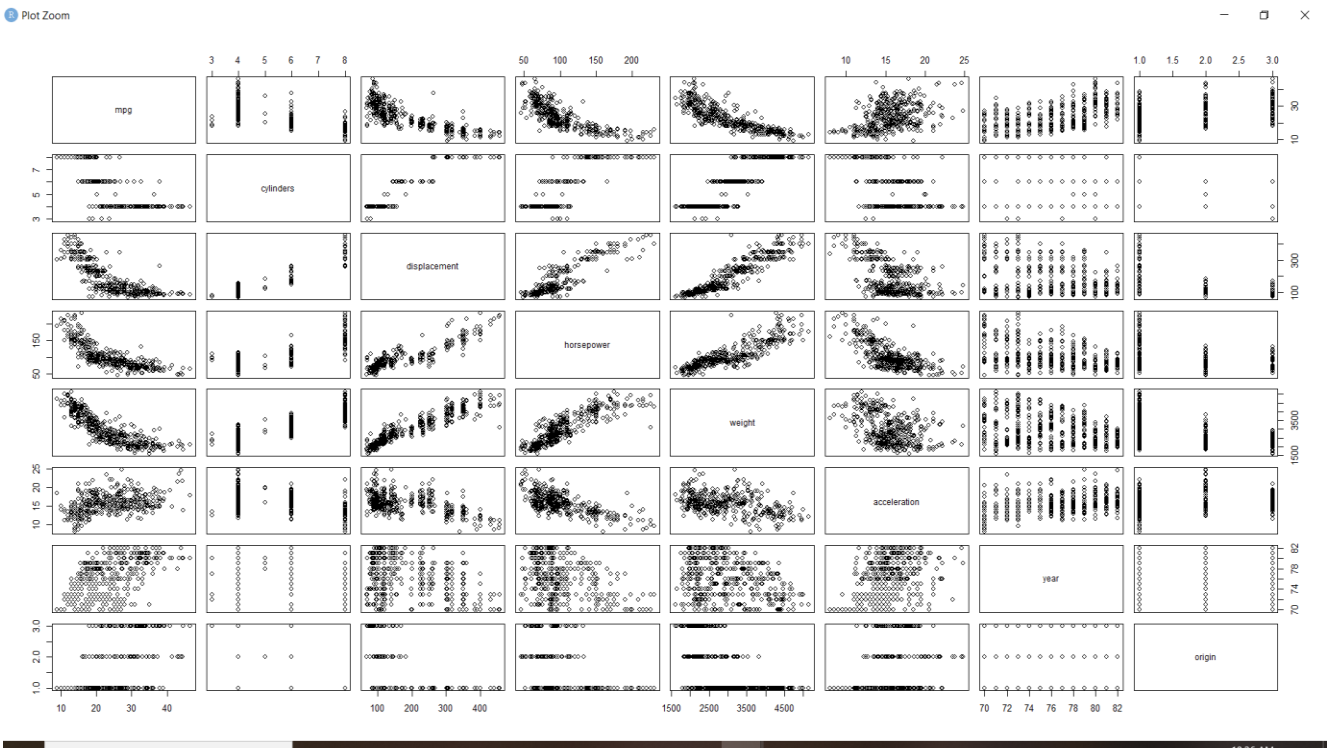
(a)



(b)

```
> cor(Auto[, sapply(Auto, is.numeric)])
                    mpg  cylinders displacement horsepower     weight acceleration       year
mpg           1.0000000 -0.7776175   -0.8051269 -0.7784268 -0.8322442    0.4233285  0.5805410
cylinders    -0.7776175  1.0000000    0.9508233  0.8429834  0.8975273   -0.5046834 -0.3456474
displacement -0.8051269  0.9508233    1.0000000  0.8972570  0.9329944   -0.5438005 -0.3698552
horsepower   -0.7784268  0.8429834    0.8972570  1.0000000  0.8645377   -0.6891955 -0.4163615
weight       -0.8322442  0.8975273    0.9329944  0.8645377  1.0000000   -0.4168392 -0.3091199
acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955 -0.4168392    1.0000000  0.2903161
year          0.5805410 -0.3456474   -0.3698552 -0.4163615 -0.3091199    0.2903161  1.0000000
origin        0.5652088 -0.5689316   -0.6145351 -0.4551715 -0.5850054    0.2127458  0.1815277
                origin
mpg           0.5652088
cylinders    -0.5689316
displacement -0.6145351
horsepower   -0.4551715
weight       -0.5850054
acceleration  0.2127458
year          0.1815277
origin        1.0000000
>
```

(c)

```
> summary(lm.fit)

Call:
lm(formula = mpg ~ . - name, data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.5903 -2.1565 -0.1169  1.8690 13.0604

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -17.218435   4.644294  -3.707  0.00024 ***
cylinders     -0.493376   0.323282  -1.526  0.12780
displacement   0.019896   0.007515   2.647  0.00844 **
horsepower    -0.016951   0.013787  -1.230  0.21963
weight        -0.006474   0.000652  -9.929  < 2e-16 ***
acceleration   0.080576   0.098845   0.815  0.41548
year           0.750773   0.050973  14.729  < 2e-16 ***
origin         1.426141   0.278136   5.127 4.67e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.328 on 384 degrees of freedom
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8182
F-statistic: 252.4 on 7 and 384 DF,  p-value: < 2.2e-16
```

(i)

Yes, there is a strong relationship between the predictors and the response.

The overall F-statistic is really high (252.4) and the p-value is extremely small (less than 2.2e-16), which basically means it's very unlikely that all the predictors have no effect. So, at least one of the variables is definitely helping explain the changes in mpg. Additionally, the high R-squared value (82.15%) indicates that the model explains a large portion of the variability in mpg.

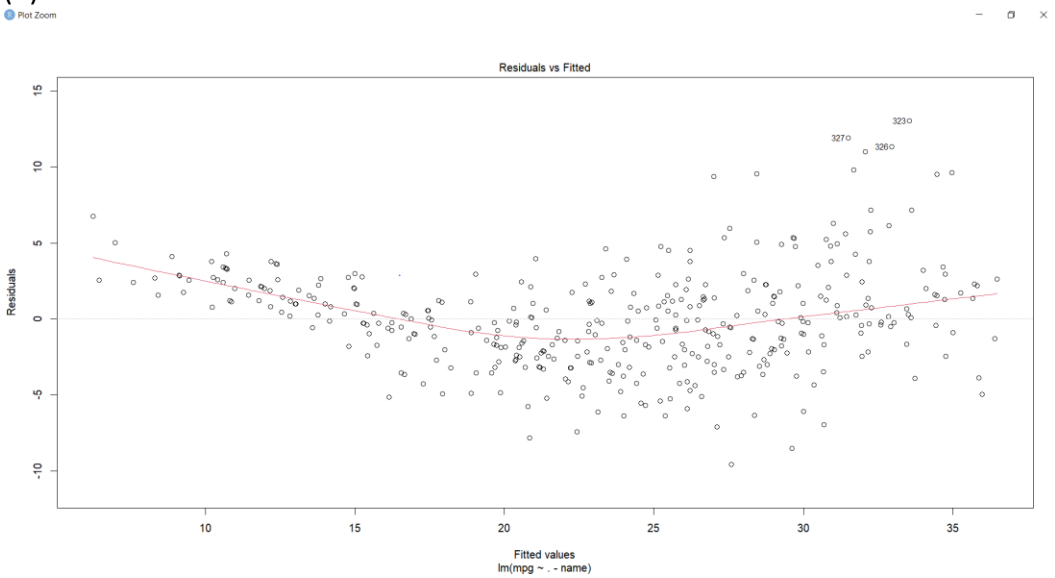
(ii)

From the regression output, 4 variables have small p-values, meaning they're statistically significant. They are following:

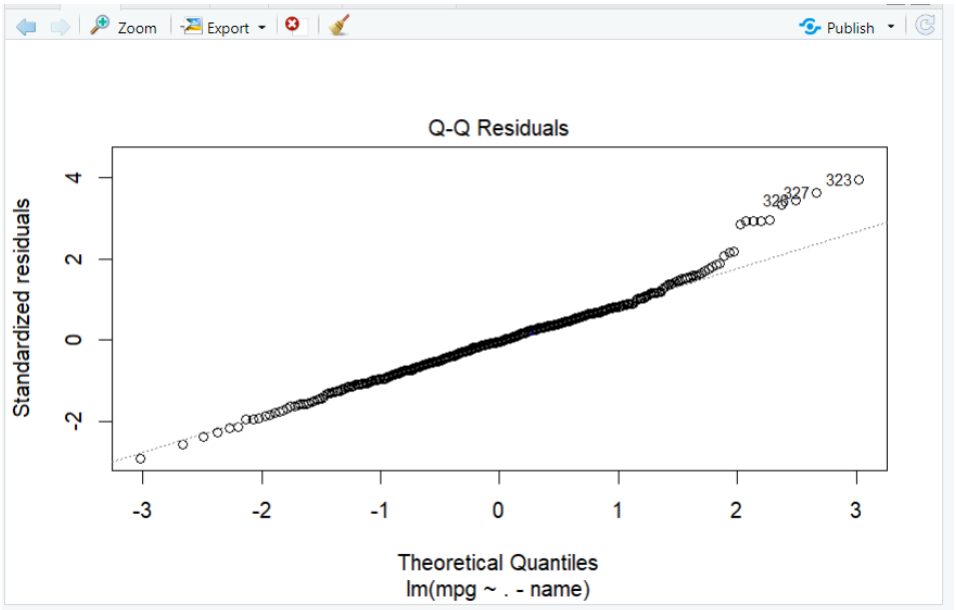
1. Displacement ($p = 0.00844$
2. Weight ($p < 2e\text{-}16$)
3. Year ($p < 2e\text{-}16$
4. Origin ($p \approx 4.67e\text{-}07$)

(iii)
The coefficient for year is 0.750773, and we know it is statistically significant (p-value < 2e-16). This means that each newer model year of a car increases its fuel efficiency by about 0.75 miles per gallon, if everything else stays the same.

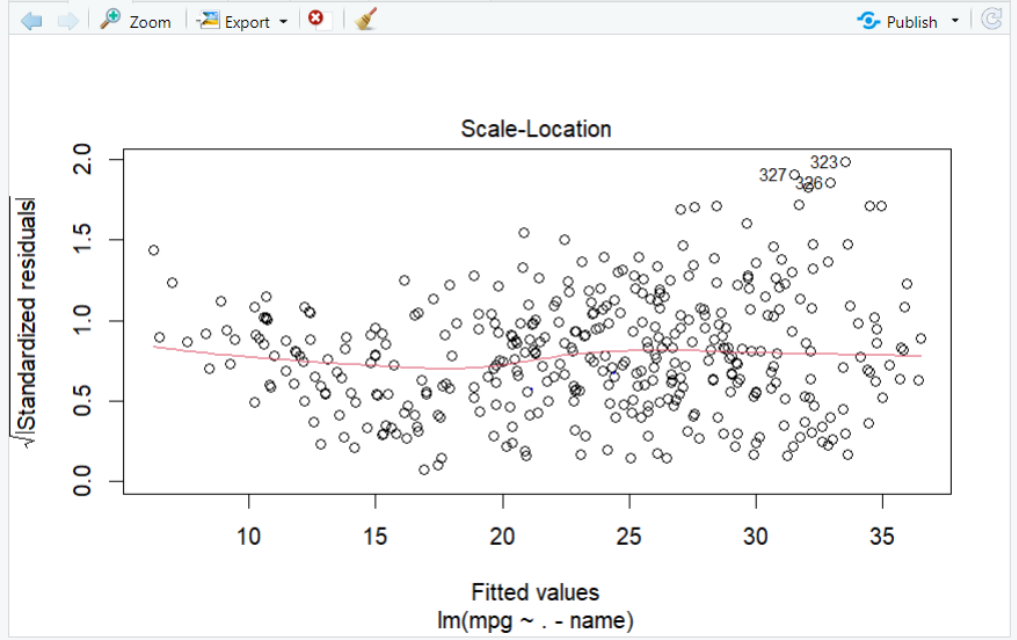
(d)



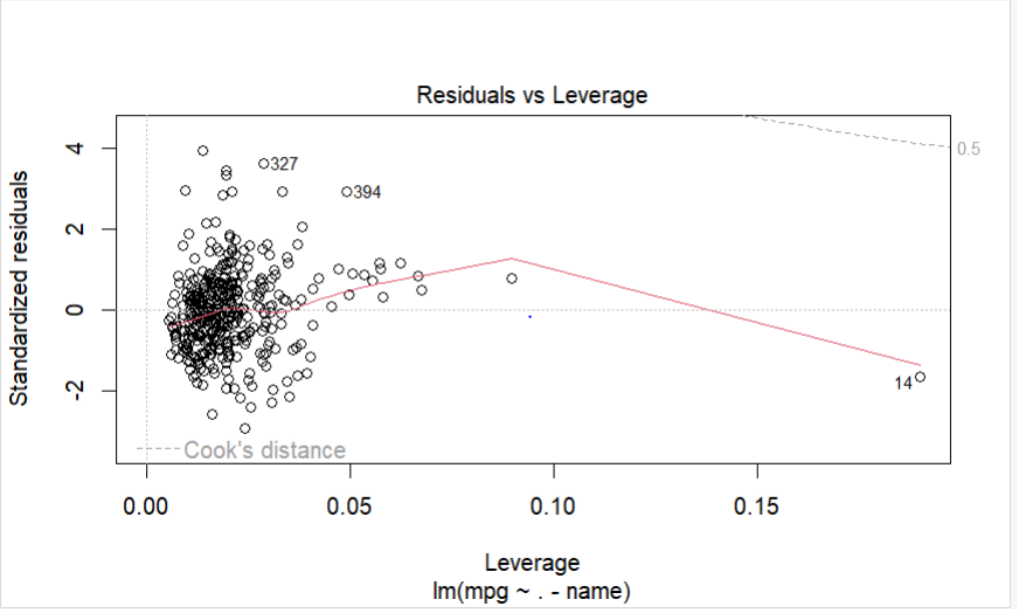
The residuals are randomly scattered around 0. So, it suggests that the linearity assumption is fairly reasonable, though a slight curve shows mild non-linearity.



Most points fall along the 45-degree line, but some points on the right (e.g., observations 323, 327, 329) deviate. Residuals are mostly normal, but there are a few moderate outliers.



The spread of points increases slightly with fitted values.



Points 14, 327, 394 show relatively high leverage. Some points are close to or beyond Cook's distance line. A few high-leverage and influential points are present. They could have a strong impact on the regression model.

The diagnostic plots suggest that the linear regression model mostly meets the assumptions of linearity, normality, and constant variance. However, there are also few outliers, slightly increasing spread, high leverage and mild non-linearity which could influence the model.

(e)

```
    Min      1Q  Median      3Q     Max
-7.6303 -1.4481  0.0596  1.2739 11.1386

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           3.548e+01  5.314e+01   0.668  0.50475
cylinders             6.989e+00  8.248e+00   0.847  0.39738
displacement         -4.785e-01  1.894e-01  -2.527  0.01192 *
horsepower            5.034e-01  3.470e-01   1.451  0.14769
weight                4.133e-03  1.759e-02   0.235  0.81442
acceleration         -5.859e+00  2.174e+00  -2.696  0.00735 **
year                  6.974e-01  6.097e-01   1.144  0.25340
origin               -2.090e+01  7.097e+00  -2.944  0.00345 **
cylinders:displacement -3.383e-03  6.455e-03  -0.524  0.60051
cylinders:horsepower   1.161e-02  2.420e-02   0.480  0.63157
cylinders:weight       3.575e-04  8.955e-04   0.399  0.69000
cylinders:acceleration 2.779e-01  1.664e-01   1.670  0.09584 .
cylinders:year        -1.741e-01  9.714e-02  -1.793  0.07389 .
cylinders:origin       4.022e-01  4.926e-01   0.816  0.41482
displacement:horsepower -8.491e-05  2.885e-04  -0.294  0.76867
displacement:weight    2.472e-05  1.470e-05   1.682  0.09342 .
displacement:acceleration -3.479e-03  3.342e-03  -1.041  0.29853
displacement:year      5.934e-03  2.391e-03   2.482  0.01352 *
displacement:origin    2.398e-02  1.947e-02   1.232  0.21875
horsepower:weight     -1.968e-05  2.924e-05  -0.673  0.50124
horsepower:acceleration -7.213e-03  3.719e-03  -1.939  0.05325 .
horsepower:year       -5.838e-03  3.938e-03  -1.482  0.13916
horsepower:origin      2.233e-03  2.930e-02   0.076  0.93931
weight:acceleration    2.346e-04  2.289e-04   1.025  0.30596
weight:year           -2.245e-04  2.127e-04  -1.056  0.29182
weight:origin         -5.789e-04  1.591e-03  -0.364  0.71623
acceleration:year      5.562e-02  2.558e-02   2.174  0.03033 *
acceleration:origin    4.583e-01  1.567e-01   2.926  0.00365 **
year:origin            1.393e-01  7.399e-02   1.882  0.06062 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.695 on 363 degrees of freedom
Multiple R-squared:  0.8893,    Adjusted R-squared:  0.8808
F-statistic: 104.2 on 28 and 363 DF,  p-value: < 2.2e-16
```

Some main effects are statistically significant (p-value < 0.05) here:
displacement (p = 0.01192)
acceleration (p = 0.00735)
origin (p = 0.00345)

Following interaction are statistically significant (p-value < 0.05):
Displacement : Year (p=0.01352)
Acceleration : Year (p=0.03033)
Acceleration : Origin (p=0.00365)

(f)

```
Call:
lm(formula = mpg ~ log(horsepower) + sqrt(displacement) + weight +
    I(weight^2) + acceleration + I(acceleration^2) + year + origin,
    data = Auto)

Residuals:
    Min      1Q  Median      3Q     Max
-9.5896 -1.6149 -0.0175  1.5273 12.5539

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         4.330e+01  8.683e+00   4.987 9.31e-07 ***
log(horsepower)    -6.246e+00  1.452e+00  -4.301 2.16e-05 ***
sqrt(displacement) -3.765e-02  1.476e-01  -0.255   0.7987
weight             -1.640e-02  1.831e-03  -8.958  < 2e-16 ***
I(weight^2)         1.878e-06  2.354e-07   7.978 1.74e-14 ***
acceleration       -2.438e+00  4.633e-01  -5.262 2.38e-07 ***
I(acceleration^2)   6.920e-02  1.353e-02   5.113 5.01e-07 ***
year                7.866e-01  4.453e-02  17.667  < 2e-16 ***
origin              6.502e-01  2.519e-01   2.581   0.0102 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.863 on 383 degrees of freedom
Multiple R-squared:  0.8682,    Adjusted R-squared:  0.8654
F-statistic: 315.3 on 8 and 383 DF,  p-value: < 2.2e-16

>
```

- R-squared = 0.8682: This means the model explains about 86.82% of the variation in mpg, which is better than previous models.
- Residual Standard Error = 2.86: Lower error means the model is making more accurate predictions.
- log(horsepower): Strong impact on mpg. As horsepower increases, mpg decreases, but not in a straight line.
- weight and weight²: Show a curved (nonlinear) effect. Heavier cars have lower mpg, but the rate of decrease changes with more weight.
- acceleration and acceleration²: Also show a curved effect. Mpg depends on acceleration in a nonlinear way.
- year: Newer cars generally have better mpg.
- origin: The place where the car was made affects its mpg.
- sqrt(displacement): This transformation didn’t help. It wasn’t statistically significant (p = 0.79), so it doesn’t add value to the model.