



國立高雄大學

課程：R 分析應用

109 學年度第一學期

期中報告——高捷運量與雨量之關係

學生：

資訊管理學系

A1063301 王薇琳

A1063329 王新賦

指導教授：曾智義

民國一零九年十一月十六日

# 目錄

壹、 專題簡介.....	1
一、 專題背景 .....	1
二、 資料建模說明 .....	1
三、 問題探討與應用 .....	2
貳、 章節之運用.....	3
一、 安裝所需使用到的套件： .....	3
二、 載入剛剛安裝的套件： .....	3
三、 將 excel 檔案讀入 R studio： .....	3
四、 使用 List 為資料命名： .....	4
五、 計算平均與加總： .....	4
六、 繪製折線圖： .....	4
參、 結論及意涵.....	5
肆、 心得.....	8

# 壹、專題簡介

## 一、專題背景

高捷為全台第二個有捷運的城市，目前已有紅、橘線以及環狀輕軌，未來將繼續持續擴增捷運黃線的部分，預計讓更多市民及遊客享受到便捷的交通服務，擺脫過去高捷只為遊客服務的印象。由於營運初期虧損非常嚴重，原本預計平均要有大約每日 45 萬人次，無奈實際每日只有約 15 萬人次搭乘，幾年下來，導致公司已經面臨破產的邊緣，只好在 2012 年和市政府修改最早的合約內容，之後營運狀況才逐漸改善，並在 2016 年首度有盈餘，而高捷也在 2019 年 12 月底突破 6.5 億人次。

因此，我們想以「運量」為出發，作為本次討論的主題，其中影響運量的關鍵之一就是天氣因素，人們可能會因為天氣因素而影響他們想使用的交通工具及出門意願，比方說：原本想騎機車出遊，但因下雨而改搭捷運。所以，我們訂定本次的主題為「雨量和運量的關係」，分析雨量大小是否會影響捷運運量。

## 二、資料建模說明

我們的資料來自高捷官網提供的歷史運量及中央氣象局提供之天氣資訊，並從中取出 6-8 月（此時間為雨季）每日的運量及雨量資料，再使用 R 語言進行分析。

數據收集來源：

中央氣象局：<https://e-service.cwb.gov.tw/HistoryDataQuery/index.jsp>

高捷運量：<https://corp.krtc.com.tw/News/statistics>

日期	紅線運量(人次)	橘線運量(人次)	總運量(人次)	雨量(mm)
2020/06/01	79,417	25,840	105,257	0
2020/06/02	77,390	25,686	103,076	0

2020/06/03	80,024	27,802	107,826	0
2020/06/04	84,119	27,598	111,717	0
2020/06/05	118,257	43,257	161,514	T
2020/06/06	120,862	36,074	156,936	42
2020/06/07	108,140	29,445	137,585	17.5
2020/06/08	85,104	26,657	111,761	12.5
2020/06/09	81,194	26,865	108,059	T
2020/06/10	84,633	28,450	113,083	0

### 三、問題探討與應用

- 探討今年 2020 年雨量和運量之間的關係為何？
- 根據之後的分析結果，希望將結果提供給企業端參考

## 貳、章節之運用

### 一、安裝所需使用到的套件：

我們運用課本 3-1 所教到的 `install.packages()` 進行套件的安裝。

```
> install.packages('readxl')  
> install.packages("ggplot2")
```

### 二、載入剛剛安裝的套件：

我們運用課本 3-2 節所提到的 `library()` 來將套件載入。

```
> library(readxl)  
> library(ggplot2)
```

### 三、將 excel 檔案讀入 R studio：

使用課本 6-2 節所教到的 `read_excel()` 來進行 Excel 檔案的讀取。

```
> car<-read_excel('C:/Users/admin/Desktop/car.xlsx')  
> rain<-read_excel('D:/大學資料D/大四上/R分析應用/雨量.xlsx')
```

下方為匯入之資料表之截圖。

	Date	RedLine	OrangeLine	Rain
3	2020-06-03	80024	27802	0.0
4	2020-06-04	84119	27598	0.0
5	2020-06-05	118257	43257	0.0
6	2020-06-06	120862	36074	42.0
7	2020-06-07	108140	29445	17.5
8	2020-06-08	85104	26657	12.5
9	2020-06-09	81194	26865	0.0
10	2020-06-10	84633	28450	0.0
11	2020-06-11	88087	29914	0.0
12	2020-06-12	110067	33681	0.0
13	2020-06-13	116352	36664	0.0
14	2020-06-14	105292	29837	7.5
15	2020-06-15	89179	28950	0.0
16	2020-06-16	89424	29306	0.0

#### 四、使用 List 為資料命名：

使用課本 5-2 所提到的 list 來為所讀取到的 Excel 資料命名。

```
> names(car)<-c('Date','RedLine','OrangeLine')
```

#### 五、計算平均與加總：

先指定 xlsx 檔裡面的分頁和該分頁表裡面的特定欄位，以下以讀取 xlsx 中七月的雨量為範例。

```
> rain7 <- read_excel('D:/大學資料D/大四上/R分析應用/雨量.xlsx',sheet=2)
> rain$mm
```

使用課本第 4 節所提到幾個常用函數 mean()和 sum()來進行雨量的平均與加總的計算。

```
> mean(rain77) > sum(rain88) > sum(rain66)
[1] 4.725806 [1] 915 [1] 915
> sum(rain77) > mean(rain88) > mean(rain66)
[1] 146.5 [1] 29.51613 [1] 29.51613
```

使用課本第 4 節，所提到幾個常用函數 mean()和 sum()來進行運量的平均與加總的計算。

```
> redsum<-car$RedLine > orangesum<-car$OrangeLine > rainsum<-car$Rain
> sum(redsum) > sum(orangesum) > sum(rainsum)
[1] 9588128 [1] 3064295 [1] 1224.5
```

#### 六、繪製折線圖：

使用課本 7-2 所教的 ggplot2 來繪製 6-8 月捷運運量折線圖，並將紅線與橘線資料一同比較。

```
> ggplot(car)+geom_line(aes(x = Date,y=Line,group=1,color="red"))+geom_line(aes(x=
Date,y=OrangeLine,group=1),color="orange")
```

使用課本 7-2 所教的 ggplot2 來繪製 6-8 月雨量折線圖。

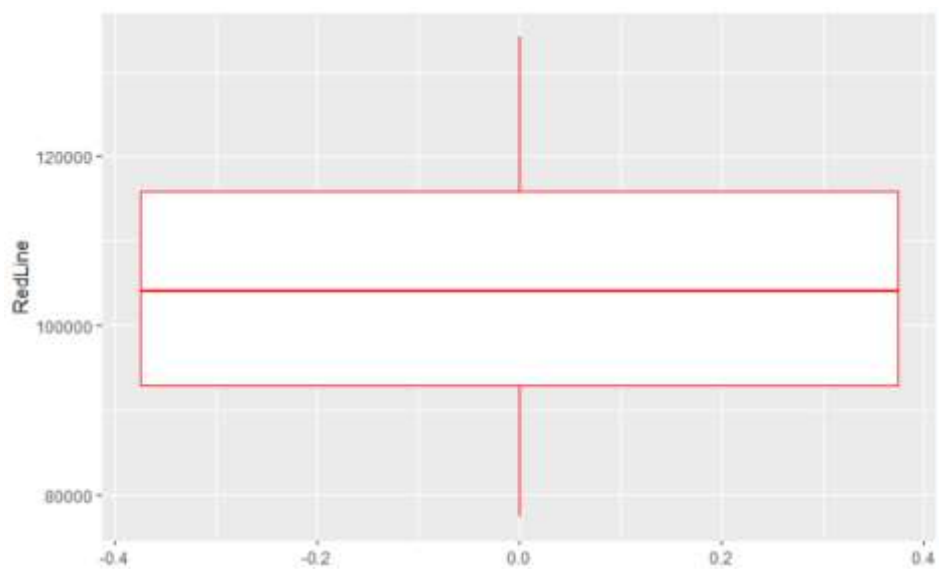
```
> ggplot(rain,aes(x = date, y =mm)) +geom_line()
```

使用課本 7-2 所教的 ggplot2 來繪製 6-8 月雨量與運量的箱型圖。

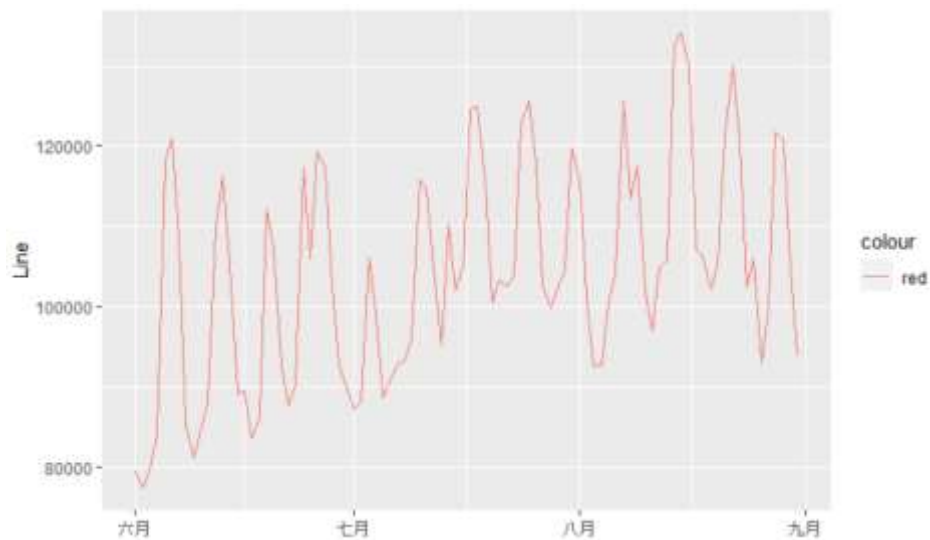
```
> ggplot(car, aes(y = RedLine)) +geom_boxplot(color = "red")
> ggplot(car, aes(y = OrangeLine)) +geom_boxplot(color = "orange")
> ggplot(car, aes(y = Rain)) +geom_boxplot(color = "blue")
```

## 參、結論及意涵

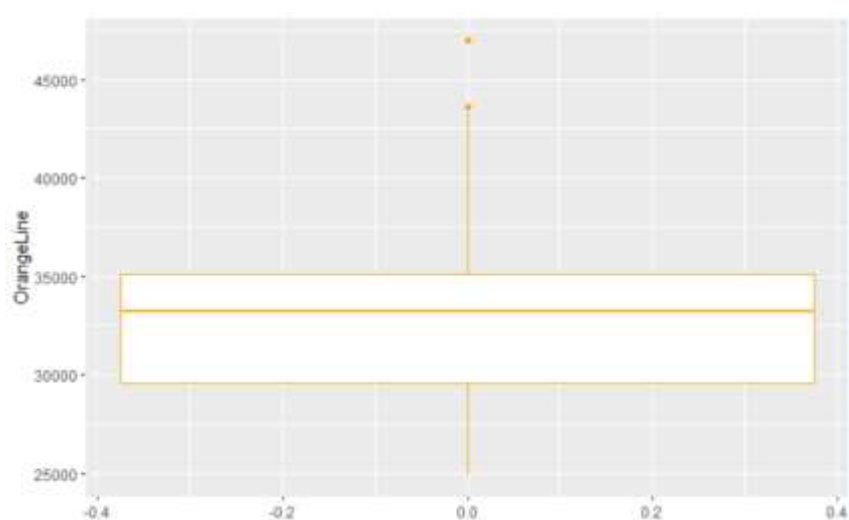
根據下方箱型圖所示，紅線運量之中位數為 104060 人次，有 50%的資料位於箱子內部，且箱子內部分布均勻，而另外 50%資料較不易觀察。



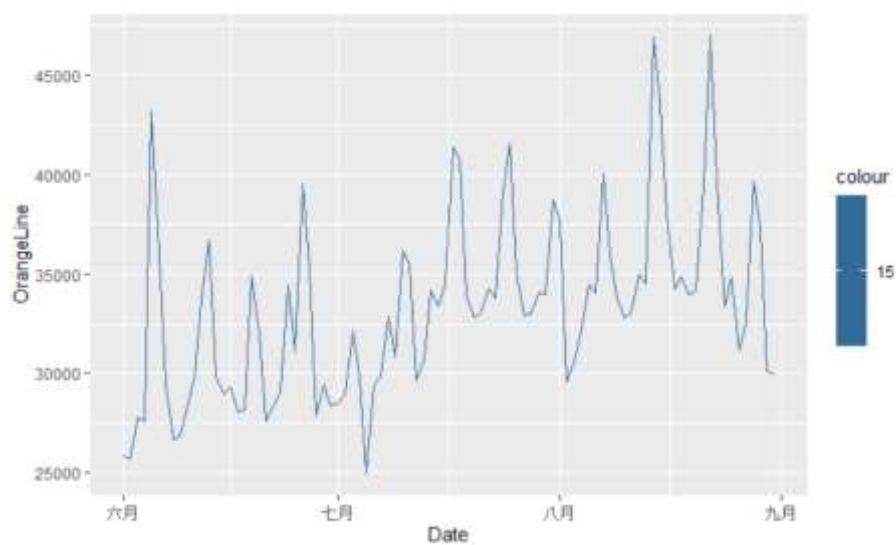
根據統計，紅線 6 月份平均每日人次為 97,396，總運量為 2,921,872 人次，七月的平均每日人次為 105,194，總運量為 3,261,007，八月的平均每日人次 109,847，總運量為 3,405,249 人次。



根據下方箱型圖所示，黃線運量之中位數為 33197 人次，有 50% 的資料位於箱子內部，且箱子內部分布不均，中位數較靠近 Q3，離 Q1 較遠，而另外 50% 資料較不易觀察。



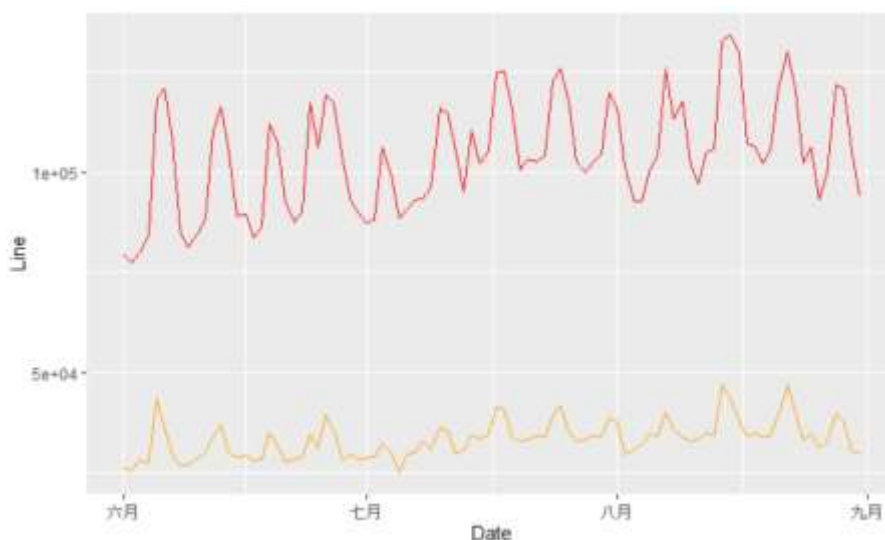
根據統計，橘線六月份平均每日人次為 30,674，總運量為 920,226 人次，七月份的平均每日人次為 33,710，總運量為 1,011,314，而八月的部分，平均每日人次為 35,621，總運量為 11,042,514 人次。



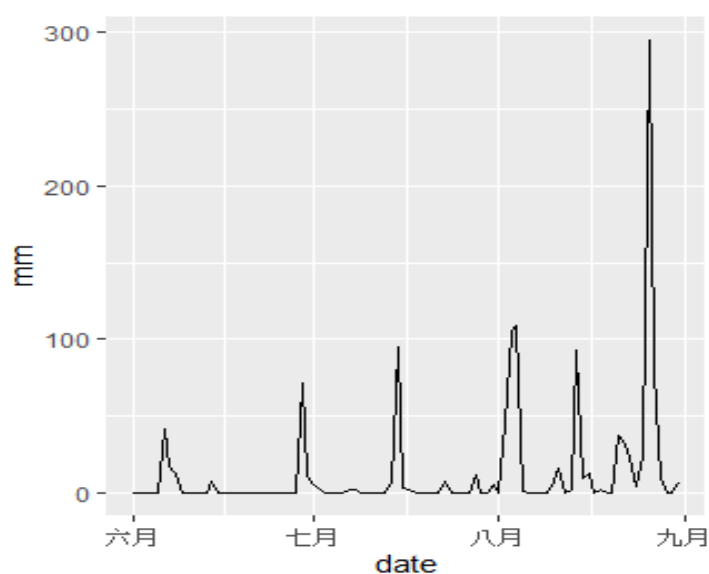


高捷在 6-8 月的運量與雨量關係：

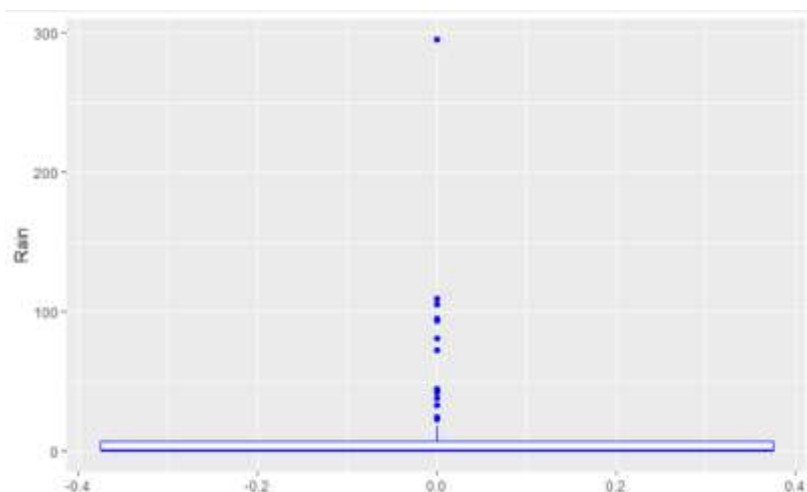
由下圖可知，當雨量特別大的時候(比方說單日雨量大於 50mm)，高捷的運量通常是下降的。民眾可能因為下雨而改變搭乘的交通工具，像是改開車避免從捷運下車到目的地前淋到雨，也可能因為下雨而直接不出門，導致捷運運量下降，因此雨量與運量之間的關係呈負相關。而雨量若不夠大(例如單日雨量小於 50mm)，則雨量對於運量的影響則不明顯。



如下圖，6 月的平均雨量為 5.43mm，7 月的平均雨量為 4.73mm，8 月的平均雨量為 29.51mm(其中在 8/26、8/27 時有西南氣流影響導致雨量暴增)。



根據下方箱型圖所示，雨量之中位數趨近於 0mm，幾乎與 Q2 重疊，有 50% 的資料位於箱子內部，且因各日雨量差異較大(例如受到西南氣流、颱風外圍環流影響)導致資料分布不均，而另外 50% 資料較不易觀察。



## 肆、心得

### 一、王薇琳：

(1) 工作：資料調查、收集數據、圖表製作、圖表分析。

(2) 心得：

雖然在系上曾修過 JAVA 等程式語言課程，但 R 語言是第一次接觸，在語法及操作上都與之前所學不同，讓我感到很新鮮。

在報告的討論過程中，我們曾遇到圖表繪製不出來的問題，在不斷查詢資料、尋找解決辦法並互相討論後，終於找出問題並解決，成功繪製同時有兩份數據的折線圖。

透過這次報告，讓我更了解資料分析相關語言的操作，也了解了 R 語言的深奧之處，相信在期中後的課程能讓我更深入地了解 R 語言，並在期末時能夠利用這學期所學將我們的報告更完整的呈現給大家。

### 二、王新賦：

(1) 工作：資料調查、收集數據、圖表製作、圖表分析。

(2) 心得：

和老師給的範例相比，我們所選定的雨量議題似乎沒有那麼大而廣闊，比較民生、貼近生活，我們試著讓議題展開，因此我們根據生活經驗，決定加入捷運方面的相關議題，讓這次的作業主題能夠有比較多發揮的空間。而在實際製作作業的過程中，其他部分沒有甚麼大問題，主要的問題發生在使用 ggplot 的時候，由於雨量方面我希望能夠使用直方圖來呈現，卻發現無法同時設定 x 軸和 y 軸的單位數值，只能選其一來

操作，因此無奈只能放棄並改以折線圖來做為呈現。最後，透過這次的期中作業，能夠實際操作 R 語言，善用課堂所學，使我獲益良多。

### 當日上台報告之照片：

組員王新賦、王薇琳(當日面試碩士班請假)

