**Візуалізація даних за допомогою бібліотеки Seaborn**

*Seaborn* - бібліотека візуалізації даних Python, що побудована на matplotlib

Повна документація по користування бібліотекою та галерея з варіантами графіки доступна за посиланням:

<http://seaborn.pydata.org/>

Установка

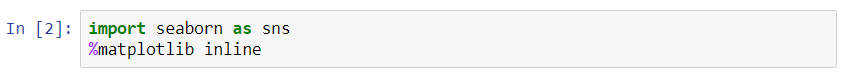
Для установки Matplotlib можна використовувати pip і conda:

conda install seaborn

pip install seaborn

Імпортування

Для імпортування бібліотеки використовується команда:

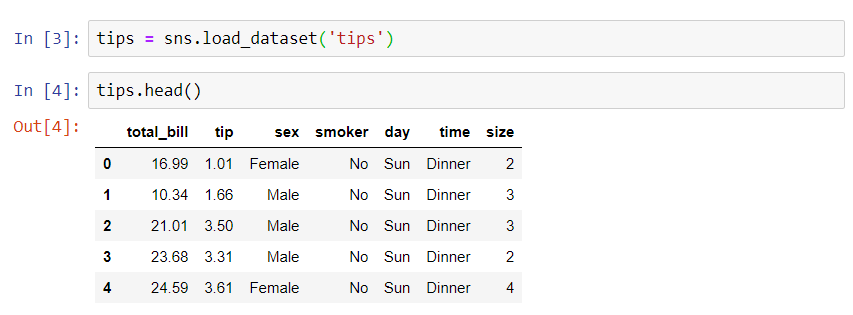


Тестові дані

Бібліотека seaborn містить в собі вбудовані набори даних, повний перелік яких можна подивитись за посиланням:

<https://github.com/mwaskom/seaborn-data>

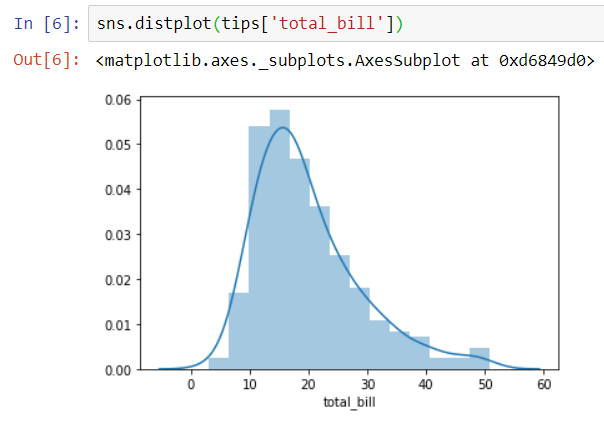
Завантаження даних з цих наборів можливе за допомогою команди:



Візуалізація даних

*Графіки для аналізу кількісних даних*

*distplot -* показує розподіл одновимірного набору спостережень.

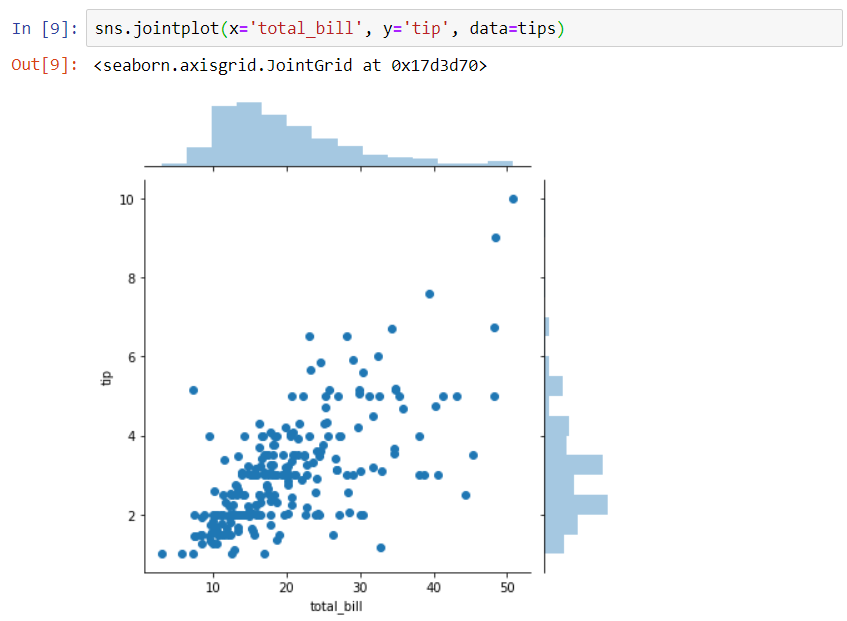


Можливі варіанти для модифікації:

kde – за замовчуванням True. Якщо False – на графіку буде відображено лише гістограму, без кривої розподілу.

bins – ширина «кроку» для гістограми.

*jointplot* – дозволяє співставити розподіл двох змінних та їх взаємозв’язок за допомогою діаграми розсіювання.

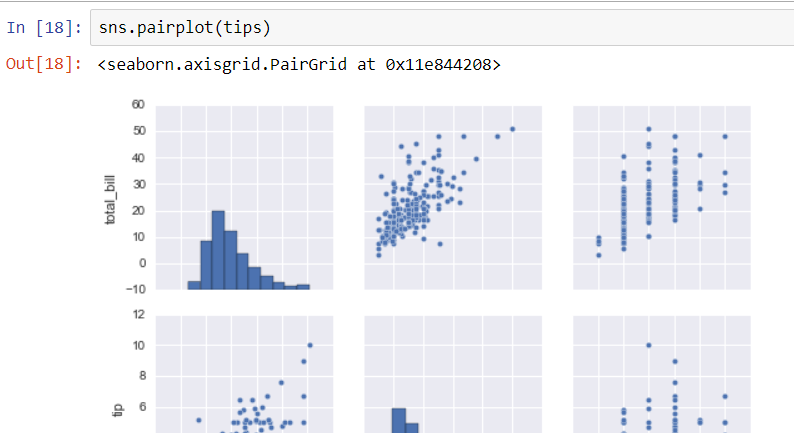


Можливі варіанти для модифікації:

kind – варіант діаграми розсіювання. Можливі варіанти – scatter (за замовчуванням), reg, resid, kde, hex.

*pairplot* – будує попарні взаємозв’язки між змінними в повному наборі даних.

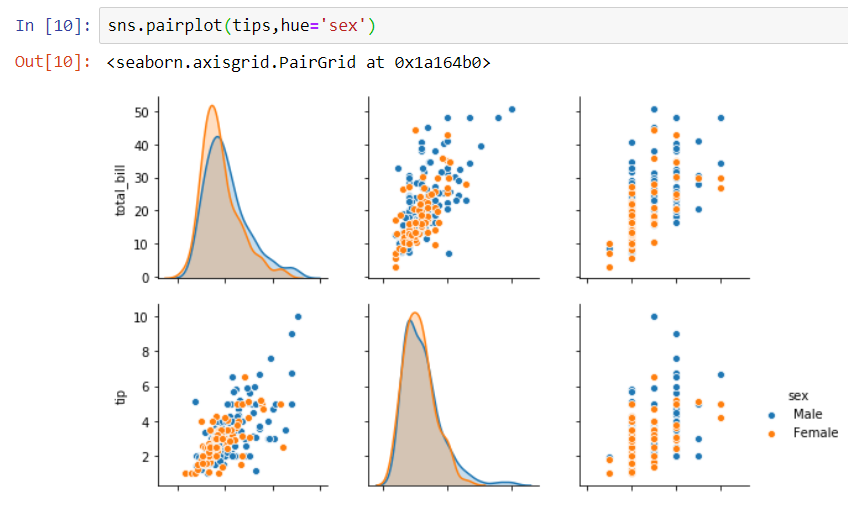
По діагоналі відображені графіки розподілу даних, в інших комірках - діаграми розсіювання (для відображення взаємозв’язку між двома змінними).



Можливі варіанти для модифікації:

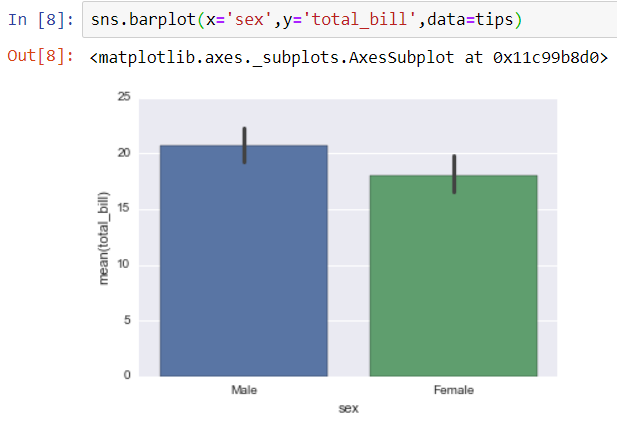
hue – вибір категоріальної змінної для розділення даних на окремі підгрупи та для їх кольорового відображення на діаграмі.

palette – кольорова схема для візуалізації (http://seaborn.pydata.org/tutorial/aesthetics.html)



*Графіки для аналізу категоріальних даних*

*barplot* – дозволяють отримувати сукупні дані по кількісним даним в залежності від категоріальної змінної.

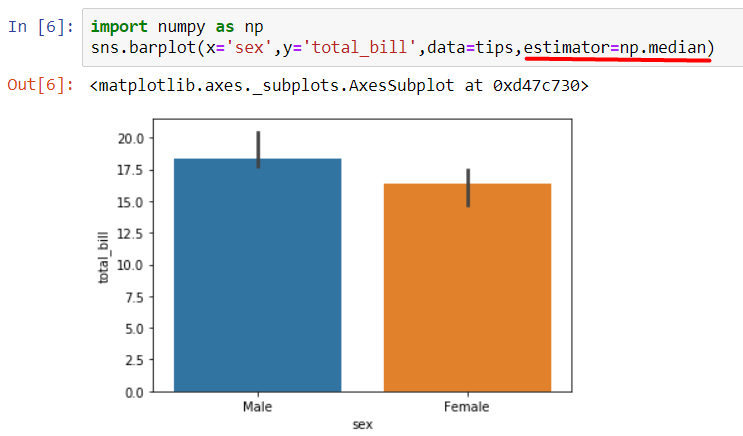


Для даного графіку х – категоріальна змінна, у – кількісна. За замовчуванням відображено середнє значення кількісної змінної.

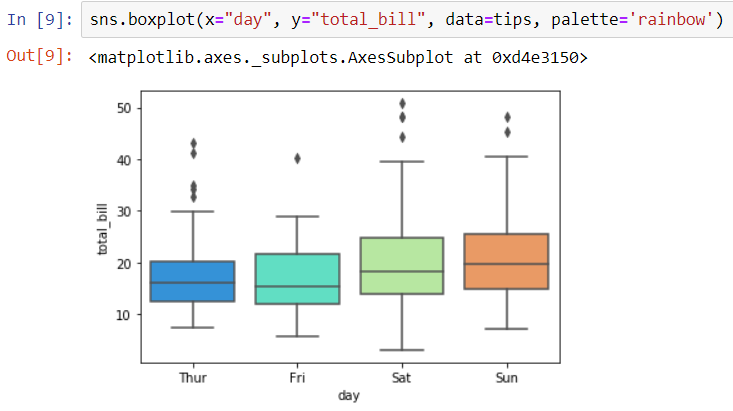
Можливі варіанти для модифікації:

hue – вибір додаткової категоріальної змінної для розділення даних на окремі підгрупи.

estimator – зміна варіанту відображення кількісної змінної. Наприклад, не середнє значення, а медіана, мода тощо. Даному параметру можна призначити як власні функції, так і функції з різних бібліотек. Наприклад:



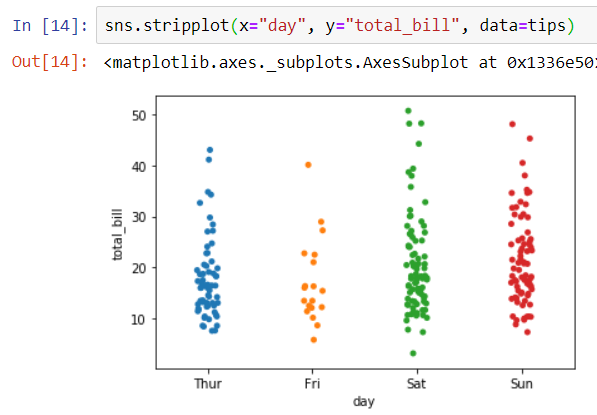
*boxplots -* використовується для візуалізації розподілу категоріальних даних за певною кількісною змінною. Кожен «ящик» демонструє інтерквартильний розмах (25%, 50% та 75%), «вуса» демонструють повний розкид даних за виключенням викидів (аномально високі чи аномально низькі значення).



Можливі варіанти для модифікації:

hue – вибір додаткової категоріальної змінної для розділення даних на окремі підгрупи.

*stripplot* – представляє собою діаграму розсіювання, де одна за змінних категоріальна.



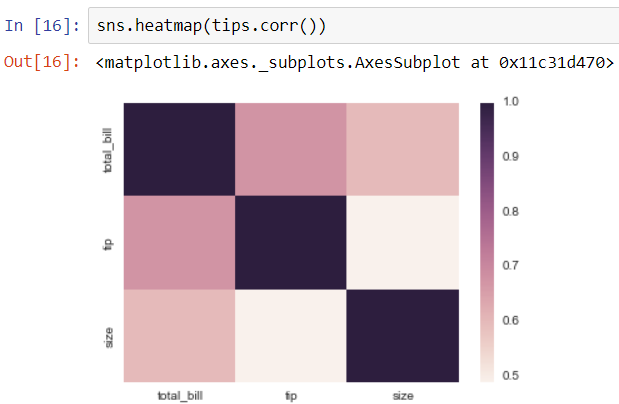
Можливі варіанти для модифікації:

hue – вибір додаткової категоріальної змінної для розділення даних на окремі підгрупи.

dodge – додається лише за умови вибору параметру hue. За замовчуванням False. Дозволяє розділити групи при візуалізації.

*Графіки для аналізу взаємозв’язку даних*

*Heatmap* – графік взаємозв’язку (кореляції) між даними у вигляді матриці. Для побудови даного графіку необхідно попередньо розрахувати матрицю кореляцій між змінними ( .corr).

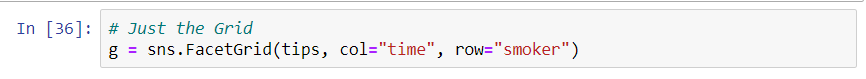


Можливі варіанти для модифікації:

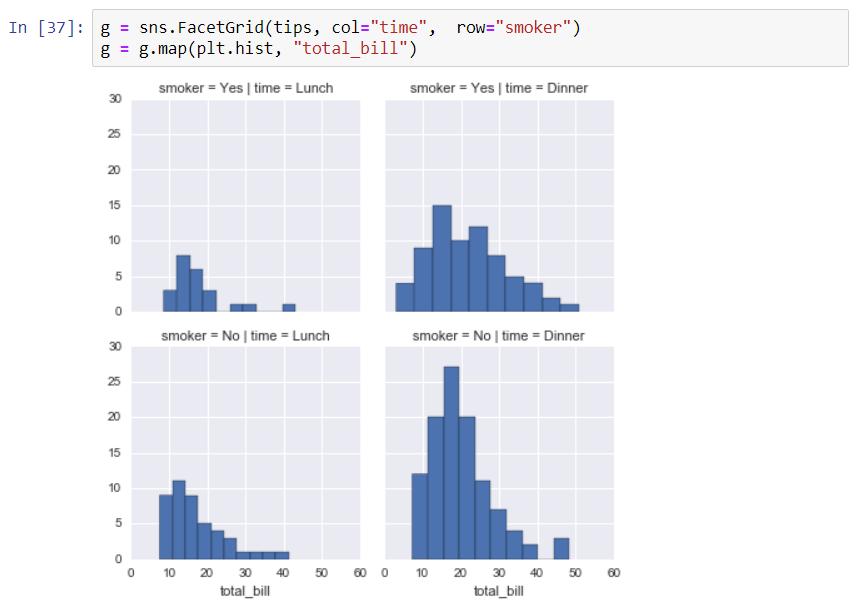
annot – додавання коефіцієнтів кореляції на діаграму.

*Графіки для відображення даних у вигляді сітки*

1. Побудуйте площу для відображення даних, обравши одну або дві кількісні змінні.



2. Додайте на площу дані для відображення:



Можливі варіанти для модифікації:

hue – вибір додаткової категоріальної змінної для розділення даних на окремі підгрупи.

.add\_legend() – додавання легенди до діаграми