By Lasantha Rathnayake

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

In the initial design, to explore the variables, histograms will be used and to reflect performance of players, bar-graphs will be generated. Scatter plots will be used to identify relationships, if any, between home-runs and other variables. Because of the data types available in the dataset, these three types of plots are most suitable over other types of visualization in order to explore the data effectively. From this study, the user will be able to explore the visualizations interactively in Tableau Story format.

The generated Tableau visualization can be obtained from the upcoming links (in Story telling format). A snap shots of Tableau Dashboards are shown in the figures below (Figure 1 and Figure 2, before and after the feedback, respectively) to get a quick idea to the reader. However, the provided links are in Tableau Story formats for easiness for user to navigate and understand.

Link to Tableau public workbook **before** feedback (Story telling format): The visualization

By Lasantha Rathnayake

Figure 1: A snap shot of Tableau public workbook (Dashboard format) – before feedback
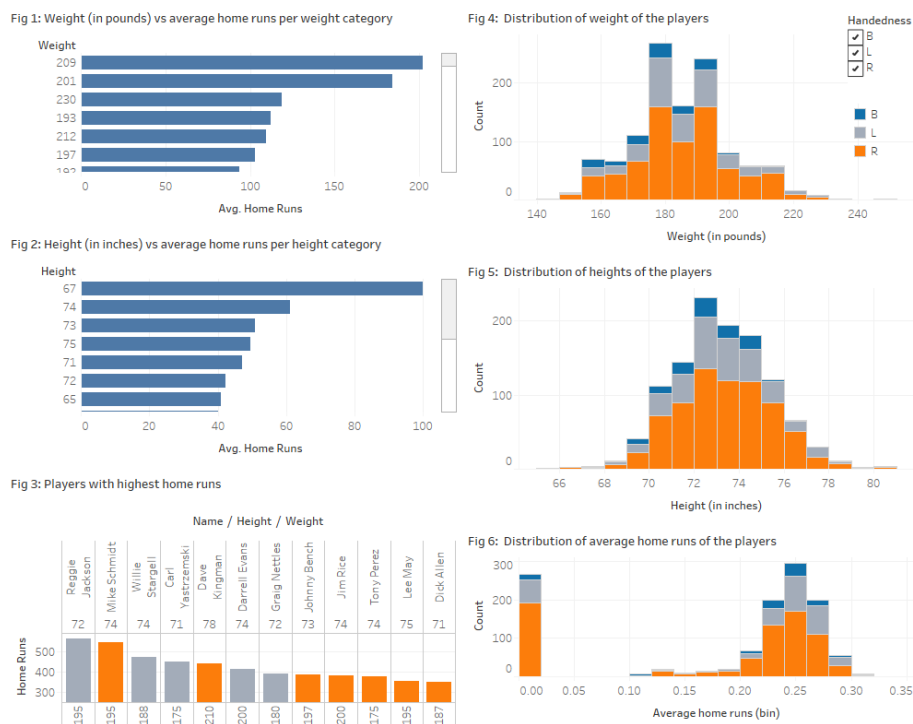

Figure 2: A snap shot of Tableau public workbook (Dashboard format) – after feedback

Link to Tableau public workbook **after** feedback (Story telling format): The visualization

In Tableau public workbook, the user can interactively visualize certain aspects by selecting different *handedness* form the legend as shown in Figure 1 and Figure 2. 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. Scatter plots generated but there are no visible linear correlations between scores and other parameters. [Note: 'Figure' refers to figures in this report and 'Fig' refers to the figures in Tableau workbook (in individual plots and Dashboard format)]

In the Tableau workbook provided in here (link), Fig 4, Fig 5 and Fig 6 show the distribution of weight, height 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.

Final remarks

When examining the data in the Figs 1 to 3 in Tableau workbook, it is understood that all the players, who performed well, fall into average sizes [height (70 to 76 inches: Fig 4) and weight (160 to 200 pounds, Fig 5) categories]. In addition, left-handed players seemed scored more compare to right-handed players (Fig 3).

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

Feedback 1:

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 bar 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.

Feedback 2:

Red-Green colors are not a good choice for using in comparisons since they are not color-blind friendly colors. The people with this disorder are not able to distinguish these two colors.

Addressing to the feedback

Most of the suggestions indicates in the feedback were corrected in the visualizations. Such changes are aligning corresponding plots, setting up a similar bar sizes in histograms and adding units wherever possible in plots. However, addressing the last part of the feedback is impossible due to lack of time information in the dataset.

The colors of the plots were changed with colorblind friendly colors.

Resources

Udacity course materials for Telling Stories with Tableau

Dataset: https://docs.google.com/document/d/1w7KhqotVi5eoKE3I_AZHbsxdr-NmcWsLTIiZrpxWx4w/pub?embedded=true