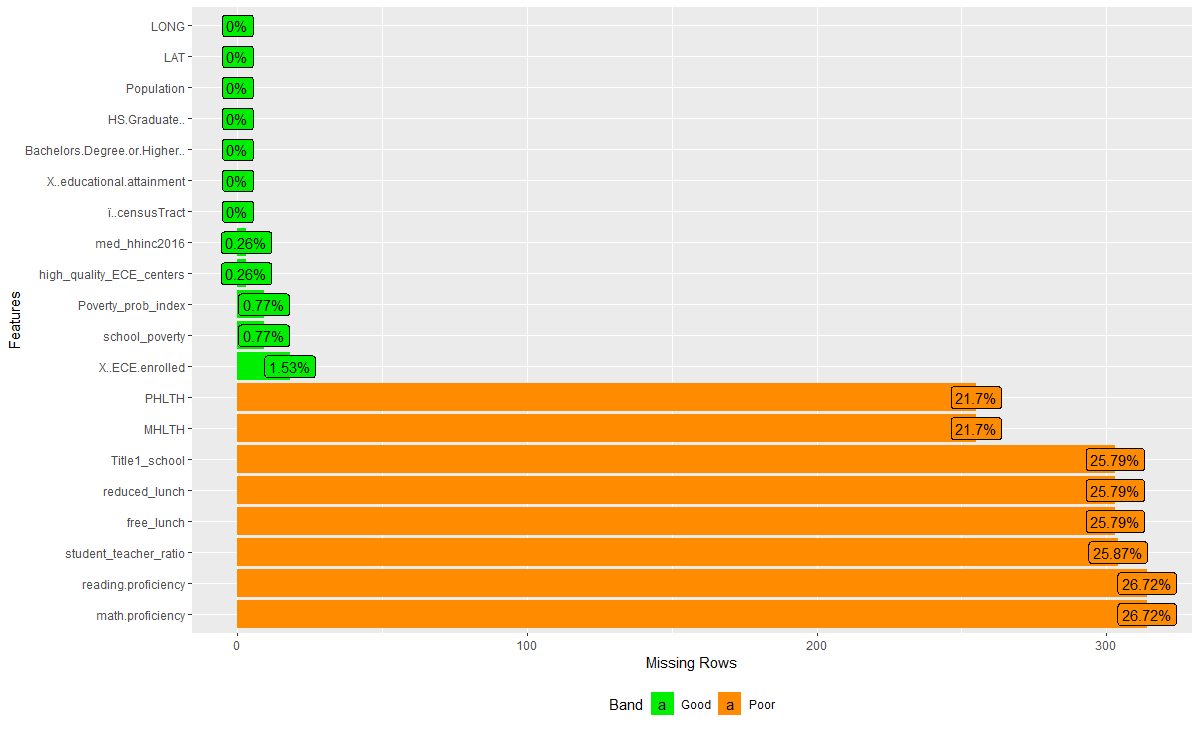
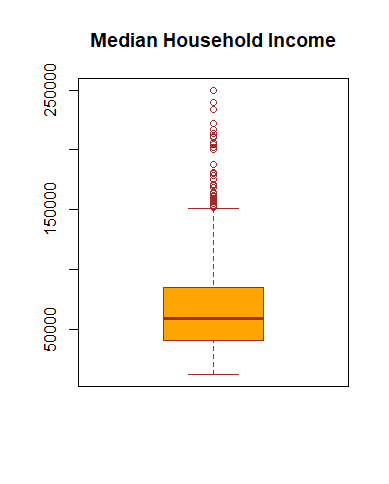
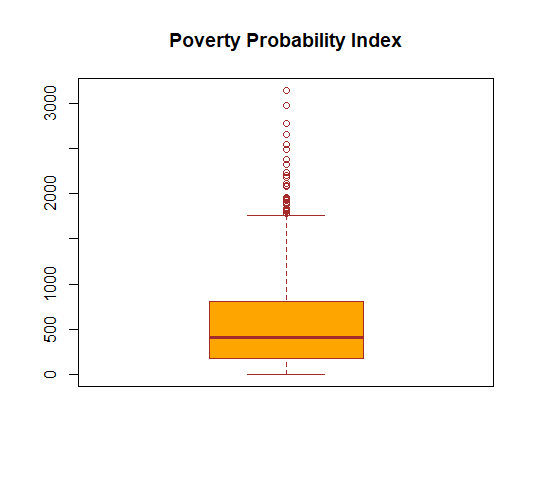
# HANDLING MISSING VALUES - KNN IMPUTATION

* Our dataset has a total of 11 features and 4 target variables. All 4 target variables and 9 features had missing data



* We filtered out tracts that had zero population
* We analyzed the data distribution of all the missing features and target variables using a box plot. We observed there were outliers.



* We ran the KNN algorithm to get the nearest neighbors based on the latitude and longitude of the tracts.
* We found 5 nearest neighbors for our features and 4 nearest neighbors for our target variables.
* Based on the boxplot distribution we imputed the median value of the neighbors where the variables had outliers while the others were imputed with the mean of the nearest neighbors