R ggplot2 Assignment 1

Yan Li

July 31 2015

We first have to load the Required packages.

#Load required packages in the user-defined library location  
.libPaths("C:/work/Rpackages/R3.2.1")  
suppressMessages(library(ggplot2, quietly = T))  
suppressMessages(library(dplyr, quietly = T))  
suppressMessages(library(mosaicData, quietly = T))

The data used for this assignment is mpg from ggplot2 package.

#Show the top 6 rows of the mpg data to know the data information  
head(mpg)

## manufacturer model displ year cyl trans drv cty hwy fl class  
## 1 audi a4 1.8 1999 4 auto(l5) f 18 29 p compact  
## 2 audi a4 1.8 1999 4 manual(m5) f 21 29 p compact  
## 3 audi a4 2.0 2008 4 manual(m6) f 20 31 p compact  
## 4 audi a4 2.0 2008 4 auto(av) f 21 30 p compact  
## 5 audi a4 2.8 1999 6 auto(l5) f 16 26 p compact  
## 6 audi a4 2.8 1999 6 manual(m5) f 18 26 p compact

Question 1: Which manufacturers makes which types of cars?

This is about the relationship between manufacture and class.

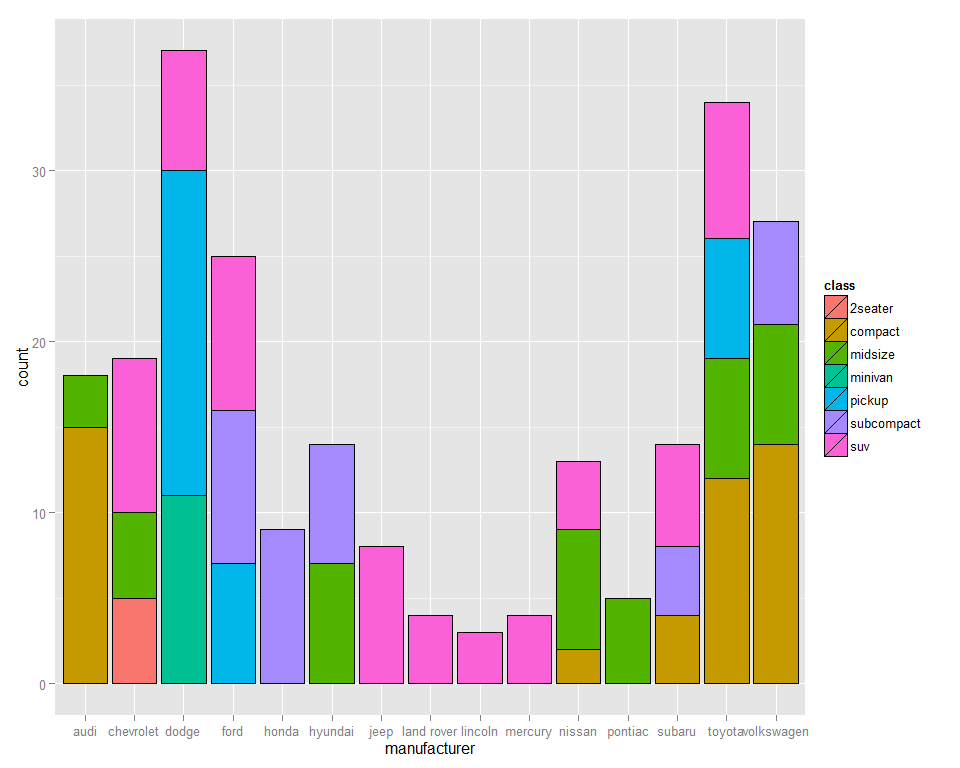
From the following table we will see the car type in each manufacturer for the mpg data.

table(mpg$manufacturer, mpg$class)

##   
## 2seater compact midsize minivan pickup subcompact suv  
## audi 0 15 3 0 0 0 0  
## chevrolet 5 0 5 0 0 0 9  
## dodge 0 0 0 11 19 0 7  
## ford 0 0 0 0 7 9 9  
## honda 0 0 0 0 0 9 0  
## hyundai 0 0 7 0 0 7 0  
## jeep 0 0 0 0 0 0 8  
## land rover 0 0 0 0 0 0 4  
## lincoln 0 0 0 0 0 0 3  
## mercury 0 0 0 0 0 0 4  
## nissan 0 2 7 0 0 0 4  
## pontiac 0 0 5 0 0 0 0  
## subaru 0 4 0 0 0 4 6  
## toyota 0 12 7 0 7 0 8  
## volkswagen 0 14 7 0 0 6 0

The bar plot summary the above in graph:

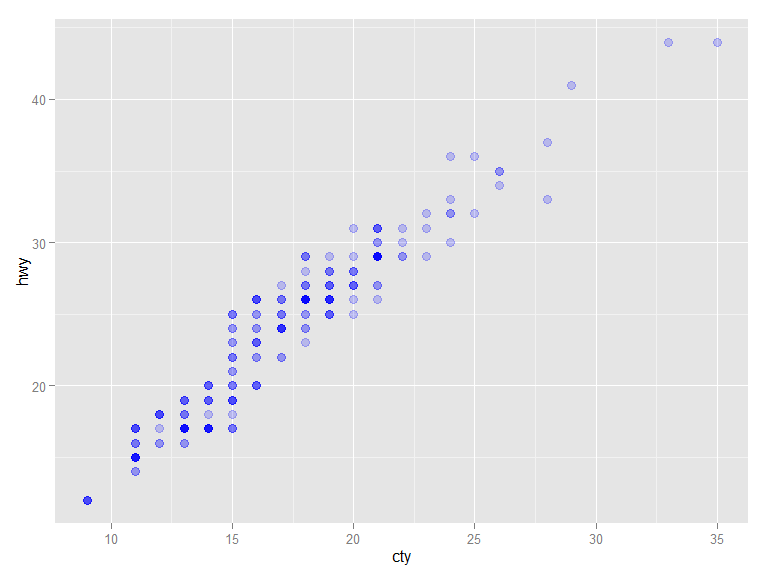
# bar plot by ggplot()  
ggplot(data = mpg, aes(x=manufacturer,fill=class))+ geom\_bar(binwidth = 0.2, color = "black")



Question 2: What is the relationship between city and highway mpg?

We can make a scatter plot to check the relationship of two variables.

qplot(cty,hwy,data = mpg, alpha = I(.2),size = I(5),color = I("blue"), shape = I(20))

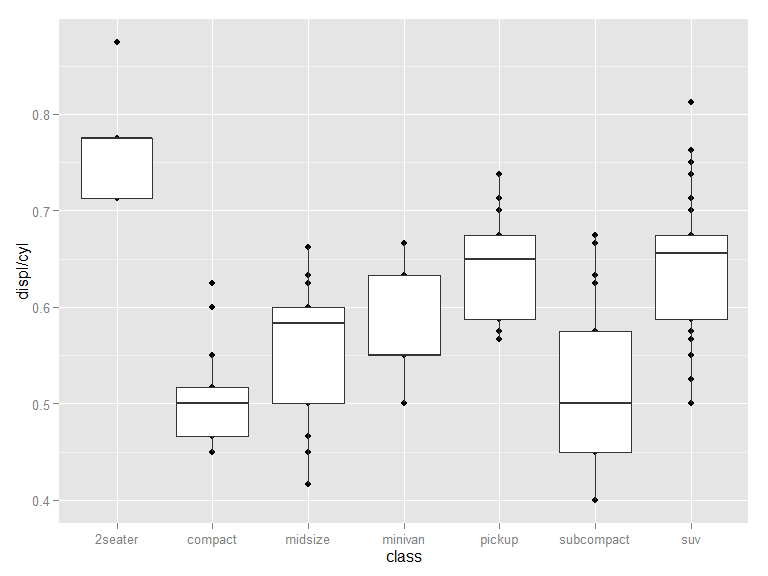


The plot shows a linear (increasing) relationship between city and highway mpg.

Question3: How does displacement per cylinder vary across car types?

This needs to include a calculation : displ/cyl in the plot.

# bar plot by qplot()  
qplot(class,displ/cyl,data = mpg)+geom\_boxplot()



The above plot shows: 2seater has the highest displacement per cylinder; compact and subcompact have the lowest displacement per cylinder.

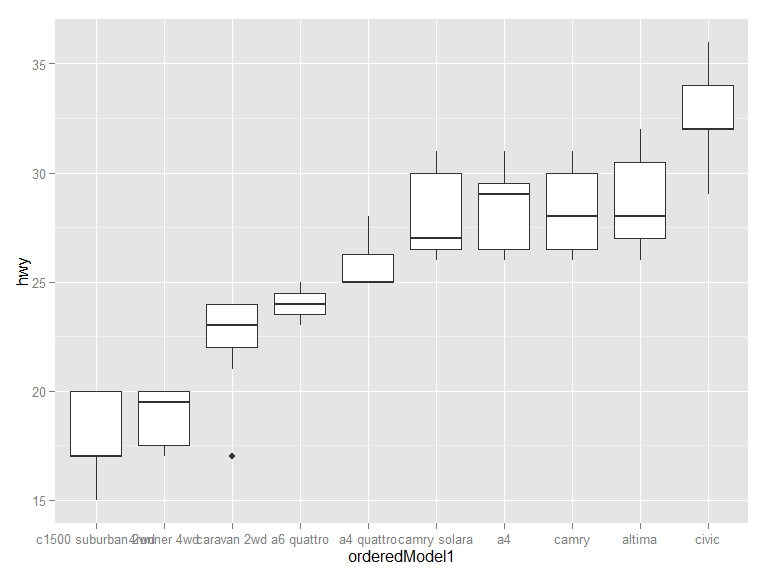
Question 4: Make a graphical league table ranking models by their fuel economy.

First, I rank all models alphabetically and found there are 38 of them. I divided them into 4 leagues, each of which has about 10 models.

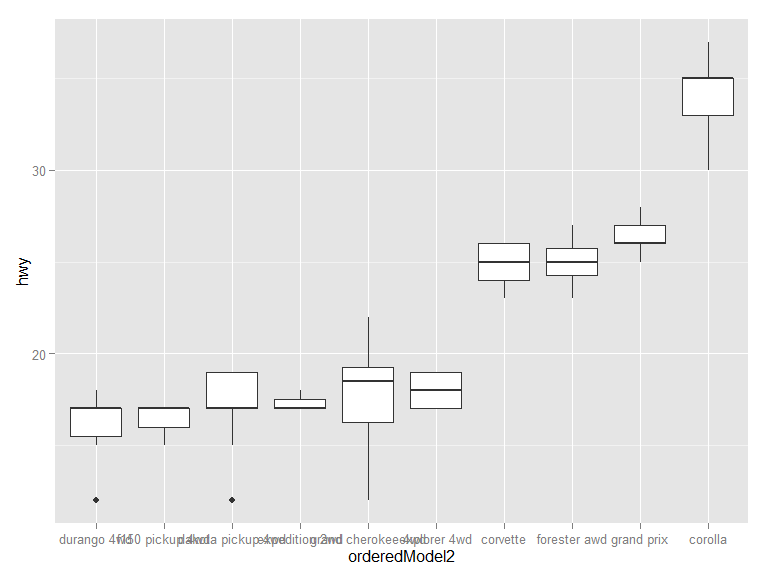
#Divide all models into 4 leagues:  
modelLevels <- levels(mpg$model)  
League1 <-modelLevels[1:10]   
League2 <-modelLevels[11:20]   
League3 <-modelLevels[21:30]   
League4 <-modelLevels[31:38]

Then I subset the mpg data according to the model names in each league and rank them based on hwy (high way miles per gallon).

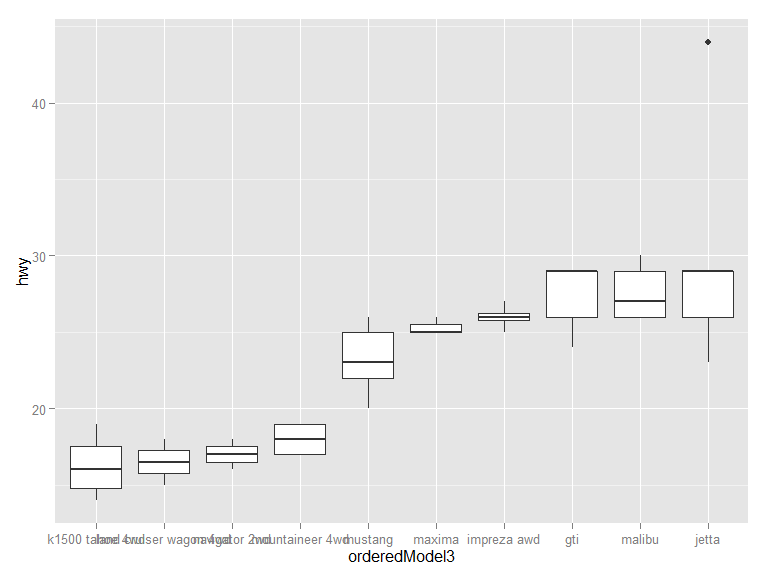
#The mdoel ranking of the first league  
data1 <- mpg[which(mpg$model %in% League1),c("model","hwy","fl")]  
orderedModel1 <- reorder(data1$model,data1$hwy)  
qplot(x=orderedModel1, y=hwy,data = data1,  
 geom =c( "boxplot"))



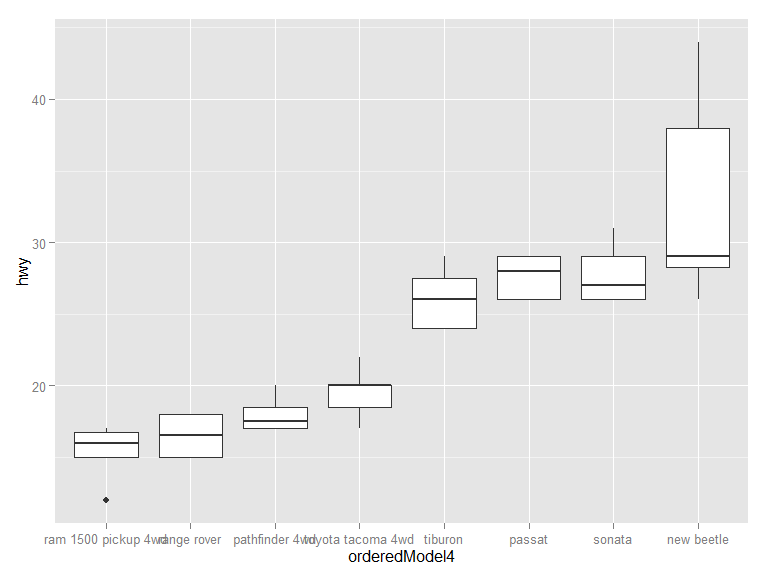
#The mdoel ranking of the second league  
data2 <- mpg[which(mpg$model %in% League2),c("model","hwy","fl")]  
orderedModel2 <- reorder(data2$model,data2$hwy)  
qplot(x=orderedModel2, y=hwy,data = data2,  
 geom =c( "boxplot"))



#The mdoel ranking of the third league  
data3 <- mpg[which(mpg$model %in% League3),c("model","hwy","fl")]  
orderedModel3 <- reorder(data3$model,data3$hwy)  
qplot(x=orderedModel3, y=hwy,data = data3,  
 geom =c( "boxplot"))



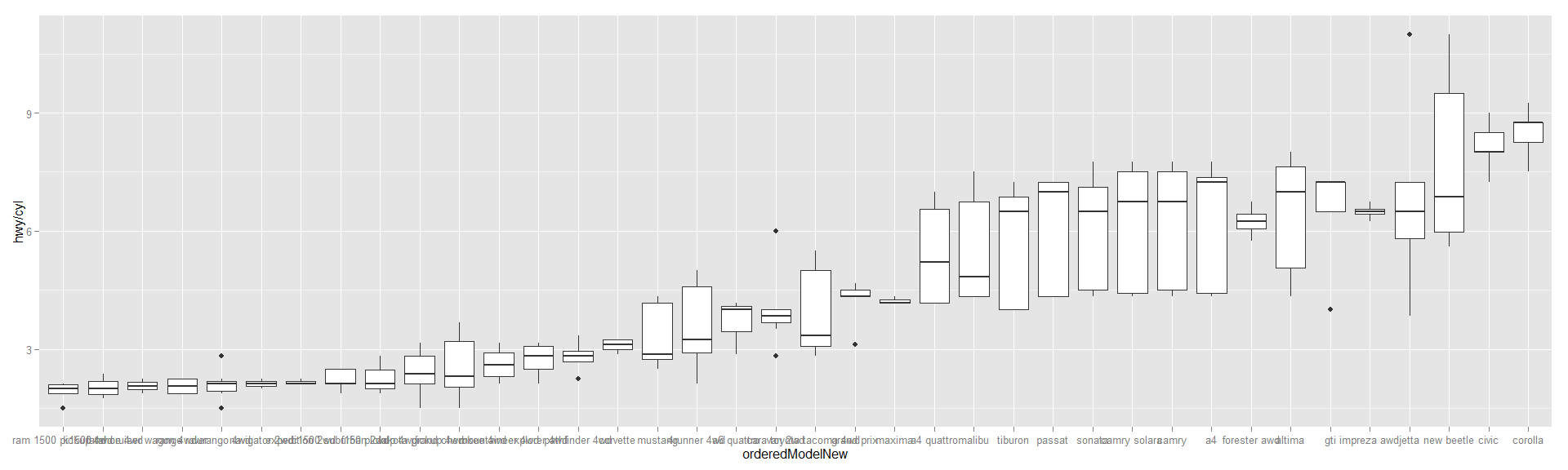
#The mdoel ranking of the fourth league  
data4 <- mpg[which(mpg$model %in% League4),c("model","hwy","fl")]  
orderedModel4 <- reorder(data4$model,data4$hwy)  
qplot(x=orderedModel4, y=hwy,data = data4,  
 geom =c( "boxplot"))



Question 5: Which cars have particularly good fuel economy given their engine size? What sort of cars are they?

I use hwy/cyl as a new measure to reorder the model.

orderedModelNew <- reorder(mpg$model,mpg$hwy/mpg$cyl)  
qplot(x=orderedModelNew, y=hwy/cyl,data = mpg,  
 geom =c( "boxplot"))



The high hwy/cyl means better fuel efficiency per cylinder. The top six ones are : corola, civic, new bettle, jetta,impreza awd and gti.Inorder to examine these cars, I extract them from the mpg data and check each feature:

#Best model names:  
bestModels <- c("corola","civic", "new bettle", "jetta","impreza awd","gti")  
  
#extract data containing these cars  
bestModels\_data<-mpg[which(mpg$model %in% bestModels),]

#Check the car class  
table(bestModels\_data$class)

##   
## 2seater compact midsize minivan pickup subcompact   
## 0 18 0 0 0 13   
## suv   
## 0

#Check the cylinder  
table(bestModels\_data$cyl)

##   
## 4 5 6   
## 26 2 3

#Check the fuel type  
table(bestModels\_data$fl)

##   
## c d e p r   
## 1 1 0 8 21

#Check the transmission  
table(bestModels\_data$trans)

##   
## auto(av) auto(l3) auto(l4) auto(l5) auto(l6) auto(s4)   
## 0 0 7 2 0 2   
## auto(s5) auto(s6) manual(m5) manual(m6)   
## 0 3 14 3

#Check the manufacturer  
table(bestModels\_data$manufacturer)

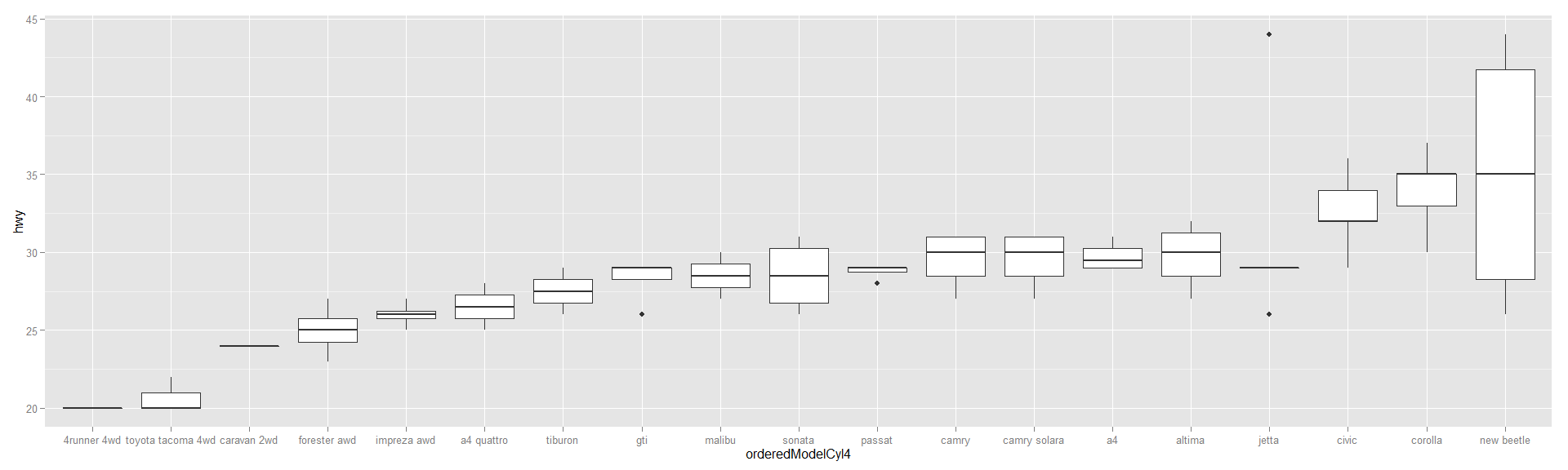
##   
## audi chevrolet dodge ford honda hyundai   
## 0 0 0 0 9 0   
## jeep land rover lincoln mercury nissan pontiac   
## 0 0 0 0 0 0   
## subaru toyota volkswagen   
## 8 0 14

Now I know these cars are most compact or subcompact cars, 4 cylinders,rugular fuel type and manual transimission. They are mainly from volswagen, honda and subra.

The other approach is that, we can subset the data into four groups based on cylinders then rank the fuel efficiency individualy.

The rank of 4 cylinders group:

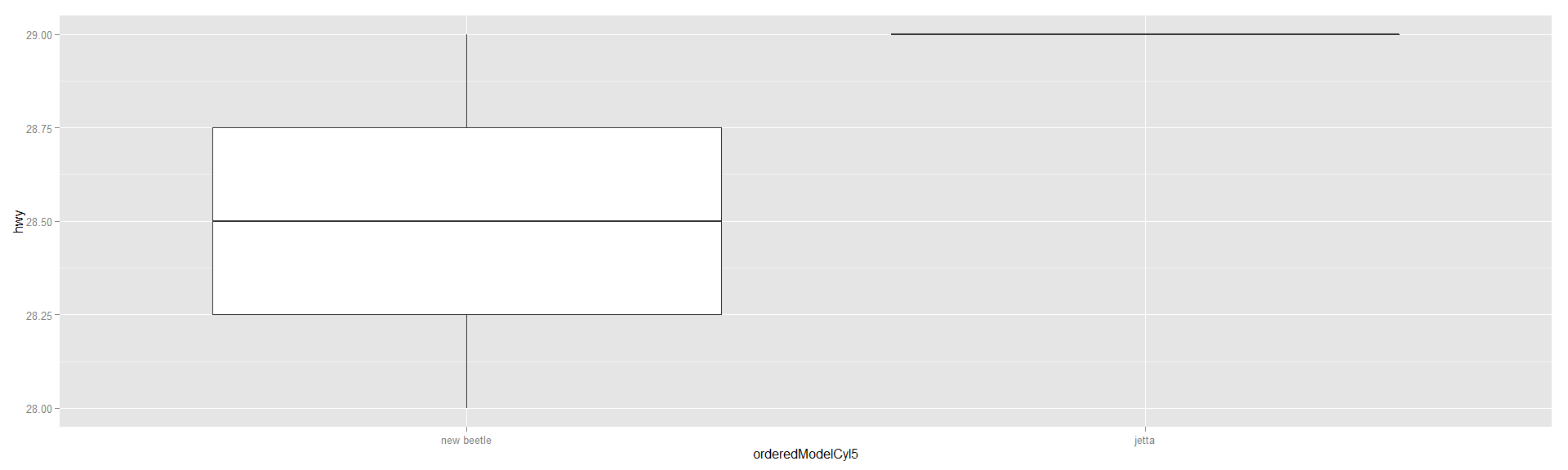
datacyl4<- mpg[which(mpg$cyl == 4),]  
orderedModelCyl4 <- reorder(datacyl4$model,datacyl4$hwy)  
qplot(x=orderedModelCyl4, y=hwy,data = datacyl4,  
 geom =c( "boxplot"))



The best car is new beetle.

The rank of 5 cylinders group:

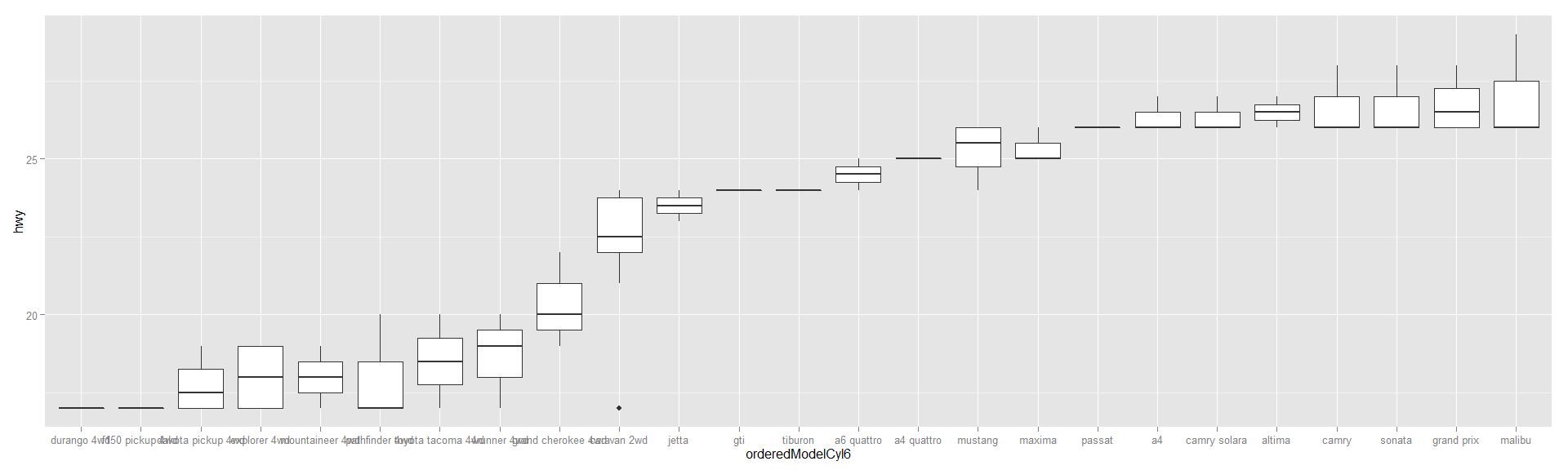
datacyl5<- mpg[which(mpg$cyl == 5),]  
orderedModelCyl5 <- reorder(datacyl5$model,datacyl5$hwy)  
qplot(x=orderedModelCyl5, y=hwy,data = datacyl5,  
 geom =c( "boxplot"))



The best one is jetta.

The rank of 6 cylinders group:

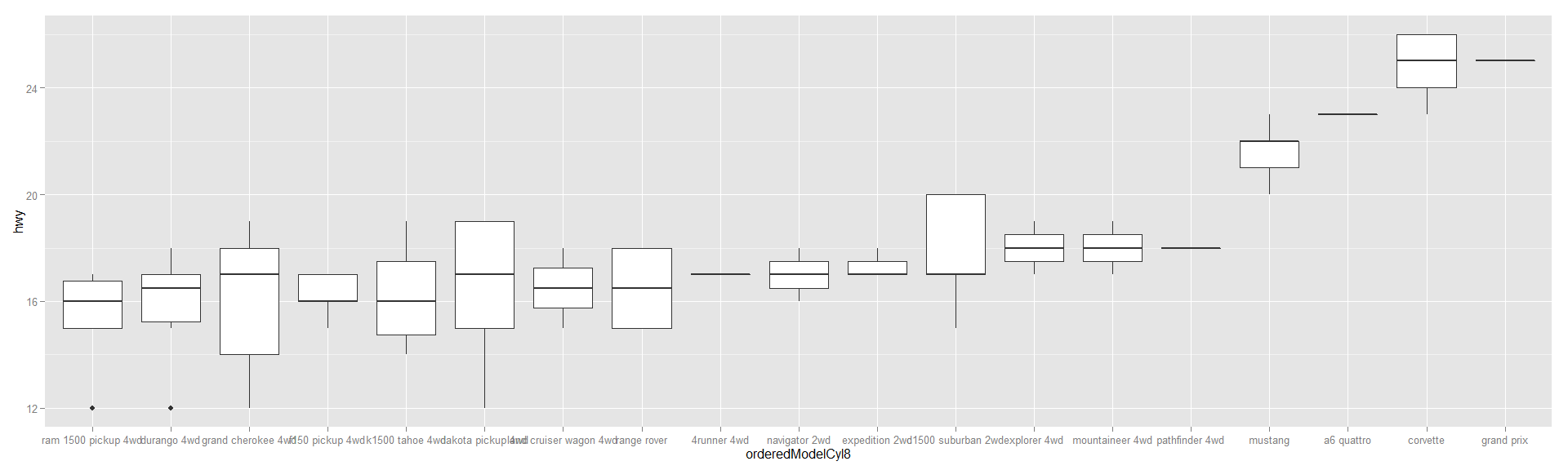
datacyl6<- mpg[which(mpg$cyl == 6),]  
orderedModelCyl6 <- reorder(datacyl6$model,datacyl6$hwy)  
qplot(x=orderedModelCyl6, y=hwy,data = datacyl6,  
 geom =c( "boxplot"))



The best one in 6 cylinders is malibu.

The rank of 8 cylinders group:

datacyl8<- mpg[which(mpg$cyl == 8),]  
orderedModelCyl8 <- reorder(datacyl8$model,datacyl8$hwy)  
qplot(x=orderedModelCyl8, y=hwy,data = datacyl8,  
 geom =c( "boxplot"))



The best one in 8 cylinders group is grand pix.

We can follow similar way to check these cars:

bestModels <- c("new bettle", "jetta",  
 "malibu","grand pix")  
  
bestModels\_data<-mpg[which(mpg$model %in% bestModels),]

table(bestModels\_data$class)

##   
## 2seater compact midsize minivan pickup subcompact   
## 0 9 5 0 0 0   
## suv   
## 0

table(bestModels\_data$cyl)

##   
## 4 5 6   
## 7 2 5

table(bestModels\_data$fl)

##   
## c d e p r   
## 0 1 0 2 11

table(bestModels\_data$trans)

##   
## auto(av) auto(l3) auto(l4) auto(l5) auto(l6) auto(s4)   
## 0 0 6 0 0 0   
## auto(s5) auto(s6) manual(m5) manual(m6)   
## 0 3 4 1

table(bestModels\_data$manufacturer)

##   
## audi chevrolet dodge ford honda hyundai   
## 0 5 0 0 0 0   
## jeep land rover lincoln mercury nissan pontiac   
## 0 0 0 0 0 0   
## subaru toyota volkswagen   
## 0 0 9

Question 6:Does any manufacturer stick out because of the fuel efficiency of their cars? What makes their fuel economy so good?

From question 5, we know volswagen is the best in fuel economy.