R Basics

陳曾基

國立陽明交通大學醫學院醫務管理研究所
國立陽明交通大學醫學院急重症醫學研究所
國立陽明交通大學醫學院醫學系家庭醫學科
臺北榮民總醫院家庭醫學部
臺北榮民總醫院醫學研究部大數據中心

課前作業

上課前請先安裝套件： tidyverse gridExtra car

install.packages(c("tidyverse", "gridExtra", "car"))

課前說明 (1)

- 示範程式所用的資料(教學檔)，係依據舊版國家衛生研究院全民健康保險研究資料庫格式 (https://nhird.nhri.org.tw/date_02.html)，隨機仿製的虛擬資料，僅供程式練習用，內容可能有很多不合理的地方，例如：男性有婦科疾病、費用兜不攏。

課前說明 (2)

- 請將示範資料(教學檔)複製於D槽SampleData子目錄(D:/SampleData/)下，將可不用更改課程提供的相關程式碼。如果個人電腦沒有D槽或習慣將檔案擺放於其他位置，則在匯入資料時，需自行於程式碼內更改檔案位置。

- 匯入資料程式與示範資料置於相同的子目錄下

注意

- 複製程式碼至MS PowerPoint或Word時，或在MS PowerPoint或Word裡更動文字時，MS Office往往會自動將程式碼內直的單引號(')與雙引號(")改為斜的單引號(‘)與雙引號(“)。若未注意，將程式碼從MS PowerPoint或Word複製至R時，R將無法判讀，形成Error，需自行更正。

Appetizer Only

R you ready

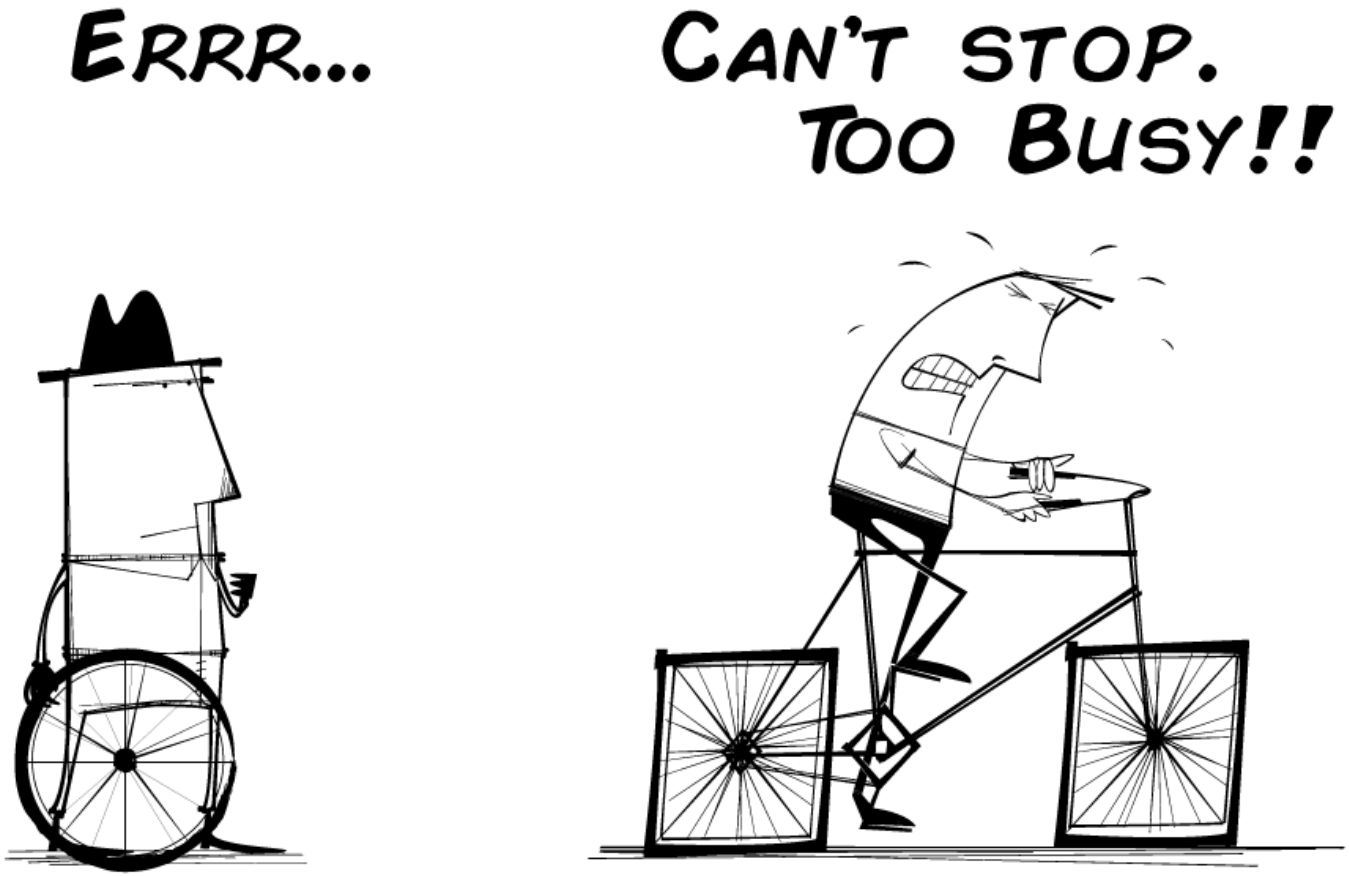
沒有人是修完游泳學分後，才進游泳池。

被丟進游泳池，才學會游泳；
被放逐到海洋，才發現寶藏。(鄧紫棋)

Do learn swimming in the pool

https://pixabay.com/illustrations/kawaii-animals-swimming-pool-kawaii-4090024/

Don’t reinvent the wheel



https://www.flickr.com/photos/toddle_email_newsletters/15412982829

https://en.wikipedia.org/wiki/R_(programming_language)

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes
Upload file
Special pages
Permanent link

R (programming language)

From Wikipedia, the free encyclopedia

R is a programming language and free software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing.[6] The R language is widely used among statisticians and data miners for developing statistical software[7] and data analysis.[8] Polls, surveys of data miners, and studies of scholarly literature databases show that R's popularity has increased substantially in recent years.[9] As of April 2018, R ranks 12th in the TIOBE index.[10]

R is a GNU package.[11] The source code for the R software environment is written primarily in C, Fortran, and R.[12] R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems. While R has a command line interface, there are several graphical front-ends available.[13]

Contents [hide]

R

Paradigms	Multi-paradigm: Array, object-oriented, imperative, functional, procedural, reflective
Designed by	Ross Ihaka and Robert Gentleman
Developer	R Core Team[1]
First appeared	August 1993; 24 years ago[2]
Stable release	3.4.4 (Someone to Lean On)[3] / March 15, 2018; 26 days ago

Ross Ihaka

Robert Gentleman

https://www.computerhope.com/people/pictures/robert_gentleman.jpg

https://en.wikipedia.org/wiki/Ross_Ihaka
https://en.wikipedia.org/wiki/Robert_Gentleman_(statistician)

R
as a Swiss knife or one part of a Swiss knife (Language - Software - Platform)

https://www.flickr.com/photos/ajc1/4663140532

Strengths of R

- Statistics
- Data processing
- Graphics

Outline of this course

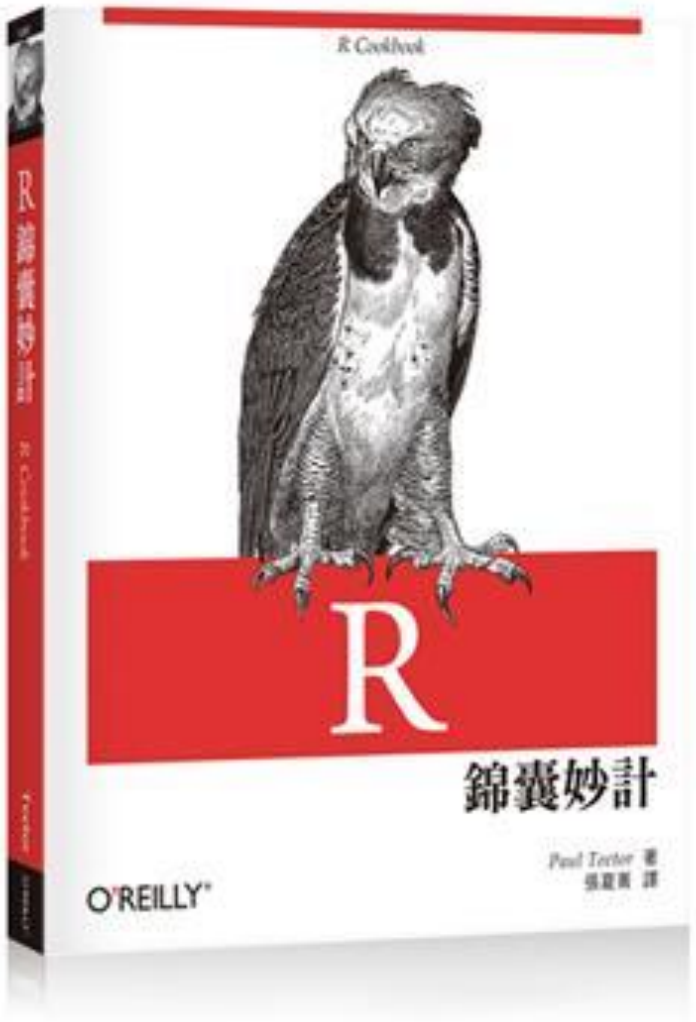
- R Basics : data processing + statistics
- R Graphics : graphics
- tidyverse : data processing
- ggplot2 : graphics

Topics

- R basics
- Data import
- Simple data processing
- R statistics
- Real examples
- Resources

Omission of functions and data merge/select/recode/transform (which will be dealt with in tidyverse)

Lecture Source :



https://rc2e.com/

第二版 全文免費 隨時更新內容

R Cookbook, 2nd Edition

R Cookbook, 2nd Edition

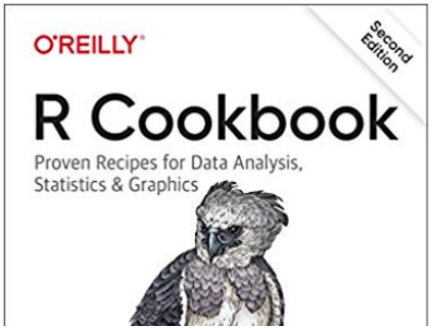
James (JD) Long

Paul Teetor

2019-09-26

Welcome to the R Cookbook 2nd Edition

R is a powerful tool for statistics, graphics, and statistical programming. It is used by tens of thousands of people daily to perform serious statistical analyses. It is a free, open source system whose implementation is the collective accomplishment of many intelligent, hard-working people. There are more than 10,000 available add-on packages, and R is a serious rival to all commercial statistical packages.



SECTION I

R BASICS

Key Concepts

- 變數名稱大小寫有差別 [case sensitive]
- 賦值 [value assignment] <- (少用 = 或 ->)
- 註解 [comment] # ……
- 缺失值 [missing value] NA

< 與 - 緊連在一起
兩者間不能有空格

- 變數名稱命名原則
  - 只能由字母、數字、點、底線所組成
  - 起首只能是 [字母] 或 [點[第一字]與非數字[第二字]]

Example

- R Console視窗內 > 右側輸入一行指令後，按下 Enter鍵，結果顯示於下一行 [1] 右側 (* [1]表示是第一組答案)

```
1 + 2    # 按下Enter鍵
```

```
> 1 + 2
[1] 3
>
```

- 如果指令未完整輸入即按下 Enter鍵，下一行首顯示 +，等候繼續將指令輸完

```
1 +    # 按下Enter鍵
```

```
> 1 +
+ 2
[1] 3
>
```

- R Console可當作電子計算機使用

賦值

- 變數無需先宣告(declare)，也可隨時覆寫

```
x <- 1 + 2     # 按下Enter鍵
x   # 輸入變數名稱 會顯示值 3
print(x)   # 另一種作法
```

- 如果想省略上述步驟，直接顯示變數內容，可在指令前後加上括號

```
(x <- 1 + 2)     # 按下Enter鍵
(x <- "ABC")     # 按下Enter鍵
(x <- 'ABC')     # 按下Enter鍵
```

重要觀念

Data Structure

- **Vector 向量 [R的核心]**
  - Matrix 矩陣 [二維的向量]
  - Array 陣列 [多維的向量]
  - **Factor 因子** [儲存類別資料的變數，同時具有字串與整數的特性] [用以將類別資料分組]
  - Scalar 純量 [僅包含一個元素的向量]
- **List 串列 (或譯 列表)**
- **Data frame 資料框架 [資料處理存放處]**

Vector : 所屬的元素(element)有相同形式
List : 所屬的元素可有不同形式(mode)

僅供參考

Mode 形式

- 實際的資料類型 (physical type)

- numeric
- character
- logical
  - TRUE (T) : 可轉換為 1
  - FALSE (F) : 可轉換為 0
- list
- function
- …

僅供參考

Class 類別

- 抽象的資料類型 (abstract type)
- R會根據物件(object)的類別而決定如何處理物件

- numeric
- integer
- Date
- list
- factor
- …

僅供參考

Example

```
x <- 1
mode(x)    # "numeric"
class(x)   # "numeric"

x <- 1L    # 數字後面加L 表示是整數
mode(x)    # "numeric"  實際
class(x)   # "integer"  外顯
```

僅供參考

Example

```
x <- 1
mode(as.list(x))    # "list"
class(as.list(x))    # "list"
mode(as.factor(x))    # "numeric"  !!!  實際
class(as.factor(x))    # "factor"  外顯
# factor 在 R 裡面是以數字的方式儲存
```

僅供參考

Example

```
x <- c("a", "b", "c")
mode(x)    # "character"
class(x)    # "character"
mode(as.list(x))    # "list"
class(as.list(x))    # "list"
mode(as.factor(x))    # "numeric" !!!
class(as.factor(x))    # "factor"
# factor 在 R 裡面是以數字的方式儲存
```

僅供參考

Example

```
x <- c(T, F, T)    # 或寫成 c(TRUE, FALSE, TRUE)
mode(x)    # "logical"
class(x)    # "logical"
mode(as.list(x))    # "list"
class(as.list(x))    # "list"
mode(as.factor(x))    # "numeric" !!!
class(as.factor(x))    # "factor"
# factor 在 R 裡面是以數字的方式儲存
```

僅供參考

Example

```
x <- as.Date("2018-4-11")
mode(x)   # "numeric" !!!  實際
class(x)   # "Date"  外顯
# Date 在 R 裡面是以數字的方式儲存
mode(as.list(x))   # "list"
class(as.list(x))   # "list"
mode(as.factor(x))   # "numeric" !!!
class(as.factor(x))   # "factor"
```

重要觀念

Vector 向量

- 一串相同形式元素的集合

```
x <- 1:12   # 1至12 共12個元素 的 向量
y <- c(1, 3, 5, 7, 9, 11)   # c : concatenate
z <- c("John", "Mary", "Paul", "Jane")
mean(x)   # 所有元素的平均值
summary(x)
x + 1   # 向量裡 每個元素 加 1
x ^ 2   # 向量裡 每個元素 開平方
```

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    3.75    6.50    6.50    9.25   12.00
```

```
 [1]  2  3  4  5  6  7  8  9 10 11 12 13
```

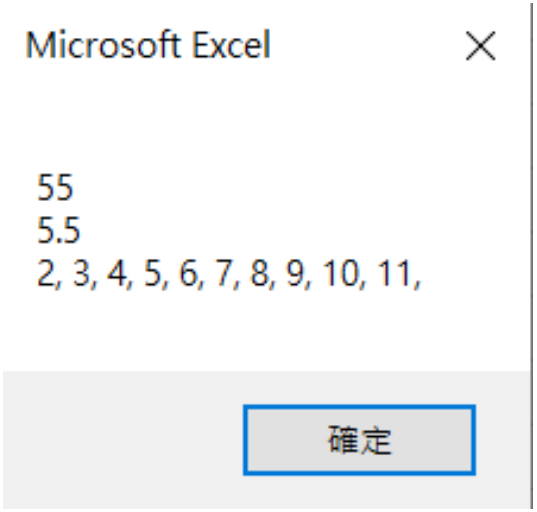
```
 [1]   1   4   9  16  25  36  49  64  81 100 121 144
```


c : combine, construct, create

傳統解法 (VBA)

- 1到10共10個數，總和？平均值？每個數各加1

```
Sub Z()
   For i = 1 To 10
      Count = Count + 1
      Total = Total + i
      Members = Members & (i + 1) & ", "
   Next i
   MsgBox Total & vbCrLf & (Total / Count) & vbCrLf & Members
End Sub
```



R式解法

- 1到10共10個數，總和？平均值？每個數各加1

x <- 1:10

sum(x); mean(x); x + 1

不同指令可寫在同一行，彼此用分號(;)隔開

```
> x <- 1:10
> sum(x); mean(x); x + 1
[1] 55
[1] 5.5
 [1]  2  3  4  5  6  7  8  9 10 11
```

Vector 向量 – 自動轉換形式

- 一個向量內所有元素必須屬於相同mode，如不依規則來賦值，R會自動轉換其值

(x <- c(1, 2))　# 1 2

(x <- c(1, "a"))　# "1" "a"

```
> (x <- c(1, 2))
[1] 1 2
> (x <- c(1, "a"))
[1] "1" "a"
```

Vector 向量 – 索引

- 向量可依位置編製索引

```
z <- c("John", "Mary", "Paul", "Jane")
z[2]   # "Mary"  表示向量z的第二個元素
```

- 向量可依多個索引，傳回一個子向量 (subvector)

```
z[c(2, 3)]   # "Mary", "Paul"
```

Vector 向量 – 純量

- 純量(scalar) : 僅包含一個元素的向量

```
x <- "John"
x[1]   # "John"
x[2]   # NA   # 該vector無第二個元素

x <- c("John")
x[1]   # "John"
x[2]   # NA
```

Vector 向量 – 名稱

- 個別向量元素，除了索引，還可擁有名稱 (name)

```
x <- c(10, 20, 30, 40)
names(x) <- c("John", "Mary", "Paul", "Jane")
x   #  John Mary Paul Jane
       10   20   30   40
```

- 向量元素命名後，可依名稱來呼叫

```
x["Mary"]   #  Mary
               20
x[c("Mary", "Paul")]   #  Mary Paul
                          20   30
```

Vector 向量 – 運算

- 向量運算時，R在兩個向量間執行元素對元素計算

```
x <- 1:6
y <- 1:6
x + y    # 2  4  6  8 10 12
# 1+1  2+2  3+3  4+4  5+5  6+6
```

Vector 向量－運算

- 向量運算時，R在兩個向量間執行元素對元素計算
- 當兩個向量元素個數不相等(length不等)時，較少元素個數的向量會依序重複(循環)使用元素

```
x <- 1:6
y <- 1:3
x + y    # 2 4 6 5 7 9
# 1+1  2+2  3+3  4+1  5+2  6+3
```

Vector 向量 – 運算

- 向量運算時，R在兩個向量間執行元素對元素計算
- 當兩個向量元素個數不相等(length不等)且無倍數關係時，R會出現警告訊息，但照樣執行循環使用元素

```
x <- 1:6
y <- 1:4
x + y    # 2 4 6 8 6 8
# 1+1  2+2  3+3  4+4  5+1  6+2
Warning message:
In x + y : 較長的物件長度並非較短物件長度的倍數
```

Vector 向量 – 運算

- 向量與純量間的運算，只是循環規則的運用而已 (複習：純量(scalar)為僅包含一個元素的向量)

```
x <- 1:6
x + 1   # 2 3 4 5 6 7
# 1+1  2+1  3+1  4+1  5+1  6+1
y <- 1
x + y
# 1+1  2+1  3+1  4+1  5+1  6+1
```

Vector 向量 – 邏輯比較

- 向量運算時，R在兩個向量間執行元素對元素計算

```
x <- 1:6   # 1 2 3 4 5 6
Y <- 6:1   # 6 5 4 3 2 1
x < y   # TRUE  TRUE  TRUE FALSE FALSE FALSE
x[x < y]   # 1 2 3  只留下序位為TRUE的元素
y[x < y]   # 6 5 4
```

Vector 向量 – 邏輯比較

- 向量運算時，R在兩個向量間執行元素對元素計算

```
x <- 11:17
x %% 2 == 0    # 除以2 餘數為0 => 偶數
# FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE
x[x %% 2 == 0]   # 12 14 16
# 只留下序位為TRUE的元素

x[ x > median(x) ]  # 留下值大於中位數的元素
```

邏輯比較運算符號

== ： 等於

!= ： 不等於

< ： 小於

> ： 大於

<= ： 小於等於

>= ： 大於等於

A %in% B ： A 是否在 B 中

&& ： 交集 (適用於單一值的邏輯判斷)

& ： 交集 (適用於向量式的邏輯判斷)

|| ： 聯集 (適用於單一值的邏輯判斷)

| ： 聯集 (適用於向量式的邏輯判斷)

Vector 向量 – 複習

```
x <- c(1,2,3,1,2)
names(x) <- c("A","B","C","D","E")
x[1]
x["A"]
x[1,4]   # Error : 維度數目不正確
x[c(1,4)]
x[c("A","D")]

x[x==1]   # 將值為 1 的元素集在一起成為一個向量
> x[x==1]
A D
1 1
```

僅供參考

Matrix (Matrices) 矩陣

- 具有維度(dimension, n = 2)的向量

```
x <- 1:15
dim(x)    # NULL
x         #  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
dim(x) <- c(3, 5)
x      # 元素先往下再往右排列
```

```
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    4    7   10   13
[2,]    2    5    8   11   14
[3,]    3    6    9   12   15
```

僅供參考

Array 陣列

- 具有維度(dimension, n >= 3)的向量

```
x <- 1:24
dim(x) <- c(2, 3, 4)
x
```

```
, , 1

     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6

, , 2

     [,1] [,2] [,3]
[1,]    7    9   11
[2,]    8   10   12

, , 3

     [,1] [,2] [,3]
[1,]   13   15   17
[2,]   14   16   18

, , 4

     [,1] [,2] [,3]
[1,]   19   21   23
[2,]   20   22   24
```

重要觀念

List 串列

- 一串未必相同形式元素的集合
- 元素可以是純量、向量、串列、資料框架
- 使用list函數建立

```
(x <- list(1:12))   # 含單一元素 [向量]
```

```
[[1]]
 [1]  1  2  3  4  5  6  7  8  9 10 11 12
```

```
(x <- list(1:12, c("A", "B", "C"), pi))   # 三個元素 向量 向量 純量
```

```
[[1]]
 [1]  1  2  3  4  5  6  7  8  9 10 11 12

[[2]]
[1] "A" "B" "C"

[[3]]
[1] 3.141593
```

List 串列 – 索引

- 串列可依位置編製索引

```
z <- list("John", "Mary", "Paul", "Jane")
z[[2]]   # "Mary" 表示串列z的第二個元素
```

```
[[1]]
[1] "Mary"
```

- 串列可依多個索引，傳回一個子串列 (sublist)

```
z[c(2, 3)]   #
```

```
[[1]]
[1] "Mary"

[[2]]
[1] "Paul"
```

List 串列 – 名稱

- 串列的個別元素可擁有名稱(name)

```
x <- list(10, 20, 30, 40)
names(x) <- c("John", "Mary", "Paul", "Jane")
x   #
```

```
$John
[1] 10

$Mary
[1] 20

$Paul
[1] 30

$Jane
[1] 40
```

- 串列元素命名後，可依名稱呼叫

```
x[["Mary"]]   # 20
x$Mary   # 20   此處名稱不需加雙引號
```

注意

List 串列－子串列

```
x <- list(10, 20, 30, 40)

x[[2]]   # 20 傳回一個元素
mode(x[[2]])   # "numeric"
```

```
[1] 20
```

```
x[c(2, 3)]   # 傳回一個(子)串列
mode(x[c(2, 3)])   # "list"
```

```
$Mary
[1] 20

$Paul
[1] 30
```

```
x[2]   # 傳回一個(子)串列
mode(x[2])   # "list"
```

```
$Mary
[1] 20
```

List 串列 – 命名

- 串列元素的命名，可在使用list函數建立串列時，以name=value方式為之，此時名字不需加雙引號

```
(x <- list(e1 = 1:12, e2 = c("A", "B", "C"), e3 = pi))
```

```
$e1
 [1]  1  2  3  4  5  6  7  8  9 10 11 12

$e2
[1] "A" "B" "C"

$e3
[1] 3.141593
```

重要觀念

Data Frame 資料框架

- 相當於MS Excel的Sheet或MS Access的Table
- 可視為一個排列整齊的串列(list)：每個元素為一個向量(vector)，每個向量長度相同

```
x <- 1:4
y <- c(11, 22, 33, 44)
z <- c("John", "Mary", "Paul", "Jane")
df <- data.frame(x, y, z)
df
colnames(df) <- c("XXX", "YYY", "ZZZ")    # 更(欄位)名
df
```

```
  x  y    z
1 1 11 John
2 2 22 Mary
3 3 33 Paul
4 4 44 Jane
```

```
  XXX YYY  ZZZ
1   1  11 John
2   2  22 Mary
3   3  33 Paul
4   4  44 Jane
```

Data Frame 資料框架 – 本質

- 可簡易視為一個排列整齊的串列：每個元素為一個向量，每個向量長度相同

```
x <- 1:4
y <- c(11, 22, 33, 44)
z <- c("John", "Mary", "Paul", "Jane")
df <- data.frame(x, y, z)
mode(df)   # "list"
class(df)   # "data.frame"
# data frame 外顯是個資料框架 內部實體為串列(list)
```

```
> mode(df)
[1] "list"
> class(df)
[1] "data.frame"
```

Data Frame 資料框架 – 索引

- 可依據索引位置或名稱來選取資料

```
df[[2]]   # 傳回第2直行(column)(單行)的資料 (=> 向量)
df[, 2]   # 同上                                          [1] 11 22 33 44
df[["y"]]   # 同上
df[, "y"]   # 同上
df$y   # 同上 (欄位名稱不需加引號) (此寫法較易記憶)
df[, c(2, 3)]   # 傳回第2至3直行的資料 (=>資料框架)

df[3, 2]   # 傳回第3橫列(row)第2直行(column)的元素
df[3, 2:3]   # 傳回第3橫列第2至3直行的元素 (=> 資料框架)
df[c(1, 3), 2]   # 傳回第1與3橫列第2直行的元素 (=> 向量)
df[2, ]   # 傳回第2橫列的元素 (=> 資料框架)
```

Data Frame 資料框架 – 名稱

- 可依據索引位置或名稱來選取資料

```
df[2]   # 傳回一個資料框架
df["y"]   # 傳回一個資料框架
```

```
  y
1 11
2 22
3 33
4 44
```

```
df[c(2, 3)]   # 傳回一個資料框架
df[, c(2, 3)]   # 傳回一個資料框架
df[c("y", "z")]   # 傳回一個資料框架
df[, c("y", "z")]   # 傳回一個資料框架
```

```
  y    z
1 11 John
2 22 Mary
3 33 Paul
4 44 Jane
```

Data Frame 資料框架 – 自動命名

- 也可先不用替元素命名，但是R會自動給個奇怪的名稱

```
df <- data.frame(1:4, c(11, 22, 33, 44), c("John", "Mary", "Paul", "Jane"))
df
```

```
  X1.4 c.11..22..33..44. c..John....Mary....Paul....Jane..
1    1                11                              John
2    2                22                              Mary
3    3                33                              Paul
4    4                44                              Jane
```

重要觀念

Factor 因子

- 用來將資料(類別變數)分組，方便統計運算

```
( x <- c("甲","乙","丙","丁","甲","甲","乙") )
class(x)
summary(x)
( y <- as.factor(x) )
mode(y)
class(y)
summary(y)
as.numeric(y)
# factor 在 R 內部係以整數形式儲存
factor(x, levels = c("甲","乙","丙","丁"))
```

```
> ( x <- c("甲", "乙", "丙", "丁", "甲", "甲", "乙") )
[1] "甲" "乙" "丙" "丁" "甲" "甲" "乙"
> class(x)
[1] "character"
> summary(x)
   Length     Class      Mode
        7 character character
> ( y <- as.factor(x) )
[1] 甲 乙 丙 丁 甲 甲 乙
Levels: 乙 丁 丙 甲
> mode(y)
[1] "numeric"
> class(y)
[1] "factor"
> summary(y)
乙 丁 丙 甲
 2  1  1  3
> as.numeric(y)
[1] 4 1 3 2 4 4 1
```

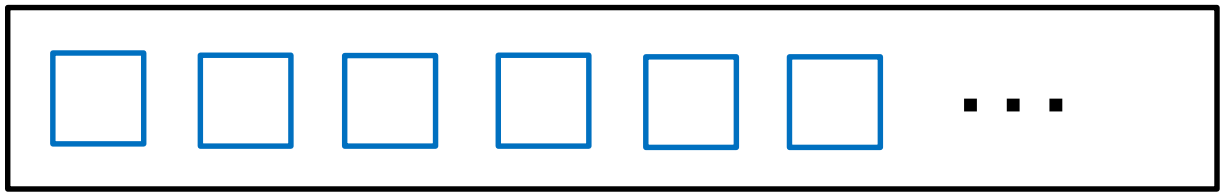
levels !!!

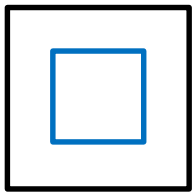
```
> factor(x, levels = c("甲", "乙", "丙", "丁"))
[1] 甲 乙 丙 丁 甲 甲 乙
Levels: 甲 乙 丙 丁
```

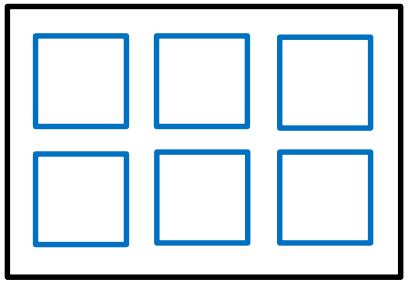
將 factor 排序 !!!

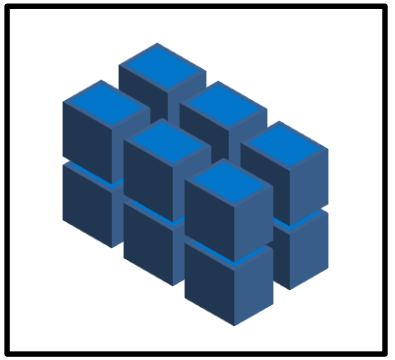
重要観念

複習-1

vector 

scalar  (= one-element vector)

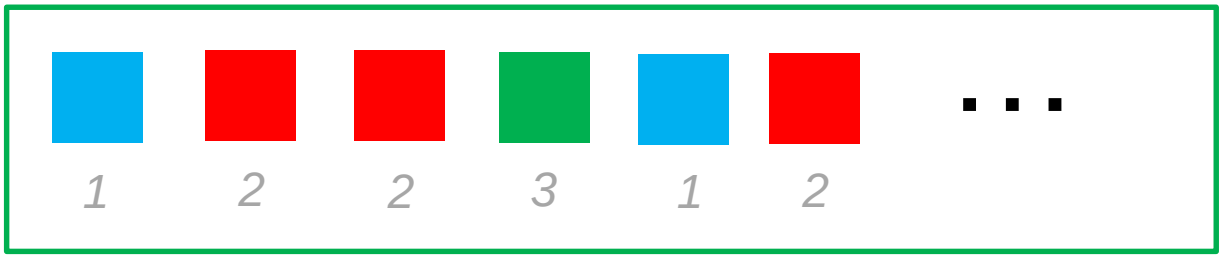
matrix (2D) 

array (3D …) 

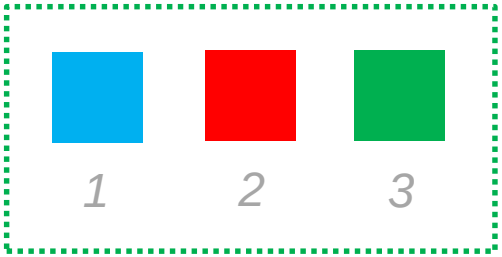
重要觀念

複習-2

factor



level



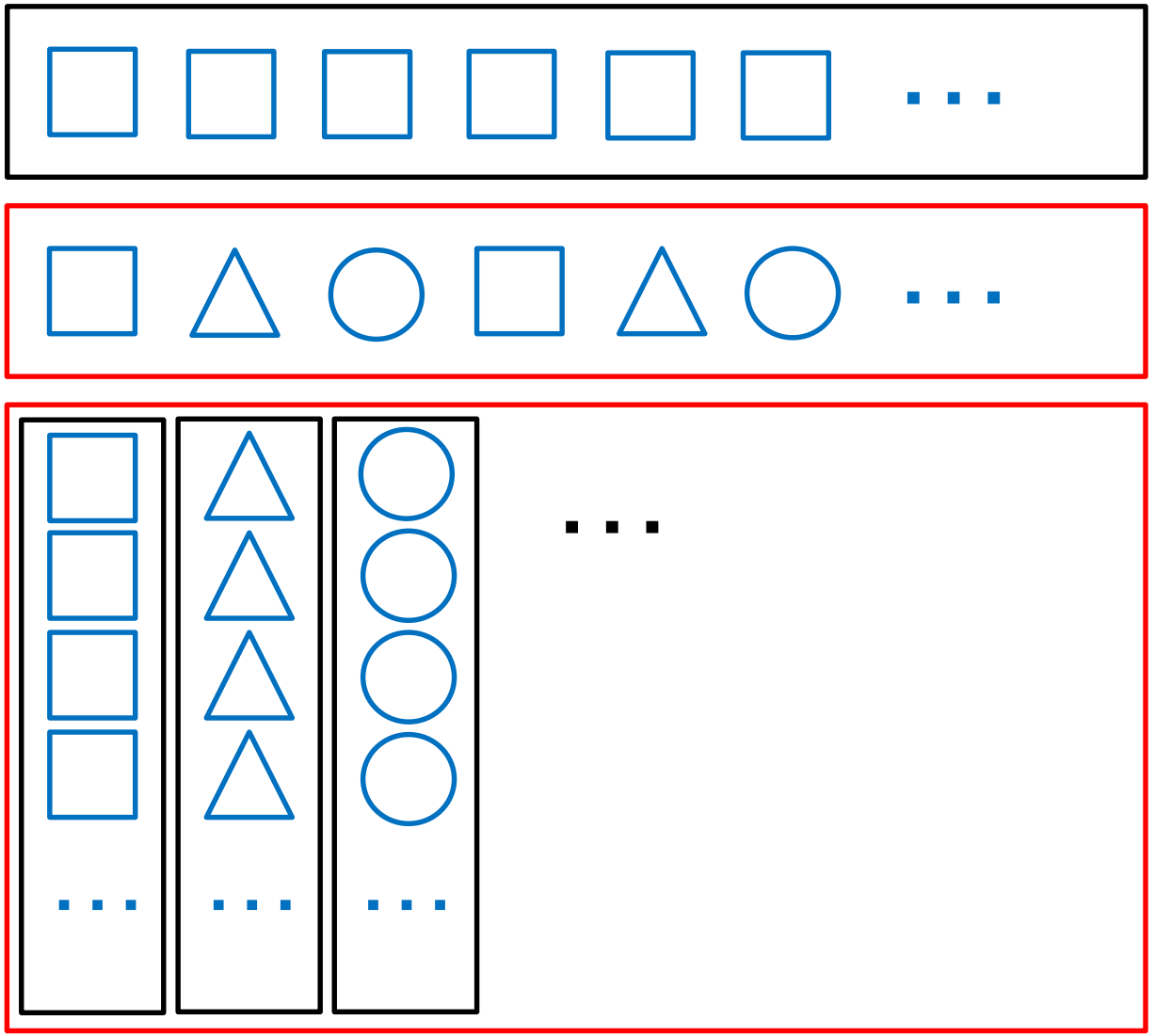
重要觀念

複習-3

vector

list

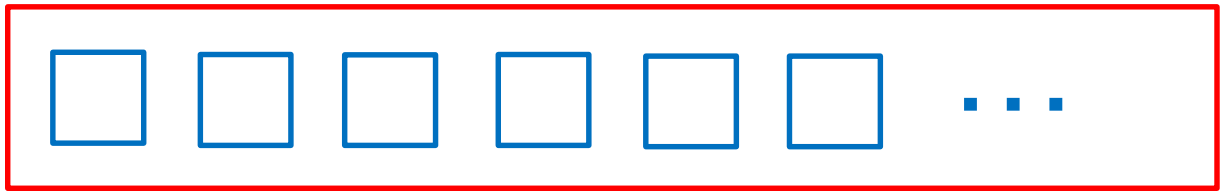
data frame



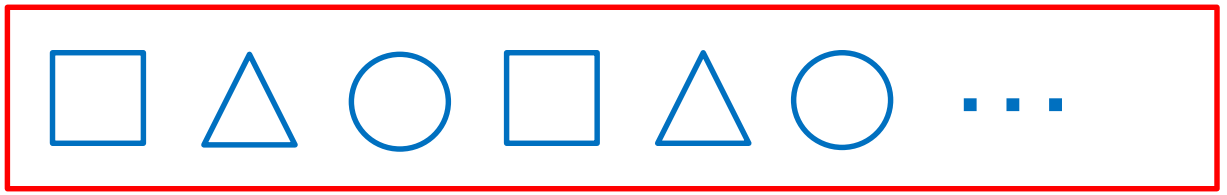
重要觀念

複習-4

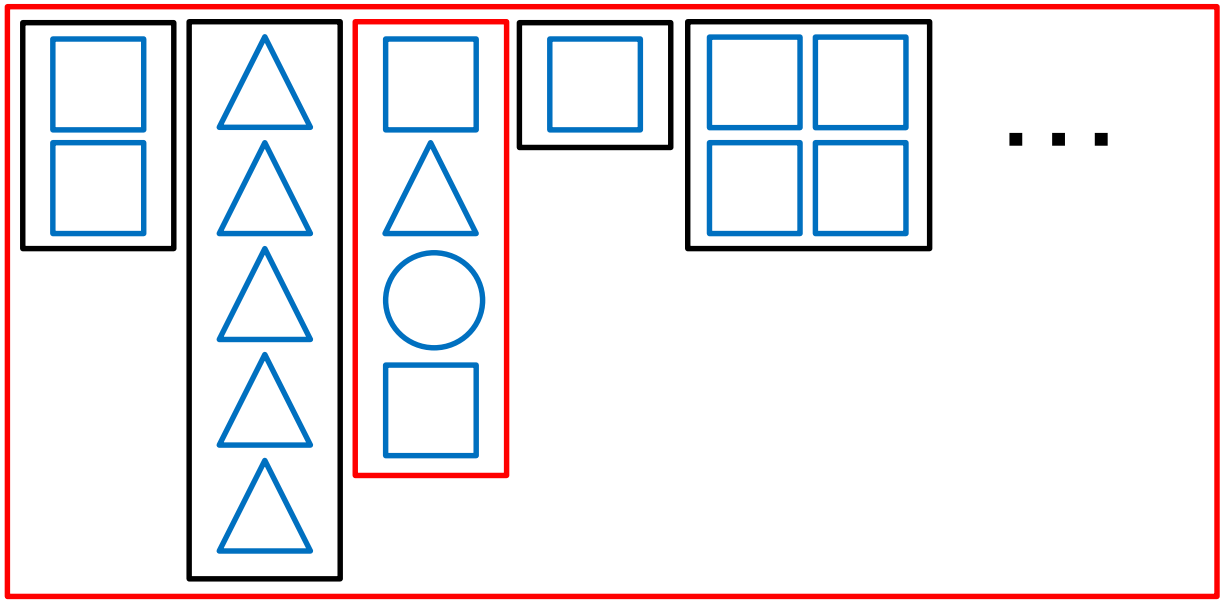
list



list



list



SECTION II

DATA IMPORT

CSV file

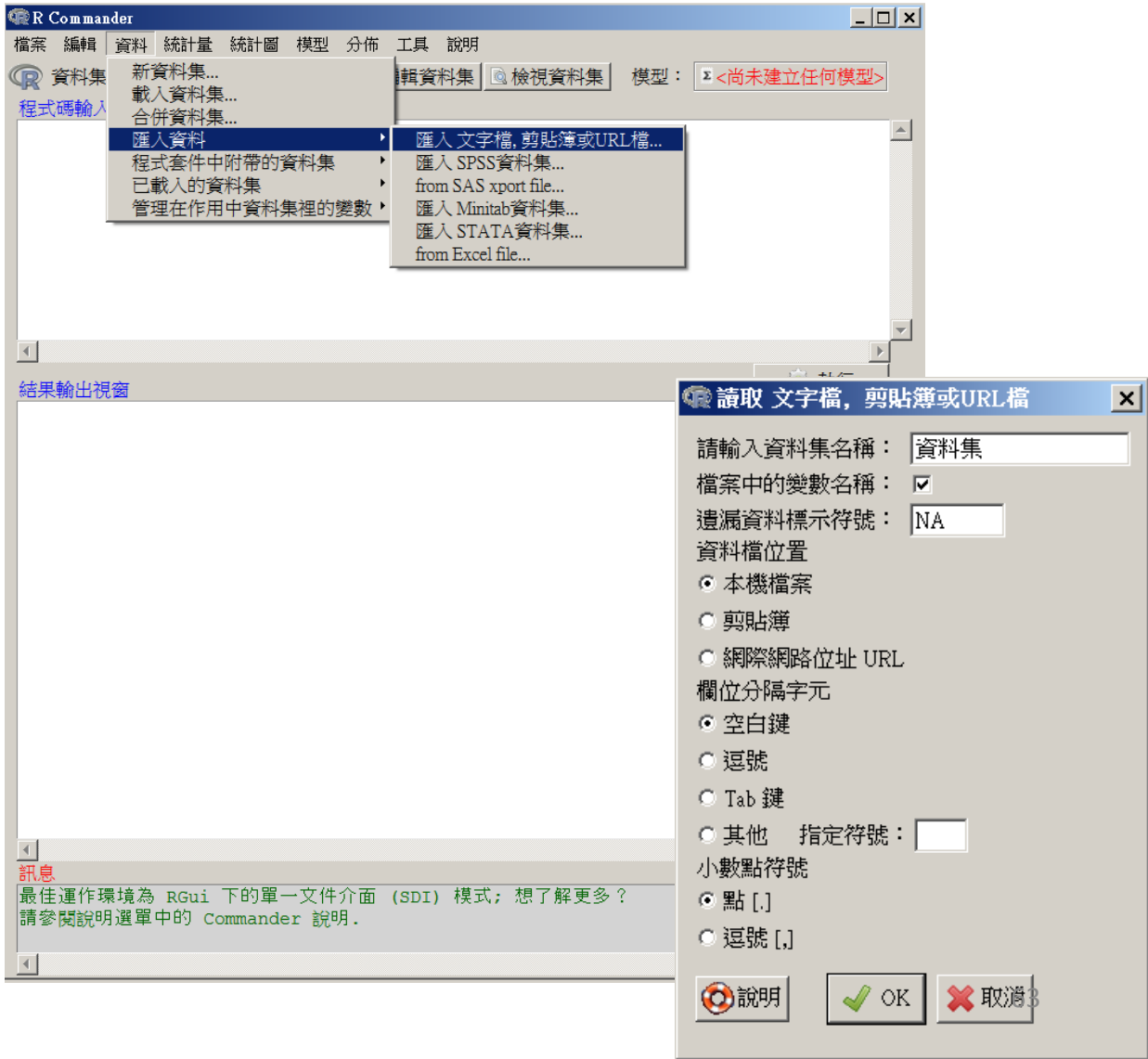
- x <- read.csv(file = "...", header = TRUE, colClasses = ..., ...)

- 匯入後以 data frame 型態儲存
- 如果第一列(row)並非欄位(column)名稱，則 header = FALSE，R會自動將欄位命名為V1, V2, V3, ... (V表示variable)
- 如果檔案非典型的CSV檔，可加入sep = "..."，告知分割欄位的符號
- 其他參數參見Reference (fullrefman.pdf) (=> I => The utils package => read.table)

僅供參考

With R Commander : Import data from text files

- 資料 ->
  匯入資料 ->
  匯入 文字檔, ...
- 設定
- 選取欲匯入檔案



Fixed-width file

- x <- read.fwf(file = "...", widths = c(...))

- 匯入後以 data frame 型態儲存
- 如果想略過前面幾(a)列，則加上參數skip = a
- 如果只想匯入幾(b)行，則加上參數n = b
- 如果想加快匯入速度，可加入參數 stringsAsFactors = FALSE，讓R先不要自行轉換
- 其他參數與 read.table 相同
- Reference (fullrefman.pdf) => I => The utils package => read.fwf

僅供參考

內建的資料集

data()

mtcars

head(mtcars)

str(mtcars)

str : structure

```
Data sets in package 'datasets':

AirPassengers              Monthly Airline Passenger Numbers 1949-1960
BJsales                    Sales Data with Leading Indicator
BJsales.lead (BJsales)     Sales Data with Leading Indicator
BOD                        Biochemical Oxygen Demand
CO2                        Carbon Dioxide Uptake in Grass Plants
ChickWeight                Weight versus age of chicks on different diets
DNase                      Elisa assay of DNase
EuStockMarkets             Daily Closing Prices of Major European Stock Indices, 1991-1998
Formaldehyde               Determination of Formaldehyde
HairEyeColor               Hair and Eye Color of Statistics Students
Harman23.cor               Harman Example 2.3
Harman74.cor               Harman Example 7.4
Indometh                   Pharmacokinetics of Indomethacin
InsectSprays               Effectiveness of Insect Sprays
JohnsonJohnson             Quarterly Earnings per Johnson & Johnson Share
LakeHuron                  Level of Lake Huron 1875-1972
LifeCycleSavings           Intercountry Life-Cycle Savings Data
Loblolly                   Growth of Loblolly pine trees
Nile                       Flow of the River Nile
Orange                     Growth of Orange Trees
OrchardSprays              Potency of Orchard Sprays
PlantGrowth                Results from an Experiment on Plant Growth
Puromycin                  Reaction Velocity of an Enzymatic Reaction
Seatbelts                  Road Casualties in Great Britain 1969-84
Theoph                     Pharmacokinetics of Theophylline
Titanic                    Survival of passengers on the Titanic
ToothGrowth                The Effect of Vitamin C on Tooth Growth in Guinea Pigs
UCBAdmissions              Student Admissions at UC Berkeley
UKDriverDeaths             Road Casualties in Great Britain 1969-84
UKgas                      UK Quarterly Gas Consumption
USAccDeaths                Accidental Deaths in the US 1973-1978
USArrests                  Violent Crime Rates by US State
USJudgeRatings             Lawyers' Ratings of State Judges in the US Superior Court
USPersonalExpenditure      Personal Expenditure Data
UScitiesD                  Distances Between European Cities and Between US Cities
VADeaths                   Death Rates in Virginia (1940)
WWWusage                   Internet Usage per Minute
WorldPhones                The World's Telephones
ability.cov                Ability and Intelligence Tests
airmiles                   Passenger Miles on Commercial US Airlines, 1937-1960
airquality                 New York Air Quality Measurements
anscombe                   Anscombe's Quartet of 'Identical' Simple Linear Regressions
```

暫勿執行

匯入示範資料

- 可利用 Import_BaseR.r，匯入資料
- 請將資料檔複製於D槽SampleData子目錄下，如果欲放置於其他位置，請同時更改Import_BaseR.r

資料及程式碼置於SampleData子目錄

注意 !!!
複製程式碼至R時，注意直的引號不要被MS Office改成斜的引號

暫勿執行

匯入示範資料

注意 !!!
複製程式碼至R時，注意直的引號不要被MS Office改成斜的引號

```
directory <- "D:/SampleData/"    # CHANGE HERE !!! 於此處指出檔案擺放的位置

cd <- read.fwf(
  file = paste(directory, "CD2009.DAT", sep = ""),
  strip.white = TRUE,    # strip leading and trailing white spaces from unquoted character fields
  widths = c(6, 1, 34, 8, 2, 6, 2, 2, 2, 2, 2, 8, 8, 8, 32, 4, 1, 3, 5, 5, 5, 4, 2, 1, 32, 32, 8, 8, 12, 8, 12,
8, 2, 8, 8, 8, 1),
  col.names = c("FEE_YM", "APPL_TYPE", "HOSP_ID", "APPL_DATE", "CASE_TYPE",
"SEQ_NO", "CURE_ITEM_NO1", "CURE_ITEM_NO2", "CURE_ITEM_NO3",
"CURE_ITEM_NO4", "FUNC_TYPE", "FUNC_DATE", "TREAT_END_DATE", "ID_BIRTHDAY",
"ID", "CARD_SEQ_NO", "GAVE_KIND", "PART_NO", "ACODE_ICD9_1", "ACODE_ICD9_2",
"ACODE_ICD9_3", "ICD_OP_CODE", "DRUG_DAY", "MED_TYPE", "PRSN_ID", "PHAR_ID",
"DRUG_AMT", "TREAT_AMT", "TREAT_CODE", "DIAG_AMT", "DSVC_NO", "DSVC_AMT",
"BY_PASS_CODE", "T_AMT", "PART_AMT", "T_APPL_AMT", "ID_SEX"),
  colClasses = c("character", "character", "character", "character", "character", "numeric",
"character", "character", "character", "character", "character", "character", "character", "character",
"character", "character", "character", "character", "character", "character", "character", "character",
"numeric", "character", "character", "character", "numeric", "numeric", "character", "numeric",
"character", "numeric", "character", "numeric", "numeric", "numeric", "character")
)

cd   # 查看是否匯入
```

以 Base R 的函數匯入大量資料

速度太慢了 !!!

改用 tidyverse 套件

精確地說是 readr 套件

匯入示範資料

- 利用 Import_tidyverse.r，匯入資料
- 將資料複製於D槽下，如果欲放置於其他位置，請同時更改Import_tidyverse.r
- 利用readr套件的 read_fwf 函數匯入
- 匯入速度相當快
- 不會將文字欄位轉換為因子(factor)

注意 !!!
以read_fwf函數匯入的資料集，如果資料集某處並無資料(空白)，R將其視為NA，而非長度為零的資料。與read.csv函數迥異

資料與程式碼置於SampleData子目錄

```
source("D:/SampleData/Import_tidyverse.r")
# CHANGE HERE !!! 更動.r檔的位置
```

```
if (! require(tidyverse)) { install.packages("tidyverse"); require(tidyverse) }

directory <- "D:/SampleData/"    # CHANGE HERE !!!

columnnames <- c("FEE_YM", "APPL_TYPE", "HOSP_ID", "APPL_DATE", "CASE_TYPE",
"SEQ_NO", "CURE_ITEM_NO1", "CURE_ITEM_NO2", "CURE_ITEM_NO3",
"CURE_ITEM_NO4", "FUNC_TYPE", "FUNC_DATE", "TREAT_END_DATE",
"ID_BIRTHDAY", "ID", "CARD_SEQ_NO", "GAVE_KIND", "PART_NO", "ACODE_ICD9_1",
"ACODE_ICD9_2", "ACODE_ICD9_3", "ICD_OP_CODE", "DRUG_DAY", "MED_TYPE",
"PRSN_ID", "PHAR_ID", "DRUG_AMT", "TREAT_AMT", "TREAT_CODE", "DIAG_AMT",
"DSVC_NO", "DSVC_AMT", "BY_PASS_CODE", "T_AMT", "PART_AMT", "T_APPL_AMT",
"ID_SEX")

columnwidths <- c(6, 1, 34, 8, 2, 6, 2, 2, 2, 2, 2, 8, 8, 8, 32, 4, 1, 3, 5, 5, 5, 4, 2, 1, 32, 32, 8, 8,
12, 8, 12, 8, 2, 8, 8, 8, 1)

columntypes <- "cccccncccccccccccccccccncccnncncncnnnc"

cd <- read_fwf(
  paste(directory, "CD2009.DAT", sep = ""),
  fwf_widths(columnwidths, columnnames),
  col_types = columntypes,
  progress = TRUE
)
```

自行再匯入OO2009.DAT 。另外CD2010.DAT與OO2010.DAT，請如法炮製。

```
> if (! require(tidyverse)) { install.packages("tidyverse"); require(tidyverse) }
>
> directory <- "D:/SampleData/"    # CHANGE HERE !!!
>
> columnnames <- c("FEE_YM", "APPL_TYPE", "HOSP_ID", "APPL_DATE", "CASE_TYPE", "SE
>
> columnwidths <- c(6, 1, 34, 8, 2, 6, 2, 2, 2, 2, 2, 8, 8, 8, 32, 4, 1, 3, 5, 5,
>
> columntypes <- "ccccncccccccccccccccccncccnncncncnnnc"
>
> cd <- read_fwf(
+   paste(directory, "CD2009.DAT", sep = ""),
+   fwf_widths(columnwidths, columnnames),
+   col_types = columntypes,
+   progress = TRUE
+ )
|=============================================================| 100% 171 MB
> cd
# A tibble: 598,574 x 37
   FEE_YM APPL_TYPE HOSP_ID          APPL_DATE CASE_TYPE SEQ_NO CURE_ITEM_NO1
   <chr>  <chr>     <chr>            <chr>     <chr>      <dbl> <chr>
 1 200901 1         00000000000000~ 20090204  01          3866 <NA>
 2 200901 1         00000000000000~ 20090212  01           247 <NA>
 3 200901 1         00000000000000~ 20090212  01           919 <NA>
 4 200901 1         00000000000000~ 20090212  01           752 <NA>
 5 200901 1         00000000000000~ 20090212  01           844 <NA>
 6 200901 1         00000000000000~ 20090212  01           852 <NA>
 7 200901 1         00000000000000~ 20090209  01          6542 <NA>
 8 200901 1         00000000000000~ 20090207  09           389 D0
 9 200901 1         00000000000000~ 20090207  09           369 <NA>
10 200901 1         00000000000000~ 20090211  08          3196 33
# ... with 598,564 more rows, and 30 more variables: CURE_ITEM_NO2 <chr>,
#   CURE_ITEM_NO3 <chr>, CURE_ITEM_NO4 <chr>, FUNC_TYPE <chr>,
#   FUNC_DATE <chr>, TREAT_END_DATE <chr>, ID_BIRTHDAY <chr>, ID <chr>,
#   CARD_SEQ_NO <chr>, GAVE_KIND <chr>, PART_NO <chr>, ACODE_ICD9_1 <chr>,
#   ACODE_ICD9_2 <chr>, ACODE_ICD9_3 <chr>, ICD_OP_CODE <chr>,
#   DRUG_DAY <dbl>, MED_TYPE <chr>, PRSN_ID <chr>, PHAR_ID <chr>,
#   DRUG_AMT <dbl>, TREAT_AMT <dbl>, TREAT_CODE <chr>, DIAG_AMT <dbl>,
#   DSVC_NO <chr>, DSVC_AMT <dbl>, BY_PASS_CODE <chr>, T_AMT <dbl>,
#   PART_AMT <dbl>, T_APPL_AMT <dbl>, ID_SEX <chr>
```

注意 !!!
以read_fwf函數匯入的資料集，如果資料集某處並無資料(空白)，R將其視為NA

匯入完畢後請測試

```
head(cd)   # 顯示前6列
tail(cd)   # 顯示最後6列
summary(cd)

summary(cd$ID_SEX)
table(cd$ID_SEX)
table(cd$CURE_ITEM_NO1)

summary(cd$T_AMT)
class(cd$TREAT_CODE)
summary(cd$TREAT_CODE)
```

重要觀念

匯入完畢後了解資料性質

```
mode(cd)   # "list"
class(cd)   # "spec_tbl_df" "tbl_df"     "tbl"        "data.frame"
# class(cd)   # "data.frame"  (如果係用Base R的 read.fwf 匯入)

mode(cd$FEE_YM)   # "character"
class(cd$FEE_YM)   # "character"  # 此欄位 either 向量 or 因子

is.vector(cd$FEE_YM)   # TRUE
is.factor(cd$FEE_YM)   # FALSE
is.list(cd$FEE_YM)   # FALSE
```

- 檔案匯入後為 data frame/tibble，每個欄位(直行)為 vector

插播

下載網路檔案示範 - 1

- 政府資料開放平臺 (https://data.gov.tw/)
- 醫療機構與人員基本資料 (https://data.gov.tw/dataset/15393)
- CSV 檔 (https://www.mohw.gov.tw/dl-61786-4bd02cd5-3d91-4f99-9556-363233aa8059.html)

醫療機構與人員基本資料

醫療機構之基本資料(含名稱、地址等)

主要欄位說明	機構代碼、機構名稱、權屬別、型態別、縣市鄉鎮、電話、地址、診療科別、醫師、中醫師、牙醫師、藥師、藥劑生、護理師、護士、助產士、助產師、醫事檢驗師、醫事檢驗生、物理治療師、職能治療師、醫事放射師、醫事放射士、物理治療生、職能治療生、呼吸治療師、諮商心理師、臨床心理師、營養師、語言治療師、牙體技術師、聽力師、牙體技術生、驗光師、驗光生
資料資源下載網址	CSV下載按鈕 CSV 檢視資料 醫療機構與人員基本資料
提供機關	衛生福利部醫事司

Accessed Apr. 9, 2021

插播

下載網路檔案示範 - 2

```
f <- read.csv(file = "https://www.mohw.gov.tw/dl-61786-4bd02cd5-3d91-4f99-9556-363233aa8059.html")

head(f)

View(f)
```

```
> head(f)
   機構代碼       機構名稱       權屬別       型態別      縣市鄉鎮
1 3846011179   道民中醫診所 私立中醫診所 中醫一般診所 臺東縣台東市
2 3846011197   大春中醫診所 私立中醫診所 中醫一般診所 臺東縣台東市
3 3846011222 陳靜婉中醫診所 私立中醫診所 中醫一般診所 臺東縣台東市
4 3846011231   景春中醫診所 私立中醫診所   中醫診所 臺東縣台東市
5 3846011268   普賢中醫診所 私立中醫診所   中醫診所 臺東縣台東市
6 3846021022   宏霖中醫診所 私立中醫診所 中醫一般診所 臺東縣成功鎮
       電話                  地址 診療科別 醫師 中醫師 牙醫師 藥師
1 089/323991   臺東縣台東市開封街449號             0       1       0     0
2   89355125 臺東縣台東市中華路二段一七六號             0       2       0     0
3 089-331812   臺東縣台東市中正路271號             0       1       0     0
4 089-359629 臺東縣台東市中華路一段277號1樓             0       2       0     0
5 089-516885   臺東縣台東市知本路三段583號             0       1       0     0
6 089-852553   臺東縣成功鎮三民路三二號             0       1       0     0
  藥劑生 護理師 護士 助產士 助產師 醫事檢驗師 醫事檢驗生 物理治療師
1      0       0    0       0       0          0          0          0
2      0       0    1       0       0          0          0          0
3      0       0    0       0       0          0          0          0
4      0       0    0       0       0          0          0          0
5      0       0    0       0       0          0          0          0
6      0       0    0       0       0          0          0          0
  職能治療師 醫事放射師 醫事放射士 物理治療生 職能治療生 呼吸治療師
1          0          0          0          0          0          0
2          0          0          0          0          0          0
3          0          0          0          0          0          0
4          0          0          0          0          0          0
5          0          0          0          0          0          0
6          0          0          0          0          0          0
  諮商心理師 臨床心理師 營養師 語言治療師 牙體技術師 聽力師 牙體技術生
1          0          0       0          0          0       0          0
2          0          0       0          0          0       0          0
3          0          0       0          0          0       0          0
4          0          0       0          0          0       0          0
5          0          0       0          0          0       0          0
6          0          0       0          0          0       0          0
  驗光師 驗光生
1      0      0
2      0      0
3      0      0
4      0      0
5      0      0
6      0      0
>
```

Accessed Apr. 9, 2021

插播

下載網路檔案示範 - 3

- 有時候，擬下載檔案的編碼格式為 UTF-8-BOM，而無法成功匯入 R。例如：
- 健康保險資料開放服務 (https://data.nhi.gov.tw/)
- 急診轉住院暫留急診超過四十八小時案件比率(醫院總額指標) (https://data.nhi.gov.tw/Datasets/DatasetDetail.aspx?id=267&Mid=A110100)
- CSV 檔 (https://data.nhi.gov.tw/Datasets/DatasetResource.ashx?rId=A21030000I-E3000Y-001)

Accessed Apr. 9, 2021

插播

下載網路檔案示範 - 4

```
f <- read.csv(file =
"https://data.nhi.gov.tw/Datasets/DatasetResource.ashx?rId=A21030000I-E3000Y-001", fileEncoding = "UTF-8-BOM")

head(f)
View(f)
# 如果檔案編碼格式為 BIG5 而無法成功匯入 R，可試加上 fileEncoding = "BIG5" 參數
```

Accessed Apr. 9, 2021

SECTION III

SOME SIMPLE DATA PROCESSING

Descriptive/Summary Stats Functions

```
mean(cd$T_AMT)
median(cd$T_AMT)
sd(cd$T_AMT)
var(cd$T_AMT)
sum(cd$T_AMT)
min(cd$T_AMT)
max(cd$T_AMT)
range(cd$T_AMT)
summary(cd$T_AMT)
```

重要觀念 !!!

絕大多數R的程式
係利用函數(function)運作

函數多為內建(核心函數)
或他建(附屬於外掛套件)
也可自建(自行建立)

函數運用須注意參數設定

```
> summary(cd$T_AMT)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0   320.0   425.0   934.3   776.0 576870.0
```

資料集裡有多少筆紀錄

```
nrow(cd)

length(cd$HOSP_ID)

sum(complete.cases(cd))
# complete.cases : Return a logical vector indicating which cases are complete, i.e., have no missing values

NROW(na.omit(cd))
# na.omit : returns the object with incomplete cases removed
```

某欄位裡有多少相異的資料

```
unique(cd$ID)
length(unique(cd$ID))

View(unique(cd$ID))
# 在RGui裡View需大寫
# 若載入tidyverse套件，則也可寫成view (小寫)，但只顯示前1000筆(row)資料。當資料輸出量大時，請勿任意執行 view()，否則會造成 RStudio 近乎當掉。
```

挑出符合某些條件的資料

```
subset(cd, FUNC_DATE == "20090101")

edit(subset(cd, FUNC_DATE == "20090101"))

cd[ which(cd$FUNC_DATE == "20090101"), ]

subset(cd, FUNC_DATE == "20090101" & ID_SEX
== "M")
```

每個月各有幾筆紀錄

```
table(cd$FEE_YM)

View(table(cd$FEE_YM))


table(cd$FEE_YM, cd$ID_SEX)

View(table(cd$FEE_YM, cd$ID_SEX))
```

僅供參考

每個月各有幾筆紀錄(grid顯示)

```
if (! require(gridExtra)) { install.packages("gridExtra") }
library(gridExtra)
d <- table(cd$FEE_YM, cd$ID_SEX)
grid.table(d)

d <- table(substring(cd$FEE_YM, 5, 6), cd$ID_SEX)
grid.table(d)
```

substring 也可寫為 substr

Recode - 1 (有很多種作法)

```
head(cd$ID_SEX, 30)
head( ifelse(cd$ID_SEX == "M", "1",
ifelse(cd$ID_SEX == "F", "2", "U")), 30 )
```

Nested if-else

```
GENDER <- ifelse(cd$ID_SEX == "M", "1",
ifelse(cd$ID_SEX == "F", "2", "U"))
head(GENDER, 30)
```

* 相當於 Excel 的 IF 函數

Recode - 2

```
GENDER <- cd$ID_SEX
head(GENDER, 15)

GENDER[ 1 : length(GENDER) ] <- "U"
head(GENDER, 15)
GENDER[ cd$ID_SEX == "M" ] <- "1"
head(GENDER, 15)
GENDER[ cd$ID_SEX == "F" ] <- "2"
head(GENDER, 15)
```

先將所有元素的值設為最後的選項

一一對相對應位置內符合條件者賦值

- 先複製ID_SEX欄位至新欄位，以免變更原始資料

Recode - 3

- 利用 car 套件裡的 recode 函數

```
if (! require(car)) { install.packages("car"); require(car) }
# 也可寫成
# install.packages("car")  # install the car package
# library(car)  # load the car package

head( recode( cd$ID_SEX, " 'M' = '1'; 'F' = '2'; else = 'U' " ), 15 )

head( cd$ID_SEX, 15 )   # 原始資料沒變
```

Recode (range) - 1

```
head(
  ifelse( cd$T_AMT <= 500,
    "normal",
    ifelse( cd$T_AMT <= 1000,
      "high",
      "extreme"
    )
  ), 15
)
head( cd$T_AMT, 15 )   # 比較一下
```

Recode (range) - 2

```
FEE <- cd$T_AMT
head(FEE, 15)

FEE[ cd$T_AMT <= 500 ] <- "normal"
head(FEE, 15)
FEE[ cd$T_AMT > 500 & cd$T_AMT <= 1000 ] <- "high"
head(FEE, 15)
FEE[ cd$T_AMT > 1000 ] <- "extreme"
head(FEE, 15)
```

- 先複製T_AMT欄位至新欄位，以免變更原始資料

Recode (range) - 3

- 利用 car 套件裡的 recode 函數

```
library(car)

head( recode(cd$T_AMT, " 0:500 = 'normal'; 501:1000 = 'high'; else = 'extreme' "), 15)

head( cd$T_AMT, 15 )   # 原始資料沒變
```

Recode (range) - 4

- 利用 base 套件裡的 cut 函數，新增一個factor變數

0 ~ <= 500; > 500 ~ <= 1000; > 1000 ~ Inf

```
FEE <- cut( cd$T_AMT,
            breaks = c(0, 500, 1000, Inf),
            labels = c("normal", "high", "extreme") )
head(FEE, 15)

head(cd$T_AMT, 15)   # 比較一下

class(FEE)   #  "factor" !!!
```

SECTION IV

R STATISTICS

不想每次拷貝貼上相同程式碼

```
source("D:/SampleData/Import_tidyverse.r")
# CHANGE HERE !!! 更動.r檔的位置
```

- 將寫好的程式碼儲存在指令稿（.r 檔）檔案中，再呼叫進來利用
- 相當於在目前的程式碼，貼上已有的程式碼
- 可將常用的自建函數儲存在.r 檔，呼叫使用之，可算是簡易版的套件(package)
- 預設是適用目前程式碼的工作環境，可更改之
- 缺點：不會顯示執行過程，難以判斷是否有誤

Confidence Interval for a Mean

```
mean(cd$T_AMT)

t.test(cd$T_AMT)
```

- 假設已先利用Import_tidyverse.r匯入資料
- 計算平均醫療費用的95%信賴區間
- 最好是資料筆數大於30
- Reference (fullrefman.pdf) => I => The stats package => t.test

Student's t-Test

```
x <- subset(cd, ID_SEX == "M" | ID_SEX == "F")
t.test(x$T_AMT ~ x$ID_SEX)
t.test(T_AMT ~ ID_SEX, data = x)   # 另一種寫法
```

- 只能比較兩組，先排除ID_SEX為U者
- 假設已先利用Import_tidyverse.r匯入資料
- 檢驗男女兩組群的醫療費用是否有差異
- 嚴謹的作法，需先檢測兩組的變異數
- Reference (fullrefman.pdf) => I => The stats package => t.test

Correlation

substr 也可寫為 substring，start與stop可省略

```
Age <- 2009 - as.numeric(substr(cd$ID_BIRTHDAY, start = 1, stop = 4))
cor.test(Age, cd$T_AMT)
cor.test(Age, cd$T_AMT, method = "spearman")
```

- 假設已先利用Import_tidyverse.r匯入資料
- 檢測每次就診時病人年齡與醫療費用的相關性
- Reference (fullrefman.pdf) => I => The graphics package => cor.test

SECTION V

REAL EXAMPLES

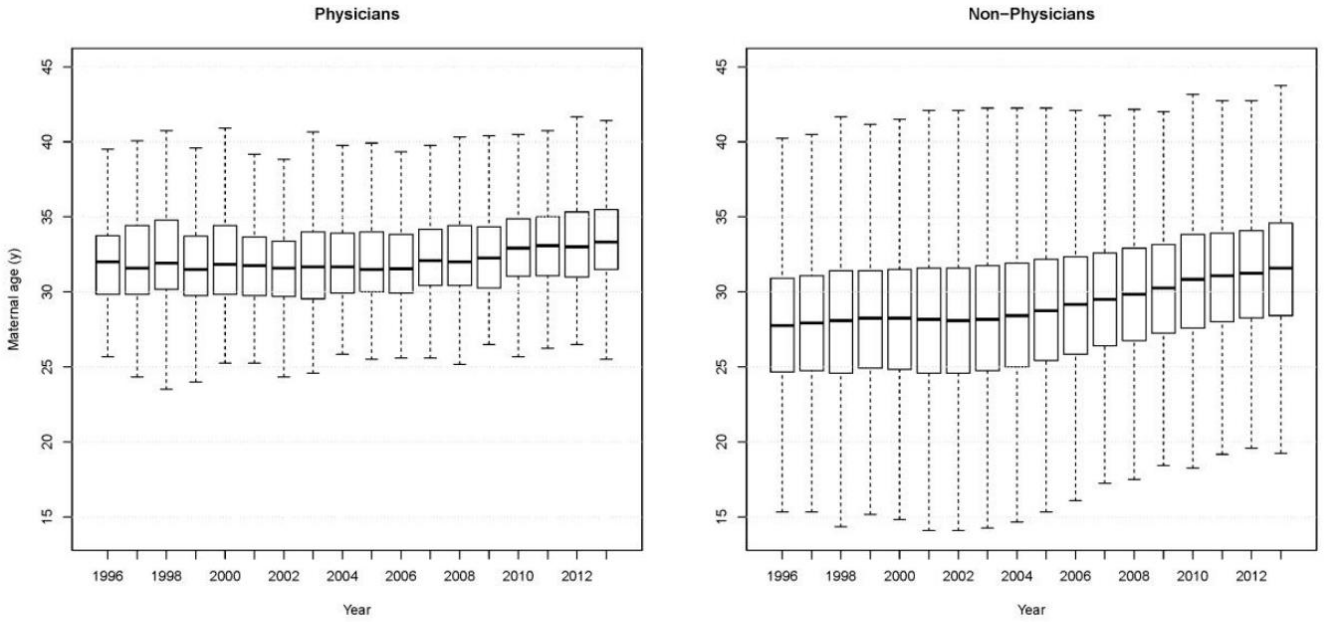


Figure 1. Distribution of maternal age at delivery among female physicians and non-physicians, 1996–2013.

SECTION V-A

BIRTH TRENDS AMONG FEMALE PHYSICIANS IN TAIWAN

女醫師與非醫師生產年齡差異

- t-test
- 發表於 Int J Environ Res Public Health 2017;14:746 (doi: 10.3390/ijerph14070746) [PMID: 28698490]
- # 資料與程式碼置於WomanDocPregnancy子目錄
- 女醫師 : 4,940 筆記錄 / 非醫師 : 4,063,023 筆記錄

Table 2. Patients, visits and disease patterns for different age groups of family physicians.

	Age Group			Kruskal–Wallis Test	Wilcoxon–Mann–Whitney Test
	<50	50–64	≥65	*p*-Value	*p*-Value
Patients, n					
Mean	4058	3855	2330	0.006	0.012 (<50 vs. ≥65)
SD	3073	2474	2019		<0.001 (50–64 vs. ≥65)
Visits per year, n					
Mean	11,019	13,371	9220	0.021	0.289 (<50 vs. ≥65)
SD	8837	8992	8600		0.006 (50–64 vs. ≥65)
Proportion of acute illness visits					
Mean	0.369	0.347	0.314	0.532	
SD	0.228	0.215	0.199		
Proportion of chronic illness visits					
Mean	0.152	0.154	0.11	0.429	
SD	0.163	0.17	0.122		

SECTION V-B

PATTERN OF VISITS TO OLDER FAMILY PHYSICIANS IN TAIWAN

老家庭醫師的病人是否比較老

- 先 Kruskal-Wallis test 三組醫師一起比較
- 再 Wilcoxon-Mann-Whitney test 兩兩比較

- 發表於 Int J Environ Res Public Health 2017;14:499 (doi: 10.3390/ijerph14050499) [PMID: 28481318]

- # 資料與程式碼置於ElderlyFamilyPhysician子目錄
- 共 2,018,440 筆記錄，一筆記錄：
  (一位)醫師代號 – 醫師年齡層 –(一位)病人年齡

僅供參考

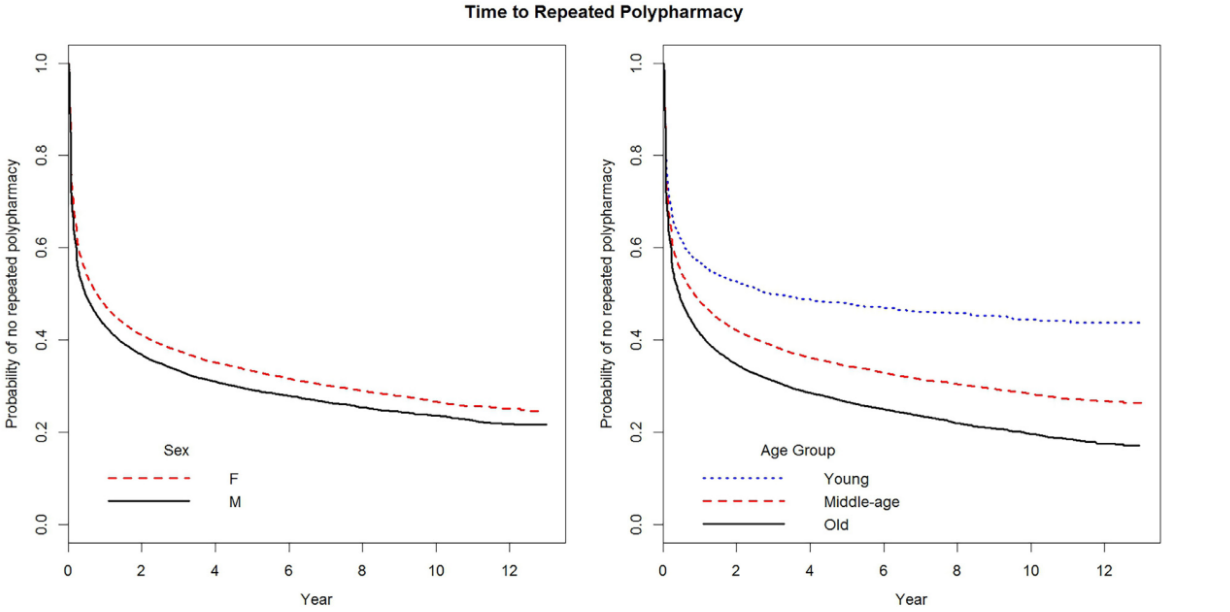


FIGURE 1 | Probabilities of having no repeated prescription of excessive polypharmacy among patients, stratified by sex or age groups (both $p < 0.001$ by log-rank test), 2001–2013.

SECTION V-C

IS EXCESSIVE POLYPHARMACY A TRANSIENT OR PERSISTENT PHENOMENON?

僅供參考

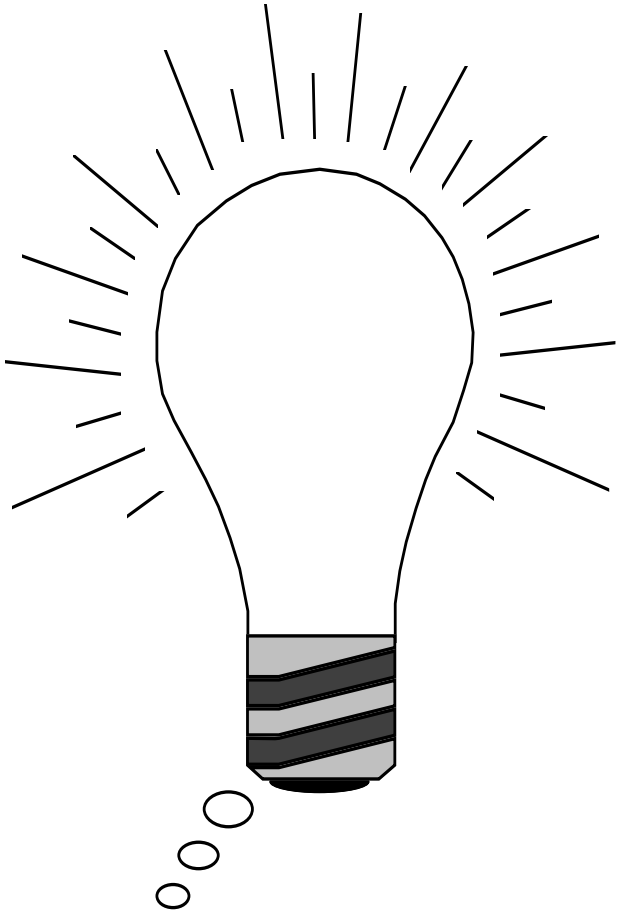
多重用藥是否持續出現

- Survival analysis – Kaplan-Meier estimate
- Log-rank test

- 發表於 Front Pharmacol 2018;9:120 (doi: 10.3389/fphar.2018.00120) [PMID: 29515446]

- # 資料與程式碼置於Deprescribing子目錄

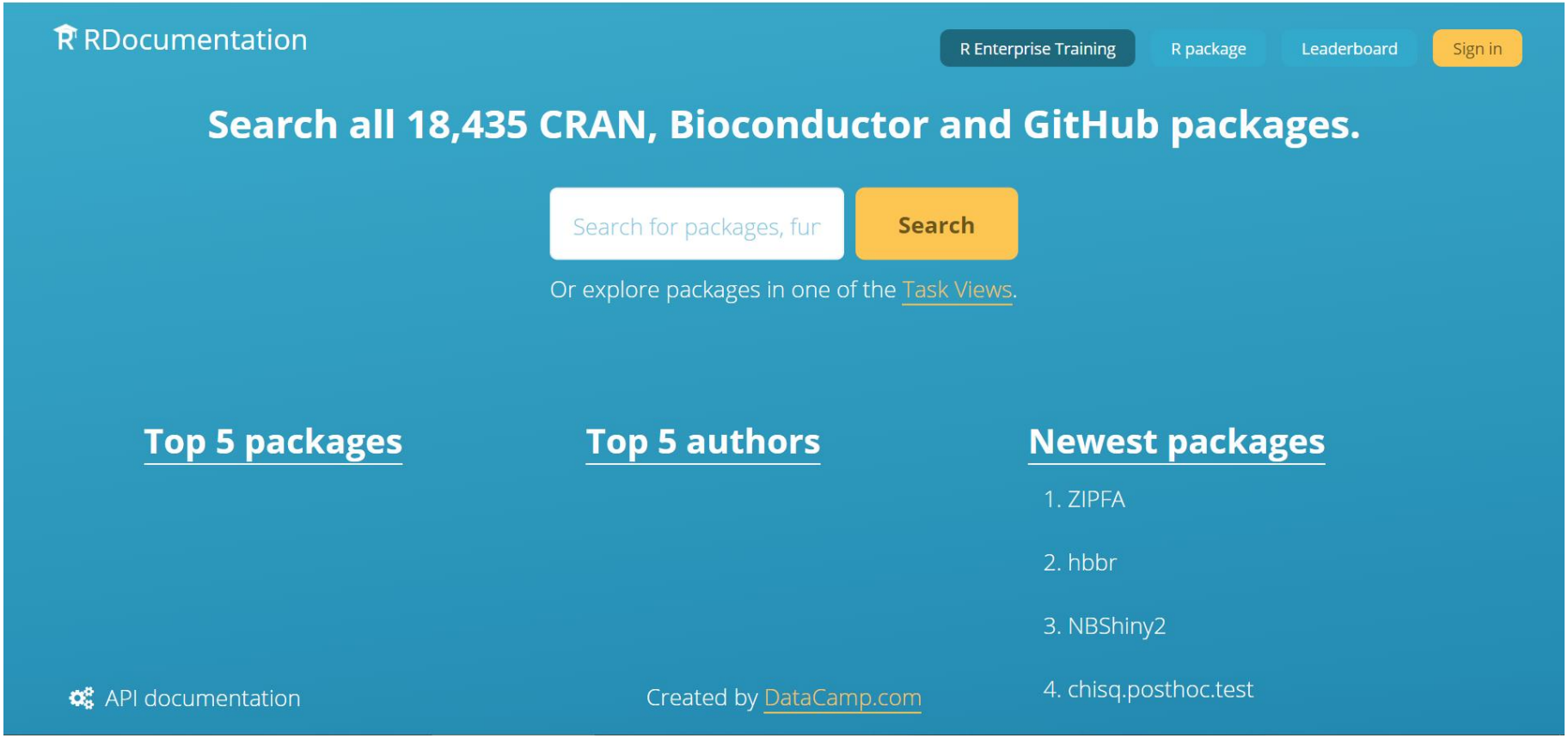
Thanks for Your Attention !



SECTION VI

VALUABLE RESOURCES

https://www.rdocumentation.org/



https://cran.r-project.org/other-docs.html

Contributed Documentation

Note: The CRAN area for contributed documentation is frozen and no longer actively maintained.

English --- Other Languages

Manuals, tutorials, etc. provided by users of R. The R core team does not take any responsibility for contents, but we appreciate the effort very much and encourage everybody to contribute to this list! To submit, follow the submission instructions on the CRAN main page. All material below is available directly from CRAN, you may also want to look at the list of other R documentation available on the Internet.

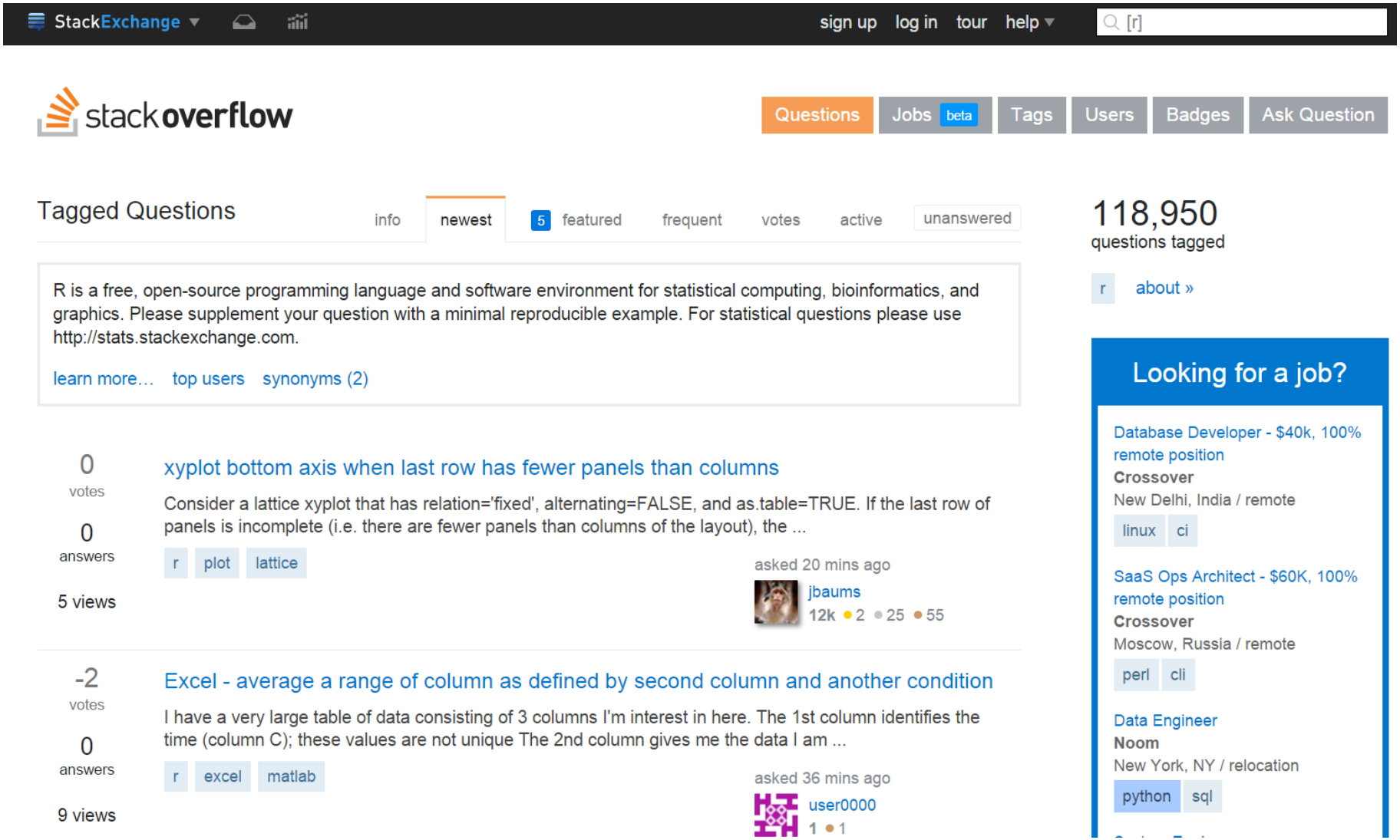
Note: Please use the directory listing to sort by name, size or date (e.g., to see which documents have been updated latetly).

English Documents

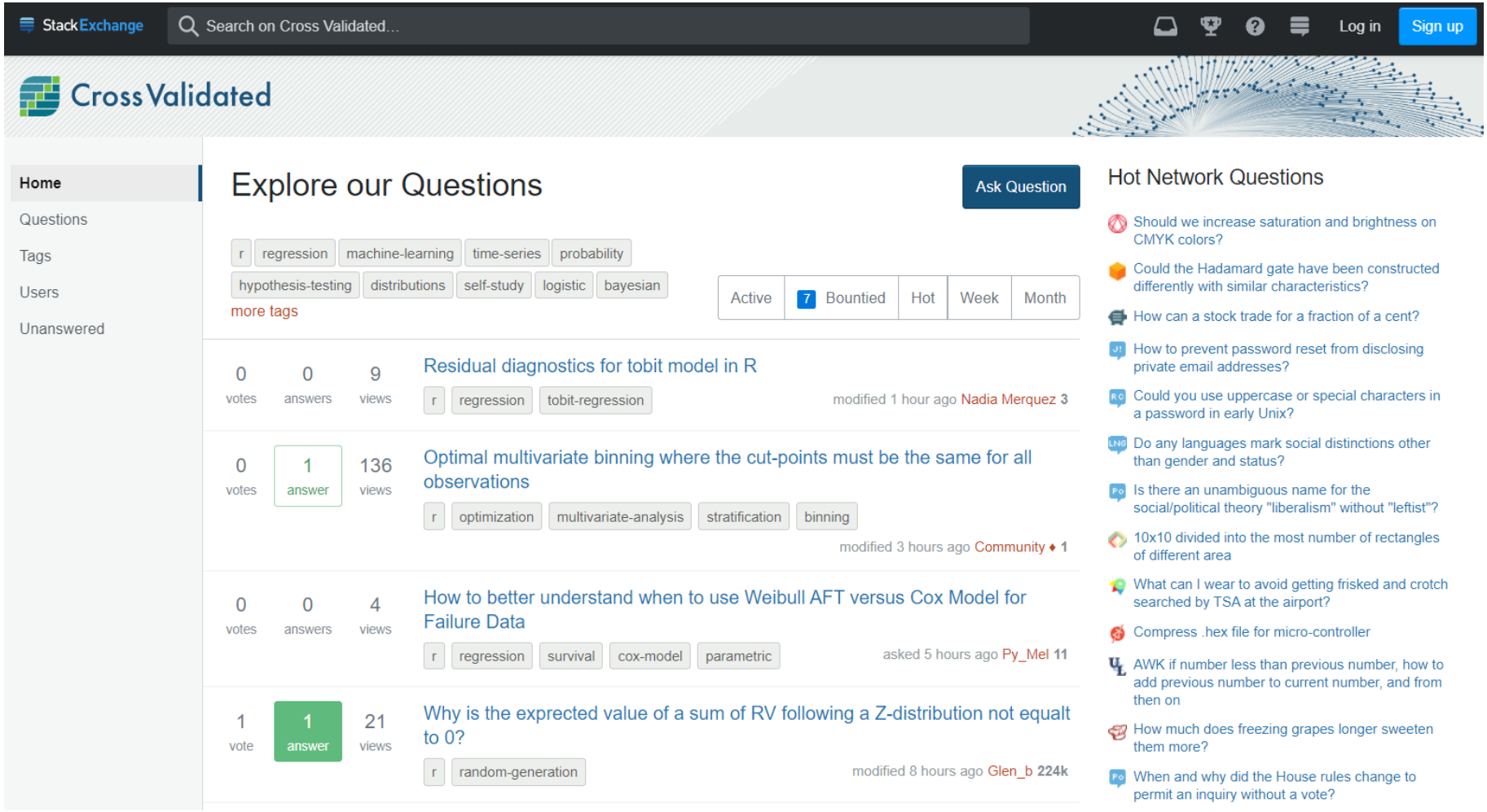
Documents with more than 100 pages:

- **"Visual Statistics. Use R!"** by Alexey Shipunov (PDF, 2019-03-17, 429 pages). All book materials are accessible from Alexey Shipunov's English R page .
- **"Using R for Data Analysis and Graphics - Introduction, Examples and Commentary"** by John Maindonald (PDF, data sets and scripts are available at JM's homepage).
- **"Practical Regression and Anova using R"** by Julian Faraway (PDF, data sets and scripts are available at the book homepage).
- The **Web Appendix** to the book "An R Companion to Applied Regression" (second edition) by John Fox and Sanford Weisberg contains information about R to fit a variety of regression models.
- **"An Introduction to S and the Hmisc and Design Libraries"** by Carlos Alzola and Frank E. Harrell, especially of interest to SAS users, users of the Hmisc or Design packages, or R users interested in data manipulation, recoding, etc. (PDF)
- **"Statistical Computing and Graphics Course Notes"** by Frank E. Harrell, includes material on S, LaTeX, reproducible research, making good

https://stackoverflow.com/questions/tagged/r



https://stats.stackexchange.com/?tags=r
https://stats.stackexchange.com/questions/tagged/r



https://www.r-bloggers.com/

https://joe11051105.gitbooks.io/r_basic/

Introduction

本書是根據本人學習 R 的經驗，加上整理歸納後所產出的筆記，有任何錯問題歡迎大家糾正與指教。

本書涵蓋內容

1. 環境建置
2. 基本運算
3. 變數與資料
4. 資料匯入與輸出
5. 流程控制
6. 資料整理
7. 自訂函數
8. 繪圖功能
9. 文件撰寫

* PDF檔 ： https://legacy.gitbook.com/download/pdf/book/joe11051105/r_basic

http://www.hmwu.idv.tw/index.php/r-software

Han-Ming Wu 吳漢銘

http://www.hmwu.idv.tw

國立臺北大學 統計學系

Department of Statistics, National Taipei University

Home | About Me | Photo Gallery | Facebook | Links | Contact Me | NTPU-108(上)課程 | 【作業考

Home ▸ R Software (R統計軟體教學)

R Software (R軟體教學/R語言程式設計)

點擊數：197008

R統計軟體教學/學習講義 (R Software Learning Materials)

https://yijutseng.github.io/DataScienceRBook/

資料科學與R語言

曾意儒 Yi-Ju Tseng

2020-07-20

本書介紹如何使用R語言完成資料讀取 (檔案、透過API擷取或爬蟲)、資料清洗與處理、探索式資料分析、資料視覺化、互動式資料呈現 (搭配Shiny) 與資料探勘等，並介紹R與Hadoop Ecosystems介接方法。

資料探勘章節尚未完成，epub版本格式微調中。

如要一次安裝所有本書會使用到的套件，可在R內執行以下程式碼：

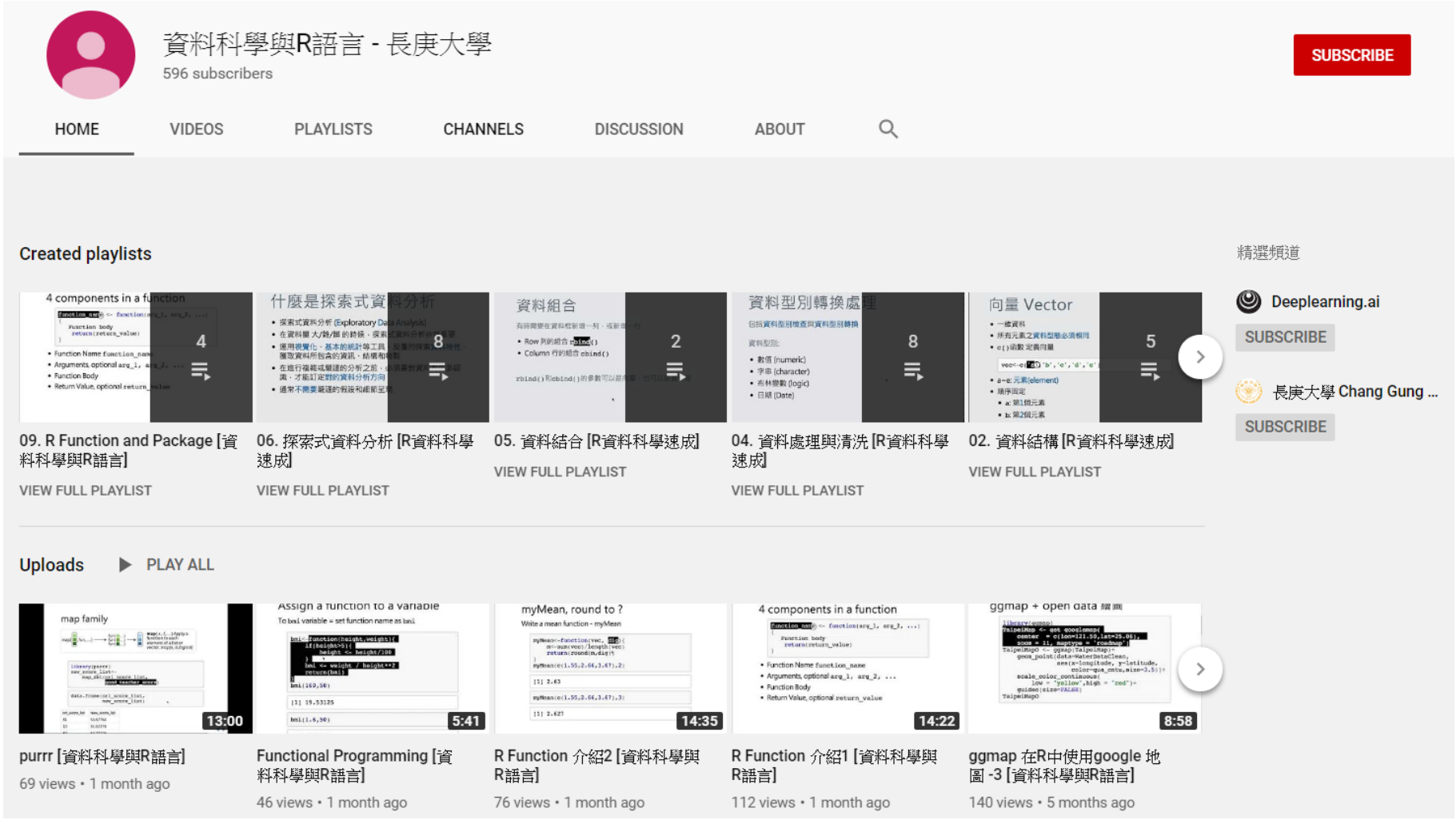
```
install.packages("devtools")
devtools::install_github("yijutseng/DataAnalyticsWithRBook")
```

本書為長庚大學資訊管理學系 大數據分析方法課程教學使用書籍，並可搭配YouTube平台上的教學影片參考使用，**影片閱讀清單**詳見本書最末章節Ch 13 教學影片資訊。

如果您想修改文字或範例，歡迎透過此連結或是透過GitHub issue提供建議與回饋。

本書程式碼執行環境：

https://www.youtube.com/c/資料科學與R語言/videos



https://mgimond.github.io/ES218/index.html

Data Manipulation | Plots | Univariate | Bivariate | Two-way tables | Misc | Datasets

Exploratory Data Analysis in R

This repo houses lecture notes used in an *Exploratory Data Analysis in R* course taught to undergraduates at Colby College. The course assumes little to no background in quantitative analysis nor in computer programming and was first taught in Spring, 2015. The course introduces students to data manipulation in R, data exploration (in the spirit of John Tukey's EDA) and the R markdown language. Many of the visualization tools are adopted from William Cleveland's *Data Visualization* book.

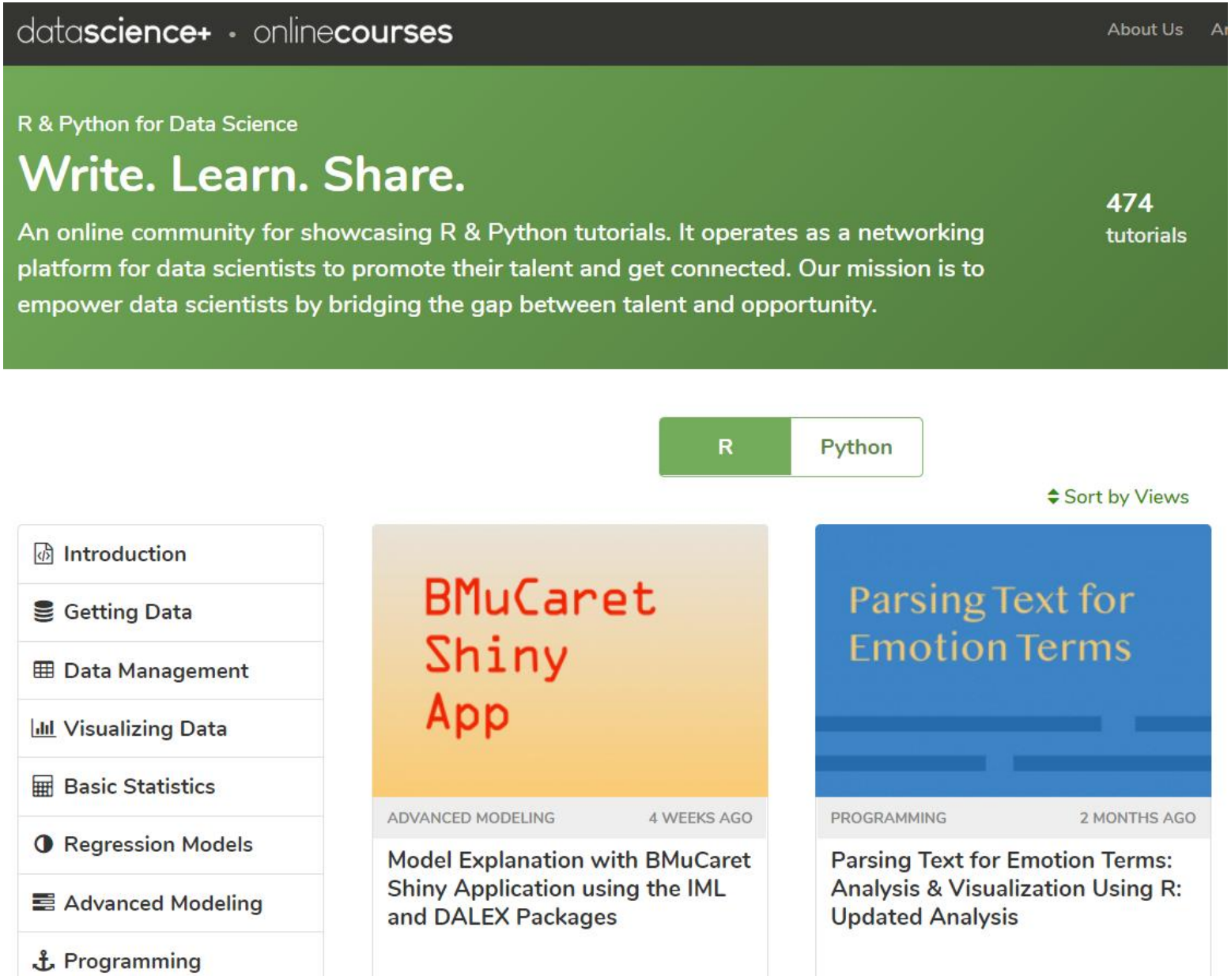
Week 1

Course introduction

The R and R Studio environments

- Command line vs. script file
- Packages
    - Base packages
    - Installing packages from CRAN
    - Installing packages from GitHub
    - Using a package in a R session
- Getting a session's info
- A brief video intro to the RStudio environment

https://datascienceplus.com/



https://rstudio-education.github.io/hopr/

Hands-On Programming with R

Garrett Grolemund

Welcome

This is the website for **"Hands-On Programming with R"**. This book will teach you how to program in R, with hands-on examples. I wrote it for non-programmers to provide a friendly introduction to the R language. You'll learn how to load data, assemble and disassemble data objects, navigate R's environment system, write your own functions, and use all of R's programming tools. Throughout the book, you'll use your newfound skills to solve practical data science problems.

If you are already comfortable with R, and would like to focus instead how to analyze data using R's Tidyverse packages, I recommend R for Data Science, a book that I co-authored with Hadley Wickham.

http://www.cookbook-r.com/

Cookbook for R

Welcome to the Cookbook for R. The goal of the cookbook is to provide solutions to common tasks and problems in analyzing data.

Most of the code in these pages can be copied and pasted into the R command window if you want to see them in action.

1. Basics
2. Numbers
3. Strings
4. Formulas
5. Data input and output
6. Manipulating data
7. Statistical analysis
8. Graphs
9. Scripts and functions
10. Tools for experiments

https://openbiox.github.io/Cookbook-for-R-Chinese/

Cookbook for R 中文版

Winston Chang（著）

王诗翔 等（译）

2019-06-27

前言

这里是《Cookbook for R》中文版的网站，本站内容的翻译由生物信息学创新协作小组 Openbiox 完成翻译并维护。这里以直观明了的问题/需求和方案为基本内容向读者介绍 R 的基础和如何解决常见的分析问题。从阅读中读者可以学习安装和使用三方包、操作基础的数据类型，学习数据的导入、操作和可视化，学习统计分析和编写脚本等内容。这本书是 R 诸多问题解决方案的参考手册，建议读者配合系统的 R 语言读物使用，如《R for Data Science》。

https://www.tutorialspoint.com/r/index.htm

tutorialspoint
SIMPLYEASYLEARNING

Categories

Library Videos Q/A eBooks Search y

LEARN R PROGRAMMING

programming language

R Tutorial

- R - Home
- R - Overview
- R - Environment Setup
- R - Basic Syntax
- R - Data Types
- R - Variables

LEARN R PROGRAMMING

simply easy learning

R Tutorial

PDF Versic | Quick Guid | Resources | Job Search | Discussion

R is a programming language and software environment for statistical analysis, graphics representation and reporting. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team. R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating

http://r-statistics.co/

r-statistics.co by Selva Prabhakaran

Search..

- **Tutorial**
- R Tutorial
- **ggplot2**
- ggplot2 Short Tutorial
- ggplot2 Tutorial 1 - Intro
- ggplot2 Tutorial 2 - Theme
- ggplot2 Tutorial 3 - Masterlist
- ggplot2 Quickref
- **Foundations**
- Linear Regression
- Statistical Tests
- Missing Value Treatment
- Outlier Analysis
- Feature Selection
- Model Selection

Welcome to r-statistics.co

An educational resource for those seeking knowledge related to machine learning and statistical computing in R. Here, you will find quality articles, with working R code and examples, where, the goal is to make the #rstats concepts clear and as simple as possible.

This is built by keeping in mind, statisticians who are new to R programming language, R programmers without a stats background, analysts who work in SAS or python, college grads and developers who are relatively new to both R and stats/ML. If you are completely new to R, this R tutorial is a good place it start.

Video Courses

Learn R By Intensive Practice: Gain clear understanding of base R programming concepts and internalise through a lot of practice.

Introduction to R Programming: Covers core programming with R, statistics fundamentals, base graphics, data manipulation with data.table and dplyr. (For Beginners)

Mastering R Programming: Covers advanced machine learning concepts and interesting programming exercises throughout the course. (For Data scientists and Machine Learning Engineers)

https://www.statmethods.net/index.html

Quick-R
powered by DataCamp

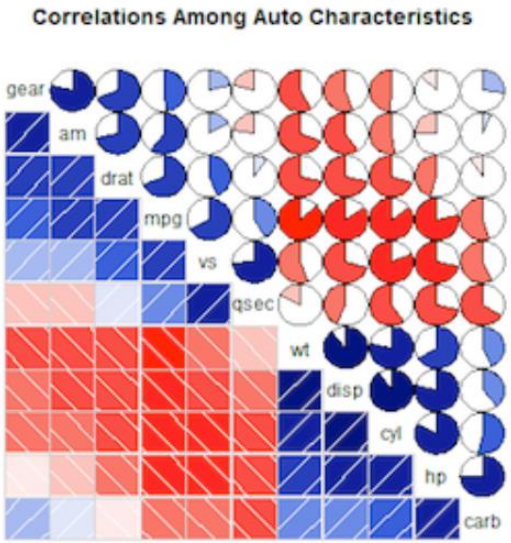
R Tutorial | R Interface | Data Input | Data Management | Statistics | Advanced Statistics | Graphs | Advanced Graphs

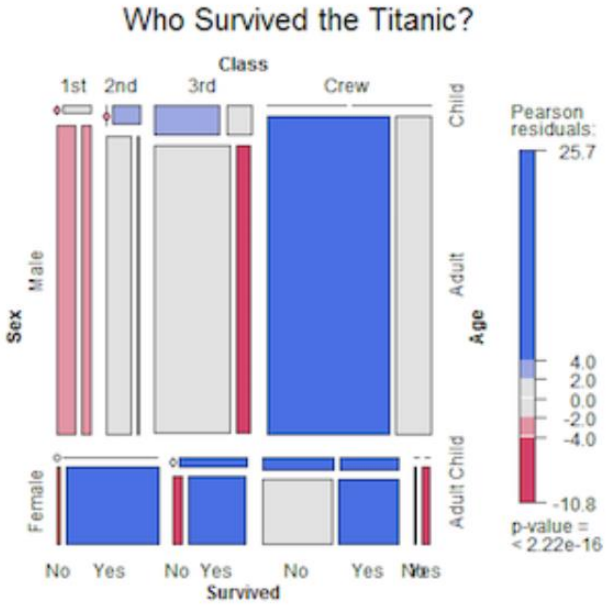
Site Contents

- Learning R
- R Tutorial
- R Interface
- Data Input
- Data Management
- Statistics
- Advanced Statistics
- Graphs
- Advanced Graphs

About this Site

About Quick-R





https://r-coder.com/

ALL ABOUT R PROGRAMMING

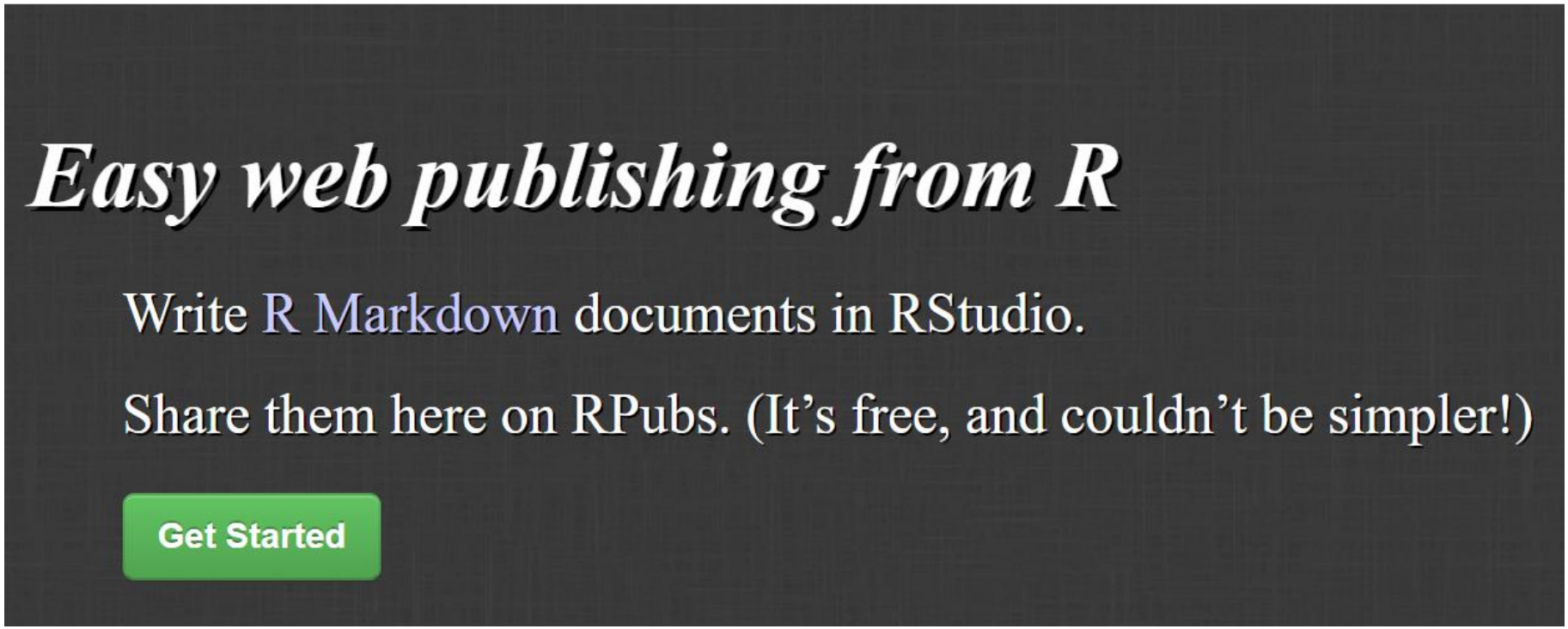
R CODER

Begin your data science career with R language!

START LEARNING NOW

INTRODUCTION	DATA STRUCTURES	DATA WRANGLING	PROGRAMMING	IMPORT & EXPORT	GRAPHICS
R installation	Vector	Sort and order	Creating functions	Read txt files	Plot function
Operators	Factor	Subset	If else statement	Import CSV files	Scatter plot
Data types	List	Categorize	For loop	Read Excel files	Barplot
Workspace	Matrix	Split	While loop	Export data	Histogram
Working directory	Data Frame	Merge data frames	apply function		Density plot
Getting help			tapply function		Boxplot
Install packages			sapply function		Save plot

http://rpubs.com/



https://bookdown.org/ndphillips/YaRrr/

YaRrr! The Pirate's Guide to R

YaRrr! The Pirate's Guide to R

Nathaniel D. Phillips

2018-01-22

Chapter 1 Preface

https://www.guru99.com/r-tutorial.html

R Tutorial for Beginners: Learning R Programming

Training Summary

R is a programming language is widely used by data scientists and major corporations like Google, Airbnb, Facebook etc. for data analysis. This is a complete course on R for beginners and covers basics to advance topics like machine learning algorithm, linear regression, time series, statistical inference etc.

What should I know?

Basic knowledge of statistical concept is a plus.

Course Syllabus

Introduction

Tutorial	What is R Programming Language? Introduction & Basics
Tutorial	How to Download & Install R, RStudio, Anaconda on Mac or Windows
Tutorial	R Data Types, Arithmetic & Logical Operators with Example
Tutorial	R Matrix Tutorial: Create, Print, add Column, Slice
Tutorial	Factor in R: Categorical & Continuous Variables

Data Preparation

https://www.datamentor.io/r-programming/

Interested in Data Science? Get Started With Data Science in R.

Sign Up Now

Learn R Programming

The Definitive Guide

R is a programming language and environment commonly used in statistical computing, data analytics and scientific research.

It is one of the most popular languages used by statisticians, data analysts, researchers and marketers to retrieve, clean, analyze, visualize and present data.

Due to its expressive syntax and easy-to-use interface, it has grown in popularity in recent years.

TABLE OF CONTENTS

https://pairach.com/2012/02/26/r-tutorials-from-universities-around-the-world/

Pairach Piboonrungroj, PhD

Supply Chain Analytics with R

About Me | Research | Teaching | Blog | Events | In Press | Contact me

R-Uni (A List of 100 Free R Tutorials and Resources in University webpages)

by Pairach on February 26, 2012

28 Votes

Search

Free 100 Online tutorials for R programming, Statistics and Graphics

Here is a list of FREE R tutorials hosted in official website of universities around the world. The tutorials are listed in no particular order, actually based on when I have discovered it. They will be categorised soon. Please kindly suggest me other university-hosted online R tutorials by email to me@pairach.com.

A list of R tutorials, which are not hosted in the webpages of academic institutes can be found here.

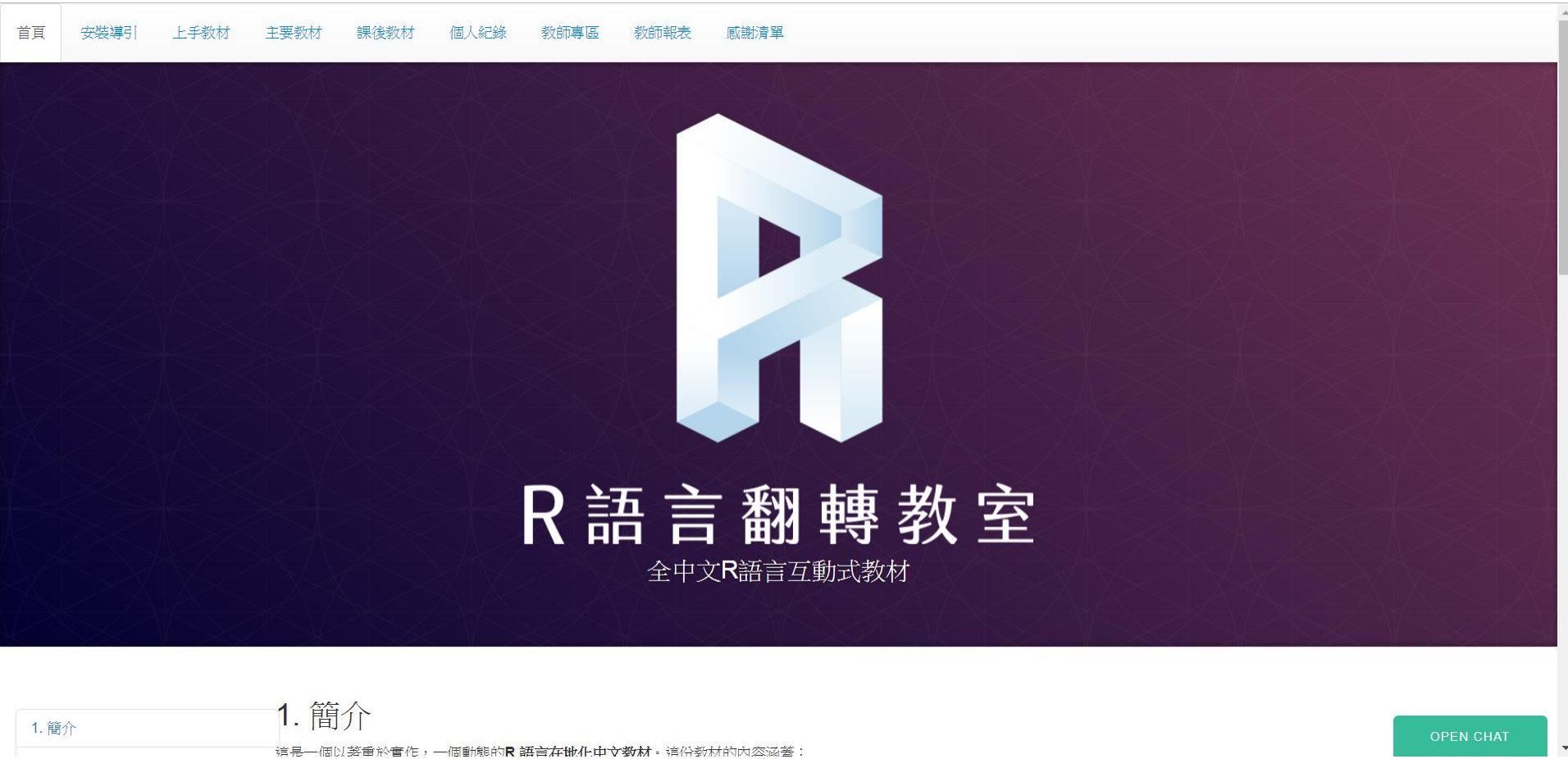
1. **University of Oxford**
   Modern Applied Statistics with S, 4edn (On-line material)[url]
   by W. N. Venables and B. D. Ripley
2. **University of California at Davis**
   Getting Started with the R Data Analysis Package [url]
   by Norm Matloff

https://static-bcrf.biochem.wisc.edu/courses/Tabular-data-analysis-with-R-and-Tidyverse/book/

Tabular data analysis with R and Tidyverse:

Environmental Health

http://datascienceandr.org/



https://rpubs.com/skydome20/Table

R系列筆記

skydome20

2019/03/16(update)

【當我開始寫R 系列筆記：緣起與心路歷程】

https://bookdown.org/tonykuoyj/eloquentr/

認識 R 的美好

郭耀仁

1 關於

1.1 關於作者

是郭耀仁，資料科學與推廣教育的愛好者，喜歡使用 R 語言與 Python 做資料科學應用，在台大資工系統訓練班開設多門 R 語言與 Python 的相關課程，亦與企業合作提供客製化的內訓課程；同時也是一個超棒的中文資料科學專欄 DataInPoint 的主編；這個專欄與波士頓的資料科學教學團隊 DataCamp 有行銷合作（Affiliate Marketing）。

如果您有 R 語言、Python、資料科學、教學、專案或顧問的需求，可以 email 與我聯絡：tonykuoyj@gmail.com

https://medium.com/datainpoint/r-essentials/home

DataInPoint 進擊的資料科學 輕鬆學習 R 語言 PYTHON 與網頁資料擷取 NUMPY 101 PANDAS 101 MATPLOTLIB 101 | DATA

輕鬆學習 R 語言

從基礎到應用，成為 R 資料分析師

輕鬆學習 R 語言：起步走

關於 R 的特性、開發環境與開始寫 R

郭耀仁 Yao-Jen Kuo

Nov 8, 2018 · 12 min read ★

https://nhsrcommunity.com/