ESPN NBA Player Data Analysis

Group 5:
Manpreet Dhindsa, Amber Curran, Robel Semunegus

Introduction

- Objective/Goal
  - Identify historical trends to determine how basketball has changed over time in the NBA
  - Wanted to answer questions that were relevant to NBA discussions from a data perspective: “As players like Steph Curry took over the league, how has the 3 point shot changed over time?” and “Has LeBron James been snubbed for the MVP award during his career?”
  - Determine the top players based on certain statistical categories in the dataset

Data

- Top 50 NBA players based on Points per Game (PPG) from the 2001-2002 to 2019-2020 season from ESPN.com
- Allows us to see many statistics of the Top 50 players throughout several seasons such as position, minutes played, points made, three point field goals made, rebounds, assists, and several more
- Allows us to analyze how trends have potentially evolved over the years, how player rankings can be impacted by different statistics, and much more

How we obtained the data

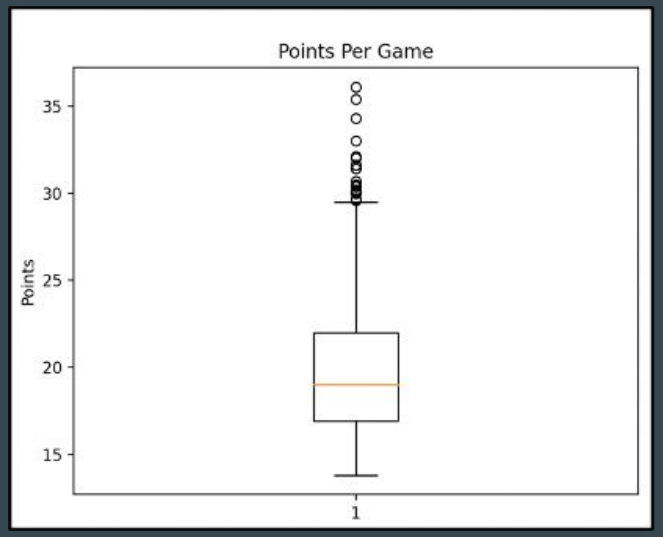
- Obtained this data by web scraping ESPN.com using the BeautifulSoup library
- Because it was for several years rather than one year, this involved developing a for loop to web scrape each web page for every year
- Then, several for loops were used to create appropriate lists of the different players in each season along with their specific statistics

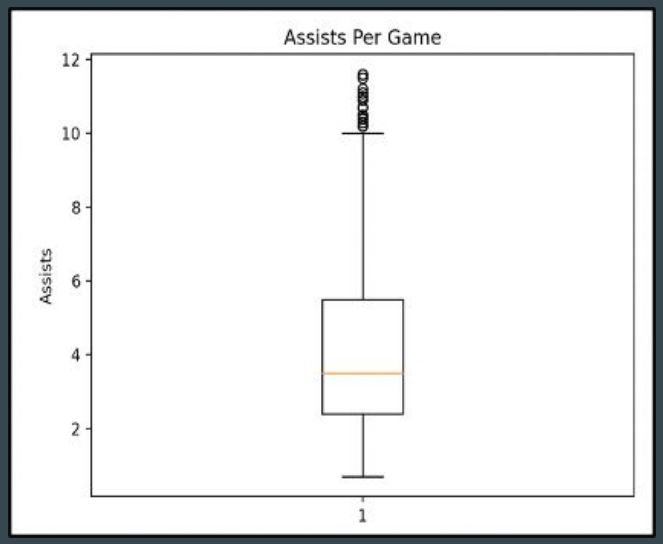
Top-5 Starting Lineup Analysis

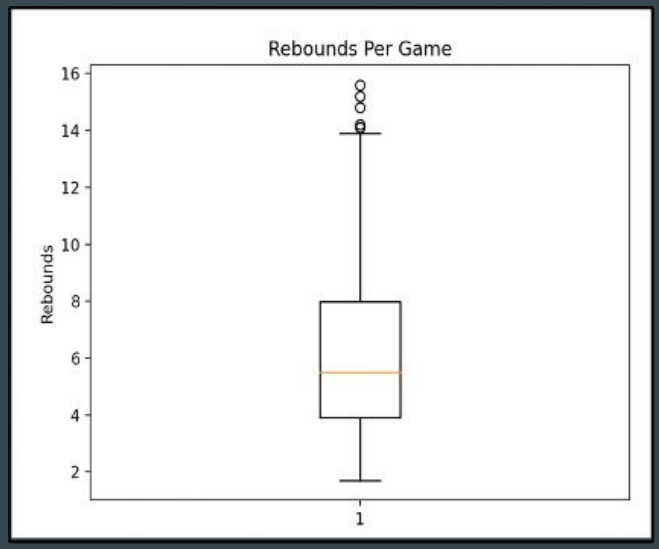
Top-5 Starting Lineup: Experimental Design

- Question: “Based on certain categories, who would be in our Top 5 starting lineup?”
  - In the NBA community, a common discussion topic is ranking the top players in the league
  - Based on the data set, we wanted to analyze which players were the best among the categories
- Our analysis was based on the best player in the following statistical categories:
  - Points Per Game (PPG)
  - Assists Per Game (APG)
  - Rebounds Per Game (RPG)
  - Blocks Per Game (BPG)
  - Player Efficiency Rating (PER)
- Using the above statistical categories allowed selection of players with a variety of skill sets
  - Some categories had multiple players at the top of an observed category; selected first player in the data set that had the top statistic (i.e. earliest season)
- Wrote several for-loops to find the maximum value in each category
  - Used the max value location as an index to find the associated player’s name, position, and season
- Developed box plots to show the distribution of values in each selected category

Top-5 Starting Lineup: Visualizations

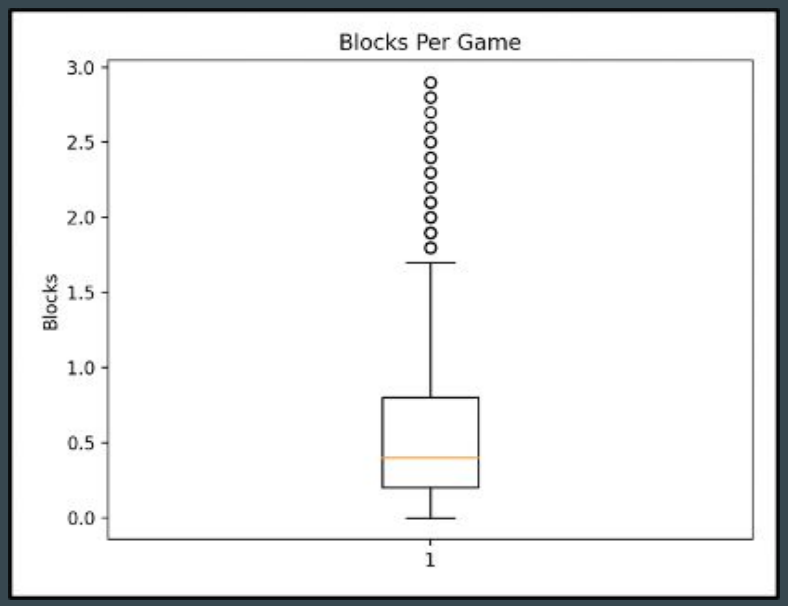


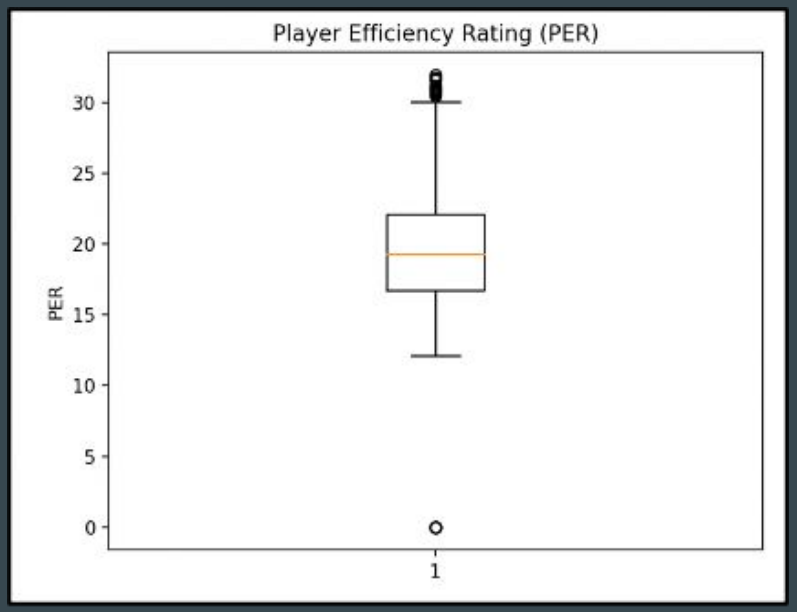




- Goal was to determine the player who had the observation at the maximum value
- Box plots demonstrate distribution and median of values for each statistical category
  - Shows outliers within the data set; outliers represent the top players in the categories

Top-5 Starting Lineup: Visualizations





- Blocks had the smallest range of values of the 5 observed categories
    - Demonstrates that blocks is probably a difficult skill for NBA players
- PER was the only observed category that had outliers below the 25% quantile
    - Outliers due to PER not being a recorded statistic in 2001-2002 season

Top-5 Starting Lineup: Results

- Table below provides results of the Top-5 starting lineup analysis
  - Players were represented as the data point with the largest value in the box plot
- Good coverage of different skill sets (offensive and defensive) in the lineup
  - Centers (2), Power Forwards (1), Shooting Guards (1), Point Guards (1)
- Four of the five players have won the NBA Most Valuable Player (MVP) award
  - 3 of those players have won the MVP award multiple times
- Results provide information to avid NBA fans, reporters, and players
  - Demonstrates who succeeded the most in major statistical categories in the past two decades
  - Provides some insight into how some of the players were able to win the MVP

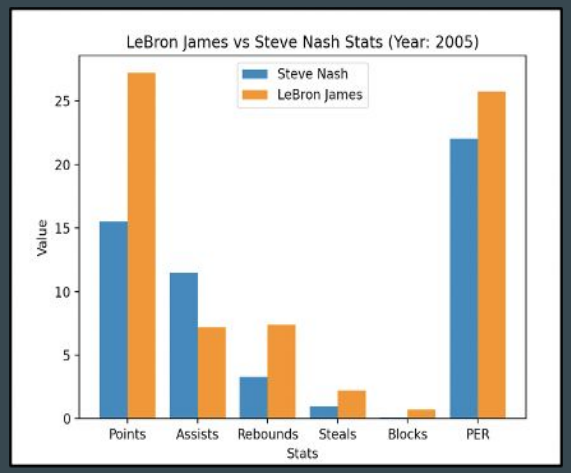
Player Name	Top Statistical Category	Value
James Harden	Points	36.1
Andre Drummond	Rebounds	15.6
Steve Nash	Assists	11.6
Tim Duncan	Blocks	2.9
Giannis Antetokounmpo	Player Efficieny Rating	31.94

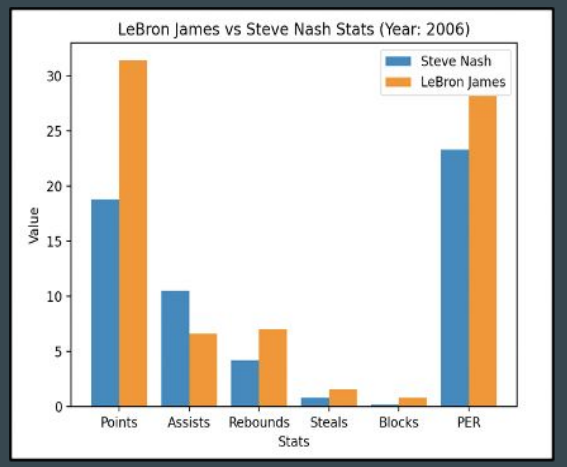
LeBron James vs MVP Analysis

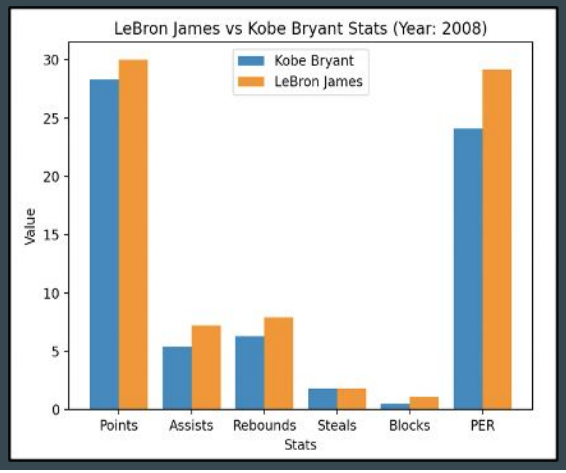
LeBron James vs MVP: Experimental Design

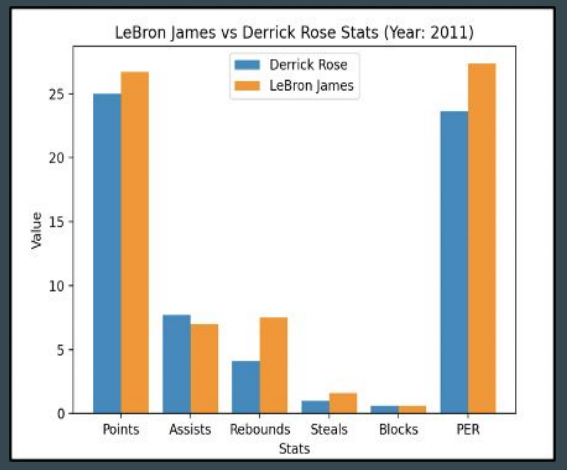
- Question: “Should LeBron James have won more Most Valuable Player (MVP) awards?”
  - In the NBA community, many people believe that LeBron James has been snubbed multiple times for MVP
  - Looking at certain statistical categories, we wanted to determine how LeBron’s stats compared to each MVP
- Analyzed the following statistical categories between LeBron James and each MVP:
  - Points Per Game (PPG)
  - Assists Per Game (APG)
  - Rebounds Per Game (RPG)
  - Steals Per Game (SPG)
  - Blocks Per Game (BPG)
  - Player Efficiency Rating (PER)
- Analysis ignored seasons where LeBron James was not in the league or was the MVP
  - Additionally, LeBron James does not appear in 2018-2019 season; did not participate in 70% of games to be included in the list
- Sorted results into 3 categories: Years where LeBron was snubbed, Years where LeBron and the MVP had similar statistics, and Years where the MVP was a “runaway” winner
- Developed a for-loop to compare observed categories between LeBron James and the MVP
  - Used bar plots to compare statistical categories between LeBron James and the MVP

LeBron James vs MVP: LeBron Performed Better - Visualizations







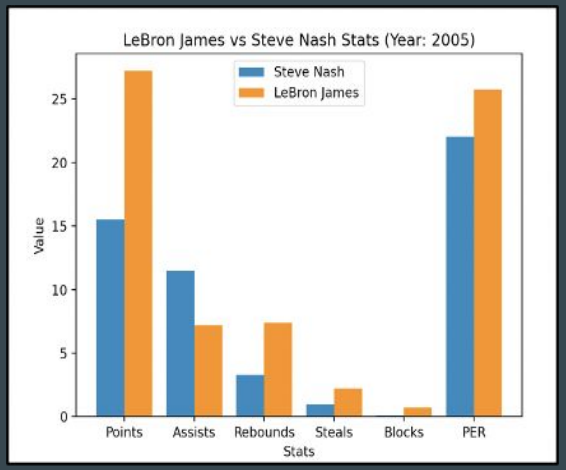


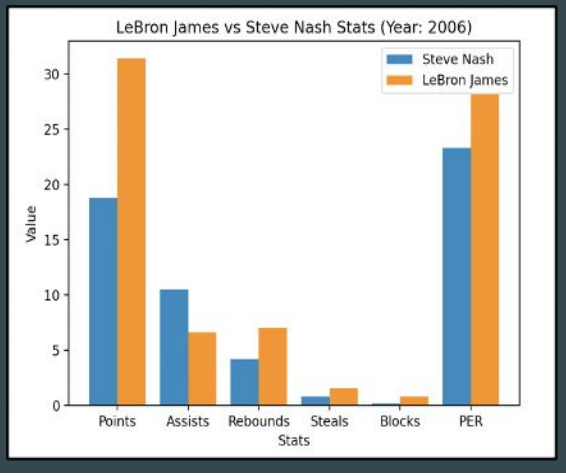
LeBron James vs MVPs: Steve Nash - Results

- Bar plots compare various statistics between LeBron James and Steve Nash in the 2004-2005 season and 2005-2006 season, respectively

- In the two years that Steve Nash won the MVP, LeBron James outperformed him in all observed statistical categories except assists

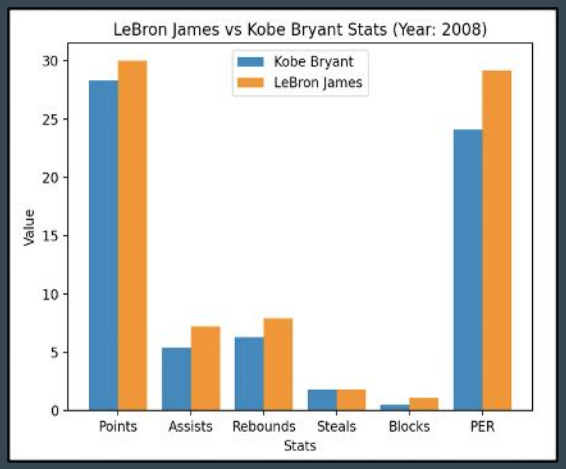
- James' PER is higher than Nash's in both seasons; indicates that James overall did better statistically

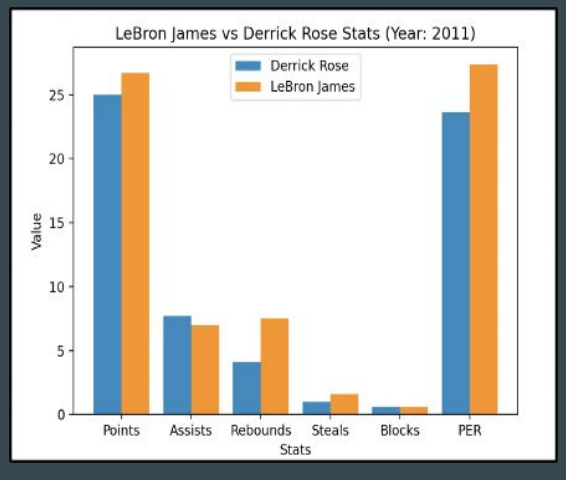




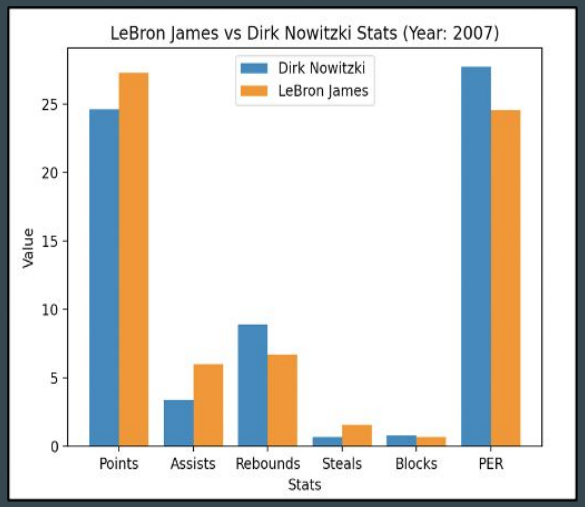
LeBron James vs MVPs: Kobe Bryant & Derrick Rose - Results

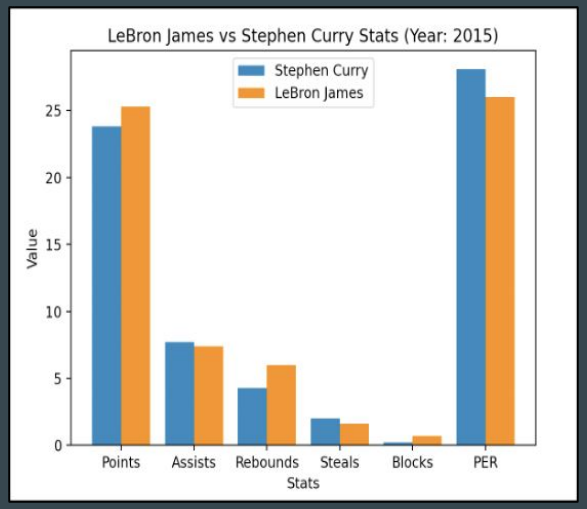
- Bar plots compare the observed statistics LeBron James to Kobe Bryant (in the 2007-2008 season) and Derrick Rose (in the 2010-2011 season)
- In the 2007-2008 season, LeBron James outperformed Kobe Bryant in all observed statistical categories except steals (averaged the same number of steals per game)
- In the 2010-2011 season, Derrick Rose only outperformed LeBron James in assists and averaged the same number of blocks
- LeBron's PER was higher than both players indicating that he performed better overall statistically in several different categories

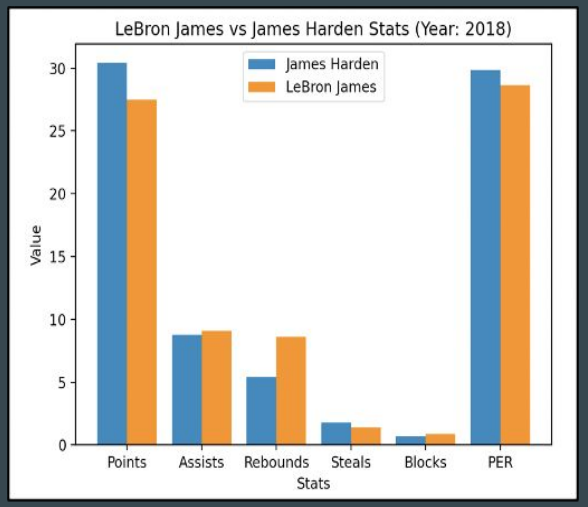




LeBron James vs MVP: Similar Statistics - Visualizations

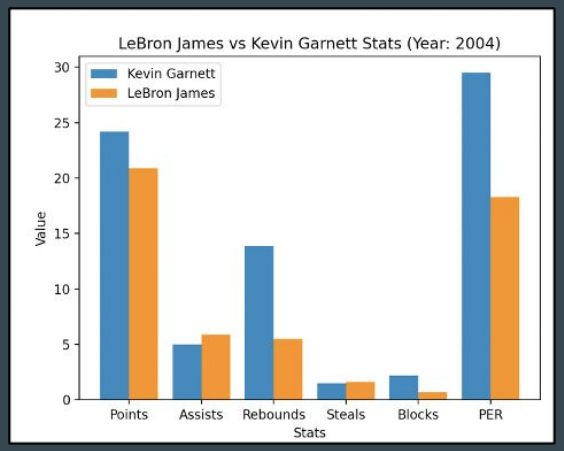


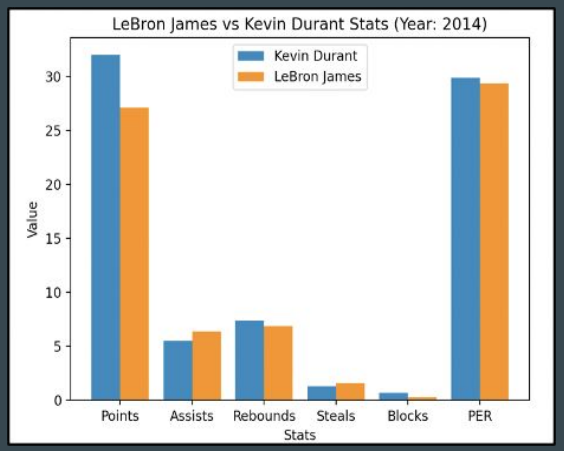


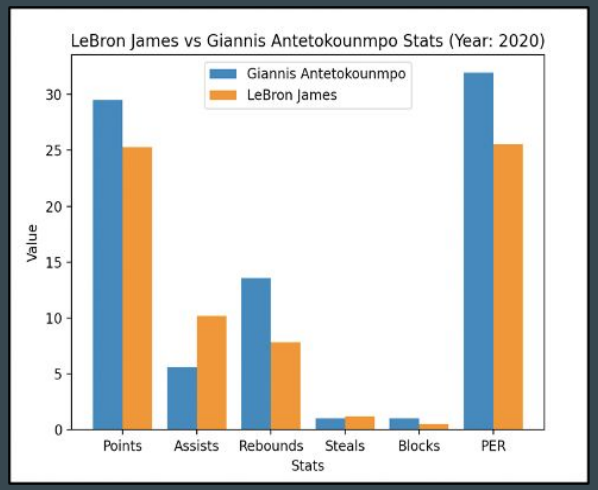


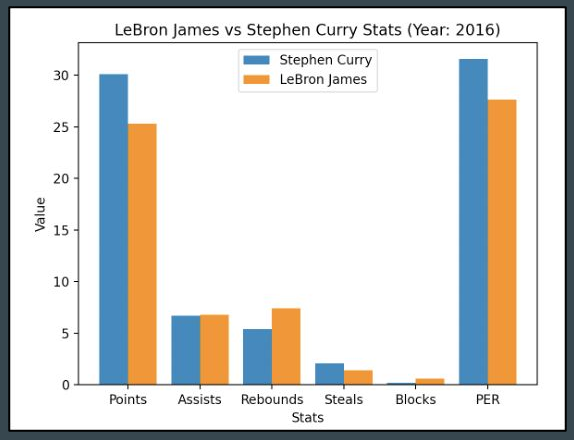
- The MVPs in 2006-2007, 2015-2016, and 2017-2018 seasons did not drastically perform better than LeBron James in the observed statistical categories
  - MVPs outperformed LeBron James but not as overwhelmingly than the “runaway” MVPs
- PER of the MVPs in these seasons were all higher than LeBron James’; indicates, statistically, the MVPs had a much better overall season performance than LeBron James

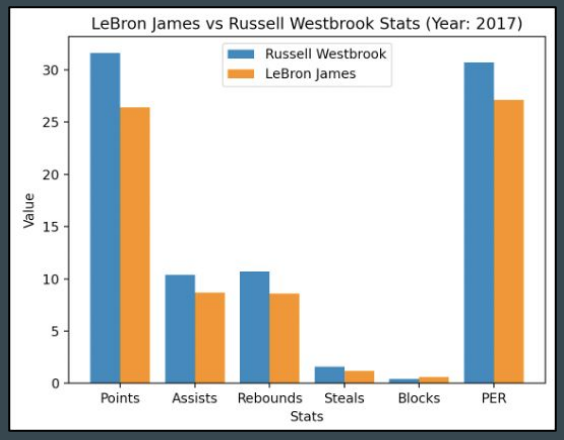
LeBron James vs MVP: “Runaway” MVPs - Visualizations







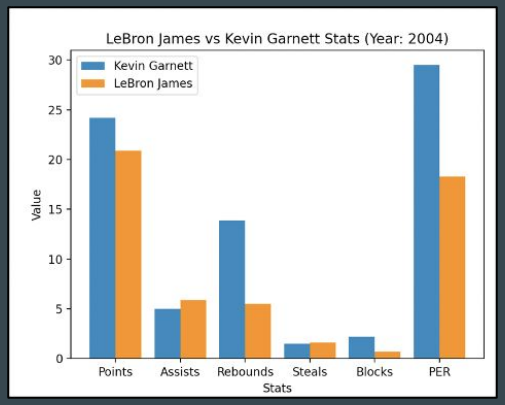


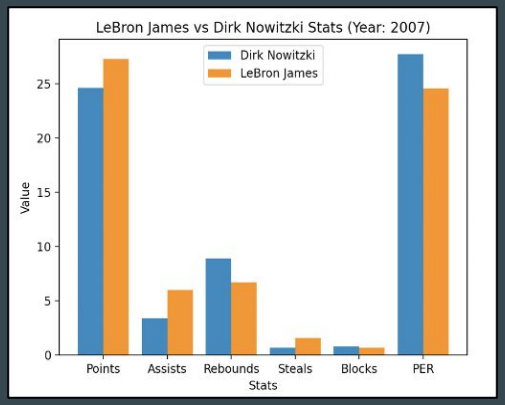


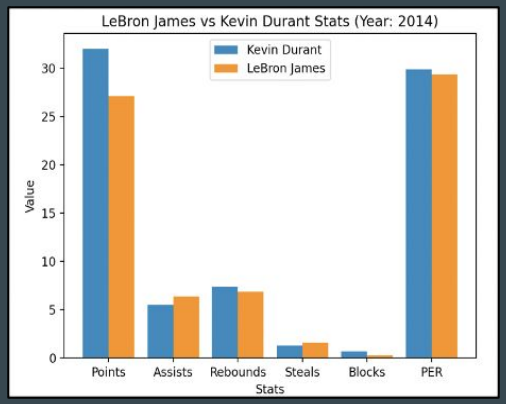
LeBron James vs MVP: “Runaway” MVPs - Results

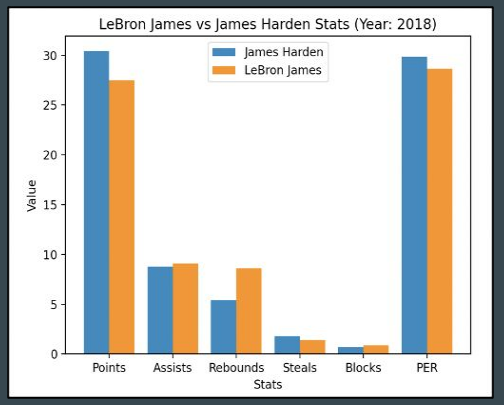
- Based on our analysis, LeBron James was outperformed in the following seasons:
  - 2004-2005, 2013-2014, 2015-2016, 2016-2017, and 2019-2020 seasons
  - The MVP of each season outperformed LeBron James in several observed statistical categories
- All “runaway” MVPs averaged more points per game (PPG) than LeBron James
  - Based on our analysis, the years LeBron James was snubbed for an MVP, he averaged a higher PPG
- All “runaway” MVPs had a higher PER than LeBron James indicated that their overall performance, statistically, was better
  - Based on our analysis, the years LeBron James was snubbed for an MVP, he had a higher PER

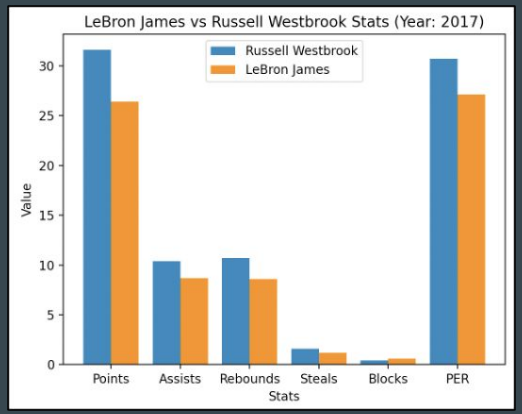
LeBron James vs MVP: Non-Snubbed Years - Visualizations

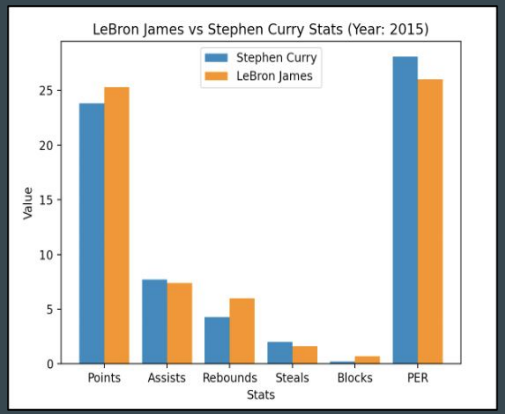


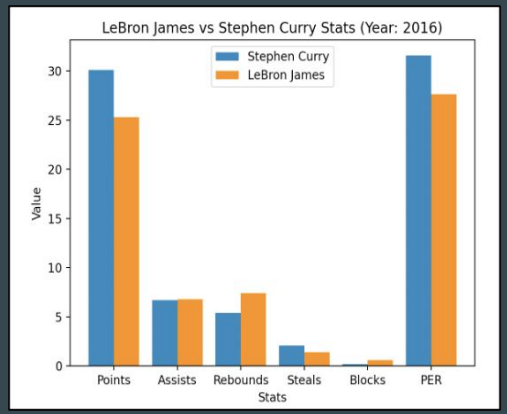


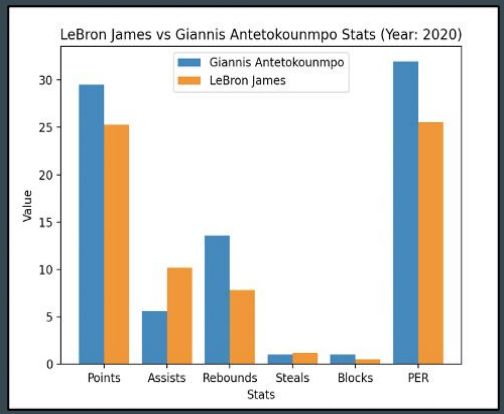










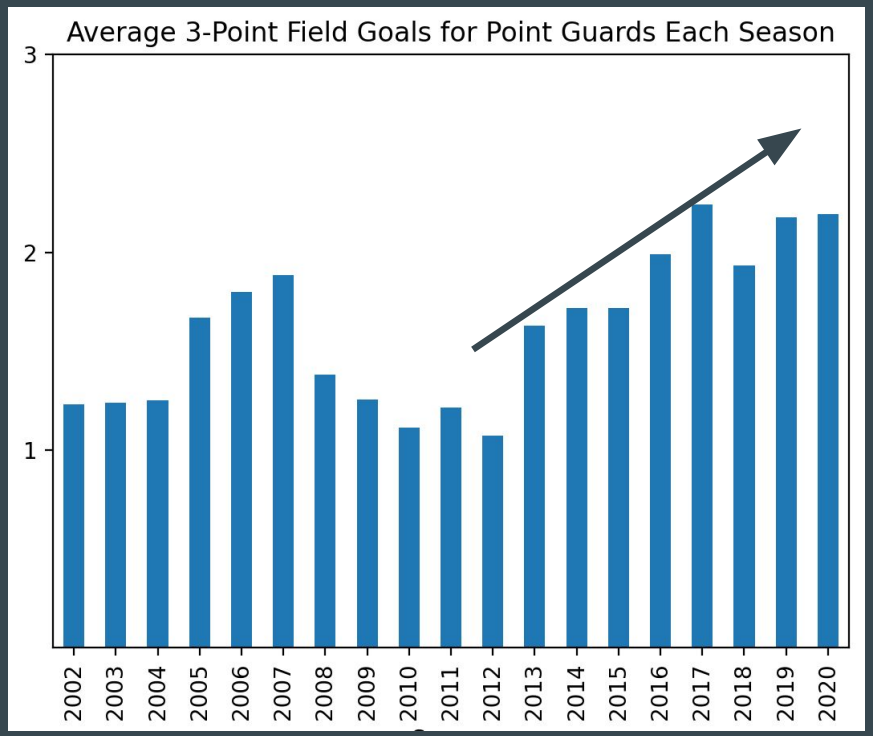


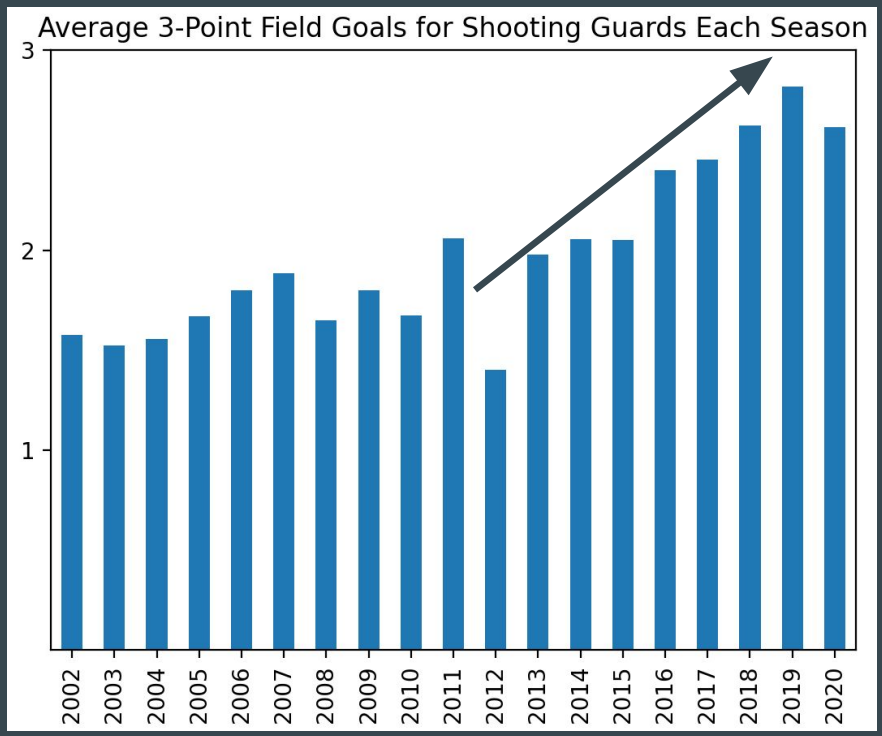
3-Point Field Goals for Each Position

3-Point Field Goals by Position

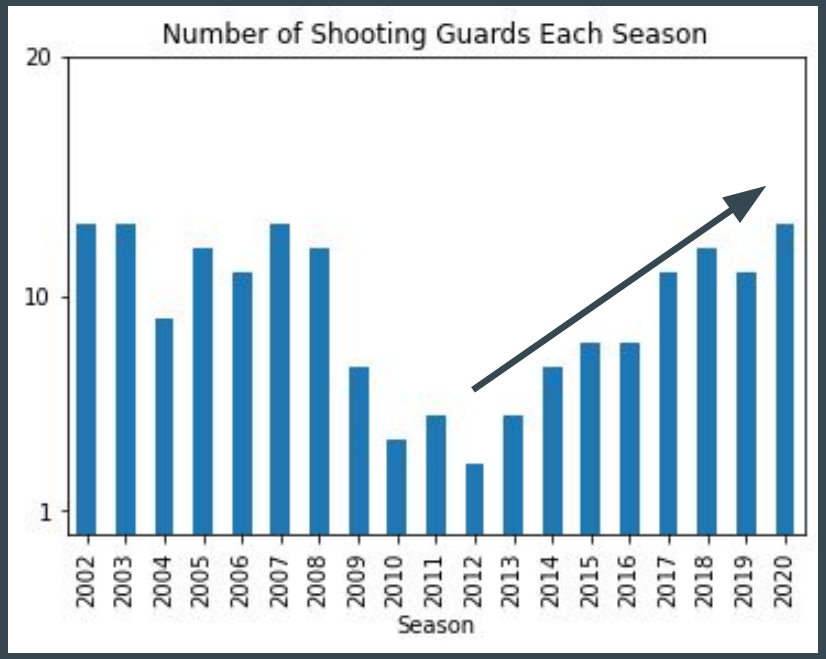
- Question: “How has the game evolved over the years in terms of three-point shooting?”
  - In the NBA community, many people believe that 3-pointers are becoming much more common among players
    - “As players like Steph Curry took over the league, how has the 3 point shot changed over time?”
  - We wanted to determine which position has increased the most in terms of 3-pointers and if any position may be less needed now that the focus has shifted
- Analyzed the average 3-point field goals for each position from 2002-2020:
- Created a series to show the average number of 3-point field goals by season and position as well as a series to show the number of players in each position by season
- Developed bar plots to compare different positions

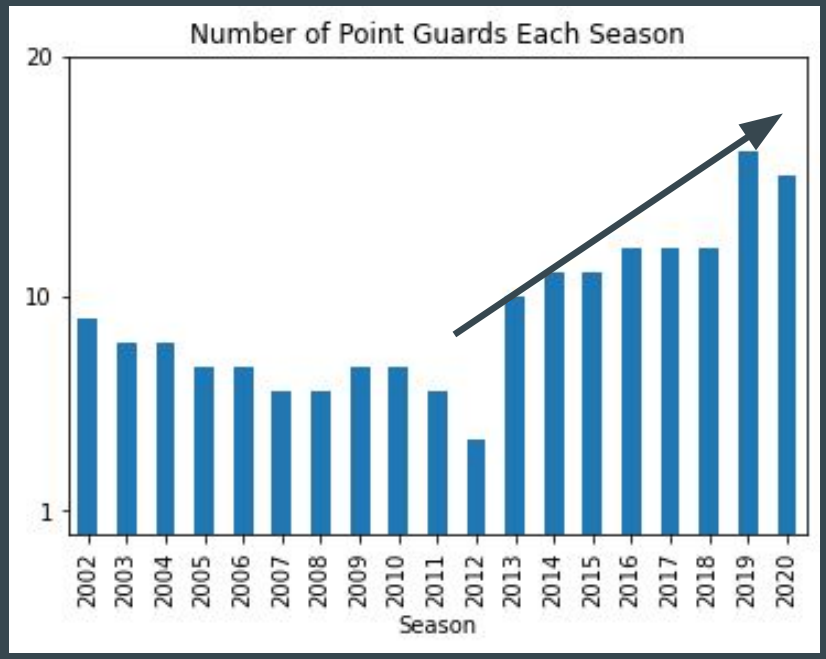
Point Guards and Shooting Guards have steadily increased the number of 3-point field goals since 2012, which may show an overall shift in the frequency and reliance of 3-point field goals in the NBA

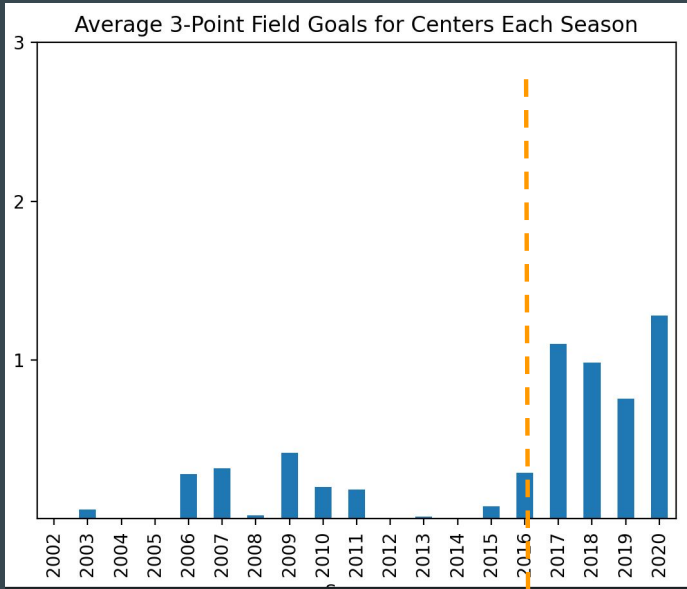


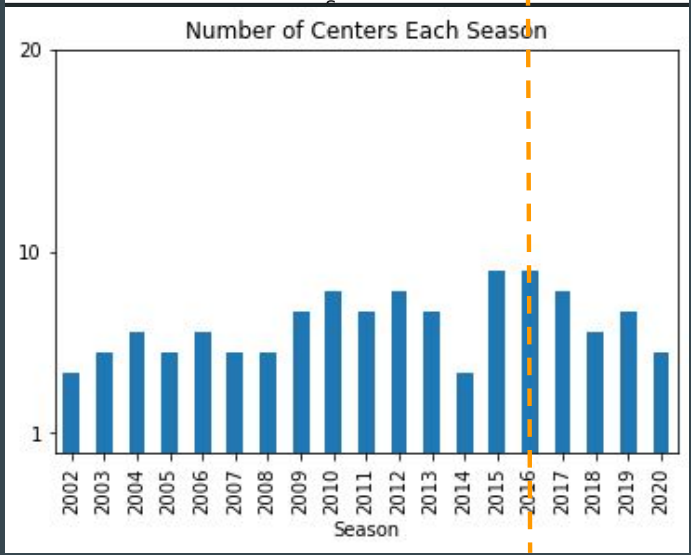


The number of Point Guards and Shooting Guards has also increased since 2012, which shows that teams may be putting more importance on PGs and SGs since they are performing the best for 3-point field goals









Traditionally, Centers do not shoot many three-pointers, which is evident in the plot from 2002-2015. Starting in 2016, however, Centers have also started to increase the number of three-pointers, indicating a shift in the game.

Additionally, the number of Centers is decreasing since 2016 as PGs and SGs are increasing, indicating the traditional definition of a Center's role is evolving as three-point shooting is becoming more sought after.

Player of Choice

Player of Choice

- Question: “How do I search statistics on a specific player?”
- User input asks “What year do you want to see the top 5 players?”
- Results display the number of times the player makes the top 50 list, first year, first position, last year, last position, average number of games played, and average number of points scored

```
What player statistics do you want? Lebron James
From 2002 to 2020, lebron james has been ranked in the top 50 list of NBA players 16 times
He first made the top 50 list in 2004 as a small forward and last made the list in 2020 as a small forward
For the years he made the list, he played on average 75.62 games and averaged 27.03 points per game
```

Top 5 All Around Ranked Players

Top All Around Ranked Players

- Question: “Who are the top all around players in the NBA?”
- Player statistics were averaged across all seasons
- A ranking system was implemented (from 1 to 254) to compare across the different categories:
  - Points Per Game, Field Goal Percentage, 3-point Field Goal Percentage, Free Throw Percentage, Rebounds Per Game, Assists Per Game, Steals Per Game, Blocks Per Game
- These rankings were averaged to give each player an overall average rank score
- Results display the top 5 all around players in order of their Average Rank

Player Name	PointsRank	FGPercentRank	ThreePointPercentRank	FTPercentRank	ReboundsRank	AssistsRank	StealsRank	BlocksRank	Average Rank
Kevin Durant	4.0	49.0	67.0	13.0	73.0	89.0	102.0	38.0	54.38
Kawhi Leonard	24.0	54.0	52.0	33.0	81.0	119.0	8.0	70.0	55.12
Stephen Curry	22.0	71.0	6.0	5.0	156.0	27.0	20.0	192.0	62.38
James Harden	2.0	147.0	96.0	31.0	114.0	19.0	23.0	93.0	65.62
LeBron James	4.0	34.0	137.0	206.0	71.0	17.0	29.0	68.0	70.75

Top All Around Ranked Players Per Season

- Question: “Who are the top all around players in the NBA for each season?”
- Within one season, values were ranked 1 to 50 in the following categories:
  - Points Per Game, Field Goal Percentage, 3-point Field Goal Percentage, Free Throw Percentage, Rebounds Per Game, Assists Per Game, Steals Per Game, Blocks Per Game
- User input asks “What year do you want to see the top 5 players?”
- Results display the top 5 all around players in order for that season based on the indicated ranking system

```
What year do you want to see the top 5 players? 2020
        Player Name Cities Position  Season  Average Rank
907   Kawhi Leonard    LAC       SF    2020         12.88
900    James Harden    HOU       SG    2020         13.25
902  Damian Lillard    POR       PG    2020         15.25
909   Anthony Davis    LAL       PF    2020         15.38
928    Nikola Jokic    DEN        C    2020         17.62
```

Testing

- Top-5 Starting Lineup Testing:
  - Tested variables and lists in the code to verify that we were getting the correct output
  - Used the “unittest” library to perform glass-box testing
  - Used assertEqual() statements to compare expected player name with the code output
  - Used assertEqual() statements to compare expected statistical value for each observed category with the code output
  - Developed individual unit tests for each variable and list element
  - Verify that the first player to achieve the highest value in an observed category was selected (for categories with multiple players with the highest value)
- LeBron James vs MVP Testing:
  - Wanted to verify that the MVP for each season was correctly identified and collected
  - Used the “unittest” library to perform glass-box testing
  - Tested a list that contained the MVPs in order by season they won the MVP
  - Used assertEqual() statements to compare expected player name with the code output

Testing cont.

- 3-Point Field Goals:
  - Assert statements to test that the series grouped by seasons and positions have the correct number of rows with all expected seasons and positions
  - Assert statements to test that the smaller subsets added up to the large data set that it was grouped from
- Top All-Around Player Ranking System:
  - First, unit test to see if the code was sorting the data set properly
  - Used assertEqual() statements to compare expected player name with the code output
    - The expected top player for mean points per game throughout all the seasons was Allen Iverson and after grouping and sorting, the code output Allen Iverson
  - Next, unit test if the the ranking system was working properly
  - Used assertEqual() statements to compare expected player name with the code output
    - The expected top player for mean field goal percentage was Clint Capella, and after grouping, sorting, and ranking within the dataframe, the code output Clint Capella as the top name

Beyond the Original Specifications

Web Scraping

- As an extension of the Web Scraping assignment, expanded code to scrape ESPN NBA player data of the Top 50 scorers from 19 different seasons
  - Did not include current season data because season is ongoing and data set is continuously updated
- Used BeautifulSoup library to perform the web scraping
- Collected data from 19 different web pages
  - Website structure for each web page was similar; had to modify year value in web address
- Developed a for-loop that collected data from each web page, stripped the data into an appropriate format, and placed into lists (based on each column attribute)
- Extracted data from web pages and exported it to a CSV file for analysis

User Input

- User input was implemented to display statistics on a player of choice
- User input was also implemented to display the top all-around ranked players for any year of the dataset

Conclusion

- Summary of findings
  - Over the years, there is a trend of increased reliance on the 3-point shot and decreased reliance on centers
  - Top all around players: Kevin Durant, Kawhi Leonard, Stephen Curry, James Harden, and LeBron James
  - Top-5 Starting Lineup: James Harden (SG) for points per game, Andre Drummond (C) for rebounds, Steve Nash (PG) for assists, Tim Duncan (C) for blocks, Giannis Antetokounmpo (PF) for PER
  - There are potentially seasons where LeBron James was snubbed of an MVP based on statistics from our dataset
- Usability
  - New NBA fans could learn some information about how players, teams, and the NBA in general has changed over the years
  - NBA fans could have a better understanding of the top players in the NBA over time
  - NBA coaches/ GM could better identify quantitatively how the NBA has changed over the years to maximize their player and team's efficiency
- Improvements/ Expansion
  - There is potential to see if other players were snubbed of a MVP based on these statistics
  - Additional unit testing could be performed
  - Expand on our analysis to predict who the top players for future seasons may be

Questions?