Jaya Vijayen
Data Analyst Nanodegree Program
Data Visualization in Tableau

Summary

I wanted to see how handedness, weight, height and batting average affects the home run performance of the baseball players. This data visualization attempts to answer these questions. The baseball data visualization showed the following results:

First, the data visualization shows left handed baseball players had a better average home run performance in their games. Left handed players have considerable advantage on the diamond. This is because left handed players are closer to first base. Hence, their swinging momentum is towards the first base. In addition, left handed players also have a good ball angle.

Second, I wanted to know if weight affected home runs. From the visualization, the top twenty five baseball players who had the most home runs weighed between 180 to 200 pounds. Hence, weight within this range resulted in greater performance.

More than 90% of baseball players have heights between 65 to 80 inches. The visualization shows that the ideal height for scoring high home runs would be between 70 to 78 inches. This conclusion is based upon the top twenty five baseball players' height.

As far as the batting average, most players are between the ranges of 0.21 to 0.30. The top twenty five baseball players fell within this range. Hence, the batting average does not impact the home runs.

Finally, I wanted to find out if there is any correlation between handedness and batting average. From the analysis, there was only 0.1 – 0.2 difference in the handedness, which is very minimal. As a result, handedness does not affect the batting average.

In conclusion, handedness, weight and height does impact home run performances. However, batting average does not affect home runs. In addition, batting average is not influenced by handedness.

Links:

i. Version 1

https://public.tableau.com/profile/jaya1357#!/vizhome/Baseball_Story_1_0/Story_1?publish=yes

ii. Final Version

https://public.tableau.com/profile/jaya1357#!/vizhome/Baseball_Story_Final_1/Finale?publish=yes

Design

A. Chart Types

Given that I had over 1000 baseball players in the dataset, I wanted to have several chart types for my project. Some of the charts I used in this project were vertical bars, scatter plots, strip charts and horizontal bars.

The classic bar chart is easy to use and easy to understand. One of the reasons why bar charts have been around for quite some time. We have the ability to process the length of bars with extreme efficiency which makes the bar chart one the most powerful data visualization of choice. In this project, I used the bar charts to find out the handedness (left hand, right hand or both) of the baseball players.

Scatter plot shows relationship between two variables. It gives a sense of trends and outliers that leads to deeper investigations of the baseball data. For example, I wanted to find out if there is any correlation on the baseball player weight and their home run performance.

Strip chart are useful when we want to see a lot of data at a glance and understand the concentration of the data. It plots the data in order along a line with each data point represented as a circle. I used this basic plot to show the concentration of the baseball player height to identify which height was the most common among the baseball players.

B. Color

Color is key for in my data visualization project. As mentioned in the Udacity lessons, “It allows us to encode another variable or split the data in groups”. For my project, I took into consideration for people who have colorblindness. I decided to use Tableau color blind palette. I also used color to bring out important summaries in each of my story line so people can follow along.

In addition, I utilized shape, size, label, details and tooltip to give the data more depth and clarity. For example, I used diamond symbol shape. I then resized the diamond for the correlation of home runs and weights for the top twenty five players. I used the tool tip to show the weight and home runs. The labels were used to showcase the number of home runs and weight.

C. Feedback

Here are some of the feedback received for my first Tableau Story. I enhanced my project by incorporating these feedback. I went to the specific sheets and made the corrections.

General Comments

1. Please remove the acronyms from your dashboard. For example, use “Left” “Right” and “Both” instead of “L”, “R” and “B”.9
2. The index does not explain what B, R and L means.
   For 1 and 2, I did a right click on the x axis. A dialog box popped up. I then selected “Edit Alias”. I changed “L” to “Left”.

3. The text on the x and y-axis is too small.
4. Titles, captions, subtitles and descriptions of the axes are essential for the chart interpretation. Please review your charts to guarantee that this information is legible.
   For 3 and 4 above, I did a right click on either x or y-axis. A dialog box appeared. I selected “Edit Axis”. If we look below there is something called “Axis Titles”, I renamed the title. Repeat the step above, instead of selecting “Edit Axis” choose “Format” to increase the size of the font.
5. The text on the legend box is too small.
   For the legend box, I clicked on the tiny triangle located on the top right corner of the legend box. I then selected “Format Legends” to fix this problem.
6. Put a border around the explanation box so it pops to the reader. As well as add color to the text so it draws the reader.
   I selected the text box. Next, I clicked on the small triangle located on the top right corner. I chose “Format description”. A format description pane appears on the left. I clicked on “Border” and chose the line I wanted. This added the border around the text box. For color, I selected the text box. Next, I clicked on the small triangle located on the top right corner. I chose “Edit description”. I then selected the color for the text.

Comment for Analysis 1 – Home runs affected by handedness

1. Consider deleting redundant words in the conclusion statements. For example, for Analysis 1 no need to say “the bar graph indicates”.
   I selected the text box. Next, I clicked on the small triangle located on the top right corner. I chose “Edit description”. I removed redundant words and reconstructed my sentences.

Comment for Analysis 2 – Does weight increase home run performance?

1. In the conclusion statement, provide indication of the range of weight of the baseball players.
   I went to the “Story” tab located in the bottom. Click the story tab and on the left you will see a pane. On that left pane, there is a “Drag to add text”. Drag that to the sheet and let go off the mouse. An “Edit description” box appears. I typed my conclusion statement.

Comment for Analysis 3 – Weight of top 25 home run hitters

1. The column with names requires sufficient width to wrap around text.
   I clicked on one of the baseball player name. For example, I clicked on “Reggie Smith” and a dialog box appears. I chose “Rotate Label” and it gave the full name of the players on the axis.

Comment for Analysis 4 – Height of players

1. Consider using a different type of plot, perhaps bar charts to compare heights of players.
2. If using scatter plot, consider changing the range of the x-axis, maybe between 60 and 80 inches.
   For 1 and 2 above, I changed the scatter plot to a strip chart. This is an option that is available on the "Show Me" tab located on the top right corner. This "Show Me" tab has various charts and graphs. I like it!
3. The x-axis should say height in inches and y-axis should say number of home runs.
   For 3 above, I did a right click on either x-axis or y-axis. A dialog box popped up. I selected "Edit Axis". If we look below there is something called "Axis Titles", I included the inches as well as renamed the "HR" to "Home Runs".

Comment for Analysis 5 – Height of players for top 25 home run hitters

1. Don't show the height with a zero decimal.
   I did a right click on the x-axis. A dialog box popped up. I selected "Format". A pane appears on the left. Go to "Scale". Next, choose "Numbers" and select "Numbers (Customs)". Reduce the decimal places from "2" to "0". This removed the decimals.
2. The column width for name needs to be widened.
   I clicked on one of the baseball player name. For example, I clicked on "Bobby Bonds" and a dialog box appears. I chose "Rotate Label" and it gave the full name of the players on the axis.

Comment for Analysis 7 – Batting Average

1. The title, average of batting average sounds incorrect.
   I did a right click on the x-axis. A dialog box popped up. I selected "Edit Axis". If we look below there is something called "Axis Titles", I renamed the title.

D. Resources

1. http://www.perceptualedge.com/articles/ie/the_right_graph.pdf
2. https://www.tableau.com/learn/tutorials/on-demand/
3. https://nehal96.github.io/PISA-2012-Data-Visualization/v4/
4. http://raul-p.github.io/pisa2012/
5. https://public.tableau.com/profile/amir.rahbaran#!/vizhome/WiP-Udacity-FP/EconomyofFreedom?publish=yes
6. https://github.com/bcko/Ud-DA-Tableau-Titanic/
7. https://public.tableau.com/profile/helen.luu#!/vizhome/UdacityDataVisualizationProjectFinal_0/Baseballperformance

E. Data Files

- Dataset: baseball_data.csv