2. Data Visualization - Box plot, scatter plot, histogram

November 15, 2022

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

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

[2]:

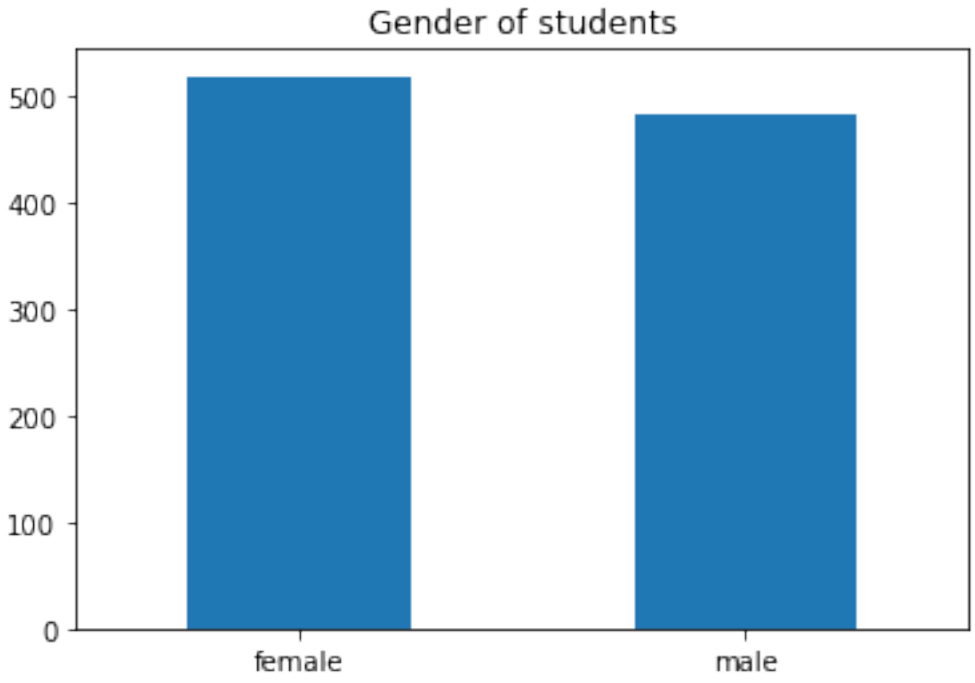
	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group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

[3]:
```
df.isnull().sum()
```

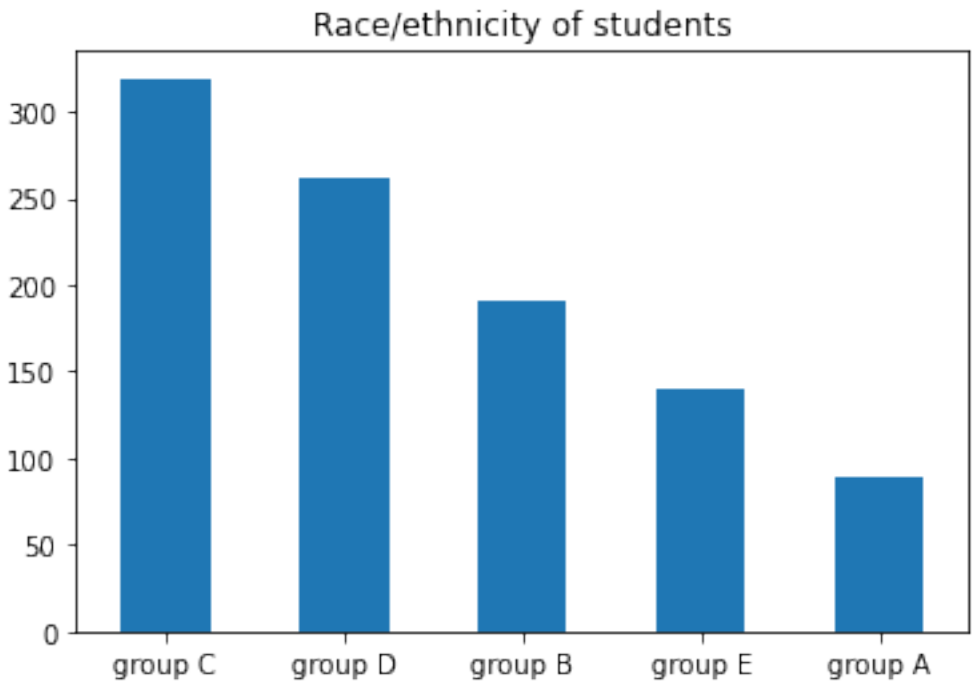
[3]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test preparation course        0
math score                     0
reading score                  0
writing score                  0
dtype: int64
```

[14]:
```
df['gender'].value_counts().plot(kind='bar', title='Gender of students')
plt.xticks(rotation=0)
```

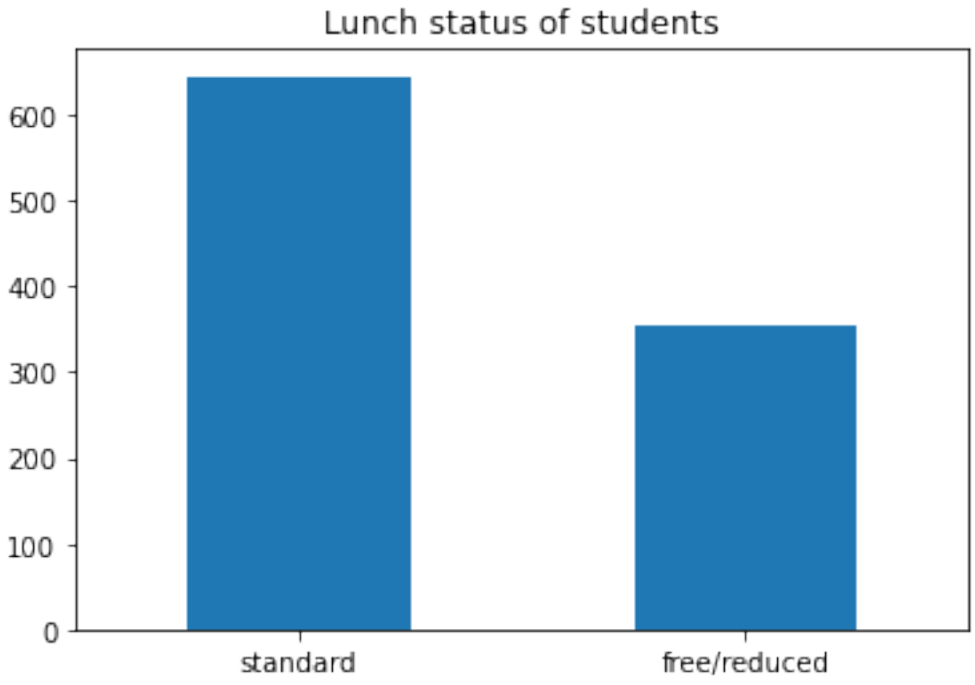
```
plt.show()
```



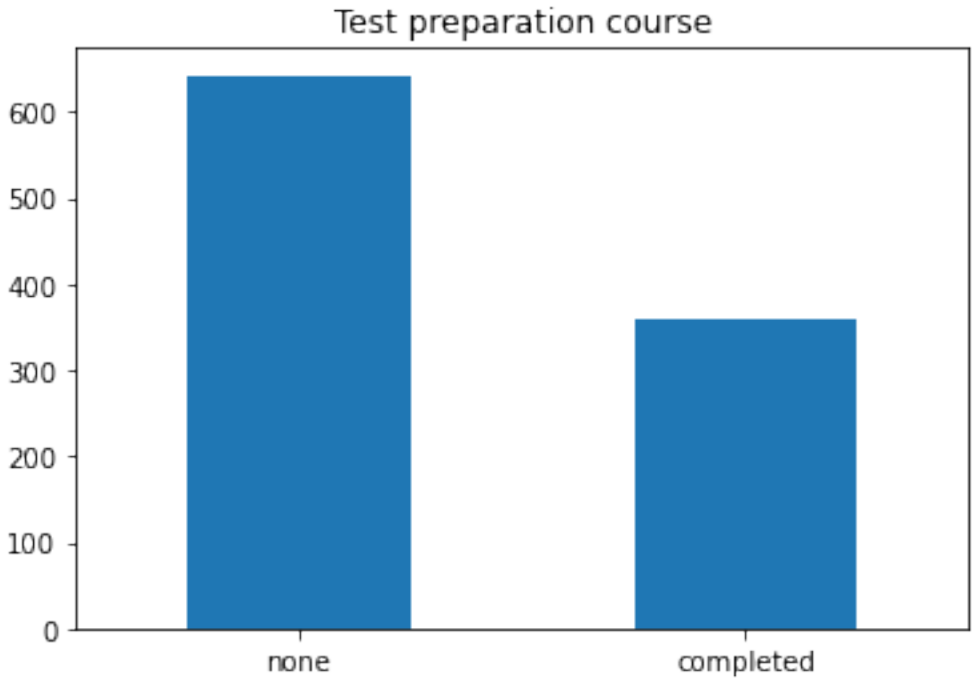
[16]:
```
df['race/ethnicity'].value_counts().plot(kind='bar', title='Race/ethnicity of students')
plt.xticks(rotation=0)
plt.show()
```



[6]:
```
df['lunch'].value_counts().plot(kind='bar', title='Lunch status of students')
plt.xticks(rotation=0)
plt.show()
```

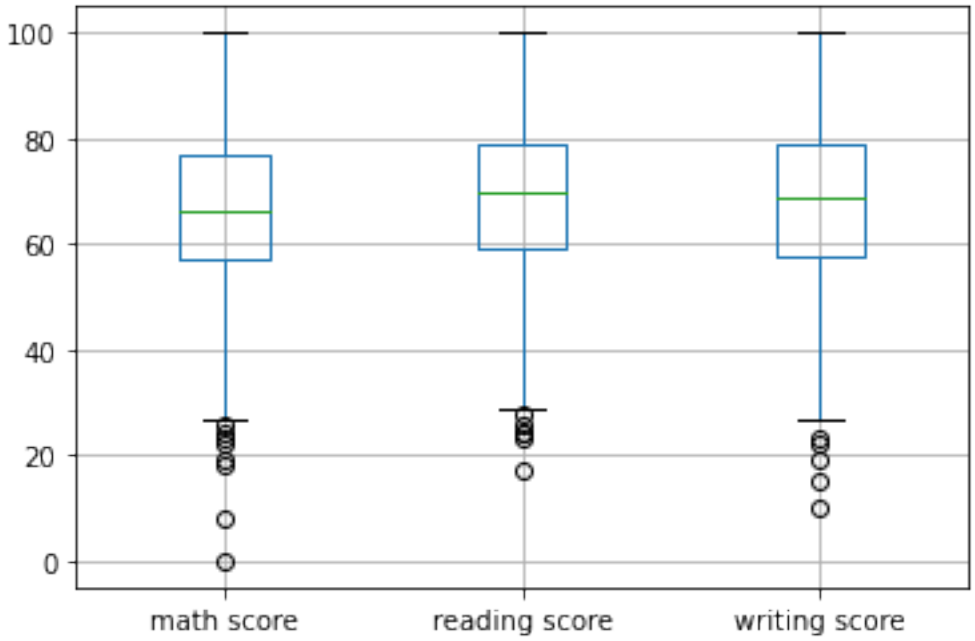


[7]:
```
df['test preparation course'].value_counts().plot(kind='bar', title='Test␣
↪preparation course')
plt.xticks(rotation=0)
plt.show()
```



[8]:
```
df.boxplot()
```

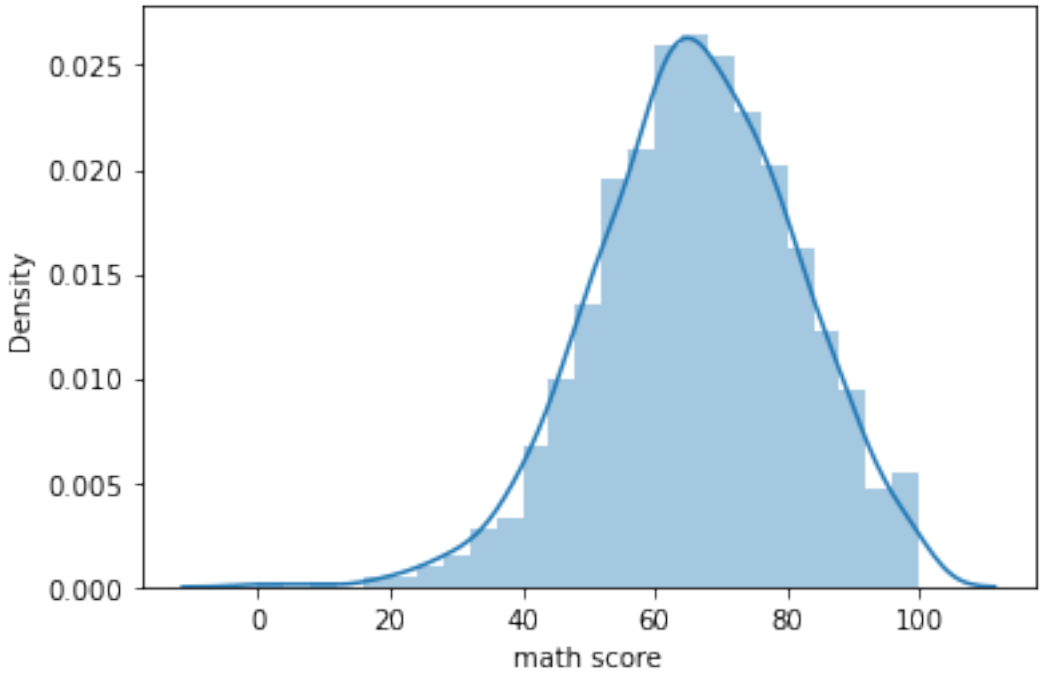
[8]:
```
<AxesSubplot:>
```



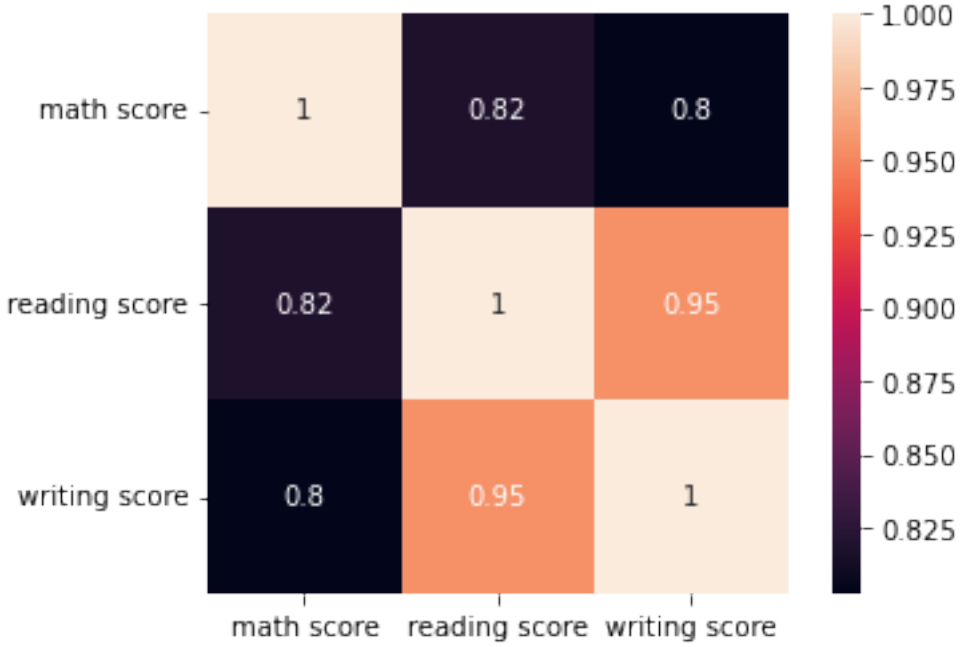
[9]:
```
sns.distplot(df['math score'])
```

```
/home/samuel-adirala/anaconda3/lib/python3.9/site-
packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar flexibility)
or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

[9]:
```
<AxesSubplot:xlabel='math score', ylabel='Density'>
```



[10]:
```
corr = df.corr()
sns.heatmap(corr, annot=True, square=True)
plt.yticks(rotation=0)
plt.show()
```



[11]:
```
sns.scatterplot(x='math score', y='writing score', hue='gender', data=df)
```

[11]: <AxesSubplot:xlabel='math score', ylabel='writing score'>