**Classification using**

**Linear Discriminant Analysis**

**Banknotes Data**

**Data Set Information**

Data were extracted from 1372 images that were taken from genuine and forged banknote-like specimens. There are no missing values in the data set. For digitization, an industrial camera usually used for print inspection was used. The final images have 400 x 400 pixels. Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained. Wavelet Transform tools were used to extract four features from the images: Variance, Skewness, Kurtosis and Entropy. I highlight each variable in purple that is also used in my R scripts as variables or outputs. I used R studio version 0.99.902 to satisfy all of my programming needs. Each script has notes and comments explaining how the following analysis was employed. I use set.seed(20) to promote replicability of the outputs.

**ATTRIBUTES**

1. Variance of Wavelet Transformed Image (Variance WTI)

Variance WTI is the Variance of the Fourier coefficients. We can assume that the forgery process would result in a different variance of the Fourier coefficients. This could be due to paper or ink composition which differs from authentic banknotes.

1. Skewness of Wavelet Transformed Image (Skewness WTI)

Skewness WTI is the Skewness of the Fourier coefficients. This is similar to Variance WTI of the Fourier coefficients in that the forgery process would result in a different skewness. Once again, the physical properties of the paper, ink, press, etcetera could account for this difference.

1. Kurtosis of Wavelet Transformed Image (Kurtosis WTI)

Kurtosis WTI is formally described as the standardized-fourth population moment about the mean. Kurtosis represents the movement of mass within a p.d.f. in such a way that the variance is unchanged. Kurtosis can be seen graphically by how much weight is placed on the tails and the height of the peak. Positive Kurtosis is when the tails are heavier and the peak is higher. Negative Kurtosis is when the shoulder and center of the p.d.f. are fattened while the tails are flattened.

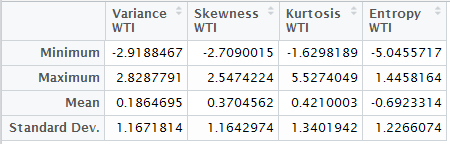
1. Entropy of Wavelet Transformed Image (Entropy WTI)

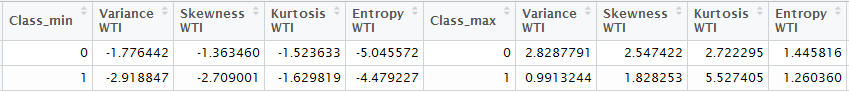
Entropy WTI is referring to the Spectral Entropy which describes the complexity of a system. In our case, with wavelets, it is describing how much noise is present given a certain resolution, or number of bins. Low Entropy is characterized by a concentration in the frequency domain (in exemplum: punctuated amplitude by a burst of energy in a tight frequency interval). High Entropy is characterized by a broad interval in the frequency domain (in exemplum: white noise or a deterministic chaotic system).

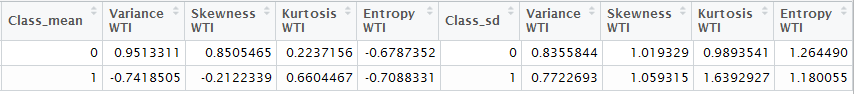
**CLASSES**

1. Forged Banknotes are represented by Class 0 with a total amount equal to 610.
2. Genuine Banknote are represented by Class 1 with a total amount equal to 762.

**PRECONDITIONING**

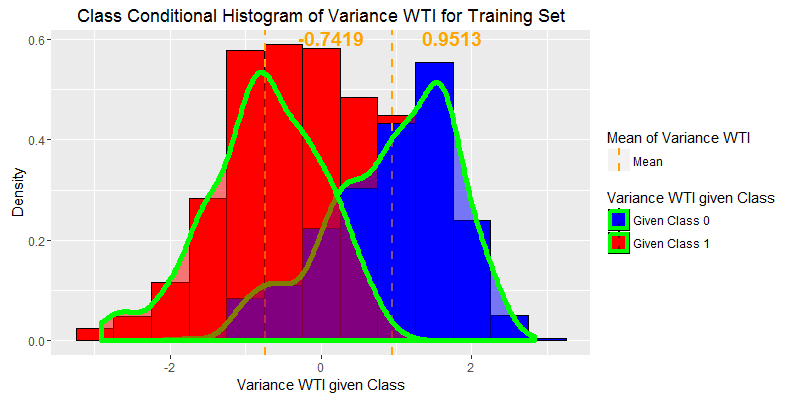
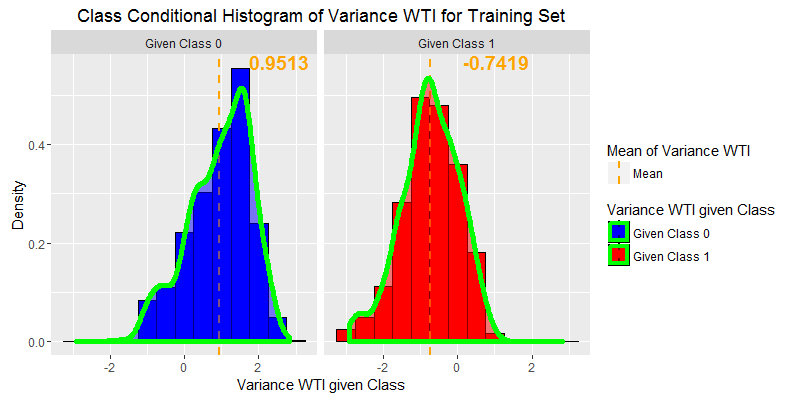
1. Before I perform LDA, I normalize the banknotes data set by dividing each point by that column’s respective absolute mean. So that for all , each = , for all . This procedure will help the covariance matrix to not be ill-conditioned.
2. Now banknotes is ready to be split into a training and a testing data frame. I use ‘createDataPartition’. This ensures that there is good class balance between Training and Testing. I chose to allocate 80% of the data points to Training and the remaining 20% to Testing. There are 1098 objects in Training, of which 602(54.83%) are Class 0 and 496(45.17%) are Class 1. There are 274 objects in Testing, of which 160(58.39%) are Class 0 and 114(41.61%) are Class 1.
3. Next, I calculate basic statistics for Training as a whole and Training given the Class:
   1. Training Statistics:  
        
      
   2. Training given Class Statistics:



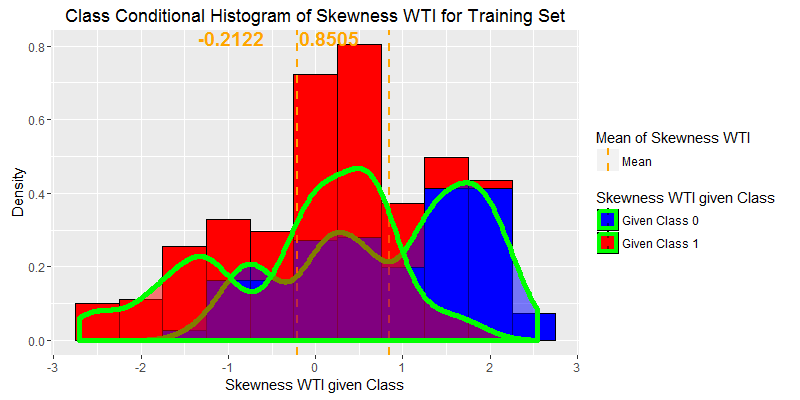
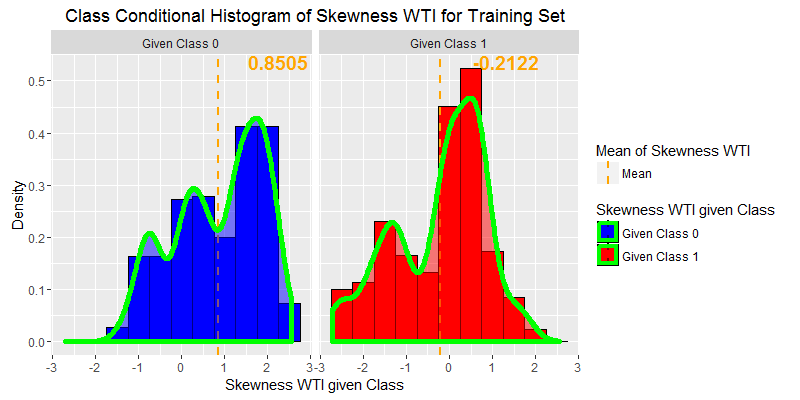


**Histogram Analysis for Training**

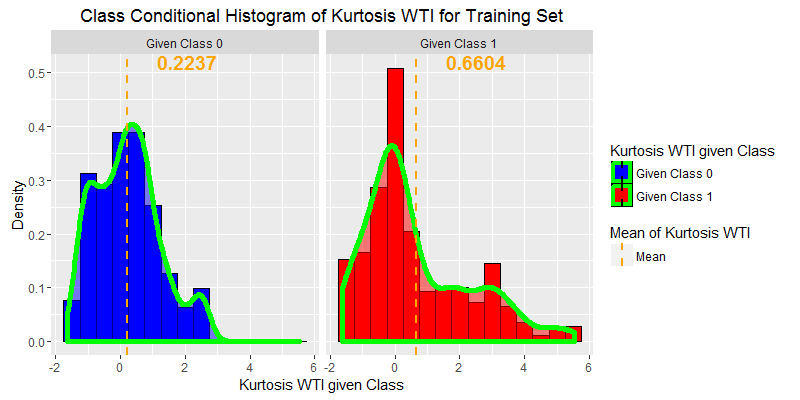
Now that I have useful statistics about my data and it’s organized in a convenient way, I can plot histograms to see how the class conditional statistics compare within each attribute. I use **library**(ggplot2) since it provides the most flexibility and has the best visual appeal. I partition Training into temporary data frames out\_zero and out\_one then plot the histograms of each attribute given Class. I plot the p.d.f. curves in green on top of the histograms to better illustrate similarities and differences amongst the distributions. I use an orange dashed line to represent the mean of the distribution. The standard deviation of the distribution can be referred to in the table. The standard deviations were close enough in value that vertical lines convoluted the graph. I decided to include faceted graphs of each attribute in addition to overlaid graphs of each attribute. Even though the scales are the same on the faceted graph, I found the overlaid graph convenient to compare the class distribution within each attribute. Objects classified as Class 0(forged) are **blue** and objects classified as Class 1(genuine) are **red**.

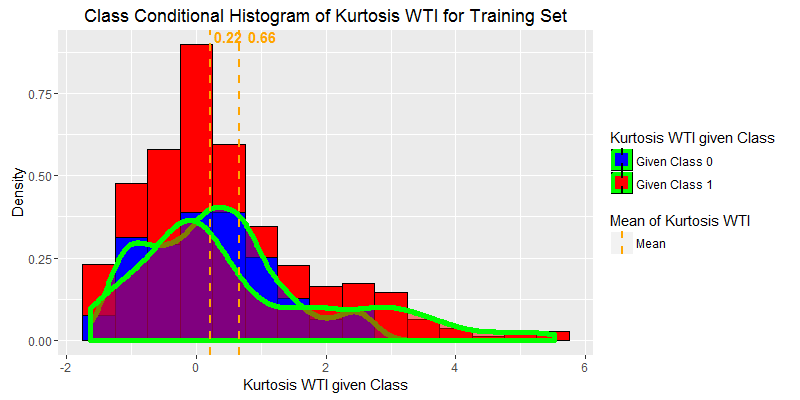


We can see that Variance WTI gives good class separation. The p.d.f. for each class is similar but there is a significant distance between the class means. Class 0 s.d. = 0.8356 and Class 1 s.d. = 0.7723.

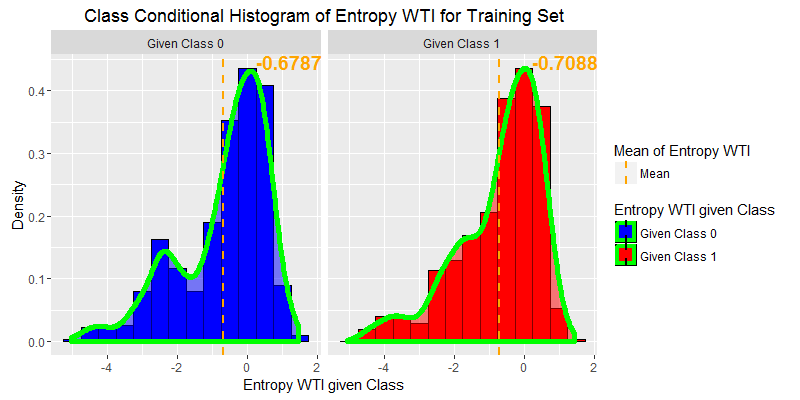


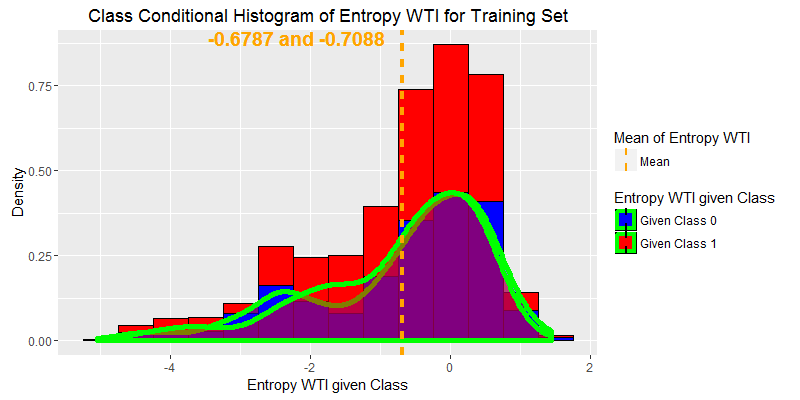
Skewness WTI does not quite have the separation as Variance WTI but there is still a noteworthy gap between the two class means. Class 1 is much more biased to negative values. I suspect Skewness is helpful in determining class. Class 0 s.d. = 1.0193 and Class 1 s.d. = 1.0593.

****



Kurtosis WTI is worse than both the above two attributes. The class means are much closer in value. Thankfully, Class 1 is skewed right, which will still aid in class discrimination. Class 0 s.d. = 0.9894 and Class 1 s.d. = 1.639.

****

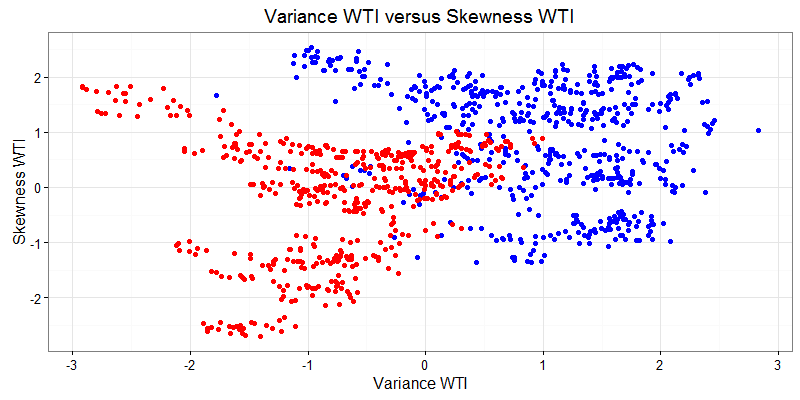


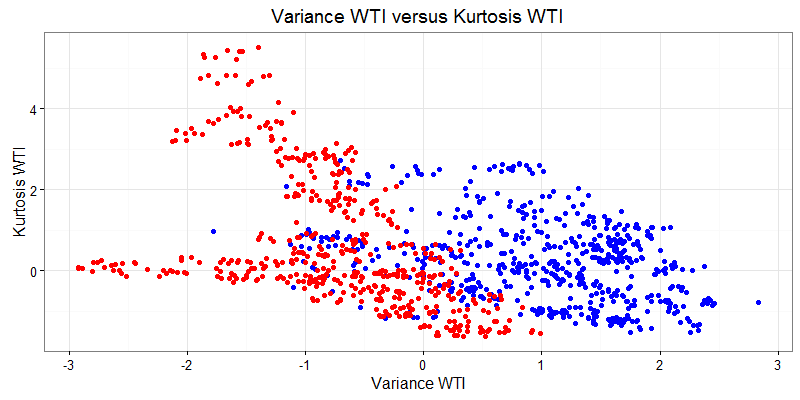
We can easily see that Entropy WTI is a terrible indicator as to which class our objects fall. The distributions sit basically right on top of each other and the class means are very close in value. Class 0 s.d. = 1.264 and Class 1 s.d. = 1.180.

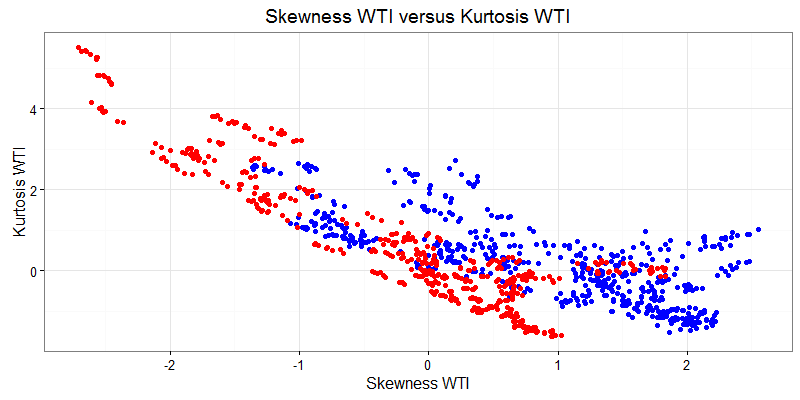
In conclusion, based on the above histograms I find it evident that the best three attributes for class discrimination are Variance WTI, Skewness WTI and Kurtosis WTI.

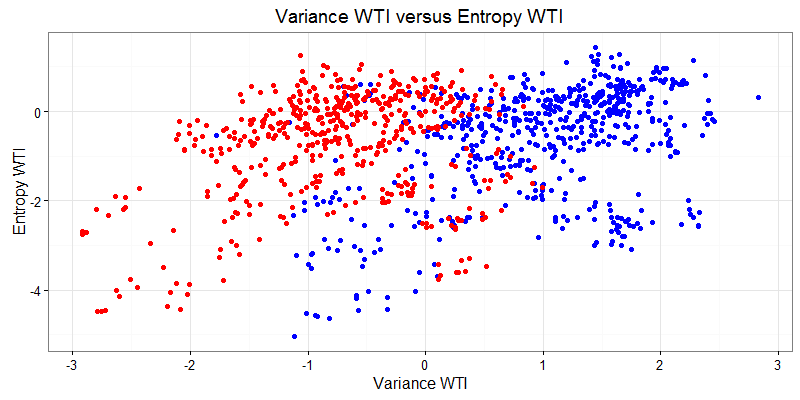
**2D Graphing of Attributes**

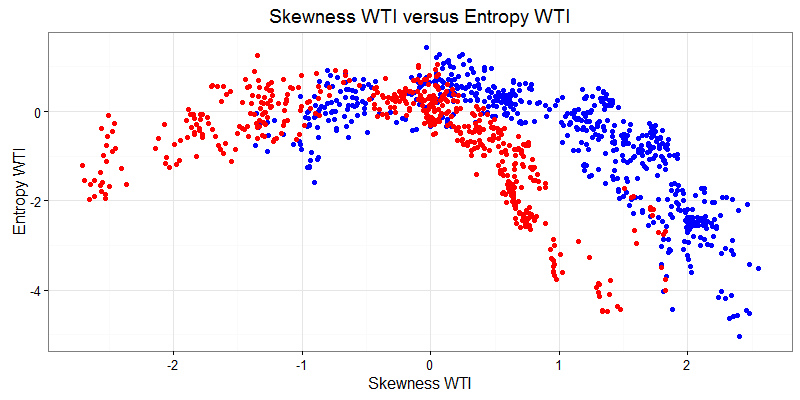
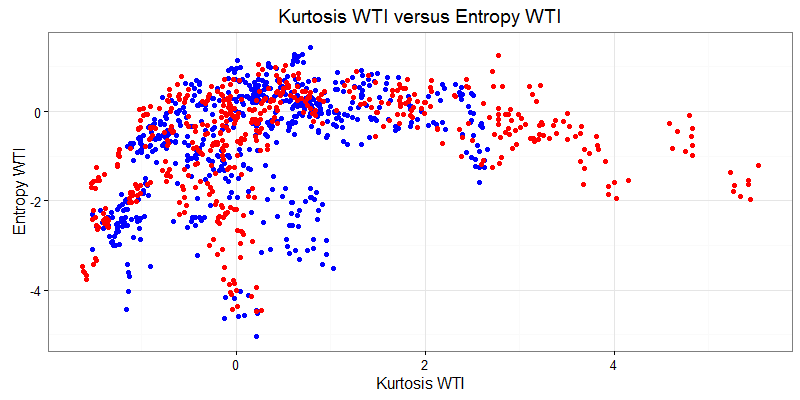
Now that we have seen class relationship amongst the attributes looking at their histograms, let’s try to see what plotting the attributes against each other tells us. Objects classified as Class 0(forged) are **blue** and objects classified as Class 1(genuine) are **red**.







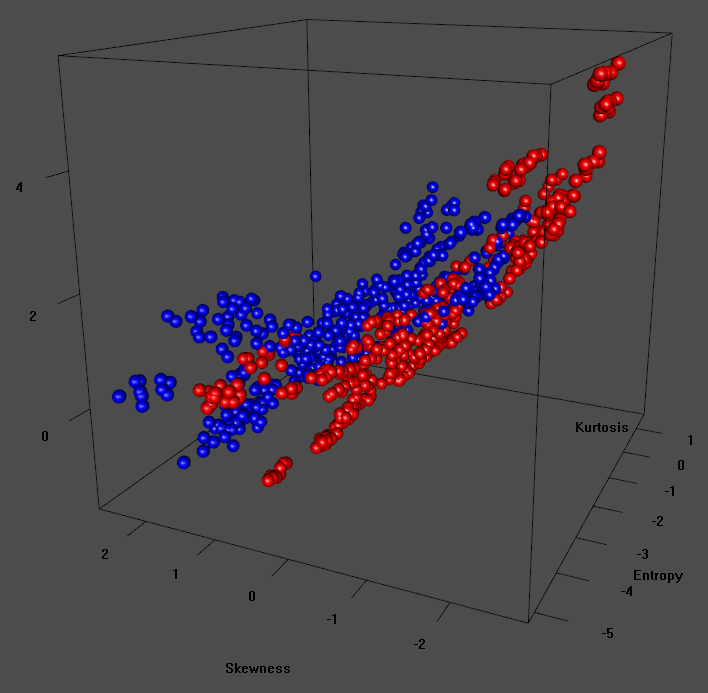


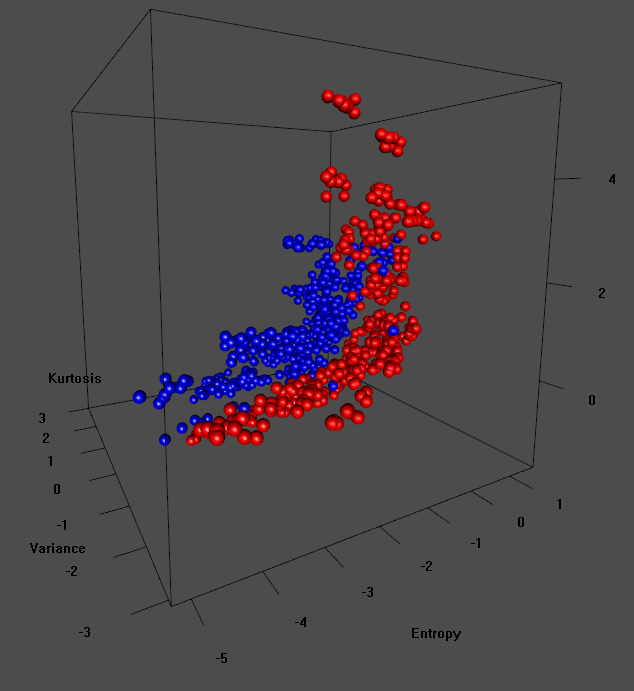
 

We can see that Variance has very good class separation against the other attributes. Skewness is decent against Kurtosis and Entropy. Kurtosis is terrible against Entropy. Entropy, as expected from our histogram analysis boasts the worst aptitude in separating the classes. Maybe graphs in will be more telling.

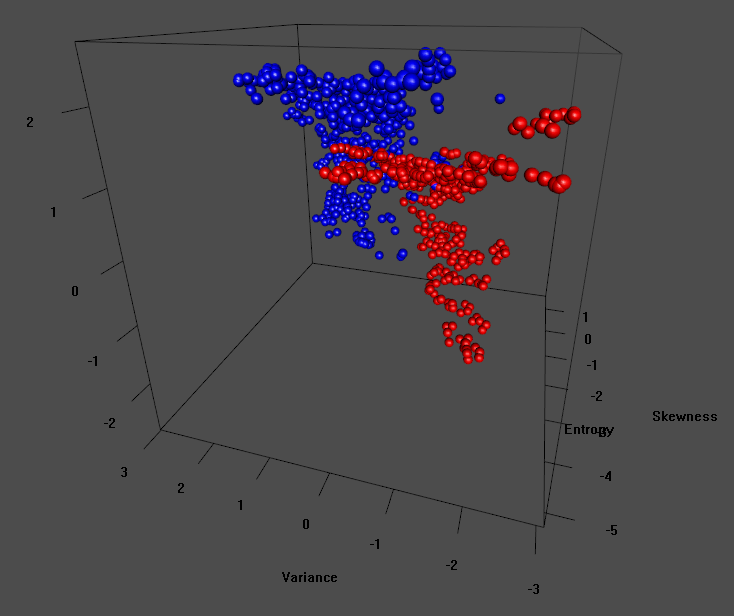
**3D Graphing of Attributes**

**Skewness vs. Kurtosis vs. Entropy**

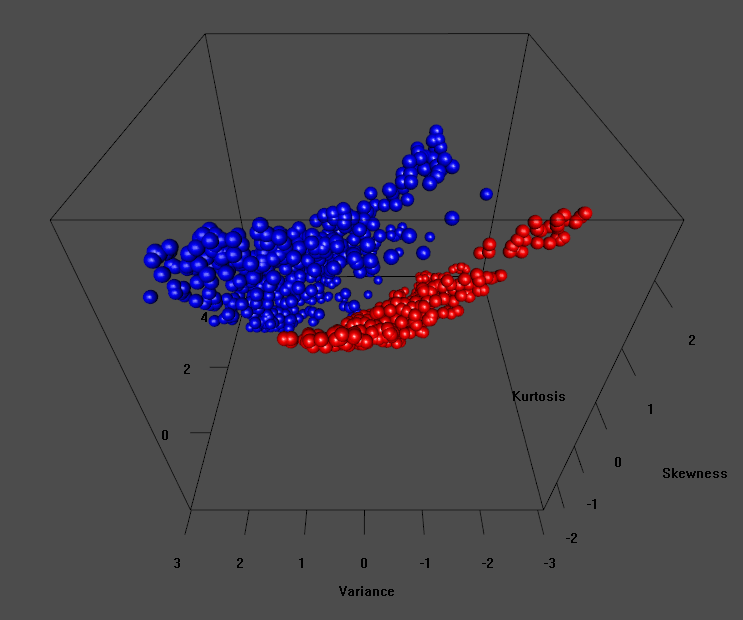


**Variance vs. Kurtosis vs. Entropy**

**Variance vs. Skewness vs. Entropy**



**Variance vs. Skewness vs. Kurtosis**



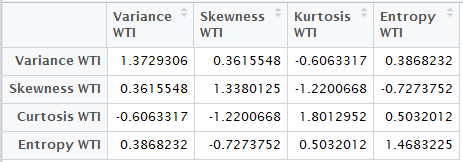
The above 3D graphs were generated in rgl and spun to reflect the best separating angle. The final graph illustrates the three best attributes’ superior class discriminating prowess.

**Test-Statistic**



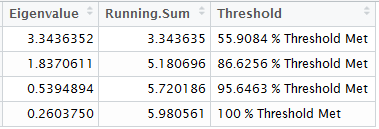
We can see from the above table that the test-statistic confirms that the first three attributes have the best discriminatory power. Variance WTI shows to have the best, as was the case in the histogram and graphing analysis. Finally, we will look at an eigenvalue analysis of the covariance matrix, ∑.

**Covariance Matrix**

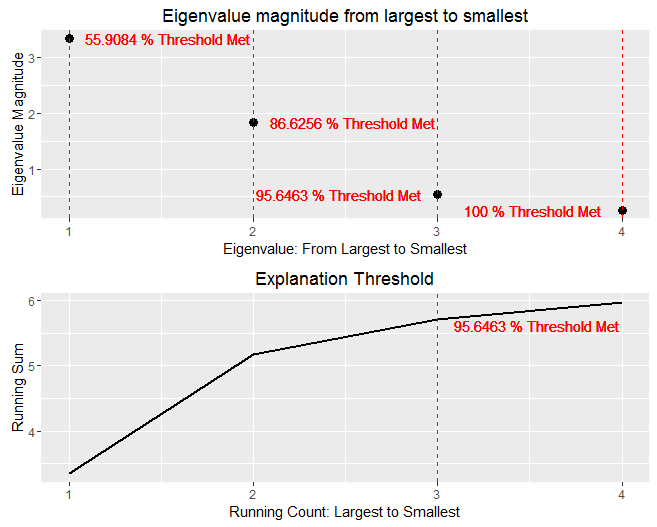


The covariance Matrix is well defined so there should be no issue manipulating this for our learning algorithm.

**Eigenvalues & Running Sum & Threshold  
for Covariance Matrix**



The following eigenvalue plots help to illustrate the percentage of the data explained by the eigenvectors of the covariance matrix.

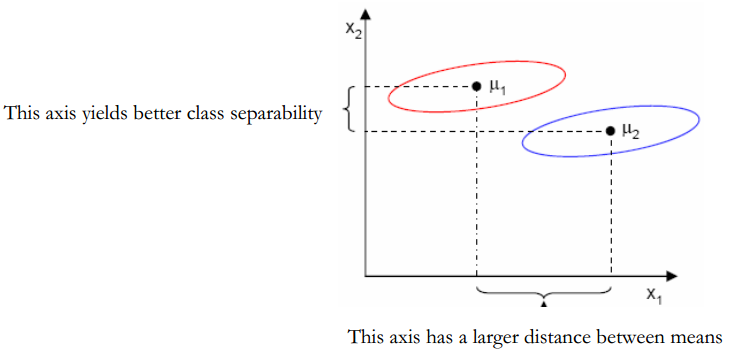


We can see that 95.65% of the information is contained in the first three eigenvectors. The remaining 4.35% of information contained in the last vector is proportional to the error of the residue if we were to use these results in principal component analysis. However, the dimensions of our data set, 1372x4, is small enough not to warrant such a technique.

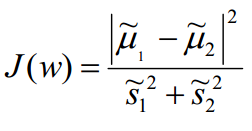
Now that we have inspected the data we can move on to classification. I will first give a very brief explanation to Linear Discriminant Analysis.

**Linear Discriminant Analysis**

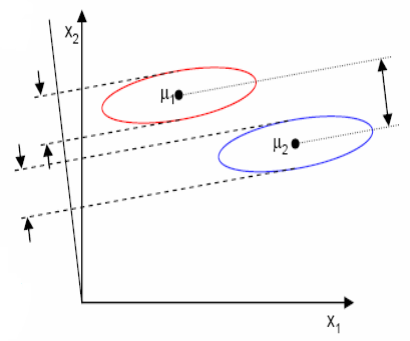
I implement Linear Discriminant Analysis (LDA) on the Banknote Authentication Data Set. LDA aims to maximize a function that represents the difference between the class means, normalized by the within-class variability, called scatter. This normalization helps to reduce error caused by means that are well-separated but with overlapping variance as in the following graph.



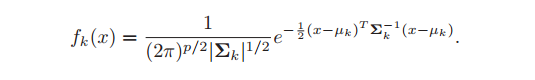
The Fischer linear discriminant function :



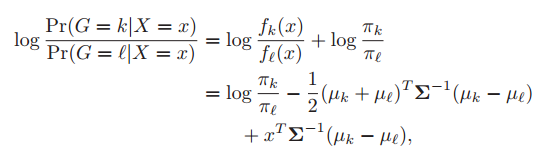
Maximizing ensures that we achieve a projection where examples from the same class are projected very close to each other while the class means are projected as far apart as possible. The following graph illustrates how the slight rotation of axes allows for better separability.



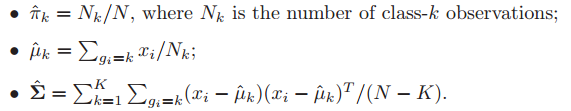
With LDA, we assume that the data arise from Gaussian distribution and that the covariance matrix is the same for all classes. This causes the decision boundaries to be linear, by looking at the log-ratio of the class- specific Gaussian densities, :



The log-ratio leads to:



Where class frequency , class mean , and the covariance matrix , are estimated by the training set as follows:



With just two classes the LDA rule classifies to class 2 if:



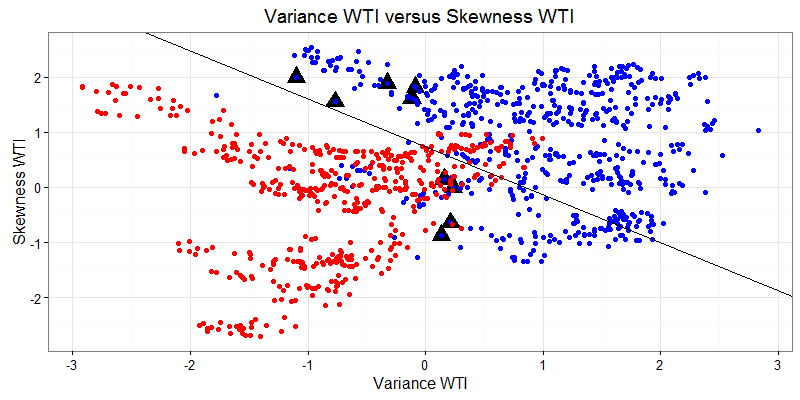
and to class 1 otherwise.

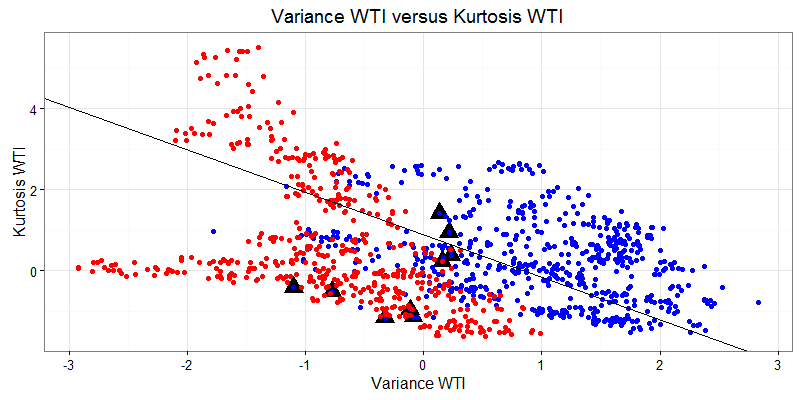
**Implementing LDA and Graphical Results**

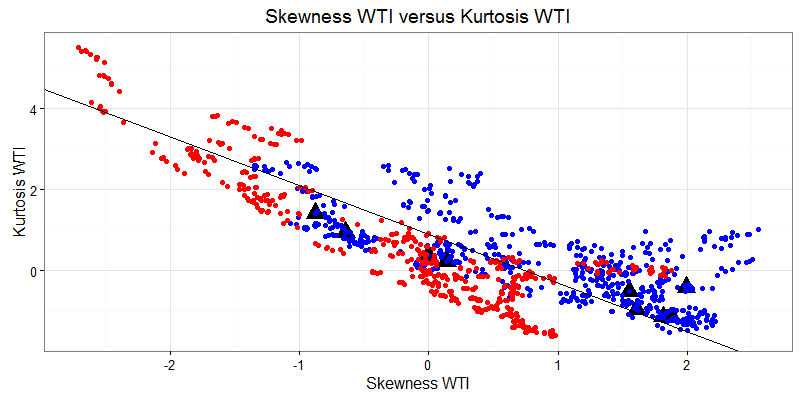
LDA is implemented in R using the MASS package. I save the predicted values into Training and Testing under the column named Pred. I assigned as prior probabilities the class priors from Training.

**2D Graphs with Hyperplanes**

First, let us take a look at the graphs in 2D. I chose to render only the three best attributes (Variance WTI, Skewness WTI and Kurtosis WTI). The black triangles represent the misclassifications after implementing LDA. Keep in mind that each of these lines is a projection on the two chosen variables from the coefficients generated by LDA.

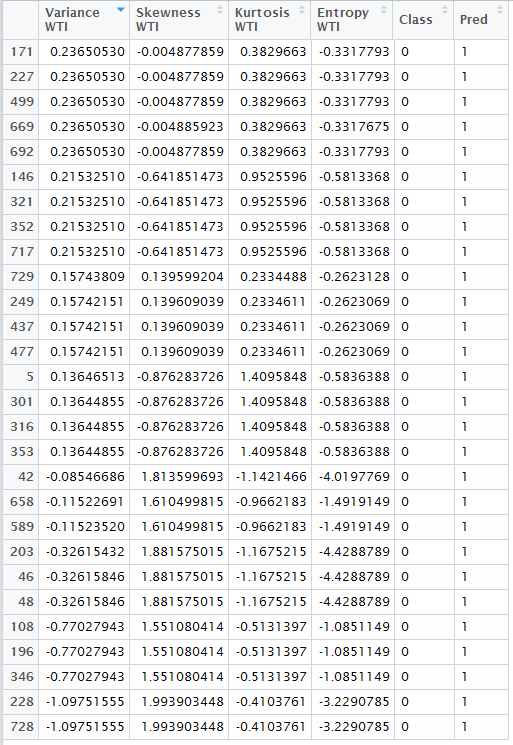






We can see from the above graphs that Variance WTI versus Skewness WTI has the best discrimination. Also, there were 28 total misclassifications but only 13 actual misclassifications\*. The remaining 15 are duplicate points. See below:

**Chart of duplicate points:**

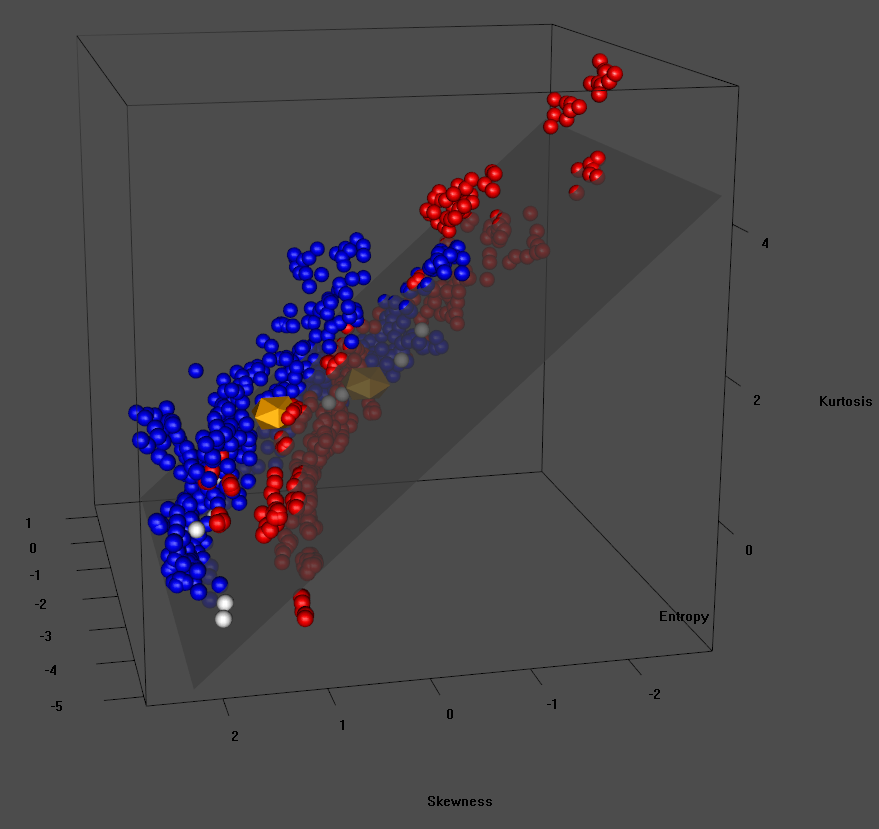


\*It may have been wise to remove these duplicate points since they are not adding any new information.

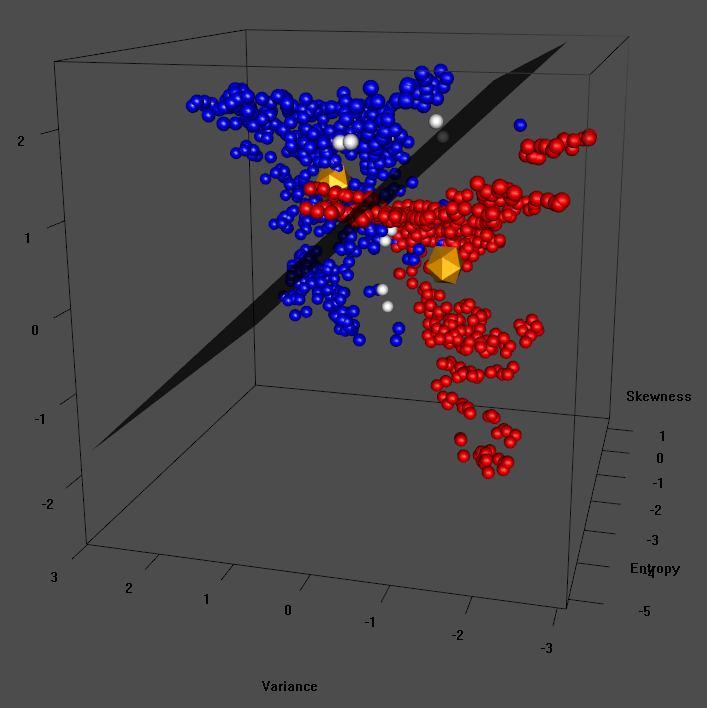
**3D Graphs with Hyperplanes**

I will include four 3D graphs consisting of all permutations of attributes. I chose not to leave out Entropy since I would only be left with one graph to study. The orange icosahedrons denote the class means. The white spheres represent the 13 misclassifications. The hyperplane equation was generated from the LDA coefficients.

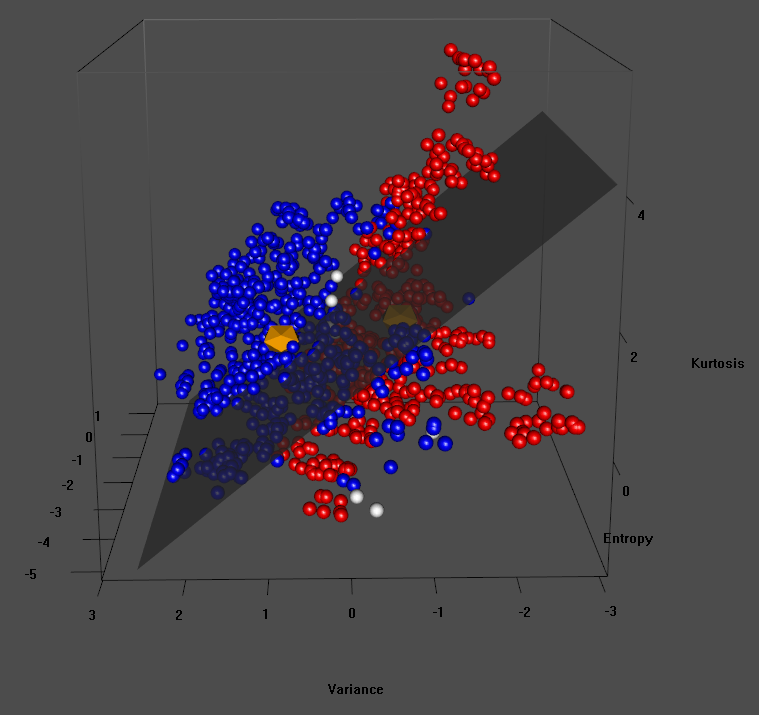
**Skewness versus Kurtosis versus Entropy**



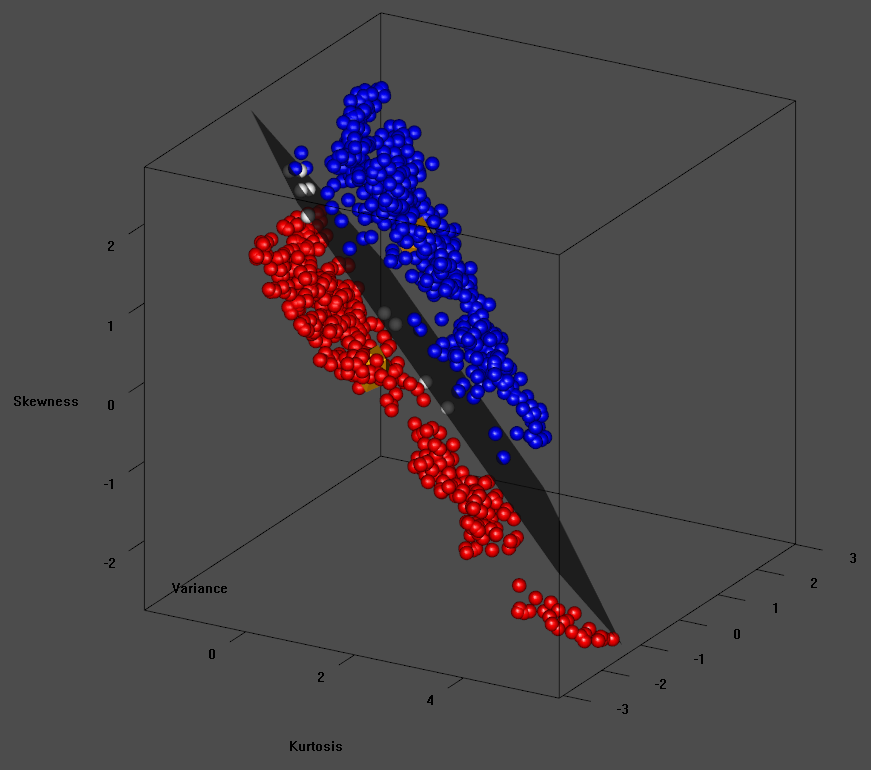
**Variance versus Skewness versus Entropy**

****

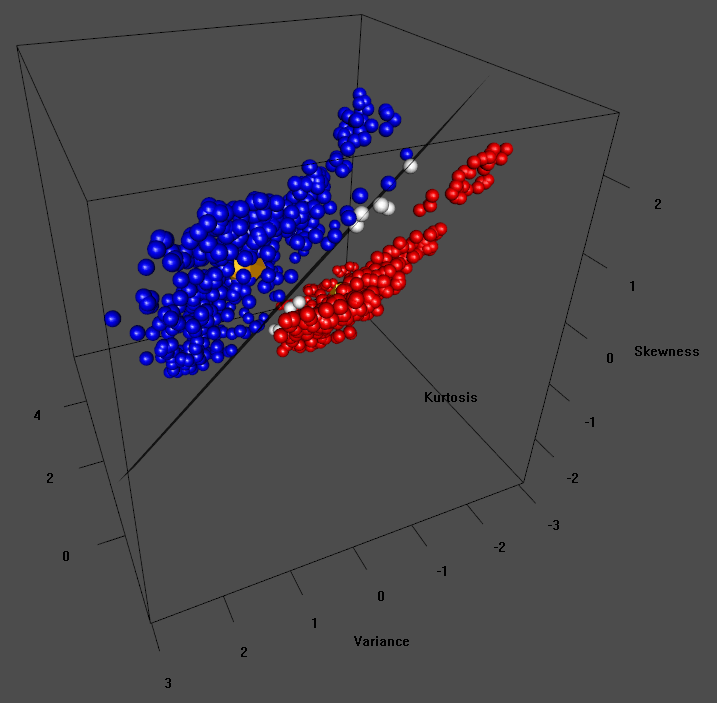
**Variance versus Kurtosis versus Entropy**



**Variance versus Skewness versus Kurtosis**



**Variance versus Skewness versus Kurtosis**



I included two 3D graphs of Variance, Skewness and Kurtosis because those are the best attributes and I wanted to illustrate how nicely the projected hyperplane split the data even though the fourth attribute’s contribution is not included.

**Performance**

For performance I look at Positive Predictive Value(PPV), Negative Predictive Value (NPV), Sensitivity, Specificity, Accuracy and Confidence Intervals with predicted frequency given Class as the parameter. I will list these results for both Training and Testing. I take Class 0 (forged) banknotes to be the positive condition and Class 1 (genuine) to be the negative condition.

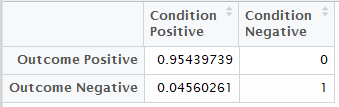
**Notation and Equation for diagnostics statistics**

; ;

;

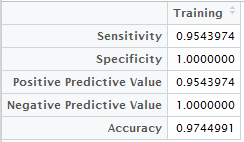
**Training Performance**

**Confusion Matrix**

****

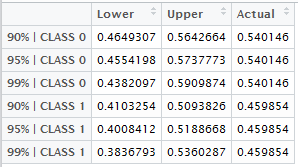
I am pleased to see that our true positive rate is 100%. This means that each genuine banknote is classified as such. But unfortunately, there is a small classification error associated with catching forfeiters. 4.56% of banknotes in Training were classified as genuine when in fact they were forged.

**Diagnostics**



The total accuracy of 97.50% is excellent. 100% NPV means that there are no genuine banknotes which were erroneously reported as forged.

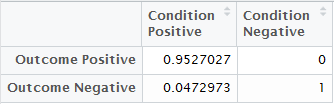
**Confidence Interval**



The above table shows three class specific confidence intervals for the class proportion in Training. For both classes, forged and genuine, we can say with high confidence that the *predicted* proportion will contain the *actual* proportion.

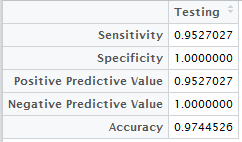
**Testing Performance**

**Confusion Matrix**



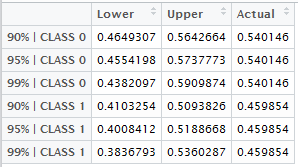
The above table illustrates that we have very good generality between Training and Testing. Testing also was able to deliver TN rate of 100%.

**Diagnostics**

****

Total accuracy in Testing of 97.45% is also excellent and very close to Training accuracy of 97.50%. 100% NPV means that there are no genuine banknotes which were erroneously reported as forged.

**Confidence Interval**



The above table shows three class specific confidence intervals for the class proportion in Training. For both classes, forged and genuine, we can say with high confidence that the *predicted* proportion will contain the *actual* proportion. Class 0 performed a tad worse when requiring 90% confidence. Overall, LDA does a very good job ensuring us that the actual proportion is contained with high confidence by the predicted proportions.

**Conclusion**

All in all, LDA performed very well as a class discriminator for this data set. The accuracy was high on both the training and testing sets while still maintaining near perfect generalization by delivering similar diagnostics. To improve the performance, Entropy WTI could potentially be dropped as a feature. It seems to offer very little contribution to discrimination. The 4D hyperplane projected onto the three best attributes was almost a perfect split; despite the noise added by Entropy WTI. It seems Entropy WTI contributed just that, entropy. Also, the tradeoff between higher specificity and lower sensitivity is preferred in this context. The ramifications of forging legal tender are life-changing to say the least. Erroneously classifying genuine notes as forged notes has a higher opportunity cost than allowing a small number of forged banknotes into our monetary system.

**Breast Cancer Wisconsin Data**

**Data Set Information**

This data set was taken from the archives of the UCI database which I name bcdf. They were donated in 1992 by physician Dr. William Wolberg. The data points came in chunks as Dr. Wolberg was able to report his clinical cases. I do not expect these discrete blocks of instances to be fundamentally different from each other. This is however implicitly assuming that Dr. Wolberg and his staff collected this information with similarly advanced equipment and methods. There are originally 699 data points, 458 which are classified as Class 2(benign) and 241 which are classified as Class 4(malignant). Each of the nine attributes were gathered range from 1-10. Mitosis has no values equal to 9. I highlight each variable in purple that is also used in my R scripts as variables or outputs. I used R studio version 0.99.902 to satisfy all of my programming needs. Each script has notes and comments explaining how the following analysis was employed. I use set.seed(20) to promote replicability of the outputs.

**ATTRIBUTES**

1. Clump Thickness (Thickness)

Thickness is measured to be the thickness of the tumor. Clump thickness in benign cells tend to be grouped in monolayers while malignant cells are often grouped in multilayers.

1. Uniformity of Cell Size (Uni.Size)

See below.

1. Uniformity of Cell Shape (Uni.Shape)

Benign cells do not have such wild variation in their size and shape. Cancerous cells on the other hand do, so these two attributes are very important in discriminating class.

1. Marginal Adhesion (Adhesion)

Adhesion is the tendency for the cells to stick together. Benign cells are better at sticking together than cancerous cells. This propensity for cells lose their stickiness is one reasons why cancerous cells spread more easily.

1. Single Epithelial Cell Size (Epithelial)

Epithelial cells are those which protect the body and act a shield. We have them on our skin, in our throat, and numerous other places where the body meets environment. In fact, carcinoma (a type of cancer) develops from epithelial cells. Epithelial cell size is closely related to Uni.Size and Uni.Shape. Epithelial cells which are particularly enlarged can be a good at predicting cancer.

1. Bare Nuclei (Nuclei)

This describes nuclei that are no longer surrounded by cytoplasm; the remainder of the cell. Benign tumors often have very few bare nuclei while cancerous cells will many.

1. Bland Chromatin (Chromatin)

Chromatin is a complex of DNA and other proteins that forms chromosomes within the nucleus. Bland chromatin is referring to the texture. In benign cells this texture tends to be smoother while in cancerous cells the texture is courser.

1. Normal Nucleoli (Nucleoli)

Nucleoli are small structures found inside the nucleus. Benign cells tend to have very small nucleoli as opposed to malignant cells. Cancerous cells tend to have more than then usual amount as well.

1. Mitosis (Mitosis)

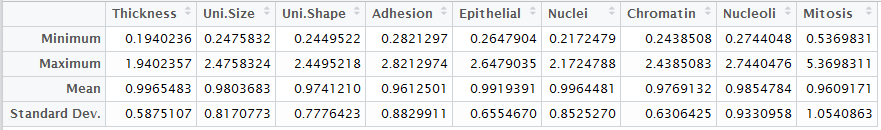
Mitosis is nuclear division plus cytokinesis, and produces two identical daughter cells during prophase, prometaphase, metaphase, anaphase, and telophase. The number of Mitosis which is taking place in a tumor mass can be a significant indicator. After all, cancer is seemingly unfettered cell division.

**CLASSES**

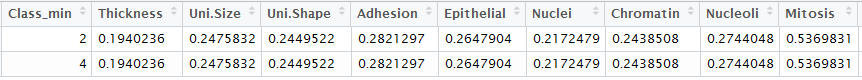
1. Patients with benign tumors are represented by Class = 2 with a total amount equal to.
2. Patients with malignant tumors are represented by Class = 4 with a total amount equal to.

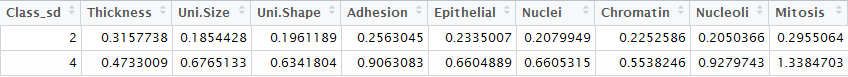
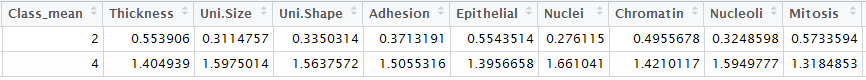
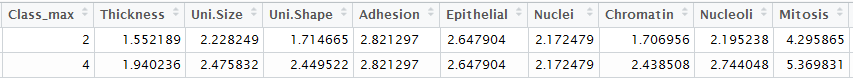
**PRECONDITIONING**

1. Before I perform LDA, I normalize bcdf by dividing each point by that column’s respective absolute mean. So that for all , each = , for all . We have nine attributes, so . This procedure will help the covariance matrix to not be ill-conditioned.
2. bcdf does not have an equal split between the two classes. Benign has 444 cases and malignant has 239 cases. I am going to artificially increase the number of malignant cases to 478 but doubling each object with Class 4. The total number of objects in bcdf is now 922.
3. There are 16 data points with missing values. I chose to flat out remove the objects with the NA values. I believe this is the best route since most of the NA valued rows were Class 2 of which there are far more instances. The alternative is to replace the missing values with some statistic. NOTE: all of the missing values were from attribute 6, Epithelial. Now bcdf is ready to be split into a training and a testing data frame. I use ‘createDataPartition’. This ensures that there is good class balance between Training and Testing. I chose to allocate 70% of the data points to Training and the remaining 30% to Testing. There are 646 objects in Training, of which 310(47.99%) are Class 2 and 336(52.01%) are Class 4. There are 276 objects in Testing, of which 134(48.55%) are Class 2 and 142(51.45%) are Class 4.
4. Next, I calculate basic statistics for Training as a whole and Training given Class:
   1. Training Statistics:



* 1. Training given Class statistics:

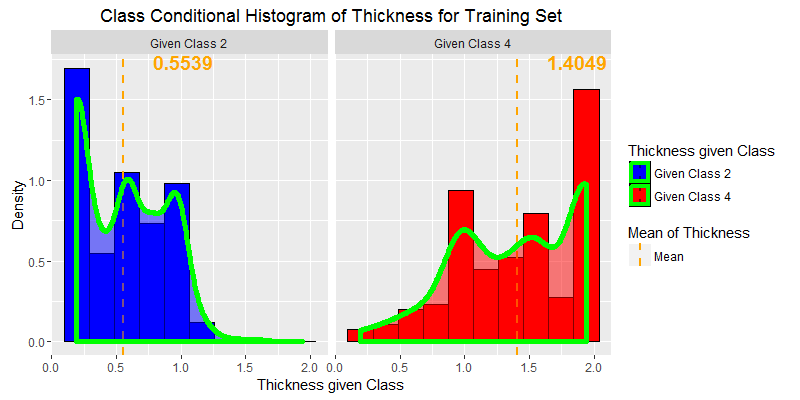


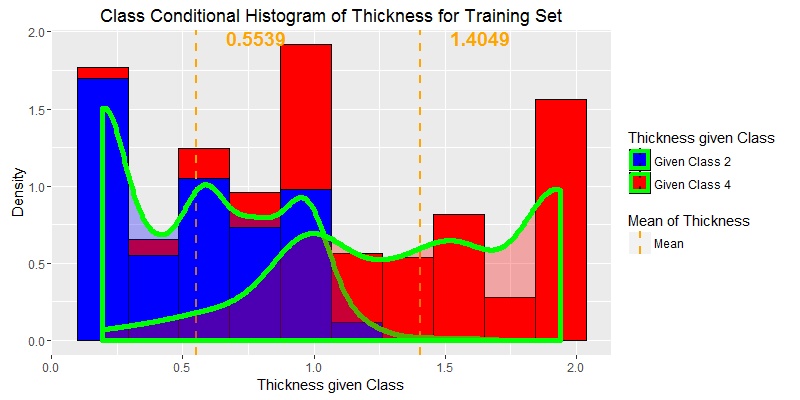


We can see from the above tables that both classes take the lowest possible value for each attribute. Even benign cells take many of the maximum values. But, we see how nicely separated the class means are. In many cases they are separated by 2 standard deviations. I suspect that this will allow LDA to discriminate benign from cancerous nicely.

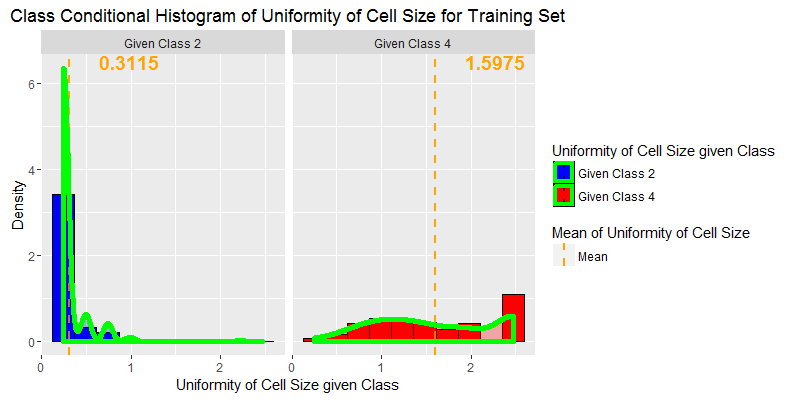
**Histogram Analysis for Training**

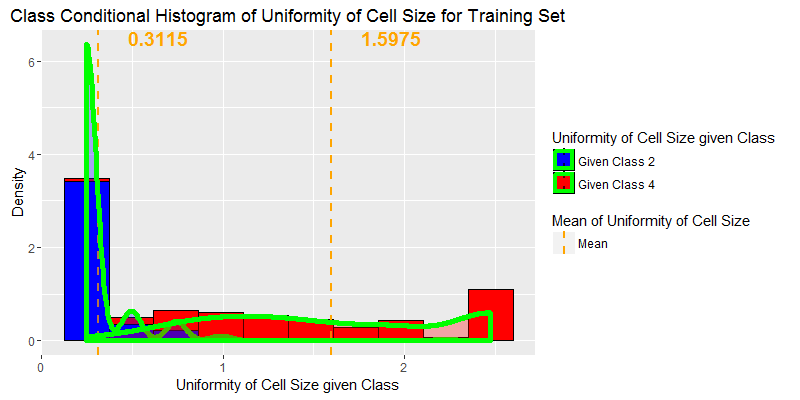
Now that I have useful statistics about my data and it’s organized in a convenient way, I can plot histograms to see how the class conditional statistics compare within each attribute. I use **library**(ggplot2) since it provides the most flexibility and has the best visual appeal. I partition Training into temporary data frames out\_two and out\_four then plot the histograms of each attribute given Class. I plot the p.d.f. curves in green on top of the histograms to better illustrate similarities and differences amongst the distributions. I use an orange dashed line to represent the mean of the distribution. The standard deviation of the distribution can be referred to in the table. The standard deviations were close enough in value that vertical lines convoluted the graph. I decided to include faceted graphs of each attribute in addition to overlaid graphs of each attribute. Even though the scales are the same on the faceted graph, I found the overlaid graph convenient to compare the class distribution within each attribute. Objects classified as Class = 2(benign) are **blue** and objects classified as Class = 4(malignant) are **red**.



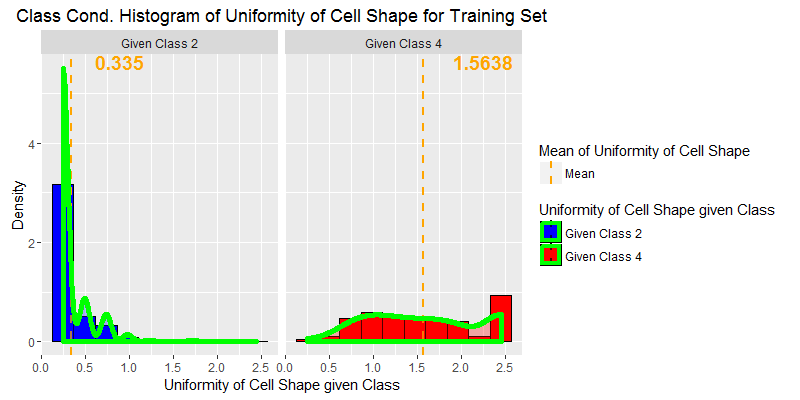


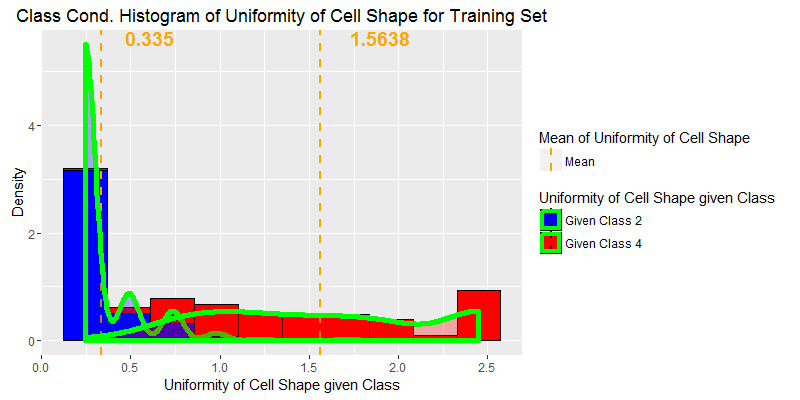
As expected, Thickness has good separation between the classes. Even with good separation in means, there is more overlap between the classes than in other attributes. Class 2 s.d. = 0.3158 and Class 4 s.d. = 0.4733.



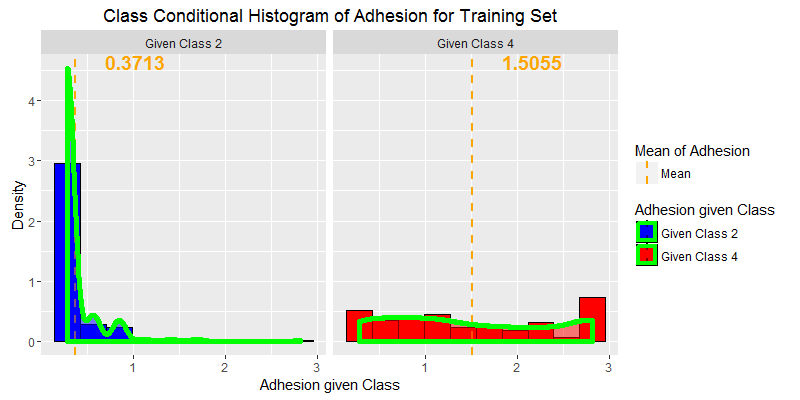


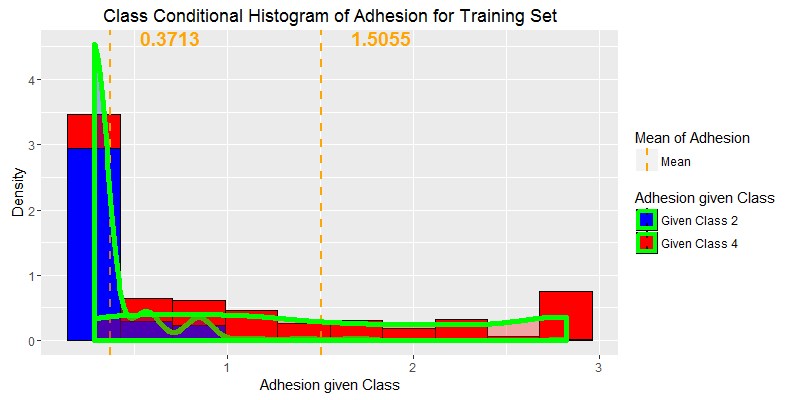
We can see that benign cells tend to have much lower uniformity in their cell size. Malignant cells have more of a uniform distribution but with a slight favor to being very large.  
Class 2 s.d. = 0.1854 and Class 4 s.d. = 0.6765.



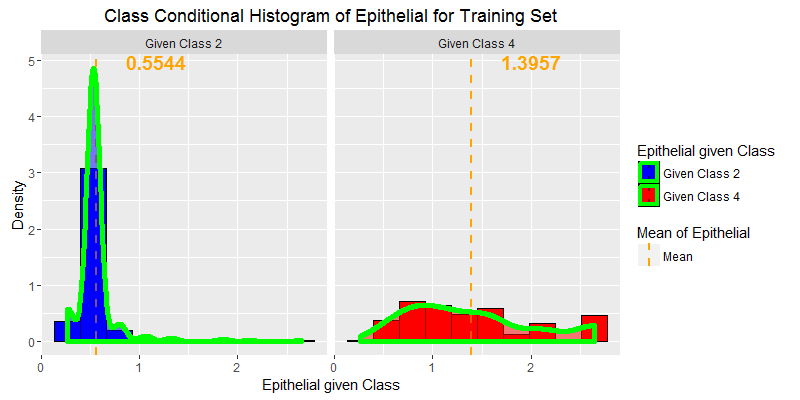


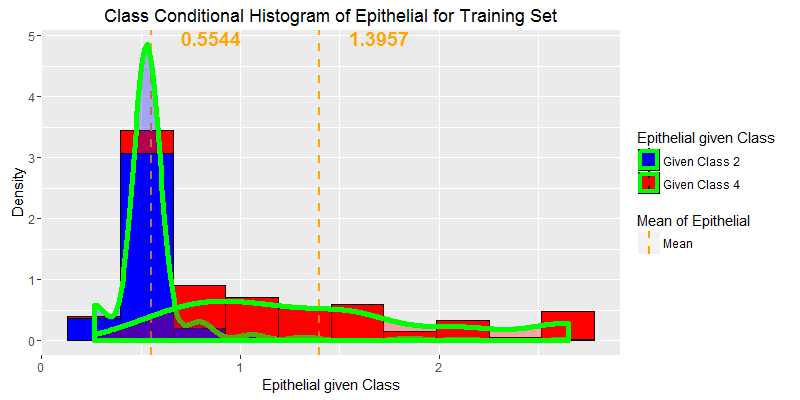
These distributions are VERY similar to Uni.Size. I expect that they go hand in hand. I removed Uni.Shape from my LDA calculation but the accuracy improvement was a wash for multiple random seeds. I also tried combining Uni.Size and Uni.Shape by taking their average and there was no real increase in generalization.   
Class 2 s.d. = 0.1961 and Class 4 s.d. = 0.6342.



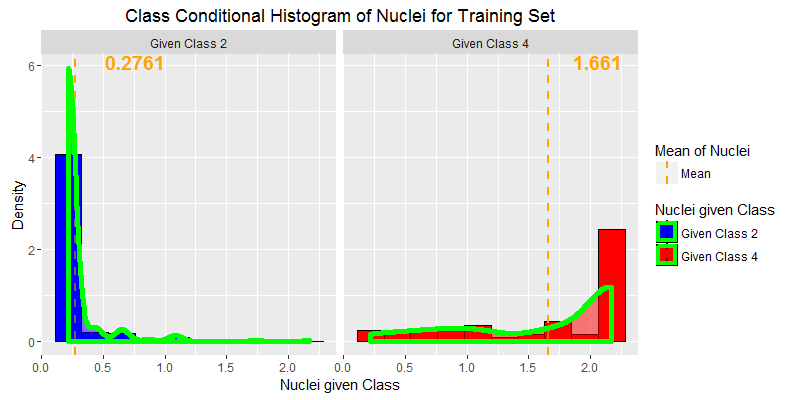


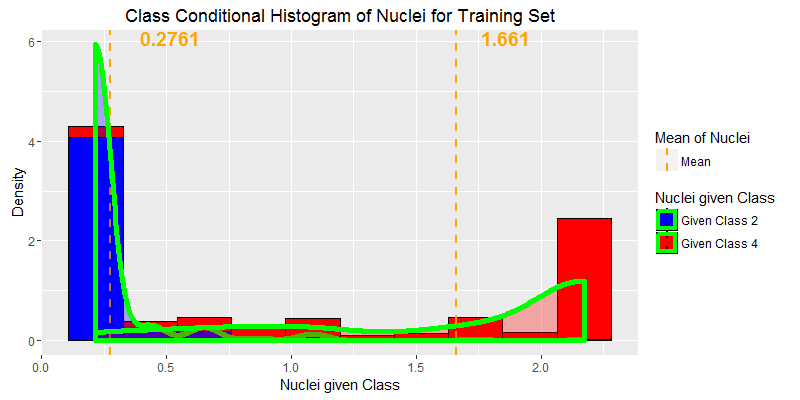
We can see that Adhesion is similar to Uni.Size and Uni.Shape in that benign cases are skewed right and malignant cases are uniformly distributed. Still, it is telling that there are virtually no benign cases that take high values for Adhesion.  
Class 2 s.d. = 0.2563 and Class 4 s.d. = 0.9063.



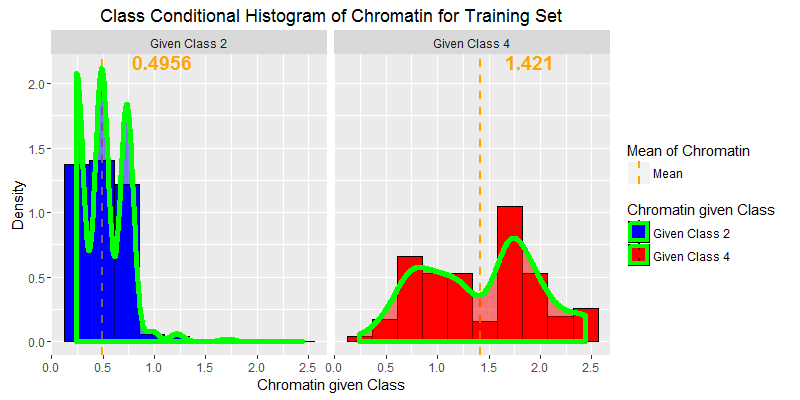


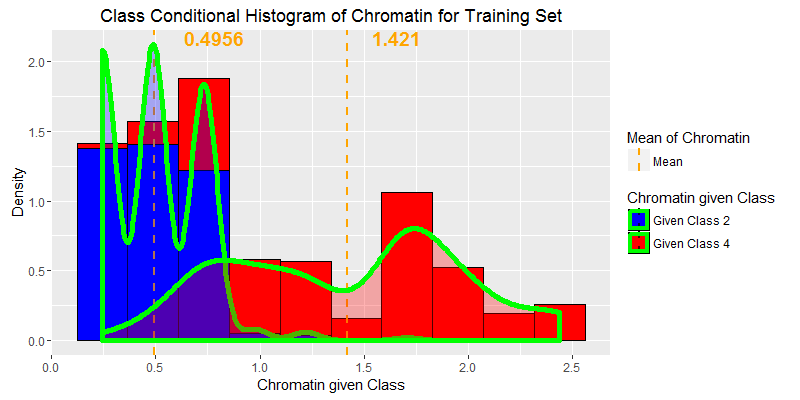
Benign cells tend to have smaller values of Epithelial but the means and distributions are not quite as separated as the above attributes. Class 2 s.d. = 0.2335 and Class 4 s.d. = 0.6605.



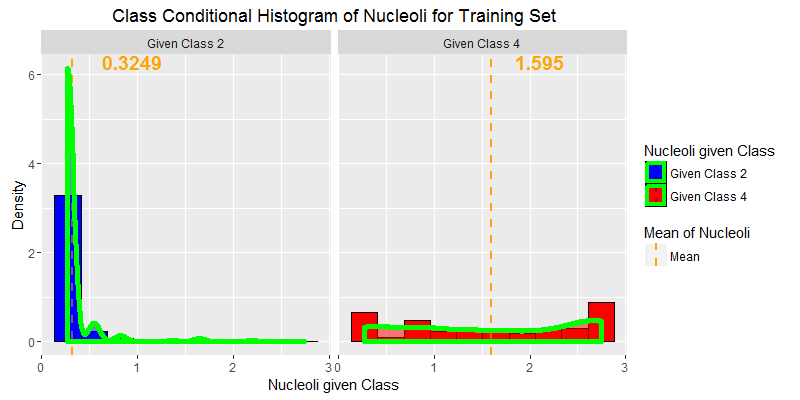


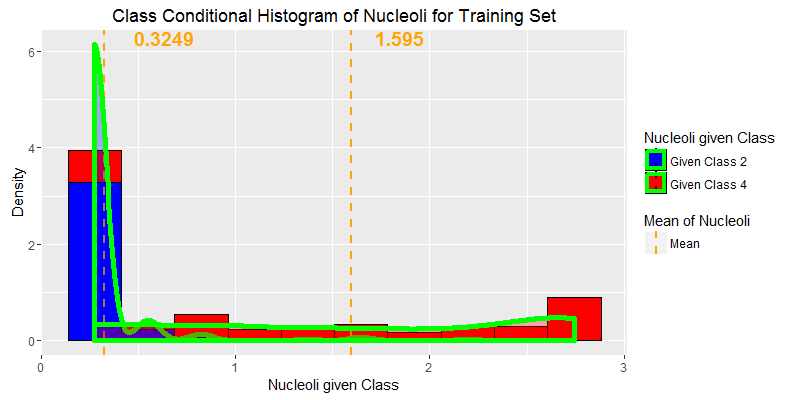
The class means are very well separated. Even though Class 4 takes values over the entire range, I expect Nuclei to be very important to classification. Class 2 s.d. = 0.2080 and Class 4 s.d. = 0.6605.



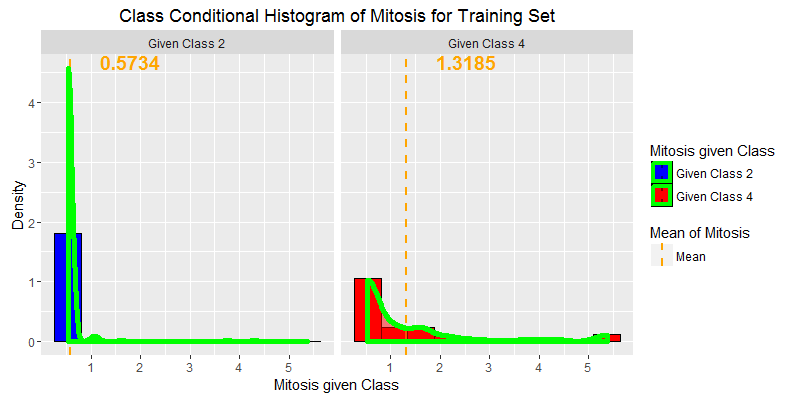


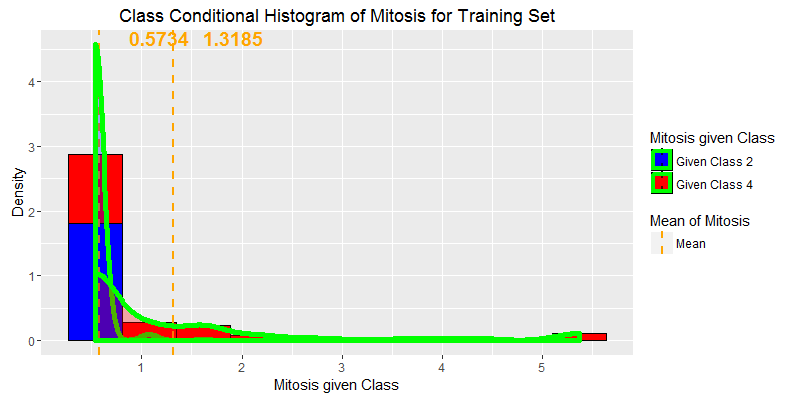
Chromatin yields fairly poor classification compared to the previous attributes. Benign cells tend to have smoother chromatin but malignant cells are spread over all the possible values save the lowest. Class 2 s.d. = 0.2253 and Class 4 s.d. = 0.5538.





Nucleoli appears to be the weaker at separating the classes. The class means are well separated but Class 4’s mean is equal to its median. In most other attributes malignant cells are shifted more to the right. Benign cases are heavily shifted to the lowest value, which will help in discrimination. 305 (98.39%) of benign cases take the three lowest values. Class 2 s.d. = 0.2050 and Class 4 s.d. = 0.9280.





At first glance Mitosis does not seem to be contributing much but closer inspection shows that in benign cells, 301(97.10%) instances take the lowest value and 310 (99.35%) taking the lowest three values.  
Malignant cells are still heavily skewed left with many cases taking the lowest value. Class 2 s.d. = 0.2955 and Class 4 s.d. = 1.338.

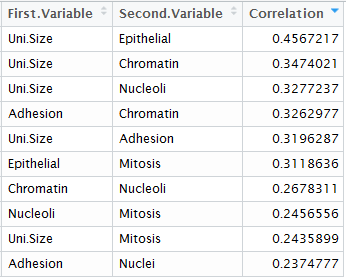
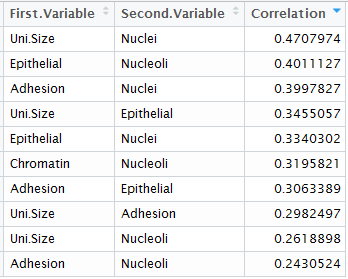
**Test-Statistic**

ttest

We can see that many factors have similar t-test scores. Mitosis is by far the weakest, but still it is significant.

**Correlation Table Given Class**

**Class = 2 ---------------------- Class = 4**

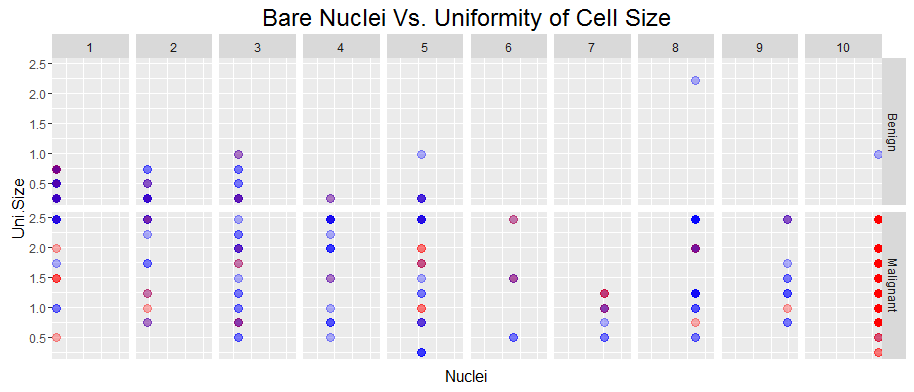


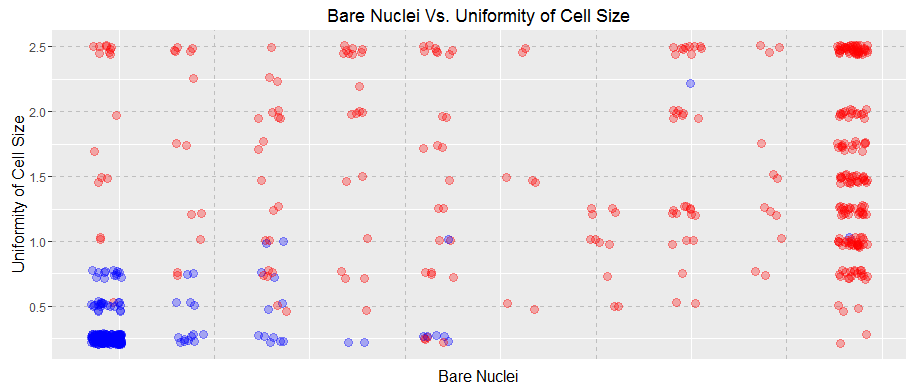
The above tables list the highest correlations between the attributes given Class. I have omitted Uni.Shape because the distribution is so similar to Uni.Size.

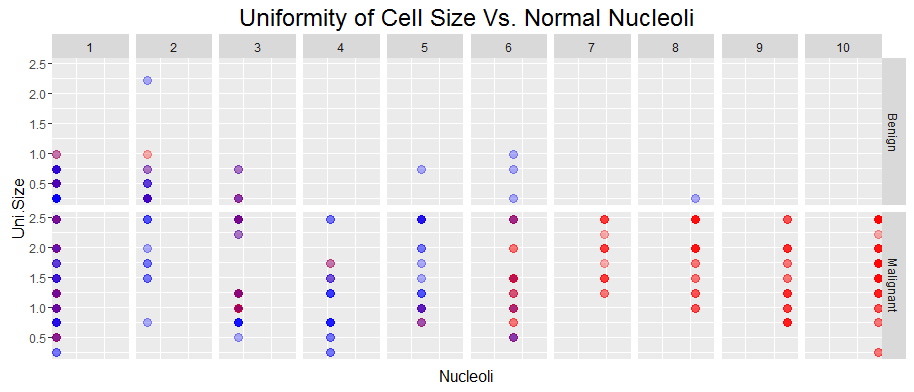
Best Attributes: Given the distributions for the histograms, the separation of means between classes, t-test statistics, and the correlation table I have chosen the following four attributes as the best suited for discriminating between benign and malignant: Uni.Size, Nuclei, Nucleoli, and Epithelial. I will now represent these attributes with 2D and 3D graphs.

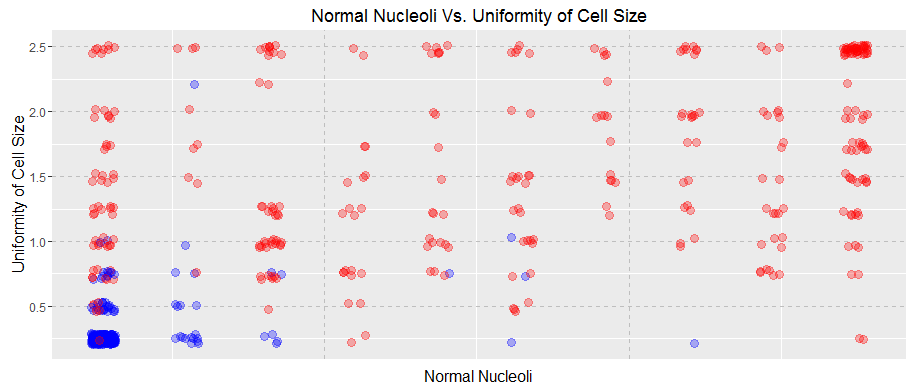
**2D Graphing of Attributes**

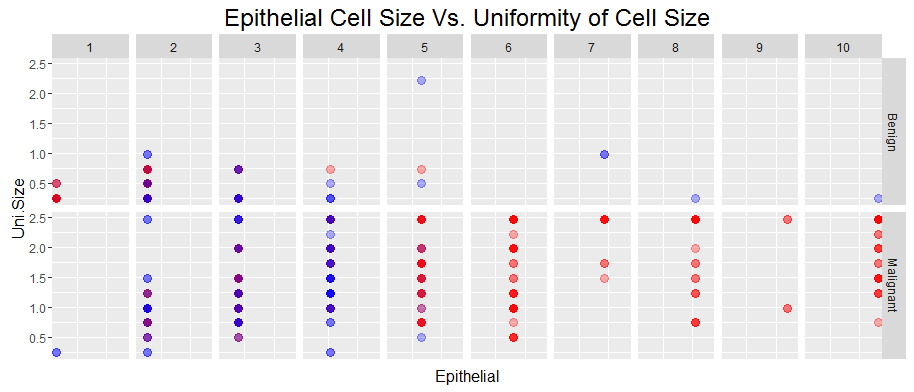
Objects classified as Class 2(benign) are **blue** and objects classified as Class 4(malignant) are **red**. Since the values for each attribute can only take 10 discrete values, I have chosen to jitter the points in the overlaid graph. Jitter with minor transparency allows the trends to be seen more quickly.

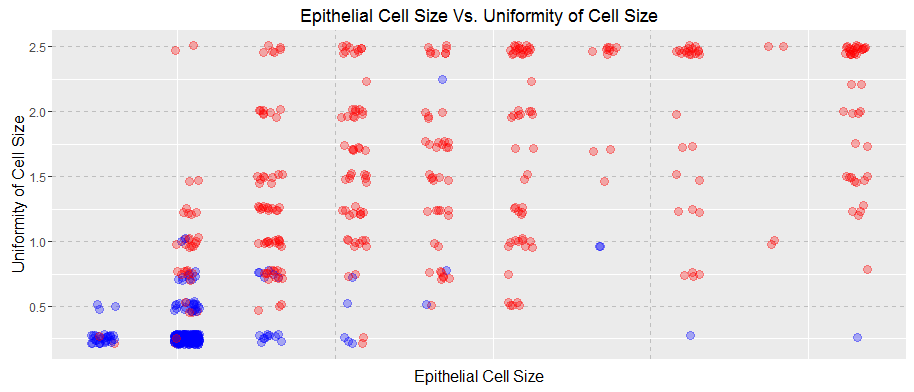


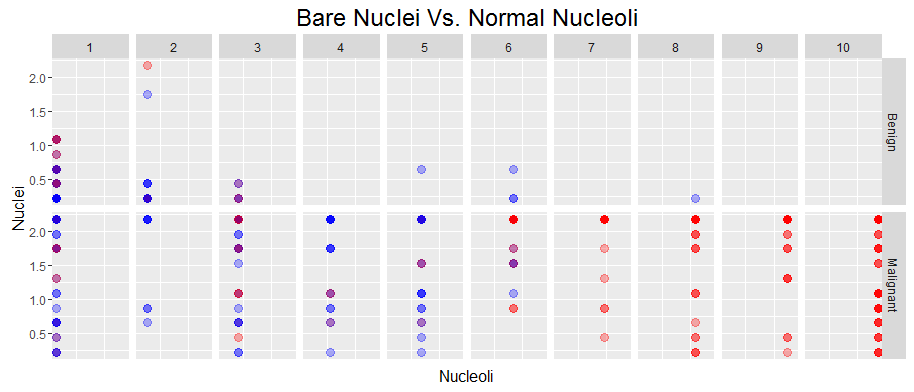


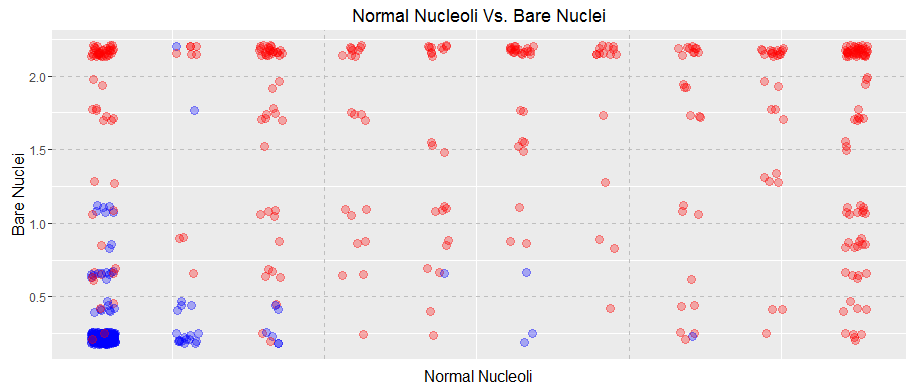


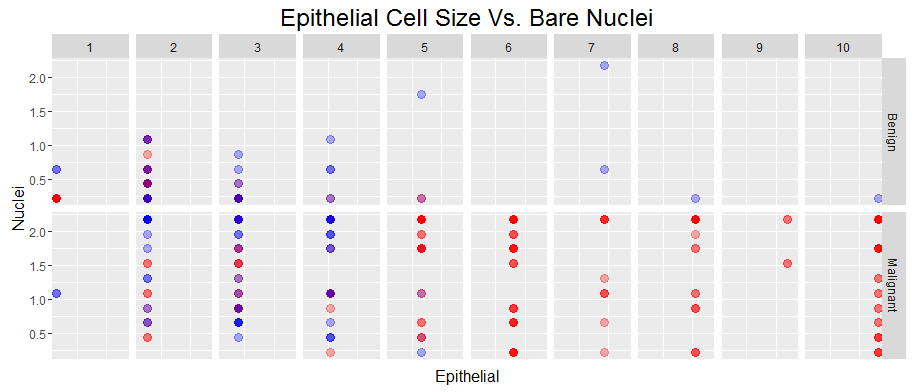


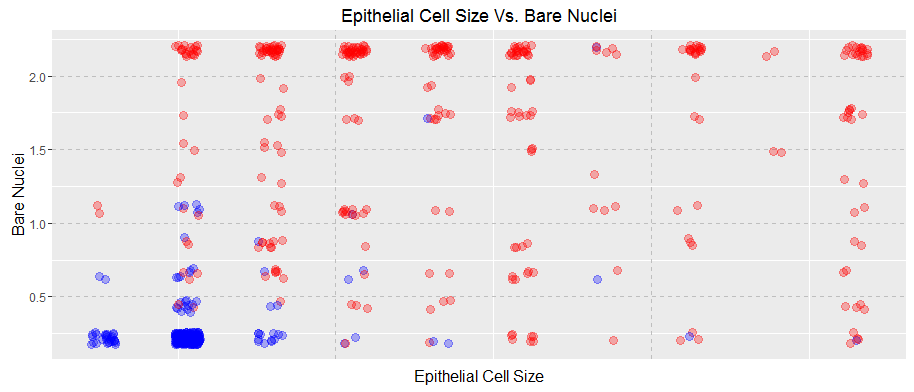


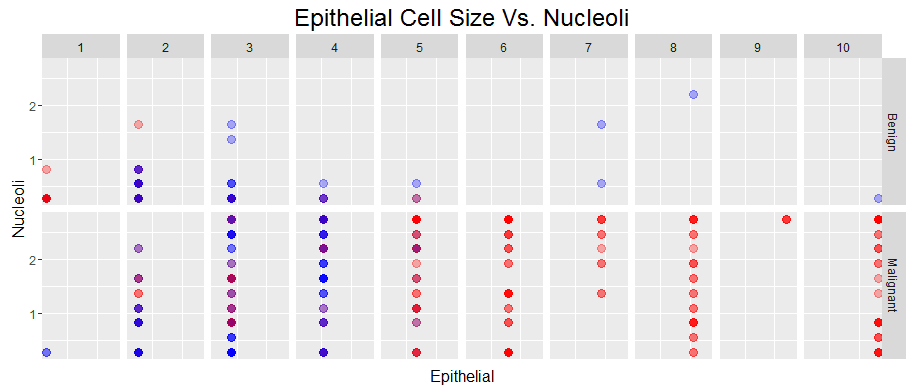


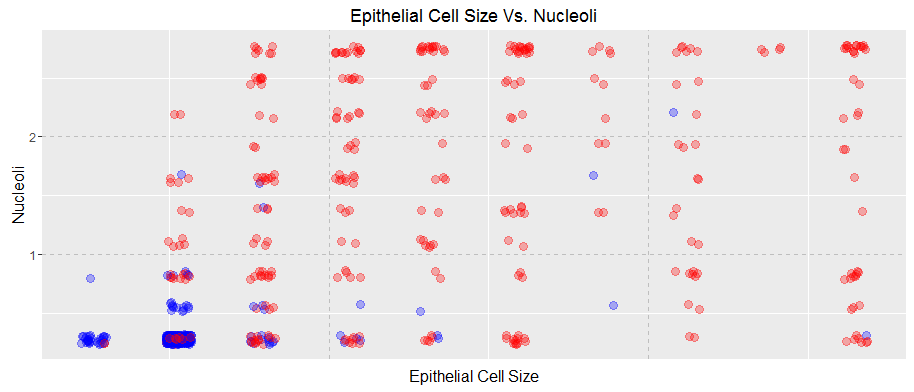








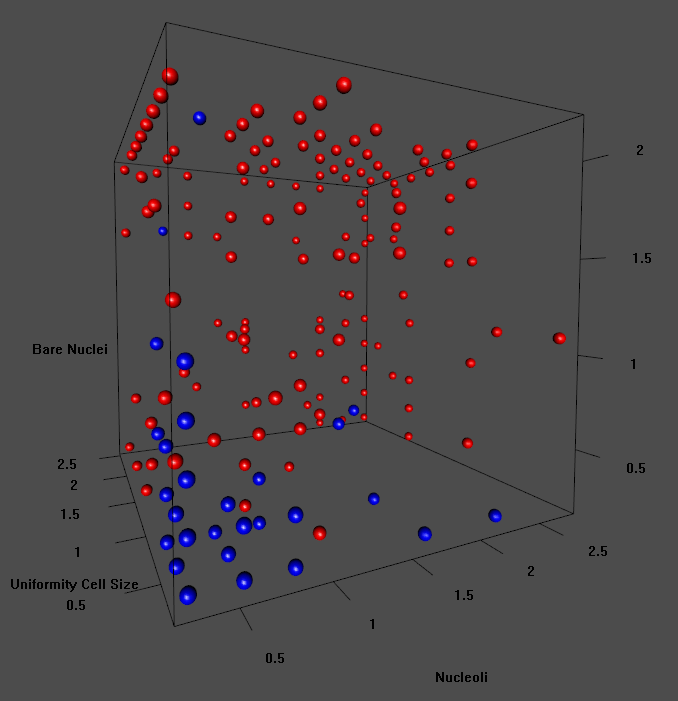




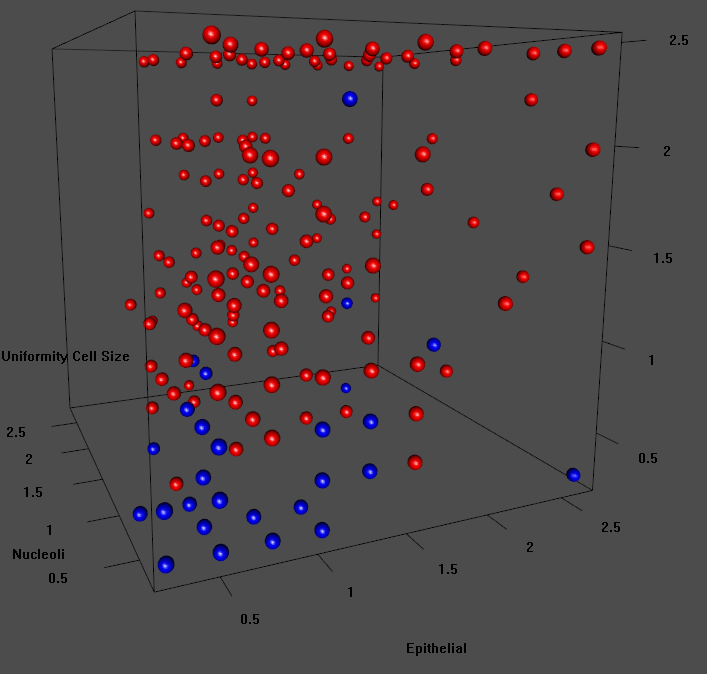
All of the above graphs confirm what our histograms have already shown us. Namely, that as values for the attributes increase, the likelihood that the subject has cancer drastically increases. This is especially true for the four attributes I chose to visualize.

**3D Graphing of Attributes**

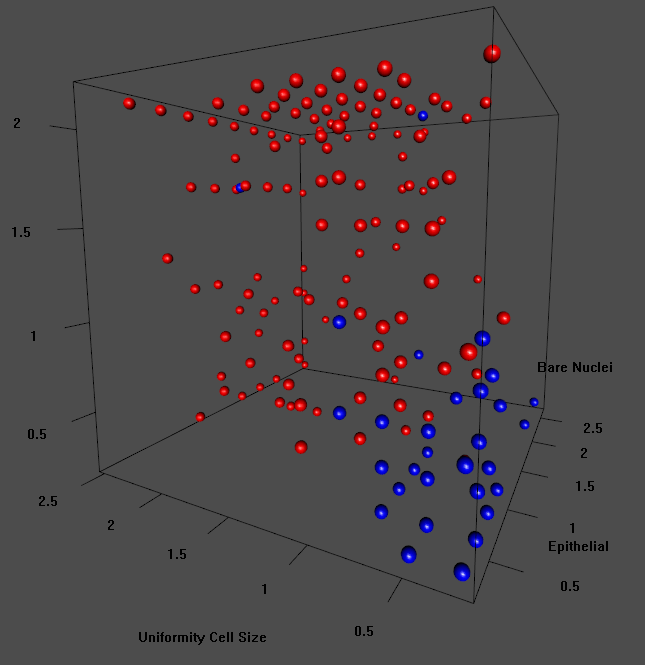
**Uni.Size vs. Nuclei vs. Nucleoli**

****

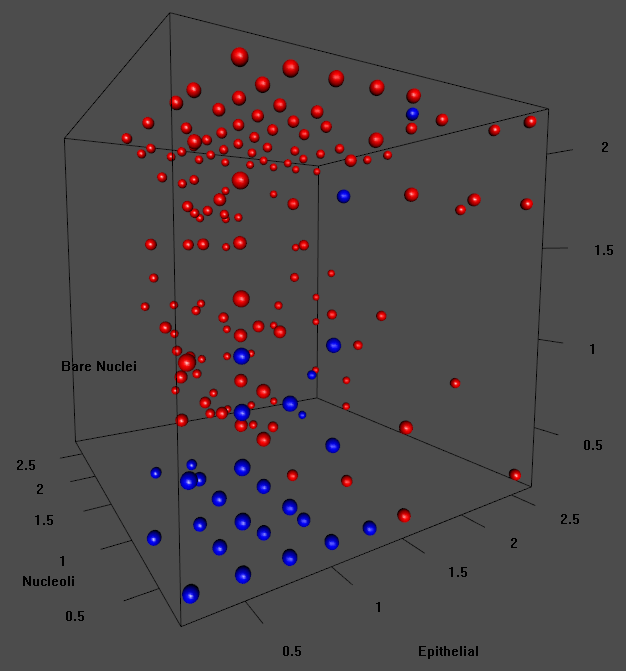
**Uni.Size vs. Nucleoli vs. Epithelial**

****

**Uni.Size vs. Nuclei vs. Epithelial**

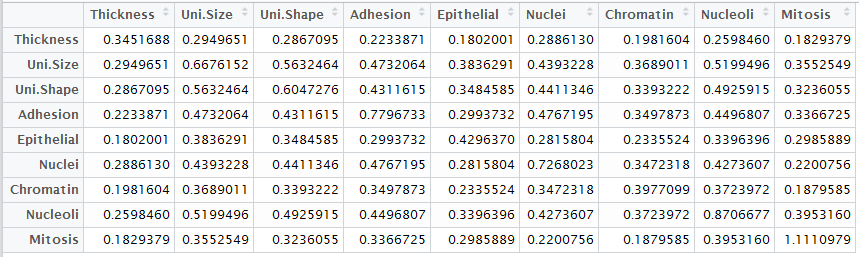


**Nuclei vs. Nucleoli vs. Epithelial**



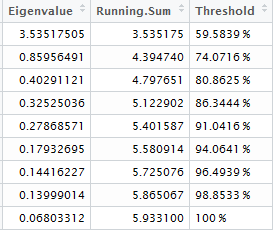
We can see from the above 3D graphs that they conform in pattern  
 to the corresponding 2D graphs.

**Covariance Matrix**

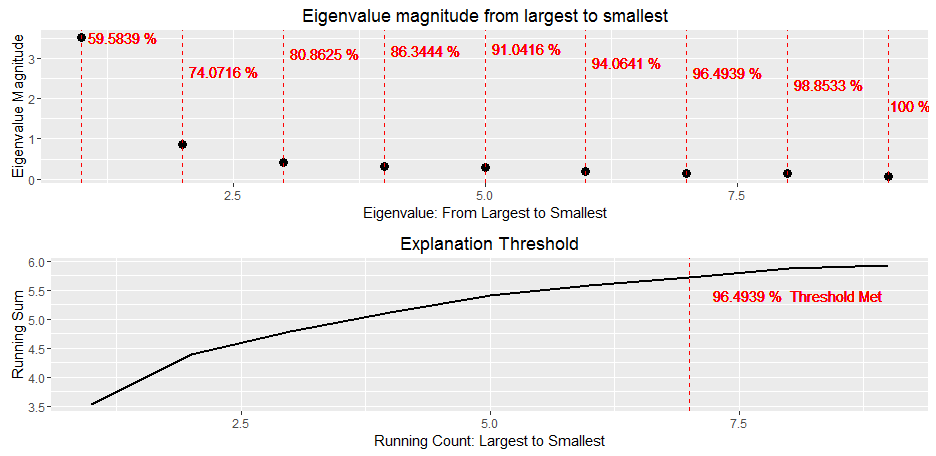


The covariance Matrix is well defined, there should be no issue manipulating this for our learning algorithm.

**Eigenvalues & Running Sum & Threshold  
for Covariance Matrix**



The following eigenvalue plots help to illustrate the percentage of the data explained by the eigenvectors of the covariance matrix.



We can see that 96.49% of the information is contained in the first seven eigenvectors. The remaining 3.51% of information contained in the last vector is proportional to the error of the residue if we were to use these results in Principal Component Analysis. However, the dimensions of our data set, 922x9, is small enough not to warrant dimensionality reduction.

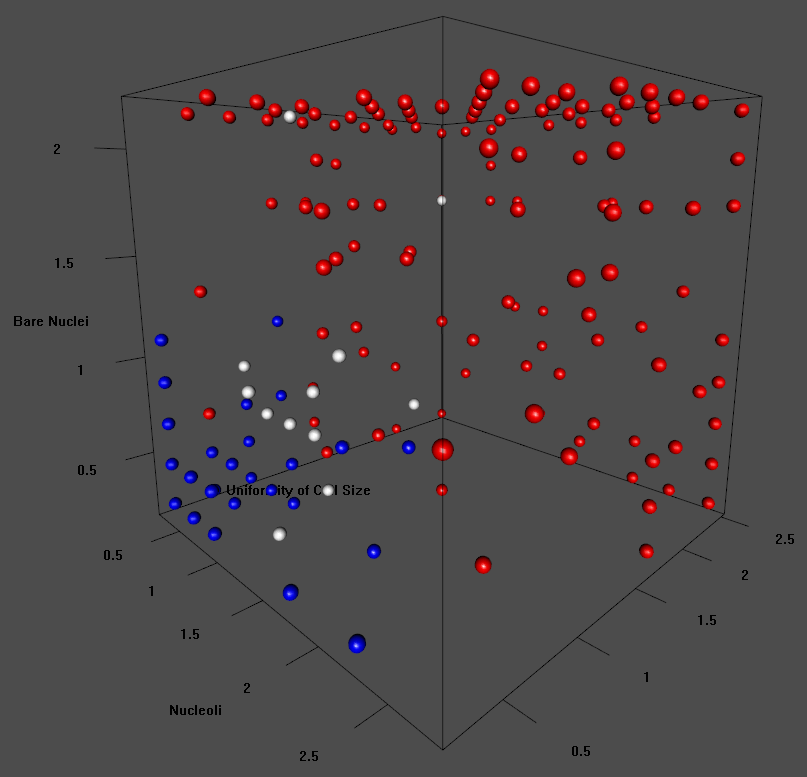
**Implementing LDA and Graphical Results**

LDA is implemented in R using the {MASS} package. I save the predicted values into Training and Testing under the column named Pred. I assigned as prior probabilities the class priors from Training. I have given a brief overview of LDA above in the banknotes report.

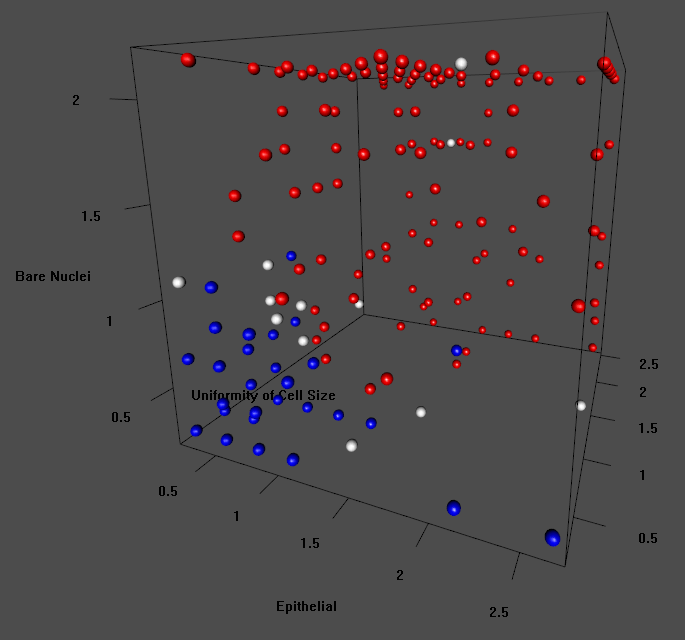
**3D Graphs with Misclassifications**

First, let us take a look at the graphs in 3D. I chose to render only the four best attributes (Uni.Size, Nuclei, Nucleoli, Epithelial). The white points are misclassifications.

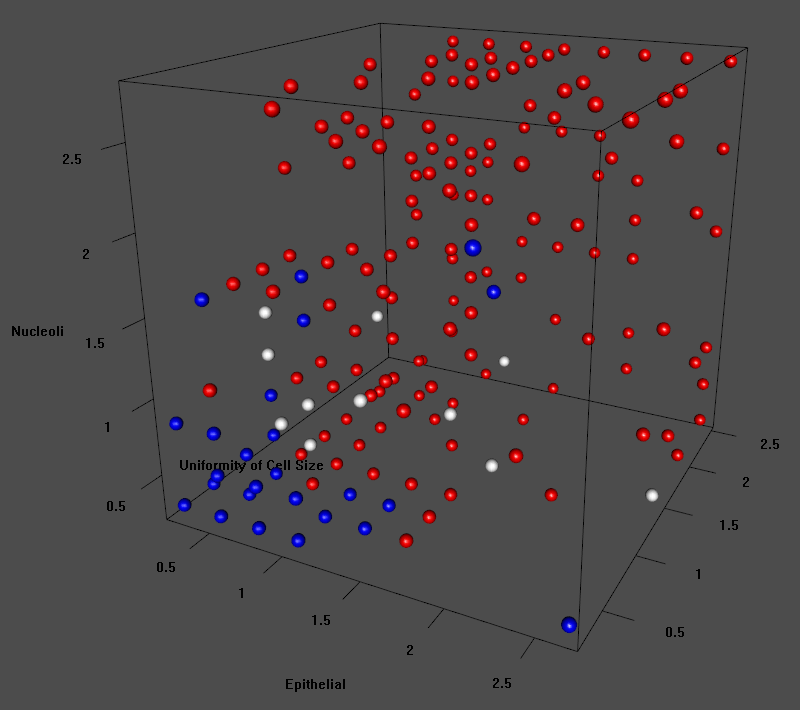
**Uni.Size vs Nuclei vs Nucleoli**



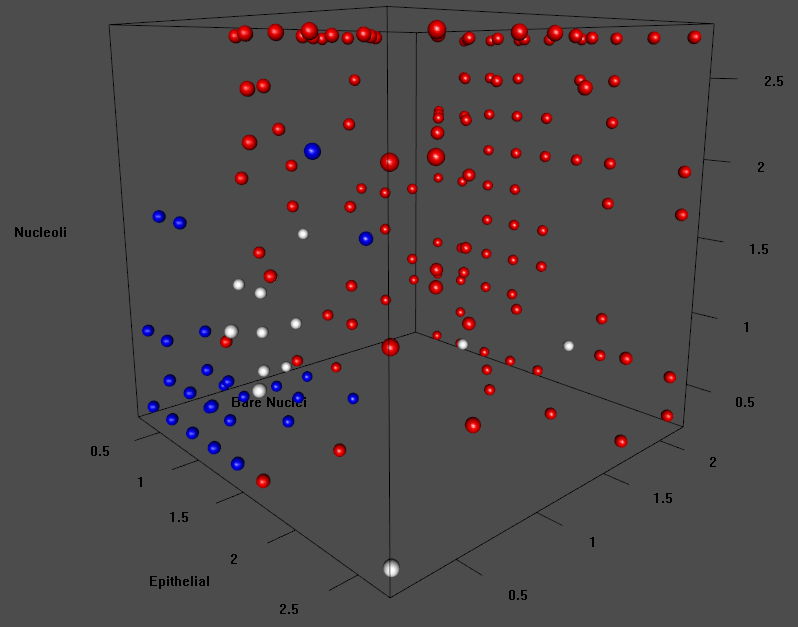
**Uni.Size vs Nuclei vs Epithelial**

****

**Uni. Size vs Nucleoli vs Epithelial**

****

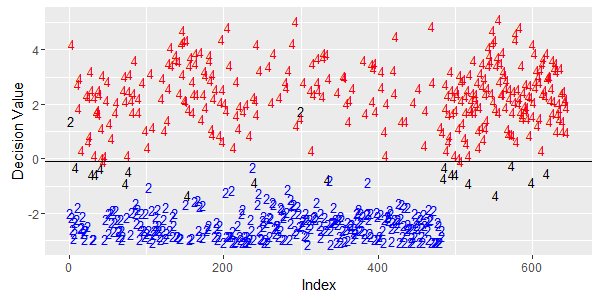
**Nuclei vs Nucleoli vs Epithelial**

****

**Performance**

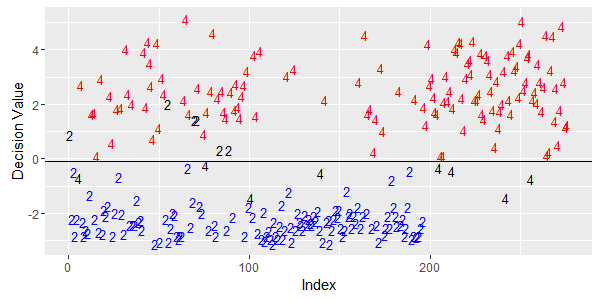
For performance I look at Positive Predictive Value(PPV), Negative Predictive Value (NPV), Sensitivity, Specificity, Accuracy and Confidence Intervals with predicted frequency given Class as the parameter. I will list these results for both Training and Testing. I am associating a positive condition as having a benign tumor and a negative condition as having a malignant tumor.

**Incorrect Prediction in Training**



20 misclassifications: 2 false negatives and 18 false positives

**Incorrect Prediction in Testing**



14 misclassifications: 6 false negatives and 8 false positives

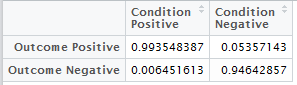
**Notation and Equation for diagnostics statistics**

; ;

;

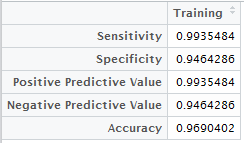
**Training Performance**

**Confusion Matrix**

****

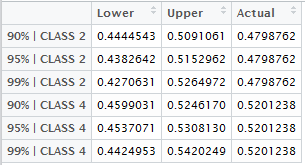
We can see from the above confusion matrix that the number of true positives is extremely high while the number of true negatives is very high. The following table shows the diagnostics in more detail.

**Diagnostics**



In the above table, we observe that 96.90% accuracy is very high. Since we are dealing with cancer diagnosis, it is crucial that our accuracy be as absolutely high as possible. There is a bit of an imbalance between sensitivity and specificity. This algorithm gives excellent results when predicting if a patient actually has a benign tumor. However, the specificity is lower which will result in more patients thinking that they have cancer when they don’t. In my opinion, this is preferable. I believe that most people would rather be told that they have cancer and actually don’t than be told that they don’t and in reality do.

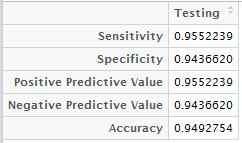
**Confidence Intervals**

****

The above table shows three class specific confidence intervals for the class proportion in Training. We can see that the actual proportion of both Class 2 and Class 4 fall inside the confidence intervals. The 90% CI for Class 4 barely contains the actual proportion. This observation is in line with the fact that specificity is lower than sensitivity.

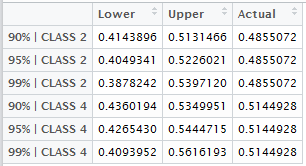
**Testing Performance**

**Diagnostics**

****

We can see that with an accuracy of 94.93% on Testing that LDA was a good generalizer for this data set. The specificity is lower than sensitivity like the training set but this time we see that they are more balanced.

**Confidence Intervals**

****

We see from the above table that the actual class proportion falls within all of the confidence intervals. This tells us that LDA yields satisfactory results with respect to predicting accurate class proportions.

**Conclusion**

All in all, LDA performed very well as a class discriminator for this data set. The accuracy was high on both the training and testing sets while still maintaining good generalization by delivering similar diagnostics. To improve the performance, one could try PCA dimensionality reduction. Some of the attributes have fairly strong correlation and thus could benefit from PCA or other forms of feature selection. The data also appears to be clustered on either end of the value spectrum, 1-10. This implies that clustering algorithms may prove to be successful.