110306081 資管三甲 蒲依蓮 第一次作業

1.

請讀入資料：stock = read_csv(“stock.csv”)

資料為 2023/12/1~2024/3/7，5 個股票的股價，變數為：

id_name：每檔股票的證券代碼與名稱，如：2330 台積電，表示證券代碼 2330，公司名稱為台積電。

type：open 代表開盤價，close 代表收盤價。

2023/12/1：為該天交易價格（剩下日期變數依此類推）。

請用 tidyr 提到的 gather, spread, sperate 等函數指令，將資料整理成下方「tidy」格式（這裡示意圖只秀出前 18 筆）：

```
> stock <- stock %>% separate(id_name, into = c("id", "name"),sep = " ") %>%
+   gather(year_month_date, money, '2023/12/1':'2024/3/7') %>%
+   separate (year_month_date, into = c("year", "month", "date"),sep = "/") %>%
+   spread(type,money)
> stock
# A tibble: 305 × 7
   id    name  year  month date  close  open
   <chr> <chr> <chr> <chr> <chr> <dbl> <dbl>
 1 2002  中鋼  2023  12    1      26    25.9
 2 2002  中鋼  2023  12    11     25.7  26.0
 3 2002  中鋼  2023  12    12     25.4  25.7
 4 2002  中鋼  2023  12    13     25.2  25.4
 5 2002  中鋼  2023  12    14     25.4  25.3
 6 2002  中鋼  2023  12    15     26.4  25.5
 7 2002  中鋼  2023  12    18     26.6  27.0
 8 2002  中鋼  2023  12    19     26.6  26.6
 9 2002  中鋼  2023  12    20     26.6  26.9
10 2002  中鋼  2023  12    21     26.6  26.4
# ℹ 295 more rows
# ℹ Use `print(n = ...)` to see more rows
```

2. 電商公司，有三個資料集合：

sales.df：產品銷售狀況（“salesID”銷售紀錄編號，“Store”店家編號， “Product”產品編號，“Client”顧客編號，“UnitPrice”單價，“Quantity”購買數量，“Region”顧客國家 ）

client.df：顧客的個人資料 (“Client”顧客編號，“Age”年紀，“Membership”會員等級， “Gender”性別)

prod.df：產品的相關資料（“Item” 代號__產品）

請用 tidyverse 套件裡學到的方法，分析

##1. prod.df 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 prod.df。

```
> ##2.1
> prod.df<- prod.df %>% separate(Item, into = c("Product", "Item"), sep = "_", convert = TRUE)
> prod.df
# A tibble: 6 × 2
  Product Item
    <int> <chr>
1     101 iPhone
2     102 iPad
3     103 MacBook
4     104 iMac
5     105 AirPods
6     106 AppleWatch
```

##2. 將 3 個報表合併為 full.table

```
> full.table<- sales.df %>%
+   left_join(prod.df, by = "Product") %>%
+   left_join(client.df, by="Client", suffix=c("",".y"))%>%
+   select(-ends_with(".y"))
>
> full.table <-na.omit(full.table)
> full.table
# A tibble: 100 × 11
    ...1 Store Product Client UnitPrice Quantity Region Item      Age Membership Gender
   <dbl> <chr>   <dbl>  <dbl>     <dbl>    <dbl> <chr>  <chr>   <dbl> <chr>      <chr>
 1     1 A         103      1        10       72 Brazil MacBook    36 basic      female
 2     2 A         102     13        14        1 France iPad       41 diamond    female
 3     3 B         104     16        20       64 Korea  iMac      50 diamond    female
 4     4 B         105      5         4       62 USA    AirPods   37 diamond    female
 5     5 C         104     11        11       97 China  iMac      21 basic      male
 6     6 A         105     18         6       14 USA    AirPods   52 basic      female
 7     7 A         105      4        13       74 Taiwan AirPods   58 silver     male
 8     8 A         105     12         8       47 Korea  AirPods   56 basic      male
 9     9 C         103     20         6       76 Spain  MacBook   21 gold       female
10    10 A         103      2        14       58 China  MacBook   46 diamond    male
# ℹ 90 more rows
# ℹ Use `print(n = ...)` to see more rows
```

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

```
> ##2.3
> full.table<-full.table%>%
+   mutate(spend = UnitPrice*Quantity)
> full.table
# A tibble: 100 × 12
    ...1 Store Product Client UnitPrice Quantity Region Item      Age Membership Gender spend
   <dbl> <chr>   <dbl>  <dbl>     <dbl>    <dbl> <chr>  <chr>   <dbl> <chr>      <chr>  <dbl>
 1     1 A         103      1        10       72 Brazil MacBook    36 basic      female   720
 2     2 A         102     13        14        1 France iPad       41 diamond    female    14
 3     3 B         104     16        20       64 Korea  iMac      50 diamond    female  1280
 4     4 B         105      5         4       62 USA    AirPods   37 diamond    female   248
 5     5 C         104     11        11       97 China  iMac      21 basic      male    1067
 6     6 A         105     18         6       14 USA    AirPods   52 basic      female    84
 7     7 A         105      4        13       74 Taiwan AirPods   58 silver     male     962
 8     8 A         105     12         8       47 Korea  AirPods   56 basic      male     376
 9     9 C         103     20         6       76 Spain  MacBook   21 gold       female   456
10    10 A         103      2        14       58 China  MacBook   46 diamond    male     812
# ℹ 90 more rows
# ℹ Use `print(n = ...)` to see more rows
```

##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，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）。

```
> full.table$Group <- ifelse(full.table$Membership %in% c("gold", "diamond"), "Gold/Diamond", "Other")
> full.table%>%
+   group_by(Group)%>%
+   distinct(Client)%>%
+   summarize(n())
# A tibble: 2 × 2
  Group        `n()`
  <chr>        <int>
1 Gold/Diamond    11
2 Other            8
```

=> Gold/Diamond 組的會員有 11 人，其他有 8 人。

110306081 資管三甲 蒲依蓮 第一次作業

a. (年齡)

```
> full.table%>%
+   group_by(Group)%>%
+   distinct(Client,Age)%>%
+   summarize(avg_age = mean(Age))
# A tibble: 2 × 2
  Group         avg_age
  <chr>           <dbl>
1 Gold/Diamond     39.5
2 Other            40.1
```

=> 以平均年齡而言，Gold/Diamond 組有較高的平均年齡為 39.5 歲，而 Other 的平均年齡為 40.1 歲。

b.(性別)

```
> full.table%>%
+   group_by(Group, Gender)%>%
+   distinct(Client)%>%
+   summarize(Number_of_people = n())
# A tibble: 4 × 3
# Groups:   Group [2]
  Group        Gender Number_of_people
  <chr>        <chr>              <int>
1 Gold/Diamond female                9
2 Gold/Diamond male                  2
3 Other        female                4
4 Other        male                  4
```

=> 以性別人數而言，Gold/Diamond 組有 9 位女性、2 位男性，而 Other 組則有 4 位女性、4 位男性。

c.(國家)

```
> full.table%>%
+   group_by(Group, Region)%>%
+   summarize(Number_of_people = n())
# A tibble: 20 × 3
# Groups:   Group [2]
   Group        Region   Number_of_people
   <chr>        <chr>               <int>
 1 Gold/Diamond Brazil                  5
 2 Gold/Diamond China                   8
 3 Gold/Diamond France                  5
 4 Gold/Diamond Germany                 6
 5 Gold/Diamond Japan                   3
 6 Gold/Diamond Korea                   8
 7 Gold/Diamond Spain                   5
 8 Gold/Diamond Taiwan                  6
 9 Gold/Diamond Thailand                6
10 Gold/Diamond USA                     5
11 Other        Brazil                  4
12 Other        China                   3
13 Other        France                  4
14 Other        Germany                 2
15 Other        Japan                   1
16 Other        Korea                  11
17 Other        Spain                   5
18 Other        Taiwan                  6
19 Other        Thailand                2
20 Other        USA                     5
```

=> 以國家人數而言，Gold/Diamond 組最多消費的地區在中國(8 人)和韓國(8 人)，最少的地區在日本(3 人)，Other 組最多的地區在韓國(11 人)，最少的地區在日本(1 人)。

d.(消費情況)

```
> full.table%>%
+   group_by(Group)%>%
+   summarize(total_spend = sum(spend))
# A tibble: 2 × 2
  Group        total_spend
  <chr>              <dbl>
1 Gold/Diamond       33427
2 Other              25767
```

=> 以總消費情況而言 Gold/Diamond 組有較高的總花費，Gold/Diamond 組共花費 33,427 元，Other 組共花費 25,767 元。

```
> full.table%>%
+   group_by(Group)%>%
+   summarize(mean_spend = mean(spend))
# A tibble: 2 × 2
  Group        mean_spend
  <chr>             <dbl>
1 Gold/Diamond       586.
2 Other              599.
```

=> 以平均消費情況而言 Other 組有較高的平均花費，Gold/Diamond 組平均花費 586 元，Other 組平均花費 599 元。

##5. 在 full.table 針對女性客戶進行分析（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。

```
> f.table<-full.table%>%
+   filter(Gender %in% "female")
> f.table
# A tibble: 64 × 13
    ...1 Store Product Client UnitPrice Quantity Region  Item         Age Membership Gender spend Group
   <dbl> <chr>   <dbl>  <dbl>     <dbl>    <dbl> <chr>   <chr>      <dbl> <chr>      <chr>  <dbl> <chr>
 1     1 A         103      1        10       72 Brazil  MacBook       36 basic      female   720 Other
 2     2 A         102     13        14        1 France  iPad          41 diamond    female    14 Gold/Diamond
 3     3 B         104     16        20       64 Korea   iMac          50 diamond    female  1280 Gold/Diamond
 4     4 B         105      5         4       62 USA     AirPods       37 diamond    female   248 Gold/Diamond
 5     6 A         105     18         6       14 USA     AirPods       52 basic      female    84 Other
 6     9 C         103     20         6       76 Spain   MacBook       21 gold       female   456 Gold/Diamond
 7    12 C         104      8        15      100 Spain   iMac          54 gold       female  1500 Gold/Diamond
 8    14 B         102      9        13       71 USA     iPad          33 silver     female   923 Other
 9    15 C         103     14        15       25 Germany MacBook       50 diamond    female   375 Gold/Diamond
10    17 B         106      5         8       13 USA     AppleWatch    37 diamond    female   104 Gold/Diamond
# ℹ 54 more rows
# ℹ Use `print(n = ...)` to see more rows
```

110306081 資管三甲 蒲依蓮 第一次作業

a.

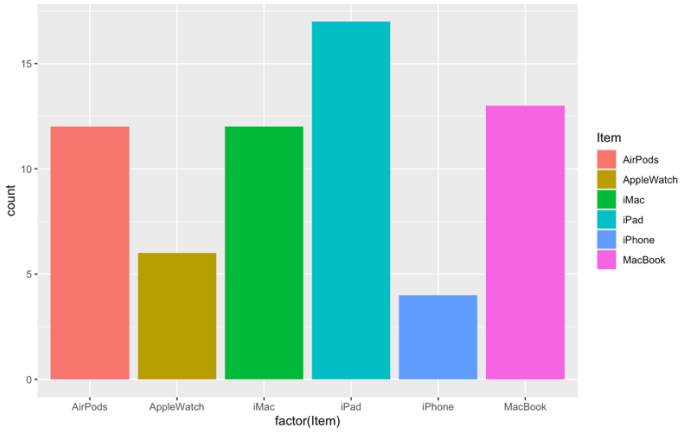
```
> f.table%>%
+   group_by(Group)%>%
+   distinct(Client)%>%
+   summarise(Number_of_people = n())
# A tibble: 2 × 2
  Group         Number_of_people
  <chr>                    <int>
1 Gold/Diamond                 9
2 Other                        4
```

=>女性會員數而言，Gold/Diamond 組有 9 位女會員，Other 組有 4 位女會員。因此可知，Gold/Diamond 組有較多女會員。

b.

```
> table(f.table$Item)

   AirPods AppleWatch       iMac       iPad     iPhone    MacBook
        12          6         12         17          4         13
```



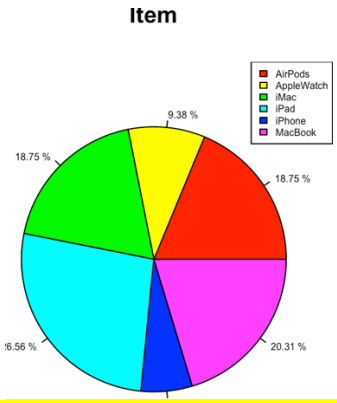
=>以消費數量而言，全部女性會員消費最多的產品為 iPad，共 17 台，最少為 iPhone，共 4 台。

c.

```
> f.table%>%
+   group_by(Group)%>%
+   distinct(Client,Age)%>%
+   summarise(Average_age = sum(Age)/n())
# A tibble: 2 × 2
  Group         Average_age
  <chr>               <dbl>
1 Gold/Diamond         39.1
2 Other                40
```

=> 以平均年齡而言，Gold/Diamond 組為 39.1 歲，低於 Other 組的 40 歲。

d.



=>以各產品的總銷售比例而言，由高到低為 ，最高 iPad 佔總體的 26.56%，其次為 Macbook 佔 20.31%，AirPods 和 iMac 並列第三，各佔 18.75%，AppleWatch 佔 9.38%，最後為 iPhone，佔整體的 6.25%。

e.

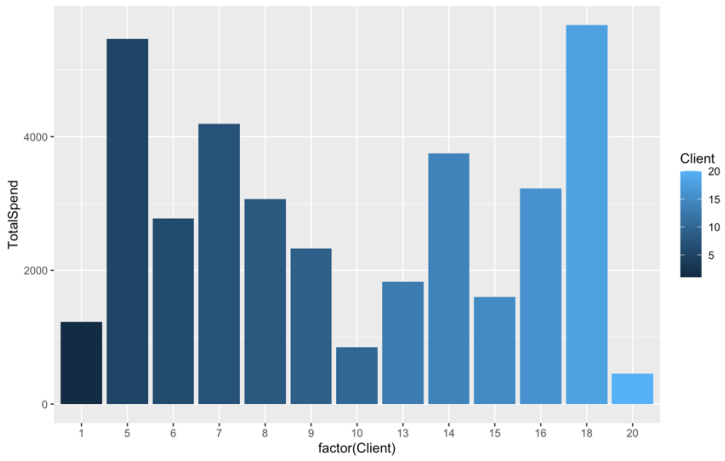
```
> table(f.table$Region)

   Brazil     China    France   Germany     Japan     Korea     Spain    Taiwan  Thailand       USA 
        6         6         5         4         3        11         7         8         5         9 
```

=> 以地區而言，韓國佔全體最高，日本佔最低。

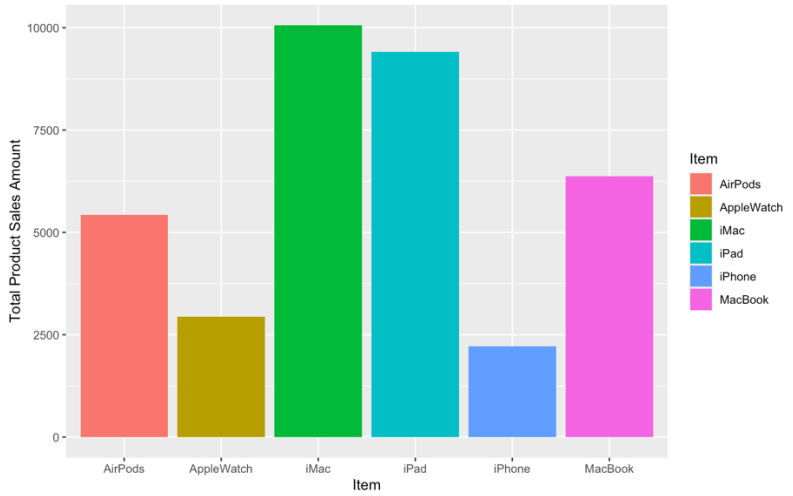
f.

```
> TotalSales <- f.table %>%
+   group_by(Client) %>%
+   summarise(TotalSpend = sum(spend)) %>%
+   arrange(desc(TotalSpend))
> TotalSales
# A tibble: 13 × 2
   Client TotalSpend
    <dbl>      <dbl>
 1     18       5671
 2      5       5465
 3      7       4189
 4     14       3747
 5     16       3224
 6      8       3065
 7      6       2779
 8      9       2325
 9     13       1830
10     15       1599
11      1       1232
12     10        851
13     20        456
```



=> 以上為各個女顧客的總花費金額，可知 18 號女顧客花費最多。

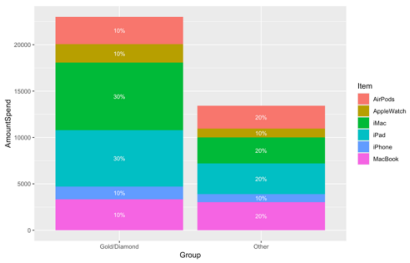
g.



=>iMac 有最高的銷售金額。

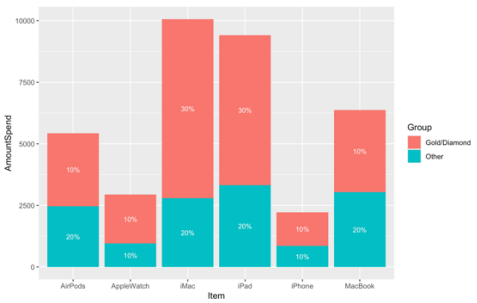
h.

```
> Product
# A tibble: 12 × 4
# Groups:   Group [2]
   Group        Item       AmountSpend Proportion
   <chr>        <chr>            <dbl>      <dbl>
 1 Gold/Diamond AirPods           2963         10
 2 Gold/Diamond AppleWatch        1992         10
 3 Gold/Diamond MacBook           3336         10
 4 Gold/Diamond iMac              7273         30
 5 Gold/Diamond iPad              6090         30
 6 Gold/Diamond iPhone            1362         10
 7 Other        AirPods           2468         20
 8 Other        AppleWatch         950         10
 9 Other        MacBook           3033         20
10 Other        iMac              2790         20
11 Other        iPad              3319         20
12 Other        iPhone             857         10
```



=> 以總消費情況而言 Gold/Diamond 組有較高的總花費，Gold/Diamond 其中最高花費的比例落在 iMac。

i.



=> 以產品消費情況而言，iMac 有最高的銷售額，Gold/Diamond 在 iMac 花費比 Other 多。