

# Mining NBA

David Gavrilović

October 27, 2019

## Contents

<b>1</b>	<b>Stats 101</b>	<b>3</b>
1.1	Traditional stats . . . . .	3
1.2	Advanced stats . . . . .	4
<b>2</b>	<b>Calculating prime of an average player</b>	<b>5</b>
	<b>References</b>	<b>11</b>

# 1 Stats 101

## 1.1 Traditional stats

- **Pos** - Position. Traditionally, position can be one of the following: *PG* - point guard, *SG* - shooting guard, *SF* - small forward, *PF* - power forward and *C* - center. Nowadays, one player usually plays multiple positions and usually is one of the: *Point* - primarily PG, *Combo guard* - plays PG and SG, *Wing* - SF and SG, *Forward* - PF and SF, *Big* - usually C but can also be PF.
- **G** - Games. Number of games player played in during a season.
- **GS** - Games started. Number of games player started. Cannot be greater than G.
- **MP** - Minutes played (Per game or total in a season).
- **FG** - Field goals.
- **FGA** - Field goals attempts.
- **FG%** - Field goal percentage. Calculated as  $FG / FGA$ .
- **3P** - 3-point field goals.
- **3PA** - 3-point field goal attempts.
- **3P%** - 3-point percentage. Calculated as  $3P / 3PA$ .
- **2P** - 2-point field goals.
- **2PA** - 2-point field goal attempts.
- **2P%** - 2-point percentage. Calculated as  $2P / 2PA$ .
- **FT** - Free throws.
- **FTA** - Free throw attempts.
- **FT%** - Free throws percentage. Calculated as  $FT / FTA$ .
- **eFG%** - Field goal percentage that takes into account that a 3-point field goal is, by one point, worth more than a 2-point field goal. Calculated as  $(FG + 0.5 * 3P) / FGA$ .
- **ORB** - Offensive rebounds.
- **TRB** - Defensive rebounds.
- **AST** - Assists.
- **STL** - Steals.
- **BLK** - Blocks.
- **TOV** - Turnovers.
- **PF** - Personal fouls.
- **PTS or PPG** - Points or Points per game.

## 1.2 Advanced stats

- **ORtg** - Offensive rating. An estimate of points produced/scored by a player/team per 100 possessions [1].
- **DRtg** - Defensive rating. An estimate of point allowed per 100 possessions [1].
- **PER** - Player efficiency rating. A measure of a per minute production standardized such that the league average is 15 [2].
- **TS%** - True shooting percentage. Points per scoring attempt converted to the 2-point field goal percentage needed to score that many points per attempt. Calculated as  $PTS / (2 * FGA + 0.44 * FTA)$ .
- **3PAr** - 3-Point attempt rate. Percentage of FGA from 3-point range.
- **FTr** - Free throw attempt rate. Number of FTA per FGA.
- **ORD%** - Offensive rebound percentage. An estimate of the percentage of available offensive rebounds a player grabbed while he was on the floor.
- **DRB%** - Defensive rebound percentage. An estimate of the percentage of available defensive rebounds a player grabbed while he was on the floor.
- **TRB%** - Total rebound percentage. An estimate of the percentage of available rebounds a player grabbed while he was on the floor.
- **AST%** - Assist percentage. An estimate of the percentage of teammate field goals a player assisted while he was on the floor.
- **STL%** - Steal Percentage. An estimate of the percentage of opponent possessions that end with a steal by the player while he was on the floor.
- **BLK%** - Block percentage. An estimate of the percentage of opponent two-point field goal attempts blocked by the player while he was on the floor.
- **TOV%** - Turnover percentage. An estimate of turnovers per 100 plays.
- **USG%** - Usage percentage. An estimate of the percentage of team plays used by a player while he was on the floor.
- **OWS** - Offensive win shares. An estimate of the number of wins contributed by a player due to his offense [3].
- **DWS** - Defensive win shares. An estimate of the number of wins contributed by a player due to his defense [3].
- **WS** - Win shares. An estimate number of wins contibuted by a player [3].
- **WS/48** - Win shares per 48 minutes. League average is around 0.100.
- **OBPM** - Offensive Box plus/minus. A box score estimate of the offensive points per 100 possessions that a player contributed above a league-average player, translated to an average team [4].

- **DBPM** - Defensive Box plus/minus. A box score estimate of the defensive points per 100 possessions that a player contributed above a league-average player, translated to an average team [4].
- **BPM** - Box plus/minus. A box score estimate of the points per 100 possessions that a player contributed above a league-average player, translated to an average team [4].
- **VORP** - Value over replacement player. An estimate of the points per 100 team possessions that a player contributed above a replacement-level (-2.0) player, translated to an average team and prorated to an 82-game season [4].

## 2 Calculating prime of an average player

In this section we will try to determine when does an average player hit their *prime*. But first, what is prime? In order to answer that question we must first define something called *peak*. Peak is the best season player had in his career. Knowing that, prime can be defined as player's seasons in which he is close, or somewhat close, to his peak. Knowing length of a player's prime is important, because prime is probably the most important thing when discussing player's career.

In this paper only individual statistics will be used in process of estimating prime years of a player. Because the ultimate stat that can correctly measure when do players hit their prime does not exist (yet), multiple other stats will be used, both traditional and advanced. Those stats are, mostly, PTS, FG and FGA from traditional and PER, WS, BPM and VORP from advanced stats. Reasoning behind first three is, well, player in his prime should be best version of himself, so he will probably score and shoot more than in his non-prime seasons. Stats such as PER, BPM and VORP exist so we could determine impact on a game by player, or how good a player is. Prime season should have higher values for this stat than for other seasons. WS is used to determine players contribution to winning. It's also expected for this stat to be higher for seasons in player's prime than for the other seasons.

Of course, different players can hit their prime at different age. Let's check prime years of some of the players inducted, or soon to be inducted, into the Naismith memorial basketball hall of fame. Selected players are chosen without any particular reason. Players used are Bird 1, Shaq 2 and Duncan 3. As we can see from the shown box plots, mentioned players have the best stats when they were around 26 to 30 years old. From that we can conclude that that was the length of their prime.

In the next step of determining prime of players, I checked at what age do players win awards, such as *regular season MVP*, *FMVP*, *DPOY* and *Sixth man of the year*. Plots are shown in figure 4. Every plot but plot for Sixth man of the year is somewhat similar. The number of awards players won while in their late twenties and early thirties is greater than number of awards in other years. Defensive player of the year award plot is different, where players aged from 23

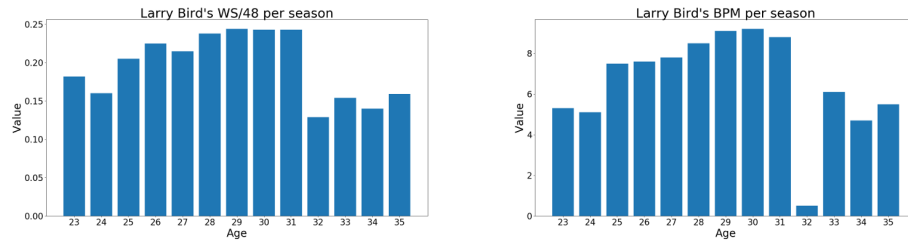


Figure 1: Bird's WS/48 and BPM per season

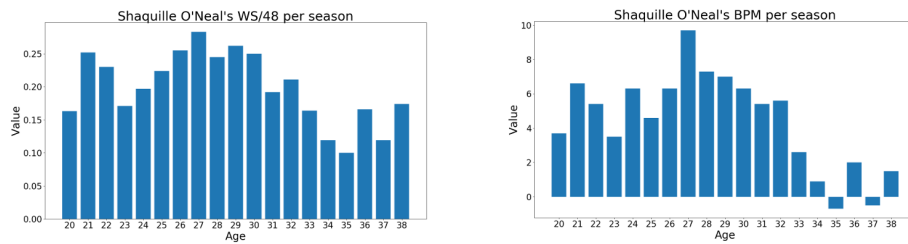


Figure 2: Shaq's WS/48 and BPM per season

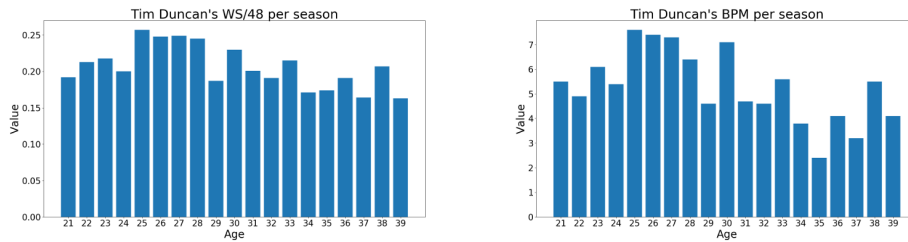


Figure 3: Duncan's WS/48 and BPM per season

to 31 won with roughly the same frequency, with an exception of players who were 28 years old when they won, and the ones who were 27. After that there is a significant drop. These plots are showing us that the best players in the season are usually ones between 26 and 31 years old. But those players are stars and superstars. What about an average player?

First, I checked age of players who played in the NBA. Results can be seen in figure 5. Note that there are two histograms. One represents every player who played in the NBA (in the three point era, from 1979/80 season). The other is filtering out every player (from the same era) that has played in less than 25 games and less then 15 minutes per game. Those players are filtered out because I don't consider them regular contributors for the team they are playing for. The reason behind that is that they are probably not skilled enough, and

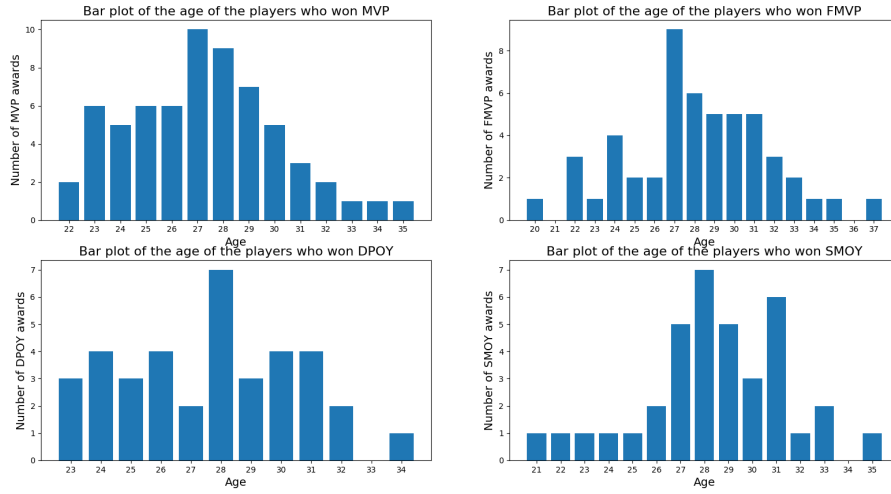


Figure 4: Age of award winners

possibly not in their prime, so they might not be important. Also, I just wanted to check on their histogram as well.

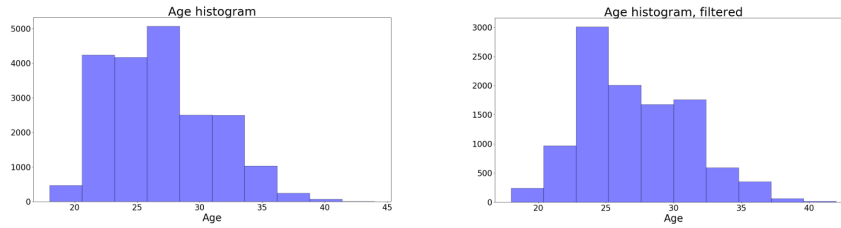


Figure 5: Age histograms

In the left histogram, unfiltered one, most of the players are in between 21 and 28 years old. That makes sense because players are usually drafted when they are younger than 25, and after rookie contracts expire and they are not good enough, they are out of the league. Histogram on the right has way lower number of players younger than around 23 years old. The reason behind that is that rookies aren't usually contributors right away because they need some time to develop. The difference in number of players that are above 28 years old in unfiltered and filtered data is not that large. That is the case because players above that age are usually good players, and they might be good because they are in their prime. That trend continues on both histograms until the significant drop somewhere after players reach 32 or 33. That might happen because they are simply not serviceable anymore, and because of that cannot find teams to sign with. To conclude this paragraph, number of players is rising at first, until the age of 28. After that, there are two significant drops.

Now, I can try to estimate prime of an average player with statistics. First, I checked traditional stats. Box plots of four stats are shown, PTS, FG, FGA and FT. Simply, it is expected from a player in his prime to shoot and score more than in his non-prime seasons. Note that values represented on a y-axis are averages per season, not per game! In figure 6 every player from the three-point era is taken into account while players represented in the figure 7 are the ones I consider contributors (minutes per game > 15 and games played > 35).

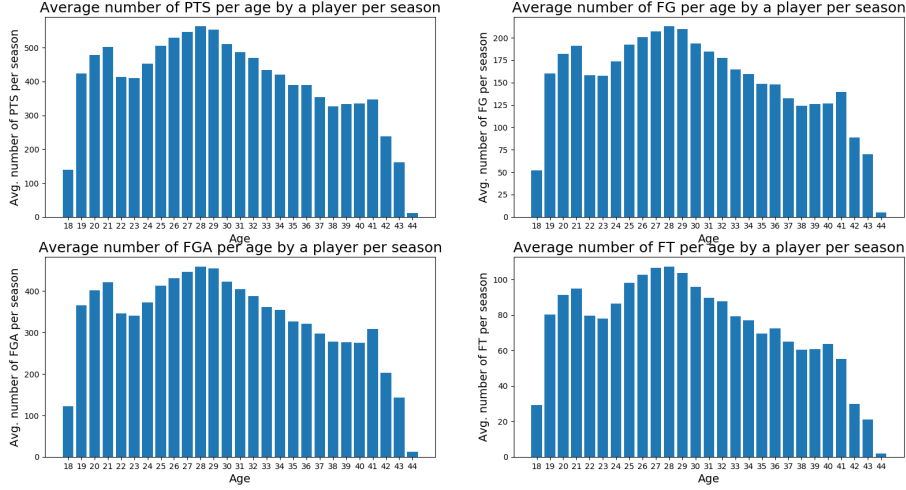


Figure 6: Season totals averages per age unfiltered

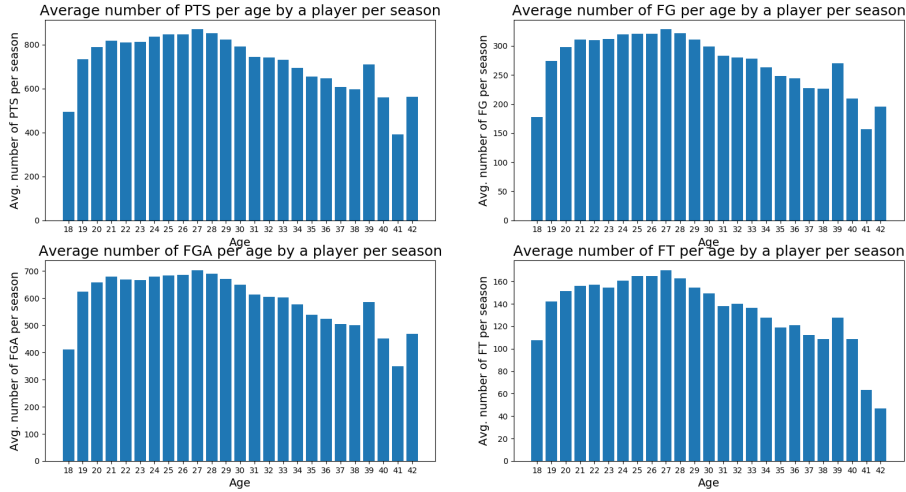


Figure 7: Season totals averages per age filtered

Both plots are somewhat similar. Values for players age from 26 to 30 or maybe 31 are higher than for the any other age, with a significant drop after that.



Both plots show peaks at a similar age, 28 for unfiltered data and 27 for filtered. So it's good thing to consider that prime of an average player is actually in late twenties/early thirties. The main difference can be seen in values for players that are younger than 25. Unfiltered data has a easily noticeable difference between values for players 21 and players 22 years old. Reason behind that is probably draft. Promising players are drafted earlier, often before they are 20, while a lot of players with lower upside decide to develop in college and try to enter the league later. Those players are the reason why the values are lower for certain age. From plots with filtered data we can observe that those players are not contributing right away, so lower values are, while still there, not so emphasized.

Advanced stats are better than traditional and hold more information about how good a certain player is. I did the same thing as in the paragraph before, but now with some of the advanced stats. Stats chosen are PER, WS, BPM and VORP. Box plots for unfiltered players are shown in figure 8, while box plots after filtering are in figure 9.

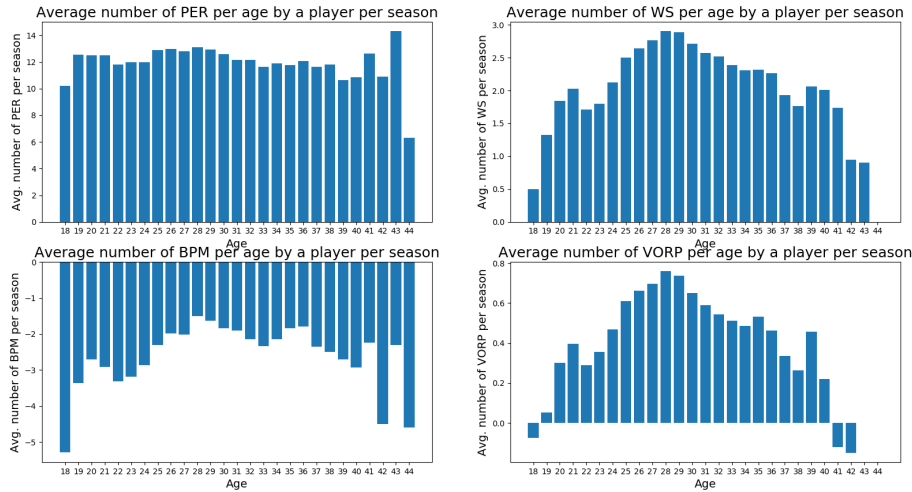


Figure 8: Advanced stats averages per age unfiltered

These plots are way more interesting. First, PER box plot, for both filtered and unfiltered data, is somewhat similar to the ones with traditional stats, with the same results. Plots for WS and VORP are normal distribution-like, with spike from 26 to 31 in y-axis values and lower values from the sides. That radius with higher values contains prime years of an average player. Plot for BPM is, interestingly, negative for unfiltered data, and with some positive values for filtered data. Highest values for BPM are from players from 25 to 36 years old. Some of the plots that are showing advanced data has high value for players 39 years old, which is interesting. Those players were usually very good in their prime, and because of that they were usually good when they were older, in limited minutes.

Taking all mentioned things into a consideration, age when players are usu-

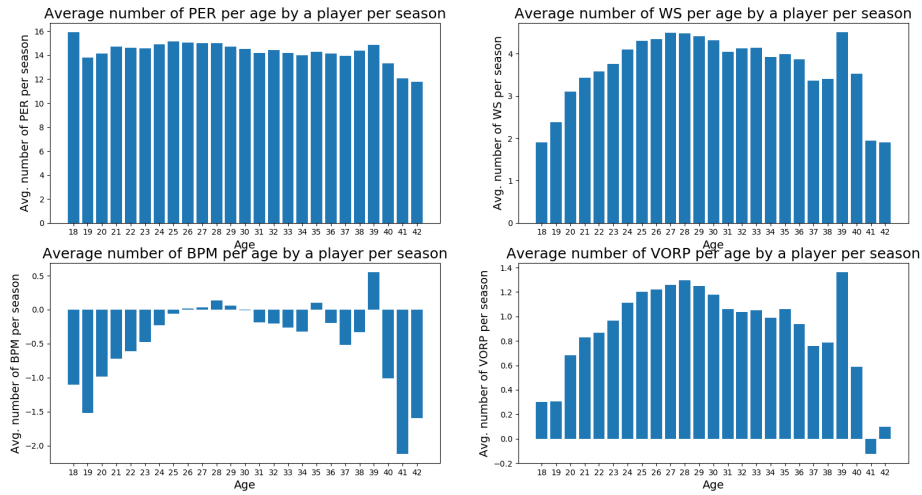


Figure 9: Advanced stats averages per age filtered

ally winning awards and age when the players are statistically better than in any other age, conclusion is that an average player enters into prime at around 25 year old, and exits when he is around 31, with kinda significant drop after that. Peak season happens when player is from 27 to 29 years of age, somewhere in the middle of their prime.

## References

- [1] Basketball-Reference. Calculating individual offensive and defensive ratings. on-line at: <https://www.basketball-reference.com/about/ratings.html>.
- [2] Basketball-Reference. Calculating per. on-line at: <https://www.basketball-reference.com/about/per.html>.
- [3] Basketball-Reference. Nba win shares. on-line at: <https://www.basketball-reference.com/about/ws.html>.
- [4] Basketball-Reference. About box plus/minus (bpm). on-line at: <https://www.basketball-reference.com/about/bpm.html>.