DMPM Assignment 2 Part 1

Name: Rushikesh Jyoti

Division: A

Roll no: 27

SRN: 201901139

Question: Build a simple linear regression model using the given data-set

(a) Print model summary and (b) Plot residual plot (c) Plot scatter plot showing the fitted line.

Perform this for 2 cases as below

Case 1: response = Wt and predictor = Ht

Case 2: response = Wt and predictor = Age

Code

```
df = read.csv("HT-WT-Age.csv")[2:4]


head(df)

summary(df)



model1 = lm(df$Weight ~ df$Height)

model1



pred1 = predict(model1)

resd1 = residuals(model1)
```

```
summary(model1)



plot(df$Weight, df$Height,
    main = "Height and Weight",
    abline(lm(df$Height ~ df$Weight)),
    ylab = "Height in cm",
    xlab = "Weight in kg"
)

plot(df$Height, resd1,
    main = "Residual Plot(HT and WT)",
    abline(0,0),
    ylab = "Residuals",
    xlab = "Height in cm"
)




model2 <- lm(df$Weight ~ df$Age)

print(model2)
print(summary(model2))

pred2 <- predict(model2)
resd2 <- residuals(model2)

print(pred2)
```

```
print(resd2)


plot(df$Weight,
    df$Age,
    main = "Age and Weight",
    abline(lm(df$Age~df$Weight)),
    ylab = "Age in years",
    xlab = "Weight in kg"
)


plot(df$Age,
    resd2,
    main = "Residual Plot(Age and WT)",
    abline(0,0),
    ylab = "Residuals",
    xlab = "Age in years"
)
```

Output

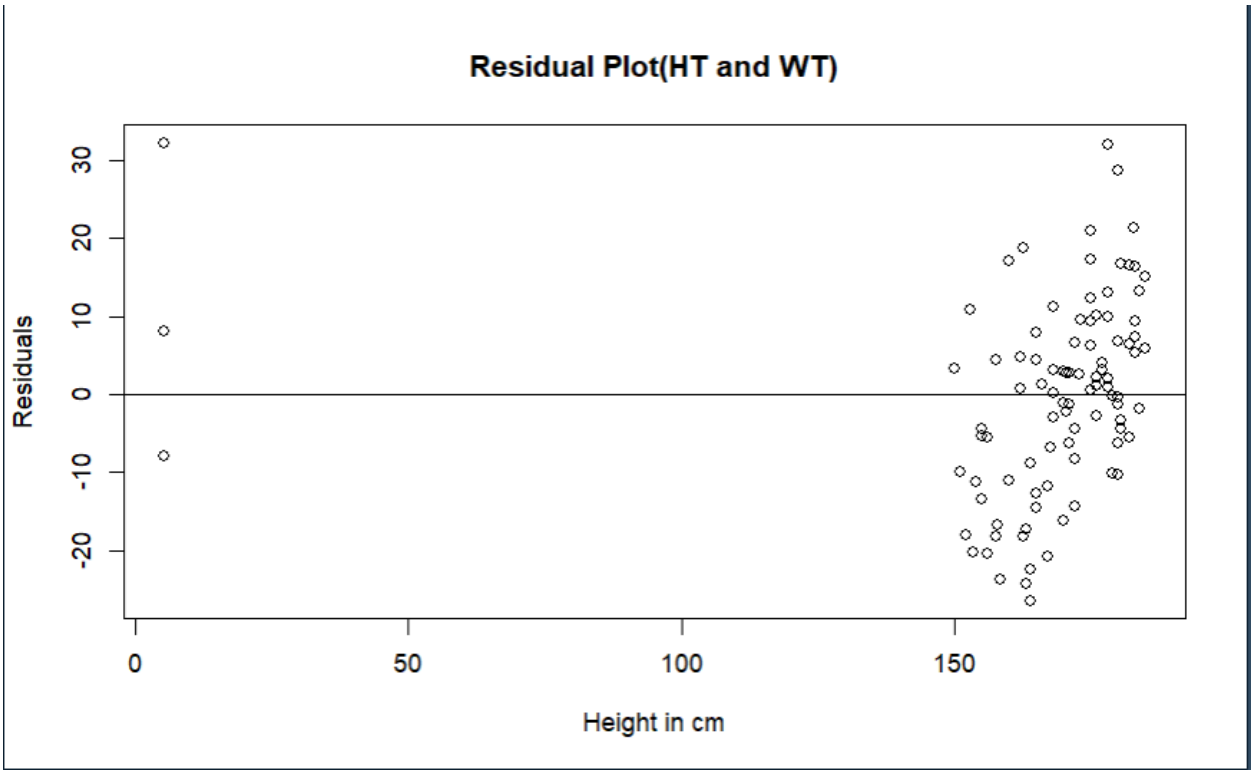
a) Summary

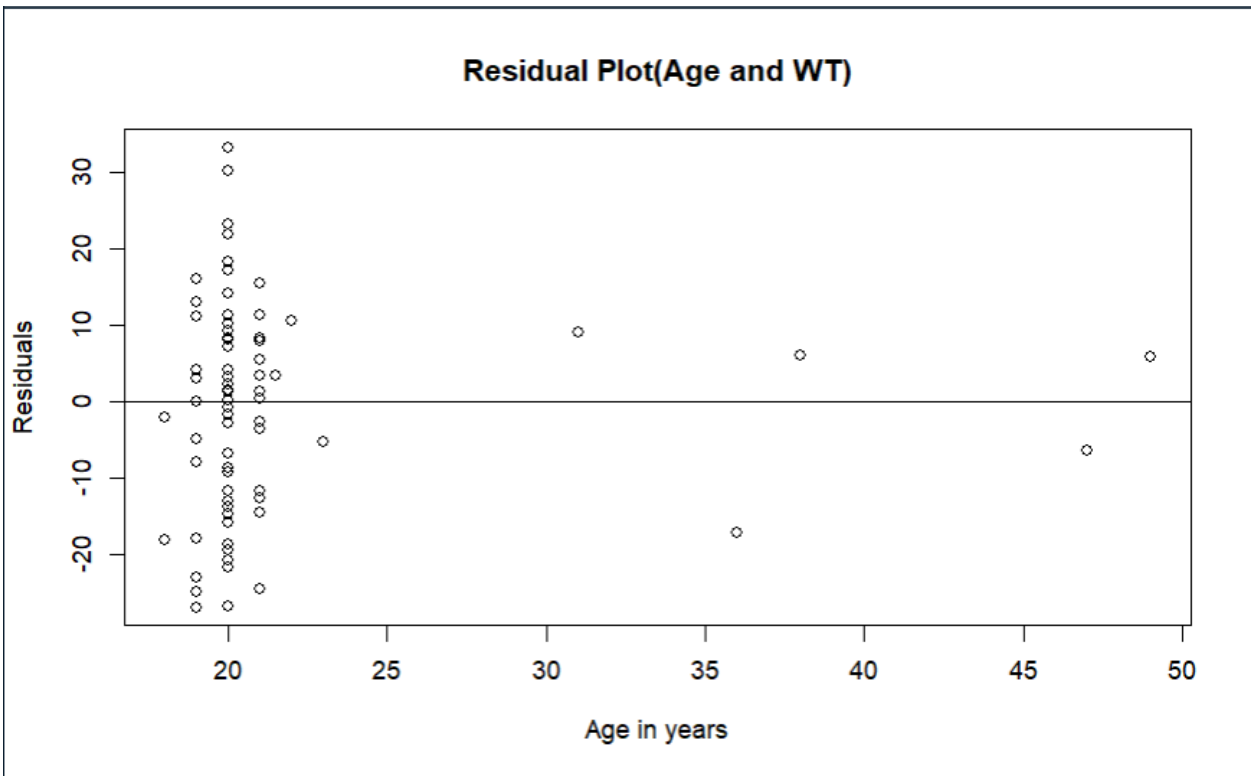
```
> summary(df)
     Height          Weight          Age       
 Min.   :  5.11   Min.   : 40.00   Min.   :18.00  
 1st Qu.:163.00   1st Qu.: 58.00   1st Qu.:20.00  
 Median :171.00   Median : 68.00   Median :20.00  
 Mean   :165.44   Mean   : 66.58   Mean   :21.14  
 3rd Qu.:178.00   3rd Qu.: 75.00   3rd Qu.:21.00  
 Max.   :185.00   Max.   :100.00   Max.   :49.00  
```

```
> model1 = lm(df$Weight ~ df$Height)
> summary(model1)

Call:
lm(formula = df$Weight ~ df$Height)

Residuals:
    Min      1Q  Median      3Q     Max 
-26.409  -8.344   0.825   7.369  32.176 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 47.22645    7.12076   6.632 2.01e-09 ***
df$Height    0.11696    0.04235   2.762  0.00689 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.55 on 95 degrees of freedom
Multiple R-squared:  0.07434,   Adjusted R-squared:  0.06459 
F-statistic: 7.629 on 1 and 95 DF,  p-value: 0.006895
```

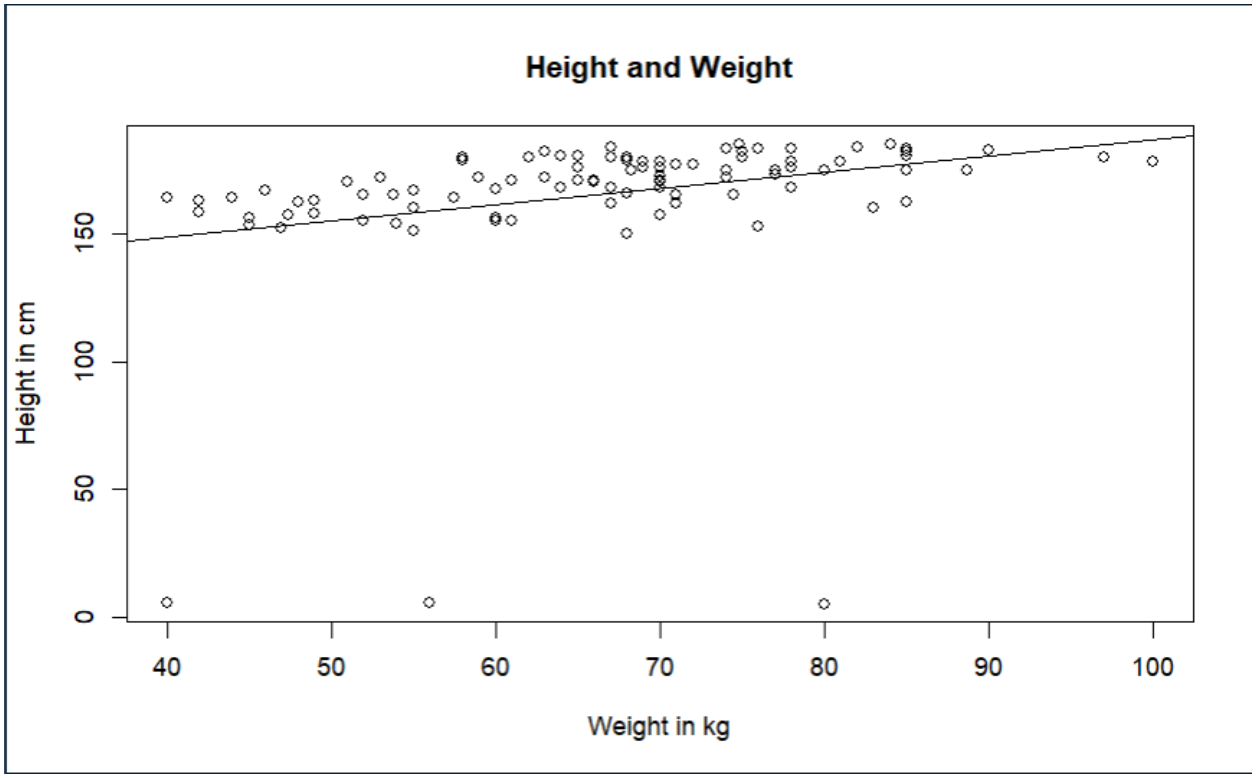
```
> model2 ← lm(df$Weight ~ df$Age)
> print(summary(model2))

Call:
lm(formula = df$Weight ~ df$Age)

Residuals:
    Min      1Q  Median      3Q     Max
-26.917  -8.758   1.242   8.401  33.242

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   69.939      6.004  11.649   <2e-16 ***
df$Age        -0.159      0.277  -0.574    0.567
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.02 on 95 degrees of freedom
Multiple R-squared:  0.003456,  Adjusted R-squared:  -0.007034
F-statistic: 0.3294 on 1 and 95 DF,  p-value: 0.5674
```

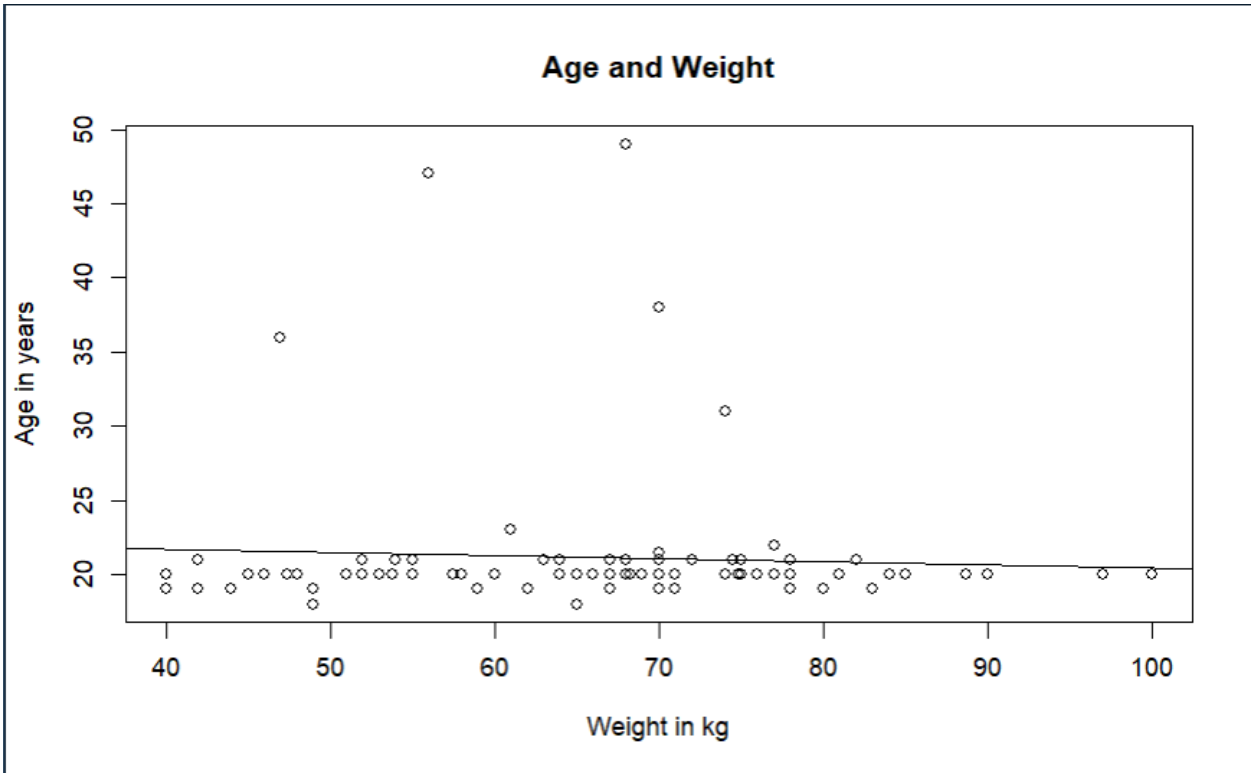
b) Residual Plot





c) Scatter Plot




Age and Weight
Age in years
20
25
30
35
40
45
50
40
50
60
70
80
90
100
Weight in kg