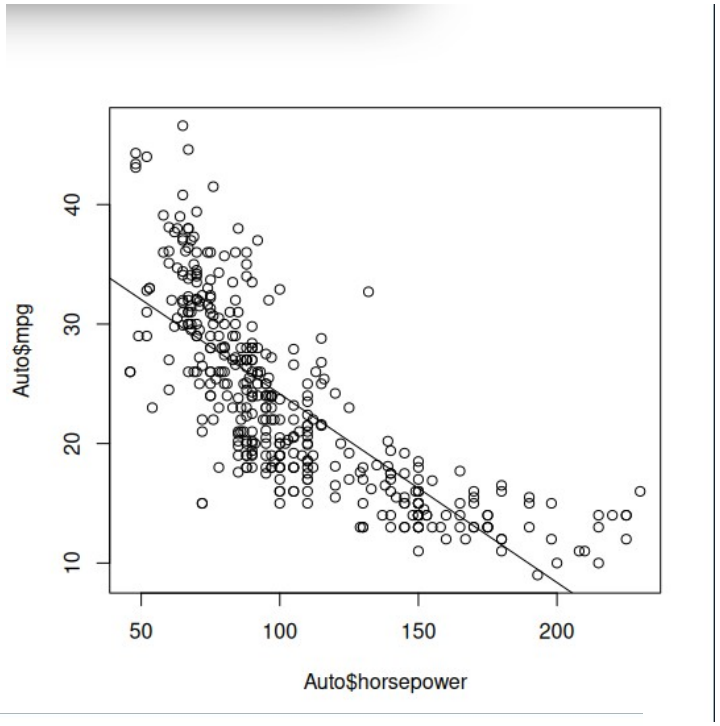
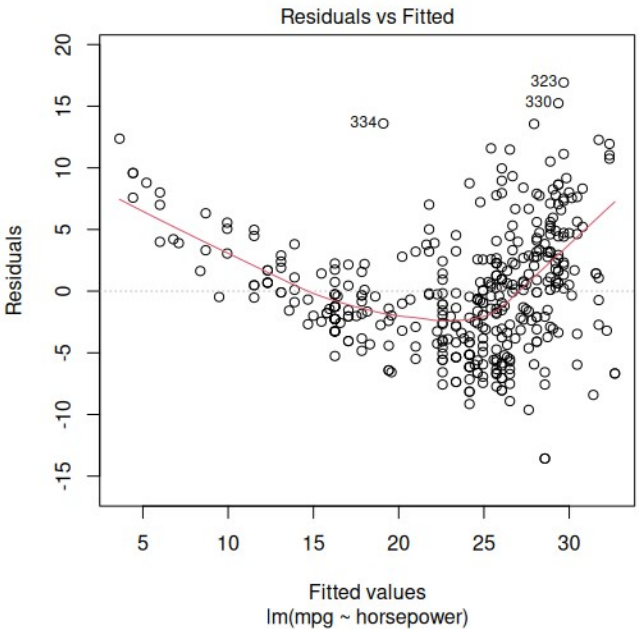
Question 8:

b) I plotted the predictor on the x-axis and the response on the y-axis. Here is the resulting plot:

```
# b)

plot(Auto$horsepower, Auto$mpg, abline(model))
```



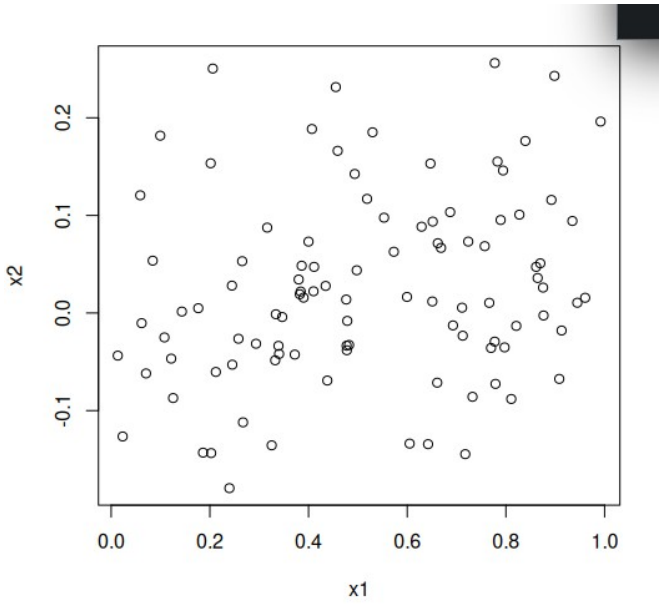
c) Here is the resulting plot of the residuals of the least squared fit. This doesn't look very great. As we increase along the x-axis in this plot, the residuals get much more spread out. We would theoretically like this to remain roughly consistent, however in this case it seems as if a linear model would not be the best fit



Question 14:

a) y = 2 + 2*x1 + .3*x2: B0 = 2, B1 = 2, B2 = .3

b) The correlation is Cor(x1,x2)=.226.
Plot:



c) B0 = 2.13
B1 = 1.75
B2 = .74
These are roughly fairly close to the true B values. The B2 value is pretty off

We can reject the null that B0 and B1 = 0, as we have very low p-values. We cannot reject the null B2 = 0, since we have a p-value of .52

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8311 -0.7273 -0.0537  0.6338  2.3359

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.1305     0.2319   9.188 7.61e-15 ***
x1            1.7589     0.4072   4.319 3.79e-05 ***
x2            0.7397     1.1337   0.652    0.516
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.056 on 97 degrees of freedom
Multiple R-squared:  0.1811,    Adjusted R-squared:  0.1642
F-statistic: 10.73 on 2 and 97 DF,  p-value: 6.191e-05
```

d) The R^2 is slightly worse than model with x1 and x2. Both the intercept and B1 coefficients are both statistically significant

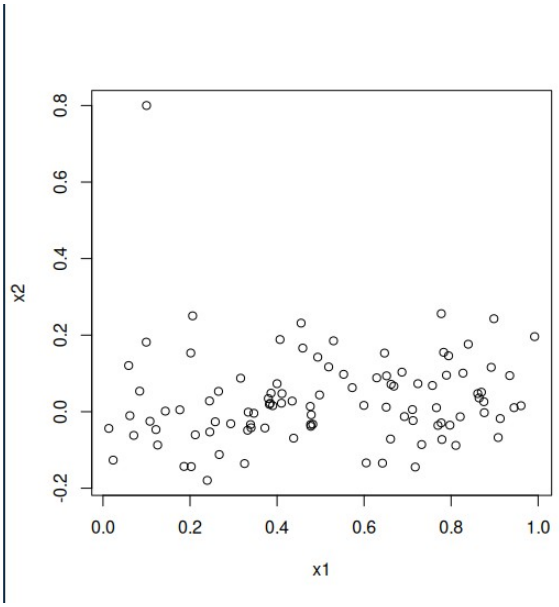
```
Call:
lm(formula = y ~ x1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.87789 -0.68357 -0.07517  0.61429  2.40388

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.1172     0.2303   9.193 6.83e-15 ***
x1            1.8190     0.3955   4.599 1.27e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.053 on 98 degrees of freedom
Multiple R-squared:  0.1775,    Adjusted R-squared:  0.1691
F-statistic: 21.15 on 1 and 98 DF,  p-value: 1.27e-05
```

e) The R^2 is not very good for this model, and B2 is not statistically significant, as we have a very high p-value. This model is not great at all.

```
Call:
lm(formula = y ~ x2)

Residuals:
     Min       1Q   Median       3Q      Max
-3.04598 -0.74058 -0.03127  0.71168  2.99145

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0146     0.1183   25.48   <2e-16 ***
x2            1.8470     1.1997    1.54    0.127
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.147 on 98 degrees of freedom
Multiple R-squared:  0.02361,   Adjusted R-squared:  0.01365
F-statistic:  2.37 on 1 and 98 DF,  p-value: 0.1269
```

f) These results do not contradict each other, B2 is not statistically significant in all instances. The B1 coefficient has been the closest to the actual value and has performed best in the linear model

g) The B2 coefficient now becomes statistically significant in all the models where it is used. Interestingly, in the model with x1 and x2, the estimates for the coefficient are now very different from the actual values, but the models seem to fit a lot better than what they did before adding the point. The new point is definitely an outlier to the rest of the set for the x2 model, and it has a lot of pull on the linear model fit that only uses x2.




```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
     Min       1Q   Median       3Q      Max
-2.72638 -0.68667 -0.04457  0.66586  2.30250

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.2297     0.2318   9.618 8.16e-16 ***
x1            1.5254     0.4007   3.807 0.000245 ***
x2            2.3564     0.8785   2.682 0.008583 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.077 on 98 degrees of freedom
Multiple R-squared:  0.1927,    Adjusted R-squared:  0.1763
F-statistic:  11.7 on 2 and 98 DF,  p-value: 2.778e-05
```