

109-1: EE4052

通識課程： 計算機程式設計 之旅

Computer Programming

Unit 09: 檔案資料輸入與輸出

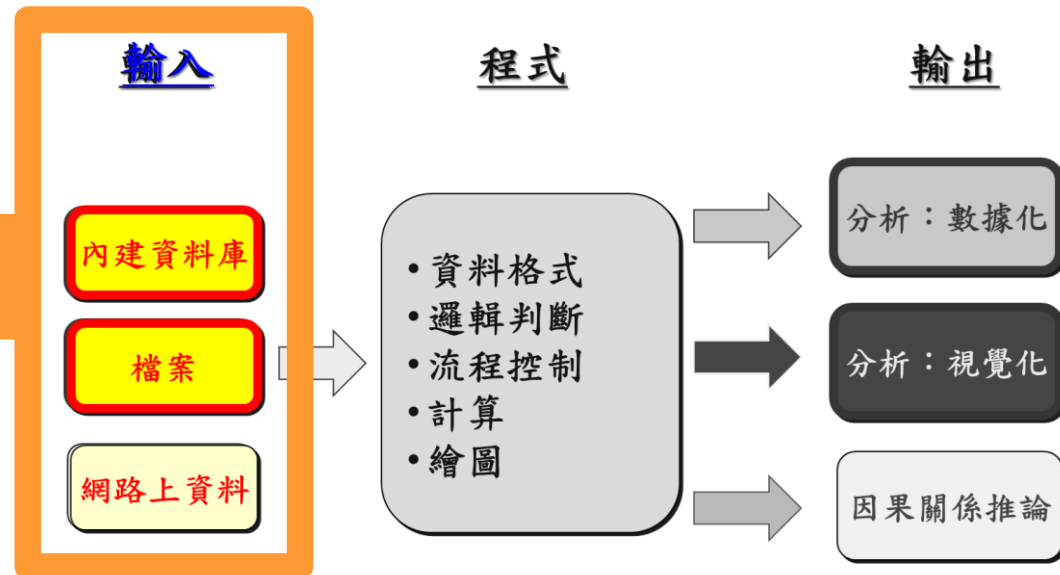
連 豐 力

臺大電機系

Sep 2020 - Jan 2021

課程主題進度

- **U01:** 課程介紹：討論主題，作業，報告，進行方式
- **U02:** 主題，案例，程式，演算法，資源
- **U03:** 設定軟體 **R** 與 **Rstudio**
- **U04:** 數據處理與繪圖指令功能
- **U05:** 資料類別與基本運算
- **U06:** 邏輯判斷與流程控制
- **U07:** 函數：計算與排序
- **U08:** 多維度資料格式
- **U09:** 檔案資料輸入與輸出
- **U10:** 繪圖功能與文字
- **U11:** 多重繪圖與顏色
- **U12:** 函數：動畫與動作
- **U13:** 探索性資料分析
- **U14:** 資料間的相關性
- **U15:** 資料連結分析



作業

HW07：檔案資料輸入與輸出

On 11/18, 2020

- 假設您預計購買一個手機，
從網路上找到一些手機的資料，
如下表所示：

廠牌	Brand	Apple	HTC	HTC	ASUS	ASUS
型號	Model	iPhone7	OneM8	OneS9	ZenFoneDeluxe	ZenFoneZoom
價格 (元)	Price	24500	21900	9990	8990	15990
螢幕 (吋)	Screen	4.7	5	5	5.5	5.5
重量 (克)	Weight	138	160	158	170	185
記憶體 (GB)	GB	32	16	32	16	64
日期 (年月日)	Date	20160916	20140328	20160617	20150827	20151201

- 這筆資料，已經整理成四個檔案：(請到課程網站下載到您的工作目錄)
 - [HW07_Phone.txt](#), [HW07_Phone.csv](#), [HW07_Phone.web](#), [HW07_Phone.html](#)

HW07：檔案資料輸入與輸出

On 11/18, 2020

- 編輯一個程式於 .R 檔，完成下面的工作：
 - 用 `read.table` 取讀取 HW07_Phone.txt 中的資料，指定到：`myPhoneTxt`
 - 用 `read.csv` 取讀取 HW07_Phone.csv 中的資料，指定到：`myPhoneCsv`
 - 用 `read.table` 去取讀取下面網址的檔案，指定到：`myPhoneWebTab`
`https://goo.gl/XHtXw8`
 - 用 `read.csv` 去取讀取下面網址的檔案，指定到：`myPhoneWebCsv`
`https://goo.gl/XHtXw8`
 - 用 `readLines()` 去取讀取檔案，指定到：`myPhoneHTML`
 - 比較一下：`myPhoneTxt`，`myPhoneCsv`，`myPhoneWebTab` 與 `myPhoneWebCsv`，`myPhoneHTML`，這幾個物件，看看：內容是否相同？格式是否相同？
 - 如果不同，可否重新讀入（例如：設定不同的參數），或者重新指定到新的物件，使得這四個的內容與格式都相同。
 - 請利用讀進來的數據，建立一個 5x3 的矩陣 (matrix)：`Number`，放置五個手機的價格，螢幕，重量三種資料。
- 把執行的過程，以及產生的數據等，整理到報告檔 (pdf)。

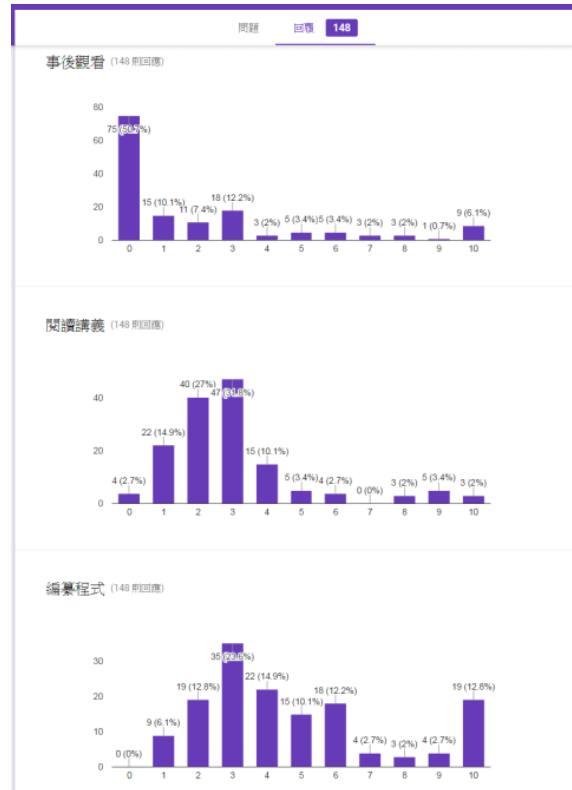
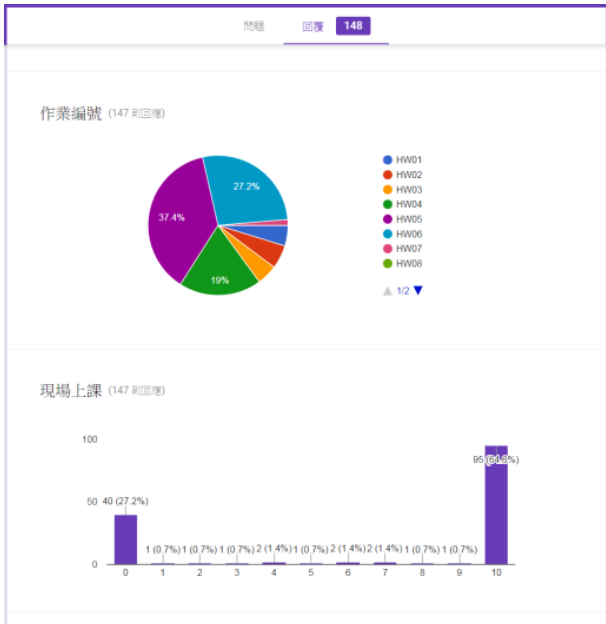
HW07：檔案資料輸入與輸出

On 11/18, 2020

- 繳交下面檔案，檔案名稱：[HW07_學號_關鍵字.xxx](#)
- 您可以繳交下面格式之中的任何一種格式的檔案：
 - 程式檔：[HW07_B01921001_ReadData.R](#)
 - 程式與結果檔：[HW07_B01921001_ReadData.Rmd](#)
 - 或轉成：[HW07_B01921001_ReadData.html](#)
 - 報告檔案：[HW07_B01921001_Phone.pdf](#)
- 繳交方式與期限：
 - 上傳檔案到：<https://cool.ntu.edu.tw>
 - 繳交期限：[11/23 \(Mon\), 11pm 以前](#)
- 學習方式：
 - 請至下面網址輸入此次的學習方式所花的時間：
 - <https://forms.gle/TGYXj2uLoL4HwqLHA>

HW07++A : 課程學習時間資料分析

課程學習時間資料分析



- 1 時間戳記, 學號, 系級, 作業編號, 現場上課, 同步觀看, 事後觀看, 閱讀講義, 編寫程式, 整理作業
- 2 2017/3/22 上午 7:41:30, B05921005, 機械三, HW02, 4, 5, 1, 3, 5, 2
- 3 2017/3/25 下午 3:16:19, B04202010, 物理二, HW04, 10, 0, 0, 3, 4, 3
- 4 2017/3/25 下午 4:51:44, B05104026, 地質一, HW04, 10, 0, 6, 6, 10, 6
- 5 2017/3/25 下午 5:51:29, B02703024, 財金四, HW01, 7, 0, 3, 9, 9, 9
- 6 2017/3/25 下午 5:52:00, B02703024, 財金四, HW02, 5, 0, 4, 9, 10, 10
- 7 2017/3/25 下午 5:52:27, B02703024, 財金四, HW03, 8, 0, 7, 9, 10, 10
- 8 2017/3/25 下午 5:52:58, B02703024, 財金四, HW04, 0, 0, 8, 10, 10, 10
- 9 2017/3/26 上午 10:59:38, B04902076, 資工二, HW04, 0, 9, 0, 1, 2, 3
- 10 2017/3/26 上午 11:38:17, B05409004, 職治一, HW04, 10, 0, 2, 4, 4, 2
- 11 2017/3/26 下午 1:30:55, B03705043, 資管三, HW04, 0, 10, 0, 2, 3, 6
- 12 2017/3/26 下午 2:15:34, B03601043, 農藝三, HW01, 3, 0, 1, 2, 4, 6
- 13 2017/3/26 下午 2:16:15, B03601043, 農藝三, HW02, 1, 0, 4, 4, 7, 10
- 14 2017/3/26 下午 2:16:45, B03601043, 農藝三, HW03, 0, 4, 0, 1, 3, 4
- 15 2017/3/26 下午 2:17:30, B03601043, 農藝三, HW04, 0, 0, 5, 1, 4, 3
- 16 2017/3/26 下午 4:36:43, B05603043, 農化一, HW04, 0, 10, 0, 2, 1, 6
- 17 2017/3/26 下午 7:52:40, B02B02054, 生技四, HW05, 10, 0, 3, 3, 3, 3
- 18 2017/3/26 下午 7:56:11, B02B02054, 生技四, HW04, 10, 3, 3, 3, 2, 3
- 19 2017/3/26 下午 10:18:00, B02203013, 化學四, HW04, 0, 0, 0, 1, 1, 3
- 20 2017/3/26 下午 11:04:31, B03105034, 人類三, HW04, 6, 6, 3, 3, 4, 4
- 21 2017/3/26 下午 11:07:20, B01B02057, 生技四, HW04, 10, 0, 2, 4, 6, 5
- 22 2017/3/26 下午 11:35:55, B03208008, 地理三, HW04, 0, 0, 10, 2, 6, 10
- 23 2017/3/28 上午 8:45:37, B04902058, 資工二, HW04, 0, 10, 3, 4, 5, 4
- 24 2017/3/28 上午 9:05:30, B01502013, 機械五, HW04, 2, 0, 0, 1, 1, 1
- 25 2017/3/28 上午 10:21:20, B04401125, 醫學二, HW04, 10, 0, 0, 2, 3, 4
- 26 2017/3/28 上午 10:21:51, B02302333, 會計四, HW04, 10, 0, 0, 3, 3, 6
- 27 2017/3/28 上午 10:22:11, B04401094, 醫學二, HW01, 10, 0, 0, 3, 6, 6
- 28 2017/3/28 上午 10:22:13, B03703011, 財金三, HW04, 10, 0, 3, 3, 10, 4
- 29 2017/3/28 上午 10:23:10, B04401094, 普泰中, HW02, 10, 0, 0, 3, 5, 5
- 30 2017/3/28 上午 10:24:52, B04401094, 醫學二, HW03, 10, 0, 0, 1, 3, 3
- 31 2017/3/28 上午 10:26:05, B04401094, 醫學二, HW04, 10, 0, 0, 2, 10, 3
- 32 2017/3/28 上午 10:32:28, B04401125, 醫學二, HW05, 10, 0, 0, 2, 2, 2

HW07++A : 課程學習時間資料分析

On 11/18, 2020

- 請參考範例程式，ReadData_LearnTime.R，完成下面工作：
 - 資料：從網路上，讀取檔案
 - 統整：用 `summary` 瞭解所讀取的資料的基本統計或分析
 - 計算：計算數據的總筆數
 - 計算：計算每一個 HW 個別的筆數
 - 計算：計算每一個 學生 ID 個別的筆數
 - 計算：計算每一個人，按照不同作業的每一個作業，所花費的總時間
 - 計算：計算每一個作業，每一個活動，分別所花費的時間分布狀況
 - 排序：比較您所花費的時間，在所有學生所花費時間，所佔的順序的百分比
 - 排序：分析您在哪些活動所花費的時間，所佔的順序比較前面
 - 其他的，您覺得想要進一步瞭解的資訊！！！！
- 把執行的過程，以及產生的數據等，整理一個檔案 (.R or .Rmd or .html or .pdf)。

HW07++A：課程學習時間資料分析

On 11/18, 2020

- 繳交下面檔案，檔案名稱：[HW07_學號_關鍵字.xxx](#)
- 您可以繳交下面格式之中的任何一種格式的檔案：
 - 程式檔：[HW07_B01921001_LearnTime.R](#)
 - 程式與結果檔：[HW07_B01921001_LearnTime.Rmd](#)
 - 或轉成：[HW07_B01921001_LearnTime.html](#)
 - 報告檔案：[HW07_B01921001_LearnTime.pdf](#)
- 繳交方式與期限：
 - 上傳檔案到：<https://cool.ntu.edu.tw>
 - 繳交期限：[11/23 \(Mon\), 11pm 以前](#)
- 學習方式：
 - 請至下面網址輸入此次的學習方式所花的時間：
 - <https://forms.gle/TGYXj2uLoL4HwqLHA>

HW07++B : 2016年美國總統選舉

■ 2016年美國總統選舉

競選資金 [編輯]

這裡列出2016年9月聯邦選舉委員會 (FEC) 公布各候選人於競選運動中使用的資金。外部團體是獨立支出專門委員會，或稱政治行動委員會及超級政治行動委員會。數據取自FEC和政治反應中心^[203]部分支出總額由於在FEC載數日期前取款而無法使用。截至2016年9月，10名列入選票的候選人已向FEC提交財務報告。

候選人	選舉委員會 (截至9月30日)				外部團體 (截至10月16日)			總計
	募集資金	使用資金	手頭資金	債務	募集資金	使用資金	手頭資金	
希拉蕊·柯林頓 ^{[204][205]}	\$460,168,401	\$400,504,099	\$59,664,302	\$626,094	\$171,240,103	\$148,604,471	\$22,635,633	\$534,352,332
唐納·川普 ^{[206][207]}	\$224,449,710	\$189,673,422	\$34,776,287	\$0	\$214,496,514	\$183,418,431	\$31,078,083	\$367,405,384
蓋瑞·強生 ^{[208][209]}	\$10,573,731	\$9,463,272	\$1,217,539	\$1,538,118	\$1,378,510	\$917,521	\$460,988	\$10,349,663
洛基·德拉富恩特 ^[210]	\$7,351,270	\$7,354,663	-\$3,392	\$7,334,250	\$0	\$0	\$0	\$7,354,663
吉爾·史坦 ^{[211][212]}	\$3,218,525	\$3,144,843	\$73,681	\$87,740	\$0	\$0	\$0	\$3,144,843
埃文·麥克馬林 ^[213]	\$501,093	\$496,776	\$4,316	\$0	\$0	\$0	\$0	\$496,776
達雷爾·卡斯爾 ^[214]	\$52,234	\$51,365	\$869	\$2,500	\$0	\$0	\$0	\$51,365
格洛麗亞·拉里瓦 ^[215]	\$29,243	\$24,207	\$5,034	\$0	\$0	\$0	\$0	\$24,207
莫妮卡·穆爾黑德 ^[216]	\$11,547	\$9,127	\$2,419	\$4,500	\$0	\$0	\$0	\$9,127
彼德·斯奎斯 ^[217]	\$7,966	\$4,238	\$7,454	\$8,000	\$0	\$0	\$0	\$4,238

```

3090 href="https://zh.wikipedia.org/wiki/2016%E5%B9%B4%E7%B8%8E%E5%9C%8B%E7%B8%BD%E7%B5%B1%E9%81%B8%E8%88%89#cite_note-203">[203] </A></SUP>部分支出
3091 <TABLE class="wikitable sortable"
3092 style="text-align: center; font-size: 12px;"><TBODY>
3093 <TR>
3094 <TH rowspan="2">候選人</TH>
3095 <TH colspan="4">選舉委員會<SMALL> (截至9月30日) </SMALL></TH>
3096 <TH colspan="3">外部團體<SMALL> (截至10月16日) </SMALL></TH>
3097 <TH rowspan="2">總計</TH></TR>
3098 <TR>
3099 <TH>募集資金</TH>
3100 <TH>使用資金</TH>
3101 <TH>手頭資金</TH>
3102 <TH>債務</TH>
3103 <TH>募集資金</TH>
3104 <TH>使用資金</TH>
3105 <TH>手頭資金</TH></TR>
3106 <TR>
3107 <TD data-sort-value="Clinton, Hillary"><A title="希拉蕊·柯林頓" class="mw-redirect"
3108 href="https://zh.wikipedia.org/wiki/%E5%B8%8C%E6%B8%89%E8%8E%89%C2%B7%E5%85%B8%E6%9E%97%E9%A0%92">希拉蕊·柯林頓</A><SUP>
3109 class="reference" id="cite_ref-204"><A href="https://zh.wikipedia.org/wiki/2016%E5%B9%B4%E7%B8%8E%E5%9C%8B%E7%B8%BD%E7%B5%B1%E9%81%B8%E8%
3110 class="reference" id="cite_ref-205"><A href="https://zh.wikipedia.org/wiki/2016%E5%B9%B4%E7%B8%8E%E5%9C%8B%E7%B8%BD%E7%B5%B1%E9%81%B8%E8%
3111 <TD>$460,168,401</TD>
3112 <TD>$400,504,099</TD>
3113 <TD>$59,664,302</TD>
3114 <TD>$626,094</TD>
3115 <TD>$171,240,103</TD>
3116 <TD>$148,604,471</TD>
3117 <TD>$22,635,633</TD>
3118 <TD><b>$534,352,332</b></TD></TR>
3119 <TR>
3120 <TD data-sort-value="Trump, Donald"><A title="唐納·川普" href="https://zh.wikipedia.org/wiki/%E5%94%90%E7%B4%8D%C2%B7%E5%B7%9D%E6%99%AE">
3121 class="reference" id="cite_ref-206"><A href="https://zh.wikipedia.org/wiki/2016%E5%B9%B4%E7%B8%8E%E5%9C%8B%E7%B8%BD%E7%B5%B1%E9%81%B8%E8%
3122 class="reference" id="cite_ref-207"><A href="https://zh.wikipedia.org/wiki/2016%E5%B9%B4%E7%B8%8E%E5%9C%8B%E7%B8%BD%E7%B5%B1%E9%81%B8%E8%
3123 <TD>$224,449,710</TD>

```

HW07++B : 2016年美國總統選舉

On 11/18, 2020

- 請參考範例程式，ReadData_USA2016.R，完成下面工作：
 - 資料：從網路上，讀取該網頁的資料，
 - 請試著從此讀取進來的資料，整理出一個資料框 (data.frame) 的變數或物件，然後，將“競選經費”的表格內的：候選人，募資資金，使用資金等各項資金的金額等資料放入該資料框內。
- 把執行的過程，以及產生的數據等，整理一個檔案 (.R or .Rmd or .html or .pdf)。

競選資金 [編輯]

這裡列出2016年9月聯邦選舉委員會 (FEC) 公布各候選人於競選運動中使用的資金。外部團體是獨立支出專門委員會，或稱政治行動委員會及超級政治行動委員會。數據取自FEC和政治反應中心^[203]部分支出總額由於在FEC截數日期前取款而無法使用。截至2016年9月，10名列入選票的候選人已向FEC提交財務報告。

候選人	選舉委員會 (截至9月30日)				外部團體 (截至10月16日)			總計
	募集資金	使用資金	手頭資金	債務	募集資金	使用資金	手頭資金	
希拉蕊·柯林頓 ^{[204][205]}	\$460,168,401	\$400,504,099	\$59,664,302	\$626,094	\$171,240,103	\$148,604,471	\$22,635,633	\$534,352,332
唐納·川普 ^{[206][207]}	\$224,449,710	\$189,673,422	\$34,776,287	\$0	\$214,496,514	\$183,418,431	\$31,078,083	\$367,405,384
蓋瑞·強生 ^{[208][209]}	\$10,573,731	\$9,463,272	\$1,217,539	\$1,538,118	\$1,378,510	\$917,521	\$460,988	\$10,349,663
洛基·德拉富恩特 ^[210]	\$7,351,270	\$7,354,663	-\$3,392	\$7,334,250	\$0	\$0	\$0	\$7,354,663
吉爾·史坦 ^{[211][212]}	\$3,218,525	\$3,144,843	\$73,681	\$87,740	\$0	\$0	\$0	\$3,144,843
埃文·麥克馬林 ^[213]	\$501,093	\$496,776	\$4,316	\$0	\$0	\$0	\$0	\$496,776
達雷爾·卡斯爾 ^[214]	\$52,234	\$51,365	\$869	\$2,500	\$0	\$0	\$0	\$51,365
格洛麗亞·拉·里瓦 ^[215]	\$29,243	\$24,207	\$5,034	\$0	\$0	\$0	\$0	\$24,207
莫妮卡·穆爾黑德 ^[216]	\$11,547	\$9,127	\$2,419	\$4,500	\$0	\$0	\$0	\$9,127
彼德·斯奎斯 ^[217]	\$7,966	\$4,238	\$7,454	\$8,000	\$0	\$0	\$0	\$4,238

HW07++B：2016年美國總統選舉

On 11/18, 2020

- 繳交下面檔案，檔案名稱：[HW07_學號_關鍵字.xxx](#)
- 您可以繳交下面格式之中的任何一種格式的檔案：
 - 程式檔：[HW07_B01921001_USA2016.R](#)
 - 程式與結果檔：[HW07_B01921001_USA2016.Rmd](#)
 - 或轉成：[HW07_B01921001_USA2016.html](#)
 - 報告檔案：[HW07_B01921001_USA2016.pdf](#)
- 繳交方式與期限：
 - 上傳檔案到：<https://cool.ntu.edu.tw>
 - 繳交期限：[11/23 \(Mon\), 11pm 以前](#)
- 學習方式：
 - 請至下面網址輸入此次的學習方式所花的時間：
 - <https://forms.gle/TGYXj2uLoL4HwqLHA>

HW07++C : 餐廳帳單的小費

On 11/18, 2020

- 某餐廳的 244 筆紀錄小費金額的數據

- 原始檔案連結：

<https://raw.githubusercontent.com/mwaskom/seaborn-data/master/tips.csv>

	A	B	C	D	E	F	G
1	total_bill	tip	sex	smoker	day	time	size
2	16.99	1.01	Female	No	Sun	Dinner	2
3	10.34	1.66	Male	No	Sun	Dinner	3
4	21.01	3.5	Male	No	Sun	Dinner	3
5	23.68	3.31	Male	No	Sun	Dinner	2
6	24.59	3.61	Female	No	Sun	Dinner	4
7	25.29	4.71	Male	No	Sun	Dinner	4
8	8.77	2	Male	No	Sun	Dinner	2
9	26.88	3.12	Male	No	Sun	Dinner	4
10	15.04	1.96	Male	No	Sun	Dinner	2
11	14.78	3.23	Male	No	Sun	Dinner	2
12	10.27	1.71	Male	No	Sun	Dinner	2
13	35.26	5	Female	No	Sun	Dinner	4
14	15.42	1.57	Male	No	Sun	Dinner	2
15	18.43	3	Male	No	Sun	Dinner	4
16	14.83	3.02	Female	No	Sun	Dinner	2
17	21.58	3.92	Male	No	Sun	Dinner	2
18	10.33	1.67	Female	No	Sun	Dinner	3
19	16.29	3.71	Male	No	Sun	Dinner	3
20	16.97	3.5	Female	No	Sun	Dinner	3
21	20.65	3.35	Male	No	Sat	Dinner	3
22	17.92	4.08	Male	No	Sat	Dinner	2
23	20.29	2.75	Female	No	Sat	Dinner	2
24	15.77	2.23	Female	No	Sat	Dinner	2
25	39.42	7.58	Male	No	Sat	Dinner	4
26	19.82	3.18	Male	No	Sat	Dinner	2

HW07++C : 餐廳帳單的小費

On 11/18, 2020

- 請完成下面工作：
 - 資料：從網路上，讀取檔案
 - 統整：用 `summary` 瞭解所讀取的資料的基本統計或分析
 - 計算：計算花費金額的統計數據
 - 計算：計算小費的統計數據
 - 計算：計算每一個花費金額，小費所佔的百分比
 - 計算：計算每一個花費金額，每一個人所花費的金額的平均值
 - 計算：計算男性所花費的金額分布，以及所給小費的數據分布
 - 計算：計算女性所花費的金額分布，以及所給給小費的數據分布
 - 計算：計算午餐所花費的金額分布，以及所給給小費的數據分布
 - 計算：計算晚餐所花費的金額分布，以及所給給小費的數據分布
 - 其他的，您覺得想要進一步瞭解的資訊！！
- 把執行的過程，以及產生的數據等，整理一個檔案 (.R or .Rmd or .html or .pdf)。

HW07++C：餐廳帳單的小費

On 11/18, 2020

- 繳交下面檔案，檔案名稱：[HW07_學號_關鍵字.xxx](#)
- 您可以繳交下面格式之中的任何一種格式的檔案：
 - 程式檔：[HW07_B01921001_tips.R](#)
 - 程式與結果檔：[HW07_B01921001_tips.Rmd](#)
 - 或轉成：[HW07_B01921001_tips.html](#)
 - 報告檔案：[HW07_B01921001_tips.pdf](#)
- 繳交方式與期限：
 - 上傳檔案到：<https://cool.ntu.edu.tw>
 - 繳交期限：[11/23 \(Mon\), 11pm 以前](#)
- 學習方式：
 - 請至下面網址輸入此次的學習方式所花的時間：
 - <https://forms.gle/TGYXj2uLoL4HwqLHA>

- 取得檔案的資料：輸入與輸出
- 取得內建資料庫
- 取得其他資料庫的資料
- 取得網頁的資料

檔案資料輸入與輸出

設定工作目錄

- `mywd <- "L:/DataWD"`
- `setwd(mywd)`
- `getwd()`

工作空間內的物件

- `x <- 1:10`
- `y <- matrix(1:6, nrow = 2, ncol = 3)`

- `ls()`
- `rm()`
- `rm("x")`
- `ls()`

- `current = ls()`
- `rm(list = current)`
- `ls()`

- `rm(list = ls())`
- `ls()`

檔案指令 – dump, source

- # 儲存物件的名稱與其值 (內容)
- `x <- 1:10`
- `y <- matrix(1:6, nrow = 2, ncol = 3)`
- `ls()`
- `dump(c("x", "y"), file = "my_dump.txt")`
- `rm(x); rm(y)`
- `ls()`
- `source(file = "my_dump.txt")`
- `ls()`

```
my_dump.txt
```

```
x <-  
1:10
```

```
y <-  
structure(1:6, .Dim = 2:3)
```

檔案指令 – dput, dget

- # 儲存物件的值 (內容)
- `x <- 1:10`
- `y <- matrix(1:6, nrow = 2, ncol = 3)`
- `dput(y, file = "my_dput.txt")`
- `newy <- dget("my_dput.txt")`
- `newy`

```
my_dput.txt
```

```
structure(1:6, .Dim = 2:3)
```

檔案指令 - sink

- # 螢幕輸出轉到檔案

- sink("my_sink.txt")
- x
- y

- sink()
- x
- y

my_sink.txt

```
[1]  1  2  3  4  5  6  7  8  9 10
      [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```

檔案指令 - read.table, write.table

- my_data <- iris[c(1, 2, 51, 52, 101, 102), c(1, 2, 5)]
- my_data <- edit(my_data)
- write.table(my_data, "my_data_content.txt")
- my_data_in <- read.table("my_data_content.txt")

```
> my_data
      Sepal.Length Sepal.Width  Species
1             5.1           3.5   setosa
2             4.9           3.0   setosa
51            7.0           3.2 versicolor
52            6.4           3.2 versicolor
101           6.3           3.3  virginica
102           5.8           2.7  virginica
```

row.names	Sepal.Length	Sepal.Width	Species	var5	var6
1	5.1	3.5	setosa		
2	4.9	3	setosa		
51	7	3.2	versicolor		
52	6.4	3.2	versicolor		
101	6.3	3.3	virginica		
102	5.8	2.7	virginica		
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					

```
my_data_content.txt

"Length" "width" "Species"
"1" 5.1 3.5 "setosa"
"2" 4.9 3 "setosa"
"3" 7 3.2 "versicolor"
"4" 6.4 3.2 "versicolor"
"5" 6.3 3.3 "virginica"
"6" 5.8 2.7 "virginica"
```

```
> my_data_in
      Length width  Species
1         5.1   3.5   setosa
2         4.9   3.0   setosa
3         7.0   3.2 versicolor
4         6.4   3.2 versicolor
5         6.3   3.3  virginica
6         5.8   2.7  virginica
```


檔案指令 – read.table, read.csv

- # csv: comma separated values
- read.table("mydataf2.csv")
- read.table("mydataf2.csv", header = TRUE)
- read.table("mydataf2.csv", header = TRUE, sep = ",")
- dt1 <- read.table("mydataf2.csv", sep = ",", header = TRUE)
- dt2 <- read.csv("mydataf2.csv")

mydata2.csv

```
Length,width,Species
5.1,3.5,setosa
4.9,3,setosa
7,3.2,versicolor
6.4,3.2,versicolor
6.3,3.3,virginica
5.8,2.7,virginica
```

檔案指令 – read.table, read.csv

mydata2.csv

```
Length,Width,Species
5.1,3.5,setosa
4.9,3,setosa
7,3.2,versicolor
6.4,3.2,versicolor
6.3,3.3,virginica
5.8,2.7,virginica
```

```
> read.table( "mydataf2.csv" )
      V1
1 Length,Width,Species
2      5.1,3.5,setosa
3      4.9,3,setosa
4      7,3.2,versicolor
5      6.4,3.2,versicolor
6      6.3,3.3,virginica
7      5.8,2.7,virginica
```

```
> read.table( "mydataf2.csv", header = TRUE )
```

```
Length.Width.Species
1      5.1,3.5,setosa
2      4.9,3,setosa
3      7,3.2,versicolor
4      6.4,3.2,versicolor
5      6.3,3.3,virginica
6      5.8,2.7,virginica
```

```
> read.table( "mydataf2.csv", header = TRUE )
```

```
Length.Width.Species
1      5.1,3.5,setosa
2      4.9,3,setosa
3      7,3.2,versicolor
4      6.4,3.2,versicolor
5      6.3,3.3,virginica
6      5.8,2.7,virginica
```

```
> read.table( "mydataf2.csv", header = TRUE, sep = "," )
```

```
Length Width Species
1      5.1      3.5 setosa
2      4.9      3.0 setosa
3      7.0      3.2 versicolor
4      6.4      3.2 versicolor
5      6.3      3.3 virginica
6      5.8      2.7 virginica
```

```
> dt1 <- read.table( "mydataf2.csv", sep = ",", header = TRUE )
```

```
> dt1
  Length Width Species
1      5.1      3.5 setosa
2      4.9      3.0 setosa
3      7.0      3.2 versicolor
4      6.4      3.2 versicolor
5      6.3      3.3 virginica
6      5.8      2.7 virginica
```

```
> dt2 <- read.csv( "mydataf2.csv" )
```

```
> dt2
  SLength Width Species
1      5.1      3.5 setosa
2      4.9      3.0 setosa
3      7.0      3.2 versicolor
4      6.4      3.2 versicolor
5      6.3      3.3 virginica
6      5.8      2.7 virginica
```

檔案指令 – read.table, read.csv

- # 設定不同的目錄
- `tphdata <- read.table("L:/DataWD/Typhoon-01.txt", header = TRUE)`
- `tphdata`

```
> tphdata
```

```
      x1  x2 x3 x4 d1  d2 d3
1 23.0 130.9 970 0 23.1 130.2 970
2 23.1 130.2 970 0 23.2 129.5 970
3 23.2 129.5 970 0 23.3 128.6 965
4 23.3 128.6 965 0 23.3 128.1 960
5 23.3 128.1 960 0 23.3 127.9 960
6 23.3 127.9 960 0 23.3 127.6 960
7 23.3 127.6 960 0 23.4 126.9 960
8 23.4 126.9 960 0 23.4 126.6 960
9 23.4 126.6 960 0 23.4 126.5 960
...
...
...
46 23.9 121.5 955 1 24.0 121.4 965
47 24.0 121.4 965 1 24.1 121.4 965
48 24.1 121.4 965 1 24.2 121.3 965
49 24.2 121.3 965 1 24.3 121.3 965
50 24.3 121.3 965 1 24.4 121.3 965
```

檔案指令 - scan

- # 讀取大量資料的文字檔
- `scandata <- scan("scanlist.txt", list(Length = 0, Width = 0, Species = ""))`

scanlist.txt

```
5.1 3.5 setosa
4.9 3.0 setosa
7.0 3.2 versicolor
6.4 3.2 versicolor
6.3 3.3 virginica
5.8 2.7 virginica
```

> scandata

\$Length

```
[1] 5.1 4.9 7.0 6.4 6.3 5.8
```

\$width

```
[1] 3.5 3.0 3.2 3.2 3.3 2.7
```

\$Species

```
[1] "setosa"      "setosa"      "versicolor"
[4] "versicolor" "virginica"   "virginica"
```

檔案指令 - Insurance 例子

- # 讀取資料
- `Ins_c1 <- read.csv("Insurance.csv")`
- `head(Ins_c1)`

- `Ins_c2 <- read.table("Insurance.csv")`
- `head(Ins_c2)`

- `Ins_c3 <- read.table("Insurance.csv", header = TRUE, sep=",")`
- `head(Ins_c3)`

- `Ins_t1 <- read.table("Insurance.txt")`
- `head(Ins_t1)`

- `Ins_t2 <- read.table("Insurance.txt", header = TRUE, sep="")`
- `head(Ins_t2)`

檔案指令 - read.table, read.csv

	District	Group	Age	Holders	Claims
1	1	<1l	<25	197	38
2	1	<1l	25-29	264	35
3	1	<1l	30-35	246	20
4	1	<1l	>35	1680	156
5	1	1-1.5l	<25	284	63
6	1	1-1.5l	25-29	536	84
7	1	1-1.5l	30-35	696	89
8	1	1-1.5l	>35	3582	400
9	1	1.5-2l	<25	133	19
10	1	1.5-2l	25-29	286	52

```
> Ins_c1 <- read.csv( "Insurance.csv" )
```

```
> head( Ins_c1 )
```

```

X District Group Age Holders Claims
1 1      1 <1l <25  197  38
2 2      1 <1l 25-29 264  35
3 3      1 <1l 30-35 246  20
4 4      1 <1l >35 1680 156
5 5      1 1-1.5l <25  284  63
6 6      1 1-1.5l 25-29 536  84

```

```
> Ins_c2 <- read.table( "Insurance.csv" )
```

```
> head( Ins_c2 )
```

```

V1          V2
1 NA, "District", "Group", "Age", "Holders", "Claims"
2 1      , "1", "<1l", "<25", 197, 38
3 2      , "1", "<1l", "25-29", 264, 35
4 3      , "1", "<1l", "30-35", 246, 20
5 4      , "1", "<1l", ">35", 1680, 156
6 5      , "1", "1-1.5l", "<25", 284, 63

```

```
> Ins_c3 <- read.table( "Insurance.csv", header = TRUE, sep="," )
```

```
> head( Ins_c3 )
```

```

X District Group Age Holders Claims
1 1      1 <1l <25  197  38
2 2      1 <1l 25-29 264  35
3 3      1 <1l 30-35 246  20
4 4      1 <1l >35 1680 156
5 5      1 1-1.5l <25  284  63
6 6      1 1-1.5l 25-29 536  84

```

```
> Ins_t1 <- read.table( "Insurance.txt" )
```

```
> head( Ins_t1 )
```

```

District Group Age Holders Claims
1      1 <1l <25  197  38
2      1 <1l 25-29 264  35
3      1 <1l 30-35 246  20
4      1 <1l >35 1680 156
5      1 1-1.5l <25  284  63
6      1 1-1.5l 25-29 536  84

```

```
> Ins_t2 <- read.table( "Insurance.txt", header = TRUE, sep="" )
```

```
> head( Ins_t2 )
```

```

District Group Age Holders Claims
1      1 <1l <25  197  38
2      1 <1l 25-29 264  35
3      1 <1l 30-35 246  20
4      1 <1l >35 1680 156
5      1 1-1.5l <25  284  63
6      1 1-1.5l 25-29 536  84

```

讀取資料 - 從 xls, xlsx 檔案

- # 讀取資料
- `read.csv("tips_10.csv")`
- `read.table("tips_10.xlsx")`
- `read.csv("tips_10.xlsx")`

- # 載入套件
- `install.packages("readxl")`
- `library(readxl)`

- `read_excel("tips_10.xlsx")`
- `read_excel("tips_10.xlsx") -> my_data_xls_in`
- `my_data_xls_in`

讀取資料 - 從 xls, xlsx 檔案

- # 從不同頁面 sheet 讀取資料
- # 用編號
- `read_excel("tips_10_3.xlsx", sheet = 1)`
- `read_excel("tips_10_3.xlsx", sheet = 2)`
- `read_excel("tips_10_3.xlsx", sheet = 3)`

- # 用頁面名稱
- `read_excel("tips_10_3.xlsx", sheet = "tips")`
- `read_excel("tips_10_3.xlsx", sheet = "money")`
- `read_excel("tips_10_3.xlsx", sheet = "out")`

讀取資料 - 從其他檔案格式

- R Programming/Importing and exporting data
 - https://en.wikibooks.org/wiki/R_Programming/Importing_and_exporting_data

- Getting Data From An Online Source
 - <https://www.r-bloggers.com/getting-data-from-an-online-source/>

內建資料庫

- # R 內建資料集：datasets
- # 近百個資料集，涵蓋：醫學、自然、社會、人體等資料
- data(package = "datasets")
- help(AirPassengers)
- ?AirPassengers
- **AirPassengers** % Monthly Airline Passenger Numbers 1949-1960
- summary(AirPassengers)
- data(package = .packages(all.available = TRUE))
- # CO2 Carbon Dioxide Uptake in Grass Plants
- # uspop Populations Recorded by the US Census
- # Titanic Survival of passengers on the Titanic
- # women Average Heights and Weights for American Women

- `# CO2` Carbon Dioxide Uptake in Grass Plants
- `CO2`
- `summary(CO2)`

- `# uspop` Populations Recorded by the US Census
- `uspop`
- `summary(uspop)`

- `# Titanic` Survival of passengers on the Titanic
- `Titanic`
- `summary(Titanic)`

- `# women` Average Heights and Weights for American Women
- `women`
- `summary(women)`

其他軟體套件的資料集

- # MASS
- library(MASS)
- data(Insurance)

- ?Insurance
- head(Insurance)
- tail(Insurance)
- dim(Insurance)
- names(Insurance)
- attributes(Insurance)

- class(Insurance\$District)
- class(Insurance\$Age)
- class(Insurance\$Holders)
- levels(Insurance\$Age)

```

> head( Insurance )
  District Group Age Holders Claims
1      1  <1| <25  197   38
2      1  <1| 25-29  264   35
3      1  <1| 30-35  246   20
4      1  <1| >35 1680  156
5      1 1-1.5| <25  284   63
6      1 1-1.5| 25-29  536   84

> tail( Insurance )
  District Group Age Holders Claims
59      4 1.5-2| 30-35   68   16
60      4 1.5-2| >35  344   63
61      4 >2| <25    3    0
62      4 >2| 25-29  16    6
63      4 >2| 30-35  25    8
64      4 >2| >35  114   33

> dim( Insurance )
[1] 64 5

> names( Insurance )
[1] "District" "Group" "Age" "Holders" "Claims"

> attributes( Insurance )
$names
[1] "District" "Group" "Age" "Holders" "Claims"

$class
[1] "data.frame"

$row.names
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
[23] 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44
[45] 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64

> class( Insurance$District )
[1] "factor"

> class( Insurance$Age )
[1] "ordered" "factor"

> class( Insurance$Holders )
[1] "integer"

> levels( Insurance$Age )
[1] "<25" "25-29" "30-35" ">35"

```

其他軟體套件的資料集

- # arules 軟體套件中，Groceries 資料集
- `install.packages("arules")`
- `library(arules)`
- `data(Groceries)`
- ?Groceries
- `Groceries[1:10]`
- `inspect(Groceries[1:10])`
- # 10位消費者購物車中的商品

```
> Groceries[1:10]
```

```
transactions in sparse format with
10 transactions (rows) and
169 items (columns)
```

```
> inspect( Groceries[1:10] )
```

```

items
[1] {citrus fruit,
    semi-finished bread,
    margarine,
    ready soups}
[2] {tropical fruit,
    yogurt,
    coffee}
[3] {whole milk}
[4] {pip fruit,
    yogurt,
    cream cheese ,
    meat spreads}
[5] {other vegetables,
    whole milk,
    condensed milk,
    long life bakery product}
[6] {whole milk,
    butter,
    yogurt,
    rice,
    abrasive cleaner}
[7] {rolls/buns}
[8] {other vegetables,
    UHT-milk,
    rolls/buns,
    bottled beer,
    liquor (appetizer)}
[9] {pot plants}
[10] {whole milk,
    cereals}
```

其他資料庫的資料

- The R Datasets Package
 - <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html>
- UC Irvine Machine Learning Repository
 - <https://archive.ics.uci.edu/ml/datasets.html>
- The Free Datasets at r-dir.com
 - <http://r-dir.com/reference/datasets.html>
- Rdatasets: An archive of datasets distributed with R
 - <http://vincentarelbundock.github.io/Rdatasets>
- Datasets in R packages (IAState)
 - http://www.public.iastate.edu/~hofmann/data_in_r_sortable.html

網頁的資料

- Reading a CSV-file from an URL
- The number of police officers in Scotland over time
- http://www.quandl.com/api/v1/datasets/EUROSTAT/CRIM_PLCE_42.csv
- `read.csv("http://www.quandl.com/api/v1/datasets/EUROSTAT/CRIM_PLCE_42.csv")`

	Date	Value
1	2010-12-31	17263
2	2009-12-31	17409
3	2008-12-31	17048
4	2007-12-31	16221
5	2006-12-31	16234
6	2005-12-31	16221
7	2004-12-31	16001
8	2003-12-31	15482
9	2002-12-31	15287
10	2001-12-31	15093
11	2000-12-31	14948
12	1999-12-31	14684
13	1998-12-31	14854
14	1997-12-31	15050
15	1996-12-31	14672
16	1995-12-31	14479
17	1994-12-31	14313
18	1993-12-31	14139"

- # 和訊網 萬科 股票的相關金融資料
- `http://stockdata.stock.hexun.com/2008en/zxcwzb.aspX?stockid=000002&type=1&date=2013.06.30`
- `install.packages("XML")`
- `library(XML)`
- `url =`
`"http://stockdata.stock.hexun.com/2008en/zxcwzb.aspX?stockid=000002&type=1&date=2013.06.30"`
- `tales1 = readHTMLTable(url)`
- `tales1`

讀取資料 - 從網路的檔案

> tales1

```

$`NULL`
  v1
1 Annual

$`NULL`
      Period End Date      June 30 2013      March 31 2013
1      Operating Income  41,390,345,567.72  13,999,905,876.13
2      Net Profit        4,556,304,906.89   1,613,904,228.33
3      Total Profit      7,133,412,305.51   2,394,885,503.42
4      Net Profit Excluding Extraordinary Items  4,536,753,831.66   1,615,472,700.40
5      Total Assets      432,241,960,220.85  417,894,248,034.75
6      Shareholders' Equity  66,644,627,234.02  65,578,003,059.72
7      Net Cash Flows From Operating Activities  -9,792,399,309.57  -2,383,260,770.37
8      Basic Earnings Per Share      0.41      0.15
9      The Rate Of Return On Equity   6.84      2.46
10     Net Cash Flows From Operating Activities Per Share  -0.89     -0.22
11     Net Assets Value Per Share     6.05      5.96
12     Net Assets Per Share After Adjusted  0.00      0.00
13     Foreign Financial Accounting Standard Net Profit  0.00      0.00
14     EPS Excluding Extraordinary Items  0.41      0.15
15     Report Start Time      2013-01-01      2013-01-01
16     Report End Time       2013-06-30      2013-03-31

$`NULL`
      Period End Date      December 31 2016      December 31 2013      December 31 2014      December 31 2013
1      Operating Income  240,477,236,923.34  195,549,130,020.90  146,388,004,498.44  135,418,791,080.35
2      Net Profit        21,022,606,256.56   18,119,406,249.27   15,745,454,144.70   15,118,549,405.78
3      Total Profit      39,253,611,726.28   33,802,617,619.10   25,252,363,233.49   24,291,011,249.30
4      Net Profit Excluding Extraordinary Items  20,929,278,864.67   17,615,950,216.12   15,576,596,101.66   15,113,721,585.36
5      Total Assets      830,674,213,924.14  611,295,567,689.29  508,408,755,415.65  479,205,323,490.54
6      Shareholders' Equity  113,444,766,722.65  100,183,517,822.33  88,164,569,909.35   76,895,983,339.70
7      Net Cash Flows From Operating Activities  39,566,129,021.69   16,046,020,691.50   41,724,819,113.36   1,923,868,889.89
8      Basic Earnings Per Share      1.90      1.64      1.43      1.37
9      The Rate Of Return On Equity   18.53     18.09     17.86     19.66
10     Net Cash Flows From Operating Activities Per Share  3.58      1.45      3.79      0.18
11     Net Assets Value Per Share     10.28     9.08      7.99      6.98
12     Net Assets Per Share After Adjusted  0.00      0.00      0.00      0.00
13     Foreign Financial Accounting Standard Net Profit  0.00      0.00      0.00      0.00
14     EPS Excluding Extraordinary Items  1.90      1.60      1.41      1.37
15     Report Start Time      2016-01-01      2015-01-01      2014-01-01      2013-01-01
16     Report End Time       2016-12-31      2015-12-31      2014-12-31      2013-12-31

$`NULL`
NULL

```

讀取資料 – 從網路的檔案

- # 和訊網 萬科 股票的相關金融資料
- ls()
- names(tales1)
- tales1[[1]]
- tales1[[2]]
- tales1[[3]]
- tales1[[4]]

- aa <- tales1[[2]]
- aa

- class(aa)
- aa[1]
- aa[2]
- aa[3]

讀取資料 - 從網路的檔案

- # 2016年美國總統選舉

<https://goo.gl/WEKw4I>

- # load packages

- source("ReadData_2016USA.R")



這裡列出2016年9月聯邦選舉委員會 (FEC) 公布各候選人於競選運動中使用的資金。外部團體是獨立支出專門委員會，或稱政治行動委員會及超級政治行動委員會。數據取自FEC和
政治反應中心^[20]部分支出總額由於在FEC數目日期取數而無法使用。截至2016年9月，
10名列入選舉的候選人已向FEC提交財務報告。

候選人	選舉委員會 (截至9月30日)				外部團體 (截至10月16日)				總計
	募集資金	使用資金	手續費	債務	募集資金	使用資金	手續費	債務	
希拉蕊·克林頓 ^{[20][21]}	\$460,168,401	\$400,504,099	\$59,664,302	\$626,094	\$171,240,103	\$148,604,471	\$22,635,633	\$534,852,332	
唐納·川普 ^{[20][22]}	\$224,449,710	\$189,673,422	\$34,776,287	\$0	\$214,496,514	\$182,418,431	\$31,078,083	\$367,405,394	
加里·納哈爾 ^{[20][23]}	\$10,573,731	\$9,463,272	\$1,217,539	\$1,538,118	\$1,378,510	\$917,521	\$460,988	\$10,349,663	
洛基·德拉索爾 ^[22]	\$7,351,270	\$7,354,663	-\$3,392	\$7,334,250	\$0	\$0	\$0	\$7,354,663	
吉爾·斯奈德 ^{[21][22]}	\$3,218,525	\$3,144,843	\$73,681	\$87,740	\$0	\$0	\$0	\$3,144,843	
班·麥克萊斯 ^[23]	\$501,093	\$496,776	\$4,316	\$0	\$0	\$0	\$0	\$496,776	
唐爾斯·卡斯爾 ^[24]	\$52,234	\$51,365	\$869	\$2,500	\$0	\$0	\$0	\$51,365	
格羅麗亞·拉墨茲 ^[25]	\$29,243	\$24,207	\$5,034	\$0	\$0	\$0	\$0	\$24,207	
傑克·卡博爾 ^[26]	\$11,547	\$9,127	\$2,419	\$4,500	\$0	\$0	\$0	\$9,127	
傑德·博塞爾 ^[27]	\$7,966	\$4,238	\$7,454	\$8,000	\$0	\$0	\$0	\$4,238	

```
> data.main.num
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
[1,] 460168401 400504099 59664302 626094 171240103 148604471 22635633 534352332
[2,] 224449710 189673422 34776287 0 214496514 183418431 31078083 367405384
[3,] 10573731 9463272 1217539 1538118 1378510 917521 460988 10349663
[4,] 0 0 0 0 0 0 0 0
[5,] 0 0 0 0 0 0 0 0
[6,] 0 0 0 0 0 0 0 0
[7,] 0 0 0 0 0 0 0 0
[8,] 0 0 0 0 0 0 0 0
[9,] 0 0 0 0 0 0 0 0
[10,] 0 0 0 0 0 0 0 0
>
```

下課了