In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
import seaborn as sns
sns.set(color_codes=True)
```

In [2]:

```
df = pd.read_csv('housing_boston.csv')
df.head()
```

Out[2]:

	0.00632	18	2.31	0	0.538	6.575	65.2	4.09	1	296	15.3	396.9	4.98	24
0	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
1	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
2	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
3	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2
4	0.02985	0.0	2.18	0	0.458	6.430	58.7	6.0622	3	222	18.7	394.12	5.21	28.7

In [3]:

```
df.shape
df.describe()
```

Out[3]:

	0.00632	18	2.31	0	0.538	6.575	65.2	
count	451.000000	451.000000	451.000000	451.000000	451.000000	451.000000	451.000000	45
mean	1.423961	12.709534	10.322616	0.077605	0.540822	6.343024	65.558758	
std	2.497774	24.351772	6.794183	0.267847	0.113942	0.667459	28.158255	
min	0.009060	0.000000	0.460000	0.000000	0.385000	3.561000	2.900000	
25%	0.070175	0.000000	4.930000	0.000000	0.447000	5.926500	40.800000	
50%	0.191330	0.000000	8.140000	0.000000	0.518000	6.229000	71.900000	
75%	1.215500	20.000000	18.100000	0.000000	0.605000	6.635000	91.650000	
max	9.966540	100.000000	27.740000	1.000000	0.871000	8.780000	100.000000	1

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 451 entries, 0 to 450
Data columns (total 14 columns):
 #   Column   Non-Null Count  Dtype  
---  ------   --------------  -----  
 0   0.00632  451 non-null    float64
 1   18       451 non-null    float64
 2   2.31     451 non-null    float64
 3   0        451 non-null    int64  
 4   0.538    451 non-null    float64
 5   6.575    451 non-null    float64
 6   65.2     451 non-null    float64
 7   4.09     451 non-null    float64
 8   1        451 non-null    int64  
 9   296      451 non-null    int64  
 10  15.3     451 non-null    float64
 11  396.9    451 non-null    float64
 12  4.98     451 non-null    float64
 13  24       451 non-null    float64
dtypes: float64(11), int64(3)
memory usage: 49.5 KB
```

In [5]:

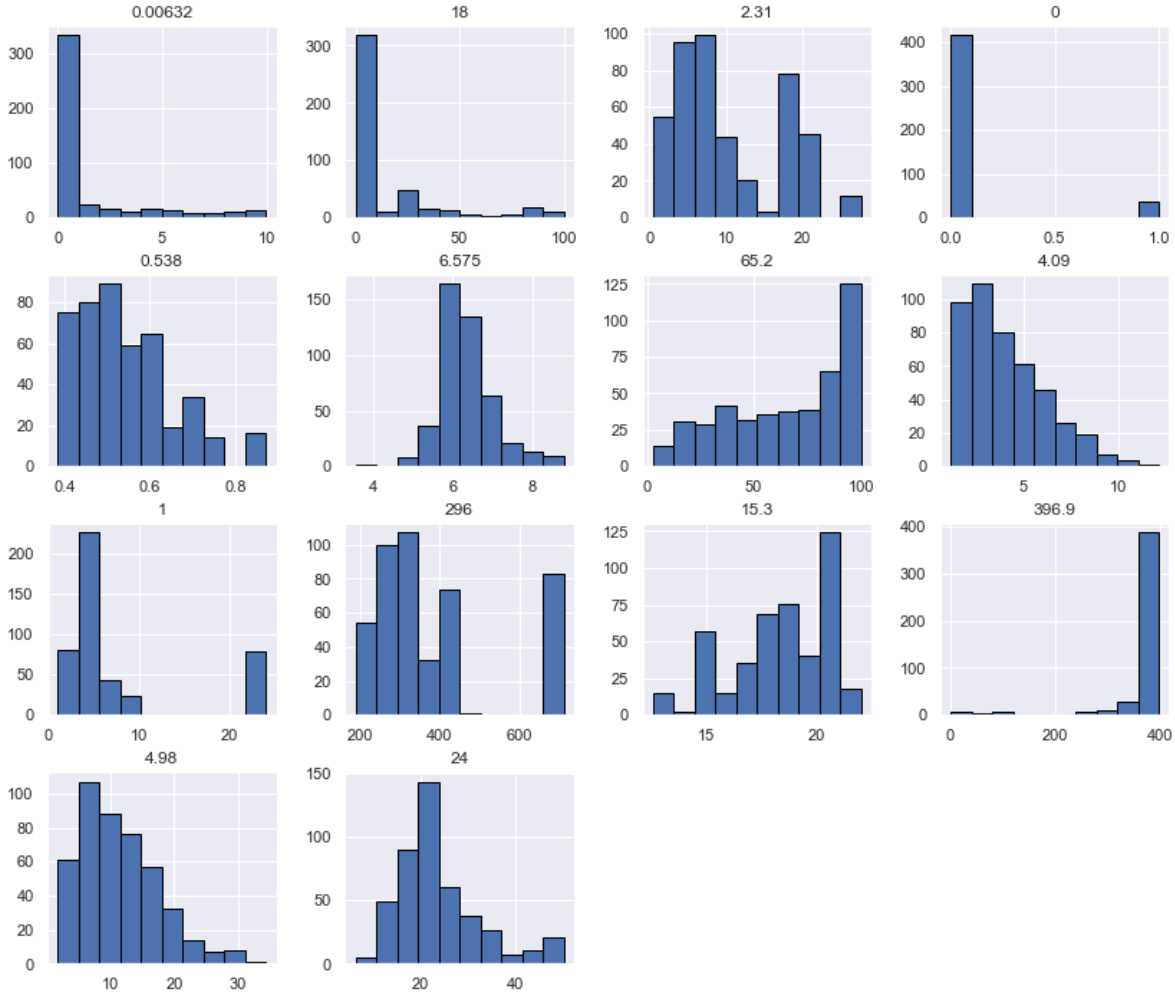
```
df.isna().sum()
```

Out[5]:

```
0.00632    0
18         0
2.31       0
0          0
0.538      0
6.575      0
65.2       0
4.09       0
1          0
296        0
15.3       0
396.9      0
4.98       0
24         0
dtype: int64
```

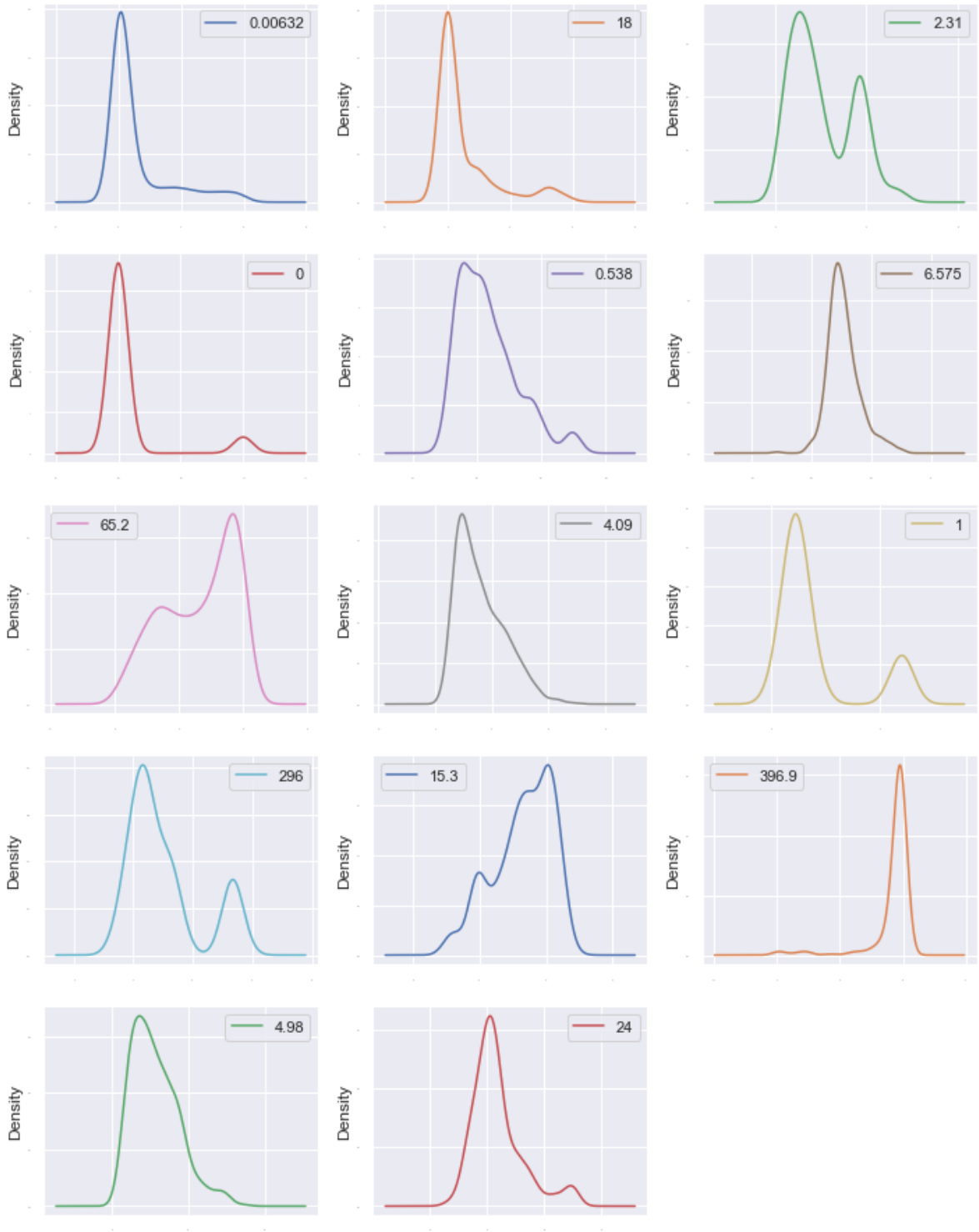
In [6]:

```
df.hist(edgecolor= 'black',figsize=(14,12))
plt.show()
```



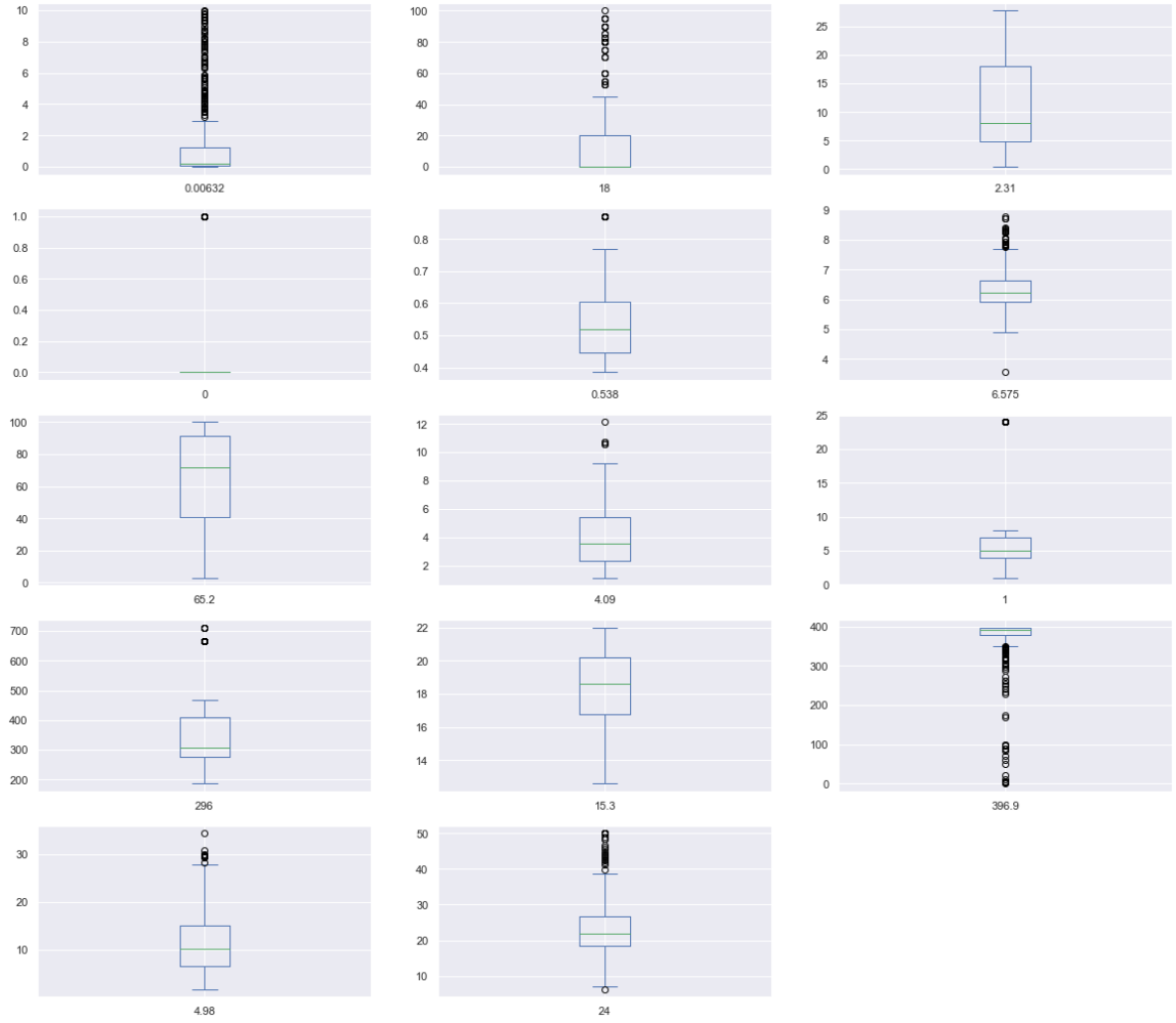
In [7]:

```
df.plot(kind='density', subplots=True, layout= (5,3), sharex=False,
legend=True, fontsize=1, figsize= (12,16))
plt.show()
```



In [8]:

```
df.plot(kind="box", subplots=True, layout=(5,3), sharex=False,
figsize=(20,18))
plt.show()
```



In [9]:

```
sns.pairplot(df, height=1.5);
plt.show()
```