

Exploring NBA Historical Data 1977-2020

Uziel Rios

2022-09-13

Introducing the dataset

This data set includes statistics of every NBA player from each team from the 1977-2020 (courtesy of FiveThirtyEight) Additionally, a table of FiveThirtyEight's glossary will accompany what each category represents. Don't worry, I did not utilize nor understand every category. You may notice that the five major stats are per 36 minutes instead of per game. One possible reason FiveThirtyEight measures this is to evaluate the player's performance on a per-minute basis. Plus, most starters average roughly 36-37 minutes per game, so rounding it to 36 minutes gives a decent representation of the player's production on the court.

```
nba_set <- read_csv("./nba-data-historical.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   player_id = col_character(),
##   name_common = col_character(),
##   pos = col_character(),
##   team_id = col_character(),
##   franch_id = col_character()
## )

## See spec(...) for full column specifications.
```

```
knitr::include_graphics("./Categories.png")
```

Category	Description
player_id	Basketball-Reference.com player ID
name_common	Name of player
year_id	Season (i.e., 2019-20 is "2000")
type	Regular season (RS) or playoffs (PO)
age	Age on Feb. 1 of season
team_id	Team played for
pos	Primary position
tmRtg	Team efficiency margin
franch_id	Franchise played for
G	Games played
Min	Minutes played
MP%	Share of team minutes played
MPG	Minutes per game
P/36	Pace-adjusted points per 36 minutes
TS%	True Shooting Percentage
A/36	Pace-adjusted assists per 36 minutes
R/36	Pace-adjusted rebounds per 36 minutes
SB/36	Pace-adjusted steals plus blocks per 36 minutes
TO/36	Pace-adjusted turnovers per 36 minutes
Raptor O	Offensive RAPTOR rating
Raptor D	Defensive RAPTOR rating
Raptor+/-	Total RAPTOR rating
Raptor WAR	Total RAPTOR wins above replacement
PIE%	Player Impact Estimate
AWS%	PIE% using Alternate Win Score
USG%	Usage Rate
AST%	Assist Rate
TOV%	Turnover Rate
ORB%	Offensive Rebound Rate
DRB%	Defensive Rebound Rate
TRB%	Total Rebound Rate
STL%	Steal Rate
BLK%	Block Rate
ORtg	Offensive Rating (points produced per 100 poss.)
%Pos	Share of team possessions used
DRtg	Defensive Rating (points allowed per 100 poss.)
2P%	2-point field goal percentage
3P%	3-point field goal percentage
FT%	Free throw percentage
3PAr	Share of field goal attempts that were 3-pointers
FTAr	Ratio of free throw attempts to field goal attempts
Pace +/-	Player's effect on team pace

Determining which Position on Average Scored the Most Points per Decade The first research question is to find out which position scores the most points on average from the decades 1980-2020. I did this by subsetting the data set where we only consider the five main positions (Point Guard (PG), Shooting Guard(SG), Small Forward(SF), Power Forward (PF), and Center(C)). This is to only consider the players who purely play on one of these positions as some have played multiple positions.

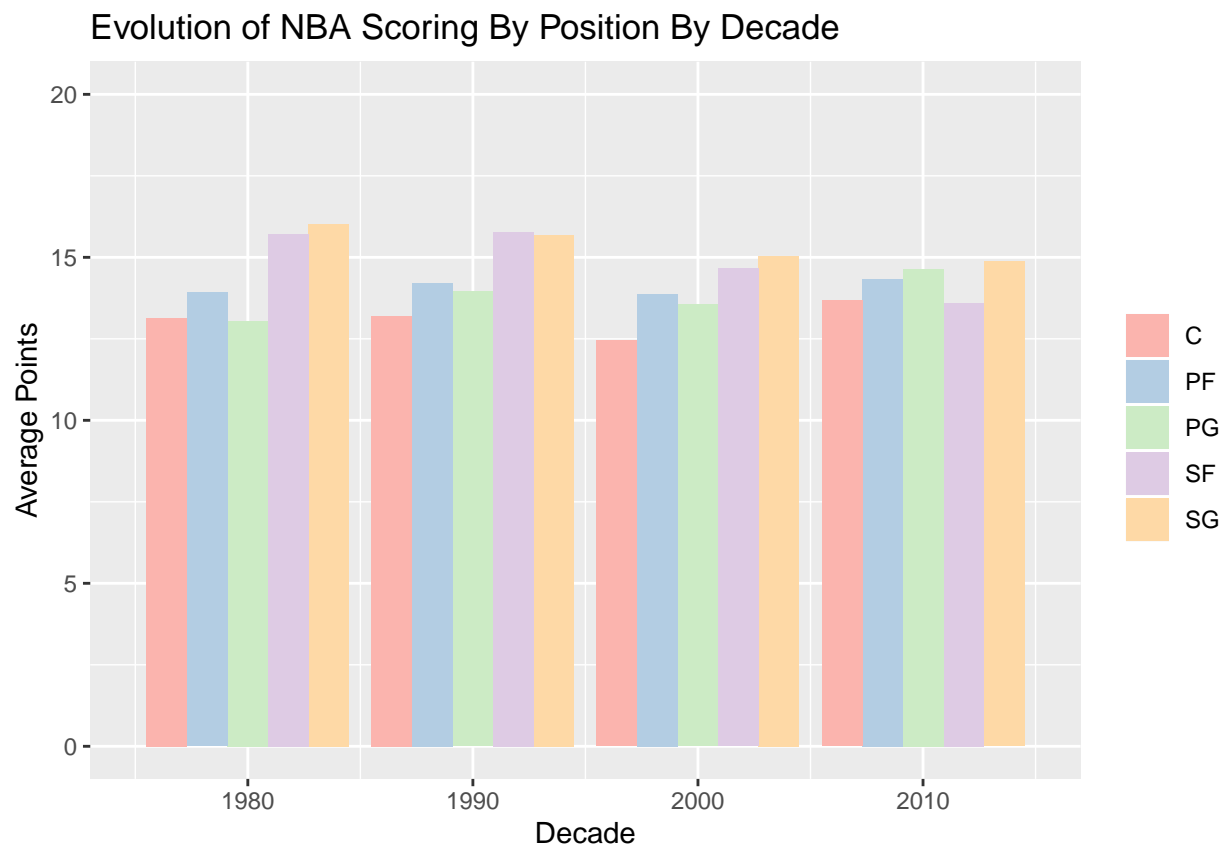
```
nba_set %>%
  select(year_id, pos, `P/36`) %>%
  subset(pos %in% c("PG", "SG", "SF", "PF", "C")) %>%
  mutate(decade = floor(year_id/10) * 10) %>%
  group_by(decade, pos) %>%
  summarise(avg_pos_pts = mean(`P/36`, na.rm = TRUE)) %>%

  ggplot(aes(decade, avg_pos_pts)) +

  geom_bar(aes(fill = pos), position = "dodge", stat = "identity") +
  labs(title = "Evolution of NBA Scoring By Position By Decade", x = "Decade", y = "Average Points") +
  xlim(1975, 2015) +
  ylim(0,20) +
  theme(legend.title = element_blank()) +
  scale_fill_brewer(palette = "Pastell1")
```

```
## 'summarise()' regrouping output by 'decade' (override with '.groups' argument)
```

```
## Warning: Removed 10 rows containing missing values (geom_bar).
```



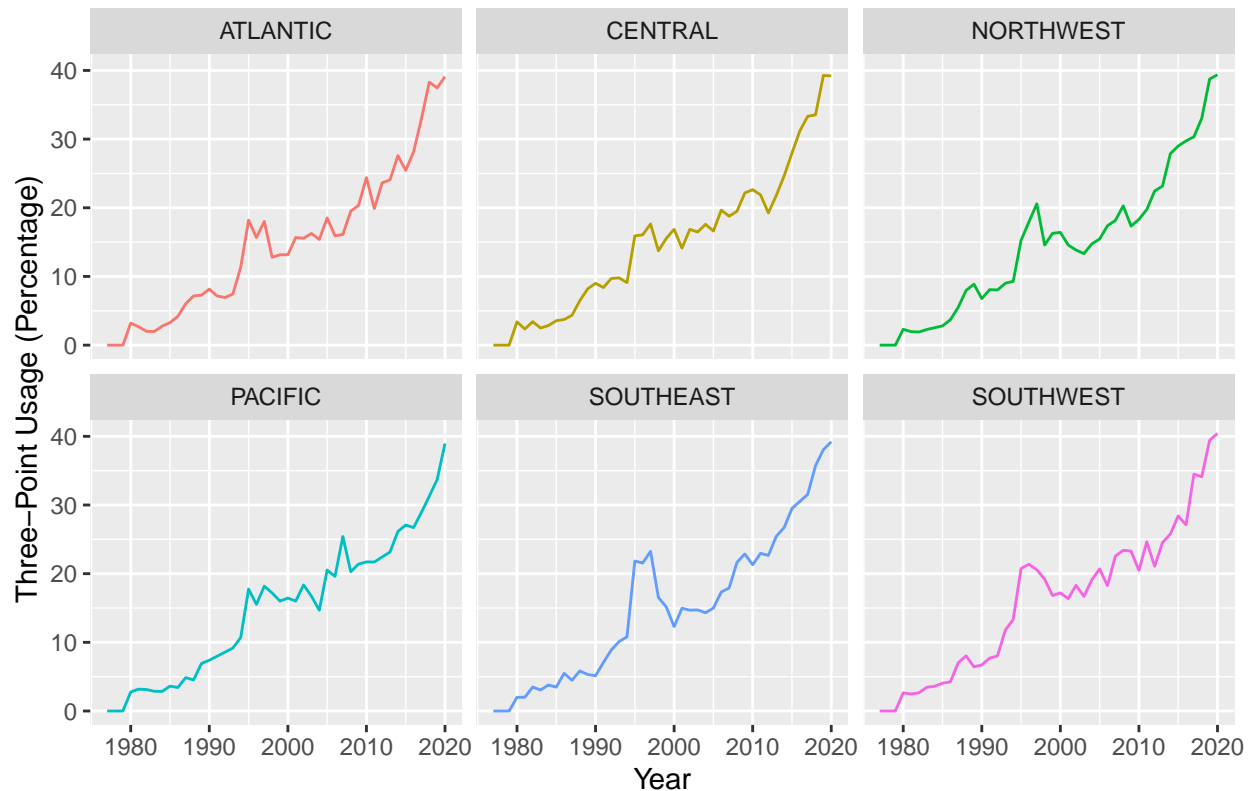
Based on this graph, we can see that there were no instances of drastic change throughout the course of this time period. The scoring average for all positions were always between the range of about 12.5 to 16 points. One trend that can be noticed is that shooting guards and small forwards scored less points on average as each decade passed by, while point guards scored more points. Another trend that is shown is that the power forward position seemed to be the most consistent throughout the four decades, as the scoring average for this position was around 14 points for all four decades. Furthermore, the 2010s is the decade in with the smallest scoring gaps among the five positions, in comparison to the other three decades. To answer the question, there was not much evolution in regards to the scoring averages of each of the five positions. Although the scoring average for shooting guards slightly decreased in every decade, this position held the highest scoring average for every decade (with the exception of the 1990s, where it was edged out slightly by small forwards). On the other hand, centers usually scored the least amounts of points during each decade.

Evolution of the Three-Point Shot The second question for this data set is how has the three-point shooting evolved over time. There are two ways we can go about answering this. One is to group the teams by division (Atlantic, Central, Southeast, Northwest, Pacific, and Southwest). Then, we can measure the average rate of three-point of attempts for each division. We can facet the plot to see how the three point usage has evolved for each division.

```
nba_set %>%
mutate(Division = case_when(franch_id %in% c("MIL", "IND", "CHI", "CLE", "DET") ~ "CENTRAL",
franch_id %in% c("MIA", "ORL", "CHA", "WAS", "ATL") ~ "SOUTHEAST",
franch_id %in% c("DEN", "OKC", "UTA", "POR", "MIN") ~ "NORTHWEST",
franch_id %in% c("LAL", "LAC", "GSW", "SAC", "PHO") ~ "PACIFIC",
franch_id %in% c("HOU", "DAL", "SAS", "MEM", "NOH") ~ "SOUTHWEST",
franch_id %in% c("NYK", "TOR", "BOS", "NJN", "PHI") ~ "ATLANTIC")) %>%
group_by(Division, year_id) %>%
summarize(avg_3Usage = mean(`3PAr`, na.rm = TRUE)) %>%
ggplot() +
geom_line(aes(x = year_id, y = avg_3Usage, color = Division)) +
labs(title = "Average Three-Point Usage by NBA Division",
x = "Year",
y = "Three-Point Usage (Percentage)") +
facet_wrap(~ Division) +
theme(legend.position = "none")
```

```
## 'summarise()' regrouping output by 'Division' (override with '.groups' argument)
```

Average Three-Point Usage by NBA Division



The graph above depicts the average three point usage for each of the six NBA divisions from 1977-2020. It is clear that for all divisions, there is a steep increase in three-point attempts since the introduction of the three-point line in 1979. We see a sizable increased rate in three-point attempts from 20% to up to 40% at around early to mid-2010's. This makes sense as the floor is more spaced and we have players like Stephen Curry, Klay Thompson, and James Harden, who have been some of the contributors of utilizing the three-point shot more often. One thing to point out is that all divisions experienced dips in three point usage in the late 1990s to early 2000s. This is probably due to the more stifling defenses along the perimeter, which will discourage teams from attempting three-pointers. Additionally, centers and forwards were the focal point of offenses, whereas guards run the offense and feeding the ball to their centers and forwards. Since then, all divisions experienced increases in three-point usage, but the Central, Northwest, and Southeast divisions experienced much steeper increases in the mid 2010s, which was when three-point shooting really became mainstream. The second method is to create a heat map that shows the three-point usage for every franchise. Note that some franchises were founded later than others, which is why there are gaps within the heat map.

```
grouped <- nba_set %>%
  filter(!is.na(`3PAr`))
summarise(group_by(grouped, franch_id, year_id), Percent3Attempts = mean(`3PAr`)) %>%
  ggplot(mapping = aes(x = year_id, y = franch_id, fill = Percent3Attempts)) +
  geom_tile() +
  labs(x = "Year",
       y = "Franchise",
       fill = "% of Attempts Were 3 Pointers")
```

```
## 'summarise()' regrouping output by 'franch_id' (override with '.groups' argument)
```

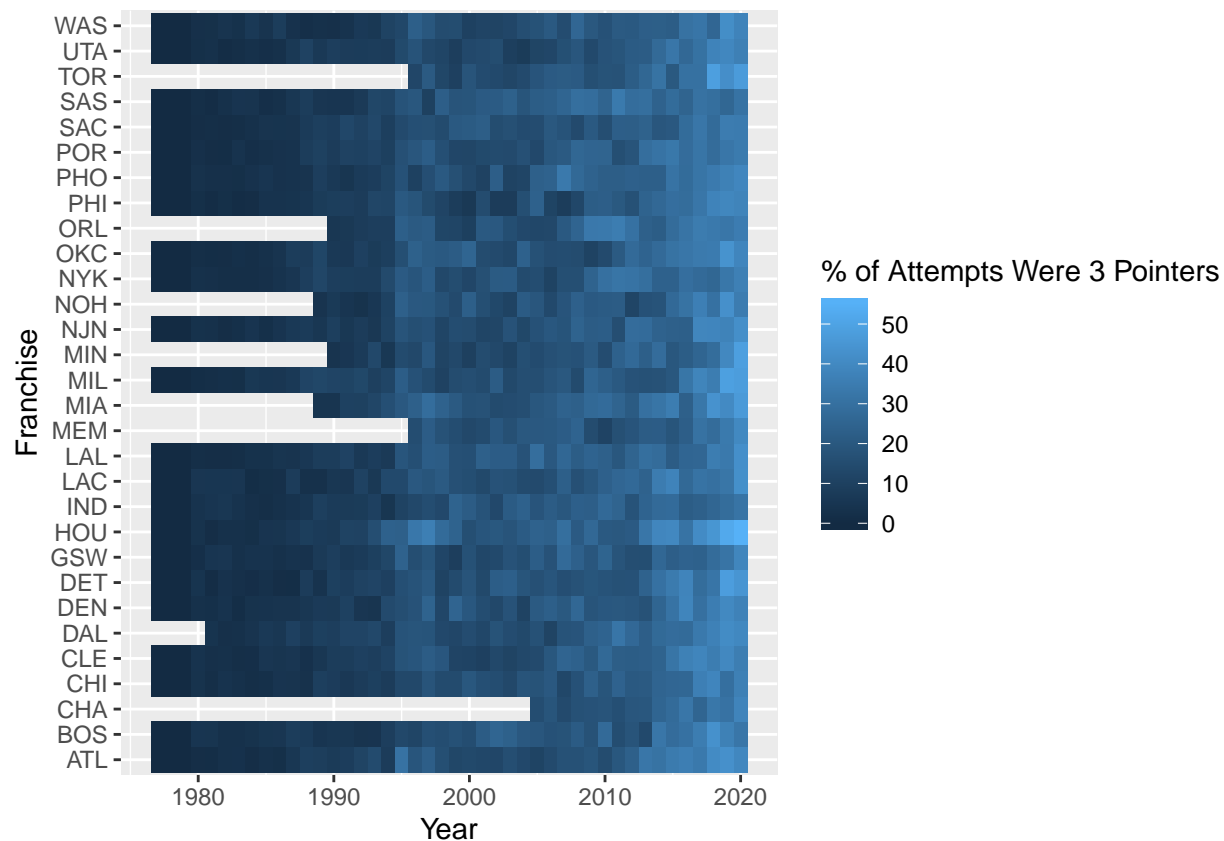


Figure Interpretation: From this figure, we can see that 3 point field goal attempts have increased drastically since the NBA was first founded. Notable franchises that rely heavily on these attempts today include HOU, MIN, and TOR.

So, how has 3 point usage evolved over time? In the beginning, the 3 point shot was virtually nonexistent with a field goal attempt percentage in the low single digits. Over time however, teams and coaches began to realize the efficiency of the shot and, as visualized in both graphs, 3 point shots increased significantly as a result across the board. Today, 3 point shots are a major part of basketball with some teams having nearly as much as 50% of their shots being 3 point attempts.