

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Matplotlib is building the font cache; this may take a moment.

In [2]:

```
emp= {'Age_Years':[10,23,22,21,23,24,27,24,29,32,25,24], 'Salary_Thosands':[15,22,52,21,34,36,44,63,37,41,23,29]}
df=pd.DataFrame(data=emp)
df
```

Out[2]:

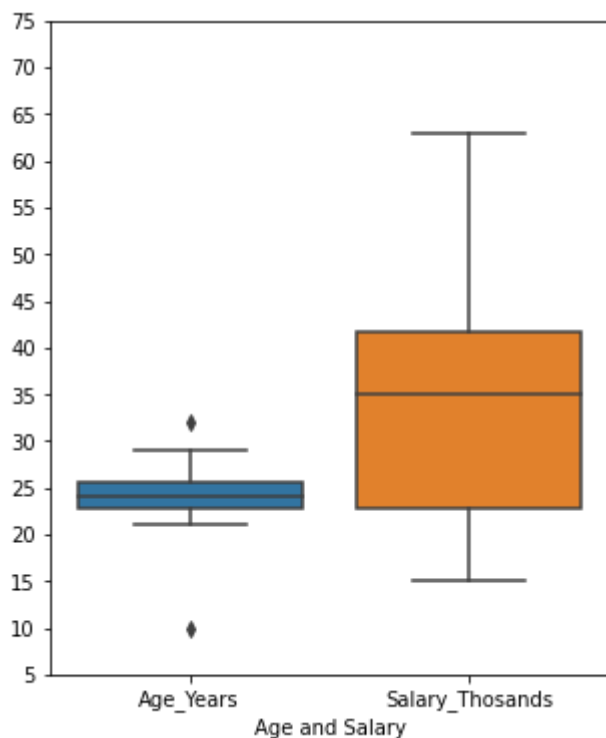
	Age_Years	Salary_Thosands
0	10	15
1	23	22
2	22	52
3	21	21
4	23	34
5	24	36
6	27	44
7	24	63
8	29	37
9	32	41
10	25	23
11	24	29

In [4]:

```
#Box plot
plt.figure(figsize=(5,6))
ax=sns.boxplot(data=df)
plt.yticks(range(5,80,5))    # range on y axis
plt.xlabel('Age and Salary')
plt.show                    # seaborn version will keep changing, if code doesn't work
                             #https://seaborn.pydata.org
```

Out[4]:

```
<function matplotlib.pyplot.show(close=None, block=None)>
```

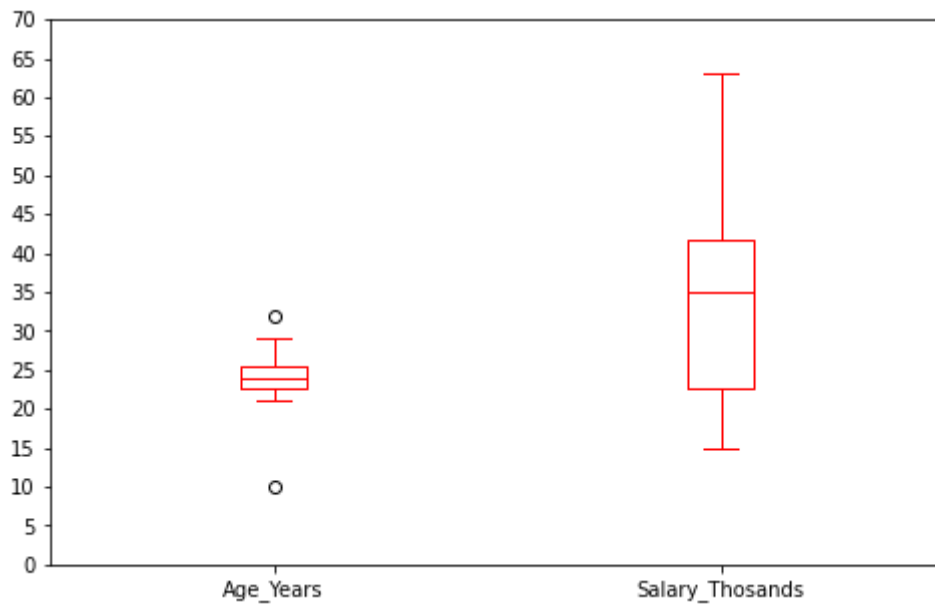


In [6]:

```
df.plot(kind='box', figsize=(8,5),color='r', yticks=range(0,75,5))
```

Out[6]:

<AxesSubplot:>



In [7]:

```
sal= [15,22,52,21,34,36,44,63,37,41,23,29]  
sal.sort()  
sal
```

Out[7]:

```
[15, 21, 22, 23, 29, 34, 36, 37, 41, 44, 52, 63]
```

In [8]:

```
# outliers  
import pandas as pd  
ser= pd.Series(sal)
```

In [9]:

```
ser
```

Out[9]:

```
0    15
1    21
2    22
3    23
4    29
5    34
6    36
7    37
8    41
9    44
10   52
11   63
dtype: int64
```

In [10]:

```
# 1st quantile
q1=ser.quantile(0.25)
print('Q1',q1)

#3rd quantile
q3=ser.quantile(0.75)
print('Q3',q3)

#Find the IQR(Inter quantile range) to identify outliers

iqr= q3 -q1
print('IQR',iqr)
```

```
Q1 22.75
Q3 41.75
IQR 19.0
```

In [11]:

```
# Outlier detection formula
# higher side= Q3+(1.5 * IQR)
# lower side = Q1-(1.5 * IQR)
```

In [12]:

```
upside= q3 + (1.5 * iqr)
print(upside)
```

```
70.25
```

In [13]:

```
#anything above 70. 25 is outlier
```

In [14]:

```
lower_side= q1-(1.5 * iqr)  
print(lower_side)
```

-5.75

In []:

```
#anything below -5.75 is outlier
```