1. Perform a two way ANOVA on the data in twan03.xls available on ecourseware. This data is the number of spruce moths found in traps classified by Factor 1: location of trap in tree – Top, Middle, lower, and ground and Factor 2: type of lure in trap – scent, sugar, chemical. You will need to rearrange the data so that each row is one observation with three variables – Factor 1, Factor 2 and Count

Answer:

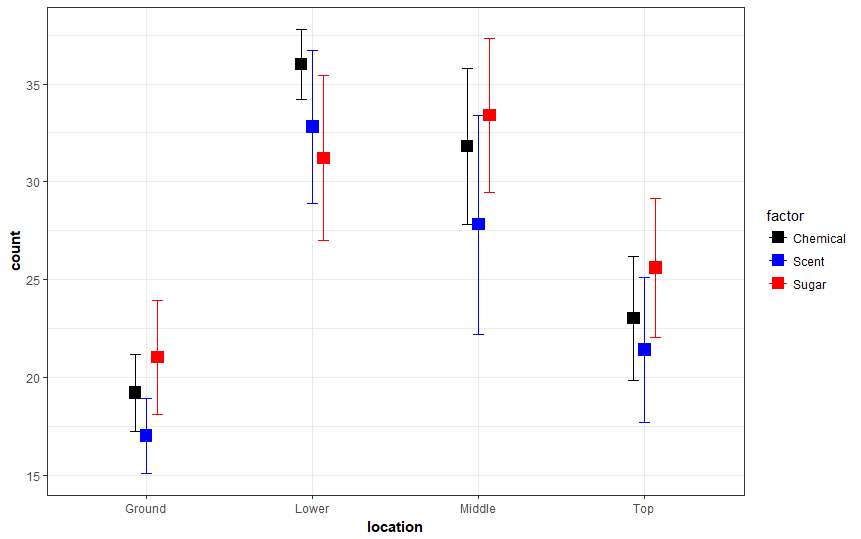
Count is a continuous variable and factor1 and factor2 are nominal.

This is the summary containing mean and standard deviation for factors

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Factor1 | Factor2 | N | Count mean | Count sd | Count variance |
| Ground | Chemical | 5 | 19.20 | 4.44 | 19.70 |
| Ground | Scent | 5 | 17.00 | 4.30 | 18.50 |
| Ground | Sugar | 5 | 21.00 | 6.52 | 42.50 |
| Lower | Chemical | 5 | 36.00 | 4.00 | 16.00 |
| Lower | Scent | 5 | 32.80 | 8.76 | 76.70 |
| Lower | Sugar | 5 | 31.20 | 9.44 | 89.20 |
| Middle | Chemical | 5 | 31.80 | 8.90 | 79.20 |
| Middle | Scent | 5 | 27.80 | 12.56 | 157.70 |
| Middle | Sugar | 5 | 33.40 | 8.82 | 77.80 |
| Top | Chemical | 5 | 23.00 | 7.07 | 50.00 |
| Top | Scent | 5 | 21.40 | 8.26 | 68.30 |
| Top | Sugar | 5 | 25.60 | 7.92 | 62.80 |

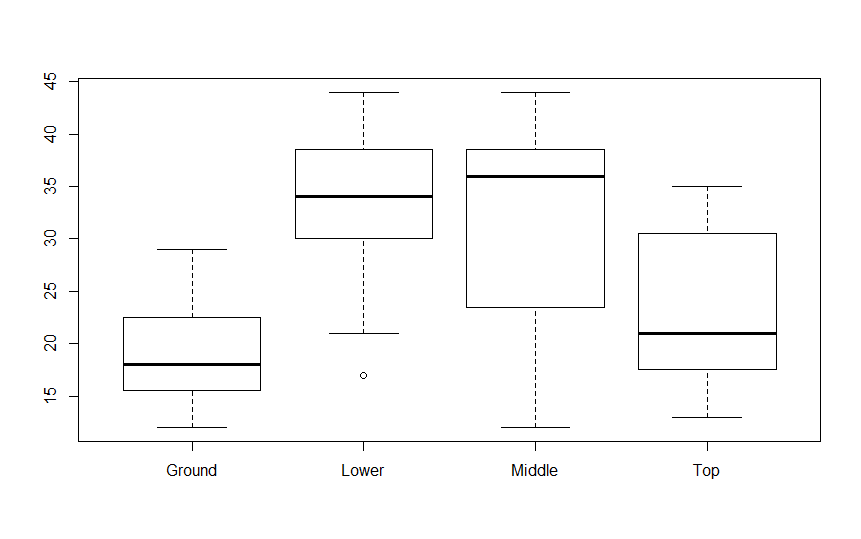
No variance value is 10 times the other so let’s assume equal variance.

This is the plot of means for different factors



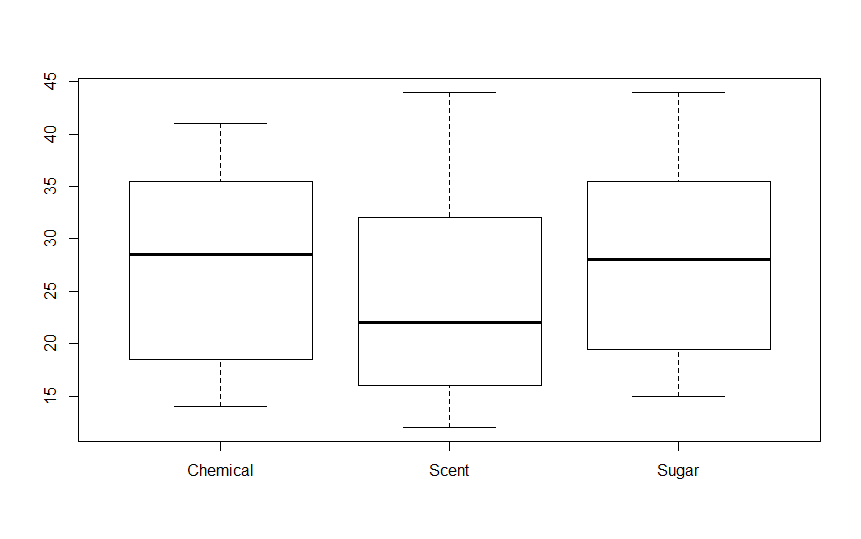
Means look different when location == lower.

This is the boxplot of count with respect to location.



Means look different

This is the boxplot of count with respect to type of lure(factor2).



Means look potentially the same.

An Anova model gives the following statistics

Df Sum Sq Mean Sq F value Pr(>F)

factor 2 113 56.5 0.894 0.416

location 3 1981 660.5 10.450 2.09e-05 \*\*\*

factor:location 6 115 19.2 0.303 0.932

Residuals 48 3034 63.2

So the interaction effect does not look significant. So we look into each factor separately.

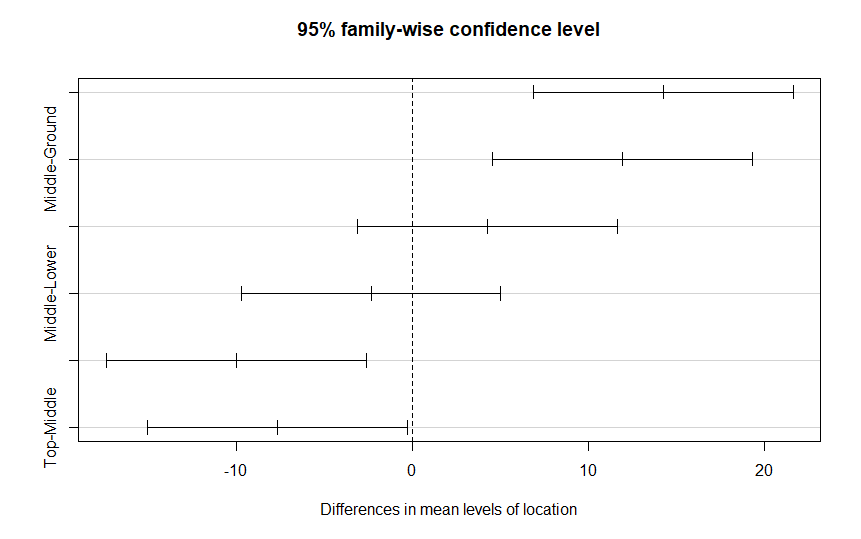
This is the summary of an Anova model when only location of the trap is considered

Df Sum Sq Mean Sq F value Pr(>F)

location 3 1981 660.5 11.34 6.46e-06 \*\*\*

The effect of Location looks significant

This is the tuckey plot which indicates there is difference in means for each location.



These are the pairwise t test values

Ground Lower Middle

Lower 2.3e-05 - -

Middle 0.00037 0.40598 -

Top 0.26276 0.00280 0.02396

Indicating difference in means for Lower-Ground, Middle-Ground,Lower-Top and Middle-Top among 6 possible pairs.

This is the summary of an Anova model when only type of lure in the trap is considered

Df Sum Sq Mean Sq F value Pr(>F)

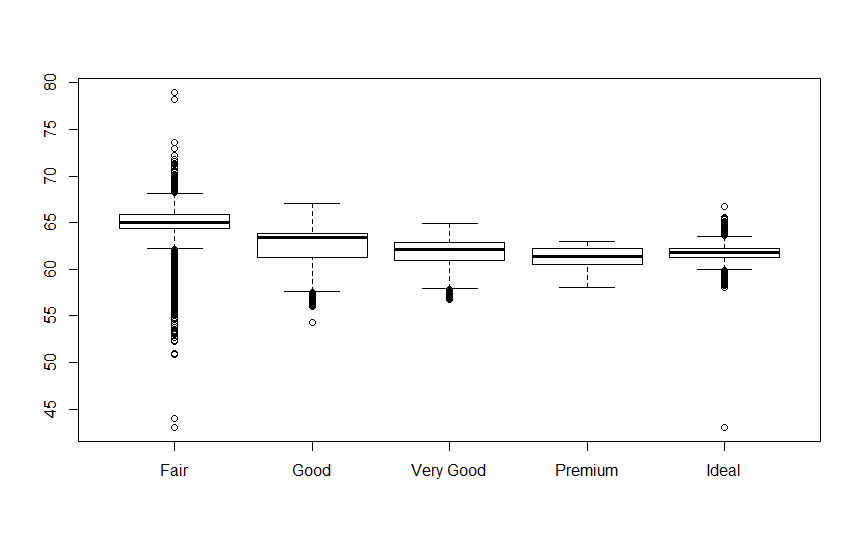
Type\_of\_lure 2 113 56.52 0.628 0.537

The effect of the type of lure does not look significant.

1. Perform a three-way ANOVA on the data in the dataset *diamonds* in R package *ggplot2*. Model depth in terms of the three factors, cut, color and clarity

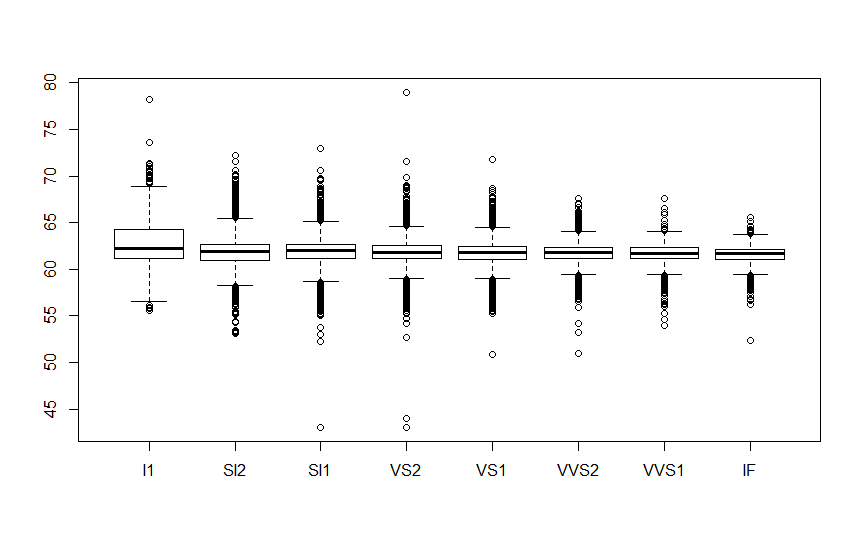
Answer:

This is the boxplot of depth by cut



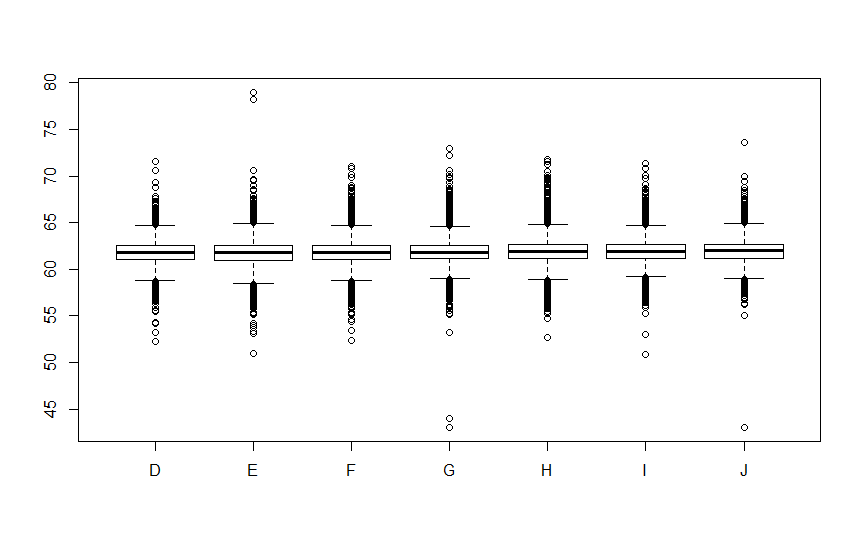
Means look almost the same.

This is the boxplot of depth by clarity



Means look same.

This is the boxplot of depth by color



Means look same.

The variance table for each factor shows that there is a difference in variance for factors combined in a three way. For example

|  |  |  |  |
| --- | --- | --- | --- |
| **Cut** | **Clarity** | **Color** | **Depth Variance** |
| Fair | VS1 | I | 21.0137333 |
| Good | VS1 | I | 4.7466343 |
| Very Good | VS1 | I | 1.643978 |

So our assumption of equal variance wont work.

A three way anova model gives the following statistics

Df Sum Sq Mean Sq F value Pr(>F)

cut 4 13656 3414 1961.056 < 2e-16 \*\*\*

color 6 285 48 27.290 < 2e-16 \*\*\*

clarity 7 380 54 31.192 < 2e-16 \*\*\*

cut:color 24 329 14 7.869 < 2e-16 \*\*\*

cut:clarity 28 1511 54 31.004 < 2e-16 \*\*\*

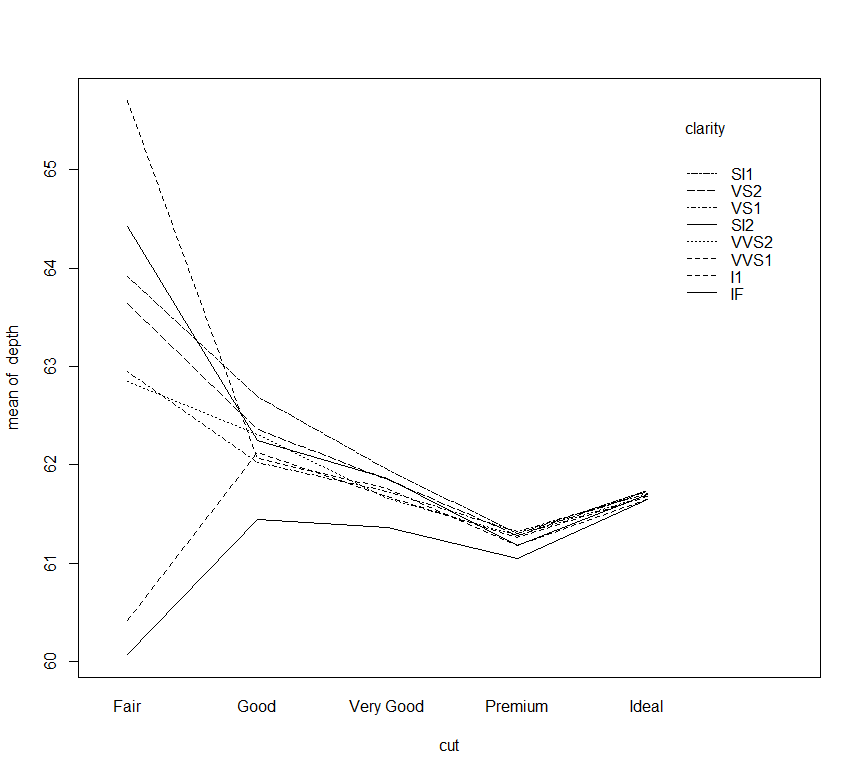
color:clarity 42 232 6 3.171 2.15e-11 \*\*\*

cut:color:clarity 164 884 5 3.098 < 2e-16 \*\*\*

Residuals 53664 93427 2

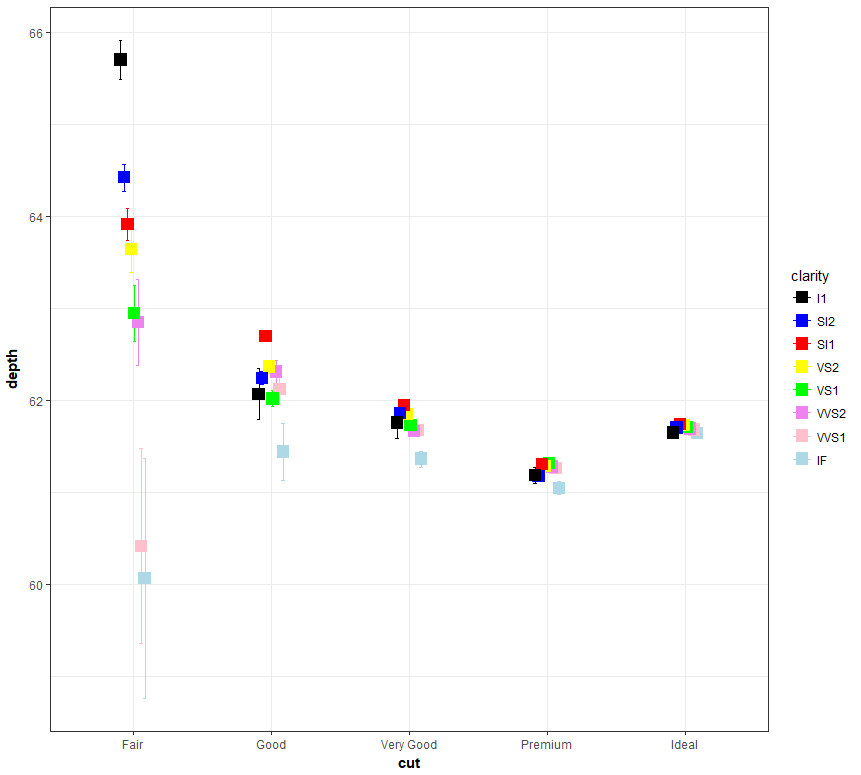
The interaction effects look significant.

This is the interaction plot for depth vs cut and clarity



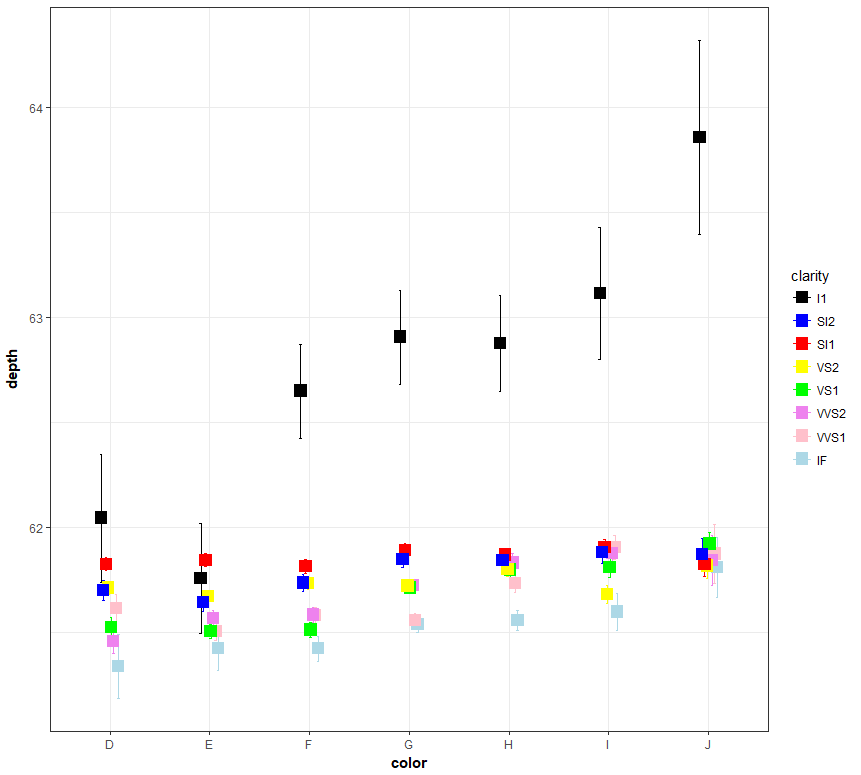
Looks like clarity = “IF ” negates the effect of cut = “fair”

This is the mean plot of depth vs cut and clarity



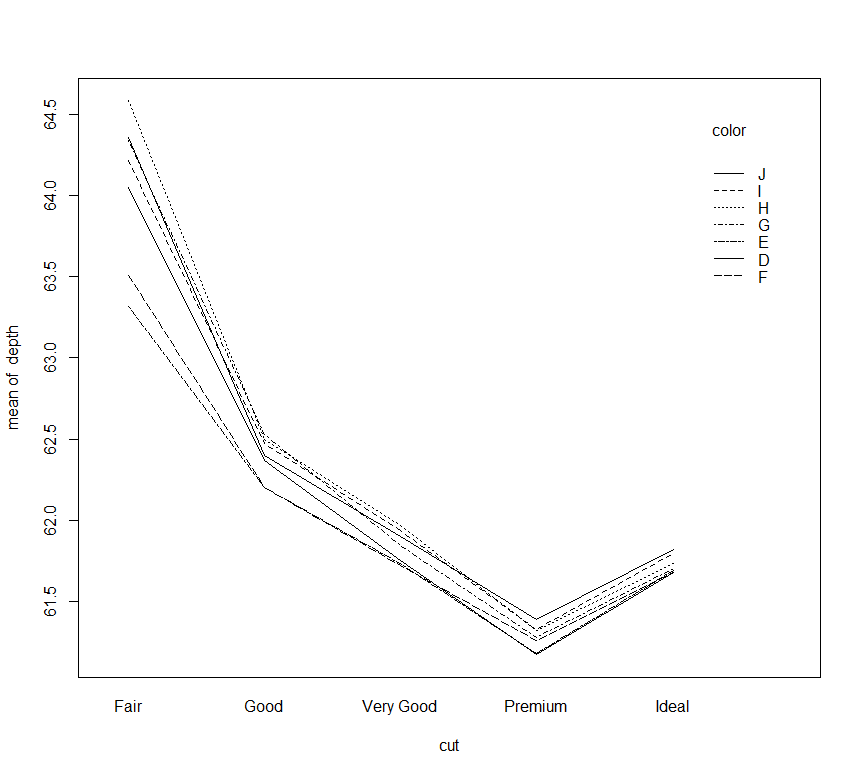
This shows clarity = “IF ” and “VVS1” negates the effect of cut = “fair”

This is the mean plot of depth vs color and clarity



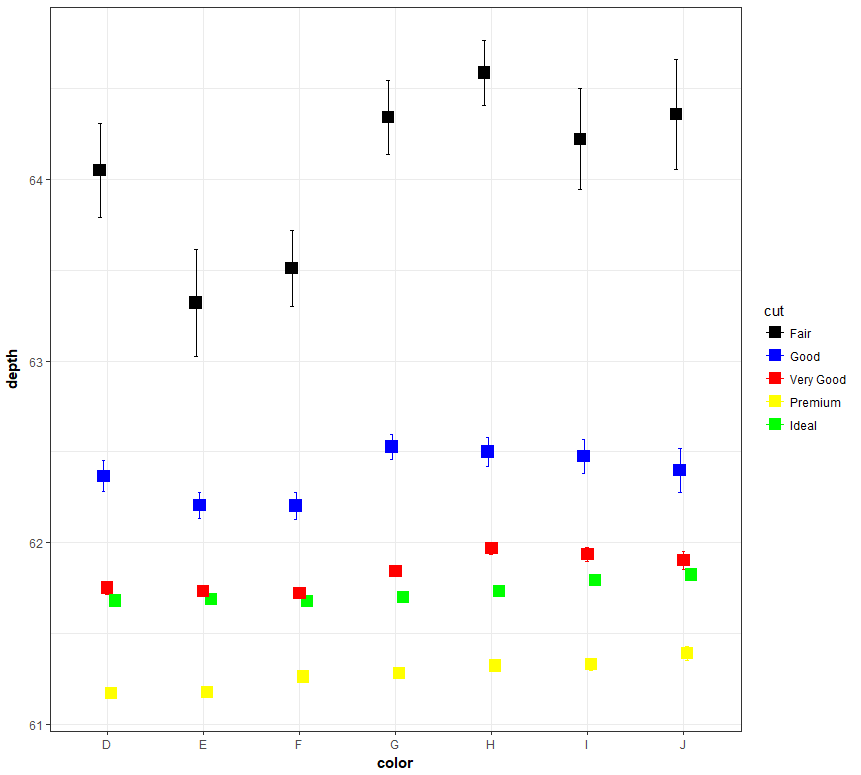
Shows Clarity = I1 increses the effect of every color.

This is the interaction plot of depth, cut and color



Shows cut = fair increases the mean depth for each color

This is the mean plot



Shows Cut = fair has a significant effect.