**Quiz 1**

*Content: matrix, data frame.*

In the R code file for Topic 1, <Topic1\_Rcode.R>, there is a part of code as below:

data1<-read.csv("C:/Data/crab.txt",sep = "", header = TRUE)

data1[1:8,] #first 8 rows. The output is below.

color spine width satell weight

1 3 3 28.3 8 3.05

2 4 3 22.5 0 1.55

3 2 1 26.0 9 2.30

4 4 3 24.8 0 2.10

5 4 3 26.0 4 2.60

6 3 3 23.8 0 2.10

7 2 1 26.5 0 2.35

8 4 2 24.7 0 1.90

**Q1.** Column “color” in the data frame above is recognized by R as a categorical variable. True or False?

Ans: FALSE. The values of column “color” are numeric, hence, by default, R considers this column as a numeric column/variable.

**Q2.** To let R recognizes the first column of data frame above as a categorical, we need to sue a function to convert this column into a factor (\*). True or False?

(\*) such as function as.factor(), or factor().

Ans: TRUE. Both as.factor() and factor() could be used.

**Q3.** Brandon uses function as.factor() to transform the first column of data frame above into a categorical variable by the code below.

color = as.factor(data1$color)

Then, now, the first column of “data1” is recognized by R as categorical. True or False?

Ans: FALSE.

This line of code by Brandon creates a new variable, named as “color”, which factorizes the column “color” of “data1”. Hence, this new variable with the name “color” is a categorical variable and it is outside of “data1”.  
To verify, you can run the code

is.numeric(data1$color)

Then R will show “TRUE” to confirm that the first column of “data1” is still NUMERIC, not categorical. Instead, you will have FALSE when running

is.numeric(color)

**Q4.** After running the line of code on Q3, Brandon continues with two lines of code below:

color\*2 # LINE 1: to multiply every element of “color” by 2

data1$color\*2 # LINE 2: to multiply every element of column “color” in “data1” by 2

Both lines of code above will not produce answers. True or False?

Ans: FALSE. LINE 1 will produce “NA”. This is because object “color” is memorized by R as categorical. Hence, multiplication is not possible for its categories.

However, for column “color” inside “data1”, data1$color, it is still recognized by R as a numeric column. Hence, multiplication is possible for it.

**Q5.** If now you run the code

spine

then you will get an error from R. True or False?

Ans: True. This is because R now knows “data1” but it doesn’t know the name of each column inside “data1” yet.

**Q6.** Fact: function attach(df) helps R to know every names inside data frame “df”, to know the values of each column, as well as the properties (categorical/numeric) of each column.

After all the code above was run, Brandon continues with function attach() as below.

attach(data1)

Now, there is no error appearing when Brandon runs the code

spine

True or False?

Ans: TRUE. Because after function attach(data1) was run, R now knows that “spine” is a column inside “data1” and R knows all the values of that column.

**Q7.** Brandon continues with the code below.

spine = as.factor(data1$spine)

After this line of code is run, column “spine” of “data1” now become a categorical variable. True or False?

Ans: False. Back to Q3 and the answer of Q3. This code creates a new object/vector, named “spine” OUTSIDE of “data1”, it uses column “spine” of “data1” and factorizes it. So, this “spine” is a categorical variable.

However, the column “spine” INSIDE “data1” is still memorized by R as a numeric column. This memory was created when function attach(data1) was run.

**Q8.** To let R recognizes column “spine” inside “data1” as a categorical column, we should use

data1$spine = as.factor(data1$spine)

True or False?

Ans: True.