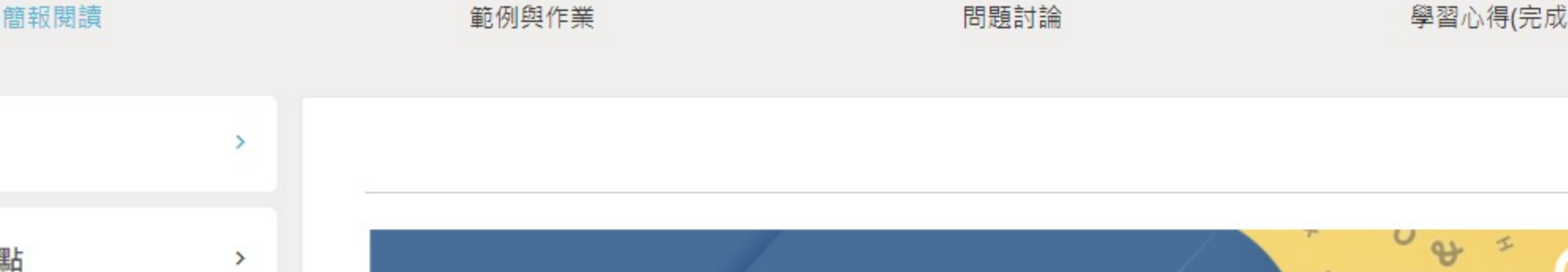


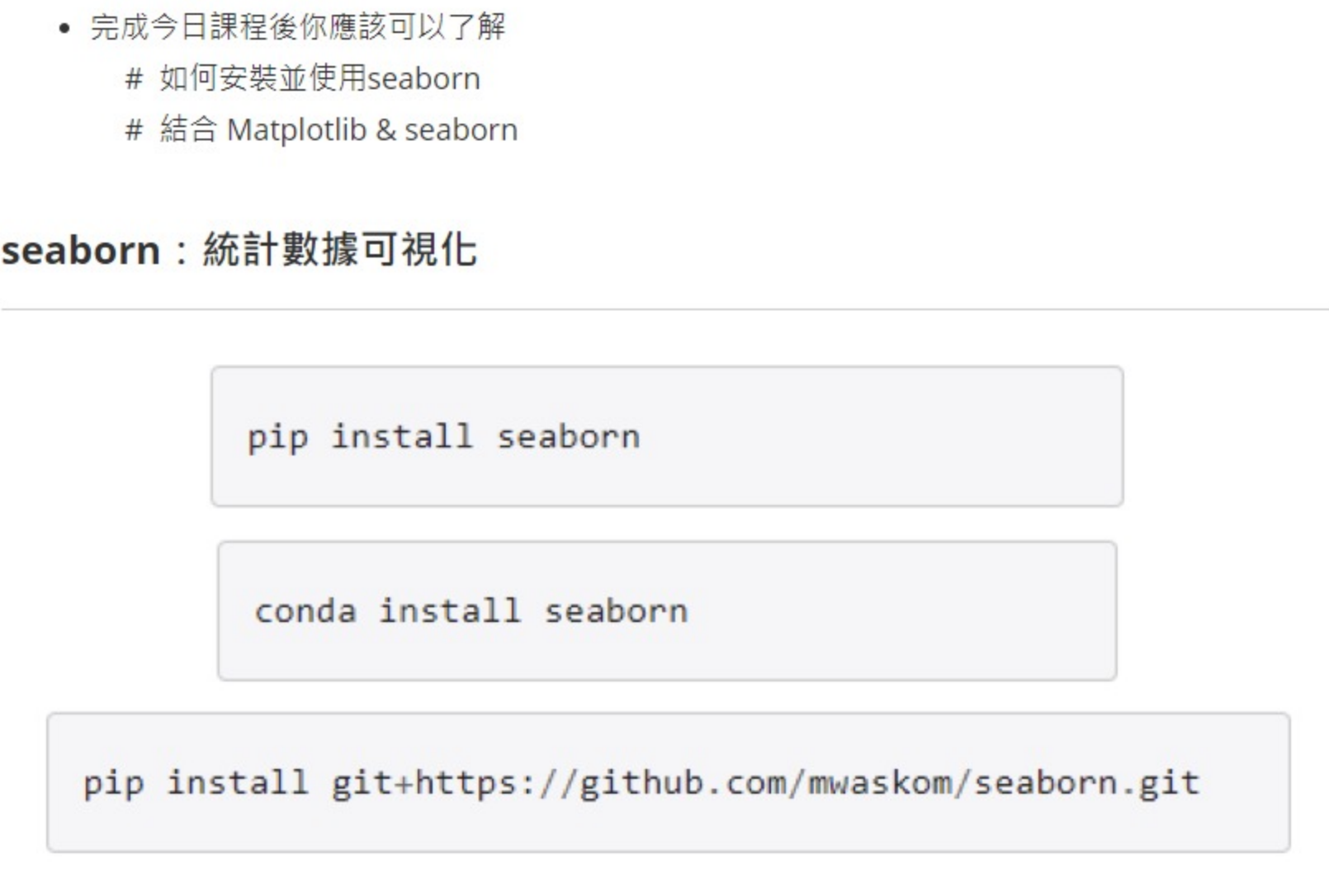
D20 使用 Seaborn 進行資料視覺化



重要知識點	>
seaborn：統計數據可視化	>
Seaborn 圖樣式	>
可視化統計關係	>



重要知識點



- 了解資料視覺化應用
- 完成今日課程後你應該可以了解
 - # 如何安裝並使用Seaborn
 - # 結合 Matplotlib 與 seaborn

seaborn：統計數據可視化

```
pip install seaborn
```

```
conda install seaborn
```

```
pip install git+https://github.com/mwaskom/seaborn.git
```

功能：

- 可視化統計關係
- 構建結構化多維圖網格
- 繪圖美學

Seaborn 圖樣式

有五種預設seaborn主題：darkgrid、whitegrid、dark、white、和ticks，它們各自適合不同的應用和個人喜好，默認主題是darkgrid

```
sns.set_style("whitegrid") # 設定圖形樣式
```

```
#設定圖示的高與寬
data = np.random.normal(size=(20, 6))
+ np.arange(6) / 2
```

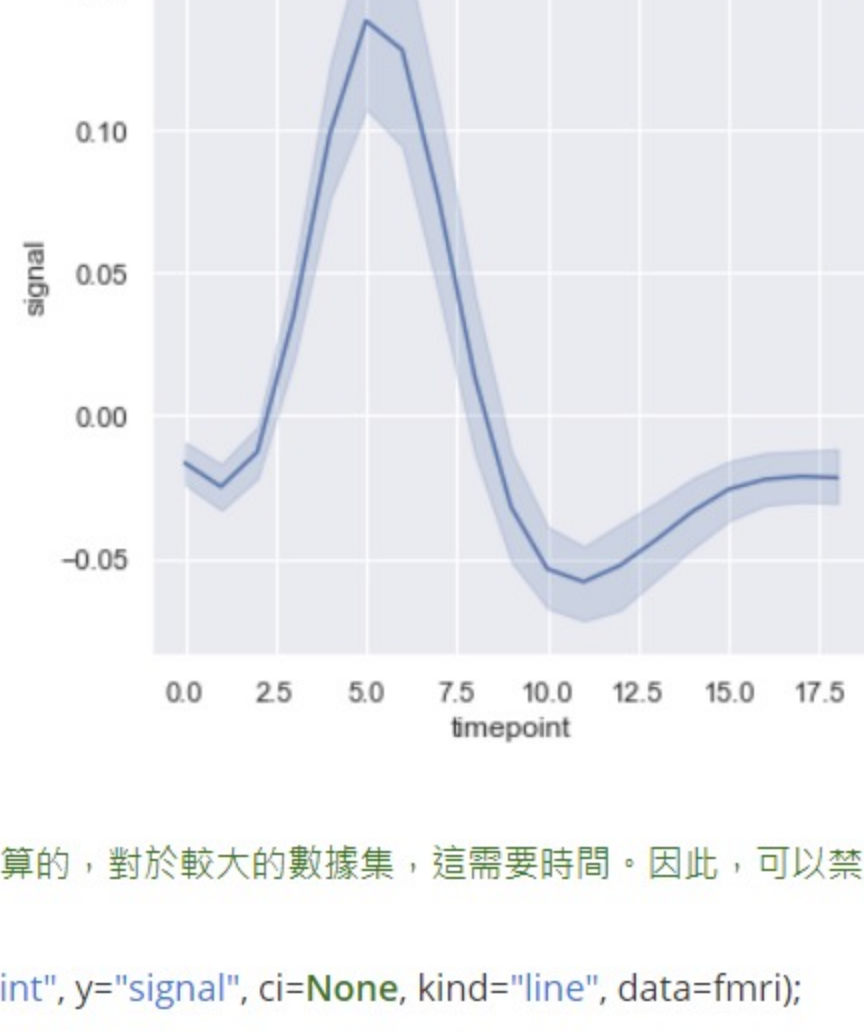
```
#列印圖示
sns.boxplot(data=data);
```



可視化統計關係

- 統計分析是一個了解數據集中變數之間如何相互關聯以及這些關係如何依賴於其他變數的過程。可視化可能是此過程的核心元件，因為當數據正確可視化時，人類可視化系統可以看到指示關係的趨勢和模式。
- 將變數與散點圖關聯，散點圖是統計可視化的支柱，它使用點雲描繪了兩個變數的聯合分佈，其中每個點表示數據集中的觀測值，這種描述允許眼睛辨識出大量有關它們之間是否有任何有意義的關係的資訊。

```
# 散點圖關聯變量
sns.set(style="darkgrid")
sns.relplot(x="total_bill", y="tip", data=tips);
```



當點以兩個維度繪製時，可以通過根據第三個變數對點進行着色，將另一個維度添加到繪圖中。這稱為使用「色調語義」，因為點的颜色獲得意義：
sns.relplot(x="total_bill", y="tip", hue="smoker", data=tips);



聚合和表示不確定性

更複雜的數據集將具有變數相同值的多個度量，Seaborn 中的預設行為是透過繪製平均值和 95% 信心區間圍繞平均值來聚合每個值的多個測量值：xx

```
fmri = sns.load_dataset("fmri")
sns.relplot(x="timepoint", y="signal", kind="line", data=fmri)
```



信心區間是使用引導計算的，對於較大的數據集，這需要時間，因此，可以禁用它們：

```
sns.relplot(x="timepoint", y="signal", ci=None, kind="line", data=fmri);
```



對於較大的數據是通過繪製標準差而不是信心區間來表示每個時間點的分佈：

```
sns.relplot(x="timepoint", y="signal", kind="line", ci="sd", data=fmri);
```



使用語義映射繪製數據子集

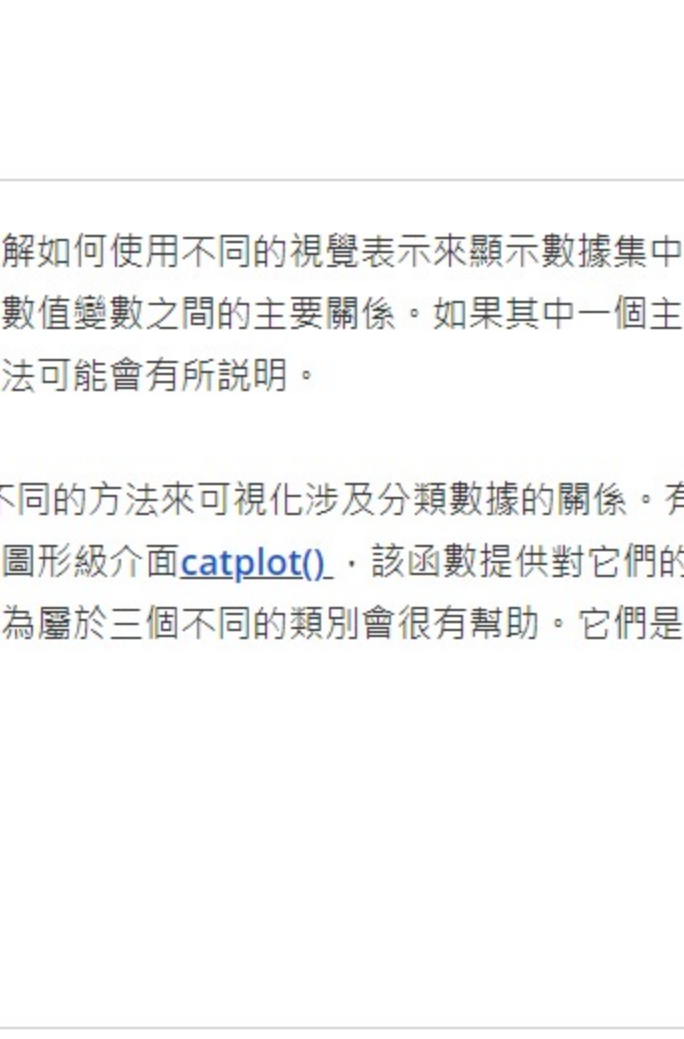
Lineplot() 具有與 Scatterplot() 相同的靈活性：它可以通過修改繪圖元素の色調、大小和樣式來顯示最多三個附加變數

```
sns.relplot(x="timepoint", y="signal", hue="event", kind="line", data=fmri);
```



預設情況下，向線條圖添加樣式語義會更改行中的破折號樣式：

```
sns.relplot(x="timepoint", y="signal", hue="region", style="event", kind="line", data=fmri);
```



通過每個觀測中使用的標記來標識子集，或者與破折號一起識別子集，或者代替它們：

```
sns.relplot(x="timepoint", y="signal", hue="region", style="event", dashes=False, markers=True, kind="line", data=fmri);
```



可視化線性關係

- 許多數據集包含多個定量變數，分析的目標通常是將這些變數相互關聯
- 可以通過顯示兩個變數的關係分佈來完成此目的函數，該函數提供對它們的統一更高級別的訪問。
- Seaborn 的目標是通過可視化快速輕鬆地瀏覽數據集，因為這樣做與通過統計表瀏覽數據集同樣重要

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(color_codes=True)
tips = sns.load_dataset("tips")
```

```
sns.regplot(x="total_bill", y="tip", data=tips);
```


調用，兩個函數繪製兩個變數的散點圖，然後拟合回歸模型並繪製生成的回歸線和該回歸的 95% 置信區間：xyy ~ x


```
sns.lmplot(x="size", y="tip", data=tips);

sns.set(style="whitegrid")
sns.relplot(x="size", y="tip", data=tips);
```

```
sns.set(color_codes=True)
x = np.random.normal(size=100)
sns.distplot(x);
sns.distplot(x, bins=20, kde=False, rug=True);
```


知識點複習

在關係圖教程中，我們瞭解如何使用不同的視覺表示來顯示數據集中的多個變數之間的關係。在示例中，我們重點介紹了兩個變數之間的主要關係，如果其中一個主要變數是“分類”(分為離散組)，則使用更專業化的可視化方法可能會有所說明，

在 Seaborn 中，有幾種不同的方法來可視化涉及分類數據的關係，有許多軸級函數用於以不同的方式繪製分類數據，以及一個圖形級介面(catplot())，該函數提供對它們的統一更高級別的訪問。

將不同的分類情節類型視為屬於三個不同的類別會很有幫助，它們是：

- 分類散點圖
- 分類分佈圖
- 分類估計圖

延伸閱讀

使用Seaborn Python包進行可視化

網站：[Visualization Using Package Seaborn Python](#)

- Seaborn是一個統計繪圖庫
- 它具有漂亮的默認樣式
- 它也被設計為與Pandas數據框對象一起很好地工作。
- 使用散點圖關聯變量
- 用線圖強調連續性
- 使用語義映射繪製數據子集

Putri Choirunisa 4 Followers About Follow

Visualization Using Package Seaborn Python

Putri Choirunisa Jan 15 · 6 min read

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics

Keys Features

- Seaborn is a statistical plotting library
- It has beautiful default styles
- It also is designed to work very well with Pandas dataframe objects.

LETS TRY!!!!

In this case using the tips data in the Seaborn library, first step install and getting started. To install the latest release of seaborn, you can use pip:

```
pip install seaborn
```

It's also possible to install the released version using conda:

```
conda install seaborn
```

[下一步：閱讀範例與完成作業](#)