

Chapter 9

BASIC DATA PROCESSING (1)

資料處理小技巧

函數	說明
head()	前 6 筆資料或前 n 筆資料
tail()	最後 6 筆或最後 n 筆資料
rm(list=ls())	清除所有物件
rm()	清除指定物件
names()	檢示物件格式 / 更名

```
1 iris_data <- iris
2
3 head(iris_data,10)
4 tail(iris_data,3)
5 names(iris_data)
6 |
```

```
> head(iris_data,10)
  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
> tail(iris_data,3)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
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
> names(iris_data)
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
> |
```

解構 data.frame

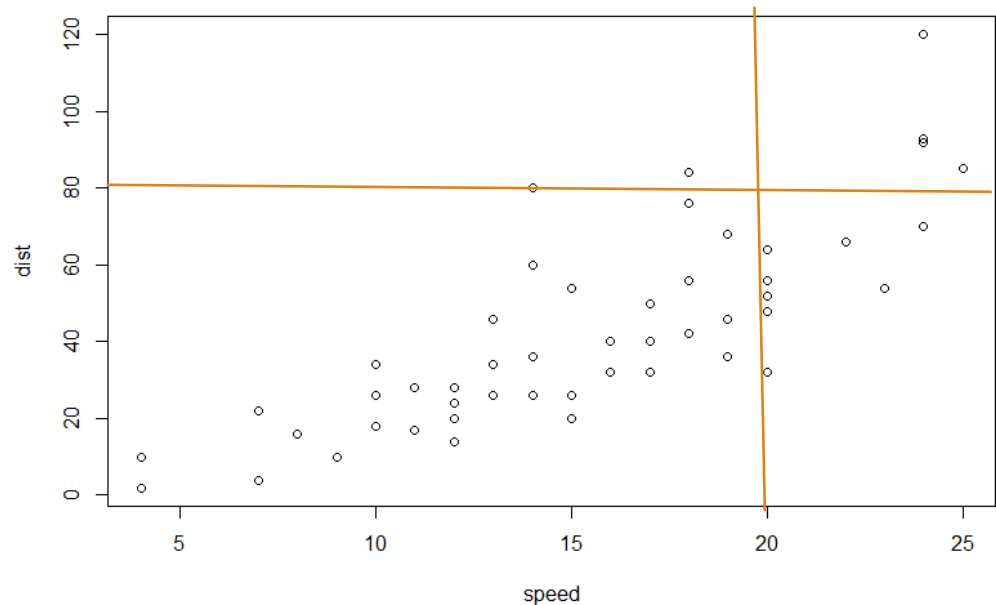
```
1 iris_data <- iris
2
3 head(iris_data,1)
4
5 print(iris_data[1,1])
6 print(iris_data[1,2])
7 print(iris_data[1,3])
8 print(iris_data[1,4])
9 print(iris_data[1,5])
10
11 print(iris_data[3:5, c("Sepal.Length", "Species")])
12 print(iris_data[120:123, c("Sepal.Length", "Species")])
13
```

```
> iris_data <- iris
>
> head(iris_data,1)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1          3.5          1.4          0.2   setosa
>
> print(iris_data[1,1])
[1] 5.1
> print(iris_data[1,2])
[1] 3.5
> print(iris_data[1,3])
[1] 1.4
> print(iris_data[1,4])
[1] 0.2
> print(iris_data[1,5])
[1] setosa
Levels: setosa versicolor virginica
>
> print(iris_data[3:5, c("Sepal.Length", "Species")])
  Sepal.Length Species
3          4.7   setosa
4          4.6   setosa
5          5.0   setosa
> print(iris_data[120:123, c("Sepal.Length", "Species")])
  Sepal.Length Species
120          6.0 virginica
121          6.9 virginica
122          5.6 virginica
123          7.7 virginica
```

解構 data.frame

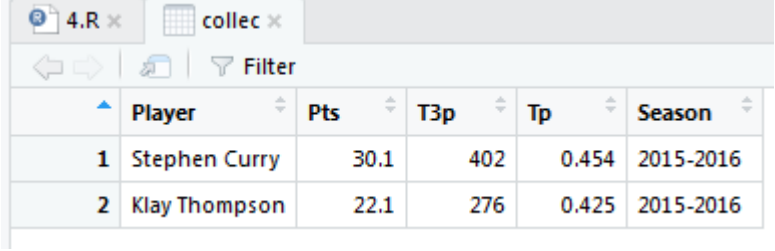
```
1 car_data <- cars
2 plot(car_data)
3
4 a <- car_data[car_data$speed>20, c("speed","dist")]
5 print(a)
6
7 b <- car_data[car_data$speed>20 &
8               car_data$dist>80,c("speed","dist")]
9 print(b)
```

```
> car_data <- cars
> plot(car_data)
>
> a <- car_data[car_data$speed>20, c("speed","dist")]
> print(a)
  speed dist
44    22   66
45    23   54
46    24   70
47    24   92
48    24   93
49    24  120
50    25   85
>
> b <- car_data[car_data$speed>20 &
+               car_data$dist>80,c("speed","dist")]
> print(b)
  speed dist
47    24   92
48    24   93
49    24  120
50    25   85
```



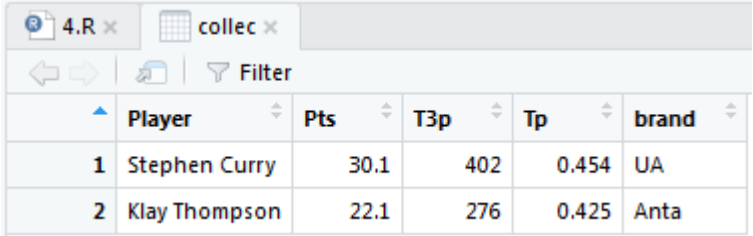
更新 data.frame 欄位 與刪除欄位

```
1 Player <- c("Stephen Curry", "Klay Thompson")
2 Pts <- c(30.1, 22.1)
3 T3p <- c(402L, 276)
4 Tp <- c(0.454, 0.425)
5 Season <- c("2015-2016", "2015-2016")
6
7 collec <- data.frame(Player, Pts, T3p, Tp, Season,
8                       stringsAsFactors = FALSE)
9
10 View(collec)|
```



	Player	Pts	T3p	Tp	Season
1	Stephen Curry	30.1	402	0.454	2015-2016
2	Klay Thompson	22.1	276	0.425	2015-2016

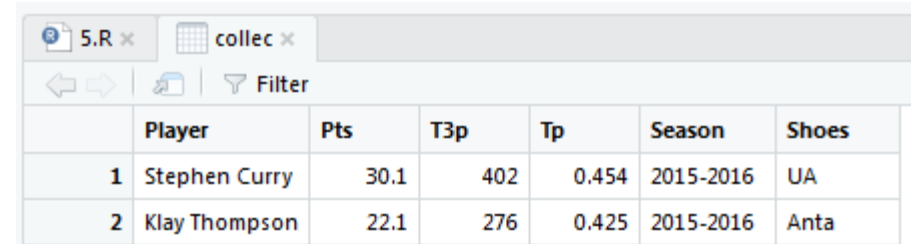
```
1 Player <- c("Stephen Curry", "Klay Thompson")
2 Pts <- c(30.1, 22.1)
3 T3p <- c(402L, 276)
4 Tp <- c(0.454, 0.425)
5 Season <- c("2015-2016", "2015-2016")
6
7 collec <- data.frame(Player, Pts, T3p, Tp, Season,
8                       stringsAsFactors = FALSE)
9 #Add new column
10 Shoes <- c("UA", "Anta")
11 #delete column
12 collec$Season <- NULL
13
14 collec$brand <- Shoes|
15 View(collec)
```



	Player	Pts	T3p	Tp	brand
1	Stephen Curry	30.1	402	0.454	UA
2	Klay Thompson	22.1	276	0.425	Anta

更新 data.frame 欄位 與刪除資料

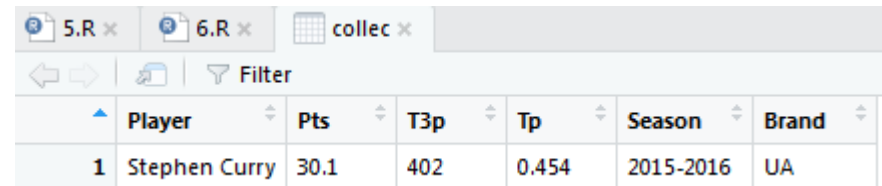
```
1 Player <- c("Stephen Curry", "Klay Thompson")
2 Pts <- c(30.1, 22.1)
3 T3p <- c(402L, 276)
4 Tp <- c(0.454, 0.425)
5 Season <- c("2015-2016", "2015-2016")
6 Shoes <- c("UA", "Anta")
7
8 collec <- data.frame(Player, Pts, T3p, Tp, Season, Shoes,
9                       stringsAsFactors = FALSE)
10 View(collec)
```



RStudio interface showing the 'collec' data frame with 2 rows and 7 columns. The columns are Player, Pts, T3p, Tp, Season, and Shoes. The data is as follows:

	Player	Pts	T3p	Tp	Season	Shoes
1	Stephen Curry	30.1	402	0.454	2015-2016	UA
2	Klay Thompson	22.1	276	0.425	2015-2016	Anta

```
1 Player <- c("Stephen Curry", "Klay Thompson")
2 Pts <- c(30.1, 22.1)
3 T3p <- c(402L, 276L)
4 Tp <- c(0.454, 0.425)
5 Season <- c("2015-2016", "2015-2016")
6 Shoes <- c("UA", "Anta")
7
8 collec <- data.frame(Player, Pts, T3p, Tp, Season, Shoes,
9                       stringsAsFactors = FALSE)
10
11 collec <- collec[-2,]
12 names(collec)[6] <- "Brand"
13 View(collec)
```

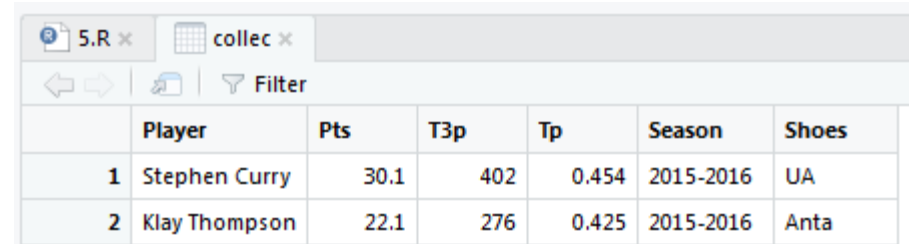


RStudio interface showing the 'collec' data frame after row deletion and column renaming. The columns are Player, Pts, T3p, Tp, Season, and Brand. The data is as follows:

	Player	Pts	T3p	Tp	Season	Brand
1	Stephen Curry	30.1	402	0.454	2015-2016	UA

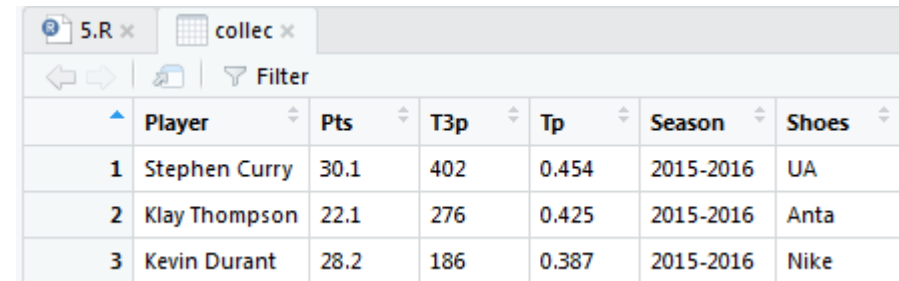
新增 data.frame 資料

```
1 Player <- c("Stephen Curry", "Klay Thompson")
2 Pts <- c(30.1, 22.1)
3 T3p <- c(402L, 276)
4 Tp <- c(0.454, 0.425)
5 Season <- c("2015-2016", "2015-2016")
6 Shoes <- c("UA", "Anta")
7
8 collec <- data.frame(Player, Pts, T3p, Tp, Season, Shoes,
9                       stringsAsFactors = FALSE)
10 View(collec)
```



	Player	Pts	T3p	Tp	Season	Shoes
1	Stephen Curry	30.1	402	0.454	2015-2016	UA
2	Klay Thompson	22.1	276	0.425	2015-2016	Anta

```
1 Player <- c("Stephen Curry", "Klay Thompson")
2 Pts <- c(30.1, 22.1)
3 T3p <- c(402L, 276L)
4 Tp <- c(0.454, 0.425)
5 Season <- c("2015-2016", "2015-2016")
6 Shoes <- c("UA", "Anta")
7
8 collec <- data.frame(Player, Pts, T3p, Tp, Season, Shoes,
9                       stringsAsFactors = FALSE)
10
11 KD <- c("Kevin Durant", 28.2, 186, 0.387,
12         "2015-2016", "Nike")
13 collec <- rbind(collec, KD)
14
15 View(collec)
```



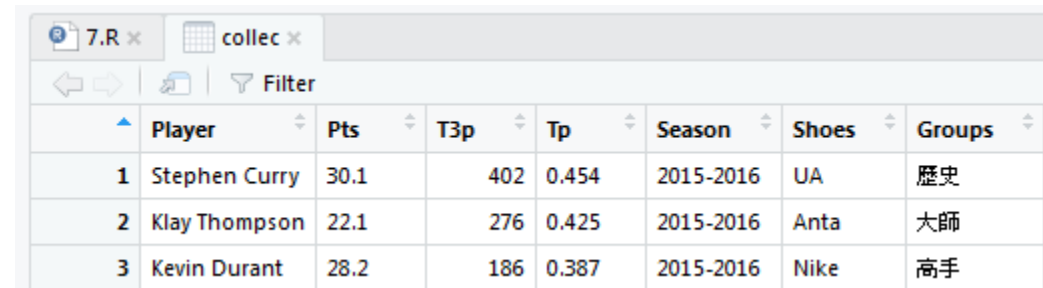
	Player	Pts	T3p	Tp	Season	Shoes
1	Stephen Curry	30.1	402	0.454	2015-2016	UA
2	Klay Thompson	22.1	276	0.425	2015-2016	Anta
3	Kevin Durant	28.2	186	0.387	2015-2016	Nike

新增 data.frame 類別變數

```
1 Player <- c("Stephen Curry", "Klay Thompson")
2 Pts <- c(30.1, 22.1)
3 T3p <- c(402, 276)
4 Tp <- c(0.454, 0.425)
5 Season <- c("2015-2016", "2015-2016")
6 Shoes <- c("UA", "Anta")
7
8 collec <- data.frame(Player, Pts, T3p, Tp, Season, Shoes,
9                       stringsAsFactors = FALSE)
10
11 KD <- c("Kevin Durant", 28.2, 186, 0.387,
12         "2015-2016", "Nike")
13 collec <- rbind(collec, KD)
14
15 collec$T3p <- as.numeric(collec$T3p)
16 class(collec$T3p)
17 collec$Groups <- cut(collec$T3p, breaks = c(0, 250, 350, Inf),
18                     labels = c("高手", "大師", "歷史"))
19 View(collec)
```

Cut 函數

使用 (breaks) 4 個區間
分 3 個類別 (labels)



	Player	Pts	T3p	Tp	Season	Shoes	Groups
1	Stephen Curry	30.1	402	0.454	2015-2016	UA	歷史
2	Klay Thompson	22.1	276	0.425	2015-2016	Anta	大師
3	Kevin Durant	28.2	186	0.387	2015-2016	Nike	高手

合并 data.frame

```
1 iris_1 <- iris[1:3,]
2 iris_2 <- iris[30:32,]
3 |
4 iris_3 <- rbind(iris_1,iris_2)
5 print(iris_1)
6 print(iris_2)
7 print(iris_3)
```

```
1 iris_1 <- iris[50:51,1]
2 iris_2 <- iris[50:51,5]
3 |
4 iris_3 <- cbind(iris_1,iris_2)
5 print(iris_1)
6 print(iris_2)
7 print(iris_3)|
```

```
Console Terminal x
~/
> iris_1 <- iris[1:3,]
> iris_2 <- iris[30:32,]
>
> iris_3 <- rbind(iris_1,iris_2)
> print(iris_1)
  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
> print(iris_2)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
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
> print(iris_3)
  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
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
```

```
> print(iris[50,])
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
50           5          3.3          1.4          0.2  setosa
> print(iris[51,])
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
51           7          3.2          4.7          1.4 versicolor
> print(iris_3)
      iris_1 iris_2
[1,]      5      1
[2,]      7      2
> |
```

merge()

```
1 name <- c("Curry","James","Jordan")
2 n2 <- c(30,23,23)
3 n <- data.frame(name,n2)
4
5 name <- c("Curry","Jordan")
6 b2 <- c("UA","Air Jordan")
7 b <- data.frame(name,b2)
8
```

```
9 m <- merge(n,b)
10 print(m)
```

```
> m <- merge(n,b)
> print(m)
      name n2      b2
1 Curry  30      UA
2 Jordan 23 Air Jordan
```

```
12 m1 <- merge(n,b, all.x=TRUE)
13 print(m1)
```

```
> m1 <- merge(n,b, all.x=TRUE)
> print(m1)
      name n2      b2
1 Curry  30      UA
2 James  23      <NA>
3 Jordan 23 Air Jordan
```

```
14 m2 <- merge(n,b, all.y=TRUE)
15 print(m2)
```

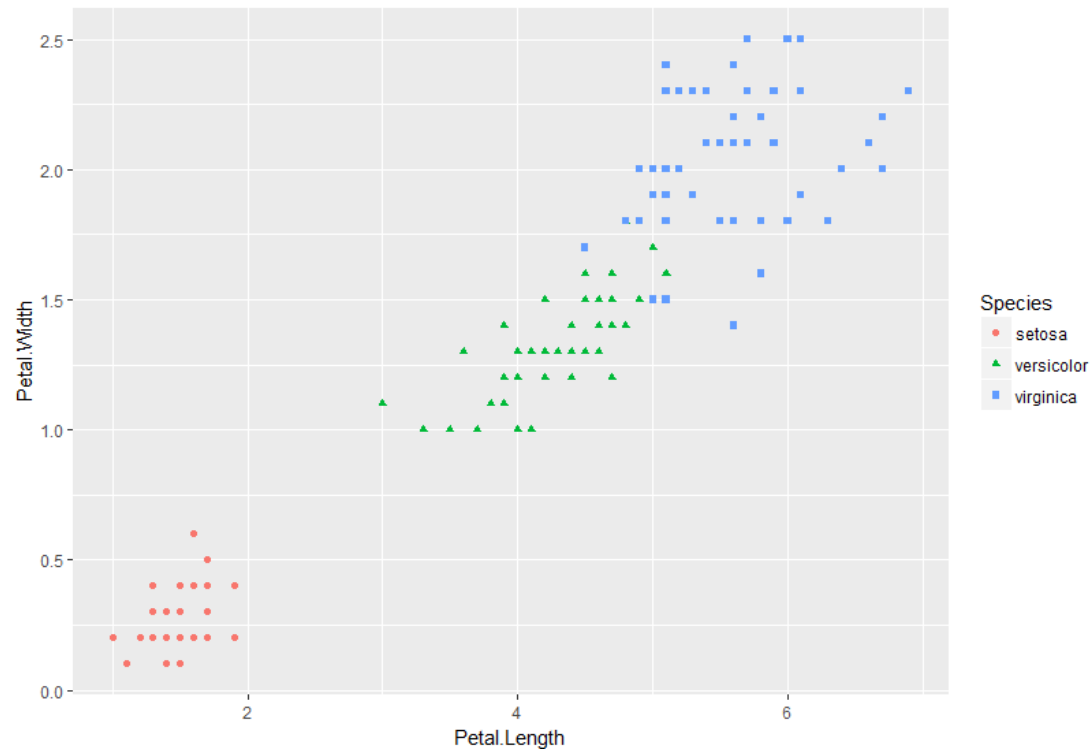
```
> m2 <- merge(n,b, all.y=TRUE)
> print(m2)
      name n2      b2
1 Curry  30      UA
2 Jordan 23 Air Jordan
```

```
16 m3 <- merge(n,b, all.x=TRUE,all.y=TRUE)
17 print(m3)
```

```
> m3 <- merge(n,b, all.x=TRUE,all.y=TRUE)
> print(m3)
      name n2      b2
1 Curry  30      UA
2 James  23      <NA>
3 Jordan 23 Air Jordan
```

隨堂練習 1

1. 取出 IRIS 當中的 Petal.Width data
2. 以 ggplot2 畫出不同分類的散佈圖



Any Questions !?