```
import pandas as pd
df = pd.read_csv('./stud.csv')
Df.shape
df.head()
```

	Roll.no	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	1	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0
1	2	female	group C	some college	standard	completed	69.0	90.0	NaN
2	3	female	NaN	master's degree	standard	none	NaN	95.0	93.0
3	4	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0
4	5	male	group C	some college	NaN	NaN	76.0	78.0	NaN

df.tail()

	Roll.no	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
5	6	female	group B	NaN	standard	none	71.0	83.0	78.0
6	7	female	group B	some college	standard	completed	88.0	NaN	92.0
7	8	male	group B	some college	free/reduced	none	NaN	43.0	39.0
8	9	male	group D	high school	NaN	completed	64.0	64.0	67.0
9	10	female	group B	high school	free/reduced	none	38.0	60.0	NaN

df.count()

```
Roll.no                        10
gender                         10
race/ethnicity                  9
parental level of education     9
lunch                           8
test preparation course         9
math score                      8
reading score                   9
writing score                   7
dtype: int64
```

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 9 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Roll.no                      10 non-null     int64
 1   gender                       10 non-null     object
 2   race/ethnicity               9 non-null      object
 3   parental level of education  9 non-null      object
 4   lunch                        8 non-null      object
 5   test preparation course      9 non-null      object
 6   math score                   8 non-null      float64
 7   reading score                9 non-null      float64
 8   writing score                7 non-null      float64
dtypes: float64(3), int64(1), object(5)
memory usage: 852.0+ bytes
```

df.isnull()

	Roll.no	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	True
2	False	False	True	False	False	False	True	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	True	True	False	False	True
5	False	False	False	True	False	False	False	False	False
6	False	False	False	False	False	False	False	True	False
7	False	False	False	False	False	False	True	False	False
8	False	False	False	False	True	False	False	False	False
9	False	False	False	False	False	False	False	False	True

df.isnull().sum()

```
Roll.no                      0
gender                       0
race/ethnicity               1
parental level of education  1
lunch                        2
test preparation course      1
math score                   2
reading score                1
writing score                3
dtype: int64
```

df.dropna()

	Roll.no	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	1	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0
3	4	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0

df.fillna(0)

	Roll.no	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	1	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0
1	2	female	group C	some college	standard	completed	69.0	90.0	0.0
2	3	female	0	master's degree	standard	none	0.0	95.0	93.0
3	4	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0
4	5	male	group C	some college	0	0	76.0	78.0	0.0
5	6	female	group B	0	standard	none	71.0	83.0	78.0
6	7	female	group B	some college	standard	completed	88.0	0.0	92.0
7	8	male	group B	some college	free/reduced	none	0.0	43.0	39.0
8	9	male	group D	high school	0	completed	64.0	64.0	67.0
9	10	female	group B	high school	free/reduced	none	38.0	60.0	0.0

df['lunch'].fillna('standard')

```
0        standard
1        standard
2        standard
3    free/reduced
4        standard
5        standard
6        standard
7    free/reduced
8        standard
9    free/reduced
Name: lunch, dtype: object
```

df['math score'].fillna(df['math score'].mean())

[17]:

```
0    72.000
1    69.000
2    65.625
3    47.000
4    76.000
5    71.000
6    88.000
7    65.625
8    64.000
9    38.000
Name: math score, dtype: float64
```

df['reading score'].fillna(df['reading score'].median())

```
0    72.0
1    90.0
2    95.0
3    57.0
4    78.0
5    83.0
6    72.0
7    43.0
8    64.0
9    60.0
Name: reading score, dtype: float64
```

df['lunch'].value_counts()

```
lunch
standard        5
free/reduced    3
Name: count, dtype: int64
```

df['lunch'].fillna(df['lunch'].mode()[0])

```
0        standard
1        standard
2        standard
3    free/reduced
4        standard
5        standard
6        standard
7    free/reduced
8        standard
9    free/reduced
Name: lunch, dtype: object
```

df.fillna(method='backfill')

	Roll.no	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	1	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0
1	2	female	group C	some college	standard	completed	69.0	90.0	93.0
2	3	female	group A	master's degree	standard	none	47.0	95.0	93.0
3	4	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0
4	5	male	group C	some college	standard	none	76.0	78.0	78.0
5	6	female	group B	some college	standard	none	71.0	83.0	78.0
6	7	female	group B	some college	standard	completed	88.0	43.0	92.0
7	8	male	group B	some college	free/reduced	none	64.0	43.0	39.0
8	9	male	group D	high school	free/reduced	completed	64.0	64.0	67.0
9	10	female	group B	high school	free/reduced	none	38.0	60.0	NaN

df.fillna(method='pad')

	Roll.no	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	1	female	group B	bachelor's degree	standard	none	72.0	72.0	74.0
1	2	female	group C	some college	standard	completed	69.0	90.0	74.0
2	3	female	group C	master's degree	standard	none	69.0	95.0	93.0
3	4	male	group A	associate's degree	free/reduced	none	47.0	57.0	44.0
4	5	male	group C	some college	free/reduced	none	76.0	78.0	44.0
5	6	female	group B	some college	standard	none	71.0	83.0	78.0
6	7	female	group B	some college	standard	completed	88.0	83.0	92.0
7	8	male	group B	some college	free/reduced	none	88.0	43.0	39.0
8	9	male	group D	high school	free/reduced	completed	64.0	64.0	67.0
9	10	female	group B	high school	free/reduced	none	38.0	60.0	67.0

df.describe()

	Roll.no	math score	reading score	writing score
count	10.00000	8.000000	9.000000	7.000000
mean	5.50000	65.625000	71.333333	69.571429
std	3.02765	16.044024	16.881943	21.360845
min	1.00000	38.000000	43.000000	39.000000
25%	3.25000	59.750000	60.000000	55.500000
50%	5.50000	70.000000	72.000000	74.000000
75%	7.75000	73.000000	83.000000	85.000000
max	10.00000	88.000000	95.000000	93.000000

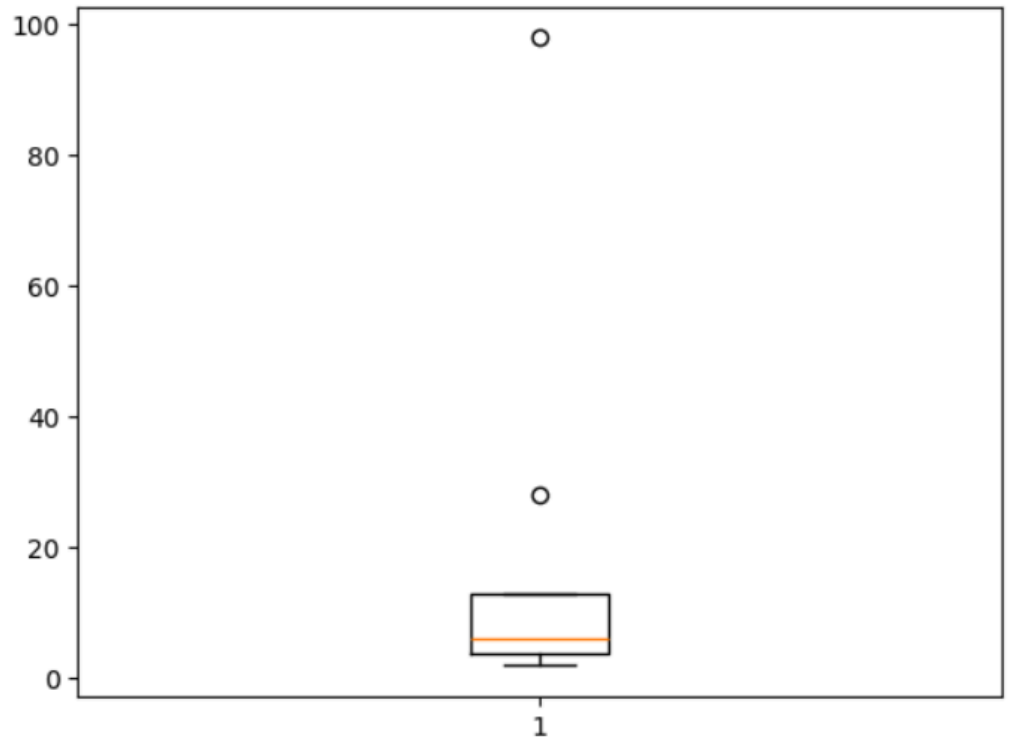
```
import numpy as np
x = np.array([5,4,3,2,7,8,98,28])
np.mean(x)
```

19.375

```
np.median(x)
```

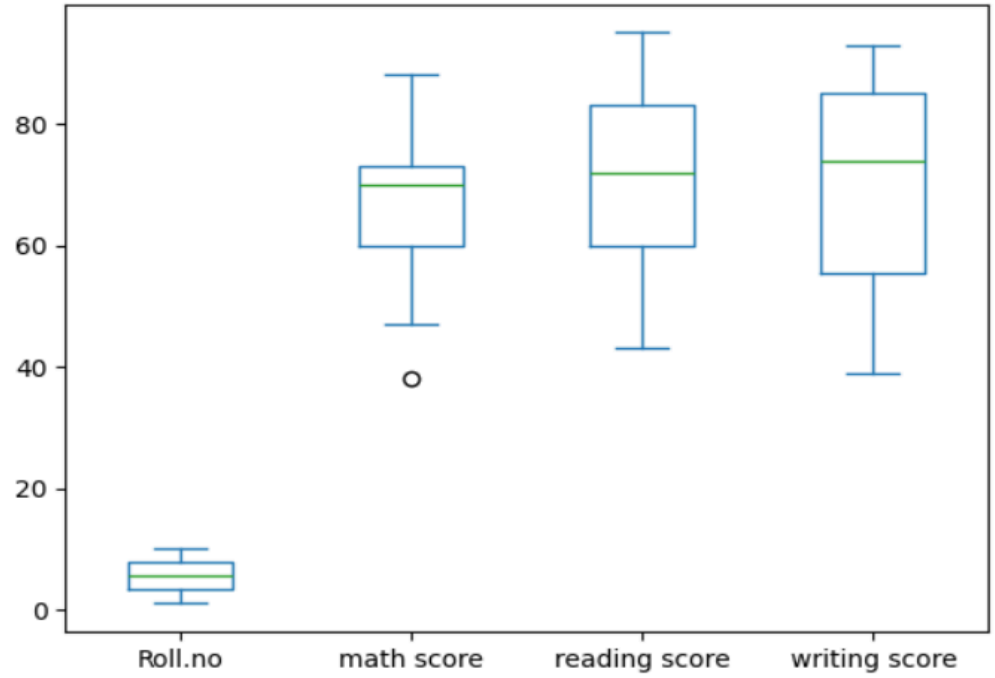
6.0

```
import matplotlib.pyplot as plt
plt.boxplot(x);
```



df.plot.box()

<Axes: >



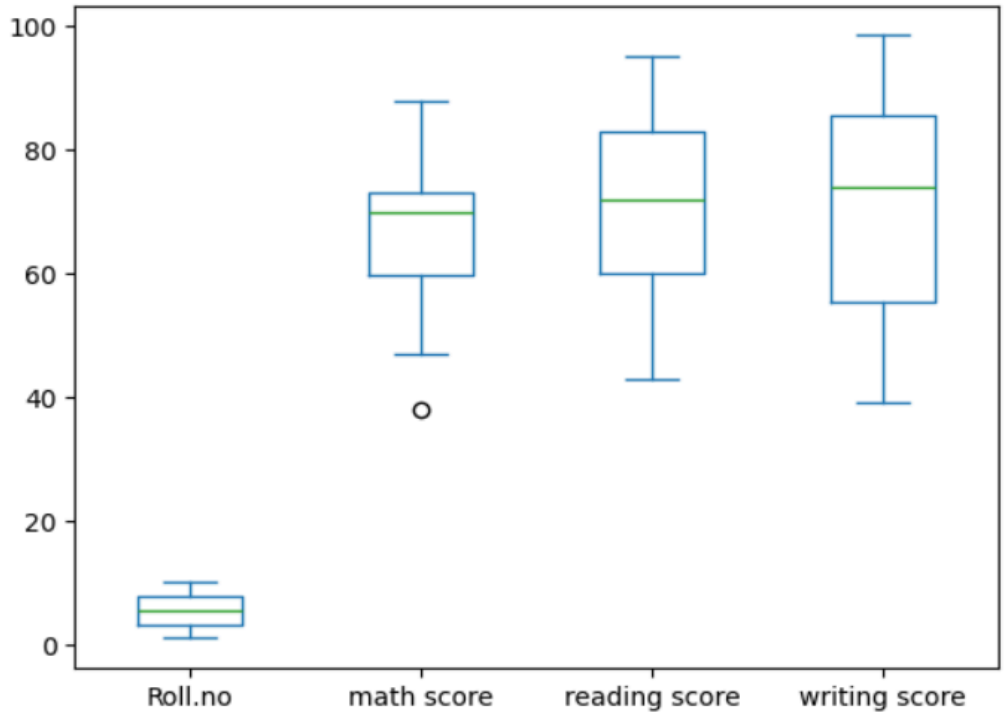
df.loc[6,'writing score']

92.0

df.loc[6,'writing score']=98.45

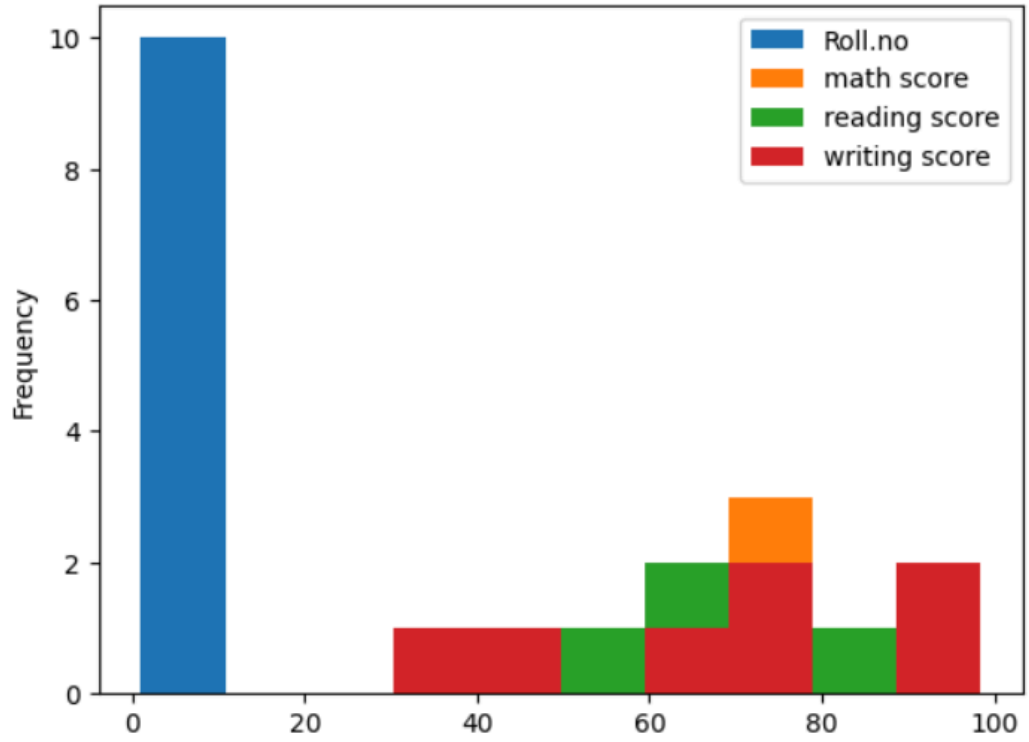
df.plot.box()

<Axes: >



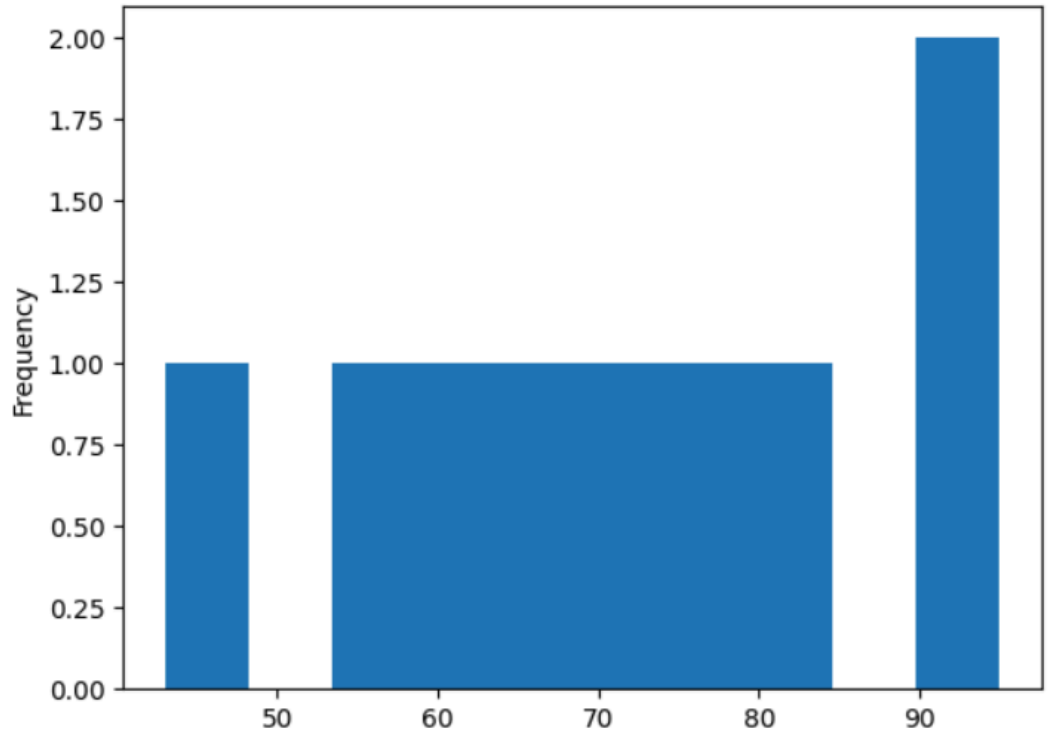
df.plot.hist()

```
<Axes: ylabel='Frequency'>
```



df['reading score'].plot.hist()

```
<Axes: ylabel='Frequency'>
```



```
x= df[['Roll.no','math score']]
x.describe()
```

	Roll.no	math score
count	10.00000	8.000000
mean	5.50000	65.625000
std	3.02765	16.044024
min	1.00000	38.000000
25%	3.25000	59.750000
50%	5.50000	70.000000
75%	7.75000	73.000000
max	10.00000	88.000000

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
x_scaled = scaler.fit_transform(x)

pd.DataFrame(x_scaled).describe()
```

	0	1
count	10.000000	8.00000
mean	0.500000	0.55250
std	0.336406	0.32088
min	0.000000	0.00000
25%	0.250000	0.43500
50%	0.500000	0.64000
75%	0.750000	0.70000
max	1.000000	1.00000