- Read the dataset from the url into a dataframe.
- Display the first few rows to make sure it was read properly.

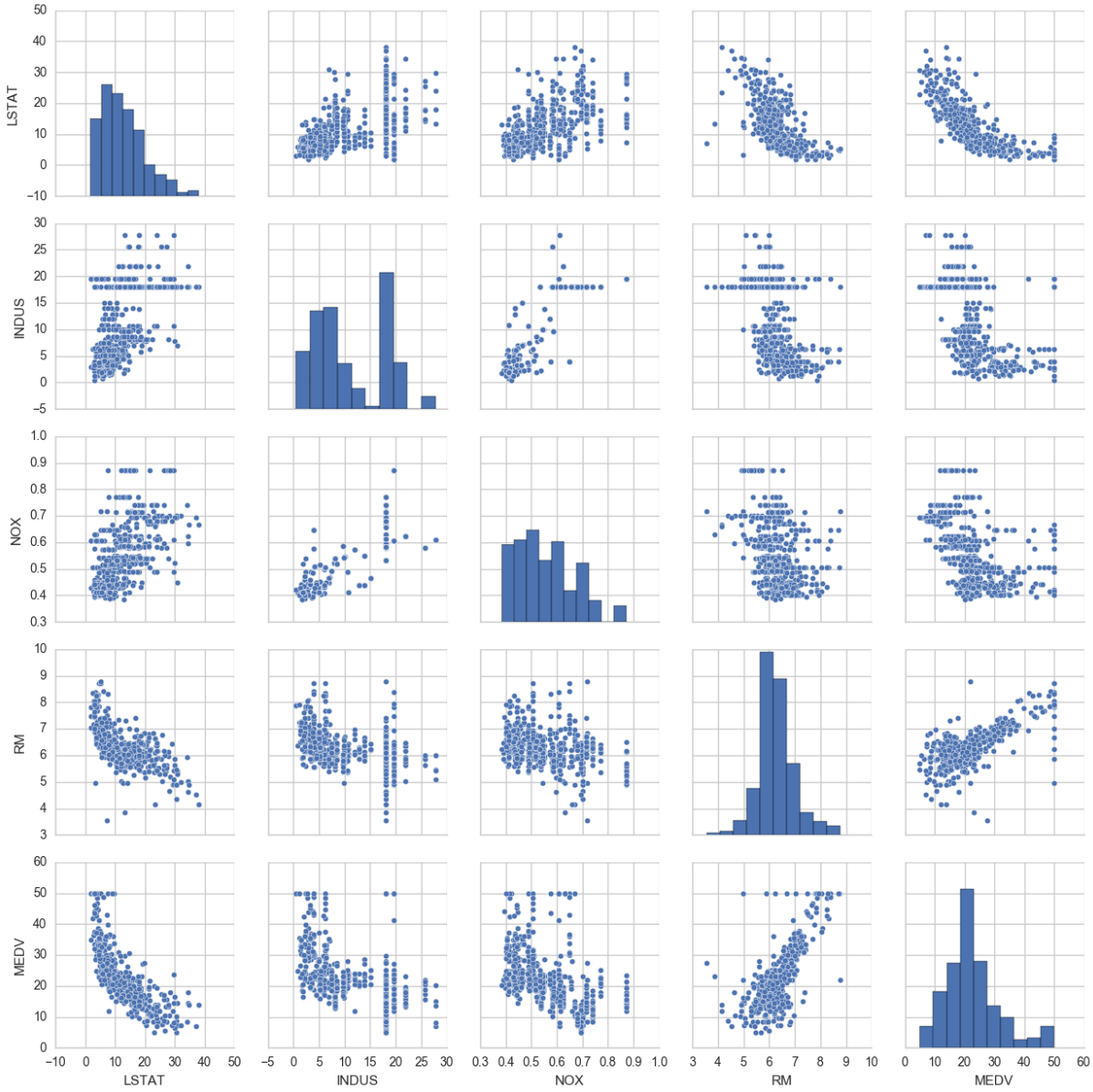
In [3]:
```
import pandas as pd
df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/housing/housing.data',
                 header=None, sep='\s+')
df.columns = ['CRIM', 'ZN', 'INDUS', 'CHAS',
              'NOX', 'RM', 'AGE', 'DIS', 'RAD',
              'TAX', 'PTRATIO', 'B', 'LSTAT', 'MEDV']
df.head()
```

Out[3]:

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296.0	15.3	396.90
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242.0	17.8	396.90
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242.0	17.8	392.83
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222.0	18.7	394.63
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222.0	18.7	396.90

- Visualize the important characteristics of the dataset before building the model - Exploratory Data Analysis (EDA). We will create a scatter matrix to visialize the pair-wise correlations between different features in te dataset.

In [5]:
```
import matplotlib.pyplot as plt
import seaborn as sb
sb.set(style='whitegrid', context='notebook')
cols = ['LSTAT', 'INDUS', 'NOX', 'RM', 'MEDV']
sb.pairplot(df[cols], size=2.5)
plt.show()
```



- Create a correlation matrix to quantify the linear relationship between features.

In [9]:
```
import numpy as np
cor_matrix = np.corrcoef(df[cols].values.T)
sb.set(font_scale=1.5)
cor_heat_map = sb.heatmap(cor_matrix, cbar=True, annot=True,square=True,
                          fmt='.2f', annot_kws={'size':15},
                          yticklabels=cols,
                          xticklabels=cols)
plt.show()
```