Vehicle_Basic_Analysis

Piyu Tiwari

February 25, 2019

Purpose

Purpose of this analysis is to look at the trend in fuel efficiency in vehicles over the years and to find out the major reasons behind the trend.

Basic data Analysis

Number of observations : 39542 observations

Number of features : 83 features

Number of unique years : 36 years

```
first_year <- min(vehicles$year)
last_year <- max(vehicles$year)
```

First year : 1984

Last year : 2019

What are the primary fuel types

```
table(vehicles$fuelType1)
```

```
           Diesel       Electricity Midgrade Gasoline       Natural Gas
             1118               163                94                60
 Premium Gasoline  Regular Gasoline
            10962             27145
```

What are the major transmissions

```
table(vehicles$trany)
```

```
                                        Automatic (A1)
                          11                       158
               Automatic (AM-S6)        Automatic (AM-S7)
                         113                       341
               Automatic (AM-S8)        Automatic (AM-S9)
                          29                         2
                 Automatic (AM5)          Automatic (AM6)
                          14                       135
                 Automatic (AM7)          Automatic (AM8)
                         201                         5
              Automatic (AV-S10)        Automatic (AV-S6)
                           3                       183
               Automatic (AV-S7)        Automatic (AV-S8)
                         113                        32
                  Automatic (L3)           Automatic (L4)
                           2                         2
                 Automatic (S10)           Automatic (S4)
                          53                       233
                  Automatic (S5)           Automatic (S6)
                         830                      2919
                  Automatic (S7)           Automatic (S8)
                         299                      1274
                  Automatic (S9) Automatic (variable gear ratios)
                          55                       754
                Automatic 10-spd           Automatic 3-spd
                           8                      3151
                 Automatic 4-spd           Automatic 5-spd
                       11045                      2198
                 Automatic 6-spd           Automatic 7-spd
                        1530                       706
                 Automatic 8-spd           Automatic 9-spd
                         330                       186
                    Manual 3-spd              Manual 4-spd
                          77                      1483
          Manual 4-spd Doubled              Manual 5-spd
                          17                      8343
```

```
                  Manual 6-spd                   Manual 7-spd
                          2612                             95
```

It seems like there are 11 blank values. I will replace blanks with NA as it is easy to handle NA.

```
vehicles$trany[vehicles$trany ==""] <- NA
```

We only care whether transmission is automatic or manual hence I am going to replace all subscript with automatic or manual.

```
vehicles$trany <- as.factor(ifelse(substr(vehicles$trany,1,4)== "Auto","Auto","Manual"))
```

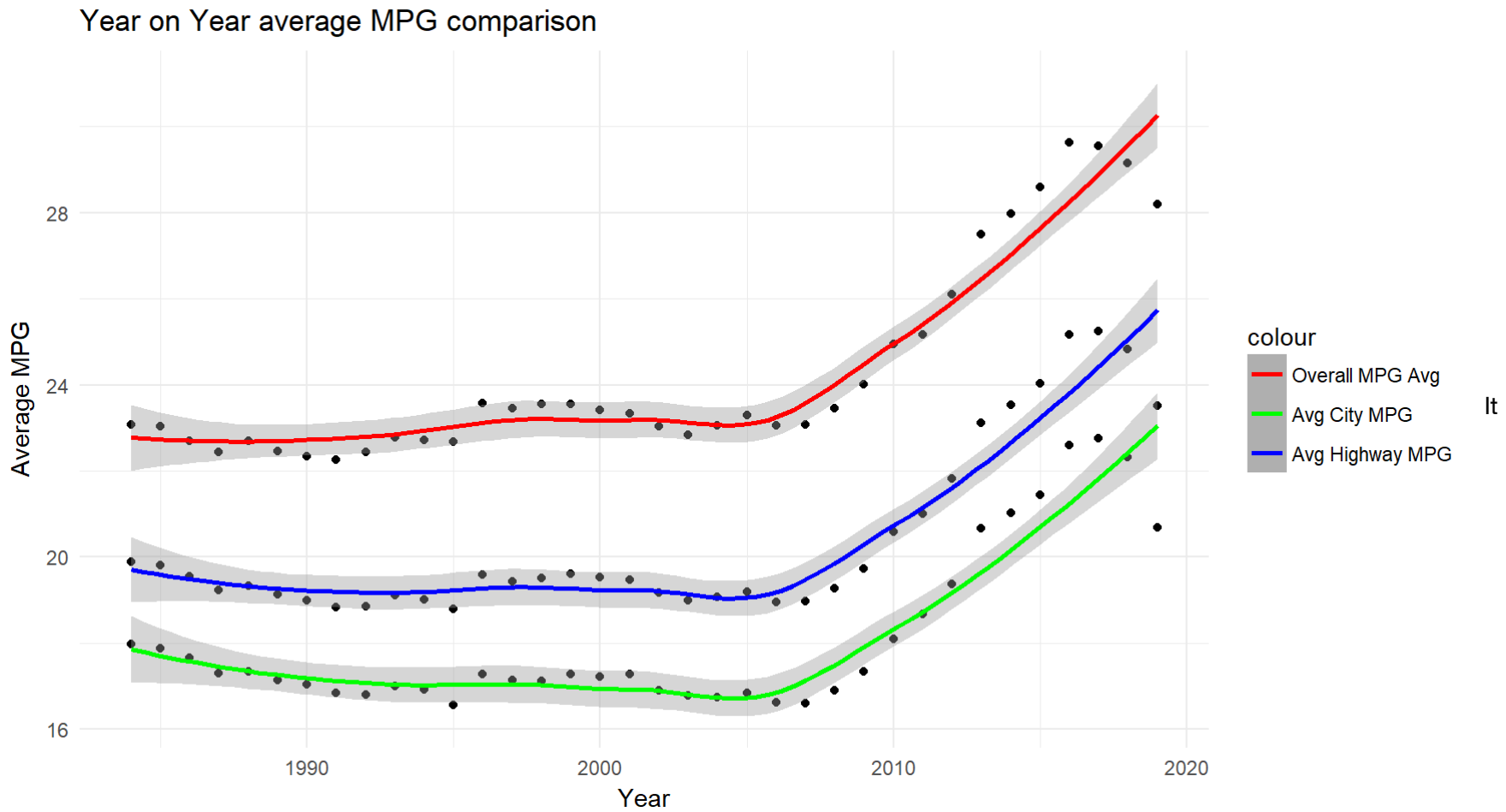
Analysis

MPG trend over the years.

First thing that I would like to know is the trend in car fuel efficiency over the years. Has it increased or decreased?

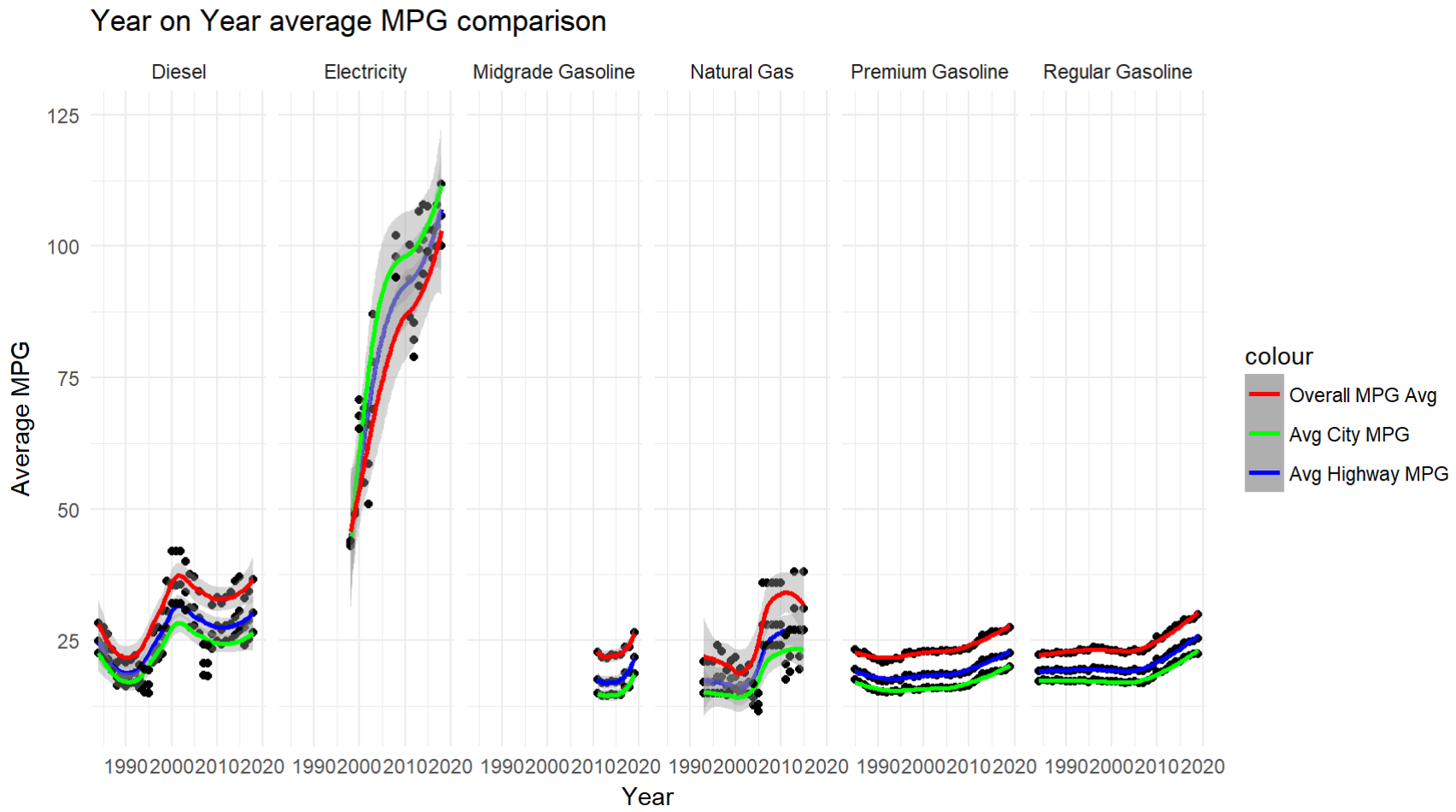
To check the trend, we will compare average overall fuel efficiency , city fuel efficiency and highway fuel efficiency.

```
mpg_by_year <-  vehicles %>% group_by(year) %>% summarise(mean_mpg =mean(comb08) ,mean_city_mpg= mean(city08),
                                                          mean_highway_mpg = mean(highway08))


p1 <- ggplot(data = mpg_by_year, aes(x=year)) + geom_point(aes(y=mean_mpg)) + geom_smooth(aes(y=mean_mpg,color="r
ed"),method = "auto") + geom_point(aes(y=mean_city_mpg)) + geom_smooth(aes(y=mean_city_mpg,color="green"),method
 = "auto") + geom_point(aes(y=mean_highway_mpg)) + geom_smooth(aes(y=mean_highway_mpg,color="blue"),method = "aut
o")+labs(x="Year",y="Average MPG",title="Year on Year average MPG comparison")+scale_color_manual(values=c("red",
"green","blue"),label=c("Overall MPG Avg","Avg City MPG","Avg Highway MPG")) + theme_minimal()

p1
```



looks like fuel efficiency has increased over the years. I think it will be worthwhile to check what happens with different fuel engine types(Electric engine or gasoline engine).

```
mpg_by_year <-  vehicles %>% group_by(year,fuelType1) %>% summarise(mean_mpg =mean(comb08) ,mean_city_mpg= mean(c
ity08),
                                                                    mean_highway_mpg = mean(highway08))


ggplot(data = mpg_by_year, aes(x=year)) + geom_point(aes(y=mean_mpg)) + geom_smooth(aes(y=mean_mpg,color="red"),m
ethod = "auto") + geom_point(aes(y=mean_city_mpg)) + geom_smooth(aes(y=mean_city_mpg,color="green"),method = "aut
o") + geom_point(aes(y=mean_highway_mpg)) + geom_smooth(aes(y=mean_highway_mpg,color="blue"),method = "auto")+lab
s(x="Year",y="Average MPG",title="Year on Year average MPG comparison")+scale_color_manual(values=c("red","green"
,"blue"),label=c("Overall MPG Avg","Avg City MPG","Avg Highway MPG"))+ facet_grid(~fuelType1)+theme_minimal()
```

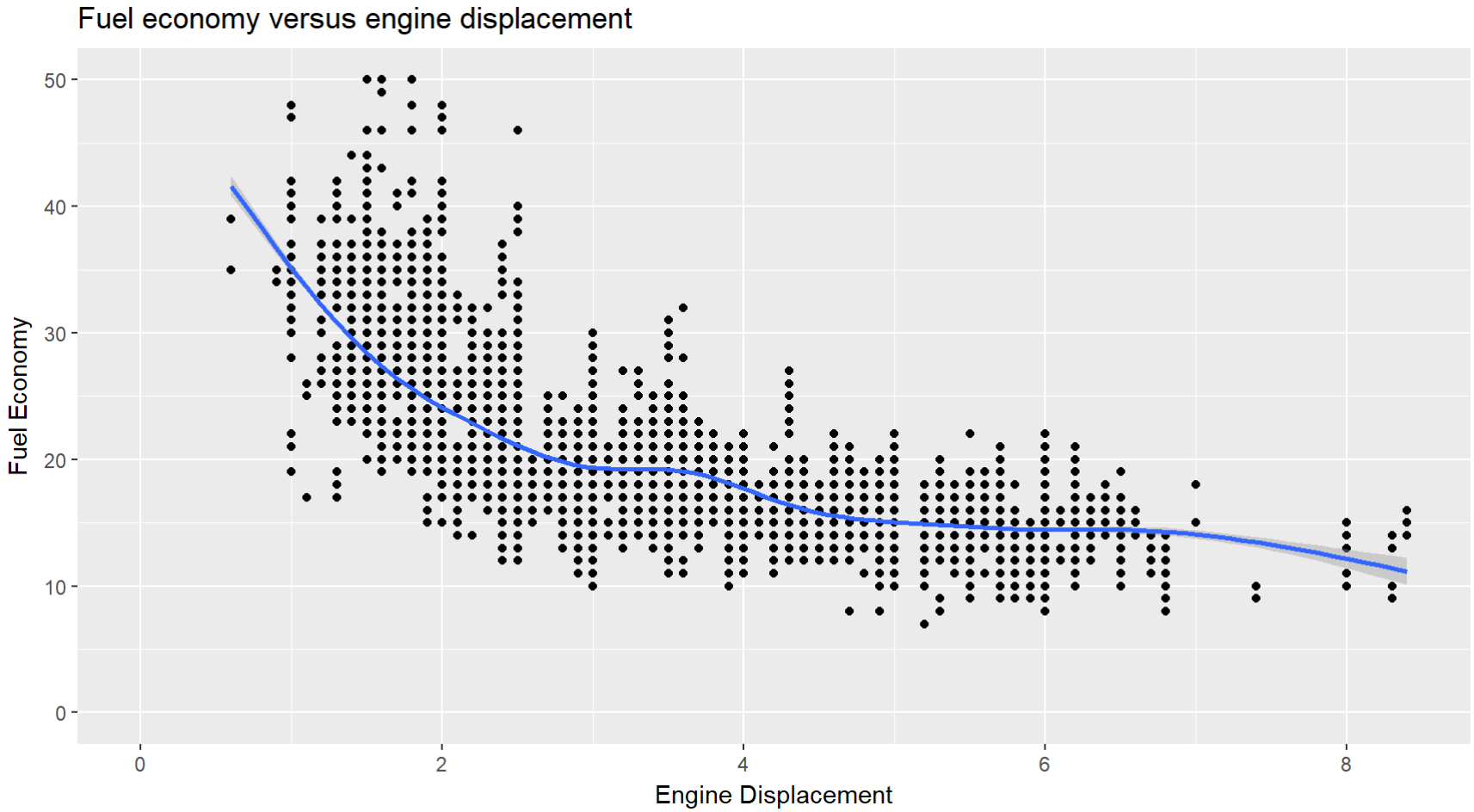


As can be seen from the previous graph that the electricity vehicles have contributed a lot in the increase in fuel efficiency. While fuel efficiency of regular gasoline has also increased. But there could be various reasons for that.Usually, cars with higher engine have lower fuel efficiency, It could be very well possible that we have decreased the production of large engine vehicles.

Fuel efficiency of Large engine vehicle versus lower engine vehicle

Before looking into whether the production of large engine vehicle has been less when compared to small engine vehicle, we will first confirm our hypothesis that large engines have lower fuel efficiency.

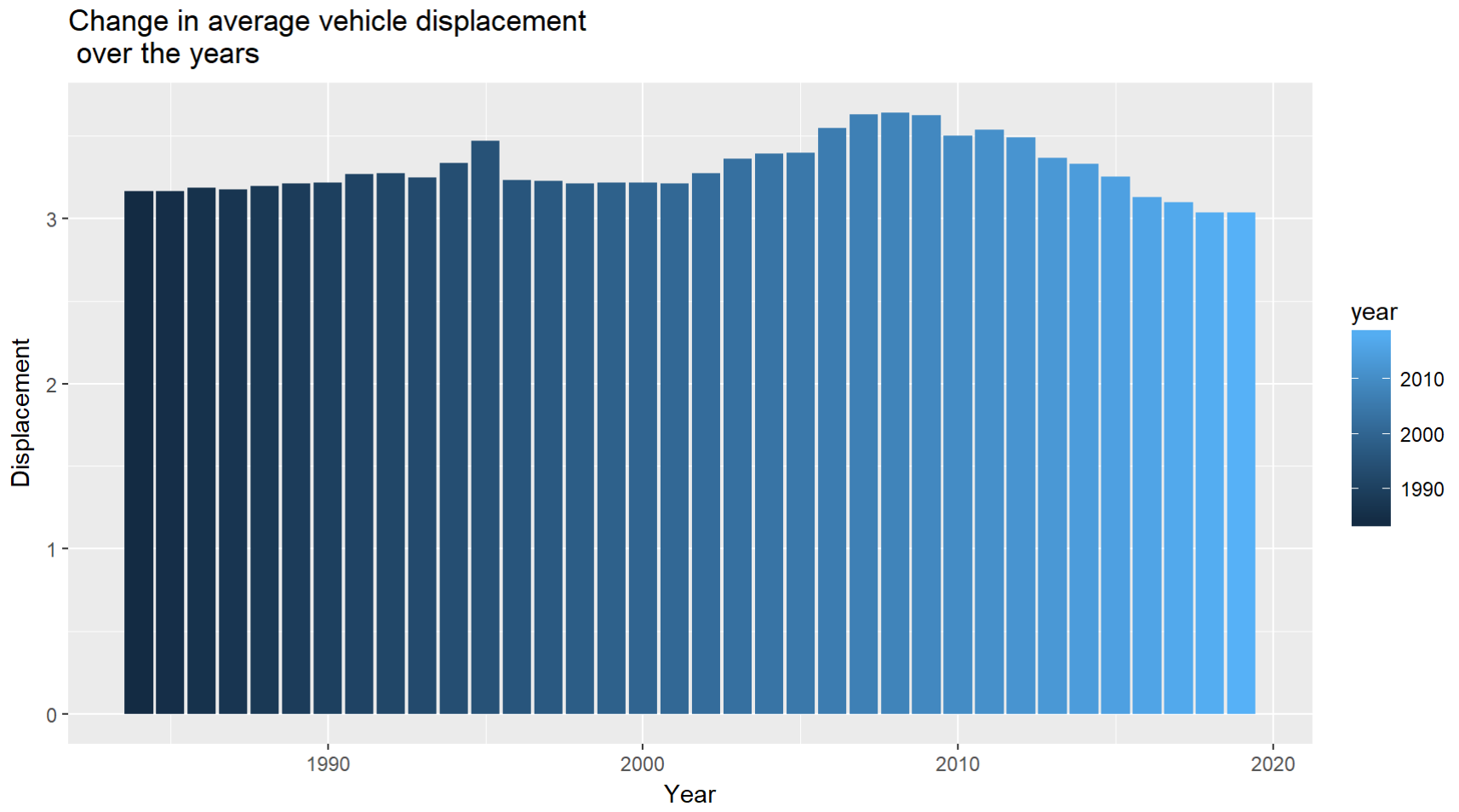
```
ggplot(data = vehicles,aes(x=displ,y=comb08))+geom_point()+ylim(0,50)+geom_smooth()+labs(x="Engine Displacement",
y="Fuel Economy",title="Fuel economy versus engine displacement")
```



It is quite evident that the fuel economy decreases with increase in fuel displacement.

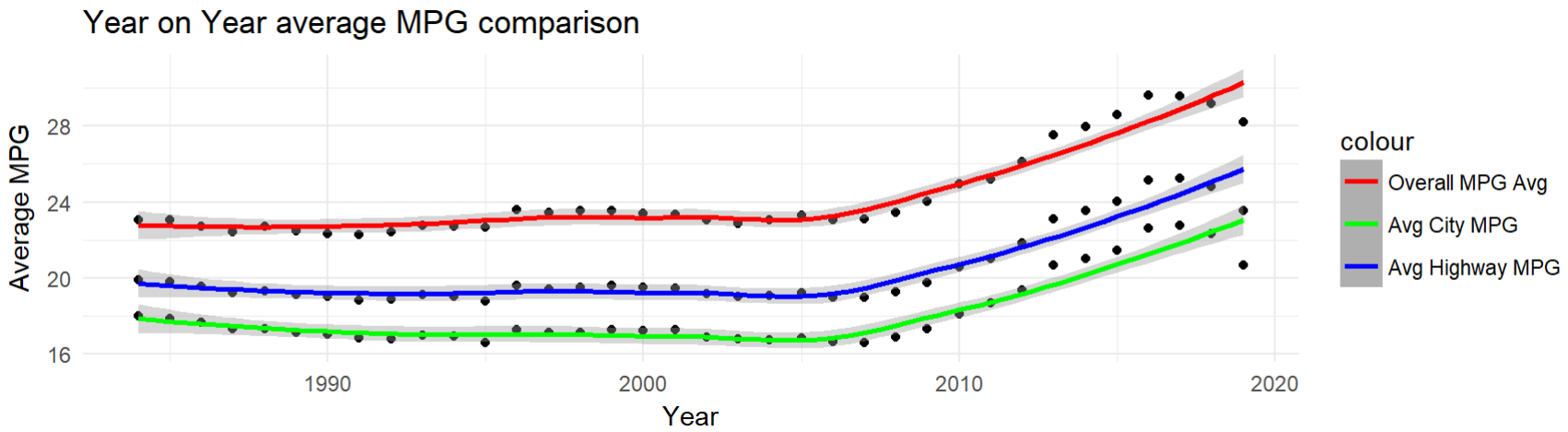
Now, we are going to check if production of larger vehicle has decreased over the years.

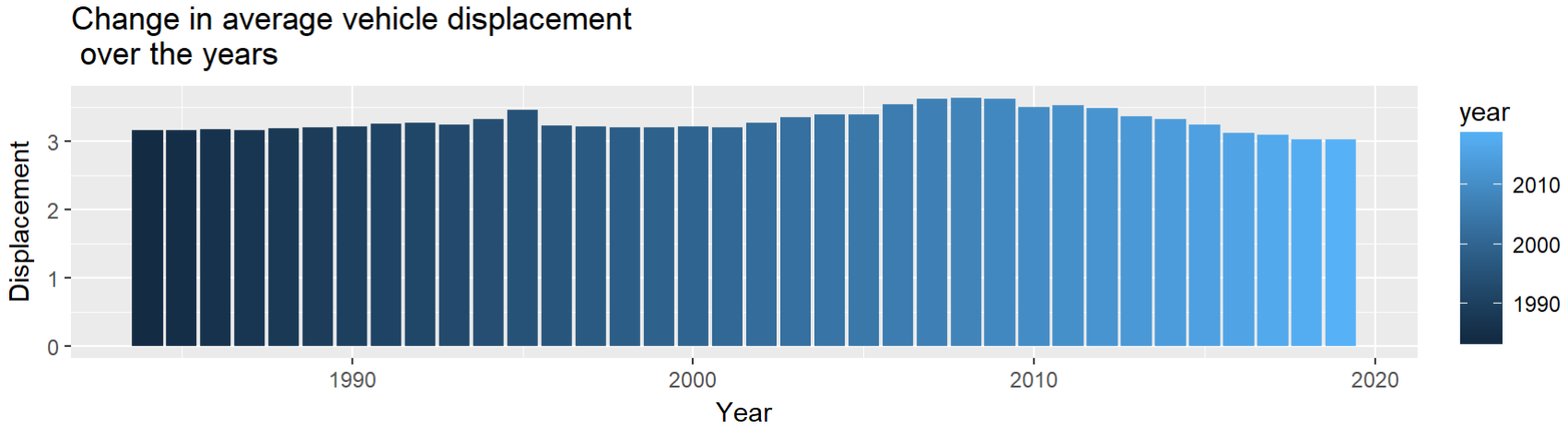
```
p2 <- ggplot(data = vehicles,aes(x=year,y=displ,fill=year))+geom_bar(stat="summary",fun.y="mean")+labs(x="Year",
 y = "Displacement", title="Change in average vehicle displacement \n over the years ")

#+geom_hline(yintercept =mean(vehicles$displ,na.rm = T),color="red")+geom_text(aes(1984,mean(vehicles$displ,na.rm
= T),label="Lifetime displ avg",vjust=-1))

p2
```



It is for sure that the higher displacement engine count has decreased considerably. For better comparison, lets us plot fuel efficiency and displacement next to each other.

```
grid.arrange(p1,p2)
```





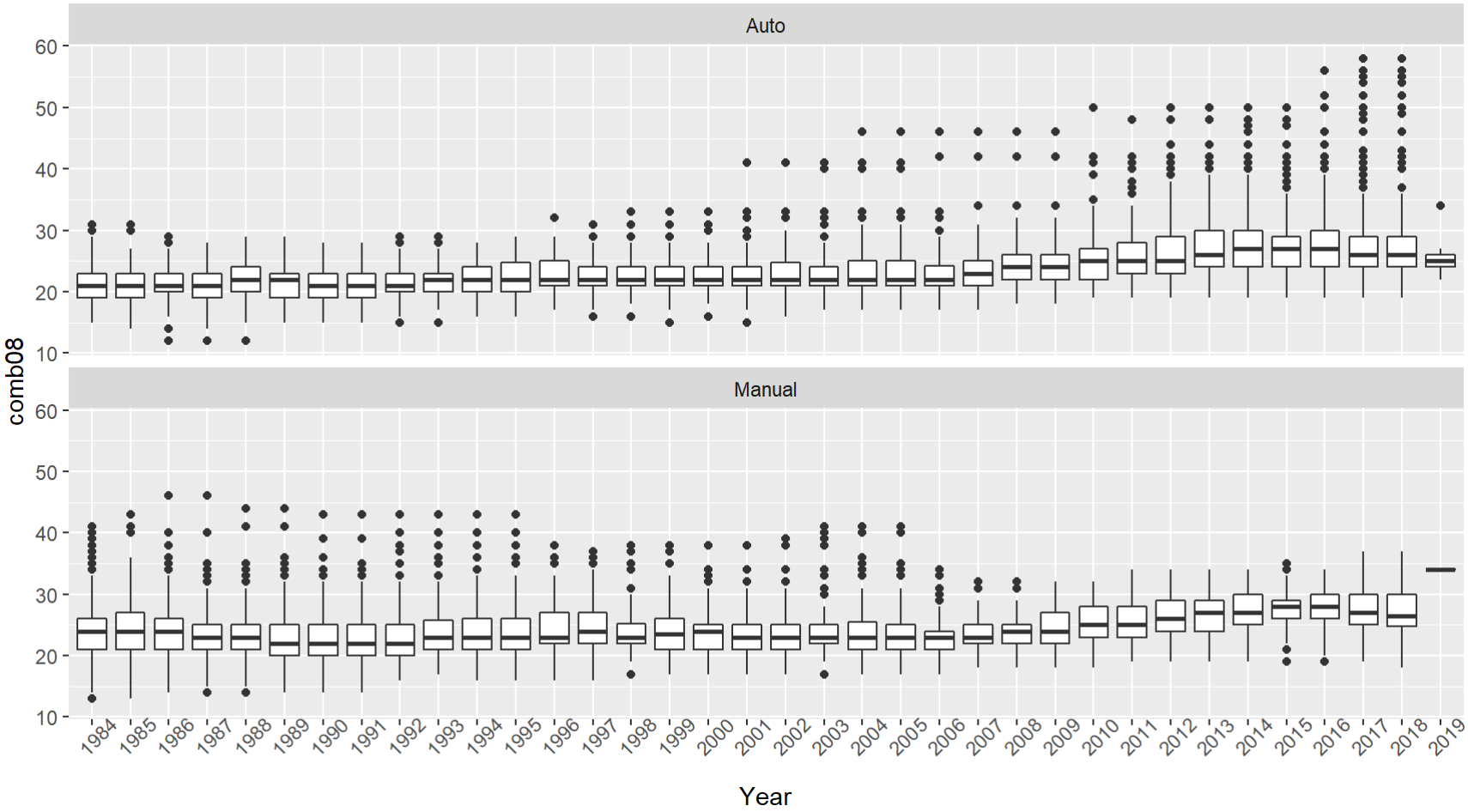
Things to notice in the above plots:

- Engine displacement has been increasing up until 2008, with a sudden increase between 2006-2008.
- Since 2009, there is considerable decrease in engine size. This partially explains the increase in fuel economy.
- Until 2005, engine displacement has increased however fuel efficiency roughly remained constant. This indicates that engine efficiency has increased over the years.
- Although there is jump in engine size between 2006 and 2008 however engine efficiency has remained constant. This look like an issue. Need to investigated more.

Look like smaller vehicles are in fashion.

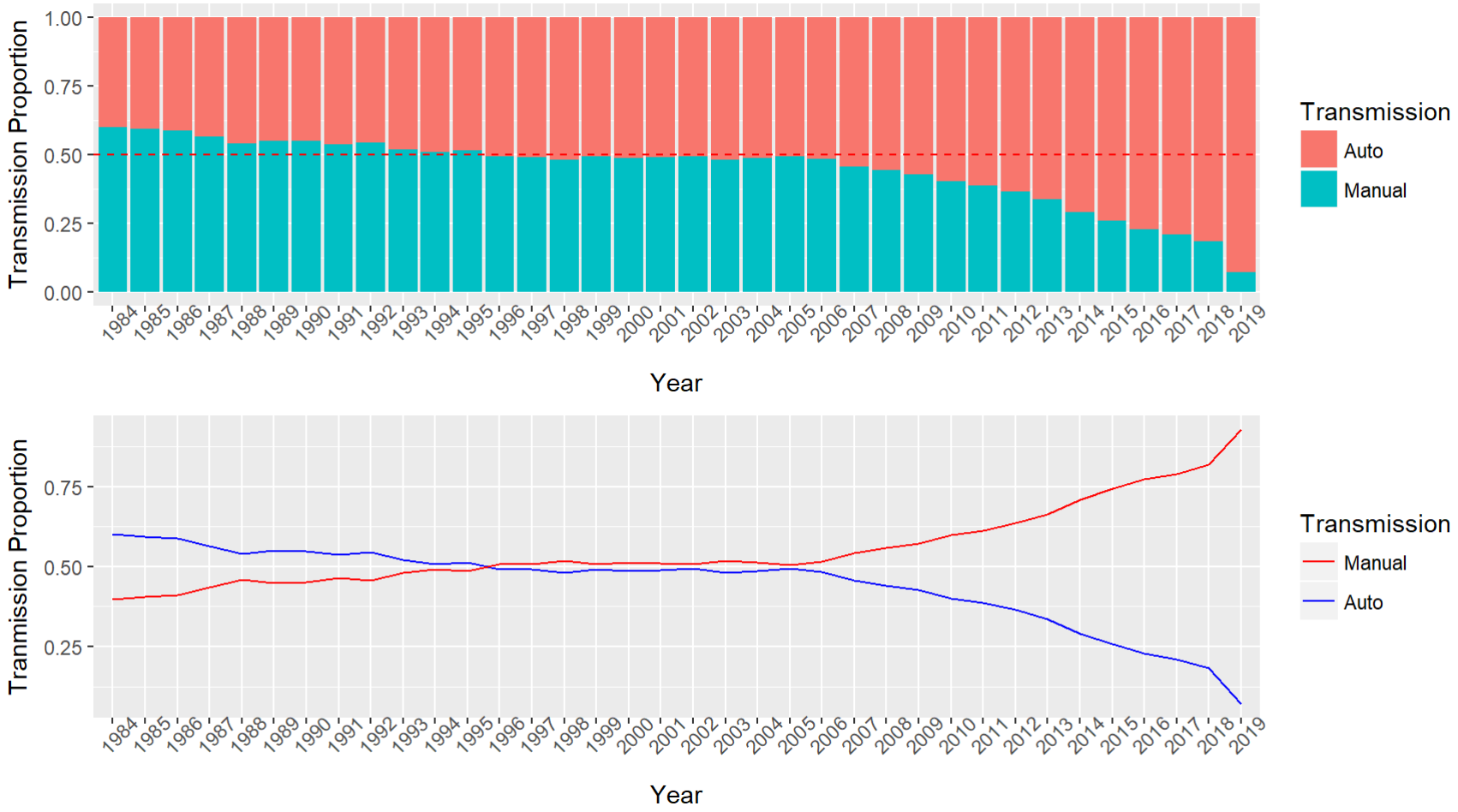
Let us see whether automatic engine are more efficient than manual engine for four cylinder engines, and how efficiency has changed over the years.

```
gascar4 <- subset(vehicles,cylinders == "4")

ggplot(data = gascar4,aes(factor(year),comb08))+geom_boxplot()+facet_wrap(~trany,ncol=1)+theme(axis.text.x = elem
ent_text(angle = 45))+labs(x="Year","Economy")
```



Looks like the manual car proportions has been decreasing over the time. Let us just visualize that.

```
p3 <- ggplot(data = gascar4,aes(factor(year),fill=trany))+geom_bar(position = "fill") + labs(x="Year",y="Transmis
sion Proportion",fill = "Transmission")+theme(axis.text.x = element_text(angle = 45))+geom_hline(yintercept = 0.5
,color="red",linetype=2)

p4 <- gascar4 %>% group_by(year) %>% summarise(proportionTrany = mean(ifelse(trany=="Manual",1,0))) %>% ggplot(ae
s(factor(year)))+geom_line(aes(y=proportionTrany,group=1,color="red"))+geom_line(aes(y=1-proportionTrany,group=1,
color="blue"))+labs(x="Year",y="Tranmission Proportion",colour="Transmission")+scale_color_manual(values=c("red",
"blue"),label=c("Manual","Auto"))+theme(axis.text.x = element_text(angle = 45))

grid.arrange(p3,p4)
```



Above plot indicates that although efficiency of manual cars is better that auto transmission cars but there proportion has been decreasing considerably.

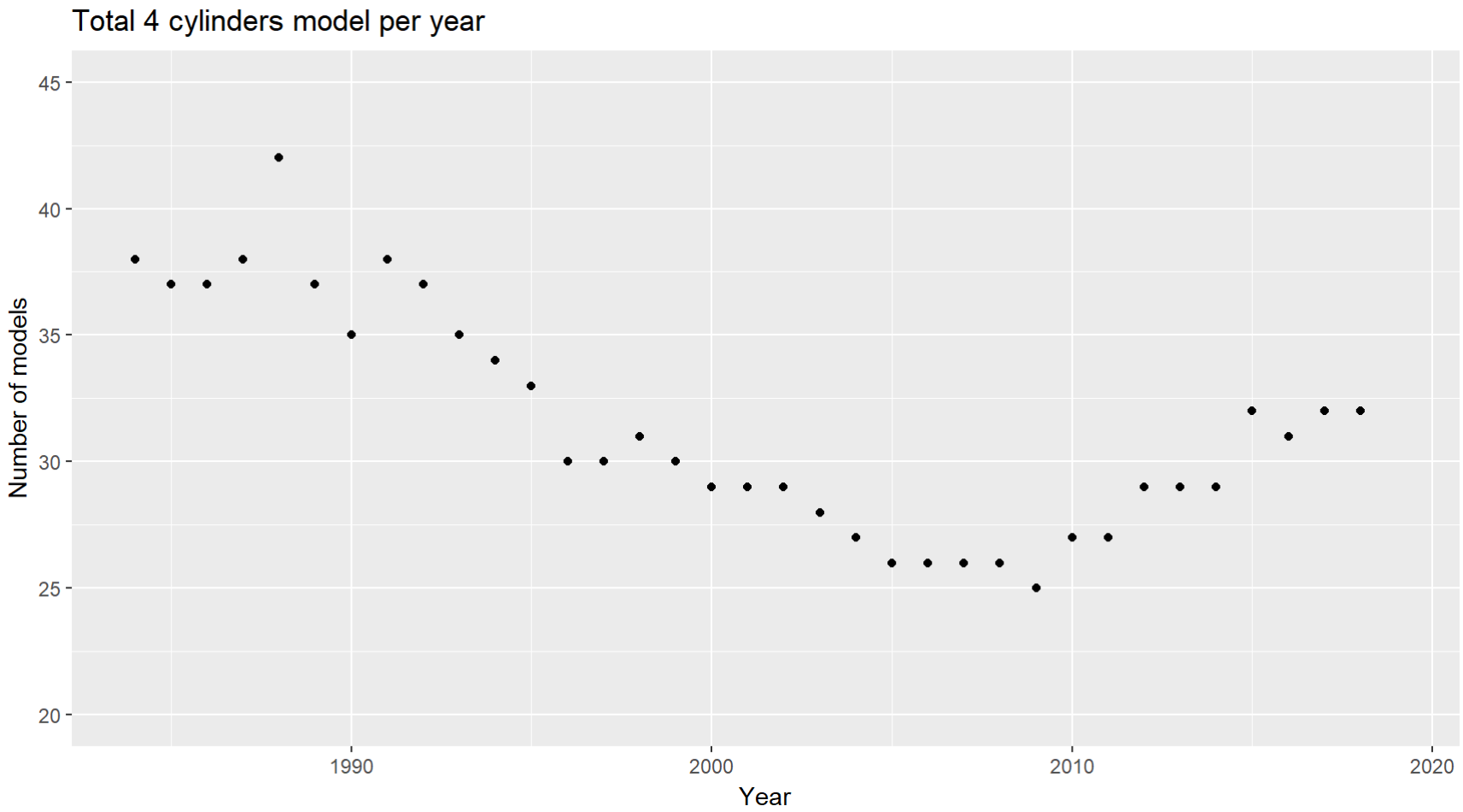
Till now, In our analysis we have identified following patters : -

- Fuel efficiency has improved over the years.
- Various reasons behind this fuel efficiency improvement could be following
- Increase in electricity motor vehicles
- Proportionally more smaller engine displacement vehicles
- Although manual cars seems to be better in fuel efficiency, there market share has been decreasing over the time.

Extra Analysis

We will look into how brands are improving when compared to 4 gas cylinders models.

```
carMakes <- gascar4 %>% group_by(year) %>% summarise(numberOfMakes = length(unique(make)))

ggplot(carMakes,aes(year))+geom_point(aes(y=numberOfMakes))+ylim(20,45)+labs(x="Year",y='Number of models',title=
"Total 4 cylinders model per year")
```



It looks like that the number of 4 cylinder models decreased consistently over the years till about 2011. After 2011, we see an upward trend.

Can we look at the brands who made 4 cylinder models every year of this study? We we select only those brands that have launched 4-gas cylinder models every year till 2016.

```
uniquemakes <- select (subset(gascar4,year<2015),c(year,make)) %>% group_by(year) %>% nest()
brands <- reduce(uniquemakes$data,intersect)
print(brands)
```

```
# A tibble: 13 x 1
   make
   <chr>
 1 Chrysler
 2 Honda
 3 Subaru
 4 Toyota
 5 Volkswagen
 6 Nissan
 7 Dodge
 8 Ford
 9 Mitsubishi
10 Chevrolet
11 Mazda
12 GMC
13 Jeep
```

How these brands have done with respect to fuel efficiency.

```
 subset(gascar4,make %in% brands$make) %>% group_by(year,make) %>% summarise(avg_fuel_effc = mean(comb08)) %>% gg
plot(aes(x=year,y=avg_fuel_effc))+geom_line()+ facet_wrap(~make)
```

We can see that overall fuel efficiency has increased over the years. However, some brands have been improving the fuel efficiency considerably more than others.