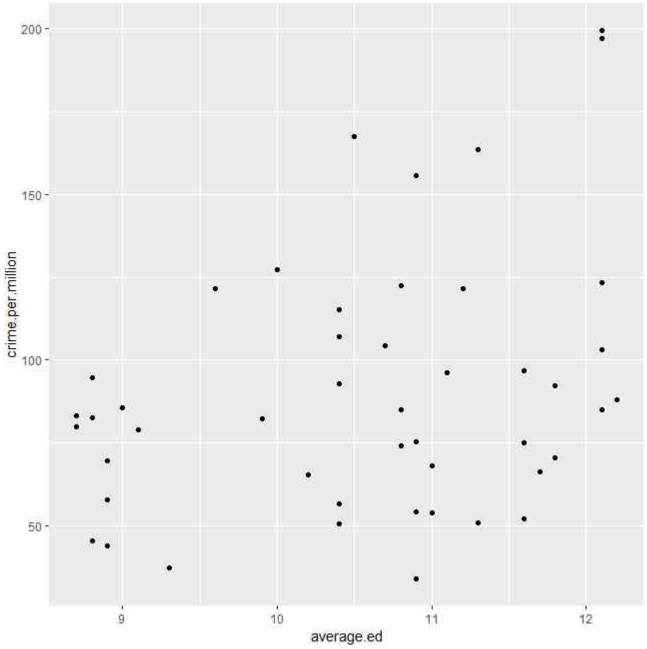
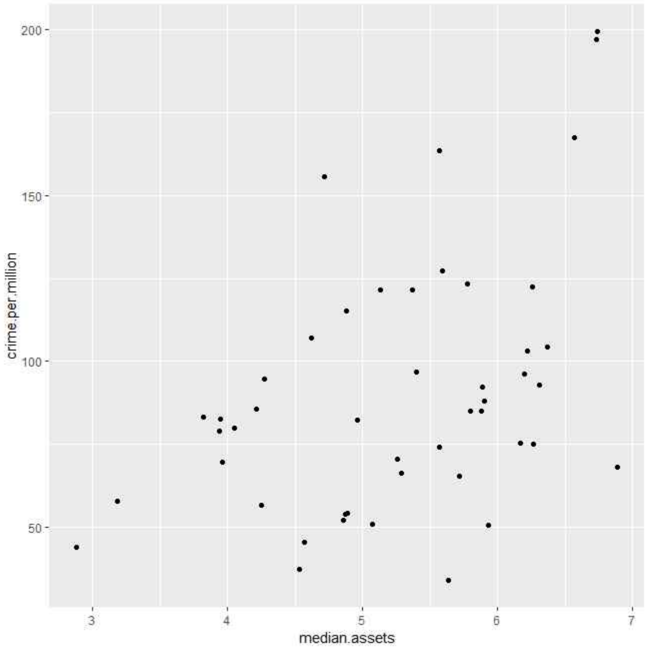
데이터 사이언스 과제7

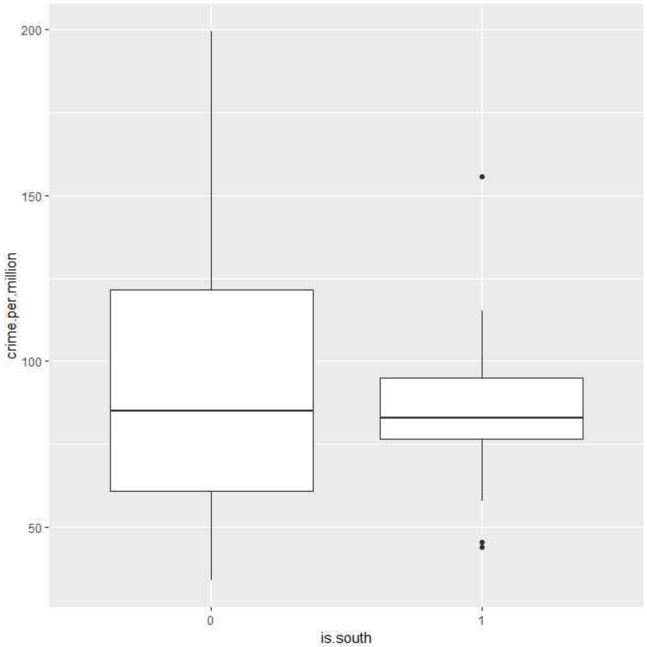
```
< Linear Regression >

> #1. Packages
> library(MASS)
> library(plyr)
> library(ggplot2)
> library(knitr)
> library(GGally)
>
>
>
> #2. Linear regression
> # Import data set
> crime <-
read.table("http://www.andrew.cmu.edu/user/achoulde/94842/data/crime_si
mple.txt", sep = "\t",
+ header = TRUE)
> # Assign more meaningful variable names
> colnames(crime) <- c("crime.per.million", "young.males", "is.south",
"average.ed",
+   "exp.per.cap.1960", "exp.per.cap.1959", "labour.part",
+   "male.per.fem", "population", "nonwhite",
+   "unemp.youth", "unemp.adult", "median.assets", "num.low.salary")
> # Convert is.south to a factor
> # Divide average.ed by 10 so that the variable is actually average
education
> # Convert median assets to 1000's of dollars instead of 10's
> crime <- transform(crime, is.south = as.factor(is.south),
+   average.ed = average.ed / 10,
+   median.assets = median.assets / 100)
> # print summary of the data
> summary(crime)
 crime.per.million  young.males     is.south    average.ed
exp.per.cap.1960
 Min.    : 34.20     Min.    :119.0    0:31       Min.    : 8.70    Min.    : 45.0
 1st Qu.: 65.85     1st Qu.:130.0    1:16       1st Qu.: 9.75    1st Qu.: 62.5
 Median : 83.10      Median :136.0                 Median :10.80     Median : 78.0

 Mean    : 90.51      Mean    :138.6                 Mean    :10.56    Mean    :
85.0
 3rd Qu.:105.75      3rd Qu.:146.0                3rd Qu.:11.45    3rd Qu.:104.5
 Max.    :199.30      Max.    :177.0                Max.    :12.20    Max.    :166.0
 exp.per.cap.1959  labour.part     male.per.fem       population
 Min.    : 41.00    Min.    :480.0    Min.    : 934.0    Min.    :   3.00
 1st Qu.: 58.50    1st Qu.:530.5    1st Qu.: 964.5    1st Qu.: 10.00
 Median : 73.00     Median :560.0    Median : 977.0     Median : 25.00
 Mean    : 80.23     Mean    :561.2    Mean    : 983.0    Mean    : 36.62
 3rd Qu.: 97.00    3rd Qu.:593.0    3rd Qu.: 992.0    3rd Qu.: 41.50
 Max.    :157.00    Max.    :641.0    Max.    :1071.0    Max.    :168.00
    nonwhite         unemp.youth          unemp.adult       median.assets
 Min.    :  2.0    Min.    : 70.00    Min.    :20.00    Min.    :2.880
 1st Qu.: 24.0    1st Qu.: 80.50    1st Qu.:27.50    1st Qu.:4.595
 Median : 76.0     Median : 92.00     Median :34.00     Median :5.370
 Mean    :101.1     Mean    : 95.47     Mean    :33.98     Mean    :5.254
 3rd Qu.:132.5    3rd Qu.:104.00    3rd Qu.:38.50    3rd Qu.:5.915
 Max.    :423.0    Max.    :142.00    Max.    :58.00    Max.    :6.890
 num.low.salary
 Min.    :126.0
 1st Qu.:165.5
 Median :176.0
 Mean    :194.0
 3rd Qu.:227.5
 Max.    :276.0
>
> # Scatter plot of outcome (crime.per.million) against average.ed
> qplot(average.ed, crime.per.million, data = crime)
```



```
> # correlation between education and crime
> with(crime, cor(average.ed, crime.per.million))
[1] 0.3228349
> # Scatter plot of outcome (crime.per.million) against median.assets
> qplot(median.assets, crime.per.million, data = crime)
```



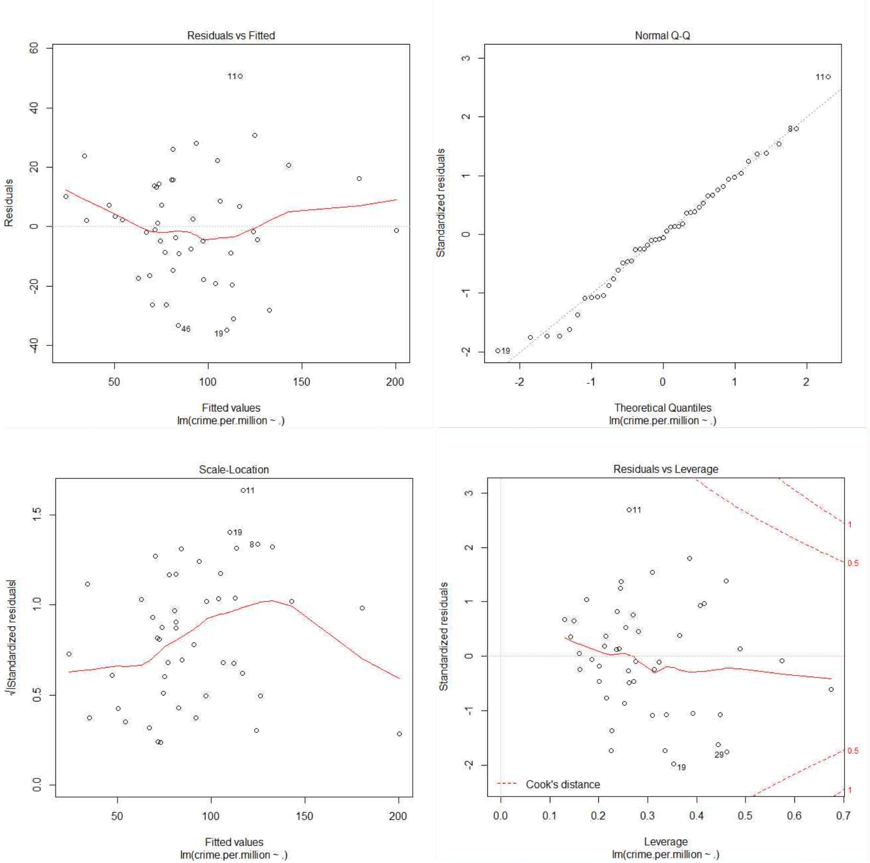
```
> # correlation between education and crime
> with(crime, cor(median.assets, crime.per.million))
[1] 0.4413199
>
> # Boxplots showing crime rate broken down by southern vs
non-southern state
> qplot(is.south, crime.per.million, geom = "boxplot", data = crime)
```



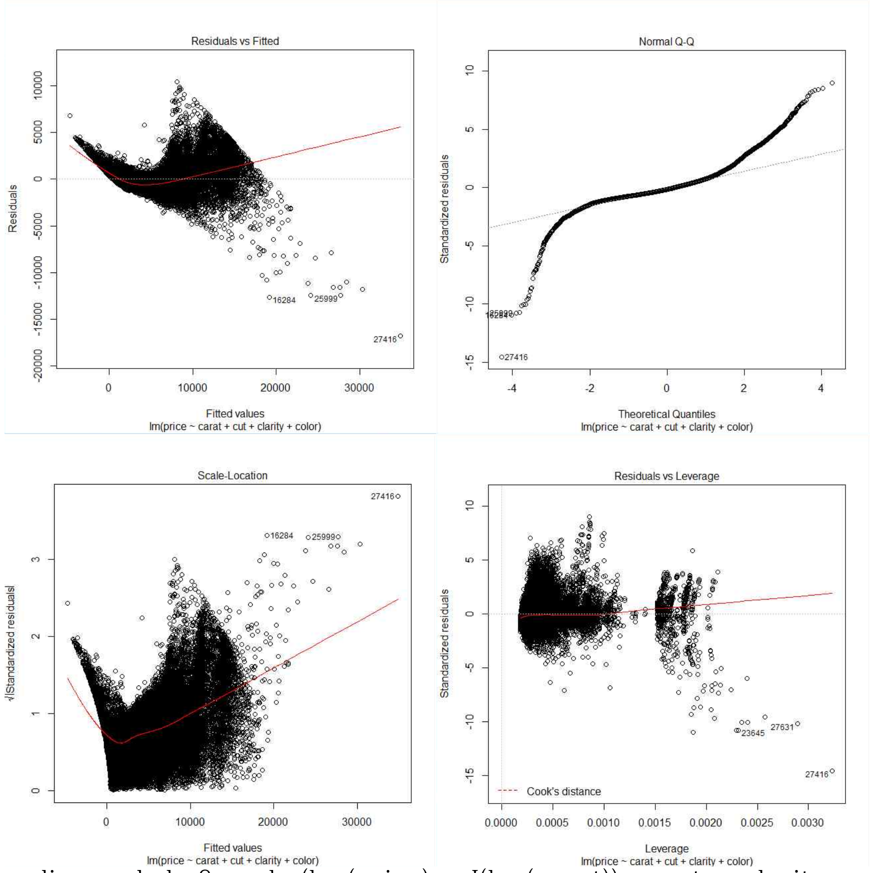
```
> crime.lm <- lm(crime.per.million ~ ., data = crime)
> # Summary of the linear regression model
> crime.lm

Call:
lm(formula = crime.per.million ~ ., data = crime)

Coefficients:
      (Intercept)        young.males           is.south1         average.ed
       -6.918e+02          1.040e+00          -8.308e+00          1.802e+01
exp.per.cap.1960   exp.per.cap.1959         labour.part       male.per.fem
        1.608e+00         -6.673e-01          -4.103e-02          1.648e-01
       population           nonwhite         unemp.youth        unemp.adult
       -4.128e-02          7.175e-03          -6.017e-01          1.792e+00
    median.assets     num.low.salary
        1.374e+01          7.929e-01

> summary(crime.lm)

Call:
lm(formula = crime.per.million ~ ., data = crime)

Residuals:
    Min       1Q   Median       3Q      Max
-34.884  -11.923  -1.135   13.495   50.560

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -6.918e+02  1.559e+02  -4.438 9.56e-05 ***
```

```
young.males         1.040e+00  4.227e-01   2.460  0.01931 *
is.south1          -8.308e+00  1.491e+01  -0.557  0.58117
average.ed          1.802e+01  6.497e+00   2.773  0.00906 **
exp.per.cap.1960    1.608e+00  1.059e+00   1.519  0.13836
exp.per.cap.1959   -6.673e-01  1.149e+00  -0.581  0.56529
labour.part        -4.103e-02  1.535e-01  -0.267  0.79087
male.per.fem        1.648e-01  2.099e-01   0.785  0.43806
population         -4.128e-02  1.295e-01  -0.319  0.75196
nonwhite            7.175e-03  6.387e-02   0.112  0.91124
unemp.youth        -6.017e-01  4.372e-01  -1.376  0.17798
unemp.adult         1.792e+00  8.561e-01   2.093  0.04407 *
median.assets       1.374e+01  1.058e+01   1.298  0.20332
num.low.salary      7.929e-01  2.351e-01   3.373  0.00191 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.94 on 33 degrees of freedom
Multiple R-squared:  0.7692,	Adjusted R-squared:  0.6783
F-statistic: 8.462 on 13 and 33 DF,  p-value: 3.686e-07

>
> options(scipen=4) # Set scipen = 0 to get back to default
> summary(crime.lm)

Call:
lm(formula = crime.per.million ~ ., data = crime)

Residuals:
    Min      1Q  Median      3Q     Max
-34.884 -11.923  -1.135  13.495  50.560

Coefficients:
                    Estimate  Std. Error t value  Pr(>|t|)
(Intercept)      -691.837588  155.887918  -4.438 0.0000956 ***
young.males         1.039810    0.422708   2.460   0.01931 *
is.south1          -8.308313   14.911588  -0.557   0.58117
average.ed         18.016011    6.496504   2.773   0.00906 **
exp.per.cap.1960    1.607818    1.058667   1.519   0.13836
exp.per.cap.1959   -0.667258    1.148773  -0.581   0.56529
labour.part        -0.041031    0.153477  -0.267   0.79087
male.per.fem        0.164795    0.209932   0.785   0.43806
population         -0.041277    0.129516  -0.319   0.75196
nonwhite            0.007175    0.063867   0.112   0.91124
unemp.youth        -0.601675    0.437154  -1.376   0.17798
unemp.adult         1.792263    0.856111   2.093   0.04407 *
median.assets      13.735847   10.583028   1.298   0.20332
num.low.salary      0.792933    0.235085   3.373   0.00191 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.94 on 33 degrees of freedom
Multiple R-squared:  0.7692,	Adjusted R-squared:  0.6783
F-statistic: 8.462 on 13 and 33 DF,  p-value: 0.0000003686
```

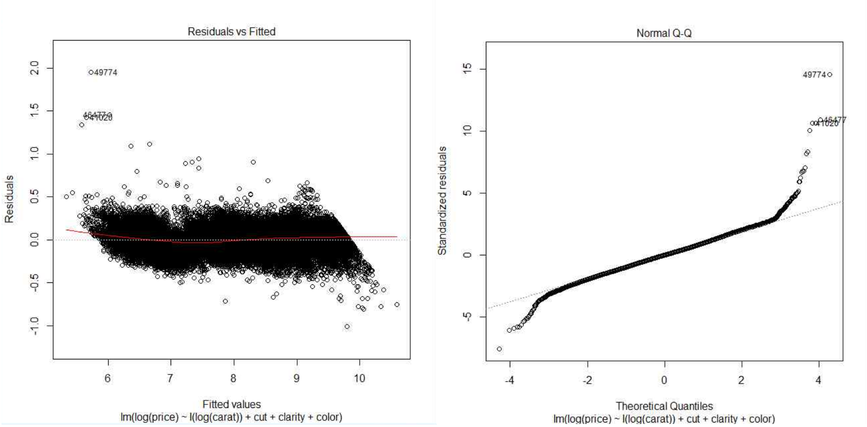
```
>
> # List all attributes of the linear model
> attributes(crime.lm)
$`names`
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "contrasts"     "xlevels"       "call"          "terms"
[13] "model"

$class
[1] "lm"

> crime.lm$coef
     (Intercept)      young.males        is.south1       average.ed
   -691.837587905      1.039809653     -8.308312889     18.016010601
exp.per.cap.1960 exp.per.cap.1959      labour.part     male.per.fem
      1.607818377     -0.667258285     -0.041031047      0.164794968
       population         nonwhite      unemp.youth      unemp.adult
     -0.041276887      0.007174688     -0.601675298      1.792262901
    median.assets   num.low.salary
     13.735847285      0.792932786
>
> # Pull coefficients element from summary(lm) object
> round(summary(crime.lm)$coef, 3)
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -691.838    155.888  -4.438    0.000
young.males         1.040      0.423   2.460    0.019
is.south1          -8.308     14.912  -0.557    0.581
average.ed         18.016      6.497   2.773    0.009
exp.per.cap.1960    1.608      1.059   1.519    0.138
exp.per.cap.1959   -0.667      1.149  -0.581    0.565
labour.part        -0.041      0.153  -0.267    0.791
male.per.fem        0.165      0.210   0.785    0.438
population         -0.041      0.130  -0.319    0.752
nonwhite            0.007      0.064   0.112    0.911
unemp.youth        -0.602      0.437  -1.376    0.178
unemp.adult         1.792      0.856   2.093    0.044
median.assets      13.736     10.583   1.298    0.203
num.low.salary      0.793      0.235   3.373    0.002
>
> # Pull the coefficients table from summary(lm)
> crime.lm.coef <- round(summary(crime.lm)$coef, 3)
> # See what this gives
> class(crime.lm.coef)
[1] "matrix"
> attributes(crime.lm.coef)
$`dim`
[1] 14  4

$dimnames
$dimnames[[1]]
 [1] "(Intercept)"      "young.males"      "is.south1"
 [4] "average.ed"          "exp.per.cap.1960"  "exp.per.cap.1959"
 [7] "labour.part"         "male.per.fem"       "population"
[10] "nonwhite"             "unemp.youth"          "unemp.adult"
[13] "median.assets"       "num.low.salary"

$dimnames[[2]]
[1] "Estimate"    "Std. Error" "t value"     "Pr(>|t|)"


> crime.lm.coef["average.ed", "Pr(>|t|)"]
[1] 0.009
>
> plot(crime.lm)
```

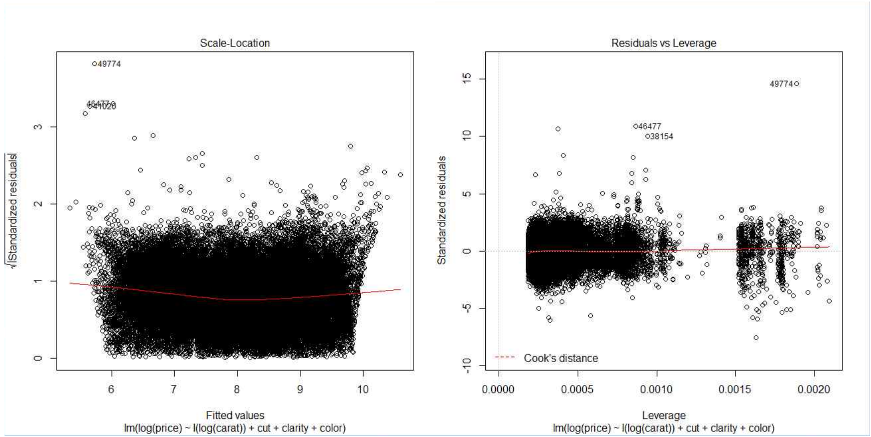


```
> diamonds.lm <- lm(price ~ carat + cut + clarity + color, data =
diamonds)
> plot(diamonds.lm)
```

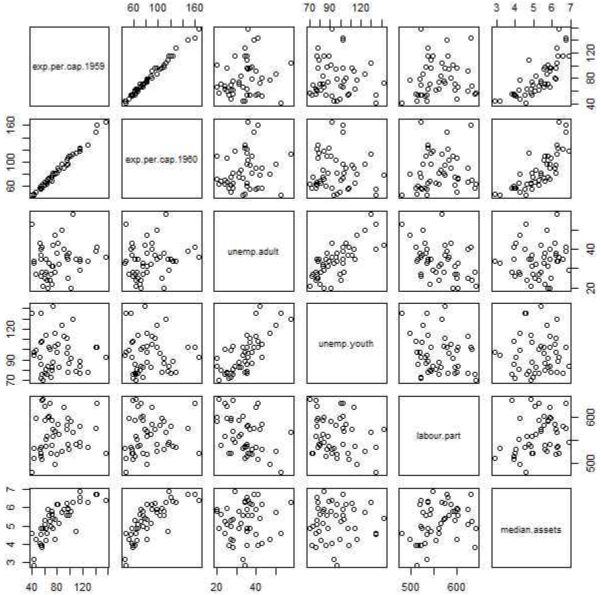


```
> diamonds.lm2 <- lm(log(price) ~ I(log(carat)) + cut + clarity + color, data
= diamonds)
> plot(diamonds.lm2)
```

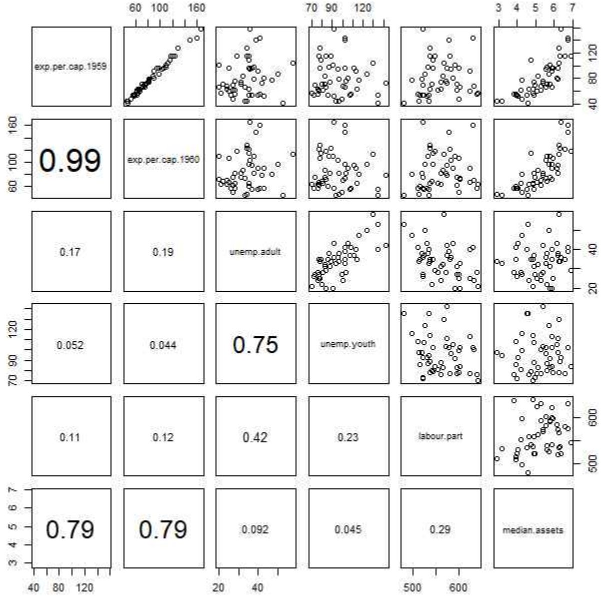




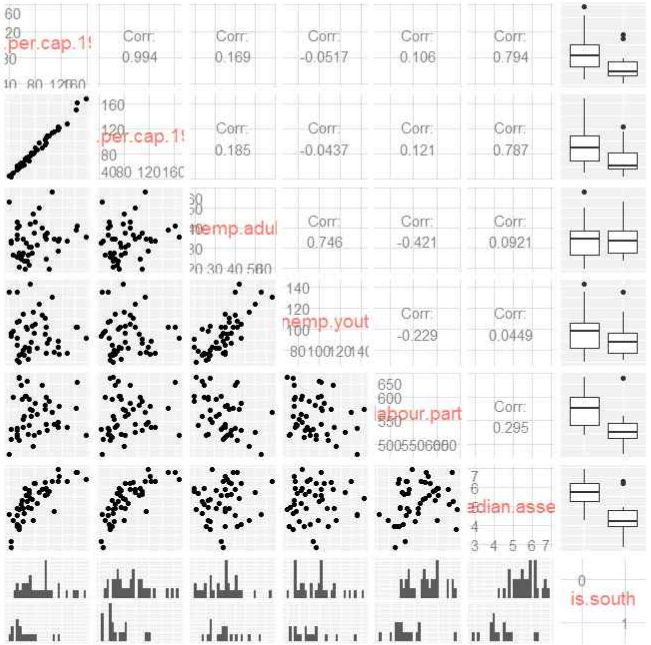
```
> economic.var.names <- c("exp.per.cap.1959", "exp.per.cap.1960",
"unemp.adult",
+ "unemp.youth", "labour.part", "median.assets")
> pairs(crime[,economic.var.names])
```



```
> round(cor(crime[,economic.var.names]), 3)
                exp.per.cap.1959 exp.per.cap.1960 unemp.adult
unemp.youth
exp.per.cap.1959           1.000            0.994       0.169      -0.052
exp.per.cap.1960           0.994            1.000       0.185      -0.044
unemp.adult                0.169            0.185       1.000       0.746
unemp.youth               -0.052           -0.044       0.746       1.000
labour.part                0.106            0.121      -0.421      -0.229
median.assets              0.794            0.787       0.092       0.045
                labour.part median.assets
exp.per.cap.1959       0.106         0.794
exp.per.cap.1960       0.121         0.787
unemp.adult           -0.421         0.092
unemp.youth           -0.229         0.045
labour.part            1.000         0.295
median.assets          0.295         1.000
>
> # Function taken from ?pairs Example section.
> panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
+ {
+   usr <- par("usr"); on.exit(par(usr))
+   par(usr = c(0, 1, 0, 1))
+   r <- abs(cor(x, y))
+   txt <- format(c(r, 0.123456789), digits = digits)[1]
+   txt <- paste0(prefix, txt)
+   if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
+   text(0.5, 0.5, txt, cex = pmax(1, cex.cor * r))
+ }
> # Use panel.cor to display correlations in lower panel.
> pairs(crime[,economic.var.names], lower.panel = panel.cor)
```



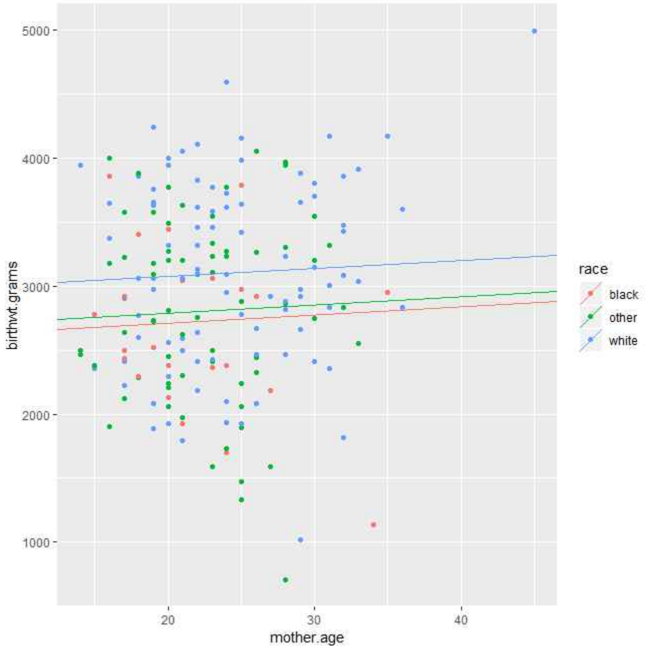
```
> # ggpairs from GGally library
> # Unlike pairs(), ggpairs() works with non-numeric
> # predictors in addition to numeric ones.
> # Consider ggpairs() for your final project
> ggpairs(crime[,c(economic.var.names, "is.south")], axisLabels = "internal")
```



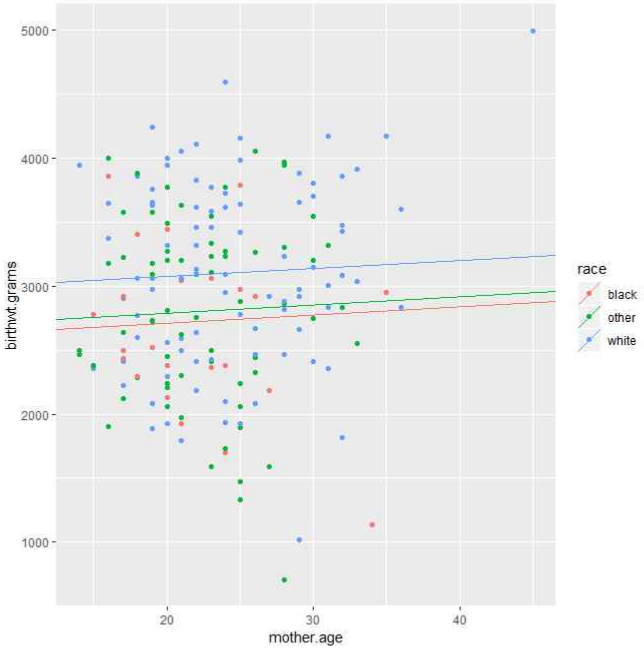
```
> crime.lm.2 <- update(crime.lm, . ~ . - exp.per.cap.1959 - unemp.youth)
> summary(crime.lm.2)

Call:
lm(formula = crime.per.million ~ young.males + is.south + average.ed +
    exp.per.cap.1960 + labour.part + male.per.fem + population +
    nonwhite + unemp.adult + median.assets + num.low.salary,
    data = crime)

Residuals:
   Min     1Q Median     3Q    Max
-35.82 -11.57  -1.51  10.63  55.02

Coefficients:
                   Estimate  Std. Error t value  Pr(>|t|)
(Intercept)     -633.438828  145.470340  -4.354  0.000111 ***
young.males        1.126883    0.418791   2.691  0.010853 *
is.south1         -0.556600   13.883248  -0.040  0.968248
average.ed        15.328028    6.202516   2.471  0.018476 *
exp.per.cap.1960   1.138299    0.226977   5.015 0.0000153 ***
labour.part        0.068716    0.133540   0.515  0.610087
male.per.fem       0.003021    0.173041   0.017  0.986172
population        -0.064477    0.128278  -0.503  0.618367
nonwhite          -0.013794    0.061901  -0.223  0.824960
unemp.adult        0.931498    0.541803   1.719  0.094402 .
median.assets     15.158975   10.524458   1.440  0.158653
num.low.salary     0.825936    0.234189   3.527  0.001197 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.98 on 35 degrees of freedom
Multiple R-squared:  0.7543,    Adjusted R-squared:  0.6771
F-statistic: 9.769 on 11 and 35 DF,  p-value: 0.00000009378

> crime.lm.summary.2 <- summary(crime.lm.2)
>
> kable(crime.lm.summary.2$coef, digits = c(3, 3, 3, 4), format =
'markdown')


|                 | Estimate| Std. Error| t value| Pr(>&#124;t&#124;)|
|:----------------|--------:|----------:|-------:|------------------:|
|(Intercept)      | -633.439|    145.470|  -4.354|             0.0001|
|young.males      |    1.127|      0.419|   2.691|             0.0109|
|is.south1        |   -0.557|     13.883|  -0.040|             0.9682|
|average.ed       |   15.328|      6.203|   2.471|             0.0185|
|exp.per.cap.1960 |    1.138|      0.227|   5.015|             0.0000|
|labour.part      |    0.069|      0.134|   0.515|             0.6101|
|male.per.fem     |    0.003|      0.173|   0.017|             0.9862|
|population       |   -0.064|      0.128|  -0.503|             0.6184|
|nonwhite         |   -0.014|      0.062|  -0.223|             0.8250|
|unemp.adult      |    0.931|      0.542|   1.719|             0.0944|
|median.assets    |   15.159|     10.524|   1.440|             0.1587|
|num.low.salary   |    0.826|      0.234|   3.527|             0.0012|
>
>
>
> #3. Thinking more critically about linear regression
> crime.lm <- lm(crime.per.million ~ ., data = crime)
> crime.lm2 <- update(crime.lm, . ~ . - exp.per.cap.1959 - unemp.youth)
>
> kable(summary(crime.lm)$coef, digits = c(3, 3, 3, 4), format =
'markdown')


|                 | Estimate| Std. Error| t value| Pr(>&#124;t&#124;)|
|:----------------|--------:|----------:|-------:|------------------:|
|(Intercept)      | -691.838|    155.888|  -4.438|             0.0001|
|young.males      |    1.040|      0.423|   2.460|             0.0193|
|is.south1        |   -8.308|     14.912|  -0.557|             0.5812|
|average.ed       |   18.016|      6.497|   2.773|             0.0091|
|exp.per.cap.1960 |    1.608|      1.059|   1.519|             0.1384|
|exp.per.cap.1959 |   -0.667|      1.149|  -0.581|             0.5653|
|labour.part      |   -0.041|      0.153|  -0.267|             0.7909|
|male.per.fem     |    0.165|      0.210|   0.785|             0.4381|
|population       |   -0.041|      0.130|  -0.319|             0.7520|
|nonwhite         |    0.007|      0.064|   0.112|             0.9112|
|unemp.youth      |   -0.602|      0.437|  -1.376|             0.1780|
|unemp.adult      |    1.792|      0.856|   2.093|             0.0441|
|median.assets    |   13.736|     10.583|   1.298|             0.2033|
|num.low.salary   |    0.793|      0.235|   3.373|             0.0019|
>
```

```
> crime.lm.summary2 <- summary(crime.lm2)
> kable(crime.lm.summary2$coef, digits = c(3, 3, 3, 4), format =
'markdown')


|                 | Estimate| Std. Error| t value| Pr(>&#124;t&#124;)|
|:----------------|--------:|----------:|-------:|------------------:|
|(Intercept)      | -633.439|    145.470|  -4.354|             0.0001|
|young.males      |    1.127|      0.419|   2.691|             0.0109|
|is.south1        |   -0.557|     13.883|  -0.040|             0.9682|
|average.ed       |   15.328|      6.203|   2.471|             0.0185|
|exp.per.cap.1960 |    1.138|      0.227|   5.015|             0.0000|
|labour.part      |    0.069|      0.134|   0.515|             0.6101|
|male.per.fem     |    0.003|      0.173|   0.017|             0.9862|
|population       |   -0.064|      0.128|  -0.503|             0.6184|
|nonwhite         |   -0.014|      0.062|  -0.223|             0.8250|
|unemp.adult      |    0.931|      0.542|   1.719|             0.0944|
|median.assets    |   15.159|     10.524|   1.440|             0.1587|
|num.low.salary   |    0.826|      0.234|   3.527|             0.0012|
>
> # all 95% confidence intervals
> confint(crime.lm2)
                        2.5 %       97.5 %
(Intercept)      -928.7593182 -338.1183387
young.males         0.2766914    1.9770739
is.south1         -28.7410920   27.6278928
average.ed          2.7362499   27.9198056
exp.per.cap.1960    0.6775118    1.5990864
labour.part        -0.2023846    0.3398163
male.per.fem       -0.3482706    0.3543119
population         -0.3248958    0.1959409
nonwhite           -0.1394591    0.1118719
unemp.adult        -0.1684209    2.0314168
median.assets      -6.2068096   36.5247604
num.low.salary      0.3505063    1.3013656
> # Just for education
> confint(crime.lm2, parm = "average.ed")
              2.5 %   97.5 %
average.ed 2.73625 27.91981
> # 75% confidence interval
> confint(crime.lm2, parm = "average.ed", level = 0.75)
              12.5 %   87.5 %
average.ed 8.072542 22.58351
> # How does 2 SE rule compare to confint output?
> # lower endpoint
> coef(crime.lm2)["average.ed"] - 2* summary(crime.lm2)$coef["average.ed",
"Std. Error"]
average.ed
  2.922995
> # upper endpoint
> coef(crime.lm2)["average.ed"] + 2* summary(crime.lm2)$coef["average.ed",
"Std. Error"]
average.ed
  27.73306
>
> my.data <- data.frame(y = c(12, 13, 10, 5, 7, 12, 15),
+  x1 = c(6, 6.5, 5, 2.5, 3.5, 6, 7.5),
+  x2 = c(6, 6.5, 5, 2.5, 3.5, 6, 7.5))
> my.data
   y  x1  x2
1 12 6.0 6.0
2 13 6.5 6.5
3 10 5.0 5.0
4  5 2.5 2.5
5  7 3.5 3.5
6 12 6.0 6.0
7 15 7.5 7.5
>
> crime.lm.summary2$coef["exp.per.cap.1960",]
     Estimate      Std. Error         t value        Pr(>|t|)
1.13829907170 0.22697675756 5.01504684417 0.00001532994
>
> crime.lm.summary2$coef["average.ed",]
   Estimate  Std. Error     t value    Pr(>|t|)
15.32802778  6.20251646  2.47125951  0.01847635
>
>
>
> #4. Factors in linear regression
> #추가
> colnames(birthwt) <- c("birthwt.below.2500", "mother.age", "mother.weight",
+       "race", "mother.smokes", "previous.prem.labor", "hypertension",
"uterine.irr",
+       "physician.visits", "birthwt.grams")
> birthwt <- transform(birthwt,
+                 race = as.factor(mapvalues(race, c(1, 2, 3),
+                                  c("white","black", "other"))),
+                 mother.smokes = as.factor(mapvalues(mother.smokes,
+                                  c(0,1), c("no", "yes"))),
+                 hypertension = as.factor(mapvalues(hypertension,
+                                  c(0,1), c("no", "yes"))),
+                 uterine.irr = as.factor(mapvalues(uterine.irr,
+                                  c(0,1), c("no", "yes")))
+                 )
The following `from` values were not present in `x`: 1, 2, 3
The following `from` values were not present in `x`: 0, 1
The following `from` values were not present in `x`: 0, 1
The following `from` values were not present in `x`: 0, 1
>
> # Fit regression model
> birthwt.lm <- lm(birthwt.grams ~ race + mother.age, data = birthwt)
> # Regression model summary
> summary(birthwt.lm)

Call:
lm(formula = birthwt.grams ~ race + mother.age, data = birthwt)
```

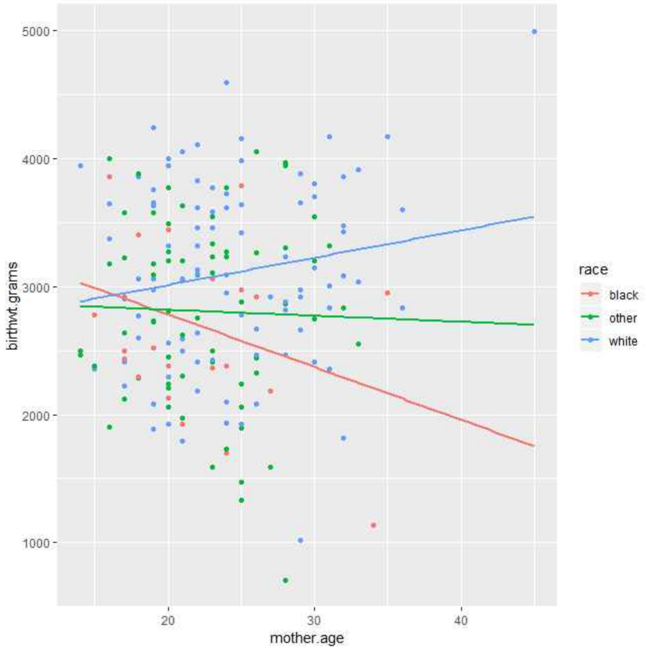
```
Residuals:
     Min       1Q   Median       3Q      Max 
-2131.57  -488.02    -1.16   521.87  1757.07 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 2584.264    258.393  10.001   <2e-16 ***
raceother     80.249    165.582   0.485    0.628    
racewhite    365.715    160.636   2.277    0.024 *  
mother.age     6.288     10.073   0.624    0.533    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 715.7 on 185 degrees of freedom
Multiple R-squared:  0.05217,	Adjusted R-squared:  0.0368 
F-statistic: 3.394 on 3 and 185 DF,  p-value: 0.01909

> 
> # Calculate race-specific intercepts
> intercepts <- c(coef(birthwt.lm)["(Intercept)"],
+   coef(birthwt.lm)["(Intercept)"] + coef(birthwt.lm)["raceother"],
+   coef(birthwt.lm)["(Intercept)"] + coef(birthwt.lm)["racewhite"])
> lines.df <- data.frame(intercepts = intercepts,
+   slopes = rep(coef(birthwt.lm)["mother.age"], 3),
+   race = levels(birthwt$race))
> qplot(x = mother.age, y = birthwt.grams, color = race, data = birthwt) +
+ geom_abline(aes(intercept = intercepts, slope = slopes, color = race),
+ data = lines.df)
```



```
> head(model.matrix(birthwt.lm), 20)
    (Intercept) raceother racewhite mother.age
85            1         0         0         19
86            1         1         0         33
87            1         0         1         20
88            1         0         1         21
89            1         0         1         18
91            1         1         0         21
92            1         0         1         22
93            1         1         0         17
94            1         0         1         29
95            1         0         1         26
96            1         1         0         19
97            1         1         0         19
98            1         1         0         22
99            1         1         0         30
100           1         0         1         18
101           1         0         1         18
102           1         0         0         15
103           1         0         1         25
104           1         1         0         20
105           1         0         1         28
>
> qplot(x = mother.age, y = birthwt.grams, color = race, data = birthwt) +
+ geom_abline(aes(intercept = intercepts, slope = slopes, color = race),
+ data = lines.df)
```



```
> qplot(x = mother.age, y = birthwt.grams, color = race, data = birthwt) +
+ stat_smooth(method = "lm", se = FALSE, fullrange = TRUE)
```



```
> birthwt.lm.interact <- lm(birthwt.grams ~ race * mother.age, data =
birthwt)
> summary(birthwt.lm.interact)

Call:
lm(formula = birthwt.grams ~ race * mother.age, data = birthwt)

Residuals:
     Min       1Q   Median       3Q      Max
-2182.35  -474.23    13.48   523.86  1496.51

Coefficients:
                     Estimate Std. Error t value    Pr(>|t|)
(Intercept)           3606.33     615.26   5.861 0.000000021 ***
raceother             -696.74     756.65  -0.921      0.3584
racewhite            -1022.79     694.21  -1.473      0.1424
mother.age             -41.17      27.82  -1.480      0.1407
raceother:mother.age    36.51      33.85   1.078      0.2823
racewhite:mother.age    62.54      30.67   2.039      0.0429 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 710.7 on 183 degrees of freedom
Multiple R-squared:  0.07541,	Adjusted R-squared:  0.05015
F-statistic: 2.985 on 5 and 183 DF,  p-value: 0.01291
```