

Class Assignment: MPG tables and charts

MA615

September 9, 2019

Class assignment:

Using the MPG dataset, which is available with the ggplot2 library, produce a document that includes the following elements: headers, text, tables, and plots.

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

This example is flawed even though the code works. Explain. And now that you know how to produce the table can you improve the code by using an **apply** functional?

```
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

class <- unique(mpg$class)     ## define a vector for car classes

cty_mean <- rep(0, length(class)) ## define variables to hold mean mileage
hwy_mean <- rep(0, length(class)) ## for each class of car

for(i in 1:length(class)){      ## calculate the mean city and
  cty_mean[i] = mean(mpg$cty[mpg$class==class[i]]) ## highway mileage for each
  hwy_mean[i] = mean(mpg$hwy[mpg$class==class[i]]) ## class of car
}

cty_mean <- round(cty_mean, 2)   ## round the results to 2 decimal places
hwy_mean <- round(hwy_mean, 2)

tbl_1 <- cbind(class, cty_mean, hwy_mean) ## form an array for the table

kable(tbl_1, digits = 2,          ## call kable to make the table
      col.names = c("Class", "City", "Highway"),
      caption = "Mean City and Highway MPG by Car Class" )
```

The next table is a bit tricky.

```
## You probably want to get started with a little investigation.
## Using unique(mpg$year), you find out that there are only two years - 1999, 2008.

mod_99 <- unique(mpg$model[mpg$year=='1999']) ## vector of models in 1999
```

Table 1: Mean City and Highway MPG by Car Class

Class	City	Highway
compact	20.13	28.3
midsize	18.76	27.29
suv	13.5	18.13
2seater	15.4	24.8
minivan	15.82	22.36
pickup	13	16.88
subcompact	20.37	28.14

```

mod_08 <- unique(mpg$model[mpg$year=='2008'])  ## vector of models in 2008

mod_99_cty_mean <- rep(0, length(mod_99))      ## setup vector for 1999 cty mileage means
mod_08_cty_mean <- rep(0, length(mod_08))      ## setup vector for 2008 cty mileage means
mod_99_hwy_mean <- rep(0, length(mod_99))      ## setup vector for 1999 hwy mileage means
mod_08_hwy_mean <- rep(0, length(mod_08))      ## setup vector for 2008 hwy mileage means

for(i in 1:length(mod_99)){

  mod_99_cty_mean[i] =
    mean(mpg$cty[mpg$year=='1999' & mpg$model == mod_99[i]])

  mod_08_cty_mean[i] =
    mean(mpg$cty[mpg$year=='2008' & mpg$model == mod_08[i]])

  ###
  mod_99_hwy_mean[i] =
    mean(mpg$hwy[mpg$year=='1999' & mpg$model == mod_99[i]])

  mod_08_hwy_mean[i] =
    mean(mpg$hwy[mpg$year=='2008' & mpg$model == mod_08[i]])

}

ind_srt_99_cty <- order(mod_99_cty_mean)
miles_cty_99 <- rev(mod_99_cty_mean[ind_srt_99_cty])
mod_cty_99 <- rev(mod_99[ind_srt_99_cty])

ind_srt_08_cty <- order(mod_08_cty_mean)
miles_cty_08 <- rev(mod_08_cty_mean[ind_srt_08_cty])
mod_cty_08 <- rev(mod_08[ind_srt_08_cty])

ind_srt_99_hwy <- order(mod_99_hwy_mean)
miles_hwy_99 <- rev(mod_99_hwy_mean[ind_srt_99_hwy])
mod_hwy_99 <- rev(mod_99[ind_srt_99_hwy])

```

Table 2: Top 3 MPG Models

cMod3l 99	cMilage 99	hmodel 99	hmilage 99	cModel 08	cMilage 08	hmodel 08	hmilage 08
new beetle	26	new beetle	35	corolla	27	corolla	36
civic	24.8	corolla	32.67	civic	24	civic	33.75
corolla	24.67	civic	31.6	gti	21.5	camry	30

```
ind_srt_08_hwy <- order(mod_08_hwy_mean)
miles_hwy_08 <- rev(mod_08_hwy_mean[ind_srt_08_hwy])
mod_hwy_08 <- rev(mod_08[ind_srt_08_hwy])
```

```
miles_cty_08 <- round(miles_cty_08, 2)
miles_cty_99 <- round(miles_cty_99, 2)
```

```
miles_hwy_08 <- round(miles_hwy_08, 2)
miles_hwy_99 <- round(miles_hwy_99, 2)
```

```
tbl_2 <- cbind(mod_cty_99[1:3],
               miles_cty_99[1:3],
               mod_hwy_99[1:3],
               miles_hwy_99[1:3],
               mod_cty_08[1:3],
               miles_cty_08[1:3],
               mod_hwy_08[1:3],
               miles_hwy_08[1:3])
```

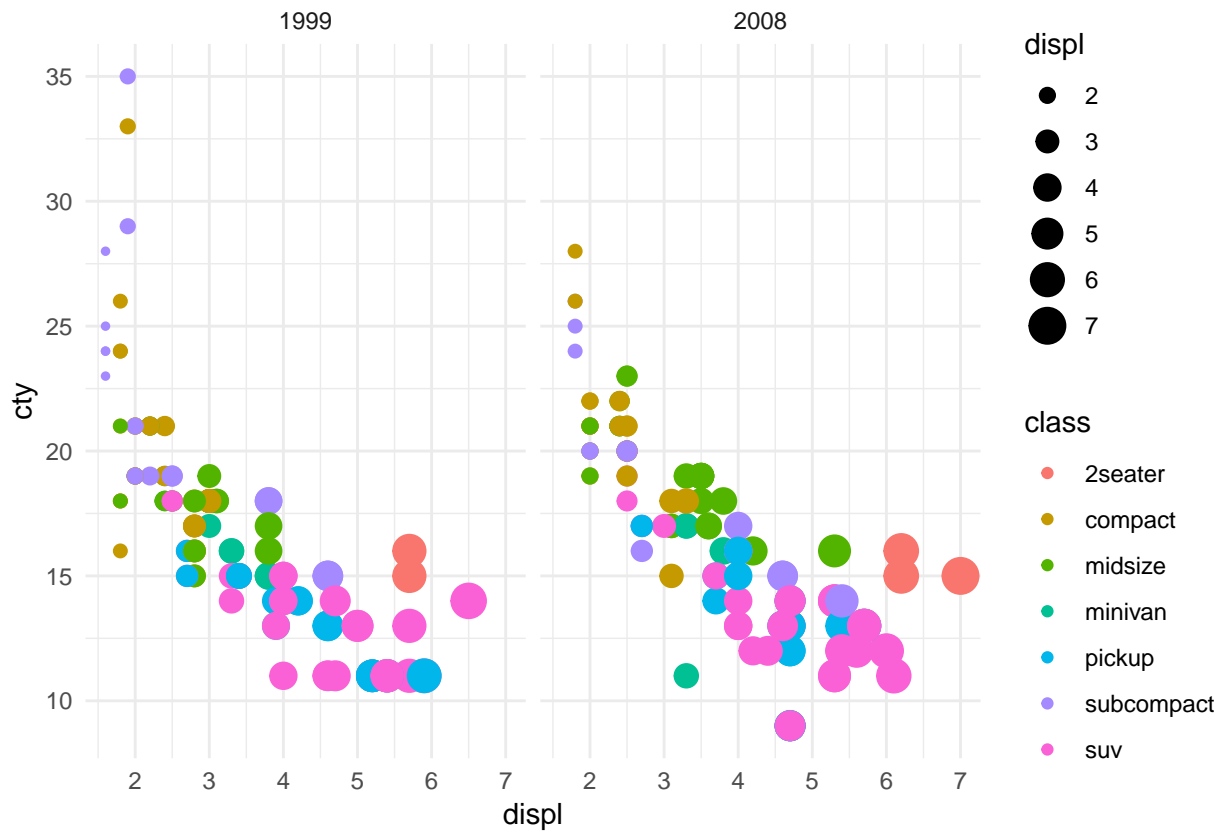
```
colnames(tbl_2) <- c('cMod3l 99', 'cMilage 99',
                    'hmodel 99', 'hmilage 99',
                    'cModel 08', 'cMilage 08',
                    'hmodel 08', 'hmilage 08'
                    )
```

```
kable(tbl_2, digits = 2, booktabs = TRUE, caption = "Top 3 MPG Models") #>% kable%>%add_header_above(c
```

```
#
#
#
```

```
library(ggplot2)
```

```
ggplot(mpg) +
  aes(x = displ, y = cty, colour = class, size = displ) +
  geom_point() +
  scale_color_hue() +
  theme_minimal() +
  facet_wrap(vars(year))
```



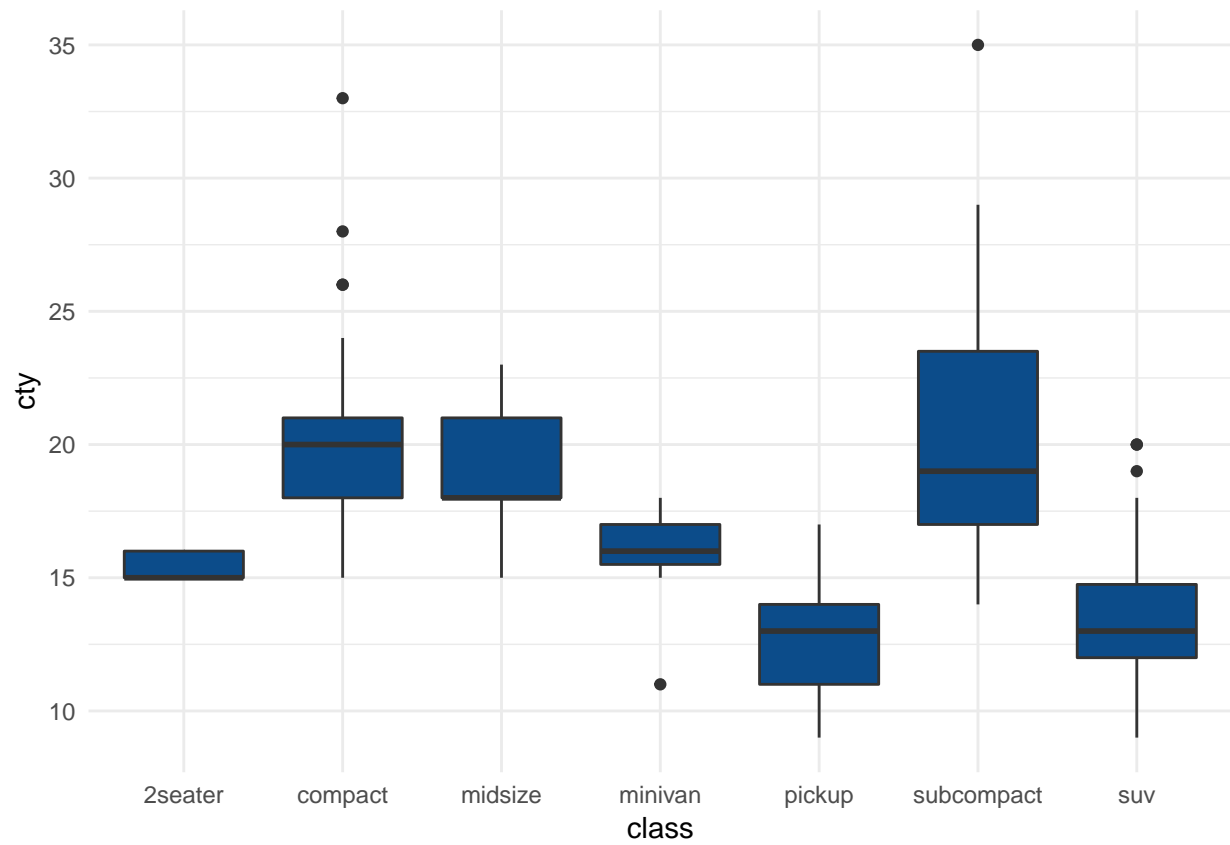
```
ggplot(mpg) +
  aes(x = displ, y = hwy, colour = class, size = displ) +
  geom_point() +
  scale_color_hue() +
  theme_minimal() +
  facet_wrap(vars(year))
```



Boxplots

```
#esquisser(mpg)

ggplot(mpg) +
  aes(x = class, y = cty) +
  geom_boxplot(fill = "#0c4c8a") +
  theme_minimal()
```



```
ggplot(mpg) +  
  aes(x = class, y = hwy) +  
  geom_boxplot(fill = "#0c4c8a") +  
  theme_minimal()
```

