**SDM Assignment 2**

Download the dataset “KoolKarmaSales.csv” posted on Canvas. The data contains information about sales of lemonade (called “Kool Karma”); sales are distributed across 25 sales districts; in each district, we record the sales amount (in thousands of dollars, denoted by “SALES”), the amount of money spent on advertising (in thousands of dollars, denoted by “ADVT”) and the median household income (also in thousands of dollars, denoted by “INCOME”). Analyze the dataset using the R software to answer the following questions.

1. Investigate the distribution of sales. First, compute a histogram of SALES. Attach the resulting graph, copy and paste the accompanying R code for computing this histogram, and listthree key observations that you see about Kool Karma sales.
2. Now, investigate the relationship between sales and advertising. Draw a scatterplot between SALES and ADVT. Attach the resulting graph, copy and paste the accompanying R code that you used to draw this scatterplot, and list three key observations about the relationship between advertising and sales that you see from this plot.
3. Finally, refine your analysis in Question 2 by conditioning the scatterplot between sales and advertising on different levels of income. To do this, install the R library “lattice” by typing *install.packages(“lattice”)*, then load it for your current session by typing *library(lattice)*. Since we want to condition the scatterplot on high vs low levels of INCOME, create a new variable, say, *INCOME.HIGH* for which

INCOME.HIGH = 1 if INCOME > 55 and

INCOME.HIGH = 0 if INCOME <= 55

and compute a scatterplot between SALES and ADVT, conditional on INCOME.HIGH. Attach the resulting graph, copy and paste the accompanying R code for computing this scatterplot, and list three key observations about the relationships between advertising, income and sales that you see from this plot.