



# 誰是...? 台大學生...

---

論成為台大學生的社會性因素

G16 (沒有名字)  
林泊里 林書敏 傅羿寧 吳俞萱



# 目錄

一、簡介.....	p01.
二、方法.....	p01.
1.資料取得.....	p01.
2.原始碼運作與說明.....	p02.
三、結果.....	p10.
四、討論與貢獻.....	p11.
五、附錄.....	p11.
1.組員分工.....	p11.
2.參考資料.....	p11.

## 一、簡介

基於時常在新聞上或者是耳語間聽聞大家以「台大」作為噱頭抑或為台大學生貼上諸多標籤，於是使我們思考並好奇，在入學成績背後，是否有更多社會性的因素影響我們能否成為台大學生。而當我們搜查並閱讀相關文本包括報導者、天下雜誌等的文章後，我們發現經濟系駱明慶教授分別在2002年推出論文《誰是台大學生——性別、省級與城鄉差異》，之後又在2016年再推出《誰是台大學生2.0——多元入學的影響》，而在拜讀完此兩篇論文後，我們認為其內容豐富且歸納完善，與我們的所求不謀而合，因此我們決定將論文內我們最認同的三大因素「居住地」、「家庭所得」、「雙親教育程度」之數據進一步整理再用本學期所學之ggplot以圖表視覺化呈現之。

## 二、方法

### 1. 資料取得

考量於2002年的《誰是台大學生——性別、省級與城鄉差異》資料太過老舊，我們的資料來源為駱明慶教授的《誰是台大學生2.0——多元入學的影響》。因其中的數據多以文字敘述或數值表格呈現之，且論文已輸出成PDF形式，因此我們手動將我們認為重要的數據輸入至EXCEL再建立CSV檔以供R讀取。

（我們曾於作業期間詢問駱明慶教授能否提供我們當時搜查的原始數據資料，或者是否有更新年度資料能提供，可惜因學籍資料為教授當年向學校申請做研究用，無法轉手第三人，而更新年度的資料則需時間研究，因此並無取得論文外更詳細之資料。）

## 2. 原始碼運作與說明

### A. 各縣市學生成為台大生之機率

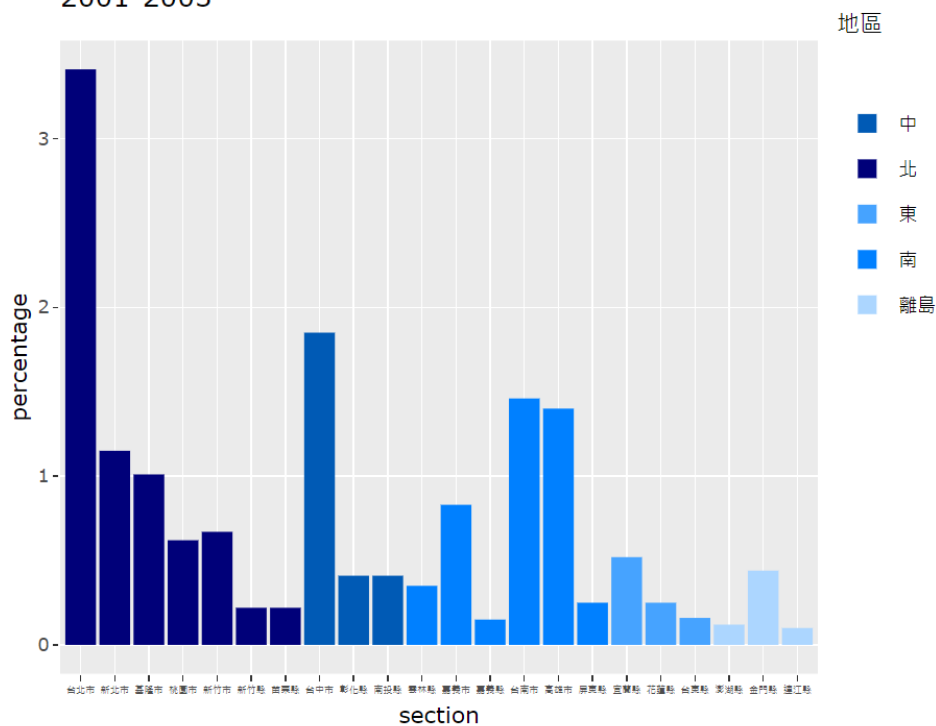
❖ 2001 - 2003

```
df<-tibble(
  section=c(
    "台北市","新北市","高雄市","台中市","桃園市","台南市","彰化縣",
    "基隆市","雲林縣","新竹市","宜蘭縣","南投縣","嘉義市","屏東縣",
    "苗栗縣","新竹縣","花蓮縣","嘉義縣","台東縣","金門縣","澎湖縣",
    "連江縣"),
  percentage=c(3.41,1.15,1.4,1.85,
    0.62,1.46,0.41,1.01,
    0.35,0.67,0.52,0.41,
    0.83,0.25,0.22,0.22,
    0.25, 0.15,0.16,0.44,
    0.12,0.1),
  number=c("北","北","南","中","北",
    "南","中","北","南","北",
    "東","中","南","南","北",
    "北","東","南","東","離島",
    "離島","離島"))

df$section = with(df,factor(section, levels=c("台北市","新北市","基隆市","桃園市",
  "新竹市","新竹縣","苗栗縣","台中市",
  "彰化縣","南投縣","雲林縣","嘉義市",
  "嘉義縣","台南市","高雄市","屏東縣",
  "宜蘭縣","花蓮縣","台東縣","澎湖縣",
  "金門縣","連江縣"))))

picture<-ggplot(data = df)+
  geom_bar(mapping=aes(x=section,y=percentage,fill=number),
    stat="identity")+
  labs(title='2001-2003')+
  theme(axis.text.x=element_text(size=4.6))+
  scale_fill_manual(name="地區",values=c("#005AB5", "#000079", "#46A3FF", "#0080FF", "#ACD6FF"))
ggplotly(picture)
```

2001-2003



## ❖ 2004 - 2006

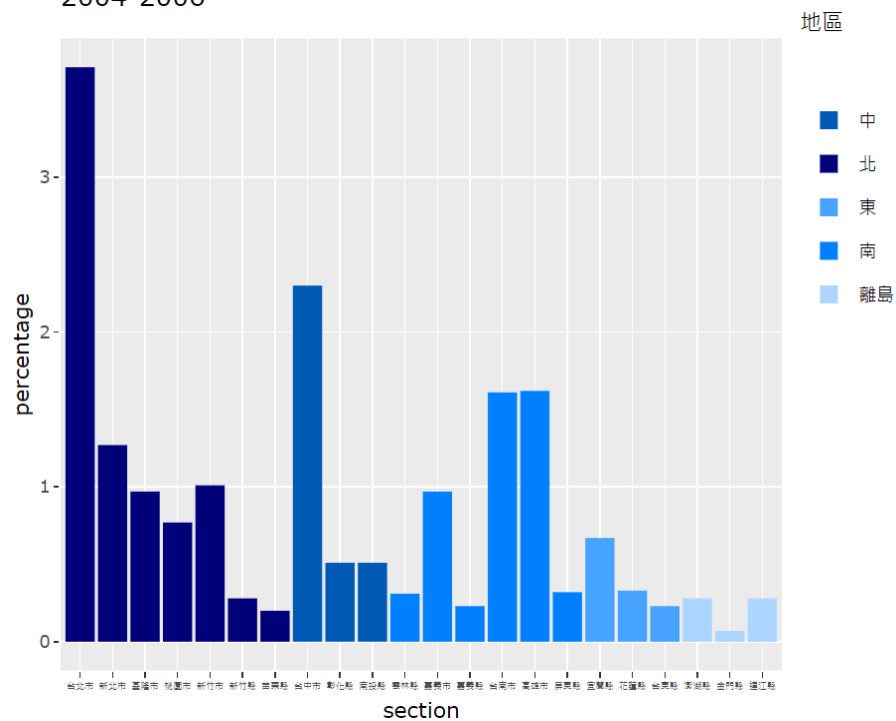
```
df_04_06<-data.frame(section=c("台北市","新北市","高雄市","台中市","桃園市","台南市","彰化縣",
,"基隆市","雲林縣","新竹市","宜蘭縣","南投縣","嘉義市","屏東縣",
,"苗栗縣","新竹縣","花蓮縣","嘉義縣","台東縣","金門縣","澎湖縣",
,"連江縣"),
percentage=c(3.71,1.27,1.62,2.3,0.77,1.61,
0.51,0.97,0.31,1.01,0.67,0.51,
0.97,0.32,0.2,0.28,0.33,0.23,
0.23,0.07,0.28,0.28),
number=c("北","北","南","中","北",
"南","中","北","南","北",
"東","中","南","南","北",
"北","東","南","東","離島",
"離島","離島"))

df_04_06$section = with(df, factor(section, levels=c(c("台北市","新北市","基隆市","桃園市",
,"新竹市","新竹縣","苗栗縣","台中市",
,"彰化縣","南投縣","雲林縣","嘉義市",
,"嘉義縣","台南市","高雄市","屏東縣",
,"宜蘭縣","花蓮縣","台東縣","澎湖縣",
,"金門縣","連江縣"))))

picture_04_06<-ggplot(data = df_04_06)+
  geom_bar(mapping=aes(x=section,y=percentage,fill=number),
  stat = "identity")+
  labs(title='2004-2006')+
  theme(axis.text.x=element_text(size=4.6))+
  scale_fill_manual(name="地區",values=c("#005AB5", "#000079", "#46A3FF", "#0080FF", "#ACD6FF"))

ggplotly(picture_04_06)
```

2004-2006



## ❖ 2007 - 2010

```
df_07_10<-data.frame(section=c("台北市","新北市","高雄市","台中市","桃園市","台南市","彰化縣",
,"基隆市","雲林縣","新竹市","宜蘭縣","南投縣","嘉義市","屏東縣",
,"苗栗縣","新竹縣","花蓮縣","嘉義縣","台東縣","金門縣","澎湖縣",
,"連江縣"),
percentage=c(3.68,1.37,1.67,2.19,
0.77,1.7,0.52,1.14,
0.36,1.2,0.77,0.52,
1.24,0.28,0.33,0.48,
0.33,0.36,0.22,0,
0.16, 0.59),
number=c("北","北","南","中","北",
"南","中","北","南","北",
"東","中","南","南","北",
"北","東","南","東","離島",
"離島","離島"))

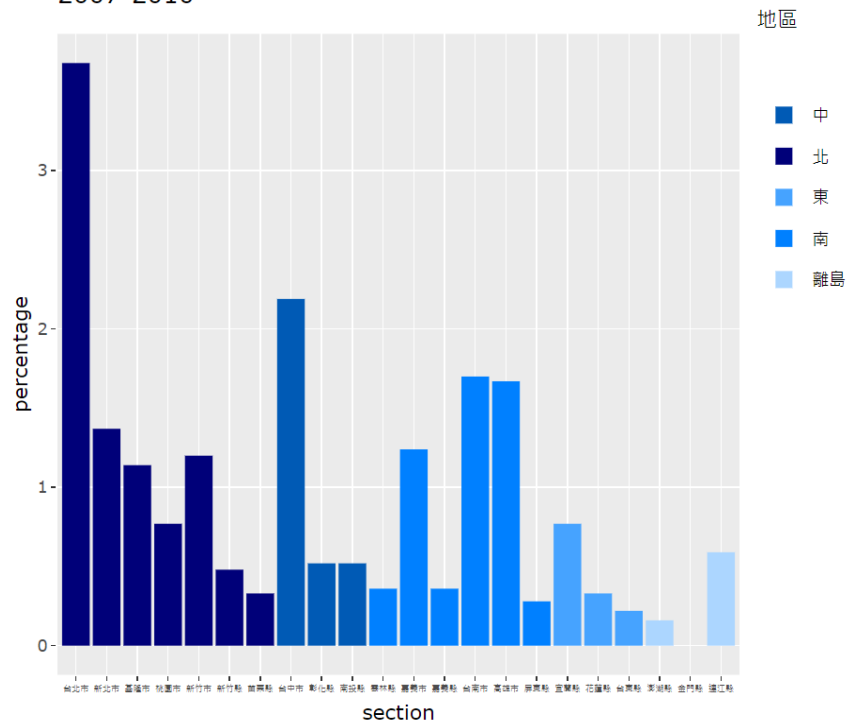
df_07_10$section = with(df, factor(section, levels=c(c("台北市","新北市","基隆市","桃園市",
,"新竹市","新竹縣","苗栗縣","台中市",
,"彰化縣","南投縣","雲林縣","嘉義市",
,"嘉義縣","台南市","高雄市","屏東縣",
,"宜蘭縣","花蓮縣","台東縣","澎湖縣",
,"金門縣","連江縣"))))

picture_07_10<-ggplot(data = df_07_10)+
geom_bar(mapping=aes(x=section,y=percentage,fill=number),
stat = "identity")+

labs(title='2007-2010')+
theme(axis.text.x=element_text(size=4.6))+
scale_fill_manual(name="地區",values=c("#005AB5", "#000079", "#46A3FF", "#0080FF", "#ACD6FF"))

ggplotly(picture_07_10)
```

## 2007-2010

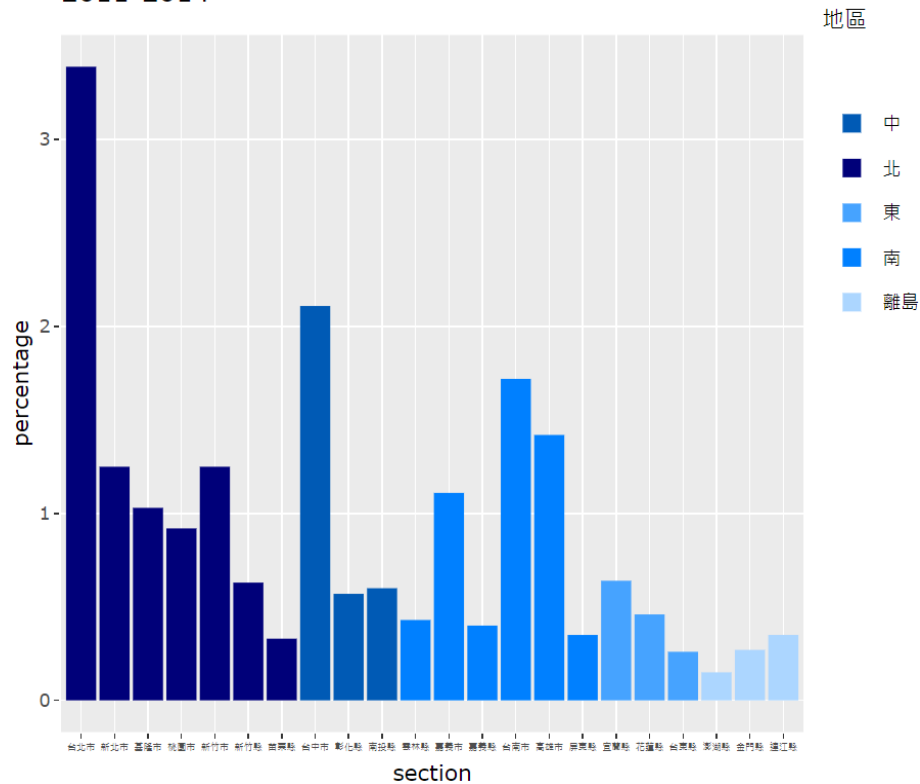


## ❖ 2011 - 2014

```
df_11_14<-data.frame(section=c("台北市","新北市","高雄市","台中市","桃園市","台南市","彰化縣",
"基隆市","雲林縣","新竹市","宜蘭縣","南投縣","嘉義市","屏東縣",
"苗栗縣","新竹縣","花蓮縣","嘉義縣","台東縣","金門縣","澎湖縣",
"連江縣"),
percentage=c(3.39,1.25,1.42,2.11,0.92,1.72,
0.57,1.03,0.43,1.25,0.64,0.6,
1.11,0.35,0.33,0.63,0.46,0.4,
0.26,0.27,0.15,0.35),
number=c("北","北","南","中","北",
"南","中","北","南","北",
"東","中","南","南","北",
"北","東","南","東","離島",
"離島","離島"))
df_11_14$section = with(df, factor(section, levels=c(c("台北市","新北市","基隆市","桃園市",
"新竹市","新竹縣","苗栗縣","台中市",
"彰化縣","南投縣","雲林縣","嘉義市",
"嘉義縣","台南市","高雄市","屏東縣",
"宜蘭縣","花蓮縣","台東縣","澎湖縣",
"金門縣","連江縣"))))

picture_11_14<-ggplot(data = df_11_14)+
  geom_bar(mapping=aes(x=section,y=percentage,fill=number),stat = "identity")+
  labs(title='2011-2014')+
  theme(axis.text.x=element_text(size=4.6))+
  scale_fill_manual(name="地區",values=c("#005AB5", "#000079", "#46A3FF", "#0080FF", "#ACD6FF"))
ggplotly(picture_11_14)
```

2011-2014



此凸顯是各縣市學生成為台大生的機率。分別將各縣市依照北、中、南、東與離島五個區域依序分為五個色塊。其中可見台北市比例最高，而其他區域的縣市中則是六都的機率較高，顯示城鄉差距依然存在。

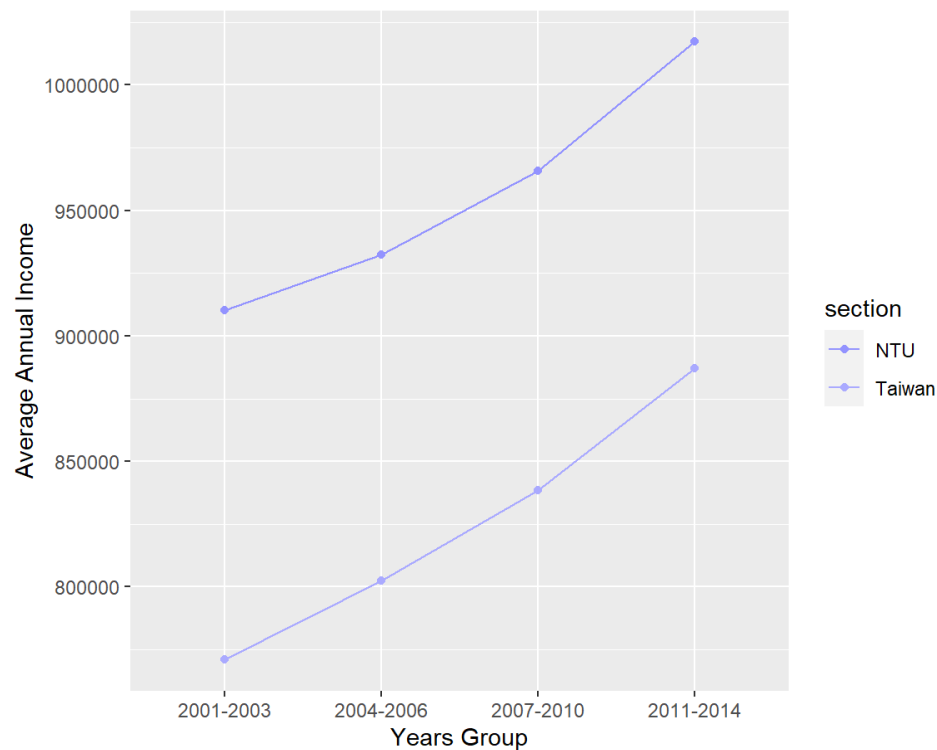
## B. 家庭平均年所得數值比較

```
```{r}
#所需套件
library(dplyr)
library(ggplot2)
library(ggrepel)

#匯入數據資料
mydata <- readr::read_csv("/Users/linshumin/Desktop/學分大作戰/R
資料庫/期末專題/台大全國平均所得比較.csv")

#繪圖
#以X軸為觀察年度，Y軸為年所得金額，使用geom_line繪製折線圖並搭配geom_point標記觀察年度的數值點
#運用color與scale_color_manual將台大學生家庭與全國家庭的平均年所得用色碼錶之顏色代碼區分
#使用labs為X軸與Y軸命名
aai <- ggplot(data = mydata) +
  geom_line(aes(x = Years, y = 台大平均年所得, group = 1, color = "NTU")) +
  geom_point(aes(x = Years, y = 台大平均年所得, color = "NTU")) +
  geom_line(aes(x = Years, y = 全國平均年所得, group = 1, color = "Taiwan")) +
  geom_point(aes(x = Years, y = 全國平均年所得, color = "Taiwan")) +
  scale_color_manual(values = c("#9393FF", "#AAAAFF"), name = "section") +
  labs(x = "Years Group", y = "Average Annual Income")

#呈現
aai
```
```



深紫色線為台大學生家庭平均年所得，而淡紫色線為全國家庭平均年所得，比較2001年至2014年各年間台大學生家庭年所得一直都高於全國家庭平均年所得，且比值約為1.15。



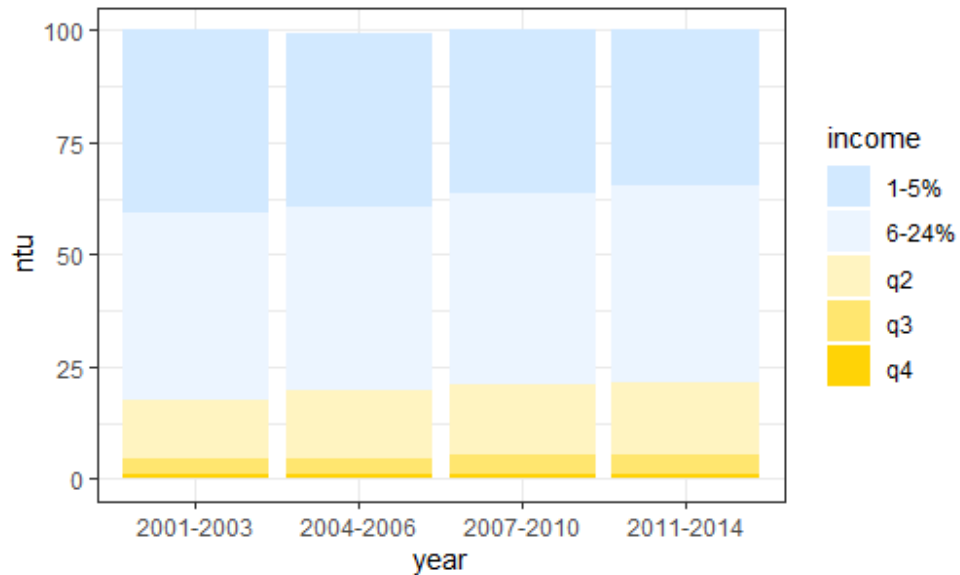
## C. 家庭所得國內排名

```
##{r}
# 所需套件
library(readxl)
library(ggplot2)
library(plotly)
library(plyr)

# 匯入數據資料
dataset <- read_excel("C:/Users/Irene/Desktop/畫圖.xlsx")

# 繪圖
# 以X軸為觀察年度、y軸為台大學生占比，使用geom_bar繪製長條圖
# 並依據家庭所得在台之排名(百分比)呈現不同顏色
# 運用scale_fill_manual使用色碼表之顏色代碼自行搭配圖表色彩
f <- ggplot(data = dataset) +
  geom_bar(mapping = aes(x = year, y = ntu, fill = income), stat = "identity") +
  scale_fill_manual(values=c("#D2E9FF", "#ECF5FF", "#FFF4C1", "#FFE66F", "#FFD306"))

# 呈現
ggplotly(f)
```



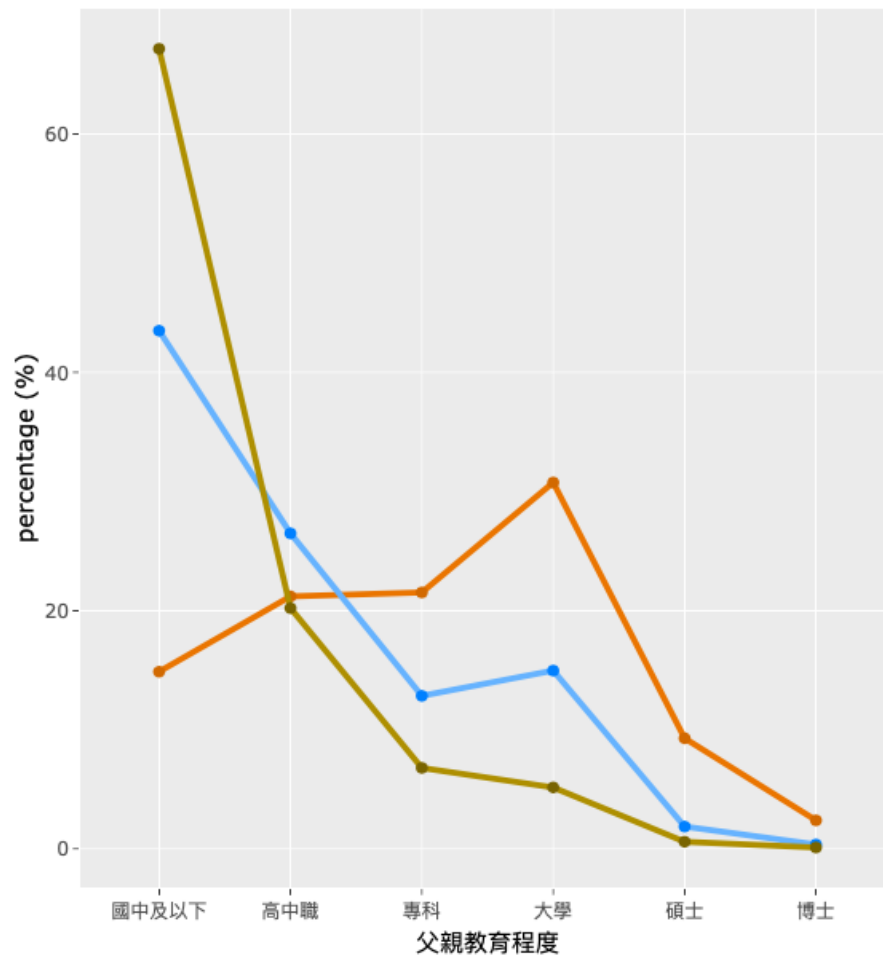
縱軸為台大生占比、橫軸為觀察年度，圖中之色塊呈現家庭所得在台之排名(以百分比計算)，可見約有四成台大學生之家庭所得在全台前5%，而有超過八成之台大學生之家庭所得在全台前25%以內，因此可推知，所得在前25%以外的全台學生家庭只能搶剩下不到19%的台大入學名額。

## D. 雙親教育程度

### ❖ 父親教育程度

```
### 父親教育程度

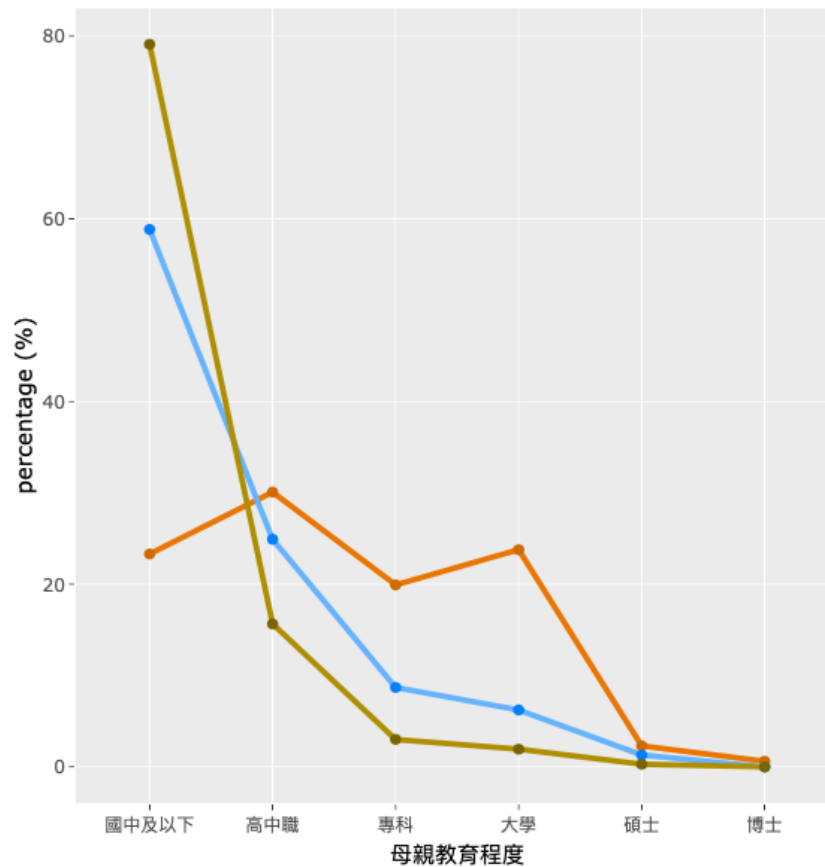
```{r}
p <- ggplot(DadData) +
  geom_line(mapping = aes(x = degree, y = law, group=1), stat = "identity",
    color="#EA7500", size=1) +
  geom_line(mapping = aes(x = degree, y = students, group=1), stat =
    "identity", color="#66B3FF", size=1) +
  geom_line(mapping = aes(x = degree, y = all, group=1), stat = "identity",
    color="#AE8F00", size=1) +
  labs(x = "父親教育程度", y = "percentage (%)") +
  geom_point(mapping = aes(x = degree, y = law, group=1), color =
    "#D26900") +
  geom_point(mapping = aes(x = degree, y = students, group=1), color =
    "#0080FF") +
  geom_point(mapping = aes(x = degree, y = all, group=1), color =
    "#796400") +
  ggplotly(p)
```



## ❖ 母親教育程度

```
### 母親教育程度

```{r}
q <- ggplot(MomData) +
  geom_line(mapping = aes(x = degree, y = law, group=1), stat = "identity",
    color="#EA7500", size=1) +
  geom_line(mapping = aes(x = degree, y = students, group=1), stat =
    "identity", color="#66B3FF", size=1) +
  geom_line(mapping = aes(x = degree, y = all, group=1), stat = "identity",
    color="#AE8F00", size=1) +
  labs(x = "母親教育程度", y = "percentage (%)") +
  geom_point(mapping = aes(x = degree, y = law, group=1), color =
    "#D26900") +
  geom_point(mapping = aes(x = degree, y = students, group=1), color =
    "#0080FF") +
  geom_point(mapping = aes(x = degree, y = all, group=1), color =
    "#796400")
ggplotly(q)
```



橘色線為台大法學院一到四年級學生數據；藍線為18~22歲大學生；土色線為18~22歲全國人民。不論是母親還是父親，橘色線(台大學生)和其他皆有明顯的差別。台大學生的父親42.42%是大學以上學歷，母親則有26.69%。

### 三、結果

「居住地」、「家庭所得」與「雙親教育程度」和能否成為台大學生據正相關。

### 四、討論與貢獻

綜觀上述，我們可得知能否成為台大生的確存在許多社會性的因素，其中我們提出討論的「居住地」、「家庭所得」與「雙親教育程度」在數據上證實其據正相關。然而深入比較有依年份區分之「各縣市成為台大學生機率」、「家庭所得排名」與「家庭年所得數值」數據，可見這些社會性的因素與成為台大生的相關性正在逐漸微幅下降，而再參考了駱明慶教授的《誰是台大學生2.0——多元入學的影響》後，我們也認同其所述，此現象因與越來越多元的入學管道相關，如：繁星計劃的出現，但是從僅「逐漸微幅下降」的數據中也可推知：追求人人機會平等的升學措施仍難以平衡階級複製的現象。此外，我們也可從已有之數據去做更深入的相關研究，例如，從各縣市成為台大學生機率之數據中可見，非位居北部也非直轄市的嘉義在此機率排名中卻一直維持大約第六的名次，與預期不同，執得深入研究。

然而，雖然我們認同區域均衡與消弭貧富差距的重要性，但是我們也同時認為為了應應此況而產生的「同分不同質」之現象的公平性有待商榷，因此，比起增設不同的入學管道，根本性的改善不同地區間的教學品質，提升弱勢地區之學習水準會是更妥善的解決之道。

## 五、附錄

### 1. 組員分工

林泊里	搜尋、整理論文資料、繪圖－各縣市學生成為台大生之機率、ppt製作與口頭報告
林書敏	搜尋、整理論文資料、繪圖－家庭平均年所得數值比較、書面報告撰寫整理
傅羿寧	搜尋、整理論文資料、繪圖－家庭所得國內排名、網頁製作
吳俞萱	搜尋、整理論文資料、繪圖－雙親教育程度、網頁製作

### 2. 參考資料

- 駱明慶（2002），《誰是台大學生——性別、省級與城鄉差異》，瀏覽於 <https://homepage.ntu.edu.tw/~luohm/NTU.pdf>
- 駱明慶（2016），《誰是台大學生2.0——多元入學的影響》，瀏覽於 <http://www.econ.ntu.edu.tw/ter/new/data/new/TER46-1/TER461-2.pdf>
- 天下雜誌，林倖妃、程晏鈴（2016），數字告訴你誰有資格上台大？，瀏覽於 <https://topic.cw.com.tw/2016multi-star/>
- 關鍵評論網，白經濟（2017），台大怎麼上：誰是台大生2.0，瀏覽於 <https://www.thenewslens.com/article/85532>
- 報導者，RedHerrings 白經濟原作者網站（2018），【數字說話】教育能翻轉人生嗎——高教的反向重分配現象，瀏覽於 <https://www.twreporter.org/a/education-opportunity-inequality-across-income>
- 教育部（2020），中華民國109年版教育統計2020，瀏覽於 [https://stats.moe.gov.tw/bookcase/Education\\_Statistics/109/index.html](https://stats.moe.gov.tw/bookcase/Education_Statistics/109/index.html)