Homework 1 – Reflection Questions

After completing the coding exercises for Homework 1, answer the following questions in short sentences (1–3 sentences each):

1. What does FG% statistic mean?

Field goal percentage is essentially the percentage of basketball shots taken that end up being made by a player or team. It is an efficiency statistic for shot-making: the higher the FG%, the better a player or team is at making shots efficiently.

2. Histogram Insights: Looking at your histogram for FG%, what patterns do you notice? Is it symmetric, skewed, or uniform?

The FG% histogram has a distinct large peak from 40% to 50%. This makes sense, as most players in the NBA shoot at this efficiency level. Outside of this peak, the histogram is mostly symmetric, with a slightly larger tail on the left of the distribution.

3. What does the points per minute statistic mean?

Points per minute is a descriptive statistic that standardizes points scored by the number of minutes that a player receives. The flaw with points and points per game is that both statistics do not show how many scoring opportunities a player receives. To adjust for this, points per minute is a rate that shows how many points a player scores per minute that they play.

4. Looking at your histogram for points per minute, what patterns do you notice? Is it symmetric or skewed?

The histogram for points per minute has a distinct peak from 0.3 to 0.4 points per minute. The histogram is mostly symmetric, but the right tail is slightly larger than the left tail. This suggests that the distribution of points per minute is slightly skewed to the left.

5. How could a coach or team analyst use points per minute and plus-minus metrics when making decisions about lineups or playing time?

A coach or team analyst can use points per minute and plus-minus metrics instead of points per game when creating lineups or assigning playing time. A metric like points per minute gives an indication of how efficient a player is at scoring for each opportunity they get, rather than

showing how many points they can score across the entire game. Plus-minus shows how well the team does at scoring differential when a certain player is on the court, which contributes directly to whether or not the team wins games.

6. Based on what you know about advanced player statistics in basketball, if you were a GM and had to pick **one** advanced statistic to prioritize for individual player evaluation – PER, NBA Efficiency Rating, BSB Win Score, or Plus-Minus – which one would you pick? As part of your answer, please define each statistic, and build a strong argumentation for outlining which would be your "favorite" statistic to use?

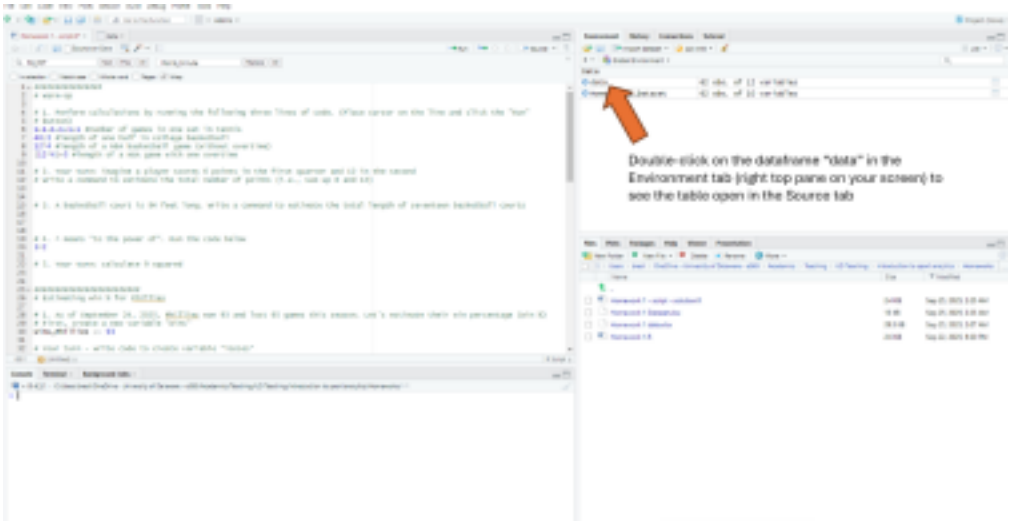
As mentioned above, plus-minus shows how well the team does at scoring differential when a certain player is on the court. This metric is flawed because an individual's plus-minus is largely dependent on how strong their teammates are when they are on the court. Hollinger's PER, which is Player Efficiency Rating, takes the major events of the game and standardizes a player's efficiency against the rest of the NBA. This metric is flawed because it rewards low efficiency scorers too much. Similarly, NBA Efficiency Rating shows an NBA player's efficiency with each of the major events of an NBA possession (missed FGs, missed FTs, points, assists, rebounds, turnovers, and steals). This metric is also flawed because it rewards high volume scorers, even if they are not efficient. Finally, BSB Win Score is a weighted efficiency statistic that applies weights to major events that can happen during a possession. I believe that this is the best of the four statistics, because it isolates one player's contributions to the game and applies weights that create a better balance between scoring efficiency and other contributions.

7. For you personally, what was the most challenging or unclear aspect of this homework?

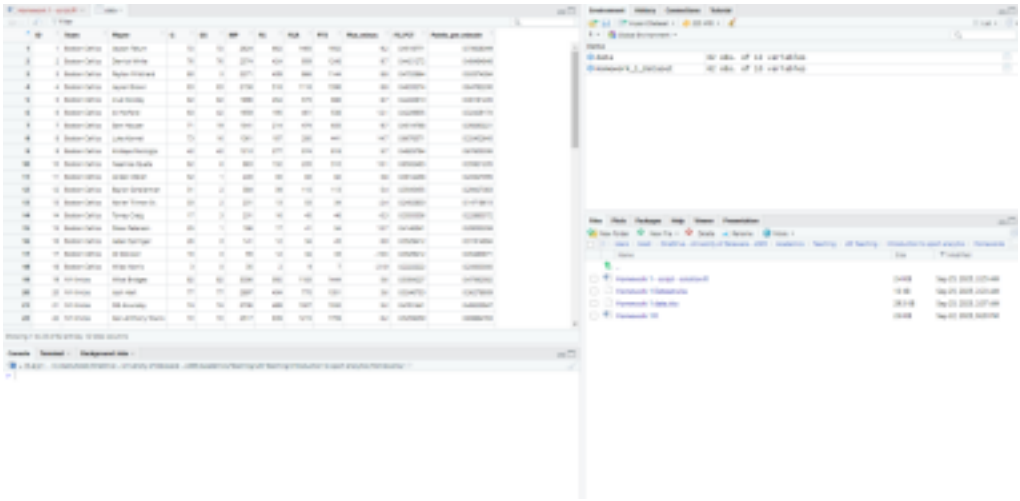
The most challenging part of this homework assignment was getting everything to work in R Studio. When I typed out the code for the histogram, there were a few syntax errors I had to fix before the code successfully generated the desired output.

Extra credit (3%)

Next, double-click on your dataframe called "data" in the Environment tab.



The dataframe should open in the Source pane (see the picture below).



This table is sortable – i.e., if you click on the column names, you should be able to sort these columns.

8. Sort the points per minute column. Which player has the highest points per minute? How many minutes did they play last season and is this measure reliable for this player? What does this tell us about the importance of time played on the reliability/accuracy of player statistics?

On my end, Marjon Beauchamp led the sample in points per minute, with just above 0.88 points per minute. However, he only played 17 minutes last season, which is an extremely low number of minutes compared to the rest of the players in the sample. Therefore, this calculated statistic for points per minute is not reliable for Marjon Beauchamp. This shows that the sample size of minutes played is important when creating an efficiency statistic with minutes played as one of the metrics considered.

9. Multiple players (e.g., Chris Boucher, RJ Barrett, Matt Ryan) have high points per minute yet a negative plus-minus. Explain why this might be the case. What does it tell us about their individual contributions relative to the team performance?

This suggests that when these players are on the floor, they score a lot per minute that they play but the team still loses the battle for point differential. This could suggest that these players are solid scorers, but the team does not play great defense when they are on the floor. This situation tells us that strong individual contributions can be misleading or meaningless if the team performance is not simultaneously strong.

10. Do you notice any apparent pattern or relationship between FG%, points per minute, and plus-minus in your dataset? Explain why/why not.

There is a slight positive relationship between FG% and points per minute in the sample. As a player's FG% improves, their points per minute generally increases. This makes natural sense, as making a higher percentage of shots results in more points scored, which improves points per minute. There is not much of a relationship between plus-minus and FG% or points per minute. This suggests that even if a player shoots efficiently from the field or by minutes played, the team performance might not necessarily be good from a point differential perspective. This proves that team-based statistics may not necessarily be positive just because a player's scoring performances are good.