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/vs/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?

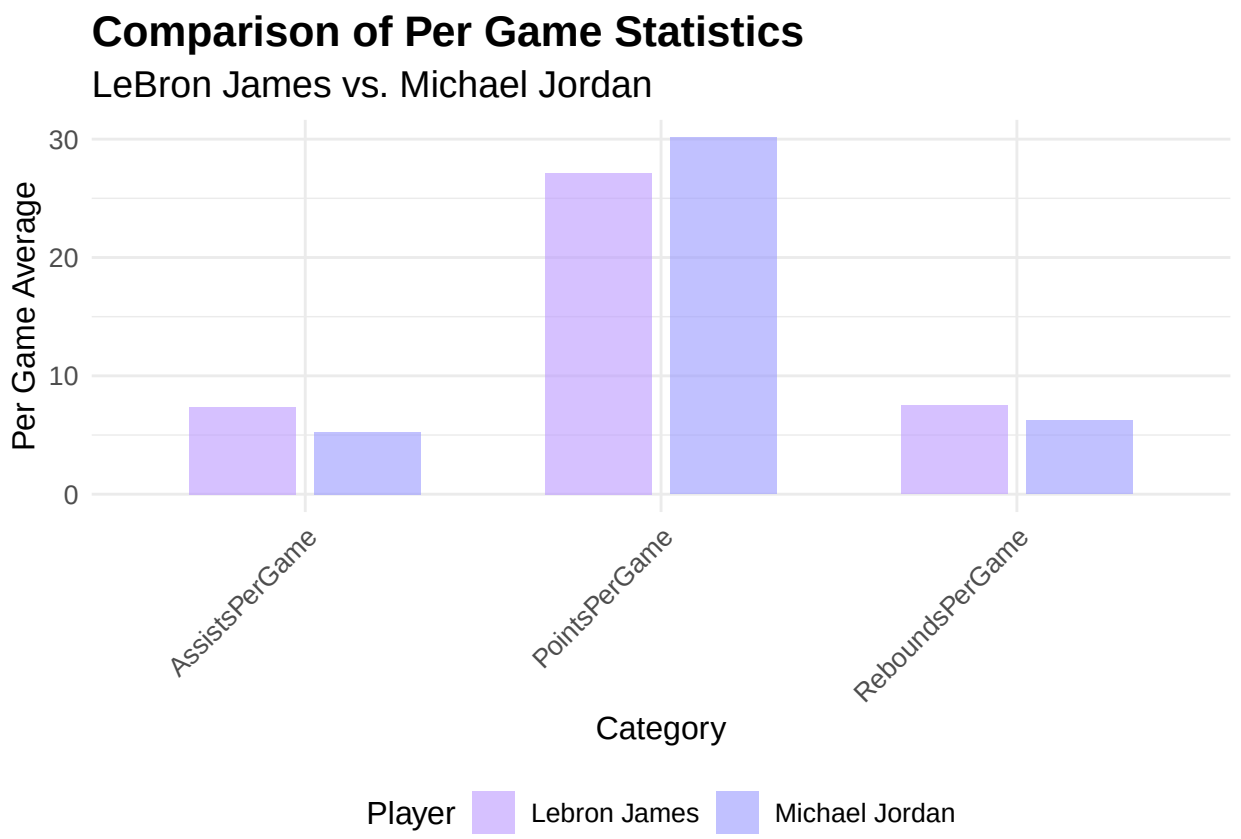
- Points per game (PointsPerGame)
- Assists per game (AssistsPerGame)
- Rebounds per game (ReboundsPerGame)
- Total Points (TotalPoints)
- Total Assists (TotalAssists)
- Total Rebounds (TotalRebounds)
- Player efficiency rating (PlayerEfficiencyRating)
- Career achievements (CareerAchievements)
- All-star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

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), ReboundsP

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

StatsComparison <- bind_rows(LebronStatsPerGame, JordanStatsPerGame)

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

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 = "Categor
  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 =
```




```
print(StatsComparisonLong)
```

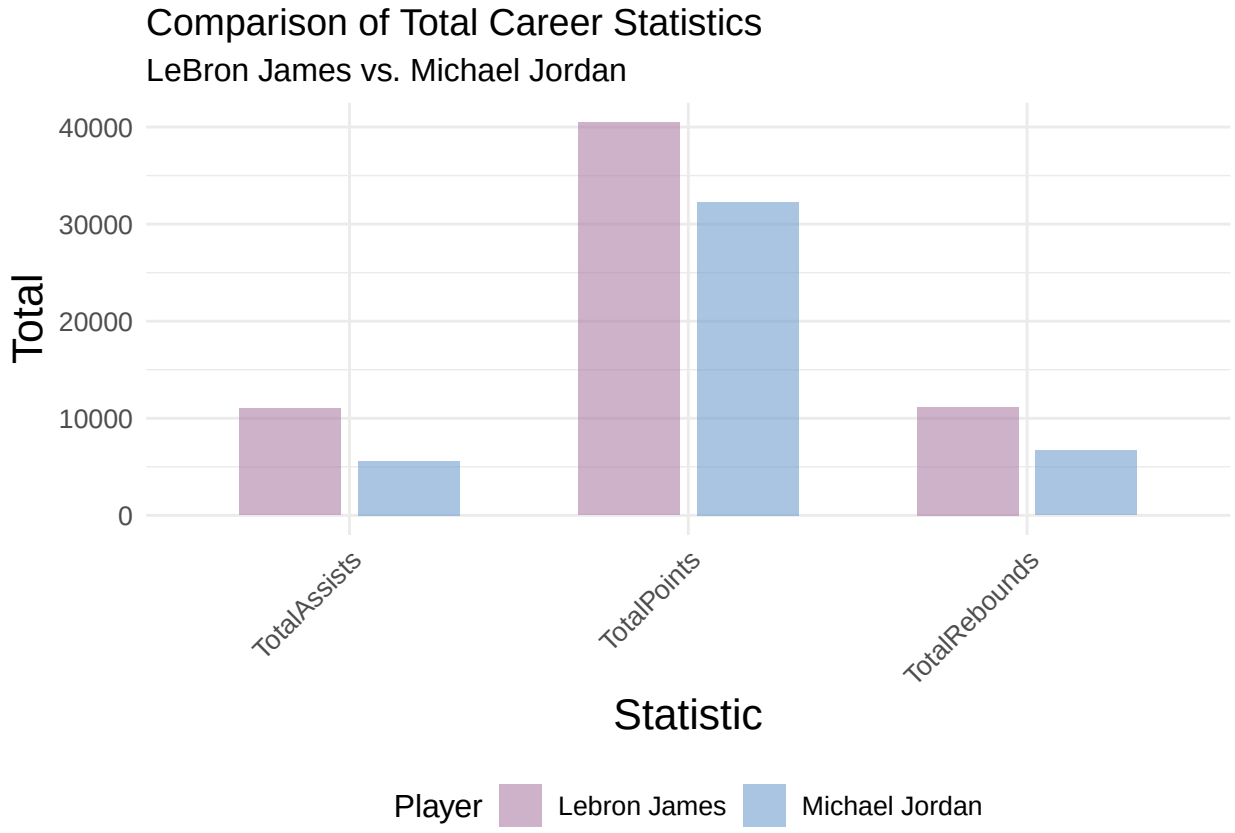
```
## # A tibble: 6 x 3
##   Player         Statistic       Value
##   <chr>          <chr>           <dbl>
## 1 Lebron James   AssistsPerGame   7.38
## 2 Lebron James   ReboundsPerGame  7.50
## 3 Lebron James   PointsPerGame   27.1
## 4 Michael Jordan AssistsPerGame   5.25
## 5 Michael Jordan ReboundsPerGame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

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

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)
```

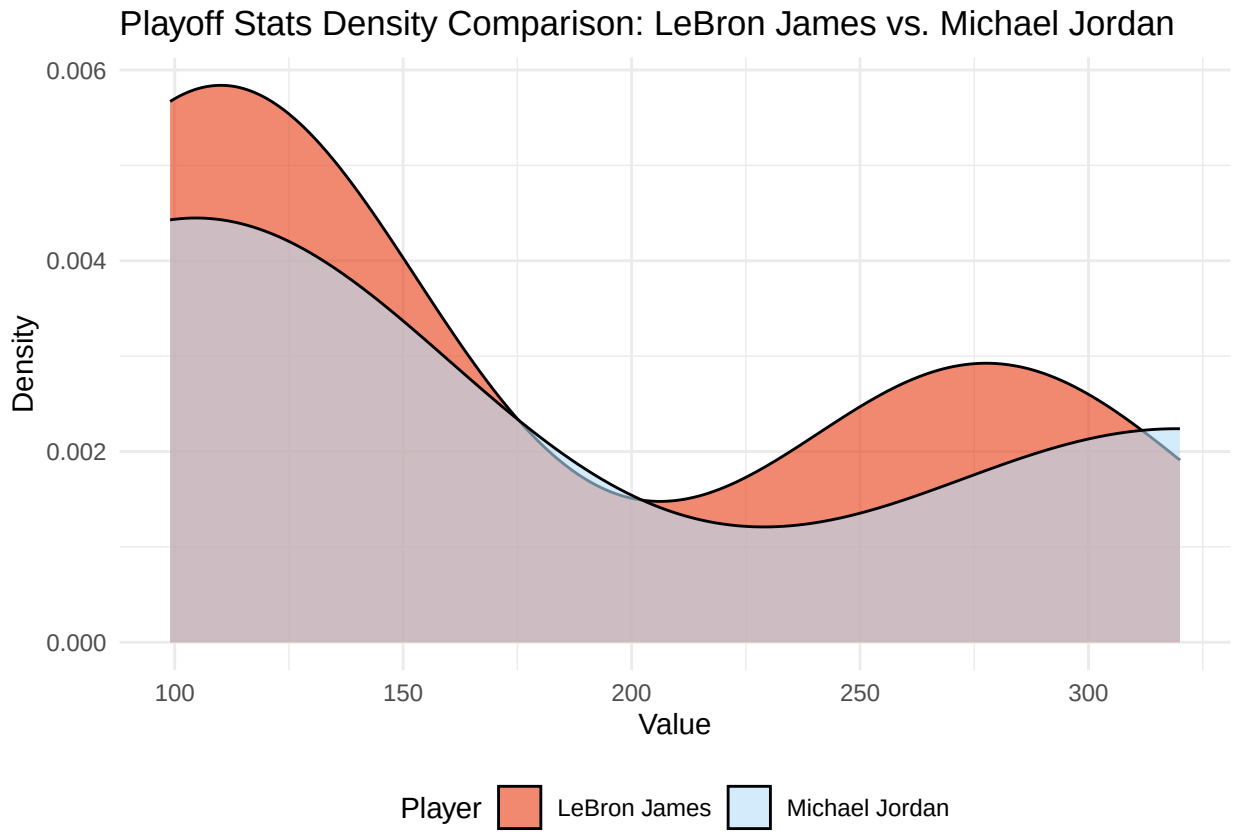
```
## # A tibble: 2 x 4
##   Player         TotalRebounds TotalAssists TotalPoints
##   <chr>                  <dbl>        <dbl>       <dbl>
## 1 Lebron James           11185        11009       40474
## 2 Michael Jordan          6672         5633       32292
```

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")
```



```r
print(StatsComparisonLong)
```

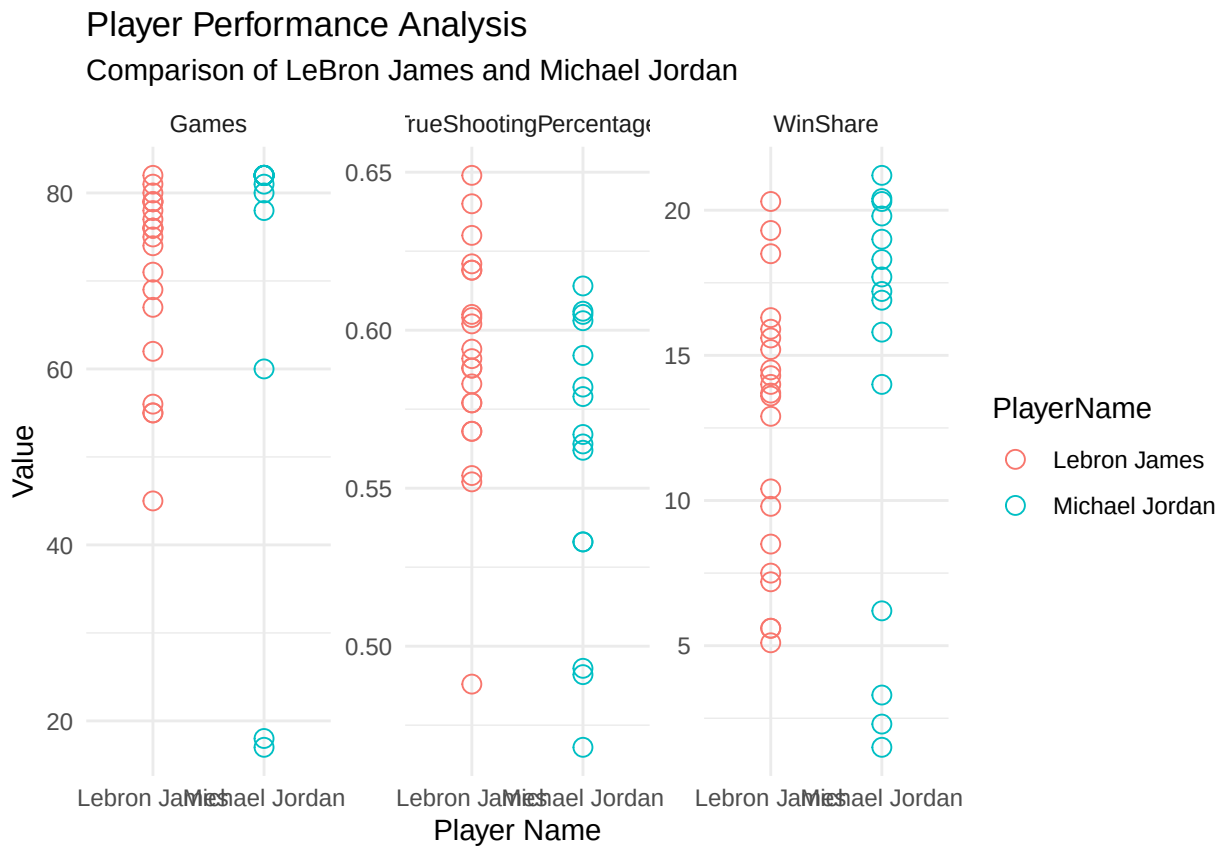
```
## # A tibble: 6 x 3
##   Player         Statistic     Value
##   <chr>          <chr>         <dbl>
## 1 LeBron James   TotalRebounds   111
## 2 LeBron James   TotalAssists    109
## 3 LeBron James   TotalPoints     278
## 4 Michael Jordan TotalRebounds    99
## 5 Michael Jordan TotalAssists    110
## 6 Michael Jordan TotalPoints     320
```

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 cleary 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

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

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

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




```
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
```

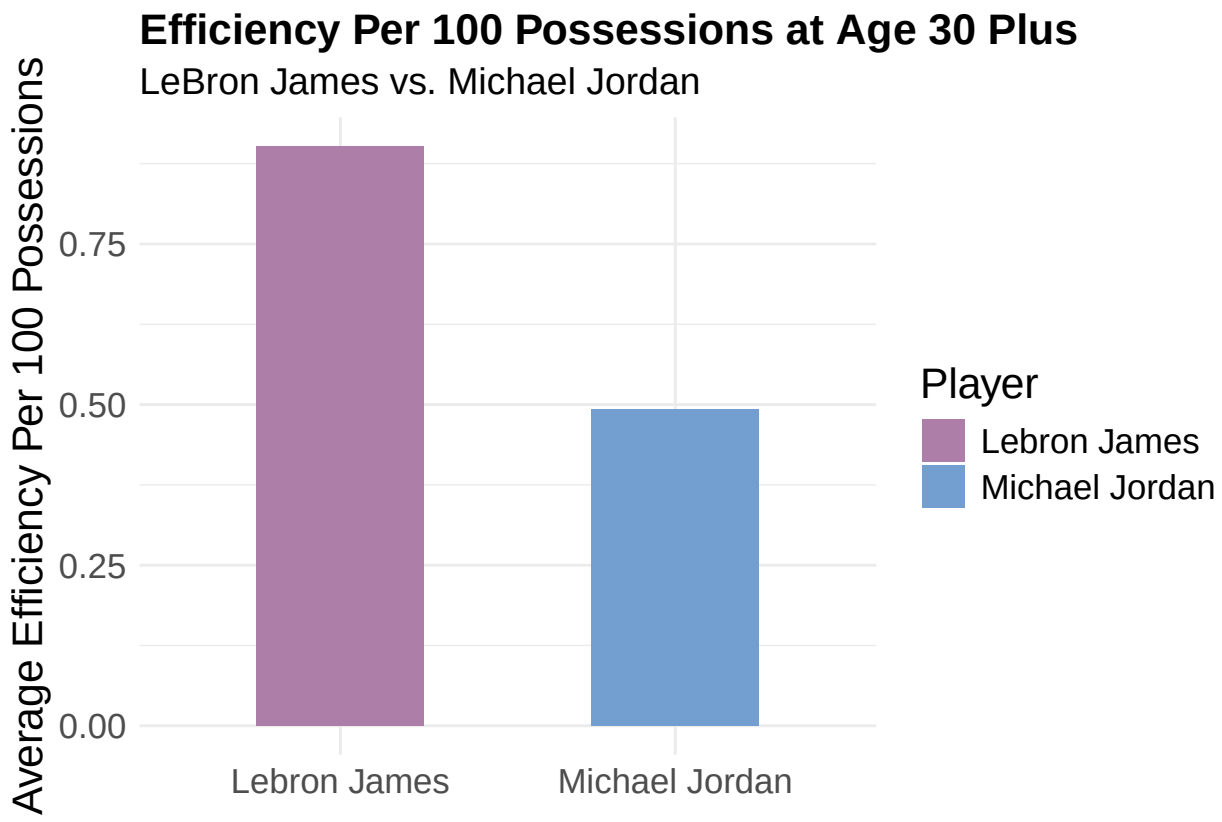
This graph compares LeBron and Jordans efficiency ratings over the course of their careers when comparing their turnover percentages to their scoring ability. The graph shows that LeBrons per 100 possessions stats makes LeBron a more efficient player.

3. Showcasing and comparing LeBron and Jordans Legacys

Barpgraph that shows LeBron and Jordans Effeciency 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")
```

```r
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
```




```r
print(OldComp)
```

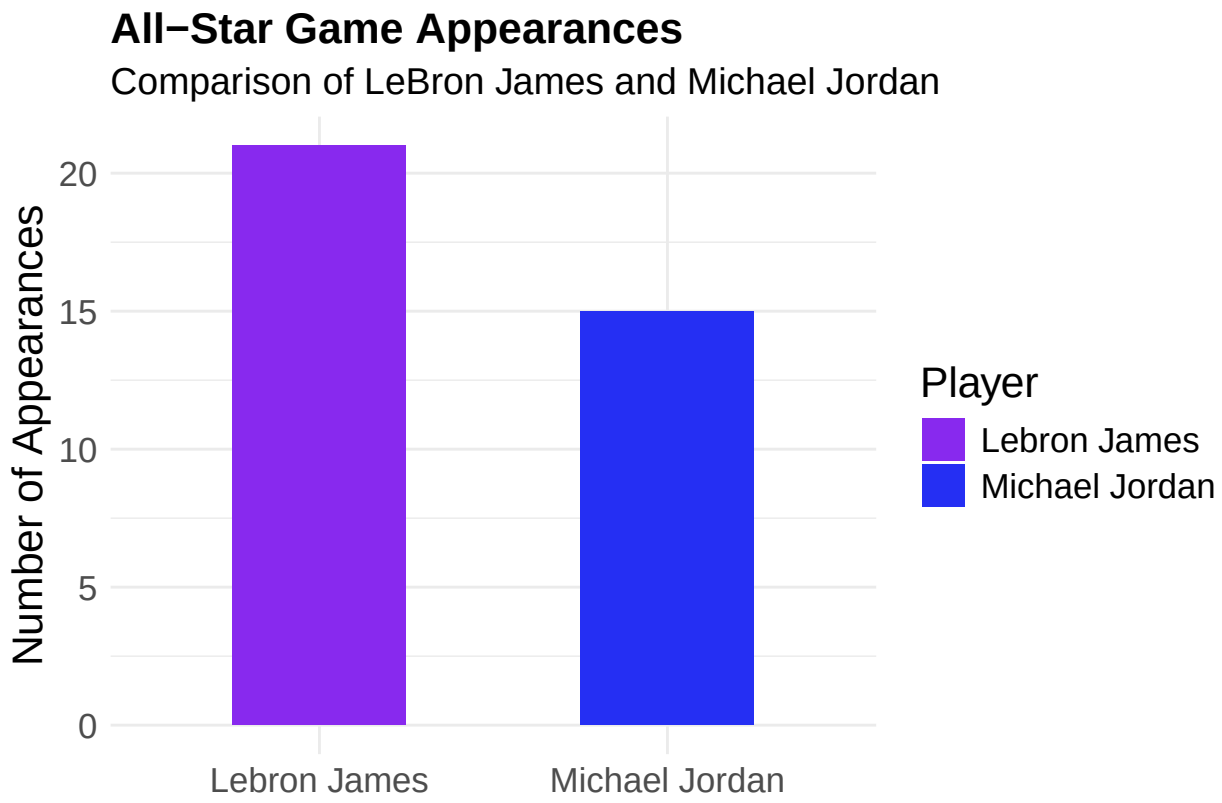
```
## # A tibble: 14 x 34
##    Player     Season   Age Tm    Lg    Pos       G    GS    MP    FG   FGA `FG%`
##    <chr>      <chr>  <dbl> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 Lebron Ja~ 2015-~    31 CLE   NBA   SF       76    76  2709  14    26.9 0.52
##  2 Lebron Ja~ 2016-~    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 appearences

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

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

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 = ""
  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 = e
```



```r
print(AllStarData)
```

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

When it comes to All-Star game appearences Lebron has been slected 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.