**Business Analytics**
**CD2008**
**Ing. Juan José Franklin Uraga, PhD.**

Homework #04

To complete this homework you will need to install the "datos" package that contains the data frame named *millas* and the package "ggplot2" that contains the data frame *diamonds*. No additional databases or packages are required:

I. **Answer the following questions using the data frame *millas*:**

1. Which geom should you use to generate **(6 points)**:

- a line graphic (a connecting line and also a trend line)
- a box plot
- a histogram

2. Execute the following code and describe in your own words what are the results showed in the graph. **(12 points)**.

```
ggplot(millas, aes(x = cilindrada, y = autopista, color = traccion)) +
geom_point() +
geom_smooth(se = FALSE)
```

3. In the previous exercise, What is shown when using show.legend = FALSE inside geom_smooth()? What happens if you remove show.legend = FALSE from the code? **(3 points)**

4. What does the argument se do in geom_smooth()? **(9 points)**

5. Do these two graphics are displayed differently? Why? **(15 points)**

```
#Graph 1
ggplot(millas, aes(x = cilindrada, y = autopista)) +
geom_point() +
geom_smooth()

#Graph 2
ggplot() +
geom_point(millas, aes(x = cilindrada, y = autopista)) +
geom_smooth(millas, aes(x = cilindrada, y = autopista))
```

## II.    Answer the following exercise using the data frame *diamonds:*

6. Recreate this graphic. The parameter that controls the transparency is 0.3. It is not necessary that you present it using the same colors, but it is mandatory that all colors must be able to differentiate. **(15 points)**
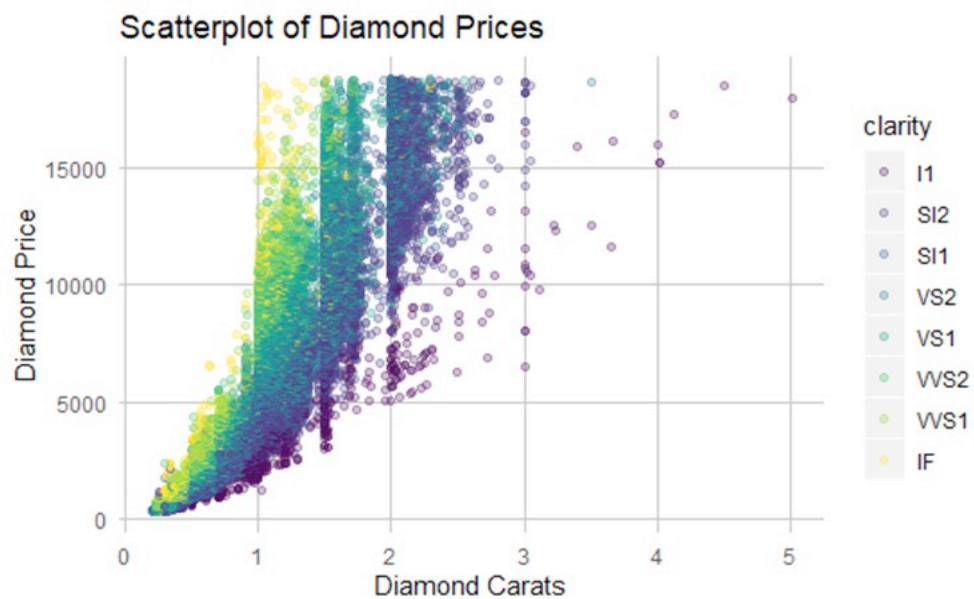


7a) and 7b) Using the information contained in the *diamonds* data set, recreate the following graphics **(25 points each):**