Lab_03

Abhinav Garg

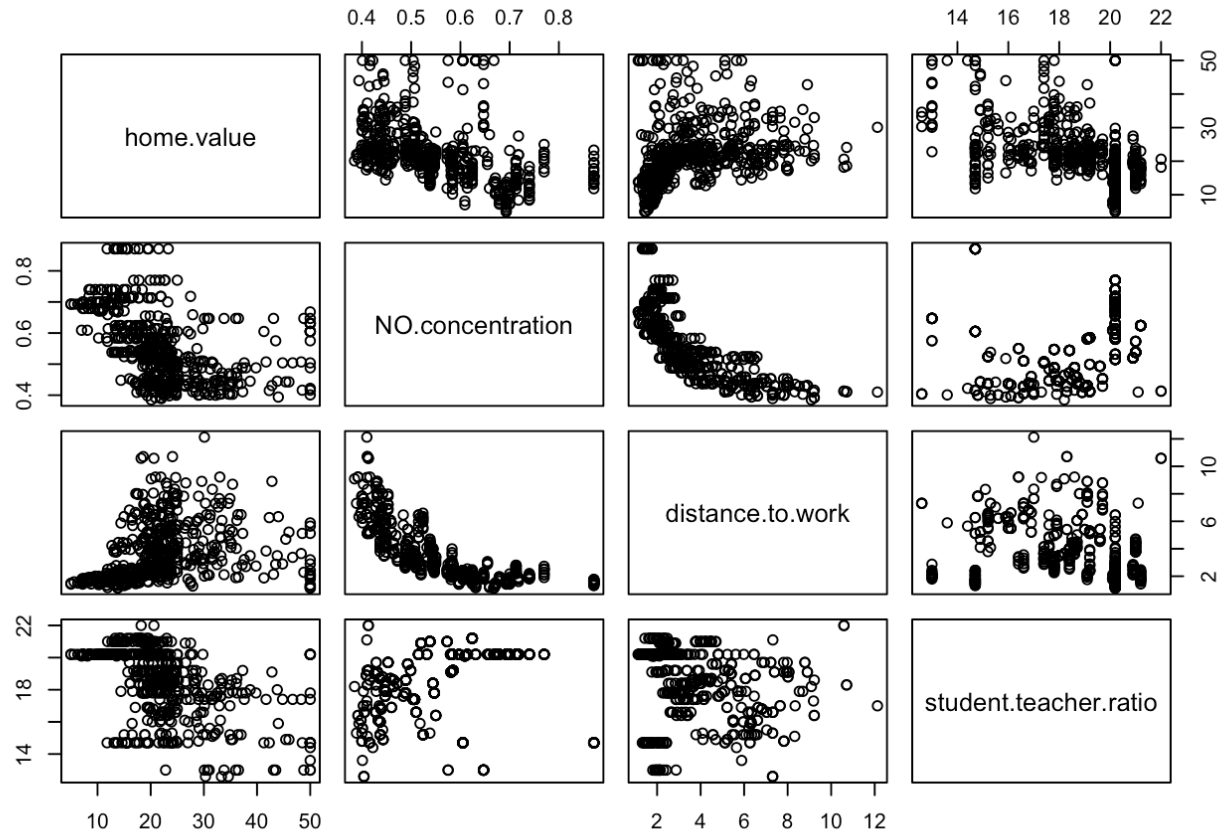
2/9/2017

Loading Data

```
load("BostonData.Rdat")
```

Plot Data

```
plot(boston)
```



0)It looks like the variables distance.to.work and NO.concentration are tightly related an increase in nitrogen oxide could mean you are closer to work. Also Nitrogen Oxide concentration and home value are related closely as increase in NO concentrtion would mean a lower home value. The increase in student-teacher ratio also shows a decrease in home value.

Fitting Linear Regression

```
mod1 <- lm(home.value ~ NO.concentration, data = boston)
summary(mod1)
```

```
##
## Call:
## lm(formula = home.value ~ NO.concentration, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.691  -5.121  -2.161   2.959  31.310
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        41.346      1.811   22.83   <2e-16 ***
## NO.concentration  -33.916      3.196  -10.61   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.323 on 504 degrees of freedom
## Multiple R-squared:  0.1826, Adjusted R-squared:  0.181
## F-statistic: 112.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

1 a)We can see that with an increase in 1 unit of Nitrogen Oxide concentration, There is a decrease in the home-value which is reasonable. As a higher concentration of Nitrogen would cause a hazardous environment and that in turn would reduce the home values

```
mod2 <- lm(home.value ~ distance.to.work, data = boston)
summary(mod2)
```

```
##
## Call:
## lm(formula = home.value ~ distance.to.work, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.016  -5.556  -1.865   2.288  30.377
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       18.3901     0.8174  22.499  < 2e-16 ***
## distance.to.work   1.0916     0.1884   5.795 1.21e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.914 on 504 degrees of freedom
## Multiple R-squared:  0.06246,    Adjusted R-squared:  0.0606
## F-statistic: 33.58 on 1 and 504 DF,  p-value: 1.207e-08
```

1 b)We can see that with an increase in 1 unit of distance, There is an increase in the home-value which is reasonable. Because if you're closer to your work place you spend less distance travelling means that the homes near the work place aren't valued highly. Homes away from the work place could have higher value. Also since the p-value for this variable is closer to 0, we can't say that this association by chance.

```
mod3 <- lm(home.value ~ student.teacher.ratio, data = boston)
summary(mod3)
```

```
##
## Call:
## lm(formula = home.value ~ student.teacher.ratio, data = boston)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.8342  -4.8262  -0.6426   3.1571  31.2303
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             62.345      3.029   20.58   <2e-16 ***
## student.teacher.ratio   -2.157      0.163  -13.23   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.931 on 504 degrees of freedom
## Multiple R-squared:  0.2578, Adjusted R-squared:  0.2564
## F-statistic: 175.1 on 1 and 504 DF,  p-value: < 2.2e-16
```

1 c) The more student teacher ratio, the more the number of students and thus the spending power for that particular region would be less as students don't spend as much on homes.

2. We see that the adjusted R-squared for the mod3 is the highest and thus explains the data best in comparison to the other two models

```
mod.full <- lm(home.value ~ distance.to.work + NO.concentration + student.teacher.
ratio, data = boston)
summary(mod.full)
```

```
##
## Call:
## lm(formula = home.value ~ distance.to.work + NO.concentration +
##     student.teacher.ratio, data = boston)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -12.434  -4.931  -1.270   2.951  32.606
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             89.0255     4.2358  21.017  < 2e-16 ***
## distance.to.work        -1.2803     0.2374  -5.393 1.07e-07 ***
## NO.concentration       -44.7740     4.2729 -10.479  < 2e-16 ***
## student.teacher.ratio   -1.9939     0.1503 -13.270  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.109 on 502 degrees of freedom
## Multiple R-squared:  0.4061, Adjusted R-squared:  0.4026
## F-statistic: 114.4 on 3 and 502 DF,  p-value: < 2.2e-16
```

3. Here we can say that with a unit increase in Nitrogen Oxide the home value decreases by 44 units, also since there is an increase in Nitrogen Oxide the further you are from work increases the home value and the increase in student-teacher ratio decreases the homevalue

4. Comaparing the R-squared with the single variable models we see that the R-squared is higher, hence the combined model with all three variables explains the data much better.

```
predict(mod.full, newdata=data.frame("distance.to.work" = 3,
"NO.concentration" = 0.35, "student.teacher.ratio" = 10),
interval="prediction")
```

```
##        fit      lwr      upr
## 1 49.57499 35.22737 63.92261
```

So as we can see for the data point where distance to work is 3 units, Nitrogen Oxide is 0.35 units and student teacher ration is 10 units, the predicted home value would be approximately between 35227 and 63922 $