STAT 355

Professor Tim Keaton

NBA G.O.A.T. Debate

Part 1

The NBA has had over 4,500 NBA players, but 2 names have been etched in history as the greatest of all time. **LeBron James** and **Michael Jordan**. These two dominated the NBA through their hard work, sheer dominance, and unbelievable mentality. This has created an ongoing debate between the LeBron followers and the Jordan followers, and this has bled through the culture of the NBA and the world of sports. With this debate between 2 of the greatest of all time, their dominance in the playoffs stands to be the most common, with Jordan's 6 championships and LeBron's 4. The playoffs are a true test of a player's ability to perform under pressure, and both Jordan and James have proven their worth time and time again. I wanted to analyze their respective statistics in their playoff game log to gain insight into one of the sports world's most heated debates. This project is made to reinforce that these claims and debates aren't without basis, **Is the G.O.A.T conversation really that close?**

Data:

G	Date	Series	Tm	Location	Opp	G_Num	Outcome	Score	MP	FG	FGA	FG_Percent
1	2006-04-22	EC1	CLE	H	WAS	1	W	+11	1900-01-02 00:00:00	12	27	0.444
2	2006-04-25	EC1	CLE	H	WAS	2	L	-5	1900-01-01 19:28:00	7	25	0.280
3	2006-04-28	EC1	CLE	A	WAS	3	W	+1	1900-01-01 23:37:00	16	28	0.571
4	2006-04-30	EC1	CLE	A	WAS	4	L	-10	1900-01-01 21:20:00	13	23	0.565
5	2006-05-03	EC1	CLE	H	WAS	5	W	+1	1900-01-01 22:07:00	14	23	0.609

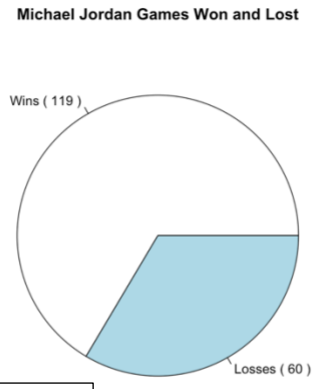
Snippet of data:

3P	3PA	3P_Percent	FT	FTA	FT_Percent	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	GmSc
1	4	0.250	7	11	0.636	3	8	11	11	0	0	4	3	32	23.3
1	6	0.167	11	15	0.733	2	7	9	2	2	3	10	5	26	6.7
3	5	0.600	6	9	0.667	1	4	5	3	2	0	4	3	41	27.4
7	12	0.583	5	7	0.714	1	5	6	5	0	0	7	5	38	23.0
0	1	0.000	17	18	0.944	5	2	7	6	2	0	4	5	45	38.4

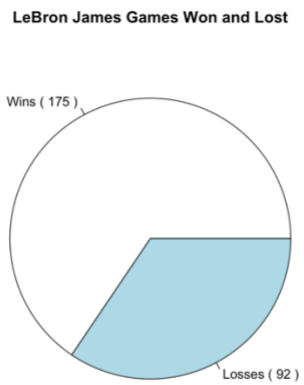
Below is a key for what the rows stand for

G: Game number of the playoff	Date	Series	Tm: Team	Location	Opp: Opponent
G_Num: Game number of the series	Outcome	Score	MP: Minutes Player	FG: Field Goals Made	FGA: Field Goals Attempted
FG_Percent: Field Goal %	3P: 3 Pointers Made	3PA: 3 Pointers Attempted	3P_Percent: 3 Pointer %	FT: Free Throws Made	FTA: Free Throws Attempted
FT_Percent: Free Throw Percentage	ORB: Offensive Rebounds	DRB: Defensive Rebounds	TRB: Total Rebounds	AST: Assists	STL: Steals
BLK: Blocks	TOV: Turnovers	PF: Personal Fouls	PTS: Points	GmSc: Game Score	

EDA:

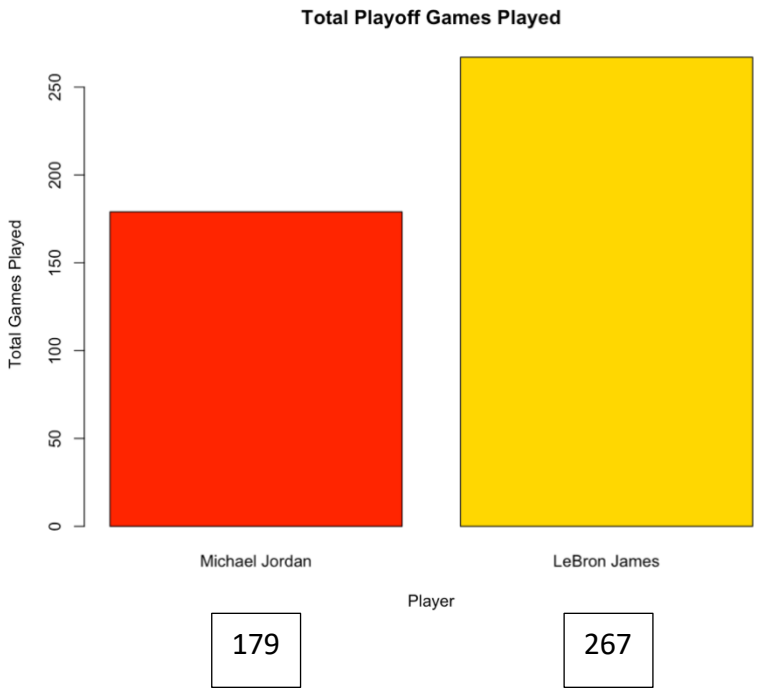


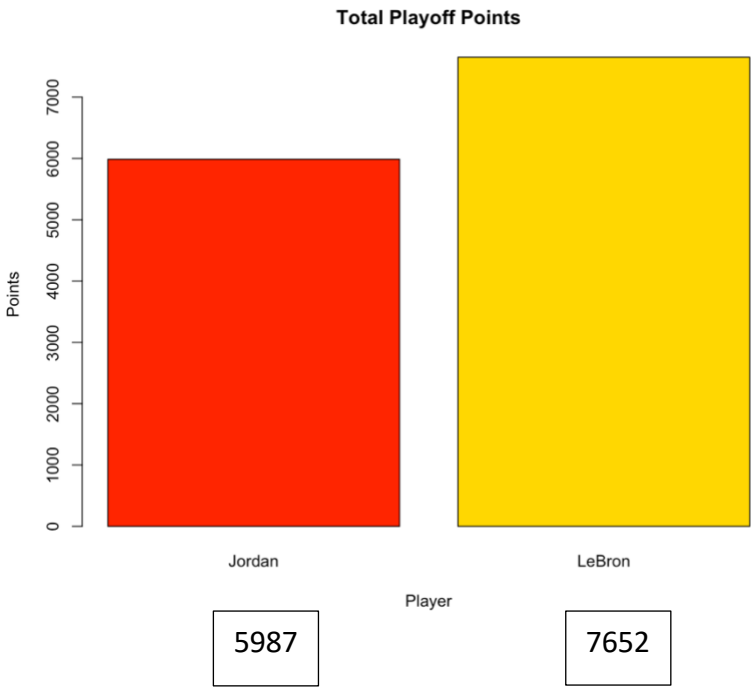
~66.5% Win %



~65.5% Win %

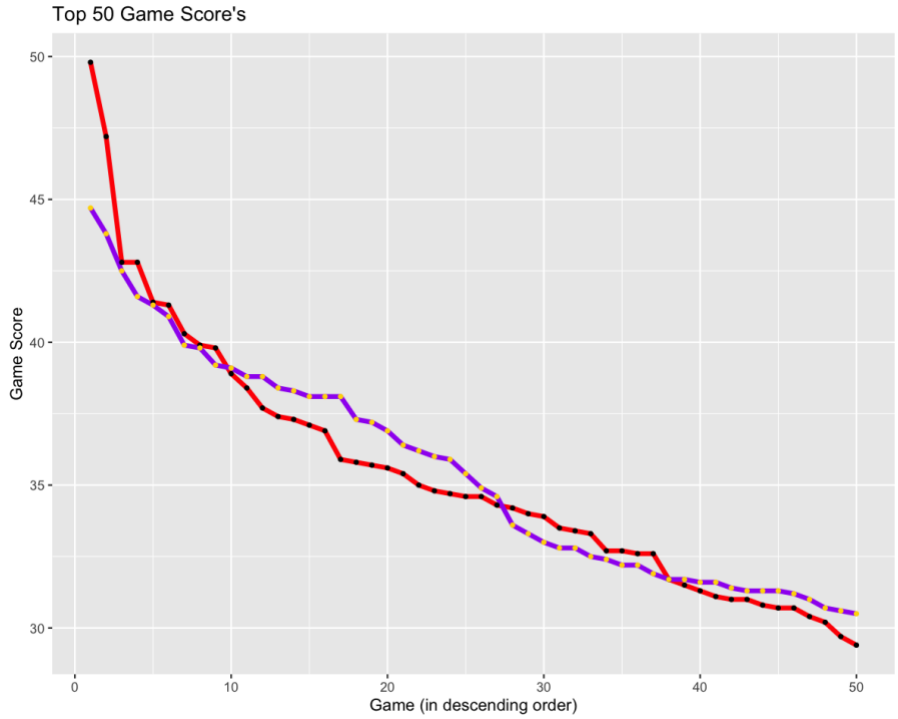
This plot shows how close the winning percentages in the playoffs for the 2 dominant players.





179	267	5987	7652

LeBron has a clear upper hand in playoff games played, even currently playing in it while Jordan is closing that gap in the total points where we can see a smaller difference in total points than total games played.



Red = Jordan
Purple = Lebron

This graph is portraying the top 50 games of Jordan and LeBron's Game Score which is calculated by the following equation:
Game Score = PTS + 0.4 * FG - 0.7 * FGA - 0.4*(FTA - FT) + 0.7 * ORB + 0.3 * DRB + STL + 0.7 * AST + 0.7 * BLK - 0.4 * PF - TOV

This graph is made using a very careful formula made by John Hollinger as a rough measure of a player's performance in a single game.

Part 2

Research Question 1: Does one player have a stronger correlation between his game score and his teams score margin than the other?

This question is one that I was very excited to answer as both players receive heavy criticism from the other's fanbase that they have always needed to play on good teams with other greats to win games. This question would give an answer on who affected their teams to a greater extent. To answer this question, let's break it into smaller parts. Is there a correlation between Jordan's Game Score and his teams score margin. Is there one for LeBron. If there is for both, is there one with a stronger effect on their team's performance?

Let's look at the Statistic:

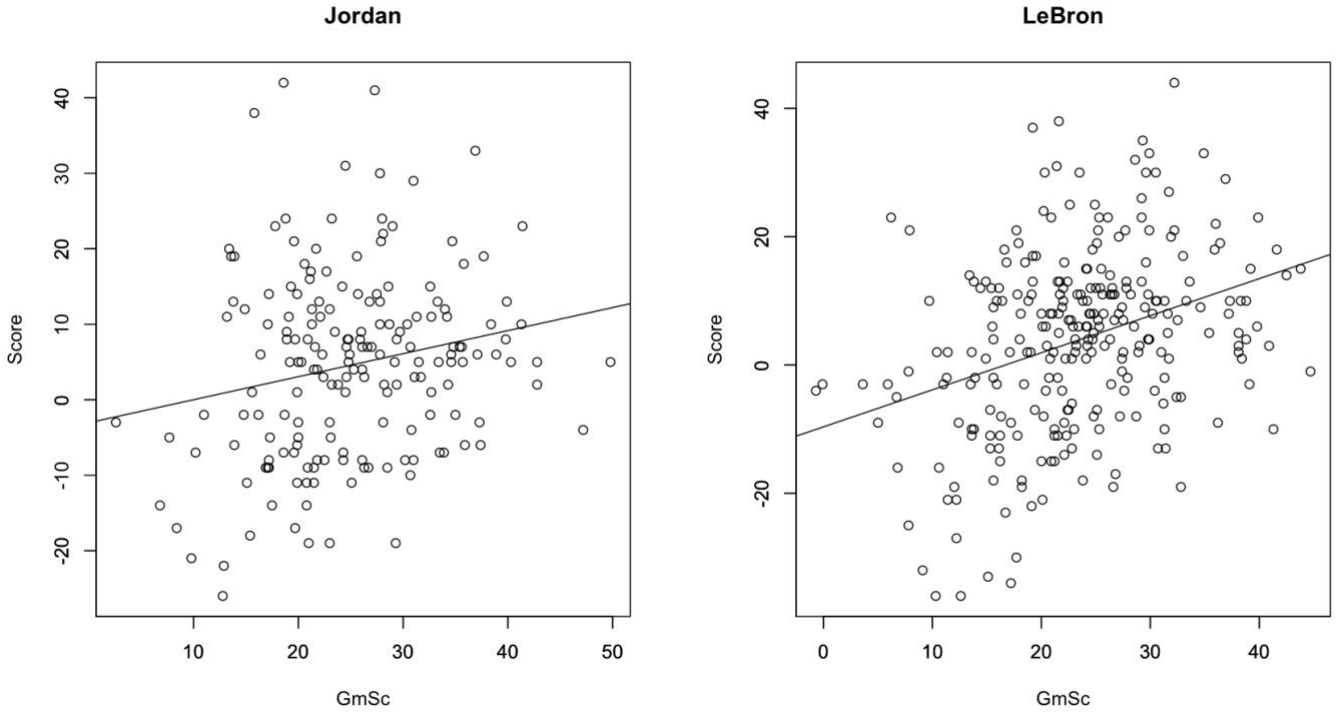
- Game Score in the NBA is a statistic developed by by John Hollinger and is meant to be a rough measure of a player's performance in a single game. The formula is as follows.
    - Game Score = PTS + 0.4 * FG - 0.7 * FGA - 0.4*(FTA - FT) + 0.7 * ORB + 0.3 * DRB + STL + 0.7 * AST + 0.7 * BLK - 0.4 * PF – TOV
- Team Score Margin is the amount a team wins or losses by, "-8" would mean that your team lost by 8 points, "+16" or "16" means that your team won by 16 points.

Let's start by looking at the game score averages for both LeBron and Jordan.

```
> mean(df_jordan$GmSc)
[1] 25.14134
> mean(df_lebron$GmSc)
[1] 23.49101
```

These averages are very similar with Michael Jordan having a slightly greater average. Let's dive into the tests,

We can first look at the data and the trends on a scatterplot with a regression line to visualize the correlation.



According to the plots for both athletes, we seem to have a positive relationship for both, but the data seems to vary fairly frequently allowing me to say that the correlation seems moderate, especially for the Jordan data.

The relationship are as follows:

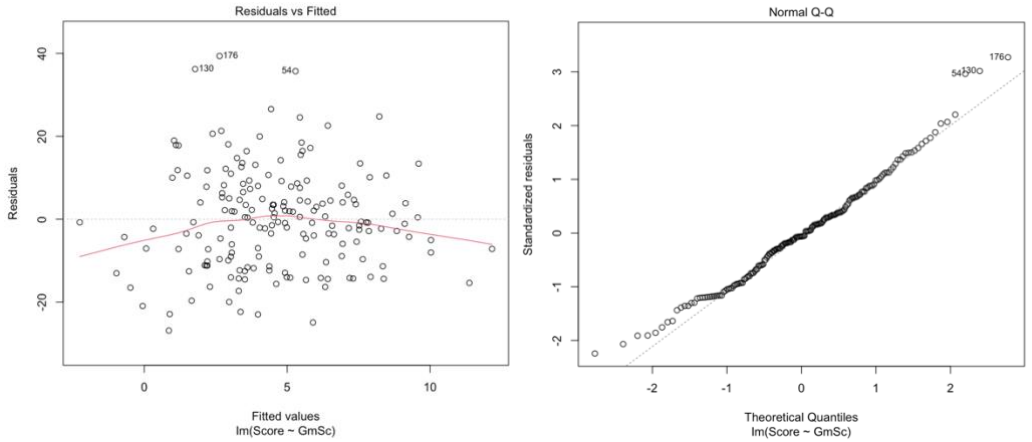
```
> reg_jordan$coef
(Intercept)        GmSc
 -3.0504861   0.3055434
> reg_lebron$coef
(Intercept)        GmSc
 -9.6296280   0.5776551
```

As Jordan's Game Score increases by 1, The score margin during that game increases by 0.3055

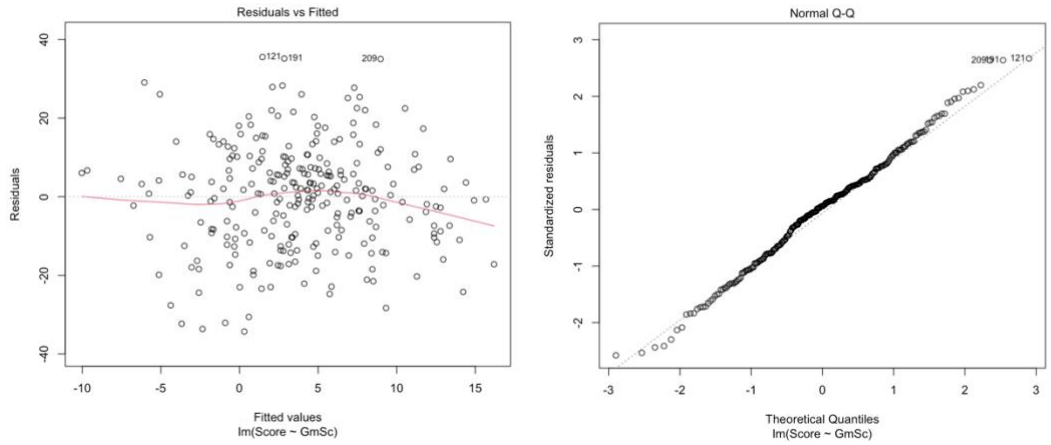
As LeBron's Game Score increases by 1, The score margin during that game increases by 0.5777

Before we run the test we need to make sure out assumptions are valid,

Jordan:



LeBron:



Given that the residual plot displays no significant trend and the fitted line appears relatively compared to the data straight, further testing is necessary to assess how well the data fits and to determine whether the correlation is statistically significant. To do this, we will check normality in the Q-Q plot which seems like it follows a relatively linear path.

Hypothesis Test at $\alpha = .05$

H_0 : There is no significant correlation between the score margin and the player's game score.

H_a: There is a significant correlation between the score margin and the player's game score.

- We will first stat with Jordan's data

```
> cor.test(df_jordan$GmSc, df_jordan$Score)

        Pearson's product-moment correlation

data:  df_jordan$GmSc and df_jordan$Score
t = 2.7592, df = 177, p-value = 0.006404
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.05813166 0.33962888
sample estimates:
      cor
0.2030725
```

  - Since $\boldsymbol{p} = 0.006404 < \alpha = .05$, reject H_0. We have sufficient evidence to conclude that $\beta_1 \neq 0$. As Jordan's Game Score increases, the score margin is also expected to change.

- We will now look at LeBron's data

```
> cor.test(df_lebron$GmSc, df_lebron$Score)

        Pearson's product-moment correlation

data:  df_lebron$GmSc and df_lebron$Score
t = 5.8295, df = 265, p-value = 1.612e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2262467 0.4393984
sample estimates:
      cor
0.3371359
```

  - Since $\boldsymbol{p}$ = 1.612e-08 < α = .05, reject H_0. We have sufficient evidence to conclude that $\beta_1 \neq 0$. As LeBron's Game Score increases, the score margin is also expected to change.

We now know that both of the players have a correlation towards the score margin, but we do not know if there is a significant difference in the amount that they affect their respective teams, so let's compute an ANOVA test that will result us in knowing if there is a significant difference between the correlations of Michael Jordan and LeBron James in their personal Game Score and that game's score.

ANOVA Test at α = .05

H_0 : There is no difference in the mean Score Margin for Michael Jordan and LeBron James and their respective Game Scores

H_a: There is a difference in the mean Score Margin for Michael Jordan and LeBron James and their respective Game Scores

```
> summary(aov(Score ~ GmSc*player, data = df_combined))
             Df Sum Sq Mean Sq F value   Pr(>F)
GmSc          1   6684    6684  40.433 5.08e-10 ***
player        1      1       1   0.005   0.9463
GmSc:player   1    532     532   3.221   0.0734 .
Residuals   442  73063     165
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

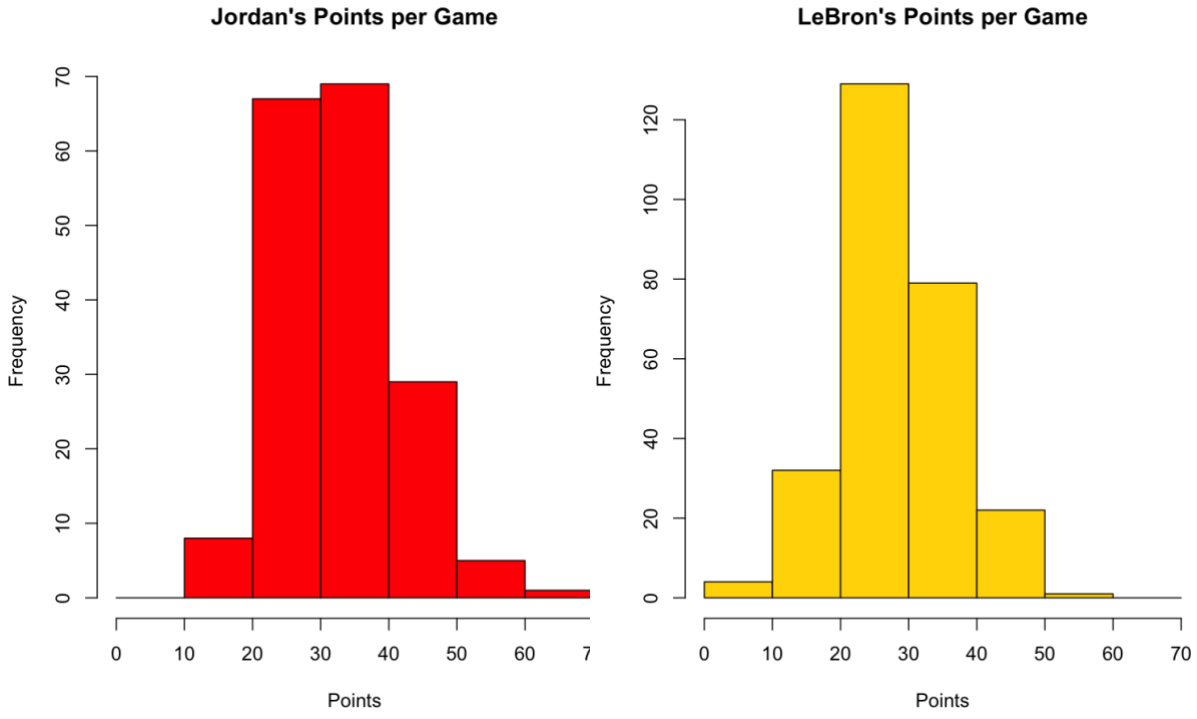
Since $\boldsymbol{p}$ = 0.0734 > α = .05, we fail to reject H_0. we do not have sufficient evidence to reject the null hypothesis that there is no difference in the mean Score for Michael Jordan and LeBron James and their respective Game Scores.

To answer the Research questions, we performed correlation tests to find out if Jordan and LeBron's Game Score affected the score margin of that game and concluded that it did. We then ran an ANOVA test to compare the relationship between personal Game Score and game score margin for Michael Jordan and LeBron James. We received the results that we did not have enough evidence to conclude that one athlete has a stronger correlation than the other.

Research Question 2: Is there a significant difference in the average number of points scored per game by LeBron James and Michael Jordan in the playoffs?

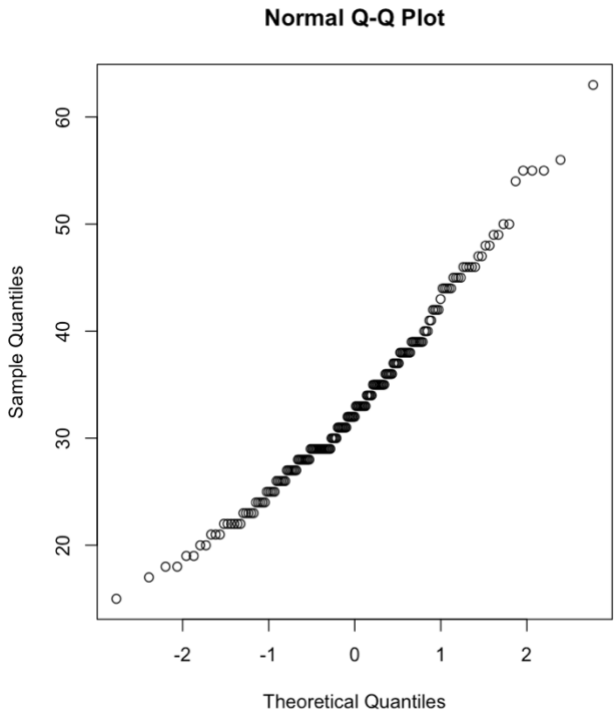
The most basic statistic in deciding who the greatest basketball player of all time is points. Let's compare the amount points that LeBron and Jordan.

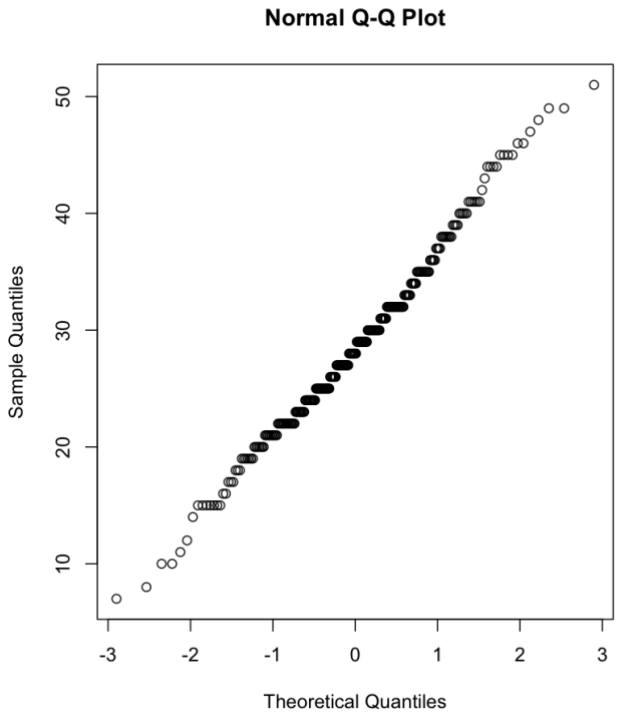
First let's look at a histogram of their total points.



These histograms show that Lebron had proportionally more 30+ points games than LeBron did. It also shows the consistency of Jordan and the longevity of LeBron.

Similar to the last question, we used Q-Q plots to assume normality by the linearity of the quantiles:





Hypothesis Test at $\alpha = .05$

H_0 : There is no significant difference in the mean points between Michael Jordan and LeBron James in the playoffs

H_a: There is a significant difference in the mean points between Michael Jordan and LeBron James in the playoffs

```
> t.test(df_lebron$PTS, df_jordan$PTS, var.equal = FALSE)

        Welch Two Sample t-test

data:  df_lebron$PTS and df_jordan$PTS
t = -5.7992, df = 359.55, p-value = 1.46e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.411336 -3.164167
sample estimates:
mean of x mean of y
 28.65918  33.44693
```

Since $\boldsymbol{p}$ = 0.1.46e-08 < α = .05, we reject H_0. we have sufficient evidence to reject the null hypothesis and conclude that there is a significant difference in the mean points between Michael Jordan and LeBron James in the playoffs

Now that we this information, we know that there is a difference in mean points between MJ and King James, but we do not know if this is because of the players effective scoring or bad shot judgement. We should test the fg% to see if there is a difference in points because one is a better (or worse scorer with a lack of judgement) scorer than the other.

H_0 : There is no significant difference in the mean field goal percentage between Michael Jordan and LeBron James in the playoffs

H_a: There is a significant difference in the mean field goal percentage between Michael Jordan and LeBron James in the playoffs

```
> t.test(df_lebron$FG_Percent, df_jordan$FG_Percent)

        Welch Two Sample t-test

data:  df_lebron$FG_Percent and df_jordan$FG_Percent
t = 1.1811, df = 423.54, p-value = 0.2382
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.00784989  0.03148925
sample estimates:
mean of x mean of y
0.4963895 0.4845698
```

Since $\boldsymbol{p}$ = 0.2382 > α = .05, we fail to reject H_0. We do not have sufficient evidence to reject the null hypothesis and conclude that there is no significant difference in the mean field goal percentage between Michael Jordan and LeBron James in the playoffs.

```
> summary(df_jordan$PTS)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  15.00   27.50   32.00   33.45   39.00   63.00
> summary(df_lebron$PTS)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   7.00   23.00   28.00   28.66   33.50   51.00
```

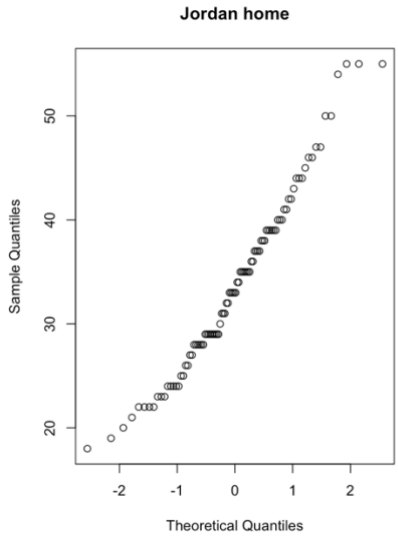
This t-test shows us that there is no difference in the skill level of the person shooting the ball or scoring more, so if we look at the above summary of the data, we can see that Jordan is better

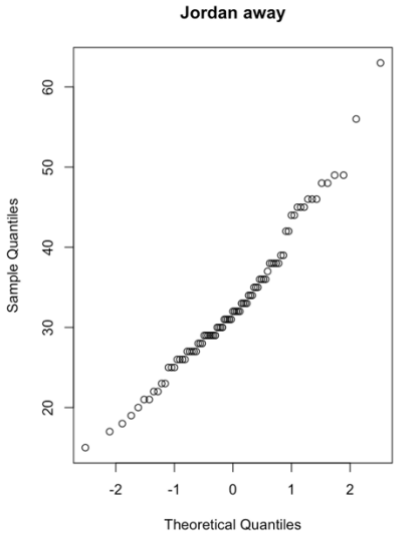
scorer than LeBron in the Playoffs. This is supported by us rejecting the null hypothesis of "no significant difference in the mean field goal percentage between Michael Jordan and LeBron James in the playoffs" from the first t-test.

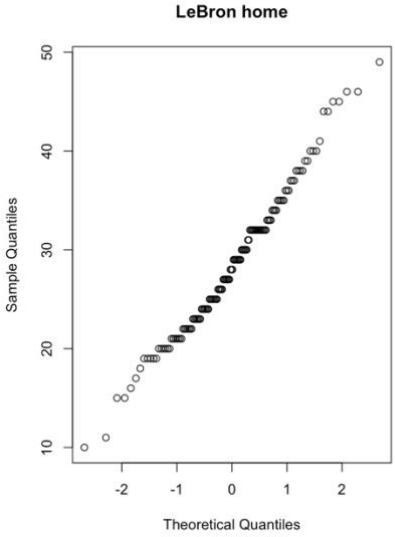
Research Question 3: Is there an affect on the location (home or away) on the variance of the points LeBron or Jordan score.

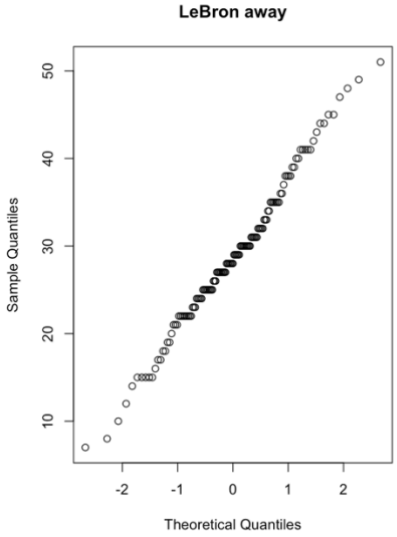
The NBA's home field advantage is said to be the most meaningful of the 4 American sports with teams winning 60% of their home court games. Does this translate to the G.O.A.T.'s? If it does which one gets affected the most?

We can start by plotting Q-Q Plots of the 4 situations to know if we need to do a Wilcoxon test instead of a F-test, LeBron at home, LeBron away, Jordan at home, Jordan away.









These q-q plots are all linear giving us a valid assumption showing that the data is normally distributed.

H_0 : There are equal variance between Jordan's points scored at home and away.

H_a: There are not equal variance between Jordan's points scored at home and away.

```
> var.test(jordan_home$PTS, jordan_away$PTS)

        F test to compare two variances

data:  jordan_home$PTS and jordan_away$PTS
F = 0.95711, num df = 93, denom df = 84, p-value = 0.8344
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6274801 1.4532852
sample estimates:
ratio of variances
          0.957107
```

Since $\boldsymbol{p} = 0.8344 > \alpha = .05$, we fail to reject H_0. We do not have sufficient evidence to reject the null hypothesis and conclude that there are equal variance between Jordan's points scored at home and away.

H_0 : There are equal variance between LeBron's points scored at home and away.

H_a: There are not equal variance between LeBron's points scored at home and away.

```
> var.test(lebron_home$PTS, lebron_away$PTS)

        F test to compare two variances

data:  lebron_home$PTS and lebron_away$PTS
F = 0.75474, num df = 135, denom df = 130, p-value = 0.1061
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5357927 1.0619831
sample estimates:
ratio of variances
          0.7547444
```

Since $\boldsymbol{p}$ = 0.1061 > α = .05, we fail to reject H_0. We do not have sufficient evidence to reject the null hypothesis and conclude that there are equal variance between LeBron's points scored at home and away.

The data shows that there is no evidence to reject the null hypothesis of equal variances for both LeBron James and Michael Jordan, indicating that the variances of points scored by each player in their home and away games are not significantly different.

Data Set: https://www.basketball-reference.com

Jordan + LeBron