

「商業資料分析與管理決策」 ——統計與大數據

Spring 2025

教師：統計系余清祥

內容：統計與生活

下載：<http://csyue.nccu.edu.tw>

Email: csyue@nccu.edu.tw

2025年3月9日

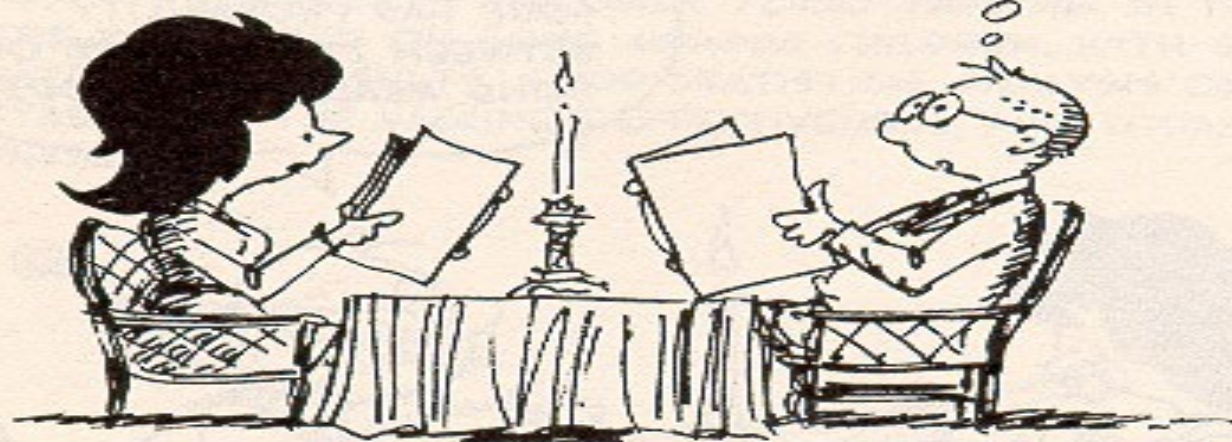


WHAT IS STATISTICS?

WE MUDDLE THROUGH LIFE MAKING CHOICES
BASED ON INCOMPLETE INFORMATION...

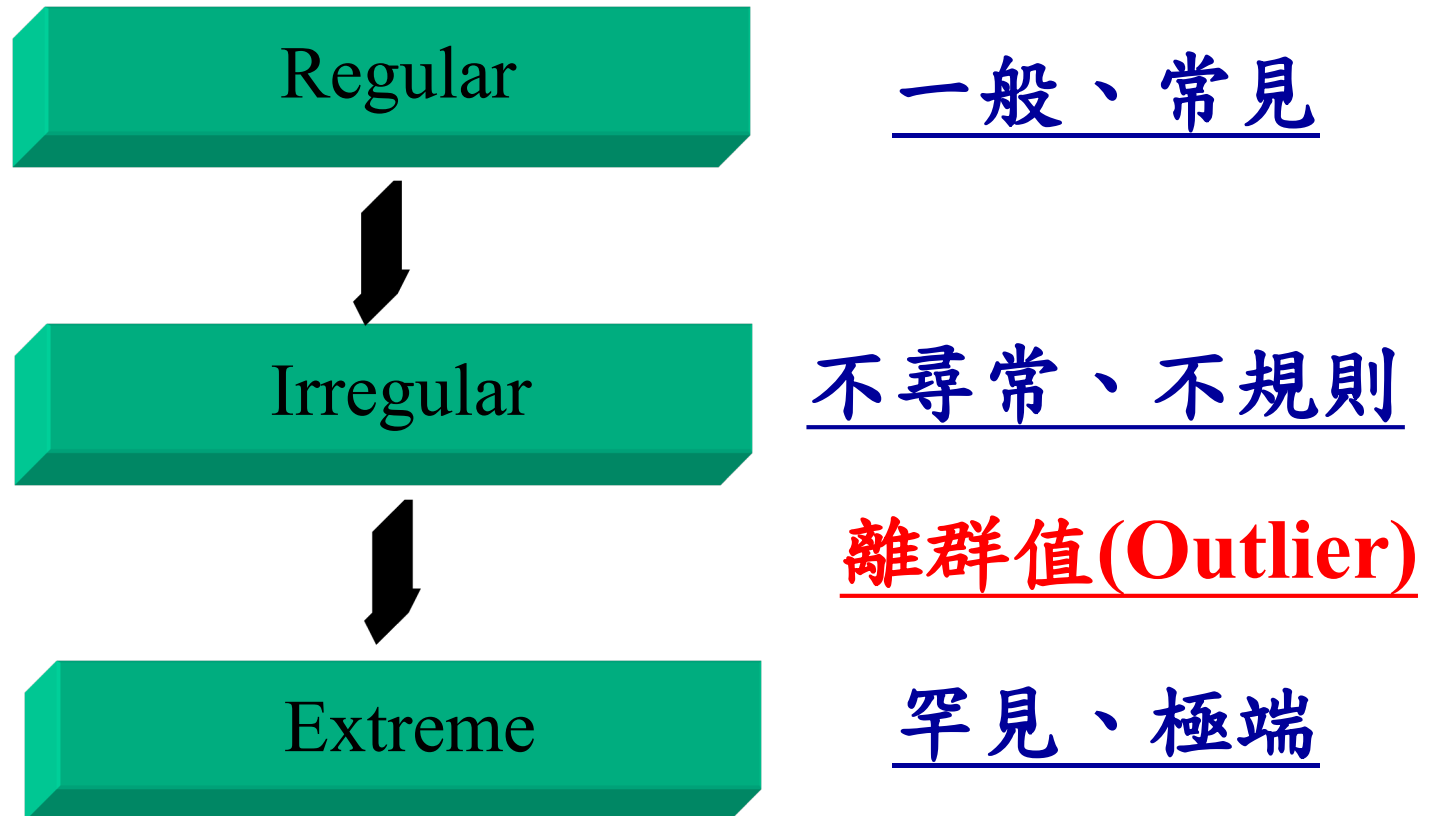
SHOULD I HAVE THE SOUP?
EVERYTHING ELSE IS SO
EXPENSIVE, AND I DON'T
KNOW WHO'S PAYING... ARE
STATISTICIANS STINGY? I'VE
NEVER GONE OUT WITH
ONE BEFORE... THOUGH I
ONCE KNEW A VERY
GENEROUS ACCOUNTANT...

SHOULD I HAVE THE SOUP?
27 OUT OF THE 36 TIMES
I'VE HAD IT, IT WAS PRETTY
GOOD... BUT IS MONDAY THE
REGULAR CHEF'S NIGHT
OFF? AND WHAT IF ALL THE
AIR MOLECULES IN THE
ROOM SUDDENLY FLY UP TO
THE CEILING?



統計與知識

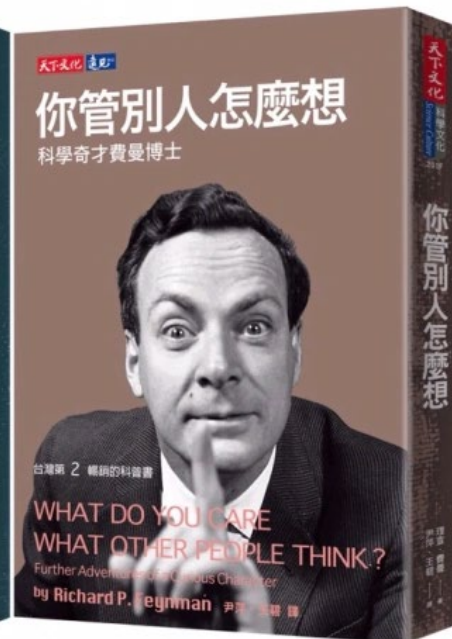
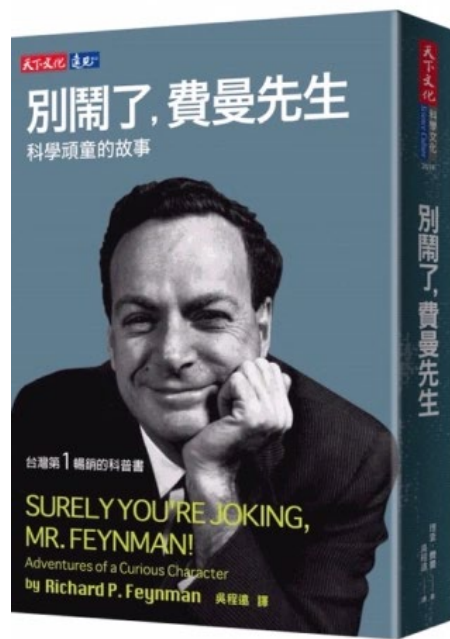
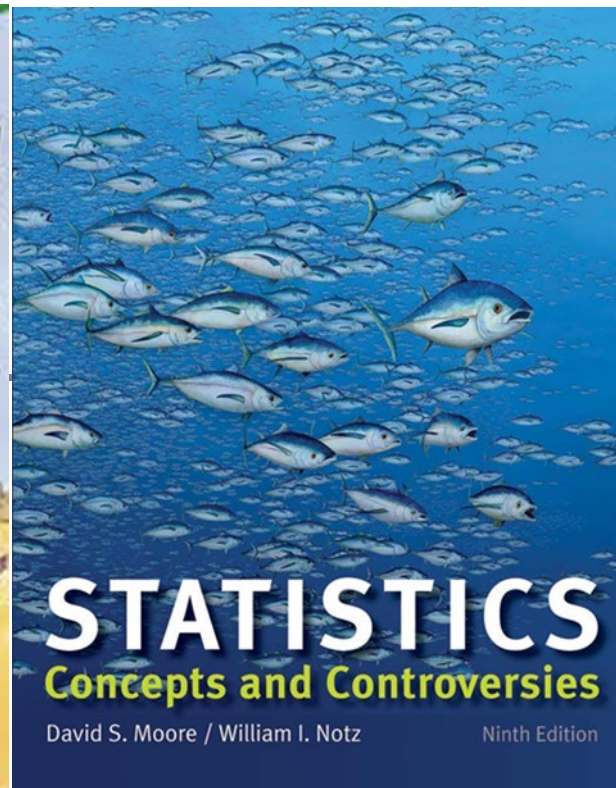
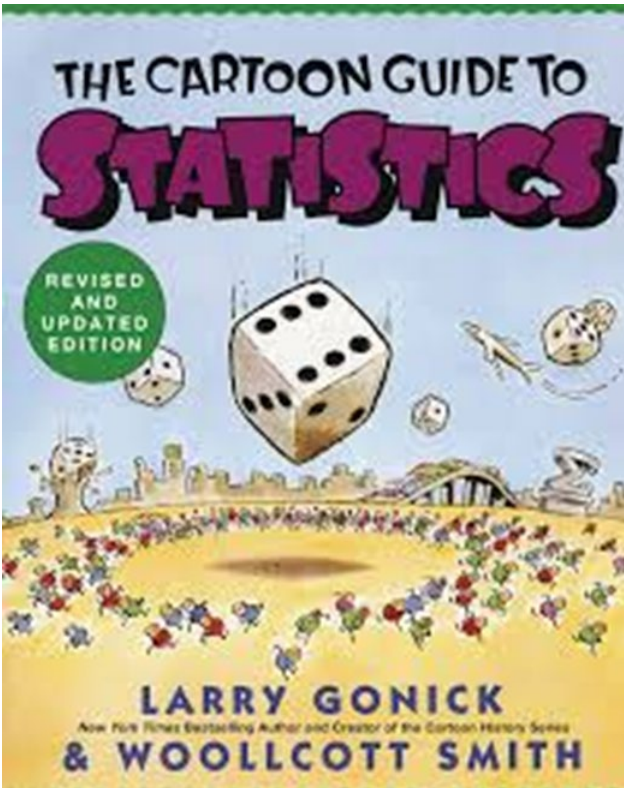
- 統計整理資訊的方法屬於歸納法(Induction)，從龐雜的資料找出共同趨勢，並區分資料具有以下哪一種特性：



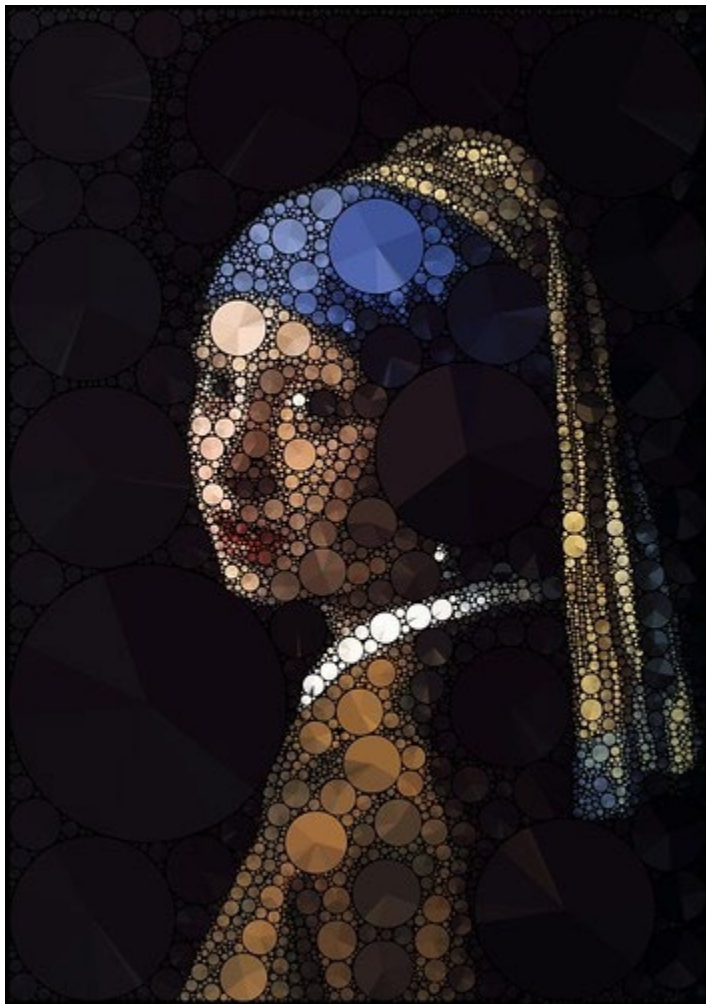


幾本有趣的參考書籍

- 《看漫畫，學統計》(2003)，天下文化。
→ The Cartoon Guide to Statistics (1993)
- 《統計，讓數字說話》(1998)，天下文化。
- 《統計，改變了世界》(2001)，天下文化。
- 《你管別人怎麼想：科學奇才費曼博士》
(2005)，天下文化。
- 《真實的謊言—揭開民調與統計的黑盒子》
(1996)，時報文化。



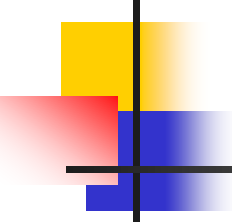
什麼是統計？



http://farm2.static.flickr.com/1281/4669551487_34496c71aa.jpg



<https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.canstockphoto.com%2Fbright-statistics-illustration-35101794.html&psig=AOvVaw10rqxE16qvmWwyShbVj6ml&ust=1672757240051000&source=images&cd=vfe&ved=0CBAQjRxqFwoTCPcagpCRqfwCFQAAAAAdAAAAABAJ>



商學院有3G（三「計」）

會計——很快忘記！

經濟——經常忘記！！

統計——通通忘記！！！！

註：為什麼這三門課特別棘手，統計
似乎最難上手？

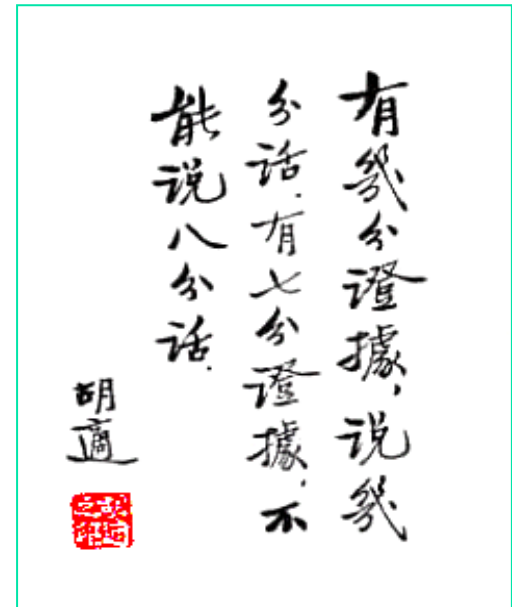


什麼是統計？

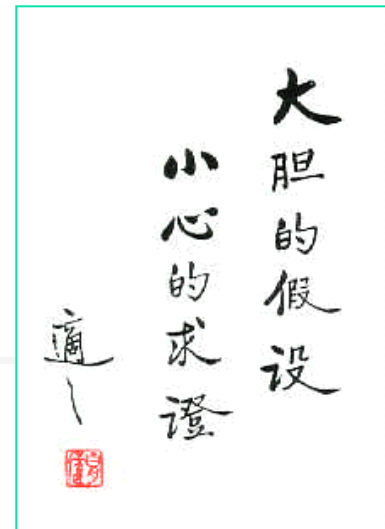
- 統計學是研究定義問題、運用資料蒐集、整理、陳示、分析與推論等科學方法，在不確定(Uncertainty)情況下，做出合理決策的科學。



https://en.wikipedia.org/wiki/File:Hu_Shih_1960_color.jpg



學習統計的幾個要素



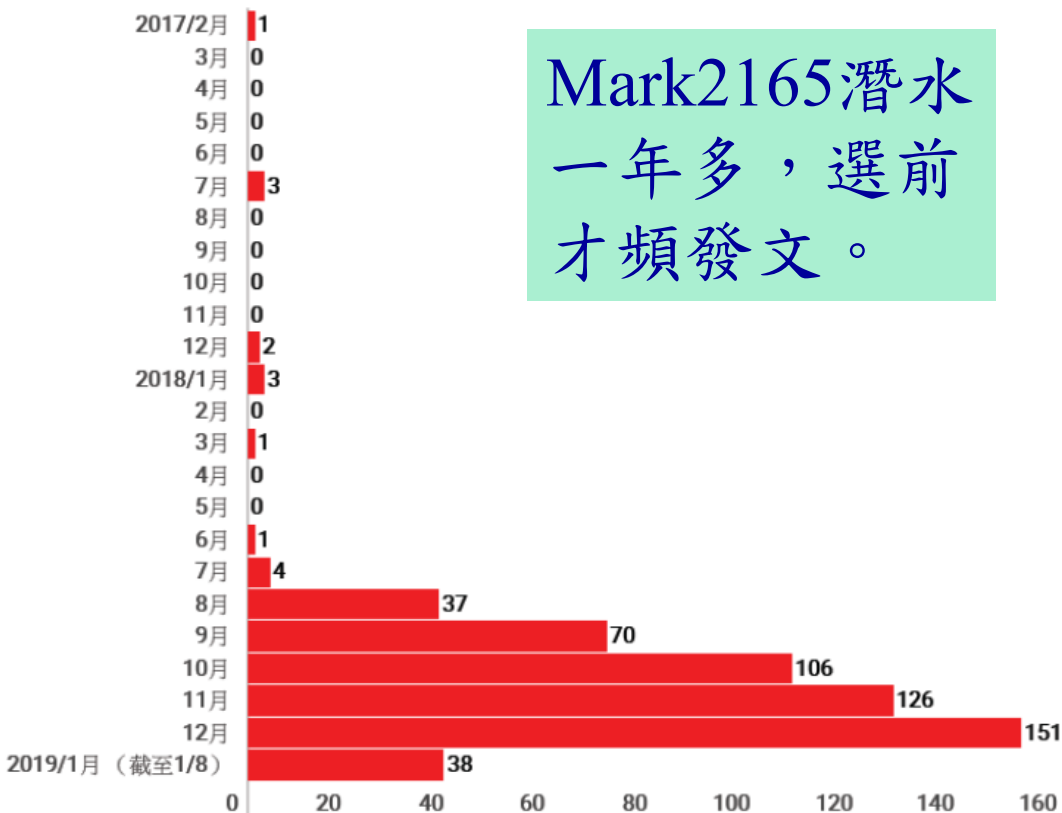
■ 解決問題大致有以下幾個要素：

- 如何定義、測量？(e.g. Variable Format, Data Collection)
- 如何判斷、取捨？(e.g. Estimation, Prediction, Testing)
- 如何詮釋、增加附加價值？(e.g. Utility, Decision)

如何從龐雜訊息中篩選資訊？

- 訊息產生及傳播方式愈趨大量多元，如何篩選、整理、分析變得更加困難。

Mark2165潛水
一年多，選前
才頻發文。



<https://www.wittenberg.edu/sites/default/files/styles/boxlarge2x/public/2020-05/Insta%20Fake%20News.png?itok=16pf5u7H>

<https://www.cw.com.tw/article/5093610>



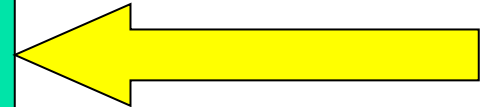
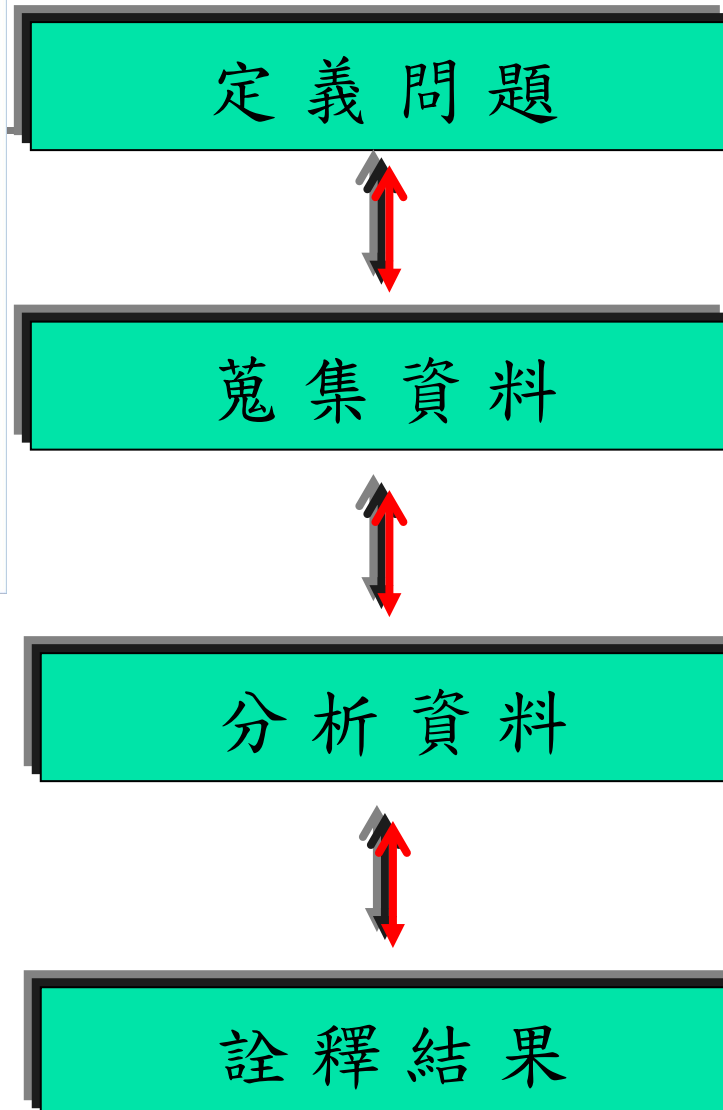
報紙上常看到的敘述！

- 2004年九月底某學校公布使用手機的肇事率，發現在車內用免持聽筒打電話，比用手持聽筒還要危險。
- 由於民眾以為使用免持聽筒就能夠一面開車、一面通話，但往往更容易分心，肇事率比用手持聽筒高出將近1倍。
- 你/妳覺得這裡有甚麼疑問？

解決問題(Problem Solving)流程



<https://www.oneeducation.org.uk/course/creativity-and-problem-solving-skills/>



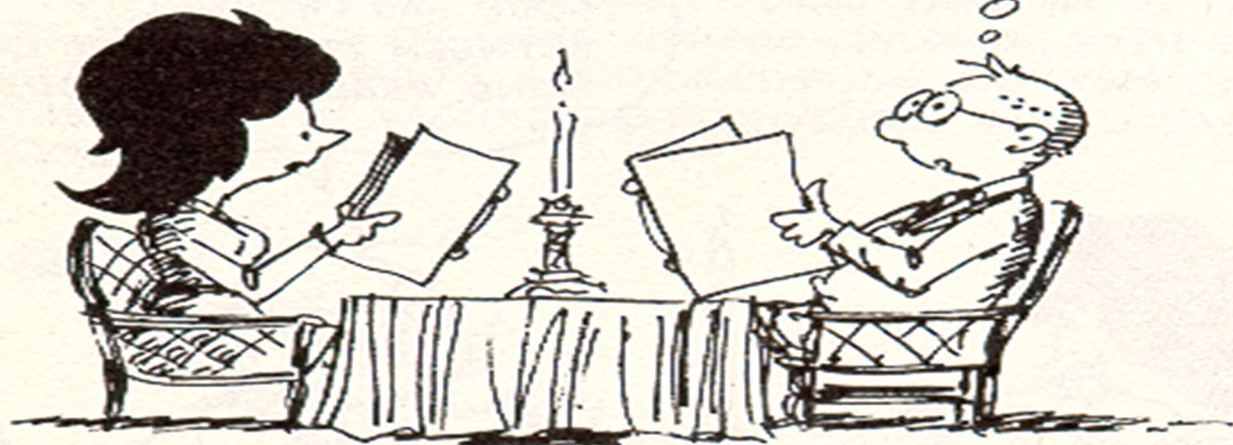
絕大多數的
資料分析重心

WHAT IS STATISTICS?

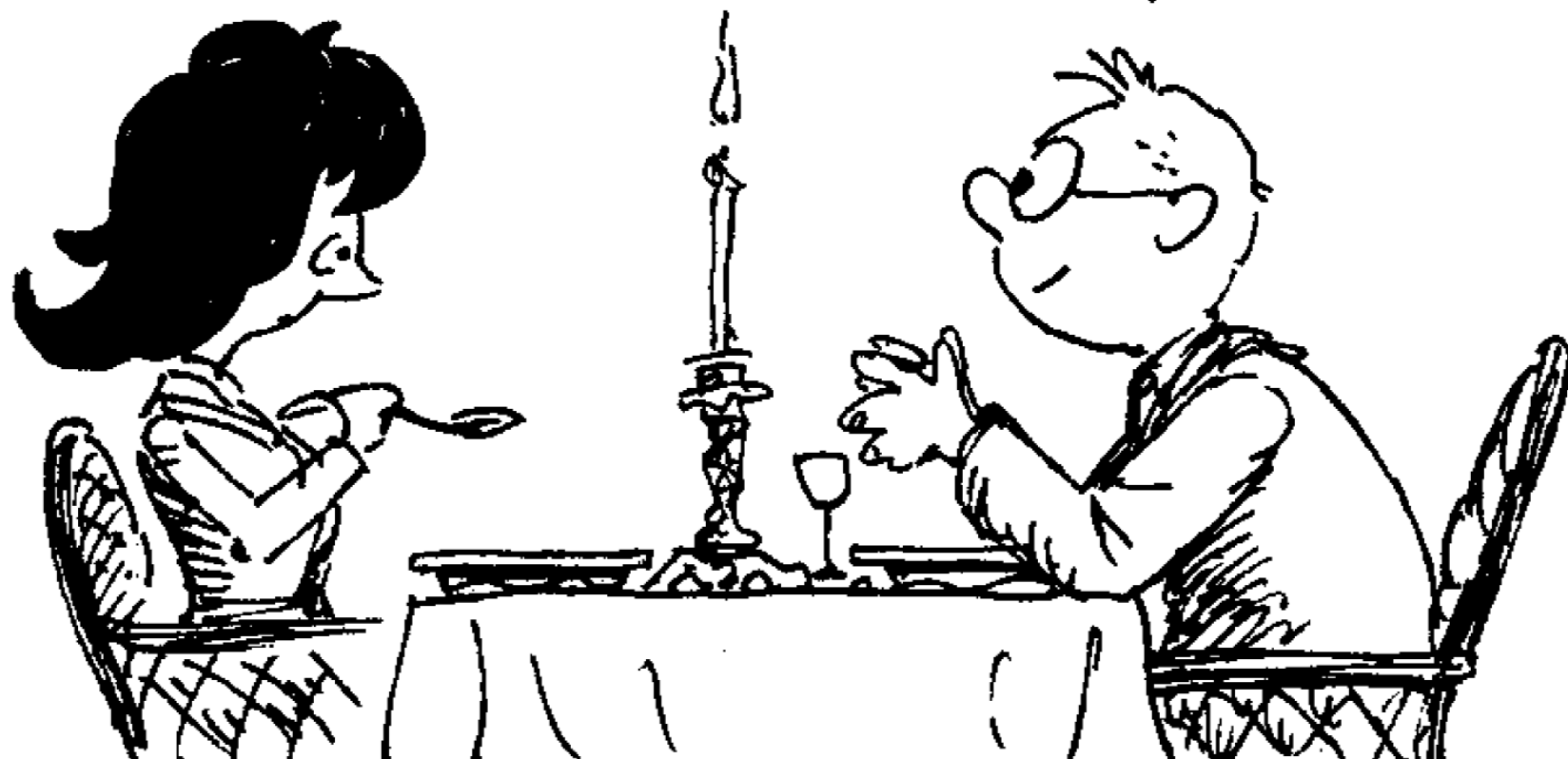
WE MUDDLE THROUGH LIFE MAKING CHOICES
BASED ON INCOMPLETE INFORMATION...

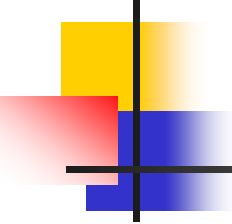
SHOULD I HAVE THE SOUP?
EVERYTHING ELSE IS SO
EXPENSIVE, AND I DON'T
KNOW WHO'S PAYING... ARE
STATISTICIANS STINGY? I'VE
NEVER GONE OUT WITH
ONE BEFORE... THOUGH I
ONCE KNEW A VERY
GENEROUS ACCOUNTANT...

SHOULD I HAVE THE SOUP?
27 OUT OF THE 36 TIMES
I'VE HAD IT, IT WAS PRETTY
GOOD... BUT IS MONDAY THE
REGULAR CHEF'S NIGHT
OFF? AND WHAT IF ALL THE
AIR MOLECULES IN THE
ROOM SUDDENLY FLY UP TO
THE CEILING?



GOOD CHOICE! I'M 95%
CONFIDENT THAT TONIGHT'S
SOUP HAS PROBABILITY
BETWEEN 73% AND 77% OF
BEING REALLY DELICIOUS!





馬克吐溫對統計的想法

There are three kinds of lies:

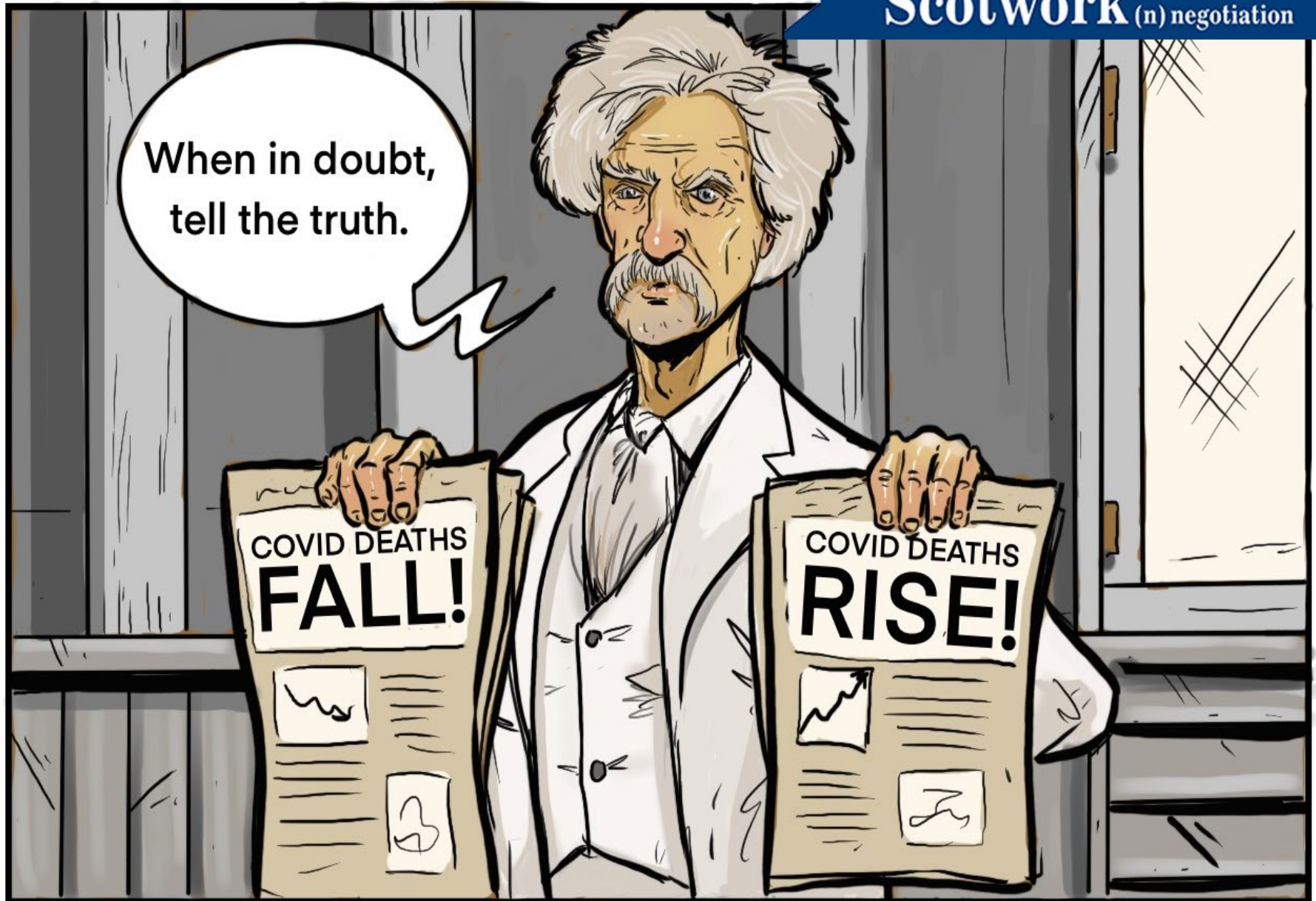
Lies,

Damned lies,

and Statistics!!

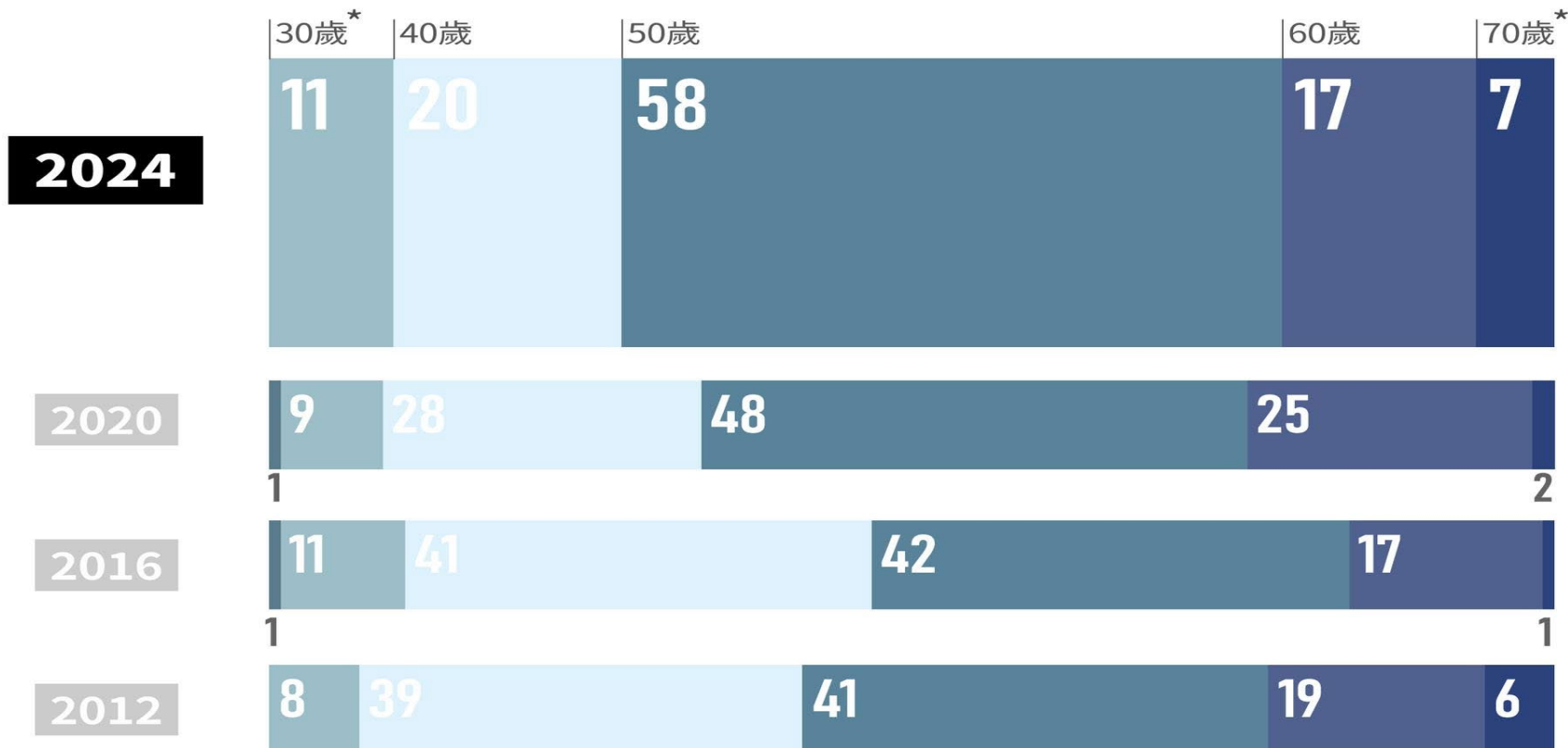
by Benjamin Disraeli (英國首相)





第11屆立委年齡分布

單位：人



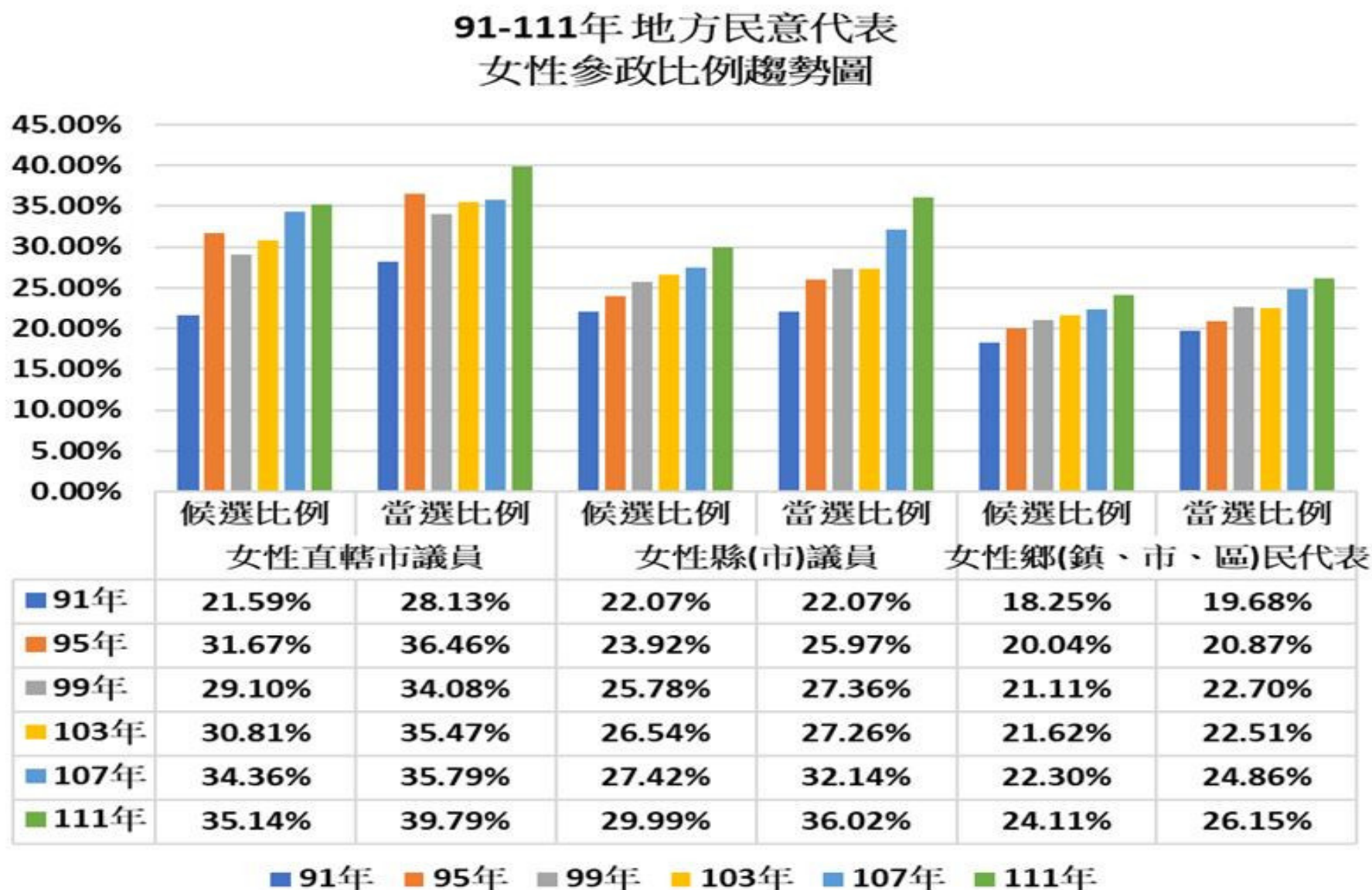
*註：70歲以上 - 游錫堃(75)、陳永康(73)、柯建銘(72)、賴士葆(72)、陳超明(72)
陳雪生(72)、林德福(70)

30至39歲 - 黃捷(30)、廖偉翔(33)、牛煦庭(33)、羅廷瑋(34)、徐巧芯(34)
廖先翔(35)、黃健豪(35)、邱若華(35)、吳沛憶(36)、陳冠廷(37)、洪申翰(39)

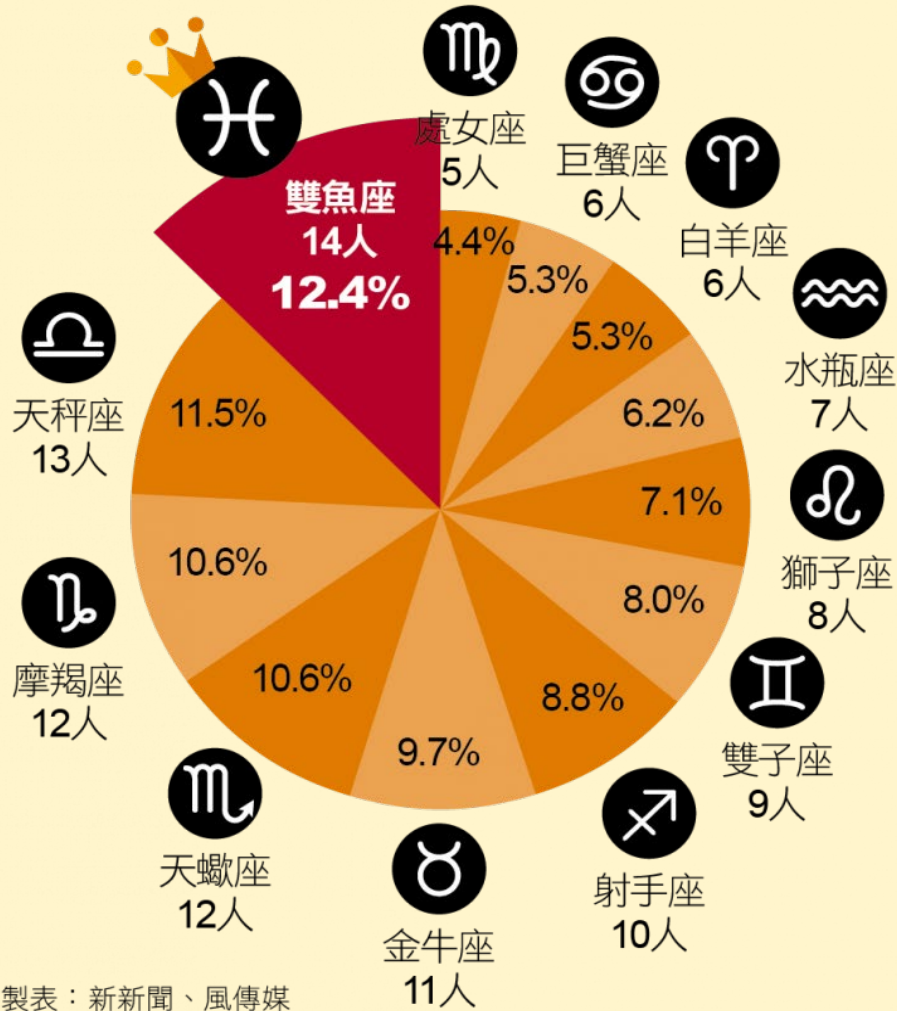
資料來源：中選會、立法院

https://scontent-tpe1-1.xx.fbcdn.net/v/t39.30808-6/425288557_770454721779762_4629537348715970133_n.jpg?_nc_cat=110&ccb=1-7&_nc_sid=3635dc&_nc_ohc=nK9p5qScxDsAX_3BIDr&_nc_ht=scontent-tpe1-1.xx&oh=00_AfD6Gfg9MnO1WINiv4RLSqkS5najZlDyN09IUmy-hZ2VLA&oe=65E036E2

表 2：近 20 年(六屆) 地方民意代表選舉結果性別統計資料



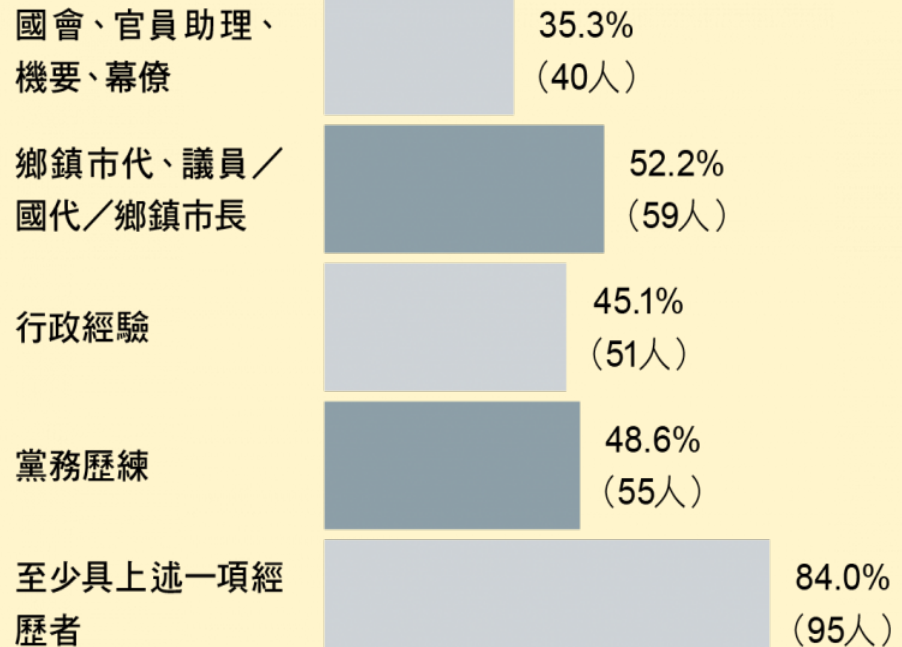
雙魚座最適合當立委？



- 適合度的定義？
- 分布均勻嗎？(0.4678)

選立委需要具備哪些履歷？

■ 本屆立委歷任》

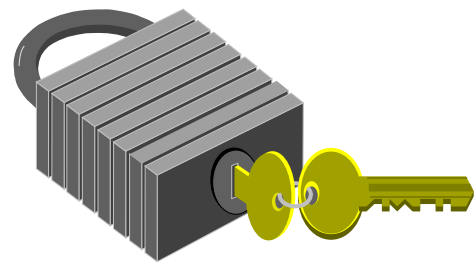


註1.統整立委當選前的經歷。連任立委以統計選上首任立委前的經歷為主；屆期不連續者，以選上最近一任之前的經歷為主。

註2：黨務經歷不計國民黨、民進黨全國黨代表、國民黨中央委員

資料來源：選舉公報、立委網頁自介、維基百科

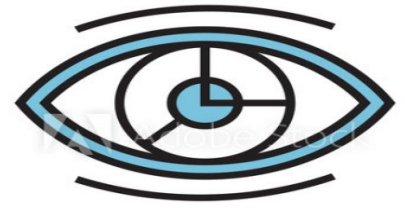
資料整理：新新聞、風傳媒



什麼是統計？

- 統計學是研究定義問題、運用資料蒐集、整理、陳示、分析與推論等科學方法，在不確定(Uncertainty)情況下，做出合理決策的科學。

註：通常由部分資料（「樣本」 Sample）推敲出整體（「母體」 Population）的特性。



Visualization

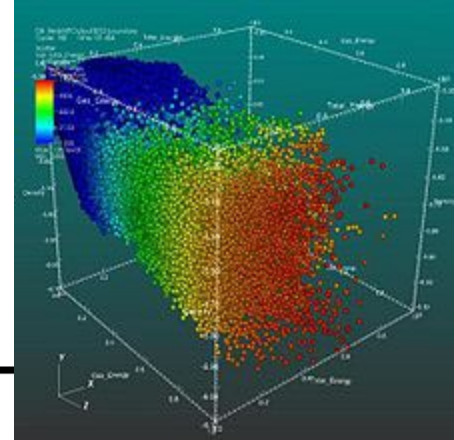
#212488495

文字分析：觀念的視覺化

為什麼需要文學？了解文學、接近文學，對我們形成價值判斷有什麼關係？如果說，文學有一百種所謂「功能」而我必須選擇一種最重要的，我的答案是：德文有一個很精確的說法「macht sichtbar」，意思是「**使看不見的東西被看見**」。在我自己的體認中，這就是文學跟藝術最重要、最實質、最核心的一個作用。

——龍應台

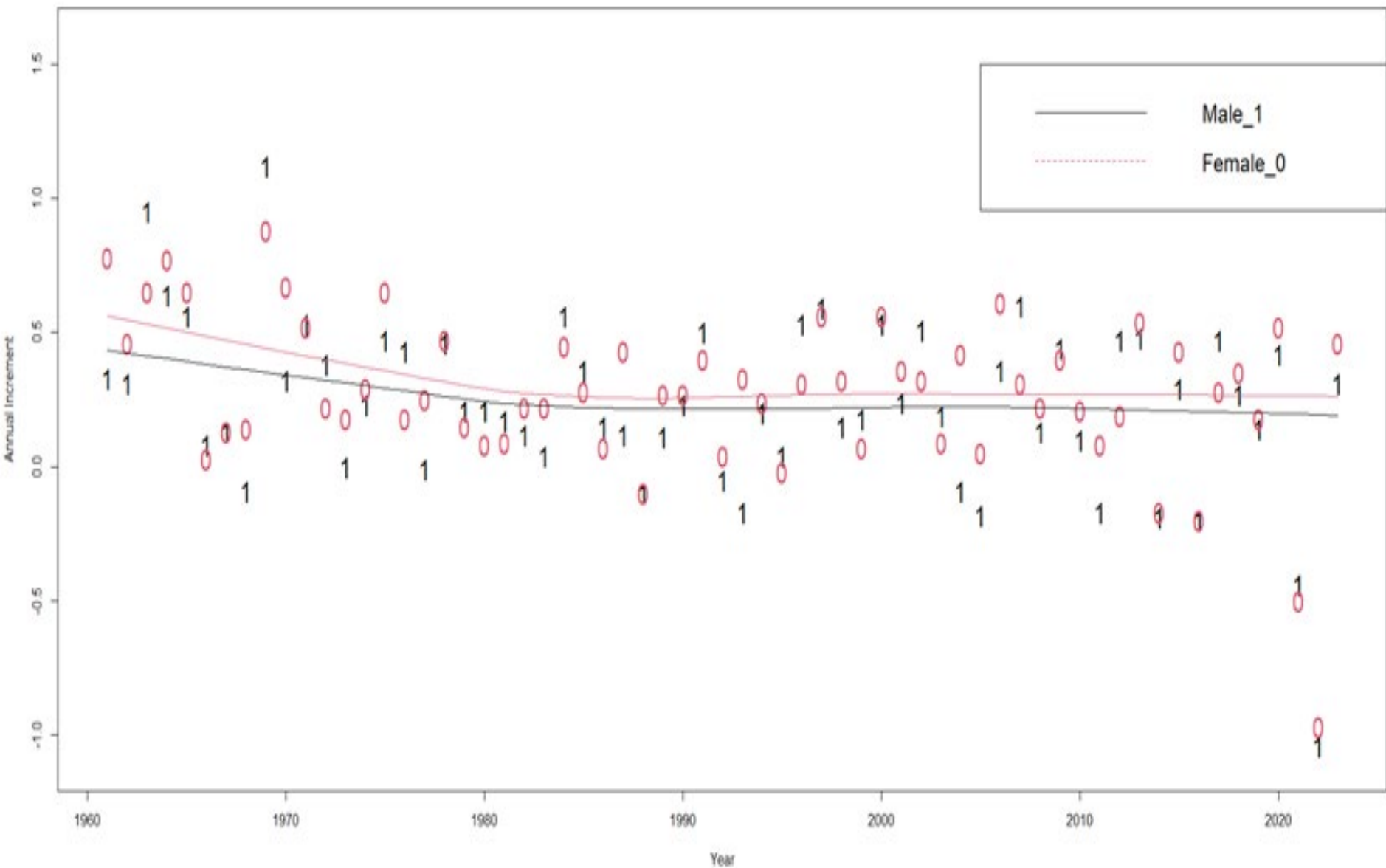
Data visualization



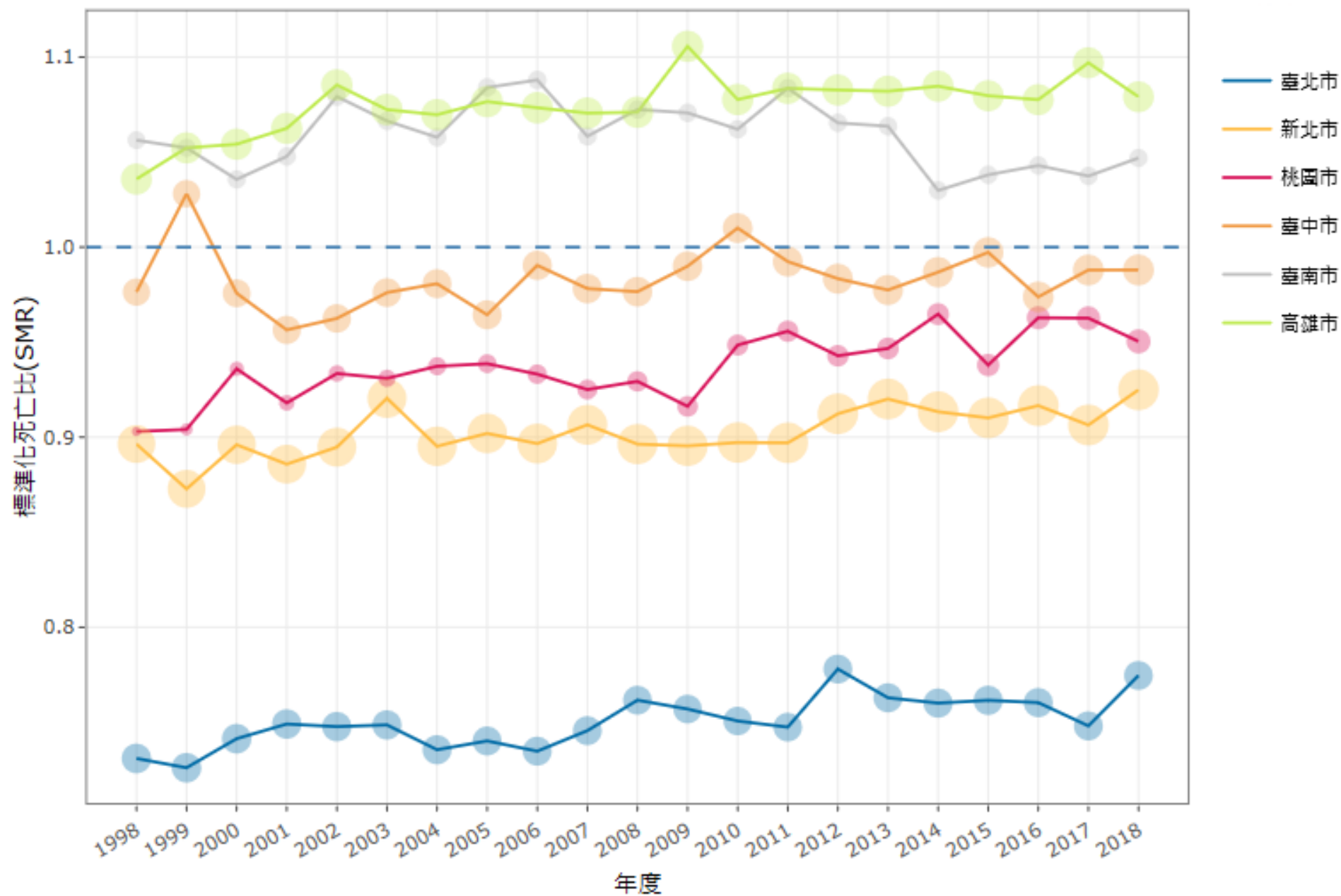
Data visualization is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewers of the images. This communication is achieved through the use of a systematic mapping between graphic marks and data values in the creation of the visualization. This mapping establishes how data values will be represented visually, determining how and to what extent a property of a graphic mark, such as size or color, will change to reflect changes in the value of a datum.

To communicate information clearly and efficiently, data visualization uses statistical graphics, plots, information graphics and other tools. Numerical data may be encoded using dots, lines, or bars, to visually communicate a quantitative message.^[1] Effective visualization helps users analyze and reason about data and evidence. It makes complex data more accessible, understandable and usable. Users may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic (i.e., showing comparisons or showing causality) follows the task. Tables are generally used where users will look up a specific measurement, while charts of various types are used to show patterns or relationships in the data for one or more variables.

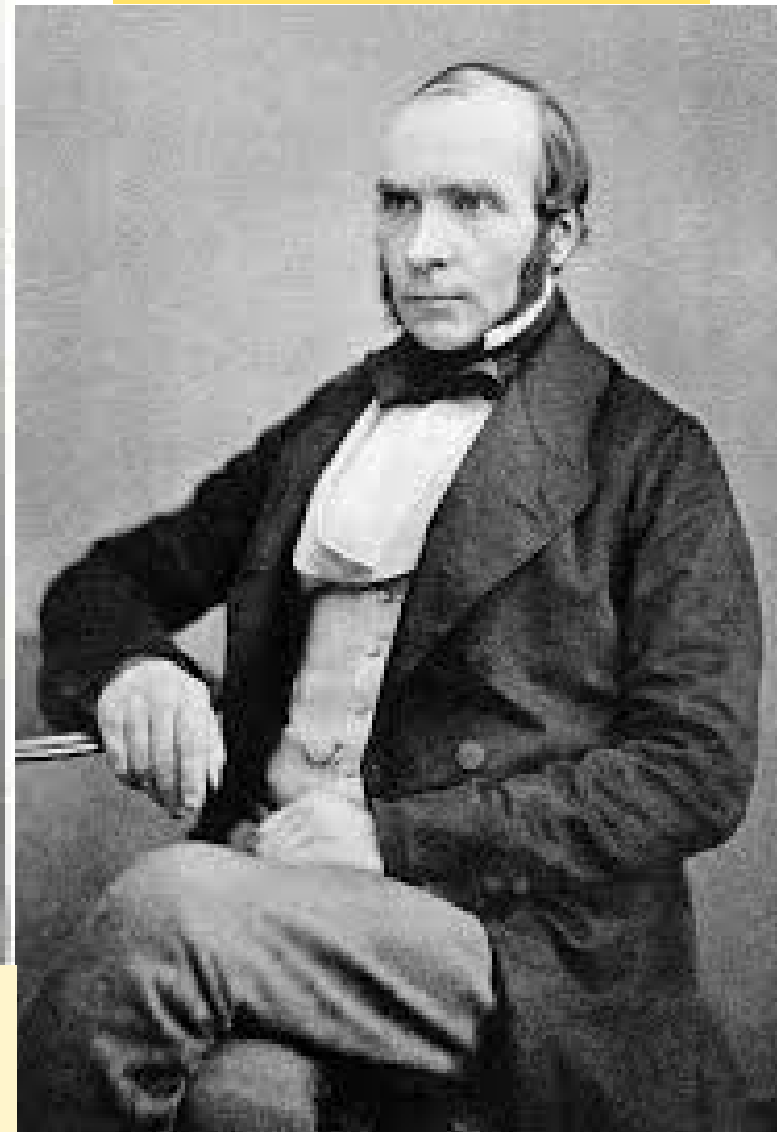
臺灣男女性平均壽命延長趨勢



六都1998-2018年男性標準死亡比



John Snow & Spatial Analysis



<https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcTUhpPXin7G9FYiZDkwcle2ZV1YIoqMC Ern2g&usqp=CAU>

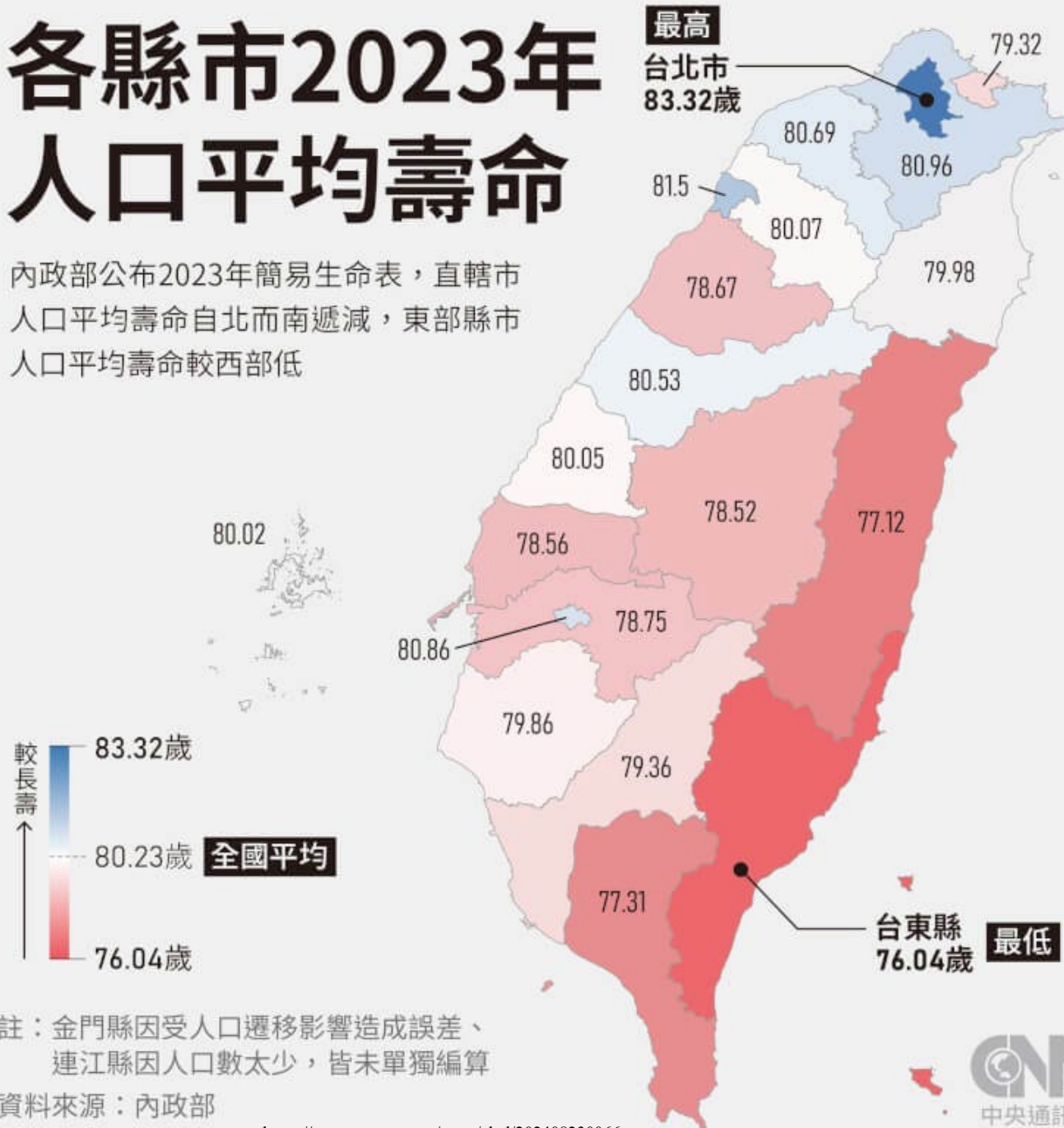
Charles Picquet

巴黎48個行政區的霍亂熱區圖

https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSh_go8YV CyyZvQsr7UBFWtJ66b1GW2cN0RIg&usqp=CAU

各縣市2023年人口平均壽命

內政部公布2023年簡易生命表，直轄市人口平均壽命自北而南遞減，東部縣市人口平均壽命較西部低



註：金門縣因受人口遷移影響造成誤差、連江縣因人口數太少，皆未單獨編算
資料來源：內政部

<https://www.cna.com.tw/news/ahel/202408230066.aspx>

112年我國10大死因

1 癌症	6 COVID-19
2 心臟疾病	7 高血壓性疾病
3 肺炎	8 事故傷害
4 腦血管疾病	9 慢性下呼吸道疾病
5 糖尿病	10 腎炎、腎病症候群及腎病變

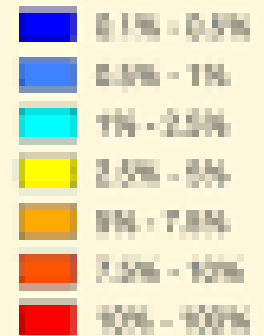
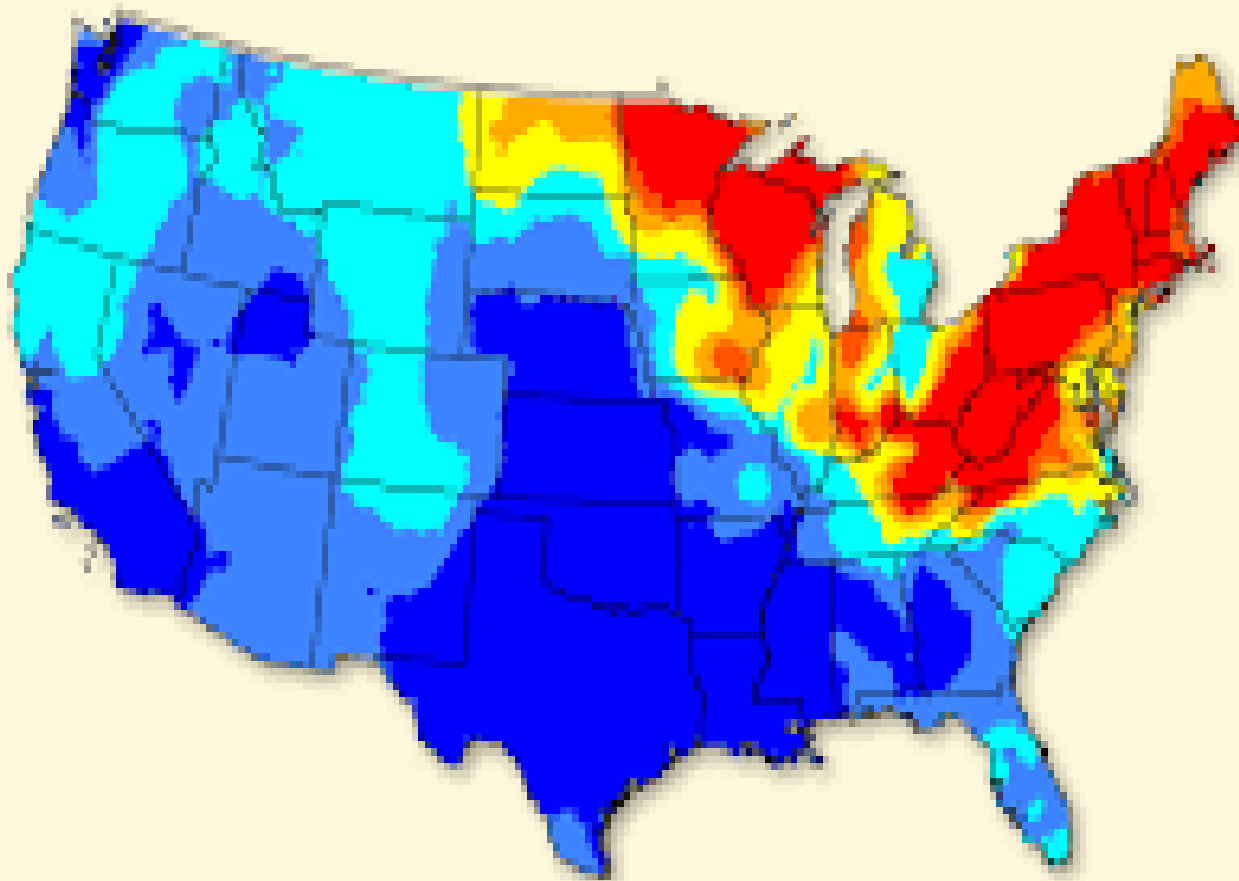
資料來源：電報部

<https://news.pts.org.tw/article/700413>

美國2024年萊姆病預測(Lyme disease, 壁蝨)

Forecasted Prevalence of Lyme Disease

2024



什麼是大數據？



http://www.abc.net.au/reslib/201204/r926508_9670000.jpg

什麼是大數據？

■ 大數據2010年由IBM所提出，涵蓋四個V：

→ 大量化(Volume)：至少TB及PB以上

→ 快速化(Velocity)：即時處理

→ 多樣化(Variety)：視頻、GIS等多樣性

→ 真實性(Veracity)：資料品質（2014年提出）

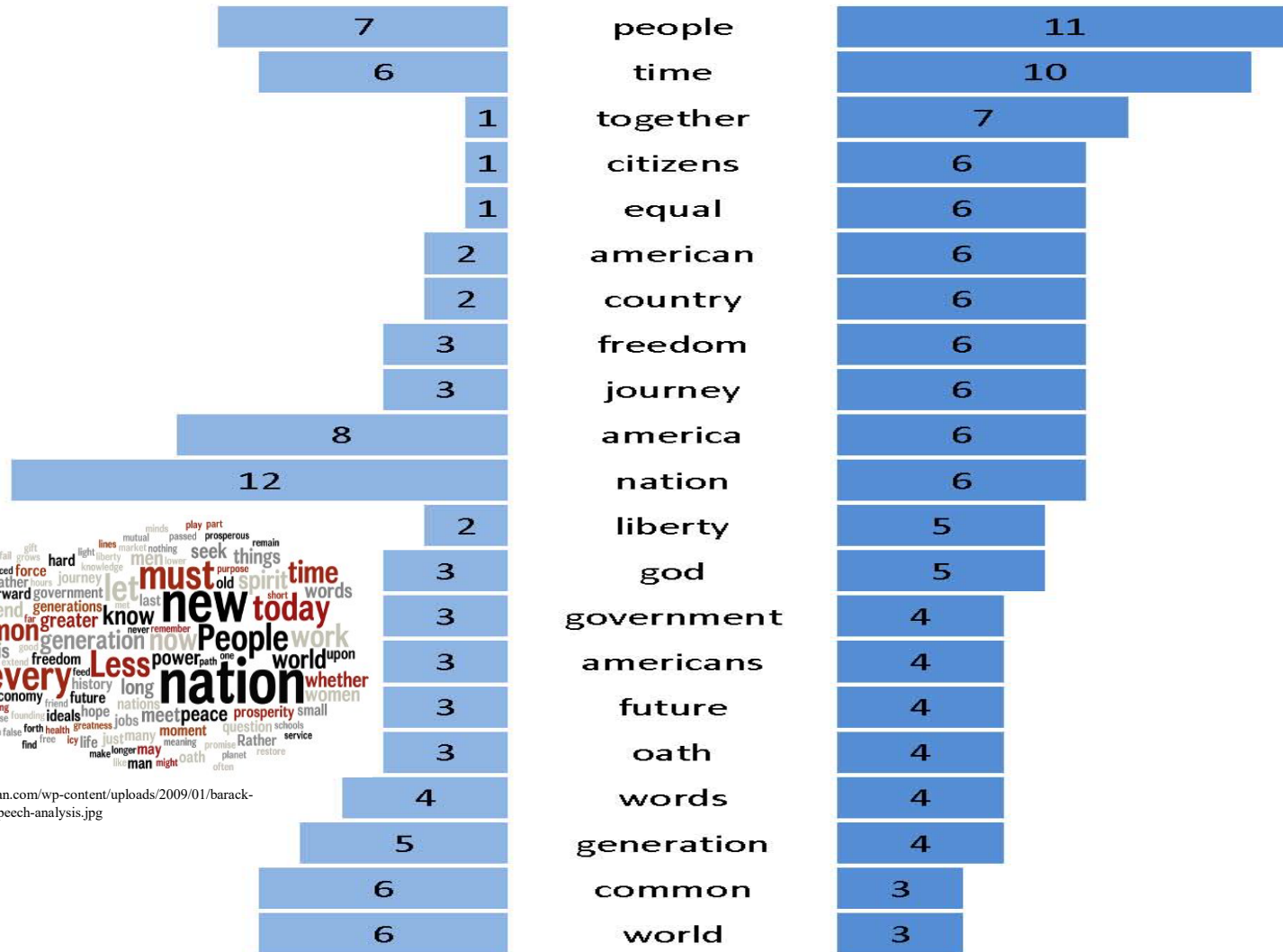
註：加入資料分享與傳遞(Visible)後成為「大、快、雜、疑、轉」；也有人加入價值(Value)成為新的5V。

電腦容量單位的演變（資料爆炸！）

單位	縮寫	意義
Bit	b	1 or 0
Byte	B	8 Bits
Kilobyte	KB	1,024 Bytes
Megabyte	MB	1,024 KB
Gigabyte	GB	1,024 MB
Terabyte	TB	1,024 GB
Petabyte	PB	1,024 TB
Exabyte	EB	1,024 PB
Zettabyte	ZB	1,024 EB
Yottabyte	YB	1,024 ZB

Comparing Inaugural Addresses

■ 2009 ■ 2013



<http://sixminutes.dlugan.com/wp-content/uploads/2009/01/barack-obama-inauguration-speech-analysis.jpg>

Analyzing the speech of President Obama (**Textmining**)

第14任蔡英文總統就職演講稿最常出現字詞

排序	單字			雙字詞		
	類別	次數	頻率	類別	次數	頻率
1	的	293	5.48%	我們	86	2.012%
2	我	114	2.13%	台灣	41	0.959%
3	們	90	1.68%	政府	37	0.866%
4	一	75	1.40%	國家	32	0.749%
5	會	74	1.38%	一個	29	0.679%
6	是	70	1.31%	新政	27	0.632%
7	個	66	1,23%	經濟	27	0.632%
8	民	63	1.18%	這個	25	0.585%
9	人	59	1.10%	民主	24	0.562%
10	國	59	1.10%	社會	22	0.515%

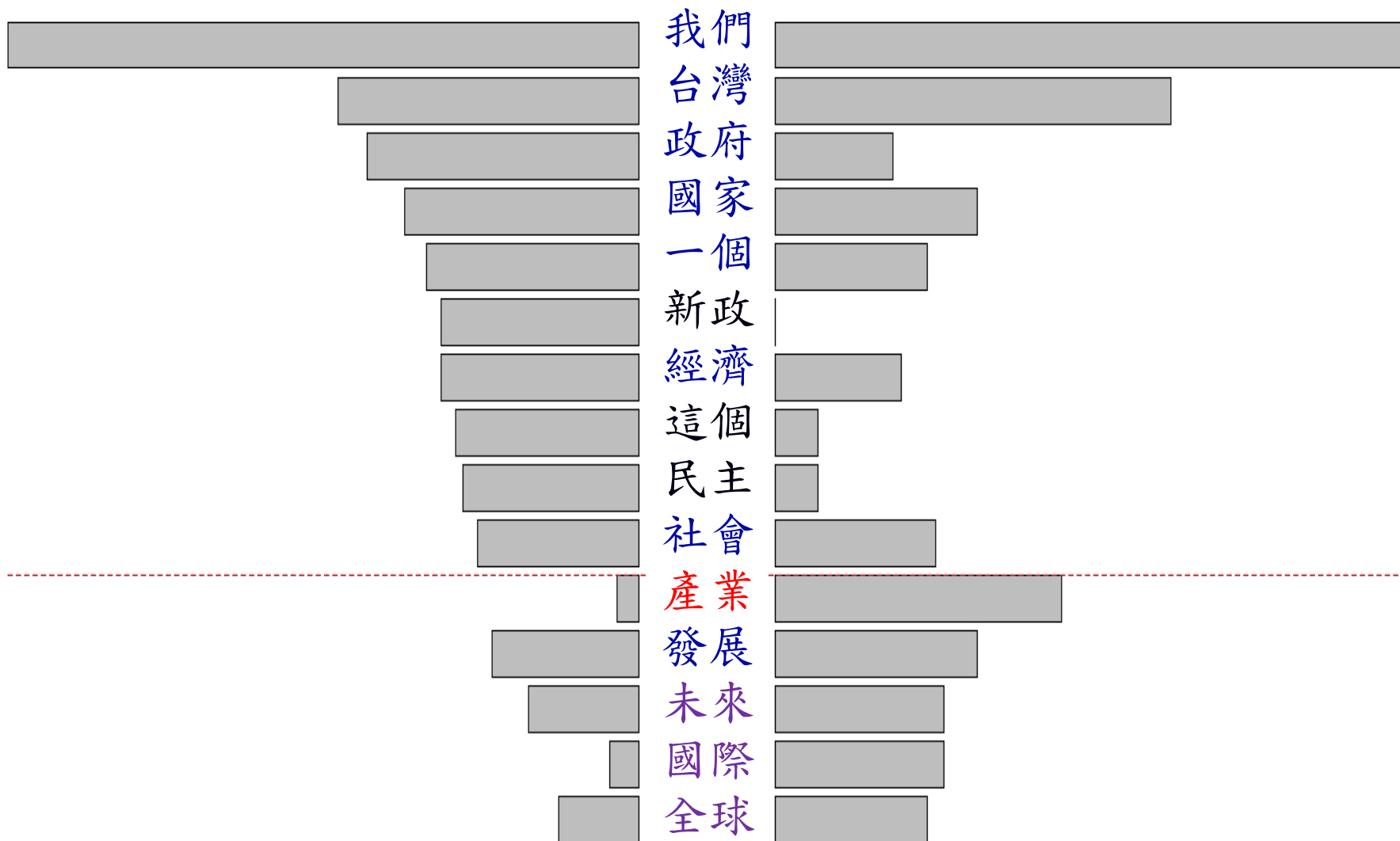
第15任蔡英文總統就職演講稿最常出現字詞

排序	單字			雙字詞		
	類別	次數	頻率	類別	次數	頻率
1	的	257	4.94%	我們	75	2.590%
2	我	114	2.10%	台灣	47	1.620%
3	們	92	1.77%	產業	34	1.170%
4	國	79	1.52%	國家	24	0.830%
5	人	68	1.31%	發展	24	0.830%
6	會	65	1.25%	未來	20	0.690%
7	在	63	1,21%	國際	20	0.690%
8	一	62	1.19%	社會	19	0.660%
9	是	55	1.06%	全球	18	0.620%
10	要	53	1.02%	一個	18	0.620%

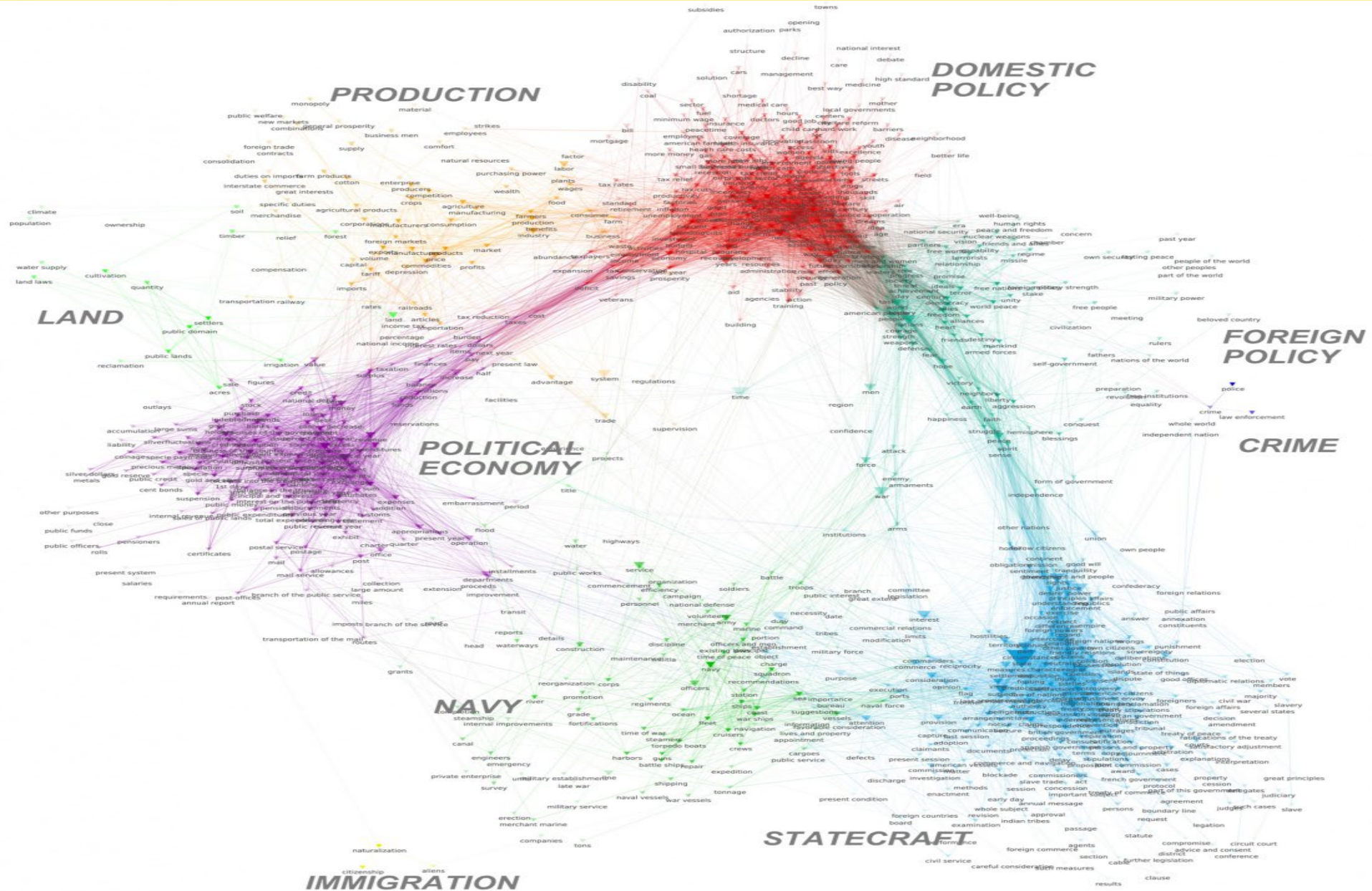
蔡英文總統就職演講稿常見雙字詞

第14任

第15任



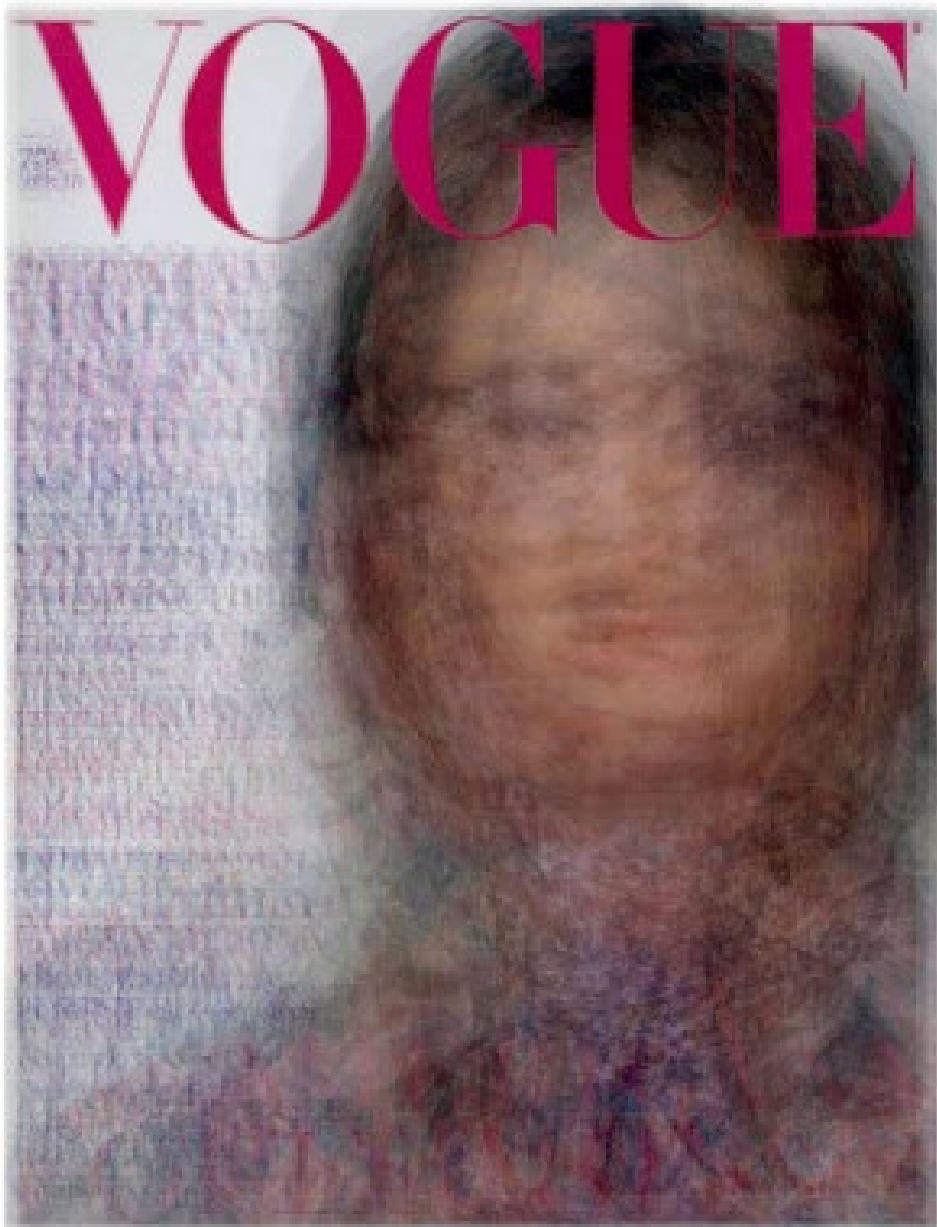
美國總統國情咨文(State of Union;1790年至今)



來源：<https://sciencenode.org/feature/text-mining-strikes-gold-in-political-discourse.php>

耶魯大學數位人文實驗室「Robots Reading Vogue」

1970



1980

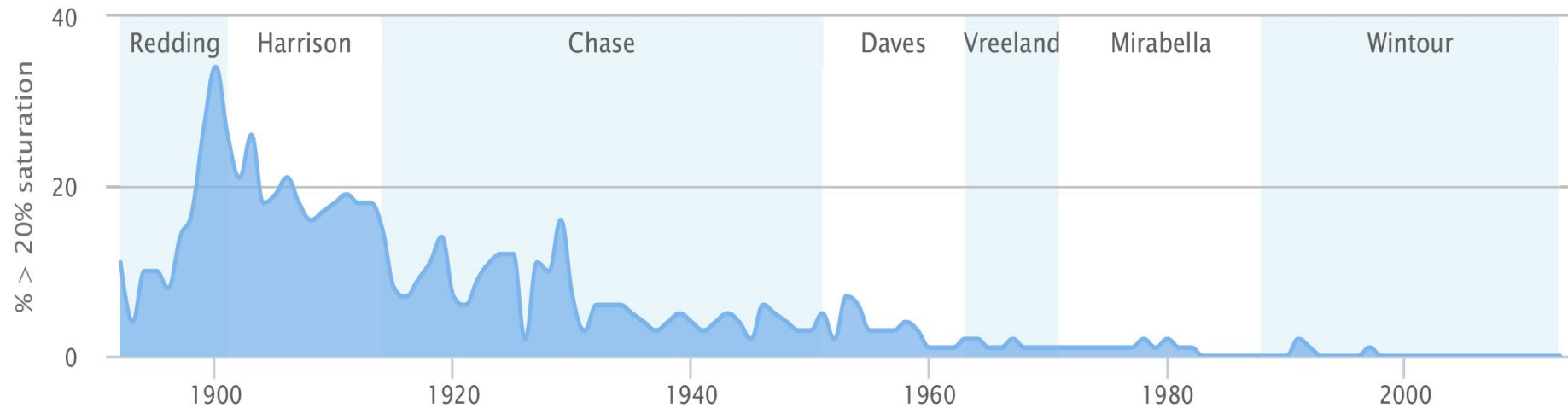




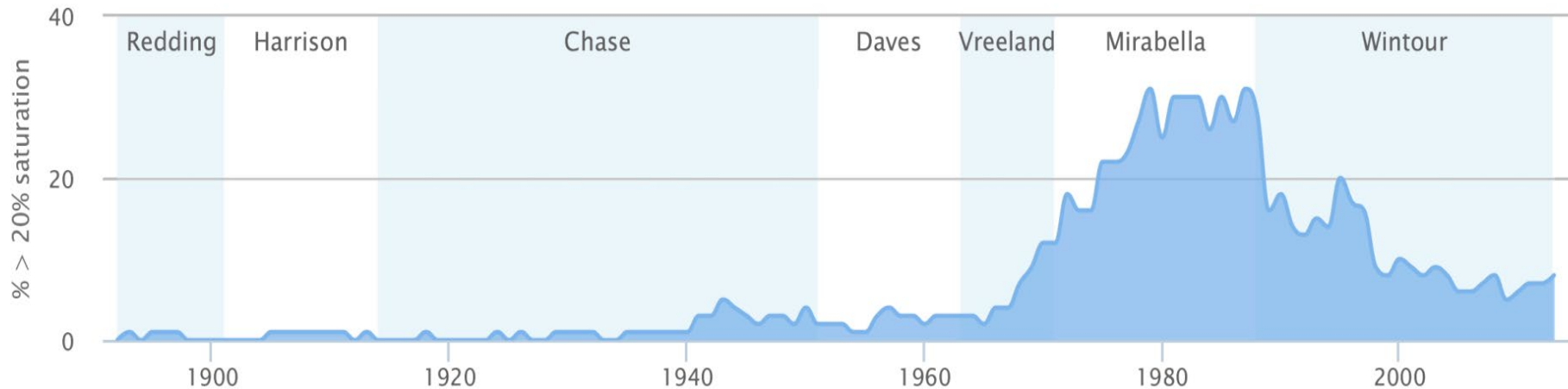
https://miro.medium.com/max/3778/1*zdoQ-oKnWAPBKbUMYYL--w.jpeg



Dressmaking over Time



Women's Health over Time



Vogue雜誌的風格趨勢變化



聊天機器人Chat GPT

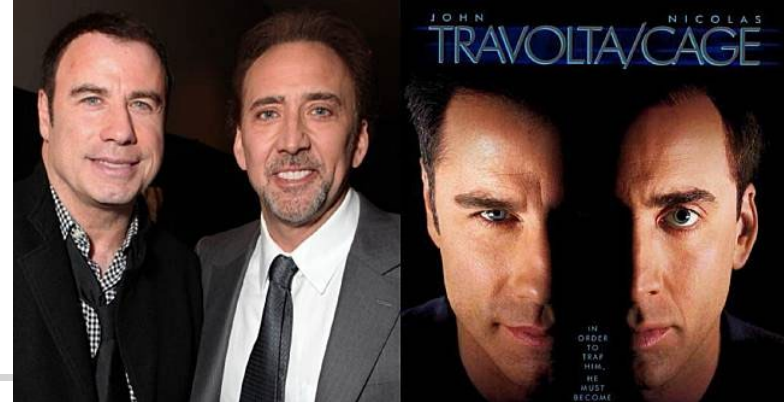
- 聊天機器人Chat GPT可生成類似人類風格的文字。(Generative Pre-Trained Transformer)
→ 「假作真時真亦假，無為有處有還無」！

商周

可以寫一段推薦《商業周刊》的話嗎？



當然可以。《商業周刊》是一份權威的商業新聞雜誌，提供最新的商業資訊和分析。它的內容涵蓋了金融、科技、消費品、能源和基礎設施等領域，為讀者提供全面而及時的商業資訊。此外，《商業周刊》還有一個強大的團隊，由一流的記者和專家組成，為讀者提供深入而獨特的視角。如果您對商業新聞感興趣，那麼《商業周刊》絕對值得您一試。



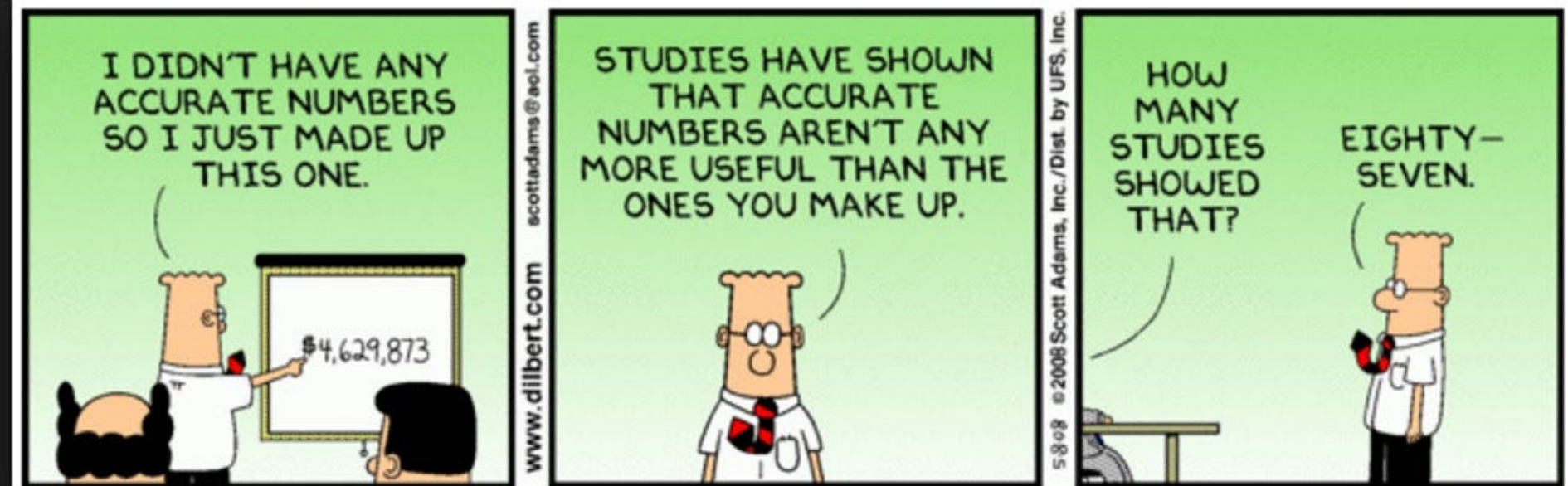
深偽技術Deepfake

- 深偽技術是「deep learning」和「fake」的混成詞，以深度學習的演算法，從現有影片資料製造出逼真的假影片。（假新聞！）



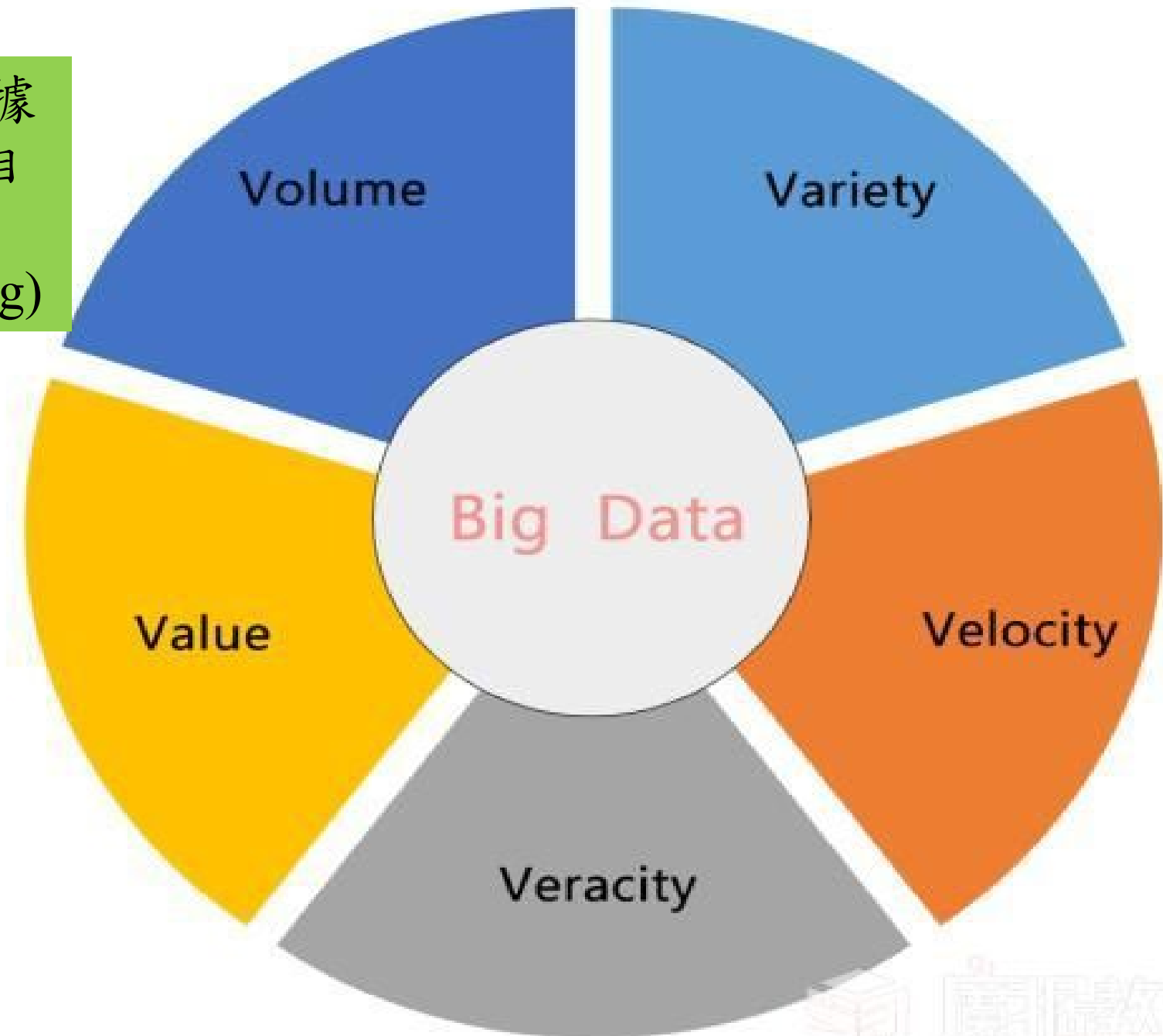


<https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.pinterest.com%2Fpin%2F167688786102533739%2F&psig=AOvVaw191BQGepIMqvj7BKEURwmp&ust=1612364346039000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCKCkqrm7y-4CFQAAAAAdAAAAABAD>



https://media-exp1.licdn.com/dms/image/C4E12AQEZaBITJLladA/article-cover_image-shrink_600_2000/0/1520178855524?e=1616025600&v=beta&t=VsCZNHDKnmoUy6sdD0t86DirLkdqM7rNLTGatwMZoSc

Value：數據
價值密度相
對較低。
(Datamining)



資訊的價值



- 1990年代後期Amazon網站雇用十幾位書評及編輯，提供推薦閱讀的書單，其銷售量卻比不上以讀者回饋產生的建議，最後解散書評團隊。（現在1/3業績來自於電腦推薦。）
- 2004美國Walmart開始運用歷史交易記錄，發現颶風來臨前，手電筒、小甜點Pop-Tarts銷售量大增。（註：「尿布與啤酒」是另一個知名範例！）

大數據的特性



大數據的幾個基本特性

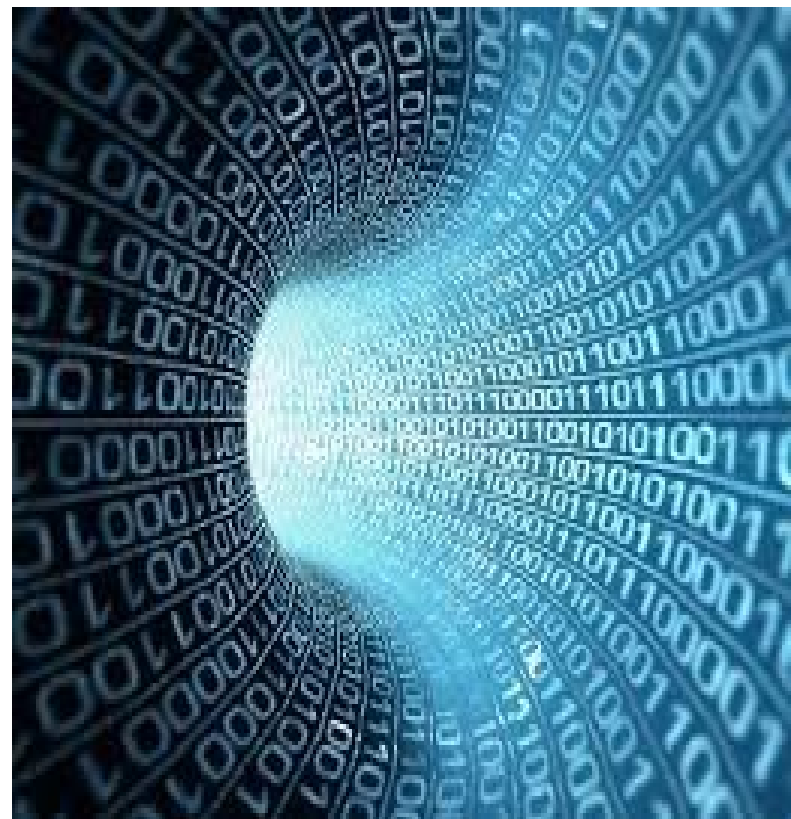
■ 除了數量龐大，有人認為大數據還有以下幾個特性：

→ 「樣本 = 母體」

→ 不精確且含有雜訊的資料

→ 相關性而非因果關係

註：你/妳認同哪些特性？



關連性vs.因果關係

- 數量分析無法決定因果關係，小學生腳丫大小與拼字能力，可能存在潛在變數(Latent Variable)。
- 巨量資料及電腦科技的進步，使得尋找關連性更為可行。
 - 美國包裹快遞公司(UPS)從2000年開始實施預防性維修(Preventive Maintenance)；
 - 紐約市根據過去人孔蓋維修記錄，預測哪些地方有較高的損壞機會(2010 Wired)。

資料分享與傳遞（特性之一）

JAN
2024

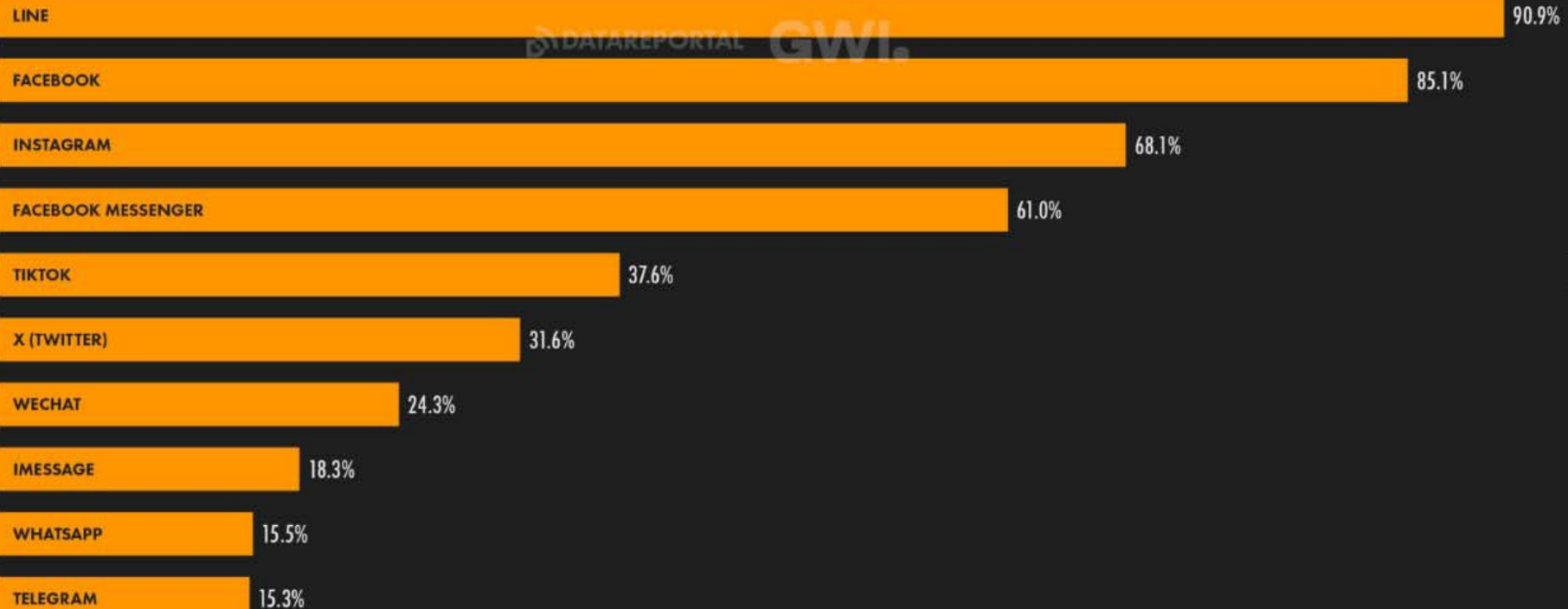
MOST USED SOCIAL MEDIA PLATFORMS

PERCENTAGE OF INTERNET USERS AGED 16 TO 64 WHO USE EACH PLATFORM EACH MONTH

NOTE: YOUTUBE IS NOT OFFERED AS AN ANSWER OPTION FOR THIS QUESTION IN GWI'S SURVEY, SO IT WILL NOT APPEAR IN THIS RANKING



TAIWAN



SOURCE: GWI (Q3 2023). FIGURES REPRESENT THE FINDINGS OF A BROAD SURVEY OF INTERNET USERS AGED 16 TO 64. SEE [GWI.COM](https://www.gwi.com). **NOTE:** YOUTUBE IS **NOT** OFFERED AS AN ANSWER OPTION FOR THIS QUESTION IN GWI'S SURVEY. **COMPARABILITY:** A VERSION OF THIS CHART THAT APPEARED IN OUR PREVIOUS REPORTS WAS BASED ON A PREVIOUS QUESTION IN GWI'S SURVEY THAT INCLUDED YOUTUBE AS AN ANSWER OPTION. GWI'S CURRENT SURVEY FEATURES A REVISED VERSION OF THIS QUESTION THAT DOES **NOT** INCLUDE YOUTUBE AS AN ANSWER OPTION. WHILE OTHER CHANGES TO THE QUESTION'S WORDING MAY MEAN THAT THE VALUES AND RANK ORDER SHOWN HERE ARE **NOT DIRECTLY COMPARABLE** WITH THOSE SHOWN ON A SIMILAR CHART IN PREVIOUS REPORTS.

we
are
social

Meltwater

JAN
2023

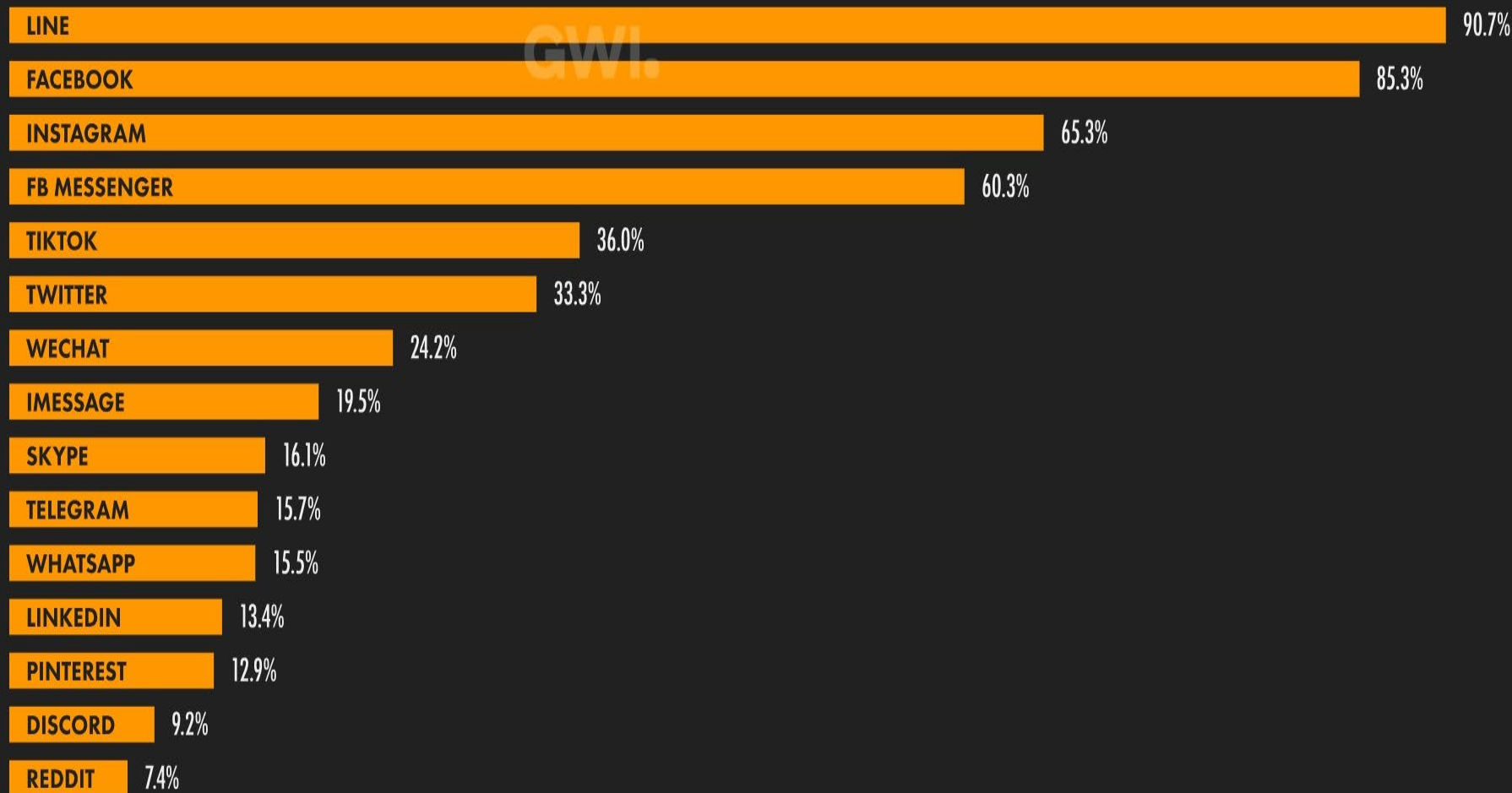
MOST USED SOCIAL MEDIA PLATFORMS

PERCENTAGE OF INTERNET USERS AGED 16 TO 64 WHO USE EACH PLATFORM EACH MONTH

NOTE: YOUTUBE IS NOT OFFERED AS AN ANSWER OPTION FOR THIS QUESTION IN GWI'S SURVEY, SO IT WILL NOT APPEAR IN THIS RANKING

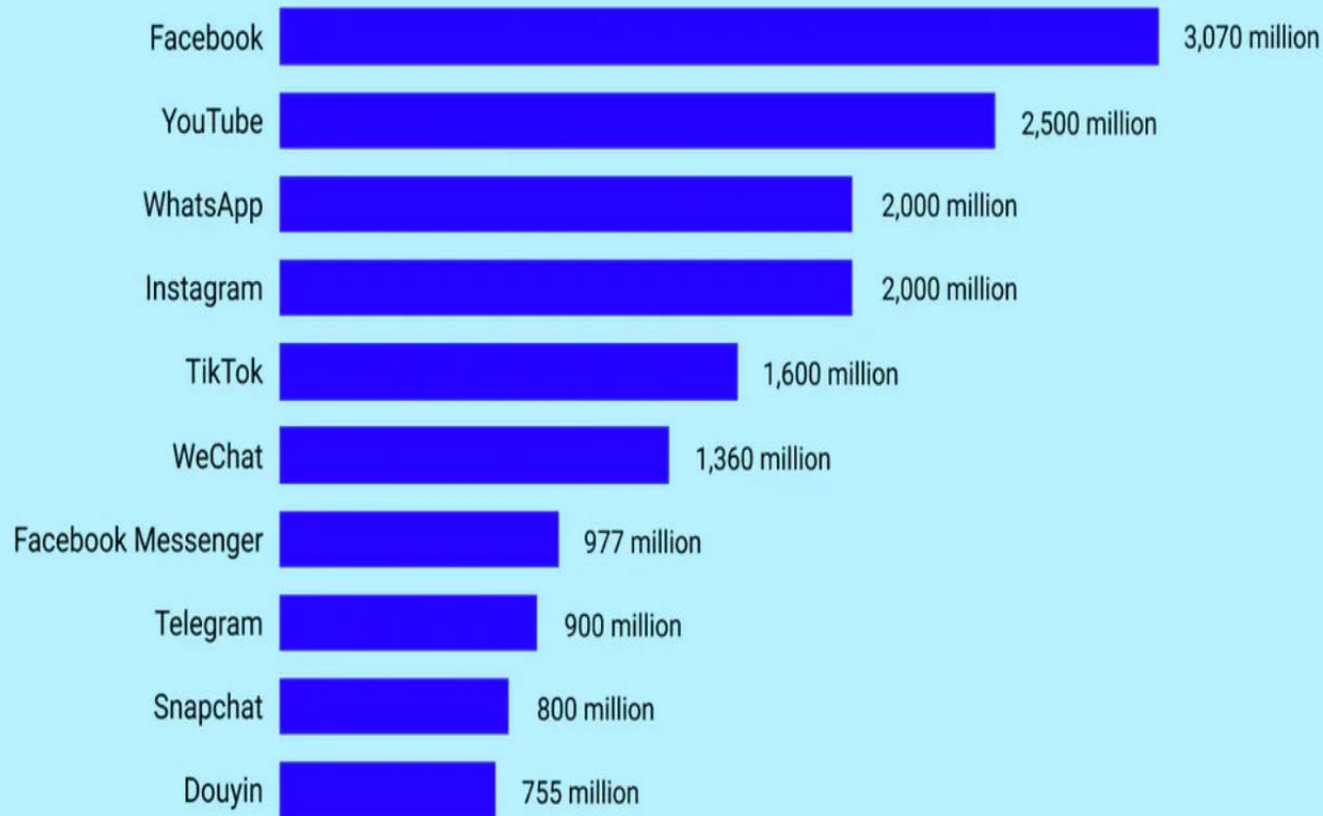


TAIWAN



全球最常使用社群軟體(2024年10月)

Most Popular Social Media Platforms in 2024



Source: DataReportal



<https://www.shopify.com/zh/blog/most-popular-social-media-platforms>

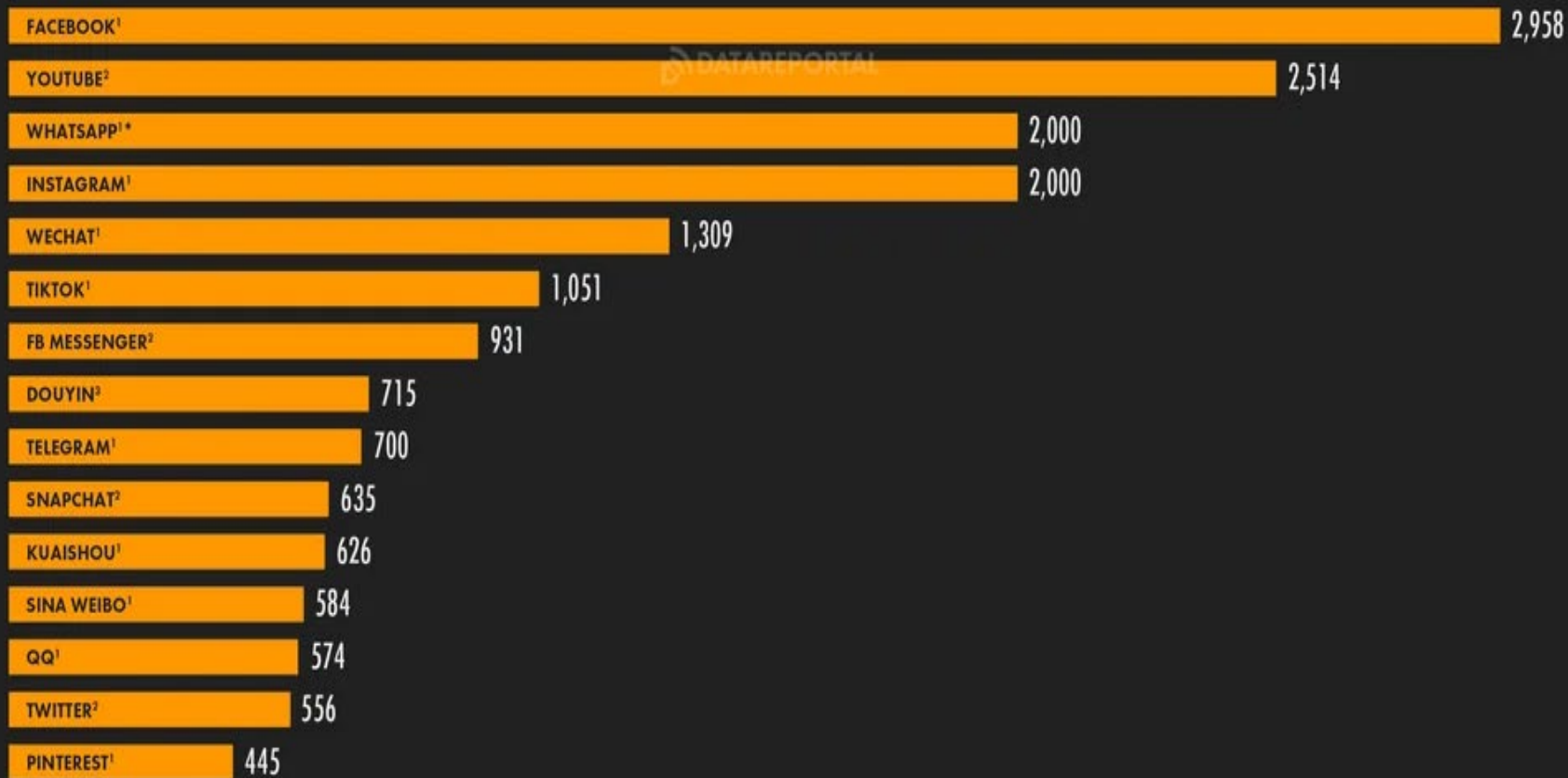
JAN
2023

THE WORLD'S MOST USED SOCIAL PLATFORMS

RANKING OF SOCIAL MEDIA PLATFORMS BY GLOBAL ACTIVE USER FIGURES (IN MILLIONS)



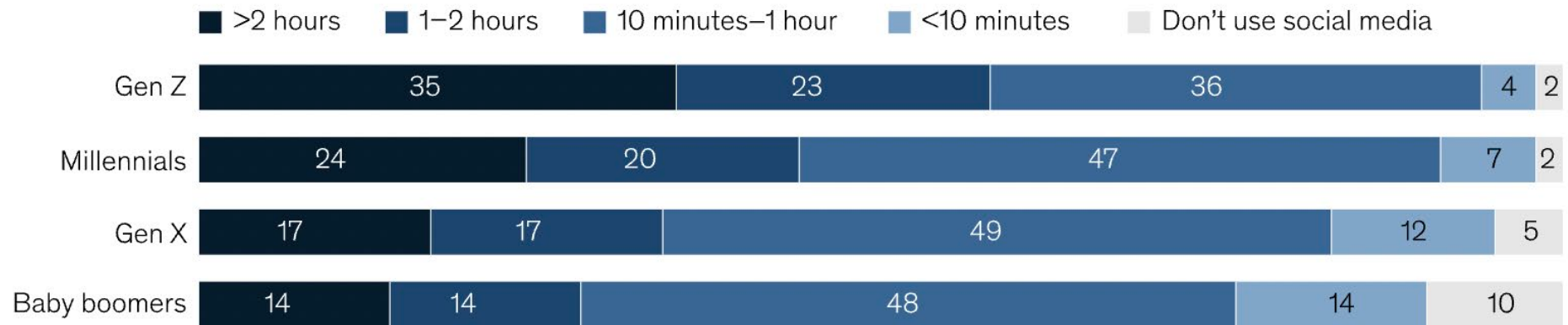
GLOBAL OVERVIEW



年齡與社群軟體使用時間(2024)

Time spent on social media daily,¹ % of respondents (n = 41,960)

<https://sproutsocial.com/insights/new-social-media-demographics/>

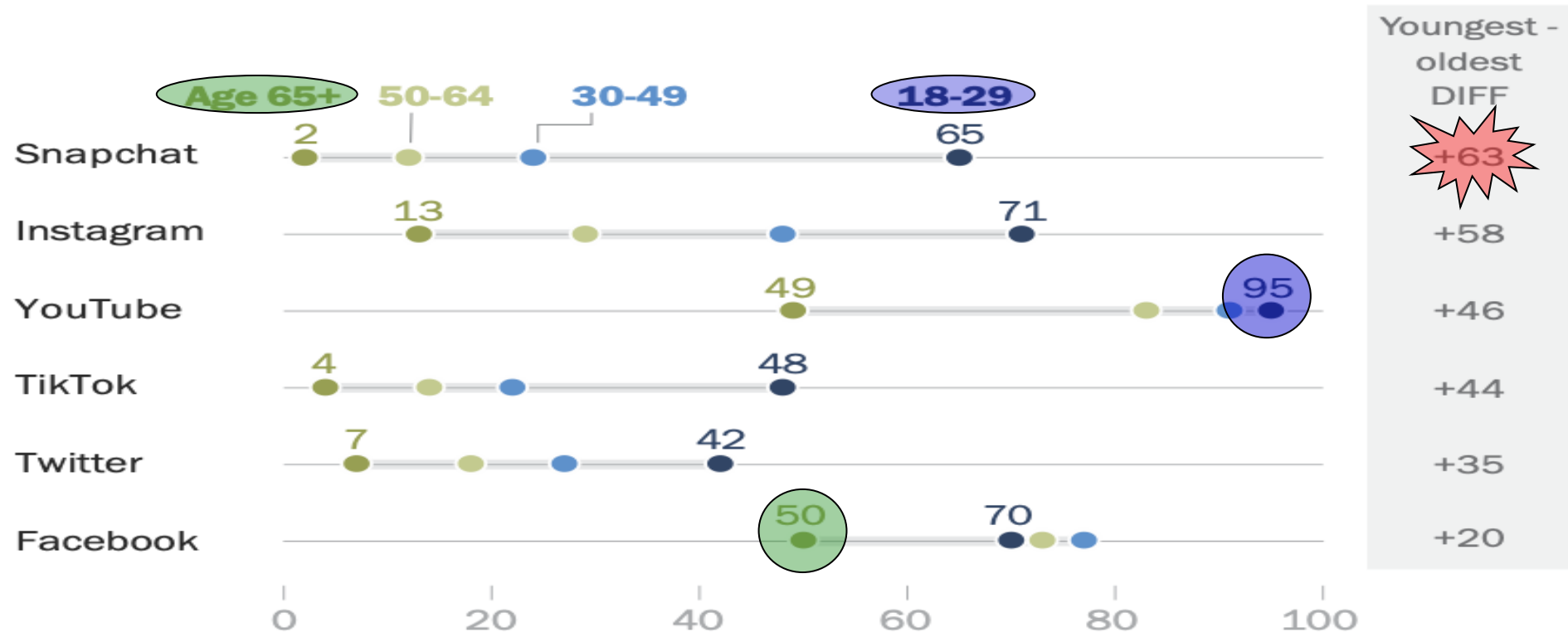


- 18-29 years – Snapchat (41%), TikTok (35%), Instagram (32%)
- 30-39 years – LinkedIn (34%), X/Twitter (34%), Snapchat (33%), Instagram (32%)
- 40-49 years – LinkedIn (25%), Facebook (22%), X/Twitter (21%)
- 50-59 years – Facebook (29%), LinkedIn (24%), Pinterest (24%)

社群軟體也有代溝？(2021)

Age gaps in Snapchat, Instagram use are particularly wide, less so for Facebook

% of U.S. adults in each age group who say they ever use ...



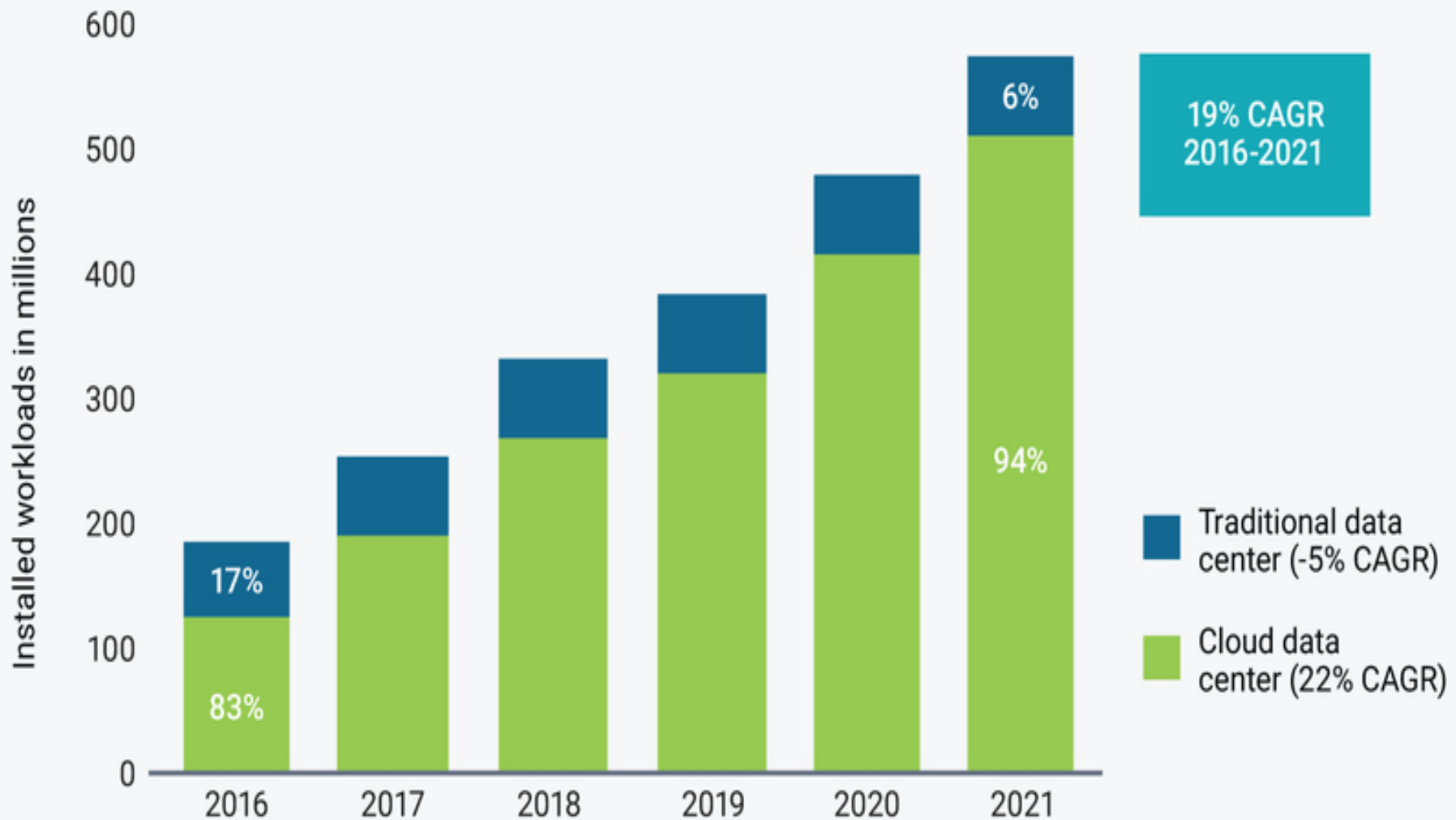
Note: All differences shown in DIFF column are statistically significant. The DIFF values shown are based on subtracting the rounded values in the chart. Respondents who did not give an answer are not shown.

Source: Survey of U.S. adults conducted Jan. 25-Feb. 8, 2021.

"Social Media Use in 2021"

PEW RESEARCH CENTER

雲端儲存資料更為普遍



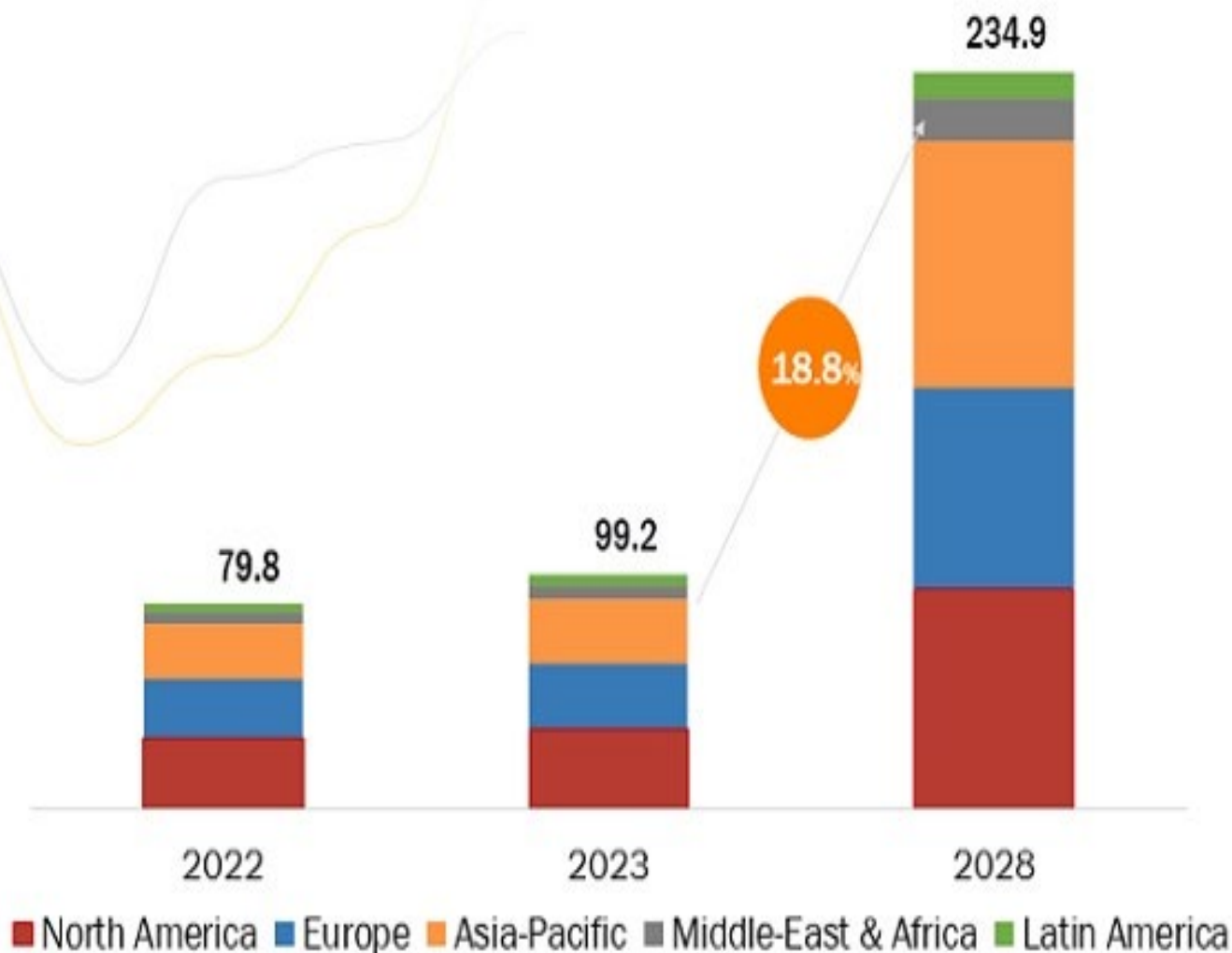
Source: Cisco Global Cloud Index, 2016-2021.

CLOUD STORAGE MARKET GLOBAL FORECAST TO 2028 (USD BN)



CAGR OF
18.8%

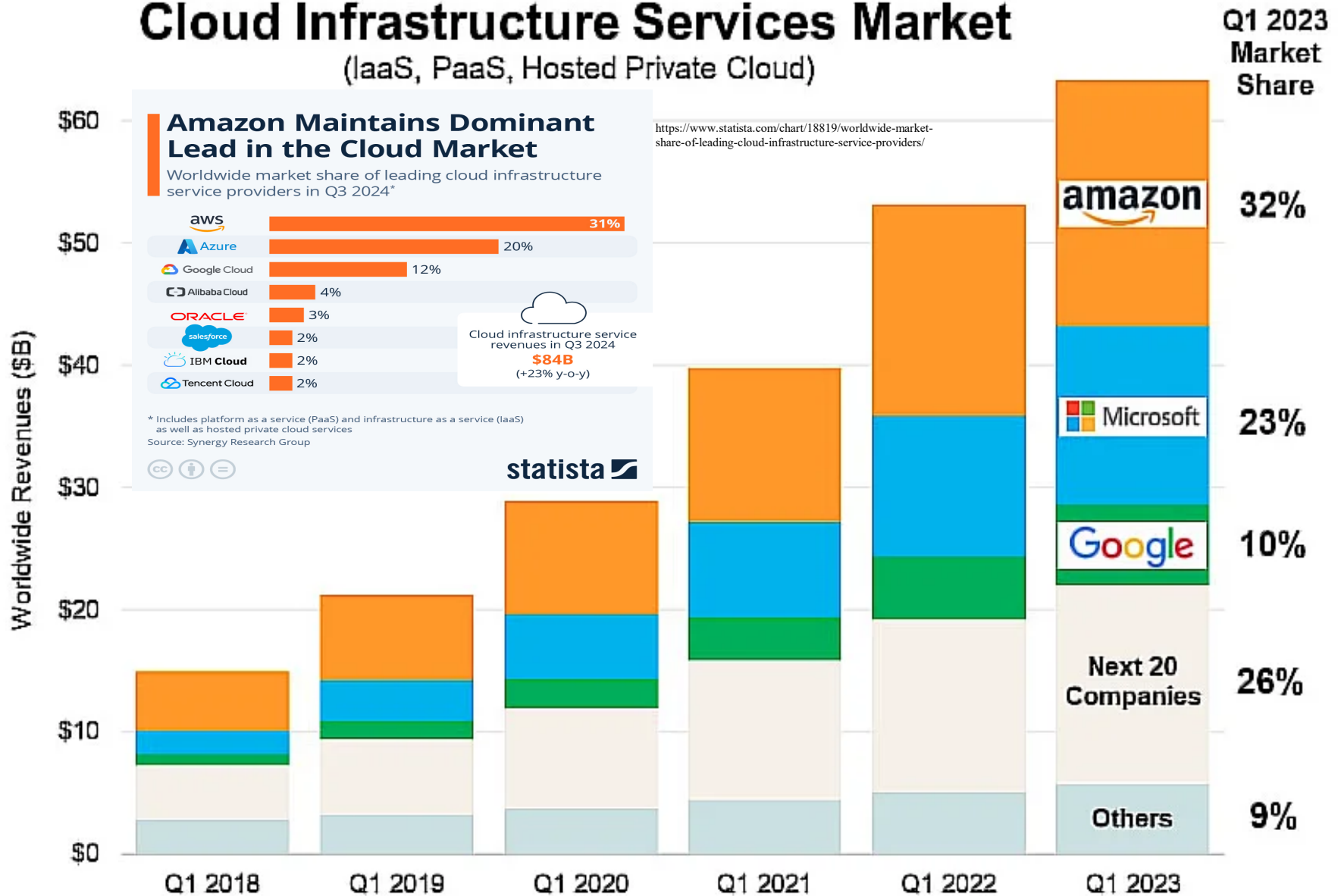
The global Cloud Storage market is expected to be worth USD 234.9 billion by 2028, growing at a CAGR of 18.8% during the forecast period.



雲端基礎設施服務的市場

Cloud Infrastructure Services Market

(IaaS, PaaS, Hosted Private Cloud)

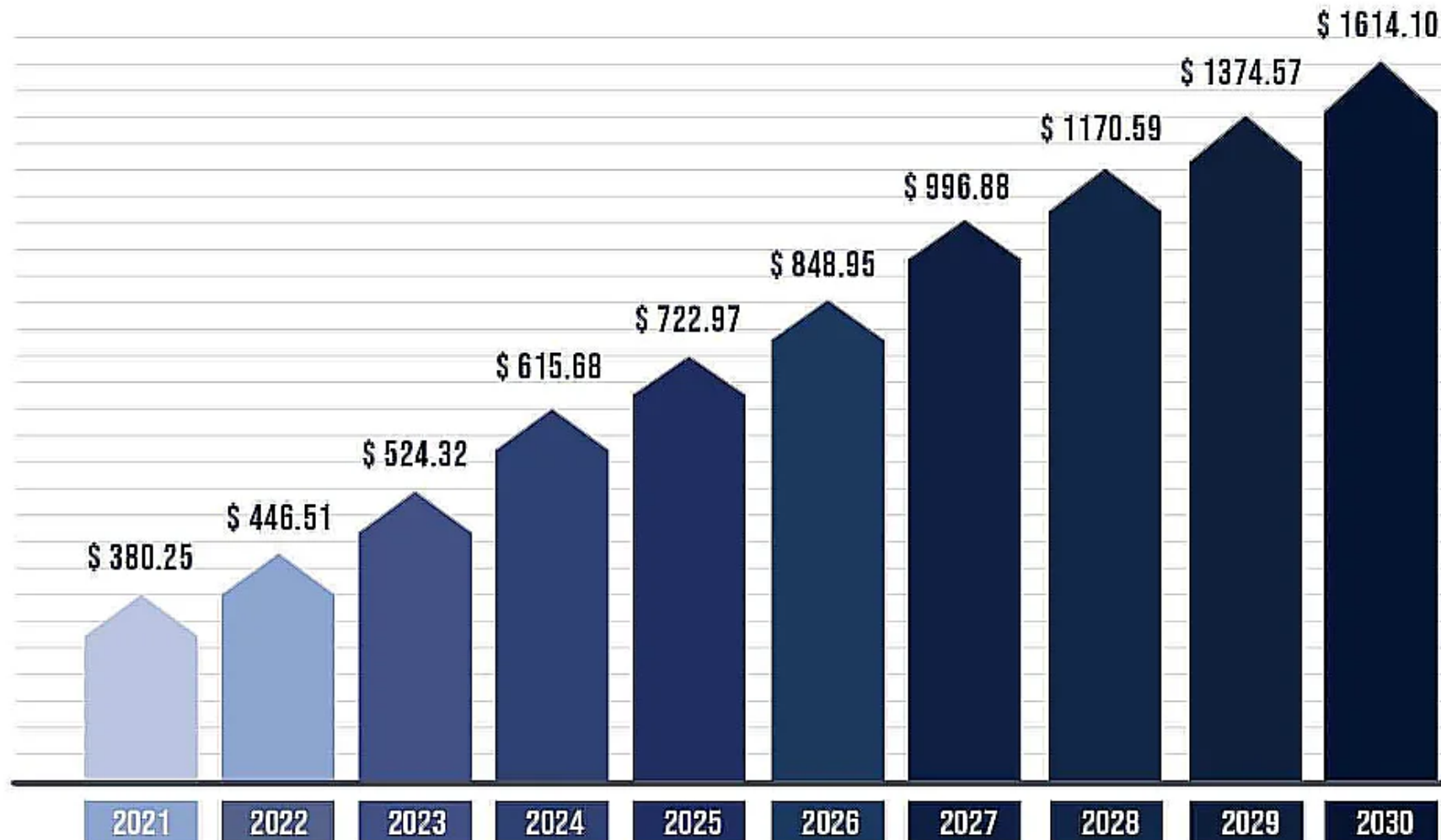


Source: Synergy Research Group

雲端計算市場發展迅速(98%企業使用)



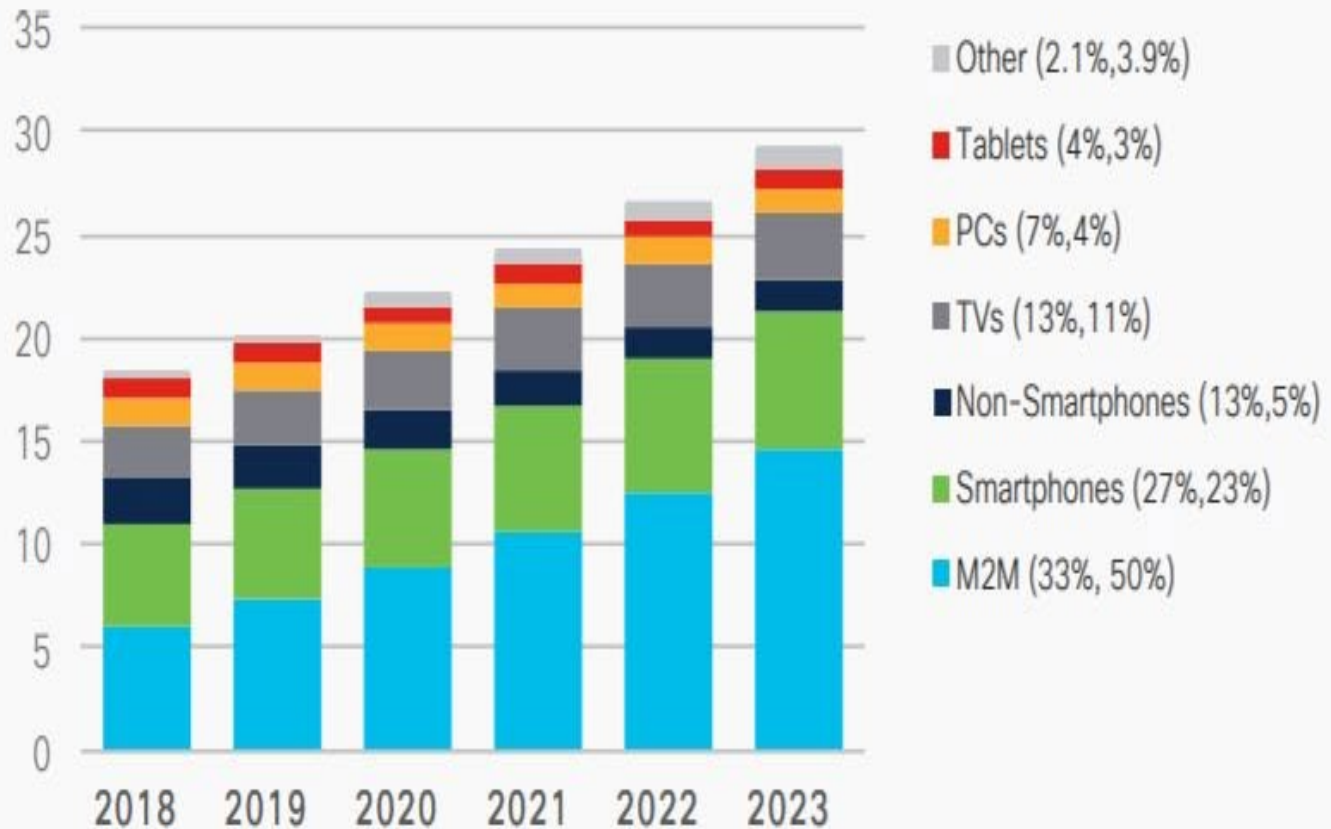
CLOUD COMPUTING MARKET SIZE, 2021 TO 2030 (USD BILLION)



未來趨勢：機器對機器。。

10% CAGR
2018-2023

Billions of
Devices



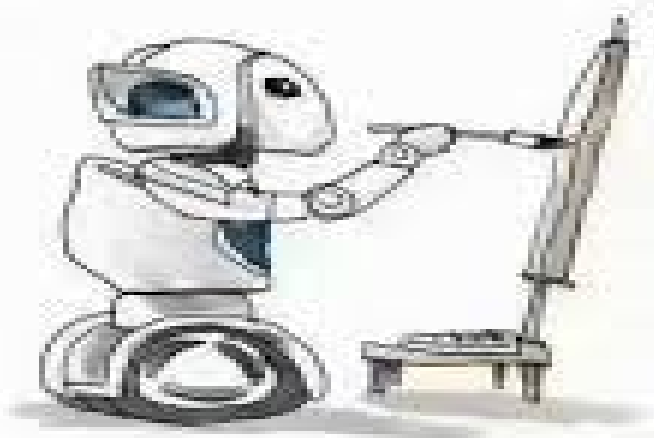
* Figures (n) refer to 2018, 2023 device share

Source: Cisco Annual Internet Report, 2018-2023

<https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>

機器對機器 (Machine to machine, M2M)

機器裝置間在無需人為干預下，直接透過網路溝通而自行完成任務。



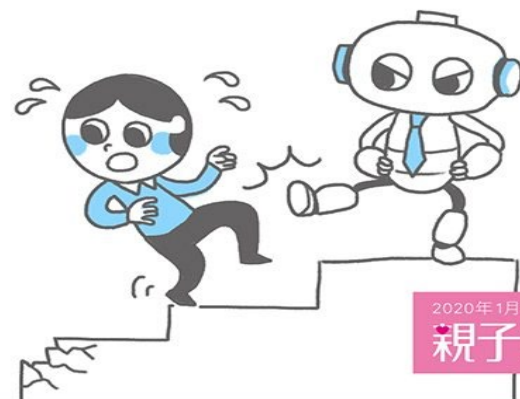
人工智能方向是不是搞錯了
本來想是AI幫我們洗碗掃地
我們就有空去寫詩画画
現在AI開始寫詩画画了
而我們還在洗碗掃地

https://www.popo8.com/host/data/202402/20/18/p1708479682_448type_jpeg_size_660_85_end.jpg_b.jpg

<https://www.parenting.com.tw/article/5080979>

⑦ AI 帶來很多挑戰和反思

許多人擔心AI什麼都會，
下棋贏過世界棋王、會寫詩、
寫新聞、挑咖啡豆。
那以後是不是工作都會被取代？
人還有什麼價值呢？



2020年1月號

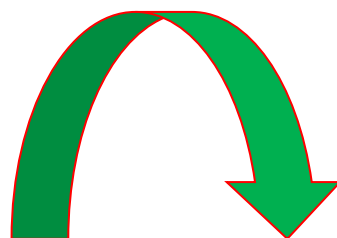
親子天下

科技來自於或是改變人性？

- 「科技始終來自於人性」是Motorola知名的廣告用詞，強調其產品根據人性因素設計。
- 科技也改變了我們的生活，但是否會改變人們的思維模式、甚至人性？



NOKIA
Connecting People



https://scontent-lga3-1.cdninstagram.com/v/t51.2885-15/sh0.08/e35/s640x640/70025765_480213932560518_6084948241454488487_n.jpg?_nc_ht=scontent-lga3-1.cdninstagram.com&_nc_cat=106&_nc_ohc=nGAHr_5D4wEAX-ZkqE8&oh=c1b19989c814908e98200e4b0759d6c5&oe=5EA54259



資訊解讀的價值

- Data Mining
- Big Data
- Predictive Analytics
- Artificial Intelligence
- Machine Learning

資訊與知識的價值(資料採礦)

- 資料挖掘(Data Mining)的範例：\$ \$ \$ \$
→ 協助超級市場促銷及陳設商品。

Milk, eggs, sugar, bread



Customer1

Milk, eggs, cereal, bread



Customer2

Eggs, sugar



Customer3

數據分析的嚆頭！？



沃馬特量販店(Wal-Mart)

- 沃馬特最先蒐集、分析顧客資料，並以整理所得的資訊，提高銷售業績。
 - 分析發現美國消費者在週末購物時，許多人會同時購買尿布及啤酒。
 - 問題：為什麼這兩種商品會一起購買？又如何將這份資訊轉變為業績？
- 註：沃馬特從美國西南部發跡，剛開始只是一家五金行，現在已是全美最大的百貨零售業者。

「尿布與啤酒」的延伸價值

- 尿布與啤酒屬於關連性(Association)的關係，與常見的因果關係(Causality)不同。
- 關連性的價值未必低於因果關係，像是尿布與啤酒的關連，可用於：
 - 商品定價與促銷；
 - 商品擺設（商場動線）；
 - 商品倉儲。





賣場如何應用資訊？

■ 商品定價與促銷

- 尿布及啤酒的定價（如：打折）
- 如何促銷尿布及啤酒以外的商品

■ 商品擺設（商場動線）

- 尿布、啤酒兩種商品的相對位置
- 熱銷商品如何擺放

■ 商品倉儲

- 何時進貨（購買資訊？）

寶可夢旋風也能帶來商機嗎？



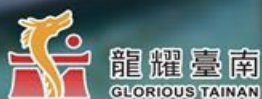
參考資料：<https://tw.news.yahoo.com/%E5%8F%B0%E5%8D%97%E5%A1%9E%E7%88%86-%E5%B0%8F%E9%BB%83%E7%B4%99%E6%A2%9D%E6%9B%9D-%E9%BE%9C%E9%80%9F%E5%8E%9F%E5%9B%A0-%E7%B6%B2%E7%AC%91-%E4%B8%8D%E5%8F%AD%E4%BD%A0%E4%BA%86-143342818.html>

Pokémon GO Safari Zone in Tainan (寶可夢台南狩獵區)：估計主場都會公園奇美博物館有8萬人，大台南全區16萬人，連續五天活動總計主場有56萬人次，台南全區100萬人次。 六億商機！！
今日新聞NOWnews 記者陳聖璋 2018年11月5日





2024 年 3 月 9-10 日
City Safari 台南



北門區域



鹽水月津港區域



中西區



安平燈區



北區及東區



台南高鐵站



高鐵燈區

原街役場



新化區



戴著狩獵帽的伊布

2024年3月9日及3月10日
Pokémon GO
City Safari：台南



https://www.youtube.com/watch?app=desktop&v=wPQxZq2s0oE&ab_channel=%E5%8F%B0%E8%A6%96%E6%96%B0%E8%81%9ETTNEWS

台南週末訂房滿 網哀嚎沒飯店

15萬人湧入抓寶！

台南寶可夢大爆發

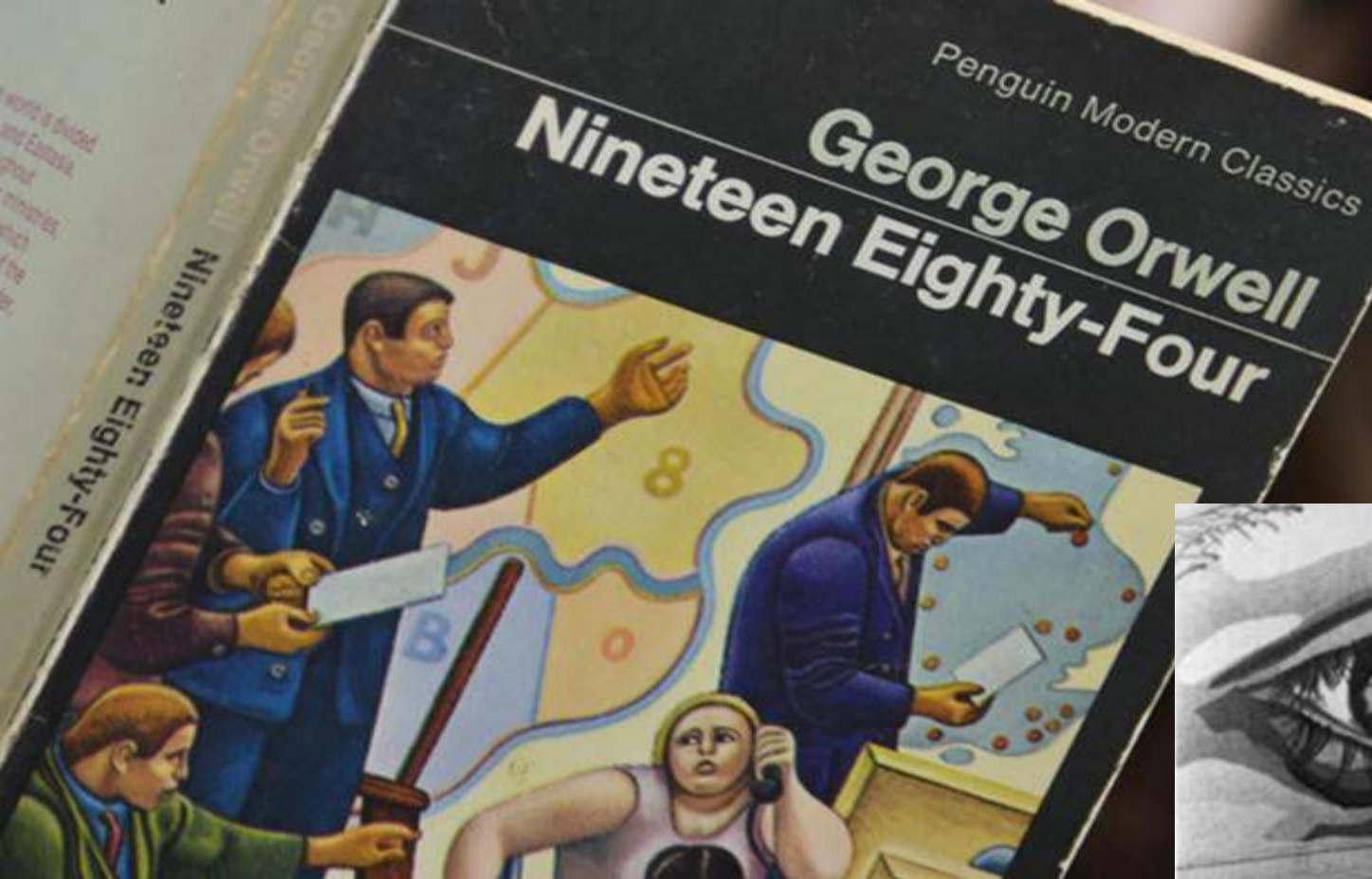
週六高鐵票已賣光



從眾（或一窩蜂）效應的典範「寶可夢現象」



自由時報



「老大哥在看著你」
(Big Brother is Watching You)





大數據也有黑暗面

- 資訊蒐集對隱私造成威脅（類似歐威爾的小說「1984」）；
 - 避免迷信因果關係而做出「預防性的懲罰」（例如：「關鍵報告」）；
 - 類似「訴諸權威」的誤謬，迷戀資料分析、濫用，以致於形成「資料獨裁」，或是「大數據的傲慢與偏見」。
- 「考試成績不等於真正學識」！



資料獨裁 vs. 資料驅動

To be or not to be, that is the question...

使用大數據的建議

■ 監管大數據的三個策略：

→ 個人資料保護由「個人同意制」，轉成資料使用者的「使用責任制」；

→ 運用大數據時，需尊重個人的「能動性」
(Human Agency; 政府對人民的判斷是基於實際的行動，而非大數據分析的預測)；

→ 培養「演算學家」以處理大數據的審計師。

註：除了量化分析人員，公司需積極培養大數據的專家（IT、分析師等）。

■ 不確定性 : It's Difficult to Make Predictions, Especially About the Future.

Predictive Tools



- *Prediction is very difficult, especially about the future.*

- Niels Bohr – Physicist (1885-1962)

"It is tough to make predictions, especially about the future."

- Yogi Berra, Baseball Savant





資料分享的副作用

- Facebook社群軟體盛行也帶來負面效應，除了個人資料安全疑慮，害怕與人接觸的社交恐懼症 (sociophobia) 日趨嚴重。
 - 擔心自己遭受社會遺忘或淘汰；
「滿紙荒唐言，一把辛酸淚！都云作者痴，誰解其中味？」
 - 人際關係的疏離（虛擬vs.實際）；
「假作真時真亦假，無為有處有還無」
 - 私刑（人肉搜索）、扮演法官。



大數據等於一切？

■ 無限母體 vs. 有限母體

→ 再多的資料還是樣本！（如消費者行為）

■ 資料品質仍是主要關鍵！

→ 人工智慧（AI）仰賴好的訓練資料，尋找背後的規則及趨勢（如 alphago）。

→ Google Flu Trends 透過搜尋引擎的「關鍵詞」，比美國疾管局（Center of Disease Control; CDC）更早偵測出流感盛行。

→ 族群及關鍵詞、從眾效應(Bandwagon effect)



大數據研究的建議

■ 統計學家（&資料科學家）的專業技能

→ 與人溝通、與資料溝通(data sense)、與專業溝通、與電腦（機器）溝通。

■ 增加專業領域間的溝通平台與機會

→ 如何結合人與機器？

→ 「多樣性」（物種、知識）是重要資產，
不少發現源自其他角度的思維。

（馬爾薩斯《人口論》與達爾文、馬克斯）

大數據的分析考量



- 分析大數據需兼顧兩個層面：技術、知識
(Technical Expertise Component & Functional Knowledge Component)，涵蓋以下四個項目：
 - Data Acquisition；（領域相關知識）
 - Data Processing；（資訊量化）
 - Data Management；（SQL等資料管理）
 - Data Mining/Analysis.（資料採礦）

祝大家有個
愉快的學期！

