

使用 Seaborn 進行資料視覺化



簡報閱讀



範例與作業



問題討論



學習心得(完成)

重要知識點

seaborn：統計數據視覺化

Seaborn 圖樣

如何表示數據關係

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 & seaborn

```
pip install seaborn
```

```
conda install seaborn
```

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

功能：

- 視覺化統計關係
- 構建結構化多繪圖網格
- 繪圖美學

Seaborn 圖樣

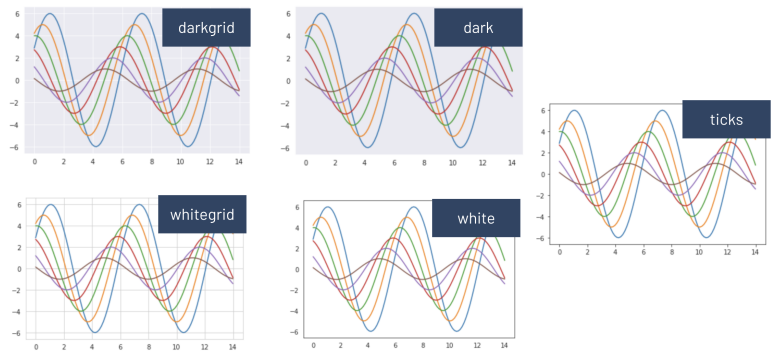
有五種預設 seaborn 主題：darkgrid，whitegrid，dark，white，和 ticks。它們各自適合不同的應用和個人喜好。默認主題是 darkgrid

`sns.set_style("darkgrid")` # 設定圖形樣式

- grid：就是圖表當中的格線，這個格線能夠有助於使用者快速地查看圖表當中的數值型資料
 - darkgrid：圖表背景顏色的有，加上格線
 - whitegrid：圖表背景顏色的無，加上格線
- dark 與 white：就是圖表背景顏色的有無
- ticks：在 x 軸及 y 軸座標的地方多了一條短短的線，輔助我們去做圖表上的位置核對

`sns.set_style("")`，分別使用自定義的 `sinplot()`

```
1 #宣告一個 sin 正弦波
2 def sinplot(flip=1):
3     x = np.linspace(0, 14, 100)
4     for i in range(1, 7):
5         plt.plot(x, np.sin(x + i * .5) * (7 -
```



如何表示數據關係

以常用的散點圖曲線圖來表示變數之間的關係：

`sns.relplot()`

關注的是統計量之間的關係。x, y 數值型數據，
關注兩個數值變數之間的關係

`relplot()` - relationship plot

`replot` 常用參數

`hue`：在同一維度上，用顏色區分不同數據

`style`：在同一維度上，用線的不同表現形式區分，如點線，虛線等

`size`：控制點大小或者線條粗細

`kind`：繪製圖的型態. `kind='scatter'` (散點圖, 預設值) `kind='line'` (線圖)，可以通過參數 `ci`：

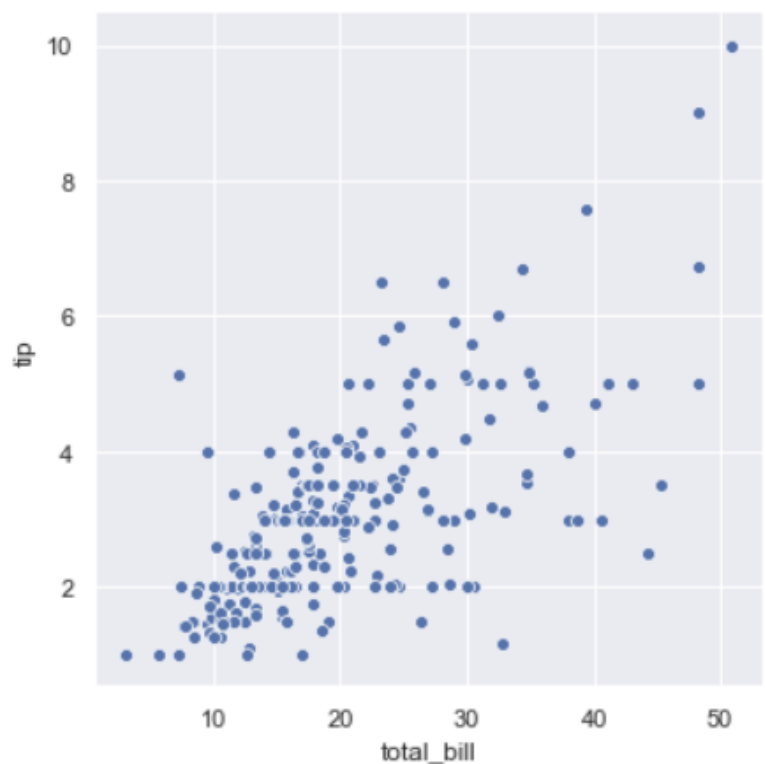
(confidence interval) 參數，來控制陰影部分

- 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。這種描述允許眼睛推斷出大量有關它們之間是否有任何有意義的關係的資訊。

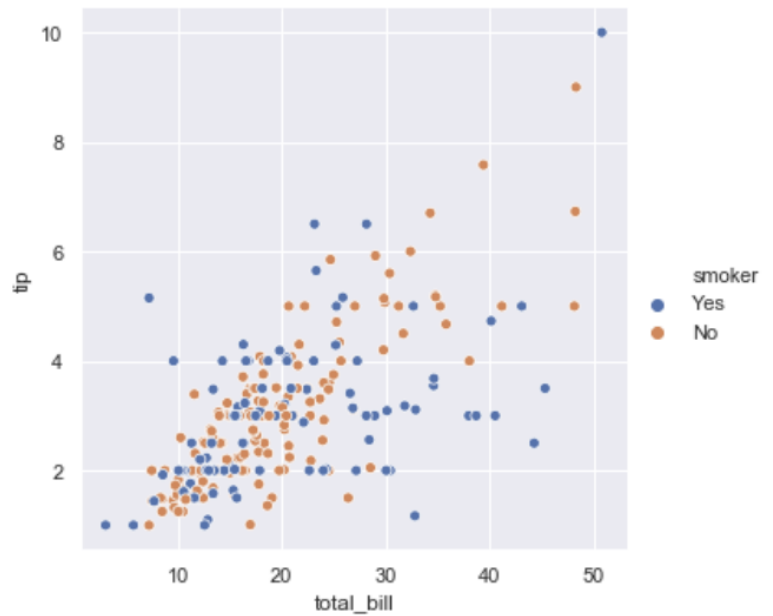
散點圖關聯變量

```
sns.set(style="darkgrid")
```

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

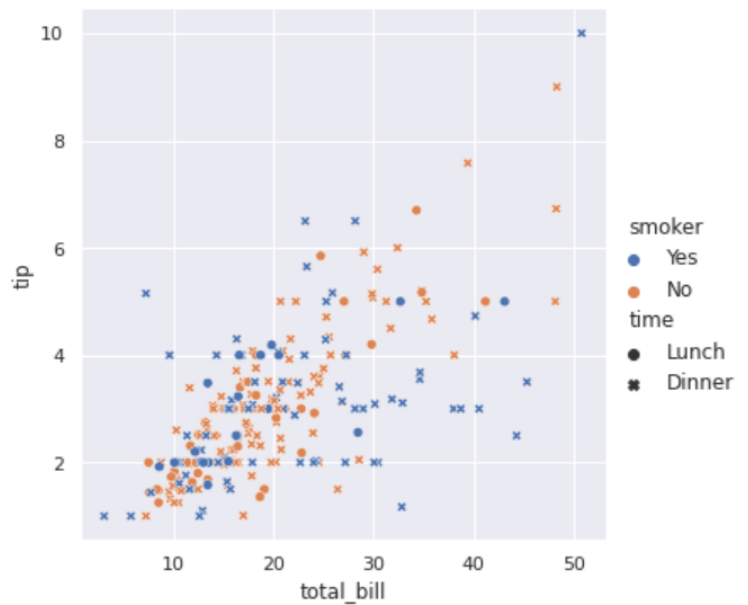


當點以兩個維度繪製時，可以通過根據第三個變數對點進行著色，將另一個維度添加到繪圖中。這稱為使用「色調語義」，因為點的顏色獲得意義：



通過獨立地改變每個點的色調和樣式，也可以表示四個變量

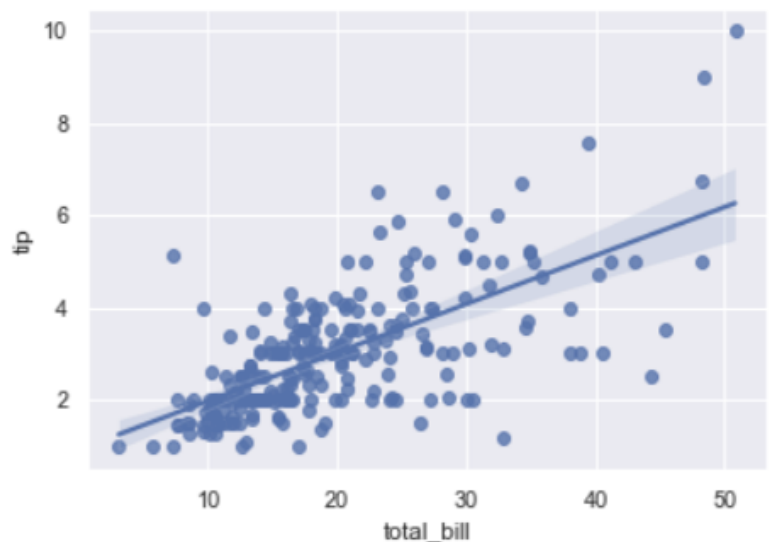
```
sns.relplot(x="total_bill", y="tip",
hue="smoker", style="time", data=tips);
```



回歸線性關係

- 可以通過顯示兩個變數的關係分佈來完成此目的的函數。可以將通過線性回歸的通用框架實現。
- Seaborn 的目標是通過可視化快速輕鬆地流覽數據集，因為這樣做與通過統計表流覽數據集同樣重要
- 調用中，兩個函數繪製兩個變數的散點圖，然後擬合回歸模型並繪製生成的回歸線和該回歸的 95% 信心區間
- regplot() 常用參數
 - x / y：特徵跟標記
 - data：Pandas dataframe
 - dropna：移除 NA 值，True/False

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



知識點回顧

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。

如果其中一個主要變數是"分類"(分為離散組)，則使用回歸方法。

在 Seaborn 中，有幾種不同的方法來可視化涉及分類數據的關係。

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
- 分類分布圖
- 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

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](#)

