

보스턴 주택 가격 데이터 변수 분석

1. 변수 정의

변수이름	변수 정의
CRIM	범죄율
ZN	25,000평방피트 기준 거주지 비율
INDUS	비 소매업종 점유 구역 비율
CHAS	찰스강의 인접 여부(더미변수 / 1=인접해 있음, 0=인접하지 않음)
NOX	이산화질소 농도(백만 개당)
RM	거주지의 평균 방의 개수
AGE	1940년 이전에 건설된 주택에 사는 비율
DIS	보스턴 대표 5개의 회사와 거리에 따른 가중 평균
RAD	고속도로 인접 가능 INDEX
TAX	재산세율(\$10,000당)
PTRATIO	학생/선생 비율
B	유색인종 비율
LSTAT	저소득층 비율
MEDV	집값의 중앙값(단위 \$1,000)

2. 기초통계량

(평균, 중앙값, 사분위수, 사분위수 범위, 범위, 표준편차, 분산)

Console	Terminal
C:/Users/changwan/Desktop/2018.1/데이터마이닝/	
> #기초통계량 조사	
> #평균	
> colMeans(Boston)	
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT MEDV CAT. MEDV	
3.61352 11.36364 11.13678 0.06917 0.55470 6.28463 68.57490 3.79504 9.54941 408.23715 18.45553 356.67403 12.65306 22.53281 0.16601	
> #중앙값	
> apply(Boston[1:15], 2, median)	
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT MEDV CAT. MEDV	
0.25651 0.00000 9.69000 0.00000 0.53800 6.20850 77.50000 3.20745 5.00000 330.00000 19.05000 391.44000 11.36000 21.20000 0.00000	
> #사분위수 q1, q2, q3, q4	
> apply(Boston[1:15], 2, quantile)	
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT MEDV CAT. MEDV	
0% 0.006320 0.0 0.46 0 0.385 3.5610 2.900 1.1296 1 187 12.60 0.32 1.730 5.000 0	
25% 0.082045 0.0 5.19 0 0.449 5.8855 45.025 2.1002 4 279 17.40 375.38 6.950 17.025 0	
50% 0.256510 0.0 9.69 0 0.538 6.2085 77.500 3.2074 5 330 19.05 391.44 11.360 21.200 0	
75% 3.677083 12.5 18.10 0 0.624 6.6235 94.075 5.1884 24 666 20.20 396.23 16.955 25.000 0	
100% 88.976200 100.0 27.74 1 0.871 8.7800 100.000 12.1265 24 711 22.00 396.90 37.970 50.000 1	
> #사분위수 범위 q3-q1 >> 중간 50%의 데이터에 대한 정보	
> apply(Boston[1:15], 2, IQR)	
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT MEDV CAT. MEDV	
3.5950 12.5000 12.9100 0.0000 0.1750 0.7380 49.0500 3.0883 20.0000 387.0000 2.8000 20.8475 10.0050 7.9750 0.0000	
> #각 변수 데이터의 범위 max - min	
> apply(Boston[1:15], 2, range)	
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT MEDV CAT. MEDV	
[1,] 0.00632 0 0.46 0 0.385 3.561 2.9 1.1296 1 187 12.6 0.32 1.73 5 0	
[2,] 88.97620 100 27.74 1 0.871 8.780 100.0 12.1265 24 711 22.0 396.90 37.97 50 1	
> #표준편차	
> apply(Boston[1:15], 2, sd)	
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT MEDV CAT. MEDV	
8.60155 23.32245 6.86035 0.25399 0.11588 0.70262 28.14886 2.10571 8.70726 168.53712 2.16495 91.29486 7.14106 9.19710 0.37246	
> #분산	
> apply(Boston[1:15], 2, var)	
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT MEDV CAT. MEDV	
7.3987e+01 5.4394e+02 4.7064e+01 6.4513e-02 1.3428e-02 4.9367e-01 7.9236e+02 4.4340e+00 7.5816e+01 2.8405e+04 4.6870e+00 8.3348e+03 5.0995e+01 8.4587e+01 1.3872e-01	

>> 기초통계량을 통해 각 변수별 데이터들이 어떠한 특성을 가지고 있는지 대략적으로 파악 할 수 있다.

평균과 중앙값을 통해 데이터의 모형을 대략적으로 짐작 할 수 있으며 사분위수와 IQR을 통해 데이터가 전체 몇 %부분에 집중되어 있는지 파악할 수 있다. 사분위수를 통해 box도표를 그리는데 활용하며 표준편차와 분산을 통해 퍼짐의 정도를 파악 할 수 있으며 이를 통해 시각화를 진행한다.

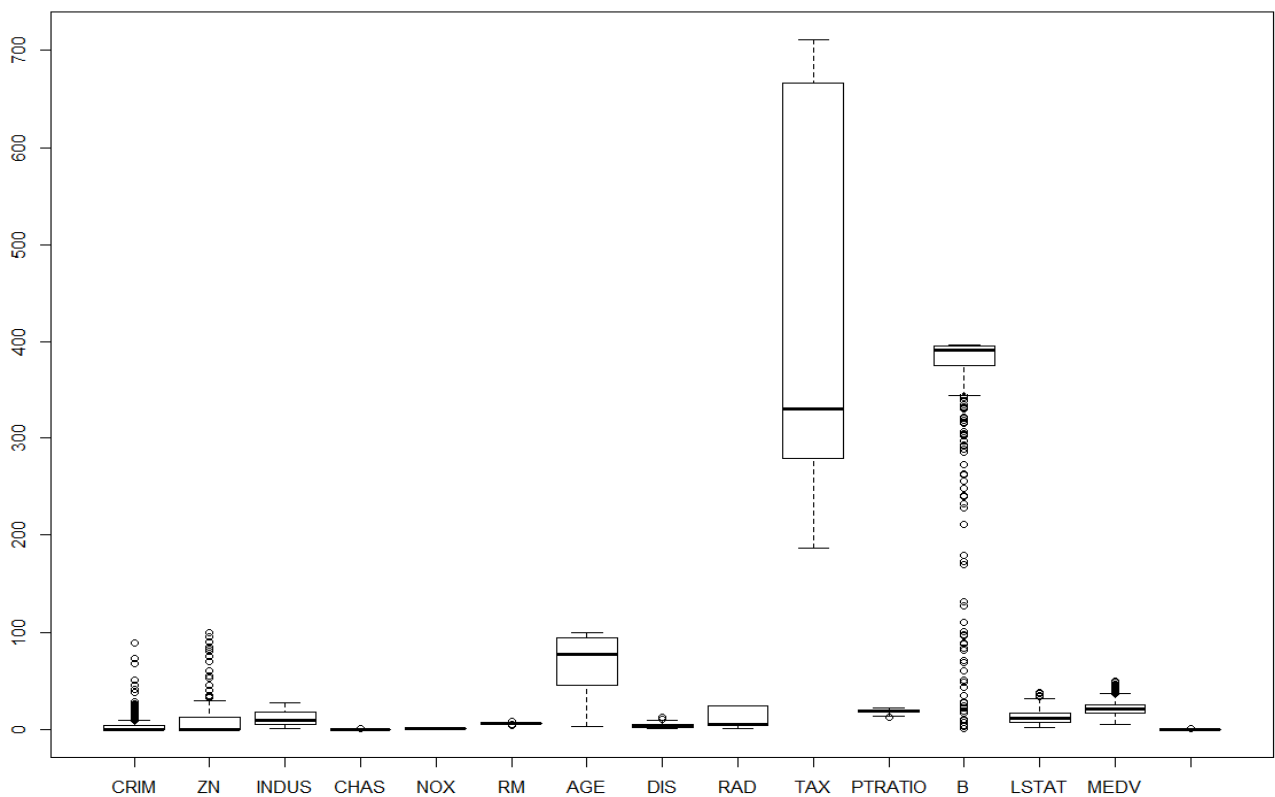
3. BOX-PLOT

각 변수별 Q1, Q2, Q3를 통해 데이터의 밀집 정도 이상점을 파악할 수 있도록 BOX도표를 그린다.

```

Console Terminal x
C:/Users/changwan/Desktop/2018.1/데이터마이닝/
> #박스도표를 그리기 위한 사분위수
> summary(Boston)
      CRIM      ZN      INDUS      CHAS      NOX      RM      AGE      DIS      RAD      TAX
Min.   :0.006  Min.   : 0.0  Min.   : 0.46  Min.   :0.0000  Min.   :0.385  Min.   :3.56  Min.   : 2.9  Min.   : 1.13  Min.   : 1.00  Min.   :187
1st Qu.:0.082  1st Qu.: 0.0  1st Qu.: 5.19  1st Qu.:0.0000  1st Qu.:0.449  1st Qu.:5.89  1st Qu.:45.0  1st Qu.: 2.10  1st Qu.: 4.00  1st Qu.:279
Median :0.257  Median : 0.0  Median : 9.69  Median :0.0000  Median :0.538  Median :6.21  Median :77.5  Median : 3.21  Median : 5.00  Median :330
Mean   :3.614  Mean   :11.4  Mean   :11.14  Mean   :0.0692  Mean   :0.555  Mean   :6.28  Mean   :68.6  Mean   : 3.79  Mean   : 9.55  Mean   :408
3rd Qu.:3.677  3rd Qu.:12.5  3rd Qu.:18.10  3rd Qu.:0.0000  3rd Qu.:0.624  3rd Qu.:6.62  3rd Qu.:94.1  3rd Qu.: 5.19  3rd Qu.:24.00  3rd Qu.:666
Max.   :88.976  Max.  :100.0  Max.   :27.74  Max.   :1.0000  Max.   :0.871  Max.   :8.78  Max.  :100.0  Max.  :12.13  Max.   :24.00  Max.   :711

      PTRATIO      B      LSTAT      MEDV      CAT. MEDV
Min.   :12.6  Min.   : 0.32  Min.   : 1.73  Min.   : 5.0  Min.   :0.000
1st Qu.:17.4  1st Qu.:375.38  1st Qu.: 6.95  1st Qu.:17.0  1st Qu.:0.000
Median :19.1  Median :391.44  Median :11.36  Median :21.2  Median :0.000
Mean   :18.5  Mean   :356.67  Mean   :12.65  Mean   :22.5  Mean   :0.166
3rd Qu.:20.2  3rd Qu.:396.23  3rd Qu.:16.95  3rd Qu.:25.0  3rd Qu.:0.000
Max.   :22.0  Max.   :396.90  Max.   :37.97  Max.   :50.0  Max.   :1.000
  
```



>> 박스도표의 경우 데이터가 한쪽으로 치우쳐져 있는지, 데이터의 밀집정도를 시각적으로 쉽게 파악 할 수 있

다. 일반적인 박스도표의 경우 일정 범위를 넘어가게 되면 이를 이상치 또는 ‘잘못된 데이터’라고도 표현하며 이를 제거하거나 오류를 찾는 과정을 거쳐야 한다. CRIM의 경우 범위를 벗어난 데이터들 중 88.976이라는 데이터를 볼 수 있다. 범칙율이 88.976이라는 것이 잘못 입력된 데이터인지, 실제 데이터 값인지 확인하여 데이터를 분석해야 한다. 이렇듯 이상치를 잘 다룬다면 기본의 기초통계량 값에 영향을 주어 보다 정확한 분석을 진행 할 수 있다.

TAX의 경우 재산세율을 뜻하는데 이는 일반적으로 MEDV(주택가격의 중앙값)을 예측할 수 있는 중요한 변수로 판단된다. 직관적으로 재산세가 낮을수록 주택 가격에는 긍정적인 영향을 끼칠 것이라는 생각을 하게 된다. 박스도표를 보게 되면 50%이하의 범위에 많이 밀집되어 있다. 이는 재산세율이 약 330이하의 전체의 50%를 차지하고 있으며, 이를 MEDV와 어떤 상관관계가 있을지 추가적으로 분석하도록 한다.

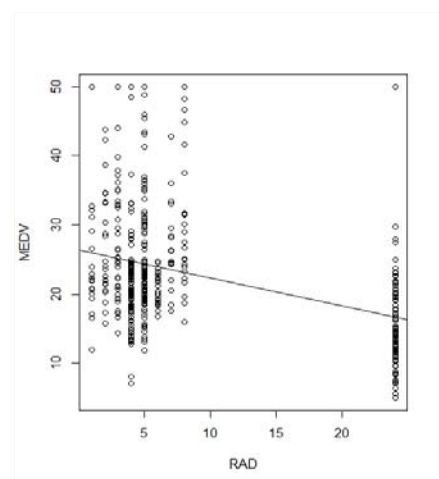
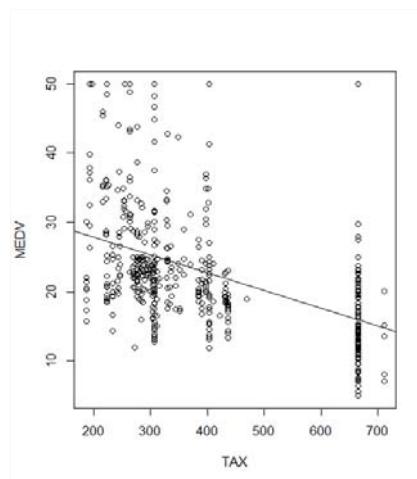
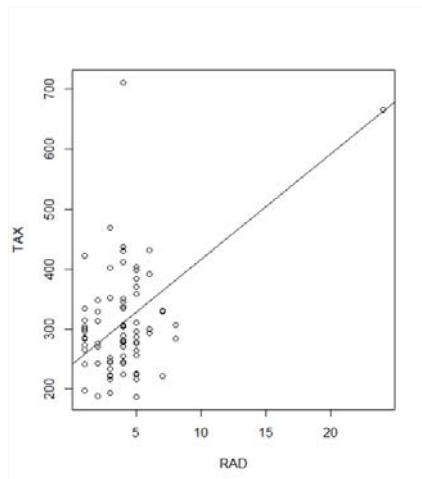
4. 변수들 간의 상관관계

Console	Terminal																																																																																																																																																																																																																																																																																
	C:/Users/changwan/Desktop/2018.1/데이터마이닝/																																																																																																																																																																																																																																																																																
	> #상관관계 파악하기																																																																																																																																																																																																																																																																																
	> cor(Boston[1:15])																																																																																																																																																																																																																																																																																
	<table><tr><th></th><th>CRIM</th><th>ZN</th><th>INDUS</th><th>CHAS</th><th>NOX</th><th>RM</th><th>AGE</th><th>DIS</th><th>RAD</th><th>TAX</th><th>PTRATIO</th><th>B</th><th>LSTAT</th><th>MEDV</th><th>CAT.</th><th>MEDV</th></tr><tr><td>CRIM</td><td>1.000000</td><td>-0.200469</td><td>0.406583</td><td>-0.0558916</td><td>0.420972</td><td>-0.219247</td><td>0.352734</td><td>-0.379670</td><td>0.6255051</td><td>0.582764</td><td>0.28995</td><td>-0.385064</td><td>0.455621</td><td>-0.38830</td><td>-0.15199</td><td></td></tr><tr><td>ZN</td><td>-0.200469</td><td>1.000000</td><td>-0.533828</td><td>-0.0426967</td><td>-0.516604</td><td>0.311991</td><td>-0.569537</td><td>0.664408</td><td>-0.3119478</td><td>-0.314563</td><td>-0.39168</td><td>0.175520</td><td>-0.412995</td><td>0.36045</td><td>0.36530</td><td></td></tr><tr><td>INDUS</td><td>0.406583</td><td>-0.533828</td><td>1.000000</td><td>0.0629380</td><td>0.763651</td><td>-0.391676</td><td>0.644779</td><td>-0.708027</td><td>0.5951293</td><td>0.720760</td><td>0.38325</td><td>-0.356977</td><td>0.603800</td><td>-0.48373</td><td>-0.36628</td><td></td></tr><tr><td>CHAS</td><td>-0.055892</td><td>-0.042697</td><td>0.062938</td><td>1.0000000</td><td>0.091203</td><td>0.091251</td><td>0.086518</td><td>-0.099176</td><td>-0.0073682</td><td>-0.035587</td><td>-0.12152</td><td>0.048788</td><td>-0.053929</td><td>0.17526</td><td>0.10863</td><td></td></tr><tr><td>NOX</td><td>0.420972</td><td>-0.516604</td><td>0.763651</td><td>0.0912028</td><td>1.000000</td><td>-0.302188</td><td>0.731470</td><td>-0.769230</td><td>0.6114406</td><td>0.668023</td><td>0.18893</td><td>-0.380051</td><td>0.590879</td><td>-0.42732</td><td>-0.23250</td><td></td></tr><tr><td>RM</td><td>-0.219247</td><td>0.311991</td><td>-0.391676</td><td>0.0912512</td><td>-0.302188</td><td>1.000000</td><td>-0.240265</td><td>0.205246</td><td>-0.2098467</td><td>-0.292048</td><td>-0.35550</td><td>0.128069</td><td>-0.613808</td><td>0.69536</td><td>0.64127</td><td></td></tr><tr><td>AGE</td><td>0.352734</td><td>-0.569537</td><td>0.644779</td><td>0.0865178</td><td>0.731470</td><td>-0.240265</td><td>1.000000</td><td>-0.747881</td><td>0.4560225</td><td>0.506456</td><td>0.26152</td><td>-0.273534</td><td>0.602339</td><td>-0.37695</td><td>-0.19120</td><td></td></tr><tr><td>DIS</td><td>-0.379670</td><td>0.664408</td><td>-0.708027</td><td>-0.0991758</td><td>-0.769230</td><td>0.205246</td><td>-0.747881</td><td>1.000000</td><td>-0.4945879</td><td>-0.534432</td><td>-0.23247</td><td>0.291512</td><td>-0.496996</td><td>0.24993</td><td>0.11889</td><td></td></tr><tr><td>RAD</td><td>0.625505</td><td>-0.311948</td><td>0.595129</td><td>-0.0073682</td><td>0.611441</td><td>-0.209847</td><td>0.456022</td><td>-0.494588</td><td>1.0000000</td><td>0.910228</td><td>0.46474</td><td>-0.444413</td><td>0.488676</td><td>-0.38163</td><td>-0.19792</td><td></td></tr><tr><td>TAX</td><td>0.582764</td><td>-0.314563</td><td>0.720760</td><td>-0.0355865</td><td>0.668023</td><td>-0.292048</td><td>0.506456</td><td>-0.534432</td><td>0.9102282</td><td>1.000000</td><td>0.46085</td><td>-0.441808</td><td>0.543993</td><td>-0.46854</td><td>-0.27369</td><td></td></tr><tr><td>PTRATIO</td><td>0.289946</td><td>-0.391679</td><td>0.383248</td><td>-0.1215152</td><td>0.188933</td><td>-0.355501</td><td>0.261515</td><td>-0.232471</td><td>0.4647412</td><td>0.460853</td><td>1.00000</td><td>-0.177383</td><td>0.374044</td><td>-0.50779</td><td>-0.44342</td><td></td></tr><tr><td>B</td><td>-0.385064</td><td>0.175520</td><td>-0.356977</td><td>0.0487885</td><td>-0.380051</td><td>0.128069</td><td>-0.273534</td><td>0.291512</td><td>-0.4444128</td><td>-0.441808</td><td>-0.17738</td><td>1.000000</td><td>-0.366087</td><td>0.33346</td><td>0.15514</td><td></td></tr><tr><td>LSTAT</td><td>0.455621</td><td>-0.412995</td><td>0.603800</td><td>-0.0539293</td><td>0.590879</td><td>-0.613808</td><td>0.602339</td><td>-0.496996</td><td>0.4886763</td><td>0.543993</td><td>0.37404</td><td>-0.366087</td><td>1.000000</td><td>-0.73766</td><td>-0.46991</td><td></td></tr><tr><td>MEDV</td><td>-0.388305</td><td>0.360445</td><td>-0.483725</td><td>0.1752602</td><td>-0.427321</td><td>0.695360</td><td>-0.376955</td><td>0.249929</td><td>-0.3816262</td><td>-0.468536</td><td>-0.50779</td><td>0.333461</td><td>-0.737663</td><td>1.00000</td><td>0.78979</td><td></td></tr><tr><td>CAT. MEDV</td><td>-0.151987</td><td>0.365296</td><td>-0.366276</td><td>0.1086312</td><td>-0.232502</td><td>0.641265</td><td>-0.191196</td><td>0.118887</td><td>-0.1979240</td><td>-0.273687</td><td>-0.44342</td><td>0.155137</td><td>-0.469911</td><td>0.78979</td><td>1.00000</td><td></td></tr></table>		CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV	CAT.	MEDV	CRIM	1.000000	-0.200469	0.406583	-0.0558916	0.420972	-0.219247	0.352734	-0.379670	0.6255051	0.582764	0.28995	-0.385064	0.455621	-0.38830	-0.15199		ZN	-0.200469	1.000000	-0.533828	-0.0426967	-0.516604	0.311991	-0.569537	0.664408	-0.3119478	-0.314563	-0.39168	0.175520	-0.412995	0.36045	0.36530		INDUS	0.406583	-0.533828	1.000000	0.0629380	0.763651	-0.391676	0.644779	-0.708027	0.5951293	0.720760	0.38325	-0.356977	0.603800	-0.48373	-0.36628		CHAS	-0.055892	-0.042697	0.062938	1.0000000	0.091203	0.091251	0.086518	-0.099176	-0.0073682	-0.035587	-0.12152	0.048788	-0.053929	0.17526	0.10863		NOX	0.420972	-0.516604	0.763651	0.0912028	1.000000	-0.302188	0.731470	-0.769230	0.6114406	0.668023	0.18893	-0.380051	0.590879	-0.42732	-0.23250		RM	-0.219247	0.311991	-0.391676	0.0912512	-0.302188	1.000000	-0.240265	0.205246	-0.2098467	-0.292048	-0.35550	0.128069	-0.613808	0.69536	0.64127		AGE	0.352734	-0.569537	0.644779	0.0865178	0.731470	-0.240265	1.000000	-0.747881	0.4560225	0.506456	0.26152	-0.273534	0.602339	-0.37695	-0.19120		DIS	-0.379670	0.664408	-0.708027	-0.0991758	-0.769230	0.205246	-0.747881	1.000000	-0.4945879	-0.534432	-0.23247	0.291512	-0.496996	0.24993	0.11889		RAD	0.625505	-0.311948	0.595129	-0.0073682	0.611441	-0.209847	0.456022	-0.494588	1.0000000	0.910228	0.46474	-0.444413	0.488676	-0.38163	-0.19792		TAX	0.582764	-0.314563	0.720760	-0.0355865	0.668023	-0.292048	0.506456	-0.534432	0.9102282	1.000000	0.46085	-0.441808	0.543993	-0.46854	-0.27369		PTRATIO	0.289946	-0.391679	0.383248	-0.1215152	0.188933	-0.355501	0.261515	-0.232471	0.4647412	0.460853	1.00000	-0.177383	0.374044	-0.50779	-0.44342		B	-0.385064	0.175520	-0.356977	0.0487885	-0.380051	0.128069	-0.273534	0.291512	-0.4444128	-0.441808	-0.17738	1.000000	-0.366087	0.33346	0.15514		LSTAT	0.455621	-0.412995	0.603800	-0.0539293	0.590879	-0.613808	0.602339	-0.496996	0.4886763	0.543993	0.37404	-0.366087	1.000000	-0.73766	-0.46991		MEDV	-0.388305	0.360445	-0.483725	0.1752602	-0.427321	0.695360	-0.376955	0.249929	-0.3816262	-0.468536	-0.50779	0.333461	-0.737663	1.00000	0.78979		CAT. MEDV	-0.151987	0.365296	-0.366276	0.1086312	-0.232502	0.641265	-0.191196	0.118887	-0.1979240	-0.273687	-0.44342	0.155137	-0.469911	0.78979	1.00000	
	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV	CAT.	MEDV																																																																																																																																																																																																																																																																	
CRIM	1.000000	-0.200469	0.406583	-0.0558916	0.420972	-0.219247	0.352734	-0.379670	0.6255051	0.582764	0.28995	-0.385064	0.455621	-0.38830	-0.15199																																																																																																																																																																																																																																																																		
ZN	-0.200469	1.000000	-0.533828	-0.0426967	-0.516604	0.311991	-0.569537	0.664408	-0.3119478	-0.314563	-0.39168	0.175520	-0.412995	0.36045	0.36530																																																																																																																																																																																																																																																																		
INDUS	0.406583	-0.533828	1.000000	0.0629380	0.763651	-0.391676	0.644779	-0.708027	0.5951293	0.720760	0.38325	-0.356977	0.603800	-0.48373	-0.36628																																																																																																																																																																																																																																																																		
CHAS	-0.055892	-0.042697	0.062938	1.0000000	0.091203	0.091251	0.086518	-0.099176	-0.0073682	-0.035587	-0.12152	0.048788	-0.053929	0.17526	0.10863																																																																																																																																																																																																																																																																		
NOX	0.420972	-0.516604	0.763651	0.0912028	1.000000	-0.302188	0.731470	-0.769230	0.6114406	0.668023	0.18893	-0.380051	0.590879	-0.42732	-0.23250																																																																																																																																																																																																																																																																		
RM	-0.219247	0.311991	-0.391676	0.0912512	-0.302188	1.000000	-0.240265	0.205246	-0.2098467	-0.292048	-0.35550	0.128069	-0.613808	0.69536	0.64127																																																																																																																																																																																																																																																																		
AGE	0.352734	-0.569537	0.644779	0.0865178	0.731470	-0.240265	1.000000	-0.747881	0.4560225	0.506456	0.26152	-0.273534	0.602339	-0.37695	-0.19120																																																																																																																																																																																																																																																																		
DIS	-0.379670	0.664408	-0.708027	-0.0991758	-0.769230	0.205246	-0.747881	1.000000	-0.4945879	-0.534432	-0.23247	0.291512	-0.496996	0.24993	0.11889																																																																																																																																																																																																																																																																		
RAD	0.625505	-0.311948	0.595129	-0.0073682	0.611441	-0.209847	0.456022	-0.494588	1.0000000	0.910228	0.46474	-0.444413	0.488676	-0.38163	-0.19792																																																																																																																																																																																																																																																																		
TAX	0.582764	-0.314563	0.720760	-0.0355865	0.668023	-0.292048	0.506456	-0.534432	0.9102282	1.000000	0.46085	-0.441808	0.543993	-0.46854	-0.27369																																																																																																																																																																																																																																																																		
PTRATIO	0.289946	-0.391679	0.383248	-0.1215152	0.188933	-0.355501	0.261515	-0.232471	0.4647412	0.460853	1.00000	-0.177383	0.374044	-0.50779	-0.44342																																																																																																																																																																																																																																																																		
B	-0.385064	0.175520	-0.356977	0.0487885	-0.380051	0.128069	-0.273534	0.291512	-0.4444128	-0.441808	-0.17738	1.000000	-0.366087	0.33346	0.15514																																																																																																																																																																																																																																																																		
LSTAT	0.455621	-0.412995	0.603800	-0.0539293	0.590879	-0.613808	0.602339	-0.496996	0.4886763	0.543993	0.37404	-0.366087	1.000000	-0.73766	-0.46991																																																																																																																																																																																																																																																																		
MEDV	-0.388305	0.360445	-0.483725	0.1752602	-0.427321	0.695360	-0.376955	0.249929	-0.3816262	-0.468536	-0.50779	0.333461	-0.737663	1.00000	0.78979																																																																																																																																																																																																																																																																		
CAT. MEDV	-0.151987	0.365296	-0.366276	0.1086312	-0.232502	0.641265	-0.191196	0.118887	-0.1979240	-0.273687	-0.44342	0.155137	-0.469911	0.78979	1.00000																																																																																																																																																																																																																																																																		

>> 상관관계는 각 변수들 간의 ‘양의 상관’, ‘음의 상관’, ‘관계없음’을 보여줄 수 있는 데이터이다. 1에 가까울수록 둘의 상관관계는 강하다고 말 할 수 있으며 -1에 가까울수록 강한 음의 상관관계라고 말할 수 있다.

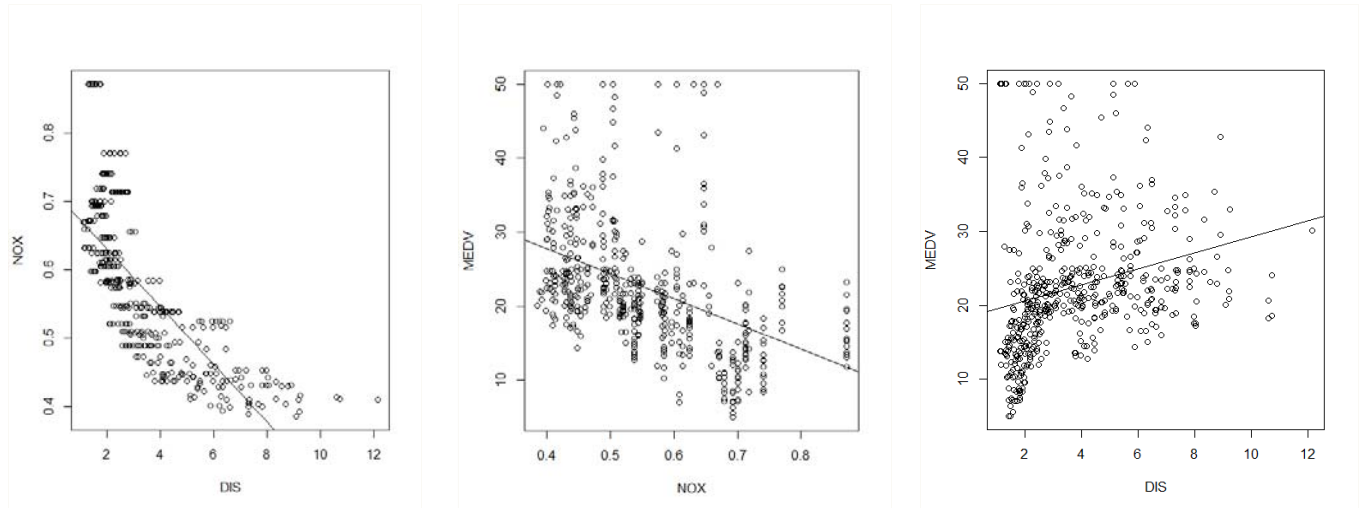
① 가장 강한 상관관계(TAX & RAD)

: 데이터 중 TAX(재산세율(\$10,000당), RAD(고속도로 인접 가능 INDEX)는 ‘0.9102’로 가장 강한 상관관계를 보이고 있다. 하지만 주택 가격과는 각각 음의 상관관계를 띄고 있으므로 두 변수의 수치가 낮은 값을 가질수록 주택 가격에는 긍정적인 영향을 끼치는 것을 알 수 있으며 앞서 박스플롯에서 TAX와 MEDV는 음의 상관관계가 있을 것이라는 생각과 일치하는 것을 알 수 있다.



② 강한 음의 상관관계(DIS & NOX)

: 두 데이터는 '-0.769230'로 변수들 중 가장 강한 음의 상관관계를 나타내고 있다. DIS(보스턴 대표 5개의 회사와 거리에 따른 가중 평균)가 멀어질수록 NOX(이산화질소 농도(백만 개당))는 떨어진다고 판단 할 수 있으며 이는 직관적으로도 인식이 가능하다. 이를 주택 가격과 비교하였을 때 NOX가 높아질수록 가격은 하락하는 것을 확인 할 수 있으며 DIS가 커질수록 주택가격에는 긍정적인 영향을 끼치는 것을 확인 할 수 있다.



③ 두 변수간의 상관관계가 없다고 할 수 있다. (RAD & CHAS)

: RAD(고속도로 인접 가능 INDEX)와 CHAS(찰스강의 인접 여부) 두 변수간의 상관계수는 '-0.0073682'로 거의 0에 가깝다. 이는 두 변수간의 상호 연관성이 없다고 말할 수 있다. 또한 CHAS 변수는 다른 변수들의 상관관계를 봐도 거의 0에 근접한 값을 보이고 있으며 이는 CHAS와 각 변수들 간의 연관성이 없을 뿐 아니라 주택 가격에도 영향을 거의 주지 않는 것을 알 수 있다. 이와 같이 연관성이 거의 없는 변수의 경우 제거하고 추가적으로 변수들 간의 상관관계를 보는 과정이 필요로 하다.

5. 결론

- 주택 가격을 결정하는데 여러 변수가 작용한다는 것을 볼 수 있었다. 강한 상관관계를 띄는 변수, 관계가 거의 없는 변수 등 데이터 값을 통해 확인 할 수 있었는데 강한 상관관계를 가진 변수들도 실제 주택가격에는 상관관계를 보이지 않은 변수들과 반대되는 상관관계들이 존재하는 것도 볼 수 있었다. 이처럼 변수들의 수가 많음에 따라 주택가격을 정확히 예측하는 것 자체가 어렵고 정확도 또한 불확실하다. 이러한 경우 변수들 간의 특성을 고려하여 군집화하고 불필요한 데이터 제거하여 분석한다면 주택 가격을 예측하는데 도움을 줄 수 있을 것이라 생각한다.