

# Data Visualization with Tableau

## Kairui Wu

Final Version of Visualization:

[https://public.tableau.com/profile/karry555#!/vizhome/FirstVersion\\_0/Story1](https://public.tableau.com/profile/karry555#!/vizhome/FirstVersion_0/Story1)

First Version of Visualization:

[https://public.tableau.com/profile/karry555#!/vizhome/Book1\\_27921/Story1](https://public.tableau.com/profile/karry555#!/vizhome/Book1_27921/Story1)

### Summary:

The dataset contains 1,157 observations on the handedness (right, left, or both), the height (in inches), the weight (in pounds), the batting average, and the home runs of baseball players. For this visualization project, I come up with several questions to explore.

1. Do weight and height influence the batting average and the home run of players?
2. Is there any relationship between the home run and the performance of players?
3. Do players have different handedness have different performances?
4. From the dataset, what kind of players can achieve good performance more easily?

### Design:

For the first question, I created a new column of Weight to height ratio and divided these values into different bins. Using bar chart, we can easily find players with which weight to height ratio will have better performance and more home run.

For the second question, I created a scatter plot on batting average and home run and made a linear line to find the relationship.

For the third question, I created boxplot of batting average and home run on different handedness.

For the last question, I made a dashboard on which there were all of the plots and summed up the findings.

For the color of these plots, I use blue and grey theme. This is because the visualization should pursue a high data to ink ratio.

### Feedback:

The initial version of the visualization was shared with my classmates at University of Toronto. Below is the received feedback:

1. Only focusing on the influence of weight or height on performance may not be accurate. The body weight should be normalized.
2. It is not accurate to use line plot to present a variable that is not continuous. Using bar chart would be better.

3. When presenting a relationship between categorical variable and value variable, using boxplot would make the visualization more clearly.
4. Here needs to be a summary dashboard to present all the findings together.

**Resources:**

The original dataset used in this project was selected from the [Data Set Options](#) by Udacity.