Project 1 Sports Analytics

Audrey Holloman and Nicole Kadosh

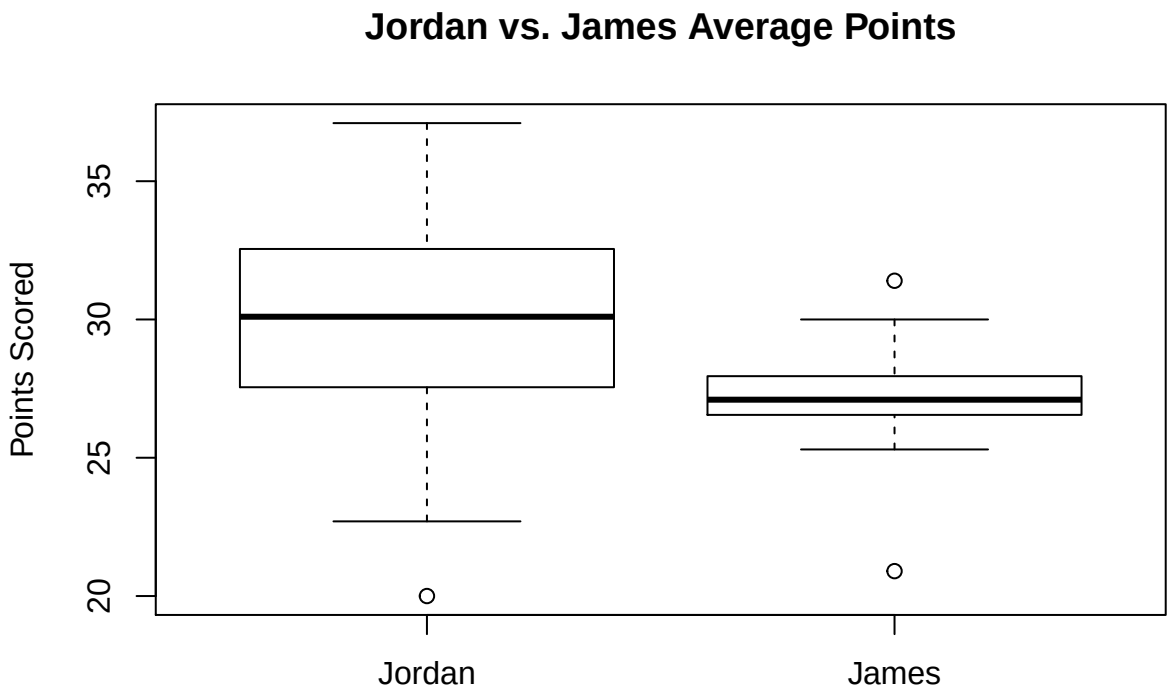
8/30/2018

```
knitr::opts_chunk$set(
  echo = TRUE,
  fig.align = "center",
  message = FALSE,
  warning = FALSE
)
```

```
library(readxl)
library(tidyverse)
new <- read_csv("~/JordanJamesNew.csv")
```

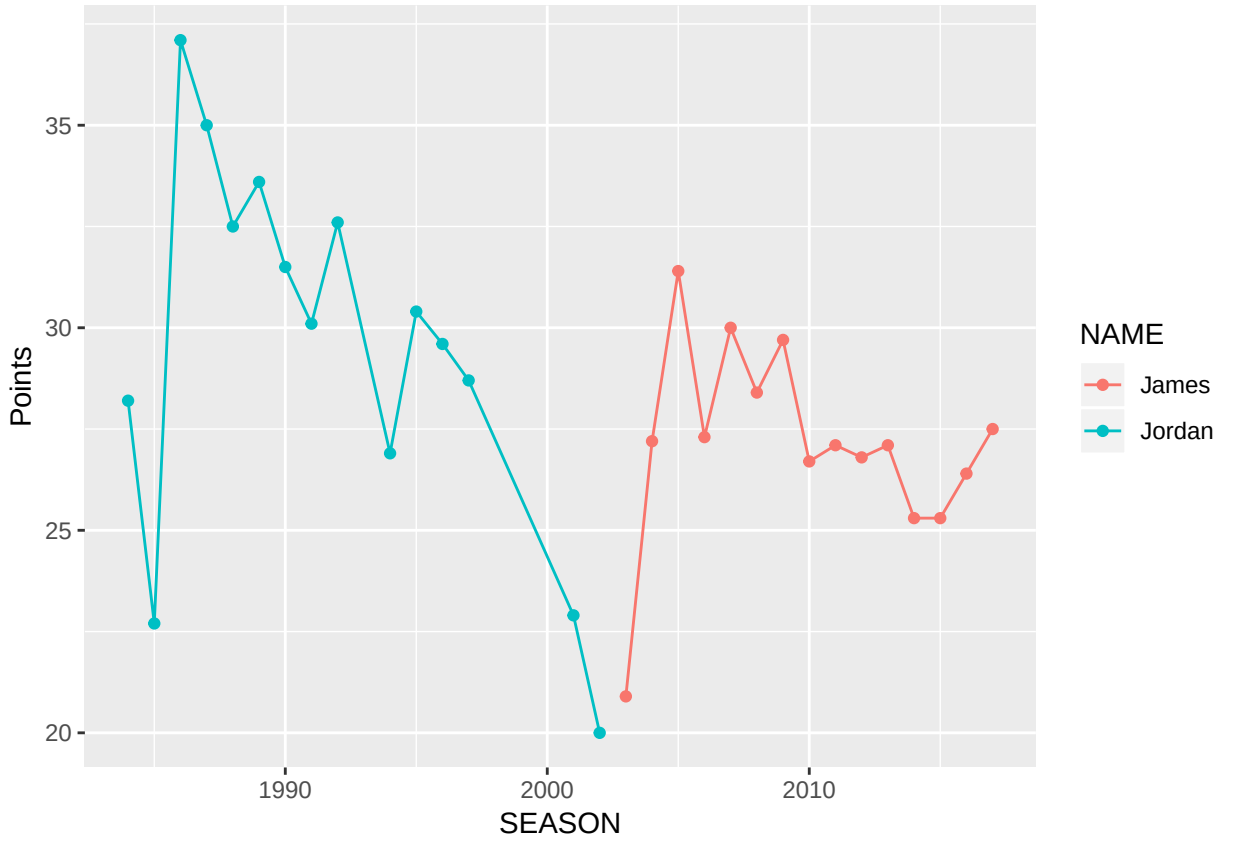
In determing whether Michael Jordan or Lebron James is the greatest player, we decided to use the varibles; points, 3 points made, free throws made, defensive rebounds and steals. We only used the regular season stats, as there are more games within a regular season, therefore, giving more data to analyze. The five variables are all positive aspects of each of the players, however, we decided they best tested the skill of each player. The variables are both defensive and offensive skills. We used box plots to compare the averages of each varaible and the outliers. For the box plots, the goal for the players is to have a higher overall weighted average and smaller range because this infers that the player is more consistent. Next we graphed each variable for the players side by side allowing for comparison. We combined all the variables together in a weighted average in order to determine, which player was the greatest.

```
 JordanJames <- read_excel("JordanJames.xls")
boxplot(JordanJames$PTS, JordanJames$PTS2, names = c("Jordan", "James"), main = "Jordan vs. James Averag
```




```
ggplot(new) +
  geom_line(aes(x=SEASON, y=PTS, col=NAME)) +
```

```
  geom_point(aes(x=SEASON, y=PTS, col=NAME)) +
  ylab("Points")
```




```
Jordan_Avg_Pts <- 5 * mean(JordanJames$PTS)
Jordan_Avg_Pts
```

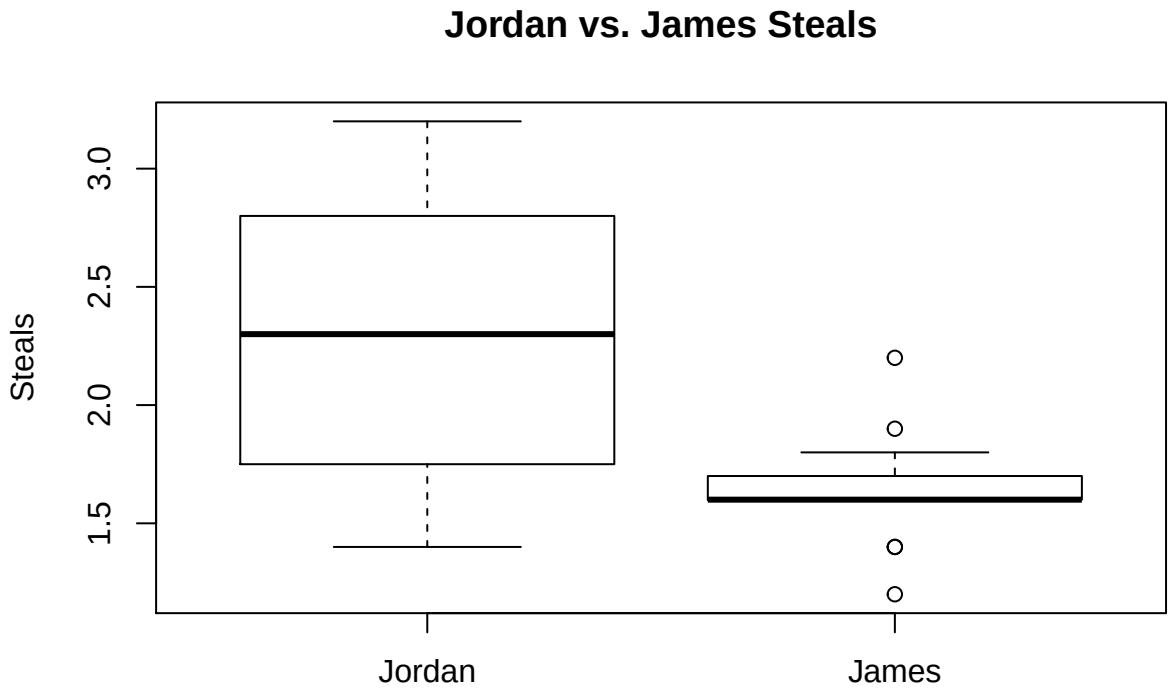
```
## [1] 147.2667
```

```
James_Avg_Pts <- 5 * mean(JordanJames$PTS2)
James_Avg_Pts
```

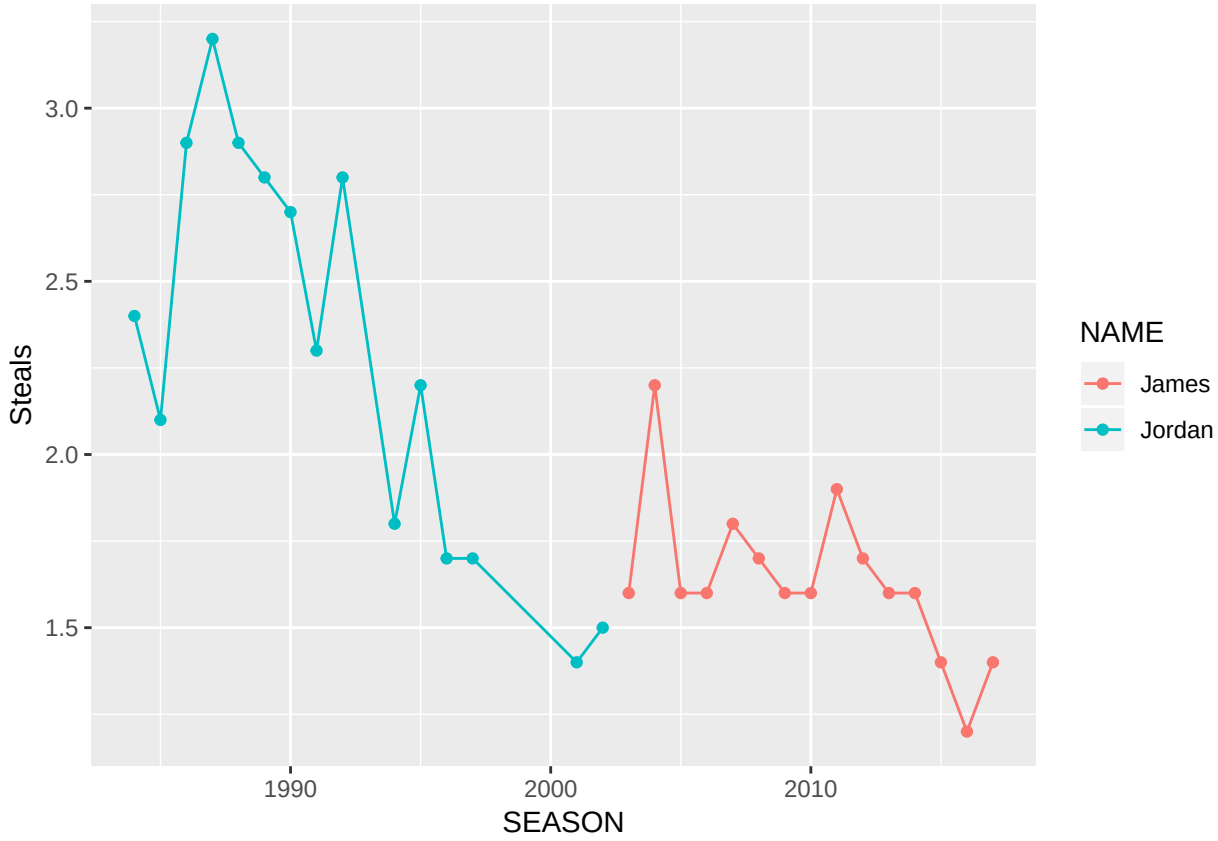
```
## [1] 135.7
```

Above is the box plot of the variable 'points', which demonstrates that Jordan has a wider range of points scored and an ovrall higher average. Where as James has a smaller ranger of points scored with his overall average lower. We decided out of our five variables that points is of the most importance, therefore gave it the highested weighted average, 5.

```
boxplot(JordanJames$STL, JordanJames$STL2, names = c("Jordan", "James"), main = "Jordan vs. James Steals
```




```
ggplot(new) +
  geom_line(aes(x=SEASON, y=STL, col=NAME)) +
  geom_point(aes(x=SEASON, y=STL, col=NAME)) +
  ylab("Steals")
```



```
Jordan_Avg_Stl <- 3 * mean(JordanJames$STL)
Jordan_Avg_Stl
```

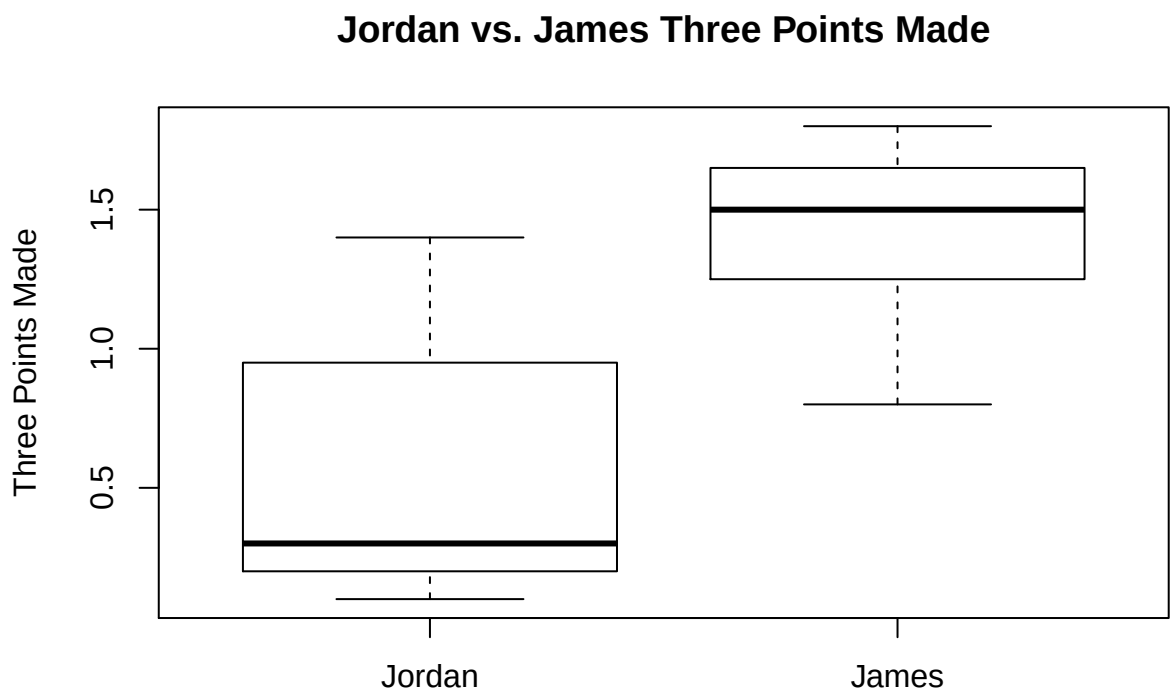
```
## [1] 6.88
```

```
James_Avg_Stl <- 3 * mean(JordanJames$STL2)
James_Avg_Stl
```

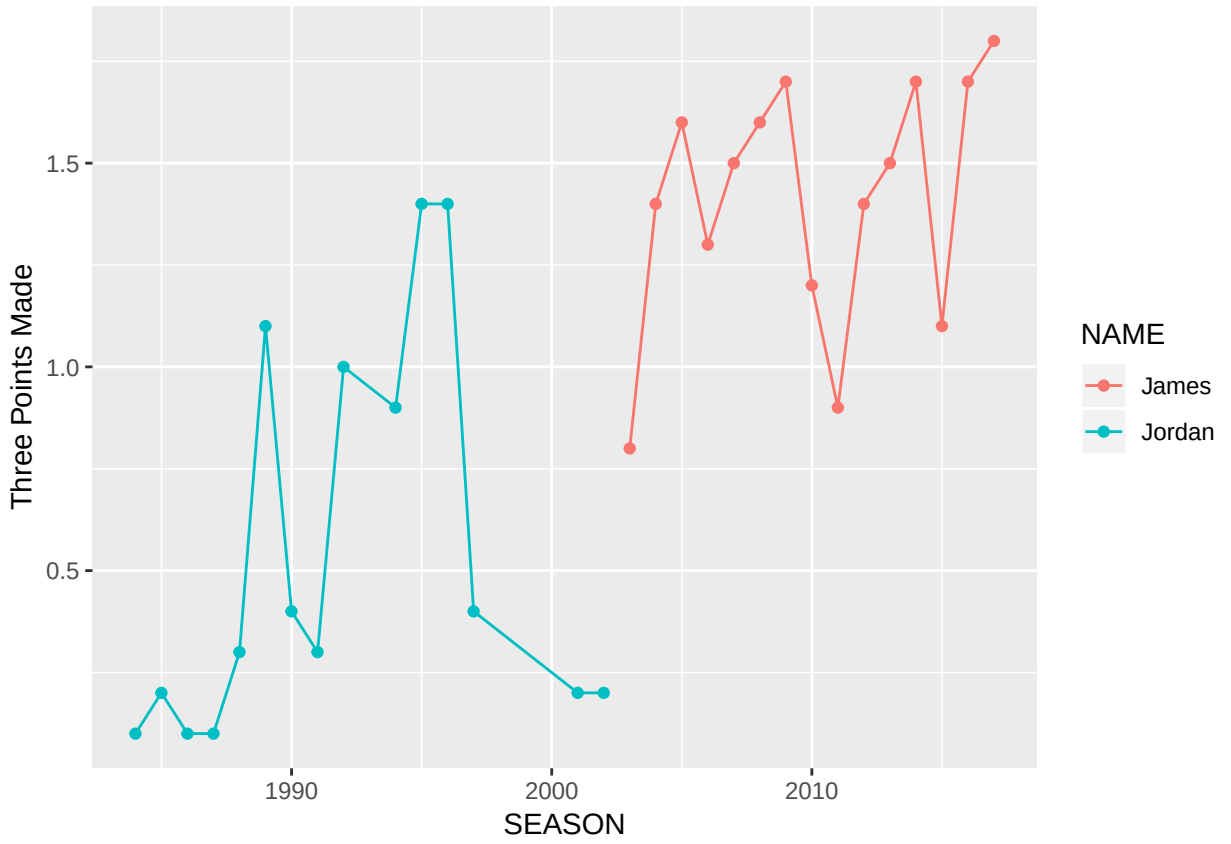
```
## [1] 4.9
```

Above is the box plot of the variable 'steals', which demonstrates that Jordan has an extremely wider range of steals and an ovrall higher average. Where as James has a extremely smaller ranger of steals and a very low average. We decided out of our five variables that steals is of the average importance, therefore gave it the middle weighted average, 3.

```
boxplot(JordanJames$aud, JordanJames$aud2, names = c("Jordan", "James"), main = "Jordan vs. James Three
```




```
ggplot(new) +
  geom_line(aes(x=SEASON, y=aud, col=NAME)) +
  geom_point(aes(x=SEASON, y=aud, col=NAME)) +
  ylab("Three Points Made")
```




```
Jordan_Avg_3pt <- 4 * mean(JordanJames$aud)
Jordan_Avg_3pt
```

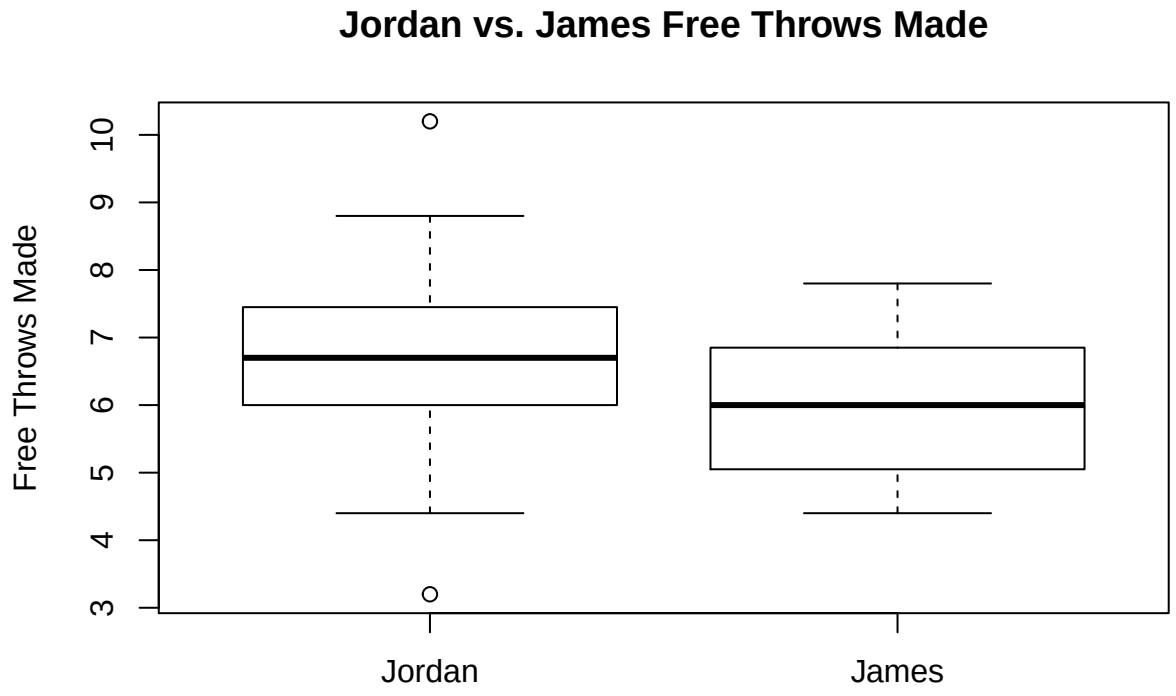
```
## [1] 2.16
```

```
James_Avg_3pt <- 4 * mean(JordanJames$aud2)
James_Avg_3pt
```

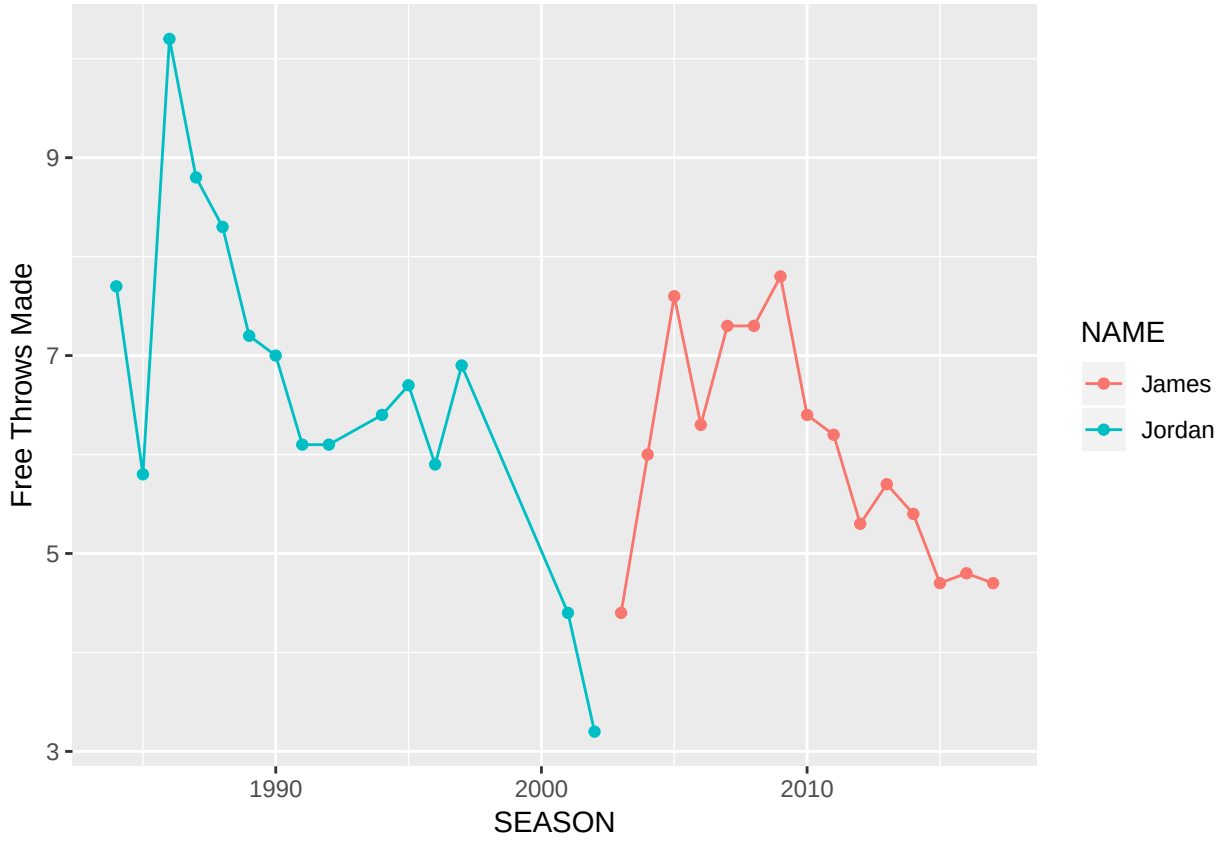
```
## [1] 5.653333
```

Above is the box plot of the variable '3 points made', which demonstrates that James has a smaller range of 3 points made, but the average is significantly higher. Therefore, being the first variable that James is higher than Jordan. Where as Jorden has a wide ranger of 3 points made with a much lower average. We decided out of our five variables that 3 points made is of the second most importance, therefore gave it a weighted average of 4.

```
boxplot(JordanJames$hol, JordanJames$hol2, names = c("Jordan", "James"), main = "Jordan vs. James Free
```




```
ggplot(new) +
  geom_line(aes(x=SEASON, y=hol, col=NAME)) +
  geom_point(aes(x=SEASON, y=hol, col=NAME)) +
  ylab("Free Throws Made")
```



```
Jordan_Avg_FT <- 1 * mean(JordanJames$hol)
Jordan_Avg_FT
```

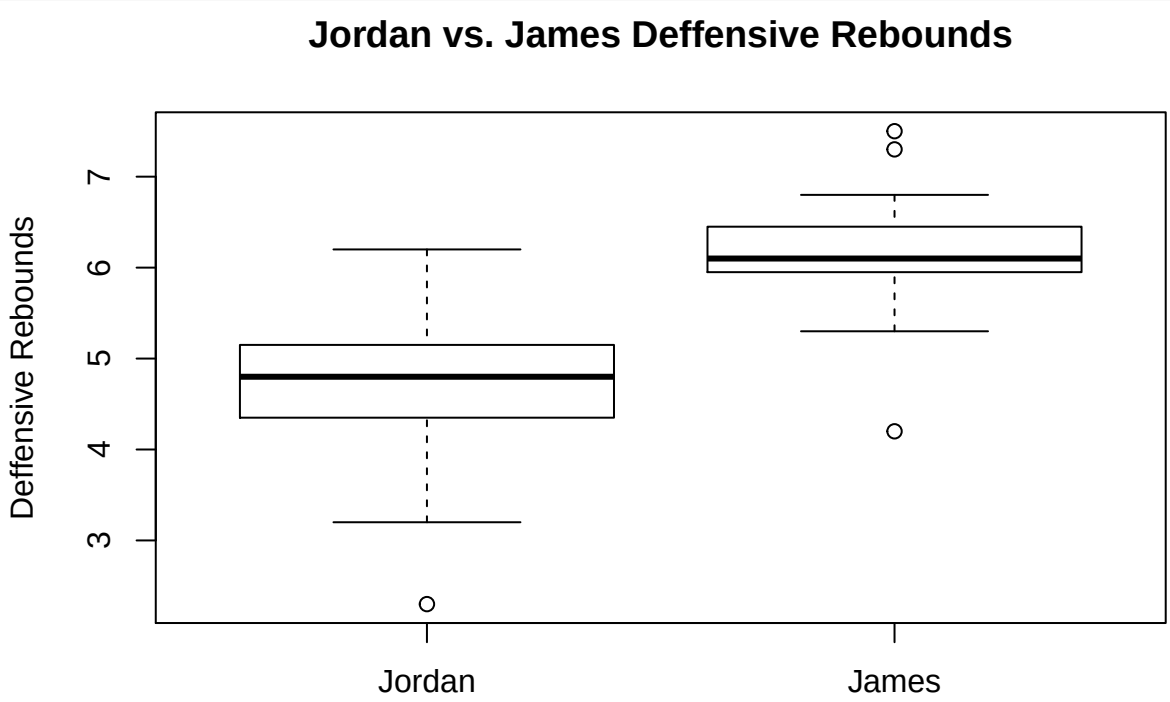
```
## [1] 6.713333
```

```
James_avg_FT <- 1 * mean(JordanJames$hol2)
James_avg_FT
```

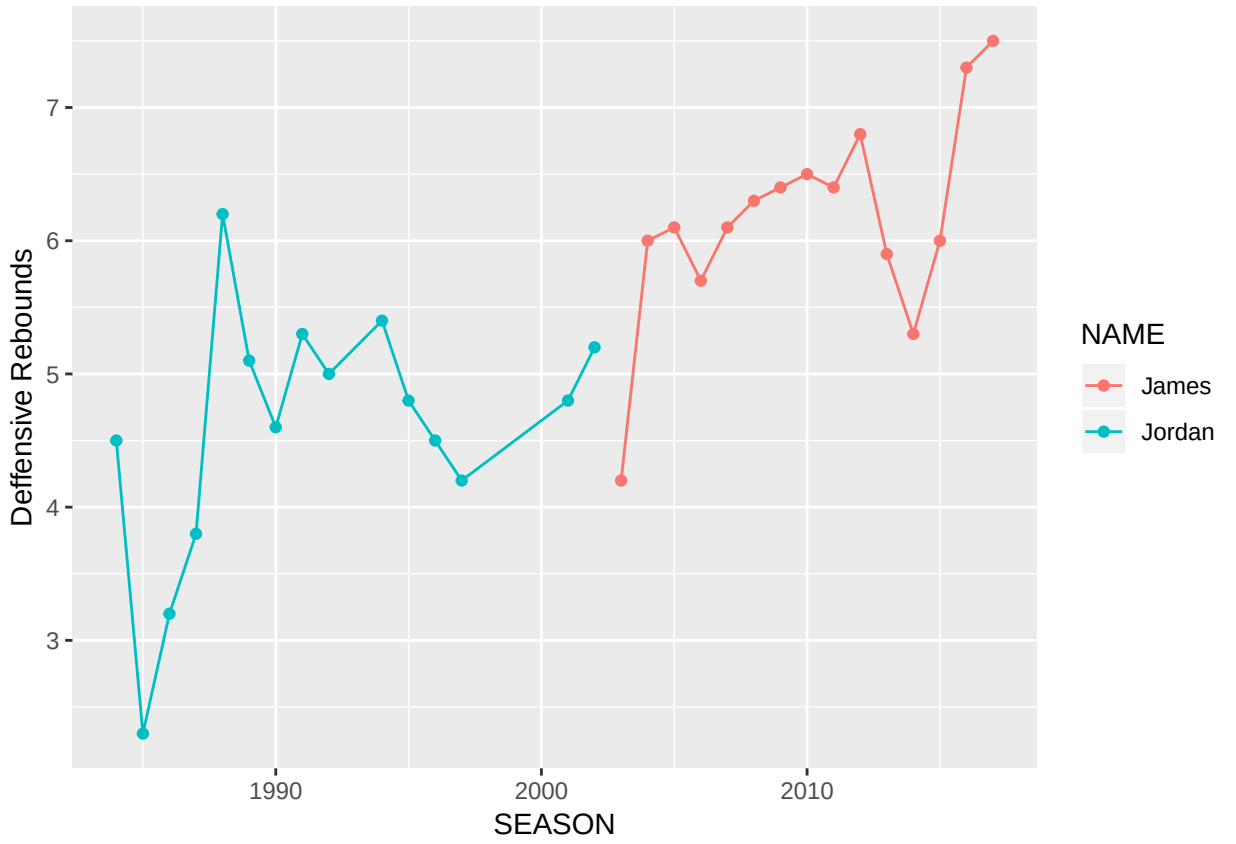
```
## [1] 5.993333
```

Above is the box plot of the variable 'free throws made', which demonstrates that Jordan has a smaller range of free throws made and an ovrall higher average. Where as James has a wider ranger of three throws made with a loweer average. We decided out of our five variables that free throws made is of the lease importance, therefore we gave it the lowest weighted average, 1.

```
boxplot(JordanJames$DREB, JordanJames$DREB2, names = c("Jordan", "James"), main = "Jordan vs. James Defi
```




```
ggplot(new) +
  geom_line(aes(x=SEASON, y=DREB, col=NAME)) +
  geom_point(aes(x=SEASON, y=DREB, col=NAME)) +
  ylab("Deffensive Rebounds")
```




```
Jordan_Avg_Dreb <- 2 * mean(JordanJames$DREB)
Jordan_Avg_Dreb
```

```
## [1] 9.186667
```

```
James_Avg_Dreb <- 2 * mean(JordanJames$DREB2)
James_Avg_Dreb
```

```
## [1] 12.33333
```

Above is the box plot of the variable 'defensive rebounds', which demonstrates that James has a smaller range of defensive rebounds and an ovrall higher average. Therefore, this is the second variable where James' avererage is higher than Jordan's. Where as Jordan has a wider ranger of defensive rebounds with a lower overal average. We decided out of our five variables that defensive rebounds is of the minimal importance, therefore we gave it the second lowest weighted average, 2.

```
Jordan_WA <- (Jordan_Avg_Pts + Jordan_Avg_3pt + Jordan_Avg_FT + Jordan_Avg_Dreb + Jordan_Avg_Stl) / 5
Jordan_WA
```

```
## [1] 34.44133
```

```
James_WA <- (James_Avg_Pts + James_Avg_3pt + James_avg_FT + James_Avg_Dreb + James_Avg_Stl) / 5
James_WA
```

```
## [1] 32.916
```

Above is the overall weighted average of each player. This weighted average contains the combination of each variable with its given weight. The scaling of the weights given to each variable was based off of our idea that offensive skills were of higher weight than defensive. The weights were further decided based off our personal opinion of importance. The overall weighted average states that Michael Jordan's score is 34.44 and Lebron James' score is 32.92. This conculdes that Michael Jordan is the greatest player.