Chapter 3 Problem 8

8a.) Use the lm() function to perform a simple linear regression with mpg as the response and horsepower as the predictor. Comment on the output.

```
> auto=read.csv("Auto.csv",header=T,na.strings="?")
> lm.fit=lm(mpg~horsepower,data=auto)
> summary(lm.fit)

Call:
lm(formula = mpg ~ horsepower, data = auto)

Residuals:
     Min       1Q   Median       3Q      Max
-13.5710  -3.2592  -0.3435   2.7630  16.9240

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 39.935861   0.717499   55.66   <2e-16 ***
horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.906 on 390 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.6059,      Adjusted R-squared:  0.6049
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

i. Is there a relationship between the predictor and the response?
H_o: No relationship between horsepower & mpg. vs. H_a: There is a relationship between horsepower & mpg.

We reject the null hypothesis. The low p-value associated with the horsepower parameter (β_1) suggests that there is a relationship between horsepower and mpg.

ii. How strong is the relationship between the predictor and the response?
r = sqrt(0.6059) = 0.7784
The correlation coefficient indicates a strong relationship, though not conventionally the strongest.

iii. Is the relationship between the predictor and the response positive or negative?
There is a negative relationship. With a one-unit increase in horsepower, mpg decreases by 0.1578.

iv. What is the predicted mpg with a horsepower of 98? What are the associated 95% CI and PI?
Our model is: mpg = 39.9359 – 0.1578(horsepower)
39.9359 – 0.1578(98) = 24.4715.

Confidence interval:

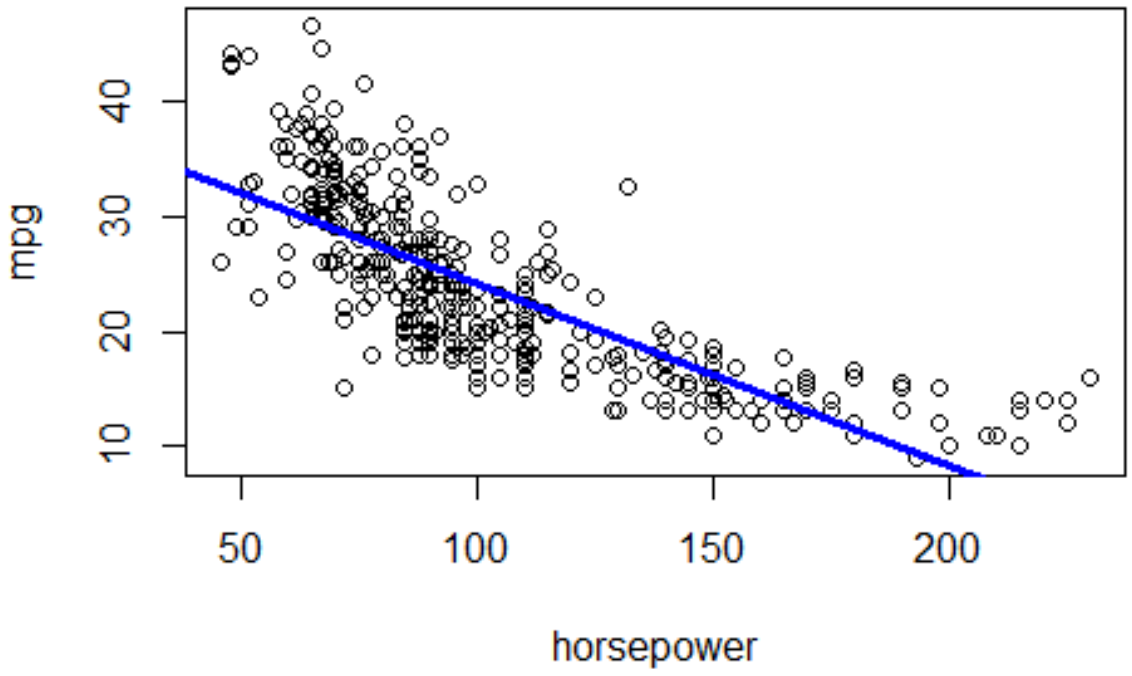
```
predict(lm.fit, data.frame(horsepower=c(98)),interval="confidence")
       fit      lwr      upr
1 24.46708 23.97308 24.96108
```

Prediction interval:

```
> predict(lm.fit, data.frame(horsepower=c(98)),interval="prediction")
       fit     lwr      upr
1 24.46708 14.8094 34.12476
```

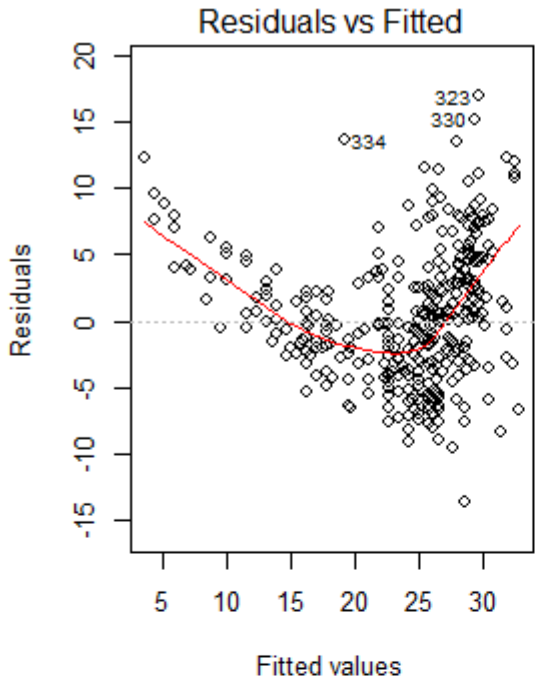
8b.) Plot the response and the predictor. Use the abline() function to display the least squares regression line.

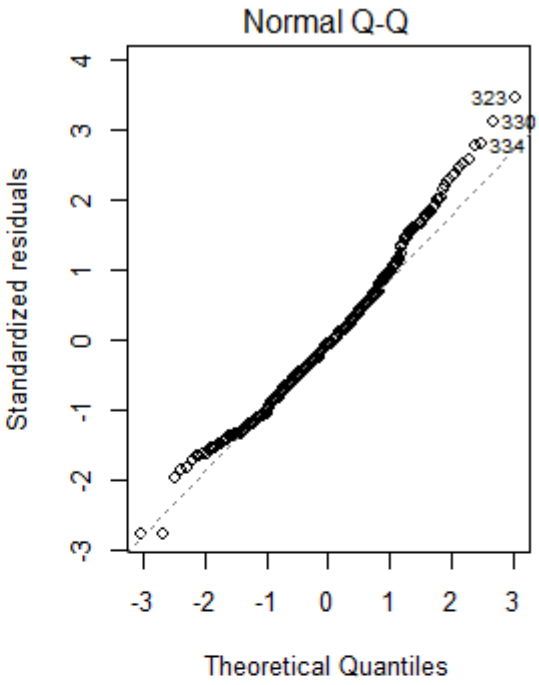
```
> plot(auto$horsepower,auto$mpg,xlab="horsepower",ylab="mpg")
> abline(lm.fit,lwd=3,col="blue:")
```

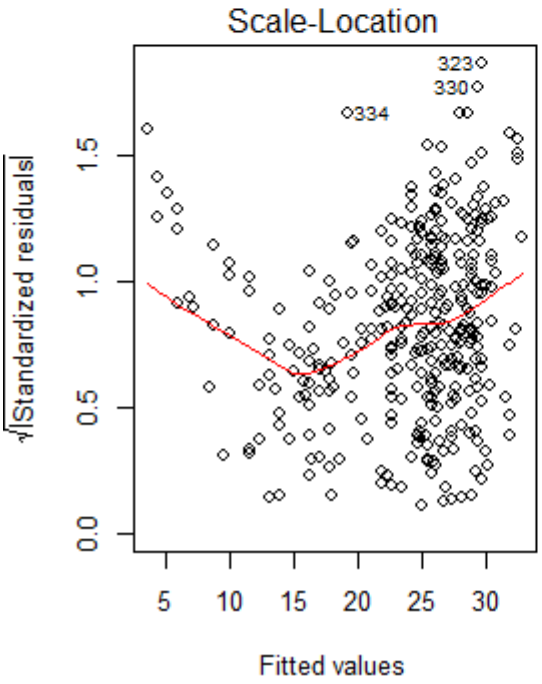


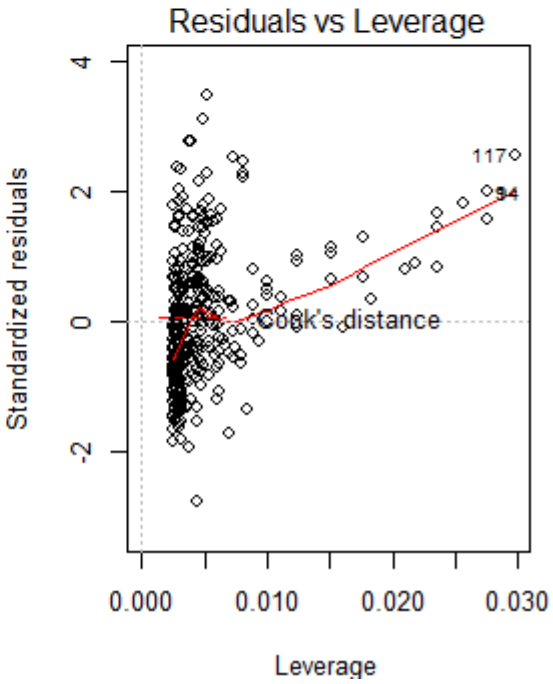
8c.) Use the plot() function to produce diagnostic plots of the least squares regression fit. Comment on any problems you see with the fit.

```
> plot(lm.fit)
```









From the Residuals vs. Leverage plot, we see that there are some high leverage points (94, 117). The Residuals vs. Fitted plot follows a quadratic pattern, which could suggest non-linearity.