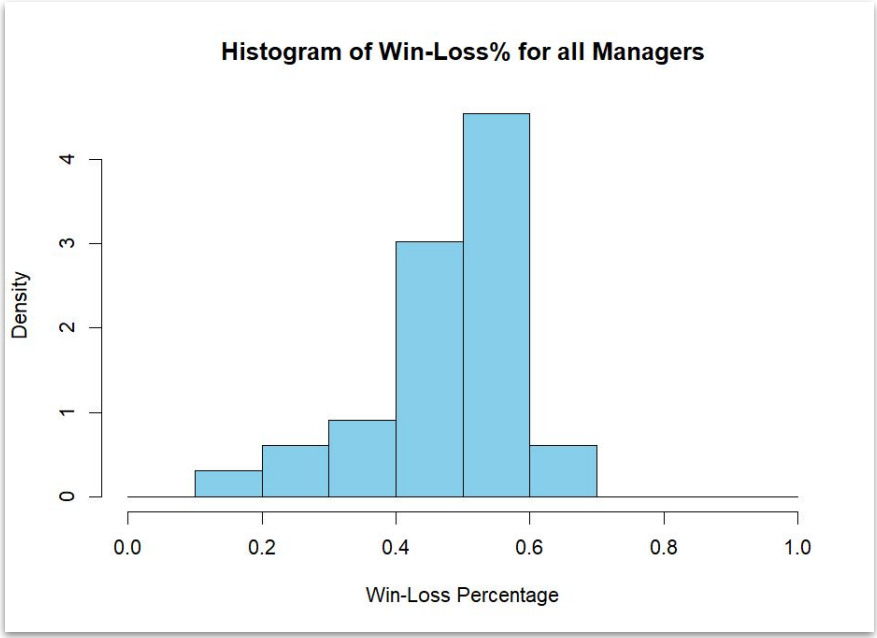
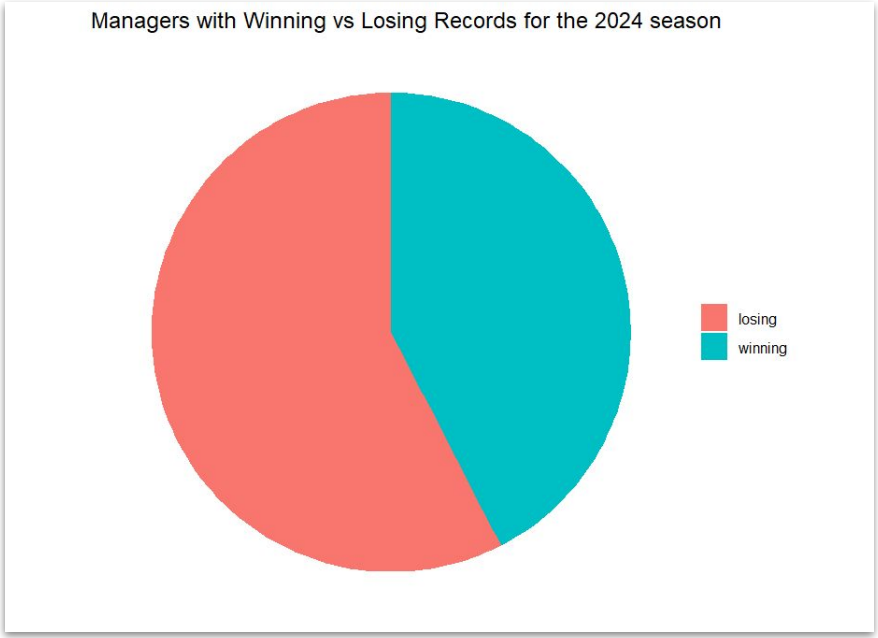
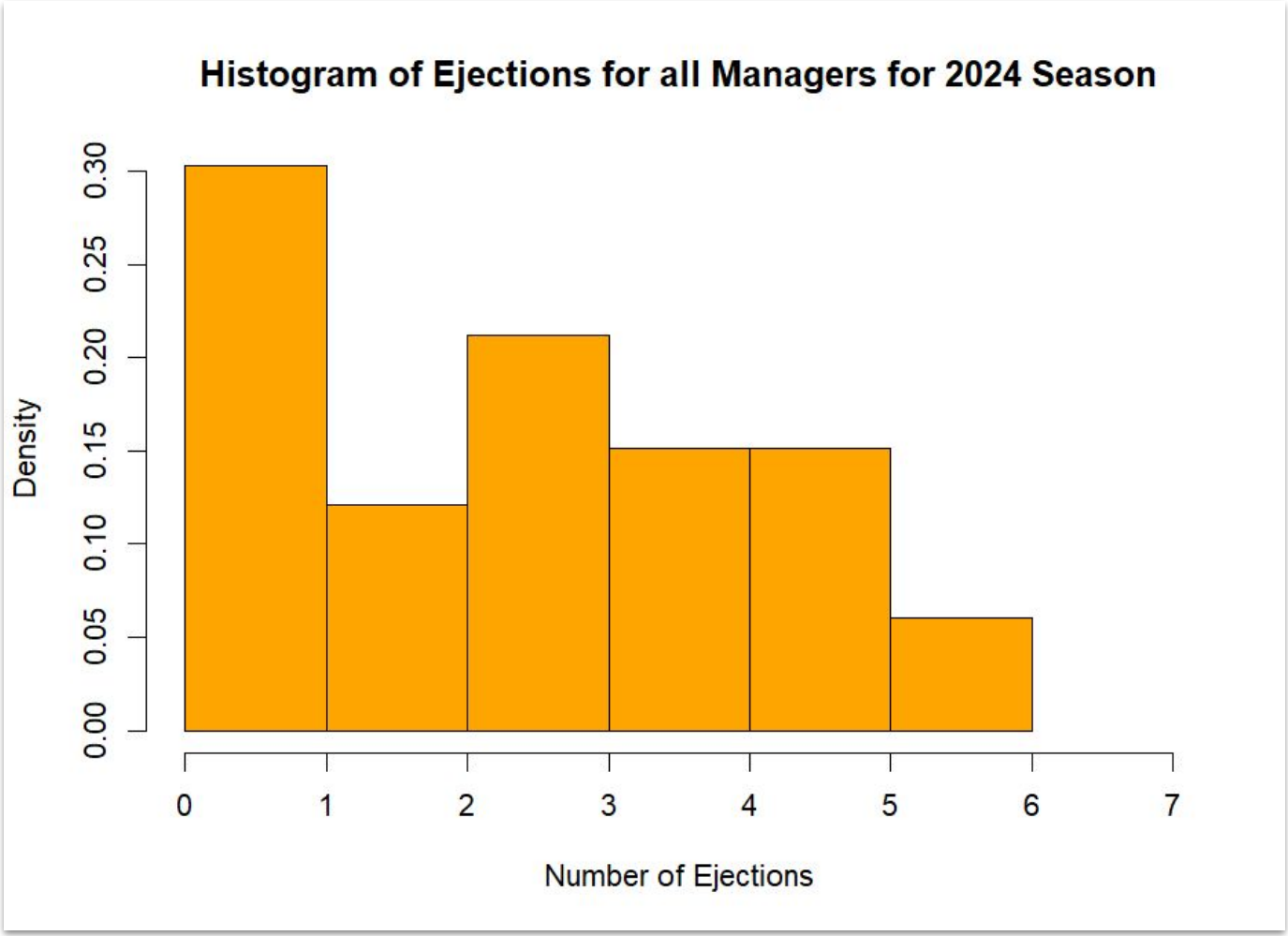
# 2024 MLB Managerial Data Analysis Visualizations

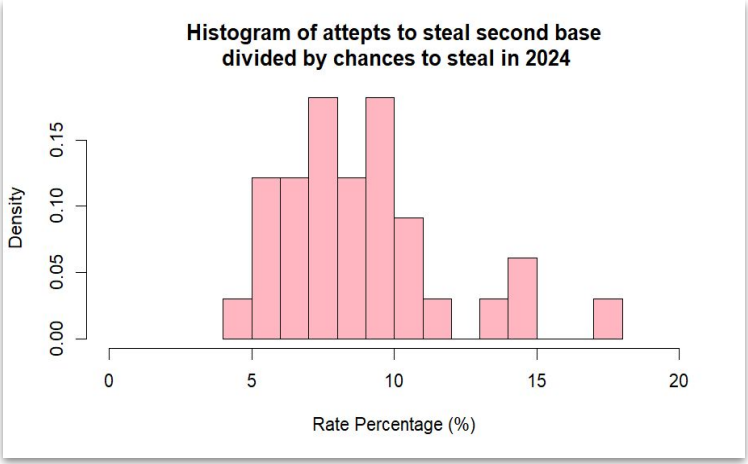
Dani DiTomasso, Data Analytics 2024 Assignment 6

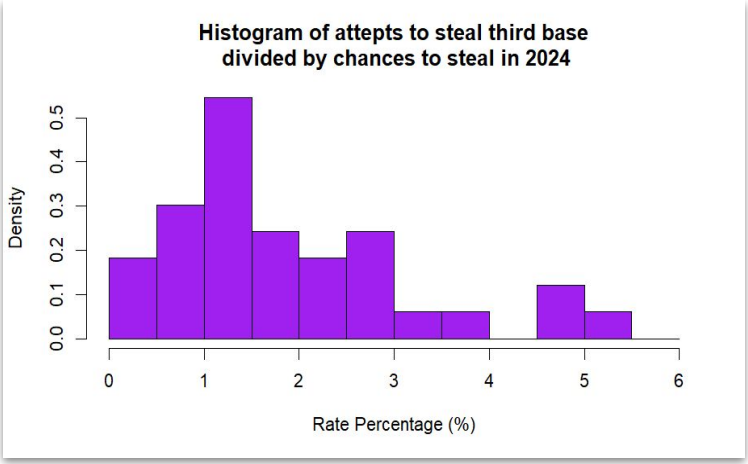
# Win-Loss Percentage EDA

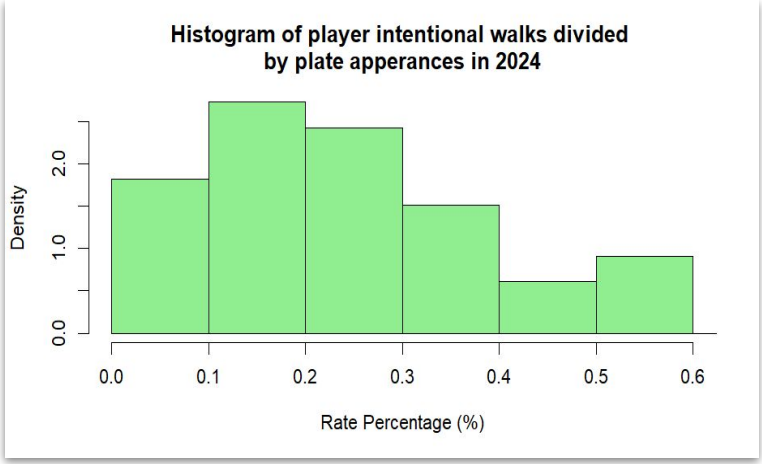


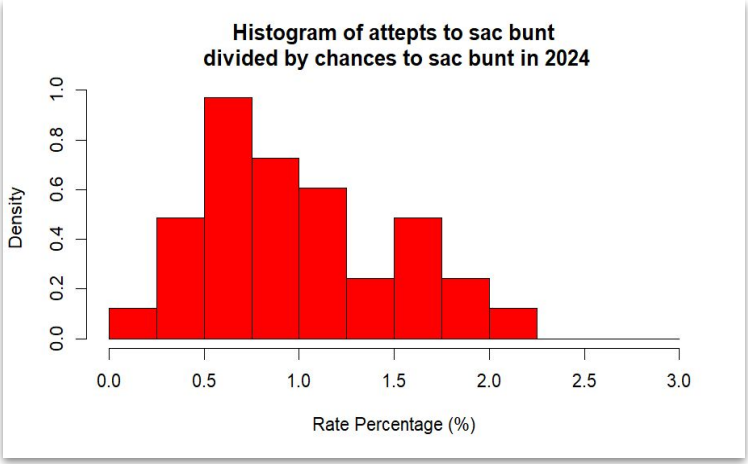




Histogram of Ejections for all Managers for 2024 Season
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1
2
3
4
5
6
7
Number of Ejections


Histogram of attepts to steal second base
divided by chances to steal in 2024
Density
0.00
0.05
0.10
0.15
0
5
10
15
20
Rate Percentage (%)


Histogram of attepts to steal third base
divided by chances to steal in 2024
Density
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5
6
Rate Percentage (%)


Histogram of player intentional walks divided
by plate apperances in 2024
Density
0.0
1.0
2.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Rate Percentage (%)


Histogram of attepts to sac bunt
divided by chances to sac bunt in 2024
Density
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Rate Percentage (%)

## Records Variables

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.365e-01  7.706e-02   6.963 1.75e-07 ***
Challenges   1.248e-03  2.627e-03   0.475    0.639
Ejections   -1.647e-03  6.298e-03  -0.261    0.796
Age         -5.915e-05  1.193e-03  -0.050    0.961
Overturned   3.339e-03  4.538e-03   0.736    0.468
Finish      -4.655e-02  7.568e-03  -6.151 1.42e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

## Second Base

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.4251533  0.3994955  -1.064   0.2963
Ch           0.0006628  0.0002853   2.323   0.0276 *
Att         -0.0059353  0.0033316  -1.782   0.0857 .
Rate         0.0700785  0.0546544   1.282   0.2103
`Rate+`      0.0011202  0.0026908   0.416   0.6804
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

## Third Base

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.551e-01  5.889e-02   2.634  0.01358 *
Ch.1         3.688e-04  6.959e-05   5.299 1.23e-05 ***
Att.1       -9.901e-03  3.215e-03  -3.080  0.00461 **
Rate.1       4.546e-02  1.116e-01   0.407  0.68695
`Rate+.1`    9.404e-04  1.948e-03   0.483  0.63306
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

## Intentional Walks

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.129e-01  1.202e-01   6.762 2.42e-07 ***
PA          -4.779e-05  2.155e-05  -2.218  0.03485 *
IBB          1.579e-02  4.602e-03   3.431  0.00188 **
Rate.3      -1.272e+00  3.916e-01  -3.248  0.00301 **
`Rate+.3`    5.894e-04  9.342e-04   0.631  0.53318
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

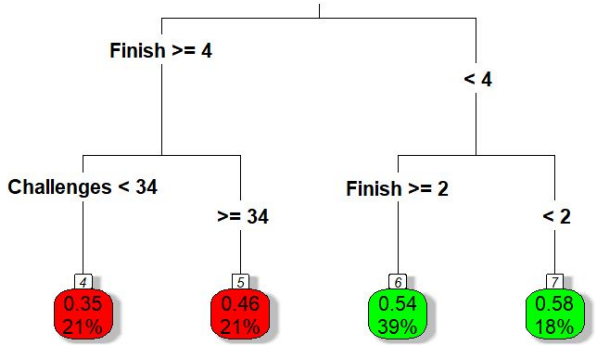
## Substitutions

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.506114   0.484316   1.045   0.3056
`PH/G`      -1.353632   0.829943  -1.631   0.1149
`PH/G+`      0.012267   0.006760   1.815   0.0811 .
`PR/G`       3.726675   2.358327   1.580   0.1261
`PR/G+`     -0.006946   0.004365  -1.591   0.1237
`P/G`        0.210244   0.724750   0.290   0.7740
`P/G+`      -0.010376   0.030110  -0.345   0.7332
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

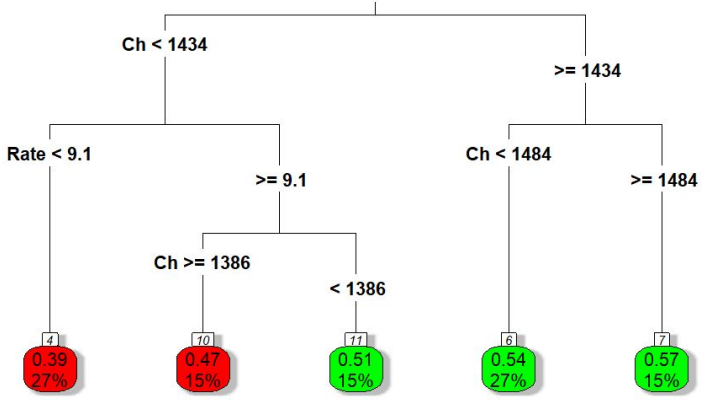
## Sacrificial Bunts

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.632e-01  7.544e-02   2.163   0.0392 *
Ch.2         2.678e-04  5.895e-05   4.543 9.66e-05 ***
Att.2       -1.539e-02  6.125e-03  -2.513   0.0180 *
Rate.2       2.244e-01  1.400e-01   1.603   0.1202
`Rate+.2`   -4.381e-04  9.548e-04  -0.459   0.6499
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

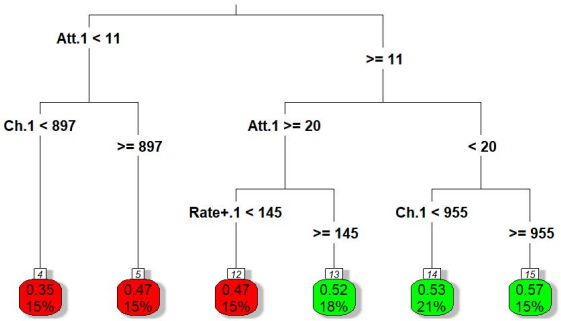
## Decision Tree for Win-Loss Predictions for Manager Record Data



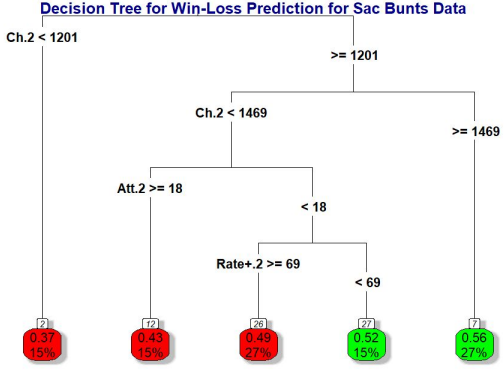

## Decision Tree for Win-Loss Predictions for Second Base Data



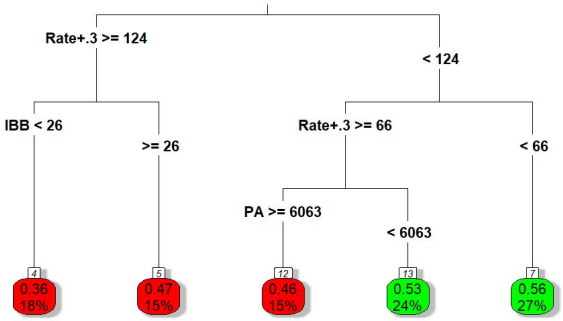

## Decision Tree for Win-Loss Predictions for Third Base Data



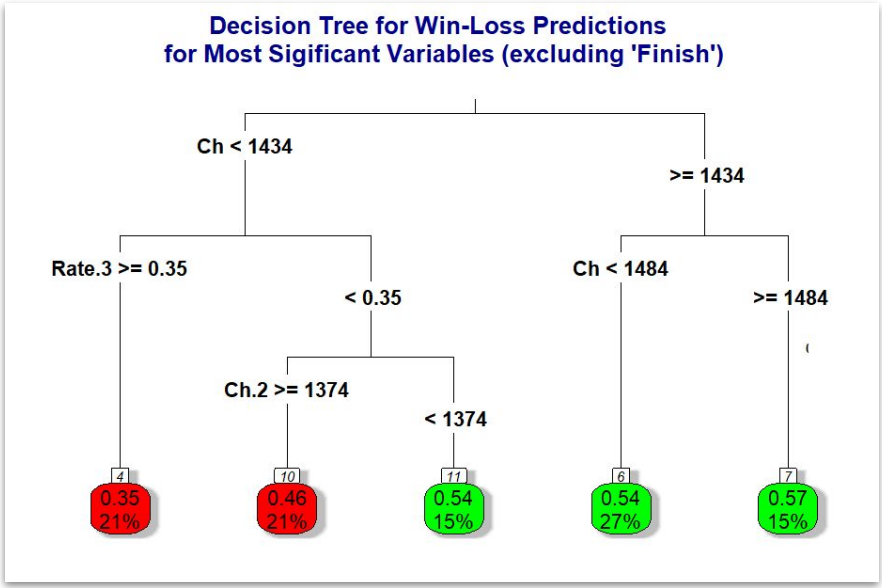

## Decision Tree for Win-Loss Prediction for Sac Bunts Data



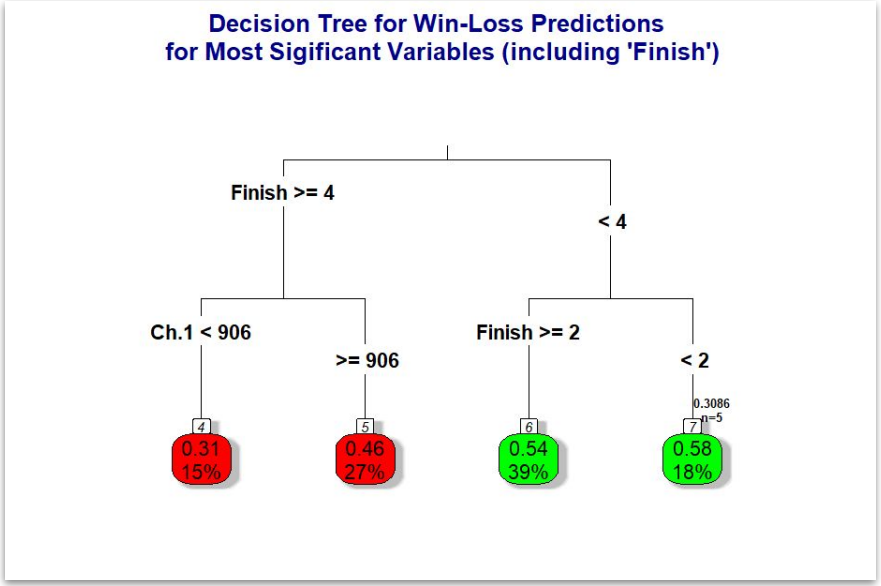

## Decision Tree for Win-Loss Predictions for Intentional Walk Data



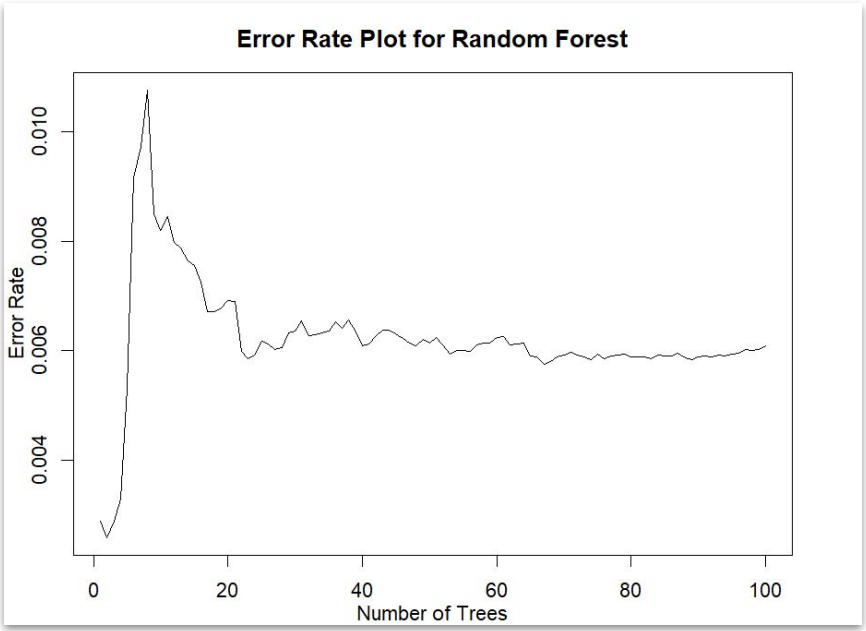
# Decision Tree Models for Most Significant Variables

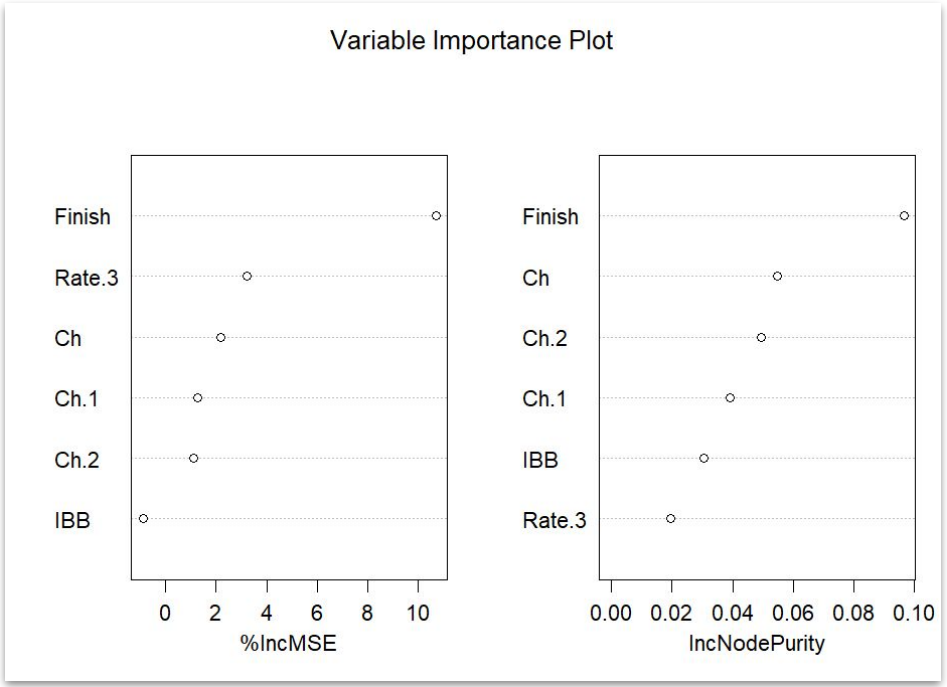



Mean Squared Error: 0.00222707845117845
R-squared: 0.771327433622756




Mean Squared Error: 0.00360962116402116
R-squared: 0.629370337273226

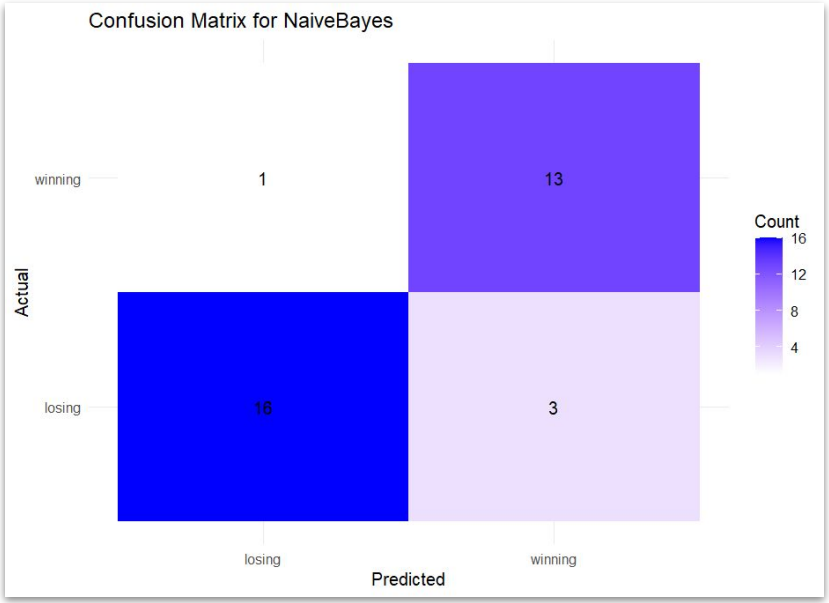


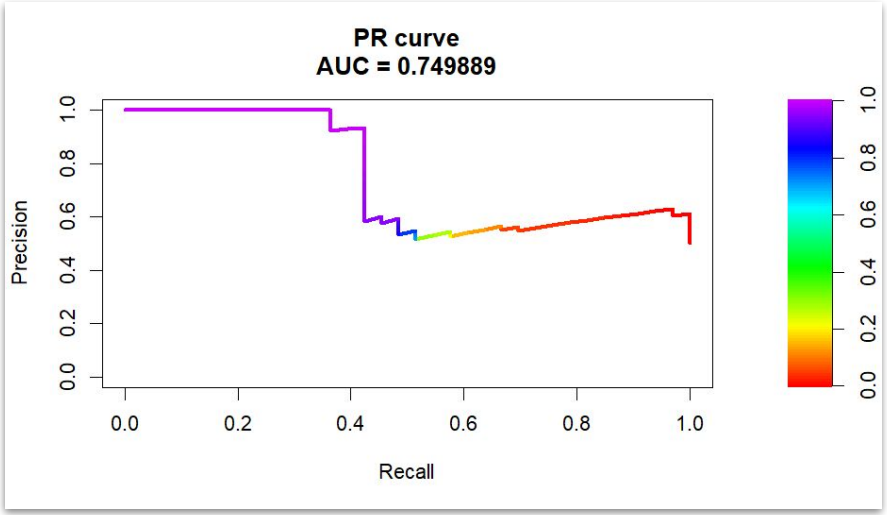



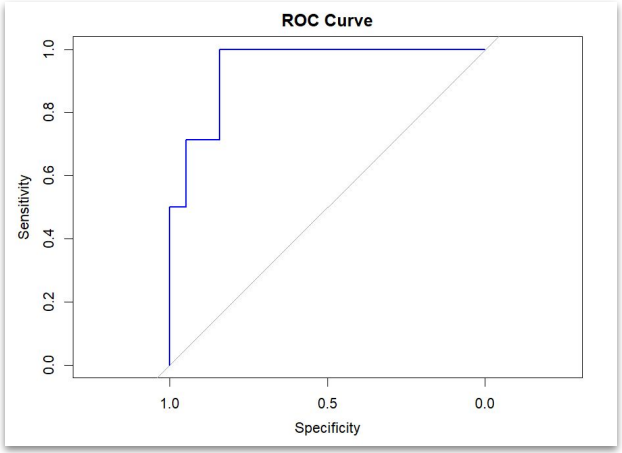

# Random Forest Data:

Mean Squared Error: 0.00136983876035943
R-squared: 0.859347323535612

# Naive Bayes


Confusion Matrix for NaiveBayes
winning
losing
1
13
16
3
Actual
losing
winning
Predicted
Count
16
12
8
4


PR curve
AUC = 0.749889
Precision
0.0
0.2
0.4
0.6
0.8
1.0
Recall
0.0
0.2
0.4
0.6
0.8
1.0


ROC Curve
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
Specificity
1.0
0.5
0.0