

## Links to Tableau Public workbooks, published online:

### Earlier version:

[https://public.tableau.com/views/Baseballdatavisualization\\_early\\_version/Story1?:embed=y&:display\\_count=yes&publish=yes](https://public.tableau.com/views/Baseballdatavisualization_early_version/Story1?:embed=y&:display_count=yes&publish=yes)

### Last version:

[https://public.tableau.com/views/Baseballdatavisualization/Story1?:embed=y&:display\\_count=yes&publish=yes](https://public.tableau.com/views/Baseballdatavisualization/Story1?:embed=y&:display_count=yes&publish=yes)

- **Summary:**

The visualization deals with baseball data set with 1,125 player, the visualization elaborates on how high count of home runs are achieved by reviewing different attributes of players, for example percentage of left handed players is higher in the category with high count of home runs than other categories, also higher home runs count is associated with average batting per player, it is also associated with players with high BMI, while BMI measurement is shown to be not the best way to know about the health of an athlete.

- **Design:**

First I made a 3 new calculated fields to help change some of the variables from continuous to categorical to achieve better visualization

1- I used the HR (number of home runs) variable and change it to a categorical variable with three levels ( high count, intermediate count, low count) I used the 25%, median and 75% percentile of the data variable to set these categories

2- I computed the BMI using the Height and weight to get an idea of the body mass index for athletes and how it differs according to number of home runs

3- I changed the BMI to a categorical variable based on its known values whether the person is (obese, normal, overweight)

I used most of my base plots as bar plots as they brought the ideas forward in more confined way as I used my new calculated categorical variables

I used categorical data instead of continuous as the values I wanted to highlight were scattered and not clear in its original continuous form when plotted

I used stacked bar plots which is the best choice when dealing with 2 categorical variables and a continuous one

Changes after feedback :

I used color as my primary visual encoding, but I used colors of green and red which is not suitable for people with color blindness , a mistake that I altered after receiving the feedback

I added the average value as a constant line to make comparison with my findings more prominent

- **Feedback:**

I received these notes as feedback to my initial visualization :

- charts don't have a title , the aim of the plot is not clear
- Titles, captions, subtitles, and descriptions of the axes are not found on charts and they are essential for the chart interpretation.
- The acronyms don't clearly refer to the variables , they must be adjusted to be interpretable.
- avoid using the colors **green** and **red** together because colorblind people will have difficulty analyzing the graph.

I readjusted and updated all these points based on the feedback