Visualisation Project

2024-08-27

Storytelling with Open Data

Van Thai Phan - s3818387

Introduction

The NBA nowadays is full of exciting talents who are able to score the basketball using a variety of different methods. This presentation is an analysis of the different methods of scoring employed by the top 20 highest scorers of the NBA's 2022/2023 season to determine what these players have in common and what makes them such effective scorers. Furthermore, the presentation aims to discover whether the volume and/or efficiency of a type of shot can set the most elite scorers in the NBA apart from the rest.

The dataset used for the presentation was published by Basketball Reference. This dataset includes the per game data of the leading scorers in the NBA who played more than 57 games, which is 70% of available games in a season. This ensures that the data accurately reflects the scoring ability of players throughout a relatively equal sample size of games.

The statistics that will be looked at are the statistics of all the ways that the players can score the ball which includes:

Field goal attempts and field goal percentage (The amount of shots taken per match excluding free throws and the efficiency of these shots)

2-point attempts and 2-point percentage (The amount of shots taken within the 3-point line per game excluding free throws and the efficiency of these shots)

3-point attempts and 3-point percentage (The amount of shots taken outside the 3-point line per game excluding free throws and the efficiency of these shots)

Free throw attempts and free throw percentage (The amount of free throw shots per game and the efficiency of these shots)

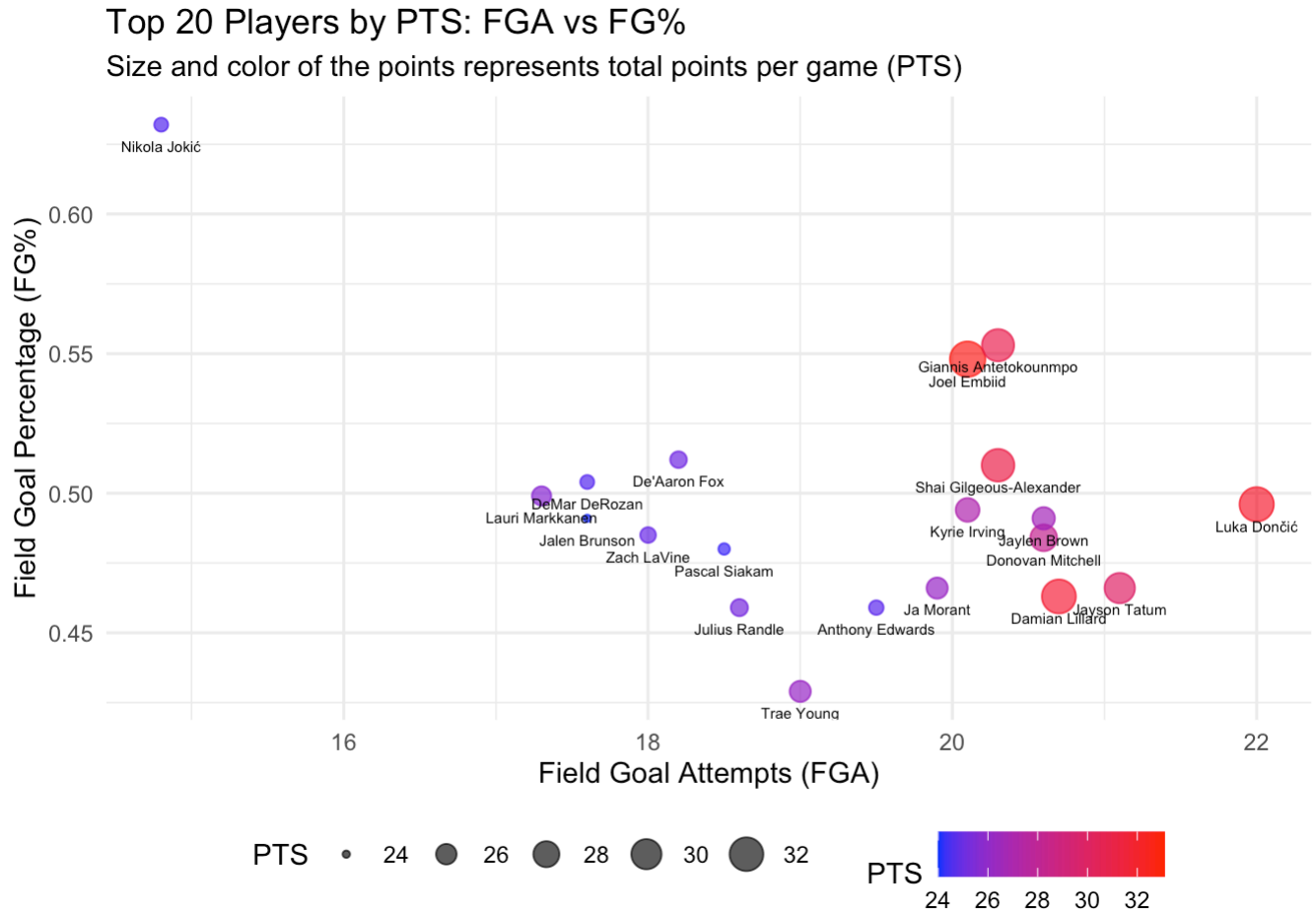
```
data <- read.csv("/Users/macbookair/Desktop/Visualisation project/sportsref_download.
csv")
```

Sort by PTS and select the top 20 players

```
top_players <- data %>%
  arrange(desc(PTS)) %>%
  head(20)
```

Scatter plot comparing FGA and FG%

```
ggplot(top_players, aes(x=FGA, y=FG., label=Player)) +
  geom_point(aes(size=PTS, color=PTS), alpha=0.7) +
  scale_color_gradient(low="blue", high="red") +
  geom_text(vjust=2.5, hjust=0.5, size=2) +
  labs(title="Top 20 Players by PTS: FGA vs FG%",
       x="Field Goal Attempts (FGA)",
       y="Field Goal Percentage (FG%)",
       subtitle="Size and color of the points represents total points per game (PT
S)") +
  theme_minimal() +
  theme(legend.position="bottom")
```



From this scatterplot, we can see that the top 20 scorers in the NBA last season all took over 17 shots per game with over 45% efficiency, except for two outliers: Nikola Jokic and Trae Young.

Nikola Jokic is the only player who took less than 17 shots, he in fact took less than 15 shots per game, more than 2 attempts less than the player with the second least field goal attempts. He made up for the lack of volume with an exceptionally high efficiency of over 62.5%. However, his low field goal attempts per game does make it difficult for him to outscore the other players in the top 20, with his scoring average only around 24 points per game.

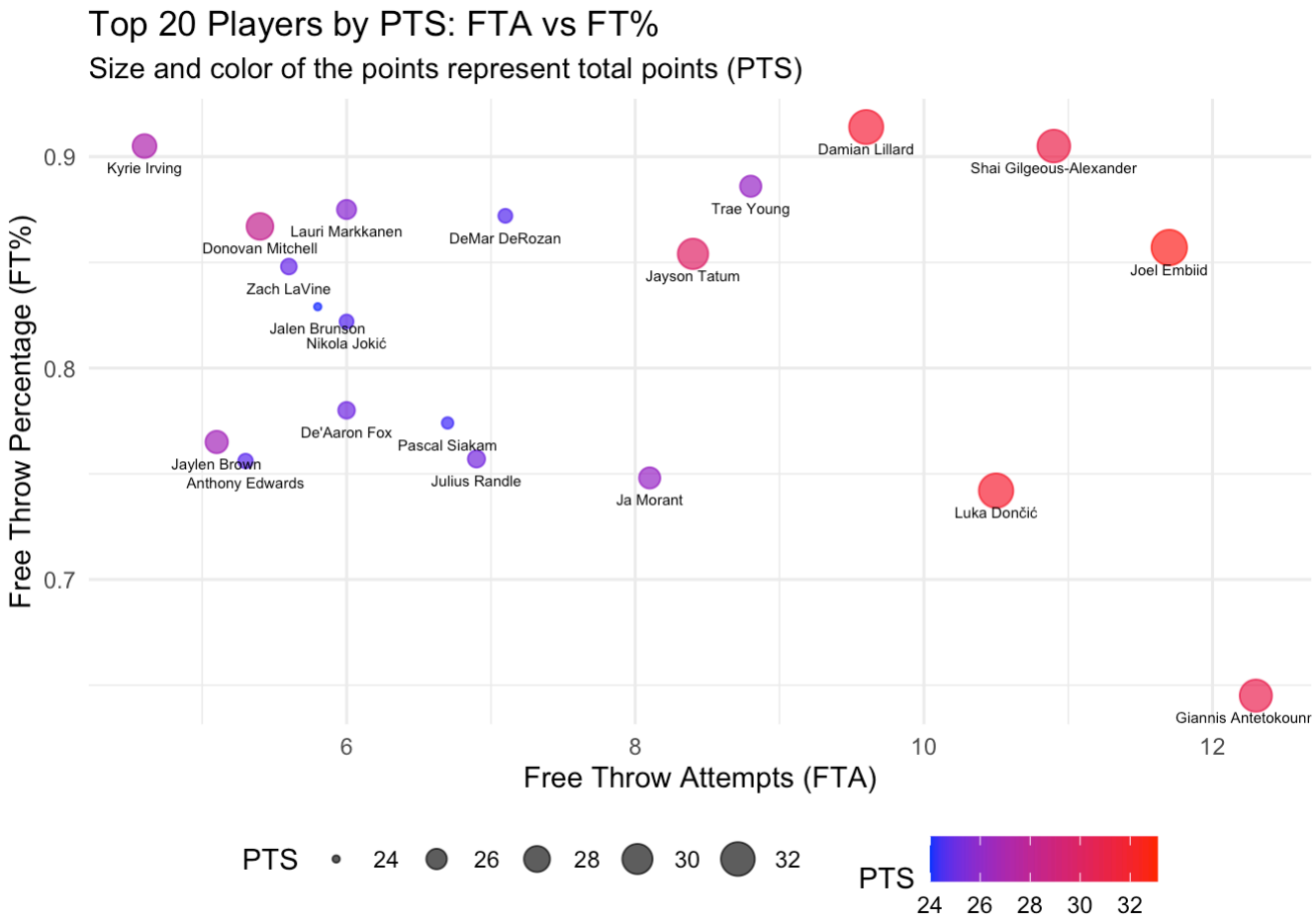
On the other hand, Trae Young had around 19 field goal attempts per game but at an efficiency lower than 45%. He made up for his below average efficiency with a relatively high amount of attempts.

The top 5 highest scoring players include: Joel Embiid, Luka Doncic, Shai Gilgeous-Alexander, Damian Lillard, and Giannis Antetokounmpo. These players all took over 20 attempts per game and maintained an efficiency of over 45% but they were not the only ones to do so. Thus, we need to look deeper to see what made them stand out from the rest.

Scatter plot comparing FTA and FT%

```
plot <- ggplot(top_players, aes(x=FTA, y=FT., label=Player)) +
  geom_point(aes(size=PTS, color=PTS), alpha=0.7) +
  scale_color_gradient(name="PTS", low="blue", high="red") +
  scale_size_continuous(name="PTS") +
  geom_text(vjust=2.5, hjust=0.5, size=2) +
  labs(title="Top 20 Players by PTS: FTA vs FT%",
       x="Free Throw Attempts (FTA)",
       y="Free Throw Percentage (FT%)",
       subtitle="Size and color of the points represent total points (PTS)") +
  theme_minimal() +
  theme(legend.position="bottom")

plot
```




```
colnames(data)
```

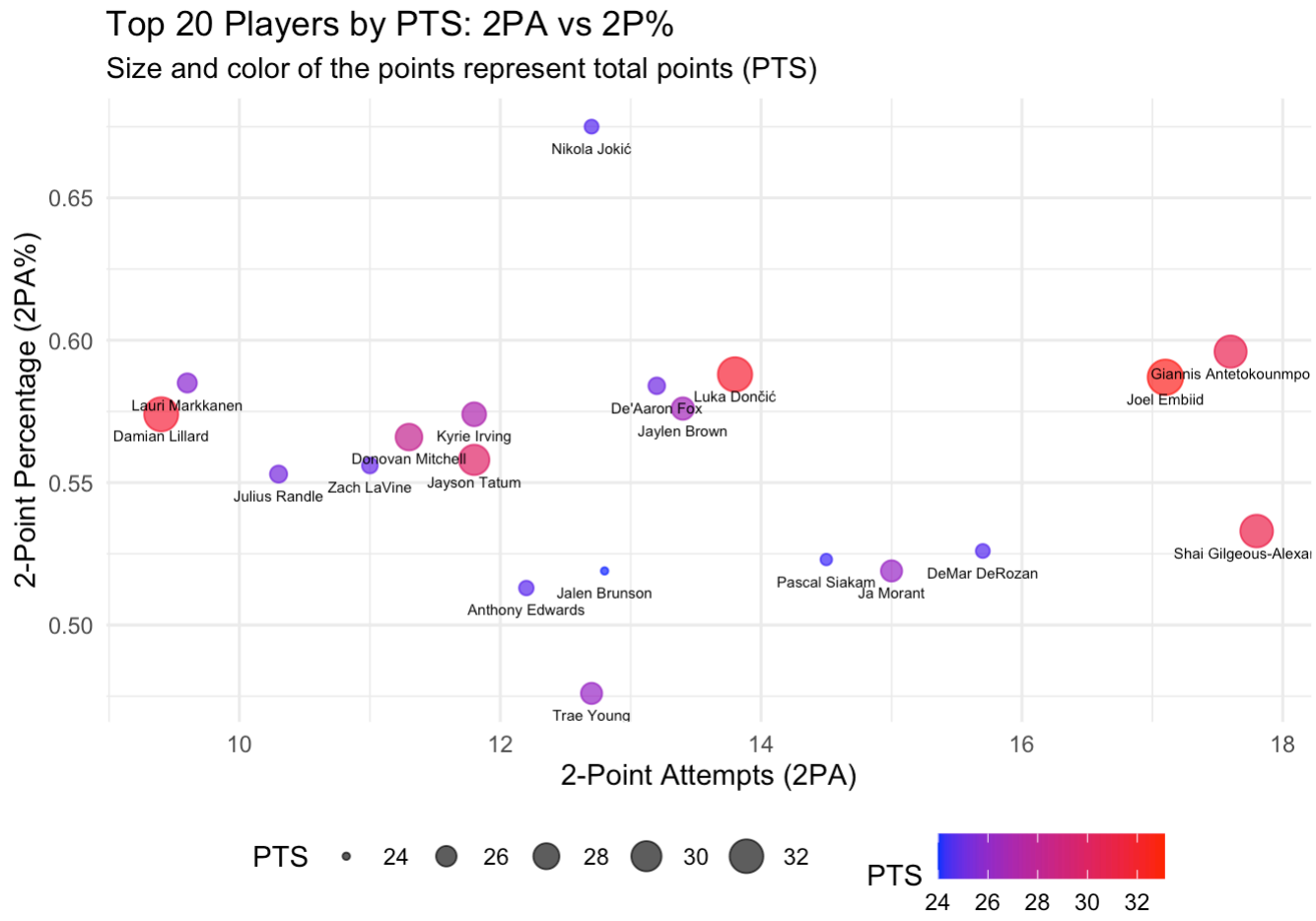
```
##  [1] "Rk"     "Player" "Pos"    "Age"    "Tm"     "G"      "GS"     "MP"
##  [9] "FG"     "FGA"    "FG."    "X3P"    "X3PA"   "X3P."   "X2P"    "X2PA"
## [17] "X2P."   "eFG."   "FT"     "FTA"    "FT."    "ORB"    "DRB"    "TRB"
## [25] "AST"    "STL"    "BLK"    "TOV"    "PF"     "PTS"
```

This scatterplot shows that the top 5 highest scorers in the NBA last season all took more than 9 attempts at the free throw line, with Giannis Antentokounmpo taking the most at over 12 free throws per game at the lowest efficiency at under 70%. Luka Doncic had the second worse free throw efficiency among the top 5 scorers at around 75%. The other 3 players all shot above 80% from the free throw line.

Scatter plot comparing 2PA and 2P%

```
plot <- ggplot(top_players, aes(x=X2PA, y=X2P., label=Player)) +
  geom_point(aes(size=PTS, color=PTS), alpha=0.7) +
  scale_color_gradient(name="PTS", low="blue", high="red") +
  scale_size_continuous(name="PTS") +
  geom_text(vjust=2.5, hjust=0.5, size=2) +
  labs(title="Top 20 Players by PTS: 2PA vs 2P%",
       x="2-Point Attempts (2PA)",
       y="2-Point Percentage (2PA%)",
       subtitle="Size and color of the points represent total points (PTS)") +
  theme_minimal() +
  theme(legend.position="bottom")

plot
```



This scatterplot shows that all of the top 20 highest scorers in the NBA apart from Trae Young shot with an efficiency of over 50% inside the 3-point line, with Nikola Jokic once again being by far the most efficient.

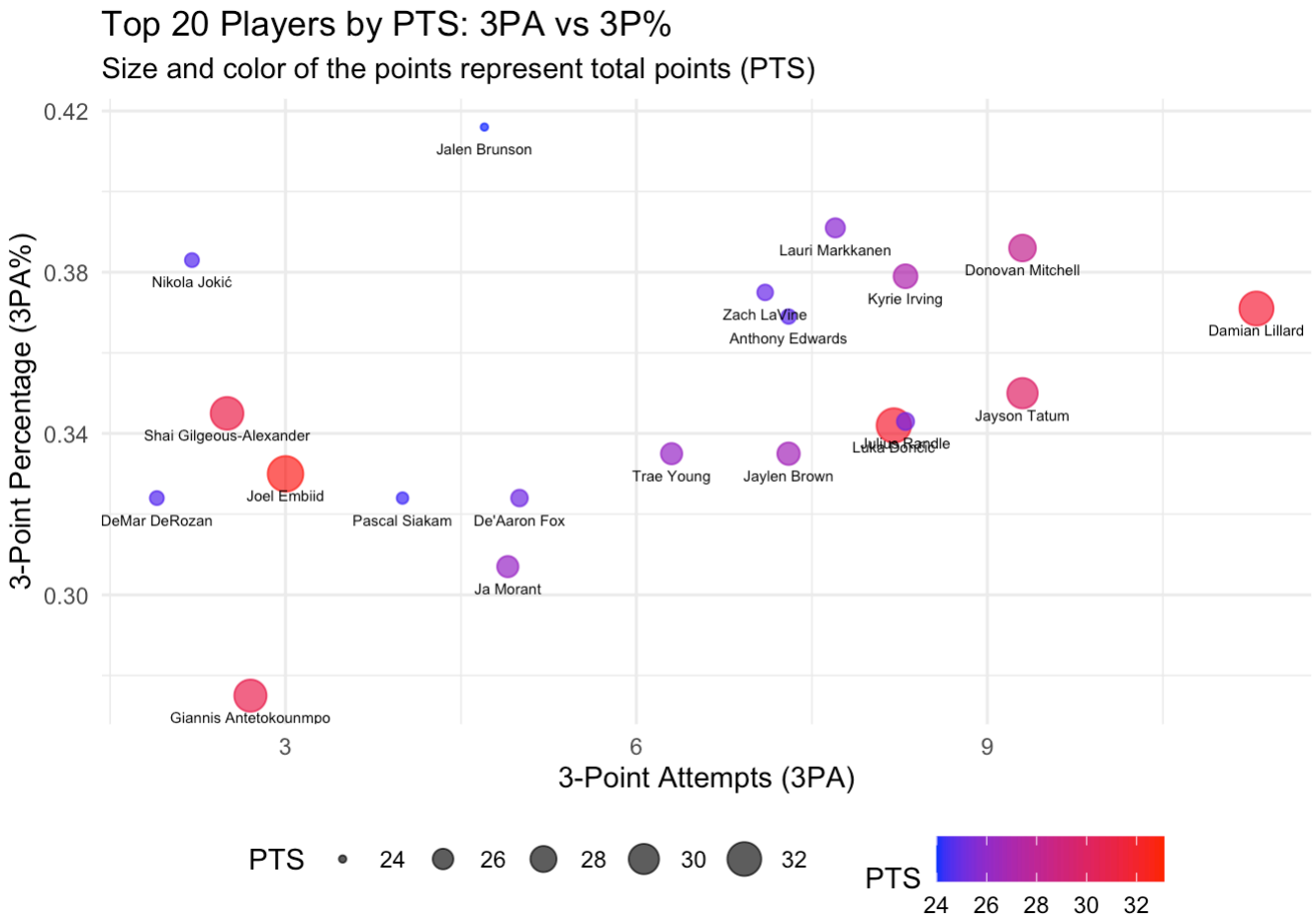
When looking at the top 5 highest scorers as mentioned before, we can see that they are spread out across the scatterplot, indicating that the amount of 2-point attempts that they took vary and not consistent with each other.

Among the top 5 highest scorers, Damian Lillard took the least amount of 2-point attempts per game by far, with less than 10 of these shots per game, much lower than the amount taken by Embiid, Giannis, and Shai who all took more than 17 2-point attempts per game.

Scatter plot comparing 3PA and 3P%

```
plot <- ggplot(top_players, aes(x=X3PA, y=X3P., label=Player)) +
  geom_point(aes(size=PTS, color=PTS), alpha=0.7) +
  scale_color_gradient(name="PTS", low="blue", high="red") +
  scale_size_continuous(name="PTS") +
  geom_text(vjust=2.5, hjust=0.5, size=2) +
  labs(title="Top 20 Players by PTS: 3PA vs 3P%",
       x="3-Point Attempts (3PA)",
       y="3-Point Percentage (3PA%)",
       subtitle="Size and color of the points represent total points (PTS)") +
  theme_minimal() +
  theme(legend.position="bottom")

plot
```



This scatterplot shows that all of the top 20 highest scorers in the NBA apart from Giannis Antetokounmpo shot with an efficiency of over 30% from behind the 3-point line, with Jalen Brunson being the most efficient by far.

Similar to the 2-point attempts scatterplot, the 3-point attempt plot also indicates that the amount of 3-point attempts taken by the top 5 highest scorers also vary significantly, with Damian Lillard taking the most three pointers and also having the highest efficiency.

Apart from the top 5, the other 15 players are also spread out very evenly in this plot, thus, we can see that the volume and efficiency of 3-point attempts are not what set the top 5 scorers apart from the rest.

Findings: All the top 20 scorers in the NBA last season have a lot in common when looking at their statistics. However, they also have vast differences in the way they score. What makes the top 5 scorers really stand out from the rest is their ability to get fouls and make free throws.

Thus, it can be concluded from the dataset, and the plots that free throws are the difference maker between good scorers and elite scorers in the NBA.