

Jordan V. Lebron

Xiang Jiang and Albion Shoshi

April 5, 2024

```
# Import libraries and data here
```

```
library(dplyr)
library(readr)
library(tidyverse)
library(rvest)
library(ggplot2)
library(tidyr)
library(esquisse)
```

```
# Downloaded Lebron Data from Basketball Reference
```

```
# Link: https://www.basketball-reference.com/players/j/jamesle01.html#per_game
```

```
LebronAdjustedShooting <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronAdjustedShooting.csv')
LebronAdvanced <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronAdvanced.csv')
LebronAllStarGames <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronAllStarGames.csv')
LebronGameHighs <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronGameHighs.csv')
LebronPer100Poss <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronPer100Poss.csv')
LebronPer36Min <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronPer36Min.csv')
LebronPerGame <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronPerGame.csv')
LebronPlayByPlay <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronPlayByPlay.csv')
LebronPlayoffsSeries <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronPlayoffsSeries.csv')
LebronShooting <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronShooting.csv')
LebronTotals <- read_csv('/Users/xiangjiang/Desktop/Lebron/LebronTotals.csv')
```

```
# Downloaded Lebron Data from Basketball Reference
```

```
# Link: https://www.basketball-reference.com/players/j/jordami01.html
```

```
JordanAdjustedShooting <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanAdjustedShooting.csv')
JordanAdvanced <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanAdvanced.csv')
JordanAllStarGames <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanAllStarGames.csv')
JordanGameHighs <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanGameHighs.csv')
JordanPer100Poss <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanPer100Poss.csv')
JordanPer36Min <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanPer36Min.csv')
JordanPerGame <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanPerGame.csv')
JordanPlayByPlay <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanPlayByPlay.csv')
JordanPlayOffsSeries <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanPlayOffsSeries.csv')
JordanShooting <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanShooting.csv')
JordanTotals <- read_csv('/Users/xiangjiang/Desktop/Jordan/JordanTotals.csv')
```

```
# Note: Not every file is needed to be used.
```

Guiding Question:

- Who is the Greatest Basketball Player of All Time in NBA History, Lebron or Jordan?

Become acquainted with your data sources:

Where did you find them?

- Data retrieved from Basketball Reference.

Who collected/maintains them?

- Sean Lahman donated much of the initial data.
- Chip Hart and Kevin Cohen provided input on the site's design and data collection, respectively.
- Tom Timmerman, Dick Pfander, Todd Spehr, Matt Shuh, Justin Kubatko, Sean Burrill, Tariq Jabbar, Mike Lynch, Michael Hamel, Mark Montieth, and Robert Bradley have all contributed significantly to the collection of NBA and ABA data.
- Matthew Maurer, an NBA Draft historian, has supplied information on the NBA Draft.

When & Why were they originally collected?

Basketball Reference was launched in April of 2004 by Justin Kubatko. The primary goal was to create a comprehensive database that tracks basketball information. This encompasses player statistics, team performances, game results, and much more, making it an invaluable resource for fans, analysts, and researchers interested in the sport's history and current events.

What does a case represent in each data source, and how many total cases are available?

- For player statistics, each row or case represents a season's worth of performance data for that player.
- Game logs would treat each game as a case.
- Draft data entries represent individual players selected in the NBA draft.
- There are thousands of cases covering decades of basketball history across the NBA and ABA.

What are some of the variables that you plan to use?

- Points per game (PointsPerGame)
- Assists per game (AssistsPerGame)
- Rebounds per game (ReboundsPerGame)
- Total Points (TotalPoints)
- Total Assists (TotalAssists)
- Total Rebounds (TotalRebounds)
- Player efficiency rating (PlayerEfficiencyRating)
- Career achievements (CareerAchievements)
- All-star game appearances (AllStarGameAppearances)
- Game scores (GameScores)

Explore intuition related to the research question

Create some informative plots and summary statistics

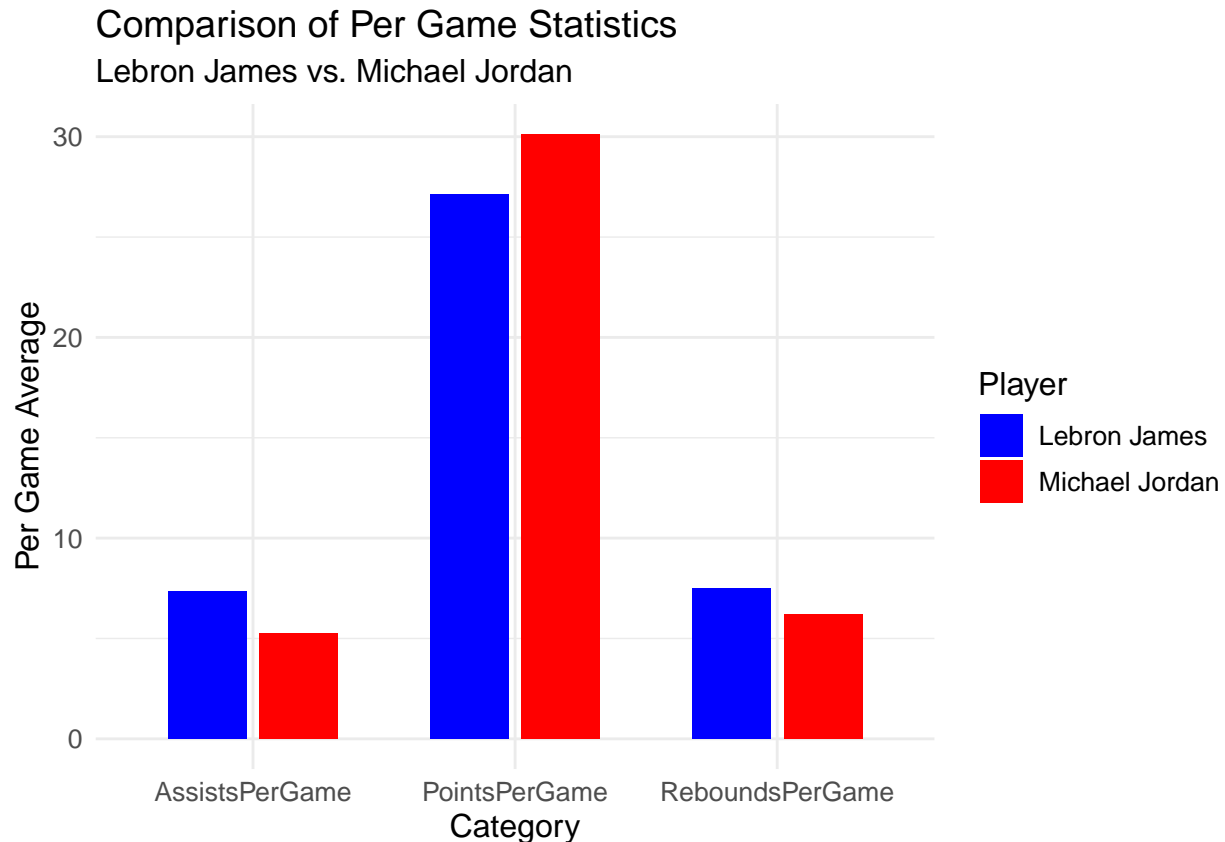
```
LebronStatsPerGame <- LebronTotals %>%
  summarise(Player = "Lebron James", AssistsPerGame = sum(AST, na.rm = TRUE) / sum(G, na.rm = TRUE), ReboundsPerGame = sum(
    REB, na.rm = TRUE) / sum(G, na.rm = TRUE))

JordanStatsPerGame <- JordanTotals %>%
  summarise(Player = "Michael Jordan", AssistsPerGame = sum(AST, na.rm = TRUE) / sum(G, na.rm = TRUE), ReboundsPerGame = sum(
    REB, na.rm = TRUE) / sum(G, na.rm = TRUE))

StatsComparison <- bind_rows(LebronStatsPerGame, JordanStatsPerGame)
```

```
StatsComparisonLong <- StatsComparison %>%
  pivot_longer(cols = c(AssistsPerGame, ReboundsPerGame, PointsPerGame), names_to = "Statistic", values_to = "Value")

ggplot(StatsComparisonLong, aes(x = Statistic, y = Value, fill = Player)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.7), width = 0.6) +
  theme_minimal() +
  labs(title = "Comparison of Per Game Statistics", subtitle = "Lebron James vs. Michael Jordan", x = "Category", y = "Per Game Average") +
  scale_fill_manual(values = c("Lebron James" = "blue", "Michael Jordan" = "red")) +
  theme(text = element_text(size = 12))
```



```
print(StatsComparison)
```

```
## # A tibble: 2 x 4
##   Player      AssistsPerGame ReboundsPerGame PointsPerGame
##   <chr>          <dbl>          <dbl>          <dbl>
## 1 Lebron James      7.37            7.49           27.1
## 2 Michael Jordan    5.25            6.22           30.1
```

```
LebronTotalsStats <- LebronTotals %>%
  summarise(Player = "Lebron James", TotalRebounds = sum(TRB, na.rm = TRUE), TotalAssists = sum(AST, na.rm = TRUE), TotalPoints = sum(P, na.rm = TRUE))

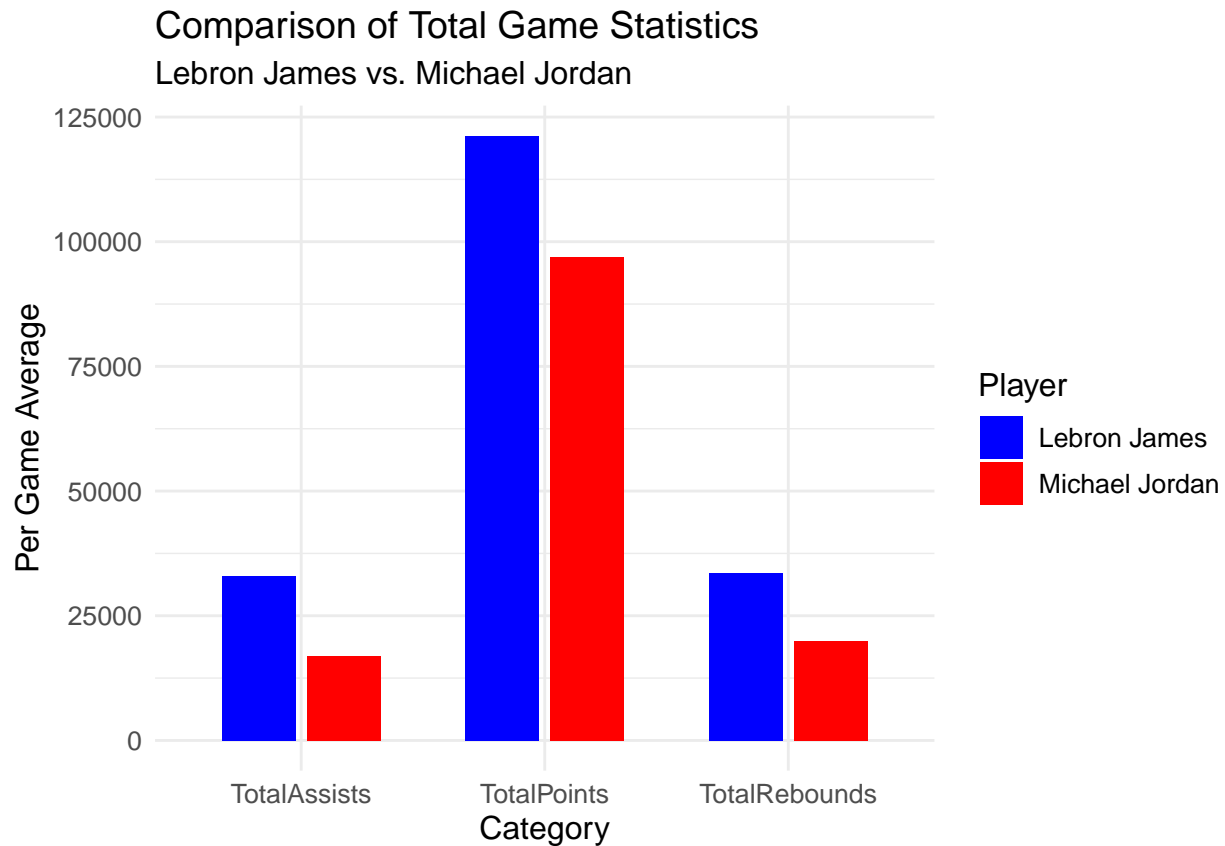
JordanTotalsStats <- JordanTotals %>%
  summarise(Player = "Michael Jordan", TotalRebounds = sum(TRB, na.rm = TRUE), TotalAssists = sum(AST, na.rm = TRUE), TotalPoints = sum(P, na.rm = TRUE))

StatsComparison <- bind_rows(LebronTotalsStats, JordanTotalsStats)

StatsComparisonLong <- StatsComparison %>%
  pivot_longer(cols = c(TotalRebounds, TotalAssists, TotalPoints), names_to = "Statistic", values_to = "Value")

ggplot(StatsComparisonLong, aes(x = Statistic, y = Value, fill = Player)) +
```

```
geom_bar(stat = "identity", position = position_dodge(width = 0.7), width = 0.6) +
theme_minimal() +
labs(title = "Comparison of Total Game Statistics", subtitle = "Lebron James vs. Michael Jordan", x = "Category") +
scale_fill_manual(values = c("Lebron James" = "blue", "Michael Jordan" = "red")) +
theme(text = element_text(size = 12))
```



```
print(StatsComparison)
```

```
## # A tibble: 2 x 4
##   Player      TotalRebounds TotalAssists TotalPoints
##   <chr>          <dbl>         <dbl>         <dbl>
## 1 Lebron James      33495          32961          121227
## 2 Michael Jordan    20016          16899           96876
```

```
PER <- function(x){
  x %>%
    mutate(uPER = (PTS + TRB + AST + STL + BLK - ((FGA - FG) + (FTA - FT) + TOV)) / G) %>%
    summarise(AveragePER = mean(uPER, na.rm = TRUE))
}
```

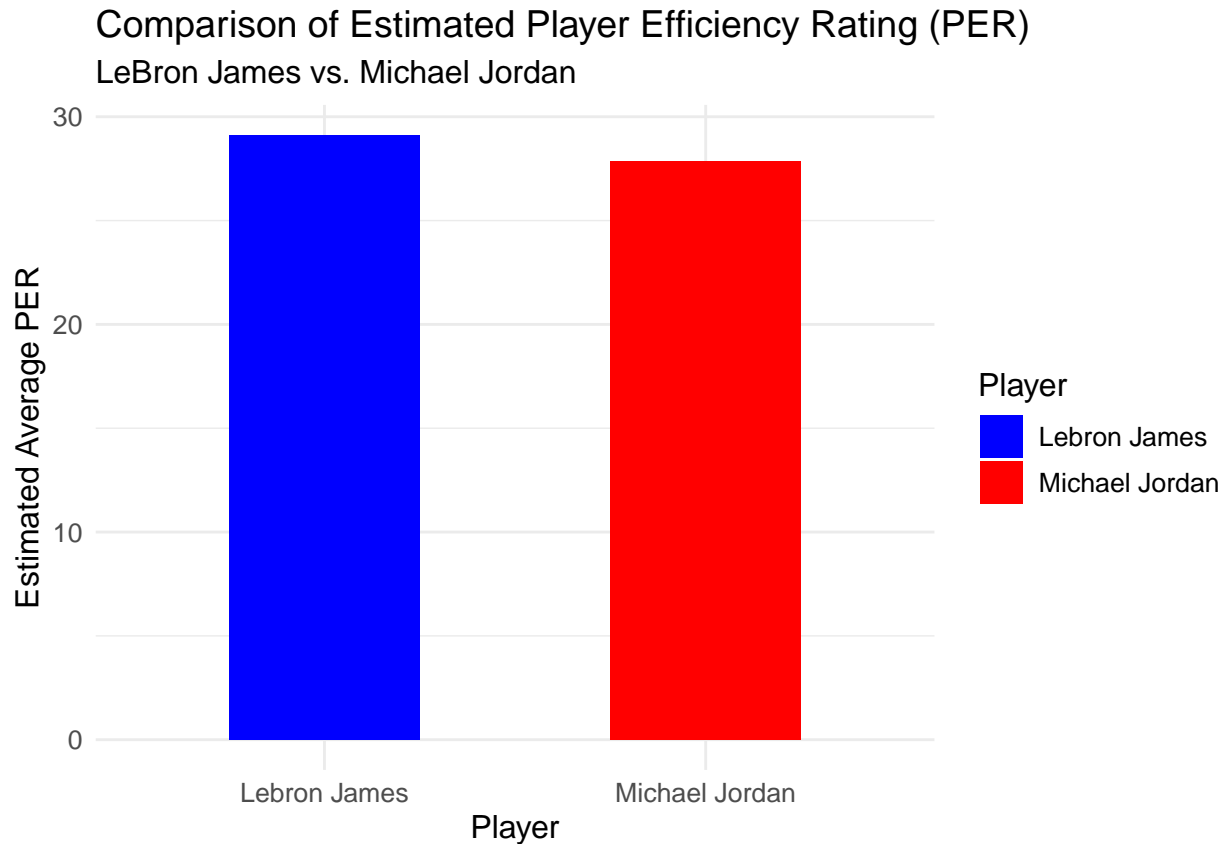
```
LebronPER <- PER(LebronTotals) %>%
  mutate(Player = "Lebron James")
```

```
JordanPER <- PER(JordanTotals) %>%
  mutate(Player = "Michael Jordan")
```

```
PERComparison <- bind_rows(LebronPER, JordanPER)
```

```
ggplot(PERComparison, aes(x = Player, y = AveragePER, fill = Player)) +
  geom_bar(stat = "identity", position = position_dodge(), width = 0.5) +
```

```
theme_minimal() +
labs(title = "Comparison of Estimated Player Efficiency Rating (PER)", subtitle = "LeBron James vs. Michael Jordan",
scale_fill_manual(values = c("LeBron James" = "blue", "Michael Jordan" = "red"))) +
theme(text = element_text(size = 12))
```



```
LebronAllStarAppearances <- nrow(LebronAllStarGames)
JordanAllStarAppearances <- nrow(JordanAllStarGames)

print(paste("Lebron James All-Star Game Appearances:", LebronAllStarAppearances))
```

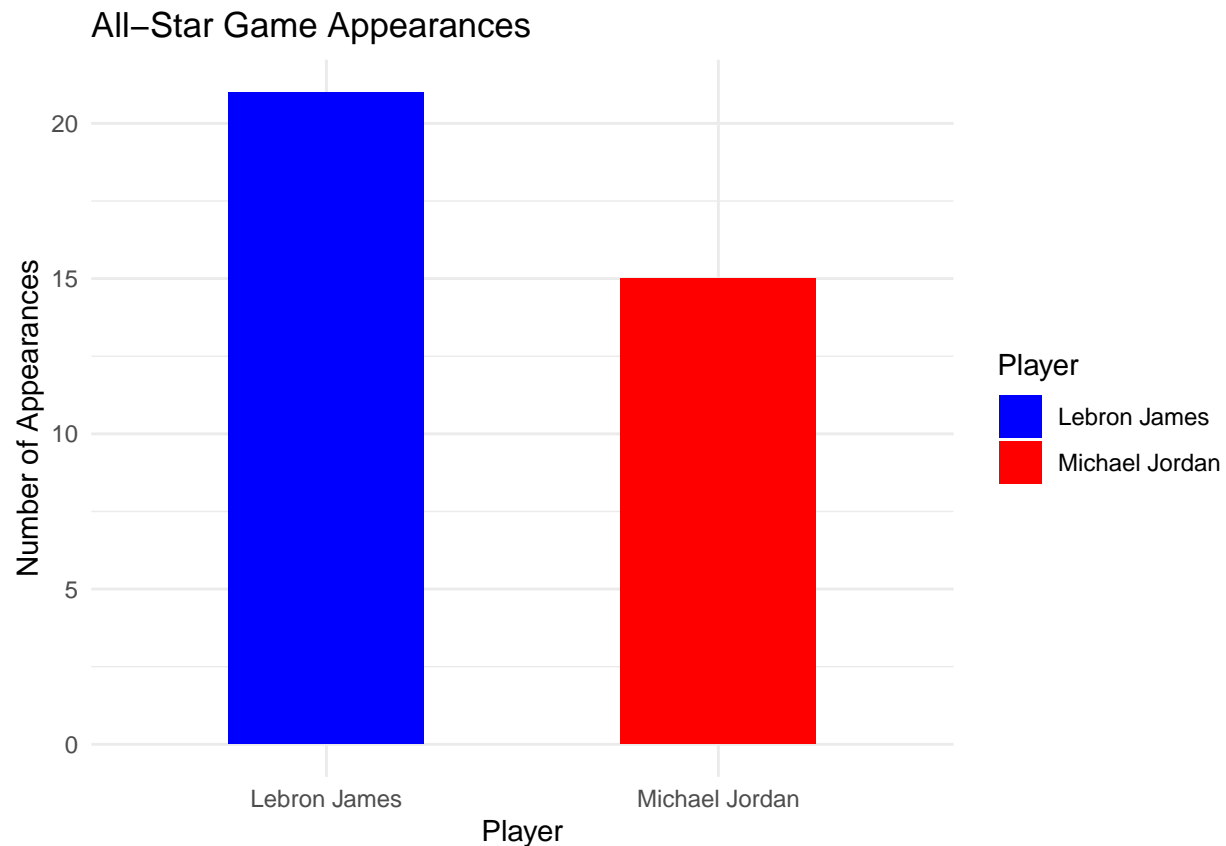
```
## [1] "Lebron James All-Star Game Appearances: 21"
```

```
print(paste("Michael Jordan All-Star Game Appearances:", JordanAllStarAppearances))
```

```
## [1] "Michael Jordan All-Star Game Appearances: 15"
```

```
AllStarData <- data.frame(Player = c("Lebron James", "Michael Jordan"), Appearances = c(LebronAllStarAppearances, JordanAllStarAppearances))

ggplot(AllStarData, aes(x = Player, y = Appearances, fill = Player)) +
  geom_bar(stat = "identity", position = position_dodge(), width = 0.5) +
  theme_minimal() +
  labs(title = "All-Star Game Appearances", y = "Number of Appearances", fill = "Player") +
  scale_fill_manual(values = c("Lebron James" = "blue", "Michael Jordan" = "red"))
```



```

LebronCareerTotals <- LebronTotals %>%
  summarise(TotalPoints = sum(PTS, na.rm = TRUE), TotalAssists = sum(AST, na.rm = TRUE), TotalRebounds = sum(T
JordanCareerTotals <- JordanTotals %>%
  summarise(TotalPoints = sum(PTS, na.rm = TRUE), TotalAssists = sum(AST, na.rm = TRUE), TotalRebounds = sum(T
MileStones <- data.frame(TotalPoints = 30000, TotalAssists = 5000, TotalRebounds = 5000, TotalSteals = 1000, T
Achievements <- function(totals, MileStones) {
  as.data.frame(lapply(totals, function(x) x >= MileStones))
}

LebronAchievements <- Achievements(LebronCareerTotals, MileStones)
JordanAchievements <- Achievements(JordanCareerTotals, MileStones)

LebronAchievements$Player <- "Lebron James"
JordanAchievements$Player <- "Michael Jordan"

CareerAchievements <- rbind(LebronAchievements, JordanAchievements)

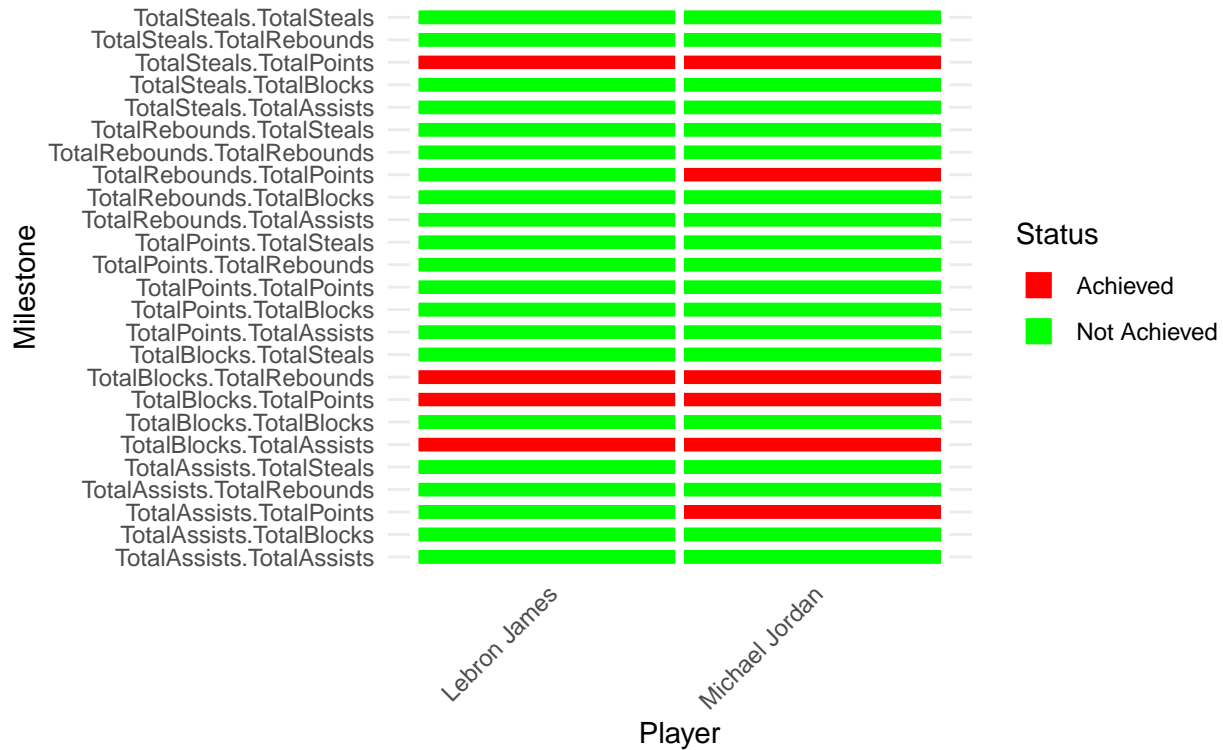
CareerAchievementsLong <- CareerAchievements %>%
  pivot_longer(cols = -Player, names_to = "Milestone", values_to = "Achieved") %>%
  mutate(Achieved = as.factor(Achieved), Labels = ifelse(Achieved == "TRUE", "Achieved", "Not Achieved"))

ggplot(CareerAchievementsLong, aes(x = Player, y = Milestone, fill = Achieved)) +
  geom_tile(color = "white", size = 1.5) +
  scale_fill_manual(values = c("TRUE" = "green", "FALSE" = "red"), labels = c("Achieved", "Not Achieved")) +
  labs(title = "Career Achievements of LeBron James and Michael Jordan",
       subtitle = "Green indicates the milestone was achieved", y = "Milestone", fill = "Status") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

Career Achievements of LeBron James and Michael Jordan

Green indicates the milestone was achieved



```
print(CareerAchievements)
```

```
## TotalPoints.TotalPoints TotalPoints.TotalAssists TotalPoints.TotalRebounds
## 1 TRUE TRUE TRUE
## 2 TRUE TRUE TRUE
## TotalPoints.TotalSteals TotalPoints.TotalBlocks TotalAssists.TotalPoints
## 1 TRUE TRUE TRUE
## 2 TRUE TRUE FALSE
## TotalAssists.TotalAssists TotalAssists.TotalRebounds TotalAssists.TotalSteals
## 1 TRUE TRUE TRUE
## 2 TRUE TRUE TRUE
## TotalAssists.TotalBlocks TotalRebounds.TotalPoints TotalRebounds.TotalAssists
## 1 TRUE TRUE TRUE
## 2 TRUE FALSE TRUE
## TotalRebounds.TotalRebounds TotalRebounds.TotalSteals
## 1 TRUE TRUE
## 2 TRUE TRUE
## TotalRebounds.TotalBlocks TotalSteals.TotalPoints TotalSteals.TotalAssists
## 1 TRUE FALSE TRUE
## 2 TRUE FALSE TRUE
## TotalSteals.TotalRebounds TotalSteals.TotalSteals TotalSteals.TotalBlocks
## 1 TRUE TRUE TRUE
## 2 TRUE TRUE TRUE
## TotalBlocks.TotalPoints TotalBlocks.TotalAssists TotalBlocks.TotalRebounds
## 1 FALSE FALSE FALSE
## 2 FALSE FALSE FALSE
## TotalBlocks.TotalSteals TotalBlocks.TotalBlocks Player
## 1 TRUE TRUE Lebron James
## 2 TRUE TRUE Michael Jordan
```

Describe preliminary observations and intuition about the research question