Harry Finch
15916 Las Planideras, Rancho Santa Fe, CA 92067 | (619) 729-3452 | haf58@cornell.edu
Cornell University – College of Agriculture and Life Sciences
Bachelor of Science in Biometry and Statistics | Minor: Mathematics
Ithaca, NY | Expected May 2027

Project Title

NBA Draft Analytics Project – Independent Research | Summer 2025

Introduction

- **Background**: Drafting an NBA superstar or star is a very difficult task. Not all drafts have superstar or star level talent. There is a common thought that a better draft pick is associated with a higher probability to draft star or superstar level talent. This project aims to test if that claim is correct as well as any other patterns that may impact the probability of drafting a superstar level player.
- **Problem Statement**: The goal is to develop a robust analytical framework that assesses player value and optimizes drafting decisions for team managers. The key question is how to use historical data to identify potential patterns or correlations for players that explain why certain players pan out and others don't as well as tackling the question of whether it is possible to predict superstar level talent.
- **Goals and Objectives**:
    1. Build predictive models to analyze NBA drafts from 2014 to 2023.
    2. Assess player value based on a combination of performance, physical traits, and advanced metrics.
    3. Apply statistical techniques like regression and clustering to uncover draft trends.
    4. Develop a risk-adjusted model to provide guidance for general managers on maximizing value at each draft position.

Data

- **Data Source**: Data was compiled from various public sources, including basketball reference, statmuse, and NBA.com, covering the 2014–2023 NBA draft classes.
- **Data Description**: The dataset includes:
    - **Performance Metrics**: Advanced statistics such as Player Efficiency Rating (PER) and True Shooting Percentage (TS%), Assist Percentage (AST%), Turnover Percentage (TOV%), Box Plus Minus (BPM), Value Over Replacement (VORP), and Win Shares (WS).
    - **Accolades**: Allstar teams, All NBA teams, All Defensive Teams, MVPS, DPOYS, and FMVPS.
    - **Draft Information**: Draft position, college/international team, and other relevant biographical data.
- **Data Cleaning and Preparation**: Data from various sources was merged and cleaned. Missing values were handled, and features were engineered to create new variables relevant to player value assessment.

Methodology

- **Exploratory Data Analysis (EDA)**: The initial phase involved visualizing relationships between draft position and various metrics. For example, a box plot was used to examine the distribution of key performance indicators for players drafted in different rounds, highlighting potential differences in talent.
- **Statistical Methods**:
  - **Regression Analysis**: Multiple linear and logistic regression models were employed to display which NBA players have statistically achieved stardom, and explained potential patterns and correlations that could explain why.
  - **CHI Squared Testing**: Chi squared testing was used to see if NBA players who panned out to become stars could've been expected based on the data or if certain players reaching stardom was not statistically explainable.
- **Tools Used**: The entire project was developed using the **R** programming language for data manipulation, statistical modeling, and visualization.

Results

- **Key Findings**:
  - **Predictive Model Insights**: This model actually showed us that there is a significant gap between the best 5 players in the last 10 years and everyone else. This is a gap from players in tier 1 to players in tier 2 and so on. The gap is actually so considerably large that without the players from tier 1 the model is far more statistically significant. It displays that stardom can be oftentimes predicted and explained yet superstardom is far more difficult to predict.
  - **Chi Squared Results**: The clustering analysis revealed distinct archetypes of players that tend to be drafted in specific ranges. For example, highly athletic, low-skill players were often drafted in the late first round but had a lower success rate compared to more skilled players.
- **Visualizations**: Visualizations included scatter plots of performance metrics versus draft position, correlations displayed for predictors of stardom, and a series of interactive charts to illustrate the risk-adjusted value of each draft pick

Discussion and Conclusion

- **Summary of Findings**: This research successfully built a data-driven framework for NBA draft analytics. The models revealed that a combination of performance metrics, award catalog, and a risk-adjusted approach can significantly improve the accuracy of player value assessment.
- **Implications**: The findings provide actionable insights for team management, allowing them to move beyond traditional scouting biases and make more strategic, value-based decisions during the draft.

- **Limitations**: The model is limited by the availability of pre-draft data. It does not account for a player's psychological profile, work ethic, or other unquantifiable factors that can impact their NBA career.
- **Future Work**: Future research could expand the dataset to include more international players, incorporate sentiment analysis of scouting reports, or use machine learning techniques like neural networks to improve predictive accuracy.

References

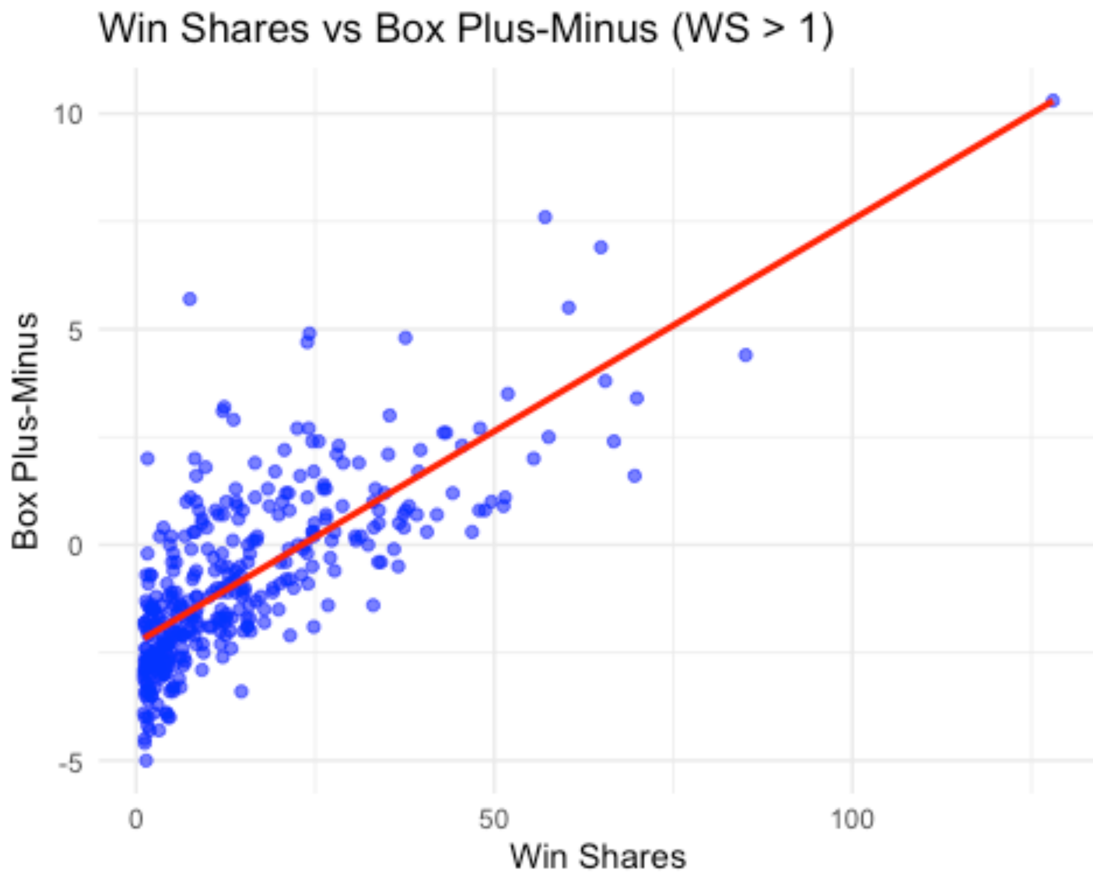
- **Data Sources**: NBA.com, Basketball-Reference.com, and StatMuse.
- **Analysis**: Harry Finch. "NBA Draft Analytics Project." **R Markdown File**. August 9, 2025.

NBA Analysis.RMD

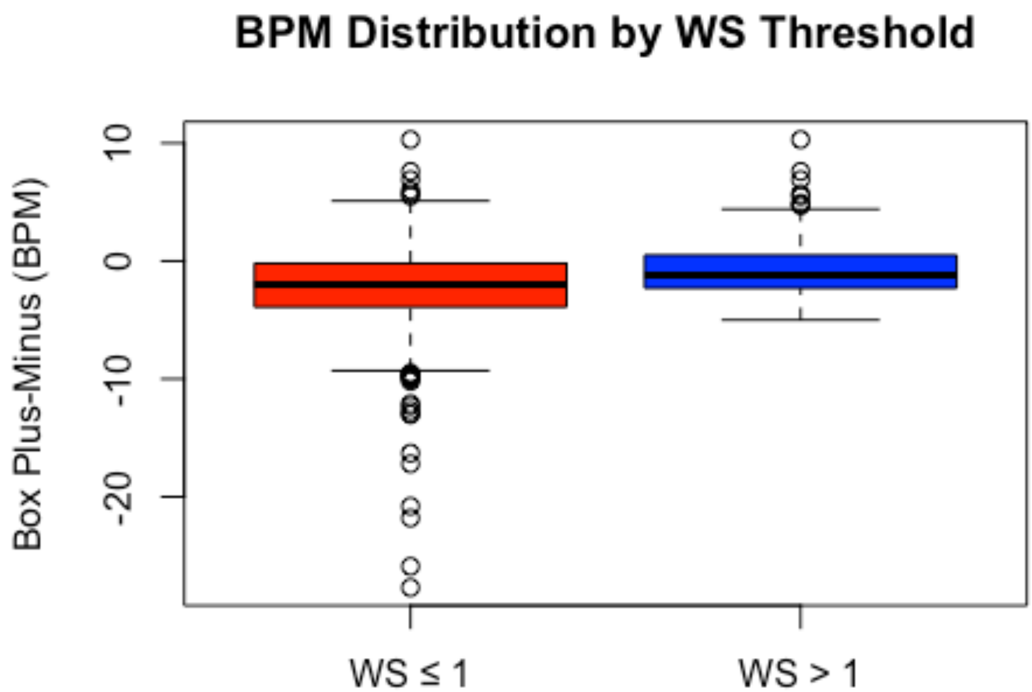
Harry Finch

2025-08-09

```
## Rows: 596 Columns: 37
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr (16): Team, Player, Height, Pre NBA, Conference, Nationality,
Position, ...
## dbl (21): Year, Pick, Weight, HS Final Rank, Years, GP, MP, PTS, TRB, AST,
F...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
## Rows: 55 Columns: 14
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (1): Player
## dbl (13): Allstar Selections, All NBA Selections, All Defensive
Selections, ...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.

## `geom_smooth()` using formula = 'y ~ x'
```




```
## Correlation (WS > 1): 0.745
```




```
## # A tibble: 10 × 5
##    Player                    Pick    WS   BPM NBA_Star_Score
##    <chr>                    <dbl> <dbl> <dbl>          <dbl>
##  1 Nikola Jokić                41 128    10.3          1318.
##  2 Joel Embiid                  3  64.9   6.9           448.
##  3 Luka Dončić                  3  57.1   7.6           434.
##  4 Karl-Anthony Towns           1  85.1   4.4           374.
##  5 Shai Gilgeous-Alexander     11  60.4   5.5           332.
##  6 Jayson Tatum                 3  65.5   3.8           249.
##  7 Domantas Sabonis            11  69.9   3.4           238.
##  8 Donovan Mitchell            13  51.9   3.5           182.
##  9 Tyrese Haliburton           12  37.6   4.8           180.
## 10 Jarrett Allen               22  66.7   2.4           160.
```

To kick off the project, we needed a starting metric to define what makes an NBA "star." We began with a very straightforward formula — multiplying Win Shares (WS) by Box Plus-Minus (BPM). These two advanced stats are among the most widely recognized: WS estimates the number of wins a player contributes to their team, while BPM measures a player's per-100 possessions impact relative to league average.

We chose to filter for players with WS > 1, excluding those with minimal NBA playing time who could distort the analysis. Players under that threshold often barely stepped on the floor, and their BPM values tend to be unstable.

With this filtered group, we found a positive correlation of 0.75 between WS and BPM — a relationship strong enough to suggest that players who accumulate more wins also tend to post better on-court impact metrics. Finally, we listed the top 10 players by this "baseline Star Score" to give an initial sense of which players rank highest under this simple definition of stardom.

```
##                       Player Pick NBA_Star_Score
## 441             Nikola Jokić   41        1493.40
## 289              Joel Embiid    3         616.81
## 379              Luka Dončić    3         558.96
## 336       Karl-Anthony Towns    1         469.44
## 511 Shai Gilgeous-Alexander   11         407.20
## 275             Jayson Tatum    3         383.90
## 150         Domantas Sabonis   11         297.66
## 151         Donovan Mitchell   13         271.65
## 562        Tyrese Haliburton   12         230.48
## 36               Bam Adebayo   14         214.00
```

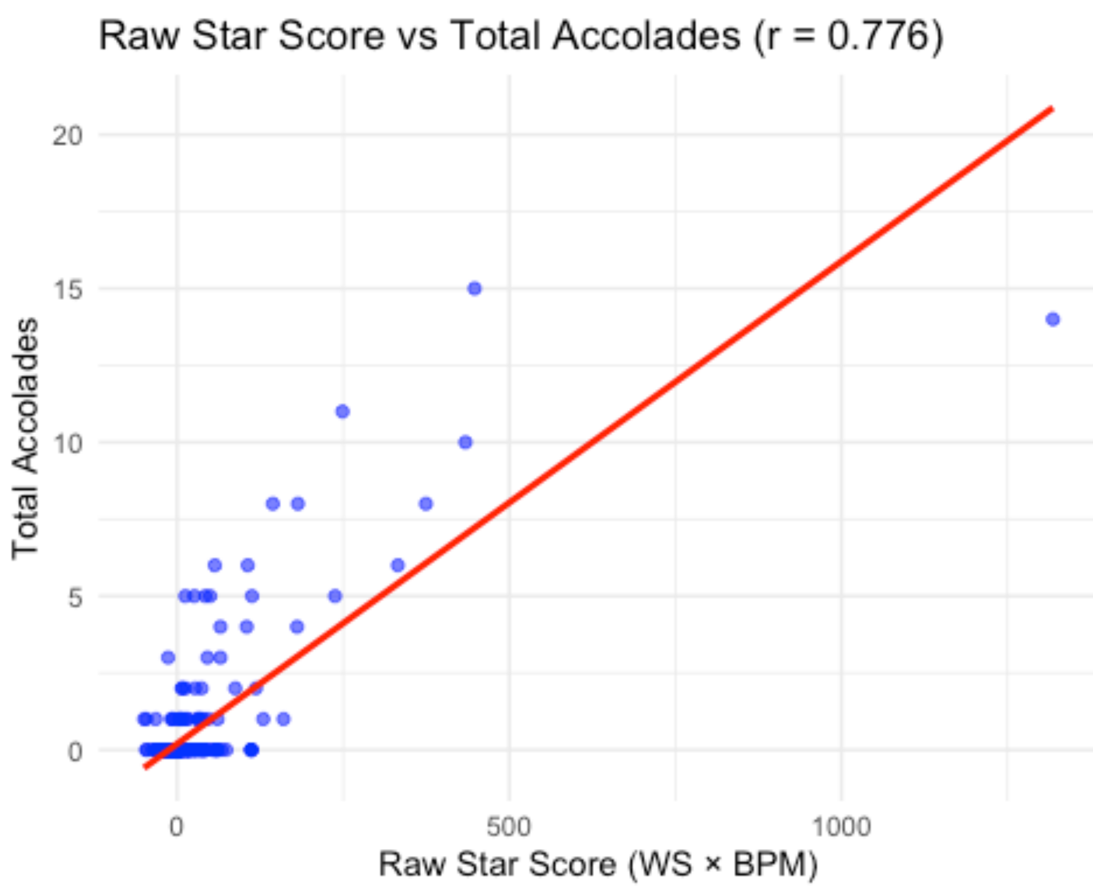
Here we built on the original Star Score, because WS × BPM by itself was good, but not really enough. This time we added in a player's accolades things like All-NBA, All-Star, and All-Defensive selections since those usually point to guys who have been great for a while and have gotten league recognition for it.

We gave each award a weight based on how important we thought it was to a player's overall impact. All-NBA got the most weight, then All-Star, then All-Defensive. That way, it values both production and recognition. It also naturally rewards guys who've done it for more than just a season or two.

When we ran it, the top 10 list shifted compared to the first version, which shows how adding accolades really changes who shows up as a "star" in this model.

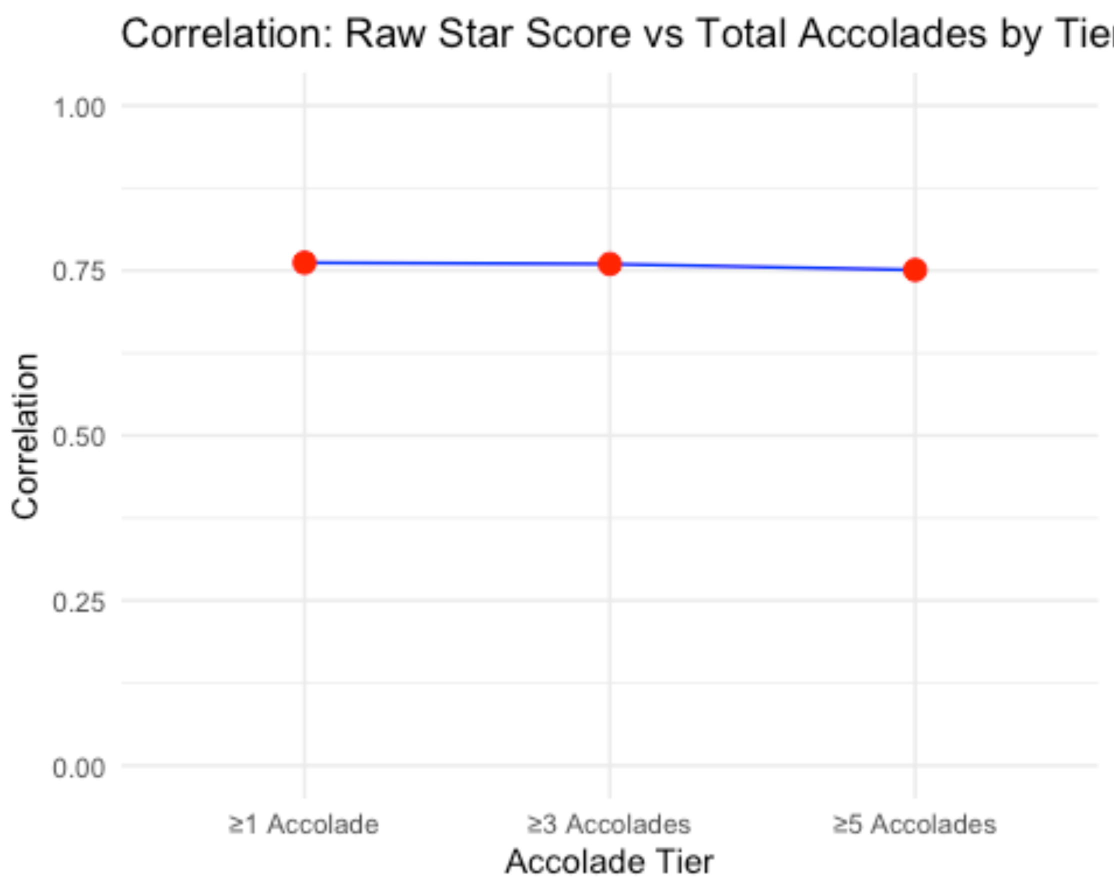
```r
# --------------------------
# 4) Raw Star Score vs Total Accolades (streamlined)
# --------------------------
if (!"Raw_Star_Score" %in% names(NBA_Drafts)) {
  NBA_Drafts$Raw_Star_Score <- NBA_Drafts$WS * NBA_Drafts$BPM
}
if (!"Total_Accolades" %in% names(NBA_Drafts)) {
  need_cols <- c("All NBA Selections","Allstar Selections","All Defensive
Selections")
  for (cn in need_cols) if (!cn %in% names(NBA_Drafts)) NBA_Drafts[[cn]] <- 0
  NBA_Drafts <- NBA_Drafts %>%
    mutate(across(all_of(need_cols), ~tidyr::replace_na(., 0))) %>%
    mutate(Total_Accolades = `All NBA Selections` + `Allstar Selections` +
`All Defensive Selections`)
}
r_all <- safe_cor(NBA_Drafts$Raw_Star_Score, NBA_Drafts$Total_Accolades)
cat("Correlation (All Players):", round(r_all, 3), "\n")

## Correlation (All Players): 0.776
```

```
p_all <- ggplot(NBA_Drafts, aes(Raw_Star_Score, Total_Accolades)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = paste0("Raw Star Score vs Total Accolades (r = ", round(r_all,
3), ")"),
       x = "Raw Star Score (WS × BPM)", y = "Total Accolades")
save_plot(p_all, "raw_star_vs_accolades_all")

## `geom_smooth()` using formula = 'y ~ x'
```




```
tiers <- list(
  "≥1 Accolade"  = subset(NBA_Drafts, Total_Accolades >= 1),
  "≥3 Accolades" = subset(NBA_Drafts, Total_Accolades >= 3),
  "≥5 Accolades" = subset(NBA_Drafts, Total_Accolades >= 5)
)

tier_correlations <- lapply(names(tiers), function(nm) {
  df <- tiers[[nm]]
  tibble(
    Tier = nm,
    Correlation = round(safe_cor(df$Raw_Star_Score, df$Total_Accolades), 3),
    N = nrow(df)
  )
```

```r
}) %>% bind_rows()

print(tier_correlations)

## # A tibble: 3 × 3
##   Tier         Correlation     N
##   <chr>              <dbl> <int>
## 1 ≥1 Accolade        0.762    50
## 2 ≥3 Accolades       0.76     22
## 3 ≥5 Accolades       0.751    16

p_tiers <- ggplot(tier_correlations, aes(Tier, Correlation, group = 1)) +
  geom_line(color = "blue") +
  geom_point(size = 3, color = "red") +
  ylim(0, 1) +
  labs(title = "Correlation: Raw Star Score vs Total Accolades by Tier",
       x = "Accolade Tier", y = "Correlation")
save_plot(p_tiers, "tiered_correlations")
```




```r
compare_cor <- function(r1, n1, r2, n2) {
  if (any(is.na(c(r1, r2))) || min(n1, n2) < 4) return(tibble(z_stat = NA,
p_value = NA))
  r1 <- max(min(r1, 0.9999), -0.9999)
  r2 <- max(min(r2, 0.9999), -0.9999)
```

```
  z1 <- 0.5 * log((1 + r1) / (1 - r1))
  z2 <- 0.5 * log((1 + r2) / (1 - r2))
  se <- sqrt(1 / (n1 - 3) + 1 / (n2 - 3))
  z  <- (z1 - z2) / se
  p  <- 2 * (1 - pnorm(abs(z)))
  tibble(z_stat = z, p_value = p)
}

r1 <- tier_correlations$Correlation[1]; n1 <- tier_correlations$N[1]
r2 <- tier_correlations$Correlation[3]; n2 <- tier_correlations$N[3]
print(compare_cor(r1, n1, r2, n2))

## # A tibble: 1 × 2
##   z_stat p_value
##    <dbl>   <dbl>
## 1 0.0821   0.935
```

In this step, we tested whether a player's raw Star Score (WS × BPM, no accolades baked in) actually lines up with their total accolades. The goal was to check if the metric is unintentionally biased or if it naturally reflects recognized player value.

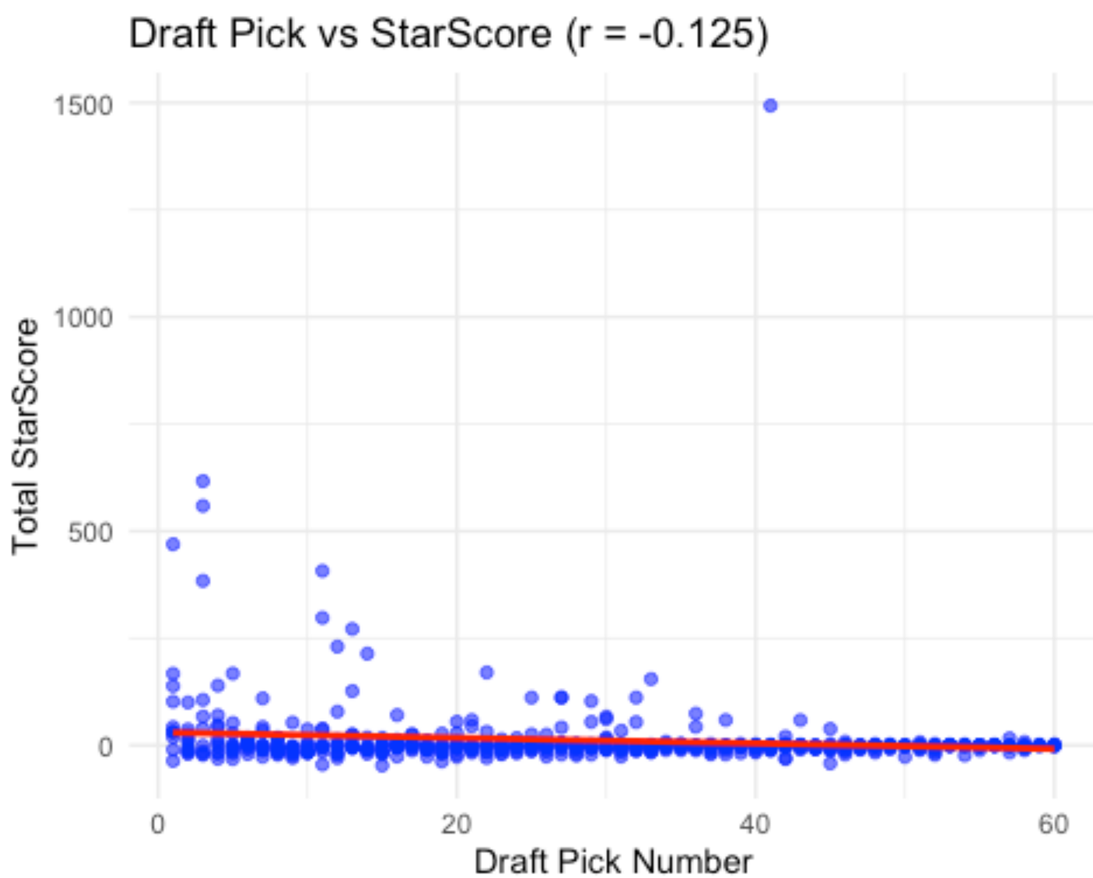
We found a strong positive correlation (r = 0.776) between total accolades and raw Star Score. This relationship holds steady across different accolade tiers ≥1 accolade, ≥3 accolades, and ≥5 accolades all produce similar correlation values. That means no matter how many accolades you set as the cutoff, the spread remains consistent, and the metric is capturing something real.

In plain terms: players with higher raw Star Scores tend to have more accolades, but the score itself isn't just a mirror of accolades it's picking up on actual production and impact. This also confirms there's no major bias introduced by using Star Score before we layer in accolade weighting.

To double-check, we ran a Fisher's z-test comparing the correlation for the ≥1 accolade group vs. the ≥5 accolade group. The z-statistic was 0.0820 and the p-value was 0.93, meaning the difference in correlation between those two groups is statistically insignificant. That reinforces the takeaway: the correlation is stable regardless of accolade tier.

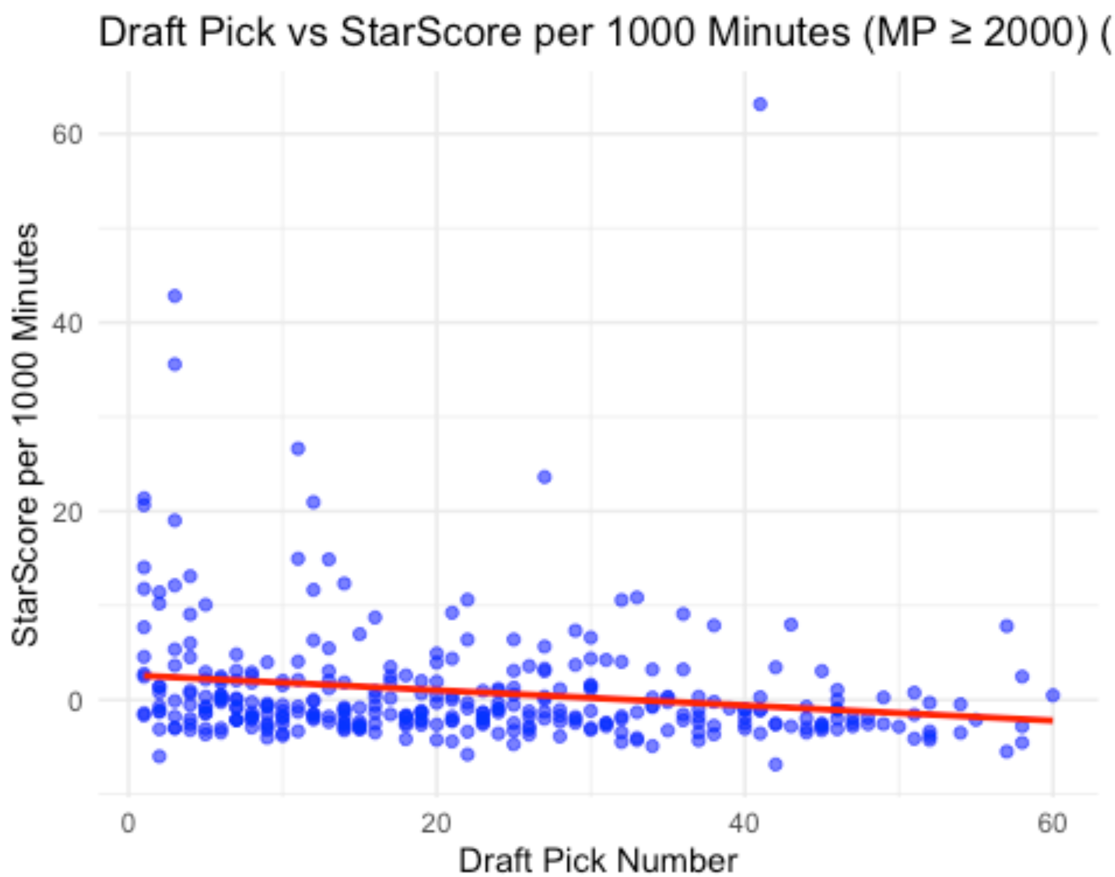
```
# --------------------------
# 5) Draft Pick vs Star Score
# --------------------------
r_overall <- safe_cor(NBA_Drafts$NBA_Star_Score, NBA_Drafts$Pick)
cat("Correlation (Pick vs StarScore):", round(r_overall, 3), "\n")

## Correlation (Pick vs StarScore): -0.125

p1 <- ggplot(NBA_Drafts, aes(Pick, NBA_Star_Score)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
```

```
    title = paste0("Draft Pick vs StarScore (r = ", round(r_overall, 3),
")"),
    x = "Draft Pick Number", y = "Total StarScore"
  )
save_plot(p1, "pick_vs_star_total")

## `geom_smooth()` using formula = 'y ~ x'
```




```
NBA_Drafts <- NBA_Drafts %>%
  mutate(Star_perKmin = ifelse(MP > 0, NBA_Star_Score / (MP / 1000),
NA_real_))

r_perK_all <- safe_cor(NBA_Drafts$Star_perKmin, NBA_Drafts$Pick)
cat("Correlation (Pick vs StarScore per 1000 min, all):", round(r_perK_all,
3), "\n")

## Correlation (Pick vs StarScore per 1000 min, all): 0.048

min_minutes <- 2000
df_min <- NBA_Drafts %>% filter(is.finite(Star_perKmin), MP >= min_minutes)

r_perK_min <- safe_cor(df_min$Star_perKmin, df_min$Pick)
cat("Correlation (Pick vs StarScore per 1000 min, MP ≥", min_minutes, "):",
round(r_perK_min, 3), "\n")
```

```
## Correlation (Pick vs StarScore per 1000 min, MP ≥ 2000 ): -0.183

p2 <- ggplot(df_min, aes(Pick, Star_perKmin)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(
    title = paste0("Draft Pick vs StarScore per 1000 Minutes (MP ≥ ",
min_minutes,
                   ") (r = ", round(r_perK_min, 3), ")"),
    x = "Draft Pick Number", y = "StarScore per 1000 Minutes"
  )
save_plot(p2, "pick_vs_star_perKmin_mp2000")

## `geom_smooth()` using formula = 'y ~ x'
```




```
summary_tbl <- tibble::tibble(
  Metric = c("StarScore (total)",
             "StarScore per 1000 min (all)",
             paste0("StarScore per 1000 min (MP ≥ ", min_minutes, ")")),
  Correlation_with_Pick = c(round(r_overall, 3),
                            round(r_perK_all, 3),
                            round(r_perK_min, 3)),
  N = c(sum(is.finite(NBA_Drafts$NBA_Star_Score) &
is.finite(NBA_Drafts$Pick)),
```

```r
        sum(is.finite(NBA_Drafts$Star_perKmin) & is.finite(NBA_Drafts$Pick)),
        nrow(df_min))
)
print(summary_tbl)
```

```
## # A tibble: 3 × 3
##   Metric                              Correlation_with_Pick     N
##   <chr>                                               <dbl> <int>
## 1 StarScore (total)                                  -0.125   546
## 2 StarScore per 1000 min (all)                        0.048   546
## 3 StarScore per 1000 min (MP ≥ 2000)                 -0.183   332
```

Here we're looking at how Star Score relates to draft pick. The overall correlation comes out to r = -0.125, which is negative but not particularly strong. That lines up with the general idea that earlier picks tend to perform better, but there's still a ton of variation. A big part of that variation is likely explained by outliers like Nikola Jokić a second-rounder who posts MVP-level numbers. We'll talk about him later.

When we shift to Star Score per 1,000 minutes, the correlation drops to r = 0.048, basically flat. That tells us draft pick isn't much of a predictor for per-minute production when you include everyone.

Things get more sensible when we filter to players with at least 2,000 minutes played. Here the correlation strengthens to r = -0.183, which still isn't massive but makes more sense—earlier picks generally do better when you remove the fringe guys who barely saw the floor.

Bottom line: Draft position has some predictive value, but between outliers like Jokić and the huge number of low-minute players dragging the numbers around, it's far from perfect. Once you remove those fringe cases, the relationship lines up more with expectations.

```r
# -------------------------
# 6) StarScore v3.6
# -------------------------

# make sure award cols exist & are numeric zeros when missing
award_cols <- c(
  "Allstar Selections","All NBA first team","All NBA second team","All NBA
thrid team",
  "All Defensive first team","All Defensive seond team","MVP","MVP Shares",
  "DPOY","DPOY Top 3 Finish","FMVP"
)
for (cn in award_cols) if (!cn %in% names(NBA_Drafts)) NBA_Drafts[[cn]] <- 0
NBA_Drafts[award_cols] <- lapply(NBA_Drafts[award_cols], \(x)
ifelse(is.na(x), 0, x))

NBA_Drafts$StarScore_v3.6 <- with(
  NBA_Drafts,
  (WS * BPM) +
    (MVP * 350) +
```

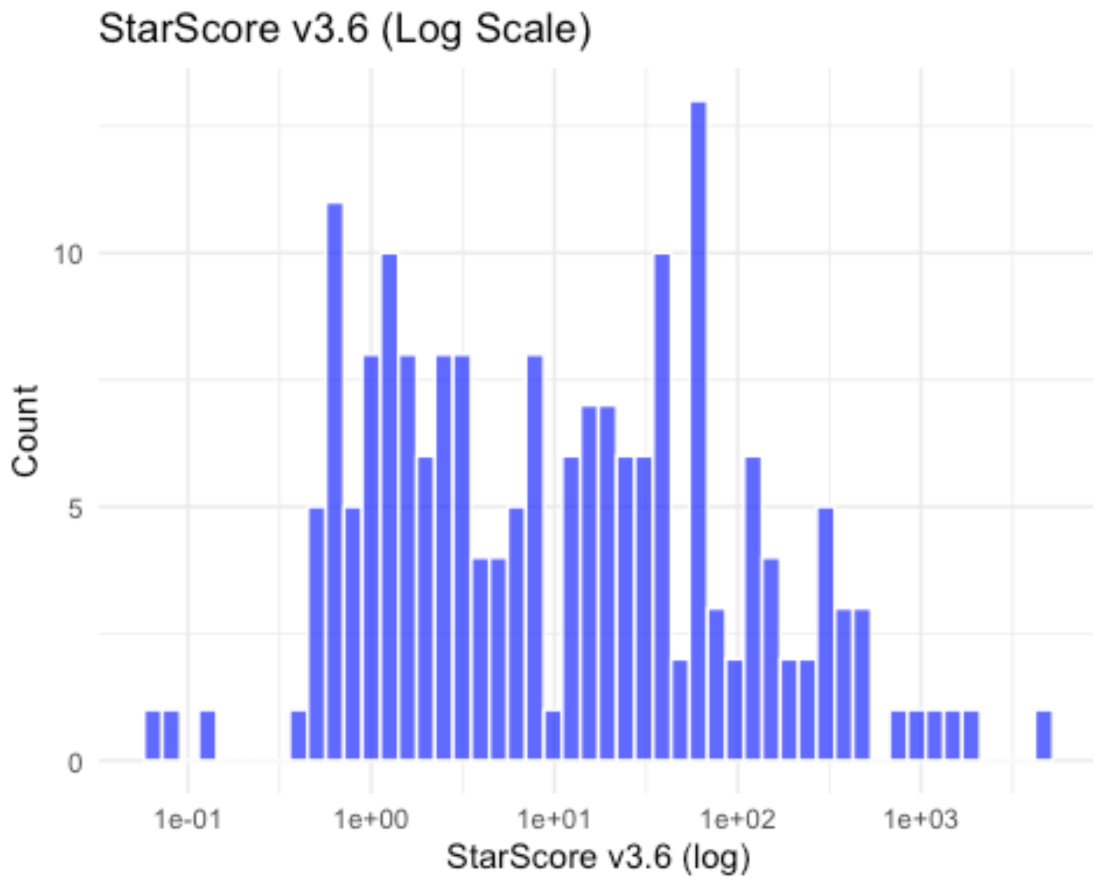
```r
    (FMVP * 275) +
    (DPOY * 60) +
    ((`DPOY Top 3 Finish` > 0) * 40) +
    (`All NBA first team` * 100) +
    (`All NBA second team` * 80) +
    (`All NBA thrid team` * 60) +
    (`Allstar Selections` * 25) +
    (`All Defensive first team` * 20) +
    (`All Defensive seond team` * 15) +
    (`MVP Shares` * 200)
)

top_new_era_stars <- NBA_Drafts[order(-NBA_Drafts$StarScore_v3.6),

c("Player","Pick","WS","BPM","StarScore_v3.6")]
print(head(top_new_era_stars, 100))
```

```
##                       Player Pick    WS  BPM StarScore_v3.6
## 441             Nikola Jokić   41 128.0 10.3        4370.40
## 289              Joel Embiid    3  64.9  6.9        1928.61
## 511 Shai Gilgeous-Alexander   11  60.4  5.5        1653.20
## 379              Luka Dončić    3  57.1  7.6        1252.36
## 275             Jayson Tatum    3  65.5  3.8        1003.30
## 336       Karl-Anthony Towns    1  85.1  4.4         679.44
## 151         Donovan Mitchell   13  51.9  3.5         532.45
## 270             Jaylen Brown    3  40.6  0.3         467.18
## 150         Domantas Sabonis   11  69.9  3.4         438.66
## 139             Devin Booker   13  51.5  1.1         359.85
## 562        Tyrese Haliburton   12  37.6  4.8         350.48
## 238            Jalen Brunson   33  45.5  2.3         343.65
## 36               Bam Adebayo   14  57.6  2.5         339.00
## 39               Ben Simmons    1  35.4  3.0         321.80
## 169              Evan Mobley    3  28.3  2.3         310.29
## 29           Anthony Edwards    1  24.8  1.7         283.16
## 543               Trae Young    5  43.3  2.6         272.58
## 323            Julius Randle    7  49.6  1.0         268.60
## 259        Jaren Jackson Jr.    4  28.8  0.9         230.92
## 222               Ja Morant     2  24.1  2.7         197.27
## 264            Jarrett Allen   22  66.7  2.4         185.08
## 596          Zion Williamson    1  24.2  4.9         168.58
## 364       Kristaps Porziņģis    4  48.0  2.7         154.60
## 394             Marcus Smart    6  34.2 -0.4         146.32
## 249           Jalen Williams   12  20.7  2.2         145.54
## 137            Derrick White   29  39.7  2.2         117.34
## 123             De'Aaron Fox    5  33.9  0.8         112.52
## 484          Robert Williams   27  23.9  4.7         112.33
## 424          Montrezl Harrell  32  42.9  2.6         111.54
## 91              Clint Capela   25  69.6  1.6         111.36
## 460            Pascal Siakam   27  55.5  2.0         111.00
## 61            Cade Cunningham   1   7.0  1.0          94.40
```

```
## 415         Mikal Bridges   10  42.0  0.7          89.40
## 17         Alperen Sengun   16  22.5  2.7          85.75
## 130       Dejounte Murray   29  26.2  1.4          76.68
## 419     Mitchell Robinson   36  35.2  2.1          73.92
## 365         Kyle Anderson   30  39.4  1.7          66.98
## 158         Dyson Daniels    8   9.2  0.5          64.60
## 138          Desmond Bane   30  25.5  2.4          61.20
## 220           Ivica Zubac   32  51.3  0.9          61.17
## 564          Tyrese Maxey   21  26.5  1.3          59.45
## 20          Amen Thompson    4  12.3  3.2          59.36
## 108        Daniel Gafford   38  31.1  1.9          59.09
## 50         Brandon Clarke   21  24.6  2.4          59.04
## 207    Isaiah Hartenstein   43  28.0  2.1          58.80
## 373       Lauri Markkanen    7  33.1  1.0          58.10
## 503        Scottie Barnes    4  19.4  1.7          57.98
## 115        Darius Garland    5  24.6  0.3          57.38
## 131          Delon Wright   20  28.9  1.9          54.91
## 237          Jakob Poeltl    9  44.2  1.2          53.04
## 388       Malcolm Brogdon   36  33.4  1.3          43.42
## 573     Victor Wembanyama    1   7.5  5.7          42.75
## 407      Matisse Thybulle   20  14.1  0.9          42.69
## 372       Larry Nance Jr.   27  34.7  1.2          41.64
## 100      D'Angelo Russell    2  26.5  0.6          40.90
## 370           LaMelo Ball    3  13.6  2.9          39.44
## 429          Myles Turner   11  48.7  0.8          38.96
## 558           Tyler Herro   13  19.9  0.7          38.93
## 156         Dwight Powell   45  47.9  0.8          38.32
## 80          Chet Holmgren    2  12.1  3.1          37.51
## 69        Cameron Johnson   11  22.9  1.6          36.64
## 251          Jamal Murray    7  38.1  0.9          34.29
## 433           Nic Claxton   31  26.1  1.3          33.93
## 459        Paolo Banchero    1  11.1  0.8          33.88
## 579        Walker Kessler   22  16.6  1.9          31.54
## 126         Deandre Ayton    1  37.6  0.8          30.08
## 51         Brandon Ingram    2  27.3  0.1          27.73
## 290          John Collins   19  39.2  0.7          27.44
## 152      Donte DiVincenzo   17  23.9  1.1          26.29
## 591           Zach LaVine   13  37.3  0.7          26.11
## 202     Immanuel Quickley   25  21.3  1.2          25.56
## 467      Payton Pritchard   26  20.9  1.2          25.08
## 550       Trey Murphy III   17  18.4  1.3          23.92
## 450            OG Anunoby   23  31.4  0.2          21.28
## 536         Thomas Bryant   42  20.4  1.0          20.40
## 414    Michael Porter Jr.   14  26.5  0.7          18.55
## 310             Josh Hart   30  36.7  0.5          18.35
## 448            Obi Toppin    8  16.6  1.1          18.26
## 308           Josh Giddey    6  13.9  1.3          18.07
## 400         Mark Williams   15   9.7  1.8          17.46
## 175          Franz Wagner    8  21.4  0.8          17.12
## 567            Tyus Jones   24  33.9  0.5          16.95
```

```
## 239              Jalen Duren   13  18.6  0.9          16.74
## 544    Trayce Jackson-Davis   57   8.2  2.0          16.40
## 351             Kevon Looney   30  37.4  0.4          14.96
## 2               Aaron Gordon    4  46.9  0.3          14.07
## 294            Jonathan Isaac    6  14.0  1.0          14.00
## 197             Herbert Jones   35  15.6 -0.4          13.76
## 136          Dereck Lively II   12   8.4  1.6          13.44
## 332             Jusuf Nurkić   16  33.2  0.4          13.28
## 377               Lonzo Ball    2  12.6  1.0          12.60
## 455           Onyeka Okongwu    6  24.9  0.5          12.45
## 186             Goga Bitadze   18  14.9  0.8          11.92
## 248              Jalen Suggs    5   5.6 -1.1           8.84
## 125         De'Anthony Melton   46  14.3  0.6           8.58
## 499             Santi Aldama   30  12.2  0.7           8.54
## 243            Jalen Johnson   20   8.4  1.0           8.40
## 274           Jaylin Williams   34   7.6  1.1           8.36
## 59               Buddy Hield    6  27.7  0.3           8.31
## 465                Paul Reed   58  11.5  0.7           8.05
```

```
# Log-scale view (shows the right tail more cleanly)
p2 <- ggplot(NBA_Drafts, aes(StarScore_v3.6)) +
  geom_histogram(bins = 50, fill = "blue", color = "white", alpha = 0.7) +
  scale_x_log10() +
  labs(title = "StarScore v3.6 (Log Scale)", x = "StarScore v3.6 (log)", y =
"Count")
save_plot(p2, "v36_log_distribution")
```



At this point we updated the Star Score into a more refined version we're calling Star Score v3.6. The base is still the same Win Shares multiplied by Box Plus-Minus but now we're incorporating awards into the equation to capture a more complete picture of impact. The logic is that accolades represent tangible recognition of a player's value, and the rarer or more prestigious the award, the more weight it carries.

MVP is worth 350 points, since only one player wins it per year and it's the single most important individual award in basketball. Finals MVP is a close second at 275, rewarding postseason dominance on the biggest stage. Defensive Player of the Year is 60 points, recognizing its significance while still accounting for the fact that defense tends to be less statistically impactful than offense. A top-three finish in DPOY voting is worth 40, acknowledging defensive excellence without a full award win.

For All-NBA teams, first team is worth 100, second team 80, and third team 60 a scale that reflects the prestige difference between tiers. All-Star selections are valued at 25, while All-Defensive first and second teams are worth 20 and 15 respectively. These defensive teams matter, but they don't shift a player's overall value as dramatically as the top-tier awards. MVP Shares are multiplied by 200, as they highlight players who played at an MVP-caliber level even if they didn't actually win.

We plotted Star Score v3.6 on a logarithmic scale. A log view lets us see the "right tail" of the distribution more clearly, where the top superstars live, instead of having those players visually drown out the rest. This makes it much easier to study how the top end compares to the middle class of players.

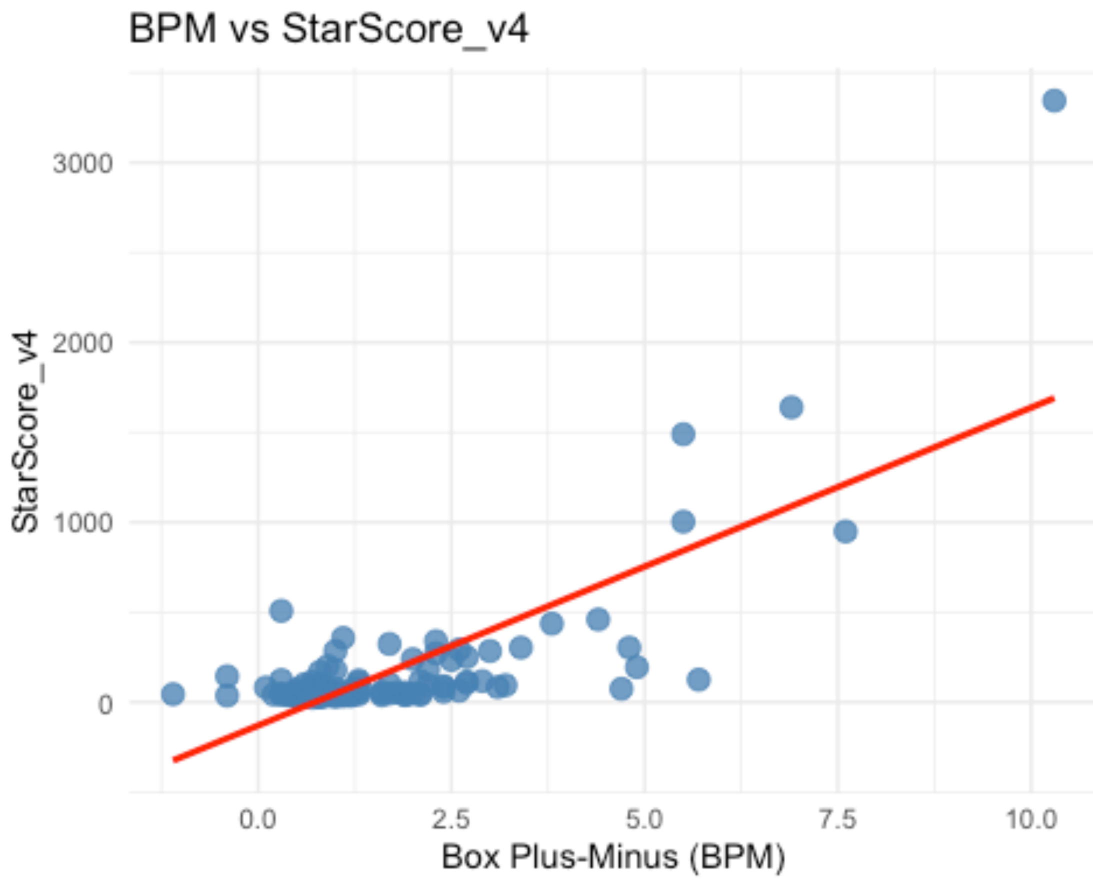
From here, we pulled the top 100 players using Star Score v3.6. This group represents what our metric considers the best in the NBA, factoring in both instant impact and career longevity. These 100 will serve as our filtered Top 100 group moving forward, where we'll run more advanced testing to refine the definition of a "star" and see how strongly draft position can predict stardom.

```
## Rows: 100 Columns: 67
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (9): Player, Team, Height, Pre NBA, Conference, Nationality,
Position, ...
## dbl (58): PER, TS%, 3PAr, FTr, ORB%, DRB%, TRB%, AST%, STL%, BLK%, TOV%,
USG...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```
## # A tibble: 100 × 5
##     Player                    Pick `WS/82g` BPM_x StarScore_v4
##     <chr>                    <dbl>    <dbl> <dbl>        <dbl>
##   1 Nikola Jokić                41     14.1  10.3        3345.
##   2 Joel Embiid                  3     11.8   6.9        1639.
##   3 Shai Gilgeous-Alexander     11     10.7   5.5        1490.
##   4 Luka Dončić                  3     10.4   5.5        1004.
##   5 Jayson Tatum                 3      9.2   7.6         949.
##   6 Jaylen Brown                 3      5.5   0.3         508.
##   7 Donovan Mitchell            13      7.9   4.4         460.
##   8 Karl-Anthony Towns           1     10.8   3.8         437.
##   9 Devin Booker                13      6.3   1.1         357.
##  10 Jalen Brunson               33      7.7   2.3         339.
##  11 Anthony Edwards              1      5.3   1.7         324.
##  12 Domantas Sabonis            11      8.9   3.4         303.
##  13 Tyrese Haliburton           12      4.7   4.8         303.
##  14 Trae Young                   5      7.3   2.6         293.
##  15 Julius Randle                7      5.7   1           285.
##  16 Ben Simmons                  1      7.6   3           284.
##  17 Evan Mobley                  3      8.6   2.3         271.
##  18 Ja Morant                    2      6.4   2.7         250.
##  19 Pascal Siakam               27      7.2   2           239.
##  20 Bam Adebayo                 14      8.3   2.5         234.
##  21 Jaren Jackson Jr.            4      5.8   0.9         200.
##  22 Zion Williamson              1      9.3   4.9         194.
##  23 Jalen Williams              12      7.9   2.2         181.
##  24 Cade Cunningham              1      2.8   1           179.
##  25 De'Aaron Fox                 5      5.2   0.8         169.
```

```
## 26 Marcus Smart               6       4.4  -0.4         144.
## 27 Victor Wembanyama          1       5.2   5.7         125.
## 28 Darius Garland             5       5.3   0.3         123.
## 29 Josh Giddey                6       4.1   1.3         121.
## 30 LaMelo Ball                3       4.8   2.9         117.
## 31 Dejounte Murray           29       4.3   2.1         117.
## 32 Zach LaVine               13       4.7   0.7         115.
## 33 Alperen Sengun            16       6.5   2.7         112.
## 34 Kristaps Porzingis         4       7.9   2.7         109.
## 35 Tyrese Maxey              21       6.8   1.3         108.
## 36 Scottie Barnes             4       5.8   1.7         101.
## 37 D'Angelo Russell           2       3.5   0.6         100.
## 38 Paolo Banchero             1       4.6   0.8          99.6
## 39 Derrick White             29       6.6   2.2          96.0
## 40 Amen Thompson              4       7.7   3.2          94.4
## 41 Tyler Herro               13       4.5   0.7          91.2
## 42 Jarrett Allen             22       9.6   2.4          88.9
## 43 Chet Holmgren              2       8.7   3.1          83.8
## 44 Desmond Bane              30       6.7   2.4          83.8
## 45 Mikal Bridges             10       6.2   0.7          83.8
## 46 Brandon Ingram             2       4.5   0.1          81.9
## 47 Lauri Markkanen            7       6     1            76.4
## 48 Robert Williams           27       8.3   4.7          74.5
## 49 Malcolm Brogdon           36       5.9   1.3          74.0
## 50 Dyson Daniels              8       3.9   0.5          71.3
## 51 Immanuel Quickley         25       5.4   1.2          70.1
## 52 Jamal Murray               7       5.8   0.9          70.1
## 53 Franz Wagner               8       6     0.8          67.3
## 54 Montrezl Harrell          32       6.8   2.6          63.4
## 55 Tyus Jones                24       3.9   0.5          59.6
## 56 Brandon Clarke            21       6.6   2.4          57.4
## 57 Cameron Johnson           11       5.5   1.6          56.4
## 58 Lonzo Ball                 2       3.6   1            56.0
## 59 Delon Wright              20       4.3   1.9          54.6
## 60 Payton Pritchard          26       4.9   1.2          54.1
## 61 Trey Murphy III           17       6     1.3          53.7
## 62 Mark Williams             15       7.5   1.8          53.6
## 63 Clint Capela              25       8.6   1.6          49.2
## 64 Kyle Anderson             30       4.5   1.7          48.1
## 65 Trayce Jackson-Davis      57       5.2   2            47.0
## 66 Donte DiVincenzo          17       4.7   1.1          46.3
## 67 Deandre Ayton              1       7.8   0.8          45.4
## 68 Jalen Johnson             20       3.7   1            45.0
## 69 Ivica Zubac               32       7.2   0.9          44.7
## 70 Buddy Hield                6       3.2   0.3          44.4
## 71 Jalen Suggs                5       2.2  -1.1          44.2
## 72 Michael Porter Jr.        14       6.3   0.7          44.0
## 73 Isaiah Hartenstein        43       6.1   2.1          43.9
## 74 Aaron Gordon               4       5.3   0.3          43.6
## 75 Nic Claxton               31       6.9   1.3          43.0
```

```
##  76 Mitchell Robinson         36      8.6   2.1         42.4
##  77 OG Anunoby                23      5.2   0.2         42.3
##  78 Matisse Thybulle          20      3.3   0.9         42.2
##  79 John Collins              19      6.8   0.7         41.7
##  80 Daniel Gafford            38      6.7   1.9         40.6
##  81 De'Anthony Melton         46      3.3   0.6         40.3
##  82 Jusuf Nurkić              16      4.6   0.4         40.2
##  83 Santi Aldama              30      4.3   0.7         40.2
##  84 Myles Turner              11      6.2   0.8         39.5
##  85 Walker Kessler            22      6.9   1.9         38.9
##  86 Obi Toppin                 8      3.8   1.1         38.5
##  87 Dereck Lively II          12      7.6   1.6         37.9
##  88 Josh Hart                 30      5.7   0.5         37.2
##  89 Jakob Poeltl               9      6.1   1.2         37.1
##  90 Kevon Looney              30      5.1   0.4         37.0
##  91 Larry Nance Jr.           27      5.2   1.2         36.6
##  92 Herbert Jones             35      5.3  -0.4         36.5
##  93 Thomas Bryant             42      5     1           36.1
##  94 Jalen Duren               13      7.4   0.9         34.5
##  95 Jaylin Williams           34      3.8   1.1         33.9
##  96 Jonathan Isaac             6      4.1   1           30.8
##  97 Onyeka Okongwu             6      6.7   0.5         28.3
##  98 Goga Bitadze              18      3.8   0.8         27.6
##  99 Dwight Powell             45      5.6   0.8         27.3
## 100 Paul Reed                 58      3.6   0.7         24.3

## `geom_smooth()` using formula = 'y ~ x'
```

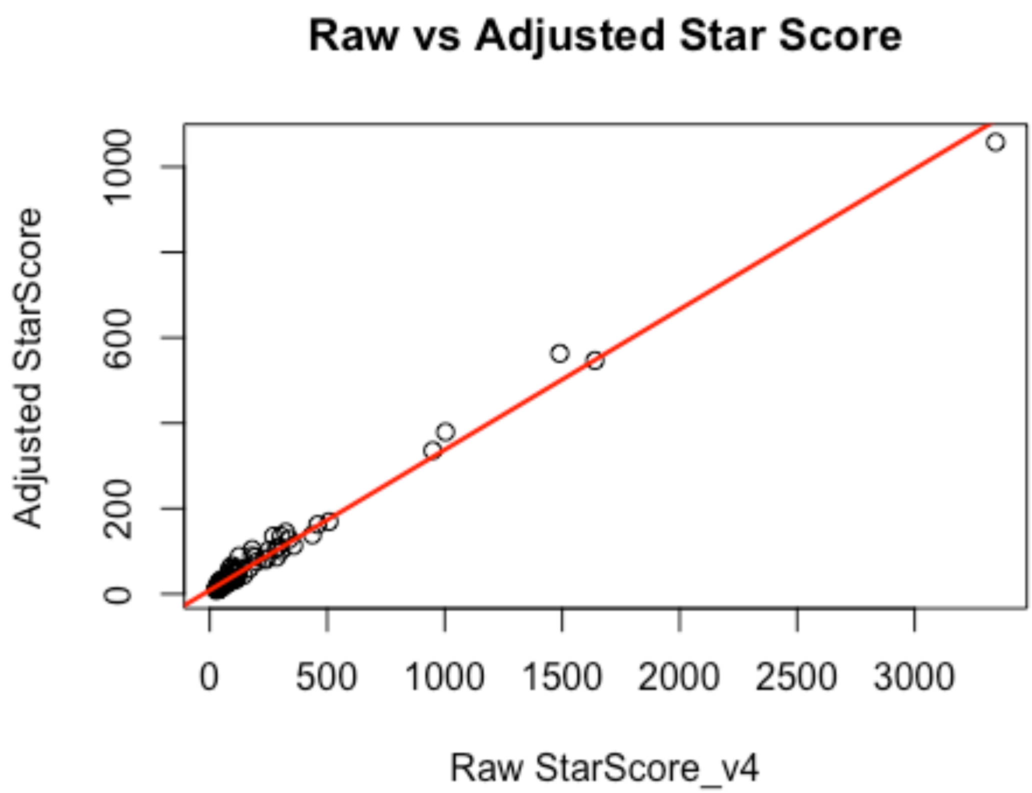



```
##                StarScore_v4      WS/82g     BPM_x        VORP        PER
TS%
## StarScore_v4    1.00000000 0.63143201 0.7539854 0.72230822 0.5866652
0.07535928
## WS/82g          0.63143201 1.00000000 0.7366213 0.66744359 0.7920189
0.53209494
## BPM_x           0.75398545 0.73662131 1.0000000 0.85396481 0.8101810
0.28072678
## VORP            0.72230822 0.66744359 0.8539648 1.00000000 0.6831783
0.07516445
## PER             0.58666520 0.79201890 0.8101810 0.68317830 1.0000000
0.51746570
## TS%             0.07535928 0.53209494 0.2807268 0.07516445 0.5174657
1.00000000
## AST%            0.38732463 0.04935636 0.3978044 0.49795983 0.1897463 -
0.44718198
##                       AST%
## StarScore_v4  0.38732463
## WS/82g        0.04935636
## BPM_x         0.39780437
## VORP          0.49795983
## PER           0.18974628
```

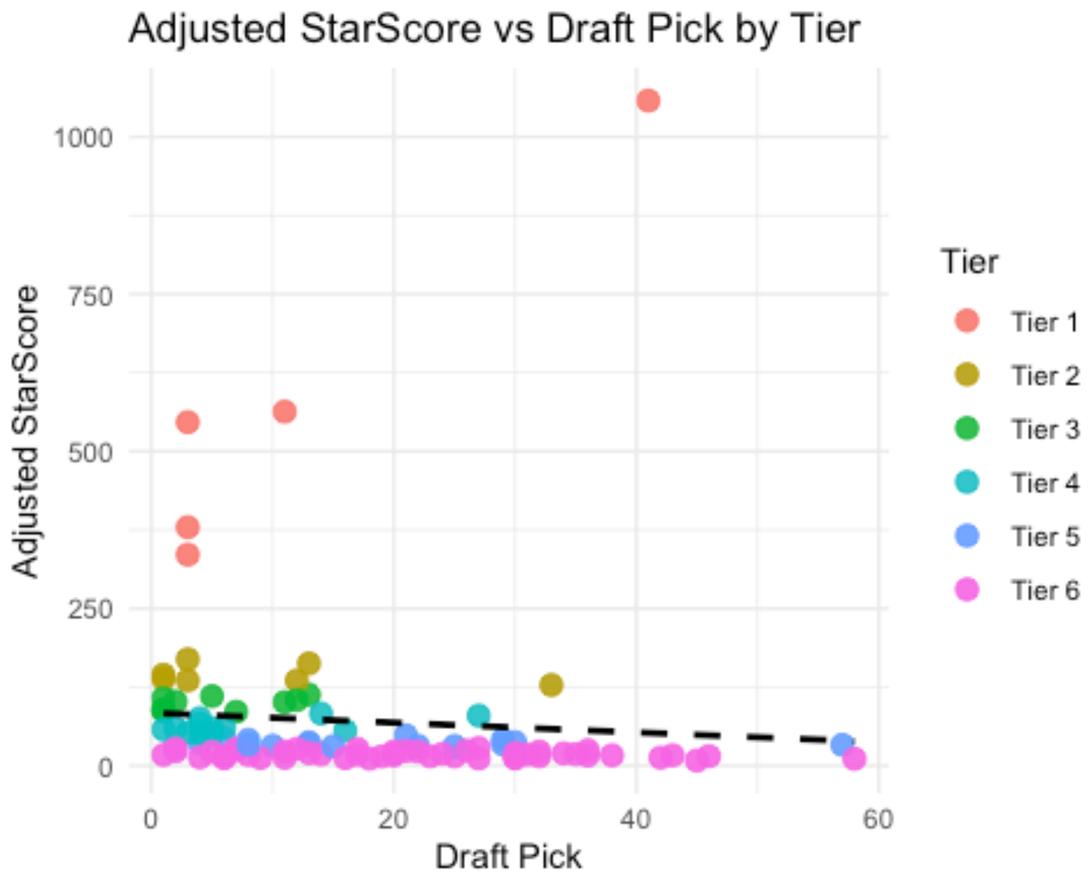
```
## TS%          -0.44718198
## AST%          1.00000000

## # A tibble: 100 × 5
##     Player                   Pick Years StarScore_v4 Adjusted_StarScore
##     <chr>                   <dbl> <dbl>        <dbl>              <dbl>
##   1 Nikola Jokić               41    10       3345.              1058.
##   2 Shai Gilgeous-Alexander    11     7       1490.               563.
##   3 Joel Embiid                 3     9       1639.               546.
##   4 Luka Dončić                 3     7       1004.               379.
##   5 Jayson Tatum                3     8        949.               336.
##   6 Jaylen Brown                3     9        508.               169.
##   7 Donovan Mitchell           13     8        460.               163.
##   8 Anthony Edwards             1     5        324.               145.
##   9 Karl-Anthony Towns          1    10        437.               138.
##  10 Evan Mobley                 3     4        271.               136.
##  11 Tyrese Haliburton          12     5        303.               136.
##  12 Jalen Brunson              33     7        339.               128.
##  13 Devin Booker               13    10        357.               113.
##  14 Trae Young                  5     7        293.               111.
##  15 Ben Simmons                 1     7        284.               107.
##  16 Jalen Williams             12     3        181.               104.
##  17 Ja Morant                   2     6        250.               102.
##  18 Domantas Sabonis           11     9        303.               101.
##  19 Cade Cunningham             1     4        179.                89.6
##  20 Victor Wembanyama           1     2        125.                88.3
##  21 Zion Williamson             1     5        194.                86.9
##  22 Julius Randle               7    11        285.                85.8
##  23 Bam Adebayo                14     8        234.                82.9
##  24 Pascal Siakam              27     9        239.                79.7
##  25 Jaren Jackson Jr.           4     7        200.                75.7
##  26 Amen Thompson               4     2         94.4               66.8
##  27 Josh Giddey                 6     4        121.                60.6
##  28 De'Aaron Fox                5     8        169.                59.6
##  29 Chet Holmgren               2     2         83.8               59.3
##  30 Paolo Banchero              1     3         99.6               57.5
##  31 Alperen Sengun             16     4        112.                55.8
##  32 LaMelo Ball                 3     5        117.                52.5
##  33 Scottie Barnes              4     4        101.                50.6
##  34 Darius Garland              5     6        123.                50.2
##  35 Tyrese Maxey               21     5        108.                48.3
##  36 Marcus Smart                6    11        144.                43.4
##  37 Dejounte Murray            29     8        117.                41.5
##  38 Dyson Daniels               8     3         71.3               41.2
##  39 Desmond Bane               30     5         83.8               37.5
##  40 Tyler Herro                13     6         91.2               37.2
##  41 Kristaps Porzingis          4     9        109.                36.5
##  42 Zach LaVine                13    11        115.                34.7
##  43 Derrick White              29     8         96.0               33.9
##  44 Franz Wagner                8     4         67.3               33.6
```

```
##   45 Trayce Jackson-Davis        57     2           47.0                33.2
##   46 Mikal Bridges              10     7           83.8                31.7
##   47 D'Angelo Russell            2    10          100.                 31.6
##   48 Jarrett Allen              22     8           88.9                31.4
##   49 Immanuel Quickley          25     5           70.1                31.3
##   50 Mark Williams              15     3           53.6                31.0
##   51 Robert Williams            27     7           74.5                28.2
##   52 Brandon Ingram              2     9           81.9                27.3
##   53 Lauri Markkanen             7     8           76.4                27.0
##   54 Trey Murphy III            17     4           53.7                26.8
##   55 Dereck Lively II           12     2           37.9                26.8
##   56 Jamal Murray                7     8           70.1                24.8
##   57 Malcolm Brogdon            36     9           74.0                24.7
##   58 Payton Pritchard           26     5           54.1                24.2
##   59 Brandon Clarke             21     6           57.4                23.4
##   60 Cameron Johnson            11     6           56.4                23.0
##   61 Lonzo Ball                  2     6           56.0                22.9
##   62 Jalen Johnson              20     4           45.0                22.5
##   63 Walker Kessler             22     3           38.9                22.4
##   64 Montrezl Harrell           32     8           63.4                22.4
##   65 Jalen Suggs                 5     4           44.2                22.1
##   66 Santi Aldama               30     4           40.2                20.1
##   67 Jalen Duren                13     3           34.5                19.9
##   68 Jaylin Williams            34     3           33.9                19.6
##   69 Tyus Jones                 24    10           59.6                18.9
##   70 Herbert Jones              35     4           36.5                18.2
##   71 Michael Porter Jr.         14     6           44.0                18.0
##   72 Nic Claxton                31     6           43.0                17.6
##   73 Donte DiVincenzo           17     7           46.3                17.5
##   74 Delon Wright               20    10           54.6                17.3
##   75 Obi Toppin                  8     5           38.5                17.2
##   76 Matisse Thybulle           20     6           42.2                17.2
##   77 Deandre Ayton               1     7           45.4                17.1
##   78 Isaiah Hartenstein         43     7           43.9                16.6
##   79 Daniel Gafford             38     6           40.6                16.6
##   80 Mitchell Robinson          36     7           42.4                16.0
##   81 De’Anthony Melton          46     7           40.3                15.2
##   82 OG Anunoby                 23     8           42.3                14.9
##   83 Ivica Zubac                32     9           44.7                14.9
##   84 Clint Capela               25    11           49.2                14.8
##   85 Buddy Hield                 6     9           44.4                14.8
##   86 John Collins               19     8           41.7                14.7
##   87 Kyle Anderson              30    11           48.1                14.5
##   88 Aaron Gordon                4    11           43.6                13.1
##   89 Josh Hart                  30     8           37.2                13.1
##   90 Thomas Bryant              42     8           36.1                12.8
##   91 Onyeka Okongwu              6     5           28.3                12.7
##   92 Jonathan Isaac              6     6           30.8                12.6
##   93 Myles Turner               11    10           39.5                12.5
##   94 Jakob Poeltl                9     9           37.1                12.4
```

```
## 95 Jusuf Nurkić                16    11          40.2               12.1
## 96 Kevon Looney                30    10          37.0               11.7
## 97 Larry Nance Jr.             27    10          36.6               11.6
## 98 Goga Bitadze                18     6          27.6               11.3
## 99 Paul Reed                   58     5          24.3               10.9
## 100 Dwight Powell              45    11          27.3                8.22
```




```
## `geom_smooth()` using formula = 'y ~ x'
```




```
## cor(Pick, Adjusted):  -0.078

## cor without Tier1 elites:  -0.416

## cor without Jokic:  -0.316

##
## Call:
## lm(formula = Adjusted_StarScore ~ Pick, data = top100)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -67.71  -50.01  -37.33   -6.18 1005.55
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  83.9583    21.3702   3.929 0.000159 ***
## Pick         -0.7755     0.9950  -0.779 0.437640
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 135.8 on 98 degrees of freedom
## Multiple R-squared:  0.00616,    Adjusted R-squared:  -0.003981
## F-statistic: 0.6074 on 1 and 98 DF,  p-value: 0.4376
```

```
##
## Call:
## lm(formula = Adjusted_StarScore ~ Pick, data = top100_no_tier1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.761 -23.420  -8.605   9.569 113.509
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  64.7194     5.8918  10.985  < 2e-16 ***
## Pick         -1.2025     0.2728  -4.407  2.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36.04 on 93 degrees of freedom
## Multiple R-squared:  0.1728, Adjusted R-squared:  0.1639
## F-statistic: 19.42 on 1 and 93 DF,  p-value: 2.803e-05

##
## Call:
## lm(formula = Adjusted_StarScore ~ Pick, data = top100_no_jokic)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -76.97 -40.30 -18.60   7.21 490.47
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   96.263     13.900   6.925 4.75e-10 ***
## Pick          -2.151      0.656  -3.278  0.00145 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 88.1 on 97 degrees of freedom
## Multiple R-squared:  0.09975,    Adjusted R-squared:  0.09047
## F-statistic: 10.75 on 1 and 97 DF,  p-value: 0.00145
```

At this stage, we filtered the player pool down and brought in everyone's advanced stats so the playing field is level. We added what I'm calling a "self-creation bonus", your assist percentage (×1.5) plus your usage percentage (×1.2) minus your turnover percentage (×1.5). This fixes a flaw in earlier versions where bigs who only score efficiently around the rim were getting inflated scores just because WS and BPM tend to like them.

We also swapped over to Win Shares per 82 games instead of total WS so younger players aren't punished just for not having long careers yet. VORP (value over replacement player) is now in there too, ×10 for weight. On the award side, we tuned some values down. All-Defensive 1st and 2nd teams are now 15 and 10 points, DPOY is 40, and a top-three DPOY finish is 20. This stops defense only guys from being overrated, while still rewarding elite defense.

The result is StarScore v4, which is way less biased toward bigs, pure defenders, or high-usage gunners. Some names at the top are the same, but there's been a good shake-up. We plotted BPM vs StarScore v4 and still see a clear correlation, which means even with the new layers, the metric is holding together.

From there, we made an Adjusted StarScore: take v4 per year (v4 ÷ years played), then multiply by the square root of years played. That way, you still get some credit for longevity but not so much that young players can't compete. It cleans the list up a lot.

We split Adjusted StarScore into tiers:

Tier 6 (<30) – good role players, add value but not special

Tier 5 (30–50) – fringe All-Stars, strong 3rd options

Tier 4 (50–85) – winning impact guys, high-end 3rd or faux 2nd options, some young risers

Tier 3 (85–120) – legit 2nd options, sometimes 1st on bad teams, might have one flaw holding them back

Tier 2 (120–300) – star-level players just shy of superstar status, or elite #2s

Tier 1 (300+) – generational/transcendent players

Only five guys hit Tier 1: Nikola Jokić, Joel Embiid, Shai Gilgeous-Alexander, Luka Dončić, and Jayson Tatum — and there's a real gap between them and everyone else. The plot of Adjusted StarScore vs draft pick shows this clearly, with tiers layered in so you can see the separation.

Finally, we ran correlations and regressions on draft pick vs Adjusted StarScore under three conditions — with the full Top 100, with Tier 1 players removed, and with just Nikola Jokić removed. With everyone included, the p-value came out to 0.4376, showing no strong relationship. Remove the Tier 1 guys, and suddenly the p-value drops to 2.803e-05 a huge jump in statistical significance. Remove just Jokić, and it's 0.0145.

The takeaway? You can generally predict the trajectory of a star — players drafted higher tend to end up better overall — but that pattern completely falls apart with true superstars. Tier 1 guys like Jokić, Embiid, Luka, Shai, and Tatum don't follow the normal rules. They're generational, once-in-a-lifetime outliers who arrive when they arrive, and no draft model is going to reliably spot them years in advance. You can predict stars. You can't predict superstars.

```
## # A tibble: 6 × 19
##   Tier                    `Total Drafted` `Pct of Total Drafted`
##   <chr>                             <int>                  <dbl>
## 1 Tier 1 and up                         5                  0.838
## 2 Tier 2 and up                        12                  2.01
## 3 Tier 3 and up                        22                  3.69
## 4 Tier 4 and up                        34                  5.70
## 5 Tier 5 and up                        50                  8.38
## 6 Tier 6 and up (Top 100)             100                 16.8
```

```
##   `1st Round (1–30)` `Pct of 1st Rounders` `2nd Round (31+)`
##                <int>                 <dbl>             <int>
## 1                  4                  1.33                 1
## 2                 10                  3.33                 2
## 3                 20                  6.67                 2
## 4                 32                 10.7                  2
## 5                 47                 15.7                  3
## 6                 84                 28                   16
##   `Pct of 2nd Rounders` `Lottery Picks (1–14)` `Pct of Lottery Picks`
##                   <dbl>                  <int>                  <dbl>
## 1                 0.337                      4                   2.86
## 2                 0.673                     10                   7.14
## 3                 0.673                     20                  14.3
## 4                 0.673                     30                  21.4
## 5                 1.01                      38                  27.1
## 6                 5.39                      55                  39.3
##   `Late 1st Round (15–30)` `Pct of Late 1st` `Top 10 Picks` `Pct of Top 
10`
##                      <int>             <dbl>          <int>          
<dbl>
## 1                        0              0                 3          
3
## 2                        0              0                 7          
7
## 3                        0              0                14          
14
## 4                        2              1.25             23          
23
## 5                        9              5.62             29          
29
## 6                       29             18.1              41          
41
##   `Top 5 Picks` `Pct of Top 5` `Top 3 Picks` `Pct of Top 3` `#1 Overall`
##           <int>          <dbl>         <int>          <dbl>        <int>
## 1             3              6             3           10              0
## 2             7             14             7           23.3            2
## 3            13             26            12           40              6
## 4            21             42            15           50              7
## 5            23             46            16           53.3            7
## 6            28             56            19           63.3            8
##   `Pct of #1 Overall`
##                 <dbl>
## 1                   0
## 2                  20
## 3                  60
## 4                  70
## 5                  70
## 6                  80
```

This table is basically our probability chart for drafting different tiers of players depending on your pick. We broke down the entire 2014–2023 draft pool into fixed buckets #1 overall, Top-3, Top-5, Top-10, Lottery (1–14), Late First (15–30), First Round (1-30), and Second Round (31+). Then for each tier cutoff, we calculated the share of those draft slots that actually produced players at that level. The denominators are fixed based on the actual number of picks in each bucket over the decade, so these percentages are real probabilities, not just proportions from our Top-100 list.

It paints a very clear picture: the higher the tier, the lower your odds — and the drop-off is steep. If you have a second-round pick, the data shows it's damn near impossible to bank on getting a Tier 2 or better player. You can get lucky and find a gem, but it's rare enough that relying on it is basically a wish.

It also really shows just how low the odds are of drafting a true star or superstar no matter what pick you have. Obviously it's harder to do it with a lower pick, but once you get into those top-tier categories, the patterns start to break down even for high picks the #1 slot included. Certain draft slots are "reliable" up to a point and then fall off. The #1 pick is the cleanest example — 80% of them landed in our Top 100, but the hit rate drops hard after Tier 2. By the time you get to Tier 1, even the top pick is no sure thing, and predicting a generational player becomes impossible.

The higher you go in the tier scale, the more the statistics flatten and become chaotic superstars and transcendent players simply don't follow the same probability rules as everyone else. On the other end, the lower tiers (like Tier 5 and Tier 6) are more uniform and predictable. The trends there are much smoother: better picks generally yield better players, and the drop-off is gradual rather than abrupt.

All in all, this part of the analysis does exactly what we wanted — it's as close as you can get to a statistical map of your odds of hitting on different types of players based on your draft position. It shows the smooth curves at the bottom tiers, the jagged unpredictability at the top, and makes it clear that while drafting a solid player can be done with some level of certainty, drafting a superstar is never guaranteed.

```
## # A tibble: 12 × 4
##    Group          NationalityGroup Draft_Hit_Rate Group_Composition_Pct
##    <chr>          <chr>                     <dbl>                 <dbl>
##  1 Tier 6 and up American                    16.9                  70  
##  2 Tier 6 and up International               16.4                  30  
##  3 Tier 5 and up American                     8.7                  72  
##  4 Tier 5 and up International                7.7                  28  
##  5 Tier 4 and up American                     5.6                  67.6
##  6 Tier 4 and up International                6                    32.4
##  7 Tier 3 and up American                     3.4                  63.6
##  8 Tier 3 and up International                4.4                  36.4
##  9 Tier 2 and up American                     1.7                  58.3
## 10 Tier 2 and up International                2.7                  41.7
## 11 Tier 1 and up American                     0.2                  20  
## 12 Tier 1 and up International                2.2                  80
```

When we break things down by nationality, the results are surprisingly uniform. Americans and international players post very similar hit rates at every tier cutoff once you adjust for how many of each group actually get drafted. The differences in raw counts mostly come from the fact that far more Americans enter the draft pool, not from one group being drastically more likely to succeed.

Unlike draft position, where the patterns start to fall apart as you move toward the superstar tiers, nationality stays pretty consistent. You don't see that same "prediction gets thrown out the window" effect — the rates remain steady from Tier 6 all the way up to Tier 2.

That said, Tier 1 once again breaks the rules. Four of the top five players in Tier 1 are internationals, which reinforces what we saw earlier: true superstars are anomalies. You can't predict them with the same tools you use for stars or role players. These generational talents have intangibles that don't show up in the numbers, and when they emerge, they disrupt the otherwise clean, predictable patterns in the data,plus a bigger global talent pool. Net: prediction works okay for "stars," but once you're talking generational guys, the patterns break.

```
## # A tibble: 18 × 4
##    Group         PositionGroup Draft_Hit_Rate Group_Composition_Pct
##    <chr>         <chr>                  <dbl>                 <dbl>
##  1 Tier 6 and up Center                   3.4                  20
##  2 Tier 6 and up Forward                  6.9                  41
##  3 Tier 6 and up Guard                    6.5                  39
##  4 Tier 5 and up Center                   1.3                  16
##  5 Tier 5 and up Forward                  2.3                  28
##  6 Tier 5 and up Guard                    4.7                  56
##  7 Tier 4 and up Center                   1                    17.6
##  8 Tier 4 and up Forward                  1.7                  29.4
##  9 Tier 4 and up Guard                    3                    52.9
## 10 Tier 3 and up Center                   0.8                  22.7
## 11 Tier 3 and up Forward                  0.7                  18.2
## 12 Tier 3 and up Guard                    2.2                  59.1
## 13 Tier 2 and up Center                   0.5                  25
## 14 Tier 2 and up Forward                  0.3                  16.7
## 15 Tier 2 and up Guard                    1.2                  58.3
## 16 Tier 1 and up Center                   0.3                  40
## 17 Tier 1 and up Forward                  0.2                  20
## 18 Tier 1 and up Guard                    0.3                  40
```

For the position breakdown, the results don't really show a strong or consistent correlation between draft tier and position — the distribution stays fairly balanced across most tiers. That said, one small but consistent pattern does pop out: guards and forwards tend to have slightly higher hit rates than centers, with guards often leading the way. This probably ties back to the higher overall skill versatility required to succeed as a guard in today's NBA ball-handling, playmaking, and shot creation are in constant demand, making it easier for elite guards to stand out and sustain long-term value. Centers, by contrast, tend to be more dependent on system fit and have fewer paths to superstardom unless they're truly exceptional.

```
## # A tibble: 12 × 4
##    Group          HS_Rank_Group Draft_Hit_Rate Group_Composition_Pct
##    <chr>          <chr>                  <dbl>                 <dbl>
##  1 Tier 6 and up Ranked                   17.8                  56
##  2 Tier 6 and up Unranked                 15.6                  44
##  3 Tier 5 and up Ranked                   10.2                  64
##  4 Tier 5 and up Unranked                  6.4                  36
##  5 Tier 4 and up Ranked                    6.7                  61.8
##  6 Tier 4 and up Unranked                  4.6                  38.2
##  7 Tier 3 and up Ranked                    4.5                  63.6
##  8 Tier 3 and up Unranked                  2.8                  36.4
##  9 Tier 2 and up Ranked                    2.5                  66.7
## 10 Tier 2 and up Unranked                  1.4                  33.3
## 11 Tier 1 and up Ranked                    1                    60
## 12 Tier 1 and up Unranked                  0.7                  40
```

For the high school ranking analysis, the data shows that players who were ranked in the HS Top 100 tend to have about a 2–3% higher hit rate in the broader, lower tiers (Tier 4–6). As we move up into higher tiers, that gap narrows — dropping to roughly 1.5% at Tier 3, about 1% at Tier 2, and disappearing almost entirely at Tier 1. This follows the same theme we've seen elsewhere: the higher the talent level, the less predictable success becomes. At the role-player to solid-starter level, there's a clear pattern that high school scouting captures well, but once you're dealing with true stars and superstars, the predictability fades. Those players tend to defy pre-draft rankings entirely, reinforcing the idea that all-time talent is more about rare, unquantifiable traits than measurable pre-draft indicators.

```
## # A tibble: 12 × 4
##    Group          OneAndDone        Draft_Hit_Rate Group_Composition_Pct
##    <chr>          <chr>                      <dbl>                 <dbl>
##  1 Tier 6 and up Not One-and-Done             12.9                  53
##  2 Tier 6 and up One-and-Done                 25.4                  47
##  3 Tier 5 and up Not One-and-Done              5.3                  44
##  4 Tier 5 and up One-and-Done                 15.1                  56
##  5 Tier 4 and up Not One-and-Done              3.4                  41.2
##  6 Tier 4 and up One-and-Done                 10.8                  58.8
##  7 Tier 3 and up Not One-and-Done              1.9                  36.4
##  8 Tier 3 and up One-and-Done                  7.6                  63.6
##  9 Tier 2 and up Not One-and-Done              1.2                  41.7
## 10 Tier 2 and up One-and-Done                  3.8                  58.3
## 11 Tier 1 and up Not One-and-Done              0.5                  40
## 12 Tier 1 and up One-and-Done                  1.6                  60
```

For the one-and-done analysis, the numbers show a clear and consistent trend — players who entered the NBA after just one college season have a noticeably higher hit rate across every tier compared to players who stayed longer. In some tiers, the gap is massive, reaching as high as 10–12%, and even at 7%. This makes sense when you consider what the one-and-done path usually signals: these are players talented enough to make the leap almost immediately, often with elite athletic tools, advanced skills for their age, and enormous developmental upside. In contrast, players who stay multiple years in college tend to be more polished but may have already hit closer to their ceiling, which lowers the likelihood of them breaking into the higher tiers of NBA success. This pattern reinforces that early entry often correlates with higher long-term star potential.

```
##
## ==============================
## Chi-squared tests for cutpoint:  Tier 2+
## ==============================
##
## --- PositionGroup vs Tier2+ ---
##          y
## x         Below2 Tier2+
##   Center       2     18
##   Forward      1     40
##   Guard        2     37
## Chi-sq = 1.62  df = NA  p = 0.4389  (simulated p-value)
##
## --- Nationality vs Tier2+ ---
##                        y
## x                       Below2 Tier2+
##   Australia                  0      4
##   Austria                    0      1
##   Bahamas                    0      2
##   Bosnia and Herzegovina     0      1
##   Cameroon                   1      1
##   Canada                     1      3
##   Croatia                    0      1
##   Dominican Republic         0      1
##   Finland                    0      1
##   France                     0      1
##   Georgia                    0      1
##   Germany                    0      2
##   Latvia                     0      1
##   Lithuania                  0      1
##   Serbia                     1      0
##   Slovenia                   1      0
##   Spain                      0      1
##   Switzerland                0      1
##   Turkey                     0      1
##   U.S. Virgin Islands        0      1
##   United Kingdom             0      1
##   United States              1     69
## Chi-sq = 52.932  df = NA  p = 0.04359  (simulated p-value)
##
## --- HS_Rank_Top100 vs Tier2+ ---
```

```
##           y
## x          Below2 Tier2+
##   Ranked        3     53
##   Unranked      2     42
## Chi-sq = 0.034  df = NA  p = 1  (simulated p-value)
##
## --- OneAndDone vs Tier2+ ---
##                   y
## x                  Below2 Tier2+
##   Not One-and-Done      2     51
##   One-and-Done          3     44
## Chi-sq = 0.357  df = NA  p = 0.6791  (simulated p-value)
##
## ==============================
## Chi-squared tests for cutpoint:  Tier 3+
## ==============================
##
## --- PositionGroup vs Tier3+ ---
##          y
## x         Below3 Tier3+
##   Center       3     17
##   Forward      2     39
##   Guard        7     32
## Chi-sq = 3.447  df = NA  p = 0.202  (simulated p-value)
##
## --- Nationality vs Tier3+ ---
##                         y
## x                        Below3 Tier3+
##   Australia                   0      4
##   Austria                     0      1
##   Bahamas                     0      2
##   Bosnia and Herzegovina      0      1
##   Cameroon                    1      1
##   Canada                      1      3
##   Croatia                     0      1
##   Dominican Republic          1      0
##   Finland                     0      1
##   France                      0      1
##   Georgia                     0      1
##   Germany                     0      2
##   Latvia                      0      1
##   Lithuania                   0      1
##   Serbia                      1      0
##   Slovenia                    1      0
##   Spain                       0      1
##   Switzerland                 0      1
##   Turkey                      0      1
##   U.S. Virgin Islands         0      1
##   United Kingdom              0      1
##   United States               7     63
```

```
## Chi-sq = 28.504  df = NA  p = 0.2016  (simulated p-value)
##
## --- HS_Rank_Top100 vs Tier3+ ---
##           y
## x          Below3 Tier3+
##   Ranked        8     48
##   Unranked      4     40
## Chi-sq = 0.63  df = NA  p = 0.5381  (simulated p-value)
##
## --- OneAndDone vs Tier3+ ---
##                   y
## x                  Below3 Tier3+
##   Not One-and-Done      5     48
##   One-and-Done          7     40
## Chi-sq = 0.281  df = 1  p = 0.5959
##
## ==============================
## Chi-squared tests for cutpoint:  Tier 4+
## ==============================
##
## --- PositionGroup vs Tier4+ ---
##          y
## x         Below4 Tier4+
##   Center       5     15
##   Forward      4     37
##   Guard       13     26
## Chi-sq = 6.606  df = NA  p = 0.03259  (simulated p-value)
##
## --- Nationality vs Tier4+ ---
##                         y
## x                        Below4 Tier4+
##   Australia                   1      3
##   Austria                     0      1
##   Bahamas                     0      2
##   Bosnia and Herzegovina      0      1
##   Cameroon                    1      1
##   Canada                      1      3
##   Croatia                     0      1
##   Dominican Republic          1      0
##   Finland                     0      1
##   France                      1      0
##   Georgia                     0      1
##   Germany                     0      2
##   Latvia                      0      1
##   Lithuania                   1      0
##   Serbia                      1      0
##   Slovenia                    1      0
##   Spain                       0      1
##   Switzerland                 0      1
##   Turkey                      0      1
```

```
##   U.S. Virgin Islands         0      1
##   United Kingdom              0      1
##   United States              14     56
## Chi-sq = 23.077  df = NA  p = 0.3575  (simulated p-value)
##
## --- HS_Rank_Top100 vs Tier4+ ---
##           y
## x          Below4 Tier4+
##   Ranked       14     42
##   Unranked      8     36
## Chi-sq = 0.329  df = 1  p = 0.5661
##
## --- OneAndDone vs Tier4+ ---
##                   y
## x                  Below4 Tier4+
##   Not One-and-Done      8     45
##   One-and-Done         14     33
## Chi-sq = 2.336  df = 1  p = 0.1264
##
## ==============================
## Chi-squared tests for cutpoint:  Tier 5+
## ==============================
##
## --- PositionGroup vs Tier5+ ---
##          y
## x         Below5 Tier5+
##   Center       6     14
##   Forward     10     31
##   Guard       18     21
## Chi-sq = 4.397  df = 2  p = 0.111
##
## --- Nationality vs Tier5+ ---
##                         y
## x                        Below5 Tier5+
##   Australia                   2      2
##   Austria                     0      1
##   Bahamas                     0      2
##   Bosnia and Herzegovina      0      1
##   Cameroon                    2      0
##   Canada                      1      3
##   Croatia                     0      1
##   Dominican Republic          1      0
##   Finland                     0      1
##   France                      1      0
##   Georgia                     0      1
##   Germany                     0      2
##   Latvia                      0      1
##   Lithuania                   1      0
##   Serbia                      1      0
##   Slovenia                    1      0
```

```
##   Spain                       0      1
##   Switzerland                 0      1
##   Turkey                      1      0
##   U.S. Virgin Islands         0      1
##   United Kingdom              0      1
##   United States              23     47
## Chi-sq = 23.383  df = NA  p = 0.2597  (simulated p-value)
##
## --- HS_Rank_Top100 vs Tier5+ ---
##           y
## x          Below5 Tier5+
##   Ranked       21     35
##   Unranked     13     31
## Chi-sq = 0.386  df = 1  p = 0.5347
##
## --- OneAndDone vs Tier5+ ---
##                   y
## x                  Below5 Tier5+
##   Not One-and-Done     14     39
##   One-and-Done         20     27
## Chi-sq = 2.217  df = 1  p = 0.1365
```

The chi-squared tests largely confirm the trends observed earlier, but they also highlight where those trends are statistically significant and where they may be due to random variation.

Position Group: Across most tier cutpoints, position group does not show a statistically significant relationship with higher-tier placement, with the exception of Tier 4 and above. In this range, guards and forwards have a notably higher representation than centers. This supports the earlier observation that the skill versatility required at guard and forward positions makes it more likely for those players to sustain performance at higher tiers.

Nationality: Nationality distributions remain relatively consistent across most tiers, but Tier 2 and above shows statistical significance. This aligns with the observation that international players are disproportionately represented at the very top of the league. These results reinforce the idea that while nationality is not a strong predictor for most tiers, superstar-level performance is harder to predict and often transcends traditional development pipelines.

High School Ranking: While descriptive statistics suggested that ranked players have a slightly higher hit rate in lower tiers, the chi-squared results show no statistically significant relationship between high school ranking and higher-tier placement. This suggests that ranking systems may help identify good players, but are less effective at identifying eventual stars.

One-and-Done Status: Although descriptive trends show one-and-done players having higher hit rates, the chi-squared tests indicate no statistically significant difference within the top 100 player pool. This is likely a sample size limitation, and while the broader draft pool may show a meaningful effect, it does not hold strongly in this narrowed dataset.

In summary, the statistical tests confirm that most observed trends are not absolute and that once players approach the star or superstar level, factors beyond measurable early-career indicators play a larger role in determining success.