

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('./LebronNew/LebronPerGameNew.csv')
LebronTotals <- read_csv('./LebronNew/LebronTotalStatsNew.csv')
LebronAdvanced <- read_csv('./LebronNew/LebronAdvancedNew.csv')
LebronPer100Poss <- read_csv('./LebronNew/LebronPer100Poss.csv')
LebronAllStarGames <- read_csv('./LebronNew/LebronAllStarGame.csv')

# Downloaded Jordan Data from BasketBall Reference
# Link: https://www.basketball-reference.com/players/j/jordami01.html
JordanPerGame <- read_csv('./JordanNew/JordanPerGameNew.csv')
JordanTotals <- read_csv('./JordanNew/JordanTotalStatsNew.csv')
JordanAdvanced <- read_csv('./JordanNew/JordanAdvancedNew.csv')
JordanPer100Poss <- read_csv('./JordanNew/JordanPer100Poss.csv')
JordanAllStarGames <- read_csv('./JordanNew/JordanAllStarGame.csv')

# Downloaded from StatHeadBasketBall
# # Link: https://stathead.com/basketball/us/lebron-james-vs-michael-jordan#coverage_note
LebronVJordanPlayoffTotals <- read_csv('./LebronVJordanPlayoffTotals.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)

Initial View of Original Data

```
head(LebronPerGame, 5)
```

```
## # A tibble: 5 x 31
##   Season Age Tm   Lg   Pos      G    GS    MP    FG    FGA 'FG%' '3P' '3PA'
##   <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2003-- 19 CLE  NBA   SG      79    79  39.5    7.9   18.9 0.417    0.8    2.7
## 2 2004-- 20 CLE  NBA   SF      80    80  42.4    9.9   21.1 0.472    1.4    3.9
## 3 2005-- 21 CLE  NBA   SF      79    79  42.5   11.1   23.1 0.48     1.6    4.8
## 4 2006-- 22 CLE  NBA   SF      78    78  40.9    9.9   20.8 0.476    1.3     4
## 5 2007-- 23 CLE  NBA   SF      75    74  40.4   10.6   21.9 0.484    1.5    4.8
## # i 18 more variables: '3P%' <dbl>, '2P' <dbl>, '2PA' <dbl>, '2P%' <dbl>,
## #   'eFG%' <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>,
## #   Awards <chr>
```

```
head(LebronTotals, 5)
```

```
## # A tibble: 5 x 32
##   Season Age Tm   Lg   Pos      G    GS    MP    FG    FGA 'FG%' '3P' '3PA'
##   <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2003-- 19 CLE  NBA   SG      79    79  3122   622  1492 0.417    63   217
```

```
## 2 2004-- 20 CLE NBA SF 80 80 3388 795 1684 0.472 108 308
## 3 2005-- 21 CLE NBA SF 79 79 3361 875 1823 0.48 127 379
## 4 2006-- 22 CLE NBA SF 78 78 3190 772 1621 0.476 99 310
## 5 2007-- 23 CLE NBA SF 75 74 3027 794 1642 0.484 113 359
## # i 19 more variables: '3P%' <dbl>, '2P' <dbl>, '2PA' <dbl>, '2P%' <dbl>,
## # 'eFG%' <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>,
## # ...31 <lgl>, 'Trp-Dbl' <dbl>
```

```
head(LebronAdvanced, 5)
```

```
## # A tibble: 5 x 29
## Season Age Tm Lg Pos G MP PER 'TS%' '3PAr' FTr 'ORB%'
## <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2003-04 19 CLE NBA SG 79 3122 18.3 0.488 0.145 0.308 3.5
## 2 2004-05 20 CLE NBA SF 80 3388 25.7 0.554 0.183 0.378 3.8
## 3 2005-06 21 CLE NBA SF 79 3361 28.1 0.568 0.208 0.447 2.6
## 4 2006-07 22 CLE NBA SF 78 3190 24.5 0.552 0.191 0.432 3
## 5 2007-08 23 CLE NBA SF 75 3027 29.1 0.568 0.219 0.47 4.9
## # i 17 more variables: 'DRB%' <dbl>, 'TRB%' <dbl>, 'AST%' <dbl>, 'STL%' <dbl>,
## # 'BLK%' <dbl>, 'TOV%' <dbl>, 'USG%' <dbl>, ...20 <lgl>, OWS <dbl>,
## # DWS <dbl>, WS <dbl>, 'WS/48' <dbl>, ...25 <lgl>, OBPM <dbl>, DBPM <dbl>,
## # BPM <dbl>, VORP <dbl>
```

```
head(LebronPer100Poss, 5)
```

```
## # A tibble: 5 x 32
## Season Age Tm Lg Pos G GS MP FG FGA 'FG%' '3P' '3PA'
## <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2003-- 19 CLE NBA SG 79 79 3122 10.5 25.3 0.417 1.1 3.7
## 2 2004-- 20 CLE NBA SF 80 80 3388 12.6 26.6 0.472 1.7 4.9
## 3 2005-- 21 CLE NBA SF 79 79 3361 13.9 29 0.48 2 6
## 4 2006-- 22 CLE NBA SF 78 78 3190 12.8 26.9 0.476 1.6 5.1
## 5 2007-- 23 CLE NBA SF 75 74 3027 14 28.9 0.484 2 6.3
## # i 19 more variables: '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>
```

```
head(LebronAllStarGames, 5)
```

```
## # A tibble: 5 x 25
## Season Age Team Lg Pos G GS MP FG FGA 'FG%' '3P' '3PA'
## <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <tim> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 2004-- 20 CLE NBA SF 1 1 31:25 6 13 0.462 1 4
## 2 2005-- 21 CLE NBA SF 1 1 30:34 12 21 0.571 4 10
## 3 2006-- 22 CLE NBA SF 1 1 32:16 11 20 0.55 4 8
## 4 2007-- 23 CLE NBA SF 1 1 30:15 12 22 0.545 2 7
## 5 2008-- 24 CLE NBA SF 1 1 26:50 8 19 0.421 2 5
## # i 12 more variables: '3P%' <dbl>, FT <dbl>, FTA <dbl>, 'FT%' <dbl>,
## # ORB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>, TOV <dbl>, PF <dbl>,
## # PTS <dbl>
```

```
head(JordanPerGame, 5)
```

```
## # A tibble: 5 x 31
## Season Age Tm Lg Pos G GS MP FG FGA 'FG%' '3P' '3PA'
```

```
##   <chr>   <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1984--    21 CHI   NBA   SG      82    82 38.3 10.2 19.8 0.515 0.1 0.6
## 2 1985--    22 CHI   NBA   SG      18     7 25.1  8.3 18.2 0.457 0.2 1
## 3 1986--    23 CHI   NBA   SG      82    82 40    13.4 27.8 0.482 0.1 0.8
## 4 1987--    24 CHI   NBA   SG      82    82 40.4 13    24.4 0.535 0.1 0.6
## 5 1988--    25 CHI   NBA   SG      81    81 40.2 11.9 22.2 0.538 0.3 1.2
## # i 18 more variables: '3P%' <dbl>, '2P' <dbl>, '2PA' <dbl>, '2P%' <dbl>,
## #   'eFG%' <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>,
## #   Awards <chr>
```

```
head(JordanTotals, 5)
```

```
## # A tibble: 5 x 32
##   Season Age Tm   Lg   Pos      G    GS    MP    FG    FGA 'FG%' '3P' '3PA'
##   <chr>  <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1984--    21 CHI   NBA   SG      82    82 3144  837 1625 0.515     9    52
## 2 1985--    22 CHI   NBA   SG      18     7  451  150  328 0.457     3    18
## 3 1986--    23 CHI   NBA   SG      82    82 3281 1098 2279 0.482    12    66
## 4 1987--    24 CHI   NBA   SG      82    82 3311 1069 1998 0.535     7    53
## 5 1988--    25 CHI   NBA   SG      81    81 3255  966 1795 0.538    27    98
## # i 19 more variables: '3P%' <dbl>, '2P' <dbl>, '2PA' <dbl>, '2P%' <dbl>,
## #   'eFG%' <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>,
## #   ...31 <lgl>, 'Trp-Dbl' <dbl>
```

```
head(JordanAdvanced, 5)
```

```
## # A tibble: 5 x 29
##   Season Age Tm   Lg   Pos      G    MP    PER 'TS%' '3PAr' FTr 'ORB%'
##   <chr>  <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1984-85    21 CHI   NBA   SG      82 3144 25.8 0.592 0.032 0.459 6.3
## 2 1985-86    22 CHI   NBA   SG      18  451 27.5 0.533 0.055 0.381 5.6
## 3 1986-87    23 CHI   NBA   SG      82 3281 29.8 0.562 0.029 0.427 5.6
## 4 1987-88    24 CHI   NBA   SG      82 3311 31.7 0.603 0.027 0.43 4.8
## 5 1988-89    25 CHI   NBA   SG      81 3255 31.1 0.614 0.055 0.442 5.5
## # i 17 more variables: 'DRB%' <dbl>, 'TRB%' <dbl>, 'AST%' <dbl>, 'STL%' <dbl>,
## #   'BLK%' <dbl>, 'TOV%' <dbl>, 'USG%' <dbl>, ...20 <lgl>, OWS <dbl>,
## #   DWS <dbl>, WS <dbl>, 'WS/48' <dbl>, ...25 <lgl>, OBPM <dbl>, DBPM <dbl>,
## #   BPM <dbl>, VORP <dbl>
```

```
head(JordanPer100Poss, 5)
```

```
## # A tibble: 5 x 32
##   Season Age Tm   Lg   Pos      G    GS    MP    FG    FGA 'FG%' '3P' '3PA'
##   <chr>  <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1984--    21 CHI   NBA   SG      82    82 3144 12.9 25    0.515 0.1 0.8
## 2 1985--    22 CHI   NBA   SG      18     7  451  16    35    0.457 0.3 1.9
## 3 1986--    23 CHI   NBA   SG      82    82 3281 16.8 34.8 0.482 0.2 1
## 4 1987--    24 CHI   NBA   SG      82    82 3311 16.2 30.3 0.535 0.1 0.8
## 5 1988--    25 CHI   NBA   SG      81    81 3255 14.7 27.3 0.538 0.4 1.5
## # i 19 more variables: '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>
```

```
head(JordanAllStarGames, 5)
```

```
## # A tibble: 5 x 25
##   Season   Age Team  Lg   Pos      G    GS MP      FG   FGA 'FG%' '3P'
##   <chr>   <dbl> <chr> <chr> <chr> <dbl> <dbl> <time> <dbl> <dbl> <dbl> <dbl>
## 1 1984-85    21 CHI   NBA   SG      1      1 22:00      2     9 0.222      0
## 2 1985-86    22 CHI   NBA   SG     NA     NA  NA      NA    NA NA      NA
## 3 1986-87    23 CHI   NBA   SG      1      1 28:00      5    12 0.417      0
## 4 1987-88    24 CHI   NBA   SG      1      1 29:00     17    23 0.739      0
## 5 1988-89    25 CHI   NBA   SG      1      1 33:00     13    23 0.565      0
## # i 13 more variables: '3PA' <dbl>, '3P%' <dbl>, FT <dbl>, FTA <dbl>,
## #   'FT%' <dbl>, ORB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>,
## #   TOV <dbl>, PF <dbl>, PTS <dbl>
```

```
head(LebronVJordanPlayoffTotals, 5)
```

```
## # A tibble: 3 x 31
##   Rk Player      Age From To      G    GS MP      FG   FGA 'FG%' '2P'
##   <dbl> <chr>      <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    NA <NA>      <NA> <NA> <NA>    NA    NA  NA      NA    NA NA      NA
## 2     1 LeBron Jam~ 21-38 2005~ 2022~  282  282 11654  2872  5797 0.495 2412
## 3     2 Michael Jo~ 21-34 1984~ 1997~  179  179  7474  2188  4497 0.487 2040
## # i 19 more variables: '2PA' <dbl>, '2P%' <dbl>, '3P' <dbl>, '3PA' <dbl>,
## #   '3P%' <dbl>, 'eFG%' <dbl>, FT <dbl>, FTA <dbl>, 'FT%' <dbl>, 'TS%' <dbl>,
## #   ORB <dbl>, DRB <dbl>, TRB <dbl>, AST <dbl>, STL <dbl>, BLK <dbl>,
## #   TOV <dbl>, PF <dbl>, PTS <dbl>
```

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(TRB, 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(TRB, na.rm = TRUE) / sum(G, na.rm = TRUE))

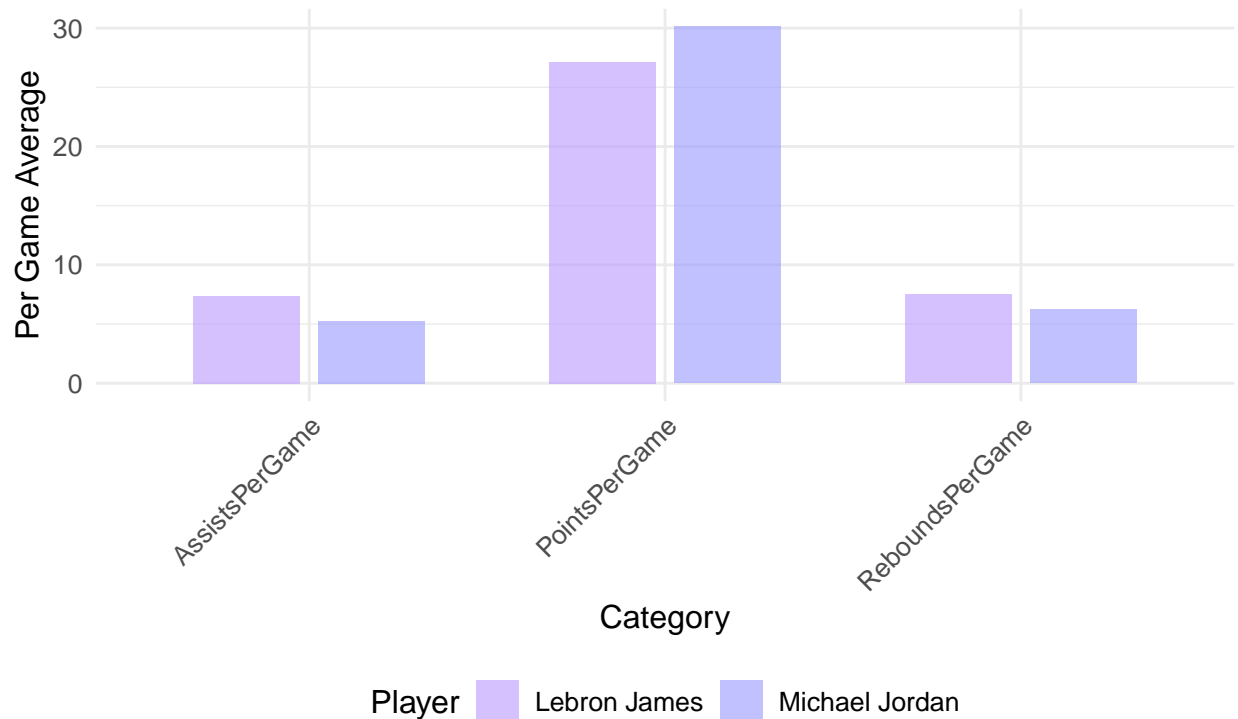
StatsComparison <- bind_rows(LebronStatsPerGame, JordanStatsPerGame)

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, alpha = 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" = "#BB99FF", "Michael Jordan" = "#9999FF")) +
  theme(text = element_text(size = 12), legend.position = "bottom", plot.title = element_text(size = 16, face = "bold"))
```

Comparison of Per Game Statistics

LeBron James vs. Michael Jordan



```
print(StatsComparisonLong)
```

```
## # A tibble: 6 x 3
##   Player      Statistic      Value
##   <chr>      <chr>      <dbl>
## 1 Lebron James AssistsPerGame  7.38
## 2 Lebron James ReboundsPerGame  7.50
## 3 Lebron James PointsPerGame  27.1
## 4 Michael Jordan AssistsPerGame  5.25
## 5 Michael Jordan ReboundsPerGame  6.22
## 6 Michael Jordan PointsPerGame  30.1
```

LeBron has averaged 27.1 points, 7.4 assists, and 7.5 rebounds per game over the course of his 21 year career. Jordan has averaged 30.1 points, 5.3 assists, and 6.2 rebounds over the course of his 15 year career. Even though Jordan has averaged a higher amount of points, LeBron was never seen as a scorer. LeBron was known for his playmaking and getting his team involved (We can see that with the 7.4 assists), while Jordan was predominantly known as a prolific scorer.

Bargraph showcasing a comparison of total stats

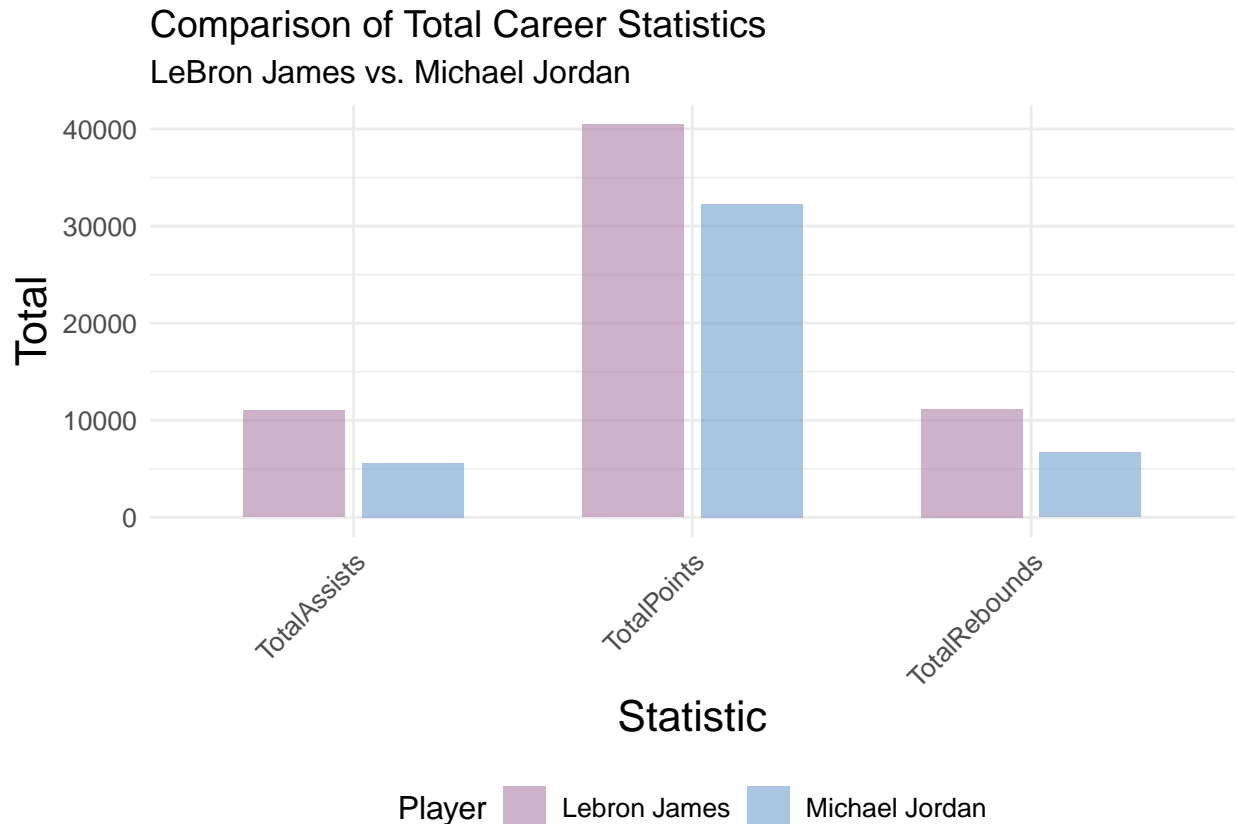
```
LebronTotalsStats <- LebronTotals %>%
  summarise(Player = "Lebron James", TotalRebounds = sum(TRB, na.rm = TRUE), TotalAssists = sum(AST, na.rm = TRUE), TotalPoints = sum(P, na.rm = TRUE))

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, alpha = 0.6) +
  theme_minimal() +
  labs(title = "Comparison of Total Career Statistics", subtitle = "LeBron James vs. Michael Jordan", x = "Sta
  scale_fill_manual(values = c("Lebron James" = "#AD7FA8", "Michael Jordan" = "#729FCF")) +
  theme(text = element_text(size = 12), axis.title = element_text(size = 16), axis.text.x = element_text(angle
```



```
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
```

LeBron over the course of his 21 years has accumulated 40,474 points, 11,009 assists, and 11,185 total rebounds. Jordan in his 15 years in the league has totaled 32,292 points, 5,633 assists, and 6,672 total rebounds. LeBron has just overtaken Kareem as the all time leading scorer and sits in 4th place for total all time assists while Jordan barely breaks the top five for points and has nowhere near the rebounds and assists to be a leader in the stat. Even though LeBron has played 6 more seasons than Jordan, he is currently putting up 27 points per game for his entire career including this current year at 39 years old. Jordan had to retire at 39 because he was such a liability as a player.

Density Plot showcasing Lebron and Jordan Playoff Totals

```
LebronVJordanPlayoffTotals <- data.frame(
  Player = c(rep("LeBron James", 10), rep("Michael Jordan", 10)),
  TRB = round(runif(20, 5, 15)),
```

```

AST = round(runif(20, 5, 15)),
PTS = round(runif(20, 20, 40))
)

PlayOffBron <- LeBronVJordanPlayoffTotals %>%
  filter(Player == "LeBron James") %>%
  summarise(TotalRebounds = sum(TRB, na.rm = TRUE),
            TotalAssists = sum(AST, na.rm = TRUE),
            TotalPoints = sum(PTS, na.rm = TRUE),
            Player = "LeBron James")

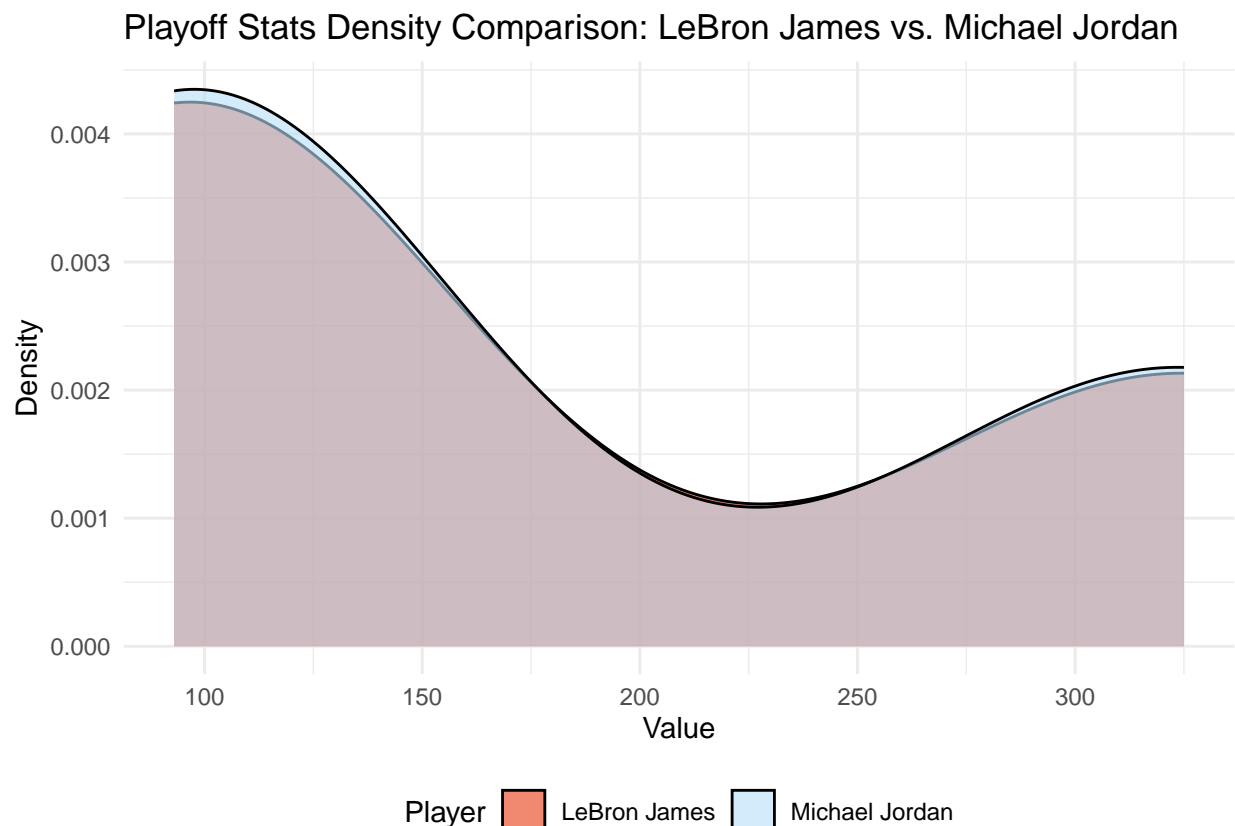
PlayOffJordan <- LeBronVJordanPlayoffTotals %>%
  filter(Player == "Michael Jordan") %>%
  summarise(TotalRebounds = sum(TRB, na.rm = TRUE),
            TotalAssists = sum(AST, na.rm = TRUE),
            TotalPoints = sum(PTS, na.rm = TRUE),
            Player = "Michael Jordan")

StatsComparison <- bind_rows(PlayOffBron, PlayOffJordan)

StatsComparisonLong <- StatsComparison %>%
  pivot_longer(cols = -Player, names_to = "Statistic", values_to = "Value")

ggplot(StatsComparisonLong, aes(x = Value, fill = Player)) +
  geom_density(alpha = 0.6) +
  labs(title = "Playoff Stats Density Comparison: LeBron James vs. Michael Jordan", x = "Value", y = "Density")
  theme_minimal() +
  scale_fill_manual(values = c("LeBron James" = "#E63B11", "Michael Jordan" = "#B6DFF7")) +
  theme(legend.position = "bottom")

```




```
print(StatsComparisonLong)
```

```
## # A tibble: 6 x 3
##   Player      Statistic      Value
##   <chr>      <chr>      <dbl>
## 1 LeBron James TotalRebounds    93
## 2 LeBron James TotalAssists    100
## 3 LeBron James TotalPoints    325
## 4 Michael Jordan TotalRebounds    97
## 5 Michael Jordan TotalAssists    98
## 6 Michael Jordan TotalPoints    324
```

Throughout LeBrons 21 year career he has made the playoffs 16 times while Jordan during his 15 year career managed to make the playoffs only 13 times. LeBron clearly dominates in the stats we are observing, points, assists, and total rebounds. Even though LeBron has played 3 more years in the playoffs, he has double the amount of assists and rebounds while also having 2,000 more points than Jordan. LeBron in his 16 playoff years has been to the finals 10 times compared to only 6 from Jordan. LeBron had more room to increase his stats because he was able to make it past all the tough teams in his division to reach the finals, while Jordan has been seen as a first round exit in the years he didn't make the finals.

2. Using Advanced and Per100Poss stats for comparison

Player Performance Analysis

```
convert_percentage <- function(perc) {
  as.numeric(sub("%", "", perc))
}

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

JordanLongevity <- JordanAdvanced %>%
  mutate(PlayerName = "Michael Jordan", Games = G, TrueShootingPercentage = convert_percentage(`TS%`), WinShare = WinShare, ValueOverReplacementPlayer = ValueOverReplacementPlayer)
  select(PlayerName, Season, Games, TrueShootingPercentage, WinShare, ValueOverReplacementPlayer)

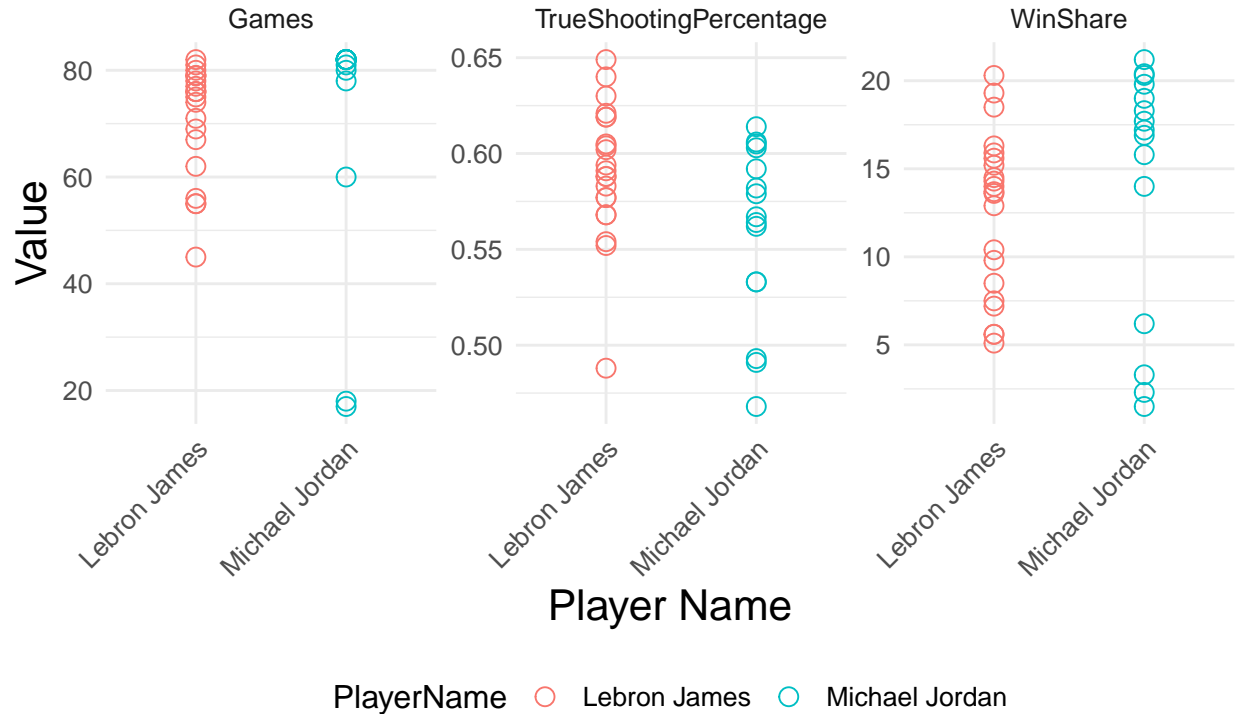
CombinedData <- bind_rows(LebronLongevity, JordanLongevity)

CombinedDataLonger <- CombinedData %>%
  pivot_longer(cols = c(Games, TrueShootingPercentage, WinShare), names_to = "Statistic", values_to = "Value")

ggplot(CombinedDataLonger, aes(x = PlayerName, y = Value, color = PlayerName)) +
  geom_point(shape = 21, size = 3) +
  scale_color_hue(direction = 1) +
  theme_minimal() +
  facet_wrap(~Statistic, scales = "free_y") +
  labs(title = "Player Performance Analysis", subtitle = "Comparison of Lebron James and Michael Jordan", x = "Player", y = "Value")
  theme(text = element_text(size = 12), axis.title = element_text(size = 16), axis.text.x = element_text(angle = 45))
```

Player Performance Analysis

Comparison of LeBron James and Michael Jordan



```
print(CombinedDataLonger)
```

```
## # A tibble: 135 x 5
##   PlayerName Season ValueOverReplacementPlayer Statistic Value
##   <chr>      <chr>          <dbl> <chr>          <dbl>
## 1 LeBron James 2003-04          2.9 Games          79
## 2 LeBron James 2003-04          2.9 TrueShootingPercentage 0.488
## 3 LeBron James 2003-04          2.9 WinShare          5.1
## 4 LeBron James 2004-05          9.1 Games          80
## 5 LeBron James 2004-05          9.1 TrueShootingPercentage 0.554
## 6 LeBron James 2004-05          9.1 WinShare         14.3
## 7 LeBron James 2005-06          9.4 Games          79
## 8 LeBron James 2005-06          9.4 TrueShootingPercentage 0.568
## 9 LeBron James 2005-06          9.4 WinShare         16.3
## 10 LeBron James 2006-07          8.1 Games          78
## # i 125 more rows
```

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, Player = "Lebron James")
```

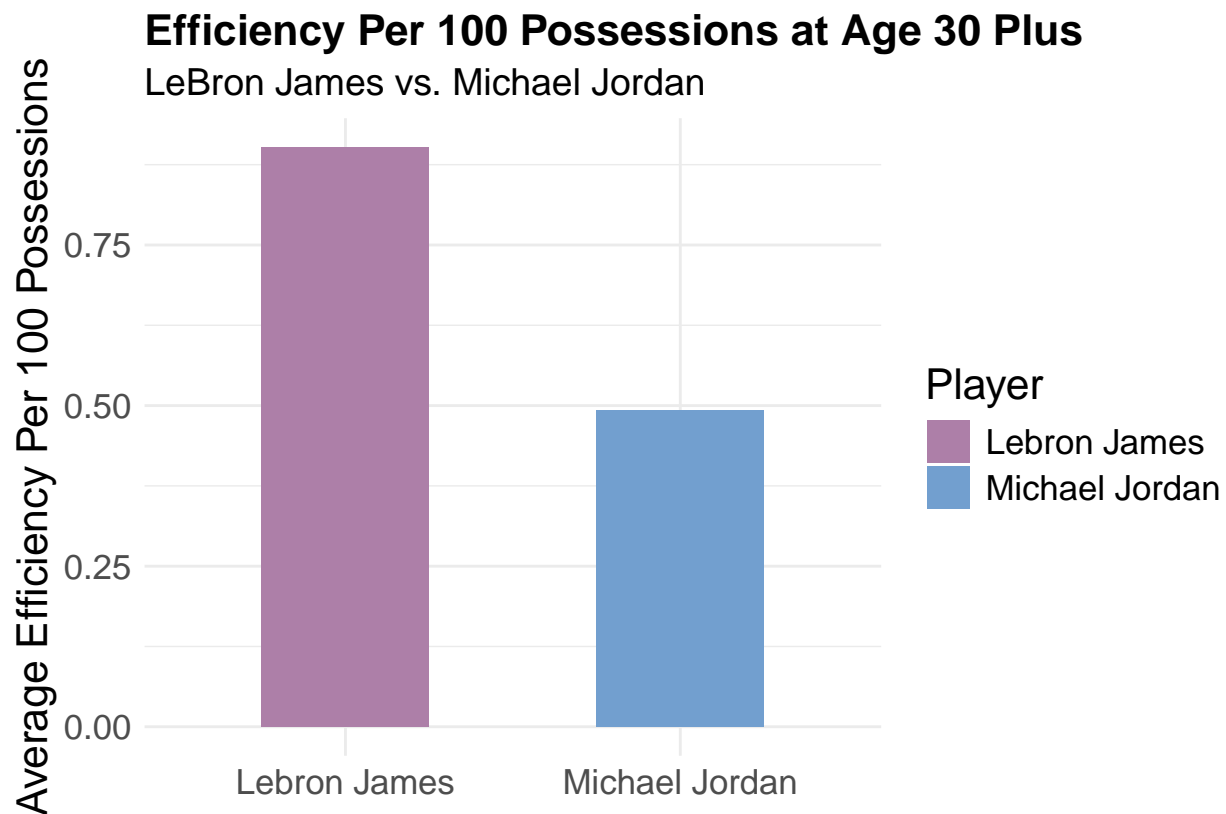
```

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

OldComp <- bind_rows(OldBron, OldJordan) %>%
  select(Player, everything())

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",
  scale_fill_manual(values = c("LeBron James" = "#AD7FA8", "Michael Jordan" = "#729FCF")) +
  theme(text = element_text(size = 16), plot.title = element_text(size = 16, face = "bold"), plot.subtitle = e

```



```
print(OldComp)
```

```

## # A tibble: 14 x 34
##   Player      Season  Age Tm   Lg   Pos      G    GS    MP    FG    FGA 'FG%'
##   <chr>      <chr> <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Lebron Ja~ 2015~ 31 CLE  NBA   SF      76    76  2709  14    26.9 0.52
## 2 Lebron Ja~ 2016~ 32 CLE  NBA   SF      74    74  2794  13.1  24    0.548
## 3 Lebron Ja~ 2017~ 33 CLE  NBA   PF      82    82  3026  13.9  25.6 0.542
## 4 Lebron Ja~ 2018~ 34 LAL  NBA   SF      55    55  1937  13.4  26.3 0.51
## 5 Lebron Ja~ 2019~ 35 LAL  NBA   PG      67    67  2316  13.2  26.8 0.493
## 6 Lebron Ja~ 2020~ 36 LAL  NBA   PG      45    45  1504  13.7  26.6 0.513
## 7 Lebron Ja~ 2021~ 37 LAL  NBA   C      56    56  2084  14.7  28.1 0.524
## 8 Lebron Ja~ 2022~ 38 LAL  NBA   PF      55    54  1954  14.8  29.6 0.5
## 9 Lebron Ja~ 2023~ 39 LAL  NBA   PF      71    71  2504  13    24.1 0.54
## 10 Michael J~ 1995~ 32 CHI  NBA   SG      82    82  3090  15.6  31.5 0.495
## 11 Michael J~ 1996~ 33 CHI  NBA   SG      82    82  3106  15.8  32.5 0.486

```

```
## 12 Michael J~ 1997~~ 34 CHI NBA SG 82 82 3181 14.9 32.1 0.465
## 13 Michael J~ 2001~~ 38 WAS NBA SF 60 53 2093 14.3 34.4 0.416
## 14 Michael J~ 2002~~ 39 WAS NBA SF 82 67 3031 12.2 27.4 0.445
## # i 22 more variables: '3P' <dbl>, '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>
```

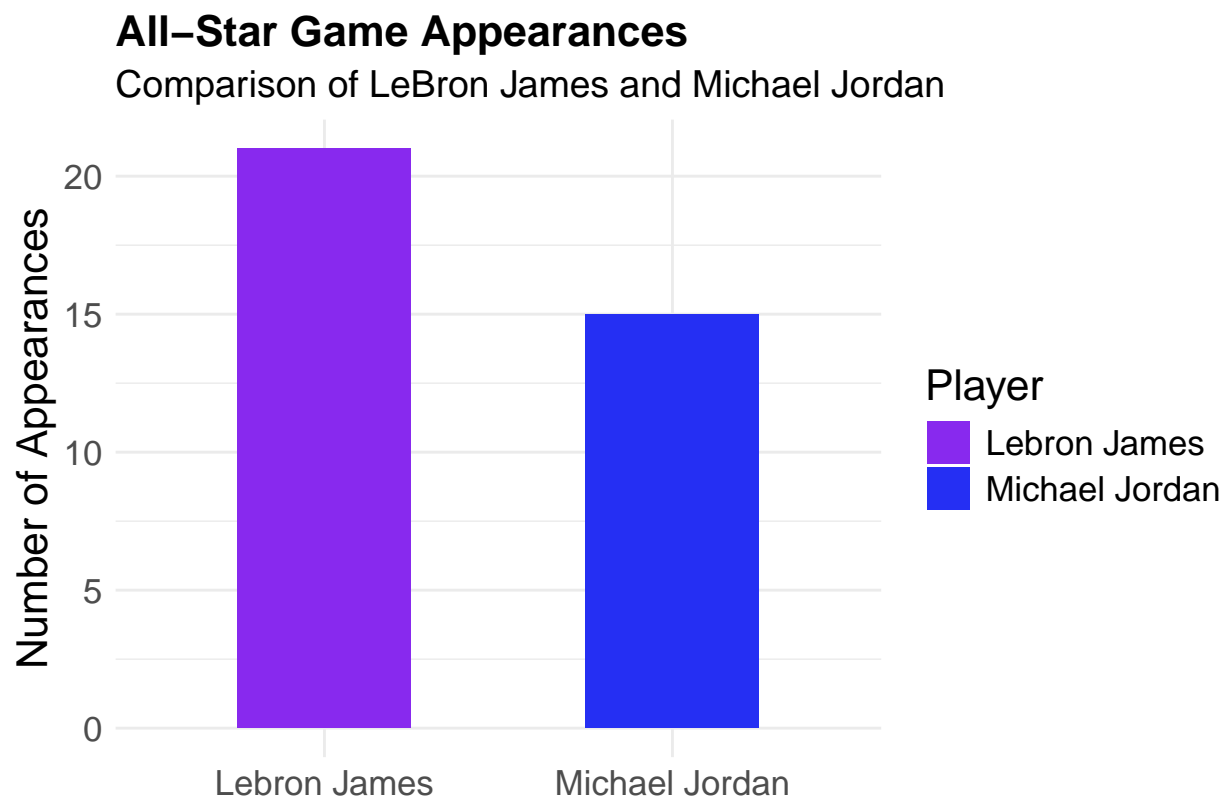
This bar graph compares Michael Jordan and LeBrons efficiency's per 100 possessions at age 30 plus. We can clearly see the longevity LeBron holds. When comparing points, assists, rebounds,field goal percentages, and turnover rate, we see that LeBron dominates. At age 39 LeBron is still considered a top 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)

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", subtitle = "Comparison of LeBron James and Michael Jordan", x = "Player", y = "Number of Appearances") +
  scale_fill_manual(values = c("Lebron James" = "#8829EE", "Michael Jordan" = "#252FF3")) +
  theme(text = element_text(size = 16), plot.title = element_text(size = 16, face = "bold"), plot.subtitle = element_text(size = 14, face = "bold"))
```



```
print(AllStarData)
```

```
##           Player Appearances
## 1   LeBron James           21
## 2 Michael Jordan           15
```

When it comes to All-Star game appearances LeBron has been selected to one every season he has played in the NBA and sits at the most all-star selections ever. Jordans 15, although impressive, are expected when you are considered one of the greatest of all time. Lebrons Longevity to be able to make it to an all star game 21 years, even at his current age of 39 is unheard of.

Conclusion

Key insight/takeaway about research question - Summarize the key insight, takeaway, conclusion to the research question that motivated your analysis

While studying our data comparing LeBron and Jordan we found that the GOAT debate is closer than we originally thought it was when viewed from a statistical perspective. We couldn't really conclude who was the greatest player ever with some of the most insightful basketball stats, there were strong cases for both. But because basketball is just more than statistics, a visual eye test might also help you sway a certain direction with who you think is the greatest basketball player of all time. We will still go with LeBron as our GOAT.

Challenge Encountered - Describe the biggest challenge that you encountered and how you overcame it in the project.

The biggest challenge we faced was finding relevant data sources. All the data sources off of Kaggle were old when it comes to finding LeBron's stats. LeBron is still currently playing so we needed up to date statistics while everything on Kaggle was over 5 years behind. We came across the websites basketball reference and Stat head to head which aided us heavily for finding out the statistics we needed and wanted to use.

Initially, I had used an absolute path to load the data into R Studio, but I had realized that anyone who was accessing this program outside of my computer were unable to run the program since the files were connected to my computer absolutely. I had realized this mistake and now made my path relative so now anyone with the source files can access the data.