



Chapter 7 - ex 3: Tips

Cho dữ liệu tips có sẵn trong seaborn library. Hãy vẽ những biểu đồ sau:

1. Vẽ violinplot cho cho cột total_bill
2. Vẽ swarmplot cho cột total_bill theo sex
3. Vẽ boxplot cho cột total_bill
4. Tạo FacetGrid với 'time' và chỉ định thứ tự của các hàng bằng row_order, ánh xạ (map) của 'total_bill' lên lưới
5. Tạo Factor plot (phiên bản mới là catplot) chứa point plot của giá trị 'total_bill'
6. Tạo PairGrid với một scatter plot "total_bill" và "tip"
7. Tạo Pairplot với một scatter plot "total_bill" và "tip", sử dụng palette color = 'day'

```
In [1]: import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
```

```
In [2]: # Load the data
tips = sns.load_dataset("tips")
tips.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
total_bill    244 non-null float64
tip           244 non-null float64
sex           244 non-null category
smoker        244 non-null category
day           244 non-null category
time          244 non-null category
size          244 non-null int64
dtypes: category(4), float64(2), int64(1)
memory usage: 7.2 KB
```

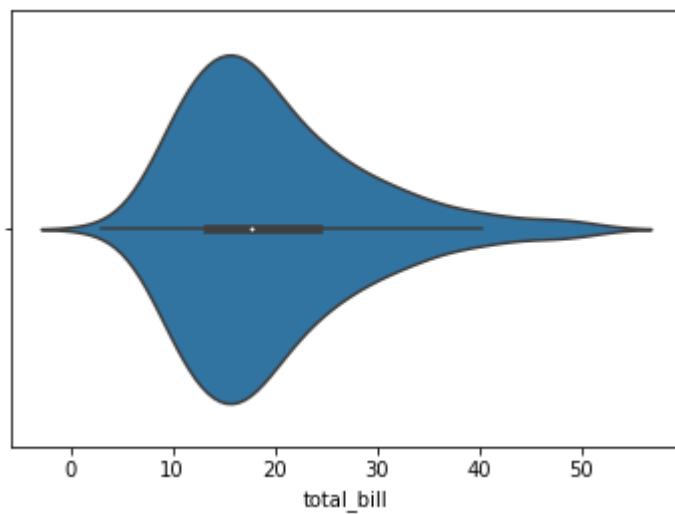
```
In [3]: tips.head()
```

```
Out[3]:
```

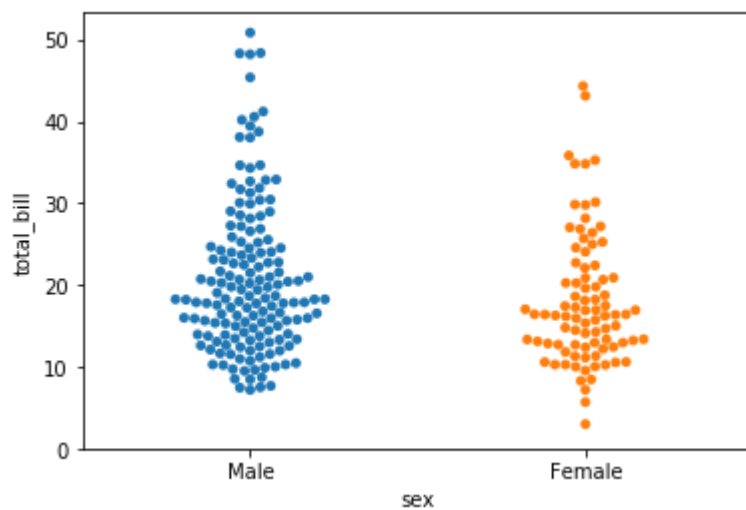
	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
In [4]: # Create violinplot
sns.violinplot(x = "total_bill", data=tips)

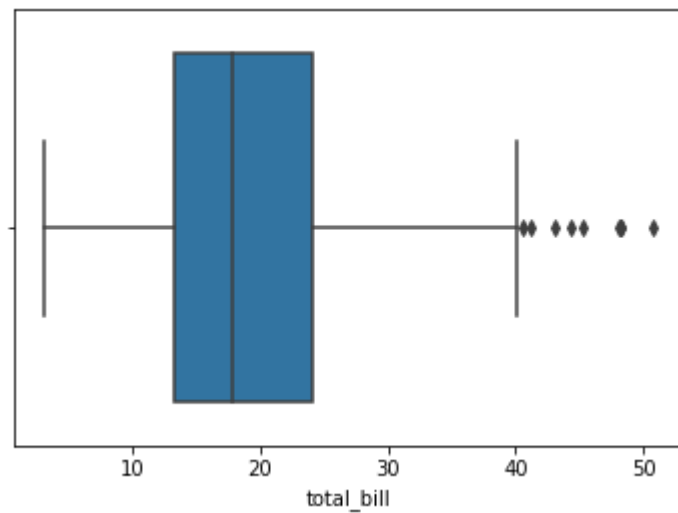
# Show the plot
plt.show()
```



```
In [5]: # Construct swarmplot
sns.swarmplot(x="sex", y="total_bill", data=tips)
plt.show()
```



```
In [6]: sns.boxplot(x="total_bill", data=tips)
plt.show()
```



```
In [7]: from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
le.fit(tips["smoker"])
tips["smoker"] = le.transform(tips["smoker"])
tips.head()
```

```
Out[7]:
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	0	Sun	Dinner	2
1	10.34	1.66	Male	0	Sun	Dinner	3
2	21.01	3.50	Male	0	Sun	Dinner	3
3	23.68	3.31	Male	0	Sun	Dinner	2
4	24.59	3.61	Female	0	Sun	Dinner	4

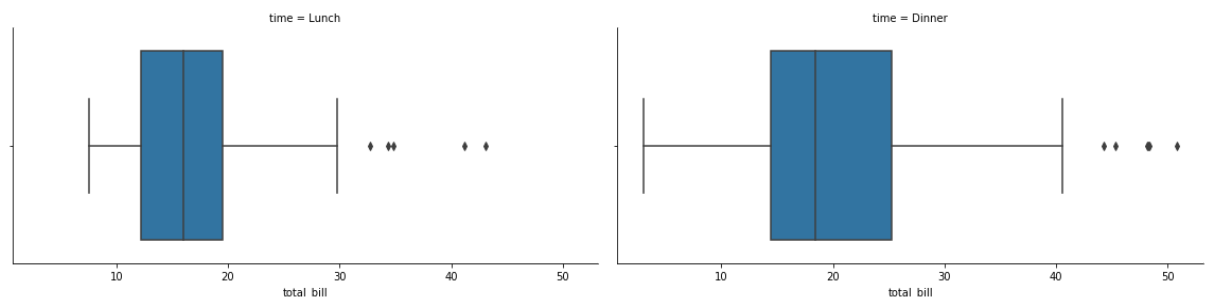
```
In [29]: # Create FacetGrid with 'time' and specify the order of the rows using row_order
plt.figure(figsize=(8,6))
g2 = sns.FacetGrid(tips,
                    col="time",
                    row_order=['Lunch', 'Dinner'],
                    height=4, aspect=2)
# Map a boxplot of 'total_bill' onto the grid
g2.map(sns.boxplot, 'total_bill')

# Show the plot
plt.show()
plt.clf()
```

c:\program files\python36\lib\site-packages\seaborn\axisgrid.py:715: UserWarning: Using the boxplot function without specifying `order` is likely to produce an incorrect plot.

warnings.warn(warning)

<Figure size 576x432 with 0 Axes>



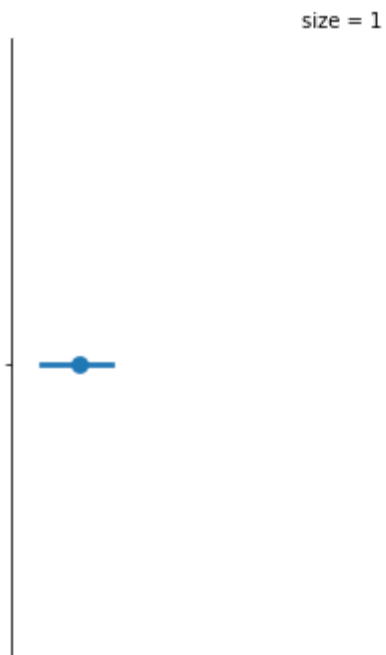
<Figure size 432x288 with 0 Axes>

Note:

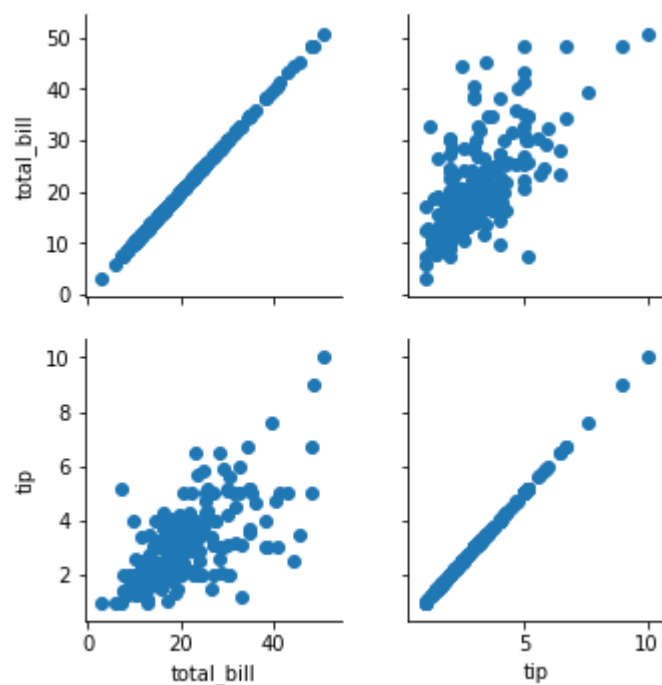
- height: changes the height, while maintaining the aspect ratio (so it will also get wider if only size is changed.)
- aspect: changes the width while keeping the height constant.

```
In [21]: # Create a factor plot that contains point plot of total_bill values
sns.catplot(data=tips,
            x='total_bill',
            kind='point',
            row='size')

plt.show()
plt.clf()
```

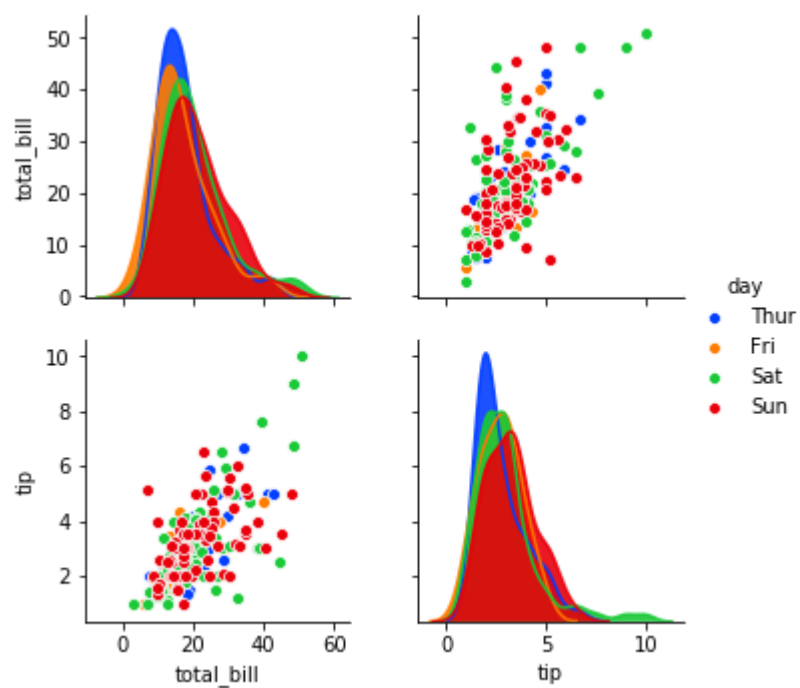


```
In [23]: # Create a PairGrid with a scatter plot "total_bill", "tip"  
g = sns.PairGrid(tips, vars=["total_bill", "tip"])  
g2 = g.map(plt.scatter)  
plt.show()  
plt.clf()
```



<Figure size 432x288 with 0 Axes>

```
In [28]: # Plot a pairplot and use a different color palette and color code by 'day'
sns.pairplot(data=tips,
             vars=["total_bill", "tip"],
             kind='scatter',
             hue='day',
             palette='bright',
             diag_kws={'alpha':.9})
plt.show()
plt.clf()
```



<Figure size 432x288 with 0 Axes>

In []: