

# 文化沙漠在哪裡？

<R 語言期末專題>

B07B01028 生科二 林哲豪

B08303111 經濟一 陳靖騏

B05608005 園藝四 姜禹安

B07611011 生機二 魏廷宇



# 目錄

## 1. 簡介

## 2. 方法

### 2.1 資料取得

### 2.2 原始碼運用說明

## 3. 結果

## 4. 組員分工

## 5. 參考資料

# 1. 簡介

動機：

由於常常聽說在台灣從事藝文產業沒有前途，藝文青年工作者收入低廉，因此我們想知道究竟是什麼造成了這樣文化沙漠的情景，又想知道在台灣，文化沙漠最嚴重處究竟在哪？

目的：

我們由政府 and 民間兩方面切入，先利用各項指標，將政府投入的程度以及人民參與度分級，將名次最後的幾名挑出，並探討可能的原因為何。

## 2. 方法

### 2.1 資料取得

自文化部文化統計網站下載 2012~2017 年間的文化統計資料，以 R 進行資料清理與指標換算，政府投入指標分別為：人才培育總小時數、文化預算佔總預算比例和每萬人均展演場地數，民眾參與指標為：人均文化場地營業額，每千人均藝文活動參與次數，和每萬人均文化場地組織個數。

## 2.2 原始碼運用說明

以 dplyr 對各縣市之指標排序並比較後，選取非直轄市和直轄市中文化指標排名倒數前三名以及排名最高者，以 ggplot2 作圖。



以 NA 取代離群值，分別依地理分區及行政層級劃分各縣市，以線性回歸之  $R^2 > 0.3$  為基準，篩選出政府投入語民眾參與指標間具有較明顯相關性的群組。



以 lm() 函數與 Partial F test 逐一檢驗線性與多項式迴歸線之適配性，決定最佳適配之迴歸方式，建立模型。

### 3. 結果

以 dplyr 對各縣市之指標排序並比較後的詳細分析：

**非直轄市比較：**

倒數 10 名分別為：金門縣、屏東縣、嘉義縣、雲林縣、彰化縣、苗栗縣、基隆市、新竹縣、澎湖縣、連江縣；多為西部或是離島縣市，大多缺乏的是政府的資源投入

新竹市：(人民參與良好、政府投入仍須配合)

藝文活動參與次數、文創產業營業額上升，民間文化組織下降

→政府投入指標總體下降，教育小時總數成下滑趨勢

→人民參與指標總體上升，各指標排名皆處前段，良好發展趨勢

金門縣：(改善政府投入)

藝文活動參與次數、文創產業營業額明顯上升

→政府投入指標微上升，教育小時總數過低，可增設課程培育人才

投入觀光文創產業

→人民參與指標上升

屏東縣：(政府發展遲緩)

→政府指標幾乎持平，教育小時總數變化大不平穩，而文化局占地

方預算占比為非直轄市最低，可調整政策平衡發展方向

→人民參與指標總體微上升，只有藝文活動參與次數有下降趨勢

嘉義縣：(上升趨勢)

→政府投入指標上升，但教育小時總數相較其他縣市偏低，可加強

文化課程培育人才，以改善文創產業較低落的現況

→人民參與指標明顯上升，文創產業營業額倒數第二

### 直轄市比較：

「臺北市」的人民參與指標排名最高，但在政府投入指標中卻倒數第二。

「臺南市」的政府投入指標最高，但在人民參與指標卻是最後一名。

相對來說，高雄在兩方面的指標中都是位居第二位，因此總體排名佔據鰲首。

縣市	人民參與(名次)	政府投入(名次)
台南市	15	6
桃園市	15	10
台中市	12	16
新北市	12	10
高雄市	5	8
台北市	4	13

高雄市：(下降趨勢)

→政府投入指標總體明顯下降趨勢

→人民參與指標明顯全數下降，但前幾年的數值相對較高，因此平均值仍勝過其他直轄市

臺中市：(改善人民參與、政府投入發展遲緩)

→政府投入指標總體微上升、歷年教育小時總數雖回歸線呈現上升，但近年仍逐年降低

→人民參與指標明顯下降

桃園市：(上升趨勢)

→政府投入指標總體上升，但各數值相較其他直轄市偏低，可能與較晚升格直轄市相關，文化資源相對落後

→人民參與指標有上升趨勢

新北市：(發展遲滯，唯文化局占地方政府預算比例排名倒數第二，可從此或是增設藝文場所來提升)

→政府投入指標總體上升，唯有教育小時總數明顯上升趨勢且數值明顯較高

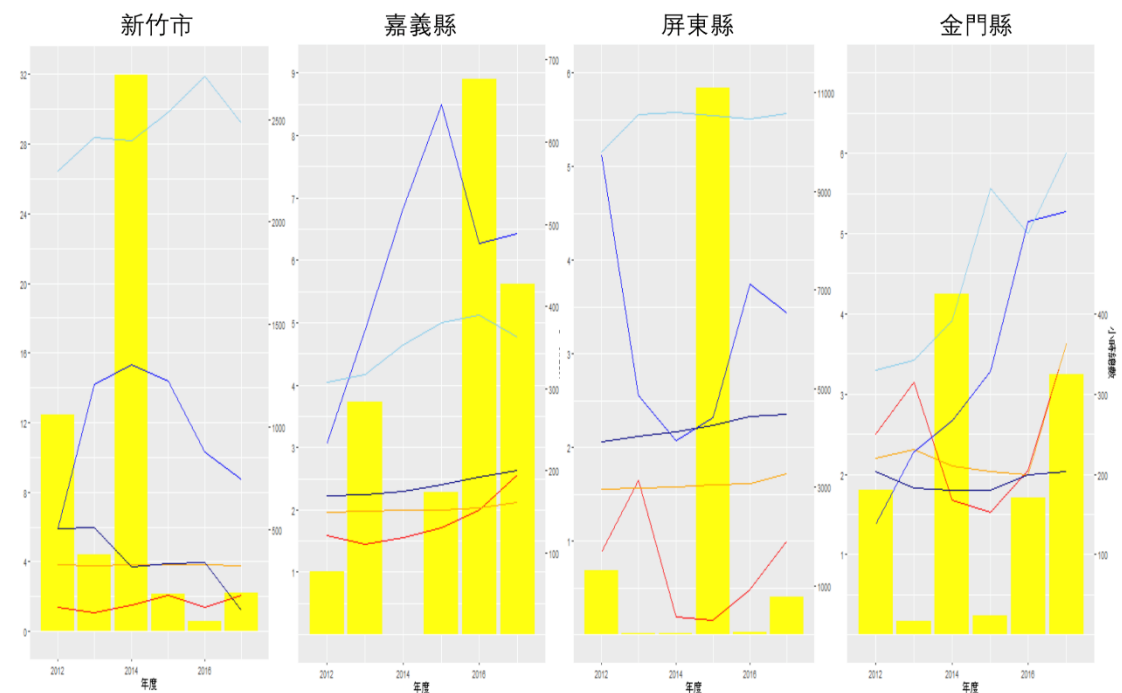
→人民參與指標幾乎持平，文創產業營業額略降

## 綜合直轄市、非直轄市比較：

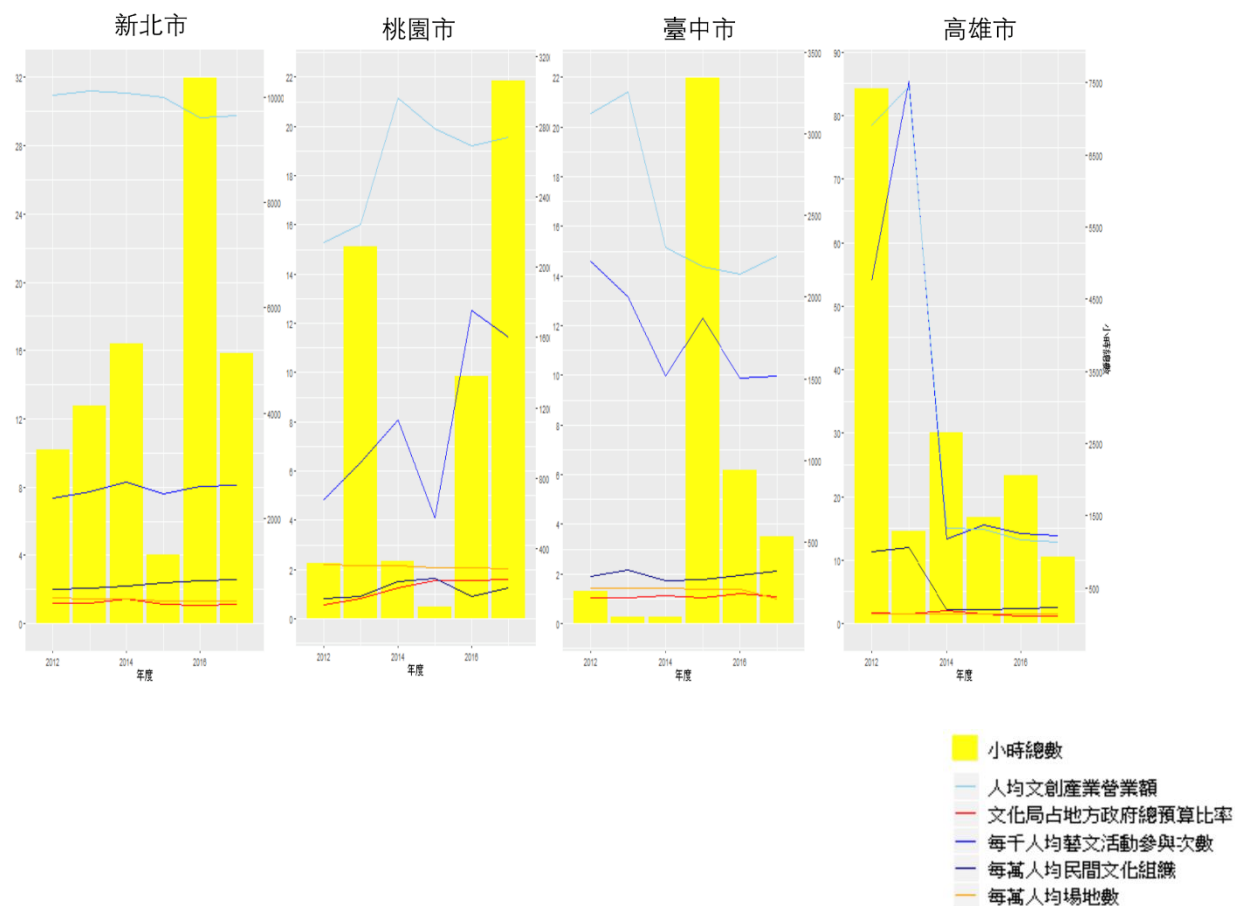
極少政府投入增加後但是人民參與卻下降的情況(臺中市的政府投入指標幾乎持平、小時總數近年大幅下降)；

均未見人民參與指標上升但政府投入明顯降低的情況，而兩指標同升或同降的情況占大多數，說明兩者之間可能呈現正相關的關係，只是資料數較少，因此未有明顯統計資料佐證，僅以圖形趨勢來判斷各縣市的发展現況。

將結果以 ggplot2 作圖：





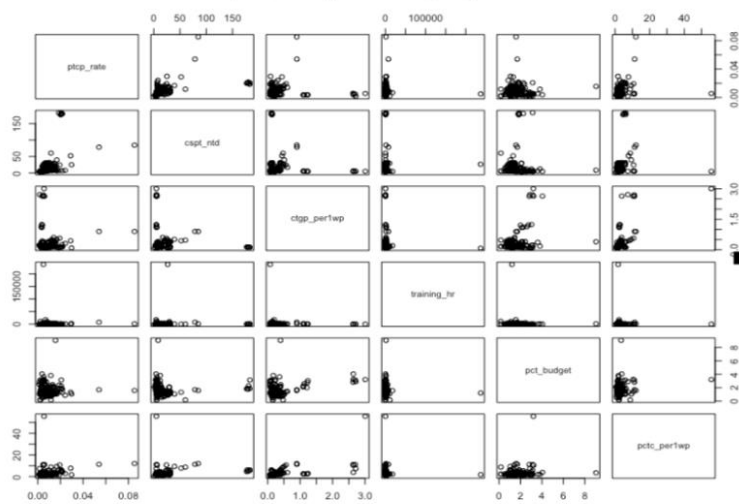


## (去除離群值、分群尋找相關性與回歸分析) 結論：

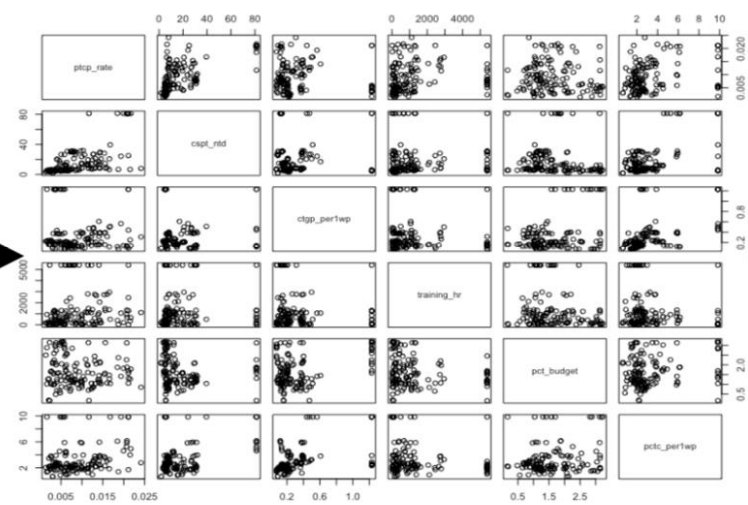
### 1. 去除離群值：

參考 <Outlier detection and treatment with R> 文中的方式，以  $Q1 - 1.5 IQR$  與  $Q3 + 1.5 IQR$  為標準( $IQR = Q3 - Q1$ )，符合兩者即判定為離群值。取代的方法則使用自訂函數，分別以(1) 5% 與 95%百分位數 (2) 缺失值 取代離群值後繪製散佈圖矩陣，觀察所有指標間之相關性，發現以缺失值取代離群值，較能顯示資料點的分散程度。總共取代了 59 個離群值，數量佔全部資料的 7.45%。

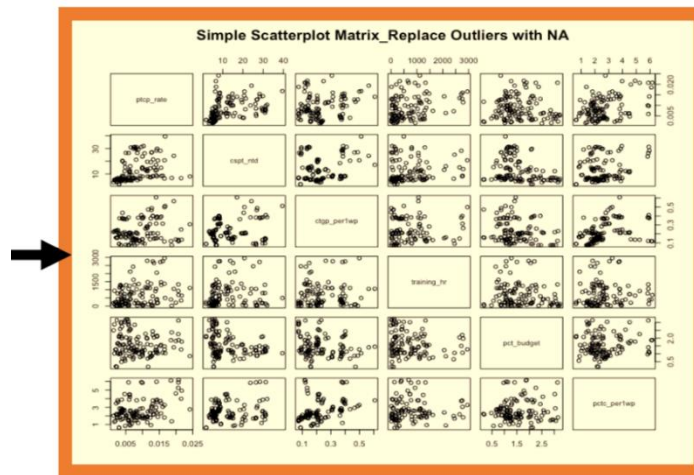
Simple Scatterplot Matrix\_Original Data



Simple Scatterplot Matrix\_Replace Outliers with 5%/95%ile



Simple Scatterplot Matrix\_Replace Outliers with NA



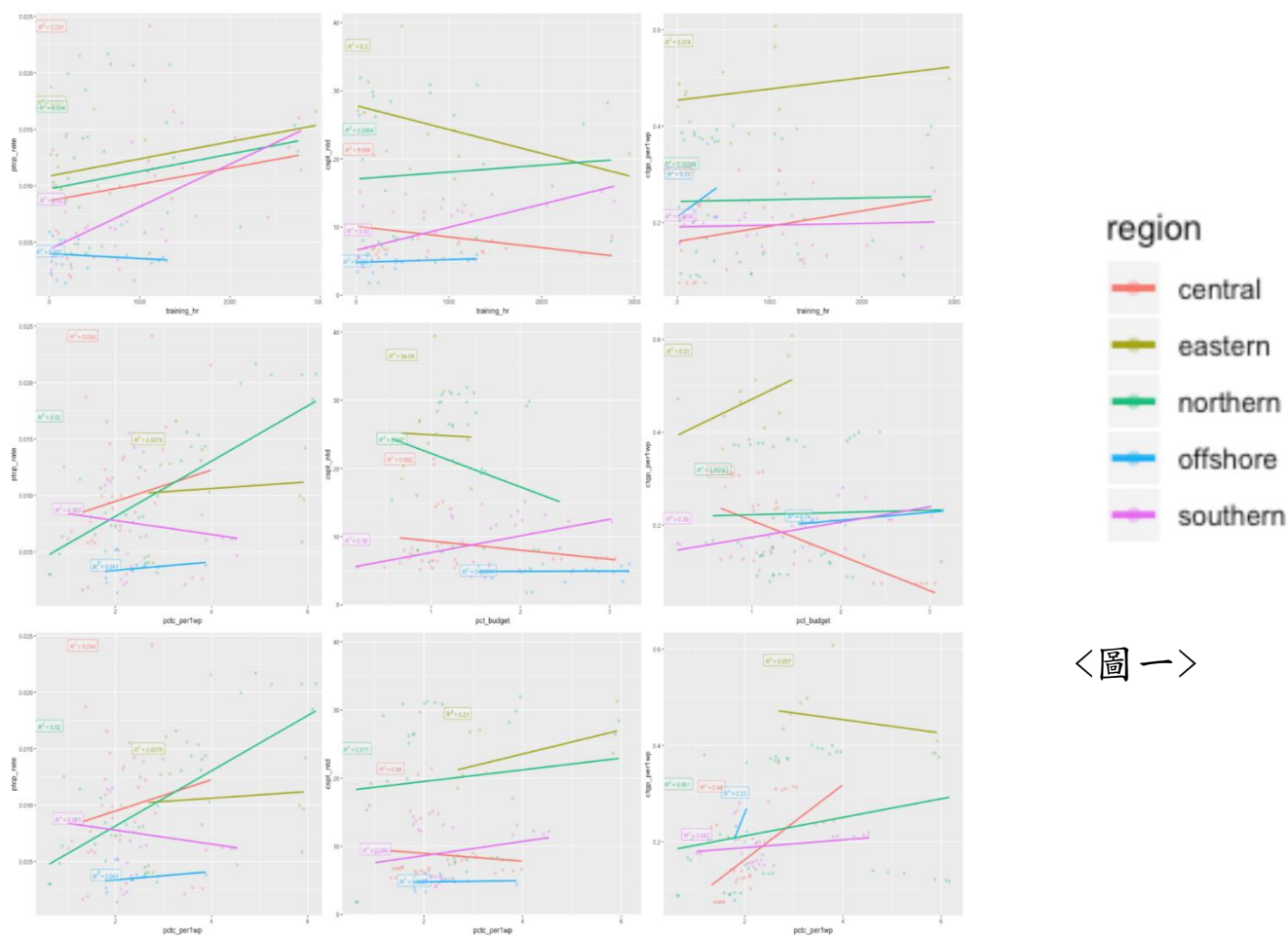
## 2. 找出具相關性群組：

以地理分區(北、中、南、東部及離島)<圖一>與行政層級(直轄市、市及縣)<圖二>將各縣市分組，以簡單線性回歸擬合各組資料點，並在圖上顯示出決定係數。以決定係數  $R^2 > 0.3$  為判斷標準，篩選出具有相關性的地理分區與行政層級群組，分別為：

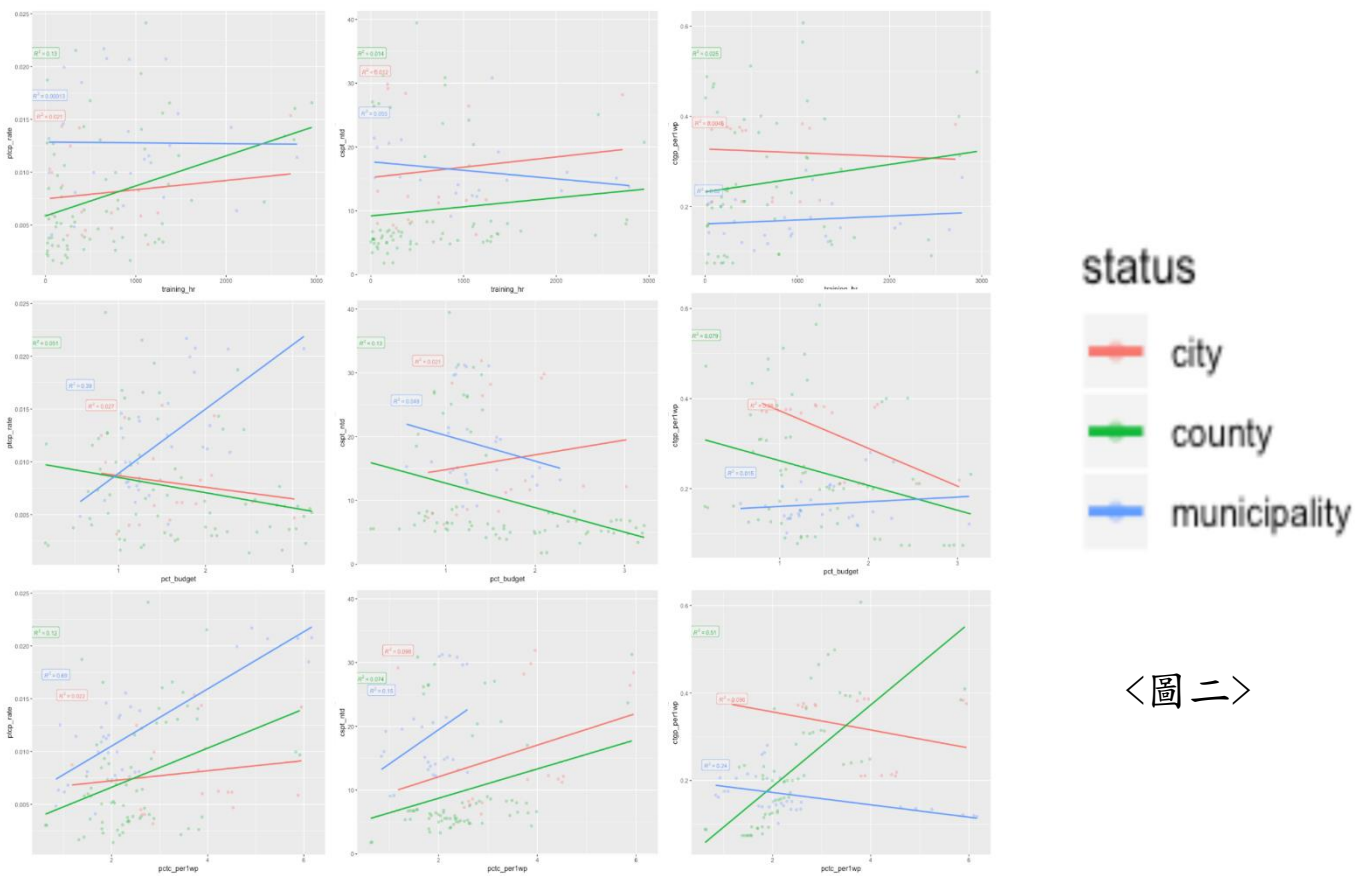
( 群組名： 政府投入指標 / 民眾參與指標 )

- A. 南臺灣：文化培訓總時數 / 每千人均藝文活動參與次數
- B. 南臺灣：文化培訓總時數 / 人均文創產業營業額
- C. 中臺灣：文化預算占總預算比例 / 每萬人均文化團體數
- D. 南臺灣：文化預算占總預算比例 / 每萬人均文化團體數
- E. 北臺灣：每萬人均藝文場地數 / 每千人均藝文活動參與次數
- F. 中臺灣：每萬人均藝文場地數 / 每萬人均文化團體數 (負相關)
- G. 直轄市：文化預算占總預算比例 / 每千人均藝文活動參與次數
- H. 一般市：文化預算占總預算比例 / 每萬人均文化團體數
- I. 直轄市：每萬人均藝文場地數 / 每千人均藝文活動參與次數
- J. 一般縣：每萬人均藝文場地數 / 每萬人均文化團體數 (負相關)

( 未標註者兩指標間呈正相關 )



<圖一>



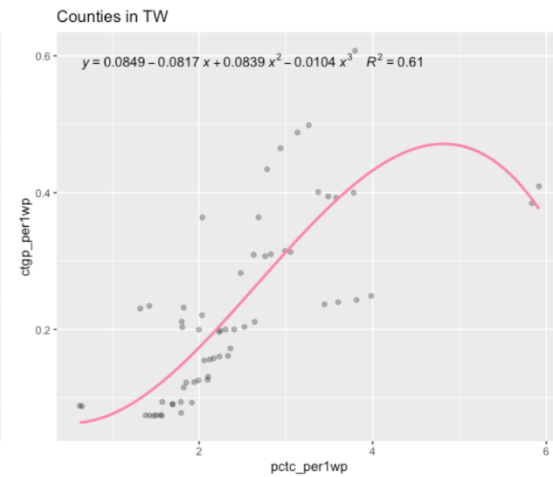
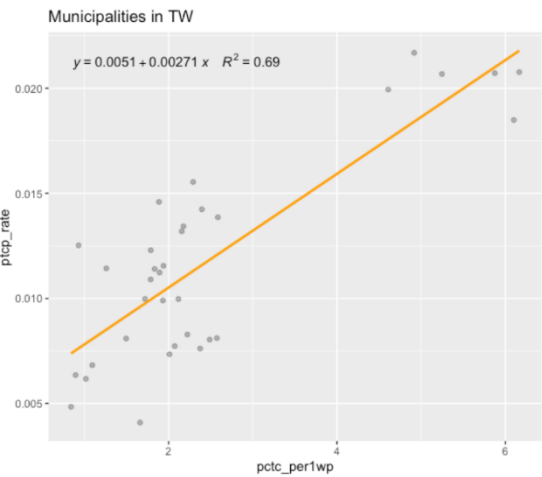
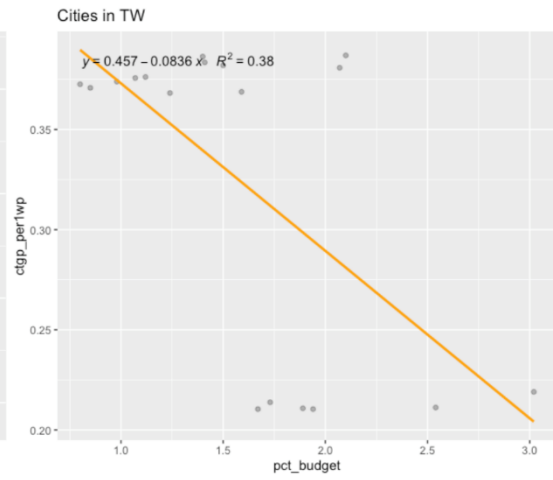
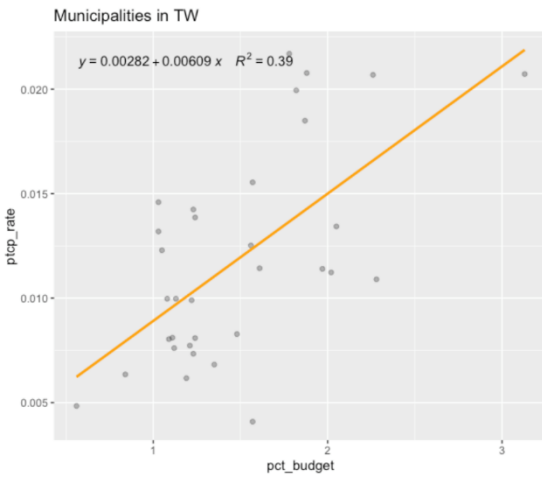
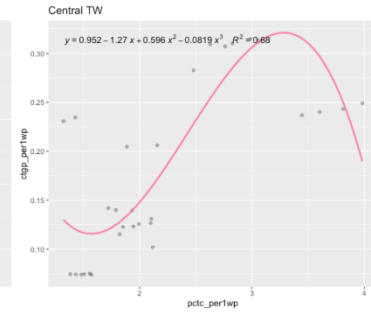
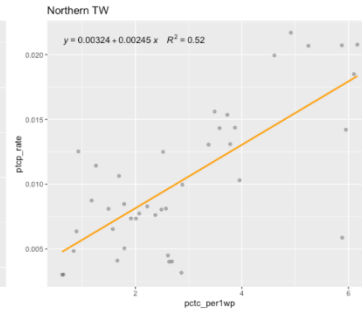
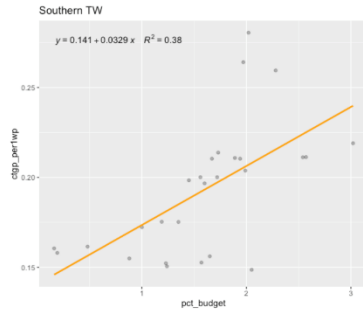
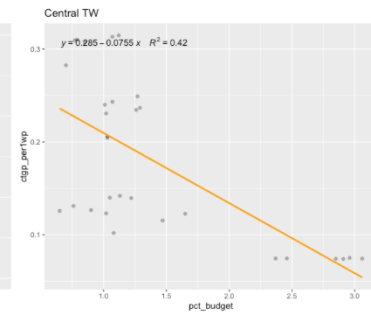
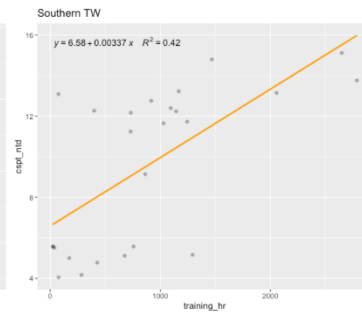
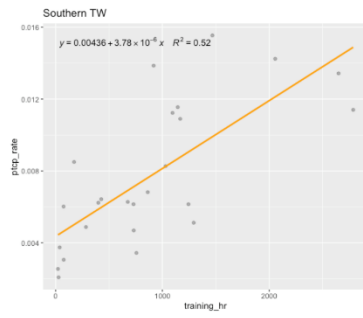
<圖二>

### 3. 尋找最佳回歸方式：

先對 `lm()` 之線性擬合結果使用 `summary()`，以檢驗提高  $x$  值之次方是否可解釋  $y$  值顯著性。後以 partial F test 檢驗提高次方是否比低次方，選取出最適合的回歸方式。結果如下：(字母代表的群組請參考上一步驟之結果)

(1) E,G,H & I：大部分政府投入與民間參與指標間具有相關性的群組，以簡單線性(一次)回歸建立模型較佳，以二次或三次回歸進行擬合無法進一步提高模型之準確度。其中 C & H 呈現線性負相關，顯示中臺灣以及一般市兩群組的民間文化團體數隨著文化預算的增加呈現下降的趨勢，其餘群組之政府投入與民間參與指標間皆呈現正相關。

(2) F & J：以三次多項式回歸建立模型，相較於二次與一次線性回歸，能提高模型之準確度。顯示中臺灣與一般縣兩群組之人均文化團體數量，隨人均文化場地數增加而上升，但在人均文化場地達到一定數量時，人均文化團體數量則會些微下降。



#### 4. 回歸分析部分總結：

以缺失值取代離群值後，依照地理分區與行政層級將臺灣各縣市劃分為數個群組後，分析三項民間文化參與指標與三項政府文化投入指標的相關性。總計 72 個群組中有 10 個群組的決定係數  $R^2 > 0.3$ ，約佔 14%。

進一步統計檢驗後發現，大多數具有相關性的群組，民間文化參與指標與政府文化投入指標間為正相關，且多可用簡單線性回歸建立模型，但有兩個群組則以三次回歸建立模型準確度較高。其中，有兩個群組為簡單線性負相關，其 y 軸皆為人均民間文化團體數，x 軸則為文化預算占比。若「政府投入能提升民間文化參與」的假設是正確的，這或許暗示著該指標並不適合作為民間文化參與的指標。

本次分析只使用了 2012-2017 六年間的資料，在以地理分區與行政層級進行分群後，資料筆數更顯不足。若能獲得更多年的資料，並選取更多相關指標進行分析，將更能看出政府投入與民間參與指標間的相關性，有助於政府文化相關部門準確訂立政策，以提高民眾的文化參與度。

## 4. 組員分工

### 分工-林哲豪：

資料初步分析，完成各指標排名以及文化總排名 挑選特定縣市對歷  
年各指標作圖，並且目測觀察其趨勢、或提出政策可修正方向

### 分工-魏廷宇：

編修文化部文化統計資料，並透過 dplyr 製作成可使用的政府投入  
指標，分別為人才培育總小時數、文化預算佔總預算比例和每萬人  
均展演場地數，以及製作書面報告。

### 分工-姜禹安：

1. 將文化統計資料添加分組欄位、選擇適當方法去除離群值後，以簡單線性回歸之決定係數，篩選出政府投入與民間參與指標間具有相關性的群組。
2. 以不同次方多項式回歸擬合資料點，並以 `summary(lm)` 檢驗提高  $x$  值之次方是否可解釋  $y$  值顯著性，後以 partial F test 檢驗提高次方是否比低次方，選取出最適合的回歸方式，並進行繪圖。
3. 彙整各部分資料，製作展示用海報。



分工-陳靖騏：

透過 dplyr 將資料編修成，方便處理的格式，製作民眾參與的指標，分成人均文化場地營業額，每千人均藝文活動參與次數，每萬人均文化場地組織個數。

## 5. 參考資料

[https://www.flaticon.com/freeicon/dunes\\_882946?term=desert&page=1&position=5](https://www.flaticon.com/freeicon/dunes_882946?term=desert&page=1&position=5)  
[https://stat.moc.gov.tw/HS\\_UserCatalogView.aspx](https://stat.moc.gov.tw/HS_UserCatalogView.aspx)  
<https://datascienceplus.com/outlier-detection-and-treatment-with-r/>