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

In [2]:
```
df = pd.read_csv("study_performance.csv")
print(df.head())
```

```
   gender race_ethnicity parental_level_of_education         lunch  \
0  female        group B           bachelor's degree      standard
1  female        group C                some college      standard
2  female        group B            master's degree      standard
3    male        group A          associate's degree  free/reduced
4    male        group C                some college      standard

  test_preparation_course  math_score  reading_score  writing_score
0                    none          72             72             74
1               completed          69             90             88
2                    none          90             95             93
3                    none          47             57             44
4                    none          76             78             75
```

In [3]:
```
df.describe()
```

Out[3]:

	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 [4]:
```
df.info
```

Out[4]:
```
<bound method DataFrame.info of      gender race_ethnicity parental_level_of_education         lunch  \
0    female        group B           bachelor's degree      standard
1    female        group C                some college      standard
2    female        group B            master's degree      standard
3      male        group A          associate's degree  free/reduced
4      male        group C                some college      standard
..      ...            ...                         ...           ...
995  female        group E            master's degree      standard
996    male        group C                 high school  free/reduced
997  female        group C                 high school  free/reduced
998  female        group D                some college      standard
999  female        group D                some college  free/reduced

    test_preparation_course  math_score  reading_score  writing_score
0                      none          72             72             74
1                 completed          69             90             88
2                      none          90             95             93
3                      none          47             57             44
4                      none          76             78             75
..                      ...         ...            ...            ...
995               completed          88             99             95
996                    none          62             55             55
997               completed          59             71             65
998               completed          68             78             77
999                    none          77             86             86

[1000 rows x 8 columns]>
```

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
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
```

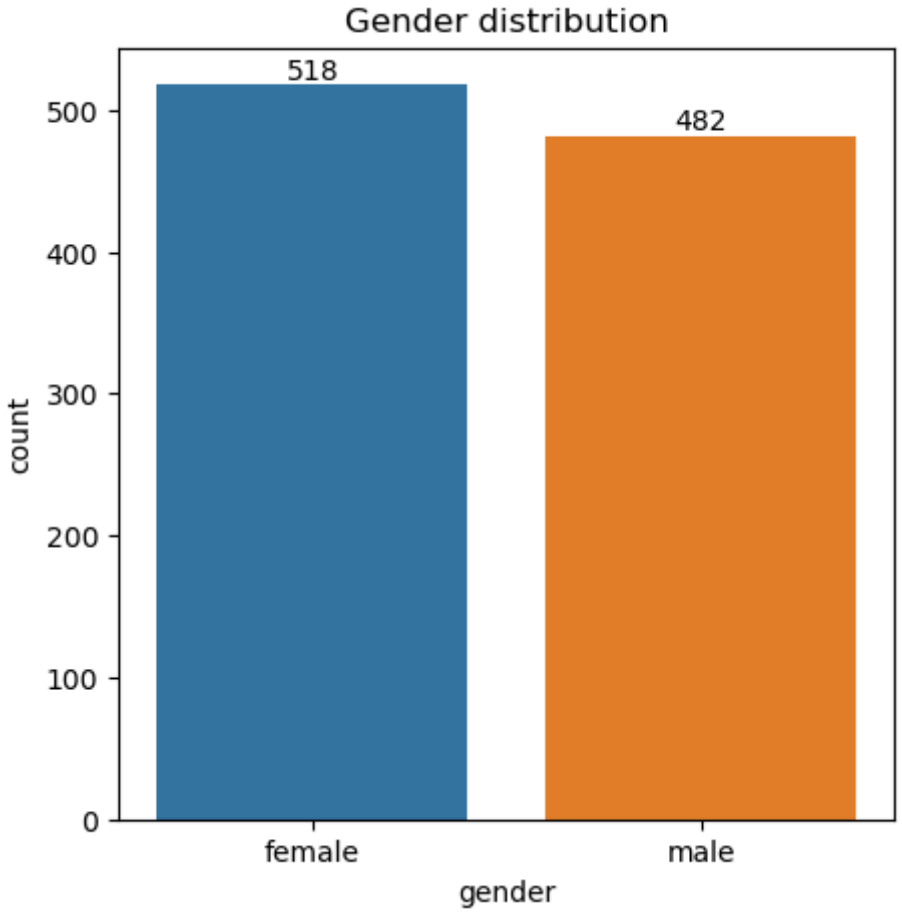
In [7]:
```
df.head()
```

Out[7]:

	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

gender Distribution

In [33]:
```
plt.figure(figsize= (5,5))
ax = sns.countplot(data = df, x = "gender")
ax.bar_label(ax.containers[0])
plt.title("Gender distribution")
plt.show()
```

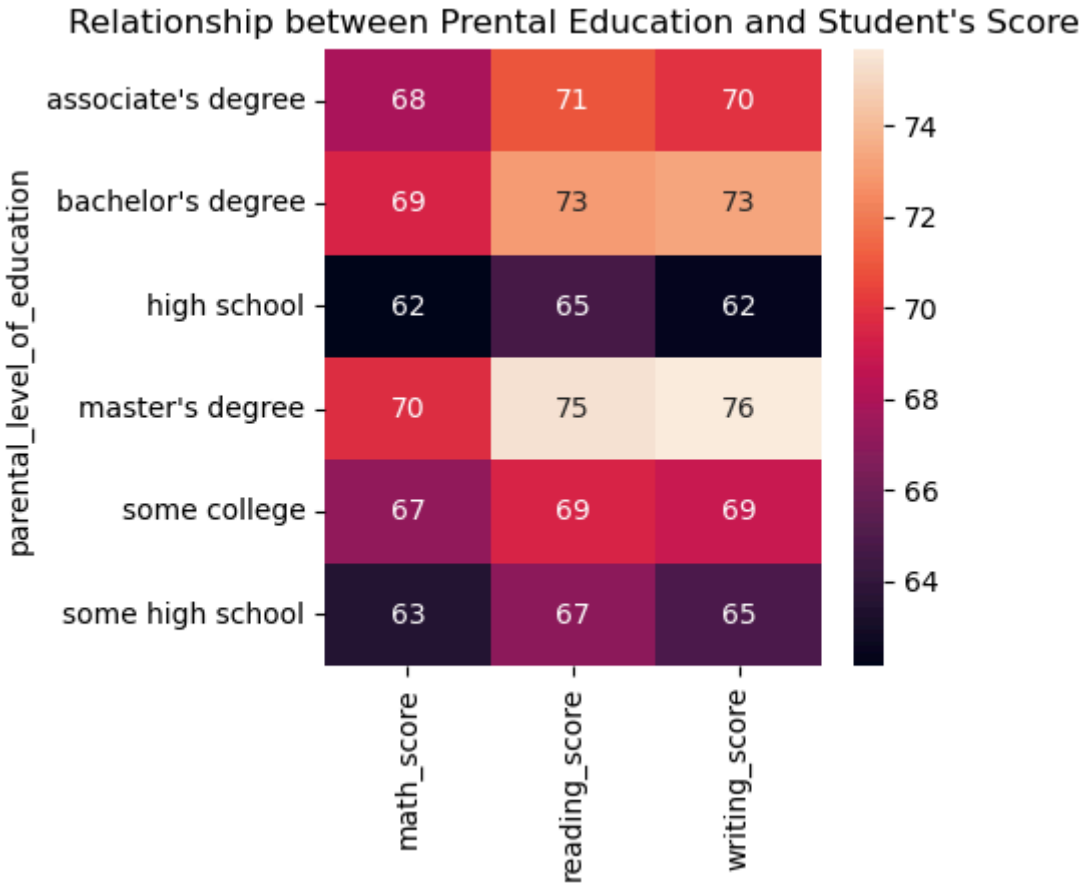


In [23]:
```
# from the above chart we have analysed that:
#the number of female in the data is more than the number of males
```

In [31]:
```
gb = df.groupby("parental_level_of_education").agg({"math_score":'mean',"reading_score":'mean',"writing_score":'mean'})
print(gb)
```

```
                             math_score  reading_score  writing_score
parental_level_of_education
associate's degree            67.882883      70.927928      69.896396
bachelor's degree             69.389831      73.000000      73.381356
high school                   62.137755      64.704082      62.448980
master's degree               69.745763      75.372881      75.677966
some college                  67.128319      69.460177      68.840708
some high school              63.497207      66.938547      64.888268
```

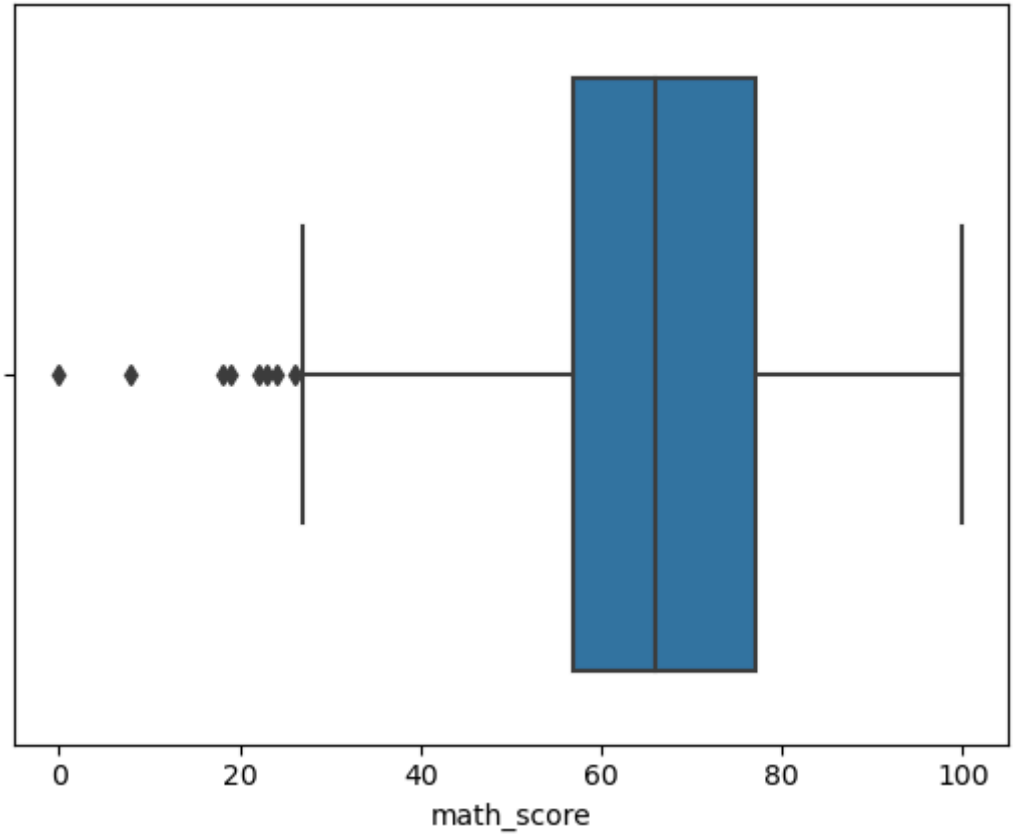
In [35]:
```
plt.figure(figsize= (4,4))
sns.heatmap(gb, annot = True)
plt.title("Relationship between Prental Education and Student's Score")
plt.show()
```



In [32]:
```
# from the above chart we have concluded that the education of the parents have a good impacts on their scores.
```

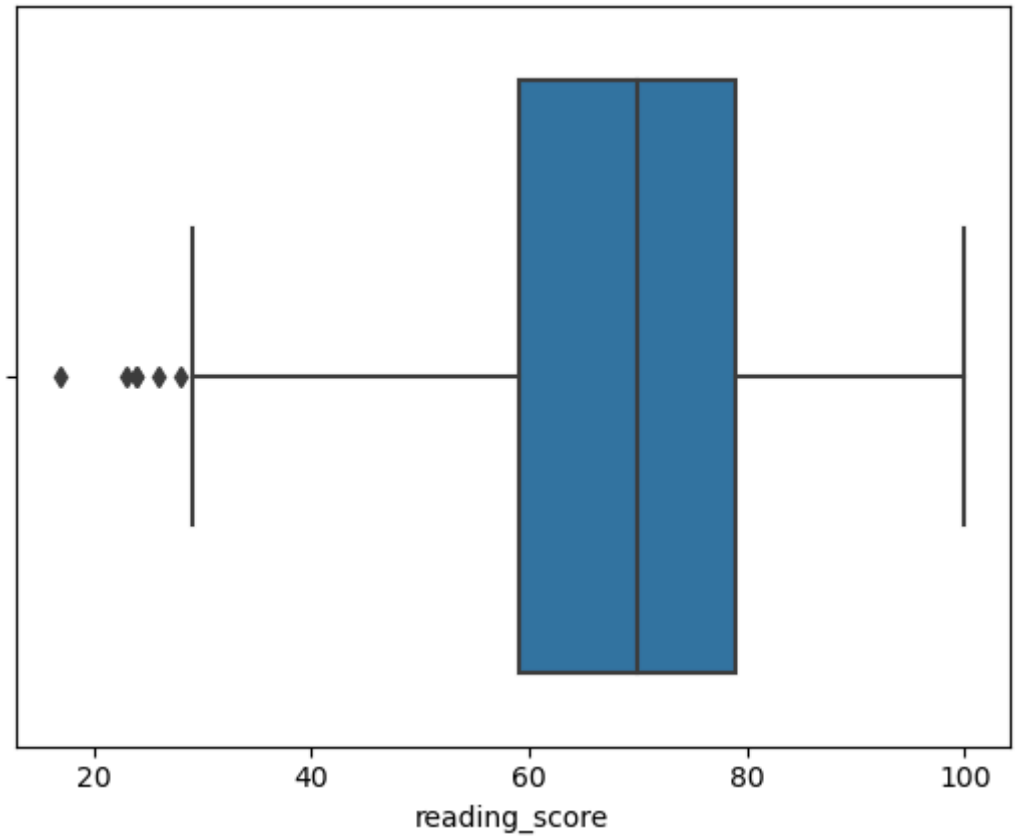
In [36]:
```
sns.boxplot(data = df, x = "math_score")
plt.show
```

Out[36]: <function matplotlib.pyplot.show(close=None, block=None)>



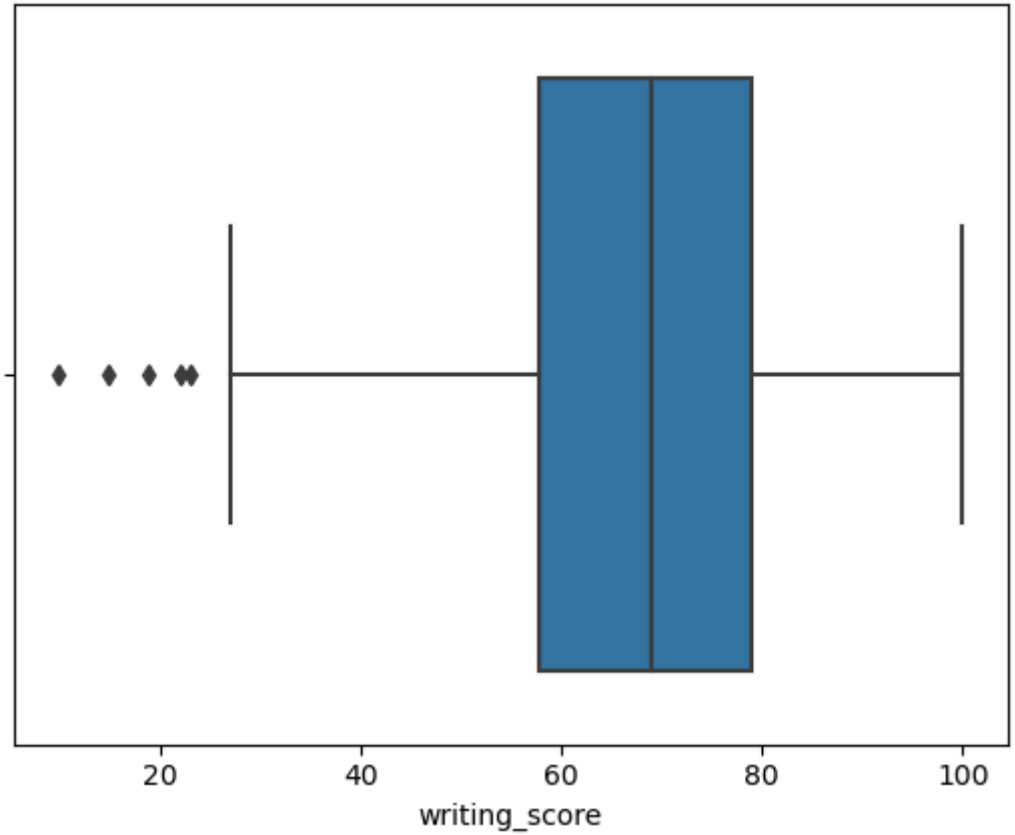
In [37]:
```
sns.boxplot(data = df, x = "reading_score")
plt.show
```

Out[37]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [38]:
```
sns.boxplot(data = df, x = "writing_score")
plt.show
```

Out[38]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



In [39]:
```
print(df["race_ethnicity"].unique())
```

```
['group B' 'group C' 'group A' 'group D' 'group E']
```

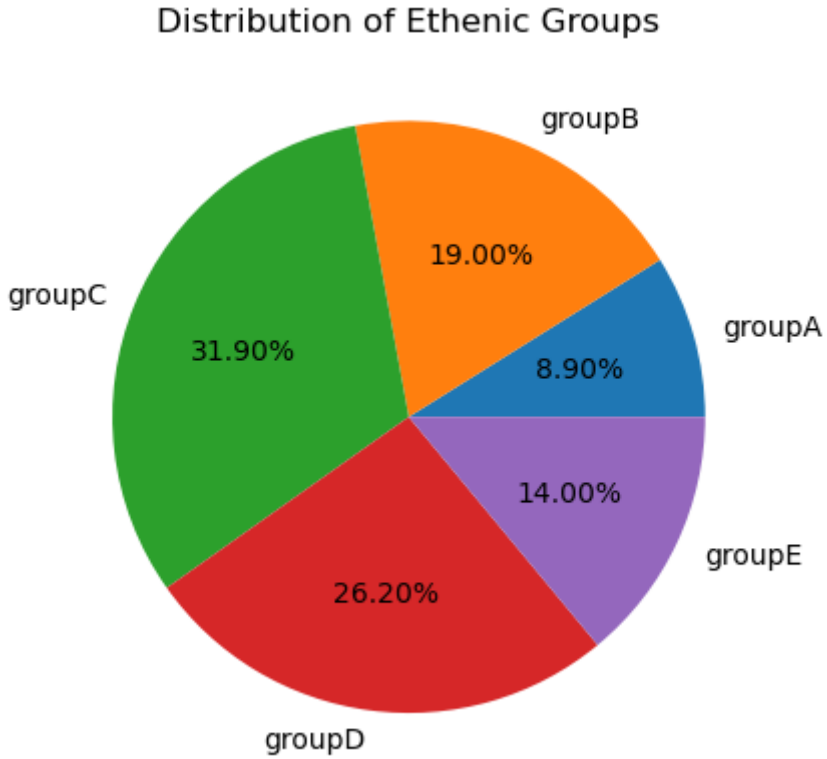
Distribution of Ethinic Groups

In [55]:
```
groupA = df.loc[(df['race_ethnicity'] == "group A")].count()
groupB = df.loc[(df['race_ethnicity'] == "group B")].count()
groupC = df.loc[(df['race_ethnicity'] == "group C")].count()
groupD = df.loc[(df['race_ethnicity'] == "group D")].count()
groupE = df.loc[(df['race_ethnicity'] == "group E")].count()

l = ["groupA", "groupB", "groupC", "groupD", "groupE"]
mlist = [groupA['race_ethnicity'], groupB['race_ethnicity'], groupC['race_ethnicity'], groupD['race_ethnicity'], group
print(mlist)
plt.pie(mlist, labels = l, autopct = "%1.2f%%")
plt.title("Distribution of Ethenic Groups")
plt.show
```

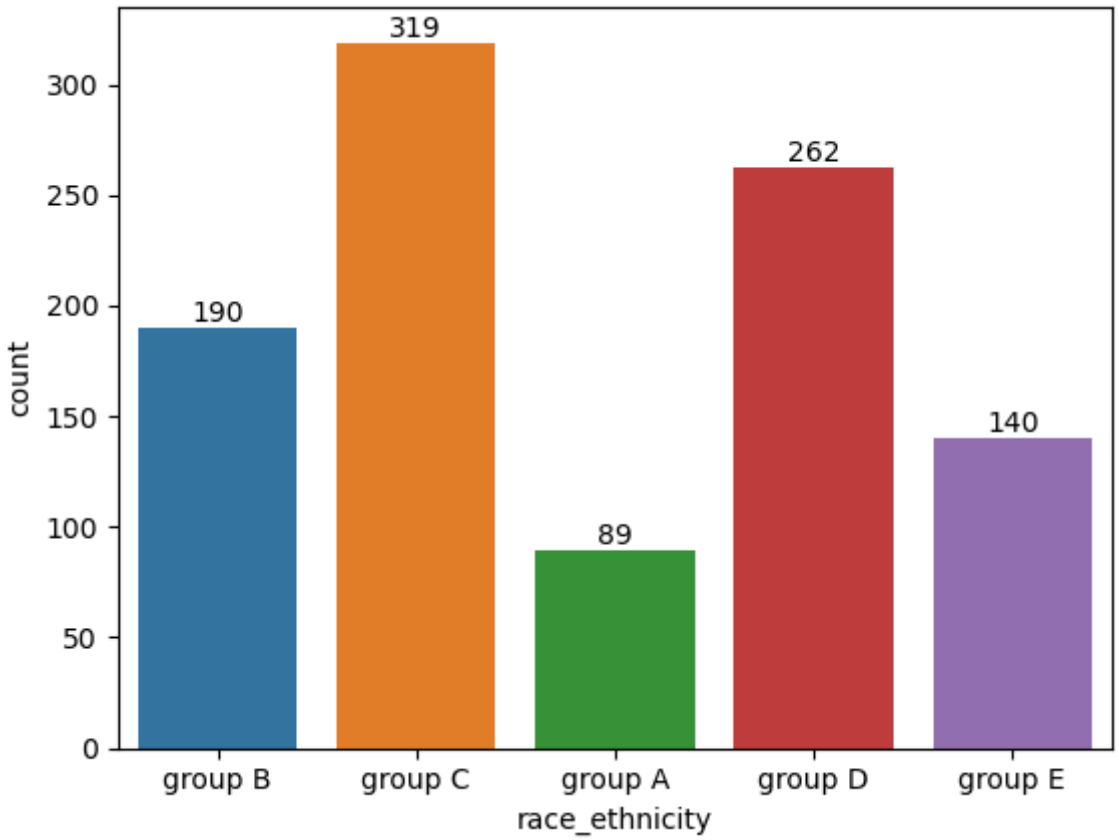
```
[89, 190, 319, 262, 140]
```

Out[55]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```


Distribution of Ethenic Groups
groupB
19.00%
groupC
31.90%
groupA
8.90%
14.00%
groupE
26.20%
groupD

In [54]:
```
ax = sns.countplot(data = df, x = 'race_ethnicity')
ax.bar_label(ax.containers[0])
```

Out[54]:
```
[Text(0, 0, '190'),
 Text(0, 0, '319'),
 Text(0, 0, '89'),
 Text(0, 0, '262'),
 Text(0, 0, '140')]
```



In [ ]: