```
In [1]:  import pandas as pd
         import numpy as np
         import seaborn as sns
```

```
In [3]:  data = pd.read_csv(r"D:\Data Analysis Project\StudentsPerformance.csv")
```

```
In [ ]:  #understanding the data
```

```
In [4]:  data.head()
```

Out[4]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75

```
In [5]:  data.tail()
```

Out[5]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

```
In [6]:  data.shape
```

Out[6]: (1000, 8)

```
In [7]:  data.describe()
```

Out[7]:

	math score	reading score	writing score
count	1000.00000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000
std	15.16308	14.600192	15.195657
min	0.00000	17.000000	10.000000
25%	57.00000	59.000000	57.750000
50%	66.00000	70.000000	69.000000
75%	77.00000	79.000000	79.000000
max	100.00000	100.000000	100.000000

In [8]:
```python
data.columns
```

Out[8]:
```
Index(['gender', 'race/ethnicity', 'parental level of education', 'lunch',
       'test preparation course', 'math score', 'reading score',
       'writing score'],
      dtype='object')
```

In [9]:
```python
data.nunique()
```

Out[9]:
```
gender                          2
race/ethnicity                  5
parental level of education     6
lunch                           2
test preparation course         2
math score                     81
reading score                  72
writing score                  77
dtype: int64
```

In [22]:
```python
data['gender'].unique()
```

Out[22]:
```
array(['female', 'male'], dtype=object)
```

In [23]:
```python
#cleaning the data
```

In [45]:
```python
data.isnull().sum()
```

Out[45]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test preparation course        0
math score                     0
reading score                  0
writing score                  0
dtype: int64
```

In [46]:
```python
student = data.drop(['race/ethnicity','parental level of education'], axis=1)
```

In [47]:
```python
student.head()
```

Out[47]:

	gender	lunch	test preparation course	math score	reading score	writing score
0	female	standard	none	72	72	74
1	female	standard	completed	69	90	88
2	female	standard	none	90	95	93
3	male	free/reduced	none	47	57	44
4	male	standard	none	76	78	75

In [48]:
```
#3 relationship analysis
```

In [49]:
```
student = student.select_dtypes(include='number')
correlation_matrix = student.corr()
```

In [50]:
```
sns.heatmap(correlation_matrix,xticklabels=correlation_matrix.columns,yticklabels=correla
```
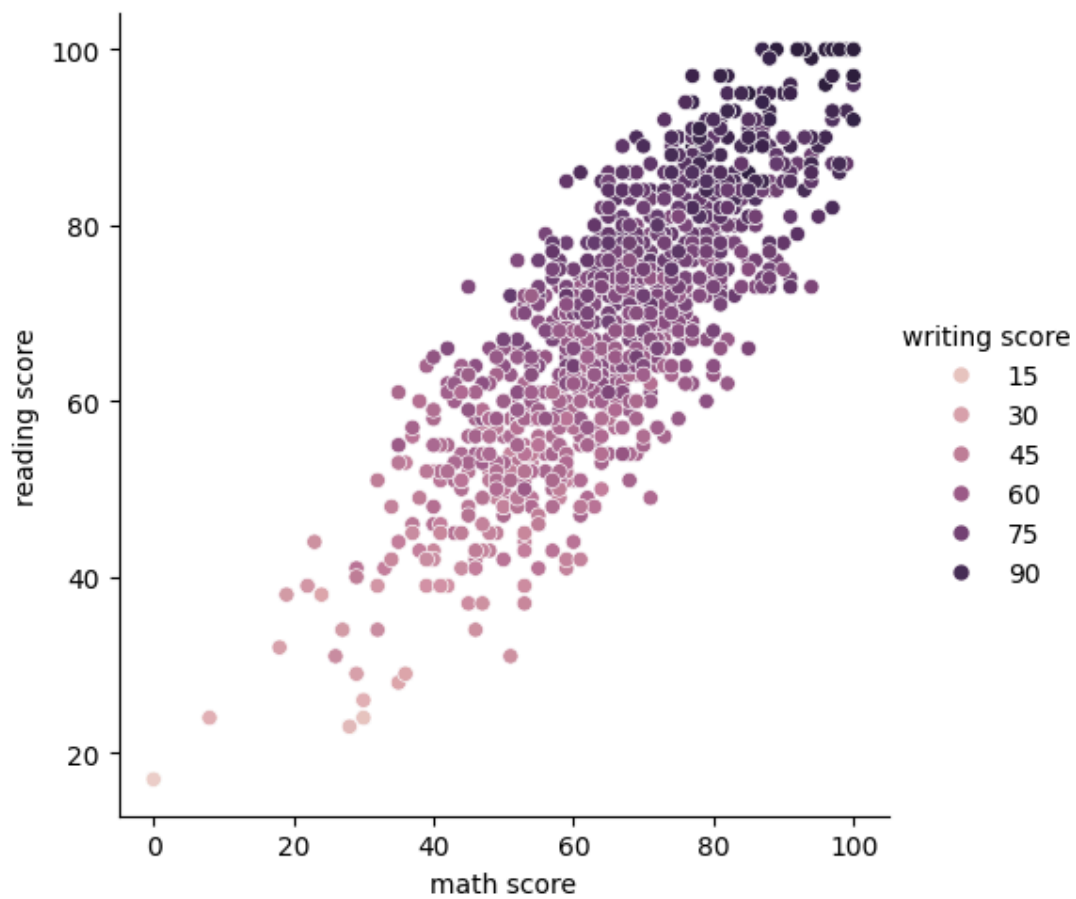
Out[50]: <Axes: >

In [51]:
```
sns.pairplot(student)
```
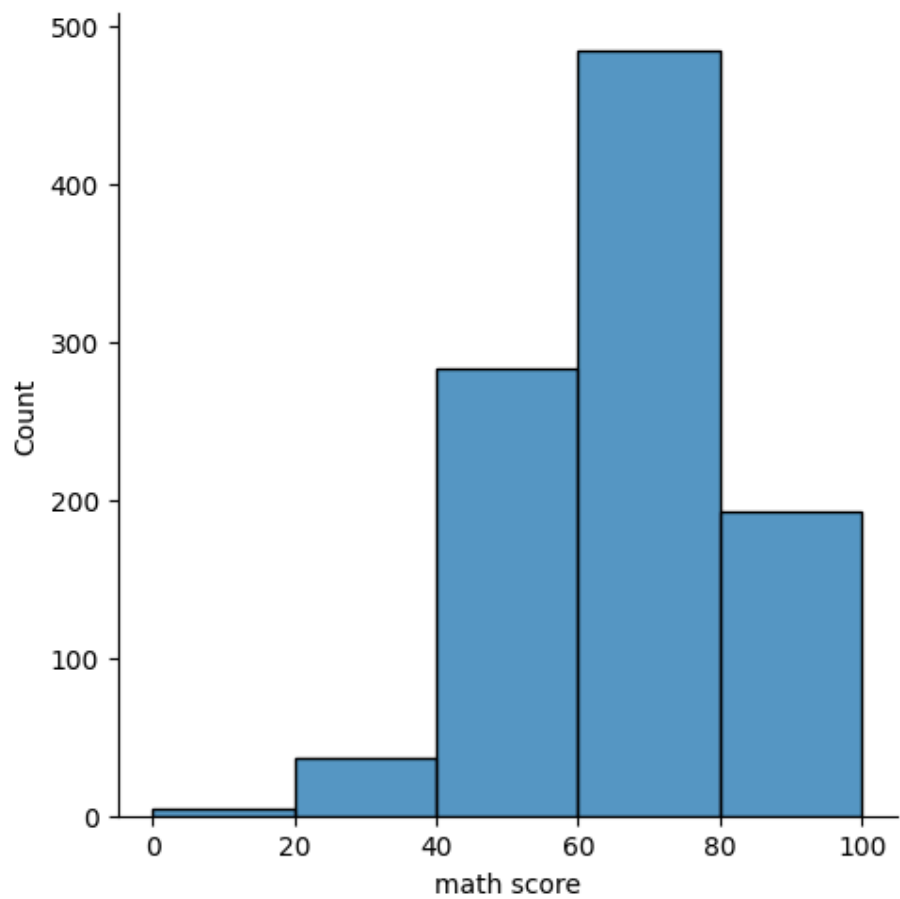
Out[51]: <seaborn.axisgrid.PairGrid at 0x21ce29265d0>
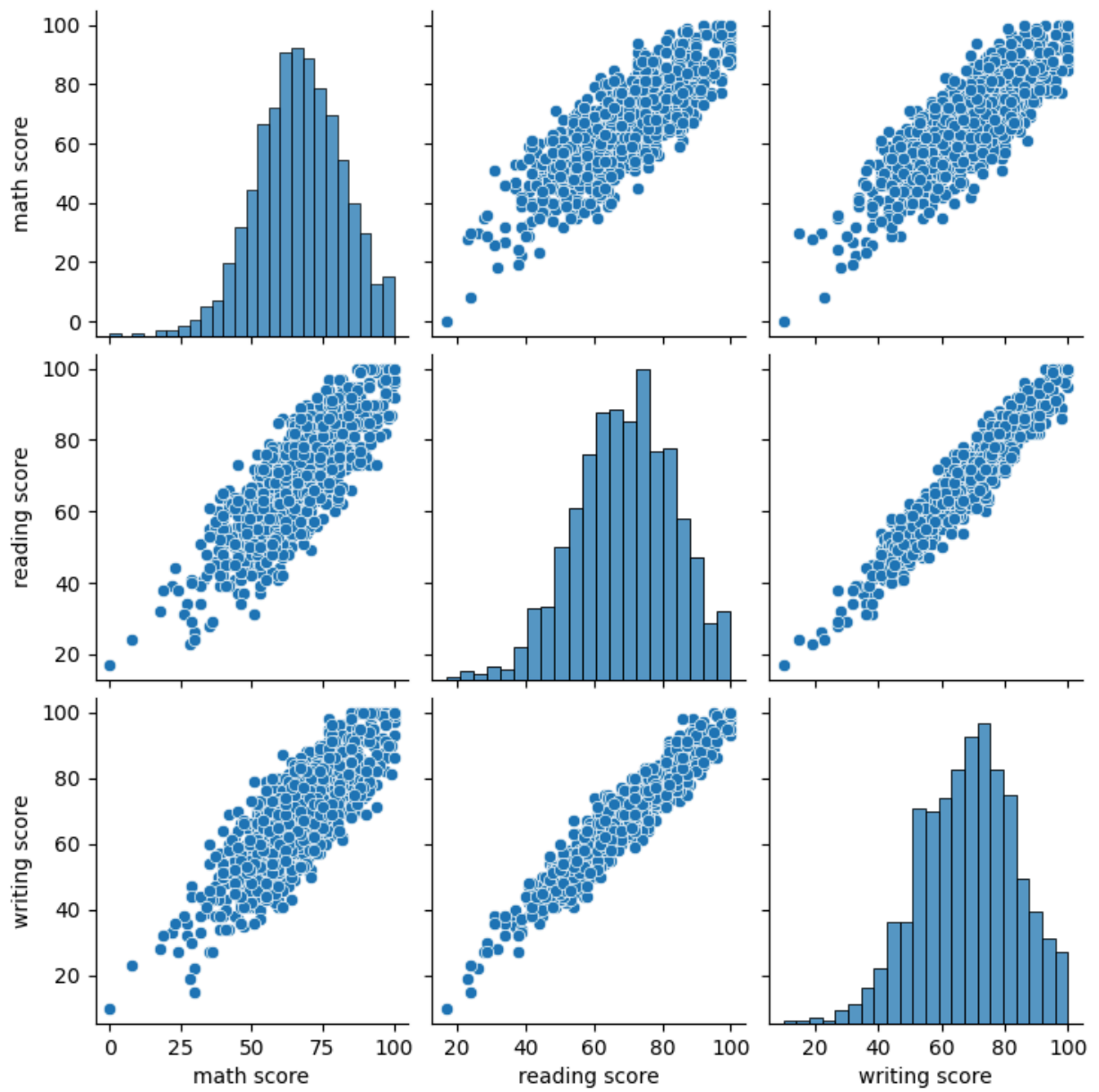
`sns.relplot(x='math score', y='reading score', hue='writing score', data=student)`

`<seaborn.axisgrid.FacetGrid at 0x21ce31a0440>`

In [65]: `sns.displot(student['math score'],bins=5)`

Out[65]: `<seaborn.axisgrid.FacetGrid at 0x21ce353c680>`