Lecture 7

May 23, 2023

1 Seaborn

1.1 Importing library

```
[1]: import seaborn as sns
```

1.2 Import bill dataset

```
[5]: print(sns.get_dataset_names())
     data = sns.load_dataset('tips')
     print(data)
```

```
['anagrams', 'anscombe', 'attention', 'brain_networks', 'car_crashes',
'diamonds', 'dots', 'dowjones', 'exercise', 'flights', 'fmri', 'geyser', 'glue',
'healthexp', 'iris', 'mpg', 'penguins', 'planets', 'seaice', 'taxis', 'tips',
'titanic']
     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
..          ...   ...     ...    ...   ...     ...   ...
239       29.03  5.92    Male     No   Sat  Dinner     3
240       27.18  2.00  Female    Yes   Sat  Dinner     2
241       22.67  2.00    Male    Yes   Sat  Dinner     2
242       17.82  1.75    Male     No   Sat  Dinner     2
243       18.78  3.00  Female     No  Thur  Dinner     2

[244 rows x 7 columns]
```

1.3 Distribution plot

```
[6]: sns.distplot(data['total_bill'])
```

```
/var/folders/03/k1p5_v6d69bg7b999gdktlgw0000gn/T/ipykernel_5033/4292925720.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```
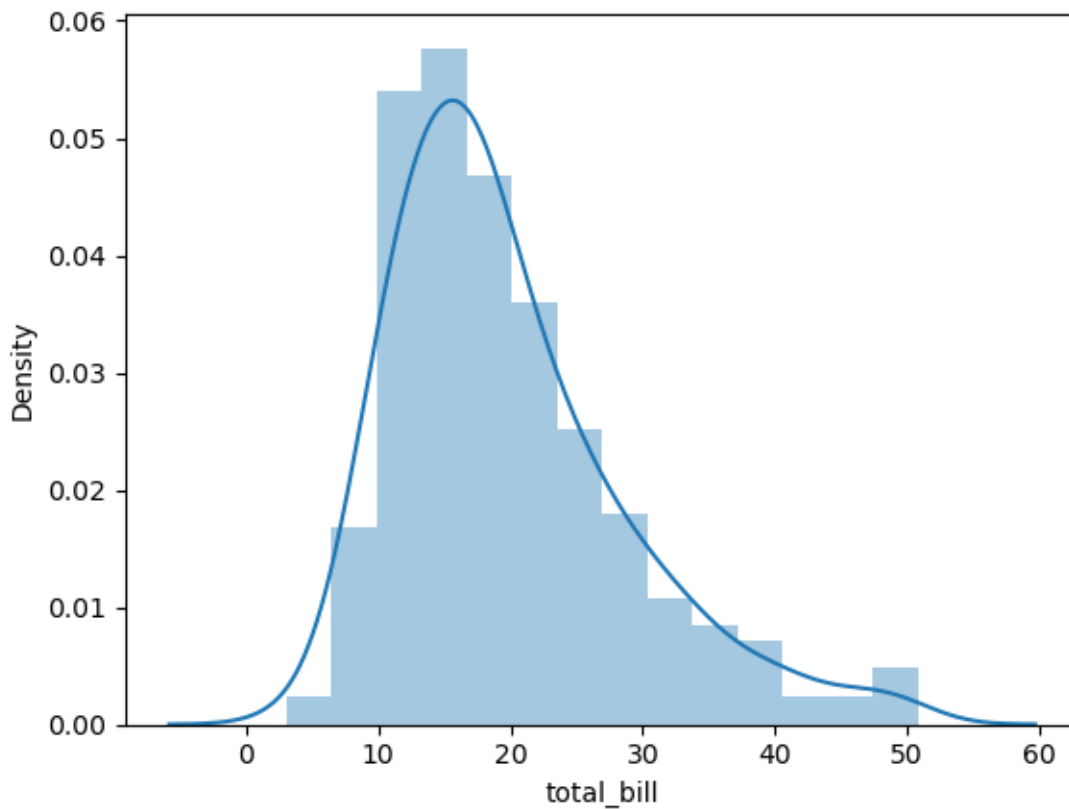
Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
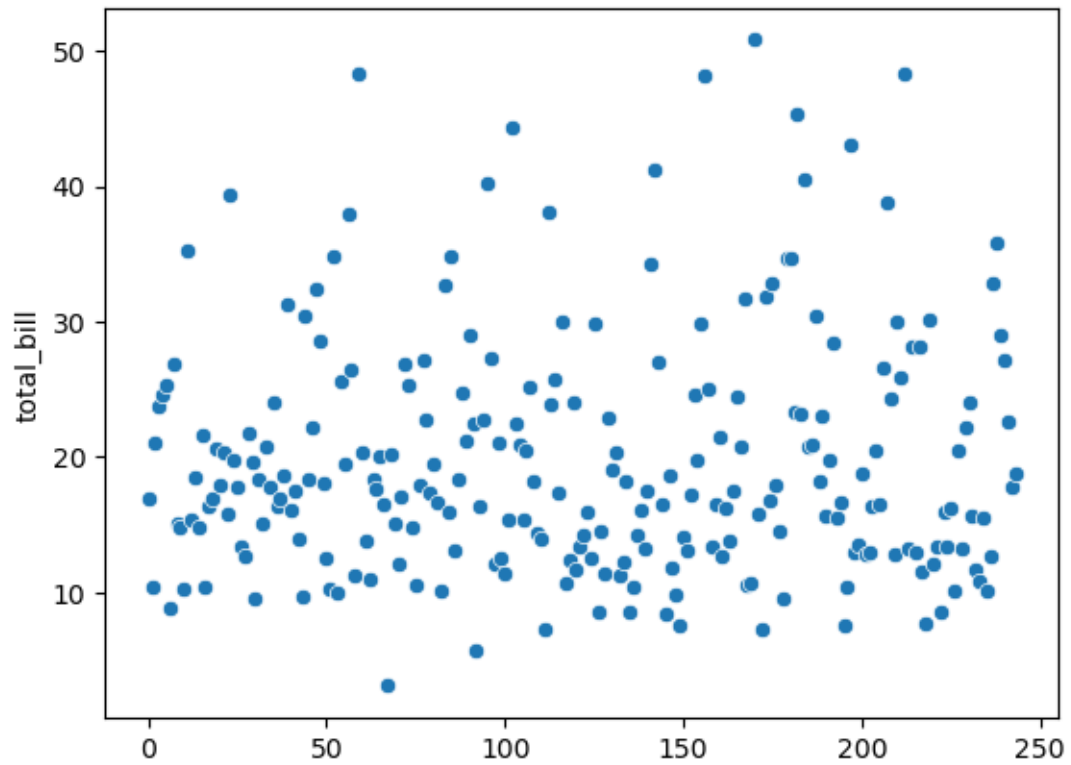
```
sns.distplot(data['total_bill'])
```

[6]: <Axes: xlabel='total_bill', ylabel='Density'>

1.4 Scatter plot

```
[7]: sns.scatterplot(data['total_bill'])
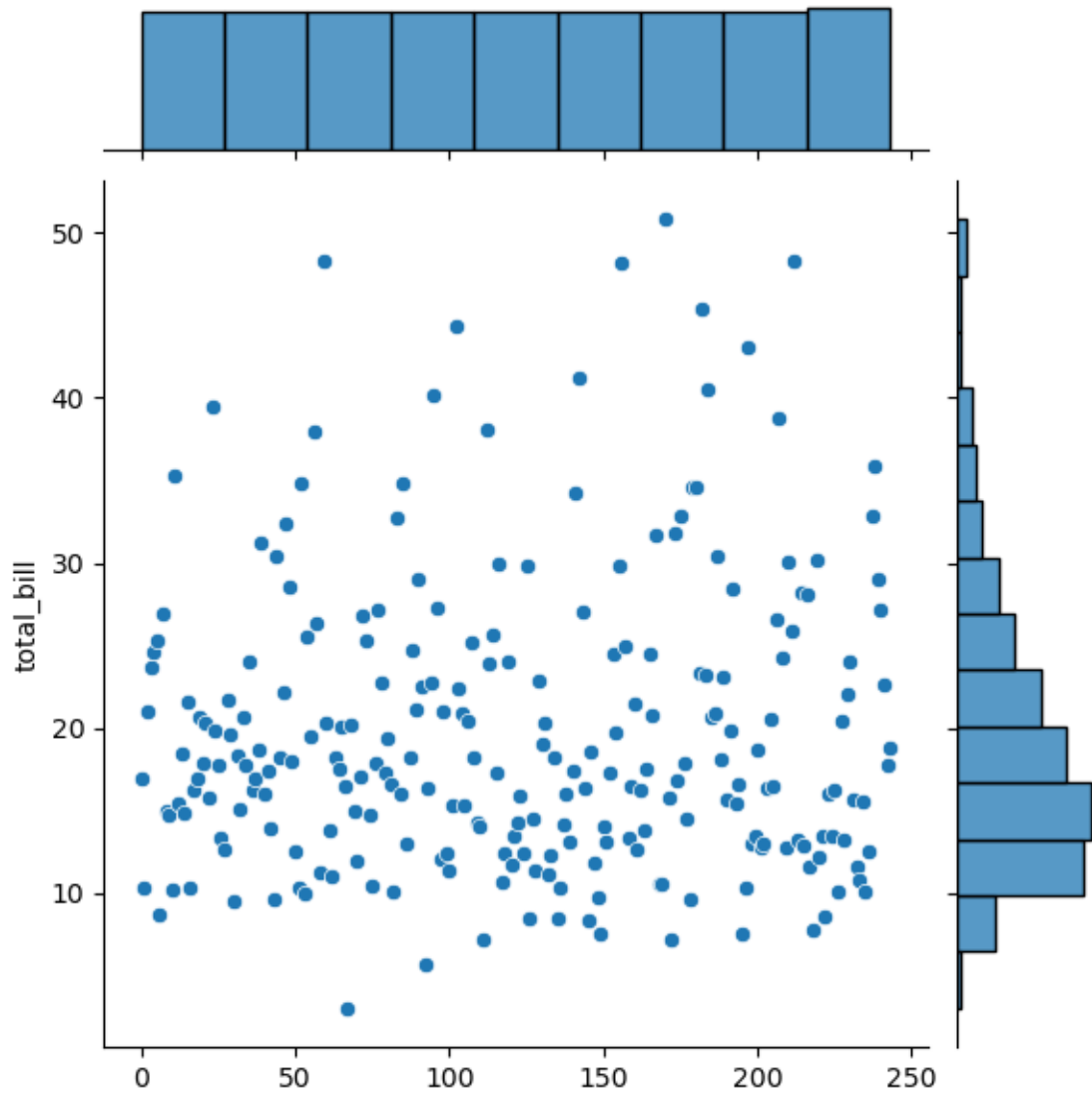```

[7]: <Axes: ylabel='total_bill'>

## 1.5 Joint plot

```
[8]: sns.jointplot(data['total_bill'])
```
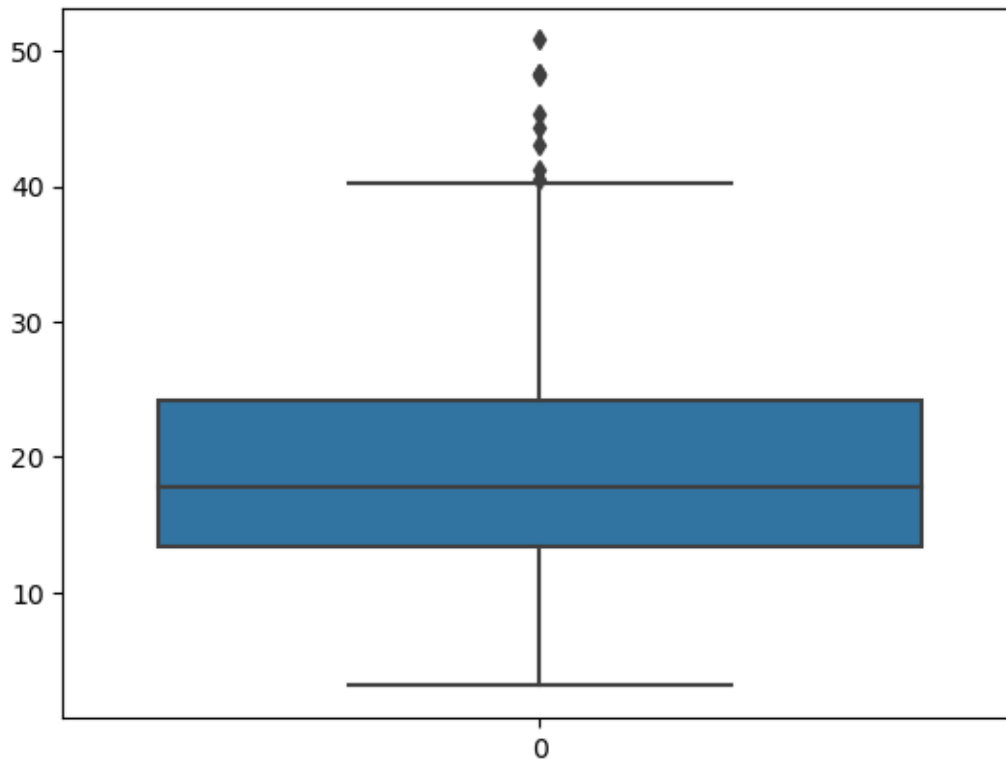
```
[8]: <seaborn.axisgrid.JointGrid at 0x13ba7c8e0>
```

## 1.6 Box plot

```
[9]: sns.boxplot(data['total_bill'])
```

```
[9]: <Axes: >
```

2 Bar plot

```
[23]: sns.barplot(x=data.day, y=data.total_bill)
```
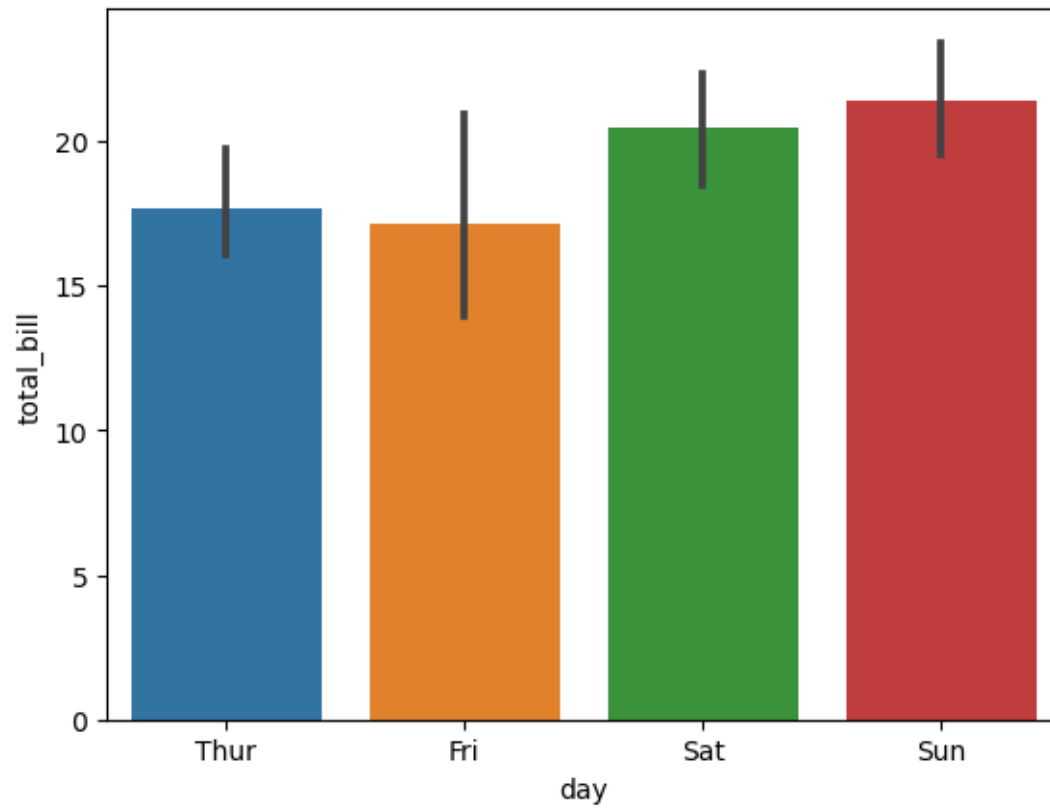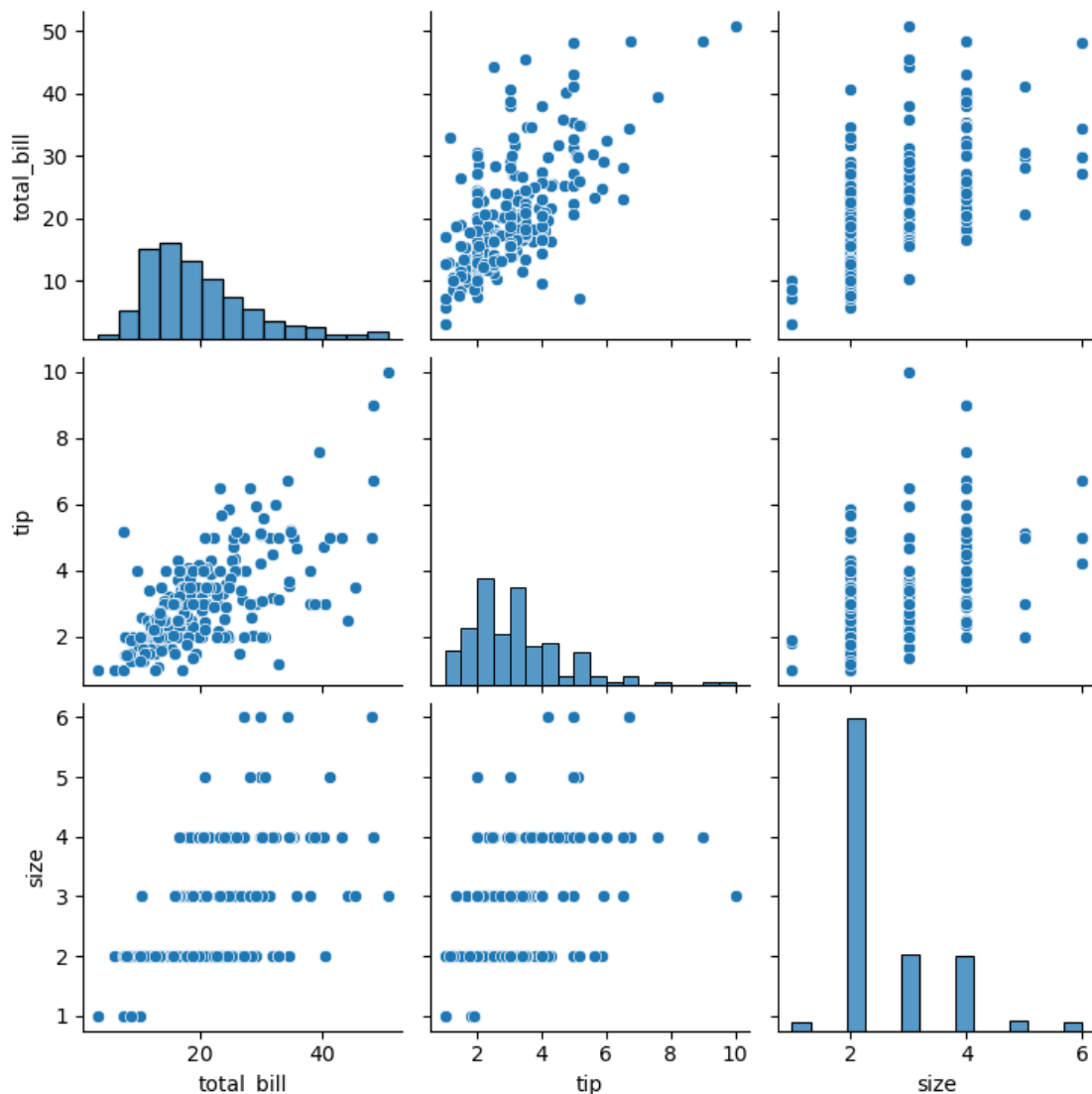
```
[23]: <Axes: xlabel='day', ylabel='total_bill'>
```

2.1 Pair plot

```
[13]: sns.pairplot(data)
```

```
[13]: <seaborn.axisgrid.PairGrid at 0x13f5ac3a0>
```

2.2 Heat map

```
[15]: data.corr()  #correlation matrix
```

/var/folders/03/k1p5_v6d69bg7b999gdktlgw0000gn/T/ipykernel_5033/2818546882.py:1:
FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only valid
columns or specify the value of numeric_only to silence this warning.
   data.corr() #correlation matrix

```
[15]:             total_bill       tip      size
      total_bill    1.000000  0.675734  0.598315
      tip           0.675734  1.000000  0.489299
```

```
size            0.598315  0.489299  1.000000
```

[20]: `sns.heatmap(data.corr())`

/var/folders/03/k1p5_v6d69bg7b999gdktlgw0000gn/T/ipykernel_5033/1794321463.py:1:
FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only valid
columns or specify the value of numeric_only to silence this warning.
    sns.heatmap(data.corr())

[20]: `<Axes: >`