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

In [2]:
```python
%matplotlib inline
```

In [3]:
```python
cd \Users\758449\OneDrive - Cognizant\Desktop
```
```
C:\Users\758449\OneDrive - Cognizant\Desktop
```

In [4]:
```python
stu_data =  pd.read_csv("StudentsPerformance.csv")
```

In [5]:
```python
stu_data
```

Out[5]:

	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
...	...	...	...	...	...	...	...	...
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

1000 rows × 8 columns

In [6]:
```python
stu_data.columns
```

Out[6]:
```
Index(['gender', 'race/ethnicity', 'parental level of education', 'lunch',
       'test preparation course', 'math score', 'reading score',
       'writing score'],
      dtype='object')
```

In [7]:
```python
stu_data.info()
```
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   gender                       1000 non-null   object
 1   race/ethnicity               1000 non-null   object
 2   parental level of education  1000 non-null   object
 3   lunch                        1000 non-null   object
 4   test preparation course      1000 non-null   object
 5   math score                   1000 non-null   int64
 6   reading score                1000 non-null   int64
 7   writing score                1000 non-null   int64
dtypes: int64(3), object(5)
memory usage: 62.6+ KB
```

In [8]:
```python
stu_data.count()
```

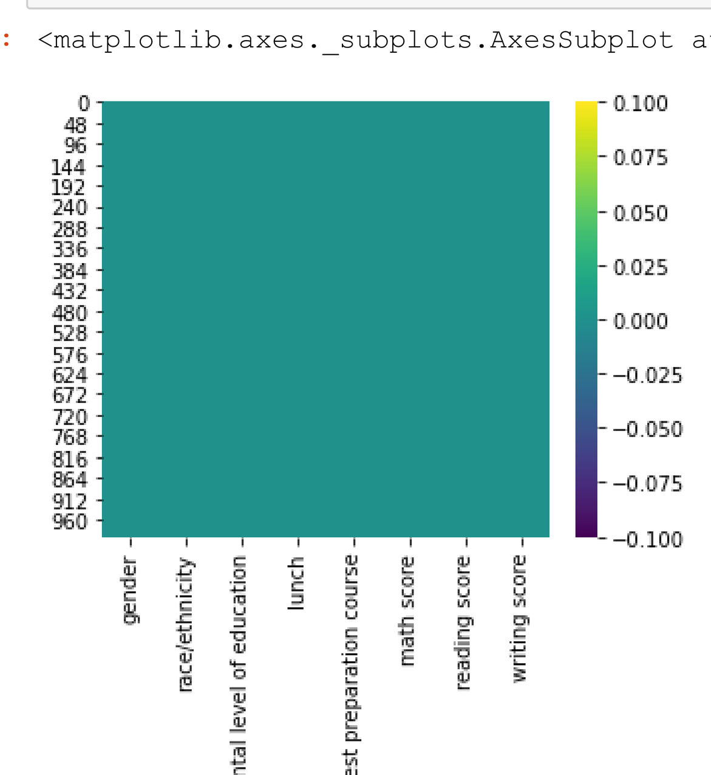
Out[8]:
```
gender                         1000
race/ethnicity                 1000
parental level of education    1000
lunch                          1000
test preparation course        1000
math score                     1000
reading score                  1000
writing score                  1000
dtype: int64
```

In [9]:
```python
stu_data.isnull().sum()
```

Out[9]:
```
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 [10]:
```python
plt.figure(figsize=(5,4))
sns.heatmap(stu_data.isnull(),cmap='viridis')
```

Out[10]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x18acd737fd0>
```



In [29]:
```python
#data is fine no null values
```

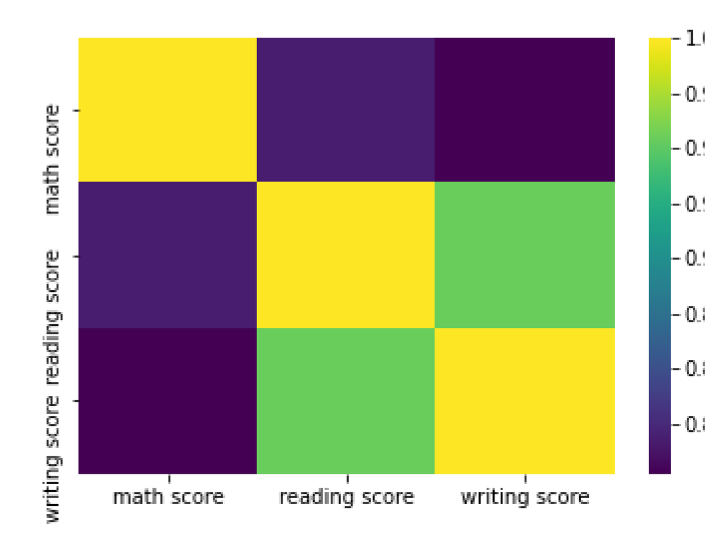
In [70]:
```python
stu_data.select_dtypes(include =np.number).head()
```

Out[70]:

	math score	reading score	writing score
0	72	72	74
1	69	90	88
2	90	95	93
3	47	57	44
4	76	78	75

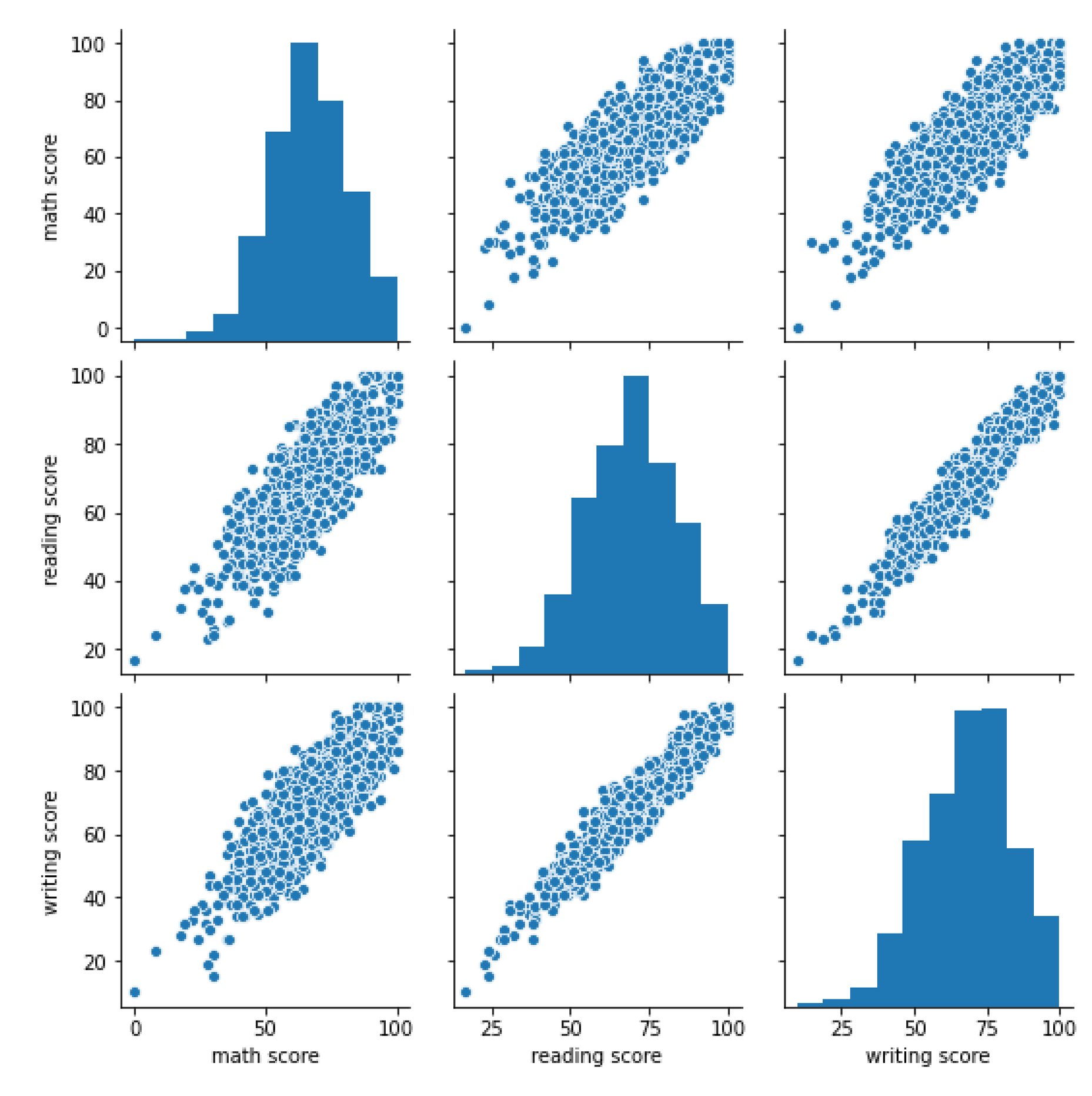
In [85]:
```python
sns.heatmap(stu_data.corr(),cmap='viridis')
```

Out[85]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x18ace06f8e0>
```



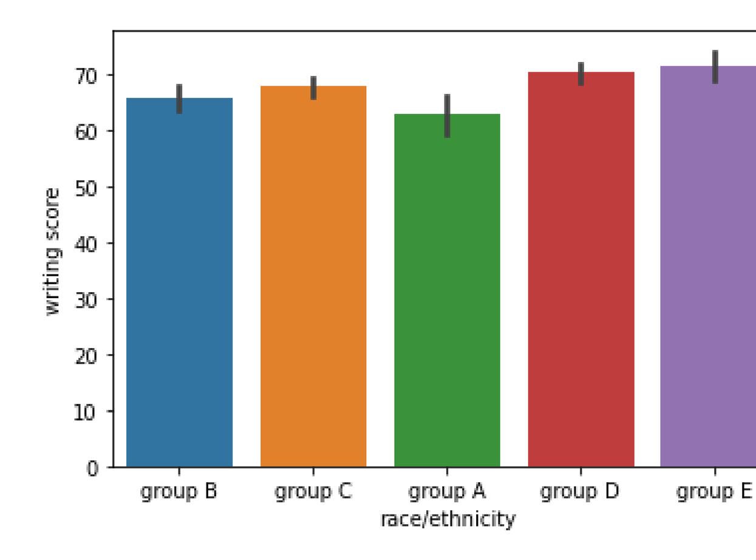
In [87]:
```python
sns.pairplot(stu_data)
```

Out[87]:
```
<seaborn.axisgrid.PairGrid at 0x18ace591430>
```



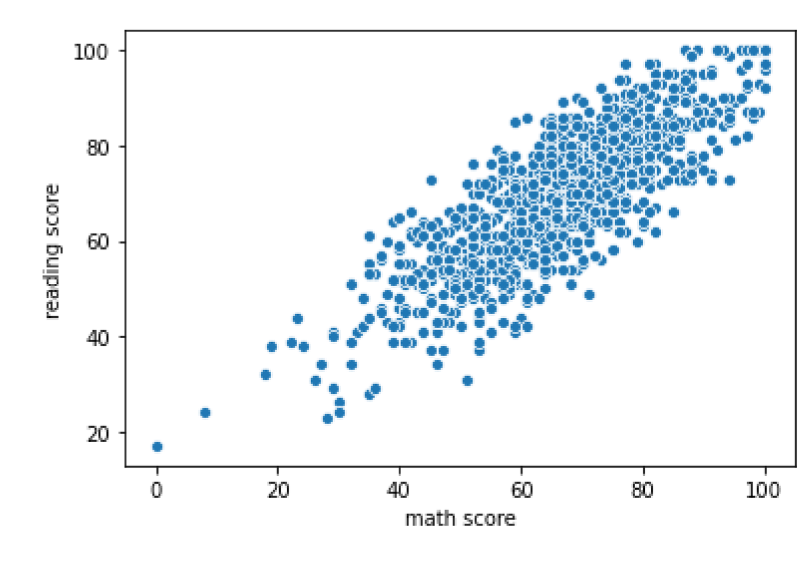
In [90]:
```python
sns.barplot(x= "race/ethnicity",y = "writing score" , data = stu_data)
```

Out[90]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x18acfd42d00>
```



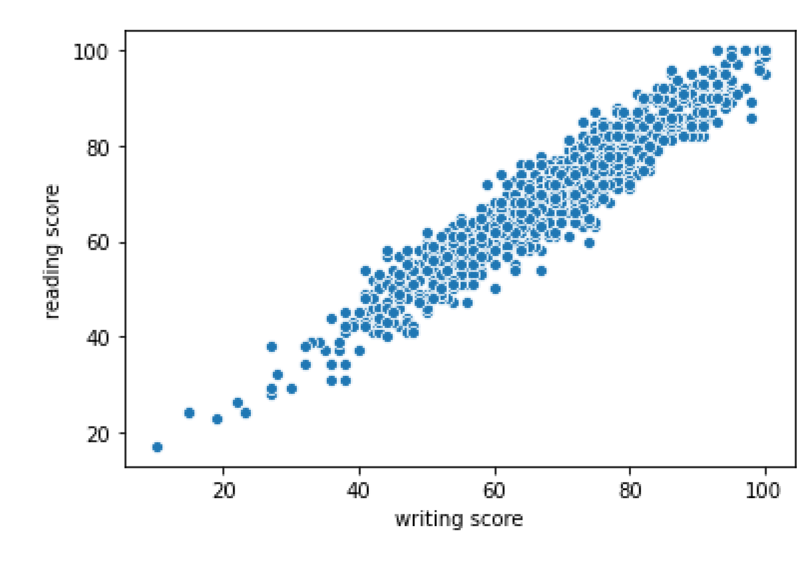
In [94]:
```python
sns.scatterplot(x= "math score" , y = "reading score", data = stu_data)
```

Out[94]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x18acfe556a0>
```



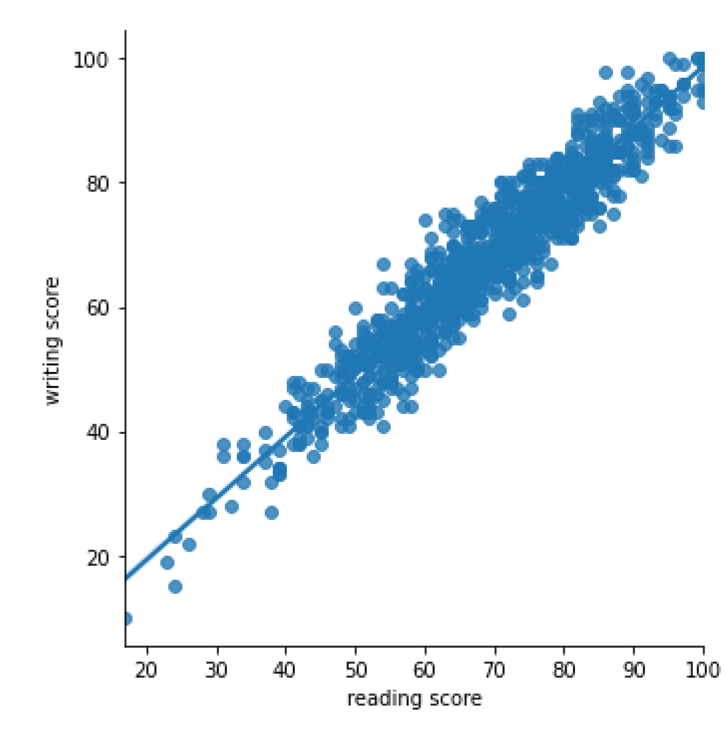
In [96]:
```python
sns.scatterplot(x ="writing score",y ="reading score",data =stu_data)
```

Out[96]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x18acfe7e970>
```



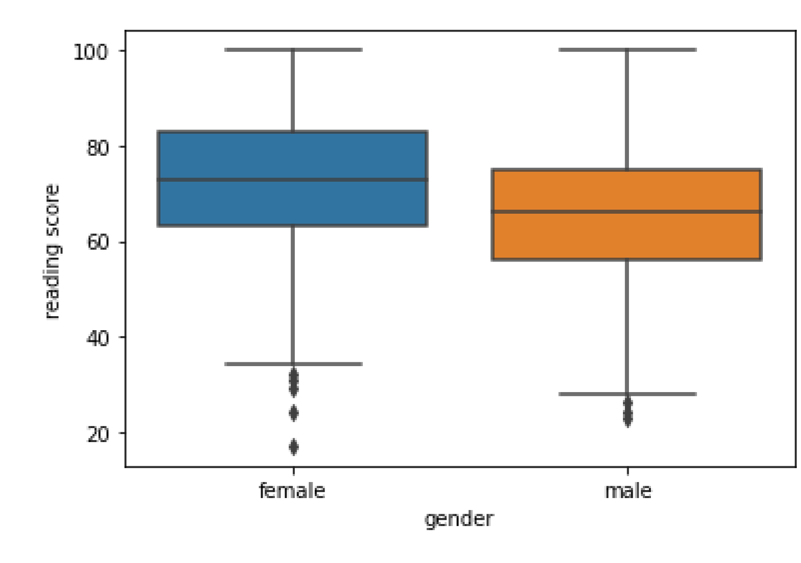
In [101]:
```python
plt.figure(figsize=(5,6))
sns.lmplot(x = "reading score" , y="writing score",data =stu_data)
```

Out[101]:
```
<seaborn.axisgrid.FacetGrid at 0x18ad15b3a30>
```
```
<Figure size 360x432 with 0 Axes>
```



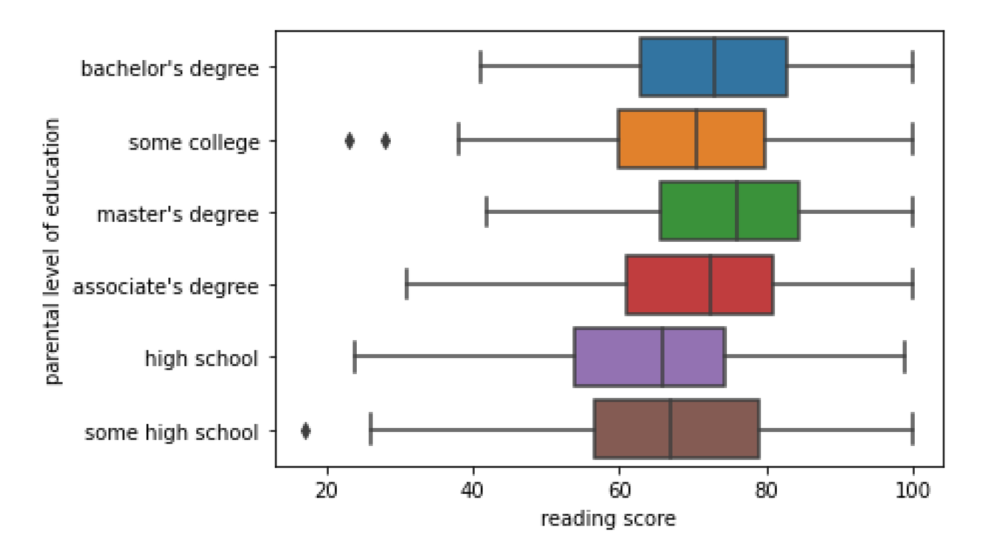
In [103]:
```python
sns.boxplot(x="gender" , y ="reading score" , data = stu_data)
```

Out[103]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x18ad123fac0>
```

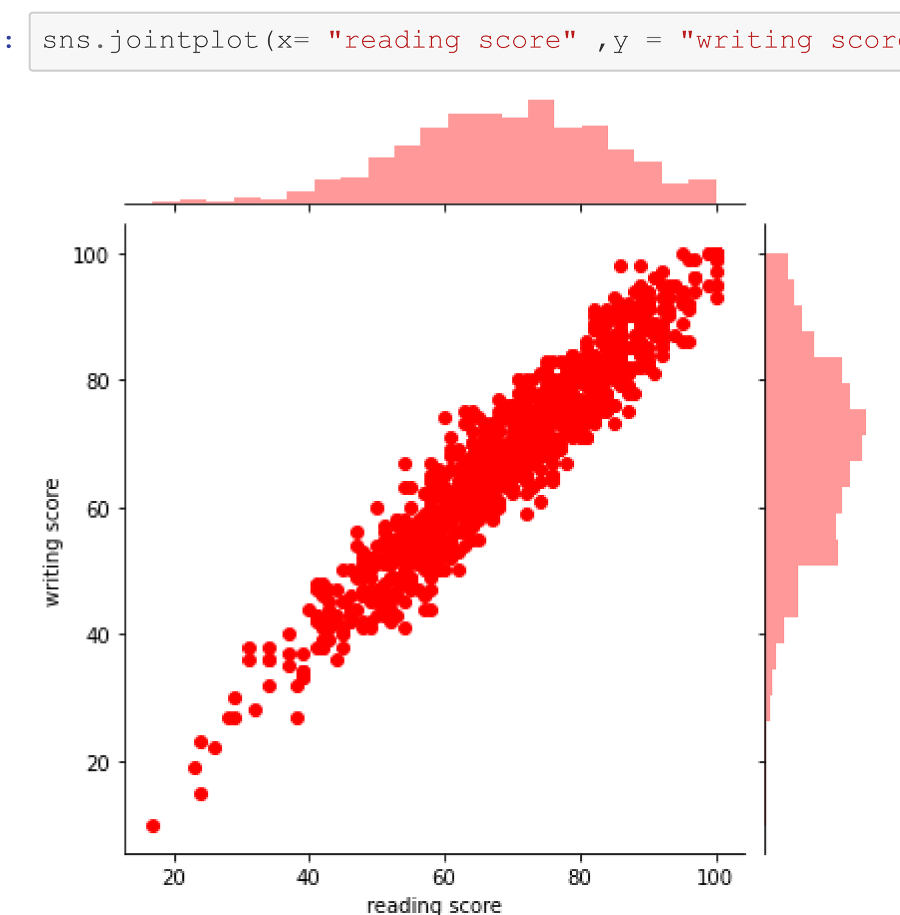


In [105]:
```python
sns.boxplot(x= "reading score" , y ="parental level of education", data = stu_data)
```

Out[105]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x18ad165b520>
```



In [121]:
```python
sns.jointplot(x= "reading score" ,y = "writing score", data = stu_data , color ="r")
```



just add meanscore for how to add new feature and find mean for numeric coulmns)

In [123]:
```python
stu_data['meanscore'] = stu_data.mean(axis=1)
```

In [125]:
```python
stu_data
```

Out[125]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	meanscore
0	female	group B	bachelor's degree	standard	none	72	72	74	72.666667
1	female	group C	some college	standard	completed	69	90	88	82.333333
2	female	group B	master's degree	standard	none	90	95	93	92.666667
3	male	group A	associate's degree	free/reduced	none	47	57	44	49.333333
4	male	group C	some college	standard	none	76	78	75	76.333333
...	...	...	...	...	...	...	...	...	...
995	female	group E	master's degree	standard	completed	88	99	95	94.000000
996	male	group C	high school	free/reduced	none	62	55	55	57.333333
997	female	group C	high school	free/reduced	completed	59	71	65	65.000000
998	female	group D	some college	standard	completed	68	78	77	74.333333
999	female	group D	some college	free/reduced	none	77	86	86	83.000000

1000 rows × 9 columns

drop meanscore using drop command.

In [130]:
```python
stu_data.drop(['meanscore'],axis=1)
```

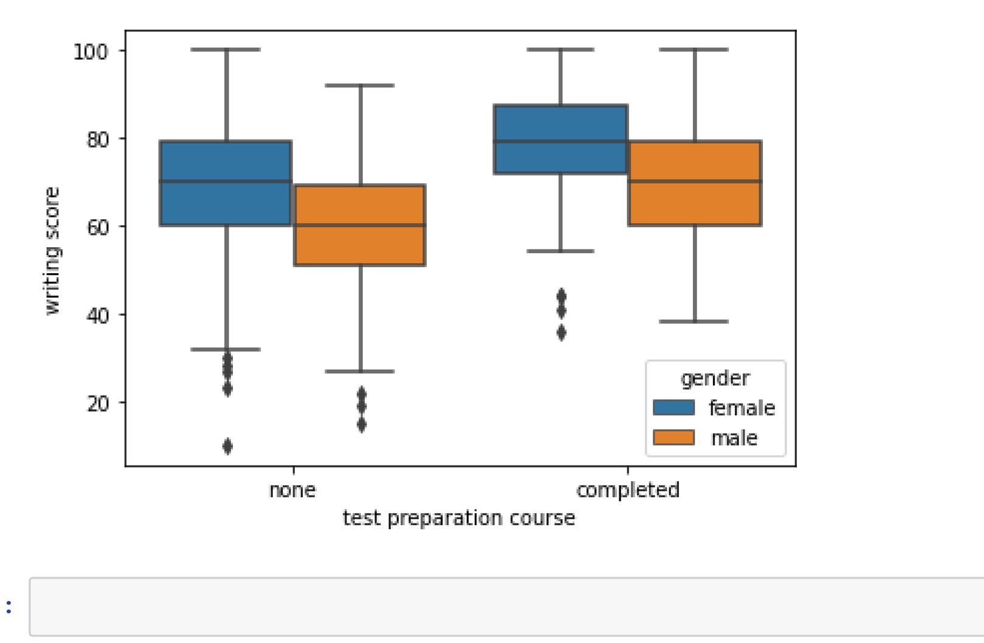
Out[130]:

	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
...	...	...	...	...	...	...	...	...
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

1000 rows × 8 columns

In [140]:
```python
sns.boxplot(x="test preparation course",y="writing score" ,data =stu_data,hue='gender',)
```

Out[140]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x18ad4584dc0>
```



In [143]:

In [ ]:

In [ ]: