

HW3_632

Prashan A. Welipitiya

3/3/2020

Exercise 1

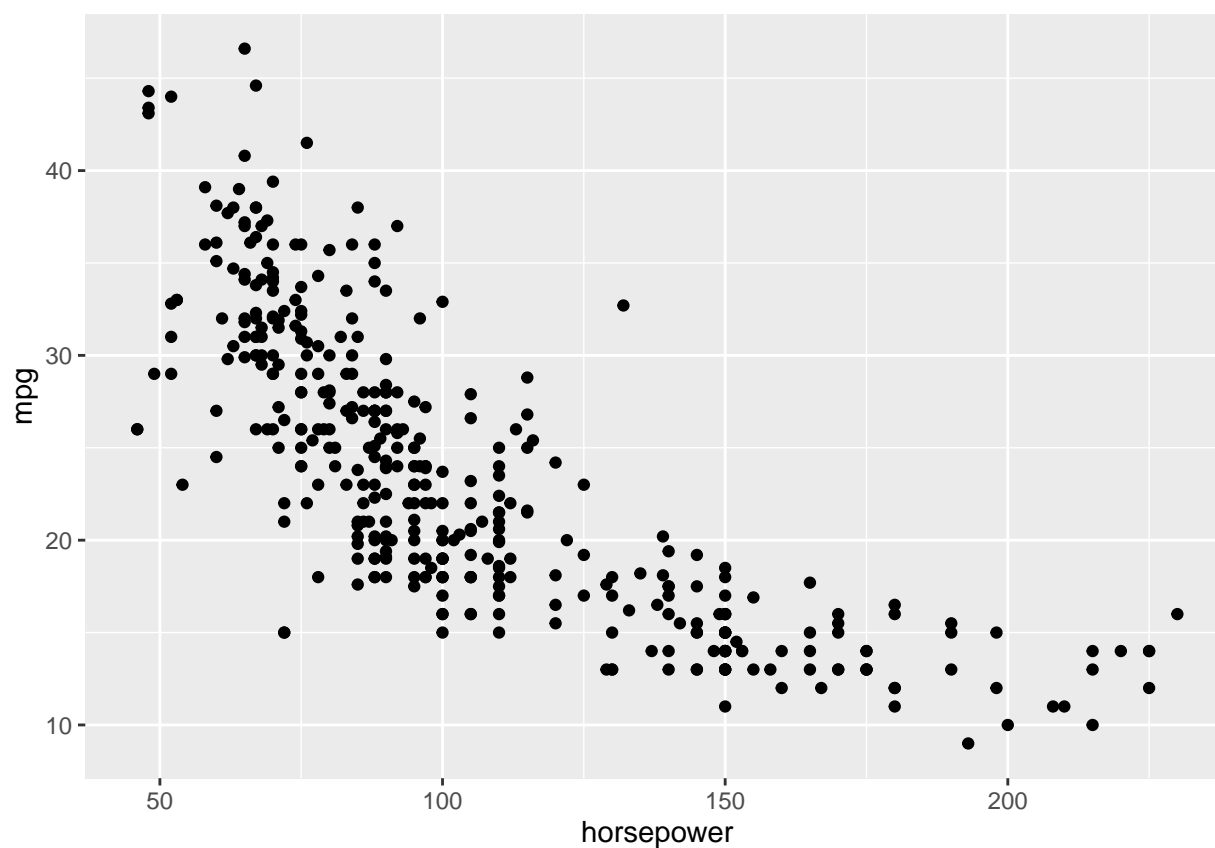
```
library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 3.6.2
```

```
#help(Auto)  
library(ggplot2)
```

a.

```
ggplot(data=Auto, aes(x = horsepower, y = mpg)) + geom_point()
```



b.

```
lm1 = lm(mpg~horsepower + I(horsepower^2), data = Auto)
summary(lm1)

##
## Call:
## lm(formula = mpg ~ horsepower + I(horsepower^2), data = Auto)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7135  -2.5943  -0.0859   2.2868  15.8961
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    56.9000997   1.8004268   31.60  <2e-16 ***
## horsepower     -0.4661896   0.0311246  -14.98  <2e-16 ***
## I(horsepower^2)  0.0012305   0.0001221   10.08  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.374 on 389 degrees of freedom
## Multiple R-squared:  0.6876, Adjusted R-squared:  0.686
## F-statistic:  428 on 2 and 389 DF,  p-value: < 2.2e-16
```

$$Y = 56.9 - 0.4x + 0.0012x^2$$

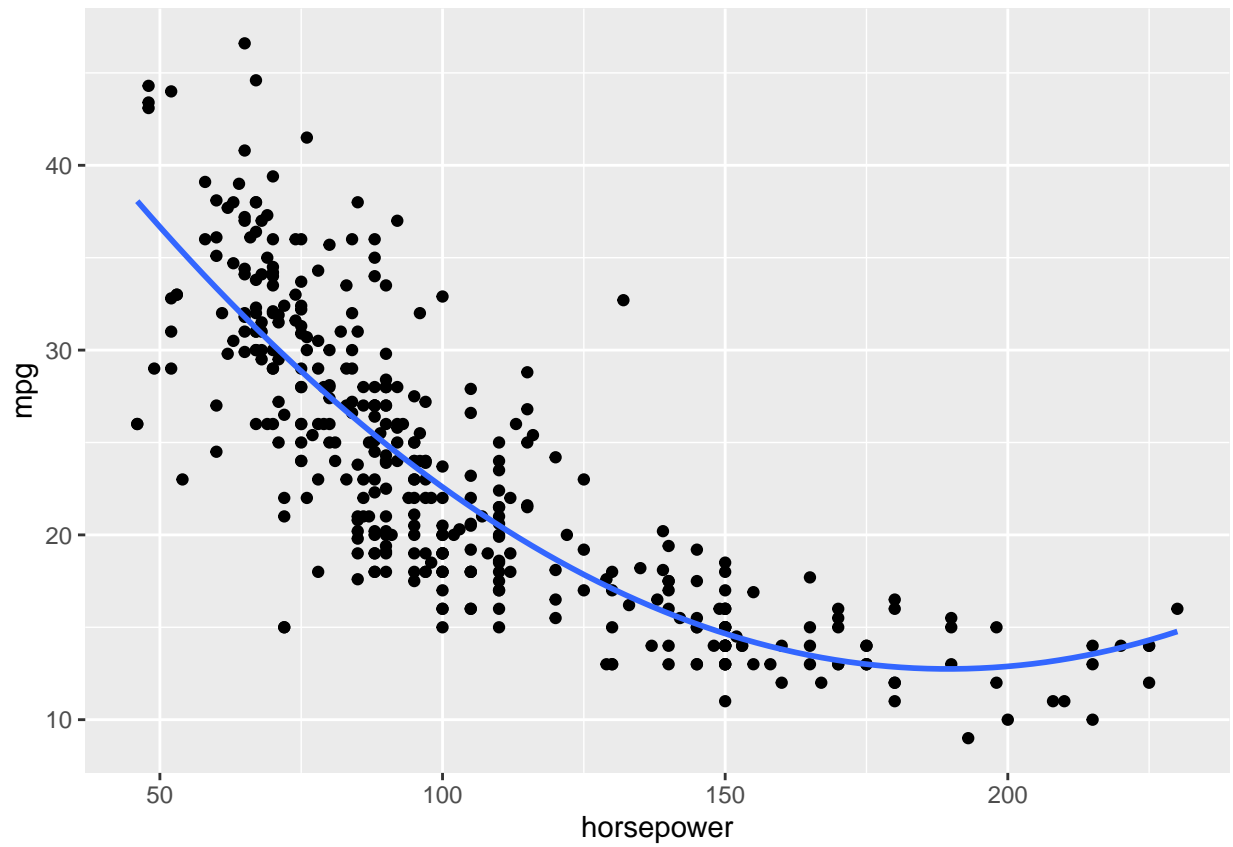
c.

```
x_new = data.frame(horsepower = 150)
predict(lm1, newdata = x_new, interval = "prediction")
```

```
##          fit          lwr          upr
## 1 14.65872  6.027273 23.29016
```

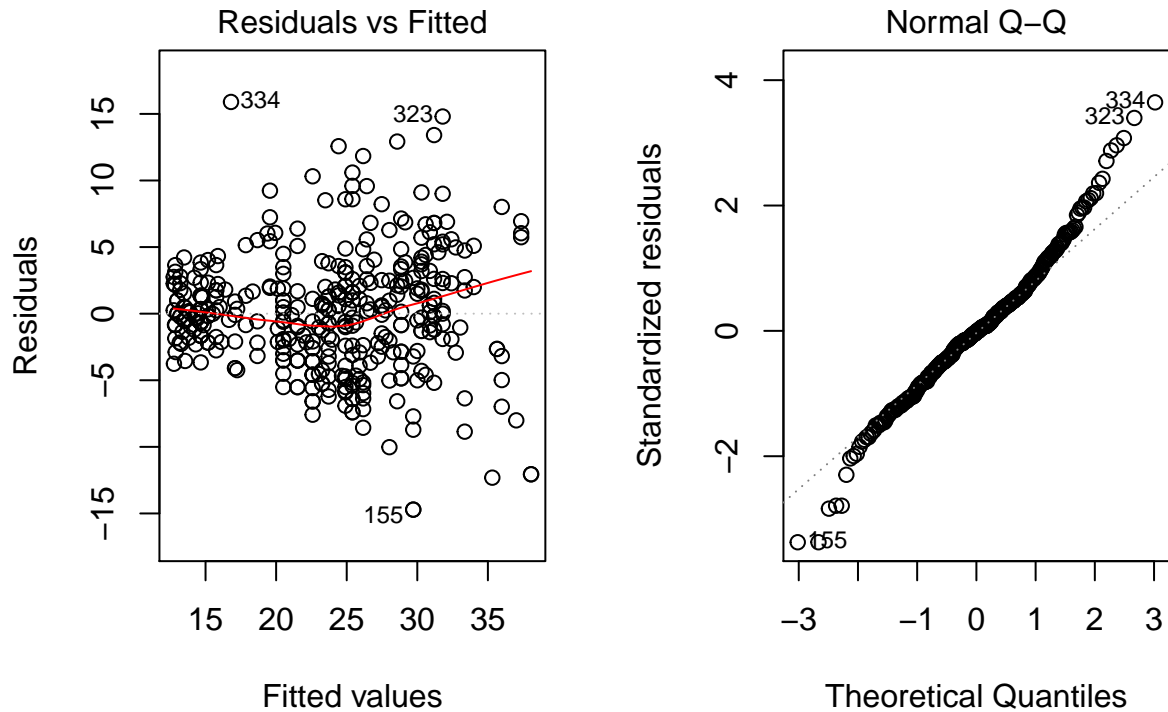
d.

```
ggplot(data=Auto, aes(x = horsepower, y = mpg)) + geom_point() + stat_smooth(method = 'lm', formula = y
```



e.

```
par(mfrow = c(1, 2))  
plot(lm1, 1:2)
```



There is fanning and non constant variance and the residuals deviate from the normal distribution.

Exercise 2

```
#help(Carseats)
```

a.

```
lm2 = lm(Sales~Price + Urban + US, data = Carseats)
summary(lm2)
```

```
##
## Call:
## lm(formula = Sales ~ Price + Urban + US, data = Carseats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.9206 -1.6220 -0.0564  1.5786  7.0581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 13.043469    0.651012   20.036 < 2e-16 ***
## Price      -0.054459    0.005242  -10.389 < 2e-16 ***
## UrbanYes   -0.021916    0.271650   -0.081    0.936
## USYes      1.200573    0.259042    4.635 4.86e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.472 on 396 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2335
## F-statistic: 41.52 on 3 and 396 DF,  p-value: < 2.2e-16
```

b.

A decrease in price by 0.05 is associated with an increase of a sale of one thousand units, when all other predictors are held firm. A location in an urban area on average will sell 20 less units compared to a non urban area. A location in the US on average will sell 1,200 units more than a store outside the US.

c.

$$\hat{y} = 13.04 - 0.05x_1 - 0.02x_2 + 1.2x_3$$

d. The Urban variable.

e.

```
lm3 = lm(Sales~Price + US, data = Carseats)
summary(lm3)
```

```
##
## Call:
## lm(formula = Sales ~ Price + US, data = Carseats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.9269 -1.6286 -0.0574  1.5766  7.0515
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.03079    0.63098  20.652 < 2e-16 ***
## Price      -0.05448    0.00523  -10.416 < 2e-16 ***
## USYes      1.19964    0.25846   4.641 4.71e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.469 on 397 degrees of freedom
## Multiple R-squared:  0.2393, Adjusted R-squared:  0.2354
## F-statistic: 62.43 on 2 and 397 DF,  p-value: < 2.2e-16
```

f.

23.35% of the variation in sales can be explained by the regression model.

g.

```
confint(lm3)
```

```
##                2.5 %      97.5 %  
## (Intercept) 11.79032020 14.27126531  
## Price      -0.06475984 -0.04419543  
## USYes       0.69151957  1.70776632
```