

Tableau Project Link

The following link is for Ver.1 and Ver.2 of the project.

<https://public.tableau.com/profile/carey.hack#!/>

Summary

This is data representing baseball player's height, weight, average, and home runs. Any entry with "0" in average and home runs has been removed to reduce noise. As I cycle through the data, I will show the optimal height and weight numbers associated with having the most and mean number of home runs and batting average. I will be looking for how the features given, height and weight, affect the performance metrics given, batting average and home runs. Then I will explore that feature a little more.

Design

I initially use scatter plots to show any relation. I first notice the weight numbers given show a heavy propensity to naturally land in 10 lbs. bins. I notice that weight may have been rounded up/down for most players after noticing this trend. I also start to notice that weight is looking more correlated to the number of home runs than the height is. As weight goes up, home runs go up, while height seems to show a more normal distribution for home runs peaking between 72 and 74 inches. On Slide 2 batting averages are compared to height and weight. The weight does not seem to correlate to a batting average in a noticeable way while height seems to show that taller players will tend to have worse batting averages.

Slides 3 and 4 will introduce line graphs showing the mean of batting average and home runs, treating weight and height as a continuous "x" scale. These charts will combine player data to give us the means of the height and weight of all players. To be honest, I include these as they were feedback given. They give great data on the means of players for home runs and batting average, but I find them to be kind of static. I feel given the context of baseball, knowing the individual stats is more telling about performance than grouping everyone into buckets of averages. Humble opinion and by all means, if I am overlooking the usefulness of this chart, please do tell me.

Slide 5 shows a scatter plot with exactly what I thought I would see. Batting average and home runs have a relationship. A higher batting average is a likely indicator of having more home runs. For example, any player with more than 150 home runs will have a batting average higher than 0.24, with only two exceptions.

Slide 6 will summarize the findings that the players with the most home runs will be 190 lbs. while also being 72 inches tall. I focus on home runs for this final slide as the slides with batting average do not show the level of relationship that I feel is expressed with home runs when compared to height and especially as they relate to weight. The bar charts for slide 6 will validate the number above. They will also give a breakdown for handedness. I believe weight and home runs have a relationship so I further investigate that, by handedness, while giving a line graph to show the mean of batting averages in the final slides top chart.

Feedback

1 Answer



1

Accepted Answer

Well done posting for feedback which shows how much you are interested in enhancing your plots and learning.

For the answer, I will comment on the visualizations which will help you to enhance the plots and to provide 2 versions of the story as required by this project requirements:

- First thing that catches the eye are the abbreviations, what does "R" and "L" stands for, please make sure to use full clear words
- In the title in Slide 2, it says 'Average vs....' Average of what? is it Batting Average?
- Try dropping the trend lines if possible
- Try replacing plots in slides 1 and 2 with line plots because I believe there are a better choice of plot type
- If you want a scatter plot, try 2 continuous variables (Home Run vs Batting Average)
- Last suggestion has to do with last slide, can you please make sure they have clear title for each plot



Wafic F Mentor 6 minutes ago

This was speedy, awesome, and very appreciated feedback. I felt silly overlooking little things like abbreviating B, L, and R. They make perfect sense to me but would not to someone who has never seen this data or cares about baseball. I followed all feedback except replacing slides 1 and 2. I feel like the relationship the scatter plots show and the tight groups of the plots give a ton of info on the Mean area that most of the players live in for the respective metric while still plotting each player individually. Instead, I included the line plots as they show great Mean info on a scale similar to a timeline. I also added the scatter plot comparing Home Runs and Batting Average as this shows great relationships as well while also allowing for a deeper dive with the various filters applied.

Resources

No resources outside of the Udacity course were consulted for this project.