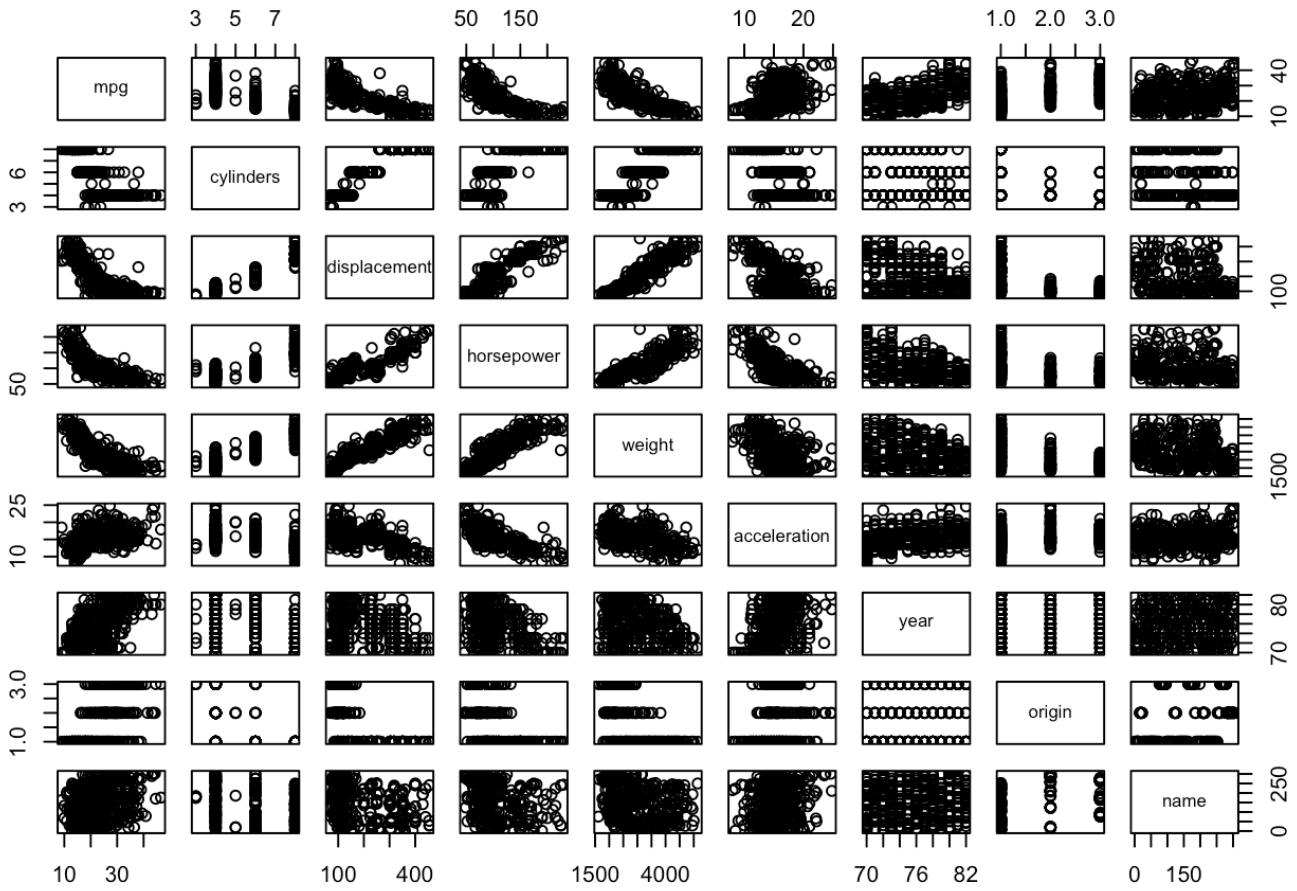
Problem 5

Part A

```
library("ISLR")
plot(Auto)
```




```
Auto.quantonly = Auto[, !names(Auto) %in% c("name")]
```

Part B

```
cor(Auto.quantonly)
```

```
##                     mpg  cylinders displacement horsepower     weight
## mpg           1.0000000 -0.7776175   -0.8051269 -0.7784268 -0.8322442
## cylinders    -0.7776175  1.0000000    0.9508233  0.8429834  0.8975273
## displacement -0.8051269  0.9508233    1.0000000  0.8972570  0.9329944
## horsepower   -0.7784268  0.8429834    0.8972570  1.0000000  0.8645377
## weight       -0.8322442  0.8975273    0.9329944  0.8645377  1.0000000
## acceleration  0.4233285 -0.5046834   -0.5438005 -0.6891955 -0.4168392
## year          0.5805410 -0.3456474   -0.3698552 -0.4163615 -0.3091199
## origin        0.5652088 -0.5689316   -0.6145351 -0.4551715 -0.5850054
##              acceleration       year     origin
## mpg             0.4233285  0.5805410  0.5652088
## cylinders      -0.5046834 -0.3456474 -0.5689316
## displacement   -0.5438005 -0.3698552 -0.6145351
## horsepower     -0.6891955 -0.4163615 -0.4551715
## weight         -0.4168392 -0.3091199 -0.5850054
## acceleration    1.0000000  0.2903161  0.2127458
## year            0.2903161  1.0000000  0.1815277
## origin          0.2127458  0.1815277  1.0000000
```

Part C

Note: For this problem, I relied heavily on this site (http://www.r-tutor.com/elementary-statistics/simple-linear-regression/significance-test-linear-regression).

```
Auto.fit <- lm(mpg ~ cylinders + displacement + horsepower + weight + acceleration
+ year + origin, data=Auto.quantonly)
summary(Auto.fit)
```

```
##
## Call:
## lm(formula = mpg ~ cylinders + displacement + horsepower + weight +
##     acceleration + year + origin, data = Auto.quantonly)
##
## 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 clear relationship between the predictors and the response. The p-value is much less than `0.05`, so we reject the null hypothesis.

(ii)

We look at the `Pr(>|t|)` column to determine which predictors appear to have a statistically significant relationship. Here we find that the following predictors have a p-value < 0.05:

- displacement
- weight
- year
- origin

(iii)

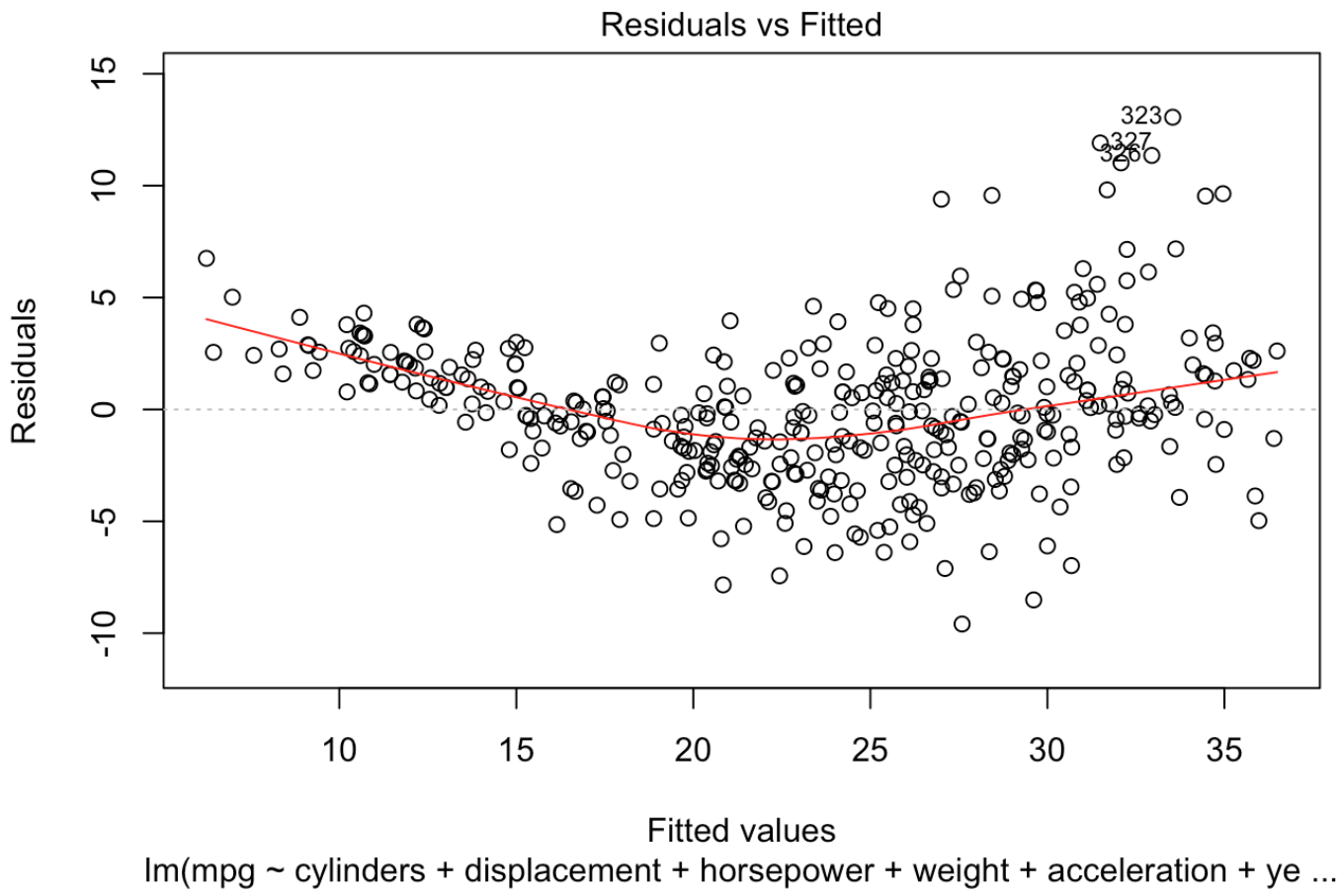
The year the car was made is one of the most important predictors for its mpg. This makes sense intuitively as a sanity check, because older cars don't have catalytic converters and they generally weren't built with efficiency in mind.

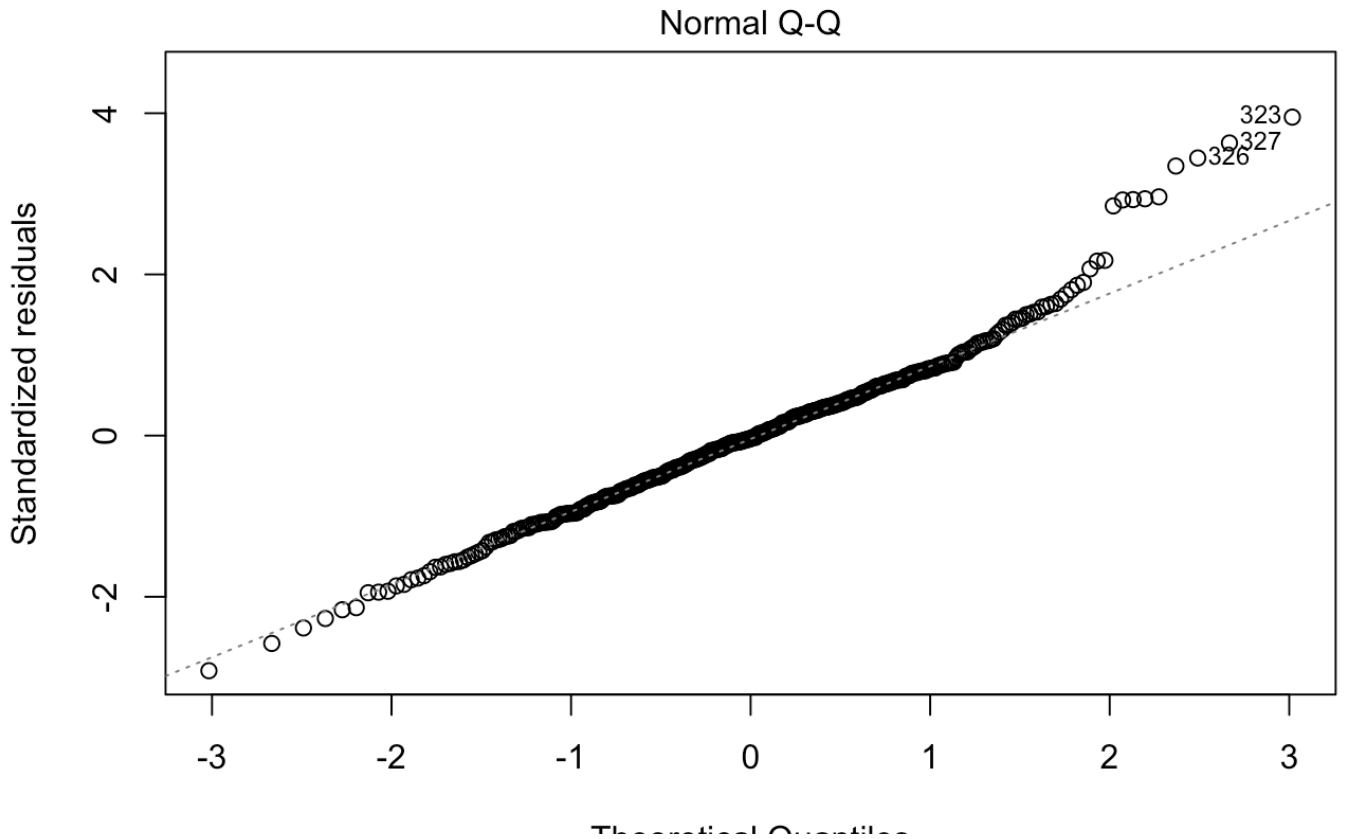
Part D

As you can see from the plot below, there are a few issues with the fit:

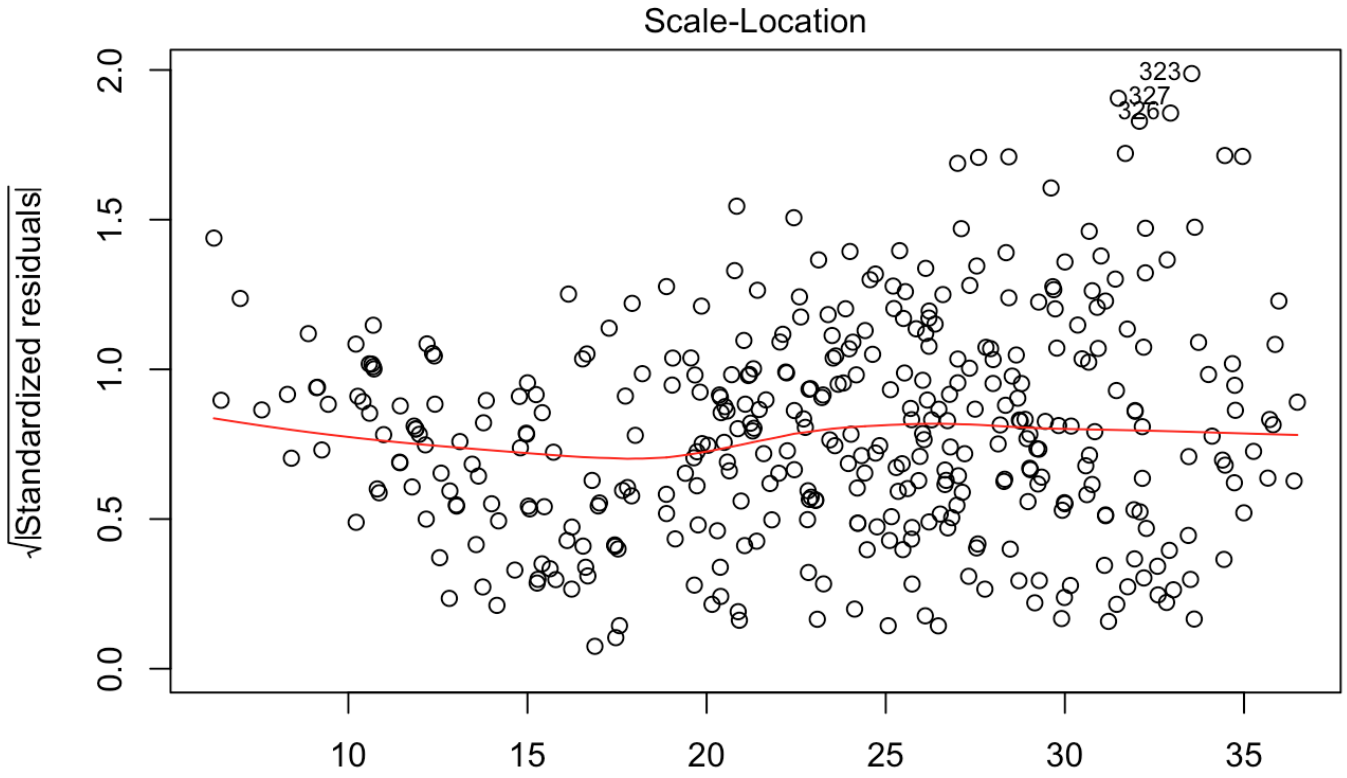
- There are some unusually large outliers in the residuals vs. the fitted (up in the top right hand corner and bottom right). The model fits our data well on the left, but the variance spreads out as we move rightwards.
- Nearly all observations are clustered on the left, but then there's one all the way off to the right with an extremely high leverage. This smells fishy…

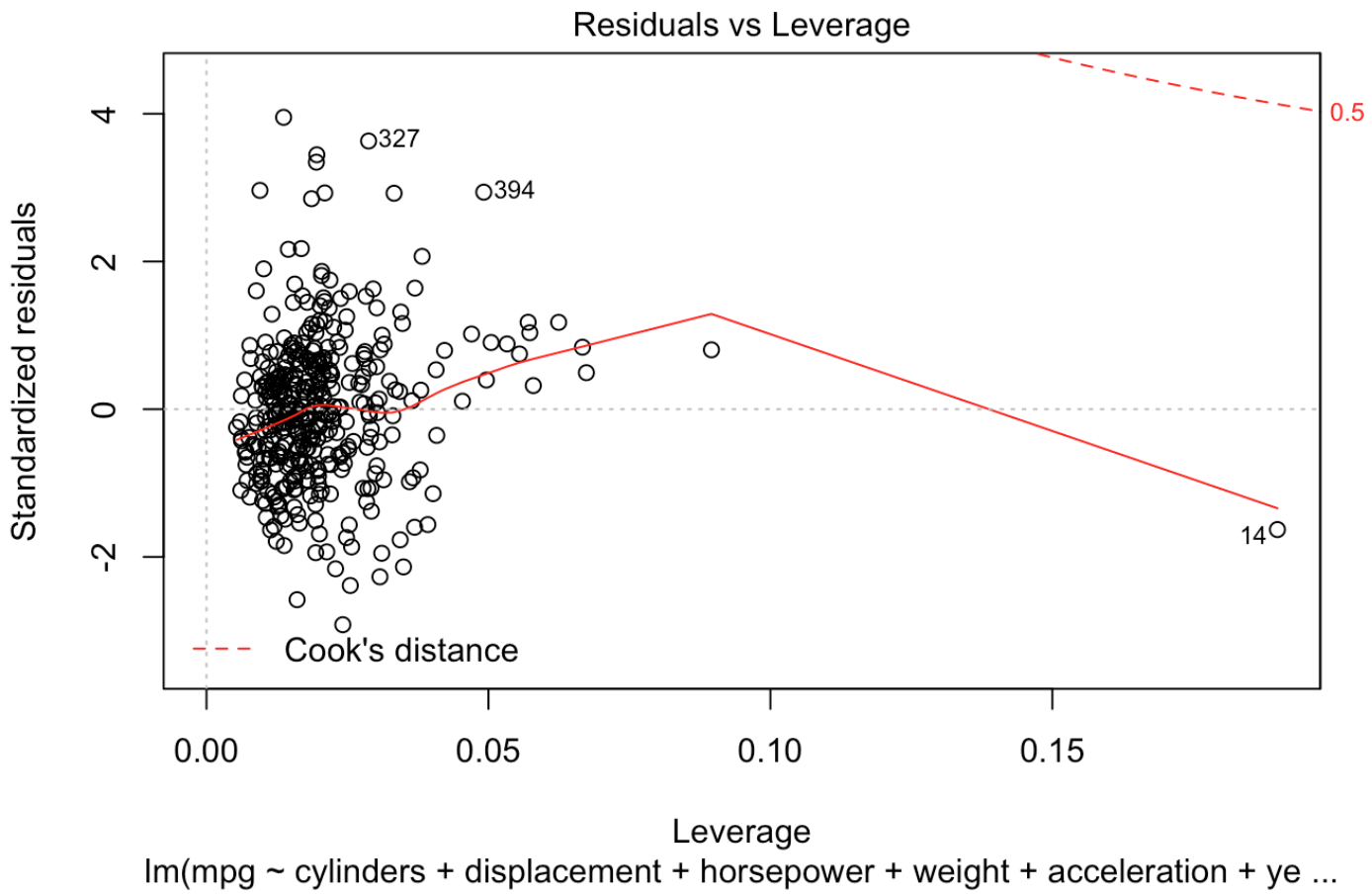
```
plot(Auto.fit)
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ cylinders + displacement + horsepower + weight + acceleration + ye ...
323
327
326
15
10
5
0
-5
-10
10
15
20
25
30
35


Normal Q-Q
Standardized residuals
Theoretical Quantiles
323
327
326
4
2
0
-2
-3
-2
-1
0
1
2
3

Theoretical Quantiles
lm(mpg ~ cylinders + displacement + horsepower + weight + acceleration + ye ...




Residuals vs Leverage
Standardized residuals
4
2
0
-2
327
394
14
0.5
Cook's distance
0.00
0.05
0.10
0.15
Leverage
lm(mpg ~ cylinders + displacement + horsepower + weight + acceleration + ye ...