Data Science Homework 2 - 6 October 2021

1. Fix each of the following common data frame subsetting errors:

mtcars[mtcars$cyl < 6] SHOULD BE **mtcars [mtcars$cyl < 6,]** mtcars[-1:3,] SHOULD BE **mtcars[1:3,]** mtcars[mtcars$cyl = 8,] SHOULD BE **mtcars[mtcars$cyl ==8,]** mtcars[mtcars$cyl == 4| 6,] SHOULD BE **mtcars[mtcars**$cyl == 4, |$mtcars$**cyl ==6,]**

2. (1 point) Why does the following code generated five missing values?

x = 1:5 x[NA]

Answer: NA is infectious; if you input NA, your output will also be NA

3. Why does mtcars[1:15] return an error? How does it differ from mtcars[1:15,]

Answer: mtcars[1:15] does not specify which columns to subset, it only specifies the rows. Contrastingly, mtcars[1:15,] includes a comma, which separates the rows to be subsetted from the columns to be subsetted (in this case, we want to subset the matrix to include rows 1 through 15 and all of the columns).

4. Explain how the following code works:

x = matrix(c(1:3, NA, 5:7, NA, NA), nrow = 3) x[is.na(x)] = 0

Answer: This code generates a matrix that initially has missing values (NAs), but then replaces those NAs with 0's.

5.

Answer: The following code was used to generate the column mpg_2 with the appropriate values:

> within(mtcars, mtcars$mpg_2 <-ifelse(mpg < 16, "Low", ifelse(mpg < 21, "Low_intermmediate", ifelse(mpg < 26, "Intermediate_high", "High"))))