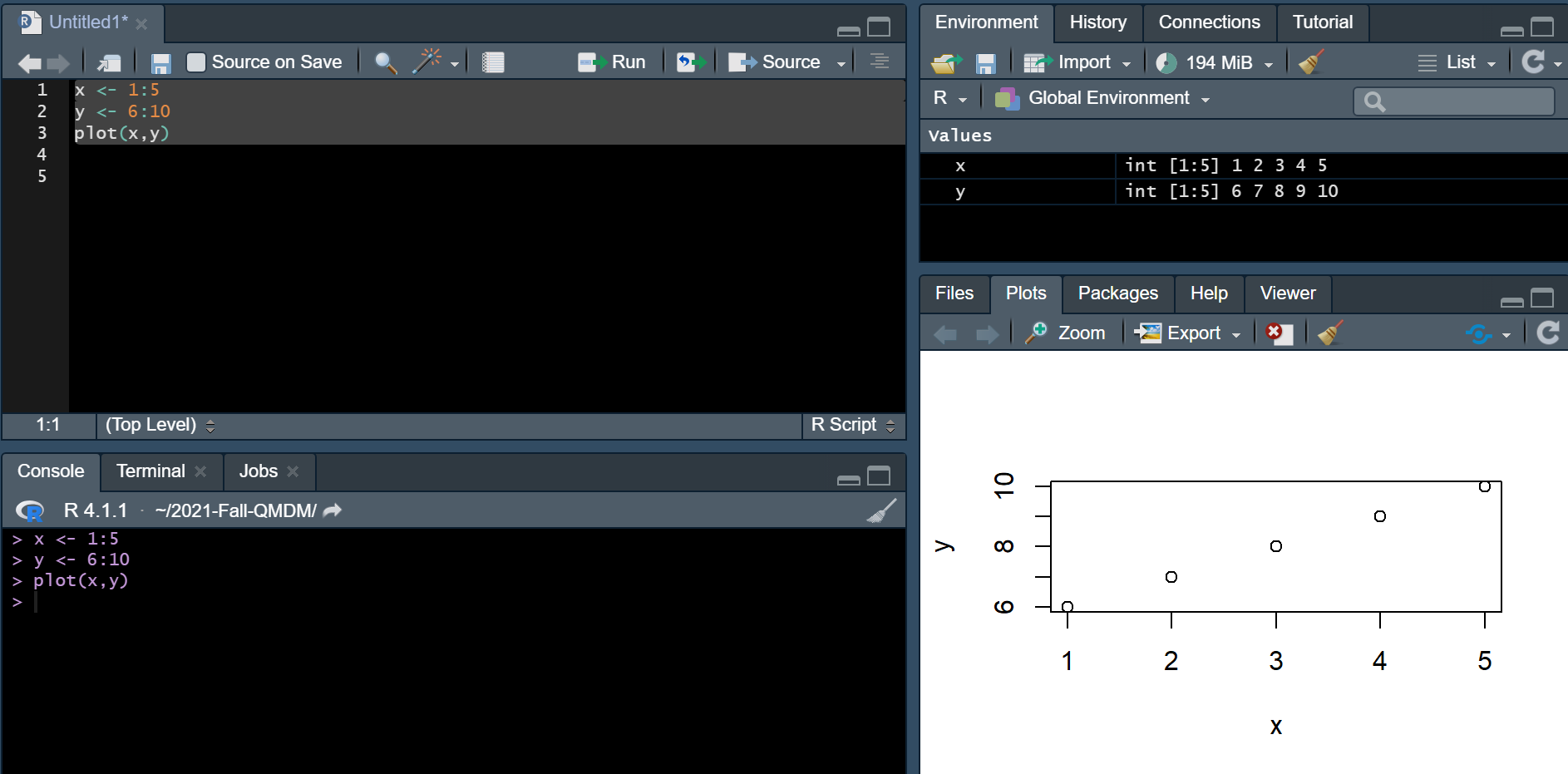
1-1︱What is RStudio and Why Should You Download It?

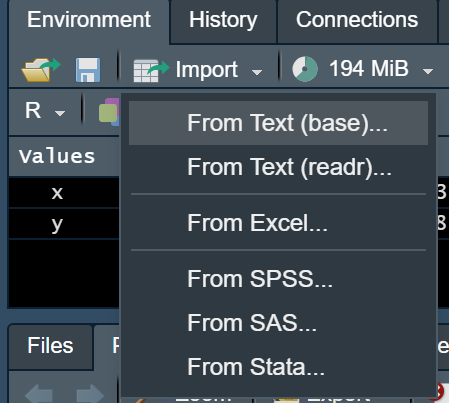
RStudio: A free, open source integrated environment for R language, helping R keep more organized and adding more functionality to it.

RStudio是為R語言設計的一種跨平台整合開發環境。其特色包括可客製化的軟體套件、視覺化介面與同團隊開發的一系列數據可視化與出版工具。

* RStudio 有分別的Console、Memory、Plots區塊，亦可直接儲存匯出視覺化的物件。

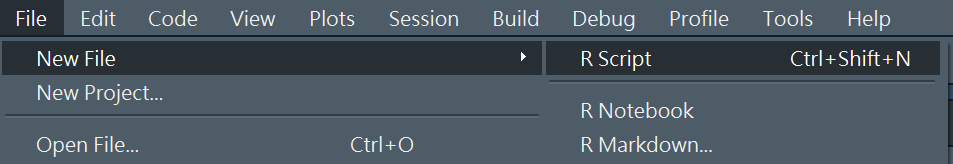


* 使用RStudio可以輕易地匯入各式資料。



* R Script: 日後可再用RStudio開啟編輯

R Markdown: 可將R程式碼嵌入其他如Word, PDF等檔案

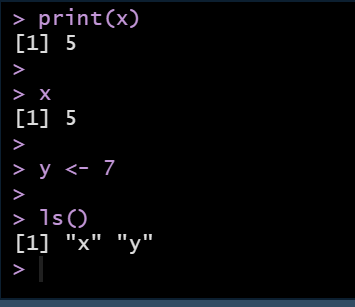


1-2︱Download and Install R and RStudio

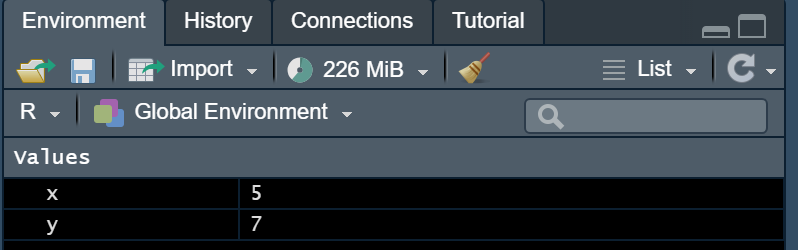
要先下載R語言，再下載RStudio。

1-3︱Getting started with R: Basic Arithmetic and Coding in R

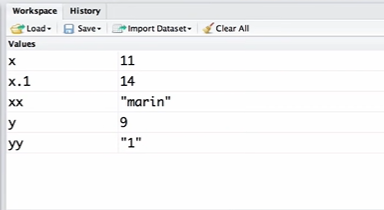
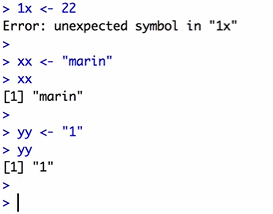
* 可使用=或<- 指派數值



* Ls() 可找出各變數被指派的數值



* rm(物件)可移除Memory中已存在的物件/變數
* 變數前面必須為「字元」開頭
* 若要指派字元/字串給某變數，必須加上引號
* 若指派的「數值」加上引號時，則會被視為指派「字元」

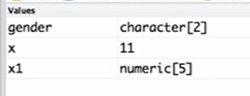
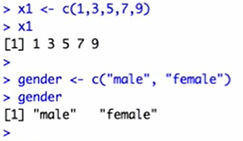


* 四則運算：加+ 減- 乘\* 除/ 次方^ 開根號sqrt() 取對數log() 取絕對值abs()

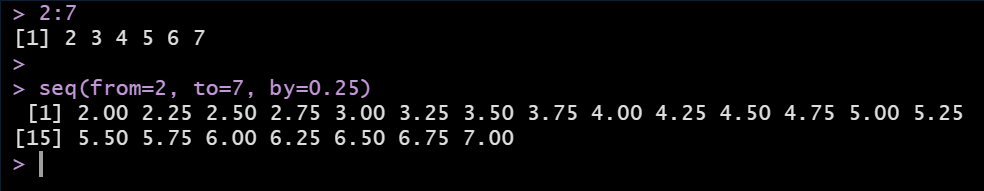
1-4︱Create and Work with Vectors and Matrices in R

* 建立/合成一個向量/column(行)： Concatenate 串接函數

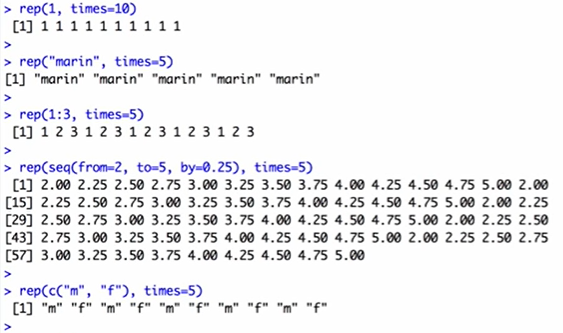
→c(數字 或 “字元”)



* 找出整數/等差數列



* 重複數字、字元、數列、字串→rep()
* 字元/非連續數字的合成→使用Concatenate 串接函數



* 向量y <- c(1, 3, 5, 7, 9)  
  找出y的第3個分量/元素→y[3]

找出y除了第3個分量以外的其他分量→y[-3]

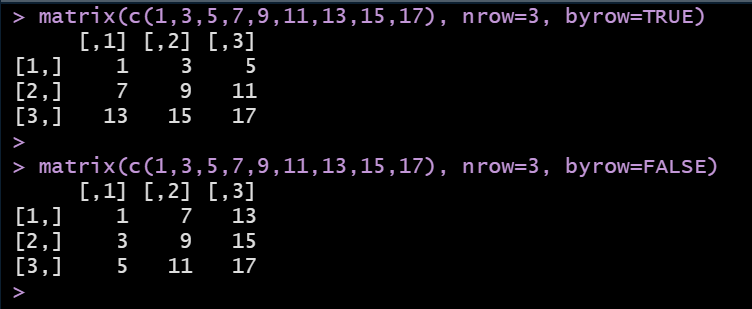
找出y的第1到3個分量→y[1:3]

找出y的第1和第5個分量→y[c(1, 5)]

找出y除了第1和第5個分量以外的其他分量→y[-c(1, 5)]

找出y的所有分量中小於6的分量→y[y<6]

* 建立矩陣：按照列排序/按照行排序，TRUE和FALSE要全大寫



* 矩陣mat <- matrix(c(1,3,5,7,9,11,13,15,17), nrow=3, byrow=TRUE)

　　[ , 1][ , 2][ , 3]  
[1, ] 1 3 5   
[2, ] 7 9 11  
[3, ] 13 15 17  
找出第2列的元素→mat[2, ]

找出第1行的元素→mat[ , 1]

找出第1列第2行的元素→mat[1, 2]

找出第1、3列第2行的元素→mat[C(1, 3), 2]

1-5-a︱Import Data, Copy Data from Excel to R CSV & TXT Files

(大部分電子表格亦適用)

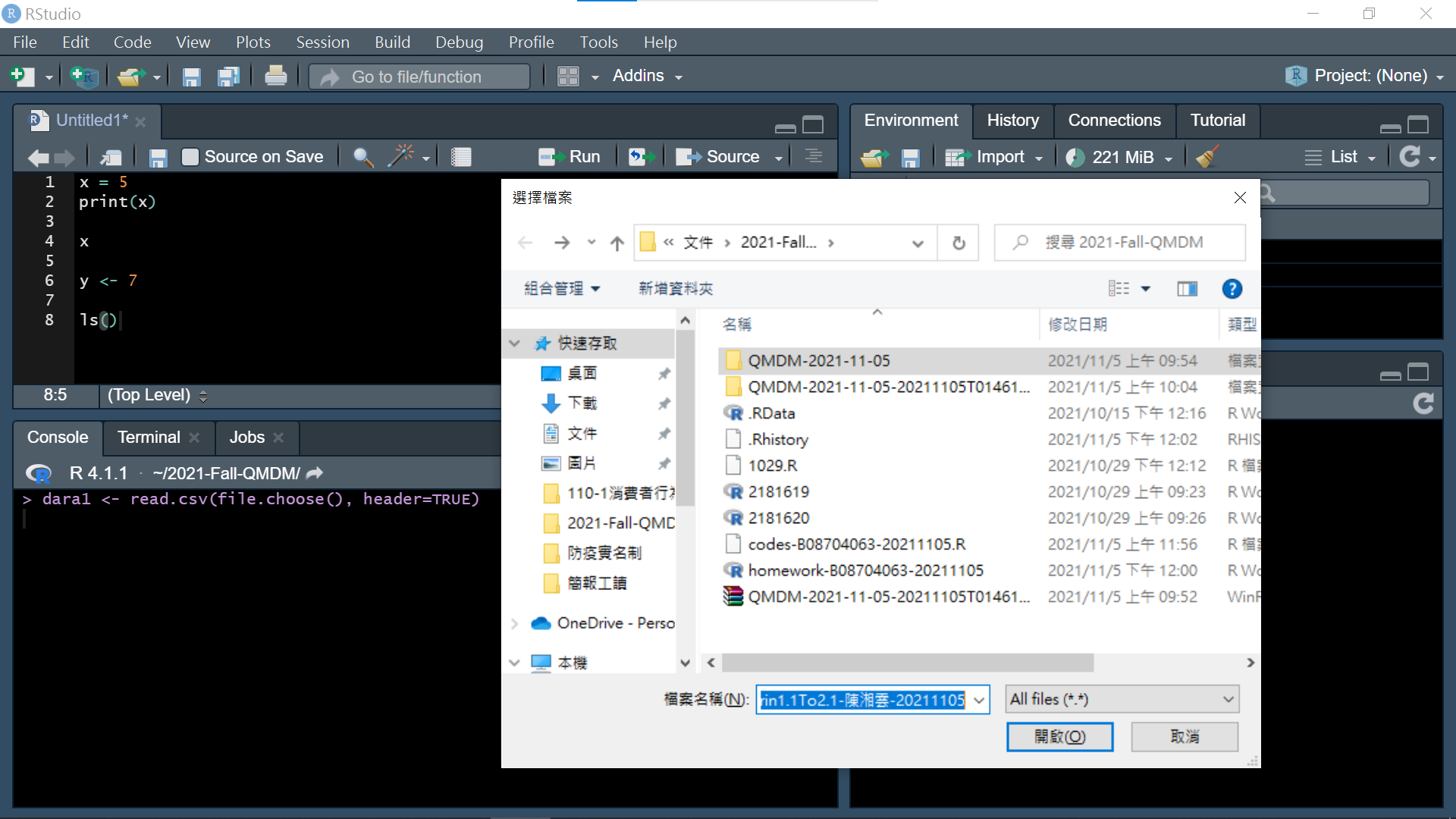
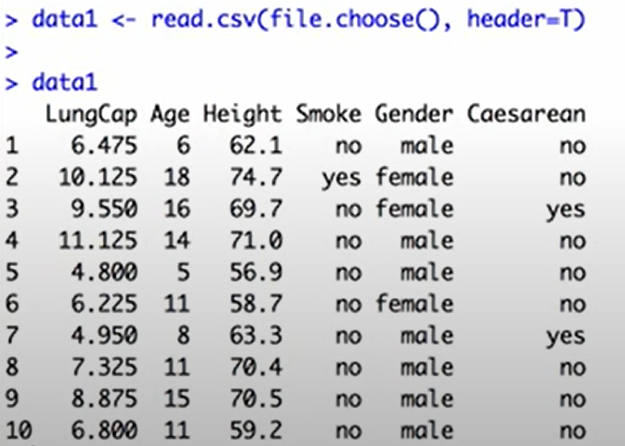
* 兩種表格資料的純文字檔：

CSV: Comma Separated Value (以逗號分隔資料)→easier and better for R

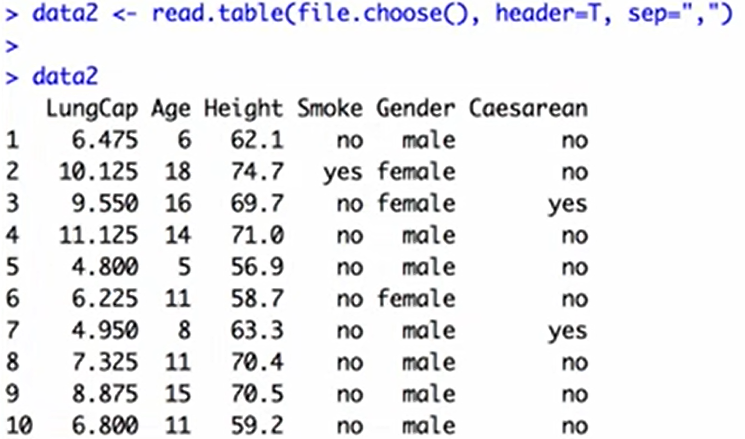
TXT: Tab Delimited Tax File (以tab、分號或其他符號分隔資料)

* 讀取檔案資料夾中的**csv檔**
* Header=TRUE → 資料檔的第一列/首標會被視為變數

<方法一>



<方法二>



* 讀取檔案資料夾中的**txt檔**

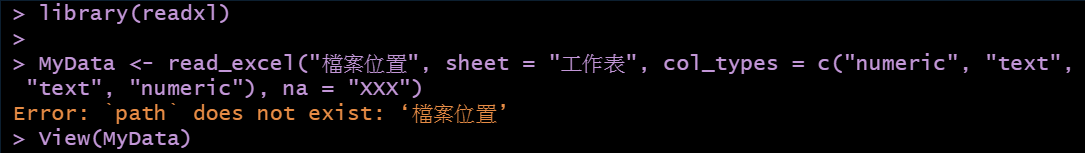
<方法一>



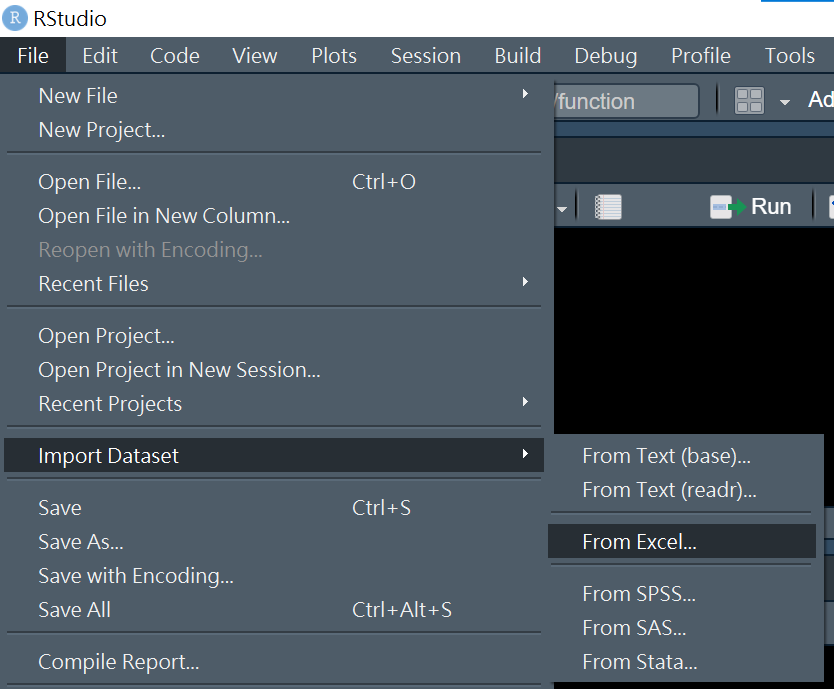
<方法二>



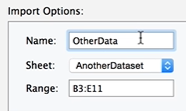
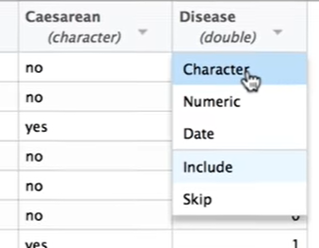
1-5-b︱Importing/Reading Excel data into R using RStudio (readxl)

* RStudio內建的選單或readxl套件可匯入.xlsx 和.xls檔：  
  

<選單>



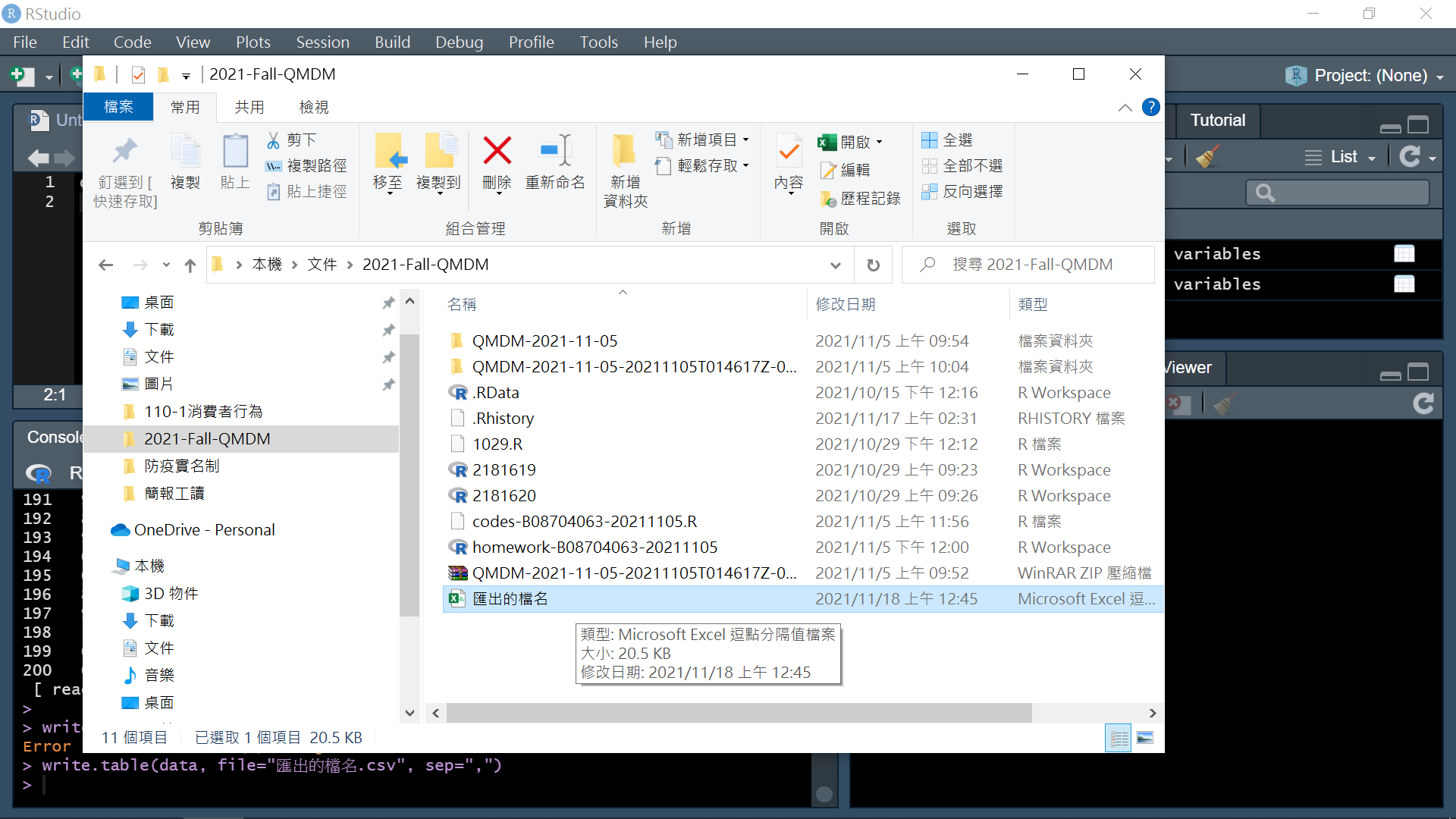
* Sheet選項：Default = 預設值 = 第一個工作表
* First row as names→First row of the data set will be the variable names
* 點選變數右下角的三角形，可改變該變數欄位的類別(column types)
* 選擇Skip，可不匯入該欄位資料
* Range: 輸入由左上儲存格到右下儲存格的欲匯入範圍



1-6︱Export Data from R (csv , txt and other formats)

* 匯出資料成CSV檔

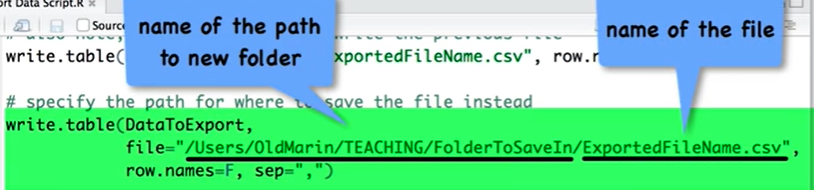




* 若不想要匯出row name，則加入row.names=F



* 將資料匯出另存在別的檔案料夾：



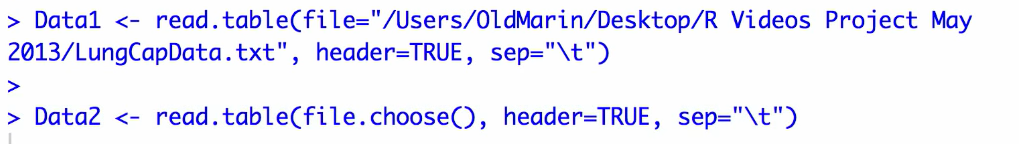
* 匯出資料成txt檔和其他格式：

write.table(R當中的資料名稱, file="匯出的檔名.txt", sep="\t")

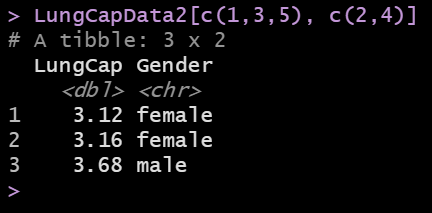
write.table(R當中的資料名稱, file="匯出的檔名", sep=" ")

1-7︱Importing , Checking and Working with Data in R

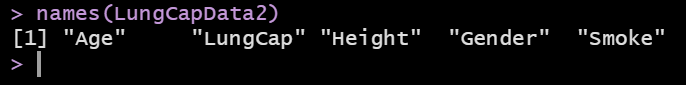
* 2種**讀取資料**的方法：



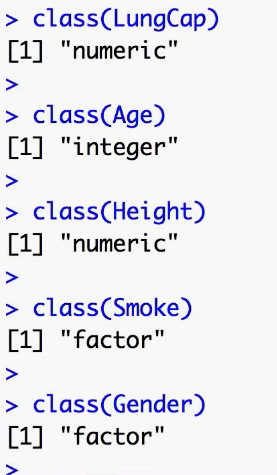
* 移除已匯入的某個data frame→rm(資料名)
* 找出某個data frame的row和column數量→dim(資料名)
* **顯示某個data frame的前6列資料→head(資料名)**
* **顯示某個data frame的後6列資料→tail(資料名)**
* 顯示data frame的某幾列某幾行：



* 顯示data frame的表頭：

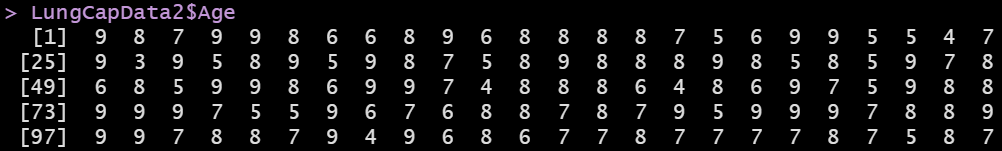


* Class()函數可找出個變數的類型



1-8︱Working with Variables and Data in R

* **使用$符號引用/取出表格的某一行**



* **用 attach()函數，可將數據框添加到R的搜索路徑中**←→**detach()函數**
* 找出某一欄的平均數：

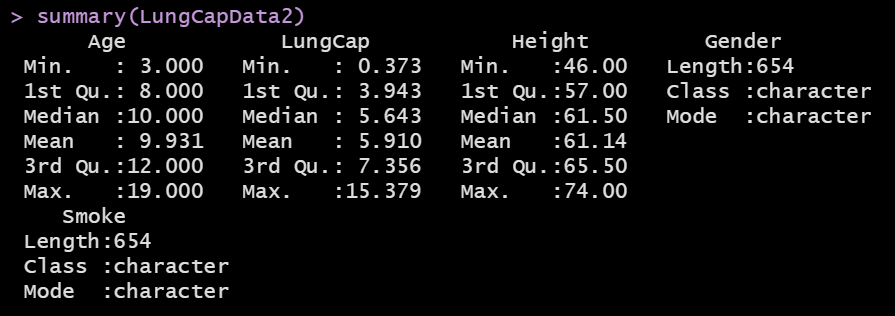
<方法1>



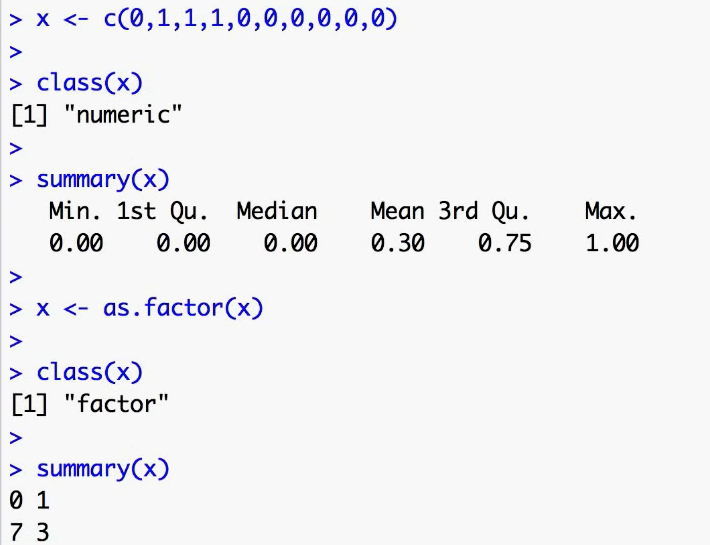
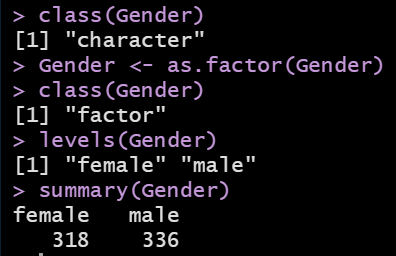
<方法2>



* Summary()函數

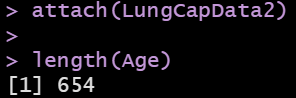


* 用as.函數將資料的類型(class)重設之後，可得到不一樣的summary結果

1-9︱Subsetting (Sort/Select) Data in R with Square Brackets

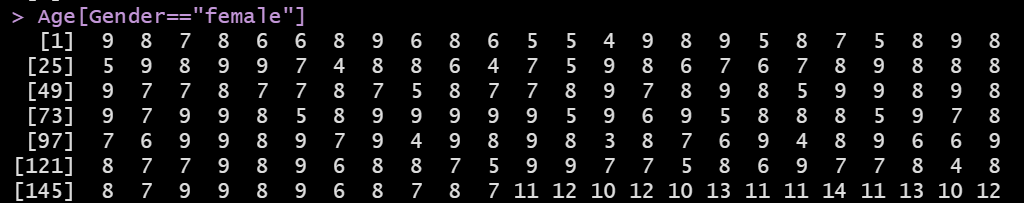
* Attach()+length()函數→找出某一欄的observation數量



* **利用[]符號，Subsetting (Sort/Select) Data (資料的子集)  
  （column=向量的概念；data frame=矩陣的概念）**

1. Age欄的第1-10列的資料：



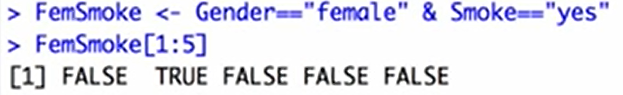
1. Age欄中與Gender為female者同一列的資料：  
   

* **=(一個等號)→指派的意思**
* **==(兩個等號)→等於/全等的意思**
* 使用&取交集(且)



1-10︱Logic Statements (TRUE/FALSE), cbind and rbind Functions in R

* 完全符合條件→TRUE；其餘者→FALSE



* **合併rows→rbind()函數**

> rbind(x, y) # 透過 row 合併。

[,1] [,2] [,3]

x 1 2 3

y 10 20 30

* **合併columns→cbind()函數**

> cbind(x, y) # 透過 column 合併。

x y

[1,] 1 10

[2,] 2 20

[3,] 3 30

1-11︱Setting Up Working Directory in R

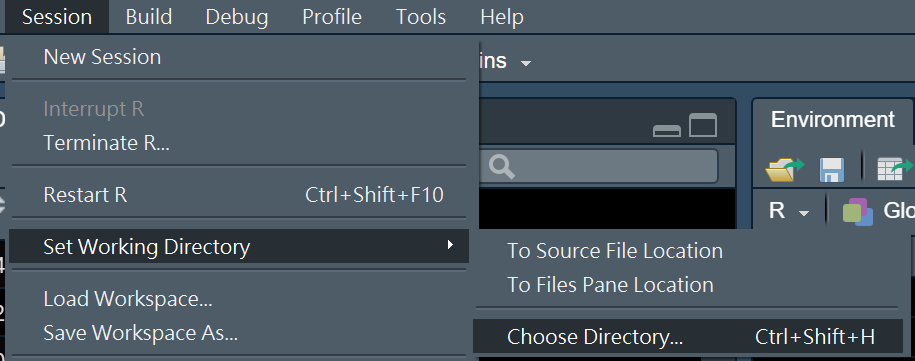
* A working directory is one spot ( e.g. a folder) created for saving all your works.
* 使用getwd()函數找出目前的working directory  
  
* 設定Working directory：  
  ＜方法一＞



＜方法二＞



＜方法三＞



* **儲存整個工作空間成RData文件**的方法：

> save.image("D:/Datatest/hello.RData")

若是要儲存在Working Directory，則檔案路徑可只輸「檔名.RData」

* **讀取RData文件**的2種方法：

＜方法一＞

> load("D:/Datatest/hello.RData")

若是要從Working Directory讀取，則檔案路徑可只輸「檔名.RData」

＜方法二＞

> load(file.choose())

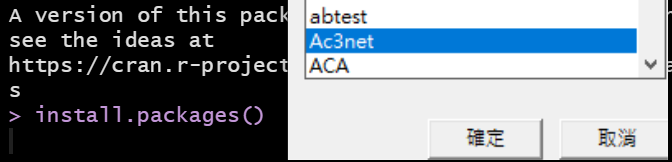
1-12︱Writing Scripts in R

* An **R script** is just a plain **text** file that you save R code in**（.R）**
* **RData**文件是整個**工作空間(workspace)（.RData）**

1-13︱How to Install Packages in R

* 安裝某套件





* **Libraries disappear when ending an R session, so must load the library for those packages want to use during each R session.**

****

1-14︱Customizing The Look of R Studio

