

## R-web 資料分析應用：圖表繪製(二)

江 奕 副統計分析師

生統 eNews 出刊囉！上期向大家介紹了一些基本的圖表繪製功能，其中，包括次數分配表、列聯表、莖葉圖、2D 散佈圖及 3D 散佈圖等功能，本期，將緊接著上一期的生統 eNews 繼續利用【雲端資料分析暨導引系統】(R-web, <http://www.r-web.com.tw>) 介紹其他的「圖表繪製」方法！

本期將依序介紹 R-web 圖表繪製模組內的二維曲線圖、三維曲線圖、直方圖、長條圖、圓餅圖及盒鬚圖，本系列圖形將統一使用源自基隆社區為基礎的整合篩檢計畫 (Keelung Community-based Integrated Screen Program, KCIS) 的心血管疾病資料作為範例資料檔，此資料的變數定義可參考下表。詳細資料介紹請參閱[首期生統 eNews](#)。

變數	定義
性別(Gender)	女性(0)、男性(1)
年齡(Age)	
腰圍(Waist)	公分(cm)
心臟收縮壓(SysBP)	毫米汞柱(mmHg)
心臟舒張壓(DiaBP)	毫米汞柱(mmHg)
空腹葡萄糖(AC)	毫克/分升(mg/dl)
高密度脂蛋白(HDL)	毫克/分升(mg/dl)
三酸甘油酯(TG)	毫克/分升(mg/dl)
嚼檳榔習慣(Betelnut)	無(0)、有(1)
飲酒習慣(Alc_Drink)	無(0)、有(1)
個人心血管疾病史(CVD)	無(0)、有(1)
家族心血管疾病史(FamilyHx)	無(0)、有(1)
抽菸習慣(Tobacco)	無(0)、有(1)
菸草消費量(Tobacco_Consumption)	無(0)、每日一包(1)、每日兩包(2)、每日三包以上(3)

## ➤ 二維曲線圖

二維曲線圖類似於 2D 散佈圖，亦可將兩變數分別置於座標圖上的 X 軸與 Y 軸，初步觀察兩變數間之相關性。不同在於，二維曲線圖提供二維函數曲線的描繪，若有樣本資料點，可同時選擇附加資料點以觀察樣本資料點和函數曲線的關係。依序點選主選單中【圖表繪製】→【曲線(面)圖】→【二維曲線圖】進行繪製。

### ● 使用資料檔

**步驟一：資料匯入**

使用者個人資料檔
檢視資料型態(開新視窗)

● 選擇要進行分析的資料檔或上傳檔案

- CVD
- CVD\_100
- CVD\_BP

您所選擇的資料檔為： CVD\_100

● 不使用資料檔

**步驟二：參數設定**

請由下方按鈕建立函數：

y = f(x) =

運算子：

數學函數：

三角函數：

截斷類型函數：

游標控制：

小算盤：

7	8	9	/
4	5	6	*
1	2	3	-
0	00	.	+

---

已儲存的函數：清除選擇的函數

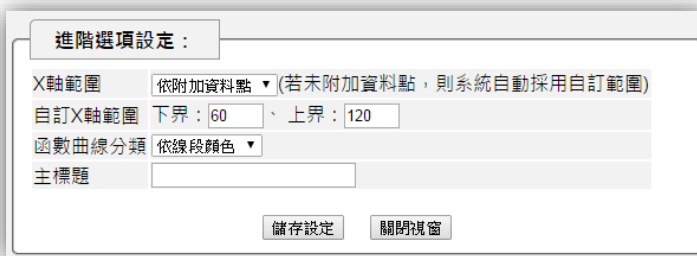
f(x) = 0.5054 \* x + 43.039

附加資料點

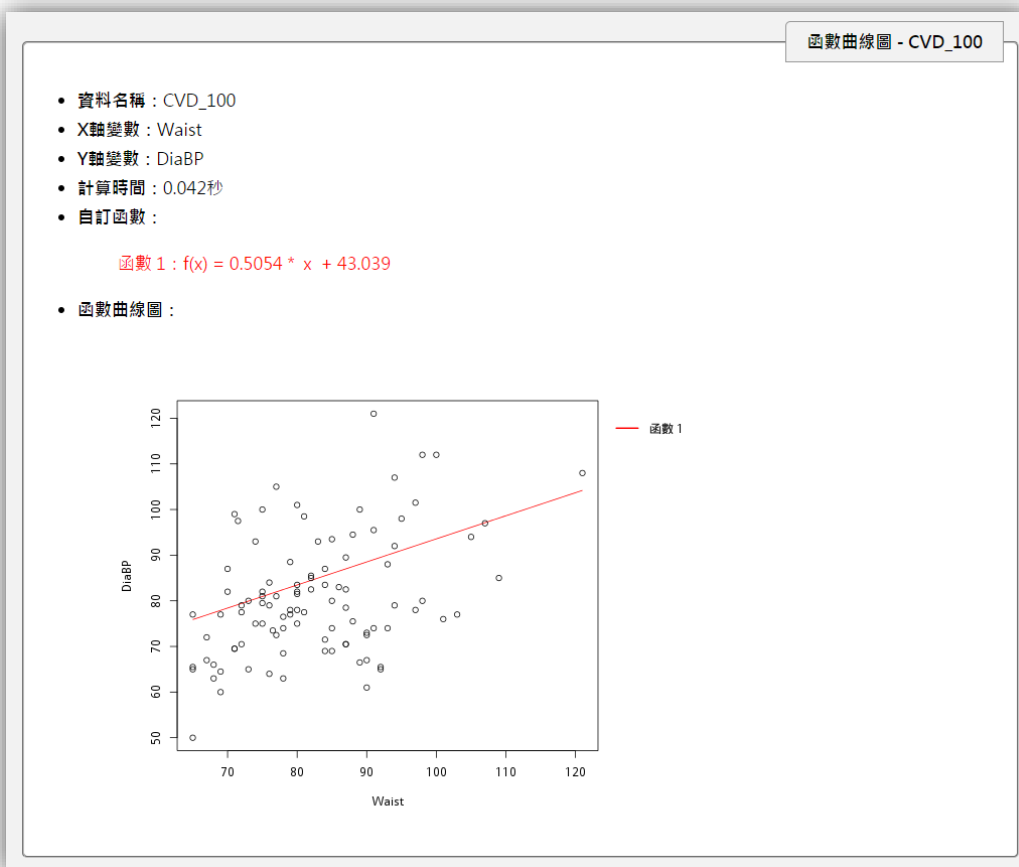
<div style="border: 1px solid #ccc; padding: 2px;"> <ul style="list-style-type: none"> <li>ID</li> <li>CVD</li> <li>Age</li> <li>Gender</li> <li>SysBP</li> <li>AC</li> <li>HDL</li> </ul> </div>	<p>X軸變數</p> <p style="text-align: center;">&lt;- -&gt; Waist</p>
	<p>Y軸變數</p> <p style="text-align: center;">&lt;- -&gt; DiaBP</p>

上圖使用分析方法中的迴歸分析，自變數為「腰圍(Waist)」，依變數為「心臟舒張壓(DiaBP)」，初步估計出兩變數的線性關係為  $y=0.447x+43.039$ 。

步驟二：參數設定中，將估計出的模型輸入並點選「儲存目前的函數」建立函數，將 X 軸變數選入「Waist」，Y 軸變數選入「DiaBP」。進階選項中，可設定 X 軸範圍、函數曲線分類等，確認無誤儲存後，點選「繪製圖形」開始進行分析。(此處使用分層抽樣取出的 100 筆樣本作為範例資料)

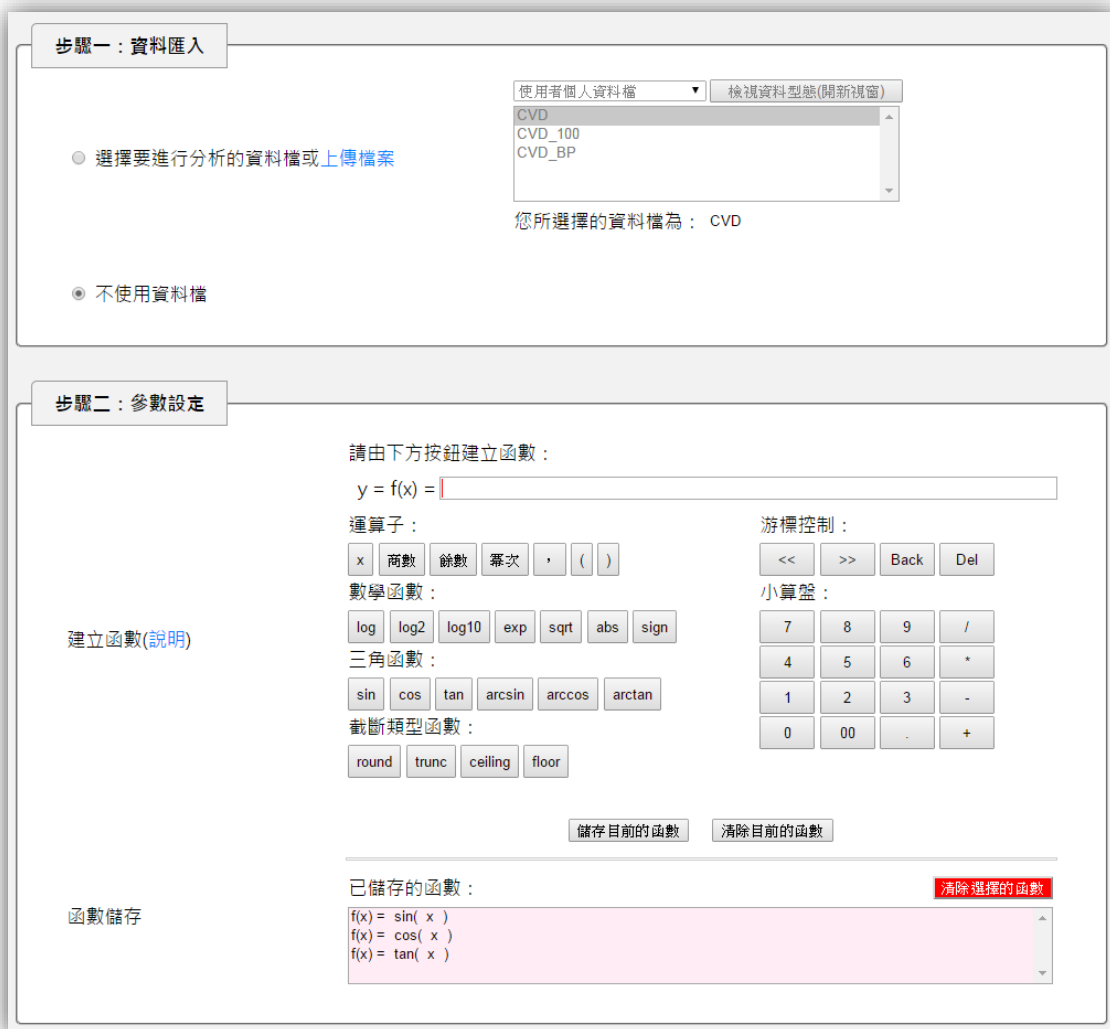
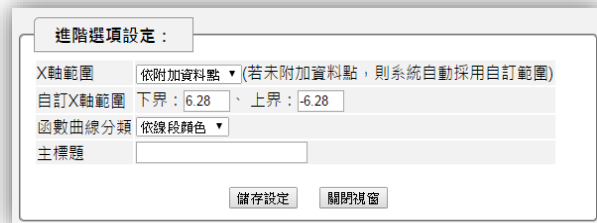


輸出結果，繪製出兩變數的散佈圖，與配適的迴歸函數。如下圖所示。

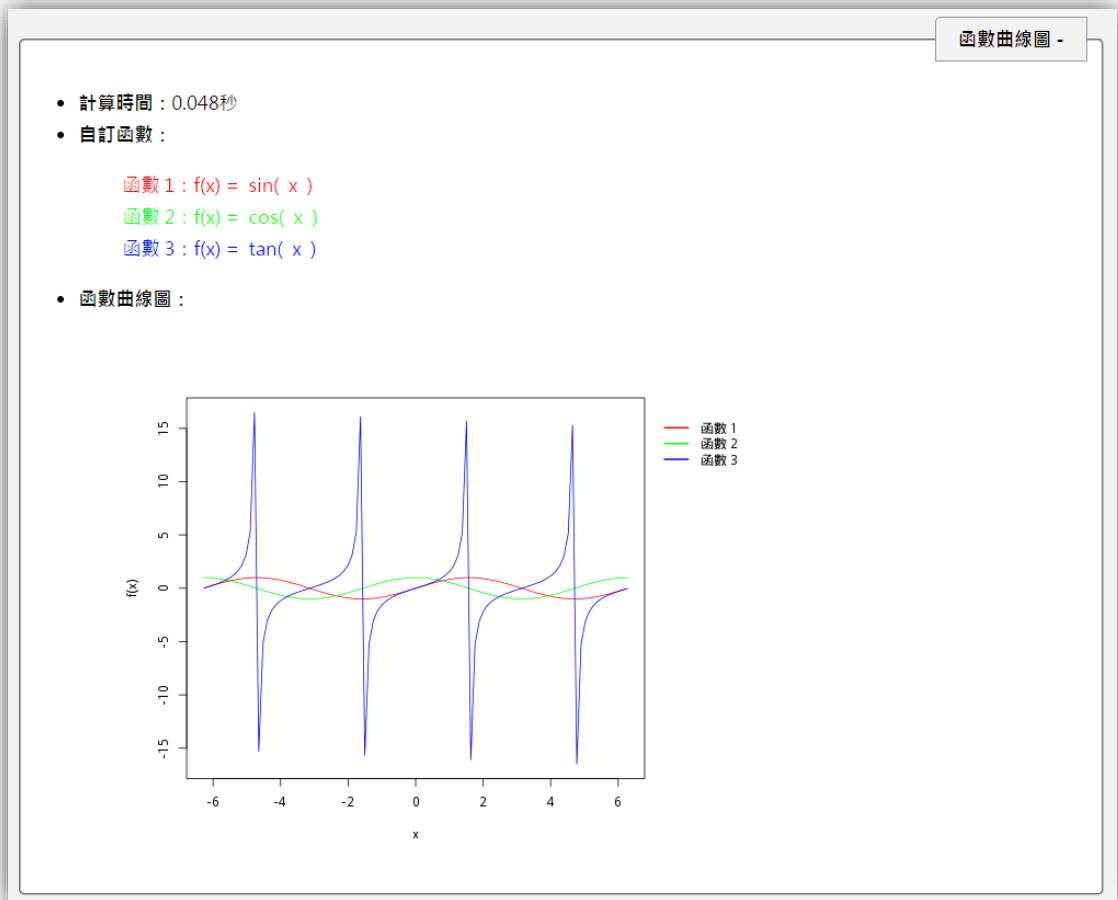


● 不使用資料檔

如下圖所示，若想了解三角函數圖形，則可依序輸入  $f(x)=\sin(x)$ ， $f(x)=\cos(x)$  與  $f(x)=\tan(x)$ ，分別點選「儲存目前的函數」即可將多函數同時儲存。進階選項設定中，我們自訂 X 軸範圍設定，下界：6.28、上界：-6.28，確認後，即可儲存並開始分析。



輸出之三角函數， $\sin(x)$ ， $\cos(x)$ 與  $\tan(x)$  圖形如下圖所示，系統會依照使用者輸入之不同函數分別依顏色做區別。



### ➤ 三維曲面圖

三維曲面圖，使用者可依不同需求，於 3D 空間中繪製出不同的曲面圖。亦可同時將資料點附於圖上，同時觀察三種變數與圖形之間其相關性。由於資料檔使用方法與二維曲面圖類似，在此僅簡單介紹三維曲面圖的繪製方法。在 R-web 內依序點選主選單中【圖表繪製】→【曲線(面)圖】→【三圍曲面圖】進行繪製。

步驟一：資料匯入

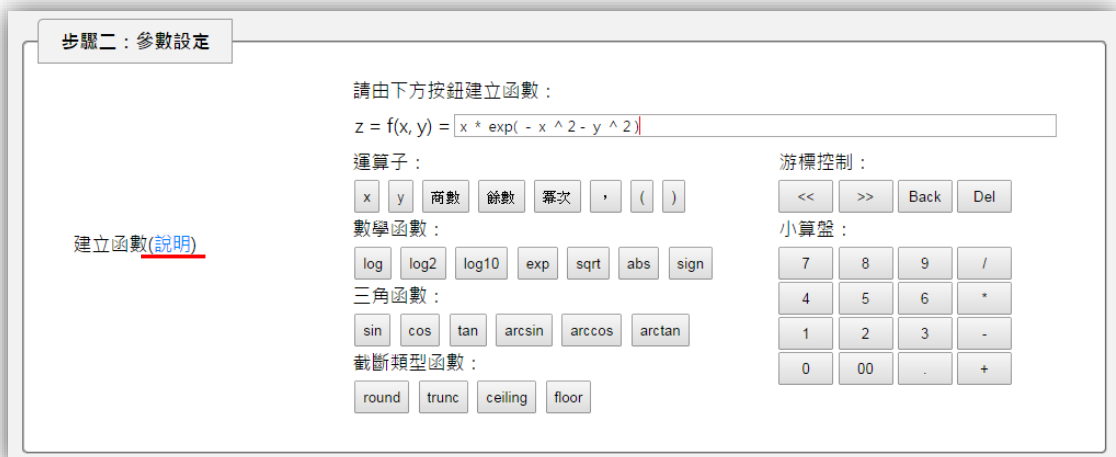
== 請選擇一個資料夾 == ▾ 檢視資料型態(開新視窗)

== 請選擇一個資料夾 ==

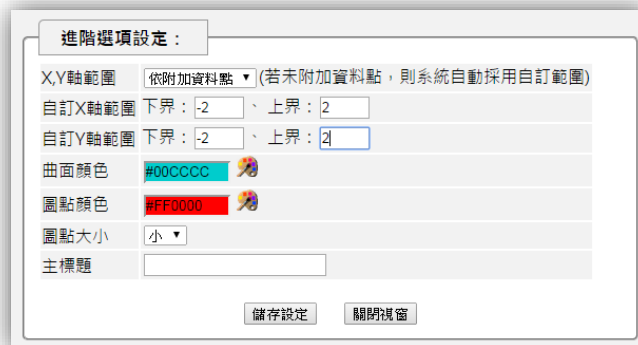
● 選擇要進行分析的資料檔或上傳檔案

● 不使用資料檔

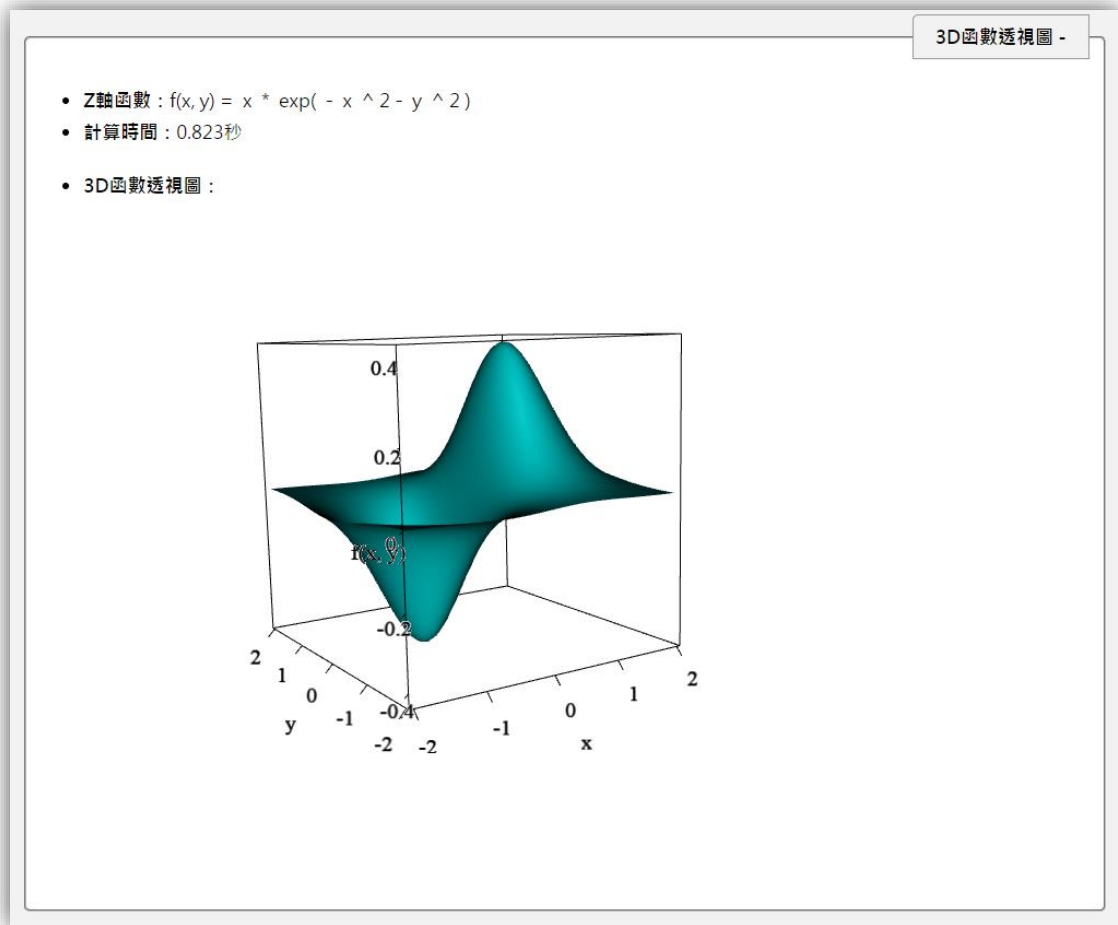
您所選擇的資料檔為： == 尚未選擇資料檔 ==



操作畫面如上圖所示，利用按鈕點選建立「 $z = f(x, y) = x * \exp(-x^2 - y^2)$ 」此函數，若有不了解按鈕的定義，可點選圖中之「說明」查看。進階選項中，除了可設定 X 軸與 Y 軸的範圍外，同時可設定個人喜好之顏色、圖點大小與主標題命名等設定。



繪製出的圖形，如下圖。繪製後的圖形為互動式，使用者可利用滑鼠滾輪調整圖形大小，或按壓住滑鼠左鍵轉換各種不同的角度以觀察函數或資料的特徵。



## ➤ 長條圖

長條圖，顧名思義是由長條方形的長度來詮釋每一個不同類別次數的大小，僅能用於離散型的變數資料，其各相鄰長條間彼此不可相連接。常用來觀測兩個或以上變數之間的關係。在 R-web 內依序點選主選單中【圖表繪製】→【長條圖】進行繪製。

**步驟一：資料匯入**

選擇要進行分析的資料檔或上傳檔案  
 直接輸入已經分組之次數

使用者個人資料檔   
 CVD  
 CVD\_100  
 您所選擇的資料檔為：CVD

欲繪製長條圖個數  個、每個長條圖各  組

---

**步驟二：參數設定**

資料格式為   
 選擇欲繪製長條圖的變數   
 選擇分組變數

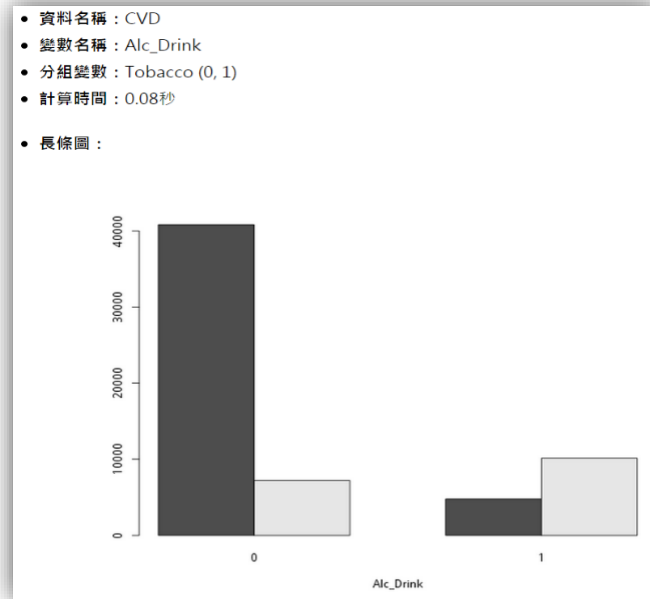
操作畫面如上圖所示，步驟一：資料匯入，可直接選取使用者之資料進行分析，亦可「使用直接輸入已經分組之次數」進行繪製。步驟二：參數設定中，可選擇資料的格式為「未經彙整之原始資料」或者為「已經彙整之分組次數」，此分析我們選取「抽菸習慣(Tobacco)」作為欲繪製長條圖的變數，分組變數則選取「有、無飲酒習慣(Alc\_Drink)」，進階選項設定中，可選擇各分組圖形呈現方式「並排」或「堆疊」，主標題、橫軸、縱軸標題等皆可修改，確認儲存後，即可點選「繪製圖形」開始繪製。

**進階選項設定：**

各分組圖形呈現方式

主標題   
 橫軸標題   
 縱軸標題   
 縱軸範圍 最小值： 最大值：

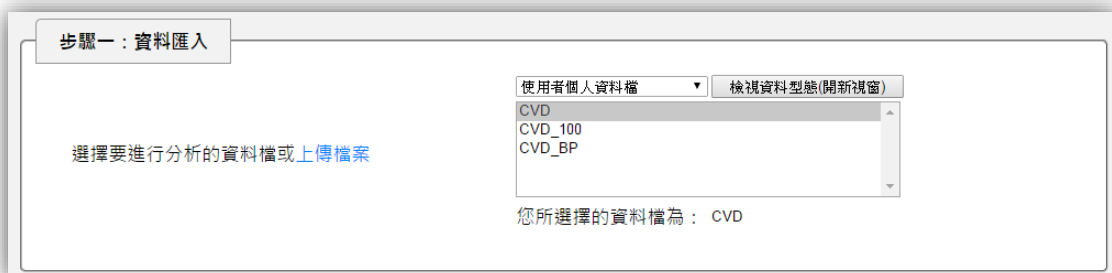




輸出圖形如上圖所示，由變數的定義知道 Tobacco – 0 表示沒抽菸習慣、1 表示有抽菸習慣，Alc\_Drink – 0 表示沒有飲酒習慣、1 表示有飲酒習慣。由上圖可以簡單的看出此筆資料，沒有飲酒習慣且同時沒有抽菸習慣的人之機率較沒有飲酒習慣且卻有抽菸習慣的人高。

## ➤ 直方圖

直方圖，非常類似長條圖，同樣是以長條方形的長度來詮釋每一個不同組別次數的大小，其與長條圖最大的不同點在於，其各相鄰長條間彼此互相連接。為研究者常使用的一種繪圖方法，其可初步了解資料分佈情形。於 R-web 內依序點選主選單中【圖表繪製】→【直方圖】進行繪製。



**步驟二：參數設定**

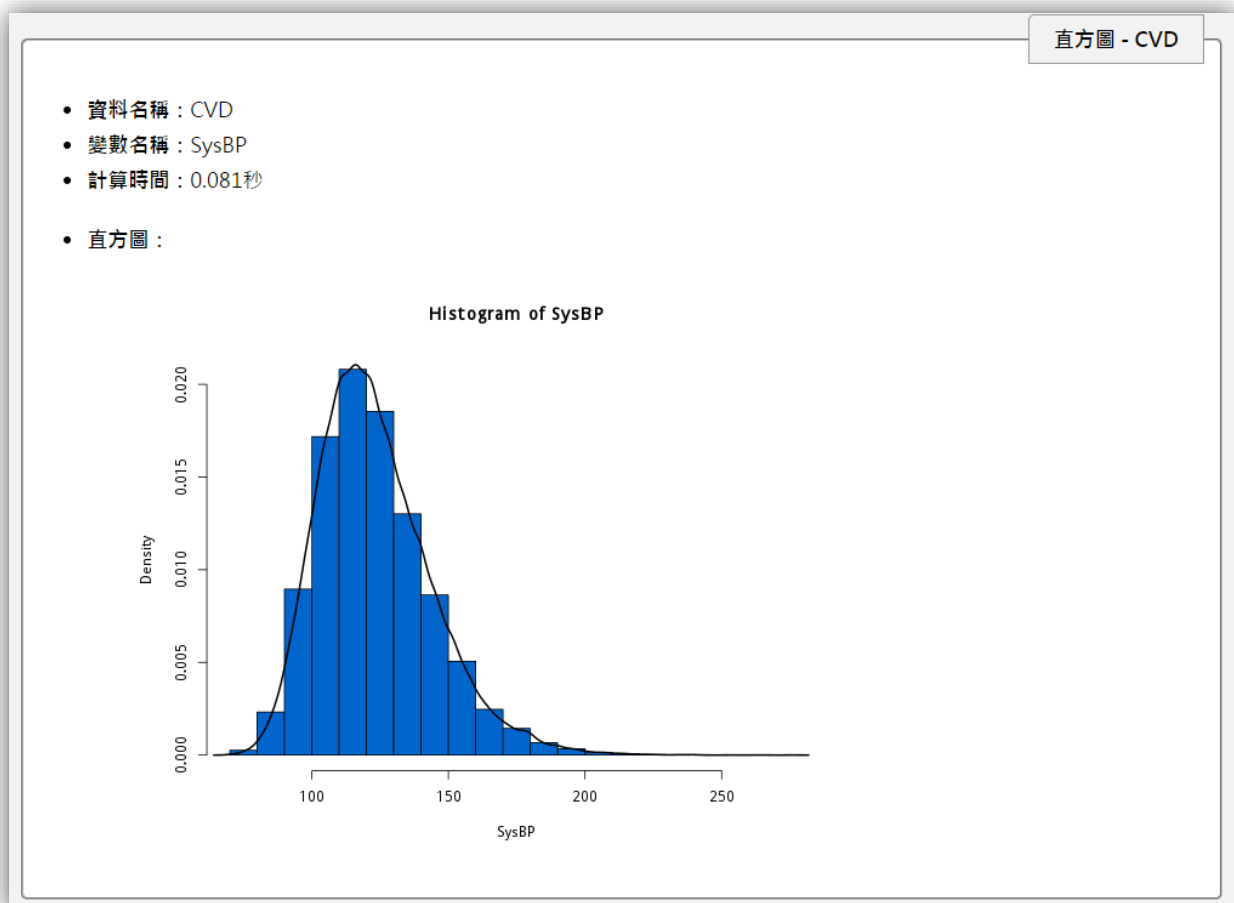
選擇變數 SysBP ▾

是否在圖形中加上密度估計曲線  是  否

是否在圖形中標示實際資料位置  是  否

操作方法如上圖，步驟二：參數設定中，選擇欲繪製之變數，亦可同時選擇是否需在圖形中「加上密度估計曲線」或「標示實際資料位置」。

假使研究人員想看看此筆資料中，「心臟收縮壓」的分佈情形，我們直接將變數 SysBP 選取放入就好囉！輸出結果如下圖所示，大部分的篩檢者「心臟收縮壓」集中於 110~130 之間，亦可看出此為一右偏分配，簡單來說，右偏分配的特性就是，平均數>中位數>眾數。



## ➤ 圓餅圖

圓餅圖是以圓當中的扇形面積比例大小來詮釋每一個不同類別次數的大小，而這些扇形區域合在一起會剛好是一個圓形，並且每個類別其扇形所張開的角度，應為 360 度乘以該類別的相對次數。此圖形在商業領域和大眾媒體中，可以說無處不在啊！R-web 中可依序點選主選單中【圖表繪製】→【圓餅圖】進行繪製。

The screenshot shows the R-web software interface for creating a pie chart. It is divided into two main sections: "步驟一：資料匯入" (Step 1: Data Import) and "步驟二：參數設定" (Step 2: Parameter Setting).

**步驟一：資料匯入**

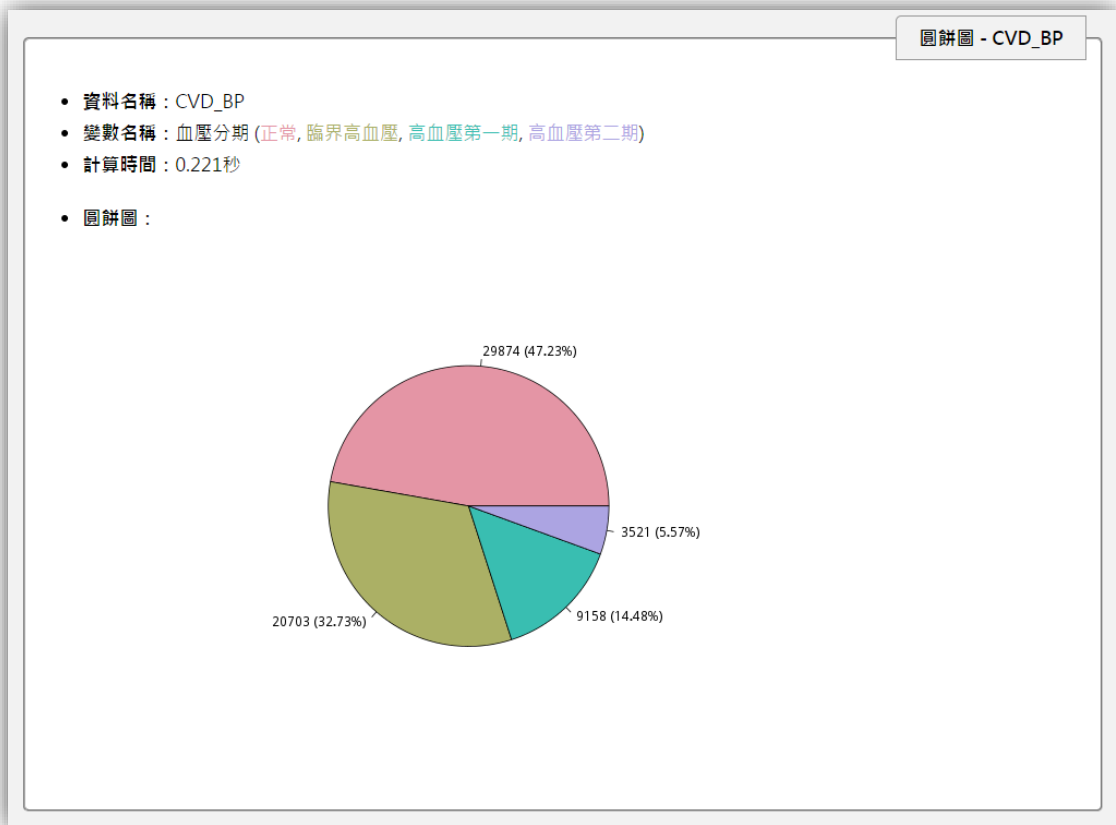
- Option 1: 選擇要進行分析的資料檔或上傳檔案 (Selected). A dropdown menu shows a list of files: CVD, CVD\_100, and CVD\_BP (highlighted). Below the menu, it says "您所選擇的資料檔為： CVD\_BP".
- Option 2: 直接輸入已經分組之次數.
- 欲繪製組數: 請選擇 ▾ 組

**步驟二：參數設定**

- 資料格式為: 未經彙整之原始資料 ▾
- 選擇欲繪製圓餅圖的變數: 血壓分期 ▾

操作畫面如上圖，步驟一：資料匯入中，我們運用的資料有先經過「資料分組」，欲操作的使用者可於 R-web 主選單中【資料處理】→【資料分組】進行分組。利用變數「心臟收縮壓(SysBP)」新增加一變數「高血壓分期」，將資料分為四類，正常(90 - 119)、臨界高血壓(120 - 139)、高血壓一期(140 - 159)、高血壓二期( $\geq 160$ )。此變數之分類方法是參考維基百科所建立的(<http://en.wikipedia.org/wiki/Hypertension>)。

將資料儲存為個人資料檔「CVD\_BP」後，步驟二：參數設定中，於「選擇欲繪製圓餅圖的變數」中選入剛剛所新增的變數「血壓分期」，而進階選項中一樣可設定主標題名稱，確認無誤後，點選「繪製圖形」進行分析吧！



輸出結果畫面如上圖，可看出此次的整合篩檢計畫中，47.2%的人「心臟收縮壓」檢測為正常值，臨界高血壓的比例佔 32.73%，高血壓第一期的比例佔 14.48%，高血壓第二期的比例佔 5.57%。有興趣的讀者可以試試使用變數「心臟舒張壓(DiaBP)」來進行分組畫畫看喔！

## ➤ 盒鬚圖

盒鬚圖的功能非常的多，可以用來看出分配的位置、分配的範圍、分配的分散程度和分配的偏態等等，其中也包含了最小值、第一四分位數、中位數、第三四分位數、最大值與 IQR 等訊息，還可以協助用來判斷離群值的資訊喔！R-web 中可依序點選主選單中【圖表繪製】→【盒鬚圖】進行繪製。

**步驟一：資料匯入**

選擇要進行分析的資料檔或 [上傳檔案](#)

使用者個人資料檔
檢視資料型態(開新視窗)

CVD  
CVD\_100  
CVD\_BP

您所選擇的資料檔為： CVD\_BP

**步驟二：參數設定**

選擇欲繪製盒鬚圖之變數

ID  
CVD  
Gender  
Waist  
SysBP

->

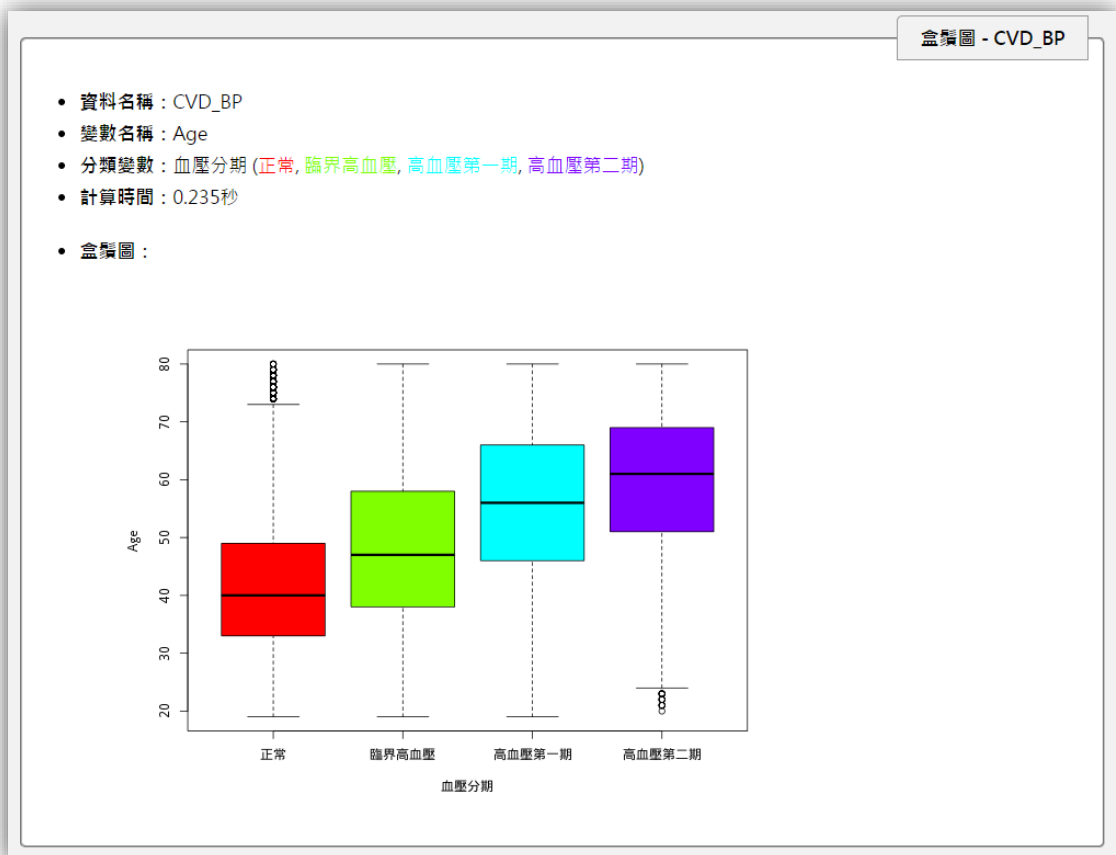
Age

<-

選擇分類變數  
(若選用此項則僅能繪製單一變數盒鬚圖)

血壓分期

上圖為我們的操作畫面，我們一樣使用剛剛所建立的資料「CVD\_BP」來做為本次的分析。步驟二：參數設定中，選擇欲繪製盒鬚圖之變數後，可同時選擇「分類變數」，選擇完畢後，點選「繪製圖形」開始繪製。



上圖為我們的輸出結果畫面，此筆資料可以明顯得看出篩檢計畫中的人群，無論是第一四分位數、中位數或是第三四分位數，「心臟舒張壓」都隨著年齡的增長而越來越高。

本期生統 eNews 就介紹到此囉！我們後半段的圖形繪製方法帶讀者多認識了 R-web 中的二維曲線圖、三維曲線圖、直方圖、長條圖、圓餅圖及盒鬚圖，以上 R-web 的基本圖形繪製方法也介紹到一個小段落。下一期的生統 eNews 將開始介紹「分析方法」在 R-web 中的應用，敬請期待！