# Leafs

*Kaisa Roggeveen, Scott Graham*

*March 22nd 2018*

Dr. Steven M. Vamosi Associate Dean, Diversity, Equity and Inclusion Professor, Population Biology 2500 University Drive NW Department of Biological Sciences University of Calgary Calgary AB T2N 1N4 Canada

## Introduction

The intent of this paper is to develop a method for classifying leaves as either Cherry or Pear, based on their measured length and width. This method was developed for Dr. Steven Vamosi, a botanist from the Universtiy of Calgary.

The classification method used was Linear discriminant analysis, developed by R.A Fischer. To .... To take training samples from a sampled population take measurements, from these measurements classification rules are created. This will then be tested against the classifying sample to see if there are any miss classifications.

Cherry and Pear leaves are both leaves from fruit trees. Cherry trees belong to the genus Prunus and Pear trees belong to the genus Pyrus [2],[3]. A common feature amongst the leaves is that they both have a midrib, which is the centreal vein of the leaf which extends along the leaf's centerline.

## Data

### Measurement Process

The first step taken in the measurement of the leaves was to give each leaf an identification number based on the species. The method used to meausre the dimensions was to create a box with the minimum length and width in which the entire leaf would be encompassed in the box.

To begin creating the sides of the box, a ruler was aligned parallel to the midrib, which is the central vein in the leaf and moved towards the left and the right of the picture until only one point on the leaf remained [1]. From the single point on the side of the leaf, a line was drawn parallel to the midrib of the leaf.

Next, the base and point of the leaf were measured, a ruler was placed perpendicular to the midrib and the ruler was moved towards to tip of the leaf until a single point remained, a line was draw perpendicular to the midrib at this point. At the base of the leaves the length of the leaf was set as the point where the leaf ends and the stem begins, at this point a line was drawn perpendicular to the midrib.

After all the boxes were created, the width (lines parallel to midrib) and the length (lines perpendicular to midrib) were measured and the results were recorded in a spread sheet.

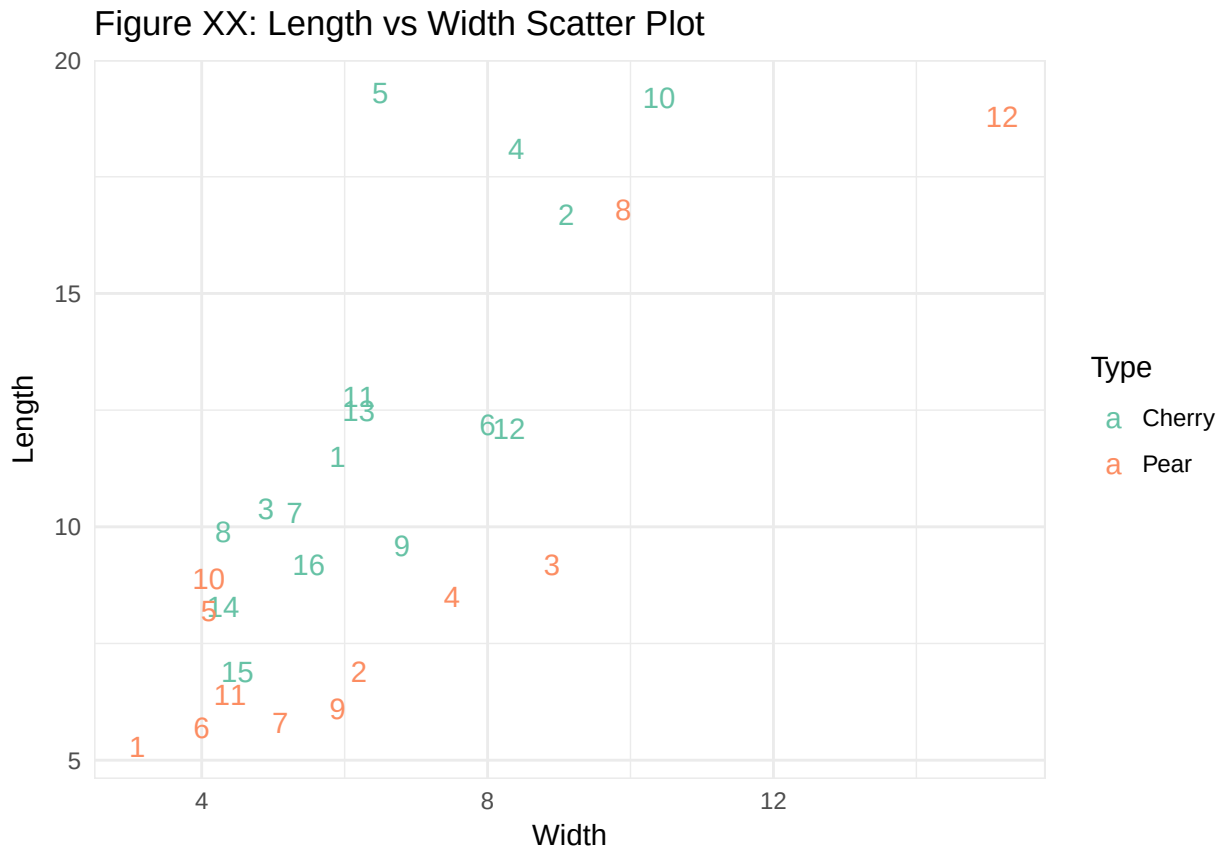
### Data Creation

| Type | Length | Width |
|---|---|---|
| Cherry | 11.5 | 5.9 |
| Cherry | 16.7 | 9.1 |
| Cherry | 10.4 | 4.9 |
| Cherry | 18.1 | 8.4 |

| Type | Length | Width |
|---|---|---|
| Cherry | 19.3 | 6.5 |
| Cherry | 12.2 | 8.0 |
| Cherry | 10.3 | 5.3 |
| Cherry | 9.9 | 4.3 |
| Cherry | 9.6 | 6.8 |
| Cherry | 19.2 | 10.4 |
| Cherry | 12.8 | 6.2 |
| Cherry | 12.1 | 8.3 |
| Cherry | 12.5 | 6.2 |
| Cherry | 8.3 | 4.3 |
| Cherry | 6.9 | 4.5 |
| Cherry | 9.2 | 5.5 |
| Pear | 5.3 | 3.1 |
| Pear | 6.9 | 6.2 |
| Pear | 9.2 | 8.9 |
| Pear | 8.5 | 7.5 |
| Pear | 8.2 | 4.1 |
| Pear | 5.7 | 4.0 |
| Pear | 5.8 | 5.1 |
| Pear | 16.8 | 9.9 |
| Pear | 6.1 | 5.9 |
| Pear | 8.9 | 4.1 |
| Pear | 6.4 | 4.4 |
| Pear | 18.8 | 15.2 |

| Type | Length | Width |
|---|---|---|
| Cherry:16 | Min. : 5.300 | Min. : 3.100 |
| Pear :12 | 1st Qu.: 7.875 | 1st Qu.: 4.475 |
| NA | Median : 9.750 | Median : 6.050 |
| NA | Mean :10.914 | Mean : 6.536 |
| NA | 3rd Qu.:12.575 | 3rd Qu.: 8.075 |
| NA | Max. :19.300 | Max. :15.200 |

In this orginal dataset there are a few issues that need to be ackowledged. The first issues that occured during the data measurements was the result of the leaves that were distributed as the training sample were images, in which the images were not to scale. This resulted in a few outliers, which much larger lengths and widths compared to the other leaves in the set. These outliers included Pear#12, Cherry#10 and Cherry#5. However, based on the nature of this project in just observing the ratio between the length and width, this should not be affected by the size of the image, unless the image was streched in either direction.




## Classification Procedure (LDA)

### Training Data

```
## Call:
## lda(Type ~ Length + Width, data = ., cv = TRUE)
##
## Prior probabilities of groups:
##    Cherry      Pear
## 0.5714286 0.4285714
##
## Group means:
##           Length    Width
## Cherry 12.437500 6.537500
## Pear    8.883333 6.533333
##
## Coefficients of linear discriminants:
##               LD1
## Length -0.4193533
## Width   0.5310692
```

Table 3: blarg

| Predicted | Actual | Length | Width | Cherry Probability | Pear Probability | Correct Prediction |
|---|---|---|---|---|---|---|
| Cherry | Cherry | 11.5 | 5.9 | 0.7882 | 0.2118 | TRUE |
| Cherry | Cherry | 16.7 | 9.1 | 0.8839 | 0.1161 | TRUE |
| Cherry | Cherry | 10.4 | 4.9 | 0.8050 | 0.1950 | TRUE |
| Cherry | Cherry | 18.1 | 8.4 | 0.9694 | 0.0306 | TRUE |
| Cherry | Cherry | 19.3 | 6.5 | 0.9967 | 0.0033 | TRUE |
| Cherry | Cherry | 12.2 | 8.0 | 0.5227 | 0.4773 | TRUE |
| Cherry | Cherry | 10.3 | 5.3 | 0.7387 | 0.2613 | TRUE |
| Cherry | Cherry | 9.9 | 4.3 | 0.8292 | 0.1708 | TRUE |
| Pear | Cherry | 9.6 | 6.8 | 0.3582 | 0.6418 | FALSE |
| Cherry | Cherry | 19.2 | 10.4 | 0.9284 | 0.0716 | TRUE |
| Cherry | Cherry | 12.8 | 6.2 | 0.8685 | 0.1315 | TRUE |
| Pear | Cherry | 12.1 | 8.3 | 0.4481 | 0.5519 | FALSE |
| Cherry | Cherry | 12.5 | 6.2 | 0.8456 | 0.1544 | TRUE |
| Cherry | Cherry | 8.3 | 4.3 | 0.6414 | 0.3586 | TRUE |
| Pear | Cherry | 6.9 | 4.5 | 0.3892 | 0.6108 | FALSE |
| Cherry | Cherry | 9.2 | 5.5 | 0.5485 | 0.4515 | TRUE |
| Pear | Pear | 5.3 | 3.1 | 0.4152 | 0.5848 | TRUE |
| Pear | Pear | 6.9 | 6.2 | 0.1426 | 0.8574 | TRUE |
| Pear | Pear | 9.2 | 8.9 | 0.0764 | 0.9236 | TRUE |
| Pear | Pear | 8.5 | 7.5 | 0.1391 | 0.8609 | TRUE |
| Cherry | Pear | 8.2 | 4.1 | 0.6631 | 0.3369 | FALSE |
| Pear | Pear | 5.7 | 4.0 | 0.3091 | 0.6909 | TRUE |
| Pear | Pear | 5.8 | 5.1 | 0.1664 | 0.8336 | TRUE |
| Cherry | Pear | 16.8 | 9.9 | 0.8116 | 0.1884 | FALSE |
| Pear | Pear | 6.1 | 5.9 | 0.1134 | 0.8866 | TRUE |
| Cherry | Pear | 8.9 | 4.1 | 0.7528 | 0.2472 | FALSE |
| Pear | Pear | 6.4 | 4.4 | 0.3355 | 0.6645 | TRUE |
| Pear | Pear | 18.8 | 15.2 | 0.1853 | 0.8147 | TRUE |

## New Data

```
## $class
## [1] Cherry Pear   Cherry
## Levels: Cherry Pear
##
## $posterior
##      Cherry      Pear
## 1 0.8003229 0.1996771
## 2 0.2772400 0.7227600
## 3 0.5942515 0.4057485
##
## $x
##           LD1
## 1 -0.63325061
## 2  0.94345071
## 3  0.04321672
```

## Observation Space

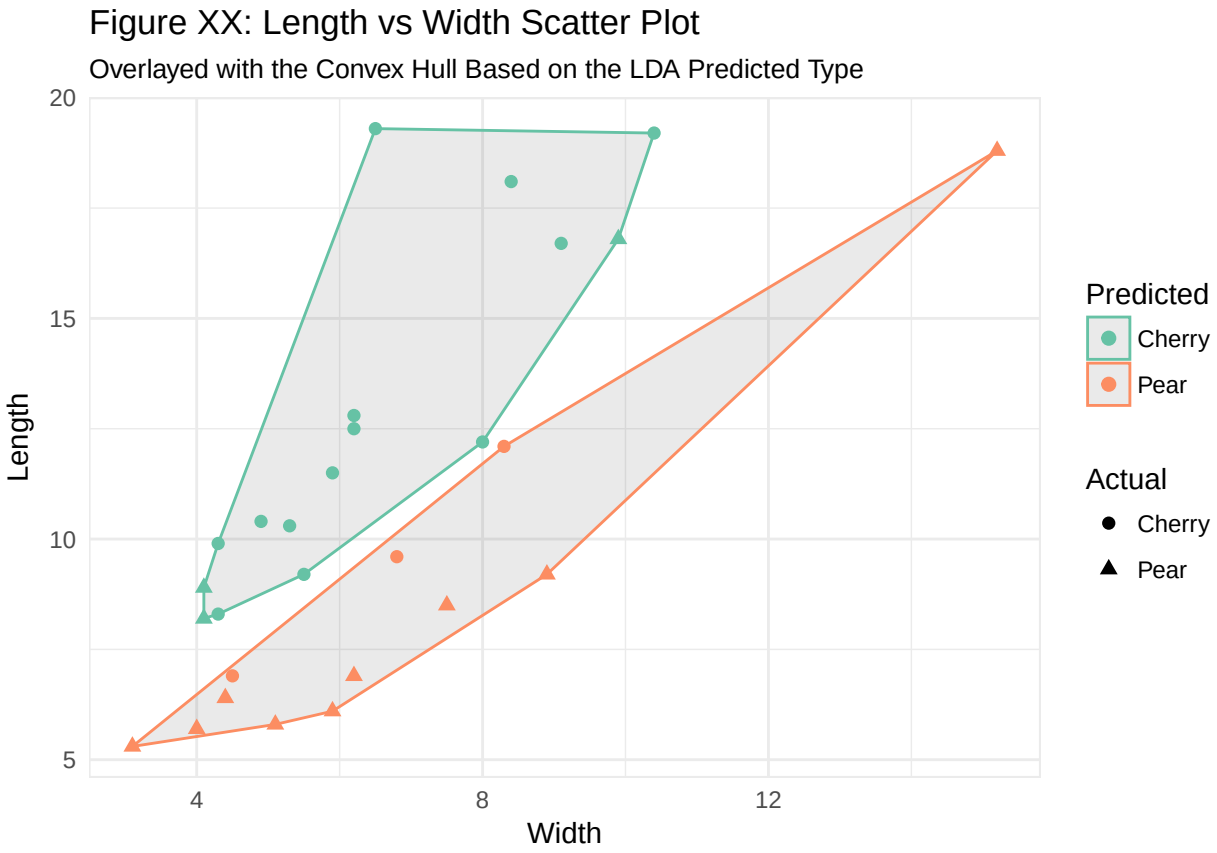
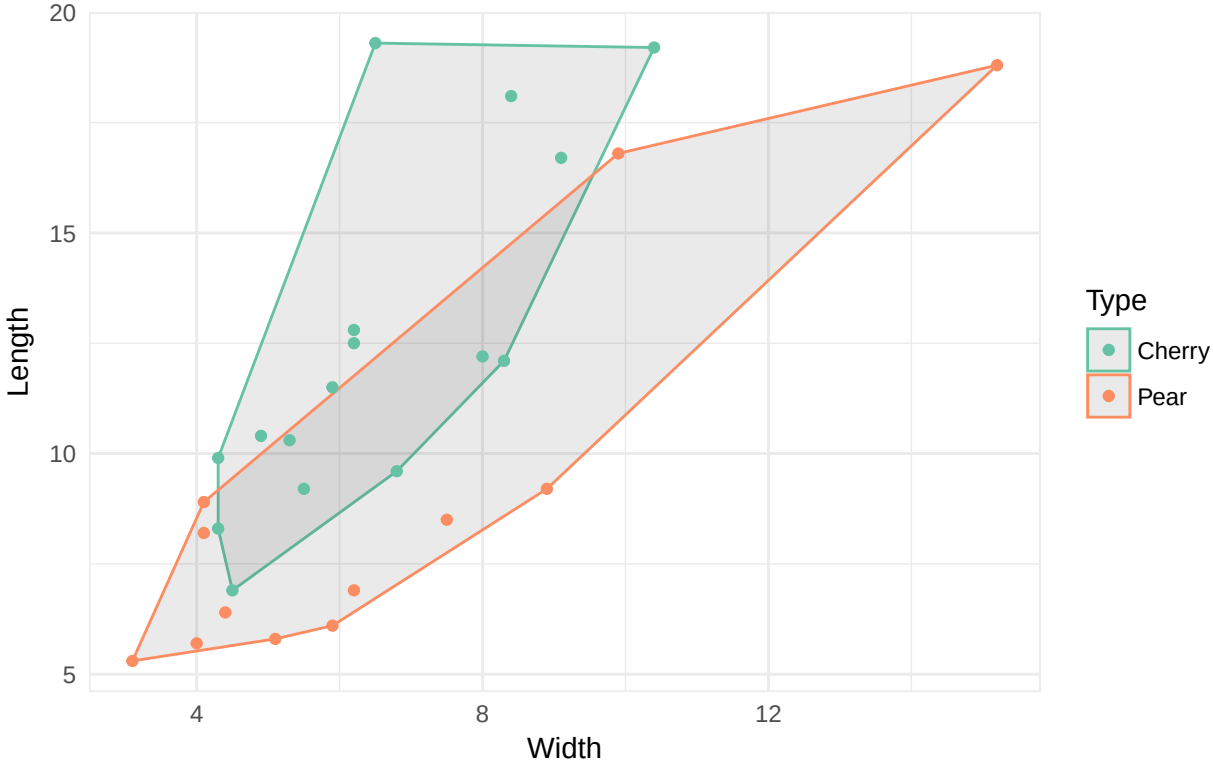


Figure XX: Length vs Width Scatter Plot

Overlayed with the Convex Hull of that Type



# Probability Distributions

## Contour

Figure XX: Length vs Width Scatter Plot

Overlayed with the Contour Plot

Figure XX: Length vs Width Scatter Plot

Overlayed with a Contour Plot of that Type

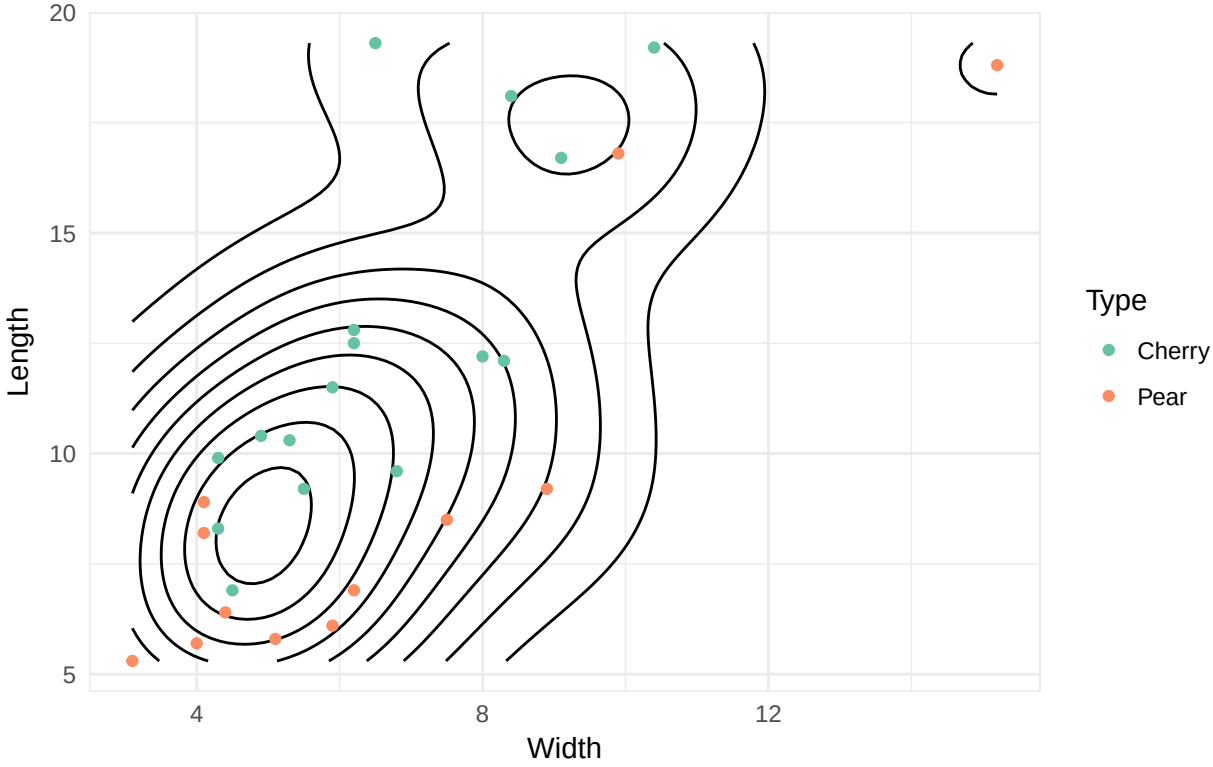


Figure XX: Density Plot by Type

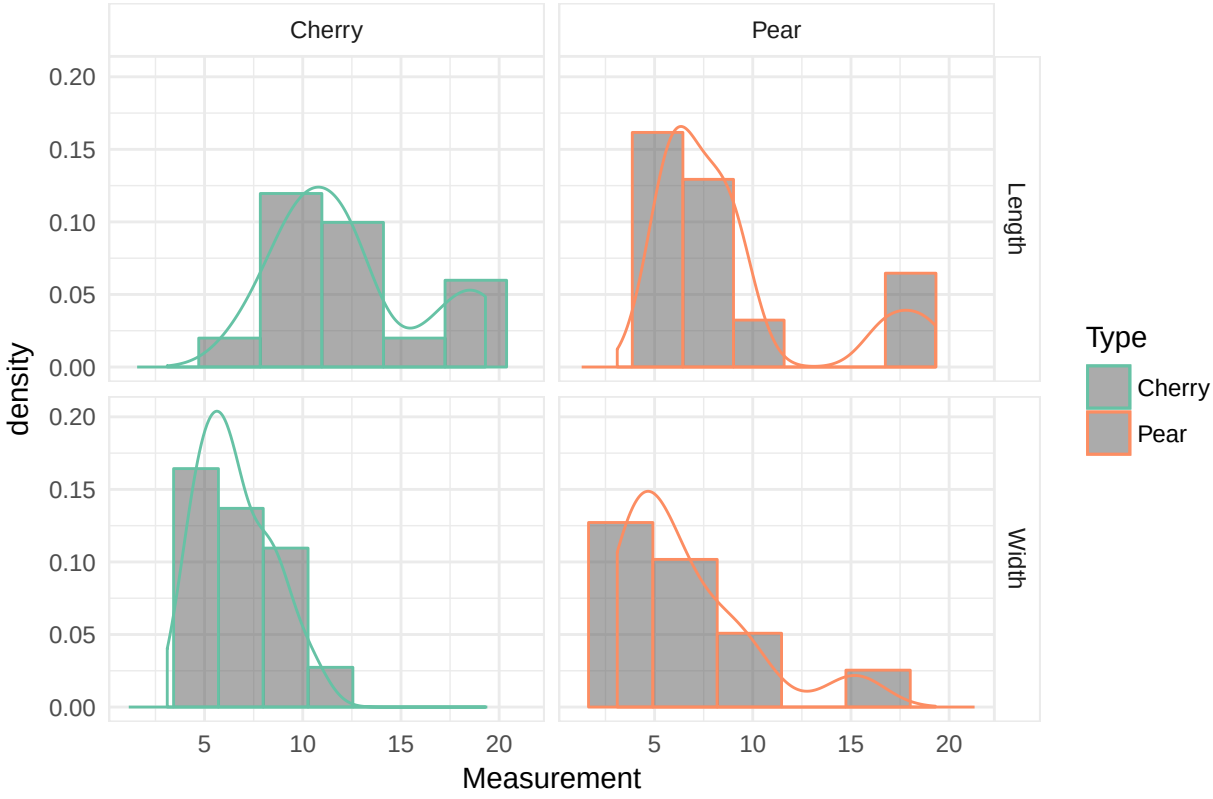
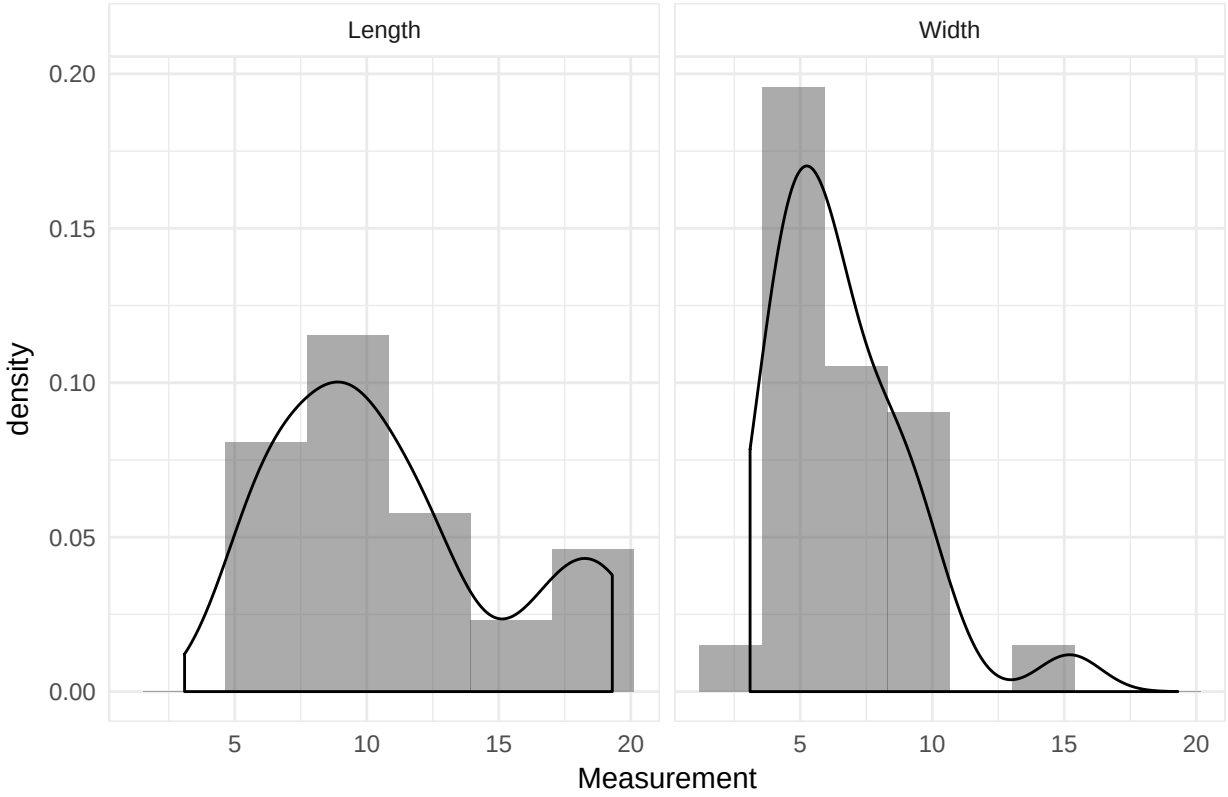


Figure XX: Density Plot




## Covariance Matrix

Table 4: Shared Covariance Matrix

| | Length | Width |
|---|---|---|
| Length | 19.422011 | 8.476508 |
| Width | 8.476508 | 6.685344 |

Table 5: Cherry Covariance Matrix

| | Length | Width |
|---|---|---|
| Length | 15.047833 | 5.443833 |
| Width | 5.443833 | 3.357167 |

Table 6: Pear Covariance Matrix

| | Length | Width |
|---|---|---|
| Length | 19.27788 | 13.37333 |
| Width | 13.37333 | 11.83152 |

# Classification Procedure (QDA)

## Training Data

```
## Call:
## qda(Type ~ Length + Width, data = ., cv = TRUE)
##
## Prior probabilities of groups:
##    Cherry      Pear
## 0.5714286 0.4285714
##
## Group means:
##           Length    Width
## Cherry 12.437500 6.537500
## Pear    8.883333 6.533333
```

Table 7: blarg

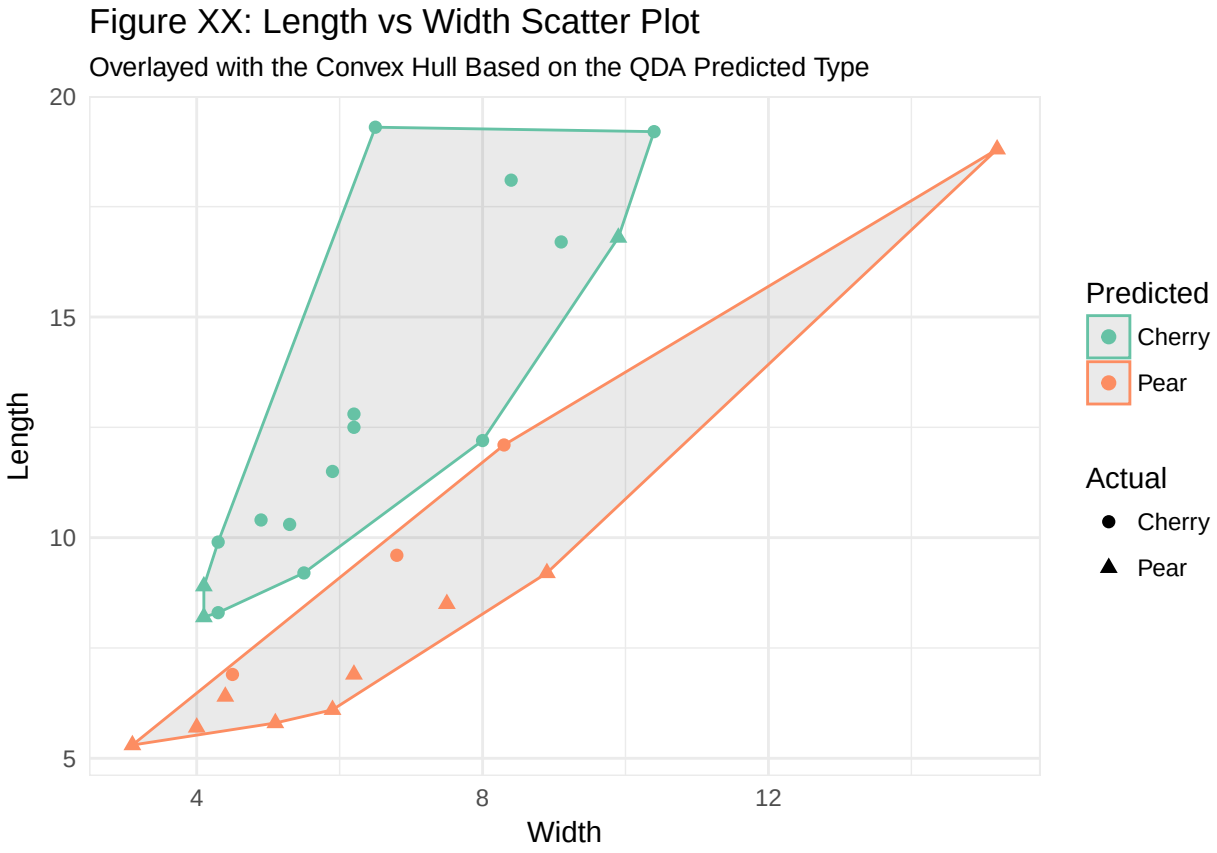
| Predicted | Actual | Length | Width | Cherry Probability | Pear Probability | Correct Prediction |
|---|---|---|---|---|---|---|
| Cherry | Cherry | 11.5 | 5.9 | 0.8815 | 0.1185 | TRUE |
| Cherry | Cherry | 16.7 | 9.1 | 0.9488 | 0.0512 | TRUE |
| Cherry | Cherry | 10.4 | 4.9 | 0.8529 | 0.1471 | TRUE |
| Cherry | Cherry | 18.1 | 8.4 | 0.9971 | 0.0029 | TRUE |
| Cherry | Cherry | 19.3 | 6.5 | 1.0000 | 0.0000 | TRUE |
| Cherry | Cherry | 12.2 | 8.0 | 0.5677 | 0.4323 | TRUE |
| Cherry | Cherry | 10.3 | 5.3 | 0.8177 | 0.1823 | TRUE |
| Cherry | Cherry | 9.9 | 4.3 | 0.8310 | 0.1690 | TRUE |
| Pear | Cherry | 9.6 | 6.8 | 0.4686 | 0.5314 | FALSE |
| Cherry | Cherry | 19.2 | 10.4 | 0.9662 | 0.0338 | TRUE |
| Cherry | Cherry | 12.8 | 6.2 | 0.9454 | 0.0546 | TRUE |
| Pear | Cherry | 12.1 | 8.3 | 0.4363 | 0.5637 | FALSE |
| Cherry | Cherry | 12.5 | 6.2 | 0.9302 | 0.0698 | TRUE |
| Cherry | Cherry | 8.3 | 4.3 | 0.6487 | 0.3513 | TRUE |
| Pear | Cherry | 6.9 | 4.5 | 0.4710 | 0.5290 | FALSE |
| Cherry | Cherry | 9.2 | 5.5 | 0.6619 | 0.3381 | TRUE |
| Pear | Pear | 5.3 | 3.1 | 0.3250 | 0.6750 | TRUE |
| Pear | Pear | 6.9 | 6.2 | 0.2714 | 0.7286 | TRUE |
| Pear | Pear | 9.2 | 8.9 | 0.0382 | 0.9618 | TRUE |
| Pear | Pear | 8.5 | 7.5 | 0.1751 | 0.8249 | TRUE |
| Cherry | Pear | 8.2 | 4.1 | 0.6431 | 0.3569 | FALSE |
| Pear | Pear | 5.7 | 4.0 | 0.3748 | 0.6252 | TRUE |
| Pear | Pear | 5.8 | 5.1 | 0.3235 | 0.6765 | TRUE |
| Cherry | Pear | 16.8 | 9.9 | 0.8095 | 0.1905 | FALSE |
| Pear | Pear | 6.1 | 5.9 | 0.2545 | 0.7455 | TRUE |
| Cherry | Pear | 8.9 | 4.1 | 0.7287 | 0.2713 | FALSE |
| Pear | Pear | 6.4 | 4.4 | 0.4250 | 0.5750 | TRUE |
| Pear | Pear | 18.8 | 15.2 | 0.0000 | 1.0000 | TRUE |

## New Data

```
## $class
## [1] Cherry Pear   Cherry
```

```
## Levels: Cherry Pear
##
## $posterior
##      Cherry      Pear
## 1 0.6362745 0.3637255
## 2 0.3382849 0.6617151
## 3 0.5712093 0.4287907
```

### Observation Space

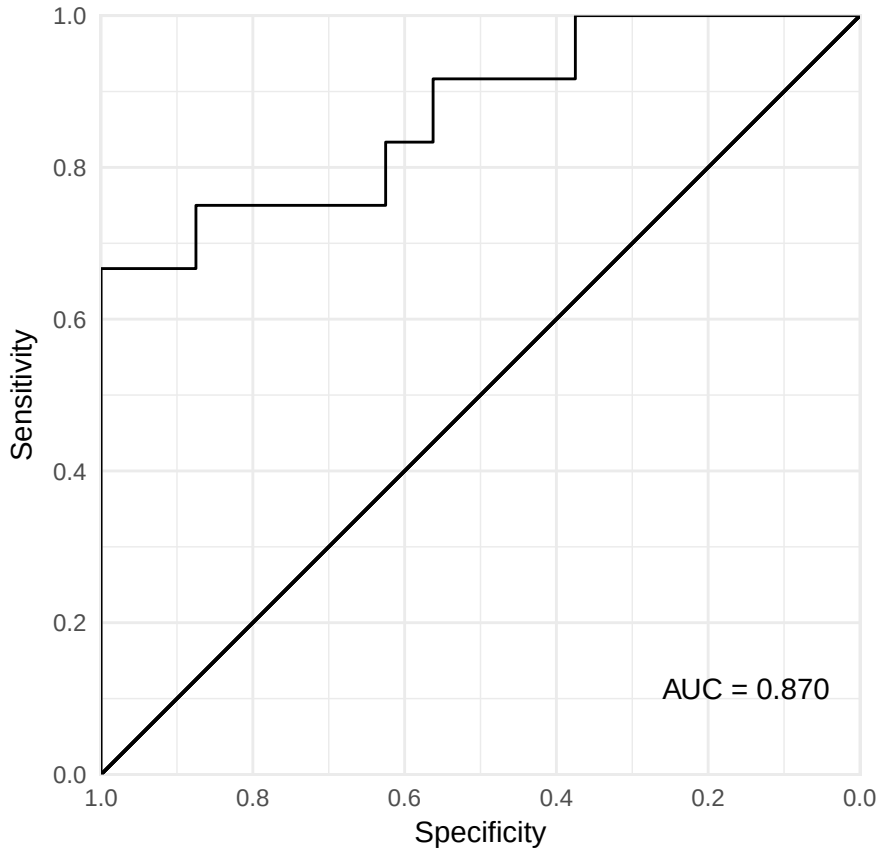



## Classification Procedure (GLM)

### Training Data

```
##
## Call:
## glm(formula = Type ~ Length + Width, family = binomial, data = .)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4568  -0.6116  -0.2002   0.5487   2.0874
##
## Coefficients:
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.7396     1.5194   1.145  0.25223
## Length       -0.7764     0.2875  -2.700  0.00693 **
## Width         0.9338     0.3815   2.448  0.01438 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 38.243  on 27  degrees of freedom
## Residual deviance: 24.257  on 25  degrees of freedom
## AIC: 30.257
##
## Number of Fisher Scoring iterations: 5
```



### New Data

### Observation Space

## Conclusion

## Appendix

### References

[1] The Parts of a Leaf. (17, October 30). Retrieved March 20, 18, from http://www.robinsonlibrary.com/science/botany/anatomy/leafparts.htm [2] Britannica, T. E. (2016, November 11). Cherry. Retrieved March 20, 2018, from https://www.britannica.com/plant/cherry

[3] Britannica, T. E. (2015, May 13). Pear. Retrieved March 20, 2018, from https://www.britannica.com/plant/pear