

NBA stats

The National Basketball Association (NBA) has changed since the introduction of the 3-point line back in the 1979-80 season. In the beginning, teams were reluctantly shooting three pointers while today, players like Steph Curry and James Harden are leading a “3-point revolution”. Basketball went from a “big guy” game to a positionless sport, based on spacing and long-range shots. But what kind of story do the stats tell? Do they agree or do they show something else?

Nomikos Skyllas, nomikos.skyllas@gmail.com, [Github](#), [LinkedIn](#)

Results

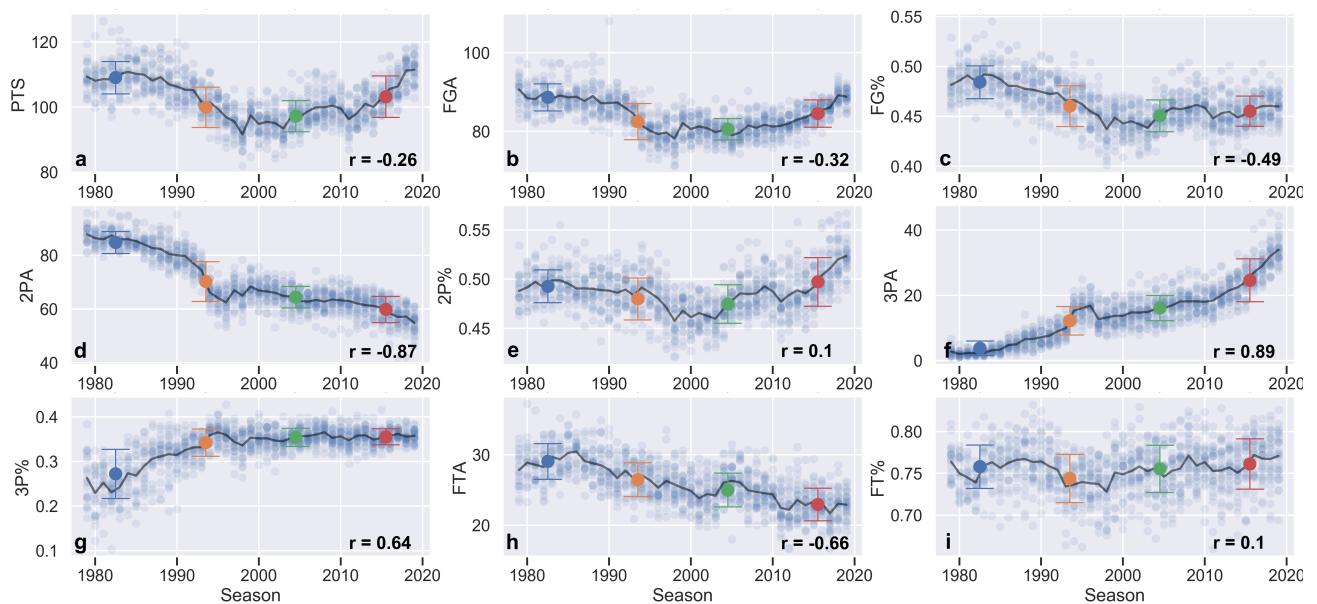


Fig 1: Time series of scoring-related statistical categories of all NBA teams since 1979: a)PTS, b) FGA, c) FG%, d) 2PA , e) 2P%, f) 3PA, g) 3P%, h) FTA and i) FT%. The black line is the mean of each year, the colored dots and error bars depict the mean and the standard deviation of each decade and the blue circles represent individual teams. The correlation coefficient of each category with time (NBA seasons) is plotted at the bottom of each sub figure.

NBA teams during the entire 80s decade were scoring close to 110 points per game (Fig 1a), a number that was reached again in recent years (111.5). In-between there was a constant decrease in scoring (90s) until it reached a minimum of less than 100 points before the 00s. The same pattern was followed by the field goal attempts (Fig 1b, correlation coefficient with points = 0.79), field goal percentage (Fig 1c, correlation coefficient with points = 0.73) and 2-point shooting percentage (Fig 1e, correlation coefficient with points = 0.73). On the other hand, the 2-point attempts (Figure 1d) show a steady decline since 1979, with a strong negative correlation with time ($r = -0.87$). Their standard deviation during the 80s, 00s and 10s is smaller than the 90s one ($\sigma = 3.9$). This is probably because of the sudden drop in 2PA the early 90s, suggesting a transitional decade during which teams tried to adapt their style of play. The 3-point attempts (Figure 1f) show the exact opposite trend: a steady increase with a strong positive correlation with time ($r = 0.89$). What's interesting in this statistical category is a peak during the years 1994, 95 and 96 (16.1 3PA on average), when the 3-point line was brought closer to the basket (22 feet). This general increase in attempts is not followed by the 3-point shot percentages (Figure 1g), as they reached the level of 35% during mid-90s and remained stable since then, suggesting that the players were already good shooters in the 90s but they did not attempt as many 3-point shots as they do nowadays. The standard deviation of 3-point percentage in the 80s ($\sigma = 0.06$) was larger than the other decades, indicating that a big difference existed between good and bad-shooting teams. Free throw attempts (Figure 1h) show a decrease of almost 7 attempts per game (21%) since the 80s while at the same time kept a stable shooting percentage of about 75% ((Figure 1i)).

During the 80s, teams were scoring more points, were having more field goal attempts with percentages close to 50% and equally high percentages for the 2-point shots. During the 90s a decrease in all of the aforementioned categories is apparent, reaching its lowest point in the 00s. During that decade, teams were scoring fewer points, taking fewer shots and had lower shooting percentages. The 2-point attempts were not at their lowest point but they had the worst percentages in the past 41 years. 3-point percentages were as high as today, but fewer attempts were made. In the past decade (10s), teams scored once again more points (and field goals in general) but relied more on 3-point shots in comparison to the past, while at the same time they shot fewer 2-pointers but under better conditions (higher percentages).

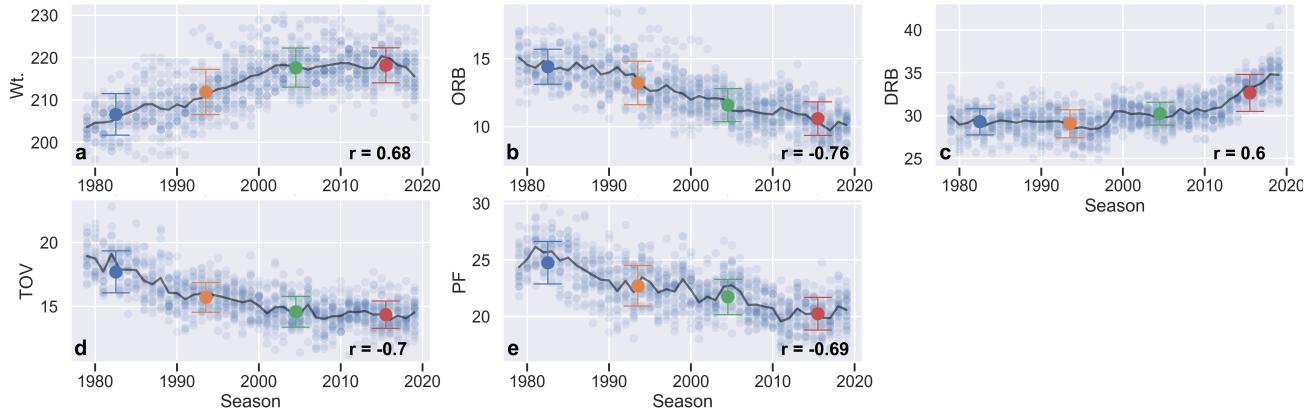


Fig 2: Time series of non scoring-related statistical categories of all NBA teams since 1979: a) Wt., b) ORB, c) DRB, d) TOV and e) PF. The black line is the mean of each year, the colored dots and error bars depict the mean and the standard deviation of each decade and the blue circles represent individual teams. The correlation coefficient of each category with time (NBA seasons) is plotted at the bottom of each sub figure.

The non scoring-related statistical categories in Fig 2 showed a positive or negative correlation with time (seasons), which was larger than 0.6 ($|r| > 0.6$). It is interesting to see that the average weight of the teams (Fig 2a) increased from 203.6 pounds in 1979 to 215.6 pounds in 2019-20. Along with the average weight, defensive rebounds (Fig 2c) showed a positive correlation ($r > 0.6$) with season. More specifically they started increasing in the late 90s from around 30 per game to almost 35 recently. In the 10s, they show a larger standard deviation ($\sigma = 2.2$) than in the previous decades, suggesting that this might be a transitional decade for the specific category (as the 90s were for the 2PA). The rest of the statistical categories, offensive rebounds, turnovers and personal fouls, showed negative correlation ($r < -0.6$) with season. Seemingly, fewer fouls lead to fewer free throws, and shooting more long-range shots in the recent years, possibly lead to fewer offensive and more defensive rebounds.

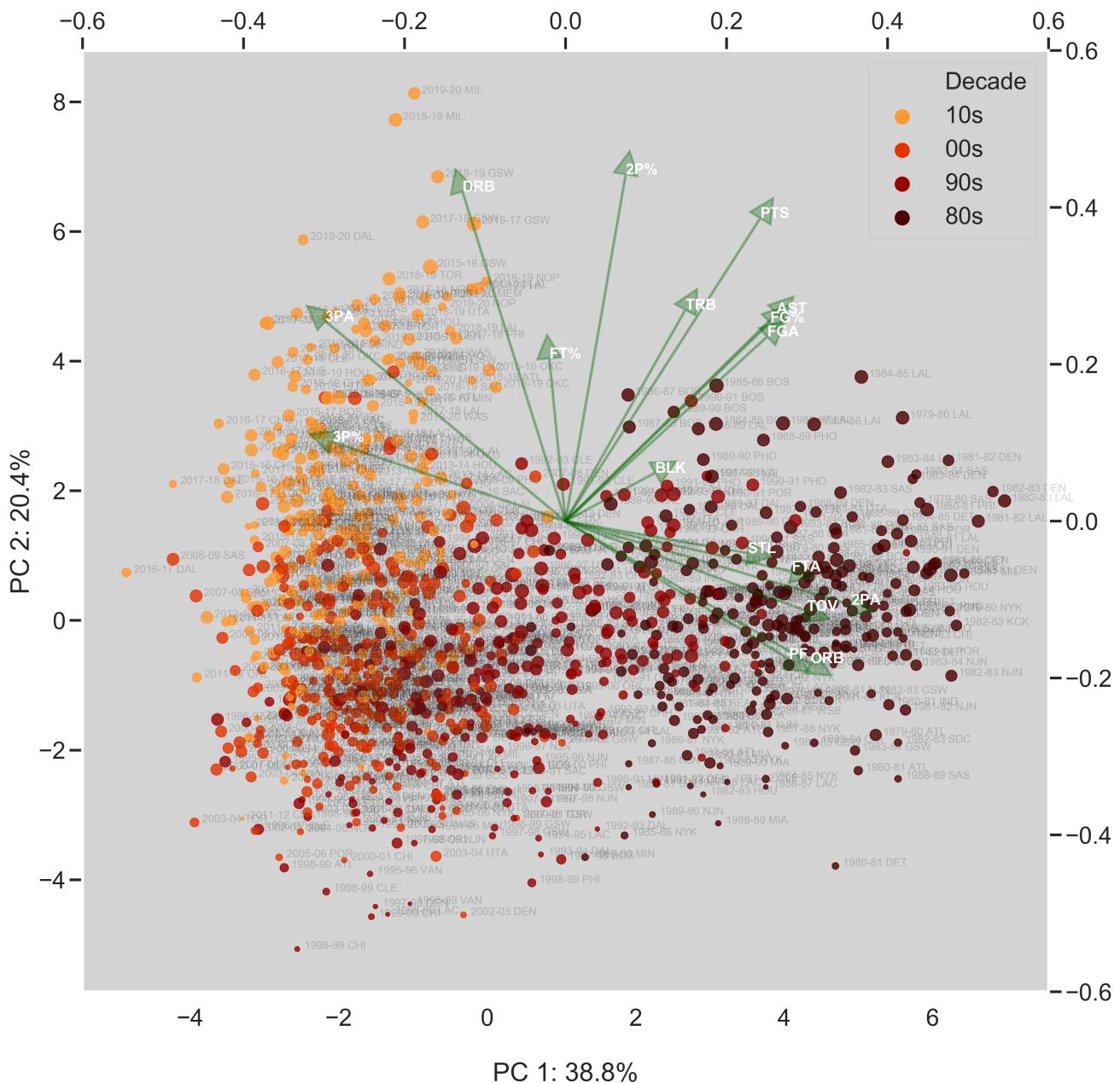


Fig 3: Principal Component Analysis (PCA) of 17 stat categories for all NBA teams since season 1979-80 and on. The size of each dot represents the amount of wins during the regular season (teams with more wins have larger dots) and the color depicts the decade of the team (see plot legend).

Principal component 1 (Fig 3) accounts for 38.8% of the variation in the data (teams). Its largest loadings come from 2PA, 3PA, 3P%, TOV and ORB, meaning that the variation of teams along the x-axis is mainly due to the changes in the aforementioned statistical categories. Along the x-axis, the teams separate mainly based on their decade, as the colors indicate. The 80s teams, relying mainly on 2-point shots committing more personal fouls, turning the ball over more often and collecting more offensive rebounds, cluster on the right side of the plot while the newest teams (00s and 10s), shooting more 3-point shots with higher percentages, cluster on the left side. The 90s teams are in-between, in the lower, middle part of the plot. Principal component 2 explains 20.4% of the variance in the data and separates the modern (10s) teams from the ones of the previous decade (00s), along the y-axis. Its highest loadings come from DRB, 2P% and PTS, meaning that in the last decade teams collected more defensive rebounds, improved their 2-point shot percentages and scored more points. The 80s teams are also higher up on the y-axis suggesting that they also had higher numbers in the aforementioned three categories compared to the 90s and 00s teams.

Among the teams worth highlighting are the recent Budenholzer-lead Bucks, that are clearly separating from the rest of the NBA teams. They are on the upper extreme of the plot, even further than the recent Warriors dynasty. The mid-00s Phoenix Suns seem like a pioneer team for its time, separating from the group of the other 00s teams and moving to the modern 3-point dominated era earlier than the rest, only to be followed almost ten years later by teams like the Rockets and the Warriors. The late 90s - early 00s Chicago Bulls (after Jordan's retirement) are on the lower left extreme of the plot, separating from the other 90s - 00s teams, by having very few points per game and wins per regular season.

Conclusions

The time series suggest that the NBA has moved from a high-scoring, 2-Point and offensive rebound-dominated era with many turnovers and personal fouls to another high-scoring era (10s). This new era is dominated by 3-point shots and defensive rebounds. In between, there was a period with fewer points, fewer shots and worse shooting percentages (from mid-90s to late 00s). The data seem to follow the narrative of the "big guy" era in the past, the "3-point revolution" that we are witnessing today and the mid-range era in between. This is better visualized in the PCA plot where the lower right side of the plot is an "ecosystem" in which "big guys" can thrive by playing closer to the basket, shooting close-range 2-point shots with high percentages, collect more offensive rebounds and play tough defense (steals, personal fouls and turnovers). The lower left side is a transitional (between the 80s and the 10s) "ecosystem" in which teams with lower scores and lower 2P% gather. This might indicate less efficient shooting decisions, possibly mid-range shots. The upper left side is "populated" by modern teams, with more 3-point attempts, better 3-point and 2-point percentages, that collect more defensive rebounds.

Basketball is a beautiful and constantly evolving game and players adapt and dominate in different eras (or "ecosystems") by developing the necessary skills. Consequently, a comparison between players from different decades, in order to find the "Greatest Of All Time" might not be valid.

Methods

The year-by-year, team stats per game of all NBA teams since season 1979-80 until the (uncompleted) season 2019-20 were collected. The data were scraped from [Basketball Reference](#), and they correspond only to regular season games. The twenty-two statistical categories used in this project are: Wins Per Season (W), Losses Per Season (L), Average Player Age (Age), Average Player Height (Ht.), Average Player Weight (Wt.), Field Goal Attempts Per Game (FGA), Field Goal Percentage (FG%), 3-Point Field Goal Attempts Per Game (3PA), 3 Point Field Goal Percent-age (3P%), 2-Point Field Goal Attempts Per Game (2PA), 2-Point Field Goal Percentage (2P%), Free Throw Attempts Per Game (FTA), Free Throw Percentage (FT%), Offensive Rebounds Per Game (ORB), Defensive Rebounds Per Game (DRB), Total Rebounds Per Game (TRB), Assists Per Game (AST), Steals Per Game (STL), Blocks Per Game (BLK), Turnovers Per Game (TOV), Personal Fouls Per Game (PF) and Points Per Game (PTS). All statistical categories were tested for correlation with season and nine of them exhibited a correlation coefficient $|r| > 0.6$. Time series were created for 14 statistical categories which either exhibited a strong ($|r| > 0.6$) correlation with season or had a clear pattern over time. Furthermore, their decadal means and standard deviations were examined. Finally, Principal Component Analysis (PCA) was performed on seventeen statistical categories, which were selected because they reflect the game style of each team, in contrast to the categories that were left out of the PCA: W, L, Age, Ht. and Wt. which are not related to the style of play. In all calculations, the season 1979-80 was included in the 80s decade since it was the only year from the 70s decade with a 3-point line.

References

1. <https://www.basketball-reference.com/>