**SDM Assignment 4**

The following questions refer to the dataset “Autoparts.csv” posted on Canvas. The data contains information about sales of auto parts from 18 sales regions around the USA. For each region, we have data on sales amount (in thousands of dollars, denoted by “Sales”), the amount of money spent on marketing (in thousands of dollars, denoted by “Mktg”) as well as the number of sales outlets, the size of the population (measured in millions), the number of registered vehicles (also measured in millions), and the number of sales reps per region. The following questions refer to this data.

1. We would like to investigate the impact of marketing efforts on sales. To that end, run a regression model with “Sales” as the response variable and “Mktg” as the (only) explanatory variable. Copy and paste the R code and the output, and writethree key findings that you can infer from this analysis.
2. Now, we would like to investigate the impact of all variables. To that end, run a regression model with “Sales” as the response variable and all other variables in the data set as predictor variables. Copy and paste the R code and the output, and writethree key findings that you can infer from this analysis. Does anything in this output strike you as “strange”?
3. Create a table listing the multiple R-squared, adjusted R-squared, AIC, and BIC of the two models above. Which is the better model and why?