Cut (Week 1)

Katie Dunning (umu5hh)

2025-03-01

Cut (Week 1 Analysis)

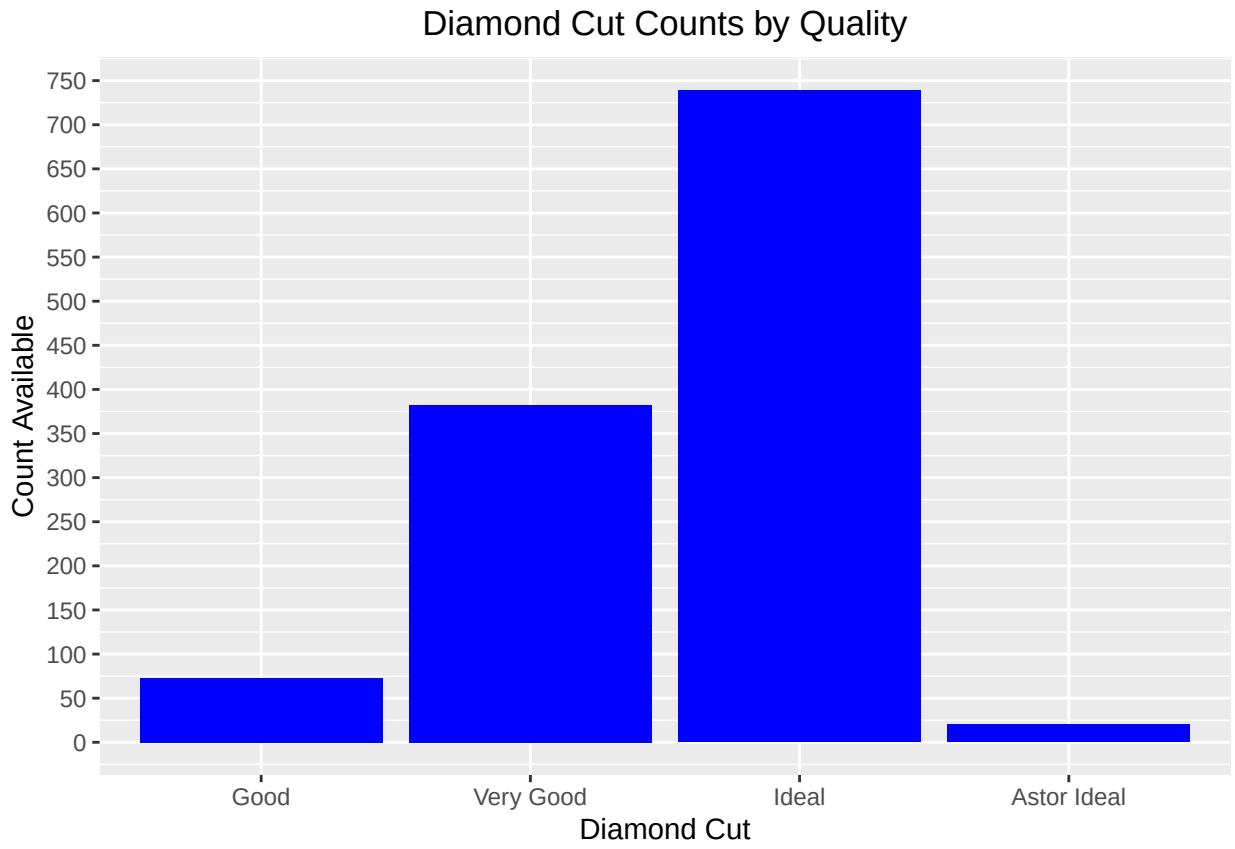
Description of Cuts

Diamond cuts refer to the proportion, dimensions, and faceting of a diamond. The cuts of a diamond are descriptive of the quality of the gem's faceting, proportion, and polish, as well as how symmetrical the piece is. Higher grade cuts are more symmetrical and have better light performance. Diamonds in the "good" category are considered to be in the top 25% of diamond cut quality, while "very good" is in the top 15%, and "ideal" is in the top 3%. Astor by Blue Nile is touted to "reflect the most light possible" and are grade/certified by a number of third parties.

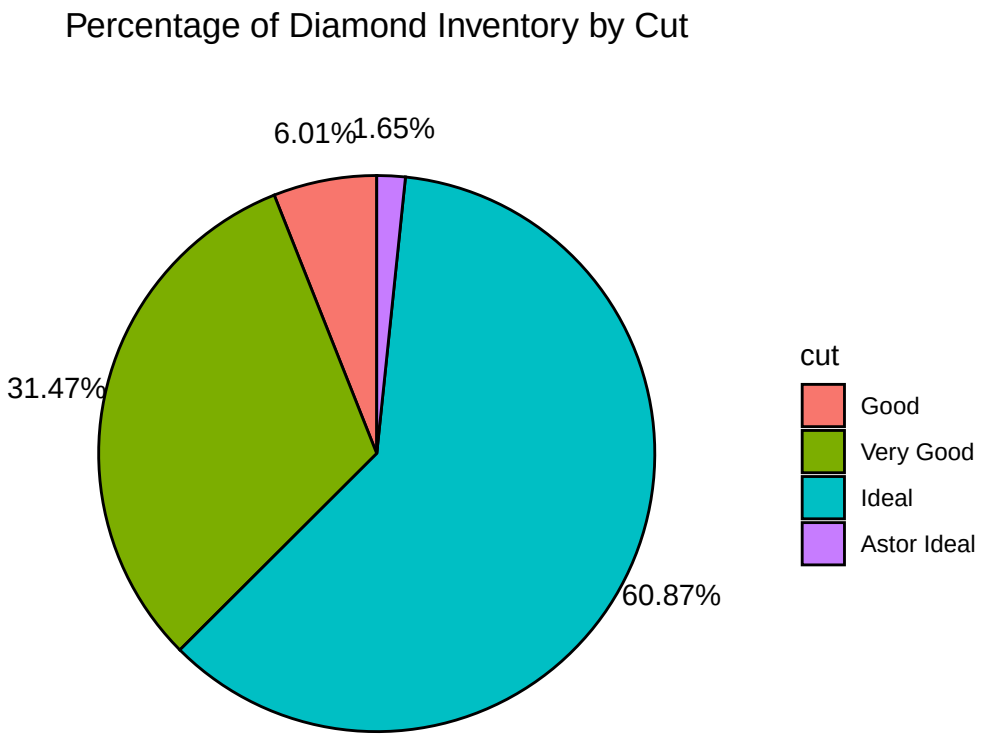
Citation: Blue Nile. (n.d.). Diamond cut: Grading scale and buying tips. https://www.bluenile.com/education/diamonds/cut?srsltid=AfmBOop9PWytZgjMIGYvLwuojS7LFcIV_5Pwh_pNHS44fwFoimTfBnbC

Available Inventory of Cuts

```
##
##      Good   Very Good       Ideal Astor Ideal
##        73         382         739          20
```

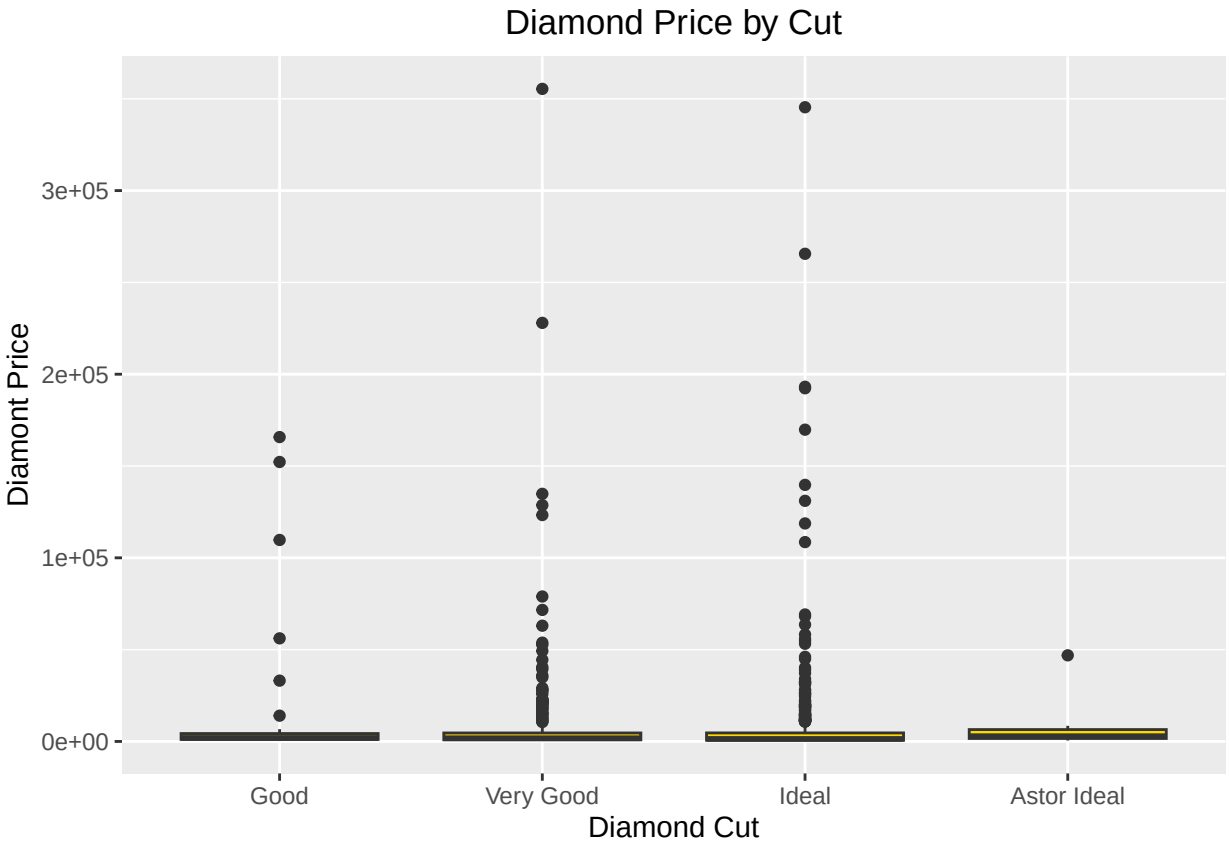



```
##           cut percent
## 1        Good    6.01
## 2   Very Good   31.47
## 3       Ideal   60.87
## 4 Astor Ideal    1.65
```

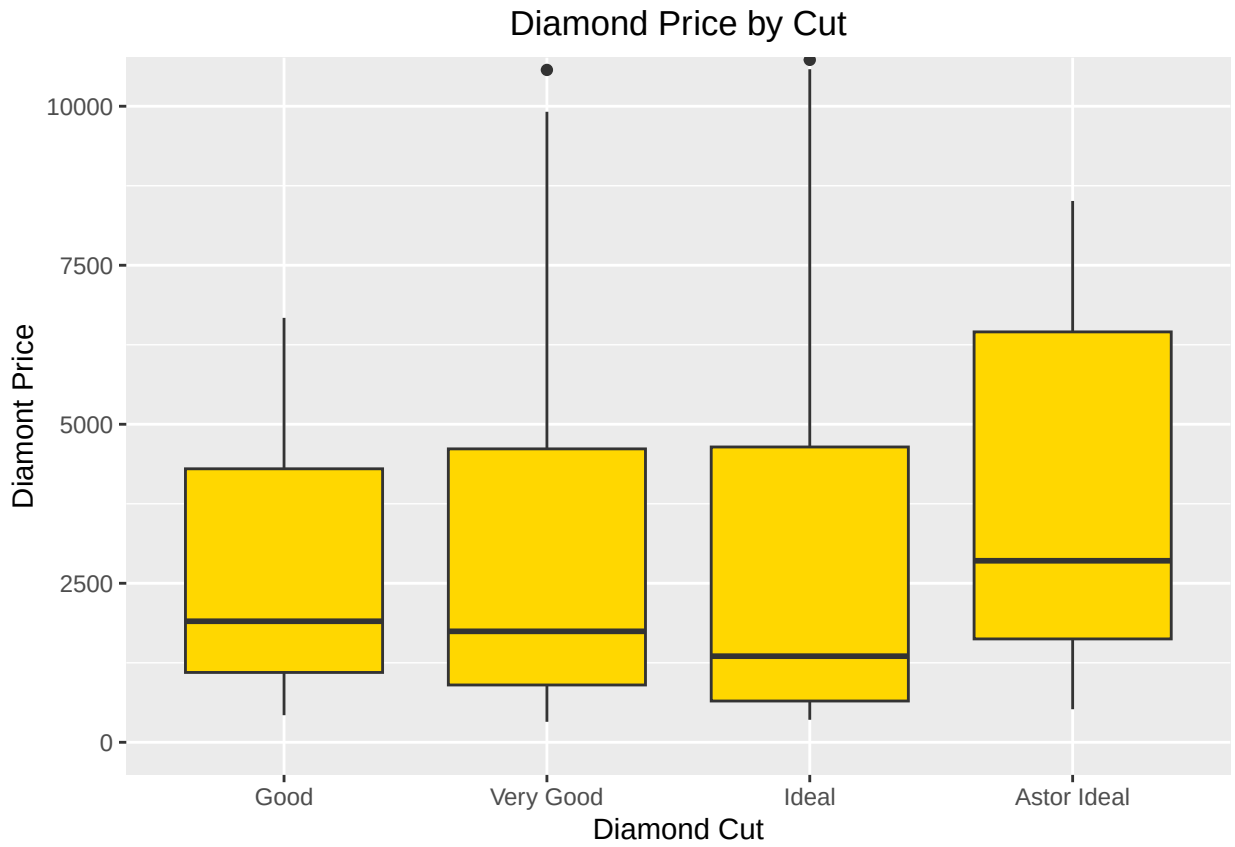


The cut is highly skewed regarding available quantity of each type. There are vastly more ideal cuts (739) and very good cuts (382) compared to the good cuts (73) and the Astor ideal cuts (20).

Cut versus Price Graphs



When viewing these original side-by-side boxplots, it is obvious that cut has a great deal of price variation and many outliers. This could suggest the other variables at play (carat, color, and clarity) causing deviations. The most outliers are witnessed in the very good and ideal ranges, however, they are the cuts with the largest quantity.



When viewing the boxed ranges, which represent 50% of the data, the values are surprisingly close between "good", "very good", and "ideal". "Astor ideal" has a pretty wide range and trends more expensive in the 50% range, but the sample size is very low at only 20 diamonds, and it shares more than half of its range with very good and ideal cuts.

```
## # A tibble: 4 x 7
##   cut            q0   q25   q50   q75   q100  mean
##   <fct>       <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl>
## 1 Good          426 1098  1903  4300  165766 9467.
## 2 Very Good     322  901  1744. 4613  355403 7758.
## 3 Ideal         354  649  1354  4643  345397 6489.
## 4 Astor Ideal   520 1625. 2854  6452.  46893 5852.
```

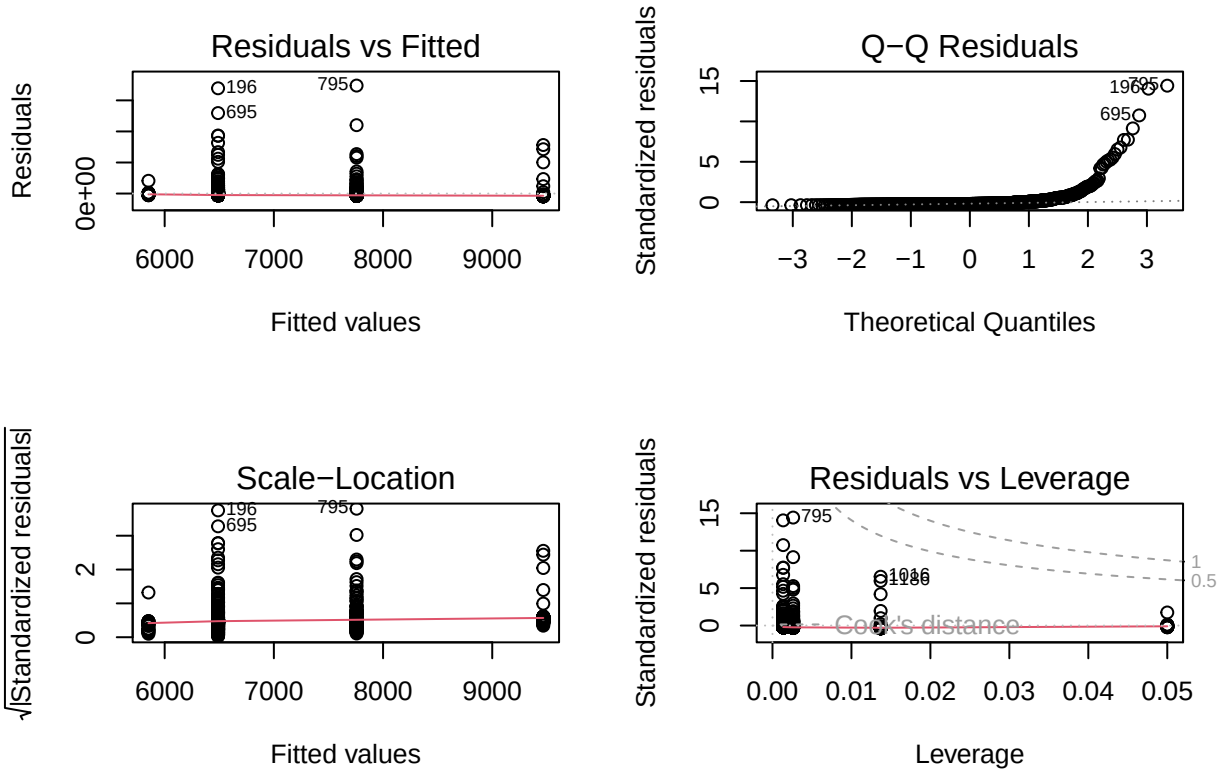
Viewing the quantiles reinforces the similarity in the q75(the value in which 75% of the data lies under) for all cuts.

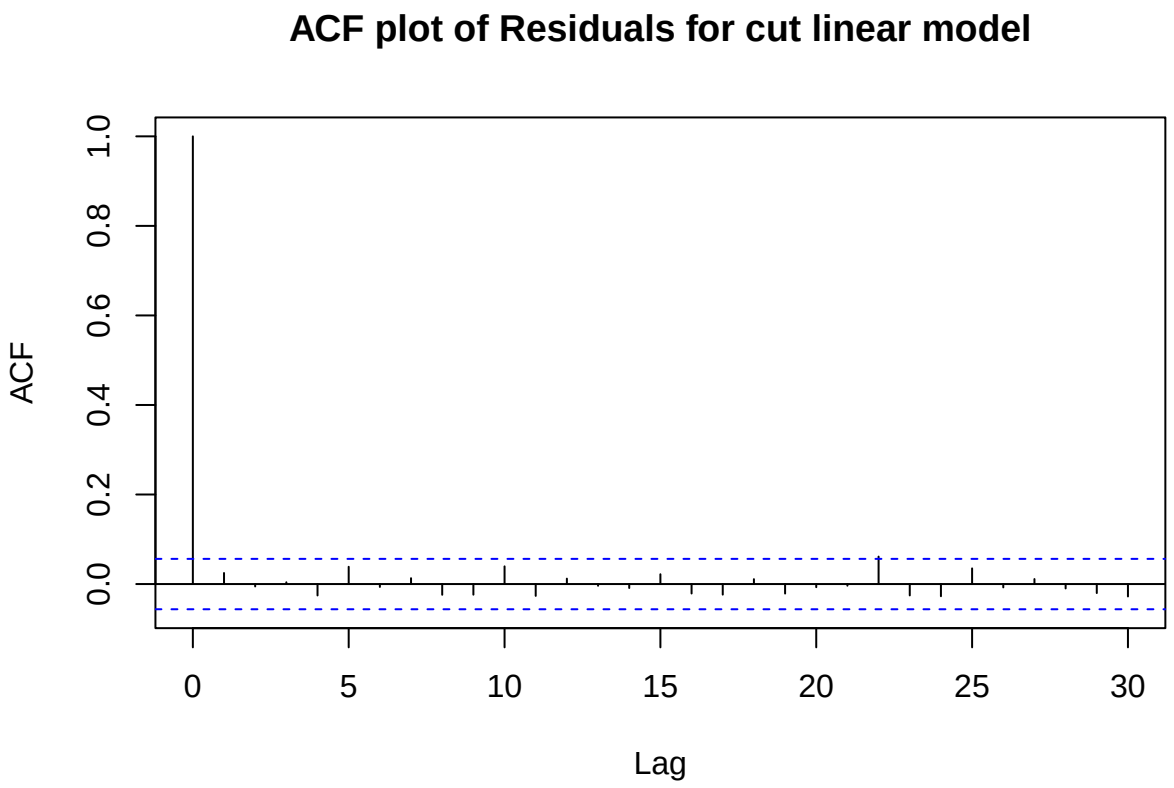
Linear regression for Cut and Price

```
##
## Call:
## lm(formula = price ~ cut, data = Diamonds)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -9041  -5987  -5330  -2787 347645
##
## Coefficients:
```

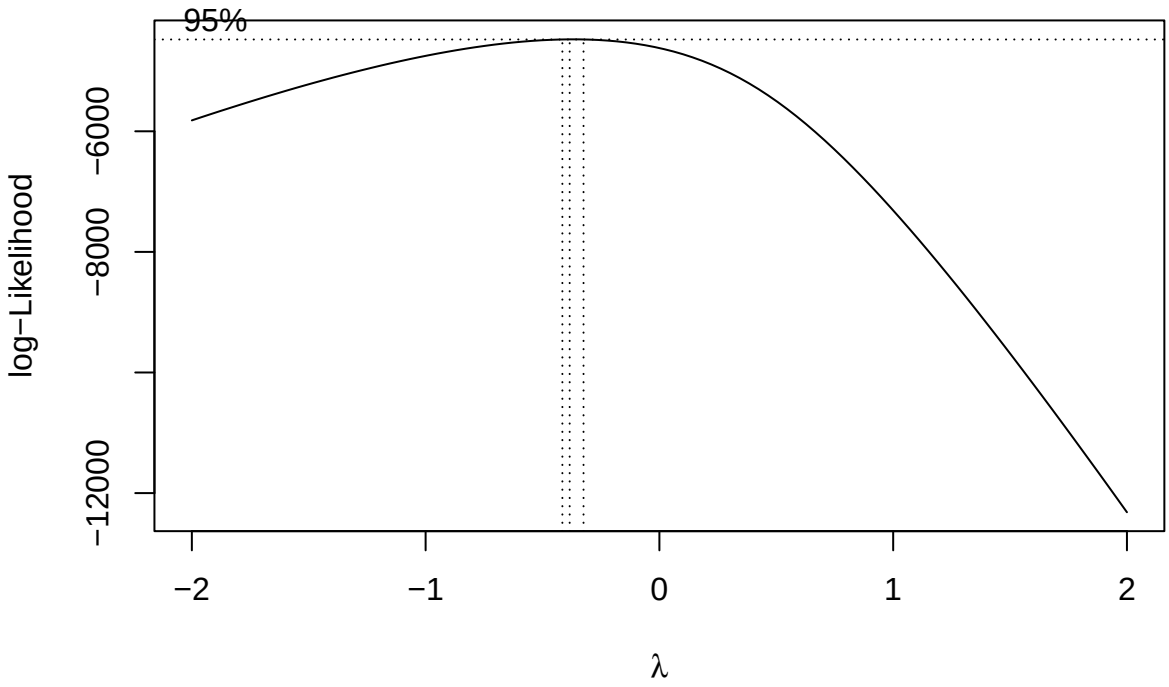
```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5851.6     5397.3   1.084    0.279
## cutGood        3615.7     6091.9   0.594    0.553
## cutIdeal        637.4     5469.8   0.117    0.907
## cutVery Good   1906.1     5536.8   0.344    0.731
##
## Residual standard error: 24140 on 1210 degrees of freedom
## Multiple R-squared:  0.001246,   Adjusted R-squared:  -0.001231
## F-statistic: 0.503 on 3 and 1210 DF,  p-value: 0.6803
```

$\hat{y} = 5851.6 + 3615.7\,Good + 637.4\,Ideal + 1906.1\,VeryGood$ or $\hat{y} = 5851.6 + 3615.7\,Good + 1906.1\,VeryGood + 637.4\,Ideal$

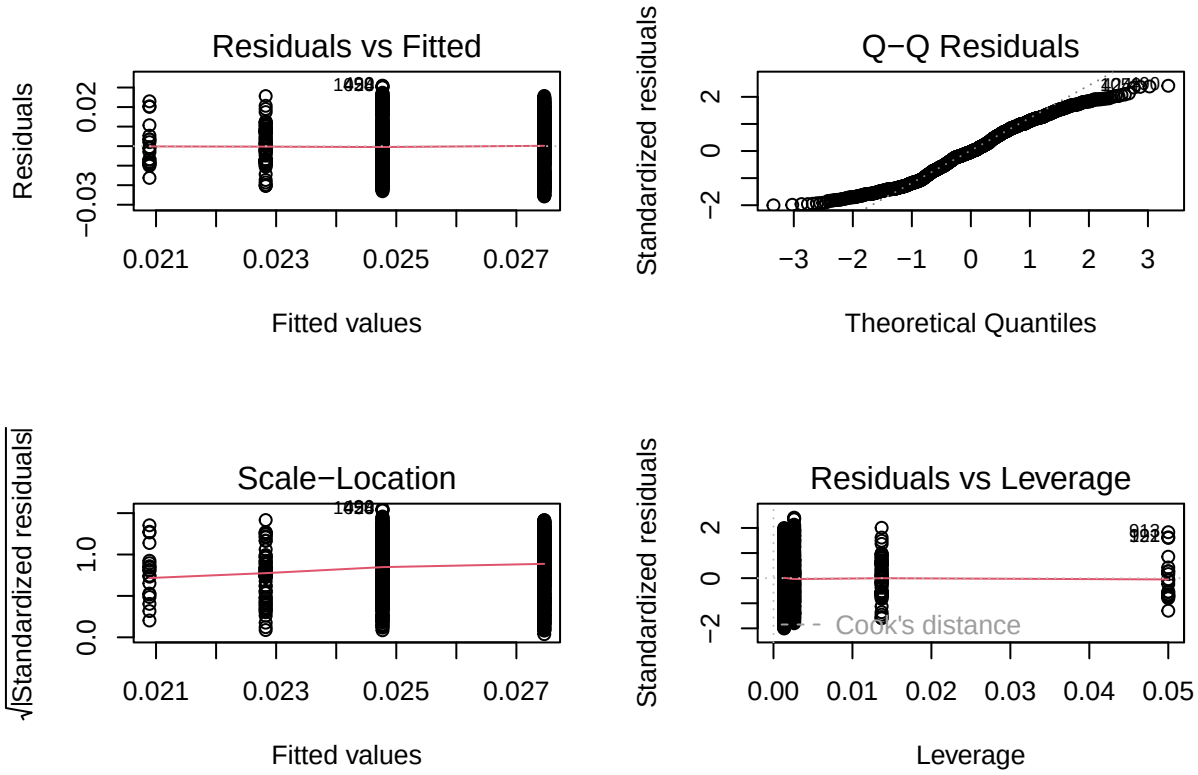




From the Residuals vs. Fitted plot, we can see that assumption 1 is partially met, with a mostly flat red line, but partially violated with all of the points sitting above the line. Additionally, the vertical variance is not very even, so assumption 2 is violated. Though less important, assumption 4 may be violated due to the strangely arcing points away from the line in the QQ Residuals plot. Based on the ACF plot, assumption 3 is also met, due to no drastic lags except for the typical line at 0.



The Boxcox test suggests a lambda of 1/sqrt(y), since it is close to -0.5.



The transformation fixes both assumption 1 and 2 violations. Points sit evenly above and below the line and the vertical variance is very similar now. Additionally, the QQ plot has improved somewhat.

```
##
## Call:
## lm(formula = ystar2 ~ cut, data = Diamonds)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0257629 -0.0100970 -0.0004244  0.0106793  0.0309569
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.020890   0.002877   7.261 6.85e-13 ***
## cutGood      0.001935   0.003247   0.596   0.5513
## cutIdeal     0.006574   0.002916   2.255   0.0243 *
## cutVery Good 0.003880   0.002951   1.315   0.1888
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01287 on 1210 degrees of freedom
## Multiple R-squared:  0.01648,    Adjusted R-squared:  0.01404
## F-statistic: 6.757 on 3 and 1210 DF,  p-value: 0.000161
```

The R2 of 0.01648 shows that this is an extremely poor fit for a model. Based on that value, and the rest of the analysis, very little of the price might be explained by the cut alone.