R Notebook

```
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()
```

mtcars is a dataset in which each observation is a car model and has 11 variables. The data is from 1974.

```
# 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]
```

```
# Now use the function from hand_functions.R to produce summary statistics

sp_out <- sum_special(mtcars_mpg2)
sp_out
```

```
## $sp_means
##        mpg        cyl       disp         hp       gear
##  15.900000   7.555556 313.811111 191.944444   3.444444
##
## $sp_var
##          mpg          cyl         disp           hp         gear
##    7.5258824    0.7320261 9438.7645752 3253.5849673    0.6143791
##
## $sp_cov
##               mpg        cyl       disp         hp        gear
```

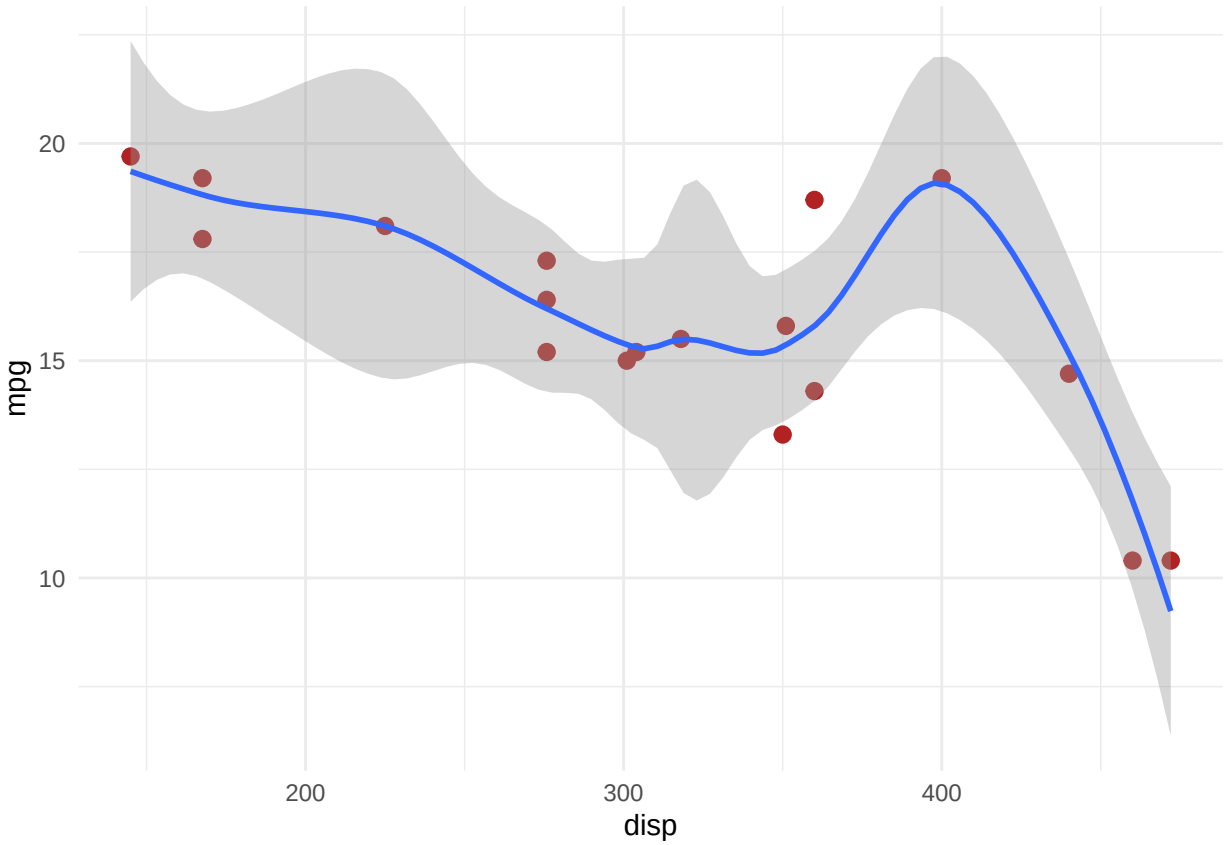
```
## mpg     7.5258824 -1.3176471 -188.79529  -75.81176   0.6352941
## cyl    -1.3176471  0.7320261   64.71111   28.44444  -0.2614379
## disp -188.7952941 64.7111111 9438.76458 2679.60065 -34.1934641
## hp    -75.8117647 28.4444444 2679.60065 3253.58497  15.2026144
## gear    0.6352941 -0.2614379  -34.19346   15.20261   0.6143791
##
## $sp_cor
##             mpg        cyl       disp         hp       gear
## mpg   1.0000000 -0.5613802 -0.7083614 -0.4844811  0.2954459
## cyl  -0.5613802  1.0000000  0.7784989  0.5828450 -0.3898406
## disp -0.7083614  0.7784989  1.0000000  0.4835389 -0.4490217
## hp   -0.4844811  0.5828450  0.4835389  1.0000000  0.3400314
## gear  0.2954459 -0.3898406 -0.4490217  0.3400314  1.0000000
```

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

The first plot shows mpg vs. engine displacement. The plot shows that mpg decreases with displacement. Displacement is the combined swept volume of the pistons inside the cylinders of an engine and directly affects power output and fuel efficiency.

```
ggplot(mtcars_mpg2) +
  aes(x = disp, y = mpg) +
  geom_point(shape = "bullet", size = 4L, colour = "#B22222") + #scatterplot
  geom_smooth(span = 0.5) + #line plot, span controls the amount of smoothing for the default loess smo
  theme_minimal() #theme, how the plot looks
```

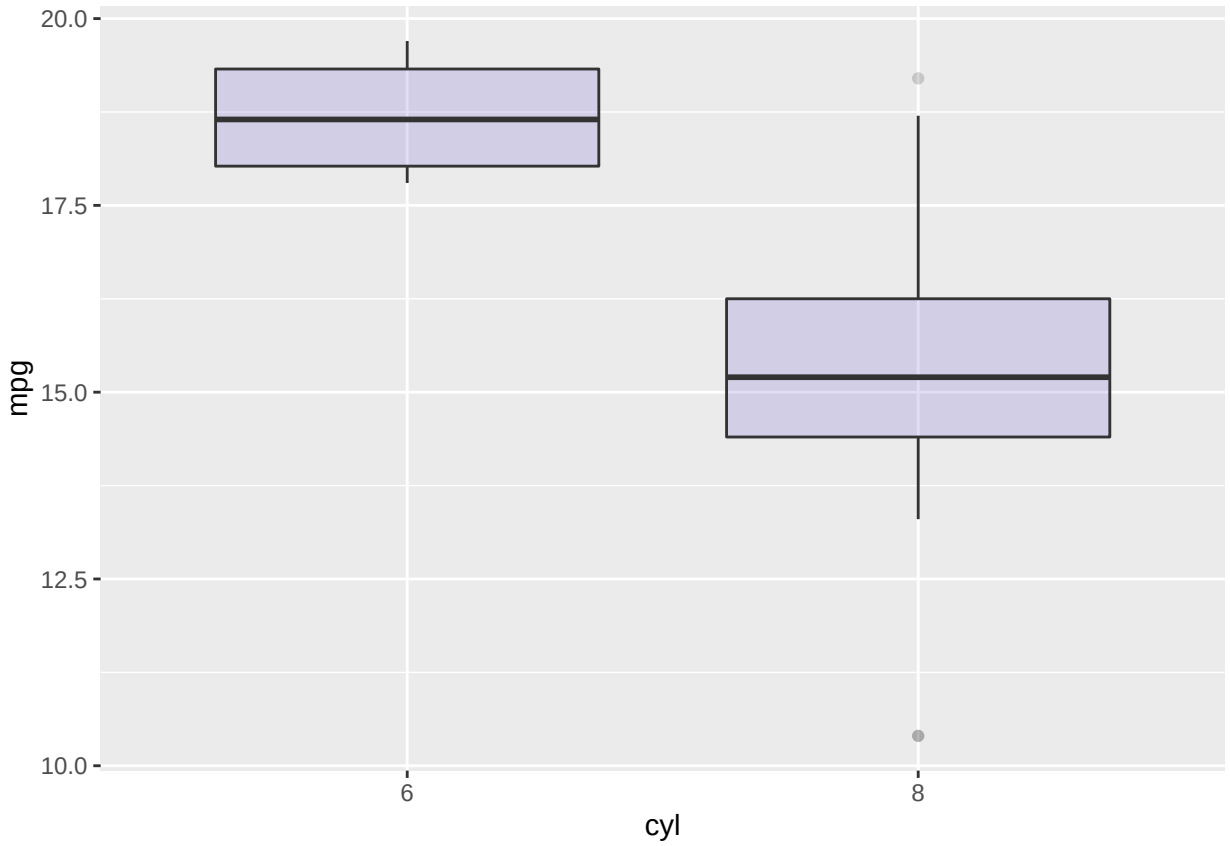
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The second plot shows that cars with 6 cylinders have higher mpg than cars with 8 cylinders.

```
# note that this boxplot cannot be made with esquisse() unless
# `cyl` is changed to a factor using `as.factor(mtcars_mpg2$cyl)`.
# I verified this with this code:
# mtcars_mpg3 <- mtcars_mpg2
# mtcars_mpg3$cyl <- as.factor(mtcars_mpg3$cyl)
# esquisser(data = mtcars_mpg3)

ggplot(mtcars_mpg2, aes(x=as.factor(cyl), y=mpg)) + #`as.factor` changes `cyl` to a factor or category
  geom_boxplot(fill="slateblue", alpha=0.2) + #`fill` fills the boxes with color, `alpha` changes trans
  xlab("cyl")
```


20.0
17.5
15.0
12.5
10.0
mpg
6
8
cyl