Master of MVPs

An interactive data-driven story by
Chris Traver, Chris Vrabel, Rohit Biswas

Introduction

For this project, we were interested in looking at trends in most valuable players (MVPs) across the NHL, NBA, MLB, and NFL. In order to keep the datasets manageable, we decided to narrow down the data focus to only MVPs during years between 2000 to 2015. In doing so, we were able to utilize the data to develop visualizations that give us insights into common characteristics and distinguishing factors of MVPs throughout the NHL, NBA, MLB, and NFL, but we were also able to explore a variety of other possible trends such as how salaries compare between sports and by age.

Data Description

We utilized a combination of datasets for this project, all extracted from various subdirectories of reference.com websites for each sport. The specifics of each website for the different sports are detailed below:

NFL: www.pro-football.reference.com
MLB: www.baseball.reference.com
NHL: www.hockey.reference.com
NBA: www.basketball.reference.com

We began by compiling a master list of all of the MVPs across the four sports. We scraped their demographic data, including their Name, Sport, Team, Position, Age at MVP, Height, Weight, College, Salary, Hometown, and Latitude and Longitude for their hometown to plot points on a map. Using this data we were able to develop the interactive map and populate the Biography tooltips that appear when you hover over data points throughout the page.

Next, we collected headshot images from reference.com for each player to include in the Biography tooltips. All of these images were saved locally in our project directory and linked using relative paths in our code to ensure that links would not be broken if our project is moved to another directory later on.

Finally, we collected individual statistics for each MVP depending on their sport. We saved each of these into separate CSV files that are loaded on demand using an onclick event in our

visualizations. We did for two reasons: to prevent delayed loading issues that may be caused by loading all files at once during the initial page load, as well as for clarity for our sake when sorting through the data.

Data source and accuracy. We were able to download all of our datasets directly from the reference sites using their CSV export functionality. Reference.com is known to be a reputable repository of sports statistics for multiple sports, therefore we beleive that our data is extremely reliable.

Variables that we used. Our datasets were extremely detailed and contained many different variables pertinent to each MVP depending on their specific sport. As mentioned above, we collected demographic data on each MVP including their Name, Sport, Team, Position, Age at MVP, Height, Weight, College, Salary, Hometown, and Latitude and Longitude of their hometown. The variables varied for each specific sport, but for MLB players for example we used variables containing information on a player's on base percentage (OBP), hitting average, and several others.

Dataset Integration

As mentioned above, for this project we integrated multiple datasets including a master list of MVPs across all sports, as well as individual datasets containing statistics for each player. We integrated these datasets together using a common field that each dataset shared: Player Name. By using the player's name, we were able to link together all of our datasets and load statistics about a player on demand instead of all at once on the initial page load.

Additional Data

In order to develop our visualization for the interactive map, we relied on a JSON shape file of coordinates for the world in order to project our map in the visualization. This JSON file was adapted from Mike Bostock at https://bl.ocks.org. We opted to use the 110m shape file instead of the 50m shape file because the latter was too high resolution and caused our interactive map to lag significantly when trying to zoom in and pan around the map. We also used JQuery for some interactive elements, Bootstrap for a responsive layout, Google Fonts for their Roboto Font, and D3 for the interactive elements.

Mapping Data to Visual Elements

Description

Our project contains a number of integrated visualizations. However, each has different corresponding mappings of data to visualization elements. We outline the details of each below:

Interactive World Map. The interactive map plots the hometowns of each MVP. Each MVP is represented by a single point on the map, even if they have been awarded MVP multiple times between 2000 to 2015. We utilized descriptive icons of a basketball, baseball, football, and hockey skate to signify which sport each MVP plays. If you hover over a data point, a tooltip will appear that displays biographical information about that MVP.

Player Specific Graphs. When you click a player's hometown on the world map, two player specific graphs will appear below. On the left, is a bar graph comparing the player's stats during his MVP season, to the league averages. It shows how much better that player performed in comparison to the average player. On the right, is a graphic showing the awards that the player has won over his career. There are four awards displayed for each sport, all being highly respected awards. By displaying this graphic, the viewer can get a sense of how that player performed over his career, and where he may stand in comparison to the other MVPs.

Age vs. Salary Scatterplot. This scatterplot shows the correlation between the age of each MVP at the time they were awarded MVP, compared to their salary that current year. The different colored dots distinguish between the different sports of each MVP, and are detailed in the legend on the top right corner of the graph. We chose this color scale to be a categorical one since the leagues are analogous to categories to place the players in. We utilized linear scales for both the x and y axes for this graph, with the x-axis representing a range of ages and the y-axis representing a range of salaries.

Interactive Histogram. This histogram shows the frequencies of ages, heights, weights, and salaries across all sports. Since the graph displays multiple datasets in one visualization, the scale for each x-axis differs in values, while still remaining linear in each graph, while the y-axis remains the frequency for each graph. The coloring was kept consistent with the scatterplot to create the same association with league and the respective color.

The Story

The purpose of our integrated visualizations in this project was to elucidate trends across MVPs from several different sports.

The map shows the hometowns of all of the MVPs across the different sports. It is interesting to see that there are several clusterings of hometowns based on specific regions and sports. For example, we noticed that most basketball MVPs are originally from the east coast, with several MVPs sharing the same hometown. We also noticed that most of the hockey MVPs were from Canadian and Russian hometowns or northern regions of the United States where hockey is more popular. Additionally you notice that most players (not including hockey) come from the coastal and warmer areas of the country. This is likely due to the warmer weather allowing for them to play more often.

When a user clicks on a specific player's icon on the map. Their detailed statistics are pulled up in the panel below the map. We noticed that the MVPs vastly outplay the average players. There are cases where the MVP stat is lower than the average, but these are all cases where the statistic is not a part of their play style. For example, guys like Shaquille O'Neal, or Tim Duncan, are below average for the number of threes made per game, because those are shots that they do not take. With regards to their career awards, we first notice that a good number of the players are multiple time winners. This is something obvious, as we would expect that the top players play at a high level for several years. Also, most of the MVPs have won a championship at some point during their careers. This shows that to win a championship, a team typically needs to have a superstar. In baseball, and somewhat in hockey, a player has less of an impact than a single player important player in football and basketball. This correlates with less of the MVPs winning championships in baseball and hockey. Those players may be playing at a high level, but if their team is poor, they cannot single-handedly carry them to the championship.

We also noticed several interesting trends about MVP ages and salaries through our interactive scatterplot. Specifically, we were surprised to notice that NFL MVPs tended to be older than MVPs from all of the other sports. We were also surprised to see that the salaries of all of the NHL MVPs were much lower than all of the other sports.

Interesting Conclusions:

- MVPs are clustered in parts of the USA (East Coast for all, Southwest for MLB).
- MVPs tend to specialize in scoring points rather than other stats, as not all of their stats are above the league average for that season.
- MVPs in the NHL are not paid as much as MVPs in other leagues
- MVPs in the NHL and NBA are much younger than those in the MLB and NFL
- MVPs in the NHL never went to college

- MVPs’ heights have a relatively normal distribution across all leagues, with a mean of about 74 inches
- DONE woo.