1. 爬蟲

先至「104」家教網上勾選欲選取的條件,從104家教網上看到的資料如下,包括聯絡人,科目,對象,上課地點,經驗,時薪,將符合條件之所有資料透過R語言網路 爬蟲下載為csv檔。



104 家教網顯示資料

$contact perso\hat{\boldsymbol{n}}$	region [‡]	grade [‡]	salary [‡]	subject $^{\circ}$
田先生	一百十十十二十二十二十二十二十二十二十二十二十二十二十二十二十二十二十二十二十二	社會人士	無經驗可 \$1000	英文
張小姐	台北市士林區	學齡前兒童	至少兩年 \$1200	英文
謝小姐	台北市信義區	社會人士	至少一年 \$700	英文
張小姐	台北市士林區	國小四年級	至少一年 \$400	英文
謝小姐	台北市中正區	社會人士	無經驗可 \$1150	英文
禁先生	台北市士林區	社會人士	無經驗可 \$800	英文
賴小姐	台北市北投區	高中二年級	至少五年 \$700	英文
李小姐	台北市文山區	社會人士	無經驗可 \$800	英文
溫小姐	台北市大同區	社會人士	至少三年 \$650	英文
陳先生	台北市中山區	國小四年級	一年以下 \$500	英文
吳小姐	台北市中山區	國中二年級	至少四年 \$1400	英文
楊小姐	台北市萬華區	國中三年級	一年以下 \$800	英文
陳小姐	台北市内湖區	國中二年級	無經驗可 \$500	英文
黃先生	台北市士林區	大專生以上	無經驗可 \$900	英文
王小姐	台北市高華區	社會人士	至少一年 \$1500	英文

下載後的 data.frame 資料格式(.csv 檔)

下載後的 data.frame 還要進一步處理,像是在資料的 salary 欄位,需要把經驗和願付價格分開,使用 strsplit 將 salary 欄中的資料以\$分成兩個字串:

