FLS 10 Haitie Liu

Question1:

```
Call:
lm(formula = MPG ~ Weight, data = cars)

Residuals:
     Min       1Q   Median       3Q      Max 
-12.0008  -2.7684  -0.3342   2.1245  16.4920 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 46.2734424  0.7974987   58.02   <2e-16 ***
Weight      -0.0076613  0.0002577  -29.73   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.332 on 392 degrees of freedom
Multiple R-squared:  0.6927,	Adjusted R-squared:  0.6919 
F-statistic: 883.6 on 1 and 392 DF,  p-value: < 2.2e-16

                   2.5 %       97.5 %
(Intercept) 44.705532760 47.841351974
Weight      -0.008168061 -0.007154609
```

After fitting the model, we have the image to the left, now the following the 6 step hypothesis test.

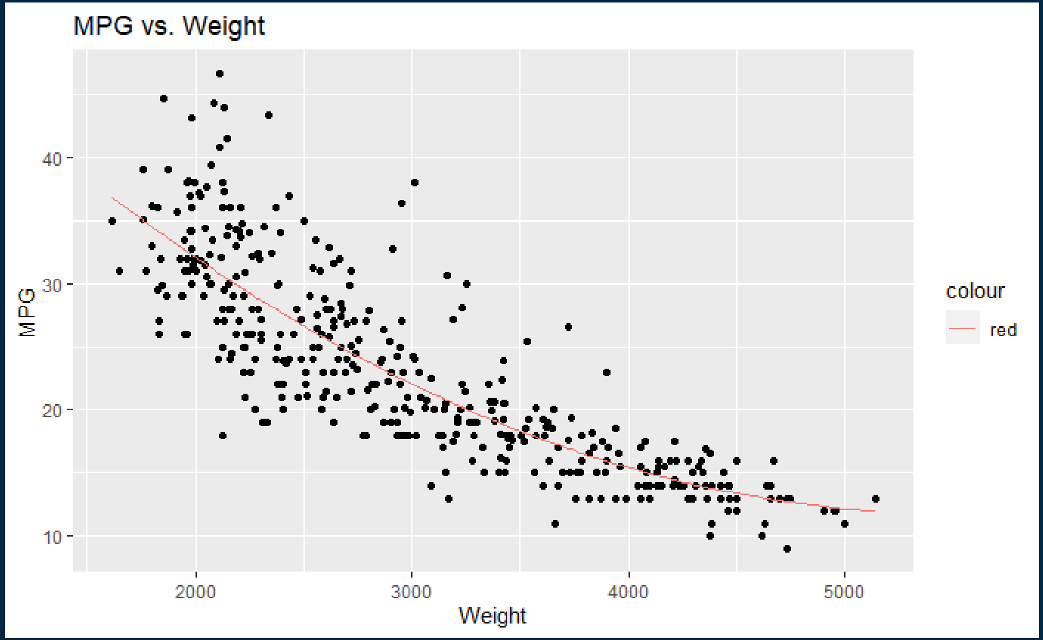
Step1: H0: slope = 0, H1: slope ≠ 0
Step2: Significance level 5%
Step3: Find test statistics t=-29.73
Step4: Find P value P= 2e-16 ***
Step5: Reject H0
Step 6: Overwhelming evidence suggests that slope ≠ 0

95% Confidence interval:

We are 95% confident that when Weight increase by 1 unit, the mpg of the according vehicle decreases by (0.008168061, 0.007154609)

Question 2:

```
#a)
fit=train(MPG~Weight,method ="lm", data = cars, trControl=trainControl(method = "LOOCV"))
summary(fit)
#b)
cars=cars%>%mutate(Weight2=Weight**2)
fit2=train(MPG~Weight+Weight2,method="lm",data= cars,trControl=trainControl(method="LOOCV"))
summary(fit2)
```



A) Fitting model using leave one out by (method="LOOCV") in the caret packet.

B) Fit2 has lower RMSE, therefore more preferrable.

C) Description: when Weight^2 is constant, for every unit increase of Weight, mpg decreases 0.01848, when Weight is constant, for every unit increase in Weight^2, mpg increases 0.000001692.

D) When weight is 2000, mpg is estimated to be **32.07914**

Question 3:

```
#Impute (predict and insert) the missing horsepowers by fitting a regression model.
is.na(cars$Horsepower)
fit3=lm(Horsepower~Weight,data = cars)
summary(fit3)

test=data.frame(Weight=c(cars$Weight[351],cars$Weight[371]))
imputation=predict(fit3,newdata =test)

#78.7072 106.7188
cars$Horsepower[351]=imputation[1]
cars$Horsepower[371]=imputation[2]
```

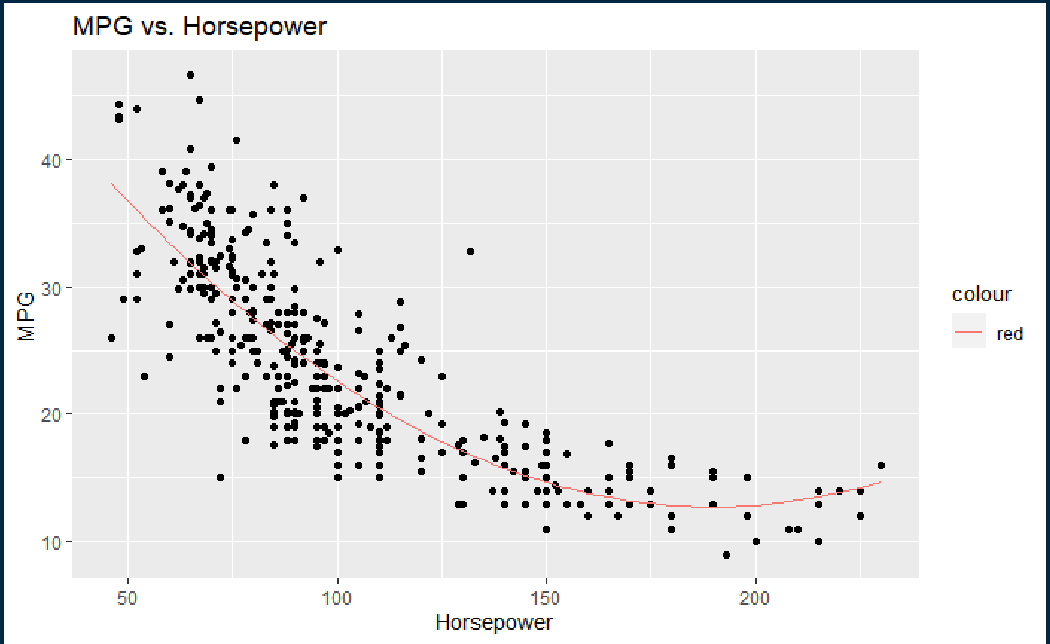
```
Call:
lm(formula = MPG ~ Horsepower + Horsepower2, data = cars)

Residuals:
     Min       1Q   Median       3Q      Max 
-14.7437  -2.6105  -0.0678   2.2685  15.8808 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 56.939707   1.799882   31.64   <2e-16 ***
Horsepower  -0.466258   0.031112  -14.99   <2e-16 ***
Horsepower2  0.001230   0.000122   10.08   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 4.377 on 391 degrees of freedom
Multiple R-squared:  0.6872,	Adjusted R-squared:  0.6856 
F-statistic: 429.4 on 2 and 391 DF,  p-value: < 2.2e-16

                    2.5 %       97.5 %
(Intercept) 53.4010501967 60.478364588
Horsepower  -0.5274259571 -0.405090819
Horsepower2  0.0009897377  0.001469582
```

A)Imputing using model between Horsepower and Weight, insert value of 78.7 and 106.7 as Horsepower using weight

B)Fitting the model using horsepower and horsepower^2

C)When Horsepwer2 is constant, when Hoursepower increases 1 unit, mpg is likely to decrease between 0.5274259571, 0.405090819 with 95% confidence level, when horsepower is constant, every 1 unit of increase in horsepower2, mpg is likely to increase between 0.0009897377 0.001469582, with 95% confidence interval.

Plot (Horsepower vs MPG



D)Here is the fitment of my model, (MPG ~ Horsepower + Horsepower2)

We can see that using a two degree model, when Horsepower is over 200, the line starting to bend upwards.