

國立高雄大學

課程： R分析應用

106學年度第二學期

期中報告: 以R語言分析台灣觀光產業

學生：A1043338 郭上暐

W1063217 許秀平

指導教授： 曾智義 教授

民國 106年 6月 21日

**目錄**

[壹、專題介紹 1](#_Toc517684643)

[一、議題背景 1](#_Toc517684644)

[二、資料建模 1](#_Toc517684645)

[三、欲探討與應用之問題 1](#_Toc517684646)

[貳、章節之應用 2](#_Toc517684647)

[一、第三章之運用 2](#_Toc517684648)

[二、第四章之運用 3](#_Toc517684649)

[三、第五章使用 3](#_Toc517684650)

[四、第六章之運用 4](#_Toc517684651)

[五、第七章之運用 4](#_Toc517684652)

[六、第八、九章之運用 6](#_Toc517684653)

[七、第十一章之運用 7](#_Toc517684654)

[八、第十二章之運用 7](#_Toc517684655)

[九、第十三章運用 9](#_Toc517684656)

[十、第十五章運用 10](#_Toc517684657)

[參、結論與心得 12](#_Toc517684658)

**圖目錄**

[圖 1 建模範例 1](#_Toc517685673)

[圖 2 readxl安裝程式碼 2](#_Toc517685674)

[圖 3 ggplot2套件安裝程式碼 3](#_Toc517685675)

[圖 4 lubridate套件安裝程式碼 3](#_Toc517685676)

[圖 5 Data.frame建構程式碼 4](#_Toc517685677)

[圖 6 資料匯入之程式碼 4](#_Toc517685678)

[圖 7 折線圖程式碼 5](#_Toc517685679)

[圖 8 102至106年度觀光人次變化趨勢的折線圖 5](#_Toc517685680)

[圖 9 雙縱軸折線圖程式碼 5](#_Toc517685681)

[圖 10 人數與人數百分比雙縱軸折線圖 6](#_Toc517685682)

[圖 11 觀光景點資訊函數程式碼 6](#_Toc517685683)

[圖 12 觀光景點資訊函數運行成果 7](#_Toc517685684)

[圖 13 關鍵字的程式碼與執行成果 7](#_Toc517685685)

[圖 14 dylar與magrittr套件的執行成果 8](#_Toc517685686)

[圖 15 select函數程式碼與執行成果 8](#_Toc517685687)

[圖 16 mutate函數的執行成果 8](#_Toc517685688)

[圖 17 cbind和rbind函數的程式碼與執行成果 9](#_Toc517685689)

[圖 18 merge函數的程式碼與執行結果 9](#_Toc517685690)

[圖 19 melt函數的程式碼與執行結果 9](#_Toc517685691)

[圖 20 str\_detect()函數程式碼與執行成果 10](#_Toc517685692)

[圖 21 str\_sub()函數程式碼與執行成果 10](#_Toc517685693)

[圖 22 熱力圖製作程式碼 10](#_Toc517685694)

[圖 23 「年份」、「人數」和「收入」熱力圖 11](#_Toc517685695)

[圖 24 ddply程式碼 11](#_Toc517685696)

[圖 25 信賴區間圖程式碼 11](#_Toc517685697)

[圖 26 新北、高雄與桃園市觀光收入之信賴區間圖 12](#_Toc517685698)

**表目錄**

[表 1 變數名稱與定義欄位 3](#_Toc517685713)

# 壹、專題介紹

## 一、議題背景

觀光業是台灣經濟一個重要的領域。想要更好更快的發展，必須立足現在，回望過去，汲取經驗和教訓。可以藉由R軟體，分析新北市、桃園市、高雄市、苗栗縣、花蓮縣等四個地區需要門票的景點的觀光大數據，通過兩兩比較，來分析其發展趨勢、面臨的困難和存在的優勢。究其因，結其果，找其長，補其短。希望可以通過此次分析結果可適用於全體範圍的觀光業的發展，故本小組旨在通過分析大量的旅遊數據和相關景點評價文本，找出可供借鑒的經驗。

## 二、資料建模

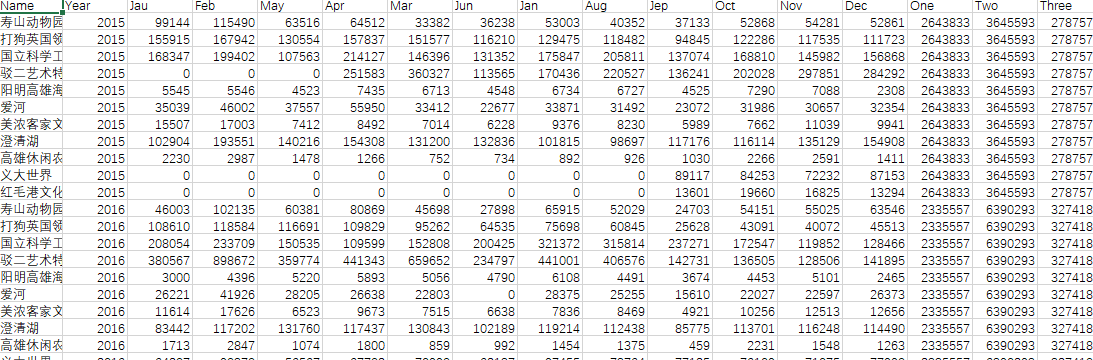


圖 1 建模範例

## 三、欲探討與應用之問題

期中:

(一) 新北市、桃園市兩市發展趨勢比較

(二) 花蓮縣、苗栗縣兩縣發展趨勢比較

(三) 瞭解需要門票的旅遊景點的發展趨勢

(四) 通過最受歡迎的有票景點的對比，探討可持續發展的優勢

期末：

(一)觀光人數與觀光收入的關聯性

(二)南北部城市間的觀光收入差距

# 貳、章節之應用

## 一、第三章之運用

1. 由於一般常用的讀取函數read.table()無法讀取.xls或.xlsx檔，因此首要任務便是安裝read\_excel()函數和其程式庫library(readxl)，安裝過程如圖2所示。

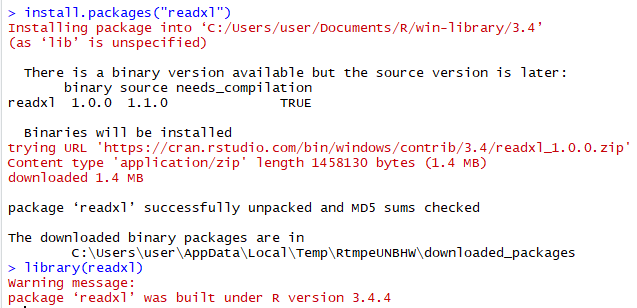


圖 2 readxl安裝程式碼

1. 為了呈現本專題所需的圖表，故安裝ggplot2套件。安裝過程如圖3所示。

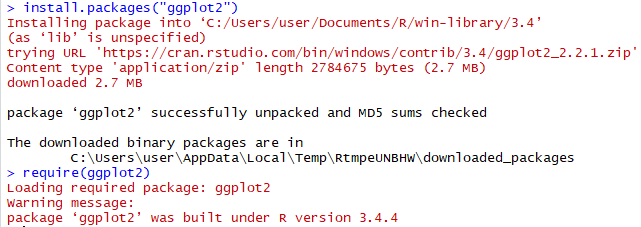


圖 3 ggplot2套件安裝程式碼

1. 圖4為安裝lubridate套件的過程。

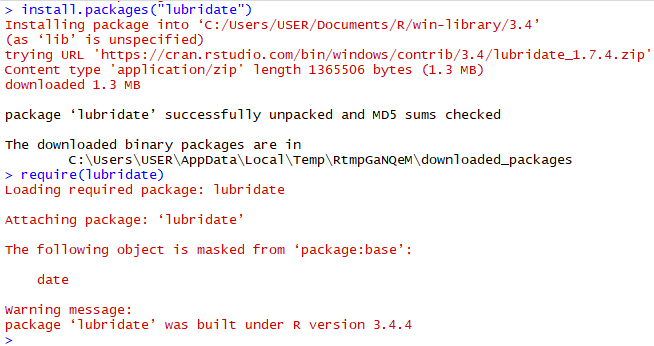


圖 4 lubridate套件安裝程式碼

## 二、第四章之運用

根據我們蒐集與整理後的數據，指派變數並定義其數值型別。(表1)

表 1 變數名稱與定義欄位

|  |  |  |  |
| --- | --- | --- | --- |
| 資料欄位 | 變數名稱 | 資料型態 | 範例 |
| 年份 | Year | int | 2013 |
| 月份 | Month | int | 1 |
| 縣市 | County | char | 新北市 |
| 觀光景點總人次 | Number | int | 1,768,218 |
| 該景點的觀光收益 | Revenue | int | 24,902,136 |

## 三、第五章使用

Data.frame的構建

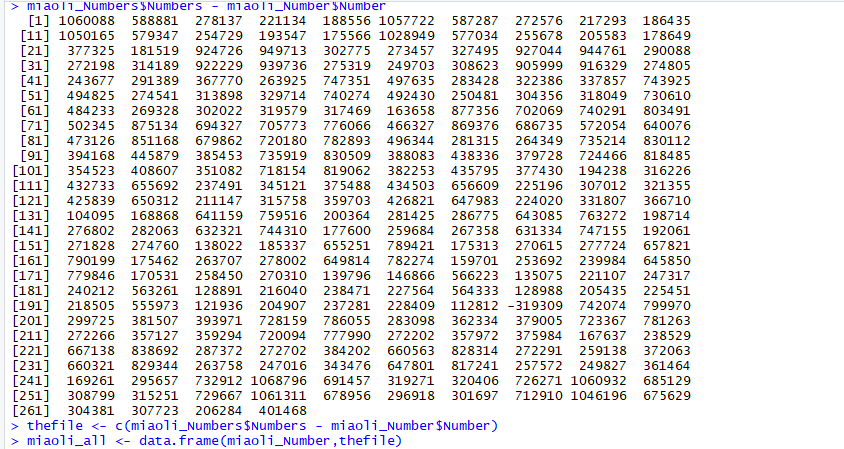


圖 5 Data.frame建構程式碼

## 四、第六章之運用

安裝完readxl()後，我們運用其函數庫中的read\_excel(“路徑”)匯入.xlsx檔的資料到RStudio。(如圖6)

C:\Users\上暐\Downloads\30421908_995097093974605_1631718551_n.png

圖 6 資料匯入之程式碼

## 五、第七章之運用

1. 依據課本7-15至7-16關於線形圖的作法，編寫出四個縣市於102至106年度觀光人次變化趨勢的折線圖。

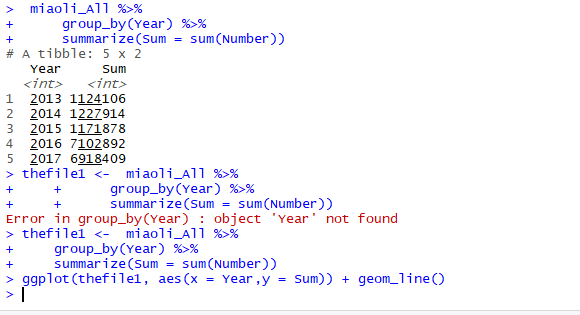


圖 7 折線圖程式碼

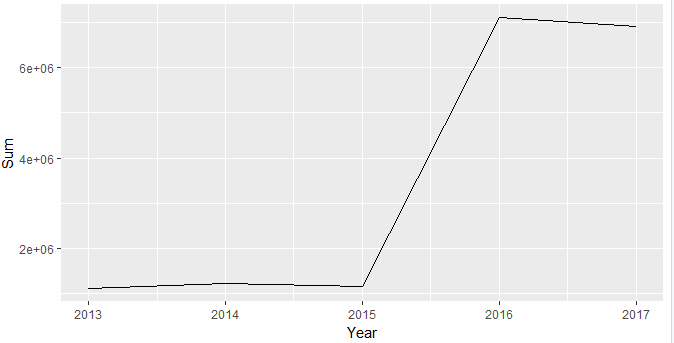


圖 8 102至106年度觀光人次變化趨勢的折線圖

1. 利用雙縱軸(人數、百分比)的方法調查該縣市在某年度的每月觀光人數分佈，同時得知幾月份的遊客數佔該年的最大宗。

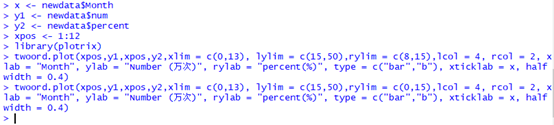


圖 9 雙縱軸折線圖程式碼

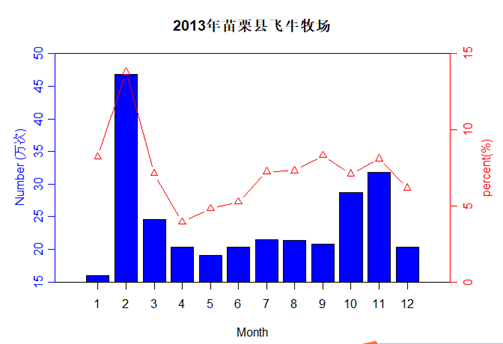


圖 10 人數與人數百分比雙縱軸折線圖

## 六、第八、九章之運用

除了折線圖之外，本小組結合了課本第八章R函數的建立，以及第九章的if和else指令，從而建立新的函數。

使用者只要輸入「新北市」、「桃園市」、「苗栗縣」或「花蓮縣」，即可得知該縣市於102至105年度的觀光景點資訊。

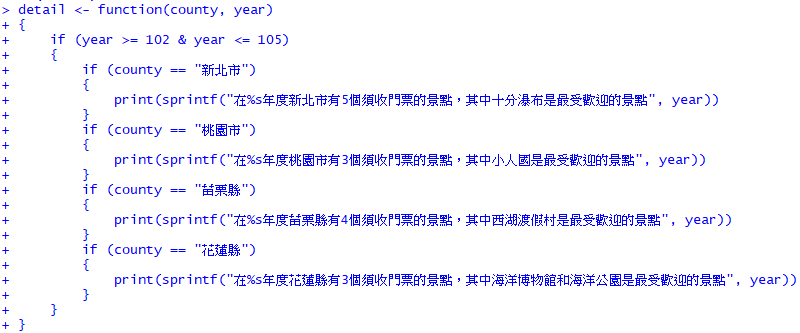


圖 11 觀光景點資訊函數程式碼

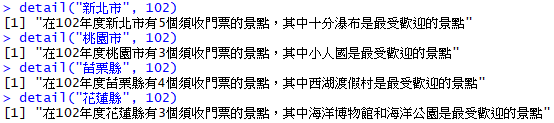


圖 12 觀光景點資訊函數運行成果

## 七、第十一章之運用

關鍵字的設定

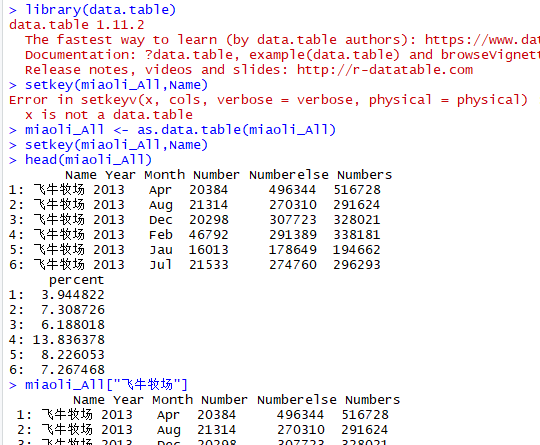


圖 13 關鍵字的程式碼與執行成果

## 八、第十二章之運用

1. dylar與magrittr套件的使用

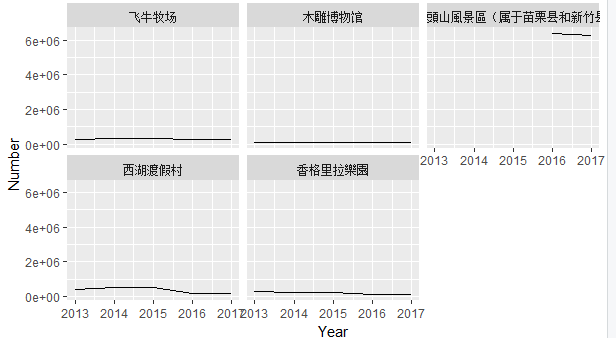


圖 14 dylar與magrittr套件的執行成果

1. select 使用

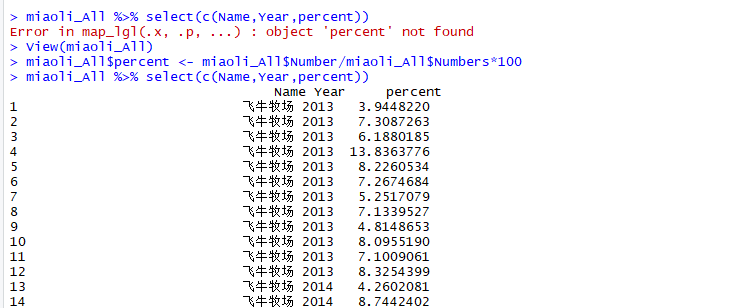


圖 15 select函數程式碼與執行成果

1. mutate的使用

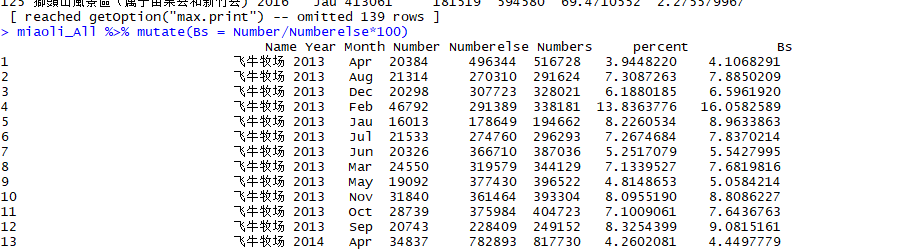


圖 16 mutate函數的執行成果

1. cbind和rbind的使用

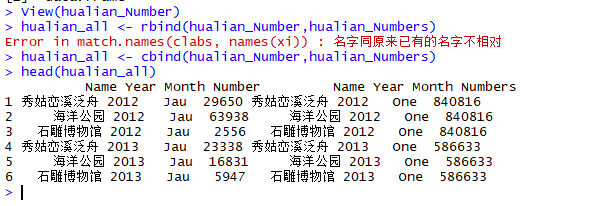


圖 17 cbind和rbind函數的程式碼與執行成果

1. Merge的使用

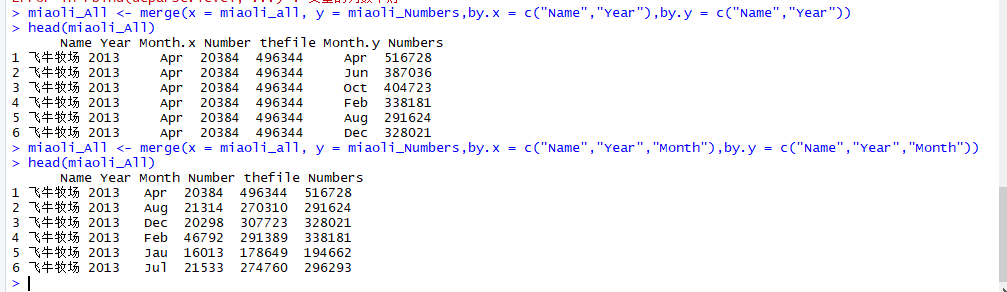


圖 18 merge函數的程式碼與執行結果

1. 安裝Reshape2套件並使用其中的melt函數

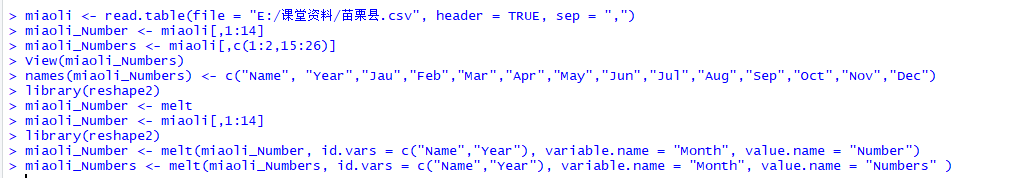


圖 19 melt函數的程式碼與執行結果

## 九、第十三章運用

1. 由於剛得到的原始數據雜亂無章，所以我們先利用str\_detect()函數找出變數名稱，並找出其相關的數值資料，以便進行後續整理。

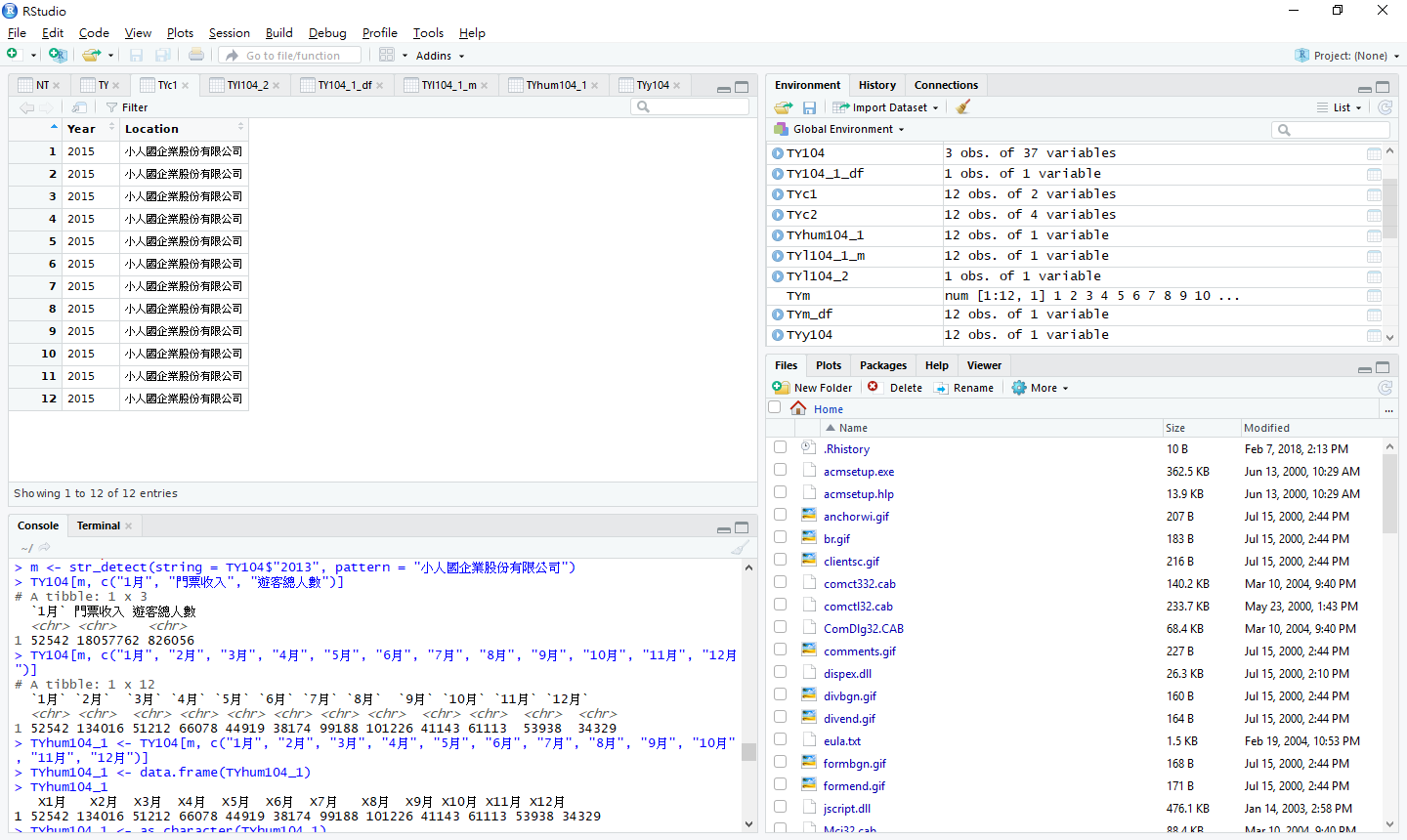


圖 20 str\_detect()函數程式碼與執行成果

1. 除此之外，我們也嘗利用找尋數值資料的方式(str\_sub)，整理出我們需要的資料集。

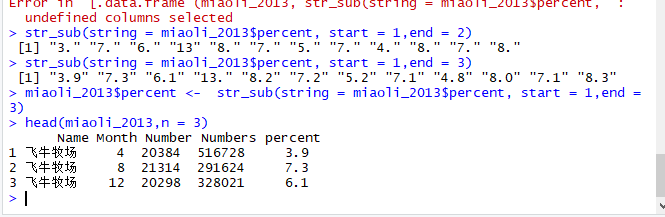


圖 21 str\_sub()函數程式碼與執行成果

## 十、第十五章運用

1. 我們利用cor()函數了解「年份」、「人數」和「收入」三者之間的相關性，並繪製成熱力圖以方便知悉三者間的關聯。由熱力圖上的顯示結果可以得知人數與收入之間呈現低度相關，亦即人數並非決定收入的主要因素。

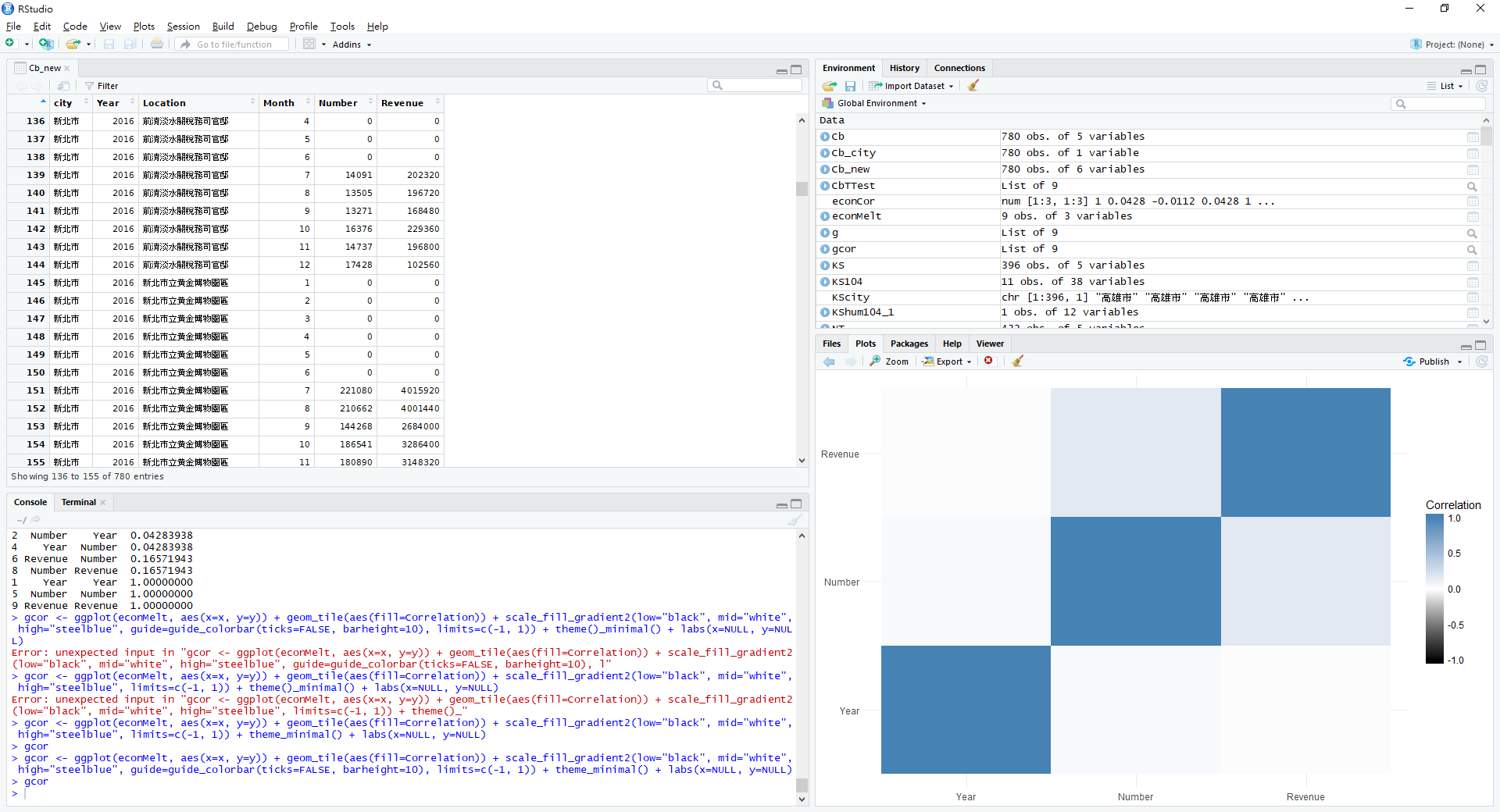


圖 22 熱力圖製作程式碼

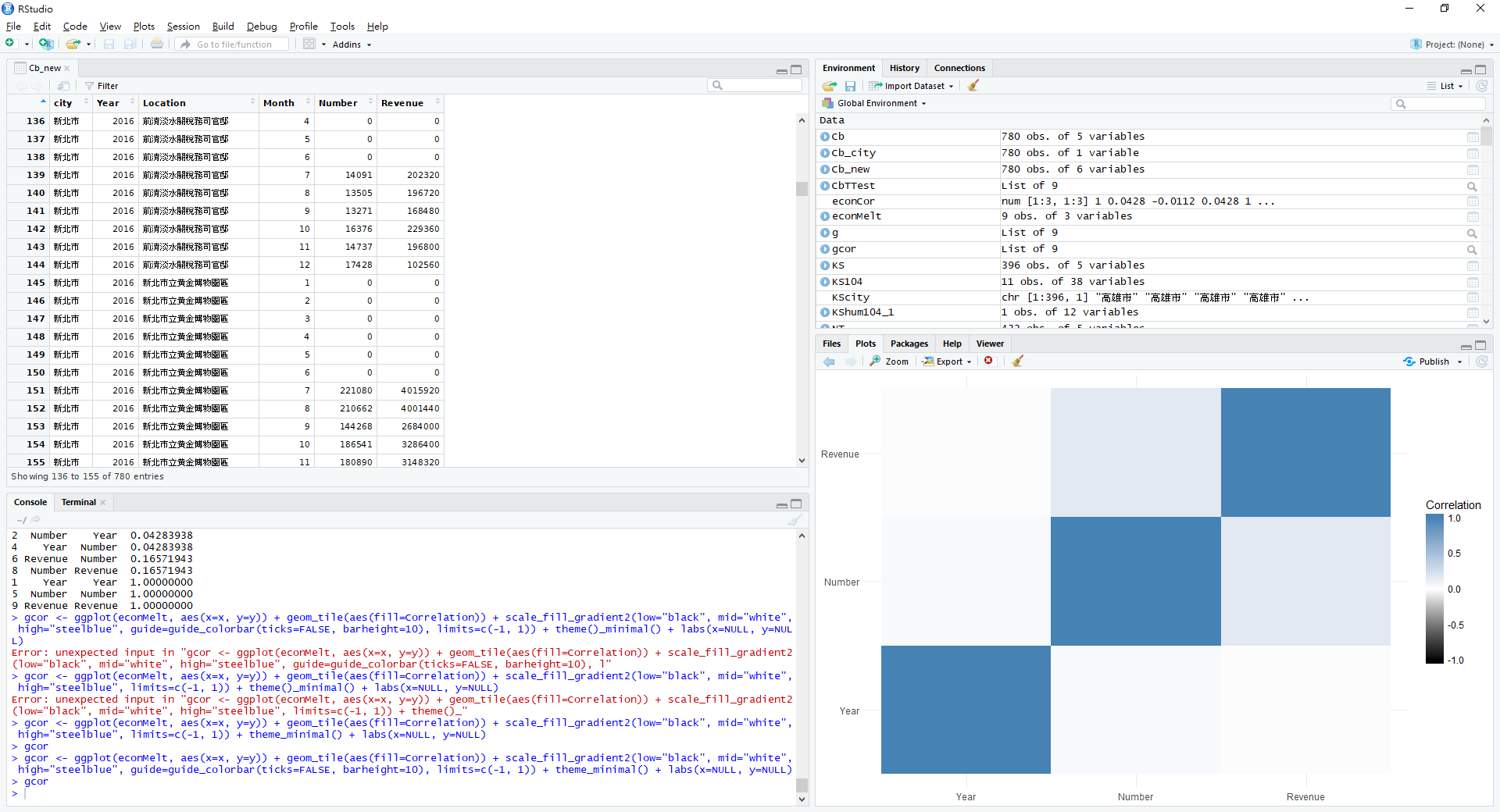


圖 23 「年份」、「人數」和「收入」熱力圖

1. 我們先利用ddply對資料集Cb\_month進行分群並計算出內部資料的平均數、全距的最大和最小值，之後將結果儲存於變數revByCity。接著我們再依據revByCity畫出信賴區間圖。



圖 24 ddply程式碼

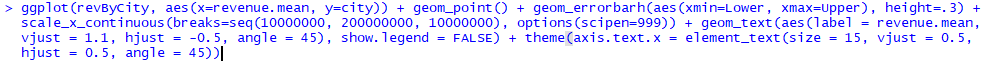


圖 25 信賴區間圖程式碼

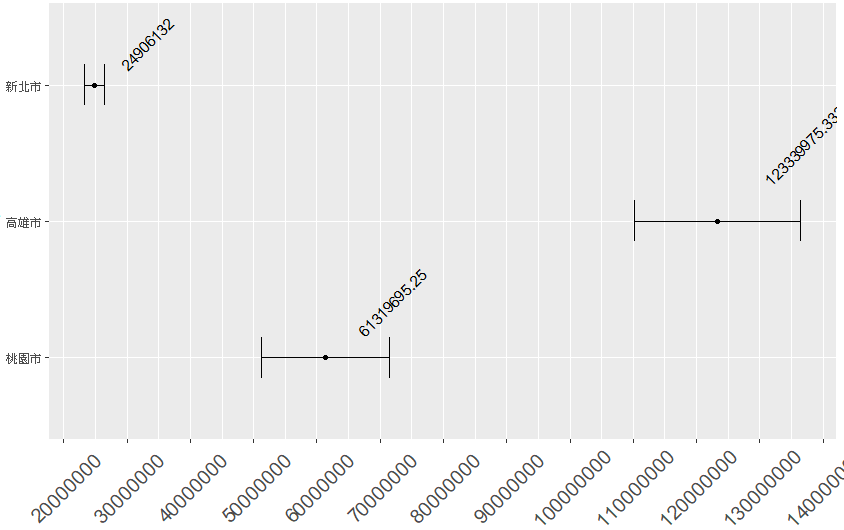


圖 26 新北、高雄與桃園市觀光收入之信賴區間圖

# 參、結論與心得

大眾常有南北差距的認知，不論是經濟、人口、或是各種資源等方面，許多人普遍認為北部優於南部。再者，我們常會將觀光人數的多寡與收入的多少畫上等號。由上述的分析中，可以得知在各縣市的觀光收入比較上，南部城市(高雄市)明顯高於北部城市(新北市、桃園市)，且從熱力圖的呈現可看出觀光人數與觀光收入的關聯為低度相關，也就是說觀光人數多不一定代表連帶的收入也會高於其他縣市。

不同於期中時找資料的順利，後續發展顯得異常困難。由於不同縣市的旅歐觀光景點的資料並沒有得到統一規範，使得我們一籌莫展。後續通過對資料的分析和主題的確定，選定了新北市、桃園市、花蓮縣、苗栗縣和高雄市等五個縣市，並確定了需要用到的工具，這正是之前所說：數據和議題的雙向分析。後經由R軟體的參與，減少了部分工作量，提高了效率。