2a. Was not completed

2b. For the 4% missing dataset the mean was the most accurate, because not much data was missing so averaging out the values yielded a small value.

2c. For the 20% missing dataset the HD_Conditional was the most accurate, because it compares every value to one another.

2d. The two datasets were overall different because with 4% missing there is more known data, so there is more to work with. Whereas when working with the 20% data more of the data would have to be imputed being less accurate.

2e. Mean is definitely faster for both 4 and 20 % missing data.