CS 5301-Assignment 3

Purpose: The purpose of this assignment is to test concepts of data frame, logical subsetting, and file import.

What to submit: For this assignment, you should submit a single RMD file that contains the code and results of the questions below. Your code should be organized in proper sections and comments to match the tasks in this assignment. Include your name in the beginning of the RMD file.

You are a given a text file "**patient.csv**" that has three columns. The first row should be treated as the names of the columns. There are no row names. This file contains measurement of various physiological metrics of a patient admitted to a hospital. The first column represents the time (in hours) the measurements were taken, the second column indicates the name of the metric, and the third column indicates the recorded value. All missing values are marked as NA.

A. Read this file into a data frame and write R code to answer the questions B, C, D. [10 points]
B. At how many different time points, the person's temperature ("Temp" in column 2) was taken. Ignore missing items. What are those time points? [10 points]
C. At how many different time points, the person's Heart rate ("HR" in column 2) was measured. Ignore missing items. What are those time points? [10 points]
D. Calculate the average temperature ("Temp" in column 2), and average Heart rate ("HR" in column 2) of the patient during the stay. Ignore missing items. [10 points]
E. In the first column of the data frame, replace the values by following the rules in the table below. After modifying, write the resulting data frame to a text file. [10 points]

Between 0 and 8.0 (0 included, 8.0 included)	Replace by 1
Between 8.1 and 16.0	Replace by 2
Between 16.1 and 24.0	Replace by 3
Between 24.1 and 32.0	Replace by 4
Between 32.1 and 40.0	Replace by 5
Between 40.1 and 48.0	Replace by 6

2. You are given a csv file (**outcomes.csv**) that represents the outcomes of patients admitted to ICU. There are 4 columns
    A. RecordID: record id of a patient.

B. Length of stay (days): number of days between the patient's admission to the ICU and the end of hospitalization. Missing values are indicated as NA.
C. Survival (days): If the patient's death was recorded (in or out of hospital), then *Survival* is the number of days between ICU admission and death; otherwise, *Survival* is assigned the value -1.
D. In-hospital death (0: survived during hospital stay, or 1: died-in-hospital)

Write R code to find the following: [50 points]

A. Number of patients who survived during their stay in hospital.
B. Number of patients who died during their stay in hospital.
C. Average of **number of survival days** among the patients who died-in-hospital.
D. Among the patients who survived during their stay in hospital, how many of them stayed in the hospital for least 5 days.
E. Among the patients who survived during their stay in hospital, how many of them stayed in the hospital less than 7 days.