**R套件教學簡介**

\*\*\*在正式進行實作課程之前，以下幾點建議提供學員們前置作業時參考：

1. 建議攜帶個人筆電，事先安裝完成R及R studio，安裝細節可參考以下網頁

<http://www.cc.ntu.edu.tw/chinese/epaper/0030/20140920_3006.html>

<http://cran.csie.ntu.edu.tw/bin/windows/base/>

<http://joe11051105.gitbooks.io/r_basic/content/environment_settings/RStudio_introduction.html>

<https://www.rstudio.com/products/rstudio/download/>

<http://www.r-tutor.com/elementary-statistics/numerical-measures/skewness>

1. 以下指令設定工作路徑，在讀取或寫入檔案時，可以省掉很多麻煩。

setwd("C:\\Users\\Chih-Hsuan Wang\\Desktop\\R檔案\\")

setwd("C:/Users/chtti/Desktop/R檔案")

1. 請先將個人檔案存成DAT、CSV或TXT檔，在R環境下可用以下指令抓取

heart<- read.table("heart.dat", header=T, sep=" ")

glass<- read.table("glass.txt", header=T, sep=",")

bank<- read.table("bank.csv", header=T, sep=";")

rm(list=ls()) #清除記憶體指令，避免先前的殘留資料

1. 上課時不會對外部資料庫進行直接連結，也不會進行資料清理或格式不一致的處理；同時建議欄位名稱盡量用英文命名，避免許多原因未明的操作干擾
2. 若有缺失值欄位(屬性)的資料請事先補齊(平均數、中位數、眾數)或直接刪除，避免屆時R套件處理發生問題(查詢缺失值之指令如下)，亦可用迴歸方式進行差補

\*which(is.na(data[, 4])==TRUE)

1. R操作的範例檔案已壓縮存在資料夾中，有興趣的學員可以事先感覺一下

\*\*\*2006年在香港舉辦的IEEE的ICDM國際會議選出了10大最具影響力的資料採礦演算法，列舉如下：C4.5, K-means, SVM, Apriori, EM, PageRank, AdaBoost, KNN, Naïve Bayes, CART，除PageRank外，其他皆在本次授課範圍