

# ASSIGNMENT 3: INTERACTIVE VISUALIZATION

Ming-Yi Peng

CS-GY 6313 B

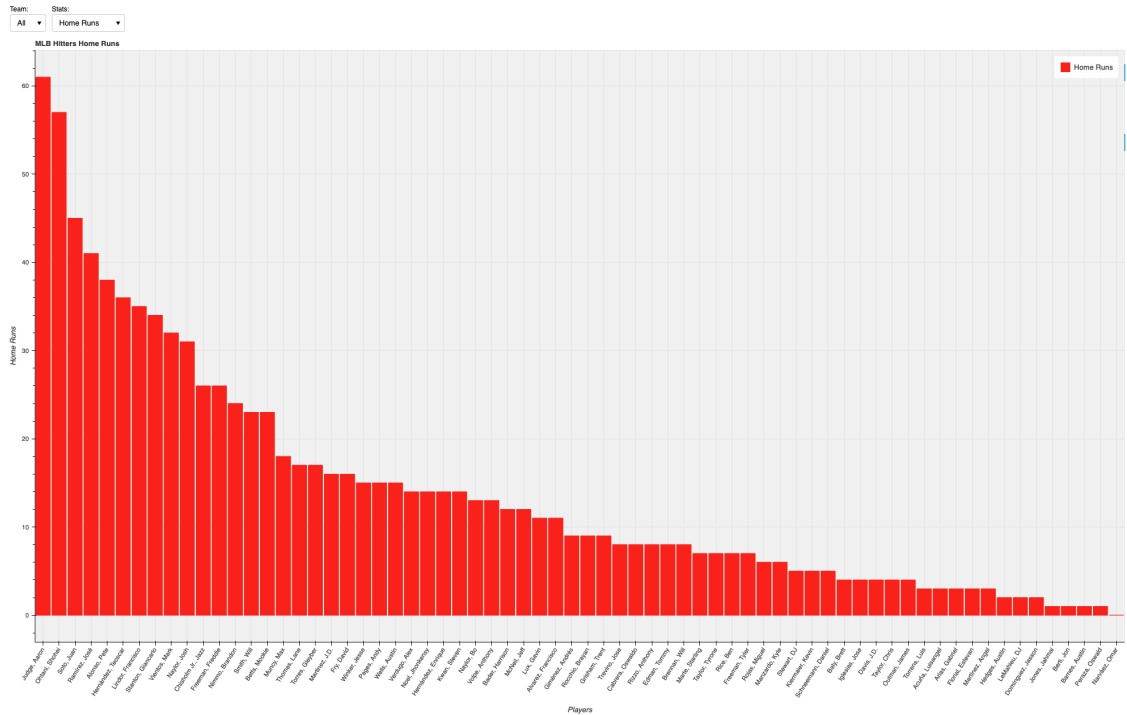


Figure 1: A snippet of the plot from the interactive visualization

## 1 Question

Major League Baseball (MLB) 2024 World Series concluded on October 30, 2024, wrapping up an exciting season where every team performed remarkably. This year was particularly thrilling for New York fans, as the Yankees and Mets advanced to the American League and National League Championship Series, respectively. The World Series ended with the Yankees going head-to-head against the Los Angeles Dodgers, creating a buzz around baseball in New York. As a huge baseball fan, I found it fascinating to explore player abilities through data visualization, which motivated me to use data from the 2024 MLB season as the foundation for this project.

**Main Question:** Among the four teams in the ALCS and NLCS, what were the batting abilities of the hitters? How did these players perform this season in terms of batting average and home runs?

To answer this question, I obtained the latest season's player data from MLB's official website and analyzed the hitters from these four teams, focusing on metrics such as batting average and home run counts.

## 2 Design Decisions

Player statistics are extensive and complex, with some metrics requiring intricate formulas. To more intuitively convey the relationship between the data and player performance, I focused on the two most straightforward indicators: home runs and batting average. This choice ensures that even viewers unfamiliar with baseball can understand the visualization's message.

Home runs represent the total number of home runs a player hits during the season, while batting average (AVG), calculated by dividing a player's hits (H) by their total at-bats (AB), measures hitting efficiency. Both metrics correlate positively with player skill, making them excellent indicators of performance.

$$AVG = \frac{H}{AB}$$

First, I gathered home run and batting average data for players from the four championship series teams. Since the raw dataset included unrelated information, I pre-processed the data, selecting and merging only relevant statistics. I then sorted the data in descending order for clarity. To organize the presentation further, I incorporated interactive visualization, allowing viewers to filter and rank player data by team.

Interaction Features:

- **Filtering:** Users can select a specific team (NYY, NYM, LAD, or CLE) to display only the players' stats for that team, and the default "All" option shows rankings across players from all four teams.
- **Metric Selection:** Users can toggle between viewing either home runs or batting averages as the primary metric.
- **Hovering:** When hovering over the data points, the tool displays player details, including the team and their batting average or home run count.

Note: All data is pre-sorted in descending order, allowing for a clear ranking of player performance across these metrics.

These features, though basic, presented several implementation challenges. The primary difficulty stemmed from inconsistencies in the raw data format, which complicated the merging of player statistics. Secondly, some player statistics were imprecise due to varying numbers of at-bats, which could skew batting averages—players with fewer at-bats might appear to have higher averages. To address this, I set a minimum at-bat threshold for players.

### 3 Strengths and Weaknesses

Strengths:

- The visualization provides the strengths that enhance its usability and clarity
- The use of bar heights for metrics allows for quick, visual comparison of player performance, making it easy to interpret differences among players.
- Sorting the data in descending order aids in ranking players, while the option to filter by team enables focused comparisons within each team.
- The hover tool provides detailed information which improves user engagement and allows for closer inspection without cluttering the main visualization.

Weaknesses:

- The unstructured nature of the raw data made the implementation challenging. As a result, the presentation lacks polish and sophistication.
- If more metrics were to be added, preprocessing and implementing new interactive methods would require significant time and a redesign of the current algorithms. The current implementation lacks modularity, making it difficult to seamlessly incorporate new features.
- With only two metrics displayed, the visualization feels somewhat sparse. Adding more teams and player statistics in the future could enhance its depth and make the content more engaging.

## 4 User Study (Extra)

To evaluate the usability and effectiveness of the visualization, I conducted a simple user review with five participants, asking them to explore the tool and provide feedback on two questions:

- Can you identify the top players using this visualization?
- Can you determine which team is the most competitive?

### **Results:**

The feedback was unanimous. Users identified Aaron Judge from the New York Yankees as the top player, ranking first and second across the four teams in home runs and batting average, respectively. Shohei Ohtani from the Los Angeles Dodgers followed closely, ranking second in home runs and third in batting average. Based on the players' performance, participants concluded that the Yankees and Dodgers were the most competitive teams. These findings align with real-life data, as both Judge and Ohtani were recently named MVPs in the American and National Leagues, respectively.

### **Suggestions for Improvement:**

They expressed interest in seeing the full dataset, including players from all 30 teams, and suggested adding an option to view pitcher statistics.

In terms of visualization features, they would like to see more color variation to enhance visual appeal. Additionally, users recommended implementing slider functions, enabling them to filter players by batting average or other statistics rather than solely by team.

This feedback highlights areas for improvement that will be valuable for refining the visualization and expanding its functionality in the future.