

Tableau Story of Baseball Data

Prepared by: Snehal Murkute

August 19, 2018

Objective

This project depicts a visualization story created in tableau using different graphs for baseball players data.

Objective of this project is to analyze various factors impacting players performance. This data set containing 1,157 baseball players including their handedness (right or left handed), height (in inches), weight (in pounds), batting average, and home runs. I newly created Performance_1 factor using average of Scaled feature HR and Bat Avg.

I have created two versions of story.

First Story :- https://public.tableau.com/profile/snehal1376#!/vizhome/Baseball_data_story_v1/BaseballData_Story_V1

Final Story :- https://public.tableau.com/profile/snehal1376#!/vizhome/Baseball_data_story_v2/BaseballData_Story_V2

Summary

Using tableau, I have created visualizations to explore the relationships between home runs scored & Bat Avg of baseball players with height, weight, handedness, and new custom feature BMI. I also created custom feature Performance_1 to compare players overall Performance with other features.

My analysis draws conclusion that left handed players are better batters as on average they scored most number of Home Runs and their average Bat Avg is only lower than both handed players by very small margin. There is also some correlation of Height & Weight with HR, Bat Avg and performance of players as explained in Tableau story.

Design

1. Created new variables Performance_1 using average of scaled value(i.e between 0-1) of HR and Bat Avg.

$$\text{Feature scaling formula} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

2. Created new measures(BMI) in tableau and used to create relationships between players BMI and Performance.

$$\text{BMI formula} = 703 * \text{weight (LBS)} / [\text{height (in)}]^2$$

3. Created Top performers scale and Top performers set for dynamic lookup of Top N Players. N can be updated using Top performers scale.

SNEHAL MURKUTE

4. Explored all the data using bar charts, histograms and scatter plots.
 - a. *Bar Charts* : Bar graph is used to compare things between different groups. Handedness dimension have three groups in it i.e. Left, Right and Both. Hence, Bar chart is ideal choice to compare players performances by handedness as I want to compare players performance between different groups.
 - b. *Histograms*: A histogram is used to discover, and show the underlying frequency distribution (shape) of a set of continuous data. Height and Weight data is continuous, hence Histograms are ideal choice to show distribution of players performance by Height and Weight.
 - c. *Scatter Plots*: Scatter Plot is used to determine relationships between the two different things, I wanted to explore if there is any relationship between Players BMI and performance, so scatter plots are ideal choice to unearth such relation if any present.
5. Used simple measures and small multiples for creating graphs to understand the data in easier way.
6. Added annotation and details on data fields in graphs to get information on hover over.
7. Initially I used line graphs, and used Moving Average to smooth out graph. But I removed Line graphs from project as Histogram is more ideal choice to represent continuous data i.e Height and Weight, same was promptly pointed by Udacity reviewer too.

Feedback

1. Use Moving Average in line graphs to smooth out graph.
2. It will be great to see new feature Performance derived using HR and Bat Avg to compare players overall performance.
3. Find relationships of new feature Performance with Height, Weight and handedness.
4. Create new feature BMI using Height & weight of Players and compare with players performance.
5. Show list of top performers by handedness and overall.
6. Use Histograms instead of Line charts to show distribution of players performance for continuous data.

Resources

https://en.wikipedia.org/wiki/Feature_scaling

https://en.wikipedia.org/wiki/Body_mass_index

https://onlinehelp.tableau.com/current/pro/desktop/en-us/publish_workbooks_tableaupublic.html

<https://onlinehelp.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-connect.html>

<https://interworks.com/blog/ccapitula/2015/05/07/tableau-essentials-calculated-fields-aggregate-functions/>

<https://www.evolytics.com/blog/tableau-201-how-to-make-small-multiples/>

<https://kb.tableau.com/articles/HowTo/dual-axis-bar-chart-multiple-measures>

<https://kb.tableau.com/articles/howto/finding-the-top-n-within-a-category>

https://onlinehelp.tableau.com/current/pro/desktop/en-us/sortgroup_sets_topn.html