Tableau Story Write-Up

(Baseball Dataset)

Introduction

The purpose of this write-up is to better communicate the findings in each version of the Tableau workbook. The storytelling and visualizations help showcase the analysis performed on the baseball dataset (provided by Udacity), the dataset contains names for 1,157 baseball players and details specific attributes and statistics associated with each player. These attributes and statistics, also known as variables, include handedness (left or right), height (in.), weight (lbs.), batting avg, and home runs. Below are links to both versions of the Tableau storytelling visualizations.

Original Version:

https://public.tableau.com/shared/JF4WS28M9?:display_count=y&:origin=viz_share_link

Final Version:

https://public.tableau.com/views/AttributesthatincreaseaBaseballPlayersPerformance/PlayerPerformanceonAvgHomeRunsvsBattingAvg?:language=en&:display_count=y&publish=yes&:origin=viz_share_link

Summary

The visualizations that were constructed were designed to effectively communicate the findings regarding the baseball dataset. The project demonstrates correlations between various variables like height and weight using plots and graphs. Through data visualization, we can learn about how different variables relate to each other and discover specific factors that drive certain statistics within a dataset. The main purpose of this work is to visualize which of these attributes contribute the most to the peak performance of baseball players.

Design

Original version of the Project

The first revision of the project consisted of seven worksheets with visualizations to determine the relationships between the variables. Below you will find the name for each worksheet in this initial revision. Beside each name, an explanation is provided which outlines the reasoning behind the design choices made for the visualization.

Number of Players per Handedness – This bar chart was created to count the number of players the fell into each of the three categories of Handedness. I chose a bar chart because it was important to visualize the difference between each category (Right, Left, or Ambidextrous) in the most effective method possible. A uniform and consistent color was chosen to shade all three plots because it was important to visualize to the audience that each plot was portraying the same information, which was the number of players from the baseball dataset belonging to each Handedness. The bar chart was sorted by largest count to smallest count, so that the reader could quickly perceive which category contained the greatest number of baseball players.

Mean Batting Avg per Weight (lbs.) – This line plot was created to depict the relationship between a baseball players weight and the mean batting average for that weight. To reduce the number of unique datapoints for weight, I binned the weight variables values using a bin size of five pounds. The line plot contains two lines each representing a different manner in calculating the average batting average based on the values used in the calculation. Since the collection of the data or values was not performed by me, I felt unsure what a zero value for the variable Batting Avg. represents, and due to this uncertainty it was important to me that I provided both calculations of the mean batting average per weight, one calculation to include these zero values, and the other to exclude them. The color choice was done to keep in mind the colorblind audience, so blue and orange were chosen.

Mean Batting Avg per Height (in.) – The design decisions of this line plot reflect the same reasoning behind the choices made for the "Mean Batting Avg per Weight (lbs.)" line plot, except this visualized the relationship with height instead of weight. The height for each player was grouped in bins using a metric of one inch. The design and coloring choices for each line on the line plot are the exact same as above.

Home Runs per Weight (lbs.) – The design of this bar chart was to visualize in the most effective manner the total number of Home Runs using the same weight categories that were established previously, those being every five pounds. Since the bar chart visualized the same data across all categories, a uniform color of blue was used for each plot.

Home Runs per Height (in.) – The design for this bar chart mimics the bar chart above. All design choices used the same reasoning as the "Home Runs per Weight (lbs.)", except that this chart represented the total number of Home Runs using the height categories we established earlier.

Average Home Runs vs Handedness (colored by Height & sized by Weight) – The design of this bubble chart was used to portray the relationship between the Average amount of Home Runs and a players Handedness. Varying shades a blue ranging from very light blue to very dark blue were included to represent the height of the player for which the datapoint represented. The size of the bubble or datapoint was representative of the weight of the player. I wanted to provide enough information to show that the weight and height of a player is less correlated with the Average amount of Home Runs, than the correlation of the players Handedness.

Mean Non-Zero Batting Avg. vs Sum of Home Runs per Handedness – The design of this scatter plot was revolving the importance to visualize the correlation between the Batting Avg of a player and the amount of Home Runs they hit. I further increased the granularity by using the Handedness of the players to color the scatter plot points. The main objective was to show that regardless of the Handedness of the player, there is no correlation between the Batting Avg and the amount of Home Runs they hit.

Final version of the Project

The first thing I changed in my design after receiving feedback was the title. As the feedback mentioned, it was important to command the viewers' attention and hint toward the focus of the data visualization. The title "Attributes that increase a Baseball Players Performance" directly relates to the what the visuals are intended to depict. It also grabs the readers attention as it subtly creates wonder and desire to learn which attributes of a baseball player most impact their performance.

Next, I removed the worksheet “Number of Players per Handedness”, as advised in the feedback. It made sense to me as to why the audience did not find this visualization meaningful. It is common knowledge that more people are born with right-handed than left-handed. A bar chart to show the audience that most players are right-handed provided no real value or impact to the story the data is meant to provide.

Following the feedback on how informative the “Mean Batting Avg per Weight (lbs.)” and “Mean Batting Avg per Height (in.)” worksheets were to the viewer; the following actions were taken to highlight their importance. Removal of the orange line plot, which represented the calculation of the mean batting average per weight or height, using the inclusion of zero values found in batting averages, was removed. The reason behind the removal of this line plot was taken after doing research on batting averages for baseball players. It is far too rare for a player to ever have a 0.00 batting average; thus, it was determined these averages to be erroneous and to exclude them from the calculation. The bin sizes for both worksheets were modified to better demonstrate the correlation between the variables. “Mean Batting Avg per Weight (lbs.)” had its weight bin size changed to a size of ten pounds, and “Mean Batting Avg per Height (in.)” had its height bin size changed to a size of two inches. Also, the title for each line graph was changed to provide information concerning the bin sizes.

The worksheets “Home Runs per Weight(lbs.)” and “Home Runs per Height(in.)” were impacted by the category bin sizes being modified. To better visualize the correlation between the total number of home runs and the weight or height of a player, the bar chart was changed to an area plot to showcase the trend as the players weight or height increases. According to the feedback, I removed the explanation at the bottom that contained confusing information which Tableau had automatically populated concerning the action “Handedness” which was applying a filter action to Handedness across the entire data source.

Realizing that the “Mean Non-Zero Avg. vs Sum of Home Runs per Handedness scatterplot was confusing and crowded as the feedback suggest, the following modifications were made to transform this scatter plot into the “Player Performance based on Avg Home Runs vs Batting Avg” worksheet. Using the calculated field “Player Performance Ratings”, we categorized players into three groups based on their Home Runs and Batting Avg when compared against the overall average for each of those variables. This was an effective means to visualize which players were bad, average, or good.

According to the feedback, a new bar chart was created to combine weight and height into a single graph in order to find the optimal combination at hitting home runs. The worksheet titled “Optimal Weight & Height Combination for Hitting Home Runs” demonstrates this information. Filters were added for height and weight in order to remove small and insignificant data values that took away from the objective of the worksheet.

Next, the dashboard was removed although it showed pizazz as the feedback remarks, it did not depict or emphasize any specific information concerning the relationships.

All the worksheets have had annotations added to explain the relationship being displayed by the visualizations.

Finally, the story was updated to highlight important finding to coincide with the conclusion of the analysis in this write-up. As we progress to through each point in the storytelling, an explanatory caption is written so that the audience understands the importance of the visualization.

Feedback

- A great title commands the viewers' attention and suggests the main point of the visualization. How can you improve on your current title Baseball Players Data? What relationship(s) are you depicting through the data?
- I do not understand why you made a chart of the number of players per handedness. These charts seem to reflect the obvious: many players are right-handed, and only a handful of special players are ambidextrous. I am not sure that this chart adds much value.
- The mean batting average per weight graph seems to suggest that weight has no effect on batting average. This result seems to make sense to me since light players should be able to aim at the ball equally as well as heavier players. However, I am somewhat surprised by the fact that the Mean Batting Avg per Height visualization seems to suggest that taller players tend to have lower averages. As with weight, I would expect short and tall players to be equally good at aiming at the ball. Perhaps taller players must swing lower to hit it, a process that negatively affects their performance. What do you think the reason is?
- Because hitting home runs requires power, I would expect that the heavier a player, the more likely he will hit a home run. However, your histogram suggests this relationship only holds true up to 190 lbs. Beyond this "ideal" weight, increased weight is associated with a declining number of home runs. I wonder why since these heavy weights should have the most power.
- The explanations at the bottom of the Home Runs per Weight histogram are confusing to me.
- As with weight, the number of home runs increases with height to a certain level, then declines. Is there any way that you can combine the height and weight into a single graph depicting the number of home runs per height-weight combination? Doing so might yield and optimal height and weight for hitting home runs.
- The Mean Non-Zero Avg. vs Sum of Home Runs per Handedness graph is quite crowded and confusing. I am not sure what relationship it depicts.
- Have you considered writing an explanation of the relationships that each graph depicts under each graph?
- The dashboard has visual pizazz, but I am not sure what relationships you are trying to depict or emphasize by using it.

Conclusion

In conclusion we can found that a player's weight will increase their ability to hit a home run, however, at a certain height threshold a player's batting average will suffer, and this will affect their likelihood to

hit a home run. Left-Handed players are slightly more likely to hit a home run, but weight and height seem to be more correlated to that factor. It was discovered that the optimal weight and height combination to be a home run hitting machine is 72-73 inches, weighing in at 190-199 pounds. If only I could grow a foot and a half, than perhaps I could make it in the big leagues!