

Class Assignment: MPG tables and charts (edited with dplyr functions)

Jacob Burke

September 20th, 2019

Using the tidyverse (especially dplyr) for MPG

Recall assignment:

Tables should include a comparison of city and highway mileage by the class of car and the best three car models for city and highway mileage for all the years in which data is available.

Plot the data displaying as much of the data as you can. Put continuous variables on the axes and include a locally smoothed regression line to show the relationship with mileage. Also make a box plot comparing city and highway MPG by class of car.

Tables

```
data(mpg)      ## load the data

mpg$cyl <- as.factor(mpg$cyl)  ## convert discrete variables to factors so that they plot
mpg$drv <- as.factor(mpg$drv)  ## as different colors, not gradations of one color

mpg$cty <- as.double(mpg$cty)

mpg_1 <- select(mpg, cty, hwy, class )

mpg_1 %<>% group_by(class) %>%
  summarise(City=mean(cty),
            Highway=mean(hwy)) %>%
  arrange(class, City, Highway)

kable(mpg_1, format = "latex", booktabs=TRUE, digits = 2,      ## call kable to make the table
      col.names = c("Class", "City", "Highway"),
      caption = "Mean City and Highway MPG by Car Class" )

## Selecting city, year, and model, variables
mpgcity <- select(mpg, model, cty, year)

## filtering down to respective year
mpgcity_1999 <- filter(mpgcity, year == 1999)

## grouping this subset by the model, and getting each model's average Milage
mpgcity_1999 %<>% group_by(model) %>%
  summarise(Milage = mean(cty)) %>%
  arrange(model, Milage)

## ordering data frame in order of top Milage models
```

Table 1: Mean City and Highway MPG by Car Class

Class	City	Highway
2seater	15.40	24.80
compact	20.13	28.30
midsize	18.76	27.29
minivan	15.82	22.36
pickup	13.00	16.88
subcompact	20.37	28.14
suv	13.50	18.13

```

mpgcity_1999 <- mpgcity_1999[
  with(mpgcity_1999, order(-Milage)),
]

## rounding milage
mpgcity_1999$Milage <- round(mpgcity_1999$Milage, 2)

## Repeating process for mpghwy_1999
mpghwy <- select(mpg, model, hwy, year)

mpghwy_1999 <- filter(mpghwy, year == 1999)

mpghwy_1999 %<>% group_by(model) %>%
  summarise(Milage = mean(hwy)) %>%
  arrange(model, Milage)

mpghwy_1999 <- mpghwy_1999[
  with(mpghwy_1999, order(-Milage)),
]
mpghwy_1999$Milage <- round(mpghwy_1999$Milage, 2)

## ...mpgcity_2008
mpgcity_2008 <- filter(mpgcity, year == 2008)

mpgcity_2008 %<>% group_by(model) %>%
  summarise(Milage = mean(cty)) %>%
  arrange(model, Milage)

mpgcity_2008 <- mpgcity_2008[
  with(mpgcity_2008, order(-Milage)),
]
mpgcity_2008$Milage <- round(mpgcity_2008$Milage, 2)

## ...mpghwy_2008
mpghwy_2008 <- filter(mpghwy, year == 2008)

mpghwy_2008 %<>% group_by(model) %>%
  summarise(Milage = mean(hwy)) %>%

```

Table 2: Top 3 MPG Performing Cars: 1999, 2008

City 1999		Highway 1999		City 2008		Highway 2008	
Model	Milage	Model	Milage	Model	Milage	Model	Milage
new beetle	26.00	new beetle	35.00	corolla	27.0	corolla	36.00
civic	24.80	corolla	32.67	civic	24.0	civic	33.75
corolla	24.67	civic	31.60	gti	21.5	camry	30.00

```

      arrange(model, Milage)

mpghwy_2008 <- mpghwy_2008[
  with(mpghwy_2008, order(-Milage)),
]
mpghwy_2008$Milage <- round(mpghwy_2008$Milage, 2)

table3 <- cbind(mpgcity_1999 [1:3, ],
                mpghwy_1999[1:3, ],
                mpgcity_2008[1:3, ],
                mpghwy_2008[1:3, ])

colnames(table3) <- c('Model', 'Milage',
                     "Model", "Milage",
                     'Model', 'Milage',
                     "Model", "Milage"
                     )

kable(table3, digits = 2, format = "latex", booktabs=TRUE, ,caption = "Top 3 MPG Performing Cars: 1999,
  add_header_above(c("City 1999"=2,
                     "Highway 1999"=2,
                     "City 2008"=2,
                     "Highway 2008"=2))

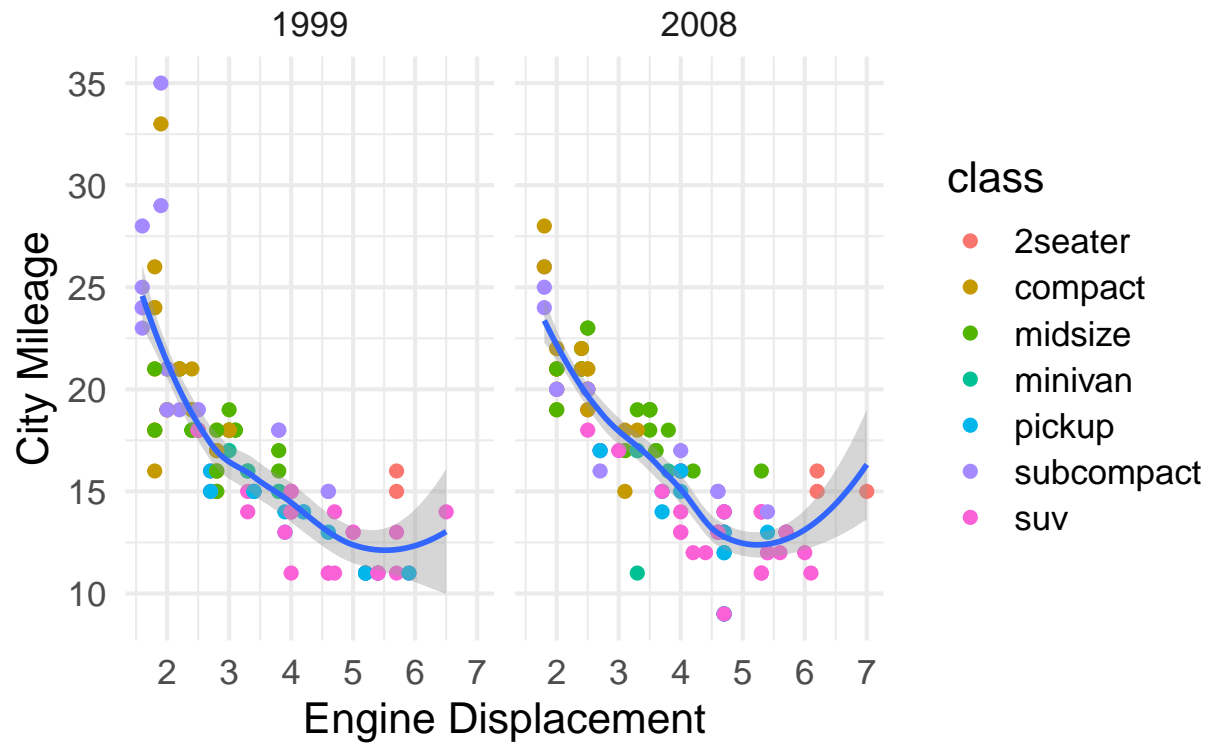
library(ggplot2)

ggplot(mpg) +
  aes(x = displ, y = cty) +
  geom_point(aes(color=class),size=2) + geom_smooth() +
  scale_color_hue() +
  theme_minimal(base_size = 16) +
  facet_wrap(vars(year)) +
  labs(x = "Engine Displacement",y = "City Mileage" ,title = "City MPG by Class of Car: 1999, 2008")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

```

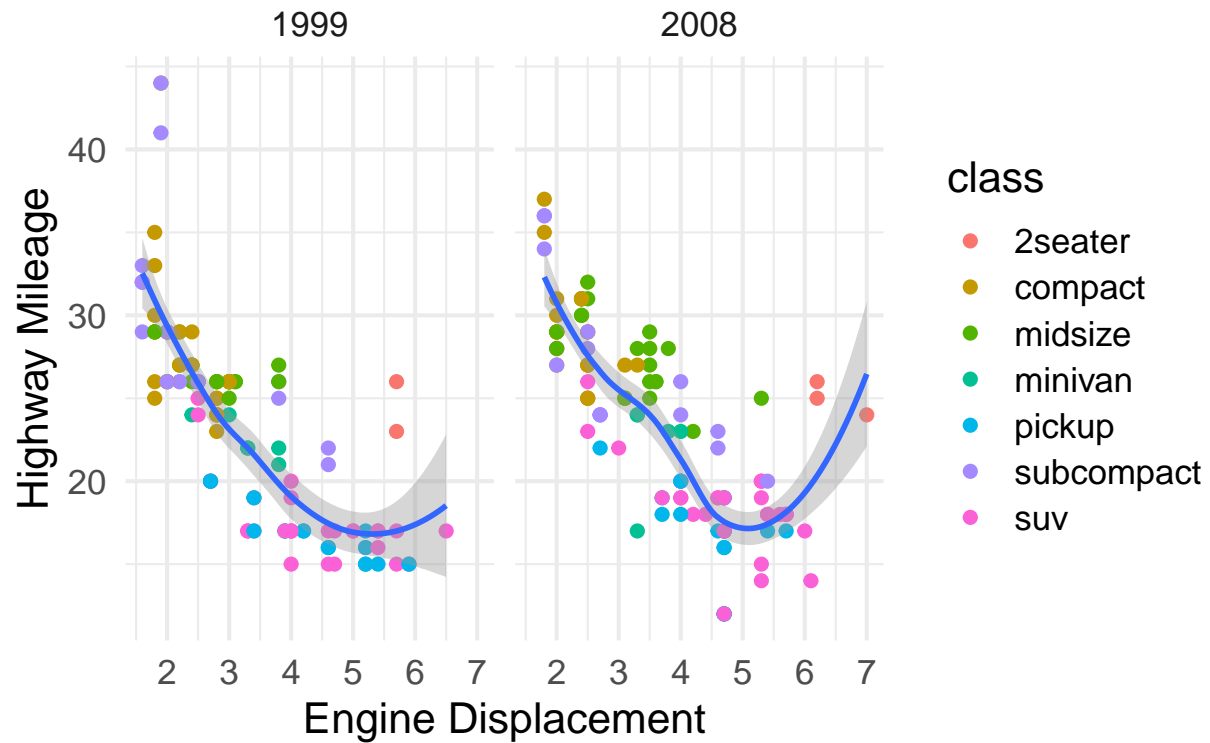
City MPG by Class of Car: 1999, 2008



```
ggplot(mpg) +
  aes(x = displ, y = hwy) +
  geom_point(aes(color=class),size=2) + geom_smooth() +
  scale_color_hue() +
  theme_minimal(base_size = 16) +
  facet_wrap(vars(year)) +
  labs(x = "Engine Displacement", y = "Highway Mileage", title = "Highway MPG by Class of Car: 1999, 2008")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Highway MPG by Class of Car: 1999, 2008

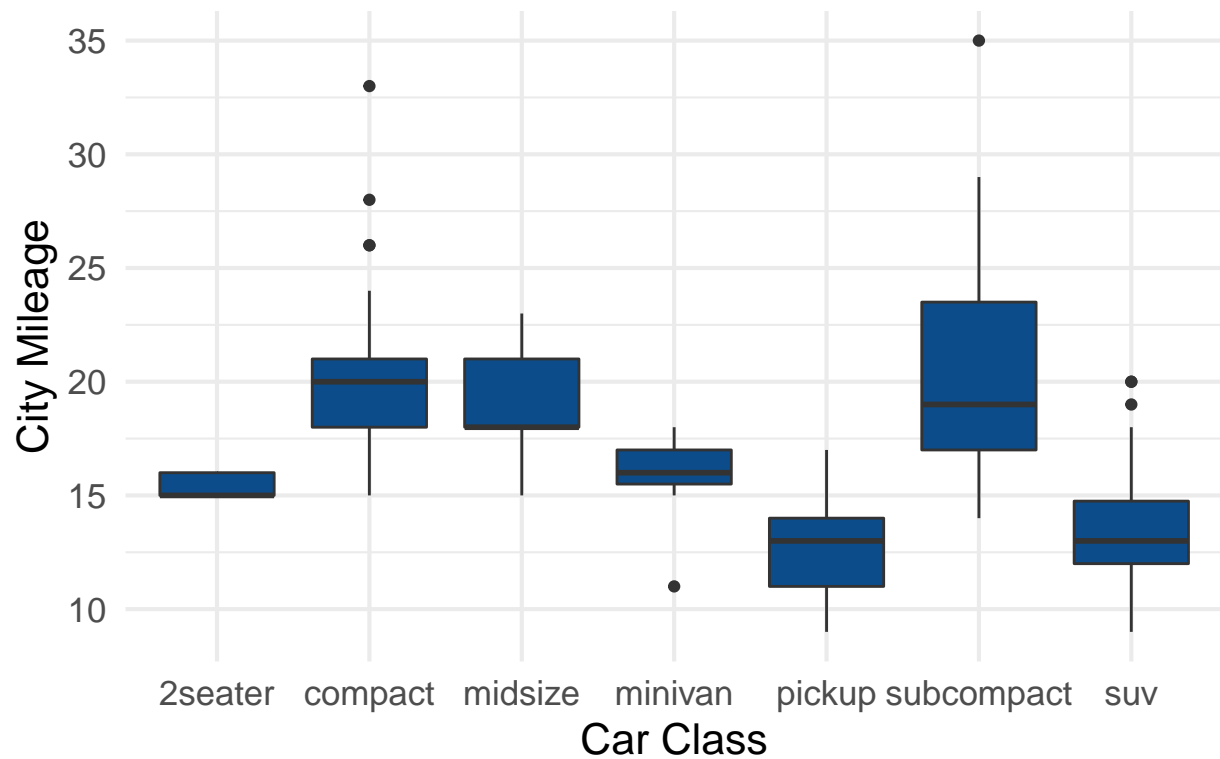


```
##esquisser(mpg)
```

Boxplots

```
ggplot(mpg) +  
  aes(x = class, y = cty) +  
  geom_boxplot(fill = "#0c4c8a") +  
  theme_minimal(base_size=16) +  
  labs(x = "Car Class", y = "City Mileage", title = "City MPG by Class of Car: 1999, 2008")
```

City MPG by Class of Car: 1999, 2008



```
ggplot(mpg) +  
  aes(x = class, y = hwy) +  
  geom_boxplot(fill = "#0c4c8a") +  
  theme_minimal(base_size=16) +  
  labs(x = "Car Class", y = "Highway Mileage", title = "Highway MPG by Class of Car: 1999, 2008")
```

Highway MPG by Class of Car: 1999, 2008

