NBA longitudinal efficiency analysis

Will Tackett, Jacob Vidergar

2024-12-15

```
library(dplyr)
library(ggplot2)
library(lme4)
library(lmerTest)
library(performance)
library(RColorBrewer)
library(sjPlot)
library(table1)
```

Exploratory Data Analysis

```
data <- read.csv("all_seasons.csv")
data <- data %>% select(-X)
head(data)
```

```
##        player_name team_abbreviation age player_height player_weight
## 1 Randy Livingston               HOU  22        193.04      94.80073
## 2 Gaylon Nickerson               WAS  28        190.50      86.18248
## 3     George Lynch               VAN  26        203.20     103.41898
## 4   George McCloud               LAL  30        203.20     102.05820
## 5     George Zidek               DEN  23        213.36     119.74829
## 6   Gerald Wilkins               ORL  33        198.12     102.05820
##                 college country draft_year draft_round draft_number gp  pts reb
## 1       Louisiana State     USA       1996           2           42 64  3.9 1.5
## 2 Northwestern Oklahoma     USA       1994           2           34  4  3.8 1.3
## 3        North Carolina     USA       1993           1           12 41  8.3 6.4
## 4         Florida State     USA       1989           1            7 64 10.2 2.8
## 5                  UCLA     USA       1995           1           22 52  2.8 1.7
## 6 Tennessee-Chattanooga     USA       1985           2           47 80 10.6 2.2
##   ast net_rating oreb_pct dreb_pct usg_pct ts_pct ast_pct  season
## 1 2.4        0.3    0.042    0.071   0.169  0.487   0.248 1996-97
## 2 0.3        8.9    0.030    0.111   0.174  0.497   0.043 1996-97
## 3 1.9       -8.2    0.106    0.185   0.175  0.512   0.125 1996-97
## 4 1.7       -2.7    0.027    0.111   0.206  0.527   0.125 1996-97
## 5 0.3      -14.1    0.102    0.169   0.195  0.500   0.064 1996-97
## 6 2.2       -5.8    0.031    0.064   0.203  0.503   0.143 1996-97
```

```
summary(data)
```

```
##  player_name        team_abbreviation       age        player_height
##  Length:12844       Length:12844       Min.   :18.00   Min.   :160.0
##  Class :character   Class :character   1st Qu.:24.00   1st Qu.:193.0
##  Mode  :character   Mode  :character   Median :26.00   Median :200.7
```

```
##                                       Mean   :27.05   Mean   :200.6
##                                       3rd Qu.:30.00   3rd Qu.:208.3
##                                       Max.   :44.00   Max.   :231.1
##  player_weight      college            country           draft_year
##  Min.   : 60.33   Length:12844       Length:12844       Length:12844
##  1st Qu.: 90.72   Class :character   Class :character   Class :character
##  Median : 99.79   Mode  :character   Mode  :character   Mode  :character
##  Mean   :100.26
##  3rd Qu.:108.86
##  Max.   :163.29
##  draft_round        draft_number            gp             pts
##  Length:12844       Length:12844       Min.   : 1.00   Min.   : 0.000
##  Class :character   Class :character   1st Qu.:31.00   1st Qu.: 3.600
##  Mode  :character   Mode  :character   Median :57.00   Median : 6.700
##                                        Mean   :51.15   Mean   : 8.213
##                                        3rd Qu.:73.00   3rd Qu.:11.500
##                                        Max.   :85.00   Max.   :36.100
##       reb              ast           net_rating          oreb_pct
##  Min.   : 0.000   Min.   : 0.000   Min.   :-250.000   Min.   :0.00000
##  1st Qu.: 1.800   1st Qu.: 0.600   1st Qu.:  -6.400   1st Qu.:0.02100
##  Median : 3.000   Median : 1.200   Median :  -1.300   Median :0.04000
##  Mean   : 3.558   Mean   : 1.825   Mean   :  -2.226   Mean   :0.05407
##  3rd Qu.: 4.700   3rd Qu.: 2.400   3rd Qu.:   3.200   3rd Qu.:0.08300
##  Max.   :16.300   Max.   :11.700   Max.   : 300.000   Max.   :1.00000
##     dreb_pct         usg_pct           ts_pct          ast_pct
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0960   1st Qu.:0.1490   1st Qu.:0.4820   1st Qu.:0.0660
##  Median :0.1305   Median :0.1810   Median :0.5250   Median :0.1030
##  Mean   :0.1406   Mean   :0.1846   Mean   :0.5131   Mean   :0.1316
##  3rd Qu.:0.1790   3rd Qu.:0.2170   3rd Qu.:0.5630   3rd Qu.:0.1790
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.5000   Max.   :1.0000
##     season
##  Length:12844
##  Class :character
##  Mode  :character
##
##
##
```

```
str(data)
```

```
## 'data.frame':    12844 obs. of  21 variables:
##  $ player_name      : chr  "Randy Livingston" "Gaylon Nickerson" "George Lynch" "George McCloud" ...
##  $ team_abbreviation: chr  "HOU" "WAS" "VAN" "LAL" ...
##  $ age              : num  22 28 26 30 23 33 26 30 24 24 ...
##  $ player_height    : num  193 190 203 203 213 ...
##  $ player_weight    : num  94.8 86.2 103.4 102.1 119.7 ...
##  $ college          : chr  "Louisiana State" "Northwestern Oklahoma" "North Carolina" "Florida State"
##  $ country          : chr  "USA" "USA" "USA" "USA" ...
##  $ draft_year       : chr  "1996" "1994" "1993" "1989" ...
##  $ draft_round      : chr  "2" "2" "1" "1" ...
##  $ draft_number     : chr  "42" "34" "12" "7" ...
##  $ gp               : int  64 4 41 64 52 80 73 79 80 80 ...
##  $ pts              : num  3.9 3.8 8.3 10.2 2.8 10.6 10.6 26.8 21.1 21.4 ...
##  $ reb              : num  1.5 1.3 6.4 2.8 1.7 2.2 6.6 4 6.3 9 ...
```

```
##  $ ast              : num  2.4 0.3 1.9 1.7 0.3 2.2 0.4 2 3.1 7.3 ...
##  $ net_rating       : num  0.3 8.9 -8.2 -2.7 -14.1 -5.8 6.9 3.2 -2.9 6.9 ...
##  $ oreb_pct         : num  0.042 0.03 0.106 0.027 0.102 0.031 0.098 0.025 0.051 0.049 ...
##  $ dreb_pct         : num  0.071 0.111 0.185 0.111 0.169 0.064 0.217 0.087 0.144 0.232 ...
##  $ usg_pct          : num  0.169 0.174 0.175 0.206 0.195 0.203 0.185 0.272 0.278 0.283 ...
##  $ ts_pct           : num  0.487 0.497 0.512 0.527 0.5 0.503 0.618 0.605 0.528 0.556 ...
##  $ ast_pct          : num  0.248 0.043 0.125 0.125 0.064 0.143 0.024 0.088 0.146 0.356 ...
##  $ season           : chr  "1996-97" "1996-97" "1996-97" "1996-97" ...
```

Filter data

```
filtered_data <- data %>%
  # Add a column for games played in the season
  mutate(gp_season = case_when(
    season == "1998-99" ~ 50,
    season == "2011-12" ~ 66,
    TRUE ~ 82
  )) %>%
  # adjust naming of draft round categories
  mutate(draft_round_combined = case_when(
    draft_round == 0 ~ "Undrafted",
    draft_round == "Undrafted" ~ "Undrafted",
    draft_round %in% c(1, 2) ~ as.character(draft_round),
    TRUE ~ NA_character_
  )) %>%
  # Filter for relevant draft round categories
  filter(draft_round_combined %in% c("Undrafted", "1", "2")) %>%
  # Calculate the percentage of games played
  mutate(gp_pct = gp / gp_season) %>%
  # Filter for players who played at least 50% of games
  filter(gp_pct > 0.5)

# make season continuous
filtered_data <- filtered_data %>%
  mutate(season_continuous = as.numeric(substr(season, 1, 4)))

# Create a new column combining 2 and Undrafted into one category
filtered_data <- filtered_data %>%
  mutate(draft_round_combined_new = ifelse(draft_round_combined %in% c("2", "Undrafted"),
                                           "2_or_Undrafted",
                                           draft_round_combined))
```

```
# convert important variables from chr to factor
filtered_data$draft_round_combined <- as.factor(filtered_data$draft_round_combined)
# filtered_data$season <- as.factor(filtered_data$season)
filtered_data$player_name <- as.factor(filtered_data$player_name)
```

```
# set min threshold for games played (50%)
print("before threshold")
```

```
## [1] "before threshold"
```

```
# Count the number of observations for each level
table(filtered_data$draft_round_combined)
```

```
##
##         1         2 Undrafted
##      5795      1851      1060
```

```
print("after threshold")
```

```
## [1] "after threshold"
```

```
# Keep observations where gp / gp_season >= 0.5
filtered_data <- filtered_data %>%
  filter(gp / gp_season >= 0.5)

# Count the number of observations for each level
table(filtered_data$draft_round_combined)
```

```
##
##         1         2 Undrafted
##      5795      1851      1060
```

```
# Prepare the data for Table 1
table_data <- filtered_data %>%
  select(draft_round_combined, ts_pct, player_height, player_weight, age, gp_pct) %>%
  mutate(draft_round_combined = factor(
    draft_round_combined,
    levels = c("1", "2", "Undrafted"),
    labels = c("1st Round", "2nd Round", "Undrafted")
  ))

# Apply variable labels using label()
label(table_data$draft_round_combined) <- "Draft Round"
label(table_data$ts_pct) <- "True Shooting Percentage"
label(table_data$player_height) <- "Height (inches)"
label(table_data$player_weight) <- "Weight (lbs)"
label(table_data$age) <- "Age"
label(table_data$gp_pct) <- "Games Played (%)"

# Create the Table 1
table1(
  ~ ts_pct + player_height + player_weight + age + gp_pct | draft_round_combined,
  data = table_data,
  overall = "Overall",
  render = function(x, name, ...) {
    if (is.numeric(x)) {
      # Customize numeric summaries: Mean (SD)
      sprintf("%0.2f (%0.2f)", mean(x, na.rm = TRUE), sd(x, na.rm = TRUE))
    } else {
      # Show counts for categorical variables
      table1::render.default(x, name, ...)
    }
  }
)
```

	1st Round	2nd Round	Undrafted	Overall
	(N=5795)	(N=1851)	(N=1060)	(N=8706)
True Shooting Percentage				
	0.54 (0.05)	0.54 (0.06)	0.53 (0.06)	0.54 (0.05)
Height (inches)				
	201.31 (9.04)	200.02 (8.51)	197.18 (9.29)	200.53 (9.06)
Weight (lbs)				
	101.13 (12.33)	100.23 (12.08)	96.53 (11.66)	100.38 (12.29)
Age				
	27.03 (4.50)	27.30 (3.88)	28.10 (3.41)	27.22 (4.27)
Games Played (%)				
	0.84 (0.14)	0.81 (0.15)	0.79 (0.15)	0.82 (0.14)

Preliminary Plots

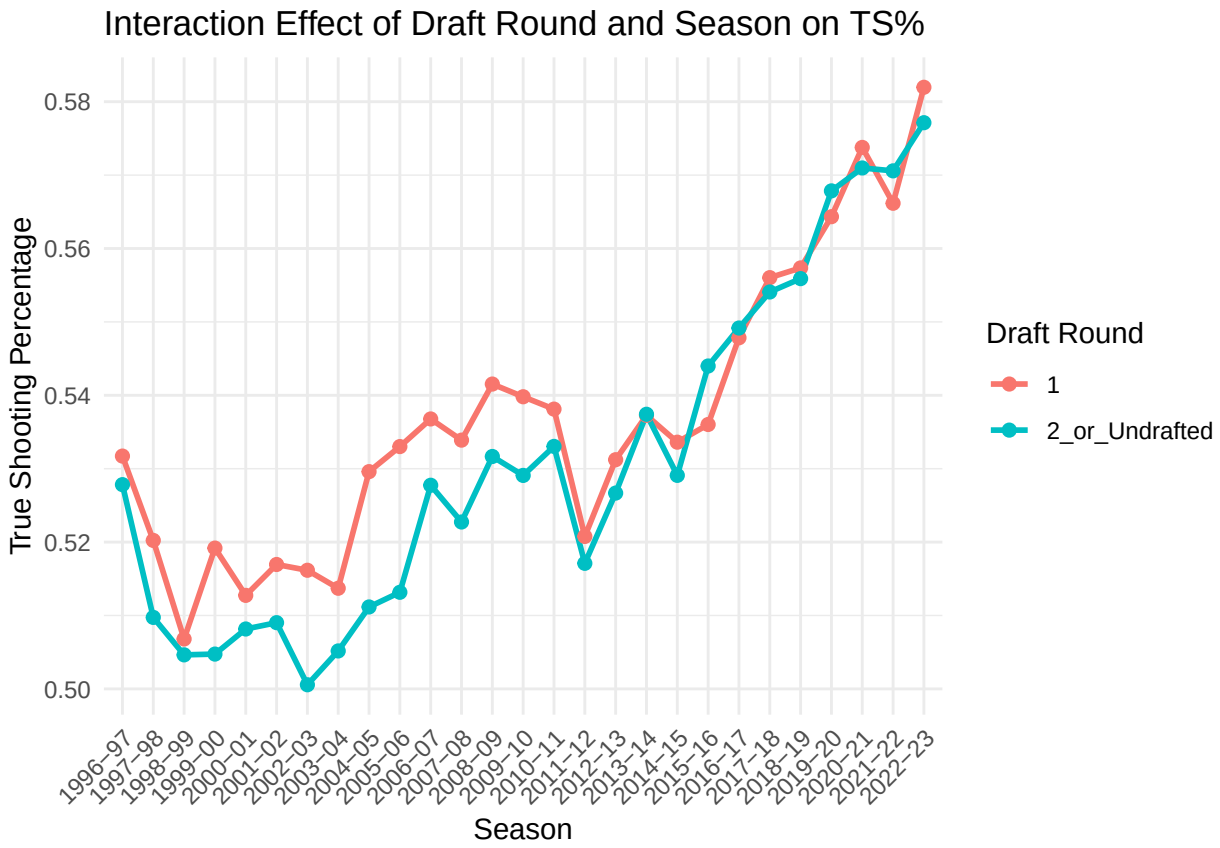
```
# Average `ts_pct` by the new combined draft_round category and season
draft_round_ts_pct_combined <- filtered_data %>%
    group_by(draft_round_combined, season) %>%
    summarize(avg_ts_pct = mean(ts_pct, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'draft_round_combined'. You can override
## using the `.groups` argument.
```

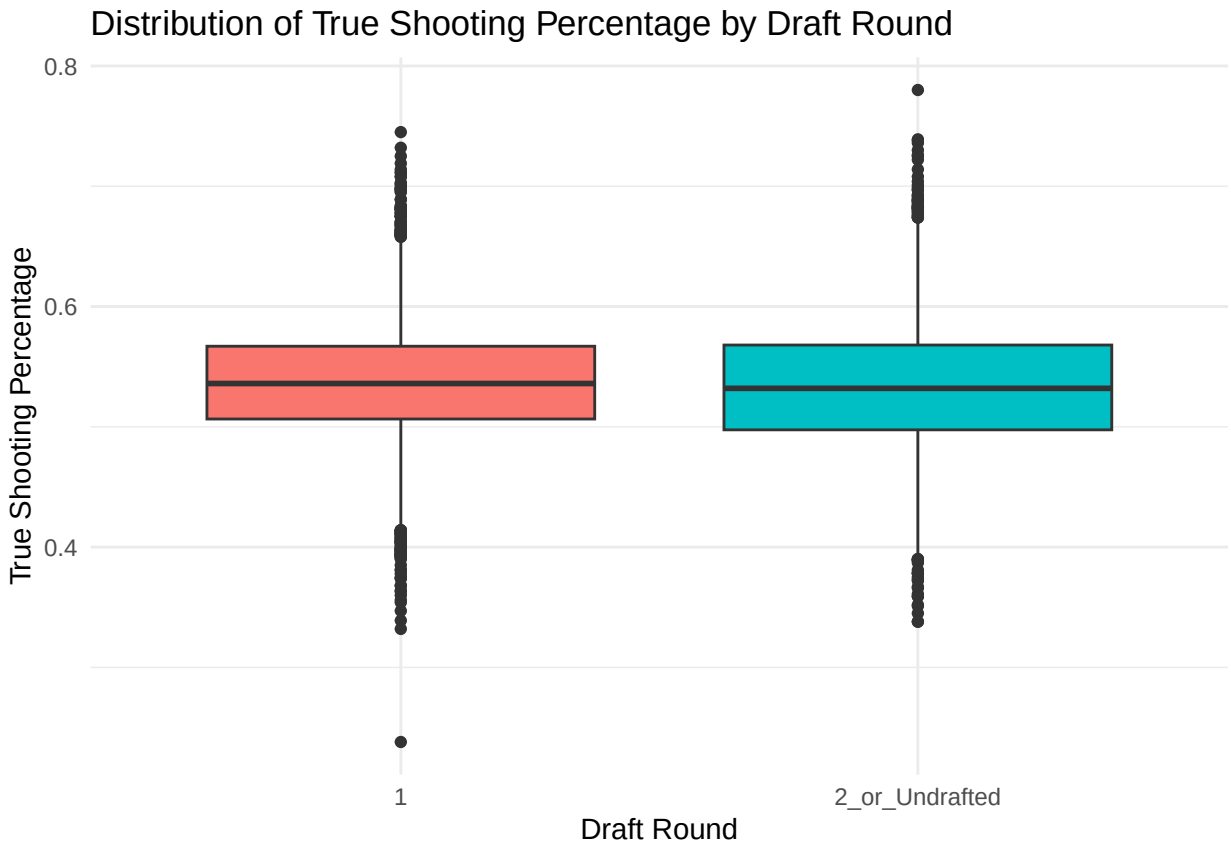
```
# Ensure `season` is treated as a factor with the correct order
draft_round_ts_pct_combined$season <- factor(draft_round_ts_pct_combined$season, levels = sort(unique(d
```

```
ggplot(filtered_data, aes(x = season, y = ts_pct, color = draft_round_combined_new, group = draft_round
  stat_summary(fun = mean, geom = "line", size = 1) + # Line plot for mean TS% by season
  stat_summary(fun = mean, geom = "point", size = 2) + # Points for clarity
  labs(
    title = "Interaction Effect of Draft Round and Season on TS%",
    x = "Season",
    y = "True Shooting Percentage",
    color = "Draft Round"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate x-axis labels
```

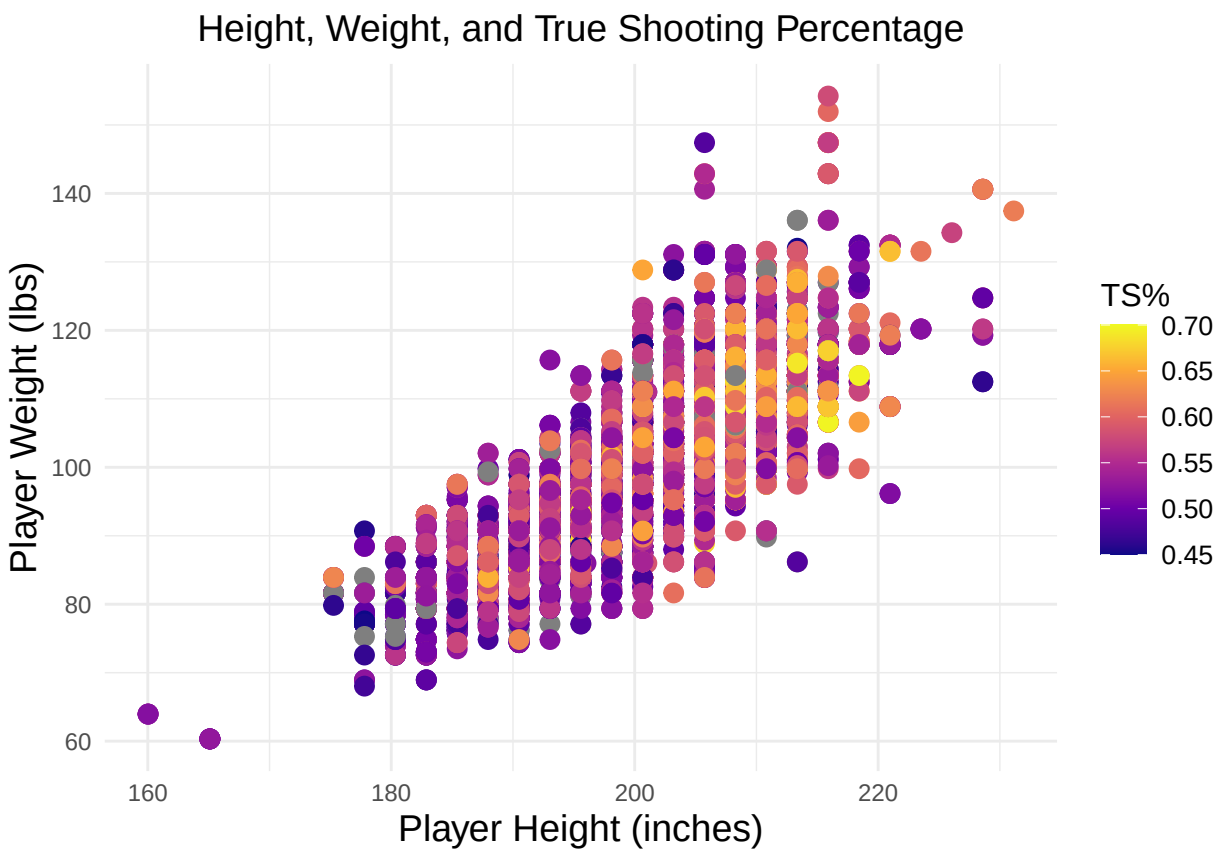
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```r
ggplot(filtered_data, aes(x = draft_round_combined_new, y = ts_pct, fill = draft_round_combined_new)) +
  geom_boxplot() +
  labs(
    title = "Distribution of True Shooting Percentage by Draft Round",
    x = "Draft Round",
    y = "True Shooting Percentage"
  ) +
  theme_minimal() +
  theme(legend.position = "none")
```




```r
ggplot(filtered_data, aes(x = player_height, y = player_weight, color = ts_pct)) +
  geom_point(alpha = 1, size = 3) +
  scale_color_viridis_c(
    name = "TS%",
    option = "plasma",         # Use a vibrant "plasma" color scheme
    limits = c(0.45, 0.7),     # Adjust the range for TS%
    breaks = seq(0.45, 0.7, by = 0.05)  # Add meaningful breaks
  ) +
  labs(
    title = "Height, Weight, and True Shooting Percentage",
    x = "Player Height (inches)",
    y = "Player Weight (lbs)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 14, hjust = 0.5),
    axis.title = element_text(size = 14),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10)
  )
```



Modeling

```
# linear_model <- lm(ts_pct ~ draft_round_combined, data = filtered_data)
# summary(linear_model)

# Fit a mixed effects model where:
# - `season` is modeled as a fixed effect to assess the change in `ts_pct` over time.
# - `draft_round_combined_new` is modeled as a fixed effect to adjust for round differences.
# - A random intercept for `player_id` to account for player-specific variability.

model <- lmer(ts_pct ~ season * draft_round_combined + player_height + player_weight + age + (1 | playe
              data = filtered_data)

# View the model summary
summary(model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## ts_pct ~ season * draft_round_combined + player_height + player_weight +
##     age + (1 | player_name)
##    Data: filtered_data
##
## REML criterion at convergence: -30235.4
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -5.9211 -0.5548  0.0230  0.5917  4.4404
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  player_name (Intercept) 0.001122 0.03350
##  Residual                0.001250 0.03535
## Number of obs: 8706, groups:  player_name, 1654
##
## Fixed effects:
##                                              Estimate Std. Error         df
## (Intercept)                                 3.517e-01  2.613e-02  1.991e+03
## season1997-98                              -1.443e-02  3.617e-03  7.268e+03
## season1998-99                              -2.728e-02  3.615e-03  7.405e+03
## season1999-00                              -1.561e-02  3.612e-03  7.481e+03
## season2000-01                              -2.184e-02  3.712e-03  7.612e+03
## season2001-02                              -1.923e-02  3.701e-03  7.750e+03
## season2002-03                              -1.959e-02  3.755e-03  7.915e+03
## season2003-04                              -2.137e-02  3.785e-03  8.032e+03
## season2004-05                              -8.023e-03  3.803e-03  8.163e+03
## season2005-06                              -4.274e-03  3.865e-03  8.322e+03
## season2006-07                              -2.324e-05  3.875e-03  8.410e+03
## season2007-08                              -2.225e-03  3.916e-03  8.499e+03
## season2008-09                               2.872e-03  3.944e-03  8.564e+03
## season2009-10                               1.276e-03  3.948e-03  8.608e+03
## season2010-11                              -1.760e-03  3.959e-03  8.621e+03
## season2011-12                              -1.830e-02  3.973e-03  8.595e+03
## season2012-13                              -9.340e-03  4.002e-03  8.525e+03
## season2013-14                              -3.541e-03  4.061e-03  8.450e+03
## season2014-15                              -8.106e-03  4.088e-03  8.341e+03
## season2015-16                              -4.248e-03  4.074e-03  8.118e+03
## season2016-17                               7.549e-03  4.112e-03  7.948e+03
## season2017-18                               1.671e-02  4.193e-03  7.810e+03
## season2018-19                               1.817e-02  4.196e-03  7.496e+03
## season2019-20                               2.773e-02  4.360e-03  7.478e+03
## season2020-21                               3.646e-02  4.366e-03  7.166e+03
## season2021-22                               3.236e-02  4.329e-03  6.718e+03
## season2022-23                               4.758e-02  4.343e-03  6.398e+03
## draft_round_combined2                       3.622e-03  6.115e-03  7.413e+03
## draft_round_combinedUndrafted               8.994e-03  8.576e-03  7.415e+03
## player_height                               7.566e-04  1.636e-04  2.415e+03
## player_weight                               3.102e-04  1.138e-04  3.272e+03
## age                                        -1.923e-04  1.355e-04  4.299e+03
## season1997-98:draft_round_combined2        -5.320e-03  7.368e-03  7.486e+03
## season1998-99:draft_round_combined2         6.098e-03  7.556e-03  7.816e+03
## season1999-00:draft_round_combined2        -1.377e-02  7.633e-03  7.628e+03
## season2000-01:draft_round_combined2         2.398e-04  7.591e-03  7.906e+03
## season2001-02:draft_round_combined2        -7.447e-03  7.817e-03  8.092e+03
## season2002-03:draft_round_combined2        -7.273e-03  7.886e-03  8.181e+03
## season2003-04:draft_round_combined2        -9.629e-03  7.720e-03  8.258e+03
## season2004-05:draft_round_combined2        -1.437e-02  7.692e-03  8.414e+03
## season2005-06:draft_round_combined2        -2.187e-02  7.961e-03  8.461e+03
## season2006-07:draft_round_combined2        -1.478e-02  7.946e-03  8.558e+03
## season2007-08:draft_round_combined2        -1.452e-02  8.000e-03  8.566e+03
```

```
## season2008-09:draft_round_combined2         -7.946e-03  8.060e-03  8.615e+03
## season2009-10:draft_round_combined2         -1.057e-02  8.035e-03  8.621e+03
## season2010-11:draft_round_combined2         -6.331e-03  8.073e-03  8.622e+03
## season2011-12:draft_round_combined2         -4.374e-03  8.103e-03  8.615e+03
## season2012-13:draft_round_combined2         -7.053e-03  8.071e-03  8.607e+03
## season2013-14:draft_round_combined2          7.047e-04  8.103e-03  8.588e+03
## season2014-15:draft_round_combined2         -7.933e-03  8.068e-03  8.564e+03
## season2015-16:draft_round_combined2         -1.410e-04  8.157e-03  8.581e+03
## season2016-17:draft_round_combined2          1.632e-03  8.236e-03  8.526e+03
## season2017-18:draft_round_combined2         -7.392e-03  8.169e-03  8.439e+03
## season2018-19:draft_round_combined2         -4.519e-03  8.150e-03  8.376e+03
## season2019-20:draft_round_combined2         -1.786e-04  8.420e-03  8.404e+03
## season2020-21:draft_round_combined2         -4.174e-03  8.536e-03  8.342e+03
## season2021-22:draft_round_combined2         -9.320e-04  8.304e-03  8.066e+03
## season2022-23:draft_round_combined2         -5.332e-03  8.405e-03  7.899e+03
## season1997-98:draft_round_combinedUndrafted -8.380e-03  1.039e-02  8.220e+03
## season1998-99:draft_round_combinedUndrafted -6.046e-03  1.054e-02  8.324e+03
## season1999-00:draft_round_combinedUndrafted -9.089e-03  1.053e-02  8.333e+03
## season2000-01:draft_round_combinedUndrafted -1.356e-02  1.043e-02  8.393e+03
## season2001-02:draft_round_combinedUndrafted -1.225e-02  1.089e-02  8.483e+03
## season2002-03:draft_round_combinedUndrafted -2.840e-02  1.057e-02  8.590e+03
## season2003-04:draft_round_combinedUndrafted -1.878e-02  1.082e-02  8.529e+03
## season2004-05:draft_round_combinedUndrafted -2.409e-02  1.079e-02  8.613e+03
## season2005-06:draft_round_combinedUndrafted -1.158e-02  1.064e-02  8.621e+03
## season2006-07:draft_round_combinedUndrafted -6.960e-03  1.064e-02  8.605e+03
## season2007-08:draft_round_combinedUndrafted -1.378e-02  1.085e-02  8.612e+03
## season2008-09:draft_round_combinedUndrafted -1.752e-02  1.080e-02  8.602e+03
## season2009-10:draft_round_combinedUndrafted -2.087e-02  1.098e-02  8.603e+03
## season2010-11:draft_round_combinedUndrafted -1.604e-02  1.113e-02  8.577e+03
## season2011-12:draft_round_combinedUndrafted -1.821e-02  1.126e-02  8.528e+03
## season2012-13:draft_round_combinedUndrafted -1.125e-02  1.126e-02  8.524e+03
## season2013-14:draft_round_combinedUndrafted -8.556e-03  1.114e-02  8.394e+03
## season2014-15:draft_round_combinedUndrafted -5.043e-03  1.086e-02  8.359e+03
## season2015-16:draft_round_combinedUndrafted  6.859e-03  1.121e-02  8.391e+03
## season2016-17:draft_round_combinedUndrafted -1.209e-02  1.113e-02  8.169e+03
## season2017-18:draft_round_combinedUndrafted -1.825e-03  1.111e-02  8.133e+03
## season2018-19:draft_round_combinedUndrafted -1.325e-02  1.086e-02  8.004e+03
## season2019-20:draft_round_combinedUndrafted -1.091e-02  1.097e-02  7.984e+03
## season2020-21:draft_round_combinedUndrafted -9.529e-03  1.096e-02  7.804e+03
## season2021-22:draft_round_combinedUndrafted -7.636e-03  1.080e-02  7.620e+03
## season2022-23:draft_round_combinedUndrafted -2.588e-02  1.090e-02  7.556e+03
##                                             t value Pr(>|t|)
## (Intercept)                                  13.463  < 2e-16 ***
## season1997-98                                -3.989 6.69e-05 ***
## season1998-99                                -7.546 5.03e-14 ***
## season1999-00                                -4.320 1.58e-05 ***
## season2000-01                                -5.885 4.15e-09 ***
## season2001-02                                -5.195 2.10e-07 ***
## season2002-03                                -5.217 1.86e-07 ***
## season2003-04                                -5.644 1.71e-08 ***
## season2004-05                                -2.110  0.03491 *
## season2005-06                                -1.106  0.26886
## season2006-07                                -0.006  0.99522
## season2007-08                                -0.568  0.57000
```

```
## season2008-09                                    0.728  0.46648
## season2009-10                                    0.323  0.74661
## season2010-11                                   -0.445  0.65657
## season2011-12                                   -4.605 4.19e-06 ***
## season2012-13                                   -2.334  0.01962 *
## season2013-14                                   -0.872  0.38320
## season2014-15                                   -1.983  0.04741 *
## season2015-16                                   -1.043  0.29713
## season2016-17                                    1.836  0.06643 .
## season2017-18                                    3.985 6.82e-05 ***
## season2018-19                                    4.330 1.51e-05 ***
## season2019-20                                    6.360 2.13e-10 ***
## season2020-21                                    8.351  < 2e-16 ***
## season2021-22                                    7.476 8.65e-14 ***
## season2022-23                                   10.957  < 2e-16 ***
## draft_round_combined2                            0.592  0.55361
## draft_round_combinedUndrafted                    1.049  0.29432
## player_height                                    4.624 3.97e-06 ***
## player_weight                                    2.726  0.00645 **
## age                                             -1.420  0.15577
## season1997-98:draft_round_combined2             -0.722  0.47032
## season1998-99:draft_round_combined2              0.807  0.41967
## season1999-00:draft_round_combined2             -1.804  0.07123 .
## season2000-01:draft_round_combined2              0.032  0.97480
## season2001-02:draft_round_combined2             -0.953  0.34075
## season2002-03:draft_round_combined2             -0.922  0.35638
## season2003-04:draft_round_combined2             -1.247  0.21235
## season2004-05:draft_round_combined2             -1.868  0.06174 .
## season2005-06:draft_round_combined2             -2.747  0.00602 **
## season2006-07:draft_round_combined2             -1.860  0.06293 .
## season2007-08:draft_round_combined2             -1.816  0.06945 .
## season2008-09:draft_round_combined2             -0.986  0.32423
## season2009-10:draft_round_combined2             -1.315  0.18849
## season2010-11:draft_round_combined2             -0.784  0.43293
## season2011-12:draft_round_combined2             -0.540  0.58934
## season2012-13:draft_round_combined2             -0.874  0.38224
## season2013-14:draft_round_combined2              0.087  0.93070
## season2014-15:draft_round_combined2             -0.983  0.32552
## season2015-16:draft_round_combined2             -0.017  0.98621
## season2016-17:draft_round_combined2              0.198  0.84291
## season2017-18:draft_round_combined2             -0.905  0.36559
## season2018-19:draft_round_combined2             -0.554  0.57931
## season2019-20:draft_round_combined2             -0.021  0.98307
## season2020-21:draft_round_combined2             -0.489  0.62490
## season2021-22:draft_round_combined2             -0.112  0.91064
## season2022-23:draft_round_combined2             -0.634  0.52587
## season1997-98:draft_round_combinedUndrafted     -0.806  0.42000
## season1998-99:draft_round_combinedUndrafted     -0.574  0.56616
## season1999-00:draft_round_combinedUndrafted     -0.863  0.38816
## season2000-01:draft_round_combinedUndrafted     -1.300  0.19352
## season2001-02:draft_round_combinedUndrafted     -1.125  0.26051
## season2002-03:draft_round_combinedUndrafted     -2.688  0.00721 **
## season2003-04:draft_round_combinedUndrafted     -1.735  0.08276 .
## season2004-05:draft_round_combinedUndrafted     -2.232  0.02561 *
```

```
## season2005-06:draft_round_combinedUndrafted  -1.088  0.27674
## season2006-07:draft_round_combinedUndrafted  -0.654  0.51303
## season2007-08:draft_round_combinedUndrafted  -1.270  0.20401
## season2008-09:draft_round_combinedUndrafted  -1.622  0.10474
## season2009-10:draft_round_combinedUndrafted  -1.901  0.05739 .
## season2010-11:draft_round_combinedUndrafted  -1.442  0.14946
## season2011-12:draft_round_combinedUndrafted  -1.618  0.10577
## season2012-13:draft_round_combinedUndrafted  -0.999  0.31777
## season2013-14:draft_round_combinedUndrafted  -0.768  0.44266
## season2014-15:draft_round_combinedUndrafted  -0.464  0.64249
## season2015-16:draft_round_combinedUndrafted   0.612  0.54062
## season2016-17:draft_round_combinedUndrafted  -1.086  0.27740
## season2017-18:draft_round_combinedUndrafted  -0.164  0.86954
## season2018-19:draft_round_combinedUndrafted  -1.220  0.22248
## season2019-20:draft_round_combinedUndrafted  -0.995  0.31996
## season2020-21:draft_round_combinedUndrafted  -0.870  0.38444
## season2021-22:draft_round_combinedUndrafted  -0.707  0.47954
## season2022-23:draft_round_combinedUndrafted  -2.373  0.01766 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation matrix not shown by default, as p = 84 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

Try season as continuous variable

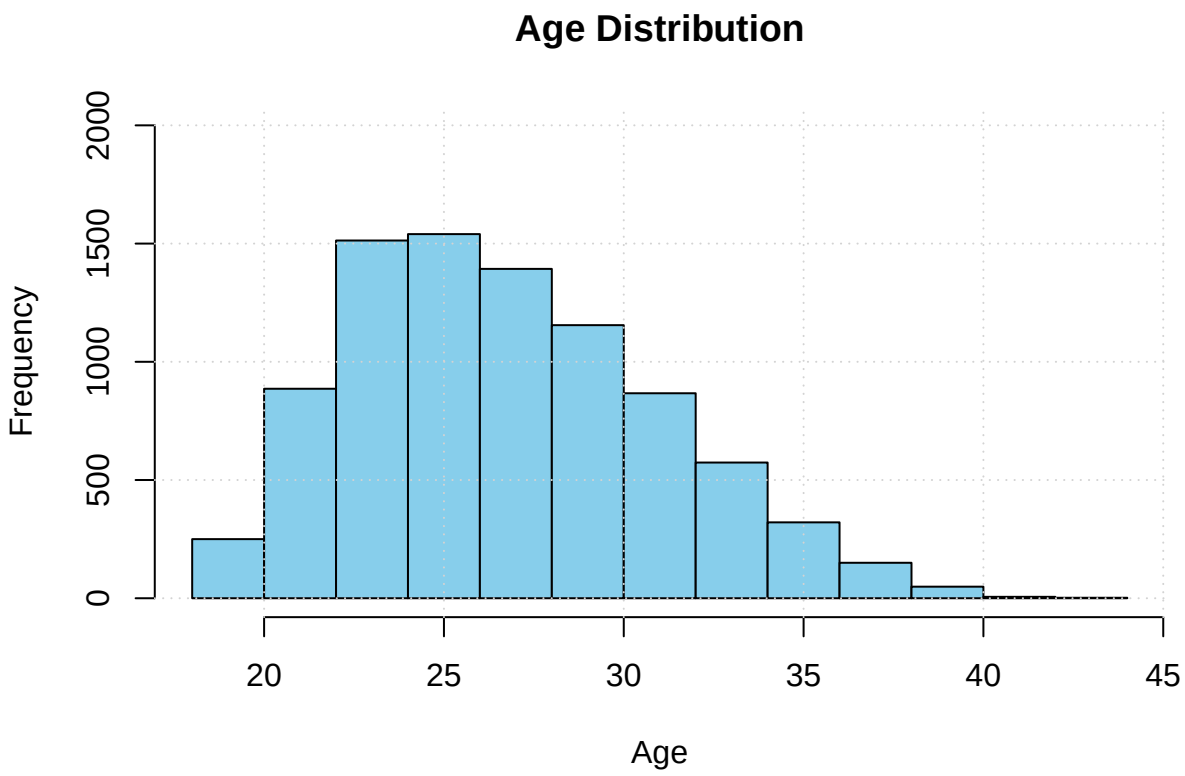
```
filtered_data <- filtered_data %>%
  mutate(season_continuous = as.numeric(substr(season, 1, 4)))

model <- lmer(ts_pct ~ season_continuous * draft_round_combined + player_height + player_weight + age +

# View the model summary
summary(model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: ts_pct ~ season_continuous * draft_round_combined + player_height +
##     player_weight + age + (1 | player_name)
##    Data: filtered_data
##
## REML criterion at convergence: -30251.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.9311 -0.5565  0.0277  0.6044  4.3320
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  player_name (Intercept) 0.001117 0.03343
##  Residual                0.001348 0.03671
## Number of obs: 8706, groups:  player_name, 1654
##
## Fixed effects:
##                                                 Estimate Std. Error
```

```
## (Intercept)                                        -3.612e+00  2.422e-01
## season_continuous                                   1.990e-03  1.203e-04
## draft_round_combined2                              -2.757e-01  4.532e-01
## draft_round_combinedUndrafted                      -1.073e-01  5.611e-01
## player_height                                       5.732e-04  1.641e-04
## player_weight                                       3.591e-04  1.148e-04
## age                                                -2.342e-04  1.378e-04
## season_continuous:draft_round_combined2             1.362e-04  2.255e-04
## season_continuous:draft_round_combinedUndrafted     5.185e-05  2.791e-04
##                                                           df t value Pr(>|t|)
## (Intercept)                                         2.298e+03 -14.911  < 2e-16 ***
## season_continuous                                   2.276e+03  16.544  < 2e-16 ***
## draft_round_combined2                               3.784e+03  -0.608 0.542959
## draft_round_combinedUndrafted                       3.167e+03  -0.191 0.848376
## player_height                                       2.430e+03   3.493 0.000485 ***
## player_weight                                       3.207e+03   3.129 0.001770 **
## age                                                 4.395e+03  -1.699 0.089359 .
## season_continuous:draft_round_combined2             3.785e+03   0.604 0.546028
## season_continuous:draft_round_combinedUndrafted     3.168e+03   0.186 0.852627
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) ssn_cn drf__2 drf__U plyr_h plyr_w age    s_:__2
## seasn_cntns -0.994
## drft_rnd_c2 -0.483  0.484
## drft_rnd_cU -0.397  0.400  0.201
## player_hght -0.108  0.006  0.015 -0.012
## player_wght  0.025  0.034 -0.006  0.024 -0.758
## age          0.373 -0.399 -0.090 -0.099  0.156 -0.221
## ssn_cnt:__2  0.483 -0.484 -1.000 -0.201 -0.015  0.006  0.091
## ssn_cnt:__U  0.397 -0.400 -0.201 -1.000  0.013 -0.024  0.099  0.201
```

Histogram for age

```r
# Basic histogram
hist(filtered_data$age,
     breaks = 10, # Number of bins
     col = "skyblue", # Bar color
     border = "black", # Border color
     main = "Age Distribution", # Title
     xlab = "Age", # X-axis label
     ylab = "Frequency", # Y-axis label
     ylim = c(0, 2000)) # Adjust Y-axis limits for better view

# Add a grid for better readability (optional)
grid()
```



Recode age based on histogram

```
# Create a new categorical variable based on age
filtered_data$career_stage <- cut(filtered_data$age,
                                  breaks = c(-Inf, 25, 30, Inf), # Define age ranges
                                  labels = c("Rookie", "Mid-Career", "Veteran"), # Labels
                                  right = TRUE) # Include the upper boundary in each interval

# Check the distribution of the new variable
table(filtered_data$career_stage)
```

```
##
##     Rookie Mid-Career    Veteran
##       3436       3301       1969
```

New model with career_stage instead of age

```
model <- lmer(ts_pct ~ season_continuous * draft_round_combined + player_height + player_weight + caree

# View the model summary
summary(model)
```

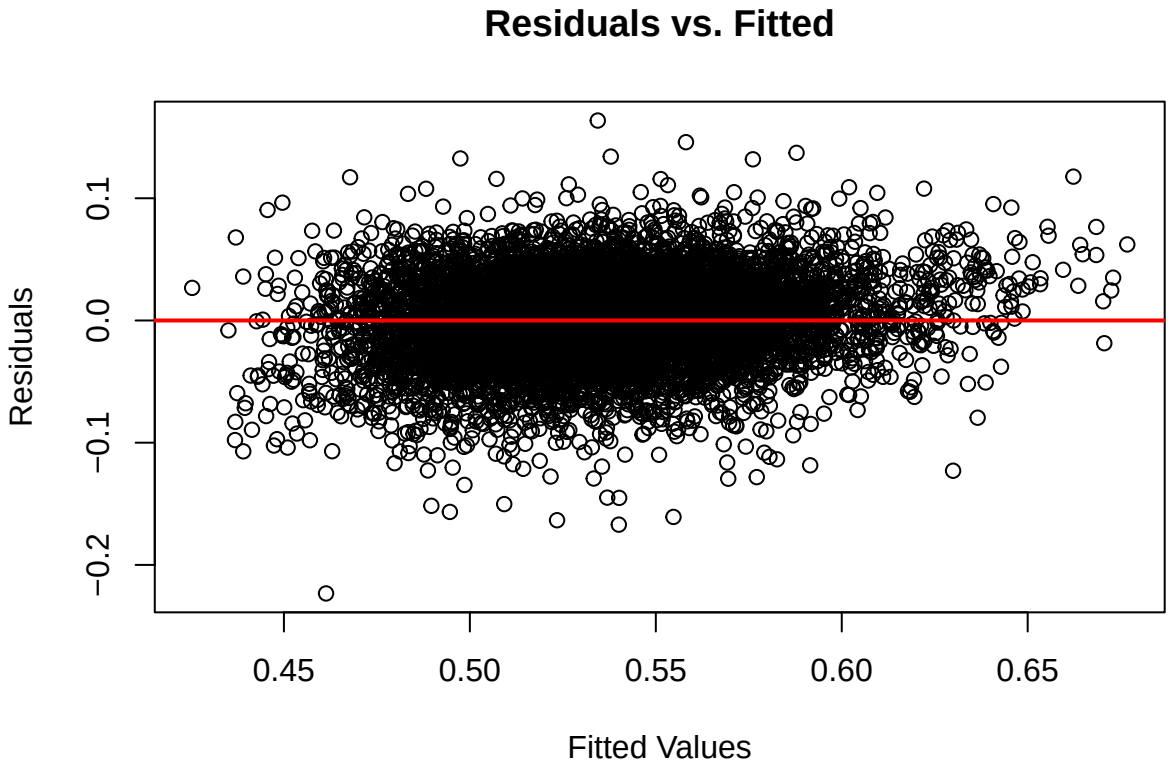
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: ts_pct ~ season_continuous * draft_round_combined + player_height +
##     player_weight + career_stage + (1 | player_name)
##    Data: filtered_data
##
## REML criterion at convergence: -30364.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -6.1301 -0.5550  0.0284  0.5999  4.4906
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  player_name (Intercept) 0.001114 0.03337
##  Residual                0.001327 0.03643
## Number of obs: 8706, groups:  player_name, 1654
##
## Fixed effects:
##                                                Estimate Std. Error
## (Intercept)                                  -3.631e+00  2.377e-01
## season_continuous                             1.992e-03  1.173e-04
## draft_round_combined2                        -3.412e-01  4.500e-01
## draft_round_combinedUndrafted                -2.849e-01  5.577e-01
## player_height                                 6.482e-04  1.630e-04
## player_weight                                 2.895e-04  1.136e-04
## career_stageMid-Career                        7.759e-03  1.057e-03
## career_stageVeteran                          -4.419e-03  1.425e-03
## season_continuous:draft_round_combined2       1.685e-04  2.239e-04
## season_continuous:draft_round_combinedUndrafted  1.392e-04  2.774e-04
##                                                      df t value Pr(>|t|)
## (Intercept)                                   2.801e+03 -15.274  < 2e-16 ***
## season_continuous                             2.851e+03  16.981  < 2e-16 ***
## draft_round_combined2                         3.885e+03  -0.758  0.44839
## draft_round_combinedUndrafted                 3.245e+03  -0.511  0.60946
## player_height                                 2.476e+03   3.978 7.15e-05 ***
## player_weight                                 3.318e+03   2.548  0.01087 *
## career_stageMid-Career                        8.665e+03   7.338 2.36e-13 ***
## career_stageVeteran                           6.411e+03  -3.102  0.00193 **
## season_continuous:draft_round_combined2       3.885e+03   0.752  0.45194
## season_continuous:draft_round_combinedUndrafted  3.246e+03   0.502  0.61581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) ssn_cn drf__2 drf__U plyr_h plyr_w cr_M-C crr_sV s_:__2
## seasn_cntns -0.994
## drft_rnd_c2 -0.477  0.479
## drft_rnd_cU -0.392  0.396  0.198
## player_hght -0.125  0.023  0.021 -0.008
## player_wght  0.047  0.010 -0.015  0.018 -0.756
## crr_stgMd-C  0.207 -0.219 -0.048 -0.073  0.115 -0.161
## crr_stgVtrn  0.339 -0.353 -0.061 -0.077  0.127 -0.180  0.541
## ssn_cnt:__2  0.477 -0.479 -1.000 -0.198 -0.021  0.014  0.048  0.061
## ssn_cnt:__U  0.392 -0.396 -0.198 -1.000  0.009 -0.018  0.073  0.077  0.198
```

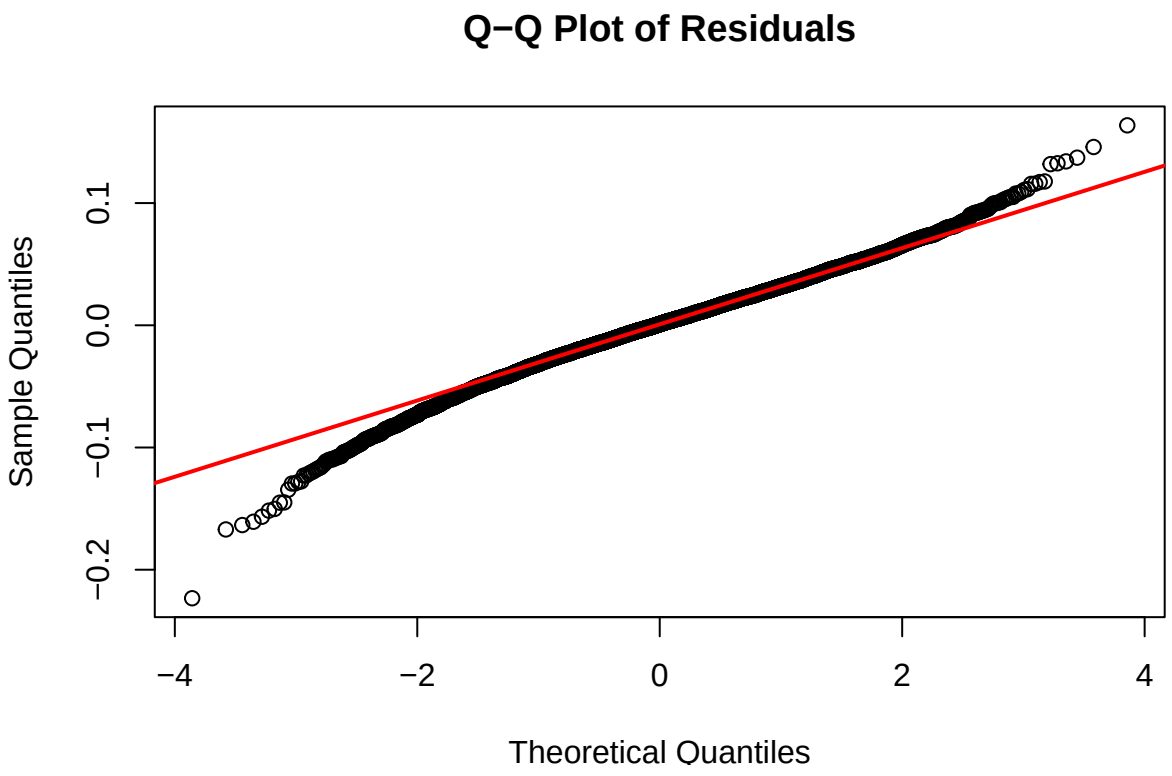
Residual Diagnostics

```r
# Extract residuals
residuals <- residuals(model)

# Plot residuals vs. fitted values
plot(fitted(model), residuals,
     main = "Residuals vs. Fitted",
     xlab = "Fitted Values",
```

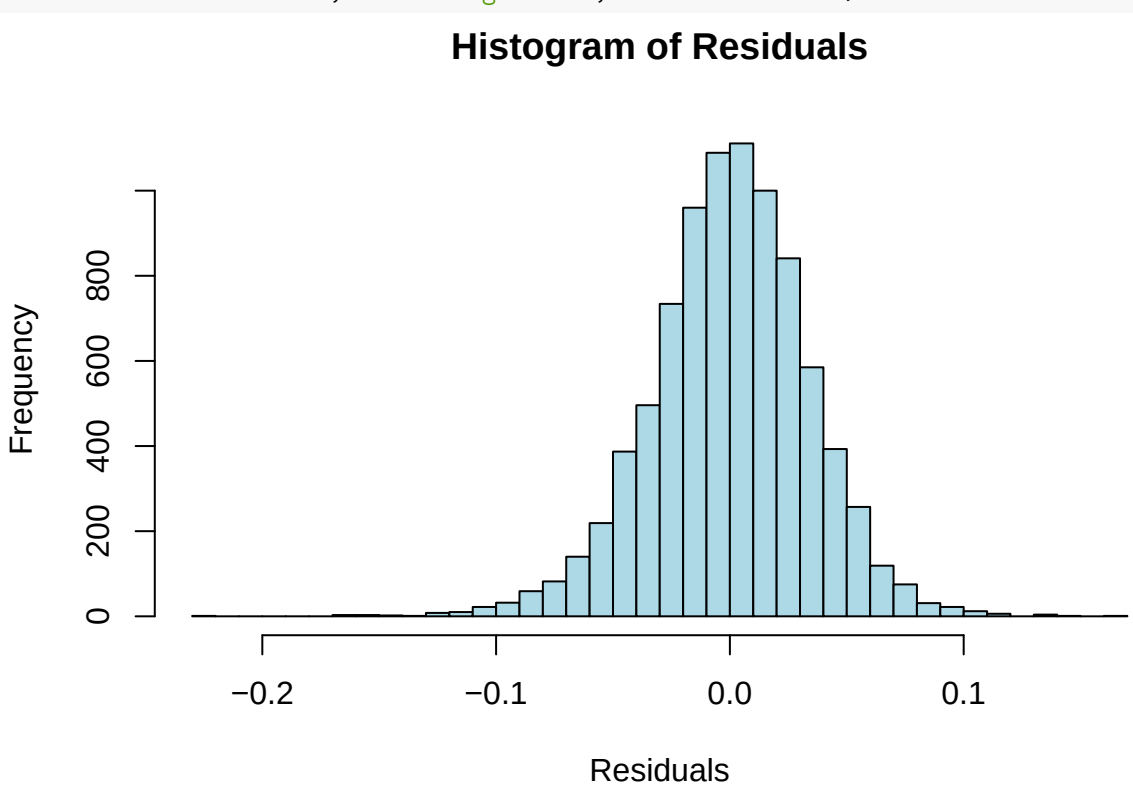
```
     ylab = "Residuals")
abline(h = 0, col = "red", lwd = 2)
```




```
# Check normality of residuals
qqnorm(residuals, main = "Q-Q Plot of Residuals")
qqline(residuals, col = "red", lwd = 2)
```

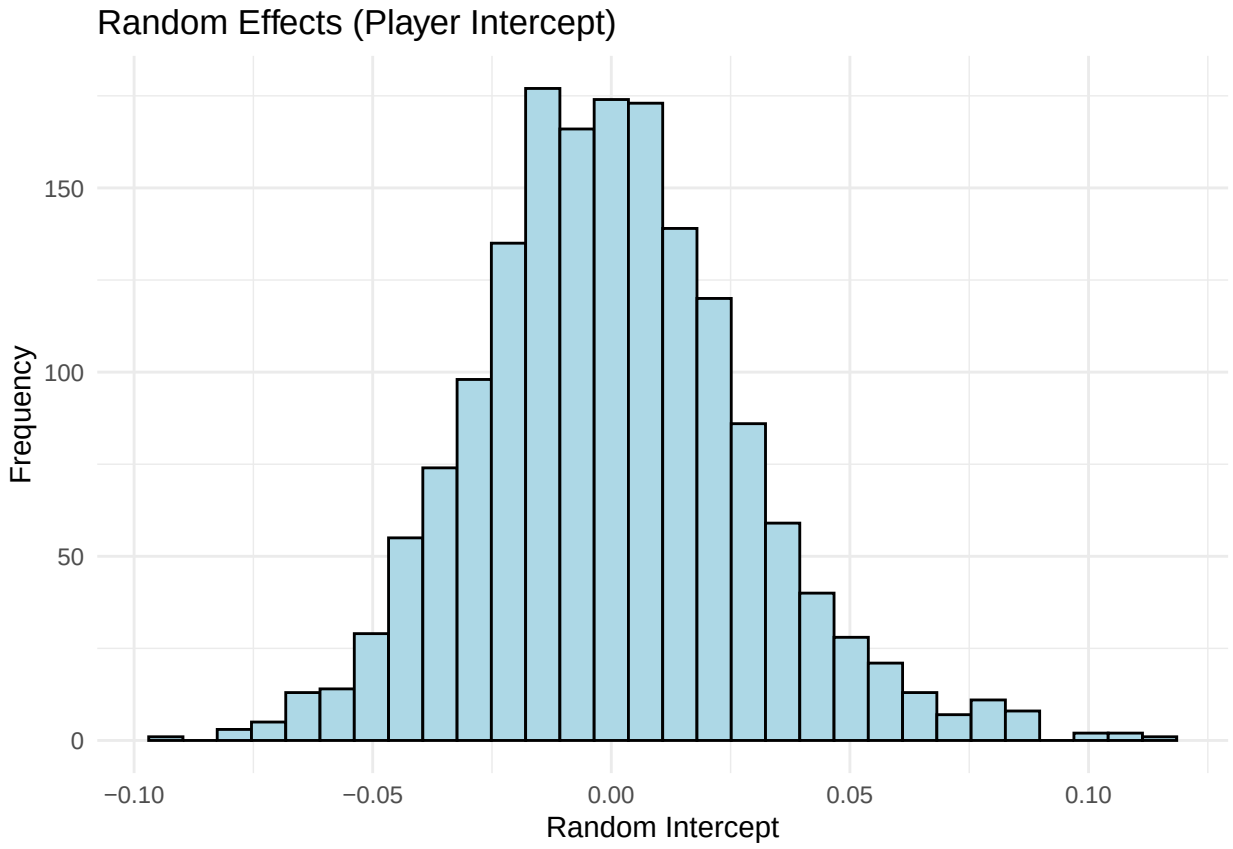



```r
# Histogram of residuals
hist(residuals, breaks = 30, main = "Histogram of Residuals",
     xlab = "Residuals", col = "lightblue", border = "black")
```



Random Effects Diagnostics

```r
# Extract random effects
ranef_data <- as.data.frame(ranef(model)$player_name)

# Plot random effects
ggplot(ranef_data, aes(x = `(Intercept)`)) +
  geom_histogram(bins = 30, fill = "lightblue", color = "black") +
  labs(title = "Random Effects (Player Intercept)", x = "Random Intercept", y = "Frequency") +
  theme_minimal()
```



Multicollinearity Check

```r
# Using performance package
# Check collinearity for your mixed-effects model
check_collinearity(model)
```

```
## Model has interaction terms. VIFs might be inflated.
##   You may check multicollinearity among predictors of a model without
##   interaction terms.
```

```
## # Check for Multicollinearity
##
## Low Correlation
##
##               Term  VIF          VIF 95% CI Increased SE Tolerance
##  season_continuous 1.73 [    1.68,     1.78]         1.31      0.58
##      player_height 2.39 [    2.31,     2.47]         1.55      0.42
##      player_weight 2.41 [    2.33,     2.49]         1.55      0.41
##       career_stage 1.23 [    1.20,     1.26]         1.11      0.81
```

```
##  Tolerance 95% CI
##      [0.56, 0.60]
##      [0.40, 0.43]
##      [0.40, 0.43]
##      [0.79, 0.83]
##
## High Correlation
##
##                                    Term      VIF           VIF 95% CI
##                     draft_round_combined 1.70e+09 [1.63e+09, 1.77e+09]
##  season_continuous:draft_round_combined 1.70e+09 [1.63e+09, 1.77e+09]
##  Increased SE Tolerance Tolerance 95% CI
##      41195.20  5.89e-10     [0.00, 0.00]
##      41198.60  5.89e-10     [0.00, 0.00]
```

Potential Logit Transform

```
filtered_data$ts_pct_logit <- log(filtered_data$ts_pct / (1 - filtered_data$ts_pct))
model_logit <- lmer(ts_pct_logit ~ season_continuous * draft_round_combined +
                      player_height + player_weight + career_stage + (1 | player_name), data = filtered

summary(model_logit)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## ts_pct_logit ~ season_continuous * draft_round_combined + player_height +
##     player_weight + career_stage + (1 | player_name)
##    Data: filtered_data
##
## REML criterion at convergence: -5921.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.6834 -0.5530  0.0265  0.5926  4.6679
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  player_name (Intercept) 0.01871  0.1368
##  Residual                0.02203  0.1484
## Number of obs: 8706, groups:  player_name, 1654
##
## Fixed effects:
##                                                Estimate Std. Error
## (Intercept)                                  -1.690e+01  9.716e-01
## season_continuous                             8.147e-03  4.794e-04
## draft_round_combined2                        -1.498e+00  1.838e+00
## draft_round_combinedUndrafted                -1.112e+00  2.279e+00
## player_height                                 2.670e-03  6.662e-04
## player_weight                                 1.190e-03  4.641e-04
## career_stageMid-Career                        3.140e-02  4.312e-03
## career_stageVeteran                          -1.776e-02  5.815e-03
## season_continuous:draft_round_combined2       7.404e-04  9.147e-04
## season_continuous:draft_round_combinedUndrafted  5.430e-04  1.133e-03
##                                                      df t value Pr(>|t|)
```

```
## (Intercept)                                        2.805e+03 -17.397  < 2e-16 ***
## season_continuous                                  2.855e+03  16.997  < 2e-16 ***
## draft_round_combined2                              3.899e+03  -0.815  0.41522
## draft_round_combinedUndrafted                      3.250e+03  -0.488  0.62568
## player_height                                      2.481e+03   4.008 6.29e-05 ***
## player_weight                                      3.332e+03   2.564  0.01039 *
## career_stageMid-Career                             8.665e+03   7.283 3.56e-13 ***
## career_stageVeteran                                6.403e+03  -3.054  0.00227 **
## season_continuous:draft_round_combined2            3.900e+03   0.809  0.41832
## season_continuous:draft_round_combinedUndrafted    3.251e+03   0.479  0.63191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) ssn_cn drf__2 drf__U plyr_h plyr_w cr_M-C crr_sV s_:__2
## seasn_cntns -0.994
## drft_rnd_c2 -0.477  0.479
## drft_rnd_cU -0.392  0.396  0.198
## player_hght -0.125  0.024  0.021 -0.008
## player_wght  0.047  0.010 -0.015  0.018 -0.756
## crr_stgMd-C  0.209 -0.221 -0.048 -0.073  0.115 -0.161
## crr_stgVtrn  0.341 -0.355 -0.061 -0.077  0.126 -0.180  0.541
## ssn_cnt:__2  0.477 -0.479 -1.000 -0.198 -0.021  0.015  0.048  0.061
## ssn_cnt:__U  0.392 -0.396 -0.198 -1.000  0.009 -0.018  0.073  0.077  0.198
```

Compare models

```
AIC(model, model_logit)
```

```
##             df        AIC
## model       12 -30340.683
## model_logit 12  -5897.477
```

Non linear effects of height and weight?

```
# Quadtratic terms
model_poly <- lmer(ts_pct ~ season_continuous * draft_round_combined +
                     poly(player_height, 2) + poly(player_weight, 2) +
                     career_stage + (1 | player_name), data = filtered_data)
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```
summary(model_poly)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## ts_pct ~ season_continuous * draft_round_combined + poly(player_height,
##     2) + poly(player_weight, 2) + career_stage + (1 | player_name)
##    Data: filtered_data
##
## REML criterion at convergence: -30393.2
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.1444 -0.5572  0.0289  0.6011  4.4884
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  player_name (Intercept) 0.001108 0.03328
##  Residual                0.001328 0.03644
## Number of obs: 8706, groups:  player_name, 1654
##
## Fixed effects:
##                                                Estimate Std. Error
## (Intercept)                                   -3.500e+00  2.360e-01
## season_continuous                              2.005e-03  1.176e-04
## draft_round_combined2                         -3.355e-01  4.495e-01
## draft_round_combinedUndrafted                 -2.921e-01  5.570e-01
## poly(player_height, 2)1                        4.819e-01  1.440e-01
## poly(player_height, 2)2                        2.555e-01  9.822e-02
## poly(player_weight, 2)1                        3.786e-01  1.343e-01
## poly(player_weight, 2)2                       -1.442e-01  9.197e-02
## career_stageMid-Career                         7.650e-03  1.058e-03
## career_stageVeteran                           -4.573e-03  1.426e-03
## season_continuous:draft_round_combined2        1.657e-04  2.237e-04
## season_continuous:draft_round_combinedUndrafted  1.427e-04  2.770e-04
##                                                       df t value Pr(>|t|)
## (Intercept)                                   2.852e+03 -14.826  < 2e-16 ***
## season_continuous                             2.847e+03  17.060  < 2e-16 ***
## draft_round_combined2                         3.871e+03  -0.746 0.455509
## draft_round_combinedUndrafted                 3.237e+03  -0.524 0.600054
## poly(player_height, 2)1                       2.538e+03   3.345 0.000833 ***
## poly(player_height, 2)2                       1.920e+03   2.602 0.009345 **
## poly(player_weight, 2)1                       3.274e+03   2.820 0.004827 **
## poly(player_weight, 2)2                       3.609e+03  -1.568 0.116977
## career_stageMid-Career                        8.660e+03   7.230 5.25e-13 ***
## career_stageVeteran                           6.391e+03  -3.206 0.001351 **
## season_continuous:draft_round_combined2       3.871e+03   0.741 0.458838
## season_continuous:draft_round_combinedUndrafted  3.237e+03   0.515 0.606512
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##               (Intr) ssn_cn drf__2 drf__U ply(plyr_h,2)1 ply(plyr_h,2)2
## seasn_cntns    -1.000
## drft_rnd_c2    -0.477  0.477
## drft_rnd_cU    -0.395  0.395  0.198
## ply(plyr_h,2)1 -0.032  0.032  0.015 -0.008
## ply(plyr_h,2)2 -0.054  0.054  0.005 -0.003 -0.140
## ply(plyr_w,2)1  0.000  0.000 -0.010  0.018 -0.774          0.103
## ply(plyr_w,2)2 -0.033  0.032 -0.016  0.001  0.296         -0.472
## crr_stgMd-C     0.219 -0.221 -0.048 -0.072  0.112         -0.045
## crr_stgVtrn     0.353 -0.355 -0.060 -0.077  0.120         -0.050
## ssn_cnt:__2     0.477 -0.477 -1.000 -0.198 -0.015         -0.005
## ssn_cnt:__U     0.395 -0.395 -0.198 -1.000  0.008          0.003
##               ply(plyr_w,2)1 ply(plyr_w,2)2 cr_M-C crr_sV s_:__2
```

```
## seasn_cntns
## drft_rnd_c2
## drft_rnd_cU
## ply(plyr_h,2)1
## ply(plyr_h,2)2
## ply(plyr_w,2)1
## ply(plyr_w,2)2 -0.248
## crr_stgMd-C    -0.156          0.008
## crr_stgVtrn    -0.172         -0.004          0.541
## ssn_cnt:__2     0.010          0.015          0.048  0.060
## ssn_cnt:__U    -0.018         -0.001          0.072  0.077  0.198
## fit warnings:
## Some predictor variables are on very different scales: consider rescaling
```

```
# Spline model
library(splines)
model_spline <- lmer(ts_pct ~ season_continuous * draft_round_combined +
                       ns(player_height, df = 3) + ns(player_weight, df = 3) +
                       career_stage + (1 | player_name), data = filtered_data)
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling

## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

```
summary(model_spline)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: ts_pct ~ season_continuous * draft_round_combined + ns(player_height,
##     df = 3) + ns(player_weight, df = 3) + career_stage + (1 |      player_name)
##    Data: filtered_data
##
## REML criterion at convergence: -30362.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.1449 -0.5561  0.0285  0.5993  4.4861
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  player_name (Intercept) 0.001110 0.03331
##  Residual                0.001328 0.03643
## Number of obs: 8706, groups:  player_name, 1654
##
## Fixed effects:
##                                              Estimate Std. Error
## (Intercept)                                -3.513e+00  2.358e-01
## season_continuous                           2.006e-03  1.177e-04
## draft_round_combined2                      -3.395e-01  4.499e-01
## draft_round_combinedUndrafted              -3.033e-01  5.574e-01
## ns(player_height, df = 3)1                  7.140e-03  1.046e-02
## ns(player_height, df = 3)2                  3.485e-04  4.036e-02
## ns(player_height, df = 3)3                  3.746e-02  1.727e-02
```

```
## ns(player_weight, df = 3)1                         2.307e-02  7.933e-03
## ns(player_weight, df = 3)2                         3.207e-02  2.700e-02
## ns(player_weight, df = 3)3                         6.353e-03  1.639e-02
## career_stageMid-Career                             7.638e-03  1.059e-03
## career_stageVeteran                               -4.609e-03  1.429e-03
## season_continuous:draft_round_combined2            1.677e-04  2.239e-04
## season_continuous:draft_round_combinedUndrafted    1.483e-04  2.773e-04
##                                                           df t value Pr(>|t|)
## (Intercept)                                        2.852e+03 -14.897  < 2e-16 ***
## season_continuous                                  2.845e+03  17.041  < 2e-16 ***
## draft_round_combined2                              3.872e+03  -0.755  0.45051
## draft_round_combinedUndrafted                      3.234e+03  -0.544  0.58636
## ns(player_height, df = 3)1                         2.340e+03   0.683  0.49486
## ns(player_height, df = 3)2                         2.084e+03   0.009  0.99311
## ns(player_height, df = 3)3                         1.892e+03   2.170  0.03016 *
## ns(player_weight, df = 3)1                         3.381e+03   2.908  0.00366 **
## ns(player_weight, df = 3)2                         3.851e+03   1.188  0.23498
## ns(player_weight, df = 3)3                         3.862e+03   0.388  0.69835
## career_stageMid-Career                             8.658e+03   7.214 5.89e-13 ***
## career_stageVeteran                                6.397e+03  -3.227  0.00126 **
## season_continuous:draft_round_combined2            3.872e+03   0.749  0.45387
## season_continuous:draft_round_combinedUndrafted    3.235e+03   0.535  0.59272
## ---
## 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
```

```
## fit warnings:
## Some predictor variables are on very different scales: consider rescaling
```

Compare models

```
anova(model, model_poly, model_spline)
```

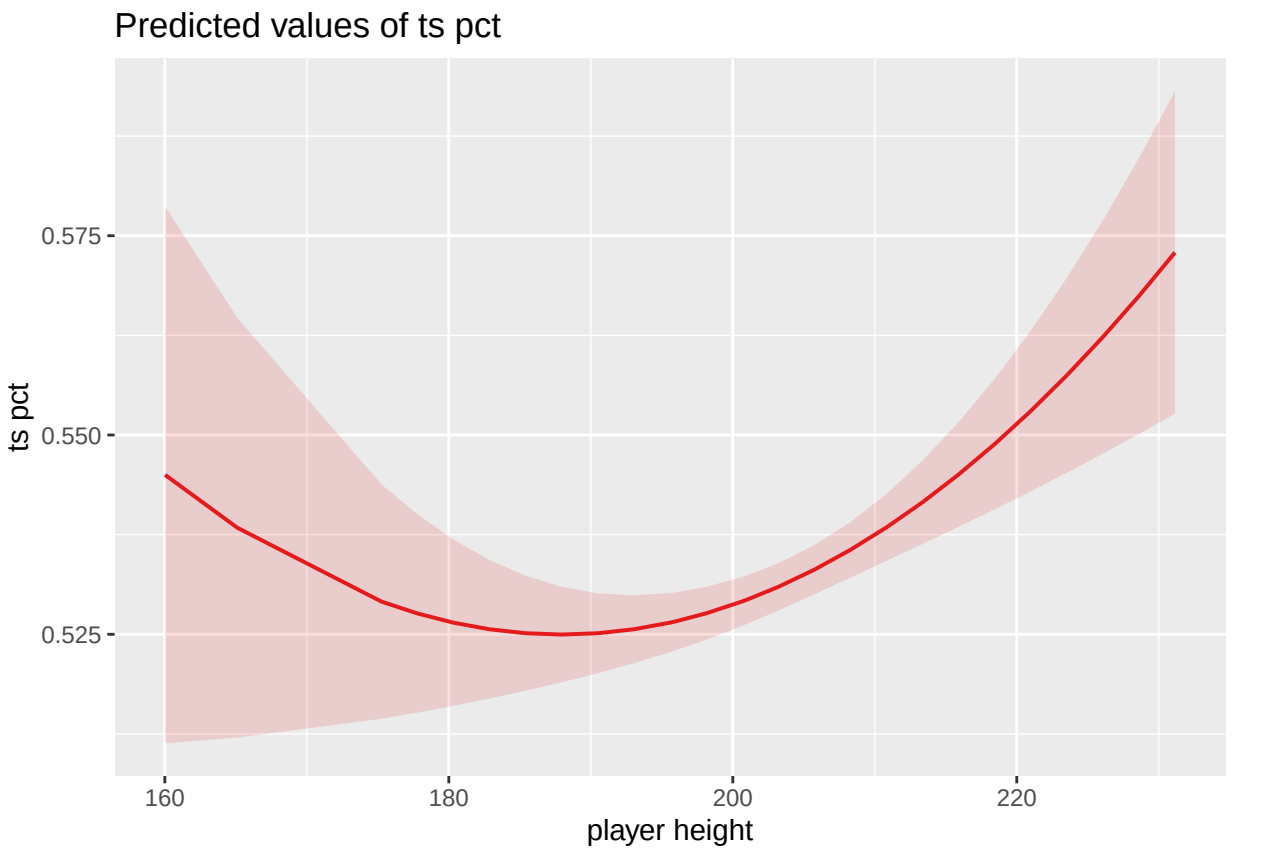
```
## refitting model(s) with ML (instead of REML)
```

```
## Data: filtered_data
## Models:
## model: ts_pct ~ season_continuous * draft_round_combined + player_height + player_weight + career_sta
## model_poly: ts_pct ~ season_continuous * draft_round_combined + poly(player_height, 2) + poly(player_
## model_spline: ts_pct ~ season_continuous * draft_round_combined + ns(player_height, df = 3) + ns(play
##              npar    AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
## model          12 -30476 -30391  15250   -30500
## model_poly     14 -30479 -30380  15253   -30507 6.9417  2    0.03109 *
## model_spline   16 -30475 -30362  15254   -30507 0.2338  2    0.88967
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

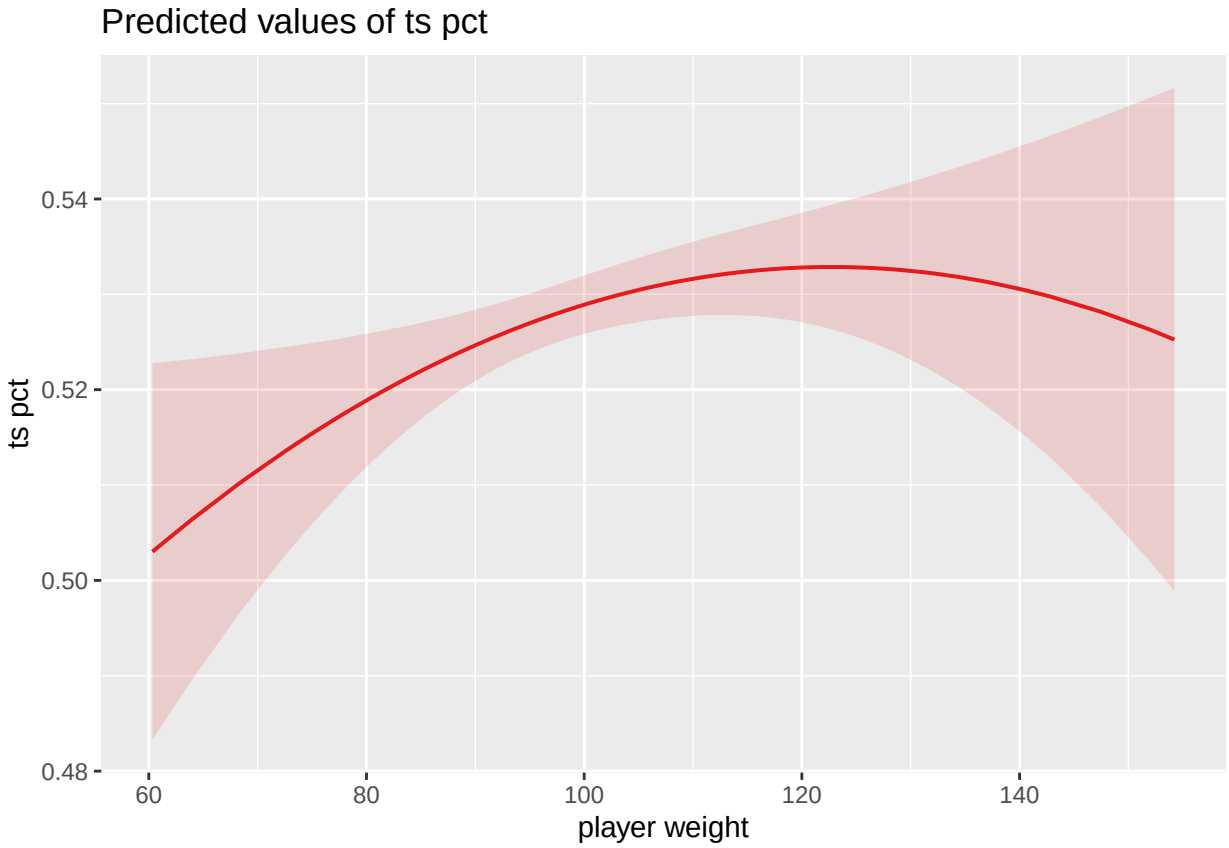
```
AIC(model, model_poly, model_spline)
```

```
##              df       AIC
## model        12 -30340.68
## model_poly   14 -30365.17
## model_spline 16 -30330.12
```

```
library(sjPlot)
plot_model(model_poly, type = "pred", terms = "player_height [all]")
```



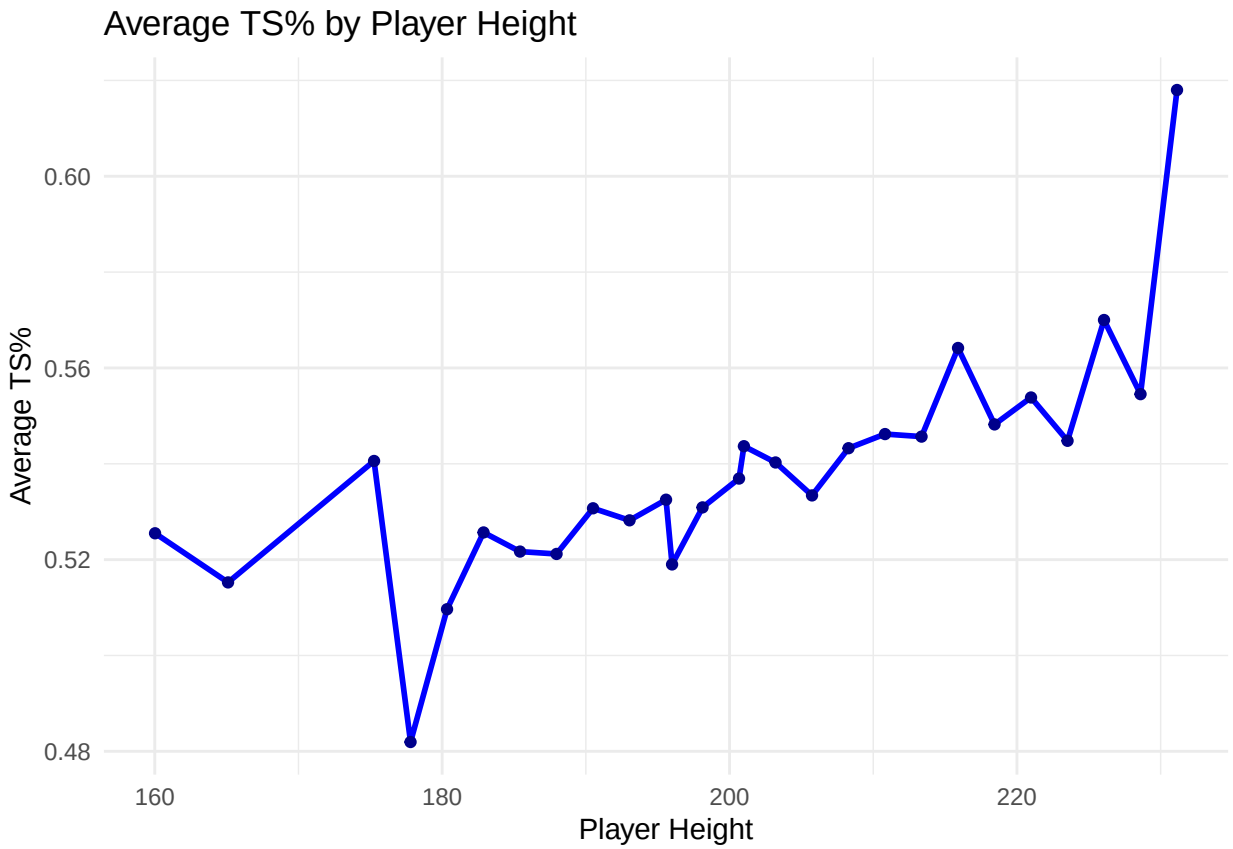

```
plot_model(model_poly, type = "pred", terms = "player_weight [all]")
```




```
# Group by player_height and calculate the mean ts_pct
avg_ts_by_height <- filtered_data %>%
  group_by(player_height) %>%
  summarize(avg_ts_pct = mean(ts_pct, na.rm = TRUE))

# View the summarized data
head(avg_ts_by_height)
```

```
## # A tibble: 6 x 2
##   player_height avg_ts_pct
##           <dbl>      <dbl>
## 1          160.      0.526
## 2          165.      0.515
## 3          175.      0.541
## 4          178.      0.482
## 5          180.      0.510
## 6          183.      0.526
```

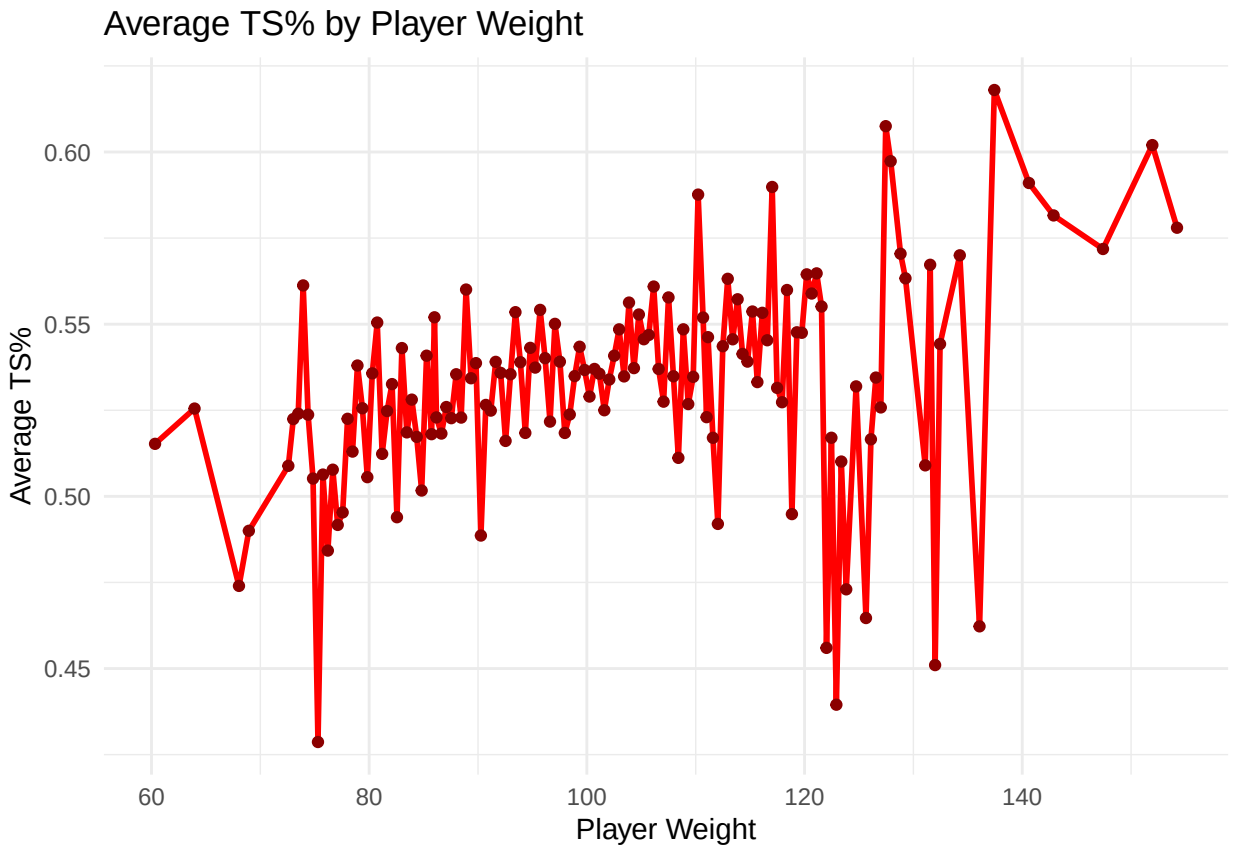
```
ggplot(avg_ts_by_height, aes(x = player_height, y = avg_ts_pct)) +
  geom_line(color = "blue", size = 1) +
  geom_point(color = "darkblue") +
  labs(
    title = "Average TS% by Player Height",
    x = "Player Height",
    y = "Average TS%"
  ) +
  theme_minimal()
```




```
# Group by player_weight and calculate the mean ts_pct
avg_ts_by_weight <- filtered_data %>%
  group_by(player_weight) %>%
  summarize(avg_ts_pct = mean(ts_pct, na.rm = TRUE))

# View the summarized data
head(avg_ts_by_weight)
```

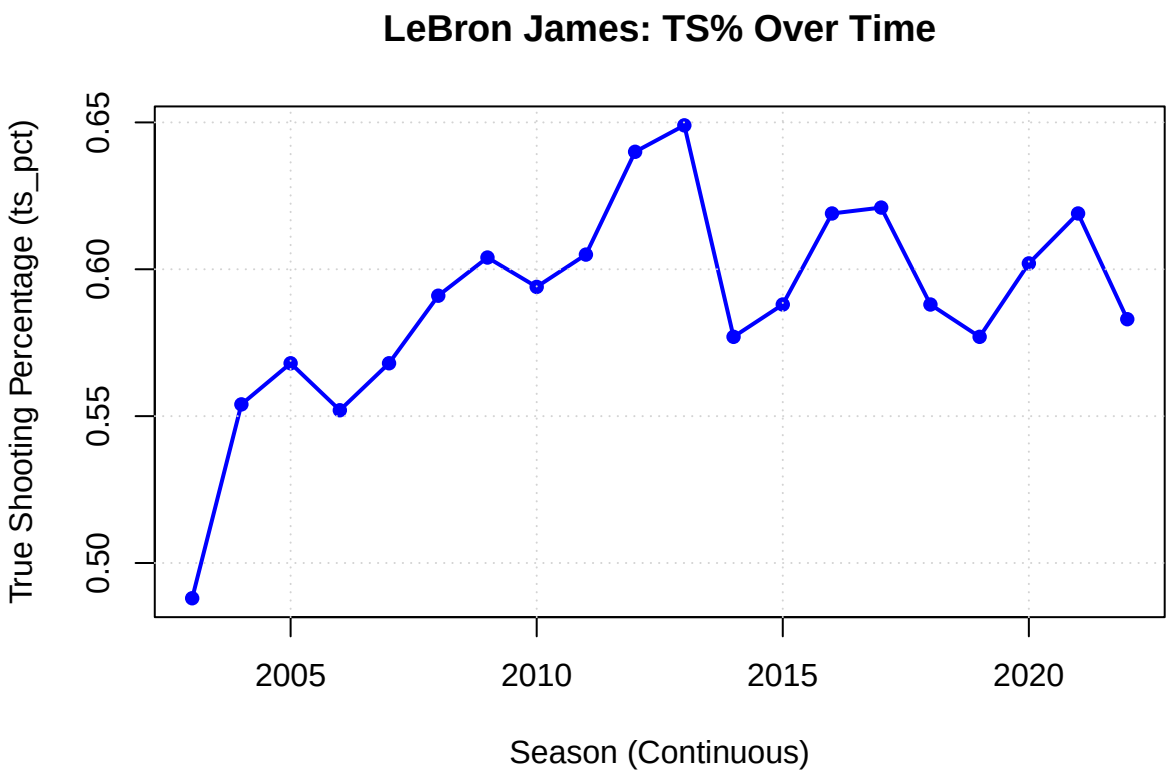
```
## # A tibble: 6 x 2
##   player_weight avg_ts_pct
##           <dbl>      <dbl>
## 1          60.3      0.515
## 2          64.0      0.526
## 3          68.0      0.474
## 4          68.9      0.49
## 5          72.6      0.509
## 6          73.0      0.522
```

```
ggplot(avg_ts_by_weight, aes(x = player_weight, y = avg_ts_pct)) +
  geom_line(color = "red", size = 1) +
  geom_point(color = "darkred") +
  labs(
    title = "Average TS% by Player Weight",
    x = "Player Weight",
    y = "Average TS%"
  ) +
  theme_minimal()
```

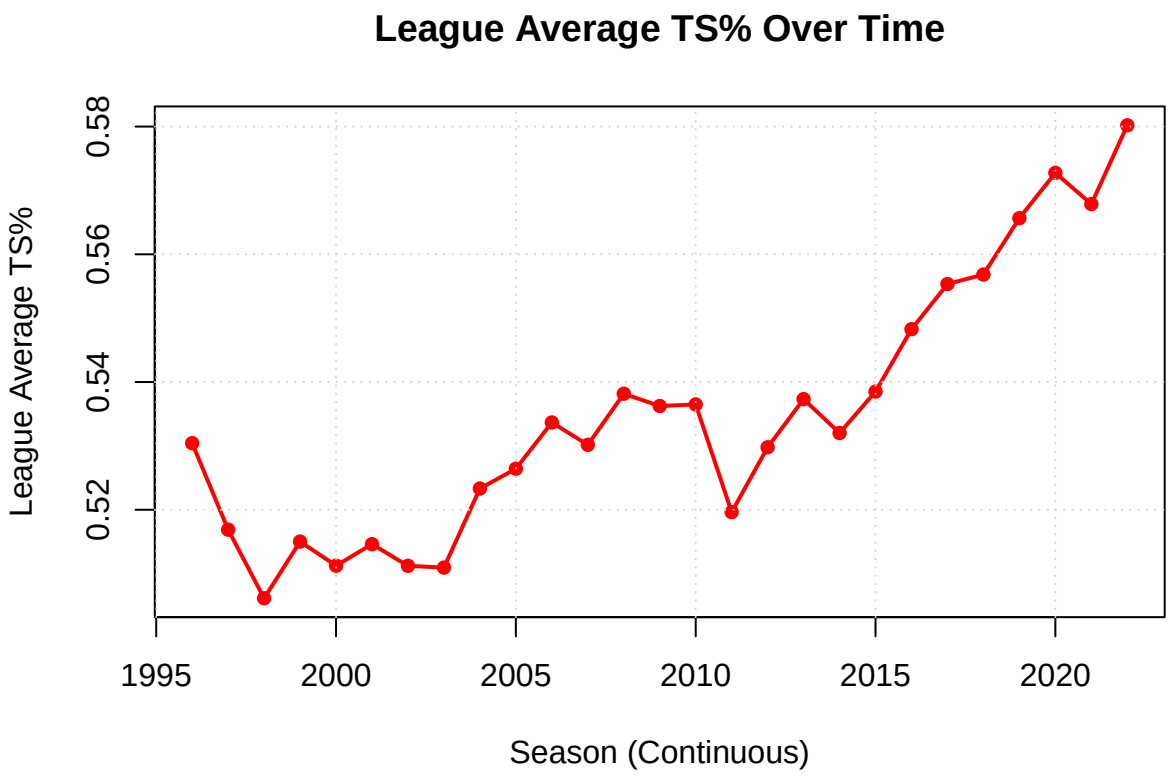


LeBron James sub-analysis

```r
# Filter the data for LeBron James
leData <- subset(filtered_data, player_name == "LeBron James")

# Plot ts_pct over season_continuous for LeBron James
plot(leData$season_continuous, leData$ts_pct,
     type = "o", # Line plot with points
     col = "blue", # Line color
     pch = 16, # Point style
     lwd = 2, # Line width
     xlab = "Season (Continuous)", # X-axis label
     ylab = "True Shooting Percentage (ts_pct)", # Y-axis label
     main = "LeBron James: TS% Over Time") # Title

# Add grid lines for better readability (optional)
grid()
```




```r
# Calculate the league average ts_pct for each season
league_avg <- aggregate(ts_pct ~ season_continuous, data = filtered_data, mean)

# Plot league average ts_pct over season_continuous
plot(league_avg$season_continuous, league_avg$ts_pct,
     type = "o", # Line plot with points
     col = "red", # Line color
     pch = 16, # Point style
     lwd = 2, # Line width
     xlab = "Season (Continuous)", # X-axis label
     ylab = "League Average TS%", # Y-axis label
     main = "League Average TS% Over Time") # Title

# Add grid lines for better readability (optional)
grid()
```




```
leModel <- lm(ts_pct ~ career_stage, data = leData)

# View the model summary
summary(leModel)
```

```
##
## Call:
## lm(formula = ts_pct ~ career_stage, data = leData)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.07271 -0.01287 -0.00217  0.01988  0.04329
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              0.56071    0.01097  51.109  < 2e-16 ***
## career_stageMid-Career   0.05229    0.01700   3.076  0.00684 **
## career_stageVeteran      0.03891    0.01502   2.590  0.01907 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02903 on 17 degrees of freedom
## Multiple R-squared:  0.3958, Adjusted R-squared:  0.3247
## F-statistic: 5.567 on 2 and 17 DF,  p-value: 0.01381
```