## Week 1 - Discussion

**Problem 1.** This problem will be based on the HW2 dataset COVID19Data.csv. Rather than considering a subset, we want to analyze the whole dataset.

- a. Use head(), tail() and View() functions to get a quick look of this dataset.
- b. Use the summary() function to get some basic descriptions of this dataset. Explain the outcome.
- c. How to get more statistics such as std, IQR for all numerical variables in a wise way? One way to do this is combining apply(), sd() and IQR() functions.
- d. What other worthy data manipulations or data analyses you can propose.

- **Problem 2.** Consider a set  $B = \{\sqrt{1}, \sqrt{3}, \sqrt{5}, \cdots, \sqrt{2k-1}\}$ , where k is a positive integer.
  - a. Write R code to generate the set B for k = 10.
  - b. Write a R function to generate the set B for any k. Test your function by comparing this result with part (a) when k = 10.
  - c. Compute the mean, median, IQR and standard deviation of  $B_{k=10}$ .
  - d. Compute the mean, median, IQR and standard deviation of  $-2 * B_{k=10}$  which is  $\{-2\sqrt{1}, \cdots, -2\sqrt{19}\}$  and  $2 * B_{k=10}$  which is  $\{2\sqrt{1}, \cdots, 2\sqrt{19}\}$ .
  - e. Compute the mean, median, IQR and standard deviation of  $-2 + B_{k=10}$  which is  $\{-2 + \sqrt{1}, \dots, -2 + \sqrt{19}\}$  and  $2 + B_{k=10}$  which is  $\{2 + \sqrt{1}, \dots, 2 + \sqrt{19}\}$ .
  - f. Compare results from parts (c), (d) and (e), what can you find?
  - g. Based on the rule you find from part (f) and the result of part (c), answer the mean, median, IQR and standard deviation of  $-10 * B_{k=10} + 10$ . Check your answer by direct calculation.