# 107-2 大數據分析方法 作業一

put your name here

搞不清楚各行各業的薪資差異嗎? 念研究所到底對第一份工作的薪資影響有多大? CP值高嗎? 透過分析**初任人員平均經常性薪資**- [開放資料連結](https://data.gov.tw/dataset/6647)，可初步了解台灣近幾年各行各業、各學歷的起薪。

## 比較103年度和106年度大學畢業者的薪資資料

### 資料匯入與處理

library(jsonlite)  
library(dplyr)  
library(readr)  
s106<-read\_csv("/Users/chris/Downloads/A17000000J-020066-Qod/106年各教育程度別初任人員每人每月經常性薪資─按大職類分.csv")  
s103<-read\_csv("/Users/chris/Downloads/A17000000J-020066-Qod/103年各教育程度別初任人員經常性薪資─按大職類分.csv")  
s106$`大學-薪資`<-as.numeric(s106$`大學-薪資`)  
s103$`大學-薪資`<-as.numeric(s103$`大學-薪資`)  
s103$大職業別<-gsub("、","-",s103$大職業別)  
New\_table<-inner\_join(s106,s103,by="大職業別")  
  
sJ103 <- read.csv("/Users/chris/Downloads/A17000000J-020066-Qod/103年各教育程度別初任人員經常性薪資─按大職類分.csv",stringsAsFactors = F)  
sJ104 <- read.csv("/Users/chris/Downloads/A17000000J-020066-Qod/104年各教育程度別初任人員經常性薪資─按大職類分.csv",stringsAsFactors = F)  
sJ105 <- read.csv("/Users/chris/Downloads/A17000000J-020066-Qod/105年各教育程度別初任人員每人每月經常性薪資─按大職類分.csv",stringsAsFactors = F)  
sJ106 <- read.csv("/Users/chris/Downloads/A17000000J-020066-Qod/106年各教育程度別初任人員每人每月經常性薪資─按大職類分.csv",stringsAsFactors = F)  
d103 <- sJ103[,c(2,12)]  
d104 <- sJ104[,c(2,12)]  
d105 <- sJ105[,c(2,12)]  
d106 <- sJ106[,c(2,12)]  
d103\_106 <- inner\_join(d103, d104, by="大職業別")  
d103\_106 <- inner\_join(d103\_106, d105, by="大職業別")  
d103\_106 <- inner\_join(d103\_106, d106, by="大職業別")  
names(df\_103\_106) <- c("大職業別","College\_f/m\_103","College\_f/m\_104","College\_f/m\_105","College\_f/m\_106")  
d103\_106[is.na(d103\_106)] <- 0  
d103\_106$`College\_f/m\_103` <- as.numeric(d103\_106$`College\_f/m\_103`)  
d103\_106$`College\_f/m\_104` <- as.numeric(d103\_106$`College\_f/m\_104`)  
d103\_106$`College\_f/m\_105` <- as.numeric(d103\_106$`College\_f/m\_105`)  
d103\_106$`College\_f/m\_106` <- as.numeric(d103\_106$`College\_f/m\_106`)

#fromjson()  
#inner\_join()

### 106年度薪資較103年度薪資高的職業有哪些?

New\_table$precent<-(New\_table$`大學-薪資.x`/New\_table$`大學-薪資.y`)#兩年度相除  
New\_table<-arrange(New\_table,desc(precent)) #遞減排序  
A<-filter(New\_table,precent>1) %>% arrange(desc(precent)) #排序  
head(A$大職業別,10) #前十資料

### 提高超過5%的的職業有哪些?

A<-filter(A,precent>1.05) %>% arrange(desc(precent))#遞減排序顯示超過5％職業  
A$大職業別#列出超過5％職業

* 有 “其他服務業-技術員及助理專業人員”(1.131278)、“住宿及餐飲業-服務及銷售工作人員”(1.129498)、“用水供應及污染整治業-技術員及助理專業人員”(1.129402)、“營造業-服務及銷售工作人員”(1.109005)、“其他服務業-專業人員”(1.103448) ### 主要的職業種別是哪些種類呢?

filter\_1<-A[grepl("-",A$大職業別),]  
filter\_1$new<-strsplit(filter\_1$大職業別,"-")  
temp<-strsplit(filter\_1$大職業別,"-")  
table(unlist(lapply(temp,"[",1))) ##出現次數，次數多的就重要唄OwO

* 主要為教育服務業(5)、用水供應及污染整治業(5)、資訊及通訊傳播業(5)、服務業部門(4)、其他服務業(4)、營造業(3)、運輸及倉儲業(3)、住宿及餐飲業(3)

## 男女同工不同酬現況分析

男女同工不同酬一直是性別平等中很重要的問題，分析資料來源為103到106年度的大學畢業薪資。

### 103到106年度的大學畢業薪資資料，哪些行業男生薪資比女生薪資多?

d103\_106mf <- arrange(d103\_106,d103\_106$`College\_f/m\_103`)%>%filter(`College\_f/m\_103`>0)  
head(d103\_106mf,10)#要錢還是要臉XD

* 取前十筆排序(其他服務業-技術員及助理專業人員、礦業及土石採取業-服務及銷售工作人員、營造業、教育服務業-事務支援人員、教育服務業、其他服務業、礦業及土石採取業、營造業-專業人員、其他服務業-服務及銷售工作人員、營造業-事務支援人員)在 **大學畢業薪資男生比女生多**這麼多，我覺得可能發薪水的根本不覺得男女同工了吧= =?能少發點就少發點

### 哪些行業女生薪資比男生薪資多?

d103\_106fm <- arrange(d103\_106,desc(d103\_106$`College\_f/m\_103`))  
head(df\_103\_106fm,10)#說實話我覺得這啥都跑不出來(因為表格不包含(消音)職業= =)

* 在105年度資料中只有**金融及保險業-專業人員** 女生薪資比男生多(100.11) 你成功地引起了我的注意= =+

## 研究所薪資差異

以106年度的資料來看，哪個職業別念研究所最划算呢 (研究所學歷薪資與大學學歷薪資增加比例最多)?

s106$`大學-薪資`<-as.numeric(s106$`大學-薪資`)  
s106$`研究所及以上-薪資`<-as.numeric(s106$`研究所及以上-薪資`)  
s106$precent<-(s106$`研究所及以上-薪資`/s106$`大學-薪資`)  
s106<- arrange(s106,desc(precent))  
head(s106$大職業別,10)

* 以下是研究所薪資大於大學薪資增加%數多前十名： **“礦業及土石採取業-事務支援人員”、“專業\_科學及技術服務業”、“其他服務業-技術員及助理專業人員”、“專業\_科學及技術服務業-事務支援人員”、“批發及零售業”、“製造業”、“藝術\_娛樂及休閒服務業-事務支援人員”、“工業部門”、“工業及服務業部門”、“服務業部門”**

## 我有興趣的職業別薪資狀況分析

### 有興趣的職業別篩選，呈現薪資

B<-filter(s106,大職業別 %in% c("藝術、娛樂及休閒服務業-技術員及助理專業人員","藝術、娛樂及休閒服務業-專業人員","金融及保險業-專業人員"))%>%select(大職業別,starts\_with("大學-薪資"),starts\_with("研究所及以上-薪資"))  
head(B)

* 以下是我較有興趣的職業別：**藝術、娛樂及休閒服務業-技術員及助理專業人員、藝術、娛樂及休閒服務業-專業人員、金融及保險業-專業人員** 結果跟我想得有些不同，但這畢竟就不是很穩定的工作，而金融及保險業我覺得可能工作經驗比學歷更重要0.0?

### 這些職業別研究所薪資與大學薪資差多少呢？

B$percent<-round(((B$`研究所及以上-薪資`/B$`大學-薪資`)-1)\*100,2)  
B$diff<-B$`研究所及以上-薪資`-B$`大學-薪資`  
head(B)

* 在diff欄位為研究所與大學的差異，而percent欄位為兩者差異%，結果顯示低於5%，實際薪資成長不到1000塊，所以我覺得研究所念不念差別不大(-~-)