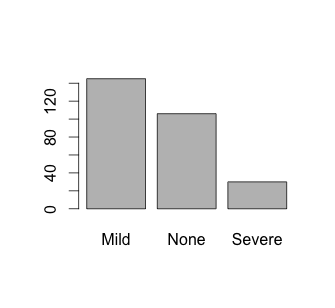
Ex12

1.

a. The classes in the data set is imbalanced. so stratified samplring to split the data in to training and testing set should be used.



b. Kappa statistic is used because the class imbalance and having more than two classes.

c. Biological predictors

Pre-Processing steps

1. Near Zero variance check
2. High correlation between predictors
3. Linear Combination between predictors
4. Centering and scaling
5. Data splitting with createDataPartition() function with 80% training set.