NBA's Hot Hand Theory

In basketball, many casual viewers have seen when a star player enters a 'zone', or seemingly is able to make any shot. To us, players that make a couple of shots get more comfortable and confident and are able to continue to make shots that are more and more difficult. There has been a long-standing debate about whether or not this 'hot hand theory' is a fallacy or truth. This project will answer the question: Does knowing the outcome of previous shots that a player has taken help us determine whether or not a certain shot will go in?

Milestones:

1. Project Selection: Form teams of 2 or 3 and select a project from the provided list.

2. Literature Study: Go through the following resources for background on the project and write a half to 1 page summary for each one:

- Gilovich, Thomas, 1985. "The Hot Hand in Basketball: On the Misperception of Random Sequences", Cognitive Psychology

  URL: http://www.bbn-school.org/us/math/ap_stats/investigations_folder/apsi_mmsi_bentley_2010/best%20practices/Hot%20Hand%20in%20Hoops%20Study.pdf

- Andrew Bocskocsky, John Ezekowitz, and Carolyn Stein, 2014. "The Hot Hand: A New Approach to an Old "Fallacy", MIT Sloan Sports Analytics Conference

  URL: http://www.sloansportsconference.com/wp-content/uploads/2014/02/2014_SSAC_The-Hot-Hand-A-New-Approach.pdf

3. Data Exploration and Cleaning: The primary source of data for this project is a publicly available dataset from Kaggle of the 2013-2014 season:

- https://www.kaggle.com/dansbecker/nba-shot-logs

  Perform the following exploration steps:

- Determine the players with a minimum number of shots taken per game to be included in the analysis.

- Exclude players not determined to be part of the analysis.

- Visualize the shot distribution total, and for the top 5 shooters (determine simple metric for determining top shooters for now such as field goal %)

- Wrangle the data to calculate the last "streak" (how many shots did they *make* or *miss* in a row in that game, start over at 0 if direction is switched) per shot taken
- Remove any non-valid data rows

4. Implement Basics:

Goal setting: Decide on what results you would need to see to be convinced of the hot hand theory.

- *Feature extraction*: Extract a set of basic features from dataset that would be relevant to the shot being made.
- Implement the following baseline techniques:
  - *Logistic regression*: Train a simple logistic regression model on the 'streak' metric created during exploratory analysis and whether or not shot is made.
  - *Confounding variables:* One by one, add in what you believe could be confounding variables and see if the 'streak' metric turns/still is significant.

5. Proposal:

Other streak metrics: Try other streak metrics from intuition about what hot hand theory believes.

Scraping: Scrape other features that could be relevant confounding or predictive features to the model.

- *Additional tasks*: Train the best possible logistic regression and determine what are the most important features for a player making the shot.

You make the call: Hot hand theory - fact or fallacy?