Lab 3 tasks–Wrangling and visualizing data

(Assignment 1, 15 marks)

This lab task requires the mammals sleep data given in "msleep" dataset under package "ggplot2". You can directly use msleep dataset after loading package "ggplot2". It contains the sleep times and weights for a set of 83 mammals. Details of 11 variables are given below:

column name	Description
name	common name
genus	taxonomic (classifying) rank
vore	carnivore, omnivore or herbivore
order	taxonomic rank
conservation	the conservation status of the mammal
sleep_total	total amount of sleep, in hours
sleep_rem	rem sleep, in hours
sleep_cycle	length of sleep cycle, in hours
awake	amount of time spent awake, in hours
brainwt	brain weight in kilograms
bodywt	body weight in kilograms

Before doing the following tasks, you are suggested to use summary() to view a descriptive summary for each variable in the dataset.

1. Create a new dataframe named as "sleep3", which includes all variables in the original msleep dataset except "name", "genus", "conservation" and "awake", for mammals with body weight falling into interval [0.1, 200].
2. In the dataframe "sleep3", add a new variable: "rem_ratio", which is the ratio of rem sleep to total sleep.
3. For the dataframe "sleep3", find summary statistics, including the number of mammals, average of rem_ratio and average of body weight over different vore groups. Based on results you obtained, briefly comment on the comparison between vore groups (write your comment in the code using "#").
4. Use an appropriate plot to compare the distributions of rem_ratio for different vore groups of mammals. Briefly comment on how different rem_ratio distributes over different vore groups (write your comment in the code using "#").
5. Use an appropriate plot to visualize the relationship between rem_ratio and body weight. Briefly comment on (write in the code using "#") the relationship that you find from the plot.