

**NAME: MUGANGA CHARLES**

**REGISTRATION NUMBER: J22B23/032**

**ACCESSS NUMBER: A96447**

**COURSE UNIT: DATA SCIENCE**

Number 2.

The perception is false because the sum and the average of the variables is positive.

Sum = 212135217

Average = 3933.529

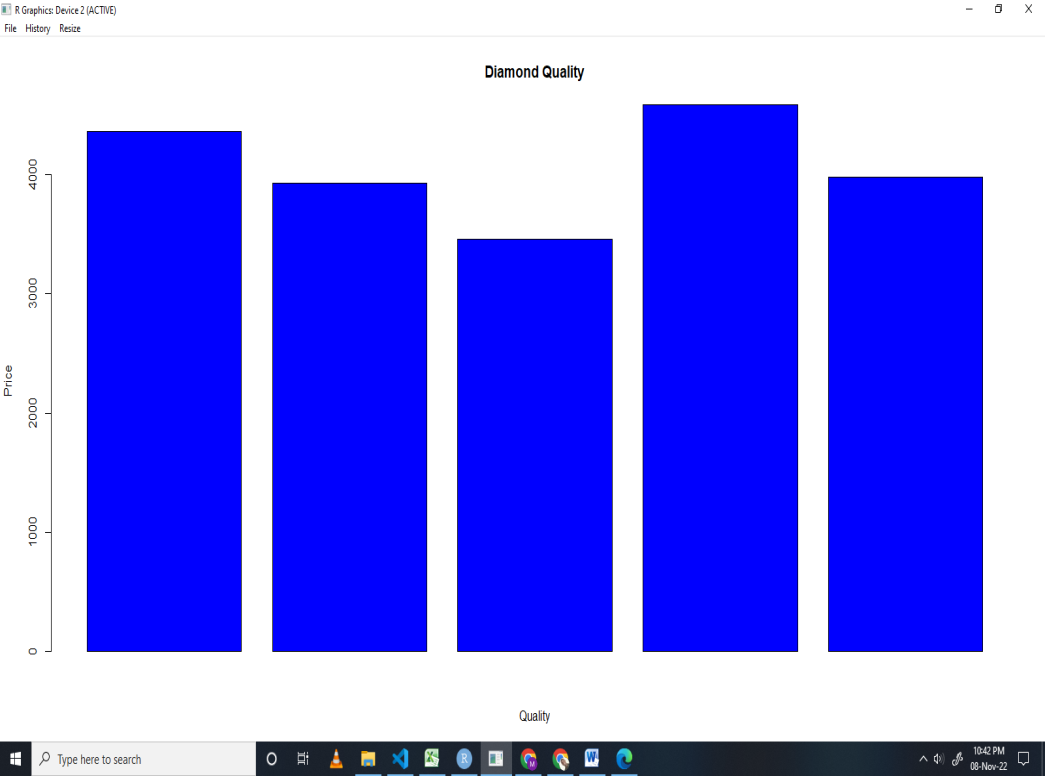
Number 3.

This is true because the occurrence changes. And from the hypothesis, the following table was generated in the process of going around the problem.

Fair Good Ideal Premium Very Good

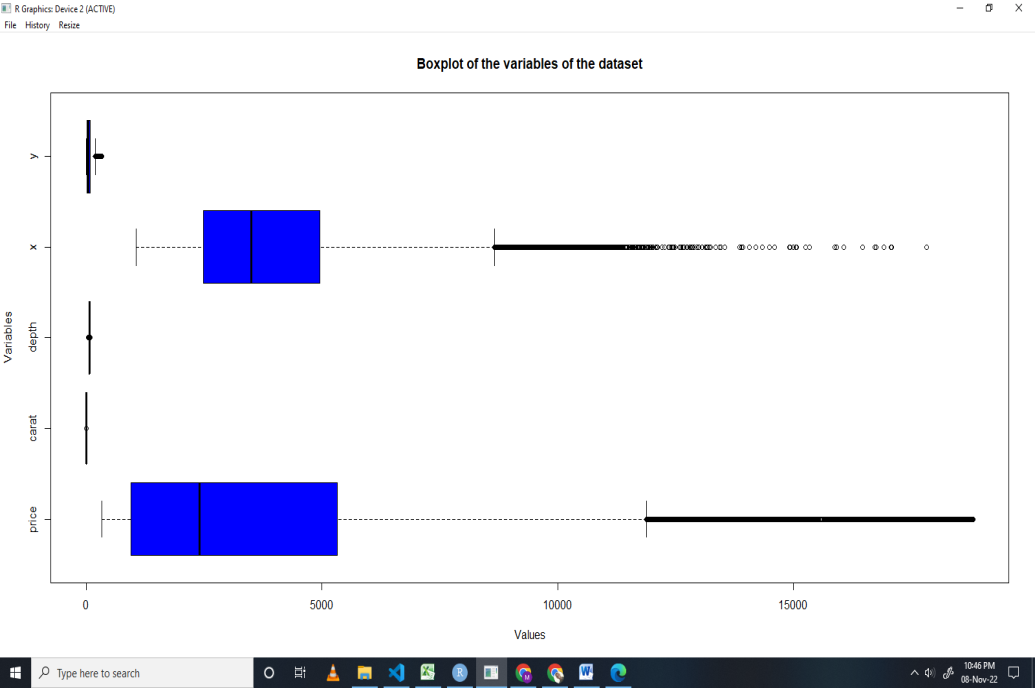
1610 4906 21551 13791 12082

The values are graphically represented To prove the hypothesis and also solve the problem



Number4.

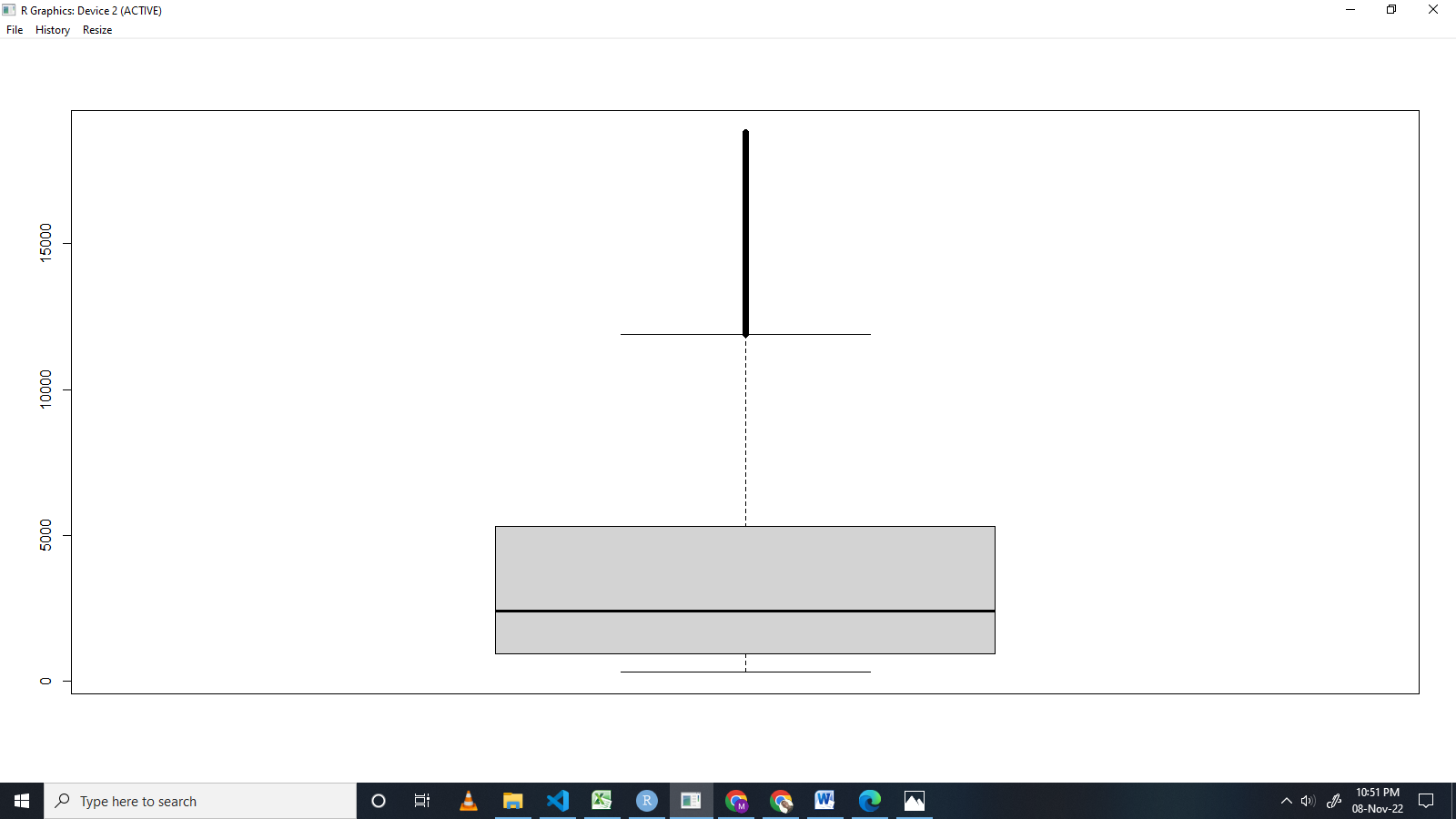
Display of the box plot of all the variable showing the outliers



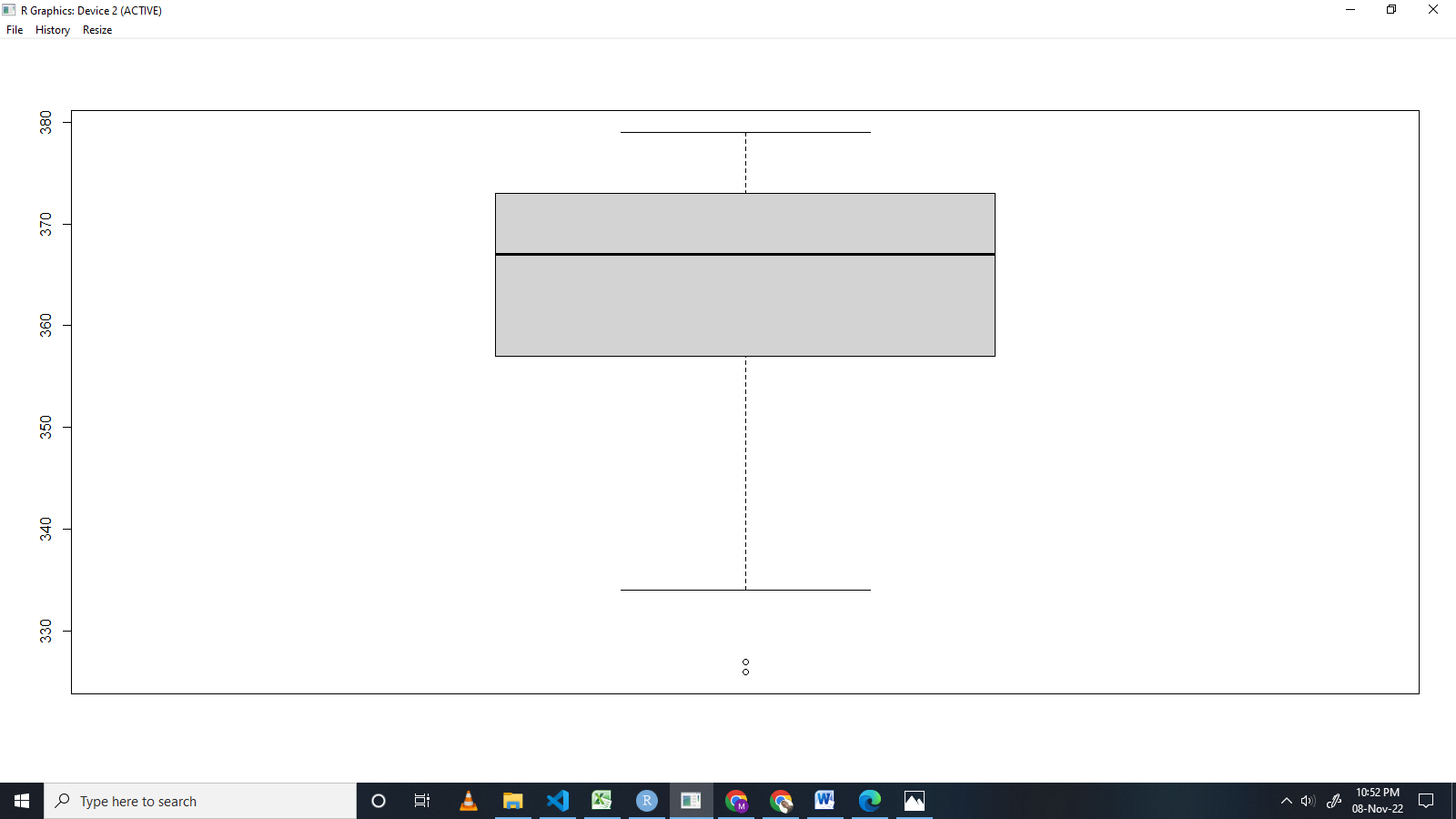
Number 5.

Taking price as my case study.

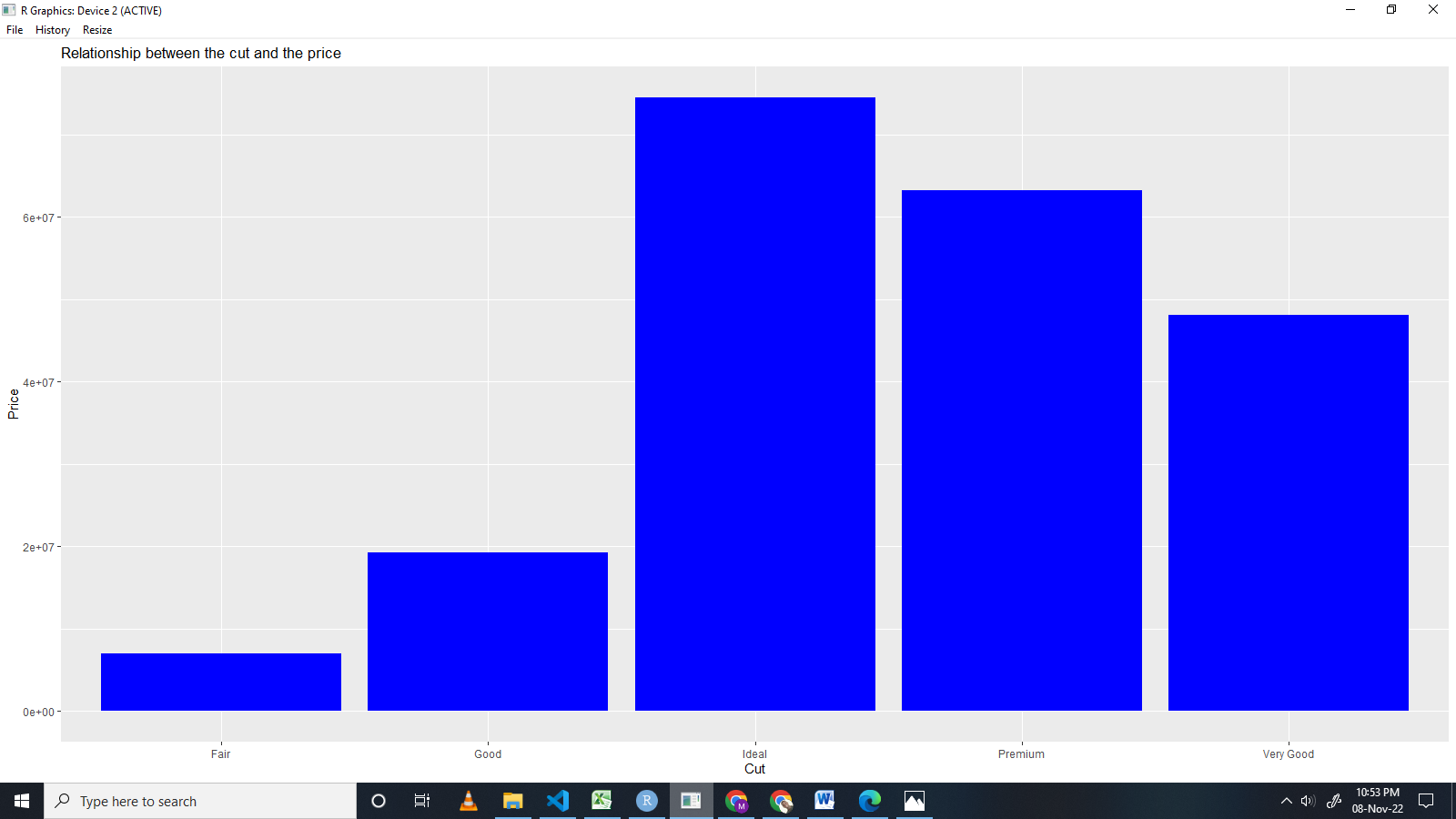
The method I used was the Inter Quatile Range method which enabled me remove he outliers using the limits generated to compare with the data.



Then it we have to removing the outliers.



Number 7



Number 8.

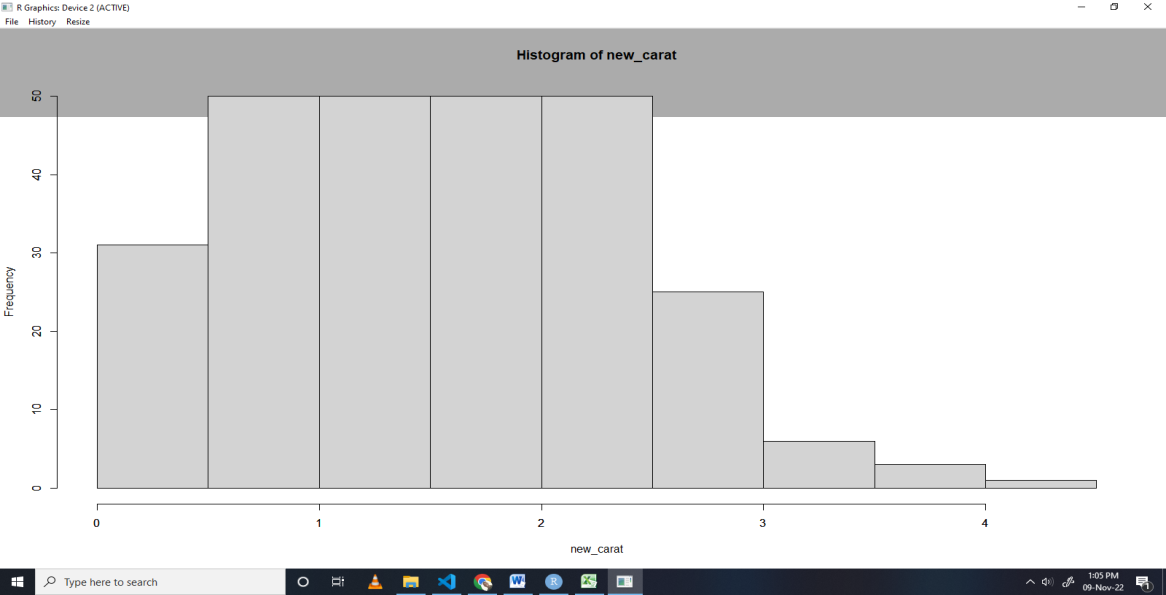
I used Kolmogorov-Smirnov test because the data set has values that are greater than 5000. The values are also ties for this case and then aimed at removing the ties from the data by removing duplicates.

The variables where all called new\_<variable name > after removing the duplicates.

The Kolmogorov-Smirnov test was used hence the results are;

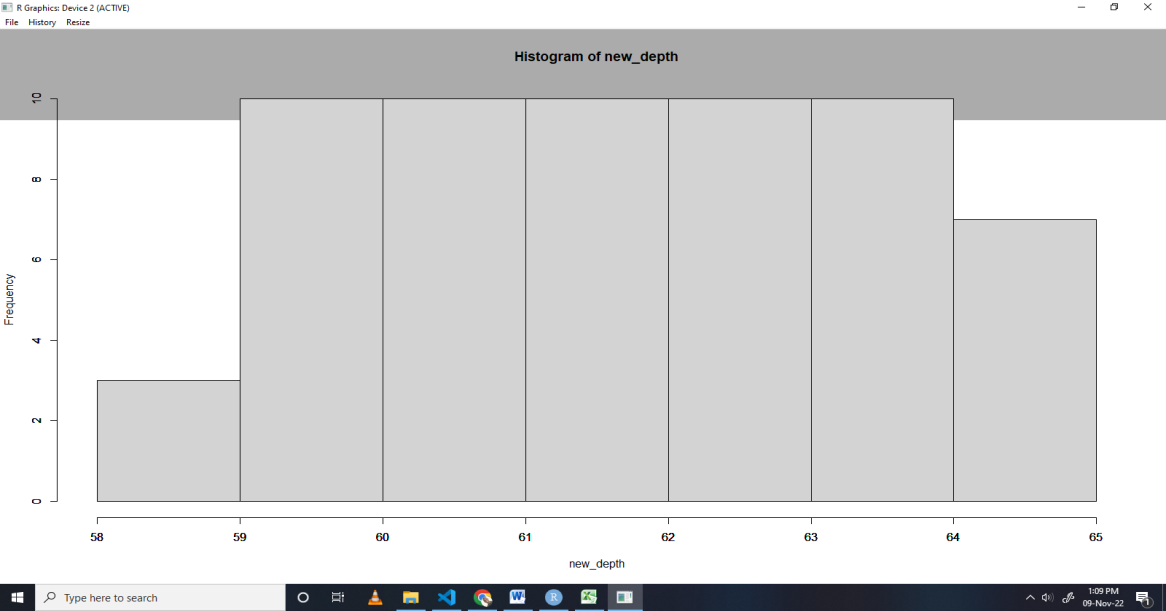
For **carat** the p value is 0.4397 thus normally distributed.

Histogram



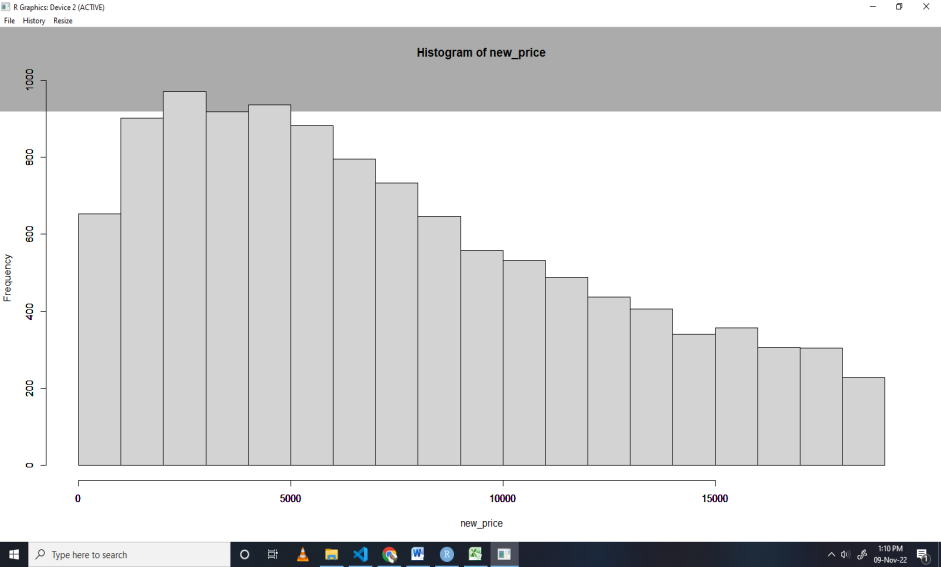
For **depth** the p value is 0.9558 thus normally distributed.

Histogram



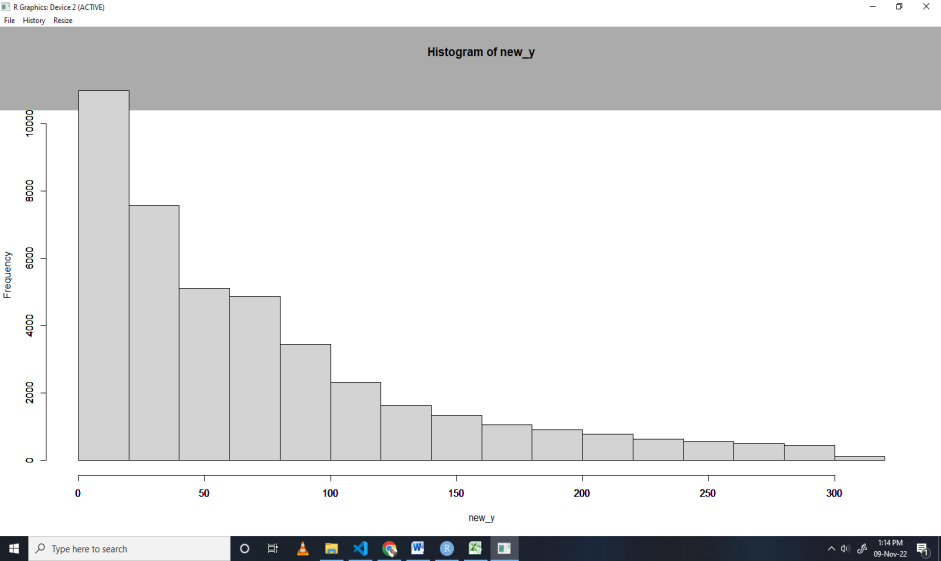
For **price**, the p value is 2.2e-16 thus not normally distributed.

Histogram



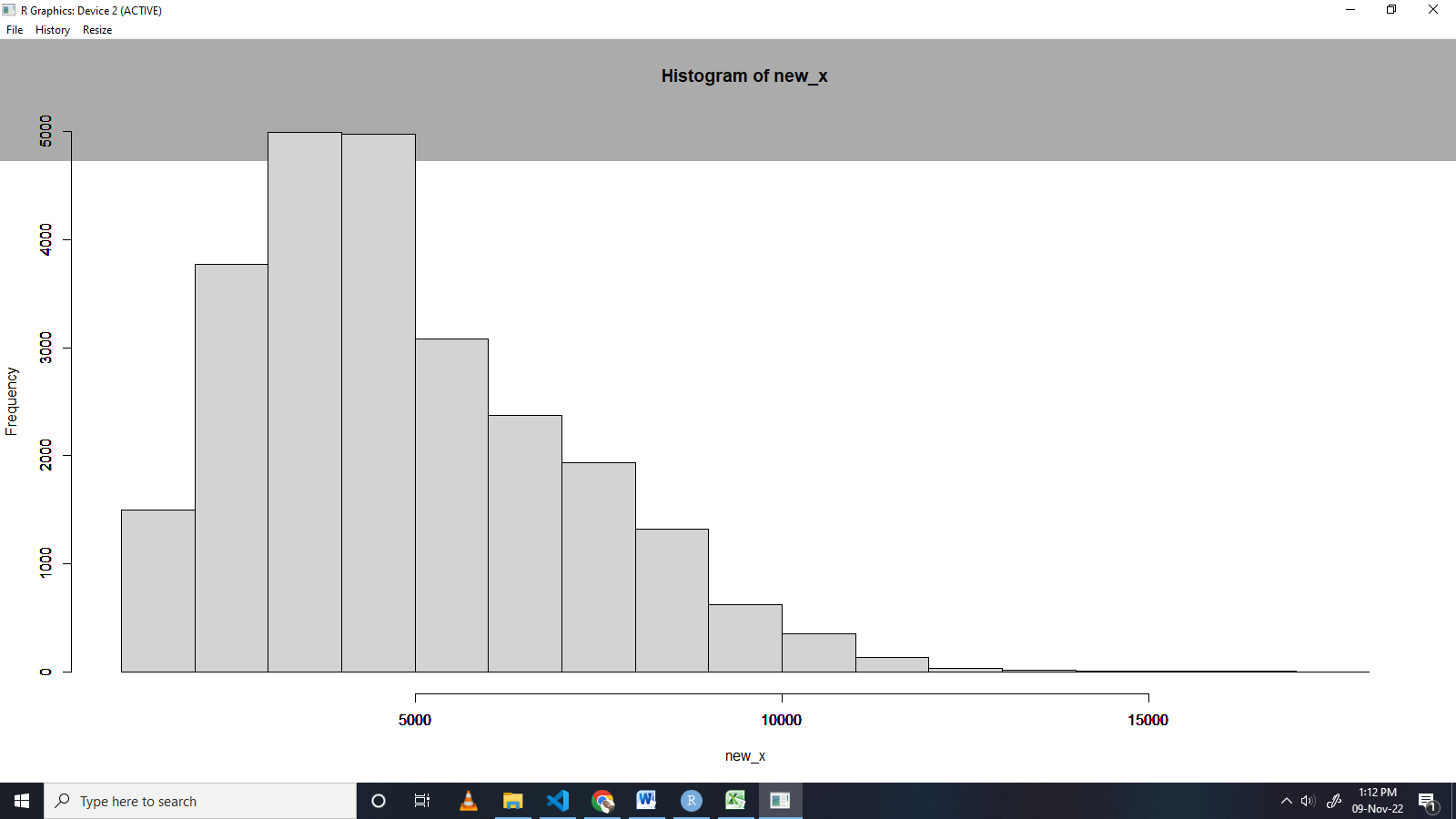
For y, the p value is 2.2e-16. Thus not normally distributed.

Histogram



For x, the p value is 2.2e-16. Thus not normally distributed

Histogram

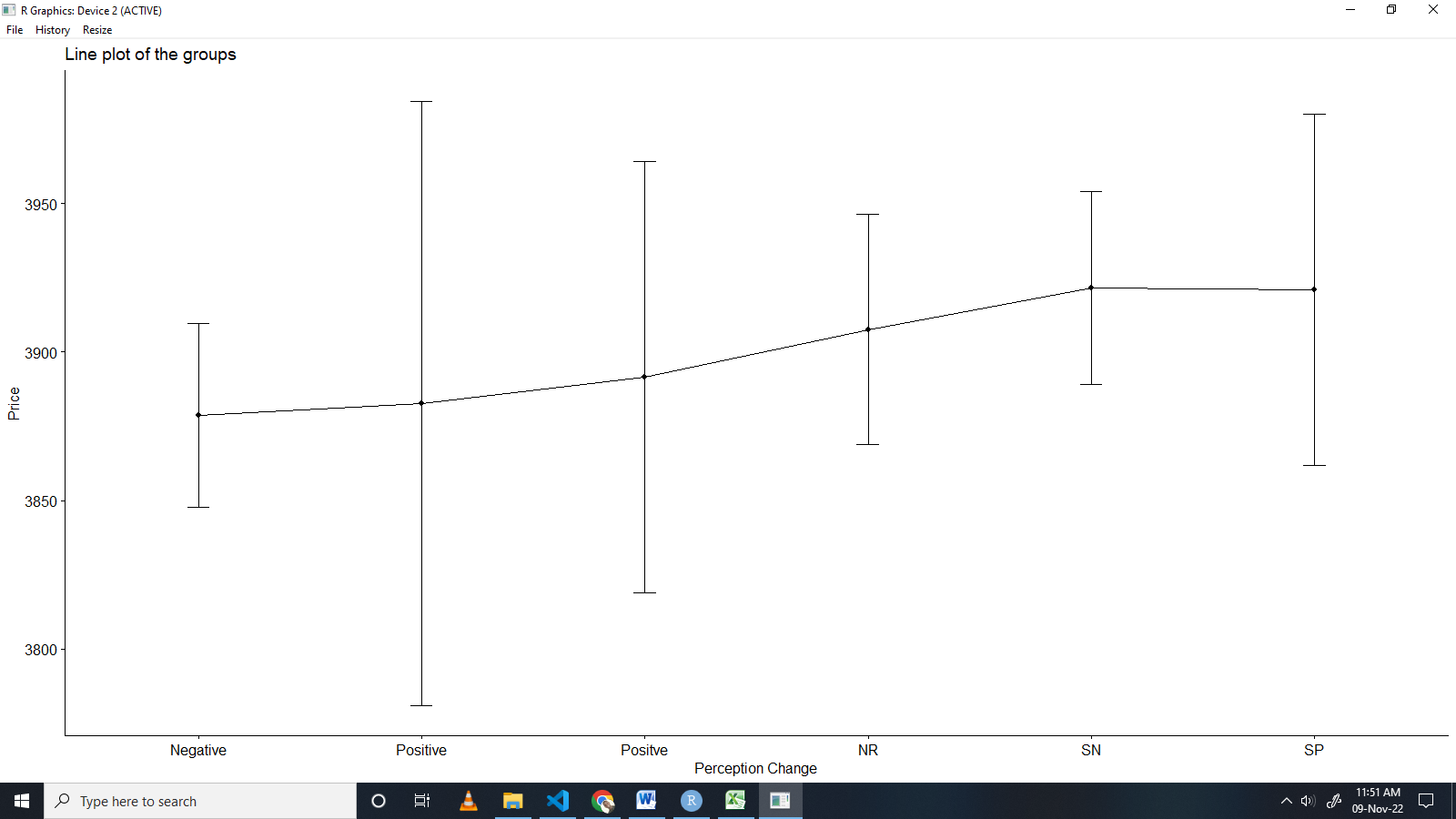


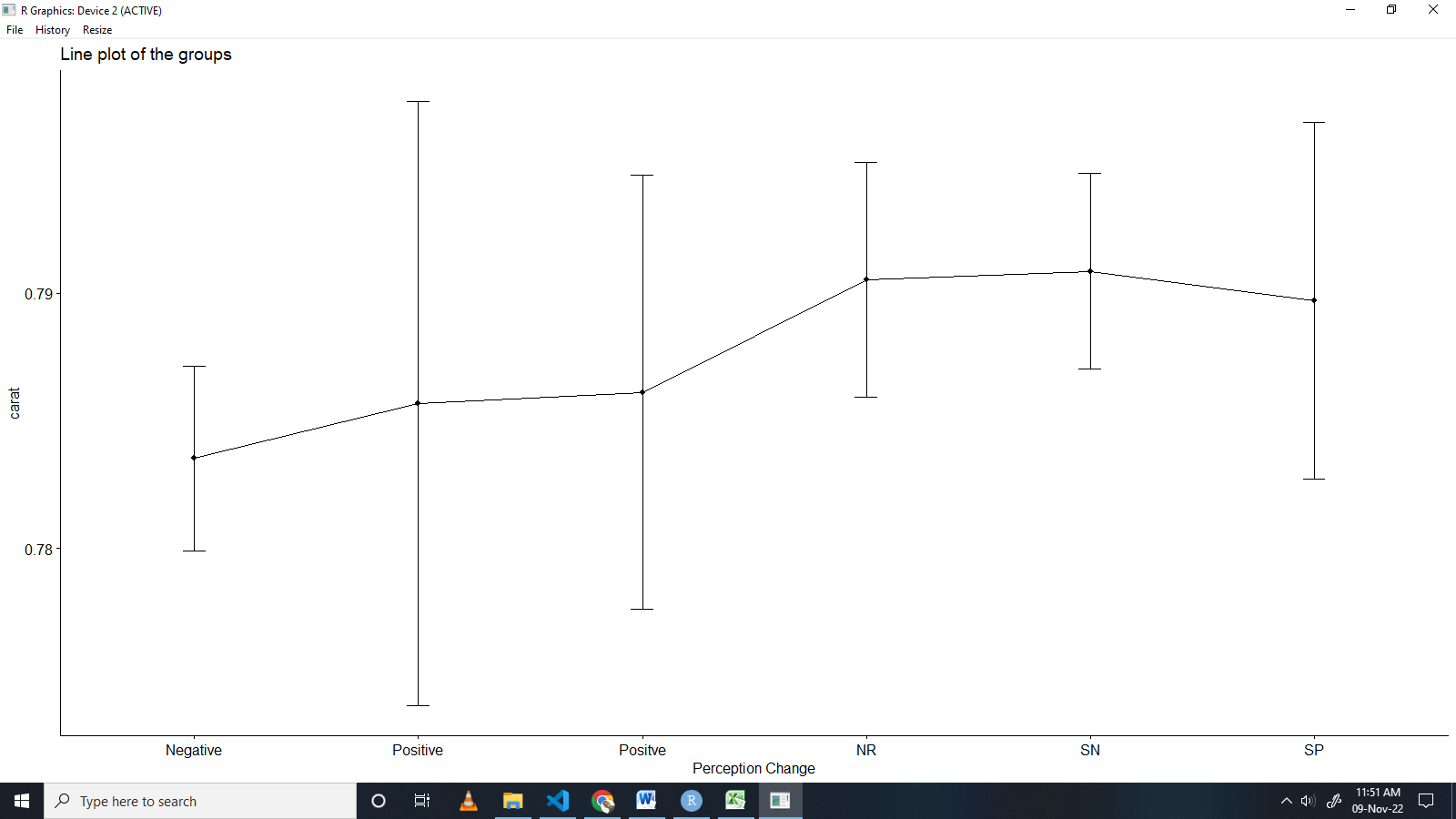
Number9.

From the groups given, The null hypothesis is that the groups have the same variance.

The alternative hypothesis is that the groups have different variances by at least one group having a variance not equal to the others groups

The test used is the one-way ANOVA test because the groups are more than two.





The values after running the anova test are as follows;

The P- value is 0.954 which is greater than 0.05 and therefore we fail to reject the null hypothesis therefore statistically not significant.