Assignment 1 - MA615 2021

Yongrong Chai

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.

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──
```

```
## ✓ ggplot2 3.3.2     ✓ purrr   0.3.3
## ✓ tibble  3.0.3     ✓ dplyr   1.0.2
## ✓ tidyr   1.1.1     ✓ stringr 1.4.0
## ✓ readr   1.3.1     ✓ forcats 0.5.0
```

```
## ── 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)]

# 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]
```

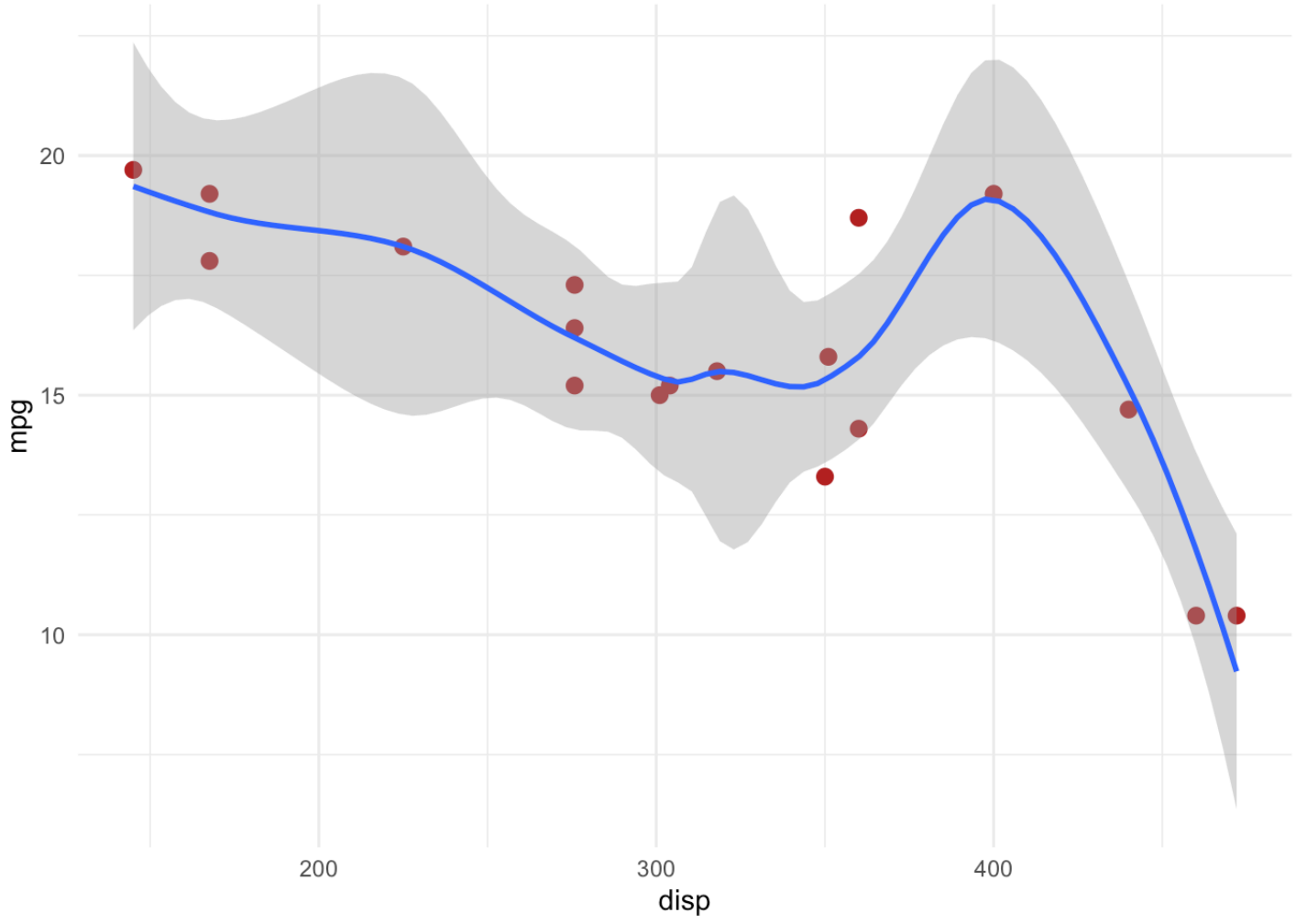
```
# Now use the function from hand_functions.R

sp_out <- sum_special(mtcars_mpg2)


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

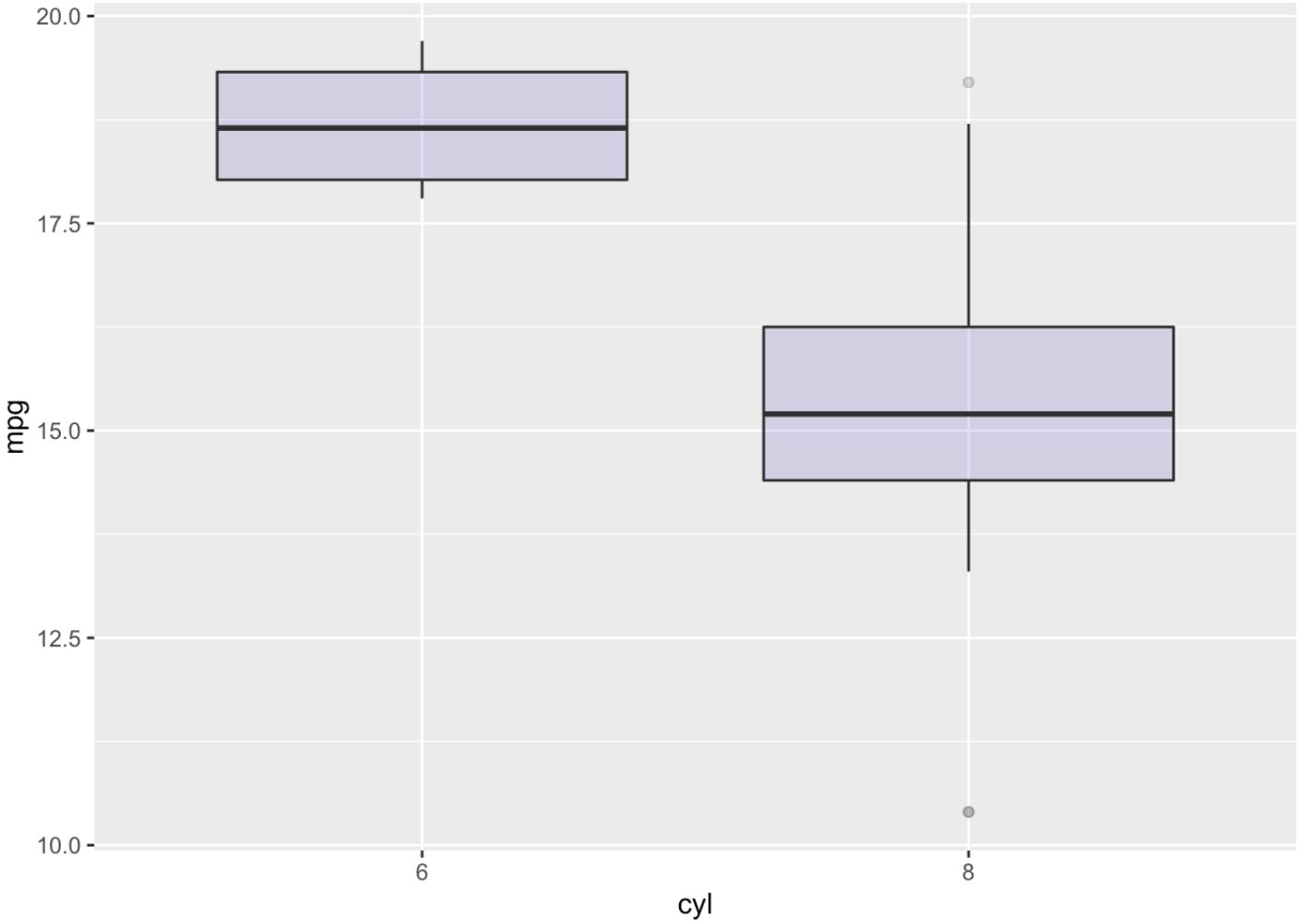
```
ggplot(mtcars_mpg2) +
  aes(x = disp, y = mpg) +
  geom_point(shape = "bullet", size = 4L, colour = "#B22222") +
  geom_smooth(span = 0.5) +
  theme_minimal()
```

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



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


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




```
mtcars_mpg2$cyl_factor = as.factor(mtcars_mpg2$cyl)
#adjusting cyl to be a categorical variable
```

Esquisse allows me to quickly visualize the data by using drag-and-drop. It is really helpful because I can retrieve the code used to produce the plot, as well as the code used to filter the data.

Besides, it is much easier to add plot's title, subtitle, caption and axis labels and make color adjustment.

And I prefer to use esquisser(viewer = "browser").

The reason I cannot made boxplot with esquisse() is that cyl is a numerical variable.

Adjustment: transform it into a categorical variable. I add a new column variable cyl_factor = as.factor(mtcars_mpg2$cyl), thus I can drag this into x and use boxplot option to generate the plot sucessfully with esquisse.