STAT 345 Midterm Project

Due April 4

Step #1

library(nbastatR)  
library(dplyr)

##   
## Attaching package: 'dplyr'

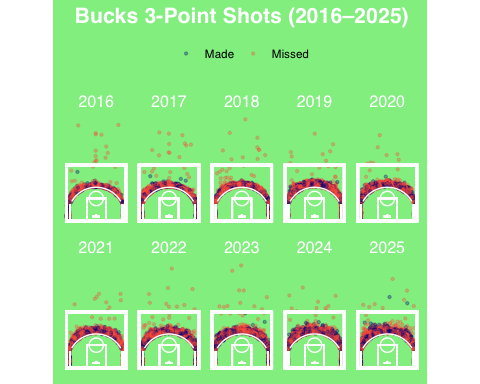
## The following objects are masked from 'package:stats':  
##   
## filter, lag

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

library(ggplot2)  
  
# Pull and combine all Bucks seasons from 2016–2025  
seasons <- 2016:2025  
bucks\_shots\_all <- lapply(seasons, function(yr) {  
 df <- teams\_shots(teams = "Milwaukee Bucks", seasons = yr)  
 df$season <- yr  
 return(df)  
}) %>% bind\_rows()

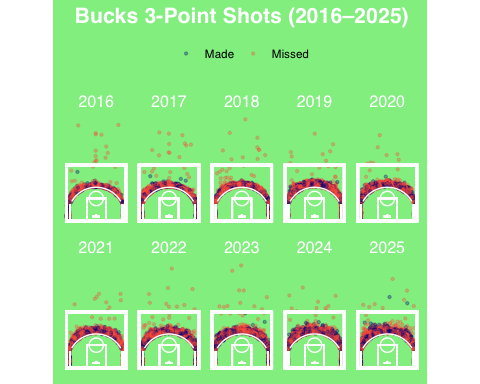
## Milwaukee Bucks 2015-16 shot data  
## Milwaukee Bucks 2016-17 shot data  
## Milwaukee Bucks 2017-18 shot data  
## Milwaukee Bucks 2018-19 shot data  
## Milwaukee Bucks 2019-20 shot data  
## Milwaukee Bucks 2020-21 shot data  
## Milwaukee Bucks 2021-22 shot data  
## Milwaukee Bucks 2022-23 shot data  
## Milwaukee Bucks 2023-24 shot data  
## Milwaukee Bucks 2024-25 shot data

# Filter to 3PT shots only  
three\_point\_shots <- bucks\_shots\_all %>%  
 filter(!is.na(locationX) & !is.na(locationY)) %>%  
 filter(typeShot == "3PT Field Goal") %>%  
 mutate(  
 shot\_result = ifelse(isShotMade, "Made", "Missed"),  
 season = as.factor(season)  
 )  
  
# Circle helper using ChatGPT  
circle\_points <- function(center = c(0, 0), radius = 1, npoints = 100, start = 0, end = 2\*pi) {  
 theta <- seq(start, end, length.out = npoints)  
 data.frame(  
 x = center[1] + radius \* cos(theta),  
 y = center[2] + radius \* sin(theta)  
 )  
}  
  
# Nice half-court function using ChatGPT  
draw\_pretty\_half\_court <- function() {  
 list(  
 geom\_rect(aes(xmin = -250, xmax = 250, ymin = -47.5, ymax = 422.5),  
 fill = NA, color = "white", inherit.aes = FALSE, linewidth = 1),  
 geom\_rect(aes(xmin = -80, xmax = 80, ymin = -47.5, ymax = 143.5),  
 fill = NA, color = "white", inherit.aes = FALSE),  
 geom\_path(data = circle\_points(c(0, 143.5), 60),  
 aes(x = x, y = y), color = "white", inherit.aes = FALSE),  
 geom\_path(data = circle\_points(c(0, 0), 7.5),  
 aes(x = x, y = y), color = "white", inherit.aes = FALSE),  
 geom\_segment(aes(x = -30, xend = 30, y = -7.5, yend = -7.5),  
 color = "white", inherit.aes = FALSE),  
 geom\_path(data = circle\_points(c(0, 0), 237.5, start = pi/6, end = 5\*pi/6),  
 aes(x = x, y = y), color = "white", inherit.aes = FALSE),  
 geom\_segment(aes(x = -220, xend = -220, y = -47.5, yend = 92.5),  
 color = "white", inherit.aes = FALSE),  
 geom\_segment(aes(x = 220, xend = 220, y = -47.5, yend = 92.5),  
 color = "white", inherit.aes = FALSE)  
 )  
}  
  
# Plot all seasons using facets  
ggplot(three\_point\_shots, aes(x = locationX, y = locationY, color = shot\_result)) +  
 geom\_point(alpha = 0.3, size = 0.8) +  
 scale\_color\_manual(values = c("Made" = "darkblue", "Missed" = "tomato")) +  
 draw\_pretty\_half\_court() +  
 coord\_fixed() +  
 facet\_wrap(~ season, ncol = 5) + # wider layout  
 theme\_minimal(base\_family = "Helvetica") +  
 theme(  
 panel.background = element\_rect(fill = "lightgreen", color = NA),  
 plot.background = element\_rect(fill = "lightgreen", color = NA),  
 panel.grid = element\_blank(),  
 legend.position = "top",  
 legend.title = element\_blank(),  
 plot.title = element\_text(color = "white", size = 16, face = "bold", hjust = 0.5),  
 strip.text = element\_text(color = "white", size = 12),  
 axis.text = element\_blank(),  
 axis.title = element\_blank(),  
 axis.ticks = element\_blank()  
 ) +  
 labs(  
 title = "Bucks 3-Point Shots (2016–2025)"  
 )



Step #2

## Milwaukee Bucks 2015-16 shot data  
## Milwaukee Bucks 2016-17 shot data  
## Milwaukee Bucks 2017-18 shot data  
## Milwaukee Bucks 2018-19 shot data  
## Milwaukee Bucks 2019-20 shot data  
## Milwaukee Bucks 2020-21 shot data  
## Milwaukee Bucks 2021-22 shot data  
## Milwaukee Bucks 2022-23 shot data  
## Milwaukee Bucks 2023-24 shot data  
## Milwaukee Bucks 2024-25 shot data

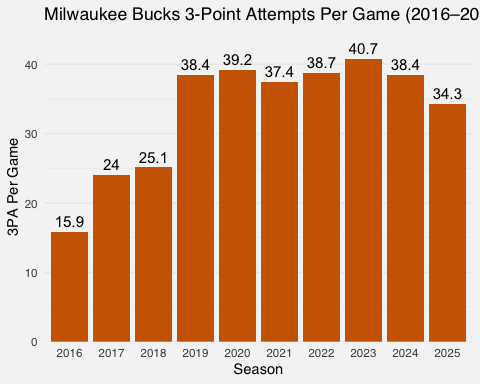


* For this project, I wanted to focus on three pointers from the Milwaukee Bucks over the past 10 seasons.
* With the red dots for misses and blue for makes, you can see each year how the team did shooting the 3 ball.
* 2025 seems like the year with the most makes, followed by 2021.
* The only year the bucks have made shots from behind half-court was 2025, where they made two.
* 2024 has a wide range of three pointers taken compared to the other years

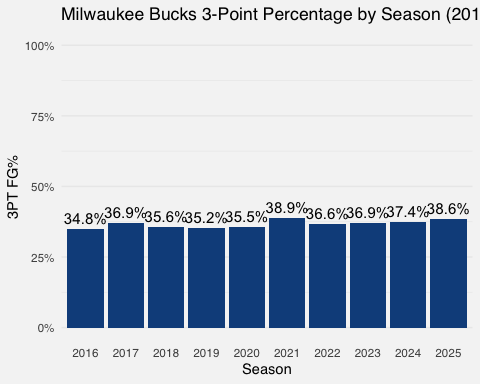
Step #4

One of the greatest 3 point shooters of all time, Damian Lillard, joined the Milwaukee Bucks in 2023. With the addition of a hall of fame caliber player known for shootings threes, how did this affect the Bucks? Did the addition of Lillard increase the number of 3’s shot per game by the bucks? Did the Bucks 3 point percentage go up after the addition of Lillard? I plan to find out with new plots and use of statistics.

# Create a lookup for number of games per season (adjusted for COVID seasons)  
games\_per\_season <- data.frame(  
 season = 2016:2025,  
 games\_played = c(82, 82, 82, 82, 73, 72, 82, 82, 82, 82) # 2020=73, 2021=72  
)  
  
# Summarize 3PA  
three\_pt\_attempts\_per\_game <- bucks\_shots\_all %>%  
 filter(!is.na(locationX) & !is.na(locationY)) %>%  
 filter(distanceShot >= 22) %>%  
 group\_by(season) %>%  
 summarise(total\_3pa = n(), .groups = "drop") %>%  
 left\_join(games\_per\_season, by = "season") %>%  
 mutate(three\_pa\_per\_game = total\_3pa / games\_played)  
  
# Plot 3PA per game  
ggplot(three\_pt\_attempts\_per\_game, aes(x = factor(season), y = three\_pa\_per\_game)) +  
 geom\_col(fill = "darkorange3") +  
 geom\_text(aes(label = round(three\_pa\_per\_game, 1)),  
 vjust = -0.5, color = "black", size = 4) +  
 scale\_y\_continuous(expand = expansion(mult = c(0, 0.1))) +  
 theme\_minimal(base\_family = "Helvetica") +  
 labs(  
 title = "Milwaukee Bucks 3-Point Attempts Per Game (2016–2025)",  
 x = "Season",  
 y = "3PA Per Game"  
 ) +  
 theme(  
 plot.background = element\_rect(fill = "#F5F5F5", color = NA),  
 panel.grid.major.x = element\_blank()  
 )

 This graphic shows an incline of 3 point attempts per game until the year 2023, when it declines. So with Lillard on the team, the Bucks did not shoot more 3’s per game.

library(dplyr)  
library(ggplot2)  
  
# Filter for 3-point shots only (assumes `bucks\_shots\_all` has all seasons)  
three\_pt\_summary <- bucks\_shots\_all %>%  
 filter(!is.na(locationX) & !is.na(locationY)) %>%  
 filter(distanceShot >= 22) %>%  
 group\_by(season) %>%  
 summarise(  
 threes\_made = sum(isShotMade, na.rm = TRUE),  
 threes\_attempted = n(),  
 three\_pt\_pct = threes\_made / threes\_attempted  
 ) %>%  
 ungroup()  
  
# Plot 3PT% by season  
ggplot(three\_pt\_summary, aes(x = factor(season), y = three\_pt\_pct)) +  
 geom\_col(fill = "dodgerblue4") +  
 geom\_text(aes(label = scales::percent(three\_pt\_pct, accuracy = 0.1)),  
 vjust = -0.5, color = "black", size = 4) +  
 scale\_y\_continuous(labels = scales::percent\_format(accuracy = 1), limits = c(0, 1)) +  
 theme\_minimal(base\_family = "Helvetica") +  
 labs(  
 title = "Milwaukee Bucks 3-Point Percentage by Season (2016–2025)",  
 x = "Season",  
 y = "3PT FG%"  
 ) +  
 theme(  
 plot.background = element\_rect(fill = "#F5F5F5", color = NA),  
 panel.grid.major.x = element\_blank()  
 )

 This graph shows the 3 point percentage by season. Since this does not clearly answer my question regarding Lillard and 3 point percentage, I will run a hypothesis test at the 95% confidence level to find out if the 3 point percentage changed with the addition of Lillard.

# Group the data  
before\_group <- filter(three\_pt\_summary, season %in% 2016:2023)  
after\_group <- filter(three\_pt\_summary, season %in% 2024:2025)  
  
# Combine shots made and attempted  
before\_made <- sum(before\_group$threes\_made)  
before\_attempts <- sum(before\_group$threes\_attempted)  
  
after\_made <- sum(after\_group$threes\_made)  
after\_attempts <- sum(after\_group$threes\_attempted)  
  
# Run a one-sided 2-sample proportion test (z-test)  
test\_result <- prop.test(  
 x = c(before\_made, after\_made),  
 n = c(before\_attempts, after\_attempts),  
 alternative = "less", # H₁: after > before  
 correct = FALSE  
)  
  
# Print results  
cat("P-value:", round(test\_result$p.value, 5), "\n")

## P-value: 0.01428

if (test\_result$p.value < 0.05) {  
 cat("✅ Conclusion: 3PT% is significantly higher in 2024–2025 compared to 2016–2023.\n")  
} else {  
 cat("❌ Conclusion: No significant increase in 3PT% in 2024–2025 at the 95% confidence level.\n")  
}

## ✅ Conclusion: 3PT% is significantly higher in 2024–2025 compared to 2016–2023.

With a p-value less than 0.05, we have significant evidence to prove the 3 point percentage is higher after 2023 than before. While it may not be directly attributed to the addition of Lillard, adding one of the greatest shooters of all time likely has something to do with it.