**Identification of differences among the performance of the baseball players using Tableau**

Summary

The dataset used in this project contains information of 1,157 baseball players. The details in the dataset are confined into following variables.

* Handedness: right handed (R), left (L) handed or both (B) handed
* Height: height in inches
* Weight: weight in pounds
* Avg: batting average,
* HR: home runs (HR)

Purpose of this exploratory data analysis (EDA) is to create a visualization that shows differences among the performance of the baseball players. To fulfil this purpose, the facilities and techniques available in Tableau suite are used.

Design

The generated Tableau visualization can be obtained from the following link. A snap shot of it is shown in the Figure below (Figure 1) to get a quick idea to the reader.

Link to Tableau public workbook: [The visualization](https://public.tableau.com/profile/lasa6459#!/vizhome/Basebal_palyers_performance_Tableau/Summary_dashboard)



Figure 1: A snap shot of Tableau public workbook

In Tableau public workbook, the user can interactively visualize certain aspects by selecting different *handedness* form the legend as shown in Figure 1. In addition, the user is capable of scrolling down to invisible regions on some plots, and hovel over the plots to see some additional information.

In the Tableau workbook provided above, Fig 4, Fig 5 and Fig 6 show the distribution of height, weight and average scores of players. As shown in Fig 4, Fig 5 and Fig 6, majority of the players are right handed. The heights and weights of majority of players in the ranges of 70 to 76 inches and 160 to 200 pounds, respectively. There are about 260 players has average score of 0.0 (Fig 6). May be those players are not batters or there might be some errors in the data entry process for those players. Ignoring the zero average for some players, the median of average home runs is 0.25 for players.

Fig 1 shows the average home runs for each weight category. In Fig 1, the data separated based on handedness (B, L and R) and visualize each group separately or all together by selecting the corresponding group/s in the legend. Based on the data in Fig 1, the highest number of home runs scored by players falls into weight category 209 pounds. The next highest is the weight category 201 pounds.

The average of home runs for each height category is shown in Fig 2. Similar to Fig 1, in Fig 2 also, the data separated based on handedness (B, L and R). The highest number of home runs scored by players falls into height category 67 inches, and the next highest is in height category 74 inches.

Fig 3 shows the home runs scored by each player. This is broken down by weight and height of the player. Color shows details about handedness. The data is filtered on home runs, which includes values greater than or equal to 350. The player who scored most runs was Reggie Jackson (height: 72 inches, weight: 95 pounds, left handed), followed by Mike Schmidt (height: 74 inches, weight: 195 pounds, right handed). Surprisingly, majority of top five players are left handed despite the fact that majority of players in the dataset are right handed.

The duration of data collection is not mentioned with the dataset. Therefore, it is assumed that the data was collected for the player’s entire career as a batter on major games.

Feedback

The graphs are easy to read. When hovering over the tabs it clearly shows what details the bar represents.

The weight comparison of players and homeruns by weight of player should be side by side to interpret easily. Similarly, the homeruns by height and players by height should be side by side. In addition, it is better if you could include units into plots. Also, use same bin width for all the histograms to make them look nice.

Should you not also consider how many games each players played? That makes a difference in the homeruns they made. Or did all plyers play the same number of games? This might play a vital role than weight or height.