**MOOC: R Exercise 1**

1. ‘Varsity’ is currently stored as type ‘double’. Change this continuous variable to a categorical variable. R denotes categorical variables as factors.
2. Sort data, first in ascending order by ‘Varsity’, then in decreasing order by ‘WritingLove’.
3. Filter the data such that you create a dataset with only records of students who have ‘MathLove’ scores greater than or equal to 70.
4. Click on Extras -> Data Summary for a quick summary of all the variables in your dataset. How many missing values does the variable ‘carter’ have?
5. Create a histogram for the variable ‘coffee’ in dark blue.
6. Create a scatterplot of hours of exercise (‘exercise’) vs. hours of homework (‘homework’). Include a smoothed loess curve. Comment on what you see (trends, outliers, etc.).
7. Create a histogram for ‘milk’. You notice that there is one person who indicated ‘50’ that seems like an outlier. After going back to the data collection forms, you verify that 50 was indeed miscoded, and that the true value is 5. You decide to fix this problem by creating a new variable called ‘milkFixed’. (Hint: Save before attempting. Go to Data 🡪 Recode Variables. Be sure to hit the “🡪 Target” button while you have the item in “Variables to Recode” highlighted. This will allow you to rename the new variable so you do not overwrite your existing values. Specify conditions in “Define Recode”)
8. What is the mean, standard deviation, and median number of hours of sleep in the dataset?
9. Perform a stratified analysis of mean outlook on life by alcohol consumption (i.e., What is the mean score for outlook on life for each amount of alcohol consumed?). What is the mean score for those who indicated that they have 5 drinks?
10. How many people consider themselves street smart? (Hint: Another way of phrasing this question is ‘what is the frequency of people who consider themselves street smart?’)
11. Create a box plot for ‘fruitveg’. What does the black dot in the figure represent?
12. Re-examine the relationship of mean outlook on life by alcohol consumption using a group box plot.