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

In [12]:
```
a=pd.read_csv(r"C:\Users\user\Downloads\4_drug200 - 4_drug200.csv")
a
```

Out[12]:

| | 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]:
```
a.isnull()
```

Out[13]:

| | 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 [14]:
```
a.fillna(value=0)
```

Out[14]:

| | 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 [16]:
```
a.dropna()
```

Out[16]:

| | 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 [8]:
```
a.head(8)
```

Out[8]:

| | 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 |

In [6]:
```
a.tail()
```

Out[6]:

| | 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 [9]:
```
a.describe()
```

Out[9]:

| | 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 [17]:
```
a.loc[0:4]
```

Out[17]:

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

| | Age | Sex | BP | Cholesterol | Na_to_K | Drug |
|---|---|---|---|---|---|---|
| **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 [18]:
```
a.iloc[0:4]
```

Out[18]:

| | 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 |

In [20]:
```
np.shape(a)
```

Out[20]:
```
(200, 6)
```

In [21]:
```
np.size(a)
```

Out[21]:
```
1200
```

In [23]:
```
np.ndim(a)
```

Out[23]:
```
2
```

In [25]:
```
a.columns
```

Out[25]:
```
Index(['Age', 'Sex', 'BP', 'Cholesterol', 'Na_to_K', 'Drug'], dtype='object')
```

In [26]:
```
a.index
```

Out[26]:
```
RangeIndex(start=0, stop=200, step=1)
```

In [43]:
```
a1=a[["Age","Na_to_K"]]
a1
```

Out[43]:

| | Age | Na_to_K |
|---|---|---|
| **0** | 23 | 25.355 |
| **1** | 47 | 13.093 |
| **2** | 47 | 10.114 |

| | Age | Na_to_K |
|---|---|---|
| **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 [47]:
```
a1.mean()
```

Out[47]:
```
Age        44.315000
Na_to_K    16.084485
dtype: float64
```

In [45]:
```
a1.median()
```

Out[45]:
```
Age        45.0000
Na_to_K    13.9365
dtype: float64
```

In [44]:
```
a1.mode()
```

Out[44]:

| | Age | Na_to_K |
|---|---|---|
| **0** | 47.0 | 12.006 |
| **1** | NaN | 18.295 |

In [48]:
```
a1.sum()
```

Out[48]:
```
Age        8863.000
Na_to_K    3216.897
dtype: float64
```

In [49]:
```
a1.cumsum()
```

Out[49]:

| | Age | Na_to_K |
|---|---|---|
| **0** | 23 | 25.355 |
| **1** | 70 | 38.448 |
| **2** | 117 | 48.562 |
| **3** | 145 | 56.360 |

| | Age | Na_to_K |
|---|---|---|
| **4** | 206 | 74.403 |
| **...** | ... | ... |
| **195** | 8732 | 3169.628 |
| **196** | 8748 | 3181.634 |
| **197** | 8800 | 3191.528 |
| **198** | 8823 | 3205.548 |
| **199** | 8863 | 3216.897 |

200 rows × 2 columns

In [50]:
```
a1.max()
```

Out[50]:
```
Age        74.000
Na_to_K    38.247
dtype: float64
```

In [51]:
```
a1.min()
```

Out[51]:
```
Age        15.000
Na_to_K     6.269
dtype: float64
```

In [52]:
```
a1.count()
```

Out[52]:
```
Age        200
Na_to_K    200
dtype: int64
```

In [53]:
```
from numpy import cov
from scipy.stats import pearsonr
from scipy.stats import spearmanr
```

In [55]:
```
b1=a["Age"].values
b2=a["Na_to_K"].values
print(np.cov(b1,b2))
```

```
[[273.71434673  -7.54375153]
 [ -7.54375153  52.18553348]]
```

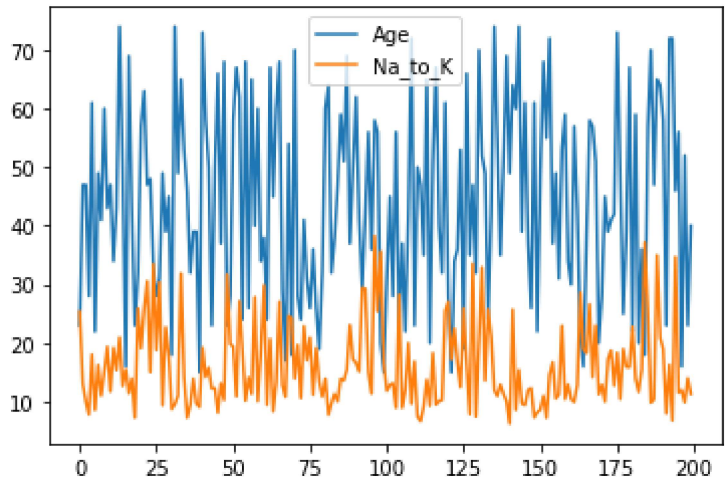
In [56]:
```
spearman =pd.Series(b1).corr(pd.Series(b2),method='spearman')
pearson =pd.Series(b1).corr(pd.Series(b2),method='pearson')
print(spearman)
print(pearson)
```

```
-0.047273882688479915
-0.06311949726772588
```

# visualization

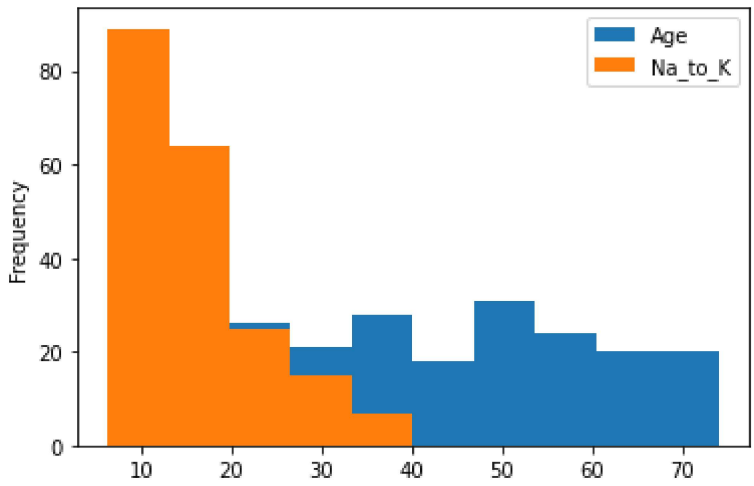
In [57]:
```
a1.plot.line()
```

Out[57]: <AxesSubplot:>



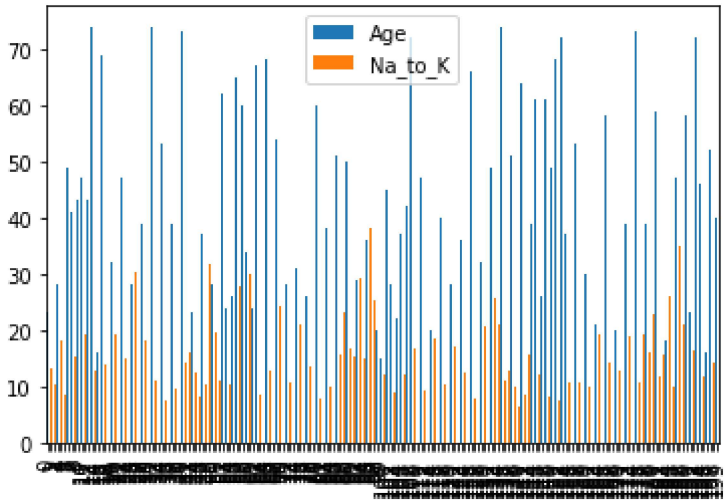

In [58]:
```
a1.plot.hist()
```

Out[58]: <AxesSubplot:ylabel='Frequency'>



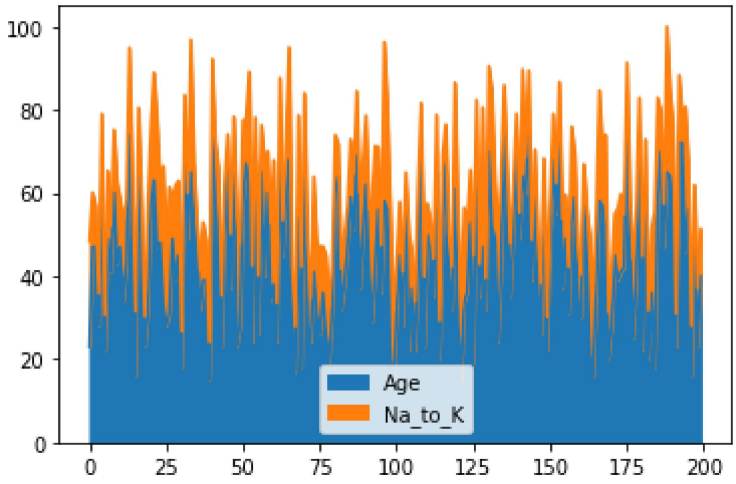

In [59]:
```
a1.plot.bar()
```

Out[59]: <AxesSubplot:>



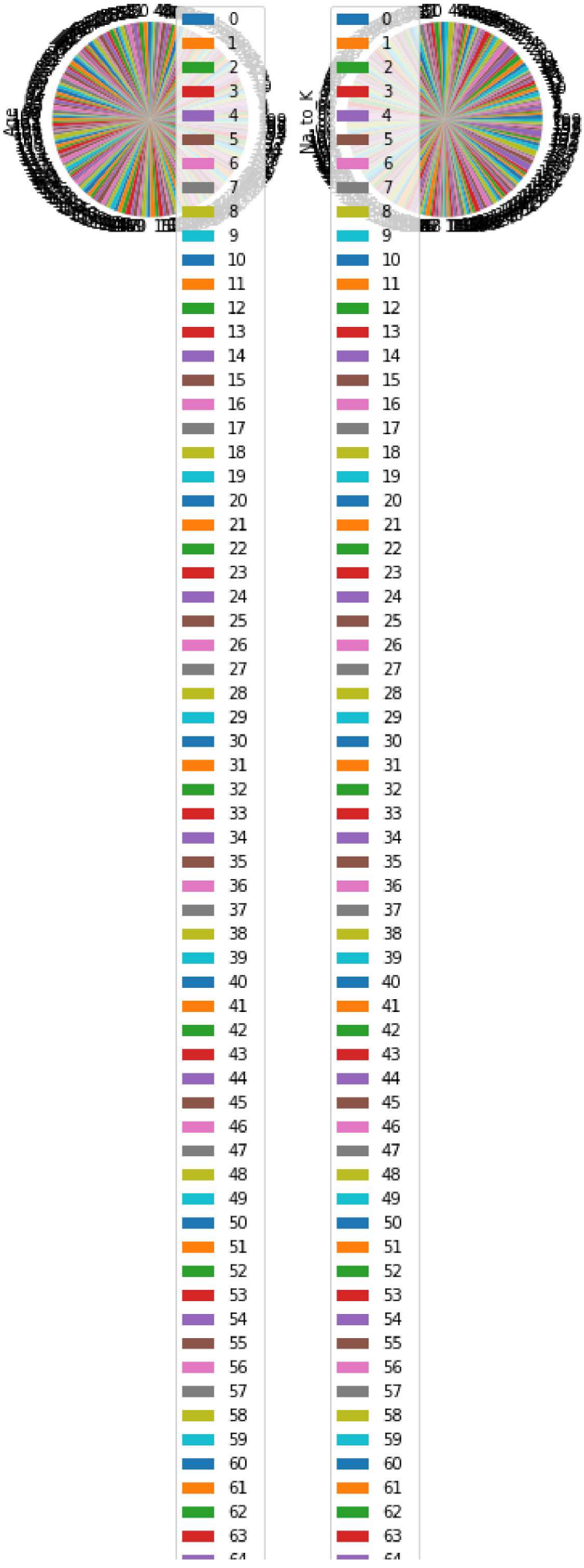

In [60]:
```
a1.plot.area()
```

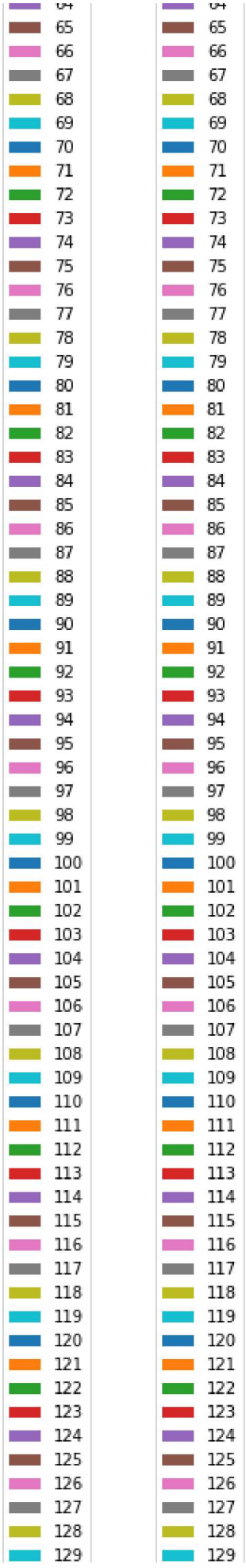
Out[60]: <AxesSubplot:>




In [61]:
```
a1.plot.pie(subplots=True)
```

Out[61]:
```
array([<AxesSubplot:ylabel='Age'>, <AxesSubplot:ylabel='Na_to_K'>],
      dtype=object)
```

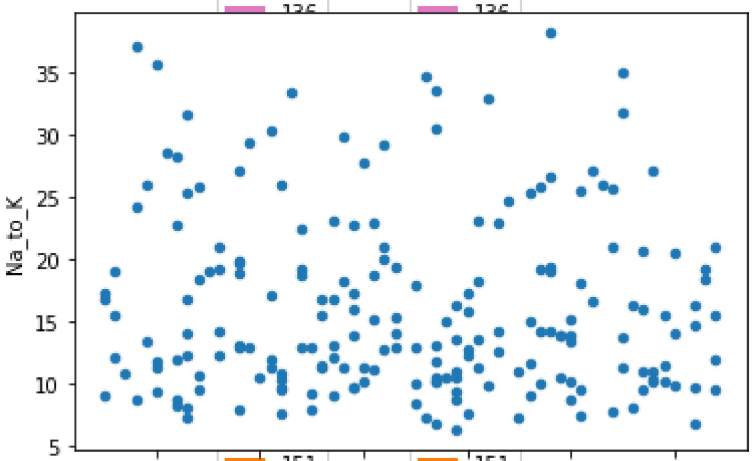

Age
Na_to_K
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63

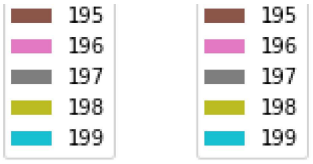


64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129

In [62]:
```
a1.plot.scatter(x="Age",y="Na_to_K")
```

Out[62]: <AxesSubplot:xlabel='Age', ylabel='Na_to_K'>




195
196
197
198
199

195
196
197
198
199