi + 87.130
```

- lm_chl_bmi <- lm( chl ~ bmi, data = dataOK )
  - # 利用 dataOK 中 bmi 為引數，chl 為因變數，建構線性回歸模型

# 畫 y = b x + a 的直線

- abline( )　　　　# 畫 y = b x + a 的直線

- plot( dataOK[ , 2 ], dataOK[ , 4 ], xlim = c( 0, 40 ), ylim = c( 0, 400 ) )
- abline( a = 87.130, b = 3.963, col = "red", lwd = 8 )
- abline( lm_chl_bmi, col = "green", lwd = 4 )

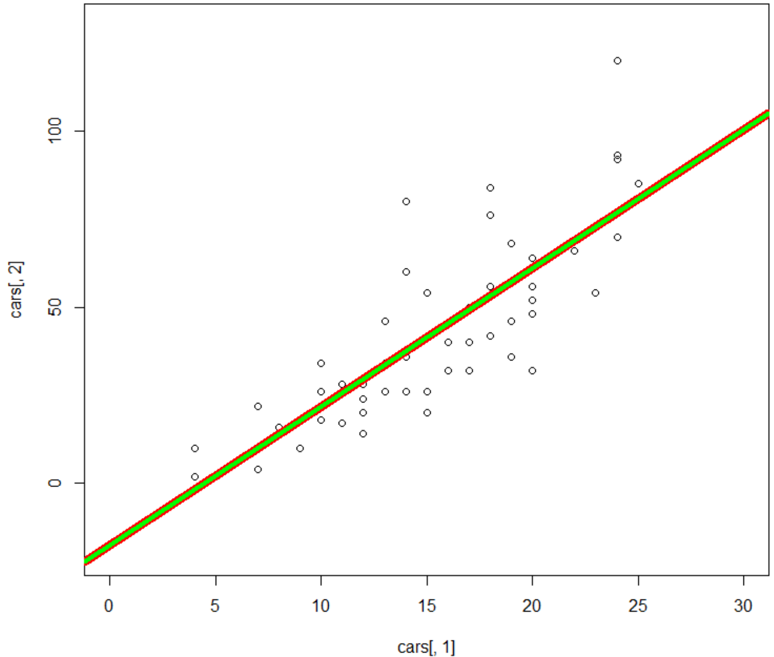



chl = 3.963 * bmi + 87.130

```
> lm_chl_bmi

Call:
lm(formula = chl ~ bmi, data = dataOK)

Coefficients:
(Intercept)          bmi
     87.130        3.963
```

# 另一個資料：cars

- cars
- plot( cars[ , 1 ], cars[ , 2 ], xlim = c(0, 30 ), ylim = c(-20, 130 ) )
- lm_cars <- lm( dist ~ speed, data = cars )
- lm_cars
- abline( a = -17.579, b = 3.932, col = "red", lwd = 8 )
- abline( lm_cars, col = "green", lwd = 4 )

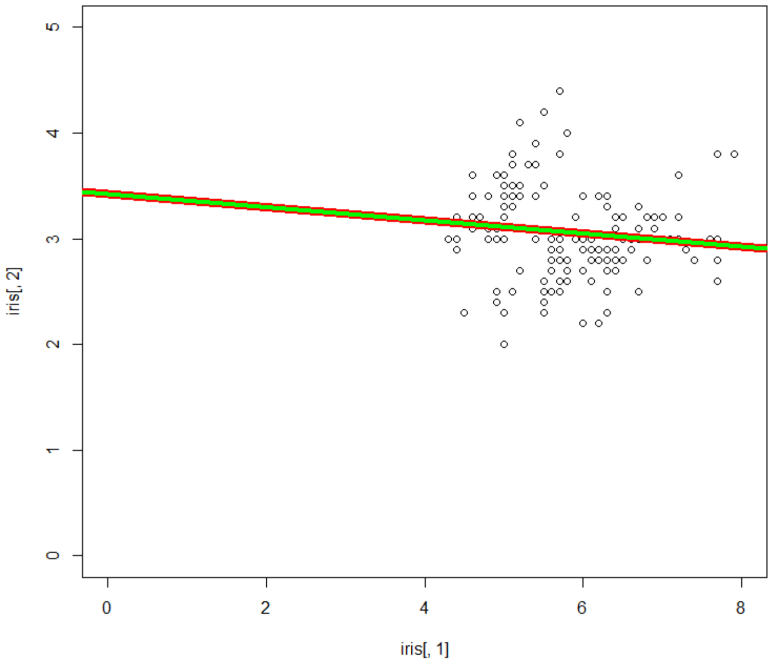



```
chl = 3.932 * speed - 17.579
```

```
> lm_cars

Call:
lm(formula = dist ~ speed, data = cars)

Coefficients:
(Intercept)        speed
    -17.579        3.932
```

# 另一個資料：iris

- iris
- plot( iris[ , 1 ], iris[ , 2 ], xlim = c( 0, 8 ), ylim = c( 0, 5 ) )
- lm_iris_1 <- lm( Sepal.Width ~ Sepal.Length, data = iris )
- lm_iris_1
- abline( a = 3.41895, b = -0.06188 , col = "red", lwd = 8 )
- abline( lm_iris_1, col = "green", lwd = 4 )

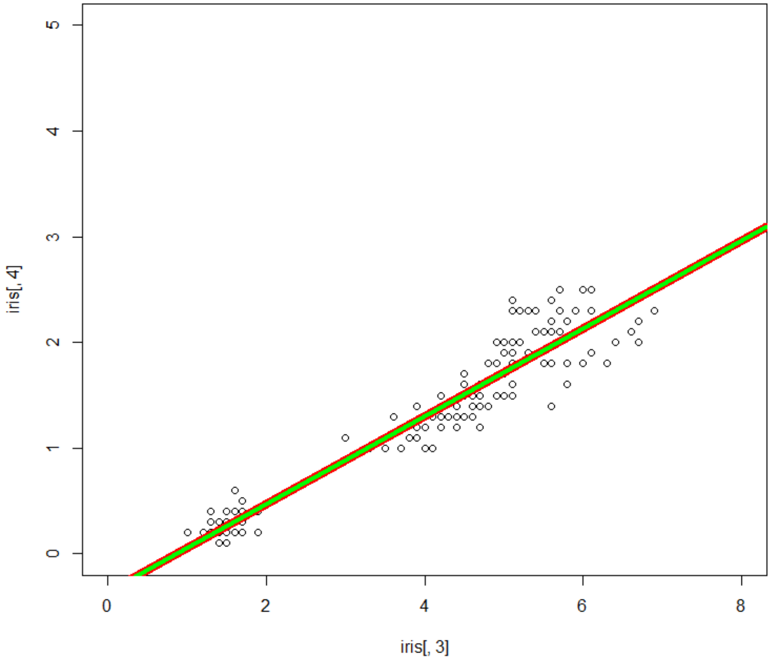



```
Sepal.Width = -0.06188 * Sepal.Length + 3.41895
```

```
> lm_iris_1

Call:
lm(formula = Sepal.Width ~ Sepal.Length,
data = iris)

Coefficients:
 (Intercept)  Sepal.Length
     3.41895      -0.06188
```

# 另一個資料：iris

- iris
- plot( iris[ , 3 ], iris[ , 4 ], xlim = c( 0, 8 ), ylim = c( 0, 5 ) )
- lm_iris_2 <- lm( Petal.Width ~ Petal.Length, data = iris )
- lm_iris_2
- abline( a = -0.3631, b = 0.4158, col = "red", lwd = 8 )
- abline( lm_iris_2, col = "green", lwd = 4 )




```
Petal.Width = 0.4158 * Petal.Length - 0.3631
```

```
> lm_iris_2

Call:
lm(formula = Petal.Width ~ Petal.Length,
data = iris)

Coefficients:
 (Intercept)  Petal.Length
     -0.3631        0.4158
```

# 另一個資料：iris

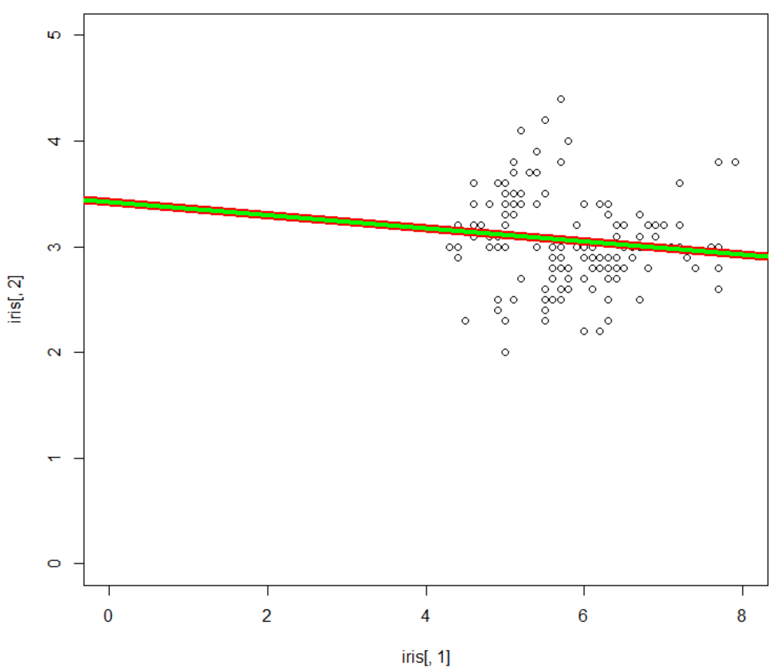
- iris

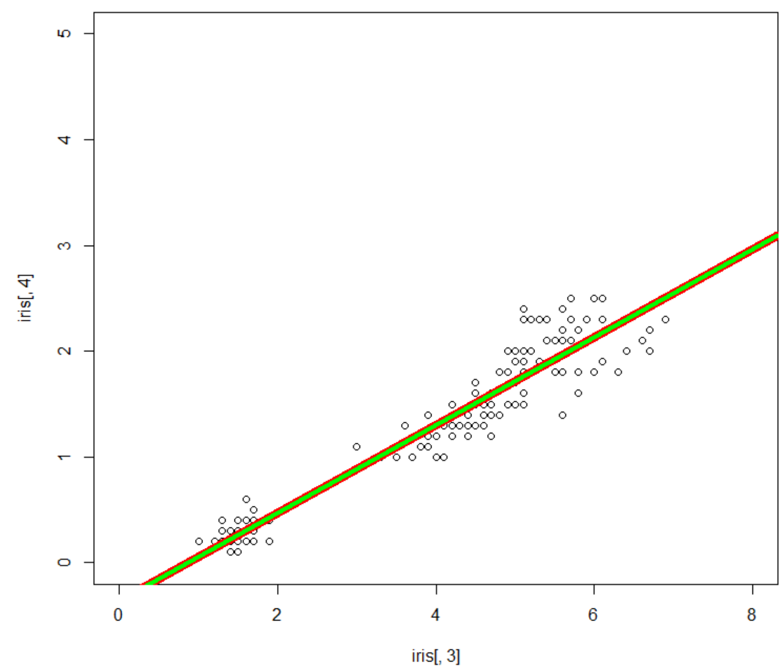
```
> lm_iris_1

Call:
lm(formula = Sepal.Width ~ Sepal.Length,
data = iris)

Coefficients:
 (Intercept)  Sepal.Length
     3.41895      -0.06188
```

```
> lm_iris_2

Call:
lm(formula = Petal.Width ~ Petal.Length,
data = iris)

Coefficients:
 (Intercept)  Petal.Length
     -0.3631        0.4158
```

```
Sepal.Width = -0.06188 * Sepal.Length + 3.41895
```

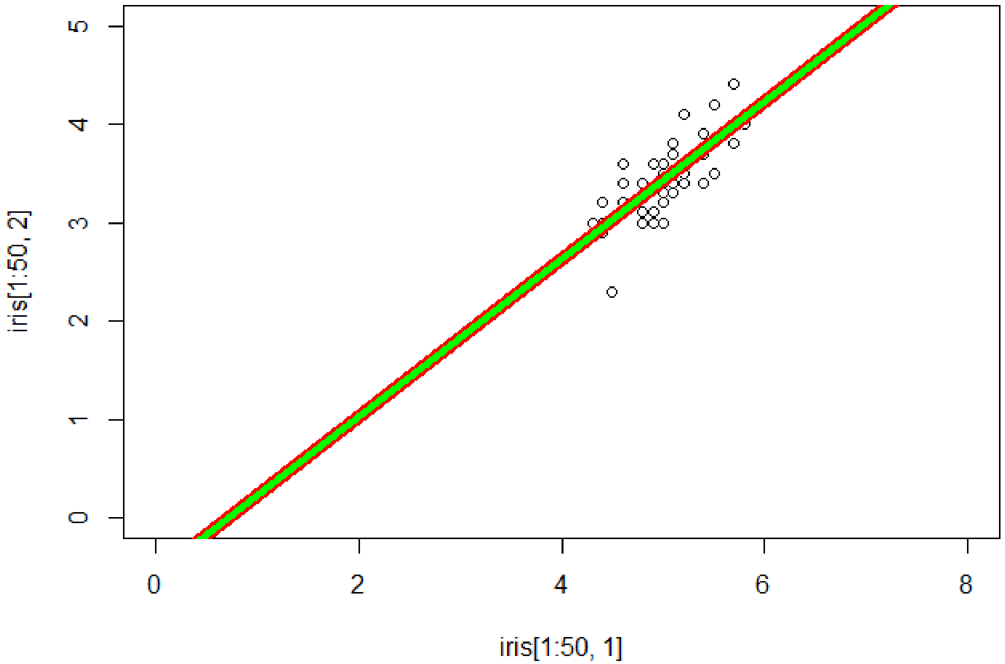
```
Petal.Width = 0.4158 * Petal.Length - 0.3631
```



# 另一個資料：iris, 依照種類

- iris
- plot( iris[ 1:50, 1 ], iris[ 1:50, 2 ], xlim = c( 0, 8 ), ylim = c( 0, 5 ) )
- lm_iris_11 <- lm( Sepal.Width ~ Sepal.Length, data = iris[ 1:50, ] )
- abline( a = -0.5694, b = 0.7985, col = "red", lwd = 8 )
- abline( lm_iris_11, col = "green", lwd = 4 )

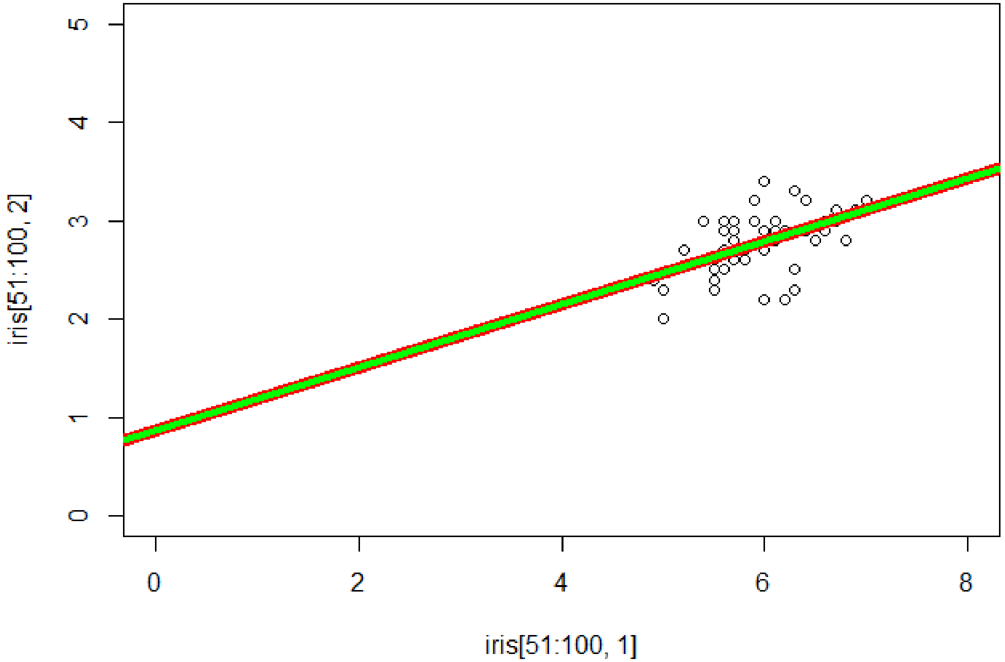



```
Sepal.Width = 0.7985 * Sepal.Length - 0.5694
```

```
> lm_iris_11

Call:
lm(formula = Sepal.Width ~ Sepal.Length,
data = iris[1:50, ] )

Coefficients:
 (Intercept)  Sepal.Length
     -0.5694        0.7985
```

# 另一個資料：iris, 依照種類

- iris
- plot( iris[ 51:100, 1 ], iris[ 51:100, 2 ], xlim = c( 0, 8 ), ylim = c( 0, 5 ) )
- lm_iris_12 <- lm( Sepal.Width ~ Sepal.Length, data = iris[ 51:100, ] )
- abline( a = 0.8721, b = 0.3197, col = "red", lwd = 8 )
- abline( lm_iris_12, col = "green", lwd = 4 )

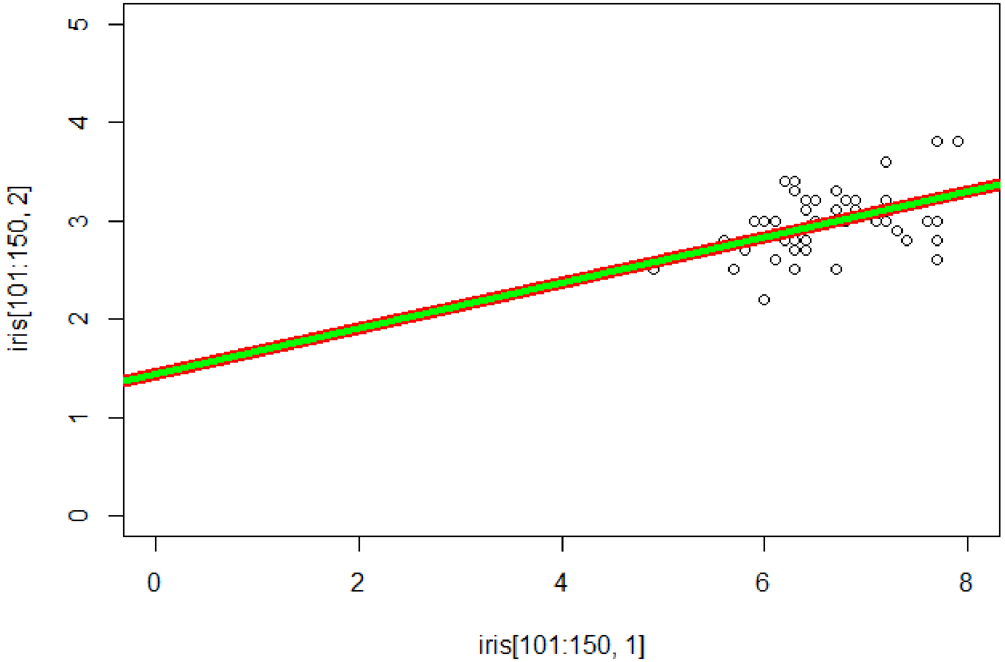



```
Sepal.Width = 0.3197 * Sepal.Length + 0.8721
```

```
> lm_iris_12

Call:
lm(formula = Sepal.Width ~ Sepal.Length,
data = iris[51:100, ] )

Coefficients:
 (Intercept)  Sepal.Length
      0.8721        0.3197
```

# 另一個資料：iris, 依照種類

- iris
- plot( iris[ 101:150, 1 ], iris[ 101:150, 2 ], xlim = c(0,8), ylim = c(0,5) )
- lm_iris_13 <- lm( Sepal.Width ~ Sepal.Length, data = iris[ 101:150, ] )
- abline( a = 1.4463, b = 0.2319, col = "red", lwd = 8 )
- abline( lm_iris_13, col = "green", lwd = 4 )




```
Sepal.Width = 0.3197 * Sepal.Length + 1.4463
```

```
> lm_iris_13

Call:
lm(formula = Sepal.Width ~ Sepal.Length,
data = iris[101:150, ] )

Coefficients:
 (Intercept)  Sepal.Length
      1.4463        0.2319
```

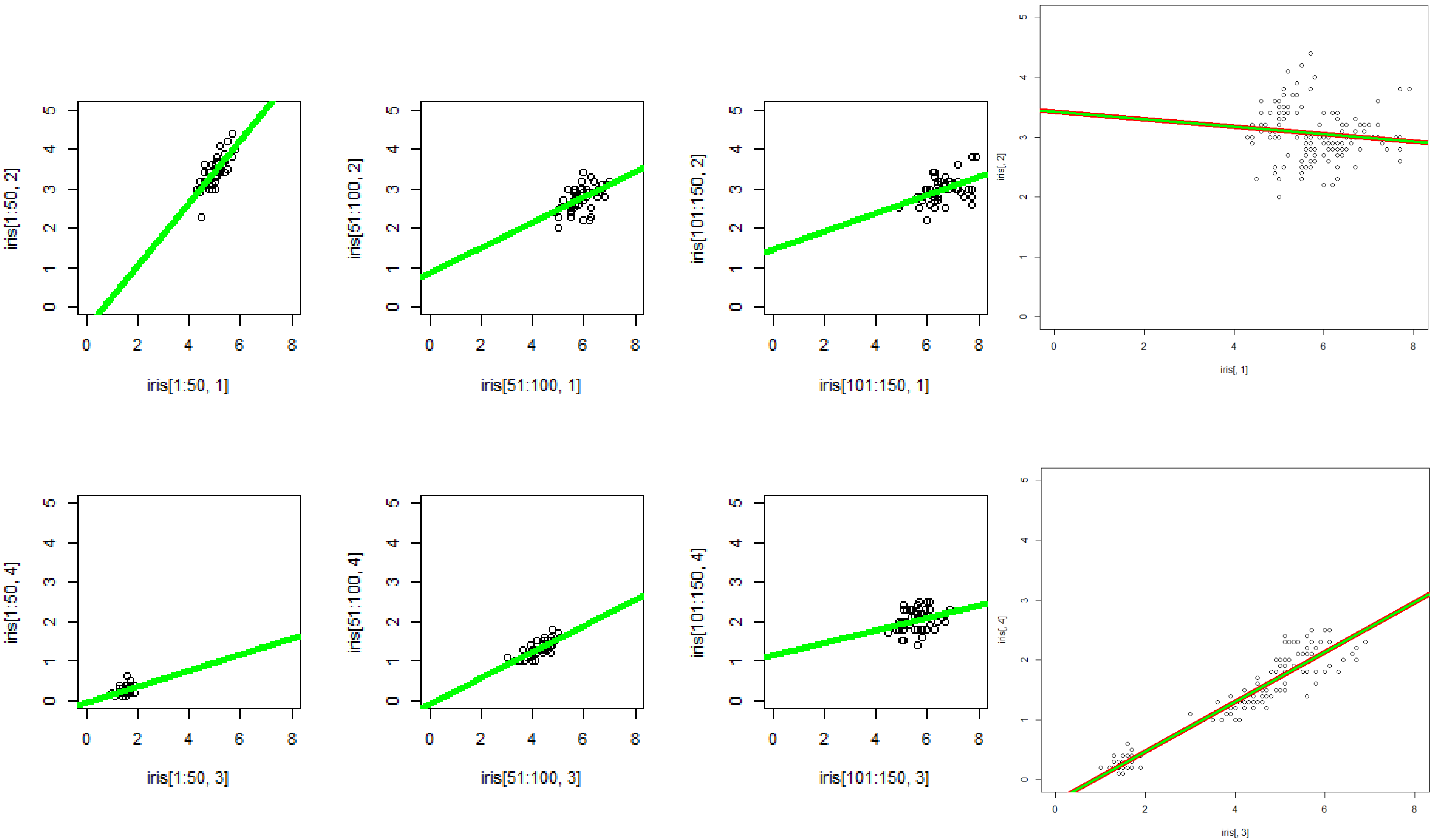
# 另一個資料：iris, 依照種類

- lm_iris_11 <- lm( Sepal.Width ~ Sepal.Length, data = iris[ 1:50, ] )
- lm_iris_12 <- lm( Sepal.Width ~ Sepal.Length, data = iris[ 51:100, ] )
- lm_iris_13 <- lm( Sepal.Width ~ Sepal.Length, data = iris[ 101:150, ] )
- lm_iris_21 <- lm( Petal.Width ~ Petal.Length, data = iris[ 1:50, ] )
- lm_iris_22 <- lm( Petal.Width ~ Petal.Length, data = iris[ 51:100, ] )
- lm_iris_23 <- lm( Petal.Width ~ Petal.Length, data = iris[ 101:150, ] )

# 另一個資料：iris, 依照種類

- layout( matrix( 1:6, nrow = 2, byrow = T ) )
- plot( iris[ 1:50, 1 ], iris[ 1:50, 2 ], xlim = c( 0, 8 ), ylim = c( 0, 5 ) )
- abline( lm_iris_11, col = "green", lwd = 4 )
- plot( iris[ 51:100, 1 ], iris[ 51:100, 2 ], xlim = c( 0, 8 ), ylim = c( 0, 5 ) )
- abline( lm_iris_12, col = "green", lwd = 4 )
- plot( iris[ 101:150, 1 ], iris[ 101:150, 2 ], xlim = c(0,8), ylim = c(0,5) )
- abline( lm_iris_13, col = "green", lwd = 4 )
- plot( iris[ 1:50, 3 ], iris[ 1:50, 4 ], xlim = c( 0, 8 ), ylim = c( 0, 5 ) )
- abline( lm_iris_21, col = "green", lwd = 4 )
- plot( iris[ 51:100, 3 ], iris[ 51:100, 4 ], xlim = c( 0, 8 ), ylim = c( 0, 5 ) )
- abline( lm_iris_22, col = "green", lwd = 4 )
- plot( iris[ 101:150, 3 ], iris[ 101:150, 4 ], xlim = c(0,8), ylim = c(0,5) )
- abline( lm_iris_23, col = "green", lwd = 4 )

# 另一個資料：iris, 依照種類



# 大綱

## 資料間的相關性

# 相關性

# cor( ), correlation 相關係數

cor( x, y )

cor_matrix <- cor( data_all, use = "pairwise" )

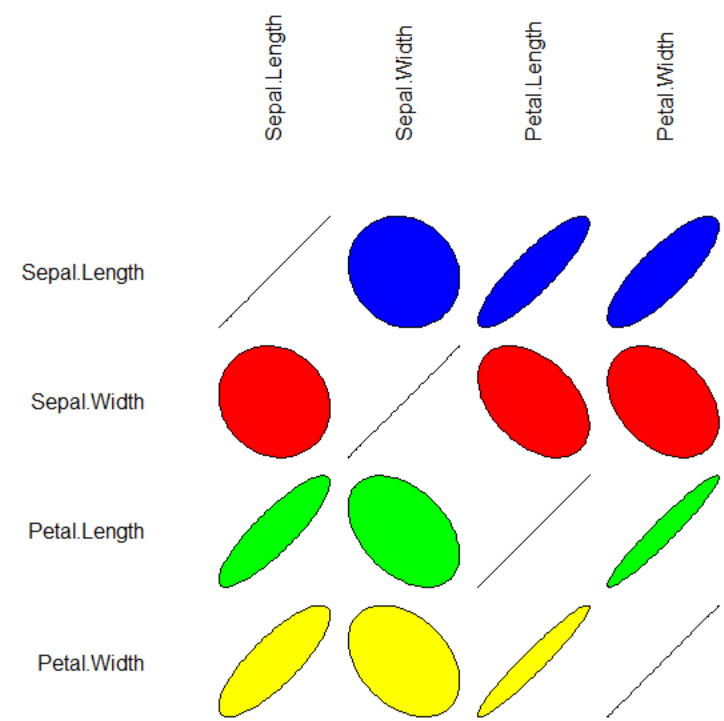
cor_iris <- cor( iris[, 1:4], use = "pairwise")

cor_iris

```
> cor_iris

             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```

# 相關性

```
# plotcoor( ), 繪製相關圖

install.packages( "ellipse" )

library( ellipse )

plotcorr( cor_iris, col = c( "blue", "red", "green", "yellow" ) )
```




```
> cor_iris

             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```

# 相關性

```
# use weather dataset

install.packages( "rattle.data" )

library( rattle.data )

data( weather )

head( weather[ , 12:21] )          # 12 to 21 variable names, values
```

```
> head( weather[ , 12:21] )

  WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm
1            6           20          68          29      1019.7      1015.0        7        7    14.4    23.6
2            4           17          80          36      1012.4      1008.4        5        3    17.5    25.7
3            6            6          82          69      1009.5      1007.2        8        7    15.4    20.2
4           30           24          62          56      1005.5      1007.0        2        7    13.5    14.1
5           20           28          68          49      1018.3      1018.5        7        7    11.1    15.4
6           20           24          70          57      1023.8      1021.7        7        5    10.9    14.8
```

# 相關性

# correlation matrix 相關係數矩陣

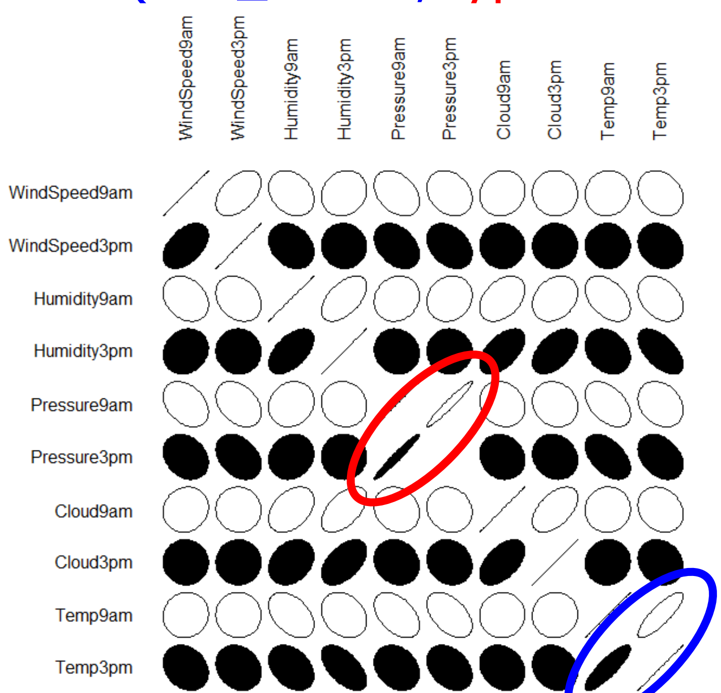
```
var <- c( 12:21 )

cor_matrix <- cor( weather[ var ], use = "pairwise" )
```

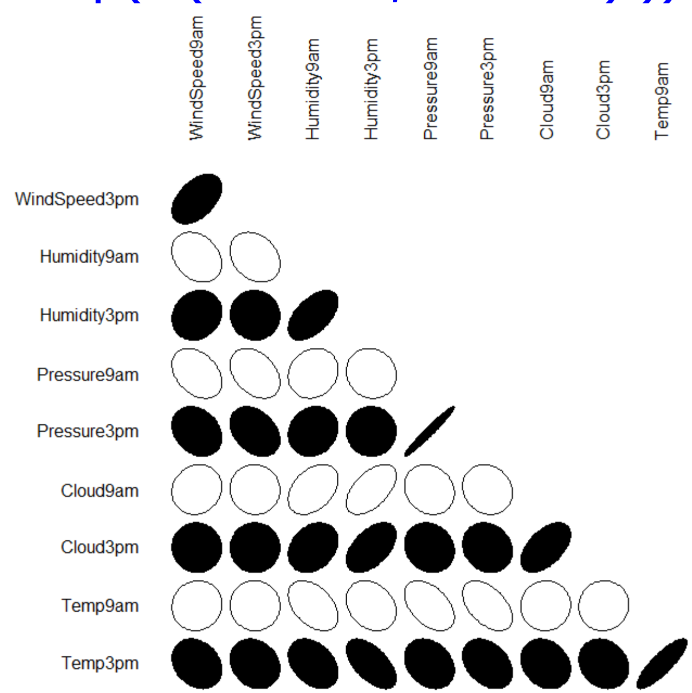
```
> cor_matrix
```

| | WindSpeed9am | WindSpeed3pm | Humidity9am | Humidity3pm | Pressure9am | Pressure3pm | Cloud9am | Cloud3pm | Temp9am | Temp3pm |
|---|---|---|---|---|---|---|---|---|---|---|
| WindSpeed9am | 1.00000000 | 0.47296617 | -0.2706229 | 0.14665712 | -0.35633183 | -0.24795238 | 0.10184246 | -0.02247149 | 0.06407405 | -0.2351864 |
| WindSpeed3pm | 0.47296617 | 1.00000000 | -0.2660925 | -0.02636775 | -0.35980011 | -0.33732535 | -0.02642642 | 0.00720724 | -0.01776636 | -0.1875697 |
| Humidity9am | -0.27062286 | -0.26609247 | 1.0000000 | 0.54671844 | 0.13572697 | 0.13442050 | 0.39284158 | 0.27193809 | -0.43655057 | -0.3551186 |
| Humidity3pm | 0.14665712 | -0.02636775 | 0.5467184 | 1.00000000 | -0.08794614 | -0.01005189 | 0.55163264 | 0.51010790 | -0.25568147 | -0.5816761 |
| Pressure9am | -0.35633183 | -0.35980011 | 0.1357270 | -0.08794614 | 1.00000000 | 0.96789496 | -0.15755279 | -0.14100043 | -0.46041819 | -0.2536738 |
| Pressure3pm | -0.24795238 | -0.33732535 | 0.1344205 | -0.01005189 | 0.96789496 | 1.00000000 | -0.12894408 | -0.14383718 | -0.49263629 | -0.3454853 |
| Cloud9am | 0.10184246 | -0.02642642 | 0.3928416 | 0.55163264 | -0.15755279 | -0.12894408 | 1.00000000 | 0.52521793 | 0.02104135 | -0.2023440 |
| Cloud3pm | -0.02247149 | 0.00720724 | 0.2719381 | 0.51010790 | -0.14100043 | -0.14383718 | 0.52521793 | 1.00000000 | 0.04094519 | -0.1728142 |
| Temp9am | 0.06407405 | -0.01776636 | -0.4365506 | -0.25568147 | -0.46041819 | -0.49263629 | 0.02104135 | 0.04094519 | 1.00000000 | 0.8444058 |
| Temp3pm | -0.23518635 | -0.18756965 | -0.3551186 | -0.58167615 | -0.25367375 | -0.34548531 | -0.20234405 | -0.17281423 | 0.84440581 | 1.0000000 |

# 相關性

```
# plotcoor( ), 繪製相關圖
install.packages( "ellipse" )
library( ellipse )
plotcorr( cor_matrix, col = rep( c( "white", "black" ) ))
plotcorr( cor_matrix, type = "lower", col = rep( c( "white", "black" ) ))
```



# 大綱

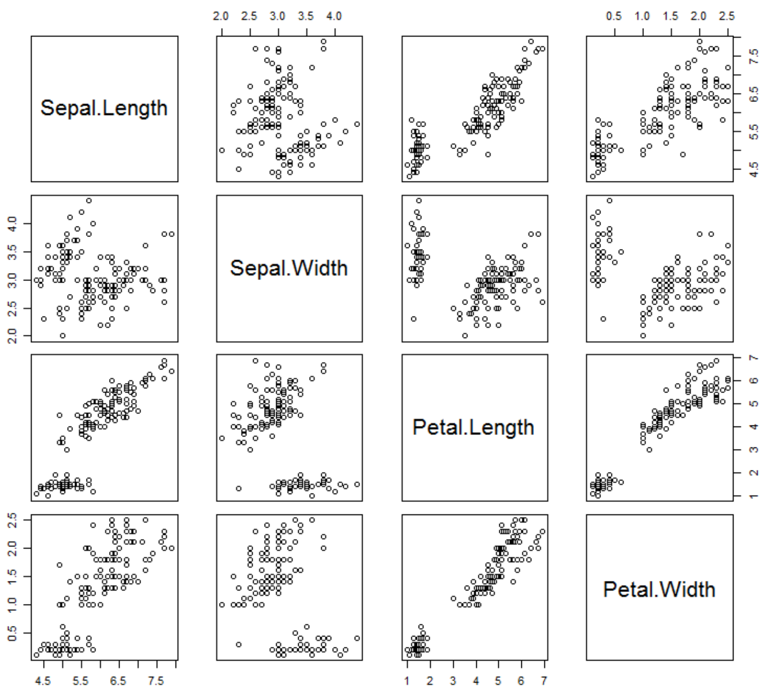


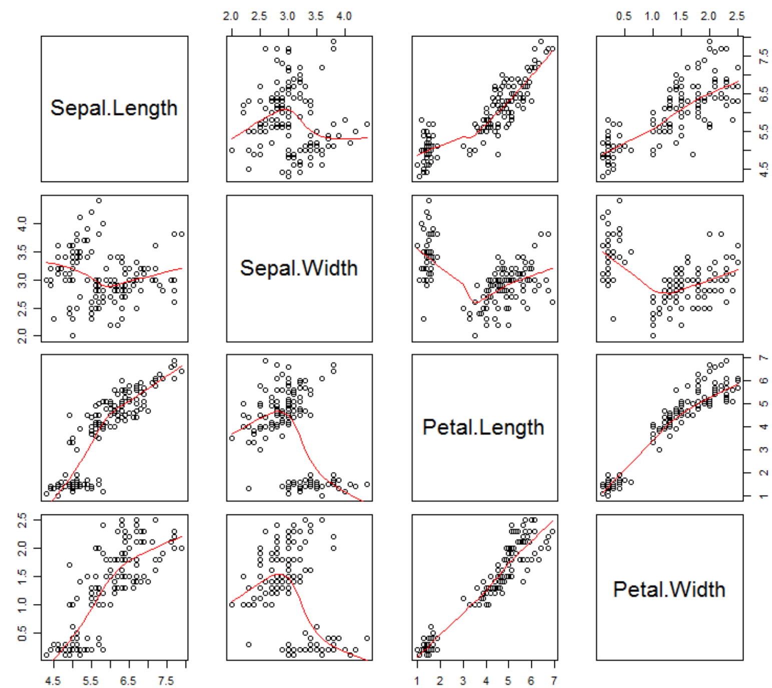
## 多維關係繪圖

# 資料庫：iris

- iris
- x <- iris[ , 1:4 ]

- plot( x )
- pairs( x )
- pairs( x, panel = panel.smooth )



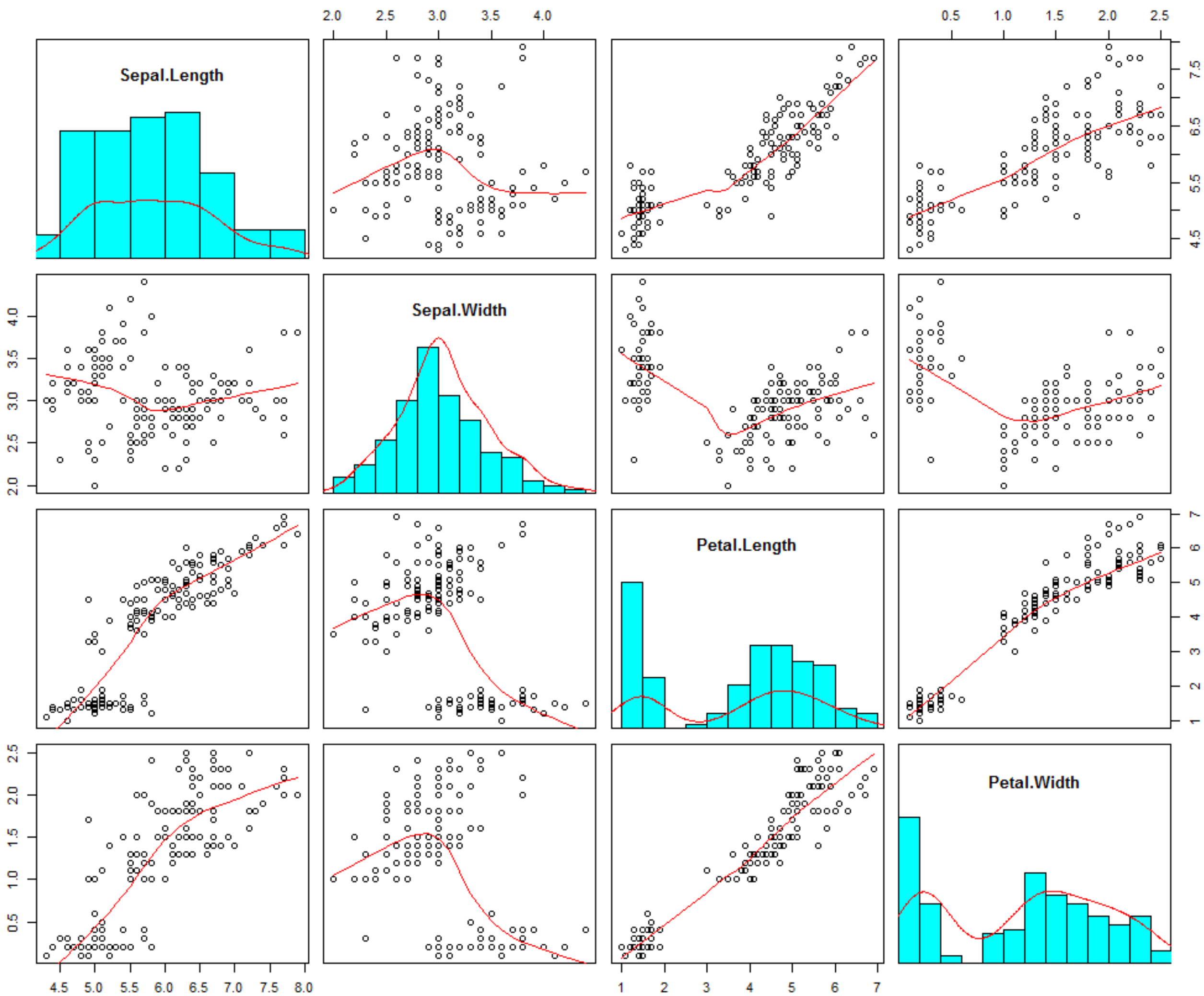
# 多維繪圖 – 散點 直方 核密度

scatterplot

```
iris
x <- iris[ , 1:4 ]

panel.hist <- function(x, ...) {
    usr <- par("usr"); on.exit(par(usr))
    par(usr = c(usr[1:2], 0, 1.5) )
    h <- hist(x, plot = FALSE)
    breaks <- h$breaks; nB <- length(breaks)
    y <- h$counts; y <- y / max(y)
    rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
      lines(density(x, na.rm = TRUE), col = "red")
}

pairs( x, panel = panel.smooth, pch = 1, bg = "lightcyan",
     diag.panel = panel.hist, font.labels = 2, cex.labels = 1.2 )
```

# 多維繪圖 – 散點 直方 核密度

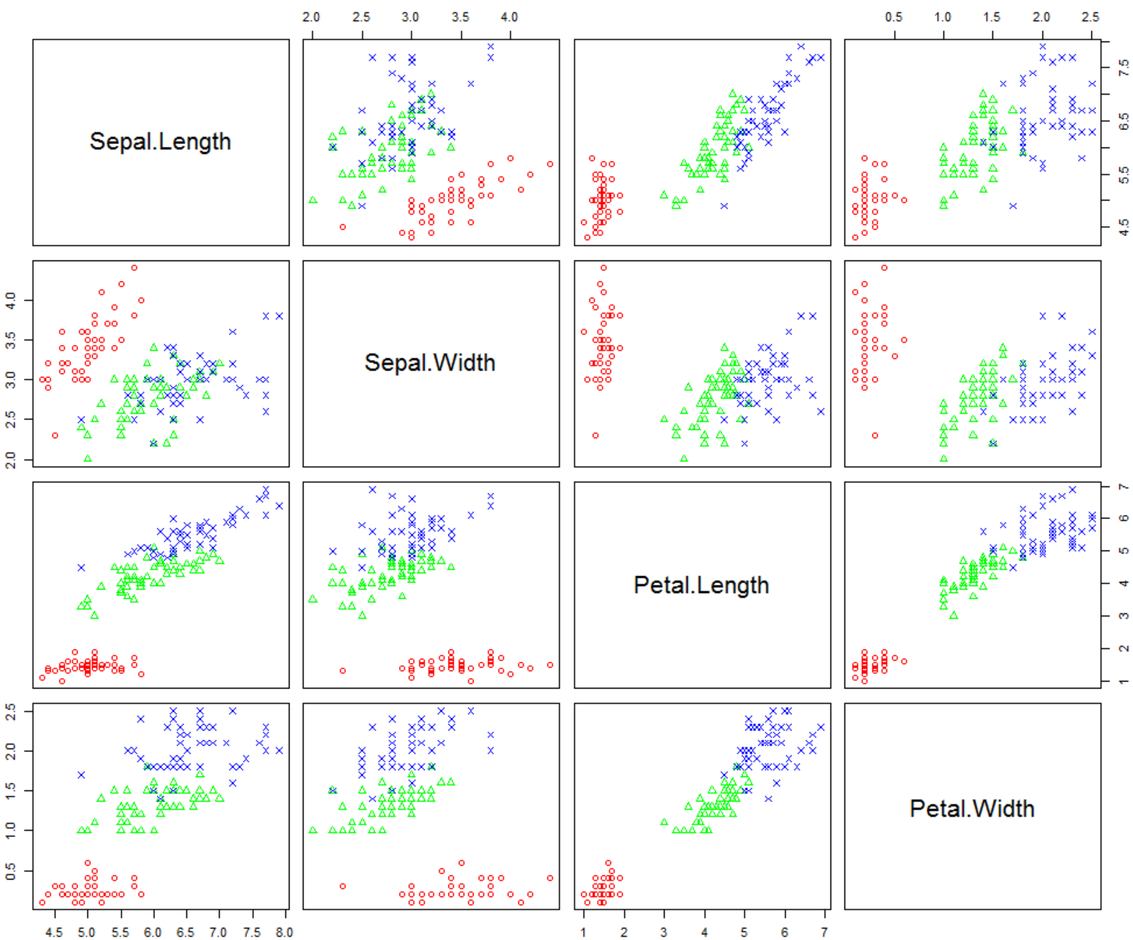
## scatterplot



# 多維繪圖 – 散點 直方 核密度

- iris
- pairs( iris[ , 1:4 ], pch = c(1, 2, 4)[iris$Species], col = c("red", "green", "blue")[iris$Species] )

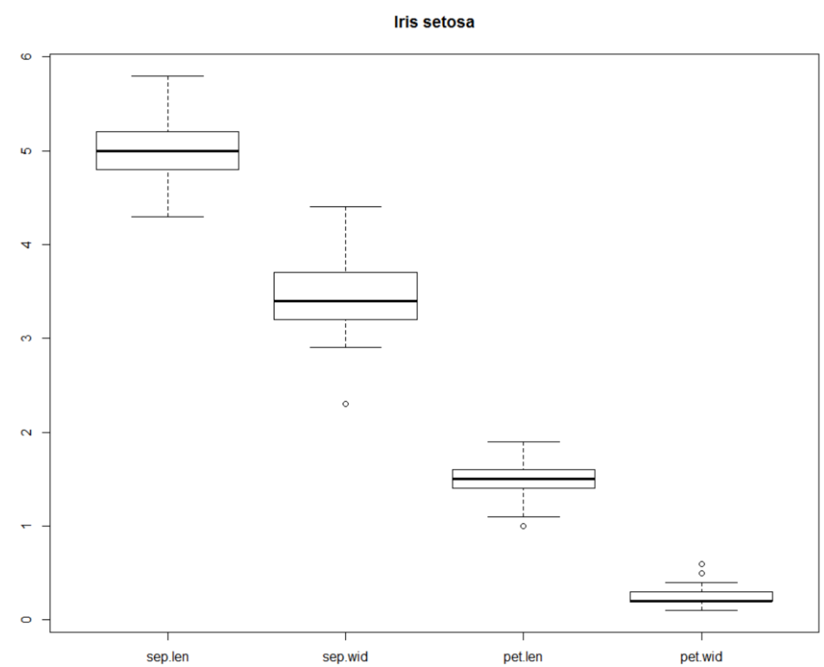
scatterplot
不同品種之散點圖



# 多維繪圖 – 多重分布

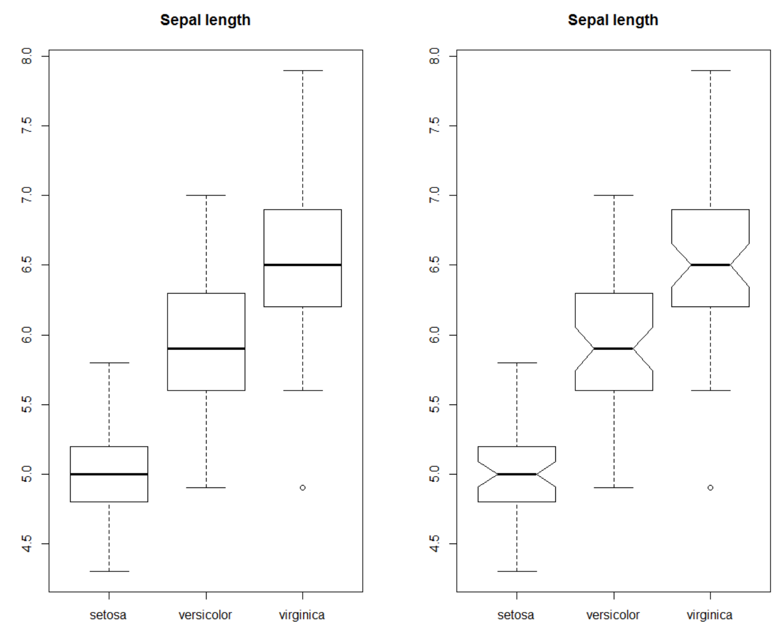
- 第一品種之中，
  花萼長度, 花萼寬度, 花瓣長度, 花瓣寬度，分布情形

- setosa <- iris[ iris$Species == "setosa", 1:4 ]
- boxplot( setosa, names = c( "sep.len", "sep.wid", "pet.len", "pet.wid" ), main = "Iris setosa" )



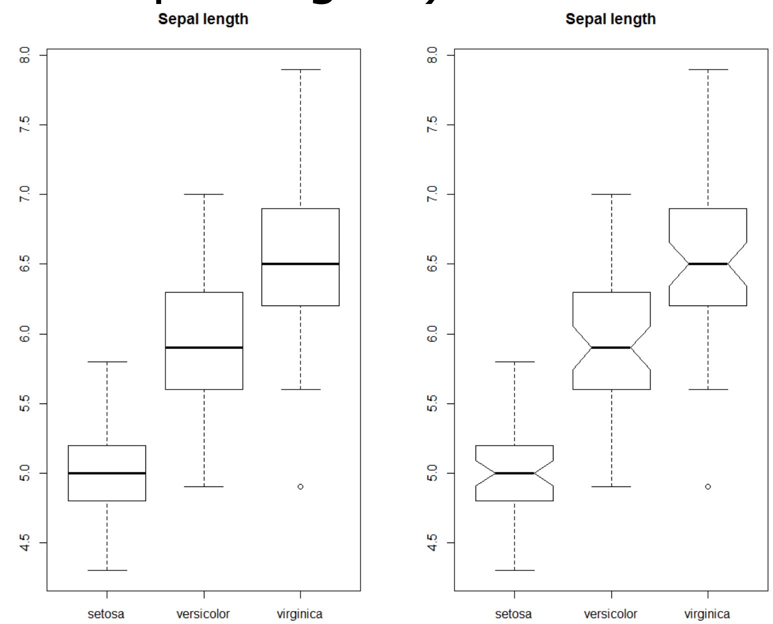
# 多維繪圖 – 多重分布

- 三個品種，
  花萼長度, 花萼寬度, 花瓣長度, 花瓣寬度，分布情形
- par( mfrow = c(1, 2) )
- with( iris, boxplot( Sepal.Length ~ Species, main = "Sepal length" ) )
- with( iris, boxplot( Sepal.Length ~ Species, notch = TRUE, main = "Sepal length" ) )



# 多維繪圖 – 多重分布

- 三個品種，
  花萼長度, 花萼寬度, 花瓣長度, 花瓣寬度，分布情形
- 依照不同種類，先分成三群
- par(mfrow = c(1, 2) )
- sx <- with( iris, split( Sepal.Length, Species ) )
- boxplot( sx, main = "Sepal length" )
- boxplot( sx, notch = TRUE, main = "Sepal length" )



# 多維繪圖 – 多重分布

不同品種之散點圖

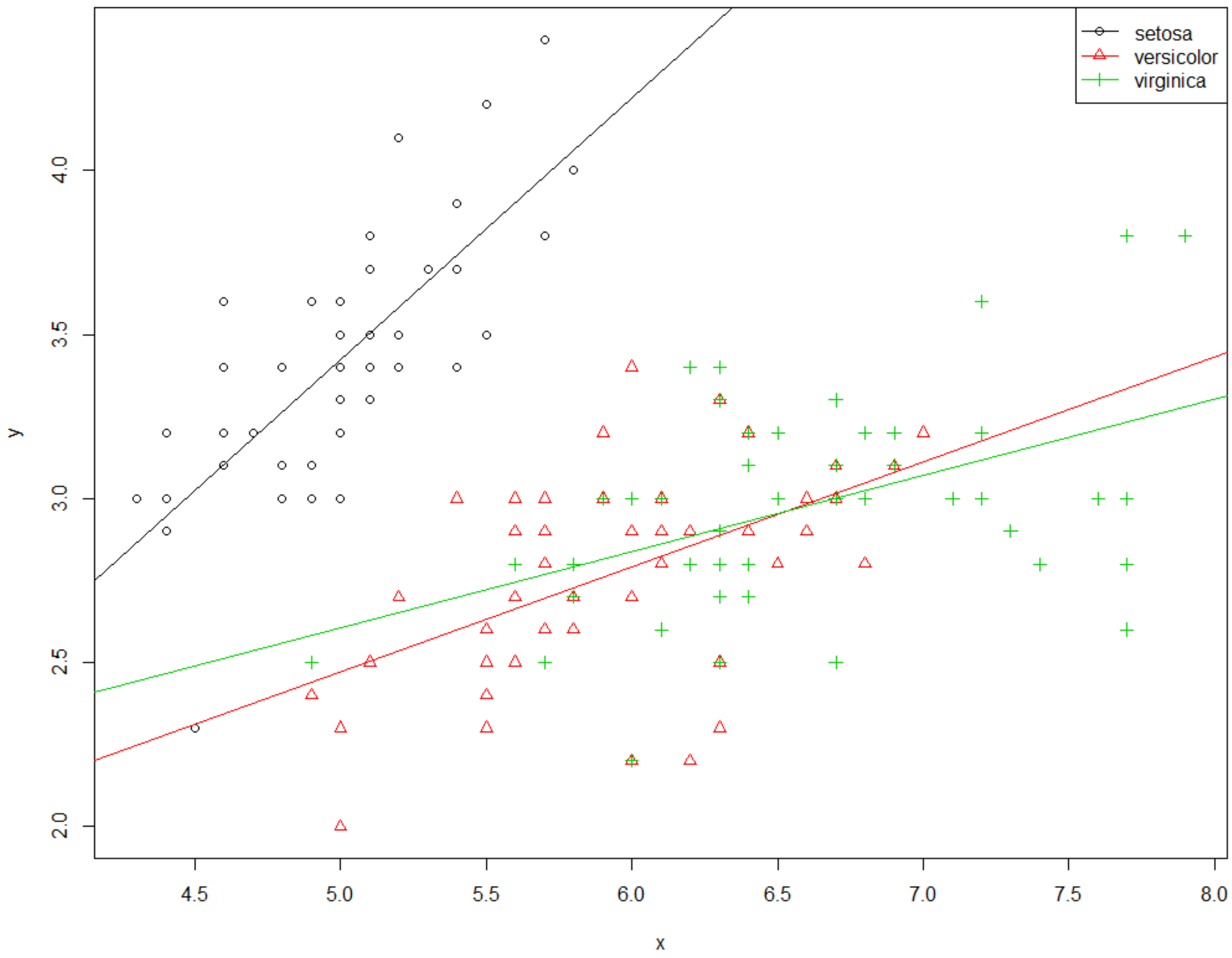
- 花萼長度 與 花萼寬度 之間的關係
- 依照不同種類，先分成三群
- sx <- with( iris, split( Sepal.Length, Species ) )
- sy <- with( iris, split( Sepal.Width, Species ) )

- par( mfrow = c(1, 1) )
- plot( 0, xlim = range(sx), ylim = range(sy), type = "n", xlab = "x", ylab = "y")
- points( sx[[1]], sy[[1]], pch = 1, col = 1)
- points( sx[[2]], sy[[2]], pch = 2, col = 2)
- points( sx[[3]], sy[[3]], pch = 3, col = 3)

- for (i in 1:3) abline( lm(sy[[i]] ~ sx[[i]]), col = i )
- legend( "topright", legend = c("setosa", "versicolor", "virginica"), lty = 1, pch = 1:3, col = 1:3 )

# 多維繪圖 – 多重分布

不同品種之散點圖

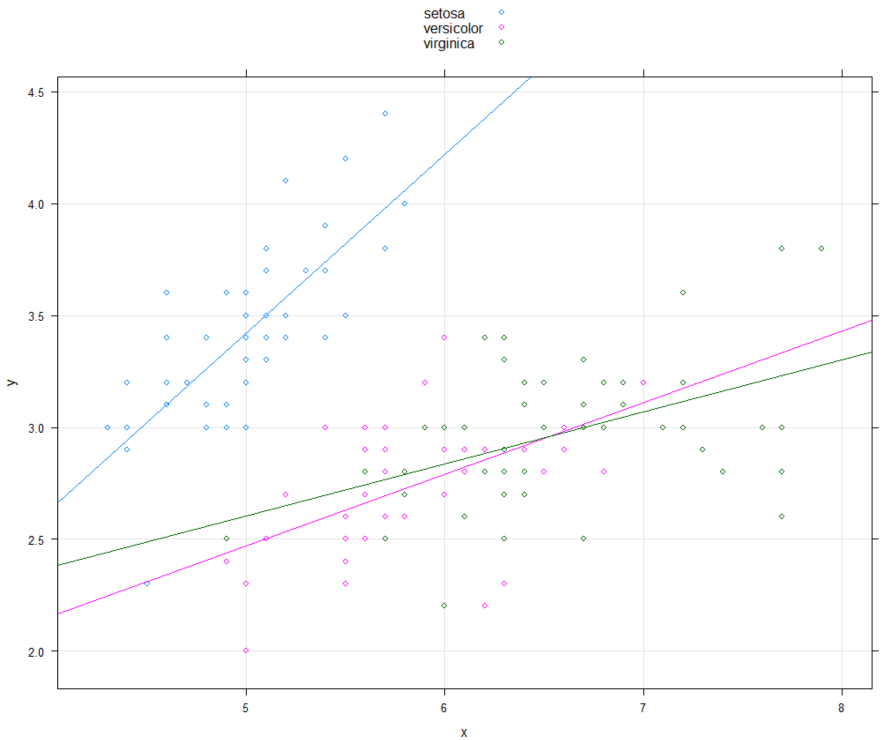


# 多維繪圖 – 多重分布

- 花萼長度 與 花萼寬度 之間的關係
- 依照不同種類，先分成三群

不同品種之散點圖

- x <- iris[[1]]
- y <- iris[[2]]
- species <- iris[[5]]
- library(lattice)



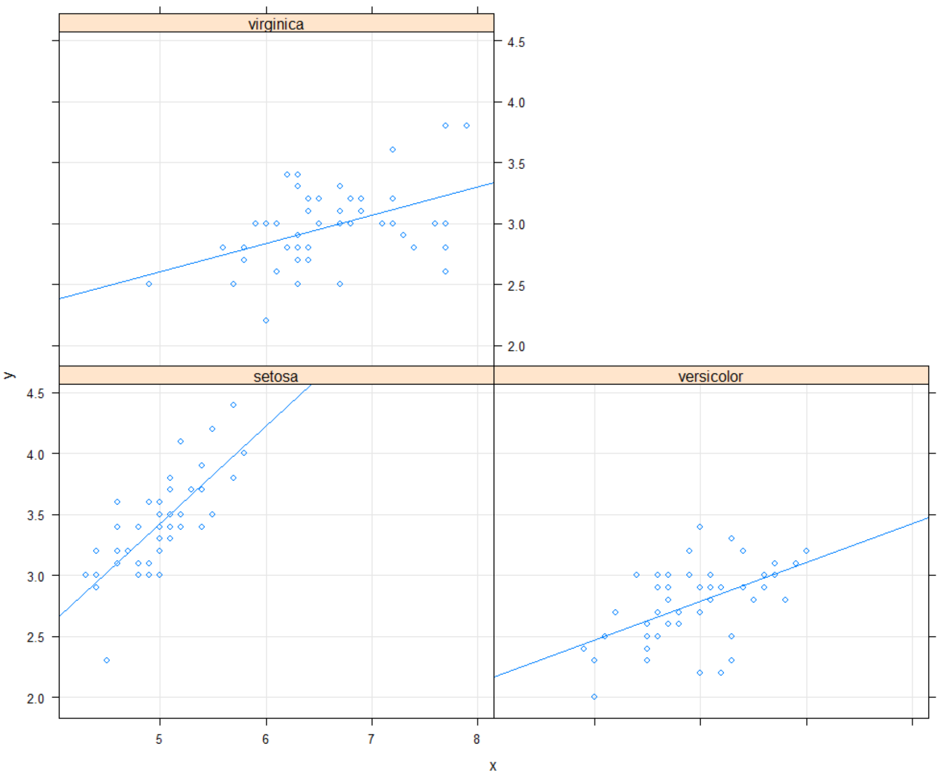

- xyplot( y ~ x, groups = species, type = c("g", "p", "r"), auto.key = TRUE)

# 多維繪圖 – 多重分布

- 花萼長度 與 花萼寬度 之間的關係
- 依照不同種類，先分成三群

不同品種分開之散點圖

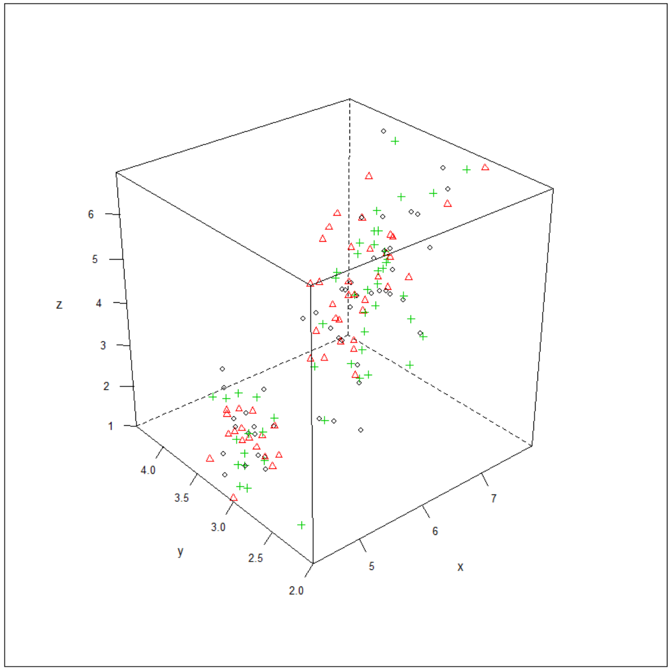
- x <- iris[[1]]
- y <- iris[[2]]
- species <- iris[[5]]
- library(lattice)




- xyplot( y ~ x | species, type = c("g", "p", "r"), auto.key = TRUE )

# 多維繪圖 – 三維散點圖

- ( 花萼長度, 花萼寬度, 花瓣長度 )




- data(iris)
- x <- iris[, 1]
- y <- iris[, 2]
- z <- iris[, 3]

- library(lattice)

- cloud( z ~ x * y, groups = iris$Spieces, pch = 1:3, col = 1:3,
- scales = list(arrows = FALSE),
- light.source = c(10, 0, 10) )