WQD7004 Programming for Data Science
Lab 6 Package (dplyr)

1. Write R scripts for each of the following

    a. Create a data frame using the data below.

    id={7,11,8,14,17,13,5,9,10,1}
    read={47,44,57,60,73,44,68,34,52,60}
    write={59,33,44,46,57,46,65,60,52,55}

    b. Return a subset of rows, where read more than or equal to 55 and order the rows by read column.
    c. Return the id (descending) in vector, where write more than or equal to 50.
    d. Return a subset of rows, where duplicate value in read.
    e. Create a new column RW that measure the different between read and write.
    f. Create a new column Status that return R if RW is positive, E if RW is 0 and W if RW is negative.

2. Write R scripts for each of the following

    a. Create a data frame using the data below.

    matric={"S111", "S112", "S115", "S124", "S245", "S331"}
    name={"Ahmad", "Chong", "Mutu", "Paul", "Siti", "Zaidi"}
    test={26, 15, 21, 24, 24, 19}
    asgn={17, 11, 13, 15, 14, 15}
    post={18, 15, 17, 12, 17, 16}
    final={29, 10, 19, 14, 28, 27}

    b. Return the name in vector, where matric begin with S1.
    c. Create a new column CA after the asgn column that add the value of test and asgn.
    d. Create a new column SA after the final column that add the value of post and final.
    e. Create a new column Total that add the value SA and CA.
    f. Create a new column Grade that return the grade based on the Total value using the table below

90 – 100	A+
80 – 89	A
75 – 79	A-
70 – 74	B+
65 – 69	B
60 – 64	B-

55 – 59	C+
50 – 54	C
45 – 49	C-
40 - 44	D+
35 – 39	D
0 - 34	F

g. Calculate the mean and standard deviation of the subject.
h. Get the matric and name of student with highest score and lowest score.
i. Get the number of students that passed and failed the subject given the passing mark is at least 65.

3. Write R scripts for each of the following using **iris** dataset.
    a. Create a data frame from the dataset and then display the data frame in a spreadsheet-style.
    b. Select the versicolor species from the dataset and remove the unnecessary columns.
    c. Sort the data frame by sepal length and width.
    d. Sort the data frame by petal length and width.
    e. Get the summary values for each columns.

4. Write R scripts for each of the following using **ToothGrowth** dataset.
    a. Create a data frame from the dataset and then display the data frame in a spreadsheet-style.
    b. Split the data frame into two, one with supplement type VC and one with OJ. Display both the data frame in a spreadsheet-style.
    c. Select the tooth length and dose from the VC data frame and then sort the data frame by tooth length.
    d. Select the tooth length and dose from the OJ data frame and then sort the data frame by tooth length (descending).
    e. Get the minimum, maximum and average tooth length by dose for VC and OJ. Show the results in a spreadsheet-style.