Gasoline Mileage Data - Tree vs Linear Regression

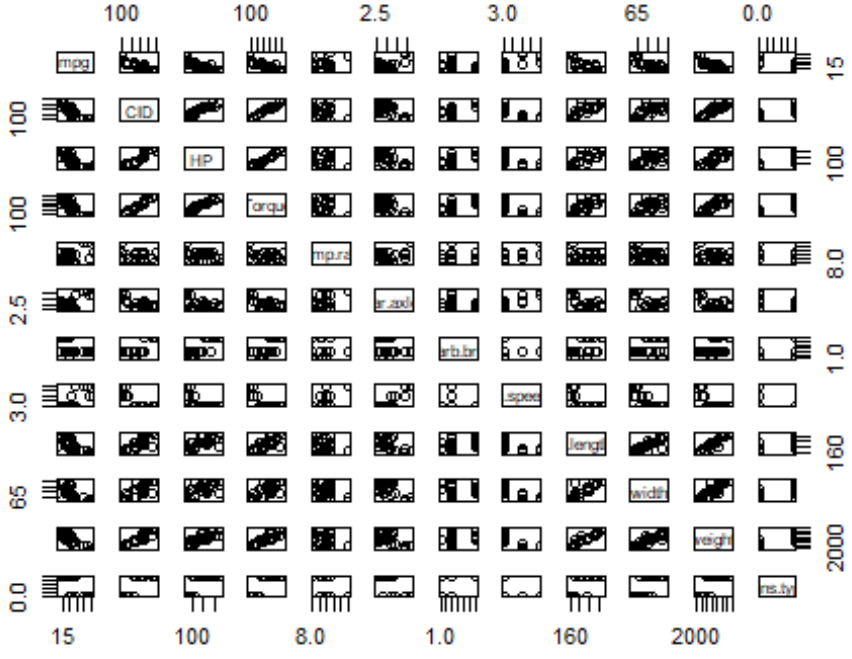
Dr. Phil

December 7, 2019

```
load ("~/Fall 2019/Math 327 Fa19/Montgomery data sets/Appendices/
Appendix_B_data.Rdata")

## Warning: namespace 'emmeans' is not available and has been replaced
## by .GlobalEnv when processing object '.Last.ref_grid'

names (dataB.3) = c("mpg", "CID", "HP", "Torque", "Comp.ratio",
"Rear.axle.R",
                    "Carb.brls", "trn.speeds", "o.length", "width",
"weight", "trans.type")
db3 = dataB.3 [complete.cases (dataB.3), ]
plot (db3)
```



First order model.

```
mpg1 = lm (mpg ~ HP + Comp.ratio + as.factor (Carb.brls) + CID +
trans.type +
                Torque + width + Rear.axle.R + as.factor (trn.speeds)
+ o.length +
                weight, data=db3)
summary (mpg1)
```

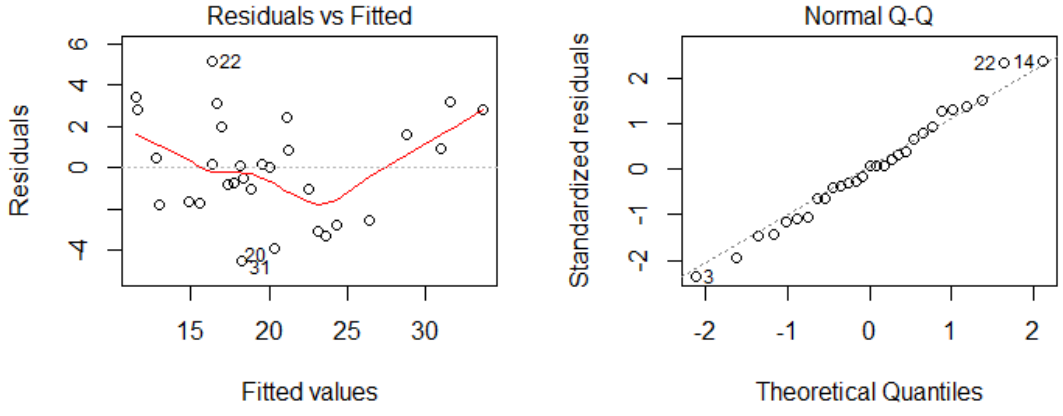
```
##
## Call:
## lm(formula = mpg ~ HP + Comp.ratio + as.factor(Carb.brls) + CID +
##     trans.type + Torque + width + Rear.axle.R +
as.factor(trn.speeds) +
##     o.length + weight, data = db3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4635 -1.6727  0.0636  1.8887  5.1371
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -2.061584  32.717000  -0.063    0.951
## HP                      -0.074636   0.088594  -0.842    0.412
## Comp.ratio               2.551209   3.272293   0.780    0.447
## as.factor(Carb.brls)2   -2.346722   3.491761  -0.672    0.511
## as.factor(Carb.brls)4   -2.224126   4.892532  -0.455    0.656
## CID                     -0.060596   0.057951  -1.046    0.311
## trans.type               4.614526   4.413927   1.045    0.311
## Torque                   0.108786   0.088985   1.223    0.239
## width                   -0.605035   0.379956  -1.592    0.131
## Rear.axle.R              6.287702   3.213166   1.957    0.068 .
## as.factor(trn.speeds)4   3.521393   6.087678   0.578    0.571
## as.factor(trn.speeds)5  -2.619626   7.042871  -0.372    0.715
## o.length                 0.242178   0.144674   1.674    0.114
## weight                  -0.005345   0.005954  -0.898    0.383
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 3.252 on 16 degrees of freedom
## Multiple R-squared:  0.8514, Adjusted R-squared:  0.7307
## F-statistic: 7.053 on 13 and 16 DF,  p-value: 0.0002176
```

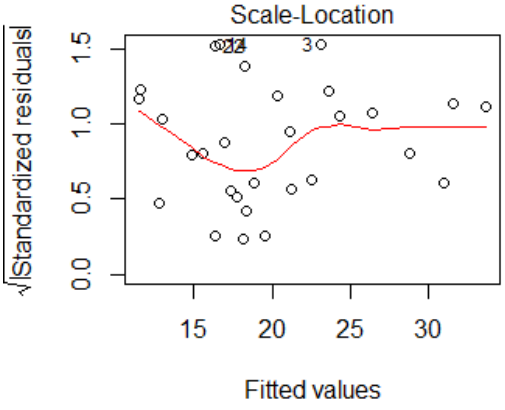
First order model residuals.

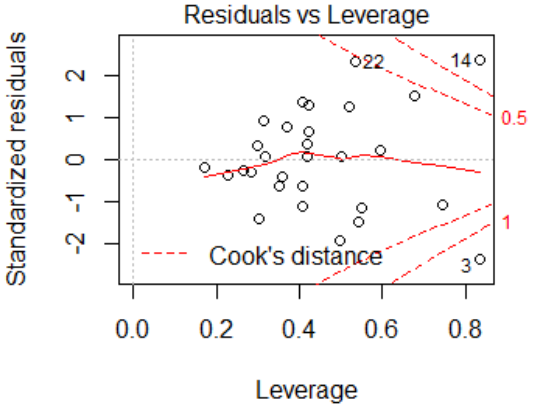
```
par (mfrow=c(1,2))
plot (mpg1)
```

```
## Warning: not plotting observations with leverage one:
##   5
```



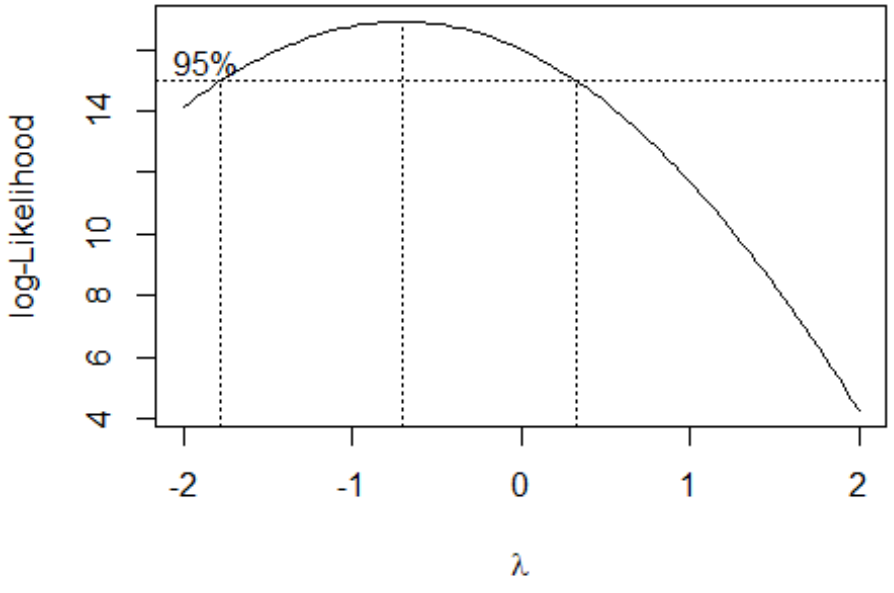

```
## Warning: not plotting observations with leverage one:
##   5
```





Box-Cox for first order model.

```
MASS::boxcox (mpg1)
```



Try power = -1.

```
mpg1.bc = lm (1/mpg ~ HP + Comp.ratio + as.factor (Carb.brls) + CID +
trans.type +
                Torque + width + Rear.axle.R + as.factor (trn.speeds)
+ o.length +
                weight, data=db3)
summary (mpg1.bc)
```

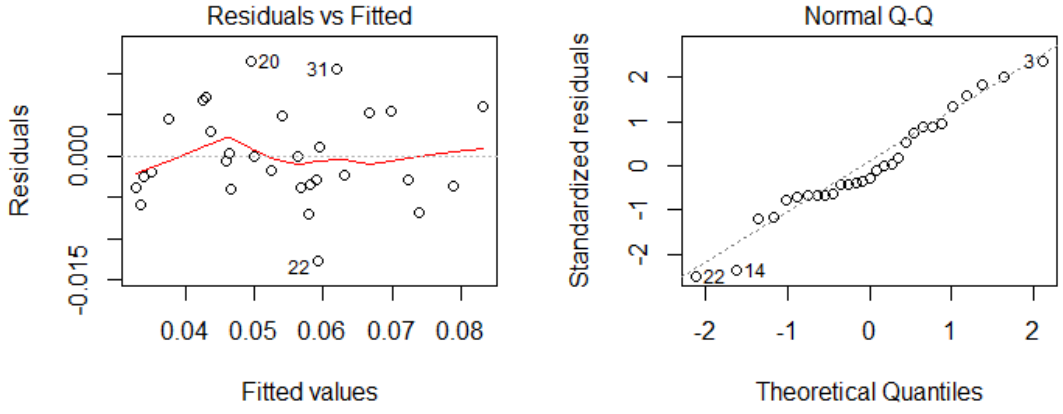
```
##
## Call:
## lm(formula = 1/mpg ~ HP + Comp.ratio + as.factor(Carb.brls) +
##     CID + trans.type + Torque + width + Rear.axle.R +
as.factor(trn.speeds) +
##     o.length + weight, data = db3)
##
## Residuals:
##       Min         1Q     Median         3Q        Max
## -0.012645 -0.003546 -0.001196  0.004730  0.011431
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              8.674e-02  7.491e-02   1.158   0.2639
## HP                       3.053e-04  2.028e-04   1.505   0.1518
## Comp.ratio              -1.548e-03  7.492e-03  -0.207   0.8389
## as.factor(Carb.brls)2    9.223e-04  7.994e-03   0.115   0.9096
## as.factor(Carb.brls)4   -4.576e-03  1.120e-02  -0.408   0.6883
## CID                      1.051e-04  1.327e-04   0.792   0.4398
## trans.type              -1.090e-02  1.011e-02  -1.078   0.2970
## Torque                  -2.121e-04  2.037e-04  -1.041   0.3133
## width                    5.930e-05  8.699e-04   0.068   0.9465
## Rear.axle.R              3.001e-03  7.357e-03   0.408   0.6887
## as.factor(trn.speeds)4 -1.277e-02  1.394e-02  -0.916   0.3733
## as.factor(trn.speeds)5 -1.493e-02  1.612e-02  -0.926   0.3682
## o.length                -6.235e-04  3.312e-04  -1.883   0.0781 .
## weight                   2.021e-05  1.363e-05   1.482   0.1577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.007446 on 16 degrees of freedom
## Multiple R-squared:  0.8594, Adjusted R-squared:  0.7452
## F-statistic: 7.524 on 13 and 16 DF,  p-value: 0.0001461
```

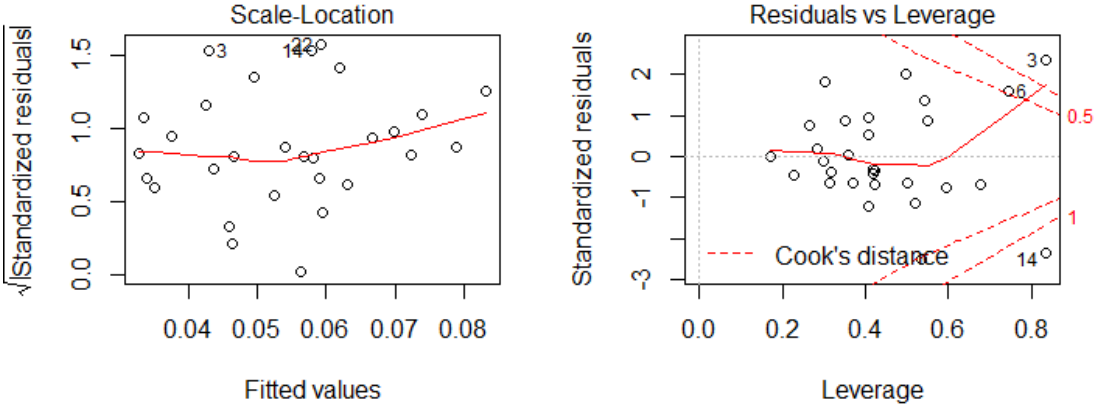
Box-Cox model residuals.

```
par (mfrow=c(1,2))
plot (mpg1.bc)
```

```
## Warning: not plotting observations with leverage one:
##   5
```




```
## Warning: not plotting observations with leverage one:
##   5
```



Stepwise regression on Box-Cox model.

```
mpg1bc.stp = step (mpg1.bc, direction='both', k=log(30))

## Start:  AIC=-265.24
## 1/mpg ~ HP + Comp.ratio + as.factor(Carb.brls) + CID + trans.type +
##     Torque + width + Rear.axle.R + as.factor(trn.speeds) + o.length
+
##     weight
##
##                          Df  Sum of Sq        RSS     AIC
## - as.factor(Carb.brls)    2 4.7434e-05 0.00093459 -270.48
## - as.factor(trn.speeds)   2 5.1280e-05 0.00093844 -270.36
## - width                   1 2.5800e-07 0.00088742 -268.64
## - Comp.ratio              1 2.3680e-06 0.00088953 -268.56
## - Rear.axle.R             1 9.2290e-06 0.00089639 -268.33
## - CID                     1 3.4803e-05 0.00092196 -267.49
## - Torque                  1 6.0103e-05 0.00094726 -266.68
## - trans.type              1 6.4450e-05 0.00095161 -266.54
## <none>                                 0.00088716 -265.24
## - weight                  1 1.2181e-04 0.00100897 -264.79
## - HP                      1 1.2559e-04 0.00101275 -264.67
## - o.length                1 1.9650e-04 0.00108366 -262.64
##
```

```
## Step:  AIC=-270.48
## 1/mpg ~ HP + Comp.ratio + CID + trans.type + Torque + width +
##     Rear.axle.R + as.factor(trn.speeds) + o.length + weight
##
##                          Df  Sum of Sq        RSS     AIC
## - as.factor(trn.speeds)  2 4.9675e-05 0.00098427 -275.73
## - width                   1 3.0000e-09 0.00093460 -273.88
## - Rear.axle.R             1 6.4390e-06 0.00094103 -273.68
## - Comp.ratio              1 1.6199e-05 0.00095079 -273.37
## - CID                     1 1.6486e-05 0.00095108 -273.36
## - Torque                  1 3.7472e-05 0.00097207 -272.70
## - trans.type              1 6.1846e-05 0.00099644 -271.96
## - HP                      1 8.7624e-05 0.00102222 -271.20
## <none>                                 0.00093459 -270.48
## - weight                  1 1.5314e-04 0.00108773 -269.33
## - o.length                1 2.1915e-04 0.00115374 -267.56
## + as.factor(Carb.brls)   2 4.7434e-05 0.00088716 -265.24
##
## Step:  AIC=-275.73
## 1/mpg ~ HP + Comp.ratio + CID + trans.type + Torque + width +
##     Rear.axle.R + o.length + weight
##
##                          Df  Sum of Sq        RSS     AIC
## - Rear.axle.R             1 6.4700e-07 0.00098492 -279.11
## - width                   1 1.7610e-06 0.00098603 -279.08
## - trans.type              1 2.2091e-05 0.00100636 -278.47
## - Comp.ratio              1 2.6416e-05 0.00101069 -278.34
## - CID                     1 4.4404e-05 0.00102867 -277.81
## - Torque                  1 4.7160e-05 0.00103143 -277.73
## - HP                      1 6.7154e-05 0.00105142 -277.15
## <none>                                 0.00098427 -275.73
## - weight                  1 1.4249e-04 0.00112676 -275.08
## - o.length                1 1.9748e-04 0.00118175 -273.65
## + as.factor(trn.speeds)  2 4.9675e-05 0.00093459 -270.48
## + as.factor(Carb.brls)   2 4.5829e-05 0.00093844 -270.36
##
## Step:  AIC=-279.11
```

```
## 1/mpg ~ HP + Comp.ratio + CID + trans.type + Torque + width +
##     o.length + weight
##
##                          Df  Sum of Sq        RSS     AIC
## - width                   1 1.4270e-06 0.00098634 -282.47
## - trans.type              1 2.1537e-05 0.00100645 -281.87
## - Comp.ratio              1 3.4959e-05 0.00101988 -281.47
## - CID                     1 4.4675e-05 0.00102959 -281.18
## - Torque                  1 5.5296e-05 0.00104021 -280.88
## - HP                      1 8.5999e-05 0.00107092 -280.00
## <none>                                 0.00098492 -279.11
## - weight                  1 1.5836e-04 0.00114328 -278.04
## - o.length                1 2.2035e-04 0.00120526 -276.46
## + Rear.axle.R             1 6.4700e-07 0.00098427 -275.73
## + as.factor(Carb.brls)    2 4.6308e-05 0.00093861 -273.76
## + as.factor(trn.speeds)   2 4.3882e-05 0.00094103 -273.68
##
## Step:  AIC=-282.47
## 1/mpg ~ HP + Comp.ratio + CID + trans.type + Torque + o.length +
##     weight
##
##                          Df  Sum of Sq        RSS     AIC
## - trans.type              1 2.1917e-05 0.00100826 -285.21
## - Comp.ratio              1 3.7848e-05 0.00102419 -284.74
## - CID                     1 4.3286e-05 0.00102963 -284.58
## - Torque                  1 5.3901e-05 0.00104024 -284.28
## - HP                      1 8.5583e-05 0.00107193 -283.38
## <none>                                 0.00098634 -282.47
## - o.length                1 2.2308e-04 0.00120942 -279.76
## - weight                  1 2.3430e-04 0.00122064 -279.48
## + width                   1 1.4270e-06 0.00098492 -279.11
## + Rear.axle.R             1 3.1200e-07 0.00098603 -279.08
## + as.factor(Carb.brls)    2 4.7502e-05 0.00093884 -277.15
## + as.factor(trn.speeds)   2 4.4165e-05 0.00094218 -277.04
##
## Step:  AIC=-285.21
## 1/mpg ~ HP + Comp.ratio + CID + Torque + o.length + weight
```

```
##
##                          Df  Sum of Sq        RSS     AIC
## - Comp.ratio              1 2.3818e-05 0.00103208 -287.91
## - CID                     1 2.8138e-05 0.00103640 -287.79
## - Torque                  1 6.2994e-05 0.00107125 -286.80
## <none>                                 0.00100826 -285.21
## - HP                      1 1.3314e-04 0.00114140 -284.89
## + trans.type              1 2.1917e-05 0.00098634 -282.47
## - o.length                1 2.5124e-04 0.00125950 -281.94
## + width                   1 1.8080e-06 0.00100645 -281.87
## + Rear.axle.R             1 3.1300e-07 0.00100795 -281.82
## - weight                  1 2.9586e-04 0.00130412 -280.89
## + as.factor(Carb.brls)    2 4.0867e-05 0.00096739 -279.65
## + as.factor(trn.speeds)   2 5.8810e-06 0.00100238 -278.59
##
## Step:  AIC=-287.91
## 1/mpg ~ HP + CID + Torque + o.length + weight
##
##                          Df  Sum of Sq        RSS     AIC
## - CID                     1 0.00003049 0.00106257 -290.44
## - Torque                  1 0.00006617 0.00109825 -289.45
## <none>                                 0.00103208 -287.91
## - HP                      1 0.00012789 0.00115997 -287.81
## + Comp.ratio              1 0.00002382 0.00100826 -285.21
## + trans.type              1 0.00000789 0.00102419 -284.74
## + width                   1 0.00000412 0.00102796 -284.63
## + Rear.axle.R             1 0.00000261 0.00102947 -284.59
## - o.length                1 0.00027713 0.00130920 -284.18
## + as.factor(Carb.brls)    2 0.00006194 0.00097014 -282.97
## - weight                  1 0.00033322 0.00136530 -282.92
## + as.factor(trn.speeds)   2 0.00002296 0.00100912 -281.79
##
## Step:  AIC=-290.44
## 1/mpg ~ HP + Torque + o.length + weight
##
##                          Df  Sum of Sq       RSS     AIC
## - Torque                  1 0.00003987 0.0011024 -292.74
```

```
## - HP                      1 0.00010166 0.0011642 -291.10
## <none>                                 0.0010626 -290.44
## + CID                     1 0.00003049 0.0010321 -287.91
## + Comp.ratio              1 0.00002617 0.0010364 -287.79
## + Rear.axle.R             1 0.00000597 0.0010566 -287.21
## + trans.type              1 0.00000063 0.0010619 -287.06
## + width                   1 0.00000056 0.0010620 -287.06
## + as.factor(trn.speeds)   2 0.00004137 0.0010212 -284.83
## + as.factor(Carb.brls)    2 0.00002541 0.0010372 -284.37
## - o.length                1 0.00045945 0.0015220 -283.06
## - weight                  1 0.00062754 0.0016901 -279.92
##
## Step:  AIC=-292.74
## 1/mpg ~ HP + o.length + weight
##
##                          Df  Sum of Sq       RSS     AIC
## - HP                      1 0.00007131 0.0011738 -294.26
## <none>                                 0.0011024 -292.74
## + Torque                  1 0.00003987 0.0010626 -290.44
## + Comp.ratio              1 0.00002656 0.0010759 -290.07
## + trans.type              1 0.00001672 0.0010857 -289.80
## + Rear.axle.R             1 0.00000838 0.0010941 -289.57
## + CID                     1 0.00000419 0.0010983 -289.45
## + width                   1 0.00000191 0.0011005 -289.39
## - o.length                1 0.00042422 0.0015267 -286.37
## + as.factor(Carb.brls)    2 0.00001373 0.0010887 -286.31
## + as.factor(trn.speeds)   2 0.00000404 0.0010984 -286.05
## - weight                  1 0.00067025 0.0017727 -281.89
##
## Step:  AIC=-294.26
## 1/mpg ~ o.length + weight
##
##                          Df  Sum of Sq       RSS     AIC
## <none>                                 0.0011738 -294.26
## + HP                      1 0.00007131 0.0011024 -292.74
## + Comp.ratio              1 0.00001677 0.0011570 -291.29
## + CID                     1 0.00001363 0.0011601 -291.21
```

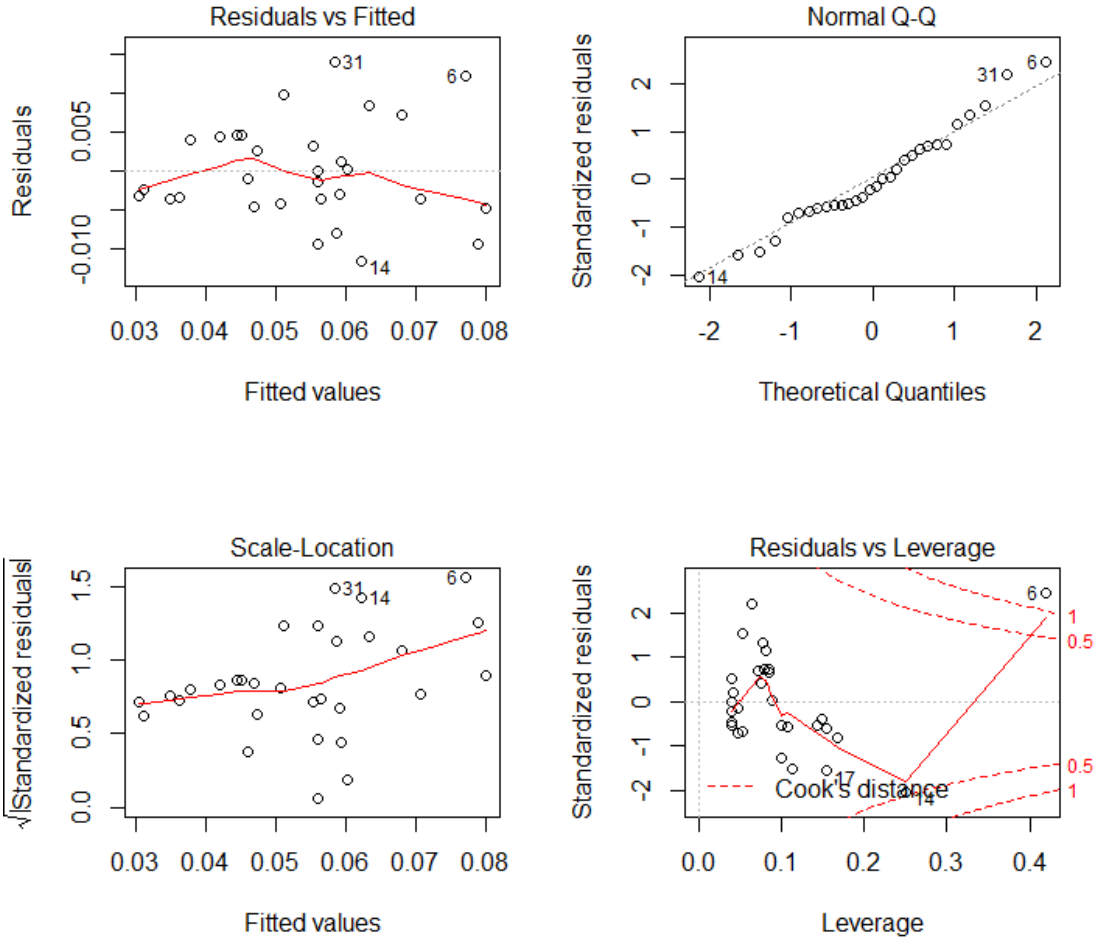
```
## + trans.type              1 0.00001193 0.0011618 -291.16
## + Torque                  1 0.00000951 0.0011642 -291.10
## + width                   1 0.00000471 0.0011690 -290.98
## + Rear.axle.R             1 0.00000369 0.0011701 -290.95
## + as.factor(Carb.brls)    2 0.00001421 0.0011595 -287.82
## + as.factor(trn.speeds)   2 0.00000600 0.0011677 -287.61
## - o.length                1 0.00061882 0.0017926 -284.96
## - weight                  1 0.00190156 0.0030753 -268.76
```

```
# Note:  BIC gives the same answer
summary (mpg1bc.stp)
```

```
##
## Call:
## lm(formula = 1/mpg ~ o.length + weight, data = db3)
##
## Residuals:
##       Min         1Q     Median         3Q        Max
## -0.011517 -0.003597 -0.001139  0.004320  0.013982
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.150e-02  2.318e-02   3.947 0.000509 ***
## o.length    -7.442e-04  1.973e-04  -3.773 0.000805 ***
## weight       2.916e-05  4.409e-06   6.614 4.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006593 on 27 degrees of freedom
## Multiple R-squared:  0.814,  Adjusted R-squared:  0.8002
## F-statistic: 59.09 on 2 and 27 DF,  p-value: 1.374e-10
```

Residuals for stepwise model.

```
par (mfrow=c(1,2))
plot (mpg1bc.stp)
```



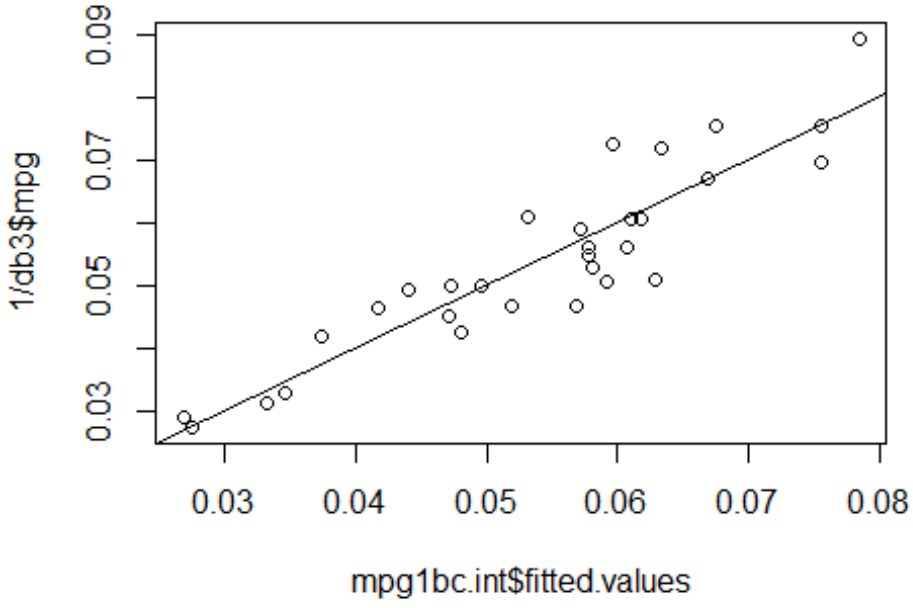
Add interaction between o.length and weight.

```
db3$o.length.c = db3$o.length - mean (db3$o.length)
db3$weight.c = db3$weight - mean (db3$weight)
mpg1bc.int = lm (1/mpg ~ o.length.c * weight.c, data=db3)
summary (mpg1bc.int)

##
## Call:
## lm(formula = 1/mpg ~ o.length.c * weight.c, data = db3)
##
```

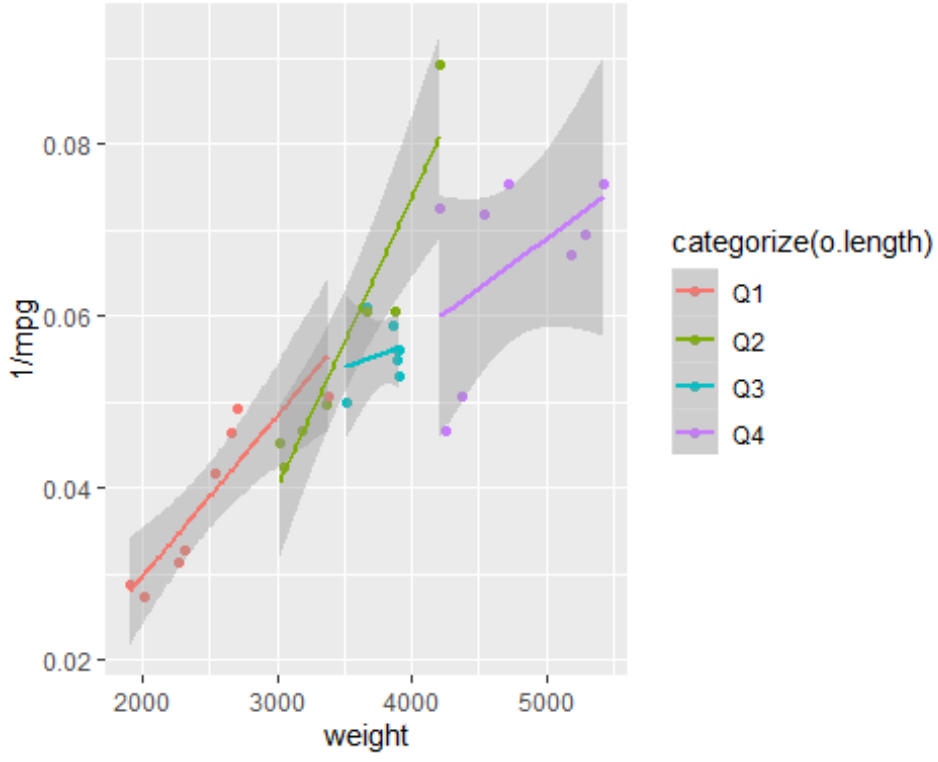
```
## Residuals:
##        Min          1Q      Median          3Q         Max 
## -0.0121158 -0.0041547 -0.0003018  0.0039427  0.0129777 
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)    
## (Intercept)          5.585e-02  1.549e-03  36.067  < 2e-16 ***
## o.length.c          -7.016e-04  1.919e-04  -3.656  0.00114 ** 
## weight.c             2.827e-05  4.285e-06   6.599 5.35e-07 ***
## o.length.c:weight.c -9.692e-08  5.603e-08  -1.730  0.09552 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.006363 on 26 degrees of freedom
## Multiple R-squared:  0.8332, Adjusted R-squared:  0.814 
## F-statistic: 43.29 on 3 and 26 DF,  p-value: 2.974e-10
```

```
plot (1/db3$mpg ~ mpg1bc.int$fitted.values)
abline (0, 1)
```



Overall length by weight interaction is probably not real.

```
categorize = function (x) {
  quartiles = summary (x) [c(2, 3, 5)]
  result = rep ("Q1", length (x))
  result [(quartiles[1] < x) & (x <= quartiles [2])] = "Q2"
  result [(quartiles[2] < x) & (x <= quartiles [3])] = "Q3"
  result [quartiles[3] < x] = "Q4"
  return (result)
}
ggplot2::qplot (x=weight, y=1/mpg, col=categorize (o.length),
data=db3) +
  ggplot2::geom_smooth (method="lm")
```



Tree regression.

```
library (rpart)
tree1 = rpart (mpg ~ ., maxdepth=5, data=db3)
print (tree1$cptable)
```

```
##           CP nsplit rel error    xerror      xstd
## 1 0.61720652      0 1.0000000 1.0398090 0.3197878
## 2 0.07174392      1 0.3827935 0.5368460 0.1107627
## 3 0.01000000      2 0.3110496 0.6150171 0.1195161
```

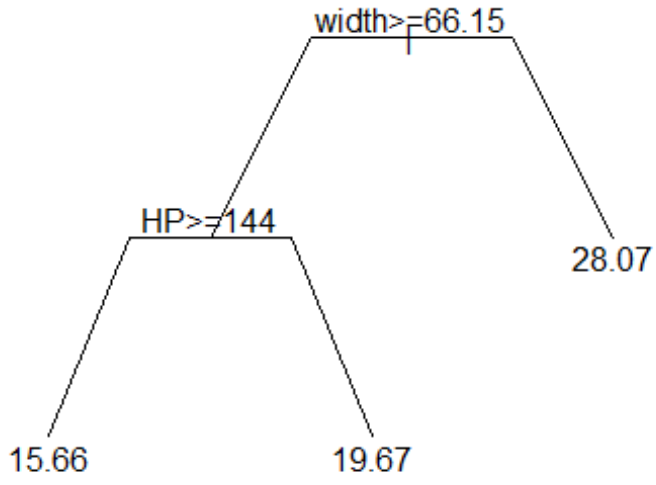
Find the tree with the smallest xerror:

```
opt1 = which.min (tree1$cptable [,"xerror"])
opt1
```

```
## 2
## 2
```

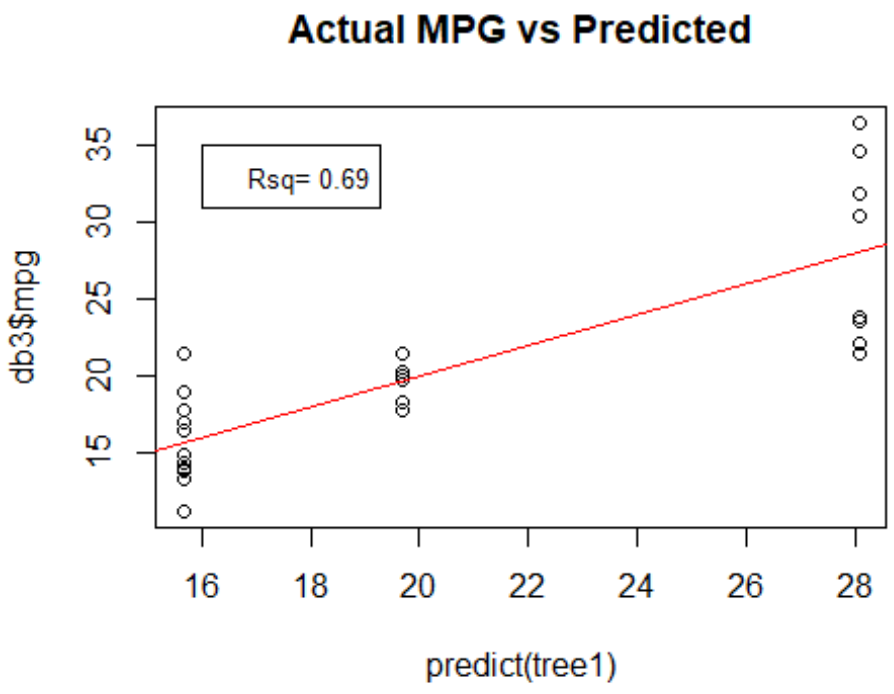
Plot the tree:

```
par (mfrow=c(1,1))
plot(tree1, uniform = TRUE, margin = 0.1, branch = 0.5,
     compress = TRUE)
text(tree1)
```



This tree says that car width is the single most important predictor. Cars have the highest MPG (mean = 28.07 mpg) when width is less than 66.15 inches. When car width is greater than 66.15 inches, horsepower matters. In that case, cars with horsepower greater than or equal 144 have mean MPG of 15.66 mpg, and cars with horsepower less than 144 have mean MPG of 19.67 mpg.

```
plot (predict(tree1), db3$mpg, main="Actual MPG vs Predicted")
abline (0, 1, col='red')
rsq1 = cor (predict(tree1), db3$mpg)^2
legend (16, 35, c(paste("Rsq=", round (rsq1, 2))), cex=0.8)
```

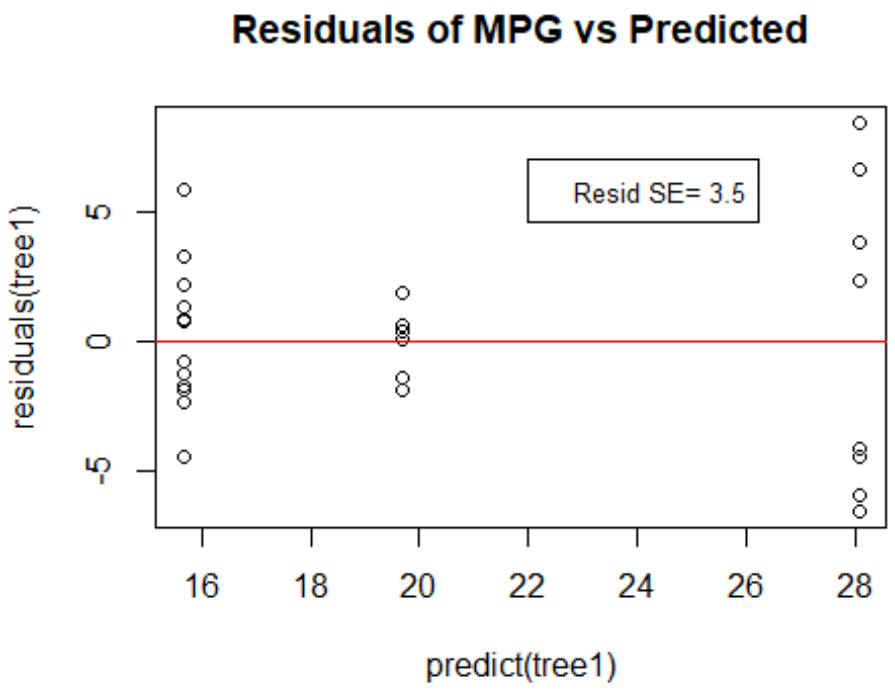



```
rsq1
```

```
## [1] 0.6889504
```

Residual plot

```
plot (predict(tree1), residuals(tree1), main="Residuals of MPG vs
Predicted")
abline (0, 0, col='red')
resid.se = sd (residuals (tree1))
legend (22, 7, c(paste ("Resid SE=", round (resid.se, 2))), cex=0.8)
```



NEW:

What if we fit a tree with only CID and HP as predictors? Will we get the tree model that's in the textbook?

```
tree2 = rpart (mpg ~ CID + HP, data=db3, method='anova')
print (tree2)
```

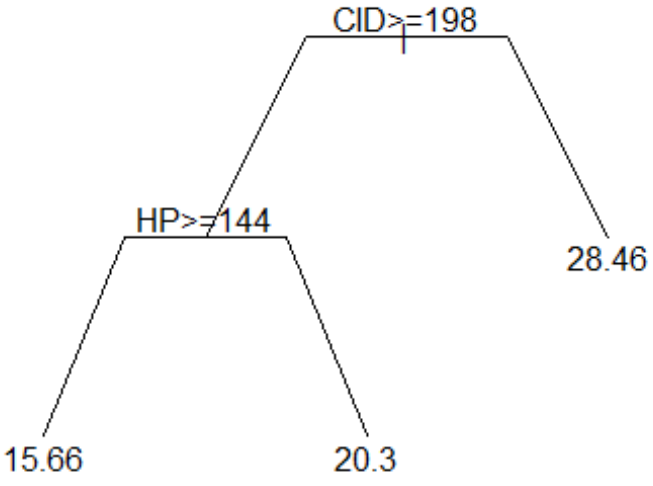
```
## n= 30
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
## 1) root 30 1139.10500 20.03833
##   2) CID>=198 23  236.97590 17.47609
##     4) HP>=144 14   92.79384 15.66214 *
##     5) HP< 144 9   26.45916 20.29778 *
##   3) CID< 198 7  254.99710 28.45714 *
```

```
print (tree2$cptable)
```

```
##           CP nsplit rel error    xerror      xstd
## 1 0.5681056      0 1.0000000 1.0649195 0.3297613
## 2 0.1033469      1 0.4318944 0.5313967 0.1227478
## 3 0.0100000      2 0.3285475 0.5446442 0.1219327
```

Draw the tree.

```
par (mfrow=c(1,1))
plot(tree2, uniform = TRUE, margin = 0.1, branch = 0.5,
     compress = TRUE)
text(tree2)
```



This tree is not the same as in the textbook.

```
cor (predict(tree2), db3$mpg)^2
```

```
## [1] 0.6714525
```

```
sd (residuals (tree2))
```

```
## [1] 3.592378
```

The correlation is lower, and the residual standard error is higher for this tree compared to the first one.

Test some example splits:

```
testsplit = function (y, x, cutoff) {
  t.test (y [x < cutoff], y [x >= cutoff], var.equal=T)
}
```

```
testsplit (db3$mpg, db3$CID, 115.25)
```

```
##
##  Two Sample t-test
##
## data:  y[x < cutoff] and y[x >= cutoff]
## t = 8.4995, df = 28, p-value = 3.06e-09
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  11.67979 19.09714
## sample estimates:
## mean of x mean of y
##  33.37500  17.98654
```

```
testsplit (db3$mpg, db3$CID, 198)
```

```
##
##  Two Sample t-test
##
## data:  y[x < cutoff] and y[x >= cutoff]
## t = 6.0688, df = 28, p-value = 1.521e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   7.274627 14.687484
## sample estimates:
```

```
## mean of x mean of y
##  28.45714  17.47609
```

Try the function, tree, from the package, tree.

```
tree3 = tree::tree (mpg ~ HP + Comp.ratio + as.factor (Carb.brls) +
CID +
                      trans.type + Torque + width + Rear.axle.R +
                      as.factor (trn.speeds) + o.length + weight,
data=db3)
print (tree3)
```

```
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 30 1139.000 20.04
##    2) CID < 136.8 5   94.370 31.48 *
##    3) CID > 136.8 25  259.300 17.75
##      6) HP < 144 11   27.770 20.41
##       12) weight < 3272.5 5    5.578 21.79 *
##       13) weight > 3272.5 6    4.770 19.26 *
##      7) HP > 144 14   92.790 15.66
##       14) HP < 185 9   48.430 16.86 *
##       15) HP > 185 5    8.140 13.50 *
```

This is a bigger tree!

What about that first split, CID < 136.8?

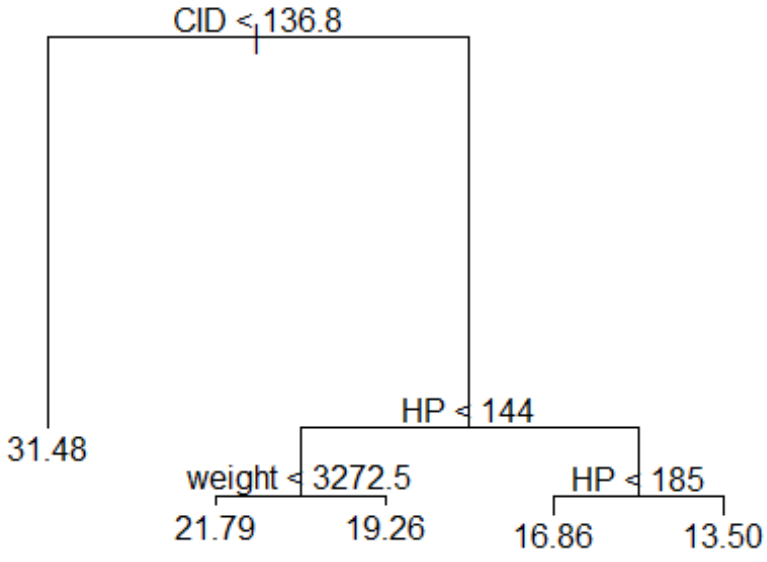
```
testsplit (db3$mpg, db3$CID, 136.8)
```

```
##
##  Two Sample t-test
##
## data:  y[x < cutoff] and y[x >= cutoff]
## t = 7.8862, df = 28, p-value = 1.37e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  10.16368 17.29632
## sample estimates:
```

```
## mean of x mean of y
##      31.48      17.75
```

Plot this tree.

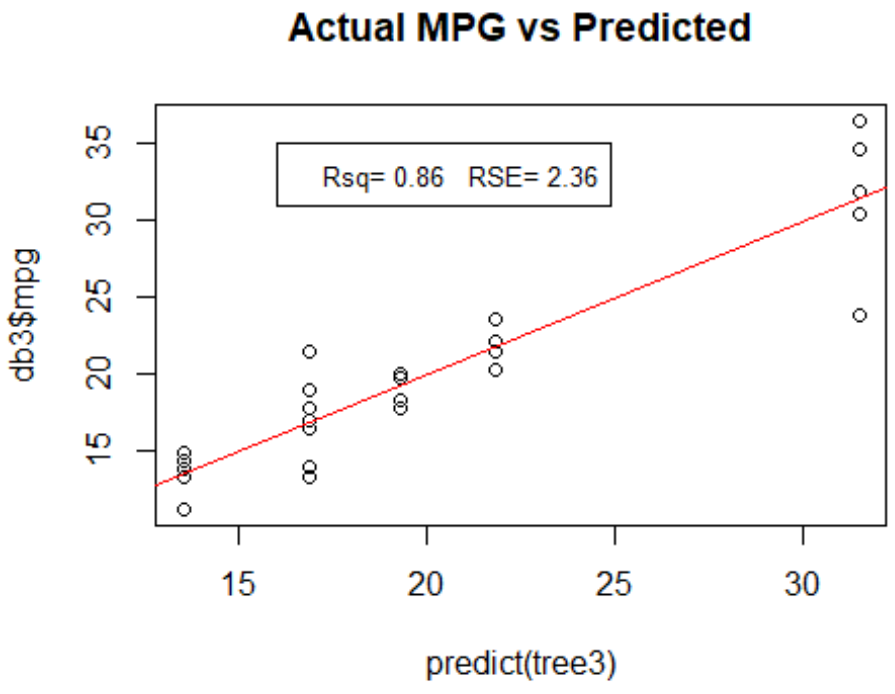
```
par (mfrow=c(1,1))
plot(tree3) #, uniform = TRUE, margin = 0.1, branch = 0.5, compress =
TRUE)
text(tree3)
```



Plot observed vs predicted.

```
plot (predict(tree3), db3$mpg, main="Actual MPG vs Predicted")
abline (0, 1, col='red')
rsq3 = cor (predict(tree3), db3$mpg)^2
rse3 = sd (residuals (tree3))
legend (16, 35, c(paste("Rsq=", round (rsq3, 2), "  RSE=", round
```

```
(rse3, 2))),
                    cex=0.8)
```




```
rsq3
```

```
## [1] 0.8584116
```

The tree function does not have some of the features that are in the rpart function, such as the cptable, and certain plotting options.

Regression trees do not provide prediction intervals, but one could use bootstrapping to calculate them.