# Enterprise Database Technologies

# CA1

## Getting to know the Dataset using R

Next I moved on to find the percentage of missing values within the data. I ran a command on the dataset to query how many values were Na in the dataset. This returned me a result of 7 Na’s which gave me a percentage of 0.1517. I then ran a summary command on the dataset in order to quickly see which values were in Na. This showed me that the Cholesterol, Restecg and Class all had missing values and would need to have values imputed later on.

Next I found the max, min, mean, mode, median and the Standard deviation of the data. The standard deviation was used to tell me how dispersed the data would be; A low standard deviation indicated it was closer to the mean and less dispersed, while a higher standard deviation indicated it was more dispersed.