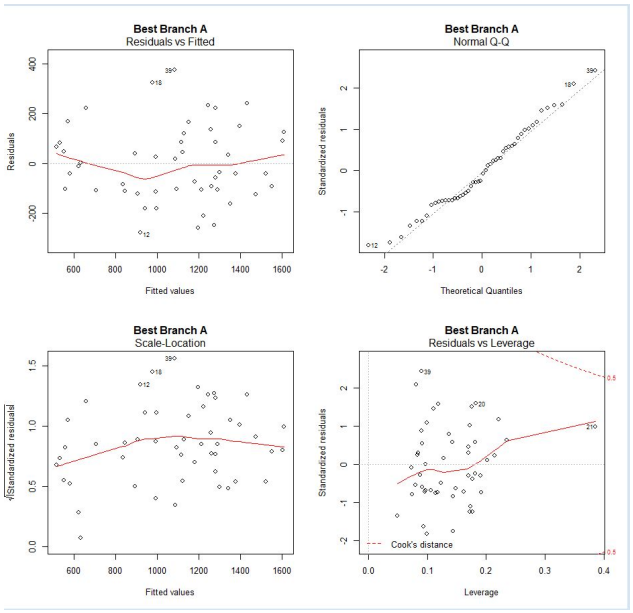
Eric Brown
ECE 3710
Class Project

I chose to pursue several different processes for creating a model for this project. Branch A was created by choosing the highest p value terms from the most complex interactions and working down to the two way interactions. The model produced by this procedure was of middling success. Learning from branch A, Branch B was produced by eliminated the highest p value terms regardless of complexity. Branch B's model resulted in more significant terms, but a less normal Q-Q plot. Branch C attempted to correct this by using a step algorithm once some of the more complicated interactions were removed. The resulting model was significantly more normal in it's Q-Q plot and appeared to have better residuals. Branch D attempted to produce a better model by using a process of favoring the removal of terms which either had no effect or improved the adjusted R squared value. It was quickly learned that a high R squared value does not necessarily make a good model. The process for branch E was the use a recursive tree instead of a normal tree. The process used by branch F was to favor the addition or subtraction of terms as long as it decreased the prediction error of the model. At the end of branch F a step function is used to optimize the model. From branch F, an outlier point was identified and so branch G used the same process as that from branch F, but excluding the outlier point. The model produced by branch G was the most accurate in its predicting power, but was the weakest in terms of the significance of its terms.

Branch A:

From branch A the best model produced had no apparent pattern in its residuals, was very normal in its Q-Q plot, no pattern to its scale-location graph, and finally no outliers, although point 21 is quite close to Cook's distance. The model was quite simple with only one interaction term, although two of the terms had quite high p values. The prediction power of the



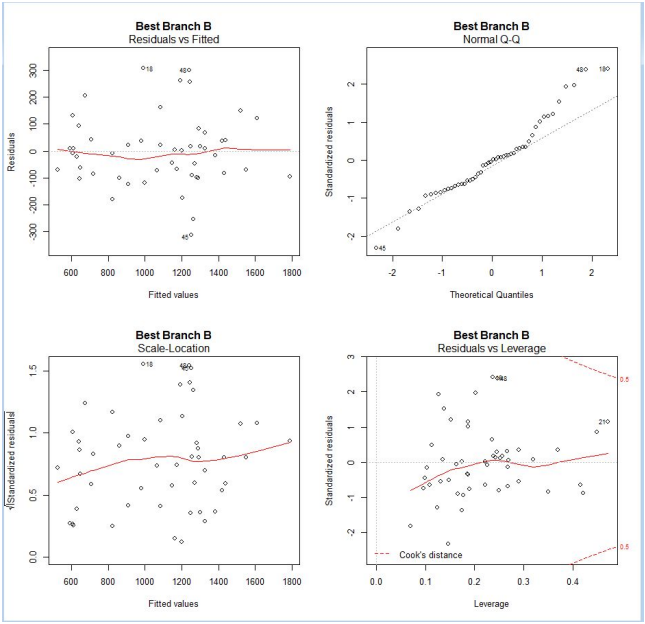
model was fairly good, with an average error of %14.46154.

```
Call:
lm(formula = M ~ M1 + M2 + M3 + M4 + W + M3:W)

Residuals:
    Min      1Q  Median      3Q     Max
-276.70 -105.65  -24.06   89.36  373.47

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2092.09277  453.79241   4.610 3.59e-05 ***
M1             0.63013    0.14117   4.464 5.74e-05 ***
M2            -0.04916    0.16972  -0.290 0.773500
M3            -1.37867    0.39646  -3.477 0.001171 **
M4            -0.10260    0.14398  -0.713 0.479955
W            -81.69907   18.94616  -4.312 9.27e-05 ***
M3:W           0.08101    0.02074   3.906 0.000327 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 160.8 on 43 degrees of freedom
Multiple R-squared:  0.8066,    Adjusted R-squared:  0.7796
F-statistic: 29.89 on 6 and 43 DF,  p-value: 7.99e-14
```

	fit	lwr	upr	TM
1	849.6973	454.5110	1244.884	1181
2	983.5427	607.1145	1359.971	1343
3	1287.4036	938.4734	1636.334	1073
4	1072.1956	703.9941	1440.397	1110
5	982.8330	598.3459	1367.320	1021
6	1095.1249	725.9067	1464.343	1472
7	1378.3615	992.0735	1764.650	1300
8	1324.4692	952.6094	1696.329	1297

Branch B:

Like branch A, the best model produced did not appear to have a pattern to its residuals, but the Q-Q plot is quite poorly matched at the higher levels. The scale location graph appears to have a slight increasing trend as well. Finally, from the residuals vs. leverage graph, there appears to be no outliers. The model itself has a higher number of significant terms, and those that are not significant are of a lower p value than the non-significant terms from branch A. The prediction power of the model is modest, at best, with an average error of %17.84363.




```
Call:
lm(formula = M ~ M1 + M2 + M3 + M4 + W + M1:M2 + M1:W + M2:W +
    M4:W + M1:M2:W)

Residuals:
     Min       1Q   Median       3Q      Max
-312.328  -85.154   -3.108   41.655  307.345

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.687e+02  1.496e+03  -0.246  0.80663
M1           3.606e+00  1.370e+00   2.631  0.01211 *
M2           3.406e+00  1.523e+00   2.237  0.03108 *
M3          -9.126e-02  1.599e-01  -0.571  0.57144
M4          -1.708e+00  5.103e-01  -3.346  0.00182 **
W            5.242e+01  5.906e+01   0.888  0.38014
M1:M2       -3.448e-03  1.203e-03  -2.865  0.00668 **
M1:W        -1.672e-01  6.209e-02  -2.693  0.01038 *
M2:W        -1.771e-01  6.829e-02  -2.593  0.01331 *
M4:W         8.844e-02  2.965e-02   2.983  0.00491 **
M1:M2:W      1.941e-04  5.839e-05   3.324  0.00194 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 146 on 39 degrees of freedom
Multiple R-squared:  0.8555,    Adjusted R-squared:  0.8184
F-statistic: 23.08 on 10 and 39 DF,  p-value: 2.24e-13
```

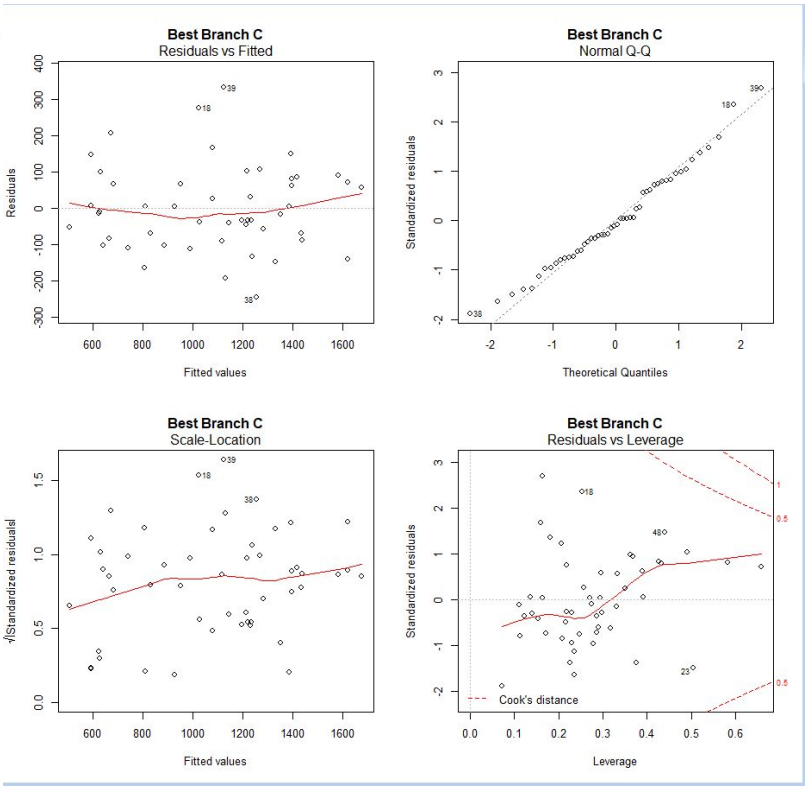
	fit	lwr	upr	TM
1	989.6914	627.0804	1352.302	1181
2	786.3576	367.8079	1204.907	1343
3	1034.8883	685.5401	1384.237	1073
4	1210.0914	838.6587	1581.524	1110
5	1141.8583	796.8836	1486.833	1021
6	937.9338	517.4008	1358.467	1472
7	1373.4300	973.0580	1773.802	1300
8	1053.6057	672.8526	1434.359	1297

Branch C:

The model produced by branch C, did not appear to have a pattern to its residuals, nor did the residuals appear to have a pattern in the Scale-Location graph. The Q-Q graph appeared to be quite normal and there did not appear to be any outliers from the Residuals vs Leverage graph. The model produced had 7 terms that were not significant at the p<=0.05 level. The prediction power of the model declined from that of branch A with an average error of

```
Call:
lm(formula = M ~ M1 + M2 + M3 + M4 + W + M1:M2 + M2:M3 + M1:M4 +
    M1:W + M2:W + M3:W + M4:W + M1:M2:W)

Residuals:
    Min      1Q  Median      3Q     Max
-244.72  -79.47  -12.65   68.73  331.50

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.492e+02  1.527e+03   0.163 0.871337
M1           6.069e+00  1.559e+00   3.892 0.000412 ***
M2           1.967e+00  1.554e+00   1.266 0.213694
M3          -3.005e+00  1.008e+00  -2.980 0.005133 **
M4          -3.214e-01  9.274e-01  -0.347 0.730903
W            4.249e+01  5.644e+01   0.753 0.456439
M1:M2       -4.217e-03  1.198e-03  -3.518 0.001196 **
M2:M3        1.450e-03  4.913e-04   2.952 0.005528 **
M1:M4       -7.699e-04  4.763e-04  -1.616 0.114741
M1:W        -2.396e-01  6.406e-02  -3.741 0.000637 ***
M2:W        -1.568e-01  6.570e-02  -2.387 0.022382 *
M3:W         6.668e-02  3.464e-02   1.925 0.062178 .
M4:W         6.069e-02  3.358e-02   1.808 0.079031 .
M1:M2:W      2.161e-04  5.564e-05   3.883 0.000423 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 134.7 on 36 degrees of freedom
Multiple R-squared:  0.8864,    Adjusted R-squared:  0.8454
F-statistic: 21.61 on 13 and 36 DF,  p-value: 3.629e-13
```

	fit	lwr	upr	TM
1	799.4591	433.3548	1165.563	1181
2	708.7179	309.6337	1107.802	1343
3	1080.1653	747.6458	1412.685	1073
4	1064.3386	689.5933	1439.084	1110
5	954.0235	597.7587	1310.288	1021
6	996.1641	595.6068	1396.721	1472
7	1562.2692	1159.4535	1965.085	1300
8	952.9857	543.5865	1362.385	1297

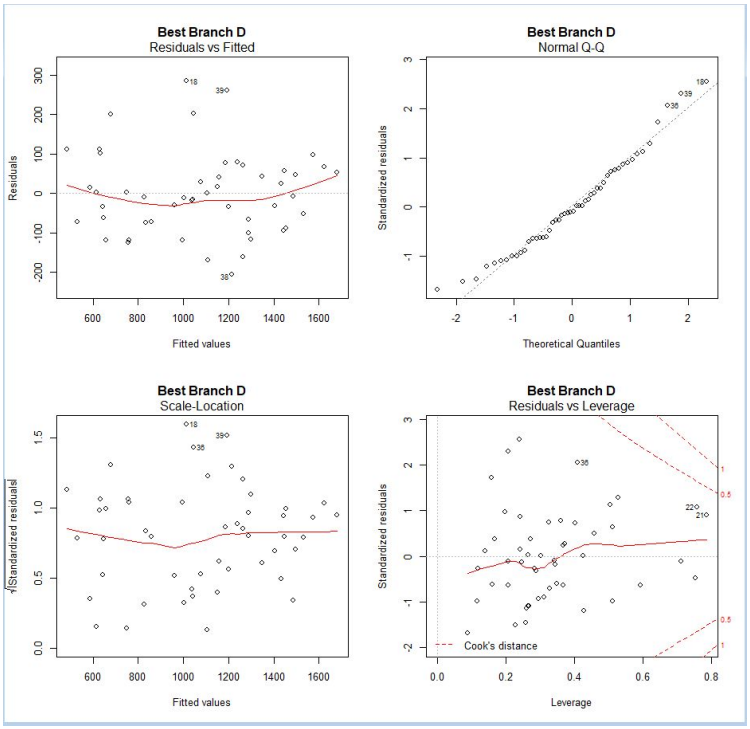


%21.23761.

Branch D:

The model produced by branch D appeared to have a slight curvature to its residuals and its residuals appeared to shrink slightly as the fitted values grew in magnitude. The Q-Q graph looked normal as did the Residuals vs Leverage graph. The model had a high number of significant terms and a high R squared value, but only had an average error from prediction of %18.21.

```
Call:
lm(formula = M ~ M1 + M2 + M3 + M4 + W + M2:M3 + M2:M4 + M3:M4 +
    M1:W + M1:M2:M3 + M1:M2:W + M2:M3:W + M1:M4:W + M3:M4:W +
    M1:M2:M3:M4 + M1:M2:M3:W)

Residuals:
     Min       1Q   Median       3Q      Max
-204.774  -72.682   -8.948   55.876  285.036

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.087e+03  2.046e+03  -1.020 0.315148
M1           4.539e+00  9.971e-01   4.552 6.86e-05 ***
M2          -1.056e+00  1.641e+00  -0.644 0.524264
M3          -8.576e-01  1.131e+00  -0.758 0.453712
M4           3.636e+00  1.248e+00   2.913 0.006379 **
W            6.906e+01  4.460e+01   1.548 0.131052
M2:M3        6.620e-03  1.503e-03   4.406 0.000105 ***
M2:M4       -2.096e-03  8.879e-04  -2.360 0.024326 *
M3:M4       -3.411e-03  9.802e-04  -3.480 0.001432 **
M1:W        -1.449e-01  4.550e-02  -3.185 0.003159 **
M1:M2:M3    -4.762e-06  1.149e-06  -4.143 0.000224 ***
M1:M2:W      9.339e-05  5.248e-05   1.779 0.084380 .
M2:M3:W     -2.253e-04  7.546e-05  -2.986 0.005296 **
M1:M4:W     -1.127e-04  4.363e-05  -2.584 0.014385 *
M3:M4:W      1.505e-04  4.325e-05   3.481 0.001429 **
M1:M2:M3:M4  1.038e-09  3.761e-10   2.761 0.009335 **
M1:M2:M3:W   1.502e-07  5.163e-08   2.910 0.006423 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 127.8 on 33 degrees of freedom
Multiple R-squared:  0.9063,    Adjusted R-squared:  0.8609
F-statistic: 19.95 on 16 and 33 DF,  p-value: 1.65e-12
```



	fit	lwr	upr	TM
1	883.8093	463.9424	1303.676	1181
2	642.6227	187.1263	1098.119	1343
3	1150.3211	809.6899	1490.952	1073
4	1230.7694	813.5044	1648.034	1110
5	989.0335	624.8828	1353.184	1021
6	1111.4399	709.5561	1513.324	1472
7	1481.5132	1091.4029	1871.624	1300
8	1184.2403	756.0505	1612.430	1297

Branch E:

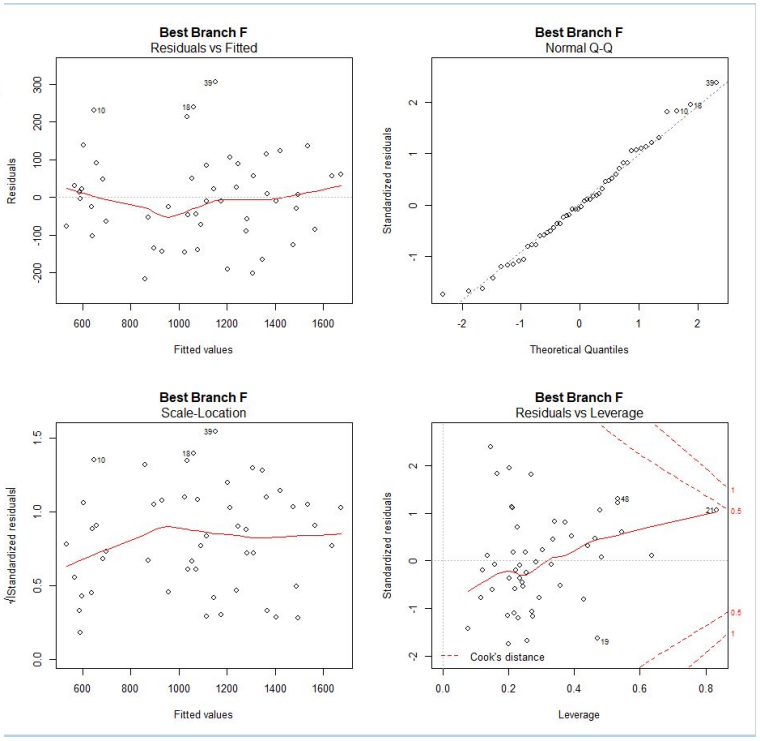
The model produced by branch E was simple but had a surprisingly low average error percentage. However, the tree model had a few quite low errors mixed with

some quite high ones, so it doesn't appear to be a reliable model for predicting data.

		TM
1	1095.083	1181
2	1095.083	1343
3	1095.083	1073
4	1253.167	1110
5	1253.167	1021
6	1095.083	1472
7	1486.200	1300
8	1253.167	1297

Branch F:

The model produced by branch F was the best by far. The residuals had an even spread and no apparent pattern. The Q-Q graph was extremely linear, and the Residuals vs Leverage graph had no outliers apart from point 21 which was very close to the 0.5 Cook's distance. The terms in the model were all significant apart from 5, but those 5 had fairly low p values with the highest being 0.27. The predictions produced by the model were accurate, and none of the actual values of moose were outside of the 0.95 confidence interval of the prediction.




```
Call:
lm(formula = M ~ M1 + M2 + M3 + M4 + W + M1:M3 + M2:M3 + M2:M4 +
    M1:W + M2:W + M2:M3:M4 + M1:M3:W + M2:M4:W + M3:M4:W)

Residuals:
     Min       1Q   Median       3Q      Max
-216.130  -76.090   -6.611   58.736  305.238

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.019e+03  2.652e+03  -1.516 0.138521
M1           4.506e+00  1.508e+00   2.989 0.005091 **
M2           5.519e+00  2.953e+00   1.869 0.070022 .
M3           2.177e+00  1.562e+00   1.394 0.172216
M4           3.752e+00  1.421e+00   2.641 0.012266 *
W            1.054e+02  6.466e+01   1.630 0.112001
M1:M3       -4.395e-03  1.242e-03  -3.538 0.001161 **
M2:M3       -1.581e-03  1.419e-03  -1.114 0.272750
M2:M4       -7.134e-03  2.370e-03  -3.010 0.004826 **
M1:W        -1.932e-01  5.708e-02  -3.385 0.001771 **
M2:W        -1.799e-01  8.281e-02  -2.172 0.036690 *
M2:M3:M4     2.706e-06  1.021e-06   2.649 0.012024 *
M1:M3:W      2.218e-04  5.591e-05   3.967 0.000343 ***
M2:M4:W      2.133e-04  7.918e-05   2.694 0.010763 *
M3:M4:W     -1.836e-04  7.731e-05  -2.374 0.023197 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 138.4 on 35 degrees of freedom
Multiple R-squared:  0.8835,    Adjusted R-squared:  0.8369
F-statistic: 18.96 on 14 and 35 DF,  p-value: 2.655e-12
```

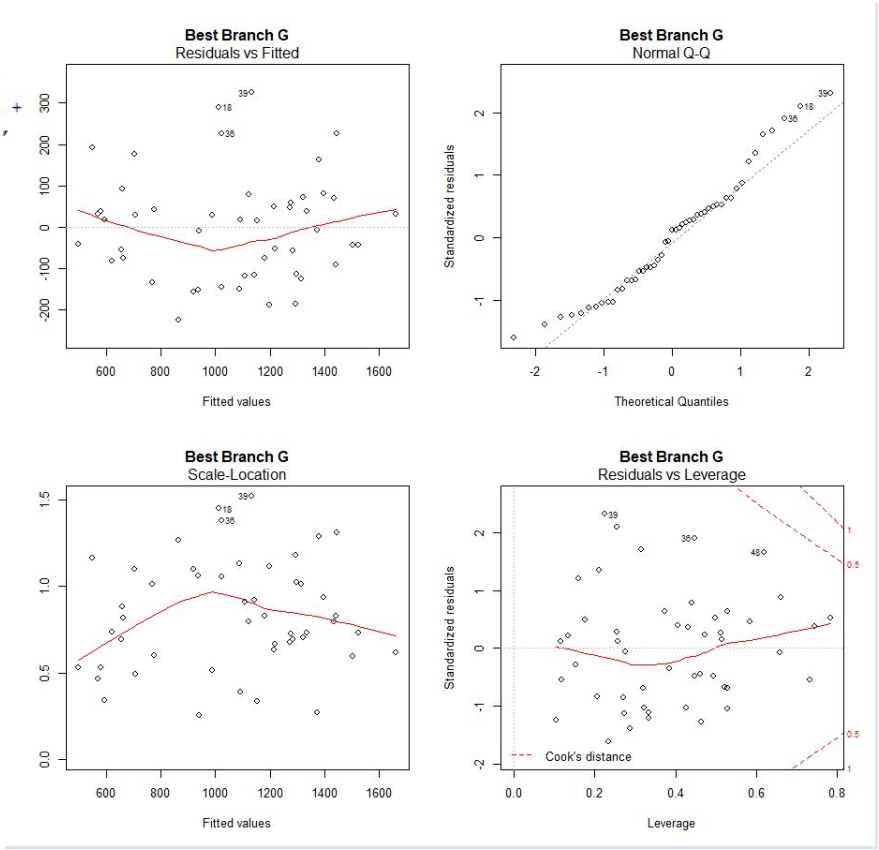
```
        fit      lwr      upr   TM
1  964.1936 526.0592 1402.328 1181
2  935.0814 454.6705 1415.492 1343
3 1088.4797 731.7547 1445.205 1073
4 1139.3299 754.1603 1524.500 1110
5 1057.0558 644.2883 1469.823 1021
6 1108.4992 674.6400 1542.358 1472
7 1282.8186 870.6948 1694.942 1300
8 1169.8549 716.0226 1623.687 1297
```

Branch G:

The model produced by branch G showed that the procedure of choosing the terms based on their prediction values does not necessarily produce a good model. The residuals had a clear parabolic shape and a curved shape in the Scale-Location graph. The Q-Q graph was not well fit, but did not have an apparent pattern. There did not appear to be any outliers from the Residuals vs Leverage graph. Few of the terms in the model were significant, and the model

had a comparatively low adjusted R squared value. This model, despite its flaws, was the best produced predictor with an average error of %9.9039. Also of note, this model had a much wider prediction interval than the other models due to its few significant terms.

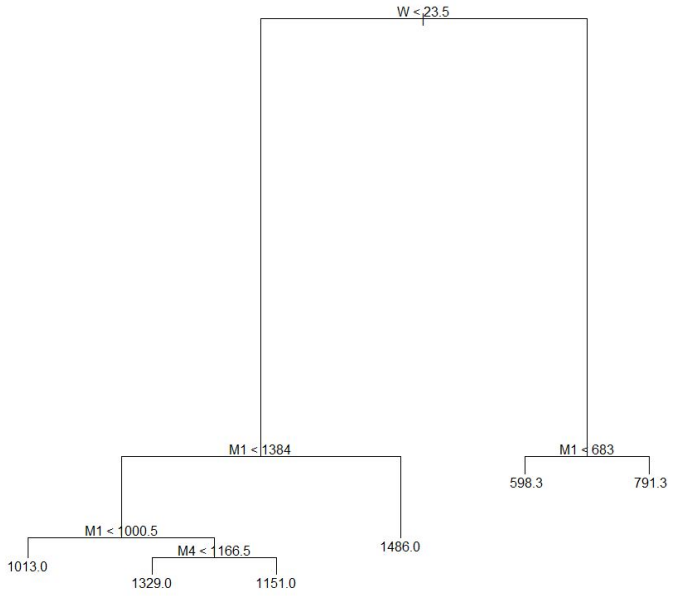
```
Call:
lm(formula = M ~ M3 + M4 + W + I(M3^2) + M1:M2 + M3:M2 + M4:M1 +
    M3:M4 + W:M1 + M3:W + M3:M1:M2 + M3:M4:M2 + M3:W:M2 + M4:W:M1 +
    M4:W:M2 + M3:M4:M1 + M3:W:M1 + M3:M4:M1:M2, data = mooseData1,
    subset = c(1:20, 22:nrow(mooseData1)))

Residuals:
    Min      1Q  Median      3Q     Max
-223.71  -90.23   15.67   59.22  324.67

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.519e+03  2.958e+03  -0.513   0.6114
M3           8.028e+00  5.220e+00   1.538   0.1345
M4           6.449e+00  3.216e+00   2.005   0.0540 .
W            2.064e+01  7.971e+01   0.259   0.7975
I(M3^2)     -5.556e-05  1.135e-03  -0.049   0.9613
M1:M2        2.762e-03  1.329e-03   2.079   0.0463 *
M3:M2       -7.990e-03  3.622e-03  -2.206   0.0352 *
M4:M1       -5.105e-03  2.623e-03  -1.946   0.0610 .
M3:M4       -9.021e-03  4.096e-03  -2.203   0.0354 *
W:M1        -6.482e-02  5.967e-02  -1.086   0.2860
M3:W        -1.702e-01  1.213e-01  -1.404   0.1707
M3:M1:M2     2.907e-07  1.416e-06   0.205   0.8387
M3:M4:M2     6.420e-06  2.940e-06   2.184   0.0369 *
M3:W:M2      3.494e-04  1.689e-04   2.069   0.0473 *
M4:W:M1      2.822e-04  1.393e-04   2.025   0.0518 .
M4:W:M2     -3.033e-04  1.577e-04  -1.924   0.0639 .
M3:M4:M1     3.778e-06  2.330e-06   1.621   0.1155
M3:W:M1     -1.426e-04  1.435e-04  -0.993   0.3284
M3:M4:M1:M2 -2.362e-09  1.405e-09  -1.682   0.1030
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 159.2 on 30 degrees of freedom
Multiple R-squared:  0.8569,    Adjusted R-squared:  0.771
F-statistic: 9.979 on 18 and 30 DF,  p-value: 3.36e-08
```



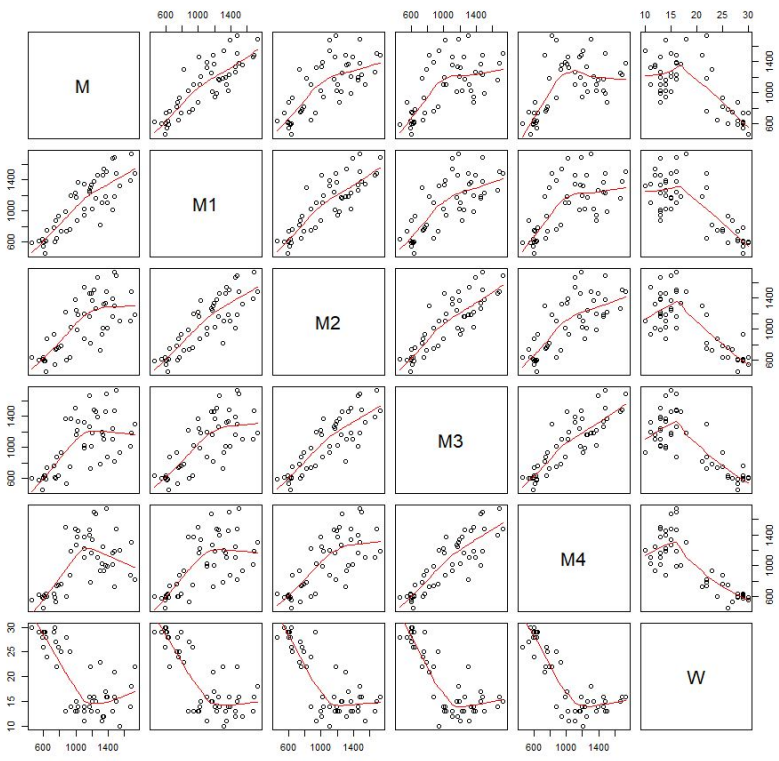
	fit	lwr	upr	TM
1	928.0964	441.7144	1414.478	1181
2	962.3696	421.2222	1503.517	1343
3	1188.5486	747.1041	1629.993	1073
4	1150.6660	632.3347	1668.997	1110
5	1105.1221	628.6357	1581.608	1021
6	1427.3133	794.8074	2059.819	1472
7	1293.3696	708.9221	1877.817	1300
8	1254.7475	710.4027	1799.092	1297

Additional Analysis:

As suggested by the pairs graph, the tree graph, and the recursive tree graph, there are important interactions between: M1 and Wolves, M2 and Wolves, M1, M4, and Wolves. From the pairs graph one can also see strong correlations between Moose and Moose of the previous


W < 23.5
M1 < 1384
M1 < 683
598.3
791.3
M1 < 1000.5
M4 < 1166.5
1013.0
1329.0
1151.0
1486.0

year.


M
M1
M2
M3
M4
W

Recursive Tree Model