

```
In [41]: import numpy as np
import pandas as pd
```

```
In [42]: x=pd.read_csv(r"C:\Users\user\Downloads\4_drug200 - 4_drug200.csv")
x
```

Out[42]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

```
In [43]: x.dtypes
```

```
Out[43]: Age          int64
Sex            object
BP             object
Cholesterol    object
Na_to_K        float64
Drug           object
dtype: object
```

```
In [44]: x.head()
```

Out[44]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY

In [45]: `x.tail()`

Out[45]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

In [46]: `x.columns`

Out[46]: Index(['Age', 'Sex', 'BP', 'Cholesterol', 'Na\_to\_K', 'Drug'], dtype='object')

In [47]: `x.index`

Out[47]: RangeIndex(start=0, stop=200, step=1)

In [48]: `x.describe()`

Out[48]:

	Age	Na_to_K
count	200.000000	200.000000
mean	44.315000	16.084485
std	16.544315	7.223956
min	15.000000	6.269000
25%	31.000000	10.445500
50%	45.000000	13.936500
75%	58.000000	19.380000
max	74.000000	38.247000

In [50]: `x["Age"]`

Out[50]:

```
0      23
1      47
2      47
3      28
4      61
..
195    56
196    16
197    52
198    23
199    40
Name: Age, Length: 200, dtype: int64
```

```
In [51]: x[0:2]
```

```
Out[51]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC

```
In [52]: x.loc[0:2]
```

```
Out[52]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC

```
In [53]: x.iloc[0:2]
```

```
Out[53]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC

```
In [54]: x.loc["Age":"Na_to_K"]
```

```
Out[54]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
--	-----	-----	----	-------------	---------	------

```
In [56]: x[x["Na_to_K"]<=2]
```

```
Out[56]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
--	-----	-----	----	-------------	---------	------

In [57]: `x.fillna(value=5)`

Out[57]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

In [58]: `x.dropna()`

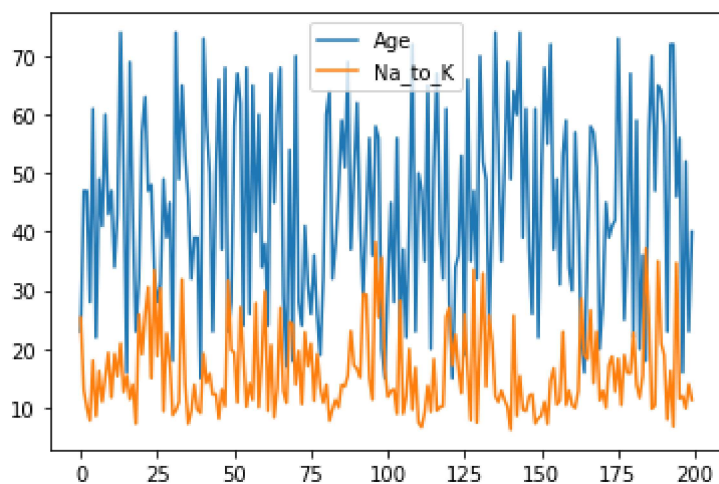
Out[58]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

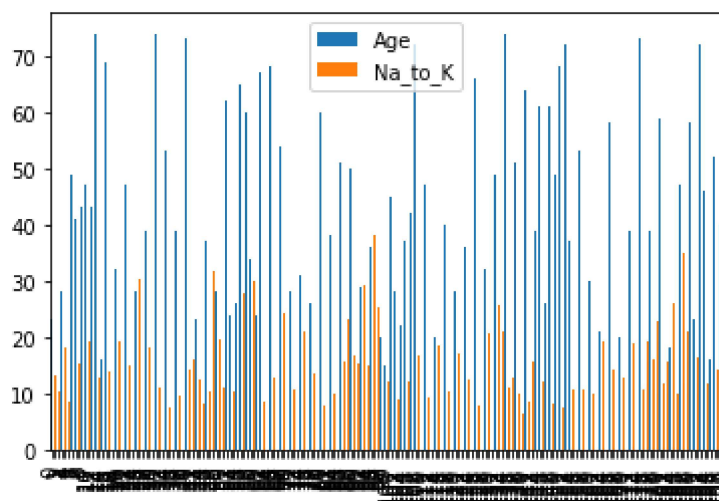
```
In [59]: x.plot.line()
```

```
Out[59]: <AxesSubplot:>
```



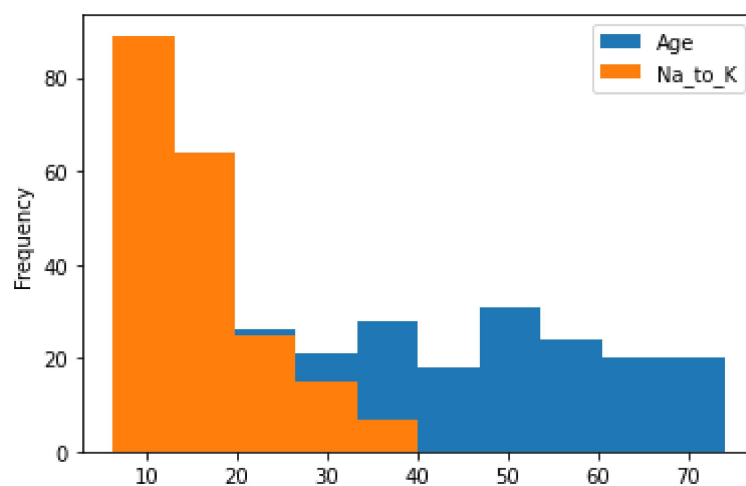
```
In [60]: x.plot.bar()
```

```
Out[60]: <AxesSubplot:>
```

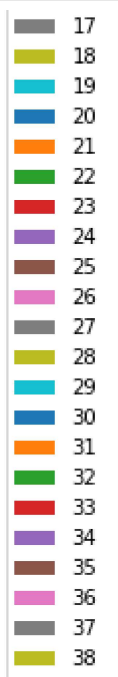


```
In [61]: x.plot.hist()
```

```
Out[61]: <AxesSubplot:ylabel='Frequency'>
```

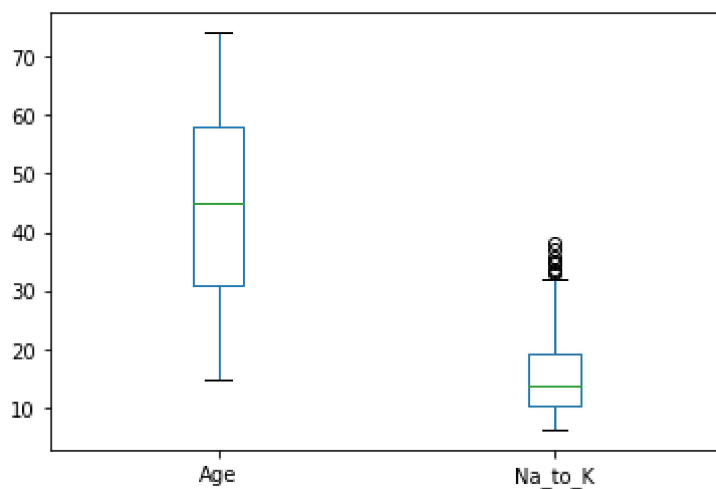


```
In [63]: x.plot.pie(y='Na_to_K')
```



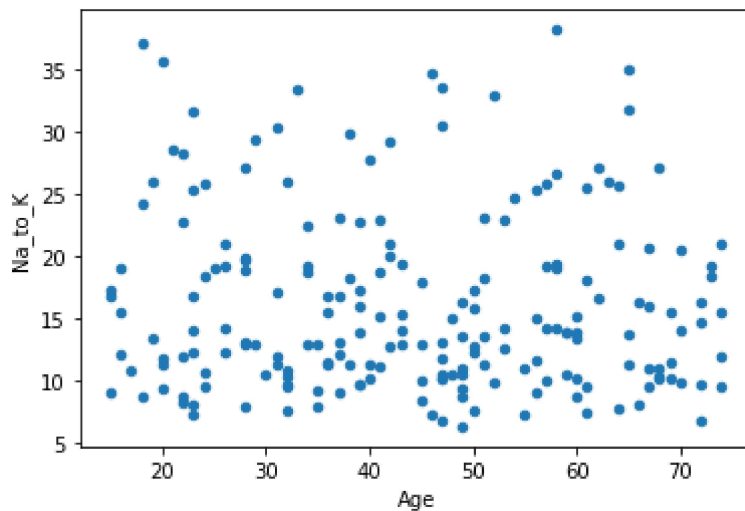
```
In [64]: x.plot.box()
```

```
Out[64]: <AxesSubplot:>
```



```
In [65]: x.plot.scatter(x='Age',y='Na_to_K')
```

```
Out[65]: <AxesSubplot:xlabel='Age', ylabel='Na_to_K'>
```



```
In [66]: x.mean()
```

```
Out[66]: Age      44.315000  
Na_to_K    16.084485  
dtype: float64
```

```
In [67]: x.median()
```

```
Out[67]: Age      45.0000  
Na_to_K    13.9365  
dtype: float64
```





```
In [71]: x.cumsum()
```

Out[71]:

	Age	Sex
0	23	F
1	70	FM
2	117	FMM
3	145	FMMF
4	206	FMMFF
...	...	...
195	8732	FMMFFFFMMMFFMFFFMMFMMFFFMFFMFMMFMMMMFMFFMMFF... HIGHLOWLOWNOR
196	8748	FMMFFFFMMMFFMFFFMMFMMFFFMFFMFMMFMMMMFMFFMMFF... HIGHLOWLOWNOR
197	8800	FMMFFFFMMMFFMFFFMMFMMFFFMFFMFMMFMMMMFMFFMMFF... HIGHLOWLOWNOR
198	8823	FMMFFFFMMMFFMFFFMMFMMFFFMFFMFMMFMMMMFMFFMMFF... HIGHLOWLOWNOR
199	8863	FMMFFFFMMMFFMFFFMMFMMFFFMFFMFMMFMMMMFMFFMMFF... HIGHLOWLOWNOR

200 rows × 6 columns

```
In [72]: x.count()
```

Out[72]:

Age	200
Sex	200
BP	200
Cholesterol	200
Na_to_K	200
Drug	200

dtype: int64

```
In [73]: x.min()
```

Out[73]:

Age	15
Sex	F
BP	HIGH
Cholesterol	HIGH
Na_to_K	6.269
Drug	drugA

dtype: object

```
In [74]: x.max()
```

Out[74]:

Age	74
Sex	M
BP	NORMAL
Cholesterol	NORMAL
Na_to_K	38.247
Drug	drugY

dtype: object

```
In [75]: from numpy import cov
d1=x['Age']
d2=x['Na_to_K']
d1
d2
```

```
Out[75]: 0      25.355
1      13.093
2      10.114
3       7.798
4      18.043
...
195    11.567
196    12.006
197     9.894
198    14.020
199    11.349
Name: Na_to_K, Length: 200, dtype: float64
```

```
In [76]: cov(d1,d2)
```

```
Out[76]: array([[273.71434673, -7.54375153],
 [ -7.54375153,  52.18553348]])
```

```
In [77]: from scipy.stats import pearsonr
print(pearsonr(d1,d2))

(-0.06311949726772592, 0.3745756399034559)
```

```
In [78]: from scipy.stats import spearmanr
print(spearmanr(d1,d2))

SpearmanrResult(correlation=-0.047273882688479915, pvalue=0.5062200581387418)
```

```
In [ ]:
```