MA615Assignment1

```
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)
```

Tidyverse is a coherent packet operating system for data manipulation, exploration and visualization with a common design philosophy.

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

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
```

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

```
##                      mpg cyl  disp  hp gear
## Hornet Sportabout   18.7   8 360.0 175    3
## Valiant             18.1   6 225.0 105    3
## Duster 360          14.3   8 360.0 245    3
## Merc 280            19.2   6 167.6 123    4
## Merc 280C           17.8   6 167.6 123    4
## Merc 450SE          16.4   8 275.8 180    3
## Merc 450SL          17.3   8 275.8 180    3
## Merc 450SLC         15.2   8 275.8 180    3
## Cadillac Fleetwood  10.4   8 472.0 205    3
## Lincoln Continental 10.4   8 460.0 215    3
## Chrysler Imperial   14.7   8 440.0 230    3
## Dodge Challenger    15.5   8 318.0 150    3
## AMC Javelin         15.2   8 304.0 150    3
## Camaro Z28          13.3   8 350.0 245    3
## Pontiac Firebird    19.2   8 400.0 175    3
## Ford Pantera L      15.8   8 351.0 264    5
## Ferrari Dino        19.7   6 145.0 175    5
## Maserati Bora       15.0   8 301.0 335    5
```

```
#just a wrong try: class(mtcars$cyl)
#just a wrong try: x1<-as.factor(mtcars$cyl)
#just a wrong try: or x1<-as.factor(mtcars$cyl)
#just a wrong try: class(x1)
#just a wrong try: mtcars_mpg2<-data.frame(mtcars_mpg2)
#just a wrong try: x2<-data.frame(x1)
#just a wrong try: names(x2)<-c("factor")
#just a wrong try: x3<-cbind(mtcars_mpg2,xx)
#cannot bind, because the number are not same.
#just a wrong try: x3

#wrong try again: mtcars_mpg2<-data.frame(mtcars_mpg2)
#wrong try again: xx<-as.factor(mtcars_mpg2$cyl)
#wrong try again: x1<-data.frame(xx)
#wrong try again: x2<-cbind(mtcars_mpg2,x1)
```

Select uses dataframe[dataframe$onedata < x,] Reduce uses dataframe[,c(residual)]

```
# 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]
```

“hand_functions.R” is a function and you need to run it so that you can operate.

Echo =TRUE means they will be included in the final rendered version. It will specify any global settings to be applied to the R Markdown script.

Echo =FALSE also is a parameter that Knitr will recognize, means the code itself will not appear in the final document.

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

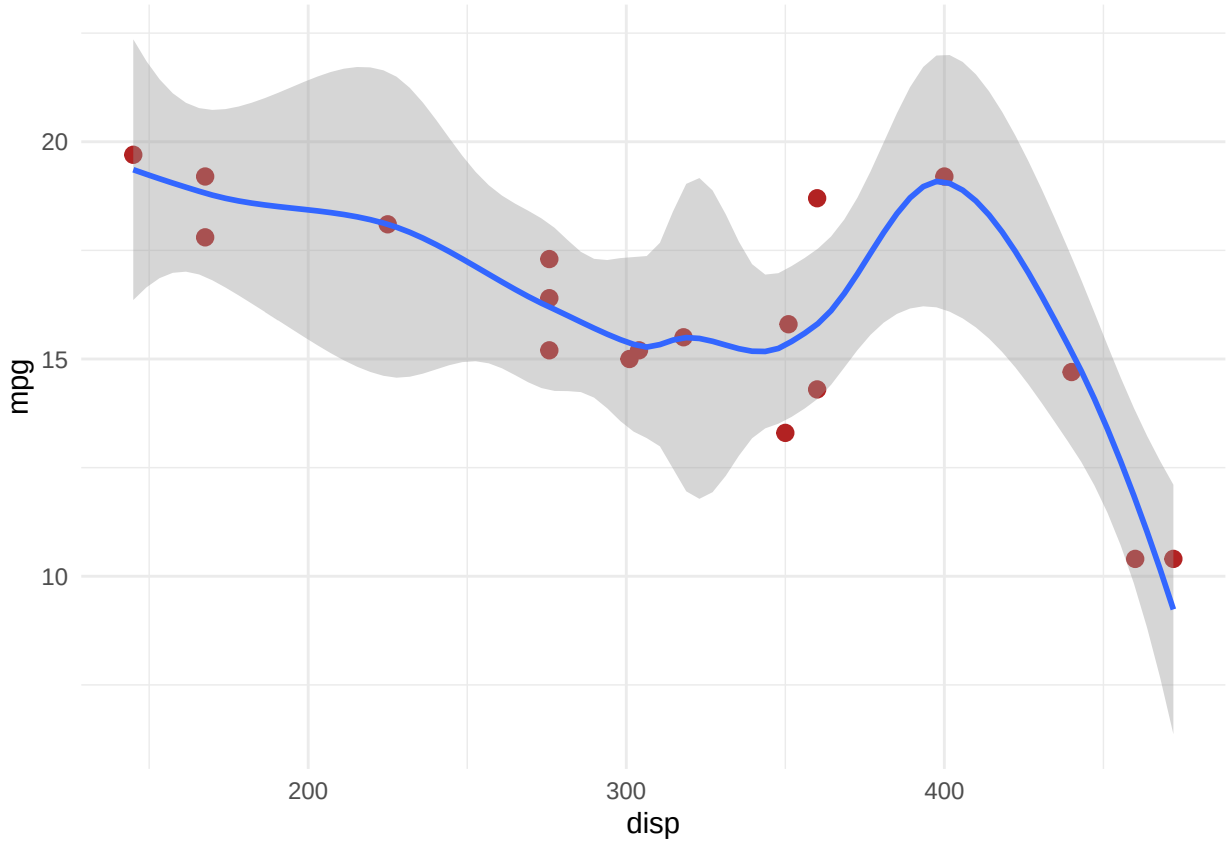
Run this code to use the function.

```
#library(esquisse)

#answer: adjustment of the data
cyl_adjust<-as.factor(mtcars_mpg2$cyl)
mtcars_mpg2<-cbind(mtcars_mpg2,cyl_adjust)

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

When you need to use Esquisse, input “library(esquisse)” and “esquisser(data = mtcars_mpg2, viewer =”browser“)”. If you don’t use it, just add “#” before it, otherwise it will run and get stuck.

```
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'
```



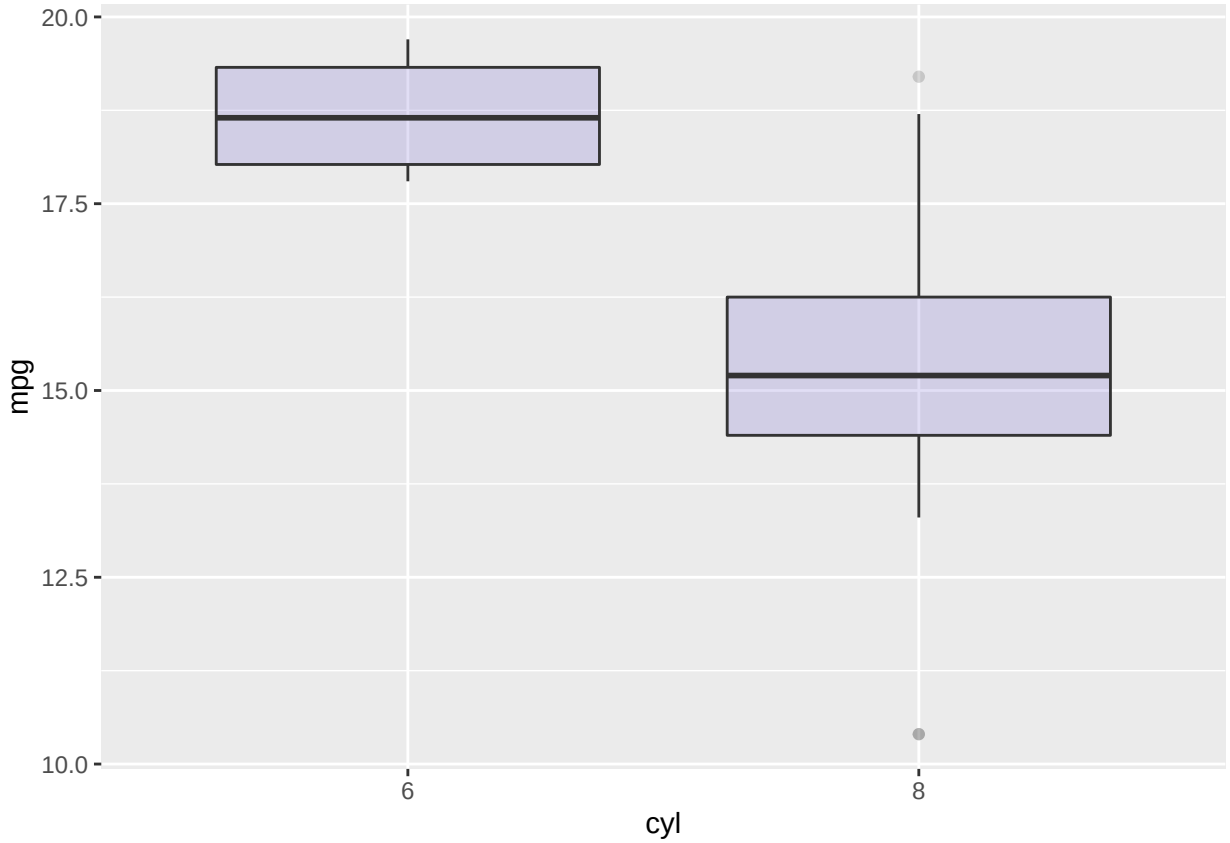
These operations are manually adjusted in Esquisse.

"geom_point" represents point plot.

"geom_smooth" represents a smoothing line in order to see what the trends look like.

"theme_minimal()" is a minimalistic theme with no background annotations.

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

#answer: create cyl_adjust<-as.factor(cyl), then add cyl_adjust to mtcars_mpg2
ggplot(mtcars_mpg2, aes(x=as.factor(cyl), y=mpg)) +
  geom_boxplot(fill="slateblue", alpha=0.2) +
  xlab("cyl")
```




```
#use adjustment in RStudio or Esquisse:
#ggplot(mtcars_mpg2, aes(x=cyl_adjust, y=mpg))
#+ geom_boxplot(fill="slateblue", alpha=0.2) + xlab("cyl")
```

"geom_boxplot" represents box plot.

In my opinion, Esquisse can't draw this boxplot because there is no as.factor(cyl) in Esquisse. In order to solve this issue, I create a cyl_adjust in mtcars_mpg2 so that Esquisse appears cyl_adjust which means as.factor(cyl) now. Put cyl_adjust in x and mpg in y, then choose boxplot mode, the issue works out.