Homework 2

Please submit the solution on Canvas into the corresponding assignment (e.g. "Homework #1") in the form of R Markdown report, knitted into either of the available formats (HTML, pdf or Word). Provide only code and output. NO NEED TO COPY THE PROBLEM FORMULATION (!)

Problem #1 (make sure to include code with outputs)

For the *fl_student_survey.csv* file, proceed to pick a quantitative variable (can be the same as in previous homework):

1. Describe its center via two main corresponding measures introduced in class. Which do you think is more appropriate? Why?
2. Describe its variability via at least two methods (that *do not* use measures of position) from the class. Advantages/Downsides of each method?
3. Provide the plot outlining the five-number summary for measures of position. Are there any outliers? Report the main variability measure. In general, what are the advantages of that measure as opposed to standard deviation?

Problem #2

2.21,

2.34,

2.37 (no cheating with *median*() function; *mean*() is fine),

2.46 (DON'T use *range*() or *sd*(), calculate those yourself, maybe using a calculator),

2.54,

2.62,

2.66,

2.76 (only parts a-b-c; no need to plot)

Problem #3

For NYC Airbnb listings data,

1. Calculate mean and median prices across burrows (*neighborhood_group* variable):

   a. Via for-loop that goes through each burrow.
   b. Via a single vectorized operation (one line of code).

2. Do the same as p.1, only now for median prices. Comment on how the mean and median prices differ for same burrows, and what that can be attributed to.

3. Define your own function $my.median(x)$ that will take vector x of numerical values as input, and output its median. Test it ("sanity check") by using $my.median()$ to calculate median prices for each burrow, and comparing the results with R's built-in $median()$ function (ones you've already got in p.2).

Problem #4

3.3, 3.61, 3.63