A Database of NBA Refereeing

**Abstract:** We build a database that collects the information released by the NBA in its *Last Two Minute* reportsand allows for player- and team-based queries.

Referees are supposed to be impartial, bringing sound judgment amid the chaotic environment of a sporting event. However, a substantial body of literature[[1]](#footnote-1) [[2]](#footnote-2) [[3]](#footnote-3) shows that the impartiality of officials is often questionable. In an effort to increase transparency on these issues, the National Basketball Association, in March 2015, started releasing its *Last Two Minute* reports*.*[[4]](#footnote-4) These documents include an assessment of all executed calls and notable non-calls for the last two minutes of any close game[[5]](#footnote-5) during the regular season and the playoffs.

While the NBA’s effort to increase transparency is commendable, seeing as it is one of very few sports league willing to publicly acknowledge officiating mistakes, the format in which the reports are published does not lend itself to statistical analysis. The reports are published as PDF files on a game-by-game basis, making their systematic use for statistical research a daunting task. It is not surprising to see, therefore, that as of today only one paper[[6]](#footnote-6) has used them as a data source.[[7]](#footnote-7) Furthermore, another issue with the data is that the Association does not release the name of the officials in charge of each individual game, making it even more burdensome to connect each call to the trio of referees that executed it, or to evaluate the performance of individual officials over multiple games.

Therefore, our project sets forth four main goals:

1. To systematically collect the reports available online and collect them into a database that can be searched by player and team.
2. To link game entries in this database to the referees that were officiating the game, so that the database can also be searched by individual referee.[[8]](#footnote-8)
3. To build a website that allows user to navigate our database .
4. Provided that the first three goals are achieved, to perform some statistical analysis on the new dataset, particularly to verify with the new, larger dataset that the conclusions of Deutscher’s paper hold.

Concerning the feasibility of the project, we have verified that the PDF documents can be converted into a manageable txt file, so that we can then automate their parsing and the construction of the database. While this is promising, the structure of the txt files that we managed to retrieve is still fairly chaotic, so their handling should still pose a significant challenge. The HTML from Basketball-Refernce can easily be handled with Beautiful Soup, so it should not be an issue. We believe that the construction of the database will also represent a significant hurdle, since during CMSC 12100 we spent relatively little time on SQL and focused mostly on how to search (rather than how to build) a database.

1. Boyko, Ryan H., Adam R. Boyko, and Mark G. Boyko. "Referee bias contributes to home advantage in English Premiership football." *Journal of sports sciences* 25, no. 11 (2007): 1185-1194. [↑](#footnote-ref-1)
2. Johnston, Ron. "On referee bias, crowd size, and home advantage in the English soccer Premiership." *Journal of Sports Sciences* 26, no. 6 (2008): 563-568. [↑](#footnote-ref-2)
3. Rodenberg, Ryan M., and Choong Hoon Lim. "Payback calls: A starting point for measuring basketball referee bias and impact on team performance." *European Sport Management Quarterly* 9, no. 4 (2009): 375-387. [↑](#footnote-ref-3)
4. NBA. “NBA Officiating Last Two Minute Report Archive”. <http://official.nba.com/nba-last-two-minute-reports-archive/> [↑](#footnote-ref-4)
5. Close games are defined by the NBA as games “which were within five points at the two-minute mark (and during overtime, where applicable)” where the two-minute mark is the 4th quarter’s. [↑](#footnote-ref-5)
6. Deutscher, Christian. "No referee bias in the NBA: New evidence with leagues’ assessment data." *Journal of Sports Analytics* 1, no. 2 (2015): 91-96. [↑](#footnote-ref-6)
7. In addition, this paper only worked on a subsection of the data that is currently available, since it was published only one year after the beginning of the program in 2015. [↑](#footnote-ref-7)
8. The referee data can be easily accessed from Basketball-Reference, a recognized source of NBA statistics. Each individual game box score is published as an HTML file on the website, and reports the officials’ names at the end of the page. The following link is an example from this weekend’s SAS@CLE game. <http://www.basketball-reference.com/boxscores/201701210CLE.html> [↑](#footnote-ref-8)