Stat 480
3/1/07

Homework

1. In this reading, I learned all about the function of qplot in R and controlling the output of the graphics. I learned how to set it up using one data frame. I can then manipulate the arguments and functions as I wish. I also learned that the qplot function will change the color, shape, or size of data points on a plot, without me having to change the data, and will then provide a legend of which piece of data is represented by which point. The qplot function can make many different types of charts, including bar charts, histograms, and box and whisker plots, simply by changing the type argument. Path and line plots are two of the most common types of plots when dealing with time series data. Two data frames can be projected on one plot, using the reshape and id arguments. The only aesthetic attribute that can be changed on a bar plot is the fill color. Box plots and jittered points plot the relationship between categorical and continuous variables. Histograms and density plots show the distribution of a single variable. You can show more than one distribution on a density plot using the group argument. You can add a smooth trend line to a data plot using the smooth type and specifying the method argument. Another important aesthetic attribute for plots is the weight aesthetic; it is possible to weight a plot of the relationship between two variables by another variable. I learned that you can always combine plots by using a vector of type names. You can also facet the data, which displays many subsets of the data on one plot. Along with faceting, it is important that you choose the margins for the plot.

This tutorial was very educational in that it was very descriptive and detailed. All of the examples of the plots were very useful, as I am a visual learner and must see how something is done, rather than just read about it. Also, the fact that it was split up into sections made it less overwhelming and easier to read. I feel as though I learned a lot about how to output nicer, cleaner looking graphics just by reading this tutorial.

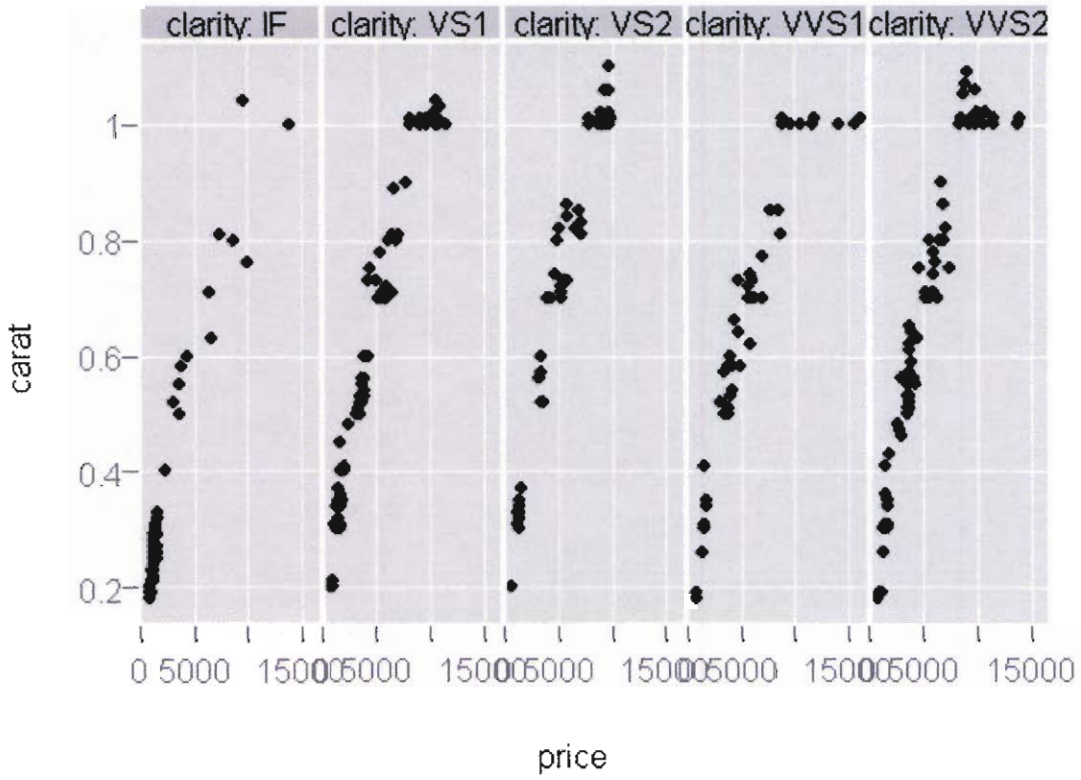
Great report,
simple clear graphics
tell an interesting story

C 5 *
5 4
0 4

2. Interesting Plots

This plot shows the relationship between price and carat, faceted by clarity. This plot is interesting because of the different clusters within each subset of clarity. As the clarity level decreases, there are larger carats produced. This is most likely because of the fact that a rare diamond of very high clarity is hard to find in mass quantities, thus, only smaller carat diamonds may be cut. ✓ good questio

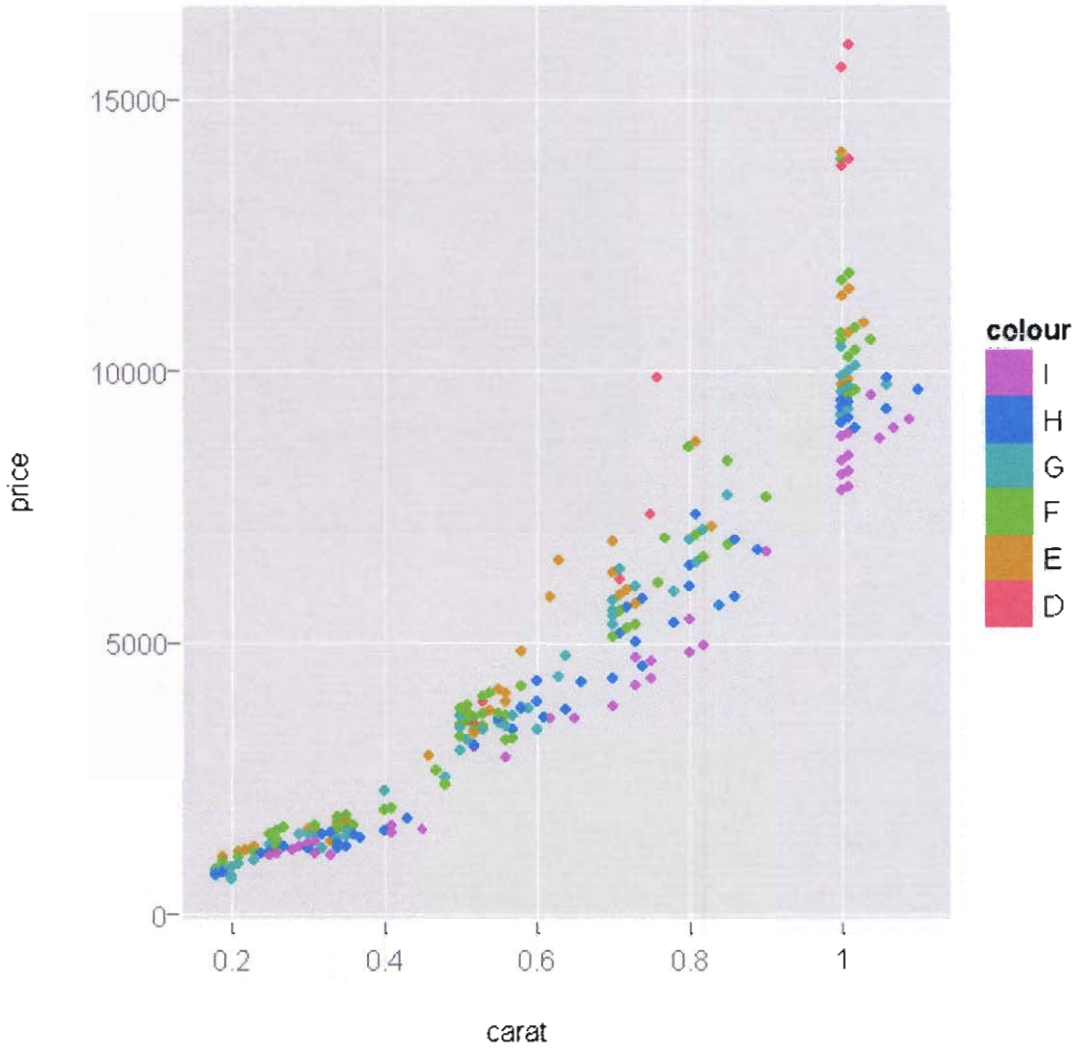
qplot(price, carat, data=diamonds, ~clarity)



This relationship shows the price per carat for each color. I thought this plot was interesting, although simple. It shows the effect that color can have on the pricing of a diamond. This type of information would be most useful to a customer. The better the color, (D→I,) the higher quality of diamond, which in turn raises the price.

great plot.

qplot(carat, price, data=diamonds, type="point", id=colour,
+ colour=colour)

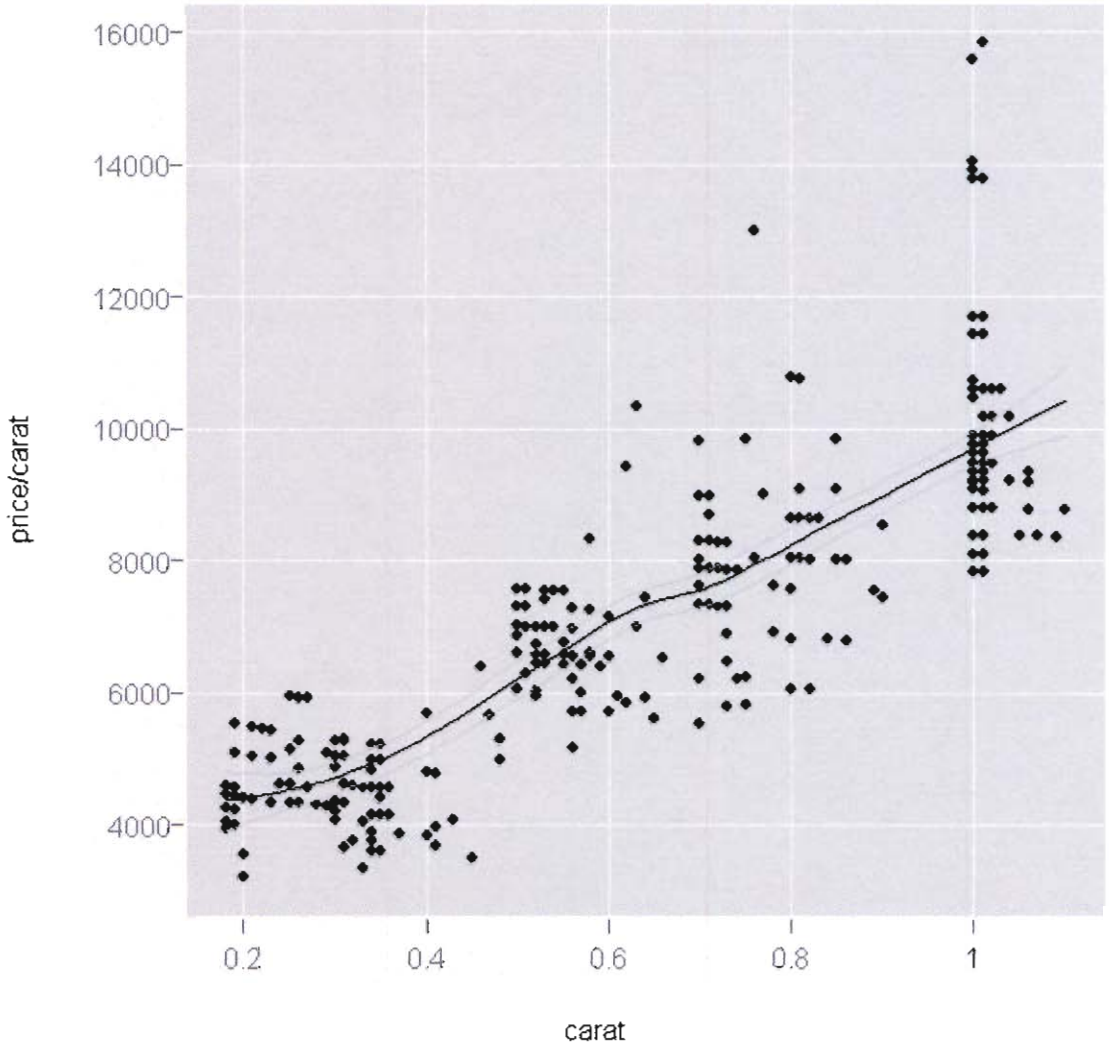


This is the relationship between carat and price per carat with a smooth best fit line displayed. There are three main clusters, which is what I thought was interesting about this plot. This shows the carat sizes at which diamonds are cut. The chance that a jeweler will cut a diamond with a carat size of .43 or .81 is highly unlikely. It is more common for a diamond to be cut at a carat size with an even fraction (1/2, 3.4, etc.). The main purpose of this plot also shows that as carat size increases, the price per carat increases, which is interesting. This could be due to the clarity of the diamond.

✓

Or just that big diamonds are worth relatively more

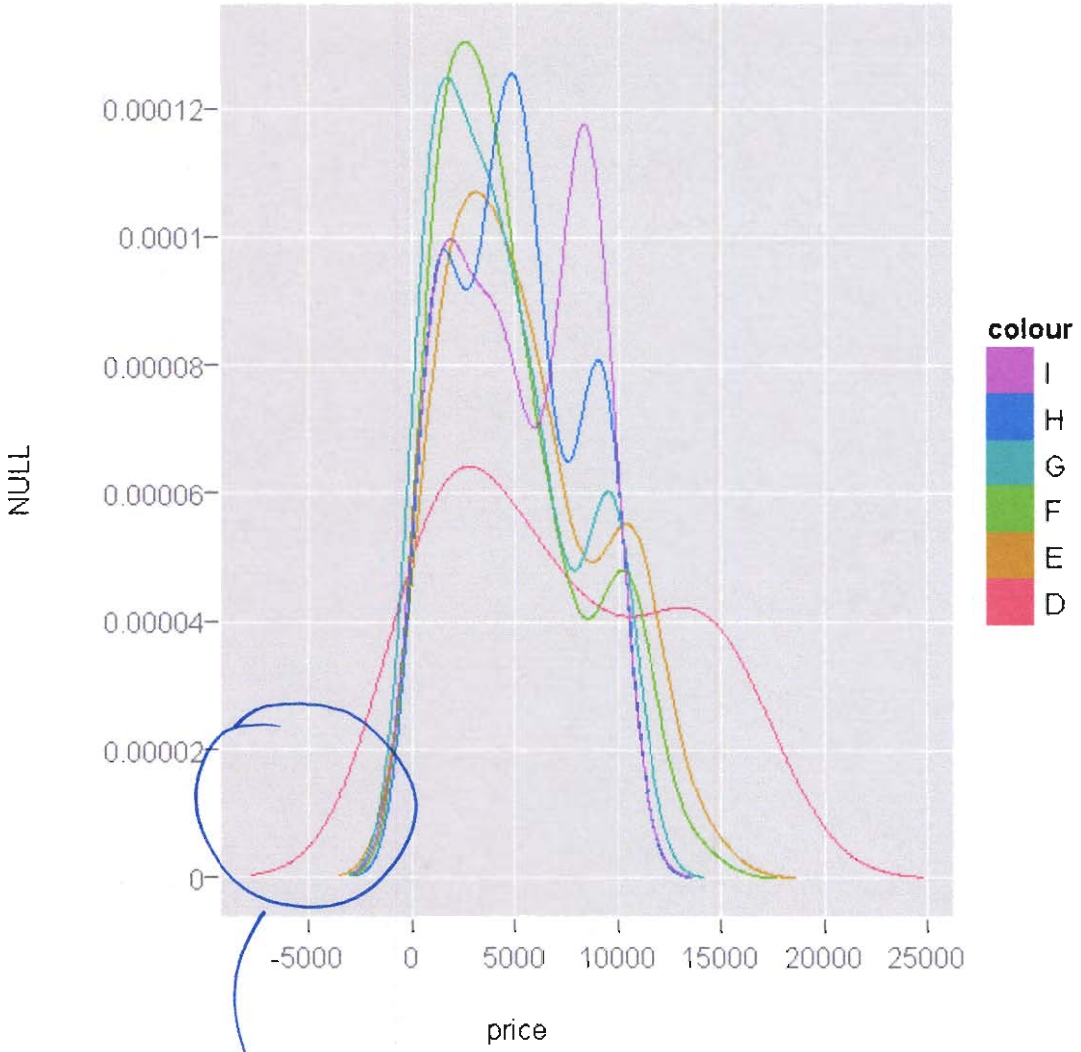
qplot(carat, price/carat, data=diamonds, type=c("point", "smooth"))



This plot shows a group of price distributions, identified by color. I liked this plot because it shows all of the data for the prices of each color on one plot. It shows the variation in price for each color, as well as an approximate mean at which the price for that color is centered around.

(median

qplot(price, data=diamonds, type="group", grob="density",
+ id=colour, colour=colour)



negative prices?

this is one reason to prefer histogram to density plots.