Assignment 1

MA615 2021

9/20/2021

Due: Friday 24 Sept

This assignment . . .

Is a continuation of the assignment we began after the first class.

As you begin using R more, teamwork, production values, internal documentation, and reproducibility will emerge as essential factors affecting the success of your projects. Learning to use the tools that support these success factors is an important part of learning R. In this exercise, you begin with the work that you did in the project where you used the esquisse package to plot some relationships in the mtcars data set.

Begin by setting up a github repo and cloning it to your local folder for MA615 assignments.

- Open your github account.
- Set up an organization for your MA615 work.
- In the organization, create a new repo for this assignment.
  - Be sure to check the boxes for (a) creating a readm, (b) creating a .gitignore file using the R template, and (c) including a licence.
  - Now, follow the instructions from Friday's class and use R to clone your github repo to your local folder – while initializing git and an rproject at the same time.

Now, copy the two R scripts in this assignment to your local disk.

- car_viz.R
- hand_functions.R

Deliverable

1. Use car_viz.R as the starting point for a two documents.

   - html
   - pdf

2. The text of the document should describe the data and discuss what you have learned in this very simple exploration.

3. The commentary should address the code itself, noting features that may be important to anyone maintaining or using the it.

4. Turn in your assignment by submitting a link to your github repo on blackboard. Make sure that the code in your repo will run to produce the documents your submit. While you might not always choose to push both the source code for a document and the document itself to your github, please do for this assignment.

```r
# read .r document and output it
source("car_viz.R", echo = T, skip.echo = 0)
```

```
## 
## > library(tidyverse)

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

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

## 
## > data(mtcars)
## 
## > mtcars_mpg2 <- mtcars[mtcars$mpg < 20, ]
## 
## > mtcars_mpg2 <- mtcars_mpg2[, c(1, 2, 3, 4, 10)]
## 
## > 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]
## 
## > sp_out <- sum_special(mtcars_mpg2)
## 
## > ggplot(mtcars_mpg2) + aes(x = disp, y = mpg) + geom_point(shape = "bullet",
## +     size = 4, colour = "#B22222") + geom_smooth(span = 0.5) +
## +      .... [TRUNCATED]

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```
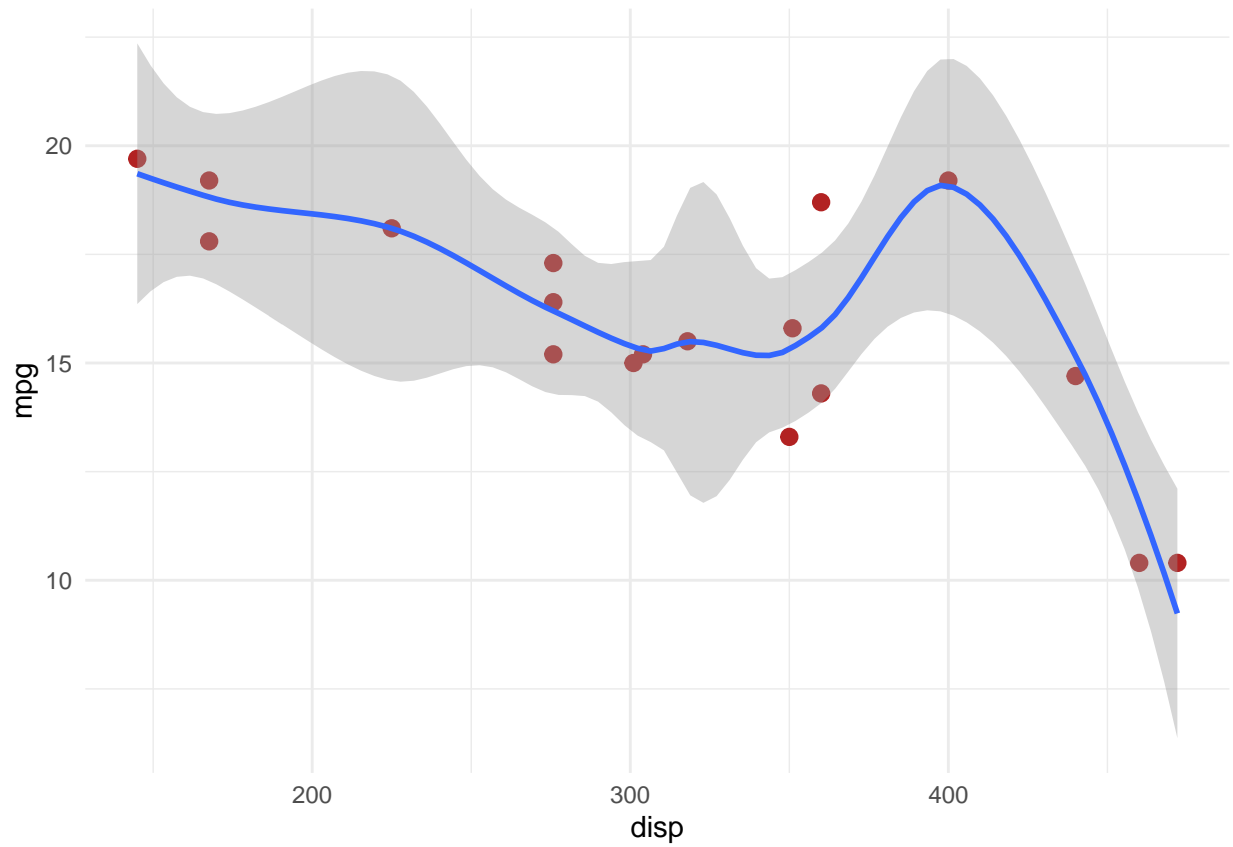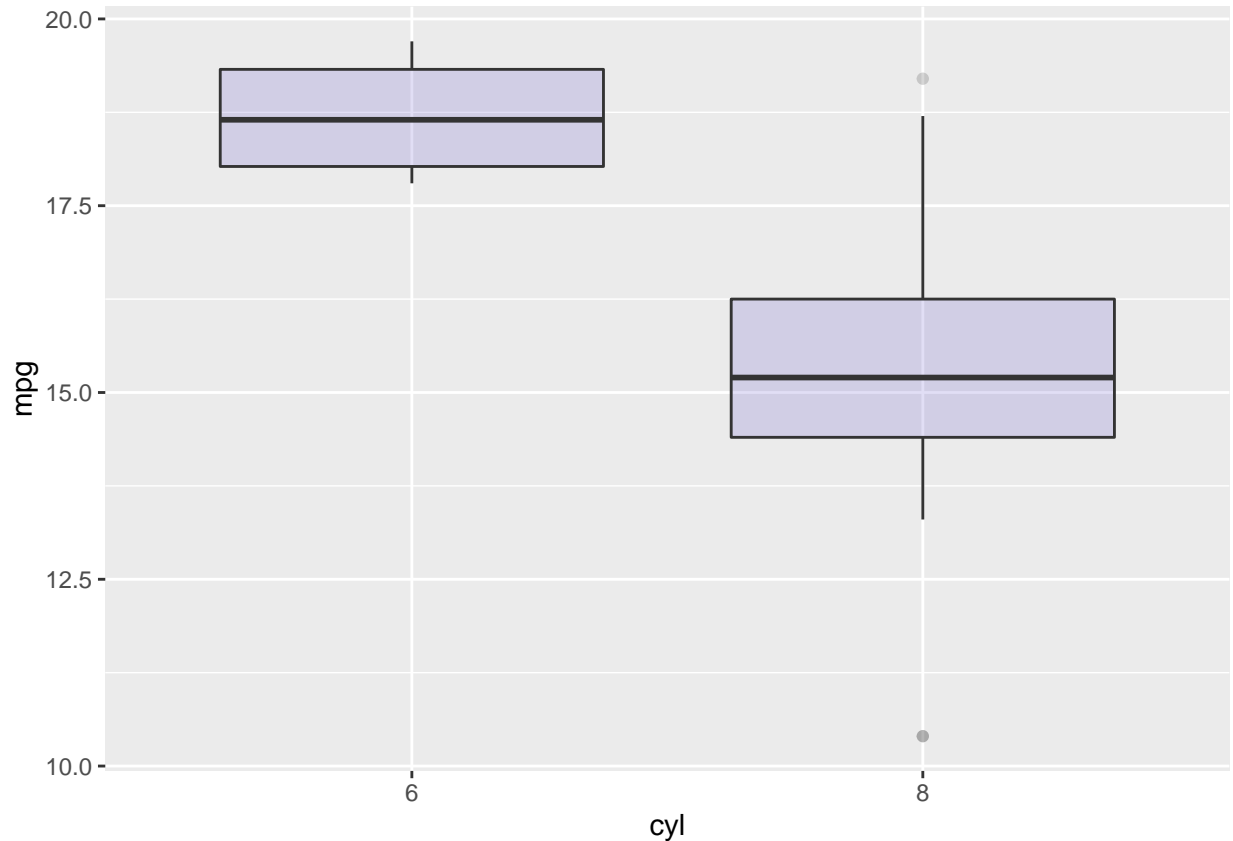
```
##
## > ggplot(mtcars_mpg2, aes(x = as.factor(cyl), y = mpg)) +
## +     geom_boxplot(fill = "slateblue", alpha = 0.2) + xlab("cyl")
```

The code for mtcars selected car models where mpg<20. Following analysis for variables include mpg, cyl, disp, hp and gears. Then use the file "hand_function.R" to produce a list include mean, variation of every column, and covariance, correlation of every two columns. Finally, the scatter point plot adds a smooth line to state the trend of the data. The smooth line uses a locally weighted regression to give a segmented fit model. The second plot is a box plot with a point, which helps the reader to compare the total mpg distribution between two groups of cyl. We can draw a conclusion that the "6" group is larger than the "8" group generally and the data in the "6" group is more centralized than the data in the "8" group.

Through learning the code, I know how to pack different R files. Also, I learned how to write a function file. Another aspect is about the basic manipulate of the data frame, how to select columns and rows follow conditions, and the use of several types of ggplot.