

Cherry Trees - Part 2

```
# Cherry tree data is already a dataframe in R

#help(trees)
#str(trees)
#head(trees)
#tail(trees)

#x <- trees$Girth (this is mislabeled and is really the tree diameter in inches)
#y <- trees$Volume (timber amount in cubic feet)

Diameter <- trees$Girth
trees1 <- cbind(trees, Diameter)

model <- lm(Volume ~ Diameter, data = trees1)
summary(model)
```

```
##
## Call:
## lm(formula = Volume ~ Diameter, data = trees1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.065 -3.107  0.152   3.495   9.587
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -36.9435     3.3651  -10.98 7.62e-12 ***
## Diameter      5.0659     0.2474   20.48 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.252 on 29 degrees of freedom
## Multiple R-squared:  0.9353, Adjusted R-squared:  0.9331
## F-statistic: 419.4 on 1 and 29 DF,  p-value: < 2.2e-16
```

```
#str(model)
```

```
anova(model)
```

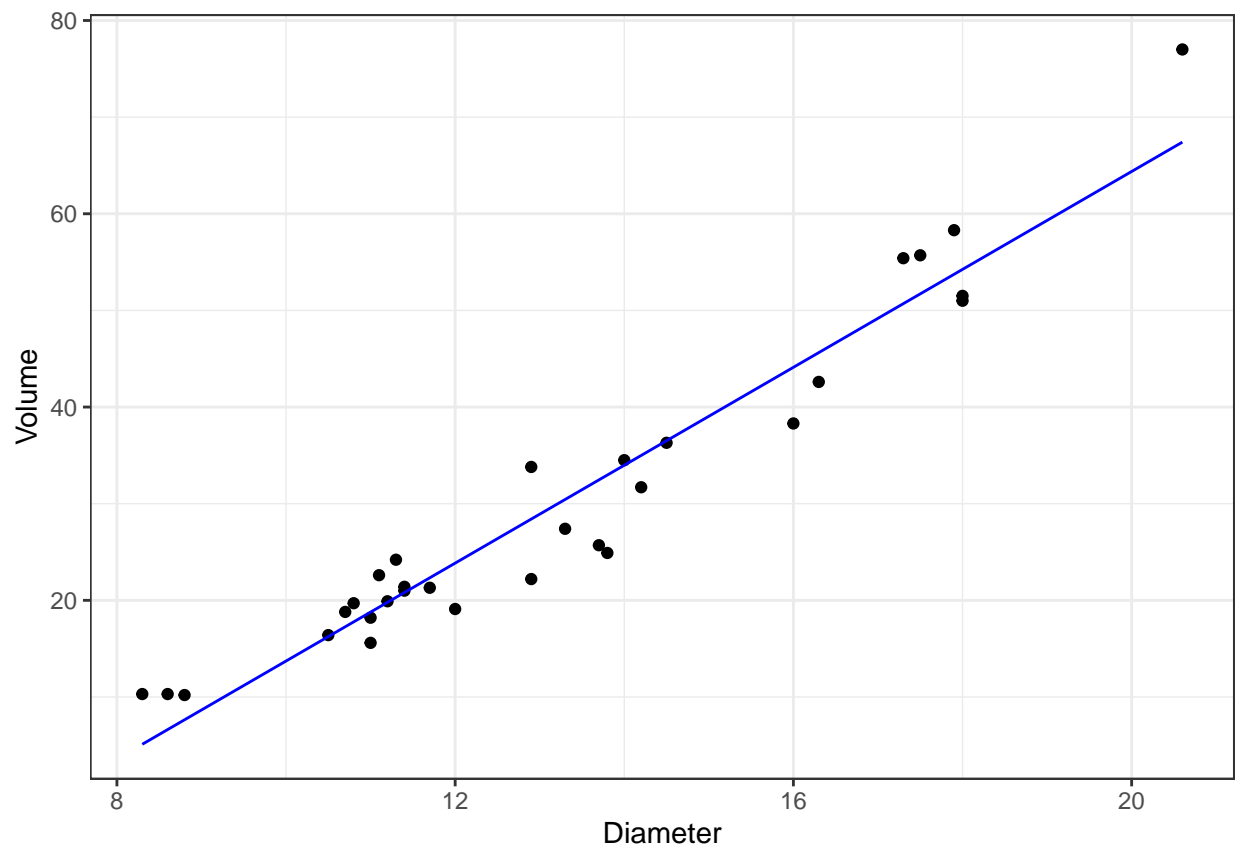
```
## Analysis of Variance Table
##
## Response: Volume
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Diameter   1 7581.8  7581.8   419.36 < 2.2e-16 ***
## Residuals 29  524.3    18.1
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(model)
```

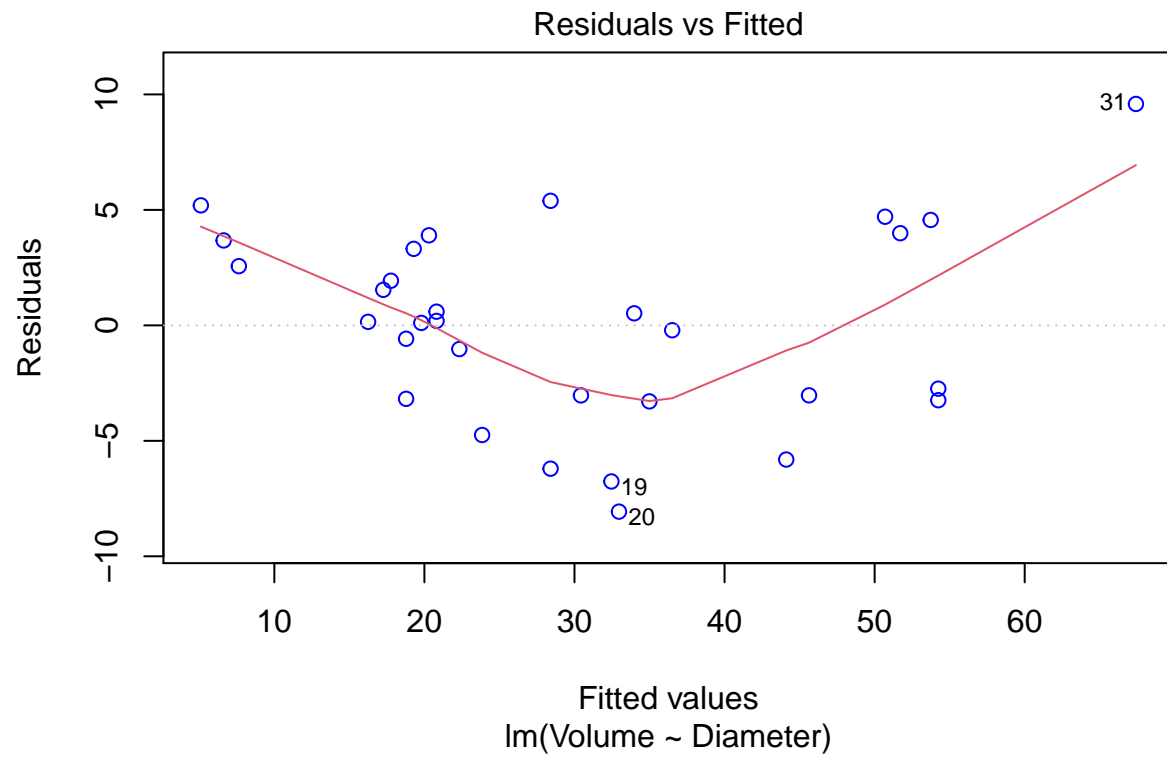
```
##           2.5 %    97.5 %
## (Intercept) -43.825953 -30.060965
## Diameter      4.559914   5.571799
```

```
library(ggplot2)
figure <- ggplot(trees1) + geom_point( aes(x=Diameter, y=Volume) ) + geom_line(aes(x=Diameter, y=model))
figure
```

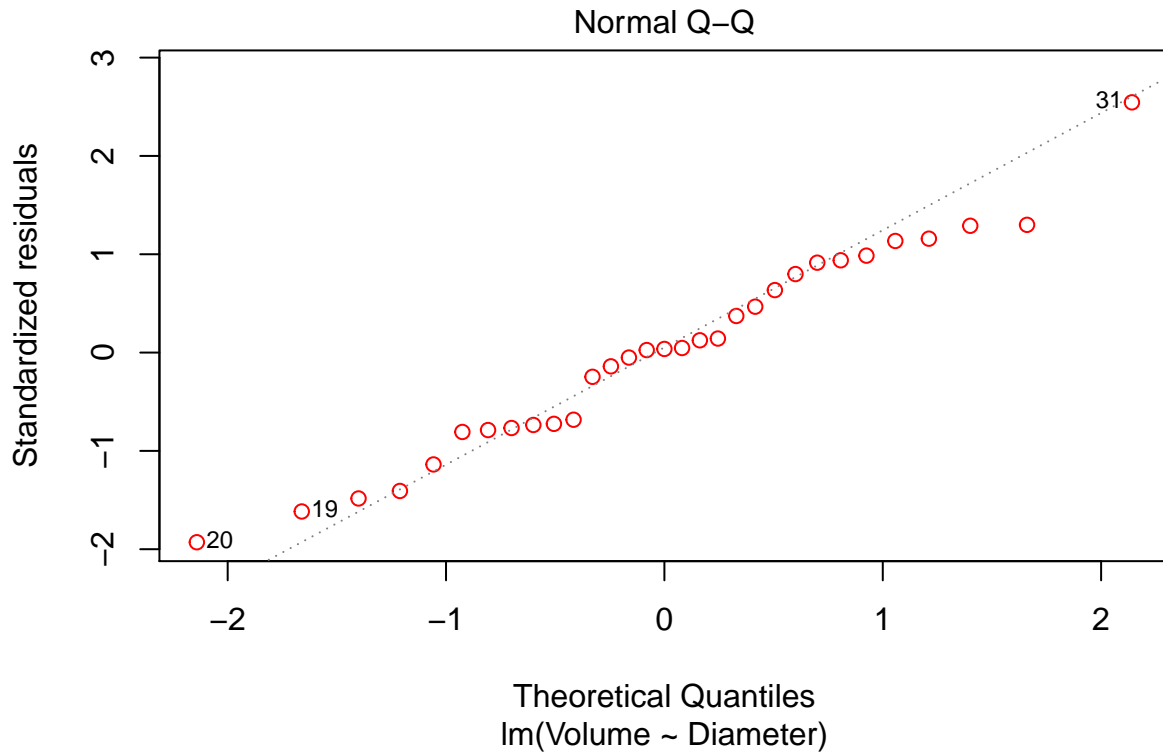


```
# residual diagnostics
```

```
plot(model, which=1, col=c("blue")) # Residuals vs Fitted Plot
```



```
plot(model, which=2, col=c("red")) # Q-Q Plot
```



```
shapiro.test(residuals(model))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(model)
## W = 0.97889, p-value = 0.7811
```

```
# Quadratic polynomial
```

```
model1 <- lm(Volume ~ Diameter + I(Diameter^2), data = trees1)
summary(model1)
```

```
##
## Call:
## lm(formula = Volume ~ Diameter + I(Diameter^2), data = trees1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.4889 -2.4293 -0.3718  2.0764  7.6447
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   10.78627    11.22282   0.961 0.344728
## Diameter      -2.09214     1.64734  -1.270 0.214534
```

```
## I(Diameter^2) 0.25454 0.05817 4.376 0.000152 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.335 on 28 degrees of freedom
## Multiple R-squared: 0.9616, Adjusted R-squared: 0.9588
## F-statistic: 350.5 on 2 and 28 DF, p-value: < 2.2e-16
```

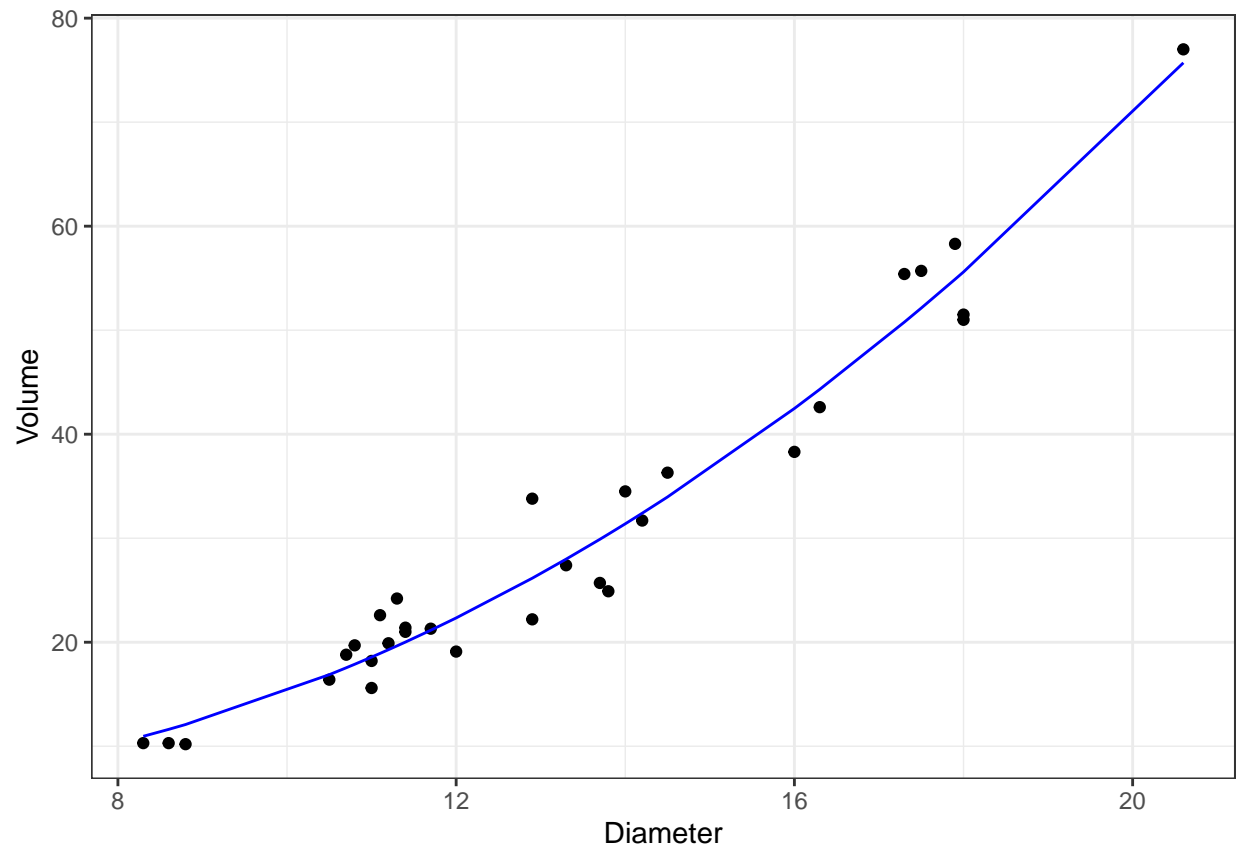
```
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: Volume
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Diameter    1 7581.8  7581.8 681.766 < 2.2e-16 ***
## I(Diameter^2) 1  212.9   212.9  19.146 0.0001524 ***
## Residuals   28  311.4    11.1
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(model1)
```

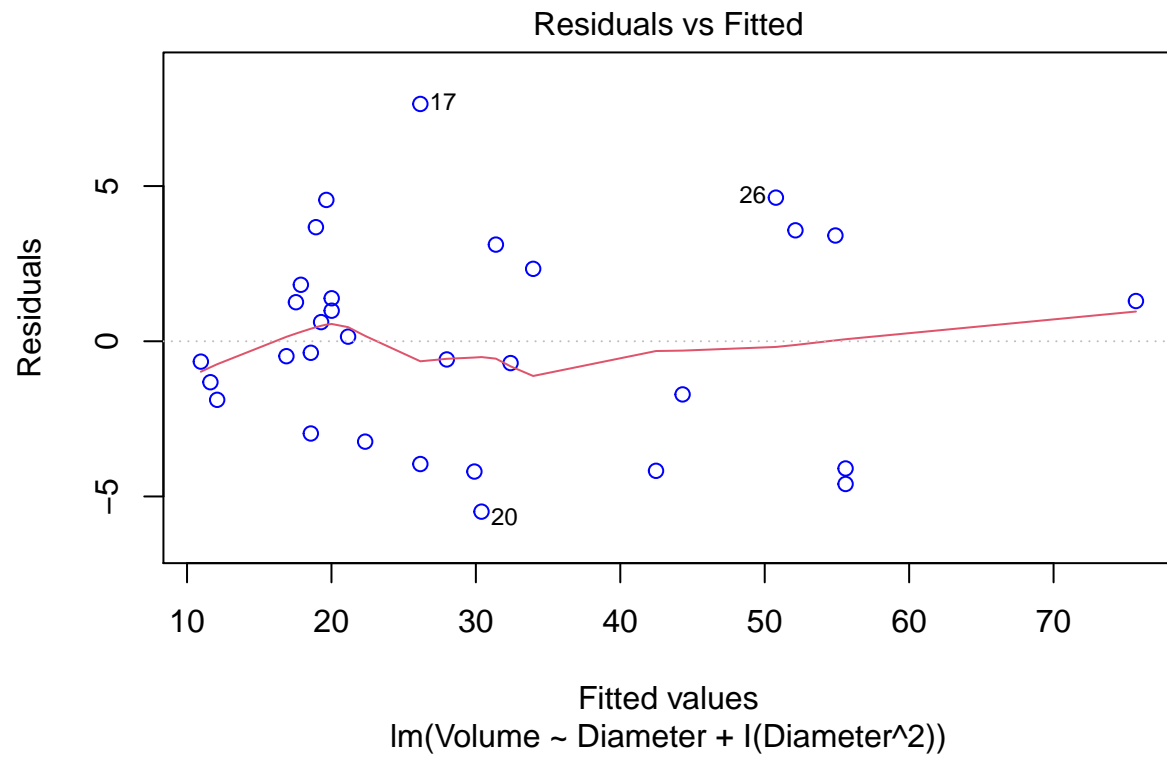
```
##          2.5 %    97.5 %
## (Intercept) -12.2026390 33.7751699
## Diameter    -5.4665661  1.2822870
## I(Diameter^2) 0.1353784  0.3736967
```

```
library(ggplot2)
figure1 <- ggplot(trees1) + geom_point( aes(x=Diameter, y=Volume) ) + geom_line(aes(x=Diameter, y=mod
figure1
```

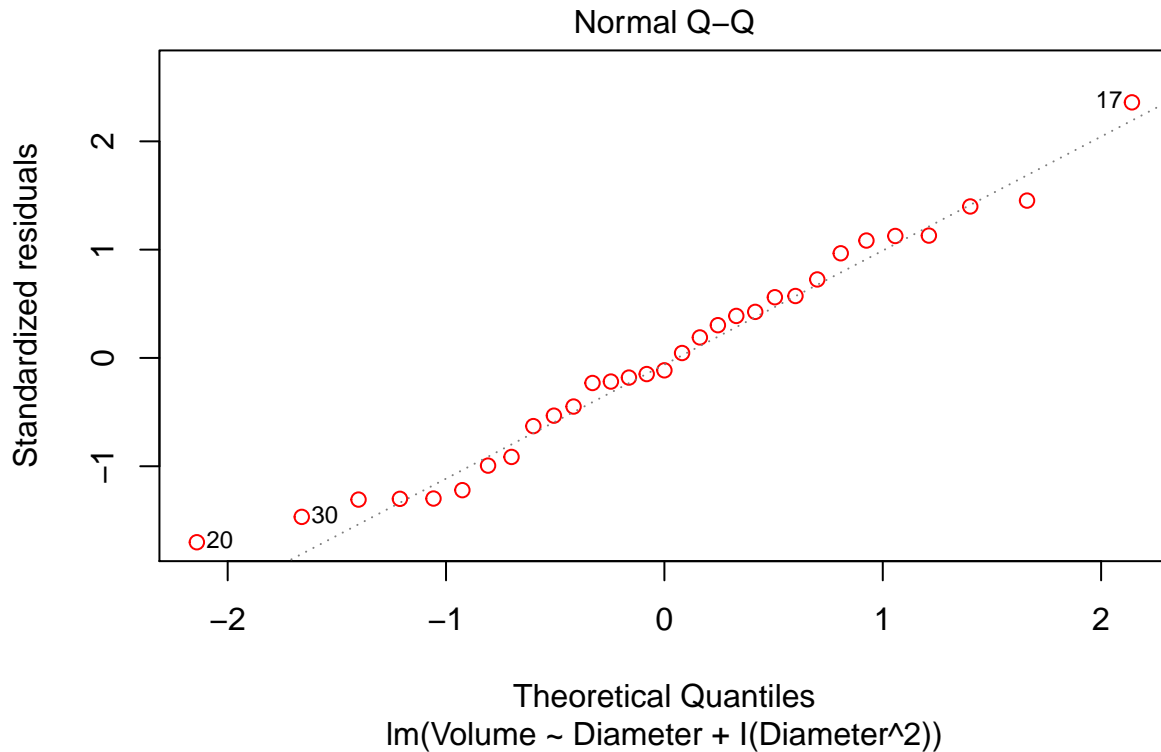


```
# residual diagnostics
```

```
plot(model1, which=1, col=c("blue")) # Residuals vs Fitted Plot
```



```
plot(model1, which=2, col=c("red")) # Q-Q Plot
```



```
shapiro.test(residuals(model1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(model1)
## W = 0.97393, p-value = 0.6327
```

```
# Quadratic polynomial using poly function
```

```
model2 <- lm(Volume ~ poly(Diameter, 2, raw=TRUE), data = trees1)
summary(model2)
```

```
##
## Call:
## lm(formula = Volume ~ poly(Diameter, 2, raw = TRUE), data = trees1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.4889 -2.4293 -0.3718  2.0764  7.6447
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    10.78627   11.22282   0.961 0.344728
## poly(Diameter, 2, raw = TRUE)1 -2.09214    1.64734  -1.270 0.214534
```



```
## poly(Diameter, 2, raw = TRUE)2  0.25454    0.05817    4.376 0.000152 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.335 on 28 degrees of freedom
## Multiple R-squared:  0.9616, Adjusted R-squared:  0.9588
## F-statistic: 350.5 on 2 and 28 DF,  p-value: < 2.2e-16
```

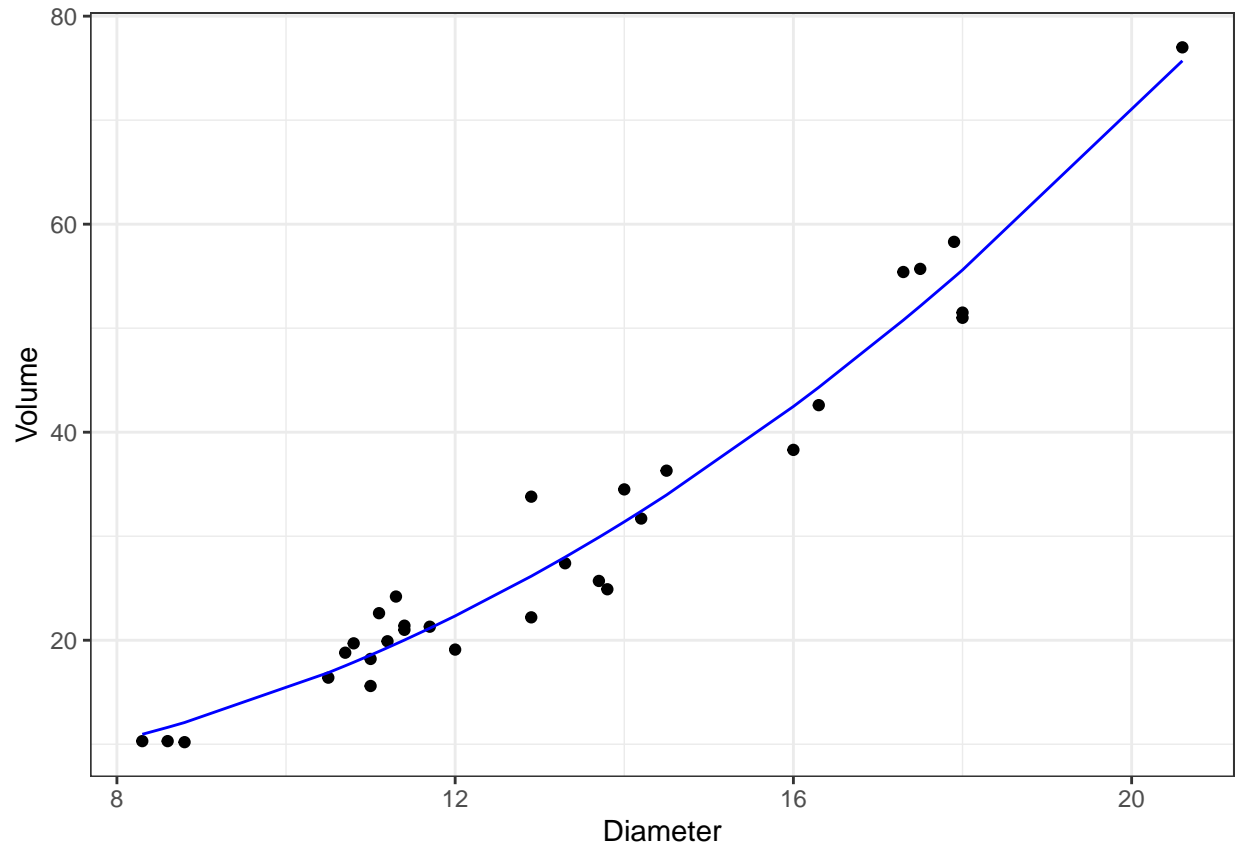
```
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: Volume
##
##              Df Sum Sq Mean Sq F value    Pr(>F)
## poly(Diameter, 2, raw = TRUE)  2 7794.7   3897.4   350.46 < 2.2e-16 ***
## Residuals                    28   311.4     11.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(model2)
```

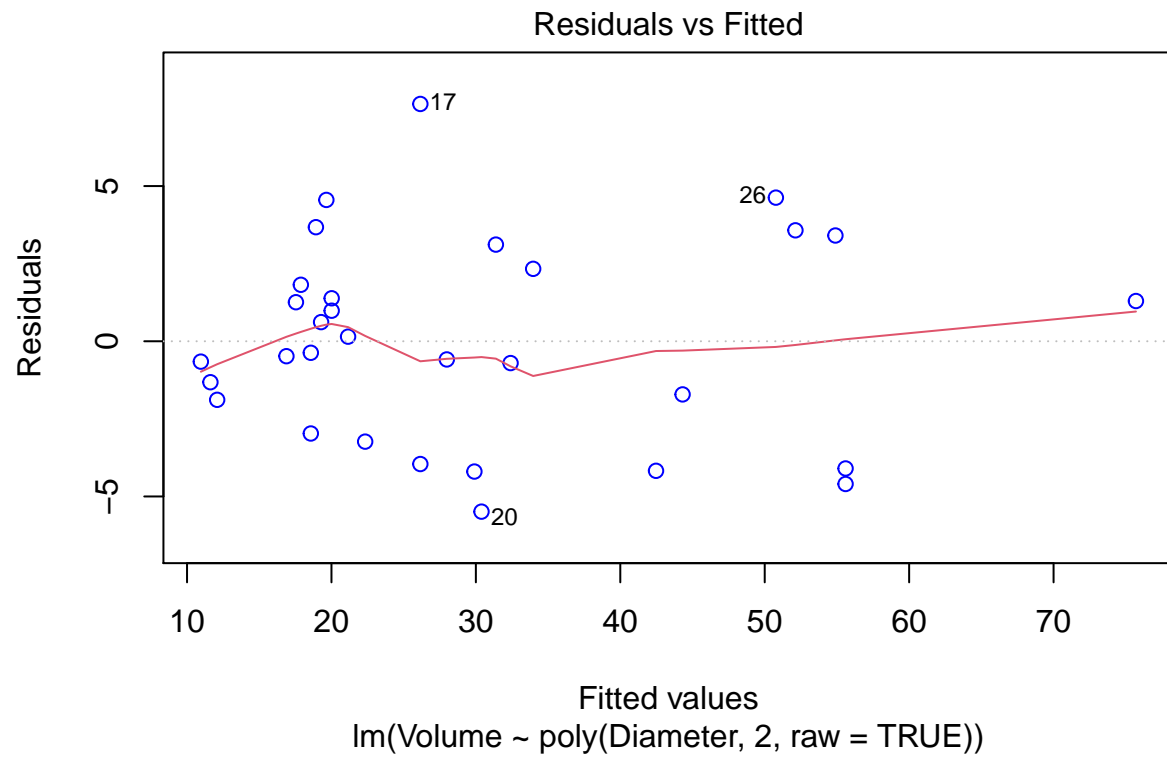
```
##              2.5 %    97.5 %
## (Intercept) -12.2026390 33.7751699
## poly(Diameter, 2, raw = TRUE)1 -5.4665661 1.2822870
## poly(Diameter, 2, raw = TRUE)2  0.1353784 0.3736967
```

```
library(ggplot2)
figure2 <- ggplot(trees1) + geom_point( aes(x=Diameter, y=Volume) ) + geom_line(aes(x=Diameter, y=mod
figure2
```

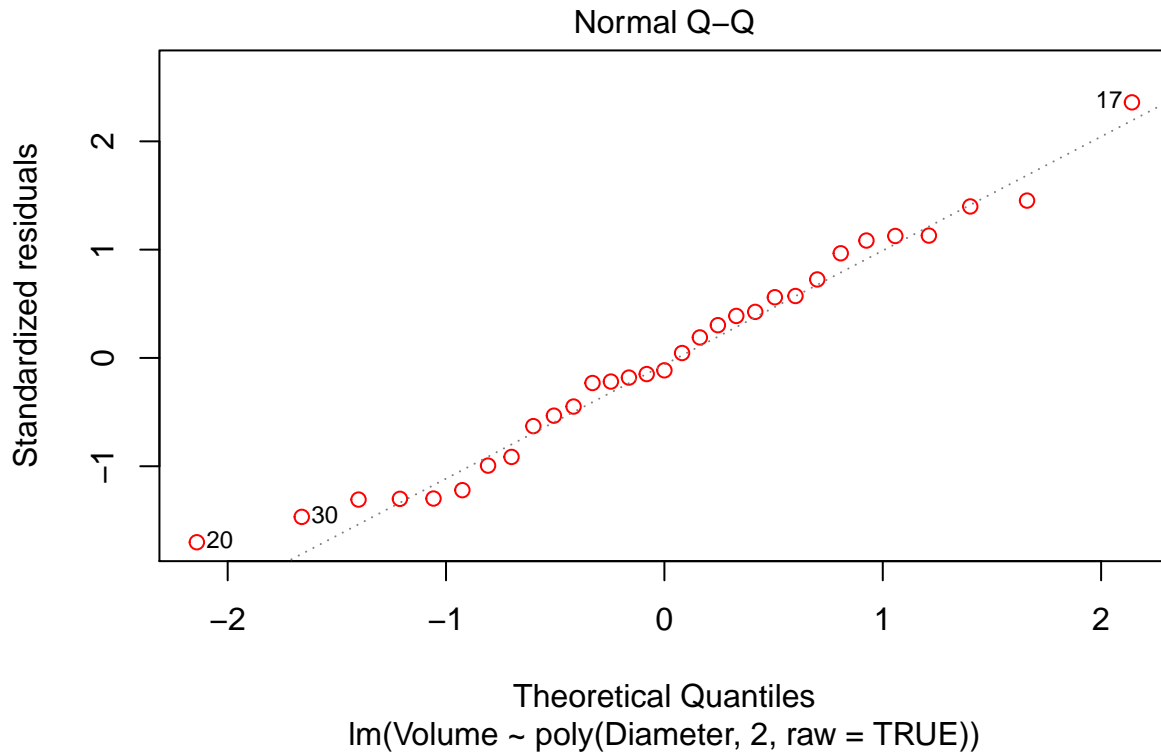


```
# residual diagnostics
```

```
plot(model2, which=1, col=c("blue")) # Residuals vs Fitted Plot
```



```
plot(model2, which=2, col=c("red")) # Q-Q Plot
```



```
shapiro.test(residuals(model2))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(model2)
## W = 0.97393, p-value = 0.6327
```

```
# Orthogonal polynomials using poly function without "raw = TRUE"
```

```
model3 <- lm(Volume ~ poly(Diameter, 2), data = trees1)
summary(model3)
```

```
##
## Call:
## lm(formula = Volume ~ poly(Diameter, 2), data = trees1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.4889 -2.4293 -0.3718  2.0764  7.6447
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    30.1710     0.5989  50.374 < 2e-16 ***
## poly(Diameter, 2)1  87.0734     3.3348  26.111 < 2e-16 ***
```

```
## poly(Diameter, 2)2 14.5918      3.3348    4.376 0.000152 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.335 on 28 degrees of freedom
## Multiple R-squared:  0.9616, Adjusted R-squared:  0.9588
## F-statistic: 350.5 on 2 and 28 DF,  p-value: < 2.2e-16
```

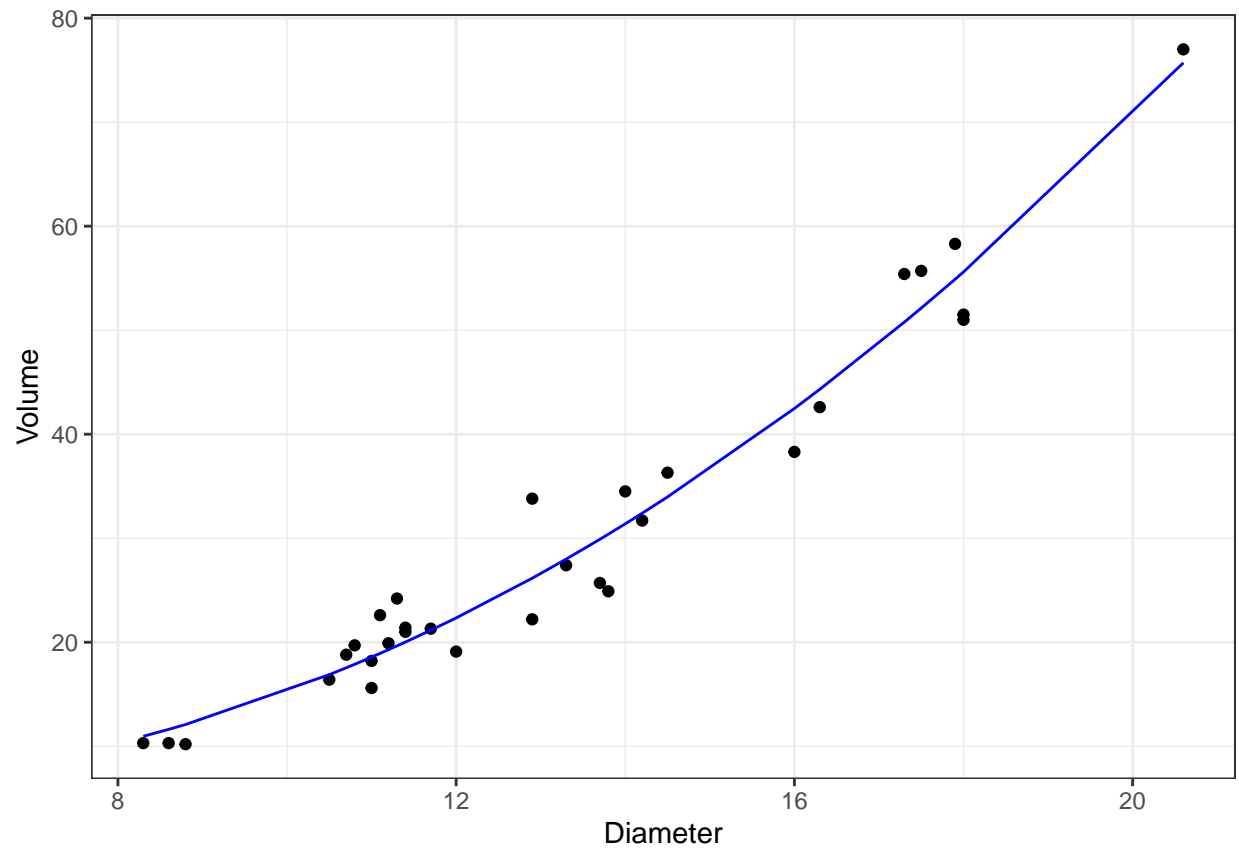
```
anova(model3)
```

```
## Analysis of Variance Table
##
## Response: Volume
##              Df Sum Sq Mean Sq F value    Pr(>F)
## poly(Diameter, 2) 2 7794.7  3897.4  350.46 < 2.2e-16 ***
## Residuals        28  311.4    11.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(model3)
```

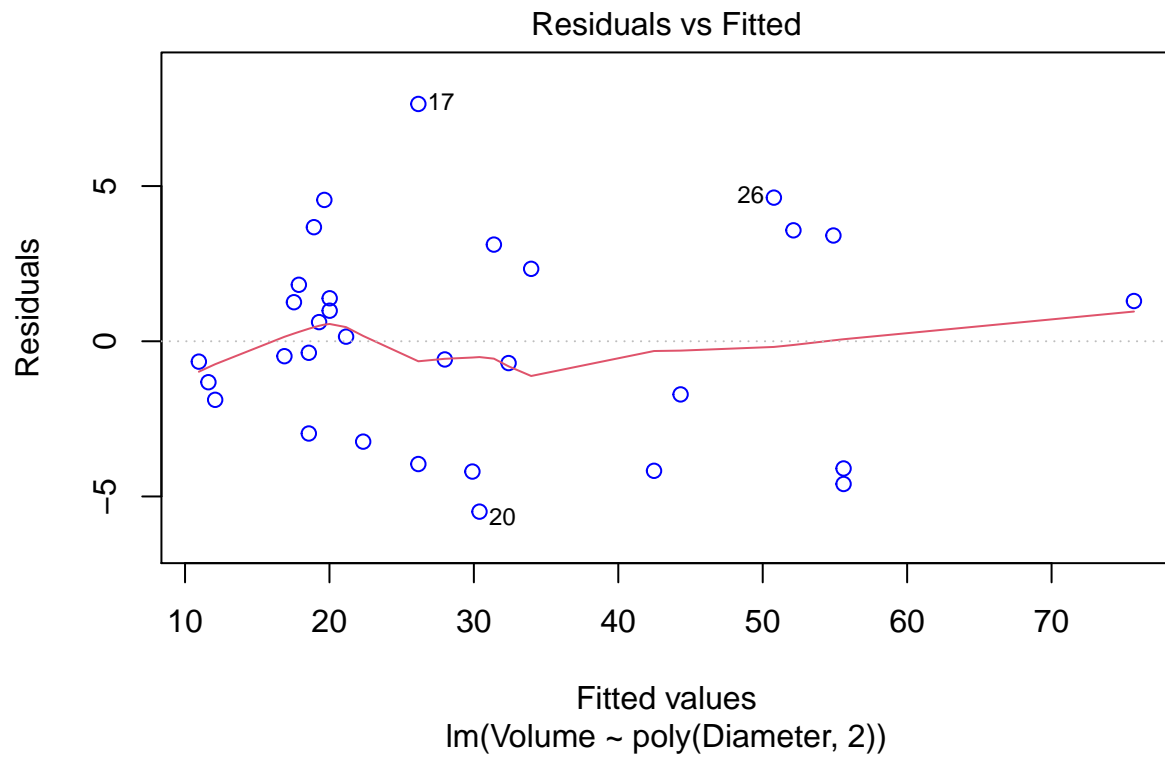
```
##              2.5 %    97.5 %
## (Intercept) 28.944084 31.39785
## poly(Diameter, 2)1 80.242428 93.90442
## poly(Diameter, 2)2  7.760796 21.42279
```

```
library(ggplot2)
figure3 <- ggplot(trees1) + geom_point( aes(x=Diameter, y=Volume) ) + geom_line(aes(x=Diameter, y=mod
figure3
```

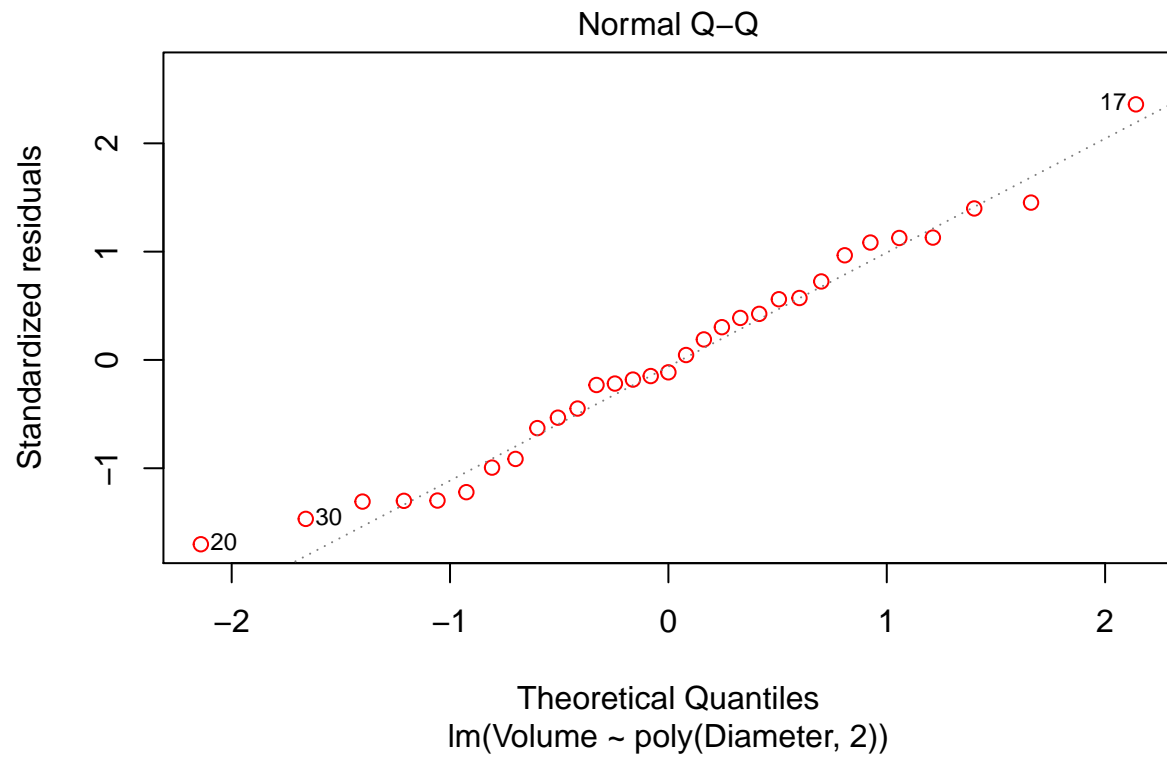


```
# residual diagnostics
```

```
plot(model3, which=1, col=c("blue")) # Residuals vs Fitted Plot
```



```
plot(model3, which=2, col=c("red")) # Q-Q Plot
```



```
shapiro.test(residuals(model3))
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: residuals(model3)  
## W = 0.97393, p-value = 0.6327
```