Exercise Problems

We will use the exercise_data.csv file to answer the following:

1. Read the exercise_data.csv in rattle.

2. The values in few columns may not read in correct data type due to presence of special characters like '$', '(', ')', '%', ','.
    a. Use rattle to convert the columns into correct data type.

3. Create a binned variable for Inventory weeks by using equal binning with 4 bins.

4. Using Rattle: Perform cross-tabulation of data:
    a. Brand Name with Dealer
    b. Items with Brand Name

5. Visualize the data using Rattle
    a. Report the correlation matrix between net sales and net cost
    b. Perform correlation plot between net sales and net cost.

Question 6 and 7 are optional with rattle. Instructor may demonstrate the use to auto-generated code to fix the issue discussed in question 6 and question 7.

6. The column names do not follow proper naming convention:
    a. There are signs like '$', '%' in the column names.
    b. Few column names are in upper case and a few of them in lower case
    c. In few columns, there are underscore '_' sign in between the words while in few column names, there are blank spaces in between the words.

    Write your code to clean up and resolve issues in the column name. You may think of using 'gsub' function to deal with the above issues. More about the function gsub() at: http://www.endmemo.com/program/R/gsub.php

7. Using the clean data, write your code to compute the following metrics:
    - Gross Margin = Net Sale – Net Cost
    - Percentage Gross Margin = (Net Sale – Net Cost)/Net Sale
    - Net Sales Per Fashion Style = Net Sale / Fashion Style Count
    - Gross Margin Per Fashion Style = Gross Margin / Fashion Style Count
    - Net Sales per Sales Units = Net Sales / Sales Units

8. After having solved question 1 to 5, use the code generated in rattle – log files and clean-up the code for further use. Explore addressing question 7 and 8 with this code.