

# College Basketball: A Deep Dive

Abhinav Yerramreddy

2024-04-11

## Adjusted Defensive Efficiency

```
# Load necessary libraries
library(dplyr)

##
## Attaching package: 'dplyr'

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

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

library(ggplot2)
library(RColorBrewer)

# Load the dataset
cbb <- read.csv("archive (1)/cbb.csv")

# Define success metric (e.g., reaching Final Four)
success_metric <- "Final Four"

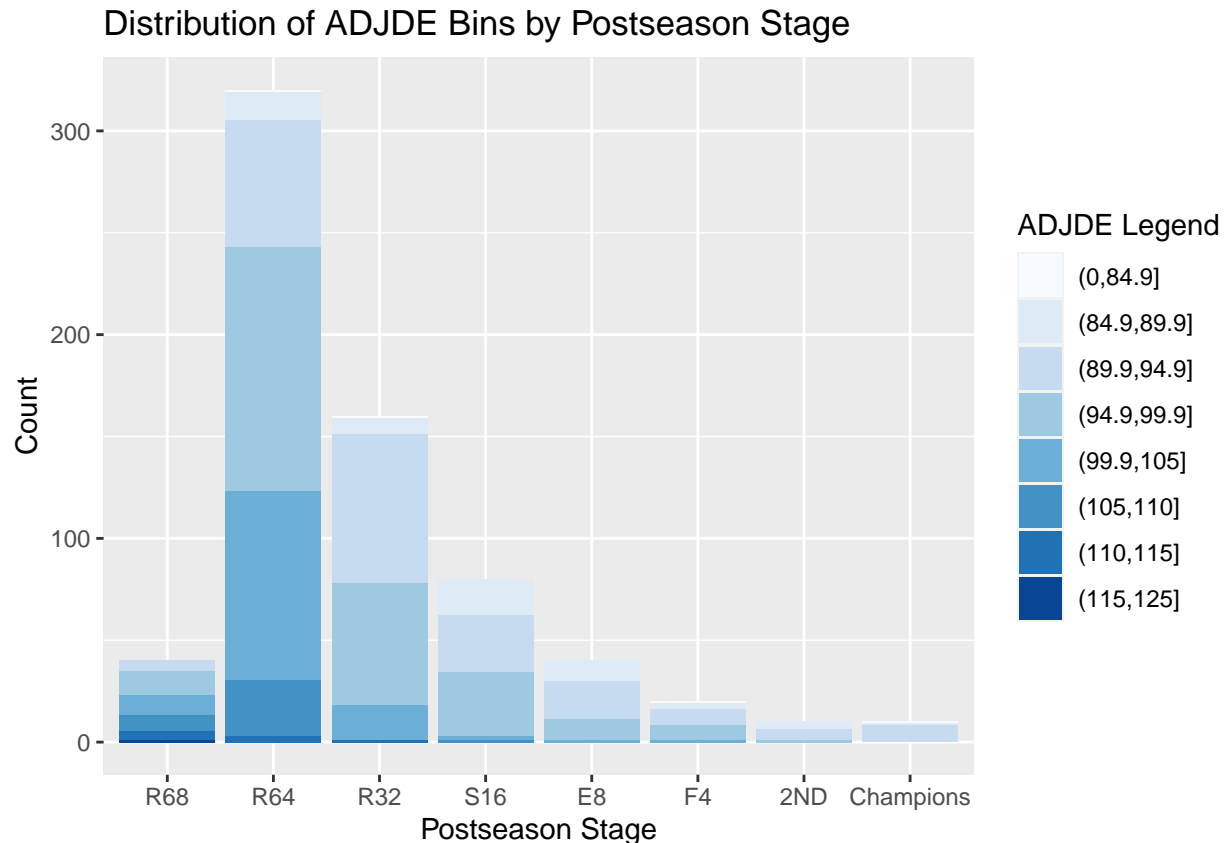
# Data Analysis
# Compare defensive statistics between successful and unsuccessful teams
defensive_stats <- c("ADJDE", "efgd%", "2pd%", "3pd%", "ftrd", "drb")

# Create Bins for Adjusted Defensive Efficiency
cbb_new <- cbb %>%
  mutate(adjde_bins = cut(ADJDE, breaks = c(0, 84.9, 89.9, 94.9,
                                           99.9, 104.9,
                                           109.9, 114.9, 125))) %>%
  filter(POSTSEASON != "N/A")

# Convert POSTSEASON into a numeric scale for correlation
# Assign numerical values based on the levels
```

```
cbb_new$POSTSEASON <- factor(cbb_new$POSTSEASON,
                             levels = c("R68", "R64", "R32", "S16", "E8", "F4", "2ND", "Champions"))
cbb_new$POSTSEASON_NUM <- as.numeric(cbb_new$POSTSEASON)

# Plot: Distribution of ADJDE bins across POSTSEASON
ggplot(data = subset(cbb_new, !is.na(POSTSEASON))) +
  geom_bar(mapping = aes(x = POSTSEASON, fill = adjde_bins)) +
  scale_fill_brewer(palette = "Blues", name = "ADJDE Legend") +
  labs(x = "Postseason Stage", y = "Count", title = "Distribution of ADJDE Bins by Postseason Stage")
```

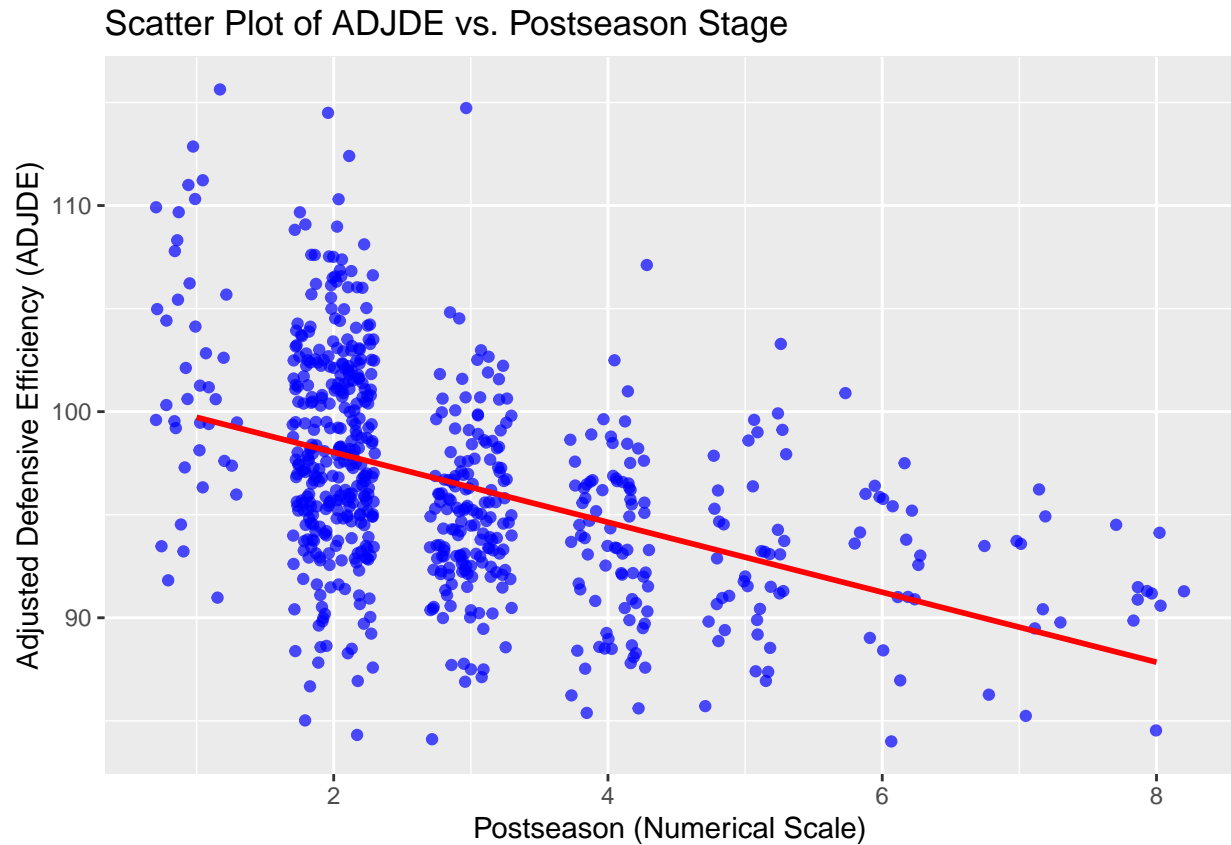


```
# Use Kendall's Tau for correlation (handles ties better)
correlation_result <- cor.test(cbb_new$ADJDE, cbb_new$POSTSEASON_NUM, method = "kendall")
print(correlation_result)
```

```
##
## Kendall's rank correlation tau
##
## data: cbb_new$ADJDE and cbb_new$POSTSEASON_NUM
## z = -12.975, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
## -0.3734283
```

```
# Scatter Plot for Visualizing Correlation
ggplot(cbb_new, aes(x = POSTSEASON_NUM, y = ADJDE)) +
  geom_jitter(width = 0.3, alpha = 0.7, color = "blue") +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(x = "Postseason (Numerical Scale)", y = "Adjusted Defensive Efficiency (ADJDE)",
       title = "Scatter Plot of ADJDE vs. Postseason Stage")
```

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



## Effective Opponent Field Goal Percentage

```
# Load necessary libraries
library(dplyr)
library(ggplot2)
library(RColorBrewer)

# Load the dataset
cbb <- read.csv("archive (1)/cbb.csv")

# Define success metric (e.g., reaching Final Four)
success_metric <- "Final Four"
```

```

# Data Analysis
# Compare defensive statistics between successful and unsuccessful teams
defensive_stats <- c("ADJDE", "efgd%", "2pd%", "3pd%", "ftrd", "drb")

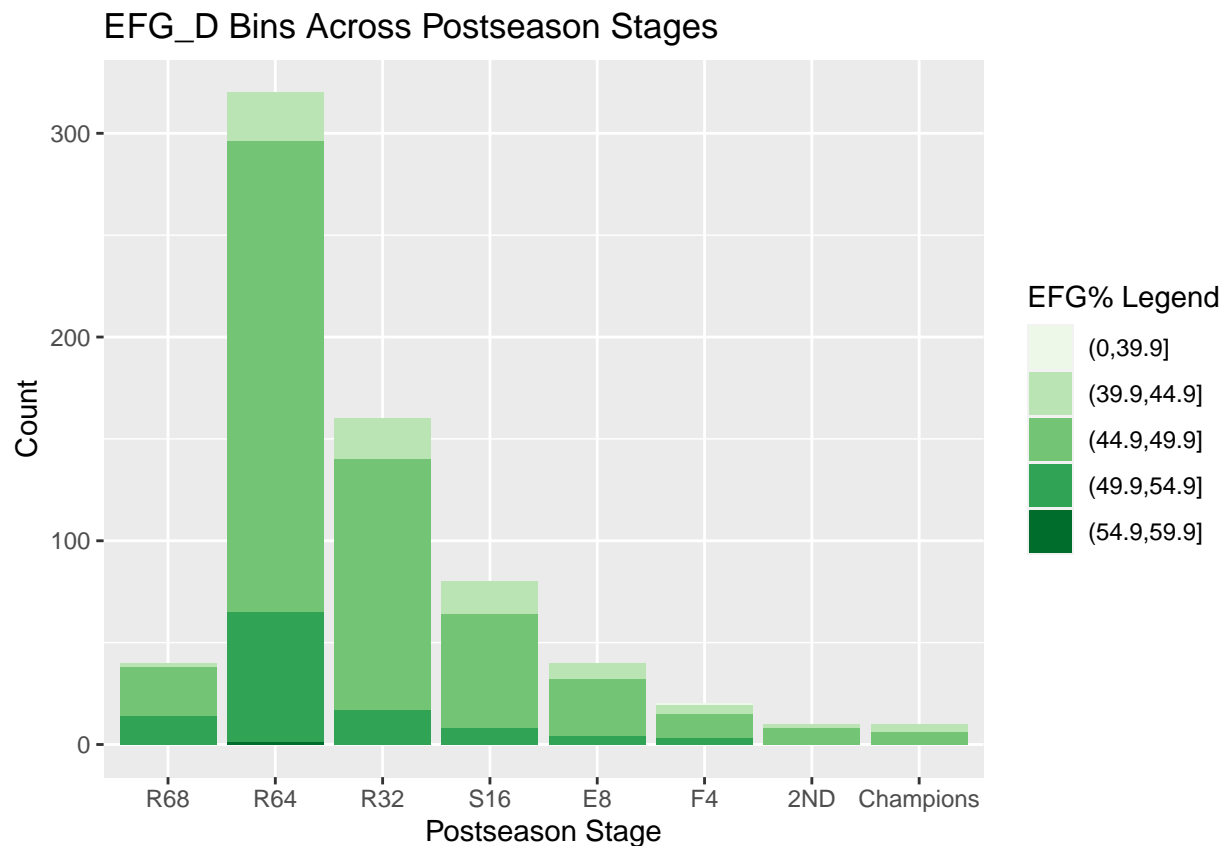
# Create Bins for Effective Opponent Field Goal Percentage
cbb_new <- cbb %>%
  mutate(efgd_bins = cut(EFG_D, breaks = c(0, 39.9, 44.9, 49.9, 54.9, 59.9, 100))) %>%
  filter(POSTSEASON != "N/A")

# Convert POSTSEASON into a factor with specific levels
cbb_new$POSTSEASON <- factor(cbb_new$POSTSEASON,
                             levels = c("R68", "R64", "R32", "S16", "E8", "F4", "2ND", "Champions"))

# Add a numerical version of POSTSEASON for correlation
cbb_new$POSTSEASON_NUM <- as.numeric(cbb_new$POSTSEASON)

# Plot Stacked Bar Plot for Effective Opponent Field Goal Percentage
ggplot(data = subset(cbb_new, !is.na(POSTSEASON))) +
  geom_bar(mapping = aes(x = POSTSEASON, fill = efgd_bins)) +
  scale_fill_brewer(palette = "Greens", name = "EFG% Legend") +
  labs(x = "Postseason Stage", y = "Count", title = "EFG_D Bins Across Postseason Stages")

```



```

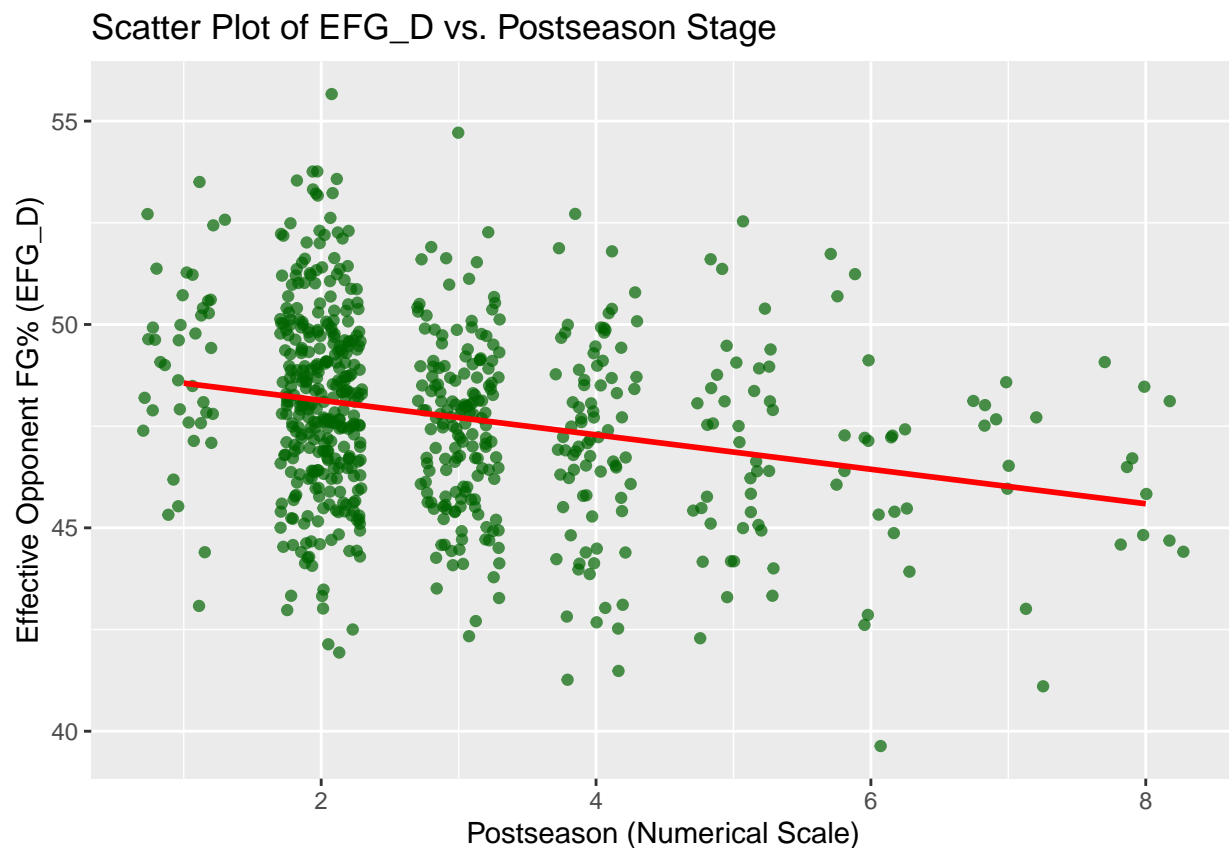
# Use Kendall's Tau for correlation (handles ties better)
correlation_result <- cor.test(cbb_new$EFG_D, cbb_new$POSTSEASON_NUM, method = "kendall")
print(correlation_result)

```

```
##
## Kendall's rank correlation tau
##
## data: cbb_new$EFG_D and cbb_new$POSTSEASON_NUM
## z = -6.5111, p-value = 7.461e-11
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1879981
```

```
# Scatter Plot for Visualizing Correlation
ggplot(cbb_new, aes(x = POSTSEASON_NUM, y = EFG_D)) +
  geom_jitter(width = 0.3, alpha = 0.7, color = "darkgreen") +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(x = "Postseason (Numerical Scale)", y = "Effective Opponent FG% (EFG_D)",
       title = "Scatter Plot of EFG_D vs. Postseason Stage")
```

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



Defensive Rebounds

```

# Load necessary libraries
library(dplyr)
library(ggplot2)
library(RColorBrewer)

# Load the dataset
cbb <- read.csv("archive (1)/cbb.csv")

# Define success metric (e.g., reaching Final Four)
success_metric <- "Final Four"

# Data Analysis
# Compare defensive statistics between successful and unsuccessful teams
defensive_stats <- c("ADJDE", "efgd%", "2pd%", "3pd%", "ftrd", "drb")

# Create Bins for Defensive Rebounds (DRB)
cbb_new <- cbb %>%
  mutate(drb_bins = cut(DRB, breaks = c(0, 14.9, 19.9, 24.9, 29.9, 34.9, 39.9, 44.9))) %>%
  filter(POSTSEASON != "N/A")

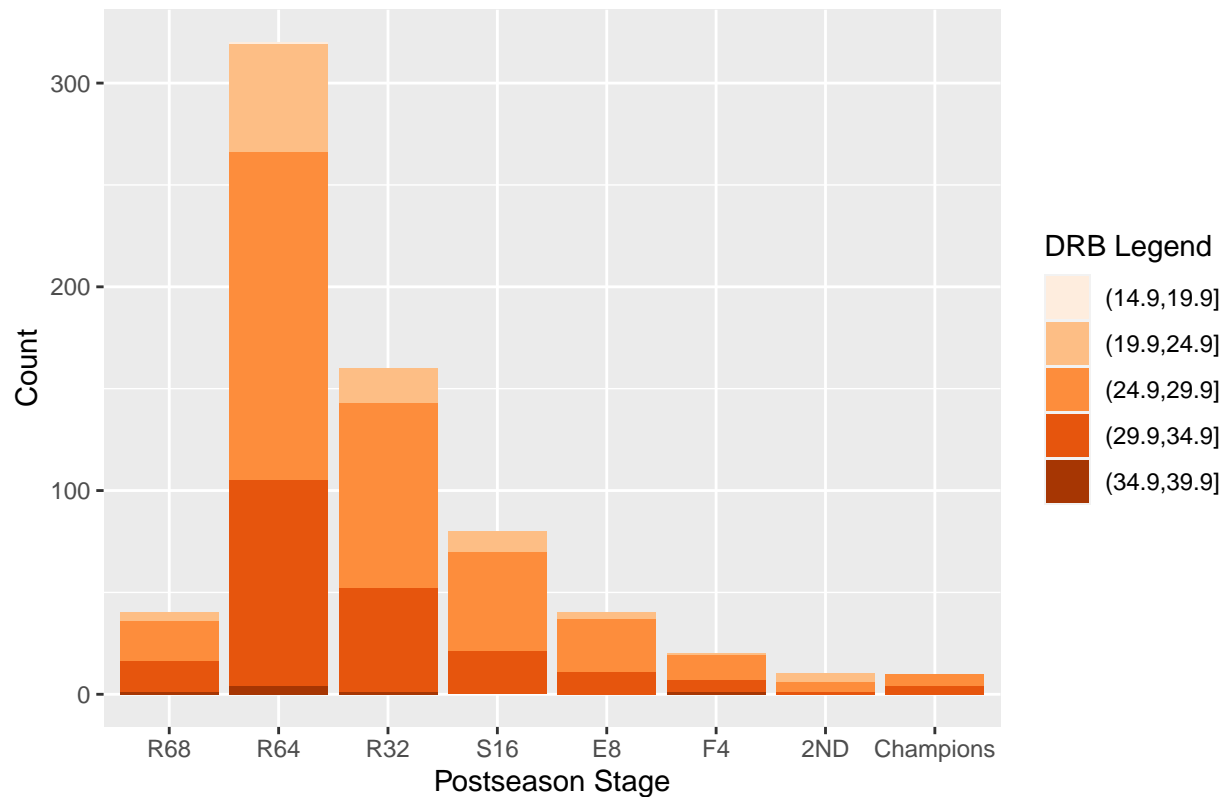
# Convert POSTSEASON into a factor with specific levels
cbb_new$POSTSEASON <- factor(cbb_new$POSTSEASON,
                             levels = c("R68", "R64", "R32", "S16", "E8", "F4", "2ND", "Champions"))

# Add a numerical version of POSTSEASON for correlation
cbb_new$POSTSEASON_NUM <- as.numeric(cbb_new$POSTSEASON)

# Plot Stacked Bar Plot for Defensive Rebounds
ggplot(data = subset(cbb_new, !is.na(POSTSEASON))) +
  geom_bar(mapping = aes(x = POSTSEASON, fill = drb_bins)) +
  scale_fill_brewer(palette = "Oranges", name = "DRB Legend") +
  labs(x = "Postseason Stage", y = "Count", title = "Defensive Rebounds (DRB) Across Postseason Stages")

```

## Defensive Rebounds (DRB) Across Postseason Stages



```
# Use Kendall's Tau for correlation (handles ties better)
correlation_result <- cor.test(cbb_new$DRB, cbb_new$POSTSEASON_NUM, method = "kendall")
print(correlation_result)
```

```
##
## Kendall's rank correlation tau
##
## data: cbb_new$DRB and cbb_new$POSTSEASON_NUM
## z = -0.82447, p-value = 0.4097
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
## tau
## -0.02376899
```

```
# Scatter Plot for Visualizing Correlation
ggplot(cbb_new, aes(x = POSTSEASON_NUM, y = DRB)) +
  geom_jitter(width = 0.3, alpha = 0.7, color = "darkorange") +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(x = "Postseason (Numerical Scale)", y = "Defensive Rebounds (DRB)",
       title = "Scatter Plot of DRB vs. Postseason Stage")
```

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

Scatter Plot of DRB vs. Postseason Stage

