

## 常見視覺化圖表與程式實作

### 統計表格介紹

統計表格是由表格與資料所組成，研究者將統計調查所取得的原始資料，經過整理、分類，排放在表格中，以方便讀者做比較。表格的主體為資料，可以幫助研究者有效地利用有限的篇幅呈現大量的資料，通常用來呈現精確的數據，常見的統計表格有次數分配表與列聯表。

#### 1. 次數分配表

##### ➤ 次數分配

所謂次數分配(frequency distribution)是指將資料依數量大小或類別種類而分成若干組，並列出各組所含次數（即各組所含觀測值的個數）。

##### ➤ 相對次數分配

相對次數分配(relative frequency distribution)是指各組次數佔總次數的比例分配。

$$\text{相對次數} = \text{組次數} / \text{總次數}$$

##### ➤ 累計次數分配

累積次數(cumulative frequency distribution)分配有兩個重要的用途，一是指出落在某一特定值之上或之下的累積次數有若干，二是方便計算某些特定的敘述統計量，如中位數、四分位數等。

#### 次數分配 R 程式實作：

```
data("iris")
#利用 cut()將連續變數 Sepal.Length 分組
fre <- cut(iris$Sepal.Length,
           breaks=c(-Inf, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, Inf),
           labels=c("小於 4", "4~4.5", "4.5~5", "5~5.5", "5.5~6",
                    "6~6.5", "6.5~7", "7~7.5", "大於 7.5"))
n <- length(iris)
f <- table(fre)#table()計算該變數的次數分配
rf <- round(f/sum(f),4)#計算相對次數，並四捨五入到小數第 4 位
cf <- cumsum(f)#計算累計次數
crf <- round(cf/sum(f),4)#計算累計相對次數，並四捨五入到小數第 4 位
freTable <- t(rbind(f, rf, cf, crf))#rbind()合併數值；t()行列互換
colnames(freTable) <- c("次數分配", "相對次數",
```

" 累計次數", " 累計相對次數")

View(freTable)

	次數分配	相對次數	累計次數	累計相對次數
小於4	0	0.0000	0	0.0000
4~4.5	5	0.0333	5	0.0333
4.5~5	27	0.1800	32	0.2133
5~5.5	27	0.1800	59	0.3933
5.5~6	30	0.2000	89	0.5933
6~6.5	31	0.2067	120	0.8000
6.5~7	18	0.1200	138	0.9200
7~7.5	6	0.0400	144	0.9600
大於7.5	6	0.0400	150	1.0000

## 2. 列聯表/交叉表

列聯表(contingency table)又稱為交叉表，將兩變量交叉分類，計算各類別的發生次數，並比較各類別的分佈狀況，以尋找兩變量間的關係。

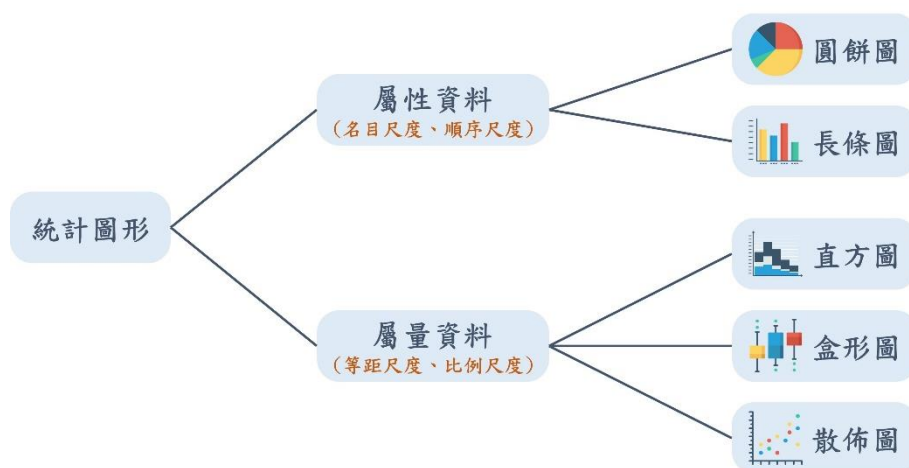
列聯表 R 程式實作：

```
library(car)#使用 car 中的 Salaries 資料集
data(Salaries)
conTable <- table(Salaries$rank, Salaries$discipline)
conTable
```

	A	B
AsstProf	24	43
AssocProf	26	38
Prof	131	135

## 統計圖形介紹

依據變數的類型可將統計圖形做適當的使用時機分類，當資料為屬性資料時（名目尺度、順序尺度），資料適合以圓餅圖(pie chart)與長條圖(bar chart)的方式呈現；當資料為屬量資料時（等距尺度、比例尺度），資料則適合以直方圖(histogram)、盒形圖(box plot)與散佈圖(scatter plot)的方式呈現。



### 1. 屬性資料

屬性資料的區別適用於資料之變數特徵屬於非數值型態時，需利用一組的標記、類別、性質或名稱以區別每個基本單位的特徵、屬性。無法以數值表示的統計資料，如航班編號、性別、學歷、旅遊同伴、頭髮顏色、宗教等。名目尺度和順序尺度通常會被歸類為屬性資料，而圓餅圖與長條圖通常也被視為屬性資料的最佳呈現方式。

#### ➤ 圓餅圖(pie chart)

圓餅圖，又稱為餅狀圖，是一個劃分為幾個扇形的圓形統計圖表，用於描述量、頻率或百分比之間的相對關係。在圓餅圖中，每個扇區的弧長（以及圓心角和面積）大小為其所表示的數量比例。這些扇區合在一起剛好是一個完全的圓形。

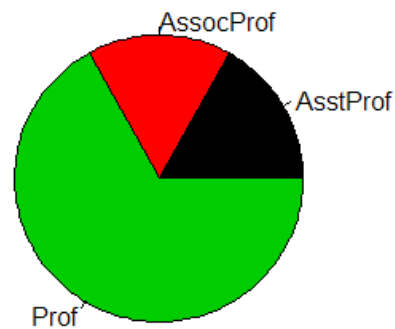
**使用時機：**通常展現整體數據中各項目所占的比重時，可用不同顏色來標示，且大多用百分比（%）顯示。例：家庭收／支占比，或者公司商品銷售占比等。

**注意事項：**資料項目不宜過多，以免過於破碎的效果；其中的每份扇形可用各類線條或空白加以區隔，讓圖的占比更清楚。

#### 圓餅圖 R 程式實作：

```
data(Salaries)
pie(table(Salaries$rank),
     main="教授等級圓餅圖",
     col=1:3)
```

教授等級圓餅圖



#### ➤ 長條圖(bar chart)

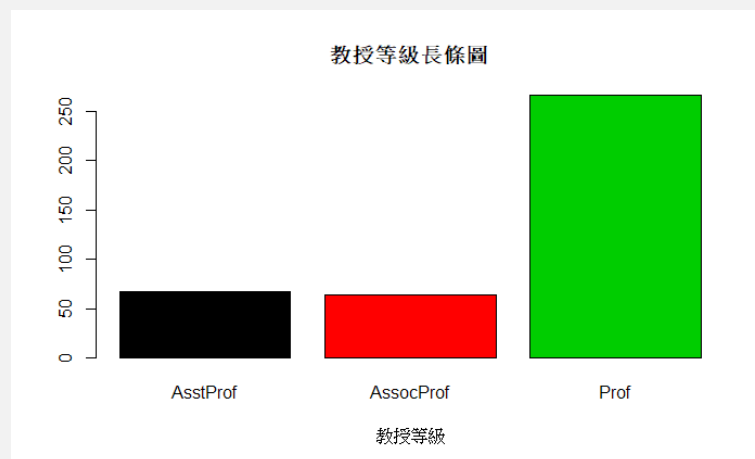
長條圖，又稱為柱狀圖，是一種以長方形的長度為變量的統計圖表。通常用來比較兩個或以上的維度資料（不同時間或者不同條件），只有一個變量，一般而言是較小的數據集分析使用的統計圖形。其特色為長方形的寬度相同，且長度與已知次數成比例，主要用於處理離散型資料。長條圖亦可橫向排列，或用多維方式表達。

**使用時機：**適用於內容較為獨立的離散型資料，用來表示有關數量的多少，特別適合於對各數量進行對比。例如：對文湖線和板南線進行客流量比較。

**注意事項：**柱與柱間是有間隔距離的，距離通常為柱寬的二分之一，若同一個項目要呈現的不只一筆資料，習慣上會用不同顏色或不同花色來區分。

#### 長條圖 R 程式實作：

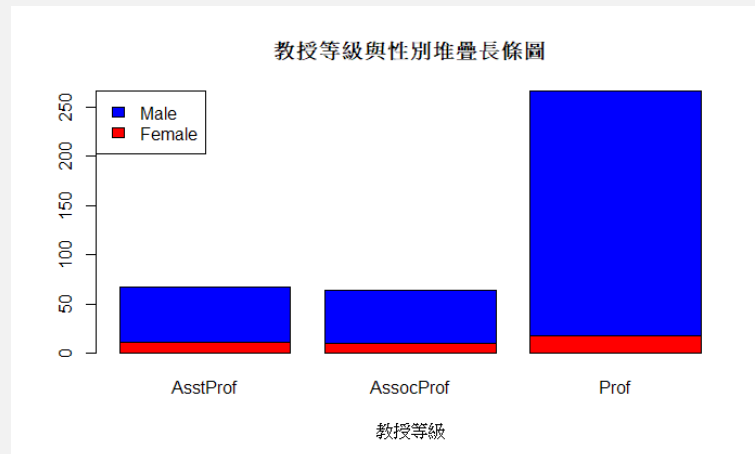
```
data(Salaries)
#單變量長條圖（屬量變數）
barplot(table(Salaries$rank), col=1:3,
        main="教授等級長條圖", xlab="教授等級")
```



### 長條圖 R 程式實作：

#雙變量堆疊長條圖（屬量變數 vs 屬量變數）

```
barplot(table(Salaries$sex, Salaries$rank), col=c(2, 4),  
         main="教授等級長條圖_按等級", xlab="教授等級",  
         legend=TRUE, args.legend=list(x="topleft"))
```



## 2. 屬量資料

描述數量和多寡的資料，資料以數值型態呈現。屬量資料可以根據資料型態再分為離散型資料與連續型資料。屬量資料適用於資料之變數特徵屬於「數值」型態，如銷售量、銷售金額、收入、支出、身高、體重、年齡、小孩數量等。通常等距尺度和比例尺度會被歸類為屬量資料。在典型的量化資料中依據數值分布是否為連續性數值型態，又可以區分為離散型資料和連續型資料兩種類型。

### ➤ 直方圖(histogram)

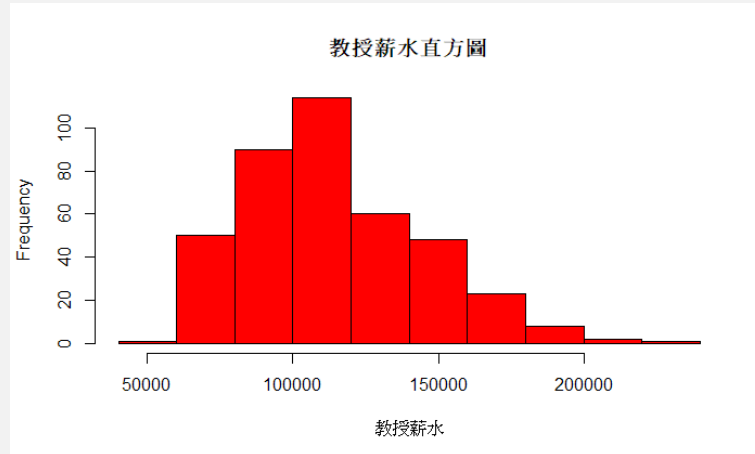
直方圖由一系列高度不等的縱向條紋或線段組成。一般用橫軸表示數據類型，縱軸表示分佈情況。是一種用來展示分組數據的分佈情況，圖中每個長方形寬度相對於組區間，面積則相對於它所代表的次數，是一種連續變量的概率分佈估計。直方圖的建構第一步是將值的範圍分段，即將整個值的範圍分成一系列間隔，然後計算每個間隔中有多少值。這些值通常被指定為連續的，不重疊的變量間隔。間隔必須相鄰，並且通常是（但非必須）相等的大小。

**使用時機：**當資料為連續的數據時，若目的為比較各類量的大小，則以直方圖表示。

**注意事項：**組距不可太少或太多，並選擇使用等寬的組距，不等寬的組距將造成視覺差異，因此儘可能採用等寬的組距。

### 直方圖 R 程式實作：

```
data(Salaries)
hist(Salaries$salary, col=2,
     main="教授薪水直方圖", xlab="教授薪水")
```



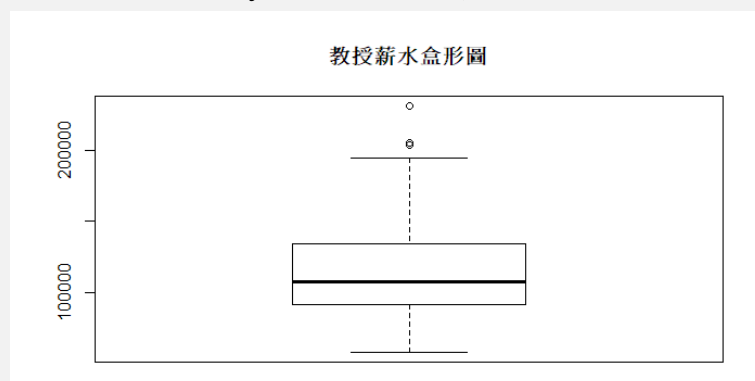
### ➤ 盒形圖(box plot)

盒形圖，又稱為箱形圖，是一種能顯示數據分散情況的統計圖形。因形狀如箱子而得名。此圖中之盒子之外，也常會有線條在上下四分位數之外延伸出去。離群值在盒形圖中會以個別點表示。盒形圖屬於無母數統計的範圍，他顯示資料的特性，但對於母體的分布並無任何假設。在各種領域也經常被使用，常見於品質管理。不過作法相對較繁瑣。它能顯示出一組數據的最大值、最小值、四分位數及離群值。

**使用時機：**當資料存在偏斜性與離群值時，盒形圖較其他統計圖表更能準確描繪出資料的離散情況。

### 盒形圖 R 程式實作：

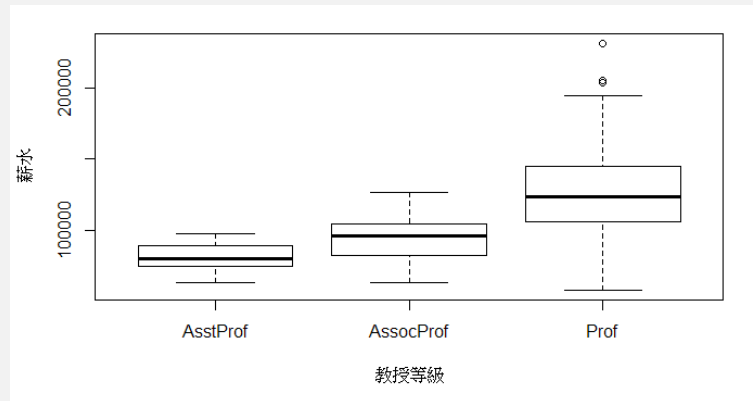
```
data(Salaries)
#單變量盒形圖（屬量變數）
boxplot(Salaries$salary, main="教授薪水盒形圖")
```



### 盒形圖 R 程式實作：

#雙變量盒形圖（屬量變數 vs 屬性變數）

```
boxplot(Salaries$salary~Salaries$rank,  
        xlab="教授等級",ylab="薪水")
```



### ➤ 散佈圖(scatter plot)

散佈圖，又稱為相關圖，圖形呈現方式是收集兩個變數成對的資料，將資料以點畫在坐標圖形上，通過對其觀察分析，加以判斷兩個變數之間的相關關係。在我們的生活及工作中，許多現象和原因，有些呈規則的關聯，有些則呈不規則。我們要瞭解它，就可藉助散佈圖統計手法來判斷它們之間的相關關係。

**使用時機：**能大概掌握原因與結果之間是否有相關及相關的程度如何，例如患者滿意度與候診時間的關係。

### 散佈圖 R 程式實作：

```
data(Salaries)  
plot(Salaries$yrs. service, Salaries$salary, col=Salaries$rank,  
      xlab="服務年數",ylab="薪水")  
legend("topright", legend=levels(Salaries$rank), fill=1:3)
```

