

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
```

```
LebronPerGame <- read_csv('/Users/xiangjiang/Desktop/LebronNew/LebronPerGameNew.csv')
LebronTotals <- read_csv('/Users/xiangjiang/Desktop/LebronNew/LebronTotalStatsNew.csv')
LebronAdvanced <- read_csv('/Users/xiangjiang/Desktop/LebronNew/LebronAdvancedNew.csv')
LebronPer100Poss <- read_csv('/Users/xiangjiang/Desktop/LebronNew/LebronPer100Poss.csv')
LebronAllStarGames <- read_csv('/Users/xiangjiang/Desktop/LebronNew/LebronAllStarGame.csv')
```

```
# Downloaded Jordan Data from Basketball Reference
```

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

```
JordanPerGame <- read_csv('/Users/xiangjiang/Desktop/JordanNew/JordanPerGameNew.csv')
JordanTotals <- read_csv('/Users/xiangjiang/Desktop/JordanNew/JordanTotalStatsNew.csv')
JordanAdvanced <- read_csv('/Users/xiangjiang/Desktop/JordanNew/JordanAdvancedNew.csv')
JordanPer100Poss <- read_csv('/Users/xiangjiang/Desktop/JordanNew/JordanPer100Poss.csv')
JordanAllStarGames <- read_csv('/Users/xiangjiang/Desktop/JordanNew/JordanAllStarGame.csv')
```

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

1. Comparing LeBron and Jordans total and per-game stats

Bargraph showcasing a comparison of per game-stats

```
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))

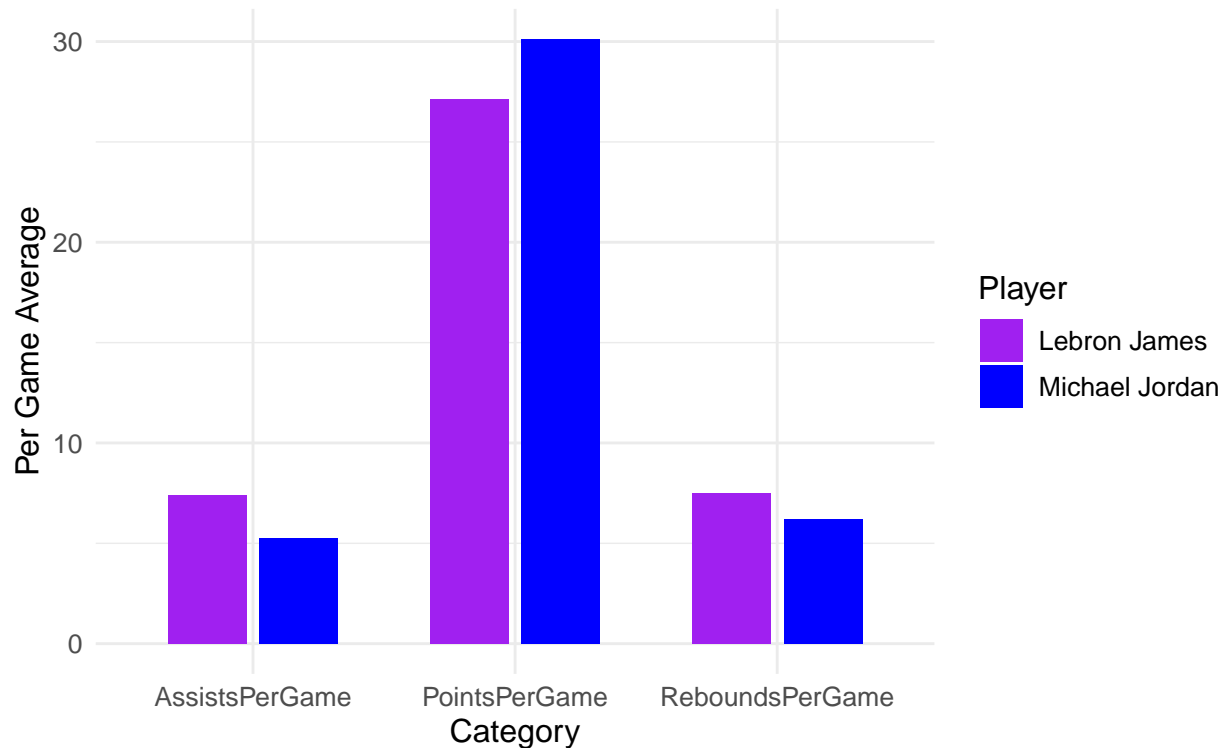
StatsComparison <- bind_rows(LebronStatsPerGame, JordanStatsPerGame)

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 = "Value") +
  scale_fill_manual(values = c("Lebron James" = "purple", "Michael Jordan" = "blue")) +
  theme(text = element_text(size = 12))
```

Comparison of Per Game Statistics

Lebron James vs. Michael Jordan



```
print(StatsComparison)
```

```
## # A tibble: 2 x 4
##   Player      AssistsPerGame ReboundsPerGame PointsPerGame
##   <chr>          <dbl>          <dbl>          <dbl>
## 1 Lebron James      7.38            7.50           27.1
## 2 Michael Jordan    5.25            6.22           30.1
```

Bargraph showcasing a comparison of total stats

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

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

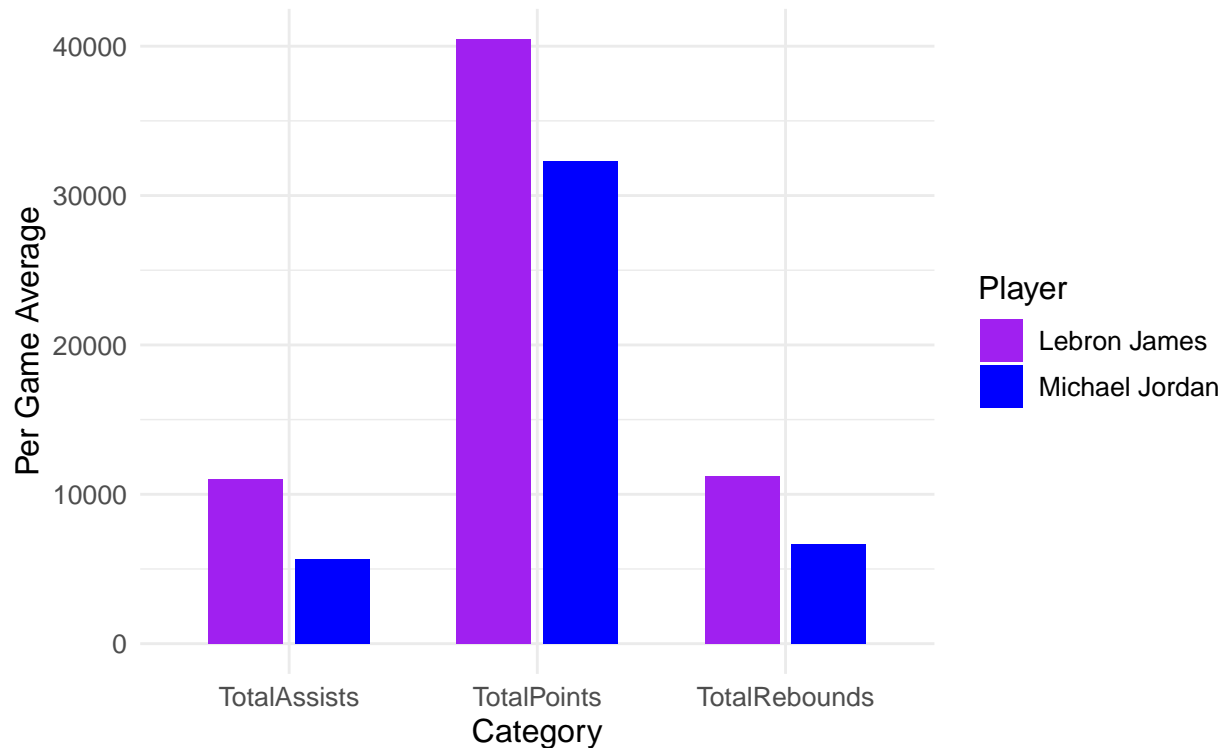
StatsComparison <- bind_rows(LebronTotalsStats, JordanTotalsStats)

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", y = "Total Value") +
  scale_fill_manual(values = c("Lebron James" = "purple", "Michael Jordan" = "blue")) +
  theme(text = element_text(size = 12))
```

Comparison of Total Game Statistics

Lebron James vs. Michael Jordan



```
print(StatsComparison)
```

```
## # A tibble: 2 x 4
##   Player      TotalRebounds TotalAssists TotalPoints
##   <chr>          <dbl>         <dbl>         <dbl>
## 1 Lebron James    11185         11009         40474
## 2 Michael Jordan   6672          5633         32292
```

This graph is showing the total career stats for assists, points, and rebounds for Jordan and Lebron's careers respectively. Not only does Lebron lead Jordan in every total career stat, but does at a significant rate. Lebron is currently the all time leading scorer accomplishing the feat earlier this year as well as in the top 4 for total all time assists only behind point guards who are known for passing.

2. Using Advanced and Per100Poss stats for comparison

Dotplot showcasing VORP (Value over replacement player)

```
LebronLongevity <- LebronAdvanced %>%
  mutate(PlayerName = "Lebron James", Games = G, TrueShootingPercentage = as.numeric("TS%"), WinShare = WS, ValueOverReplacementPlayer = VORP)
  select(PlayerName, Season, Games, TrueShootingPercentage, WinShare, ValueOverReplacementPlayer)

JordanLongevity <- JordanAdvanced %>%
  mutate(PlayerName = "Michael Jordan", Games = G, TrueShootingPercentage = as.numeric("TS%"), WinShare = WS, ValueOverReplacementPlayer = VORP)
  select(PlayerName, Season, Games, TrueShootingPercentage, WinShare, ValueOverReplacementPlayer)

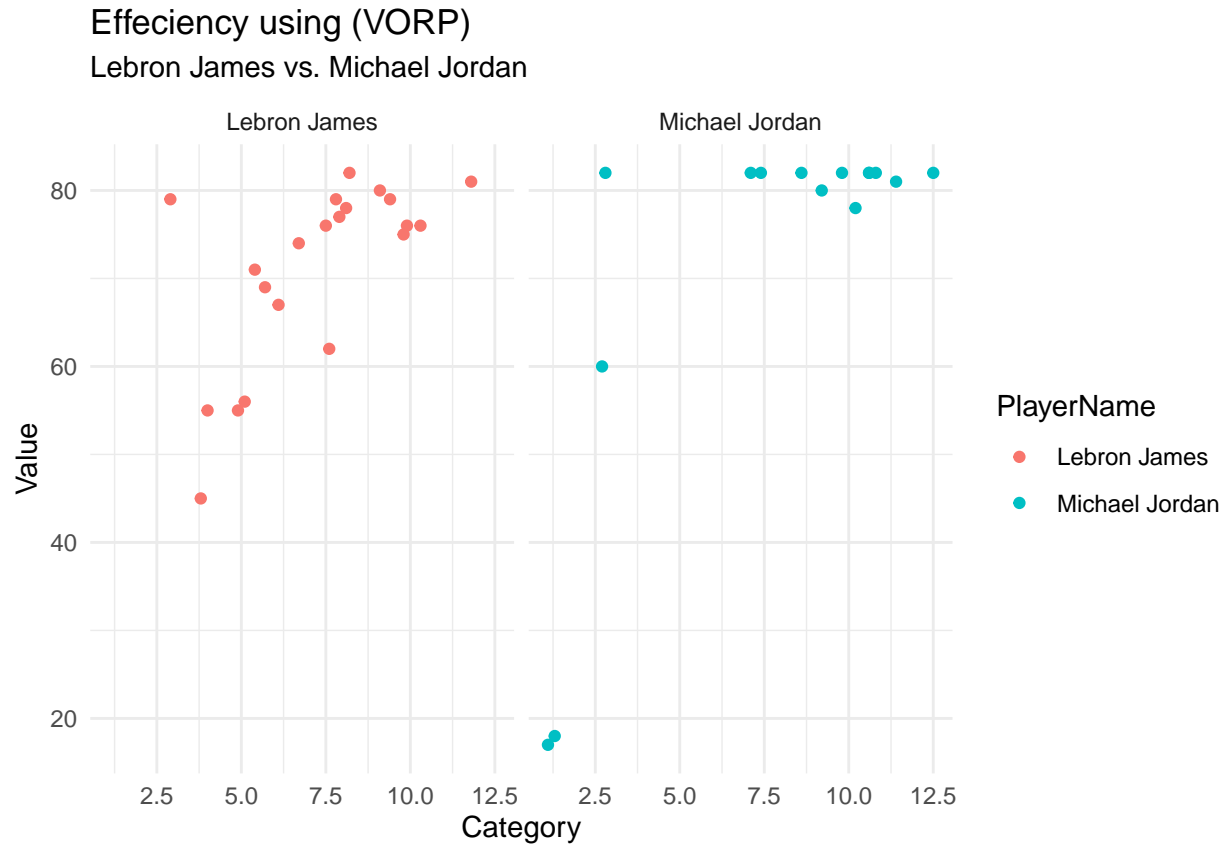
CombinedData <- bind_rows(LebronLongevity, JordanLongevity)

CombinedDataLonger <- CombinedData %>%
```

```
pivot_longer(cols = c(Games), names_to = "Statistic", values_to = "Value")
```

```
CombinedDataLonger %>%
```

```
ggplot(aes(x = ValueOverReplacementPlayer, y = Value, color = PlayerName)) +
  geom_point(shape = "circle", size = 1.5) +
  scale_color_hue(direction = 1) +
  theme_minimal() +
  facet_wrap(vars(PlayerName)) +
  labs(title = "Effeciency using (VORP)", subtitle = "Lebron James vs. Michael Jordan", x = "Category", y = "Value")
```



Bargraph showcasing Player Efficiency Ratings (PER)

```
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")
```

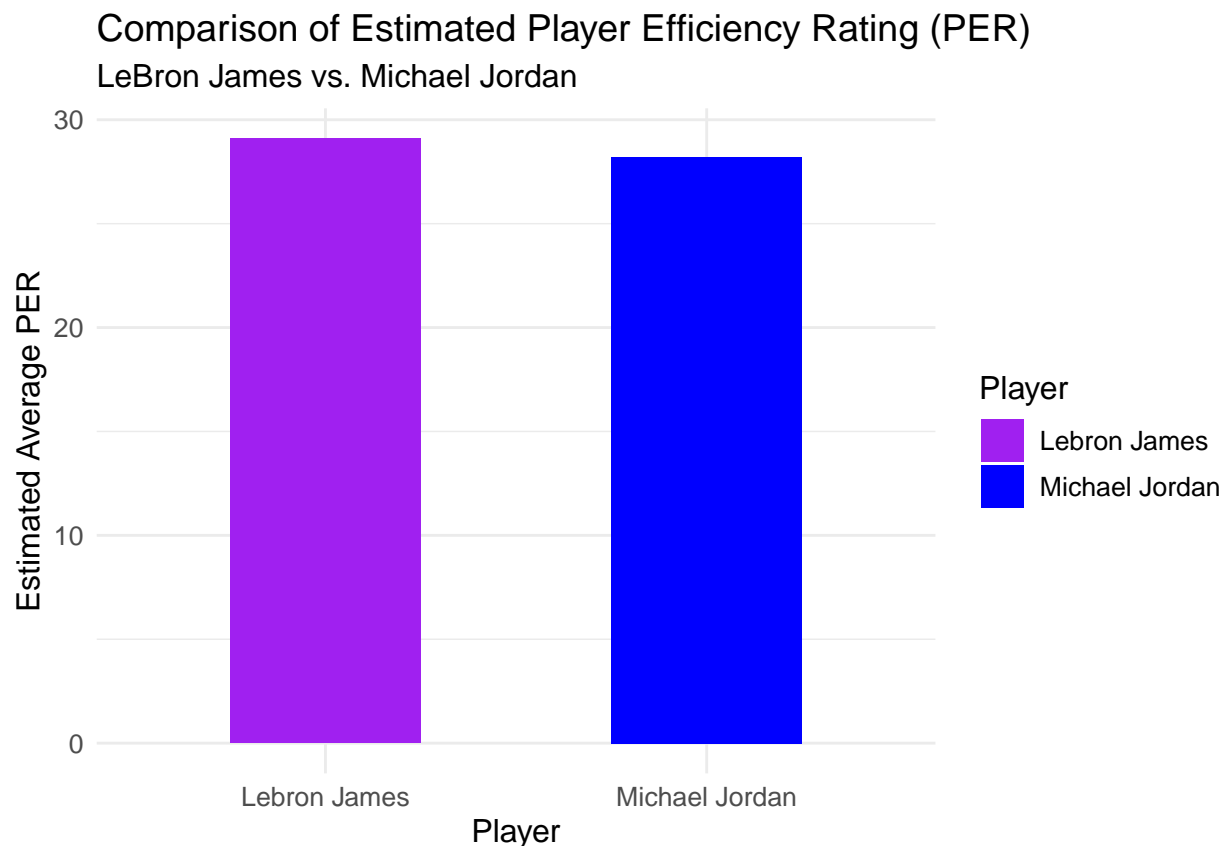
```
PERComparison <- bind_rows(LebronPER, JordanPER)
```

```
print(PERComparison)
```

```
## # A tibble: 2 x 2
```

```
## AveragePER Player
##      <dbl> <chr>
## 1      29.1 LeBron James
## 2      28.2 Michael Jordan
```

```
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" = "purple", "Michael Jordan" = "blue")) +
  theme(text = element_text(size = 12))
```



This graph compares LeBron and Jordans efficiency ratings over the course of their careers when comparing their turnover percentages to their scoring ability. The graph shows that Lebrons per 100 possessions stats makes LeBron a more efficient player.

3. Showcasing and comparing LeBron and Jordans Legacys

Barpgraph that shows LeBron and Jordans Efficiency ratings above age 30

```
OldBron <- LeBronPer100Poss %>%
  filter(Season > "2014-15") %>%
  mutate(PER = (PTS + TRB + AST + STL + BLK - ((FGA - FG) + (FTA - FT) + TOV)) / G) %>%
  mutate(Player = "LeBron James")

OldJordan <- JordanPer100Poss %>%
  filter(Season > "1994-95") %>%
  mutate(PER = (PTS + TRB + AST + STL + BLK - ((FGA - FG) + (FTA - FT) + TOV)) / G) %>%
```

```
mutate(Player = "Michael Jordan")
```

```
OldComp <- bind_rows(OldBron, OldJordan)
```

```
print(OldComp)
```

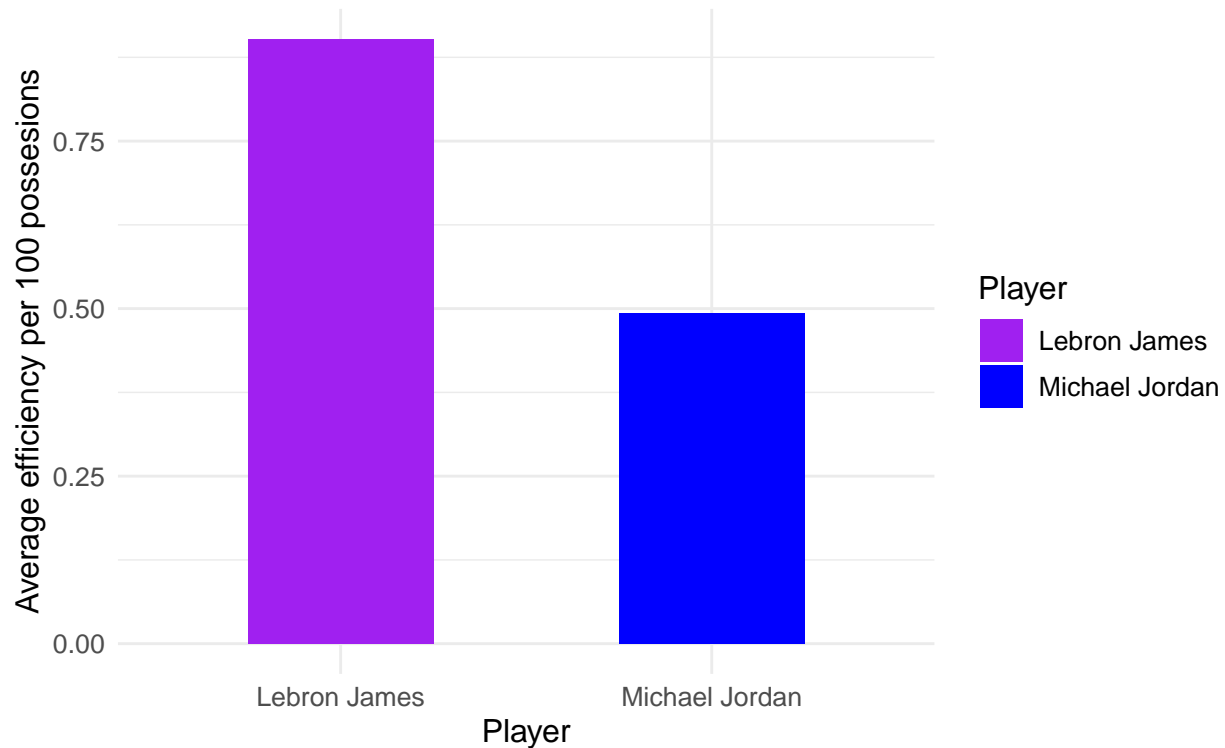
```
## # A tibble: 14 x 34
```

```
##   Season   Age Tm   Lg   Pos      G    GS    MP    FG    FGA 'FG%' '3P'
##   <chr>   <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2015-16   31 CLE  NBA   SF      76    76  2709   14    26.9 0.52   1.7
## 2 2016-17   32 CLE  NBA   SF      74    74  2794  13.1   24   0.548  2.2
## 3 2017-18   33 CLE  NBA   PF      82    82  3026  13.9   25.6 0.542  2.4
## 4 2018-19   34 LAL  NBA   SF      55    55  1937  13.4   26.3 0.51   2.7
## 5 2019-20   35 LAL  NBA   PG      67    67  2316  13.2   26.8 0.493   3
## 6 2020-21   36 LAL  NBA   PG      45    45  1504  13.7   26.6 0.513  3.4
## 7 2021-22   37 LAL  NBA    C      56    56  2084  14.7   28.1 0.524  3.7
## 8 2022-23   38 LAL  NBA   PF      55    54  1954  14.8   29.6 0.5    2.9
## 9 2023-24   39 LAL  NBA   PF      71    71  2504   13    24.1 0.54   2.8
##10 1995-96   32 CHI  NBA   SG      82    82  3090  15.6   31.5 0.495  1.9
##11 1996-97   33 CHI  NBA   SG      82    82  3106  15.8   32.5 0.486  1.9
##12 1997-98   34 CHI  NBA   SG      82    82  3181  14.9   32.1 0.465  0.5
##13 2001-02   38 WAS  NBA   SF      60    53  2093  14.3   34.4 0.416  0.3
##14 2002-03   39 WAS  NBA   SF      82    67  3031  12.2   27.4 0.445  0.3
```

```
## # i 22 more variables: '3PA' <dbl>, '3P%' <dbl>, '2P' <dbl>, '2PA' <dbl>,
## #   '2P%' <dbl>, FT <dbl>, FTA <dbl>, 'FT%' <dbl>, ORB <dbl>, DRB <dbl>,
## #   TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>, TOV <dbl>, PF <dbl>, PTS <dbl>,
## #   ...30 <lgl>, ORtg <dbl>, DRtg <dbl>, PER <dbl>, Player <chr>
```

```
ggplot(OldComp, aes(x = Player, y = PER, fill = Player)) +
  geom_bar(stat = "identity", position = position_dodge(), width = 0.5) +
  theme_minimal() +
  labs(title = "Efficiency per 100 possessions at age 30 plus", subtitle = "LeBron James vs. Michael Jordan", y
  scale_fill_manual(values = c("LeBron James" = "purple", "Michael Jordan" = "blue")) +
  theme(text = element_text(size = 12))
```

Efficiency per 100 possessions at age 30 plus LeBron James vs. Michael Jordan



This bar graph compares Michael Jordan and Lebrons efficiency's per 100 possessions at age 30 plus. We can clearly see the longevity Lebron holds. When comparing points, assists, rebounds,field goal percentages, and turnover rate, we see that Lebron dominates. At age 39 Lebron is still considered a top player in the NBA while Jordan was a dwindling old man expiring in his puny body.

Comparing total all star appearances

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

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

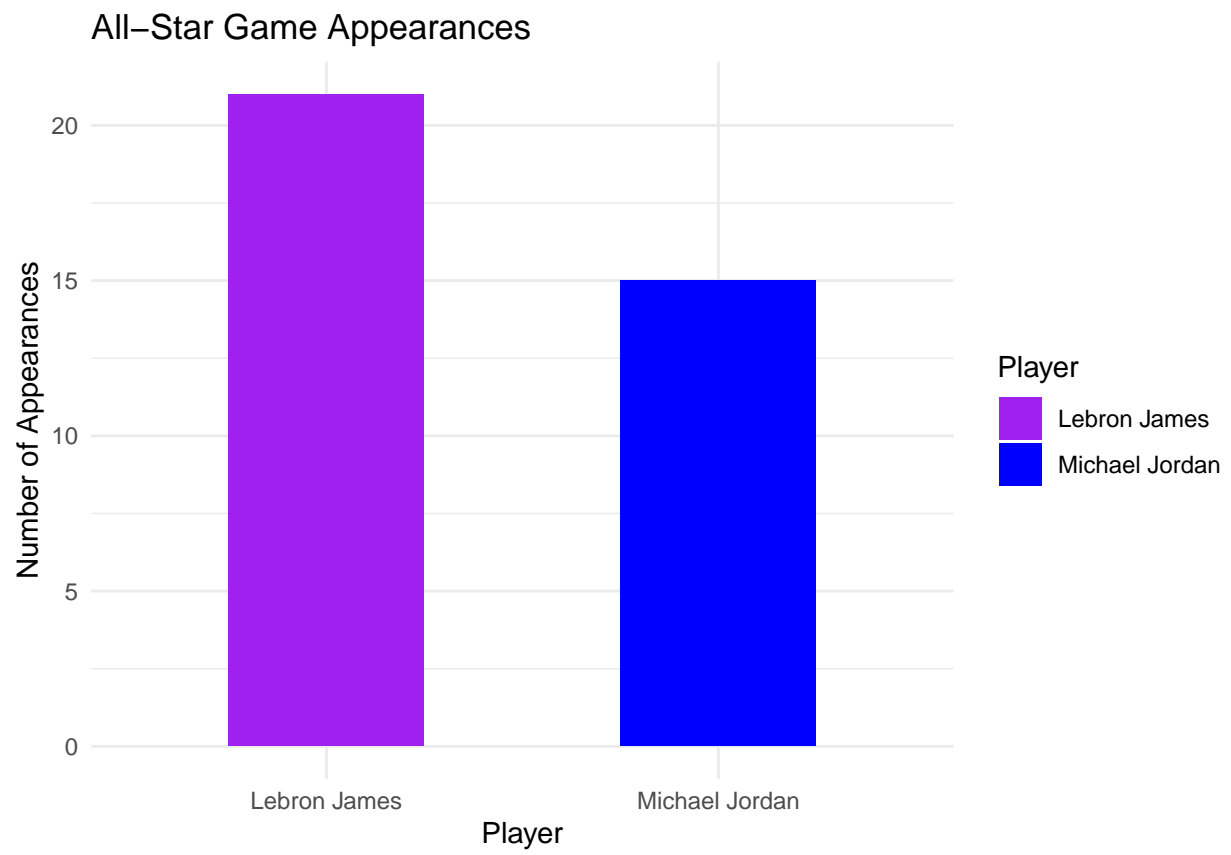
```
## [1] "Lebron James All-Star Game Appearances: 21"
```

```
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" = "purple", "Michael Jordan" = "blue"))
```

““