Write-up

Links to Tableau Public workbooks

Initial version of the data visualization

https://public.tableau.com/profile/fb8468#!/vizhome/project_4_2/Story_v1?publish=yes

Final version of the data visualization (after first udacity review)

https://public.tableau.com/profile/fb8468#!/vizhome/project_4_2/Story_v4?publish=yes

Summary

Visualizations were created in an attempt to check if a special combination of weight and height means more Hitting power and thus more Home Runs scored.

To explore the relationship between weight/height and Home Runs, "Body Mass Index" metric was used.

Also we tried to look at handedness and if it can be considered as a factor that can affect the Batting Average and Home Runs.

Conclusion

What was found is that the majority of Baseball players are right handed, but it appears that lefties have more advantage when it comes to being in the Top players category by Batting Average recorded.

Moreover being left handed with the right weight (185 to 200 lbs) to height (72-73 inches) combination could well be one of the factors when it comes to scoring more Home Runs.

Design

Intro_dashboard

1. Where I live, baseball is not a popular/known sport like in the US. And just to give some introduction for the reader about the topic.
   And Based on feedback 2, I thought it would be good to include some details and links that would help the reader familiarize with the terms used.
   I created an 'intro_dashboard'.

2. The Main idea was to highlight to the reader which Batting Average and Home Runs are the most common among the sample of baseball players.

   - A **bar chart** (**batting average Distribution_v1**) was used to show the distribution of Baseball players by Batting Average. Percentage of total was used, created bins (as dimensions) in order to breakup BA into different groups and used colors in order to highlight bands with highest percentage.
     Created a parameter "Batting Average Range" to allow the reader to change the bin size and explore with the data.

     But Based on Feedback 1 and 2, in "**batting average Distribution_v2**" I decided to include a more descriptive Y-axis title "% of baseball players" and removed the grid lines to reduce the data-ink ratio

     Also included a title for X-axis "Batting Average Range and removed the Y-axis ticks and marks and included the percentage on top of each bar as suggested in Feedback 1.

     Based on udacity review, a histogram was used instead of bar plot (batting average Distribution_v3)

   - Another **bar chart** (**HR distribution_v1**) was created for the distribution of Baseball players by Home Run scored. Created bins (as dimensions) in order to breakup HR into different groups and used colors in order to highlight bands, also created "Home Run range" parameter to allow the reader to resize the bins and check how the bar chart is updated.

     But Based on Feedback 1 and 2, I have iterated (**HR distribution_v2** and **HR distribution_v3**) twice to change from count to percentage of

total and adding a more meaningful Y-axis title and then adding an x-axis title and moving the percentage values to the top of the bar. And removing the Grid lines.

Based on udacity review, a histogram was used instead of bar plot (HR distribution_v4).

3. To highlight to the reader the relationship 'if any' between Home Runs and Batting Average (comparing two quantitative variables).
   A **scatter plot** of Home Runs vs Batting Average was created and focusing on the idea that when the number of HR increases the range of Batting Average becomes narrower and the Batting Average range borders were highlighted by reference lines.
   Y-axis and X-axis title were added, x-axis ticks were removed highlighting only the reference line values.

   - Used a filter for Batting Average to only plot Batting average greater than 0.01.
   - Decreased color opacity to 45% to take care of over-plotting as much as possible.

BMI_dashboard

4. To show if there is a relationship between weight/height and Home Runs, I used the metric Body Mass Index that relates weight to height.
   For that reason a table calculation BMI was created "([Weight]/[Height]^2)*703".

   Created a **Bar chart** to show the distribution of the baseball players by BMI range.
   For the BMI range a dimension bin was created to break up the BMI into ranges. And for more clarity "bmi bins2" table calculation was used to display the range lower bound and higher bound on the x-axis.

   X-axis and Y-axis lines and grid lines were removed.

   Percentage of total was used and the percentage was added on top of the bar chart.
   Also created a tooltip to show the reader the Average weight and Height related to each of the BMI ranges.

   Color was used to highlight bars with highest percentage.

Title was added to explain the visualization for the reader.

Also decided to include 2 images and a link to BMI in Wikipedia page in the related dashboard (**BMI_dashboard**) to give the reader a quick idea about BMI and its formula and to keep the reader involved in the reasoning.

5. In the same dashboard **(BMI_dashboard**) and in order highlight to the reader the increase in the Home Run median values when the BMI range increases, I decided to create a **box plot without the whiskers** to pull attention to the median values as well as the distribution of the data points across the BMI ranges.

   Y axis title was used, and added a caption in order to explain to the reader the visualization.

Handedness_HR_dashboard

6. To show the distribution of handedness between players, I preferred to use a **Bar chart** (**Handedness**) rather than pie chart to maintain the attention on the percentage of total rather than distracting the reader with different colors.

   I have added an action filter to the dashboard with source being the Handedness worksheet. Allowing the reader the choose what to check.

   Tried to minimize data-ink ratio, by removing the Y-axis and X-axis as well the grid lines, and without axis titles.

   Percentage of total as label on top of each bar.

   I have used the title to give the reader an idea about the visualization.

7. I used again a **bar chart** (**percentage Top N BA**) with a filer for Top players by Batting average. For interaction purpose I added a slider to allow the reader to dynamically change the values from top 1 to 100 and thus drawing attention to the percentages by handedness.
   Tried again to minimize data-ink ratio, by removing x and y axes and getting rid of grid lines.
   I used the title to ask a question to keep the reader involved.

8. First I have created a lineplot (**lineplot_median_HR)**, median home runs vs BMI range and using color to differentiate between handedness values, to show how handedness, BMI range might have an effect on the Median home run values.

   - But Based on Feedback 1, explaining that the reader will look for the overall trend in the lineplot and to check the differences between handedness each time he had to look at the Y-axis. Thus I suggested to use a bar chart instead which he preferred.
   - Then finally to convey the message that Handedness might be a factor that increases Home runs scored, I created a grouped bar chart (**grouped_barchart**) based on Handedness and BMI range.

     To highlight the BMI range color encoding was used.

     And a filter was added to allow to reader to choose which BMI range to visualize.

     Tooltip used to show median Height, median Weight.

     I had to iterate twice here:

     In **Handedness_HR_dashboard_v2,** I had the BMI legend of the grouped bar chart.
     Based on the Feedback, I removed it in **Handedness_HR_dashboard_v3**

Feedback

Feedback1

Batting average Distribution_v1

The graph is not clear:

- Y-axis title is not informative and should be changed
- X-axis title should be added
- Percentage should be on top of each bar

HR distribution_v1 and **HR distribution_v2**

- Try to use percentage instead of count this will give more meaning to the graph.
- X-axis title should be added to understand the visualization
- Y-axis title is not informative, try to use a title that include "the word players"

lineplot_median_HR in **story_v1**

- Try to replace the line plot visualization with another one or enhance it to emphasize more on the difference between handedness. I am looking at the overall trend in the lineplot and to check the differences between handedness each time I had to look at the Y-axis.

Grouped bar chart in **Story_v2**

- BMI color legend should be removed no need for it since the BMI range is added on the Y-axis already.

Feedback2

If you can add a definition for the baseball terms used.

Batting average Distribution_v1

- Something missing in the graph, probably you need to Add X-axis and Y-axis title, I did not get what you trying to convey

HR distribution_v1 and **HR distribution_v2**

- Something missing in the graph, probably you need to Add X-axis and Y-axis title, I did not get what you trying to convey

BMI_dashboard

- I like the idea of adding images to explain what is BMI

Resources

https://www.edupristine.com/blog/top-n-analysis-in-tableau

https://community.tableau.com/thread/141653

https://www.youtube.com/watch?v=xWC8_Apx2Es

https://www.youtube.com/watch?v=opfVV1maNVw

https://www.youtube.com/watch?v=Y5svmFTSdlo

https://www.thedataschool.co.uk/jevon-dacosta/the-proper-way-to-label-bin-ranges-on-a-histogram-tableau/

https://kb.tableau.com/articles/howto/excluding-values-from-a-filter