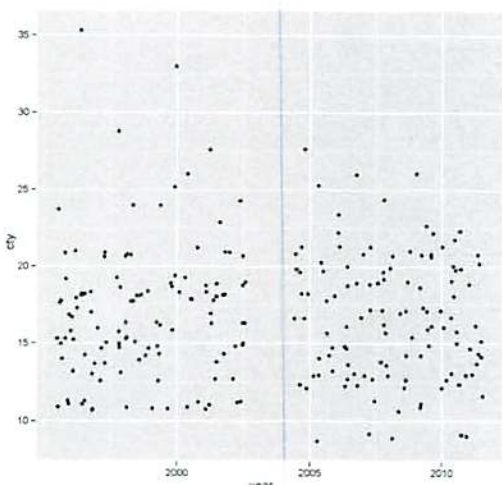
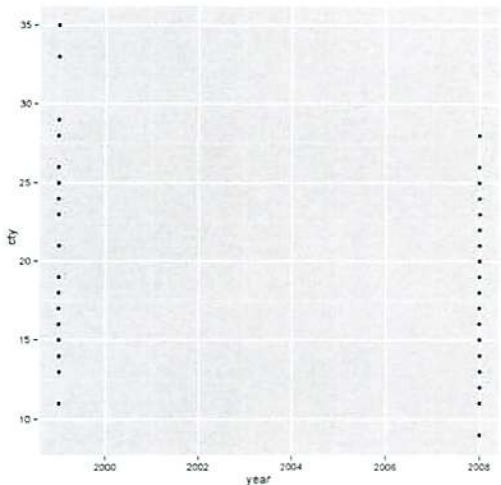


C 4
 5 4
 0 4

✓
 good use of external data

Examining the variables included in the "mpg" dataset reveals a sample limited to the years 1999 and 2008. As such, only implied trends can be established between these years. Still, given the massive increase in fuel prices over this period (nearly doubled from \$1.50, according to Department of Energy domestic averages) we would expect the city and highway fuel economy to increase. We will examine city driving trends here, as it impacts the inhabitants of the Texas Medical Center more frequently than highway economy.

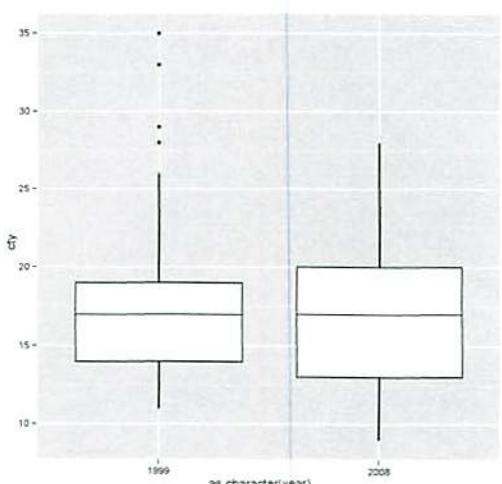
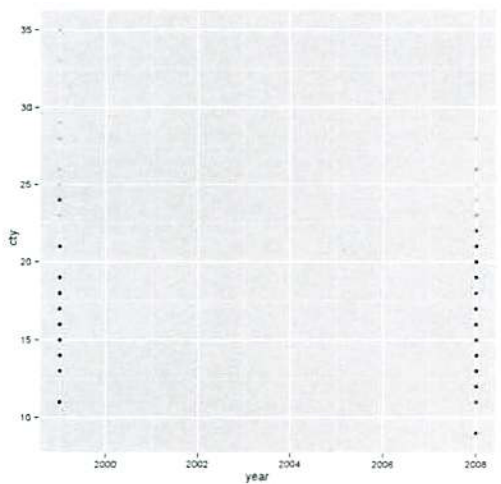


```
qplot(year, cty, data = mpg)
```

```
qplot(year, cty, data = mpg, geom="jitter")
```

As there are 234 records in this set (117 in both 1999 and 2008), it is clear that the above image suffers from overplotting.

Jitter clearly does not correct this issue as it complicates the image unreasonably. Perhaps with more refined settings this might have been more informative.



```
table(mpg$cty[mpg$year==1999])
table(mpg$cty[mpg$year==2008])
qplot(year, cty, data = mpg, alpha=1(1/21))
```

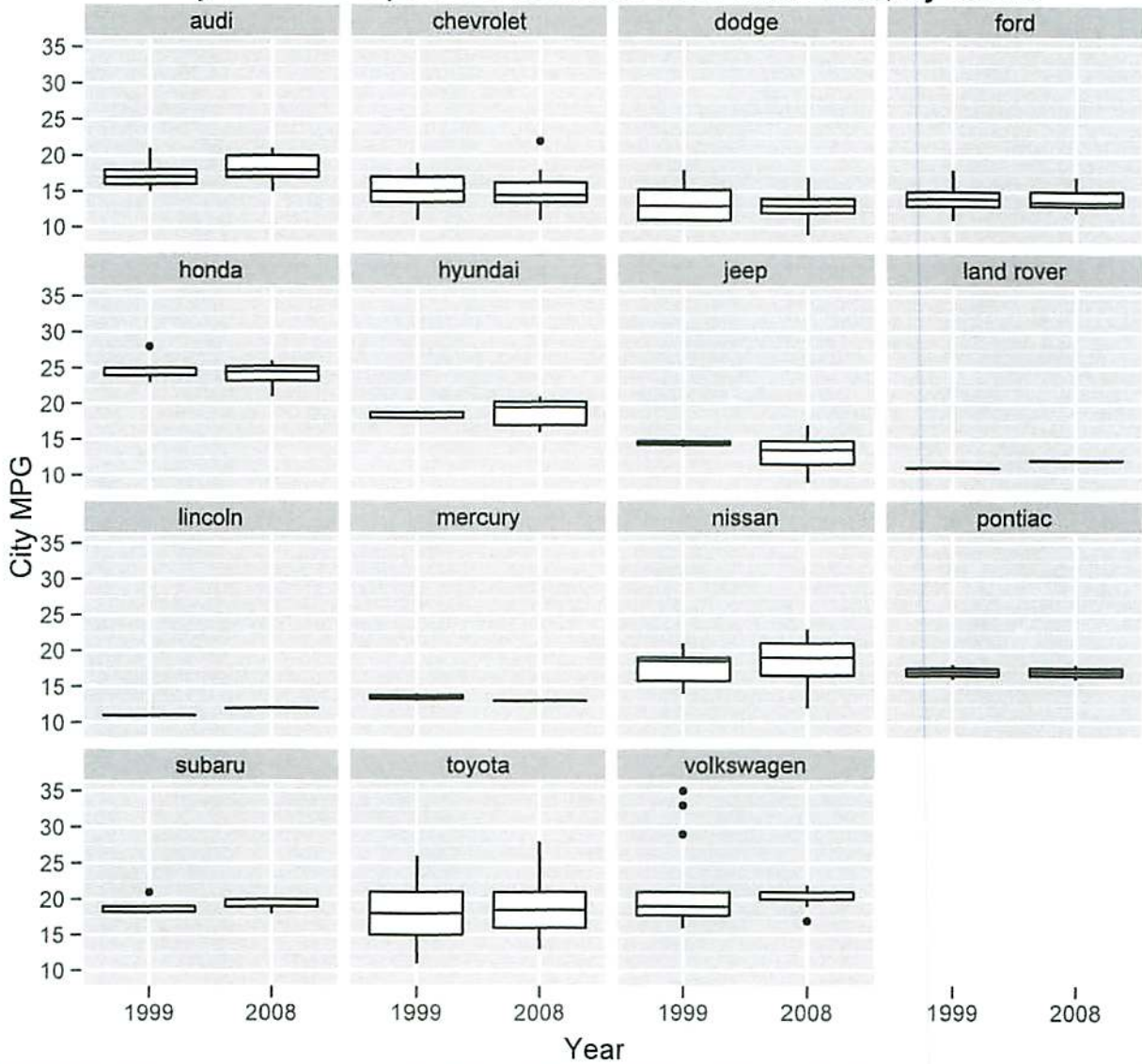
```
qplot(as.character(year), cty, data = mpg, geom="boxplot")
```

The most amount of data at any given point is 21 records. Changing the alpha to 1/21 makes the density of each point apparent, but still limits the value of the image for interpretation.

A box plot will ultimately serve us better for comparison of fuel economy between the two years. Additionally, the years between 1999 and 2008 are unnecessary and were removed. Since it is clear no real difference occurred (by median), a breakdown by brand or type of vehicle may reveal information hidden on this plot.

You have a lot of info in text - work on polishing your graphics so it's easier to see directly

City MPG Comparison between 1999 and 2008, by brand

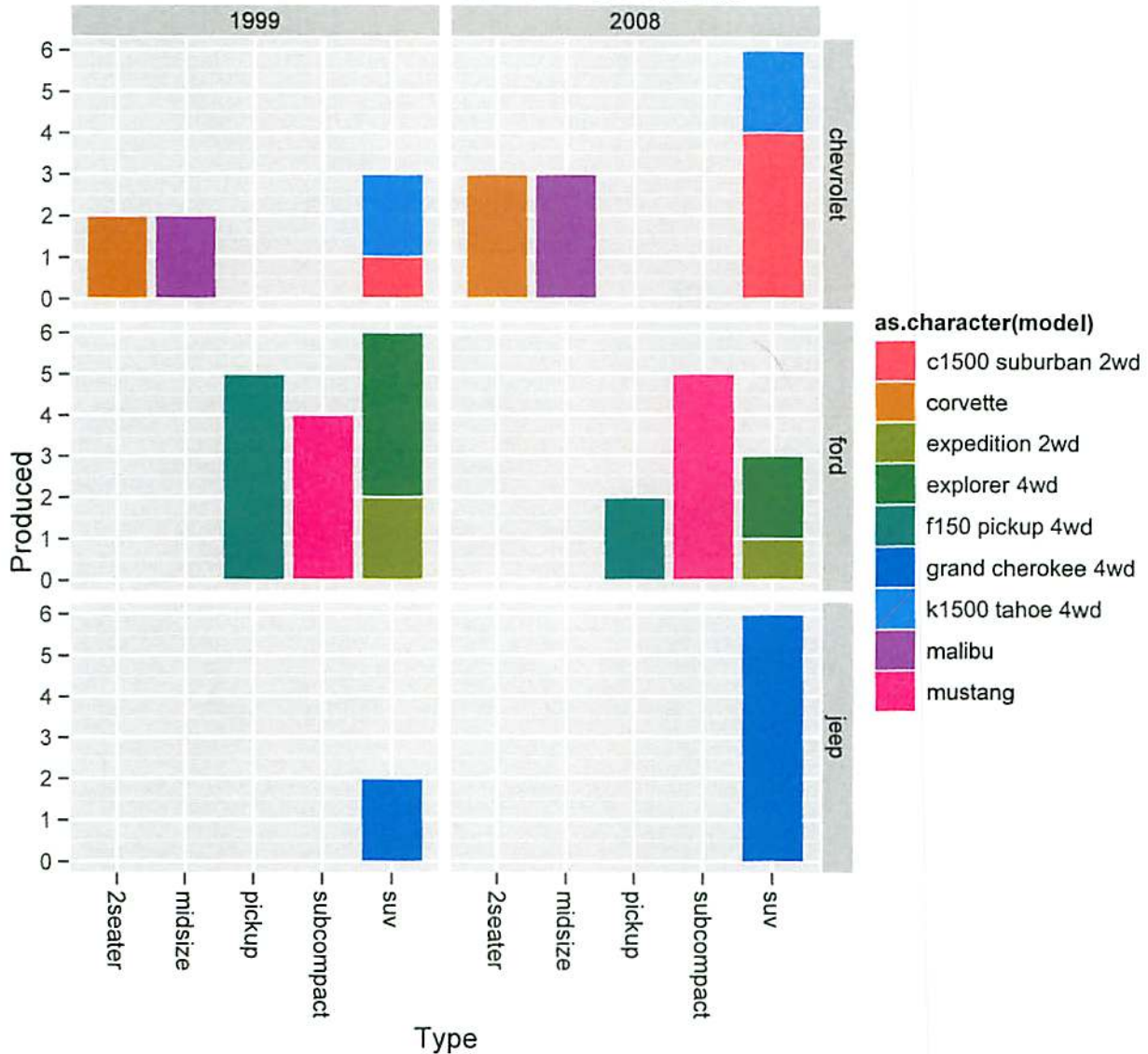


```
qplot(as.character(year), cty, data = mpg, geom="boxplot", main="City MPG Comparison between 1999 and 2008, by brand", xlab="Year", ylab="City MPG") + facet_wrap(~ manufacturer)
```

If the purpose of this graph was to compare fuel economy between brands a facet wrap would be less desirable. Since the intention is to show if brand effects the expected trend of increasing gas milage, this image is suitable. This set of box plots identifies something peculiar. For eight of the fifteen brands there was a visible increase in fuel economy as expected (though more modest than expected). However, a few brands like Chevrolet, Ford, and Jeep demonstrated decreased median fuel economy in the city. Such an observation requires an examination into the product line up of each of these manufacturers. Additionally, while Volkswagon's overall city mileage improves between 1999 and 2008, several of their very efficient models apparently are discontinued. We will examine this observation as well.



Production differences for three brands



```
mpg.filtered <- mpg[mpg$manufacturer %in% c("ford","chevrolet","jeep"),]
qplot(as.character(class), data = mpg.filtered, geom="histogram", fill=as.character(model), main="Production differences for three brands", xlab="Type", ylab="Produced", facets=manufacturer ~ year) + opts(axis.text.x=theme_text(angle=-90,hjust=0))
```

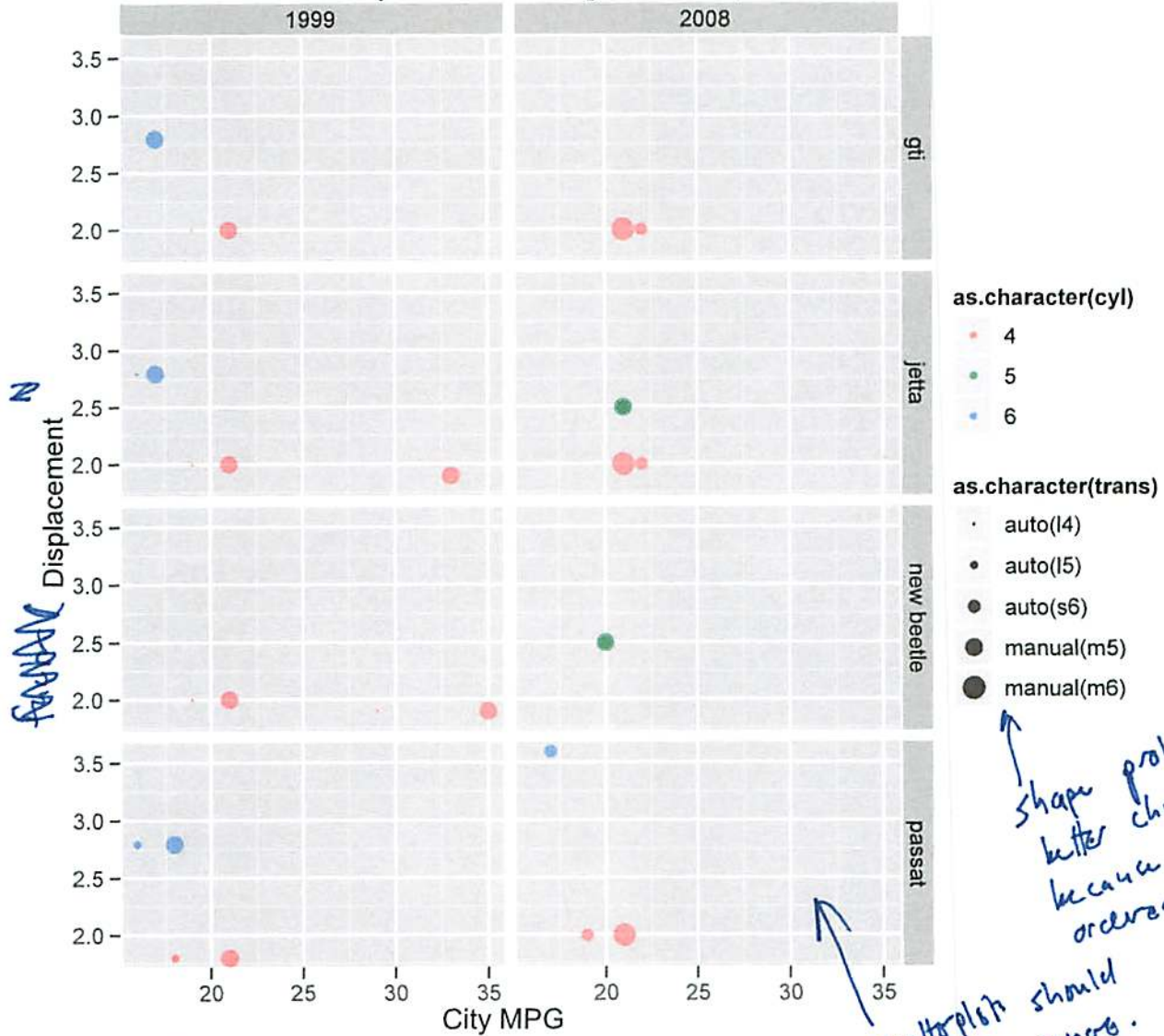
Initially the above plot was attempted using the geom "freqpoly". It is likely that the interpretation would have been easier than using facets for the years. However, as some records have identical frequencies, overplotting would have occurred.

bar chart

The histogram above clearly shows that production of models varies between 1999 and 2008 for these companies. For Chevrolet, the SUV category doubles, as do available Suburban and Corvette models (when looking at individual models), while the more modest Malibu has only one additional variant. Notably, the displacement of Chevrolet vehicles increases during the decade from 4.971 to 5.117 liters, with even the average displacement of Malibu increasing (2.75 to 3.16 liters). This likely accounts for the decrease in city fuel economy for Chevrolet, though admittedly this sample may not include all models for all car makers. Ford shows an increase in the subcompact market, with their Mustang, along with a decrease in their pickup and SUV offerings. Unfortunately for the environment, these new Mustangs have larger engines (4.2 to 4.52 L) and worse fuel economy (city 16.5 to 15.4 mpg), impacting the median of the Ford city miles per gallon negatively (seen in previous image). Finally, according to this limited dataset, Jeep released twice as many models of their SUV, the Grand Cherokee. Displacement goes from 4.35 to 4.65 L, available gears and cylinders increase to match, and city fuel usage from 14.5 to 13.17 mpg. As this is the only model in this dataset for Jeep, the Cherokee accounts totally for the drop in fuel economy seen. The trends seen here (larger diversity in SUVs, or greater engine power in smaller vehicles) are largely inverted in the other brands, where subcompacts and compacts improve fuel economy (not shown).

How could you display this graphically? faceting on model would be easier.

Vehicle Lineup for Volkswagen: 1999, 2008



```
mpg.volk <- mpg[mpg$manufacturer=="volkswagen",]
mpg.volk[order(mpg.volk$year),]
qqplot(cty, displ, data=mpg.volk, alpha=1/2, colour=as.character(cyl),size=as.character(trans),main="Vehicle Lineup for Volkswagen: 1999, 2008", xlab="City MPG", ylab="Displacement",facets=model~year)
```

Scatterplot should be square. shape probably better choice because not ordered

This image attempts to collate all the relevant data for the midsized Passat, subcompact Beetle, and the compacts Jetta and GTI. From the original series of plots it was apparent that at least 3 models produced by Volkswagen had above 28 city miles per gallon. These three models were later unavailable in the 2008 dataset. Looking at the plot above, the three models were the 1999 4 Cylinder 4 Gear Auto Beetle, the 1999 4 Cylinder 5 Gear Manual Beetle, and the 1999 4 Cylinder 5 Gear Manual Jetta. The fuel economy of these small, low displacement vehicles is profound for 1999, and requires explanation. Further inquiry into the dataset reveals that these models were diesel powered. If we are to take this dataset as complete (which it likely isn't), Volkswagen opted to discontinue these green vehicles and instead release less fuel efficient models in 2008. The only two Beetle offerings in 2008 had 31% bigger displacement, 5 cylinders instead of 4, and either a manual 5 or auto 6 gear transmission. Neither were diesel. Overall the city mpg measure dropped from an average of 26 to 20 for this vehicle model. For the Jetta, two 4 cylinder models remained in 2008 where there was once three. The average city fuel economy dropped for this subgroup from 24.33 to 21.5 mpg, despite the average displacement remaining fairly constant at 1.967 and 2.0 for 1999 and 2008, respectively. Again, the diesel fuel type appears to be the active factor that generated the Jetta outlier in 1999. Despite the loss of diesel powered vehicles, Volkswagen discontinued many of the less efficient 6 cylinder models (blue) in 2008. This helped push it's brand-wide fuel economy up despite losing it's most efficient vehicles from 1999.

not sure why they're not included may be not re-tested by the EPA.