從資料整理開始
打好精準醫學研究根基

國立中山大學
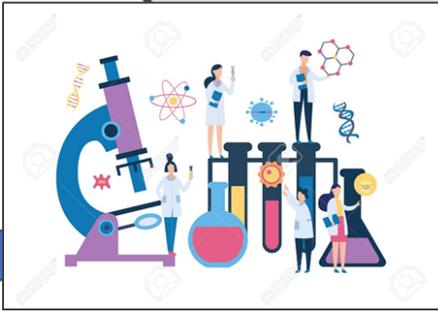
馮已榕

ijfeng@g-mail.nsysu.edu.tw

醫學生物技術暨檢驗學系（含碩士班）
Department of Medical Biotechnology and Laboratory Science, CGU
Graduate Institute of Medical Biotechnology and Laboratory Science

醫事檢驗 醫學 生物科學

CASE WESTERN RESERVE UNIVERSITY EST. 1826 Department of Population and Quantitative Health Sciences

生物統計
微積分 線性代數 數理統計
寫程式 SAS、R、Matlab
醫學影像

Main effect of face perception
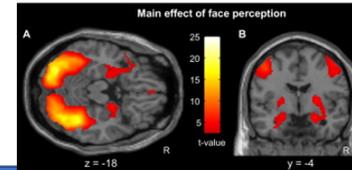
動物實驗
整形外科病例登錄

人工智慧 機器學習 深度學習
Python

台灣人工智慧學校
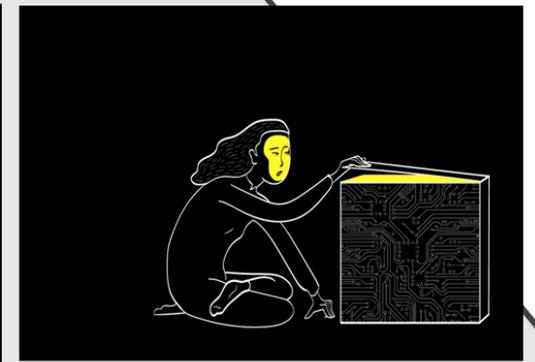
什麼是醫學(medicine)?

Predict → Disgnose → Treat → Monitor

如何預防?

| Predict | Disgnose | Treat | Monitor |

心肌梗塞的危險因子 你中了幾個？

健康吉美健檢中心
H&B HEALTH CENTER

- 年齡 男性>45歲或女性>55歲(或停經)
- 性別 男性>女性
- 三高 高血壓、高血糖、高血脂
- 家族史
- 吸菸、吸二手菸
- 久坐不動、活動量少
- 某些藥物 避孕藥和激素替代療法（HRT）

健康傳媒
Health Media

如何預防 失智症

趨吉
多動腦　多運動　維持適當體重　多社交互動　採地中海飲食

避凶
避免三高　預防頭部外傷　遠離抽菸及憂鬱

保護因子
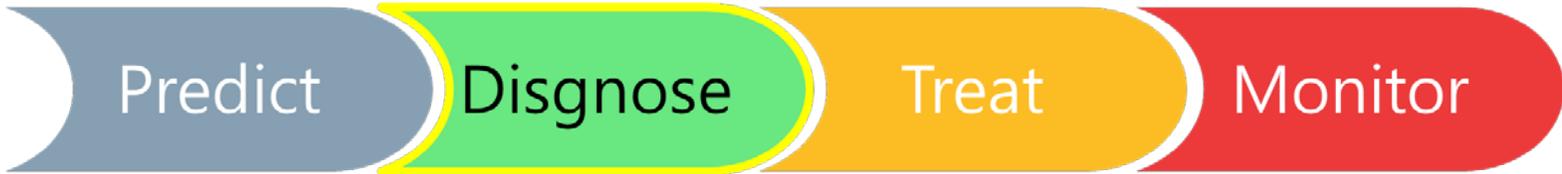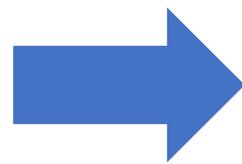
危險因子

如何診斷?

Predict → Disgnose → Treat → Monitor

- 症狀
- 身體檢查 {性別、年齡、身高、體重、心跳、血壓...}
- 病史
- 家族史
- 病理中心檢驗數據 {血液常規檢查(白血球、紅血球、血色素...)、生化檢查(糖化血色素、膽固醇...)、血清學檢查...}
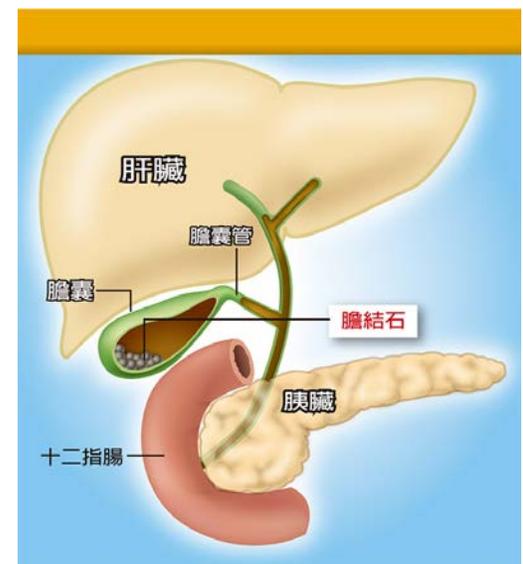- 醫療影像{x-ray 、CT 、MRI...}

如何診斷?

Predict → Disgnose → Treat → Monitor
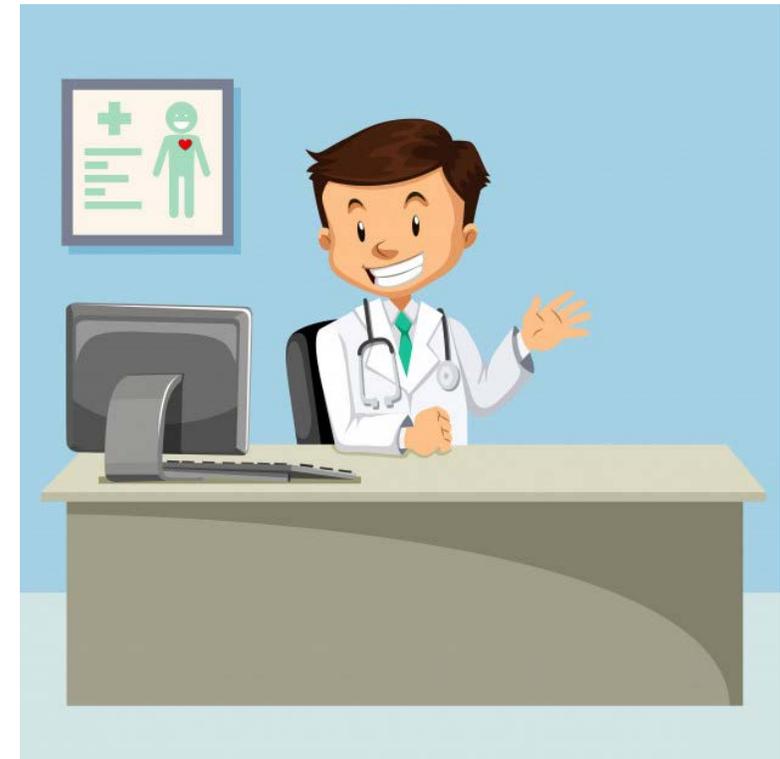
- **症狀 -** 腹部疼痛，有時甚至會感到胸部或右肩胛骨疼痛，無論如何改變姿勢都無法減輕疼痛。
- **身體檢查 –**
  大於40歲
  女性
  肥胖者（BMI>30）
- **病史**
- **家族史-**有膽結石家族
- **病理中心檢驗數據**
- **醫療影像 -** 腹部X光檢查、電腦斷層掃描或腹部超音波

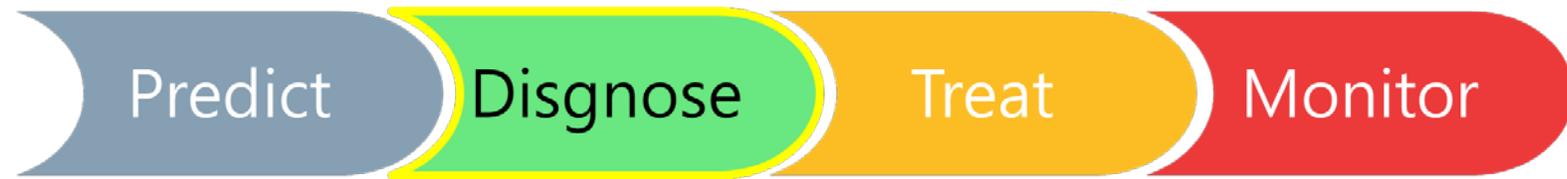→ **膽結石**

如何診斷?

- **症狀 -**尿多、口渴、飢餓、疲勞、視力模糊、體重減輕或傷口不易癒合

  吃多、喝多、尿多 + 體重減少 (三多一少)

- **身體檢查-** 40歲以上? 肥胖?

- **病史**

- **家族史-** ?

- **病理中心檢驗數據 –**

  1. 隨機血漿糖值（無論空腹與否的任意時間測得的血糖值）≥200（單位：mg/dl, 毫克/分公升）

  2. 空腹8小時後血糖 ≥126 mg/dl。

  3. 空腹口服75公克葡萄糖後測試（此檢查稱為葡萄糖耐受試驗），在2小時的血糖值≥ 200 mg/dl。

  4. 糖化血色素 (HbA1C) ≥6.5%。

- **醫療影像**

➡ 第2型糖尿病

如何診斷?

Predict | Disgnose | Treat | Monitor
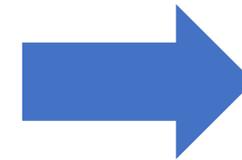
- **症狀-**
  發燒(≧38℃)
  急性呼吸道感染或嗅、味覺異常。
- **身體檢查**
- **病史**
- **家族史**
- **病理中心檢驗數據 –**
  1.臨床檢體(如鼻咽或咽喉擦拭液、痰液或下呼吸道抽取液等)分離並鑑定出新型冠狀**病毒**。
  2.臨床檢體新型冠狀**病毒**分子生物學核酸檢測陽性。
- **醫療影像**
  放射線診斷顯示有肺炎。

➡ **COVID-19**
(武漢肺炎)

Covid / Normal / Pneumonia Bacterial / Pneumonia Virus

如何治療？

Predict → Disgnose → Treat → Monitor
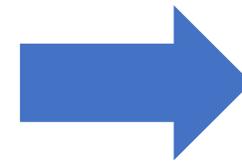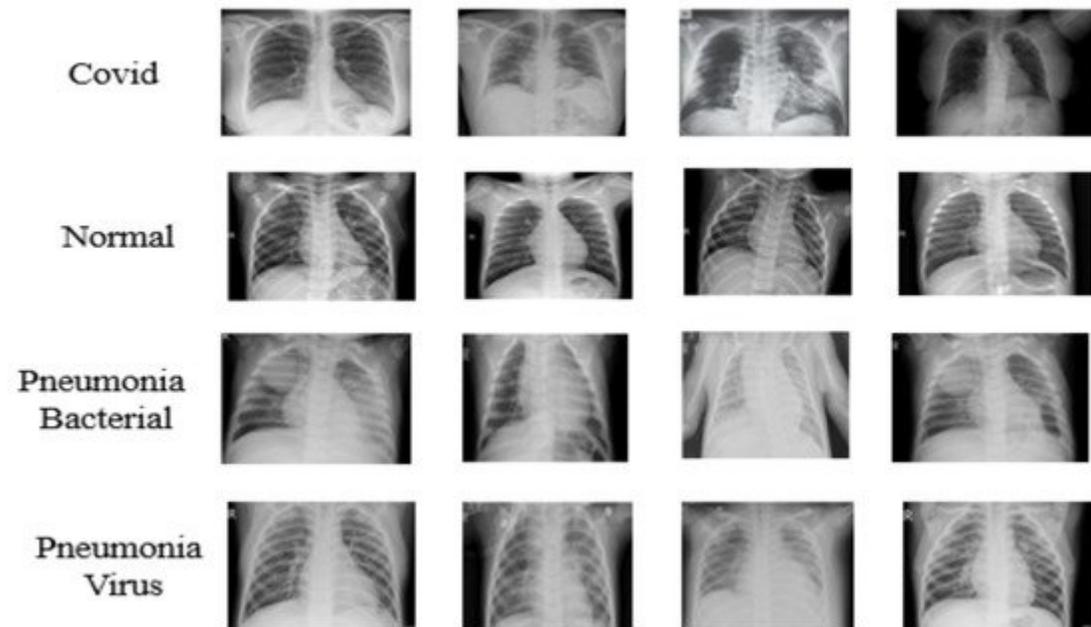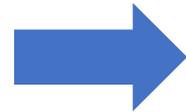
膽結石 ➡️ 膽囊切除手術
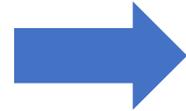
第2型糖尿病 ➡️ 胰島素/
降血糖藥物

糖尿病指的是人體內的胰臟不能製造足夠的胰島素，導致葡萄糖無法充分進入細胞內，血糖濃度就會升高形成糖尿病。
財團法人蘭陽仁愛醫院–糖尿病

COVID-19 ➡️ 呼吸器
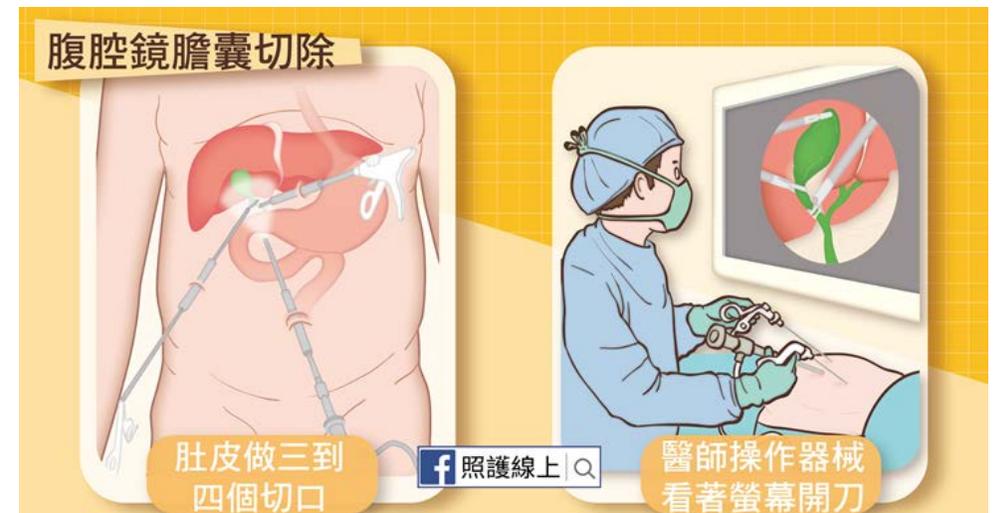
腹腔鏡膽囊切除
肚皮做三到四個切口
醫師操作器械看著螢幕開刀
照護線上

如何追蹤?

Predict	Disgnose	Treat	Monitor

- 症狀
- 身體檢查 {性別、年齡、身高、體重、心跳、血壓…}
- 病史
- 家族史
- 病理中心檢驗數據 {血液常規檢查(白血球、紅血球、血色素…)、生化檢查(糖化血色素、膽固醇…)、血清學檢查…}
- 醫療影像{x-ray 、CT 、MRI…}

傳統研究方式

Koch's postulates

描述疾病（通常是傳染病）與病原菌之間因果關係的條件

德國醫師 / 生理學家Robert Koch。
Image courtesy of Nobelprize.org

Koch's Postulates:

① The microorganism must be found in abundance in all organisms suffering from the disease, but should not be found in healthy organisms.

② The microorganism must be isolated from a diseased organism and grown in pure culture.

③ The cultured microorganism should cause disease when introduced into a healthy organism.

④ The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.

1. 病體身上可以找到大量致病病原菌，而在健康活體上找不到這些病原菌。
2. 這些病原菌可以從病體身上分離出，而且可以在適當的培養基上生長。
3. 培養出的病原菌可以造成原本健康的活體患病。
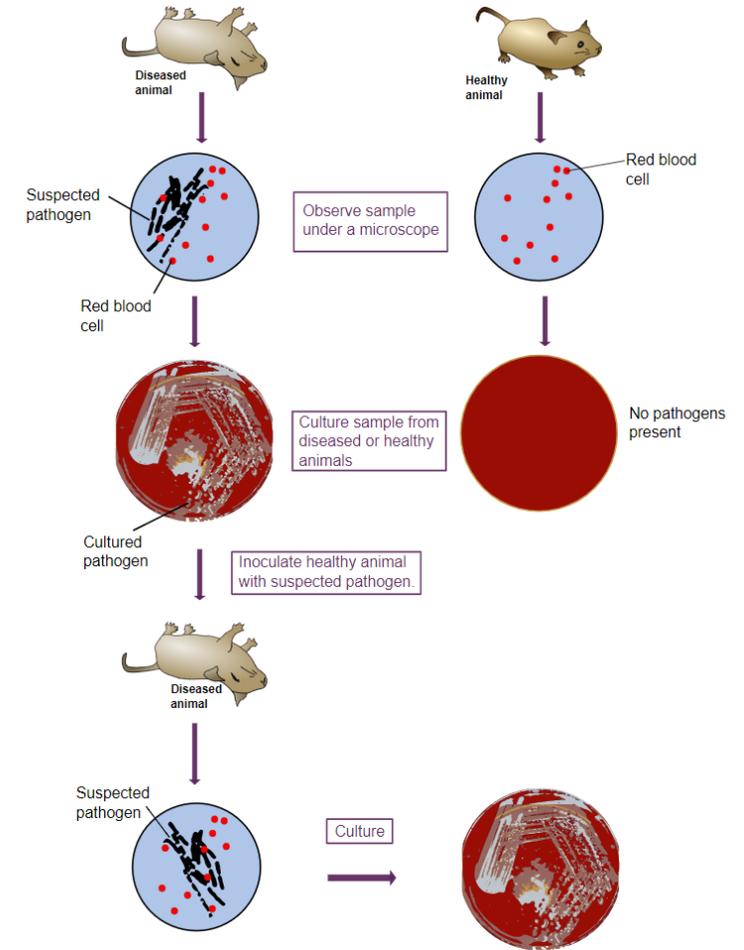4. 從這些因為接種了培養出的病原菌而患病的病體身上，可以再次分離出和原先培養一樣的病原菌菌種。

研究方式

"年齡"是否與"聽過鋼彈"相關?
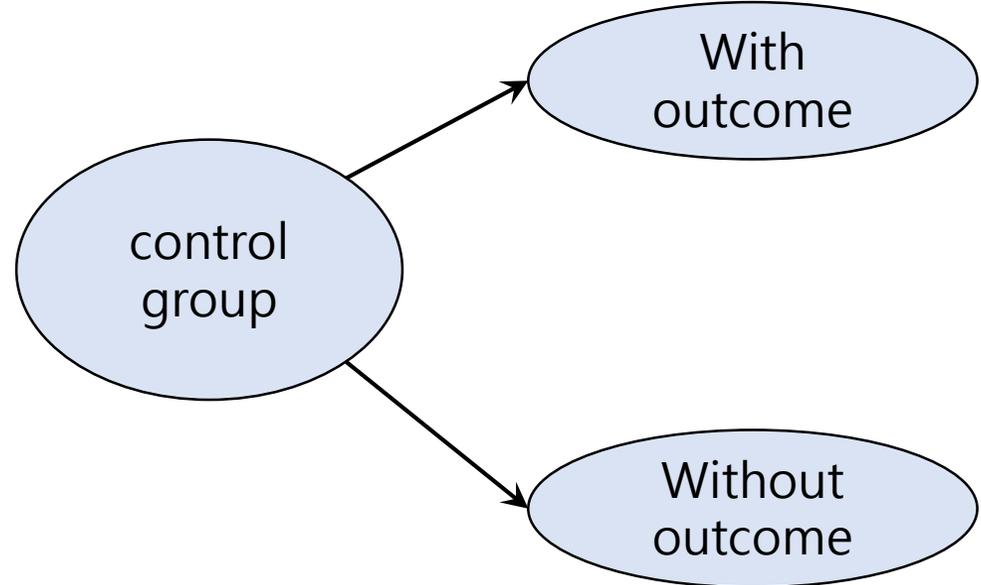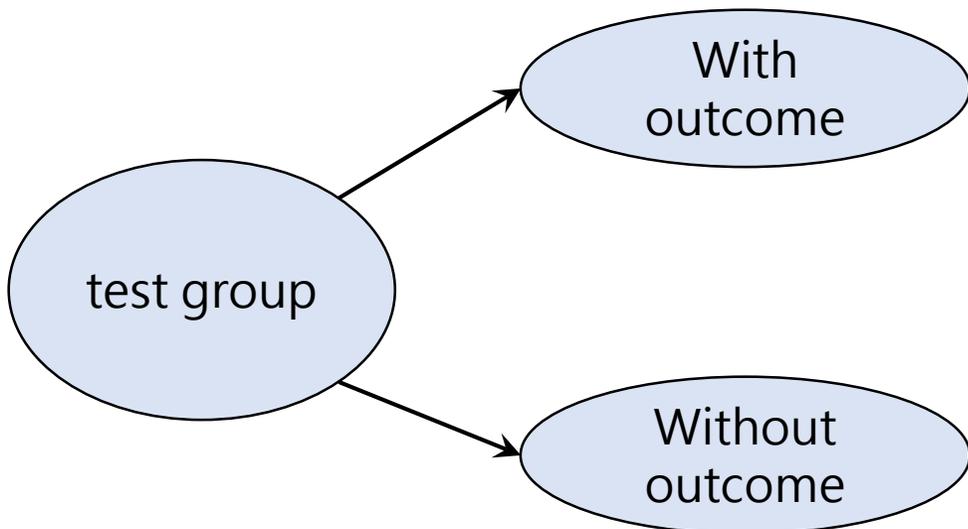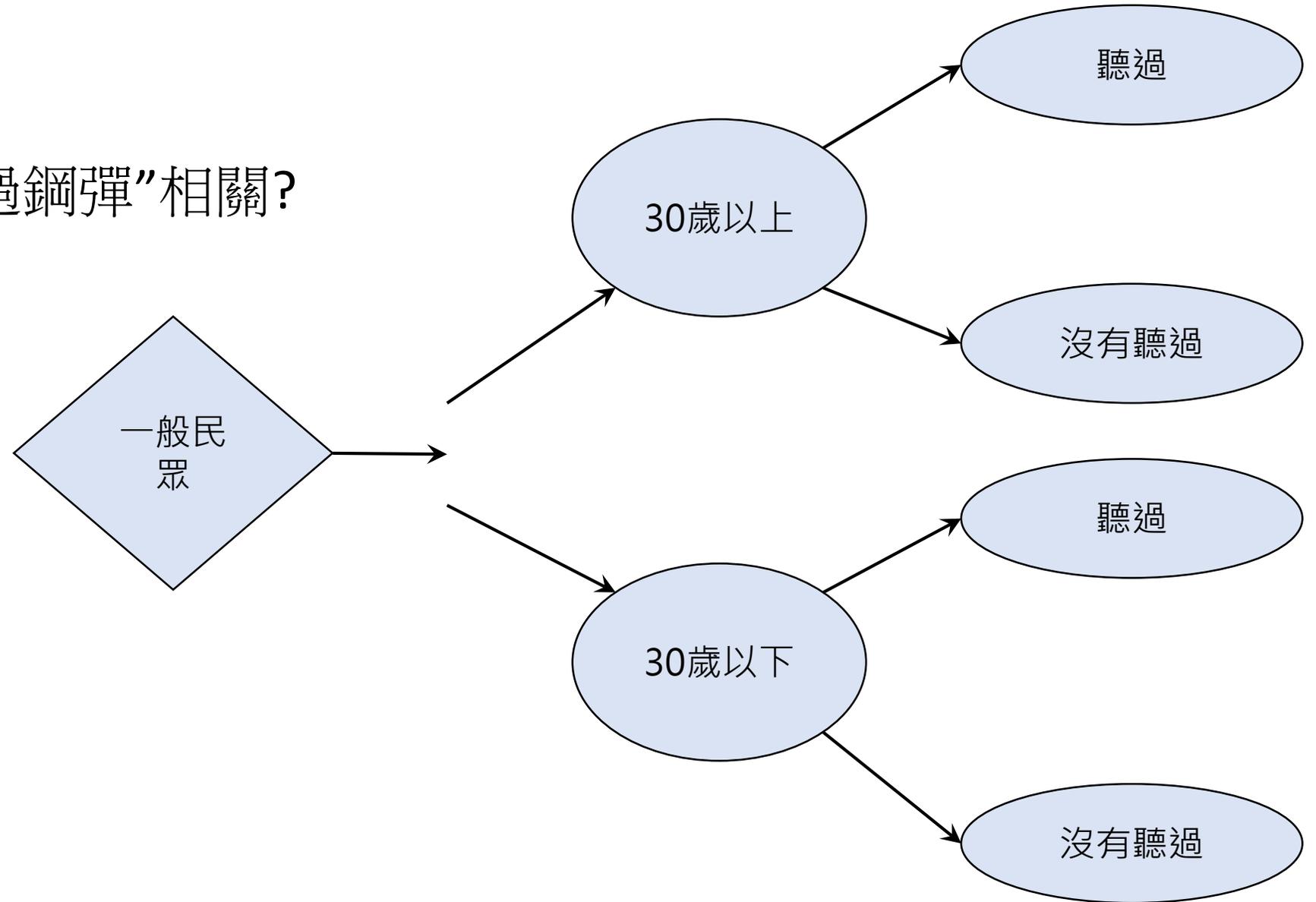
"年齡"是否與"喜歡鋼彈"相關?
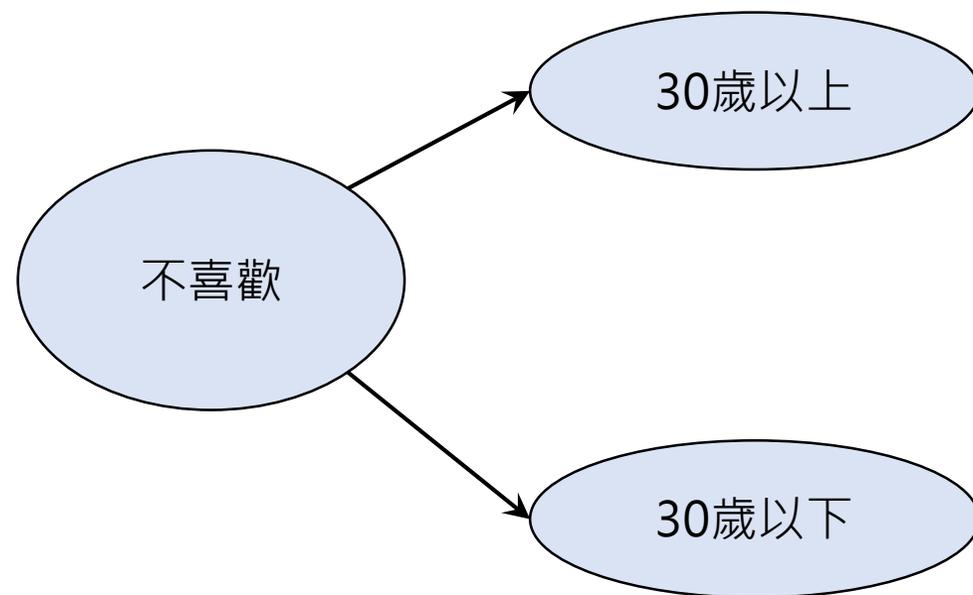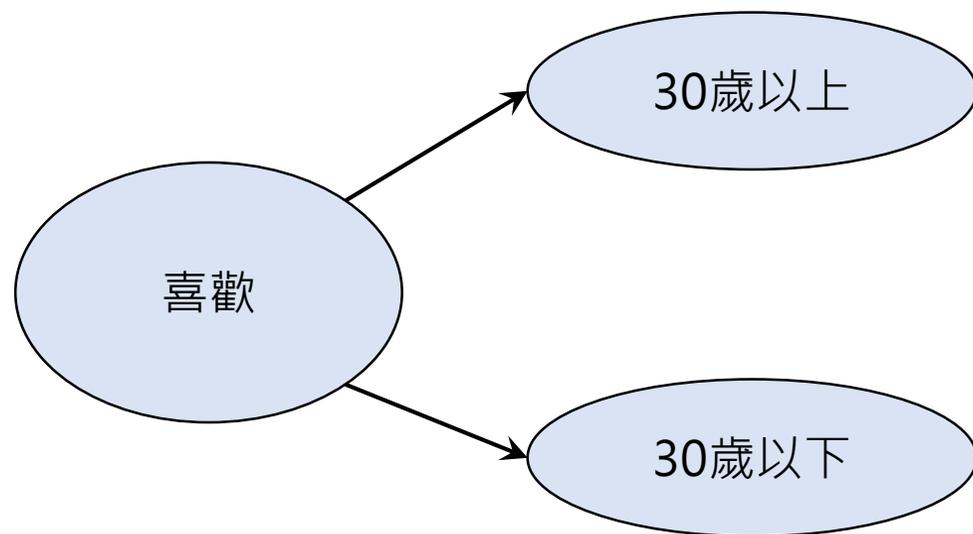
"性別"是否與"膽結石"相關?

"年齡"是否與"膽結石"相關?

"肥胖"是否與"膽結石"相關?

Multivariable thinking 多項思考

Ref: *Developing Multivariable Thinking,* Roxy Peck
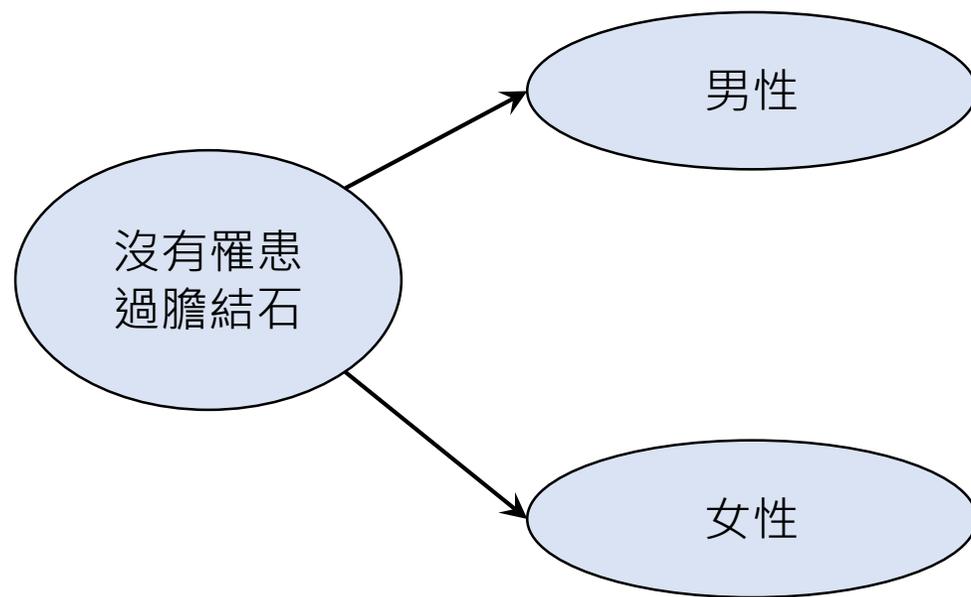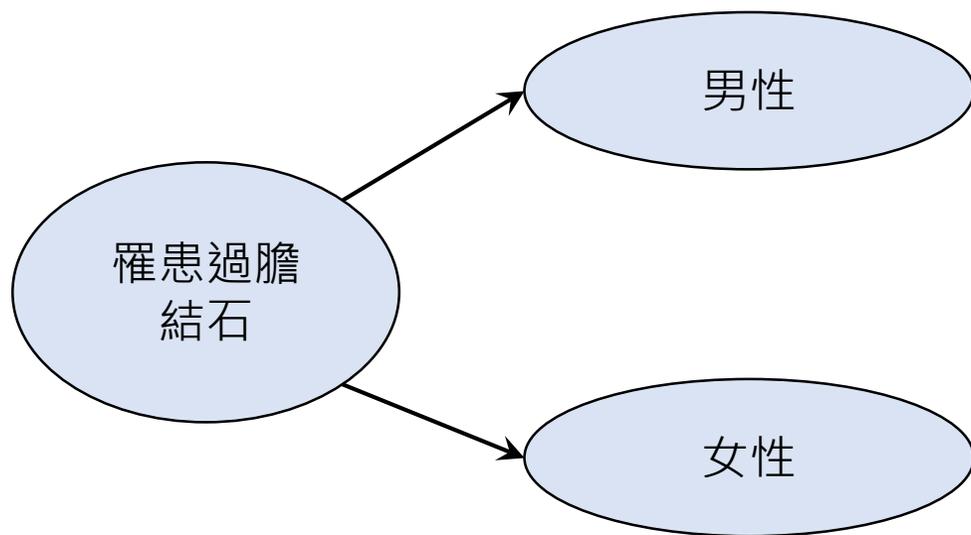
Does Taking Your Time in College Pay Off?
-- 2016 AP Statistics Exam

"A newspaper in Germany reported **that the more semesters need to complete an academic program at the university, the greater the starting salary in the first year of a job**. The report was based on a study that used a random sample of 24 people who had recently completed an academic program. Information was collected on the number of semesters each person in the sample needed to complete the program and the starting salary, in thousands of euros, for the first year of a job."
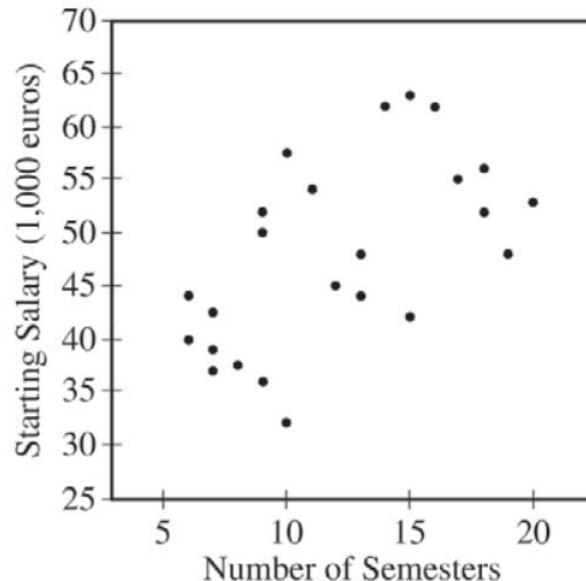
a. Does the scatterplot **support** the newspaper report about number of semesters and starting salary?

"An independent researcher received the data from the newspaper and conducted a new analysis by separating the data into three groups based on the major of each person. A revised scatterplot identifying the major of each person is shown below."

Major
∘ Business
♦ Physics
× Chemistry
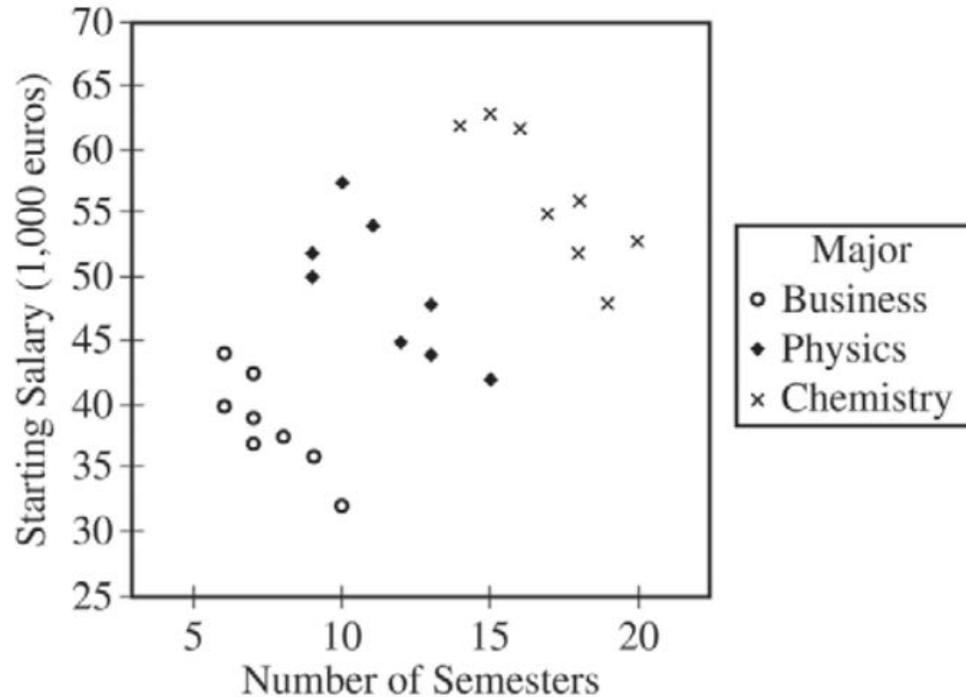
c. Based on the people in the sample, **describe the relationship between starting salary and number of semesters for the business majors**.

e. Based on the analysis conducted by the independent researcher, how could the newspaper report be **modified to give a better description** of the relationship between the number of semesters and the starting salary for the people in the sample?

Multivariable thinking

Confounding Variables

Are the variables of a child's shoe size and their reading ability correlated?

Shoe size	Reading ability
4	4
5	5
6	7
4	5
5	4
6	5
4	4
5	4
6	6
7	8
8	9
9	9
7	8
8	7
9	7
2	1
3	2
4	4
2	1
3	2
4	4

R = 0.93
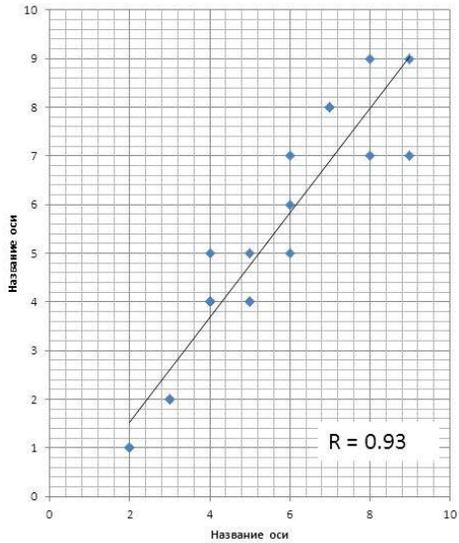
Crime rate and ice cream sales
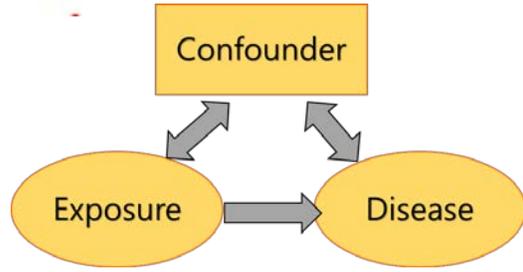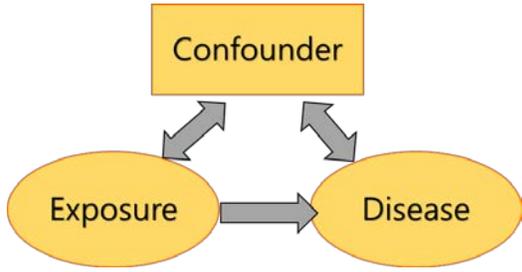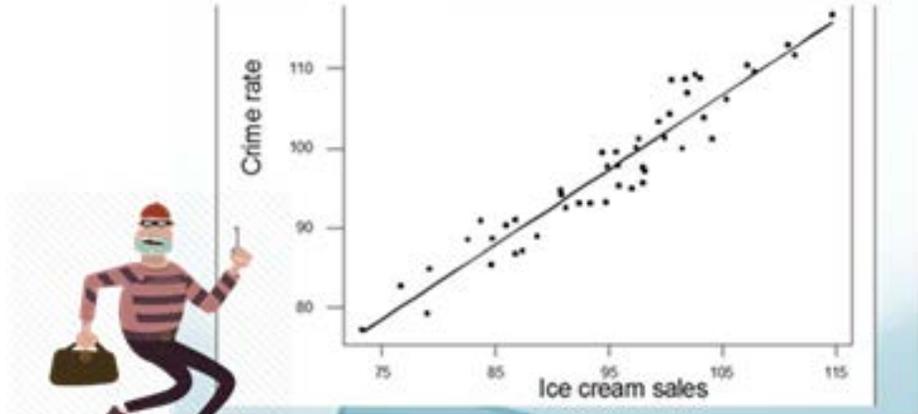
Confounding Variables
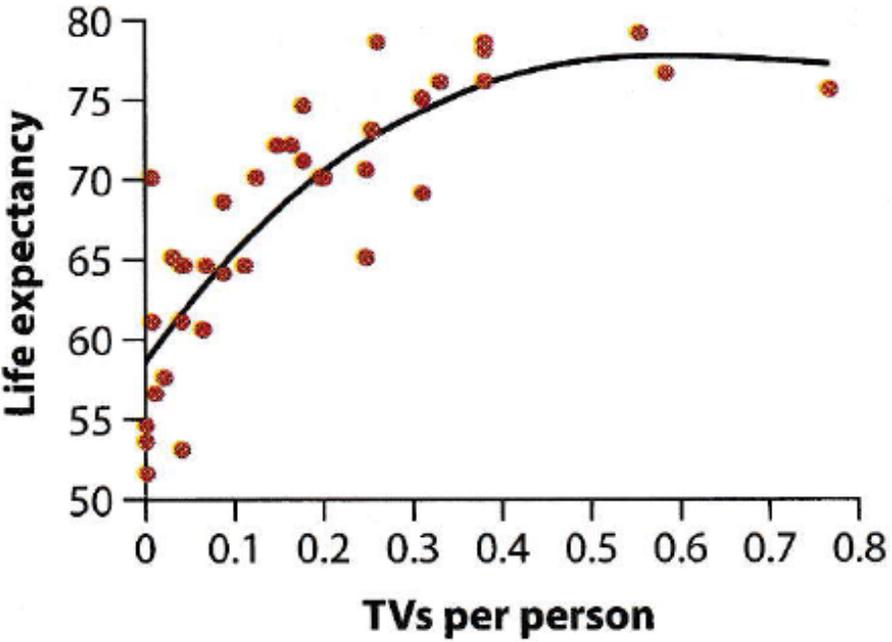
There is a **strong positive relationship** between **number of televisions per household** and **life expectancy** for countries in the world.
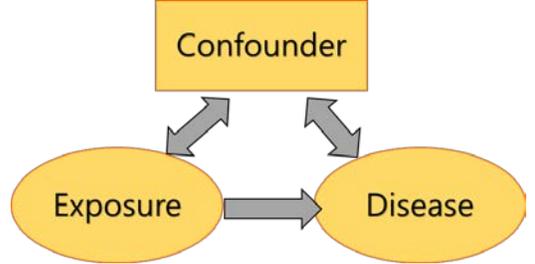
新研究工具

2016年韓國棋王李世乭代表人類出戰Google DeepMind開發的AlphaGo

【纪录片】AlphaGo世紀對決 2017_剪接

Artificial intelligence is the field of computer science that is associated with the concept of machines "**think like humans**" to perform tasks such as **learning, problem-solving, planning, reasoning and identifying patterns.**

1950
圖靈測試 (Turing test)
判斷機器是否能夠思考的實驗
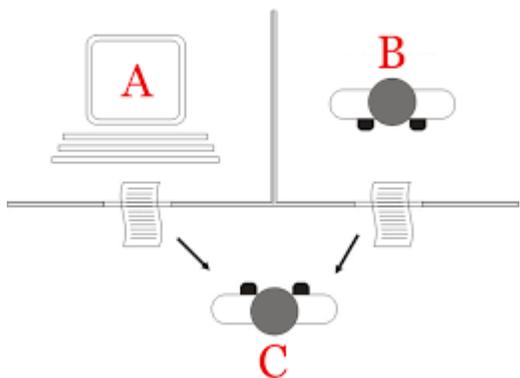
Timeline of artificial intelligence

ImageNet Error Rate

Hot terms

| Antiquity | 1950 | 1956 | 1980 | 1985 | 2010 | 2015 | 2016 | 2017 |

In Greek mythology, Talos was a giant automaton made of bronze to protect Europa

Asimov introduced the **Three Laws of Robotics**

Arthur Samuel (IBM) wrote the **first game-playing program** for checkers

John McCarthy coins "**Artificial Intelligence**" for the first time

Generation of entrepreneurs believed AI was the **wave of the future**

Computer Scientists pioneer **highly effective ML method**

'AI Winter'

ImageNet Challenge

Google DeepMind's **AlphaGo** defeats European **Go champion**

Microsoft's AI can now understand speech **better** than humans

AI 'everywhere' and accelerating
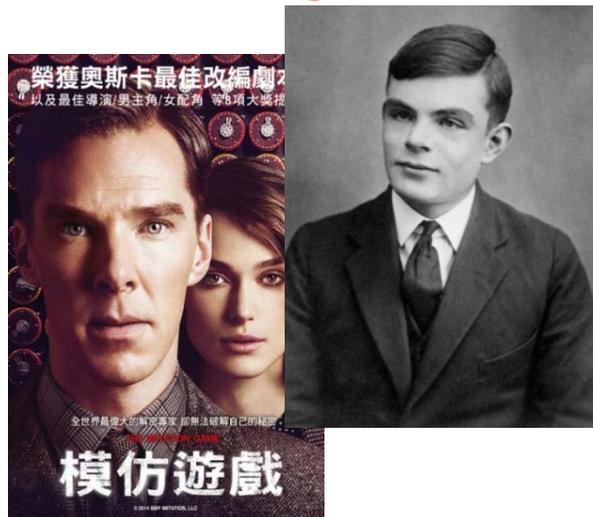
2015 AlphaGo 5:0 歐洲圍棋冠軍樊麾 二段
2016 AlphaGo 4:1 世界冠軍韓國職業棋士李世乭 九段

vmworld

榮獲奧斯卡最佳改編劇本
以及最佳導演/男主角/女配角 郭8項大獎提名

模仿遊戲

人工智慧:
計算機模仿人類思考進而模擬人類的能力/行為。

傳統規則式專家系統 (rule-based expert system)
輸入了人類的200多萬局棋譜
計算出每一步棋後面的12步變化
最強的人類只能計算10步

專家系統是一個知識庫(Knowledge-based)，設計
用來解決特定領域(Specific domain)的問題

Artifical Intelligence

Machine Learning

Deep Learning

Any technique that enables machines to mimic human behaviour

Subset of AI which uses statistical methods to learn without being explicitly programed

Subset of machine learning in which multilayered neural networks learn from large amounts of data

1960's   1980's   2000's

1950's   1970's   1990's   2010's

Early AI sparks excitement

Machine Learning starts to take off

Deep Learning drives AI boom

深藍(Deep Blue)
專門分析西洋棋

許峰雄
生於台灣基隆
1980年畢業於臺灣大學電機系
1989年獲得美國卡內基美隆大學電腦科學博士學位

分辨貓或是狗

專家系統:
狗是汪汪叫，貓是喵喵叫
狗比較大隻，貓比較小隻

鼻子形狀

腳印

機器學習(Machine Learning)

經過資料觀察，特徵萃取，模型建立的過程得到可以歸納/學習出有用規則的模型

資料結構化：

身高
體重
性別
毛髮顏色
花紋有無
叫聲
鼻子形狀
：

飼養者每月平均花費
每年上動物醫院院次數
每年上動物美容院次數
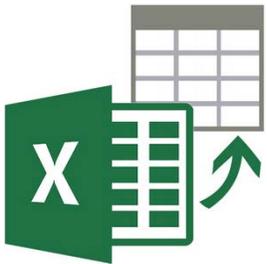
特徵提取
(feature extraction)：
選入幫助模型進行
決策(提高分類/預
測準確度)的變項

ID	H	W	label
001	8	8	DOG
002	40	50	DOG
:			
999	9	8	CAT
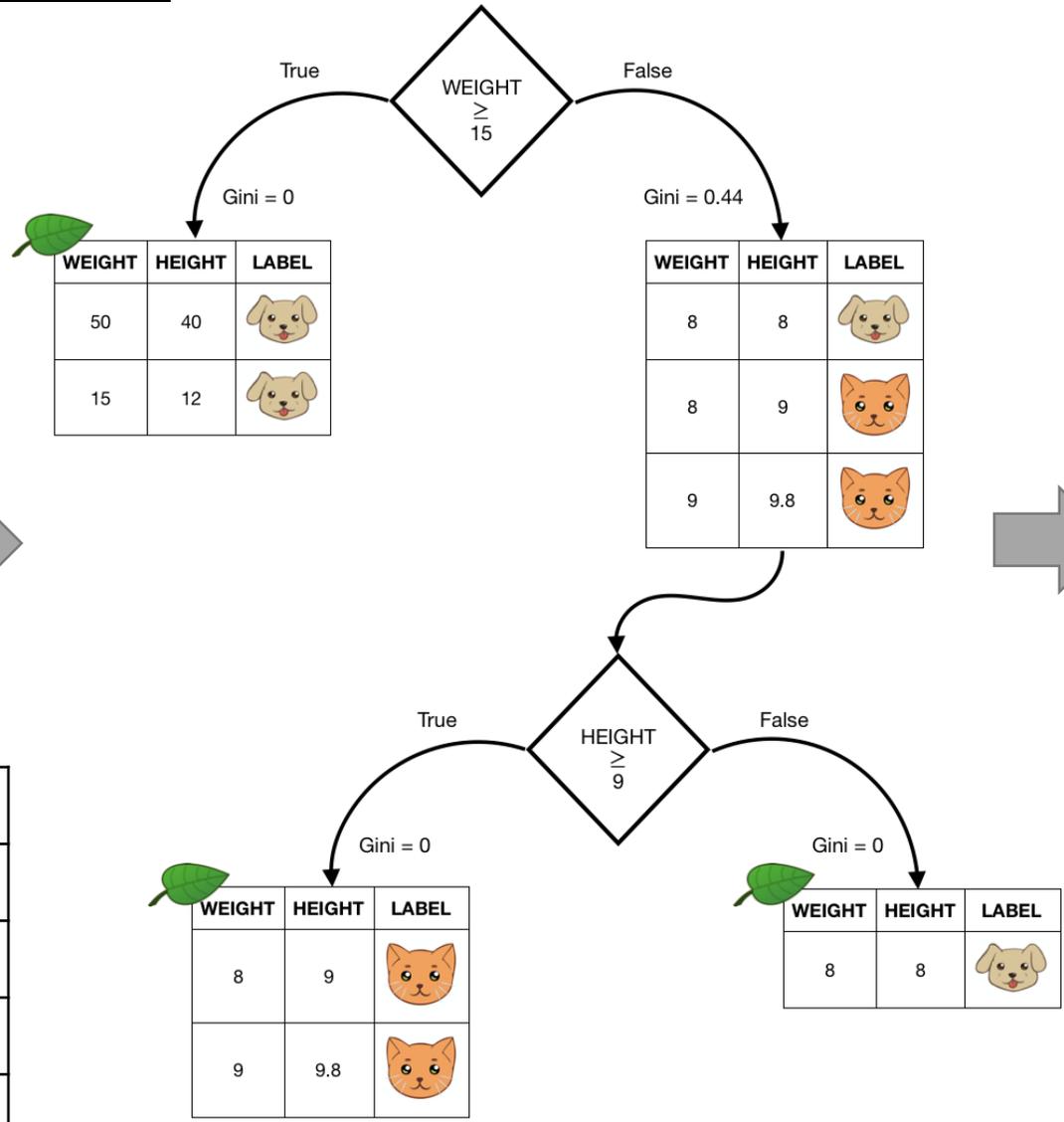
深度學習(Deep Learning)
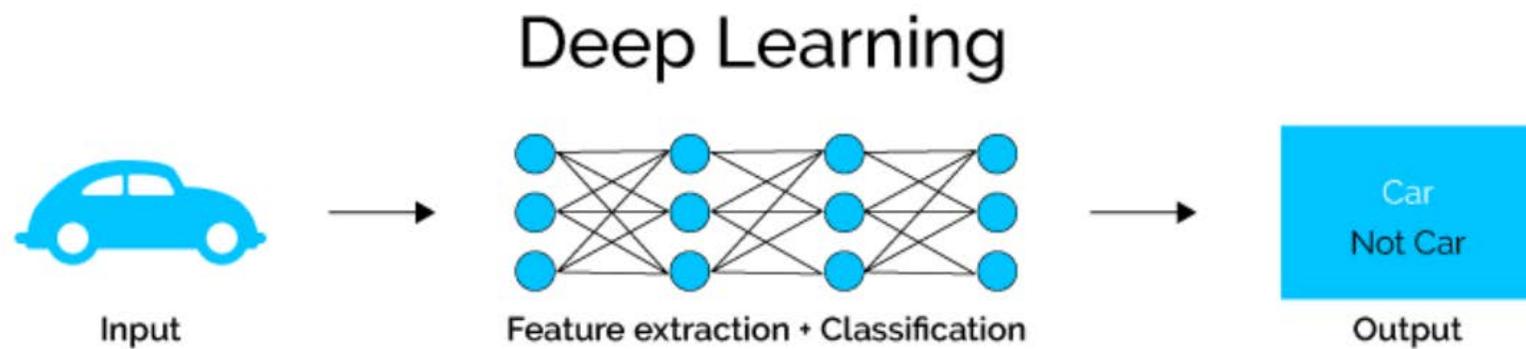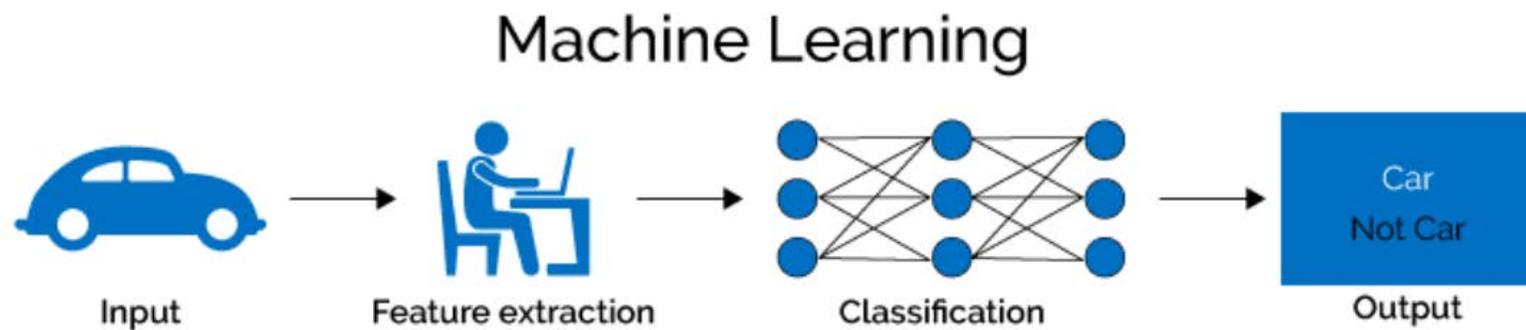
電腦看到什麼 - 電腦視覺（Computer vision）

1980S-ERA NEURAL NETWORK

1個像素 (pixel)

12 X 17 像素 (pixel)

224 X 300 像素 (pixel)

亞伯拉罕‧林肯，第十六任美國總統

1980S-ERA NEURAL NETWORK

Input layer
Hidden layer
Output layer
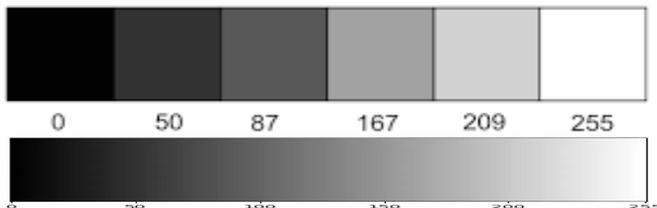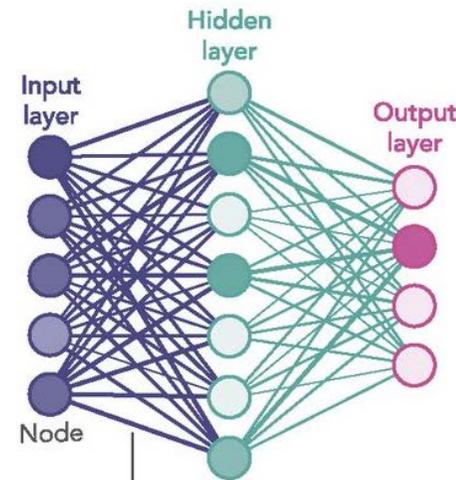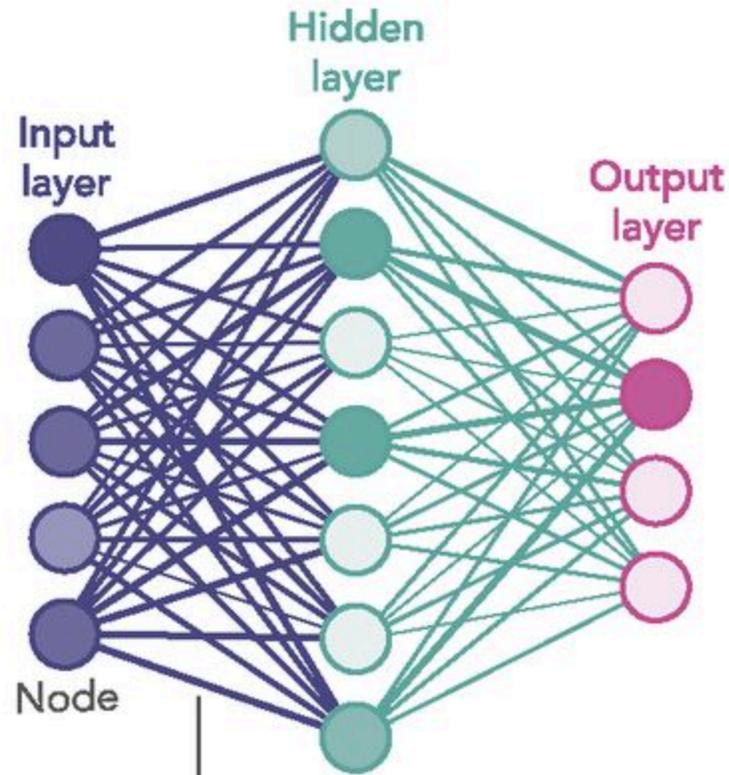
Node

Links carry signals from one node to another, boosting or damping them according to each link's 'weight'.

DEEP LEARNING NEURAL NETWORK

Multiple hidden layers process hierarchical features

Input layer

Input

Identify light/dark pixel value

Output layer

Output: 'George'

Identify combinations or features

Identify edges

Identify combinations of edges

Identify features
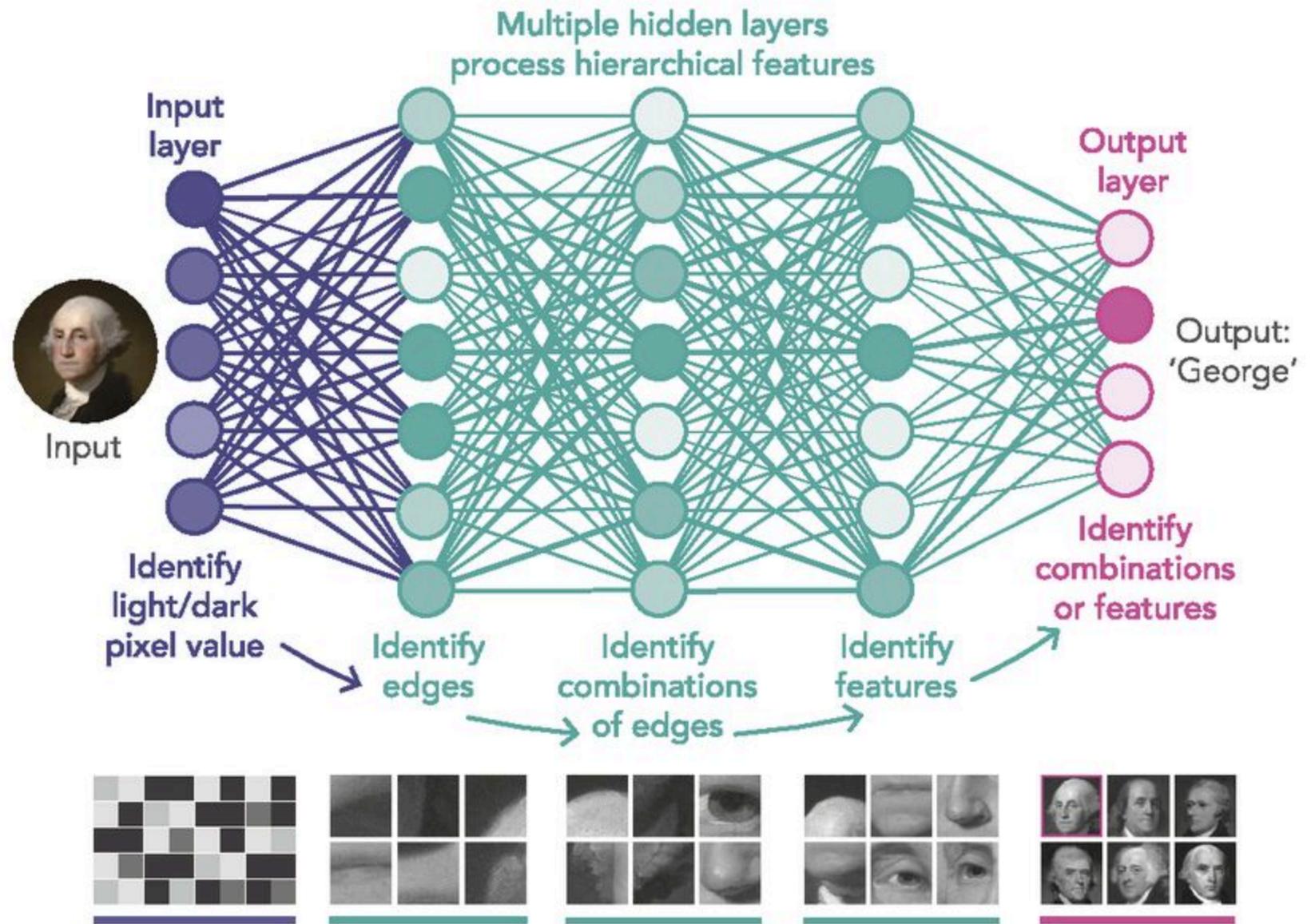
Image recognition: Pixel → edge → texton → motif → part →object

Airedale, Airedale terrier | miniature schnauzer | Boston bull, Boston terrier | standard poodle | English setter
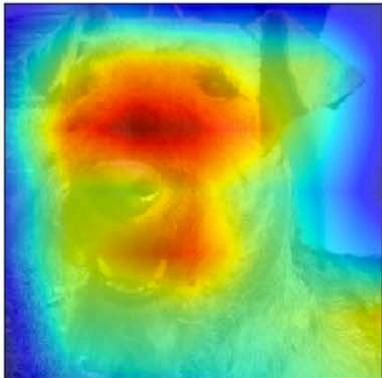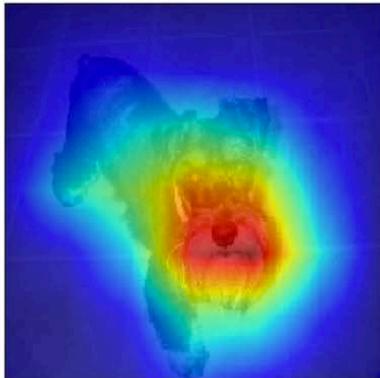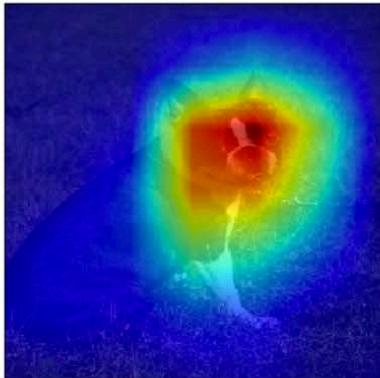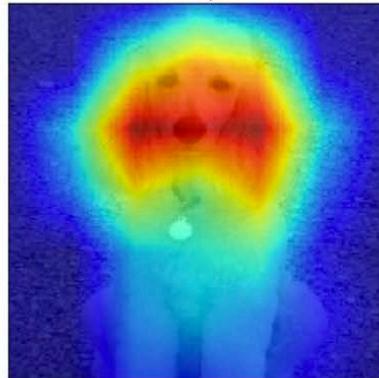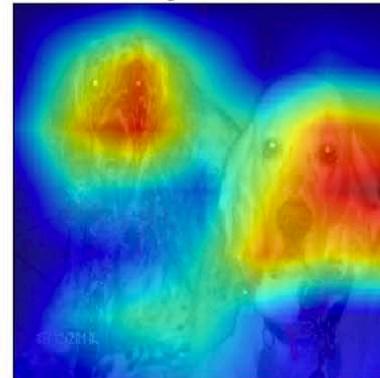
萬能㹴
(Airedale Terrier)　迷你雪納瑞
(Miniature Schnauzer)　波士頓㹴
(Boston Terrier)　標準貴賓犬
(Standard Poodle)　英國蹲獵犬
(English Setter)

Input

$16 \times 16 = 256$

Ink → 1
No ink → 0

Neural Network
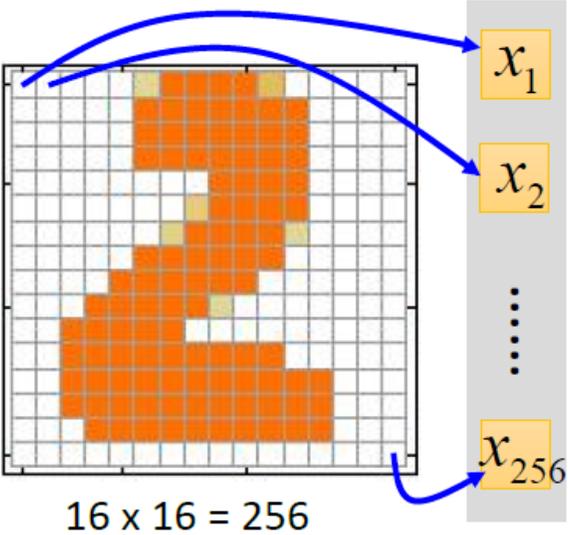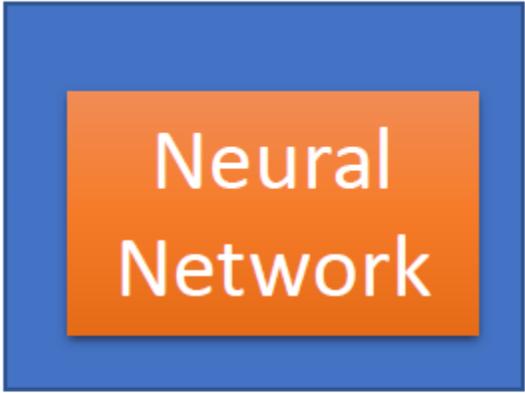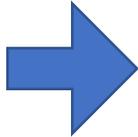
What is needed is a function ......

Output

0.1	is 1
0.7	is 2
0.2	is 0

The image is "2"
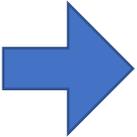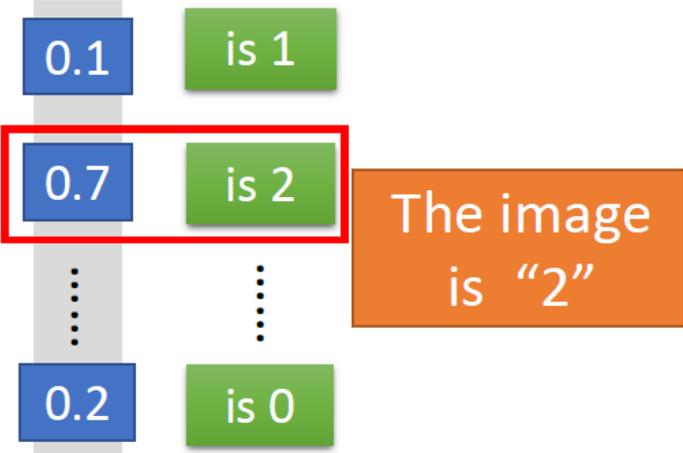
Each dimension represents the confidence of a digit.

李宏毅 Hung-Yi Lee

深度學習 學到什麼?

Deep learning can be highly flexible

- Speech Recognition

$$f*\left(\quad \right) = \text{"Morning"}$$

- Handwritten Recognition

$$f*\left(\quad \right) = \text{"2"}$$

- Playing Go

$$f*\left(\quad \right) = \text{"5-5"}$$
(step)

- Dialogue System

$$f*\left(\quad \text{"Hi"} \quad \right) = \quad \text{"Hello"}$$

(what the user said)　　(system response)

Accuracy

Deep Learning

Traditional Machine Learning

Data

Note: graph is representational only and does not depict actual data

李宏毅 Hung-Yi Lee
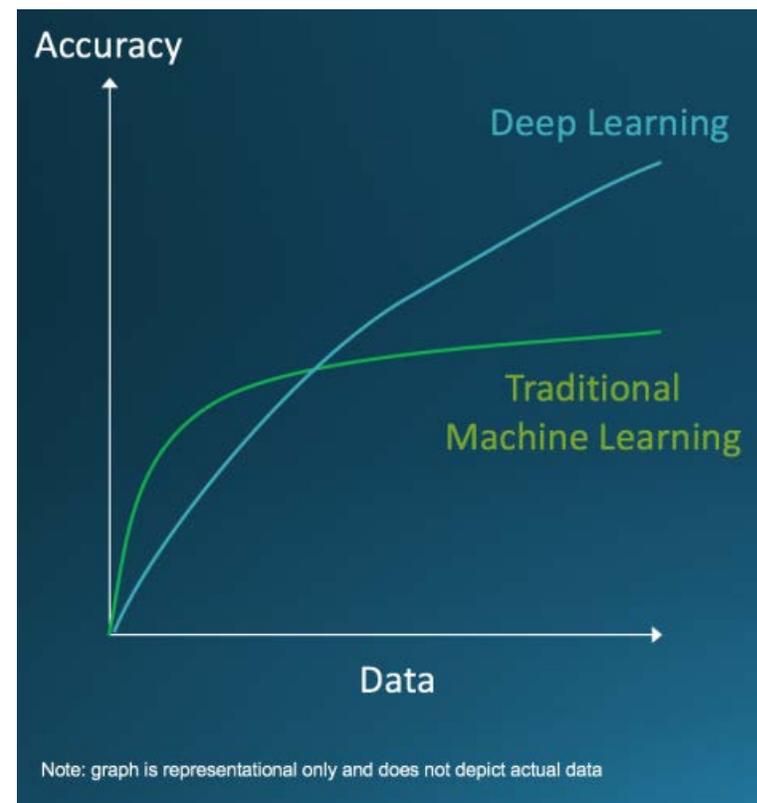
吉娃娃 OR 瑪芬

Dog or Mop?

Pug or Croissant?

New tool + Big data → New finding

馮已榕

長庚大學 醫學生物技術暨檢驗學系
凱斯西儲大學 (Case Western Reserve University)
生物統計碩士/博士

搖滾名人堂

Cuyahoga Valley National Park
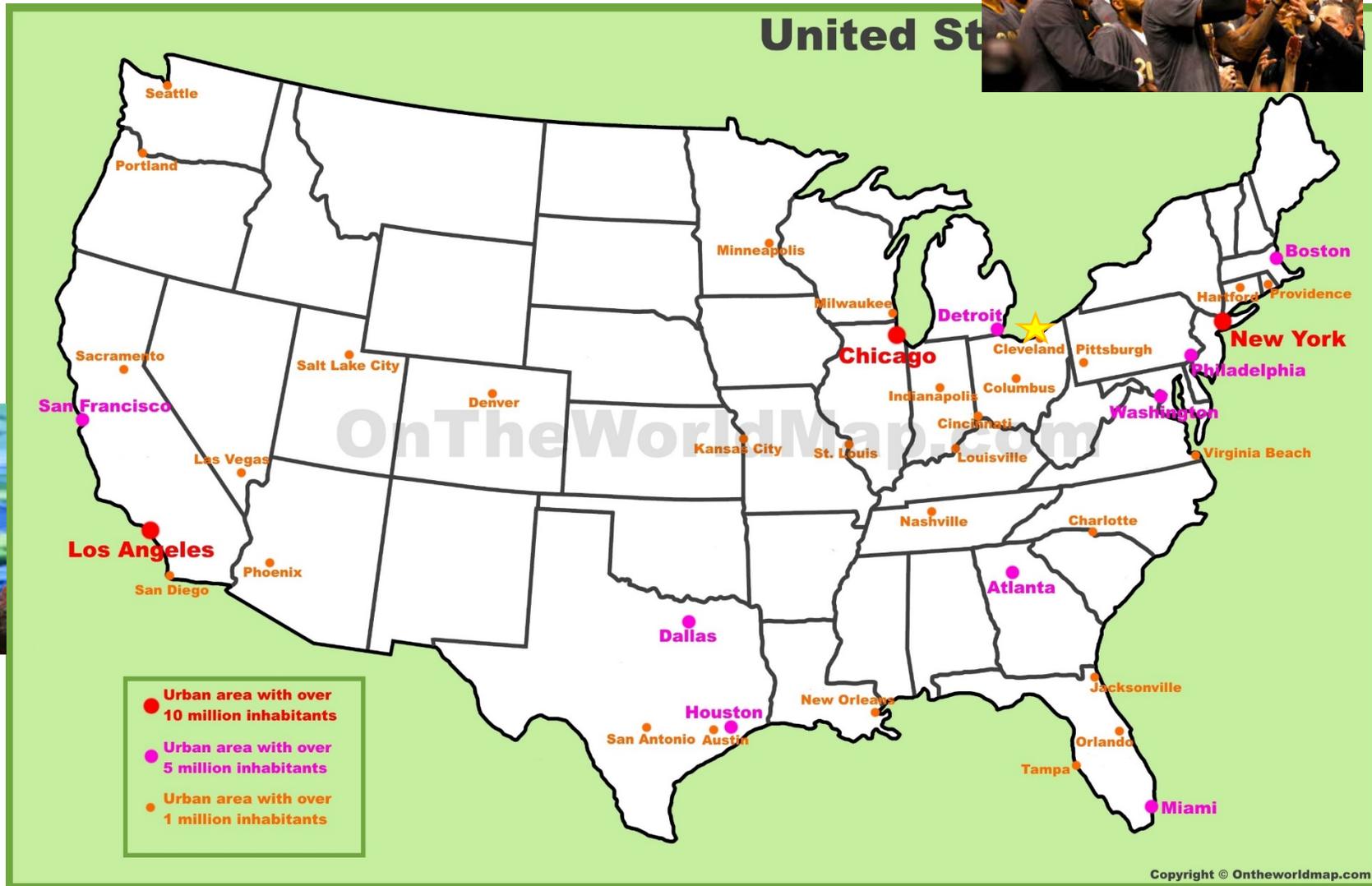
America's favorite pies

南瓜派

胡桃派

檸檬派 (ley lime pie)

蘋果派

櫻桃派
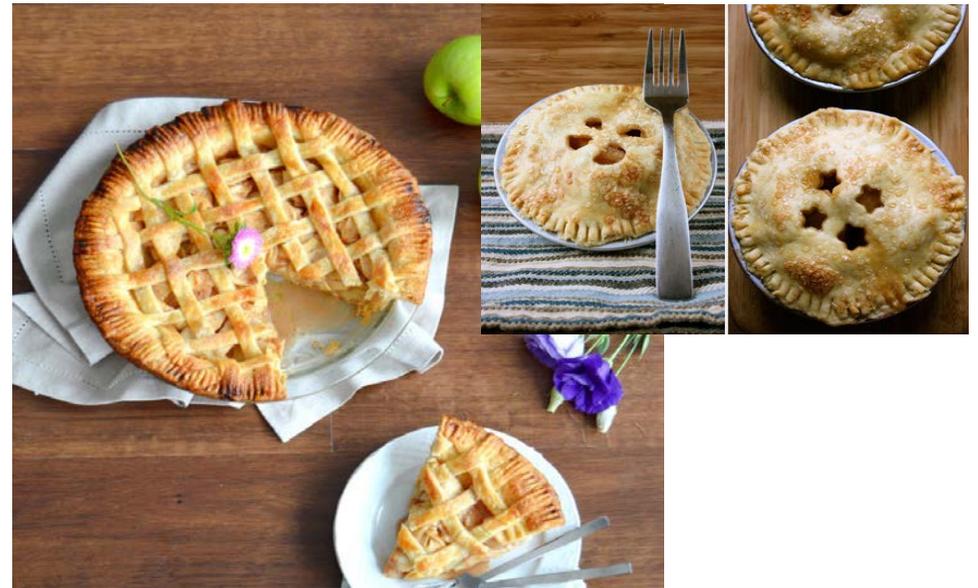
More data → New finding

定位紀錄, 心跳, 心電圖

汽車防盜系統

斐斯托斯圓盤
Phaistos Disc

2.54 cm

5D optical data storage
360TB, 189 °C 保存138億年

More data → better finding ?

王聖捷
Tricia Wang

- 知道智慧型手機的族群
- 不知道智慧型手機的族群
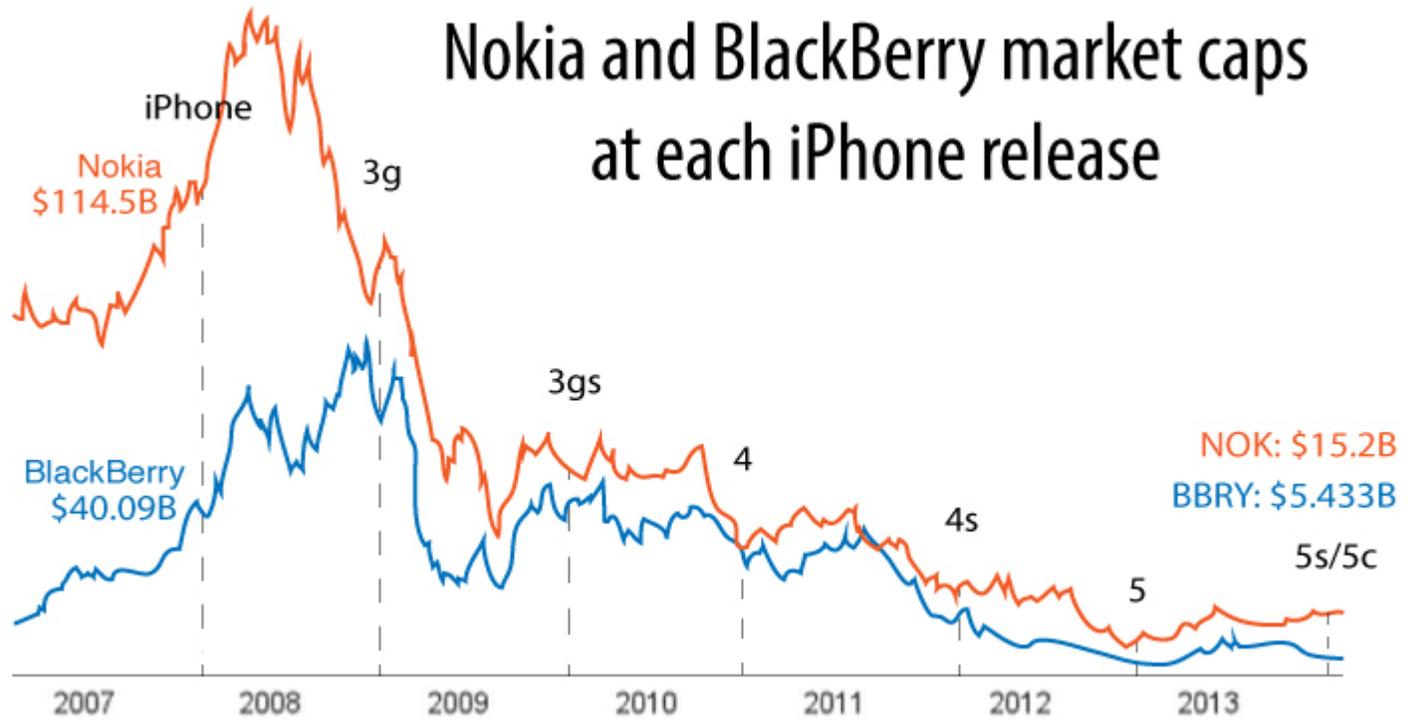
貴，一摔就壞，重

Know your data
know your question

Percent of **nutritionists** saying a food is healthy

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%

Olive oil
Kale
Almonds
Orange
鷹嘴豆泥
Eggs
Spinach
Hummus
藜麥、紅藜
Apple
Avocados
Quinoa
Oatmeal
Carrots
Cashews
Turkey   Chicken
Cottage cheese
Shrimp
Whole wheat bread
Tofu
Canned tuna
Skim. milk
Peanut butter
Sushi
脫脂牛奶
Corn
Wine
Baked potatoes
Popcorn
Whole milk
Orange juice
Steak
Pork chops
Granola
Cheddar cheese
Coconut oil
Butter
Frozen yogurt
蛋白棒
Hamburgers
Kind bar
Pizza
Crackers
Beer
燕麥棒
Granola bars
Beef jerky
Diet soda
White bread
American cheese
SlimFast Shake 代餐奶昔
SlimFast shake
Chocolate chip cookies
Bacon
Ice cream
French fries
Regular soda

Percent of **all Americans** saying a food is healthy →

20%   30%   40%   50%   60%   70%   80%   90%   100%

Ref: NYTimes.com
https://www.nytimes.com/interactive/2016/07/05/upshot/is-sushi-healthy-what-about-granola-where-americans-and-nutritionists-disagree.html#:~:text=No%20food%20elicited%20a%20greater,30%20percent%20of%20nutritionists%20did.

Percent of nutritionists saying a food is healthy

鷹嘴豆泥

藜麥、紅藜

Quinoa, lauded as a

Ordinary Americans were uncertain.

脫脂牛奶

Both nutritionists and the public were split about the healthiness of many common items, including butter, red meat, whole milk and cheddar cheese.

Foods along this line were rated equally by nutritionists and the public.

蛋白棒

No food elicited a greater di...
be...
pu...
About 70 percent of Americans called it healthy, but less than 30 percent of nutritionists did.

燕麥棒

SlimFast Shake 代餐奶昔

Percent of **all Americans** saying a food is healthy

Variables:
- Percentage o f **nutritionists** saying a food is healthy.
- Percentage o f **all Americans** saying a food is healthy.
- Food

Ref: NYTimes.com

- What do you **notice**?
- What do you **wonder**?
- What kind of **headline** would you write for this graph?
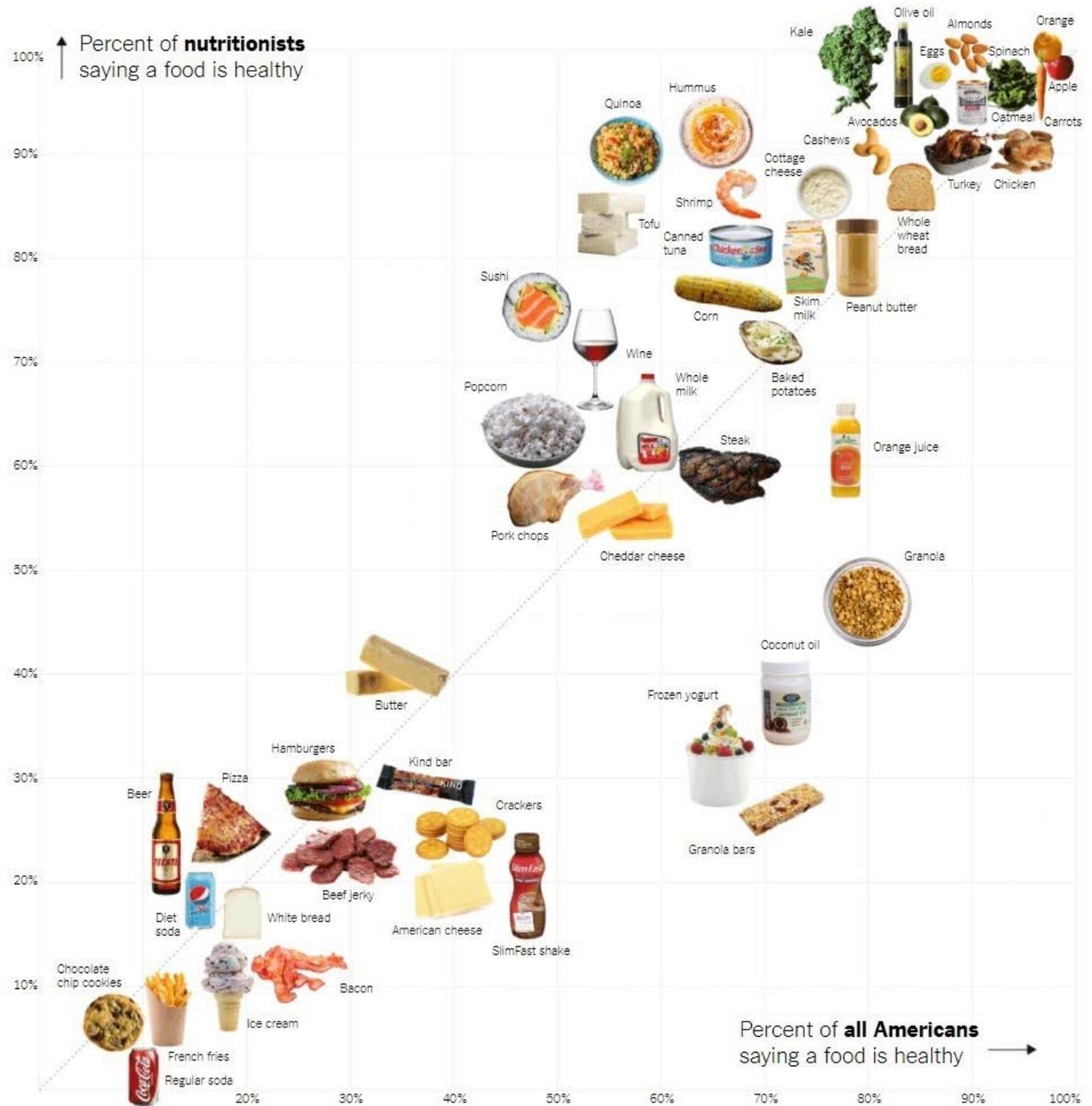
Ref: NYTimes.com

https://www.nytimes.com/interactive/2016/07/05/upshot/is-sushi-healthy-what-about-granola-where-americans-and-nutritionists-disagree.html#:~:text=No%20food%20elicited%20a%20greater,30%20percent%20of%20nutritionists%20did.
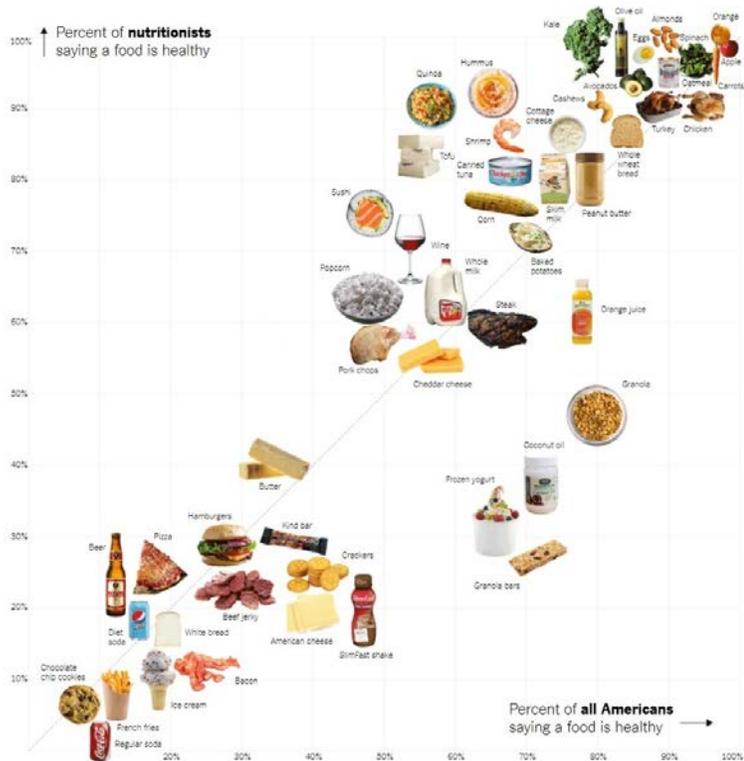
Foods considered healthier by the public than by experts

Percent describing a food as "healthy"	Nutritionists	Public	Difference
Granola bar	28%	71%	43
Coconut oil	37%	72%	35
Frozen yogurt	32%	66%	34
Granola	47%	80%	33
SlimFast shake	21%	47%	26
Orange juice	62%	78%	16
American cheese	24%	39%	15

SlimFast Shake 代餐奶昔: 4克脂肪，1克飽和脂肪，30克碳水化合物，4克纖維，22克糖，10克蛋白質

I Can't Believe It's Yogurt!

整理資料

2019 建構精準
健康照護體系高峰會議
Taiwan Precision Health System Summit

衛生福利部
Ministry of Health and Welfare

Drive progress
Precision Medicine 2030

Genome Sequencing and Big Data in Health Care
Health Care Database and Data Standardization
Health Industry Development

Garbage In Garbage Out

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

John P. A. Ioannidis, PLoS Med. **2005** Aug

醫學假設 ✔ 實驗證實 ❓ 現實生活

BAYER ❓

DATA → Model → Result

就醫流程

A Patient Approach to the hospital

B Reach to front office for Registration 掛號

C Request sent to Medical Records Department for record retrieval

調閱病歷

D Doctors Room where patient gets the consultation done

E Doctors go through the patient previous records and write new notes 看診

Simultaneously patients wait until his/her records reached to the OPD 候診

藥局領藥

F If required patient go for Investigation

G Reports of the investigation reached to the doctor

檢查(放射,抽血檢驗)

H2 After getting care and paying all the dues

Either the doctor advise for drugs

繳錢

H1 Or asked the patient to get admitted into the hospital

手術　住院

After getting care and paying all the dues

I Patient leaves the hospital 離開

http://remomd. disrup ealthcare-inf tion-technology-in-a-limited-resource-community.html

入院護理紀錄

病歷號	年齡	性別	婚姻狀況	教育程度
1234567	60	男	鰥寡	小學
1234578	64	男	新婚	10年
2546871	99	女	已婚	大專以上

入院醫生診斷

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	診斷碼1	診斷碼2	診斷碼3	診斷碼4
1	1234567	2020/06/12	1960/05/01	高血壓	腦血管疾病	心臟疾病	子宮頸惡性腫瘤
2	1234578	2020/06/31	1955/11/24	高血壓	糖尿病	氣喘	
3	2546871	2020/07/01	2008/02/09	糖尿病	心臟疾病		

入院醫生處置

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	藥碼1	藥碼2	藥碼3	藥碼4
1	1234567	2020/06/12	1960/05/01	B01	B05	B05	A03
2	1234578	2020/06/31	1955/11/24	A03	B05	A04	M04
3	2546871	2020/07/01	2008/02/09	A03	M04	A04	R06

各檔案間串檔變項說明
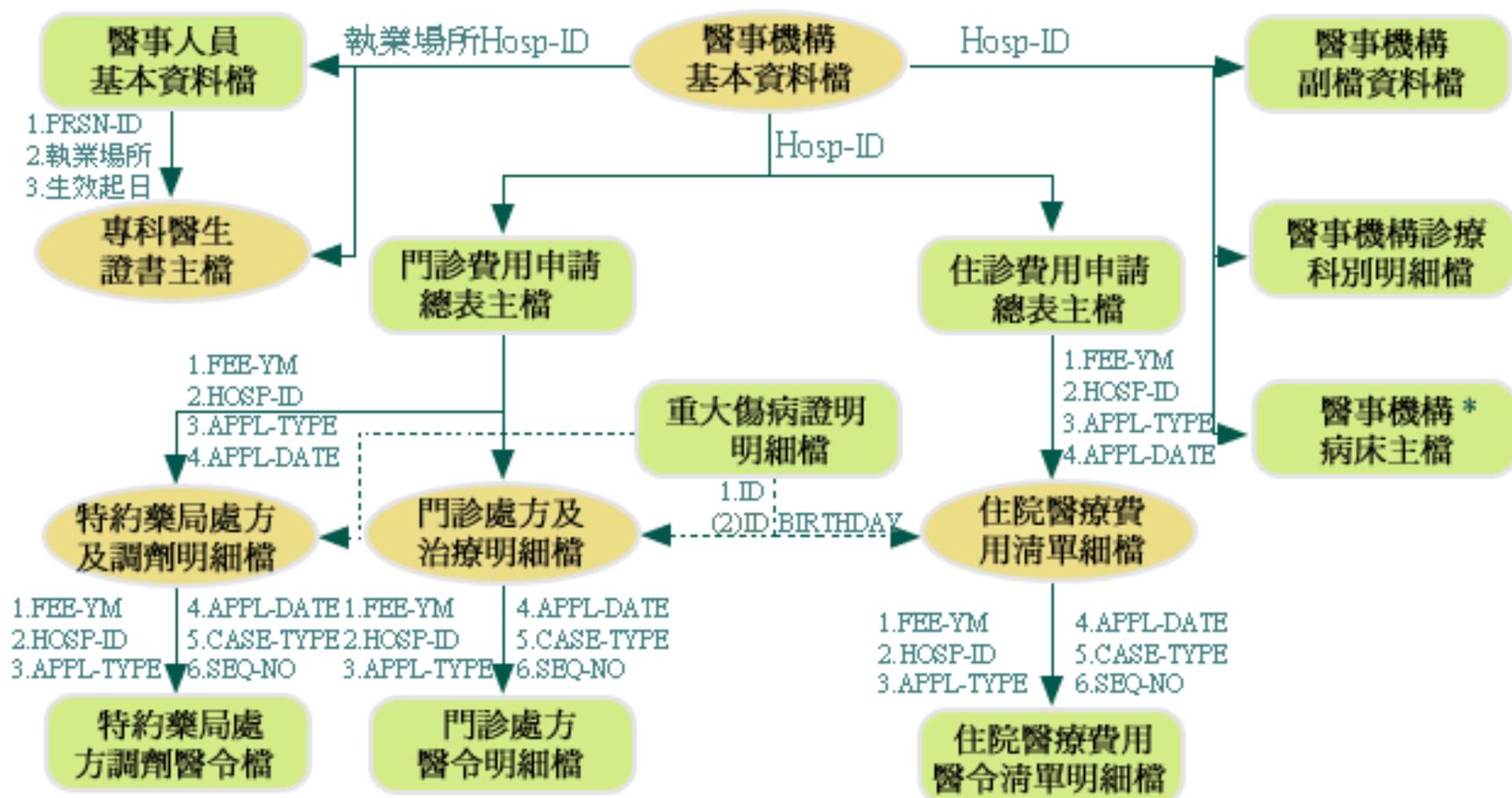
變項類型

- **連續[量] : 小數點有意義的數字**

Ex.,血清總膽固醇濃度(mg / dl), 溫度(攝氏), BMI(kg/m^2)...

- **序數: 有順序上的關係, 但沒有數學上倍數的關係**

Ex.,[冠軍, 亞軍, 季軍] [低年級, 中年級, 高年級] 疼痛指數
滿意度[很滿意, 滿意, 尚可, 不滿意, 很不滿意]

- **類別[質]: 沒有大小關係**

二項: 只有兩個類別
Ex., [死亡/存活] [男/女][有/沒有 糖尿病]
　　　　[有/沒有 抽菸]

非二項: 兩個以上類別
Ex., 血型[A型, B型, AB型, O型]
　　　種族[高加索人, 非洲人, 亞洲人]

- **時間(天) –** 開始治療後到死亡, 診斷後到復發 ....

資料整理

資料整理

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	年齡	性別	婚姻狀況	教育程度	診斷碼1	診斷碼2	診斷碼3	診斷碼4
1	1234567	2020/06/12	1960/05/01	60	男	鰥寡	小學	高血壓	腦血管疾病	心臟疾病	子宮頸惡性腫瘤
2	1234578	2020/06/31	1955/11/24	64	男	新婚	10年	高血壓	糖尿病	氣喘	
3	2546871	2020/07/01	2008/02/09	99	女	已婚	大專以上	糖尿病	心臟疾病		

1. 數值是否符合一般(醫學)認知?

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	年齡	性別	婚姻狀況	教育程度	診斷碼1	診斷碼2	診斷碼3	診斷碼4
1	1234567	2020/06/12	1960/05/01	60	男	鰥寡	小學	高血壓	腦血管疾病	心臟疾病	子宮頸惡性腫瘤
2	1234578	2020/06/31	1955/11/24	64	男	新婚	10年	高血壓	糖尿病	氣喘	
3	2546871	2020/07/01	2008/02/09	32	女	已婚	大專以上	糖尿病	心臟疾病		

12?

未婚	小學
已婚	國中
鰥寡	高中
離婚	大專以上
其他	其他

資料整理

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	年齡	性別	婚姻狀況	教育程度	診斷碼1	診斷碼2	診斷碼3	診斷碼4
1	1234567	2020/06/12	1960/05/01	60	男	鰥寡	小學	高血壓	腦血管疾病	心臟疾病	子宮頸惡性腫瘤
2	1234578	2020/06/31	1955/11/24	64	男	新婚	10年	高血壓	糖尿病	氣喘	
3	2546871	2020/07/01	2008/02/09	32	女	已婚	大專以上	糖尿病	心臟疾病		

Raw data

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	年齡	性別	婚姻狀況	教育程度	診斷碼1	診斷碼2	診斷碼3	診斷碼4
1	1234567	2020/06/12	1960/05/01	60	男	鰥寡	小學	高血壓	腦血管疾病	心臟疾病	子宮頸惡性腫瘤
2	1234578	2020/06/31	1955/11/24	64	男	已婚	國中	高血壓	糖尿病	氣喘	
3	2546871	2020/07/01	1988/02/09	32	女	已婚	大專以上	糖尿病	心臟疾病		

Step 1 除錯

資料整理

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	年齡	性別	婚姻狀況	教育程度	診斷碼1	診斷碼2	診斷碼3	診斷碼4
1	1234567	2020/06/12	1960/05/01	60	男	鰥寡	小學	高血壓	腦血管疾病	心臟疾病	
2	1234578	2020/06/31	1955/11/24	64	男	已婚	國中	高血壓	糖尿病	氣喘	
3	2546871	2020/07/01	1988/02/09	32	女	已婚	大專以上	糖尿病	心臟疾病		

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	年齡	性別	婚姻狀況	教育程度	高血壓	腦血管疾病	糖尿病	心臟疾病	氣喘
1	1234567	2020/06/12	1960/05/01	60	男	鰥寡	小學	1	1	0	1	0
2	1234578	2020/06/31	1955/11/24	64	男	已婚	國中	1	0	1	0	1
3	2546871	2020/07/01	1988/02/09	32	女	已婚	大專以上	0	0	1	1	0

Step2：整理架構

資料整理

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	年齡	性別	婚姻狀況	教育程度	診斷碼1	診斷碼2	診斷碼3	診斷碼
1	1234567	2020/06/12	1960/05/01	60	男	鰥寡	小學	高血壓	腦血管疾病	心臟疾病	
2	1234578	2020/06/31	1955/11/24	64	男	已婚	國中	高血壓	糖尿病	氣喘	
3	2546871	2020/07/01	1988/02/09	32	女	已婚	大專以上	糖尿病	心臟疾病		

⬇

床號	病歷號	住院日 (yyyy/mm/dd)	出生年月日 (yyyy/mm/dd)	年齡	性別	婚姻狀況	教育程度	高血壓	腦血管疾病	糖尿病	心臟疾病	氣喘
1	1234567	2020/06/12	1960/05/01	60	1	3	1	1	1	0	1	0
2	1234578	2020/06/31	1955/11/24	64	1	2	2	1	0	1	0	1
3	2546871	2020/07/01	1988/02/09	32	2	2	4	0	0	1	1	0

Step3 : coding book

性別	婚姻狀況	教育程度	有無
1,男	1,未婚	1,小學	0,無
2,女	2,已婚	2,國中	1,有
	3,鰥寡	3,高中	
	4,離婚	4,大專以上	
	5,其他	5,其他	

資料結構邏輯
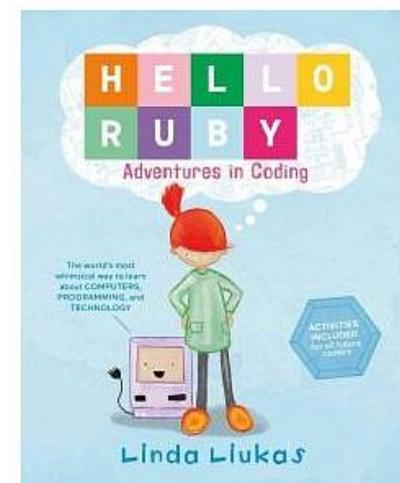
字串、數字、布林值(booleans)

字串：只要能被包括在引號裡的符號都算是字串, 他可能包含了字母、數字、空白, 或是其他電腦鍵打得出來的符號, 像是"Ruby".

數字：像是1、2、3, 或是 4.1217.

布林值(booleans)：他只有「真」(True) 或是「假」(False)兩種表達方式

I'm red and yellow.　　　True/False
I'm pink and green.　　　True/False
I'm happy.　　　　　　　True/False

My eyes are green.　　　True/False
I have six points.　　　　True/False
I am not yellow.　　　　True/False

I have legs.　　　　　　True/False
I have arms and legs.　　True/False
I have arms or legs.　　　True/False

資料結構 (data structure)

資料結構 (data structure) : 資料有許多型態. 像是字串、數字與布林值.當手上有非常多資料時, 把他們排列的有條有理, 會大大的提升使用效率. 我們將這個方法稱之為「資料結構」(data structure).

Green

veggie

Everything else

Pear

Broccoli

Carrot

Orange

Apple

Mushrooms

Zip

Fruit

Orange

Everything else

Pear

Broccoli

Carrot

Orange

Apple

Mushrooms
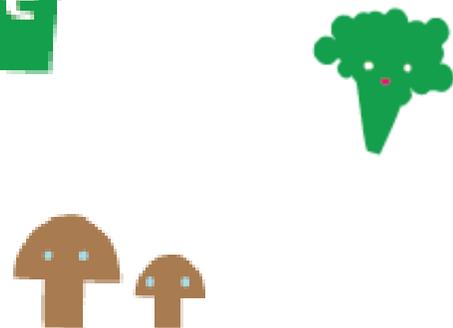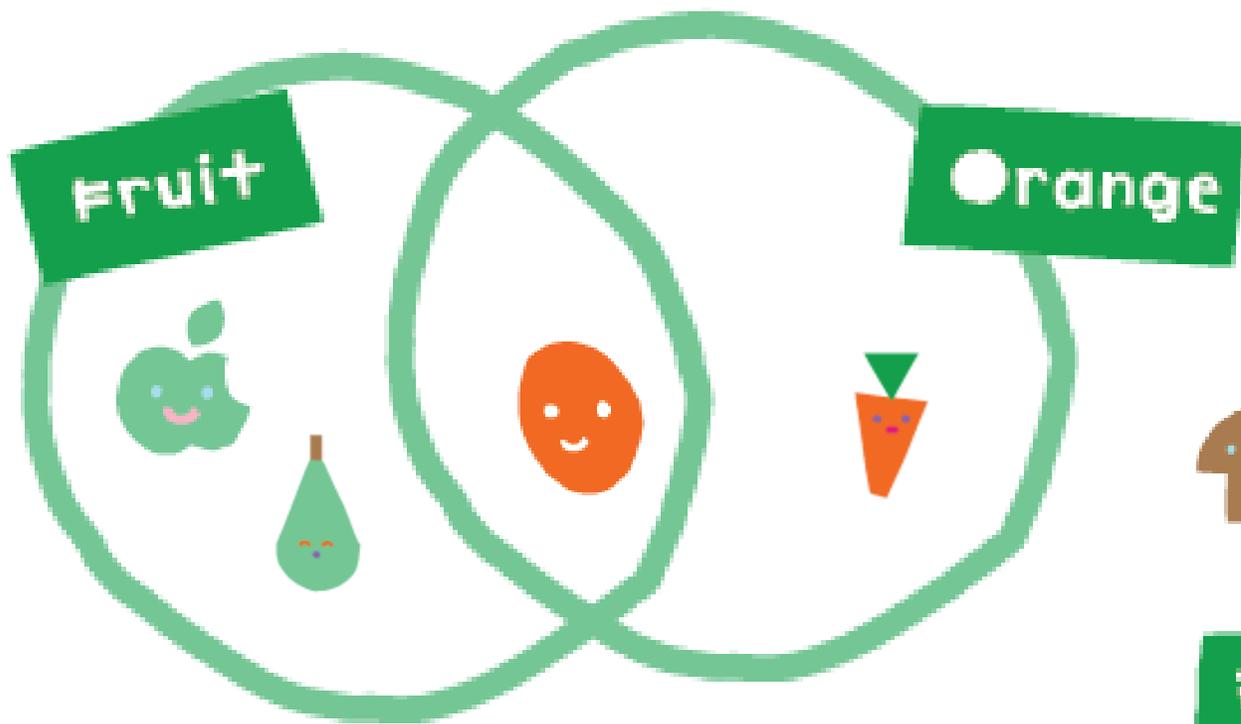
	水果	橘色
1		
2		
3		
4		
5		
6		

	水果	橘色
1	+	-
2	-	-
3	-	+
4	+	+
5	+	-
6	-	-

True / False

	水果	橘色	水果且橘色
1	+	-	
2	-	-	
3	-	+	
4	+	+	
5	+	-	
6	-	-	

True / False

	水果	橘色	水果且橘色
1	+	-	**False**
2	-	-	**False**
3	-	+	**False**
4	+	+	**True**
5	+	-	**False**
6	-	-	**False**

True / False

	水果	橘色	水果 或 橘色
1	+	-	
2	-	-	
3	-	+	
4	+	+	
5	+	-	
6	-	-	

True / False

		水果	橘色	水果 或 橘色
1		+	-	**True**
2		-	-	**False**
3		-	+	**True**
4		+	+	**True**
5		+	-	**True**
6		-	-	**False**

True / False

	水果	橘色	水果 且 不是橘色
1	+	-	
2	-	-	
3	-	+	
4	+	+	
5	+	-	
6	-	-	

True / False

		水果	橘色	水果 且 不是橘色
1		+	-	**True**
2		-	-	**False**
3		-	+	**False**
4		+	+	**False**
5		+	-	**True**
6		-	-	**False**

		水果	橘色	True / False 水果 且 橘色	True / False 水果 或 橘色	True / False 水果 且 不是橘色
1		+	-	False	True	True
2		-	-	False	False	False
3		-	+	False	True	False
4		+	+	True	True	False
5		+	-	False	True	True
6		-	-	False	False	False

迴圈 (loop)

This is how a fox would plant an entire row of carrots.

開始 → 種植 → 種植 → 種植 → 種植 → 種植 → 種植 → 結束

This is how a fox would plant an entire row of carrots.

開始 → 重複 [種植] → 結束
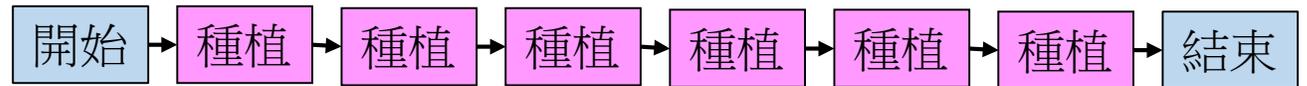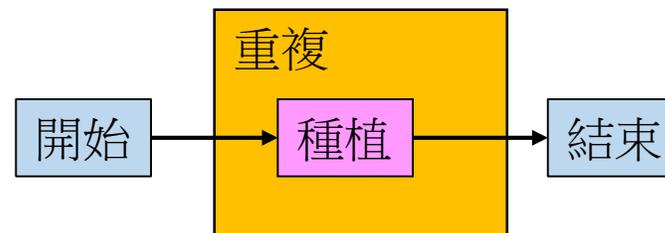
選擇

「如果...,就...」(if-then)

2

One piece of instruction has gone
missing. Can you figure out which one?

開始 → 種植 → 種植 → 種植 → 跳過 → 種植 → 種植 → 結束
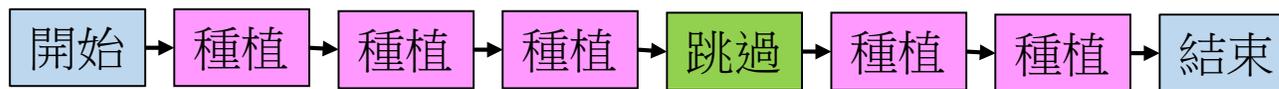
3

The foxes came up with a shorter way to write the
instructions. Can you help them fill in the code?

開始 → 重複 [種植 → 跳過] → 結束
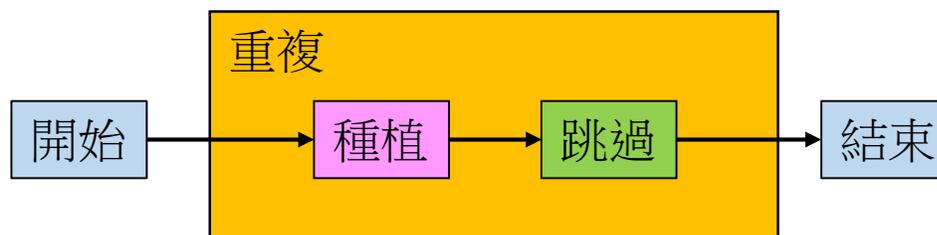
選擇

「如果...,就...」(if-then)

4 Many things are missing from this piece of instruction. Can you figure it out?

5 You know how to do this already!

選擇

「如果…,否則…」(if-else)

Oops, there's a bug in the row. What should you do to it?

7
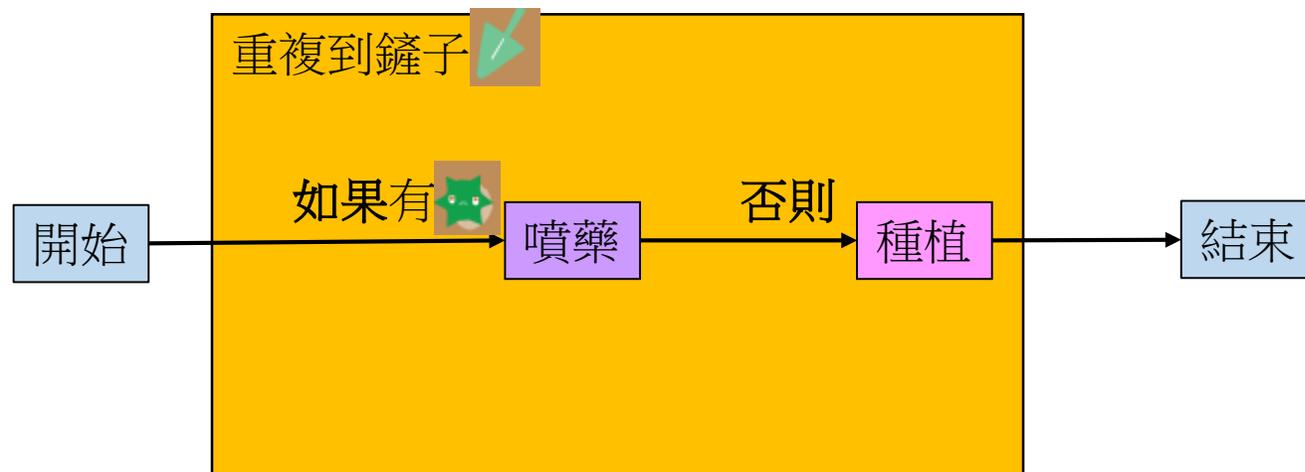
How would you instruct the Foxes to plant this row? Pay close attention to the word *not*.

7

How would you instruct the Foxes to plant this row? Pay close attention to the word *not*.
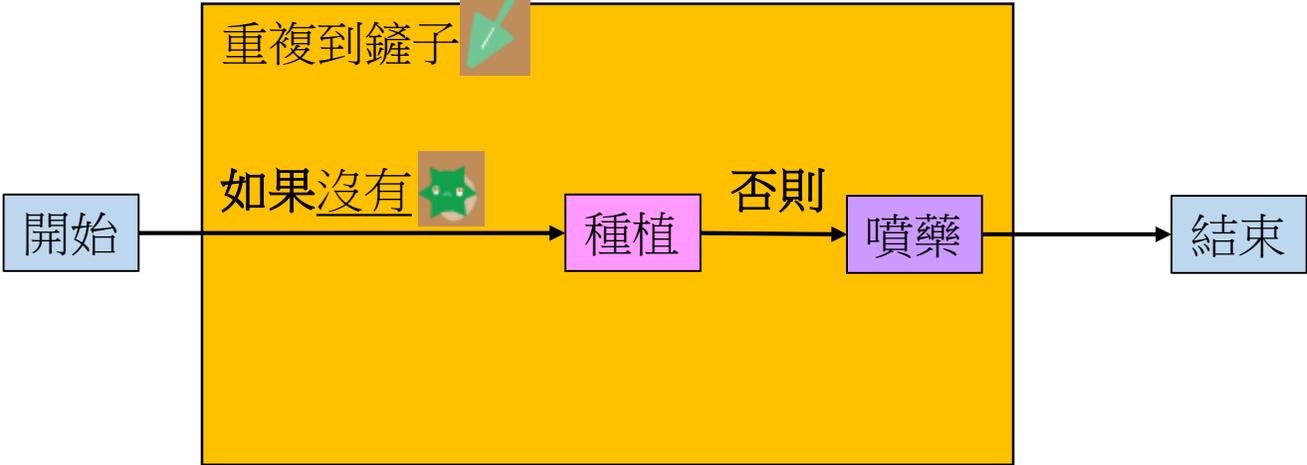
重複到鏟子

如果沒有

否則

開始 → 種植 → 噴藥 → 結束

Acute Inflammations Data Set
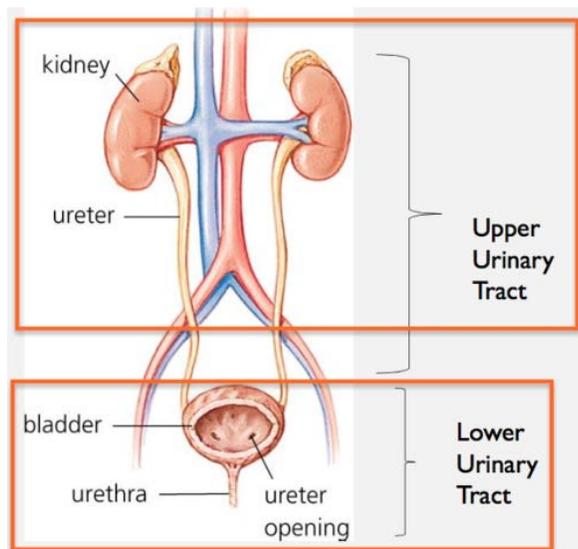
腎盂腎炎
Nephritis of renal pelvis origin

- 常發生在女性
- 突然高燒，體溫常超過40°C
- 高燒伴隨顫抖、雙邊或單側的腰椎異常疼痛
- 排尿疼痛
- 可能不規律的發生噁心、嘔吐、腹部疼痛

急性膀胱炎
Inflammation of urinary bladder

- 突發性腹部疼痛
- 常見排尿困難、排尿疼痛、頻尿
- 體溫升高，但常不超過38°C
- 排出尿液混濁，有時會有血尿

資料整理

{共120個觀察值, 6個變項, 其中一個連續型變項, 5個類別變項}

Temperature 體溫	Nausea 噁心	Lumbar pain 腰椎痛	Urine pushing 排尿困難	Micturition pains 排尿疼痛	Burning of urethra 尿道灼熱, 搔癢
35,5	no	yes	no	no	no
35,9	no	no	yes	yes	yes
35,9	no	yes	no	no	no
36,0	no	no	yes	yes	yes
36,0	no	yes	no	no	no

CTRL + F

尋找及取代

尋找(D) | 取代(P)

尋找目標(N): ,

取代成(E): .

選項(T) >>

全部取代(A) | 取代(R) | 全部尋找(I) | 找下一個(F) | 關閉

,	→	.
yes	→	1
no	→	0

Coding book

變項編號	變項英文名稱	變項中文名稱	單位	
V1	Temperature	體溫	°C	連續變項
V2	nausea	噁心		0:無(no) 1:有(yes)
V3	Lumbar pain	腰椎痛		0:無(no) 1:有(yes)
V4	Urine pushing	排尿困難		0:無(no) 1:有(yes)
V5	Micturition pains	排尿疼痛		0:無(no) 1:有(yes)
V6	Burning of urethra	尿道灼熱, 搔癢		0:無(no) 1:有(yes)

Alt+Enter : 下一行

AND、OR、NOT

Temperature 體溫	Nausea 噁心	Lumbar pain 腰椎痛	Urine pushing 排尿困難	Micturition pains 排尿疼痛	Burning of urethra 尿道灼熱, 搔癢
35,5	no	yes	no	no	no
35,9	no	no	yes	yes	yes
35,9	no	yes	no	no	no
36,0	no	no	yes	yes	yes
36,0	no	yes	no	no	no

腎盂腎炎
Nephritis of renal pelvis origin

- 常發生在女性
- 突然高燒，體溫常超過40°C
- 高燒伴隨顫抖、雙邊或單側的腰椎異常疼痛
- 排尿疼痛
- 可能不規律的發生噁心、嘔吐、腹部疼痛

急性膀胱炎
Inflammation of urinary bladder

- 突發性腹部疼痛
- 常見**排尿困難**、**排尿疼痛**、頻尿
- 體溫升高，但常不超過38°C
- 排出尿液混濁，有時會有血尿

Cystitis_1
排尿困難且排尿疼痛

Cystitis_2
排尿困難或排尿疼痛或尿道灼熱

Microsoft Excel

SPSS

R software

The R Project for Statistical Computing

Download

CRAN

R Project

About R
Logo
Contributors
What's New?
Reporting Bugs
Conferences
Search
Get Involved: Mailing Lists
Developer Pages
R Blog

R Foundation

Foundation
Board
Members
Donors
Donate

Help With R

Getting Help

Documentation

Manuals
FAQs
The R Journal
Books
Certification
Other

Links

Bioconductor
Related Projects
GSoC

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **R version 4.0.3 (Bunny-Wunnies Freak Out)** has been released on 2020-10-10.
- Thanks to the organisers of useR! 2020 for a successful online conference. Recorded tutorials and talks from the conference are available on the R Consortium YouTube channel.
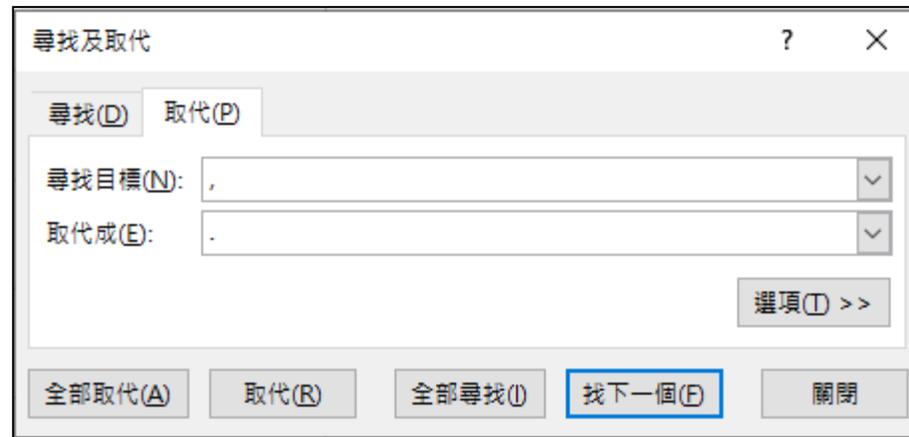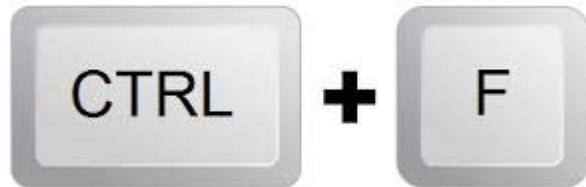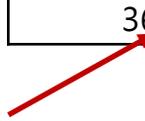- **R version 3.6.3 (Holding the Windsock)** was released on 2020-02-29.
- You can support the R Foundation with a renewable subscription as a supporting member

News via Twitter

The R Foundation
@_R_Foundation

New R blog entry by Tomas Kalibera and Simon Urbanek. Will R work on Apple Silicon?developer.r-project.org/Blog/public/20…

♡ ⤷   Nov 2, 2020

The R Foundation Retweeted

R Consortium
@RConsortium

Deadline is Oct 1! September 2020 Call for Proposals! We are soliciting progressive, pioneering projects that will benefit and serve the R community and ecosystem at large. Our goal is to foster innovation and help bring your ideas into tangible realities. bddy.me/2G9mXzA

September 2020 ISC Call for Proposal...
The deadline for submitting proposals is midnight, October 1st, 2020. The r-consortium.org

♡ ⤷   Sep 27, 2020

South Africa
    https://cran.mirror.ac.za/   TENET, Johannesburg

Spain
    https://ftp.cixug.es/CRAN/   Oficina de software libre (CIXUG)
    https://cran.rediris.es/   Spanish National Research Network, Madrid

Sweden
    https://ftpmirror1.infania.net/mirror/CRAN/   Infania Networks
    https://ftp.acc.umu.se/mirror/CRAN/   Academic Computer Club, Umeå University

Switzerland
    https://stat.ethz.ch/CRAN/   ETH Zürich

Taiwan
    https://cran.csie.ntu.edu.tw/   National Taiwan University, Taipei

Thailand
    http://mirrors.psu.ac.th/pub/cran/   Prince of Songkla University, Hatyai

Turkey
    https://cran.pau.edu.tr/   Pamukkale University, Denizli
    https://cran.gedik.edu.tr/   Istanbul Gedik University
    https://cran.ncc.metu.edu.tr/   Middle East Technical University Northern Cyprus Campus, Mersin

UK
    https://www.stats.bris.ac.uk/R/   University of Bristol
    https://cran.ma.imperial.ac.uk/   Imperial College London

USA
    https://mirror.las.iastate.edu/CRAN/   Iowa State University, Ames, IA
    http://ftp.ussg.iu.edu/CRAN/   Indiana University
    https://rweb.crmda.ku.edu/cran/   University of Kansas, Lawrence, KS
    https://repo.miserver.it.umich.edu/cran/   MBNI, University of Michigan, Ann Arbor, MI
    http://cran.wustl.edu/   Washington University, St. Louis, MO
    http://archive.linux.duke.edu/cran/   Duke University, Durham, NC
    https://cran.case.edu/   Case Western Reserve University, Cleveland, OH
    https://ftp.osuosl.org/pub/   Oregon State University
    http://lib.stat.cmu.edu/R/CRAN/   Statlib, Carnegie Mellon University, Pittsburgh, PA
    http://cran.mirrors.hoobly.com/   Hoobly Classifieds, Pittsburgh, PA
    https://mirrors.nics.utk.edu/cran/   National Institute for Computational Sciences, Oak Ridge, TN
    https://cran.revolutionanalytics.com/   Revolution Analytics, Dallas, TX

Uruguay
    https://espejito.fder.edu.uy/cran/   Facultad de Derecho, Universidad de la República

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

R for Windows

Subdirectories:

base	Binaries for base distribution. This is what you want to **install R for the first time**.
contrib	Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
old contrib	Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges).
Rtools	Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

R-4.0.3 for Windows (32/64 bit)

Download R 4.0.3 for Windows (85 megabytes, 32/64 bit)

Installation and other instructions
New features in this version

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the md5sum of the .exe to the fingerprint on the master server. You will need a version of md5sum for windows: both graphical and command line versions are available.

R-4.0.3-win

安装

```
RGui (64-bit)

檔案  編輯  看  其他  程式套件  視窗  輔助

R Console

R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R 是免費軟體，不提供任何擔保。
在某些條件下您可以將其自由散布。
用 'license()' 或 'licence()' 來獲得散布的詳細條件。

R 是個合作計劃，有許多人為之做出了貢獻。
用 'contributors()' 來看詳細的情況並且
用 'citation()' 會告訴您如何在出版品中正確地參照 R 或 R 套件。

用 'demo()' 來看一些示範程式，用 'help()' 來檢視線上輔助檔案，或
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。
用 'q()' 離開 R。

>
```

RStudio Desktop	RStudio Desktop Pro	RStudio Server	RStudio Server Pro
Open Source License	Commercial License	Open Source License	Commercial License
Free	**$995**	**Free**	**$4,975**
	/year		/year
			(5 Named Users)
DOWNLOAD	BUY	DOWNLOAD	BUY
Learn more	Learn more	Learn more	Evaluation | Learn more

RStudio Desktop 1.3.1093   - Release Notes

1. Install R.   RStudio requires R 3.0.1+.

2. Download RStudio Desktop.   Recommended for your system:

DOWNLOAD RSTUDIO FOR WINDOWS
1.3.1093 | 171.62MB
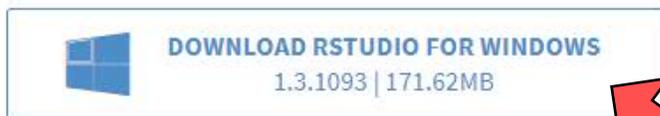
Requires Windows 10/8/7 (64-bit)
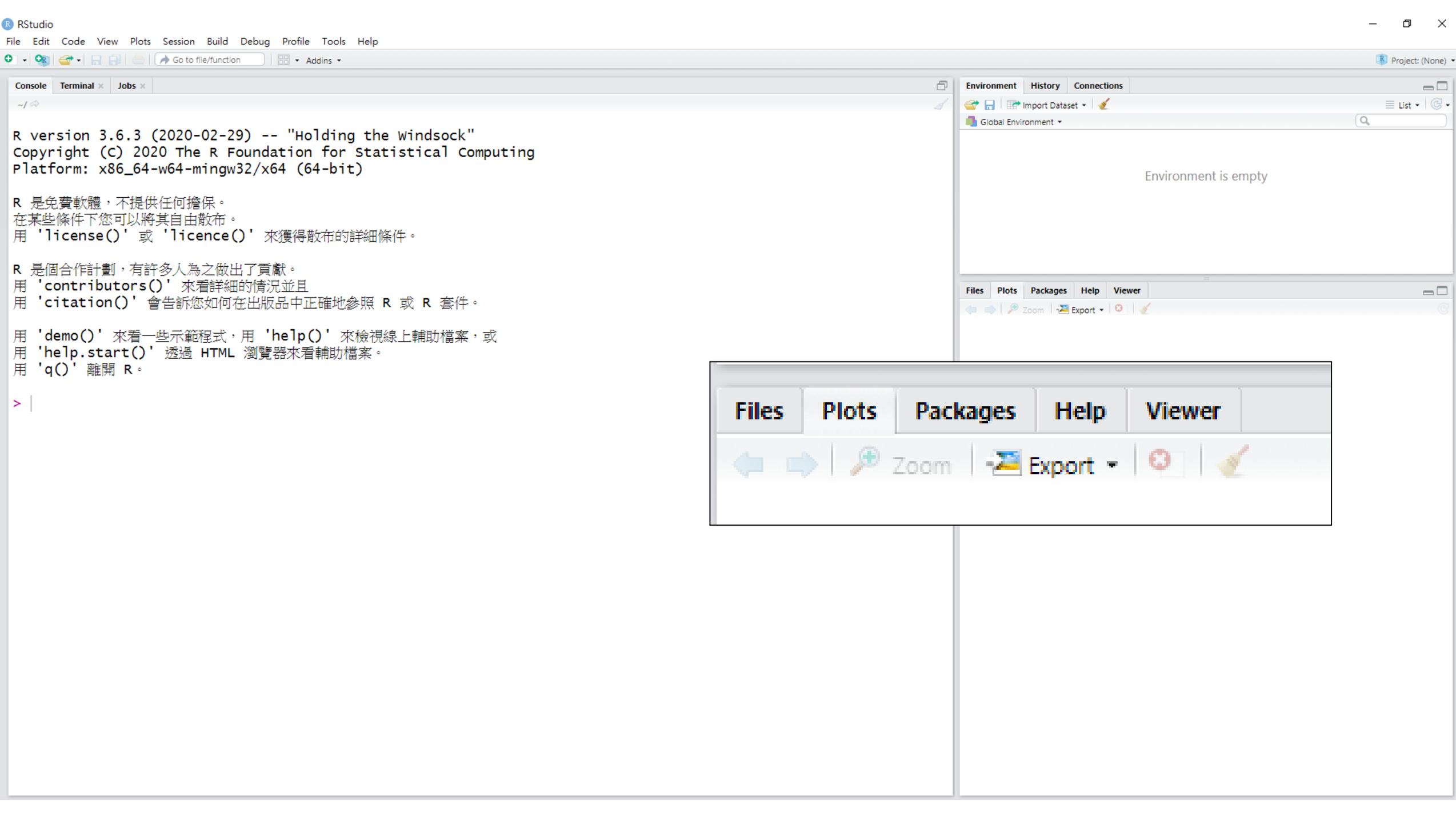
RStudio-1.3.1093

安裝

R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R 是免費軟體，不提供任何擔保。
在某些條件下您可以將其自由散布。
用 'license()' 或 'licence()' 來獲得散布的詳細條件。

R 是個合作計劃，有許多人為之做出了貢獻。
用 'contributors()' 來看詳細的情況並且
用 'citation()' 會告訴您如何在出版品中正確地參照 R 或 R 套件。

用 'demo()' 來看一些示範程式，用 'help()' 來檢視線上輔助檔案，或
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。
用 'q()' 離開 R。

>

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

New File	▶	R Script	Ctrl+Shift+N
New Project...		R Notebook	
Open File...	Ctrl+O	R Markdown...	
Recent Files	▶	Shiny Web App...	
		Plumber API...	
Open Project...			
Open Project in New Session...		Text File	
Recent Projects	▶	C++ File	
		Python Script	
Import Dataset	▶	SQL Script	
		Stan File	
Save	Ctrl+S	D3 Script	
Save As...			
Save All	Ctrl+Alt+S	R Sweave	
		R HTML	
Publish...		R Presentation	
		R Documentation	
Print...			
Close	Ctrl+W		
Close All	Ctrl+Shift+W		
Close All Except Current	Ctrl+Alt+Shift+W		
Close Project			
Quit Session...	Ctrl+Q		

ndsock"

ical Computing

R 套件。

用 `help()` 來檢視線上輔助檔案,或
劃覽器來看輔助檔案。

程式撰寫處

```
1
2  a <- 3
3  b <- 6
4
5  print(a+b)
6
7
```

```
1
2  a <- 3
3  b <- 6
4
5  print(a+b)
6
7
8
```

Console   Terminal   Jobs

D:/Dropbox/I JUNG/中山大學 課程/109-1 高中生物科學資優生培育計畫-高雄區/

```
> a <- 3
> b <- 6
>
> print(a+b)
[1] 9
>
```

Save file

Add comment

Set working directory. Read data.

```r
1
2  # First, set your working directory
3  setwd("D:/Dropbox/I JUNG/中山大學 課程/109-1 高中生物科學資優生培育計畫-高雄區/")
4
5  # Import the data and look at the first six rows
6  mydata <- read.csv(file = 'data/Acute Inflammations Data Set.csv')
7  head(mydata)
8  names(mydata)
```

```
D:/Dropbox/I JUNG/中山大學 課程/109-1 高中生物科學資優生培育計畫-高雄區/
> # First, set your working directory
> setwd("D:/Dropbox/I JUNG/中山大學 課程/109-1 高中生物科學資優生培育計畫-高雄區/")
>
> # Import the data and look at the first six rows
> mydata <- read.csv(file = 'data/Acute Inflammations Data Set.csv')
> head(mydata)
  Temperature nausea Lumbar.pain Urine.pushing Micturition.pains Burning.of.urethra
1        35.5     no         yes            no                no                 no
2        35.9     no          no           yes               yes                yes
3        35.9     no         yes            no                no                 no
4        36.0     no          no           yes               yes                yes
5        36.0     no         yes            no                no                 no
6        36.0     no         yes            no                no                 no
```

Description	Windows & Linux	Mac
Clear console	Ctrl+L	Ctrl+L

No. of variables. No. of observations.

變項 variables

Column?
Row?

觀察值
observations

```
> mydata
   Temperature nausea Lumbar.pain Urine.pushing Micturition.pains Burning.of.urethra
1         35.5     no         yes            no                no                 no
2         35.9     no          no           yes               yes                yes
3         35.9     no         yes            no                no                 no
4         36.0     no          no           yes               yes                yes
5         36.0     no         yes            no                no                 no
6         36.0     no         yes            no                no                 no
7         36.2     no          no           yes               yes                yes
8         36.2     no         yes            no                no                 no
9         36.3     no          no           yes               yes                yes
10        36.6     no          no           yes               yes                yes
11        36.6     no          no           yes               yes                yes
12        36.6     no         yes            no                no                 no
13        36.6     no         yes            no                no                 no
14        36.7     no          no           yes               yes                yes
15        36.7     no         yes            no                no                 no
16        36.7     no         yes            no                no                 no
17        36.8     no          no           yes               yes                yes
18        36.8     no          no           yes               yes                yes
19        36.9     no          no           yes               yes                yes
```

No. of variables. No. of observations.

```
> # method 1
> nrow(mydata)
[1] 120
> colnames(mydata)
[1] "Temperature"         "nausea"               "Lumbar.pain"         "Urine.pushing"       "Micturition.pains"
[6] "Burning.of.urethra"
> ncol(mydata)
[1] 6
>
> # method 2
> dim(mydata)
[1] 120   6
>
```

Matrix

Row 1 → $\begin{pmatrix} 6 & 10 \\ 5 & 3 \\ 0 & 2 \end{pmatrix}$ ← An element of the matrix
Row 2 →
Row 3 →

↑ Column 1 ↑ Column 2

Dimension of this matrix is 3 x 2

```
> mydata
   Temperature nausea Lumbar.pain Urine.pushing Micturition.pains Burning.of.urethra
1         35.5     no         yes            no                no                 no
2         35.9     no          no           yes               yes                yes
3         35.9     no         yes            no                no                 no
4         36.0     no          no           yes               yes                yes
5         36.0     no         yes            no                no                 no
6         36.0     no         yes            no                no                 no
7         36.2     no          no           yes               yes                yes
8         36.2     no         yes            no                no                 no
9         36.3     no          no           yes               yes                yes
10        36.6     no          no           yes               yes                yes
11        36.6     no          no           yes               yes                yes
12        36.6     no         yes            no                no                 no
13        36.6     no         yes            no                no                 no
14        36.7     no          no           yes               yes                yes
15        36.7     no         yes            no                no                 no
16        36.7     no         yes            no                no                 no
17        36.8     no          no           yes               yes                yes
18        36.8     no          no           yes               yes                yes
19        36.9     no          no           yes               yes                yes
```

Creating new variables.[ifelse()] [AND、OR、NOT]

ifelse(test_expression, x, y)

```
> a = c(5,7,2,9)
> ifelse(a %% 2 == 0,"even","odd")
[1] "odd"  "odd"  "even" "odd"
```

Arithmetic Operators

Operator	Description
+	addition
-	subtraction
*	multiplication
/	division
^ or **	exponentiation
x %% y	modulus (x mod y) 5%%2 is 1
x %/% y	integer division 5%/%2 is 2

Logical Operators

Operator	Description
<	less than
<=	less than or equal to
>	greater than
>=	greater than or equal to
==	exactly equal to
!=	not equal to
!x	Not x
x | y	x OR y
x & y	x AND y
isTRUE(x)	test if X is TRUE

Cystitis_1
排尿困難且排尿疼痛

=AND(D2=1,E2=1)

G2 =AND(D2=1,E2=1)

AND(logical1, [logical2], [logical3], ...)

	A	B	C	D	E	F	G	H
1	Temperature	nausea	Lumbar pain	Urine pushing	Micturition pains	Burning of urethra	Cystitis_1	
2	35.5	0	1	0	0	0	=AND(D2=1,E2=1)	
3	35.9	0	0	1	1	1		
4	35.9	0	1	0	0	0		
5	36	0	0	1	1	1		

```
10  # Creating new variables
11  # method 1
12  mydata$Cystitis_11 <- ifelse((mydata$Urine.pushing =="yes") & (mydata$Micturition.pains =="yes"), 1, 0)
13  head(mydata)
14
```

```
> head(mydata)
   Temperature nausea Lumbar.pain Urine.pushing Micturition.pains Burning.of.urethra Cystitis_11
1        35.5     no         yes            no                no                 no           0
2        35.9     no          no           yes               yes                yes           1
3        35.9     no         yes            no                no                 no           0
.        .0      no          no           yes               yes                yes           1
.        .0      no         yes            no                no                 no           0
.        .0      no         yes            no                no                 no           0
```

Logical Operators

Operator	Description
<	less than
<=	less than or equal to
>	greater than
>=	greater than or equal to
==	exactly equal to
!=	not equal to
!x	Not x
x | y	x OR y
x & y	x AND y
isTRUE(x)	test if X is TRUE

```
15 # method 2
16 attach(mydata)
17 mydata$Cystitis_12[(Urine.pushing =="yes")&(Micturition.pains =="yes")] <- 1
18 mydata$Cystitis_12[▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨] <- 0
19 detach(mydata)
20 head(mydata)
```
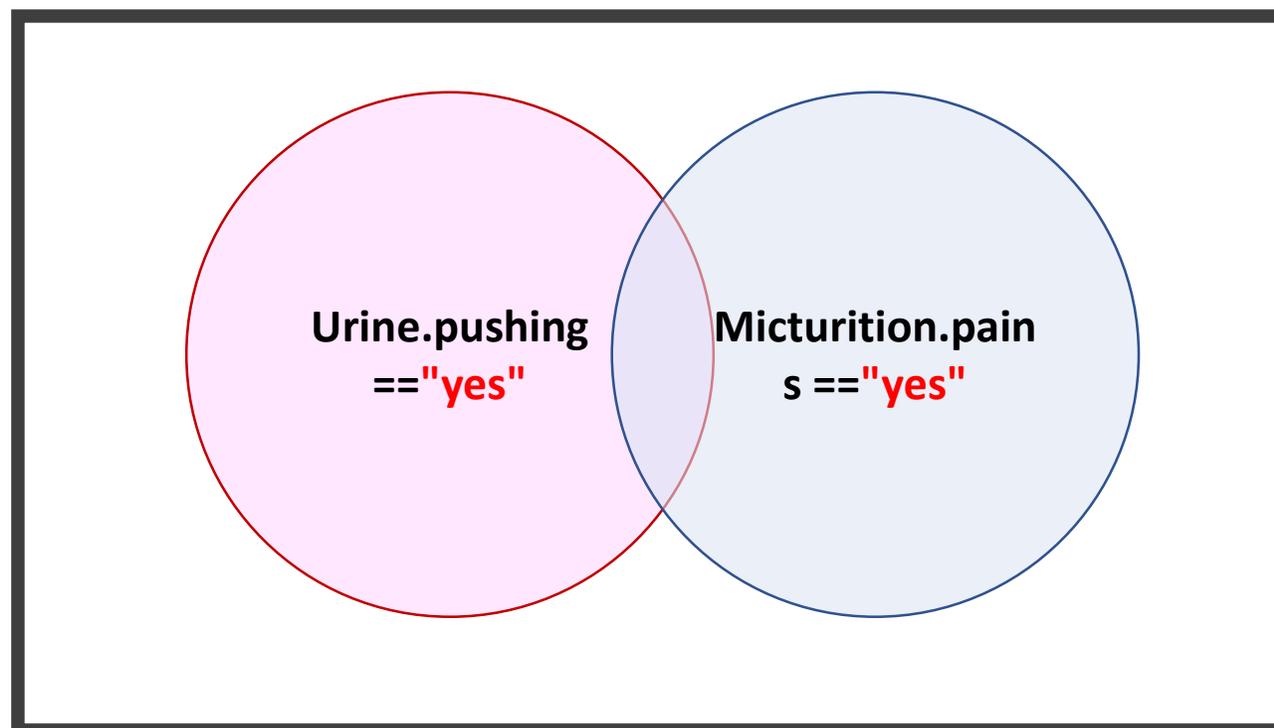
練習

```
15  # method 2
16  attach(mydata)
17  mydata$Cystitis_12[(Urine.pushing =="yes")&(Micturition.pains =="yes")] <- 1
18  mydata$Cystitis_12[(Urine.pushing =="no")|(Micturition.pains =="no")] <- 0
19  detach(mydata)
20  head(mydata)
21
```
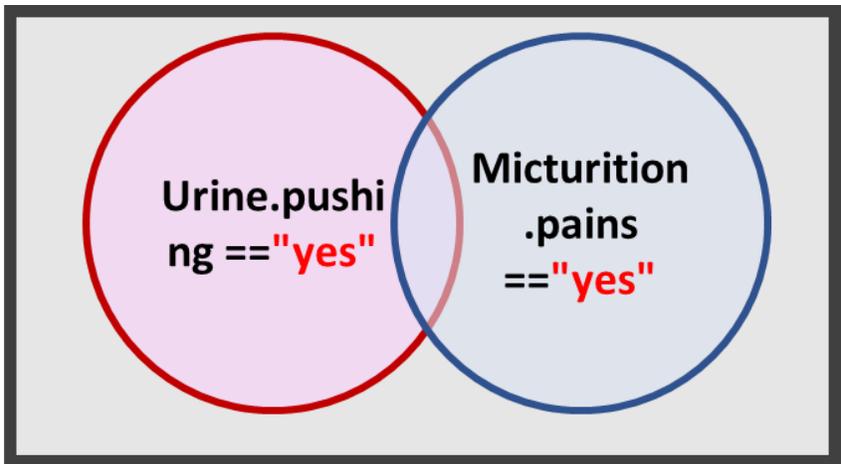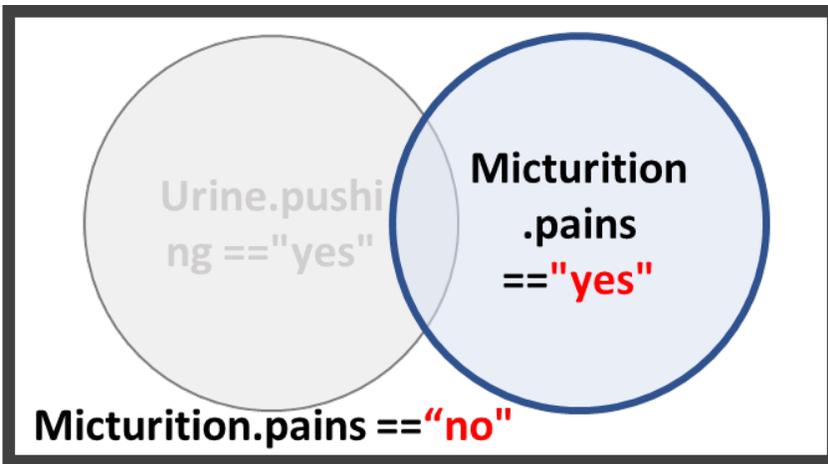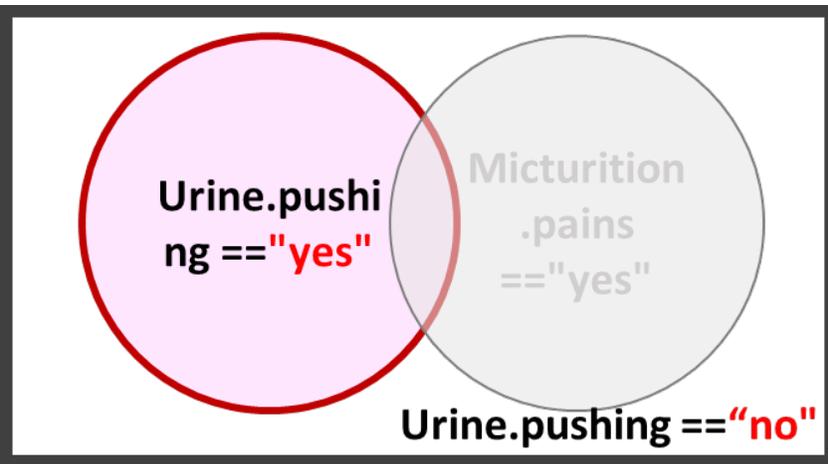

(Urine.pushing =="no") AND (Micturition.pains =="no")

```
> head(mydata)
  Temperature nausea Lumbar.pain Urine.pushing Micturition.pains Burning.of.urethra Cystitis_11 Cystitis_12
1        35.5     no         yes            no                no                 no           0           0
2        35.9     no          no           yes               yes                yes           1           1
3        35.9     no         yes            no                no                 no           0           0
4        36.0     no          no           yes               yes                yes           1           1
5        36.0     no         yes            no                no                 no           0           0
6        36.0     no         yes            no                no                 no           0           0
```

No. of conditional observations.

```
29  # method 1
30  nrow(mydata[mydata$Cystitis_11==1,])
31  nrow(mydata[mydata$Cystitis_12==1,])
32
33  # method 2
34  attach(mydata)
35  nrow(mydata[Cystitis_11==1,])
36  nrow(mydata[Cystitis_12==1,])
37  detach(mydata)
```

```
> # method 1
> nrow(mydata[mydata$Cystitis_11==1,])
[1] 49
> nrow(mydata[mydata$Cystitis_12==1,])
[1] 49
>
> # method 2
> attach(mydata)
> nrow(mydata[Cystitis_11==1,])
[1] 49
> nrow(mydata[Cystitis_12==1,])
[1] 49
> detach(mydata)
```

```
29  # method 1
30  nrow(mydata[mydata$Cystitis_11==0,])
31  nrow(mydata[mydata$Cystitis_12==0,])
32
33  # method 2
34  attach(mydata)
35  nrow(mydata[Cystitis_11==0,])
36  nrow(mydata[Cystitis_12==0,])
37  detach(mydata)
```

```
> # method 1
> nrow(mydata[mydata$Cystitis_11==0,])
[1] 71
> nrow(mydata[mydata$Cystitis_12==0,])
[1] 71
>
> # method 2
> attach(mydata)
> nrow(mydata[Cystitis_11==0,])
[1] 71
> nrow(mydata[Cystitis_12==0,])
[1] 71
> detach(mydata)
```

Cystitis_1
排尿困難且排尿疼痛
49位 膀胱炎 + 71位非膀胱炎 = 120位
(確認是否與總數相符合)

A	B	C	D	E	F	
Temperature	nausea	Lumbar pain	Urine pushing	Micturition pains	Burning of urethra	Cy
35.5	0	1	0	0	0	FALSE
35.9	0	0	1	1	1	TRUE
35.9	0	1	0	0	0	FALSE
36	0	0	1	1	1	TRUE

從 A 到 Z 排序(S)
從 Z 到 A 排序(O)
自訂排序(U)...
篩選(F)
清除(C)
重新套用(Y)

A	B	C	D	E	F	G	H
Temperature	nausea	Lumbar pain	Urine pushing	Micturition pains	Burning of urethra	Cystitis_1	

確認是否
Cystitis_1 = "TRUE" 時
Urine pushing=1 且 Micturition pains=1

Cystitis_1 = "FALSE" 時
Urine pushing=? Micturition pains=?
至少一個=0?

從最小到最大排序(S)
從最大到最小排序(O)
依色彩排序(T)
清除 "Cystitis_1" 的篩選(C)
依色彩篩選(I)
數字篩選(F)
搜尋
■ (全選)
□ FALSE
☑ TRUE

就緒 從 120中找出 49筆記錄

就緒 從 120中找出 71筆記錄

49 + 71 =120 check

10	36.3	0	0	1	1
11	36.6	0	0	1	1
12	36.6	0	0	1	1
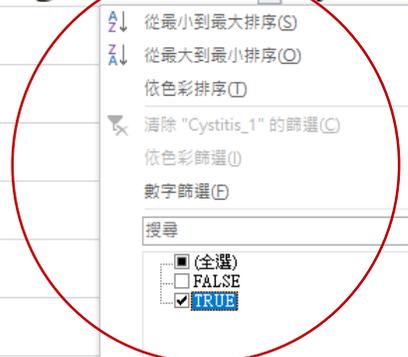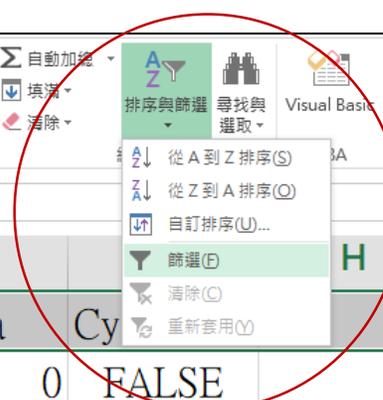
確定　取消

練習

- **Cystitis_1**
排尿困難且排尿疼痛
=AND(D2=1,E2=1)

- **Cystitis_2**
排尿困難或排尿疼痛或尿道灼熱
=OR(C2=1,D2=1,E2=1)

符合排尿困難或排尿疼痛或尿道灼熱 之**膀胱炎**與**非膀胱炎**定義患者各有幾位?

```r
#-----------------------------------------------------------------------------
# Exercise
# Creating new variables Cystitis_2
# method 1
mydata$Cystitis_21 <- ifelse( ((mydata$Urine.pushing =="yes") | (mydata$Micturition.pains =="yes")|(mydata$Burning.of.urethra =="yes")), 1, 0)
head(mydata)

# method 2
attach(mydata)
mydata$Cystitis_22 <- 0
mydata$Cystitis_22[(Burning.of.urethra =="yes")|(Urine.pushing =="yes")|(Micturition.pains =="yes")] <- 1
detach(mydata)
head(mydata)

# check no. of Cystitis_2
# method 1
nrow(mydata[mydata$Cystitis_21==1,])
nrow(mydata[mydata$Cystitis_22==1,])

# method 2
attach(mydata)
nrow(mydata[Cystitis_21==1,])
nrow(mydata[Cystitis_22==1,])
detach(mydata)


# check no. of no Cystitis_2
# method 1
nrow(mydata[mydata$Cystitis_21==0,])
nrow(mydata[mydata$Cystitis_22==0,])

# method 2
attach(mydata)
nrow(mydata[Cystitis_21==0,])
nrow(mydata[Cystitis_22==0,])
detach(mydata)
```

```
> head(mydata)
  Temperature nausea Lumbar.pain Urine.pushing Micturition.pains Burning.of.urethra Cystitis_11 Cystitis_12 Cystitis_21 Cystitis_22
1        35.5     no         yes            no                no                 no           0           0           0           0
2        35.9     no          no           yes               yes                yes           1           1           1           1
3        35.9     no         yes            no                no                 no           0           0           0           0
4        36.0     no          no           yes               yes                yes           1           1           1           1
5        36.0     no         yes            no                no                 no           0           0           0           0
6        36.0     no         yes            no                no                 no           0           0           0           0

> # check no. of Cystitis_2
> # method 1
> nrow(mydata[mydata$Cystitis_21==1,])
[1] 90
> nrow(mydata[mydata$Cystitis_22==1,])
[1] 90
>
> # method 2
> attach(mydata)
> nrow(mydata[Cystitis_21==1,])
[1] 90
> nrow(mydata[Cystitis_22==1,])
[1] 90
> detach(mydata)
>
>
> # check no. of no Cystitis_2
> # method 1
> nrow(mydata[mydata$Cystitis_21==0,])
[1] 30
> nrow(mydata[mydata$Cystitis_22==0,])
[1] 30
>
> # method 2
> attach(mydata)
> nrow(mydata[Cystitis_21==0,])
[1] 30
> nrow(mydata[Cystitis_22==0,])
[1] 30
> detach(mydata)
```

Nephritis_D

若 體溫 ≥ 38℃ 且 腰椎疼痛 [Lumber pain]
Nephritis_D 就 等於 1,
否則 Nephritis_D等於 0

= IF(AND(A2>=38,C2=1), 1, 0)

	A	B	C	D	E	F	G	H	I	J	K
	SUM				fx	= IF(AND(A2>=38,C2=1), 1, 0)					
1	Temperature	nausea	Lumbar pain	Urine pushing	Micturition pains	Burning of urethra	Cystitis_1	Cystitis_2	Nephritis_1		
2	35.5	0	1	0	0	0	FALSE	FALSE	= IF(AND(A2>=38,C2=1), 1, 0)		
3	35.9	0	0	1	1	1	TRUE	TRUE	0		
4	35.9	0	1	0	0	0	FALSE	FALSE	0		

```r
#-------------------------------------------------------------------------------

# Creating new variables Nephritis_D
# method 1
mydata$Nephritis_D1 <- ifelse((mydata$Temperature >=38) & (mydata$Lumbar.pain =="yes"), 1, 0)
head(mydata)

# method 2
attach(mydata)
mydata$Nephritis_D2 <- 0
mydata$Nephritis_D2[(Temperature >=38)&(Lumbar.pain =="yes")] <- 1
detach(mydata)
head(mydata)

# check no. of Nephritis_D1 Nephritis_D2
# method 1
nrow(mydata[mydata$Nephritis_D1==1,])
nrow(mydata[mydata$Nephritis_D2==1,])

# method 2
attach(mydata)
nrow(mydata[Nephritis_D1==1,])
nrow(mydata[Nephritis_D2==1,])
detach(mydata)

# check no. of Nephritis_D1 Nephritis_D2
# method 1
nrow(mydata[mydata$Nephritis_D1==0,])
nrow(mydata[mydata$Nephritis_D2==0,])

# method 2
attach(mydata)
nrow(mydata[Nephritis_D1==0,])
nrow(mydata[Nephritis_D2==0,])
detach(mydata)
```

```
> detach(mydata)
> # Creating new variables Nephritis_D
> # method 1
> mydata$Nephritis_D1 <- ifelse((mydata$Temperature >=38) & (mydata$Lumbar.pain =="yes"), 1, 0)
> head(mydata)
  Temperature nausea Lumbar.pain Urine.pushing Micturition.pains Burning.of.urethra Cystitis_11 Cystitis_12 Cystitis_21 Cystitis_22 Nephritis_D1 Nephritis_D2
1        35.5     no         yes            no                no                 no           0           0           0           0            0            0
2        35.9     no          no           yes               yes                yes           1           1           1           1            0            0
3        35.9     no         yes            no                no                 no           0           0           0           0            0            0
4        36.0     no          no           yes               yes                yes           1           1           1           1            0            0
5        36.0     no         yes            no                no                 no           0           0           0           0            0            0
6        36.0     no         yes            no                no                 no           0           0           0           0            0            0
>
> # method 2
> attach(mydata)
> mydata$Nephritis_D2 <- 0
> mydata$Nephritis_D2[(Temperature >=38)&(Lumbar.pain =="yes")] <- 1
> detach(mydata)
> head(mydata)
  Temperature nausea Lumbar.pain Urine.pushing Micturition.pains Burning.of.urethra Cystitis_11 Cystitis_12 Cystitis_21 Cystitis_22 Nephritis_D1 Nephritis_D2
1        35.5     no         yes            no                no                 no           0           0           0           0            0            0
2        35.9     no          no           yes               yes                yes           1           1           1           1            0            0
3        35.9     no         yes            no                no                 no           0           0           0           0            0            0
4        36.0     no          no           yes               yes                yes           1           1           1           1            0            0
5        36.0     no         yes            no                no                 no           0           0           0           0            0            0
6        36.0     no         yes            no                no                 no           0           0           0           0            0            0
```

```
> # check no. of Nephritis_D1 Nephritis_D2
> # method 1
> nrow(mydata[mydata$Nephritis_D1==1,])
[1] 50
> nrow(mydata[mydata$Nephritis_D2==1,])
[1] 50
>
> # method 2
> attach(mydata)
> nrow(mydata[Nephritis_D1==1,])
[1] 50
> nrow(mydata[Nephritis_D2==1,])
[1] 50
> detach(mydata)
>
> # check no. of Nephritis_D1 Nephritis_D2
> # method 1
> nrow(mydata[mydata$Nephritis_D1==0,])
[1] 70
> nrow(mydata[mydata$Nephritis_D2==0,])
[1] 70
>
> # method 2
> attach(mydata)
> nrow(mydata[Nephritis_D1==0,])
[1] 70
> nrow(mydata[Nephritis_D2==0,])
[1] 70
> detach(mydata)
```

練習

- **Cystitis_D**
條件1:
<u>體溫 < 38°C</u> **且** <u>排尿困難 [Urine pushing]</u>
條件2:
<u>排尿困難 [Urine pushing] 且 排尿疼痛 [Micturition pains]</u>

若 符合條件1 **或是** 條件2
則 Cystitis_D **就** 等於 1,
否則 Cystitis_D 等於 0

IF、OR、AND

```r
#-----------------------------------------------------------------------------

# Creating new variables Cystitis_D

# method 1
mydata$Cystitis_C11 <- ifelse((mydata$Temperature < 38) & (mydata$Urine.pushing =="yes"), 1, 0)
mydata$Cystitis_C21 <- ifelse((mydata$Urine.pushing =="yes") & (mydata$Micturition.pains =="yes"), 1, 0)
mydata$Cystitis_D1 <- ifelse((mydata$Cystitis_C11==1) | (mydata$Cystitis_C21==1),1,0)

head(mydata)

# method 2
attach(mydata)
mydata$Cystitis_C12 <- 0
mydata$Cystitis_C12[(Temperature < 38)&(Urine.pushing =="yes")] <- 1
mydata$Cystitis_C22 <- 0
mydata$Cystitis_C22[(Urine.pushing =="yes")&(Micturition.pains =="yes")] <- 1
mydata$Cystitis_D2 <- 0
mydata$Cystitis_D2[(Cystitis_C12 ==1)|(Cystitis_C22 == 1)] <- 1
detach(mydata)
head(mydata)

# check no. of Cystitis_D1 Cystitis_D2
# method 1
nrow(mydata[mydata$Cystitis_D1==1,])
nrow(mydata[mydata$Cystitis_D2==1,])

# method 2
attach(mydata)
nrow(mydata[Cystitis_D1==1,])
nrow(mydata[Cystitis_D2==1,])
detach(mydata)

# check no. of Nephritis_D1 Nephritis_D2
# method 1
nrow(mydata[mydata$Cystitis_D1==0,])
nrow(mydata[mydata$Cystitis_D2==0,])

# method 2
attach(mydata)
nrow(mydata[Cystitis_D1==0,])
nrow(mydata[Cystitis_D2==0,])
detach(mydata)
```

```
> # check no. of Cystitis_D1 Cystitis_D2
> # method 1
> nrow(mydata[mydata$Cystitis_D1==1,])
[1] 59
> nrow(mydata[mydata$Cystitis_D2==1,])
[1] 59
>
> # method 2
> attach(mydata)
> nrow(mydata[Cystitis_D1==1,])
[1] 59
> nrow(mydata[Cystitis_D2==1,])
[1] 59
> detach(mydata)
>
> # check no. of Nephritis_D1 Nephritis_D2
> # method 1
> nrow(mydata[mydata$Cystitis_D1==0,])
[1] 61
> nrow(mydata[mydata$Cystitis_D2==0,])
[1] 61
>
> # method 2
> attach(mydata)
> nrow(mydata[Cystitis_D1==0,])
[1] 61
> nrow(mydata[Cystitis_D2==0,])
[1] 61
> detach(mydata)
```

"**Knowledge** is the collection of skills and information a person has acquired through experience. **Intelligence** is the ability to apply knowledge."

"Knowledge is wonderful, but it fades as techniques and technologies come and go. Intelligence sustains. Its borders extend beyond any technique or technology, and that makes all the difference."

Anthony Colangelo

台北醫學大學附設醫院院長陳瑞杰｜最懂區塊鏈的院長：智慧醫院只是手段，變成「好醫院」才是目的
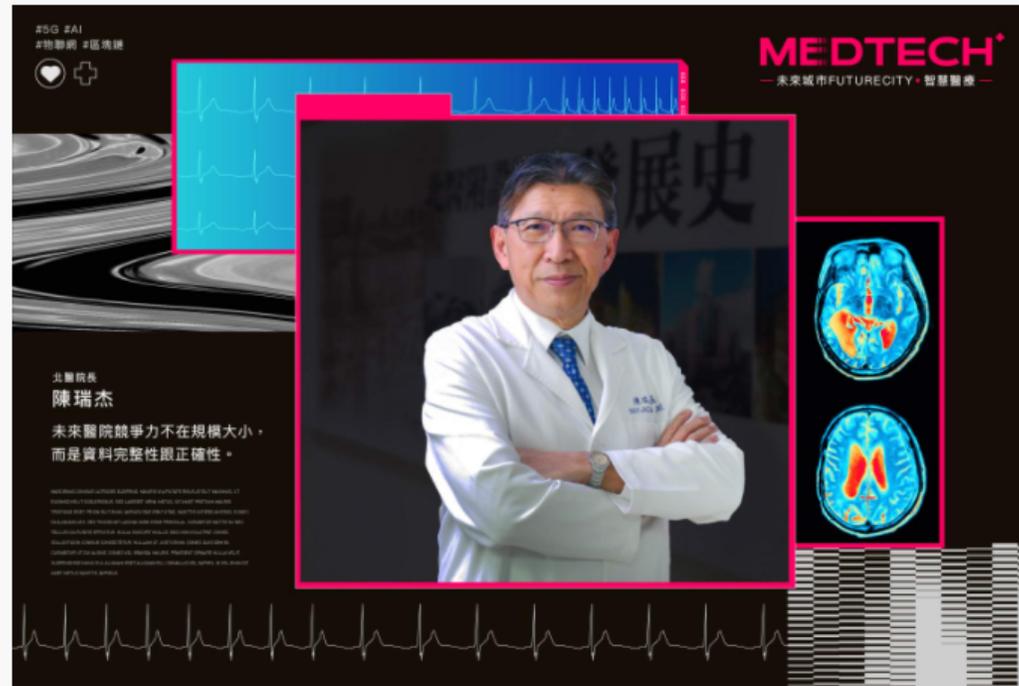
台北醫學大學附設醫院院長陳瑞杰認為，智慧醫療的最終目的是透過智慧科技變成「好醫院」。 圖片來源：台北醫學大學附設醫院　首圖設計：yao ting

2020-10-14

Thank You
For
Your Attention!

Any Questions?

回家作業

E-mail: ijfeng@g-mail.nsysu.edu.tw

Title: [109-1 高中生物科學資優生培育計畫-高雄區_姓名_學號]

答案 [PPT]

程式碼 [R]

Echocardiogram Data Set 心臟超音波資料集

Donor: Steven Salzberg and Dr. Evlin Kinney

變項編號	變項英文名稱	變項中文名稱	單位		解釋
V1	age-at-heart-attack	心臟病發病年齡	year(s)	連續變項	
V2	pericardial-effusion	心包膜積水		0:無(no) 1:有(yes)	
V3	fractional-shortening	短縮分率			左心收縮力指標之一，數值越高越不正常。
V4	epss	epss			E-point septal separation。左心收縮力指標之一，數值越高越不正常。
V5	lvdd	左心室舒張末期內徑			left ventricular end-diastolic dimension。舒張末期心臟大小，數值越大越不正常。
V6	wall-motion-index	室壁運動記分指數			室壁運動異常程度，數值越高越不正常。

1. 共有＿＿ 觀察值, 有＿＿個變項。

我們欲根據心臟超音波研究結果找出有哪些心臟病發患者心臟病發後生存年份短於1年。

研究發現年齡大於65歲老年人, **若** fractional-shortening大於等於0.15 **且** EPSS大於等於10且LVDD大於等於4.5 **且** wall-motion-index大於1.0 **則**患者存活時間短於1年。

1. 有幾位病患 心臟病發病時年齡大於等於65 歲?＿＿＿＿

2. 有幾位病患 fractional-shortening大於等於0.15?＿＿＿＿

3. 有幾位病患 EPSS大於等於10?＿＿＿＿

4. 有幾位病患 LVDD大於等於4.5?＿＿＿＿

5. 有幾位病患 wall-motion-index大於1.0?＿＿＿＿

6. 2~6條件皆符合得有幾位? ＿＿＿＿

研究另外發現年齡小於65歲者, 雖然fractional-shortening小於0.15 ，但是EPSS、LVDD與wall-motion-index 3項其中2項以上符合更為嚴苛的條件時，患者存活時間短於1年。

7. 有幾位病患 心臟病發病時年齡小於65 歲?_____

8. 有幾位病患 fractional-shortening小於0.15?_____

9. 有幾位病患 EPSS大於等於15?_____

10. 有幾位病患 LVDD大於等於4.5?_____

11. 有幾位病患 wall-motion-index大於2.0?_____

12. 有幾位病患符合條件9、10 、11中2項以上者? _____

13. 有幾位病患符合條件7 且條件8且條件12者? _____

14. 根據研究所發現特徵，請問共有幾位患者存活時間短於1年 (符合條件6或是條件13)？_____