

Final Report

Jackson Bandow

3/23/2021

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(ggplot2)  
library(plotly)
```

```
##  
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':  
##  
## last_plot
```

```
## The following object is masked from 'package:stats':  
##  
## filter
```

```
## The following object is masked from 'package:graphics':  
##  
## layout
```

```
library(data.table)
```

```
##  
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':  
##  
##   between, first, last
```

```
library(Lahman)  
library(stringr)  
library(car)
```

```
## Loading required package: carData
```

```
##  
## Attaching package: 'carData'
```

```
## The following object is masked from 'package:Lahman':  
##  
##   Salaries
```

```
##  
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':  
##  
##   recode
```

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##  
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':  
##  
##   cov, smooth, var
```

```
library(gt)
```

```
get_prob <- function(input, coefs){
  inp <- t(as.matrix(cbind(rep(1, dim(input)[1]),input)))
  cf <- t(as.matrix(coefs))
  prob_new <- (exp(cf %*% inp)) / (1 + exp(cf %*% inp))
  names(prob_new) <- NULL
  return(as.vector(prob_new))
}

mc_rate <- function(probs, actual) {
  mean(round(probs, 0) != actual)
}

low_out <- function(col){
  stats <- summary(col)
  iqr <- stats[5] - stats[2]
  res <- stats[2] - 1.5*iqr
  names(res) <- NULL
  return(res)
}

standardize <- function(col) {
  return((col - mean(col)) / sd(col))
}
```

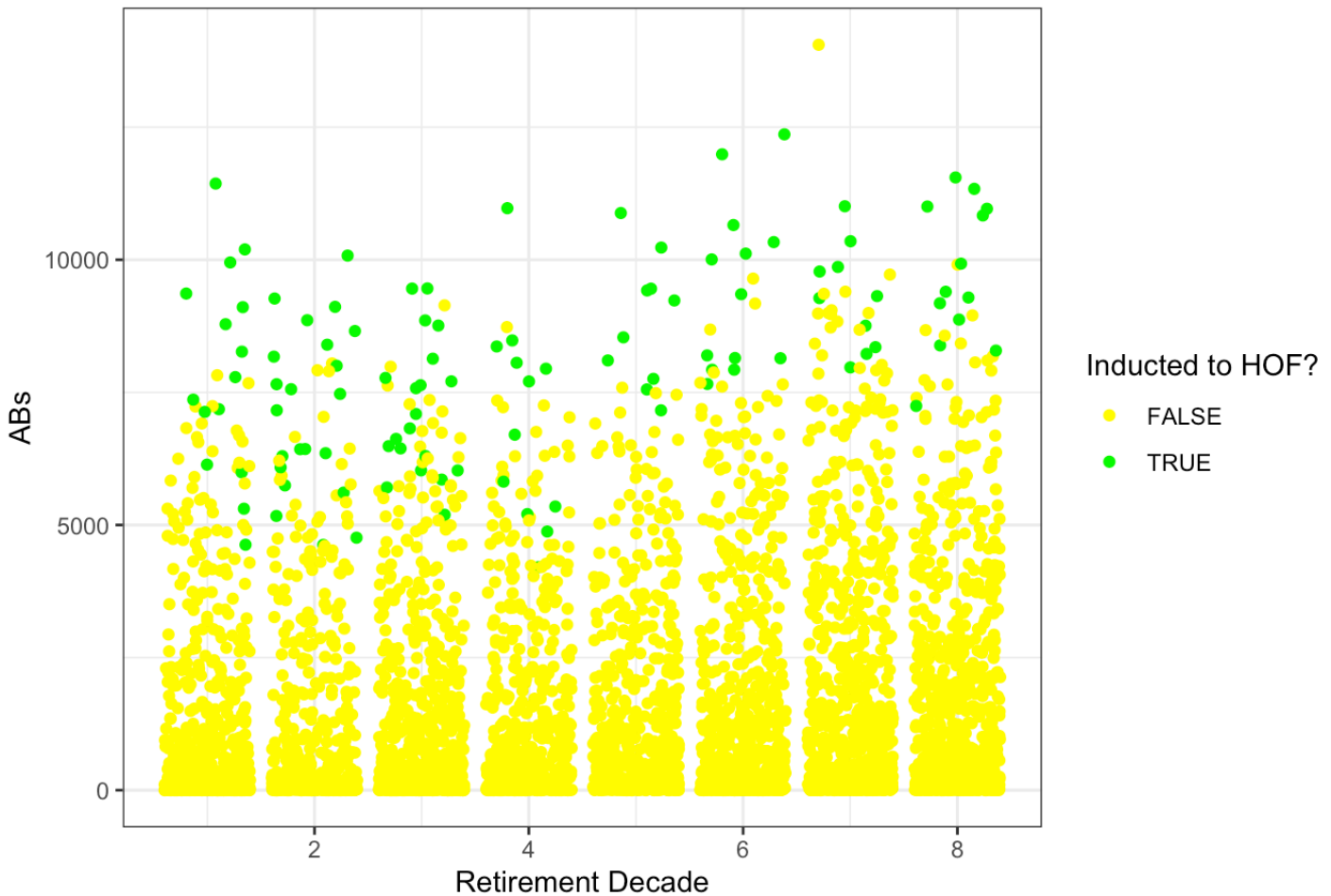
```
baseball <- fread("baseball.csv")
baseball$careerLength <- as.numeric(as.Date(baseball$finalGame) - as.Date(baseball$debut)) / 365.25
baseball$debutDecade <- round(as.numeric(format(baseball$debut, "%Y")), -1)
baseball$retireRank <- floor((as.numeric(format(baseball$finalGame, "%Y")) - 4) / 10) - 191
baseball$FP <- (baseball$PO + baseball$A) / (baseball$PO + baseball$A + baseball$E)

HOF_batters <- baseball %>% filter(inducted & position != "P" & votedBy != "Negro League" & str_sub(playerID, -2, -1) != "99" & finalGame < as.Date("2004-01-01") & finalGame > as.Date("1924-01-01"))

batters <- baseball %>% filter(position != "P" & votedBy != "Negro League" & str_sub(playerID, -2, -1) != "99" & finalGame < as.Date("2004-01-01") & finalGame > as.Date("1924-01-01"))
```

```
ggplot(batters, aes(retireRank, AB, color = inducted)) + geom_jitter() + theme_bw() +
scale_color_manual(values = c("yellow", "green")) + labs(title = "Number of At-Bats b
y Retirement Decade", x = "Retirement Decade", y = "ABs", color = "Inducted to HOF?")
```

Number of At-Bats by Retirement Decade



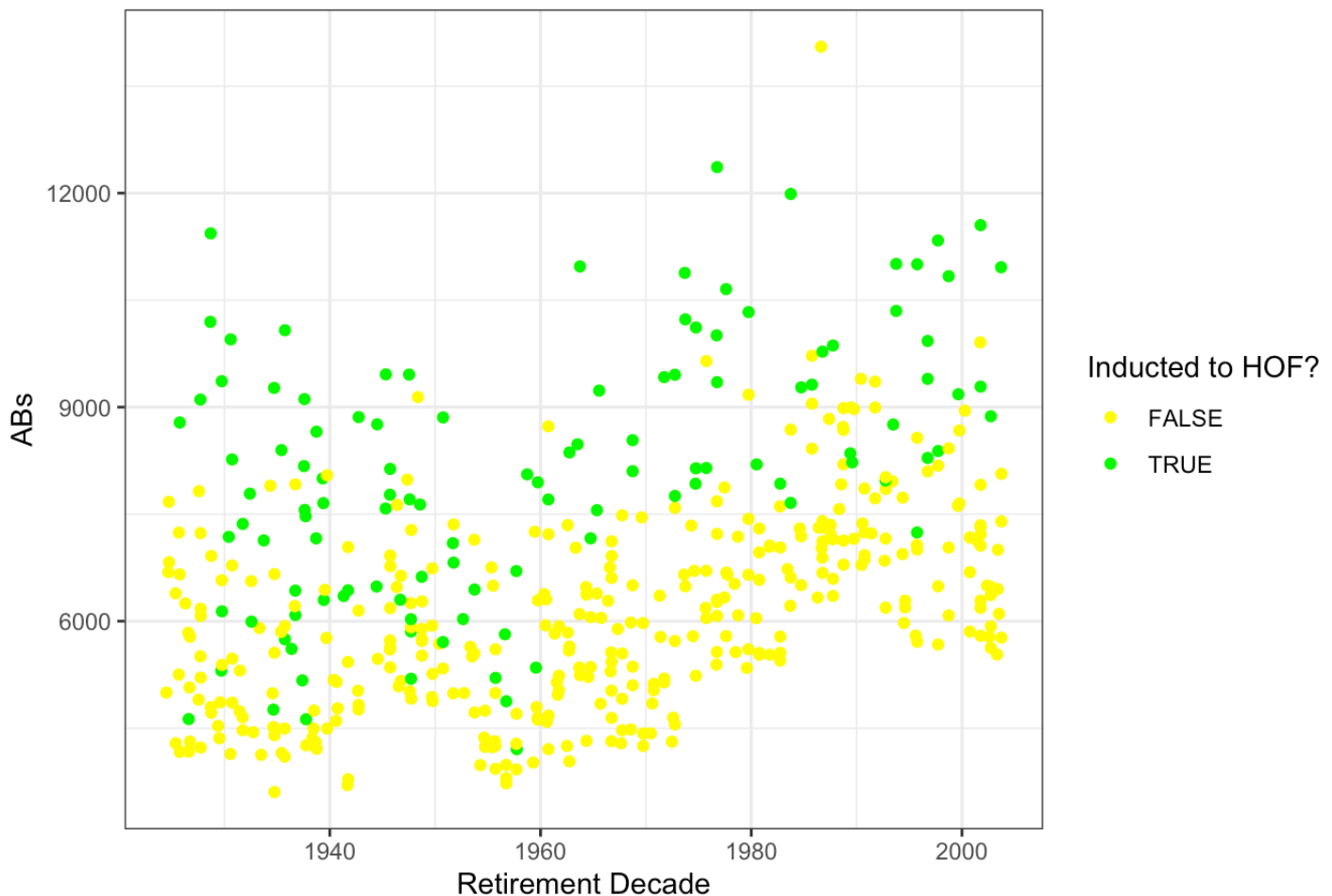
```
paste0("Composition of HOF batters in data: ", round(mean(batters$inducted)*100, 1),
"%.")
```

```
## [1] "Composition of HOF batters in data: 2.2%."
```

```
# Limit data by ABs by retirement decade
new_batters <- data.frame()
batters <- batters %>% filter(rWAR >= 10)
for (i in unique(batters$retireRank)) {
  temp <- batters %>% filter(retireRank == i)
  new_batters <- rbind(new_batters, temp[order(temp$AB)[(dim(temp)[1]-59):dim(temp)[1]],)
}

ggplot(new_batters, aes(finalGame, AB, color = inducted)) + geom_jitter() + theme_bw(
) + scale_color_manual(values = c("yellow", "green")) + labs(title = "Number of At-Bats by Retirement Decade", x = "Retirement Decade", y = "ABs", color = "Inducted to HOF?")
```

Number of At-Bats by Retirement Decade



```
paste0("Composition of HOF batters in data: ", round(mean(new_batters$inducted)*100,
1), "%.")
```

```
## [1] "Composition of HOF batters in data: 24%."
```

```
# Make k-fold indeces
rand_ind <- sample(1:dim(new_batters)[1], dim(new_batters)[1], replace = FALSE)
sets <- NULL
fold_size <- dim(new_batters)[1] / 5
for (i in 1:5) {
  sets <- append(sets, list(rand_ind[((i-1)*fold_size + 1):(i*fold_size)]))
}

# Model 1: WAR
mcrl <- NULL
auc1 <- NULL
for (i in 1:length(sets)) {
  mod <- glm(data = new_batters[!sets[[i]],], formula = induced ~ rWAR, family = "binomial")
  mcrl <- c(mcrl, mc_rate(get_prob(new_batters[sets[[i]],c("rWAR")), mod$coefficients), new_batters[sets[[i]],]$induced))
  auc1 <- c(auc1, roc(new_batters[sets[[i]],]$induced, get_prob(new_batters[sets[[i]],c("rWAR")), mod$coefficients))$auc)
}
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
# print(paste0("Model 1 -- Misclassification Rate: ", round(mean(mcr1)*100, 1), "%; Area Under the Curve: ", round(mean(auc1), 3)))

# Model 2: Standard Statistics
mcr2 <- NULL
auc2 <- NULL
for (i in 1:length(sets)) {
  mod <- glm(data = new_batters[!sets[[i]],], formula = induced ~ G.B + AB + BB + H
+ HR + R + RBI + SB + FP + OBP + SLG + BA, family = "binomial")
  mcr2 <- c(mcr2, mc_rate(get_prob(new_batters[sets[[i]],c("G.B", "AB", "BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA"))], mod$coefficients), new_batters[sets[[i]],]$induced))
  auc2 <- c(auc2, roc(new_batters[sets[[i]],]$induced, get_prob(new_batters[sets[[i]],c("G.B", "AB", "BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA"))], mod$coefficients))$auc)
}
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
# print(paste0("Model 2 -- Misclassification Rate: ", round(mean(mcr2)*100, 1), "%; A  
rea Under the Curve: ", round(mean(auc2), 3)))  
  
# Model 3: Combined  
mcr7 <- NULL  
auc7 <- NULL  
for (i in 1:length(sets)) {  
  mod <- glm(data = new_batters[!sets[[i]],], formula = induced ~ rWAR + G.B + AB +  
BB + H + HR + R + RBI + SB + FP + OBP + SLG + BA, family = "binomial")  
  mcr7 <- c(mcr7, mc_rate(get_prob(new_batters[sets[[i]],c("rWAR", "G.B", "AB", "BB",  
"H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")), mod$coefficients), new_batte  
rs[sets[[i]],]$inducted))  
  auc7 <- c(auc7, roc(new_batters[sets[[i]],]$inducted, get_prob(new_batters[sets[[i]]  
],c("rWAR", "G.B", "AB", "BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")  
], mod$coefficients))$auc)  
}
```

```
## Setting levels: control = FALSE, case = TRUE  
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```



```
# print(paste0("Model 3 -- Misclassification Rate: ", round(mean(mcr7)*100, 1), "%; Area Under the Curve: ", round(mean(auc7), 3)))

# Standardize data by retirement decade
snew_batters <- new_batters %>% select(rWAR, AB, H, HR, BB, G.B, BA, R, RBI, SB, OBP, SLG, FP) %>% log() %>% cbind(new_batters[,c("inducted", "steriods", "retireRank", "name", "career_len", "banned", "MVPs", "goldGloves", "BRaward", "CBPoY", "HAaward", "hutch", "LGMaward", "RCaward", "RoY", "SS", "TC", "allStarApps", "position", "def_pos", "winPer", "vets_comm")])
for (i in sort(unique(snew_batters$retireRank))) {
  snew_batters[snew_batters$retireRank == i, c("rWAR", "AB", "H", "HR", "BB", "G.B", "BA", "R", "RBI", "SB", "OBP", "SLG", "FP")] <- data.frame(apply(snew_batters[snew_batters$retireRank == i, c("rWAR", "AB", "H", "HR", "BB", "G.B", "BA", "R", "RBI", "SB", "OBP", "SLG", "FP")], 2, standardize))
}

# Standardized Model 1: WAR
mcr3 <- NULL
auc3 <- NULL
for (i in 1:length(sets)) {
  mod <- glm(data = snew_batters[!sets[[i]],], formula = inducted ~ rWAR, family = "binomial")
  mcr3 <- c(mcr3, mc_rate(get_prob(snew_batters[sets[[i]],c("rWAR")], mod$coefficients), snew_batters[sets[[i]],]$inducted))
  auc3 <- c(auc3, roc(snew_batters[sets[[i]],]$inducted, get_prob(snew_batters[sets[[i]],c("rWAR")], mod$coefficients))$auc)
}
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
# print(paste0("Standardized Model 1 -- Misclassification Rate: ", round(mean(mcr3)*100, 1), "%; Area Under the Curve: ", round(mean(auc3), 3)))
```

```
# Standardized Model 2: Standard Statistics
```

```
mcr4 <- NULL
```

```
auc4 <- NULL
```

```
for (i in 1:length(sets)) {  
  mod <- glm(data = sneu_batters[!sets[[i]],], formula = induced ~ G.B + AB + BB + H  
+ HR + R + RBI + SB + FP + OBP + SLG + BA, family = "binomial")  
  mcr4 <- c(mcr4, mc_rate(get_prob(sneu_batters[sets[[i]],c("G.B", "AB", "BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")), mod$coefficients), sneu_batters[sets[[i]],]$induced))  
  auc4 <- c(auc4, roc(sneu_batters[sets[[i]],]$induced, get_prob(sneu_batters[sets[[i]],c("G.B", "AB", "BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")), mod$coefficients))$auc)  
}
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
# print(paste0("Standardized Model 2 -- Misclassification Rate: ", round(mean(mcr4)*100, 1), "%; Area Under the Curve: ", round(mean(auc4), 3)))

# Standardized Model 3: Combined Stats
mcr8 <- NULL
auc8 <- NULL
for (i in 1:length(sets)) {
  mod <- glm(data = snw_batters[!sets[[i]],], formula = induced ~ rWAR + G.B + AB + BB + H + HR + R + RBI + SB + FP + OBP + SLG + BA, family = "binomial")
  mcr8 <- c(mcr8, mc_rate(get_prob(snw_batters[sets[[i]],c("rWAR", "G.B", "AB", "BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")), mod$coefficients), snw_batters[sets[[i]],]$induced))
  auc8 <- c(auc8, roc(snw_batters[sets[[i]],]$induced, get_prob(snw_batters[sets[[i]],c("rWAR", "G.B", "AB", "BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")), mod$coefficients))$auc)
}
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
# print(paste0("Standardized Model 3 -- Misclassification Rate: ", round(mean(mcr8)*100, 1), "%; Area Under the Curve: ", round(mean(auc8), 3)))

# Standardized Model 1 with Time Interactions: WAR
mcr5 <- NULL
auc5 <- NULL
for (i in 1:length(sets)) {
  mod <- glm(data = snw_batters[!sets[[i]],], formula = inducted ~ retireRank*rWAR,
family = "binomial")
  temp <- cbind(snw_batters[sets[[i]],c("retireRank", "rWAR)], snw_batters[sets[[i]],]$retireRank*snw_batters[sets[[i]],]$rWAR)
  mcr5 <- c(mcr5, mc_rate(get_prob(temp, mod$coefficients), snw_batters[sets[[i]],]$inducted))
  auc5 <- c(auc5, roc(snw_batters[sets[[i]],]$inducted, get_prob(temp, mod$coefficients))$auc)
}
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
# print(paste0("Standardized Model 1 with Time Interactions -- Misclassification Rate
: ", round(mean(mcr5)*100, 1), "%; Area Under the Curve: ", round(mean(auc5), 3)))

# Standardized Model 2 with Time Interactions: Standard Statistics
mcr6 <- NULL
auc6 <- NULL
for (i in 1:length(sets)) {
  mod <- glm(data = snw_batters[!sets[[i]],], formula = induced ~ retireRank*G.B +
retireRank*AB + retireRank*BB + retireRank*H + retireRank*HR + retireRank*R + retireR
ank*RBI + retireRank*SB + retireRank*FP + retireRank*OBP + retireRank*SLG + retireRan
k*BA, family = "binomial")
  temp <- cbind(snw_batters[sets[[i]],c("retireRank", "G.B", "AB", "BB", "H", "HR",
"R", "RBI", "SB", "FP", "OBP", "SLG", "BA")], snw_batters[sets[[i]],c("G.B", "AB", "
BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")]*snw_batters[sets[[i]],]
$retireRank)
  mcr6 <- c(mcr6, mc_rate(get_prob(temp, mod$coefficients), snw_batters[sets[[i]],]$
induced))
  auc6 <- c(auc6, roc(snw_batters[sets[[i]],]$induced, get_prob(temp, mod$coefficie
nts))$auc)
}
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
# print(paste0("Standardized Model 2 with Time Interactions -- Misclassification Rate
: ", round(mean(mcr6)*100, 1), "%; Area Under the Curve: ", round(mean(auc6), 3)))

# Standardized Model 3 with Time Interactions: Combined Stats
mcr9 <- NULL
auc9 <- NULL
for (i in 1:length(sets)) {
  mod <- glm(data = sneu_batters[!sets[[i]],], formula = inducted ~ retireRank*rWAR +
retireRank*G.B + retireRank*AB + retireRank*BB + retireRank*H + retireRank*HR + retir
eRank*R + retireRank*RBI + retireRank*SB + retireRank*FP + retireRank*OBP + retireRan
k*SLG + retireRank*BA, family = "binomial")
  temp <- cbind(sneu_batters[sets[[i]],c("retireRank", "rWAR", "G.B", "AB", "BB", "H"
, "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")], sneu_batters[sets[[i]],c("rWAR"
, "G.B", "AB", "BB", "H", "HR", "R", "RBI", "SB", "FP", "OBP", "SLG", "BA")]*sneu_bat
ters[sets[[i]],]$retireRank)
  mcr9 <- c(mcr9, mc_rate(get_prob(temp, mod$coefficients), sneu_batters[sets[[i]],]$
inducted))
  auc9 <- c(auc9, roc(sneu_batters[sets[[i]],]$inducted, get_prob(temp, mod$coefficie
nts))$auc)
}
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
# print(paste0("Standardized Model 3 with Time Interactions -- Misclassification Rate
: ", round(mean(mcr9)*100, 1), "%; Area Under the Curve: ", round(mean(auc9), 3)))

# Present Results in table
logisticModelRes <- data.frame(model = c("Model 1", "Model 2", "Model 3", "Standardiz
ed Model 1", "Standardized Model 2", "Standardized Model 3", "Standardized Model 1 wi
th Era Interactions", "Standardized Model 2 with Era Interactions", "Standardized Mod
el 3 with Era Interactions"), mc_rate = c(mean(mcr1), mean(mcr2), mean(mcr7), mean(mc
r3), mean(mcr4), mean(mcr8), mean(mcr5), mean(mcr6), mean(mcr9)), auc = c(mean(auc1),
mean(auc2), mean(auc7), mean(auc3), mean(auc4), mean(auc8), mean(auc5), mean(auc6), m
ean(auc9)))
colnames(logisticModelRes) <- c("Model", "Misclass Rate", "AUC")
gt(logisticModelRes) %>% tab_header(title = "Initial Logistic Modeling Results") %>%
fmt_percent(columns = vars("Misclass Rate"), decimals = 1) %>% fmt_number(columns = v
ars("AUC"), decimals = 3)
```

Initial Logistic Modeling Results

Model	Misclass Rate	AUC
Model 1	13.5%	0.909
Model 2	14.0%	0.889
Model 3	12.7%	0.920
Standardized Model 1	10.8%	0.931
Standardized Model 2	12.9%	0.886
Standardized Model 3	11.0%	0.930
Standardized Model 1 with Era Interactions	11.0%	0.939
Standardized Model 2 with Era Interactions	16.7%	0.855
Standardized Model 3 with Era Interactions	13.8%	0.920

```
#ggplot(new_batters, aes(retireRank, standardize(log(rWAR)), color = inducted)) + geo
m_point() + theme_bw()
```

```

ntrials <- 100
res1 <- NULL
res2 <- NULL
for (i in 1:ntrials) {
  # Make k-fold indeces
  rand_ind <- sample(1:dim(new_batters)[1], dim(new_batters)[1], replace = FALSE)
  sets <- NULL
  fold_size <- dim(new_batters)[1] / 5
  for (i in 1:5) {
    sets <- append(sets, list(rand_ind[((i-1)*fold_size + 1):(i*fold_size)]))
  }

  # Original Model
  obs <- rep(NA, dim(snew_batters)[1])
  for (i in 1:length(sets)) {
    mod <- glm(data = snew_batters[!sets[[i]],], formula = induced ~ retireRank*rWAR
, family = "binomial")
    temp <- cbind(snew_batters[sets[[i]],c("retireRank", "rWAR")], snew_batters[sets[
[i]],]$retireRank*snew_batters[sets[[i]],]$rWAR)
    obs[sets[[i]]] <- round(get_prob(temp, mod$coefficients), 0) == snew_batters[sets
[[i]],]$induced
  }
  res1 <- rbind(res1, obs)

  # New Model
  obs <- rep(NA, dim(snew_batters)[1])
  for (i in 1:length(sets)) {
    mod <- glm(data = snew_batters[!sets[[i]],], formula = induced ~ retireRank*rWAR
+ career_len + MVPs + SS + TC + RBI + winPer + vets_comm + def_pos, family = "binomia
l")
    temp <- cbind(snew_batters[sets[[i]],c("retireRank", "rWAR", "career_len", "MVPs"
, "SS", "TC", "RBI", "winPer", "vets_comm")], snew_batters[sets[[i]],]$def_pos == "2B
", snew_batters[sets[[i]],]$def_pos == "3B", snew_batters[sets[[i]],]$def_pos == "C",
snew_batters[sets[[i]],]$def_pos == "CF", snew_batters[sets[[i]],]$def_pos == "LF", s
new_batters[sets[[i]],]$def_pos == "RF", snew_batters[sets[[i]],]$def_pos == "SS", sn
ew_batters[sets[[i]],]$retireRank * snew_batters[sets[[i]],]$rWAR)
    obs[sets[[i]]] <- round(get_prob(temp, mod$coefficients), 0) == snew_batters[sets
[[i]],]$induced
  }
  res2 <- rbind(res2, obs)
}

```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Page 22 of 38

[illegible]

Page 24 of 38

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
snew_batters$mod1_accuracy <- apply(res1, 2, mean)
snew_batters$mod2_accuracy <- apply(res2, 2, mean)

g <- ggplot(snew_batters, aes(rWAR, inducted, color = mod2_accuracy, text = name)) +
  geom_jitter() + theme_bw()
ggplotly(g)
```

