Assignment1

MA_615

9/23/2021

```
#load the data
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
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)]
```

I learned how read the data that I need from a big dataset, and reduce variables to create a new dataset which is more simple and easy to read.

```
# 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]
```

```
# Now use the function from hand_functions.R

sp_out <- sum_special(mtcars_mpg2)

#using the function to calculate the summary statistics of the data frame
```

I learned that we can use formula that was save in a file to do computation which will save a lot of time. And that also means that we do not need to do a long calculation in our rmd file to find a function that fit into our analysis, which make our rmd more concise. In addition, if we need to use a formula repeatedly, a function file that is ready to use will be really helpful.

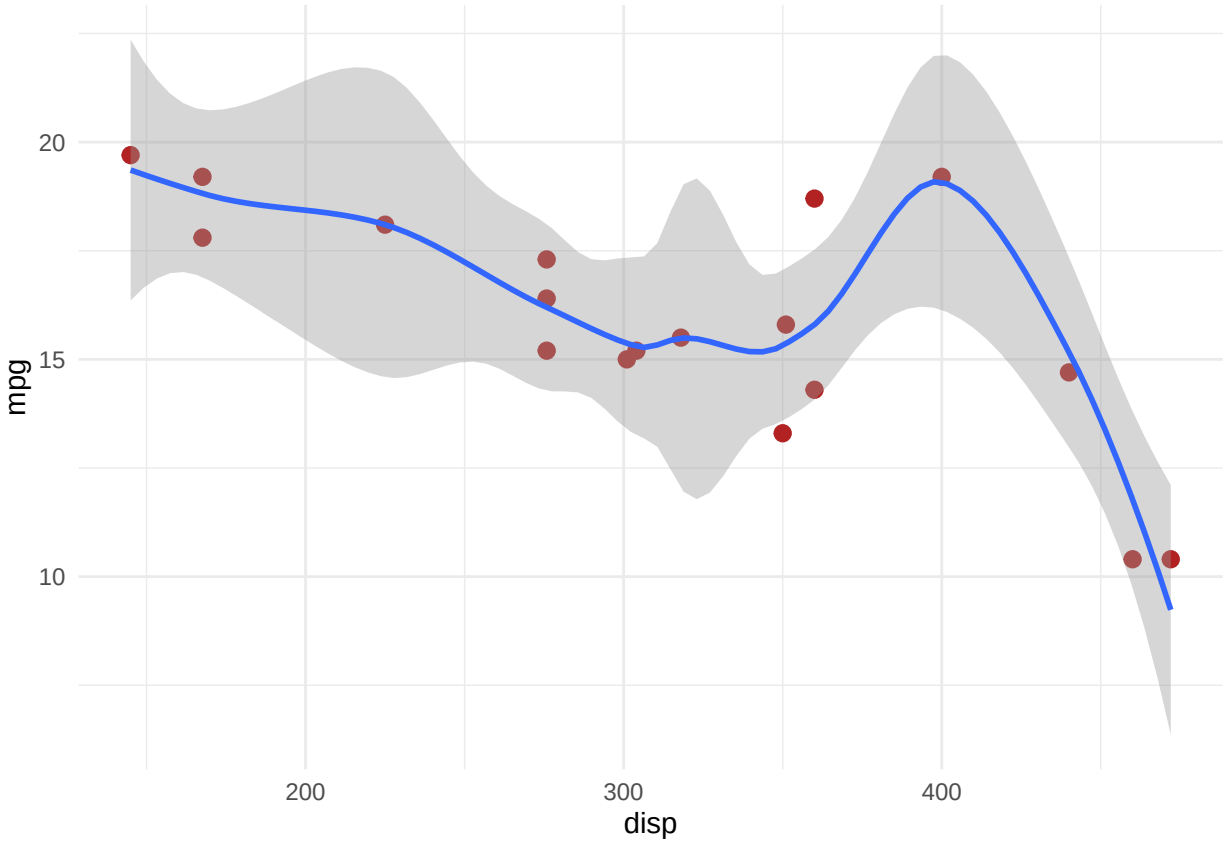
```
library(esquisse)

#esquisser(data = mtcars_mpg2, viewer = "browser")


# graph the ggplot in esquisse as same as needed
#then copy the code from esquisse and insert to my script here.

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'
```



I learned how to use esquisse to make simple plots using esquisse and apply different features or limitation to the plots.

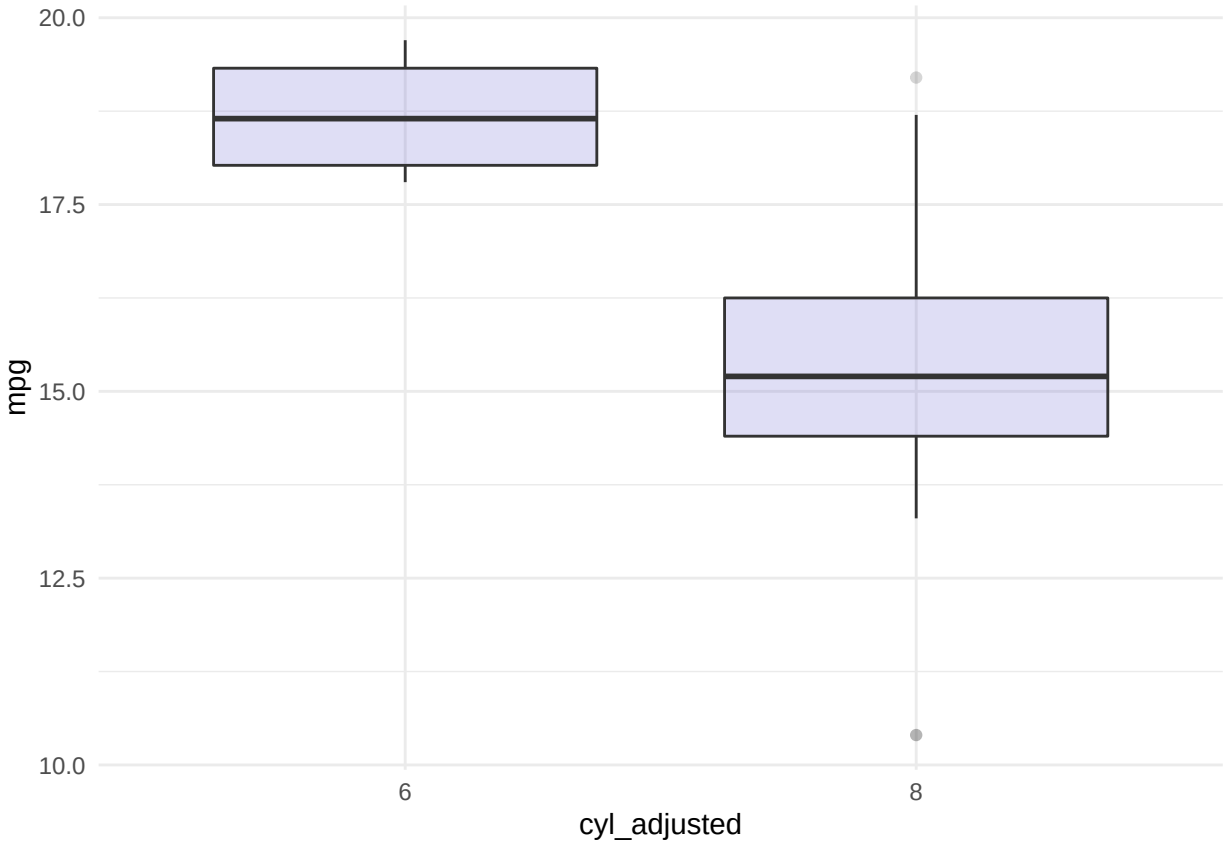
```
# note that this boxplot cannot be made with esquisse() unless
# the data is adjusted.  What adjustment is needed?
```

```r
#adjust cyl as a factor so that esquisse can read it
cyl_adjusted<- as.factor(mtcars_mpg2$cyl)

#group two data frames with same number of rows into a single data frame
mtcars_mpg2 <-cbind(mtcars_mpg2, cyl_adjusted)

#draw the plot in esquisse
#esquisser(data = mtcars_mpg2, viewer = "browser")


ggplot(mtcars_mpg2) +
 aes(x = cyl_adjusted, y = mpg) +
 geom_boxplot(shape = "circle", fill = "#625ACD", alpha=0.2) +
 theme_minimal()
```



I learned how to adjust my data frame when esquisse can read my data frame as a factor. And also learned how to combine two data frames with same length of rows into a single data frame.