

STAT 345 Midterm Project

Due April 9

“Our offense is like the Pythagorean Theorem. There is no answer!”, *Shaquille O’Neal*

The Background

Your role for the midterm project is that of data analyst intern at an NBA (professional basketball) team. Your direct supervisor (also part of the analytics team) has asked you to create a data visualization to illustrate how (or if) the team’s shots have changed over time. After some initial clarifying questions, your supervisor confessed that they had seen some pretty cool shot charts at <http://savvastjortjoglou.com/nba-shot-sharts.html> and would like to extend the ideas a bit.

Your data for the midterm project may come from a variety of sources, including the NBA directly, as well as Basketball-Reference, HoopsHype, and others. There are several ways to access the data, but perhaps the simplest is through the `nbastatR` package, which serves as a wrapper (or set of helper functions) for these various data sources. There is a lot of potentially useful information that can be found using this package; unfortunately, the documentation is somewhat inconsistent. For example, the documentation at <https://www.rdocumentation.org/packages/nbastatR/versions/0.1.110202031> is pretty well-formatted and easy to navigate, but the current function names have changed a bit (see `?nbastatR` and click “Index”).

The Tasks

1. Produce a graphic displaying the shot locations for a particular team over several years. Some notes:
 - Colors should be chosen to reflect the team, if possible.
 - There are likely many overlaid points – handle this by either binning these by location, or use opacity.
 - Incorporate information about whether the shot was made or not (shape, color, etc.).
 - The graphic should be well-labeled, titled, etc.
 - Start with a graph for a single year, then extend to several years. Up to 20 years of shot data is available. Either facet these by year or animate using the years.
 - You’ll want to figure out what the coordinates mean somehow. This might be through the documentation, but could also be determined using aspects of the data itself and the dimensions of an NBA court.
 - If possible, try to put a basketball court on the background of the image (you’ll need to scale it appropriately).
2. Summarize the graphic/series of graphics into a digestible, bullet-point brief report for front-office staff. Some notes:
 - The main body of the report should be very brief – just the graphic(s) and the bullet-pointed list of findings, which should be short and clear.
 - Include a more detailed explanation of these bullet points, for further reading by those interested. This section should follow the bullet-point section, but should be organized similarly for reference.
 - Your report to the front-office shouldn’t include any code.

- This report should be generated using RMarkdown. However, the choice of output type (Word, PDF, or HTML) is up to you (you could even make slides if you want to).
3. Write and document clean, efficient, reproducible code. Some notes:
- This code will be viewed by your direct supervisor.
 - The code file should include your code to gather, join, and clean the data; the code to generate the graphic(s) presented; and your commentary on the results (so, a single .rmd file, or an .rmd file that sources an .r file).
 - Your code should be clean, organized, and reproducible. Remove unnecessary/scratch/exploratory code.
 - Your code should be well commented. In particular, any decisions or judgement calls made in the analysis process should be explained/justified. Sections of code should be identified even if not functionalized (including purpose, data/argument inputs, analysis outputs).
4. Above and Beyond. Explore the data a bit, and create a graphic that uses (or incorporates) different information than what was used above. Some notes:
- Create an additional graphic that incorporates at least one additional variable not previously used. The additional data should be drawn from a different dataset (function call) than the original graphic used. These two (or more) datasets may need to be joined appropriately.
 - You can either add more information to the plot above, or create a different plot.
 - Formatting, labelling, etc. are all important here too.
 - Adding marginal densities or other “bells and whistles” might offer additional insight.
 - This graphic should be included at the end of the report (after the more detailed explanations).
 - You should include a brief description of the graphic (highlighting the different/additional information used).