Final Project

Dominnic Chant, Victor Yuan, Eric Zakharian, Kristina Kosis

11/10/2018

Question:

The NBA all star game took place already but we want to find a predictable analysis on the outcome of the game with different attributes in basketball. We have taken a statistical approach to this problem and here are the predictions as followed.

Create separate data frames for Team LeBron and Team Stephen

```r
library(ggplot2)
player_stats <- read.csv("NBA5.csv")
team_lebron_names <- c("LeBron James", "Kevin Durant", "Russell Westbrook",
"Kyrie Irving", "Anthony Davis", "Paul George", "Andre Drummond",
"Bradley Beal", "Victor Oladipo", "Kemba Walker",
"Goran Dragic", "LaMarcus Aldridge")
team_stephen_names <- c("James Harden", "DeMar DeRozan", "Stephen Curry",
"Giannis Antetokounmpo", "Joel Embiid", "Kyle Lowry",
"Klay Thompson", "Damian Lillard", "Draymond Green", "Karl-Anthony Towns",
"Al Horford")
starter_names <- c("LeBron James", "Kevin Durant", "Russell Westbrook", "Kyrie Irving",
"Anthony Davis", "James Harden", "DeMar DeRozan", "Stephen Curry",
"Giannis Antetokounmpo", "Joel Embiid")
# Add a description column for all star team
player_stats[, "All_Star_Team"] <- NA
player_stats[player_stats$Player %in% team_lebron_names, "All_Star_Team"] <- "L"
player_stats[player_stats$Player %in% team_stephen_names, "All_Star_Team"] <- "S"
# Add a description column for starter/reserve
player_stats[, "Starter"] <- NA
player_stats[player_stats$Player %in% starter_names, "Starter"] <- "Starter"
player_stats[(!(player_stats$Player %in% starter_names)), "Starter"] <- "Reserve"
# Separate data by team
team_lebron <- player_stats[player_stats$Player %in% team_lebron_names,]
team_stephen <- player_stats[player_stats$Player %in% team_stephen_names,]
```

Team Points per Game (Average, Maximum, Minimum)

```r
# Team LeBron
lebron_PS.G. <- mean(team_lebron$PS.G.)
lebron_total_PS <- lebron_PS.G.*5
# Team Stephen
stephen_PS.G. <- mean(team_stephen$PS.G.)
stephen_total_PS <- stephen_PS.G.*5
# Estimated total points with 5 players per team on the court at all times:
(PS.G._df <- data.frame(Team = c("Lebron", "Stephen"),
```

```
                    PS.G._Avg = c(lebron_PS.G., stephen_PS.G.),
                    PS.G._Total = c(lebron_total_PS, stephen_total_PS)))
```

```
##      Team PS.G._Avg PS.G._Total
## 1  Lebron  23.07500     115.3750
## 2 Stephen  21.62727     108.1364
```

```
ggplot(player_stats, aes(All_Star_Team, PS.G.,
                    color = All_Star_Team)) + geom_boxplot()
```
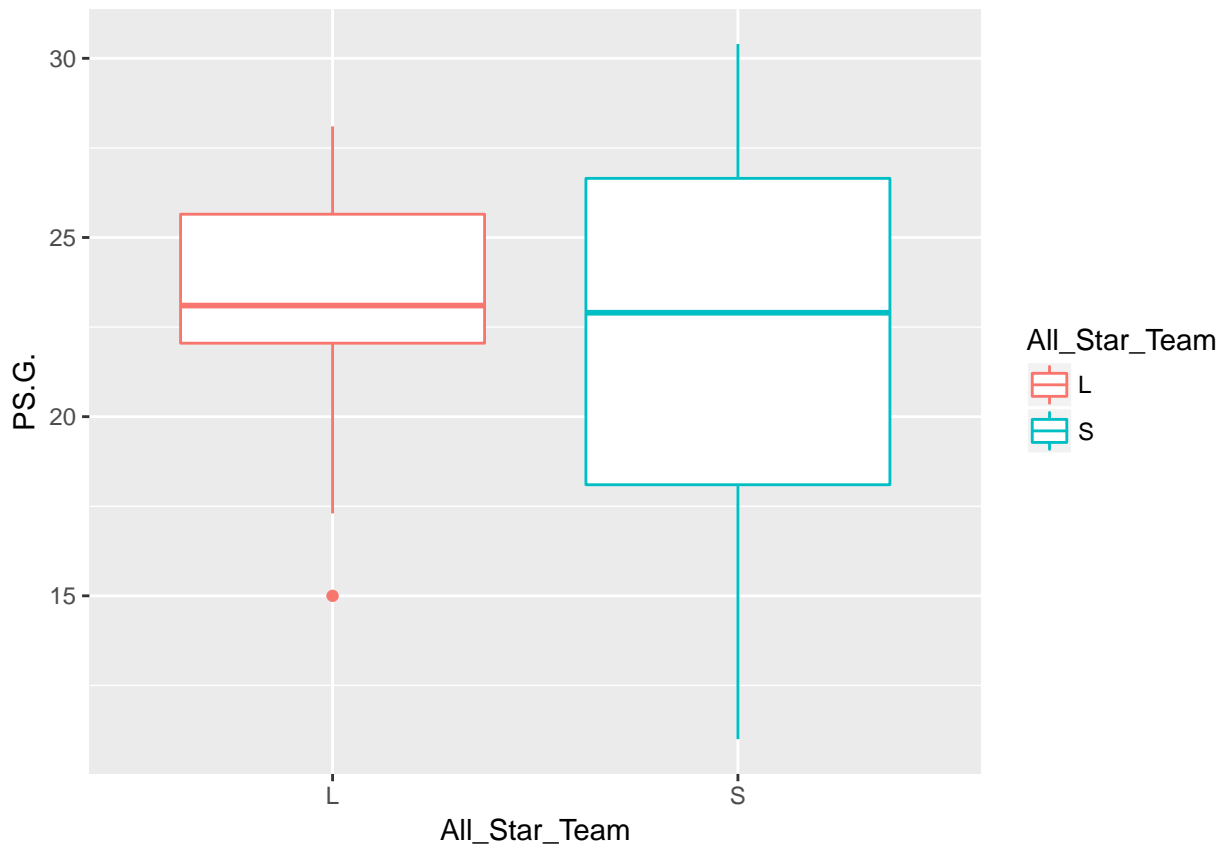

True Shooting Percentage

```
lebron_TS. <- mean(team_lebron$TS.)
stephen_TS. <- mean(team_stephen$TS.)
(TS._df <- data.frame(Team = c("LeBron", "Stephen"),
                    Avg_TS. = c(lebron_TS., stephen_TS.)))
```

```
##      Team   Avg_TS.
## 1  LeBron 0.5798333
## 2 Stephen 0.5988182
```

From the data given, you can see that Stephen's team out pointed Lebron's team in comparison to true shooting percentages. True shooting takes into consideration all types of shooting like, 2-pointers, 3-pointers, and free throws. LeBron's true shooting percentage is 57.98% and Stephen's true shooting percentage is 59.88%.

Team Field Goal Average, Field Goal Attempt Average, and Field Goal Percentage

```
lebron_FG <- mean(team_lebron$FG)
lebron_FGA <- mean(team_lebron$FGA)
lebron_FG_Success <- lebron_FG/lebron_FGA
stephen_FG <- mean(team_stephen$FG)
stephen_FGA <- mean(team_stephen$FGA)
stephen_FG_Success <- stephen_FG/stephen_FGA
FG_df <- data.frame(Team = c("Lebron", "Stephen"),
                    FG_Avg = c(lebron_FG, stephen_FG),
                    FGA_Avg = c(lebron_FGA, stephen_FGA),
                    FG_Success = c(lebron_FG_Success, stephen_FG_Success))
FG_df
```

```
##      Team   FG_Avg  FGA_Avg FG_Success
## 1  Lebron 8.483333 17.46667  0.4856870
## 2 Stephen 7.454545 15.59091  0.4781341
```

After looking at Team Field Goal Average, Field Goal Attempt Average, and Field Goal Success we have concluded that Lebron's team has a higher Team Field Goal Average, Field Goal Attempt Average, and Field Goal Success than Stephen's team has. Lebron's FG Average is 8.48 and Stephen's FG Average is 7.45. Lebron's FGA Average is 17.47 and Stephen's FG Average is 15.59. Lebron's FG Success is 49% and Stephen's FG Success is 48%.

2 Pt vs. 3 Pt Statistics

```
lebron_2P <- mean(team_lebron$X2P)
lebron_2PA <- mean(team_lebron$X2PA)
(lebron_2P_Success <- lebron_2P/lebron_2PA)
```

```
## [1] 0.5259016
```

```
stephen_2P <- mean(team_stephen$X2P)
stephen_2PA <- mean(team_stephen$X2PA)
(stephen_2P_Success <- stephen_2P/stephen_2PA)
```

```
## [1] 0.532539
```

```
lebron_3P <- mean(team_lebron$X3P)
lebron_3PA <- mean(team_lebron$X3PA)
(lebron_3P_Success <- lebron_3P/lebron_3PA)
```

```
## [1] 0.3754386
```

```
stephen_3P <- mean(team_stephen$X3P)
stephen_3PA <- mean(team_stephen$X3PA)
(stephen_3P_Success <- stephen_3P/stephen_3PA)
```

```
## [1] 0.3820225
```

From looking at the shooting statistics, Team Lebron's 2 point success is around 53% while Team Stephen's 2 point success is also around 53%, which means that both Lebron and Stephen have similar 2 point score successes. Additionally, Team Lebron's 3 point success is around 37% while Team Stephen's 3 point success is around 38%. The analysis right here concludes that Stephen's team has a higher 3 point score success than Lebron's 3 point score success.

Calculate Possessions

```r
POSS_function <- function(x){
  FGA <- sum(x$FGA)
  FTA <- sum(x$FTA)
  OREB <- sum(x$ORB)
  TO <- sum(x$TOV)
  0.976*(FGA+0.44*FTA-OREB+TO)
}


POSS_df <- data.frame(Team = c("Lebron", "Stephen"),
                      POSS = c(POSS_function(team_lebron), POSS_function(team_stephen)))
POSS_df
```

```
##      Team     POSS
## 1  Lebron 247.2716
## 2 Stephen 208.7664
```

According to this subset of the data we can see that Lebron's team has the upper hand when it comes to possession of the ball compared to Stephen's team. Statistically speaking, we see that Lebron's team has possession of the ball about **247.27** times a game, while Stephen's team has the ball **208.77** times. This number is extremely higher than what the normal possession number for a team should be, which is about **100-110**, because everyone on these teams are all all stars that usually have the majority of the possessions for their own respective teams. The reason we included this analysis is not to see how many possessions each team gets per game, but to see which team has the more possessions. Team Lebron having more possessions means that they have more chances to score the ball. The formula to find a team's possessions is **0.976(FGA+0.44FTAOREB+TO)**.

Look at the Starter Statistics

```r
# Starter Points Per Game
PS.G._starters <- function(y){
  avg_starter_mins <- mean(y$MP[y$Starter %in% "Starter"])
  perc_starter_mins <- avg_starter_mins/48
  avg_starters_PS.G. <- sum(y$PS.G.[y$Starter %in% "Starter"])
  paste(format(avg_starters_PS.G.*perc_starter_mins, digits = 4), "points")
}


# Starter Rebounds Per Game
TRB_starters <- function(y) {
  avg_starter_mins <- mean(y$MP[y$Starter %in% "Starter"])
  perc_starter_mins <- avg_starter_mins/48
  avg_starters_TRB <- sum(y$TRB[y$Starter %in% "Starter"])
  paste(format(avg_starters_TRB*perc_starter_mins, digits = 4), "rebounds")
}


# Starter Assists Per Game
AST_starters <- function(y) {
  avg_starter_mins <- mean(y$MP[y$Starter %in% "Starter"])
  perc_starter_mins <- avg_starter_mins/48
  avg_starters_AST <- sum(y$AST[y$Starter %in% "Starter"])
  paste(format(avg_starters_AST*perc_starter_mins, digits = 4), "assist")
```

```
}


# Starter Stats
starters_df <- data.frame(Team = c("Lebron", "Stephen"),
                          Points = c(PS.G._starters(team_lebron), PS.G._starters(team_stephen)),
                          Rebounds = c(TRB_starters(team_lebron),TRB_starters(team_stephen)),
                          Assists = c(AST_starters(team_lebron),AST_starters(team_stephen)))
starters_df
```

```
##      Team        Points        Rebounds        Assists
## 1  Lebron 96.71 points 29.64 rebounds 23.63 assist
## 2 Stephen 90.88 points 24.89 rebounds 19.71 assist
```

Lebron's starters slightly beat Stephen's starters in all three categories.

Look at the Reserve Statistics

```
# Reserve Points Per Game
PS.G._reserve <- function(y){
  avg_reserve_mins <- mean(y$MP[y$Starter %in% "Reserve"])
  perc_reserve_mins <- avg_reserve_mins/48
  avg_reserve_PS.G. <- sum(y$PS.G.[y$Starter %in% "Reserve"])
  paste(format(avg_reserve_PS.G.*perc_reserve_mins, digits = 4), "points")
}

# Reserve Rebounds Per Game
TRB_reserve <- function(y) {
  avg_reserve_mins <- mean(y$MP[y$Starter %in% "Reserve"])
  perc_reserve_mins <- avg_reserve_mins/48
  avg_reserve_TRB <- sum(y$TRB[y$Starter %in% "Reserve"])
  paste(format(avg_reserve_TRB*perc_reserve_mins, digits = 4), "rebounds")
}

# Reserve Assists Per Game
AST_reserve <- function(y) {
  avg_reserve_mins <- mean(y$MP[y$Starter %in% "Reserve"])
  perc_reserve_mins <- avg_reserve_mins/48
  avg_reserve_AST <- sum(y$AST[y$Starter %in% "Reserve"])
  paste(format(avg_reserve_AST*perc_reserve_mins, digits = 4), "assist")
}


# Reserve Stats
reserve_df <- data.frame(Team = c("Lebron", "Stephen"),
                         Points = c(PS.G._reserve(team_lebron), PS.G._reserve(team_stephen)),
                         Rebounds = c(TRB_reserve(team_lebron),TRB_reserve(team_stephen)),
                         Assists = c(AST_reserve(team_lebron),AST_reserve(team_stephen)))
reserve_df
```

```
##      Team        Points        Rebounds        Assists
## 1  Lebron 103.6 points 33.57 rebounds 19.64 assist
## 2 Stephen 76.34 points 29.04 rebounds 21.43 assist
```

With the reserves, Lebron dominates Stephen in points and also has a slight edge in rebounds, but Stephen surpasses Lebron in the assists.

Summary

In summary, Stephen Curry's team has the better shooting percentages, but Lebron James' teams simply score more baskets at a worse rate. Lebron's Team is able to do this with more possessions than Curry's Team. We also analyzed the starters against each other and the reserves against each other. Lebron's team dominated in both the starters vs starters and reserves vs reserves.

Prediction: Team Lebron

Actual Outcome: Team Lebron