In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as pp
```

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

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
...	...	...	...	...	...	...
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 [5]:
```
d.head()
```

Out[5]:

	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 [6]:
```
d.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

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

In [7]:
```
d.describe()
```

Out[7]:

	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 [9]:
```
np.shape(d)
```

Out[9]:
```
(200, 6)
```

In [10]:
```
np.size(d)
```

Out[10]:
```
1200
```

In [11]:
```
d.isna()
```

Out[11]:

	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

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
199	False	False	False	False	False	False

200 rows × 6 columns

In [12]:
```
d.dropna()
```

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

visualization

In [14]:
```
d=d[["Age","Na_to_K"]]
d
```

Out[14]:

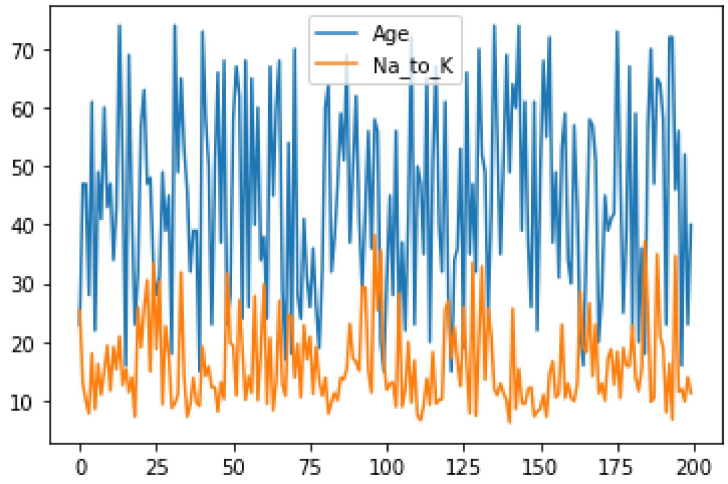
	Age	Na_to_K
0	23	25.355
1	47	13.093
2	47	10.114
3	28	7.798
4	61	18.043
...	...	...
195	56	11.567
196	16	12.006
197	52	9.894
198	23	14.020

	Age	Na_to_K
199	40	11.349

200 rows × 2 columns

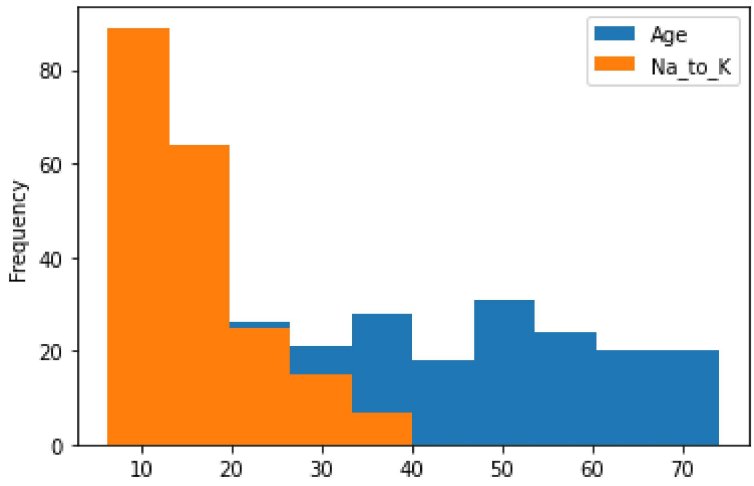
In [15]:
```
d.plot.line()
```

Out[15]: <AxesSubplot:>



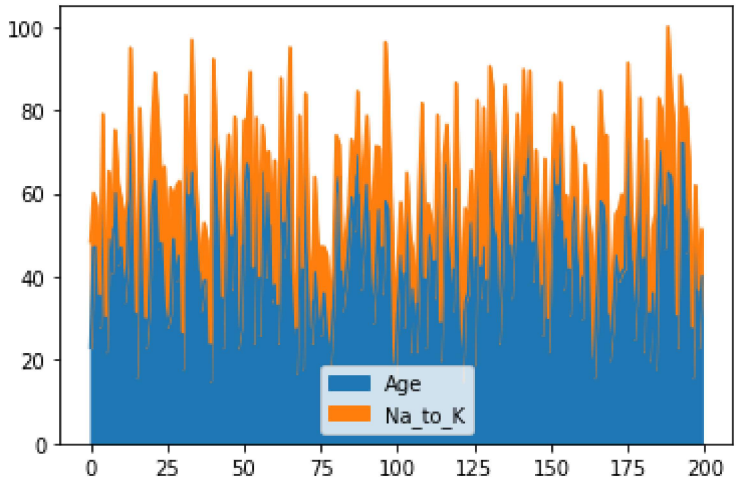
In [16]:
```
d.plot.hist()
```

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



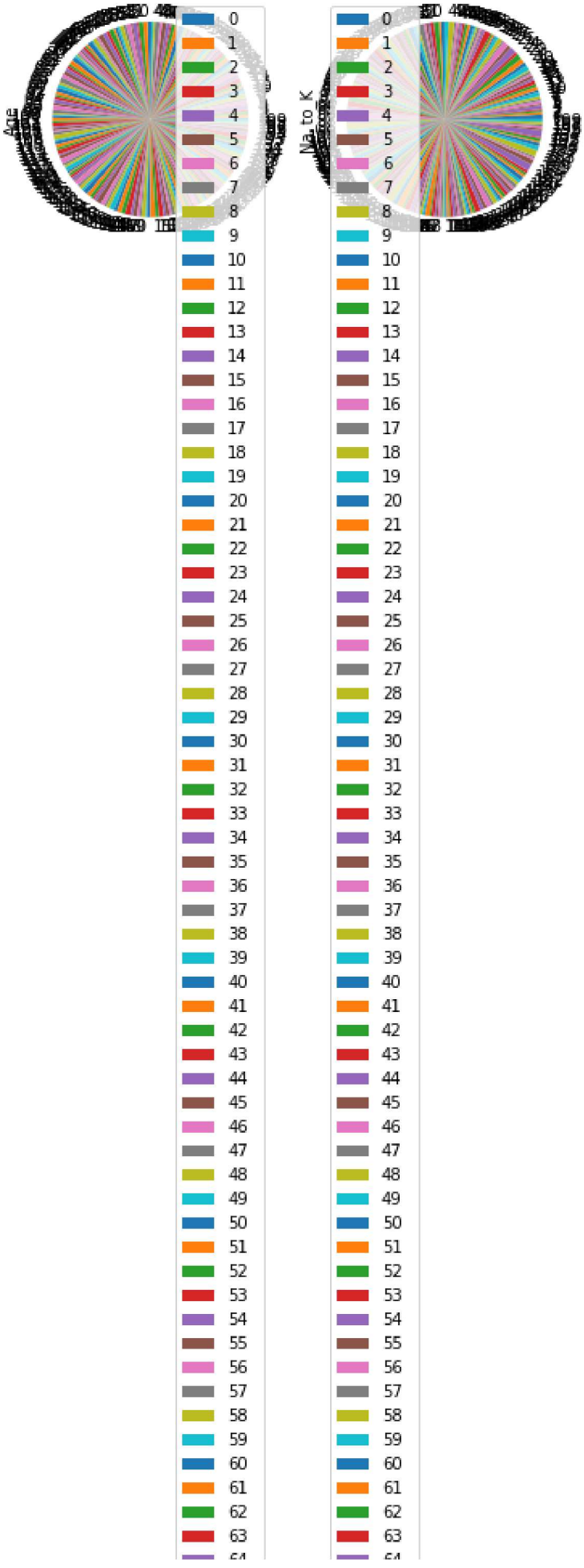
In [28]:
```
d.plot.area()
```

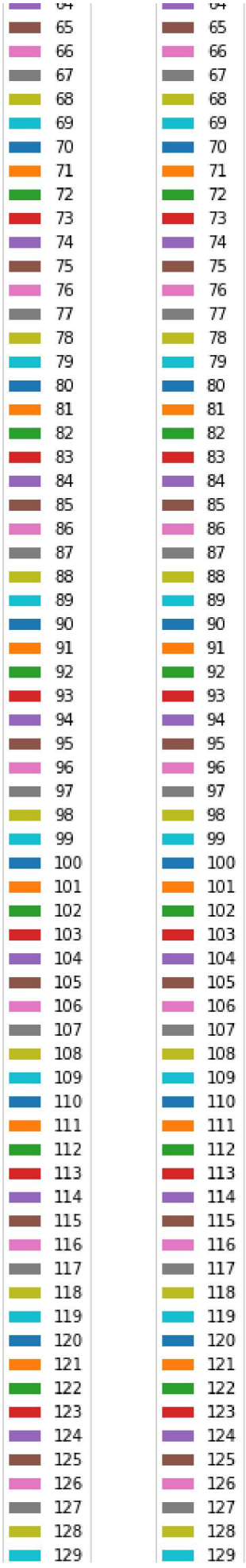
Out[28]: <AxesSubplot:>



In [32]:
```
d.plot.pie(subplots=True)
```

Out[32]:
```
array([<AxesSubplot:ylabel='Age'>, <AxesSubplot:ylabel='Na_to_K'>],
      dtype=object)
```

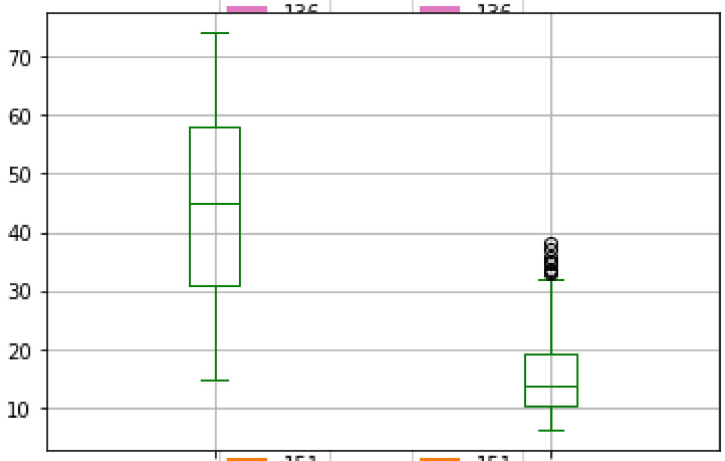

Age
Na_to_K
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63



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

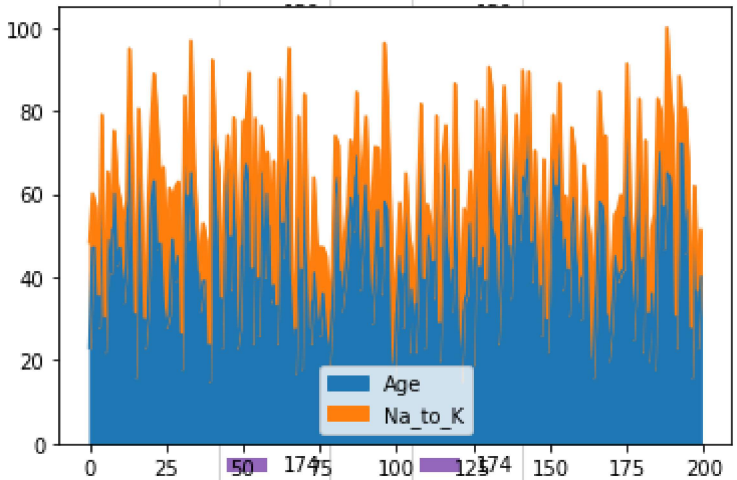
In [31]:
```
d.boxplot(color='g')
```

Out[31]: <AxesSubplot:>



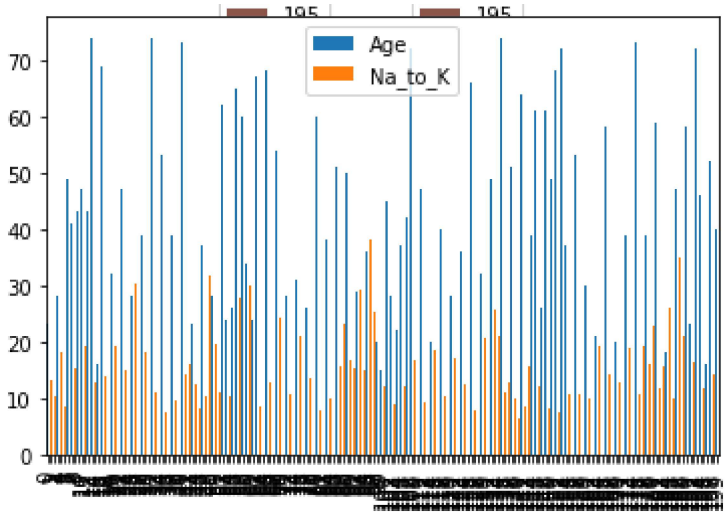
In [33]:
```
d.plot.area()
```

Out[33]: <AxesSubplot:>



In [34]:
```
d.plot.bar()
```

Out[34]: <AxesSubplot:>



In [ ]: