

Wrangling with Data 與資料糾纏

Tzeng-Ji Chen, Taipei Veterans General Hospital, Hsinchu Branch

陳曾基 臺北榮民總醫院新竹分院



<https://www.facebook.com/VHCT01/>

年輕醫師從事研究的困境

- 沒有機會拜師
- 沒有時間跟團
- 沒有銀兩補習
- 沒有材料運用
- 不會實驗技巧
- 不會統計分析
- 不會程式設計
- 不會論文寫作

<https://openclipart.org/detail/3977/stopwatch-no-shading>

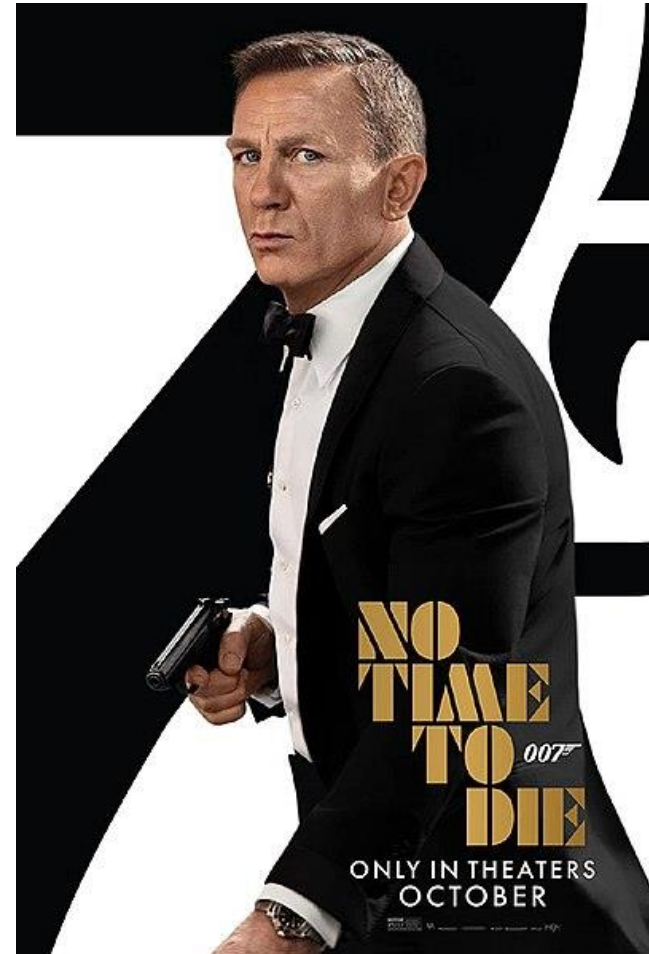
連IRB也無資格申請

年輕醫師從事研究的困境



https://live.staticflickr.com/7472/16323898735_338b00304c_b.jpg

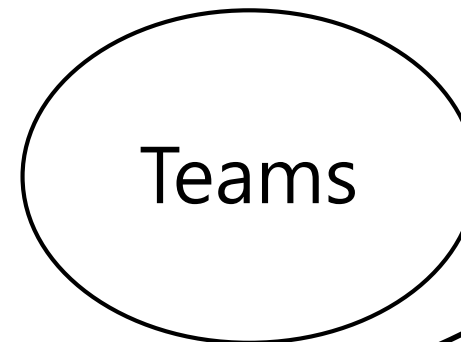
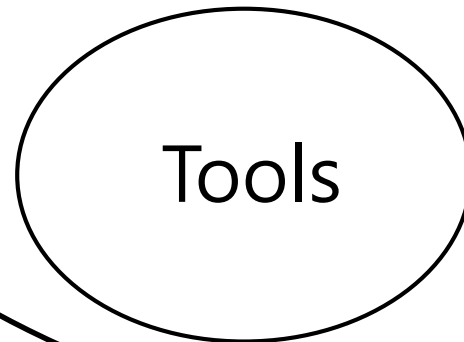
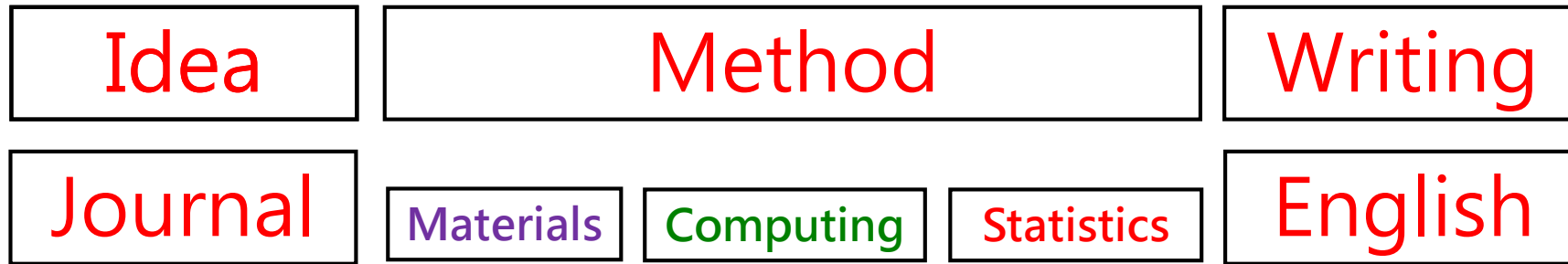
NO
TIME
TO 
WRITE



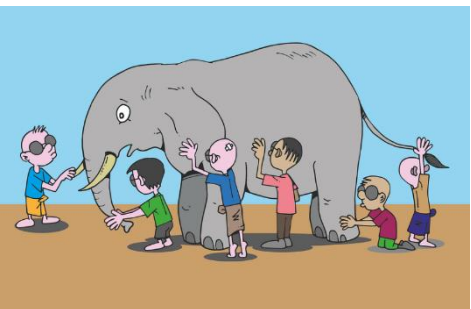
[https://azb.wikipedia.org/wiki/فيلم:No_Time_To_Die_\(Film\).jpg](https://azb.wikipedia.org/wiki/فيلم:No_Time_To_Die_(Film).jpg)

Atmosphere

Paper Production Flow



Infrastructure



Surely you have known ...

Maslow's Hierarchy of Needs

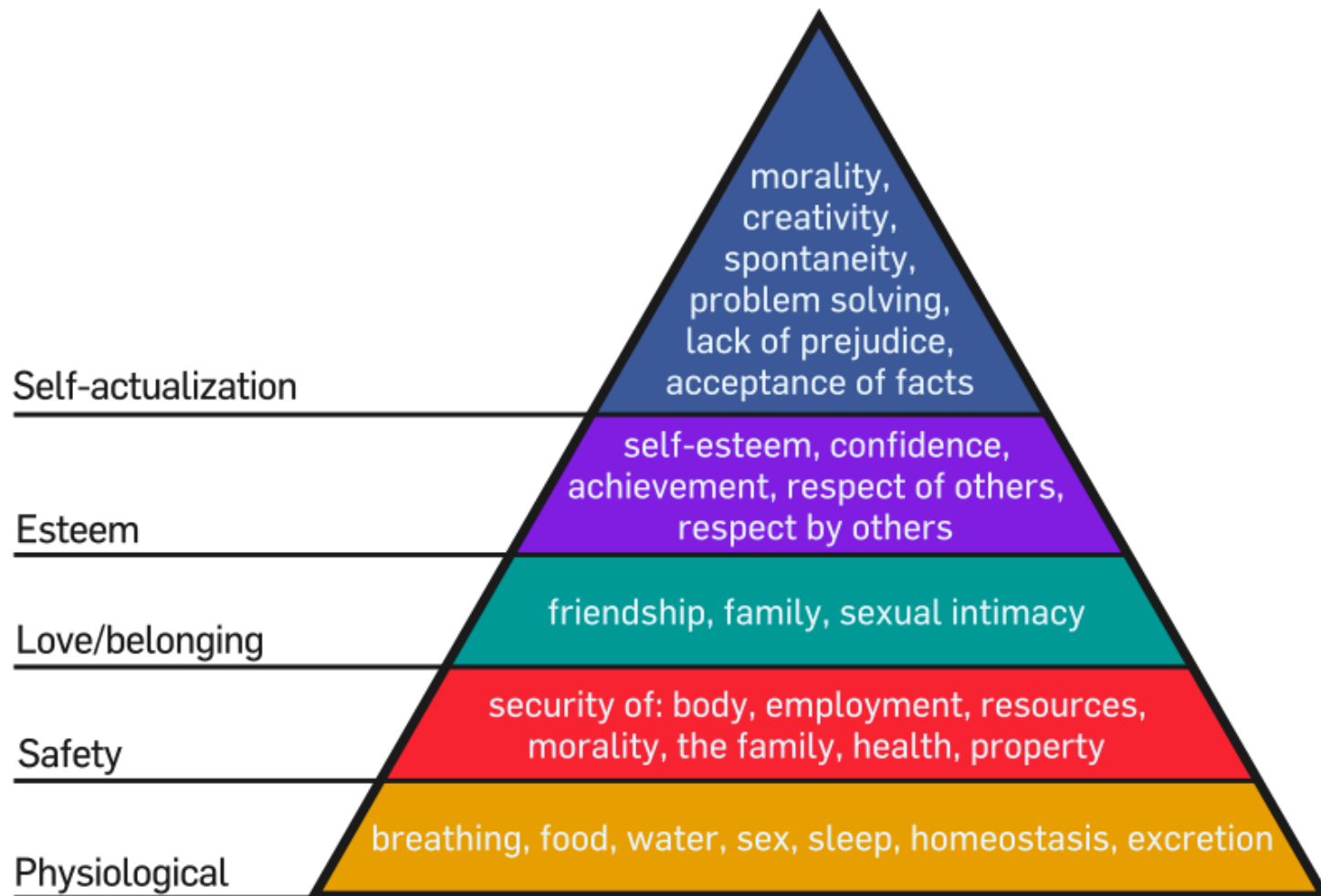
自我實現

尊嚴

社交

安全

生理



https://upload.wikimedia.org/wikipedia/en/e/e0/Abraham_Maslow.jpg

https://commons.wikimedia.org/wiki/File:Maslow%27s_Hierarchy_of_Needs.svg

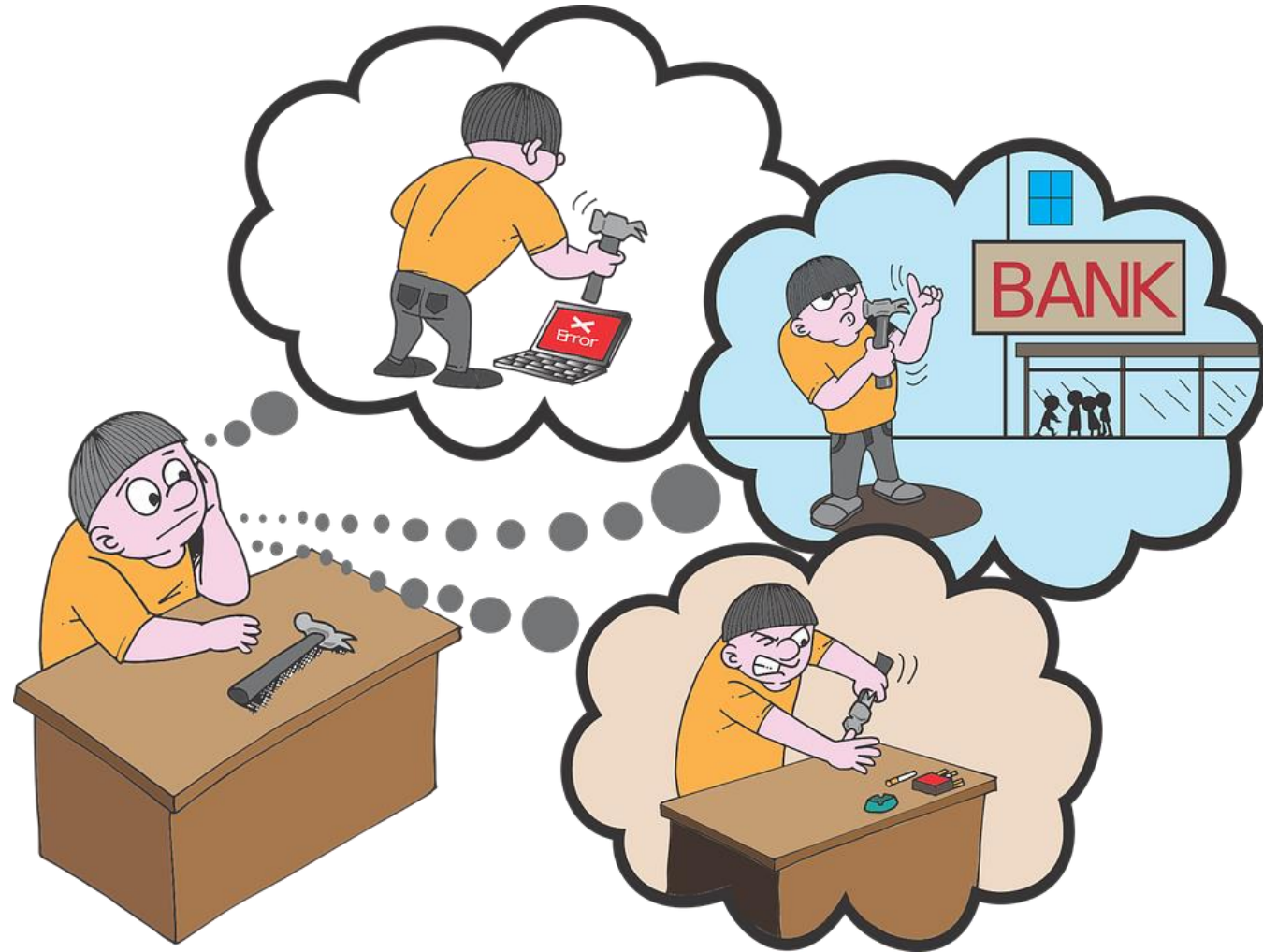
You might not have heard ...

Maslow's Hammer



<https://openclipart.org/detail/278759/caveman-remix>

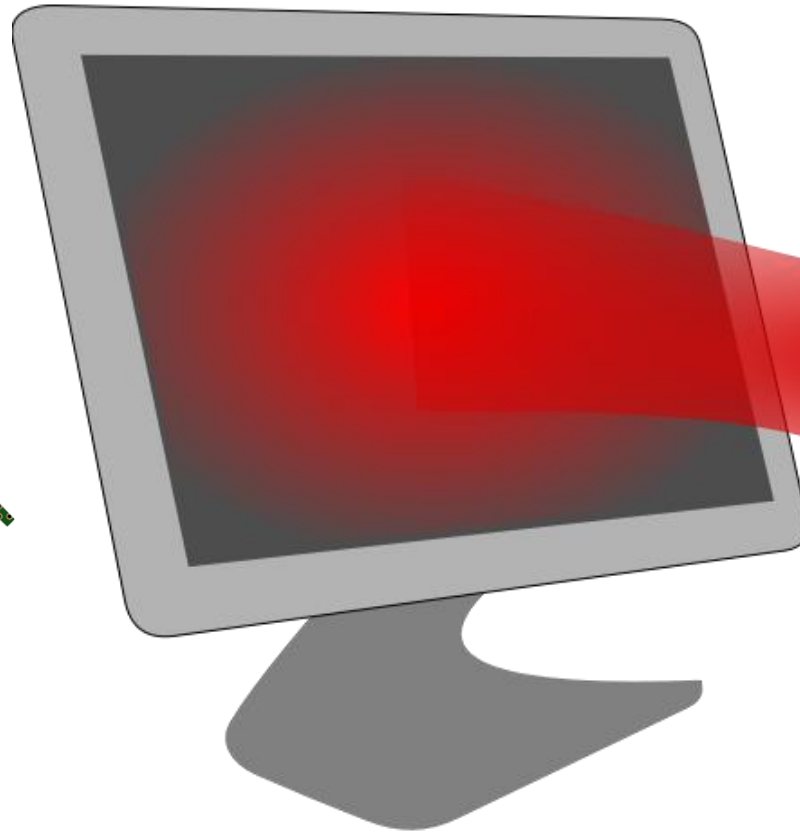
"If all you have is a hammer, everything looks like a nail."



Analogy

If you are familiar with data wrangling, everything looks like research materials.

**DATA
MINING**



Data Skills Most in Demand

1. Python
2. SQL and NoSQL
3. R
4. Data Visualization
5. Machine Learning and Natural Language Processing (NLP)
6. Amazon Web Services (AWS) and Cloud-Driven Computing
7. Extract, Transform, Load (ETL)
8. Statistics

Down to Basics

李家同／基礎不好，何來創新？

2022-09-20 02:59 聯合報／李家同

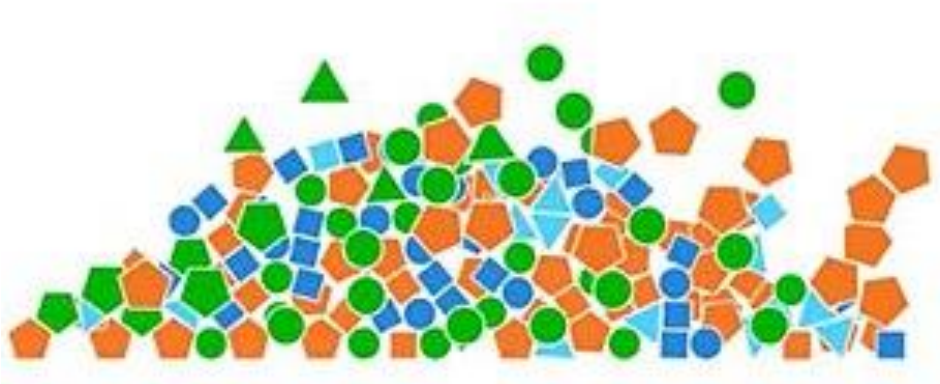
<https://udn.com/news/story/7340/6624987>

Data Wrestling

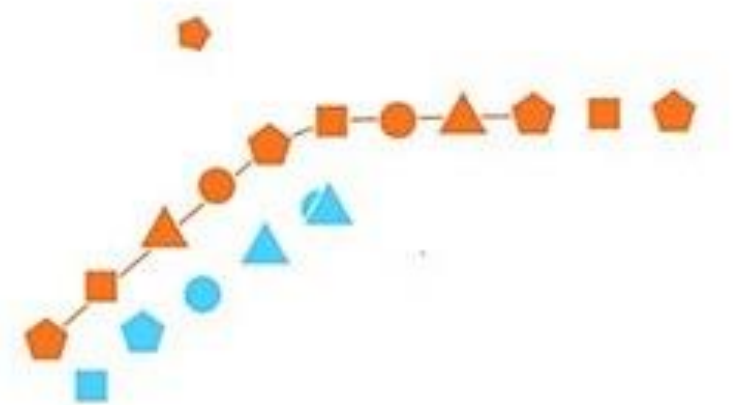
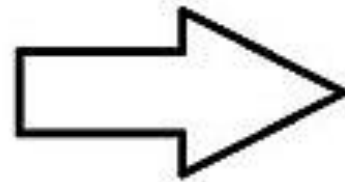


Data Wrangling

a.k.a. data cleaning, data remediation, or data munging



messy raw data



tidy data

Data Wrangling

Starting data

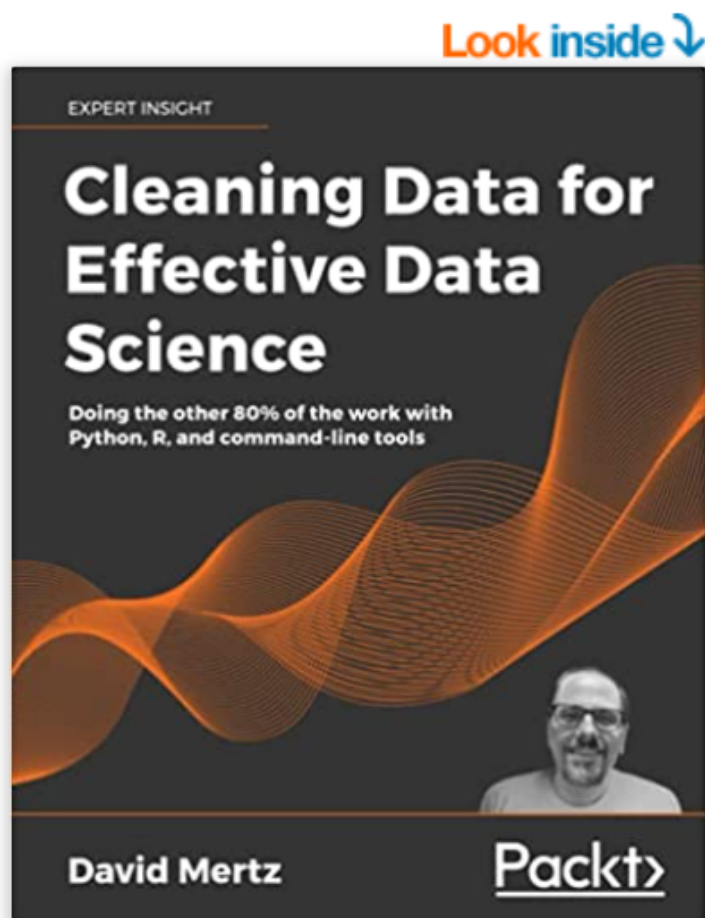
Name	Phone	Birth date	State
John, Smith	445-881-4478	August 12, 1989	Maine
Jennifer Tal	+1-189-456-4513	11/12/1965	Tx
Gates, Bill	(876)546-8165	June 15, 72	Kansas
Alan Fitch	5493156648	2-6-1985	Oh
Jacob Alan	156-4896	January 3	Alabama

Result

Name	Phone	Birth date	State
John Smith	445-881-4478	1989-08-12	Maine
Jennifer Tal	189-456-4513	1965-11-12	Texas
Bill Gates	876-546-8165	1972-06-15	Kansas
Alan Fitch	549-315-6648	1985-02-06	Ohio

80/20 Rule (Pareto Principle)





Cleaning Data for Effective Data Science: Doing the other 80% of the work with Python, R, and command-line tools

作者 [David Mertz](#) ▾ (Author)

★★★★★ ▾ 27等級

[查看所有格式和版本](#)

Kindle
US\$17.84

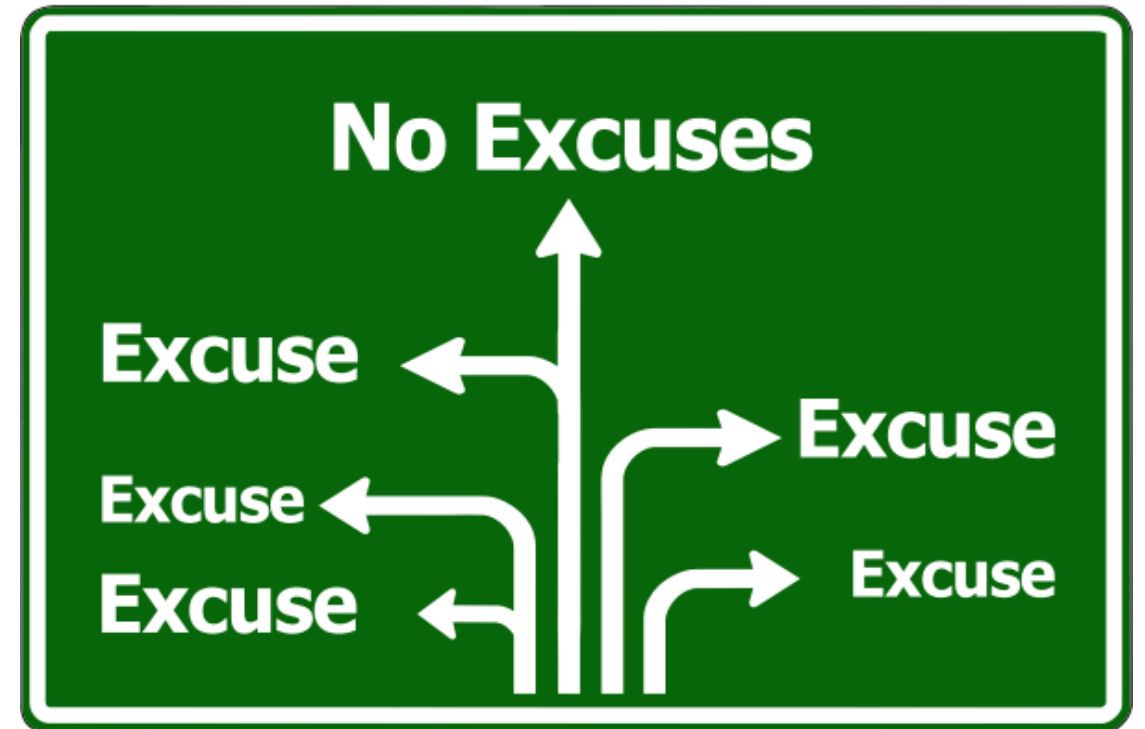
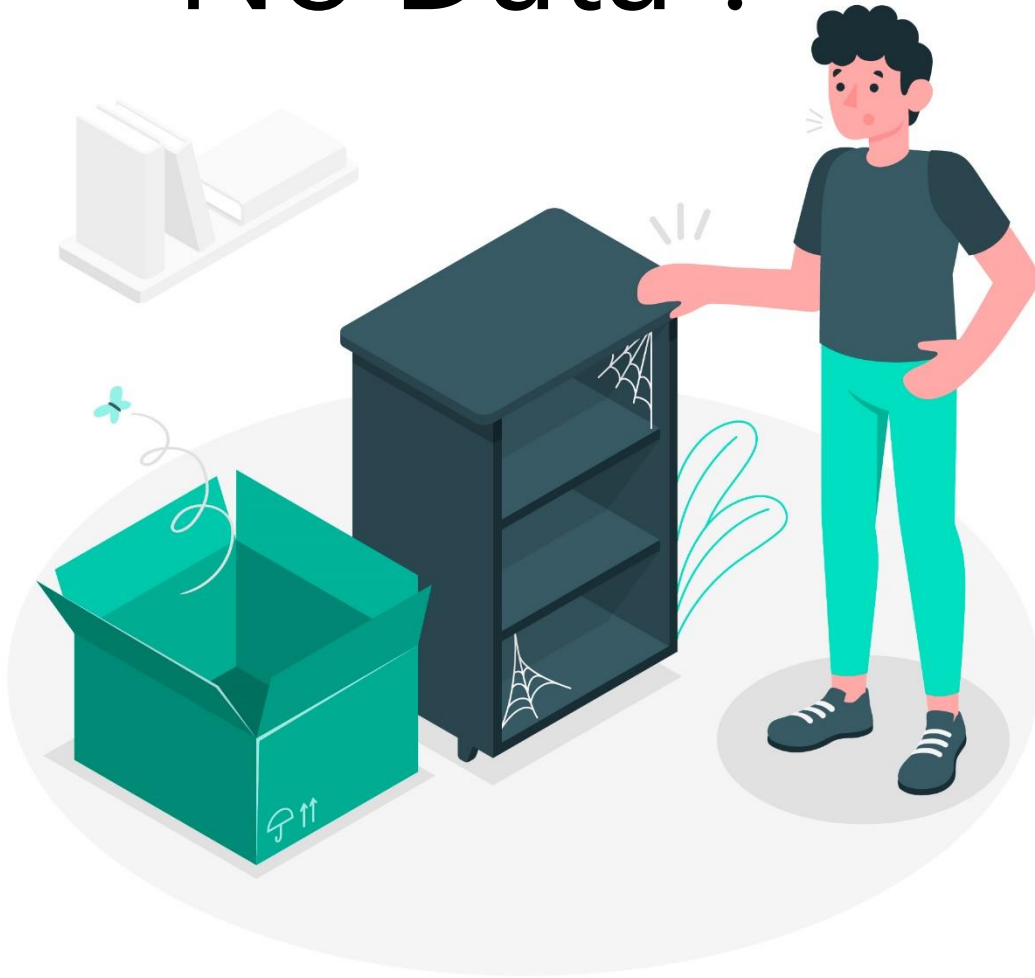
[閱讀使用我們的 免費應用程式](#)

Paperback
US\$39.99

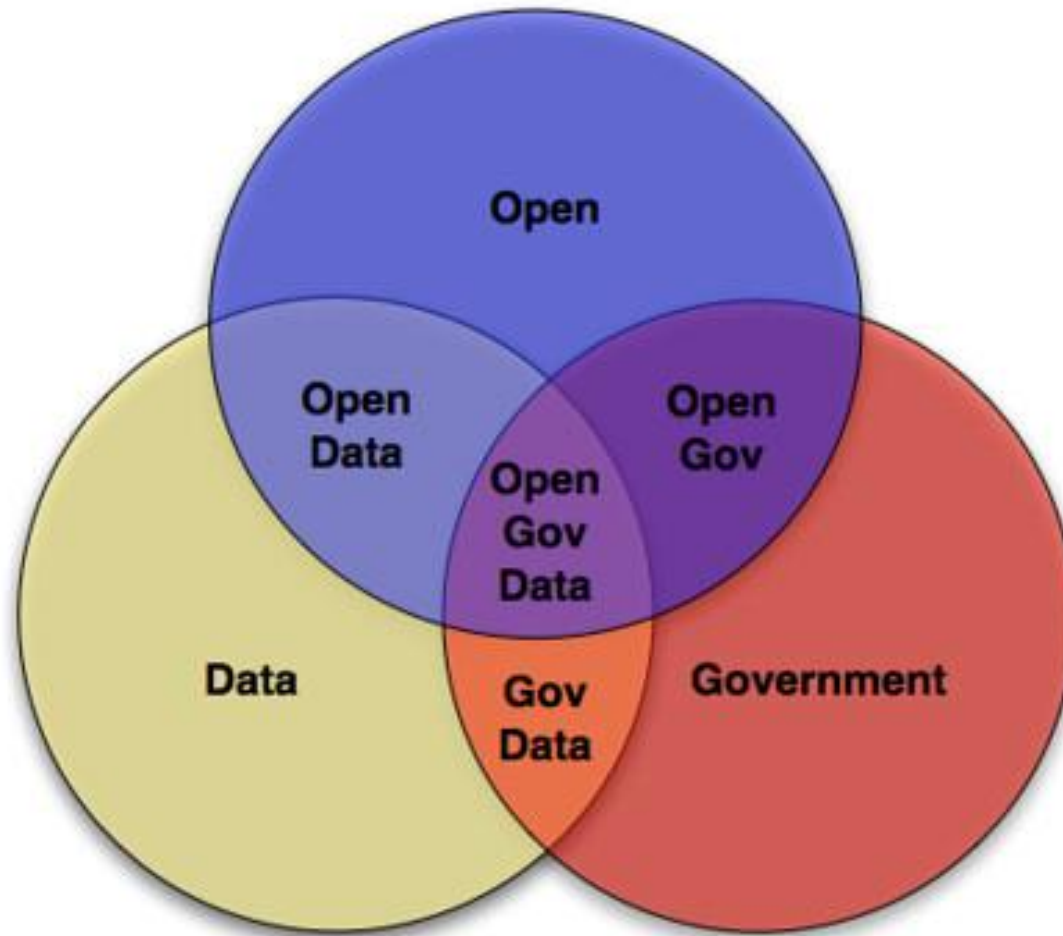
4 二手 起 US\$44.62
11 全新 起 US\$39.99

Think about your data intelligently and ask the right questions

No Data ?



Open Data



Demo



<https://openclipart.org/detail/16349/movie-clapperboard>

Q1

- 臺灣基層診所裡，有幾家家庭醫學科診所？
(執登含家醫科的診所)



首頁

理事長

理監事

醫學倫理

醫療統計

國際事務

專科醫師

醫療法規

會員信箱

醫療統計

醫療統計說明 歷年統計年鑑

首頁 / 醫療統計



公會簡介



台灣醫界雜誌



會員服務



醫師繼續教育

歷年統計年鑑

2021年台灣執業醫師・醫療機構統計(PDF)

統計圖

台灣各縣市醫師公會2021年底會員數

2021年台灣執業醫師醫療機構統計(PDF)

https://www.tma.tw/statistics/index_NYearInfo.asp?/2021.html

- ➡ 表8 各縣市各科別執業醫師數
- ➡ 表9 各鄉鎮市區各科別執業醫師數
- ➡ 表10 執業醫師年齡暨性別統計
- ➡ 表11 近年會員身故年齡暨性別統計
- ➡ 表12 執業醫師學歷統計
- ➡ 表13 執業醫師籍貫統計
- ➡ 表14 2012~2021 年首次入會會員統計
- ➡ 表15 2021年各縣市醫師公會會員移動差比表
- ➡ 表16 各級醫師公會2012 ~ 2021年會員統計
- ➡ 表17 台灣2012~2021年人口統計
- ➡ 表18 台灣人口、醫師公會會員2012~2021年動態
- ➡ 表19 各類醫療機構各科別執業醫師統計
- ➡ 表20 各縣市【醫學中心】各科別執業醫師數
- ➡ 表21 各縣市【區域醫院】各科別執業醫師數
- ➡ 表22 各縣市【地區醫院】各科別執業醫師數
- ➡ 表23 各縣市【新制精神科醫院】各科別執業醫師數
- ➡ 表24 各縣市【基層院所】各科別執業醫師數

Not
suitable

人數 ≠ 家數

請輸入關鍵字



進階搜尋

重設

熱門關鍵字：公告地價、疫情、營業稅

資料集服務分類



生育保健



出生及收養



求學及進修



服兵役



求職及就業



開創事業



婚姻



投資理財



休閒旅遊



交通及通訊



就醫



購屋及遷徙

opendata_20211231.csv
(accessed 25 Sep 2022)

23,302 records
one record for each clinic

醫療機構與人員基本資料

醫療機構之基本資料(含名稱、地址等)

評分此資料集：



平均 3.00 (1 人次投票)

👤 瀏覽次數: 36649 📄 下載次數: 7435 💬 意見數: 26

🖨 列印

主要欄位說明

*粗體欄位為資料標準欄位

機構代碼、機構名稱、權屬別、型態別、縣市區名、電話、地址、**診療科別**、1語言治療師、2牙體技術師、3聽力師、4牙體技術士、5驗光師、6驗光生、A醫師、B中醫師、C牙醫師、D藥師、E藥劑生、F護理師、G護士、H助產士、I助產師、J醫事檢驗師、K醫事檢驗生、L鑲牙生、M中藥員、N齒模員、P接骨員、Q物理治療師、R職能治療師、S醫事放射師、T醫事放射士、U物理治療生、W職能治療生、V呼吸治療師、X諮商心理師、Y臨床心理師、Z營養師

	A	B	C	D	E	F	G	H
1	機構代碼	機構名稱	權屬別	型態別	縣市區名	電話	地址	診療科別
2	3536020348	建民聯合	私立西醫診所	西醫專科診所	臺中市東勢區	04_258739	臺中市	家庭醫學科_內科_外科_
3	3536020357	東勢大學	私立西醫診所	西醫專科診所	臺中市東勢區	04_257744	臺中市	眼科_
4	3536020375	一民診所	私立西醫診所	西醫專科診所	臺中市東勢區	04-258751	臺中市	家庭醫學科_
5	3536021425	東勢蔡外	私立西醫診所	西醫專科診所	臺中市東勢區	04-258865	臺中市	家庭醫學科_內科_外科_
6	3536021434	建康耳鼻	私立西醫診所	西醫專科診所	臺中市東勢區	04-2587-9	臺中市	家庭醫學科_耳鼻喉科_
7	3536021452	長頸鹿小	私立西醫診所	西醫專科診所	臺中市東勢區	04-257718	臺中市	兒科_神經科_
8	3536021489	湯瑪仕小	私立西醫診所	西醫專科診所	臺中市東勢區	(04)25772	臺中市	外科_兒科_
9	3536010717	文華婦產	私立西醫診所	西醫專科診所	臺中市豐原區	04_252733	臺中市	婦產科_
10	3536010771	謝錦松內	私立西醫診所	西醫專科診所	臺中市豐原區	04_524131	臺中市	內科_兒科_
11	3536010799	重仁骨科	私立西醫診所	西醫專科診所	臺中市豐原區	04_252370	臺中市	骨科_復健科_
12	3536010833	洪榮信診	私立西醫診所	西醫專科診所	臺中市豐原區	04_252871	臺中市	婦產科_
13	3536010879	呂揚德內	私立西醫診所	西醫專科診所	臺中市豐原區	04-252762	臺中市	內科_
14	3536010888	紀鴻明診	私立西醫診所	西醫專科診所	臺中市豐原區	04_252601	臺中市	內科_
15	3536010897	長庚黃啟	私立西醫診所	西醫專科診所	臺中市豐原區	04_252879	臺中市	兒科_
16	3534031503	吳昌達診	私立西醫診所	西醫專科診所	宜蘭縣蘇澳鎮	03-995597	宜蘭縣	家庭醫學科_
17	3534040093	博生診所	私立西醫診所	西醫專科診所	宜蘭縣頭城鎮	03-977898	宜蘭縣	家庭醫學科_

In MS Excel:

常用

/ 排序與篩選

/ 篩選(F)

select:

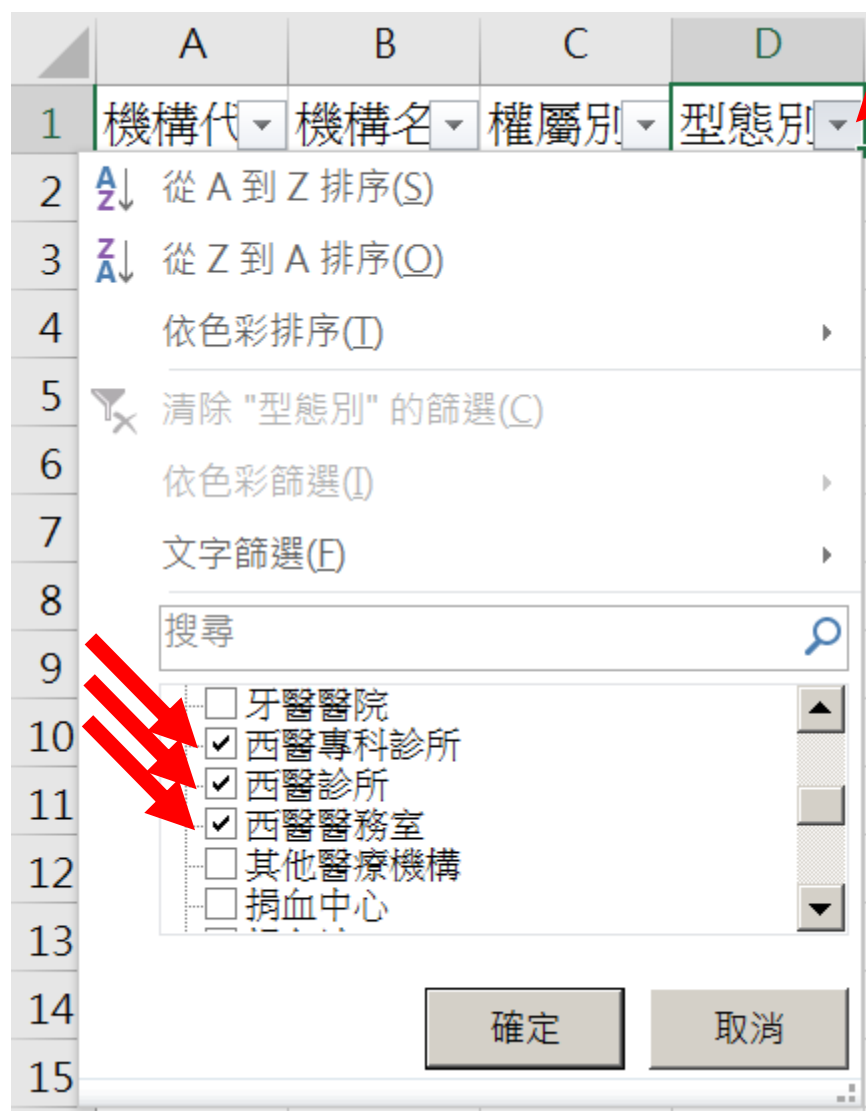
V 西醫專科診所

V 西醫診所

V 西醫醫務室

V 衛生所

=> 11,829 records



In MS Excel:
常用
/ 尋找與選取
/ 尋找(F)...

=> 1,683 records



Q2

- 挑出含家庭醫學科診所的記錄，拷貝至他處，以便進一步分析。

Processing in MS Excel

- 可先拷貝 診所資料 至新檔案
(e.g. opendata_20211231_PhysicianClinics.xlsx)
- 增加一個欄位： 常用 / [儲存格] 插入 / 插入工作表欄(C)
- 在新欄位第二格處填入公式：**=IFERROR(FIND("家庭醫學科_", \$H2), "")**
 - 在H2儲存格 尋找 FIND "家庭醫學科_"
(如果找到 儲存格會顯示 "家庭醫學科_" 出現在第幾個位元處)
如果沒找到 IFERROR 讓儲存格顯示空白 ""
 - \$H2 可只寫成 H2
\$是固定欄或列 方便其他欄位如法炮製
(絕對位址 vs. 相對位址)
- 將公式往下複製至其他儲存格
- 再篩選 新欄位裡 非空格者，供拷貝 家庭醫學科診所資料 至他處

I2									=IFERROR(FIND("家庭醫學科_", \$H2), "")
	A	B	C	D	E	F	G	H	I
1	機構代碼	機構名稱	權屬別	型態別	縣市區名	電話	地址	診療科別	家庭醫學科
2	4001020019	基督復臨	醫療財團	西醫診所	臺北市大安區	2700-2899	臺北市	家庭醫學科_內科_	1
3	3539031772	蔡鼎族小	私立西醫	西醫醫務	雲林縣虎尾鎮	05-636221	雲林縣	兒科_	
4	3539031790	陳俊誠小	私立西醫	西醫醫務	雲林縣虎尾鎮	05-631488	雲林縣	兒科_	
5	3539031825	光成診所	私立西醫	西醫醫務	雲林縣虎尾鎮	05-636612	雲林縣	兒科_	
6	3539031852	邢學平小	私立西醫	西醫專科	雲林縣虎尾鎮	05-631124	雲林縣	兒科_	
7	3539031861	端木梁診	私立西醫	西醫專科	雲林縣虎尾鎮	05-636258	雲林縣	復健科_	
8	3539031870	江重宏耳	私立西醫	西醫專科	雲林縣虎尾鎮	05-631179	雲林縣	耳鼻喉科_	
9	3539031914	大虎尾眼	私立西醫	西醫專科	雲林縣虎尾鎮	05-631410	雲林縣	眼科_	
10	3539031950	新容皮膚	私立西醫	西醫專科	雲林縣虎尾鎮	05-636575	雲林縣	皮膚科_	
11	3539031978	廖慶源婦	私立西醫	西醫專科	雲林縣虎尾鎮	05-633559	雲林縣	婦產科_	
12	3539032028	合康診所	私立西醫	西醫診所	雲林縣虎尾鎮	05-622211	雲林縣	虎尾鎮中興路72號	
13	3539032037	廖寶全診	私立西醫	西醫診所	雲林縣虎尾鎮	05-632258	雲林縣	精神科_	
14	3539032046	腎安診所	私立西醫	西醫診所	雲林縣虎尾鎮	05-633012	雲林縣	內科_	
15	3539040048	天德內小	私立西醫	西醫專科	雲林縣西螺鎮	05_86223	雲林縣	內科_兒科_	
16	3539040119	奕雄診所	私立西醫	西醫專科	雲林縣西螺鎮	55862988	雲林縣	西螺鎮延平路404號	
17	3539040146	茂杉診所	私立西醫	西醫專科	雲林縣西螺鎮	05_58622	雲林縣	家庭醫學科_	29 1

Q3

- 含家庭醫學科的診所裡，有幾家還執登其他專科？

Processing in MS Excel

- 增加一個欄位
- 在新欄位第二格處填入公式： $\text{=LEN(H2) - LEN(SUBSTITUTE(H2, "_", ""))}$
 - 計算 H2儲存格 文字字串的字元數 (LEN)
 - 將H2裡的"_" 取代 SUBSTITUTE 為空白 ""
 - 兩者相減 即為 "_" 出現次數
 - 如果每次取代字元數大於1 ($= n$)，則相減結果須再除以 n
- 將公式往下複製至其他儲存格
- 再進一步 排序 或 篩選 或 樞紐分析，...
- 也可計算 [診療科別] 欄位裡每個儲存格的文字長度 =LEN(H2) ，再排除長度等於6者
- 由於家庭醫學科都排在最前面，也可直接排序[診療科別]欄位，...

I2									=LEN(H2) - LEN(SUBSTITUTE(H2, "_", ""))
	A	B	C	D	E	F	G	H	I
1	機構代碼	機構名稱	權屬別	型態別	縣市區名	電話	地址	診療科別	診療科別數
2	4001020019	基督復臨	醫療財團	西醫診所	臺北市大安區	2700-2899	臺北市	家庭醫學科_內科_	2
3	3539040146	茂杉診所	私立西醫	西醫專科	雲林縣西螺鎮	05_58622	雲林縣	家庭醫學科_	1
4	3539040208	泓志診所	私立西醫	西醫專科	雲林縣西螺鎮	05_58646	雲林縣	家庭醫學科_內科_	2
5	3539020046	吳國猷診	私立西醫	西醫專科	雲林縣斗南鎮	05_59623	雲林縣	家庭醫學科_	1
6	3538080155	良平診所	私立西醫	西醫診所	南投縣中寮鄉	049-26919	南投縣	家庭醫學科_	1
7	3538030771	簡國仁診	私立西醫	西醫診所	南投縣草屯鎮	049-23334	南投縣	家庭醫學科_	1
8	3538030851	錦林診所	私立西醫	西醫診所	南投縣草屯鎮	049-23575	南投縣	家庭醫學科_兒科_	2
9	3537200019	恩生家醫	私立西醫	西醫專科	彰化縣二水鄉	04_87936	彰化縣	家庭醫學科_外科_	2
10	3537210079	洪明仁診	私立西醫	西醫醫務	彰化縣田尾鄉	48835135	彰化縣	家庭醫學科_	1
11	3537220011	光明診所	私立西醫	西醫醫務	彰化縣埤頭鄉	04_89246	彰化縣	家庭醫學科_	1
12	3537221134	陳嘉烈診	私立西醫	西醫專科	彰化縣埤頭鄉	04—8920	彰化縣	家庭醫學科_	1
13	3537121193	佑美診所	私立西醫	西醫診所	彰化縣秀水鄉	04-769828	彰化縣	家庭醫學科_	1
14	3537131224	德亮診所	私立西醫	西醫專科	彰化縣花壇鄉	04-786564	彰化縣	家庭醫學科_	1
15	3537050859	蕭?森診所	私立西醫	西醫專科	彰化縣員林市	04_83579	彰化縣	家庭醫學科_內科_	2
16	3537052120	育英診所	私立西醫	西醫專科	彰化縣員林市	48331458	彰化縣	家庭醫學科_耳鼻喉科_	2
17	3537052264	盧天恩診	私立西醫	西醫專科	彰化縣員林市	04-833915	彰化縣	家庭醫學科_內科_	2

Q4

- 含家庭醫學科的診所裡，還執登其他專科者，每個專科各有幾家診所？

Processing in MS Excel - 1

- 選取 [診療科別] 整個欄位
- 資料 / [資料工具] 資料剖析
 - 原 [診療科別] 欄位 被切割為 15個欄位 (H -> V)
- 在空白欄位處填入公式： **=COUNTIF**(I2:V1684, "內科")
 - 在其他專科欄位區間裡 計算 "內科" 出現的次數
- 計算其他科別出現的次數，如法炮製 (i.e. 更換 內科 為 其他科)
 - 也可在別處填入各專科名稱(一科一儲存格)，再將公式裡的科別 改為 參照位址



資料剖析精靈判定資料類型為分隔符號。

若一切設定無誤，請選取 [下一步]，或選取適當的資料類別。

原始資料類型

請選擇最適合剖析您的資料的檔案類型：

- ☒ 分隔符號(D) - 用分欄字元，如逗號或 TAB 鍵，區分每一個欄位。
- ☐ 固定寬度(W) - 每個欄位固定，欄位間以空格區分。

預覽選取的資料：

1	診療科別
2	家庭醫學科_內科_
3	家庭醫學科_
4	家庭醫學科_內科_
5	家庭醫學科_
6	家庭醫學科_

取消

< 上一步(B)

下一步(N) >

完成(E)

您可在畫面中選擇輸入資料中所包含的分隔符號，您可在預覽視窗內看到分欄的結果。

分隔符號

☐ Tab 鍵(T)☐ 分號(M)☐ 逗點(C)☐ 空格(S)☒ 其他(O):☐ 連續分隔符號視為單一處理(R)

文字辨識符號(Q):

 ▼

預覽分欄結果(P)

診療科別	
家庭醫學科	內科
家庭醫學科	內科
家庭醫學科	內科
家庭醫學科	內科
家庭醫學科	內科

取消

< 上一步(B)

下一步(N) >

完成(E)

請在此畫面選擇欲使用的欄位，並設定其資料格式。

欄位的資料格式

☐ 一般(G)

☒ 文字(T)

☐ 日期(D): YMD

☐ 不匯入此欄(I)

「一般」資料格式會使得數值被轉成數字格式，日期值被轉成日期欄格式，其餘資料則被轉成文字格式。

進階(A)...

目標儲存格(E): \$H\$1

預覽分欄結果(P)

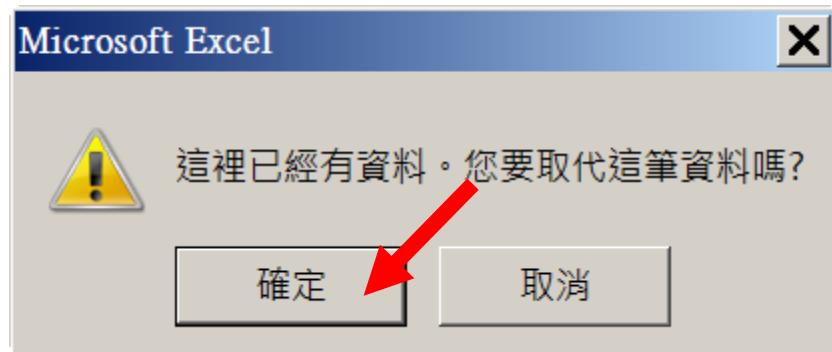
文字	一般	一般
診療科別		
家庭醫學科	內科	
家庭醫學科		
家庭醫學科	內科	
家庭醫學科		
家庭醫學科		

取消

< 上一步(B)

下一步(N) >

完成(E)



I1

✕

✓

fx

=COUNTIF(I2:V1684,"內科")

	A	B	C	D	E	F	G	H	I
1	機構代碼	機構名稱	權屬別	型態別	縣市區名	電話	地址	診療科別	190
2	4001020019	基督復臨	醫療財團	西醫診所	臺北市大安區	2700-2899	臺北市	家庭醫學科	內科
3	3539040146	茂杉診所	私立西醫	西醫專科	雲林縣西螺鎮	05_58622	雲林縣	家庭醫學科	
4	3539040208	泓志診所	私立西醫	西醫專科	雲林縣西螺鎮	05_586468	雲林縣	家庭醫學科	內科
5	3539020046	吳國猷診	私立西醫	西醫專科	雲林縣斗南鎮	05_596234	雲林縣	家庭醫學科	
6	3538080155	良平診所	私立西醫	西醫診所	南投縣中寮鄉	049-26919	南投縣	家庭醫學科	
7	3538030771	簡國仁診	私立西醫	西醫診所	南投縣草屯鎮	049-23334	南投縣	家庭醫學科	
8	3538030851	錦林診所	私立西醫	西醫診所	南投縣草屯鎮	049-23575	南投縣	家庭醫學科	兒科
9	3537200019	恩生家醫	私立西醫	西醫專科	彰化縣二水鄉	04_879363	彰化縣	家庭醫學科	外科
10	3537210079	洪明仁診	私立西醫	西醫醫務	彰化縣田尾鄉	48835135	彰化縣	家庭醫學科	
11	3537220011	光明診所	私立西醫	西醫醫務	彰化縣埤頭鄉	04-_89246	彰化縣	家庭醫學科	
12	3537221134	陳嘉烈診	私立西醫	西醫專科	彰化縣埤頭鄉	04—8920	彰化縣	家庭醫學科	
13	3537121193	佑美診所	私立西醫	西醫診所	彰化縣秀水鄉	04-769828	彰化縣	家庭醫學科	
14	3537131224	德亮診所	私立西醫	西醫專科	彰化縣花壇鄉	04-786564	彰化縣	家庭醫學科	
15	3537050859	蕭?森診所	私立西醫	西醫專科	彰化縣員林市	04_835799	彰化縣	家庭醫學科	內科
16	3537052120	育英診所	私立西醫	西醫專科	彰化縣員林市	48331458	彰化縣	家庭醫學科	耳鼻喉科
17	3537052264	盧天恩診	私立西醫	西醫專科	彰化縣員林市	04-833915	彰化縣	家庭醫學科	內科

Processing in MS Excel - 2

- 新增 22 個欄位 (AN -> BI)，第一列儲存格分別填入其他專科名稱
- 新欄位第二格處填入公式：**=IF(IFERROR(FIND(AN\$1, \$H2), 0) >= 1, 1, 0)**
 - 在H2儲存格 尋找 FIND "內科" (AN1)
(如果找到 儲存格會顯示 "內科" 出現在第幾個位元處)
如果沒找到 IFERROR 讓儲存格為0
若 IF 儲存格的值大於等於1，讓儲存格的值變為1，否則為0
- 將公式向右複製到其他專科欄位
- 將公式向下複製至其他儲存格
- 22 個欄位分別加總

AN2 × ✓ <i>fx</i> =IF(IFERROR(FIND(AN\$1, \$H2), 0) >= 1, 1, 0)											
	H	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
1	診療科別	内科	外科	兒科	婦産科	骨科	神経外科	泌尿科	耳鼻咽喉科	眼科	皮膚科
2	家庭醫學科_内科_	1	0	0	0	0	0	0	0	0	0
3	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
4	家庭醫學科_内科_	1	0	0	0	0	0	0	0	0	0
5	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
6	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
7	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
8	家庭醫學科_兒科_	0	0	1	0	0	0	0	0	0	0
9	家庭醫學科_外科_	0	1	0	0	0	0	0	0	0	0
10	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
11	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
12	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
13	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
14	家庭醫學科_	0	0	0	0	0	0	0	0	0	0
15	家庭醫學科_内科_	1	0	0	0	0	0	0	0	0	0
16	家庭醫學科_耳鼻咽喉科_	0	0	0	0	0	0	0	1	0	0
17	家庭醫學科_内科_	1	0	0	0	0	0	0	0	0	0
18	家庭醫學科_婦産科_	0	0	0	1	0	0	0	0	0	0

Processing in Perl

```
while ($line = <DATA>) {      # 迴圈方式，逐列(line)讀入資料處理之
    @s = split /_/, (split /\t/, $line)[7];    # 先依tab鍵分割，取第8欄位，再依_分割
    foreach (@s) { $seen{$_}++ }    # 將分割出來的科別，逐一儲存至hash
}
```

```
foreach ( keys %seen ) { print $_ . "\t" . $seen{$_} . "\n"; }    # 列印hash
```

```
__DATA__    # 以下擺放資料
```

Key concept: **associative array**

https://en.wikipedia.org/wiki/Associative_array

```
# FMClinics_SpecialtyCount.plx (先轉以UTF-8格式存檔)
```


家庭醫學科	1683
內科	190
外科	98
兒科	126
婦產科	79
骨科	34
神經外科	2
泌尿科	5
耳鼻喉科	48
眼科	16
皮膚科	15
神經科	21
精神科	14
復健科	38
麻醉科	2
放射診斷科	2
急診醫學科	2
職業醫學科	5
整形外科	4

Note:

In computer science, an **associative array**, **map**, **symbol table**, or **dictionary** is an abstract data type that stores a collection of **(key, value)** pairs. ... It supports 'lookup', 'remove', and 'insert' operations.

https://en.wikipedia.org/wiki/Associative_array

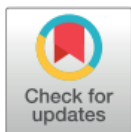
All Roads Lead to Rome



<https://pixabay.com/photos/rome-monument-colosseum-italy-2093608/>

Of Any Use?





Available online at www.sciencedirect.com

ScienceDirect

Journal of the Chinese Medical Association 81 (2018) 964–969



www.jcma-online.com

Original Article

Situation of seven-day service among family medicine clinics in Taiwan: A nationwide survey

Wei-Ting Chang^a, Yi-Jen Wang^b, Hsin-I Leu^c, Tzeng-Ji Chen^{b,d}, Shinn-Jang Hwang^{b,d},
Li-Fang Chou^e, Ming-Hwai Lin^{b,d,*}

Table 2

Number of open sessions among family medicine clinics, stratified by practice type and size.

Type of practice	Number of clinics (%, <i>n</i> = 1621)	Number of opening sessions (mean $\pm$ SD)
<u>Solo practice</u>	1080 (66.6)	15.0 $\pm$ 3.8
<u>Group practice</u>	507 (31.3)	16.9 $\pm$ 3.0
with 2 physicians	343 (21.2)	16.6 $\pm$ 3.7
with 3 physicians	104 (6.4)	17.3 $\pm$ 3.0
with ≥ 4 physicians	60 (3.7)	18.3 $\pm$ 1.9
Unknown	34 (2.1)	—



Article

In the Name of Family Medicine: A Nationwide Survey of Registered Names of Family Medicine Clinics in Taiwan

Ya-An Liu ¹, Sally Cheng ², Ya-Chuan Hsu ¹, Po-Chin Yang ¹, Hsiao-Ting Chang ^{1,3}, Ming-Hwai Lin ^{1,3}, Tzeng-Ji Chen ^{1,3,4,*}, Li-Fang Chou ⁵ and Shinn-Jang Hwang ^{1,3}

International Journal of Environmental Research and Public Health 2020;17:4062. [SCI] [SSCI]

<https://www.mdpi.com/1660-4601/17/11/4062>
<https://pubmed.ncbi.nlm.nih.gov/32517287/>
<https://wd.vghttp.gov.tw/fm/files/20200607.txt>
<https://wd.vghttp.gov.tw/fm/files/20180715.txt>

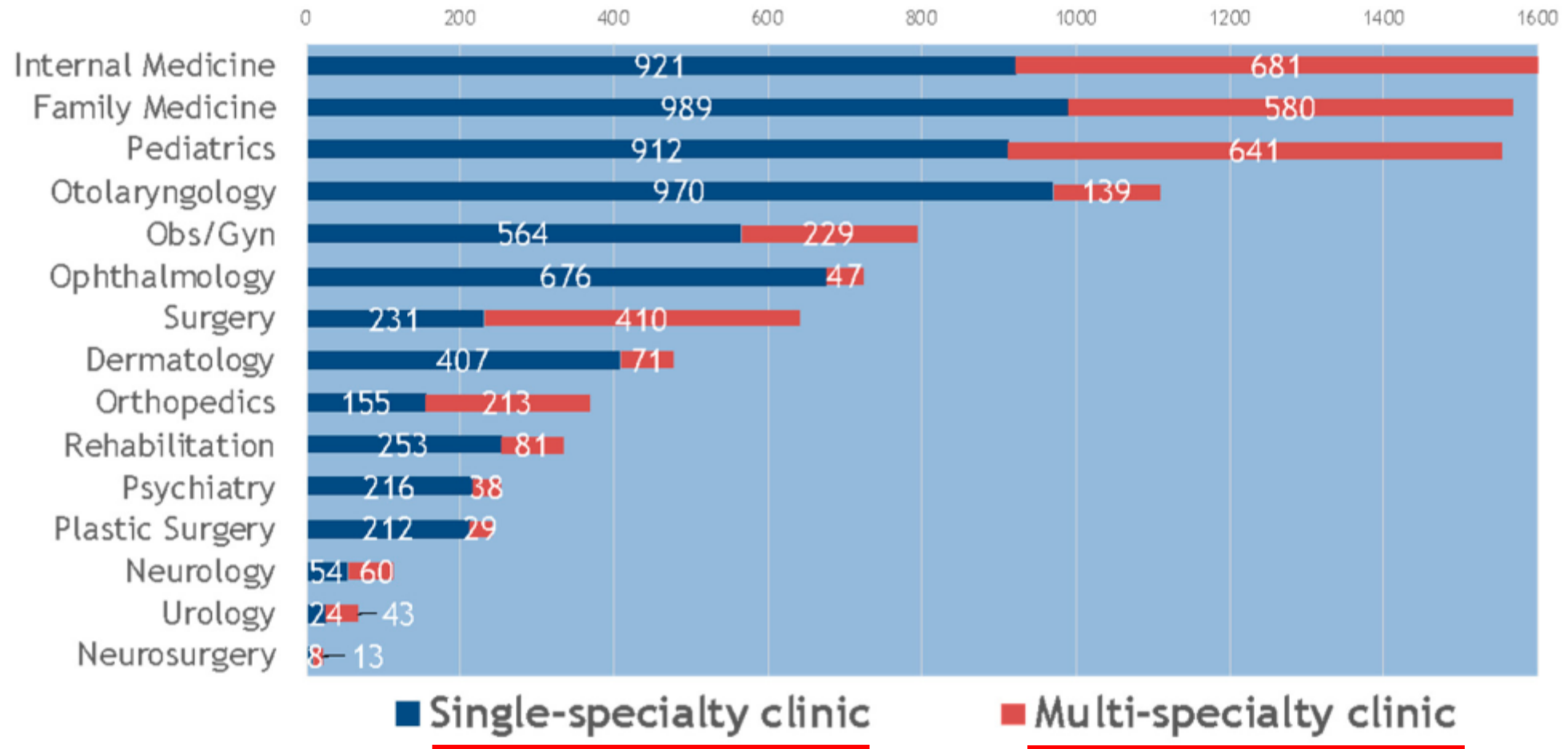


Figure 1. Numbers of single-specialty and multi-specialty clinics, stratified by specialty.

Data Wrangling Steps & Techniques

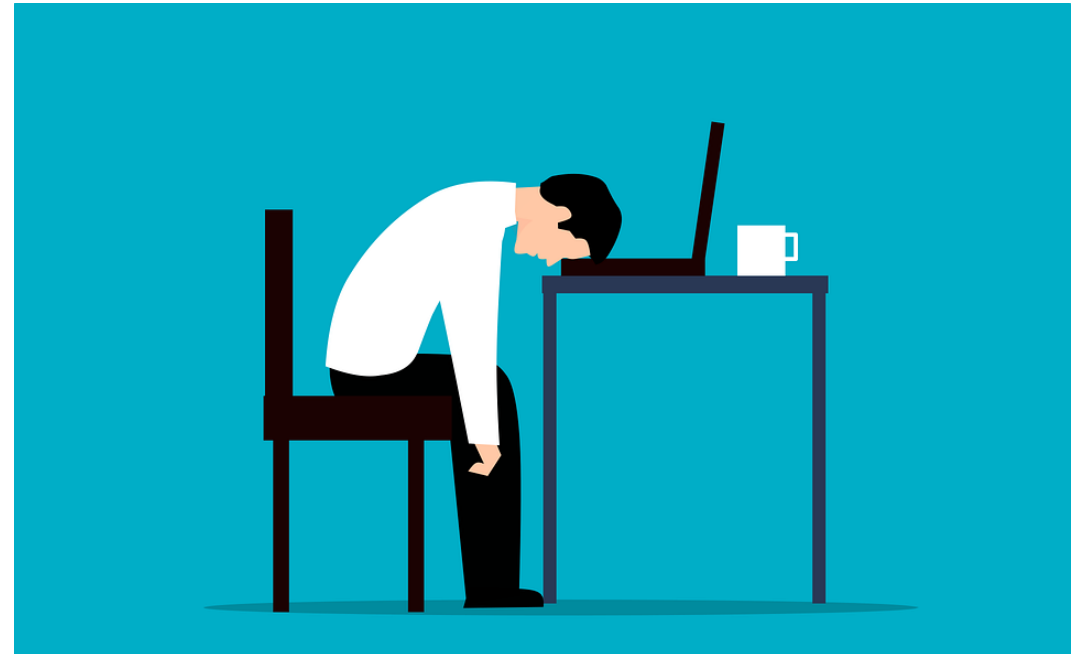
- discover
- structure
- clean
- enrich
- validate
- publish
- connect (join)
- merge (concatenate)
- transpose
- remove duplicates
- group
- export
-

https://en.wikipedia.org/wiki/Data_wrangling

<https://ithelp.ithome.com.tw/articles/10186310>

Other Tedious Works

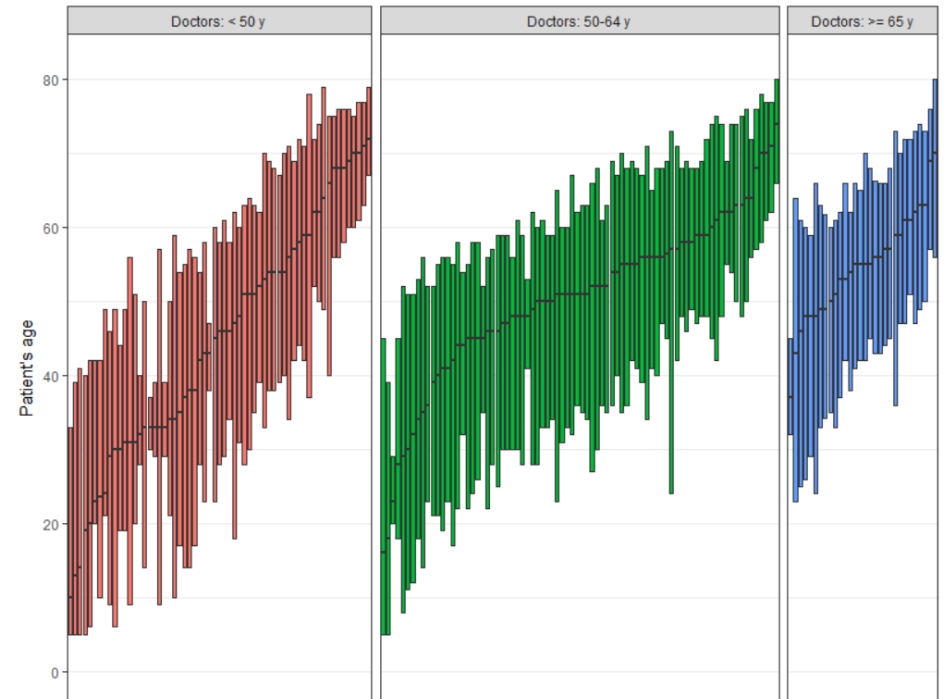
- column type inconsistency (text $\leftrightarrow$ number $\leftrightarrow$ date/time)
- encoding
- quotation marks
- line breaks
- trailing blanks
- missing values
-



<https://pixabay.com/vectors/employee-desk-stress-exhausted-6038877/>

Some ex-ante & ex-post efforts

- fetching data (from the web)
- data analysis
- data visualization
- data modelling
-



Int J Environ Res Public Health 2017;14:499.
[PMID: 28481318]

Tools for Data Wrangling ?



■ 社會

2022.07.20 05:58 臺北時間

【林百里妻遭詐騙3】電腦軟體不用太高檔 她只靠「小畫家」就拐走5億

文 | 劉文淵 林俊宏 攝影 | 李智為



<https://www.mirrormedia.mg/story/20220718inv014/>

Data Wrangling Tools

- Text editor : RE (regular expression)
- M\$ Excel : Functions, VBA (Visual Basic for Applications)
- Python
- R
- JavaScript
- SQL
-



<https://upload.wikimedia.org/wikipedia/commons/4/42/Opensource.svg>

Should data be as **BIG** as possible ?

The Tiny Clues That Uncover Huge Trends



發現大數據看不見的小細節
從消費欲望到行為分析
創造品牌商機

馬汀·林斯壯 Martin Lindstrom — 著
溫力樂、戴至中 — 譯





Original Article

I|J|S|P

Low prevalence of the use of the Chinese term for ‘psychiatry’ in the names of community psychiatry clinics: A nationwide study in Taiwan

n = 444

Ying-Xiu Dai¹, Mu-Hong Chen² and Tzeng-Ji Chen^{1,3}

International Journal of
Social Psychiatry
2016, Vol. 62(7) 601–607
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0020764016660994
isp.sagepub.com
SAGE



Article

Branding Palliative Care Units by Avoiding the Terms “Palliative” and “Hospice”: A Nationwide Study in Taiwan

n = 55

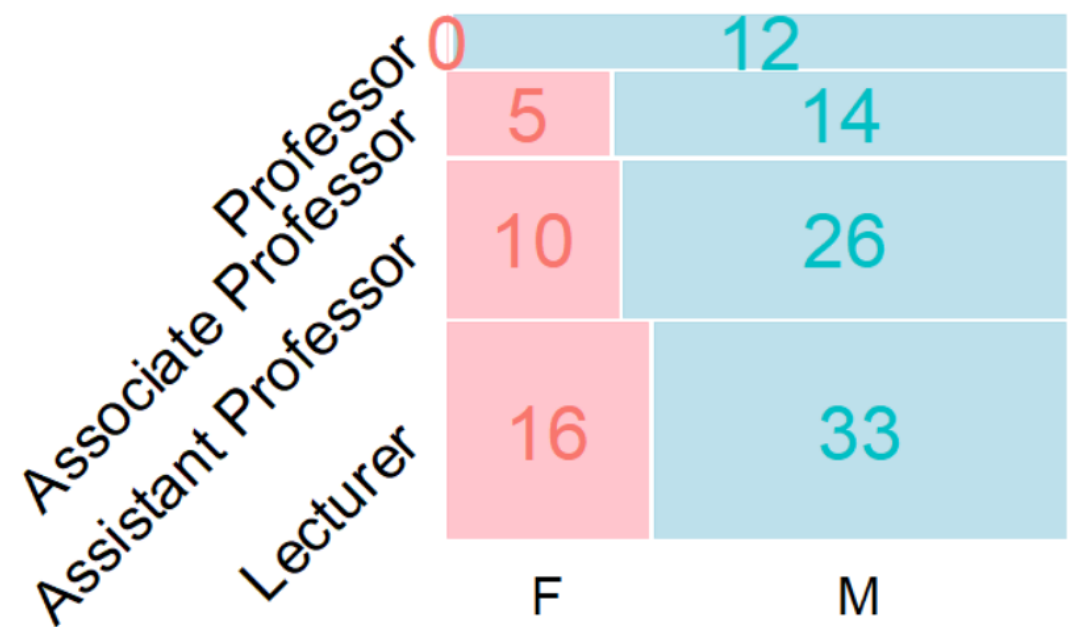
Ying-Xiu Dai, MD^{1,2}, Tzeng-Ji Chen, MD^{1,2}, and Ming-Hwai Lin, MD^{1,2}

INQUIRY: The Journal of Health Care
Organization, Provision, and Financing
Volume 54: 1–6
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0046958016686449
journals.sagepub.com/home/inq
SAGE



家庭醫學教師性別分布

n = 116



Status: August 2019

Chen SH, Chang HT, Lin MH, Chen TJ, Hwang SJ, Lin MN.
Family medicine academic workforce of medical schools in Taiwan: a nationwide survey.
Int J Environ Res Public Health 2021;18:7182. [SCI] [SSCI]

<https://www.mdpi.com/1660-4601/18/13/7182>
<https://pubmed.ncbi.nlm.nih.gov/34281118/>
<https://wd.vghttp.gov.tw/fm/files/20210705.txt>

Ideas Matter

