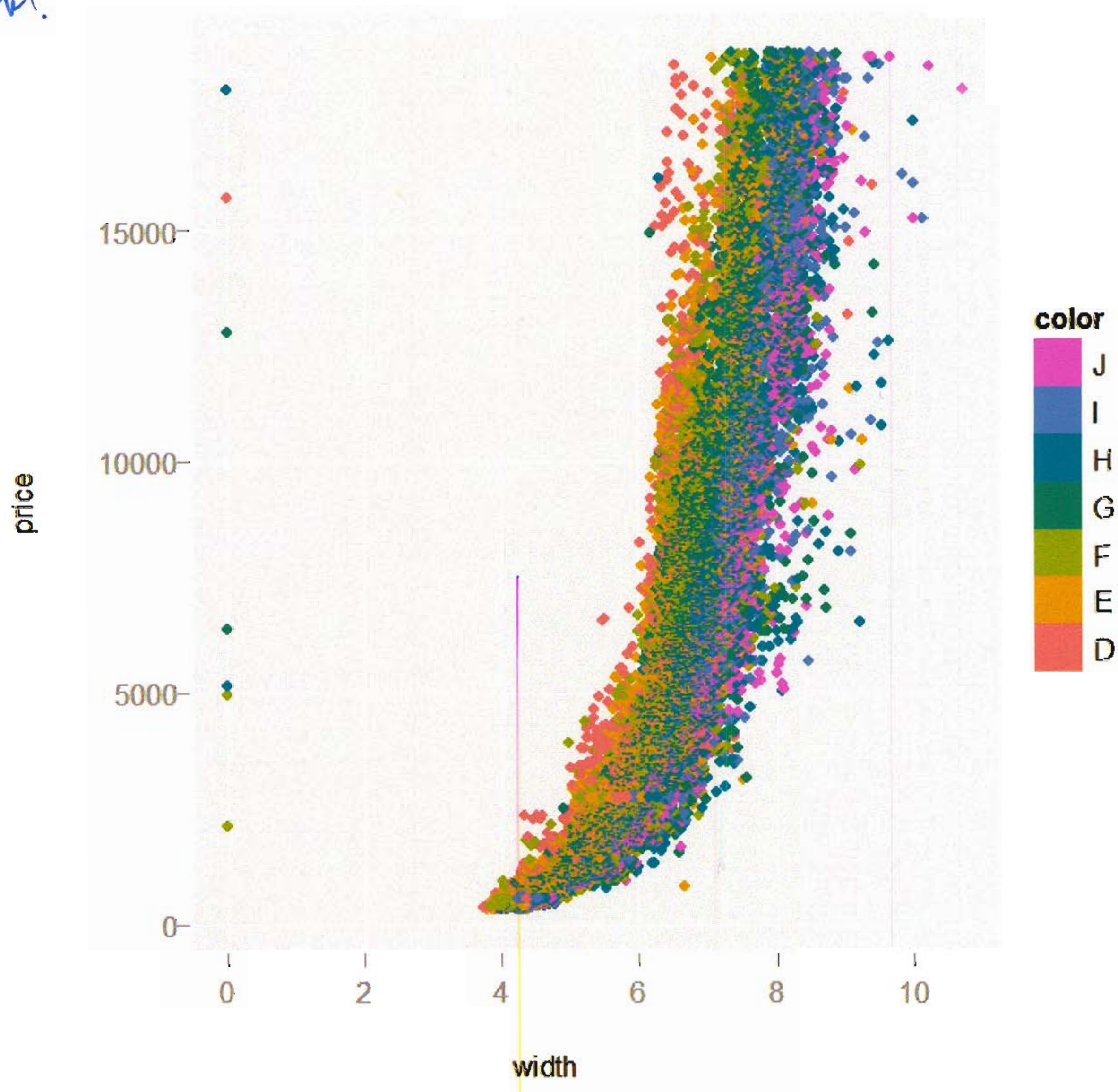
c 5
s 5
o 5

Great work!

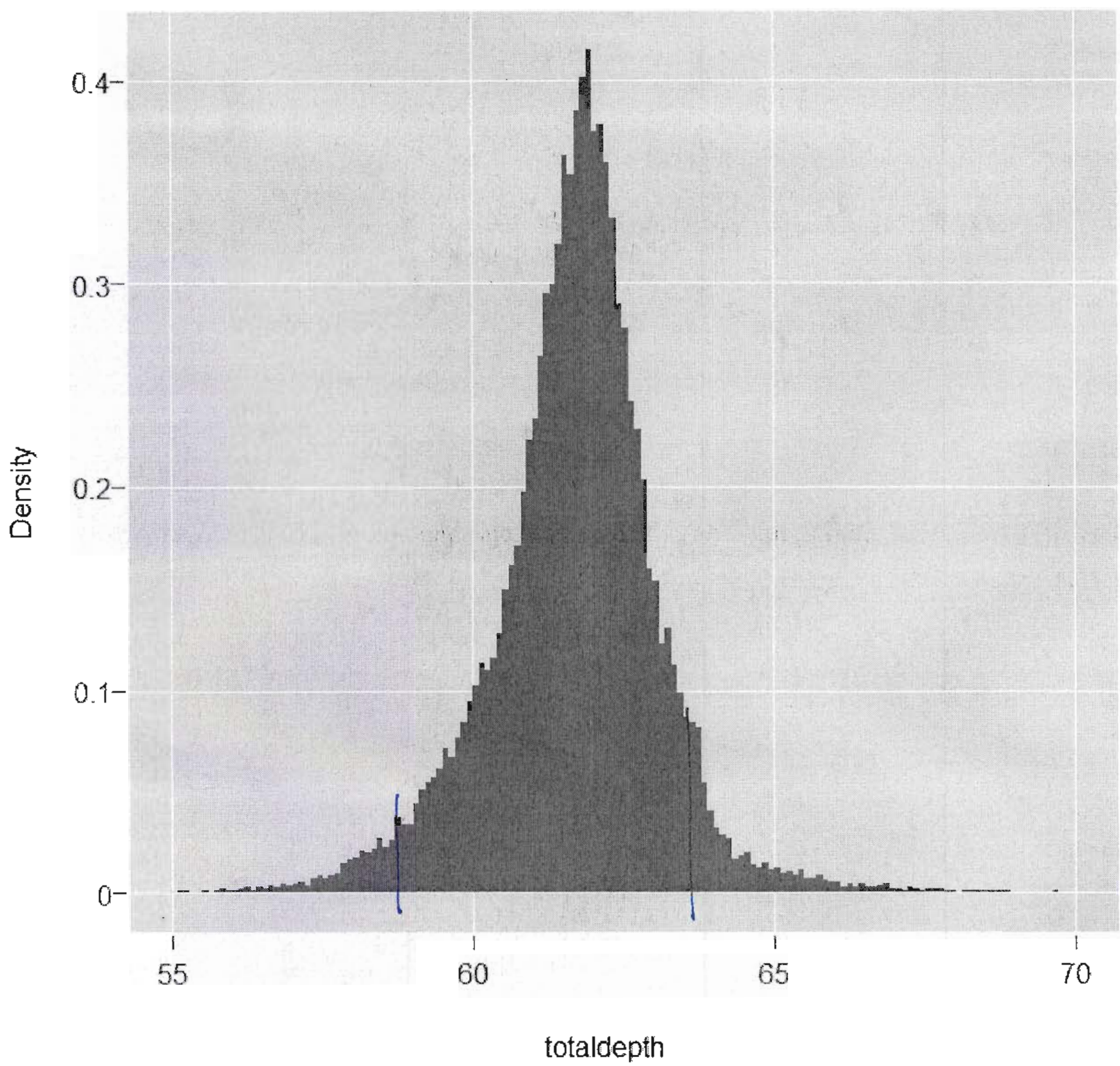
Graphs Illustrating Features of Diamond Data

Stat 480
HW Week 7



Code: qplot(width,price,data=big,colour=color)

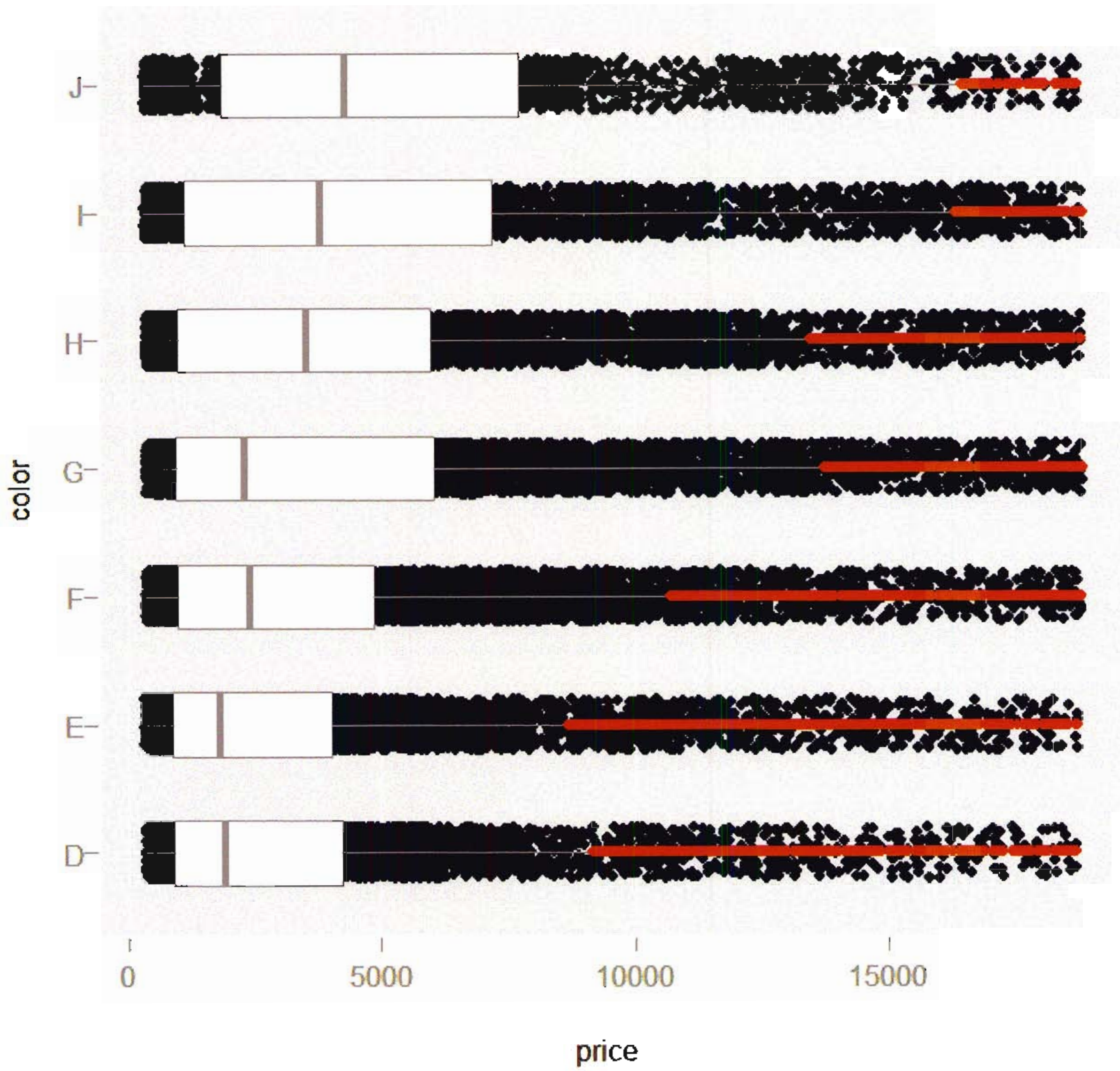
Not only do we observe an increasing trend where as diamond width increases, price also increases, but if you look carefully you can see seven vertical bands in the data. Each band represents a different color of diamond, and they progress as you go from left to right through the colors "D" to "J". I suppose this makes sense on two accounts. One, being that diamonds with color letters lower in the alphabet are more valuable, it seems reasonable to want to cut them to a smaller diamond width in order to maximize the quantity produced. At the same time, less desirable diamonds which lack the "colorless" or "icy" look that "D" colored diamonds have must make up for this shortcoming, and one way to do this is through increased diamond size or width.



Code: qplot(totaldepth,data=big,type="histogram",breaks=(by=500),xlim=c(55,70))

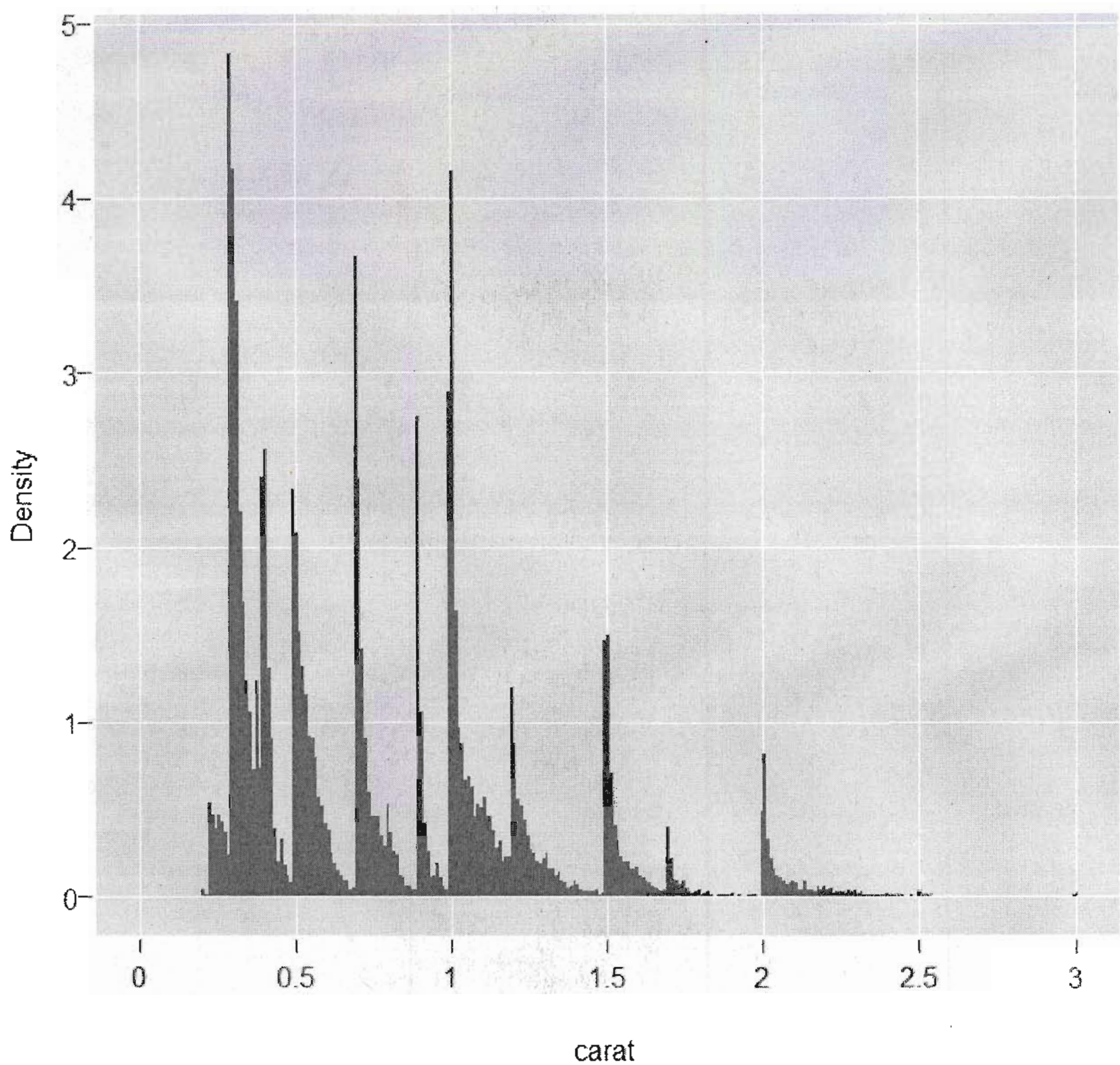
In this histogram we see that the totaldepth values come to a peak around the value of 62. This is not what I expected. I figured the graph would be similar to the histogram of diamond depth, in which we have several peaks at or near whole integer values. Looking at wikipedia, we see that ideal to very fine cut diamonds have total depth percentages between 59% and 63% and that diamonds within this range usually provide excellent brilliance. This would explain why we see a peak around 62%, trailing off rapidly on both sides. Being that diamonds with depth percentages higher than 65% lose enough brilliance that the center of the diamond becomes dark as light refracts through the bottom would explain the near cutoff of diamonds with depth percentages higher than 64%. Also, few diamonds with depth percentages less then 58% exist.

Great!



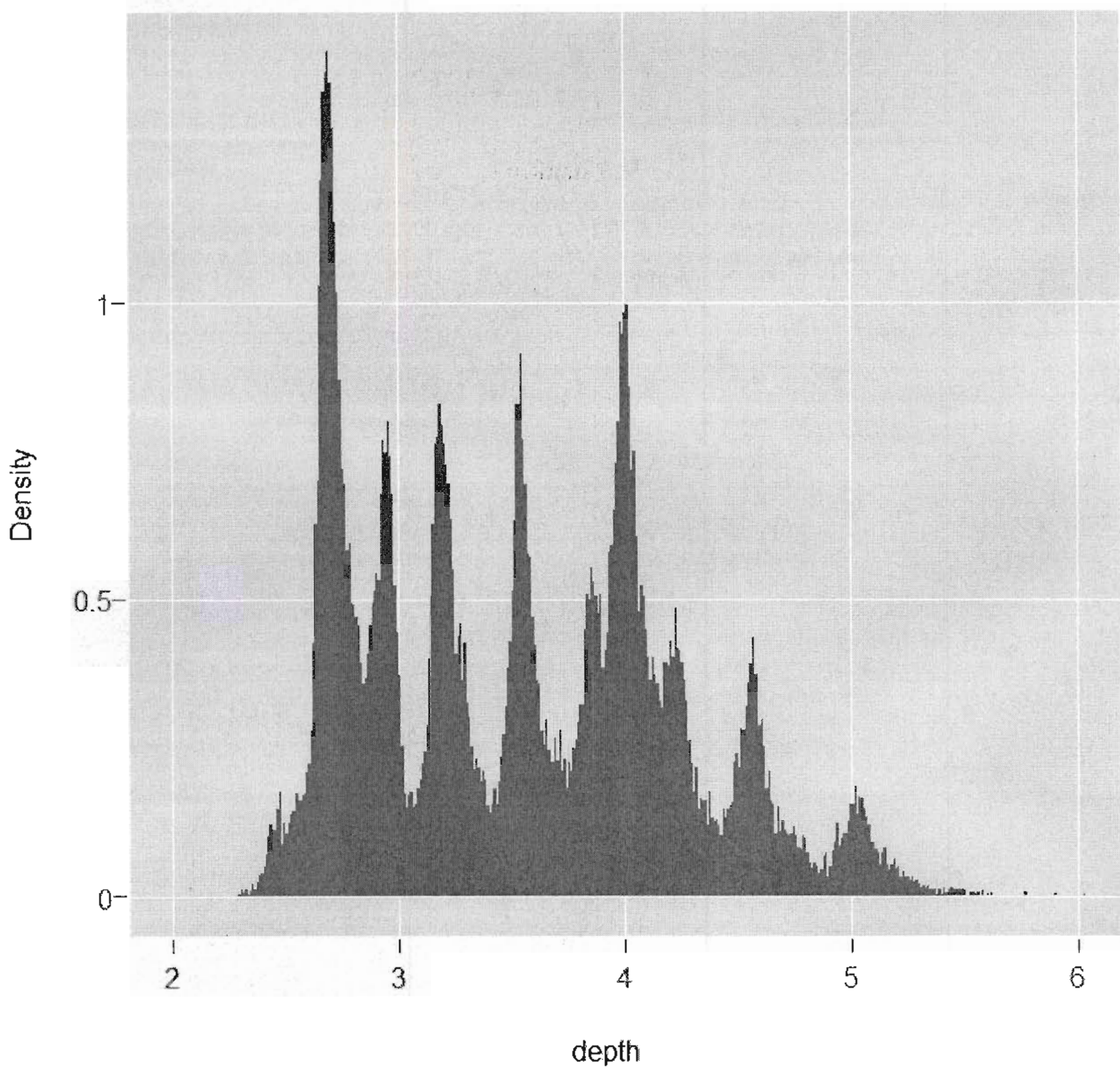
Code: qplot(price,color,data=big,type=c("jitter","boxplot"))

I found these box and whisker plots to be interesting because in each instance of color, we have an extremely small ~~mean~~ median price, with data skewed to the right as evidenced by the long whisker to the right in addition to the extreme high outliers. I thought it was odd that even with as many high outliers as these plots show, that the mean price remained relatively small. This is probably due to an abundance of small carat diamonds for each color, which are less expensive and more desirable to the normal working class individual. Another interesting quality, as the color decreases in rarity, the mean price increases, implying that fewer inexpensive diamonds of less desirable colors are produced. This makes sense because individuals wish the purchase the best quality diamond that they can afford. Ideally, this would be a small carat diamond with color D. The tradeoff between diamonds with varying colors is that you can purchase a diamond of less desirable color in a higher carat size for the same price as a diamond with rare color in a smaller carat size. Also, diamonds cut from higher colors are often cut to large carat sizes to substitute for the lack of rarity, making them more appealing to customers and thus more expensive. or is it the other way around?



Code: qplot(carat,data=big,type="histogram",breaks=(by=400),xlim=c(0,3))

The interesting trait I found with this histogram is just reaffirming what we discovered in class. Not only are diamonds with carat values just less than whole integer values not desirable, such as 0.99 carats or 1.90 carats, but this concept can also be extended to other common values as well. As you can see in the histogram, a void in the diamond data exists right below carat values of approximately, 0.50, 0.70, 0.90, 1.20, 1.50, and 1.70. This phenomenon exists because individuals more apt to lean towards purchasing a diamond just over two carats or three carats, not right below these carat values. As evidenced by the other peaks in the graph, this tendency can also be applied towards other values such as half-carat sizes.



Code: qplot(depth,data=big,type="histogram",breaks=(by=3000),xlim=c(2,6))

This last histogram demonstrates a trend similar to that in the previous histogram. Diamond depth appears to be cut to whole integer values rather than values that lie in between these whole integer values. As the graph displays, a peak exists at a diamond depth of five, four, and very close to three. Also, a peak around a depth of four and a half is present. The cause of this could be two-fold. One, individuals do not want a diamond that has a depth of 3.95 when they could have a diamond with a depth of four. Second, when diamond cutters are carving the carat size and cut, maybe measuring to a whole carat value is easier than a value in between.