2020 NCAA Tournament Prediction

Corey Maxedon

4/28/2020

Contents

Executive Summary

Sadly, the 2020 NCAA Men's Basketball tournament could not be held this year due to the coronavirus. I found data on the past four years and thought it would be interesting to see if I could accurately represent what could have been. Based on several variables in this dataset found in Appendix 1 and 2.1, is it possible to predict the teams that will make the tournament? Also, based on this information, can we give an estimate of the round this team will make it to? These are great questions to be answered using a variety of different classification type models.

First, we should examine the data (Appendix 2.1). There are several variables that could have a potentially high correlation. The variance inflation factors in Appendix 3.1 with extremely high correlation are effective field goal percentage of shots taken and allowed. This is nearly a direct calculation of other variables presented in the dataset. Another variable with a high inflation factor is the power rating. This rating is more or less a summary of a team based on several factors presented in the data already. The last two factors of potential concern are offensive and defensive efficiency. Model selection should be able to take care of this multicollinearity. We can check our final models with diagnostic plots. The last step is to view the correlation of our potential responses and regressors with a correlation matrix. My main variable of interest is post season wins. It seems there is high correlation between several variables and the response, Appendix 3.2. We can plot some of the variables with the highest correlation. The scatter plot in Appendix 3.3 shows the relationships between post season wins and the regressors. There does appear to be significant multicollinearity as we found before.

Before we begin setting up models, it is necessary to check if the regressors need transformations. The Box-Cox method provides significant evidence for transformations, Appendix 3.4. Testing the case with no transformations gives a p-value of less than 0.0001 so transformations are clearly needed, but even after testing the recommended transformations, the p-value was still below 0.0001. It seems another factor is at play such as the multicollinearity noted previously. We can again proceed with caution. Model selection and diagnostic checks should give us a better look at what is going on later. Besides, the recommended transformations hurt the interpretability of our results quite substantially. Since we want models that perform classification, we are unable to test the transformation of our response.

First, we build a model using all potentially useful variables, Appendix 4. We can check for outliers before we begin model selection. The halfnorm plot in Appendix 4.1 suggests observation 1329 is an outlier. Upon further inspection, this team made the tournament with a terrible record among other poor variablels. We will continue the analysis with this team since removing this team is unjustified. By reviewing the summary of our first model, we see several variables are insignificant. We will test removing the variable with the highest p-value in the summary of the model recursively until the ANOVA test provides evidence to accept the full model. The method is not exact, but it gives us a good idea of the variables and models that are most significant. The variables to be kept in the model were power rating, turnover rate, and wins above the bubble to name a few. Even the random effects were not significant with this final model (Appendix 4.2). Next, diagnostic checks were completed once more (Appendix 4.3). Observation 1329 was still an outlier, but the jump in trend is relatively insignificant. There is still some pretty high VIF results, but they are much lower than before. It seems this model does not drastically break any assumptions. Last, we can test the model's predictive ability. By looking at a ROC curve (Appendix 4.4), we find the best threshold for acceptance to balance the sensitivity and specificity appears to be 0.22. With this, our training error rate was 0.125. The testing error was 0.127 for 2019. Now, we can estimate the round at which a team will go out.

In this multinomial model, I will show a model that tries to predict all rounds (Appendix 5). Model selection was simple and chose a model with similar variables to the previous model except two point shooting percentage is now included in the final model. This model also allows for easy interpretation of variables. For example, power rating is still a big indictator for success in the tournament, but as a team progresses, other variables become more important. A look at the training error shows this model does better than the previous in predicting teams to make the tournament and even does a fair job in predicting the round a team will make it to. Another nice feature about this model is the ability to view important factors associated with each round of the tournament. The error is around 75%, but each team has 9 options to land

on (Appendix 5.2). This model did predict the champion and second place winner all four years (Appendix 5.1), but that could indicate the model is over fit on the training set. This model performed much better than the binomial model at predicting a team to make it to the tournament, but it may be useful to look at a model trained on the teams already known to make it to the tournament and see where they are predicted to make it in the tournament.

By setting up a model given a team has already made it into the tournament (Appendix 6), a summary of the most significant model shows power rating (both previous models biggest determining factor) is no longer included. The current model shows the highest magnitude predictor is now turnover rate which is somewhat interesting since it did not have a huge effect in the previous models and one does not typically view turnover rate as the statistic that wins games. Now, we see the error in the training set is still high (Appendix 6.2), but it is drastically reduced from the previous approach. The test set error reiterates this point. The model was actually even able to predict the correct champion in the test set. The next step would be to try out this line of thinking with a random forest.

This research question seems like it would be best fit for a categorical tree. We lose the quantitative inference capability about specific regressors the last models gave, but the main point of this analysis is prediction. We will compare the performance of a bagging and a random forest approach (Appendix 7). The bagging approach produced lower out of bag error as well as lower testing error per round. The bagging approach even correctly predicted the champion. From looking at importance we see both models put strong emphasis on wins above the bubble. This is the first model to do so. We will continue using the bagging model and try out a model trained on teams already in the tournament like before.

We fit a model in the same manner as the second multinomial model. The testing error is worse than the multinomial model (Appendix 8). We see the importance of wins above the bubble dropped as power rating became important once more. It will not be necessary to combine models on making the tournament versus round performance for the final prediction of the 2020 season. The multinomial model that considered all teams at once performed the best overall. This is nice since the model leaves interpretability of regressors intact. The table in Appendix 9 gives a comparison of error among the models created.

Each model has it own strengths. The binomial model was by far the least useful model in terms of error. Every other model was able to predict teams making the tournament much better. It appears the bag model performs the best on test data. It was also great at predicting teams to make the tournament. The multinomial model came in at a close second, but test error plays a big role when prediction is desired. We can try predicting the NCAA tournament for 2020 using both models as a comparison.

The final prediction of the tournament in 2020, which never happened due to the coronavirus, is given in Appendix 10. In this prediction, we will put more weight in the bag model's predictions due to the performance seen in the error table (Appendix 9). The summary of the predictions (Appendix 10) shows the bag model predicted 303 teams not making the tournament and 2 teams making it past the round of 32 and the multinomial model predicted 309 teams not making the tournament and 3 teams making it past the round of 32. We can view the teams the bag model predicted to make it to the Elite 8 which were Kansas and Gonzaga. The multinomial model also had Kansas and Gonzaga in the Elite 8 with Dayton coming out of nowhere and being the runner up in the tournament. The bag model predicted Dayton to go out in the round of 32. Last, but not least, we can see the comparison of teams in the Big Ten and their predicted round at which they lost in (Appendix 10.2). Indiana was predicted to go out in the first round in both models while, our rival, Purdue did not even make the tournament in either model.

Due to the error seen in the training and testing sets with these models, we cannot put much weight in their predictions, but the teams projected to make the tournament can almost be guaranteed. Every year sees drastic variability with the presence of "bracket busting" teams. It is difficult to identify winning teams without being able to compare specific matchups. This analysis gives a good measure of minimum performance based on overall team statistics alone. The main take away is the various impactful predictors an above average team possesses in order to make it late into the tournament such as power rating and wins above the bubble. The final prediction ability is less than desired, but this has been an interesting look on what could have been.

Methods and Results

Appendix 1. Variable Definitions

Team Information

YEAR: Season

TEAM: The Division I college basketball school

CONF: The Athletic Conference in which the school participates in

```
A10 = Atlantic 10

ACC = Atlantic Coast Conference

AE = America East

Amer = American

ASun = ASUN

B10 = Big Ten

B12 = Big 12

BE = Big East

BSky = Big Sky

BSth = Big South

BW = Big West

CAA = Colonial Athletic Association

CUSA = Conference USA

Horz = Horizon League

IND = Independent schools

Ivy = Ivy League

MAAC = Metro Atlantic Athletic Conference

MAC = Mid-American Conference

MEAC = Mid-Eastern Athletic Conference

MVC = Missouri Valley Conference

MWC = Mountain West
```

NEC = Northeast Conference

OVC = Ohio Valley Conference

P12 = Pac-12

Pat = Patriot League

SB = Sun Belt

SC = Southern Conference

SEC = South Eastern Conference

Slnd = Southland Conference

Sum = Summit League

SWAC = Southwestern Athletic Conference

WAC = Western Athletic Conference

WCC = West Coast Conference

Tournament Information

SEED: Seed in the NCAA March Madness Tournament

TRNMT: Made tournament, yes or no

PS_WINS: Post season wins in NCAA tournament

POSTSEASON: Round where the given team was eliminated or where their season ended

R68 = First Four

R64 = Round of 64

R32 = Round of 32

S16 = Sweet Sixteen

E8 = Elite Eight

F4 = Final Four

2ND = Runner-up

Champions = Winner of the NCAA March Madness Tournament for that given year

Team Statistics

G: Number of games played in total

W: Number of games won in total

BARTHAG: Power Rating (Chance of beating an average Division I team)

WAB: Wins Above Bubble (The bubble refers to the cut off between making the NCAA March Madness Tournament and not making it)

Offensive Statistics

ADJOE: Adjusted Offensive Efficiency (An estimate of the offensive efficiency (points scored per 100 possessions) a team would have against the average Division I defense)

EFG_O: Effective Field Goal Percentage Shot

TOR: Turnover Percentage Allowed (Turnover Rate)

ORB: Offensive Rebound Percentage

FTR : Free Throw Rate (How often the given team shoots Free Throws)

TWO_P_O: Two-Point Shooting Percentage

THREE_P_O: Three-Point Shooting Percentage

ADJ_T: Adjusted Tempo (An estimate of the tempo (possessions per 40 minutes) a team would have against the team that wants to play at an average Division I tempo)

Defensive Statistics

ADJDE: Adjusted Defensive Efficiency (An estimate of the defensive efficiency (points allowed per 100 possessions) a team would have against the average Division I offense)

EFG_D: Effective Field Goal Percentage Allowed

TORD: Turnover Percentage Committed (Steal Rate)

DRB: Defensive Rebound Percentage

FTRD: Free Throw Rate Allowed

TWO_P_D: Two-Point Shooting Percentage Allowed

THREE_P_D: Three-Point Shooting Percentage Allowed

Appendix 2. Data Manipulation

```
# Read Data into an Object
# https://www.kaggle.com/andrewsundberg/college-basketball-dataset/data
raw_data_15_19 = fread("cbb.csv")
raw_data_20 = fread("cbb20.csv")

# Combining Dataframes
raw_data_20 <- raw_data_20[,-c("RK")] #shows rank which isn't included in other years
raw_data_20 <- raw_data_20 %>%
                  mutate(POSTSEASON="No Tournament", #including arbitrary values so dataframes match co
                         SEED=99,
                         YEAR=2020)
raw_data <- bind_rows(raw_data_15_19, raw_data_20)

# Remove unneeded dataframes
rm(raw_data_15_19)
rm(raw_data_20)
```

```
# View Data
summary(raw_data) #notice NAs
```

```
##      TEAM               CONF                 G               W
##  Length:2110        Length:2110        Min.   :24.0   Min.   : 0.00
##  Class :character   Class :character   1st Qu.:30.0   1st Qu.:12.00
##  Mode  :character   Mode  :character   Median :31.0   Median :16.00
##                                        Mean   :31.3   Mean   :16.48
##                                        3rd Qu.:33.0   3rd Qu.:21.00
##                                        Max.   :40.0   Max.   :38.00
##
##      ADJOE            ADJDE           BARTHAG            EFG_O
##  Min.   : 76.7   Min.   : 84.0   Min.   :0.0077   Min.   :39.30
##  1st Qu.: 98.4   1st Qu.: 98.6   1st Qu.:0.2833   1st Qu.:48.00
##  Median :103.0   Median :103.3   Median :0.4746   Median :49.90
##  Mean   :103.3   Mean   :103.3   Mean   :0.4941   Mean   :50.03
##  3rd Qu.:107.9   3rd Qu.:107.8   3rd Qu.:0.7111   3rd Qu.:52.00
##  Max.   :129.1   Max.   :124.0   Max.   :0.9842   Max.   :59.80
##
##      EFG_D             TOR              TORD             ORB
##  Min.   :39.60   Min.   :12.40   Min.   :10.20   Min.   :14.20
##  1st Qu.:48.30   1st Qu.:17.30   1st Qu.:17.10   1st Qu.:26.30
##  Median :50.10   Median :18.60   Median :18.50   Median :29.10
##  Mean   :50.19   Mean   :18.65   Mean   :18.58   Mean   :29.04
##  3rd Qu.:52.10   3rd Qu.:19.90   3rd Qu.:20.00   3rd Qu.:31.80
##  Max.   :59.50   Max.   :26.60   Max.   :28.00   Max.   :42.10
##
##       DRB              FTR              FTRD            2P_O
##  Min.   :18.40   Min.   :21.60   Min.   :19.70   Min.   :37.70
##  1st Qu.:27.10   1st Qu.:31.30   1st Qu.:30.60   1st Qu.:46.90
##  Median :29.20   Median :34.60   Median :34.30   Median :49.10
##  Mean   :29.22   Mean   :34.69   Mean   :34.94   Mean   :49.19
##  3rd Qu.:31.30   3rd Qu.:38.00   3rd Qu.:38.80   3rd Qu.:51.40
```

```
##  Max.   :40.40   Max.   :51.00   Max.   :58.50   Max.   :62.60
##
##       2P_D            3P_O            3P_D          ADJ_T
##  Min.   :37.70   Min.   :24.80   Min.   :27.1   Min.   :57.2
##  1st Qu.:47.20   1st Qu.:32.40   1st Qu.:32.9   1st Qu.:66.4
##  Median :49.30   Median :34.30   Median :34.5   Median :68.5
##  Mean   :49.32   Mean   :34.33   Mean   :34.5   Mean   :68.4
##  3rd Qu.:51.60   3rd Qu.:36.20   3rd Qu.:36.1   3rd Qu.:70.3
##  Max.   :61.20   Max.   :44.10   Max.   :43.1   Max.   :83.4
##
##       WAB           POSTSEASON            SEED            YEAR
##  Min.   :-25.200   Length:2110        Min.   : 1.00   Min.   :2015
##  1st Qu.:-13.000   Class :character   1st Qu.: 9.00   1st Qu.:2016
##  Median : -8.300   Mode  :character   Median :99.00   Median :2018
##  Mean   : -7.814                      Mean   :54.74   Mean   :2018
##  3rd Qu.: -3.100                      3rd Qu.:99.00   3rd Qu.:2019
##  Max.   : 13.100                      Max.   :99.00   Max.   :2020
##                                       NA's   :1417
```

```
# Data Cleaning
raw_data$POSTSEASON[is.na(raw_data$POSTSEASON)] = "No Tournament" # removing NAs
raw_data$SEED[is.na(raw_data$SEED)] = 99
raw_data <- raw_data %>%
                mutate(TWO_P_O = `2P_O`,  # Not a good naming format for R
                       TWO_P_D = `2P_D`,
                       THREE_P_O = `3P_O`,
                       THREE_P_D = `3P_D`,
                       TRNMT = ifelse(POSTSEASON=="No Tournament", "No", "Yes")) %>%
                select(everything(), -c(`2P_O`, `2P_D`, `3P_O`, `3P_D`))
raw_data$POSTSEASON <- factor(raw_data$POSTSEASON, order = TRUE, levels = c('No Tournament', 'R68', 'R64
                                                                            'R32', 'S16', 'E8', 'F4', '
                                                                            'Champions'))
raw_data$PS_WINS = ifelse(as.numeric(raw_data$POSTSEASON) - 3 < 0, 0, as.numeric(raw_data$POSTSEASON) -

# Changing Data Types
raw_data$CONF <- as.factor(raw_data$CONF)
raw_data$TRNMT <- as.factor(raw_data$TRNMT)
raw_data$YEAR <- as.factor(raw_data$YEAR)          # make year start at zero

# Data Cleaning Done
clean_data <- raw_data
```

Appendix 2.1. View Data

```
# View Data Attributes
head(clean_data)
```

```
##             TEAM CONF  G  W ADJOE ADJDE BARTHAG EFG_O EFG_D  TOR TORD  ORB
## 1 North Carolina  ACC 40 33 123.3  94.9  0.9531  52.6  48.1 15.4 18.2 40.7
## 2      Wisconsin  B10 40 36 129.1  93.6  0.9758  54.8  47.7 12.4 15.8 32.1
## 3       Michigan  B10 40 33 114.4  90.4  0.9375  53.9  47.7 14.0 19.5 25.5
```

```
## 4     Texas Tech  B12 38 31 115.2  85.2  0.9696  53.5  43.0 17.7 22.8 27.4
## 5        Gonzaga  WCC 39 37 117.8  86.3  0.9728  56.6  41.1 16.2 17.1 30.0
## 6           Duke  ACC 39 35 125.2  90.6  0.9764  56.6  46.5 16.3 18.6 35.8
##    DRB  FTR FTRD ADJ_T  WAB POSTSEASON SEED YEAR TWO_P_O TWO_P_D THREE_P_O
## 1 30.0 32.3 30.4  71.7  8.6        2ND    1 2016    53.9    44.6      32.7
## 2 23.7 36.2 22.4  59.3 11.3        2ND    1 2015    54.8    44.7      36.5
## 3 24.9 30.7 30.0  65.9  6.9        2ND    3 2018    54.7    46.8      35.2
## 4 28.7 32.9 36.6  67.5  7.0        2ND    3 2019    52.8    41.9      36.5
## 5 26.2 39.0 26.9  71.5  7.7        2ND    1 2017    56.3    40.0      38.2
## 6 30.2 39.8 23.9  66.4 10.7  Champions    1 2015    55.9    46.3      38.7
##   THREE_P_D TRNMT PS_WINS
## 1      36.2   Yes       5
## 2      37.5   Yes       5
## 3      33.2   Yes       5
## 4      29.7   Yes       5
## 5      29.0   Yes       5
## 6      31.4   Yes       6
```

```
summary(clean_data)
```

```
##      TEAM                CONF           G              W
##  Length:2110        ACC    :  90   Min.   :24.0   Min.   : 0.00
##  Class :character   A10    :  84   1st Qu.:30.0   1st Qu.:12.00
##  Mode  :character   B10    :  84   Median :31.0   Median :16.00
##                     CUSA   :  84   Mean   :31.3   Mean   :16.48
##                     SEC    :  84   3rd Qu.:33.0   3rd Qu.:21.00
##                     Slnd   :  78   Max.   :40.0   Max.   :38.00
##                     (Other):1606
##      ADJOE           ADJDE          BARTHAG           EFG_O
##  Min.   : 76.7   Min.   : 84.0   Min.   :0.0077   Min.   :39.30
##  1st Qu.: 98.4   1st Qu.: 98.6   1st Qu.:0.2833   1st Qu.:48.00
##  Median :103.0   Median :103.3   Median :0.4746   Median :49.90
##  Mean   :103.3   Mean   :103.3   Mean   :0.4941   Mean   :50.03
##  3rd Qu.:107.9   3rd Qu.:107.8   3rd Qu.:0.7111   3rd Qu.:52.00
##  Max.   :129.1   Max.   :124.0   Max.   :0.9842   Max.   :59.80
##
##      EFG_D            TOR             TORD            ORB
##  Min.   :39.60   Min.   :12.40   Min.   :10.20   Min.   :14.20
##  1st Qu.:48.30   1st Qu.:17.30   1st Qu.:17.10   1st Qu.:26.30
##  Median :50.10   Median :18.60   Median :18.50   Median :29.10
##  Mean   :50.19   Mean   :18.65   Mean   :18.58   Mean   :29.04
##  3rd Qu.:52.10   3rd Qu.:19.90   3rd Qu.:20.00   3rd Qu.:31.80
##  Max.   :59.50   Max.   :26.60   Max.   :28.00   Max.   :42.10
##
##       DRB             FTR             FTRD           ADJ_T
##  Min.   :18.40   Min.   :21.60   Min.   :19.70   Min.   :57.2
##  1st Qu.:27.10   1st Qu.:31.30   1st Qu.:30.60   1st Qu.:66.4
##  Median :29.20   Median :34.60   Median :34.30   Median :68.5
##  Mean   :29.22   Mean   :34.69   Mean   :34.94   Mean   :68.4
##  3rd Qu.:31.30   3rd Qu.:38.00   3rd Qu.:38.80   3rd Qu.:70.3
##  Max.   :40.40   Max.   :51.00   Max.   :58.50   Max.   :83.4
##
##       WAB                  POSTSEASON        SEED          YEAR
##  Min.   :-25.200   No Tournament:1770   Min.   : 1.00   2015:351
```

```
##  1st Qu.:-13.000   R64          : 160   1st Qu.:99.00   2016:351
##  Median : -8.300   R32          :  80   Median :99.00   2017:351
##  Mean   : -7.814   S16          :  40   Mean   :84.46   2018:351
##  3rd Qu.: -3.100   R68          :  20   3rd Qu.:99.00   2019:353
##  Max.   : 13.100   E8           :  20   Max.   :99.00   2020:353
##                    (Other)      :  20
##     TWO_P_O          TWO_P_D        THREE_P_O       THREE_P_D    TRNMT
##  Min.   :37.70   Min.   :37.70   Min.   :24.80   Min.   :27.1   No :1770
##  1st Qu.:46.90   1st Qu.:47.20   1st Qu.:32.40   1st Qu.:32.9   Yes: 340
##  Median :49.10   Median :49.30   Median :34.30   Median :34.5
##  Mean   :49.19   Mean   :49.32   Mean   :34.33   Mean   :34.5
##  3rd Qu.:51.40   3rd Qu.:51.60   3rd Qu.:36.20   3rd Qu.:36.1
##  Max.   :62.60   Max.   :61.20   Max.   :44.10   Max.   :43.1
##
##     PS_WINS
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.1493
##  3rd Qu.:0.0000
##  Max.   :6.0000
##
```

```r
# Separate data for train and test sets
train_data <- clean_data[which(clean_data$YEAR==2015|clean_data$YEAR==2016|clean_data$YEAR==2017|clean_
test_data_19 <- clean_data[which(clean_data$YEAR==2019), ]
test_data_20 <- clean_data[which(clean_data$YEAR==2020), ]
rm(raw_data, clean_data)
```

Appendix 3. Check Assumptions

Appendix 3.1. Variance Inflation Factor

```
# Check multicollinearity
sort(faraway::vif(train_data[,-c(1:4, 18:20, 25:26)]))         # removing factors and potential respons
```

```
##      ADJ_T        FTR       FTRD        DRB        ORB       TORD
##   1.236197   1.299749   1.740173   2.191951   2.946805   3.098201
##        TOR        WAB      ADJDE      ADJOE  THREE_P_O    BARTHAG
##   3.105086  13.243053  22.690760  28.540690  33.348308  36.511312
##  THREE_P_D    TWO_P_O    TWO_P_D      EFG_O      EFG_D
##  43.064142  69.770015 121.052453 152.321249 224.862661
```

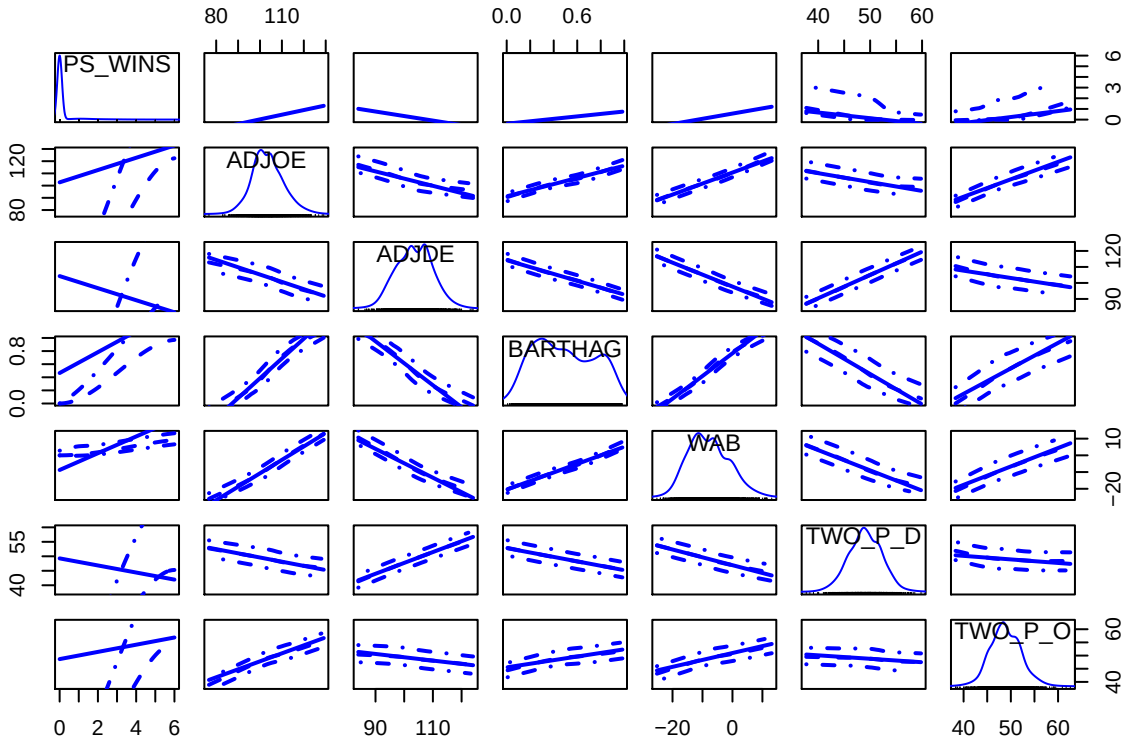
Appendix 3.2. Correlation Matrix

```
# correlation matrix
round(cor(train_data[,-c(1:4, 18:20, 25)])[,c(18)], 4)
```

```
##     ADJOE     ADJDE   BARTHAG     EFG_O     EFG_D       TOR      TORD
##    0.4683   -0.4000    0.4270    0.2836   -0.2782   -0.2472    0.0534
##       ORB       DRB       FTR      FTRD     ADJ_T       WAB   TWO_P_O
##    0.1817   -0.0908    0.0108   -0.2017   -0.0362    0.4950    0.2764
##   TWO_P_D THREE_P_O THREE_P_D   PS_WINS
##   -0.2553    0.1983   -0.1937    1.0000
```

Appendix 3.3. Scatter Plot Matrix

```
# Data Viz
car::scatterplotMatrix(~PS_WINS + ADJOE + ADJDE + BARTHAG + WAB +
                         TWO_P_D + TWO_P_O, train_data, plot.points = FALSE)
```



Appendix 3.4. Transformations

The scatter plot in Appendix 3.3 shows the relationships between post season wins and the regressors. There does appear to be pretty significant multicollinearity as we found before.

```
# Transformations
summary(bc_x <- powerTransform(cbind(ADJOE, ADJDE, BARTHAG, EFG_O, EFG_D,
                                     TOR, TORD, ORB, DRB, FTR, FTRD, ADJ_T,
                                     TWO_P_O, TWO_P_D, THREE_P_O, THREE_P_D
                                     ) ~ 1, train_data))
```

```
## bcPower Transformations to Multinormality
##           Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## ADJOE       -0.9338       -1.00      -1.1530      -0.7146
## ADJDE        1.6317        1.63       1.3267       1.9367
## BARTHAG      0.7129        0.71       0.6719       0.7539
## EFG_O        0.2583        0.33       0.0695       0.4471
## EFG_D        0.8536        1.00       0.6600       1.0472
## TOR          0.8215        1.00       0.5086       1.1344
## TORD         0.5926        0.50       0.3252       0.8601
## ORB          1.3590        1.36       1.1254       1.5927
## DRB          1.0166        1.00       0.6691       1.3641
## FTR          0.5851        0.50       0.2638       0.9065
## FTRD         0.1689        0.00      -0.0738       0.4116
## ADJ_T        0.7688        1.00       0.0218       1.5158
```

```
## TWO_P_O     -0.1324        0.00      -0.3323       0.0675
## TWO_P_D      0.6877        0.50       0.4879       0.8876
## THREE_P_O    0.9606        1.00       0.7557       1.1654
## THREE_P_D    1.0242        1.00       0.8111       1.2374
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                                                          LRT df       pval
## LR test, lambda = (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0) 2724.65 16 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                                                           LRT df       pval
## LR test, lambda = (1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1) 532.7003 16 < 2.22e-16
```

```
testTransform(bc_x, c(-1, 1.63, 0.71, 0.33, 1, 1, 0.5, 1.36, 1, 0.50, 0, 1, 0, 0.5, 1, 1))
```

```
##                                                                          LRT
## LR test, lambda = (-1 1.63 0.71 0.33 1 1 0.5 1.36 1 0.5 0 1 0 0.5 1 1) 51.98504
##                                                                           df
## LR test, lambda = (-1 1.63 0.71 0.33 1 1 0.5 1.36 1 0.5 0 1 0 0.5 1 1) 16
##                                                                         pval
## LR test, lambda = (-1 1.63 0.71 0.33 1 1 0.5 1.36 1 0.5 0 1 0 0.5 1 1) 1.1015e-05
```

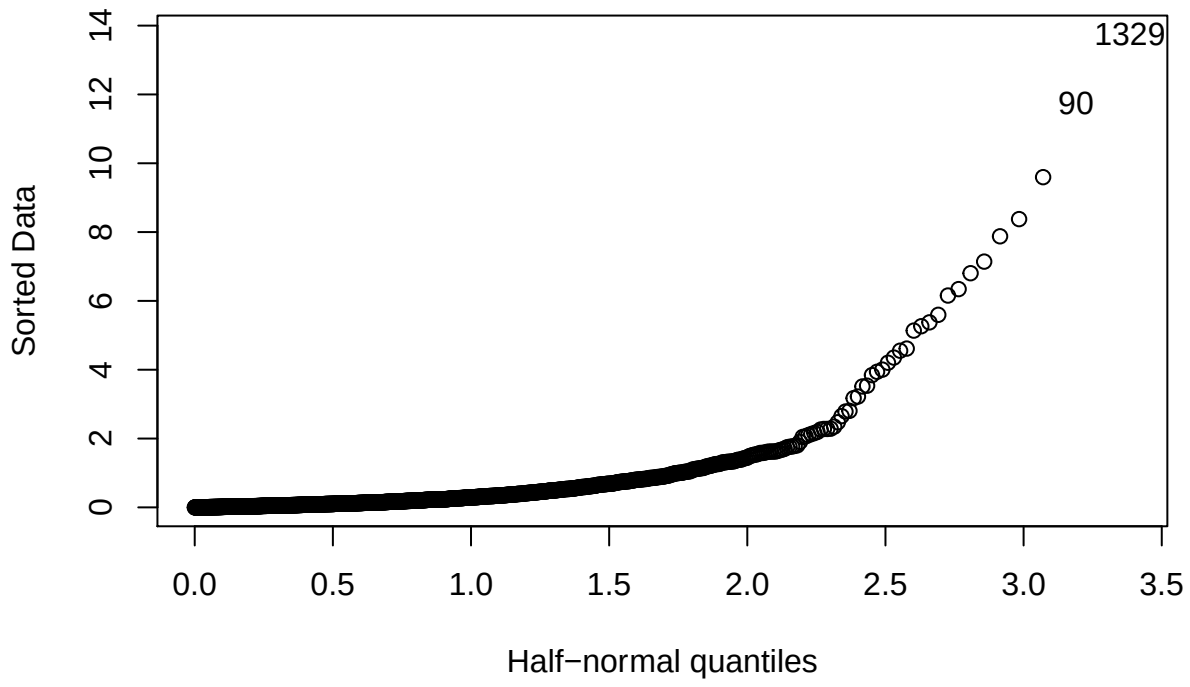
Appendix 4. Model 1: Logistic Model with Random Effects

```
model_binom <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +
                       TOR + TORD + ORB + DRB + FTR + FTRD + ADJ_T +
                       TWO_P_O + TWO_P_D + THREE_P_O + THREE_P_D +
                       WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

Appendix 4.1. Diagnostic Check

```
halfnorm(resid(model_binom, type="pearson"))
```




```
train_data[1329,]
```

```
##            TEAM CONF  G  W ADJOE ADJDE BARTHAG EFG_O EFG_D  TOR TORD  ORB
## 1662 Holy Cross  Pat 35 15  96.7 106.9  0.2398  47.9  53.2 16.8 19.6 23.1
##       DRB  FTR FTRD ADJ_T   WAB POSTSEASON SEED YEAR TWO_P_O TWO_P_D
## 1662 29.6 36.1 33.4  64.6 -14.5        R64   16 2016    47.2    52.8
##      THREE_P_O THREE_P_D TRNMT PS_WINS
## 1662      32.6      35.7   Yes       0
```

```r
summary(model_binom)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     ORB + DRB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D + THREE_P_O +
##     THREE_P_D + WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    613.1    712.8   -287.5    575.1     1385
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.7565  -0.2281  -0.0917  -0.0114  13.7424
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.641462   7.078322   2.634  0.00845 **
## ADJOE         0.524289   0.131323   3.992 6.54e-05 ***
## ADJDE        -0.643822   0.147173  -4.375 1.22e-05 ***
## BARTHAG     -31.242596   5.566700  -5.612 2.00e-08 ***
## EFG_O         0.459814   0.405906   1.133  0.25729
## EFG_D        -0.910582   0.612137  -1.488  0.13687
## TOR          -0.178962   0.105983  -1.689  0.09130 .
## TORD          0.149197   0.096552   1.545  0.12229
## ORB           0.035003   0.046874   0.747  0.45521
## DRB           0.009714   0.054970   0.177  0.85973
## FTR           0.059434   0.027141   2.190  0.02853 *
## FTRD         -0.023435   0.024953  -0.939  0.34766
## ADJ_T         0.027996   0.036494   0.767  0.44300
## TWO_P_O      -0.145603   0.253550  -0.574  0.56579
## TWO_P_D       0.535918   0.391769   1.368  0.17133
## THREE_P_O    -0.118350   0.221848  -0.533  0.59371
## THREE_P_D     0.404400   0.319731   1.265  0.20594
## WAB           0.554127   0.067668   8.189 2.64e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 18 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Appendix 4.2. Model Selection

```
# Fixed Effects Test
model_binom1 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +            # remove DRB
                      TOR + TORD + ORB + FTR + FTRD + ADJ_T +
                      TWO_P_O + TWO_P_D + THREE_P_O + THREE_P_D +
                      WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom, model_binom1)                                                  # p-value: 0.8595
```

```
## Data: train_data
## Models:
## model_binom1: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom1:     ORB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D + THREE_P_O +
## model_binom1:     THREE_P_D + WAB + (1 | CONF)
## model_binom: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom:     ORB + DRB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D + THREE_P_O +
## model_binom:     THREE_P_D + WAB + (1 | CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom1 18 611.09 705.53 -287.54   575.09
## model_binom  19 613.06 712.75 -287.53   575.06 0.0313      1     0.8595
```

```
summary(model_binom1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     ORB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D + THREE_P_O +
##     THREE_P_D + WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    611.1    705.5   -287.5    575.1     1386
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.6255  -0.2299  -0.0912  -0.0113  13.7483
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.34933    6.87232   2.670  0.00758 **
## ADJOE         0.53131    0.12562   4.229 2.34e-05 ***
## ADJDE        -0.63718    0.14207  -4.485 7.29e-06 ***
```

```
## BARTHAG     -31.27090    5.57122  -5.613 1.99e-08 ***
## EFG_O         0.45365    0.40445   1.122  0.26201
## EFG_D        -0.93198    0.59894  -1.556  0.11970
## TOR          -0.17176    0.09768  -1.758  0.07869 .
## TORD          0.16007    0.07432   2.154  0.03124 *
## ORB           0.03182    0.04323   0.736  0.46158
## FTR           0.05903    0.02705   2.182  0.02909 *
## FTRD         -0.02460    0.02406  -1.022  0.30665
## ADJ_T         0.02752    0.03641   0.756  0.44969
## TWO_P_O      -0.14786    0.25327  -0.584  0.55936
## TWO_P_D       0.54361    0.38887   1.398  0.16213
## THREE_P_O    -0.12083    0.22149  -0.546  0.58538
## THREE_P_D     0.41092    0.31715   1.296  0.19509
## WAB           0.55395    0.06763   8.191 2.59e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 17 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom2 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +            # remove THREE_P_O
                      TOR + TORD + ORB + FTR + FTRD + ADJ_T +
                      TWO_P_O + TWO_P_D + THREE_P_D +
                      WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom1, model_binom2)                                                  # p-value: 0.5832
```

```
## Data: train_data
## Models:
## model_binom2: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom2:     ORB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D + THREE_P_D +
## model_binom2:     WAB + (1 | CONF)
## model_binom1: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom1:     ORB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D + THREE_P_O +
## model_binom1:     THREE_P_D + WAB + (1 | CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom2 17 609.39 698.59 -287.69   575.39
## model_binom1 18 611.09 705.53 -287.54   575.09 0.3011      1     0.5832
```

```
summary(model_binom2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
```

```
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     ORB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D + THREE_P_D +
##     WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    609.4    698.6   -287.7    575.4     1387
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.5355  -0.2297  -0.0907  -0.0112  13.8210
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.21711    6.86923   2.652   0.0080 **
## ADJOE         0.53349    0.12579   4.241 2.22e-05 ***
## ADJDE        -0.63791    0.14232  -4.482 7.39e-06 ***
## BARTHAG     -31.29282    5.58123  -5.607 2.06e-08 ***
## EFG_O         0.24236    0.11515   2.105   0.0353 *
## EFG_D        -0.93488    0.59985  -1.559   0.1191
## TOR          -0.17005    0.09759  -1.742   0.0814 .
## TORD          0.16104    0.07416   2.171   0.0299 *
## ORB           0.02927    0.04298   0.681   0.4958
## FTR           0.05893    0.02702   2.180   0.0292 *
## FTRD         -0.02491    0.02403  -1.037   0.2999
## ADJ_T         0.02777    0.03633   0.764   0.4446
## TWO_P_O      -0.01579    0.07354  -0.215   0.8299
## TWO_P_D       0.54526    0.38953   1.400   0.1616
## THREE_P_D     0.41070    0.31764   1.293   0.1960
## WAB           0.55209    0.06744   8.186 2.69e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 16 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom3 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +            # remove TWO_P_O
                      TOR + TORD + ORB + FTR + FTRD + ADJ_T +
                      TWO_P_D + THREE_P_D +
                      WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```r
anova(model_binom2, model_binom3)                                    # p-value: 0.8298
```

```
## Data: train_data
## Models:
## model_binom3: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom3:     ORB + FTR + FTRD + ADJ_T + TWO_P_D + THREE_P_D + WAB + (1 |
## model_binom3:     CONF)
## model_binom2: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom2:     ORB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D + THREE_P_D +
## model_binom2:     WAB + (1 | CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom3 16 607.43 691.39 -287.72   575.43
## model_binom2 17 609.39 698.59 -287.69   575.39 0.0462      1     0.8298
```

```r
summary(model_binom3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     ORB + FTR + FTRD + ADJ_T + TWO_P_D + THREE_P_D + WAB + (1 |      CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    607.4    691.4   -287.7    575.4     1388
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.6171  -0.2299  -0.0913  -0.0113  13.7168
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.30302    6.86100   2.668  0.00764 **
## ADJOE         0.53646    0.12506   4.290 1.79e-05 ***
## ADJDE        -0.63877    0.14235  -4.487 7.21e-06 ***
## BARTHAG     -31.37965    5.57055  -5.633 1.77e-08 ***
## EFG_O         0.22495    0.08174   2.752  0.00592 **
## EFG_D        -0.92798    0.59860  -1.550  0.12108
## TOR          -0.17034    0.09759  -1.745  0.08091 .
## TORD          0.15949    0.07378   2.162  0.03065 *
## ORB           0.02819    0.04266   0.661  0.50866
## FTR           0.05841    0.02691   2.170  0.02998 *
## FTRD         -0.02439    0.02391  -1.020  0.30783
## ADJ_T         0.02634    0.03569   0.738  0.46052
## TWO_P_D       0.54016    0.38852   1.390  0.16443
## THREE_P_D     0.40812    0.31728   1.286  0.19834
## WAB           0.55267    0.06740   8.199 2.42e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 15 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom4 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +            # remove ORB
                        TOR + TORD + FTR + FTRD + ADJ_T +
                        TWO_P_D + THREE_P_D +
                        WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom3, model_binom4)                                                  # p-value: 0.5077
```

```
## Data: train_data
## Models:
## model_binom4: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom4:     FTR + FTRD + ADJ_T + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
## model_binom3: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom3:     ORB + FTR + FTRD + ADJ_T + TWO_P_D + THREE_P_D + WAB + (1 |
## model_binom3:     CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom4 15 605.87 684.58 -287.94   575.87
## model_binom3 16 607.43 691.39 -287.72   575.43 0.4387      1     0.5077
```

```
summary(model_binom4)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     FTR + FTRD + ADJ_T + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    605.9    684.6   -287.9    575.9     1389
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.6103  -0.2327  -0.0904  -0.0114  13.6077
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
```

```
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  17.75005    6.77933   2.618  0.00884 **
## ADJOE         0.55128    0.12296   4.484 7.34e-06 ***
## ADJDE        -0.62165    0.13924  -4.465 8.02e-06 ***
## BARTHAG     -31.29170    5.55028  -5.638 1.72e-08 ***
## EFG_O         0.19019    0.06218   3.059  0.00222 **
## EFG_D        -0.95654    0.59667  -1.603  0.10890
## TOR          -0.13285    0.07933  -1.675  0.09400 .
## TORD          0.17140    0.07141   2.400  0.01639 *
## FTR           0.05566    0.02661   2.092  0.03648 *
## FTRD         -0.02430    0.02389  -1.017  0.30895
## ADJ_T         0.02573    0.03559   0.723  0.46979
## TWO_P_D       0.54660    0.38808   1.408  0.15899
## THREE_P_D     0.41365    0.31679   1.306  0.19164
## WAB           0.56268    0.06577   8.555  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom5 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +             # remove ADJ_T
                      TOR + TORD + FTR + FTRD +
                      TWO_P_D + THREE_P_D +
                      WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom4, model_binom5)                                                  # p-value: 0.4698
```

```
## Data: train_data
## Models:
## model_binom5: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom5:     FTR + FTRD + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
## model_binom4: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom4:     FTR + FTRD + ADJ_T + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom5 14 604.40 677.85 -288.20   576.40
## model_binom4 15 605.87 684.58 -287.94   575.87 0.5223      1     0.4698
```

```
summary(model_binom5)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
```

```
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     FTR + FTRD + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    604.4    677.9   -288.2    576.4     1390
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.3533  -0.2309  -0.0920  -0.0118  12.9150
##
## Random effects:
##  Groups Name        Variance  Std.Dev.
##  CONF   (Intercept) 1.401e-17 3.743e-09
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.42403    6.72560   2.739  0.00616 **
## ADJOE         0.54672    0.12239   4.467 7.93e-06 ***
## ADJDE        -0.61378    0.13836  -4.436 9.16e-06 ***
## BARTHAG     -31.03592    5.52133  -5.621 1.90e-08 ***
## EFG_O         0.19197    0.06204   3.094  0.00197 **
## EFG_D        -0.92037    0.59287  -1.552  0.12057
## TOR          -0.13512    0.07908  -1.709  0.08750 .
## TORD          0.16674    0.07091   2.351  0.01871 *
## FTR           0.05776    0.02643   2.186  0.02884 *
## FTRD         -0.02206    0.02364  -0.933  0.35092
## TWO_P_D       0.53090    0.38631   1.374  0.16936
## THREE_P_D     0.39836    0.31527   1.264  0.20639
## WAB           0.56444    0.06571   8.590  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 13 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it

## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom6 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +            # remove FTRD
                      TOR + TORD + FTR +
                      TWO_P_D + THREE_P_D +
                      WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom5, model_binom6)                                                  # p-value: 0.3498
```

```
## Data: train_data
## Models:
## model_binom6: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom6:     FTR + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
## model_binom5: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom5:     FTR + FTRD + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom6 13 603.27 671.48 -288.63   577.27
## model_binom5 14 604.40 677.85 -288.20   576.40 0.8743      1     0.3498
```

```
summary(model_binom6)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     FTR + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    603.3    671.5   -288.6    577.3     1391
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.0055  -0.2310  -0.0925  -0.0118  13.3267
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.86651    6.70974   2.812 0.004926 **
## ADJOE         0.53968    0.12262   4.401 1.08e-05 ***
## ADJDE        -0.62169    0.13858  -4.486 7.25e-06 ***
## BARTHAG     -31.02780    5.53379  -5.607 2.06e-08 ***
## EFG_O         0.20306    0.06071   3.345 0.000823 ***
## EFG_D        -0.85127    0.58760  -1.449 0.147414
## TOR          -0.14438    0.07861  -1.837 0.066283 .
## TORD          0.13618    0.06284   2.167 0.030230 *
## FTR           0.05514    0.02624   2.102 0.035568 *
## TWO_P_D       0.49529    0.38412   1.289 0.197254
## THREE_P_D     0.36740    0.31325   1.173 0.240838
## WAB           0.57073    0.06531   8.739  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) ADJOE  ADJDE  BARTHA EFG_O  EFG_D  TOR    TORD   FTR
## ADJOE      0.422
## ADJDE     -0.701 -0.877
## BARTHAG   -0.620 -0.938  0.942
## EFG_O      0.215  0.047 -0.313 -0.201
```

```
## EFG_D      0.123  0.029 -0.073 -0.048  0.013
## TOR       -0.487 -0.064  0.248  0.165 -0.328 -0.054
## TORD      -0.406  0.119  0.114 -0.007 -0.050 -0.189  0.052
## FTR        0.056  0.132 -0.232 -0.155  0.302  0.120 -0.361 -0.087
## TWO_P_D   -0.108 -0.049  0.056  0.055 -0.014 -0.990  0.057  0.130 -0.103
## THREE_P_D -0.125 -0.060  0.066  0.060  0.001 -0.981  0.048  0.151 -0.114
## WAB        0.105 -0.094  0.063 -0.122 -0.093 -0.025  0.184 -0.042 -0.163
##           TWO_P_ THREE_
## ADJOE
## ADJDE
## BARTHAG
## EFG_O
## EFG_D
## TOR
## TORD
## FTR
## TWO_P_D
## THREE_P_D  0.976
## WAB        0.040  0.050
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom7 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +            # remove THREE_P_D
                      TOR + TORD + FTR +
                      TWO_P_D +
                      WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom6, model_binom7)                                                  # p-value: 0.2336
```

```
## Data: train_data
## Models:
## model_binom7: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom7:     FTR + TWO_P_D + WAB + (1 | CONF)
## model_binom6: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom6:     FTR + TWO_P_D + THREE_P_D + WAB + (1 | CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom7 12 602.69 665.65 -289.34   578.69
## model_binom6 13 603.27 671.48 -288.63   577.27 1.4189      1     0.2336
```

```
summary(model_binom7)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     FTR + TWO_P_D + WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    602.7    665.7   -289.3    578.7     1392
```

```
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.1463  -0.2323  -0.0930  -0.0119  12.7045
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  19.91828    6.64579   2.997 0.002725 **
## ADJOE         0.54984    0.12286   4.475 7.63e-06 ***
## ADJDE        -0.63418    0.13858  -4.576 4.73e-06 ***
## BARTHAG     -31.50604    5.54004  -5.687 1.29e-08 ***
## EFG_O         0.20360    0.06060   3.360 0.000779 ***
## EFG_D        -0.17737    0.11311  -1.568 0.116839
## TOR          -0.14911    0.07833  -1.904 0.056949 .
## TORD          0.12484    0.06171   2.023 0.043084 *
## FTR           0.05872    0.02608   2.251 0.024373 *
## TWO_P_D       0.05760    0.08315   0.693 0.488483
## WAB           0.56843    0.06538   8.694  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) ADJOE  ADJDE  BARTHA EFG_O  EFG_D  TOR    TORD   FTR
## ADJOE    0.418
## ADJDE   -0.699 -0.877
## BARTHAG -0.617 -0.938  0.942
## EFG_O    0.216  0.049 -0.315 -0.202
## EFG_D    0.005 -0.157 -0.045  0.059  0.074
## TOR     -0.486 -0.063  0.246  0.163 -0.327 -0.039
## TORD    -0.391  0.135  0.100 -0.022 -0.051 -0.218  0.048
## FTR      0.040  0.127 -0.225 -0.148  0.305  0.045 -0.359 -0.067
## TWO_P_D  0.061  0.047 -0.038 -0.019 -0.071 -0.760  0.049 -0.075  0.032
## WAB      0.115 -0.089  0.059 -0.127 -0.091  0.122  0.181 -0.052 -0.162
##         TWO_P_
## ADJOE
## ADJDE
## BARTHAG
## EFG_O
## EFG_D
## TOR
## TORD
## FTR
## TWO_P_D
## WAB     -0.042
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom8 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D +             # remove TWO_P_D
                      TOR + TORD + FTR +
```

```
                        WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom7, model_binom8)                                    # p-value: 0.488
```

```
## Data: train_data
## Models:
## model_binom8: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom8:     FTR + WAB + (1 | CONF)
## model_binom7: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom7:     FTR + TWO_P_D + WAB + (1 | CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom8 11 601.17 658.89 -289.58   579.17
## model_binom7 12 602.69 665.65 -289.34   578.69 0.4809      1      0.488
```

```
summary(model_binom8)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
##     FTR + WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    601.2    658.9   -289.6    579.2     1393
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -11.3517  -0.2339  -0.0917  -0.0122  12.8717
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  19.66107    6.62576   2.967  0.00300 **
## ADJOE         0.54657    0.12276   4.452 8.49e-06 ***
## ADJDE        -0.63132    0.13839  -4.562 5.07e-06 ***
## BARTHAG     -31.46860    5.53863  -5.682 1.33e-08 ***
## EFG_O         0.20675    0.06041   3.423  0.00062 ***
## EFG_D        -0.11798    0.07350  -1.605  0.10849
## TOR          -0.15195    0.07826  -1.942  0.05220 .
## TORD          0.12805    0.06147   2.083  0.03723 *
## FTR           0.05821    0.02604   2.235  0.02542 *
## WAB           0.57075    0.06525   8.747  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##         (Intr) ADJOE  ADJDE  BARTHA EFG_O  EFG_D  TOR    TORD   FTR
## ADJOE    0.416
## ADJDE   -0.697 -0.877
## BARTHAG -0.616 -0.939  0.942
## EFG_O    0.220  0.050 -0.316 -0.202
## EFG_D    0.076 -0.187 -0.113  0.069  0.033
## TOR     -0.490 -0.062  0.245  0.161 -0.324 -0.001
## TORD    -0.386  0.142  0.094 -0.027 -0.057 -0.424  0.053
## FTR      0.034  0.123 -0.222 -0.144  0.309  0.108 -0.358 -0.064
## WAB      0.116 -0.085  0.056 -0.129 -0.094  0.138  0.186 -0.055 -0.162
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom9 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O +                    # remove EFG_D
                        TOR + TORD + FTR +
                        WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom8, model_binom9)                                                  # p-value: 0.108
```

```
## Data: train_data
## Models:
## model_binom9: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR + TORD + FTR +
## model_binom9:     WAB + (1 | CONF)
## model_binom8: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR + TORD +
## model_binom8:     FTR + WAB + (1 | CONF)
##              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom9 10 601.75 654.22 -290.88   581.75
## model_binom8 11 601.17 658.89 -289.58   579.17 2.5831      1      0.108
```

```
summary(model_binom9)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR + TORD + FTR +
##     WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    601.8    654.2   -290.9    581.8     1394
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -10.1205  -0.2348  -0.0951  -0.0122  11.4464
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
```

```
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  20.55090    6.54723   3.139 0.001696 **
## ADJOE         0.51194    0.11968   4.277 1.89e-05 ***
## ADJDE        -0.65937    0.13594  -4.850 1.23e-06 ***
## BARTHAG     -30.98621    5.47869  -5.656 1.55e-08 ***
## EFG_O         0.21105    0.06016   3.508 0.000451 ***
## TOR          -0.15231    0.07781  -1.957 0.050299 .
## TORD          0.08614    0.05538   1.555 0.119845
## FTR           0.06296    0.02584   2.436 0.014846 *
## WAB           0.58789    0.06464   9.095  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) ADJOE  ADJDE  BARTHA EFG_O  TOR    TORD   FTR
## ADJOE    0.430
## ADJDE   -0.689 -0.919
## BARTHAG -0.618 -0.944  0.957
## EFG_O    0.212  0.050 -0.309 -0.196
## TOR     -0.487 -0.053  0.237  0.151 -0.320
## TORD    -0.391  0.075  0.049  0.001 -0.053  0.056
## FTR      0.025  0.145 -0.212 -0.151  0.310 -0.360 -0.021
## WAB      0.113 -0.055  0.069 -0.148 -0.107  0.188  0.003 -0.183
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom10 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O +                    # remove TORD
                       TOR + FTR +
                       WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom9, model_binom10)                                                  # p-value: 0.1183
```

```
## Data: train_data
## Models:
## model_binom10: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR + FTR + WAB + (1 |
## model_binom10:     CONF)
## model_binom9: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR + TORD + FTR +
## model_binom9:     WAB + (1 | CONF)
##               Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom10  9 602.19 649.42 -292.10   584.19
## model_binom9  10 601.75 654.22 -290.88   581.75 2.4394      1     0.1183
```

```
summary(model_binom10)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
```

```
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR + FTR + WAB + (1 |
##     CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    602.2    649.4   -292.1    584.2     1395
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -9.2942 -0.2370 -0.0964 -0.0121 11.5546
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  24.63257    5.99824   4.107 4.01e-05 ***
## ADJOE         0.50122    0.11901   4.212 2.54e-05 ***
## ADJDE        -0.67338    0.13582  -4.958 7.12e-07 ***
## BARTHAG     -31.15990    5.46624  -5.700 1.20e-08 ***
## EFG_O         0.21664    0.06000   3.611 0.000305 ***
## TOR          -0.15942    0.07783  -2.048 0.040528 *
## FTR           0.06424    0.02585   2.485 0.012963 *
## WAB           0.59009    0.06424   9.186  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) ADJOE  ADJDE  BARTHA EFG_O  TOR    FTR
## ADJOE    0.502
## ADJDE   -0.731 -0.926
## BARTHAG -0.672 -0.947  0.958
## EFG_O    0.224  0.063 -0.319 -0.208
## TOR     -0.503 -0.055  0.234  0.148 -0.324
## FTR      0.025  0.150 -0.215 -0.154  0.308 -0.364
## WAB      0.115 -0.060  0.074 -0.142 -0.104  0.192 -0.186
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
model_binom11 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + FTR +                # remove TOR
                       WAB + (1|CONF), nAGQ = 25, family = "binomial", train_data)
```

```
## boundary (singular) fit: see ?isSingular
```

```
anova(model_binom10, model_binom11)                                                    # p-value: 0.038
```

```
## Data: train_data
## Models:
## model_binom11: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + FTR + WAB + (1 | CONF)
## model_binom10: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR + FTR + WAB + (1 |
## model_binom10:     CONF)
```

```
##               Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_binom11  8 604.46 646.44 -294.23   588.46
## model_binom10  9 602.19 649.42 -292.10   584.19 4.2679      1    0.03884 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(model_binom11)
```

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 25) [glmerMod]
##  Family: binomial  ( logit )
## Formula: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + FTR + WAB + (1 | CONF)
##    Data: train_data
##
##      AIC      BIC   logLik deviance df.resid
##    604.5    646.4   -294.2    588.5     1396
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.6476 -0.2347 -0.0981 -0.0134 14.6925
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  CONF   (Intercept) 0        0
## Number of obs: 1404, groups:  CONF, 33
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  18.63127    5.13393   3.629 0.000284 ***
## ADJOE         0.49358    0.11701   4.218 2.46e-05 ***
## ADJDE        -0.61506    0.12965  -4.744 2.10e-06 ***
## BARTHAG     -29.79642    5.32158  -5.599 2.15e-08 ***
## EFG_O         0.17765    0.05667   3.135 0.001720 **
## FTR           0.04517    0.02392   1.888 0.058973 .
## WAB           0.61936    0.06310   9.815  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##         (Intr) ADJOE  ADJDE  BARTHA EFG_O  FTR
## ADJOE    0.538
## ADJDE   -0.724 -0.939
## BARTHAG -0.692 -0.949  0.960
## EFG_O    0.068  0.040 -0.260 -0.164
## FTR     -0.210  0.136 -0.137 -0.104  0.215
## WAB      0.259 -0.047  0.026 -0.182 -0.049 -0.125
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

```
# Random Effects
model_binom12 <- glmer(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O +                    # REML = F
                       TOR + FTR +
                       WAB + (1|CONF), REML = FALSE, family = "binomial", train_data)
```

```
## Warning: extra argument(s) 'REML' disregarded
```

```
## boundary (singular) fit: see ?isSingular
```

```
model_binom13 <- glm(TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O +                    # remove CONF
                       TOR + FTR +
                       WAB, family = "binomial", train_data)
anova(model_binom12, model_binom13)                                               # p-value: 1
```

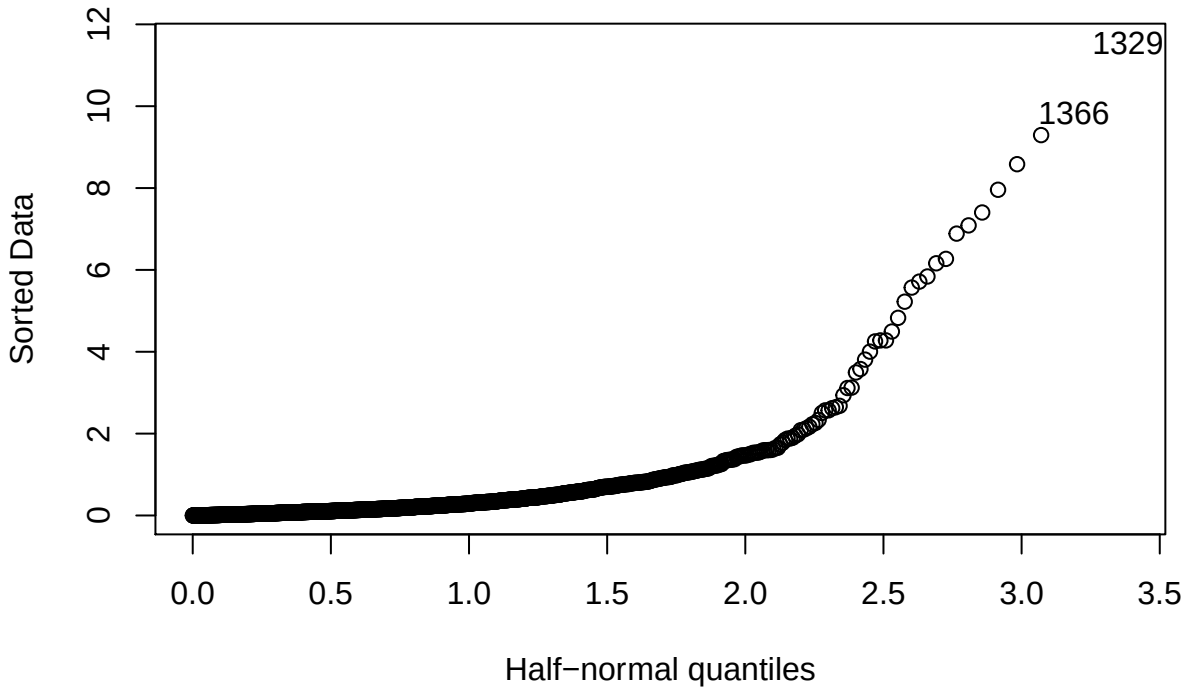
```
## Data: train_data
## Models:
## model_binom13: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR + FTR + WAB
## model_binom12: TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR + FTR + WAB + (1 |
## model_binom12:     CONF)
##               Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## model_binom13  8 600.19 642.17 -292.1   584.19
## model_binom12  9 602.19 649.42 -292.1   584.19     0      1          1
```

```
summary(model_binom13)
```

```
##
## Call:
## glm(formula = TRNMT ~ ADJOE + ADJDE + BARTHAG + EFG_O + TOR +
##     FTR + WAB, family = "binomial", data = train_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.99008  -0.33058  -0.13608  -0.01713   3.13102
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  24.63257    5.95394   4.137 3.52e-05 ***
## ADJOE         0.50122    0.11747   4.267 1.98e-05 ***
## ADJDE        -0.67338    0.13383  -5.032 4.86e-07 ***
## BARTHAG     -31.15990    5.39211  -5.779 7.52e-09 ***
## EFG_O         0.21664    0.05992   3.616   0.0003 ***
## TOR          -0.15942    0.07778  -2.050   0.0404 *
## FTR           0.06424    0.02584   2.486   0.0129 *
## WAB           0.59009    0.06423   9.187  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1380.38  on 1403  degrees of freedom
## Residual deviance:  584.19  on 1396  degrees of freedom
## AIC: 600.19
##
## Number of Fisher Scoring iterations: 7
```

Appendix 4.3. Post Diagnostic Checks

```
# Check Diagnostics
halfnorm(resid(model_binom13, type="pearson"))
```




```
sort(faraway::vif(train_data[,c(5:8, 10, 14, 17)]))
```

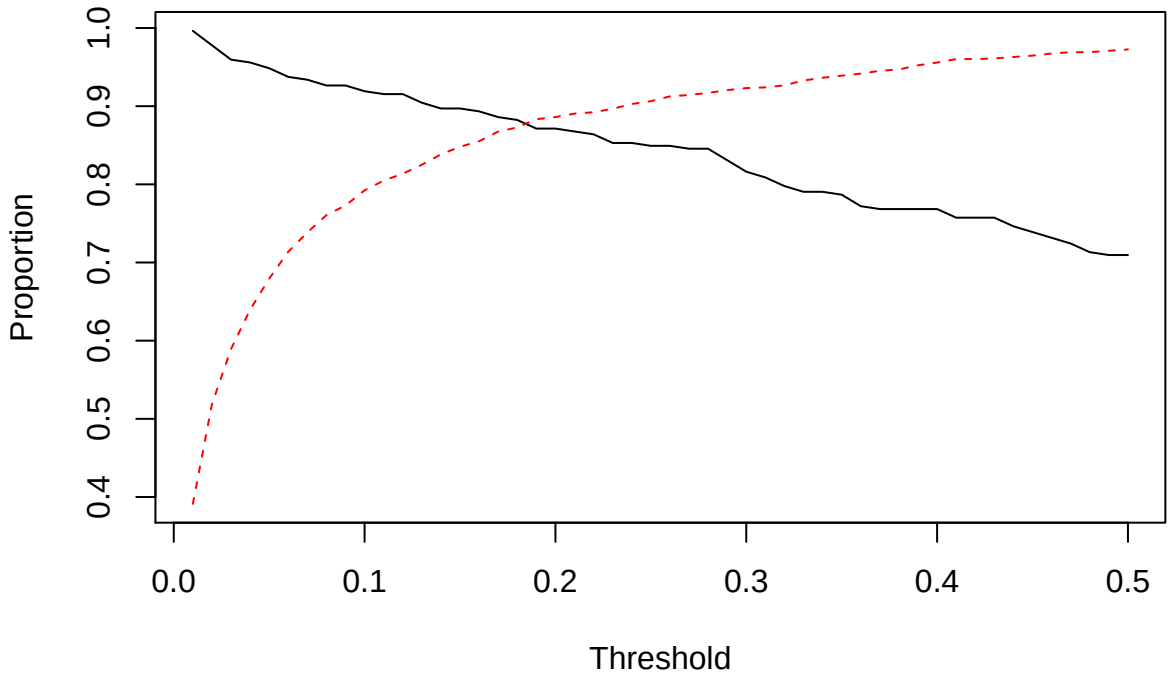
```
##       FTR       TOR     EFG_O       WAB     ADJDE     ADJOE   BARTHAG
##  1.174302  1.849976  2.771205 11.270400 14.118669 19.035144 36.023259
```

Appendix 4.4. Prediction

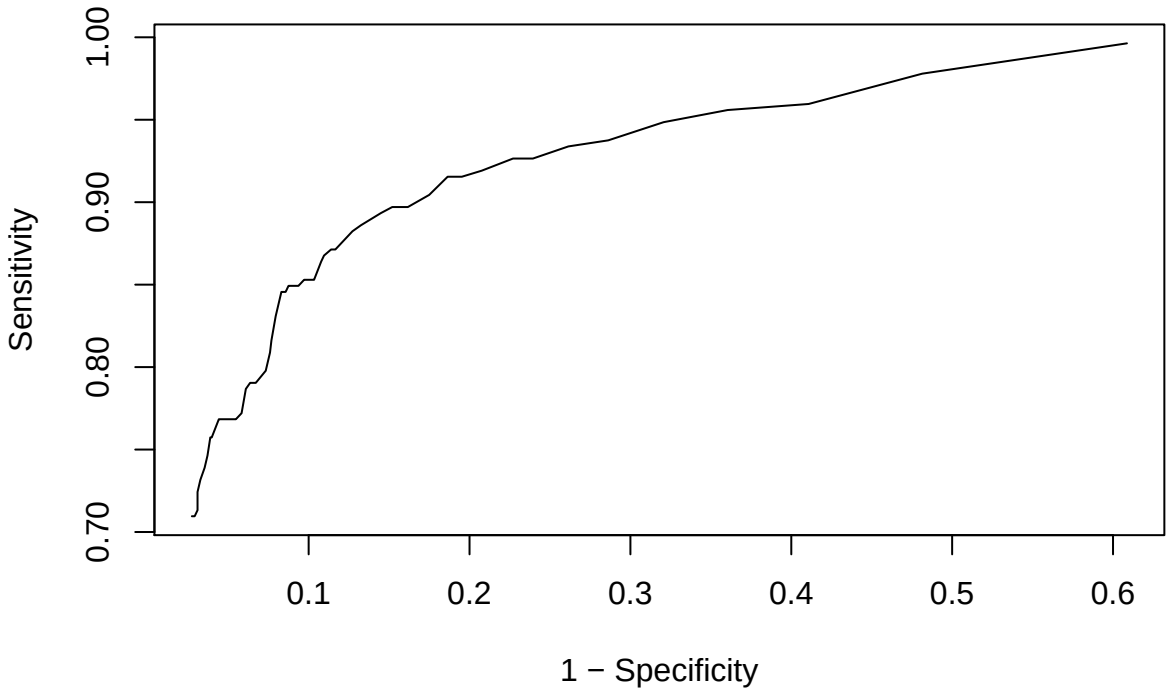
```
# Checking Performance
predprob=predict(model_binom13, train_data, type="response")

thresh <- seq(0.01,0.5,0.01)
Sensitivity <- numeric(length(thresh))
Specificity <- numeric(length(thresh))
for(j in seq(along=thresh)){
  pp <- ifelse(predprob < thresh[j],"No","Yes")
  xx <- xtabs( ~ train_data$TRNMT + pp)
  Specificity[j] <- xx[1,1]/(xx[1,1]+xx[1,2])
```

```
  Sensitivity[j] <- xx[2,2]/(xx[2,1]+xx[2,2])
}
matplot(thresh,cbind(Sensitivity,Specificity),type="l",xlab="Threshold",ylab="Proportion",lty=1:2)
```




```
plot(1-Specificity,Sensitivity,type="l")
abline(0,1,lty=2)
```




```
# Classification: Sensitivity and Specificity (ROC)
predout=ifelse(predprob < 0.18, "No", "Yes")
xtabs( ~ train_data$TRNMT + predout)
```

```
##                 predout
## train_data$TRNMT  No Yes
##              No  988 144
##              Yes  32 240
```

```
# Training Error classification rate
1-(988+240)/(988+240+144+32)
```

```
## [1] 0.1253561
```

```
# Testing Error classification rate
predprob_test=predict(model_binom10, test_data_19, type="response")
predout_test=ifelse(predprob_test < 0.18, "No", "Yes")
xtabs( ~ test_data_19$TRNMT + predout_test)
```

```
##                   predout_test
## test_data_19$TRNMT  No Yes
##                No  247  38
##                Yes   7  61
```

```r
1-(247+61)/(247+61+7+38)
```

```
## [1] 0.1274788
```

Appendix 5. Model 2: Multinomial Model - All possibilities

```
mmod <- multinom(POSTSEASON ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR +
                   TORD + ORB + DRB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D +
                   THREE_P_O + THREE_P_D + WAB + CONF, train_data, trace = FALSE)
mmod1 <- step(mmod, trace=FALSE)
```

Appendix 5.1. Prediction

```
summary(mmod1)
```

```
## Call:
## multinom(formula = POSTSEASON ~ ADJOE + ADJDE + BARTHAG + EFG_O +
##     EFG_D + TOR + TWO_P_O + WAB, data = train_data, trace = FALSE)
##
## Coefficients:
##           (Intercept)     ADJOE      ADJDE   BARTHAG       EFG_O
## R68          22.36917 0.6930545 -0.6387554 -35.55184  0.10514273
## R64          20.05135 0.3591553 -0.4764981 -25.62416  0.18339258
## R32          43.97908 0.4799637 -0.7631146 -29.61770  0.00383604
## S16         -25.97634 0.2539697 -0.3535155  24.02096 -0.16766864
## E8           25.74885 0.5936211 -0.9496349 -25.40841 -0.40395542
## F4           31.32057 0.7035848 -1.1607442 -31.56321  0.51679491
## 2ND          38.68386 1.5249836 -1.8073067 -14.92299 -4.01741403
## Champions   -26.29208 4.2496101 -5.5472676 -91.10589 -3.70585699
##                EFG_D         TOR       TWO_P_O        WAB
## R68       -0.36310116 -0.20504637  0.0856522551  0.3284524
## R64       -0.03428303 -0.04884582 -0.0006071909  0.6555698
## R32       -0.08257088 -0.38844792  0.2031092296  0.7157722
## S16        0.07755960 -0.14868146  0.3246490619  0.6036942
## E8         0.17909511 -0.28493369  0.6690769352  0.8535485
## F4         0.05044687  0.22295890 -0.1748514197  0.5565037
## 2ND       -0.42421579 -3.52135155  4.7837766041 -0.0726802
## Champions  3.37708666  0.62909477  2.9240761169 -1.2029202
##
## Std. Errors:
##           (Intercept)      ADJOE      ADJDE    BARTHAG     EFG_O
## R68        0.24180881 0.07939885 0.08769994 0.71626161 0.2176209
## R64        1.74444597 0.07222554 0.05875595 2.26236503 0.0995162
## R32        2.19415373 0.09344604 0.08559134 3.07789830 0.1635410
## S16        1.09455542 0.09604182 0.12575847 1.02116845 0.2265747
## E8         0.32485850 0.12335647 0.16585797 0.33350867 0.3171549
## F4         0.10823782 0.15291949 0.21733239 0.12743699 0.3413094
## 2ND        0.04078992 0.41198906 0.59784005 0.03145736 1.3029543
## Champions  0.04017241 0.92960089 1.48949074 0.04935019 1.3030123
##                EFG_D        TOR   TWO_P_O        WAB
## R68       0.15506872 0.13954521 0.1670340 0.13315705
## R64       0.07314185 0.06981779 0.0760650 0.06980336
## R32       0.11333051 0.10913873 0.1265307 0.10697108
## S16       0.14700707 0.13881999 0.1784446 0.13511803
```

```
## E8          0.19966822 0.20043210 0.2589957 0.18736637
## F4          0.23504649 0.27781350 0.2728535 0.21217333
## 2ND         0.64280916 1.07712118 1.4243489 0.45968890
## Champions 1.38532950 0.91461645 1.1864221 1.38636168
##
## Residual Deviance: 1142.137
## AIC: 1286.137
```

```
# Train Error
mmod1.pred <- predict(mmod1, train_data)
mmod1.table <- table(mmod1.pred, train_data[,"POSTSEASON"])
mmod1.error <- numeric(dim(mmod1.table)[1])
for(i in 1:dim(mmod1.table)[1]){
  mmod1.error[i] = round(((1-(mmod1.table[i,i])/(sum(mmod1.table[,i])))*100), 4)
}
mmod1.error.table <- data.frame(names(mmod1.table[,1]), mmod1.error)
colnames(mmod1.error.table) <- c("Round", "% Error")

# Test Error
mmod1.pred.test <- predict(mmod1, test_data_19)
mmod1.table.test <- table(mmod1.pred.test, test_data_19[,"POSTSEASON"])
mmod1.error.test <- numeric(dim(mmod1.table.test)[1])
for(i in 1:dim(mmod1.table.test)[1]){
  mmod1.error.test[i] = round(((1-(mmod1.table.test[i,i])/(sum(mmod1.table.test[,i])))*100), 4)
}
mmod1.error.test.table <- data.frame(names(mmod1.table.test[,1]), mmod1.error.test)
colnames(mmod1.error.test.table) <- c("Round", "% Error")
```

Appendix 5.2. Error Tables (%)

```
knitr::kable(mmod1.error.table)
```

Round	% Error
No Tournament	1.5901
R68	100.0000
R64	65.6250
R32	54.6875
S16	78.1250
E8	68.7500
F4	100.0000
2ND	0.0000
Champions	0.0000

```
knitr::kable(mmod1.error.test.table)
```

Round	% Error
No Tournament	1.7544
R68	100.0000

Round	% Error
R64	75.0000
R32	75.0000
S16	75.0000
E8	75.0000
F4	100.0000
2ND	100.0000
Champions	100.0000

Appendix 6. Model 3: Multinomial Model - Round Selection Given Already in Tournament

```
train_given_trnmt <- train_data[which(train_data$TRNMT=="Yes"), ]
train_given_trnmt$POSTSEASON <- as.character(train_given_trnmt$POSTSEASON)
train_given_trnmt$POSTSEASON <- as.factor(train_given_trnmt$POSTSEASON)
test_given_trnmt_19 <- test_data_19[which(test_data_19$TRNMT=="Yes"), ]
test_given_trnmt_19$POSTSEASON <- as.character(test_given_trnmt_19$POSTSEASON)
test_given_trnmt_19$POSTSEASON <- as.factor(test_given_trnmt_19$POSTSEASON)

mmod2 <- multinom(POSTSEASON ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR +
                  TORD + ORB + DRB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D +
                  THREE_P_O + THREE_P_D + WAB + CONF, train_given_trnmt, trace = FALSE)
mmod3 <- step(mmod2, trace=0)
```

Appendix 6.1. Prediction

```
summary(mmod3)
```

```
## Call:
## multinom(formula = POSTSEASON ~ ADJOE + ADJDE + TOR + ORB + DRB +
##     TWO_P_O + THREE_P_D, data = train_given_trnmt, trace = FALSE)
##
## Coefficients:
##           (Intercept)     ADJOE    ADJDE      TOR       ORB        DRB
## Champions -236.056663  2.852387 2.555557 11.78722 -2.025824 -1.3480990
## E8           8.900188 -2.173893 5.519055 19.74867 -2.692411  1.0909452
## F4           3.967330 -1.924539 5.569355 20.86565 -3.114750  1.3208299
## R32         15.016957 -2.274593 5.875645 19.84169 -2.725084  0.9152105
## R64         -8.026639 -2.307673 6.205060 20.28381 -2.765259  0.7592421
## R68        -25.960280 -2.152685 6.488939 20.72191 -2.997752  0.8513871
## S16        -11.220968 -2.103083 5.802243 20.07609 -2.719376  0.9520225
##             TWO_P_O THREE_P_D
## Champions -6.599466 -1.821630
## E8        -7.732776 -2.242833
## F4        -8.075189 -2.994908
## R32       -8.013475 -2.467972
## R64       -8.209898 -2.358930
## R68       -8.396799 -3.068424
## S16       -8.022214 -2.227293
##
## Std. Errors:
##           (Intercept)     ADJOE     ADJDE       TOR       ORB      DRB
## Champions  0.18694932 1.3967303 1.5390742 3.6535538 0.6854387 1.090490
## E8         0.44581638 0.8797126 0.8448868 0.5426949 0.6486490 1.168592
## F4         0.05398827 0.8881353 0.8596159 0.6125661 0.6667125 1.185734
## R32        3.50710706 0.8794955 0.8478554 0.5219588 0.6466125 1.170012
## R64        4.00974955 0.8798437 0.8484075 0.5212229 0.6472330 1.170833
## R68        0.74323733 0.8820179 0.8511628 0.5477912 0.6548939 1.176055
```

```
## S16         0.25546310 0.8796789 0.8476393 0.5272317 0.6467417 1.170386
##             TWO_P_O THREE_P_D
## Champions 1.141480  1.913517
## E8        1.078119  1.338540
## F4        1.085946  1.367293
## R32       1.076049  1.340593
## R64       1.076647  1.341498
## R68       1.084958  1.356640
## S16       1.077323  1.340915
##
## Residual Deviance: 538.46
## AIC: 650.46
```

```
# Train Error
mmod3.pred <- predict(mmod3, train_given_trnmt)
mmod3.table <- table(mmod3.pred, train_given_trnmt[,"POSTSEASON"])
mmod3.error <- numeric(dim(mmod3.table)[1])
for(i in 1:dim(mmod3.table)[1]){
  mmod3.error[i] = round(((1-(mmod3.table[i,i])/(sum(mmod3.table[,i])))*100), 4)
}
mmod3.error.table <- data.frame(names(mmod3.table[,1]), mmod3.error)
colnames(mmod3.error.table) <- c("Round", "% Error")

# Test Error
mmod3.pred.test <- predict(mmod3, test_given_trnmt_19)
mmod3.table.test <- table(mmod3.pred.test, test_given_trnmt_19[,"POSTSEASON"])
mmod3.error.test <- numeric(dim(mmod3.table.test)[1])
for(i in 1:dim(mmod3.table.test)[1]){
  mmod3.error.test[i] = round(((1-(mmod3.table.test[i,i])/(sum(mmod3.table.test[,i])))*100), 4)
}
mmod3.error.test.table <- data.frame(names(mmod3.table.test[,1]), mmod3.error.test)
colnames(mmod3.error.test.table) <- c("Round", "% Error")
```

Appendix 6.2. Error Tables (%)

```
knitr::kable(mmod3.error.table)
```

Round	% Error
2ND	0.0000
Champions	0.0000
E8	75.0000
F4	75.0000
R32	53.1250
R64	13.2812
R68	81.2500
S16	75.0000

```
knitr::kable(mmod3.error.test.table)
```

Round	% Error
2ND	100.0
Champions	0.0
E8	75.0
F4	100.0
R32	50.0
R64	25.0
R68	50.0
S16	62.5

Appendix 7. Model 4: Classification Tree - All possibilities

```
# Bagging
bag.cbb <- randomForest(POSTSEASON ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR +
                  TORD + ORB + DRB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D +
                  THREE_P_O + THREE_P_D + WAB + CONF, train_data, mtry=18, importance=T)
bag.cbb
```

```
##
## Call:
##  randomForest(formula = POSTSEASON ~ ADJOE + ADJDE + BARTHAG +      EFG_O + EFG_D + TOR + TORD + ORB
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 18
##
##         OOB estimate of  error rate: 15.81%
## Confusion matrix:
##               No Tournament R68 R64 R32 S16 E8 F4 2ND Champions
## No Tournament          1114   0  16   2   0  0  0   0         0
## R68                      13   0   3   0   0  0  0   0         0
## R64                      68   0  45  12   3  0  0   0         0
## R32                      12   0  21  17  10  3  0   0         1
## S16                       3   0   5  17   4  3  0   0         0
## E8                        0   0   4   5   4  2  1   0         0
## F4                        1   0   1   3   2  0  0   0         1
## 2ND                       0   0   0   1   0  2  0   0         1
## Champions                 0   0   0   3   0  1  0   0         0
##               class.error
## No Tournament  0.01590106
## R68            1.00000000
## R64            0.64843750
## R32            0.73437500
## S16            0.87500000
## E8             0.87500000
## F4             1.00000000
## 2ND            1.00000000
## Champions      1.00000000
```

```
importance(bag.cbb)
```

```
##           No Tournament         R68          R64         R32        S16
## ADJOE       15.67835597 -0.93312258  -5.02668979  1.74994572  1.9153305
## ADJDE        6.62177339 -1.77410628  -0.32637798  4.41255190 -2.7950529
## BARTHAG     20.76796719 -0.62667335 -15.00950552  2.50311592 16.3233098
## EFG_O        7.50099642  0.62936560   3.13095180 -3.87104401 -0.4123208
## EFG_D       12.41735393 -1.95194898  -1.17847550  0.65810438 -4.2414247
## TOR         13.68030969  0.86712456  -4.95771879  9.75086054 -6.8194483
## TORD         2.34339108  2.15681669  -2.14922398  1.90159055  3.5940922
## ORB          7.96752682  1.99407970   1.52415853 -0.07218137 -3.1188850
## DRB          4.65633047  2.44619065   1.74733567 -0.40556357 -1.4609951
## FTR          9.67861176  1.34458587   1.39771949 -0.58548431 -3.2252597
## FTRD        12.67805126  0.80446409  -0.06405381 -2.74702557  2.2624564
```

```
## ADJ_T       -0.09613533  0.34002065  -0.65463358 -1.59606913 -1.6566535
## TWO_P_O     12.72989310 -0.12847545   2.82041888 -2.80555642  2.7055801
## TWO_P_D      6.20676578  1.66986811  -0.20670393  2.23140896 -4.3055149
## THREE_P_O    5.16366544  0.93853795   4.23366027 -1.67706769 -2.6315510
## THREE_P_D   13.88623672  2.83704899  -0.27877529  4.08536114 -4.1695325
## WAB         81.65986130  0.36268443  28.54001214 42.25087861 28.7080906
## CONF        18.45531021 -0.04121671   5.29678822  0.46241115 -0.4153374
##                    E8          F4         2ND   Champions
## ADJOE     -1.38123177 -0.09835262  0.09491665  4.42659843
## ADJDE     -0.74821533 -0.42107688  0.63662149  1.32177554
## BARTHAG   10.42552909  3.17357304  4.01277893  9.83209996
## EFG_O      1.93531756 -1.15585603  0.41215169  1.12256457
## EFG_D      0.20771086  1.21446542  0.00000000  0.57226487
## TOR       -0.07533205 -0.73280281  2.85994166 -0.06984337
## TORD       0.31953933 -0.23250785 -0.57754282  0.10425834
## ORB        0.59151758  1.19667218 -1.41705050  0.21321041
## DRB       -2.18508589  1.73833842  1.29315150 -1.96874808
## FTR        1.63320920 -0.63120036 -1.00100150 -1.61006779
## FTRD      -1.33592151  0.87216244 -0.82255090  0.70892831
## ADJ_T     -1.46428725  0.45300375 -1.63736531  0.85581223
## TWO_P_O    4.24148103 -3.30867967  0.33337038  1.06185543
## TWO_P_D   -1.40692092 -2.18721555 -1.00100150 -1.69640277
## THREE_P_O -0.50020436  1.76204510 -0.83262693 -1.28313261
## THREE_P_D -0.75966210  2.09465131 -1.88327385  0.00000000
## WAB       17.23814775  3.48497671  3.64339626  9.07836966
## CONF       1.79054694  1.29321837  0.21822828  0.55973710
##           MeanDecreaseAccuracy MeanDecreaseGini
## ADJOE                15.191981        13.583919
## ADJDE                 6.612711        10.779949
## BARTHAG              21.623182        33.626511
## EFG_O                 7.323086        10.002384
## EFG_D                11.909420         9.436318
## TOR                  12.977863        15.972846
## TORD                  2.459812        15.275153
## ORB                   7.233455        17.196017
## DRB                   4.235551        14.382168
## FTR                   8.232703        18.120075
## FTRD                 11.321849        16.901478
## ADJ_T                -1.394040        14.341741
## TWO_P_O              12.962444        13.288670
## TWO_P_D               5.454263        10.819950
## THREE_P_O             5.051946        14.893252
## THREE_P_D            13.011488        16.129329
## WAB                  98.339254       184.488331
## CONF                 18.509861        46.594871
```

```
# Random Forest
rf.cbb <- randomForest(POSTSEASON ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR +
                  TORD + ORB + DRB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D +
                  THREE_P_O + THREE_P_D + WAB + CONF, train_data, importance=T)
rf.cbb
```

```
##
## Call:
```

```
##  randomForest(formula = POSTSEASON ~ ADJOE + ADJDE + BARTHAG +      EFG_O + EFG_D + TOR + TORD + ORB
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 4
##
##         OOB estimate of  error rate: 16.45%
## Confusion matrix:
##               No Tournament R68 R64 R32 S16 E8 F4 2ND Champions
## No Tournament          1116   0  13   2   1  0  0   0         0
## R68                      12   0   4   0   0  0  0   0         0
## R64                      74   0  35  16   3  0  0   0         0
## R32                      14   0  19  18  10  2  0   0         1
## S16                       2   0   9  18   2  1  0   0         0
## E8                        0   0   4   6   4  2  0   0         0
## F4                        1   0   1   5   1  0  0   0         0
## 2ND                       0   0   0   2   0  1  0   0         1
## Champions                 0   0   0   2   0  2  0   0         0
##               class.error
## No Tournament  0.01413428
## R68            1.00000000
## R64            0.72656250
## R32            0.71875000
## S16            0.93750000
## E8             0.87500000
## F4             1.00000000
## 2ND            1.00000000
## Champions      1.00000000
```

```
importance(rf.cbb)
```

```
##           No Tournament         R68         R64         R32        S16
## ADJOE         17.516365  0.03712023 -4.09221832  2.27568860  8.3836460
## ADJDE         16.580518 -2.60754808 -3.81142243  7.18086637  2.8849812
## BARTHAG       22.890156 -1.43707535 -5.09296854  9.40071881 15.0781601
## EFG_O         13.704443 -1.03310994  3.81633754 -2.78545688  0.8523017
## EFG_D         15.656768 -4.75616547  0.60533887  3.53555515 -3.2388041
## TOR            9.928228  0.58239392 -3.31992969  8.84716127 -3.4234049
## TORD           7.215060 -1.53338164 -0.96522612  0.02271085  2.3047714
## ORB            8.062355 -1.20516012 -0.88985411 -0.75024434 -2.5147046
## DRB            4.110359 -1.64087992  1.00294120 -3.74633436 -1.6365326
## FTR            2.281594  1.99661271  4.79745906 -2.09787117 -2.2904637
## FTRD           8.697927 -0.41623308 -0.05596516 -3.54647696  1.9577487
## ADJ_T          1.310325  0.60139223  0.67365543  0.60762716 -2.1459037
## TWO_P_O       12.082016  0.44642798  3.24267001 -3.27391669  3.3297598
## TWO_P_D       13.321296 -3.16760248 -1.91964788  1.05950355 -0.6558216
## THREE_P_O      9.489062 -0.01097076  3.40521365 -0.73080363 -1.3177905
## THREE_P_D     11.231062  0.31627742 -2.26346583  4.36052927 -3.2437657
## WAB           35.904206 -0.36663767 21.22823435 17.72189220 15.3222928
## CONF           9.864365  0.43753432  1.79536055  0.63667494 -2.0388192
##                     E8         F4           2ND  Champions
## ADJOE      5.169928685  0.2839869  2.413971e+00  6.8418715
## ADJDE      2.094628689  1.9796397  1.702664e+00  5.6774521
## BARTHAG   10.090560461  1.6202907  2.356755e+00  8.9834834
## EFG_O      3.290323527  0.4257054 -9.548675e-01  2.7096372
```

```
## EFG_D      0.763504149  0.2332974  1.480331e+00  2.8133076
## TOR       -0.294682560 -2.4947759  2.977290e+00  3.5549915
## TORD       0.559092816 -2.0189680 -5.627221e-01  0.6745065
## ORB        0.598445068  0.6577906 -2.161808e+00 -1.5819300
## DRB       -0.123490419 -2.0321987  3.492577e-01  0.1428601
## FTR       -0.509940335  0.7719138 -1.447010e+00 -0.7019921
## FTRD      -1.267969236  0.8740944  1.001002e+00  0.7883281
## ADJ_T      0.798564735 -1.0359028 -1.417051e+00 -1.0950482
## TWO_P_O    4.472930613 -3.3273379  1.116161e+00  3.1231336
## TWO_P_D    0.897222903  0.6557324  8.901588e-17  0.2515932
## THREE_P_O  0.002376116  1.1370170 -1.150272e+00  0.4790230
## THREE_P_D  0.050716978  1.9170528 -1.315053e+00  0.7530714
## WAB       13.010322056  1.2794265  4.430478e+00  7.4431495
## CONF       0.048194574 -0.2350161 -3.721557e-01  1.8774112
##           MeanDecreaseAccuracy MeanDecreaseGini
## ADJOE                17.705543         37.97875
## ADJDE                16.915173         31.15775
## BARTHAG              24.114008         57.22234
## EFG_O                13.897414         18.51597
## EFG_D                15.474244         19.25853
## TOR                   9.511572         19.29152
## TORD                  5.919104         16.07345
## ORB                   5.852788         16.09836
## DRB                   2.133066         14.52356
## FTR                   3.354277         17.60762
## FTRD                  7.554887         16.48044
## ADJ_T                 1.211641         15.51746
## TWO_P_O              12.037891         18.25599
## TWO_P_D              12.858595         16.62382
## THREE_P_O             9.534882         17.03434
## THREE_P_D            10.188479         16.65825
## WAB                  40.759544         89.99358
## CONF                  8.983196         34.24977
```

Appendix 7.1. Prediction

```
# Testing Error - Bagging
bag.pred_test <- predict(bag.cbb, test_data_19, type = "class")
bag.table <- table(bag.pred_test, test_data_19[,"POSTSEASON"])
bag.error <- numeric(dim(bag.table)[1])
for(i in 1:dim(bag.table)[1]){
  bag.error[i] = round(((1-(bag.table[i,i])/(sum(bag.table[,i])))*100), 4)
}
bag.error.table <- data.frame(names(bag.table[,1]), bag.error)
colnames(bag.error.table) <- c("Round", "% Error")

# Testing Error - Random Forest
rf.pred_test <- predict(rf.cbb, test_data_19, type = "class")
rf.table <- table(rf.pred_test, test_data_19[,"POSTSEASON"])
rf.error <- numeric(dim(rf.table)[1])
for(i in 1:dim(rf.table)[1]){
  rf.error[i] = round(((1-(rf.table[i,i])/(sum(rf.table[,i])))*100), 4)
```

```
}
rf.error.table <- data.frame(names(rf.table[,1]), rf.error)
colnames(rf.error.table) <- c("Round", "% Error")
```

Appendix 7.2. Error Tables (%)

```
knitr::kable(bag.error.table)
```

Round	% Error
No Tournament	1.4035
R68	100.0000
R64	68.7500
R32	81.2500
S16	87.5000
E8	50.0000
F4	100.0000
2ND	100.0000
Champions	0.0000

```
knitr::kable(rf.error.table)
```

Round	% Error
No Tournament	1.0526
R68	100.0000
R64	78.1250
R32	81.2500
S16	87.5000
E8	50.0000
F4	100.0000
2ND	100.0000
Champions	100.0000

Appendix 8. Model 5: Classification Tree - Round Selection Given Already in Tournament

```
# Bagging
bag.cbb_trmnt <- randomForest(POSTSEASON ~ ADJOE + ADJDE + BARTHAG + EFG_O + EFG_D + TOR +
                   TORD + ORB + DRB + FTR + FTRD + ADJ_T + TWO_P_O + TWO_P_D +
                   THREE_P_O + THREE_P_D + WAB + CONF, train_given_trnmt, mtry=18, importance=T)
bag.cbb_trmnt
```

```
##
## Call:
##  randomForest(formula = POSTSEASON ~ ADJOE + ADJDE + BARTHAG +      EFG_O + EFG_D + TOR + TORD + ORB
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 18
##
##         OOB estimate of  error rate: 47.79%
## Confusion matrix:
##           2ND Champions E8 F4 R32 R64 R68 S16 class.error
## 2ND         0         1  2  0   1   0   0   0    1.000000
## Champions   0         1  1  0   2   0   0   0    0.750000
## E8          0         0  2  1   6   4   0   3    0.875000
## F4          0         0  0  0   3   2   0   3    1.000000
## R32         0         1  3  0  24  27   0   9    0.625000
## R64         0         0  0  0  13 110   0   5    0.140625
## R68         0         0  0  0   1  15   0   0    1.000000
## S16         0         0  3  0  15   9   0   5    0.843750
```

```
importance(bag.cbb_trmnt)
```

```
##                  2ND     Champions          E8          F4        R32
## ADJOE     -1.8469846  5.342399e+00  0.06538190  0.15144294 -2.7891609
## ADJDE     -1.0010015  2.022215e+00  0.09703627  1.09909616  3.6386826
## BARTHAG    3.1940198  1.018070e+01  8.35282462  3.46821184 -3.2654904
## EFG_O     -0.2062930 -2.582161e-01  1.06309367 -1.48284067 -2.1858729
## EFG_D     -1.0010015 -4.473031e-01 -1.18092311 -2.06176847 -0.9498378
## TOR        1.7105892  2.128188e+00  0.20803596 -0.24097979  3.4468345
## TORD      -0.7849477 -1.624634e-01 -0.19697171  1.35363955  1.3800280
## ORB       -2.5416040  1.009989e-16  2.56524556  1.38319778 -2.1528073
## DRB        0.3780185 -1.344062e+00  0.20102971 -0.03886942 -1.6018445
## FTR       -0.5424857 -2.049800e+00  2.02028175  0.32280512 -0.9895899
## FTRD       0.9184367  2.437140e+00 -0.83396233 -0.24548971 -1.2866105
## ADJ_T      0.6327087 -2.847705e-01 -3.34928744 -0.75118745 -0.6730142
## TWO_P_O    1.8849590  1.610068e+00  2.87191858 -2.34759563 -3.4334246
## TWO_P_D    1.7372705 -2.210726e+00  1.06662915 -1.54650166  0.7281433
## THREE_P_O -1.0010015 -2.554979e+00  0.02034866 -0.62883578 -0.7169213
## THREE_P_D  1.1948196 -5.489050e-01  1.25585646  2.14517572  3.9491805
## WAB        1.8220696  4.592111e+00  7.60437668  0.84495448 -4.7800010
## CONF       0.3333704  2.870732e+00  2.20579247  0.18551904  1.6552898
##                  R64         R68         S16 MeanDecreaseAccuracy
## ADJOE      5.0588635  1.0763093  1.5548960            4.4570488
```

```
## ADJDE      7.2126104  5.0591854 -2.1039096           8.2830164
## BARTHAG   33.1270155  6.9511785 10.8458114          32.5141130
## EFG_O      2.3717475  1.7341707  1.3180398           1.4891280
## EFG_D      4.2595231 -2.4255351 -2.7119394           0.6870089
## TOR        3.6657920 -0.1656119 -4.3784348           3.2024541
## TORD       2.1939582 -0.9811254  2.9382190           3.3069571
## ORB        0.7502062  0.3158865 -1.8694528          -1.3484595
## DRB       -0.3453966  3.9348031 -3.1523622          -1.1640590
## FTR        2.0269341  1.5291874 -2.0345251           0.3133915
## FTRD       4.6478166 -0.9403420  1.1924899           3.0530572
## ADJ_T     -0.8902485  1.8500909 -1.8511404          -1.9486951
## TWO_P_O    1.8380679  1.0624924  1.1896894           0.7157788
## TWO_P_D    1.3288280  0.5766336 -1.4787298           0.8103443
## THREE_P_O  5.9220827  1.0169490 -1.2450765           2.9454826
## THREE_P_D  5.0852191  3.8755252 -2.4240851           6.5279816
## WAB       12.1952839  5.0416586  0.1527851          11.3413443
## CONF       3.2081878  1.4087007 -2.8567038           3.1314758
##           MeanDecreaseGini
## ADJOE             6.799858
## ADJDE             7.433371
## BARTHAG          39.271930
## EFG_O             5.377488
## EFG_D             5.206575
## TOR               9.814789
## TORD             10.071716
## ORB               8.665016
## DRB               6.831251
## FTR               7.370977
## FTRD              9.596918
## ADJ_T             6.509115
## TWO_P_O           7.994255
## TWO_P_D           5.549594
## THREE_P_O         7.921413
## THREE_P_D        11.420087
## WAB              12.671988
## CONF             21.517040
```

Appendix 8.1. Prediction

```r
# Testing Error - Bagging
bag1.pred_test <- predict(bag.cbb_trmnt, test_given_trnmt_19, type = "class")
bag1.table <- table(bag1.pred_test, test_given_trnmt_19[,"POSTSEASON"])
bag1.error <- numeric(dim(bag1.table)[1])
for(i in 1:dim(bag1.table)[1]){
  bag1.error[i] = round(((1-(bag1.table[i,i])/(sum(bag1.table[,i])))*100), 4)
}
bag1.error.table <- data.frame(names(bag1.table[,1]), bag1.error)
colnames(bag1.error.table) <- c("Round", "% Error")
```

Appendix 8.2. Error Table (%)

```
knitr::kable(bag1.error.table)
```

Round	% Error
2ND	100.000
Champions	0.000
E8	75.000
F4	100.000
R32	81.250
R64	21.875
R68	75.000
S16	87.500

Appendix 9. Full Error Table (%)

```
# Model 1: Binomial
binom.table.error <- xtabs( ~ train_data$TRNMT + predout)
binom.error = round((1-(binom.table.error[1,1]+binom.table.error[2,2])/(sum(binom.table.error)))*100, 4)
binom.error.full <- c(binom.error, rep(NA, 8))
# Testing Error classification rate
binom.table.error.test <- xtabs( ~ test_data_19$TRNMT + predout_test)
binom.error.test = round((1-(binom.table.error.test[1,1]+binom.table.error.test[2,2])/(sum(binom.table.e
binom.error.test.full <- c(binom.error.test, rep(NA, 8))

# Model 2: Multi

# Model 3: Multi
mmod3.error.full <- c(NA, mmod3.error)
mmod3.error.test.full <- c(NA, mmod3.error.test)

# Model 4: RF
bag.error.train <- round(bag.cbb$confusion[,"class.error"]*100, 4)
rf.error.train <- round(rf.cbb$confusion[,"class.error"]*100, 4)


# Model 5: RF
bag1.error.train <- round(bag.cbb_trmnt$confusion[,"class.error"]*100, 4)
bag1.error.train.full <- c(NA, bag1.error.train)
bag1.error.full <- c(NA, bag1.error)

# Final Table
mini.error.table <- data.frame(mmod3.error.full, mmod3.error.test.full, bag1.error.train.full, bag1.erro
mini.error.table <- mini.error.table[c(1, 8, 7, 6, 9, 4, 5, 2, 3),]               # correcting order to
full.error.table1 <- data.frame(binom.error.full, binom.error.test.full,
                                mmod1.error, mmod1.error.test, bag.error.train,
                                bag.error)
full.error.table2 <- data.frame(rf.error.train, rf.error, mini.error.table)
names(full.error.table1) <- c("Binom. Train", "Binom. Test", "Multi. Train", "Multi. Test",
                              "Bag Train", "Bag Test")
names(full.error.table2) <- c("RF Train", "RF Test", "Sp. Multi. Train",
                              "Sp. Multi. Test", "Sp. Bag Train", "Sp. Bag Test")

knitr::kable(full.error.table1)
```

	Binom. Train	Binom. Test	Multi. Train	Multi. Test	Bag Train	Bag Test
No Tournament	12.5356	12.7479	1.5901	1.7544	1.5901	1.4035
R68	NA	NA	100.0000	100.0000	100.0000	100.0000
R64	NA	NA	65.6250	75.0000	64.8438	68.7500
R32	NA	NA	54.6875	75.0000	73.4375	81.2500
S16	NA	NA	78.1250	75.0000	87.5000	87.5000
E8	NA	NA	68.7500	75.0000	87.5000	50.0000
F4	NA	NA	100.0000	100.0000	100.0000	100.0000
2ND	NA	NA	0.0000	100.0000	100.0000	100.0000
Champions	NA	NA	0.0000	100.0000	100.0000	0.0000

```
knitr::kable(full.error.table2)
```

	RF Train	RF Test	Sp. Multi. Train	Sp. Multi. Test	Sp. Bag Train	Sp. Bag Test
No Tournament	1.4134	1.0526	NA	NA	NA	NA
R68	100.0000	100.0000	81.2500	50.0	100.0000	75.000
R64	72.6562	78.1250	13.2812	25.0	14.0625	21.875
R32	71.8750	81.2500	53.1250	50.0	62.5000	81.250
S16	93.7500	87.5000	75.0000	62.5	84.3750	87.500
E8	87.5000	50.0000	75.0000	75.0	87.5000	75.000
F4	100.0000	100.0000	75.0000	100.0	100.0000	100.000
2ND	100.0000	100.0000	0.0000	100.0	100.0000	100.000
Champions	100.0000	100.0000	0.0000	0.0	75.0000	0.000

Appendix 10. 2020 March Madness Predictions

```
bag.pred_test_20 <- predict(bag.cbb, test_data_20, type = "class")
mmod.pred_test_20 <- predict(mmod1, test_data_20, type = "class")
final_20 <- data.frame(test_data_20, bag.pred_test_20, mmod.pred_test_20)

summary(final_20[,c("bag.pred_test_20", "mmod.pred_test_20")])
```

```
##       bag.pred_test_20     mmod.pred_test_20
##  No Tournament:302     No Tournament:309
##  R32           : 25     R32           : 26
##  R64           : 24     R64           : 15
##  E8            :  2     E8            :  2
##  R68           :  0     2ND           :  1
##  S16           :  0     R68           :  0
##  (Other)       :  0     (Other)       :  0
```

Appendix 10.1. Late Round Predictions

```
final_20[which(final_20$bag.pred_test_20=="E8"), c("TEAM", "bag.pred_test_20")]
```

```
##         TEAM bag.pred_test_20
## 1758  Kansas               E8
## 1760 Gonzaga               E8
```

```
final_20[which(final_20$mmod.pred_test_20=="E8"), c("TEAM", "mmod.pred_test_20")]
```

```
##         TEAM mmod.pred_test_20
## 1758  Kansas                E8
## 1760 Gonzaga                E8
```

```
final_20[which(final_20$mmod.pred_test_20=="2ND"), c("TEAM", "mmod.pred_test_20")]
```

```
##        TEAM mmod.pred_test_20
## 1761 Dayton               2ND
```

Appendix 10.2. Big Ten Predictions

```
final_20[which(final_20$CONF=="B10"), c("TEAM", "bag.pred_test_20", "mmod.pred_test_20")]
```

```
##              TEAM bag.pred_test_20 mmod.pred_test_20
## 1762 Michigan St.              R32               R32
## 1765     Ohio St.              R32               R64
## 1771     Michigan              R32               R32
## 1772     Penn St.              R32               R32
```

```
## 1776    Wisconsin              R32              R32
## 1780       Purdue    No Tournament    No Tournament
## 1783     Maryland              R32              R32
## 1784    Minnesota    No Tournament    No Tournament
## 1786     Illinois              R32              R32
## 1787      Rutgers              R32              R32
## 1788         Iowa              R64              R64
## 1793      Indiana              R64              R64
## 1873 Northwestern    No Tournament    No Tournament
## 1916     Nebraska    No Tournament    No Tournament
```

```
options(op)
```