```
> #Problem 4
> View(kuiper)
> price <- kuiper$Price
> mileage <- kuiper$Mileage
> make <- kuiper$Make
> type <- kuiper$Type
> cylinder <- kuiper$Cylinder
> liter <- kuiper$Liter
> model <- kuiper$Model
> trim <- kuiper$Trim
> doors <- kuiper$Doors
> cruise <- kuiper$Cruise
> sound <- kuiper$Sound
> leather <- kuiper$Leather
>
levels(make) <- 1:6
levels(model) <- 1:32
levels(type) <- 1:5
levels(trim) <- 1:47
make
model
type
trim

> #Regression Model with price of vehicle as the response (naturally) and mileage,
 make, cylinder,
> #leather interior, and sound.
> #It is assumed that there will be a strong relationship between some of these,
 though I do think
> #variables such as sound and leather will seem influential considering not everyone
 looks for
> #these properties in a vehicle.
> r <- lm(formula = price ~ mileage + make + cylinder + leather + sound, data =
 kuiper)
> summary(r)

Call:
lm(formula = price ~ mileage + make + cylinder + leather + sound,
    data = kuiper)

Residuals:
     Min       1Q   Median       3Q      Max
-11179.3  -2611.0     21.5   1581.7  24482.8

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.509e+01  1.069e+03   0.023   0.9813
mileage       -1.757e-01  1.755e-02 -10.015   <2e-16 ***
makeCadillac   1.355e+04  7.046e+02  19.236   <2e-16 ***
makeChevrolet -8.096e+02  5.431e+02  -1.491   0.1364
makePontiac   -1.112e+03  5.709e+02  -1.948   0.0517 .
makeSAAB       1.659e+04  6.677e+02  24.850   <2e-16 ***
makeSaturn    -1.761e+02  7.387e+02  -0.238   0.8116
cylinder       3.988e+03  1.417e+02  28.147   <2e-16 ***
leather        6.182e+02  3.457e+02   1.788   0.0741 .
sound          2.707e+02  3.245e+02   0.834   0.4044
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4062 on 794 degrees of freedom
Multiple R-squared:  0.833, Adjusted R-squared:  0.8311
F-statistic: 440.1 on 9 and 794 DF,  p-value: < 2.2e-16

> #The very large F-statistic value of 440.1 on 9, dof = 794, with a p-value of
 2.2e-16 suggests
> #that this model the model is very strong, and hence significant. However, it only
 tells us that
> #at least one of the predictor variables is significant, though this relationship
 would have to
> #to be considerably strong to yield such a result. Therefore, it is likely more than
 one variable
> #yielding the strength of this model. It is significant at all alpha levels of
 significance.
>
> #Removing sound and leather from the regression model to determine the resulting
 model:
> r2 <- lm(formula = price ~ mileage + make + cylinder, data = kuiper)
> summary(r2)

Call:
lm(formula = price ~ mileage + make + cylinder, data = kuiper)

Residuals:
     Min       1Q   Median       3Q      Max
-11054.1  -2629.8    -26.3   1651.4  24611.6

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.208e+02  1.024e+03   0.509   0.6112
mileage       -1.756e-01  1.757e-02  -9.994   <2e-16 ***
makeCadillac   1.388e+04  6.779e+02  20.482   <2e-16 ***
makeChevrolet -5.430e+02  5.279e+02  -1.029   0.3040
makePontiac   -1.005e+03  5.664e+02  -1.775   0.0763 .
makeSAAB       1.673e+04  6.573e+02  25.458   <2e-16 ***
makeSaturn    -2.189e+02  7.336e+02  -0.298   0.7654
cylinder       3.980e+03  1.413e+02  28.173   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4068 on 796 degrees of freedom
Multiple R-squared:  0.8321,     Adjusted R-squared:  0.8306
F-statistic: 563.5 on 7 and 796 DF,  p-value: < 2.2e-16

> #The resulting test statistic is even larger where the p-value is relatively the
 same. I assume
> #that this p-value is a limiting value for very large F-statistic values. This model
 is even
> #stronger than the original just as assumed before the model was run. This suggests
 that either
> #sound or leather had a negative impact on the linear relationship between the
 nonzero coefficients.
> #I still assume that there is more than one variable that is nonzero.
```

```
>
> #We will now determine, definitively, if one of sound or leather impacted the
 relationship, or
> #their coefficients are nonzero. We will use anova to do so.
>
>
> #Analysis of Variance
> anova(r2,r)
Analysis of Variance Table

Model 1: price ~ mileage + make + cylinder
Model 2: price ~ mileage + make + cylinder + leather + sound
  Res.Df        RSS Df Sum of Sq     F Pr(>F)
1    796 1.3175e+10
2    794 1.3102e+10  2  73097700 2.215 0.1098
> #The p-value = 0.1098. Thus, at the alpha = 0.05 significance level, we accept the
 null
> #hypothesis, that the reduced model is correct. That is to say that the model
 without sound and
> #leather is the stronger model (as assumed). It further states that we cannot reject
 the assumption
> #that both coefficients for sound and leather were zero. Thus, sound and leather
 aren't particularly
> #strong factors determining the price of a vehicle on average.
>
> #Let's try to find which variable has the strong predictive power
>
> #Deleting cylinder from the model
> r3 <- lm(formula = price ~ mileage + make, data = kuiper)
> summary(r3)

Call:
lm(formula = price ~ mileage + make, data = kuiper)

Residuals:
     Min       1Q   Median       3Q      Max
-11755.2  -3274.0   -701.8   1517.1  28174.1

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.431e+04  8.182e+02  29.705  < 2e-16 ***
mileage       -1.709e-01  2.481e-02  -6.888 1.15e-11 ***
makeCadillac   1.986e+04  9.093e+02  21.844  < 2e-16 ***
makeChevrolet -4.520e+03  7.185e+02  -6.290 5.22e-10 ***
makePontiac   -2.592e+03  7.959e+02  -3.257  0.00117 **
makeSAAB       8.771e+03  8.381e+02  10.465  < 2e-16 ***
makeSaturn    -6.852e+03  9.813e+02  -6.983 6.10e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5746 on 797 degrees of freedom
Multiple R-squared:  0.6647,     Adjusted R-squared:  0.6621
F-statistic: 263.3 on 6 and 797 DF,  p-value: < 2.2e-16

> #The F-statistic decreased in value, suggesting that cylinder was had significant
 predictive
```

```
> #power or at least had significant predictive power when paired with other mileage
 and/or make.
>
> #Simple Linear Regression with Price vs. mileage
> r4<-lm(formula = price ~ mileage, data = kuiper)
> summary(r4)

Call:
lm(formula = price ~ mileage, data = kuiper)

Residuals:
   Min     1Q Median     3Q    Max
-13905  -7254  -3520   5188  46091

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.476e+04  9.044e+02  27.383  < 2e-16 ***
mileage     -1.725e-01  4.215e-02  -4.093 4.68e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9789 on 802 degrees of freedom
Multiple R-squared:  0.02046,    Adjusted R-squared:  0.01924
F-statistic: 16.75 on 1 and 802 DF,  p-value: 4.685e-05

> #The F-statistic decreased significantly suggesting that mileage of a vehicle on
 it's own does
> #not have strong predictive power of the response variable, price of the vehicle.
>
> #This suggests that when mileage of the vehicle, make of the vehicle, and cylinder
 are
> #considered together simultaneously, we obtain a strong model with a very strong
 relationship
> #between them.
>
> #Let's run a diagnostic on our r2 regression since it was our strongest model.
> par(mfrow = c(3, 2))
> plot(r2, which = c(1:6))
> summary(influence.measures(r2))
Potentially influential observations of
       lm(formula = price ~ mileage + make + cylinder, data = kuiper) :

    dfb.1_ dfb.milg dfb.mkCd dfb.mkCh dfb.mkPn dfb.mSAA dfb.mkSt dfb.cyln dffit
 cov.r   cook.d
125  0.03   0.02    -0.18    -0.01     0.00    -0.02    -0.01    -0.05    -0.30_*
 0.95_*  0.01
126  0.03   0.00    -0.17    -0.01     0.00    -0.02    -0.01    -0.04    -0.28
 0.96_*  0.01
127  0.04  -0.02    -0.18    -0.01    -0.01    -0.02    -0.01    -0.04    -0.29
 0.96_*  0.01
128  0.06  -0.05    -0.19    -0.01    -0.01    -0.02    -0.02    -0.05    -0.32_*
 0.95_*  0.01
129  0.07  -0.08    -0.19    -0.01    -0.01    -0.02    -0.01    -0.04    -0.31_*
 0.96_*  0.01
130  0.07  -0.10    -0.18    -0.01    -0.01    -0.02    -0.01    -0.04    -0.31_*
 0.96_*  0.01
```

```
140  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
 1.03_*  0.00
151  0.08  -0.50     0.42     0.02    -0.01     0.06     0.04     0.11     0.87_*
 0.70_*  0.09
152  0.02  -0.32     0.40     0.02     0.00     0.05     0.03     0.11     0.75_*
 0.73_*  0.07
153  0.01  -0.30     0.42     0.02     0.00     0.05     0.03     0.11     0.76_*
 0.71_*  0.07
154 -0.02  -0.17     0.39     0.02     0.00     0.05     0.03     0.10     0.66_*
 0.75_*  0.05
155 -0.05  -0.08     0.38     0.02     0.01     0.04     0.03     0.10     0.63_*
 0.76_*  0.05
156 -0.07  -0.02     0.37     0.02     0.01     0.04     0.03     0.09     0.60_*
 0.78_*  0.04
157 -0.10   0.09     0.33     0.02     0.01     0.03     0.03     0.08     0.53_*
 0.83_*  0.03
158 -0.12   0.18     0.29     0.02     0.01     0.03     0.02     0.07     0.49_*
 0.87_*  0.03
159 -0.13   0.21     0.27     0.02     0.01     0.02     0.02     0.06     0.48_*
 0.89_*  0.03
160 -0.17   0.34     0.25     0.02     0.02     0.02     0.02     0.06     0.51_*
 0.93_*  0.03
341 -0.24  -0.28    -0.14     0.20     0.03     0.18     0.13     0.41     0.54_*
 0.89_*  0.04
342 -0.26  -0.26    -0.14     0.21     0.03     0.19     0.14     0.43     0.55_*
 0.87_*  0.04
343 -0.28  -0.13    -0.13     0.19     0.04     0.17     0.13     0.40     0.47_*
 0.89_*  0.03
344 -0.27  -0.08    -0.11     0.17     0.03     0.15     0.11     0.35     0.41_*
 0.92_*  0.02
345 -0.28   0.00    -0.11     0.17     0.03     0.15     0.11     0.34     0.39_*
 0.92_*  0.02
346 -0.27   0.05    -0.10     0.16     0.03     0.13     0.10     0.31     0.36_*
 0.94_*  0.02
347 -0.29   0.07    -0.10     0.16     0.03     0.14     0.10     0.32     0.37_*
 0.93_*  0.02
348 -0.28   0.08    -0.09     0.15     0.03     0.13     0.10     0.30     0.35_*
 0.94_*  0.02
349 -0.27   0.12    -0.08     0.14     0.03     0.12     0.09     0.27     0.33_*
 0.96_*  0.01
350 -0.29   0.15    -0.08     0.15     0.03     0.12     0.09     0.29     0.36_*
 0.95_*  0.02
351 -0.16  -0.13    -0.08     0.12     0.02     0.11     0.08     0.24     0.30_*
 0.97    0.01
352 -0.17  -0.11    -0.08     0.12     0.02     0.11     0.08     0.25     0.30_*
 0.97    0.01
355 -0.18  -0.09    -0.08     0.12     0.02     0.11     0.08     0.25     0.30
 0.97_*  0.01
650  0.02  -0.07     0.00     0.00     0.00    -0.03     0.00     0.00    -0.08
 1.03_*  0.00
745  0.01  -0.02     0.00     0.00     0.00     0.00     0.03     0.00     0.05
 1.03_*  0.00
764 -0.01   0.04     0.00     0.00     0.00     0.00     0.05    -0.01     0.08
 1.03_*  0.00
765 -0.01   0.02     0.00     0.00     0.00     0.00    -0.03     0.00    -0.05
 1.03_*  0.00
```

```
766 -0.01   0.01     0.00     0.00     0.00     0.00    -0.02     0.00    -0.03
 1.03_*  0.00
791  0.00   0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
 1.03_*  0.00
794  0.00   0.02     0.00     0.00     0.00     0.00     0.03     0.00     0.05
 1.03_*  0.00
    hat
125  0.01
126  0.01
127  0.01
128  0.01
129  0.01
130  0.01
140  0.02
151  0.02
152  0.02
153  0.02
154  0.01
155  0.01
156  0.01
157  0.01
158  0.01
159  0.02
160  0.02
341  0.02
342  0.02
343  0.02
344  0.01
345  0.01
346  0.01
347  0.01
348  0.01
349  0.02
350  0.02
351  0.02
352  0.02
355  0.02
650  0.02
745  0.02
764  0.02
765  0.02
766  0.02
791  0.02
794  0.02
> #There are several influential points in the dffit column which isn't too suprising
 considering
> #the market for cars. The leverage points are relatively high considering there are
 804 data points.
> #Those points with leverage of 0.02 or higher are likely to influence the data.
> #The QQ-Plot suggest that there is a sharp dispersion of normality for the error
 terms.
> #Moreover, the Residual against fitted values plot shows that the variance of the
 error
> #terms are far from constant (i.e. nonconstant). It actually seems that they are
 serially
> #correlated.
```

```
> #These results are surprising, but not completely suprising considering the
 assumption of
> #strong influence from many variables in the automobile industry.
```