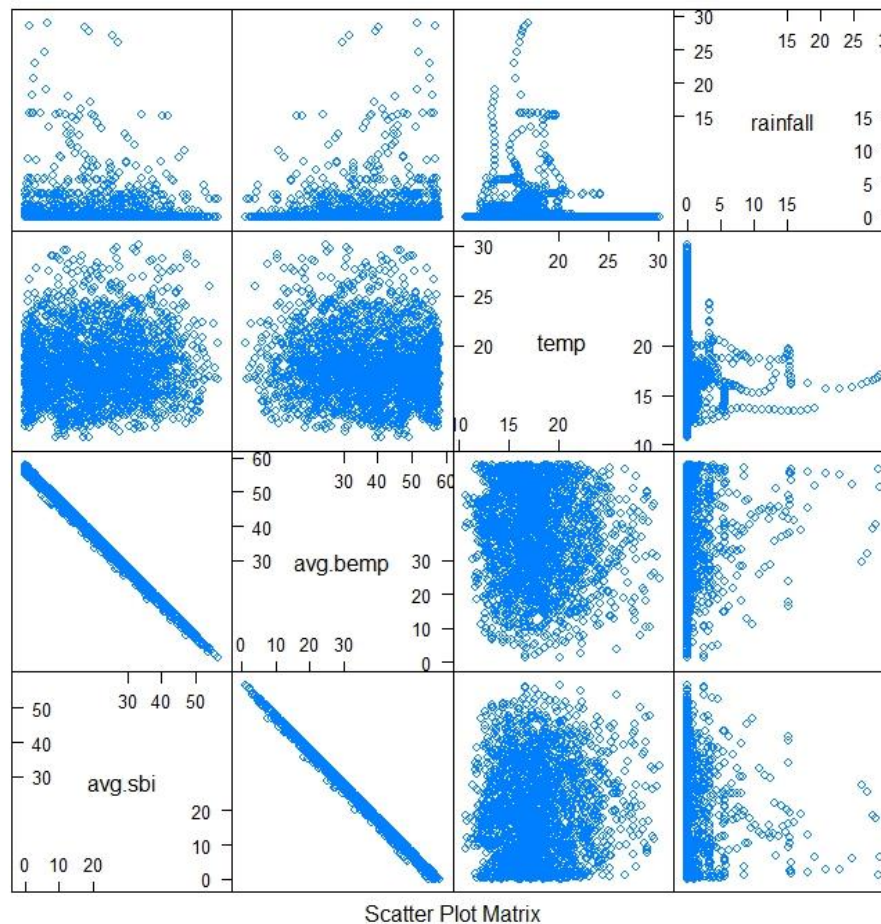


## 第 5 組 湯忠憲, 林哲丞, 簡愷均

編號	樣型/規則描述	提出假設/研究命題	預計檢驗方式
1	從 pairwise scatter plot (圖一)中可以看出雨量與平均車輛數大致呈負相關與平均空位數大致呈正相關。	假設車輛數受降雨量影響，可以基於該小時降雨量之預測來預測車輛數。	透過歷史資料的回測，衡量 RMSE 或是 MAPE 等指標。
2	從圖二中可以看出車輛數於早上 7 點、12 點以及下午六點有小高峰。	假設車輛數受上下班以及中午休息時間影響，可以利用此樣型提升車輛數預測的準確度。	透過歷史資料的回測，衡量 RMSE 或是 MAPE 等指標。
3	以標準化後的平均車輛數為特徵做場站聚類，主要可以分出公園、小學或是通勤場所等類別。(圖三)	基於場站聚類提升車輛數預測準確度。	透過歷史資料的回測，衡量 RMSE 或是 MAPE 等指標。
4	從圖四觀察平均雨量與各場站之間的關係。可以發現接近山區的場站有較高的平均雨量。	台灣北部山區多處於迎風坡面，使得降雨量相對較多。利用各場站降雨量有所不同的特點，可研究如何最佳化場站車位數以及臨時調度。	訂定車位數管制線，以該線為標準判斷是否能有效管理車位。

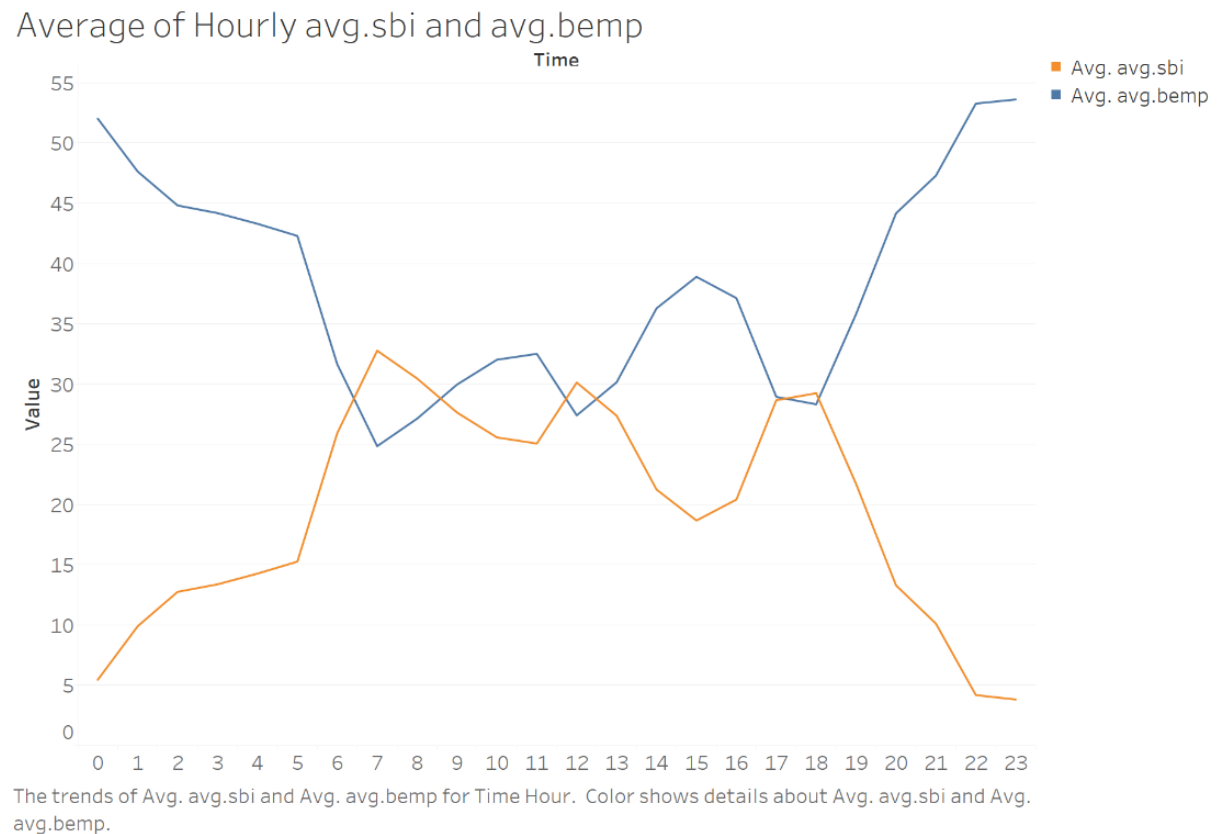
屬性名稱	說明	衡量尺度
date	日期	名目尺度
hour	時間	名目尺度
sno	場站代號	名目尺度
sarea	場站區域	名目尺度
sna	場站名稱	名目尺度
lat	緯度	間距尺度
lng	經度	間距尺度
tot	總停車格	間距尺度
avg.sbi	平均車輛數	間距尺度
max.sbi	最大出輛數	間距尺度
min.sbi	最小車輛數	間距尺度
std.sbi	車輛數標準差	間距尺度
avg.bemp	平均空位數	間距尺度
max.bemp	最大空位數	間距尺度
min.bemp	最小空位數	間距尺度
std.bemp	空位數標準差	間距尺度
temp	平均溫度	間距尺度
humidity	濕度	間距尺度
pressure	氣壓	間距尺度
max.anemo	最大風速	間距尺度
Rainfall	降雨量	間距尺度

1. 以捷運北投場站為例，提取每小時平均車輛數、平均空位數、平均氣溫以及降雨量等屬性繪製散佈圖。可以發現平均車輛數對應降雨量大致呈現負相關，平均空位數對應降雨量呈正相關。而溫度對應降雨量呈負相關。透過本樣型可以解釋下雨天比較少人騎腳踏車的自然現象。此外，也了解到雨量用於預測平均車輛數將會是有效的因子。



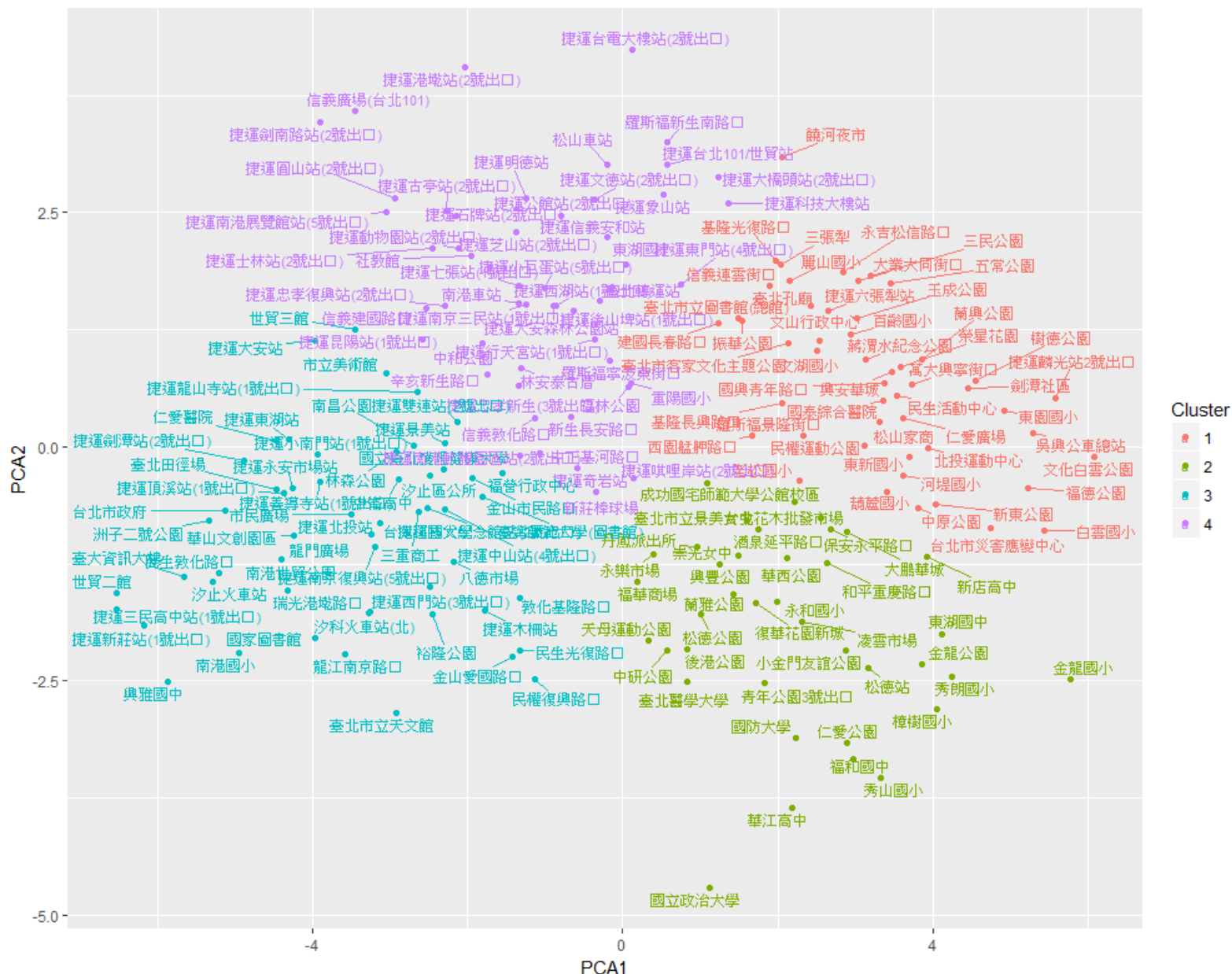
圖一、散佈圖矩陣

2. 以建國路口場站為例，利用 2014-12-08 到 2015-03-31 中，每單位小時平均車輛數以及平均空位數的平均數繪製圖二。黃線為平均平均車輛數，可以看出車輛數於早上 7 點、12 點以及下午 6 點有小高峰。對應到時間即是早上通勤、中午休息以及傍晚通勤的時段。因此，假設車輛數受上下班以及中午休息時間影響，可以利用此樣型提升車輛數預測的準確度。



圖二、每小時平均平均車輛數以及平均空位數

3. 以標準化後的平均車輛數為特徵做場站聚類。先新增欄位 `avg.sbi.ratio`，以平均車輛數除以總停車格數。該欄位主要用於標準化平均車輛數，衡量尺度主換為絕對尺度。接著將每小時標準化後的平均車輛數做為各場站的特徵，利用主成分分析 (Principal components analysis, PCA) 將特徵為度降至二維。以該二維特徵用 k-平均演算法 (k-means clustering) 聚類，並繪製圖三。可以發現第一類主要以公園、運動中心為主，第二類為國中小學，第三類集中在市區，第四類主要為商業區塊。透過聚類可以賦予各場站新的屬性，對於車輛數預測或是實際管理有一定的幫助。



### 圖三、ubike 場站聚類



4. 將各場站依照其經緯度座標標示於地圖上，點的顏色代表場站類別，大小代表該場站於十二月八日至一月九日的平均降雨量。其中，鄰近山區的場站如汐止、南港等區域點的半徑較大，相較於市中心多雨。因此，利用各場站降雨量有所不同的特點，搭配與平均車輛數的關聯性，可研究如何最佳化場站車位數以及臨時調度。

