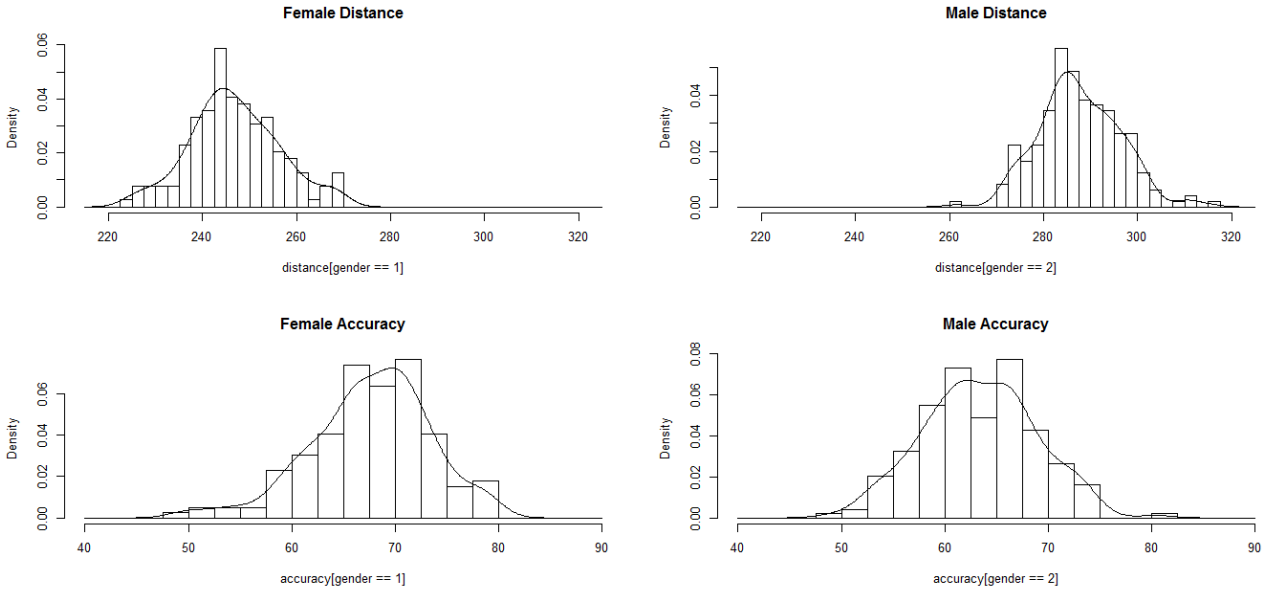
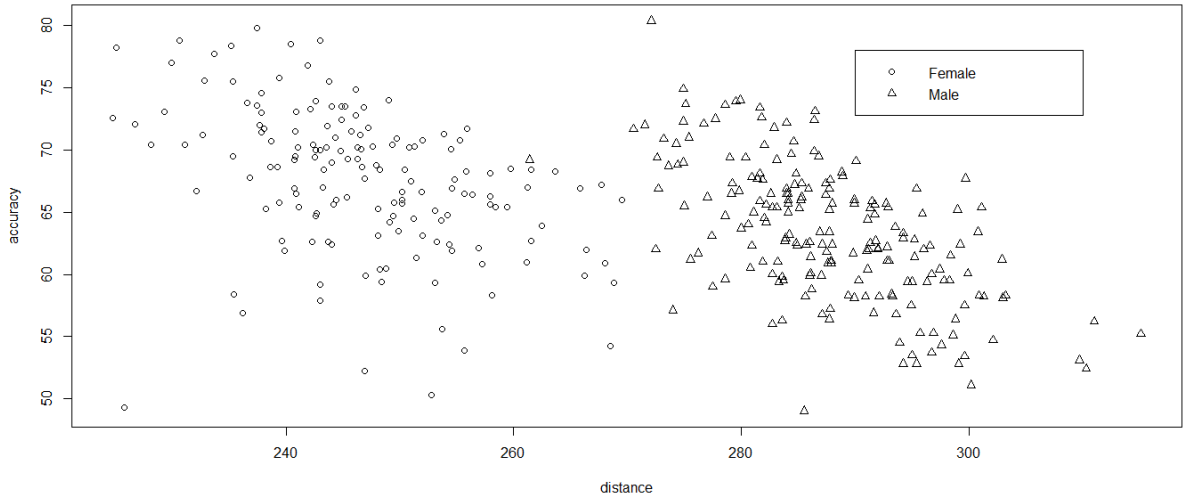
LPGA/PGA Driving Distance and Accuracy - Discriminant Analysis

Data: Groups: **F**emale ($N_F = 157$, $p_F = 0.4435$) and **M**ale ($N_M = 197$, $p_M = 0.5565$) professional golfers

Variables: Driving Distance (Yards) and Accuracy (%)

Treating data as a MVN Population with common (pooled) Variance-Covariance matrix Σ.






```
        mu.F    mu.M Sigma.F         Sigma.M         Sigma.FM
[1,] 246.801 287.611  90.132 -23.118  73.179 -28.385   80.922 -26.125
[2,]  67.591  63.365 -23.118  33.278 -28.385  29.824  -26.125  31.444
```

$$f_F(\mathbf{x}) = (2\pi)^{-2/2}\,|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}_F)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_F)\right]\right\} \qquad f_M(\mathbf{x}) = (2\pi)^{-2/2}\,|\boldsymbol{\Sigma}|^{-1/2}\exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu}_M)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_M)\right]\right\}$$

Assuming equal costs of Misclassification: C(F|M) = C(M|F), and prior probabilities: $p_F = .4435$, $p_M = .5565$, the following classification rule is used:

Classify as Female if: $\frac{f_F(\mathbf{x})}{f_M(\mathbf{x})} \geq \left(\frac{C(F|M)}{C(M|F)}\right)\left(\frac{p_M}{p_F}\right) = 1.2548$ Male otherwise

```
> for (i in 1:n.T)  {
+ fF.X[i] <- norm.c*
+     exp(-0.5*(t(X[,i]-mu.F) %*% Sigma.FM.inv %*% (X[,i]-mu.F)))
+ fM.X[i] <- norm.c*
+     exp(-0.5*(t(X[,i]-mu.M) %*% Sigma.FM.inv %*% (X[,i]-mu.M)))
+ }
> classify <- ifelse((fF.X/fM.X) >= (CF.M/CM.F)*(p.M/p.F),1,2)
>
> table(gender,classify)
      classify
gender   1   2
     1 155   2
     2   1 196
```

2 Females were misclassified and 1 male.