car_viz

kosuke

2021/9/22

In this assignment, I will make a list of mean, variance, covariance and correlation of target variables of target cars in the data “mtcars”, and then make a smooth plot and a box plot. First of all, call the data“mtcars”. Then, focus on 5 variables (mpg, cyl, disp, hp and gears) of specific cars whose mpg <20

```
#To use ggplot function, firstly load package "tidyverse"
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# call built-in data mtcars.
data(mtcars)

# Select only car models where mpg<20 with the argument [xxx,]
mtcars_mpg2 <- mtcars[mtcars$mpg < 20,]

# Reduce the variables to mpg, cyl, disp, hp, gears with the argument[,xxx]
mtcars_mpg2 <- mtcars_mpg2[, c(1,2,3,4,10)]
```

“mtcars” is a data about 32 automobiles where fuel consumption and 10 aspects of automobile design and performance are shown. As I focused on cars with mpg<20, I reduced automobiles from 32 to 18, and then reduced variables from 11 to 5.

Next, calculate the mean, variance covariance matrix, and correlation of variables specified above, and make lists of these summary statistics.

```
#read the function you have already made and stored in "hand_functions.R" so that it can be used
#print out the R script with "echo = TRUE"
source(file = "hand_functions.R", echo = TRUE)
```

```
##
## > sum_special <- function(df_x) {
## +     try(if (!is.data.frame(df_x))
## +         stop("Input data must be a data frame."))
## +     sp_means <- apply(df_ .... [TRUNCATED]
```

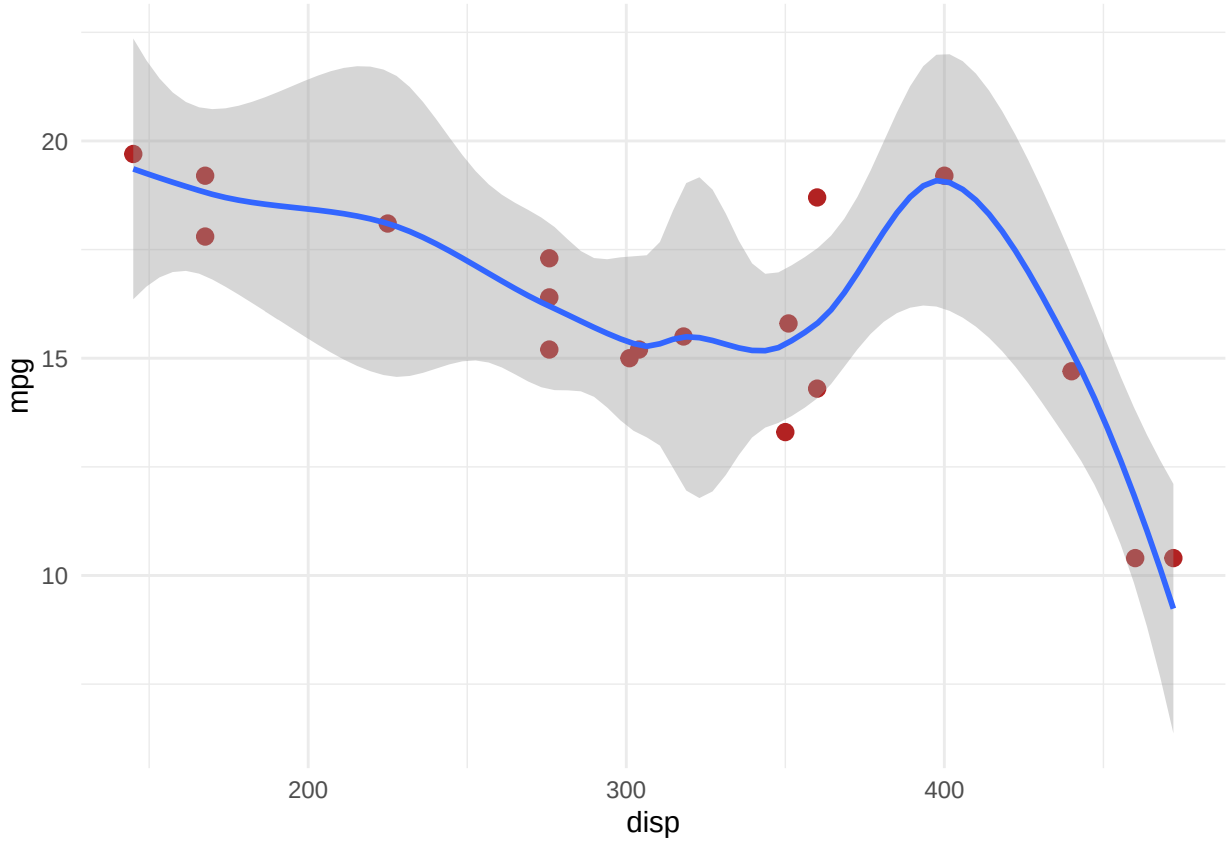
```
# Now use the function from hand_functions.R
sp_out <- sum_special(mtcars_mpg2)
```

The variance of cyl and gear were very small, because cyl takes just two values of 6 and 8, and gear also takes only 3 values of 3,4 and 5. There was a negative correlation between mpg and any of cyl, disp, and hp. From this result, we can infer that the more output power the car has, the less efficient it gets.

Next, make a dot plot in which take displacement as x axis and Miles per gallon as y axis in the mtcars_mpg2 data, and fit a smooth line.

```
#Use ggplot function to make a plots and smooth line.
ggplot(mtcars_mpg2) +
  aes(x = disp, y = mpg) +
  geom_point(shape = "bullet", size = 4L, colour = "#B22222") +
  #specify the shape, size and color of the plots
  geom_smooth(span = 0.5) +
  #specify each window where a weighted average is calculated
  theme_minimal() #specify a minimalistic theme with no background annotations
```

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

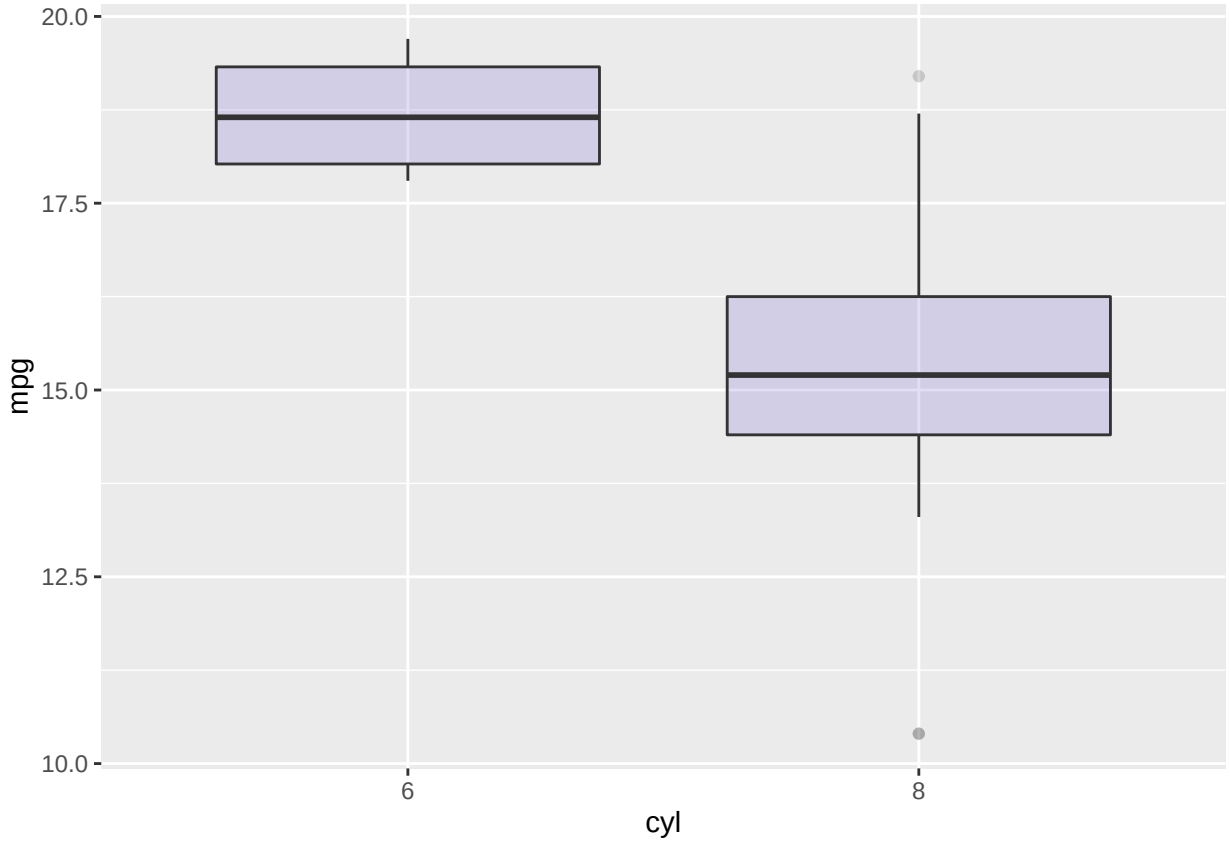


We can see a negative slope, and it gets steep at the value of or over 400 of disp. The negative slope can also be suggested by the result of the negative correlation between mpg and disp.

Finally, make a box plot of mpg for each cyl.

```
ggplot(mtcars_mpg2, aes(x=as.factor(cyl), y=mpg)) +
  #The original cyl is taken as continuous variable
  #To change the type of cyl from continuous to categorical with as.factor argument
  #With this adjustment, ggplot produce box plot for each cyl
  geom_boxplot(fill="slateblue", alpha=0.2) +
  # specify the color and transparency of the box plot
```

```
xlab("cyl") #specify the title of x axis
```



The median mpg for 6 cylinders was higher than that for 8 cylinders. This can be also suggested from the result of the negative correlation between mpg and cyl.