Untitled

iamqb

9/23/2021

```r
library(tidyverse)
```

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

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     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()
```

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

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



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

# read the R file hand_functions.R so that it can be used
# notice that 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]
```
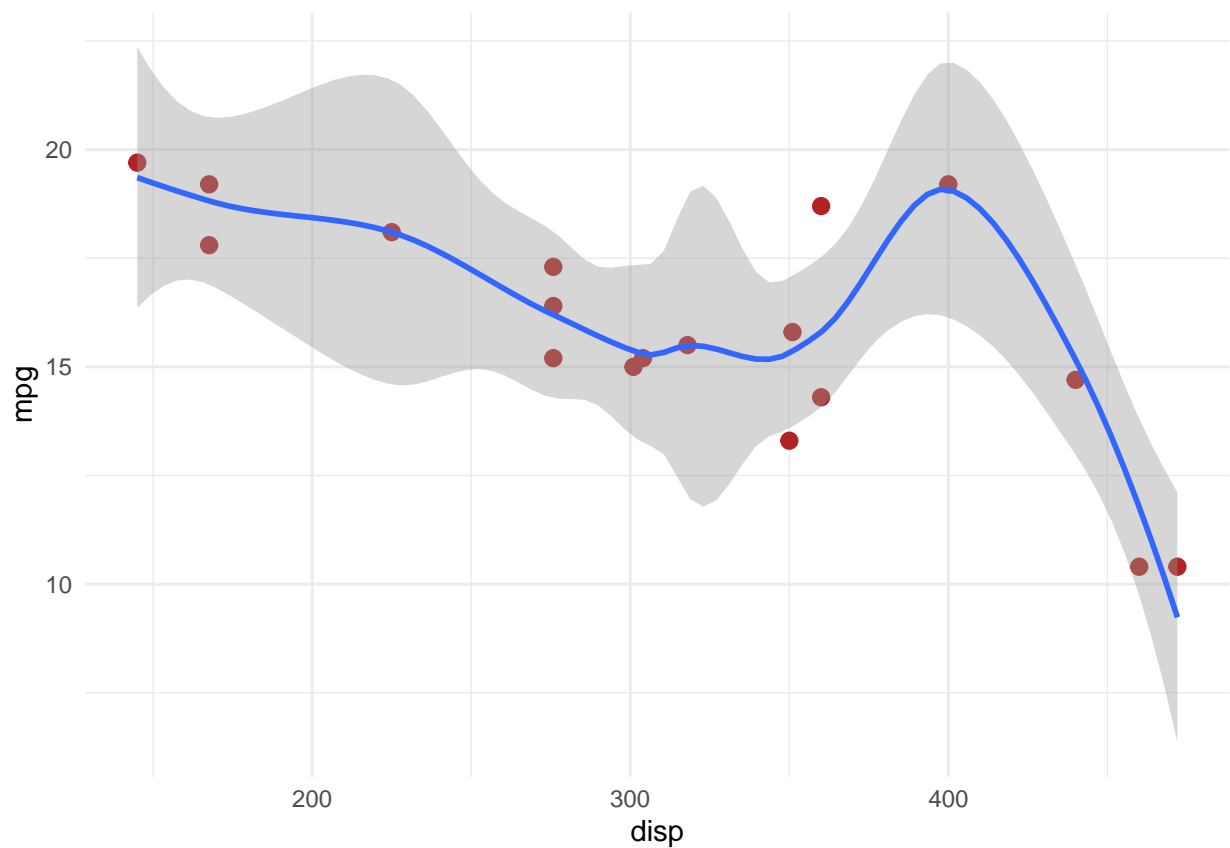
```r
# Now use the function from hand_functions.R
sp_out <- sum_special(mtcars_mpg2)
```

#First graph:

```
#library(esquisse)
#
#esquisser(data = mtcars_mpg2, viewer = "browser")

#plot that scatterplot and draw a smooth line
ggplot(mtcars_mpg2) +
  aes(x = disp, y = mpg) +
  geom_point(shape = "bullet", size = 4L, colour = "#B22222") +
  geom_smooth(span = 0.5) +
  theme_minimal()
```
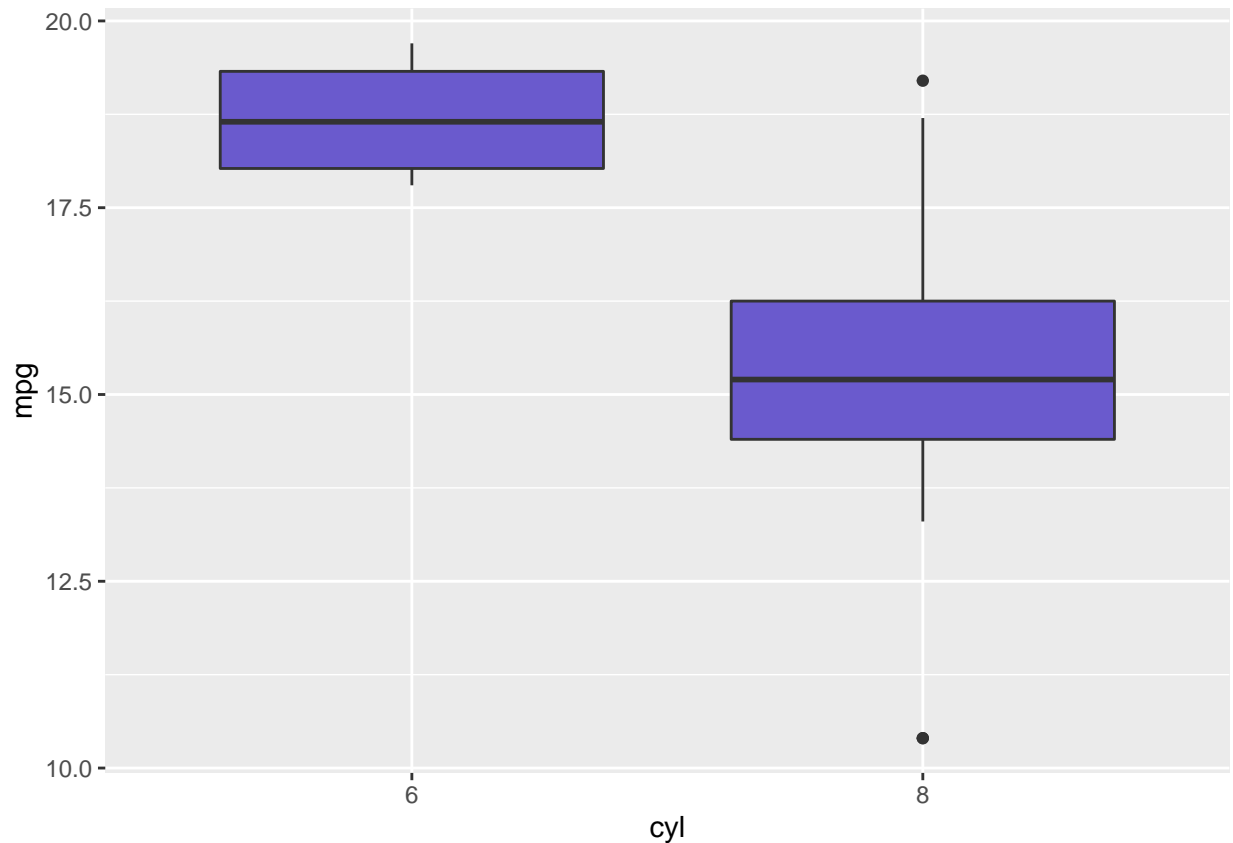
`geom_smooth()` using method = 'loess' and formula 'y ~ x'

#Second graph: boxplot mpg ~ as.factor(cyl)

```
#abjust the 'cyl' to factor, then graph the boxplot
ggplot(mtcars_mpg2, aes(x=as.factor(cyl), y=mpg)) +
  geom_boxplot(fill="slateblue", alphastr=0.2) +
  xlab("cyl")
```

Warning: Ignoring unknown parameters: alphastr

note that this boxplot cannot be made with esquisse() unless the data is adjusted. What adjustment is needed?

The data 'cyl' is numeric in the first ggplot graph, while in the second graph we use 'as.factor' to convert a vector from numeric to factor. Then the data is adjusted to the factor so that we could plot the box plot based on the factor of 'cyl'.

#What I learned in this exploration? First important thing I learned is that we need to adjust the data in an appropriate way so that we could display the data and make comparison. And there are multiple types of data like continuous, discrete, factor and so forth which we need to be careful. The ggplot is really important in R language to plot graphs.