

PSTAT 126 Project 4

Melanie Flandes

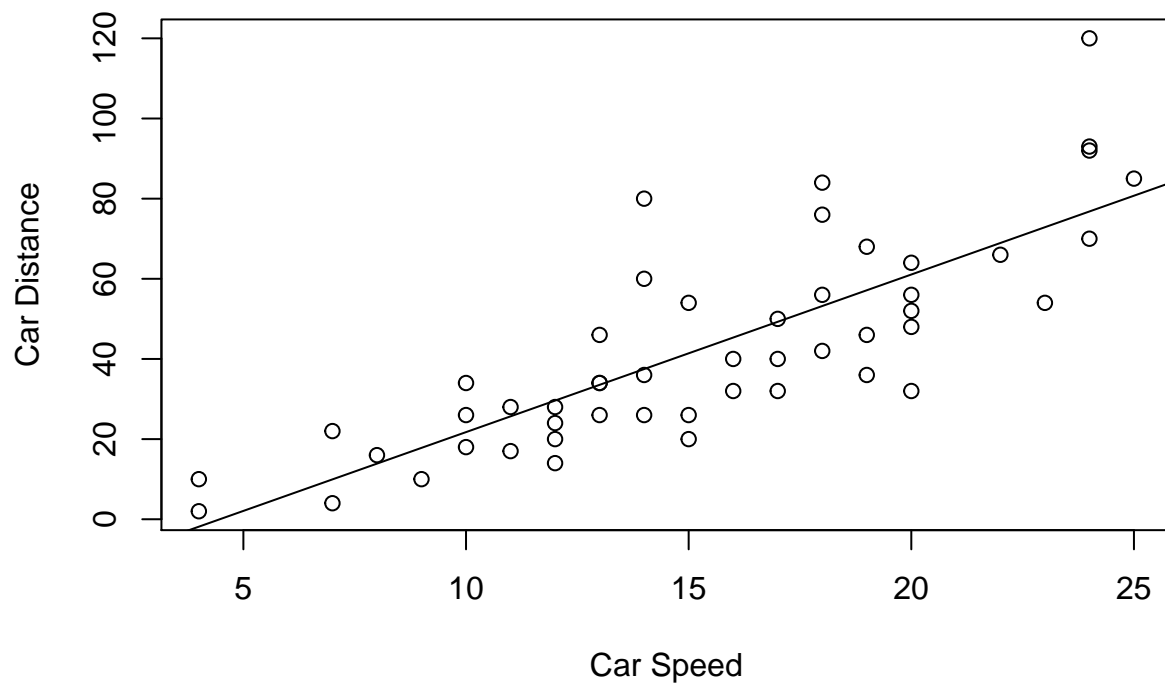
2/23/2018

Problem 1

```
library(datasets)
data(cars)
```

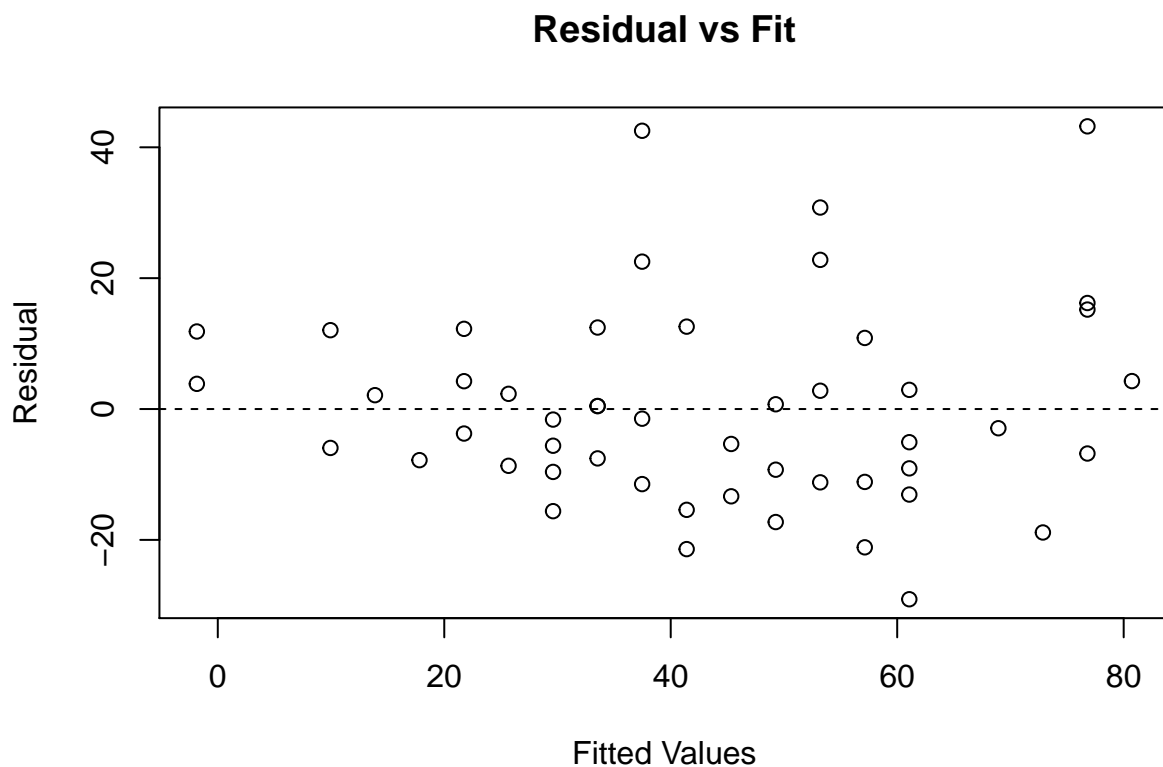
Part a | Plotting Fitness

```
y = cars$dist
x = cars$speed
fit = lm(y~x)
plot(x,y,xlab = "Car Speed", ylab = "Car Distance")
abline(fit$coefficients[1],fit$coefficients[2])
```



Part b | Residual vs Fit Plot

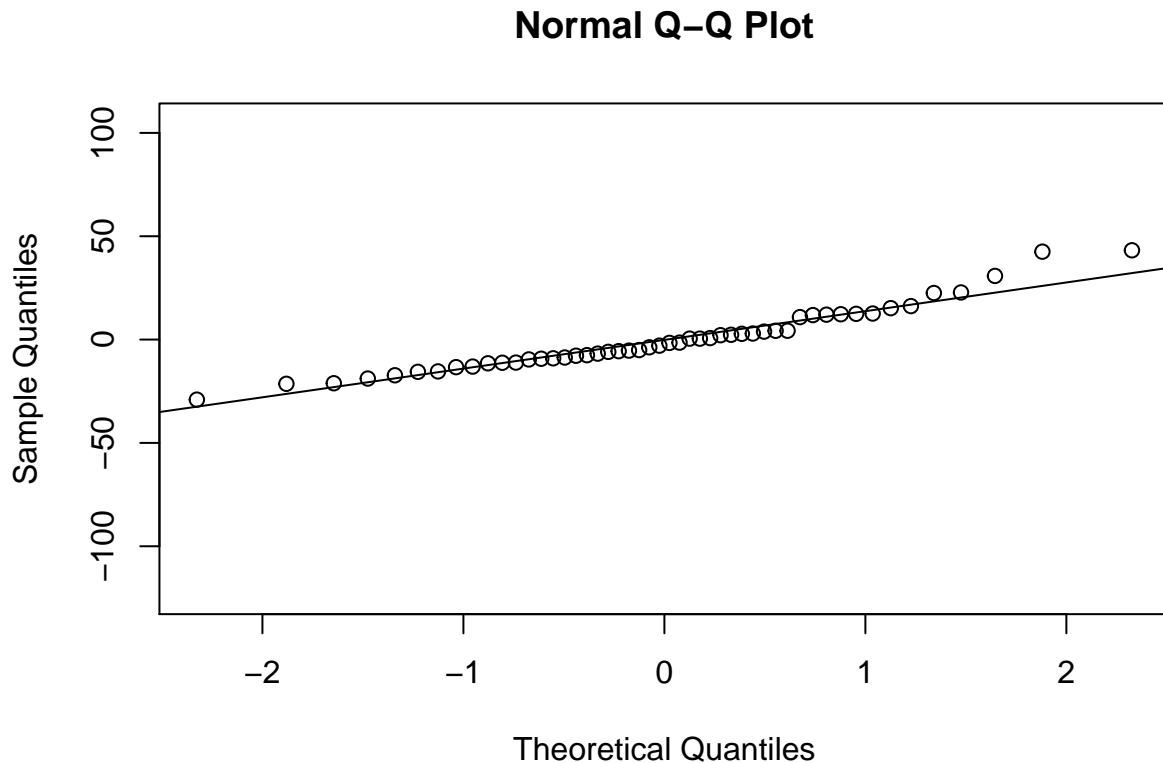
```
yhat = fitted(fit)
e = y - yhat
plot(yhat, e, xlab = "Fitted Values", ylab = "Residual", main = "Residual vs Fit")
abline(h = 0, lty = 2)
```



The “residual vs fit” plot does not have a non-constant variance and there is no non-linearity problem because it does seem to have some funneling but it is not too severe therefore we can still consider the plot to be linear. The main problem is that this plot gives us no more new information.

Part c | Q-Q Plot

```
qqnorm(e, ylim = c(-123.7, 105.1))  
qqline(e)
```



The Q-Q plot meets the normality assumption because the Residual vs Fit plot is linear and looking at the Q-Q it get more linear. Which means the normal Q-Q plot of the residuals is approximately linear supporting that the error terms are normally distributed.

Part d | Shapiro Test

H_0 : Variable is normally distributed

H_1 : Variable is not normally distributed

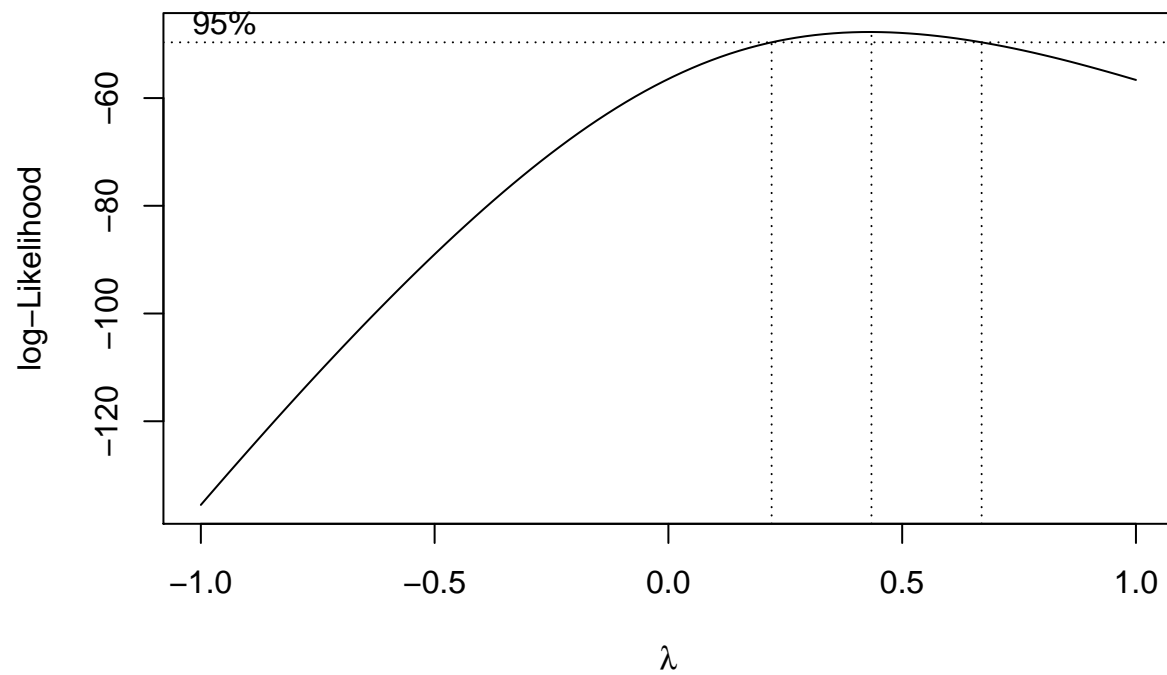
$\alpha = .05$

```
shapiro.test(e)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  e
## W = 0.94509, p-value = 0.02152
```

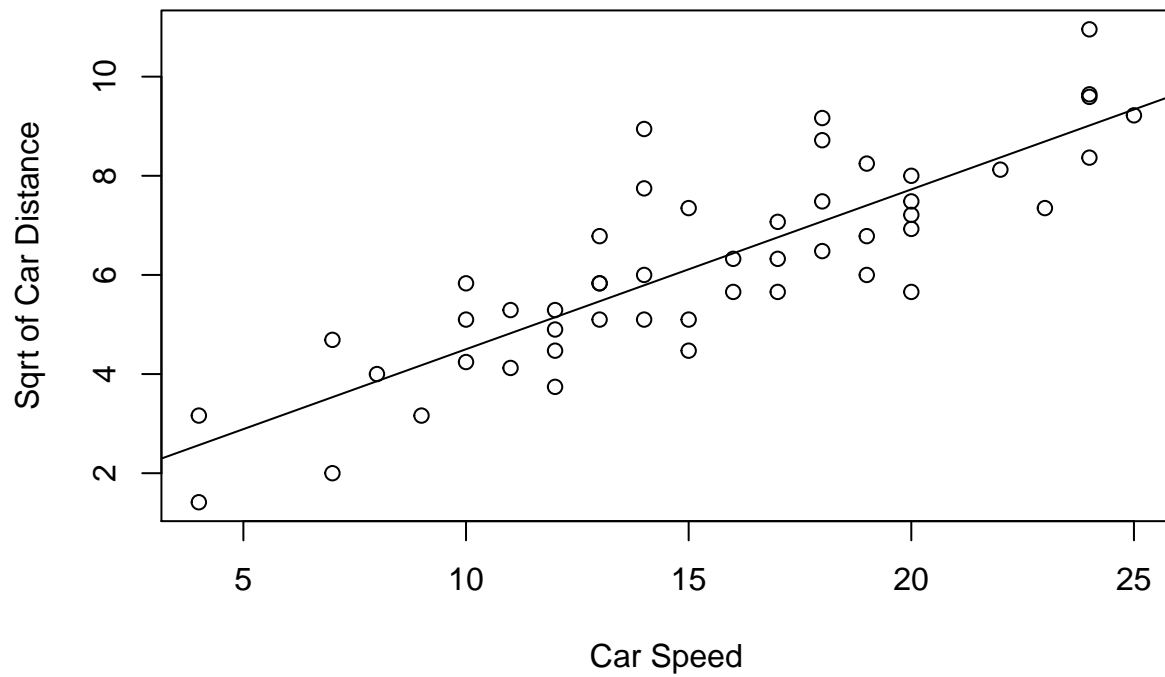
Since our p-value is 0.02152 it is smaller than our alpha value of .05, which means we fail to reject the null hypothesis. Therefore there is enough significant evidence to conclude that the variable is normally distributed. This Shapiro Test does support our conclusion from part c.

```
library(MASS)
boxcox(y ~ x, lambda = seq(-1,1,length = 10))
```



Part e | Fitness plot

```
y = sqrt(cars$dist)
fit1 = lm(y~x)
plot(x,y,xlab = "Car Speed", ylab = "Sqrt of Car Distance")
abline(fit1$coefficients[1],fit1$coefficients[2])
```

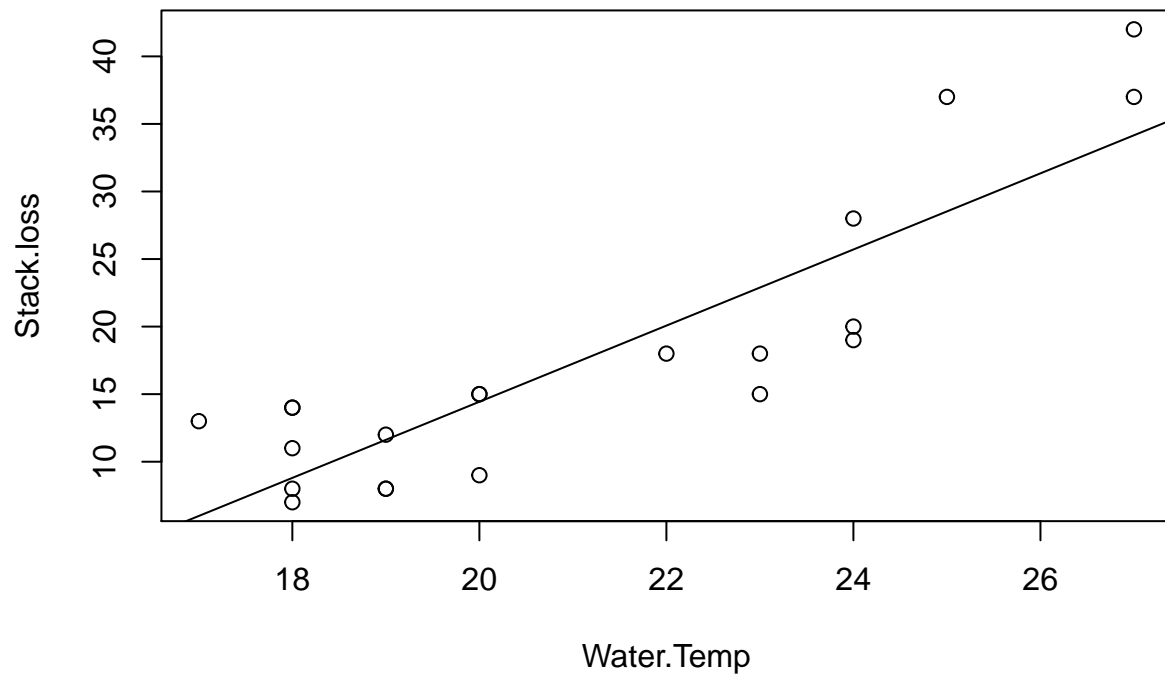


Problem 2

```
library(datasets)
data("stackloss")
```

Part a | Coefficient of Determination

```
y = stackloss$stack.loss
x = stackloss$Water.Temp
fit2 = lm(y~x)
plot(x,y,xlab = "Water.Temp", ylab = "Stack.loss")
abline(fit2$coefficients[1],fit2$coefficients[2])
```



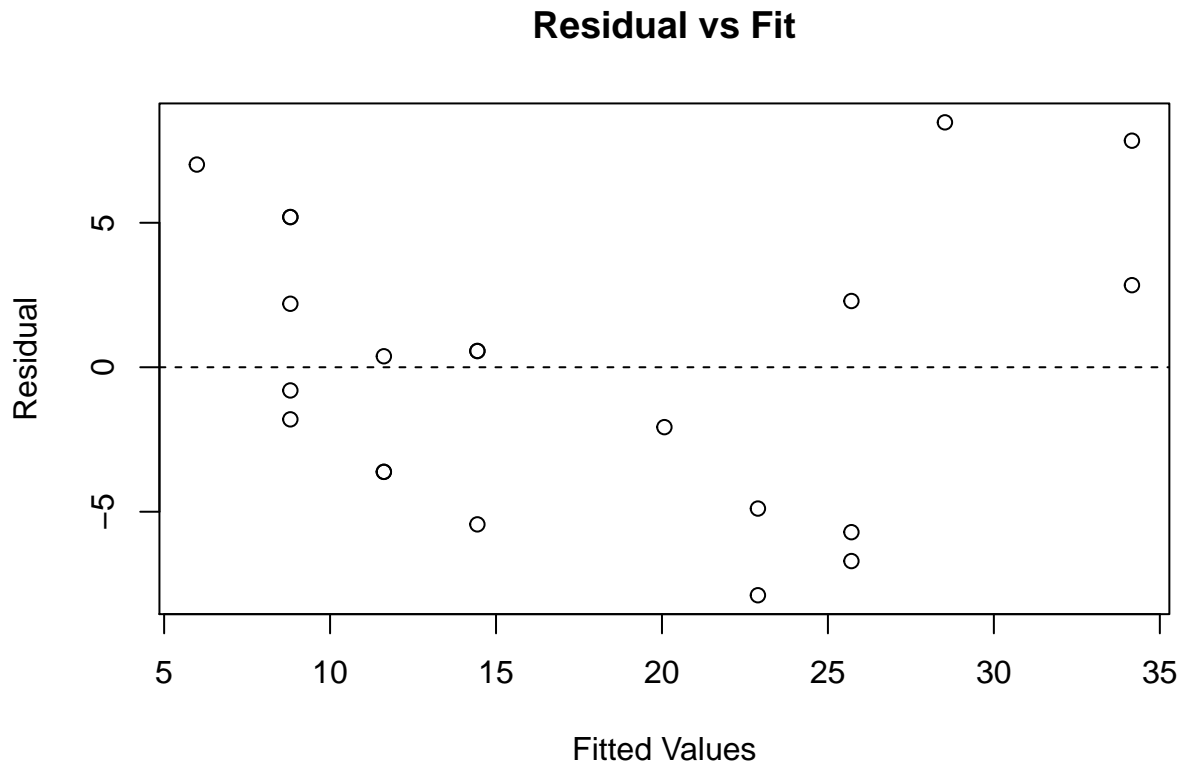
```
y.lm = lm(y~x, data = stackloss)
summary(y.lm)$r.squared
```

```
## [1] 0.766508
```

The coefficient of determination (or R squared) is 0.766508

Part b | Residual vs Fit Plot

```
yhat = fitted(fit2)
e = y - yhat
plot(yhat, e, xlab = "Fitted Values", ylab = "Residual", main = "Residual vs Fit")
abline(h = 0, lty = 2)
```



The “residual vs fit” plot has a non-constant variance and has a non-linearity problem. The main problem is that the plot shows some fanning pattern therefore there is some evidence that the error terms might not be equal.

Part c | Coefficient of Determination

```
WaterTemp2 = (stackloss$Water.Temp)^2
quadratic.model = lm(y ~ x + WaterTemp2)
summary(quadratic.model)
```

```
##
## Call:
## lm(formula = y ~ x + WaterTemp2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.396 -2.850  0.180  3.180  8.617
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  151.5654   49.0572   3.090 0.006322 **
## x           -15.2555    4.5593  -3.346 0.003596 **
## WaterTemp2     0.4131    0.1040   3.971 0.000896 ***
```

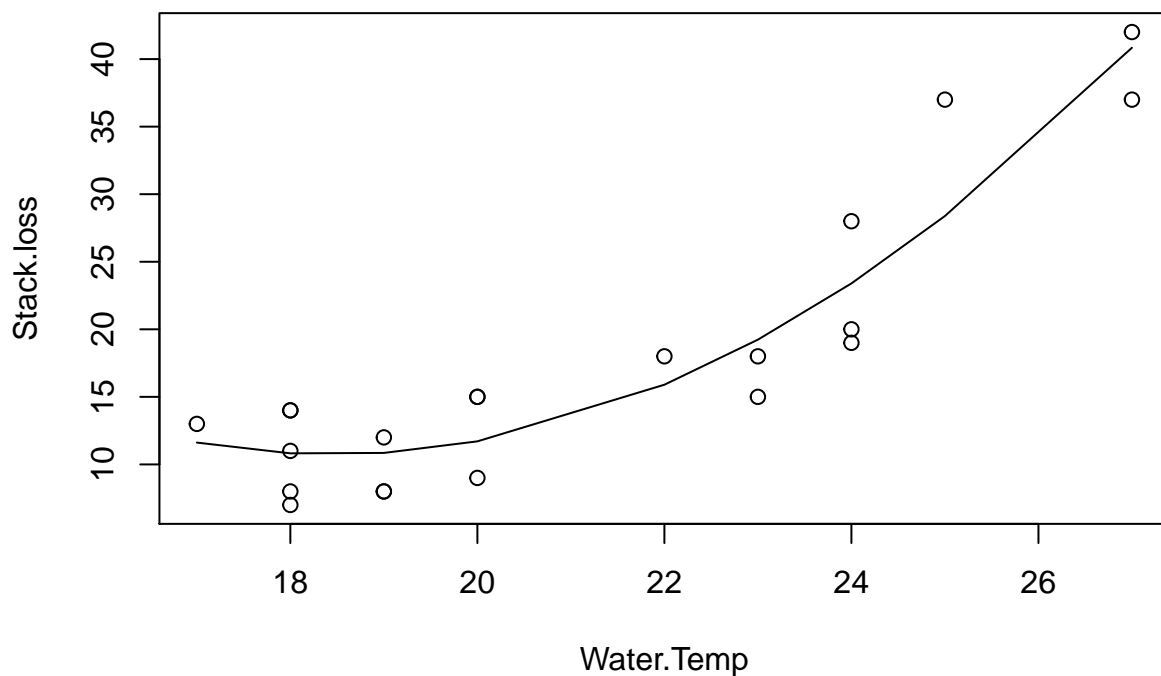


```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.783 on 18 degrees of freedom
## Multiple R-squared:  0.8755, Adjusted R-squared:  0.8617
## F-statistic: 63.31 on 2 and 18 DF,  p-value: 7.168e-09
```

The coefficient of determination (R squared) is 0.8755

Part d | Plot

```
plot(x,y,xlab = "Water.Temp",ylab = "Stack.loss")
lines(sort(x),(fitted(quadratic.model))[order(x)])
```



Part e | Analysis of Variance Table

```
x1 = stackloss$Water.Temp
x2 = (stackloss$Water.Temp)^2
fit3 = lm(x2~x1)
anov = anova(fit3)
anov
```

```
## Analysis of Variance Table
##
## Response: x2
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1           1 382385   382385   5496.5 < 2.2e-16 ***
## Residuals  19   1322         70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The $SSR(x_2|x_1) = 382385$