Introduction

Tennis is one of the most popular sports around the world. The game is very dynamic and has a lot of moving parts: serving, baseline play, and net play. Each match consists of many points, and thus a lot of data can be collected about each player's performance over the course of an entire match. The goal of this analysis is to understand which match and/or player statistics have the strongest associations with winning a match and to quantify those associations.

The dataset we are analyzing contains match statistics from 943 men's and women's matches which occurred in one of four grand slam tournaments in 2013: the Australian Open, the French Open, Wimbledon, and the US Open. Each observation is an individual match. The response variable is a binary indicator: a value of 1 indicates that player 1 won the match, and a value of 0 indicates that player 1 lost the match (the player numbers are arbitrarily assigned and correspond to which player name is listed first for each observation). The match statistics recorded for each observation are First Serve Percentage, First Serves Won, Second Serve Percentage, Second Serves Won, Aces, Double Faults, Winners, Unforced Errors, Break Points Created, Break Points Won, Net Points Attempted, Net Points Won, Total Points Won, Results for each set, and Final Number of Games Won (recorded for each player). The Australian Open and the US Open are both played on hard court, the French Open is played on clay, and Wimbledon is played on grass.

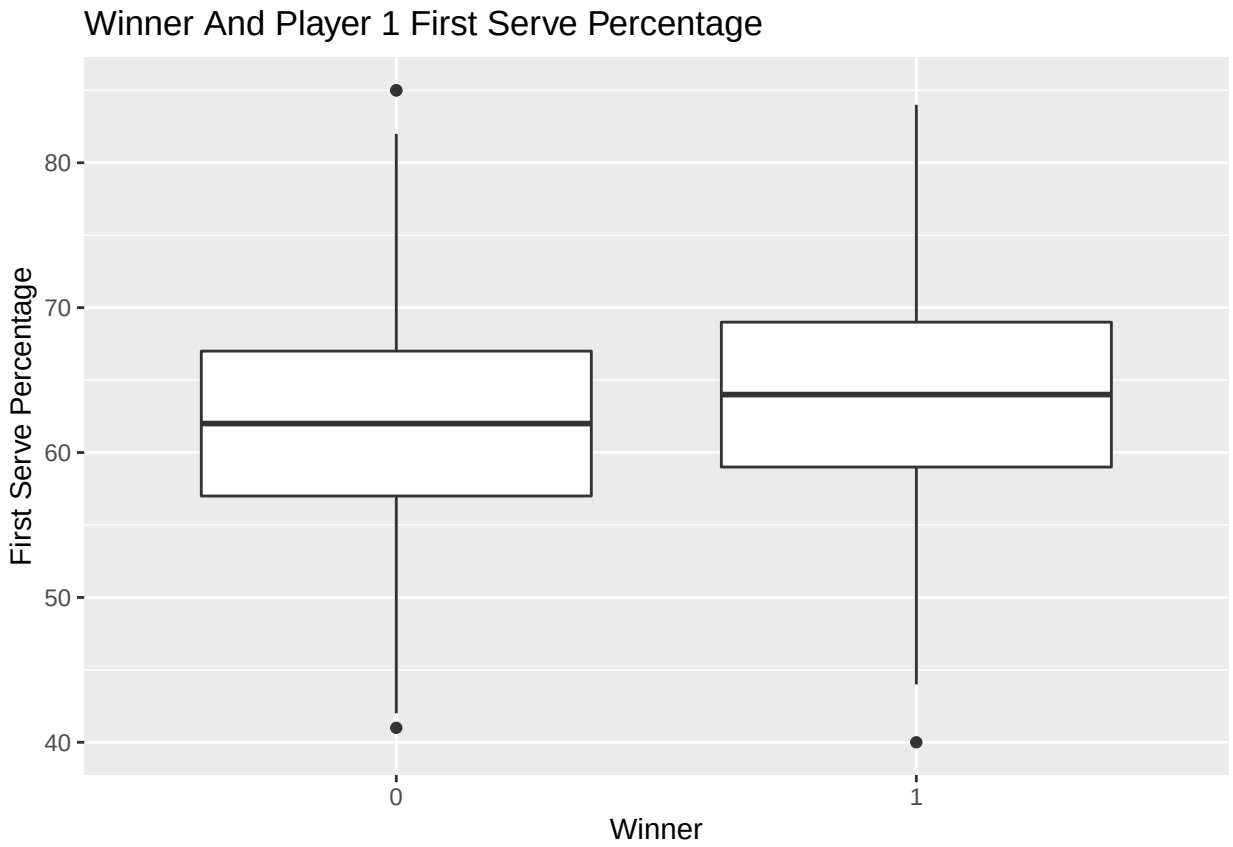
Our primary research question for this analysis is: Which tennis match statistics are most strongly associated with winning, and does that vary by gender and/or court surface?

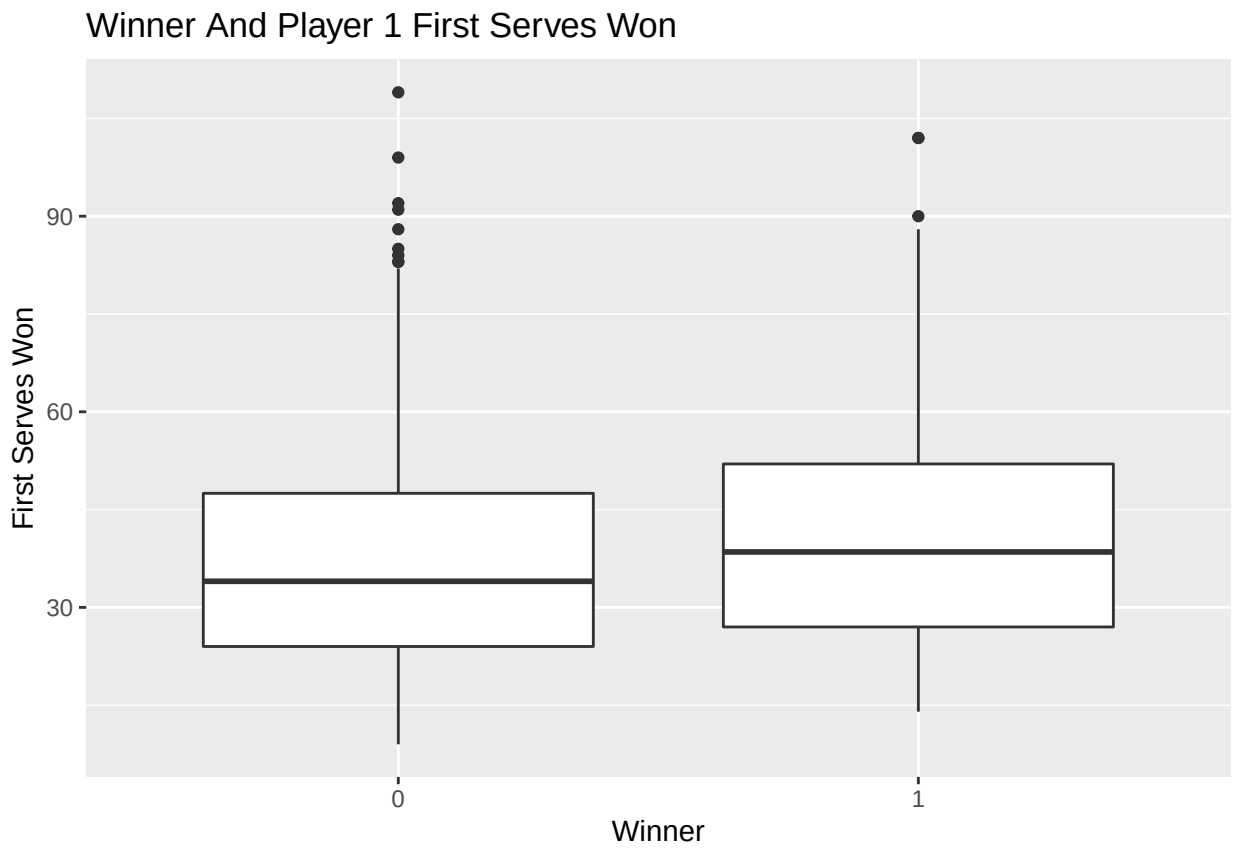
Data

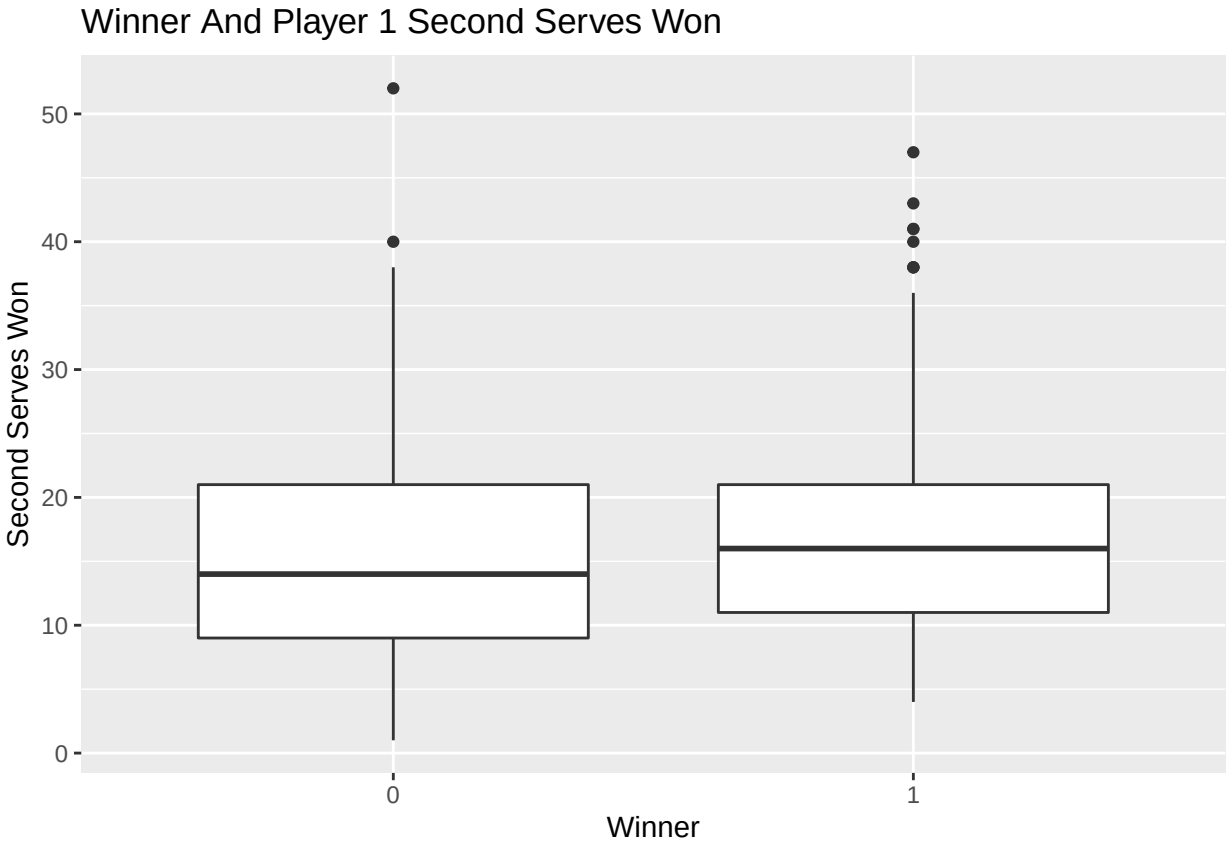
After examining the data, I first selected only for relevant match statistics: player names were removed because I am not considering any background information about specific players in my analysis, and the results at the end of each set and the total games won were removed because I am only interested in purely how players' performance during the points is associated with the final winner.

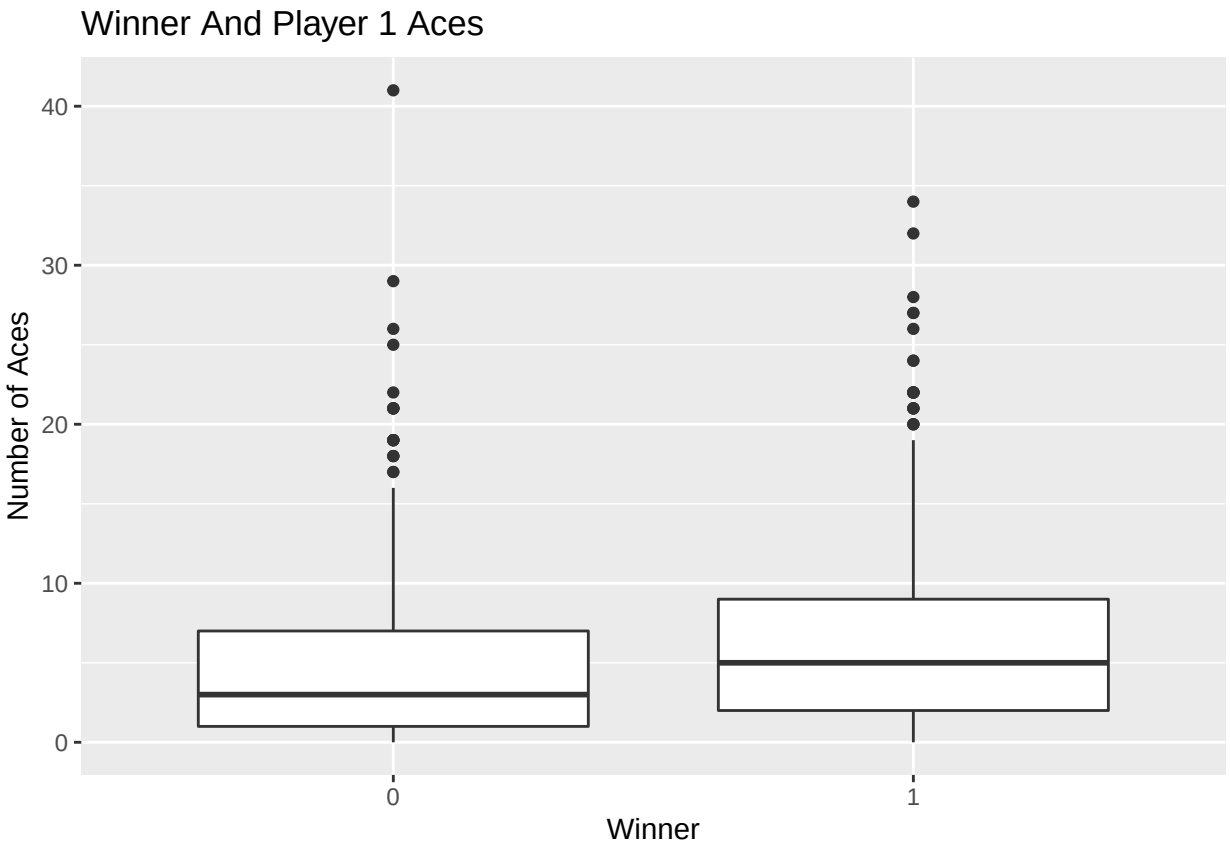
I also noticed that virtually every observation from the US Open had some missing data. Thus, I completely removed data from the US Open from my analysis and only analyzed data from the other three grand slam tournaments (one of each surface). In the remaining data, 74 out of 741 observations had some missingness with no clear pattern. Because this was a relatively small number of observations with no clear pattern of missingness, I dropped the observations with missing data from my analysis. However, it is worth noting I have not verified the assumption that the missing data is missing completely at random.

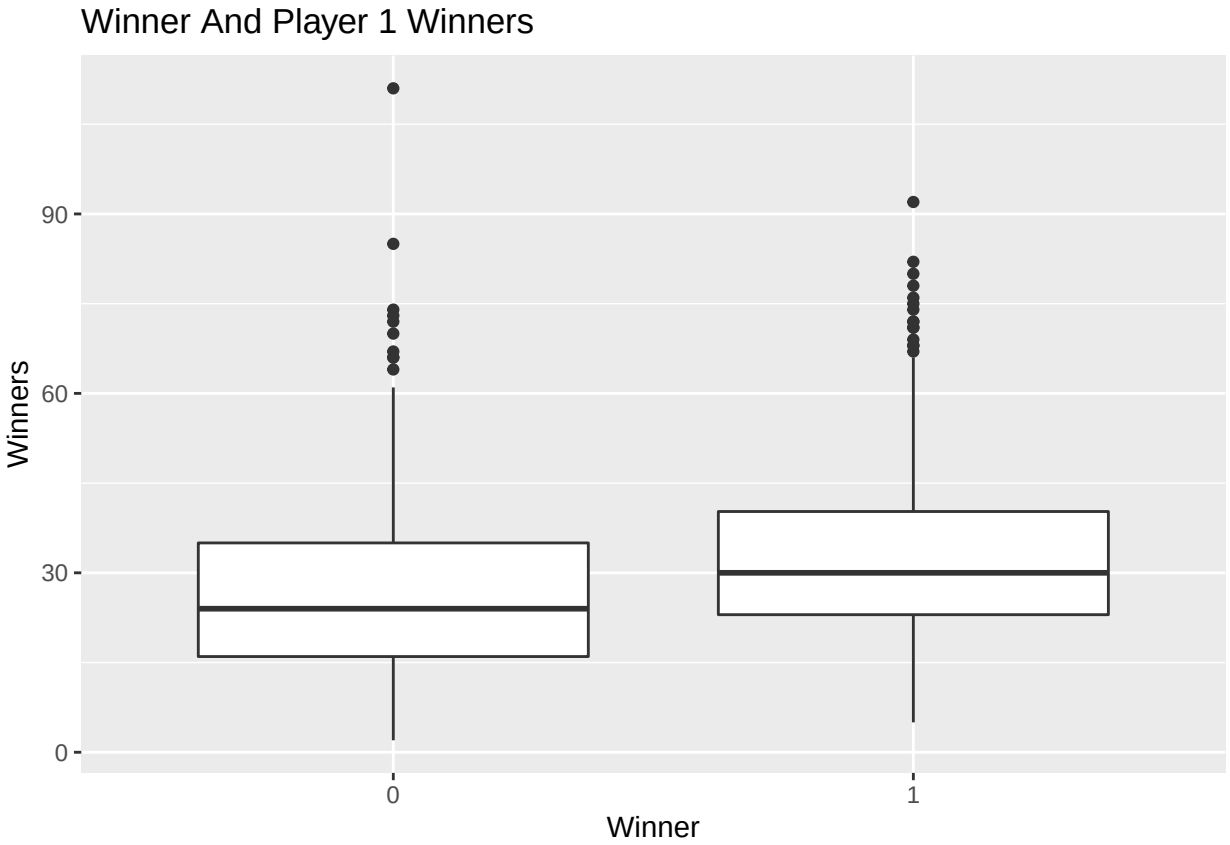
To visually investigate the associations between match statistics and match winners, I created bar plots comparing player 1 statistics when player 1 won and when player 1 lost.

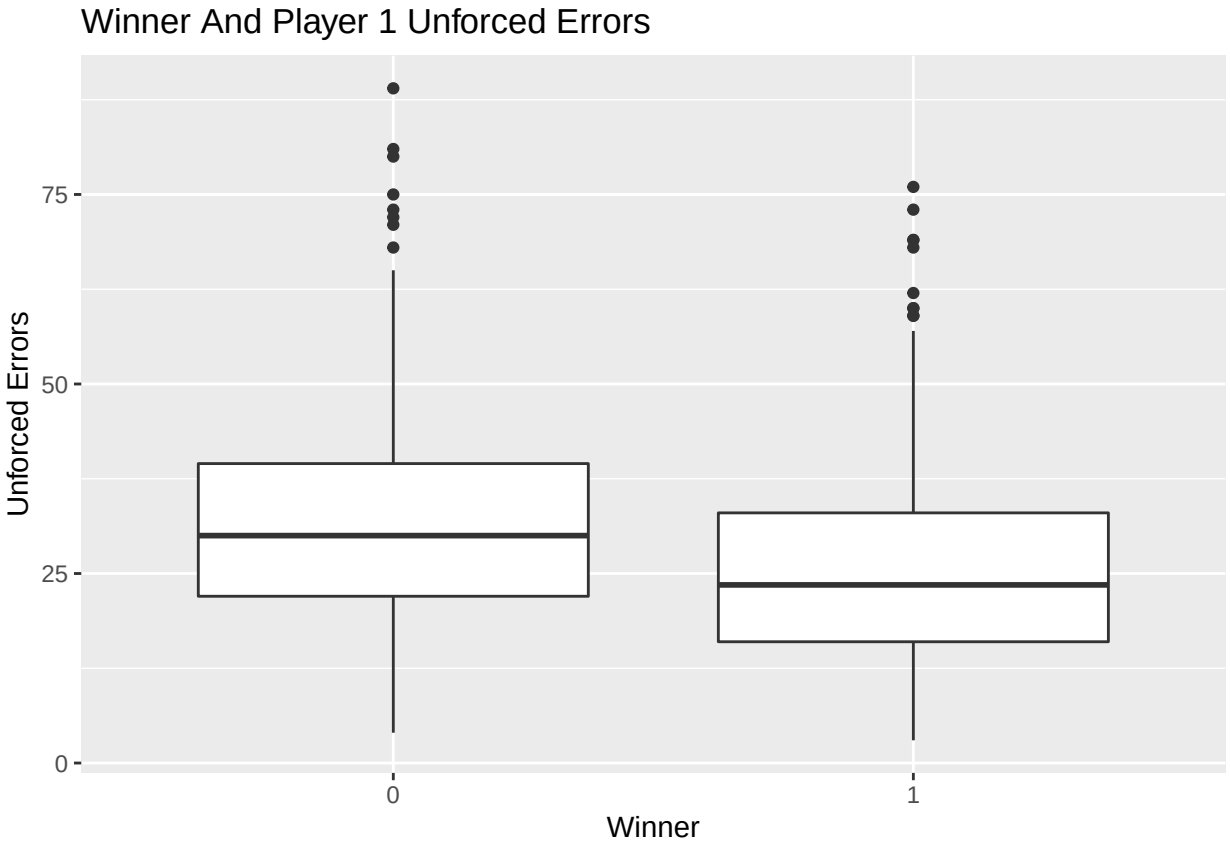

Winner And Player 1 First Serve Percentage
First Serve Percentage
80
70
60
50
40
0
1
Winner

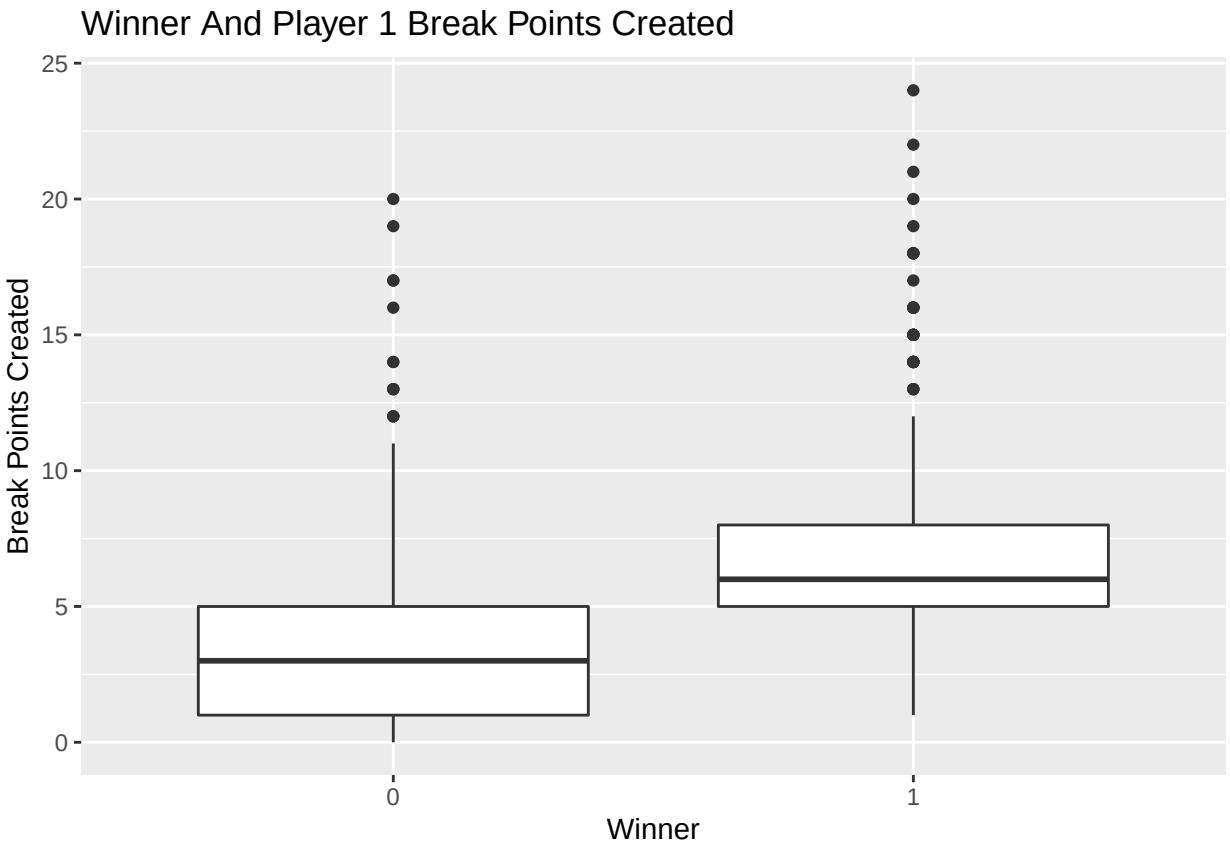

Winner And Player 1 First Serves Won
First Serves Won
90
60
30
0
1
Winner

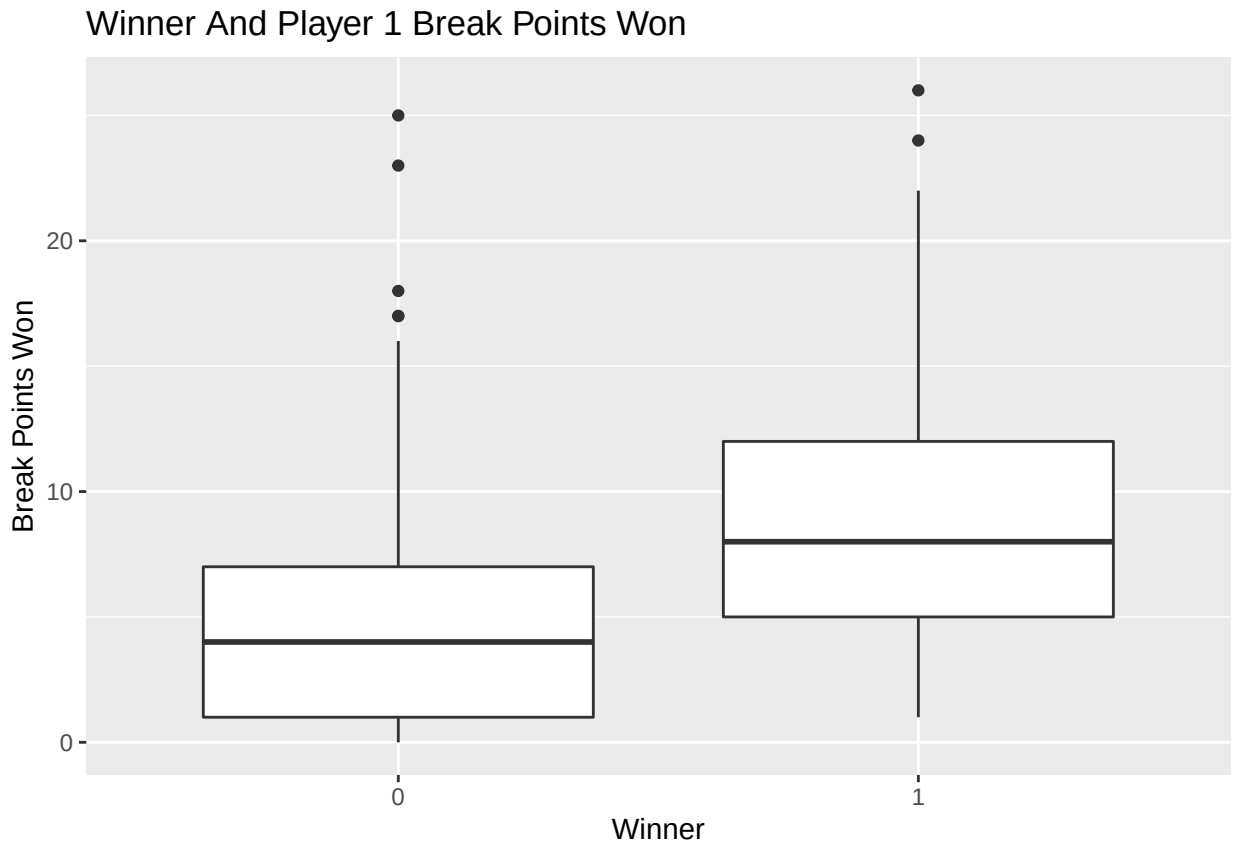

Winner And Player 1 Second Serves Won
Second Serves Won
50
40
30
20
10
0
0
1
Winner


Winner And Player 1 Aces
Number of Aces
40
30
20
10
0
0
1
Winner


Winner And Player 1 Winners
Winners
90
60
30
0
0
1
Winner


Winner And Player 1 Unforced Errors
Unforced Errors
75
50
25
0
0
1
Winner


Winner And Player 1 Break Points Created
Break Points Created
25
20
15
10
5
0
0
1
Winner



Methodology

With a set of predictors and a binary response variable, I started my analysis with logistic regression modelling,

Evaluating the correlation of all the predictor variables showed that Second Serve Percentage is perfectly negatively correlated with First Serve Percentage for both players (as shown in Appendix B). Therefore, I removed Second Serve Percentage entirely from my models.

I began with building a logistic regression model from the full dataset.

Overall

```
##
## Call:
## lm(formula = Result ~ . - gender - tournament - Result - SSP.1 -
##     SSP.2, data = data_final, family = "binomial")
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.83118 -0.17292 -0.00493  0.15446  0.93324
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.4638764  0.2023231   2.293 0.022184 *
## FSP.1       -0.0014896  0.0022105  -0.674 0.500637
## FSP.2        0.0020308  0.0022300   0.911 0.362805
## FSW.1        0.0210486  0.0018804  11.194  < 2e-16 ***
```

```
## FSW.2       -0.0212643  0.0020767 -10.239  < 2e-16 ***
## SSW.1        0.0156799  0.0030444   5.150 3.46e-07 ***
## SSW.2       -0.0164605  0.0030092  -5.470 6.45e-08 ***
## ACE.1       -0.0001632  0.0028407  -0.057 0.954214
## ACE.2       -0.0031558  0.0029615  -1.066 0.286990
## DBF.1        0.0009832  0.0050166   0.196 0.844679
## DBF.2        0.0018205  0.0047812   0.381 0.703503
## WNR.1        0.0033712  0.0018414   1.831 0.067587 .
## WNR.2       -0.0038987  0.0018864  -2.067 0.039154 *
## UFE.1       -0.0051863  0.0016382  -3.166 0.001619 **
## UFE.2        0.0035135  0.0015434   2.276 0.023147 *
## BPC.1        0.0318515  0.0039661   8.031 4.61e-15 ***
## BPC.2       -0.0318471  0.0041469  -7.680 5.95e-14 ***
## BPW.1        0.0235367  0.0034100   6.902 1.23e-11 ***
## BPW.2       -0.0131252  0.0036095  -3.636 0.000299 ***
## NPA.1        0.0047491  0.0021593   2.199 0.028208 *
## NPA.2       -0.0013301  0.0022323  -0.596 0.551496
## NPW.1       -0.0044641  0.0022555  -1.979 0.048221 *
## NPW.2        0.0027300  0.0023853   1.145 0.252836
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2576 on 644 degrees of freedom
## Multiple R-squared:  0.7437, Adjusted R-squared:  0.7349
## F-statistic: 84.94 on 22 and 644 DF,  p-value: < 2.2e-16
```

Women

```
##
## Call:
## glm(formula = Result ~ FSP.1 + FSP.2 + FSW.1 + FSW.2 + SSW.1 +
##     SSW.2 + ACE.1 + ACE.2 + DBF.1 + DBF.2 + WNR.1 + WNR.2 + UFE.1 +
##     UFE.2 + BPC.1 + BPC.2 + BPW.1 + BPW.2 + NPA.1 + NPA.2 + NPW.1 +
##     NPW.2, family = "binomial", data = data_women)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.15441  -0.00931   0.00028   0.01681   2.43599
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -12.752651  12.297487  -1.037 0.299730
## FSP.1         0.064247   0.146062   0.440 0.660039
## FSP.2         0.026912   0.118435   0.227 0.820245
## FSW.1         0.720396   0.185704   3.879 0.000105 ***
## FSW.2        -0.435214   0.143967  -3.023 0.002503 **
## SSW.1         0.494056   0.216845   2.278 0.022704 *
## SSW.2        -0.395578   0.202434  -1.954 0.050688 .
## ACE.1         0.168282   0.219268   0.767 0.442800
## ACE.2         0.025473   0.189495   0.134 0.893065
## DBF.1        -0.007332   0.320780  -0.023 0.981763
## DBF.2         0.376439   0.221004   1.703 0.088509 .
## WNR.1         0.036922   0.089534   0.412 0.680060
## WNR.2        -0.375733   0.132168  -2.843 0.004471 **
```

```
## UFE.1        -0.280885   0.104255  -2.694 0.007055 **
## UFE.2         0.129494   0.082316   1.573 0.115691
## BPC.1         0.852501   0.253618   3.361 0.000776 ***
## BPC.2        -0.460724   0.199638  -2.308 0.021011 *
## BPW.1         0.793397   0.236690   3.352 0.000802 ***
## BPW.2         0.048474   0.167286   0.290 0.771994
## NPA.1        -0.026421   0.138195  -0.191 0.848377
## NPA.2         0.013967   0.126210   0.111 0.911882
## NPW.1         0.015071   0.152353   0.099 0.921202
## NPW.2        -0.010827   0.161583  -0.067 0.946577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 453.92  on 327  degrees of freedom
## Residual deviance:  46.51  on 305  degrees of freedom
## AIC: 92.51
##
## Number of Fisher Scoring iterations: 10
```

Men

```
##
## Call:
## glm(formula = Result ~ FSP.1 + FSP.2 + FSW.1 + FSW.2 + SSW.1 +
##     SSW.2 + ACE.1 + ACE.2 + DBF.1 + DBF.2 + WNR.1 + WNR.2 + UFE.1 +
##     UFE.2 + BPC.1 + BPC.2 + BPW.1 + BPW.2 + NPA.1 + NPA.2 + NPW.1 +
##     NPW.2, family = "binomial", data = data_women)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.15441  -0.00931   0.00028   0.01681   2.43599
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -12.752651  12.297487  -1.037 0.299730
## FSP.1         0.064247   0.146062   0.440 0.660039
## FSP.2         0.026912   0.118435   0.227 0.820245
## FSW.1         0.720396   0.185704   3.879 0.000105 ***
## FSW.2        -0.435214   0.143967  -3.023 0.002503 **
## SSW.1         0.494056   0.216845   2.278 0.022704 *
## SSW.2        -0.395578   0.202434  -1.954 0.050688 .
## ACE.1         0.168282   0.219268   0.767 0.442800
## ACE.2         0.025473   0.189495   0.134 0.893065
## DBF.1        -0.007332   0.320780  -0.023 0.981763
## DBF.2         0.376439   0.221004   1.703 0.088509 .
## WNR.1         0.036922   0.089534   0.412 0.680060
## WNR.2        -0.375733   0.132168  -2.843 0.004471 **
## UFE.1        -0.280885   0.104255  -2.694 0.007055 **
## UFE.2         0.129494   0.082316   1.573 0.115691
## BPC.1         0.852501   0.253618   3.361 0.000776 ***
## BPC.2        -0.460724   0.199638  -2.308 0.021011 *
## BPW.1         0.793397   0.236690   3.352 0.000802 ***
```

```
## BPW.2         0.048474   0.167286   0.290 0.771994
## NPA.1        -0.026421   0.138195  -0.191 0.848377
## NPA.2         0.013967   0.126210   0.111 0.911882
## NPW.1         0.015071   0.152353   0.099 0.921202
## NPW.2        -0.010827   0.161583  -0.067 0.946577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 453.92  on 327  degrees of freedom
## Residual deviance:  46.51  on 305  degrees of freedom
## AIC: 92.51
##
## Number of Fisher Scoring iterations: 10
```

Australian Open

```
##
## Call:
## glm(formula = Result ~ FSP.1 + FSP.2 + FSW.1 + FSW.2 + SSW.1 +
##     SSW.2 + ACE.1 + ACE.2 + DBF.1 + DBF.2 + WNR.1 + WNR.2 + UFE.1 +
##     UFE.2 + BPC.1 + BPC.2 + BPW.1 + BPW.2 + NPA.1 + NPA.2 + NPW.1 +
##     NPW.2, family = "binomial", data = aus_open)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -1.265e-04  -2.100e-08   2.100e-08   2.100e-08   9.410e-05
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.550e+02  2.744e+05   0.001    0.999
## FSP.1        8.769e-01  1.583e+03   0.001    1.000
## FSP.2       -6.707e+00  3.424e+03  -0.002    0.998
## FSW.1        1.234e+01  2.071e+03   0.006    0.995
## FSW.2        1.738e+00  1.183e+03   0.001    0.999
## SSW.1        1.542e+01  3.706e+03   0.004    0.997
## SSW.2       -1.663e+01  2.605e+03  -0.006    0.995
## ACE.1       -6.734e+00  3.236e+03  -0.002    0.998
## ACE.2        3.323e+00  2.812e+03   0.001    0.999
## DBF.1       -1.468e+01  4.755e+03  -0.003    0.998
## DBF.2        2.792e+01  5.143e+03   0.005    0.996
## WNR.1       -4.853e+00  1.696e+03  -0.003    0.998
## WNR.2       -3.526e+00  3.064e+03  -0.001    0.999
## UFE.1       -2.160e+00  1.144e+03  -0.002    0.998
## UFE.2       -7.375e+00  2.072e+03  -0.004    0.997
## BPC.1        1.433e+02  2.150e+04   0.007    0.995
## BPC.2       -5.228e+01  1.188e+04  -0.004    0.996
## BPW.1       -1.473e+01  3.235e+03  -0.005    0.996
## BPW.2       -9.857e+00  2.834e+03  -0.003    0.997
## NPA.1        3.141e+01  5.828e+03   0.005    0.996
## NPA.2        4.164e+00  4.213e+03   0.001    0.999
## NPW.1       -2.088e+01  3.916e+03  -0.005    0.996
## NPW.2       -2.120e+00  3.023e+03  -0.001    0.999
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.7668e+02  on 201  degrees of freedom
## Residual deviance: 1.1413e-07  on 179  degrees of freedom
##   (51 observations deleted due to missingness)
## AIC: 46
##
## Number of Fisher Scoring iterations: 25
```

French Open

```
##
## Call:
## glm(formula = Result ~ FSP.1 + FSP.2 + FSW.1 + FSW.2 + SSW.1 +
##     SSW.2 + ACE.1 + ACE.2 + DBF.1 + DBF.2 + WNR.1 + WNR.2 + UFE.1 +
##     UFE.2 + BPC.1 + BPC.2 + BPW.1 + BPW.2 + NPA.1 + NPA.2 + NPW.1 +
##     NPW.2, family = "binomial", data = french_open)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -1.447e-04  -2.100e-08  -2.100e-08   2.100e-08   1.433e-04
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -9.642e+02  3.181e+05  -0.003    0.998
## FSP.1        3.703e+00  3.207e+03   0.001    0.999
## FSP.2        9.694e+00  2.233e+03   0.004    0.997
## FSW.1        6.254e-01  1.710e+03   0.000    1.000
## FSW.2        4.050e+00  3.526e+03   0.001    0.999
## SSW.1       -1.639e+00  4.560e+03   0.000    1.000
## SSW.2        1.061e+01  2.747e+03   0.004    0.997
## ACE.1       -6.179e+00  3.022e+03  -0.002    0.998
## ACE.2       -2.165e+01  6.069e+03  -0.004    0.997
## DBF.1       -2.919e-01  3.080e+03   0.000    1.000
## DBF.2        1.849e+01  3.713e+03   0.005    0.996
## WNR.1       -1.338e+00  1.236e+03  -0.001    0.999
## WNR.2       -3.655e+00  1.616e+03  -0.002    0.998
## UFE.1       -2.401e+00  1.285e+03  -0.002    0.999
## UFE.2       -2.403e+00  8.041e+02  -0.003    0.998
## BPC.1        1.831e+02  2.175e+04   0.008    0.993
## BPC.2       -1.569e+02  1.833e+04  -0.009    0.993
## BPW.1       -2.089e+01  3.573e+03  -0.006    0.995
## BPW.2        2.151e+01  3.458e+03   0.006    0.995
## NPA.1       -1.042e+01  5.438e+03  -0.002    0.998
## NPA.2       -2.395e+01  7.039e+03  -0.003    0.997
## NPW.1        2.543e+00  2.474e+03   0.001    0.999
## NPW.2        2.324e+01  4.748e+03   0.005    0.996
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3.2330e+02  on 233  degrees of freedom
## Residual deviance: 2.1246e-07  on 211  degrees of freedom
##   (18 observations deleted due to missingness)
```

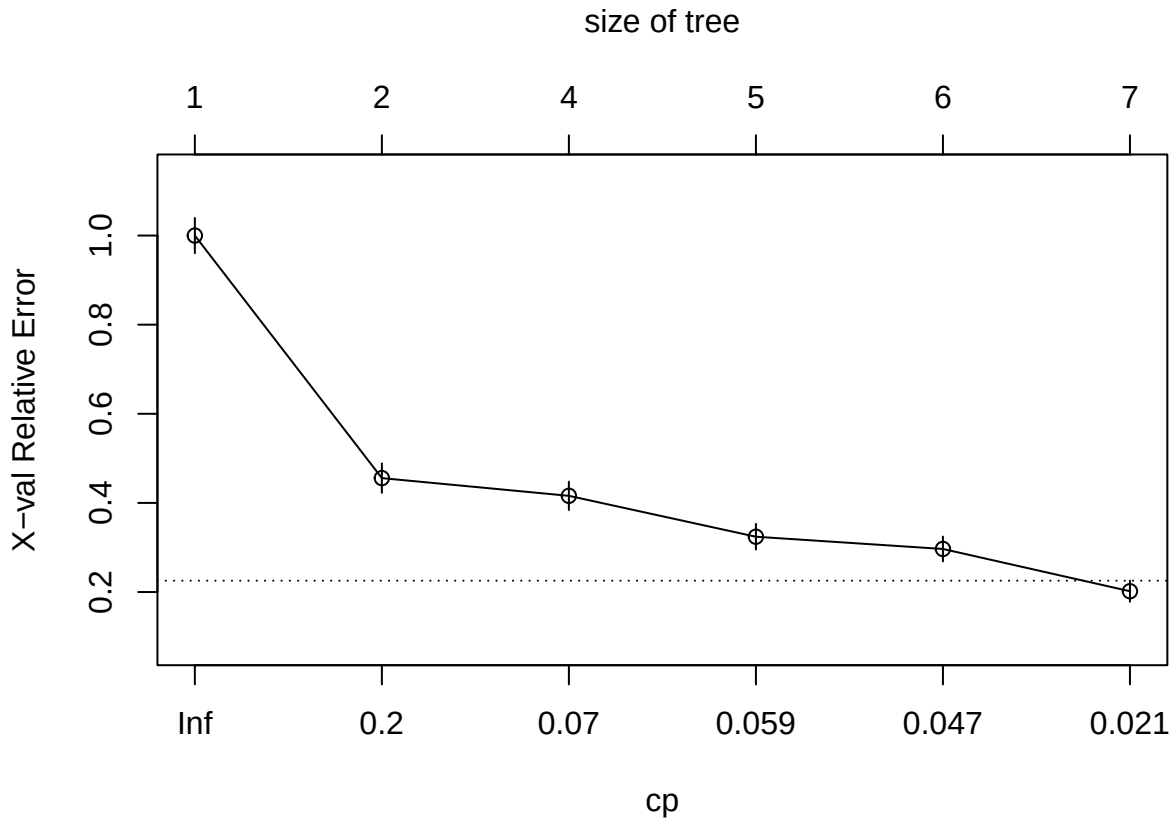
```
## AIC: 46
##
## Number of Fisher Scoring iterations: 25
```

Wimbledon

```
##
## Call:
## glm(formula = Result ~ FSP.1 + FSP.2 + FSW.1 + FSW.2 + SSW.1 +
##     SSW.2 + ACE.1 + ACE.2 + DBF.1 + DBF.2 + WNR.1 + WNR.2 + UFE.1 +
##     UFE.2 + BPC.1 + BPC.2 + BPW.1 + BPW.2 + NPA.1 + NPA.2 + NPW.1 +
##     NPW.2, family = "binomial", data = wimbledon)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.78938  -0.00777   0.00001   0.00372   2.66866
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.627847  14.275004  -0.044  0.96492
## FSP.1       -0.118397   0.217765  -0.544  0.58665
## FSP.2        0.108795   0.226943   0.479  0.63166
## FSW.1        0.289253   0.165860   1.744  0.08117 .
## FSW.2       -0.322289   0.167805  -1.921  0.05478 .
## SSW.1       -0.004789   0.211095  -0.023  0.98190
## SSW.2       -0.107042   0.251698  -0.425  0.67063
## ACE.1       -0.198079   0.198091  -1.000  0.31734
## ACE.2        0.190511   0.206792   0.921  0.35691
## DBF.1        0.302941   0.413011   0.733  0.46326
## DBF.2        0.380243   0.413005   0.921  0.35722
## WNR.1        0.084634   0.151907   0.557  0.57743
## WNR.2        0.062174   0.118029   0.527  0.59835
## UFE.1       -0.103157   0.217830  -0.474  0.63581
## UFE.2        0.028767   0.106156   0.271  0.78640
## BPC.1       -0.288525   0.337793  -0.854  0.39302
## BPC.2        0.293113   0.351949   0.833  0.40494
## BPW.1        4.009704   1.497378   2.678  0.00741 **
## BPW.2       -3.746394   1.437215  -2.607  0.00914 **
## NPA.1        0.103696   0.218237   0.475  0.63468
## NPA.2        0.136018   0.197099   0.690  0.49013
## NPW.1       -0.108786   0.334239  -0.325  0.74482
## NPW.2       -0.302885   0.323229  -0.937  0.34873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 320.195  on 230  degrees of freedom
## Residual deviance:  32.396  on 208  degrees of freedom
##   (5 observations deleted due to missingness)
## AIC: 78.396
##
## Number of Fisher Scoring iterations: 10
```

Classification Tree

```
##
## Classification tree:
## rpart(formula = Result ~ . - gender - tournament - Result, data = data_final,
##     method = "class")
##
## Variables actually used in tree construction:
## [1] BPC.1 BPC.2 BPW.1 BPW.2
##
## Root node error: 327/667 = 0.49025
##
## n= 667
##
##         CP nsplit rel error  xerror     xstd
## 1 0.544343      0   1.00000 1.00000 0.039482
## 2 0.071865      1   0.45566 0.45566 0.032896
## 3 0.067278      3   0.31193 0.41590 0.031820
## 4 0.051988      4   0.24465 0.32416 0.028875
## 5 0.042813      5   0.19266 0.29664 0.027843
## 6 0.010000      6   0.14985 0.20183 0.023583
```



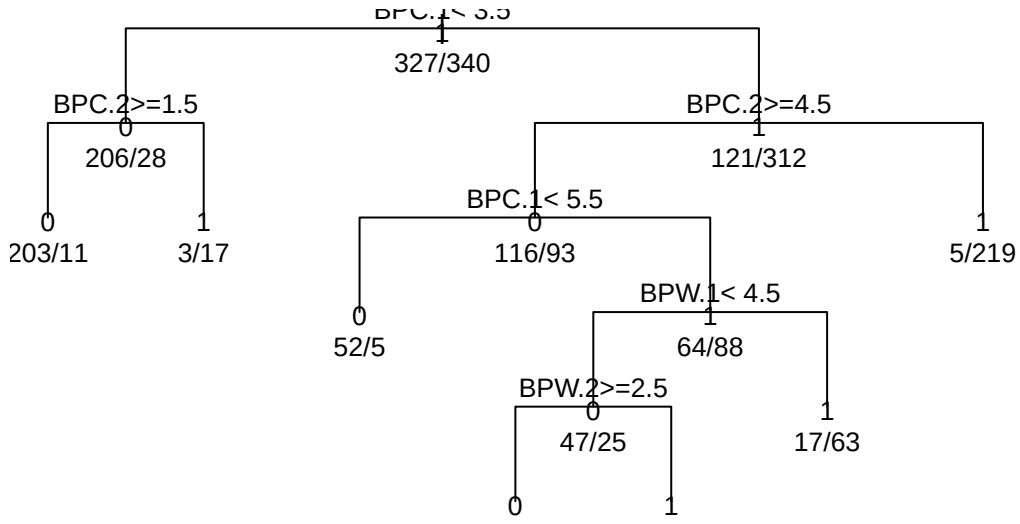

```
## Call:
## rpart(formula = Result ~ . - gender - tournament - Result, data = data_final,
##     method = "class")
##   n= 667
##
```

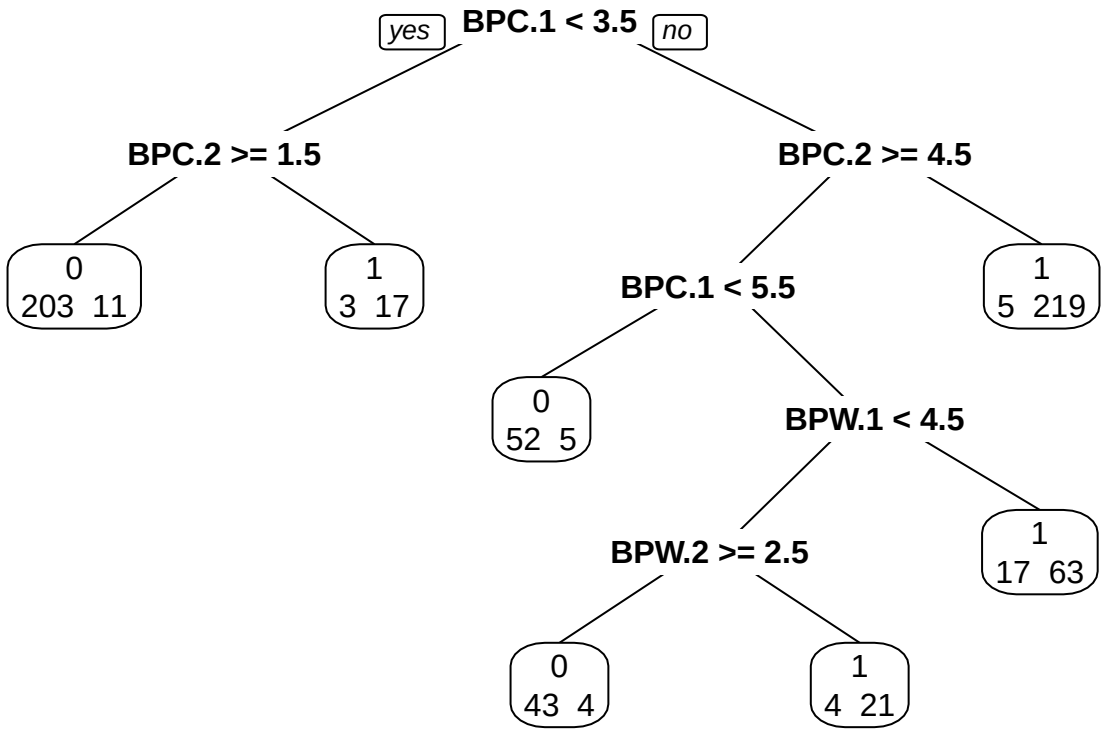
```
##           CP nsplit rel error    xerror       xstd
## 1 0.54434251      0 1.0000000 1.0000000 0.03948232
## 2 0.07186544      1 0.4556575 0.4556575 0.03289632
## 3 0.06727829      3 0.3119266 0.4159021 0.03182043
## 4 0.05198777      4 0.2446483 0.3241590 0.02887511
## 5 0.04281346      5 0.1926606 0.2966361 0.02784278
## 6 0.01000000      6 0.1498471 0.2018349 0.02358297
##
## Variable importance
## BPC.1 BPC.2 BPW.1 BPW.2 WNR.2 WNR.1 NPA.2 FSW.2 FSP.2 FSW.1 DBF.2 SSP.2 ACE.2
##    32    22    11     6     5     5     4     4     2     2     1     1     1
## NPW.1 UFE.2 ACE.1
##     1     1     1
##
## Node number 1: 667 observations,    complexity param=0.5443425
##   predicted class=1  expected loss=0.4902549  P(node) =1
##     class counts:   327   340
##    probabilities: 0.490 0.510
##   left son=2 (234 obs) right son=3 (433 obs)
##   Primary splits:
##       BPC.1 < 3.5  to the left,  improve=109.70000, (0 missing)
##       BPC.2 < 3.5  to the right, improve=100.00640, (0 missing)
##       BPW.1 < 3.5  to the left,  improve= 59.34256, (0 missing)
##       BPW.2 < 2.5  to the right, improve= 58.65435, (0 missing)
##       WNR.2 < 23.5 to the right, improve= 23.98184, (0 missing)
##   Surrogate splits:
##       WNR.1 < 17.5 to the left,  agree=0.699, adj=0.141, (0 split)
##       BPW.1 < 1.5  to the left,  agree=0.699, adj=0.141, (0 split)
##       FSW.1 < 15.5 to the left,  agree=0.679, adj=0.085, (0 split)
##       DBF.2 < 1.5  to the left,  agree=0.669, adj=0.056, (0 split)
##       FSP.2 < 77.5 to the right, agree=0.661, adj=0.034, (0 split)
##
## Node number 2: 234 observations,    complexity param=0.04281346
##   predicted class=0  expected loss=0.1196581  P(node) =0.3508246
##     class counts:   206    28
##    probabilities: 0.880 0.120
##   left son=4 (214 obs) right son=5 (20 obs)
##   Primary splits:
##       BPC.2 < 1.5  to the right, improve=23.329990, (0 missing)
##       BPW.2 < 4.5  to the right, improve= 4.491350, (0 missing)
##       BPC.1 < 2.5  to the left,  improve= 4.326923, (0 missing)
##       ACE.1 < 13.5 to the left,  improve= 4.106838, (0 missing)
##       BPW.1 < 4.5  to the left,  improve= 3.580697, (0 missing)
##   Surrogate splits:
##       BPW.2 < 0.5  to the right, agree=0.923, adj=0.10, (0 split)
##       FSP.2 < 45.5 to the right, agree=0.919, adj=0.05, (0 split)
##       SSP.2 < 54.5 to the left,  agree=0.919, adj=0.05, (0 split)
##
## Node number 3: 433 observations,    complexity param=0.07186544
##   predicted class=1  expected loss=0.2794457  P(node) =0.6491754
##     class counts:   121   312
##    probabilities: 0.279 0.721
##   left son=6 (209 obs) right son=7 (224 obs)
##   Primary splits:
```

```
##       BPC.2 < 4.5  to the right, improve=61.36290, (0 missing)
##       BPW.1 < 2.5  to the left,  improve=35.33965, (0 missing)
##       WNR.2 < 28.5 to the right, improve=19.65387, (0 missing)
##       BPW.2 < 2.5  to the right, improve=19.37897, (0 missing)
##       FSW.2 < 26.5 to the right, improve=12.88156, (0 missing)
##   Surrogate splits:
##       WNR.2 < 28.5 to the right, agree=0.704, adj=0.388, (0 split)
##       BPW.1 < 6.5  to the left,  agree=0.670, adj=0.316, (0 split)
##       NPA.2 < 15.5 to the right, agree=0.654, adj=0.282, (0 split)
##       FSW.2 < 36.5 to the right, agree=0.651, adj=0.278, (0 split)
##       BPC.1 < 6.5  to the right, agree=0.649, adj=0.273, (0 split)
##
## Node number 4: 214 observations
##   predicted class=0  expected loss=0.05140187  P(node) =0.3208396
##     class counts:   203    11
##    probabilities: 0.949 0.051
##
## Node number 5: 20 observations
##   predicted class=1  expected loss=0.15  P(node) =0.02998501
##     class counts:     3    17
##    probabilities: 0.150 0.850
##
## Node number 6: 209 observations,    complexity param=0.07186544
##   predicted class=0  expected loss=0.4449761  P(node) =0.3133433
##     class counts:   116    93
##    probabilities: 0.555 0.445
##   left son=12 (57 obs) right son=13 (152 obs)
##   Primary splits:
##       BPC.1 < 5.5  to the left,  improve=20.006380, (0 missing)
##       BPW.1 < 2.5  to the left,  improve=15.705420, (0 missing)
##       BPW.2 < 2.5  to the right, improve=15.433540, (0 missing)
##       FSW.2 < 19.5 to the right, improve= 5.124997, (0 missing)
##       BPC.2 < 7.5  to the right, improve= 4.600106, (0 missing)
##   Surrogate splits:
##       UFE.1 < 72.5 to the right, agree=0.751, adj=0.088, (0 split)
##       SSW.1 < 4    to the left,  agree=0.746, adj=0.070, (0 split)
##       BPW.1 < 1.5  to the left,  agree=0.746, adj=0.070, (0 split)
##       WNR.1 < 11.5 to the left,  agree=0.742, adj=0.053, (0 split)
##       FSP.1 < 45.5 to the left,  agree=0.737, adj=0.035, (0 split)
##
## Node number 7: 224 observations
##   predicted class=1  expected loss=0.02232143  P(node) =0.3358321
##     class counts:     5   219
##    probabilities: 0.022 0.978
##
## Node number 12: 57 observations
##   predicted class=0  expected loss=0.0877193  P(node) =0.08545727
##     class counts:    52     5
##    probabilities: 0.912 0.088
##
## Node number 13: 152 observations,    complexity param=0.06727829
##   predicted class=1  expected loss=0.4210526  P(node) =0.2278861
##     class counts:    64    88
##    probabilities: 0.421 0.579
```

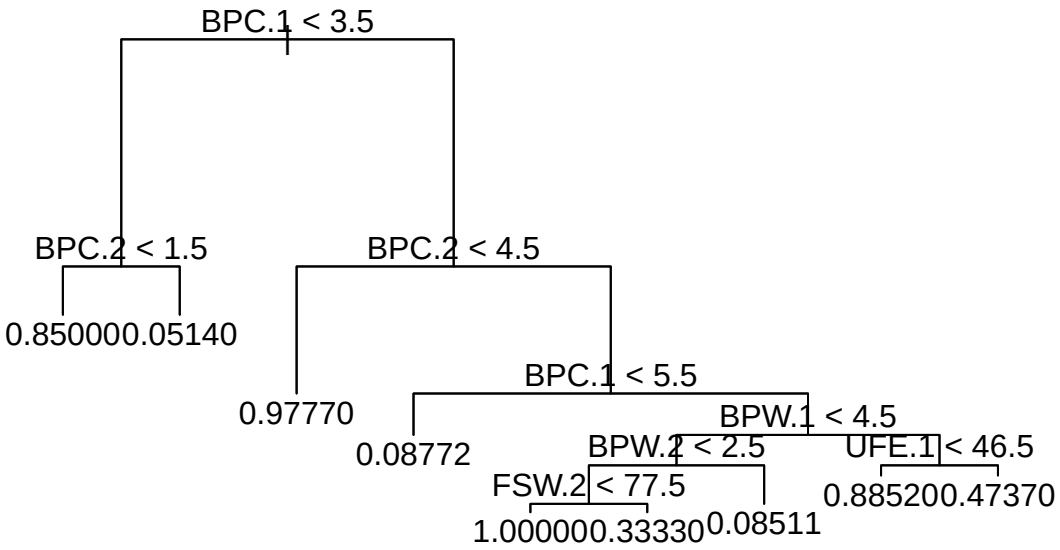
```
##   left son=26 (72 obs) right son=27 (80 obs)
##   Primary splits:
##       BPW.1 < 4.5  to the left,  improve=14.691370, (0 missing)
##       BPC.2 < 10.5 to the right, improve=12.439610, (0 missing)
##       BPW.2 < 3.5  to the right, improve=10.272980, (0 missing)
##       FSW.2 < 23.5 to the right, improve= 5.040515, (0 missing)
##       FSP.2 < 58.5 to the right, improve= 4.491228, (0 missing)
##   Surrogate splits:
##       BPW.2 < 7.5  to the left,  agree=0.717, adj=0.403, (0 split)
##       ACE.2 < 5.5  to the right, agree=0.691, adj=0.347, (0 split)
##       BPC.2 < 10.5 to the right, agree=0.664, adj=0.292, (0 split)
##       UFE.2 < 28.5 to the left,  agree=0.658, adj=0.278, (0 split)
##       ACE.1 < 12.5 to the right, agree=0.645, adj=0.250, (0 split)
##
## Node number 26: 72 observations,    complexity param=0.05198777
##   predicted class=0  expected loss=0.3472222  P(node) =0.107946
##     class counts:    47    25
##    probabilities: 0.653 0.347
##   left son=52 (47 obs) right son=53 (25 obs)
##   Primary splits:
##       BPW.2 < 2.5  to the right, improve=18.599740, (0 missing)
##       BPC.2 < 6.5  to the right, improve= 9.037904, (0 missing)
##       FSP.2 < 60.5 to the right, improve= 4.294407, (0 missing)
##       SSP.2 < 39.5 to the left,  improve= 4.294407, (0 missing)
##       DBF.1 < 2.5  to the right, improve= 4.294407, (0 missing)
##   Surrogate splits:
##       BPC.2 < 7.5  to the right, agree=0.861, adj=0.60, (0 split)
##       WNR.1 < 55.5 to the left,  agree=0.764, adj=0.32, (0 split)
##       FSP.2 < 57.5 to the right, agree=0.736, adj=0.24, (0 split)
##       SSP.2 < 41.5 to the left,  agree=0.736, adj=0.24, (0 split)
##       NPW.1 < 30.5 to the left,  agree=0.736, adj=0.24, (0 split)
##
## Node number 27: 80 observations
##   predicted class=1  expected loss=0.2125  P(node) =0.11994
##     class counts:    17    63
##    probabilities: 0.212 0.787
##
## Node number 52: 47 observations
##   predicted class=0  expected loss=0.08510638  P(node) =0.07046477
##     class counts:    43     4
##    probabilities: 0.915 0.085
##
## Node number 53: 25 observations
##   predicted class=1  expected loss=0.16  P(node) =0.03748126
##     class counts:     4    21
##    probabilities: 0.160 0.840
```

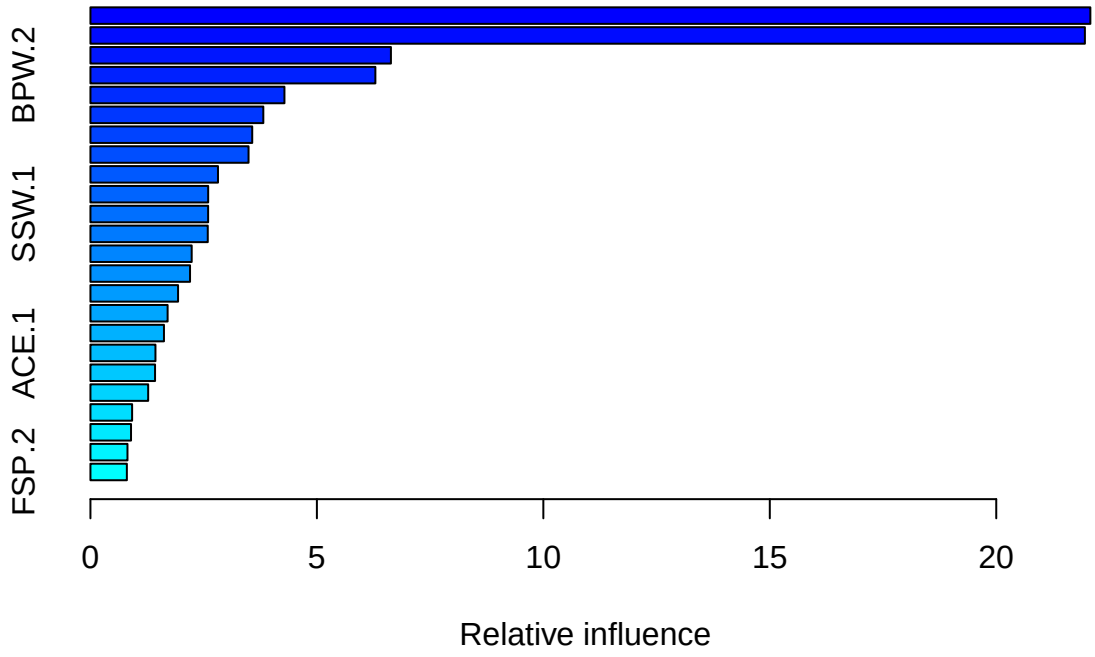
Classification Tree for Match Winner



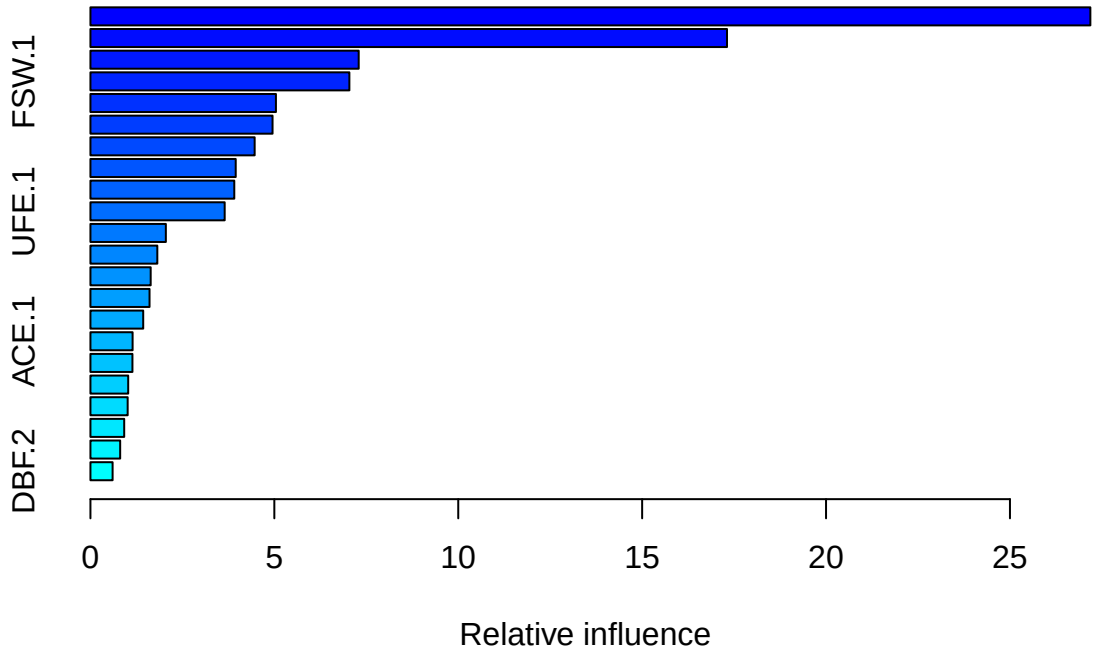



```
##
## Regression tree:
## tree(formula = Result ~ . - gender - tournament - Result, data = data_final)
## Variables actually used in tree construction:
## [1] "BPC.1" "BPC.2" "BPW.1" "BPW.2" "FSW.2" "UFE.1"
## Number of terminal nodes:  9
## Residual mean deviance:  0.0583 = 38.36 / 658
## Distribution of residuals:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.97770 -0.05140  0.00000  0.00000  0.02232  0.94860
```

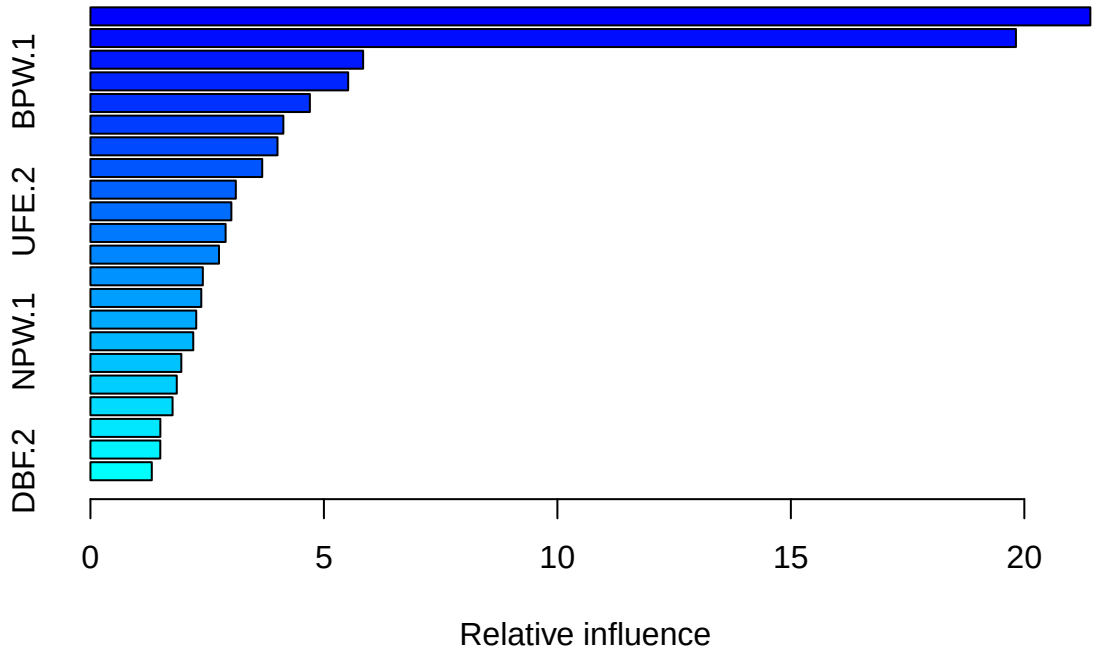

BPC.1 < 3.5
BPC.2 < 1.5
BPC.2 < 4.5
0.850000.05140
0.97770
BPC.1 < 5.5
0.08772
BPW.1 < 4.5
BPW.2 < 2.5
UFE.1 < 46.5
FSW.2 < 77.5
0.885200.47370
1.000000.33330
0.08511



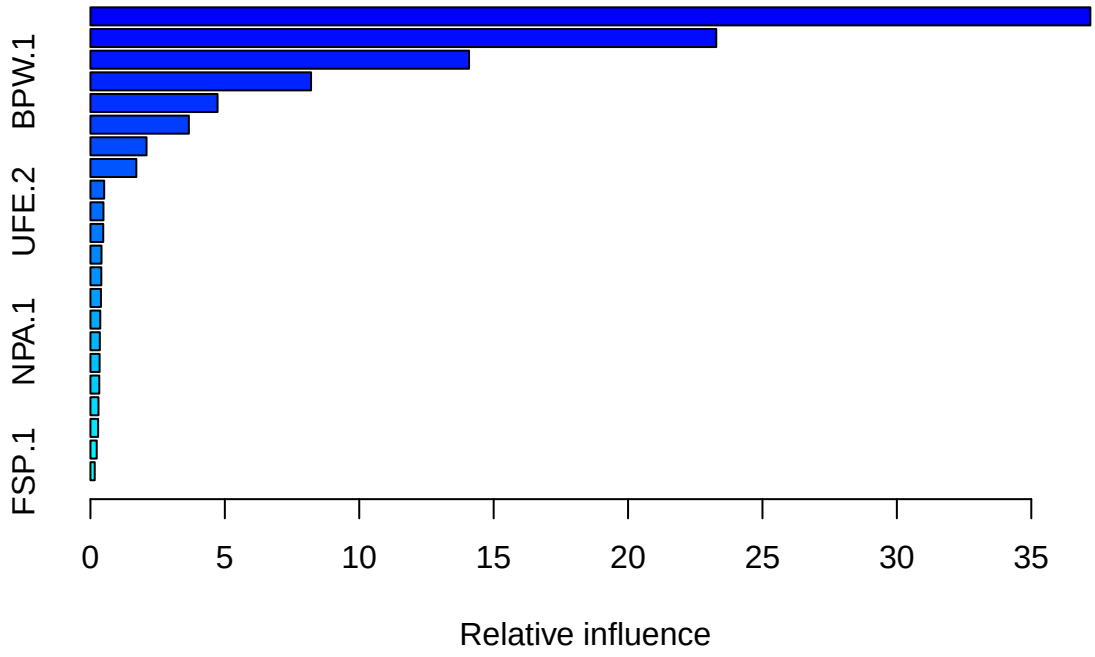

```
##         var    rel.inf
## BPC.1 BPC.1 22.0777087
## BPC.2 BPC.2 21.9571703
## BPW.1 BPW.1  6.6346823
## BPW.2 BPW.2  6.2911031
## FSW.1 FSW.1  4.2831873
## FSW.2 FSW.2  3.8169659
## UFE.2 UFE.2  3.5722966
## UFE.1 UFE.1  3.4908675
## WNR.2 WNR.2  2.8162701
## WNR.1 WNR.1  2.5996564
## SSW.1 SSW.1  2.5985702
## SSW.2 SSW.2  2.5906856
## NPA.2 NPA.2  2.2354649
## NPW.2 NPW.2  2.1990506
## ACE.2 ACE.2  1.9346380
## NPA.1 NPA.1  1.7036757
## DBF.1 DBF.1  1.6230946
## ACE.1 ACE.1  1.4354121
## NPW.1 NPW.1  1.4265412
## FSP.1 FSP.1  1.2746431
## DBF.2 DBF.2  0.9204290
## SSP.1 SSP.1  0.8958622
## SSP.2 SSP.2  0.8180098
## FSP.2 FSP.2  0.8040148
```



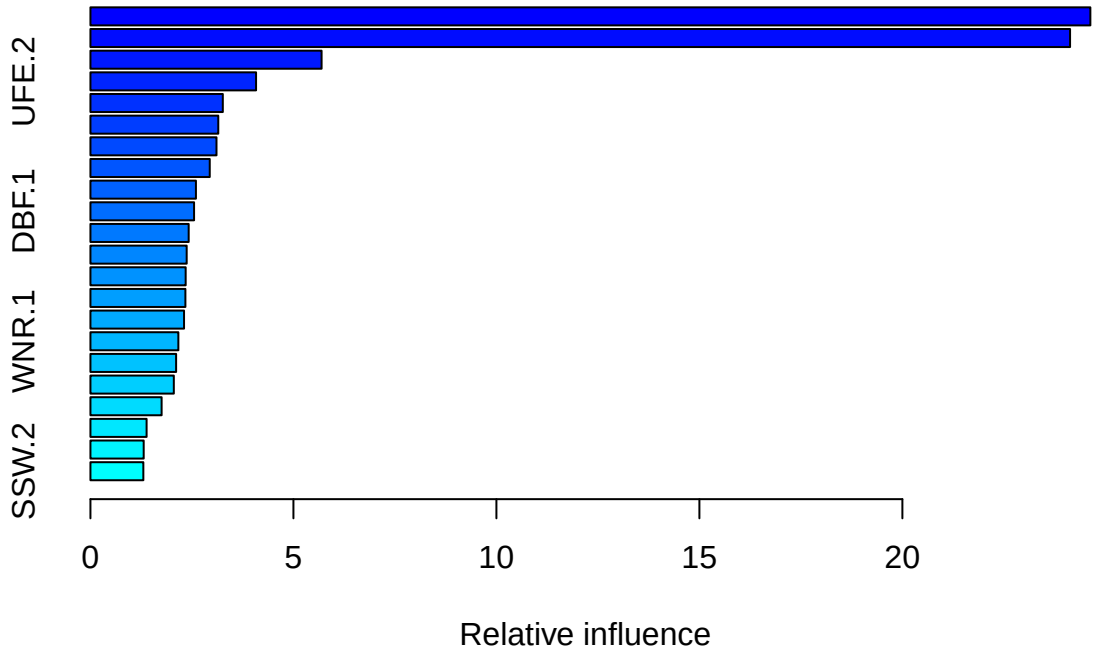

```
##         var    rel.inf
## BPC.1 BPC.1 27.1861146
## BPC.2 BPC.2 17.3072624
## BPW.1 BPW.1  7.2936826
## FSW.1 FSW.1  7.0396037
## BPW.2 BPW.2  5.0430083
## FSW.2 FSW.2  4.9512231
## SSW.1 SSW.1  4.4644381
## UFE.2 UFE.2  3.9502144
## WNR.2 WNR.2  3.9105926
## UFE.1 UFE.1  3.6500698
## SSW.2 SSW.2  2.0505981
## WNR.1 WNR.1  1.8182251
## ACE.2 ACE.2  1.6377385
## FSP.1 FSP.1  1.6069673
## NPA.2 NPA.2  1.4345412
## ACE.1 ACE.1  1.1459840
## NPA.1 NPA.1  1.1418329
## FSP.2 FSP.2  1.0266370
## NPW.1 NPW.1  1.0123413
## NPW.2 NPW.2  0.9179846
## DBF.1 DBF.1  0.8080878
## DBF.2 DBF.2  0.6028524
```



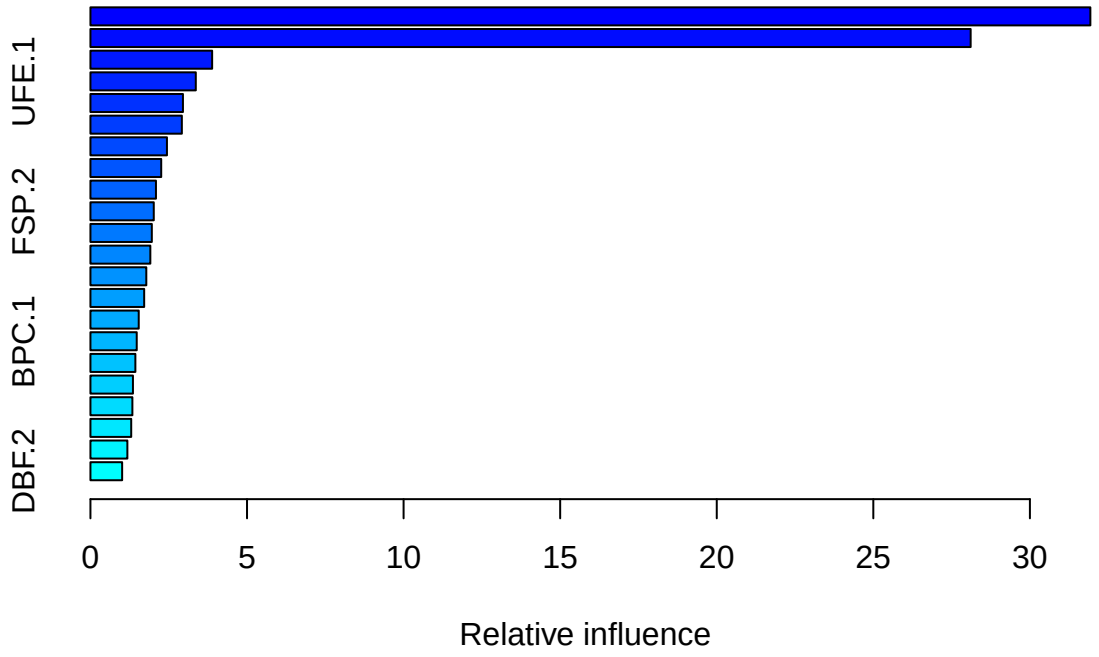

```
##         var   rel.inf
## BPC.2 BPC.2 21.413808
## BPC.1 BPC.1 19.820153
## BPW.2 BPW.2  5.839720
## BPW.1 BPW.1  5.519228
## FSW.1 FSW.1  4.699383
## FSW.2 FSW.2  4.131480
## WNR.2 WNR.2  4.006270
## WNR.1 WNR.1  3.678754
## NPA.2 NPA.2  3.113660
## UFE.2 UFE.2  3.018423
## SSW.2 SSW.2  2.894813
## UFE.1 UFE.1  2.753152
## ACE.2 ACE.2  2.407129
## SSW.1 SSW.1  2.374060
## NPW.2 NPW.2  2.265912
## NPW.1 NPW.1  2.201876
## ACE.1 ACE.1  1.946331
## FSP.2 FSP.2  1.848967
## NPA.1 NPA.1  1.759188
## DBF.1 DBF.1  1.497551
## FSP.1 FSP.1  1.495977
## DBF.2 DBF.2  1.314165
```




```
##         var    rel.inf
## BPC.2 BPC.2 37.1977006
## BPW.2 BPW.2 23.2817455
## BPC.1 BPC.1 14.0850819
## BPW.1 BPW.1  8.2072522
## DBF.2 DBF.2  4.7288018
## DBF.1 DBF.1  3.6615100
## ACE.1 ACE.1  2.0876352
## ACE.2 ACE.2  1.7069420
## UFE.1 UFE.1  0.5104477
## UFE.2 UFE.2  0.4842769
## NPW.1 NPW.1  0.4761848
## WNR.2 WNR.2  0.4112692
## FSW.2 FSW.2  0.4049195
## FSW.1 FSW.1  0.3908229
## FSP.2 FSP.2  0.3664325
## NPA.1 NPA.1  0.3492557
## WNR.1 WNR.1  0.3415011
## SSW.2 SSW.2  0.3289536
## SSW.1 SSW.1  0.2995208
## NPW.2 NPW.2  0.2854181
## NPA.2 NPA.2  0.2343649
## FSP.1 FSP.1  0.1599630
```




```
##         var   rel.inf
## BPC.1 BPC.1 24.630008
## BPC.2 BPC.2 24.129812
## BPW.1 BPW.1  5.693319
## UFE.2 UFE.2  4.079797
## ACE.1 ACE.1  3.260683
## WNR.2 WNR.2  3.148793
## UFE.1 UFE.1  3.105193
## ACE.2 ACE.2  2.938443
## FSW.2 FSW.2  2.598591
## DBF.1 DBF.1  2.552426
## FSW.1 FSW.1  2.420085
## NPW.2 NPW.2  2.370725
## NPW.1 NPW.1  2.345765
## BPW.2 BPW.2  2.339245
## FSP.1 FSP.1  2.305634
## WNR.1 WNR.1  2.165919
## NPA.1 NPA.1  2.110272
## DBF.2 DBF.2  2.054017
## FSP.2 FSP.2  1.753146
## NPA.2 NPA.2  1.383095
## SSW.1 SSW.1  1.312023
## SSW.2 SSW.2  1.303008
```




```
##         var   rel.inf
## BPW.1 BPW.1 31.930288
## BPW.2 BPW.2 28.108025
## WNR.2 WNR.2  3.886491
## UFE.1 UFE.1  3.364414
## ACE.2 ACE.2  2.954157
## FSW.1 FSW.1  2.919160
## UFE.2 UFE.2  2.443488
## FSW.2 FSW.2  2.259739
## FSP.1 FSP.1  2.091952
## FSP.2 FSP.2  2.022579
## NPA.2 NPA.2  1.960615
## BPC.2 BPC.2  1.912730
## WNR.1 WNR.1  1.784118
## SSW.2 SSW.2  1.715538
## SSW.1 SSW.1  1.542320
## BPC.1 BPC.1  1.478366
## DBF.1 DBF.1  1.434824
## ACE.1 ACE.1  1.357306
## NPW.2 NPW.2  1.339381
## NPA.1 NPA.1  1.302260
## NPW.1 NPW.1  1.179627
## DBF.2 DBF.2  1.012621
```

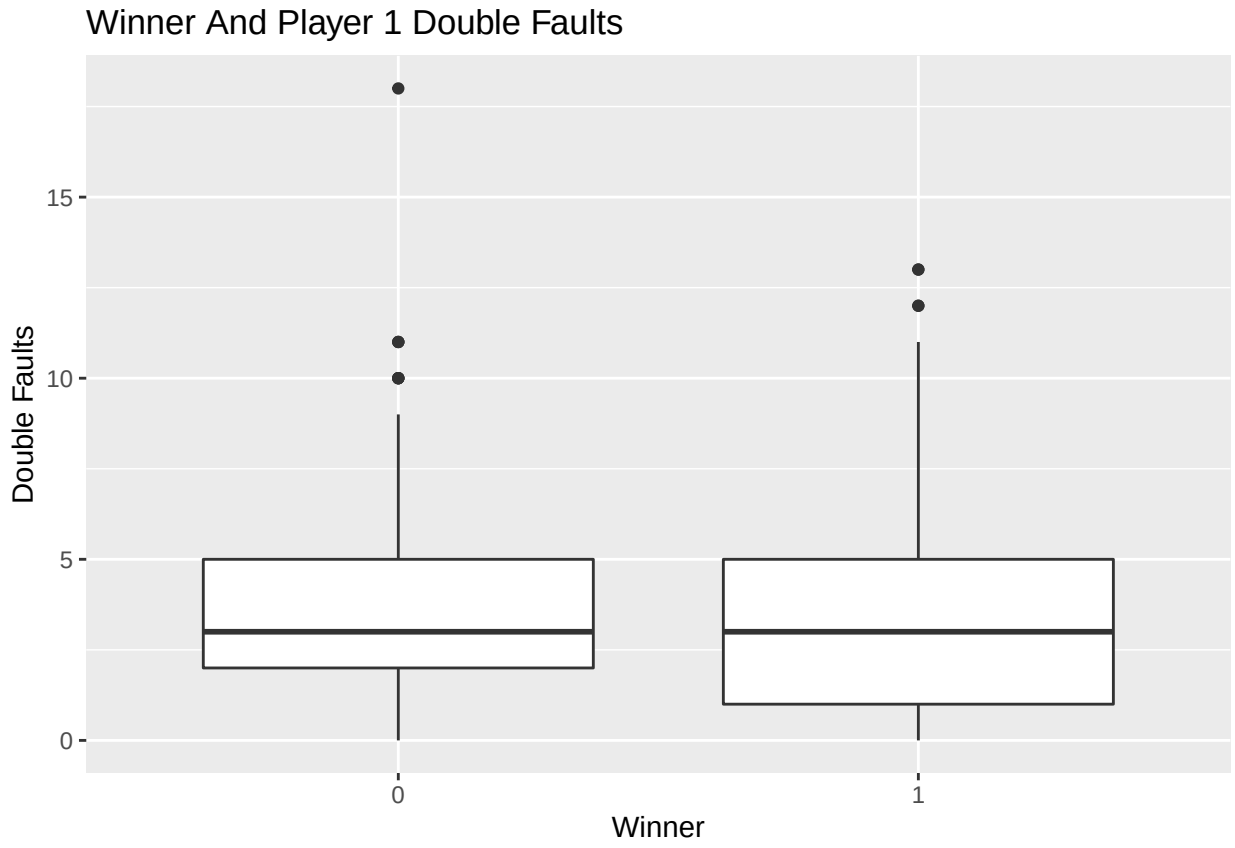
Results

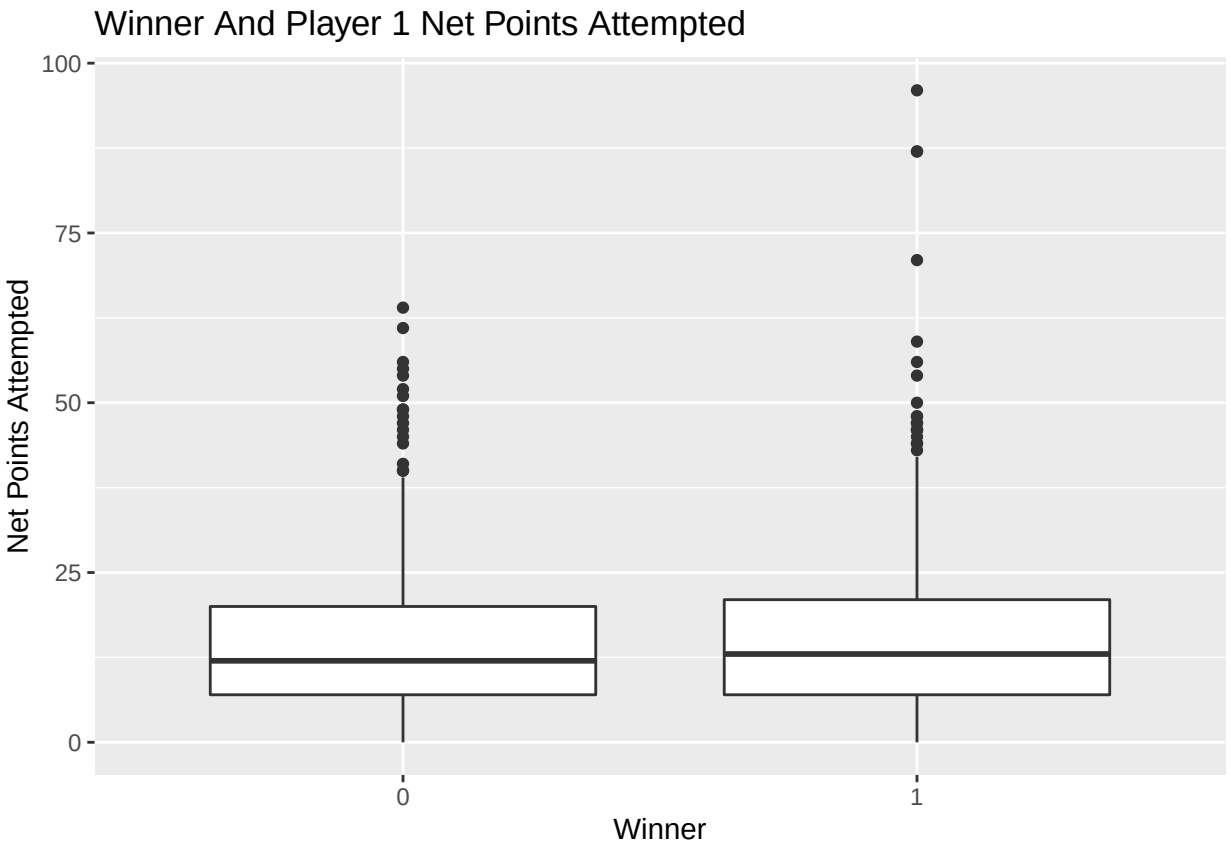
Discussion

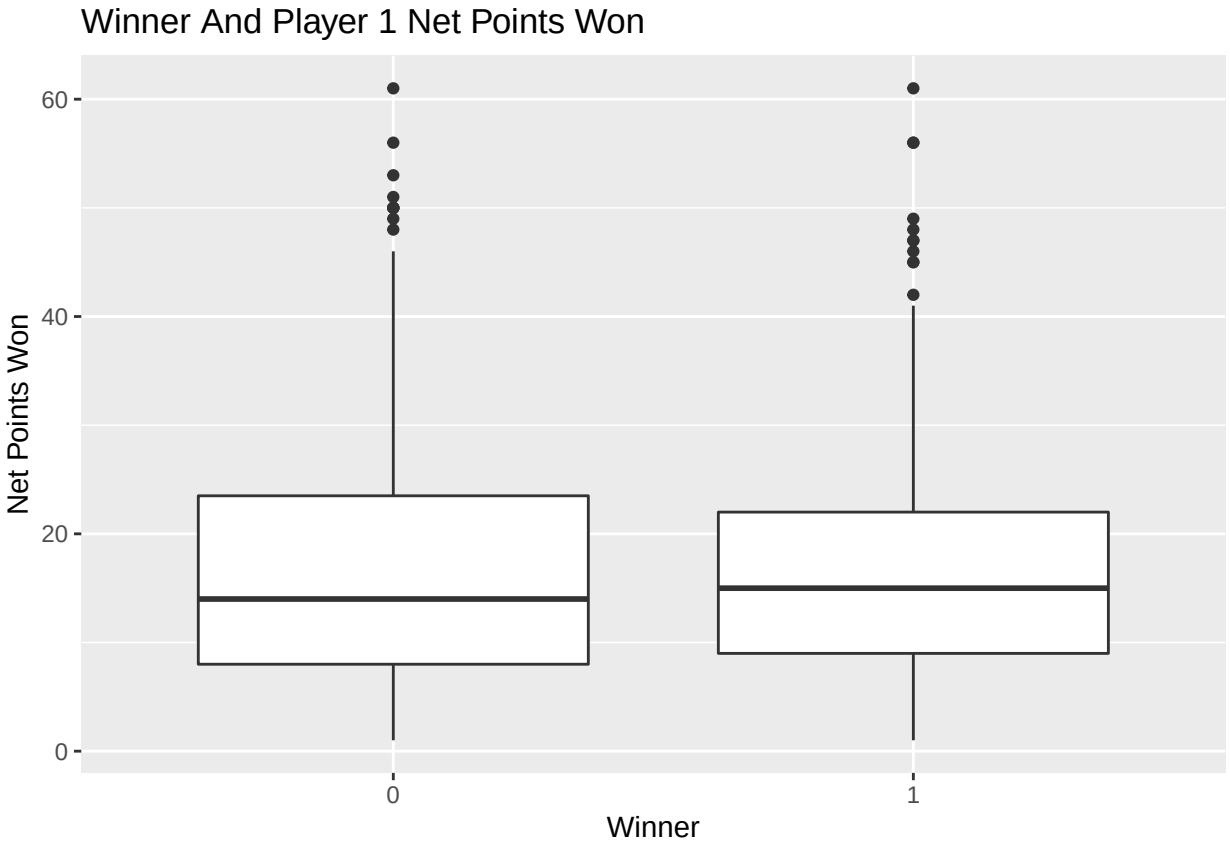
LIMITATIONS - split by gender AND surface type - not missing completely at random - not enough observations - high correlation between variables - coefficients in model not statistically significant - did not account for player demographics, skill level, ranking, etc. - from 2013, could be outdated

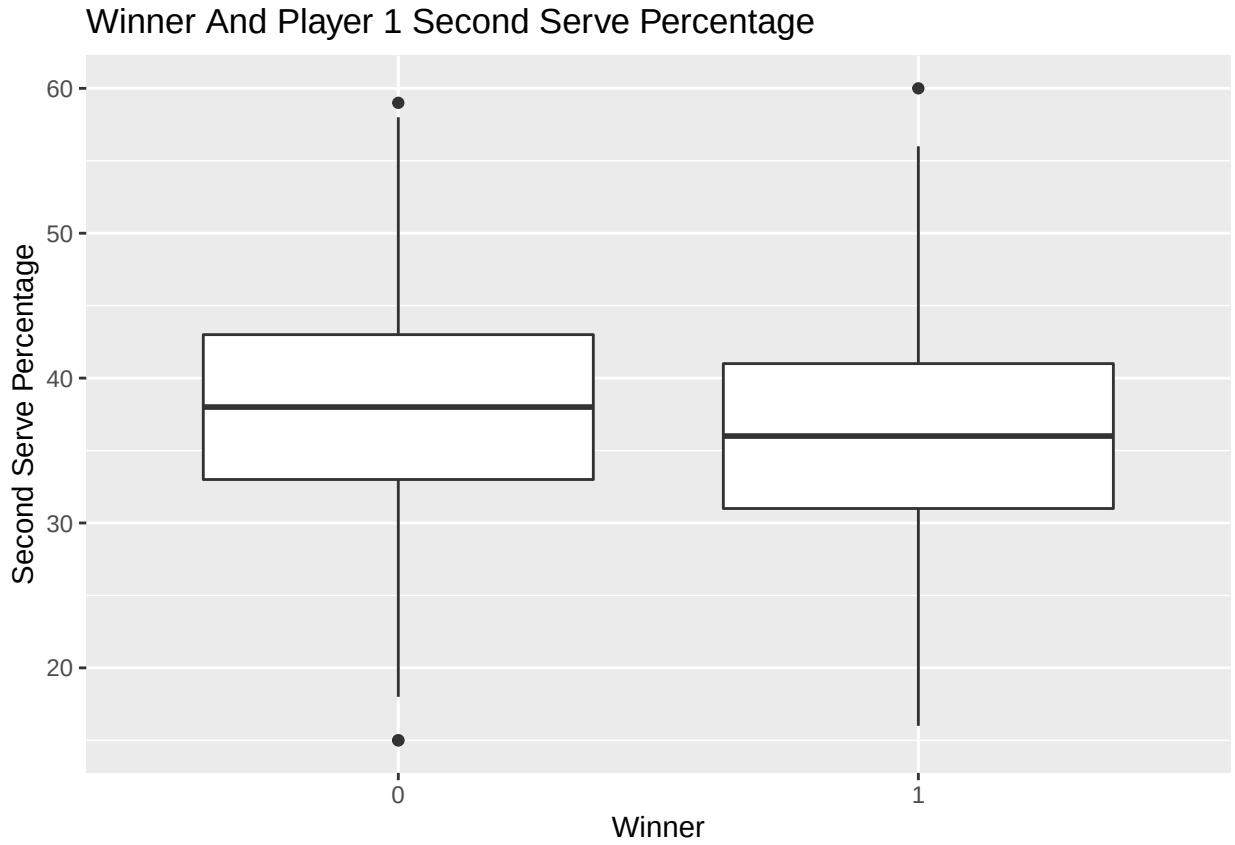
Appendix

Appendix A




Winner And Player 1 Net Points Attempted
Net Points Attempted
100
75
50
25
0
0
1
Winner


Winner And Player 1 Net Points Won
Net Points Won
60
40
20
0
0
1
Winner



Appendix B

```
##              FSP.1        FSP.2       FSW.1       FSW.2        SSP.1
## FSP.1  1.000000000  0.069236672  0.20519290  0.01732612 -1.000000000
## FSP.2  0.069236672  1.000000000  0.02647702  0.23787588 -0.069236672
## FSW.1  0.205192903  0.026477020  1.00000000  0.86550426 -0.205192903
## FSW.2  0.017326125  0.237875879  0.86550426  1.00000000 -0.017326125
## SSP.1 -1.000000000 -0.069236672 -0.20519290 -0.01732612  1.000000000
## SSP.2 -0.069236672 -1.000000000 -0.02647702 -0.23787588  0.069236672
## SSW.1 -0.358258637 -0.017989622  0.71284806  0.76677742  0.358258637
## SSW.2 -0.014847107 -0.328421436  0.77338549  0.71288529  0.014847107
## ACE.1 -0.069796998  0.007789595  0.62511821  0.56219141  0.069796998
## ACE.2 -0.007838096  0.014721986  0.57741884  0.68359337  0.007838096
## DBF.1 -0.303834348  0.029285176  0.28079521  0.30472619  0.303834348
## DBF.2 -0.087462058 -0.356327727  0.19804417  0.16709620  0.087462058
## WNR.1 -0.036118336  0.085550941  0.79674404  0.75015592  0.036118336
## WNR.2  0.065128085  0.015430079  0.73898203  0.81290964 -0.065128085
## UFE.1 -0.171009538  0.064349403  0.50234103  0.55731982  0.171009538
## UFE.2  0.013721539 -0.164998407  0.51021271  0.43922687 -0.013721539
## BPC.1  0.065141705 -0.013869657  0.23288095  0.12751960 -0.065141705
## BPC.2 -0.028616702  0.093913845  0.12175389  0.23236200  0.028616702
## BPW.1 -0.047031826 -0.124665727  0.17943862  0.08442571  0.047031826
## BPW.2 -0.118377871  0.022515349  0.13643027  0.22162957  0.118377871
## NPA.1  0.047513892  0.083971962  0.54821388  0.54062028 -0.047513892
## NPA.2  0.082835066  0.086388219  0.54896058  0.56813675 -0.082835066
## NPW.1 -0.015180338  0.053842178  0.61778334  0.60624905  0.015180338
```

```
## NPW.2  0.057183893  0.018026865  0.63850717  0.64725724 -0.057183893
##              SSP.2       SSW.1       SSW.2        ACE.1        ACE.2
## FSP.1 -0.069236672 -0.35825864 -0.01484711 -0.069796998 -0.007838096
## FSP.2 -1.000000000 -0.01798962 -0.32842144  0.007789595  0.014721986
## FSW.1 -0.026477020  0.71284806  0.77338549  0.625118209  0.577418845
## FSW.2 -0.237875879  0.76677742  0.71288529  0.562191408  0.683593370
## SSP.1  0.069236672  0.35825864  0.01484711  0.069796998  0.007838096
## SSP.2  1.000000000  0.01798962  0.32842144 -0.007789595 -0.014721986
## SSW.1  0.017989622  1.00000000  0.67262384  0.544528756  0.527116371
## SSW.2  0.328421436  0.67262384  1.00000000  0.499352980  0.521146877
## ACE.1 -0.007789595  0.54452876  0.49935298  1.000000000  0.444735528
## ACE.2 -0.014721986  0.52711637  0.52114688  0.444735528  1.000000000
## DBF.1 -0.029285176  0.35254120  0.26934857  0.208345437  0.164374713
## DBF.2  0.356327727  0.24688234  0.25369552  0.134827243  0.146705955
## WNR.1 -0.085550941  0.71979150  0.61069778  0.757264922  0.462027445
## WNR.2 -0.015430079  0.62012595  0.71093100  0.363771125  0.715952721
## UFE.1 -0.064349403  0.52327490  0.53472220  0.249984196  0.175434325
## UFE.2  0.164998407  0.51485136  0.48441792  0.137081442  0.206106819
## BPC.1  0.013869657  0.18274121  0.09809433  0.172447162  0.021665033
## BPC.2 -0.093913845  0.06589093  0.18109702 -0.088607069  0.211845597
## BPW.1  0.124665727  0.23418967  0.14164837  0.066597562 -0.132703511
## BPW.2 -0.022515349  0.18461714  0.22985206 -0.138465515 -0.006053911
## NPA.1 -0.083971962  0.44400363  0.44322360  0.424437985  0.404043426
## NPA.2 -0.086388219  0.43660236  0.45225573  0.313579394  0.428697216
## NPW.1 -0.053842178  0.55543517  0.55109280  0.410605970  0.332626518
## NPW.2 -0.018026865  0.56730001  0.58029678  0.308888526  0.411862181
##             DBF.1       DBF.2        WNR.1       WNR.2       UFE.1        UFE.2
## FSP.1 -0.30383435 -0.08746206 -0.036118336  0.06512808 -0.17100954  0.01372154
## FSP.2  0.02928518 -0.35632773  0.085550941  0.01543008  0.06434940 -0.16499841
## FSW.1  0.28079521  0.19804417  0.796744042  0.73898203  0.50234103  0.51021271
## FSW.2  0.30472619  0.16709620  0.750155916  0.81290964  0.55731982  0.43922687
## SSP.1  0.30383435  0.08746206  0.036118336 -0.06512808  0.17100954 -0.01372154
## SSP.2 -0.02928518  0.35632773 -0.085550941 -0.01543008 -0.06434940  0.16499841
## SSW.1  0.35254120  0.24688234  0.719791498  0.62012595  0.52327490  0.51485136
## SSW.2  0.26934857  0.25369552  0.610697783  0.71093100  0.53472220  0.48441792
## ACE.1  0.20834544  0.13482724  0.757264922  0.36377113  0.24998420  0.13708144
## ACE.2  0.16437471  0.14670595  0.462027445  0.71595272  0.17543432  0.20610682
## DBF.1  1.00000000  0.15343384  0.282672704  0.20103683  0.47525485  0.19096225
## DBF.2  0.15343384  1.00000000  0.119082790  0.22923944  0.16564150  0.49749179
## WNR.1  0.28267270  0.11908279  1.000000000  0.48741541  0.49624347  0.25101495
## WNR.2  0.20103683  0.22923944  0.487415407  1.00000000  0.34747388  0.50913114
## UFE.1  0.47525485  0.16564150  0.496243469  0.34747388  1.00000000  0.47163120
## UFE.2  0.19096225  0.49749179  0.251014955  0.50913114  0.47163120  1.00000000
## BPC.1  0.07490079  0.19028815  0.274979748  0.07625509 -0.07421620  0.16068369
## BPC.2  0.18562536 -0.00984233  0.005414015  0.33998218  0.17454969 -0.05404401
## BPW.1  0.10628574  0.31932085  0.254156197  0.03965773  0.32361394  0.56656054
## BPW.2  0.27278305  0.09518417  0.042757011  0.29630665  0.61687905  0.34478301
## NPA.1  0.13717369 -0.02929878  0.586876713  0.42714512  0.14152850 -0.03506010
## NPA.2  0.07628768  0.06774053  0.440125572  0.59009316  0.06004704  0.17644282
## NPW.1  0.19682970  0.02438604  0.665445378  0.48160597  0.45899576  0.18429684
## NPW.2  0.12546887  0.19403140  0.504827185  0.68988834  0.31627622  0.49203130
##             BPC.1        BPC.2       BPW.1        BPW.2       NPA.1       NPA.2
## FSP.1  0.06514170 -0.028616702 -0.04703183 -0.118377871  0.04751389  0.08283507
## FSP.2 -0.01386966  0.093913845 -0.12466573  0.022515349  0.08397196  0.08638822
```

```
## FSW.1  0.23288095  0.121753894  0.17943862  0.136430274  0.54821388  0.54896058
## FSW.2  0.12751960  0.232361998  0.08442571  0.221629573  0.54062028  0.56813675
## SSP.1 -0.06514170  0.028616702  0.04703183  0.118377871 -0.04751389 -0.08283507
## SSP.2  0.01386966 -0.093913845  0.12466573 -0.022515349 -0.08397196 -0.08638822
## SSW.1  0.18274121  0.065890931  0.23418967  0.184617143  0.44400363  0.43660236
## SSW.2  0.09809433  0.181097018  0.14164837  0.229852065  0.44322360  0.45225573
## ACE.1  0.17244716 -0.088607069  0.06659756 -0.138465515  0.42443799  0.31357939
## ACE.2  0.02166503  0.211845597 -0.13270351 -0.006053911  0.40404343  0.42869722
## DBF.1  0.07490079  0.185625357  0.10628574  0.272783046  0.13717369  0.07628768
## DBF.2  0.19028815 -0.009842330  0.31932085  0.095184166 -0.02929878  0.06774053
## WNR.1  0.27497975  0.005414015  0.25415620  0.042757011  0.58687671  0.44012557
## WNR.2  0.07625509  0.339982177  0.03965773  0.296306650  0.42714512  0.59009316
## UFE.1 -0.07421620  0.174549687  0.32361394  0.616879047  0.14152850  0.06004704
## UFE.2  0.16068369 -0.054044008  0.56656054  0.344783008 -0.03506010  0.17644282
## BPC.1  1.00000000  0.254106506  0.10615134 -0.319422025  0.42245369  0.41777464
## BPC.2  0.25410651  1.000000000 -0.32365896  0.170882959  0.34267734  0.39841732
## BPW.1  0.10615134 -0.323658962  1.00000000  0.264845802 -0.16683530 -0.16072088
## BPW.2 -0.31942202  0.170882959  0.26484580  1.000000000 -0.19312428 -0.10748119
## NPA.1  0.42245369  0.342677336 -0.16683530 -0.193124284  1.00000000  0.63332221
## NPA.2  0.41777464  0.398417315 -0.16072088 -0.107481194  0.63332221  1.00000000
## NPW.1  0.09859423  0.061294468  0.16258475  0.210302595  0.73123023  0.39429669
## NPW.2  0.11009950  0.126985395  0.23555154  0.233778847  0.37965186  0.75553637
##             NPW.1       NPW.2
## FSP.1 -0.01518034  0.05718389
## FSP.2  0.05384218  0.01802687
## FSW.1  0.61778334  0.63850717
## FSW.2  0.60624905  0.64725724
## SSP.1  0.01518034 -0.05718389
## SSP.2 -0.05384218 -0.01802687
## SSW.1  0.55543517  0.56730001
## SSW.2  0.55109280  0.58029678
## ACE.1  0.41060597  0.30888853
## ACE.2  0.33262652  0.41186218
## DBF.1  0.19682970  0.12546887
## DBF.2  0.02438604  0.19403140
## WNR.1  0.66544538  0.50482718
## WNR.2  0.48160597  0.68988834
## UFE.1  0.45899576  0.31627622
## UFE.2  0.18429684  0.49203130
## BPC.1  0.09859423  0.11009950
## BPC.2  0.06129447  0.12698540
## BPW.1  0.16258475  0.23555154
## BPW.2  0.21030259  0.23377885
## NPA.1  0.73123023  0.37965186
## NPA.2  0.39429669  0.75553637
## NPW.1  1.00000000  0.52402330
## NPW.2  0.52402330  1.00000000
```