

110年衛生福利資料科學中心

R軟體推廣課程《進階篇》

多重資料操作

講師：劉品崧 統計分析師

花蓮慈濟醫院

課程大綱

- 前言
- for迴圈
- 衛福資料庫簡介
- 清單物件
- 自訂函數與apply

前言

- 使用程式指令進行資料處理
- 為什麼要寫迴圈

使用程式指令進行資料處理

- 以難度來區分階段

- 小學生(單一物件 & 單一處理)：計算2加上3的數學計算

- > `print(2+2)`

- 國中生(多個物件 & 單一處理)：計算2 ~ 17這個數列的2次方

- > `print(c(2:17)**2)`

- 高中生(多個物件 & 多個處理)：計算2 ~ 17這個數列的2次方、3次方 ... 1028次方

- > `print(c(2:17)**2)`

- > `print(c(2:17)**3)`

- ... (中間應該還有1025行QAQ)

- > `print(c(2:17)**1028)`

- 大學生(多個流程)

為什麼要寫迴圈(1)

- 降低錯誤

- > `print(c(2:17)**2)`
 - > `print(c(2:17)**3)`
 - > `print(c(2:17)**4)`
 - ...
 - > `print(c(2:17)**1026)`
 - > `print(c(2:17)**1072)`
 - > `print(c(2:17)**1028)`

- 常見錯誤類型

- 打錯字
- 少打字
- 多打字
- 胖手指
- 複製貼上之後忘記要改 ==

為什麼要寫迴圈(2)

- 化繁為簡

```
> for (i in c(1:1028)) {  
    print(c(2:17)**i)  
}
```

- 核心動作不變

- 剩下的交給迴圈去逐一處理

為什麼要寫迴圈(3)

- 容易修改(迴圈內容)

```
> for (i in c(1:1028)) {  
    print(c(2:17)**i)  
}
```

```
> for (i in c(1:1028)) {  
    print(c(3:18)**i)  
}
```

- 容易修改(迴圈向量)

```
> for (i in c(1:1028)) {  
    print(c(2:17)**i)  
}
```

```
> for (i in c(1029:2048)) {  
    print(c(2:17)**i)  
}
```

for迴圈

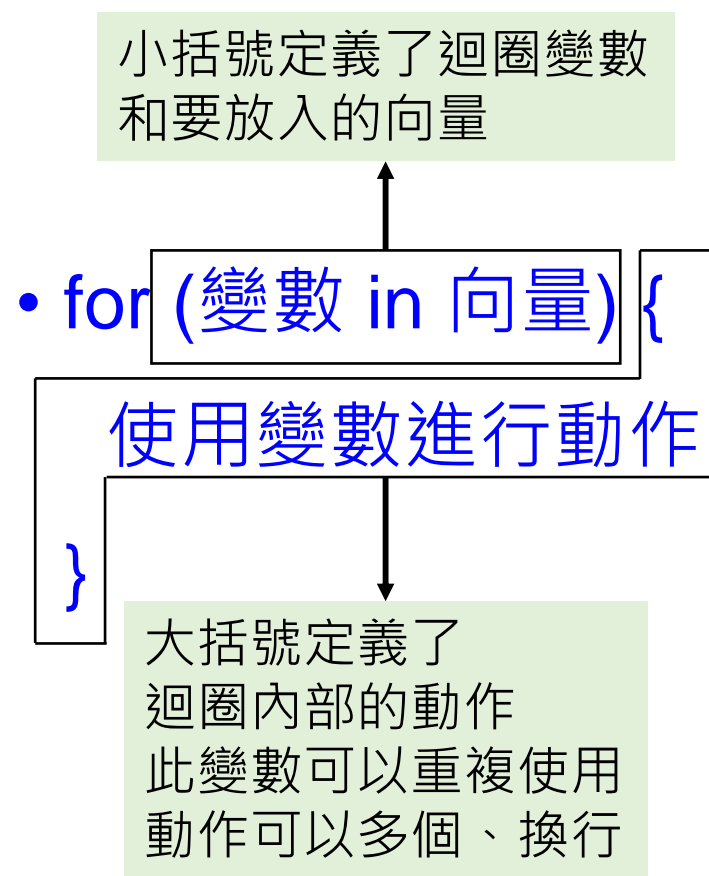
- 核心
- 迴圈
- 複雜動作
- 多層迴圈
- 練習時間

核心

- 核心是指程式指令的主要目的
 - 列印
 - 摘要報表
 - 擷取data
- 初學者務必謹記
 - 先寫一個對的核心
 - 再寫一個好的迴圈
- 建立物件希望逐一列印
 - `a1 <- sample(x = LETTERS, size = 10)`
- 如果土炮手工
 - `print(a1[1])`
 - `print(a1[2])`
 - `print(a1[3])`
 - ...
 - `print(a1[8])`
 - `print(a1[9])`
 - `print(a1[10])`
- 量多的時候會很悲劇 QAQ

for迴圈

- 基本架構



- 把前一頁的範例改成迴圈

```
• for (i in a1) {  
    print(i)  
}
```

- a1向量元素會逐一放入迴圈中
- 進去成為變數 i
- 就這麼簡單 ~

複雜動作

i 可以重複被使用

```
for (i in 1:length(a1)) {  
    print(paste0("第", i, "個抽出來的字母是", a1[i]))  
}
```

被作為物件使用

被作為索引使用

指令也可以用在多行指令

```
for (i in 1:length(a1)) {  
    print(paste0("第", i, "個抽出來的字母是"))  
    print(a1[i])  
}
```

程式範例使用函數簡介

- `Sys.time` : 回傳當下日期 & 時間
- `sample(x=樣本空間, size=抽樣個數)`
- `table` : 進行向量的次數分析
- `Sort` : 排序
- `names` : 取出元素的名稱

多層迴圈(1)迴圈A內包含迴圈B

```
for (i in c(2015:2018)) {  
  for (j in c(1:12)) {  
    print(paste("讀取",i,"年" j "月西醫門診費用檔"))  
  }  
}
```

迴圈B

迴圈A

多層迴圈(2)迴圈A內包含迴圈B1+B2

```
for (i in c(1:3)) {
```

```
  for (j in c("A", "B", "C")) {
```

```
    print(paste(i, j))
```

```
  }
```

迴圈B1

```
  for (j in c("x", "y", "z")) {
```

```
    print(paste(i, j))
```

```
  }
```

迴圈B2

```
}
```

迴圈A

練習時間

- 使用迴圈的寫法列印出以下文字
 - 今年農曆正月1日是2021-2-12，星期五
 - 今年農曆正月2日是2021-2-13，星期六
 - 今年農曆正月3日是2021-2-14，星期日
 - ...
 - 今年農曆正月15日是2021-2-26，星期五

衛福資料庫簡介

- 概論
- 承保檔
- 門診檔

衛生福利模擬資料檔

- 虛擬10萬人資料
 - 今日使用資料庫：健保門診檔、健保承保檔
- 優點：很棒的練習材料
 - 未來有可能進入資科中心分析資料
 - 助於研究樣本數估計、決定實際資料購買策略
- 缺點：部分重要欄位無法使用/轉譯
 - func_date/in_date：就診日期，無法轉換成實際日期

門診檔費用/醫令檔

門診費用檔(CD檔/OPDTE)

id	id_s	age	func_date	func_type	prsn_id	icd9cm_1	icd9cm_2	icd9cm_3	t_dot
{%{{{/ - /]""/{&	1	62歲	21842	12	:]:*_>\$~~[* (43400	40200	2720	1232

hosp_id	fee_ym	appl_date	appl_type	case_type	seq_no
3737859	201401	69898	1	4	69649

門診醫令檔(OO檔/OPDTO)

order_type	drug_no	total_dot
9	05209A	69
0	00171A	443
1	A013382100	20
1	AC43862100	353
1	B017218124	224
1	B021571100	123

1. 一名62歲男性(`id_s == "1"`)在某日前往醫院神經科(`func_type == "12"`)就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(`icd9cm_1 == "43400"`)，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(`05209A`)，1筆為診察費(`00171A`)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(`AC43862100`)、胃藥(`A013382100`)、解便秘藥物(`B017218124`)、高血壓或心臟用藥(`B021571100`)
4. 醫令檔6項申報點數(`total_dot`)加總，等於費用檔的總申報點數(`t_dot`)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

全民健保處方及治療明細檔_門急診—西醫、中醫及牙醫

(H_NHI_OPDTE)

模擬資料庫使用手冊

一、檔案內容說明

檔案代號	H_NHI_OPDTE	資料筆數	1. 114,910 筆 (西醫) 2. 12,438 筆 (牙醫) 3. 14,591 筆 (中醫) (以 2014/12 為例)
中文檔名	全民健保處方及治療明細檔	英文檔名	AMBULATORY CARE EXPENDITURES BY VISITS
檔案大小	1. 30.9MB (西醫) 2. 3.43MB (牙醫) 3. 4MB (中醫) (以 2014/12 為例)	欄位數	51
屬性	月份歷史檔	週期	月
資料描述	1. 原則上一次就醫 (同一醫院同一天同一科別一筆資料) 一筆資料。 2. 內容包含病人當次就醫之疾病、醫院及費用等相關資料。 3. 資料年度包含 2014 年		
注意事項	1. 計算人次時，補報資料不應重覆計算。 2. 計算費用時，補報資料應合併。		
主鍵與比對欄位	1. 個人身分證字號、醫師身分證字號、藥師身分證字號及醫療機構代碼經加密處理。 2. 本檔與門急診醫令明細檔 (H_NHI_OPDTE) 比對之唯一鍵值為就醫代碼：FEE_YM+APPL_TYPE+HOSP_ID+APPL_DATE+CASE_TYPE+SEQ_NO。 3. 個人歸戶及個人資料比對的鍵值：ID。 4. 醫事機構歸戶與資料比對的鍵值：HOSP_ID。		

(1)檔案名稱

(2)時間間隔

(3)資料產生方式

(4)如何連結

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診

(HEALTH-04: H_NHI_OPDTE)

模擬資料庫使用手冊

一、檔案內容說明

(1)檔案名稱				
檔案代號	H_NHI_OPDTE	資料筆數	127,491 筆(西醫) 34,545 筆(牙醫) 108,607 筆(中醫) (以 2014/12 為例)	
中文檔名	全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診	英文檔名	Details of Ambulatory Care Orders	
檔案大小	18.6MB(西醫) 5.12MB(牙醫) 15.8MB(中醫) (以 2014/12 為例)	欄位數	21	
(2)時間間隔				
屬性	月份歷史檔	週期	月	譯碼簿 更新日期 2016/09/25
資料描述	1. 為一次門診多筆醫令資料。 2. 相同之 FEE_YM + APPL_TYPE + APPL_DATE + CASE_TYPE + SEQ_NO + HOSP_ID 視為同一次就醫之醫令。 3. 每一筆資料包括醫令碼、數量及點數等。			
注意事項	同一次門診之各筆醫令點數欄位(TOTAL_DOT)加總等於門診明細檔(H_NHI_OPDTE)之合計點數(T_DOT)。			
主鍵與比對欄位	1. 醫療機構代碼(HOSP_ID)經加密處理。 2. 本檔無病人身份證字號，以 FEE_YM+APPL_TYPE+APPL_DATE+CASE_TYPE+SEQ_NO+HOSP_ID 為鍵值與門診明細檔(H_NHI_OPDTE)對照。 3. 醫療機構歸戶及醫療機構資料對照的鍵值：HOSP_ID。			
(4)如何連				

(3)資料產生方式

(4)如何連結

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療明細檔 (H_NHI_OPDTE)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
1	身分證字號	ID	Char	12	經加密處理。國民身分證統一編號，或外籍居留證號碼（如無居留證號碼請填護照號碼），左靠不足補空白
38	性別	ID_S	Char	1	1：男性 2：女性 9：不詳
16	就醫年齡	AGE	Char	20	以就醫日期及生日計算。 「0-14天」、「15-28天」、「29天-未滿一歲」、1歲、2歲、...、84歲、85歲以上
14	就醫科別	FUNC_TYPE	Char	2	詳見欄位就醫科別(FUNC_TYPE)譯碼說明

3.就醫科別 (FUNC_TYPE)

科別	代碼
眼科	10
皮膚科	11
神經科	12
精神科	13
復健科	14

通常代碼繁多的欄位
會另外整理成附錄
請參考譯碼簿 p.13

1. 一名62歲男性(`id_s == "1"`)在某日前往醫院神經科(`func_type == "12"`)就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(`icd9cm_1 == "43400"`)，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(`total_dot`)加總，等於費用檔的總申報點數(`t_dot`)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療明細檔 (H_NHI_OPDTE)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
20	國際疾病分類號一	ICD9CM_1	Char	5	左靠不足補空白， 詳見欄位國際疾病分類號 (ICD9CM_1~ICD9CM_3)譯碼說明
21	國際疾病分類號二	ICD9CM_2	Char	5	
22	國際疾病分類號三	ICD9CM_3	Char	5	

CEREBROVASCULAR DISEASE (430-438)

434 Occlusion of cerebral arteries

The following fifth-digit subclassification is for use with category 434:

0 without mention of cerebral infarction

1 with cerebral infarction

Use additional code, if applicable, to identify status post administration of tPA (rtPA) in a different facility within the last 24 hours prior to admission to current facility (V45.88)

434.0 Cerebral thrombosis

[0-1]

Thrombosis of cerebral arteries

434.1 Cerebral embolism

[0-1]

434.9 Cerebral artery occlusion, unspecified

[0-1]

(icd9.chrisendres.com)

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診 (H_NHI_OPDTO)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
7	醫令類別	ORDER_TYPE	Char	1	詳醫令類別(ORDER_TYPE) 譯碼說明。

2. 醫令類別(ORDER_TYPE)

醫令類別代碼	說明
0	診察費
1	用藥明細
2	診療明細
3	特殊材料
4	不得另計價之藥品、檢驗(查)或診療項目
5	EPO 注射
6	HCT 檢驗
7	代檢(95.07 刪除)
8	器官捐贈
9	藥事服務費

請參考譯碼簿 p.7

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診 (H_NHI_OPDTO)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
8	藥品(項目)代號	DRUG_NO	Char	12	參考中央健保局網頁公告之「全民健保藥品」、「醫療費用支付標準碼」或特殊材料代碼。

醫療服務給付項目及支付標準網路查詢服務

資料更新日期：1100701 共2856817人



衛生福利部中央健康保險署

查詢條件

中文項目名稱：	<input type="text" value="請輸入中文項目名稱"/>
英文項目名稱：	<input type="text" value="請輸入英文項目名稱"/>
診療項目代碼：	<input type="text" value="05209A"/>
支付價生效日期：	<input type="text" value="▼"/> <input type="text" value="▼"/> 月



項目代碼	診療項目	支付點數	參考起迄日
05209A	門診藥事服務費 - 慢性病處方給藥二十八天以上	83	110.03.01 ~ 迄今

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

譯碼簿(codebook)的使用：門診檔費用/醫令檔

二、檔案格式及資料描述

全民健保處方及治療醫令明細檔—西醫、中醫及牙醫門診 (H_NHI_OPDТО)					
序號	中文欄位名稱	英文欄位名稱	型態	長度	資料描述
8	藥品(項目)代號	DRUG_NO	Char	12	參考中央健保局網頁公告之「全民健保藥品」、「醫療費用支付標準碼」或特殊材料代碼。

健保用藥品項查詢

藥品名稱(中文)：	<input type="text"/>
藥品名稱(英文)：	<input type="text"/>
藥品代碼：	<input type="text" value="AC43862100"/>
劑型：	<input type="text"/>

藥品代碼	藥品名稱 (英文)	藥品名稱 (中文)	成分
AC43862100	COFARIN TAB 1MG "GENTLE"	“政德”可化凝錠 1 毫克	WARFARIN SODIUM

1. 一名62歲男性(id_s == "1")在某日前往醫院神經科(func_type == "12")就診，申報總點數為1,232點。
2. 醫師給予的主診斷為中風(icd9cm_1 == "43400")，同時有其他診斷和共病。
3. 本次申報醫令共6筆，1筆為藥事服務費(05209A)，1筆為診察費(00171A)，4筆為用藥明細，本次醫師開立的處方有抗凝血劑(AC43862100)、胃藥(A013382100)、解便秘藥物(B017218124)、高血壓或心臟用藥(B021571100)
4. 醫令檔6項申報點數(total_dot)加總，等於費用檔的總申報點數(t_dot)

實際衛福資料庫疾病診斷碼版本：以中風為例

• ICD-9-CM，2015 (含)以前



Home > 2015 ICD-9-CM Diagnosis Codes > Diseases Of The Circulatory System 390-459 >

Cerebrovascular Disease 430-438 >

- [430](#) Subarachnoid hemorrhage
- [431](#) Intracerebral hemorrhage
- [432](#) Other and unspecified intracranial hemorrhage
- [433](#) Occlusion and stenosis of precerebral arteries
- [434](#) Occlusion of cerebral arteries
- [435](#) Transient cerebral ischemia
- [436](#) Acute, but ill-defined, cerebrovascular disease
- [437](#) Other and ill-defined cerebrovascular disease
- [438](#) Late effects of cerebrovascular disease

• ICD-10-CM，2016(含)以後



ICD-10-CM Codes > I00-I99 Cerebrovascular diseases > Cerebrovascular diseases I60-I69

Cerebrovascular diseases I60-I69

Codes

- [I60](#) Nontraumatic subarachnoid hemorrhage
- [I61](#) Nontraumatic intracerebral hemorrhage
- [I62](#) Other and unspecified nontraumatic intracranial hemorrhage
- [I63](#) Cerebral infarction
- [I65](#) Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction
- [I66](#) Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction
- [I67](#) Other cerebrovascular diseases
- [I68](#) Cerebrovascular disorders in diseases classified elsewhere
- [I69](#) Sequelae of cerebrovascular disease

健保承保檔

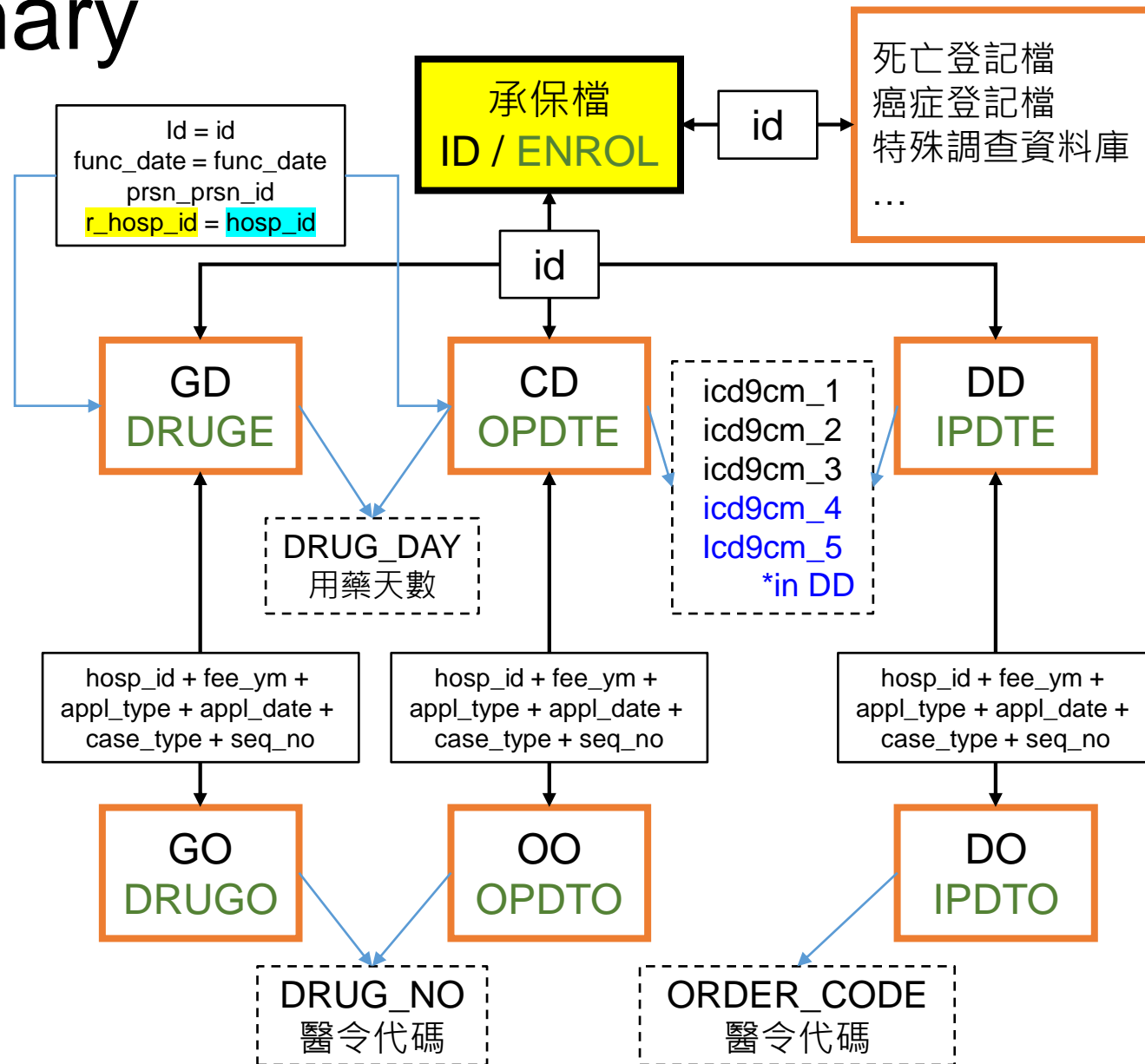
健保承保檔(ID檔/ENROL)

id	id_s	id_birth_y	prem_ym	id1_amt
@]@"&]_;@%)	2	1949	201401	38200
@]@"&]_;@%)	2	1949	201403	1249

1. 一名65歲女性(`id_s == "2"`)的投保紀錄變化。
2. 2014年1月(`prem_ym == "201401"`)的投保紀錄顯示，她的投保金額(`id1_amt`)為38,200元。
3. 2014年3月(`prem_ym == "201403"`)的投保紀錄顯示，她的投保金額(`id1_amt`)為1,249元。
4. 承保檔就如同疾病，是會變化的，在撈取資料的時候務必要小心挑選以及限制時間。

衛生福利資料庫summary

- 本次使用檔案 (現今檔名/國衛院舊稱)
 - 門診費用檔(OPDTE / CD) , 一月三檔
 - 門診醫令檔(OPDTO / OO) , 一月三檔
 - 檔名結尾：西醫10、牙醫20、中醫30
 - 住院費用檔(IPDTE / DD) , 一年一檔
 - 住院醫令檔(IPDTO / DO) , 一月一檔
 - 健保承保檔(ENROL / ID) , 一月一檔
 - 死因統計檔(DEATH / DR) , 一年一檔
- 資料庫手冊
 - 確認你要的資料需要使用什麼欄位
 - 確認跨資料庫之間要怎麼連接
 - 身分證號、時間、空間、串聯欄位
 - 瞭解資料庫的能與不能



清單(list)

- 清單元素的特性
- 清單基礎使用
- 清單進階應用
- 練習時間

為什麼需要使用清單物件

- 前面的範例比較單純
- 最終只是把結果print出來
- 但是這樣無法讓未來繼續使用
 - 當你的output是data的時候
沒辦法繼續使用會很糟糕
- 清單就是一個解決方案



清單元素的特性(1)

可以容納多種不同屬性的物件

```
teacher <- list(  
  last_name = "Liu", ----- character  
  lecture_date = ymd(20210819), ----- Date  
  student_number = 40 ----- numeric  
)
```

清單元素的特性(2)

如果清單裡面全部都是data.table

```
draw <- list(  
  data.table(pick = sample(letters, 50, replace = T)),  
  data.table(pick = sample(letters, 50, replace = T)),  
  data.table(pick = sample(letters, 50, replace = T)),  
  data.table(pick = sample(letters, 50, replace = T)),  
  data.table(pick = sample(letters, 50, replace = T))  
)
```

可以使用rbindlist()將元素一起堆疊起來

```
draw_rbind <- rbindlist(draw)
```


基礎使用：健保承保人數

- 核心理念

- 研究目標：探討每月健保納保人數量的趨勢
- 預期結果：月份(x軸)與人數(y軸)的摺線圖

- 工作心流

- 取得(`list.files`) 1-12月份承保檔的檔名清單
- 創造(`vector`)一個長度(`length`)與input相同的清單(`mode = "list"`)
- 使用迴圈(`for`)逐一讀取(`fread`)，分析資料筆數(`nrow`)，回傳出去
- 整理數據，繪製折線圖(`ggplot`)

迴圈內部的主要動作

infile

insnumber

Variable

Vector

Main function

Store output

i = 1	h_nhi_enrol10301.csv
i = 2	h_nhi_enrol10302.csv
i = 3	h_nhi_enrol10303.csv
⋮	⋮
i = 10	h_nhi_enrol10310.csv
i = 11	h_nhi_enrol10311.csv
i = 12	h_nhi_enrol10312.csv

```
# 讀取資料
temp <- fread(infile[[i]])

# 中間處理
result <- nrow(temp)

# 將資料筆數回傳
insnumber[[i]] <- result
```

100,000
100,000
98,843
⋮
97,469
97,253
97,229

技術實踐

```
setwd(full)

insfile <- list.files(pattern = "^h_nhi_enrol")
insnumber <- vector(mode = "list", length = length(insfile))

for (i in 1:length(insfile)) {
  temp <- fread(insfile[[i]]) # 讀取檔案
  result <- nrow(temp)        # 中間處理
  insnumber[[i]] <- result    # 將資料筆數回傳
  rm(temp, result)           # 整理記憶體，避免卡住下一個檔案的讀取
}

insnumber_out <- data.table(month = c(1:12), people = unlist(insnumber))

ggplot(data = insnumber_out, aes(x = month, y = people)) + geom_line() +
  scale_y_continuous(limits = c(95000, 100000))
```

進階應用：糖尿病病人的人數

- 核心理念

- 研究目標：探討每月西醫門診糖尿病病人數量的趨勢
 - 糖尿病病人：第一診斷碼(ICD9CM_1)為250開頭
- 預期結果：月份(x軸)與人數(y軸)的摺線圖

- 工作心流

- 取得1-12月份西醫門診申報檔的檔名清單
- 創造一個長度與input相同的清單
- 逐一讀取，找出糖尿病病人，分析資料筆數，回傳出去
- 整理數據，繪製折線圖

練習時間：糖尿病病人的性別比例

- 核心理念

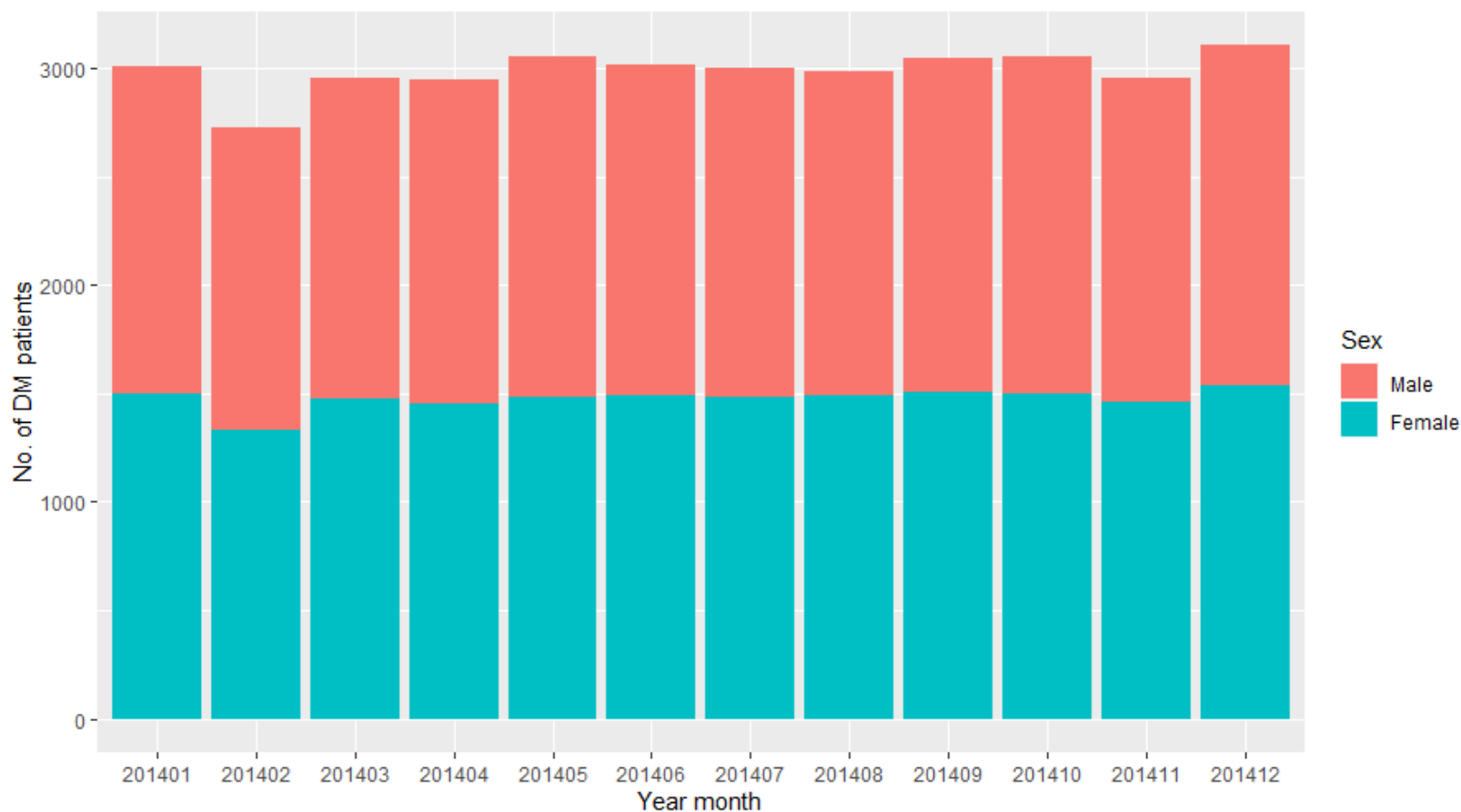
- 研究目標：探討每月西醫門診糖尿病病人數量的趨勢
 - 糖尿病病人：第一診斷碼(ICD9CM_1)為250開頭
- 預期結果：月份(x軸)與性別(y軸)的堆疊長條圖

- 工作心流

- 取得1-12月份西醫門診申報檔的檔名清單
- 創造一個長度與input相同的清單
- 逐一讀取，找出糖尿病病人，帶出性別欄位，回傳出去
- 整理數據，繪製堆疊長條圖

參考解答：糖尿病病人的性別比例

- 先講求方向正確，美編細節不要太執著～



函數與apply

- 自訂函數
- 使用函數
- 函數搭配apply

自訂函數

- 框架

```
name <- function(variables) {  
  ... (函數內容、想要執行的動作)  
}
```

- 範例：計算input的3次方後再開跟號的結果

```
my_func_1 <- function(x){  
  y <- sqrt(x**3)  
  return(y)  
}
```

1. **臨時變數**：在此設定為x，被放入function使用之物件的會被指派為此名稱。不用x也可以，隨心所欲自訂。

2. **return**：function完成後將會回傳什麼，務必要指定，避免回傳錯誤的東西

使用函數

給數值元素

```
my_func_1(3)
```

```
[1] 5.196152
```

給數值向量

```
my_input_1 <- c(10:27)
```

```
my_func_1(my_input_1)
```

```
[1] 31.62278 36.48287 41.56922 46.87217 52.38320 58.09475
```

```
[7] 64.00000 70.09280 76.36753 82.81908 89.44272 96.23409
```

```
[13] 103.18915 110.30413 117.57551 125.00000 132.57451 140.29612
```

函數搭配lapply

```
# 西醫門診清單
```

```
datalist_cd
```

```
# 自訂函數
```

```
find_dm <- function(m){  
  temp <- fread(m) # 讀取資料  
  temp <- temp[grepl("^250", ICD9CM_1)] # 擷取第一診斷碼為糖尿病的人  
  temp <- temp[, .(ID, ID_S, FEE_YM)][ID_S %in% c(1, 2)] # 保留ID和SEX欄位  
  result <- unique(temp[, .(ID, ID_S, month = as.character(FEE_YM))]) # 依照ID歸人  
  return(result) # 輸出資訊  
}
```

```
# lapply(list, function)
```

```
outlist_cd <- lapply(datalist_cd, find_dm)
```

課程討論 & Final remark

- Summary

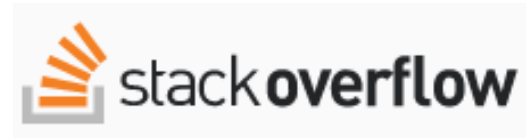
- 先組織核心任務
寫對最重要
- 再撰寫迴圈程式
將流程和效率優化

- 劉品崧

- **Email** : pslu520@gmail.com
- **PubMed** : Peter Pin-Sung Liu

- See more

- Stack Overflow
- STHDA
 - Build by A. Kassambara
creator of survminer



使用圖片版權來源



Created by Adrien Coquet
from Noun Project