

資料視覺化

數據交點 | 郭耀仁 yaojenkuo@datainpoint.com

The simple graph has brought more information to the data analyst's mind than any other device.

John Tukey

瞭解資料視覺化

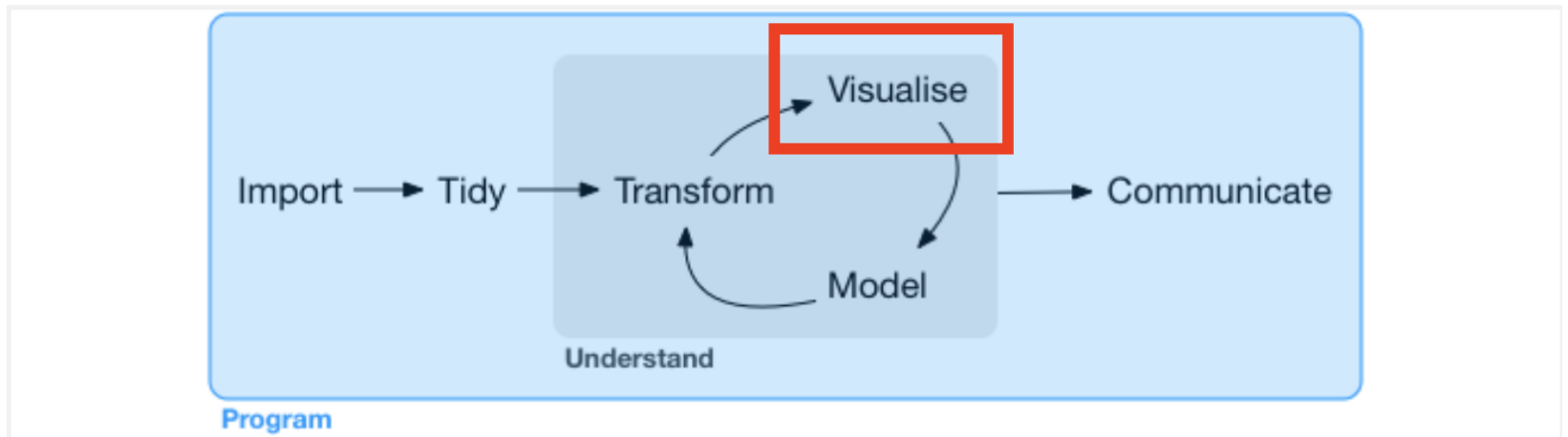
1. 資料視覺化的 What 與 Why。
2. 資料視覺化的工具箱。
3. 用四個面向考量資料視覺化的工具。

資料視覺化的 What 與 Why

何謂資料視覺化

資料視覺化是一門致力於將抽象性概念具體化的學科，透過圖形、顏色、形狀或大小等元素把原始資料或函式等所蘊含的特徵傳達給瀏覽者，進而將抽象的資訊轉換為溝通對象能快速掌握理解的精簡內容。

資料視覺化是現代資料科學應用的一個環節



來源：[R for Data Science](#)

為何資料視覺化

- 原始數列資料、函數對瀏覽者而言是抽象、不具資訊價值的。
- 資料視覺化對瀏覽者而言是具體、具有資訊價值的。

抽象的原始數列資料

In [3]:

```
print(arr)
```

```
[-1.07758271  1.55224998 -0.7295749  ... -0.18354096  0.51732877  
 2.35621533]
```


函數是抽象的

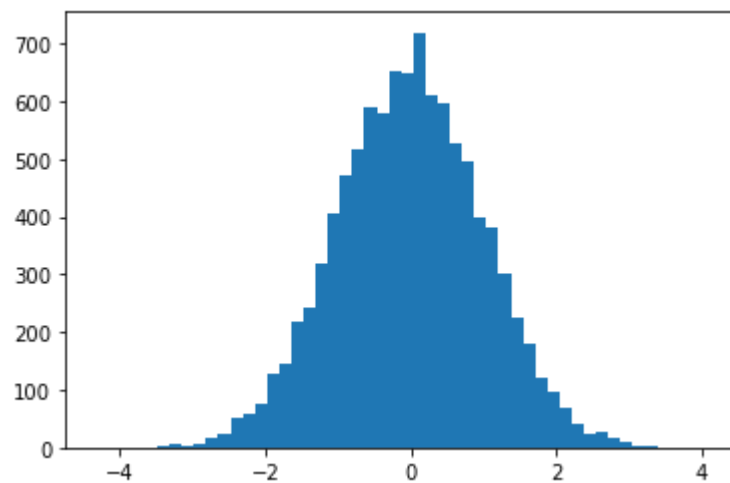
$$f(p) = -y_{true}\log(p) - (1 - y_{true})\log(1 - p) \quad (1)$$
$$y_{true} \in \{0, 1\}$$

函數是抽象的（續）

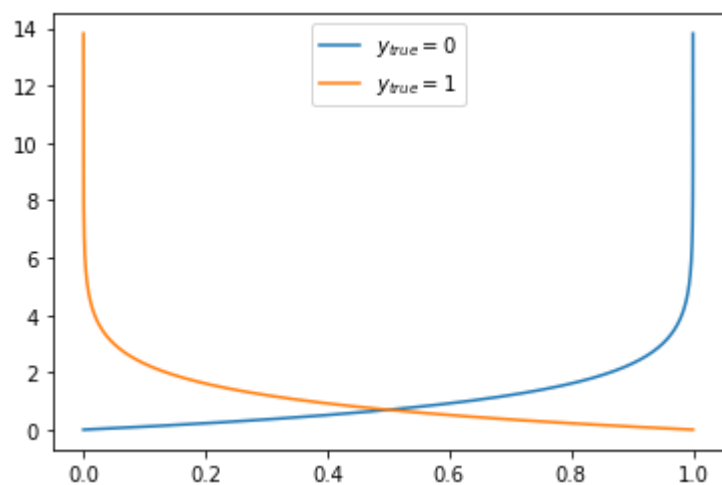
$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

資料視覺化對瀏覽者而言是具體、具有資訊價值的

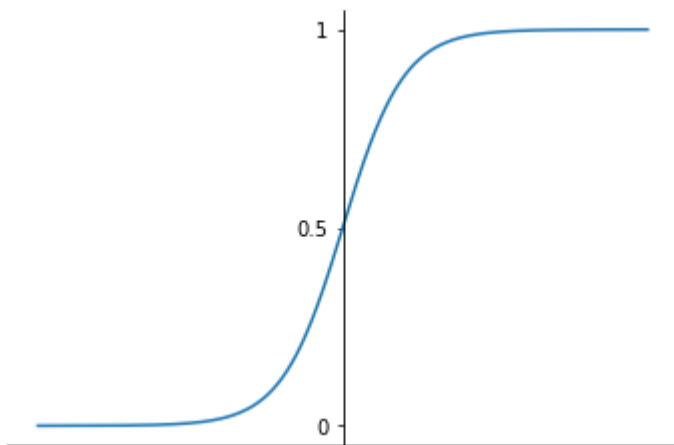
```
In [5]: plot_normal(arr)
```



```
In [7]: plot_cross_entropy(p, log_loss_0, log_loss_1)
```



```
In [9]: plot_sigmoid(x, y)
```



經典視覺化：有資訊價值、簡潔以及美觀

- [Charles Minard's map of Napoleon's disastrous Russian campaign of 1812](#)
- [Hans Rosling's 200 Countries, 200 Years, 4 Minutes](#)

資料視覺化的工具箱

如何做資料視覺化？



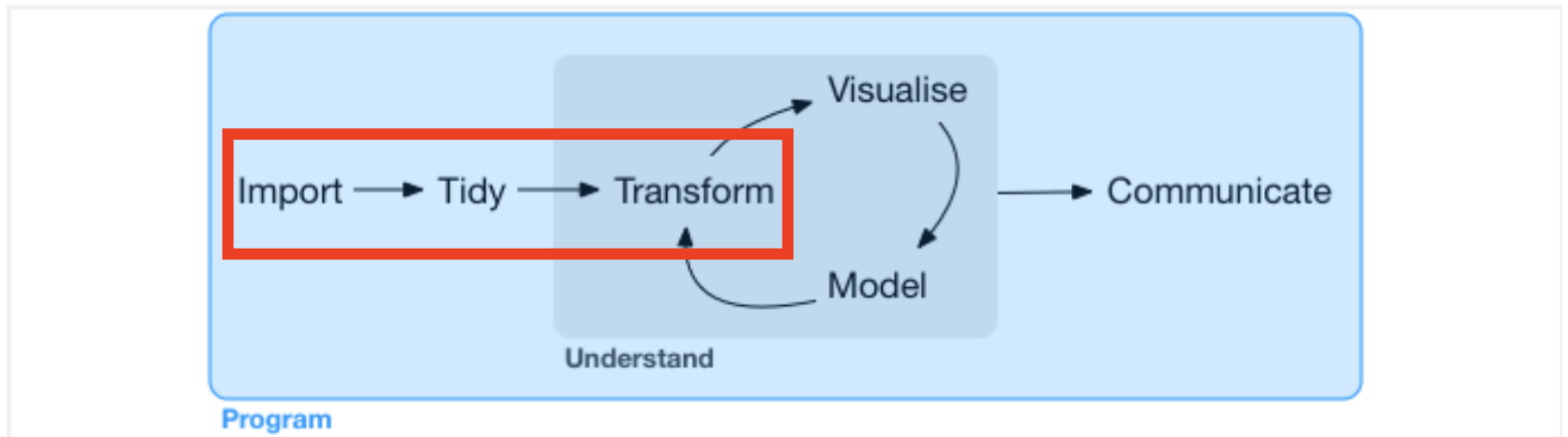
來源：<https://giphy.com>

有一半的難關不是在作圖

When it comes to making graphs, half the battle occurs before you call any graphing commands.

Winston Chang

有一半的難關在於資料視覺化的前置作業



來源：[R for Data Science](#)

資料視覺化三步驟

1. 獲取資料。
2. 將資料整理為合適的格式與類型。
3. 輸入資料視覺化套件所定義好的函數。

各據山頭的資料視覺化工具

- 商業智慧軟體
- 程式語言
 - Python
 - R

商業智慧軟體

- PowerBI
- Tableau

Python 的資料視覺化模組

- [Matplotlib](#)
- [Seaborn](#)
- [Plotly](#)
- [Folium](#)
- [Dash](#)
- [Streamlit](#)

R 的資料視覺化模組

- Base Plotting System
- Ggplot2
- Plotly
- Shiny

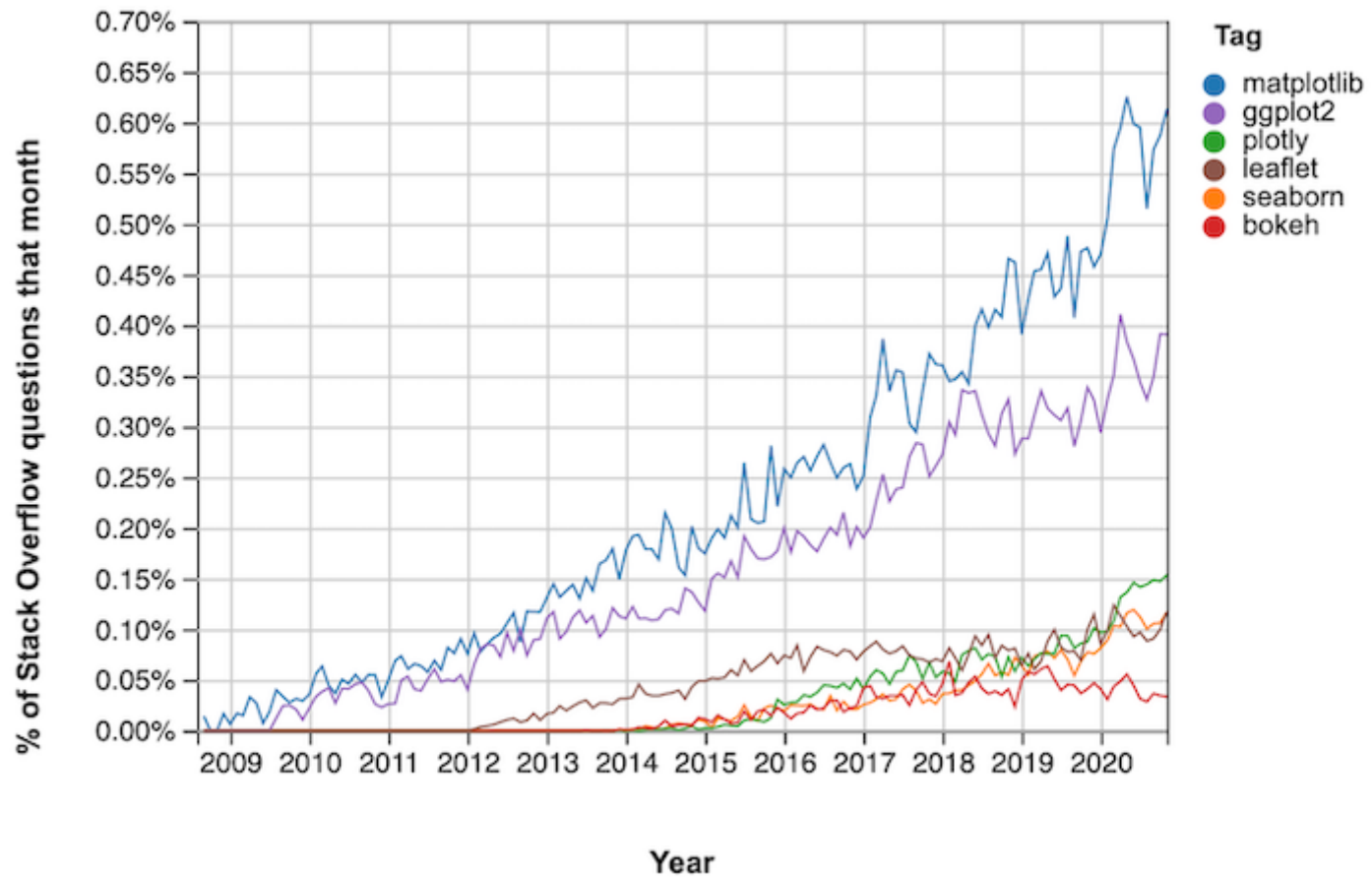
用四個面向考量資料視覺化的工具

資料視覺化工具這麼多，該怎麼選擇？

我建議可以考量四個面向

1. 使用者數量（「西瓜偎大邊」與「衝量」哲學）。
2. 互動性。
3. 語法與彈性。
4. 資料格式與視覺化類型。

考量使用者數量

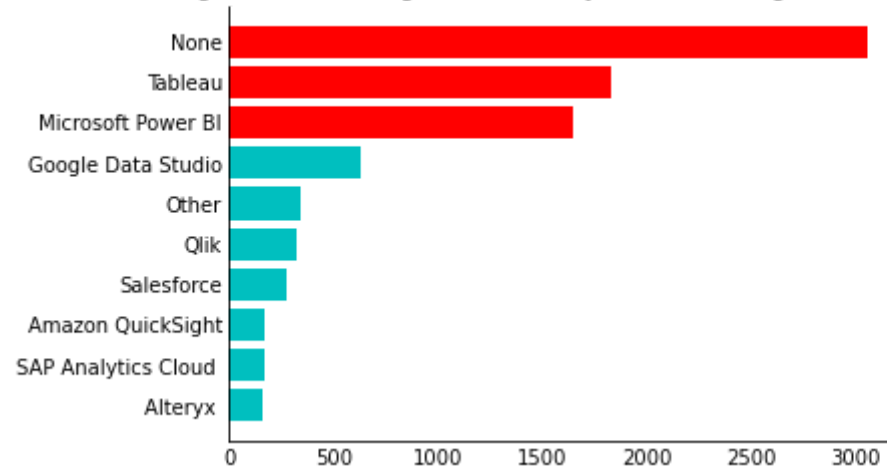


來源: [Stack Overflow Trends](#)

```
In [11]: ks.plot_summary('Q31A')
```

Which of the following business intelligence tools do you use on a regular basis? (Select all that apply)
Too many categories, only showing the top 10.

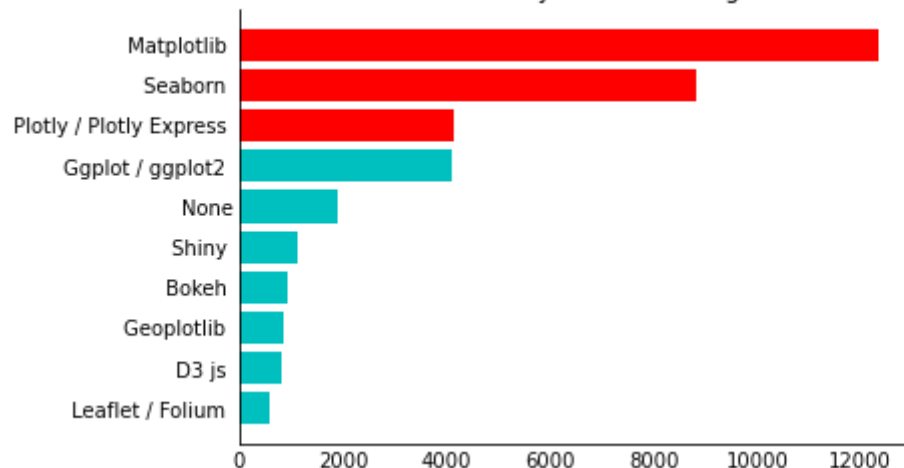
Which of the following business intelligence tools do you use on a regular basis? (Select all that apply)



```
In [12]: ks.plot_summary("Q14")
```

What data visualization libraries or tools do you use on a regular basis?
(Select all that apply)
Too many categories, only showing the top 10.

What data visualization libraries or tools do you use on a regular basis? (Select all that apply)



考量互動性

靜態 vs. 動態。

考量語法與彈性

- 物件導向型 vs. 陳述型。
- 低階 vs. 高階。

考量資料格式與視覺化類型

- 資料格式：陣列、資料框或 JSON 等。
- 特定的視覺化：樹狀圖、地理資訊或 Sankey Diagram 等。

簡而言之，商業智慧軟體

- Windows 作業系統
 - Tableau
 - PowerBI
- macOS
 - Tableau

簡而言之，Python 的視覺化模組

- 靜態
 - 低階： [Matplotlib](#)
 - 高階： [Seaborn](#)
- 動態
 - 在筆記本中： [Plotly](#)
 - 獨立網頁應用： [Dash](#) / [Streamlit](#)

簡而言之，R 的視覺化模組

- 靜態
 - 低階： [Base Plotting System](#)
 - 高階： [Ggplot2](#)
- 動態
 - 在筆記本中： [Plotly](#)
 - 獨立網頁應用： [Shiny](#)

試用看看、參考範例

- [Matplotlib](#)
- [Seaborn](#)
- [Plotly](#)
- [Dash](#)
- [Streamlit](#)

試用看看、參考範例（續）

- [Base Plotting System](#)
- [Ggplot2](#)
- [Plotly](#)
- [Shiny](#)