

newassignment1

Xiang Li

9/23/2021

```
library(tidyverse)
```

```
## — Attaching packages ————— tidyverse 1.3.1 —
```

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

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

#Import the database and then choose the data needed to be analyzed. #mtcars #Format #A data frame with 32 observations on 11 (numeric) variables. #[, 1] mpg Miles/(US) gallon #[, 2] cyl Number of cylinders #[, 3] disp Displacement (cu.in.) #[, 4] hp Gross horsepower #[, 5] drat Rear axle ratio #[, 6] wt Weight (1000 lbs) #[, 7] qsec 1/4 mile time #[, 8] vs Engine (0 = V-shaped, 1 = straight) #[, 9] am Transmission (0 = automatic, 1 = manual) #[,10] gear Number of forward gears #[,11] carb Number of carburetors

```
# 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)]
```

#Use the function from hand_function.R

```
# 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)
```

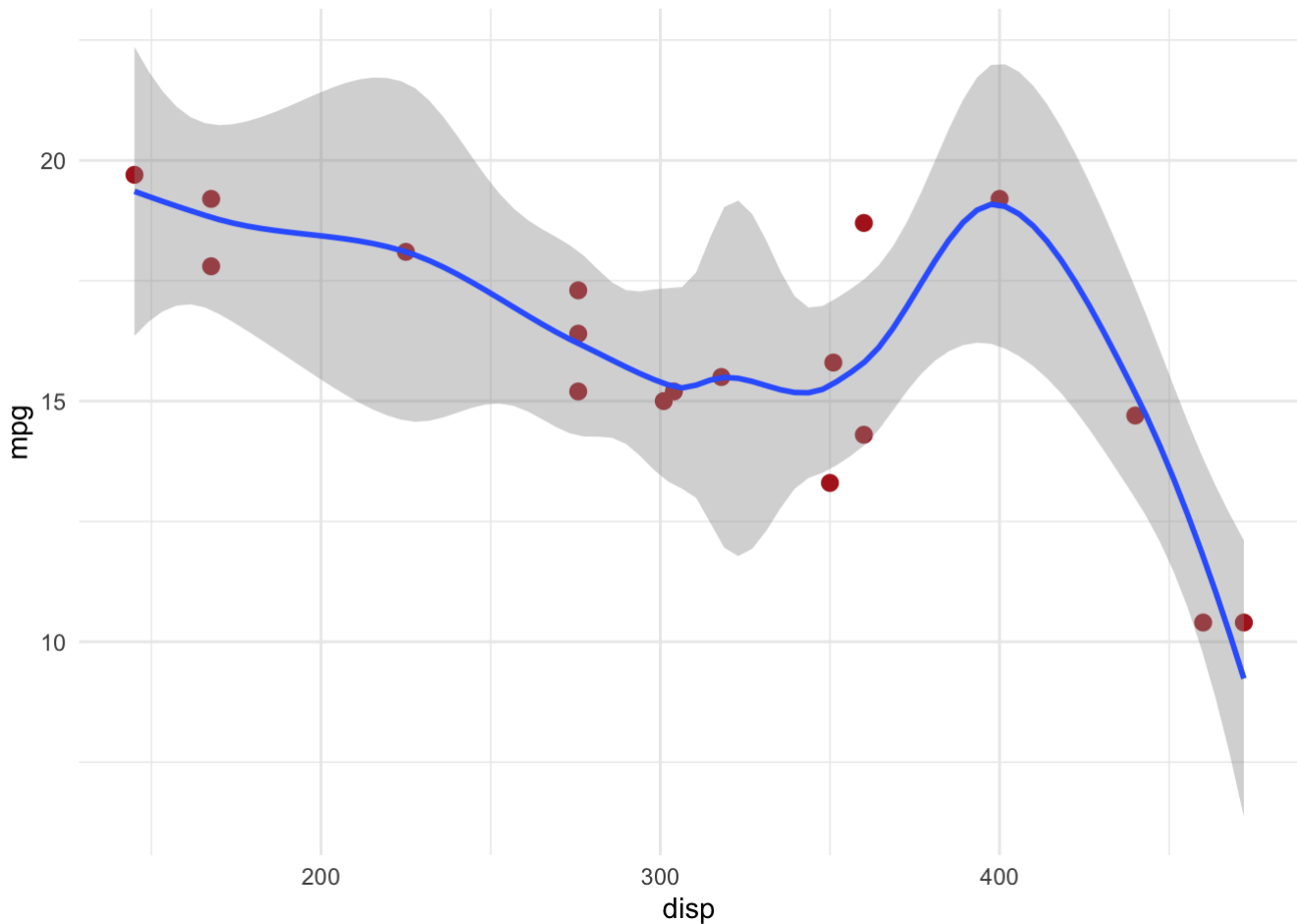
```
##
## > sum_special <- function(df_x) {
## +   try(if (!is.data.frame(df_x))
## +     stop("Input data must be a data frame."))
## +   sp_means <- apply(df_ .... [TRUNCATED]
```

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

#Plot mtcars with selected data to illustrate that how does the mpg changing trend while displacement is increasing. #theme_minimal() :A minimalistic theme with no background annotations. #theme_classic(): A classic-looking theme, with x and y axis lines and no gridlines.

```
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'
```



```
#library(esquisse)
#mtcars_mpg2<-mtcars[as.factor(mtcars$cyl)]
#esquisser(mtcars_mpg2)
```

note that this boxplot cannot be made with `esquisse()` unless

the data is adjusted. What adjustment is needed?

#ANSWER: Use `as.factor(cyl)` instead of `cyl` because a boxplot is a method for graphically depicting groups of numerical data through their quartiles. Therefore, I need to use `as.factor` to change `cyl` to a classification variable.

```
ggplot(mtcars_mpg2, aes(x=as.factor(cyl), y=mpg)) +  
  geom_boxplot(fill="slateblue", alpha=0.2) +  
  xlab("cyl")
```

