In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

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

In [5]:
```
df.head()
```

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

In [6]:
```
df.tail()
```

Out[6]:

	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 [7]:
```
df.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]:
```
df.describe()
```

Out[8]:

	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 [10]:
```
df.shape
```

Out[10]:
```
(1000, 8)
```

In [11]:
```
df.dtypes
```

Out[11]:
```
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
```

In [13]:
```
df.isnull().sum()
```

Out[13]:
```
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
```

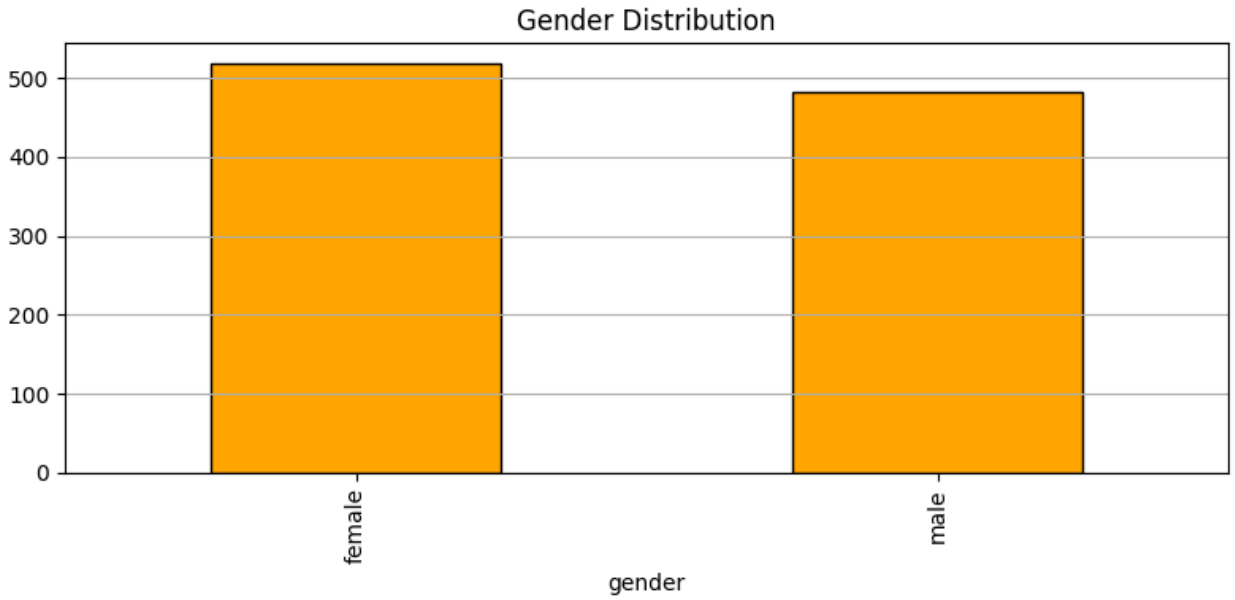
In [14]:
```
df.duplicated().sum()
```

Out[14]:
```
np.int64(0)
```

In [15]:
```
df.columns
```

Out[15]:
```
Index(['gender', 'race/ethnicity', 'parental level of education', 'lunch',
       'test preparation course', 'math score', 'reading score',
       'writing score'],
      dtype='object')
```

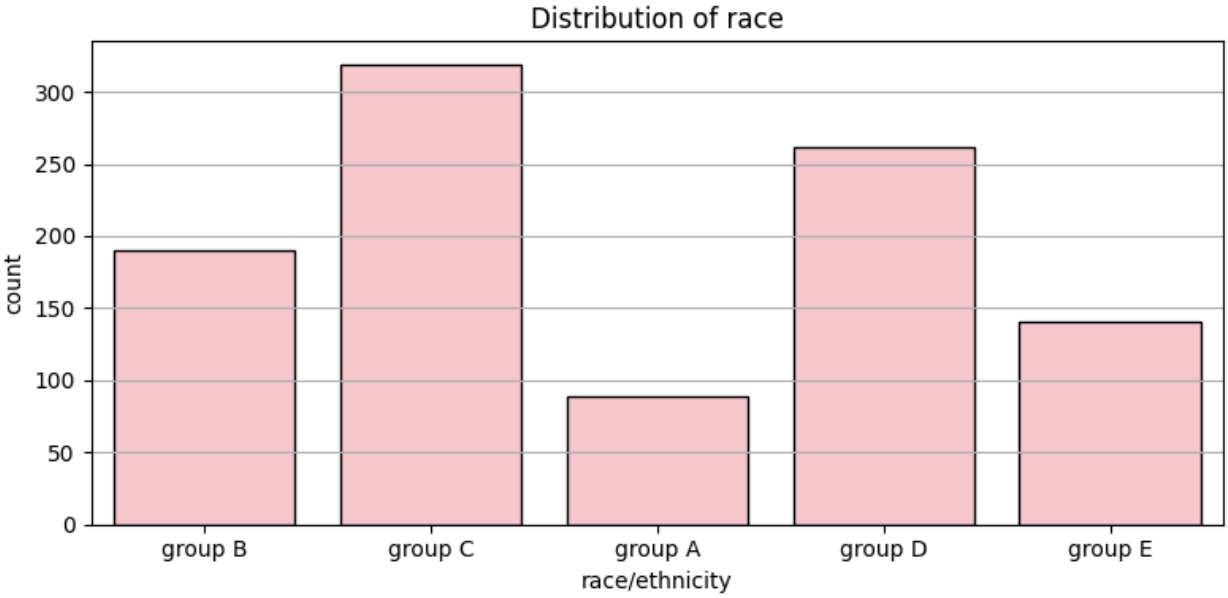
In [26]:
```
plt.figure(figsize = (8,4))
df["gender"].value_counts().plot(kind = "bar", color = "orange", edgecolor = "
plt.title("Gender Distribution")
plt.grid(axis = "y")
plt.tight_layout()
plt.show()
```



The proportion of female students is around 52%, while male students make up about 48%.

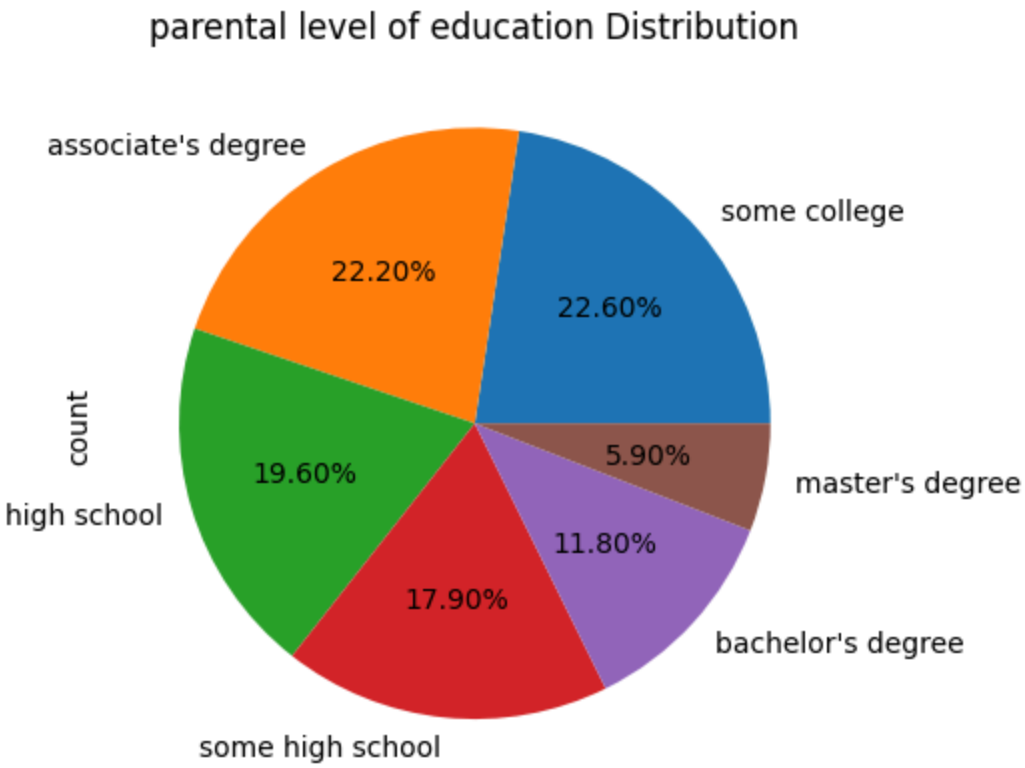
In [38]:
```
plt.figure(figsize = (8,4))
sns.countplot(data = df , x = "race/ethnicity", color = "pink", edgecolor = "b
plt.grid(axis = "y")
```

```
plt.title("Distribution of race")
plt.tight_layout()
plt.show()
```



Most students belong to Group C, while Group A has the least representation.

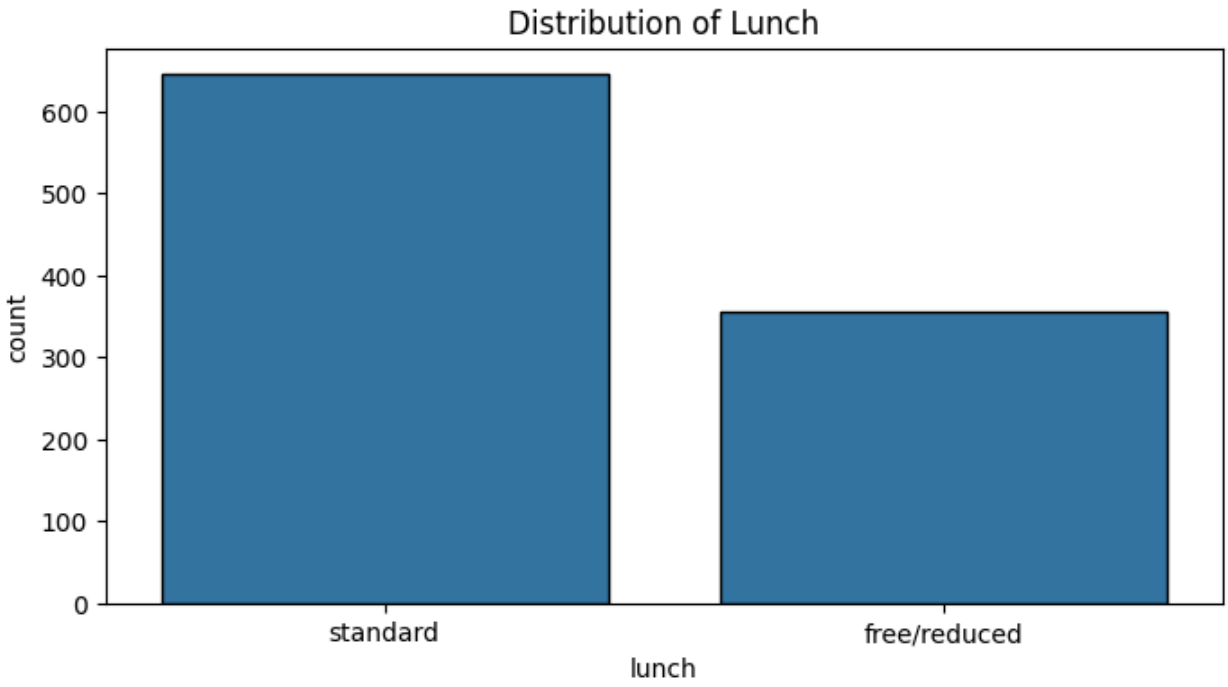
In [42]:
```
df["parental level of education"].value_counts().plot(kind = "pie", autopct =
plt.title("parental level of education Distribution")
plt.show()
```



The majority of parents have completed some college education or hold an associate’s degree,

indicating a moderately educated parent group, while only a small percentage have attained a master’s degree.

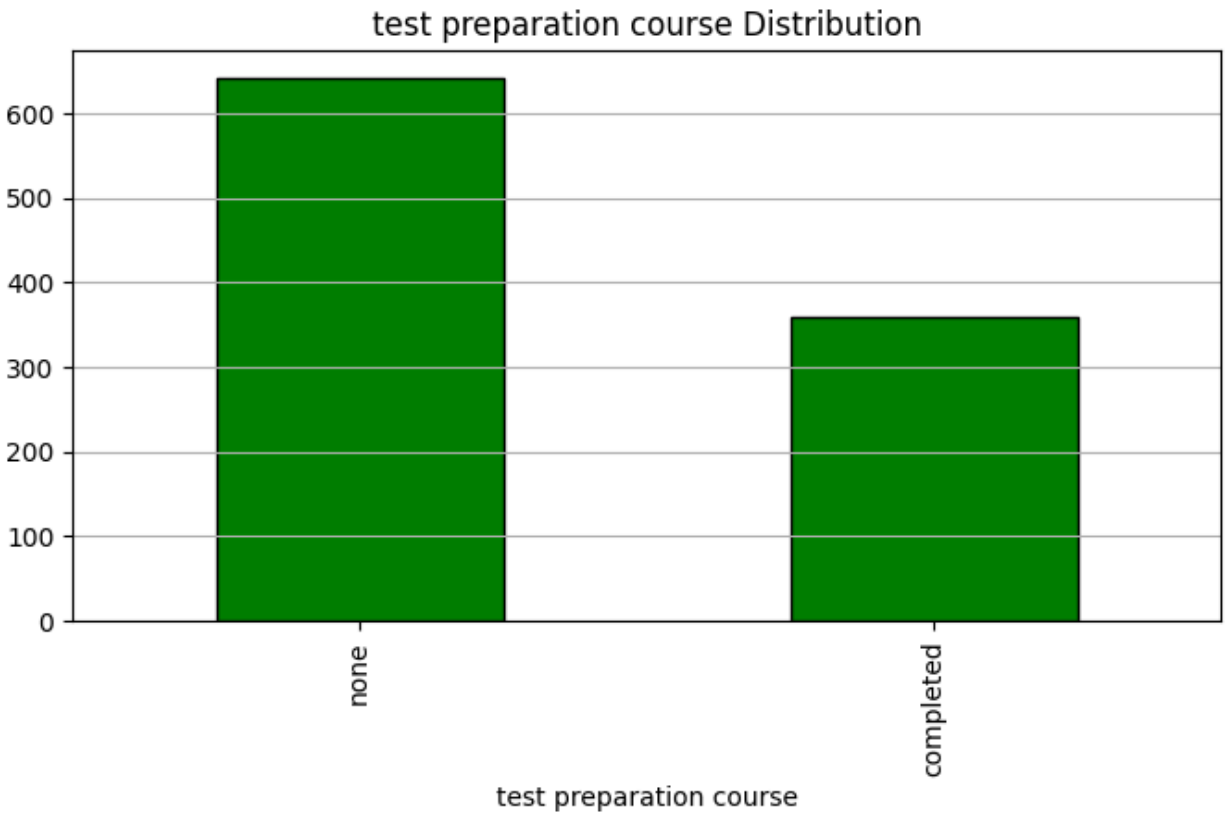
In [49]:
```
plt.figure(figsize = (8,4))
sns.countplot(data = df , x = "lunch", edgecolor = "black")
plt.title("Distribution of Lunch")
plt.show()
```



A majority of students receive a standard lunch, while a smaller group benefits from free or reduced lunch,

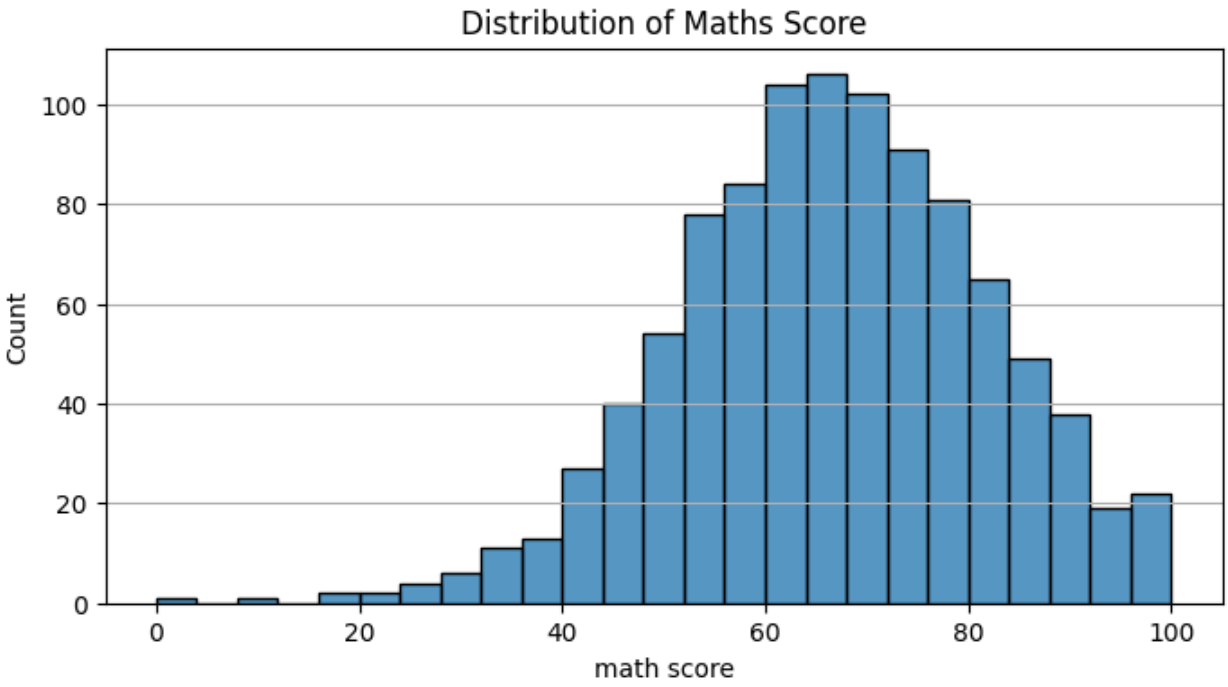
indicating varying economic backgrounds.

In [59]:
```
plt.figure(figsize = (8,4))
df["test preparation course"].value_counts().plot(kind = "bar", color = "green
plt.title("test preparation course Distribution")
plt.grid(axis = "y")
plt.show()
```



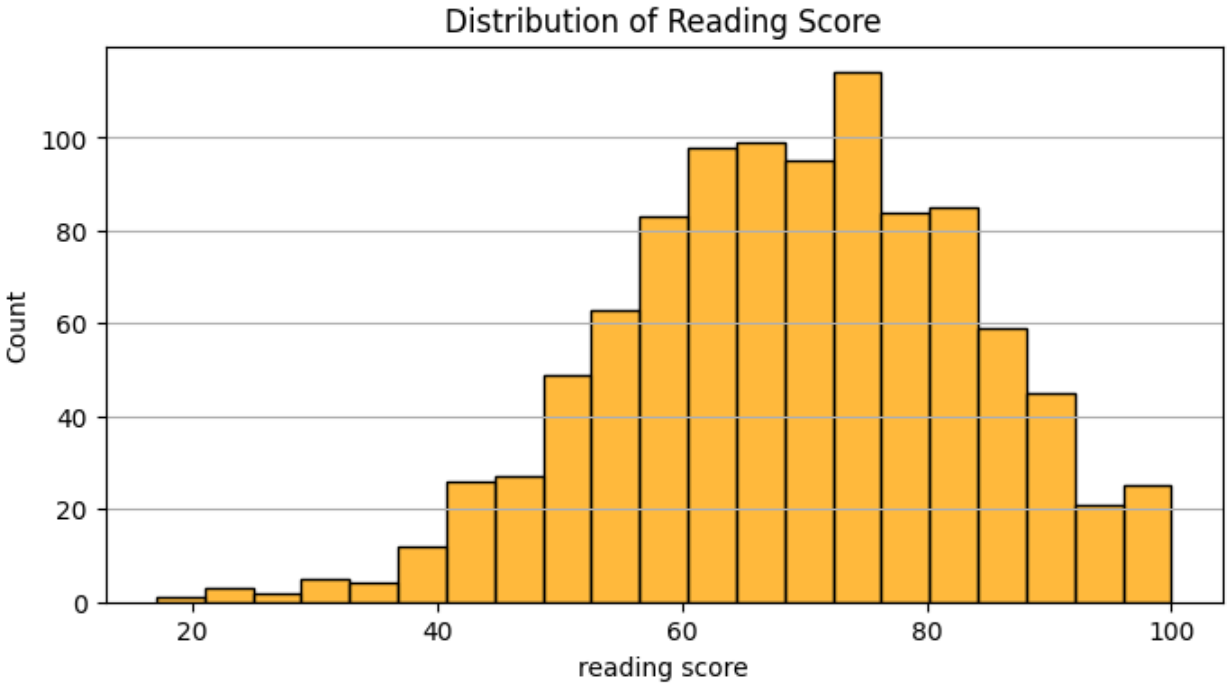
Most students did not complete the test preparation course, suggesting limited participation in extra academic support programs.

In [63]:
```
plt.figure(figsize = (8,4))
sns.histplot(data = df , x = "math score", edgecolor = "black")
plt.title("Distribution of Maths Score")
plt.grid(axis = "y")
plt.show()
```



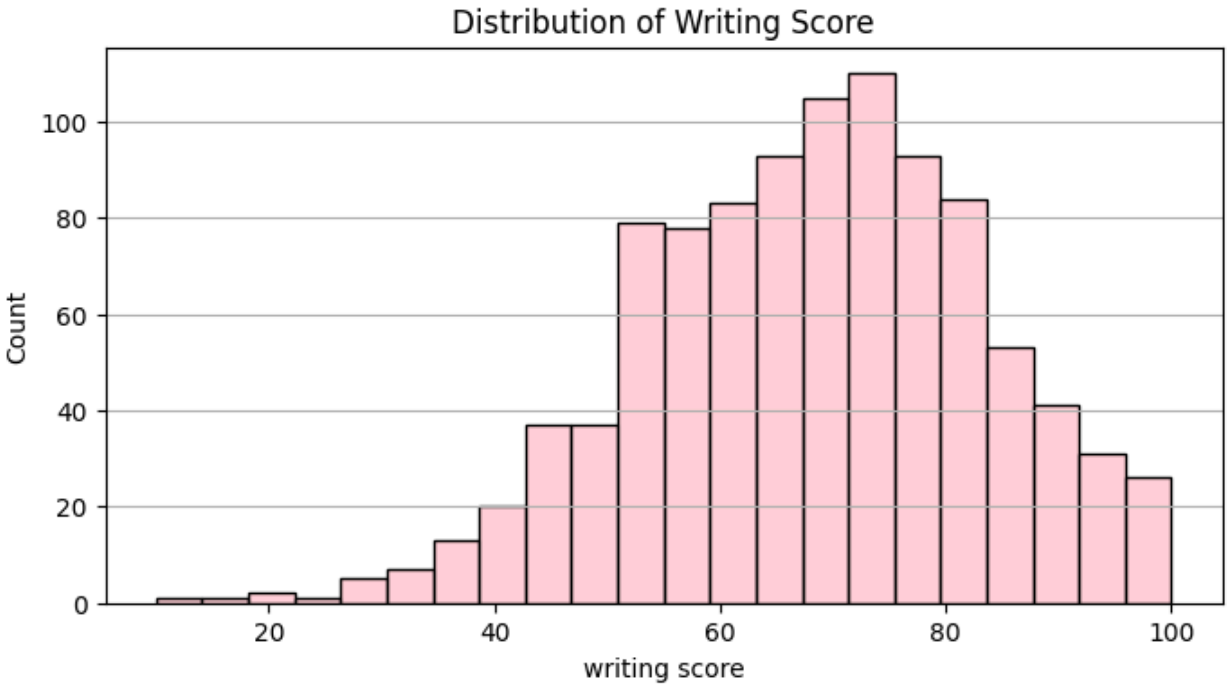
The distribution of maths scores is approximately normal, with most students scoring between 55 and 75.

In [65]:
```
plt.figure(figsize = (8,4))
sns.histplot(data = df , x = "reading score", edgecolor = "black", color = "Or
plt.title("Distribution of Reading Score")
plt.grid(axis = "y")
plt.show()
```



Most students performed well in reading, with scores mainly between 60 and 80, showing generally strong reading skills.

In [68]:
```
plt.figure(figsize = (8,4))
sns.histplot(data = df , x = "writing score", edgecolor = "black", color = "Pi
plt.title("Distribution of Writing Score")
plt.grid(axis = "y")
plt.show()
```



Most students performed well in Writing, with scores mainly between 60 and 80, showing generally strong reading skills.

In [72]:
```
df.groupby("gender")["math score"].mean()
```

Out[72]:
```
gender
female    63.633205
male      68.728216
Name: math score, dtype: float64
```

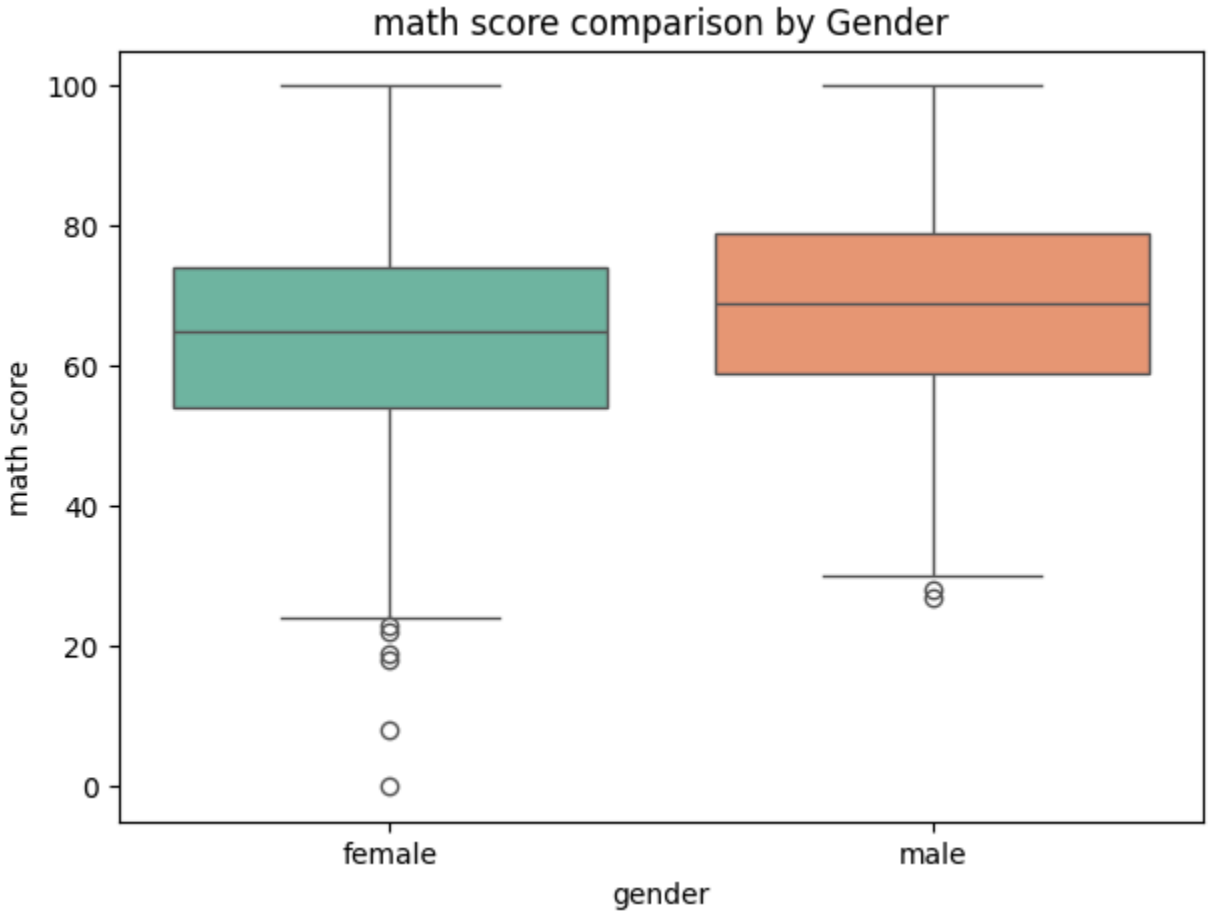
In [73]:
```
df.groupby("gender")["reading score"].mean()
```

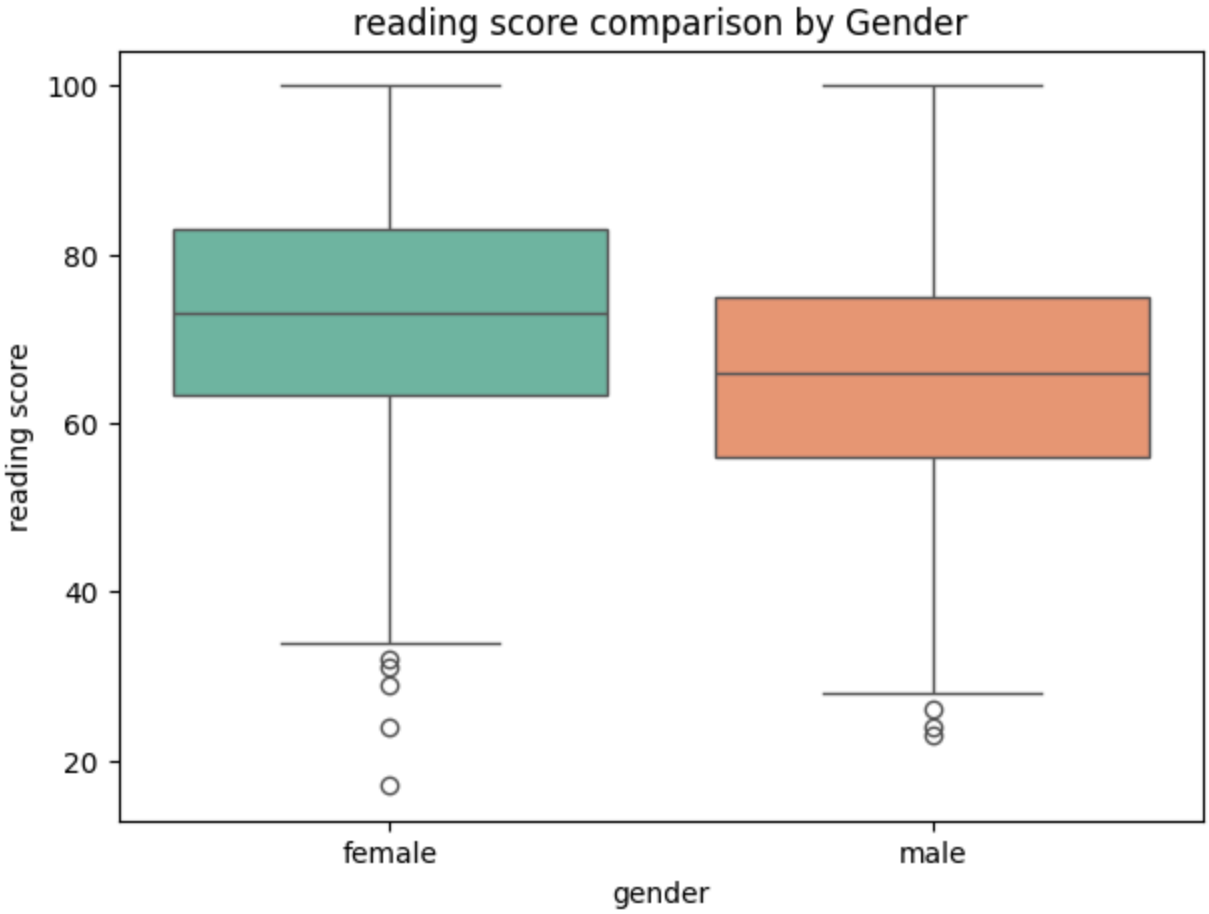
Out[73]:
```
gender
female    72.608108
male      65.473029
Name: reading score, dtype: float64
```

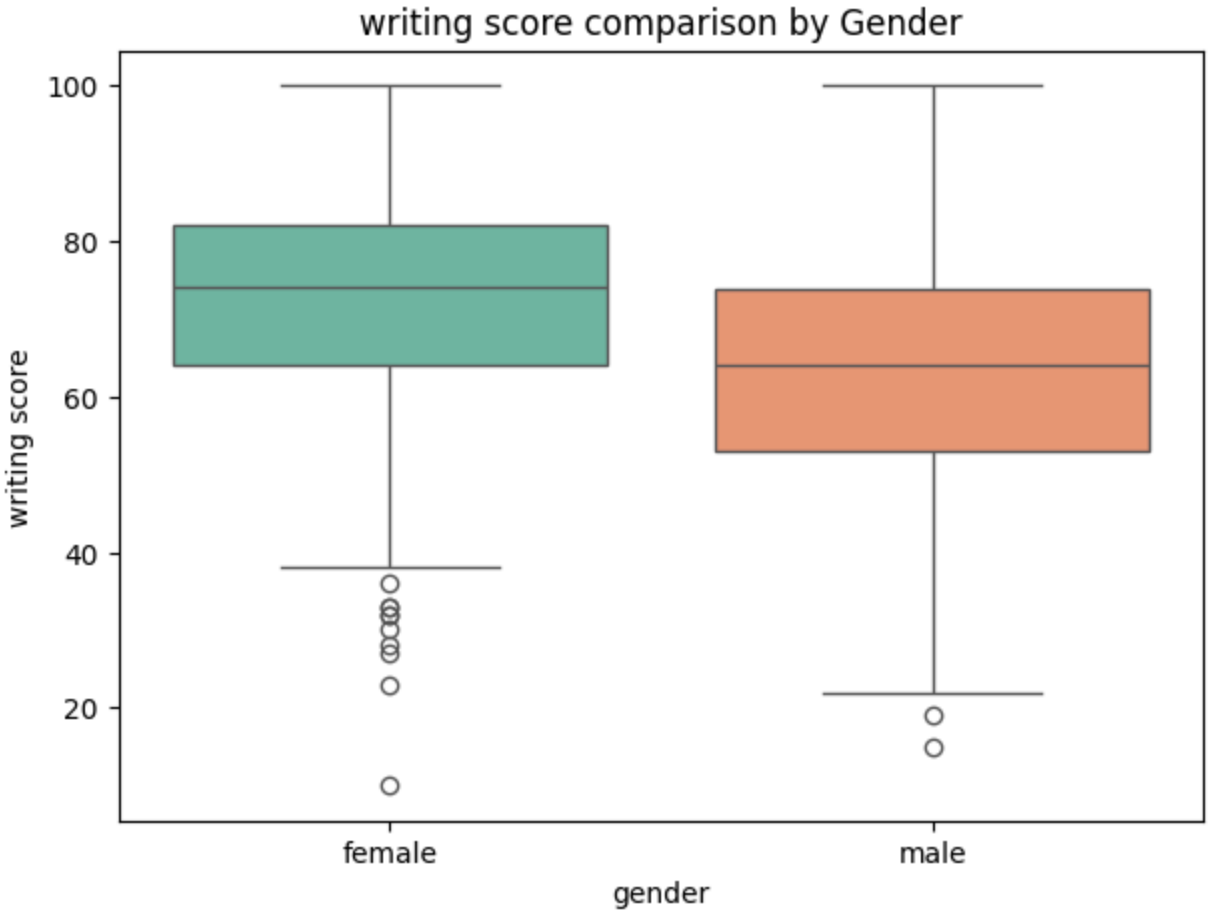
In [74]:
```
df.groupby("gender")["writing score"].mean()
```

Out[74]:
```
gender
female    72.467181
male      63.311203
Name: writing score, dtype: float64
```

In [82]:
```
for col in ["math score", "reading score", "writing score"]:
    plt.figure(figsize=(7,5))
    sns.boxplot(data=df, x="gender", y=col, palette="Set2")
    plt.title(f"{col} comparison by Gender")
    plt.show()
```




reading score comparison by Gender
100
80
60
40
20
reading score
female
male
gender

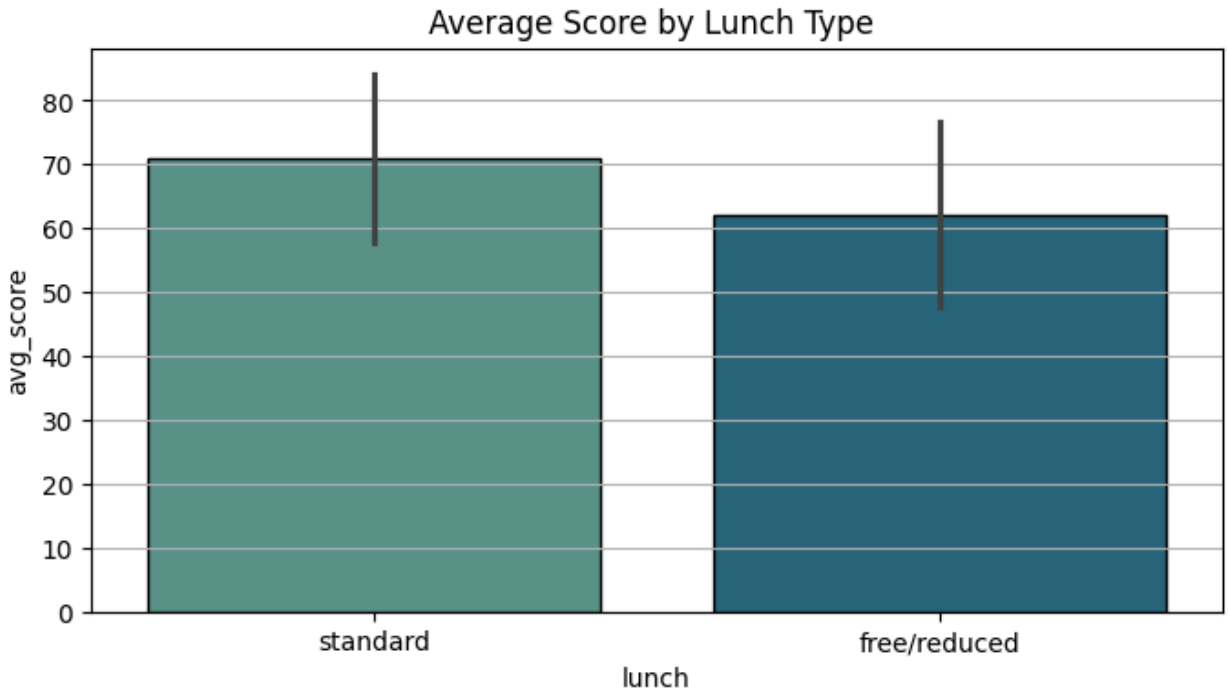


Overall, Male(boys) tend to do better in maths, while Females(girls) perform better in reading and writing, showing different strengths across subjects.

In [78]:
```
df["avg_score"] = df[["math score", "reading score", "writing score"]].mean(ax
```

In [80]:
```
import warnings
warnings.filterwarnings("ignore")
```

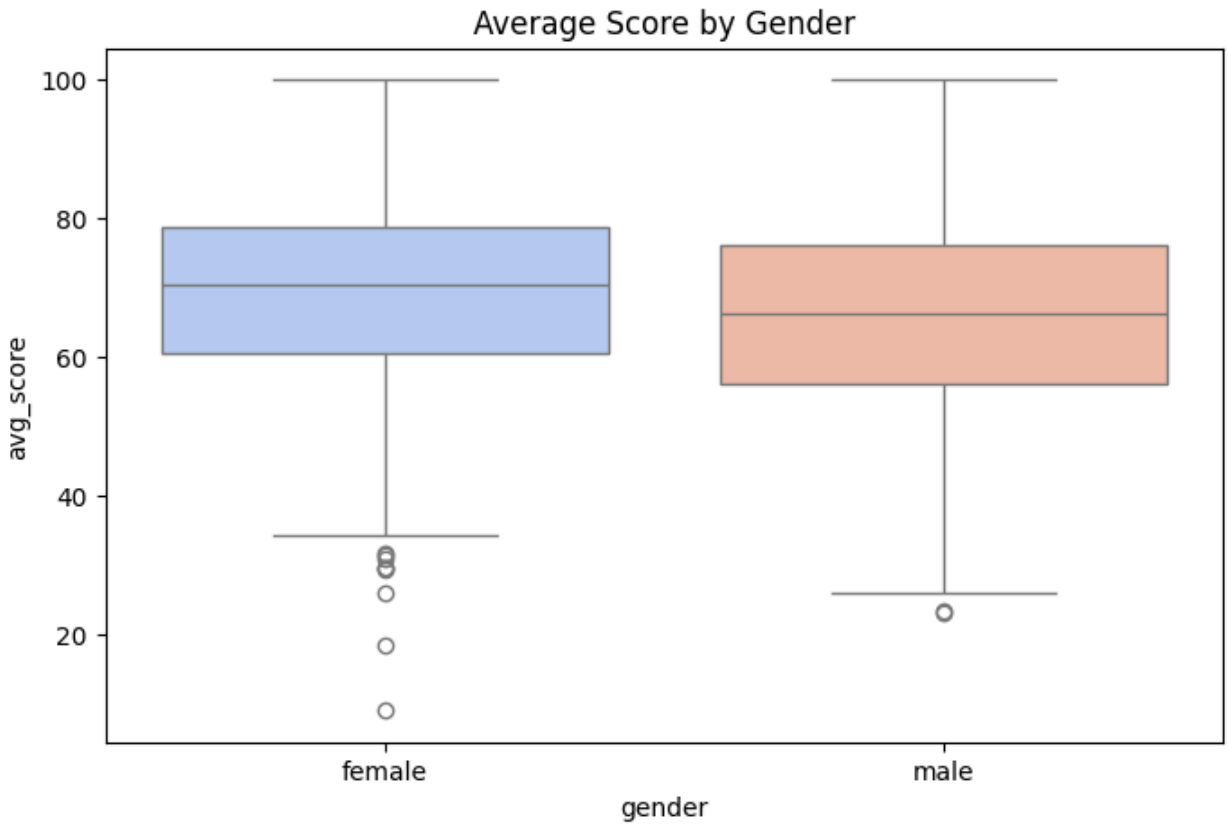
In [85]:
```
plt.figure(figsize=(8,4))
sns.barplot(data=df, x="lunch", y="avg_score", palette="crest", estimator='mea
plt.title("Average Score by Lunch Type ")
plt.grid(axis = "y")
plt.show()
```



Students who have a standard lunch generally score higher on average than those with free or reduced lunch,

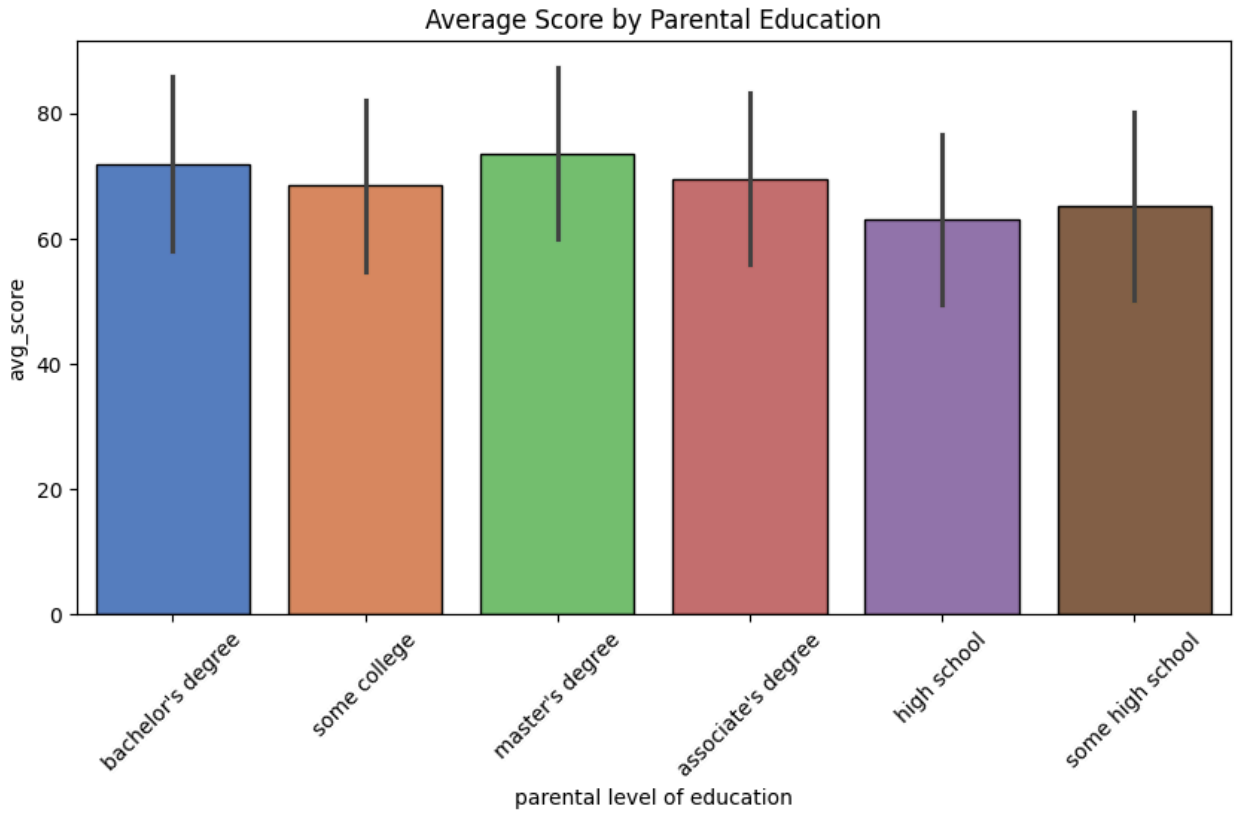
suggesting that better lunch options might be linked to better academic performance.

In [88]:
```
plt.figure(figsize=(8,5))
sns.boxplot(data=df, x='gender', y='avg_score', palette='coolwarm')
plt.title("Average Score by Gender")
plt.show()
```



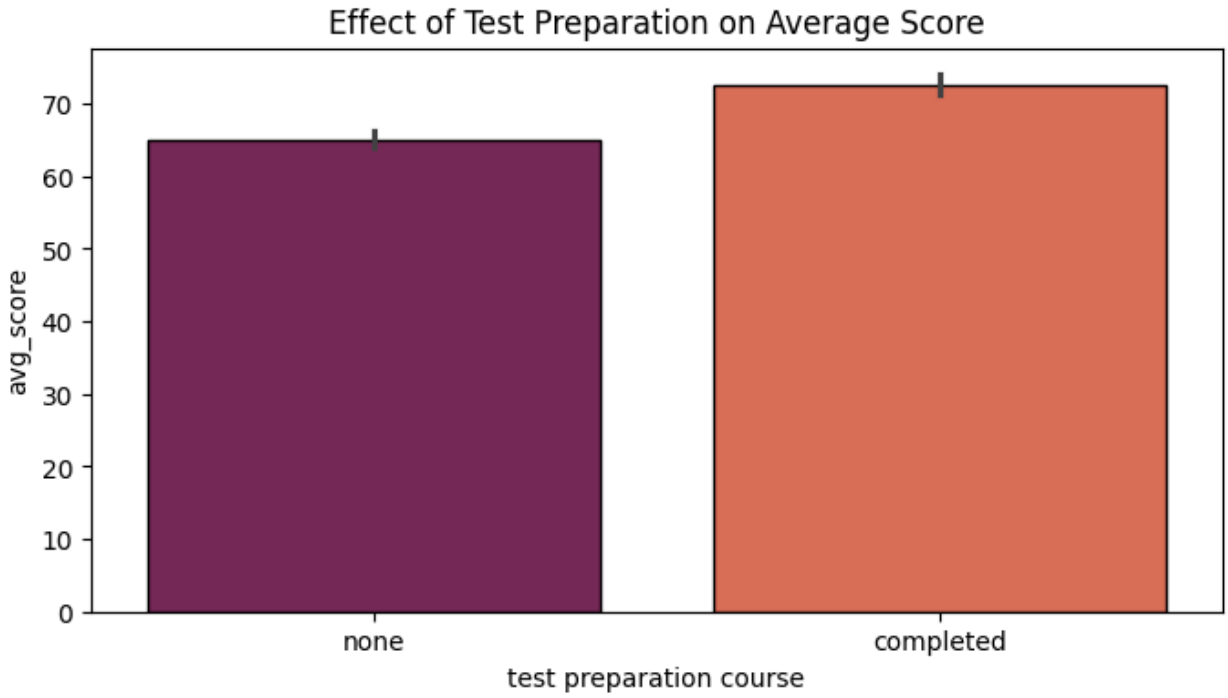
The average scores for both genders are quite similar, though females show slightly higher median scores, while males display a wider score range with more variation.

In [92]:
```
plt.figure(figsize=(10,5))
sns.barplot(data=df, x='parental level of education', y='avg_score', estimator
plt.xticks(rotation=45)
plt.title("Average Score by Parental Education")
plt.show()
```



Students whose parents have a master’s or bachelor’s degree tend to score slightly higher on average, suggesting that higher parental education may positively influence student performance.

In [96]:
```
plt.figure(figsize=(8,4))
sns.barplot(data=df, x='test preparation course', y='avg_score', palette='rock
plt.title("Effect of Test Preparation on Average Score")
plt.show()
```



Students who completed the test preparation course scored notably higher on average than those who didn’t, showing that test prep has a clear positive impact on performance.

In [ ]:

Over all Conclusion

In [ ]:

From this analysis of the Student Academic Performance dataset, we can see some clear patterns.

Student results are affected by a mix of personal, social, and family factors. The number of boys and girls is almost equal,

but boys usually score higher in math, while girls do better in reading and writing.

Family background also matters — students whose parents have higher education levels, like a bachelor’s or master’s degree,

tend to score better. Those who get a standard lunch also perform better than students with free or reduced lunch,

showing that family income and nutrition can impact learning.

Students who completed the test preparation course scored higher as well, proving that extra study support really helps.

Overall, student performance depends on gender, parental education, economic background, and academic support,

and improving these areas can help every student do their best.

In [ ]: