Data Story Telling – Baseball Data

Introduction

This report aims to present the explanatory data visualization steps and to communicate the findings and patterns on the performance of baseball players. The analysis is done using Tableau and a data set containing 1,157 baseball players including their handedness (right or left handed or both), height (in inches), weight (in pounds), batting average, and home runs.

1ˢᵗ Story and 2ⁿᵈ Story:

https://public.tableau.com/profile/mohamed8293#!/vizhome/BaseballProject-Initial/Story2?publish=yes

Summary

In this project I have used different visualization techniques to communicate insights on baseball data. I have analyzed the relationships between variables, such as: handedness, batting average, height, weight and home runs.

Design

Initial version

For the first analysis, *Number of Records by Handedness,* I chose a pie chart to easily see the handedness percentage and table of how many observations we have by the type of hand (left, right or both).

Name vs Home Runs and Batting Average *is* represented by bar chart to see the performance for each player.

Home Runs vs Batting Average is represented by a scatter plot colored by handedness. The plot shows no clear relationship between these variables just only trend.

Height and weight distribution are represented by bar chart to see the distribution of the data for height and weight vs number of records.

Handedness vs Height and weight are represented by bar chart to see the average weight and height for each player handedness.

Handedness vs Bat Avg. and HR represented by bar chart to see the average bat Avg and HR for each player handedness.

Batting Average vs Height is represented by bar chart. This plot is examining the maximum batting average per height. It seems that highest batting average have the players who have a height of 67 inches.

Batting Average vs Weight is represented by bar chart as well. In this plot we can see that the maximum batting average have the players who have a weight of 225 pounds.

For ***Home Runs vs Height,*** I chose bar chart to see which height leads to the most home runs. From this visualization we can easily see that players with height of 78 inches had the most home runs.

For ***Home Runs vs Weight,*** I chose an area chart to see which weight leads to the most home runs. It seems that players who have a weight of 230 pounds had the most home runs.

Final version

In the final version I changed some charts to emphasize the insights I found in the first version.

New design for plotting Avg HR vs Avg Bat Avg and Avg Weight vs Avg Height using scatter plot with handedness as a color filter, make the analysis easier and simpler plot

Use box plot for the analysis of HR and Bat Avg vs Handedness and the relationship between HR and Bat Avg.

Mix line plot with bar chart to analyze the effect of body shape as (weight and height) on performance as (Bat Avg and HR)

Feedback

1. My story is named Story 1, the relationship between HR and Bat Avg. with handedness number and percentage plot, should be separate or moved to HR and Bat Avg analysis dashboard

Action: I moved the relationship between HR and Bat Avg. to HR and Bat Avg analysis dashboard, it is in the story under this title (How Distribution & Statistical values are looks like for Batting Average (Avg) and Home Run (HR))

2. Handedness vs (Height & Weight) and (Bat Avg vs HR) to much details (2 dashboard) and plot can be not understanding easily

Action: use only one dashboard using scatter plot instead of charting plot that explain all information with out two much plots as before.

3. Impact of weight and height on performance in Story 1 not covered everything, just only performance without data distribution

Action: I used bar chart for the data distribution for body shape (weight and height) and line plot for the performance (HR and Bat Avg.), now we have more details and the plot explain all information needed.

Conclusion

The purpose of looking into the handedness data was to see if there were any differences from players who hit left, right or with both hands. It appears that the players who can hit left handed are performing better than their right-handed counterparts. Although the population of left-handed humans are 1 in 10, there is a much higher percentage of baseball players who are left handed (25%).

Left handed players also seem to hit slightly better as well based on batting average and hit more runs as well compared to other players.

Left handed players also have higher height and weight in our dataset

In the dataset we have 3 players best of the best based on the performance analysis

Generally, we have very good performance for players who have higher weight and height and for lower weight and shorter height