```
library(tidyverse)
library(readr)

#一、

watch.table <- read_csv("watch_table.csv")
user.table <- read_csv("user_table.csv")
drama.table <- read_csv("drama_table.csv")

## 1. 將 watch.table 與其他兩個報表合併為full.table
full.table <- watch.table %>%
  left_join(user.table, by = "user_id") %>%
  left_join(drama.table, by = "drama_id")

full.table

## 2. 分析full.table，計算每部劇男生、女生觀看次數
full.table %>%
  group_by(drama_name) %>%
  summarise(female_number = length(which(gender == "female")),
            male_number = length(which(gender == "male")))

## 3. 找出用Android系統的，針對這類客戶進行分析。
full.table %>%
  filter(device == "Android") %>%
  summarise(Avg_age = mean(age),
            total_number = n())

full.table %>%
  group_by(gender) %>%
  filter(device == "Android") %>%
  summarise(gender_distribute = n())

full.table %>%
  group_by(drama_name) %>%
  filter(device == "Android") %>%
  summarise(drama_distribute = n())

full.table %>%
  group_by(location) %>%
  filter(device == "Android") %>%
  summarise(location_distribute = n())

## 4. 針對台北男性這類客戶進行分析。
full.table %>%
  filter(location == "Taipei" & gender == "male") %>%
  summarise(Avg_age = mean(age),
            total_number = n())

full.table %>%
  group_by(drama_name) %>%
  filter(location == "Taipei" & gender == "male") %>%
  summarise(drama_distribute = n())
```

```
full.table %>%
  group_by(device) %>%
  filter(location == "Taipei" & gender == "male") %>%
  summarise(device_distribute = n())


#二、

abnyc.table <- read.csv("AB_NYC_2019.csv")

## 1. 找出 neighbourhood_group == "Manhattan"的資料，利用ggplot畫經緯度的
scatter plot。
abnyc.table %>%
  filter(neighbourhood_group == "Manhattan") %>%
  ggplot(aes(x=latitude, y=longitude, color=room_type)) +
  geom_point()

## 2. 針對曼哈頓資料，對number_of_reviews >=400的畫bar chart。
abnyc.table %>%
  filter(neighbourhood_group == "Manhattan" & number_of_reviews >= 400) %>%
  ggplot(aes(x=number_of_reviews)) +
  geom_bar(fill="blue")

## 3. 針對曼哈頓資料，number_of_reviews >=400的中，哪個neighbourhood擁有最多
number_of_reviews。
k <- abnyc.table %>%
  filter(neighbourhood_group == "Manhattan" & number_of_reviews >= 400)
i_max <- which.max(k$number_of_reviews)
k$neighbourhood[i_max]

## 4.建立一筆新資料，將3.找出的neighbourhood篩選出來，去除掉NA值後，進行EDA分析，並簡
單介紹最高房價及最低房價分別的類型。
new.table <- na.omit(abnyc.table %>%
  filter(neighbourhood == "Harlem"))

new.table %>%
  group_by(room_type) %>%
  arrange(desc(price)) %>%
  ggplot(aes(x=room_type, y=price, fill=room_type)) +
  geom_bar(stat = "identity")
```