Short Report

The Major League Baseball draft is one of the most unpredictable sports drafts every year. This is due to several factors, such as the inconsistent number of rounds in recent years, and the different playing levels that can be drafted via High School and College. With our data, we wanted to evaluate different metrics across certain categories to extract information pertaining to positions, schools, and the variation of rounds by any given year.

Our data set consisted of the entire MLB draft from the years 2018-2022. In the data set, it included rounds, player, team, year, position, and school that they attended. We retrieved this data set from Baseballsavant.com, which contains a plethora of information regarding statistics ranging from all teams to individual players. Our analysis that we conducted could be useful to not only MLB teams, but to draft prospects as well to provide them insight as to what their draft rank will possibly look like.

The first function, HSvCollege, calculates the percentage of high school versus college picks for each year and creates a bar chart for each year using the matplotlib library. The second function, TeamPosition, calculates the breakdown of positions for each team's draft picks and prints the results.The third function, TopPositions, filters out LHP, RHP, and SHP positions and calculates the top five positions with the most picks. The TopSchools function analyzes data on college baseball players drafted by MLB teams and displays the top 10 schools with the most players drafted in the first 5 rounds. It also generates a bar chart of these top 10 schools and prints out the count of picks for each of these schools.