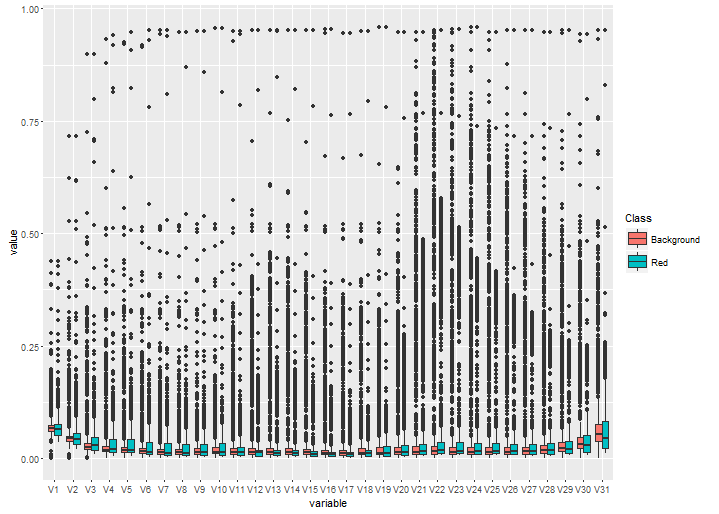
## Question 1



**Figure1**. plot of the distributions of data from the two pixel classes in the reference library training.csv

## Question 2

**Training accuracy of the classifier(LDA)**

Confusion Matrix and Statistics

Reference

Prediction Background Red

Background 664 277

Red 140 527

Accuracy : 0.7407

95% CI : (0.7185, 0.7619)

No Information Rate : 0.5

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.4813

Mcnemar's Test P-Value : 2.739e-11

Sensitivity : 0.8259

Specificity : 0.6555

Pos Pred Value : 0.7056

Neg Pred Value : 0.7901

Prevalence : 0.5000

Detection Rate : 0.4129

Detection Prevalence : 0.5852

Balanced Accuracy : 0.7407

'Positive' Class : Background

## Question 3

**Testing accuracy of the classifier(LDA)**

Confusion Matrix and Statistics

Reference

Prediction Background Red

Background 819 328

Red 170 661

Accuracy : 0.7482

95% CI : (0.7285, 0.7672)

No Information Rate : 0.5

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.4965

Mcnemar's Test P-Value : 1.988e-12

Sensitivity : 0.8281

Specificity : 0.6684

Pos Pred Value : 0.7140

Neg Pred Value : 0.7954

Prevalence : 0.5000

Detection Rate : 0.4141

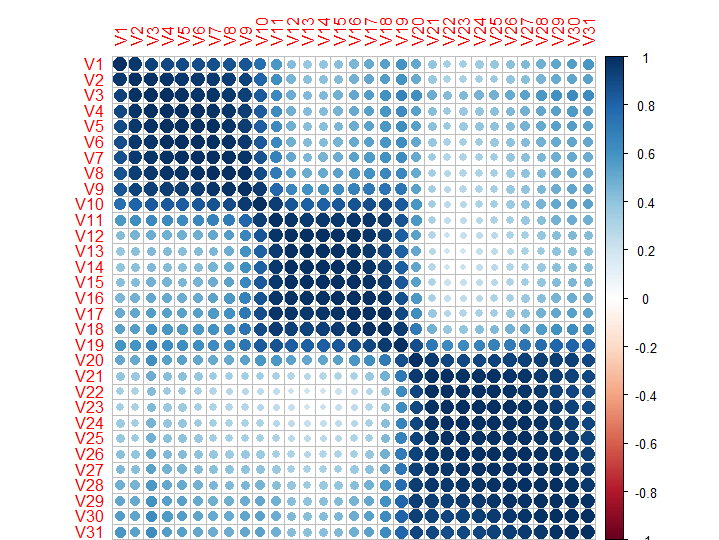
Detection Prevalence : 0.5799

Balanced Accuracy : 0.7482

'Positive' Class : Background

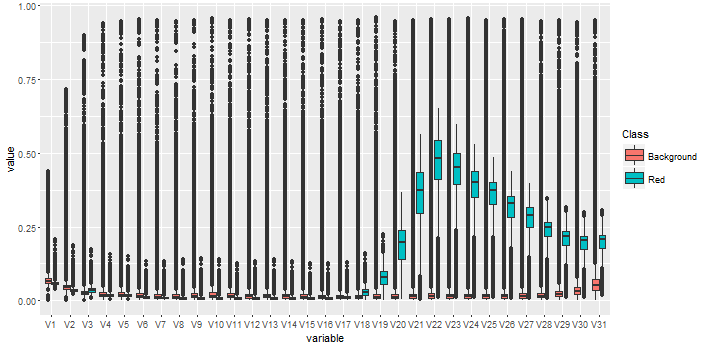
## Question 4

### Question 4.1



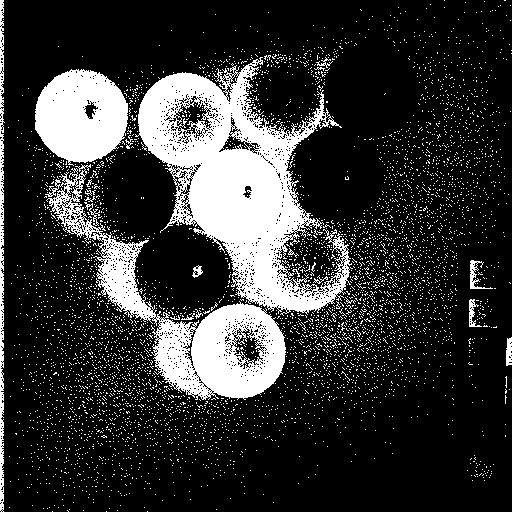
**Figure2**. plot of the correlation between each feature

### Question 4.2



**Figure3**.Distribution of pixels in the Raw Image Data

## Question 5



**Figure4.** classified results image indicating which pixels were classified as belonging to the Red superballs in white ;and which pixels were classified as background in Black

## Question 6

**The accuracy comparing the predictive algorithm to the actual status of the pixels in the image**

Confusion Matrix and Statistics

Reference

Prediction Background Red

Background 207151 1704

Red 41740 11549

Accuracy : 0.8343

95% CI : (0.8328, 0.8357)

No Information Rate : 0.9494

P-Value [Acc > NIR] : 1

Kappa : 0.2896

Mcnemar's Test P-Value : <2e-16

Sensitivity : 0.8323

Specificity : 0.8714

Pos Pred Value : 0.9918

Neg Pred Value : 0.2167

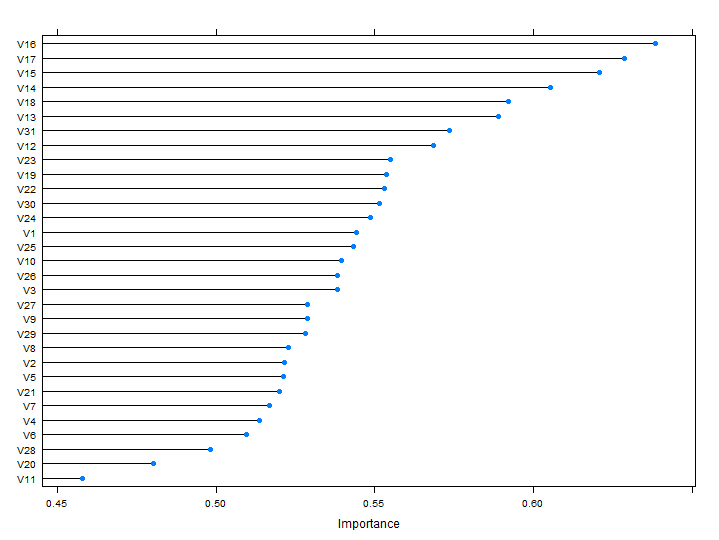
Prevalence : 0.9494

Detection Rate : 0.7902

Detection Prevalence : 0.7967

Balanced Accuracy : 0.8519

## Question 7



**Figure 5** Importance of each feature used in the LDA algorithm

## Question 8

**Compare the accuracy of three classifier types(LDA,Decision Tree,KNN)**

Accuracy

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

lda 0.7239 0.7267 0.7540 0.7464 0.7587 0.7687 0

DT 0.7425 0.7463 0.7575 0.7556 0.7627 0.7689 0

knn 0.8520 0.8596 0.8634 0.8650 0.8657 0.8843 0

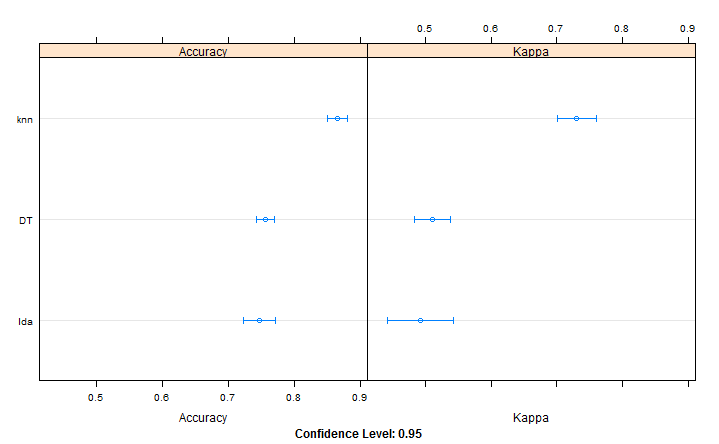
Kappa

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

lda 0.4478 0.4533 0.5081 0.4928 0.5174 0.5373 0

DT 0.4851 0.4925 0.5149 0.5112 0.5254 0.5379 0

knn 0.7040 0.7192 0.7267 0.7300 0.7313 0.7687 0



## Question 9

**Compare the accuracy of the KNN classifier with and without a pre-processing method(centering,scale)**

Accuracy

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

knn 0.8520 0.8596 0.8634 0.8650 0.8657 0.8843 0

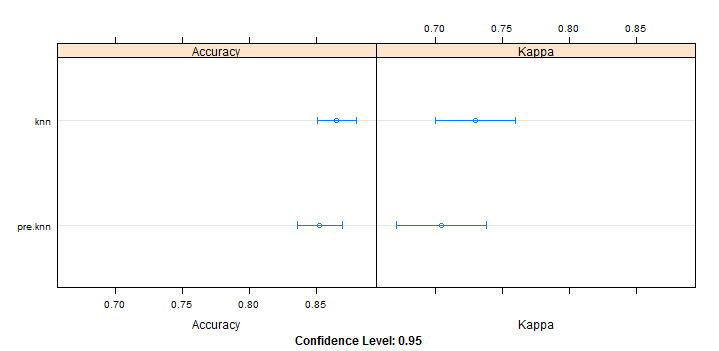
pre.knn 0.8408 0.8420 0.8460 0.8523 0.8609 0.8719 0

Kappa

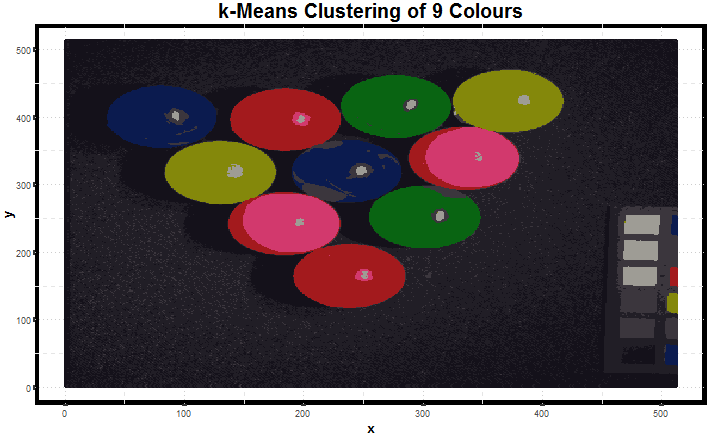
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

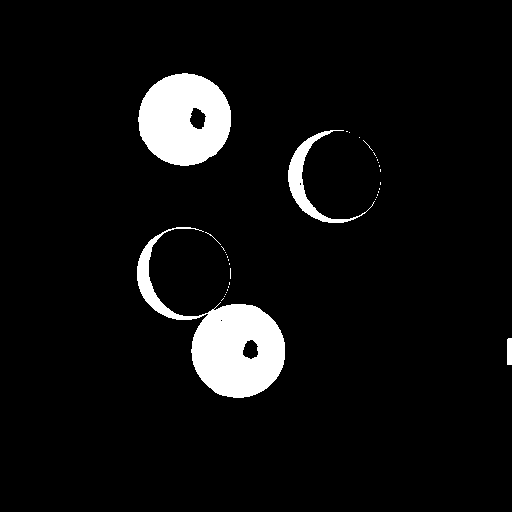
knn 0.7040 0.7192 0.7267 0.7300 0.7313 0.7687 0

pre.knn 0.6816 0.6841 0.6919 0.7046 0.7217 0.7438 0



## Question 10





**Figure** classified results image used in the **k-Means** algorithm indicating which pixels were classified as belonging to the Red superballs in white ;and which pixels were classified as background in Black

Confusion Matrix and Statistics

Reference

Prediction Background Red

Background 245655 484

Red 3103 12902

Accuracy : 0.9863

95% CI : (0.9859, 0.9868)

No Information Rate : 0.9489

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.8708

Mcnemar's Test P-Value : < 2.2e-16

Sensitivity : 0.9875

Specificity : 0.9638

Pos Pred Value : 0.9980

Neg Pred Value : 0.8061

Prevalence : 0.9489

Detection Rate : 0.9371

Detection Prevalence : 0.9389

Balanced Accuracy : 0.9757

'Positive' Class : Background