MTH 443: Lab Problem Set 6

[1] The dataset "CHD.csv" gives data on
(i) total cholesterol (totChol),
(ii) systolic blood pressure (sysBP),
(iii) diastolic blood pressure (diaBP);
of two different groups of individuals
(a) individuals who have 10 year risk of Coronary Heart Disease (CHD); value of the variable TenYearCHD in the dataset is "1" for this group and
(b) individuals who do not have 10 year risk of Coronary Heart Disease (CHD); value of the variable TenYearCHD in the dataset is "0" for this group.

(A) Obtain density estimates of the three variables "totChol", "sysBP", "diaBP" for the two different groups using k-nearest neighbor approach and kernel density estimation (with Gaussian kernel, rectangular kernel, triangular kernel and epanechnikov kernel).
(B) Do you observe any difference in the estimated densities of the two groups?
(C) Find the estimate of the tail probabilities of the variables at different threshold levels and compare the estimated probabilities of the two groups.
(D) Find an estimate of joint density of "sysBP" and "diaBP".

NOTE: You would observe in the dataset that some observations have "*NA*" as their value; use class average to impute these missing values.