

REVANTH BHARGAVA BOYIDI

MLB1-MINOR-APRIL

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

READING THE DATASET

```
In [2]: df=pd.read_csv('StudentsPerformance.csv')
df.head()
```

Out[2]:

	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 [3]: df.shape
```

Out[3]: (1000, 8)

```
In [4]: df.dtypes
```

```
Out[4]: gender                object
        race/ethnicity         object
        parental level of education  object
        lunch                  object
        test preparation course    object
        math score               int64
        reading score            int64
        writing score             int64
        dtype: object
```

NULL VALUES

```
In [5]: df.isnull().sum()
```

```
Out[5]: 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 [6]: df.dtypes
```

```
Out[6]: gender                object
        race/ethnicity         object
        parental level of education  object
        lunch                  object
        test preparation course    object
        math score               int64
        reading score            int64
        writing score             int64
        dtype: object
```

AS THERE ARE 3 MARKS, THERE AVERAGE IS TAKEN FOR FOR BETTER UNDERSTANDING

```
In [7]: df['Avg score']=(df['math score'] + df['reading score'] + df['writing score'])/3
df.head()
```

Out[7]:

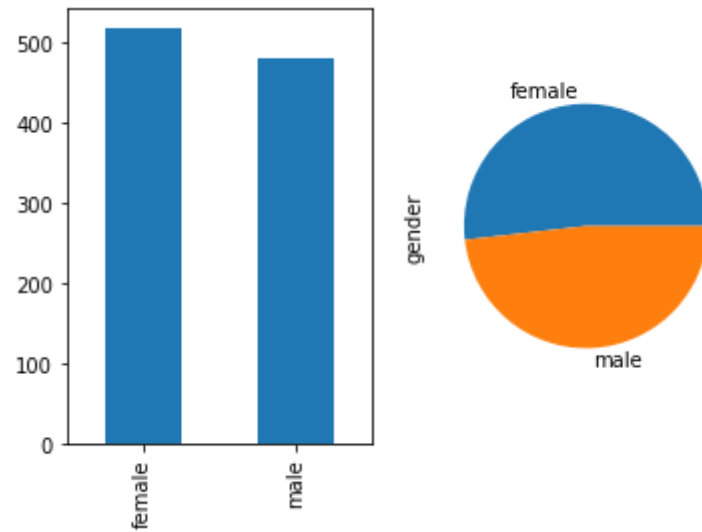
	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	Avg score
0	female	group B	bachelor's degree	standard	none	72	72	74	72.666667
1	female	group C	some college	standard	completed	69	90	88	82.333333
2	female	group B	master's degree	standard	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

```
In [8]: df.gender.value_counts()
```

```
Out[8]: female    518
male          482
Name: gender, dtype: int64
```

```
In [9]: plt.subplot(1,2,1)
df.gender.value_counts().plot(kind='bar')
plt.subplot(1,2,2)
df.gender.value_counts().plot(kind='pie')
```

Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa28729c8>



```
In [10]: df.rename(columns={'race/ethnicity': 'race'}, inplace=True)
df.race.value_counts()
```

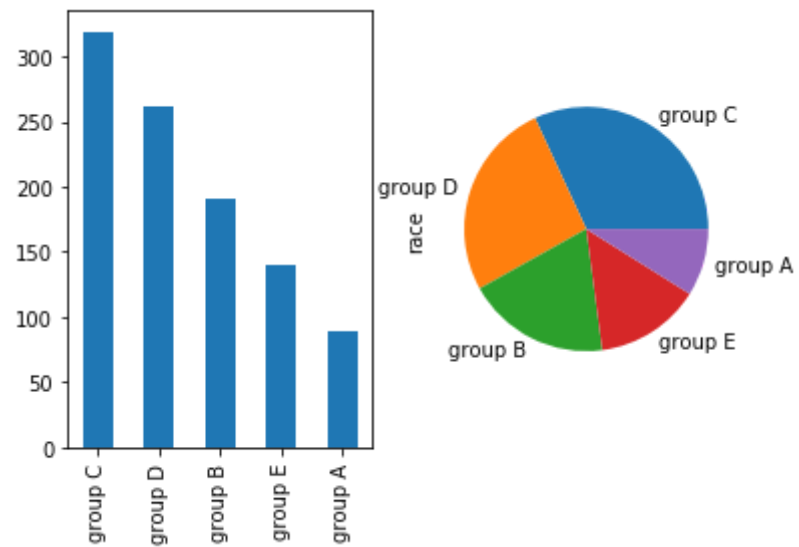
Out[10]:

group C	319
group D	262
group B	190
group E	140
group A	89

Name: race, dtype: int64

```
In [11]: plt.subplot(1,2,1)
df.race.value_counts().plot(kind='bar')
plt.subplot(1,2,2)
df.race.value_counts().plot(kind='pie')
```

Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa29c2408>

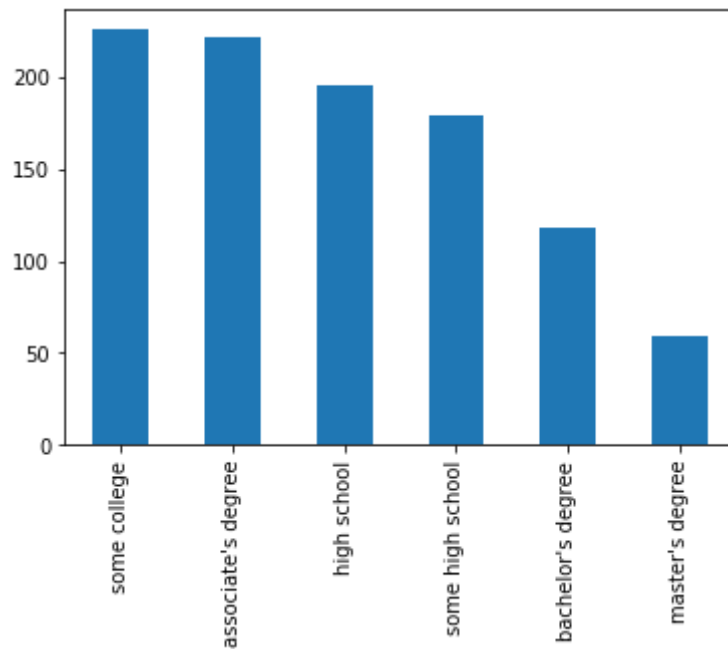


```
In [12]: df.rename(columns={'parental level of education':'parental_level_of_education'},inplace=True)
df.parental_level_of_education.value_counts()
```

```
Out[12]: some college      226
associate's degree      222
high school            196
some high school       179
bachelor's degree      118
master's degree         59
Name: parental_level_of_education, dtype: int64
```

```
In [13]: df.parental_level_of_education.value_counts().plot(kind='bar')
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa2a56588>
```

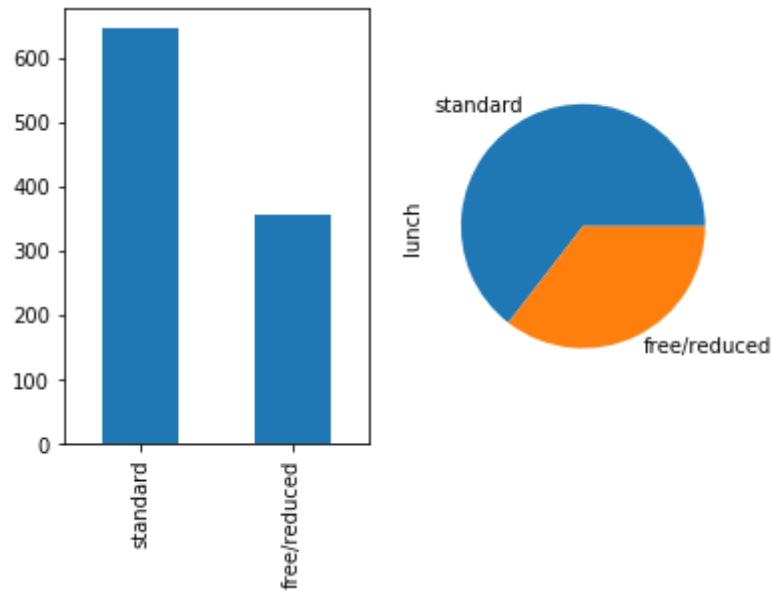


```
In [14]: df.lunch.value_counts()
```

```
Out[14]: standard      645
free/reduced      355
Name: lunch, dtype: int64
```

```
In [15]: plt.subplot(1,2,1)
df.lunch.value_counts().plot(kind='bar')
plt.subplot(1,2,2)
df.lunch.value_counts().plot(kind='pie')
```

Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa2b00908>

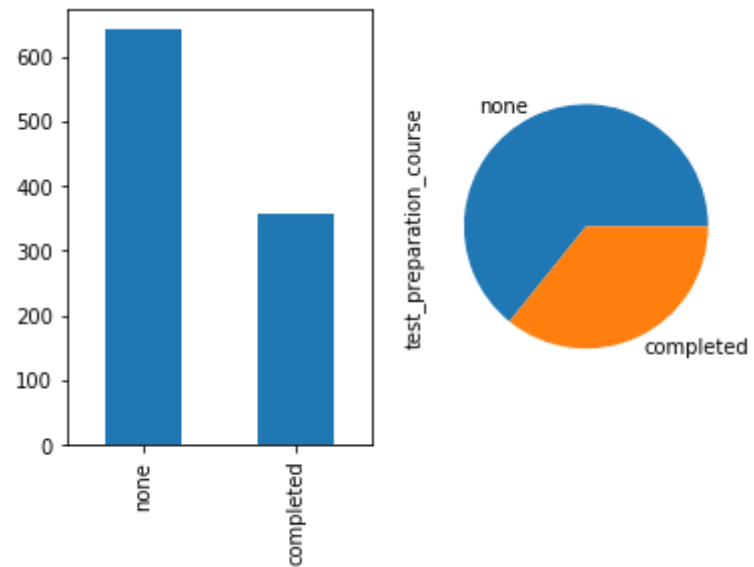


```
In [16]: df.rename(columns={'test preparation course':'test_preparation_course'},inplace=True)
df.test_preparation_course.value_counts()
```

Out[16]: none 642
completed 358
Name: test_preparation_course, dtype: int64

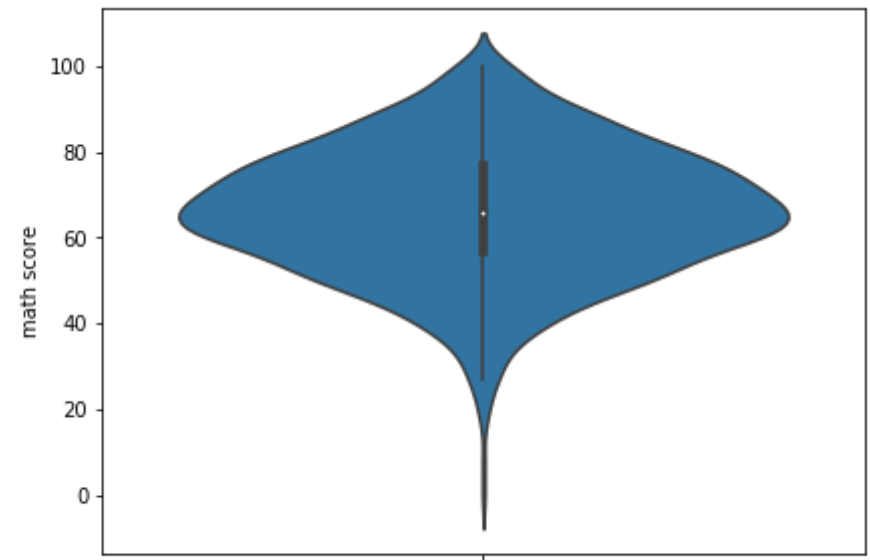
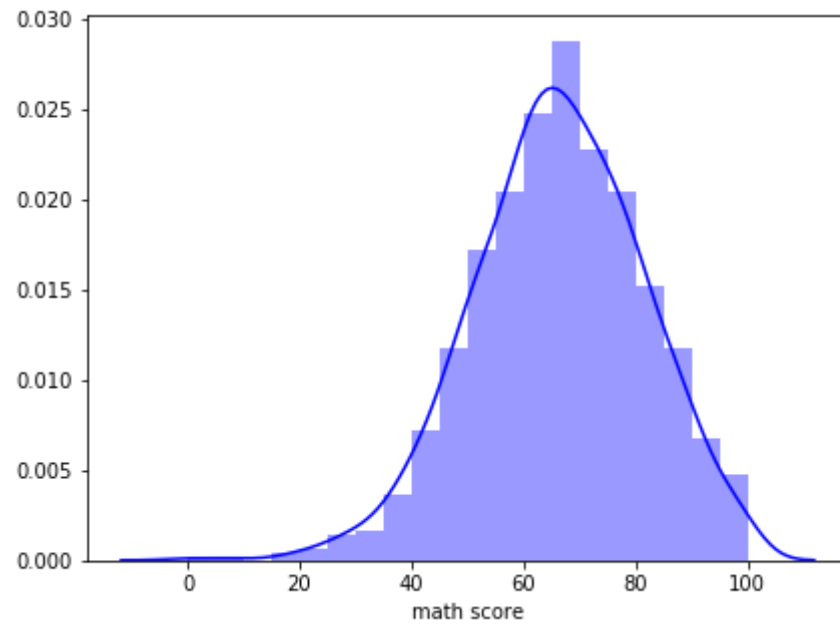
```
In [17]: plt.subplot(1,2,1)
df.test_preparation_course.value_counts().plot(kind='bar')
plt.subplot(1,2,2)
df.test_preparation_course.value_counts().plot(kind='pie')
```

Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa2ba4048>




```
In [18]: plt.figure(figsize=(15,5))
plt.subplot(1,2,1)
sns.distplot(df['math score'],kde=True,color='b',bins=list(range(0,105,5)))
plt.subplot(1,2,2)
sns.violinplot(y='math score',data=df)
```

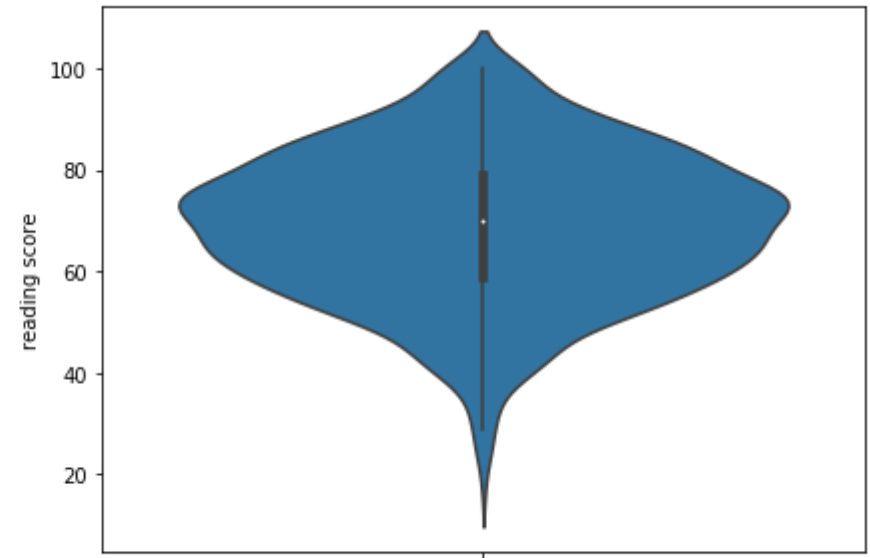
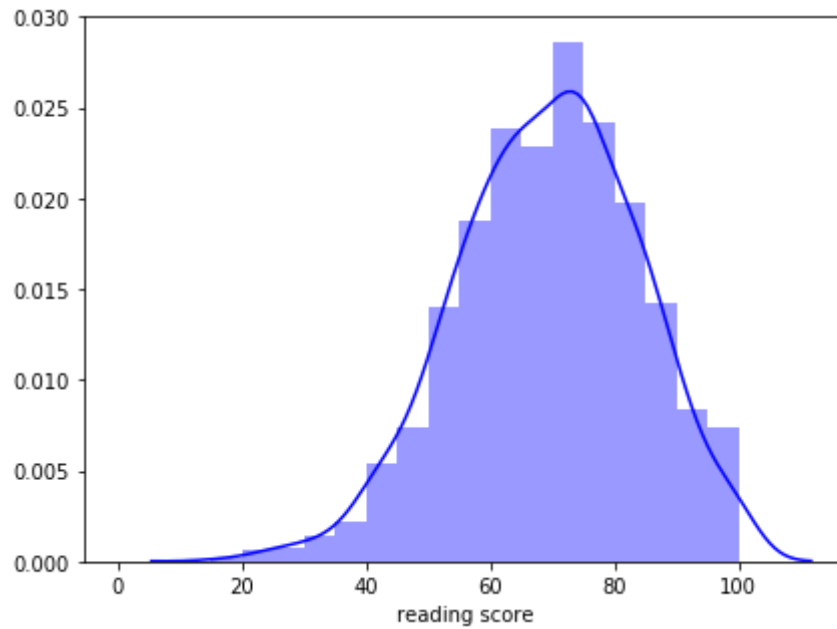
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa2c48608>



MOST OF THE STUDENTS HAVE SCORED IN THE RANGE 60 - 80 IN MATHS

```
In [19]: plt.figure(figsize=(15,5))
plt.subplot(1,2,1)
sns.distplot(df['reading score'],kde=True,color='b',bins=list(range(0,105,5)))
plt.subplot(1,2,2)
sns.violinplot(y='reading score',data=df)
```

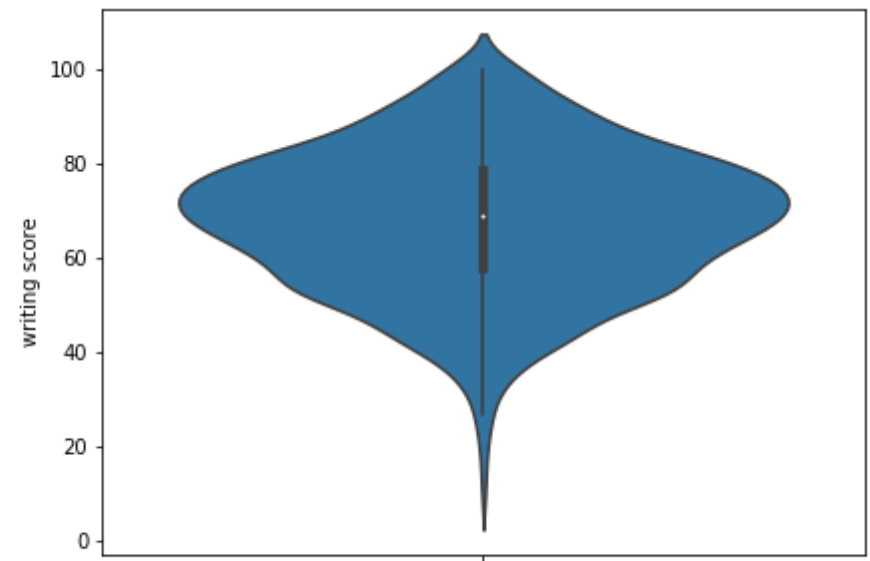
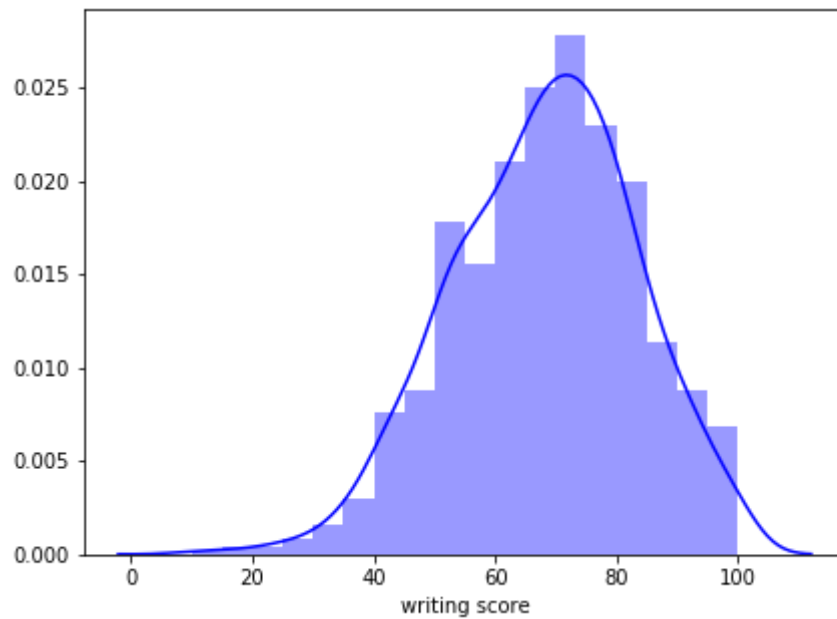
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa2d570c8>



MOST OF THE STUDENTS HAVE SCORED IN THE RANGE 70 - 85 IN READING

```
In [20]: plt.figure(figsize=(15,5))
plt.subplot(1,2,1)
sns.distplot(df['writing score'],kde=True,color='b',bins=list(range(0,105,5)))
plt.subplot(1,2,2)
sns.violinplot(y='writing score',data=df)
```

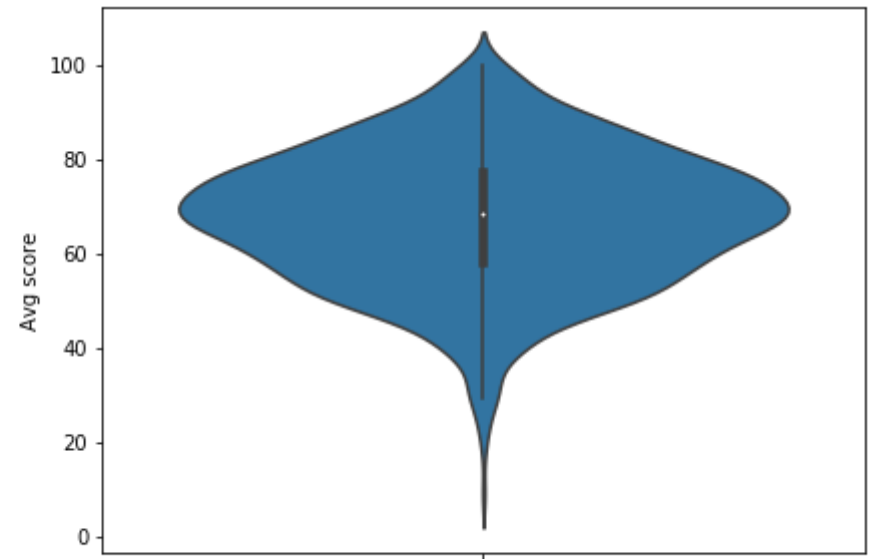
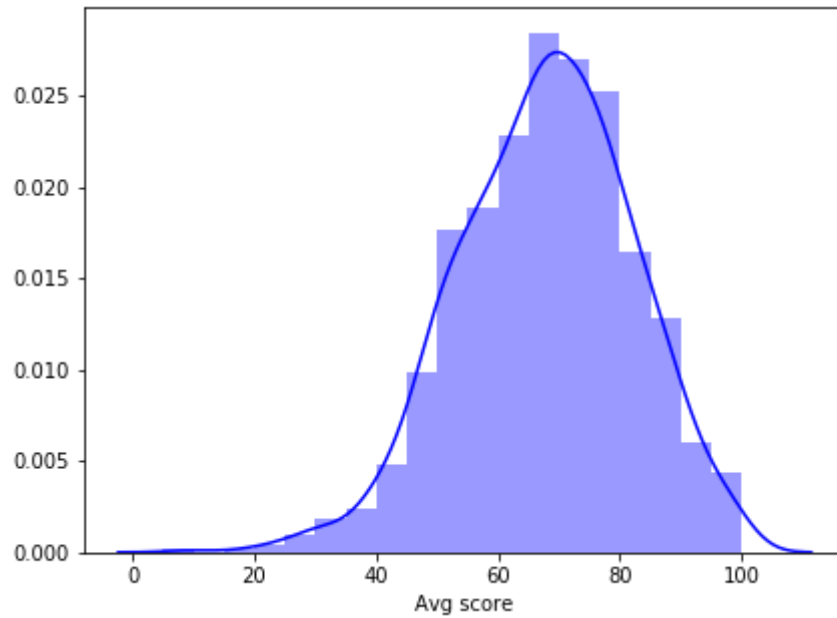
```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa2e50148>
```



MOST OF THE STUDENTS HAVE SCORED IN THE RANGE 65 - 85 IN WRITING

```
In [21]: plt.figure(figsize=(15,5))
plt.subplot(1,2,1)
sns.distplot(df['Avg score'],kde=True,color='b',bins=list(range(0,105,5)))
plt.subplot(1,2,2)
sns.violinplot(y='Avg score',data=df)
```

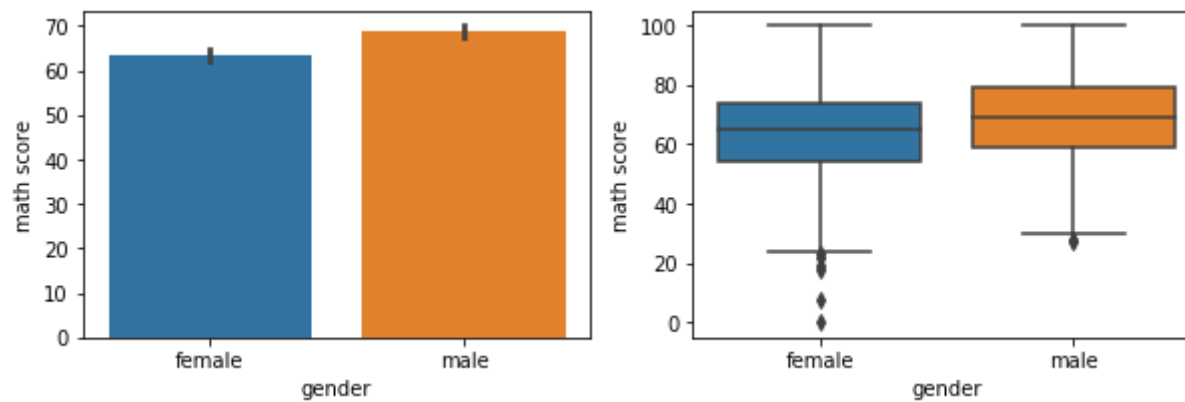
Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa3082608>



MOST OF THE STUDENTS HAVE SCORED IN THE RANGE 65 - 80 IN TOTAL

```
In [22]: plt.figure(figsize=(10,3))  
plt.subplot(1,2,1)  
sns.barplot(x='gender',y='math score',data=df)  
plt.subplot(1,2,2)  
sns.boxplot(x='gender',y='math score',data=df)
```

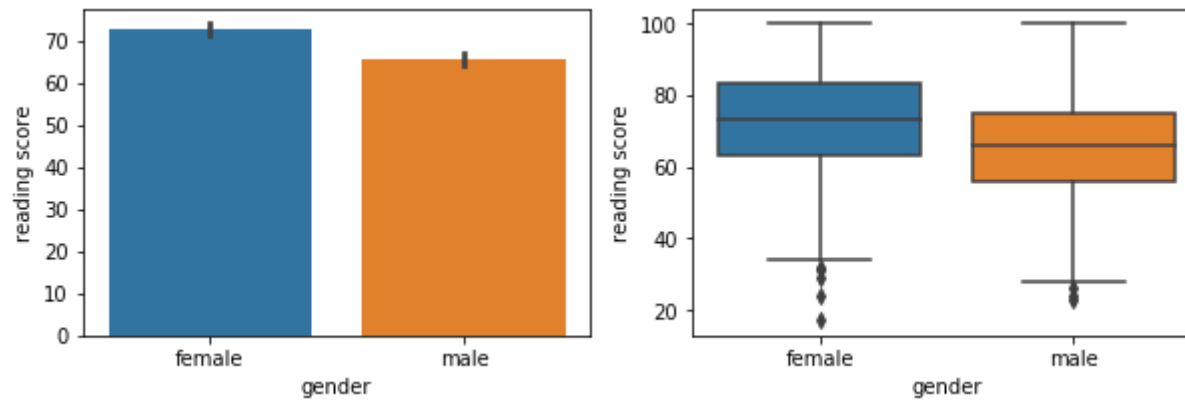
Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa3174e08>



MALE STUDENTS HAVE PERFORMED BETTER IN MATHS

```
In [23]: plt.figure(figsize=(10,3))  
plt.subplot(1,2,1)  
sns.barplot(x='gender',y='reading score',data=df)  
plt.subplot(1,2,2)  
sns.boxplot(x='gender',y='reading score',data=df)
```

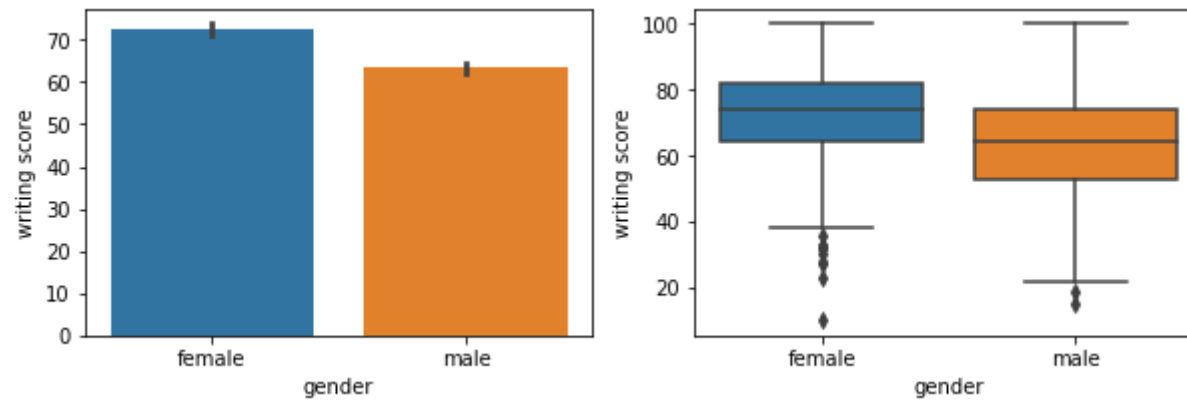
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa3397ac8>



FEMALE STDENTS HAVE PERFORMED BETTER IN READING

```
In [24]: plt.figure(figsize=(10,3))  
plt.subplot(1,2,1)  
sns.barplot(x='gender',y='writing score',data=df)  
plt.subplot(1,2,2)  
sns.boxplot(x='gender',y='writing score',data=df)
```

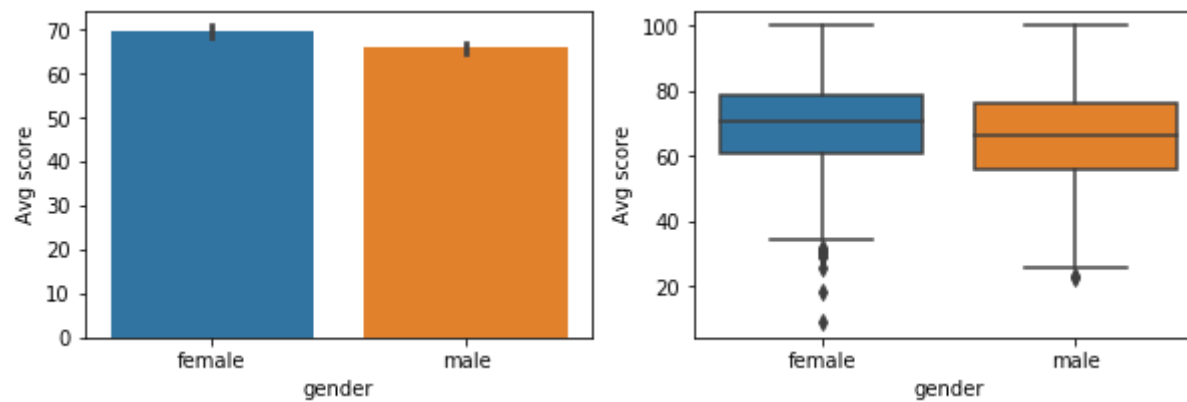
Out[24]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa36aa1c8>



FEMALE STDENTS HAVE PERFORMED BETTER IN WRITING

```
In [25]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='gender',y='Avg score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='gender',y='Avg score',data=df)
```

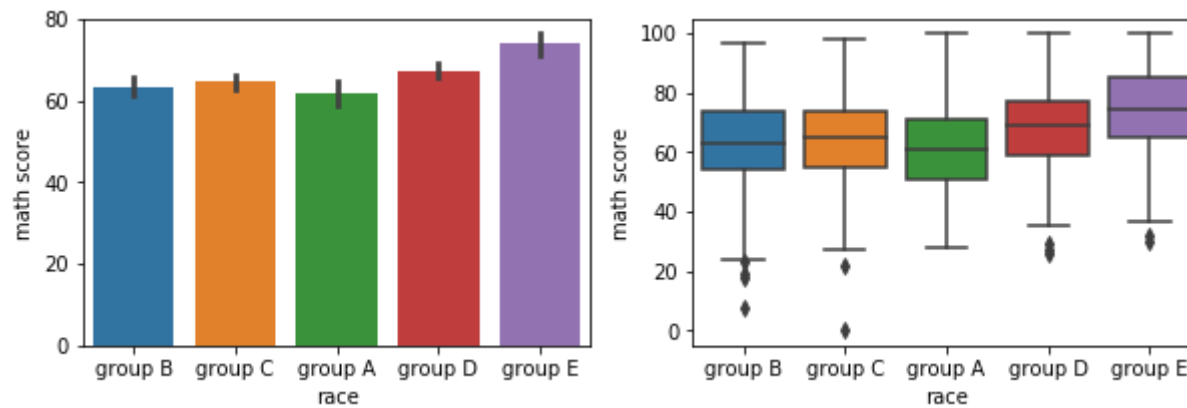
Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa3757dc8>



FEMALE STDENTS ARE PERFORMING BETTER THAN MALE STUDENTS


```
In [26]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='race',y='math score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='race',y='math score',data=df)
```

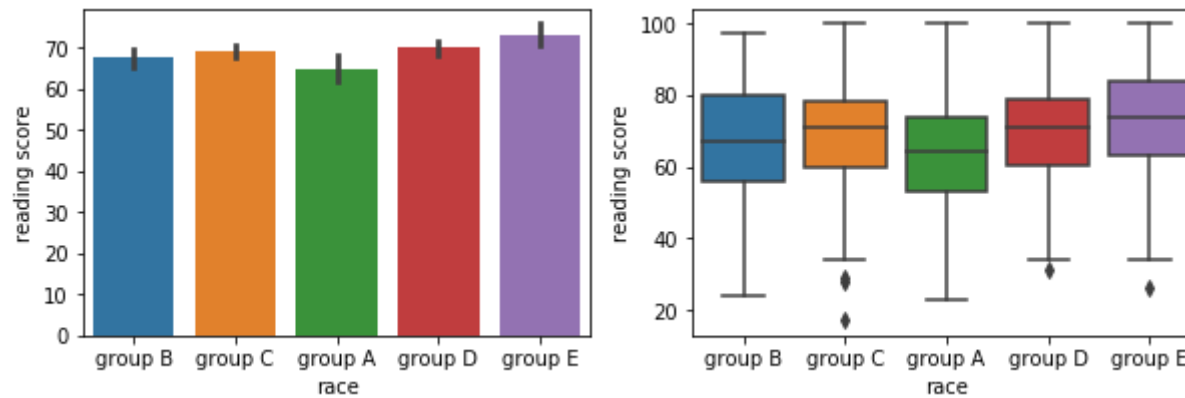
Out[26]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa384dac8>



STUDENTS OF GROUP E HAVE PERFORMED BETTER IN MATHS WHILE GROUP A HAVE SHOWN THE LEAST PERFORMANCE

```
In [27]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='race',y='reading score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='race',y='reading score',data=df)
```

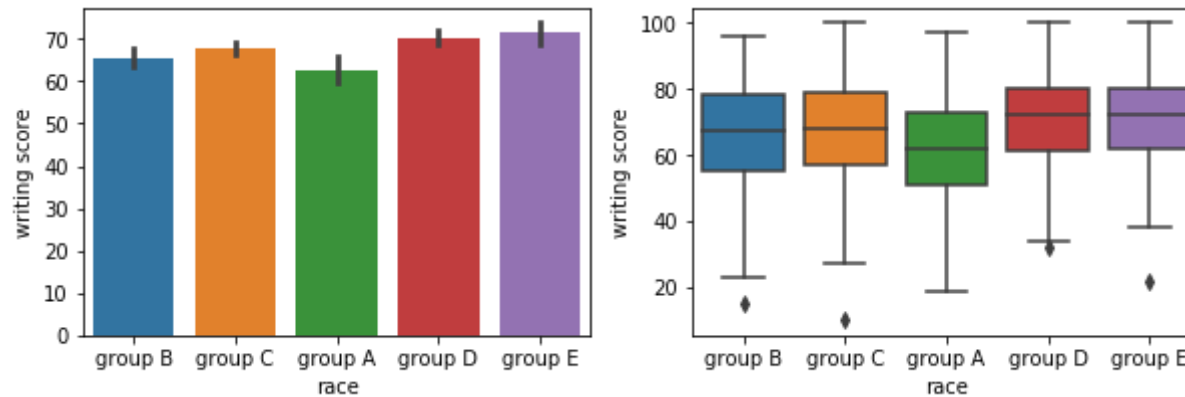
Out[27]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa397d848>



STUDENTS OF GROUP E HAVE PERFORMED BETTER IN READING WHILE GROUP A HAVE SHOWN THE LEAST PERFORMANCE

```
In [28]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='race',y='writing score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='race',y='writing score',data=df)
```

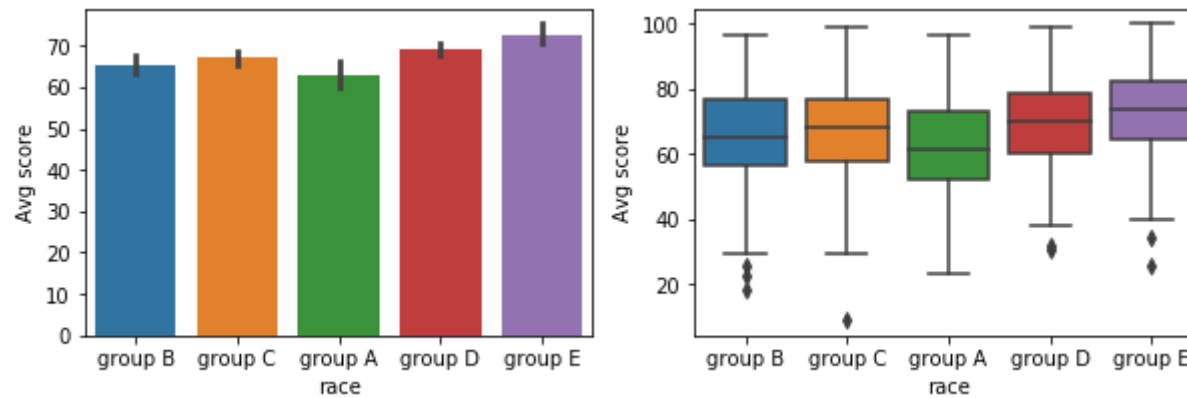
Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa3aa1348>



STUDENTS OF GROUP E AND GROUP D HAVE PERFORMED BETTER IN WRITING WHILE GROUP A HAVE SHOWN THE LEAST PERFORMANCE

```
In [29]: plt.figure(figsize=(10,3))  
plt.subplot(1,2,1)  
sns.barplot(x='race',y='Avg score',data=df)  
plt.subplot(1,2,2)  
sns.boxplot(x='race',y='Avg score',data=df)
```

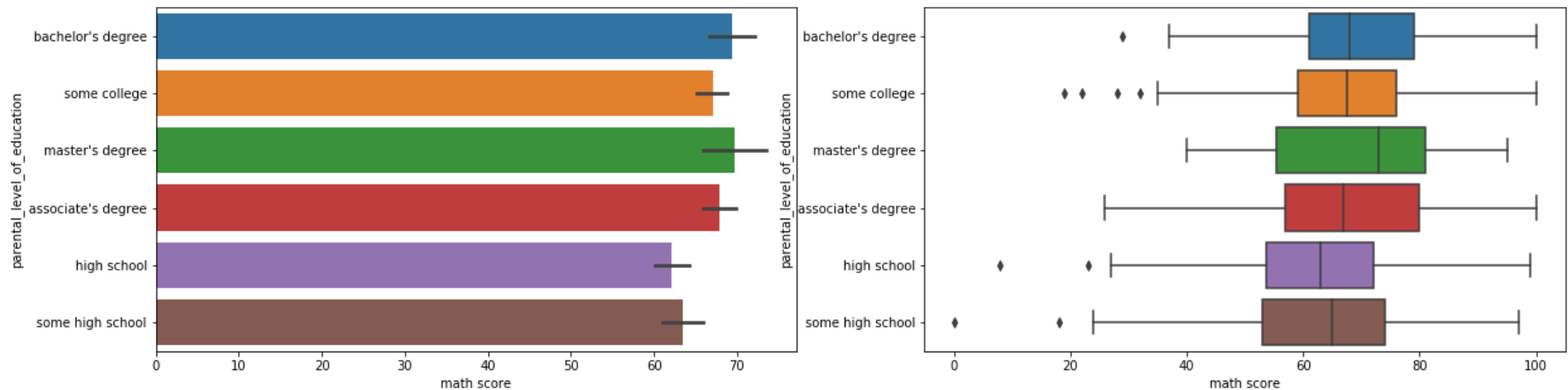
Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa4ba4ac8>



STUDENTS OF GROUP E HAVE PERFORMED BETTER IN OVERALL WHILE GROUP A HAVE SHOWN THE LEAST PERFORMANCE

```
In [30]: plt.figure(figsize=(20,5))
plt.subplot(1,2,1)
sns.barplot(y='parental_level_of_education',x='math score',data=df)
plt.subplot(1,2,2)
sns.boxplot(y='parental_level_of_education',x='math score',data=df)
```

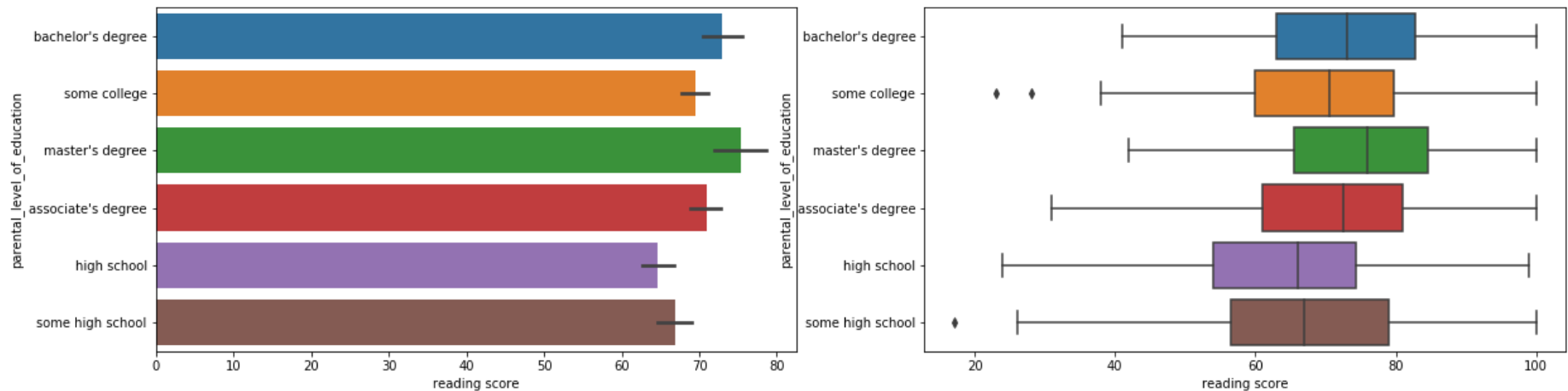
Out[30]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa4cd6c48>



STUDENTS WHOSE PARENTS HAVE A MASTER'S DEGREE OR BACHELOR'S DEGREE HAVE SHOWN THE BEST PERFORMANCE IN MATHS STUDENTS WHOSE PARENTS HAVE A GONE TILL HIGH SCHOOL HAVE SHOWN THE LEAST PERFORMANCE IN MATHS

```
In [31]: plt.figure(figsize=(20,5))
plt.subplot(1,2,1)
sns.barplot(y='parental_level_of_education',x='reading score',data=df)
plt.subplot(1,2,2)
sns.boxplot(y='parental_level_of_education',x='reading score',data=df)
```

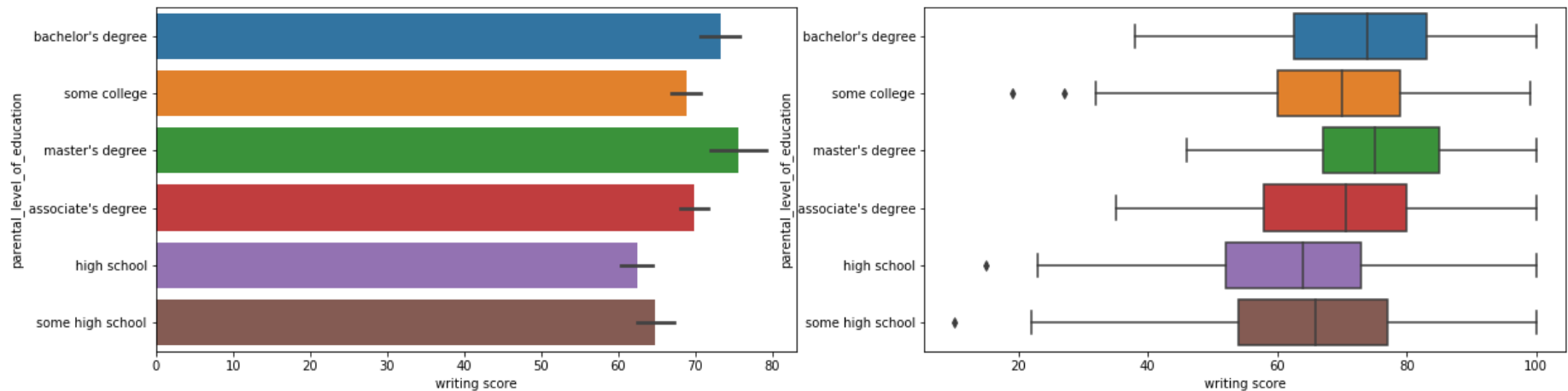
Out[31]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa502ca08>



STUDENTS WHOSE PARENTS HAVE A MASTER DEGREE HAVE SHOWN THE BEST PERFORMANCE IN READING STUDENTS WHOSE PARENTS HAVE A GONE TILL HIGH SCHOOL HAVE SHOWN THE LEAST PERFORMANCE IN READING

```
In [32]: plt.figure(figsize=(20,5))
plt.subplot(1,2,1)
sns.barplot(y='parental_level_of_education',x='writing score',data=df)
plt.subplot(1,2,2)
sns.boxplot(y='parental_level_of_education',x='writing score',data=df)
```

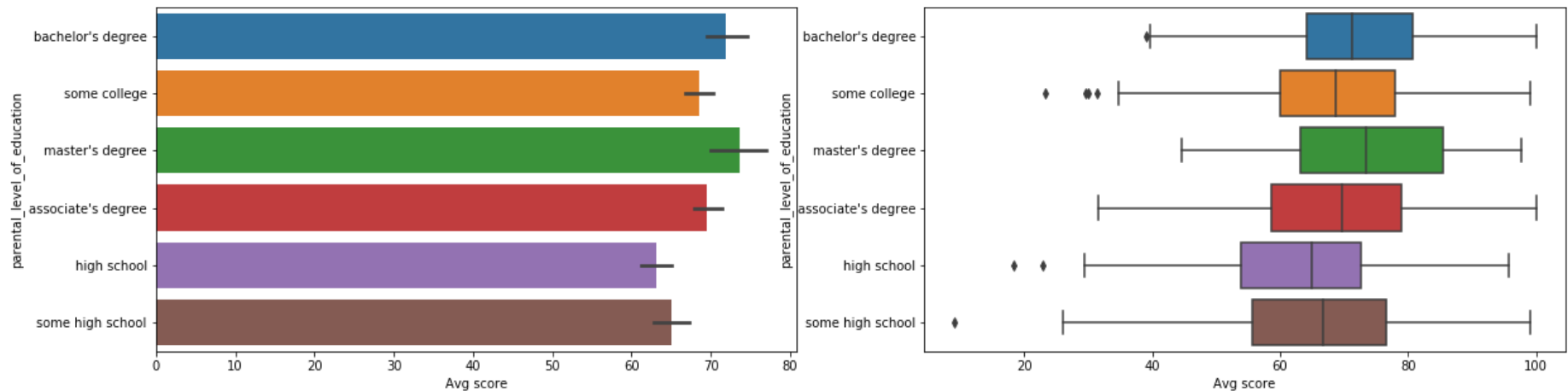
Out[32]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa4f40f48>



STUDENTS WHOSE PARENTS HAVE A MASTER DEGREE HAVE SHOWN THE BEST PERFORMANCE IN WRITING STUDENTS WHOSE PARENTS HAVE A GONE TILL HIGH SCHOOL HAVE SHOWN THE LEAST PERFORMANCE IN WRITING

```
In [33]: plt.figure(figsize=(20,5))
plt.subplot(1,2,1)
sns.barplot(y='parental_level_of_education',x='Avg score',data=df)
plt.subplot(1,2,2)
sns.boxplot(y='parental_level_of_education',x='Avg score',data=df)
```

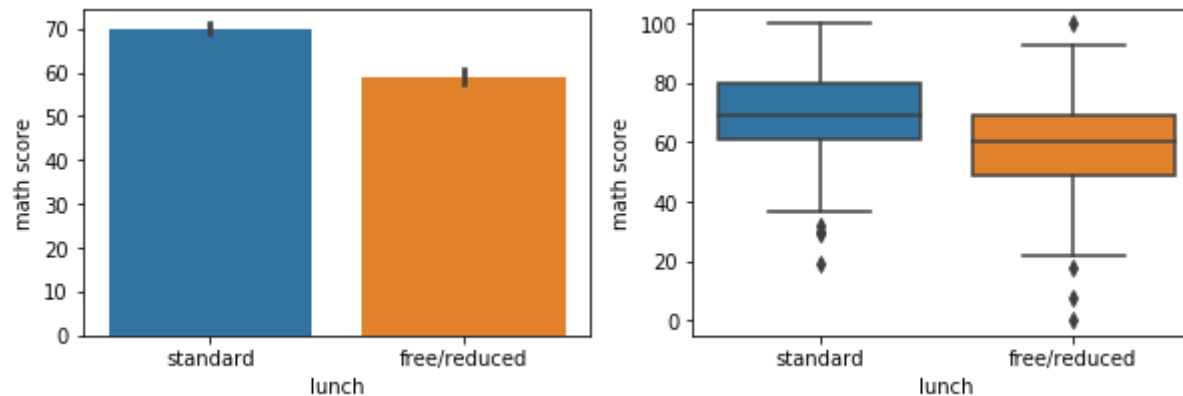
Out[33]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa559af48>



STUDENTS WHOSE PARENTS HAVE A MASTER DEGREE HAVE SHOWN THE BEST PERFORMANCE IN ALL THE TESTS STUDENTS WHOSE PARENTS HAVE A GONE TILL HIGH SCHOOL HAVE SHOWN THE LEAST PERFORMANCE IN TESTS


```
In [34]: plt.figure(figsize=(10,3))  
plt.subplot(1,2,1)  
sns.barplot(x='lunch',y='math score',data=df)  
plt.subplot(1,2,2)  
sns.boxplot(x='lunch',y='math score',data=df)
```

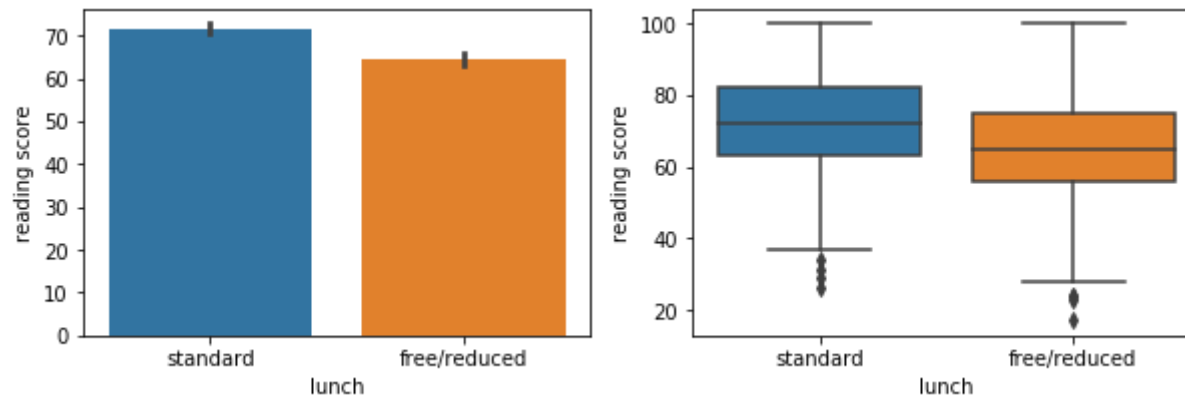
Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5901dc8>



STUDENTS WHO CAN AFFORD STANDARD LUNCH ARE PERFORMING BETTER IN MATHS COMPARED TO STUDENTS WHO GET FREE/REDUCED LUNCH

```
In [35]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='lunch',y='reading score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='lunch',y='reading score',data=df)
```

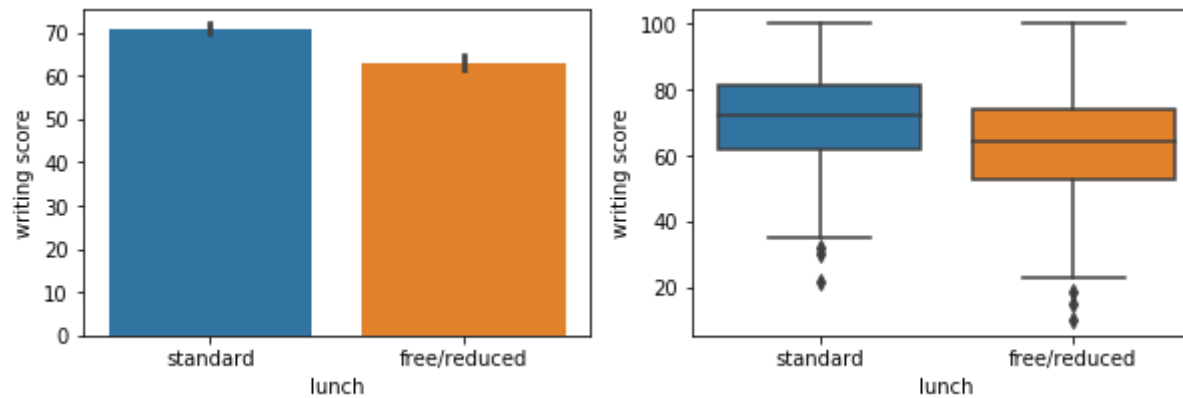
Out[35]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5782088>



STUDENTS WHO CAN AFFORD STANDARD LUNCH ARE PERFORMING BETTER IN READING COMPARED TO STUDENTS WHO GET FREE/REDUCED LUNCH

```
In [36]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='lunch',y='writing score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='lunch',y='writing score',data=df)
```

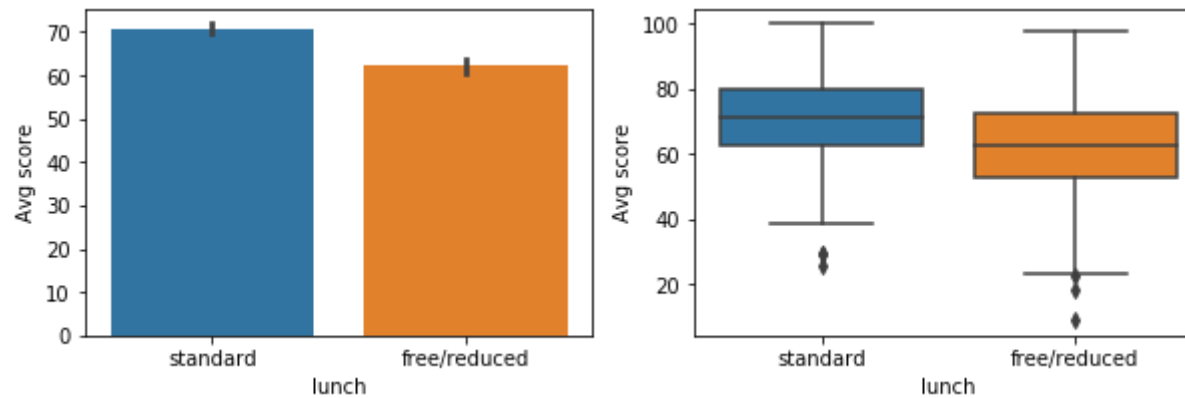
Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5856208>



STUDENTS WHO CAN AFFORD STANDARD LUNCH ARE PERFORMING BETTER IN WRITING COMPARED TO STUDENTS WHO GET FREE/REDUCED LUNCH

```
In [37]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='lunch',y='Avg score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='lunch',y='Avg score',data=df)
```

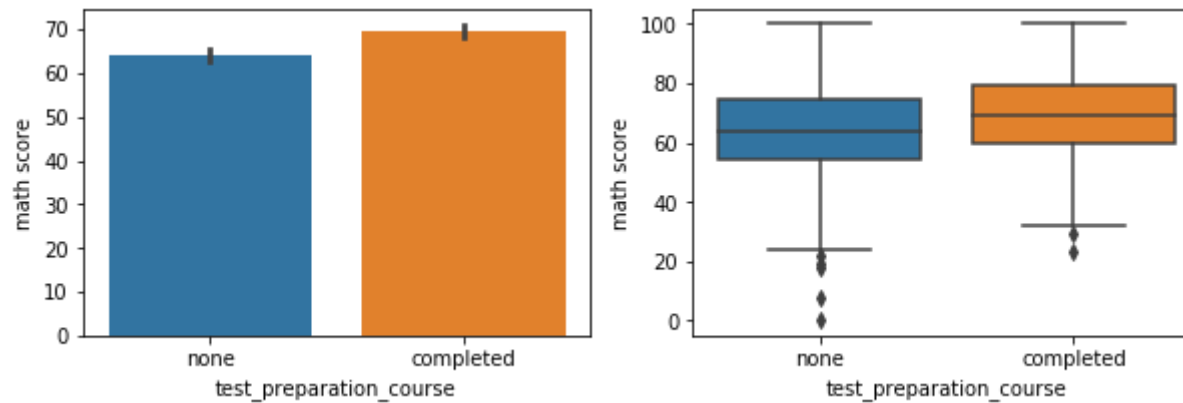
Out[37]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5b03fc8>



STUDENTS WHO CAN AFFORD STANDARD LUNCH ARE PERFORMING BETTER IN ALL TESTS COMPARED TO STUDENTS WHO GET FREE/REDUCED LUNCH

```
In [38]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='test_preparation_course',y='math score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='test_preparation_course',y='math score',data=df)
```

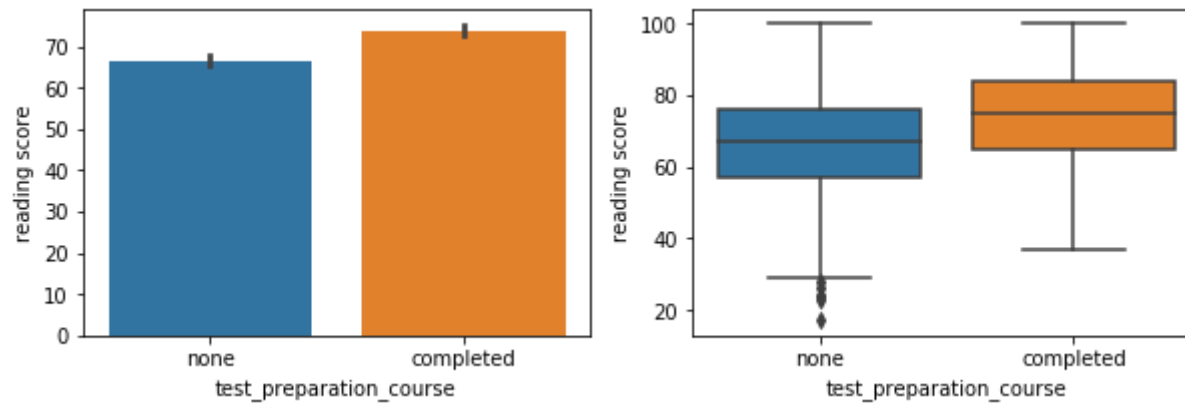
Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5bde308>



STUDENTS WHO COMPLETED THE TEST PREPARATION COURSE ARE PERFORMING BETTER IN MATHS THAN OTHER STUDENTS WHO HAVEN'T

```
In [39]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='test_preparation_course',y='reading score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='test_preparation_course',y='reading score',data=df)
```

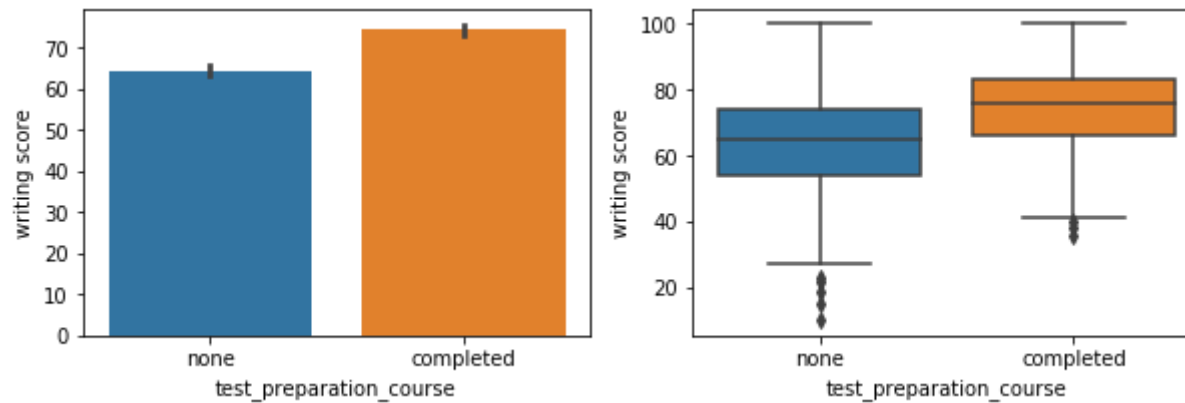
Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5caff08>



STUDENTS WHO COMPLETED THE TEST PREPARATION COURSE ARE PERFORMING BETTER IN READING THAN OTHER STUDENTS WHO HAVEN'T

```
In [40]: plt.figure(figsize=(10,3))
plt.subplot(1,2,1)
sns.barplot(x='test_preparation_course',y='writing score',data=df)
plt.subplot(1,2,2)
sns.boxplot(x='test_preparation_course',y='writing score',data=df)
```

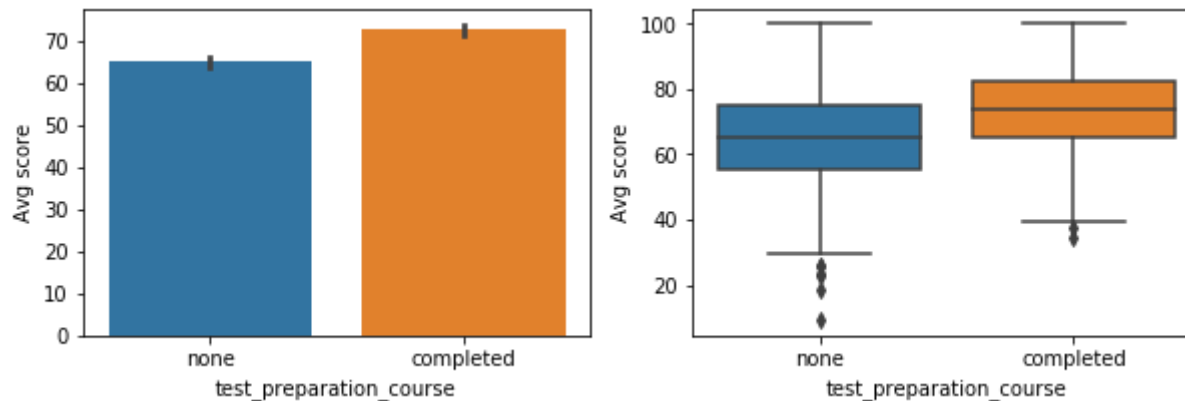
Out[40]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5d84b48>



STUDENTS WHO COMPLETED THE TEST PREPARATION COURSE ARE PERFORMING BETTER IN WRITING THAN OTHER STUDENTS WHO HAVEN'T

```
In [41]: plt.figure(figsize=(10,3))  
plt.subplot(1,2,1)  
sns.barplot(x='test_preparation_course',y='Avg score',data=df)  
plt.subplot(1,2,2)  
sns.boxplot(x='test_preparation_course',y='Avg score',data=df)
```

Out[41]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5e5c048>



STUDENTS WHO COMPLETED THE TEST PREPARATION COURSE ARE PERFORMING BETTER IN ALL TESTS THAN OTHER STUDENTS WHO HAVEN'T

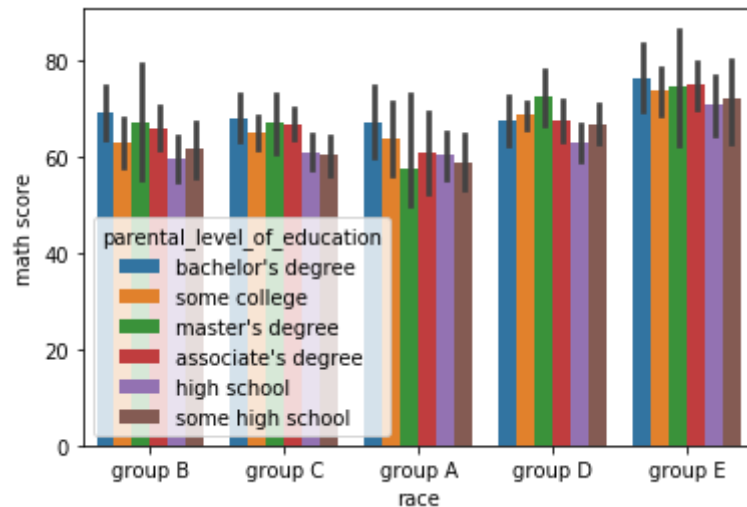
INFERENCES:-

PARENTS'S EDUCATION LEVEL IS AFFECTING THE STUDENTS PERFORMANCE AS THEY THEY CAN GUIDE THEM.

TEST PREPARATION COURSE IS REALLY USEFUL IN IMPROVING THE TEST PERFORMANCE.


```
In [42]: sns.barplot(x='race',y='math score',data=df,hue='parental_level_of_education')
```

```
Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa5ed4448>
```



IN GROUP A, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN MATHS WHILE, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE NOT PERFORMING WELL

IN GROUP B, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN MATHS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

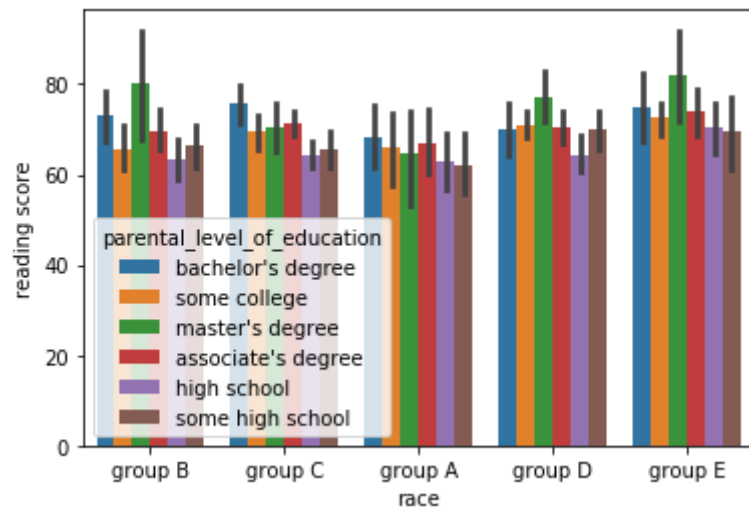
IN GROUP C, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE, MASTER'S DEGREE OR ASSOCIATE'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN MATHS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP D, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN MATHS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP E, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE, MASTER'S DEGREE OR ASSOCIATE'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN MATHS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

```
In [43]: sns.barplot(x='race',y='reading score',data=df,hue='parental_level_of_education')
```

```
Out[43]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa6009dc8>
```



IN GROUP A, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE OR ASSOCIATE'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN READING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP B, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN READING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

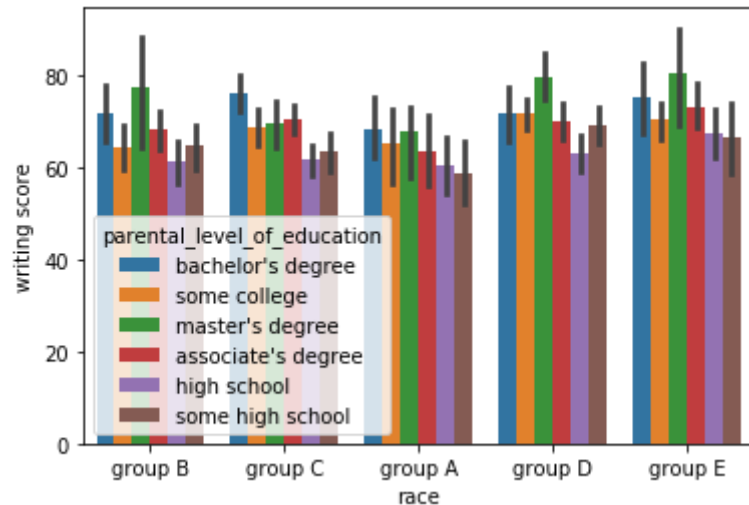
IN GROUP C, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN READING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP D, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN READING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP E, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN READING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

```
In [44]: sns.barplot(x='race',y='writing score',data=df,hue='parental_level_of_education')
```

```
Out[44]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa6141d88>
```



IN GROUP A, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE OR MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN WRITING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP B, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN WRITING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

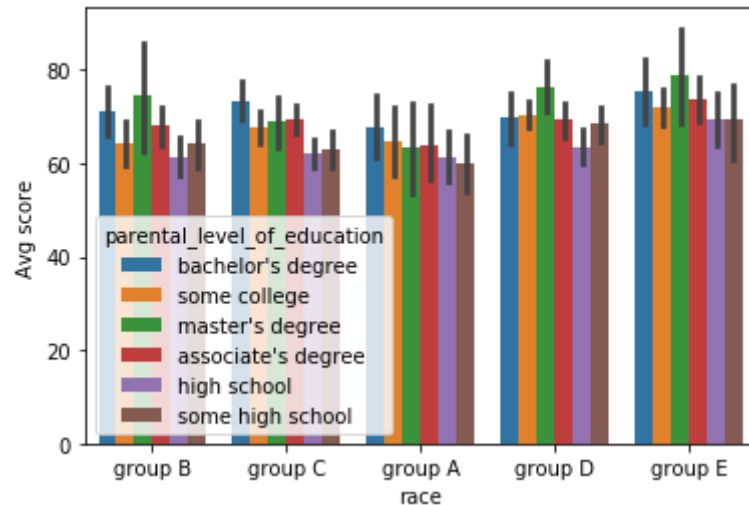
IN GROUP C, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN WRITING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP D, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN WRITING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP E, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS IN WRITING WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

```
In [45]: sns.barplot(x='race',y='Avg score',data=df,hue='parental_level_of_education')
```

```
Out[45]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa0bb6388>
```



IN GROUP A, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE ARE PERFORMING BETTER THAN OTHERS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP B, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP C, STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE ARE PERFORMING BETTER THAN OTHERS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP D, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

IN GROUP E, STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER THAN OTHERS WHILE, STUDENTS WHOSE PARENTS HAVE GONE TILL HIGH SCHOOL ARE NOT PERFORMING WELL

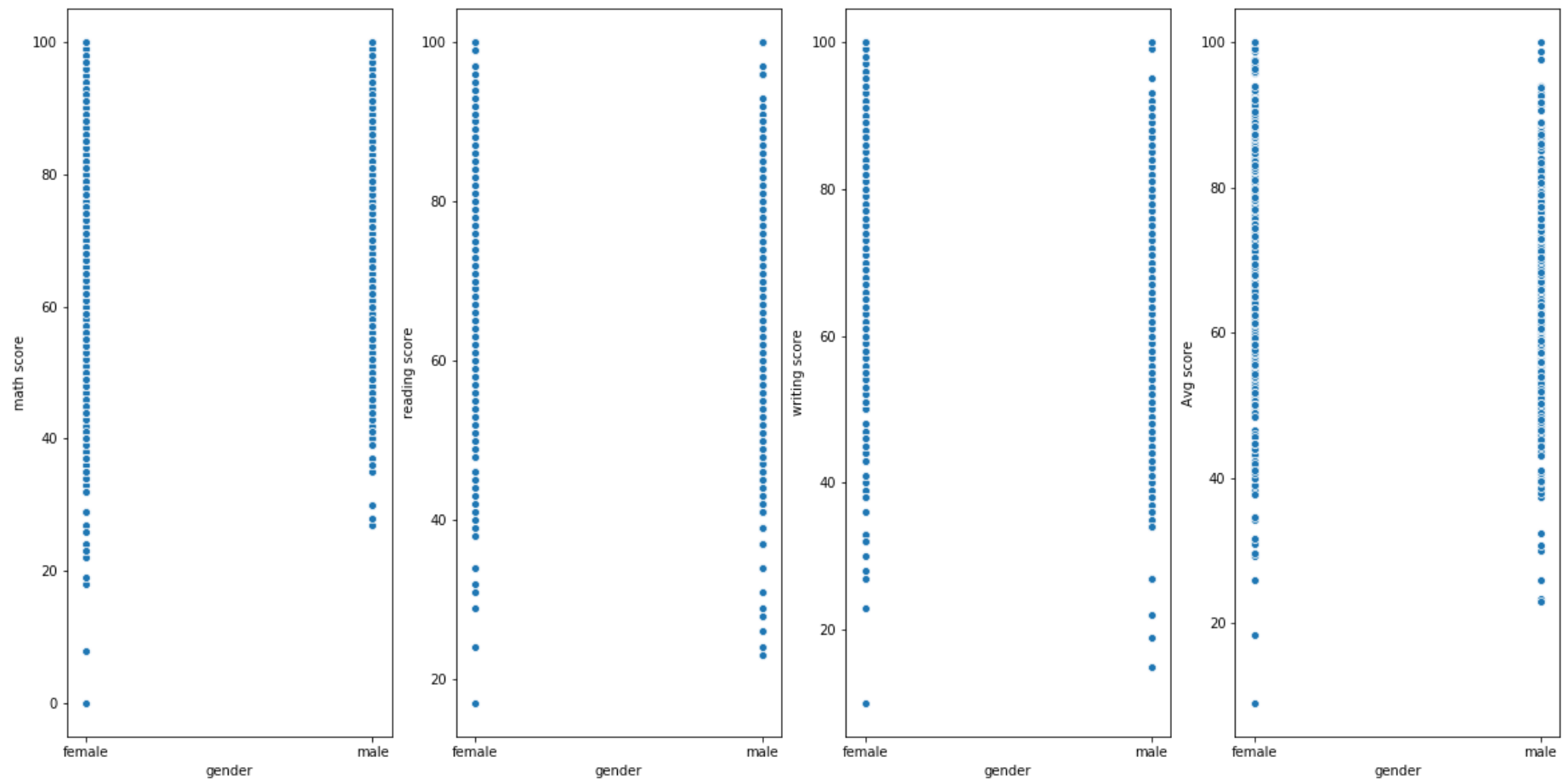
INFERENCE:-

PARENTAL EDUCATIONAL LEVEL HAS A HIGH IMPACT ON STUDENT'S PERFORMANCE.

STUDENTS WHOSE PARENTS HAVE BACHELOR'S DEGREE ARE PERFORMING BETTER IN MATHS COMPARED TO OTHERS, WHILE STUDENTS WHOSE PARENTS HAVE MASTER'S DEGREE ARE PERFORMING BETTER IN READING AND WRITING

```
In [46]: plt.figure(figsize=(20,10))
plt.subplot(1,4,1)
sns.scatterplot(x="gender",y='math score',data=df)
plt.subplot(1,4,2)
sns.scatterplot(x="gender",y='reading score',data=df)
plt.subplot(1,4,3)
sns.scatterplot(x="gender",y='writing score',data=df)
plt.subplot(1,4,4)
sns.scatterplot(x="gender",y='Avg score',data=df)
```

Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa644de88>



THERE ARE SOME FEMALE STUDENTS WHO ARE PERFORMING POORLY, BOTH GENDERS ARE ALMOST PERFORMING EQUALLY WITH

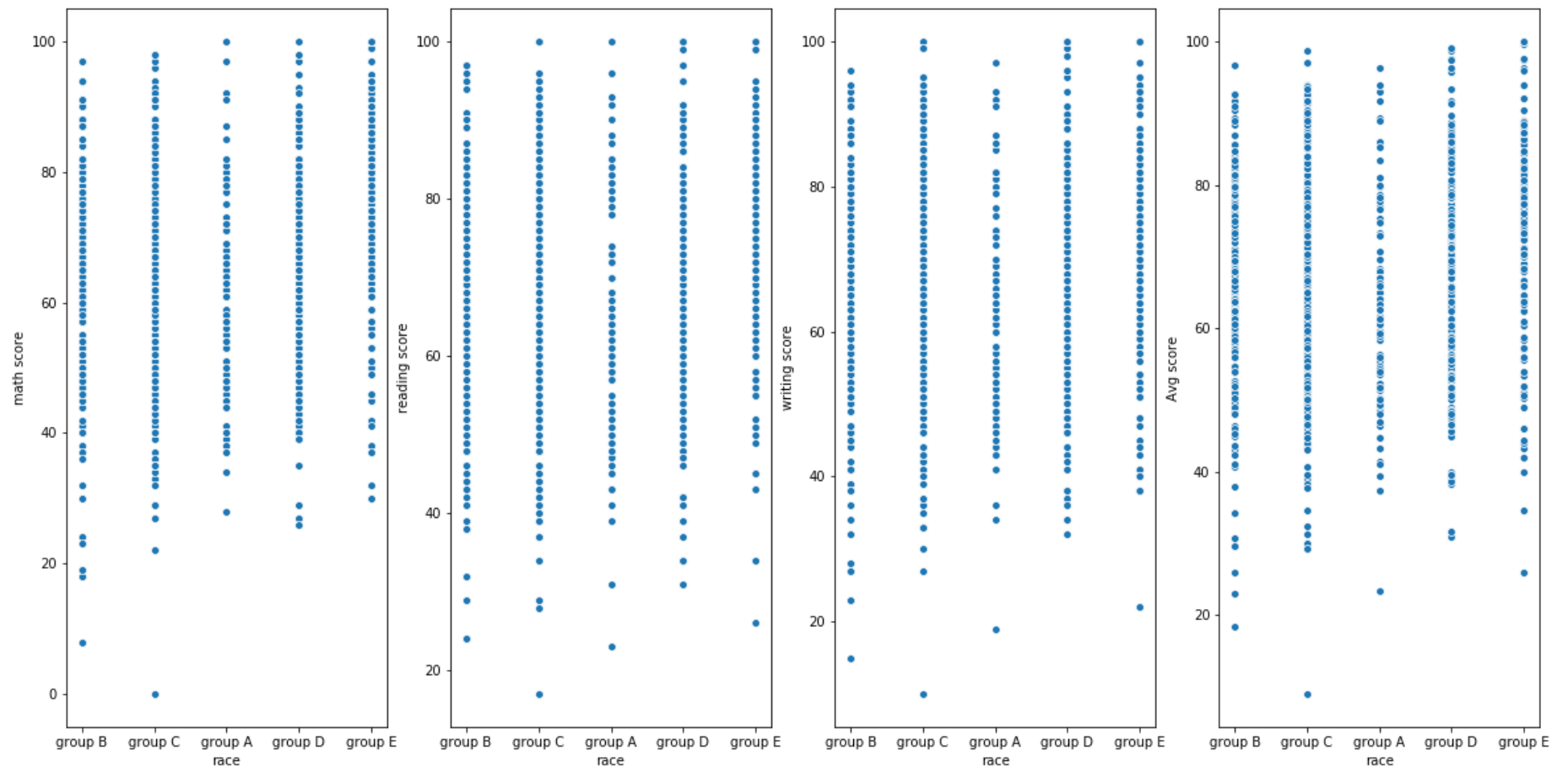
SOME EXPECTATIONS

```

In [47]: plt.figure(figsize=(20,10))
plt.subplot(1,4,1)
sns.scatterplot(x="race",y='math score',data=df)
plt.subplot(1,4,2)
sns.scatterplot(x="race",y='reading score',data=df)
plt.subplot(1,4,3)
sns.scatterplot(x="race",y='writing score',data=df)
plt.subplot(1,4,4)
sns.scatterplot(x="race",y='Avg score',data=df)

```

Out[47]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa6593348>



ALMOST EVERY GROUP A AND GROUP D STUDENT IS PERFORMING GOOD

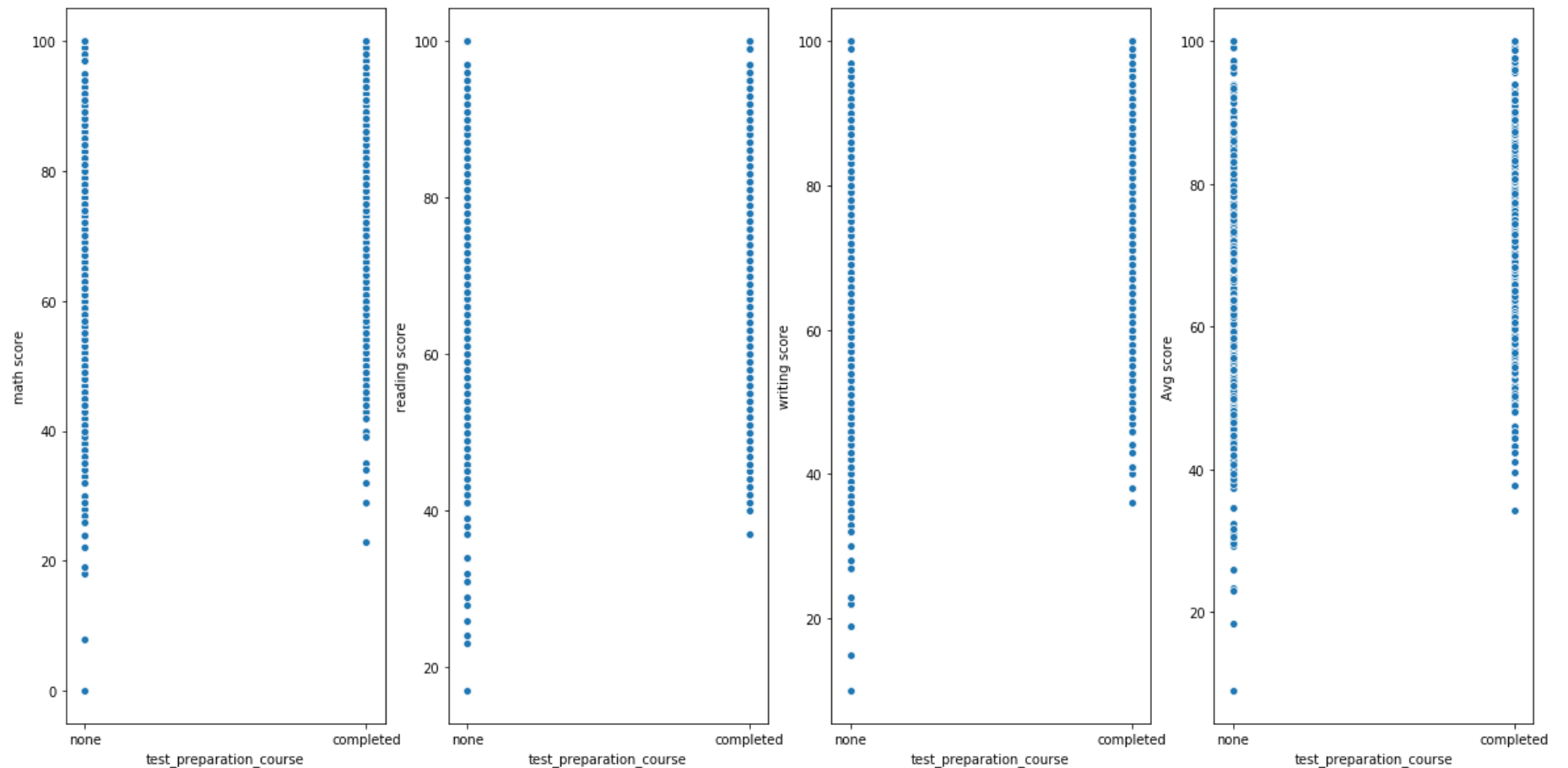
MOST OF THE STUDENTS WHO ARE SCORING LESS ARE FROM MOSTLY GROUP B AND GROUP C

```

In [48]: plt.figure(figsize=(20,10))
plt.subplot(1,4,1)
sns.scatterplot(x="test_preparation_course",y='math score',data=df)
plt.subplot(1,4,2)
sns.scatterplot(x="test_preparation_course",y='reading score',data=df)
plt.subplot(1,4,3)
sns.scatterplot(x="test_preparation_course",y='writing score',data=df)
plt.subplot(1,4,4)
sns.scatterplot(x="test_preparation_course",y='Avg score',data=df)

```

Out[48]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa66e3908>

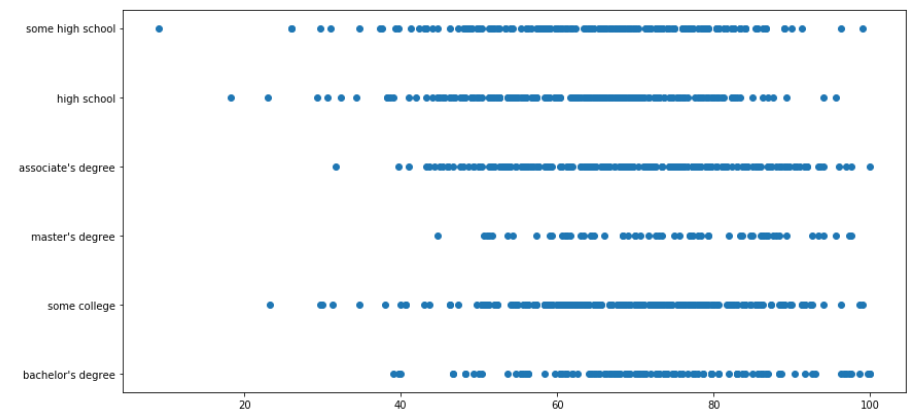
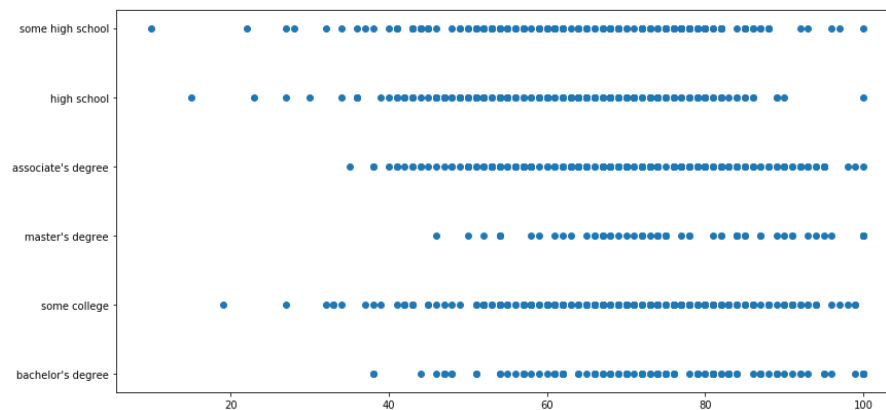
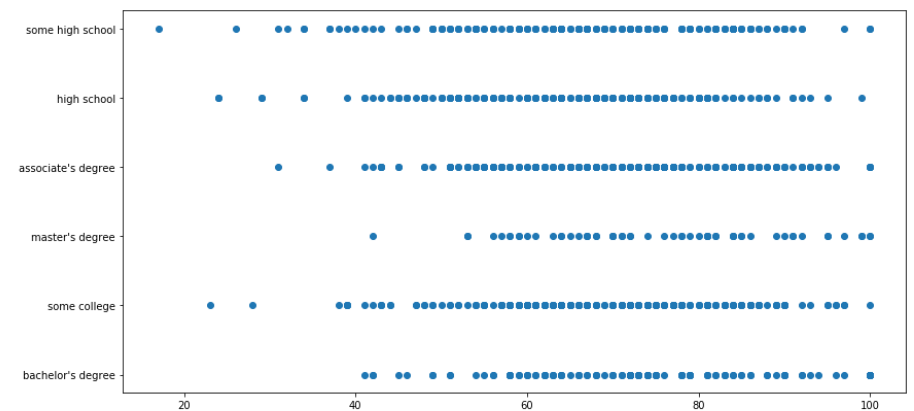
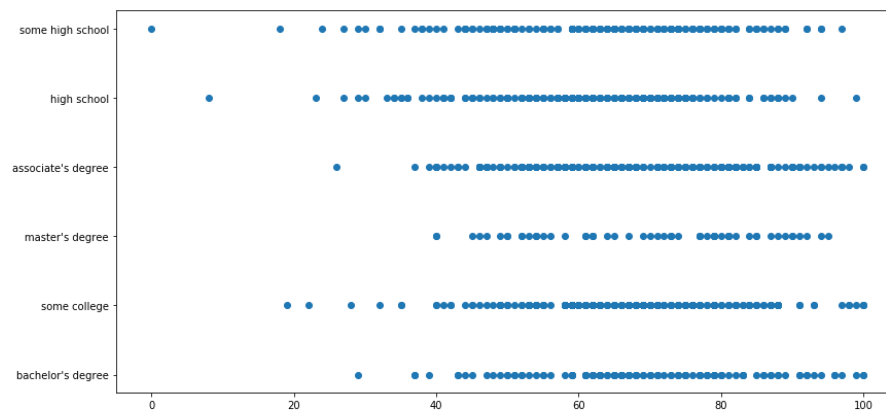


NONE OF THE STUDENTS WHO HAVE TAKEN THE COURSE HAVE SCORED LESS

SOME STUDENTS WHO HAVE NOT TAKEN THE COURSE ARE ALSO PERFORMING BETTER

```
In [49]: fig = plt.figure(figsize=(30,15))
sp1 = plt.subplot2grid((2,2),(0,0))
sp2 = plt.subplot2grid((2,2),(0,1))
sp3 = plt.subplot2grid((2,2),(1,0))
sp4 = plt.subplot2grid((2,2),(1,1))
sp1.scatter(y="parental_level_of_education",x='math score',data=df)
sp2.scatter(y="parental_level_of_education",x='reading score',data=df)
sp3.scatter(y="parental_level_of_education",x='writing score',data=df)
sp4.scatter(y="parental_level_of_education",x='Avg score',data=df)
```

Out[49]: <matplotlib.collections.PathCollection at 0x18fa68571c8>

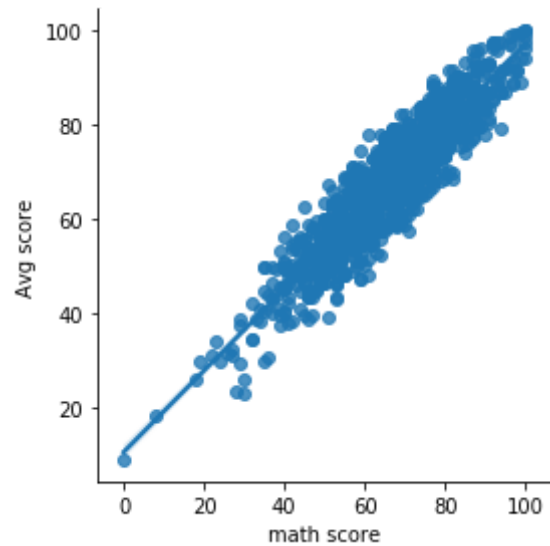


STUDENTS WHOSE PARENTS WHO HAVE GONE TILL HIGH SCHOOL OR COLLEGE HAVE PERFORMED POORLY, BUT VERY FEW OF THEM HAVE PERFORMED EXCEPTIONALLY GOOD

SO PARENT'S EDUCATIONAL LEVEL IS A MAJOR FACTOR BUT IT IS NOT THE ONLY FACTOR

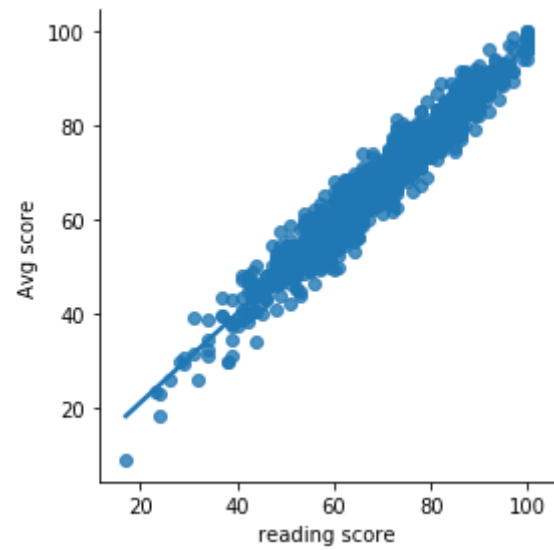
```
In [50]: sns.lmplot(x="math score",y="Avg score",data=df,height=4)
```

```
Out[50]: <seaborn.axisgrid.FacetGrid at 0x18fa68c4688>
```



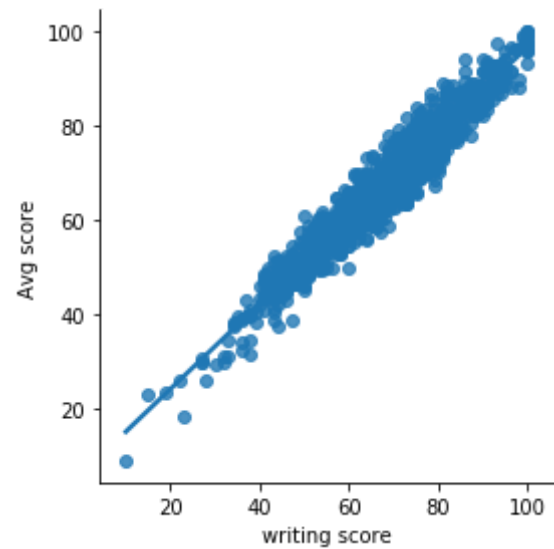
```
In [51]: sns.lmplot(x="reading score", y="Avg score", data=df,height=4)
```

```
Out[51]: <seaborn.axisgrid.FacetGrid at 0x18fa6be9488>
```



```
In [52]: sns.lmplot(x="writing score", y="Avg score", data=df,height=4)
```

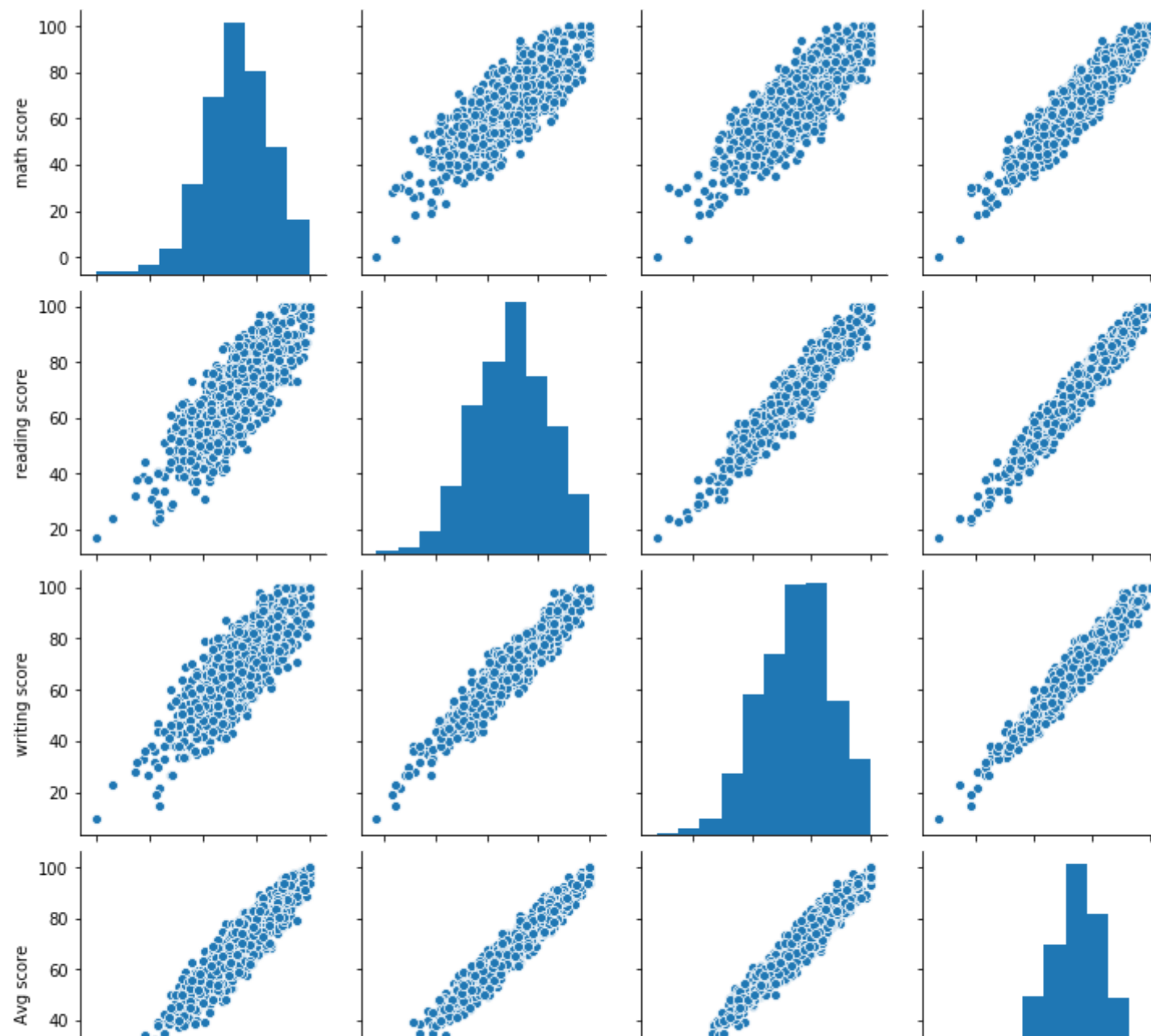
```
Out[52]: <seaborn.axisgrid.FacetGrid at 0x18fa6c42b48>
```

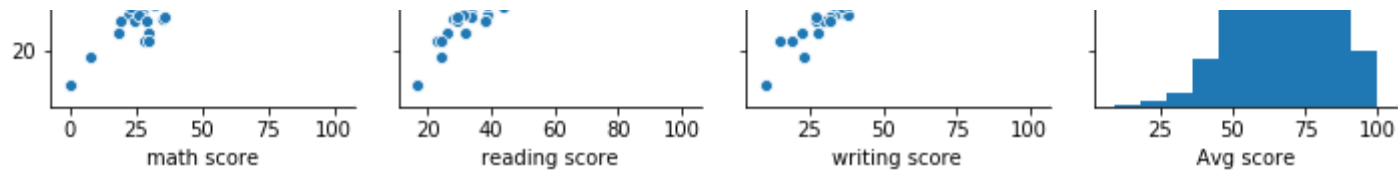


IF A STUDENT IS PERFORMING WELL IN ONE TEST THEN HE IS PERFORMING EQUALLY WELL IN ALL THE TESTS AS THE REGRESSION IS LINEARLY INCREASING

```
In [53]: sns.pairplot(df)
```

```
Out[53]: <seaborn.axisgrid.PairGrid at 0x18fa6cc9788>
```





```
In [54]: df.describe()
```

Out[54]:

	math score	reading score	writing score	Avg score
count	1000.00000	1000.000000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000	67.770667
std	15.16308	14.600192	15.195657	14.257326
min	0.00000	17.000000	10.000000	9.000000
25%	57.00000	59.000000	57.750000	58.333333
50%	66.00000	70.000000	69.000000	68.333333
75%	77.00000	79.000000	79.000000	77.666667
max	100.00000	100.000000	100.000000	100.000000

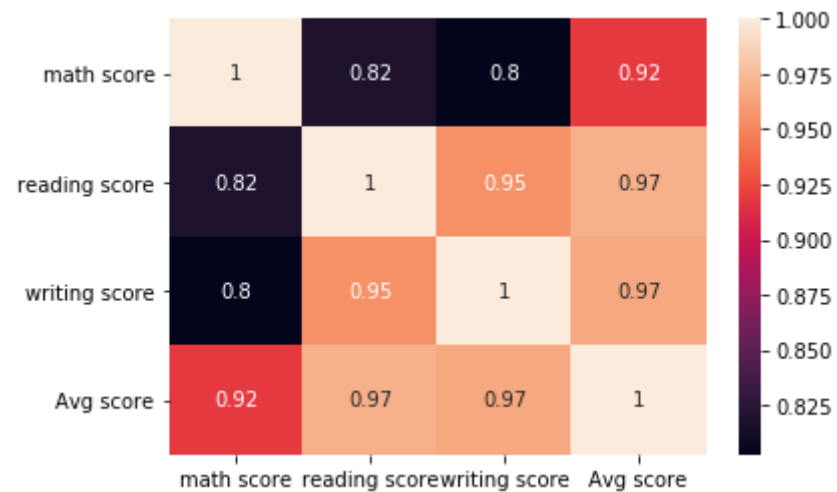
```
In [55]: df.corr()
```

Out[55]:

	math score	reading score	writing score	Avg score
math score	1.000000	0.817580	0.802642	0.918746
reading score	0.817580	1.000000	0.954598	0.970331
writing score	0.802642	0.954598	1.000000	0.965667
Avg score	0.918746	0.970331	0.965667	1.000000


```
In [56]: sns.heatmap(df.corr(),annot=True)
```

```
Out[56]: <matplotlib.axes._subplots.AxesSubplot at 0x18fa8ae45c8>
```



STUDENTS WHO PERFORM WELL IN READING TEST, GIVE THEIR BEST PERFORMANCE IN ALL THE TESTS, FOLLOWED BY STUDENTES WHO PERFORM WELL IN WRITING TEST THEN MATHS TEST