

data visualization

(a) Import library

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

b) Import dataset

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

```
In [3]: data
```

Out[3]:

	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

c)head

```
In [4]: data.head(10)
```

```
Out[4]:
```

	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
5	22	F	NORMAL	HIGH	8.607	drugX
6	49	F	NORMAL	HIGH	16.275	drugY
7	41	M	LOW	HIGH	11.037	drugC
8	60	M	NORMAL	HIGH	15.171	drugY
9	43	M	LOW	NORMAL	19.368	drugY

d) tail

```
In [5]: data.tail(10)
```

```
Out[5]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
190	58	M	HIGH	HIGH	18.991	drugY
191	23	M	HIGH	HIGH	8.011	drugA
192	72	M	LOW	HIGH	16.310	drugY
193	72	M	LOW	HIGH	6.769	drugC
194	46	F	HIGH	HIGH	34.686	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

e) describe

In [6]: `data.describe()`

Out[6]:

	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

f) shape

In [7]: `data.shape`

Out[7]: (200, 6)

g) size

In [8]: `data.size`

Out[8]: 1200

h) find missing values

```
In [9]: data.isna()
```

```
Out[9]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...
195	False	False	False	False	False	False
196	False	False	False	False	False	False
197	False	False	False	False	False	False
198	False	False	False	False	False	False
199	False	False	False	False	False	False

200 rows × 6 columns

```
In [23]: data.isnull().sum()
```

```
Out[23]: Age          0
Sex          0
BP           0
Cholesterol  0
Na_to_K      0
Drug         0
dtype: int64
```

```
In [11]: data.dropna(axis=1,how='any')
```

Out[11]:

	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 [13]: data1 = data[['Age', 'Na_to_K']]
data1
```

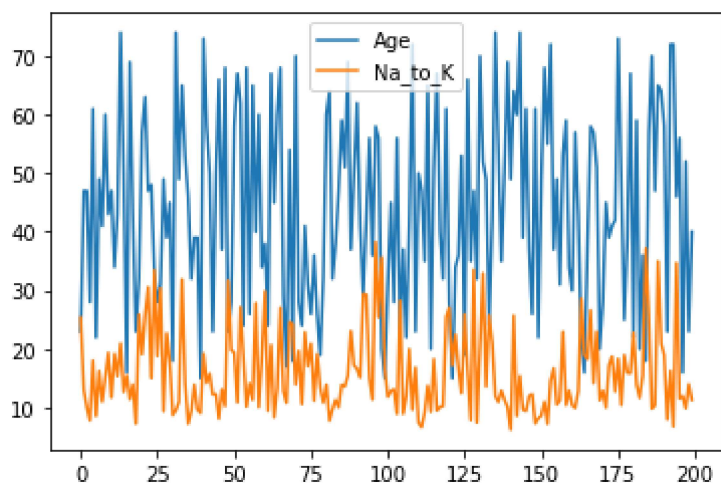
Out[13]:

	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

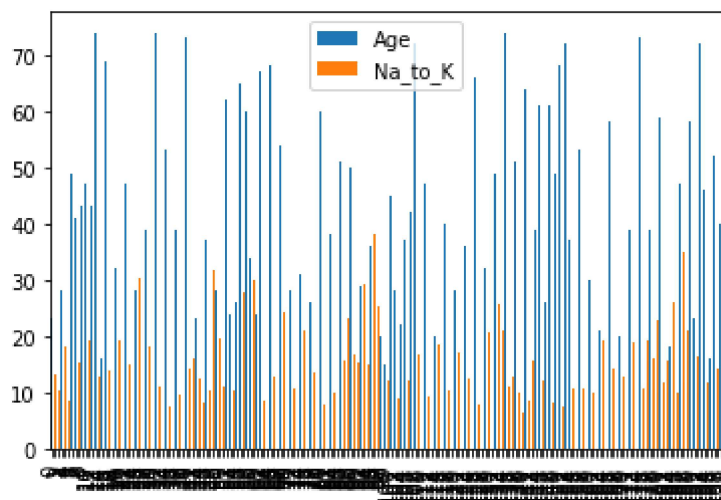
```
In [14]: data1.plot.line()
```

```
Out[14]: <AxesSubplot:>
```



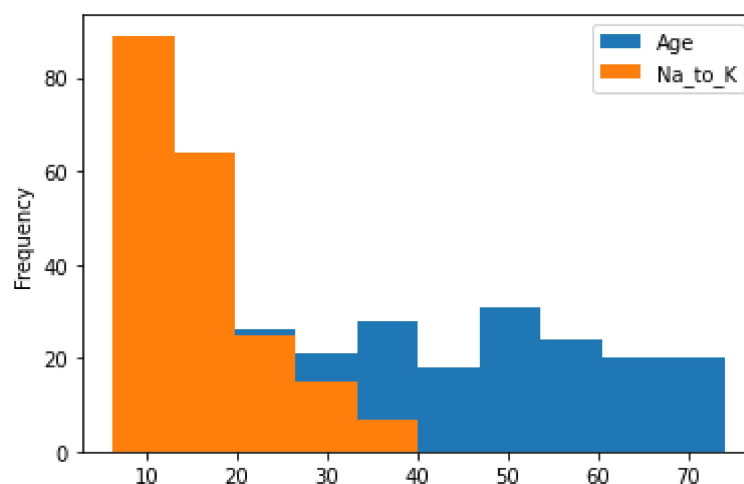
```
In [15]: data1.plot.bar()
```

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



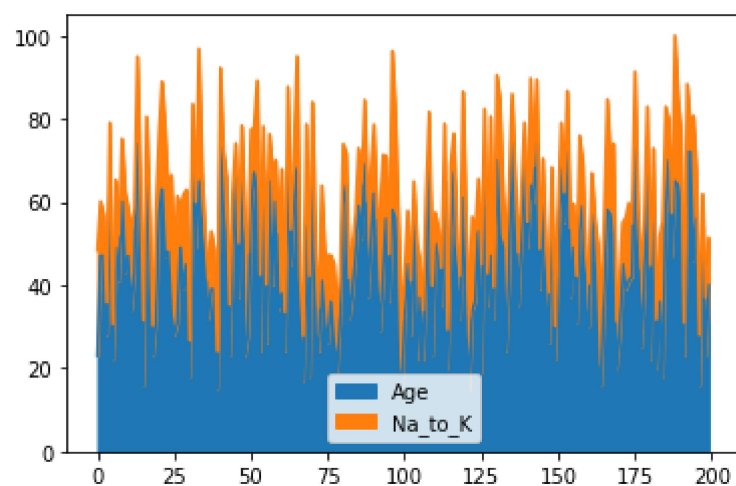
```
In [16]: data1.plot.hist()
```

```
Out[16]: <AxesSubplot:ylabel='Frequency'>
```



```
In [17]: data1.plot.area()
```

```
Out[17]: <AxesSubplot:>
```

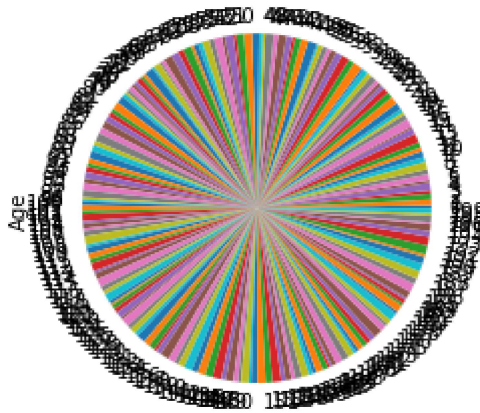


```
In [18]: data2 = data1['Age']  
data2
```

```
Out[18]: 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 [19]: data2.plot.pie()
```

```
Out[19]: <AxesSubplot:ylabel='Age'>
```



```
In [ ]:
```