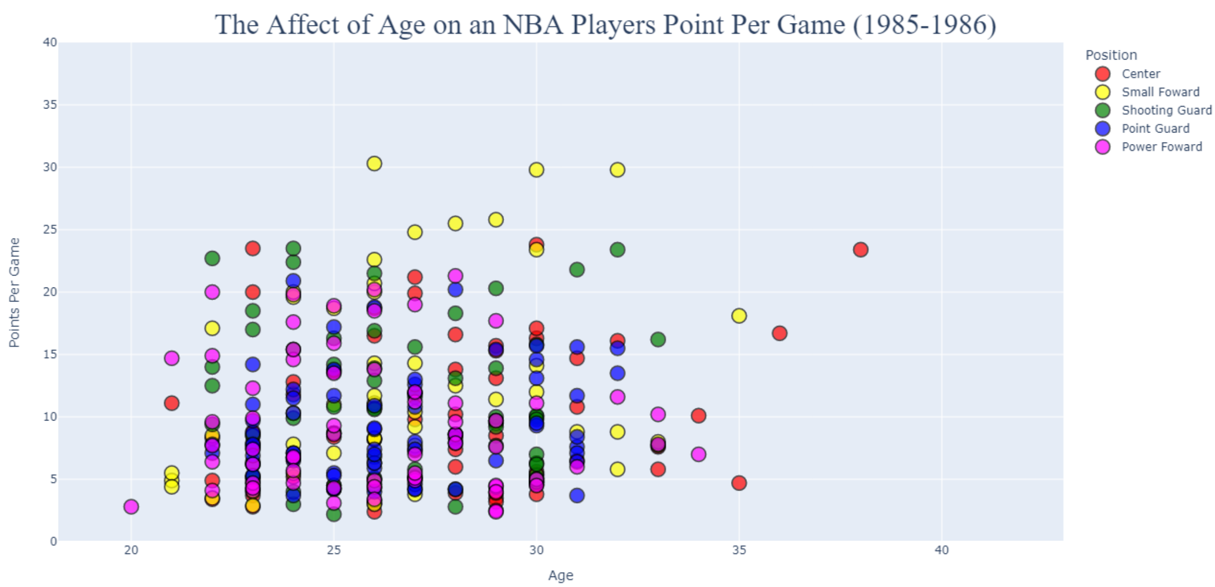
Does an NBA player's age affects their production on the court?

Hope Zobinou

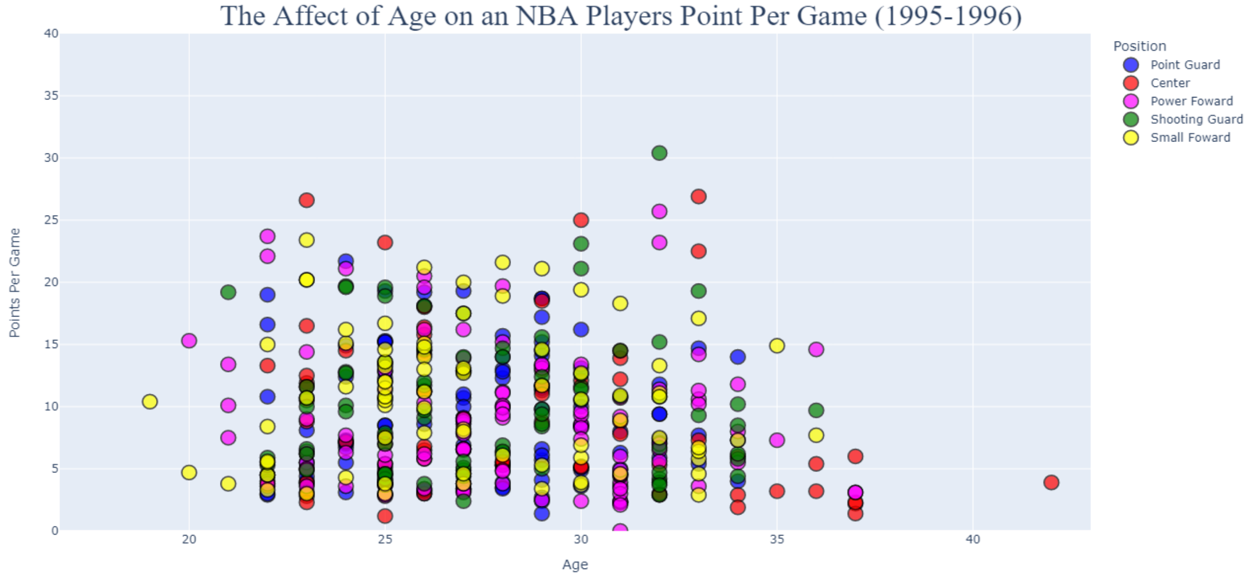
Does an NBA player's age affects their production on the court and how does it differ from era to era? What drew me into analyzing the age and points per game of NBA players is my interest in my favorite athletes playing when they are old and playing when they are young. I want to see how strong Father Time is. Even though a player's productivity can be measured with many variables, I'm interested in the player's average points per game just to keep it simple. I collected the data from a website called https://www.basketball-reference.com The site has all the stats from every player from any season in the NBA. After converting the CSV into a data frame, I cleaned up the data by removing rows that contain the column's name, removing players that average less than 10 minutes per game, adding a year column, and changing column names to make it easy for the reader to read. The data frame now contains only the things that will impact my visualization like the player's name, Position (The position they play in a game), current age, average minutes per game, and average points per game. I also changed the names of some values in the data frame to help the reader understand what a specific stat represents.

For my visualizations, I opted to use a scatter plot since the x and y are discreet and continuous respectively. The y represents the player's average points per game and the x represents the player's age from their respective seasons. I used another scatter plot to show if there is some kind of trend between the player's age and minutes. I then used a few bar graphs to better show the distribution of age as a whole for specific points per game range. A heatmap is used to see the correlation between age, minutes, and points per game numerically.

From all the seasons I got data on (1985-86, 1995-96, 2005-06, 2015-16), it looks like the player's points per game are normally distributed. At the young and old ages, the points are not as dense and not as high as the points in the Middle Ages. The young ages are 18-24, the Middle Ages (Player's prime) are 25-34, and the old ages are 35-42. This makes sense because when a player is in their prime, they are playing the best basketball in their career.

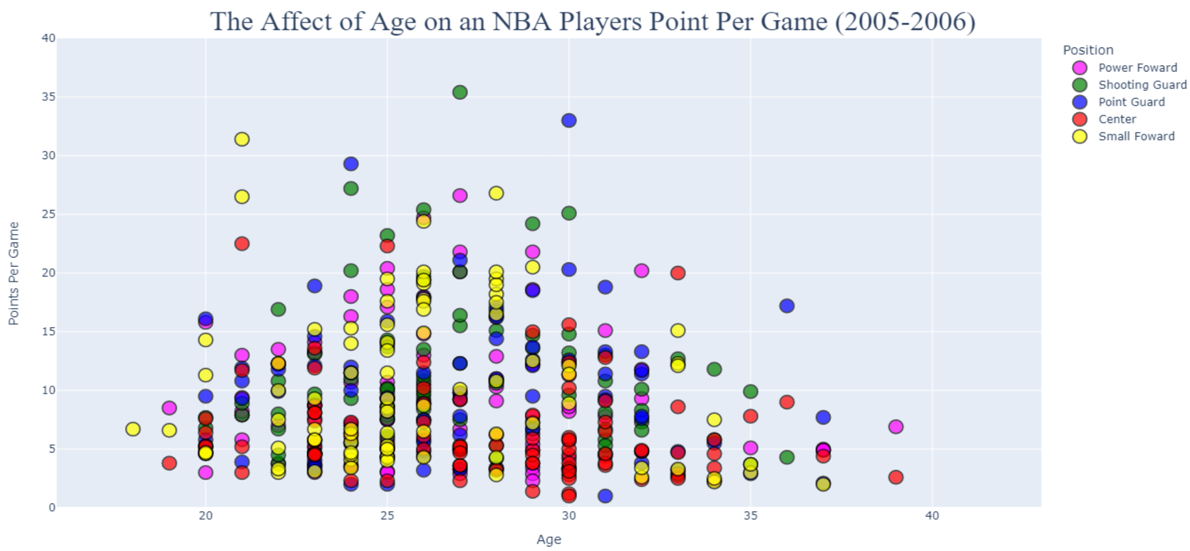


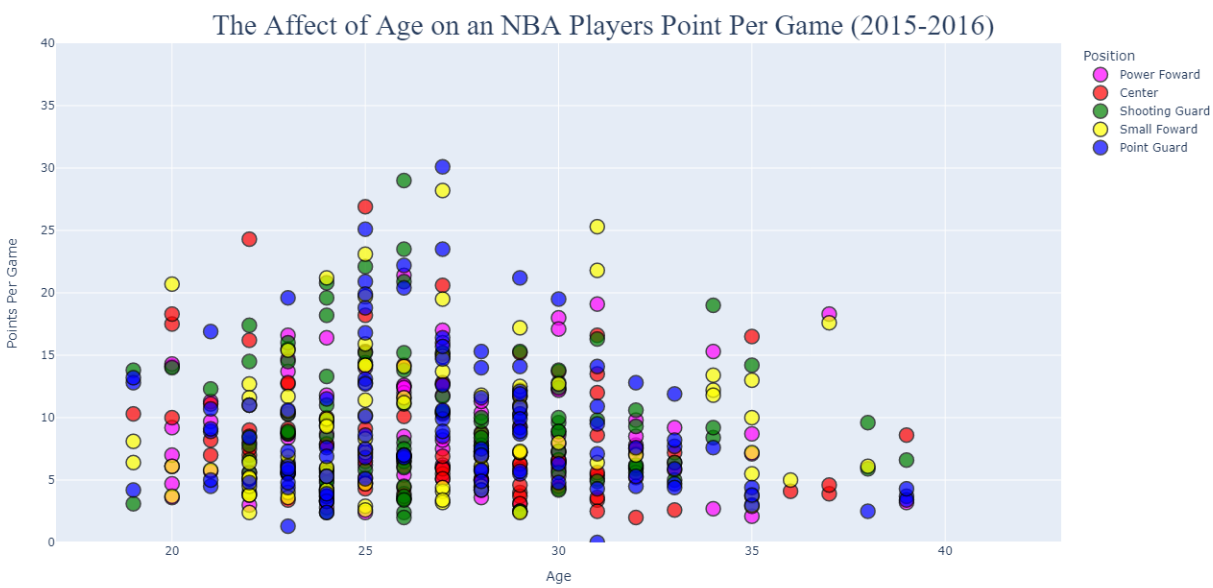
In the 1985-1986 season, we can see the normal distribution. The players in the prime age range are reaching higher point per game averages than the older and younger ages. We can also see that Small Forward and centers are the better scorers. In the 1995-1996 season we can see that the normal distribution is more solidified from the previous season. We are now also seeing that players are playing younger than 20 years old and the oldest player is 42. The better scorers are the forwards and the centers. In the 80s and 90s, forwards and centers where more dominant than guards. The game centered around close-range shots and forwards and centers are more equipped for that since most are at least 6'8".



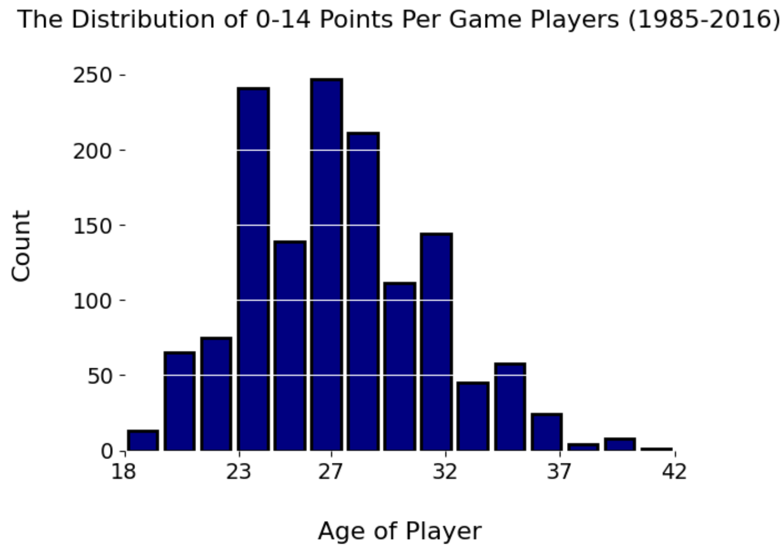
In the 2005-2006 season, the distribution looks the most normal. If you try to draw and trend line, it will look pretty similar to the one in a statistics textbook. After the 90s era, the game shifted from close-range shots to far-range shots. Guards were the best at shooting 3 points and more plays were created to allow guards to shoot more 3-pointers. You can see this effect on the 2005-2006 graph. Guards are the ones that look to be the better scorers. We can also see that the youngest player is 18 and that the majority of high points per game averages come from players in their prime.

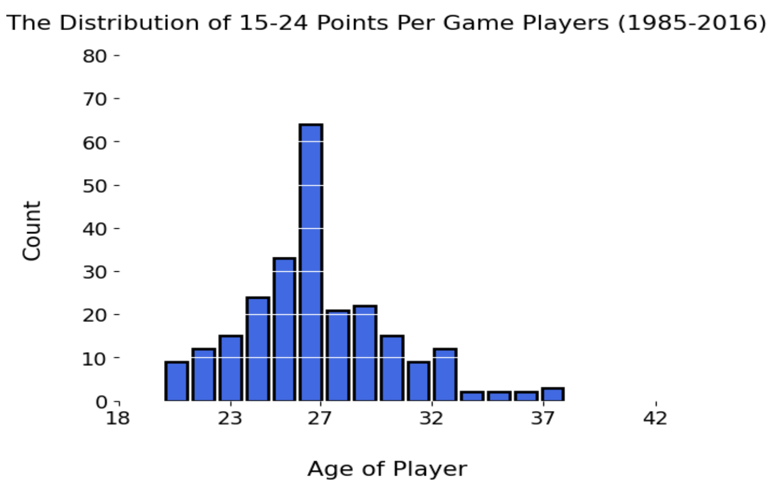
In the 2015-2016 season, again we see the trend of normal distribution between the player's age and points per game. I big change I did notice is that high points per game averages are occupied by all positions. This shows that the game is starting to become positionless since all positions are able to dominate on the court.

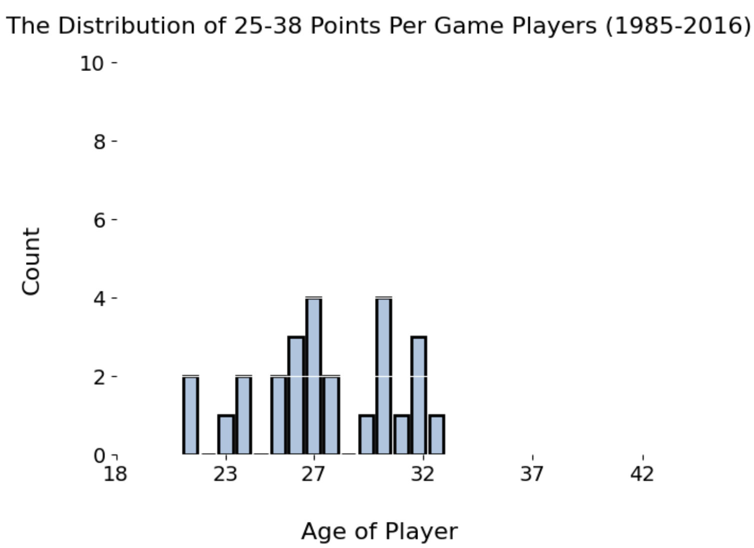




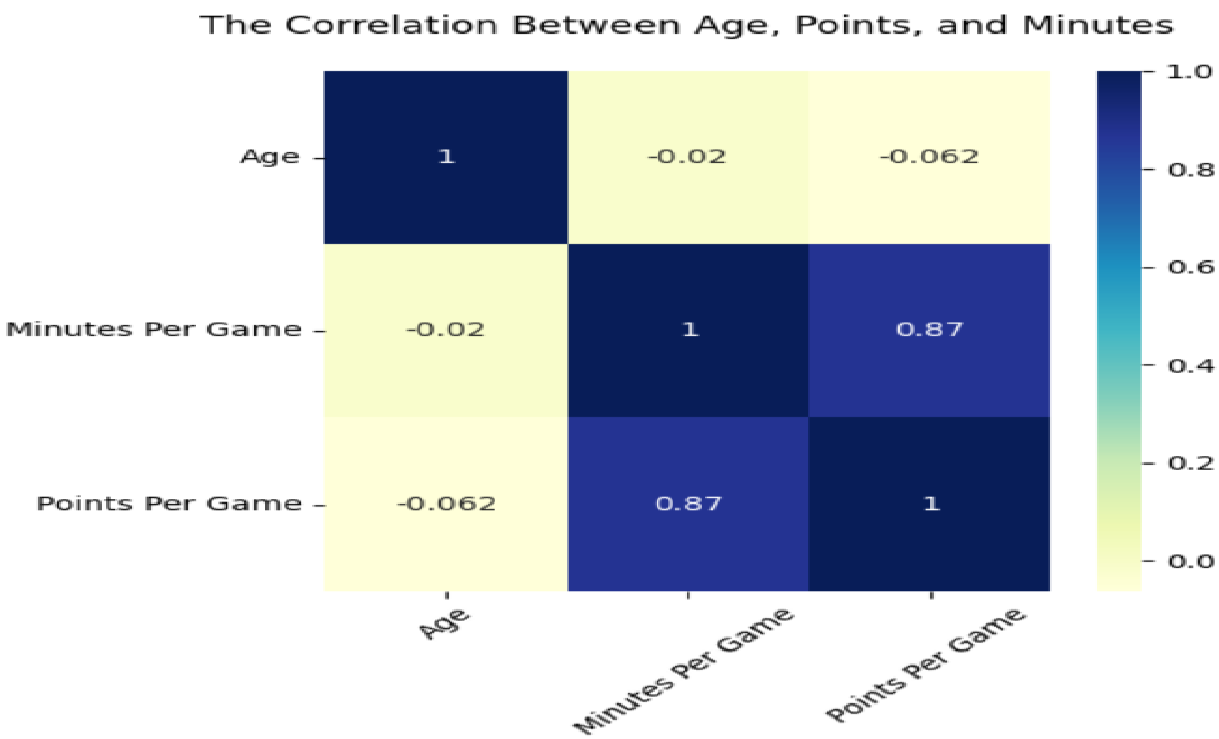
So, it looks like age does affect a player's points per game averages. But there is something else that can easily affect a player's production on the court. It's the minutes spent on the court. It's easy to score a high number of points when you spend the most time on the court. And players in their prime get the most playing time because they are playing the best and their bodies are built for it. An older player's body (ex: 40 y/o) can't play 40 minutes a night like a 26-year-old player can.



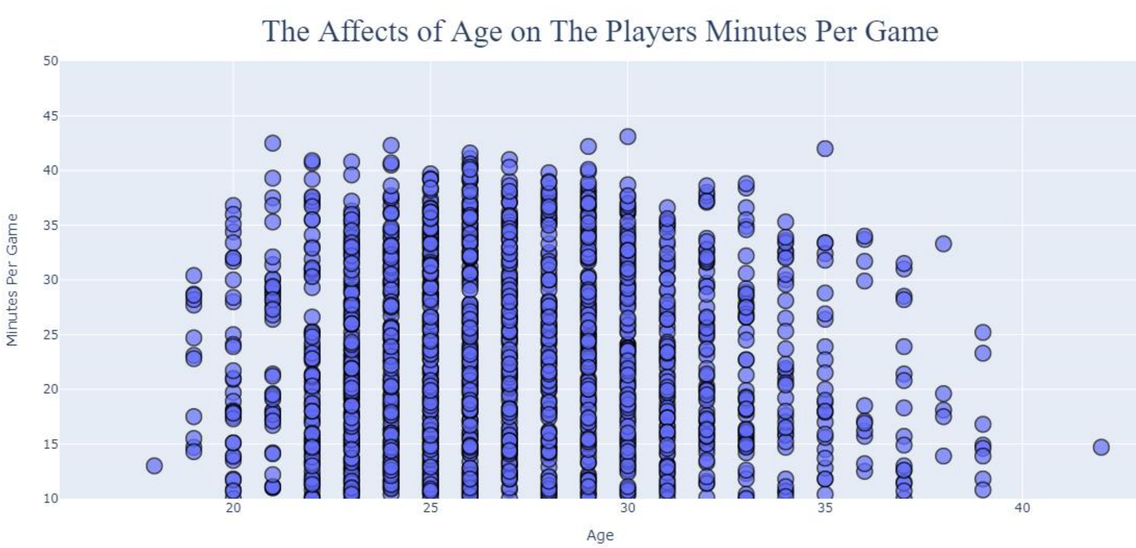




From the 3 graphs you can see that high point scorers are not really old or really young. The high point scorers are players that are in their prime. They are given more minutes to play, due to their bodies, which equates to more chances to produce on the court. We can also see that high point averages are hard to achieve. The 25-28 points per game range has the least amount of players.



From the correlation map we can see that minutes play a major role in an NBA player's production. In conclusion minutes spent on the court affect an NBA player's points per game the most but there is a correlation between minutes and age. Prime players get more playing time.



The process to complete the final project went smoothly. I didn’t change my topics; I just stuck with the first one that popped into my head. I did put in a good amount of work in advance. That work was mainly data collecting and cleaning. I didn’t learn anything new in the process. My project meets expectations because my visualizations show trends, they are readable and easy to understand. My written portion also explains what’s happening in the visualizations well and explains why that’s the case. The visualizations also helped me come to a valid conclusion.