Lab_13

```r
library(MATH4753GALEXFORCELAB13)
```

```r
library(dplyr)
#>
#> Attaching package: 'dplyr'
#> The following objects are masked from 'package:stats':
#>
#>     filter, lag
#> The following objects are masked from 'package:base':
#>
#>     intersect, setdiff, setequal, union
library(ggplot2)
```

Task 1

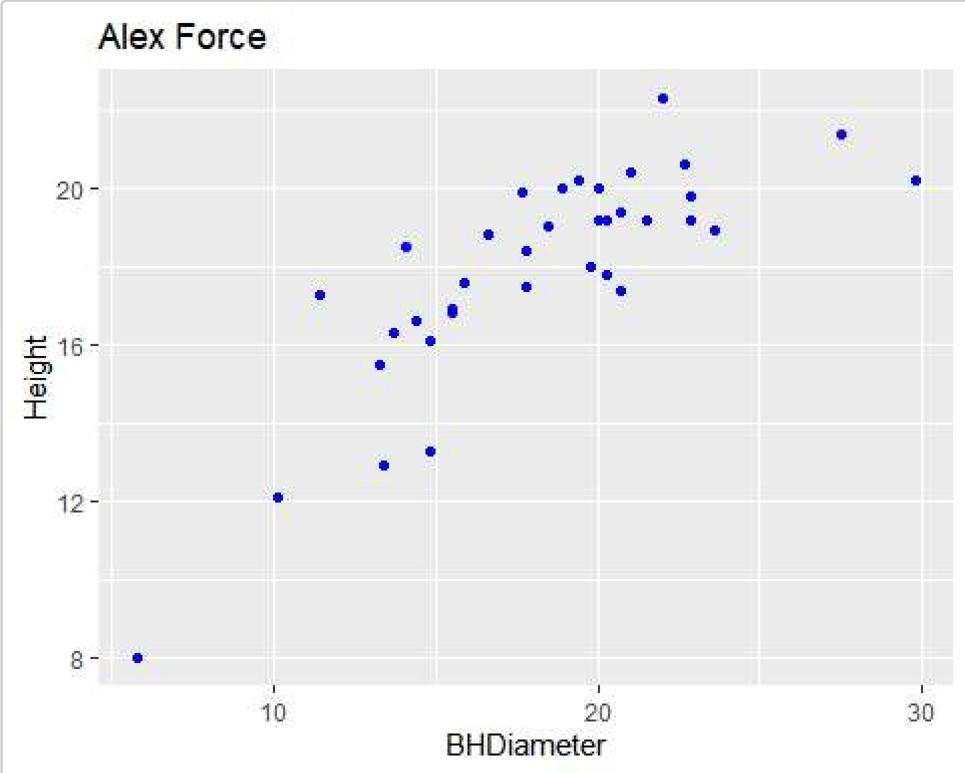
```r
spruce <- read.csv("C:/Users/tcaal/Desktop/R/4753/Module 13/SPRUCE.csv")

head(spruce, 6)
#>   BHDiameter Height
#> 1       18.9   20.0
#> 2       16.6   18.8
#> 3       15.5   16.8
#> 4       15.5   16.9
#> 5       19.4   20.2
#> 6       13.7   16.3
```

Task 2

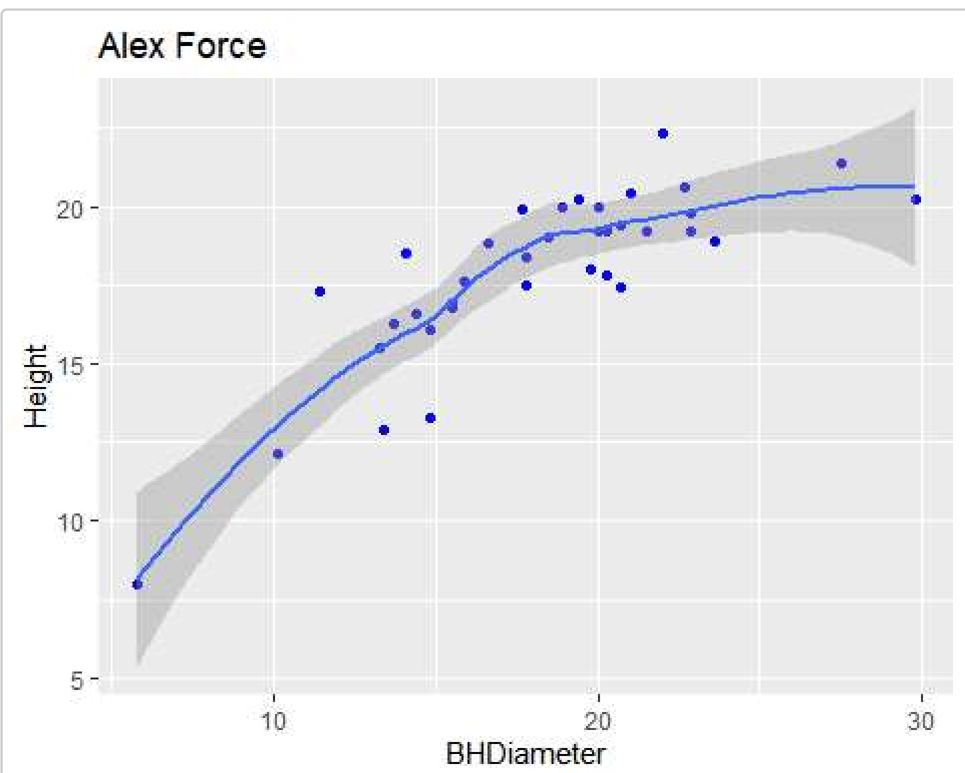
```r
plt <- spruce |>
  ggplot(aes(x = BHDiameter, y = Height)) +
  geom_point(col = 'blue') +
  ggtitle("Alex Force")

plt
```



Task 3

```
plt2 <- plt + geom_smooth(method = 'loess', formula = 'y ~ x')
plt2
```



From the plot, the data suggests a curve rather than a straight line.

Task 4

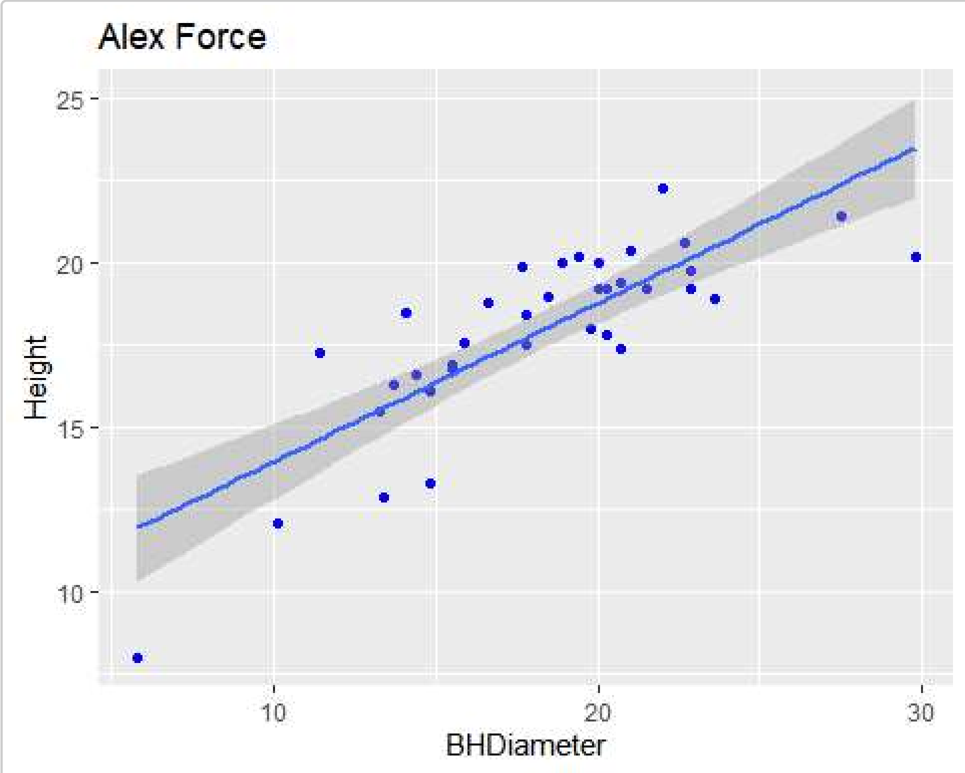
```
ylm <- lm(formula = Height ~ BHDiameter, data = spruce)
summary(ylm)
#> 
#> Call:
#> lm(formula = Height ~ BHDiameter, data = spruce)
#> 
#> Residuals:
#>     Min      1Q  Median      3Q     Max 
#> -3.9394 -0.9763  0.2829  0.9950  2.6644 
#> 
#> Coefficients:
#>             Estimate Std. Error t value Pr(>|t|)    
#> (Intercept)  9.14684    1.12131   8.157 1.63e-09 ***
#> BHDiameter   0.48147    0.05967   8.069 2.09e-09 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#> 
#> Residual standard error: 1.678 on 34 degrees of freedom
#> Multiple R-squared:  0.6569, Adjusted R-squared:  0.6468 
#> F-statistic:  65.1 on 1 and 34 DF,  p-value: 2.089e-09
```

$$\hat{\beta}_0 = 9.14684$$
$$\hat{\beta}_1 = 0.48147$$

$$\begin{aligned} \hat{Height} &= \hat{\beta}_0 + \hat{\beta}_1 BHDiameter \\ &= 9.14684 + 0.48147 BHDiameter \end{aligned}$$

Task 5

```
plt + stat_smooth(method = 'lm', formula = 'y ~ x')
```

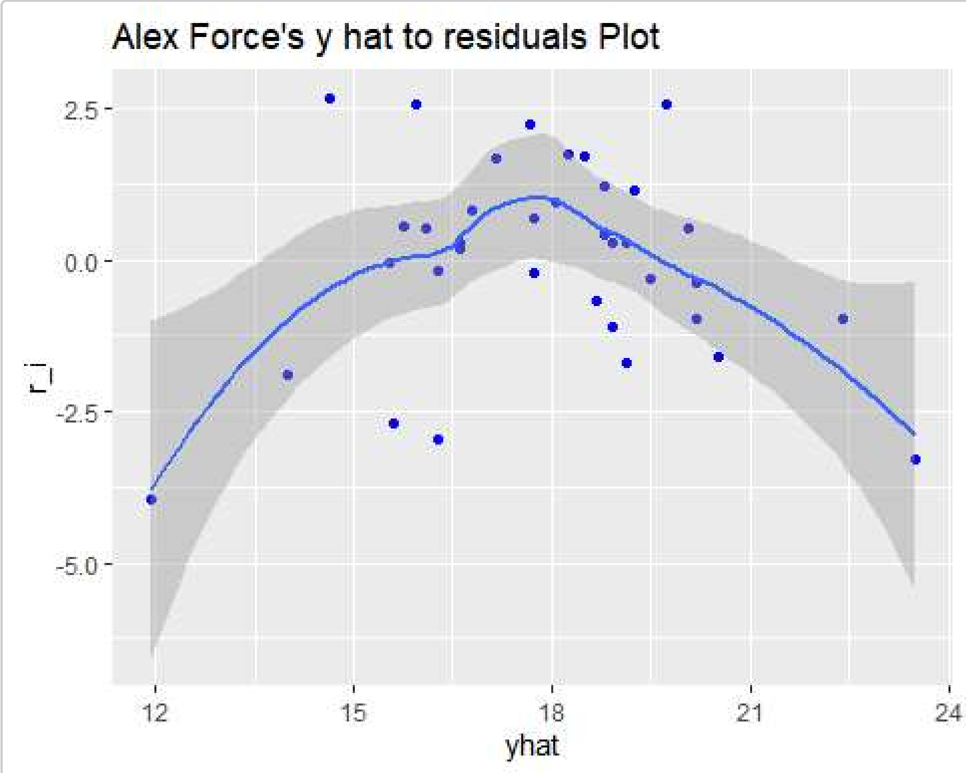


Task 6

```
spruce2 <- spruce |>
  mutate(r_i = residuals(ylm), yhat = 9.14684 + 0.48147 * BHDiameter)

plt3 <- spruce2 |>
  ggplot(aes(x = yhat, y = r_i)) +
  geom_point(col = 'blue') +
  geom_smooth(method = 'loess', formula = 'y ~ x') +
  ggtitle("Alex Force's y hat to residuals Plot")

plt3
```



(a)

The first assumption made in the linear model states that the mean of the probability distribution for the random error (epsilon) is 0, this then implies that the mean value of y is:

$$E(y) = \beta_0 + \beta_1 x$$

This in turn forms a linear combination of betas. The second assumption states that the variance for epsilon is constant for all independent x values, this is equivalent to a constant for all x values too. These 2 assumptions contribute to epsilon then takes a normal distribution. These assumptions are contrasted with the diagnostic shown above where a non normal distribution is taken by epsilon.

(b)

The plot takes on a quadratic rather than a linear fit.

Task 7

(a)

The “I” function change the class of an object indicating it is treated “as is”, when applied in terms of model formulae the I() function ensures that the operator is being used in the arithmetic way. the lm() function is structured as “response ~ terms” where the term component is structured as a series of terms which are used as a predictor to response. For the lm() function to operate properly, I() is used to interpret each component properly (intercept, BH Diameter, and BH Diameter^2).

(b)

```
ylm2 <- lm(formula = Height ~ BHDiameter + I(BHDiameter^2), data = spruce)

summary(ylm2)
#>
#> Call:
#> lm(formula = Height ~ BHDiameter + I(BHDiameter^2), data = spruce)
#>
#> Residuals:
#>     Min      1Q  Median      3Q     Max
#> -3.2966 -0.6245 -0.0707  0.7442  3.2541
#>
#> Coefficients:
#>                  Estimate Std. Error t value Pr(>|t|)
#> (Intercept)      0.860896   2.205022   0.390 0.698731
#> BHDiameter       1.469592   0.243786   6.028 8.88e-07 ***
#> I(BHDiameter^2) -0.027457   0.006635  -4.138 0.000227 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 1.382 on 33 degrees of freedom
#> Multiple R-squared:  0.7741, Adjusted R-squared:  0.7604
#> F-statistic: 56.55 on 2 and 33 DF,  p-value: 2.182e-11
```

Task 8

(a)

R^2 value for model #1: 0.6468 R^2 value for model #2: 0.7604

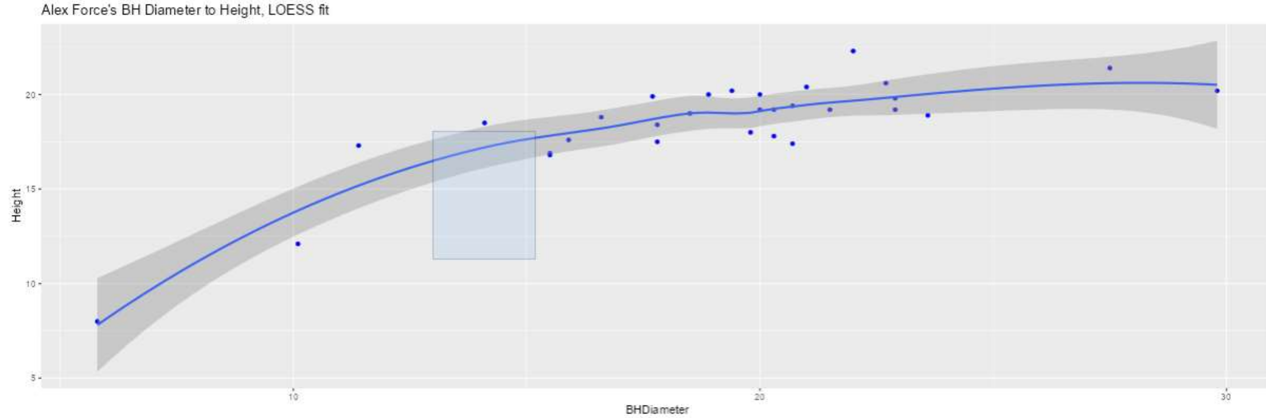
The second model, using the quadratic approach, has the R^2 value closest to 1 indicating that this model is the “best” of the 2.

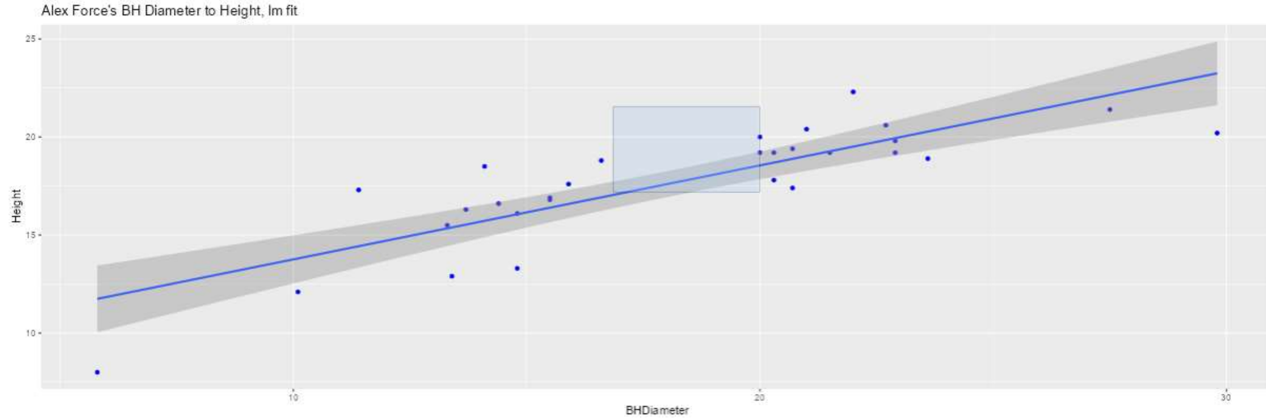
(b)

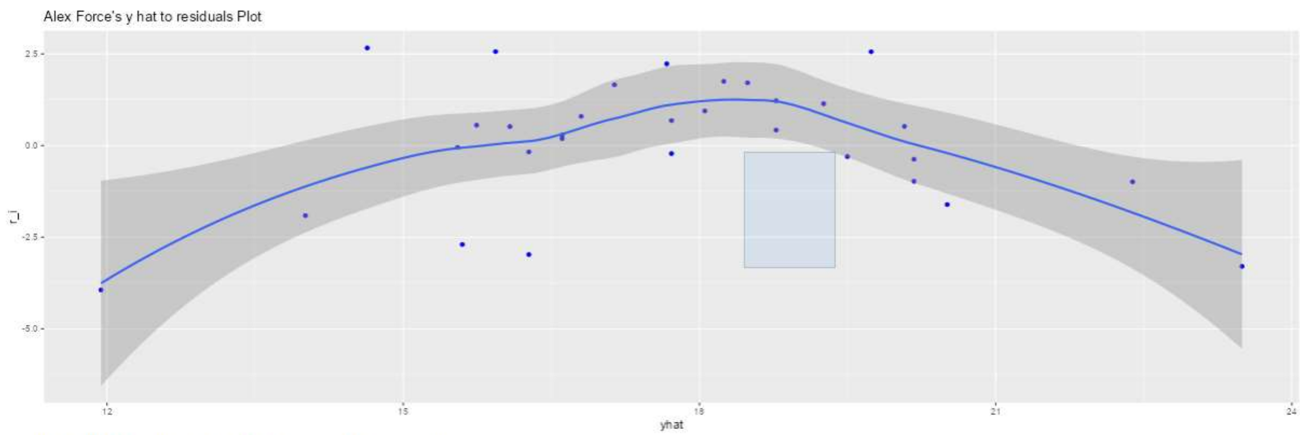
$$\hat{\beta}_0 = 0.860896$$
$$\hat{\beta}_1 = 1.469592$$
$$\hat{\beta}_2 = -0.027457$$

$$\begin{aligned} \hat{Height} &= \hat{\beta}_0 + \hat{\beta}_1 BHDiameter + \hat{\beta}_2 BHDiameter^2 \\ &= 0.860896 + 1.469592 BHDiameter - 0.027457 BHDiameter^2 \end{aligned}$$

Task 9







Estimating formula for linear fit: Height = 8.9656 + 0.4792 BH Diameter

Estimating formula for quadratic fit: Height = -0.0149 + 1.5748 BH Diameter-0.0299 BH Diameter^2