In [1]:
```
import numpy as np
import pandas as pd
import statsmodels.api as sm
from sklearn.datasets import load_boston
import matplotlib.pyplot as plt
%matplotlib inline

boston_data = load_boston()
df = pd.DataFrame()
df['MedianHomePrice'] = boston_data.target
df2 = pd.DataFrame(boston_data.data)
df['CrimePerCapita'] = df2.iloc[:,0];
df.head()
```

```
/opt/conda/lib/python3.6/site-packages/statsmodels/compat/pandas.py:56:
FutureWarning: The pandas.core.datetools module is deprecated and will
be removed in a future version. Please use the pandas.tseries module in
stead.
  from pandas.core import datetools
```

Out[1]:

	MedianHomePrice	CrimePerCapita
0	24.0	0.00632
1	21.6	0.02731
2	34.7	0.02729
3	33.4	0.03237
4	36.2	0.06905

The Boston housing data is a built in dataset in the sklearn library of python. You will be using two of the variables from this dataset, which are stored in **df**. The median home price in thousands of dollars and the crime per capita in the area of the home are shown above.

1. Use this dataframe to fit a linear model to predict the home price based on the crime rate. Use your output to answer the first quiz below. Don't forget an intercept.

In [3]:
```
df['intercept'] = 1

lm = sm.OLS(df['MedianHomePrice'], df[['intercept', 'CrimePerCapita']])
results = lm.fit()
results.summary()
```

Out[3]:

OLS Regression Results

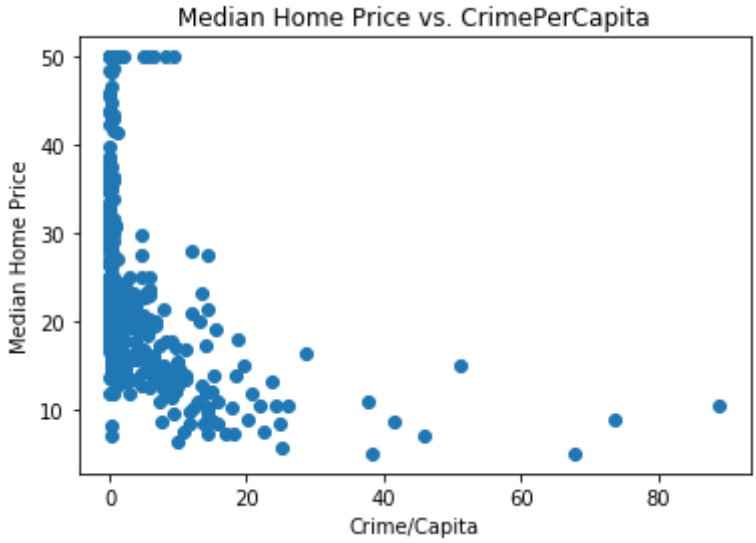
Dep. Variable:	MedianHomePrice	**R-squared:**	0.149
Model:	OLS	**Adj. R-squared:**	0.147
Method:	Least Squares	**F-statistic:**	88.15
Date:	Tue, 14 Apr 2020	**Prob (F-statistic):**	2.08e-19
Time:	02:44:13	**Log-Likelihood:**	-1799.5
No. Observations:	506	**AIC:**	3603.
Df Residuals:	504	**BIC:**	3611.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
intercept	24.0162	0.409	58.676	0.000	23.212	24.820
CrimePerCapita	-0.4128	0.044	-9.389	0.000	-0.499	-0.326

Omnibus:	138.965	**Durbin-Watson:**	0.712
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	292.343
Skew:	1.483	**Prob(JB):**	3.30e-64
Kurtosis:	5.251	**Cond. No.**	10.1

2. Plot the relationship between the crime rate and median home price below. Use your plot and the results from the first question as necessary to answer the remaining quiz questions below.

In [4]:
```
plt.scatter(df['CrimePerCapita'], df['MedianHomePrice']);
plt.xlabel('Crime/Capita');
plt.ylabel('Median Home Price');
plt.title('Median Home Price vs. CrimePerCapita');
```



In [ ]: