

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv('https://github.com/an-2-an/data/raw/master/StudentsPerformance.csv')
df['total score'] = df['math score'] + df['reading score'] + df['writing score']
df
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	total score
0	female	group B	bachelor's degree	standard	none	72	72	74	218
1	female	group C	some college	standard	completed	69	90	88	247
2	female	group B	master's degree	standard	none	90	95	93	278
3	male	group A	associate's degree	free/reduced	none	47	57	44	148
4	male	group C	some college	standard	none	76	78	75	229
...
995	female	group E	master's degree	standard	completed	88	99	95	282
996	male	group C	high school	free/reduced	none	62	55	55	172
997	female	group C	high school	free/reduced	completed	59	71	65	195
998	female	group D	some college	standard	completed	68	78	77	223
999	female	group D	some college	free/reduced	none	77	86	86	249

1000 rows × 9 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 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
8   total score                                1000 non-null   int64
dtypes: int64(4), object(5)
memory usage: 70.4+ KB
```

```
df.describe()
```

	math score	reading score	writing score	total score
count	1000.00000	1000.000000	1000.000000	1000.000000
mean	66.08900	69.169000	68.054000	203.312000
std	15.16308	14.600192	15.195657	42.771978
min	0.00000	17.000000	10.000000	27.000000
25%	57.00000	59.000000	57.750000	175.000000
50%	66.00000	70.000000	69.000000	205.000000
75%	77.00000	79.000000	79.000000	233.000000
max	100.00000	100.000000	100.000000	300.000000

```
df.sort_values(by='total score', ascending=False)[:5]
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	total score
916	male	group E	bachelor's degree	standard	completed	100	100	100	300
458	female	group E	bachelor's degree	standard	none	100	100	100	300
962	female	group E	associate's degree	standard	none	100	100	100	300
114	female	group E	bachelor's degree	standard	completed	99	100	100	299
179	female	group D	some high school	standard	completed	97	100	100	297

```
df.columns
```

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

```
df['parental level of education'].unique()
```

```
array(["bachelor's degree", 'some college', "master's degree",
      "associate's degree", 'high school', 'some high school'],
      dtype=object)
```

```
df['gender'].value_counts()
```

```
female    518
male      482
Name: gender, dtype: int64
```

```
df['gender'].value_counts(normalize=True)
```

```
female    0.518
male      0.482
Name: gender, dtype: float64
```

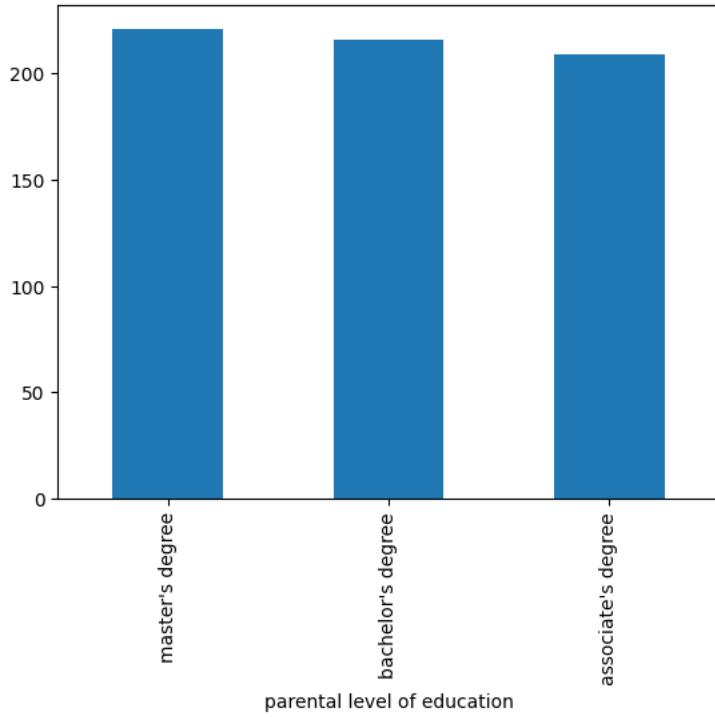
```
df.groupby('gender')['math score'].mean()
```

```
gender
female    63.633205
male      68.728216
Name: math score, dtype: float64
```

```
df.groupby('parental level of education')['total score'].mean().sort_values(ascending=False)[:3]
```

```
parental level of education
master's degree    220.796610
bachelor's degree  215.771186
associate's degree  208.707207
Name: total score, dtype: float64
```

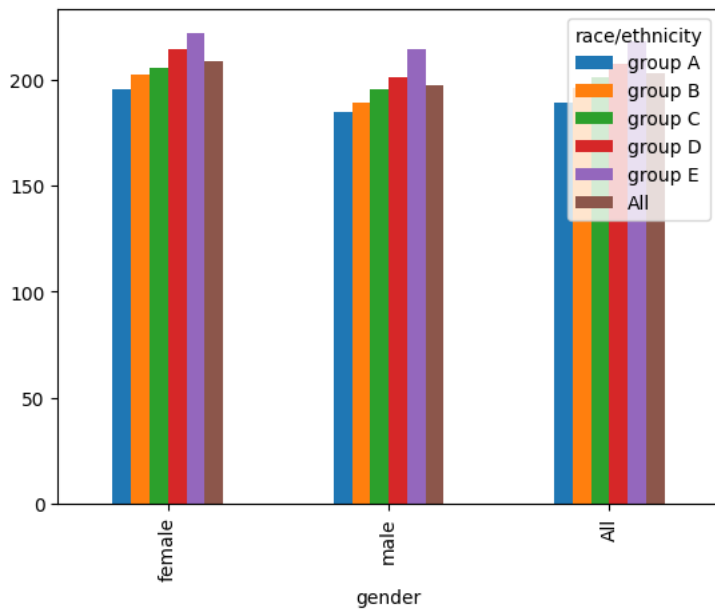
```
df.groupby('parental level of education')['total score'].mean().sort_values(ascending=False)[:3].plot(kind='bar');
```



```
df.pivot_table(index='gender', columns='race/ethnicity', values='total score', aggfunc='mean', margins=True)
```

race/ethnicity	group A	group B	group C	group D	group E	All
gender						
female	195.388889	202.528846	205.755556	214.317829	222.188406	208.708494
male	184.622642	189.000000	195.748201	200.962406	214.436620	197.512448
All	188.977528	196.405263	201.394984	207.538168	218.257143	203.312000

```
df.pivot_table(index='gender', columns='race/ethnicity', values='total score', aggfunc='mean', margins=True).plot(kind='bar');
```



```
df[df['parental level of education'].str.endswith('degree')]
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	total score
0	female	group B	bachelor's degree	standard	none	72	72	74	218
2	female	group B	master's degree	standard	none	90	95	93	278
3	male	group A	associate's degree	free/reduced	none	47	57	44	148
5	female	group B	associate's degree	standard	none	71	83	78	232
10	male	group C	associate's degree	standard	none	58	54	52	164
...
979	female	group C	associate's degree	standard	none	91	95	94	280
986	female	group C	associate's degree	standard	none	40	59	51	150
992	female	group D	associate's degree	free/reduced	none	55	76	76	207
993	female	group D	bachelor's degree	free/reduced	none	62	72	74	208
995	female	group E	master's degree	standard	completed	88	99	95	282

399 rows x 9 columns

```
df[df['parental level of education'].str[0] == 's']
```

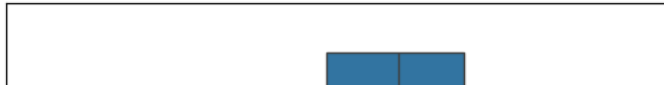
	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	total score
1	female	group C	some college	standard	completed	69	90	88	247
4	male	group C	some college	standard	none	76	78	75	229
6	female	group B	some college	standard	completed	88	95	92	275
7	male	group B	some college	free/reduced	none	40	43	39	122
13	male	group A	some college	standard	completed	78	72	70	220
...
988	female	group A	some high school	free/reduced	none	44	45	45	134
989	female	group D	some college	free/reduced	completed	67	86	83	236
991	female	group B	some high school	standard	completed	65	82	78	225
998	female	group D	some college	standard	completed	68	78	77	223
999	female	group D	some college	free/reduced	none	77	86	86	249

405 rows x 9 columns

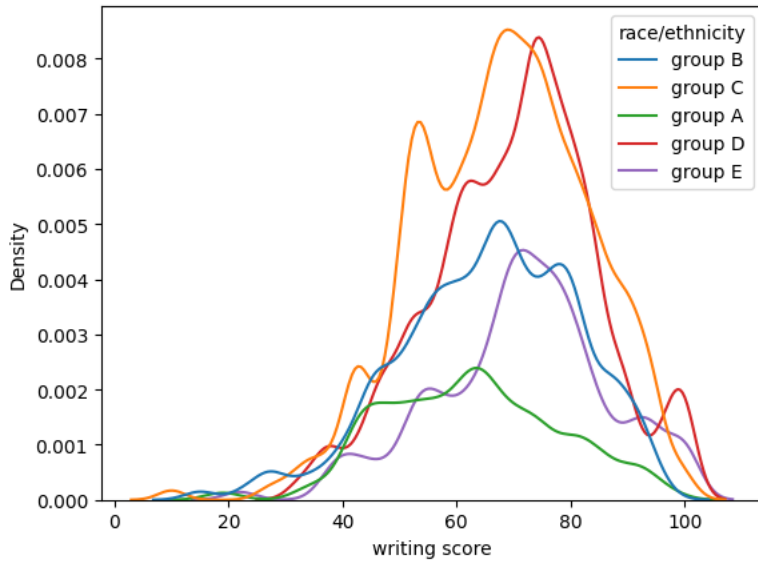
```
df.columns
```

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

```
sns.boxplot(data=df, x='reading score', hue='test preparation course');
```



```
sns.kdeplot(data=df, x='writing score', hue='race/ethnicity', bw_adjust=.5);
```



```
sns.kdeplot(data=df, x='writing score', y='reading score', fill=True);
```

