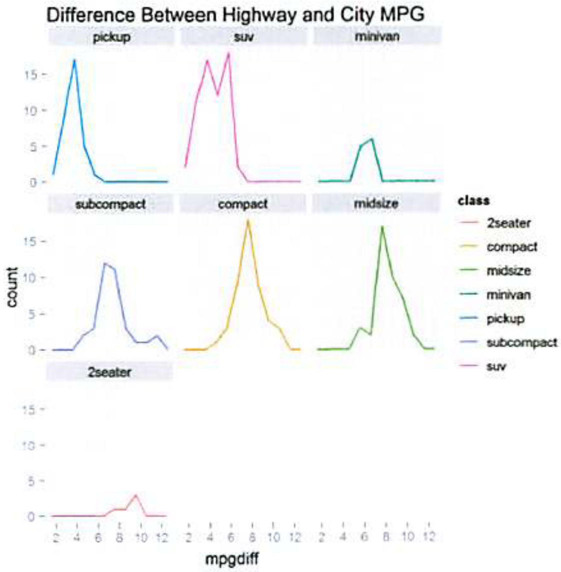
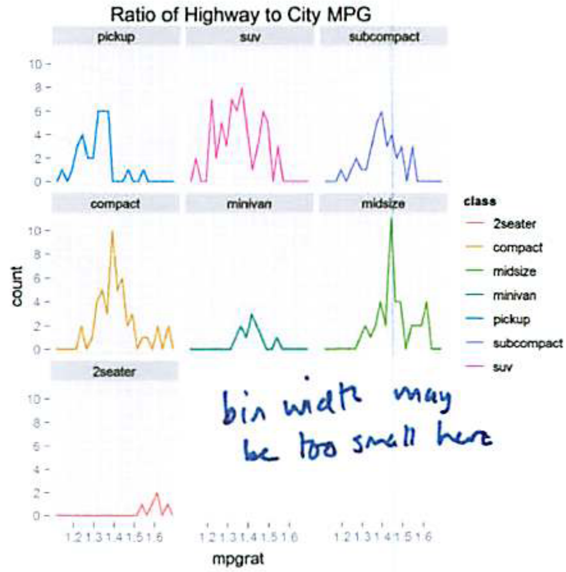
* C 4
S 4
O 4

Stat 405
Homework 1

Relationship Between Highway and City MPG





bin width may be too small here

```
>mpg$mpgdiff = mpg$hwy-mpg$cty
>mpg$mpgrat = mpg$hwy/mpg$cty
>attach(mpg)
(left)>qplot(mpgdiff,binwidth=1,color=class,geom="freqpoly",main="Difference Between Highway and City MPG")+facet_wrap(~reorder(mpg$class,mpgdiff,FUN="mean"))
(right)>qplot(mpgrat,binwidth=.025,color=class,geom="freqpoly",main="Ratio of Highway to City MPG")+facet_wrap(~reorder(mpg$class,mpgrat,FUN="mean"))
```

There's an interesting interrelationship between the class of car, and the highway and city mileage of the car. Every car is given a miles per gallon rating for both highway and city driving. The highway one is the higher of the two, as there is less braking and accelerating. I decided that it would be interesting to investigate how exactly the two were related, and what factors influenced this relationship. The results are shown above, and they are rather interesting.

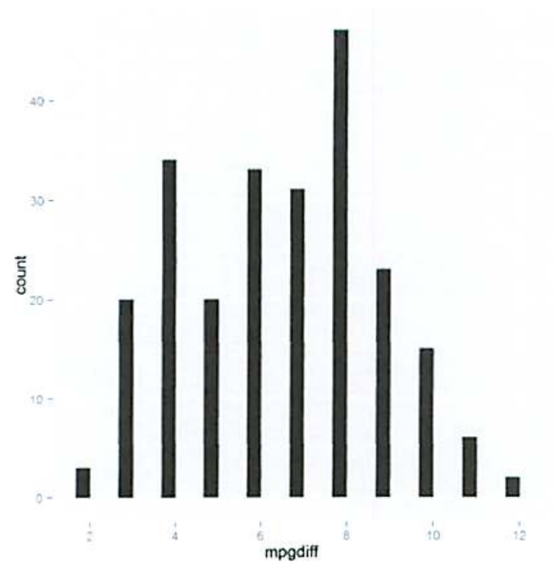
interesting finding

In order to more easily examine the two variables, I combined them into one variable by taking the difference of the highway and city mpg rating for every car. When I facetted the subsequent histogram of mpg differences by class of car, I found that smaller classes of cars, like the compact and subcompacts, had on average a higher difference between their highway and city gas mileage, which was intriguing. I thought about why this might be, and realized that it may be because the smaller cars typically have higher gas mileage to begin with. So I decided to control for the differences that may occur because of differing baseline levels by taking a ratio instead of a difference. As I'd expected, the difference didn't seem quite as big, however, it was undoubtedly still there. ✓

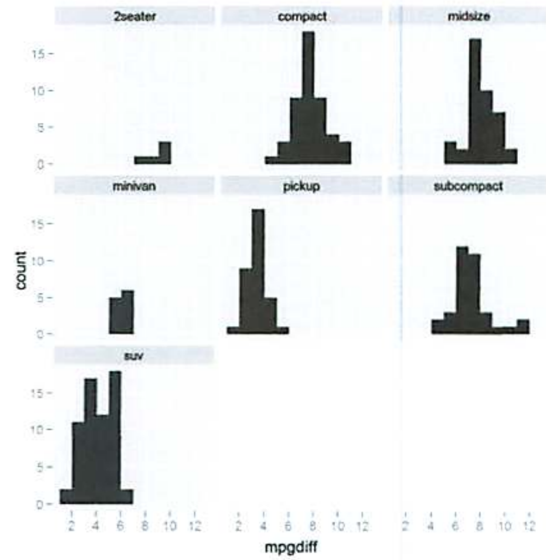
The differences in the mpg relationships of cars may have something to do with the work required by the engine to keep the vehicle moving at a constant speed on the highway. It may be that smaller cars improve more between city and highway because they are lighter, and therefore easier for the engine to keep moving once they've reach a constant speed. However, I'm not a mechanical engineer; this is pure and utter speculation.

what about qplot(hwy, hwy-cty)
qplot(hwy, hwy/cty)

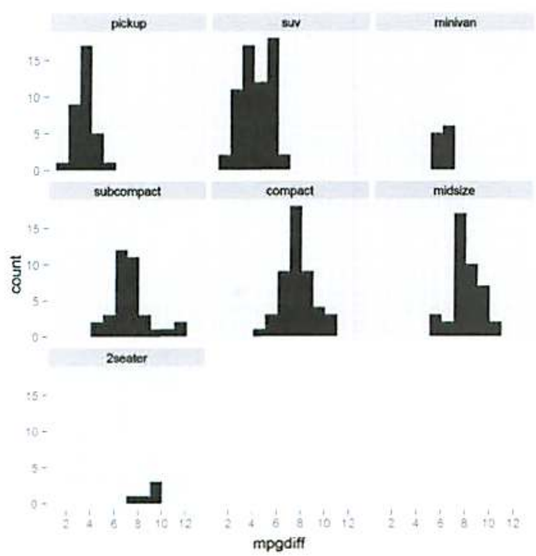
<u>Steps for Reaching the MPG comparison Plots</u>



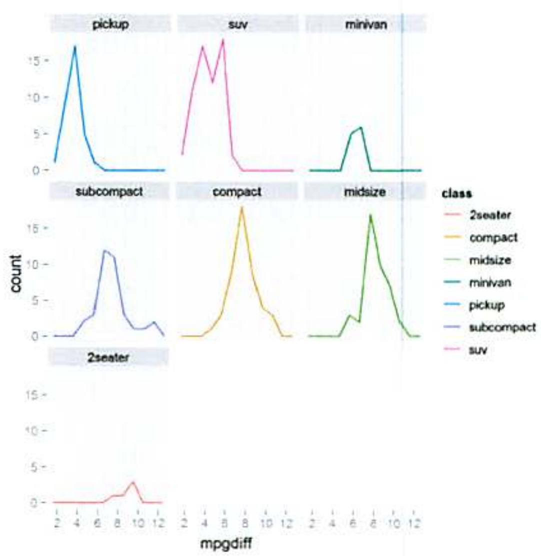
(1)



(2)



(3)



(4)

(1) After creating my new variable combining city with highway mileage I plotted it with a histogram. It was a pretty lackluster plot. The bins were unnecessarily small and it didn't show anything that I couldn't have shown with the original variables.

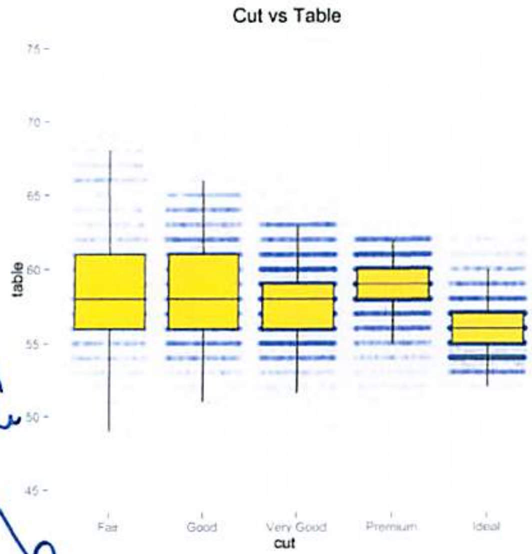
(2) With some appropriate bins and facetting, the plot became much more informative. I could see how the differences changed depending on the class of car, but it was hard to compare them because of the distance. This problem could be mitigated by ordering the classes by their mean mpg difference.

(3) The reordering made the plots easier to compare, but they still looked a bit blocky and dull, so I added color and the frequency polygon geom to make it more aesthetic and smooth.

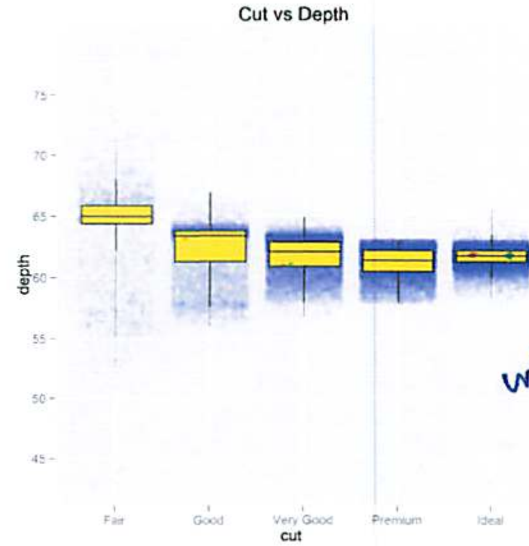
(4) The plot is pretty much complete at this stage, all I did after this was add the title.

is that a good reason.

Conformity of High Quality Diamonds



don't need to include



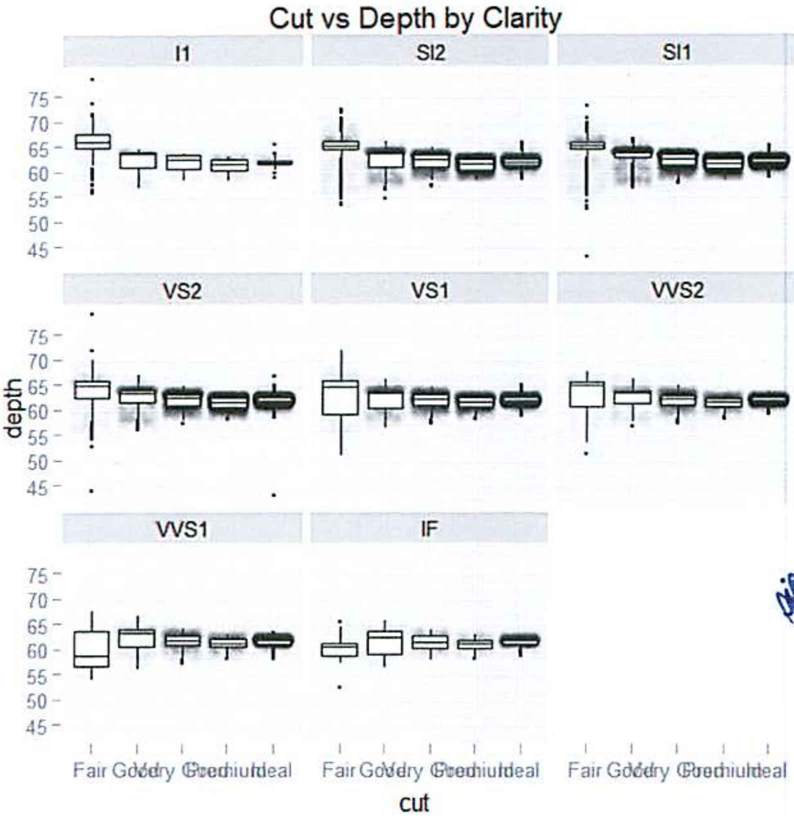
why these colours?

(right): >qplot(cut,table,geom="jitter",alpha=I(1/50),color=I("blue"),main="Cut vs Table")+geom_boxplot(outlier.size=.1,fill=I("yellow"))+ylim(45,75)
(left): >qplot(cut,depth,geom="jitter",color=I("blue"),alpha=I(1/50),main="Cut vs Depth")+geom_boxplot(outlier.size=.1,fill=I("yellow"))

This plot reveals an interesting trend towards conformity in table size as well as depth as the diamonds increase in quality of the cut. As you can see from looking at the boxplots for each quality of cut, the quartiles grow smaller, indicating a decrease in the variance of the dimensions as the cut tends towards higher qualities. Part of this can be explained by the higher frequency of diamonds of higher quality; there are far fewer diamonds of a 'fair' cut than of an 'ideal' cut. However, if that were the only contributing factor, we would expect to see some outliers for the 'ideal' cut in a similar range as the lower quality diamonds, which we don't quite see convincing evidence of.

This trend towards conformity could possibly be a result of the way the diamond market works. The average consumer will probably not be completely aware of the subtleties of diamond shopping, and will look for something of a high quality cut, that seems to be about average in size and dimension. As a result, the high quality diamonds may be geared towards the average consumer, while the more eccentric diamonds tend to be of lower quality, and may attract bargain hunters or very specific shoppers.

Reversal of Relationship Between Depth and Cut



>qplot(cut,depth,geom="jitter",alpha=I(1/50),main="Cut vs Depth by Clarity")+facet_wrap(~clarity)+geom_boxplot(outlier.size=.1)

are the jittered points necessary here?

This plot shows an interesting reversal in the relationship between depth and the quality of cut of the diamond as the clarity increases. For lower clarities, there is a clear trend towards a decrease in depth as the quality of the cut increases. However, towards the middle they grow to be about even, and when we look at the highest clarity diamonds, we can see there is a slight tendency for the depth to *increase* alongside the cut quality.

The 'Ideal' cut diamonds change the least. They remain pretty stable around a depth of just over 60. As the quality of the cut decreases, though, the depth as clarity increases decreases more and more noticeably. For instance, the 'Fair' quality diamonds drop from an average of over 65 at a clarity of I1, down to an average of below 60 for an IF clarity, while the higher quality cuts show much less prominent changes.

could this be just random variation?