Brigham Michaelis Avery Bolingbroke Thomas Baskin

PROJECT 1 REPORT

Introduction

Baseball. A confusing game of stats, steals and all-stars. In this report we attempt to show and discuss various analyses, ranging from compositions of the Hall of Fame by birth country(Excluding USA) to Team salaries through the years(1985-2018) to the most popular birth years in the Hall of Fame. Through it all, it must be stated that we can only say for certain what the graphs explicitly show. (LinkToSlides)

Dataset

The dataset we used was obtained from http://www.seanlahman.com/baseball-archive/statistics. We show here the description of the dataset.(Taken from the above website.) "***The updated version of the database contains complete batting and pitching statistics from 1871 to 2020, plus fielding statistics, standings, team stats, managerial records, post-season data, and more.***"

The method of getting this data into a usable format was dependent on each individual analysis. Generally the process involved the selections of one or multiple individual datasets and combining them and then paring them down in some manner to give a voice to our analyses. For example, we did one analysis to determine the composition of the Hall of Fame based on Country of Birth. To obtain the visual data we had to merge the people table with the HallOfFame table, and then remove unnecessary columns and exclude all members born in the USA.

Analysis Technique

Composition of the Hall of Fame by Birth Country.(Excluding the USA)

We wanted to visually show the composition of the Hall of Fame based on Country of birth. At first we merged the people table and the hall of fame table and matched on playerID's. Once that was done we displayed the composition of the Hall of Fame by Country of Birth. The main problem with this was that the USA overwhelmingly dominated all the other countries, so we excluded it from our data set which provided insight into a much more interesting data visualization than with the USA included.

Team Salaries through the years 1985-2018.

We wanted to visualize (for the given timeframe) how team salaries changed and evolved over time. Note that when we are talking about team salaries we are referring to the sum total of all the individual salaries of the members of that team for a given year. We then "line plot" those cumulative salaries through the years and make observations. We also identified and chose team salaries that looked "interesting".

Most frequently occurring birth years of Hall of Fame members

We wanted to visualize who's in the Hall of Fame by what decade they were born in. We combined the Hall of Fame information with each player, and then only considered each player

once, even though they may have been in the Hall of Fame more than once. This helps correct any skew from players that are too Famous.

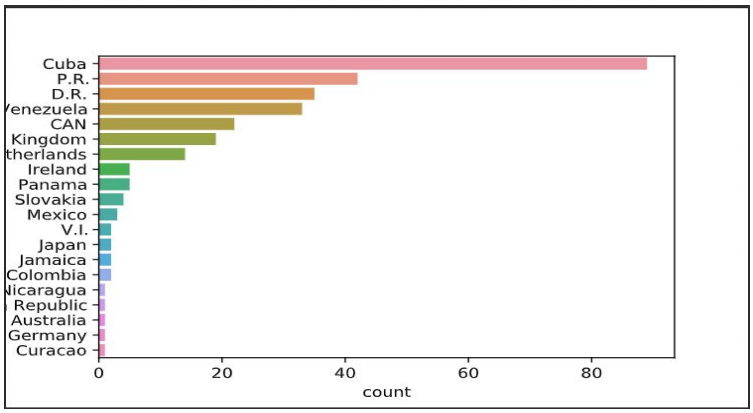
Which colleges sent the most players to the Hall of Fame?

We wanted to visualize which colleges sent the most players to the Hall of Fame, regardless of any other factors. As long as a player attended the college, they counted. We removed duplicate players from the Hall of Fame, and discounted all the colleges that sent under 4 players, in order to present a more valuable (and readable) analysis chart.

Results

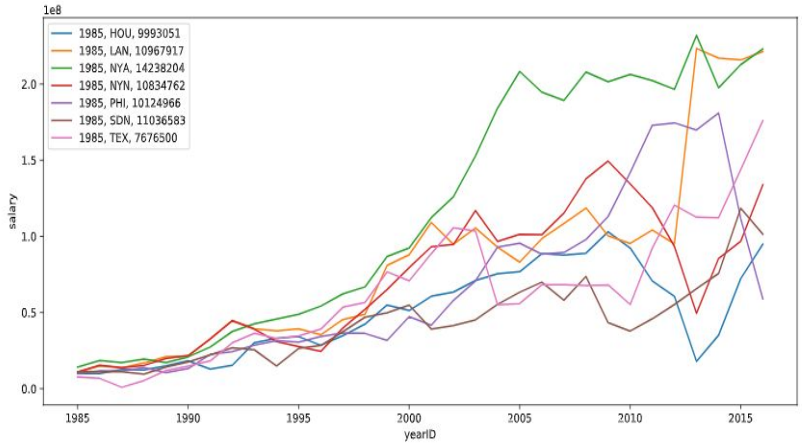
Our results for each separate analysis are given here.

1. Composition of the Hall of Fame by Birth Country.(Excluding the USA)



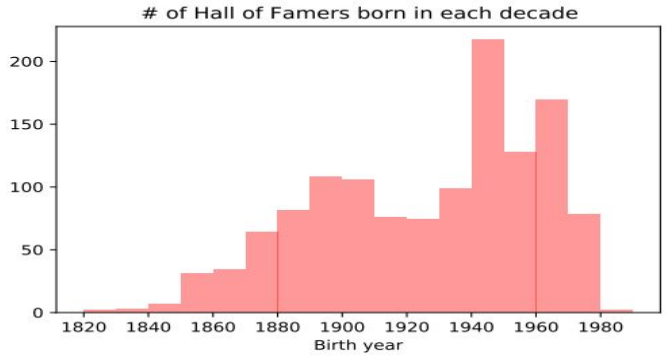
Please excuse the cut off names of Venezuela, United Kingdom, Netherlands, and Nicaragua. Here we show all the Hall of Famers for the given history of the Hall of Fame. From here, it would be really interesting to further discover and learn about why Cuba, and the D.R. and C.R. are the top 3 countries outside the US that produce players that then end up in the Hall of Fame.

2. Team Salaries through the years 1985-2018.



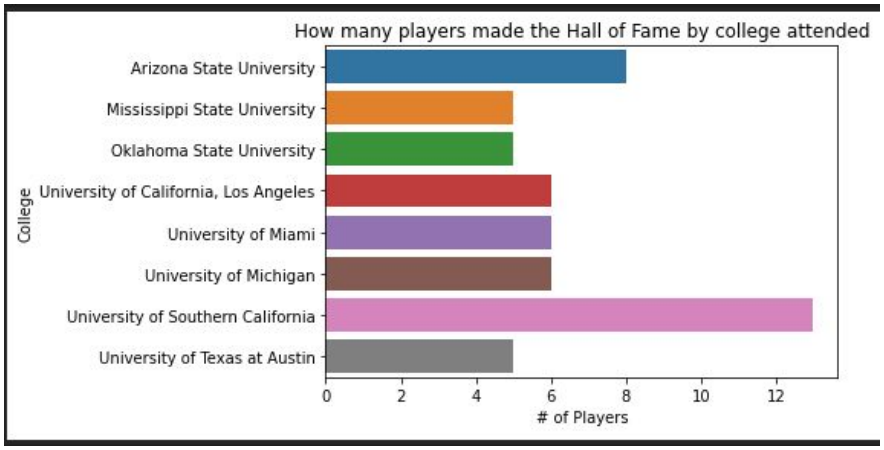
2012 was a bad year for team salaries, several team took significant cuts to their pay in that year. A general trend of increasing the pay 10 times over was observed for nearly all the teams. This is fascinating even the teams that started in the middle of the years and ended in the middle had to be competitive in their pay…. Or maybe it is better to say “There is and was a threshold you would have to meet to start a team.”

3. Most frequently occurring birth years of Hall of Fame members



From the chart we can see that the biggest group of Hall of Famers was born in the 1940’s, closely followed by the 60’s. That would mean the players are in their 60’s or 80’s today, respectively. The baseball Hall of Fame wasn’t established until 1936, so that explains the steady increase of baseball players born from the 1840’s to the 1900’s. But why the sudden jump for players born in the 1940’s? It would be interesting to see if there’s any correlation between outside events and the players’ nominations or if there were just a lot more players that were Hall of Fame worthy born that decade.

4. Which colleges sent the most players to the Hall of Fame?



The chart clearly shows that University of Southern California sent the most players to the Hall of Fame. It would be interesting to conduct further analysis on why that is. Is USC just better at training their players? Are voters for the Hall of Fame more biased towards USC? And Arizona State University surprised us as well.