Final Project

Omar Shatrat, Ali Kamel, Abhishek Ramesh

Introduction

Throughout NBA history, there has been one award that is so coveted, so elusive, and so important that it has defined the legacies of some of the greatest basketball players to ever set foot on the hardwood. The MVP is determined by a group of over 100 acclaimed sports journalists who vote on the winner at the end of each season. Given the profound impact such an award can have on a player's career, we are interested in learning about how past athletes have achieved this historic feat. Through analyzing a plethora of variables such as statistical averages, advanced metrics, physical features, and more, we will further investigate just what it takes to be crowned the Most Valuable Player of the National Basketball Association.

Preparing the Data

To begin our investigation, we load our MVP dataset as well as several libraries that will be of use to us throughout the analysis. We are analyzing data since 1976 only because many statistical measures were not tracked prior to then.

```
location <- "C:\\Users\\Abhishek Ramesh\\Google Drive\\1USF\\Fall 2018\\BSDS 100\\Final Project\\bsds.cs
mvp_data <- read.csv(location, header = T, sep = ",")

str(mvp_data)
```

```
## 'data.frame':    43 obs. of  29 variables:
##  $ Season              : int  2006 2005 2008 1999 1997 1981 2007 2015 2016 1984 ...
##  $ Lg                  : Factor w/ 1 level "NBA": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Player              : Factor w/ 24 levels "Allen Iverson",..: 23 23 14 11 11 9 6 22 22 15 ...
##  $ Draft.Year          : int  1996 1996 1996 1985 1985 1972 1998 2009 2009 1978 ...
##  $ Draft.Pick          : int  15 15 13 13 13 12 9 7 7 6 ...
##  $ Years.Before.1st.MVP: int  NA 9 12 NA 12 9 9 6 NA 6 ...
##  $ Age                 : int  31 30 29 35 33 30 28 26 27 27 ...
##  $ Tm                  : Factor w/ 15 levels "BOS","CHI","CLE",..: 12 12 7 15 15 11 4 5 5 1 ...
##  $ G                   : int  79 75 82 49 82 82 78 80 79 79 ...
##  $ MP                  : num  35.4 34.3 38.9 37.4 36.6 35 36.2 32.7 34.2 38.3 ...
##  $ PTS                 : num  18.8 15.5 28.3 23.8 27.4 24.6 24.6 23.8 30.1 24.2 ...
##  $ TRB                 : num  4.2 3.3 6.3 9.4 9.9 8 8.9 4.3 5.4 10.1 ...
##  $ AST                 : num  10.5 11.5 5.4 4.1 4.5 4.4 3.4 7.7 6.7 6.6 ...
##  $ STL                 : num  0.8 1 1.8 1.3 1.4 2.1 0.7 2 2.1 1.8 ...
##  $ BLK                 : num  0.2 0.1 0.5 0.6 0.6 1.8 0.8 0.2 0.2 0.9 ...
##  $ FG.                 : num  0.512 0.502 0.459 0.493 0.55 0.521 0.502 0.487 0.504 0.492 ...
##  $ X3P.                : num  0.439 0.431 0.361 0 0 0.222 0.416 0.443 0.454 0.247 ...
##  $ FT.                 : num  0.921 0.887 0.84 0.788 0.755 0.787 0.904 0.914 0.908 0.888 ...
##  $ WS                  : num  12.4 10.9 13.8 9.6 16.7 13.8 16.3 15.7 17.9 13.6 ...
##  $ WS.48               : num  0.212 0.203 0.208 0.252 0.268 0.231 0.278 0.288 0.318 0.215 ...
##  $ MVP.Position        : Factor w/ 3 levels "Center","Forward",..: 3 3 3 2 2 2 2 3 3 2 ...
##  $ MVP.Nationality     : Factor w/ 4 levels "Canada","Germany",..: 1 1 4 4 4 4 2 4 4 4 ...
##  $ Finals.Finish       : Factor w/ 3 levels "Did not reach Finals",..: 1 1 2 1 2 1 1 3 2 3 ...
##  $ MVP.Height..ft.     : num  6.27 6.27 6.5 6.76 6.76 6.59 6.99 6.27 6.27 6.76 ...
##  $ Wins                : int  54 62 57 37 64 62 67 67 73 62 ...
```

```
##  $ Losses              : int  28 20 25 13 18 20 15 15 9 20 ...
##  $ Win..               : num  0.659 0.756 0.695 0.74 0.78 0.756 0.817 0.817 0.89 0.756 ...
##  $ Conf.Seed           : int  2 1 1 2 1 3 1 1 1 1 ...
##  $ College             : Factor w/ 18 levels "Arizona State University",..: 10 10 9 6 6 13 9 3 3 5 .
```

```
library(ggplot2)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------- tidyverse 1.2.1 --
```

```
## v tibble  2.1.1     v purrr   0.2.5
## v readr   1.1.1     v stringr 1.3.1
## v tibble  2.1.1     v forcats 0.3.0
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

```
## -- Conflicts ------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tibble)
library(stringr)
```

Statistical Analysis

We begin our analysis by examining several relevant statistical variables to highlight the individual performance of past MVPs. This will help us get an idea of how well a player must perform to have a chance at winning the award.

To answer the first four qustions, we calculated averages, benchmarked against commonly accepted rubrics, and examined how counts changed based on position. We also created graphs and wrangled data to help us better visualize the relationship between the variables.

(Generally accepted) scoring rubric in points per game (ppg):

0-5 ppg: Fringe NBA Player
5-10 ppg: Role player
10-15 ppg: Starting caliber player
15-20 ppg: Fringe All-Star
20-25 ppg: All-Star
25+ ppg: Elite Caliber Player

(Generally accepted) rebounding rubric in rebounds per game (rpg):

0-3 rpg: Minimal Rebounder
4-6 rpg: Effective Rebounder
7-9 rpg: Upper-Echelon Rebounder
10+ rpg: Elite Rebounder

(Generally accepted) assist rubric in assists per game (apg):

0-3 apg: Minimal Assister
4-6 apg: Effective Assister
7-9 apg: Upper-Echelon Assister
10+ apg: Elite Assister

(Generally accepted) free-throw percentage rubric:

< 60%: Abyssmal Free-Throw Shooter
60% - 70%: Poor Free-Throw Shooter
70% - 80%: Average Free-Throw Shooter
80% - 90%: Good Free-Throw Shooter
90% - 1000%: Excellent Free-Throw Shooter

Keep in mind that assist and rebounding rubric values usually vary by position. Genrally, forwards and centers are expected to have more rebounds than guards because they are bigger and play closer to the basket. Guards are expected to have more assists than forwards and centers because they are smaller, faster, and handle the ball more.

For free throws, guards generally are expected to have higher free-throw percentages because they shoot more. Centers usually have lower percentages because they play closer to the basket and do not shoot from outside very often.

Do MVPs tend to have a high scoring average?

```
mean(mvp_data$PTS)
```

```
## [1] 26.58837
```

```
min(mvp_data$PTS)
```

```
## [1] 15.5
```

```
max(mvp_data$PTS)
```

```
## [1] 35
```

```
mvp_data$Player[mvp_data$PTS < 20]
```

```
## [1] Steve Nash  Steve Nash  Bill Walton
## 24 Levels: Allen Iverson Bill Walton Charles Barkley ... Tim Duncan
```

Only 3 times has the MVP winner averaged less than 20 points, and only two players have won it averaging less than 20 points. One of which was Bill Walton, who did it in the late 70's when scoring was not as abundant as it is in today's game. This demonstrates that the MVP Winner almost always must be able to score in bunches. In the rare case of Steve Nash, he averaged double-digit assists and a very high shooting percentage in addition to leading his team deep into the playoffs, a feat that may have allowed him to snag the award two consecutive years.

Do MVPs tend to have a high rebounding average?

```
mean(mvp_data$TRB)
```

```
## [1] 9.144186
```

```r
mean(mvp_data$TRB[mvp_data$MVP.Position == "Guard"])
```

```
## [1] 5.8
```

```r
mean(mvp_data$TRB[mvp_data$MVP.Position == "Forward"])
```

```
## [1] 9.78125
```

```r
mean(mvp_data$TRB[mvp_data$MVP.Position == "Center"])
```

```
## [1] 13.81
```

```r
length(mvp_data$Player[mvp_data$TRB < 7 & mvp_data$MVP.Position == "Guard"]) #out of 17
```

```
## [1] 15
```

```r
length(mvp_data$Player[mvp_data$TRB < 7 & mvp_data$MVP.Position == "Forward"]) #out of 16
```

```
## [1] 0
```

```r
length(mvp_data$Player[mvp_data$TRB < 10 & mvp_data$MVP.Position == "Center"]) #out of 10
```

```
## [1] 0
```

```r
select(spread(mvp_data, MVP.Position, TRB), Player, Center, Forward, Guard)
```

```
##                 Player Center Forward Guard
## 1           Steve Nash     NA      NA   4.2
## 2           Steve Nash     NA      NA   3.3
## 3          Kobe Bryant     NA      NA   6.3
## 4          Karl Malone     NA     9.4    NA
## 5          Karl Malone     NA     9.9    NA
## 6        Julius Erving     NA     8.0    NA
## 7        Dirk Nowitzki     NA     8.9    NA
## 8        Stephen Curry     NA      NA   4.3
## 9        Stephen Curry     NA      NA   5.4
## 10          Larry Bird     NA    10.1    NA
## 11          Larry Bird     NA    10.5    NA
## 12          Larry Bird     NA     9.8    NA
## 13       Kevin Garnett     NA    13.9    NA
## 14     Charles Barkley     NA    12.2    NA
## 15   Russell Westbrook     NA      NA  10.7
## 16        Moses Malone   14.7      NA    NA
## 17        Moses Malone   17.6      NA    NA
## 18      Michael Jordan     NA      NA   5.5
## 19      Michael Jordan     NA      NA   6.0
## 20      Michael Jordan     NA      NA   5.8
## 21        James Harden     NA      NA   5.4
## 22        Moses Malone   15.3      NA    NA
## 23      Michael Jordan     NA      NA   6.4
## 24      Michael Jordan     NA      NA   6.6
## 25        Kevin Durant     NA     7.4    NA
## 26 Kareem Abdul-Jabbar   16.9      NA    NA
## 27 Kareem Abdul-Jabbar   13.3      NA    NA
## 28       Allen Iverson     NA      NA   3.8
## 29       Magic Johnson     NA      NA   7.9
## 30        LeBron James     NA     7.9    NA
## 31          Tim Duncan     NA    12.7    NA
```

```
## 32     Hakeem Olajuwon   11.9      NA    NA
## 33         Bill Walton   13.2      NA    NA
## 34          Tim Duncan     NA    12.9    NA
## 35 Kareem Abdul-Jabbar   10.8      NA    NA
## 36        LeBron James     NA     7.3    NA
## 37        Derrick Rose     NA      NA   4.1
## 38      David Robinson   10.8      NA    NA
## 39       Magic Johnson     NA      NA   6.6
## 40       Magic Johnson     NA      NA   6.3
## 41        LeBron James     NA     8.0    NA
## 42        LeBron James     NA     7.6    NA
## 43    Shaquille O'Neal   13.6      NA    NA
```

The average rebounding total per game for an MVP is just above 9, which is exceptional. However, when breaking down the averages by position, we notice some interesting trends. The average among guards is 5.8 rebounds, for forwards it is 9.8, and for centers it is 13.8. These are all above average values for their positions, however it clearly seems as though bigger positions have had to grab more rebounds to win the MVP award. 15 out of 17 guards have won the MVP averaging less than 7 rpg. 0 forwards have accomplished the same feat. 0 centers have ever won averaging less than 10 rebounds. What this shows is that for bigger players, grabbing a lot rebounds is essential in trying to win the MVP.

Do MVPs tend to have a high assist average?

```
mean(mvp_data$AST)
```

```
## [1] 5.990698
```

```
mean(mvp_data$AST[mvp_data$MVP.Position == "Guard"])
```

```
## [1] 7.947059
```

```
mean(mvp_data$AST[mvp_data$MVP.Position == "Forward"])
```

```
## [1] 5.55625
```

```
mean(mvp_data$AST[mvp_data$MVP.Position == "Center"])
```

```
## [1] 3.36
```

```
length(mvp_data$Player[mvp_data$AST > 7 & mvp_data$MVP.Position == "Guard"]) #out of 17
```

```
## [1] 9
```

```
length(mvp_data$Player[mvp_data$AST > 7 & mvp_data$MVP.Position == "Forward"]) #out of 16
```

```
## [1] 3
```

```
length(mvp_data$Player[mvp_data$AST > 7 & mvp_data$MVP.Position == "Center"]) #out of 10
```

```
## [1] 0
```

```
select(spread(mvp_data, MVP.Position, AST), Player, Center, Forward, Guard)
```

```
##                Player Center Forward Guard
## 1          Steve Nash     NA      NA  10.5
## 2          Steve Nash     NA      NA  11.5
## 3         Kobe Bryant     NA      NA   5.4
## 4         Karl Malone     NA     4.1    NA
## 5         Karl Malone     NA     4.5    NA
## 6       Julius Erving     NA     4.4    NA
## 7       Dirk Nowitzki     NA     3.4    NA
```

```
## 8         Stephen Curry     NA      NA   7.7
## 9         Stephen Curry     NA      NA   6.7
## 10           Larry Bird     NA     6.6    NA
## 11           Larry Bird     NA     6.6    NA
## 12           Larry Bird     NA     6.8    NA
## 13        Kevin Garnett     NA     5.0    NA
## 14      Charles Barkley     NA     5.1    NA
## 15   Russell Westbrook     NA      NA  10.4
## 16         Moses Malone    1.8      NA    NA
## 17         Moses Malone    1.8      NA    NA
## 18       Michael Jordan     NA      NA   5.9
## 19       Michael Jordan     NA      NA   5.5
## 20       Michael Jordan     NA      NA   3.5
## 21         James Harden     NA      NA   8.8
## 22         Moses Malone    1.3      NA    NA
## 23       Michael Jordan     NA      NA   6.1
## 24       Michael Jordan     NA      NA   4.3
## 25         Kevin Durant     NA     5.5    NA
## 26 Kareem Abdul-Jabbar    5.0      NA    NA
## 27 Kareem Abdul-Jabbar    3.9      NA    NA
## 28        Allen Iverson     NA      NA   4.6
## 29        Magic Johnson     NA      NA  12.8
## 30         LeBron James     NA     6.2    NA
## 31           Tim Duncan     NA     3.7    NA
## 32     Hakeem Olajuwon    3.6      NA    NA
## 33          Bill Walton    5.0      NA    NA
## 34           Tim Duncan     NA     3.9    NA
## 35 Kareem Abdul-Jabbar    4.5      NA    NA
## 36         LeBron James     NA     8.6    NA
## 37         Derrick Rose     NA      NA   7.7
## 38       David Robinson    2.9      NA    NA
## 39        Magic Johnson     NA      NA  11.5
## 40        Magic Johnson     NA      NA  12.2
## 41         LeBron James     NA     7.3    NA
## 42         LeBron James     NA     7.2    NA
## 43    Shaquille O'Neal    3.8      NA    NA
```

The assist average for MVP winners is a respectable 6 apg, which is nice but not extraordinary. Among guards, the average is 7.9, for forwards it is 5.5, and for centers it is 3.3. Generally, smaller players accumulate more assists. 9 of 17 guards have won the MVP averaging greater than 7 assists while 3 forwards and 0 centers have accomplished the same feat. 8 guards have won MVP with less-than-superb assist numbers compared to 0 centers winning the award with less-than-stellar rebounding numbers. This demonstrates that assist averages are not weighed as heavily as rebounding averages when selecting an MVP. Although smaller players are expected to have more assists, it does not appear to be an absolute pre-requisite for winning an MVP the way rebounding is for centers and forwards.

Do MVPs make their free throws?

```
mean(mvp_data$FT.)
```

```
## [1] 0.8072093
```

```
mean(mvp_data$FT.[mvp_data$MVP.Position == "Guard"])
```

```
## [1] 0.8609412
```

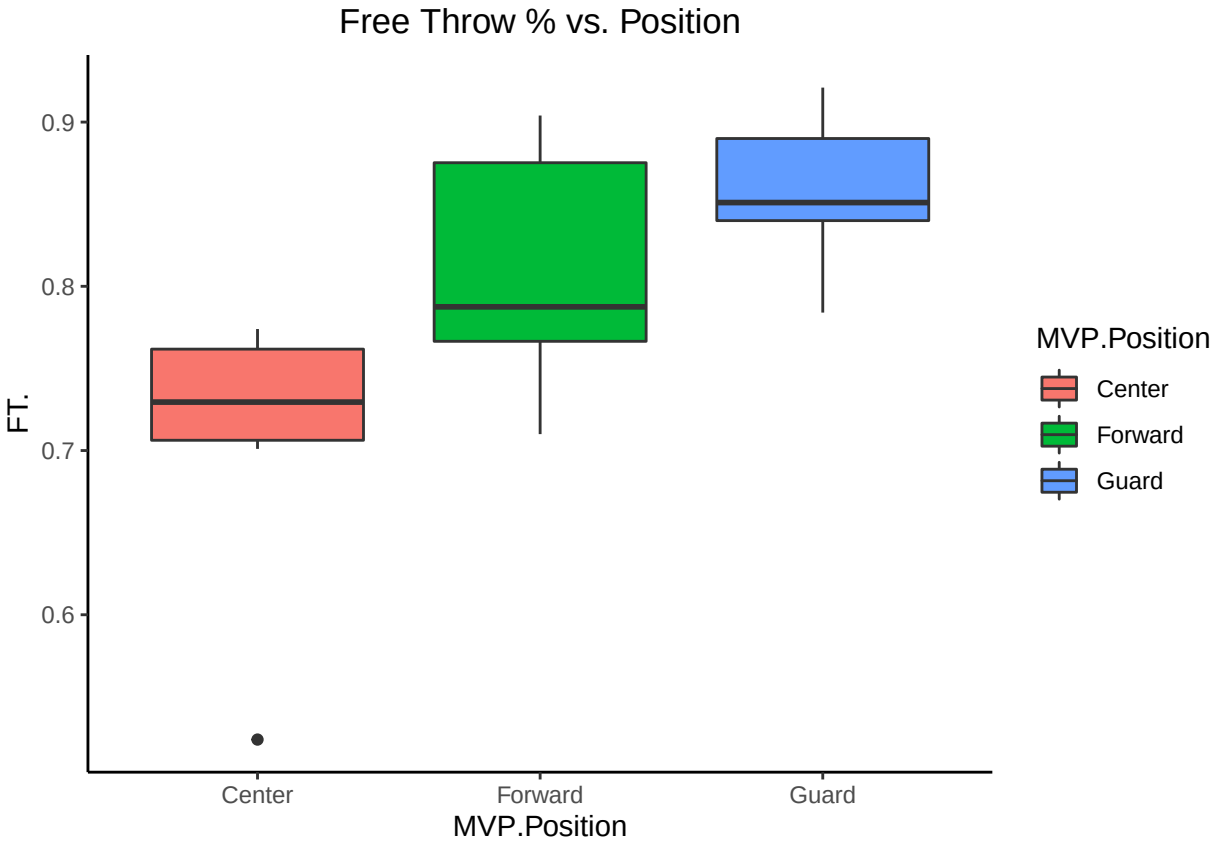
```r
mean(mvp_data$FT.[mvp_data$MVP.Position == "Forward"])
```

```
## [1] 0.8068125
```

```r
mean(mvp_data$FT.[mvp_data$MVP.Position == "Center"])
```

```
## [1] 0.7165
```

```r
ggplot(mvp_data, aes(x = MVP.Position, y = FT., fill = MVP.Position)) +
  geom_boxplot() + theme_classic() +
  ggtitle("Free Throw % vs. Position") +
  theme(plot.title = element_text(hjust = 0.5))
```




```r
mvp_data$Player[mvp_data$FT. == min(mvp_data$FT.)]
```

```
## [1] Shaquille O'Neal
## 24 Levels: Allen Iverson Bill Walton Charles Barkley ... Tim Duncan
```

```r
min(mvp_data$FT.)
```

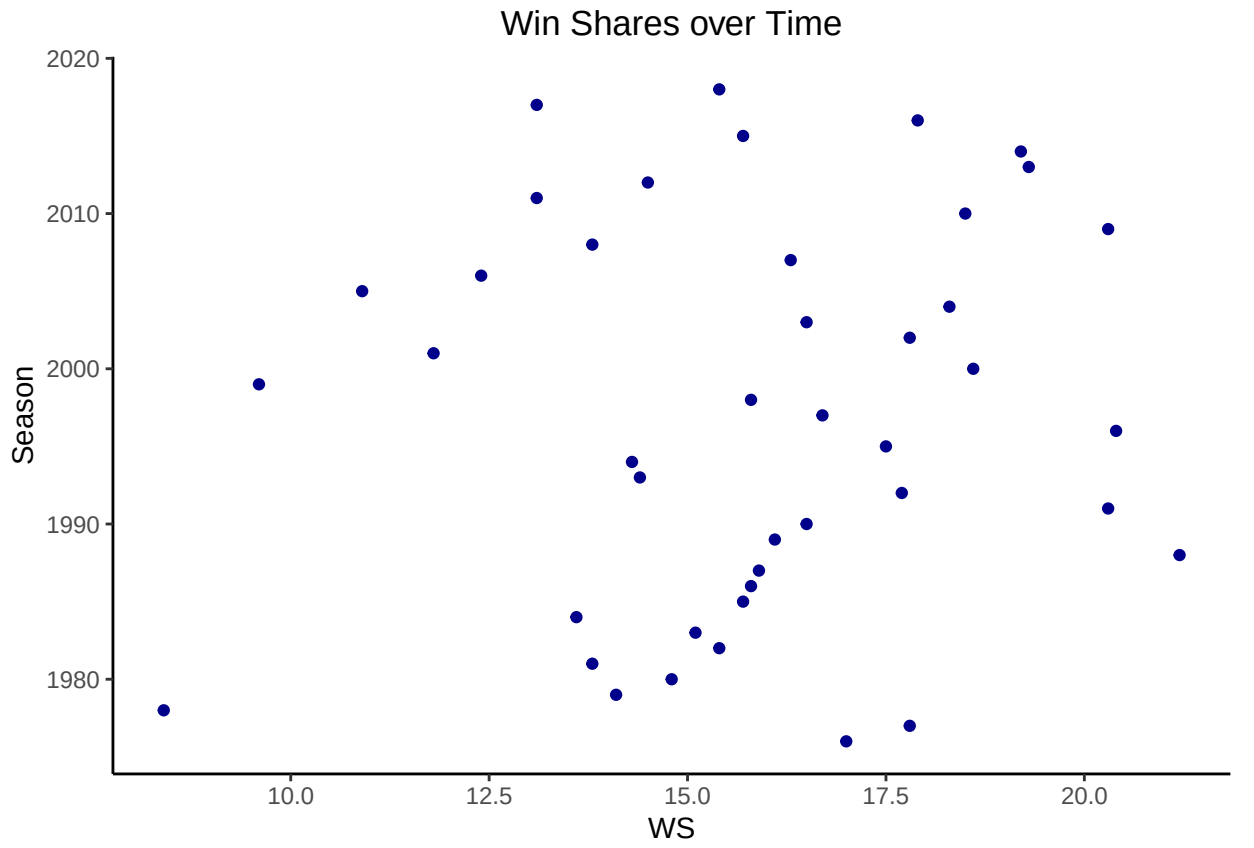
```
## [1] 0.524
```

The average free-throw percentage among MVPs is 81%; for guards it is 86%, for forwards it is 81%, and for centers it is 72%. The boxplots illustrate that smaller players tend to make more free throws. Regardless of this variance, MVPs are at worst average free throw shooters, except for Shaquille O'Neal, who shot 52.4% from the line when he won the MVP Award.

Analyze how Win Shares have changed among MVPs since 1976.

To answer this, we created a scatterplot of Season vs. Win Shares to observe any possible relationship.

```
ggplot(mvp_data, aes(x= WS, y = Season)) +
  geom_point(fill='#A4B4A4', color="blue4") +
  theme_classic() +
  ggtitle("Win Shares over Time") +
  theme(plot.title = element_text(hjust = 0.5))
```




```
cor(mvp_data$WS, mvp_data$Season)
```

```
## [1] 0.1034462
```

Over the years, it appears as though MVP winners are accumulating slightly higher Win Share totals. There is a fairly weak, positive correlation of .1.

Are MVPs generally exceptional in one statistical category or do they excel across the board?

To answer this, we found the proportion of players who met various statistical benchmarks.

```
length(mvp_data$Player[mvp_data$PTS > 20 & mvp_data$TRB > 10])/length(mvp_data$Season)
```

```
## [1] 0.372093
```

```
length(mvp_data$Player[mvp_data$PTS > 20 & mvp_data$AST > 7])/length(mvp_data$Season)
```

```
## [1] 0.2325581
```

```
length(mvp_data$Player[mvp_data$PTS > 20 & mvp_data$TRB > 5 & mvp_data$AST > 5])/
  length(mvp_data$Season)
```

```
## [1] 0.4418605
```

```
length(mvp_data$Player[(mvp_data$PTS > 20 & mvp_data$TRB > 10) |
                         (mvp_data$PTS > 20 & mvp_data$AST > 7) |
                         (mvp_data$PTS > 20 & mvp_data$TRB > 5 & mvp_data$AST > 5)])/
  length(mvp_data$Season)
```

```
## [1] 0.7674419
```

About 77% of the time, the MVP winner has demonstrated significant achievement across multiple statistical categories by averaging either (20 ppg & 10 rpg), (20 ppg & 7 apg), or (20 ppg, 5 rpg & 5 apg) at least. It is not uncommon for an MVP to only be an exceptional scorer or passer or rebounder, but they usually are adept in more than one statistical category and thus, are able to do more for their teams.

Team Success Analysis

The second major component that is factored into selecting an MVP is team success. It is not unsreasonable to believe that more successful teams have a better chance at producing an MVP. Through analyzing variables relating to team performance in the regular season and the playoffs, we hope to find out just how well a team needs to perform for one of its players to win an MVP.

Is a playoff apperance something that is common among NBA MVPs?

In order to make the playoffs, a team must have a seed of 8 or lower in their conference. We answered the question by subsetting Conference Seeds and finding the count that was greater than 8.

```
mvp_data$Conf.Seed[mvp_data$Conf.Seed > 8]
```

```
## integer(0)
```

There have been 0 MVPs in the history of the NBA whose Conference Seed was lower than 8. This means every NBA MVP since 1976 has made the playoffs that season.

Are MVP Winners on teams that are usually seeded high in their respective conference?

To visualize this answer easily, we created a table of MVP winner Conference Seeds.

```
table(mvp_data$Conf.Seed)
```

```
##
##  1  2  3  4  6
## 31  6  2  1  3
```

The overwhelming majority of MVP winners have been seeded #1 in their conference that same season. Occaisionally, the MVP will be on a lower seeded team, but that has only happened 12 times out of 43. Interestingly, 3 MVPs been seeded #6; I predict this is because they displayed heculean efforts for their teams, who otherwise would have performed terribly.

What percentage of MVPs have gone on to win a championship that season?

To answer this, we found the total number of seasons with a Finals Finish of "Won Finals" and divided that by the total number of seasons.

```
length(mvp_data$Season[mvp_data$Finals.Finish == "Won Finals"])/length(mvp_data$Season)
```

```
## [1] 0.3488372
```

The MVP goes on to win the championship that same season just over 1/3 of the time. This shows that MVPs do not necesarilly need to be champions, but rather players who do a lot for their teams even if it is in a losing effort.

What teams produce the most MVPs? Why do you think this is?

To answer, we made a table of MVP counts by team and observed which players won for the most frequent teams.

```
table(mvp_data$Tm)
```

```
##
## BOS CHI CLE DAL GSW HOU LAL MIA MIN OKC PHI PHO POR SAS UTA
##   3   6   2   1   2   4   8   2   1   2   3   3   1   3   2
```

```
mvp_data$Player[mvp_data$Tm == "CHI"]
```

```
## [1] Michael Jordan Michael Jordan Michael Jordan Michael Jordan
## [5] Michael Jordan Derrick Rose
## 24 Levels: Allen Iverson Bill Walton Charles Barkley ... Tim Duncan
```

```
mvp_data$Player[mvp_data$Tm == "LAL"]
```

```
## [1] Kobe Bryant         Kareem Abdul-Jabbar Kareem Abdul-Jabbar
## [4] Magic Johnson       Kareem Abdul-Jabbar Magic Johnson
## [7] Magic Johnson       Shaquille O'Neal
## 24 Levels: Allen Iverson Bill Walton Charles Barkley ... Tim Duncan
```

The teams with the most MVPs appear to be the Chicago Bulls and Los Angeles Lakers. I predict the reason these teams have a large amount of MVP awards is due to their stature as big markets and also because they have outstanding players who win the award multiple times. Chicago has 6 MVP awards, 5 of which were won by Michael Jordan. Los Angeles' 8 awards were shared among 4 players, however two of them won the award 3 times and two won it only once. Teams usually acquire MVP awards within focused spans of time when they have incredibly talented players, rarely are MVPS accumulated over a dispersed period.

Have MVPs succeeded more in the regular season, the playoffs, or both?

To answer, we found the mean win percentage and well as the mean and minimum win percentage for teams with varying playoff success. We depicted this data in a boxplot as well to help better visualize the relationship between regular season win percentage and playoff success.

```
mean(mvp_data$Win..)
```

```
## [1] 0.734186
```

```
mean(mvp_data$Win..[mvp_data$Finals.Finish == "Did not reach Finals"])
```

```
## [1] 0.6972381
```

```
mean(mvp_data$Win..[mvp_data$Finals.Finish == "Lost in Finals"])
```

```
## [1] 0.7524286
```

```
mean(mvp_data$Win..[mvp_data$Finals.Finish == "Won Finals"])
```

```
## [1] 0.7774
```

```
min(mvp_data$Win..[mvp_data$Finals.Finish == "Did not reach Finals"])
```

```
## [1] 0.488
```

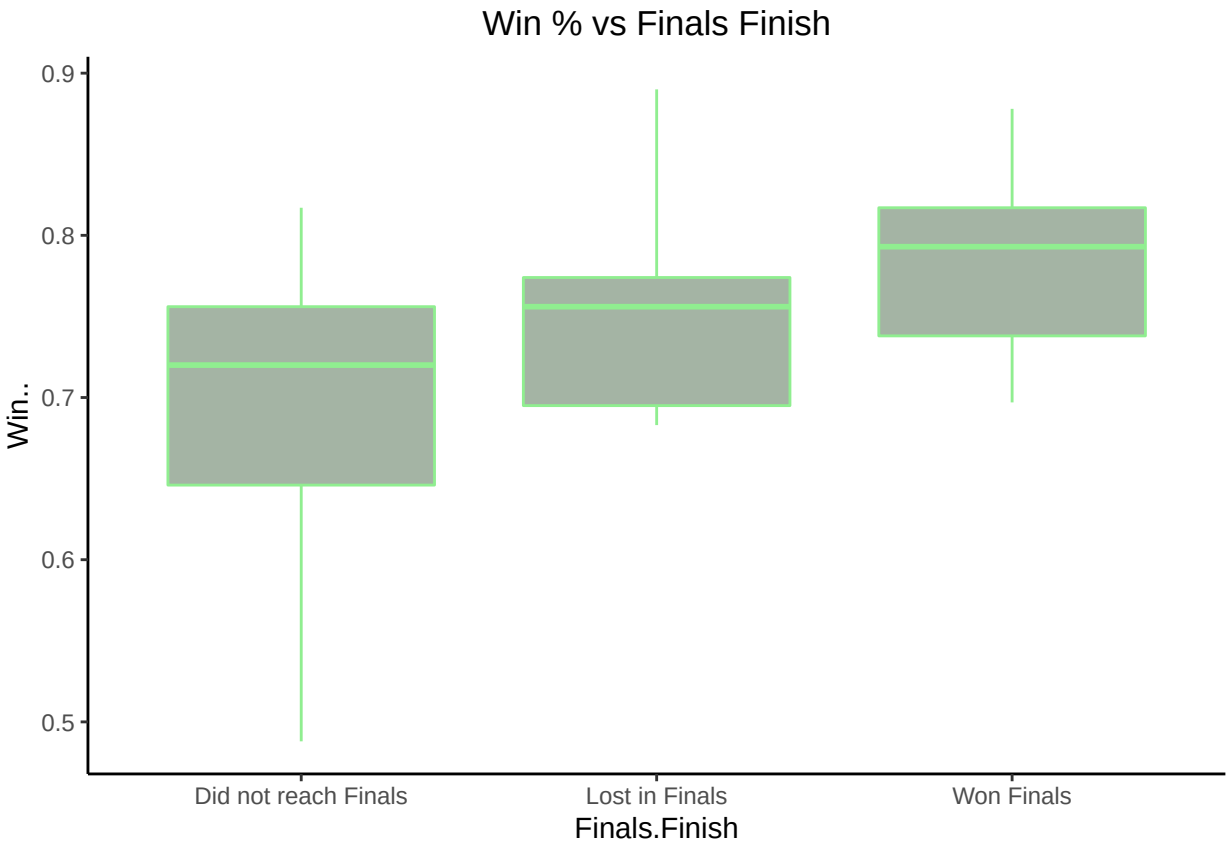
```
min(mvp_data$Win..[mvp_data$Finals.Finish == "Lost in Finals"])
```

```
## [1] 0.683
```

```
min(mvp_data$Win..[mvp_data$Finals.Finish == "Won Finals"])
```

```
## [1] 0.697
```

```
ggplot(mvp_data, aes(x = Finals.Finish, y = Win..)) +
  geom_boxplot(fill='#A4B4A4', color="lightgreen") +
  theme_classic() +
  ggtitle("Win % vs Finals Finish") +
  theme(plot.title = element_text(hjust = 0.5))
```



MVPs who have had a higher Win Percentage in the regular season tend to advance further in the Playoffs. MVPs with low regular season Win Percentages likely were given the award for carrying subpar teams into the playoffs with their historic performances. Once they reach the playoffs, the teams become no match for teams with more depth. An example would be Russell Westbrook in 2017; his team was seeded 6th out of 8 but he was the first player to ever average a triple-double (at least 10ppg, rpg, and apg) over the course of a season. In the playoffs, he did not reach the Finals because his efforts were not enough against more talented teams.

Miscellaneous Factors Analysis

There are a few external variables that may be useful in trying to determine the MVP. We take a look at past colleges and player nationality to see if there are any unforseen trends.

What school has produced the most MVPs? How many has it produced?

To answer, we used summarise and group_by to make a data frame showing us all unique winners by school.

```
summarise(group_by(mvp_data, College), unique_mvps = length(unique(Player)))
```

```
## # A tibble: 18 x 2
```

```
##    College                                    unique_mvps
##    <fct>                                            <int>
##  1 Arizona State University                             1
##  2 Auburn University                                    1
##  3 Davidson College                                     1
##  4 Georgetown University                                1
##  5 Indiana State University                             1
##  6 Louisiana Tech University                            1
##  7 Lousiana State University                            1
##  8 Michigan State University                            1
##  9 No College                                           5
## 10 Santa Clara University                               1
## 11 University of California, Los Angeles                3
## 12 University of Houston                                1
## 13 University of Massachusetts, Amherst                 1
## 14 University of Memphis                                1
## 15 University of North Carolina, Chapel Hill            1
## 16 University of Texas, Austin                          1
## 17 University of the Incarnate Word                     1
## 18 Wake Forest University                               1
```

Every college has produced 1 MVP except for UCLA, which has produced 3. 5 MVPs within this span did not attend college. Perhaps UCLA has superior traning methods.

Which countries produce the most MVPs? Is it common for an international player to win the award?

To answer, we created a table showing MVP winners' countries and found the proportion of MVPs who were not US born.

```
table(mvp_data$MVP.Nationality)
```

```
##
##  Canada Germany Nigeria      US
##       2       1       1      39
```

```
unique(mvp_data$Player[mvp_data$MVP.Nationality != "US"])
```

```
## [1] Steve Nash      Dirk Nowitzki   Hakeem Olajuwon
## 24 Levels: Allen Iverson Bill Walton Charles Barkley ... Tim Duncan
```

```
length(mvp_data$Player[mvp_data$MVP.Nationality != "US"])/length(mvp_data$Player)
```

```
## [1] 0.09302326
```

Generally speaking, it is very rare for an international player to win the MVP. It happens about 9% of the time and has only been done by 3 different players. The overwhelming majority of MVP winners have been US-born, which makes sense because the NBA is an American league.

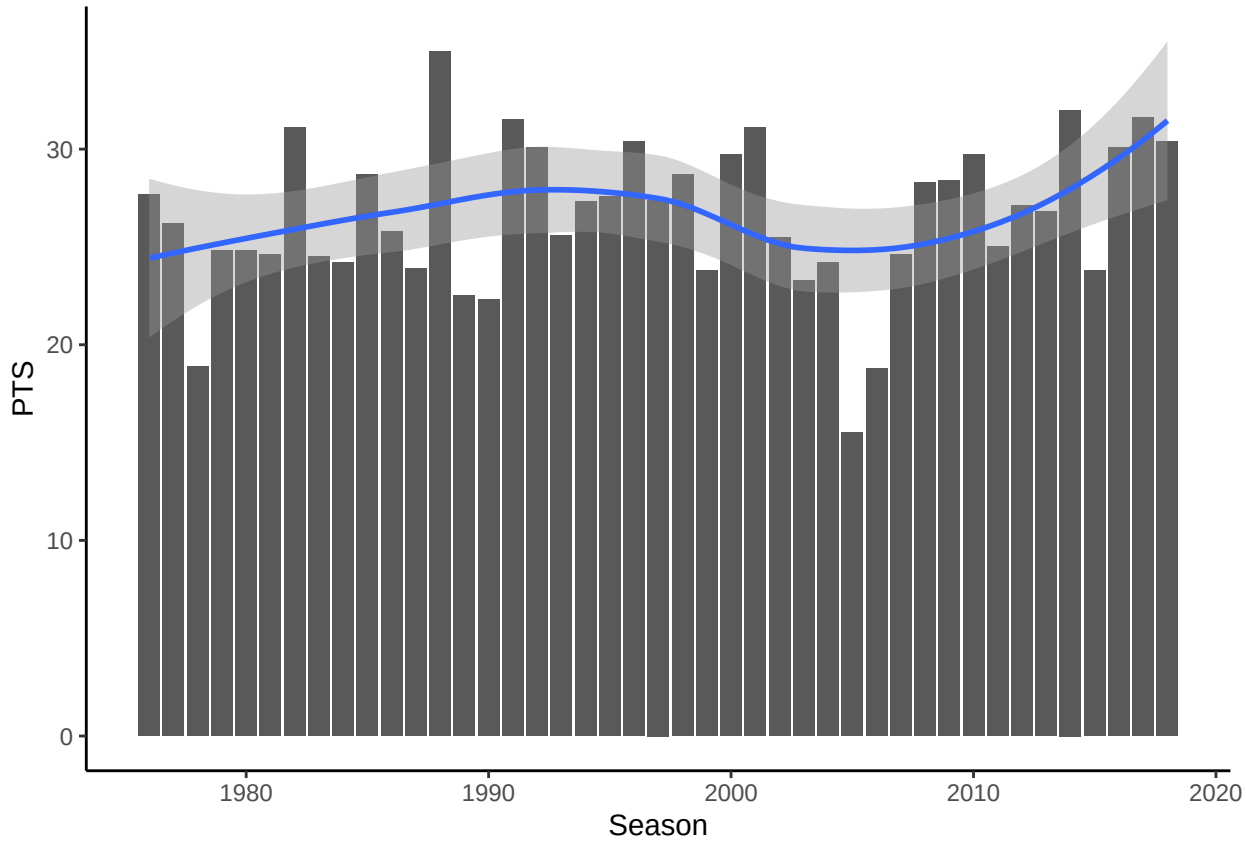
Team vs. Individual Statistics Analysis

Now that we have examined trends in individual statistics and team success, we want to know what information can be garnered from comparing both factors directly against each other. Through oberving changes in various statistical categories and team performance, we hope to determine which factor is more important in considering an MVP and to what extent each factor plays a role in MVP determination.

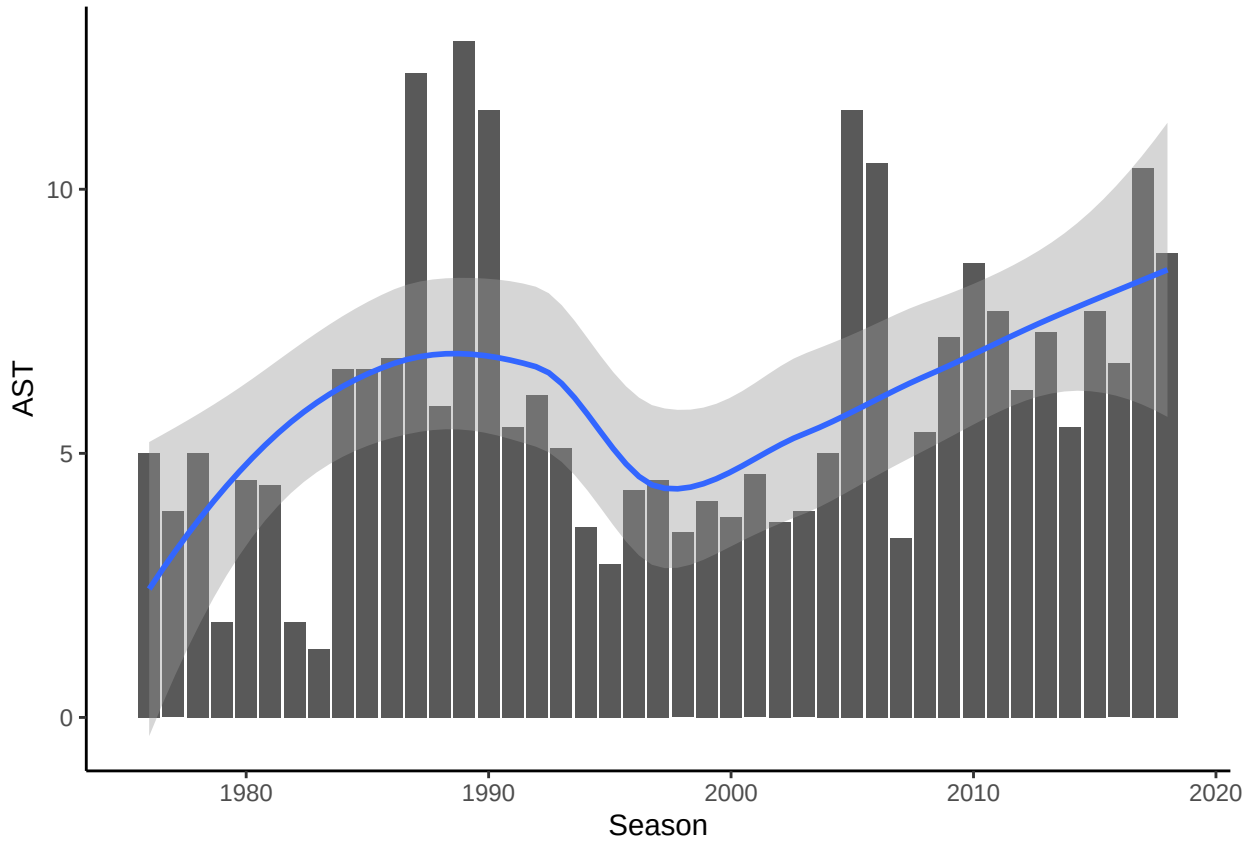
Across each decade, how have MVP trends changed? Consider changes in regular season success, playoff success, position, seeding, height, and statistical measures.

To answer, we made various plots with season as the x-axis in order to help us see how variables have changed over years.

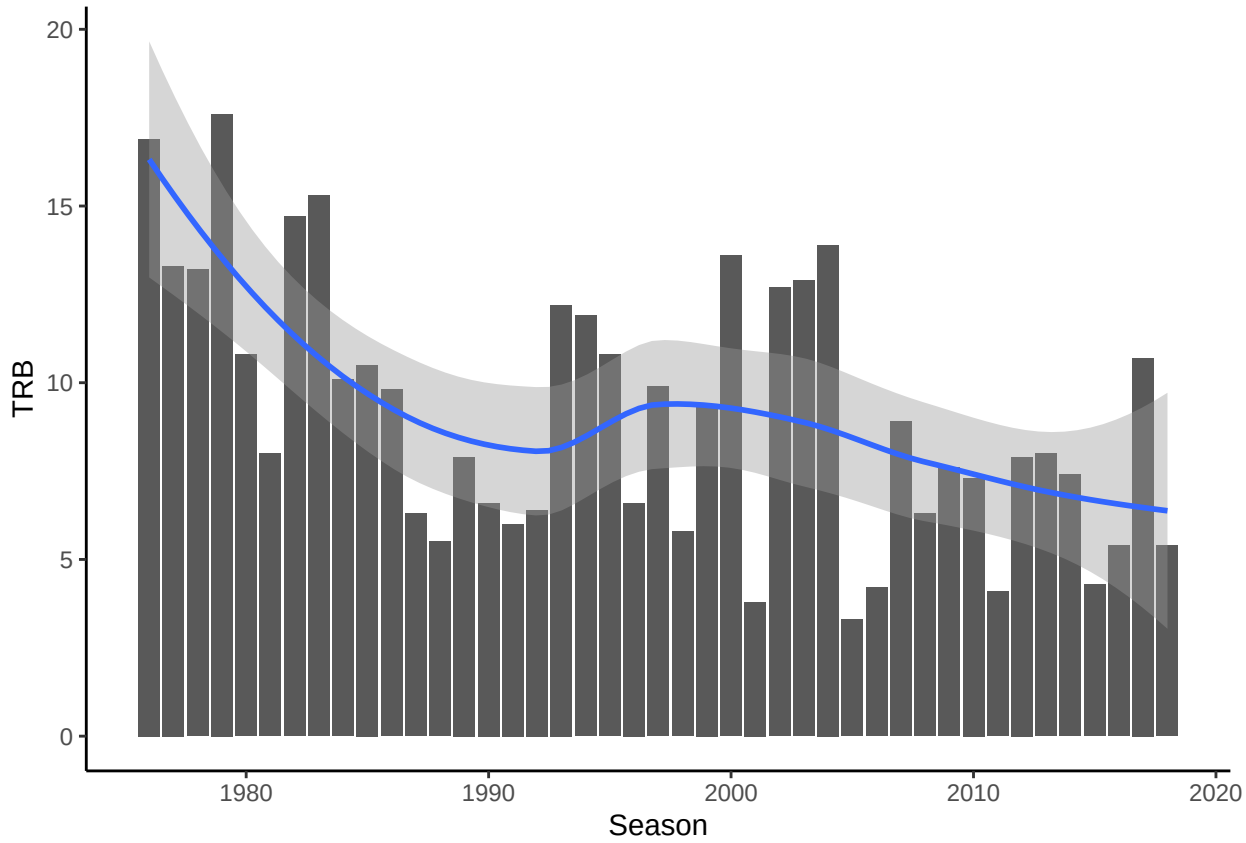
```
ggplot(mvp_data, aes(x = Season, y = PTS)) +
  geom_col() + geom_smooth(method = "loess", formula = y ~x) + theme_classic()
```



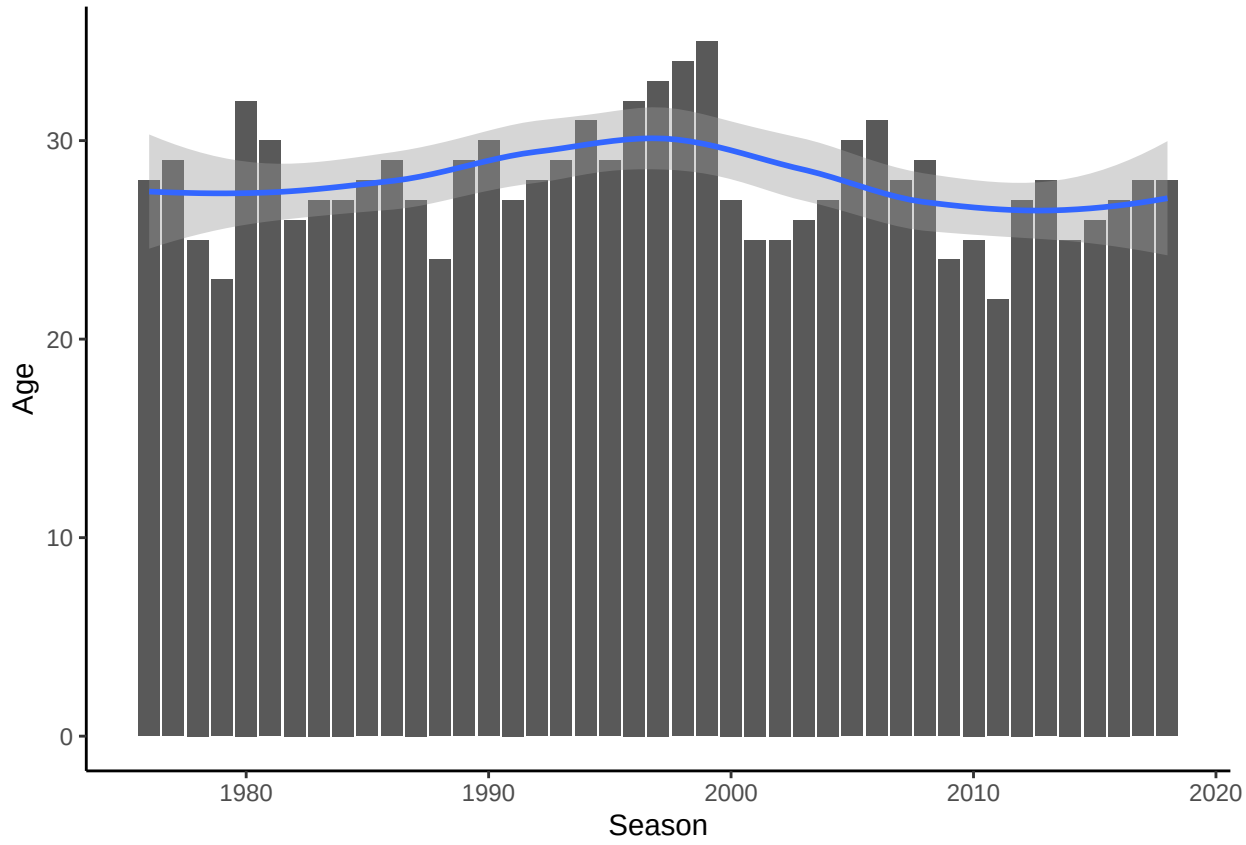

```
ggplot(mvp_data, aes(x = Season, y = AST)) +
  geom_col() + geom_smooth(method = "loess", formula = y ~x) + theme_classic()
```



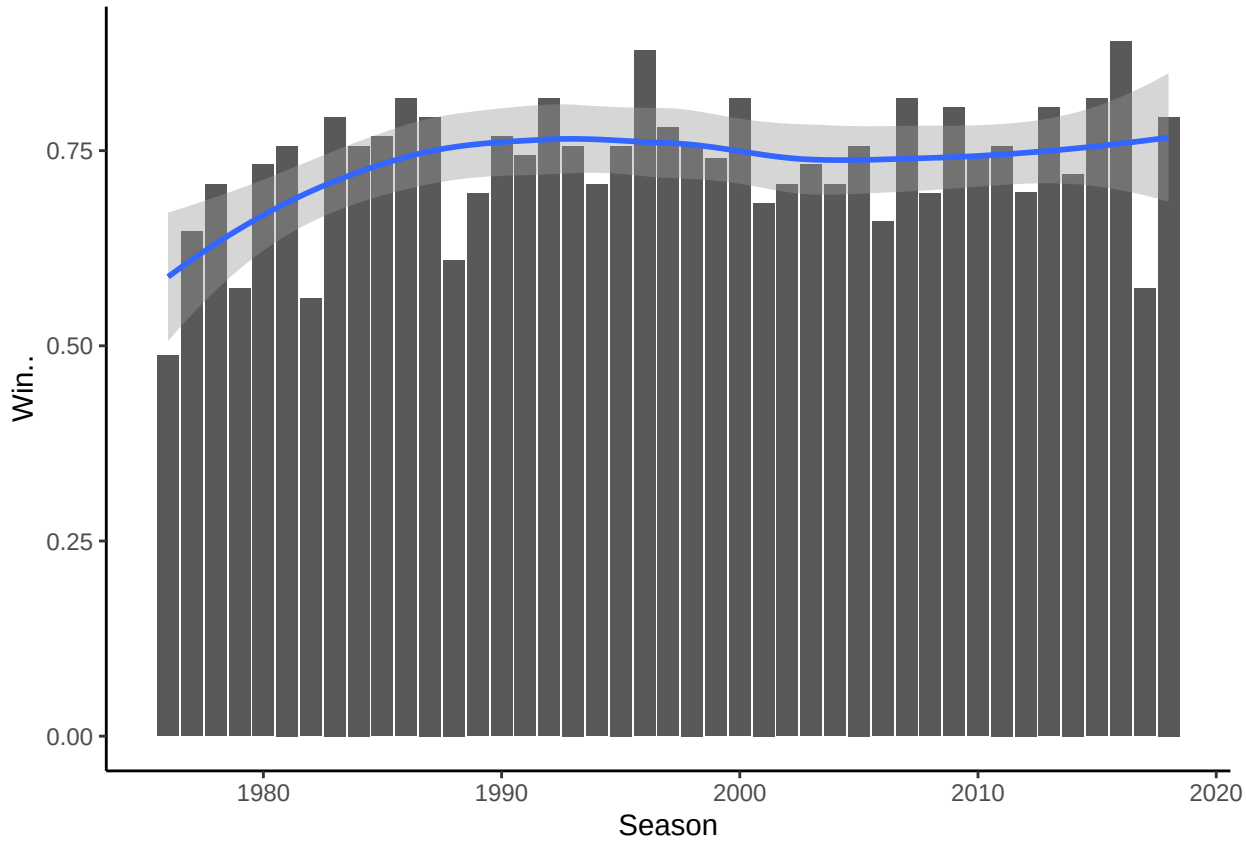

```
ggplot(mvp_data, aes(x = Season, y = TRB)) +
  geom_col() + geom_smooth(method = "loess", formula = y ~x) + theme_classic()
```



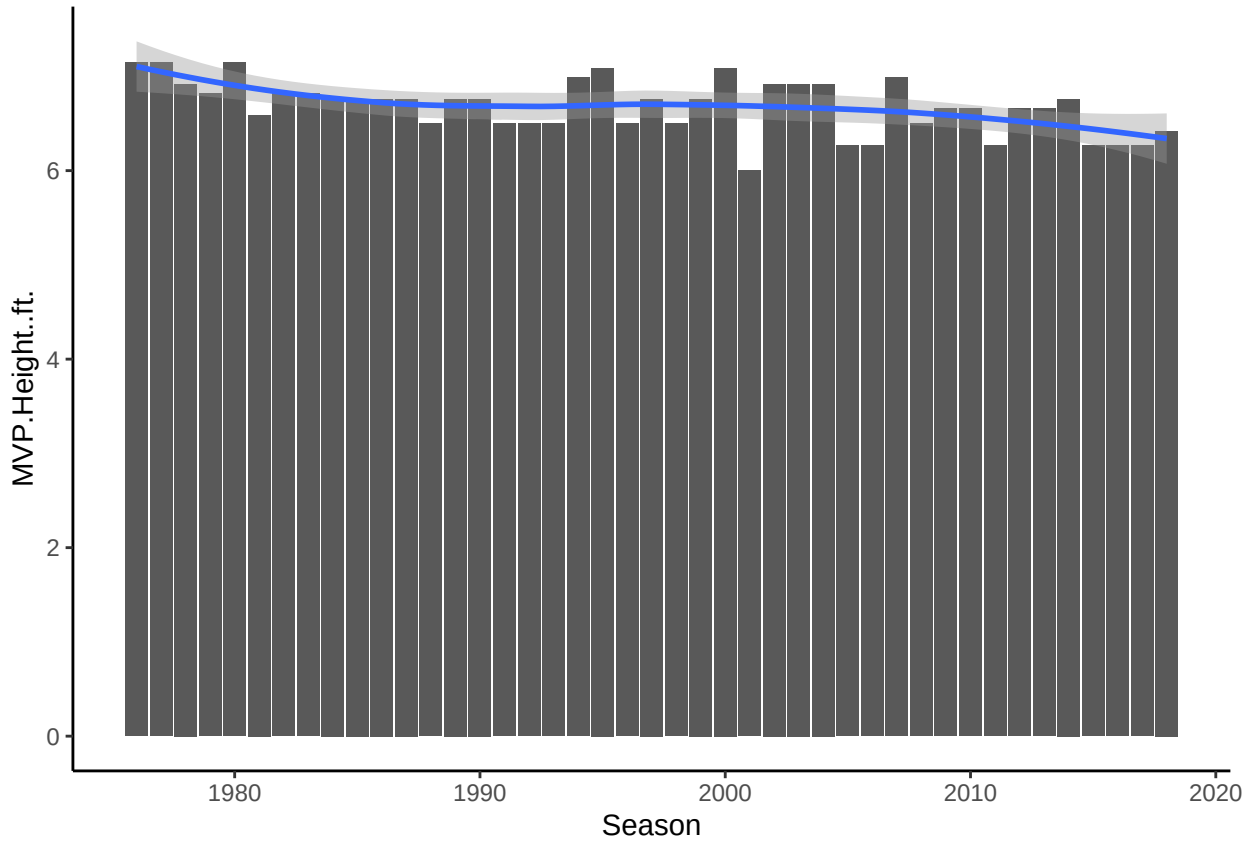

```
ggplot(mvp_data, aes(x = Season, y = Age)) +
  geom_col() + geom_smooth(method = "loess", formula = y ~x) + theme_classic()
```



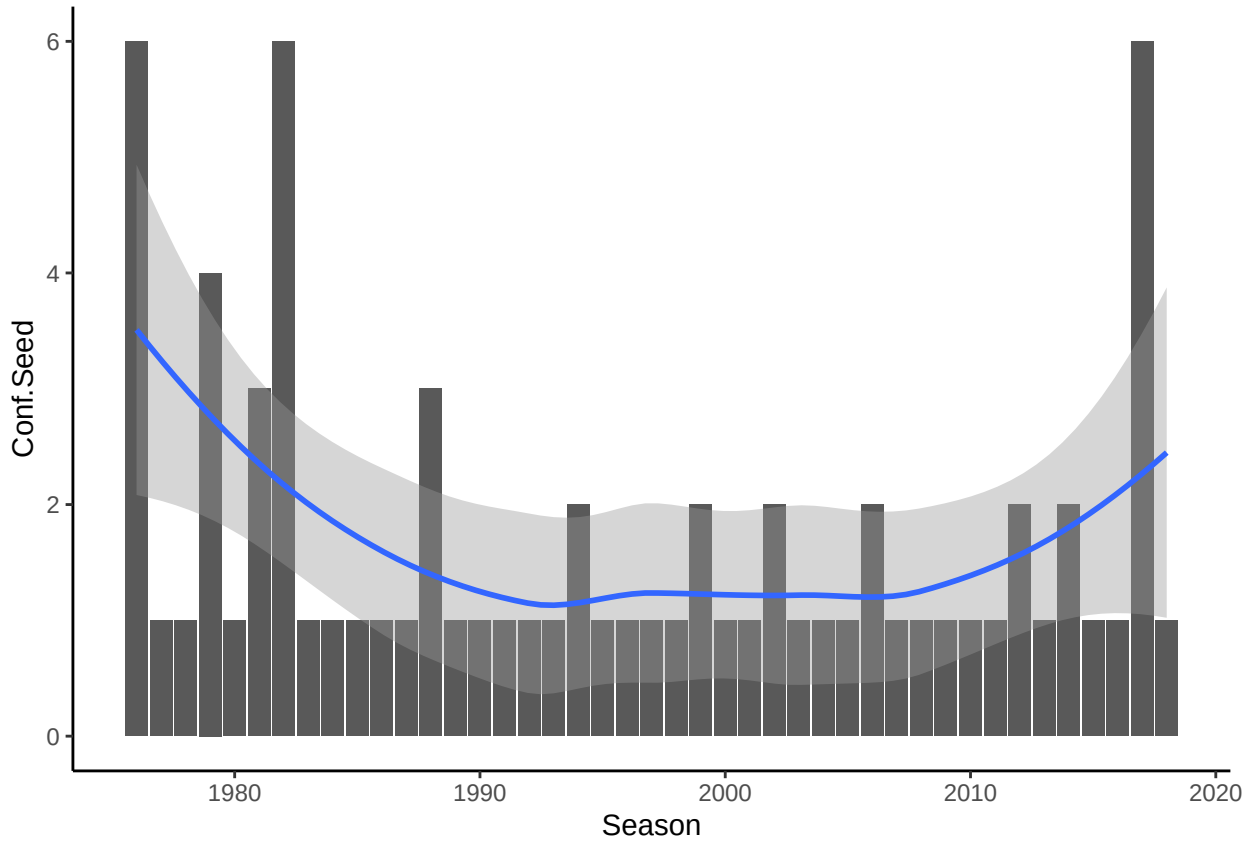

```
ggplot(mvp_data, aes(x = Season, y = Win..)) +
  geom_col() + geom_smooth(method = "loess", formula = y ~x) + theme_classic()
```



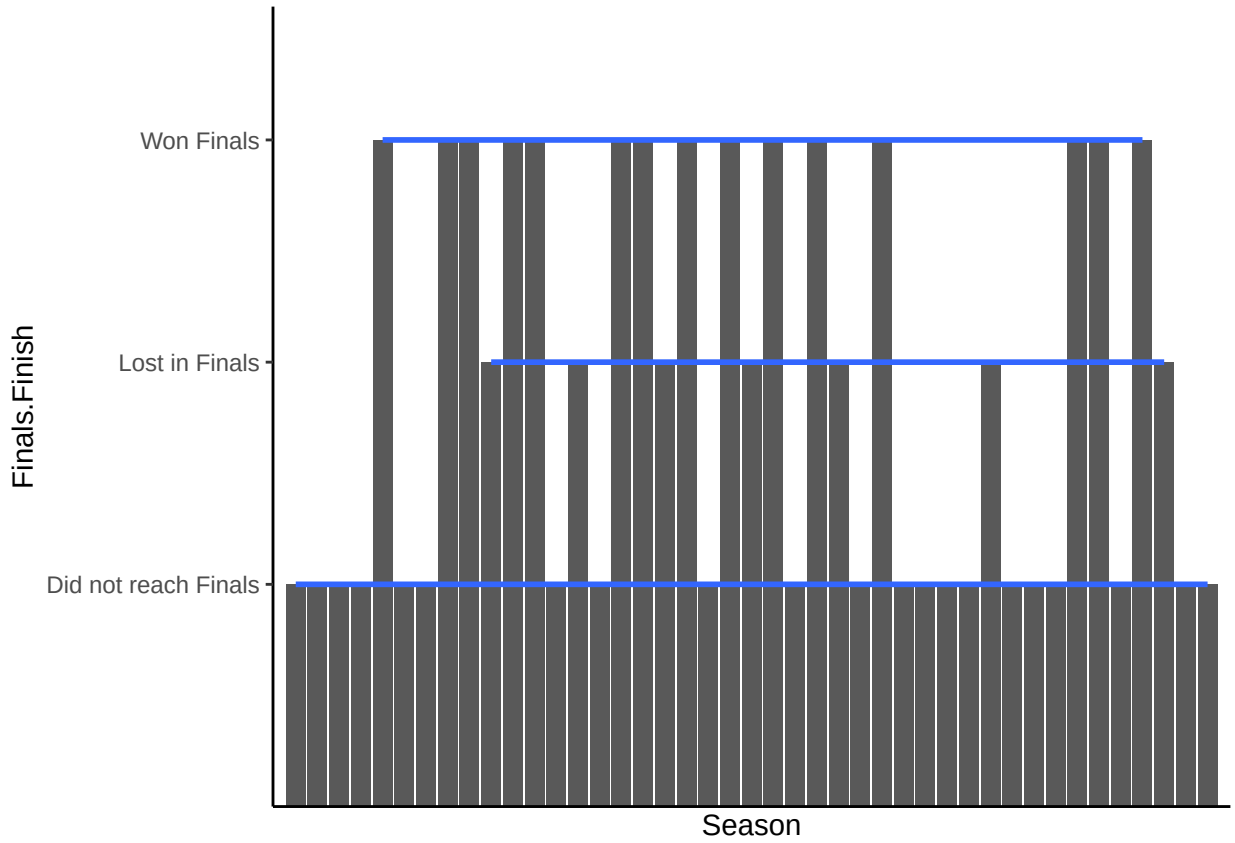

```
ggplot(mvp_data, aes(x = Season, y = MVP.Height..ft.)) +
  geom_col() + geom_smooth(method = "loess", formula = y ~x) + theme_classic()
```



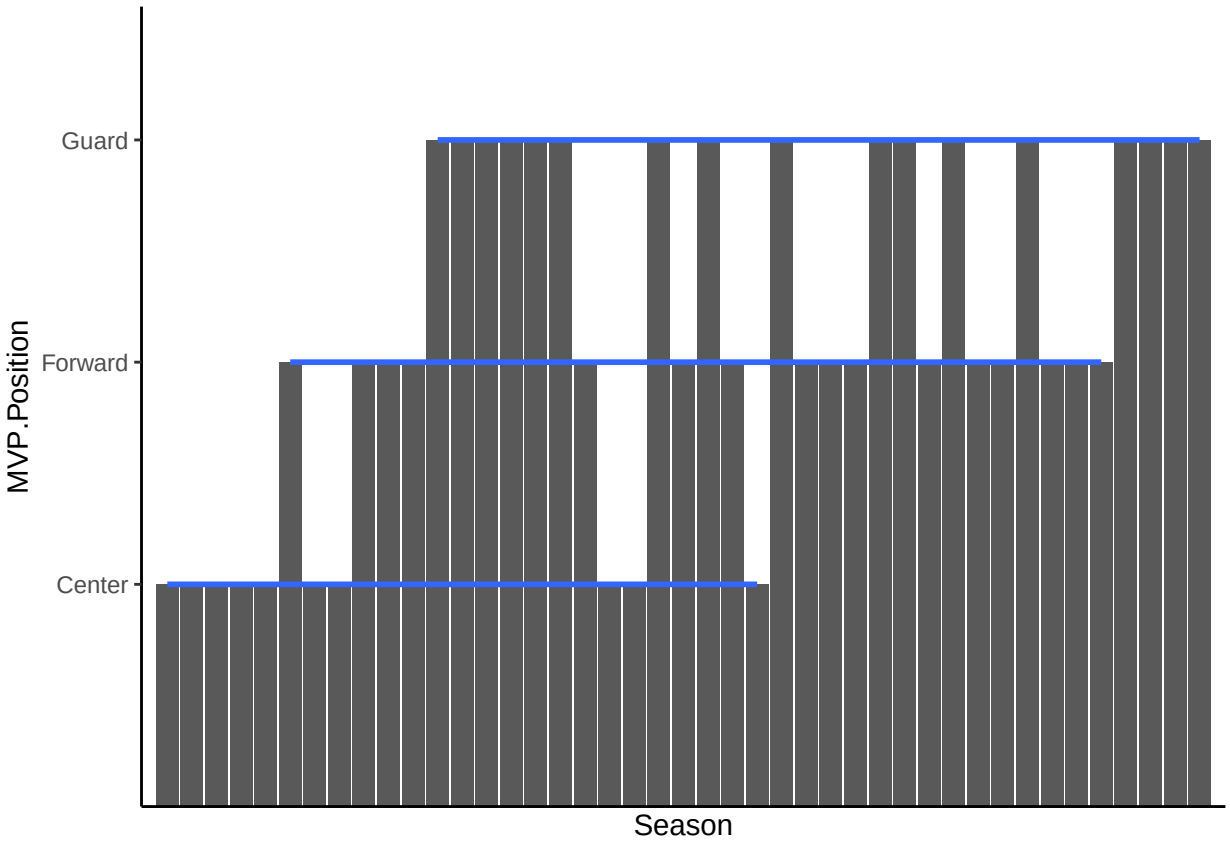

```
ggplot(mvp_data, aes(x = Season, y = Conf.Seed)) +
  geom_col() + geom_smooth(method = "loess", formula = y ~x) + theme_classic()
```




```
ggplot(mvp_data, aes(x = Season, y = Finals.Finish)) +
  scale_x_discrete() + geom_col() + geom_smooth(method = "loess", formula = y ~x) +
  theme_classic()
```




```
ggplot(mvp_data, aes(x = Season, y = MVP.Position)) +
  scale_x_discrete() + geom_col() + geom_smooth(method = "loess", formula = y ~x) +
  theme_classic()
```



Over the decades, some MVP trends have remained constant and some have changed significantly. In terms of scoring, there was a gradual increase into the late 90's and then a dip around the 2000's. Scoring has been rapidly increasing since 2010.

Assist rates rose constantly and then plummeted in the late 90's. They have been gradually increasing ever since.

Rebounding started off very high and was in constant decline until the late 90's were there was a slight resurgence. It has continued to fall ever since.

MVP age has remained relatively constant accross every decade, however there was a slight increase in age during the late 90's.

Win Percentage initially started off low and has plateaued after rising slightly in the 80's.

MVP height has very slightly decreased over time.

Seeding after the early 80's has been relatively constant at 1st or occaisionally 2nd place except for recently when a 6th seeded team produced an MVP in 2017.

There does not appear to be much variation in playoff finishes among MVPs across the decades, but there is one glaring gap in the data. Beginning in the early 2000's, there was a span of around 10 years where the MVP did not win a championship. In fact, only one time throughout that span did an MVP even reach the Finals. Throughout every other decade, there was typically a one or two year gap at most between MVPs winning championships.

Initially, centers most commonly won MVP. In the late 80s to early 90s, it became a guard dominated award. Ever since then, it has been won mostly by forwards with an occaisional guard winning the award. It should be noted that the last 4 winners were guards and a center has not won the award in nearly two decades (since 2000).

What all these trends show us is that the NBA has become a guard-dominated league. In the early 80's, the MVP was an award for bigger players but as time passed, rebounding averages went down, assist averages went up, and height declined. These are all characteristics of guards. There appeared to be a slight resurgence in the late 90s and early 2000s for centers and bigger players as we saw factors like height, age, and rebounding go up and factors like assisting and free throw percentage go down. After this slight renaissance, these trends vanished and a center has not won MVP in nearly 20 years. It clearly seems as though guards who can shoot, pass, and space the floor are becoming more valuable than bigger players who may be seen as 'one dimensional' and can only rebound and score when close to the basket.

Do MVPs get picked because of their statistical achievements or rather for their contributions and overall success of their team? Examine relevant variables.

To answer, we compared average regular season wins with playoff finishes. We also examined two unique cases in regards to statistical achievement and age (Russell Westbrook and Derrick Rose).

```
table(mvp_data$Conf.Seed)
```

```
##
##  1  2  3  4  6
## 31  6  2  1  3
```

```
table(mvp_data$Finals.Finish)
```

```
##
## Did not reach Finals       Lost in Finals           Won Finals
##                   21                    7                   15
```

```
mean(mvp_data$Win..)*82
```

```
## [1] 60.20326
```

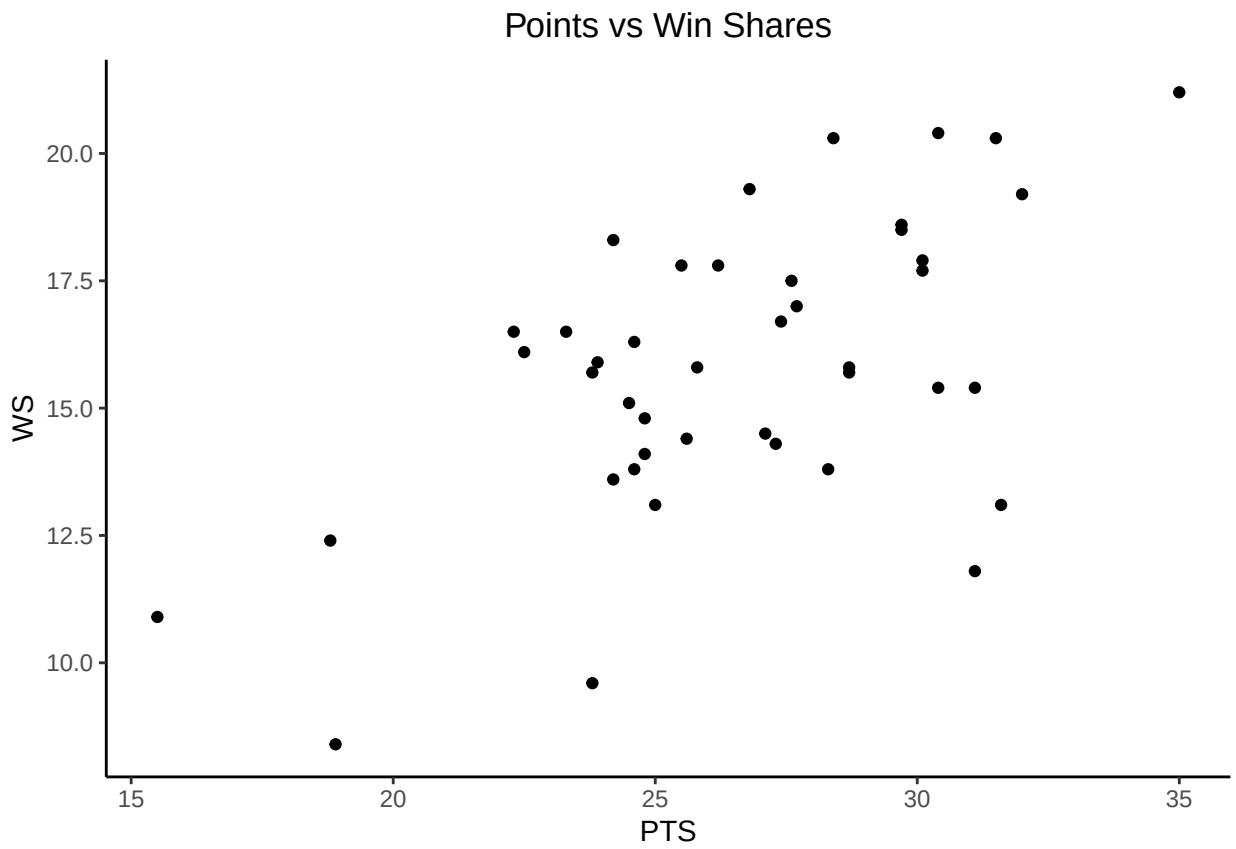
```
mean(mvp_data$Conf.Seed)
```

```
## [1] 1.651163
```

```
mean(mvp_data$WS)
```

```
## [1] 15.84419
```

```
ggplot(mvp_data, aes(x = PTS, y = WS)) +
  geom_point() + theme_classic() +
  ggtitle("Points vs Win Shares") +
  theme(plot.title = element_text(hjust = 0.5))
```




```
mean(mvp_data$WS[mvp_data$Finals.Finish == "Did not reach Finals"])
```

```
## [1] 15.55238
```

```
mean(mvp_data$WS[mvp_data$Finals.Finish == "Lost in Finals"])
```

```
## [1] 15.2
```

```
mean(mvp_data$WS[mvp_data$Finals.Finish == "Won Finals"])
```

```
## [1] 16.55333
```

```
mvp_data$Player[mvp_data$PTS > 10 & mvp_data$TRB > 10 & mvp_data$AST > 10]
```

```
## [1] Russell Westbrook
## 24 Levels: Allen Iverson Bill Walton Charles Barkley ... Tim Duncan
```

```
mvp_data$Conf.Seed[mvp_data$PTS > 10 & mvp_data$TRB > 10 & mvp_data$AST > 10]
```

```
## [1] 6
```

```
mvp_data$Finals.Finish[mvp_data$PTS > 10 & mvp_data$TRB > 10 & mvp_data$AST > 10]
```

```
## [1] Did not reach Finals
## Levels: Did not reach Finals Lost in Finals Won Finals
```

```
mvp_data$Player[mvp_data$Age == min(mvp_data$Age)]
```

```
## [1] Derrick Rose
## 24 Levels: Allen Iverson Bill Walton Charles Barkley ... Tim Duncan
```

```r
d_rose_df <- data.frame(mvp_data$Player[mvp_data$Season > 2008 & mvp_data$Season < 2014],
                        mvp_data$Age[mvp_data$Season > 2008 & mvp_data$Season < 2014],
                        mvp_data$Conf.Seed[mvp_data$Season > 2008 & mvp_data$Season < 2014],
                        mvp_data$Win..[mvp_data$Season > 2008 & mvp_data$Season < 2014],
                        mvp_data$PTS[mvp_data$Season > 2008 & mvp_data$Season < 2014])

names(d_rose_df)[names(d_rose_df) ==
                   "mvp_data.Player.mvp_data.Season...2008...mvp_data.Season...2014."] <-
  "Player"

names(d_rose_df)[names(d_rose_df) ==
                   "mvp_data.Age.mvp_data.Season...2008...mvp_data.Season...2014."] <-
  "Age"

names(d_rose_df)[names(d_rose_df) ==
                   "mvp_data.Conf.Seed.mvp_data.Season...2008...mvp_data.Season..."] <-
  "Conf Seed"

names(d_rose_df)[names(d_rose_df) ==
                   "mvp_data.Win...mvp_data.Season...2008...mvp_data.Season...2014."] <-
  "Win %"

names(d_rose_df)[names(d_rose_df) ==
                   "mvp_data.PTS.mvp_data.Season...2008...mvp_data.Season...2014."] <-
  "Pts"

d_rose_df
```

```
##          Player Age Conf Seed Win %  Pts
## 1 LeBron James  27         2 0.697 27.1
## 2 LeBron James  25         1 0.744 29.7
## 3 Derrick Rose  22         1 0.756 25.0
## 4 LeBron James  28         1 0.805 26.8
## 5 LeBron James  24         1 0.805 28.4
```

Since 1976, most MVP winners have been seeded one or two in their conference, with the infrequent winner being a lower seed. The average wins per season is 60, which is a very high amount. In addition to this, the average conference seed is 1.7, which means MVPs are regularly expected to finish near the top of their conference. All of this evidence points to the MVP being awarded on the basis of team success, however there does appear to be contradictory data.

Win Shares, a metric that measures the total regular season wins a player is responsible for, has an average of 15.8 among MVP winners. Additionally, there appears to be a moderate, positive correlation between Win Shares and PPG. Compared to the league average of 4.1 Win Shares, this indicates that MVPs get recognized when they put up large Win Share values, even if their team is not 'successful'. This is evidenced by the fact that only about 1/3 of MVP winners win the Finals while the majority do not even reach the Finals.

Over the course of NBA history, only one player has averaged a triple double (at least 10 ppg, rpg, and apg) over the course of an entire season, Russell Westbrook. He did it in the 2017 season and was on the first 6th seed team to produce an MVP in nearly 40 years. His team did not reach the Finals. Ironically, he did the exact same thing the very next season, yet he was not even considered for the MVP award.

The youngest MVP in league history was Derrick Rose, who won the award at age 22. He did so against LeBron James even though James posted much more impressive numbers across the board. Rose's win came off of two consecutive MVP seasons from LeBron, and was prior to two more consecutive MVP seasons from LeBron. As a result, it has been speculated that Rose's MVP was awarded simply due to voter fatigue.

Voters got tired of seeing James win every year and were eager to give the award to someone different.

Conclusion

The MVP is an award that demands a lot in terms of individual statistics, but not as much from a team past the regular season. Do not count on winning if you average fewer than 20 points. Additionally, an MVP should display some adeptness in other statistical facets such as rebounding and assists. The MVP will usually be on a number 1 seeded team, unless they drag an otherwise pitiful team to a better-than-expected season. Furthermore, the MVP in today's NBA will very likely not be a center, highlighting the league's admiration for guards and forwards who can lead their teams and have much more control over the outcome of each game.

It appears that the MVP award factors in team success, yet a team does not have to perform superbly for one of its players to win the MVP. In most cases, the MVP will not even reach the NBA Finals. Usually, the team is expected to have a great regular season with many wins and a high seed, but not much more is expected beyond that. When it comes to Win Shares, they rise as player points rise, meaning that MVPs are contributing to their team's success by bolstering their own statistical numbers, potentially giving off the impression that MVPs are team players when this may not always be the case. Additionally, if a player is on the brink of doing something historic such as Russell Westbrook or Derrick Rose, the success of their team becomes even less of a factor in MVP consideration. These observations make the MVP a largely individual award, but not an entirely individual award as some team success is needed to justify giving the award to a particular player.