**Module 5 – Data Manipulation/Clean In-class Exercises**

Purpose: This in-class exercise gives you practice performing data manipulations, including creating new columns/variables using simple and complex logic and recording variables.

Part 1: Loading your data set

1. Import and merge the bmx, demo, paq\_j, and sleep datasets into R
2. Label the variables and values

Part 2: Creating new columns/variables

1. Create a new variable called sleep\_duration\_med with a fixed value = 7.5
2. Create a new variable called sleep\_duration\_mean with a fixed value equal to the sample mean
3. Create a new variable, sleep\_duration\_avg, to equal the weighted average of weekday and weekend sleep duration, i.e., weekday (5/7) and weekend (2/7)
   1. Weekday = sld012; Weekend = sld013 in the original sleep.csv dataset
4. Create a variable, sleep\_duration\_cat
   1. First level = very short, average sleep duration <5
   2. Second level = short, average sleep duration >5 to <7
   3. Third level = medium, average sleep duration >7 to <9
   4. Fourth level = long, average sleep duration >9
5. Create a binary variable, sufficient\_sleep, set value equal to yes if sleep\_duration\_avg >7 and <9 and no otherwise
6. Create a binary variable, single, using the married variable. Set single = No if “Married” or “Living w/partner” and Yes otherwise
7. Convert sleep\_duration\_cat, sufficient\_sleep, and single into factor variable

Part 3. Recoding values

1. Explore the pad660 "Leisure Vigorous PA Minutes" and paq650 “Leisure Vigorous PA, yes/no” variables
2. Recode pad660 to 0 if paq650 is set to 2
3. Explore the dpq010 variable
4. Recode dpq010 to “Missing” if dpq010 equals “Refused”