Fall 2016
MIS 413/572 - Introduction to Big Data Analytics

Quiz 1

1. Please convert built-in dataset "*Titanic*" into an R data frame. Consider the following SQL/R codes.

   1) Replace below SQL code with similar R data aggregation functions.

      ```
      select Class, Sex ,Survived , SUM(Freq)
      from Titanic group by Class, Sex ,Survived
      ```

   2) Replace below R code with SQL code that does similar split-apply-combine operations. Suppose that our *Titanic* data frame is named "tdf".

      ```
      by(data=tdf$Freq, INDICES=list("Sex"= tdf$Sex, "Age"=tdf$Age), FUN=sum)
      ```

2. Consider the following student score data represented as R vectors.

   ```
   Stu_name = c("Reuven Ytzrhak", "Bullwinkle Moose","David Jones",
            "Janice Hammer", "Cheryl Cushing","John Davis", "Greg Knox",
            "Joel England", "Mary Rayburn", "Angela Williams")
   Chinese = c(50, 60, 41, 85, 69, 51, 40, 65, 87, 92)
   Math = c(95, 99, 80, 82, 75, 85, 80, 95, 89, 86)
   English = c(25, 22, 18, 15, 20, 28, 15, 30, 27, 18)
   ```

   1) Please create an R data frame "*score"* by combining these 4 vectors.

   2) Calculate a weighted score "ws" with weights (0.4, 0.4, 0.2) for *Chinese*, *Math*, and *English* respectively. Then, add the new column "ws" to the data frame.

   3) Split and replace the students' name *Stu_name* with two new columns : "First_Name" and "Last_Name". (Hint: You may use *base::strsplit*()).