0. 匯入資料

```
# install.packages("tidyverse")
library(tidyverse)

setwd("C:\\Users\\ASUS\\Desktop\\五 234 R\\HW1") salesData <-
read.csv("salesdata.csv")
clientList <- read.csv("client_list.csv")
productList <- read.csv("product_list.csv")
```

1. **product_list** 裡將兩個變數，誤紀錄為在同一個 **column**，其將其分開為兩個變數 **Product**（數字部分）及 **Item**（商品部分），取代原 **product_list**。

```
productList <- productList %>%
    separate("Item", into=c("Product", "Item"), sep = "_")
productList$Product = as.integer(productList$Product)
```

2. 將 **3** 個報表合併為 **full.table**

```
clientList <- clientList %>%
    select(Client, Age, Membership, Gender)

full.table <- salesData %>%
    left_join(clientList, by = "Client") %>%
    left_join(productList, by = "Product")
```

3. 在 **full.table.** 新增一個變數「總消費」為 **spend = UnitPrice*Quantity**

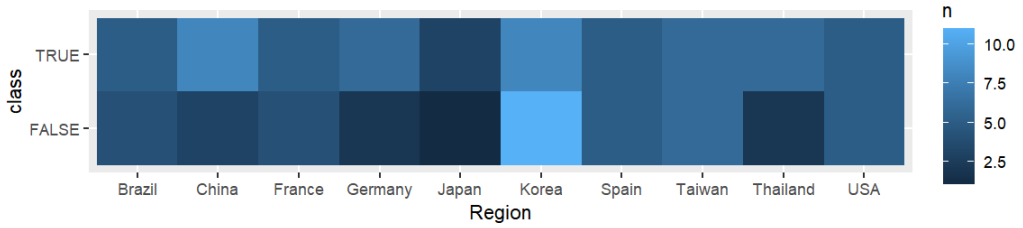
```
full.table <- full.table %>%
  mutate( spend = UnitPrice * Quantity )
```

```
> full.table
    X Store Product Client UnitPrice Quantity   Region Age Membership Gender       Item spend
1   1     A     103      1        10       72   Brazil  36      basic female    MacBook   720
2   2     A     102     13        14        1   France  41    diamond female       iPad    14
3   3     B     104     16        20       64    Korea  50    diamond female       iMac  1280
4   4     B     105      5         4       62      USA  37    diamond female    AirPods   248
5   5     C     104     11        11       97    China  21      basic   male       iMac  1067
6   6     A     105     18         6       14      USA  52      basic female    AirPods    84
7   7     A     105      4        13       74   Taiwan  58     silver   male    AirPods   962
8   8     A     105     12         8       47    Korea  56      basic   male    AirPods   376
9   9     C     103     20         6       76    Spain  21       gold female    MacBook   456
10 10     A     103      2        14       58    China  46    diamond   male    MacBook   812
```

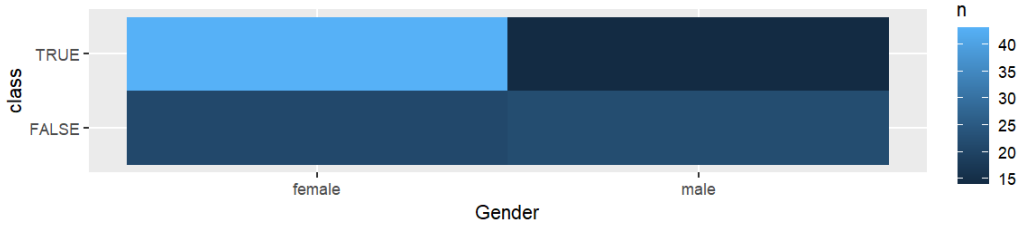
4. 在 **full.table** 將會員等級分組，其中 **gold** 和 **diamond** 的顧客為一組，其他等級的為一組，以敘述統計針對兩組客戶進行比較介紹（例如平均年紀、性別、國家、消費情況差異等）。

```
## class 中 True 代表屬於 High class(diamond/gold)
full.table <- full.table %>%
  mutate( class = (Membership=="gold" | Membership=="diamond") )
```

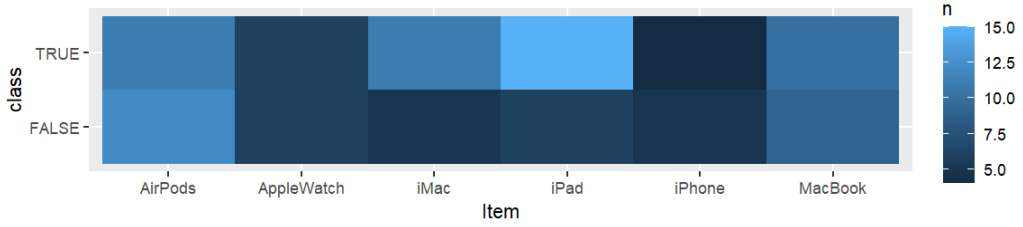
```
## 國家&會員等級
full.table %>%
  count( Region, class ) %>%
  ggplot( mapping = aes(x = Region, y = class) ) +
  geom_tile( mapping = aes(fill = n) )
```




```
## 性別&會員等級
full.table %>%
  count( Gender, class ) %>%
  ggplot( mapping = aes(x = Gender, y = class) ) +
  geom_tile( mapping = aes(fill = n) )
```




```
## 品項&會員等級
full.table %>%
  count( Item, class ) %>%
  ggplot( mapping = aes(x = Item, y = class) ) +
  geom_tile( mapping = aes(fill = n) )
```



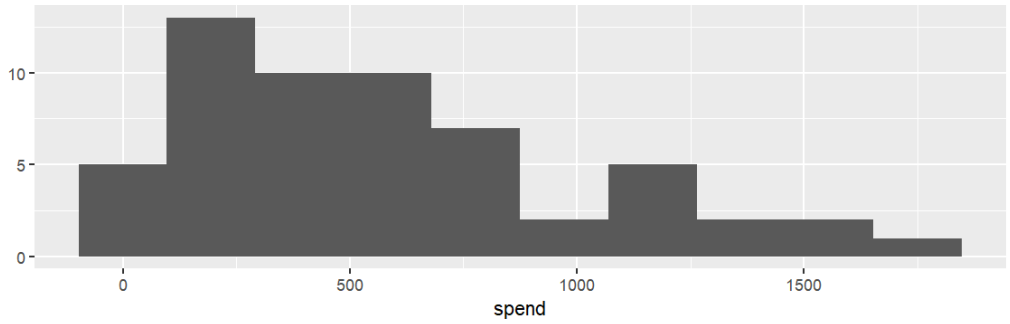

```
## 敘述統計 包含年紀和消費狀況
table1a <- full.table %>%
  subset(class) %>%
  select(UnitPrice, Quantity, Age, spend)

table1b <- full.table %>%
  subset(!class) %>%
  select(UnitPrice, Quantity, Age, spend)
```

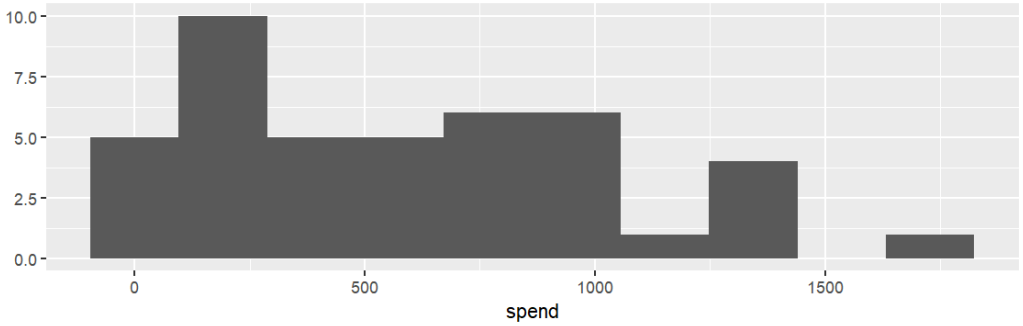
```
> summary(table1a)
   UnitPrice        Quantity          Age             spend
 Min.   : 4.00   Min.   :  1.00   Min.   :21.00   Min.   :  14.0
 1st Qu.: 8.00   1st Qu.: 26.00   1st Qu.:37.00   1st Qu.: 248.0
 Median :11.00   Median : 58.00   Median :41.00   Median : 490.0
 Mean   :11.28   Mean   : 54.74   Mean   :41.82   Mean   : 586.4
 3rd Qu.:15.00   3rd Qu.: 81.00   3rd Qu.:50.00   3rd Qu.: 816.0
 Max.   :20.00   Max.   :100.00   Max.   :54.00   Max.   :1764.0

> summary(table1b)
   UnitPrice        Quantity          Age             spend
 Min.   : 4.00   Min.   : 2.00   Min.   :21.00   Min.   :  36.0
 1st Qu.: 7.50   1st Qu.:23.50   1st Qu.:29.50   1st Qu.: 237.0
 Median :11.00   Median :57.00   Median :39.00   Median : 504.0
 Mean   :11.77   Mean   :52.23   Mean   :40.86   Mean   : 599.2
 3rd Qu.:16.00   3rd Qu.:76.50   3rd Qu.:54.00   3rd Qu.: 901.5
 Max.   :19.00   Max.   :98.00   Max.   :58.00   Max.   :1764.0
```

```
## 總消費&會員等級
table1a %>%
  qplot( spend, geom="histogram", data=., bins=10 )
```




```
table1b %>%
 qplot( spend, geom="histogram", data=., bins=10 )
```



	Gold & Diamond	Others
平均年紀	41	39
性別	絕大部分為女性	男性較多
國家	Korea 和 China 較多	Korea 特別多
消費品項	IPad 最多、IMac 次之	AirPods 為主
平均消費金額	586.4	599.2

5. 在 full.table 針對男性客戶進行分析（例如平均年紀、國家、消費情況等），並對他們在不同產品的「總消費」畫圖分析。

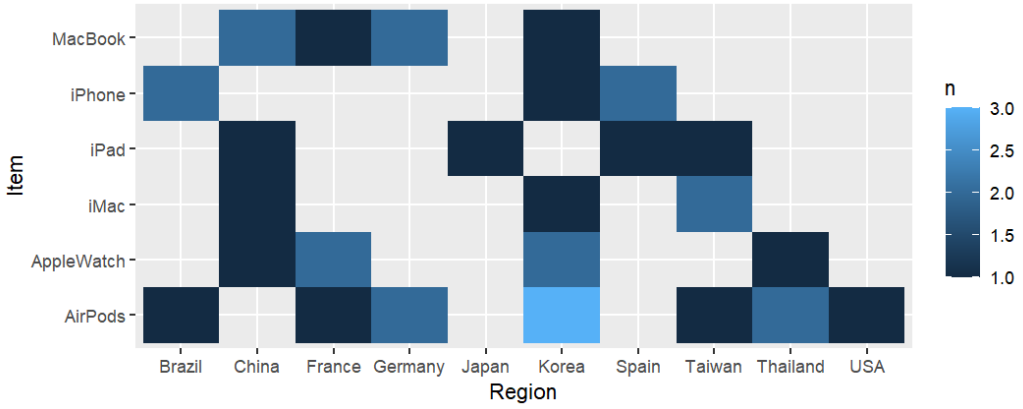
```
table2 <- full.table %>%
    filter(Gender == 'male')

table2$Membership <- factor(table2$Membership, levels =
c("diamond", "gold", "silver", "basic"))
```

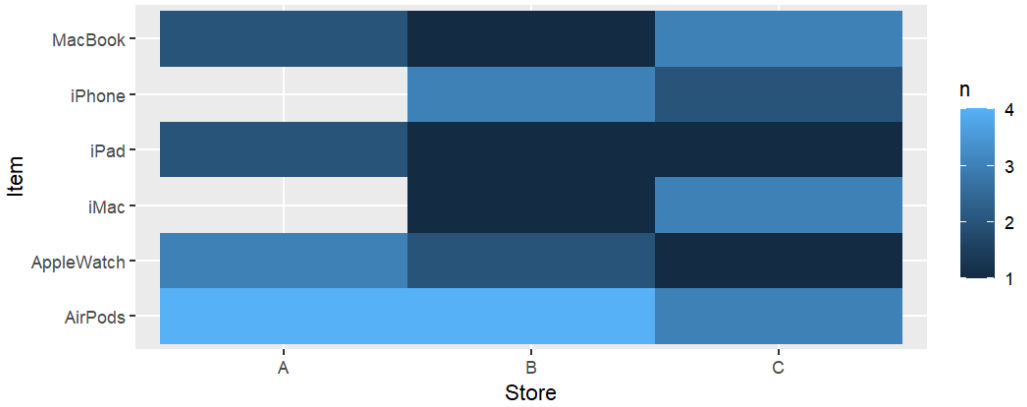
```
## 平均年紀
table2 %>%
   group_by(Item) %>%
   summarize(
      age_mean = mean(Age),
      age_max = max(Age),
      age_min = min(Age))
```

```
# A tibble: 6 × 4
  Item       age_mean age_max age_min
  <chr>         <dbl>   <int>   <int>
1 AirPods        45.1      58      21
2 AppleWatch     44.5      58      21
3 MacBook        41.2      56      26
4 iMac           26.2      37      21
5 iPad           30.2      37      21
6 iPhone         46.4      58      26
```

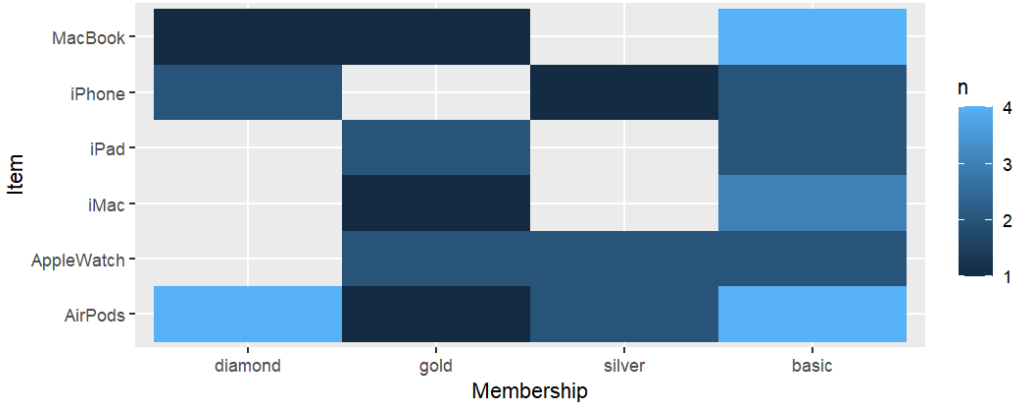
```
## 購買項目&國家
table2 %>%
     count( Region, Item ) %>%
   ggplot( mapping = aes(x = Region, y = Item) ) +
   geom_tile( mapping = aes(fill = n) )
```



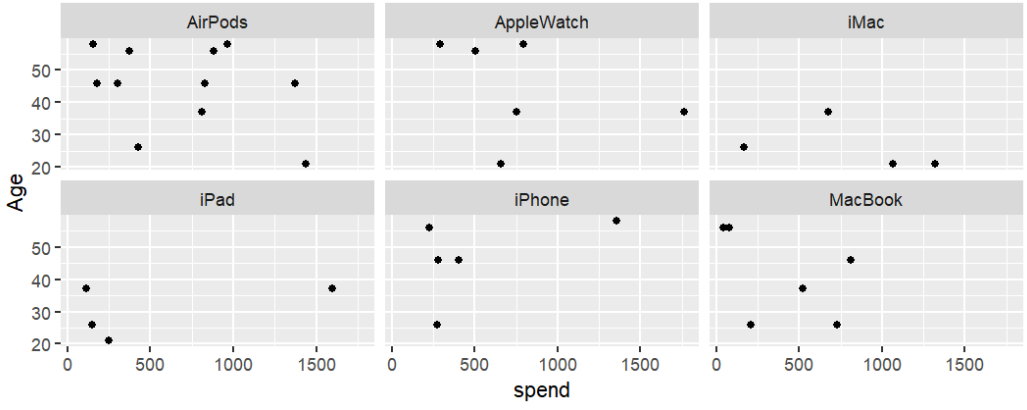

```
## 購買項目&店家
table2 %>%
     count( Store, Item ) %>%
   ggplot( mapping = aes(x = Store, y = Item) ) +
   geom_tile( mapping = aes(fill = n) )
```




```
## 購買項目&會員等級
table2 %>%
    count( Membership, Item ) %>%
    ggplot( mapping = aes(x = Membership, y = Item) ) +
    geom_tile( mapping = aes(fill = n) )
```



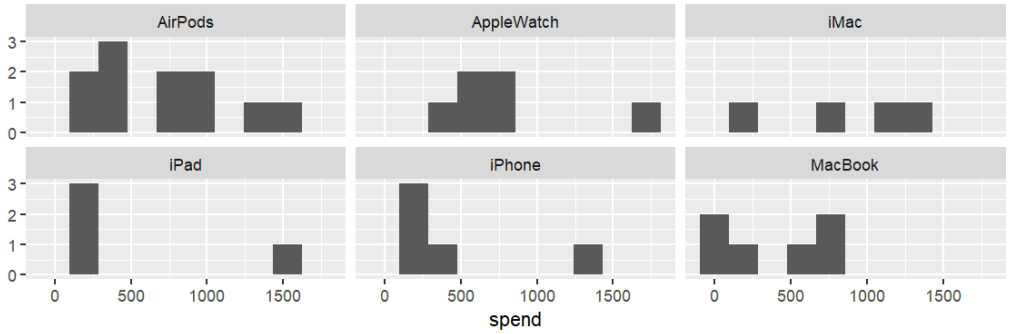

```
## 依購買項目分類 年紀&總消費分布
table2 %>%
    ggplot(aes(x = spend, y = Age)) +
    geom_point() +
    facet_wrap(~Item)
```



```
## 依購買項目分類 統計數量、總額、平均
table2 %>%
    group_by(Item) %>%
    summarize(
        spend_count = n(),
        spend_total = sum(spend),
        spend_mean = mean(spend),
        spend_max = max(spend),
        spend_min = min(spend))
```

```
# A tibble: 6 × 6
  Item       spend_count spend_total spend_mean spend_max spend_min
  <chr>            <int>       <int>      <dbl>     <int>     <int>
1 AirPods             11        7733        703      1440       156
2 AppleWatch           6        4756        793.     1764       288
3 MacBook              6        2393        399.      812        44
4 iMac                 4        3230        808.     1320       168
5 iPad                 4        2105        526.     1598       114
6 iPhone               5        2544        509.     1358       228
```

```
## 依購買項目分類 總消費長方圖
table2 %>%
  qplot( spend, geom="histogram", data=., bins=10 ) +
  facet_wrap(~Item)
```



	AirPods	AppleWatch	iMac	iPad	iPhone	MacBook
數量	11	6	6	4	4	5
平均年紀	45.1	44.5	41.2	26.2	30.2	46.4
國家	集中在 Korea	France、Korea 較多	Tiawan 較多	在 China、Japan、Spain、Taiwan 各一	Brazil、Spain 較多	China、Germany 較多
店家	主要在 A、B	主要在 A	主要在 C	主要在 A	主要在 B	主要在 C
會員等級	主要在 Diamond、Basic	沒有 Diamond	主要在 Basic	Gold、Basic 較多	Diamond、Basic 較多	主要在 Basic
平均消費金額	703	793	399	808	526	509