The MVP of MVPs (2013-2018)
Comparing the Statistics of NBA Most Valuable Players in the Steph Curry Era

by Isaac Chau and Julian Pelzner

http://people.ischool.berkeley.edu/~jbpelzner/NBA-MVP-comparison/

Data Collection:

Early into the semester, we created an iPython Jupyter Notebook in which we derived two functions to scrape statistics from Basketball-Reference.com. The code for these functions implemented the BeautifulSoup package adapted from an article written by Oscar Sanchez [1]. We used this code to scrape 'Per Game' stats, as well as 'Advanced' stats, for the seasons 2012-2013 through 2017-2018 that were provided by Basketball Reference on their website.

After some iteration on our project theme, we decided to also utilize a pre-compiled dataset of some more complex statistics that was arranged by Paul Rosotti and provided on Kaggle [2]. Most of these statistics were already included in the Basketball Reference dataset, but we were able to see some definitions of more advanced statistics to refer to and to create calculated fields for some exploratory analysis.

Lastly, after another set of iterations through our visualization process, we decided to hand pick some more stats from the NBA.com advanced player stats page and compile it into a separate data file for our analysis[3].

Data Analysis and Visualization:

For the flow of our project, Isaac handled the first level of analysis regarding identifying which of the MVPs had the greatest impact on their respective team, and thus which player could be the best MVP. To follow, Julian explored some of the statistics which that player excelled in compared to the other MVPs, and dissected those statistics to understand the component statistics which contributed to that player's success.

So, Isaac started by creating a Radial Chart which overlays all 6 MVPs along 6 different statistical parameters that are used in the judgment of each MVP. The color scheme used for this chart was monochrome, which allows for interpretation by users with colorblindness.

Their next viz was a set of adjacent plots that compares each MVP's Offensive Rating and Defensive Rating to that of their respective team's, including a trend line for comparison. They paid attention to the preattentive features regarding the sign of the trends, and flipped the y-axis on the Defensive Rating graph so that the plots would describe better statistical values in a consistent direction.

Then, the next viz adds Net Rating into the view, plotting adjacent bar charts of the residual values of the MVP above their team for Offensive, Defensive, and Net Ratings. Through analysis of this last viz, the user could identify that Stephen Curry performed much better in his 2015-2016 MVP season that any of the other MVPs included in the comparison.

With this insight, Julian returned back to the Radial Chart, highlighting Stephen Curry's values for the 2015-2016 season, drawing user attention to the statistics in which Stephen Curry had higher values than the other MVPs, namely Box Plus/Minus (BPM) and True Shooting

Percentage. The investigative visualizations that followthere are two adjacent plots, with the plot on the left showing all players who played more than 40 games in a given season, and the plot on the right showing all the MVPs.

Focusing on BPM first, he created an adjacent set of bar plots that orders players in ascending order by their BPM values. This viz allows for the users to select between BPM, Offensive BPM, and Defensive BPM in the view through the use of a generated Calculated Field and Parameter filter for these different components. Through their position on the x-axis the user can identify the rank of the players in the sample, and the height of the bar (aided by a tooltip) lets the user identify the key value of interest and compare values across players. Additionally, the MVPs on both plots are color-coded.

For the next viz, attention was turned to True Shooting Percentage, where the adjacent plots similarly represent the two different samples of players. In this viz, however, the plots are scatterplots showing how a players True Shooting Percentage correlates with one of its component statistics. This user selection of either 3-Point Field Goal Percentage or Free Throw Percentage as the field to be plotted on the x-axis was also enabled by the creation of a Calculated Field and Parameter filter for these different components. Filtering is allowed along other relevant statistics such as the volume of 3-Point Attempts and Free Throws per game, as well as productivity in terms in Points Per Game. Position on the x and y axes lets the user identify the respective values of interest. Additionally, both plots represent the points of the MVPs through images of their faces

Lastly, the style of the True Shooting Percentage scatterplots are replicated for Effective Field Goal Percentage, except 3-Point Field Goal Percentage is the only value that is plotted on the x-axis and there are fewer filters. Position on the x and y axes lets the user identify the respective values of interest, and the faces of MVPs signify their position in this viz as well.

As we move through these visualizations, we are able to confidently identify that Stephen Curry's 2015-2016 year proved to be statistically superior to that of the other MVPs. Using the tools provided by Tableau, we were able to create meaningful visualizations using limited coding, incorporating several types of visual channels to communicate our data. In the future, we would like to experiment with more D3 code to implement more transitions and interactive features in our respective visualizations, and in our website as a whole.

References:

[1] https://towardsdatascience.com/web-scraping-nba-stats-4b4f8c525994
[2] https://www.kaggle.com/pablote/nba-enhanced-stats
[3] https://stats.nba.com/players/advanced/?sort=GP&dir=-1