

# Import libraries

```
In [1]: import numpy as np
import pandas as pd
```

```
In [2]: data = pd.read_csv(r"C:\Users\user\Downloads\4_drug200.csv")
data
```

Out[2]:

	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 [3]: print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Age             200 non-null   int64
1   Sex             200 non-null   object
2   BP              200 non-null   object
3   Cholesterol     200 non-null   object
4   Na_to_K         200 non-null   float64
5   Drug            200 non-null   object
dtypes: float64(1), int64(1), object(4)
memory usage: 9.5+ KB
None
```

In [4]: `print(data.describe())`

	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 [5]: `data.size`

Out[5]: 1200

In [6]: `data.shape`

Out[6]: (200, 6)

In [7]: `data1 = data.drop(["Sex", "BP", "Cholesterol", "Drug"],axis=1)`  
`data1`

Out[7]:

	Age	Na_to_K
0	23	25.355
1	47	13.093
2	47	10.114
3	28	7.798
4	61	18.043
...	...	...
195	56	11.567
196	16	12.006
197	52	9.894
198	23	14.020
199	40	11.349

200 rows × 2 columns

In [8]: `print(data1.mean())`

```
Age      44.315000
Na_to_K  16.084485
dtype: float64
```

```
In [9]: print(data.median())
```

```
Age      45.0000  
Na_to_K  13.9365  
dtype: float64
```

```
In [10]: print(data1.mode())
```

```
   Age  Na_to_K  
0  47.0   12.006  
1   NaN   18.295
```

```
In [11]: print(data1.isna().sum())
```

```
Age      0  
Na_to_K  0  
dtype: int64
```

```
In [12]: print(data1.isna().sum())
```

```
Age      0  
Na_to_K  0  
dtype: int64
```

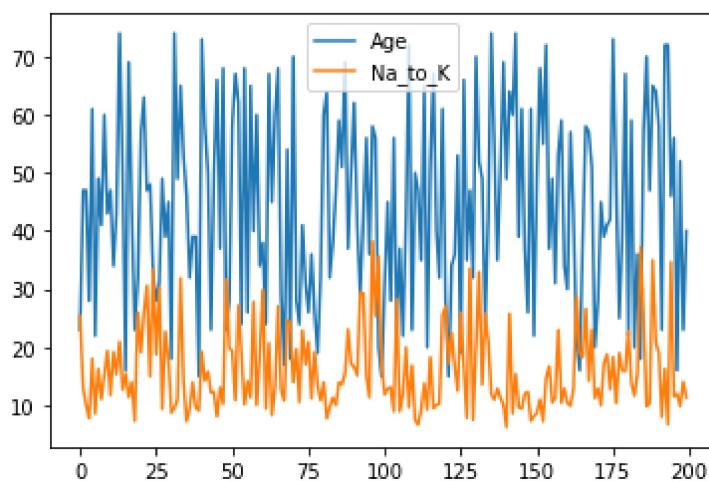
```
In [13]: print(data1.sum())
```

```
Age      8863.000  
Na_to_K  3216.897  
dtype: float64
```

```
In [14]: import matplotlib.pyplot as plot
```

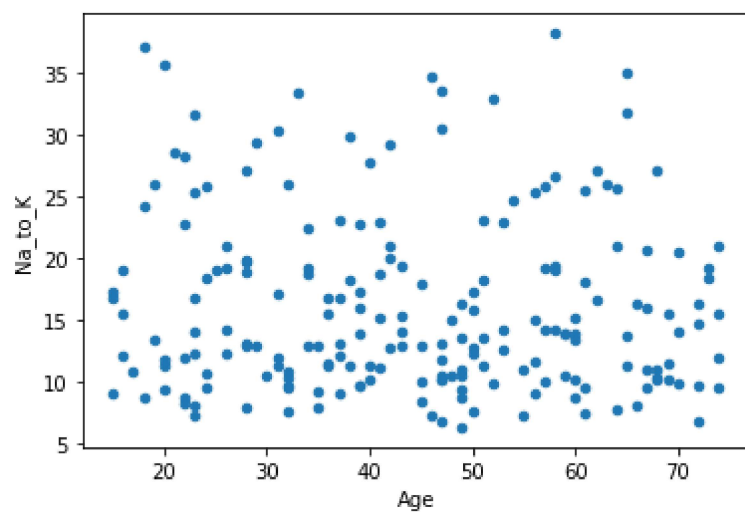
```
In [15]: data1.plot.line()
```

```
Out[15]: <AxesSubplot:>
```



```
In [20]: data.plot.scatter("Age", "Na_to_K")
```

```
Out[20]: <AxesSubplot:xlabel='Age', ylabel='Na_to_K'>
```

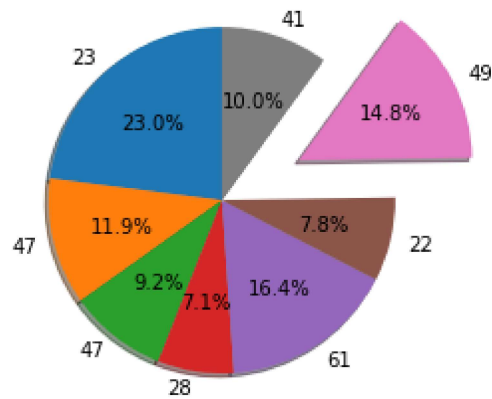


```
In [22]: data3 = data[["Age", "Na_to_K"]][0:8]
data3
```

```
Out[22]:
```

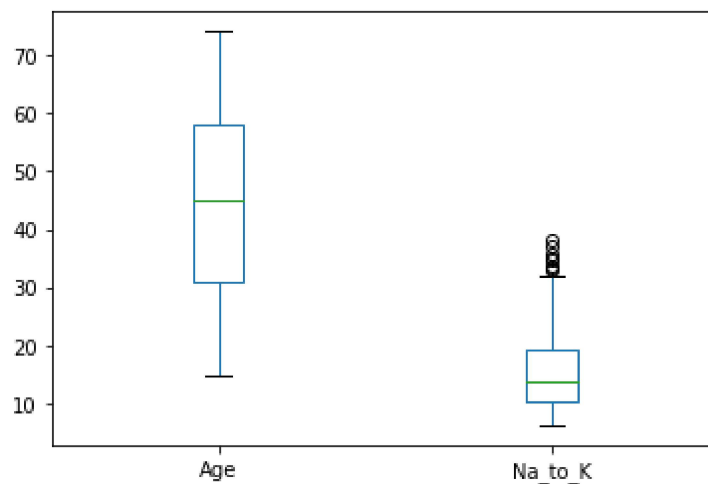
	Age	Na_to_K
0	23	25.355
1	47	13.093
2	47	10.114
3	28	7.798
4	61	18.043
5	22	8.607
6	49	16.275
7	41	11.037

```
In [24]: slice = data3["Na_to_K"]  
country = data3["Age"]  
#col = ["r","g","b","orange","y","black","white","violet","brown"]  
plot.pie(slice,labels=country,startangle=90,shadow=True,explode=(0,0,0,0,0,0,0.5,0,0),  
plot.show()
```



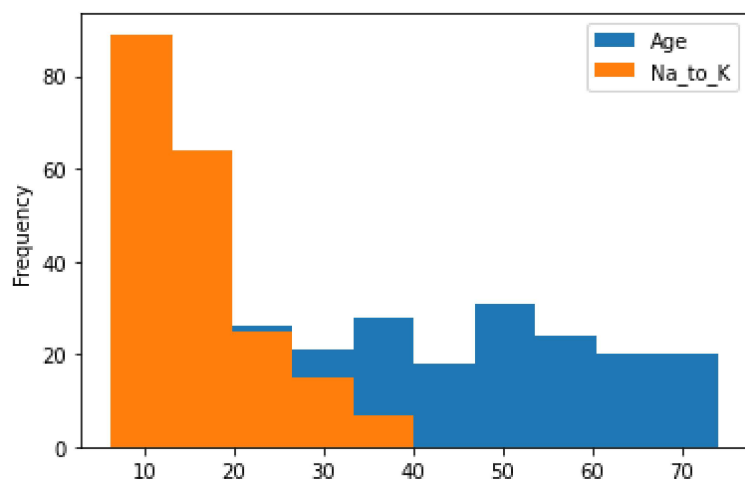
```
In [25]: data.plot.box()
```

Out[25]: <AxesSubplot:>



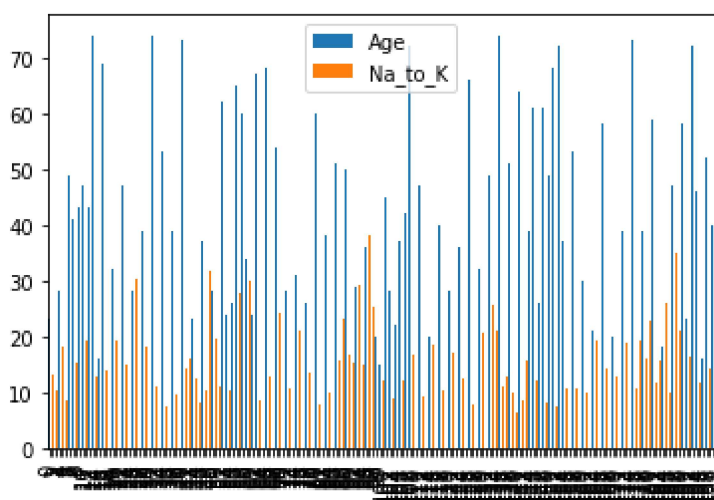
```
In [27]: data1.plot.hist()
```

```
Out[27]: <AxesSubplot:ylabel='Frequency'>
```



```
In [28]: data1.plot.bar()
```

```
Out[28]: <AxesSubplot:>
```



```
In [ ]:
```