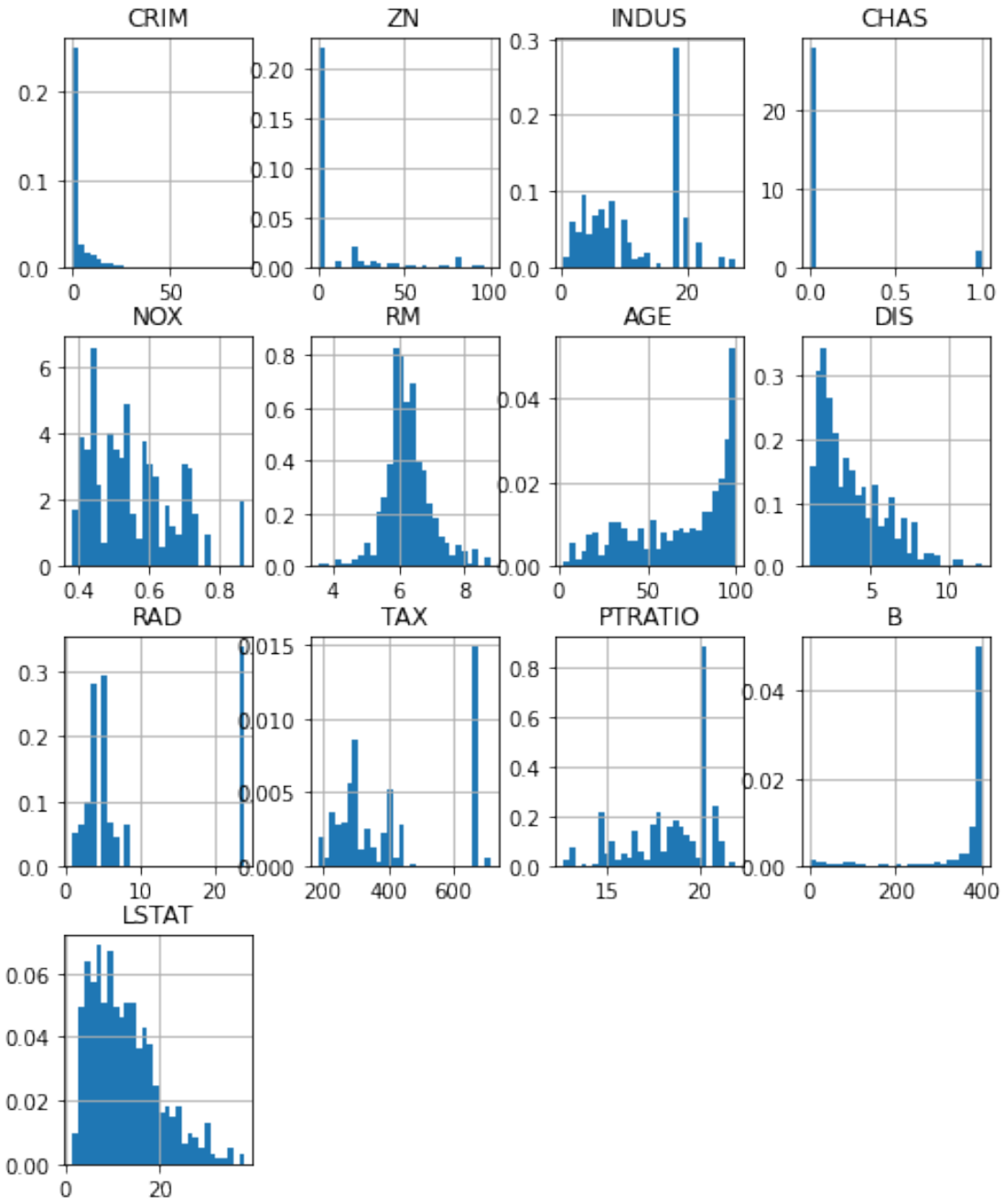
Distinguishing variable distribution

A probability distribution is a function that describes the likelihood of obtaining the possible values of a variable. There are many well-described variable distributions, such as the normal, binomial, or Poisson distributions. Some machine learning algorithms assume that the independent variables are normally distributed. Other models make no assumptions about the distribution of the variables, but a better spread of these values may improve their performance.so here, we will learn how to create plots to distinguish the variable distributions in the entire dataset by using the Boston House Prices dataset from scikit-learn.

```
# importing the required python libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.datasets import load_boston

# load boston dataset
boston_dataset = load_boston()
boston =
pd.DataFrame(boston_dataset.data,columns=boston_dataset.feature_names)

boston.head()
```

```
      CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD    TAX
\
0  0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900  1.0  296.0

1  0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671  2.0  242.0

2  0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671  2.0  242.0

3  0.03237   0.0   2.18   0.0  0.458  6.998  45.8  6.0622  3.0  222.0

4  0.06905   0.0   2.18   0.0  0.458  7.147  54.2  6.0622  3.0  222.0


   PTRATIO       B  LSTAT
0     15.3  396.90   4.98
1     17.8  396.90   9.14
2     17.8  392.83   4.03
3     18.7  394.63   2.94
4     18.7  396.90   5.33
```

```
# visualize the variable distribution
boston.hist(bins=30,figsize=(8,10),density=True)
plt.show()
```



`observation`: Most of the numerical variables in the dataset are skewed.