Measuring The Hot Hand Effect Using NBA Shooting Data

CS1951-A Capstone Project

Group Members: Alex Wainger (awainger), XiaoYi Mao (xmao), and Siyi Zhang (sz22)

In sports — and particularly in basketball — the hot hand effect is the belief that past success will lead to an increased chance at success in the future. In other words, a player who has made a few shots in a row is considered to be "hot" and is thus more likely to make the shots that follow than if he were not hot. While NBA coaches and players seem to generally believe in the hot hand, most early studies determined that the hot hand was just a fallacy, and that future shots were independent of previous hots. However, more recent work has shown that there might be some truth to the effect. So we constructed a web application that allows us to explore the hot hand effect and determine whether or not it is just a fallacy.

We scraped basketball-reference.com and obtained data on over three million shots from the last 16 NBA seasons. Our web app allows users to filter our dataset to explore whichever subset they are interested in, as well as define the hot hand however they like. We then provide a number of visualizations (shown below) that give the user some insight into whether or not the hot hand exists for that subset of the data and hot hand definition.

We used our web app to experiment with different hot hand definitions and came to a very interesting conclusion. With the hot hand defined as making two, three, or four shots in a row, there is a very strong effect on players' shooting percentages. Represented in the table below, having a "hot hand" can actually decrease a player's field goal percentage by up to two percentage points, which runs completely counter to conventional wisdom.

P-Values		Minutes between each shot							
		1	2	3	4	5	6	7	8
Shots in a row	1	.6012	.2576	.4119	.4703	.5472	.6508	.5836	.5460
	2	.0245	.0005	.0013	.0087	.0193	.0195	.0192	.0114
	3	N/A	.0257	.0001	.0004	.0003	.0005	.0008	.0012
	4	N/A	N/A	.0159	.0516	.0067	.0065	.0007	.0012

Avg Diff		Minutes between each shot							
		1	2	3	4	5	6	7	8
Shots in a row	1	-.0014	-.0027	-.0019	-.0017	-.0014	-.0010	-.0013	-.0014
	2	-.0109	-.0112	-.0093	-.0073	-.0063	-.0062	-.0062	-.0066
	3	N/A	-.0130	-.0154	-.0136	-.0130	-.0124	-.0116	-.0109
	4	N/A	N/A	-.0189	-.0116	-.0152	-.0150	-.0174	-.0170

Figure 1: The results of our permutation test with varying definitions of the hot hand. Significant p-values in the first table are marked by increasing shades of red, with the darkest red representing the most significant results. The second table shows the original difference observed between the average hot and regular shooting percentages, which are uniformly negative and marked with the same coloring scheme.

Screenshots From the Web App

Figure 2: The form users submit to define the subset of data they are interested in as well as how they would like to define the hot hand.

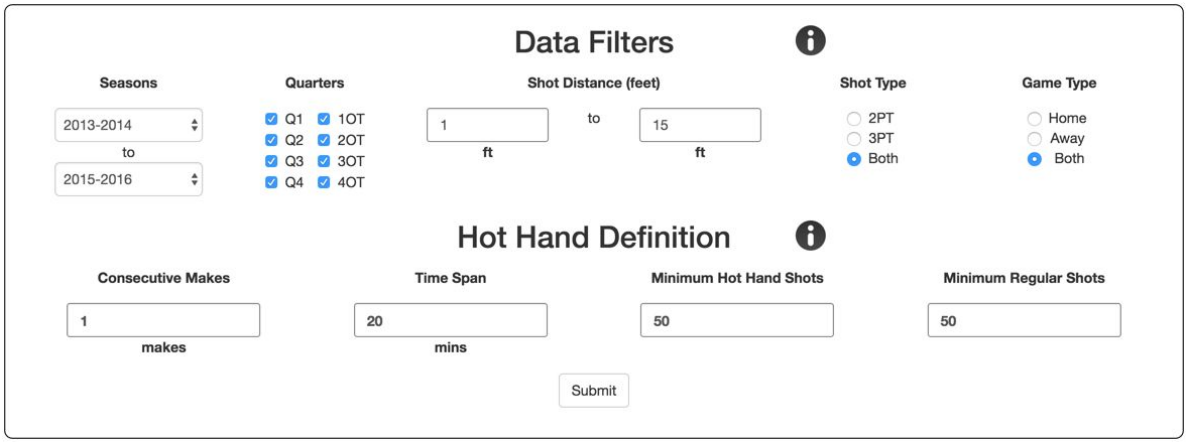

Figure 3: Each player that has taken enough shot and regular shots based on the filters and minimums defined in the form above will be plotted on the scatter plot below, which shows the relationship between hot and regular field goal percentage. The dotted line represents equivalent hot and regular percentages — any point above the line is a better hot shooter, while any point below the line is a better regular shooter. The points can be colored based on different attributes such as position (shown below), height, weight, or average shot distance.

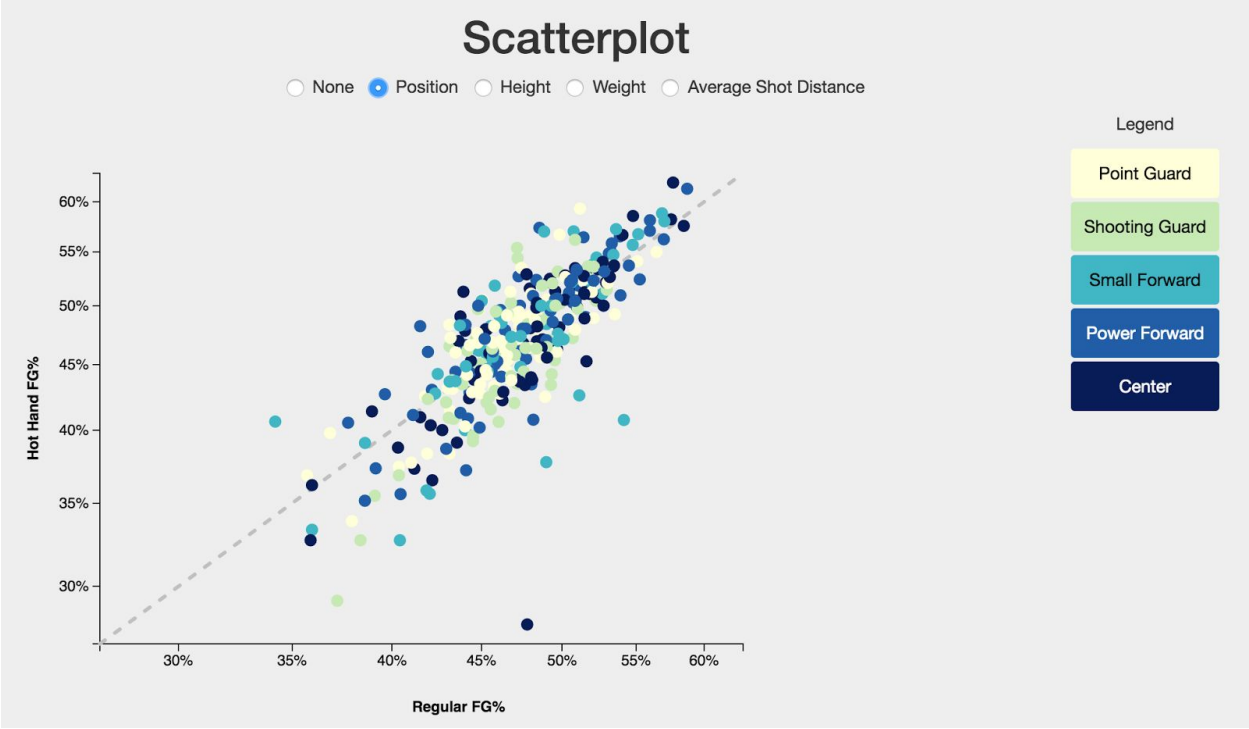

Figure 4: Clicking on a point on the scatter plot will open the player information popup, which gives a more detailed breakdown of that particular player's stats. Here, we see the popup for Paul George. The two line graphs illustrate the frequency that George takes shots from each distance (left) and his success rate from each distance (right). These two graphs serve to identify some reasons why a player would be a better hot or regular shooter. In George's case, he takes more long-distance twos and deep threes when he is hot (which are not high percentage shots), and gets to the basket at a much lower rate, even though he is extremely effective there.

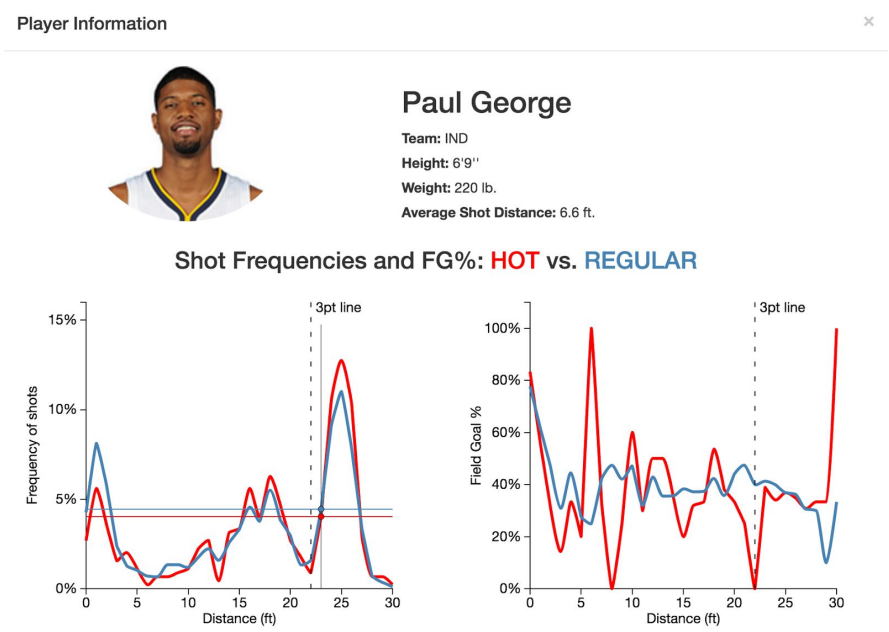

Figure 5: The histogram summarizes the data presented in the scatter plot. Each player is binned based on the difference between their hot and regular shooting percentages. The bars of the histogram can be colored in the same way that the scatter plot is, to show the demographics of each group. In the plot below, we can see that there are more points with a negative difference than a positive, indicating there may be a negative effect with this hot hand definition.

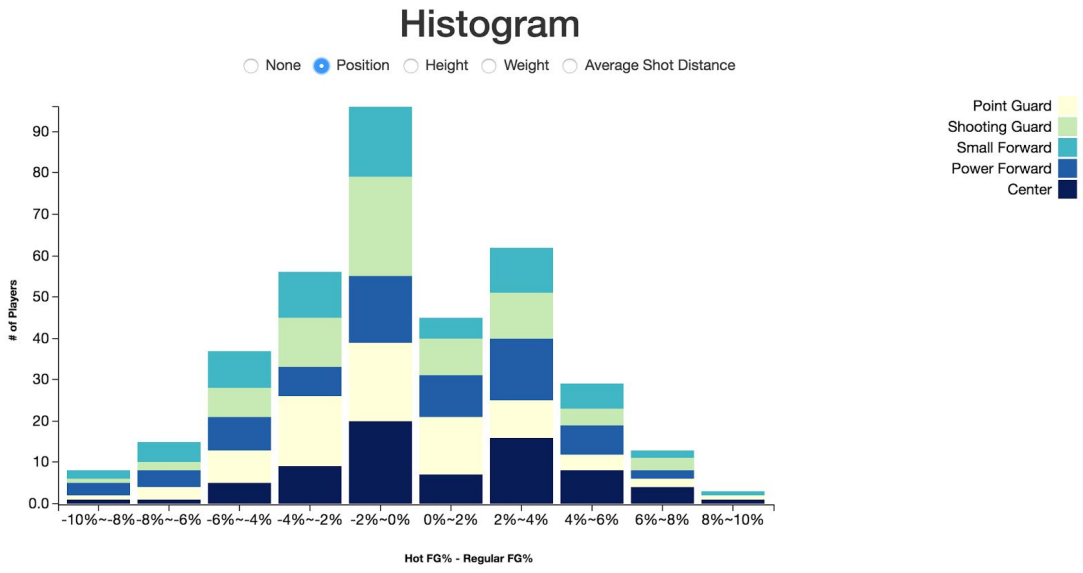

Figure 6: Our last visualization shows the result of running the permutation test with 100,000 iterations over the percentages calculated from the specified data and hot hand definition. The histogram below bins the difference between the two reshuffled samples, with blue bars indicating differences that are less extreme than the originally observed difference, and red bars showing more extreme differences. The original difference between the hot and regular percentages, as well as the p-value are displayed so users can definitively know whether this hot hand definition produced a statistically significant difference in shooting percentages, and how large that difference was.