Project Part 1

Name - blank for Part 1

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Background information

The chosen data set revolves on the best NBA players this season. I am an avid NBA fan and have constant arguments with my friends on who the best player is. What other way to debate it than using statistics??? The data was made by 538.com is a website used to opinion polls, politics, economics, and sports blogging using statistical analysis. This data is constantly updated and shows very cleanly who the best NBA player is. The beautiful way it chooses who is the best is using their "RAPTOR" rating which shall be explained in the next section.

Explanation of the rows and variables

RAPTOR is a plus-minus statistic that measures the number of points a player contributes to his team's offense and defense per 100 possessions, relative to a league-average player.

Here are the variables in the dataset:

Player: Name of the NBA athlete.

Team: The professional basketball team the player is associated with.

Position(s): There are five positions in basketball, and sometimes a player can play multiple.

Minutes: The amount of minutes they played in the 2019-2020 NBA season.

Box Score Raptor Offense: General offensive metrics composed in these four categories: scoring and usage; passing; rebounding; and space creation all used in 538 regression algorithm.

Box Score Raptor Defense: Defensive metrics that include: steals, offensive fouls drawn, Opponents' field goals made and attempted, Enhanced defensive rebounds,Positional opponents' points scored, Positional opponents' offensive rebounds, Distance traveled, for perimeter defenders only, Opponents' free throws made, Fastbreak turnovers committed, Penalty fouls committed, Opponents' offensive rating. These are then all accounted for in their regression algorithm.

Box Score Raptor Total: The two scores of the BSRO and BSRD added up.

ON/OFF Raptor Offense:The player's offensive rating, The player's courtmates' weighted average offensive ratings, the player's courtmates'(bench players) other courtmates' offensive ratings all using the regression algorithm.

ON/OFF Raptor Defense: The player's defensive rating, The player's courtmates' weighted average defensive ratings, the player's courtmates'(bench players) other courtmates' defensive ratings all using the regression algorithm.

ON/OFF Raptor Total:The two scores of the OFRO and OFRD added up.

Overall Raptor Offense: Overall RAPTOR Offense is equal to roughly 85 percent of "Box Offense" RAPTOR, plus 21 percent of "On-Off Offense" RAPTOR.

Overall Raptor Defense: Overall RAPTOR Defense is equal to roughly 85 percent of "Box Dffense" RAPTOR, plus 21 percent of "On-Off Defense" RAPTOR.

Overall Raptor Total: Overall RAPTOR TOal is equal to roughly 85 percent of "Box Total" RAPTOR, plus 21 percent of "On-Off Total" RAPTOR.

WAR: RAPTOR calculates wins above replacement level using a replacement level of -2.75 points per 100 possessions. This is similar to the method used in the movie Moneyball.Wins Above Replacement Player (WARP) evaluates a player who belongs to a team that is made up of him and 4 average players.

How the data was collected

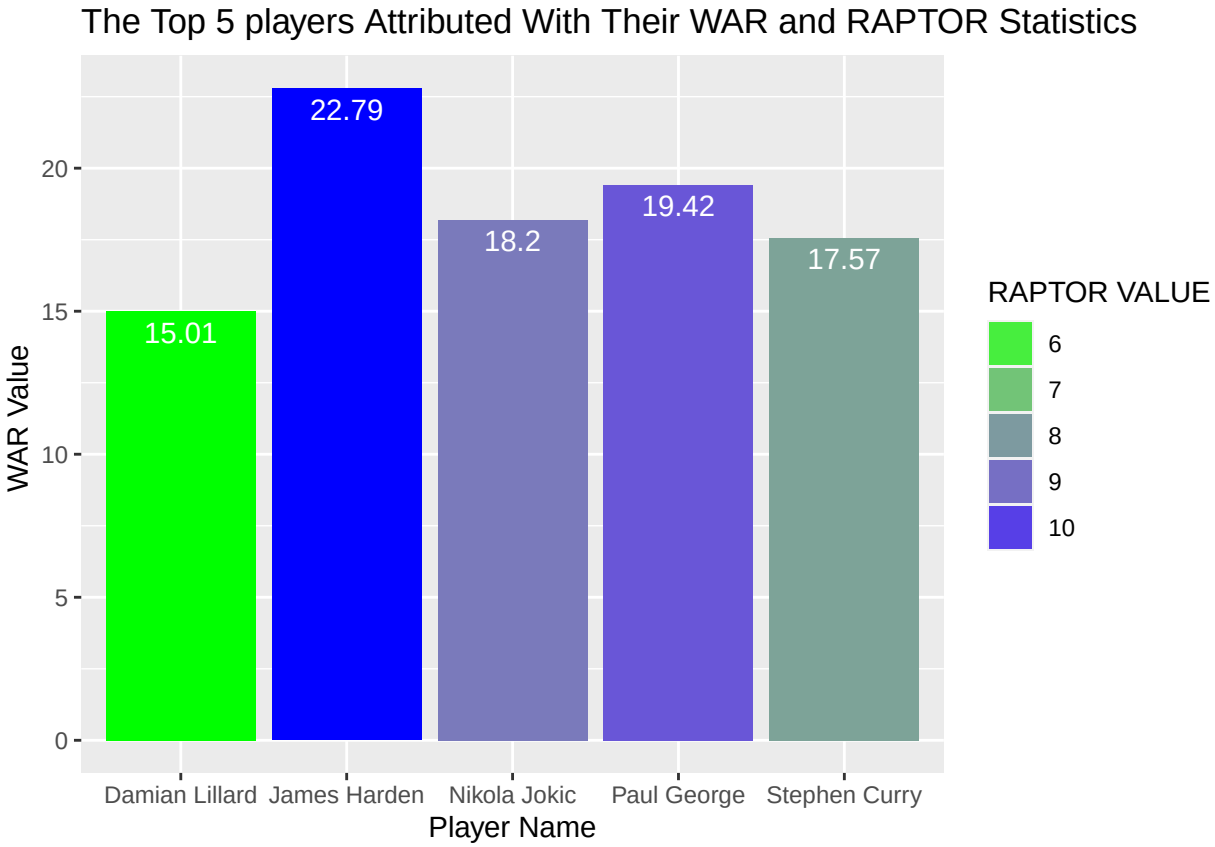
The data collected is from a population of the NBA's advanced statistics website in which they are even more categories to categorize the players. From their, they created the "Raptor" metric and sampled only players with at least 1,000 minutes played in a season since 2013-14.

Potential issues

Some potential issues that arise and are addressed in 538's data background. For example, they state that, "RAPTOR is a descriptive statistic. Descriptive RAPTOR is based solely on a player's on-court performance and the performance of the player's teammates, as described above. It does not use priors based on a player's height, weight, age or any other factor." The whole point of the Raptor statistic is to somehow synthesize so many variables to come to conclsuion on who the best player is. So though it is a limitation it still might be the most cutting edge statistic. Another limiation is that RAPTOR is based exclusively on publicly available data so we could always add to the regression algorithm and increase its efficieny in rating. Finally, in general plus minus statistics can falsely show that a players' improvement is linear, but does not take into account who the coach is, the type of basketball offense and defense system and other "synergies" as they call. They are so many real life metrics that can be hard to measure, but as statisticians we try our best to be as true to the data as possible.

Combined Numerical and Graphical Summary

```
#Numerical and GraphicalSummary
by_top5_player<-filter(RAPTOR19, RAPTOR19$war_total>15)
ggplot(by_top5_player, aes(x=player_name, y=war_total,fill=raptor_total))+
  geom_bar(stat="identity")+
  stat_summary(fun=mean, geom="text", aes(label=round(..y..,2)), vjust=1.5,
               position=position_dodge(0.9), color="white")+
  scale_fill_gradient(low="green", high="blue")+
  labs(title=
         "The Top 5 players Attributed With Their WAR and RAPTOR Statistics",
       x="Player Name",y = "WAR Value ")+
  guides(fill=guide_legend(title="RAPTOR VALUE"))
```



Conclusion

Throughout part 1 of the report, we can clearly see that James Harden had the best 2019-2020 season with the highest WAR value of 22.9 but also the highest RAPTOR value of 10.7. This goes to show that he is extremely valued to the team because he simply cannot be subsituted with four other mediocre players and that he has the best balance of offense and defensive abilties to maintain such a high RAPTOR and WAR value. The second conclusion, is that Lebron James is not mentioned anywhere in this data. This is shocking as he might be the greatest player of all time and it would be disrespectful to not include him in the data. So, this could potentially change up the best player in the NBA given that I would have to include his statistics to the dataset in part 2.

References

1. https://projects.fivethirtyeight.com/2020-nba-player-ratings/.
2. https://fivethirtyeight.com/features/how-our-raptor-metric-works/.