Homework 3

Richard Albright
ISYE6414
Spring 2020

Background

Predicting the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope – a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age. Further information, such as weather patterns and location (hence food availability) may be required to solve the problem.

From the original data examples with missing values were removed (the majority having the predicted value missing), and the ranges of the continuous values have been scaled for use with an ANN (by dividing by 200).

Data Description

The data consists of the following variables:

1. *Sex*: M, F, and I (infant) (categorical)
2. *Length*: Longest shell measurement in mm (continuous)
3. *Diameter*: Perpendicular to length in mm (continuous)
4. *Height*: Height with meat in shell in mm (continuous)
5. *Whole*: Weight of whole abalone in grams (continuous)
6. *Viscera*: Gut weight (after bleeding) in grams (continuous)
7. *Shell*: Shell weight after being dried in grams (continuous)
8. *Rings*: Number of rings of the abalone – corresponds with the age (continuous)

Read the data

```
# Import library you may need
library(car)
# Read the data set
abaloneFull = read.csv("~/Dropbox/GaTech/ISYE6414/abalone.csv",head=T)
row.cnt = nrow(abaloneFull)
# Split the data into training and testing sets
abaloneTest = abaloneFull[(row.cnt-9):row.cnt,]
abalone = abaloneFull[1:(row.cnt-10),]
```

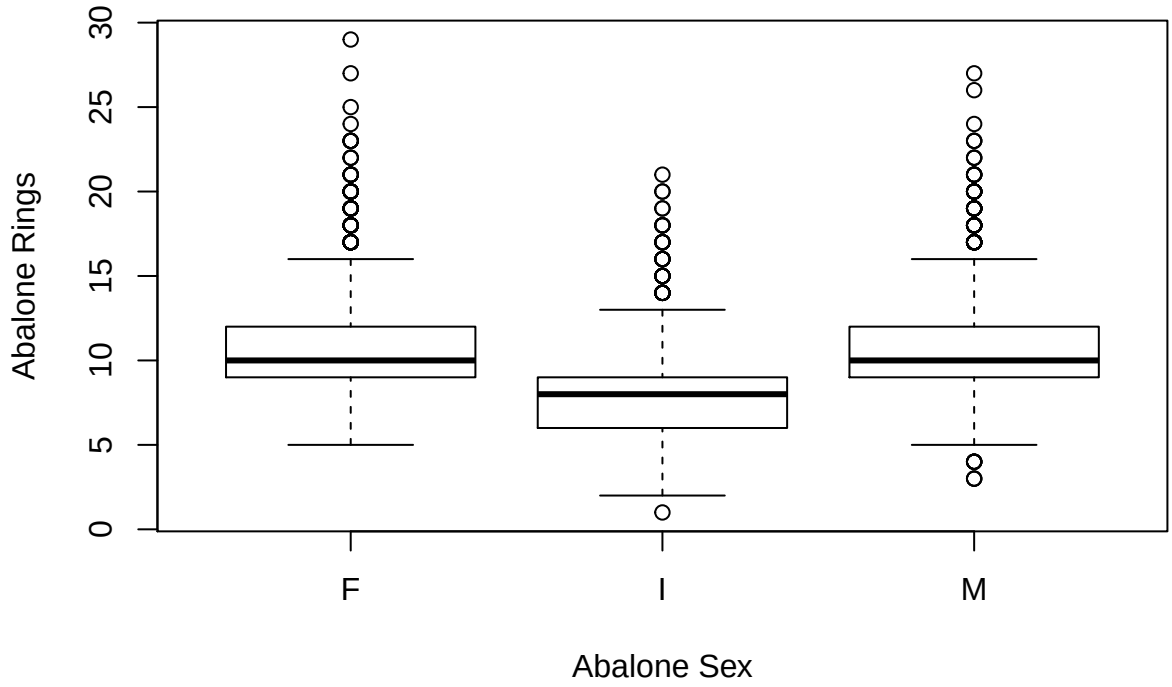
Please use abalone as your data set for the following questions unless otherwise stated.

Question 1: Exploratory Data Analysis [16 points]

Please use your best judgement when grading this question. Credit should be given to submissions that use evidence from the graphs to support the conclusion, even if it does not exactly match this solution.

(a) Create a box plot comparing the response variable, *Rings*, across the three sex categories. Based on this box plot, does there appear to be a relationship between the predictor and the response?

```
sex = as.factor(abalone$Sex)
boxplot(abalone$Rings~sex, xlab='Abalone Sex', ylab='Abalone Rings')
```



There appears to be a relationship between the predictor and response variable in that the Infant category may predict less rings. Male vs Female looks like they may not be significantly different from each other.

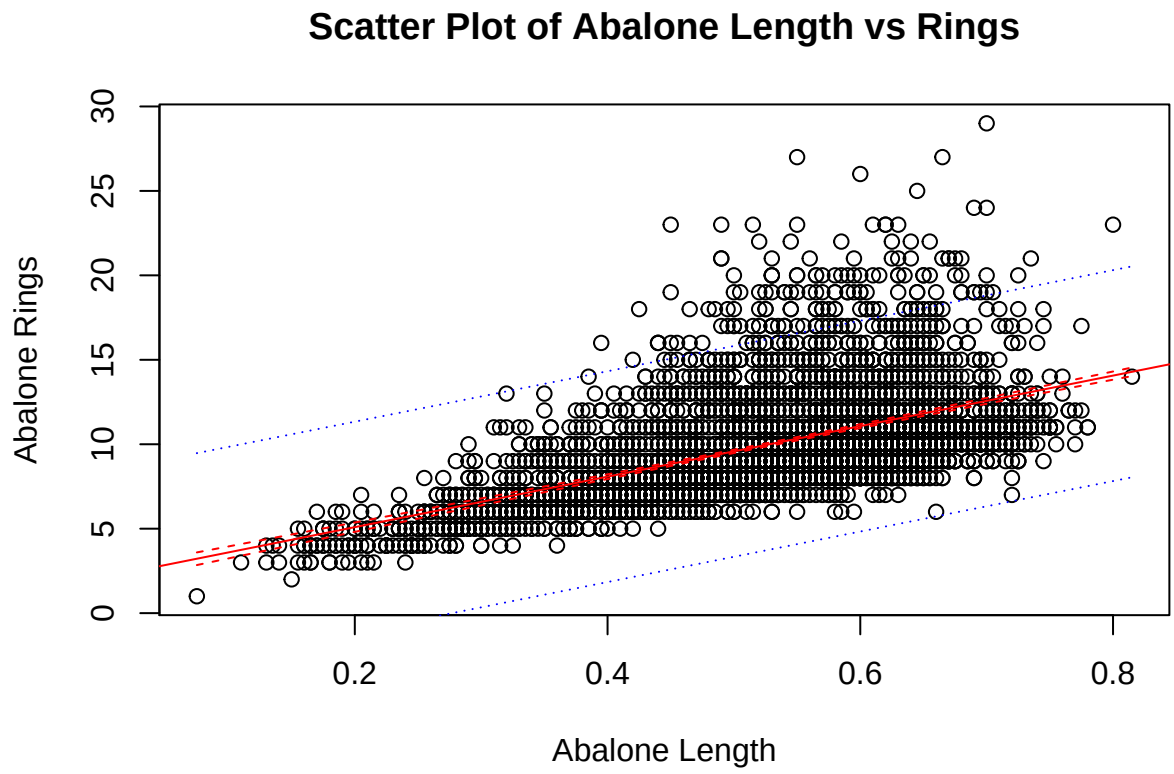
The function below is a modified function taken from Module 1 and used to answer the question below the function definition. I used a confidence level of 99% on the confidence band scatter plots.

```
plot.confbands <- function(x,y,conf=.95,CImean=T,PI=T,CIregline=F,legend=F, xlab='X Values', ylab='Y Va
    #### Modified from a function written by Sandra McBride, Duke University
    #### For a simple linear regression line, this function
    #### will plot the line, CI for mean response, prediction intervals,
    #### and (optionally) a simulataneous CI for the regression line.
    if (!is.vector(x)) {
      x <-as.vector(x)
    }
    if (!is.vector(y)) {
      y <-as.vector(y)
    }
    xx <- x[order(x)]
    yy <- y[order(x)]
    lm1 <- lm(yy~xx)
```

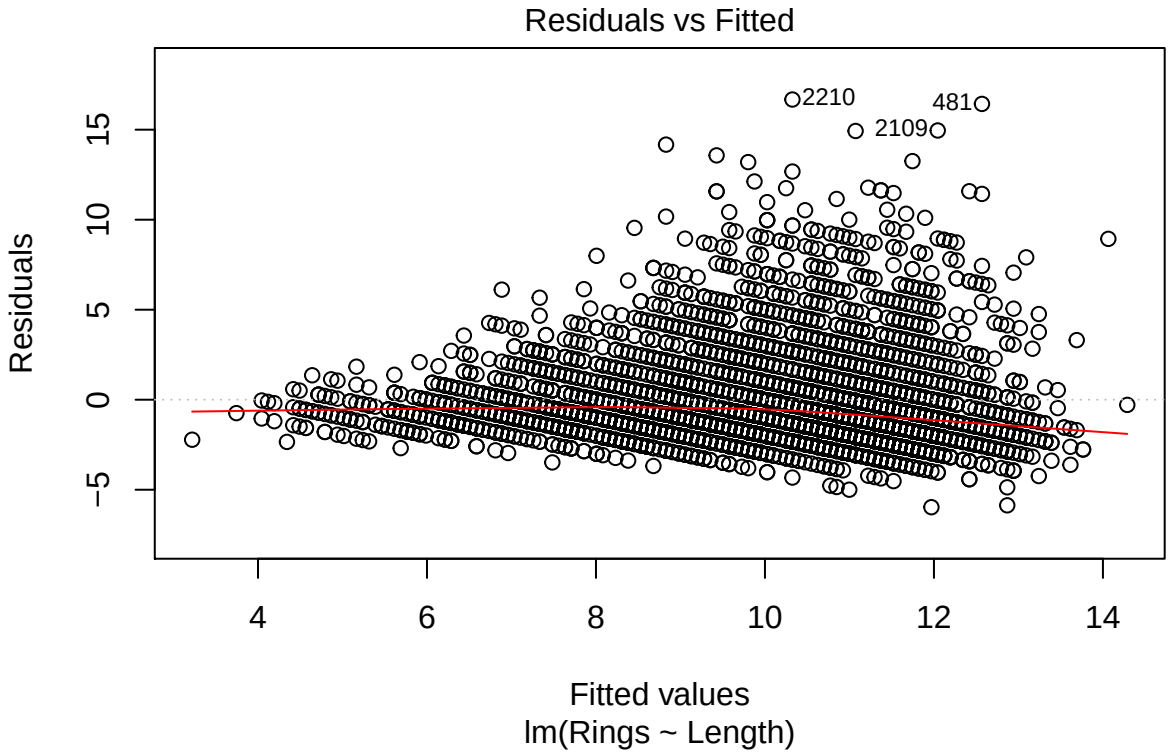
```
    plot(xx,yy,ylim=c(min(yy),max(yy)),main=main,ylab=ylab,
         xlab=xlab)
    abline(lm1$coefficients, col='red')
    #### calculation of components of intervals ####
    n <- length(yy)
    sx2 <- (var(xx))
    shat <- summary(lm1)$sigma
    s2hat <- shat^2
    SEmuhat <- shat*sqrt(1/n+ ((xx-mean(xx))^2)/((n-1)*sx2))
    SEpred <- sqrt(s2hat+SEmuhat^2)
    t.quantile <- qt(conf,lm1$df.residual)
    ####
    if (CImean==T){
        mean.up <- lm1$fitted+t.quantile*SEmuhat
        mean.down <- lm1$fitted-t.quantile*SEmuhat
        lines(xx,mean.up,lty=2, col='red')
        lines(xx,mean.down,lty=2, col='red')
    }
    if (PI==T){
        PI.up <- lm1$fitted+t.quantile*SEpred
        PI.down <- lm1$fitted-t.quantile*SEpred
        lines(xx,PI.up,lty=3, col='blue')
        lines(xx,PI.down,lty=3, col='blue')
    }
    if (CIregline==T){
        HW <- sqrt(2*qf(conf,n-lm1$df.residual,lm1$df.residual))*SEmuhat
        CIreg.up <- lm1$fitted+HW
        CIreg.down <- lm1$fitted-HW
        lines(xx,CIreg.up,lty=4, col='green')
        lines(xx,CIreg.down,lty=4, col='green')
    }
      if (legend==T){
        choices <- c(CImean,PI,CIregline)
        line.type <- c(2,3,4)
             names.line <- c("CI for mean resp.","Prediction Int.","CI for reg. line")
        legend(max(xx)-.2*max(xx),max(yy)+.2*max(yy),legend=names.line[choices],lty=line.type[choices])
    }
}
```

(b) Create plots of the response, *Rings*, against each quantitative predictor, namely *Length*, *Diameter*, *Height*, *Whole*, *Viscera*, and *Shell*. Describe the general trend of each plot. Are there any potential outliers?

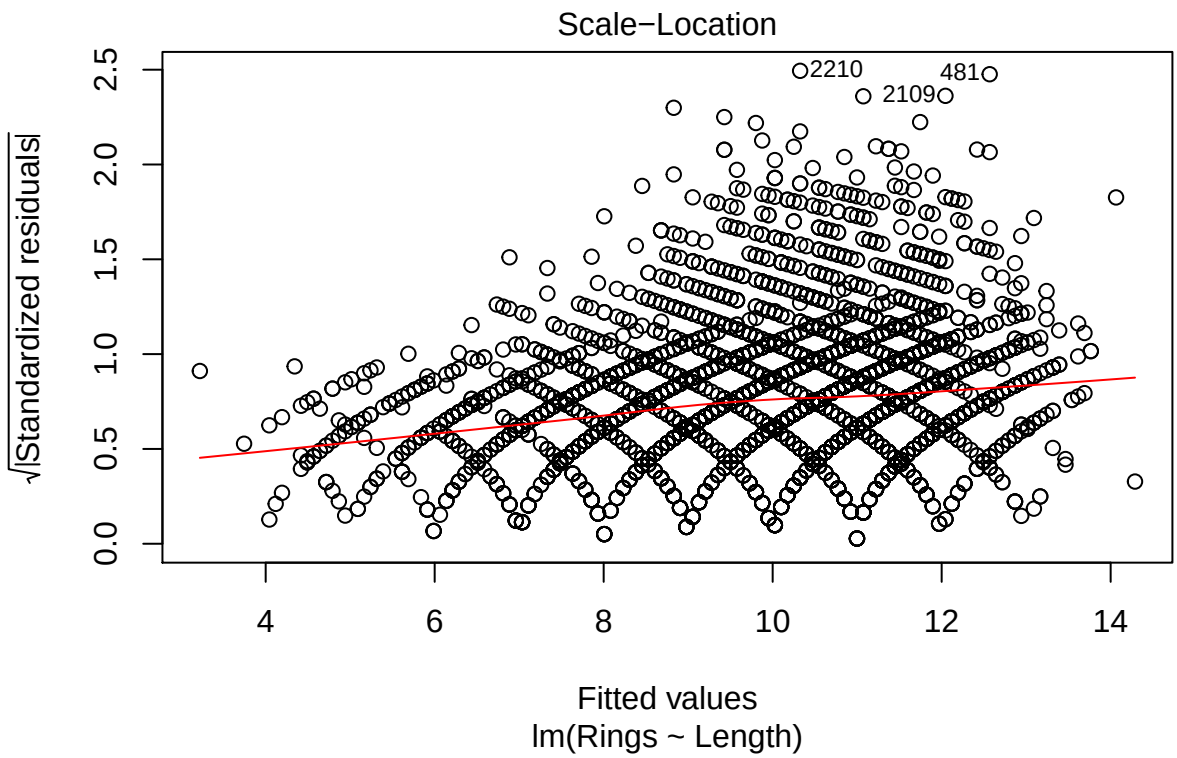
```
plot.confbands(
  abalone$Length,
  abalone$Rings,
  conf=0.99,
  xlab='Abalone Length',
  ylab='Abalone Rings',
  main='Scatter Plot of Abalone Length vs Rings')
```



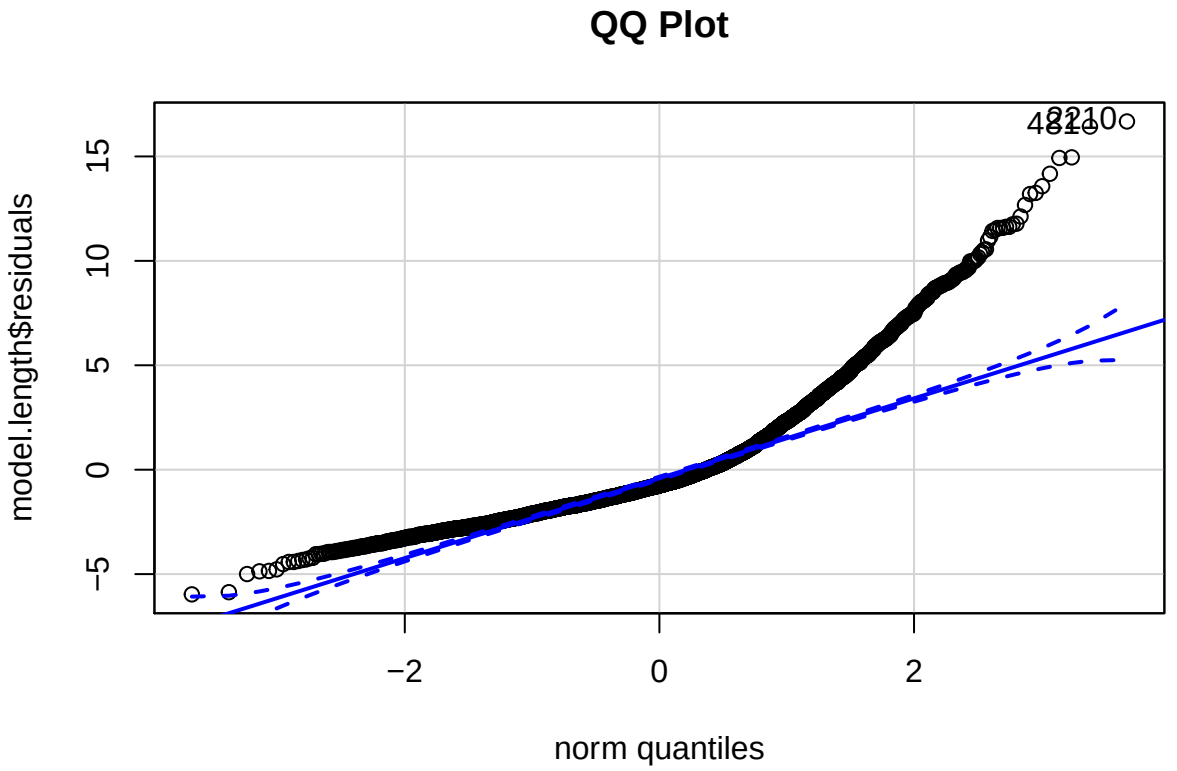

```
model.length = lm(Rings ~ Length, data=abalone)
plot(model.length, 1)
```



```r
plot(model.length, 3)
```




```r
car::qqPlot(model.length$residuals, main='QQ Plot', pch=1)
```

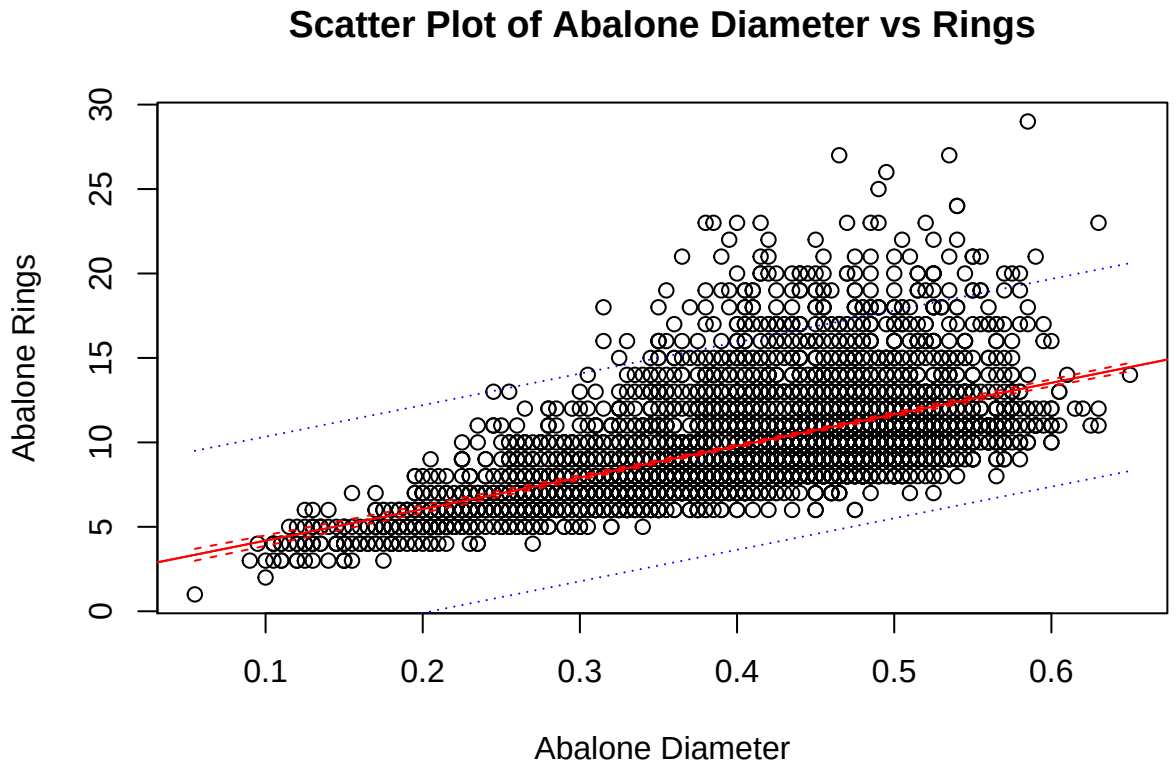



```
2210 481
```

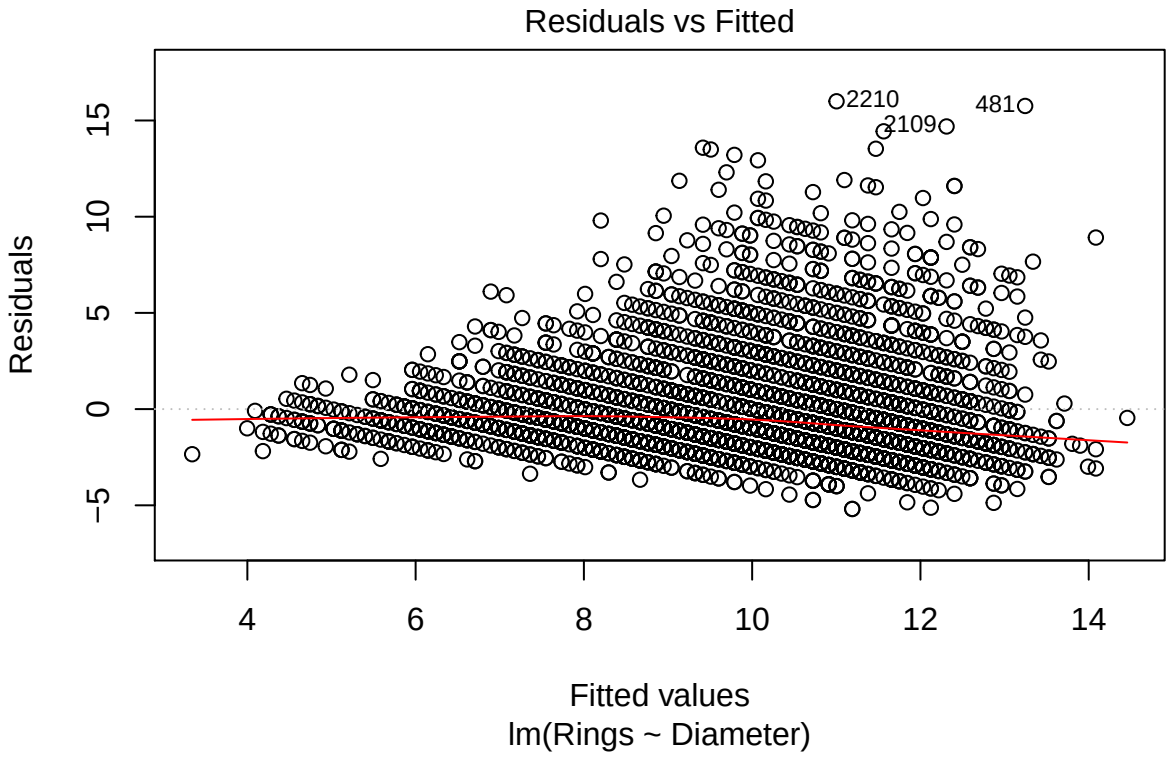
[1]

There is a moderate positive relationship between Rings and Length. The left side of the plots exhibit more variation in the Rings response variable as the Length increases. There are many potential outliers in the scatter plot above the prediction interval (dotted blue line). Rows 481, 2109, and 2210 are specifically identified as potential outliers from the fitted values vs residuals plot. The assumption of constant variance appears to hold. The QQ Plot of the residuals indicates the relationship may by curvilinear and not normally distributed.

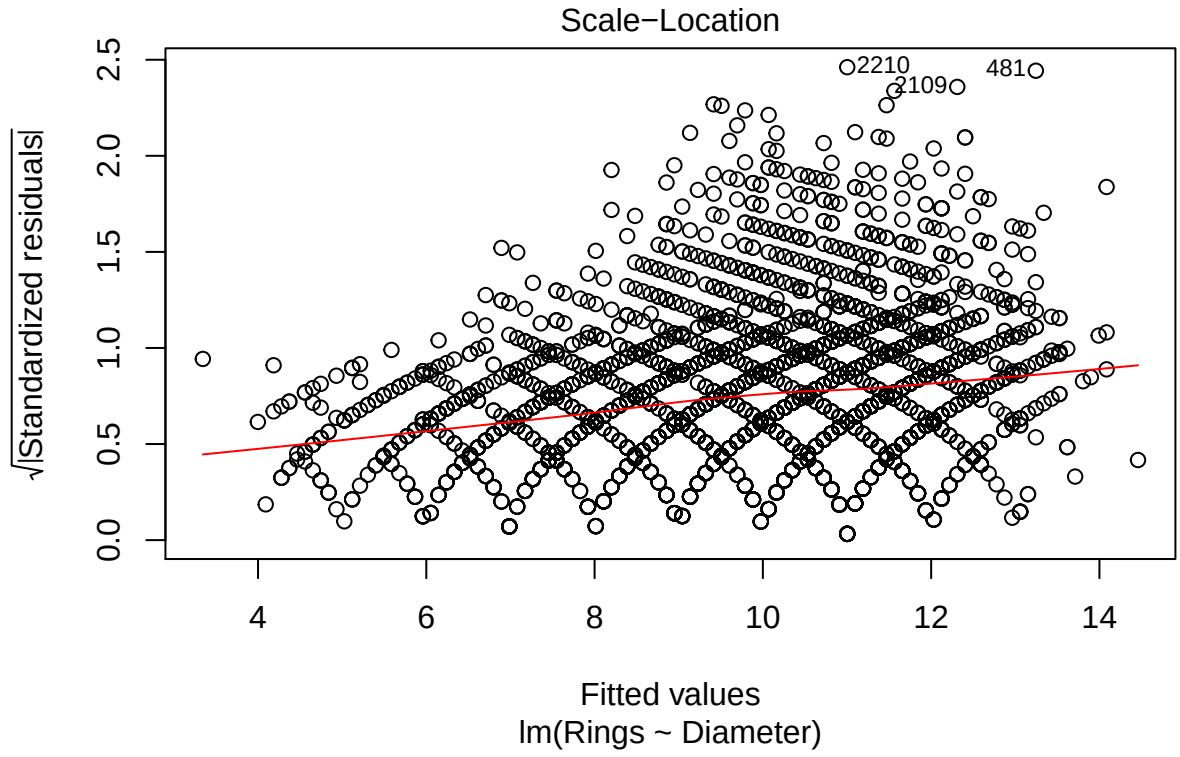
```
plot.confbands(
  abalone$Diameter,
  abalone$Rings,
  conf=0.99,
  xlab='Abalone Diameter',
  ylab='Abalone Rings',
  main='Scatter Plot of Abalone Diameter vs Rings')
model.diameter = lm(Rings ~ Diameter, data=abalone)
abline(model.diameter, col='red')
```



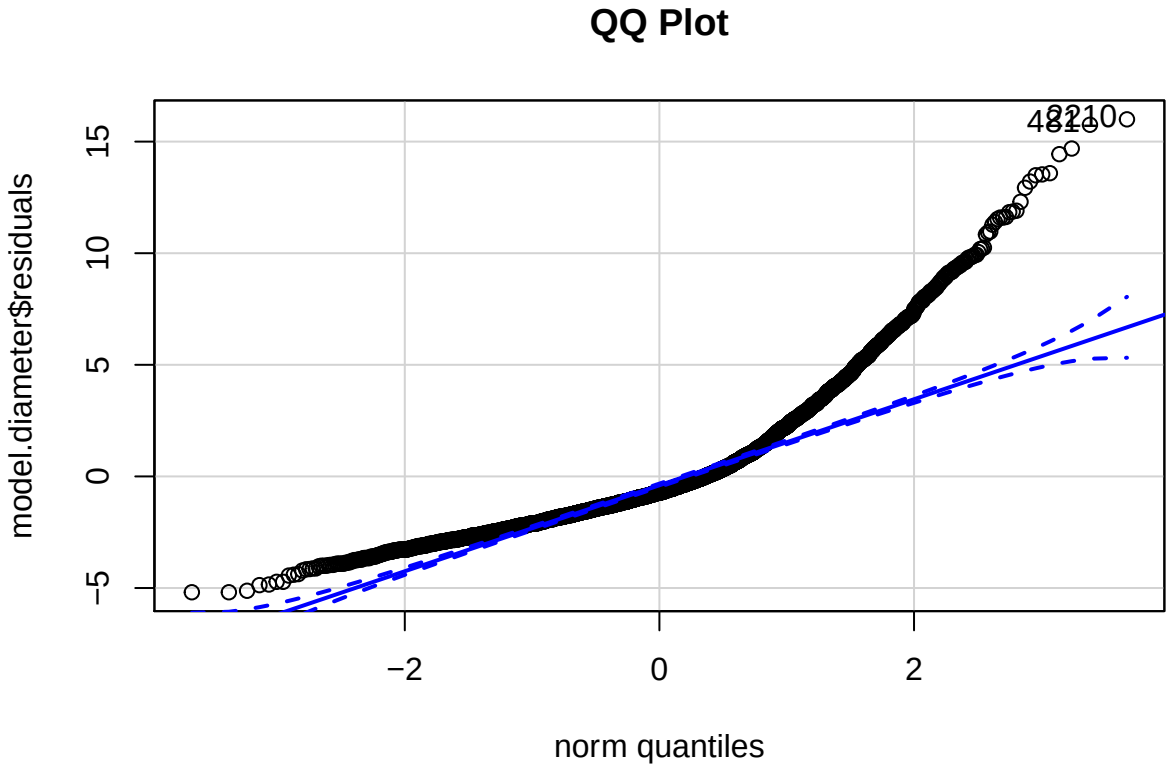

```
plot(model.diameter, 1)
```




```
plot(model.diameter, 3)
```




```
car::qqPlot(model.diameter$residuals, main='QQ Plot', pch=1)
```

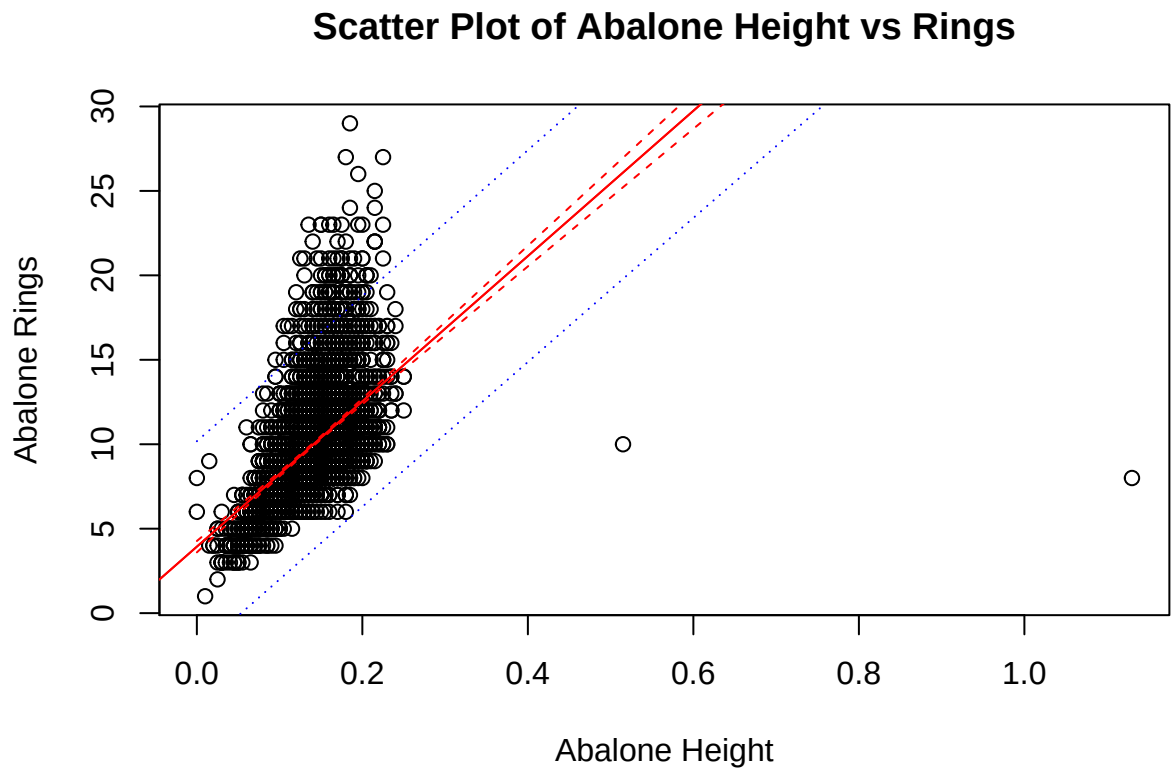


[1]

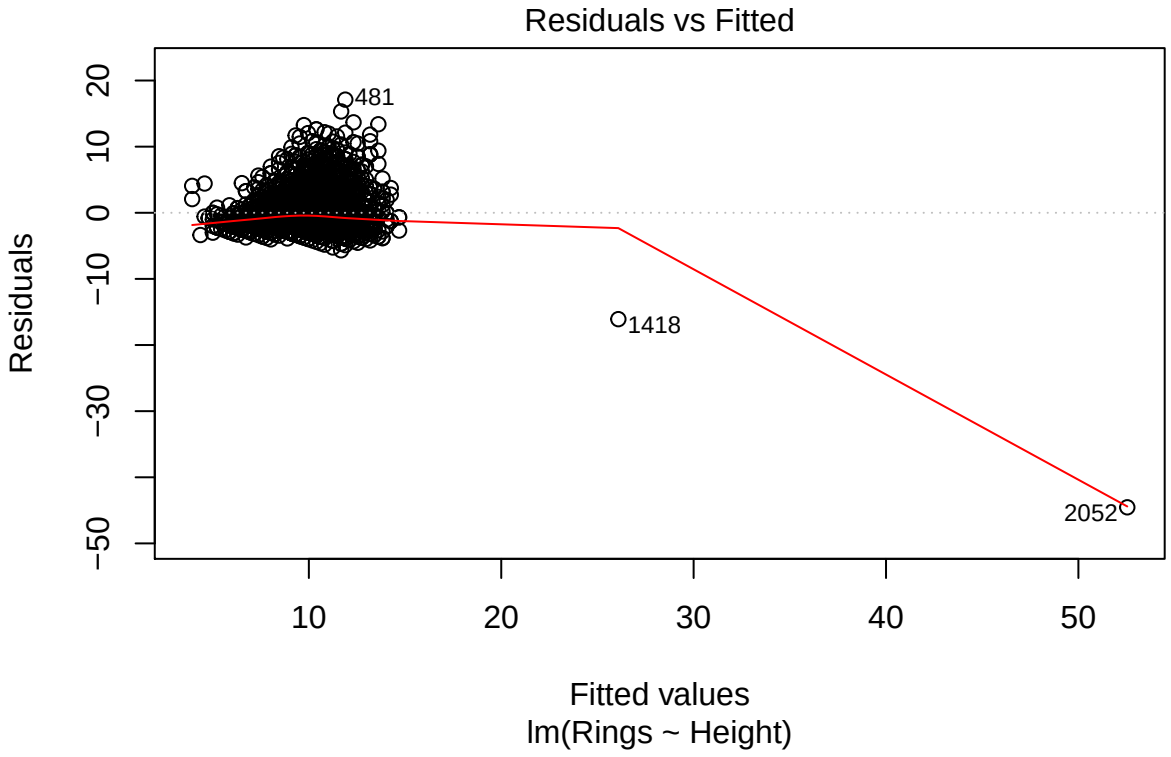
2210 481

There is a moderate positive relationship between Rings and Diameter. The left side of the plots exhibit more variation in the Rings response variable as the Diameter increases. There are many potential outliers in the scatter plot above the prediction interval (dotted blue line). Again, as identified by the fitted values vs the residuals, rows 481, 2109, and 2210 are specifically identified as potential outliers. The assumption of constant variance appears to hold. The QQ Plot of the residuals indicates the relationship may by curvilinear and not normally distributed.

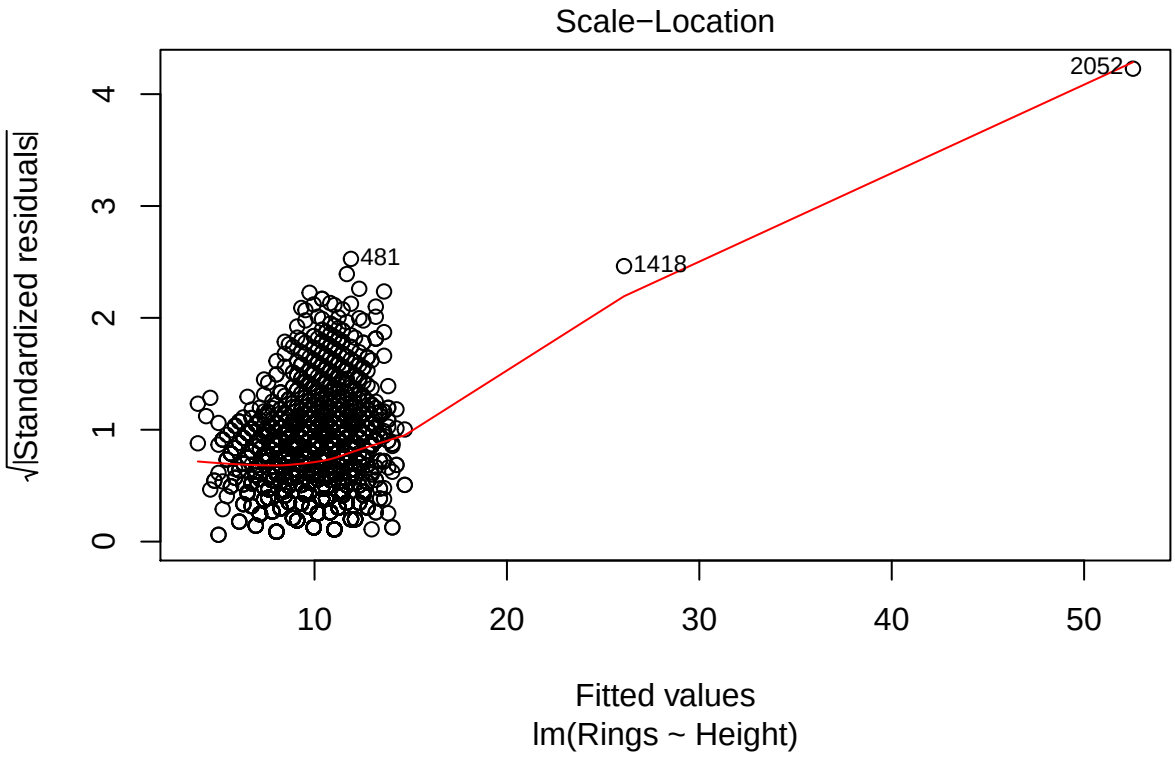
```
plot.confbands(
  abalone$Height,
  abalone$Rings,
  conf=0.99,
  xlab='Abalone Height',
  ylab='Abalone Rings',
  main='Scatter Plot of Abalone Height vs Rings')
model.height = lm(Rings ~ Height, data=abalone)
abline(model.height, col='red')
```



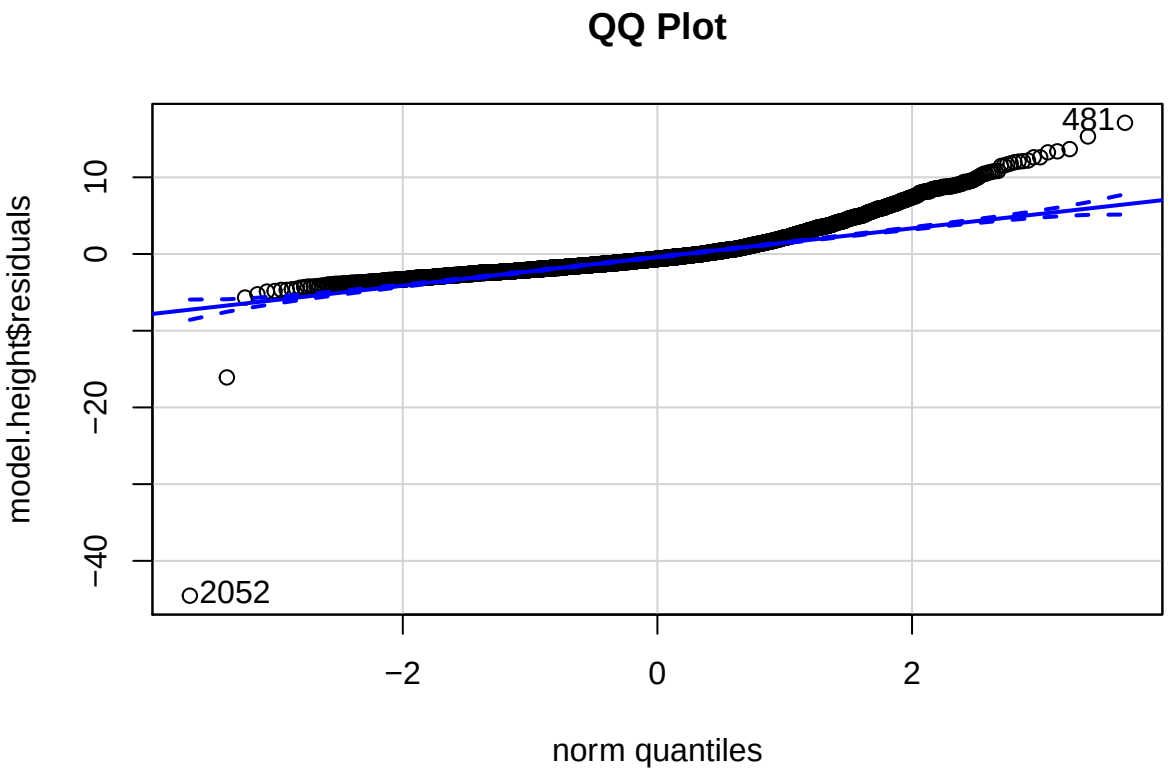

```
plot(model.height, 1)
```




```
plot(model.height, 3)
```




```
car::qqPlot(model.height$residuals, main='QQ Plot', pch=1)
```



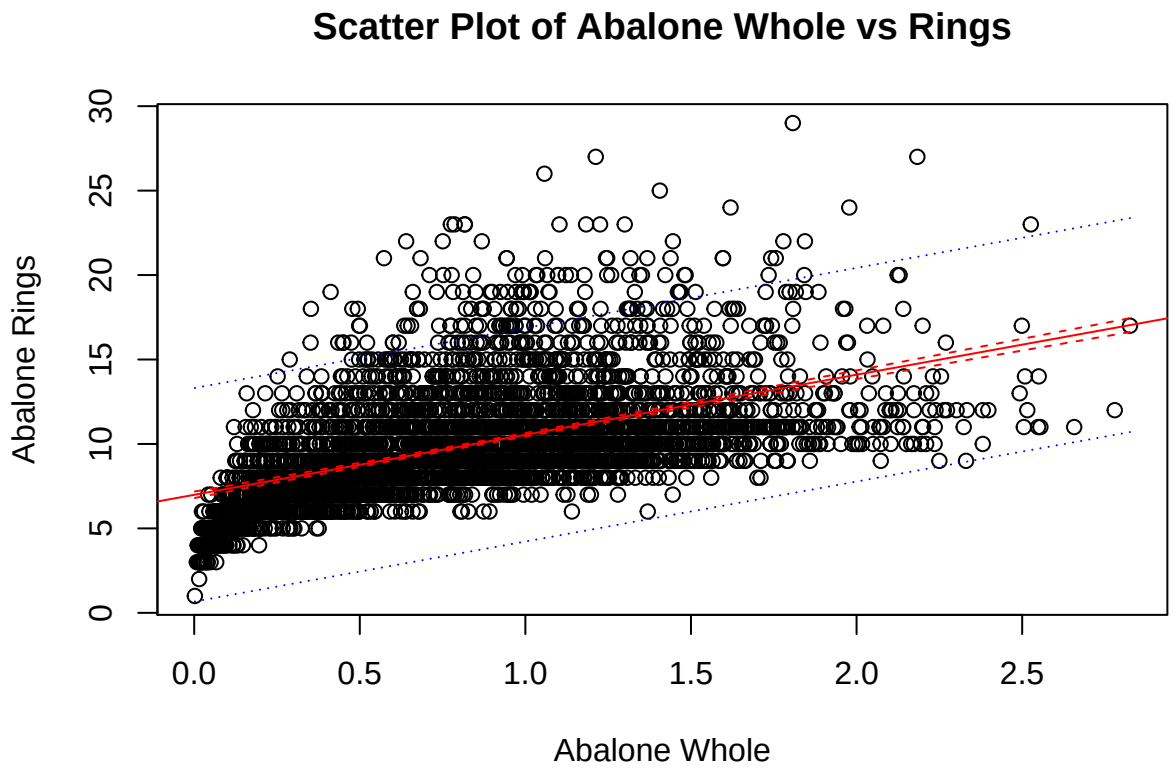
[1]

2052 481

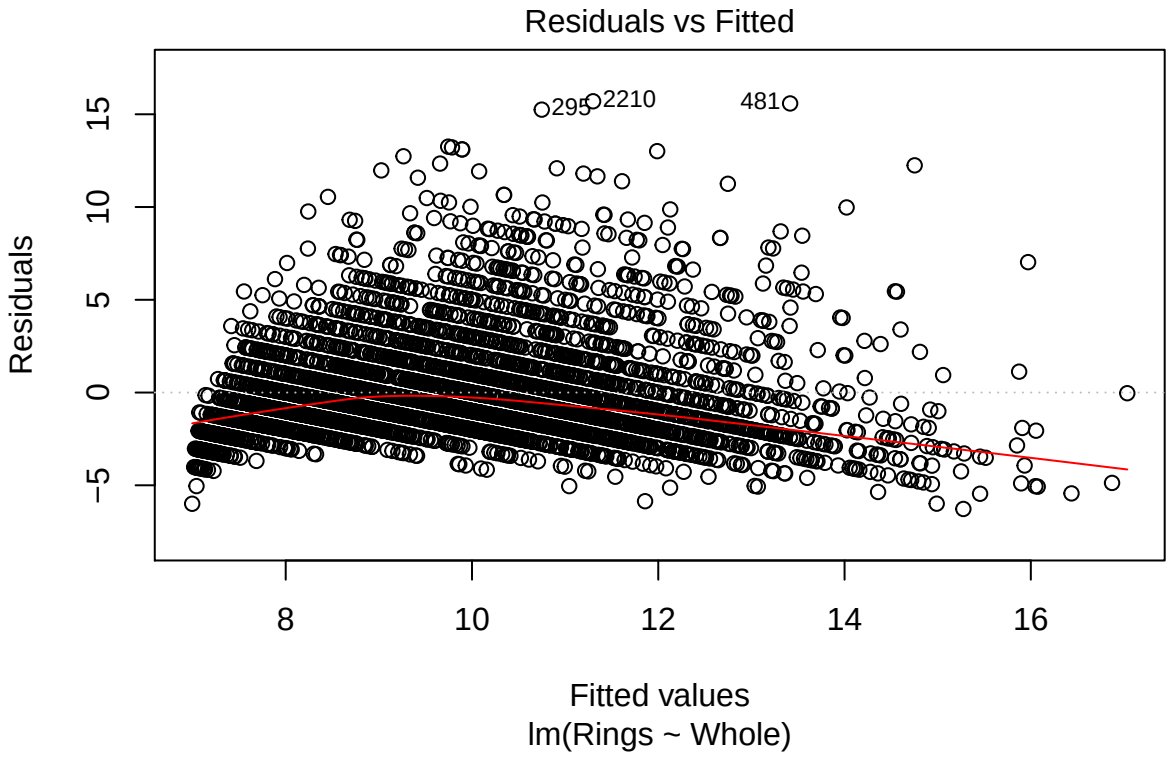
There is a strong positive relationship between Rings and Height. There are many potential outliers in the scatter plot above the prediction interval (dotted blue line), and 2 below. As identified by the fitted values

vs residuals scatter plot, rows 481, 1418, and 2052 specficially appear to be outliers. Furthermore, rows 1418 and 2052 adversely affect the constant variance assumption and should likely be removed. The assumption of constant variance doest not hold with rows 1418 and 2052 included, those outliers should likely be removed. The residuals are mostly normally distributed with rows 1418 and 2052 in the lower tail, and many outliers in the right tail.

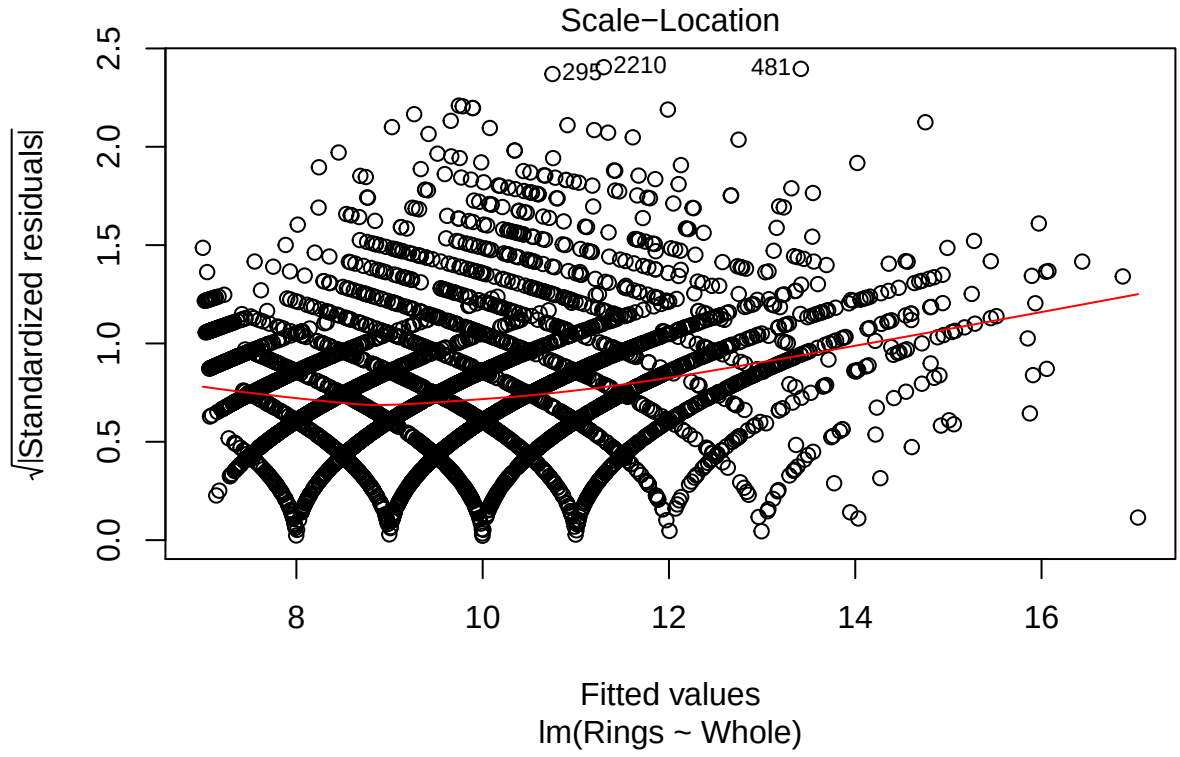
```
plot.confbands(
  abalone$Whole,
  abalone$Rings,
  conf=0.99,
  xlab='Abalone Whole',
  ylab='Abalone Rings',
  main='Scatter Plot of Abalone Whole vs Rings')
```



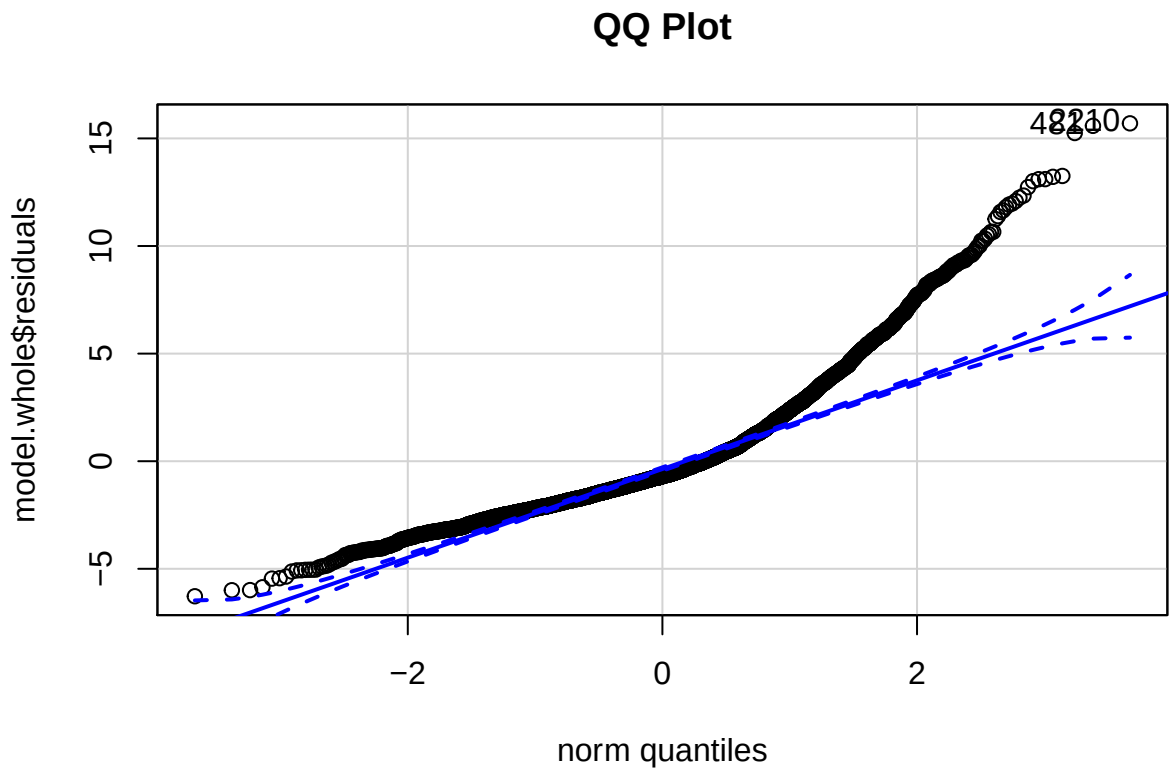

```
model.whole = lm(Rings ~ Whole, data=abalone)
plot(model.whole, 1)
```




```
plot(model.whole, 3)
```




```
car::qqPlot(model.whole$residuals, main='QQ Plot', pch=1)
```

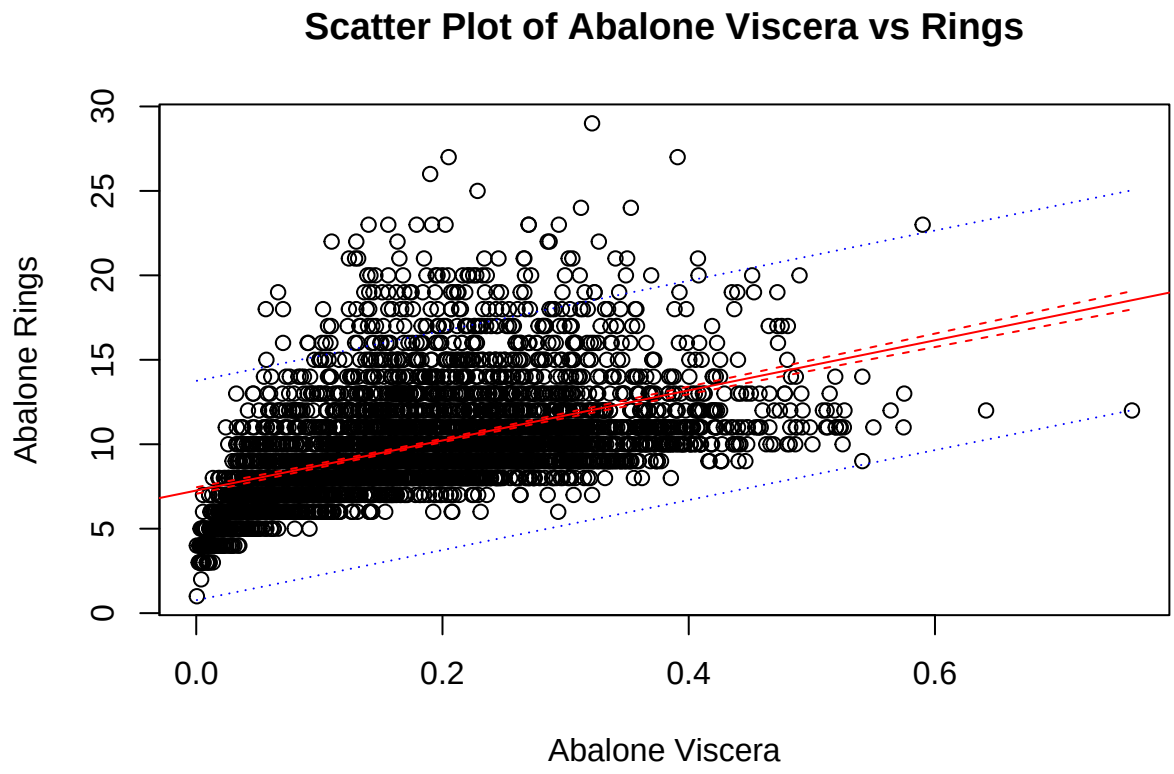


[1]

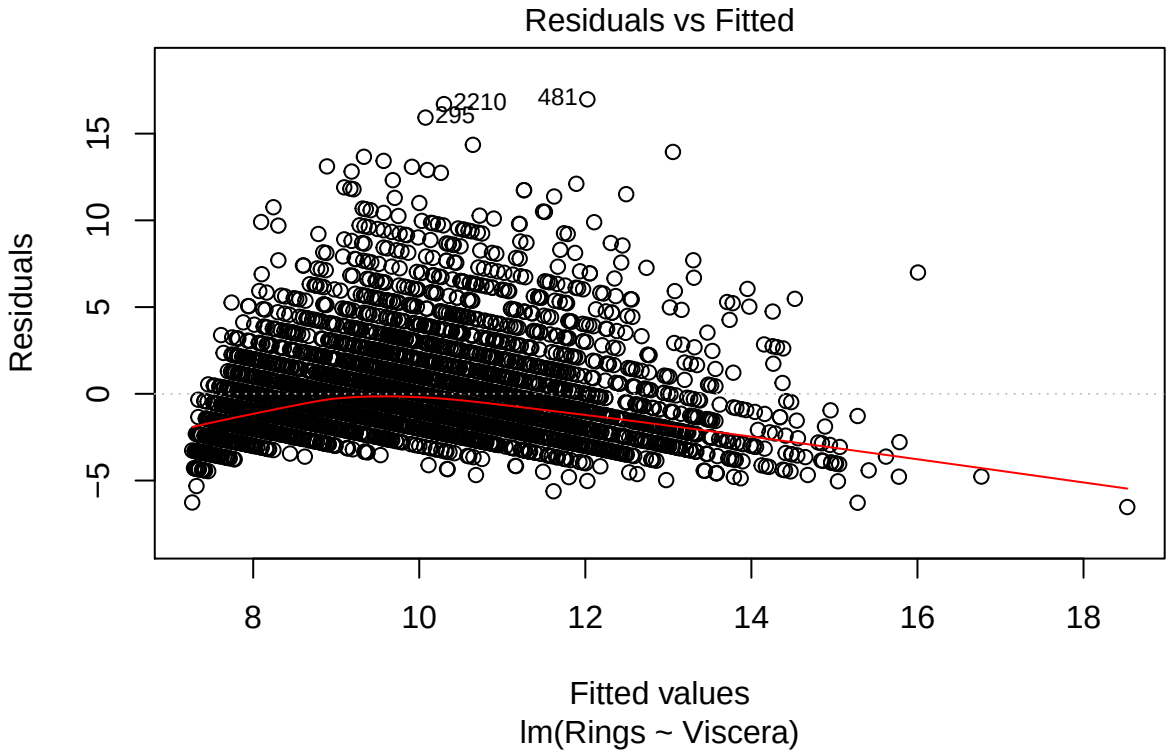
2210 481

There is a moderate positive relationship between Rings and Whole The left side of the plots exhibit more variation in the Rings response variable as Whole increases. There are many potential outliers in the scatter plot above the prediction interval (dotted blue line). Rows 295, 481, and 2210 are specifically identified as potential outliers from the fitted values vs residuals plot. The assumption of constant variance appears to hold. The QQ Plot of the residuals indicates the relationship may by curvilinear and not normally distributed.

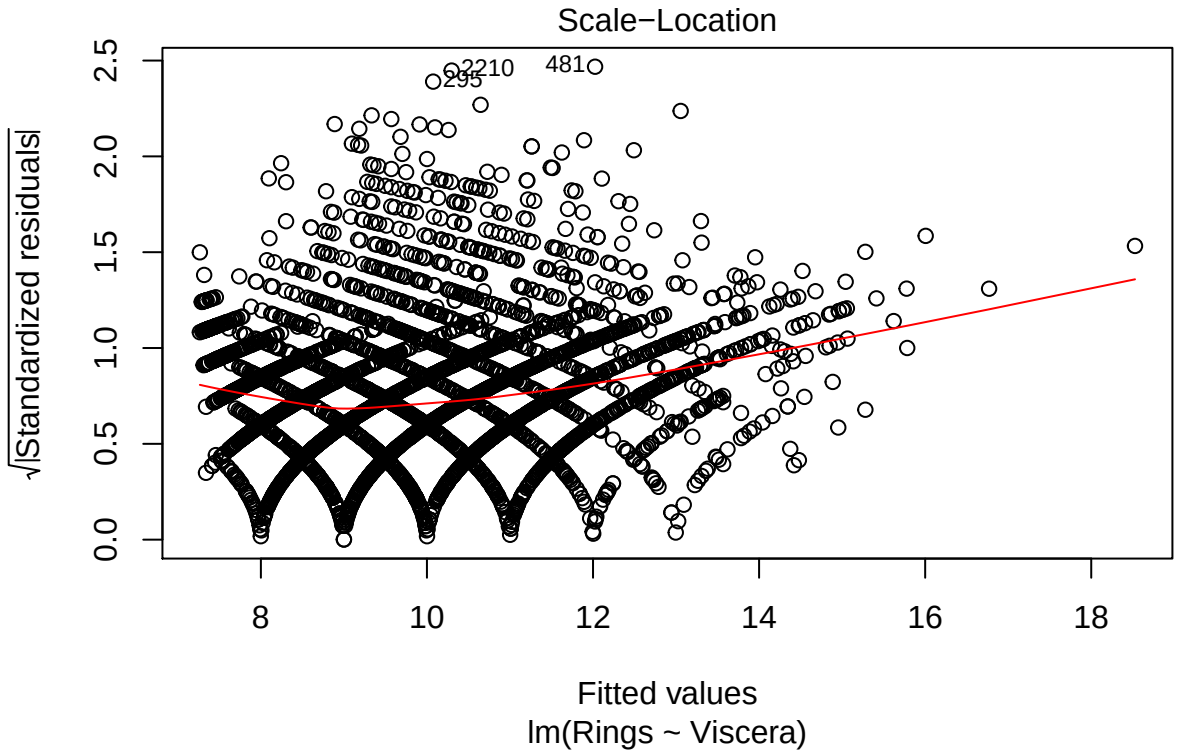
```
plot.confbands(
  abalone$Viscera,
  abalone$Rings,
  conf=0.99,
  xlab='Abalone Viscera',
  ylab='Abalone Rings',
  main='Scatter Plot of Abalone Viscera vs Rings')
```



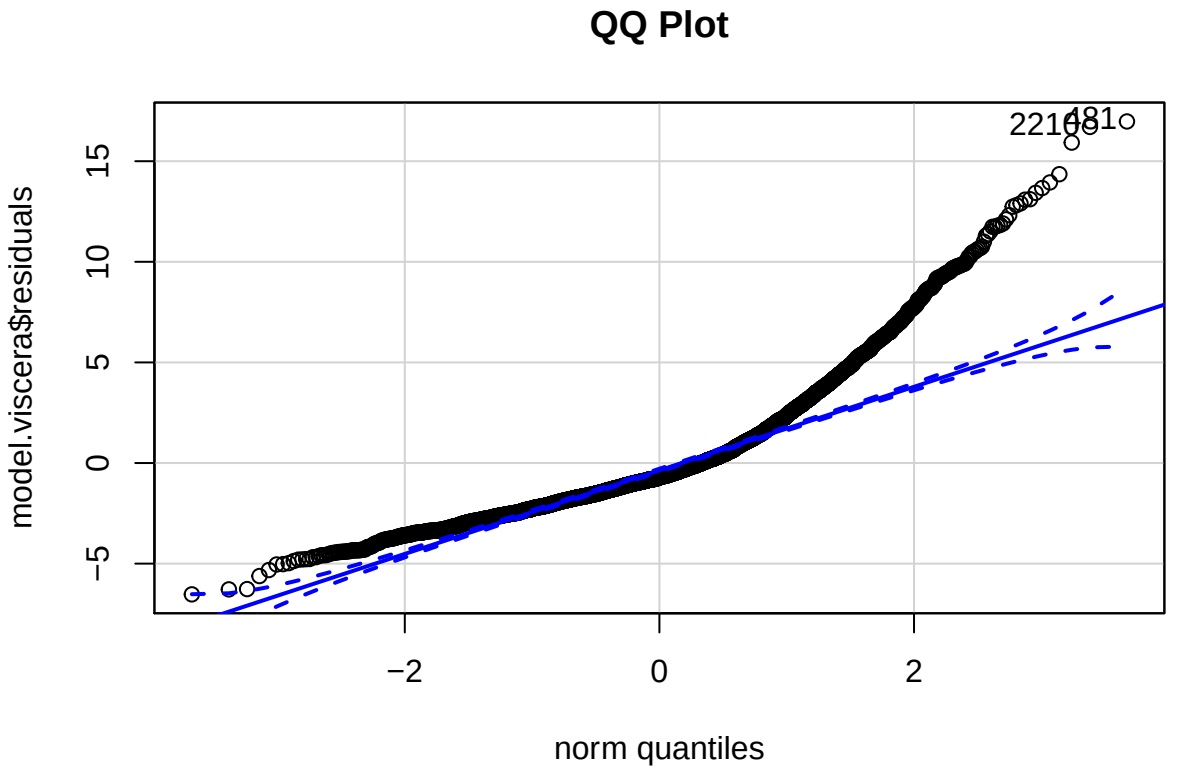

```
model.viscera = lm(Rings ~ Viscera, data=abalone)
plot(model.viscera, 1)
```



```
plot(model.viscera, 3)
```



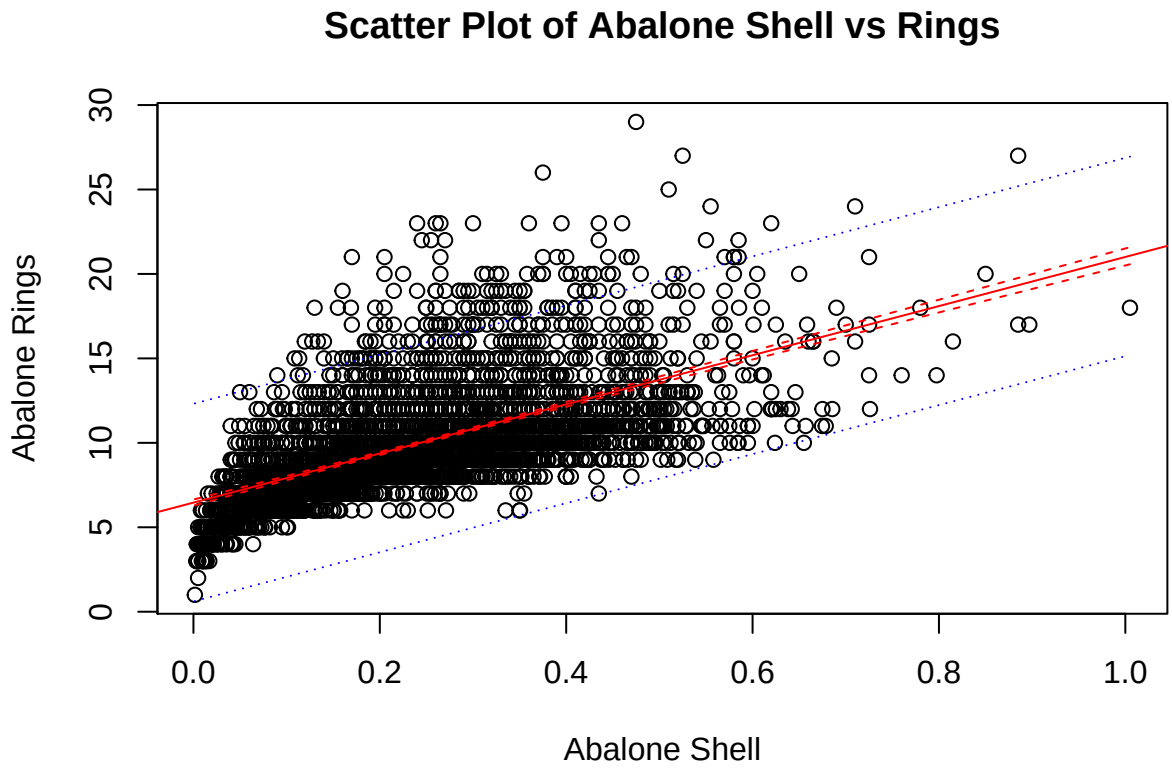

```
car::qqPlot(model.viscera$residuals, main='QQ Plot', pch=1)
```



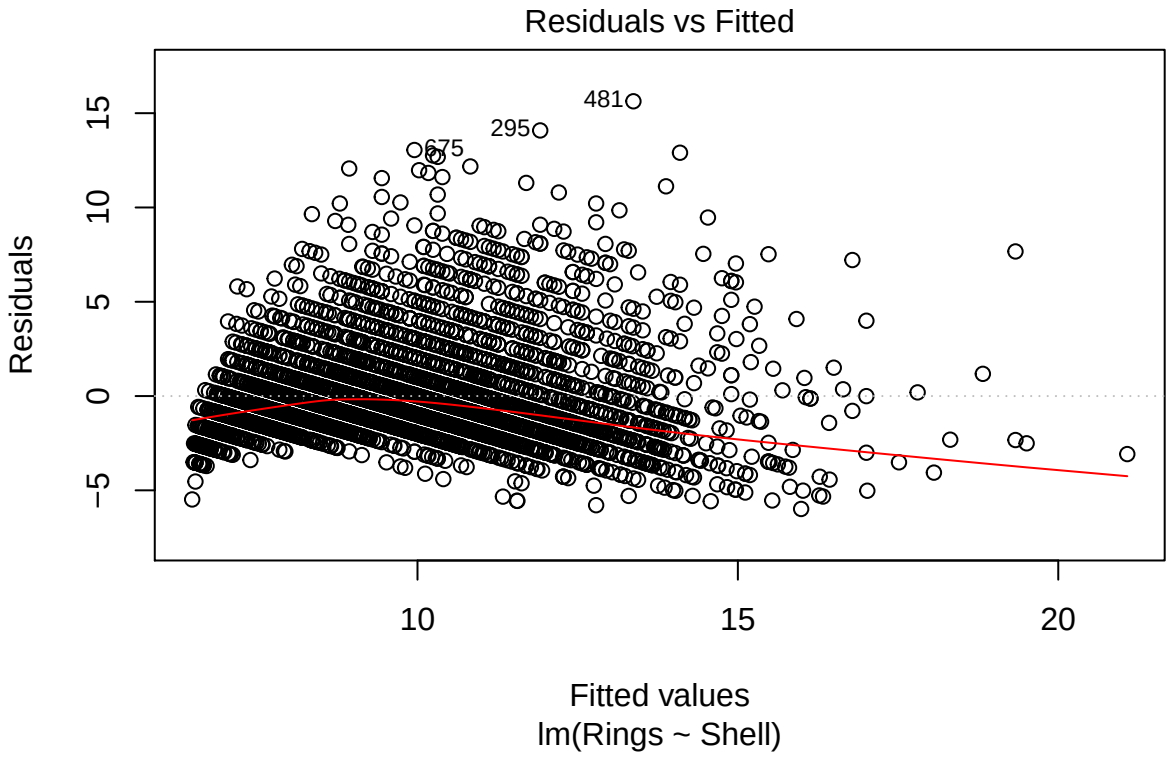

```
[1] 481
2210
```

There is a moderate positive relationship between Rings and Viscera The cnnter of the plots exhibits more variation in the Rings response variable as Viscera increases. There are many potential outliers in the scatter plot above the prediction interval (dotted blue line). Rows 295, 481, and 2210 again are specifically identified as potential outliers from the fitted values vs residuals plot. The assumption of constant variance appears to hold. The QQ Plot of the residuals indicates the relationship may by curvilinear and not normally distributed.

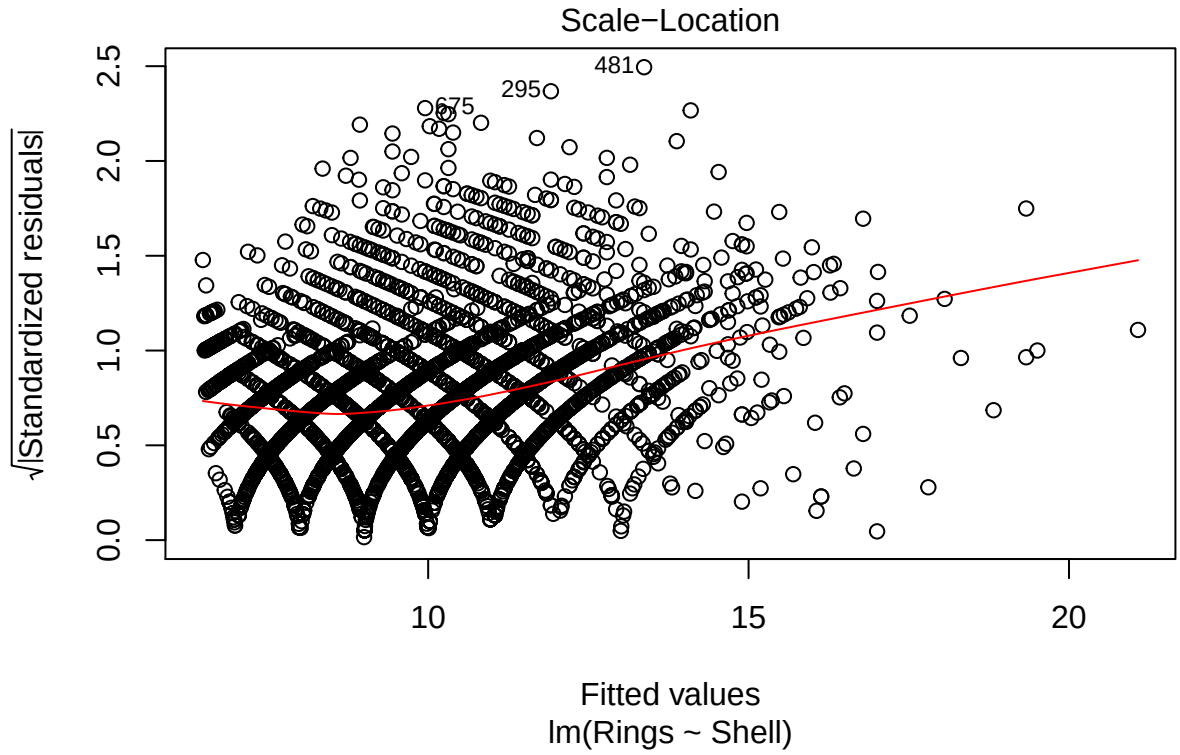
```
plot.confbands(
  abalone$Shell,
  abalone$Rings,
  conf=0.99,
  xlab='Abalone Shell',
  ylab='Abalone Rings',
  main='Scatter Plot of Abalone Shell vs Rings')
```



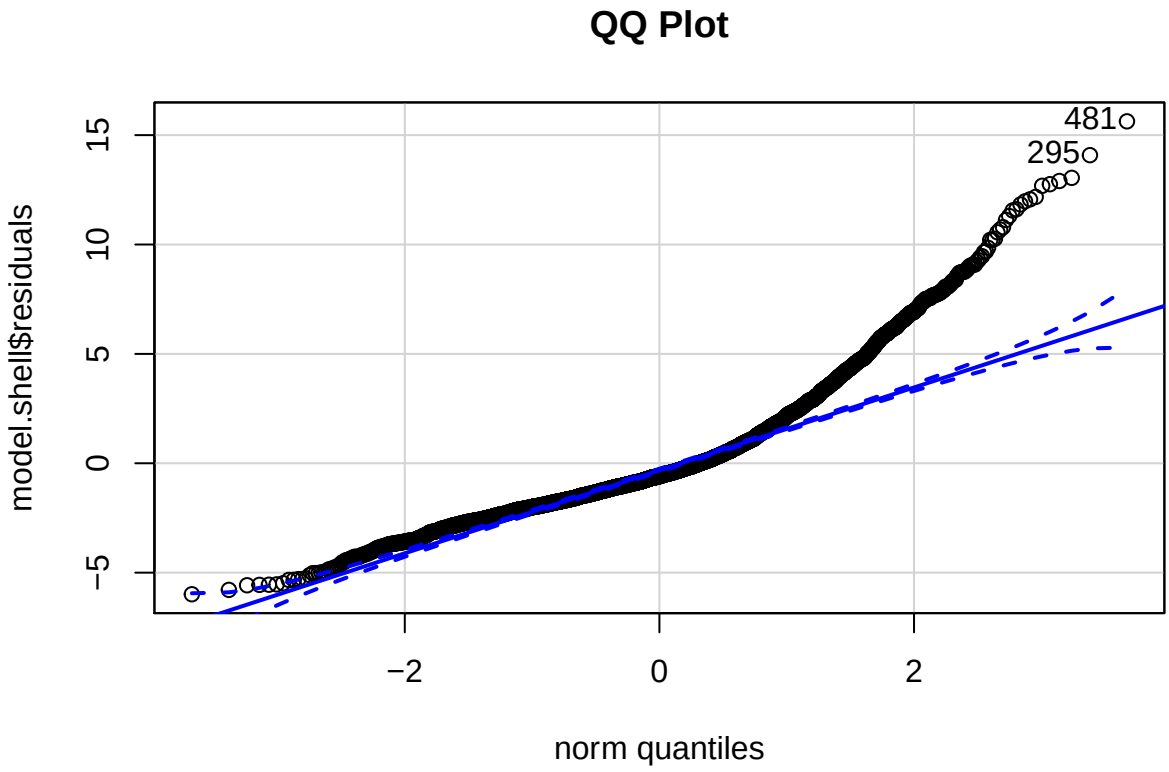

```
model.shell = lm(Rings ~ Shell, data=abalone)
plot(model.shell, 1)
```




```
plot(model.shell, 3)
```




```
car::qqPlot(model.shell$residuals, main='QQ Plot', pch=1)
```



[1] 481

295

There is a moderate positive relationship between Rings and Shell The cnnter of the plots exhibits more variation in the Rings response variable as Shell size increases. There are many potential outliers in the scatter plot above the prediction interval (dotted blue line). Rows 295, 481, and 675 are specifically identified as potential outliers from the fitted values vs residuals plot. The assumption of constant variance appears to hold. The QQ Plot of the residuals indicates the relationship may by curvilinear and not normally distributed.

(c) Display the correlations between each of the variables. Interpret the correlations in the context of the relationships of the predictors to the response and in the context of multi-collinearity.

```
correlations = cor(abalone[,2:8])
xtable(correlations)
```

	Length	Diameter	Height	Whole	Viscera	Shell	Rings
Length	1.00	0.99	0.83	0.93	0.90	0.90	0.56
Diameter	0.99	1.00	0.83	0.93	0.90	0.91	0.57
Height	0.83	0.83	1.00	0.82	0.80	0.82	0.56
Whole	0.93	0.93	0.82	1.00	0.97	0.96	0.54
Viscera	0.90	0.90	0.80	0.97	1.00	0.91	0.50
Shell	0.90	0.91	0.82	0.96	0.91	1.00	0.63
Rings	0.56	0.57	0.56	0.54	0.50	0.63	1.00

The maximum correlation between predictor variables is 0.9868135 (Length vs Diameter). The minimum correlation between predictor variables is 0.7982600 (Height vs Viscera). The predictor variables are highly correlated. The maximum correlation between the predictor vs the reponse variable is 0.6276529 (Shell v Rings). The correlation between the predictors and response variable is much less significant than the correlations between predictor variables.

```
model=lm(Rings ~ ., data=abalone)
varinf = car::vif(model)
xtable(varinf)
```

	GVIF	Df	GVIF^(1/(2*Df))
Sex	1.53	2.00	1.11
Length	40.72	1.00	6.38
Diameter	42.38	1.00	6.51
Height	3.58	1.00	1.89
Whole	34.45	1.00	5.87
Viscera	16.26	1.00	4.03
Shell	13.00	1.00	3.60

Using a VIF cutoff of 5, there is multicolinearity present between the Length, Diameter, and Whole predictor variables. A VIF of > 2.5 for Viscera and Shell also indicate they may also have multicolinearity issues.

(d) Based on this exploratory analysis, is it reasonable to assume a multiple linear regression model for the relationship between *Rings* and the predictor variables?

No there is a curvilinear relationship in the predictor variables vs the Rings response variable, transforming the data should be explored.

Question 2: Fitting the Multiple Linear Regression Model [16 points]

Plot the full model for Rings without transforming the response variable or predicting variables.

(a) Build a multiple linear regression model, called model1, using the response and all predictors. Display the summary table of the model.

```
model1=lm(Rings ~ ., data=abalone)
summ(model1, digits = 6)
```

Observations	4167
Dependent variable	Rings
Type	OLS linear regression

F(8,4158)	466.540322
R^2	0.473025
Adj. R^2	0.472011

(b) Is the overall regression significant at an α level of 0.01?

The overall regression is significant at an α level of 0.01.

(c) What is the coefficient estimate for *Viscera*? Interpret this coefficient.

The coefficient estimate for Viscera is -2.389647. Holding all other predictor variables constant. A decrease in -2.389647 units in Viscera results in a 1 Unit increase in rings.

(d) What is the coefficient estimate for the *Sex* category Male? Interpret this coefficient.

The coefficient estmate for *Sex* category Male is -0.060532. Holding all other predictor variables constant. If the sex is Male there is a -0.060532 avg difference in predicted rings vs the other 2 sex categories of Infant (SexI=1) and Female(SexI=0 and SexM=0).

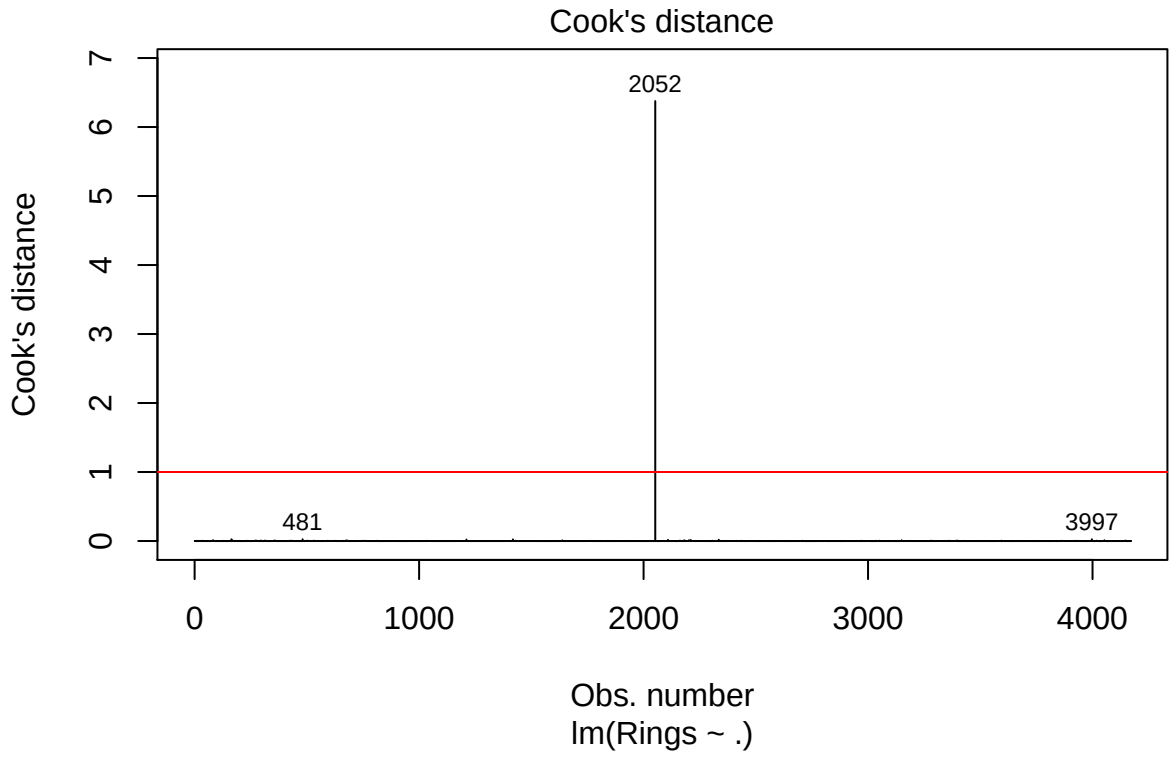
	Est.	S.E.	t val.	p
(Intercept)	4.794316	0.309116	15.509746	0.000000
SexI	-1.024825	0.109171	-9.387374	0.000000
SexM	-0.060532	0.089072	-0.679583	0.496806
Length	-3.777370	1.929218	-1.957980	0.050299
Diameter	11.587164	2.381649	4.865185	0.000001
Height	11.885169	1.642827	7.234583	0.000000
Whole	-5.574969	0.434721	-12.824241	0.000000
Viscera	-2.389647	1.335858	-1.788847	0.073712
Shell	25.703606	0.940330	27.334656	0.000000

Standard errors: OLS

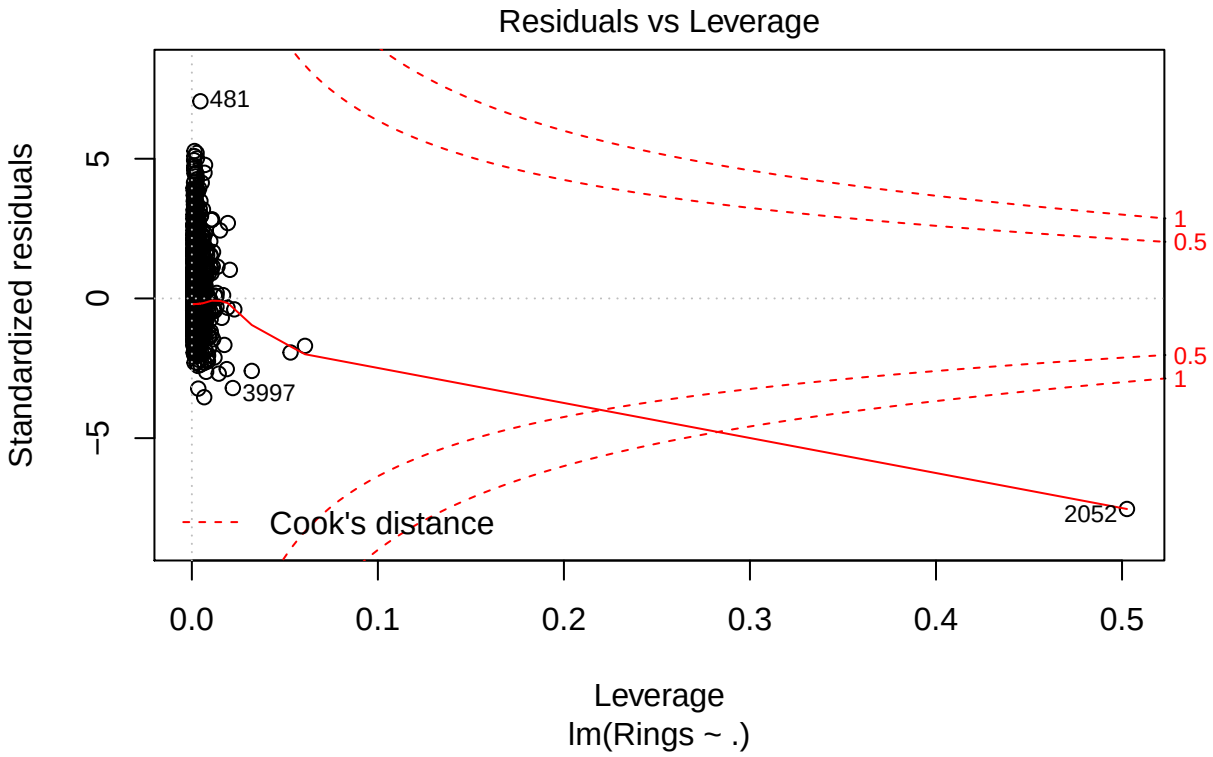
Question 3: Checking for Outliers and Multicollinearity [12 points]

(a) Create a plot for the Cook's Distances. Using a threshold Cook's Distance of 1, identify the row numbers of any outliers.

```
plot(model1, 4)
abline(h=1, col='red')
```




```
plot(model1, 5)
```



The only outlier with a cook's distance > 1 is Row 2052.

(b) Remove the outlier(s) from the data set and create a new model, called model2, using all predictors with *Rings* as the response. Display the summary of this model.

```
abalone2 <- abalone[-c(2052),]
model2=lm(Rings ~ ., data=abalone2)
summ(model2, digits = 6)
```

Observations	4166
Dependent variable	Rings
Type	OLS linear regression

F(8,4157)	479.992510
R^2	0.480176
Adj. R^2	0.479176

(c) Display the VIF of each predictor for model2. Using a threshold of 10 what conclusions can you draw?

```
varinf = car::vif(model2)
varinf
```

```
          GVIF Df GVIF^(1/(2*Df))
```

Sex 1.535728 2 1.113214 Length 40.767004 1 6.384904 Diameter 42.830157 1 6.544475 Height 6.220743 1 2.494142 Whole 34.503384 1 5.873958 Viscera 16.340894 1 4.042387 Shell 13.449515 1 3.667358

	Est.	S.E.	t val.	p
(Intercept)	4.434403	0.310680	14.273236	0.000000
SexI	-0.993005	0.108517	-9.150699	0.000000
SexM	-0.065384	0.088475	-0.739015	0.459940
Length	-4.303302	1.917485	-2.244243	0.024869
Diameter	9.727433	2.378286	4.090101	0.000044
Height	24.309732	2.311950	10.514817	0.000000
Whole	-5.447168	0.432123	-12.605584	0.000000
Viscera	-3.112687	1.330285	-2.339865	0.019338
Shell	24.376202	0.950250	25.652411	0.000000

Standard errors: OLS

```
xtable(varinf)
```

	GVIF	Df	GVIF^(1/(2*Df))
Sex	1.54	2.00	1.11
Length	40.77	1.00	6.38
Diameter	42.83	1.00	6.54
Height	6.22	1.00	2.49
Whole	34.50	1.00	5.87
Viscera	16.34	1.00	4.04
Shell	13.45	1.00	3.67

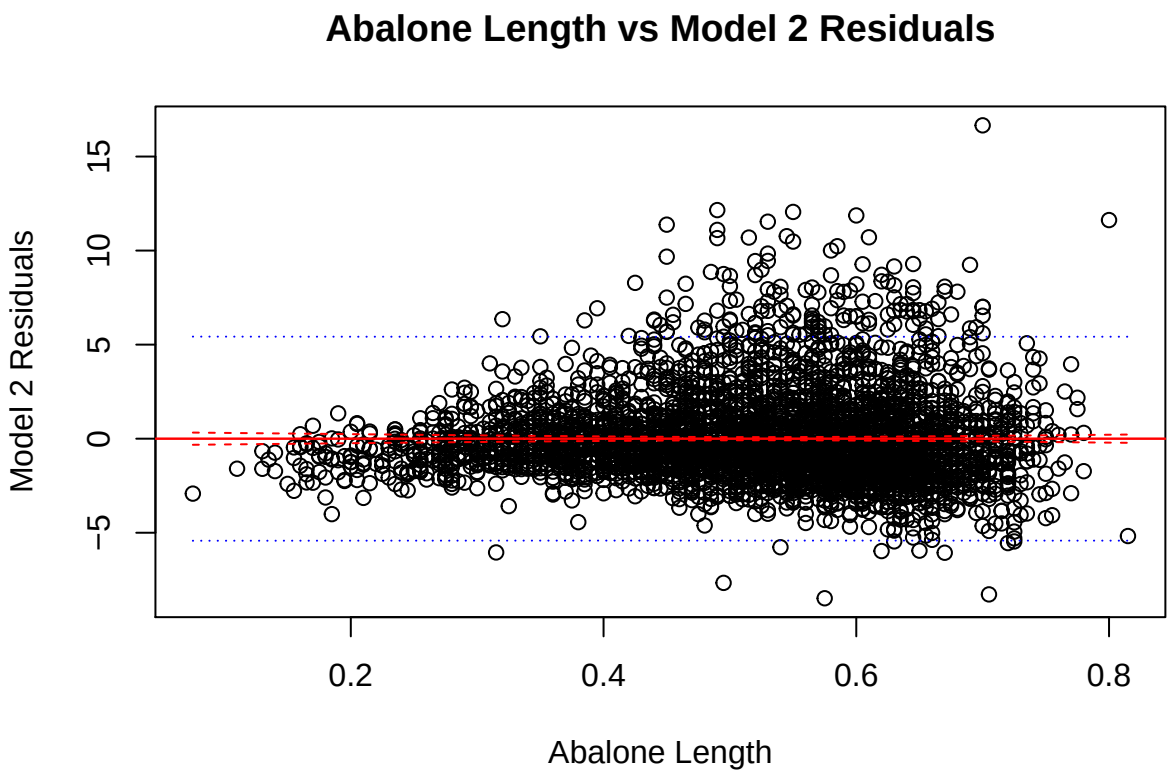
Using a VIF threshold of 10. There is not any significant multi-colinearity. Using a threshold of 10 is rather high in my opinion. I would prefer the use of 5, and would also be open to investigaating anything higher than 2.5 in more detail.

Question 4: Checking Model Assumptions [12 points]

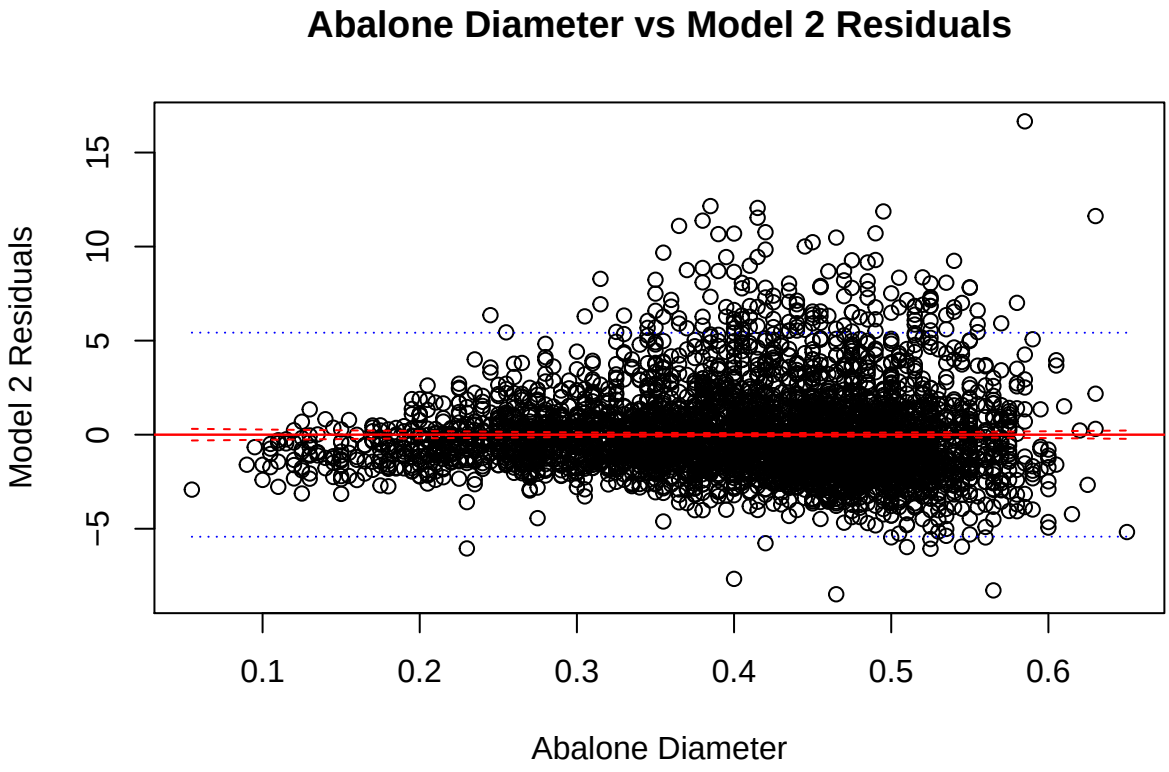
Please also use the cleaned data set and model2, which have the outlier(s) removed for the following questions.

(a) Create scatterplots of the standardized residuals of model2 versus each quantitative predictor. Does the linearity assumption appear to hold for all predictors?

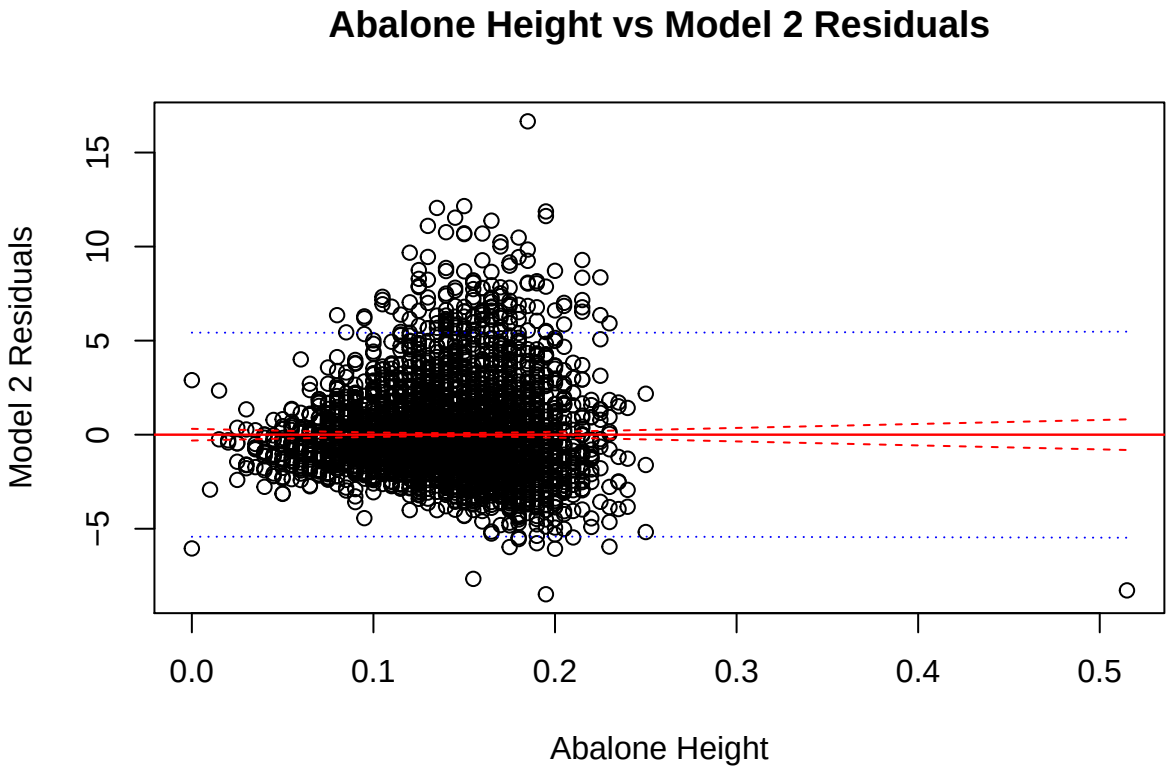
```
plot.confbands(
  abalone2$Length,
  residuals(model2),
  conf=0.99,
  xlab='Abalone Length',
  ylab='Model 2 Residuals',
  main='Abalone Length vs Model 2 Residuals')
abline(0,0, col='red')
```



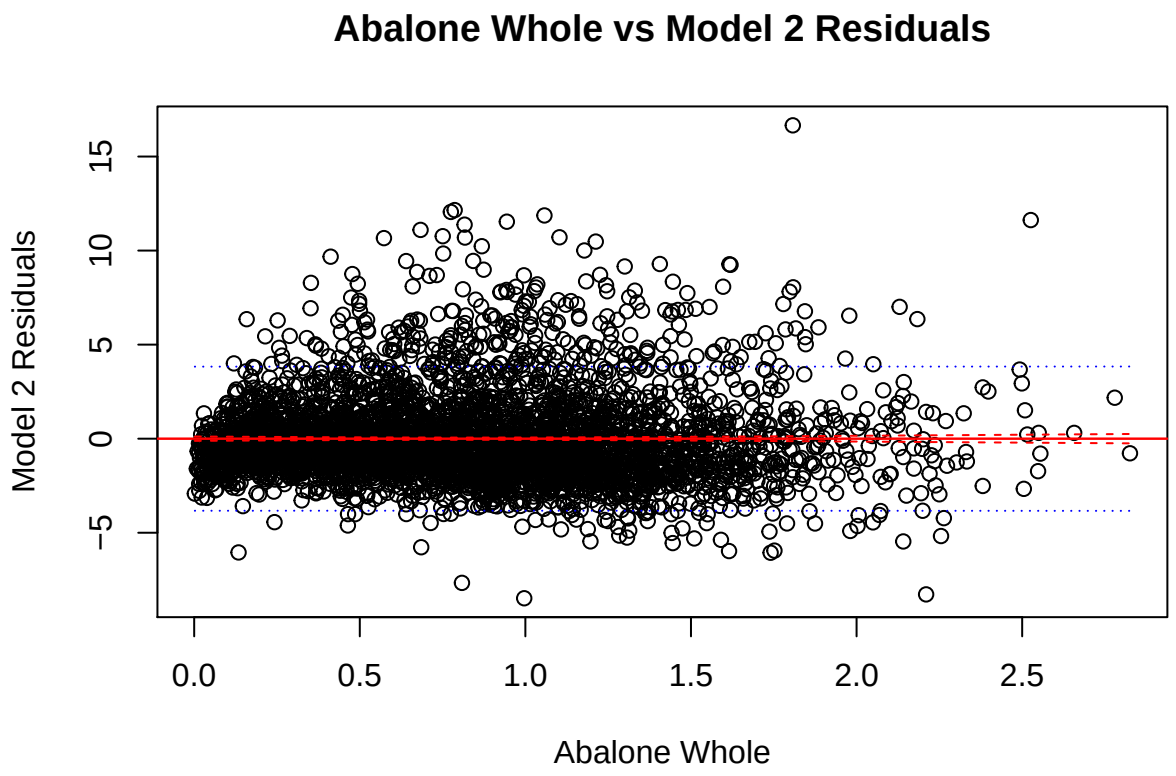

```
plot.confbands(
  abalone2$Diameter,
  residuals(model2),
  conf=0.99,
  xlab='Abalone Diameter',
  ylab='Model 2 Residuals',
  main='Abalone Diameter vs Model 2 Residuals')
abline(0,0, col='red')
```



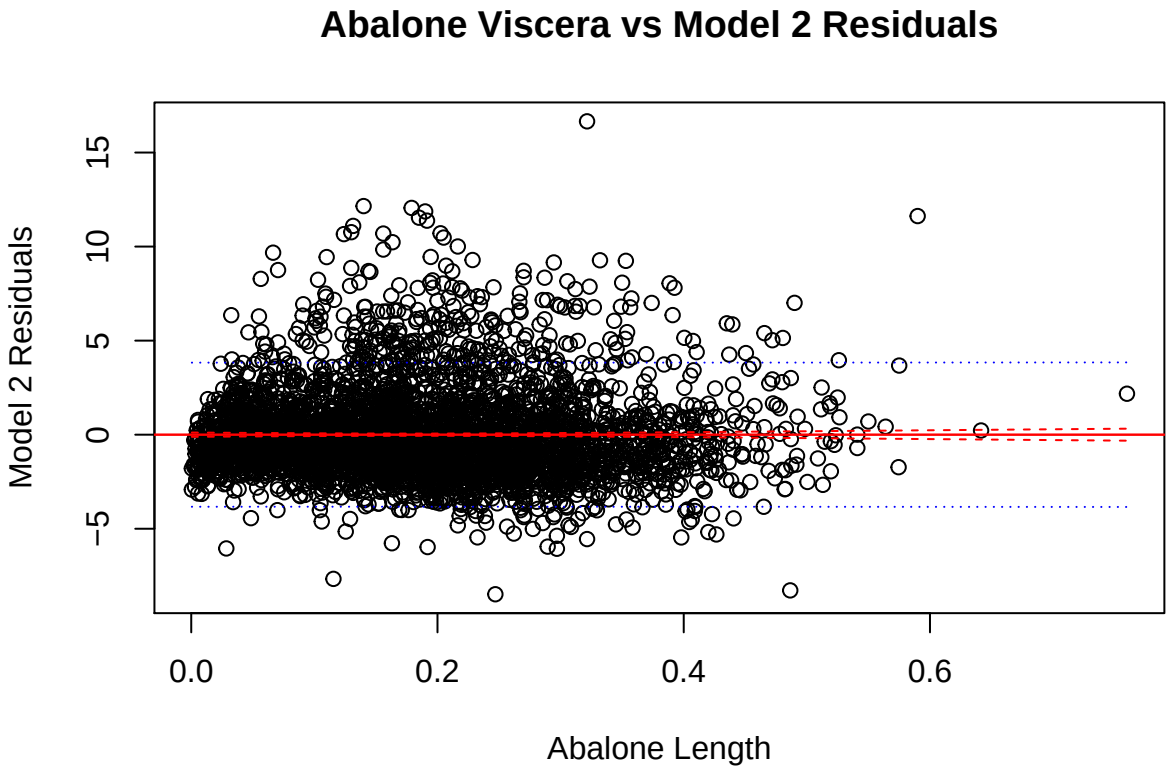

```
plot.confbands(
  abalone2$Height,
  residuals(model2),
  conf=0.99,
  xlab='Abalone Height',
  ylab='Model 2 Residuals',
  main='Abalone Height vs Model 2 Residuals')
abline(0,0, col='red')
```



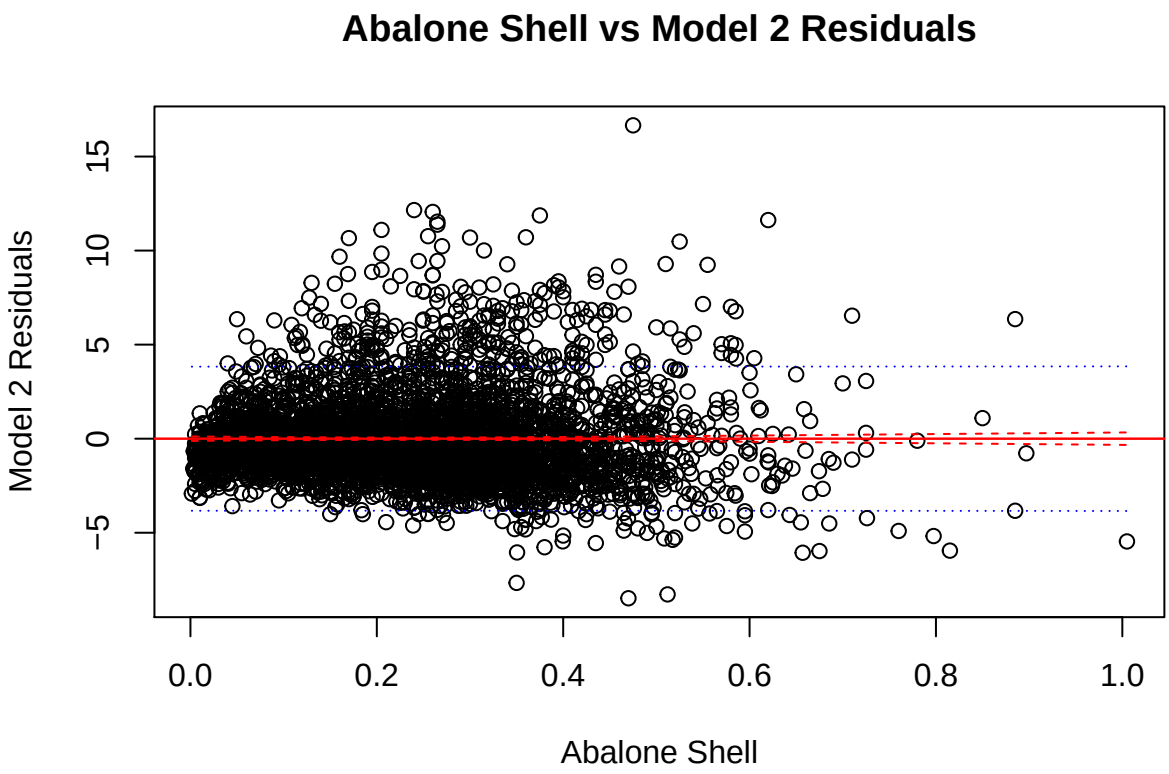

```
plot.confbands(
  abalone2$Whole,
  residuals(model2),
  xlab='Abalone Whole',
  ylab='Model 2 Residuals',
  main='Abalone Whole vs Model 2 Residuals')
abline(0,0, col='red')
```




```
plot.confbands(
  abalone2$Viscera,
  residuals(model2),
  xlab='Abalone Length',
  ylab='Model 2 Residuals',
  main='Abalone Viscera vs Model 2 Residuals')
abline(0,0, col='red')
```



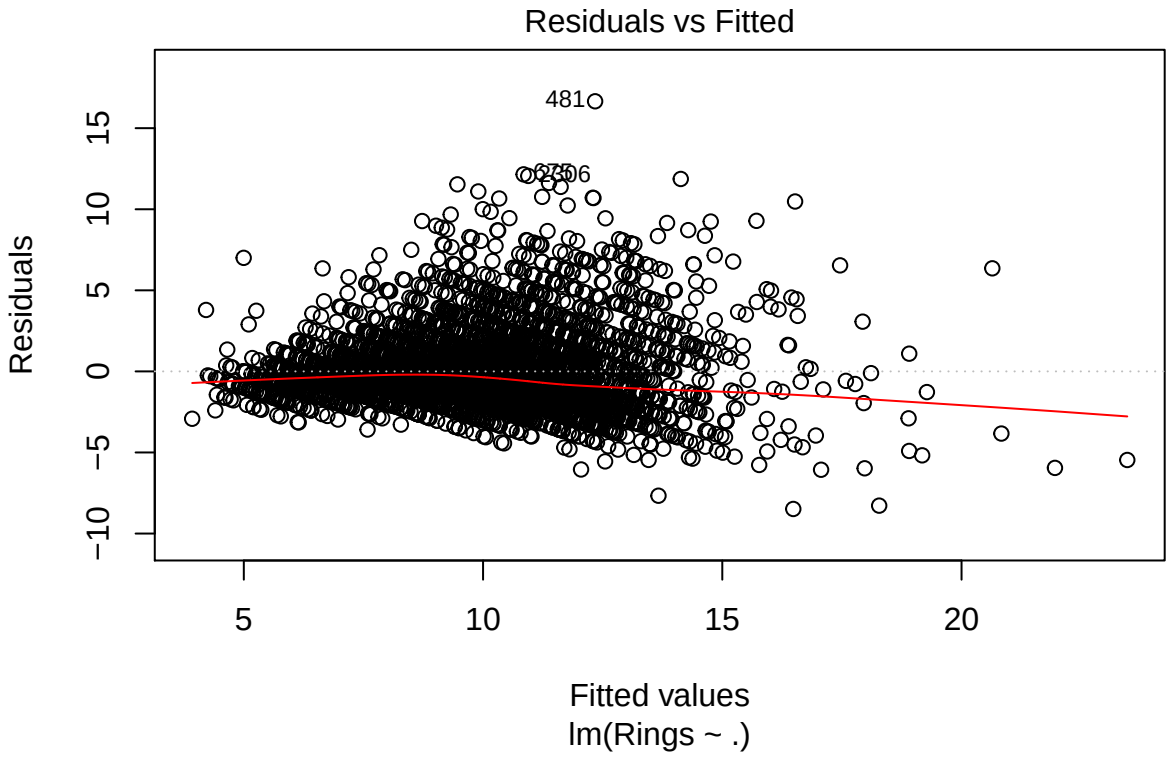

```
plot.confbands(
  abalone2$Shell,
  residuals(model2),
  xlab='Abalone Shell',
  ylab='Model 2 Residuals',
  main='Abalone Shell vs Model 2 Residuals')
abline(0,0, col='red')
```



The linearity Assumption holds for all quantitative predictor variables vs the models residuals.

(b) Create a scatter plot of the standardized residuals of model2 versus the fitted values of model2. Does the constant variance assumption appear to hold for all predictors? Do the errors appear uncorrelated?

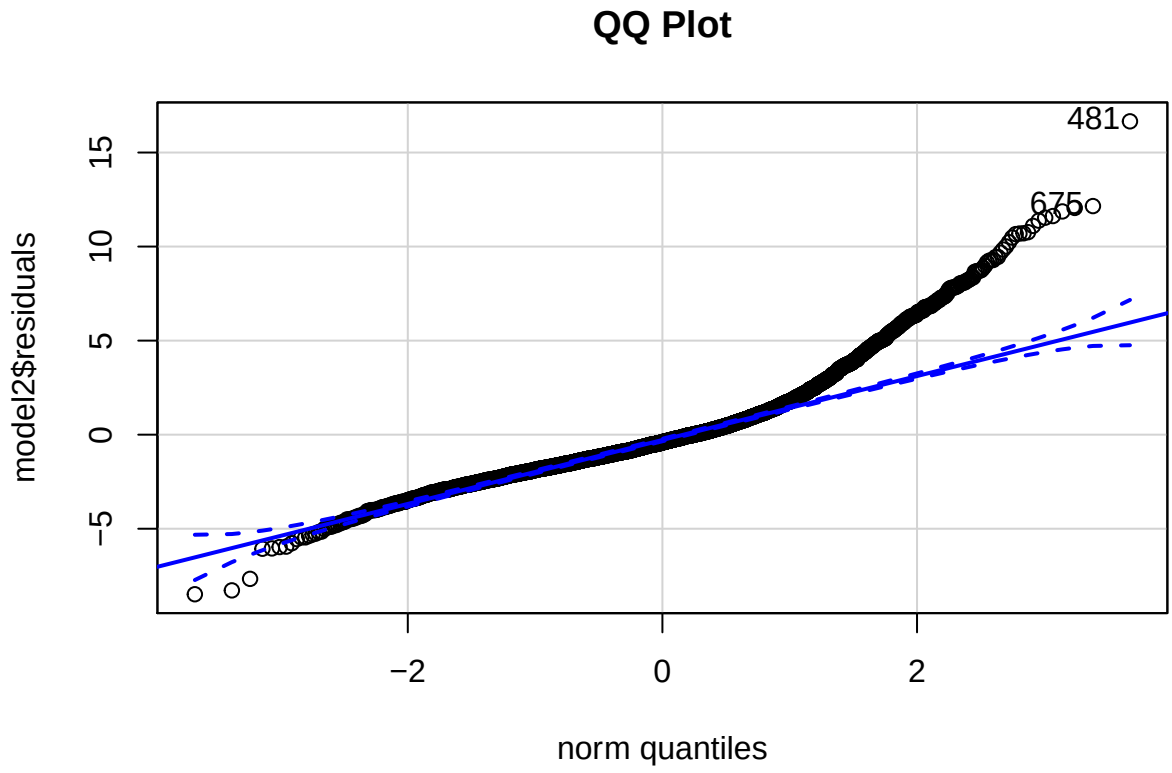
```
plot(model2, 1)
```



Constant variance appears to hold for all predictors.

(c) Create a histogram and normal QQ plot for the standardized residuals. What conclusions can you draw from these plots?

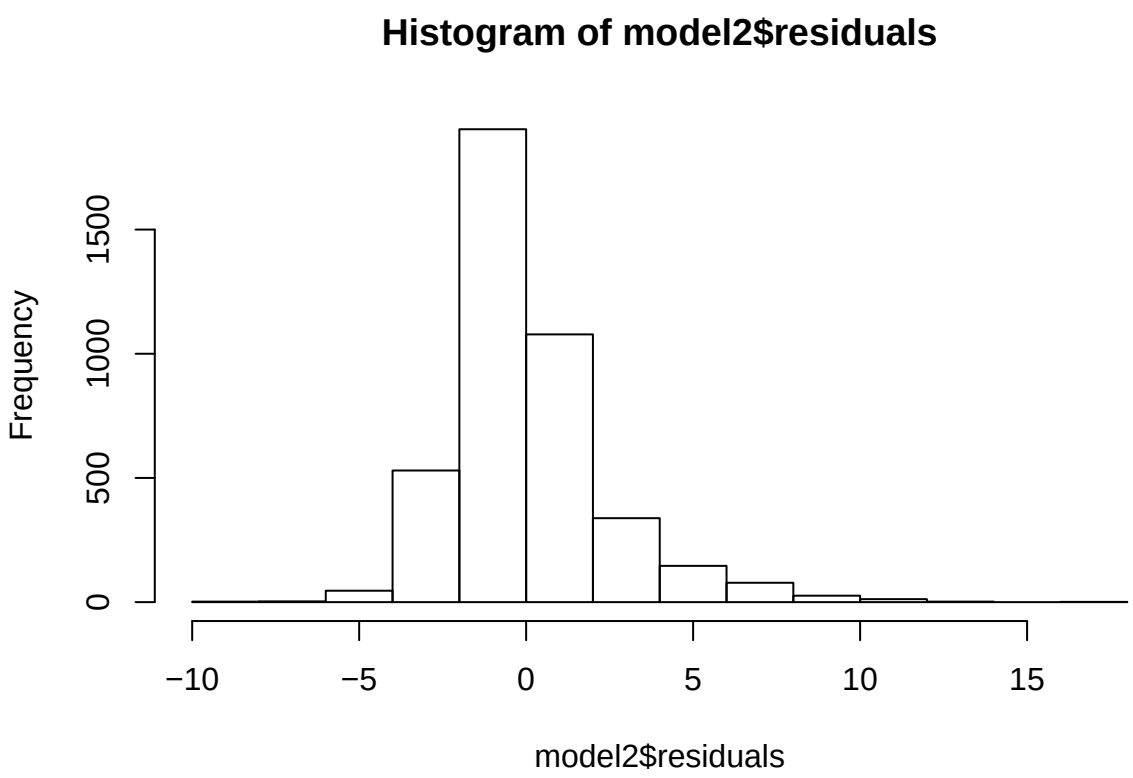
```
car::qqPlot(model2$residuals, main='QQ Plot', pch=1)
```



[1] 481

675

```
hist(model2$residuals)
```



The residuals of the model is not normally distributed and the data is skewed with a tail on the right hand side of the data. A transformation of the data should be explored.

Question 5 Model Comparison [12 points]

(a) Build a third multiple linear regression model using the cleaned data set without the outlier(s), called model3, using only *Sex*, *Length*, *Whole*, and *Shell*. Display the summary table of the model.

```
model3 <- lm(Rings~Sex+Length+Whole+Shell, data=abalone2)
summ(model3)
```

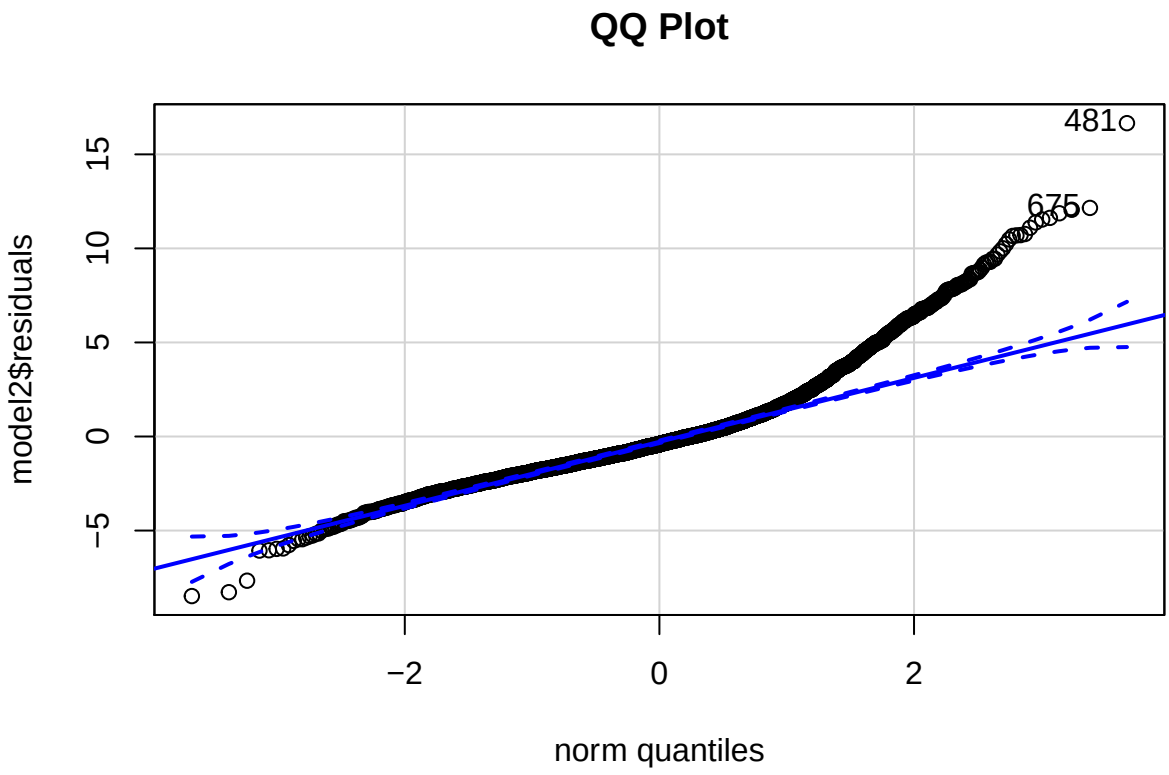
Observations	4166
Dependent variable	Rings
Type	OLS linear regression

F(5,4160)	714.29
R^2	0.46
Adj. R^2	0.46

	Est.	S.E.	t val.	p
(Intercept)	5.32	0.31	17.38	0.00
SexI	-1.14	0.11	-10.52	0.00
SexM	-0.08	0.09	-0.93	0.35
Length	6.41	0.82	7.85	0.00
Whole	-6.08	0.30	-20.45	0.00
Shell	28.01	0.90	31.10	0.00

Standard errors: OLS

```
car::qqPlot(model2$residuals, main='QQ Plot', pch=1)
```



[1] 481
675

(b) Compare and discuss the R-squared and Adjusted R-squared of model3 with model2.

model2 R-squared higher due to more variables included in the model vs model3.

(c) Conduct a partial F-test comparing model3 with model2. What can you conclude using an α level of 0.01?

```
anova(model3, model2)
```

Analysis of Variance Table

Model 1: Rings ~ Sex + Length + Whole + Shell Model 2: Rings ~ Sex + Length + Diameter + Height + Whole + Viscera + Shell Res.Df RSS Df Sum of Sq F Pr(>F)
1 4160 23348
2 4157 22556 3 791.33 48.613 < 2.2e-16 *** — Signif. codes: 0 ‘’ ***0.001*** ’’ *0.01* ’’ 0.05 ‘.’ 0.1 ’ ’ 1

We reject the null hypothesis that the coefficients added by extra predictors in model2 are 0. At least one added predictor in Model 2 has improved the predictive power of the model.

Question 6: Transforming the data [4 points]

(a) Find the optimal lambda, rounded to the nearest 0.5, for a Box-Cox transformation on model2. What transformation, if any, should be applied according to the lambda value?

```
boxcox(model2, lambda=seq(-1, 1, 0.5))
```

Results of Box-Cox Transformation

Objective Name: PPCC

Linear Model: model2

Sample Size: 4166

lambda PPCC -1.0 0.9068158 -0.5 0.9796076 0.0 0.9872216 0.5 0.9730151 1.0 0.9515636

The max PPCC of 0.9872216 occurs at lambda = 0. We should not pursue any transformation.

Question 7: Estimation and Prediction [8 points]

(a) Estimate Rings for the last 10 rows of data (abaloneTest) using both model2 and model3. Compare and discuss the mean squared prediction error of both models.

Model 2 predictions at a 95% prediction interval

```
model2.predict <- predict(model2, abaloneTest, interval='predict')
model2.predict
```

```
       fit      lwr      upr
```

4168 9.157485 4.588069 13.72690 4169 9.717423 5.147497 14.28735 4170 9.442757 4.870259 14.01525 4171 10.026313 5.457104 14.59552 4172 10.231324 5.661828 14.80082 4173 10.885592 6.314978 15.45621 4174 9.812320 5.242143 14.38250 4175 11.598167 7.024861 16.17147 4176 10.550124 5.980353 15.11989 4177 11.733285 7.158727 16.30784

Model 2 mean squared prediction error

```
model2.predict.mspe <- mean((model2.predict - abaloneTest$Rings)^2)
model2.predict.mspe
```

[1] 15.47243

Model 3 predictions at a 95% prediction interval

```
model3.predict <- predict(model3, abaloneTest, interval='predict')
model3.predict
```

```
       fit      lwr      upr
```

4168 9.236279 4.589180 13.88338 4169 9.829395 5.182057 14.47673 4170 8.847329 4.200306 13.49435 4171 10.397222 5.750871 15.04357 4172 9.968944 5.322401 14.61549 4173 10.526435 5.879782 15.17309 4174 10.444523 5.797951 15.09110 4175 10.561970 5.915534 15.20841 4176 10.965595 6.318681 15.61251 4177 11.806942 7.156238 16.45765

Model 3 mean squared prediction error.

```
model3.predict.mspe <- mean((model3.predict - abaloneTest$Rings)^2)
model3.predict.mspe
```

[1] 15.65711

Model 2 has a MSPE of 15.47243 and Model 3 has a MSPE of 15.65711, both at a 95% prediction interval. Model 2 is a better predictor since its MSPE is lower than Model 3's MSPE.

(b) Suppose you have found an adult female abalone with a length of 0.5mm, a whole weight of 0.4 grams, and a shell weight of 0.3 grams. Using model3, predict the number of rings on this abalone with a 90% prediction interval.

```
Length <- c(0.5)
Whole <- c(0.4)
Shell <- c(0.3)
Sex <- c('F')
predict(model3, data.frame(Length, Whole, Shell, Sex), interval='predict', level=0.9)
```

```
   fit      lwr      upr
```

1 14.49984 10.59104 18.40863

Rounding to the nearest integer, the prediction is 14 rings with 90% confidence the actual rings are somewhere between 11 and 18.