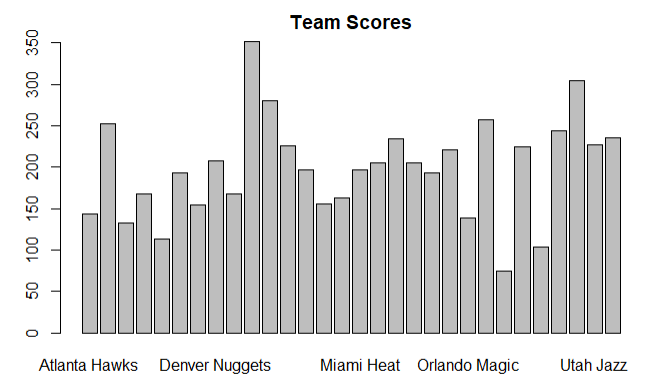
1. To begin the project, I found team statistics data for the 2017-2018 season from basketball reference and imported it into R. In order to automate this process in the future, a spider program could be set up in order to do web scraping to write into csv which could then be used for data analytics in R. To calculate a teams probability of winning, I calculated a general performance metric for each team which is called “Score.” The model takes the home team’s score and divides it by the total score of both teams to calculate the likelihood of winning. Then, to factor in home court advantage, I add (1-likelihood of winning)\*0.2 to the probability of the home team winning. For example, if the probability of winning is 50-50, then factoring in home court advantage my model predicts that the home team has a 60% chance of winning.



The visualization above shows a bar chart of the aggregated score using team statistics for each team. This score metric is calculated by weighting certain vital team statistics and summing these results. In order to maintain some degree of normalization, each team statistic was subtracted by the team’s average opponents’ team statistic (as in rebounds given up to the opponent or opponents’ 3pt percentage). Additionally, since there are negatives, the absolute value of the minimum score is added to each team’s score along with an arbitrary constant so that no values are zero.

1. The main challenges that this model faces are that it is not flexible. It does not account for player trends or status so injuries and trades will drastically reduce the accuracy of this model.
2. Based on this model, both the underdogs and favorites should try to improve the statistics with the highest variance as well as those that have been weighted more heavily. In the future, weights should be adjusted and tuned better to reflect correlation, however this was difficult without using test and train data. The two categories that benefitted and hurt teams the most were assists and turnovers so improving these statistics would result in higher win probability.