

# 109-2-電腦概論與程式設計

## 期中考

學號: 姓名:

2021/05/04

- 注意事項
  - 考試期間
  - 上傳答題檔案
  - 其它事項
- 1 簡易資料處理、以ggplot2做基礎繪圖
  - 1.1 讀入資料 (5分)
  - 1.2 資料型態轉換 (5分)
  - 1.3 資料處理 (10分)
  - 1.4 索引圖 (10分)
  - 1.5 盒形圖 (10分)
  - 1.6 散佈圖 (10分)
- 2 畫激勵函數 (20分)
- 3 股票月成交資訊
  - 3.1 讀入資料 (5分)
  - 3.2 資料型態轉換 (10分)
  - 3.3 線圖 (30分)
- 4 格式 (5分)

---

## 注意事項

### 考試期間

- 請按照平時上課之座位入座。兩台電腦螢幕中間應使用隔板(或拿包包)擋著。
- 於課程網站(<http://www.hmwu.idv.tw>)下載題目卷。
- 可參考課本、上課講義(包含電子檔)及其它資料，但不能與別人討論。
- 可使用計算機、自己的筆記型電腦及平板電腦，不可使用手機。
- 全程可上網查詢，但不能用通訊軟體FB/LINE等討論。
- 程式設計題，若程式碼直接複製(或照抄)講義上的以不給分為原則。
- 有問題者，請舉手發問。勿與同學交談
- 程式直接寫在本Rmd檔。經 knitr 編譯，產生 .html 檔，需印出 R 程式碼及執行結果。
- 不按照規定作答者，酌量扣分。
- 程式請隨時存檔，避免突然意外發生，程式檔不見。

### 上傳答題檔案

- 於教師網站首頁登入[作業考試上傳區]，帳號: r1092。密碼: xxx。
- 選取「正確的」資料夾上傳，若傳錯，請最終要上傳一份正確的的答題檔案。
- 請上傳「學號-姓名-R-MidtermExam.Rmd」、「學號-姓名-R-MidtermExam.html」。(學號及姓名，改成自己)
- 若上傳檔案格式錯誤，內容亂碼，空檔等等問題。請自行負責。
- 若要重覆上傳(第2次以上)，請在檔名最後加「-2」、「-3」，例如:「學號-姓名-R-MidtermExam.Rmd」等等。
- 上傳兩次(含)以上、格式不合等等酌量扣分。
- 如果上傳網站出現「You can modify the html file, but please keep the link [www.wftpserver.com](http://www.wftpserver.com) at least.」，請將滑鼠移至「網址列」後，按「Enter」即可。若再不行，請換其它瀏覽器(IE/Edge/Firefox/Chrome)。
- 有問題者，請FB私訊老師。
- 總分共120分。

### 其它事項

- 若有題目不會寫、或只會寫一半、或結果是有錯的，導致 knitr 無法編譯產生文件，則可以「不執行有錯的程式碼」，但必需列印此段程式

碼。助教會依照狀況部份給分。

# 1 簡易資料處理、以ggplot2做基礎繪圖

「diabetes.csv」為一記錄糖尿病的資料集，其變數解釋如下：

- 編號 (id)
- 性別 (sex): 1 = 女生, 2 = 男生
- 年齡 (age)
- 教育程度 (edu): 1 = 不識字, 2 = 小學, 3 = 國(初)中, 4 = 高中(職), 5 = 大專以上
- 糖化血紅素 (a1c)
- 體重 (wt)
- 膽固醇 (ldl)
- 收縮壓 (sbp)
- 舒張壓 (dbp)
- 嚴重度 (a1cgp): 1 = 輕, 2 = 重

## 1.1 讀入資料 (5分)

讀入資料集「diabetes.csv」(命名為diabetes)，使用xtable套件，印出資料前後各5筆資料及其結構。(指令提示: xtable)

	id	sex	age	edu	a1c	wt	ldl	sbp	dbp	a1cgp
1	1	2	63	3	8.80	74.00	114	162	73	2
2	2	1	67	5	10.60	66.50	135	155	67	1
3	3	2	80	1	7.80	57.90	144	136	96	2
4	4	1	63	2	7.50	70.20	95	126	74	1
5	5	1	36	2	7.10	87.70	156	123	77	2

diabetes data

	id	sex	age	edu	a1c	wt	ldl	sbp	dbp	a1cgp
131	131	1	65	2	7.90	76.10	111	125	74	2
132	132	2	79	1	8.90	49.50	99	137	71	2
133	133	2	57	2	10.40	58.00	56	152	93	2
134	134	1	61	2	7.60	72.00		148	94	2
135	135	2	48	5	6.60	51.90	153	116	67	2

diabetes data

```
## 'data.frame':    135 obs. of  10 variables:
## $ id      : int   1 2 3 4 5 6 7 8 9 10 ...
## $ sex     : int   2 1 2 1 1 2 2 1 1 2 ...
## $ age     : int   63 67 80 63 36 61 80 70 65 70 ...
## $ edu     : int   3 5 1 2 2 3 2 3 4 1 ...
## $ a1c     : num   8.8 10.6 7.8 7.5 7.1 7 6.7 7.7 8.3 8.8 ...
## $ wt      : num   74 66.5 57.9 70.2 87.7 55 49.3 65.7 63.6 74.6 ...
## $ ldl     : int   114 135 144 95 156 122 NA 56 94 79 ...
## $ sbp     : int   162 155 136 126 123 130 137 123 124 158 ...
## $ dbp     : int   73 67 96 74 77 65 97 70 64 69 ...
## $ a1cgp   : int   2 1 2 1 2 2 NA 1 1 2 ...
```

## 1.2 資料型態轉換 (5分)

判斷變數 sex, edu, a1cgp 是否為factor類別? 若不是，請轉換成 factor 類別變數，其中 edu 為有順序的 factor 類別變數。(要求: 需將上述三變數中的類別編碼轉回文字，例如: 性別 (sex): 1 轉為女生, 2 轉為男生)。更改完後再印出資料結構。(指令提示: supply, factor)

```
##      id  sex  age  edu  alc    wt   ldl   sbp   dbp alcgp
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
## 'data.frame':    135 obs. of  10 variables:
## $ id      : int   1 2 3 4 5 6 7 8 9 10 ...
## $ sex     : Factor w/ 2 levels "女生","男生": 2 1 2 1 1 2 2 1 1 2 ...
## $ age     : int   63 67 80 63 36 61 80 70 65 70 ...
## $ edu     : Ord.factor w/ 5 levels "不識字"<"小學"<...: 3 5 1 2 2 3 2 3 4 1 ...
## $ alc     : num   8.8 10.6 7.8 7.5 7.1 7 6.7 7.7 8.3 8.8 ...
## $ wt      : num   74 66.5 57.9 70.2 87.7 55 49.3 65.7 63.6 74.6 ...
## $ ldl     : int  114 135 144 95 156 122 NA 56 94 79 ...
## $ sbp     : int  162 155 136 126 123 130 137 123 124 158 ...
## $ dbp     : int   73 67 96 74 77 65 97 70 64 69 ...
## $ alcgp   : Factor w/ 2 levels "輕","重": 2 1 2 1 2 2 NA 1 1 2 ...
```

## 1.3 資料處理 (10分)

印出哪些編號 ( id )的紀錄具有遺失值( NA )? 刪除具有 NA 的紀錄，並將剩餘之完整資料另存成一 R 資料框(命名為 diabetes.complete )，其資料維度為何? 印出前10筆紀錄。(指令提示: apply ， is.na )

```
## [1]    7   12   20   41   50   60   62   64   69   74   75   77   87   88   91   95 101 109 112
## [20] 117 119 126 128 130 134
```

```
## [1] 110   10
```

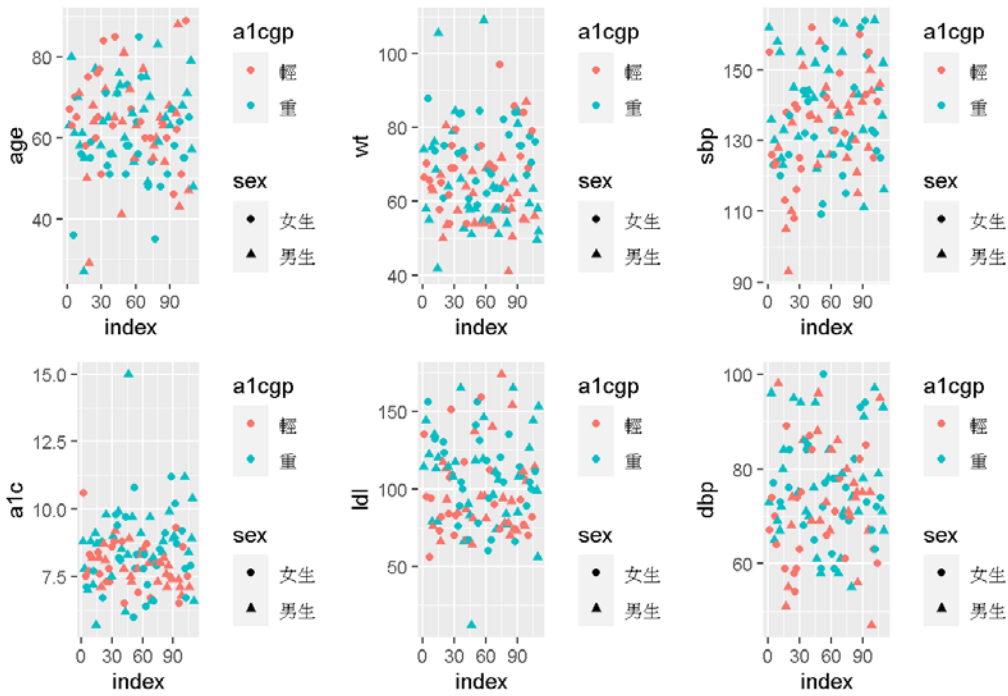
	id	sex	age	edu	a1c	wt	ldl	sbp	dbp	a1cgp
1	1	男生	63	國(初)中	8.80	74.00	114	162	73	重
2	2	女生	67	大專以上	10.60	66.50	135	155	67	輕
3	3	男生	80	不識字	7.80	57.90	144	136	96	重
4	4	女生	63	小學	7.50	70.20	95	126	74	輕
5	5	女生	36	小學	7.10	87.70	156	123	77	重
6	6	男生	61	國(初)中	7.00	55.00	122	130	65	重
8	8	女生	70	國(初)中	7.70	65.70	56	123	70	輕
9	9	女生	65	高中(職)	8.30	63.60	94	124	64	輕
10	10	男生	70	不識字	8.80	74.60	79	158	69	重
11	11	男生	71	不識字	8.20	63.00	76	128	98	輕

diabetes data

## 1.4 索引圖 (10分)

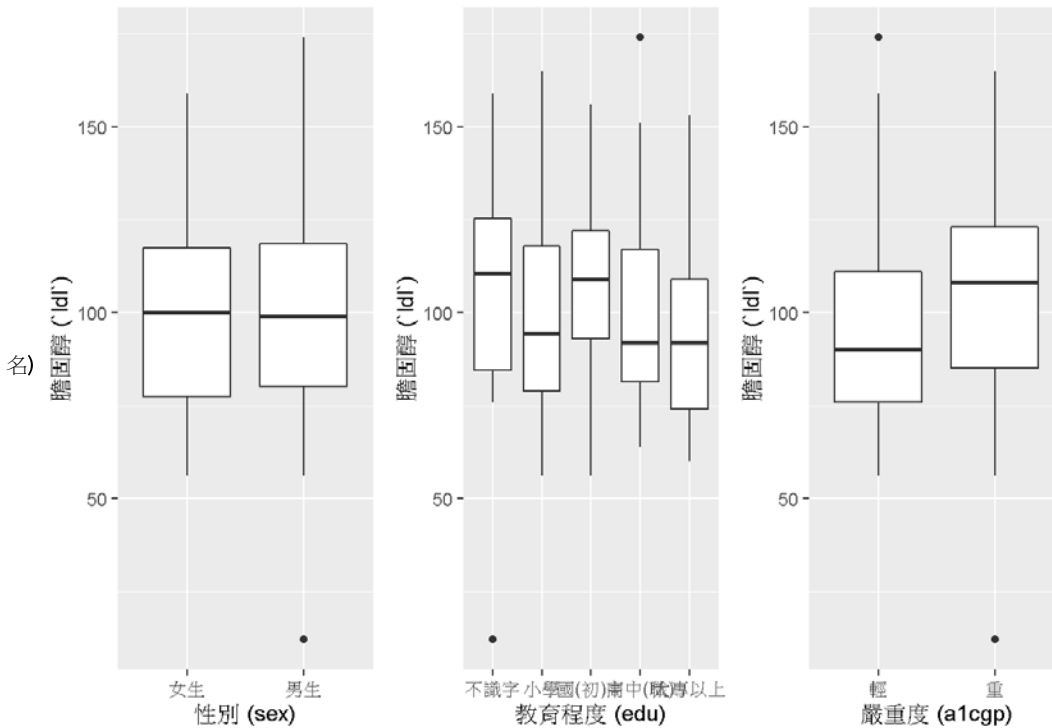
繪出 diabetes.omplete 中連續變數的索引圖( id 除外)。(要求: (1)請六張圖一頁(3 by 2); (2) 需加標題名(“索引圖”)及座標名( $x$ 軸名為 index ,  $y$ 軸名為變數名); (3) 圖上點之顏色以性別( sex )的類別為依據; (4) 圖上點之符號型狀以嚴重度( alcgp )的類別為依據。)(指令提示: lapply ， sapply ， is.numeric ， marrangeGrob )

索引圖



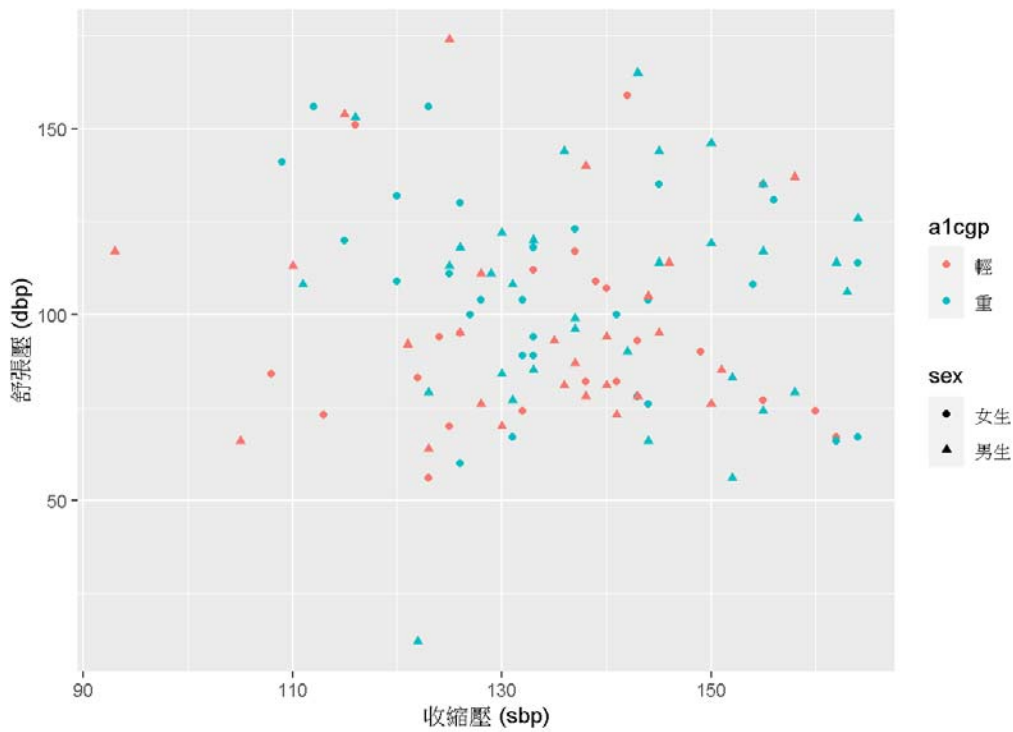
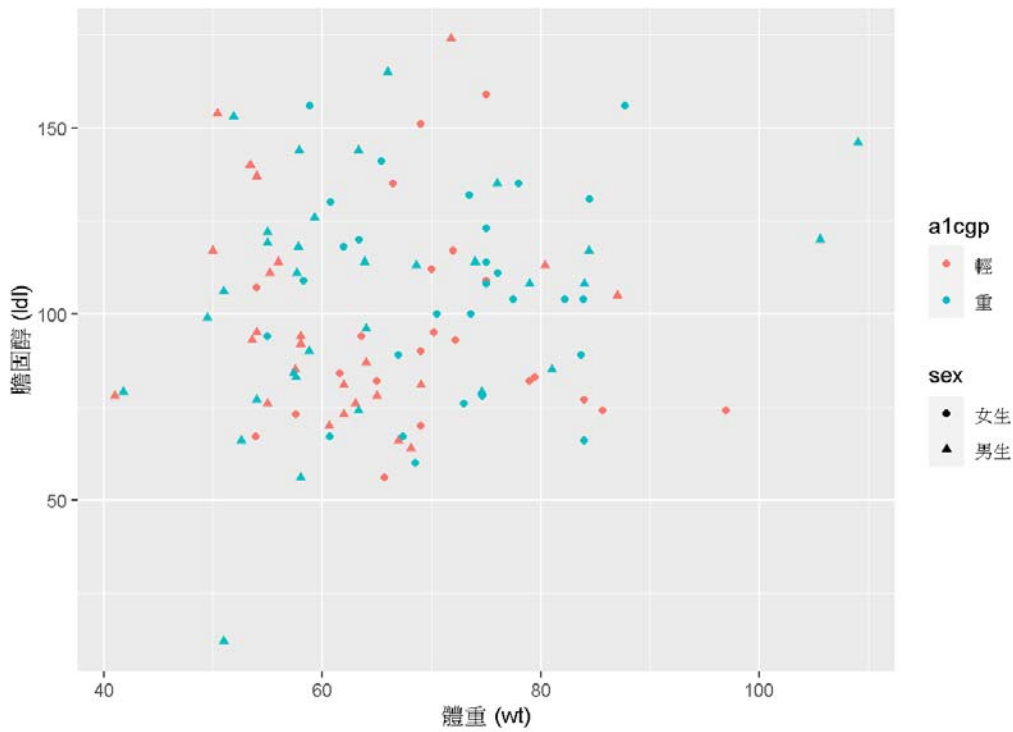
## 1.5 盒形圖 (10分)

各依照變數 `sex`, `edu` 及 `a1cgp` 之類別，繪出 `diabetes.omplete` 中，膽固醇 (`ldl`) 之盒形圖。(要求: (1)請三張圖並列一頁; (2)需加標題及座標)



## 1.6 散佈圖 (10分)

繪出 `diabetes.omplete` 中，兩個並列的散佈圖 ( $x$  vs  $y$ ): (1) 體重 (`wt`) vs 膽固醇 (`ldl`); (2) 收縮壓 (`sbp`) vs 舒張壓 (`dbp`)。圖上點之顏色以性別 (`sex`) 之類別為依據; (4) 圖上點之符號形狀以嚴重度 (`a1cgp`) 之類別為依據。



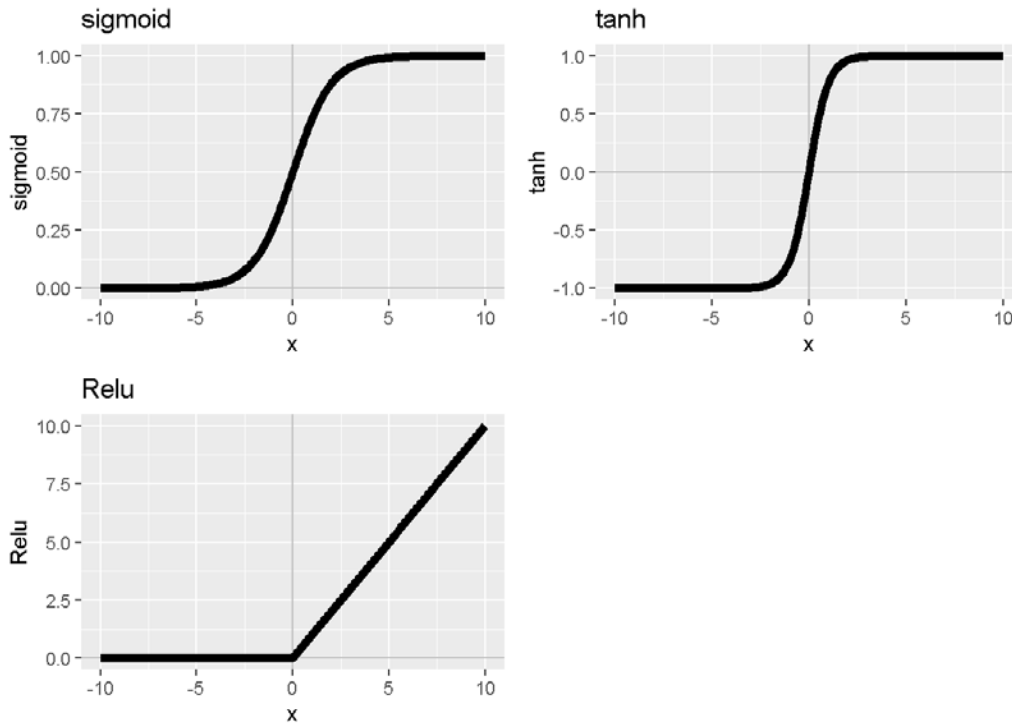
## 2 畫激勵函數 (20分)

深度學習(deep learning)領域中常使用的激勵函數(activation function) ([https://en.wikipedia.org/wiki/Activation\\_function](https://en.wikipedia.org/wiki/Activation_function) ([https://en.wikipedia.org/wiki/Activation\\_function](https://en.wikipedia.org/wiki/Activation_function)))有以下三種，請用 `ggplot2` 套件畫出它們的圖形。

$$\text{Sigmoid or logistic function (sigmoid)} : f(x) = \frac{1}{1 + e^{-x}}, x$$

$$\text{Hyperbolic tangent function (tanh)} : f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Rectified linear unit function (Relu) :  $f(x) = xI(x \geq 0)$



## 3 股票月成交資訊

有民國100年5家半導體公司股票月成交資訊紀錄於資料檔( stock-data.txt )中。

### 3.1 讀入資料 (5分)

請讀入此資料並使用 xtable 套件，印出資料前5筆資料及其結構。

半導體公司	年度	月份	最高價	最低價	加權平均價	成交筆數	成交金額	成交股數	週轉率百分比
1台積電	100	1	78.30	69.60	74.30	263,999	100,578,274,926	1,353,616,348	5.22
2台積電	100	2	77.00	69.90	72.54	235,159	74,985,055,548	1,033,654,452	3.98
3台積電	100	3	72.20	65.70	69.74	276,434	88,459,924,495	1,268,289,393	4.89
4台積電	100	4	73.90	68.00	71.37	211,611	70,177,023,098	983,177,475	3.79
5台積電	100	5	76.90	73.00	74.96	213,185	74,005,599,560	987,256,484	3.80

stockdata data

```
## 'data.frame':    60 obs. of  10 variables:
## $ 半導體公司 : chr  "台積電" "台積電" "台積電" "台積電" ...
## $ 年度 : int  100 100 100 100 100 100 100 100 100 100 ...
## $ 月份 : int  1 2 3 4 5 6 7 8 9 10 ...
## $ 最高價 : num  78.3 77 72.2 73.9 76.9 78.2 73.9 72.8 72.1 74 ...
## $ 最低價 : num  69.6 69.9 65.7 68 73 70.4 68.5 62.2 65.9 68.1 ...
## $ 加權平均價 : num  74.3 72.5 69.7 71.4 75 ...
## $ 成交筆數 : chr  "263,999" "235,159" "276,434" "211,611" ...
## $ 成交金額 : chr  "100,578,274,926" "74,985,055,548" "88,459,924,495" "70,177,023,098" ...
## $ 成交股數 : chr  "1,353,616,348" "1,033,654,452" "1,268,289,393" "983,177,475" ...
## $ 週轉率百分比: num  5.22 3.98 4.89 3.79 3.8 4.99 3.96 4.9 4.14 3.27 ...
```

### 3.2 資料型態轉換 (10分)

請將資料中的「成交筆數」、「成交金額」及「成交股數」轉成數值型變數後，請使用 xtable 套件，印出資料前5筆資料及其結構。

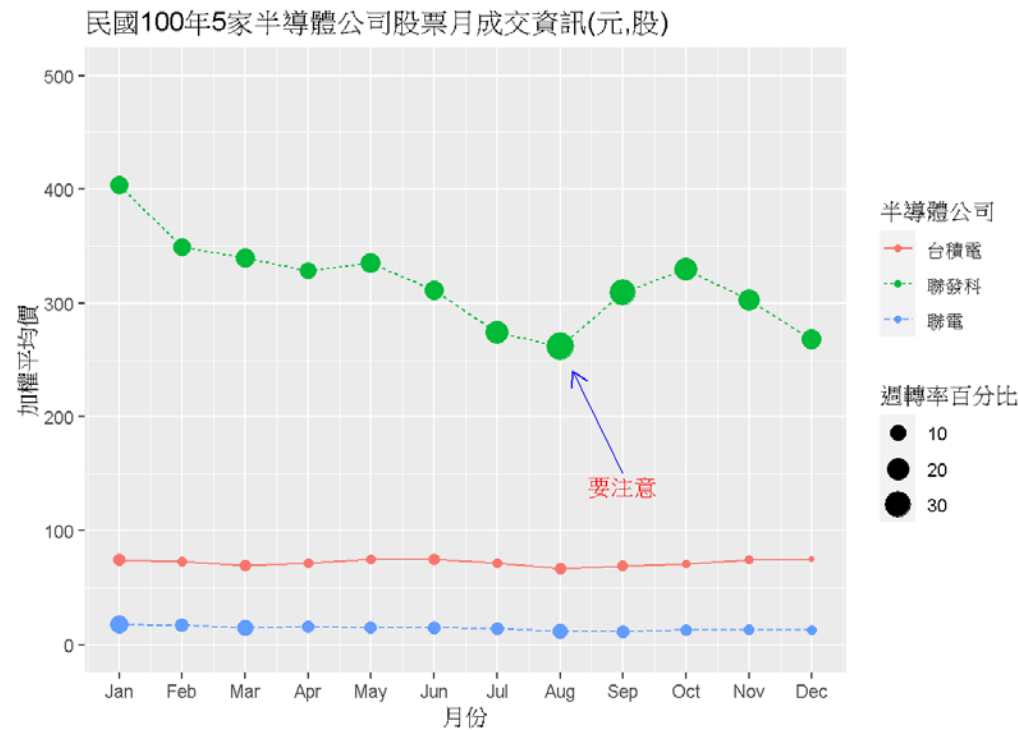
半導體公司	年度	月份	最高價	最低價	加權平均價	成交筆數	成交金額	成交股數	週轉率百分比
1台積電	100	1	78.30	69.60	74.30	263999.00	100578274926.00	1353616348.00	5.22
2台積電	100	2	77.00	69.90	72.54	235159.00	74985055548.00	1033654452.00	3.98
3台積電	100	3	72.20	65.70	69.74	276434.00	88459924495.00	1268289393.00	4.89
4台積電	100	4	73.90	68.00	71.37	211611.00	70177023098.00	983177475.00	3.79
5台積電	100	5	76.90	73.00	74.96	213185.00	74005599560.00	987256484.00	3.80

stockdata data (corrected)

```
## 'data.frame':    60 obs. of  10 variables:
## $ 半導體公司 : chr  "台積電" "台積電" "台積電" "台積電" ...
## $ 年度       : int   100 100 100 100 100 100 100 100 100 100 ...
## $ 月份       : int    1 2 3 4 5 6 7 8 9 10 ...
## $ 最高價     : num   78.3 77 72.2 73.9 76.9 78.2 73.9 72.8 72.1 74 ...
## $ 最低價     : num   69.6 69.9 65.7 68 73 70.4 68.5 62.2 65.9 68.1 ...
## $ 加權平均價 : num   74.3 72.5 69.7 71.4 75 ...
## $ 成交筆數   : num   263999 235159 276434 211611 213185 ...
## $ 成交金額   : num   1.01e+11 7.50e+10 8.85e+10 7.02e+10 7.40e+10 ...
## $ 成交股數   : num   1.35e+09 1.03e+09 1.27e+09 9.83e+08 9.87e+08 ...
## $ 週轉率百分比: num    5.22 3.98 4.89 3.79 3.8 4.99 3.96 4.9 4.14 3.27 ...
```

### 3.3 線圖 (30分)

請繪出以下三家公司之線圖。(要求: (1) *x*軸標記為月份英文簡記。(2) *y*軸的範圍為0~500)



### 4 格式 (5分)

有成功編譯出正確的「學號-姓名-R-MidtermExam.pdf」及「學號-姓名-R-MidtermExam.doc」，並上傳。以下數學式是測試MikTeX/LaTeX，請勿刪。這是常態分佈的機率密度函數：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$