1. **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)

1. **將3個報表合併為full.table**

clientList <- clientList %>%

select(Client, Age, Membership, Gender)

full.table <- salesData %>%

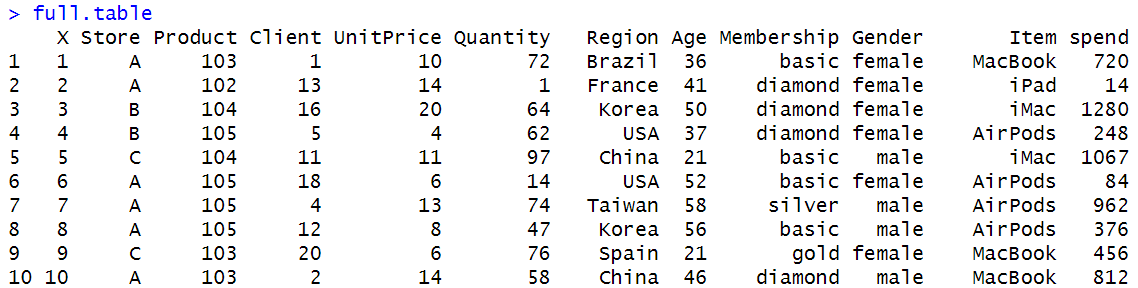
left\_join(clientList, by = "Client") %>%

left\_join(productList, by = "Product")

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

full.table <- full.table %>%

mutate( spend = UnitPrice \* Quantity )



1. **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") )

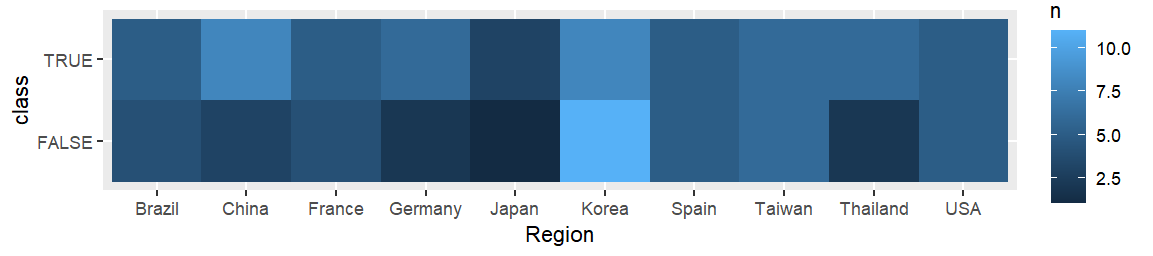
## 國家&會員等級

full.table %>%

count( Region, class ) %>%

ggplot( mapping = aes(x = Region, y = class) ) +

geom\_tile( mapping = aes(fill = n) )



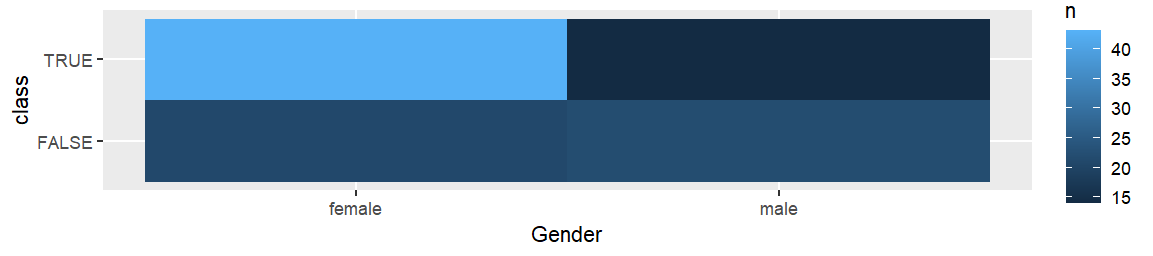
## 性別&會員等級

full.table %>%

count( Gender, class ) %>%

ggplot( mapping = aes(x = Gender, y = class) ) +

geom\_tile( mapping = aes(fill = n) )



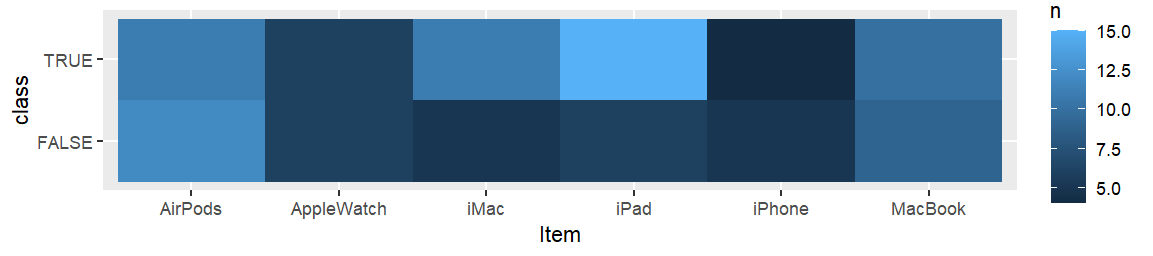
## 品項&會員等級

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 %>%

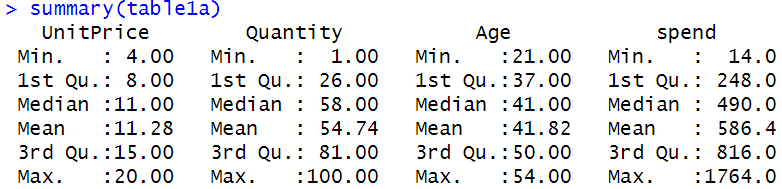
subset(class) %>%

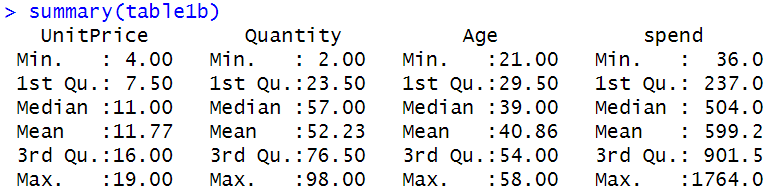
select(UnitPrice, Quantity, Age, spend)

table1b <- full.table %>%

subset(!class) %>%

select(UnitPrice, Quantity, Age, spend)





## 總消費&會員等級

table1a %>%

qplot( spend, geom="histogram", data=., bins=10 )

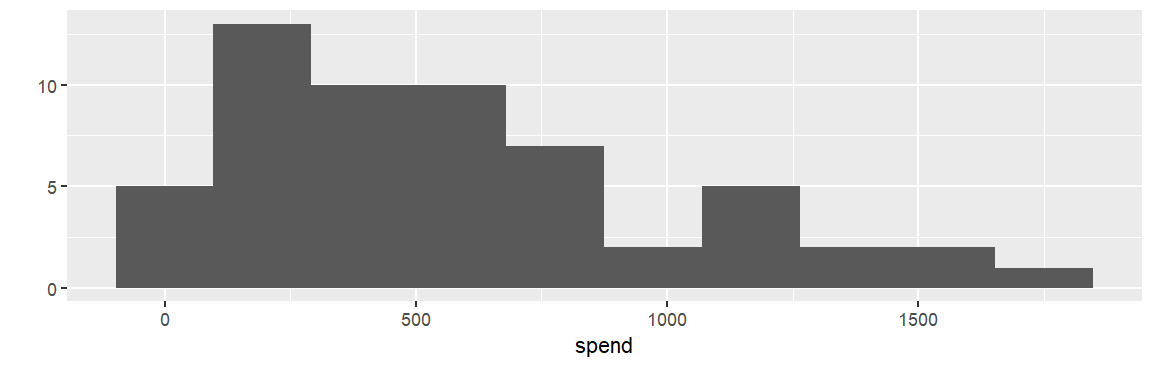
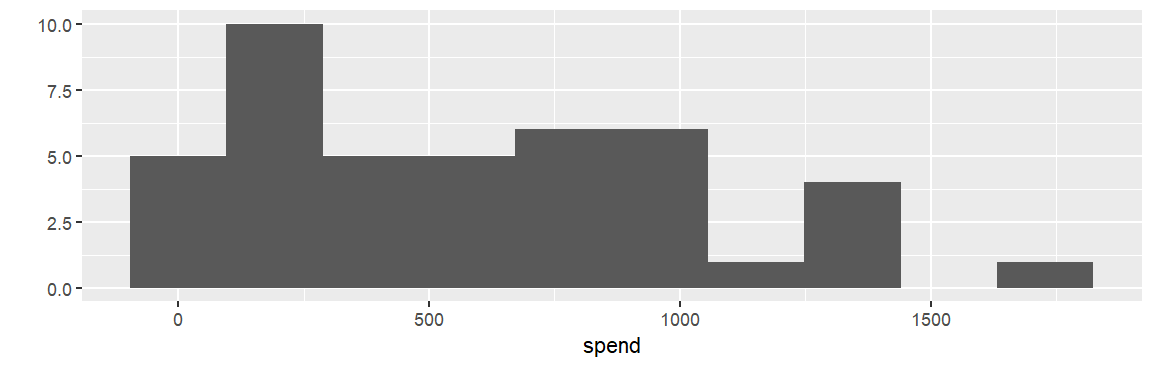


table1b %>%

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"))

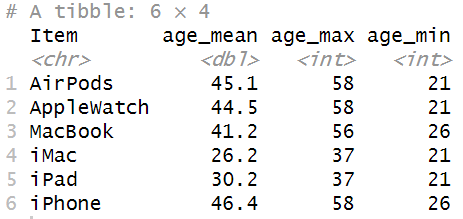
## 平均年紀

table2 %>%

group\_by(Item) %>%

summarize(

age\_mean = mean(Age),

age\_max = max(Age),

age\_min = min(Age))

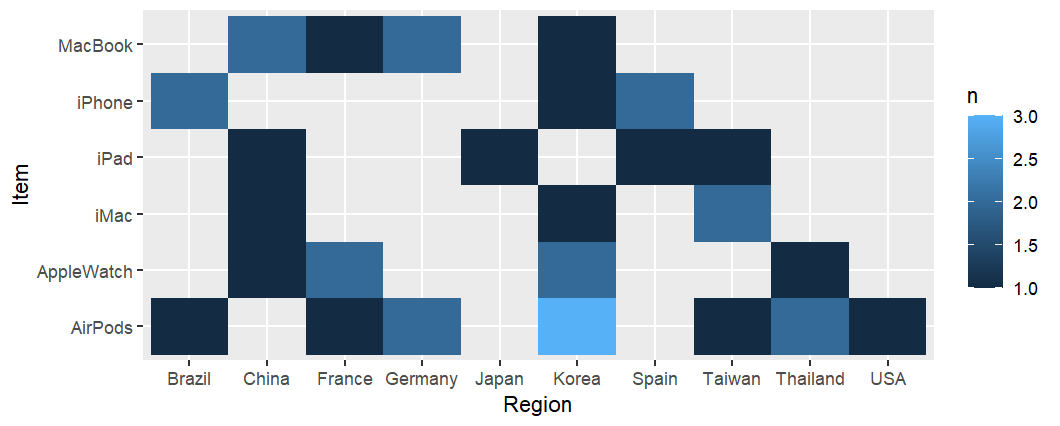
## 購買項目&國家

table2 %>%

count( Region, Item ) %>%

ggplot( mapping = aes(x = Region, y = Item) ) +

geom\_tile( mapping = aes(fill = n) )



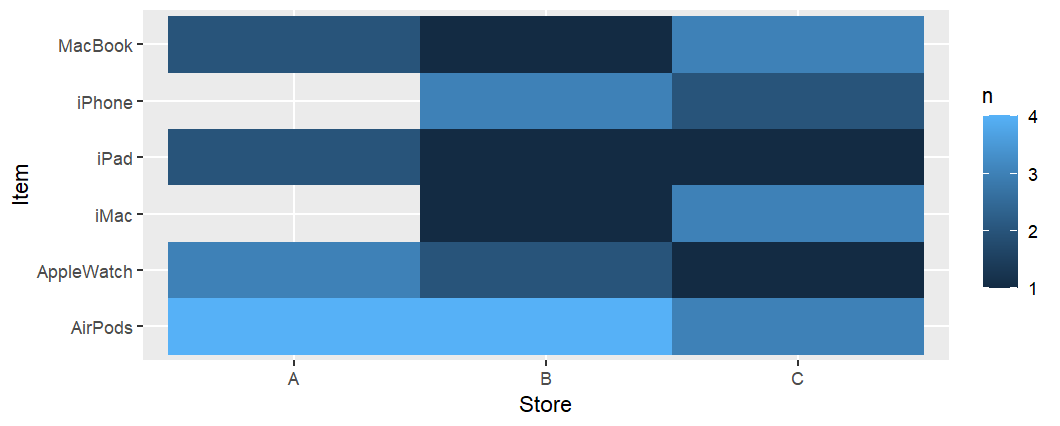
## 購買項目&店家

table2 %>%

count( Store, Item ) %>%

ggplot( mapping = aes(x = Store, y = Item) ) +

geom\_tile( mapping = aes(fill = n) )



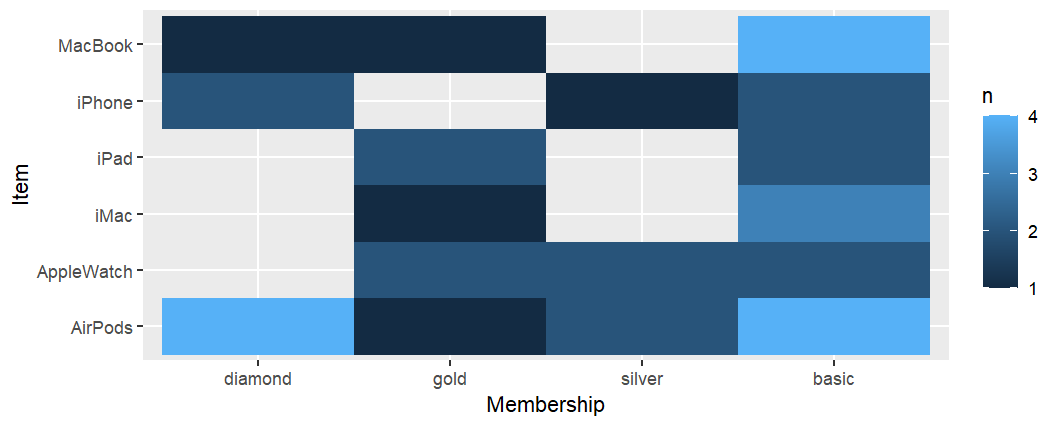
## 購買項目&會員等級

table2 %>%

count( Membership, Item ) %>%

ggplot( mapping = aes(x = Membership, y = Item) ) +

geom\_tile( mapping = aes(fill = n) )



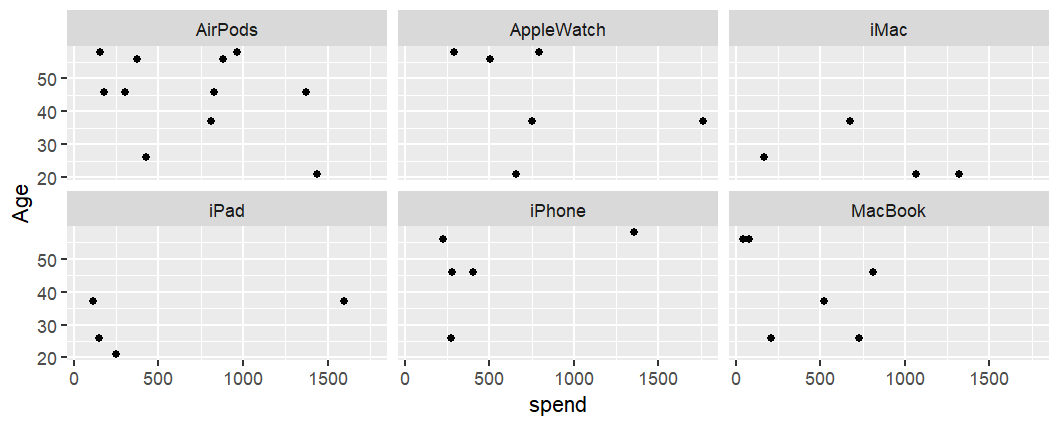
## 依購買項目分類 年紀&總消費分布

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(

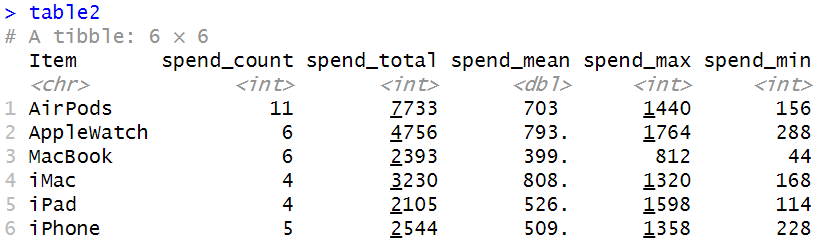
spend\_count = n(),

spend\_total = sum(spend),

spend\_mean = mean(spend),

spend\_max = max(spend),

spend\_min = min(spend))

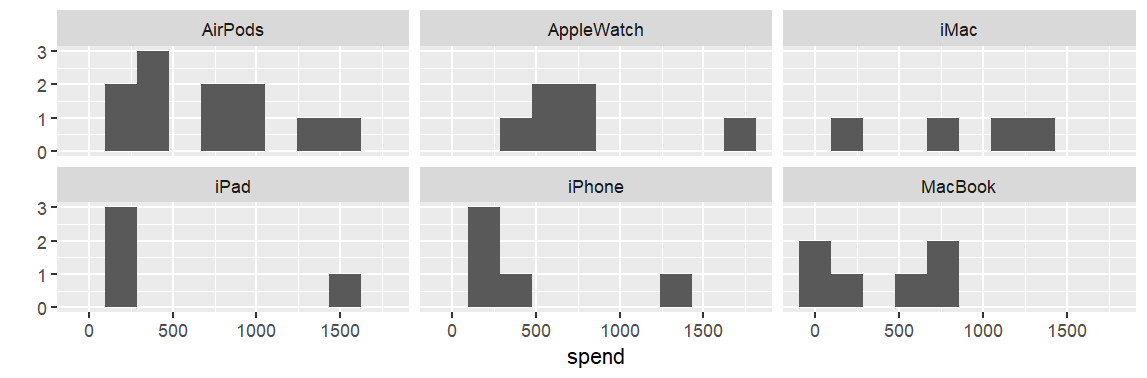


## 依購買項目分類 總消費長方圖

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 |