seaborn

July 7, 2024

Is a data visualization library

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

Importing dataset directly from seaborn

total_bill vs tip data set

[2]:
```
tips = sns.load_dataset('tips')
```

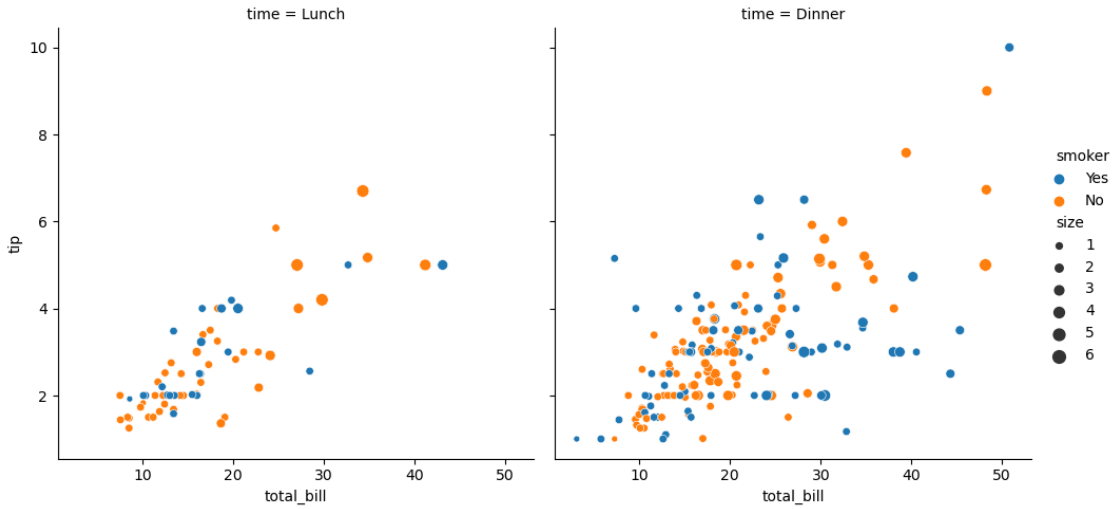
[3]:
```
tips.head()
```

[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

[4]:
```
sns.relplot(data=tips, x='total_bill', y='tip', col='time', hue='smoker',␣
↪size='size');
```

```
C:\Users\DONATUS\anaconda24\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

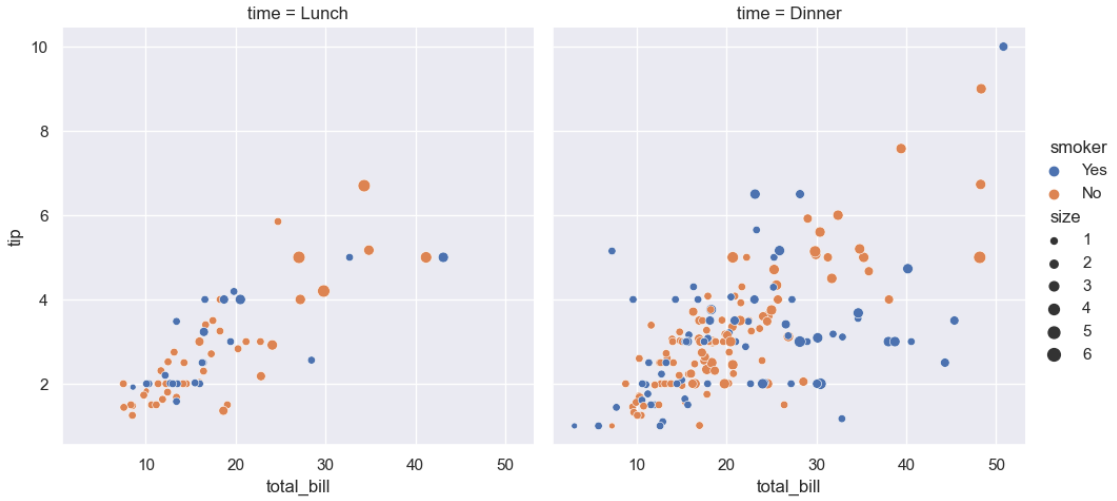


Setting Teams for the plot

[5]:
```
sns.set_theme()
```

[6]:
```
sns.relplot(data=tips, x='total_bill', y='tip', col='time', hue='smoker',␣
↪size='size');
```

```
C:\Users\DONATUS\anaconda24\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



[7]:
```
#loadidng iris data set
```

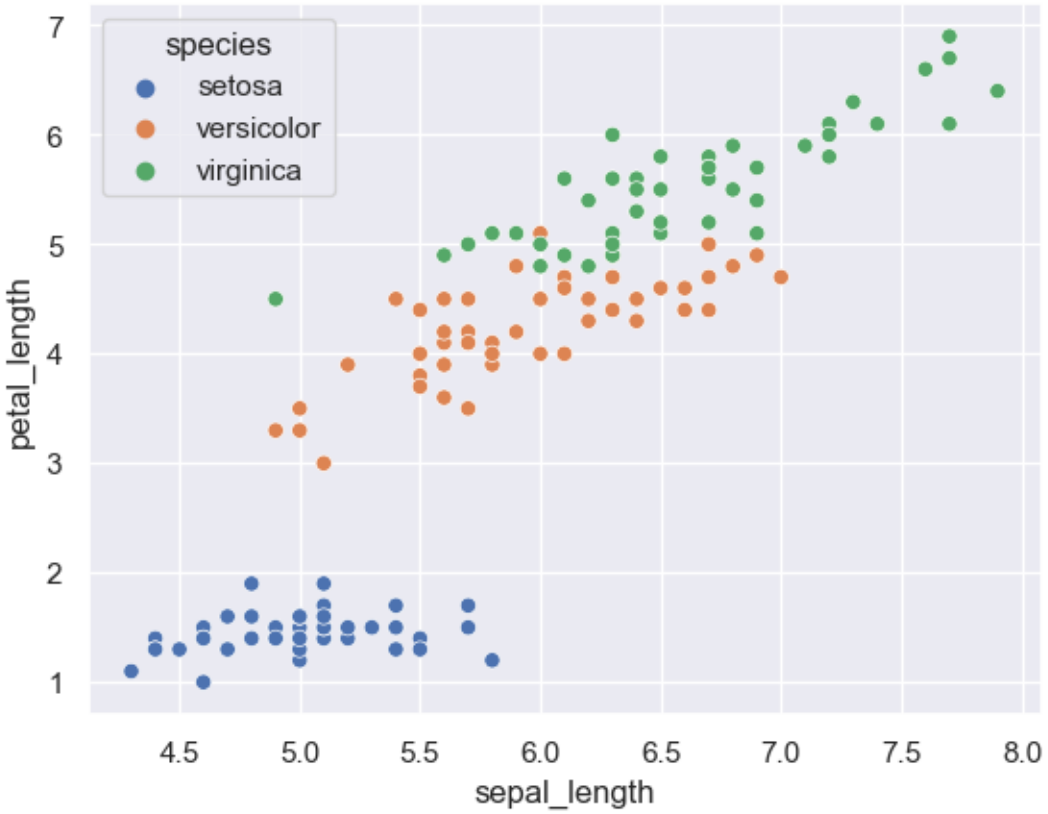
```
iris =sns.load_dataset('iris')
iris.head()
```

[7]:

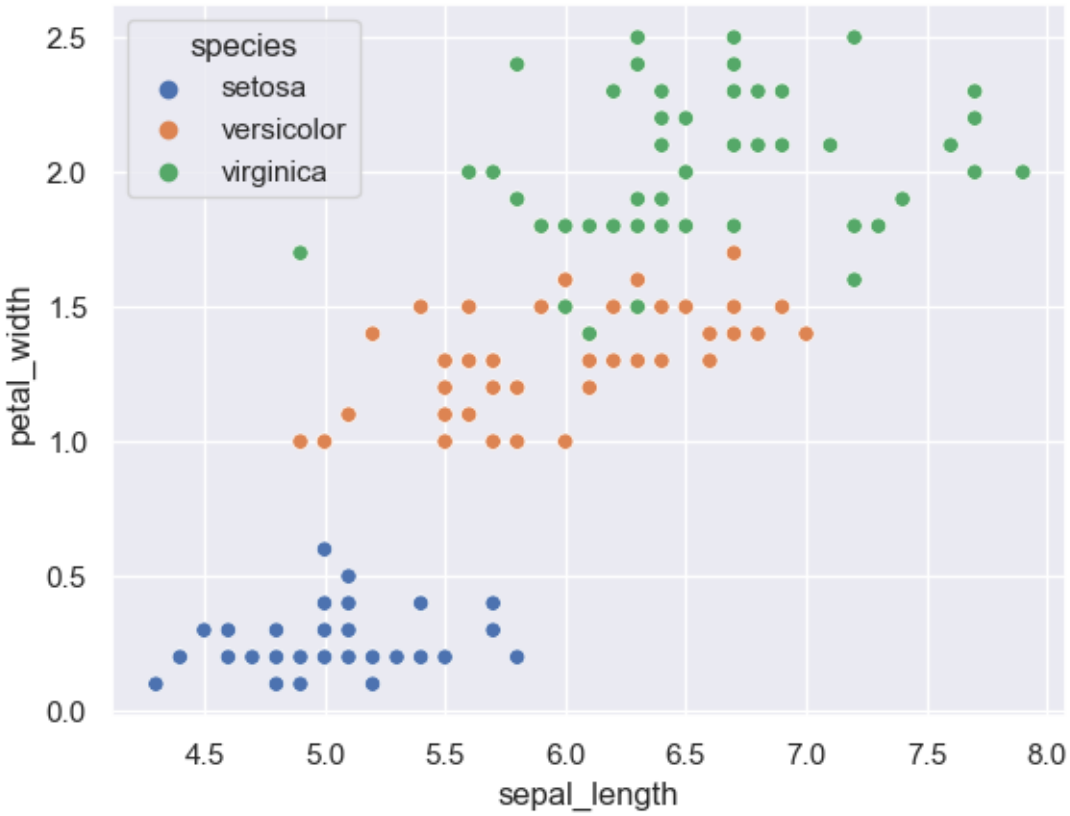
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Using a scatterplot

[8]:
```
sns.scatterplot(x='sepal_length', y='petal_length', hue = 'species', data =␣
↪iris);
```



[9]:
```
sns.scatterplot(x='sepal_length', y='petal_width', hue ='species', data = iris);
```



[10]:
```
#loading the titanic datasset

titanic = sns.load_dataset('titanic')
titanic.head()
```

[10]:

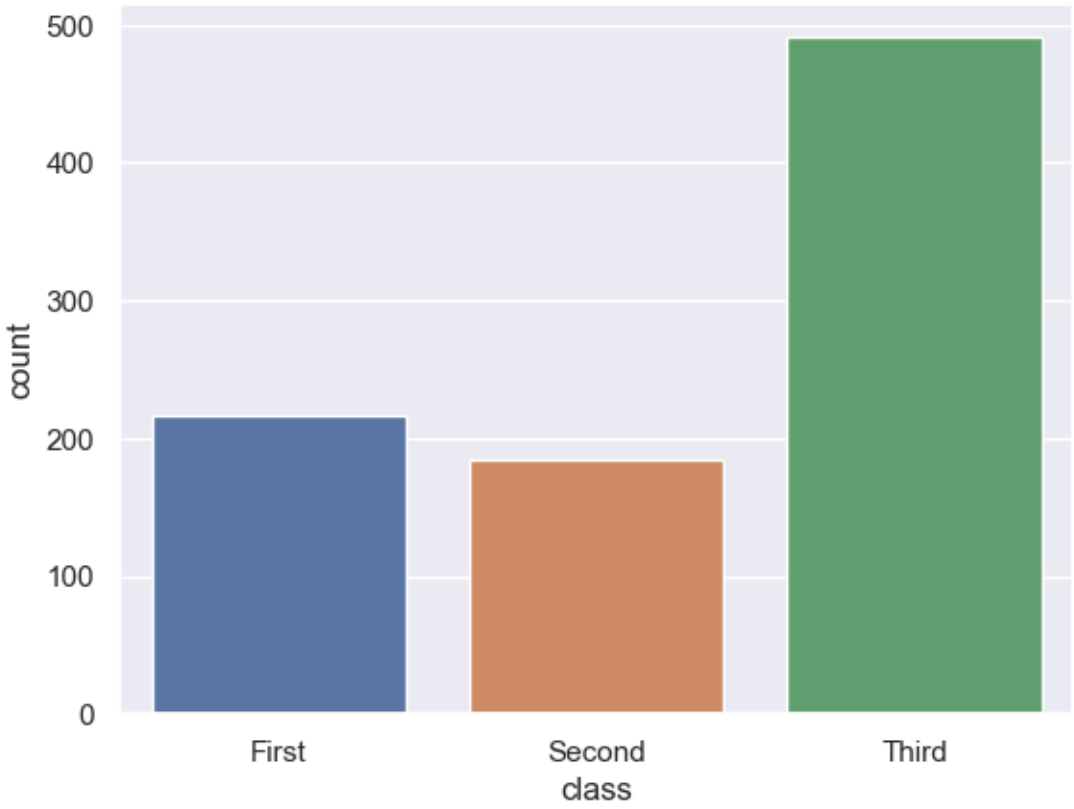
	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True

[11]:
```
titanic.shape
```

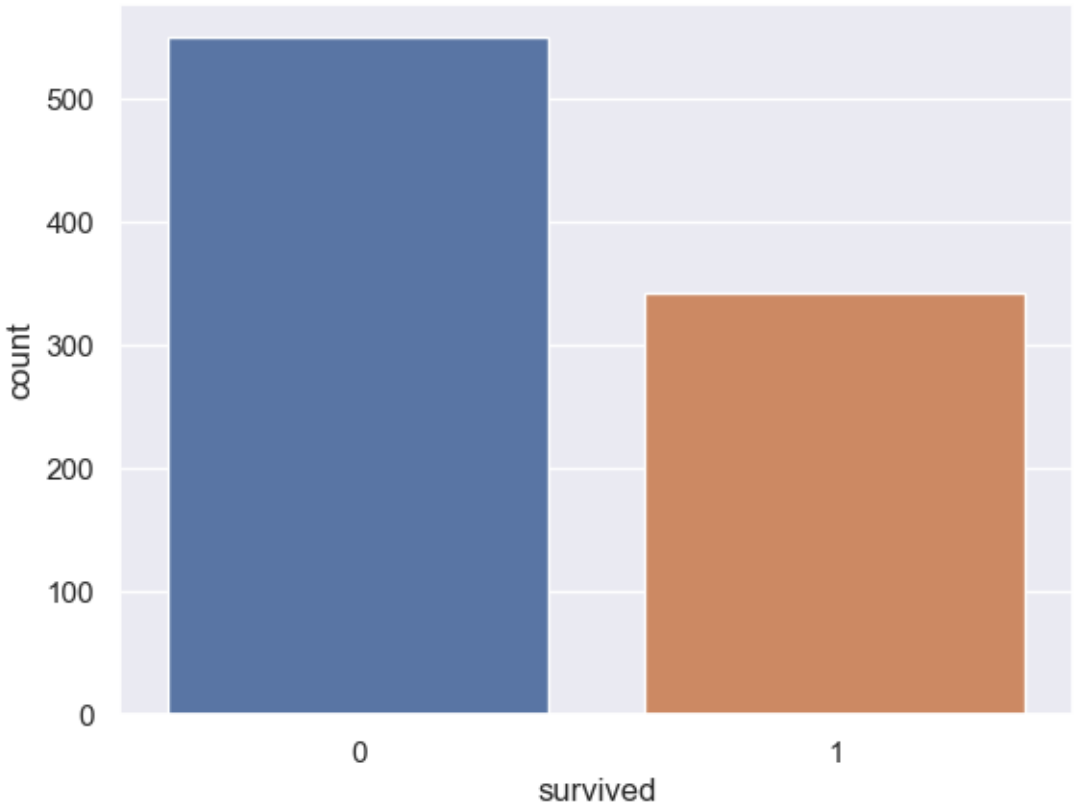
[11]: (891, 15)

Making a Count Plot

[12]:
```
sns.countplot(x='class', data = titanic);
```

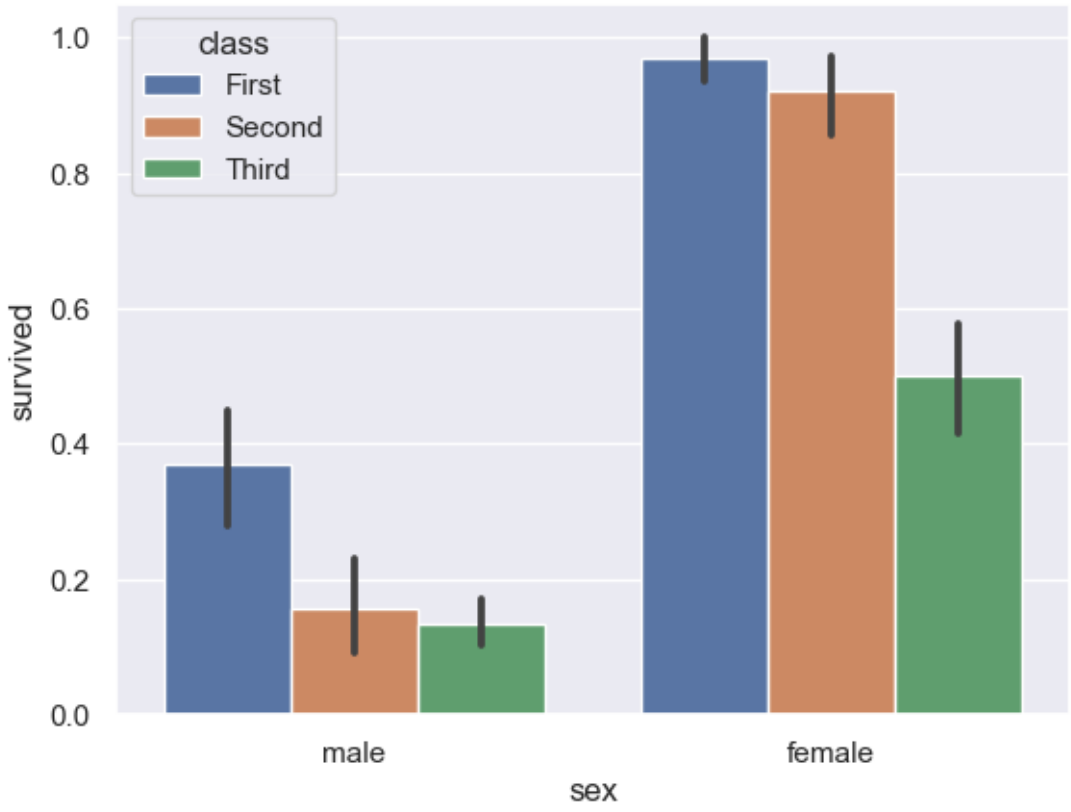


[13]:
```
sns.countplot(x='survived', data = titanic);
```



Bar Chart

[14]:
```
sns.barplot(x='sex', y='survived', hue ='class', data=titanic);
```



Using house price dataset

[15]:
```
boston = pd.read_csv('boston_house_prices.csv', header = 1)
boston.head()
```

[15]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	\
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	

	B	LSTAT	MEDV
0	396.90	4.98	24.0
1	396.90	9.14	21.6
2	392.83	4.03	34.7
3	394.63	2.94	33.4
4	396.90	5.33	36.2

[16]:
```
# Renaming a column in-place
boston.rename(columns={'MEDV': 'PRICE'}, inplace=True)
```

[25]:
```
boston.head()
```

[25]:

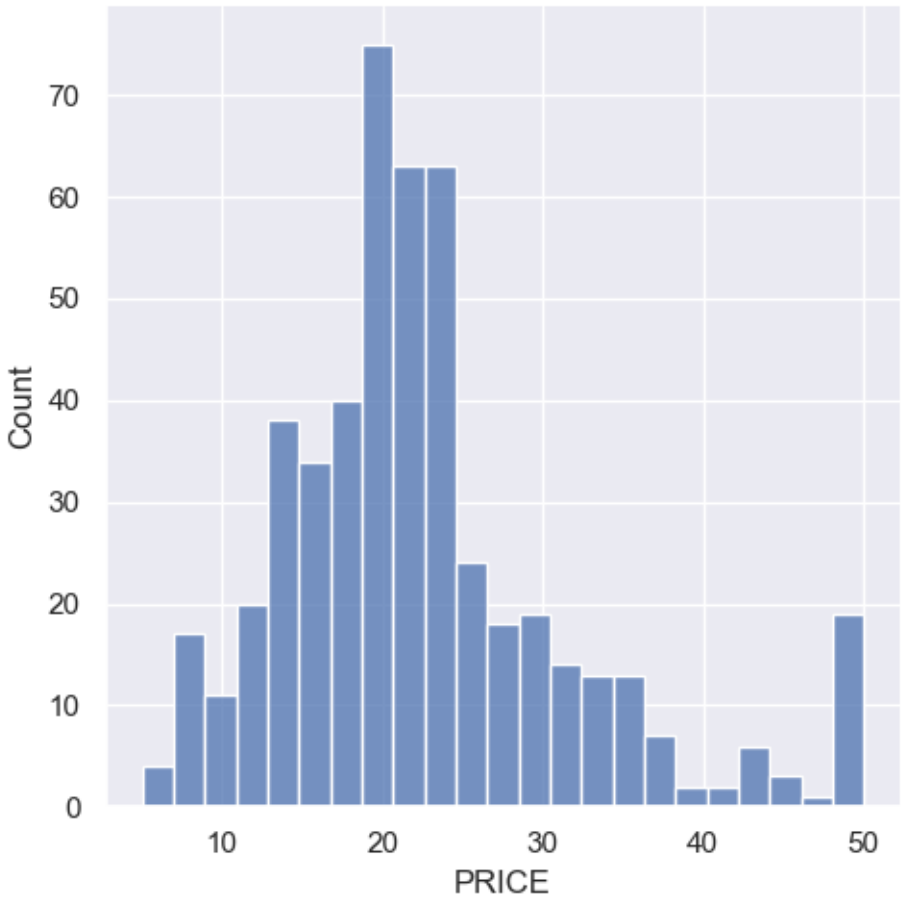
	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	\
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	

	B	LSTAT	PRICE
0	396.90	4.98	24.0
1	396.90	9.14	21.6
2	392.83	4.03	34.7
3	394.63	2.94	33.4
4	396.90	5.33	36.2

Looking at the distribution plot

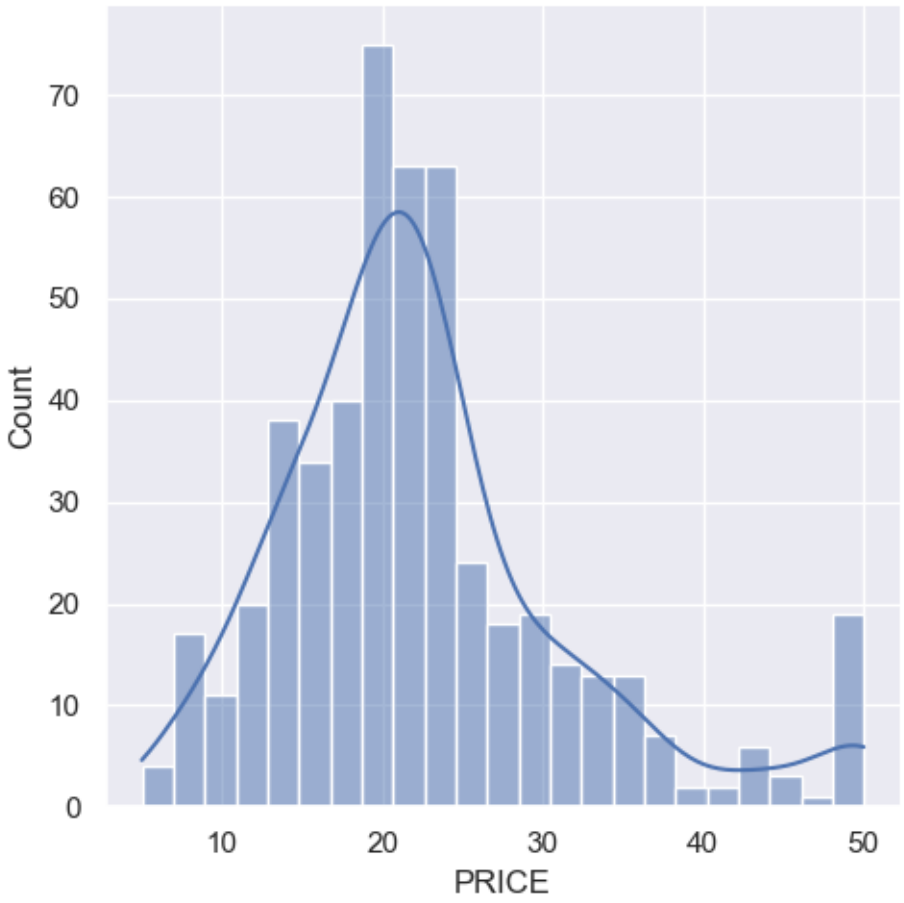
[18]:
```
sns.displot(boston['PRICE']);
```

```
C:\Users\DONATUS\anaconda24\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



[24]:
```
sns.displot(boston['PRICE'], kde=True);
```

```
C:\Users\DONATUS\anaconda24\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



CORRELATION -Positive -Negative

Using heat map to plot the correlation

[28]:
```
correlation = boston.corr()
correlation
```

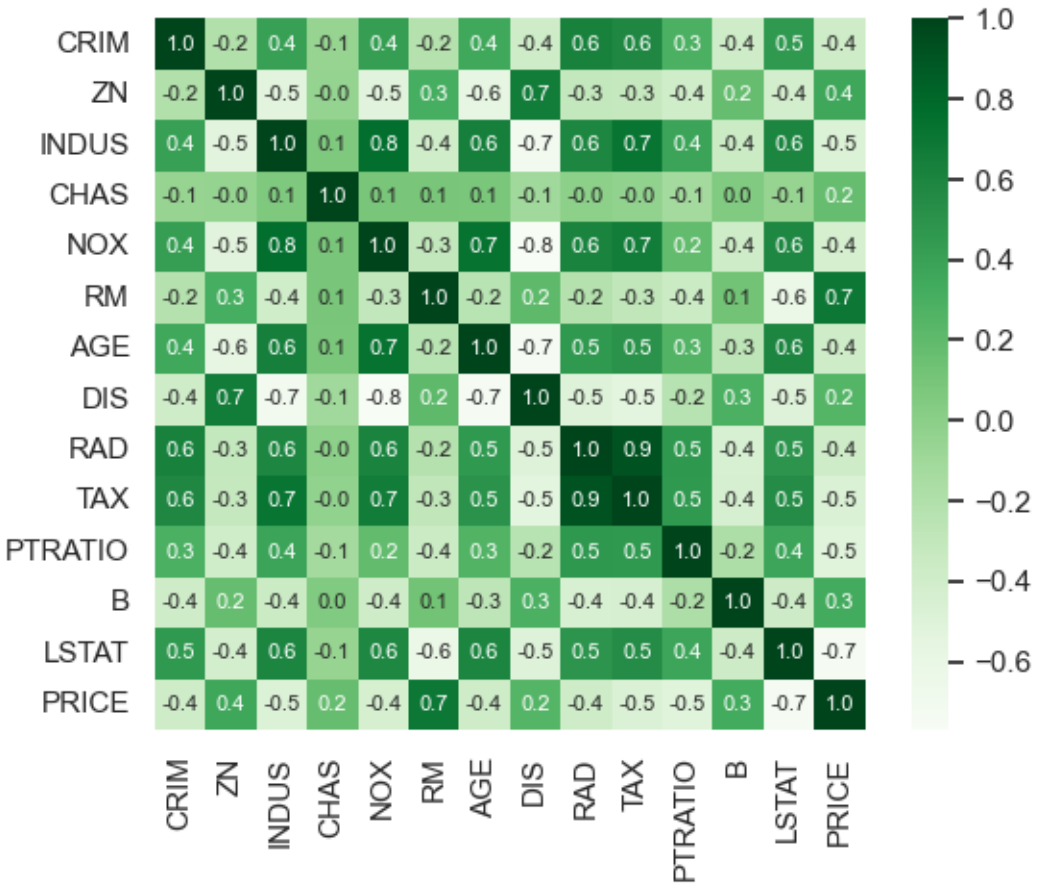
[28]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	\
CRIM	1.000000	-0.200469	0.406583	-0.055892	0.420972	-0.219247	0.352734	
ZN	-0.200469	1.000000	-0.533828	-0.042697	-0.516604	0.311991	-0.569537	
INDUS	0.406583	-0.533828	1.000000	0.062938	0.763651	-0.391676	0.644779	
CHAS	-0.055892	-0.042697	0.062938	1.000000	0.091203	0.091251	0.086518	
NOX	0.420972	-0.516604	0.763651	0.091203	1.000000	-0.302188	0.731470	
RM	-0.219247	0.311991	-0.391676	0.091251	-0.302188	1.000000	-0.240265	
AGE	0.352734	-0.569537	0.644779	0.086518	0.731470	-0.240265	1.000000	
DIS	-0.379670	0.664408	-0.708027	-0.099176	-0.769230	0.205246	-0.747881	
RAD	0.625505	-0.311948	0.595129	-0.007368	0.611441	-0.209847	0.456022	
TAX	0.582764	-0.314563	0.720760	-0.035587	0.668023	-0.292048	0.506456	
PTRATIO	0.289946	-0.391679	0.383248	-0.121515	0.188933	-0.355501	0.261515	

```
B       -0.385064  0.175520 -0.356977  0.048788 -0.380051  0.128069 -0.273534
LSTAT    0.455621 -0.412995  0.603800 -0.053929  0.590879 -0.613808  0.602339
PRICE   -0.388305  0.360445 -0.483725  0.175260 -0.427321  0.695360 -0.376955

               DIS       RAD       TAX   PTRATIO         B     LSTAT     PRICE
CRIM    -0.379670  0.625505  0.582764  0.289946 -0.385064  0.455621 -0.388305
ZN       0.664408 -0.311948 -0.314563 -0.391679  0.175520 -0.412995  0.360445
INDUS   -0.708027  0.595129  0.720760  0.383248 -0.356977  0.603800 -0.483725
CHAS    -0.099176 -0.007368 -0.035587 -0.121515  0.048788 -0.053929  0.175260
NOX     -0.769230  0.611441  0.668023  0.188933 -0.380051  0.590879 -0.427321
RM       0.205246 -0.209847 -0.292048 -0.355501  0.128069 -0.613808  0.695360
AGE     -0.747881  0.456022  0.506456  0.261515 -0.273534  0.602339 -0.376955
DIS      1.000000 -0.494588 -0.534432 -0.232471  0.291512 -0.496996  0.249929
RAD     -0.494588  1.000000  0.910228  0.464741 -0.444413  0.488676 -0.381626
TAX     -0.534432  0.910228  1.000000  0.460853 -0.441808  0.543993 -0.468536
PTRATIO -0.232471  0.464741  0.460853  1.000000 -0.177383  0.374044 -0.507787
B        0.291512 -0.444413 -0.441808 -0.177383  1.000000 -0.366087  0.333461
LSTAT   -0.496996  0.488676  0.543993  0.374044 -0.366087  1.000000 -0.737663
PRICE    0.249929 -0.381626 -0.468536 -0.507787  0.333461 -0.737663  1.000000
```

[49]:
```
sns.heatmap(correlation,
            cbar=True,  # Show the colorbar
            square=True,  # Make the heatmap square
            fmt='.1f',  # Format the values in the heatmap
            annot=True,  # Add annotations to the heatmap
            annot_kws={'size': 8},  # Set the annotation font size
            cmap='Greens');  # Set the colormap to Blues
```


CRIM 1.0 -0.2 0.4 -0.1 0.4 -0.2 0.4 -0.4 0.6 0.6 0.3 -0.4 0.5 -0.4
ZN -0.2 1.0 -0.5 -0.0 -0.5 0.3 -0.6 0.7 -0.3 -0.3 -0.4 0.2 -0.4 0.4
INDUS 0.4 -0.5 1.0 0.1 0.8 -0.4 0.6 -0.7 0.6 0.7 0.4 -0.4 0.6 -0.5
CHAS -0.1 -0.0 0.1 1.0 0.1 0.1 0.1 -0.1 -0.0 -0.0 -0.1 0.0 -0.1 0.2
NOX 0.4 -0.5 0.8 0.1 1.0 -0.3 0.7 -0.8 0.6 0.7 0.2 -0.4 0.6 -0.4
RM -0.2 0.3 -0.4 0.1 -0.3 1.0 -0.2 0.2 -0.2 -0.3 -0.4 0.1 -0.6 0.7
AGE 0.4 -0.6 0.6 0.1 0.7 -0.2 1.0 -0.7 0.5 0.5 0.3 -0.3 0.6 -0.4
DIS -0.4 0.7 -0.7 -0.1 -0.8 0.2 -0.7 1.0 -0.5 -0.5 -0.2 0.3 -0.5 0.2
RAD 0.6 -0.3 0.6 -0.0 0.6 -0.2 0.5 -0.5 1.0 0.9 0.5 -0.4 0.5 -0.4
TAX 0.6 -0.3 0.7 -0.0 0.7 -0.3 0.5 -0.5 0.9 1.0 0.5 -0.4 0.5 -0.5
PTRATIO 0.3 -0.4 0.4 -0.1 0.2 -0.4 0.3 -0.2 0.5 0.5 1.0 -0.2 0.4 -0.5
B -0.4 0.2 -0.4 0.0 -0.4 0.1 -0.3 0.3 -0.4 -0.4 -0.2 1.0 -0.4 0.3
LSTAT 0.5 -0.4 0.6 -0.1 0.6 -0.6 0.6 -0.5 0.5 0.5 0.4 -0.4 1.0 -0.7
PRICE -0.4 0.4 -0.5 0.2 -0.4 0.7 -0.4 0.2 -0.4 -0.5 -0.5 0.3 -0.7 1.0
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT PRICE
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4 −0.6

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: