Udacity Data Visualization with Tableau

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.

Data points found during this exercise show that players with the most home runs will be 190 lbs. while also being 72 inches tall. As weight increases so does the likelihood of having higher total homeruns and height will show a more normalized distribution between 72 and 74 inches. Batting average and home runs share a relationship as well as shown on slide 5, any player with more than 150 total home runs will have a batting average higher than 0.24, with only two exceptions.

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 seeing the heavy buckets in any bin ending in “0”. I also start to notice that weight is looking more correlated to the number of home runs than the height is. On Slide 2 batting averages are compared to height and weight.

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.

The final slide will give the bar charts that were used for the summary numbers given above. These charts can also be broken down by handedness.

I chose scatter plots to answer my questions about home runs. The scatter plot from slide 5 shows the power of scatter plots while giving the viewer an easy-to-understand view of the relationship of batting average to home runs. By breaking any scatter plot down into it’s respective quadrants; high:high, high:low, low:low, and low:high, the viewer can get a clearer understanding where the dispersion of points lie. I like to think of this as resetting the x,y axis for optimization. When this is done, our plots will give a clear view of the high performers.

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.