Stats 101A Final Project

1. Introduction

In crab farming, the ability to tell the age of a given crab is essential in order for farmers to know the ideal time to harvest them. In particular, after attaining a certain age, crabs grow at a negligible rate, so the ability to accurately estimate their age can help to minimize the costs associated with their upkeep and thereby maximize profits. The goal of this report is to estimate the age of crabs based on readily-observable physical attributes.

In order to complete this task, analysis was performed using linear regression on the dataset `CrabsAgePrediction.csv` which was found on Kaggle, uploaded by Gursewak Singh Sidhu. This dataset includes measurements of 3893 mud crabs found in the Boston area. These measurements include age, sex, height, length, diameter, weight, shucked weight, shell weight, and viscera weight. While the original dataset does include the weight of the crabs, this was not used in the analysis as shucked weight, shell weight, and viscera weight add up to equal the total weight. It was found that the model that best predicts the age of the mud crabs in the Boston area uses the `Shucked Weight`, `Shell Weight`, `Height`, `Diameter`, and an additional variable `is.intermediate`, which categorizes the sex of the crabs into either "intermediate" or "not intermediate," in order to predict the logarithm of the `Age`. This model eliminated variables that were not statistically significant and used a log transformation on the variable `Age` to reduce the nonlinearity issue it had with other variables.

The structure of this paper is as follows. In Section 2, the data is described. Specifically, in Section 2.1, summary statistics are provided and the distribution of each variable is determined graphically. In Section 2.2 the relationship between the variables is shown using both a scatterplot matrix as well as a correlation matrix. In Section 3, the results and their interpretation is discussed. In particular, in Section 3.1 three of the model candidates are introduced. In Section 3.2, the three models are compared and there is an explanation of how it was determined which predictive model is best. In Section 3.3 the `R` results of this predictive model are interpreted. In Section 3.4 the chosen predictive model is assessed using a few diagnostic tests. Finally, in Section 4 the project is summarized, limitations of this project are discussed, and the applicability of this model in a real world context is discussed. The final section of this report contains the link to the dataset, links to the articles that were referenced during analysis, and many of the graphs and tables that were not included in the body of the paper.

2. Data Description

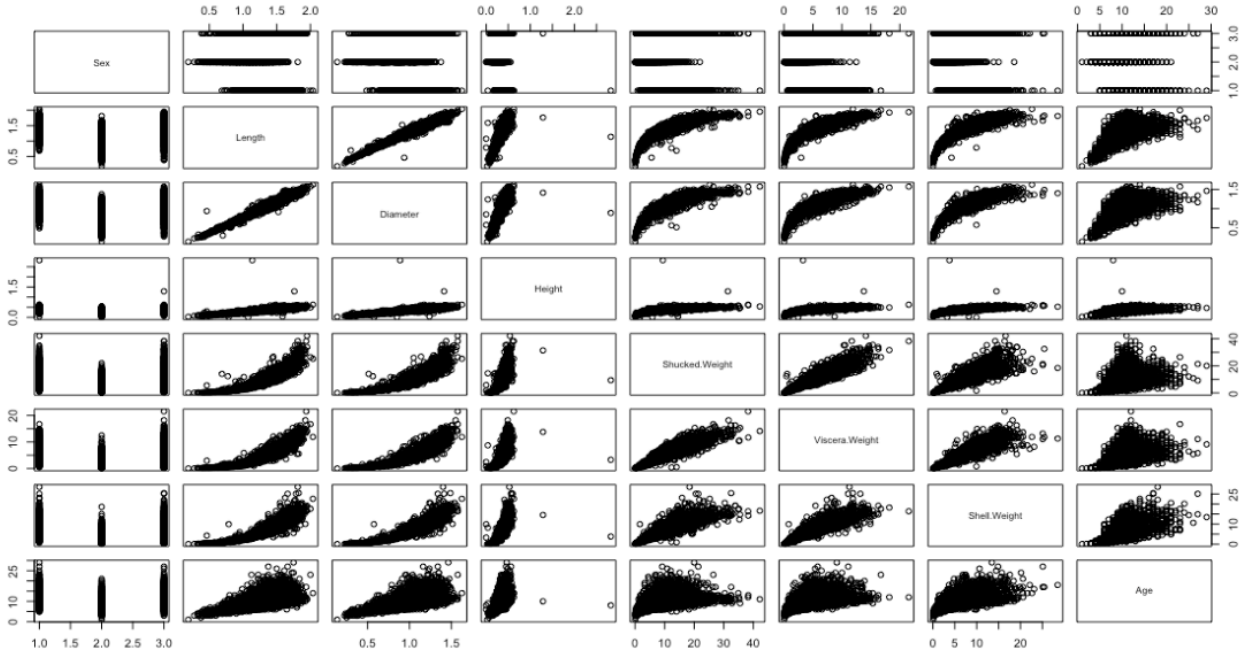
2.1 Summary Statistics and Distributions

The variable this report aims to predict is the age of mud crabs found in the Boston area. The seven original explanatory variables used to predict the age of the mud crabs were the sex, height, diameter, length, shell weight, shucked weight, and viscera weight of the crabs. In the appendix, there is a matrix of the summary statistics for each variable (`Sex` is listed separately as it is categorical and does not have the same summary statistics). Additionally, there is a histogram showing the distribution of each variable (instead of a histogram a bar graph was used for `Sex` since `Sex` is categorical).

It is evident that most of the variables have a unimodal distribution that is slightly skewed. While `Length` and `Diameter` are left skewed, `Shucked Weight`, `Shell Weight`, `Viscera Weight`, and `Age` are right skewed. A possible explanation for this may be that in the location that this particular sample was taken from it may be more unusual to find heavier crabs and older crabs as well as physically smaller crabs. The `Sex` variable has a uniform distribution with there being similar numbers of all three sexes and only slightly more males than intermediate or females. The histogram of `Height` indicates that there are outliers as while they are not visible on the graph, the range of the x axis indicates that there may be a few stand alone observations of a higher height. No other variables have extreme outliers, but since they do have skewed distributions, there could potentially still be less extreme outliers. However, for the purposes of our analysis, it was sufficient to just remove the most extreme outliers.

2.2 Correlation and association

In order to visualize the relationship between variables, the correlation matrix may be viewed in the appendix and a matrix of scatterplots of each variable plotted against each other variables may be viewed below.



The correlation matrix reflects how all of the explanatory variables (excluding `Sex` as it is categorical) have a strong positive association with each other as they all have correlations

larger than 0.85. The correlation matrix also shows that the response variable, `Age`, does not seem to have a strong correlation with any one response variable. The bottom row of the scatterplot matrix indicates that several of the explanatory variables, specifically `Height`, shucked weight, shell weight, and `Viscera Weight` have a nonlinear relationship with `Age` as the scatterplots appear somewhat curved. The scatterplot of length against `Age` as well as diameter against `Age` appear fan shaped. The scatterplot matrix also reflects how several of the explanatory variables have a strong, positive, linear association with each other.

3 Results and Interpretation

3.1 Three Model Candidates

In order to determine what model would best predict the age of mud crabs in the Boston area, several potential models were investigated before determining the best one. The three that will be discussed in depth are the original linear model with all seven of the original *x* variables, the reduced model that only considers statistically significant *x* variables and uses the version of the dataset after the `Height` outliers were removed, and the reduced model using the dataset after `Height` outliers were removed that includes a log transformation on `Age` and a conversion of the variable Sex from a categorical variable with three levels into the categorical variable with two levels `is.intermediate`. Out of all of the models that were produced, these three in particular will be spoken about in depth as they highlight the largest improvements that were made in order to create the best predictive model. See the Appendix for the summary table outputs of each model.

3.2 Selecting the Best Predictive Model

Before creating the original full model, the `Sex` variable was converted into a factor, whereas the original dataset had it as a character. This made running models using this variable go smoother and was key for being able to make graphs containing the variable. The original full model had several insignificant variables (`SexM`, `Length`, and `Viscera Weight`), so to begin with improving the model the variables `Length` and `Viscera Weight` were removed to produce a reduced model. SexM was not removed as it was one level of the `Sex` variable and the other level SexI was significant. The partial F-test (see Appendix) had a *p*-value that was greater than 0.05, so we fail to reject the null that the reduced model is better, so it was concluded that the reduced model was better than the full model.

Then, based on the plots of the *x* variables against `Age` it became evident that there were two potential outliers in the `Height` variable (see Appendix for graphs). After investigating these outliers, it became evident that these outliers should be removed as their values did not make sense in a real world context as one belonged to a crab that was reported as being nearly three feet tall and the other belonged to a crab reported as being nearly a foot and a half tall. Not only were both of these values over two standard deviations away from the mean, but they also do not make sense as mud crabs do not grow to be that tall, so these values were likely an error in the dataset. A new dataset was then created that did not include these outliers. Without these height outliers, the adjusted R-squared value of the model was larger than the adjusted R-squared value of the original full model and the adjusted R-squared value of the reduced model using the original dataset. The graph of `Height` vs. `Age` also significantly improved, however there was still fanning in many of the graphs of *x* variables against y (including `Height` vs. `Age`), and

some issues with nonlinearity. There was also a lot of fanning in the graphs of the *x* variables vs the standardized residuals. Additionally, `SexM` was still not significant.

In order to fix these issues, `Sex` was converted into a categorical variable with two levels called "`is.intermediate`" instead of the original three based on if the crab was intermediate or not since `SexI` was significant. While the adjusted R-squared stayed the same after altering this variable, this change made it so that all of the variables included in the model were statistically significant. In order to reduce fanning and nonlinearity the inverseResponsePlot() and powerTransform() functions were used to investigate possible transformations (see Appendix for output). Both indicated that a log transformation should be done on the response variable, `Age`. The lambda values produced using the powerTransform() function also indicated that a log transformation on `Shucked Weight` and `Shell Weight` as well as a square root transformation on `Diameter` may potentially improve the model but may not as the lambda values were not actually equal to 0 or 0.5. The model with a log transformation on `Age` that used the model without the `Height` outliers and used is.intermediate instead of Sex had a higher adjusted R-squared than all of the previous models as well as a residuals vs fitted plot with less fanning and a normal QQ plot that was straighter than before the transformation (see appendix for plots). The standard residual vs *x* variable plots also had less fanning. Overall, this transformation led to several improvements.

Afterwards, a model was created that also used a log transformation on shell weight and shucked weight, however while the adjusted R-squared did increase, the diagnostic plots were worse overall, especially the residual vs fitted plot, which has a significant increase in fanning (see Appendix for diagnostic plots). Performing a square root transformation on diameter slightly increased the adjusted R-squared, but it also produced worse diagnostic plots than the previous model. Based on the adjusted R-squared and the diagnostic plots, it was determined that the model using the dataset without `Height` outliers, a log transformation on `Age`, and the variable `is.intermediate` instead of `Sex` was the best of the predictive models that were tested.

3.3 R Results and Interpretation

The summary and ANOVA tables for the selected model may be found in the Appendix. This output is associated with the regression equation:

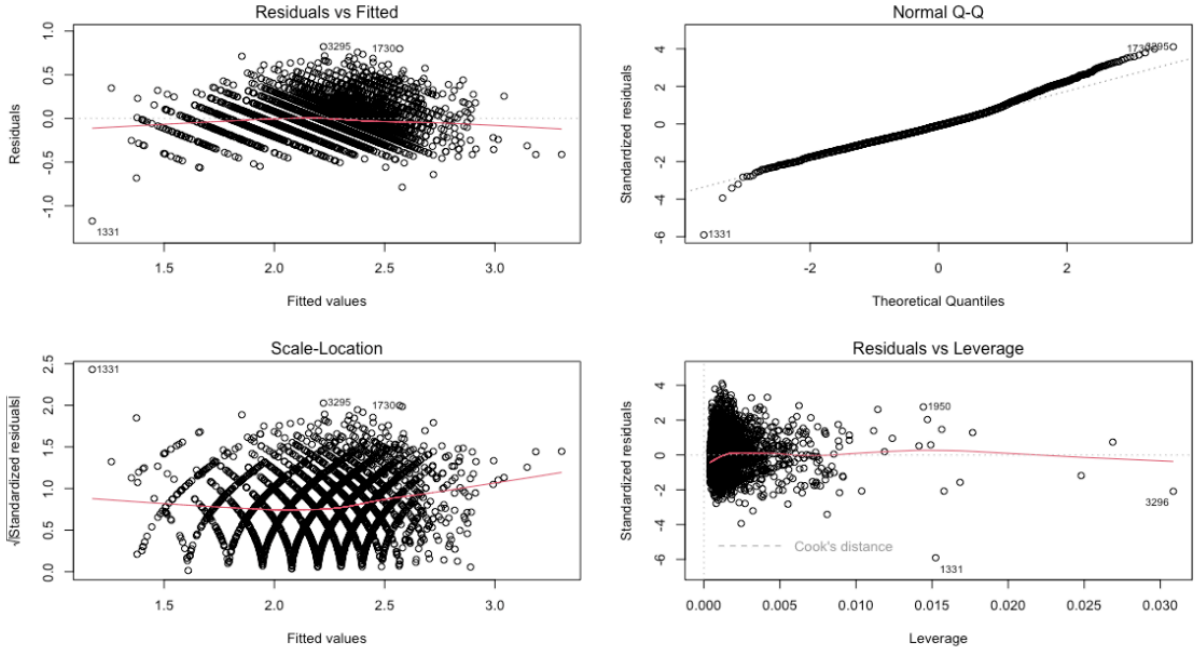
$$\widehat{log(Age)} = 1.336518 + \text{is.intermediate}(-0.096199) + \texttt{Diameter}(0.689296) + \texttt{Height}(1.033489) + \texttt{Shucked.Weight}(-0.038948) + \texttt{Shell.Weight}(0.040550).$$

The adjusted R-squared value of 0.5962 indicates that 59.62% of the variability in the log of the `Age` of mud crabs in the Boston area can be explained by Diameter, Sex (intermediate or not), Height, Shucked Weight and Shell Weight of the crabs. This is relatively low, which is likely due to both human error in obtaining the measurements as well as the fact that the physical characteristics of an animal alone are not always a good indication of its age.

The intercept value of 1.3336518 indicates that a crab that is not intermediate, has a diameter, height, shucked weight, and shell weight all equal to zero will have a log `Age` of

1.3336518 months. It does not make sense to interpret the intercept of this model in a real world context. When all other variables are held constant, an intermediate crab is predicted to be 0.00096% younger than a non intermediate crab. This is similar with all the other variables meaning that when all other variables are held constant when `Diameter` is increased by one foot, `Age` is predicted to increase by 0.00683%, when `Height` increases by 1 foot, `Age` is predicted to increase by 0.01034%, when `Shucked Weight` increases by 1 ounce, `Age` is predicted to decrease by 0.00039%, and when `Shell Weight` increases by 1 ounce, `Age` is predicted to increase by 0.00041%.

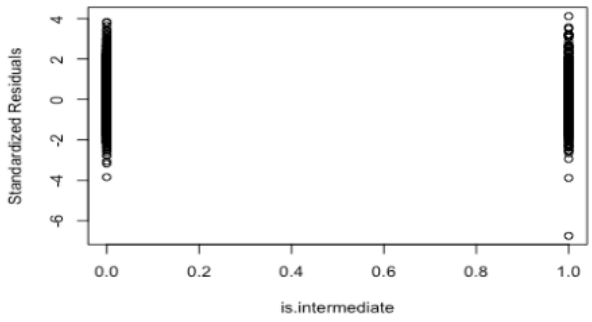
3.4 Assessing the Model - Diagnostic Plots and Multicollinearity
Below is the diagnostic plot for the model.

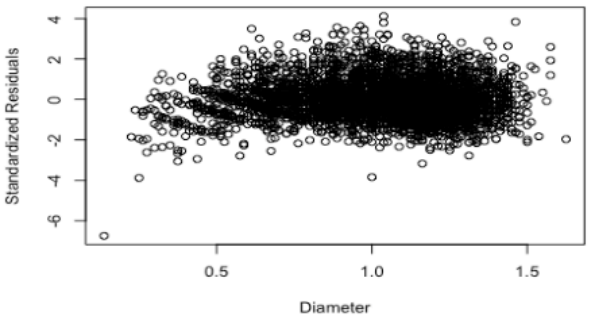


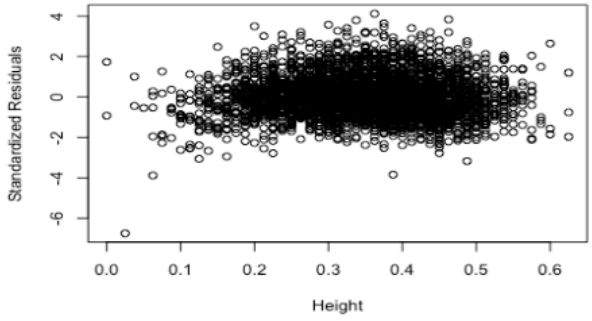
The red line in the fitted vs residuals plots is relatively flat, so the linear condition is satisfied. Of all of the models that were tested, the residuals vs fitted of this model has the least amount of fanning, and the constant variance condition appears to be satisfied. The normal QQ plot is mostly in line with the straight line, so the distribution of errors is nearly normal and only slightly tailed. The plot of the fitted values against the square root of standardized residuals indicates potential non-constant variance as the red line is curved but only slightly. The leverage vs. standardized residuals does indicate several potential high leverage points, but most of the points do lie in between standardized residuals = 4 and standardized residuals = -4. Overall, these diagnostic plots show a significant improvement over those associated with many of the other models tested, which largely led to this model being selected as the best one.

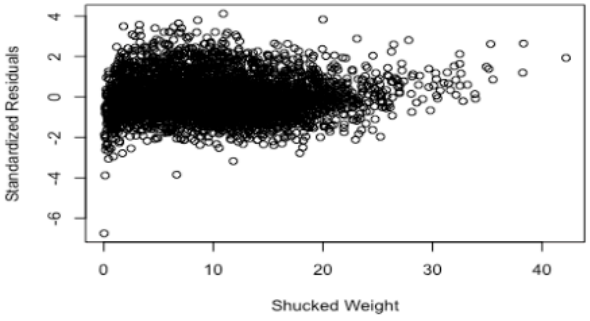
The *x* variable against standardized residual plots mostly have random scatter except the Shell Weight vs Standardized Residuals and the Shucked Weight vs standardized residuals have

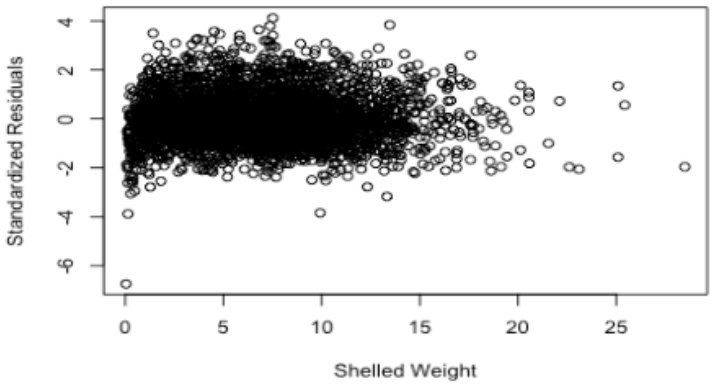
some slight fanning (see plots below). This indicates that this model mostly satisfies the constant variance condition.





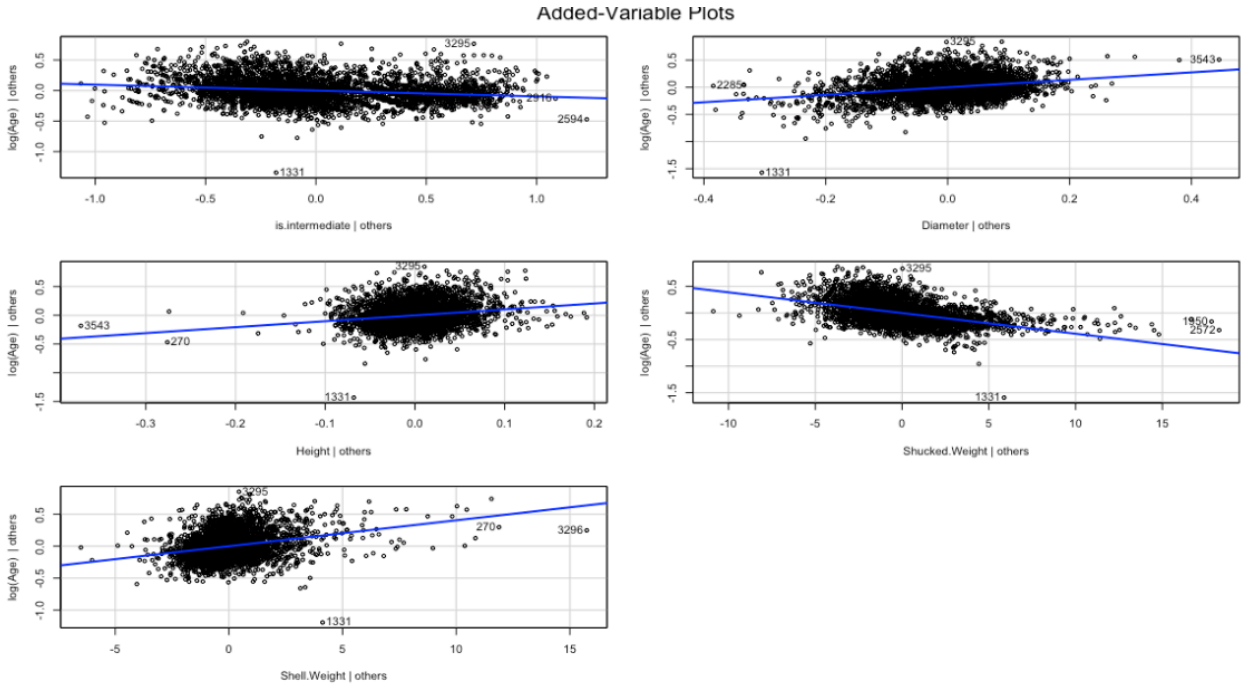






Since there was an issue with multicollinearity as shown by the VIF values of the *x* variables as well as the correlation matrix (see Appendix for values), added variable plots, backwards elimination from *p*-values, backwards AIC, backwards BIC, forward AIC, and backward BIC were used in order to determine if any of the variables should be removed. The added variable plots indicated that `is.intermediate` adds little to the prediction of the log of `Age` (See plots below). Backward elimination based on *p*-values could not be performed as all of the variables that were not statistically significant were removed and the remaining variables were equally statistically significant. The backward AIC, backward BIC, forward AIC, and backward BIC all indicated that the best model was the one that used all five of the predictor variables (see Appendix for outputs). In conclusion, even though the added variable plots

indicated `is.intermediate` added little to the prediction of the response variable, it was kept in the model.



4 Discussion

Through the use of various tools in R, it was possible to make a linear model to predict the age of mud crabs based on several physical attributes. Though the adjusted R-squared remained relatively low regardless of the changes made between different models, it did improve as different models were tested. In a real world context, this model does make sense as when crabs age, their size is expected to increase, and the sex of a crab also influences the size, so size and sex could be useful tools in predicting the age of crabs. However, since the growth of mud crabs is negligible after a certain point, it does make sense that these physical attributes alone may not be the best way to predict age. This sentiment is noted in the article "How Old is that Crab" from the Alaska Department of Fish and Game as it says that many fisheries do use size to estimate the age of crabs, but this method is unreliable as similar to humans, there is a large amount of variation in the size of crabs that are the same ages. One of the main limitations of this analysis is that only three models were spoken about in depth while many more were produced in order to determine the best model. For more of the models produced, see the separate Rmd file. Additionally, since the data came from a dataset online and was not collected manually, it is hard to tell how much human error these measurements have as the data collection process is largely unknown.

5 References and Appendix

Dataset: https://www.kaggle.com/datasets/sidhus/crab-age-prediction?resource=download

Articles:

https://medium.com/geekculture/predicting-age-of-a-crab-in-mud-crab-farming-using-machine-learning-1ae3bf030426
https://www.fish.wa.gov.au/Documents/recreational_fishing/fact_sheets/fact_sheet_mud_crab.pdf
https://www.adfg.alaska.gov/index.cfm?adfg=wildlifenews.view_article&articles_id=845

Tables and Graphs:

Summary Statistics Matrix:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Length	0.18750000	1.125000	1.362500	1.3113055	1.537500	2.03750
Diameter	0.13750000	0.875000	1.062500	1.0208933	1.200000	1.62500
Height	0.00000000	0.287500	0.362500	0.3493739	0.412500	2.82500
Shucked.Wght	0.02834950	5.343881	9.539607	10.2073420	14.273973	42.18406
Shell.Weight	0.04252425	3.713785	6.662133	6.7958441	9.355335	28.49125
Viscera.Wght	0.01417475	2.664853	4.861939	5.1365464	7.200773	21.54562
Age	1.00000000	8.000000	10.000000	9.9547906	11.000000	29.00000

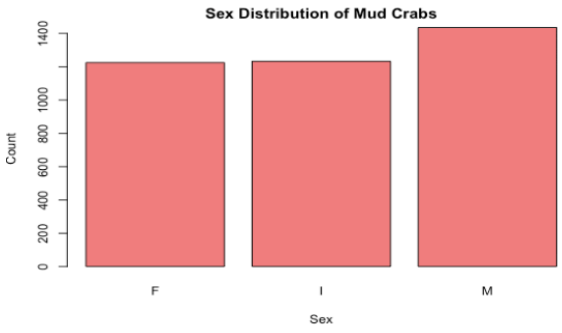
	F	I	M
Sex	1225	1233	1435

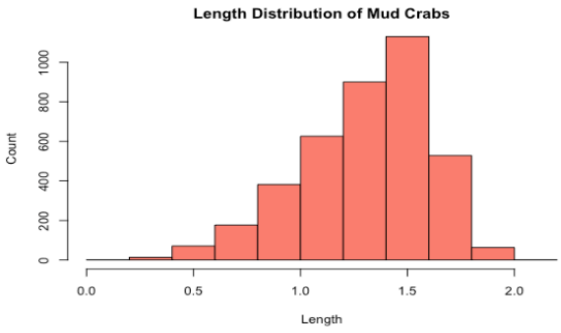
Correlation Matrix:

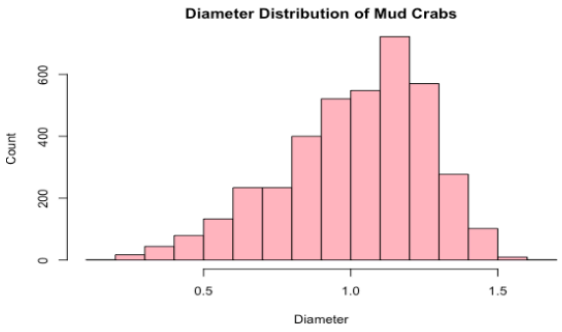
	Length	Diameter	Height	Shucked.Weight
Length	1.0000000	0.9866532	0.8230810	0.8981807
Diameter	0.9866532	1.0000000	0.8295315	0.8936257
Height	0.8230810	0.8295315	1.0000000	0.7709611
Shucked.Weight	0.8981807	0.8936257	0.7709611	1.0000000
Viscera.Weight	0.9032528	0.8998103	0.7932717	0.9312796
Shell.Weight	0.8977363	0.9055611	0.8122905	0.8824063
Age	0.5549733	0.5738443	0.5519564	0.4187598

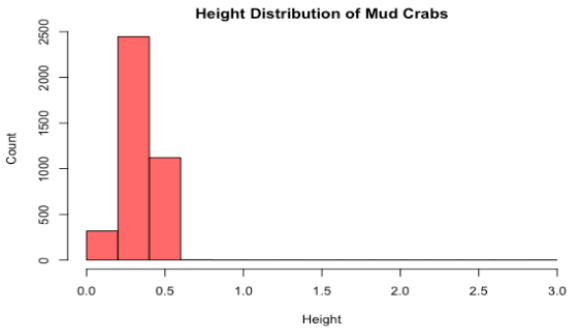
	Viscera.Weight	Shell.Weight	Age
Length	0.9032528	0.8977363	0.5549733
Diameter	0.8998103	0.9055611	0.5738443
Height	0.7932717	0.8122905	0.5519564
Shucked.Weight	0.9312796	0.8824063	0.4187598
Viscera.Weight	1.0000000	0.9061047	0.5013278
Shell.Weight	0.9061047	1.0000000	0.6251950
Age	0.5013278	0.6251950	1.0000000

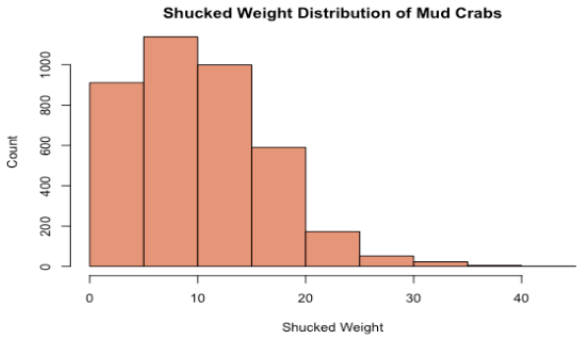
Graphs of the Distributions of Each Variable:

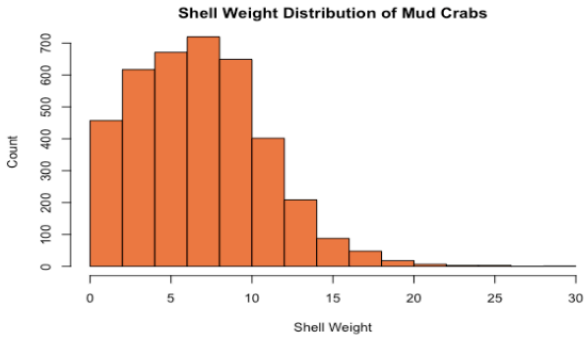


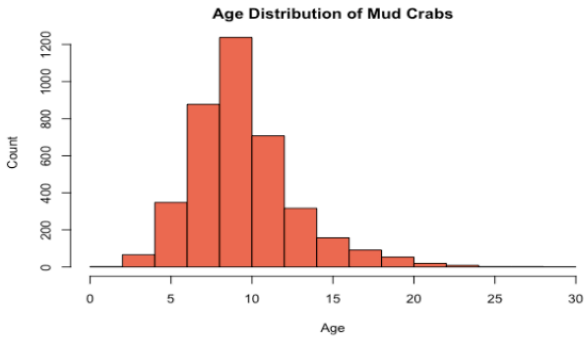












Summary Table of the Original Full Model

```
Call:
lm(formula = Age ~ Sex + Length + Diameter + Height + Shucked.Weight +
    Viscera.Weight + Shell.Weight, data = crabs)

Residuals:
     Min       1Q   Median       3Q      Max
-10.3232  -1.3337  -0.3309   0.8608  15.7347

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.80448    0.30765  12.366  < 2e-16 ***
SexI          -0.88258    0.10825  -8.153 4.73e-16 ***
SexM           0.05924    0.08768   0.676    0.499
Length        -0.57636    0.75906  -0.759    0.448
Diameter       5.12419    0.93583   5.476 4.64e-08 ***
Height         4.20740    0.63690   6.606 4.48e-11 ***
Shucked.Weight -0.40463    0.01705 -23.738  < 2e-16 ***
Viscera.Weight -0.02028    0.03776  -0.537    0.591
Shell.Weight    0.69115    0.02529  27.326  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.236 on 3884 degrees of freedom
Multiple R-squared:  0.519,     Adjusted R-squared:  0.518
F-statistic: 523.9 on 8 and 3884 DF,  p-value: < 2.2e-16
```

Summary Table of the Reduced Model Without Height Outliers

```
Call:
lm(formula = Age ~ Sex + Diameter + Height + Shucked.Weight +
    Shell.Weight, data = crabs2)

Residuals:
    Min      1Q  Median      3Q     Max
-8.0066 -1.3435 -0.3351  0.8435 16.0220

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.40145    0.29120  11.681  < 2e-16 ***
SexI          -0.83876    0.10676  -7.856 5.09e-15 ***
SexM           0.06127    0.08701   0.704    0.481
Diameter       3.32390    0.44119   7.534 6.09e-14 ***
Height         9.29230    0.95561   9.724  < 2e-16 ***
Shucked.Weight -0.40936    0.01377 -29.721  < 2e-16 ***
Shell.Weight    0.63903    0.02473  25.845  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.222 on 3884 degrees of freedom
Multiple R-squared:  0.5251,    Adjusted R-squared:  0.5244
F-statistic: 715.8 on 6 and 3884 DF,  p-value: < 2.2e-16
```

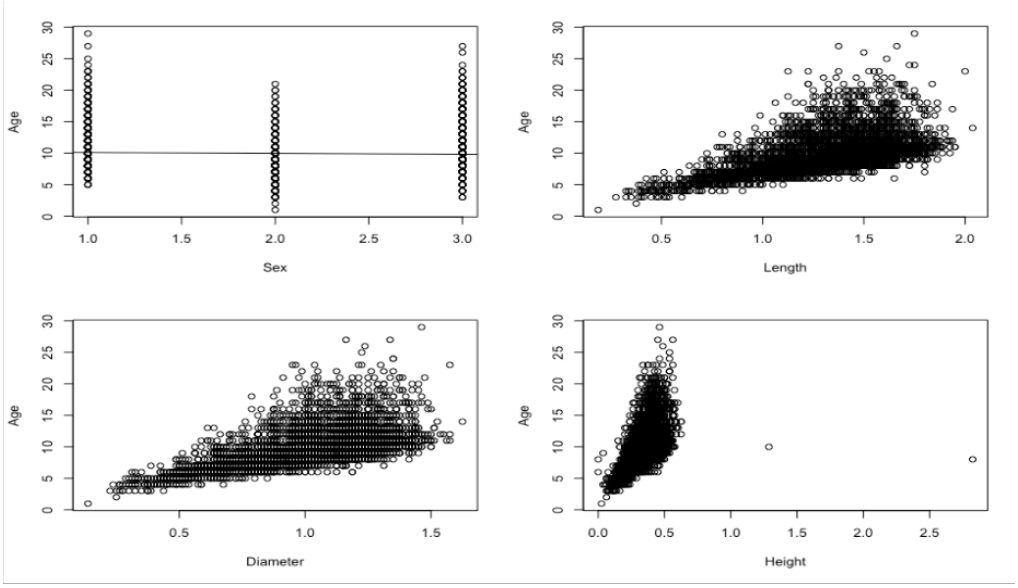
Summary Table of “Best” Predictive Model

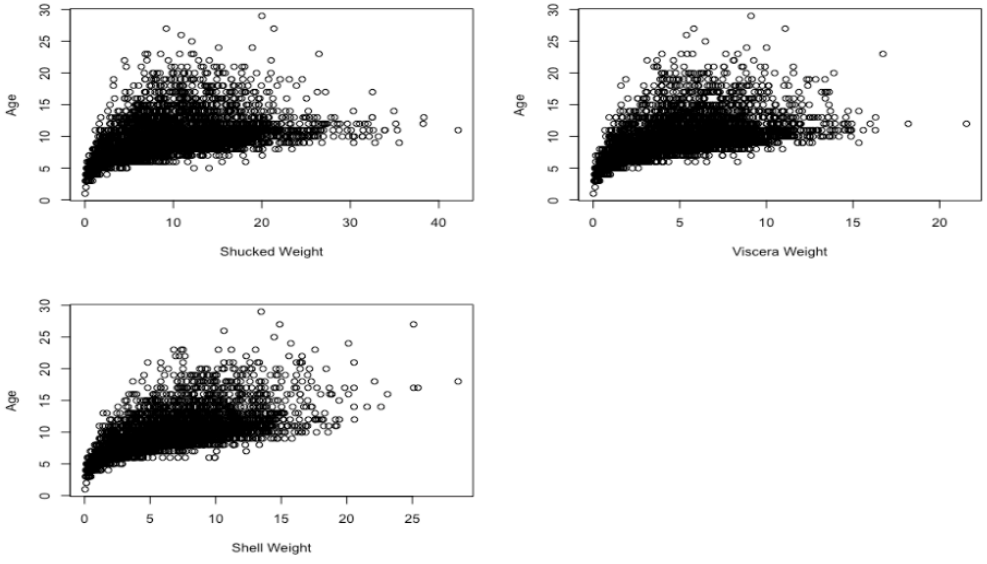
```
Call:
lm(formula = log(Age) ~ is.intermediate + Diameter + Height +
    Shucked.Weight + Shell.Weight, data = crabs2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.36155 -0.13361 -0.01763  0.11177  0.83433

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.336518   0.025809   51.79   <2e-16 ***
is.intermediate -0.096199   0.008547  -11.26   <2e-16 ***
Diameter         0.689296   0.040230   17.13   <2e-16 ***
Height           1.033489   0.087222   11.85   <2e-16 ***
Shucked.Weight  -0.038948   0.001254  -31.07   <2e-16 ***
Shell.Weight     0.040550   0.002256   17.98   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2028 on 3885 degrees of freedom
Multiple R-squared:  0.5967,    Adjusted R-squared:  0.5962
F-statistic:  1150 on 5 and 3885 DF,  p-value: < 2.2e-16
```

Partial F-Test Comparing Full and Reduced Model:

```
Analysis of Variance Table

Model 1: Age ~ Sex + Diameter + Height + Shucked.Weight + Shell.Weight
Model 2: Age ~ Sex + Length + Diameter + Height + Shucked.Weight +
Viscera.Weight +
    Shell.Weight
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1   3886 19426
2   3884 19421  2    4.8717 0.4871 0.6144
```

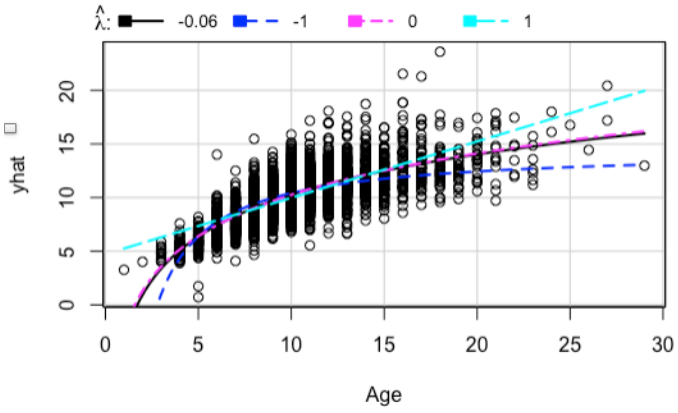
Graphs of X Variables vs Age:



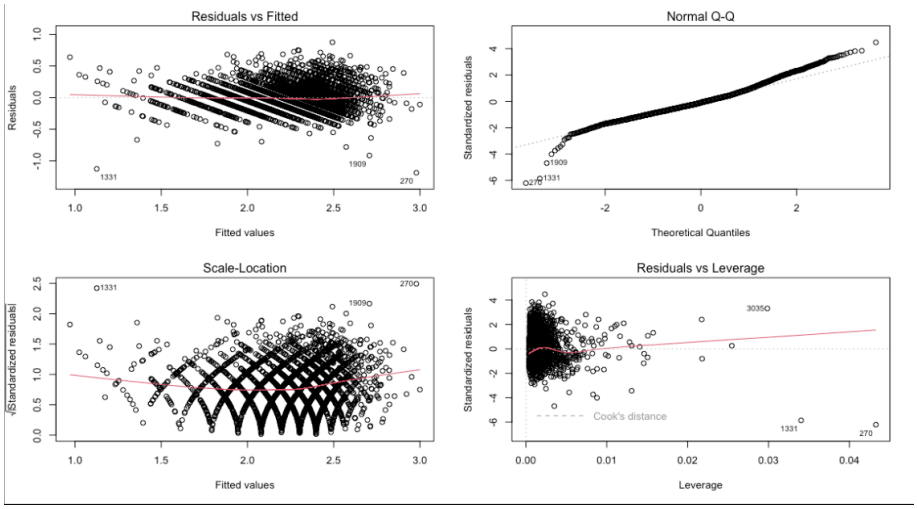


Results of powerTransform() and inverseResponsePlot():

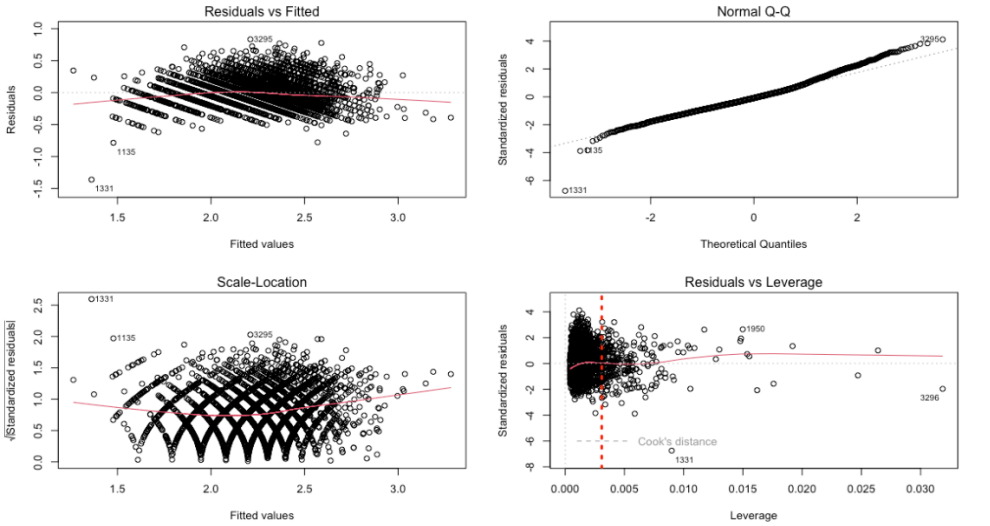
```
bcPower Transformations to Multinormality
   Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
Y1   -0.1229       -0.12      -0.1750      -0.0708
Y2    0.5955        0.60       0.5414       0.6496
Y3    0.1936        0.19       0.1745       0.2128
Y4    0.1617        0.16       0.1444       0.1791

Likelihood ratio test that transformation parameters are equal to 0
 (all log transformations)

Likelihood ratio test that no transformations are needed
```



Diagnostic Plot after Log Transformation on Shucked Weight and Shell Weight:



Diagnostic Plot after Sqrt Transformation on Diameter:



ANOVA Table of "Best" Predictive Model

```
Analysis of Variance Table

Response: log(Age)
                  Df  Sum Sq Mean Sq F value    Pr(>F)
is.intermediate    1  96.001  96.001 2333.66 < 2.2e-16 ***
Diameter           1  88.125  88.125 2142.20 < 2.2e-16 ***
Height             1  10.475  10.475  254.62 < 2.2e-16 ***
Shucked.Weight     1  28.567  28.567  694.43 < 2.2e-16 ***
Shell.Weight       1  13.292  13.292  323.10 < 2.2e-16 ***
Residuals       3885 159.819   0.041
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

VIF Values:

```
is.intermediate        Diameter          Height  Shucked.Weight
       1.495700        9.428518        6.634963        5.836997
   Shell.Weight
       7.478602
```

Backward AIC Output:

```
Start:  AIC=-12409.54
log(Age) ~ is.intermediate + Diameter + Height + Shucked.Weight +
    Shell.Weight

                 Df Sum of Sq    RSS    AIC
<none>                       159.82 -12410
- is.intermediate  1     5.211 165.03 -12287
- Height           1     5.776 165.59 -12273
- Diameter         1    12.077 171.90 -12128
- Shell.Weight     1    13.292 173.11 -12101
- Shucked.Weight   1    39.713 199.53 -11548
```

Backward BIC Output:

```
Start:  AIC=-12371.94
log(Age) ~ is.intermediate + Diameter + Height + Shucked.Weight +
    Shell.Weight

                Df Sum of Sq    RSS    AIC
<none>                       159.82 -12372
- is.intermediate  1     5.211 165.03 -12255
- Height           1     5.776 165.59 -12242
- Diameter         1    12.077 171.90 -12097
- Shell.Weight     1    13.292 173.11 -12069
- Shucked.Weight   1    39.713 199.53 -11517
```

Forward AIC Output:

```
Start:  AIC=-8886.23
log(Age) ~ 1

                  Df Sum of Sq    RSS      AIC
+ Height           1   183.463 212.82 -11303.2
+ Diameter         1   176.313 219.97 -11174.6
+ Shell.Weight     1   174.828 221.45 -11148.5
+ is.intermediate  1    96.001 300.28  -9963.6
+ Shucked.Weight   1    94.827 301.45  -9948.4
<none>                       396.28  -8886.2

Step:  AIC=-11303.23
log(Age) ~ Height

                  Df Sum of Sq    RSS    AIC
+ Shucked.Weight   1    8.4486 204.37 -11459
+ is.intermediate  1    7.4426 205.37 -11440
+ Shell.Weight     1    6.6726 206.14 -11425
+ Diameter         1    5.5707 207.25 -11404
<none>                       212.81 -11303

Step:  AIC=-11458.85
log(Age) ~ Height + Shucked.Weight

                  Df Sum of Sq    RSS    AIC
+ Shell.Weight     1   25.3320 179.03 -11972
+ Diameter         1   25.1270 179.24 -11967
+ is.intermediate  1    9.5569 194.81 -11643
<none>                       204.37 -11459

Step:  AIC=-11971.77
log(Age) ~ Height + Shucked.Weight + Shell.Weight

                  Df Sum of Sq    RSS    AIC
+ Diameter         1   14.0040 165.03 -12287
+ is.intermediate  1    7.1387 171.90 -12128
<none>                       179.03 -11972

Step:  AIC=-12286.69
```

Forward BIC Output:

```
Start:  AIC=-8879.96
log(Age) ~ 1

                  Df Sum of Sq    RSS      AIC
+ Height           1   183.463 212.82 -11290.7
+ Diameter         1   176.313 219.97 -11162.1
+ Shell.Weight     1   174.828 221.45 -11135.9
+ is.intermediate  1    96.001 300.28  -9951.1
+ Shucked.Weight   1    94.827 301.45  -9935.9
<none>                         396.28  -8880.0

Step:  AIC=-11290.7
log(Age) ~ Height

                  Df Sum of Sq    RSS    AIC
+ Shucked.Weight   1    8.4486 204.37 -11440
+ is.intermediate  1    7.4426 205.37 -11421
+ Shell.Weight     1    6.6726 206.14 -11406
+ Diameter         1    5.5707 207.25 -11386
<none>                         212.81 -11291

Step:  AIC=-11440.05
log(Age) ~ Height + Shucked.Weight

                  Df Sum of Sq    RSS    AIC
+ Shell.Weight     1   25.3320 179.03 -11947
+ Diameter         1   25.1270 179.24 -11942
+ is.intermediate  1    9.5569 194.81 -11618
<none>                         204.37 -11440

Step:  AIC=-11946.71
log(Age) ~ Height + Shucked.Weight + Shell.Weight

                  Df Sum of Sq    RSS    AIC
+ Diameter         1   14.0040 165.03 -12255
+ is.intermediate  1    7.1387 171.90 -12097
<none>                         179.03 -11947

Step:  AIC=-12255.36
```