The Effects of Technology on the Guzzler Tax and the Fuel Economy of American, German, and Japanese Cars

Nivriti Chowdhry, Matt Delhey, Jiandi Mo

September 25, 2012

Contents

1 Introduction

The primary dataset used in this exploration is the ”mpg2” dataset. The data was first used to explore the effect of fuel type on gas mileage to determine whether premium fuel results in higher gas mileage and is worth a higher price. While on average the use of premium fuel increases the engine displacement of a car and reduces gas mileage, if, however, the number of cylinders in the engine is held constant, then premium fuel has a higher gas mileage than regular fuel.

Second we looked at the Gas Guzzler Tax and its effect on the lifetime cost of operating a ”guzzler” versus a larger car with a gas mileage of less than 22.5 miles per gallon (MPG). We found that the added tax on guzzlers makes them around 5,000 dollars more expensive than larger, low gas mileage cars over a lifetime of 100,000 miles. This could work against the EPA’s goals of reducing the use of cars with poor fuel economies as it might encourage people to purchase less expensive and bigger cars.

Third, an analysis of American car manufacturers is conducted to determine which manufacturer out of the Big Three manufactuerers, that is, Chrysler, Ford, and GM has the greenest cars based on combined gas mileage and averages air pollution and greenhouse gas emissions. In order to make this comparison regarding air pollution and greenhouse gas emissions, the Green Vehicle Guide data set from the EPA had to be utilized. Out of the Big Three we found that among consumer-level engine sizes General Motors tended to have somewhat greener vehicles while at performance-level engine sizes Chrysler seemed to be the greenest producer.

Lastly, an analysis of German and Japanese manufacturers was conducted to determine which country produces cars with better fuel economies. Japanese cars were found to have a higher gas mileage for all cars except those with 6-cylinder engines, which are the most popular for consumer-level vehicles, and those with higher engine displacements, especially over 4 liters.

2 The Effect of Fuel Type on Gas Mileage

Fuels are denoted as being different types based on their Motor Octane Number (MON), which is the number displayed on gas station fuel pumps. Manufacturers specify which fuel type is to be used based on the requirements of each car's engine. Higher MON fuels are typically used in vehicles with high-compression engines, but are sometimes also said to have higher fuel economies.

An exploration into the differences in combined gas mileage with different fuel types on a fixed variable, like the number of cylinders in a car's engine, allows a better understanding of whether the claim of improved fuel economy is true. To produce this study, the data first had to be cleaned. The subsets of fuel types to be combined into four final categories: regular, premium, diesel, NA, or other. Regular fuels are those with a MON of 82 or 83, premium fuels are those with a MON of 87 or 88, and diesel is made of cetane with a rating of 40. A close look at the differences in the fuel economies of vehicles with the same number of cylinders might help us show whether fuel type truly makes a difference on fuel economy.

Different vehicle classes still contain similar engine structures and cylinders, which is why the data is faceted based on the number of cylinders in the engine. Large and performance vehicles, like trucks and sports cars respectively, have larger engines with more cylinders than other vehicle classes. However, before deciding to look at the effect of fuel type with respect to cylinders, the effects of fuel type based on a vehicle's status as a guzzler, engine displacement, and drive type were also analyzed.

Comparing only cars reported as guzzlers and the effect of fuel type on combined gas mileage showed us that there is not a noticeable difference in the gas mileage with premium fuels. However, people still pay more for premium gas and it is typically required in high-performance cars. This raised the question of whether high-performance cars have engines that simply require higher fuel grades based on high compression rather than for better fuel economy. Including other vehicles like pickup trucks and vans that are not reported as guzzlers but have similarly low combined gas mileage showed us that cars using premium fuels tend to have a lower gas mileage than those using regular fuel.

Since the effect of fuel grade on the fuel economy of a car could not be determined based solely on a car's status as a guzzler or not, we were still left with our original question regarding the reason why fuel type affects gas mileage. The relationship between fuel type and engine displacement shows that cars using premium fuel tend to have higher engine displacement than cars using regular or diesel fuels. This is probably because engines requiring premium fuel are more powerful. However, the true reason could also include a variety of other factors that cannot be pin-pointed with the given data, such as the technology the manufacturer employed or a driver's driving style. Displacement is negatively correlated with gas mileage, but this analysis shows no differnce between the combined gas mileage of premium and regular fuels where one would expect premium fuels with higher displacements to have lower gas mileage.

The relationship between fuel type and drive type could possibly provide insight into where the differences in fuel types truly manifest. The effect of drive type on the combined gas mileage faceted based on fuel type shows that combined gas mileage does not vary over different drive types. However, premium gas tends to produce lower gas mileage than regular.

Since the effect of fuel type on displacement was more noticeable, the concept of the engine being where fuel type differences manifested was revisited, but looked at from the perspective of cylinders. This allowed an understanding of how gas mileage is affected when different fuel types are put into similar engines. If

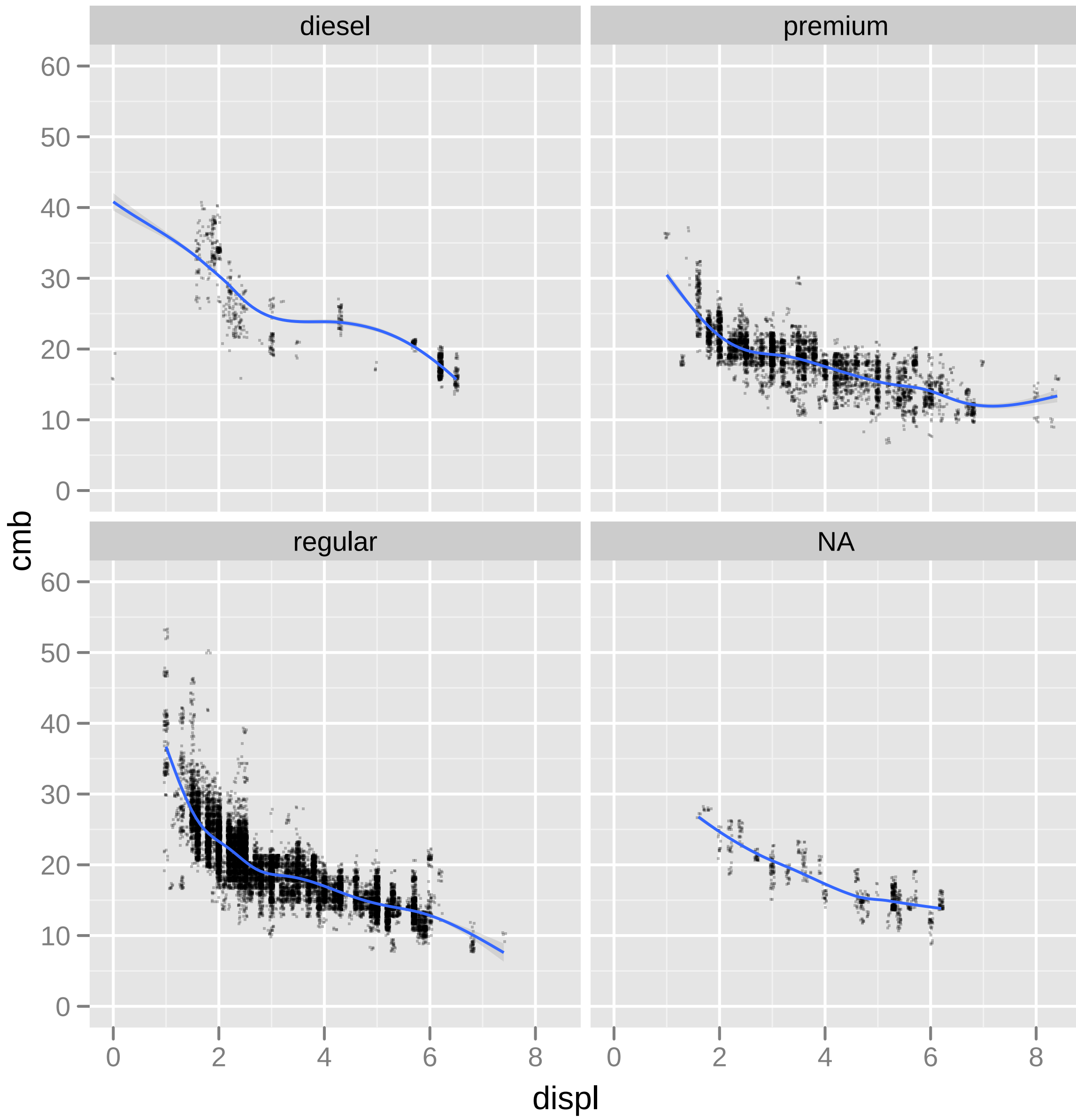


Figure 1: A scatterplot of engine displacement vs. combined gas mileage. The graphs are separated by fuel type. The blue line shows the mean combined gas mileage for the respective engine.

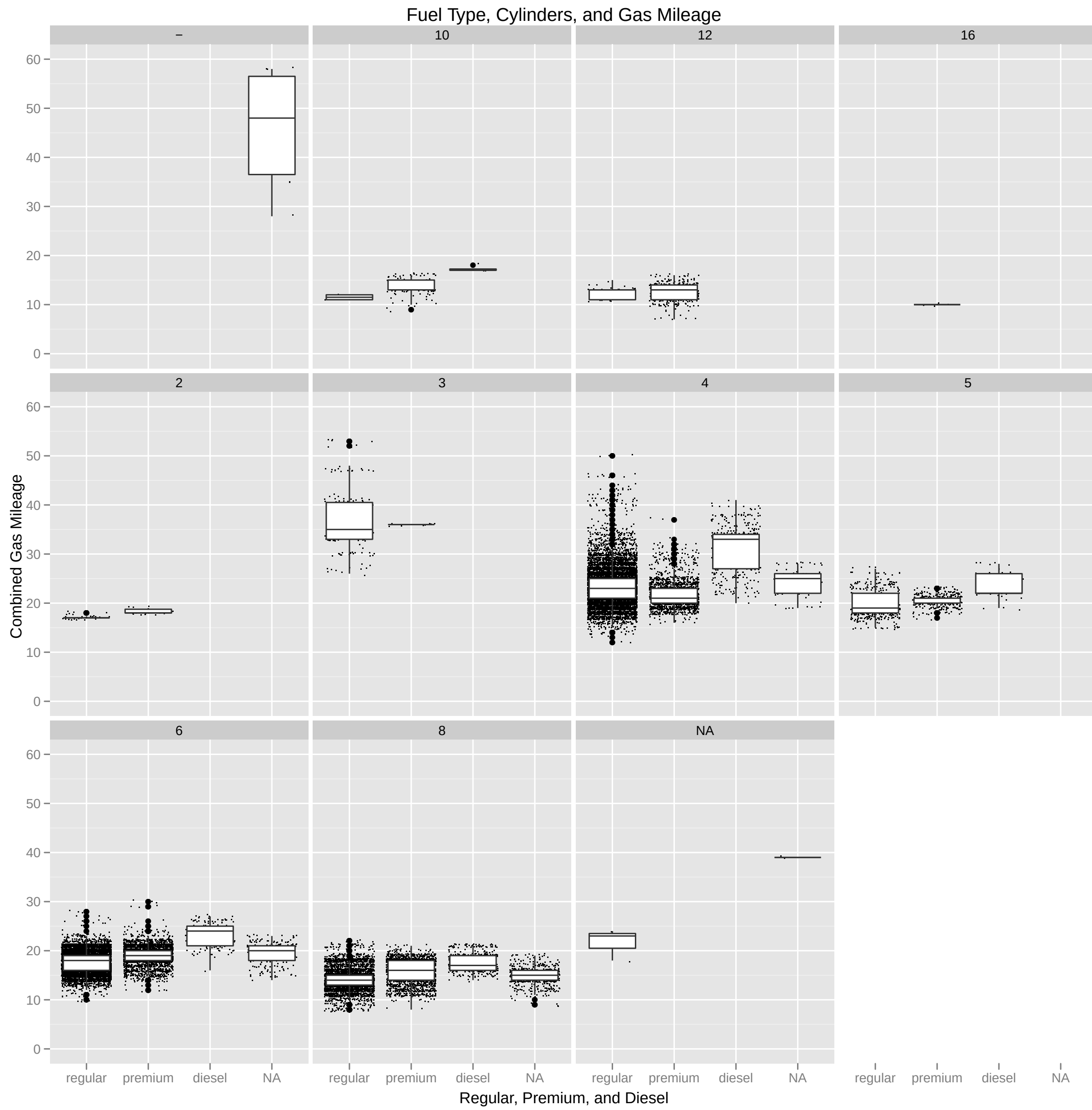


Figure 2: Boxplots of fuel type vs. combined gas mileage. The graphs are separated by the number of cylinders in the engine. When holding the cylinders constant, premium fuel has a higher combined gas mileage than regular fuel, but lower than diesel.

the cylinder is considered the primary characteristic of an engine, sorting based on the number of cylinders within the engine allows an analysis of different kinds of engine to find that premium fuels produce a higher gas mileage than regular fuel types as long as the number of cylinders is held constant.

3 The Guzzler Tax

While cars reported as guzzlers and others with low fuel economy outside of the guzzler category did not show differences in combined gas mileage based on fuel type, they did raise an interesting question as to what other costs are associated with owning a gas guzzler. One point to emphasize in this analysis is that not all cars with a combined gas mileage below 22.5 MPG are considered guzzlers. ”Working vehicles” like trucks, SUVs, SPVs, and minivans are not considered guzzlers despite their low gas mileage.

The classification of vehicles as guzzlers originates from the establishment of the Gas Guzzler Tax as part of the Energy Act of 1978 with the purpose of reducing the number of cars with poor fuel economies on the road. Under this tax, cars with a combined gas mileage of 22.5 MPG or lower were subject to a tax upon purchase. The tax amount is negatively related to a vehicle's gas mileage. The tax amount can be considerable, reaching up to 7,700 dollars at some of the lowest reported combined gas mileages. This raises an interesting question as to why consumers purchase vehicles that are subject to this extra tax considering they will already have to purchase more fuel over the lifetime of the car given its poor fuel economy. This could be because customers are not fully aware that they are being subject to a tax if it is already a component of a car's list price.

The distribution of vehicles that should recieve the tax can be found in the following table. Notice that a large portion of the vehicles do not receive any tax. These are the vehicles that should have received a tax based on their fuel economy but did not because of their vehicle class.

Tax	Count
0	980
1000	1
2600	5
3000	94
3700	230
4500	140
5400	130
6400	114
7700	165

An analysis of the cost of owning any reported gas guzzler or other vehicle with a combined gas mileage of less than 22.5 MPG can help us better understand whether the Guzzler Tax is an effective measure toward preventing people from burning excessive amounts of fuel. The first step towards this analysis was to create a new variable that reported the guzzler tax amount for each reported guzzler based on the combined gas mileage. Next, these taxable guzzlers were put into one subset while all the other vehicles with a combined gas mileage lower than 22.5 MPG were put into another. General vehicles classes were still separated within these subsets.

The formula to analyze the lifetime cost of any of these subsetted vehicles is:

```
100,000 miles * (3.85 dollars / mpg2$cmb) + mpg2$tax
```

The 100,000 miles represents the average lifetime of a vehicle in the United States as reported by the EPA. 3.85 dollars is the average gallon price of fuel over the past 20 years, while cmb is the combined gas mileage specific to that entry and tax is the guzzler tax on the vehicle, if there is one.

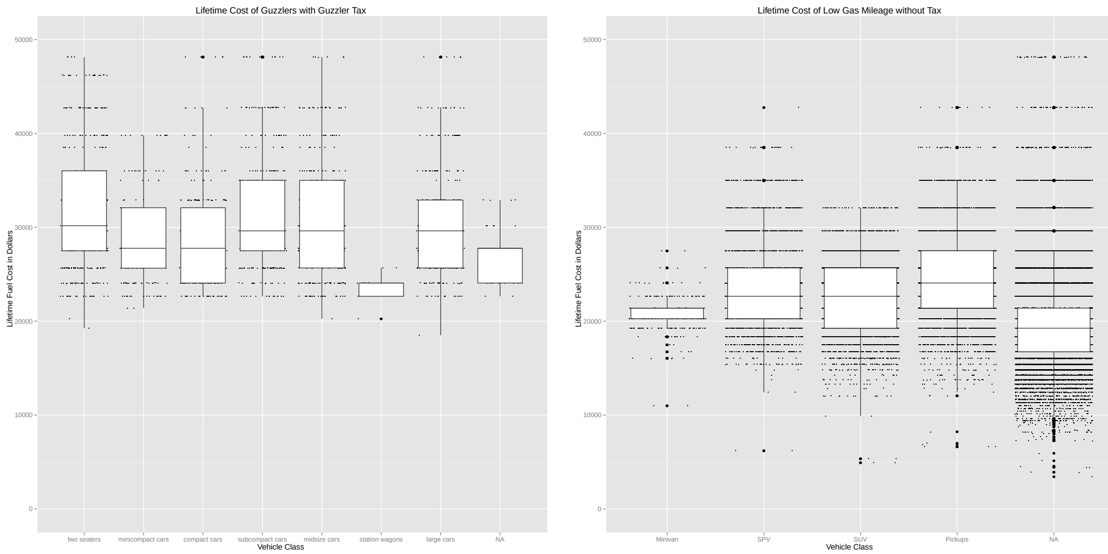


Figure 3: (Left) Boxplots of vehicle class vs. lifetime fuel cost in dollars for vehicles classified as guzzlers. (Right) Boxplots of vehicle class vs. lifetime fuel cost in dollars for vehicles with a combined gas mileage less than 22.5 MPG, but not reported as guzzlers. The side-by-side comparison shows that larger cars have a lower lifetime cost than cars actually classified as guzzlers.

A side by side comparison of the two graphs shows that even larger cars like pickups, SUVs, SPVs, and minivans end up costing about 5,000 dollars less over their lifetime than smaller cars that have a guzzler tax on them. While the tax was initiated to discourage people from buying cars with poor fuel economies, the tax could instead be pushing consumers towards purchasing larger cars that may have even worse repercussions for the environment. One way to get a better understanding of how the Guzzler Tax affects the environment could be to look at the difference in pollution due to cars reported as guzzlers and large, low gas mileage cars. It is likely that larger cars cause more pollution and that the true higher lifetime cost for guzzlers is encourages consumers to buy larger cars than they otherwise would. The Gas Guzzler Tax might be working against its mission to reduce pollution, because it is instead pushing people towards purchasing larger vehicles that emit more pollution.

4 American Vehicles (Big Three)

4.1 Introduction

The Big Three are the three largest American automobile manufacturers, Ford, General Motors, and Chrysler. While these three manufacturers have dominated American automobile sales since the beginning of the twentieth century, in the last forty years they have seen increasing competition from foreign manufacturers. In particular, foreign manufacturers have gained a reputation for producing more fuel-efficient and greener cars. Despite these complaints, the Big Three have remained very popular in the American market and, more recently, have also begun to combat the foreign manufactures by putting a greater emphasis on fuel economy. Coupled with the increasing demand for more environmentally friendly consumer automobiles, the Big Three have made a conscious effort to increase both the fuel economy and the greenness of their vehicles over the last ten years. But who has done a better job? Or, in other words, which manufacturer of the Big Three over the last ten years has been producing the greenest cars?

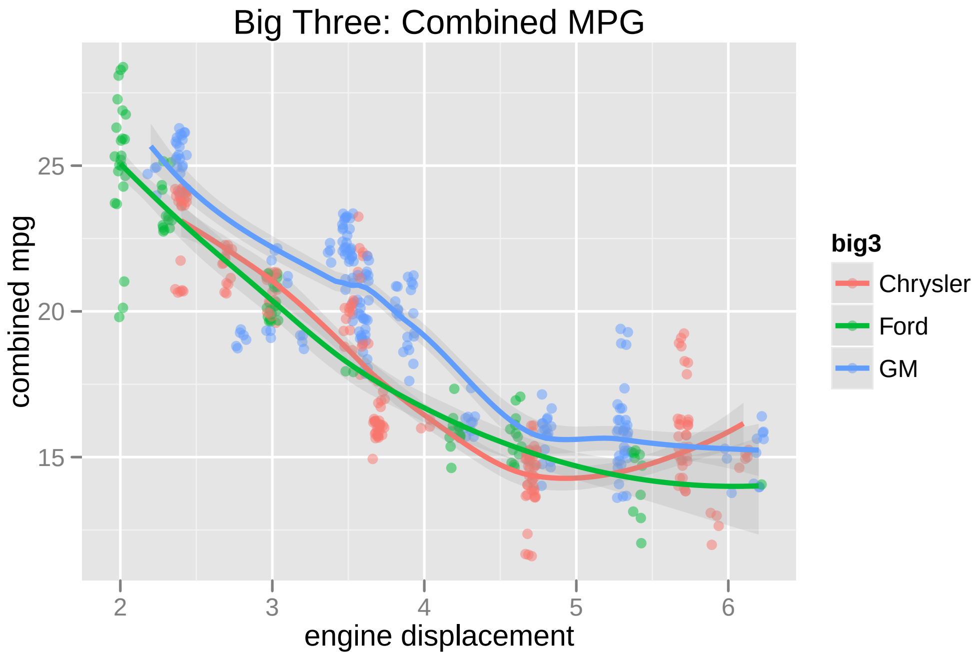
The first complication to the question is what do we mean by green? The term, as used in common language, is hard to define. However, in our analysis we answered the question in two ways, for which we said that a green vehicle will both (1) have a high combined gas mileage and (2) produce only a small amount of air pollution and greenhouse gasses, both *with respect to the power of its engine*. While the data for the analysis of (1) was available within the original data set, the analysis of (2) required more external data. The data used to get information about the air pollution and the greenhouse gasses came from the Green Vehicle Guide created by the EPA. In this guide, the EPA assigns both a Greenhouse Gas Score and an Air Pollution Score to all car models sold in the United States each year. These scores both range the integers from 0 to 10, with 10 being the greenest possible score. The Greenhouse Gas Score reflects the emissions of carbon dioxide and other greenhouse gases and the Air Pollution Score reflects the emissions of hazardous chemicals responsible for regional air pollution, smog, haze, and health issues. In order to simplify the analysis, we took the mean of the two scores to create a ”green” score, working off the assumption that greenhouse gas emissions and pollutant emissions are equally important in assessing the greenness of a vehicle.

4.2 Procedure

In order to do a practical analysis of the Big Three manufacturers and produce meaningful results, a significant amount of subsetting of the data was required. The first subset was created to only look at vehicles produced by the Big Three and their divisions and marquees. Additionally, a new variable was created to denote which of the Big Three companies each car was made by even if they were officially released under a division or marquee. The second subset, as mentioned previously, was to only look at vehicles manufactured in the last ten years, or from 2002 to the present, in order to put emphasis on the recent trends in the Big Three. Finally, in order to deal with the variation in car models across different manufacturers, only the flagship models for the popular vehicle classes of the Big Three were chosen. In choosing the popular vehicle classes, we tried to strike a balance between diversity and consumer popularity and decided upon: standard pickup trucks, sport utility vehicles, large cars, mid-size cars, compact cars, and entry-level mid-size luxury cars.

At this point we had to determine how we were going to compare the Big Three on their production of green cars. It is not a good comparison to simply look at the average green score or combined gas mileage across the different manufacturers because this ignores a very important component of each car, namely the power of its engine. This is where the final clause of our definition of a green vehicle comes into play, namely that a car must have a higher combined gas mileage and a high green score *with respect to the power of the engine*. If we ignore this clause, then the manufacturer with the greenest vehicles would simply be the one who produced the greatest amount of small automobiles and the fewest amount of large automobiles, which is neither interesting nor informative. Instead, we want to look at the manufacturer that produced the most fuel-efficient and greenest vehicles with respect to the ability of the engine. Manufacturers will always produce small and large size cars and the size of the car is the biggest factor in determining its combined gas mileage and green score, but this does not tell us anything about the ability of the manufacturer to create environmentally friendly vehicles relative to its peers. In a sense, we wish to hold the power of the engine constant and then look at the combined gas mileage and green score.

In order to accomplish this task, we needed to look at the engine displacement for each car in addition to its combined gas mileage and green score. By plotting combined miles per gallon and green score against engine displacement, we can analyze the ability of each of the Big Three manufacturers to create fuel-efficient and green vehicles at each amount of engine power and thus answer our original question. The caveat here is that the engine displacement variable may not fully capture engine power in the way that we want it to. While this is a fair complaint, engine displacement is a standard method for measuring engine power in the industry and should be accurate enough for our purposes.



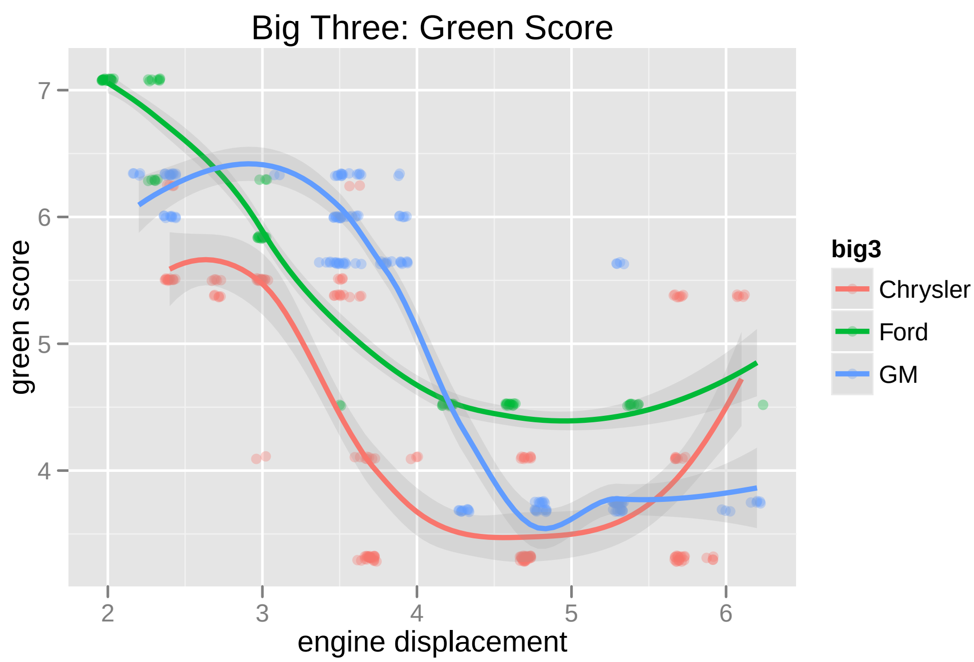


Figure 4: (Top) A scatterplot of engine displacement vs. combined mpg with the Big Three manufacturers layered on top. All data seems to follow a similar pattern. (Bottom) The same as above, except with green score instead of combined mpg. Green scores seem to be condensed around specific clusters with erratic variation among the clusters.

4.3 Analysis

Figure 4 contains both the combined gas mileage plot and the green score plot. From the combined gas mileage plot, we can see that General Motors has the highest fuel economy for the majority of engine displacements. At the upper echelon of engine sizes, Chrysler beats out General Motors by a small margin.

One interesting facet of both plots is the range of Ford vehicles around engine displacements of two liters. From exploration of the data, we know that these values are mostly the result of the Ford Focus. This car model has one of the highest combined gas mileage and green scores out of any compact car and is also very popular among its class. We might be able to explain some of the discrepancies seen in the combined miles per gallon plot by claiming that the popularity of the car has led Ford to manufacture high performance, non-green versions of the car to be sold alongside the much more environmentally friendly models.

In the green score plot, we see a more competitive picture between the manufactures. General Motors only has the highest green score between engine displacements around three and four liters. Outside of this range, Ford seems to have the highest green score. Once again Chrysler is most competitive at the higher echelon of engine sizes and possibly the greenest at the engine displacements above six liters. Thus we can conclude at engine displacements at six or higher, Chrysler is the greenest manufacturer out of the Big Three. At engine displacements below six, the greenest manufacturer is clearly either Ford or General Motors but no conclusion jumps out at us. If we had to choose one between the two, we would have to side with General Motors due to their significantly higher combined gas mileage across most engine displacements and their higher green score at mid-size engine displacements.

5 German and Japanese Vehicles

Apart from the Big Three manufacturers, German and Japanese cars are also very popular in the United States motor vehicle market. German cars are known for their durability, high technology, and high performance luxury cars. Japanese cars are known for being more pragmatic with good fuel economies and are sold at lower prices. Fuel efficiency is often one of the first things a consumer considers when choosing a vehicle, which is why we first looked at the relationship between German and Japanese manufacturers and combined gas mileage. We found that Japanese cars have a higher combined gas mileage than German cars on average.

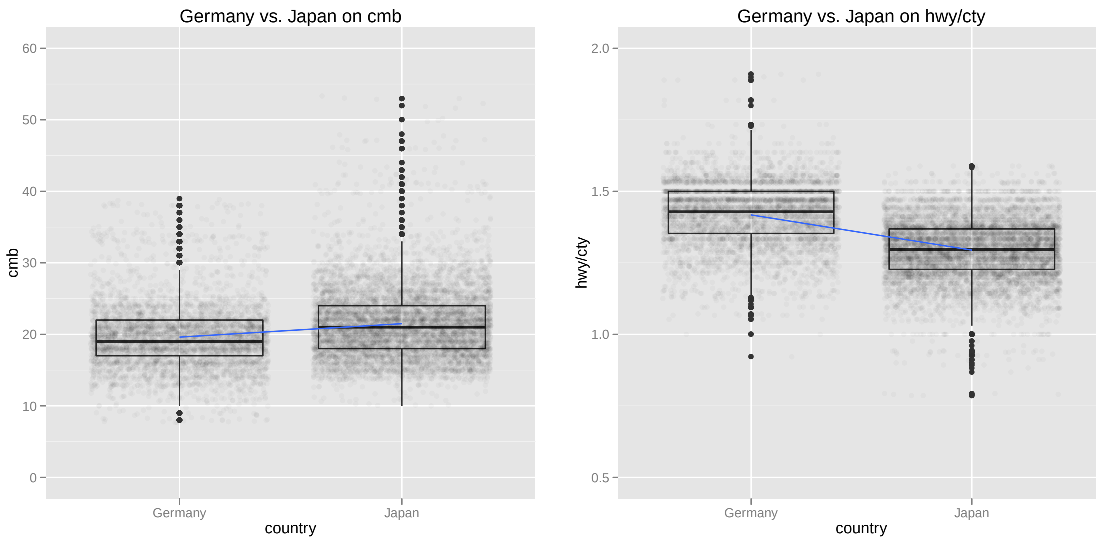


Figure 5: (Left) Boxplots overlaying scatterplots of manufacturing country vs. combined gas mileage. (Right) Boxplots overlaying scatterplots of the manufacturing country vs. the ratio of highway gas mileage to city gas mileage. Japanese cars have a higher combined gas mileage than German cars, but lower highway to city gas mileage ratio.

Since the difference between combined gas mileage could be a result of many different factors, we decided to look at the ratio between highway and city gas mileage for each country. We found that German cars have a much higher highway to city ratio than Japanese cars. This indicates larger fuel efficiency distinctions between highway and city gas mileage for German cars than Japanese cars, which could be the result of many different factors. One factor is the weight of the car and Japanese cars usually weigh less than German cars in the same vehicle class. Another factor is the environment the car is driven in. In the case of Japanese cars, a lighter vehicle weight has an advantage in city driving where there are many stop signals and traffic lights. On the highway, however, German cars have good fuel efficiency despite their heavier bodies.

This gained efficiency could be due to engine technology, which is why we chose to look at engine displacement and its relation to combined gas mileage. Typically, bigger displacement implies the consumption of more gasoline and lower gas mileage. We found that Japanese cars are more fuel-efficient than German cars when engine displacement is less than or equal to 1.6 liters. German cars have a higher combined gas mileage when displacement is greater than or equal to 2.5 liters. German cars are the only cars with reported displacements of over four liters. The lack of Japanese cars with displacements of over four liters could be due to Japanese manufacturers not designing cars with high displacements. Japanese cars are known to be smaller, lighter, and cheaper than their American and German counterparts.

The number of cylinders in the engine is another technological aspect that could affect combined gas

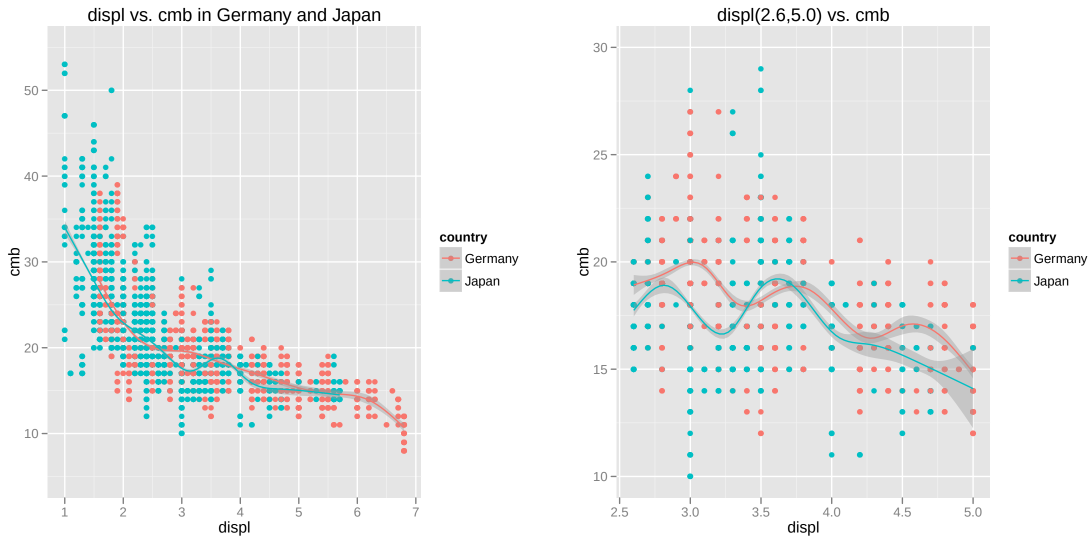


Figure 6: (Left) A scatterplot of engine displacement vs. combined gas mileage with lines indicating the mean combined gas mileage for the respective displacement. German cars are in orange; Japanese cars are in green. (Right) Zoom in on the graph on the left to more closely show the interplay between German and Japanese cars. German cars are in orange; Japanese cars are in green.

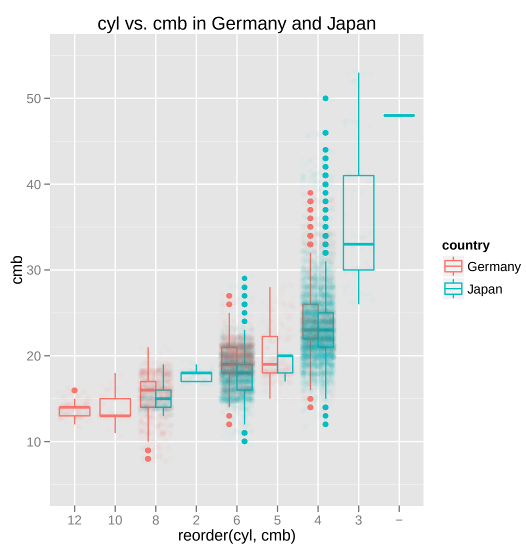


Figure 7: Boxplots overlaying scatterplots of the number of cylinders in an engine vs. combined gas mileage. German cars are in orange; Japanese cars are in green.

mileage. All the different numbers of cylinders in an engine were plotted in ascending order against combined gas mileage. In this initial plot, there is no significant difference between the combined gas mileage of German and Japanese four-cylinder cars. These four-cylinder engines are among the most popular, falling behind only six-cylinder engines. However, German cars have a significantly higher combined gas mileage than Japanese six-cylinder cars.

Only the Japanese manufacture two-cylinder cars with low gas mileage. A closer look at two-cylinder cars shows that all those vehicles are made by Mazda and are either a RX7 or RX8. The engines of these models employ a different working mechanism, which makes their engine as powerful as a six-cylinder engine and also brings their gas mileage down to the level of a powerful six-cylinder engine.

While German cars have a lower gas mileage than Japanese cars on average, they have a significantly higher gas mileage in the most popular engine type - the six-cylinder engine. This could be because the Japanese manufacture smaller vehicles with lower displacement and fewer cylinders that increase their fuel economy. They may also employ other technology that makes those cars more fuel-efficient. In the lower end of the motor vehicle market, Japanese cars take advantage of their lower production costs and retail prices to make them a strong competitor. Similarly, German cars may take advantage of their stronger bodies, more powerful engine, and luxurious features to make them competitive in the higher end of the motor vehicle market.

6 Conclusion

The findings of this study, from the advantages of premium fuel and the result of the Gas Guzzler Tax to the sustainability of cars manufactured in America to the differences in foreign cars, can help us as consumers make a more educated decision when purchasing a car. Although using premium gas results in higher gas mileage in different cars with the same number of cylinders, it may not be worth the higher price. A further exploration into the change in price from regular to premium gas would better explain whether premium gas would be worth the extra expenditure, especially if a manufacturer does not require a car to run on premium gas.

Similarly, though the Gas Guzzler Tax has made smaller ”guzzlers” more expensive over their lifetime than larger cars with low gas mileage, a further study could be conducted to see whether the Gas Guzzler Tax achieved its environmental goals. If more large cars with low gas mileage are being purchased and used, they could be emitting more greenhouse gasses. However, the switch to larger cars may also reduce traffic if the cars carry more passengers on average than the smaller ones.

For further study of the Big Three manufacturers, we could expand our analysis to more than just the flagship models for specific vehicle classes. Additionally, a further analysis of the relationship between green score and combined miles per gallon would be beneficial in order to explain some of the discrepancies between the two plots in Figure 4. In other words, a question for future study might be to ask why is it that some models have a high combined MPG at a given engine displacement but not a high green score, and vice-versa?

Lastly, while Japanese cars clearly have a better fuel economy due to their lighter body weight, it can be argued that German cars use more advanced technology that allows their heavier cars with six-cylinder engines to have better fuel economies than the Japanese alternatives. Another factor that could differentiate Japanese and German cars is the cost of maintenance over the life of the car and the length of the car's life. Though Japanese cars have a better fuel economy and likely allow their owners to save money on gas, they may not have as long a life, resulting in the purchase of a new car long before the end of a German car's life. This additional cost could be offset by the higher maintenance cost of German cars. However, if Japanese cars have more problems and need to be serviced more often, then they could have have high maintenance costs.

A Code

```
#####
##### Stat410 - Project 1
##### Final Code
##### 9/25/12
#####

# Load required packages
library(ggplot2)
library(xtable)

# Load strings as factors as FALSE to make it easier to subset
mpg2 <- read.csv("mpg2.csv.bz2", stringsAsFactors = FALSE)

#####
##### Goal: Look into relationship between fuel type gas mileage
#####

# Fuel type and gas mileage -----
# Look into the relationship between guzzlers and fuel type
# Subset for guzzlers
guzzlers_2 <- subset(mpg2, guzzler == "G")

# Determine the relevant fuel types
qplot(year, cty, data = guzzlers_2, geom = "smooth") + facet_wrap(~ fueltype)
# Interesting because it shows that only 3 fuel types  affect guzzlers
# come into play with guzzlers: regular, premium, and premium or E85

# Create new blank variable for more general fuel types
mpg2$newfuel <- NA

# Code certain fuel types and combine others (diesel, premium, regular)
mpg2[mpg2$fueltype == "Diesel", ]$newfuel <- "diesel"
mpg2[mpg2$fueltype == "Premium", ]$newfuel <- "premium"
mpg2[mpg2$fueltype == "Premium Gas or Electricity", ]$newfuel <- "premium"
mpg2[mpg2$fueltype == "Premium or E85", ]$newfuel <- "premium"
mpg2[mpg2$fueltype == "Regular", ]$newfuel <- "regular"

# Create levels for newfuel
levels_newfuel <- c("diesel", "premium", "regular")

# Assign new levels to make. There are no new labels.
mpg2$newfuel <- factor(mpg2$newfuel, levels = levels_newfuel)

# Fix overplotting with jitter and a change in alpha
# Tried 1/2, 1/4, 1/8 alphas, but changing shape works better
# Add boxplot to better see mean
```

```
# Want to zoom, so change ylim
qplot(newfuel, cmb, data = mpg2,
  geom = c("jitter", "boxplot"), shape = I(".")) +
  facet_wrap(~ guzzler) +
  ylim(0, 75)
# Interesting that there is really no difference in the gas mileage with
# higher grades of fuel even though people pay more for that gas and premium
# gas is often used in higher quality/more expensive cars.
# These higher quality and more expensive cars may just have bigger
# engines and more powerful engines.

# Create another group of guzzlers based on gas mileage
mpg2$newguzzler <- NA
mpg2[mpg2$cmb <= 22, ]$newguzzler <- "G"

# Make newguzzler a factor variable so we can plot with it
levels_newguzzler <- c("G")
mpg2$newguzzler <- factor(mpg2$newguzzler, levels = levels_newguzzler)

qplot(newfuel, cmb, data = mpg2, geom = c("jitter", "boxplot"),
  shape = I(".")) +
  facet_wrap(~ newguzzler)
# This shows us exactly what we expect to see
# Guzzlers, by definition, have lower gas mileage
# Interesting because if you look at non-guzzlers,
# Premium has a lower gas mileage than regular.

# Save to show lack of difference in gas mileage from fuel type
ggsave(file = "fuel_cmb_guz", width = 6, height = 6)

# Where does the difference in gas mileage based on fuel come from?
# What is the driving force behind fuel type's effect on gas mileage?
# Check fuel type and displacement. See if premium fuel has lower
# displacement
qplot(newfuel, displ, data = mpg2, geom = c("jitter", "boxplot"),
  shape = I("."))
# Usually higher, probably because engines requiring premium fuel are
# just more powerful and use more fuel.

# What about displacement and gas milage, split by fuel type?
qplot(displ, cmb, data = mpg2, geom = c("jitter", "smooth"),
  shape = I("."), alpha = I(1/4)) +
  facet_wrap(~ newfuel) +
  ylim(0, 60)
# Premium stays flat for all displacements pretty much. There is still a
# negative correlation between displacement and gas mileage.
# Regualr has a stronger, and therefore steeper curve that shows,
# negative correlation between displacement and gas mileage.
```

```
# Save plot for use in paper
ggsave(file = "displ_cmb_fuel.pdf", width = 6, height = 6)

qplot(drive <- ordered(drive,
                       levels = c("2-Wheel Drive", "Front-Wheel Drive",
                                  "Rear-Wheel Drive", "4-Wheel Drive",
                                  "4-Wheel or All-Wheel Drive",
                                  "All-Wheel Drive",
                                  "Part-time 4-Wheel Drive")),
                       cmb, data = mpg2,
                       geom = c("jitter", "boxplot"),
                       shape = I(".")) +
                       facet_wrap(~ newfuel) + ylim(10, 55)

# Regular fuel bounces around based on drive type
# premium fuel is generally lower than others
# Save plot for reference in paper
ggsave(file = "drive_cmb_fuel.pdf", width = 6, height = 6)

# Engine has higher displacement and is more powerful,
# requiring a higher grade of fuel
# Engine displacement is positively correlated to engine power
# Maybe split them up by cylinders and see if bigger engines need higher fuel

# Are engines with more cylinders cylinders more powerful?
# If so, would an increased number of cylinders lower gas mileage?

# Final plot: Fuel type and gas mileage -----
qplot(newfuel <- ordered(newfuel, levels = c("regular", "premium", "diesel")),
      cmb, data = mpg2, geom = c("jitter", "boxplot"),
      shape = I("."),
      xlab = "Regular, Premium, and Diesel", ylab = "Combined Gas Mileage",
      main ="Fuel Type, Cylinders, and Gas Mileage") +
        facet_wrap(~ cyl) +
        ylim(0, 60)
# Also tried different orders of fuel to show the trend clearly
# Fixed overplotting by adding jitter, boxplots, and changing shape
# Set ylim to zoom into the graphs
ggsave(file = "fuel_cmb_FINAL.pdf", width = 12, height = 12)

#####
##### Goal: Look at relationship between guzzlers, guzzler tax,
#####       and the cost of owning a guzzler over a lifetime
#####

# Guzzler Tax vs. Low Gas Mileage -----
```

```
# Is a guzzler tax equivalent to the amount spent on gas on cars with low gas
# mileage that are exempt from the tax?

# Add the tax value for each vehicle coded as a guzzler
# Taxes vary based on gas mileage, so taxes must be conditional on gas mileage

mpg2$tax <- ifelse(
  (mpg2$guzzler == "G") & (mpg2$cmb <= "12"), 7700,
    ifelse((mpg2$guzzler == "G") &
          (mpg2$cmb == "13"), 6400,
      ifelse((mpg2$guzzler == "G") &
            (mpg2$cmb == "14"), 5400,
        ifelse((mpg2$guzzler == "G") &
              (mpg2$cmb == "15"), 4500,
          ifelse((mpg2$guzzler == "G") &
                (mpg2$cmb == "16"), 3700,
            ifelse((mpg2$guzzler == "G") &
                  (mpg2$cmb == "17"), 3000,
              ifelse((mpg2$guzzler == "G") &
                    (mpg2$cmb == "18"), 2600,
                ifelse((mpg2$guzzler == "G") &
                      (mpg2$cmb == "19"), 2100,
                  ifelse((mpg2$guzzler == "G") &
                        (mpg2$cmb == "20"), 1700,
                    ifelse((mpg2$guzzler == "G") &
                          (mpg2$cmb == "21"), 1300,
                      ifelse((mpg2$guzzler == "G") &
                            (mpg2$cmb == "22"), 1000,
                        ifelse((mpg2$guzzler == "G") &
                              (mpg2$cmb > "22"), 0, 0))))))))))))

# Check to see if tax is a variable
str(mpg2)
# It is. Check whether the values distributed correctly.
table(mpg2$tax)

# Differentiate between guzzlers, which are taxed and nonguzzlers
notguzzlers <- mpg2[mpg2$guzzler != "G" && mpg2$cmb <= 22.5, ]
guzzlers <- mpg2[mpg2$guzzler == "G", ]

# Create new blank variable for vehicle classes of guzzlers
mpg2$gclass <- NA

# Check to see which classes to include
table(mpg2$guzzler, mpg2$vclass)

# Create new general categories for certain vehicle classes
mpg2[mpg2$vclass == "Compact Cars", ]$gclass <- "compact cars"
```

```
mpg2[mpg2$vclass == "Large Cars", ]$gclass <- "large cars"
mpg2[mpg2$vclass == "Midsize Cars", ]$gclass <- "midsize cars"
mpg2[mpg2$vclass == "Station Wagons", ]$gclass <- "station wagons"
mpg2[mpg2$vclass == "Midsize-Large Station Wagons", ]$gclass <- "station wagons"
mpg2[mpg2$vclass == "Minicompact Cars", ]$gclass <- "minicompact cars"
mpg2[mpg2$vclass == "Subcompact Cars", ]$gclass <- "subcompact cars"
mpg2[mpg2$vclass == "Two Seaters", ]$gclass <- "two seaters"

mpg2$ngclass <- NA

mpg2[mpg2$vclass == "Minivan - 2WD", ]$ngclass <- "Minivan"
mpg2[mpg2$vclass == "Minivan - 4WD", ]$ngclass <- "Minivan"
mpg2[mpg2$vclass == "Small Pickup Trucks", ]$ngclass <- "Pickups"
mpg2[mpg2$vclass == "Small Pickup Trucks 2WD", ]$ngclass <- "Pickups"
mpg2[mpg2$vclass == "Small Pickup Trucks 4WD", ]$ngclass <- "Pickups"
mpg2[mpg2$vclass == "Special Purpose Vehicle 2WD", ]$ngclass <- "SPV"
mpg2[mpg2$vclass == "Special Purpose Vehicle 4WD", ]$ngclass <- "SPV"
mpg2[mpg2$vclass == "Special Purpose Vehicles", ]$ngclass <- "SPV"
mpg2[mpg2$vclass == "Sport Utility Vehicle - 2WD", ]$ngclass <- "SUV"
mpg2[mpg2$vclass == "Sport Utility Vehicle - 4WD", ]$ngclass <- "SUV"
mpg2[mpg2$vclass == "Standard Pickup Trucks", ]$ngclass <- "Pickups"
mpg2[mpg2$vclass == "Standard Pickup Trucks 2WD", ]$ngclass <- "Pickups"
mpg2[mpg2$vclass == "Standard Pickup Trucks 4WD", ]$ngclass <- "Pickups"
mpg2[mpg2$vclass == "Standard Pickup Trucks/2wd", ]$ngclass <- "Pickups"

qplot(mpg2$gclass <- ordered(mpg2$gclass,
                             levels = c("two seaters", "minicompact cars",
                                        "compact cars", "subcompact cars",
                                        "midsize cars", "station wagons",
                                        "large cars")),
                            100000 * 3.85 / mpg2$cmb + mpg2$tax,
                            data = guzzlers,
                            geom = c("jitter", "boxplot"),
                            shape = I("."),
                            xlab = "Vehicle Class",
                            ylab = "Lifetime Fuel Cost in Dollars",
                            main = "Lifetime Cost of Guzzlers with Guzzler Tax") +
                            ylim(0, 50000)

# Save and insert into paper later
ggsave(file = "guzzlercost.pdf", width = 12, height = 12)

qplot(mpg2$ngclass <- ordered(mpg2$ngclass,
                              levels = c("Minivan", "SPV", "SUV", "Pickups")),
                              100000 * 3.85 / mpg2$cmb,
                              data = notguzzlers,
                              geom = c("jitter", "boxplot"),
                              shape = I("."),
```

```
                              xlab = "Vehicle Class",
                              ylab = "Lifetime Fuel Cost in Dollars",
                              main = "Lifetime Cost of Low Gas Mileage
                              without Tax") +
                              ylim(0, 50000)

# Save for later use
ggsave(file = "notguzzlercost.pdf", width = 12, height = 12)

# Table used in latex
our.table <- table(mpg2$tax)
xtable(our.table)

# Very interesting that even larger cars with low gas mileage end up
# costing about $5,000 less over their lifetime
# Lifetime estimated at 100,000 miles
# than smaller cars that have a guzzler tax on them

# Got rid of the smooth line because "NA" was skewing it.
# It showed a negative trend.

#####
##### Goal: Compare the Big Three (USA manufacturers)
#####       on the Greeness of their vehicles
#####

# Look at last 10 years in order to make our comparisions more impactful

mpg2.recent <- subset(mpg2, year >= 2002)

# Create a subset for the entire Big Three
mpg2.big3 <- subset(mpg2.recent,
                    make == "Ford" |
                    make == "Lincoln" |
                    make == "Chevrolet" |
                    make == "GMC" |
                    make == "Buick" |
                    make == "Cadillac" |
                    make == "Chrysler" |
                    make == "Dodge" |
                    make == "Jeep" |
                    make == "Ram" |
                    make == "Pontiac" |
                    make == "Mercury" |
                    make == "Saturn" |
                    make == "Oldsmobile" |
                    make == "Hummer")
```

```
# Create subset for each of the Big Three
mpg2.ford <- subset(mpg2.big3,
                    make == "Ford" |
                    make == "Lincoln" |
                    make == "Mercury")
mpg2.gm <- subset(mpg2.big3,
                  make == "Chevrolet" |
                  make == "GMC" |
                  make == "Buick" |
                  make == "Cadillac" |
                  make == "Pontiac" |
                  make == "Saturn" |
                  make == "Oldsmobile" |
                  make == "Hummer")
mpg2.chrysler <- subset(mpg2.big3,
                        make == "Chrysler" |
                        make == "Dodge" |
                        make == "Jeep" |
                        make == "Ram")

# Subset for displacements under 8
# because anything over 8 is outside of typical analysis
mpg2.big3 <- subset(mpg2.big3, displ <= 8)

# Create new variable/column for big3
# so that we can analyze it later
mpg2.big3$big3 <- NA
mpg2.big3[mpg2.big3$make %in%
  c("Ford", "Lincoln"), ]$big3 <- "Ford"
mpg2.big3[mpg2.big3$make %in%
  c("Chevrolet", "GMC", "Buick", "Cadillac", "Pontiac",
    "Oldsmobile", "Hummer"), ]$big3 <- "GM"
mpg2.big3[mpg2.big3$make %in%
  c("Chrysler", "Dodge", "Jeep", "Ram"), ]$big3 <- "Chrysler"

# Look at the ability of the manufacturers to create high cmb cars
# at each displacement, with displacement being a proxy for engine power
qplot(displ, cmb, data = mpg2.big3, color = big3, geom = "jitter",
      alpha = I(1/10)) + geom_smooth(alpha = .2, size = 1)
ggsave("matt-plot1.pdf", height = 6, width = 6)

# Look at a linear model to check for linear relationship
qplot(displ, cmb, data = mpg2.big3, color = big3, geom = "jitter",
      alpha = I(1/10)) + geom_smooth(method = lm, alpha = .2, size = 1)
ggsave("matt-plot2.pdf", height = 6, width = 6)

###
### Try the same thing while holding vclass constant across the Big Three
```

```
###

# First, using facet_wrap(), then using subsetting
# Subcompact Cars are not worth comparying because Chrysler only has 4
mpg2.big3.vclass <- subset(mpg2.big3,
                           vclass == "Standard Pickup Trucks 2WD" |
                           vclass == "Sport Utility Vehicle - 2WD" |
                           vclass == "Large Cars" |
                           vclass == "Midsize Cars" |
                           vclass == "Compact Cars")

# Get a general feel for cmb vs. displacemnt for each vehicle class
qplot(displ, cmb, data = mpg2.big3.vclass, color = big3, geom = "jitter",
  alpha = I(1/10)) + geom_smooth(alpha = .2, size = 1) + facet_wrap(~ vclass)

# Now, individually

# Standard Pickup Trucks 2WD
mpg2.big3.trucks <- subset(mpg2.big3, vclass == "Standard Pickup Trucks 2WD")
qplot(displ, cmb, data = mpg2.big3.trucks, color = big3, geom = "jitter",
      alpha = I(1/2)) + geom_smooth(alpha = .2, size = 1)
ggsave("matt-trucks.pdf", height = 6, width = 6)

# Sport Utility Vehicle - 2WD
mpg2.big3.suv <- subset(mpg2.big3, vclass == "Sport Utility Vehicle - 2WD")
qplot(displ, cmb, data = mpg2.big3.suv, color = big3, geom = "jitter",
      alpha = I(1/2)) + geom_smooth(alpha = .2, size = 1)
ggsave("matt-suv.pdf", height = 6, width = 6)

# Large Cars
mpg2.big3.large <- subset(mpg2.big3, vclass == "Large Cars")
qplot(displ, cmb, data = mpg2.big3.large, color = big3, geom = "jitter",
      alpha = I(1/2)) + geom_smooth(alpha = .2, size = 1)
ggsave("matt-large.pdf", height = 6, width = 6)

# Midsize Cars
mpg2.big3.mid <- subset(mpg2.big3, vclass == "Midsize Cars")
qplot(displ, cmb, data = mpg2.big3.mid, color = big3, geom = "jitter",
      alpha = I(1/2)) + geom_smooth(alpha = .2, size = 1)
ggsave("matt-mid.pdf", height = 6, width = 6)

# Compact Cars
mpg2.big3.compact <- subset(mpg2.big3, vclass == "Compact Cars")
qplot(displ, cmb, data = mpg2.big3.compact, color = big3, geom = "jitter",
      alpha = I(1/2)) + geom_smooth(alpha = .2, size = 1)
ggsave("matt-compact.pdf", height = 6, width = 6)

###
```

```
### Now look at the flagship model for each Big Three for each vclass.
### Also make new class for flagship luxury vehicles.
###

# By vclass
mpg2.big3.trucks.flag <- subset(mpg2.big3.trucks,
                                model == "F150 Pickup 2WD" |    # Ford
                                model == "Ram 1500 Pickup 2WD" | # Chrysler
                                model == "Sierra 1500 2WD")      # GM
mpg2.big3.suv.flag <- subset(mpg2.big3.suv,
                             model == "Durango 2WD" |        # Ford
                             model == "Grand Cherokee 2WD" | # Chrysler
                             model == "Yukon 1500 2WD")      # GM
mpg2.big3.large.flag <- subset(mpg2.big3.large,
                               model == "Taurus" |  # Ford
                               model == "Charger" | # Chrysler
                               model == "Impala")   # GM
mpg2.big3.mid.flag <- subset(mpg2.big3.mid,
                             model == "Fusion" |  # Ford
                             model == "Avenger" | # Chrysler
                             model == "Malibu")   # GM
mpg2.big3.compact.flag <- subset(mpg2.big3.compact,
                                 model == "Focus" |   # Ford
                                 model == "Sebring" | # Chrysler
                                 model == "G6")       # GM
mpg2.big3.lux.flag <- subset(mpg2.big3,
                             model == "MKZ" | # Ford
                             model == "200" | # Chrysler
                             model == "CTS")  # GM

# All together
mpg2.big3.flags <- subset(mpg2.big3,
                          model == "F150 Pickup 2WD" |     # Ford
                          model == "Ram 1500 Pickup 2WD" | # Chrysler
                          model == "Sierra 1500 2WD" |     # GM
                          model == "Durango 2WD" |        # Ford
                          model == "Grand Cherokee 2WD" | # Chrysler
                          model == "Yukon 1500 2WD"   |   # GM
                          model == "Taurus" |  # Ford
                          model == "Charger" | # Chrysler
                          model == "Impala"  | # GM
                          model == "Fusion" |  # Ford
                          model == "Avenger" | # Chrysler
                          model == "Malibu"  | # GM
                          model == "Focus" |   # Ford
                          model == "Sebring" | # Chrysler
                          model == "G6"      | # GM
                          model == "MKZ" | # Ford
```

```
                          model == "200" | # Chrysler
                          model == "CTS")  # GM

# Plot flagships cmb vs displ
qplot(displ, cmb, data = mpg2.big3.flags.displ, color = big3,
      geom = "jitter", alpha = I(1/2),
      main = "Big Three: Combined MPG",
      xlab = "engine displacement", ylab = "combined mpg") +
        geom_smooth(alpha = .2, size = 1)
ggsave("matt-plot3.pdf", height = 4, width = 6)

# Plot flagships cmb vs displ with linear smooth
# to check for a linear relationship
qplot(displ, cmb, data = mpg2.big3.flags.displ, color = big3, geom = "jitter",
      alpha = I(1/2)) + geom_smooth(method = lm, alpha = .2, size = 1)

###
### Assign gvg values
###

# Load gvg's; Tab delim
gvg_2002 <- read.delim("gvg_2002.txt", stringsAsFactors = T)
gvg_2003 <- read.delim("gvg_2003.txt", stringsAsFactors = T)
gvg_2004 <- read.delim("gvg_2004.txt", stringsAsFactors = T)
gvg_2005 <- read.delim("gvg_2005.txt", stringsAsFactors = T)
gvg_2006 <- read.delim("gvg_2006.txt", stringsAsFactors = T)
gvg_2007 <- read.delim("gvg_2007.txt", stringsAsFactors = T)
gvg_2008 <- read.delim("gvg_2008.txt", stringsAsFactors = T)
gvg_2009 <- read.delim("gvg_2009.txt", stringsAsFactors = T)
gvg_2010 <- read.delim("gvg_2010.txt", stringsAsFactors = T)
gvg_2011 <- read.delim("gvg_2011.txt", stringsAsFactors = T)
gvg_2012 <- read.delim("gvg_2012.txt", stringsAsFactors = T)

# Add year
gvg_2002$year <- 2002
gvg_2003$year <- 2003
gvg_2004$year <- 2004
gvg_2005$year <- 2005
gvg_2006$year <- 2006
gvg_2007$year <- 2007
gvg_2008$year <- 2008
gvg_2009$year <- 2009
gvg_2010$year <- 2010
gvg_2011$year <- 2011
gvg_2012$year <- 2012

# Merge
gvg.1 <- merge(gvg_2002, gvg_2003, all = T)
```

```
gvg.2 <- merge(gvg.1, gvg_2004, all = T)
gvg.3 <- merge(gvg.2, gvg_2005, all = T)
gvg.4 <- merge(gvg.3, gvg_2006, all = T)
gvg.5 <- merge(gvg.4, gvg_2007, all = T)
gvg.6 <- merge(gvg.5, gvg_2008, all = T)
gvg.7 <- merge(gvg.6, gvg_2009, all = T)
gvg.8 <- merge(gvg.7, gvg_2010, all = T)
gvg.9 <- merge(gvg.8, gvg_2011, all = T)
gvg <- merge(gvg.8, gvg_2012, all = T)

# Clean Greenhouse.Gas.Score
gvg.clean <- subset(gvg,
                    Greenhouse.Gas.Score == "0" |
                    Greenhouse.Gas.Score == "1" |
                    Greenhouse.Gas.Score == "2" |
                    Greenhouse.Gas.Score == "3" |
                    Greenhouse.Gas.Score == "4" |
                    Greenhouse.Gas.Score == "5" |
                    Greenhouse.Gas.Score == "6" |
                    Greenhouse.Gas.Score == "7" |
                    Greenhouse.Gas.Score == "8" |
                    Greenhouse.Gas.Score == "9" |
                    Greenhouse.Gas.Score == "10")
gvg.clean$Greenhouse.Gas.Score <- factor(gvg.clean$Greenhouse.Gas.Score,
  levels = c("0", "1", "2", "3", "4", "5", "6", "7", "8", "9", "10"),
  labels = c("0", "1", "2", "3", "4", "5", "6", "7", "8", "9", "10"))

# Clean Air Pollution Score
gvg.clean <- subset(gvg,
                    Air.Pollution.Score == "0" |
                    Air.Pollution.Score == "1" |
                    Air.Pollution.Score == "2" |
                    Air.Pollution.Score == "3" |
                    Air.Pollution.Score == "4" |
                    Air.Pollution.Score == "5" |
                    Air.Pollution.Score == "6" |
                    Air.Pollution.Score == "7" |
                    Air.Pollution.Score == "8" |
                    Air.Pollution.Score == "9" |
                    Air.Pollution.Score == "10")
gvg.clean$Air.Pollution.Score <- factor(gvg.clean$Air.Pollution.Score)

# Get average Greenhouse Gas for each flagship model
gvg.F150 <- subset(gvg.clean, Model == "FORD F150")
ghg.F150 <- mean((as.numeric(gvg.F150$Greenhouse.Gas.Score)))

gvg.RAM1500 <- subset(gvg.clean, Model == "DODGE RAM 1500")
ghg.RAM1500 <- mean((as.numeric(gvg.RAM1500$Greenhouse.Gas.Score)))
```

```
gvg.GMC1500 <- subset(gvg.clean, Model == "GMC Sierra 1500")
ghg.GMC1500 <- mean((as.numeric(gvg.GMC1500$Greenhouse.Gas.Score)))

gvg.Durango <- subset(gvg.clean, Model == "DODGE Durango")
ghg.Durango <- mean((as.numeric(gvg.Durango$Greenhouse.Gas.Score)))

gvg.GC <- subset(gvg.clean, Model == "JEEP Grand Cherokee")
ghg.GC <- mean((as.numeric(gvg.GC$Greenhouse.Gas.Score)))

gvg.Yukon <- subset(gvg.clean, Model == "GMC Yukon 1500")
ghg.Yukon <- mean((as.numeric(gvg.Yukon$Greenhouse.Gas.Score)))

gvg.Taurus <- subset(gvg.clean, Model == "FORD Taurus")
ghg.Taurus <- mean((as.numeric(gvg.Taurus$Greenhouse.Gas.Score)))

gvg.Charger <- subset(gvg.clean, Model == "DODGE Charger")
ghg.Charger <- mean((as.numeric(gvg.Charger$Greenhouse.Gas.Score)))

gvg.Impala <- subset(gvg.clean, Model == "CHEVROLET Impala")
ghg.Impala <- mean((as.numeric(gvg.Impala$Greenhouse.Gas.Score)))

gvg.Fusion <- subset(gvg.clean, Model == "FORD Fusion")
ghg.Fusion <- mean((as.numeric(gvg.Fusion$Greenhouse.Gas.Score)))

gvg.Avenger <- subset(gvg.clean, Model == "DODGE Avenger")
ghg.Avenger <- mean((as.numeric(gvg.Avenger$Greenhouse.Gas.Score)))

gvg.Malibu <- subset(gvg.clean, Model == "CHEVROLET Malibu")
ghg.Malibu <- mean((as.numeric(gvg.Malibu$Greenhouse.Gas.Score)))

gvg.Focus <- subset(gvg.clean, Model == "FORD Focus")
ghg.Focus <- mean((as.numeric(gvg.Focus$Greenhouse.Gas.Score)))

gvg.Sebring <- subset(gvg.clean, Model == "CHRYSLER Sebring")
ghg.Sebring <- mean((as.numeric(gvg.Sebring$Greenhouse.Gas.Score)))

gvg.G6 <- subset(gvg.clean, Model == "PONTIAC G6")
ghg.G6 <- mean((as.numeric(gvg.G6$Greenhouse.Gas.Score)))

gvg.MKZ <- subset(gvg.clean, Model == "LINCOLN MKZ")
ghg.MKZ <- mean((as.numeric(gvg.MKZ$Greenhouse.Gas.Score)))

gvg.200 <- subset(gvg.clean, Model == "CHRYSLER 200")
ghg.200 <- mean((as.numeric(gvg.200$Greenhouse.Gas.Score)))

gvg.CTS <- subset(gvg.clean, Model == "CADILLAC CTS")
ghg.CTS <- mean((as.numeric(gvg.CTS$Greenhouse.Gas.Score)))
```

```
# Get average Air Polution Score for each flagship model
aps.F150 <- mean((as.numeric(gvg.F150$Air.Pollution.Score)))
aps.RAM1500 <- mean((as.numeric(gvg.RAM1500$Air.Pollution.Score)))
aps.GMC1500 <- mean((as.numeric(gvg.GMC1500$Air.Pollution.Score)))
aps.Durango <- mean((as.numeric(gvg.Durango$Air.Pollution.Score)))
aps.GC <- mean((as.numeric(gvg.GC$Air.Pollution.Score)))
aps.Yukon <- mean((as.numeric(gvg.Yukon$Air.Pollution.Score)))
aps.Taurus <- mean((as.numeric(gvg.Taurus$Air.Pollution.Score)))
aps.Charger <- mean((as.numeric(gvg.Charger$Air.Pollution.Score)))
aps.Impala <- mean((as.numeric(gvg.Impala$Air.Pollution.Score)))
aps.Fusion <- mean((as.numeric(gvg.Fusion$Air.Pollution.Score)))
aps.Avenger <- mean((as.numeric(gvg.Avenger$Air.Pollution.Scoree)))
aps.Malibu <- mean((as.numeric(gvg.Malibu$Air.Pollution.Score)))
aps.Focus <- mean((as.numeric(gvg.Focus$Air.Pollution.Score)))
aps.Sebring <- mean((as.numeric(gvg.Sebring$Air.Pollution.Score)))
aps.G6 <- mean((as.numeric(gvg.G6$Air.Pollution.Score)))
aps.MKZ <- mean((as.numeric(gvg.MKZ$Air.Pollution.Score)))
aps.200 <- mean((as.numeric(gvg.200$Air.Pollution.Score)))
aps.CTS <- mean((as.numeric(gvg.CTS$Air.Pollution.Score)))

# Mean them for each flagship model to create Green variable
green.F150 <- (ghg.F150 + aps.F150) / 2
green.RAM1500 <- (ghg.RAM1500 + aps.RAM1500) / 2
green.GMC1500 <- (ghg.GMC1500 + aps.GMC1500) / 2
green.Durango <- (ghg.Durango + aps.Durango) / 2
green.GC <- (ghg.GC + aps.GC) / 2
green.Yukon <- (ghg.Yukon + aps.Yukon) / 2
green.Taurus <- (ghg.Taurus + aps.Taurus) / 2
green.Charger <- (ghg.Charger + aps.Charger) / 2
green.Impala <- (ghg.Impala + aps.Impala) / 2
green.Fusion <- (ghg.Fusion + aps.Fusion) / 2
green.Avenger <- (ghg.Avenger + aps.Avenger) / 2
green.Malibu <- (ghg.Malibu + aps.Malibu) / 2
green.Focus <- (ghg.Focus + aps.Focus) / 2
green.Sebring <- (ghg.Sebring + aps.Sebring) / 2
green.G6 <- (ghg.G6 + aps.G6) / 2
green.MKZ <- (ghg.MKZ + aps.MKZ) / 2
green.200 <- (ghg.200 + aps.200) / 2
green.CTS <- (ghg.CTS + aps.CTZ) / 2

# Apply Green variable column to flagship models
mpg2.big3.flags$green <- NA
mpg2.big3.flags[mpg2.big3.flags$model == "F150 Pickup 2WD",
                ]$green <- green.F150
mpg2.big3.flags[mpg2.big3.flags$model == "Ram 1500 Pickup 2WD",
                ]$green <- green.RAM1500
mpg2.big3.flags[mpg2.big3.flags$model == "Sierra 1500 2WD",
```

```
                ]$green <- green.GMC1500
mpg2.big3.flags[mpg2.big3.flags$model == "Durango 2WD",
                ]$green <- green.Durango
mpg2.big3.flags[mpg2.big3.flags$model == "Grand Cherokee 2WD",
                ]$green <- green.GC
mpg2.big3.flags[mpg2.big3.flags$model == "Yukon 1500 2WD",
                ]$green <- green.Yukon
mpg2.big3.flags[mpg2.big3.flags$model == "Taurus",
                ]$green <- green.Taurus
mpg2.big3.flags[mpg2.big3.flags$model == "Charger",
                ]$green <- green.Charger
mpg2.big3.flags[mpg2.big3.flags$model == "Impala",
                ]$green <- green.Impala
mpg2.big3.flags[mpg2.big3.flags$model == "Fusion",
                ]$green <- green.Fusion
mpg2.big3.flags[mpg2.big3.flags$model == "Avenger",
                ]$green <- green.Avenger
mpg2.big3.flags[mpg2.big3.flags$model == "Malibu",
                ]$green <- green.Malibu
mpg2.big3.flags[mpg2.big3.flags$model == "Focus",
                ]$green <- green.Focus
mpg2.big3.flags[mpg2.big3.flags$model == "Sebring",
                ]$green <- green.Sebring
mpg2.big3.flags[mpg2.big3.flags$model == "G6",
                ]$green <- green.G6
mpg2.big3.flags[mpg2.big3.flags$model == "MKZ",
                ]$green <- green.MKZ
mpg2.big3.flags[mpg2.big3.flags$model == "200",
                ]$green <- green.200
mpg2.big3.flags[mpg2.big3.flags$model == "CTS",
                ]$green <- green.CTS

# Plot Green vs. Displacement for flagship models
qplot(displ, green, data = mpg2.big3.flags, color = big3,
  geom = "jitter", alpha = I(1/2)) + geom_smooth(alpha = .2, size = 1)

# Final plot
qplot(displ, green, data = mpg2.big3.flags.final, color = big3,
      geom = "jitter",
      alpha = I(1/3), shape == I("."), main = "Big Three: Green Score",
      xlab = "engine displacement", ylab = "green score") +
      geom_smooth(alpha = .2, size = 1)
ggsave("matt-plot4.pdf", height = 4, width = 6)

#####
##### Goal:  Compare German vs. Japanese produced vehicles
#####
```

```
# Create variable "country"
mpg2$country <- NA
mpg2[mpg2$make %in% c("Acura", "Infiniti", "Honda", "Mazda", "Mitsubishi",
                      "Nissan", "Toyota", "Subaru",
                      "Suzuki", "Isuzu"), ]$country <- "Japan"
mpg2[mpg2$make %in% c("Mercedes-Benz", "BMW", "Audi",
                      "BMW Alpina", "MINI",
                      "Porsche", "Rolls-Royce",
                      "Volkswagen"), ]$country <- "Germany"
country_2 <- mpg2[!is.na(mpg2$country), ]

# Add variable mpd (mile per dollar), dpg (dollar per gallon)
country_2$dpg <- NA
country_2$mpd <- NA
premium <- country_2[country_2$fueltype == "Premium", ]
regular <- country_2[country_2$fueltype == "Regular", ]
country_2[country_2$fueltype == "Premium" & country_2$year == 1984, ]$dpg <-
  1.366
country_2[country_2$fueltype == "Premium" & country_2$year == 1985, ]$dpg <-
  1.340
country_2[country_2$fueltype == "Premium" & country_2$year == 1986, ]$dpg <-
  1.085
country_2[country_2$fueltype == "Premium" & country_2$year == 1987, ]$dpg <-
  1.093
country_2[country_2$fueltype == "Premium" & country_2$year == 1988, ]$dpg <-
  1.107
country_2[country_2$fueltype == "Premium" & country_2$year == 1989, ]$dpg <-
  1.197
country_2[country_2$fueltype == "Premium" & country_2$year == 1990, ]$dpg <-
  1.349
country_2[country_2$fueltype == "Premium" & country_2$year == 1991, ]$dpg <-
  1.321
country_2[country_2$fueltype == "Premium" & country_2$year == 1992, ]$dpg <-
  1.316
country_2[country_2$fueltype == "Premium" & country_2$year == 1993, ]$dpg <-
  1.302
country_2[country_2$fueltype == "Premium" & country_2$year == 1994, ]$dpg <-
  1.305
country_2[country_2$fueltype == "Premium" & country_2$year == 1995, ]$dpg <-
  1.336
country_2[country_2$fueltype == "Premium" & country_2$year == 1996, ]$dpg <-
  1.413
country_2[country_2$fueltype == "Premium" & country_2$year == 1997, ]$dpg <-
  1.416
country_2[country_2$fueltype == "Premium" & country_2$year == 1998, ]$dpg <-
  1.250
country_2[country_2$fueltype == "Premium" & country_2$year == 1999, ]$dpg <-
  1.357
```

```
country_2[country_2$fueltype == "Premium" & country_2$year == 2000, ]$dpg <-
  1.693
country_2[country_2$fueltype == "Premium" & country_2$year == 2001, ]$dpg <-
  1.657
country_2[country_2$fueltype == "Premium" & country_2$year == 2002, ]$dpg <-
  1.556
country_2[country_2$fueltype == "Premium" & country_2$year == 2003, ]$dpg <-
  1.777
country_2[country_2$fueltype == "Premium" & country_2$year == 2004, ]$dpg <-
  2.068
country_2[country_2$fueltype == "Premium" & country_2$year == 2005, ]$dpg <-
  2.491
country_2[country_2$fueltype == "Premium" & country_2$year == 2006, ]$dpg <-
  2.805
country_2[country_2$fueltype == "Premium" & country_2$year == 2007, ]$dpg <-
  3.033
country_2[country_2$fueltype == "Premium" & country_2$year == 2008, ]$dpg <-
  3.591
country_2[country_2$fueltype == "Premium" & country_2$year == 2009, ]$dpg <-
  2.607
country_2[country_2$fueltype == "Premium" & country_2$year == 2010, ]$dpg <-
  3.047
country_2[country_2$fueltype == "Premium" & country_2$year == 2011, ]$dpg <-
  3.792
country_2[country_2$fueltype == "Premium" & country_2$year == 2012, ]$dpg <-
  3.930
country_2[country_2$fueltype == "Regular" & country_2$year == 1984, ]$dpg <-
  1.212
country_2[country_2$fueltype == "Regular" & country_2$year == 1985, ]$dpg <-
  1.202
country_2[country_2$fueltype == "Regular" & country_2$year == 1986, ]$dpg <-
  0.927
country_2[country_2$fueltype == "Regular" & country_2$year == 1987, ]$dpg <-
  0.948
country_2[country_2$fueltype == "Regular" & country_2$year == 1988, ]$dpg <-
  0.946
country_2[country_2$fueltype == "Regular" & country_2$year == 1989, ]$dpg <-
  1.022
country_2[country_2$fueltype == "Regular" & country_2$year == 1990, ]$dpg <-
  1.164
country_2[country_2$fueltype == "Regular" & country_2$year == 1991, ]$dpg <-
  1.140
country_2[country_2$fueltype == "Regular" & country_2$year == 1992, ]$dpg <-
  1.127
country_2[country_2$fueltype == "Regular" & country_2$year == 1993, ]$dpg <-
  1.108
country_2[country_2$fueltype == "Regular" & country_2$year == 1994, ]$dpg <-
  1.112
```

```
country_2[country_2$fueltype == "Regular" & country_2$year == 1995, ]$dpg <-
  1.147
country_2[country_2$fueltype == "Regular" & country_2$year == 1996, ]$dpg <-
  1.231
country_2[country_2$fueltype == "Regular" & country_2$year == 1997, ]$dpg <-
  1.234
country_2[country_2$fueltype == "Regular" & country_2$year == 1998, ]$dpg <-
  1.059
country_2[country_2$fueltype == "Regular" & country_2$year == 1999, ]$dpg <-
  1.165
country_2[country_2$fueltype == "Regular" & country_2$year == 2000, ]$dpg <-
  1.510
country_2[country_2$fueltype == "Regular" & country_2$year == 2001, ]$dpg <-
  1.461
country_2[country_2$fueltype == "Regular" & country_2$year == 2002, ]$dpg <-
  1.358
country_2[country_2$fueltype == "Regular" & country_2$year == 2003, ]$dpg <-
  1.591
country_2[country_2$fueltype == "Regular" & country_2$year == 2004, ]$dpg <-
  1.880
country_2[country_2$fueltype == "Regular" & country_2$year == 2005, ]$dpg <-
  2.295
country_2[country_2$fueltype == "Regular" & country_2$year == 2006, ]$dpg <-
  2.589
country_2[country_2$fueltype == "Regular" & country_2$year == 2007, ]$dpg <-
  2.801
country_2[country_2$fueltype == "Regular" & country_2$year == 2008, ]$dpg <-
  3.266
country_2[country_2$fueltype == "Regular" & country_2$year == 2009, ]$dpg <-
  2.350
country_2[country_2$fueltype == "Regular" & country_2$year == 2010, ]$dpg <-
  2.788
country_2[country_2$fueltype == "Regular" & country_2$year == 2011, ]$dpg <-
  3.527
country_2[country_2$fueltype == "Regular" & country_2$year == 2012, ]$dpg <-
  3.659
country_2$mpd <- country_2$cmb / country_2$dpg

# Which country's cars saves more fuel?
qplot(country, cmb, data = country_2, geom = c("boxplot", "jitter"),
      main = "Germany vs. Japan on cmb", alpha = I(1/40)) + ylim(0,60) +
      geom_smooth(aes(group = 1), method = "lm")
ggsave("G.J.cmb.pdf", width = 6, height = 6)

qplot(country, hwy, data = country_2, geom = c("boxplot", "jitter"),
      alpha = I(1/30)) +ylim(5, 55) +
      geom_smooth(aes(group = 1), method = "lm")
```

```
qplot(country, cty / dpg, data = country_2, geom = c("boxplot", "jitter"),
      alpha = I(1/30)) +ylim(5, 55) +
      geom_smooth(aes(group = 1), method = "lm")

qplot(country, hwy / cty, data = country_2, geom = c("boxplot", "jitter"),
      main = "Germany vs. Japan on hwy/cty", alpha = I(1/40)) + ylim(0.5, 2) +
      geom_smooth(aes(group = 1), method = "lm")
ggsave("G.J.hwy.cty.pdf", width = 6, height = 6)

qplot(country, cty / hwy, data = country_2, geom = c("boxplot", "jitter"),
      alpha = I(1/40)) + ylim(0.5, 1.3) +
      geom_smooth(aes(group = 1), method = "lm")

# Look at the displ
qplot(displ, cmb, data = country_2, main = "displ vs. cmb in Germany and Japan",
      color = country) + ylim(5, 55)+ geom_smooth()
ggsave("displ.cmb.G.J.pdf", width = 6, height = 6)

qplot(reorder(cyl, cmb), cmb, data = country_2[!is.na(country_2$cyl), ],
      color = country) + ylim(5, 55)

qplot(reorder(cyl, cmb), cmb, data = country_2[!is.na(country_2$cyl), ],
      main = "cyl vs. cmb in Germany and Japan", color = country,
      geom = c("boxplot", "jitter"), alpha = I(1/40)) + ylim(5, 55)
ggsave("cyl.cmb.G.J.pdf", width = 6, height = 6)

table(country_2[country_2$cmb >= 40, ]$make)
# Honda       MINI Mitsubishi     Nissan     Suzuki     Toyota
# 60          1          1          8          7          17

table(country_2[country_2$cmb >= 55, ]$make)
# MINI Mitsubishi     Nissan     Toyota
# 1          1          3          4

country_2[country_2$cmb >= 55, ]
# Almost Japan
high_cmb <- country_2[country_2$cmb >= 40, ]
high_cmb <- high_cmb[high_cmb$make != "MINI" & high_cmb$make != "Mitsubishi", ]
table(high_cmb$make)
# Honda Nissan Suzuki Toyota
# 60      8      7     17

qplot(displ, cmb, data = high_cmb, color = cyl,
      geom = c("jitter", "boxplot"))+ ylim(40,55) + facet_wrap(~ make)

# Exploring Germany special cars (displ)
qplot(displ, cmb, data = country_2, color = country) +
      xlim(2.5, 5) +
```

```
      ylim(5, 40) +
      geom_smooth()

qplot(displ, cmb, data = country_2, color = country) +
      xlim(2.5, 5) +
      ylim(10, 35) +
      geom_smooth()

qplot(displ, cmb, data = country_2,
      main = "displ(2.6, 5.0) vs. cmb",
      color = country) +
      xlim(2.6, 5.0) +
      ylim(10, 30) +
      geom_smooth()
ggsave("displ.2.6.5.0.cmb.pdf", width = 6, height = 6)

# Tendancy chart for different countries through years
qplot(year, cmb, data = country_2,
      geom = c("boxplot", "jitter"),
      group = year, alpha = I(1/10)) +
      ylim(5, 55) +
      facet_wrap(~ country) +
      geom_smooth(aes(group = 1))

# New dataset with only premium and regular fueltype
new_fuel <- country_2[country_2$fueltype == "Premium" |
  country_2$fueltype == "Regular", ]

# Does premium save money?
qplot(year, mpd, data = new_fuel,
      main = "year vs. mpd for premium and regular",
      color = country) +
      facet_wrap(~ fueltype) +
      ylim(0, 40) +
      geom_smooth(aes(group = 1))
ggsave("year.mpd.premium.regular.pdf", width = 6, height = 6)

table(country_2$fueltype)
# CNG          Diesel     Electricity Gasoline or E85         Premium
# 17             308            11               54            4540
# Premium or E85         Regular
# 18                      6938

# Create new dataset with specific vclass
premium <- premium[premium$vclass == "Compact Cars" |
                   premium$vclass == "Midsize Cars" |
                   premium$vclass == "Two Seaters" |
                   premium$vclass == "Large Cars", ]
```

```
regular <- regular[regular$vclass == "Compact Cars" |
                   regular$vclass == "Midsize Cars" |
                   regular$vclass == "Two Seaters" |
                   regular$vclass == "Large Cars", ]

# Does the fueltype have relationships with vclass and displ?
qplot(displ, cmb, data = premium,
      main = "displ vs. cmb with premium") +
      facet_wrap(~ vclass) +
      geom_smooth()
ggsave("displ.cmb.premium.vclass.pdf", width = 6, height = 6)

qplot(displ, cmb, data = regular,
      main = "displ vs. cmb with regular") +
      facet_wrap(~ vclass) + ylim(0, 35) +
      geom_smooth()
ggsave("displ.cmb.regular.vclass.pdf", width = 6, height = 6)

#####
##### End Code.
#####
```