

Day 11

探索式數據分析

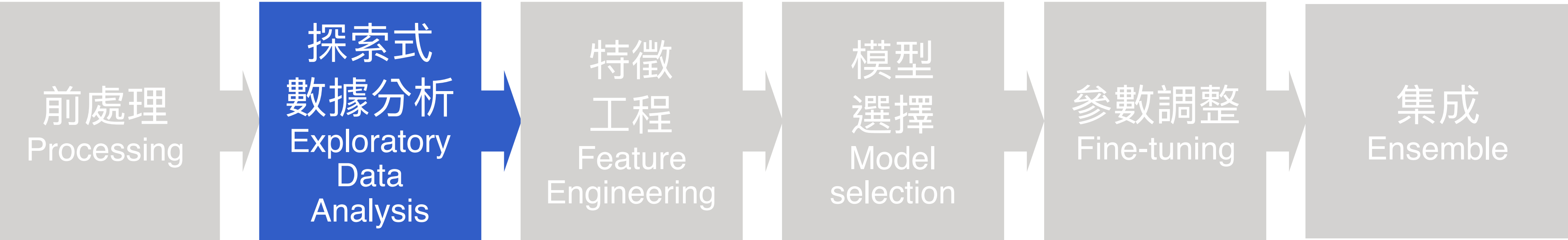
繪圖與樣式 &
Kernel Density Estimation (KDE)



知識地圖 探索式數據分析 核密度函數

機器學習概論 Introduction of Machine Learning

監督式學習 Supervised Learning

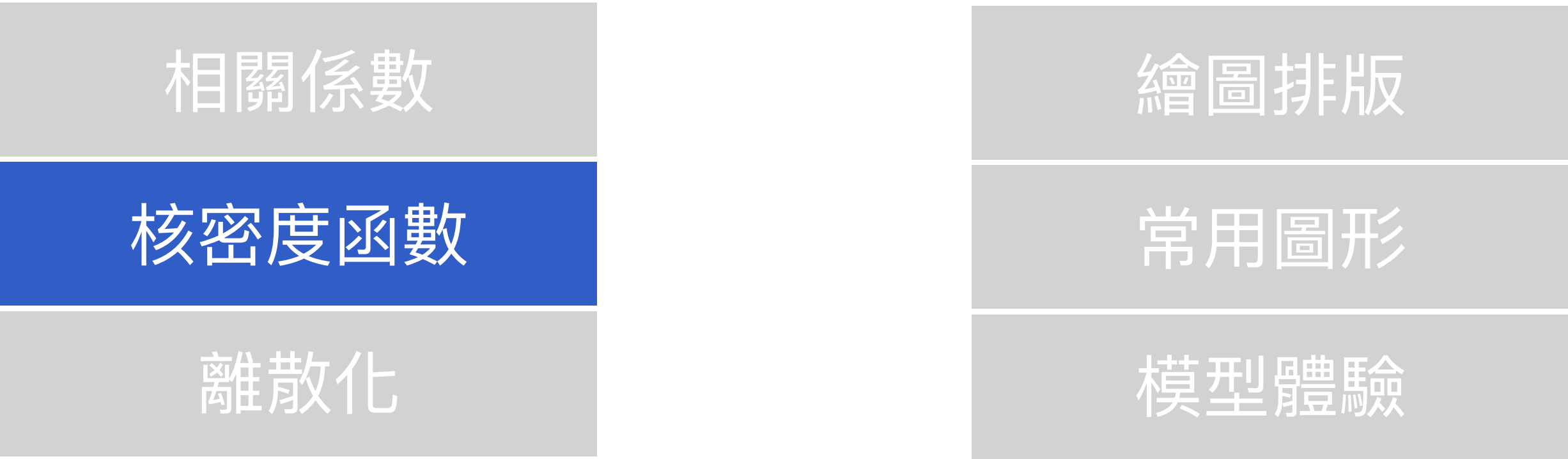


非監督式學習 Unsupervised Learning



探索式數據分析 Exploratory Data Analysis (EDA)

統計值的視覺化




本日知識點目標



目標
知識點

知道 matplotlib 的其他 theme



獲得
知識點

完成今日課程後你應該可以了解

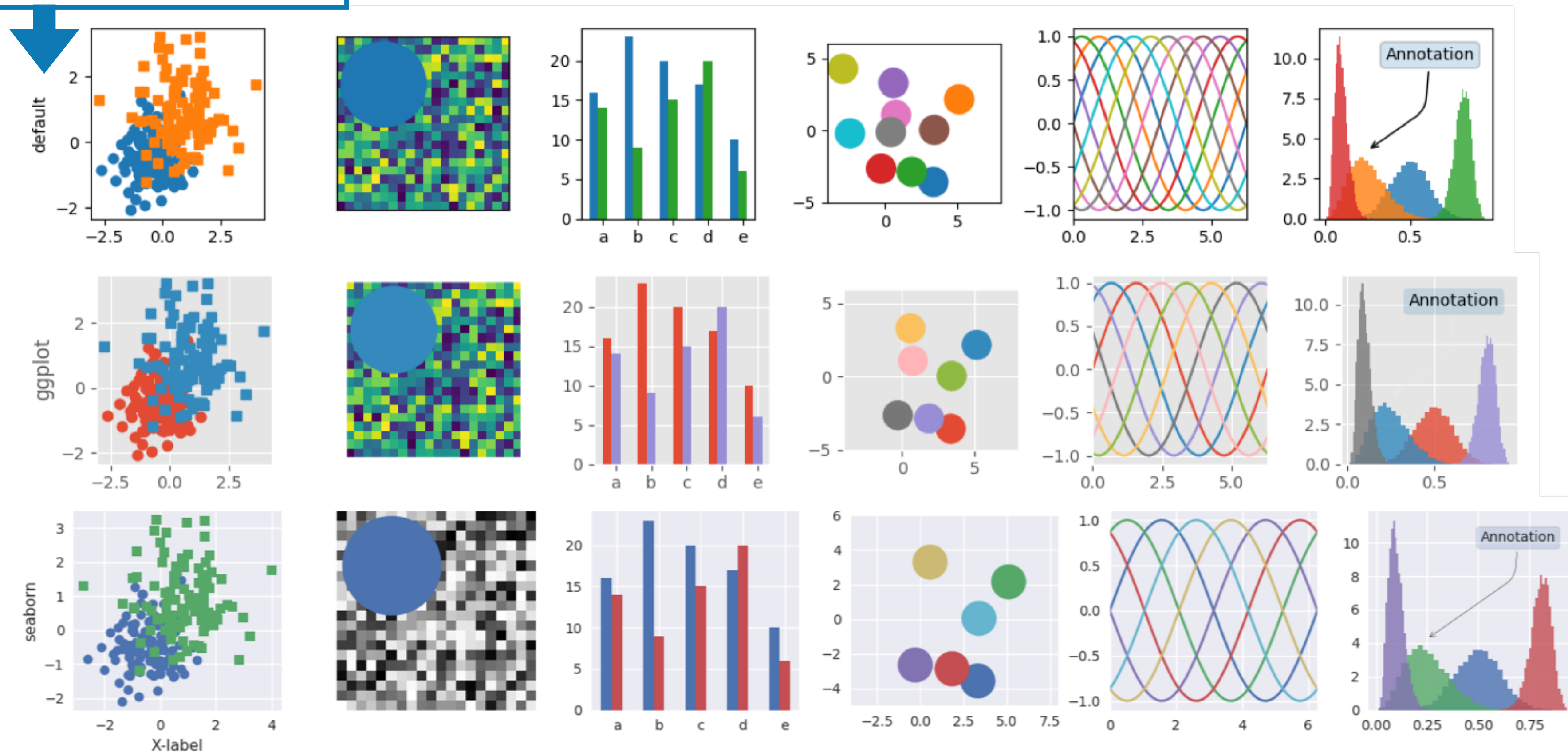
- 學會什麼是 Kernel Density Estimation (KDE) 與如何繪製

繪圖風格

```
plt.style.use('default') # 不需設定就會使用預設  
plt.style.use('ggplot')  
plt.style.use('seaborn') # 或採用 seaborn 套件繪圖
```

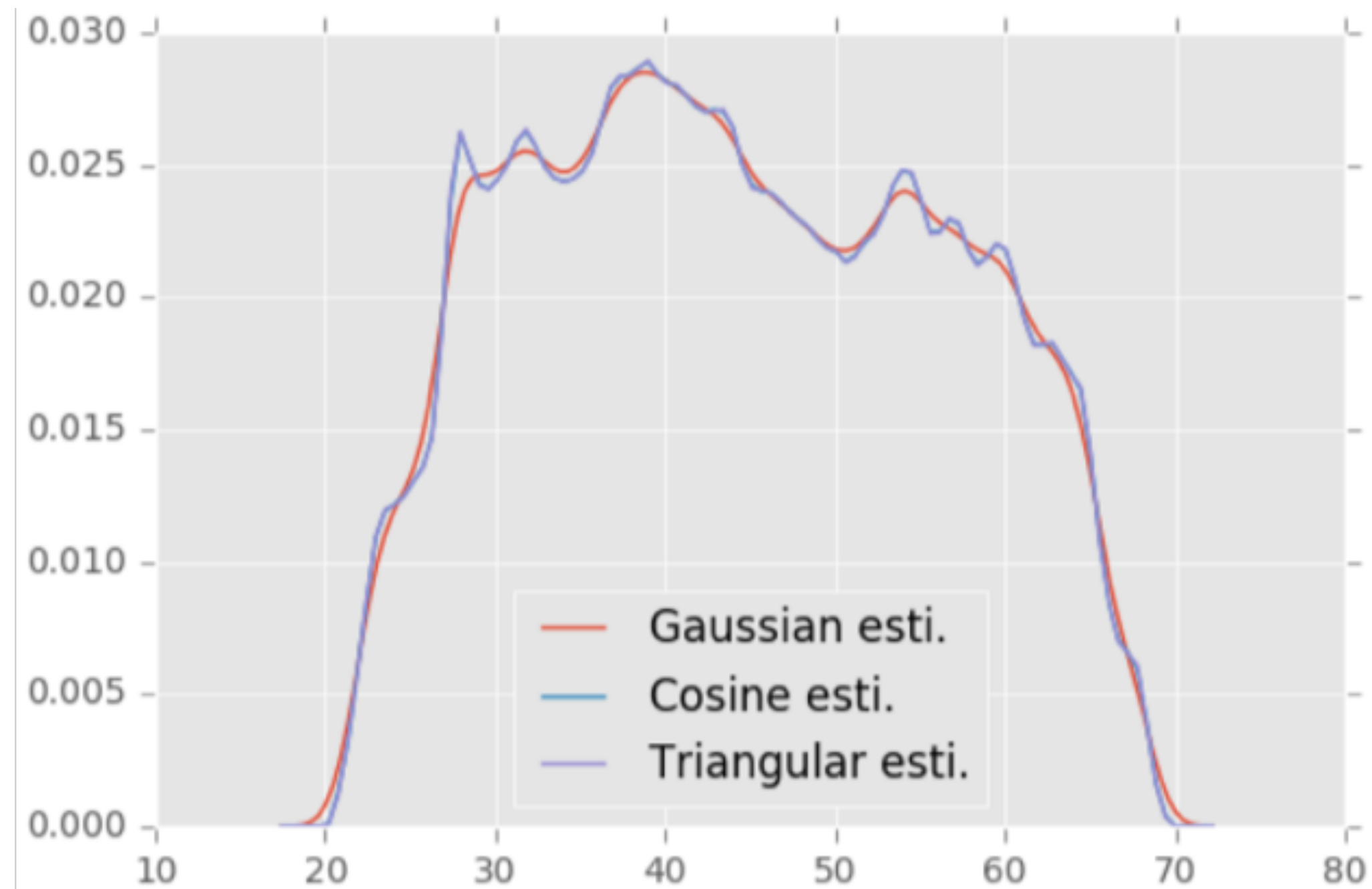
轉變繪圖風格的目的

用已經被設計過的風格，
讓觀看者更清楚明瞭，
包含色彩選擇、線條、
樣式等。



Kernel Density Estimation (KDE)

不同 kernel function 的結果



1

採用無母數方法畫出一個觀察變數的機率密度函數

某個 X 出現的機率為何

2

Density plot 的特性

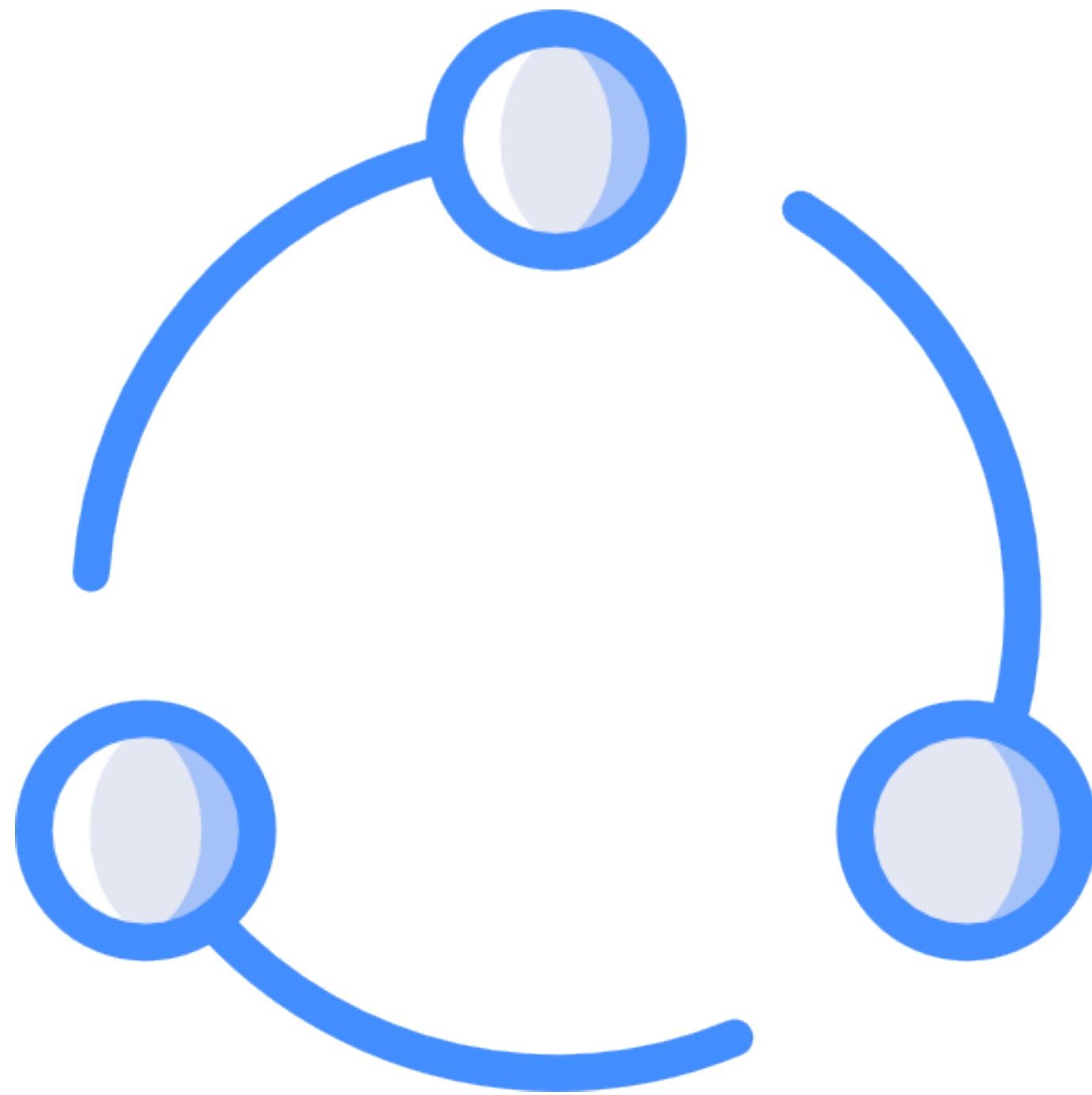
- 歸一：線下面積和為 1
- 對稱： $K(-u) = K(u)$

3

常用的 Kernel function

- Gaussian (Normal dist)
- Cosine

重要知識點複習



- KDE 的優點與缺點
 - 優：無母數方法，對分布沒有假設 (使用上不需擔心是否有一些常見的特定假設，如分布為常態)
 - 缺：計算量大，電腦不好可能跑不動
- 透過 KDE plot，我們可以較為清楚的看到不同組間的分布差異

解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

