1.

Code:

```
sales.df <- read.csv("salesdata.csv")
client.df <- read.csv("client_list.csv")
prod.df <- read.csv("product_list.csv")
library(tidyverse)
prod.df <- prod.df %>% separate(Item, into=c("Product", "Item"), sep = "_")
```

	Product	Item
1	101	iPhone
2	102	iPad
3	103	MacBook
4	104	iMac
5	105	AirPods
6	106	AppleWatch

2.

#檢查資料型態發現 sales.df 之 Product 型態為 int、prod.df 之 Product 型態為 chr 無法 join，所以將後者轉成 int

Code:

```
str(sales.df)
str(client.df)
str(prod.df)
```

```
> str(sales.df)
'data.frame':   39 obs. of  7 variables:
 $ salesID  : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Store    : chr  "B" "B" "A" "B" ...
 $ Product  : int  105 106 103 102 101 102 103 104 105 106 ...
 $ Client   : int  1 1 1 1 2 2 2 2 2 2 ...
 $ UnitPrice: int  4 5 4 6 4 5 6 7 8 9 ...
 $ Quantity : int  13 10 11 2 44 3 8 4 6 10 ...
 $ Region   : chr  "Taiwan" "Taiwan" "Taiwan" "Taiwan" ...
> str(client.df)
'data.frame':   10 obs. of  4 variables:
 $ Client    : int  1 2 3 4 5 6 7 8 9 10
 $ Age       : int  22 35 33 33 52 21 25 26 40 19
 $ Membership: chr  "basic" "silver" "gold" "diamond" ...
 $ Gender    : chr  "male" "female" "male" "female" ...
> str(prod.df)
'data.frame':   6 obs. of  2 variables:
 $ Product: chr  "101" "102" "103" "104" ...
 $ Item   : chr  "iPhone" "iPad" "MacBook" "iMac" ...
```

prod.df$Product <-as.integer(prod.df$Product)

full.table <- sales.df %>% inner_join(client.df) %>% inner_join(prod.df)

	salesID	Store	Product	Client	UnitPrice	Quantity	Region	Age	Membership	Gender	Item
1	1	B	105	1	4	13	Taiwan	22	basic	male	AirPods
2	2	B	106	1	5	10	Taiwan	22	basic	male	AppleWatch
3	3	A	103	1	4	11	Taiwan	22	basic	male	MacBook
4	4	B	102	1	6	2	Taiwan	22	basic	male	iPad
5	5	A	101	2	4	44	USA	35	silver	female	iPhone
6	6	A	102	2	5	3	USA	35	silver	female	iPad

3.

Code:

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

	salesID	Store	Product	Client	UnitPrice	Quantity	Region	Age	Membership	Gender	Item	spend
1	1	B	105	1	4	13	Taiwan	22	basic	male	AirPods	52
2	2	B	106	1	5	10	Taiwan	22	basic	male	AppleWatch	50
3	3	A	103	1	4	11	Taiwan	22	basic	male	MacBook	44
4	4	B	102	1	6	2	Taiwan	22	basic	male	iPad	12
5	5	A	101	2	4	44	USA	35	silver	female	iPhone	176
6	6	A	102	2	5	3	USA	35	silver	female	iPad	15

4.

(group1 為 gold&diamond，group2 為 others)

#平均年紀：group2 較 group1 高

#性別:group1 的 female 較 male 多出約 2 倍，group2 男女性別筆數相同

#國家:兩組內的國家都恰好不同且資料幾乎是平均分佈，除了 group2 內的 USA(6 筆)明顯較 China(2 筆)多

#消費情況差異：兩組直方圖相比，group1 明顯消費較多 iPhone，group2 則明顯消費較多 iPad 及較少 iMac

Code:

```
group1 <- full.table %>% filter(Membership=="gold" | Membership=="diamond")
group2 <- full.table %>% filter(Membership!="gold" & Membership!="diamond")
mean(group1$Age)
mean(group2$Age)
```

```
> mean(group1$Age)
[1] 27.31579
> mean(group2$Age)
[1] 32.3
>
```

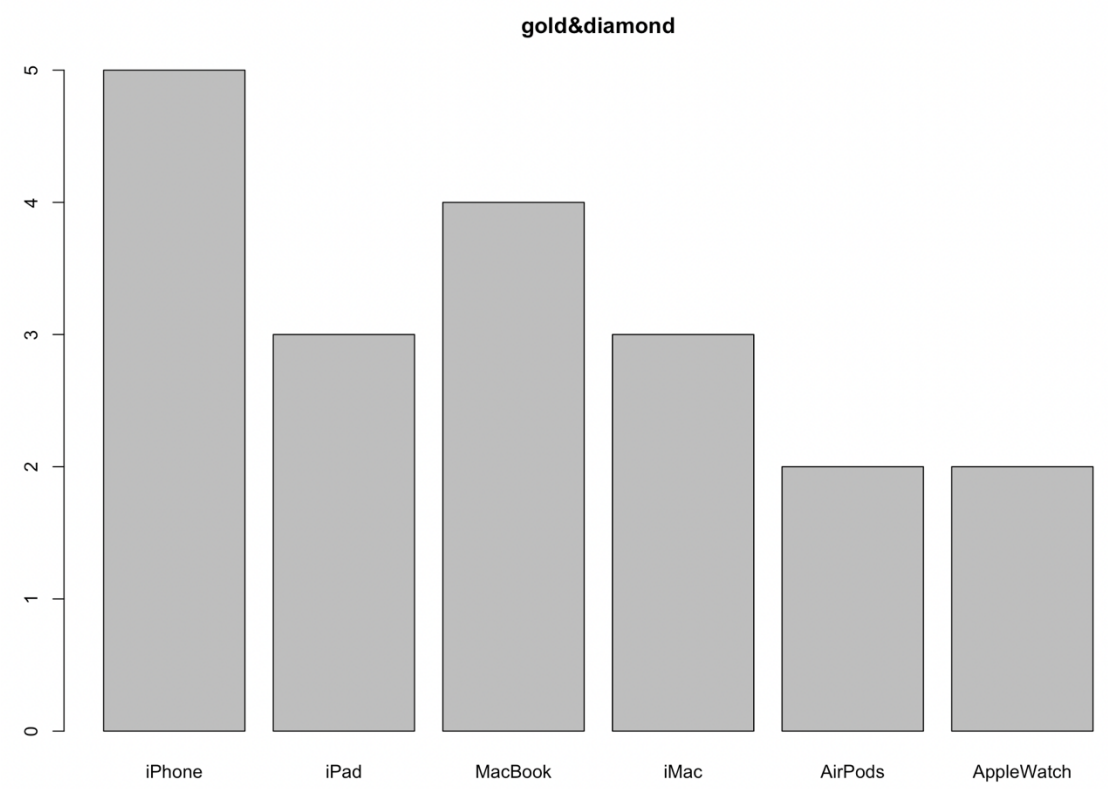
```
table(group1$Gender)
table(group2$Gender)
```

```
> table(group1$Gender)

female   male
    13      6
> table(group2$Gender)

female   male
    10     10
```

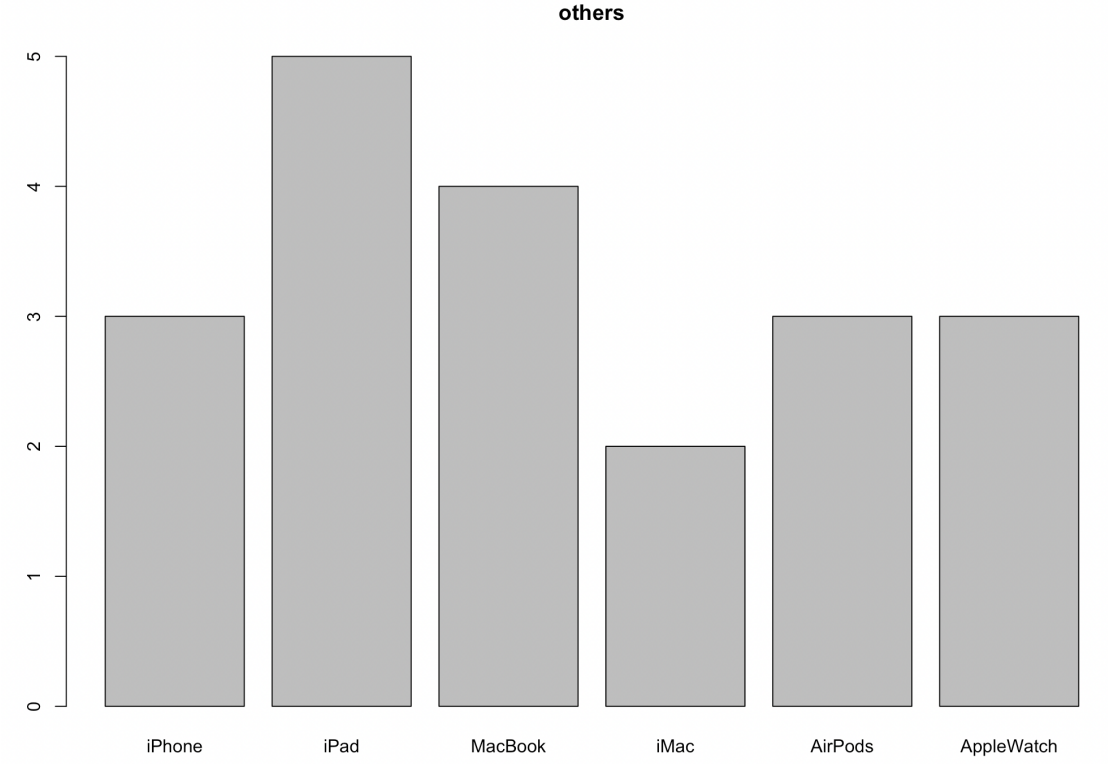
```
table(group1$Region)
table(group2$Region)
```

```
> table(group1$Region)

   Brazil    France     Korea     Spain Thailand
        3         4         5         4         3
> table(group2$Region)

   China Germany   Japan  Taiwan     USA
       2       4       4       4       6
```

```
barplot(table(group1$Product), names.arg = c("iPhone","iPad",
"MacBook","iMac","AirPods","AppleWatch"), main = "gold&diamond")
```



barplot(table(group2$Product), names.arg = c("iPhone","iPad", "MacBook","iMac","AirPods","AppleWatch"), main = "others")



5.

#平均年紀：32.875

#國家：Germany、Taiwan 有 4 筆，Brazil、Thailand 有 3 筆，China 有 2 筆
#消費情況：多數為 102(5 筆)和 101、103(4 筆)，少數則有 105(2 筆)和 106(1 筆)，104 則完全沒有消費
#不同產品的「總消費」：iPad 總消費最多，AppleWatch 總消費非常少，另外完全沒有 iMac 的消費
Code:

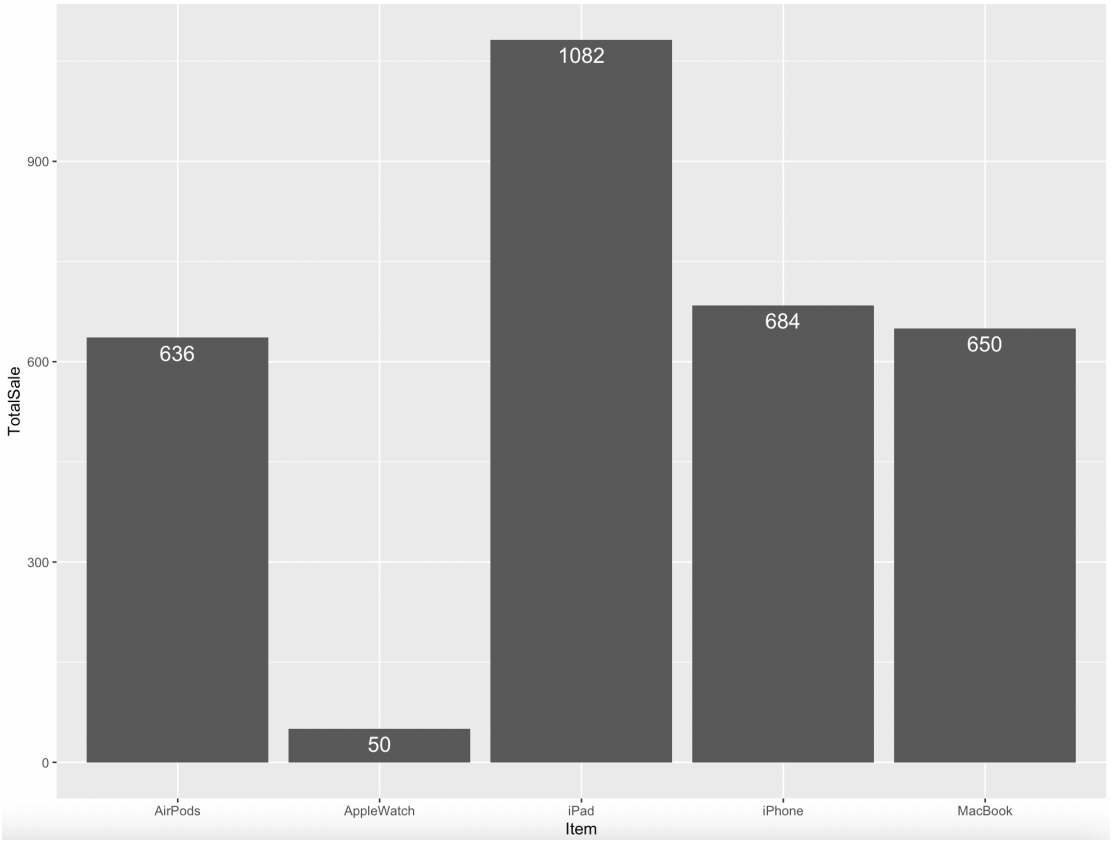
```
male.table <- full.table %>% filter(Gender=="male")
mean(male.table$Age)
table(male.table$Region)
```

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

   Brazil     China   Germany    Taiwan Thailand
        3         2         4         4         3
```

```
table(male.table$Product)
```

```
> table(male.table$Product)

101 102 103 105 106
  4   5   4   2   1
```

```
spend_item <-
   male.table %>%
   group_by(Item) %>%
   summarise(TotalSale=sum(spend))
spend_item %>%
   ggplot(aes(x=Item, y=TotalSale)) +
   geom_bar(stat="identity") +
   geom_text(aes(label=TotalSale), vjust=1.5, color="white",size=5)
```


1082
900
636
684
650
600
TotalSale
300
50
0
AirPods
AppleWatch
iPad
iPhone
MacBook
Item