Final Project: NBA Playoffs

Han Nguyen

May 5, 2018

Project Summary

This project aims to analyze real time data of the National Basketball Association in order to predict end results of a basketball game. In this project, I sit through Game 6 of the Boston Celtics vs. Milwaukee Bucks Playoffs season and gather data every two minutes to predict end results of the game as time passes on. We will look into how each team has ranked throughout the regular 2017-2018 season, player lineups and injuries, the four factors that help win a game, and real time tweets during the game. I focus more on data of the Boston Celtics in this project.

Introduction to the Celtics

The Boston Celtics have been a highly ranked team for many years, especially during the era of the Big Three: Paul Pierce, Ray Allen, Kevin Garnett. They won a championship in 2008, and have yet to win one. After the big three got traded, the Celtics have been trying to rebuild their team with the new coaching of Brad Stevens. This year, we added a top player to the roster, Kyrie Irving, which gave the Celtics some hope of going far in the Playoffs this year, until he got injured, which we will discuss later.

Four Factors

The "Four Factors of Basketball Success" are Shooting(Field Goals and 3 Pointers), Turnovers, Rebounds, and Free Throws. In this project, we focus more on these four factors when taking account of predicting results and analysis.

Data

The data used in this project were from the NBA website, and Twitter. The Twitter data was collected during the game (halftime, after 3rd quarter, and final) and saved in files. The R code for retrieving this data is also in the project folder titled, Final Project Twitter API. The data from the NBA website includes box scores of previous Celtics vs. Bucks games in the season, and Player Statistics during Game 6.

Celtics vs Bucks Regular Season Stats

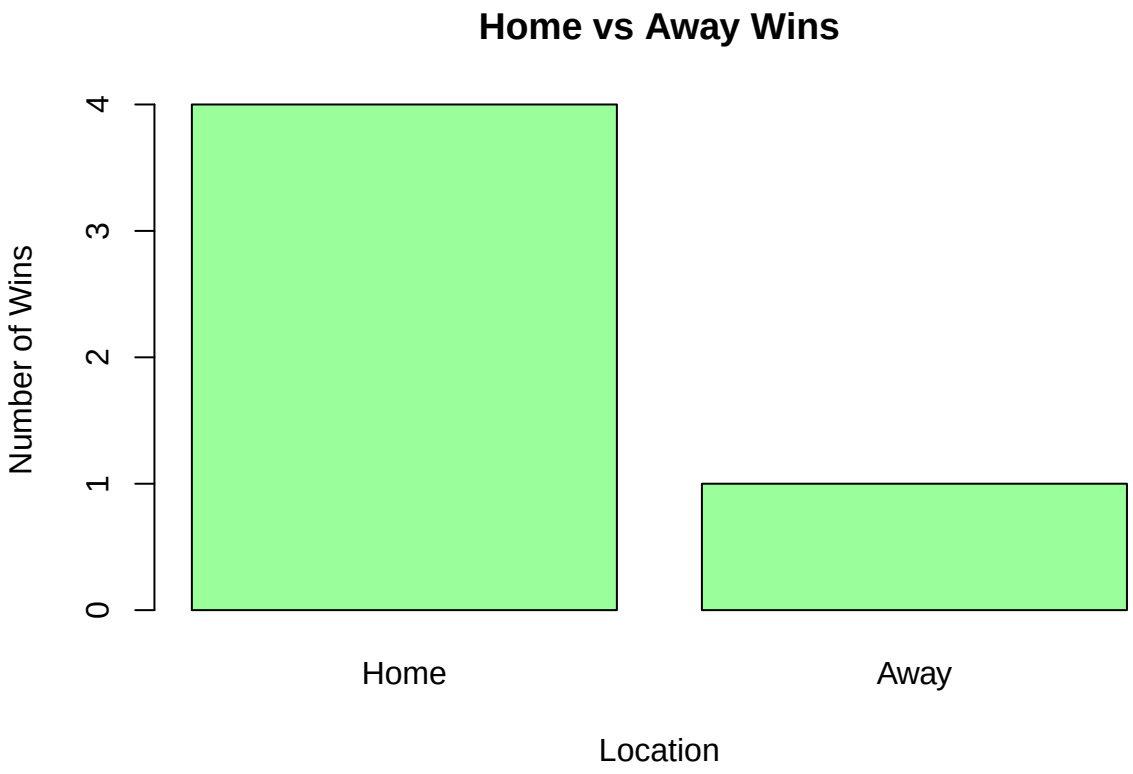
Table 1: Regular Season Stats

CELTICS	REGULAR SEASON	BUCKS
55-27	Win-Lose	44-38
2	Rank	7
67.1	Wins	53.7
32.9	Losses	46.3
1 W	Streak	1 L
4-Jun	Last 10	4-Jun
27-14	Home	25-16
28-13	Away	19-22

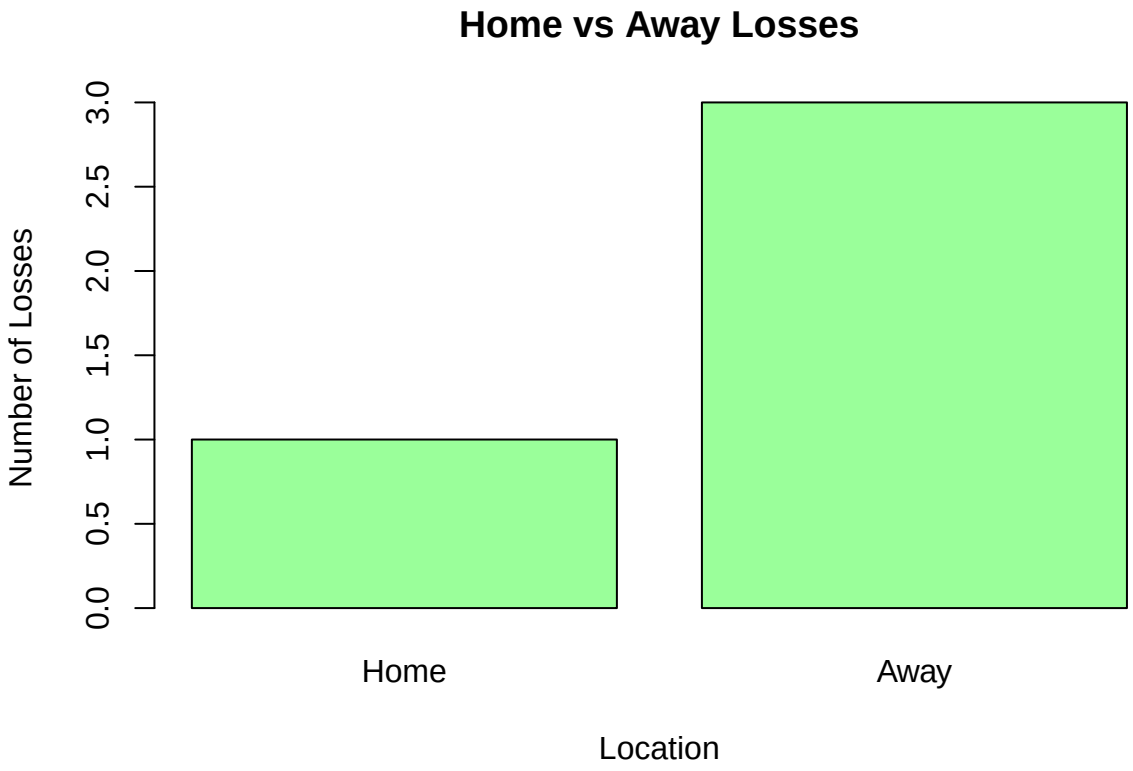
We see that the Celtics should have a better advantage in the playoffs against the Bucks because of their already high ranking during the season. The Celtics have higher statistics overall, which means they should be more likely to win the playoffs against the Bucks, which we will analyze with the real time data.

Home vs Away Games

In general, teams have a better advantage when playing in their home arena because of the "motivation" of the fans cheering their own team on. This is seen in these bar plots for the Celtics vs Bucks games throughout the season.



It is seen that the Celtics clearly win most of their home games.

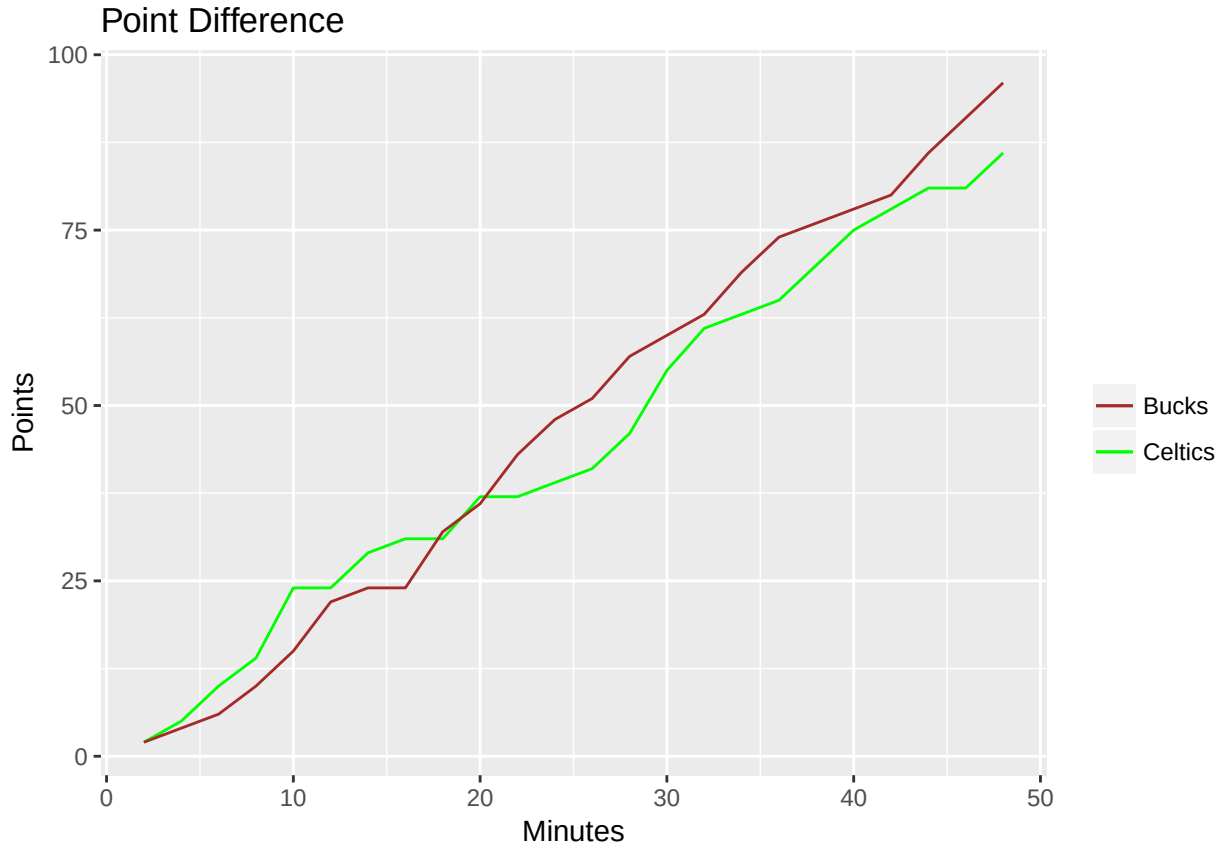


And it is seen that they lose most of their away games vs the Bucks.

With this information, one can predict before the game starts that the Celtics will not likely win game 6 of the playoffs, because game 6 will be an away game for the Celtics.

Real Time Point Differences

This plot shows the real time number of points each time has gotten every 2 minutes of the game. Obviously, the closer towards the end of the game, the easier to predict who will win the game or not.

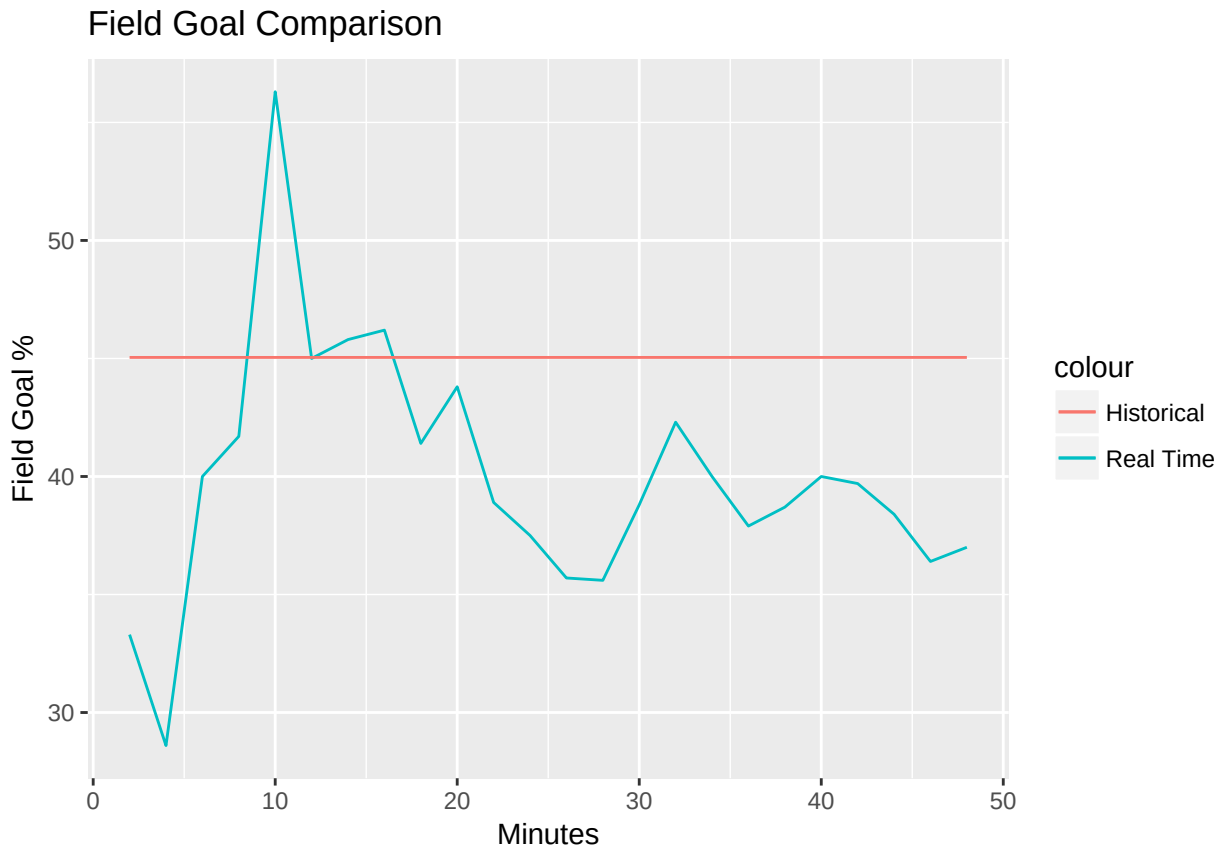


For the first quarter of the game, and into the second quarter, the Celtics were always in the lead. At this point, one can say that the Celtics may win the game because they are already holding the lead, but of course it is too early to tell since anything can happen in the remaining quarters. By half time, the Bucks caught up with the Celtics, but the scores were still close. So the game can end with a win or lose for the Celtics. As we ended the 3rd quarter, the Bucks continued to hold their lead, and it was likely to see that the Celtics will lose the game, since the Bucks were in the lead for so long already. There were times when both teams were close in points to each other, but with the Bucks in the lead for the majority of the game already, it was clear that the outcome would be a loss for the Celtics.

Historical vs Real Time Statistics

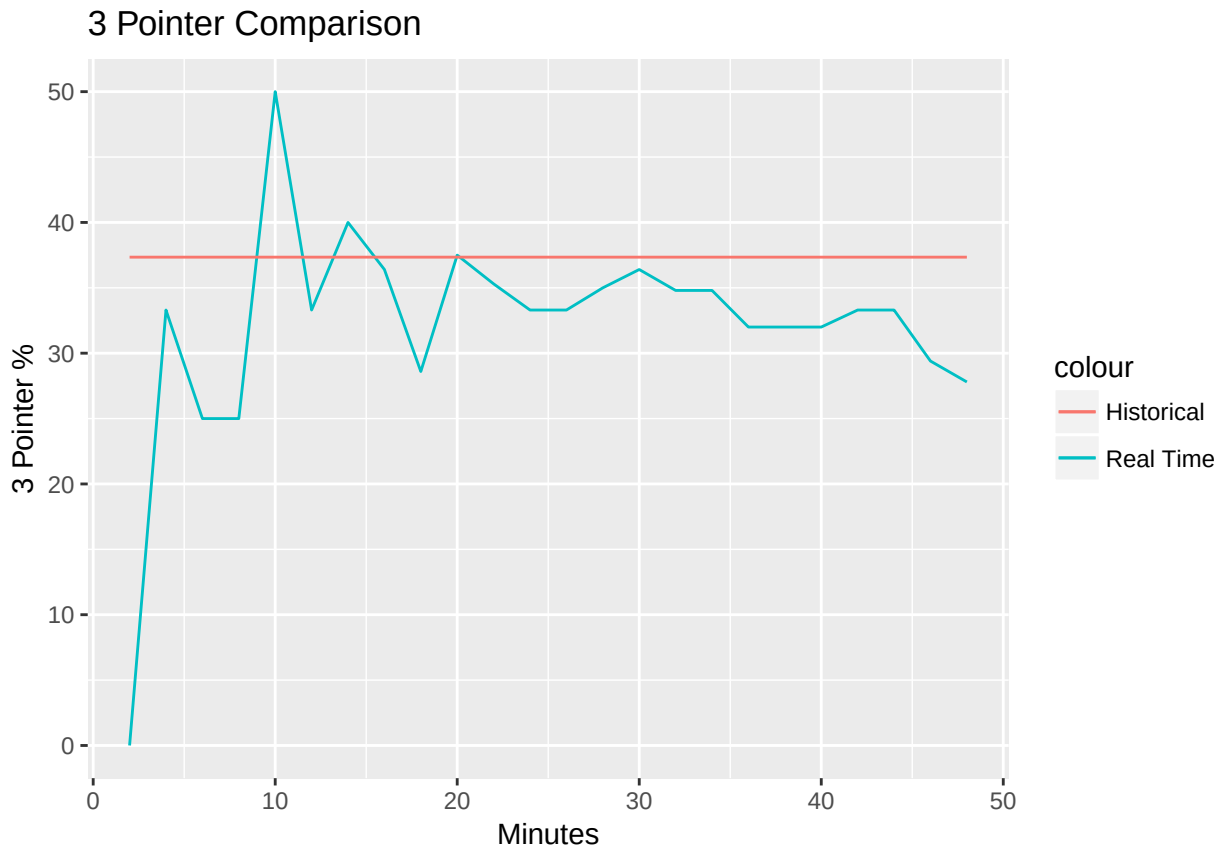
In this section, we will look at the averages of the "Four Factors",Field Goal Percentages, 3 Pointer Percentages, Free Throw Percentages, Turnovers, and Rebounds, in the past Celtics vs Bucks games, and how that average compared to the real time data of game 6, to see if the Celtics have been exceeding their average or below it. For each factor, we plot the average line representing historical data, and the real time data of game 6.

Field Goals



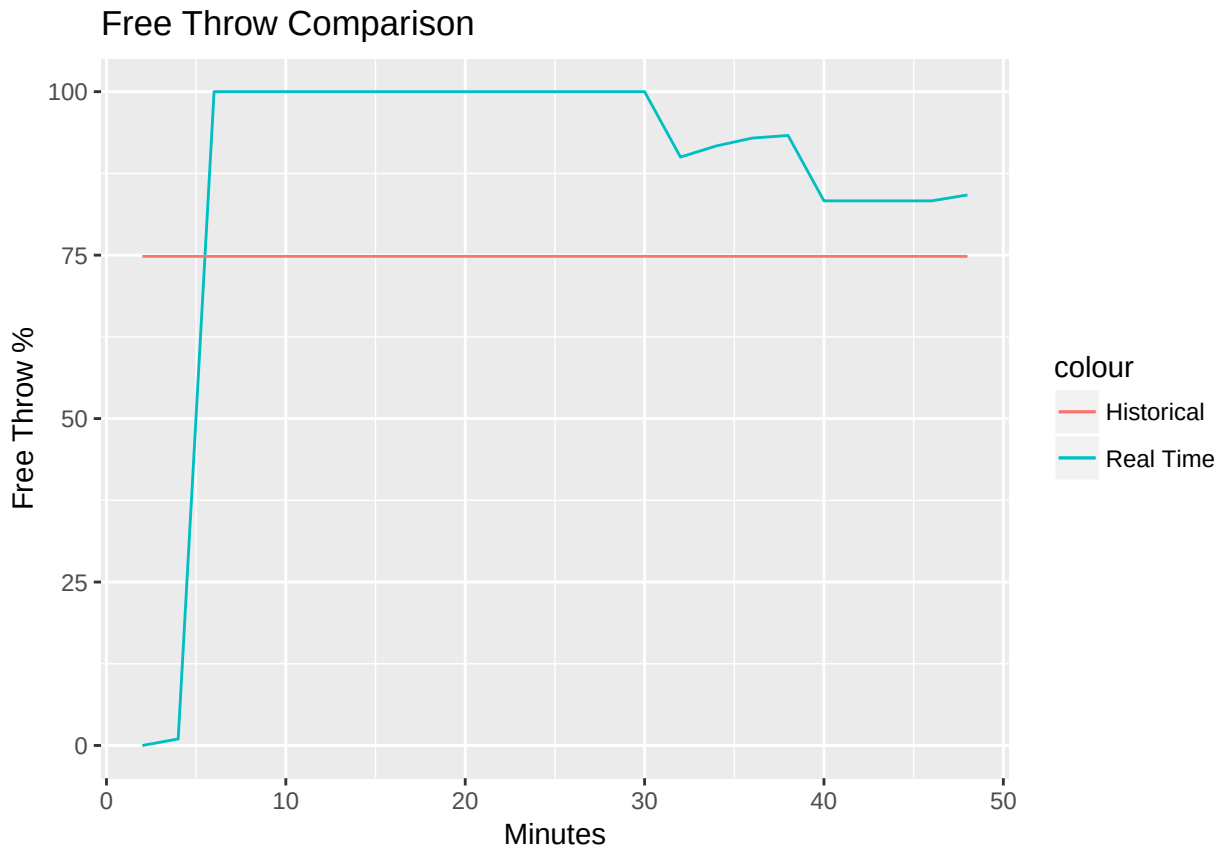
We see that the field goal percentage during the game was only above average in the first quarter into the 2nd quarter, which explains the previous plot where the Celtics were in the lead at the same time. It quickly declined and continued to stay below the average of the previous games. During the game, one can see that since the field goal percentages continued to fall below the average, we can predict that the Celtics will lose the game since they aren't shooting field goals as well as they have previously.

3 Pointers



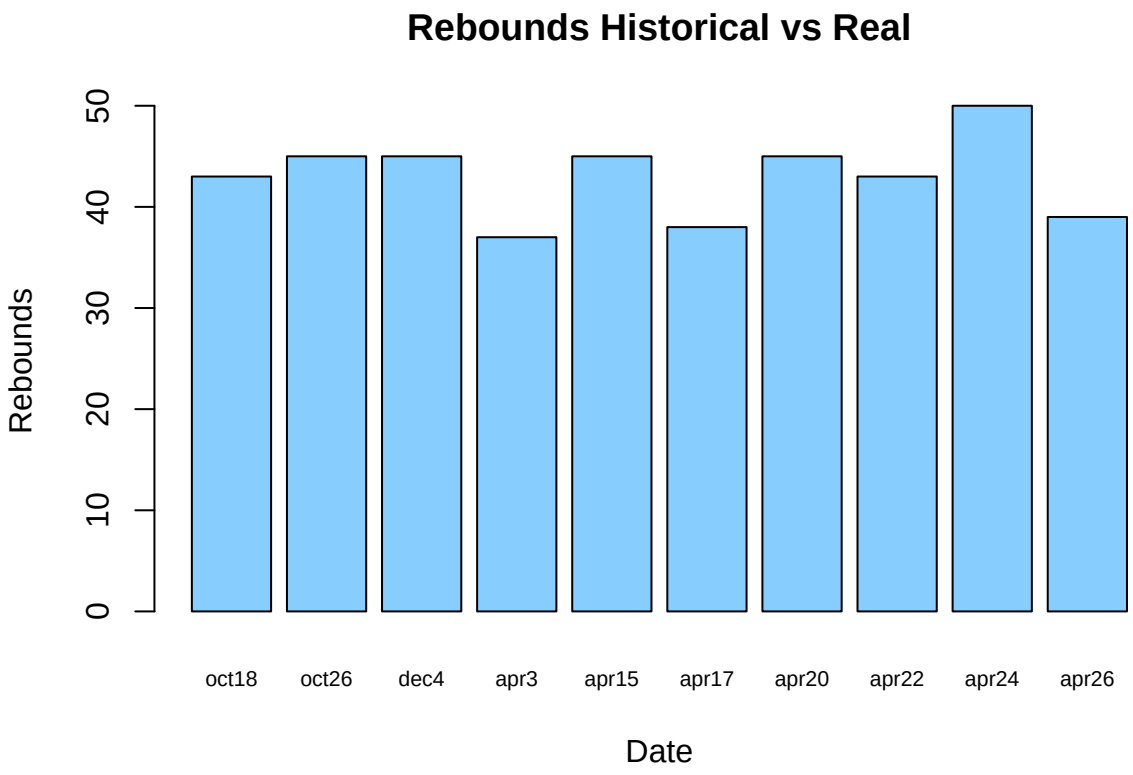
We, again, see that the only times the 3 Pointer percetages were above the average was when the Celtics were in the lead. Otherwise, throughout the game, the percentage fell below average, although it didn't fall too far below the line. Since the 3 point percentages fell mostly below the average percentage of the previous games, it is more likely that the Celtics are not shooting as well, which will lead to a lower score.

Free Throws

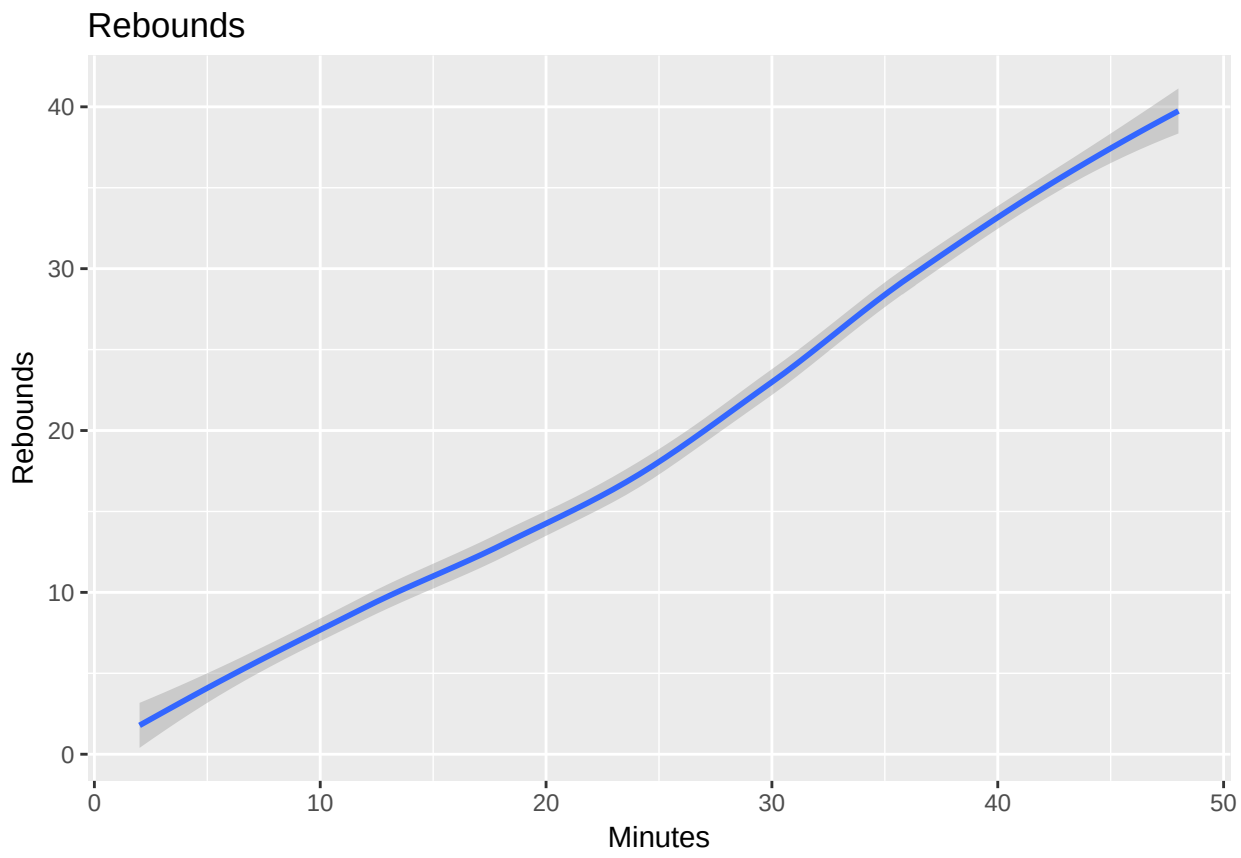


Throughout the game, the free throw percentage was mostly always above the average. Free throw percentages are generally high, and are not as important as Field Goals or 3 Pointers, since they only count as 1 point and don't occur as often as Field Goals or 3 Pointers. It isn't as important as the other factors in basketball success.

Rebounds



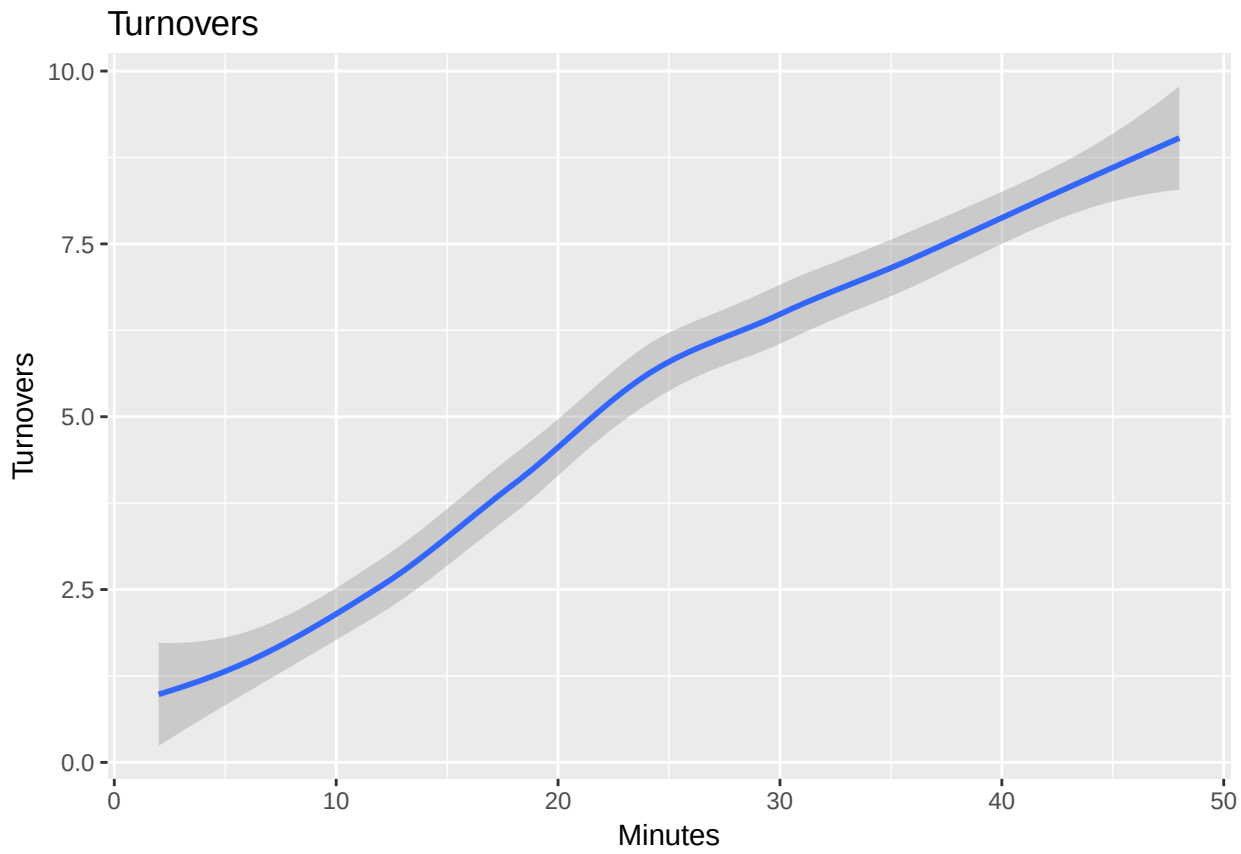
The barplot shows the number of rebounds in all previous games, as well as Game 6 which was on April 26. We see that the number of rebounds for the present game was close to the previous dates, but it is still lower than most of the other games, which contributes to the loss.



The number of rebounds also increased slowly throughout the game, as shown in the second graph.

Turnovers

The barplot shows the number of turnovers in the present game on April 26, and the number of turnovers in the previous games. It is seen that the number of turnovers is actually lower than most of those from previous games, besides two days (April 17 and April 22).



The number of turnovers also had a steady growth in the second graph.

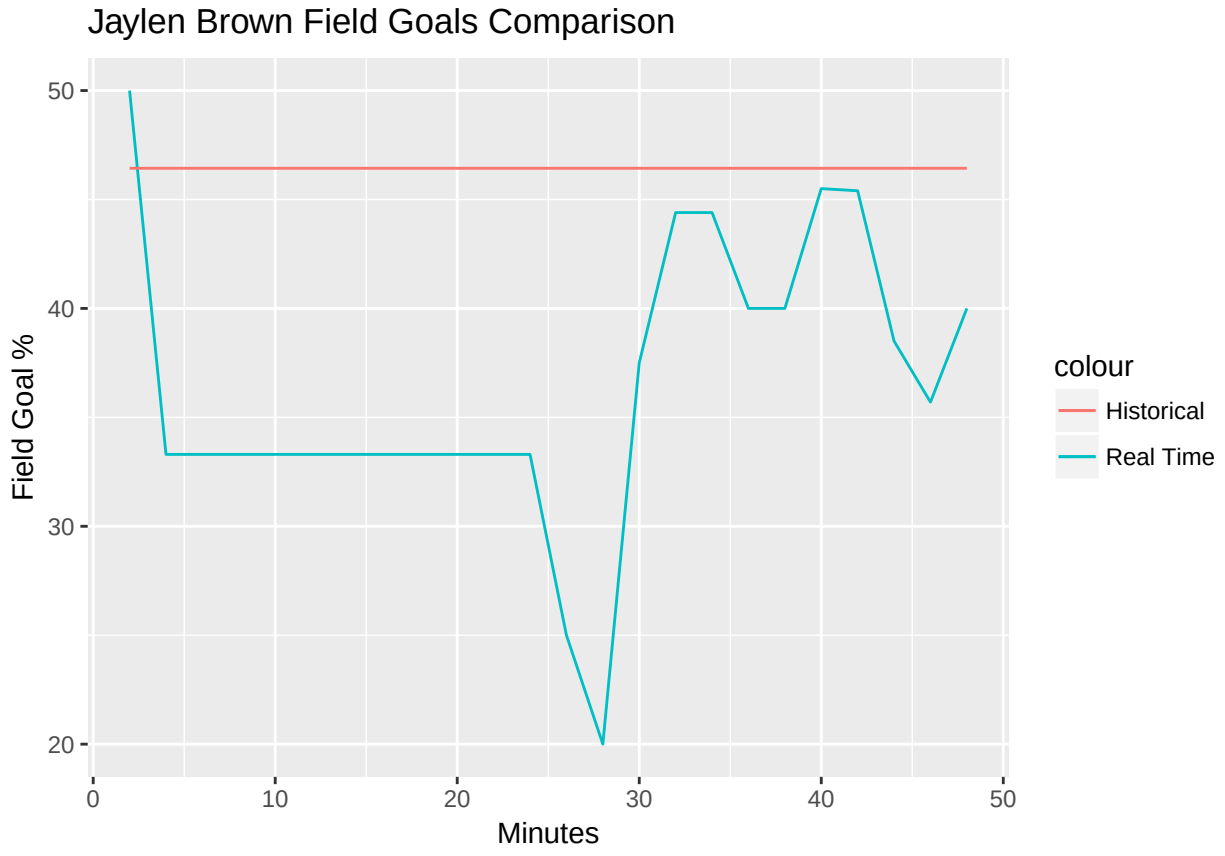
Overall of Four Factors

It is seen that the four factors (besides free throws) were below average throughout most of the entirety of the game. This all contributes to a lower score and an unwell played game, so it was straightforward to predict that the Celtics would lose the game, seeing that they weren't playing to the best of their ability and reaching their averages.

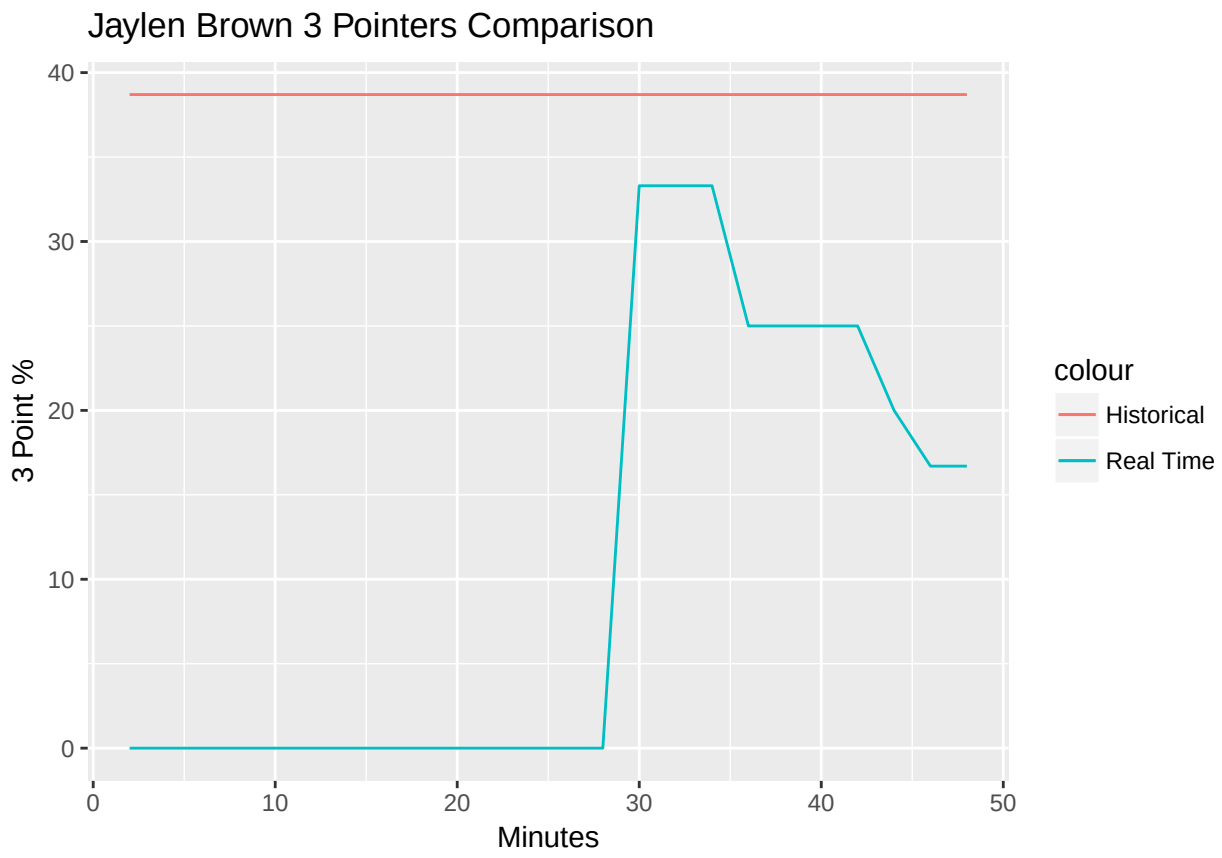
Player Statistics: Historical vs Real Time

In this section, we look into key players in the regular starting lineup of each game. As we mentioned previously, Kyrie Irving came into the Celtics to uphold the team. He was the star of the Celtics rebuild. Kyrie put up many points and great statistics in the games he played this season. Unfortunately, he had a knee injury and had to undergo surgery, meaning he would miss the remainder of the games which included all of the playoffs. This means that he missed the game that we are analyzing right now. What does this mean? Would Kyrie really have an affect on the rest of the playoffs? Or would the other players be able to reach these expectations? The key players we will be analyzing are Jaylen Brown, Jayson Tatum, Al Horford, Marcus Smart, Marcus Morris, Terry Rozier, and of course, Kyrie Irving. Similarly to the previous section, we will be analyzing each player's averages in shooting in previous games vs the real time game.

Jaylen Brown



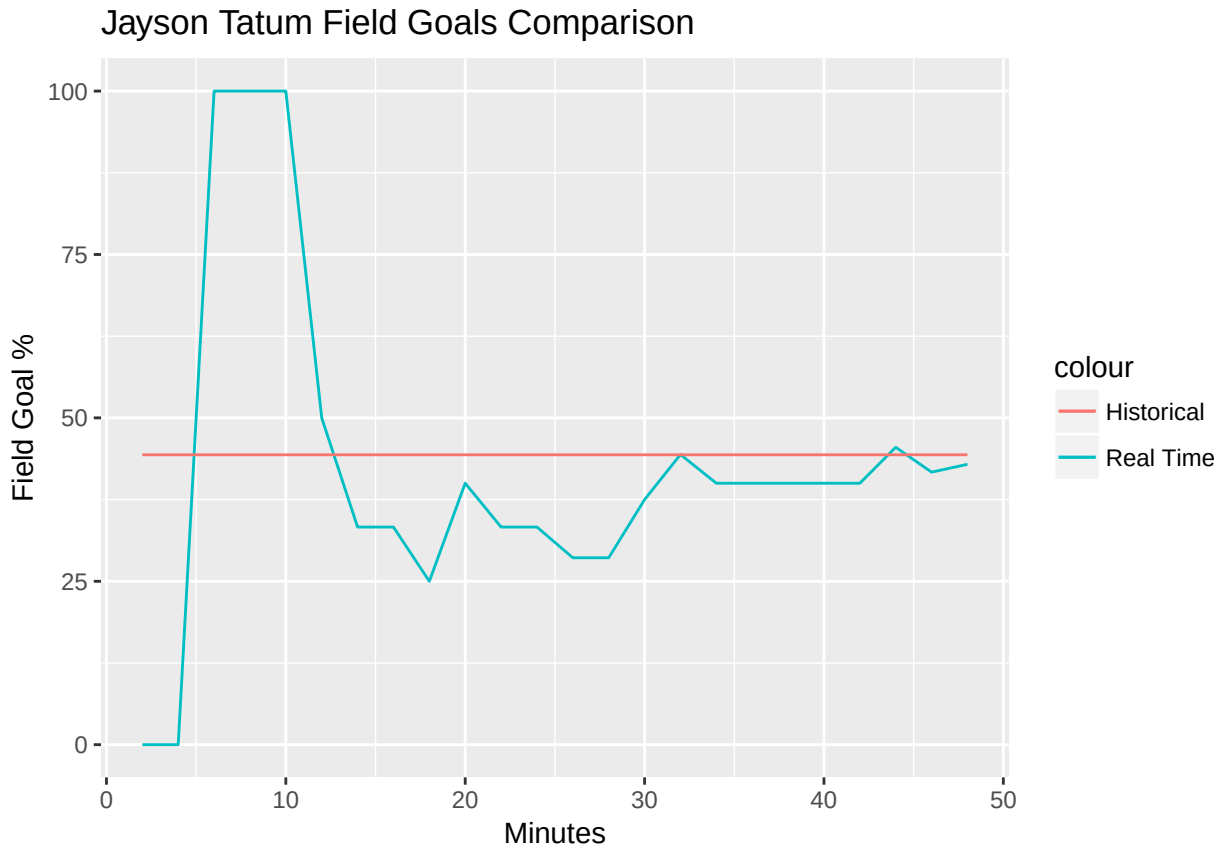
Jaylen Brown generally has a high average field goal percentage in his previous games against the Bucks. While looking at the real time data, we see that his field goal percentage throughout the game was a lot lower than his average. He was not playing as well in this game as previous games, which we may predict contribute to the end result of the overall game: a loss.



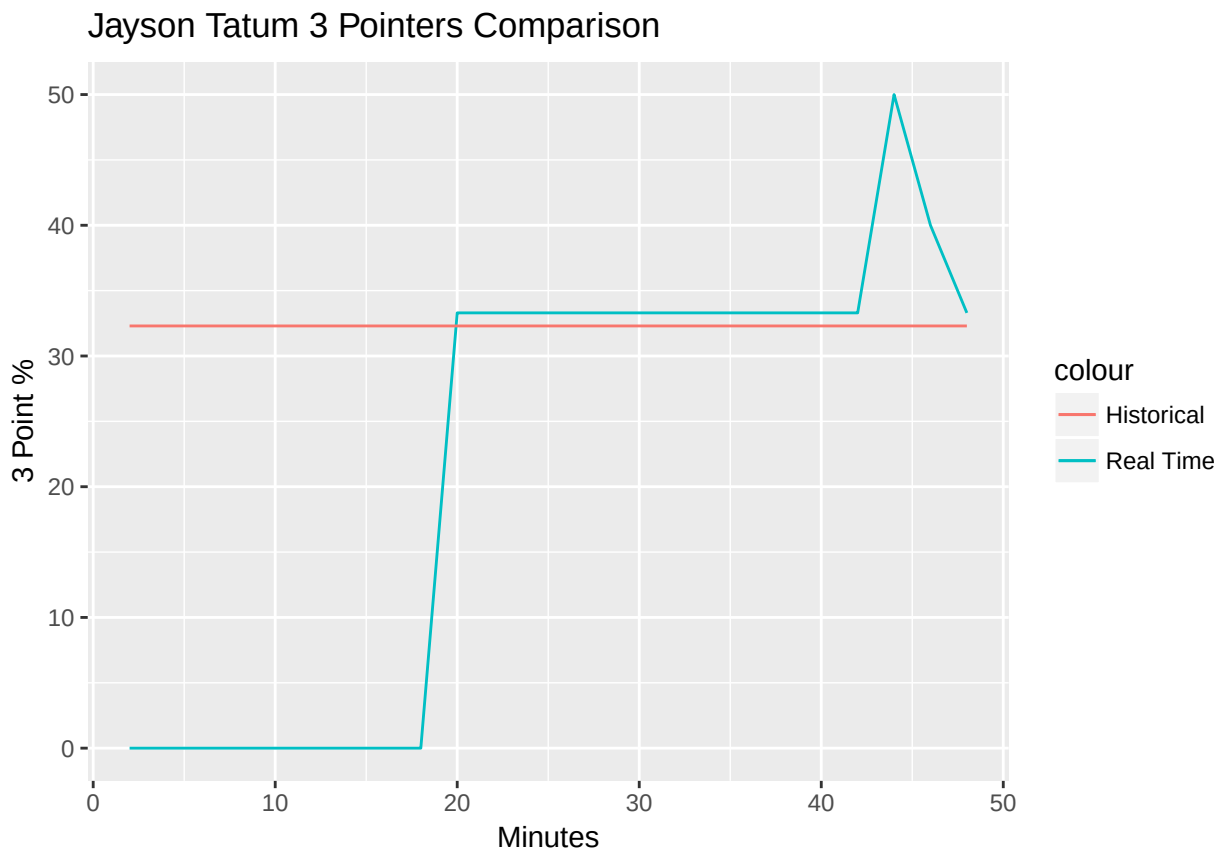
He did not make any 3 pointers for the first half of the game and way into the 3rd quarter. His average is pretty high, but during game 6, Jaylen Brown's 3 pointers fell short and below average.

With the below average field goal and 3 pointer percentages, we see that he did not put up as many points as he usually would, which would significantly contribute to the loss of the game.

Jayson Tatum

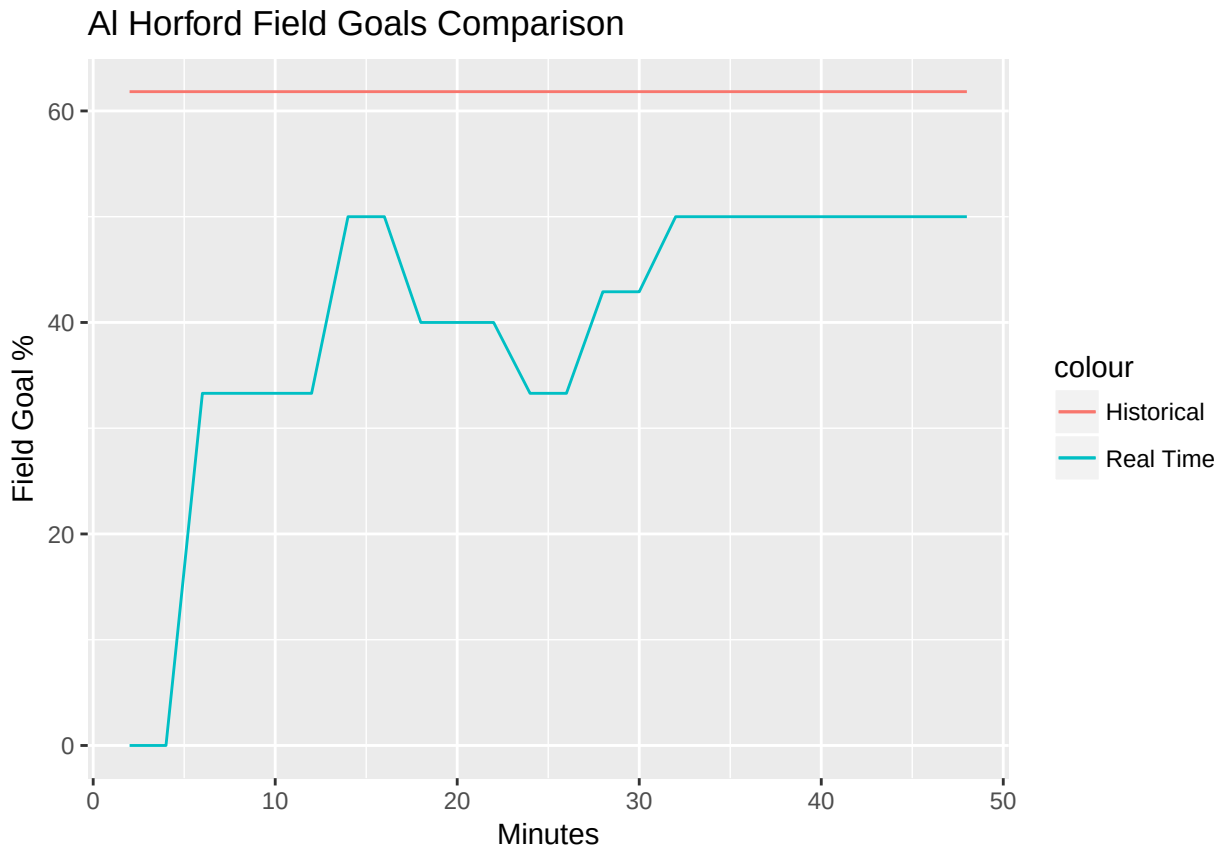


Jayson Tatum's field goal percentage throughout the game was only above average during the time the Celtics were in the lead. His contribution fell short though because he ended up being below his average for most of the remainder of the game, which would lead to the loss.

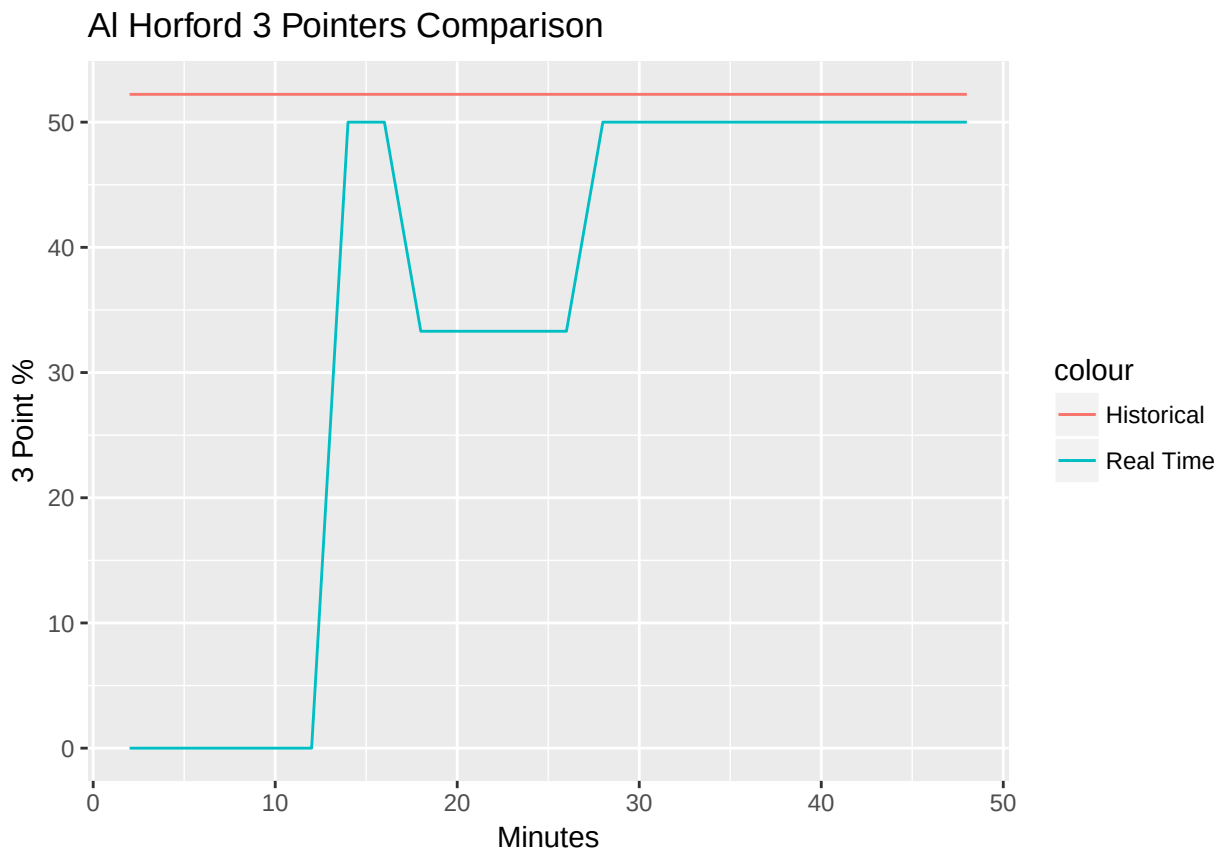


He put up less 3 pointer attempts in this game, seeing that he didn't make a shot until nearly halftime, and did not make any more 3 pointers until the end, which would not help with the significant point difference that there already was for the Celtics and the Bucks. On average, Jayson Tatum would have a higher 3 point percentage with more attempts.

Al Horford

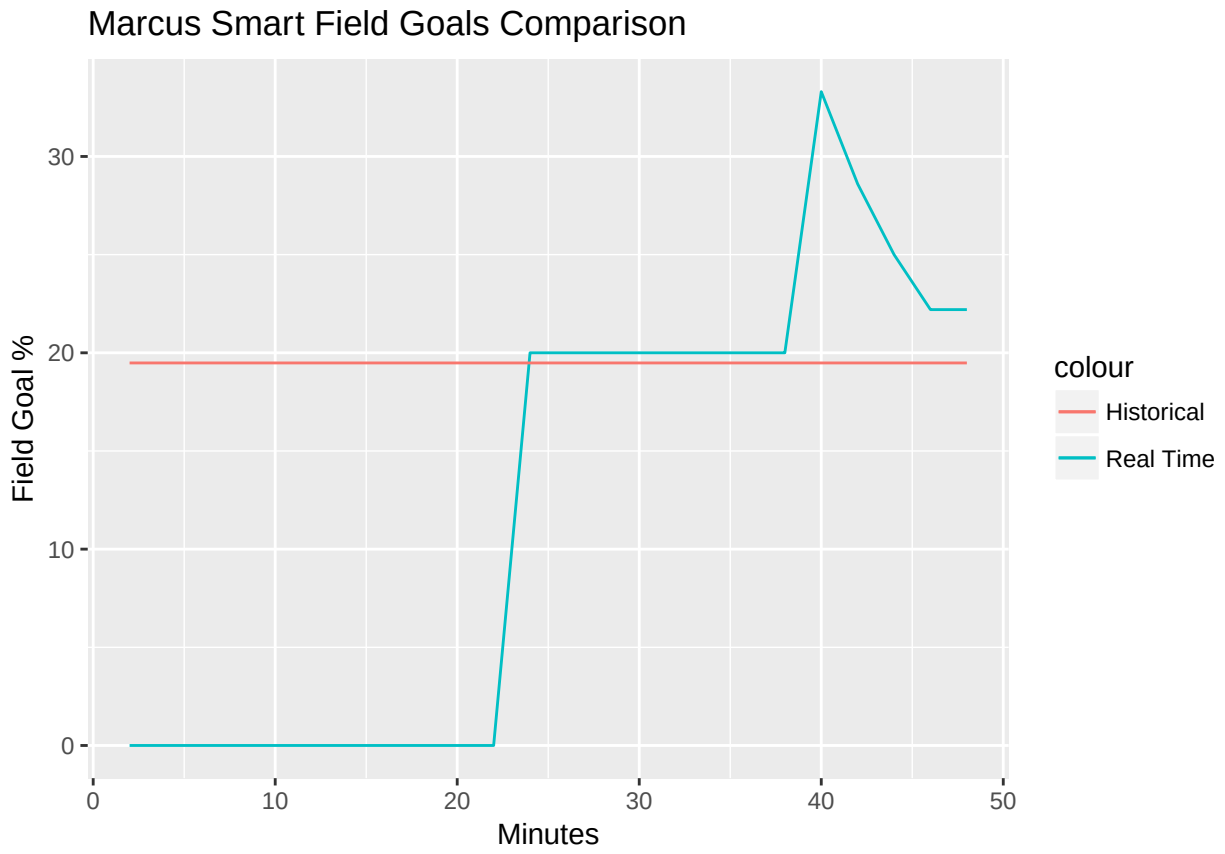


Al Horford typically, on average, has a high field goal percentage. His average is over 60%, but in game 6, we see that he doesn't come close to his average throughout the entirety of the game. His highest in the game was about 50%, so we see that he did not contribute much to the score throughout the game.

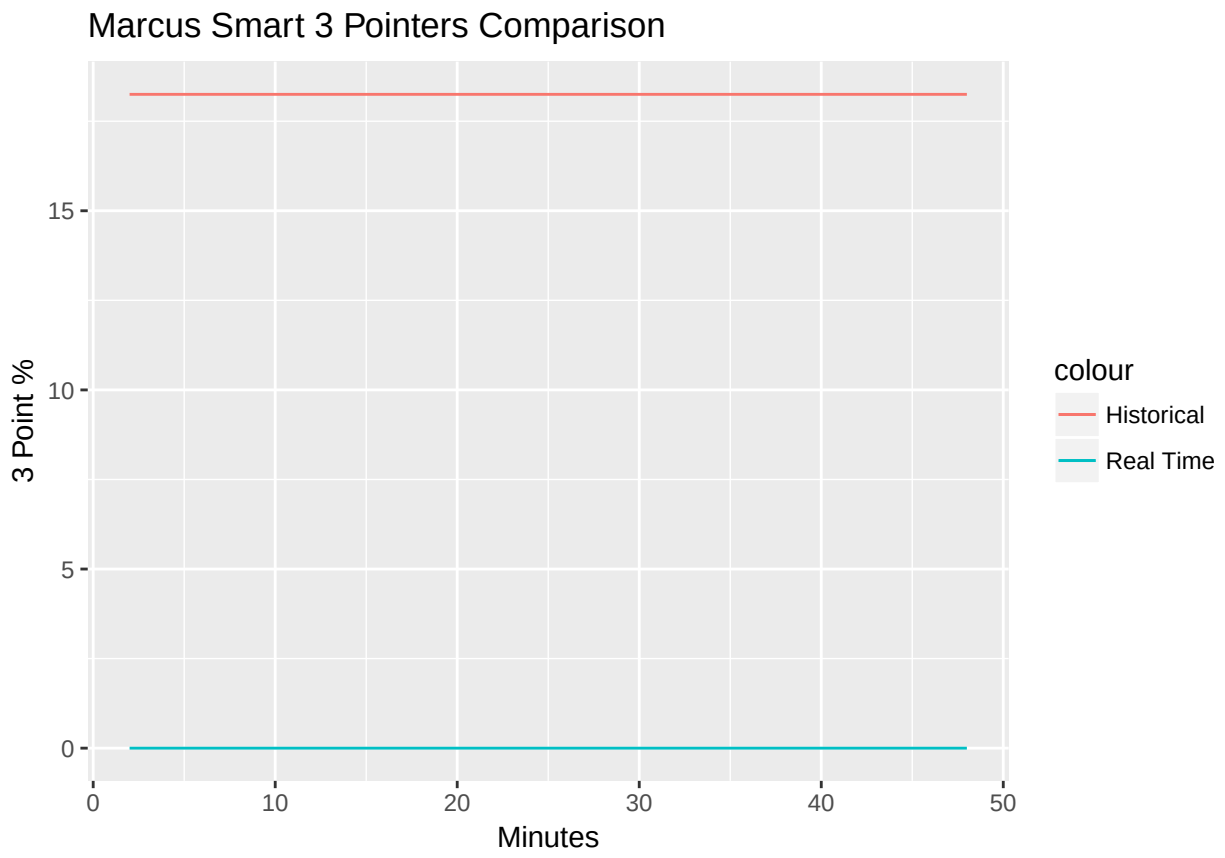


He also generally has a high 3 pointer percentage, with over 50%. Again, he did not go above his average in this game, and barely made any 3 pointers, as can be seen in the constant lines where he did not attempt to shoot any 3s.

Marcus Smart

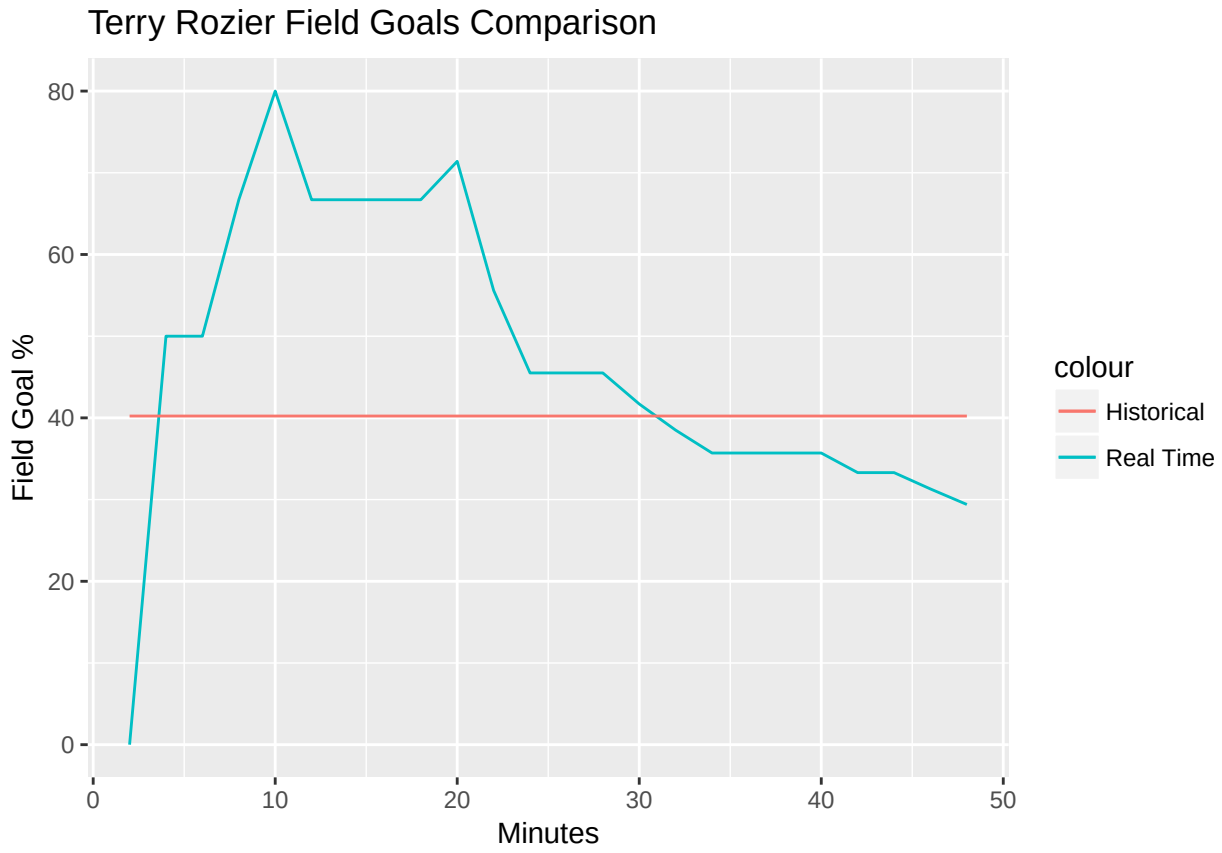


Marcus Smart is not a very strong shooter in general. He only has barely 20% for average field goal percentages. During game 6, we see that his percentages did not change much. He did go above average towards second half, but with less attempts at a field goal. So with less attempts, there was not much of a contribution to the score.

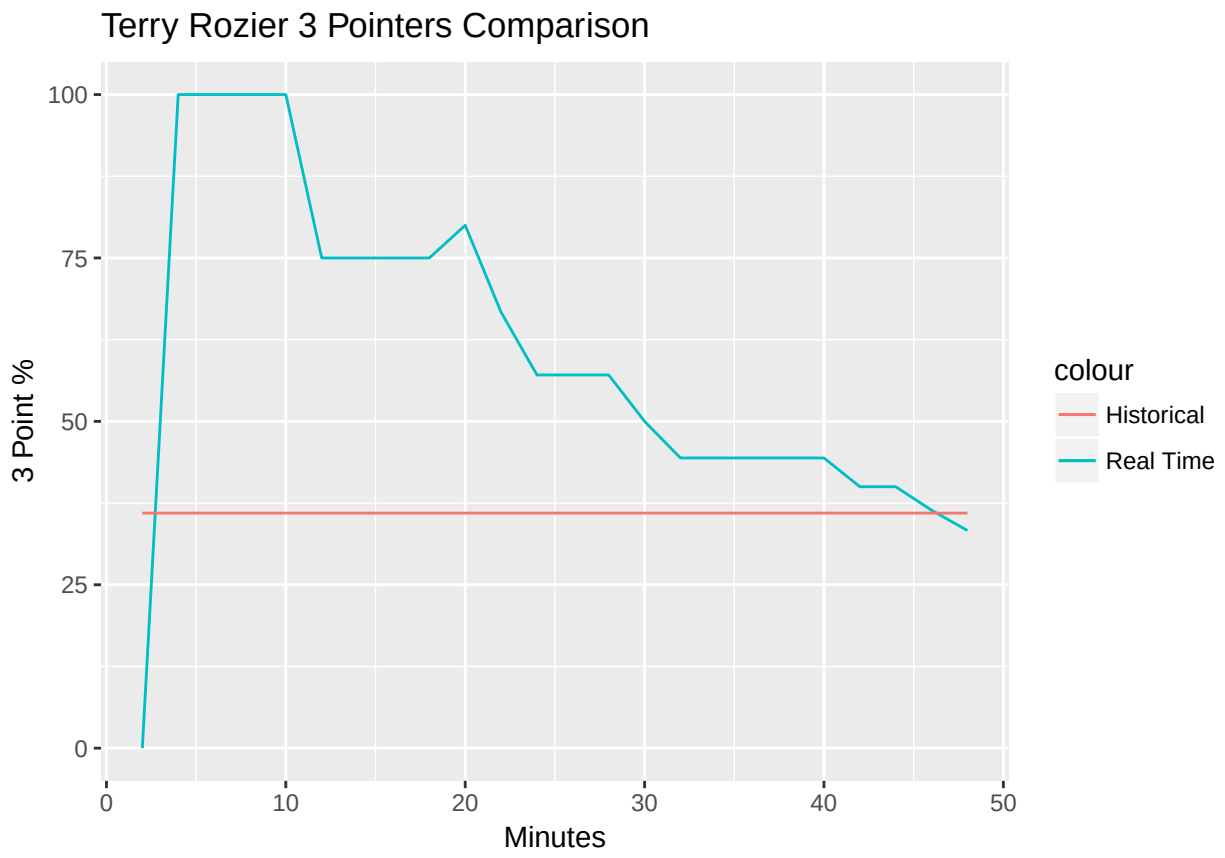


He is also not a strong 3 point shooter. On average, he is about 18% for 3 pointers, but in this game, he did not make any shoots. This is significant to the score, because he did not contribute any 3 points throughout the game.

Terry Rozier

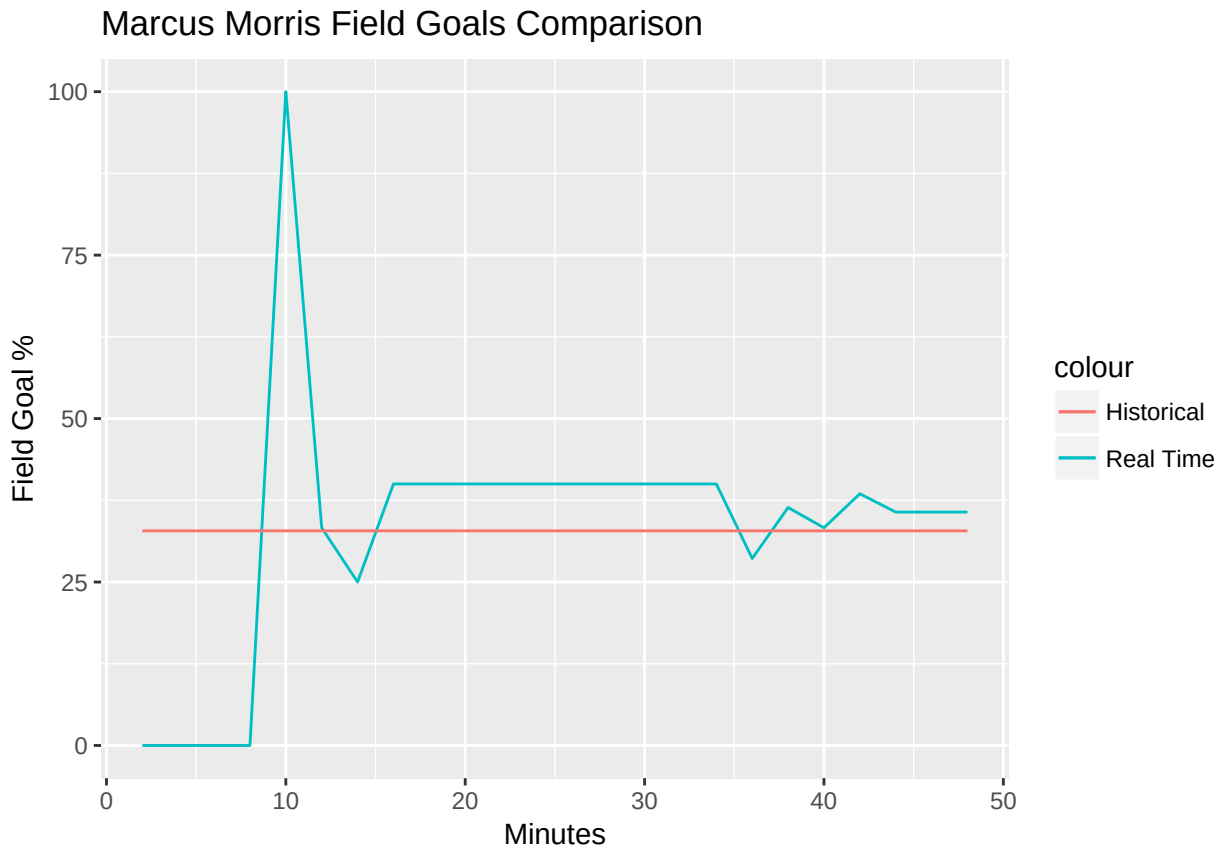


Terry Rozier, unlike the previous players, did above his average throughout the game until the last quarter. He was always around 70% for field goal percentages. Towards the last quarter, he fell below his average, which contributed to the Celtics not being able to bring up their score near the end and lose the game.

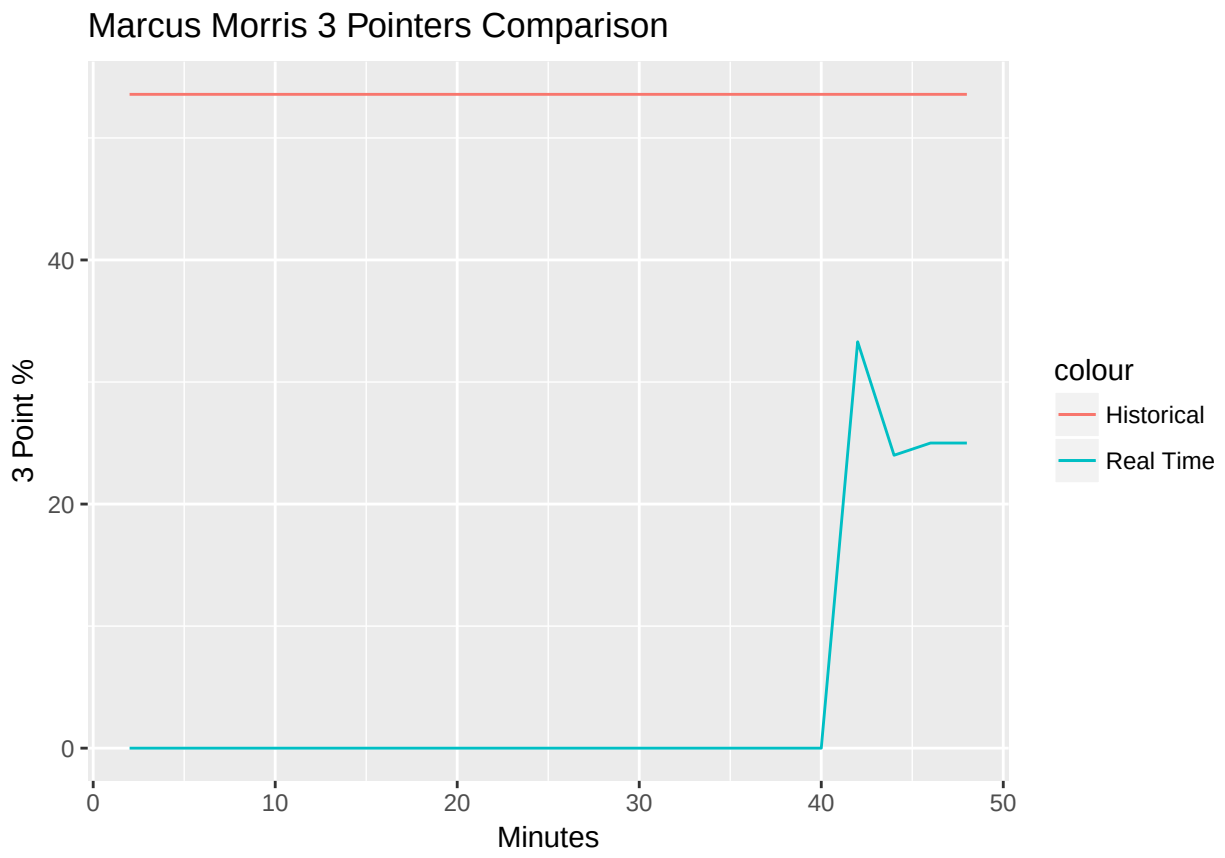


He is a very strong 3 point shooter, as can be seen in game 6. He was above average the whole time, and started strong when the Celtics were in the lead. But, his 3 point percentage declined throughout the game, which led to less points being contributed to the score, and the Celtics losing.

Marcus Morris



Marcus Morris did not play in many of the previous games. But he did go above his average throughout game 6, but did not make as many attempts. With fewer attempts, there is not much contribution to the score.



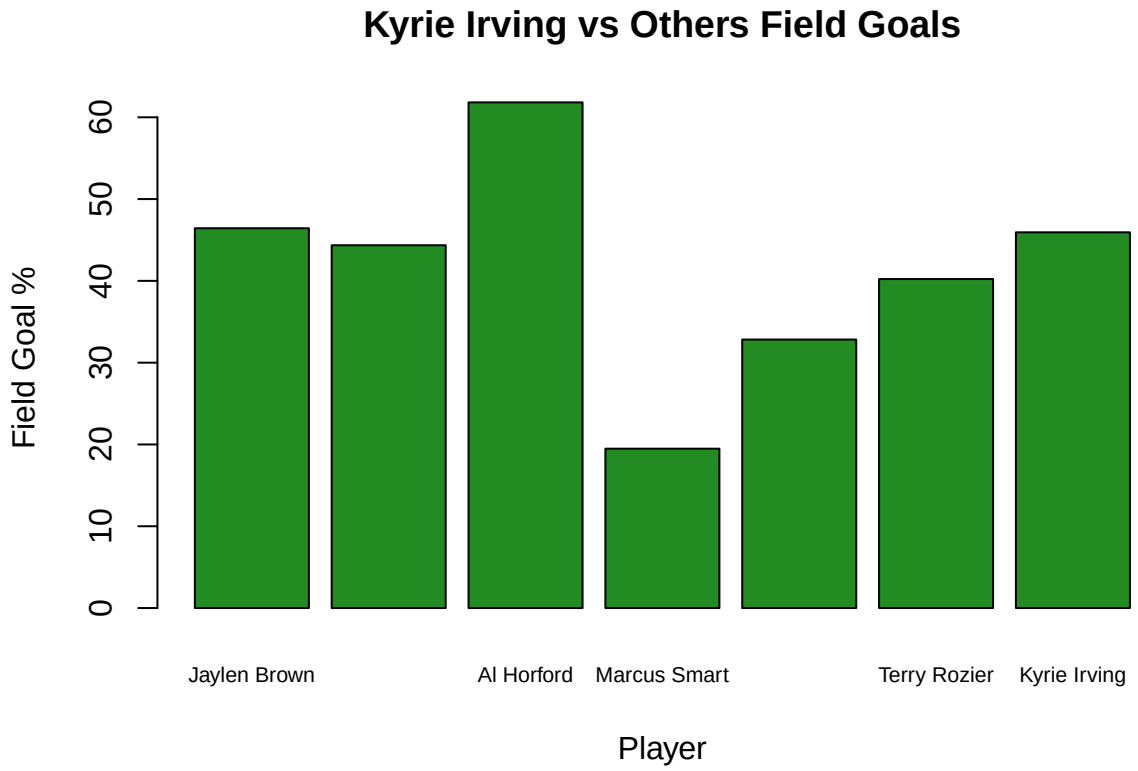
He generally has a decent 3 point shooting percentage, but did not make any shots until the very last quarter. He was well below his average even for this short period of time. With few attempts at 3 pointers and falling below average, there wasn't much contribution to the score, leading to a loss of the game.

Overall Game 6 Real Time Players

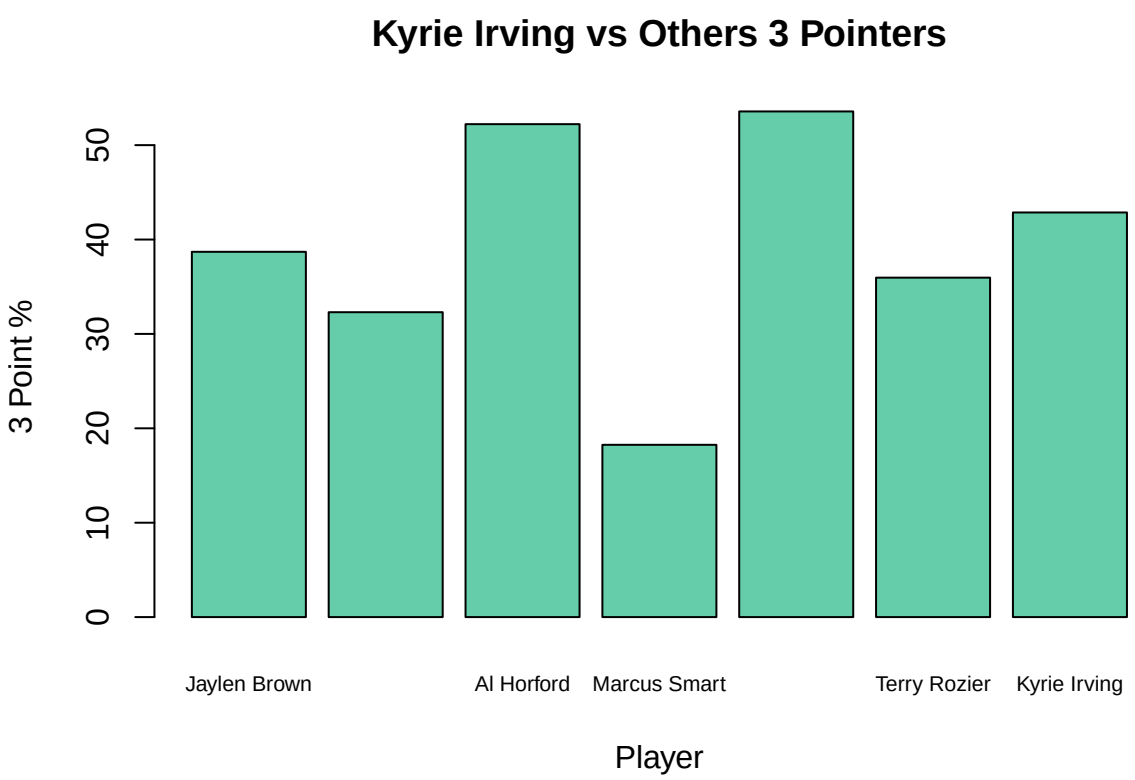
We can say that the players did not bring their best talent to game 6. Most of the players, besides Terry Rozier, played below average throughout the entirety of the game. If only one player is shooting well, he cannot hold up the entire team. Again, we only looked at the field goals and 3 pointers, which contribute mostly to the scoring of the game. Throughout the game, we can predict that the Celtics will end up losing the game, since most of the players were not shooting well compared to their averages.

Kyrie Irving

As we said before, Kyrie Irving is a strong key player to the Celtics, but unfortunately got a knee injury and had to miss the playoffs, including game 6. In this section, we will see if him missing the game had a big difference on the game outcome, and if Kyrie Irving would've made the Celtics win or still lose this game. We will compare his average to the averages of the other players on the team who we have previously looked at.



This barplot shows the averages of field goal percentages for every player, including Kyrie Irving, for previous games against the Bucks. We see that Al Horford had the highest percentage, and then it was Jaylen Brown or Kyrie Irving next. One can say that if we switched out Marcus Morris for Kyrie Irving, who replaced him for the remaining games, there would've been a better chance at winning game 6. Kyrie Irving's average field goal percentage is nearly 50%, which would've helped contribute to a better score for the game result. He doesn't have such a big significance though because the other players such as Jayson Tatum, Jaylen Brown, and Terry Rozier nearly match up to Kyrie's average.

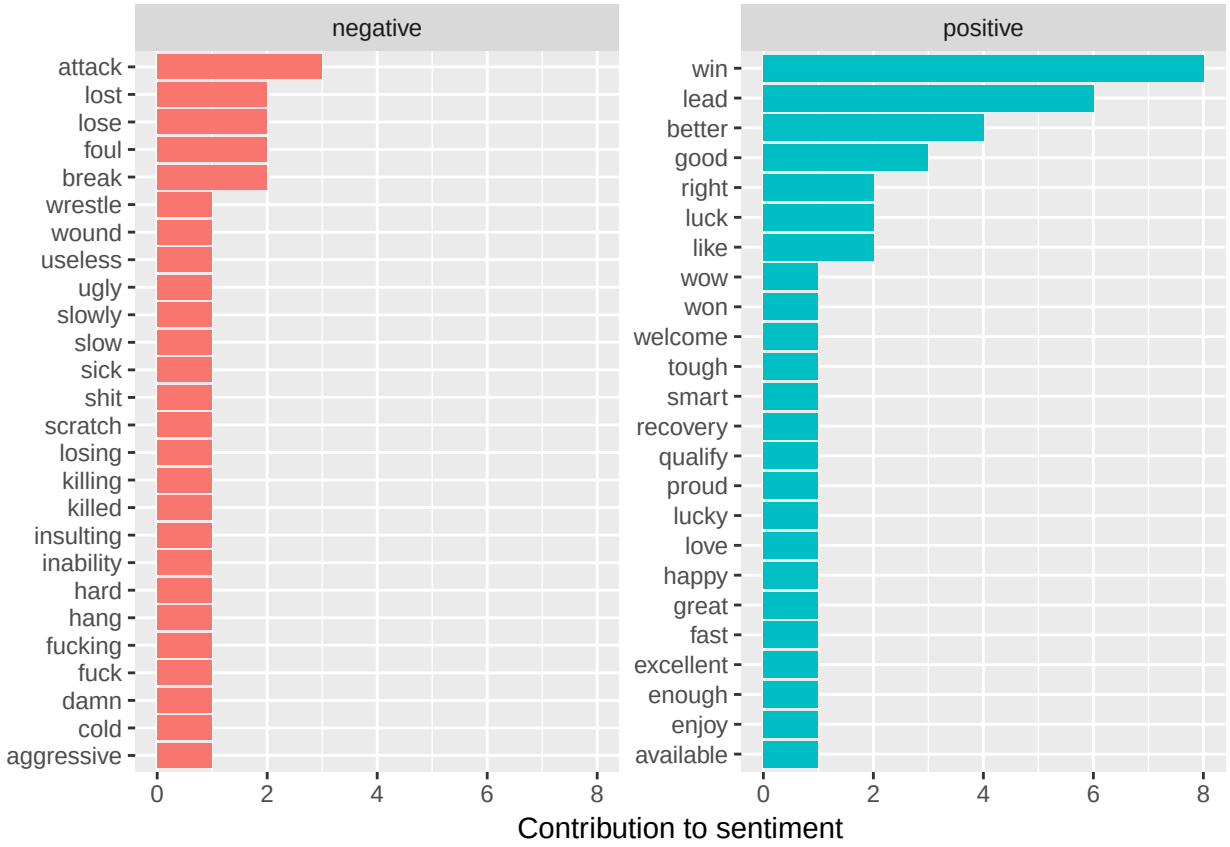


This barplot shows the averages of 3 pointer percentages for each player, including Kyrie Irving again. We see that the two highest 3 point shooters here are Al Horford and Marcus Morris, and then next would be Kyrie Irving. Kyrie does have a pretty high average 3 point percentage, so replacing him with someone else would be difficult because there aren't many players with such high 3 point percentages. This may be a significant factor in whether or not the Celtics could've won the game with or without Kyrie. He would've contributed more points than other players, althought Jaylen Brown, Jayson Tatum, and Terry Rozier nearly matched the average of Kyrie Irving.

Twitter

In this section, I use real time tweets during the game to see if there are any predictions or explanations of what is currently happening in the game. I use sentiment analysis (positive vs negative) to see what people are saying about the Celtics. I used the Twitter API to get tweets with search "celtics" or "#celtics", during halftime, end of 3rd quarter, and the final. I wanted to think that if there were more positive words than negative, it would mean the Celtics were doing well and others believe the Celtics may win.

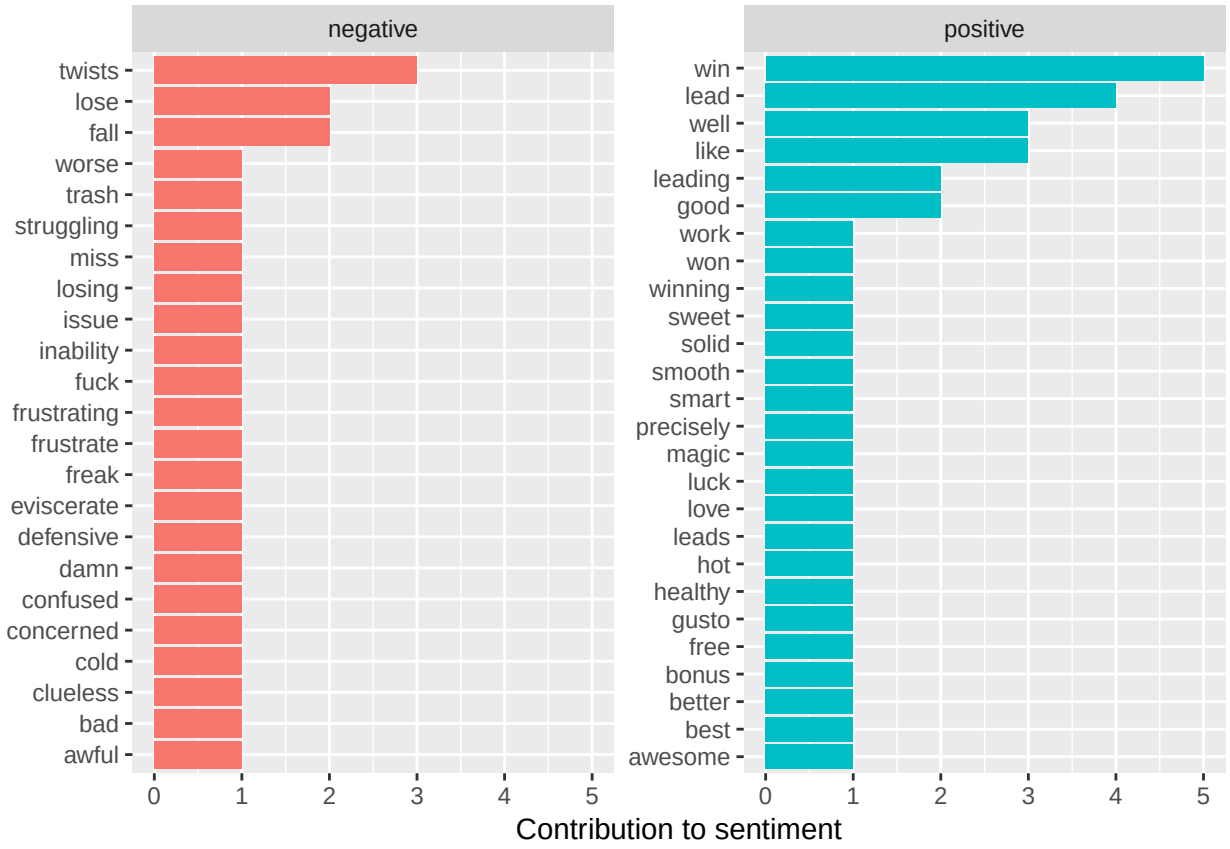
Halftime



During halftime, we can see that there were more positive words being used because the Celtics, although losing at this point, were close to having the lead. Many people believed the Celtics could've came back for the second half to win the game. The most commonly used word was "win" and "lead" on the positive side, with many people believing the Celtics would win the game. There were less negative words being used at this point, giving us the belief that the Celtics did have a chance to come back, which wouldn't have been hard to do.

Here, we see the used positive and negative words in tweets. We can see that the negative words were relative to the Celtics losing their lead against the Bucks. There are a few swear words that have been used in tweets, probably from those who believed the Celtics would hold their lead, and now afraid they may lose the game, especially since the Celtics would've closed out the first playoffs round with a win. For positive words, we see "win", "lead", "better", "proud", etc, which may relate to the Celtics having the lead before and being able to come back for the second half. At this point, it is hard to tell if the Celtics would win or not, because they had a good chance of coming back.

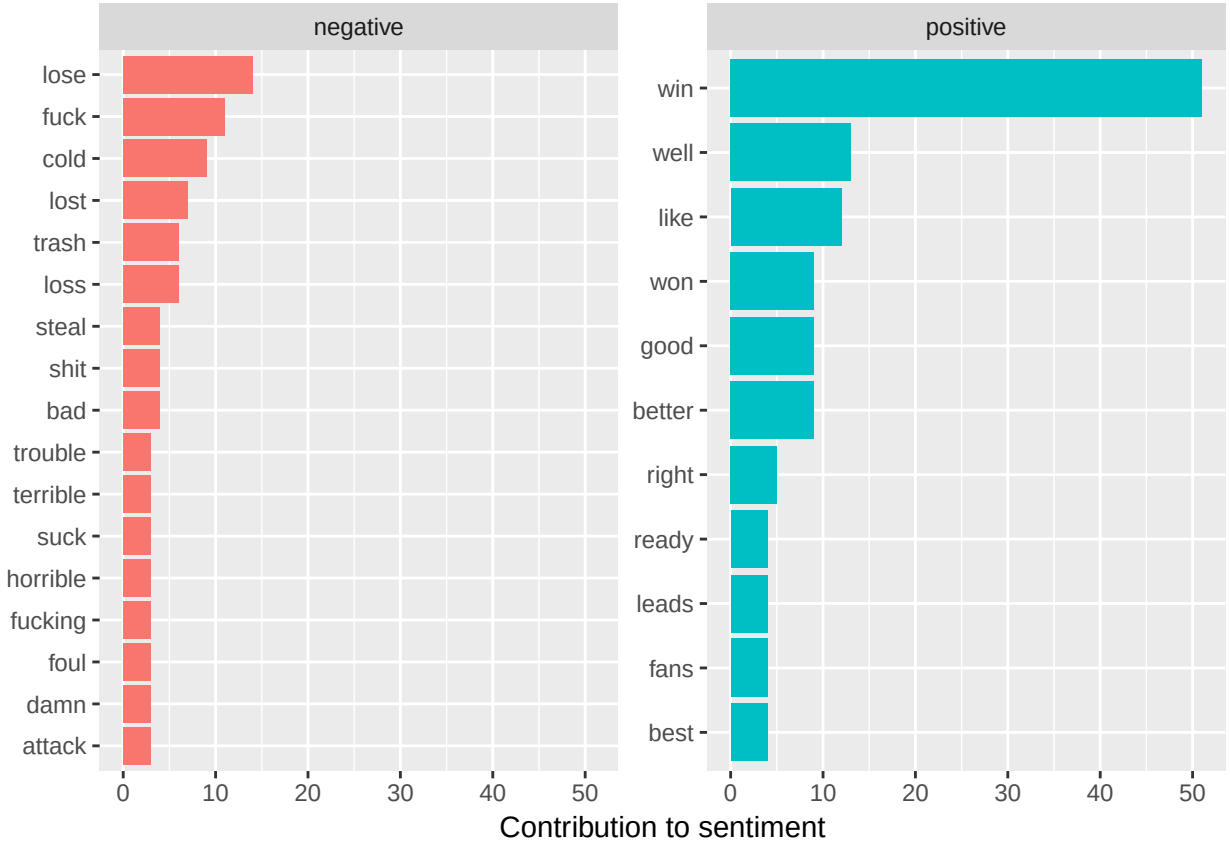
3rd Quarter



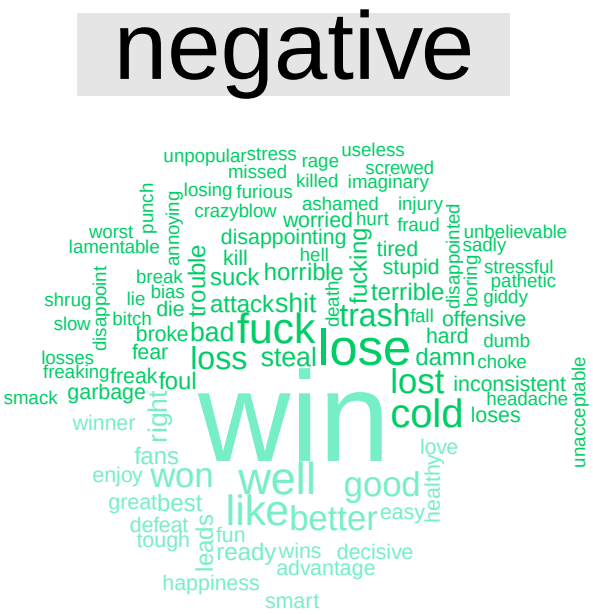
At this point, there were still more positive words that contributed to the sentiment, although the Celtics were still losing. At the same time, we cannot be sure if these words may have been used towards the Bucks, but assuming that we were searching up tweets relative to the Celtics, we cannot tell. There were more negative contributions this time around compared to the tweets during halftime, though.

In this word cloud, we see that there were more different negative words than during halftime. The most contributed negative word was “twists” which may be relative to the fact that the Celtics are losing more than before, especially after a lead. There are more negative words like “trash”, “miss”, “inability”, “frustrating”, etc, which shows how the game has changed since halftime. At halftime, there was still a chance for the Celtics to come back, but at this point, it was pretty clear that the Celtics have taken a deeper fall and would be hard to win in the last quarter.

Final



We can still see that there was more contribution of positive words, although the Celtics lost the game at this point, but there are less positive words than negative words. The negative words “lose” and “f*ck" contributed more, showing the loss of the Celtics. There are more contributed negative words to the sentiment.



In the word cloud, we see that there are significantly more negative words at this point than positive words. There are many more swear words now, since the Celtics lost. The positive words did not change from halftime. There are more negative words now like “disappoint”, “pathetic”, “inconsistent”, “ashamed”, etc, which shows that a lot of people had wanted the Celtics to win, but the loss was a disappointment to many, since they had a lead and could’ve closed out round 1 of the playoffs with this game.

Overall Sentiment Analysis of Tweets

We see that there is a big difference in sentiment from halftime to the final. During halftime, there were more positive words and contribution to sentiment, seeing that there was a good chance the Celtics would’ve come back from their shortage of points. Heading to the 3rd quarter tweets, we see that there were more negative words than positive since there was only one quarter left for the Celtics to possibly make a come back, though it was not likely. In the final, there were a lot more negative words since the Celtics ended up losing. We can see that the closer we get to the end of the game, the better the tweets explain and predict the outcome of the game.

Conclusion

It is easier to analyze the real time data when we are heading closer towards the end of the game. This is obvious, because towards the end, there is a clearer idea of how many points one time is losing to the other team, and how much time is left in the game. Throughout the game, I was able to predict that in the 1st quarter, the Celtics had a good chance of winning game 6 because we were putting up a lot of points and were in a good lead. In the 2nd quarter, although there were less points, there still could've been a chance to bring up the score. In the 3rd and 4th quarter, it was clear that the game would not end well. The Celtics were playing an away game, not playing above average for shooting, turnovers, and rebounds, and the players themselves were not playing well. The lineup could've been better with Kyrie Irving, although one player cannot hold the whole team up, like how Terry Rozier could not give the team a win while his other teammates were playing below average. The tweets gave a clear idea of how the Celtics were playing in the 2nd half of the game. Being able to analyze every 2 minutes of the game and comparing averages and tweets, we see that it was clear the Celtics would lose game 6.