

In [1]:

```
import matplotlib.pyplot as plt
import pandas as pd
import pylab as pl
import numpy as np
%matplotlib inline
```

In [10]:

```
cc=pd.read_csv("syyy.csv.csv")
```

In [11]:

```
cc.head()
```

Out[11]:

	Ye	Salary
0	1	1
1	2	2
2	3	3
3	4	4
4	5	5

In [12]:

```
cc.describe()
```

Out[12]:

	Ye	Salary
count	10.00000	10.00000
mean	5.50000	5.50000
std	3.02765	3.02765
min	1.00000	1.00000
25%	3.25000	3.25000
50%	5.50000	5.50000
75%	7.75000	7.75000
max	10.00000	10.00000

In [13]:

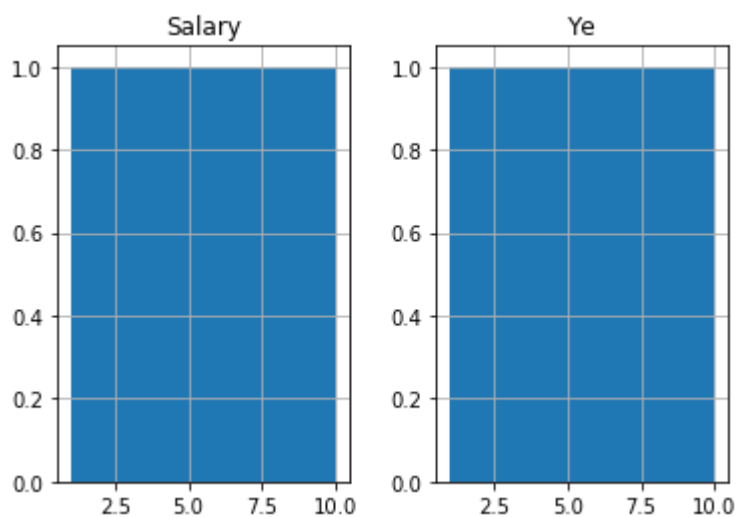
```
cd=cc[['Ye','Salary']]  
cd.head()
```

Out[13]:

	Ye	Salary
0	1	1
1	2	2
2	3	3
3	4	4
4	5	5

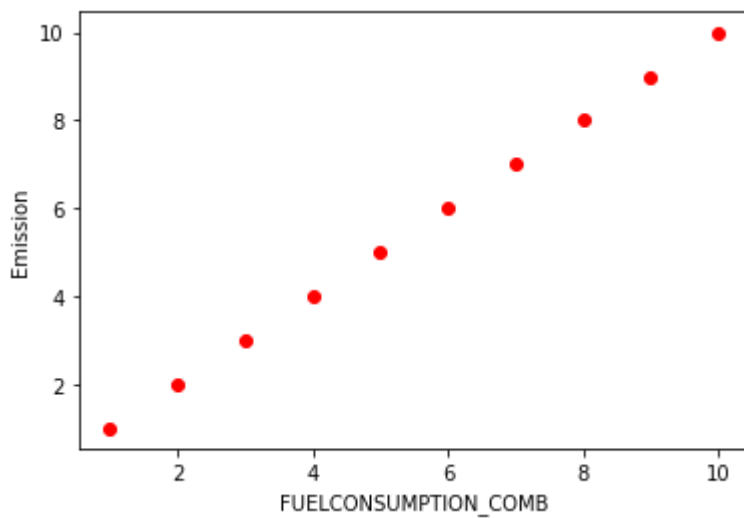
In [15]:

```
viz = cd[['Ye','Salary']]  
viz.hist()  
plt.show()
```



In [17]:

```
plt.scatter(cd.Ye, cd.Salary, color='r')
plt.xlabel("FUELCONSUMPTION_COMB")
plt.ylabel("Emission")
plt.show()
```



In [20]:

```
msk = np.random.rand(len(cc)) < 0.8
train = cd[msk]
test = cd[~msk]
```

In [21]:

```
from sklearn import linear_model
regr = linear_model.LinearRegression()
train_x = np.asanyarray(train[['Ye']])
train_y = np.asanyarray(train[['Salary']])
regr.fit(train_x, train_y)
# The coefficients
print('Coefficients: ', regr.coef_)
print('Intercept: ', regr.intercept_)
```

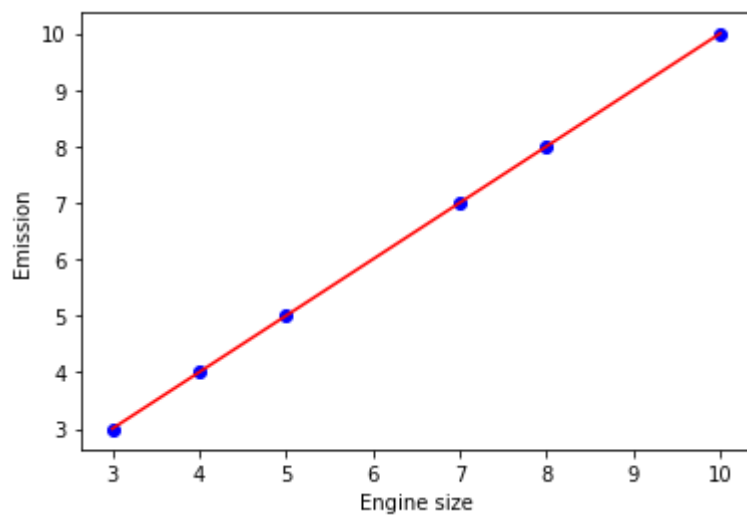
```
Coefficients: [[1.]]
Intercept: [-1.77635684e-15]
```

In [22]:

```
plt.scatter(train.Ye, train.Salary, color='blue')  
plt.plot(train_x, regr.coef_[0][0]*train_x + regr.intercept_[0], '-r')  
plt.xlabel("Engine size")  
plt.ylabel("Emission")
```

Out[22]:

Text(0, 0.5, 'Emission')



In []: