**Data Visualization: Baseball players Performance**

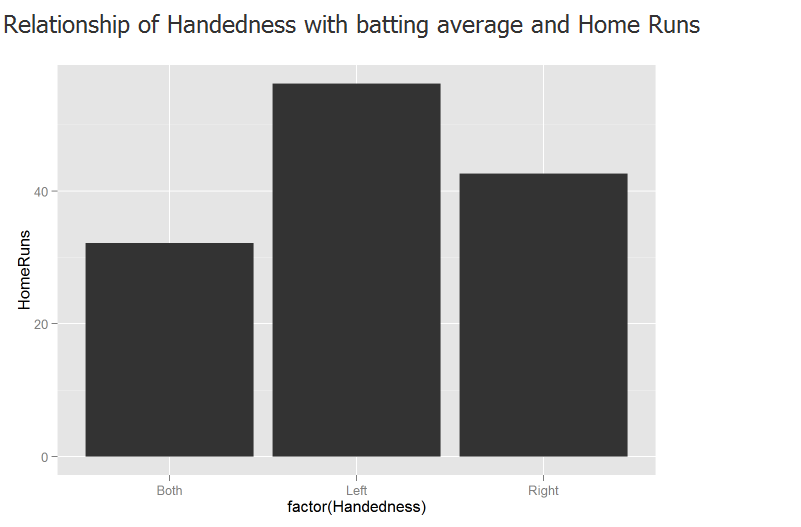
**Summary**

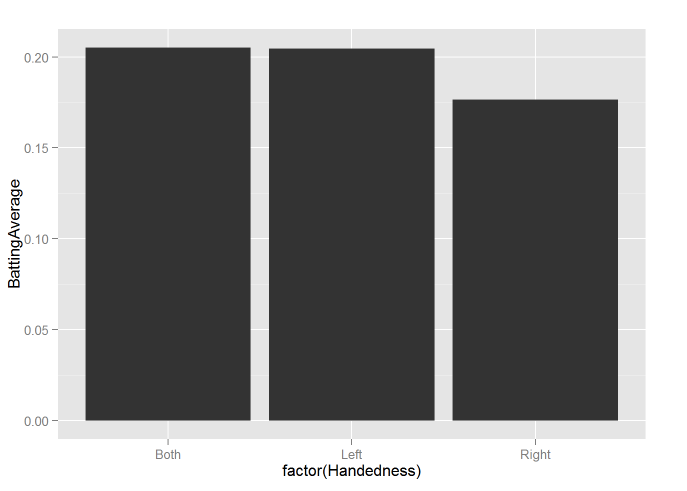
The goal of this visualization is to compare the performance of 1157 baseball player in terms of their Handedness and how the performance measures (batting average and home runs) varies with player’s Height.

# Design

#### Exploratory Data Analysis in R

I downloaded the Baseball Data from Udacity’s dataset options that included performance and physical aspects like height, weight and handedness of 1157 players. I have analyzed the data using R. Some basic statistics and plots revealed a clear performance difference between left and right handed players as shown below:





The remain analysis can be found in PlayerDataExpolration.Rmd

#### Data Visualization (dimple.js)

To emphasize my finding in exploratory analysis that **Left handers usually have better performance** in terms of batting average and Home Runs, I introduced one more variable **Height to explain how the performance of left and right hand players varies by Height**. Height was showing an interesting trend in my exploratory analysis. Exploratory Analysis shows that players in certain Height range (70-75) have a better average home run.

Line chart helps to determine the relationship between two sets of values. As my attempt was to show the performance trend by Handedness line chart was selected.

A players performance can be measured by batting Average as well as home runs. Hence two separate visualizations were created:

* Mean Batting average variance by Height faceted by Handed ness
* Mean Homeruns variance by Height faceted by Handed ness

I also included two bar charts to show the performance (Batting Average & Home Runs) summary by Handedness before the trend chart as it will give a clear idea to audience before looking at the trend. A bar chart is usually used for comparison of categorical data. Here Bar chart was selected as I’m trying to compare the left and right hand player’s performance. The visualization can be found in index\_2.html.

I interviewed 3 individuals in person, and asked for their feedback on the data visualization after prompting them with the background information and a small set of questions. Highlighted comments from them are listed below:

## Feedback

I interviewed 3 individuals in person, and asked for their feedback on the data visualization after prompting them with the background information and a small set of questions. Highlighted comments from them are listed below:

##### **Interview #1**

I have showed this graph to my colleague and he gave the following comments:

The graph is clearly explaining the performance differences between left & Right handed players. Some improvements I see are:

* Y-axis of first chart is showing repeated values. Rounding it to 2 decimals will solve the issue
* The line graph shows close trends which is difficult to follow for some ranges where the mean values are very close. Highlighting the Left Handers trend to emphasize the better performance .
* Also adding title to the summary charts will help to convey the message effectively.

**Post-feedback Design -1**

* Y-axis values were corrected by rounding to 2 decimal places
* A mouse over event was added to the trend lines to highlight the focused trend line .
* Added title to bar charts

##### **Interview #2**

Second person who reviewed the improved graph was another Nano-degree student in my organization and her commend was:

I loved the visualization which very clearly conveying the message. Color combinations are really nice. However one thing I noted is the color coding used to encode handedness in the bar chart and line chart are different. This will mislead the viewer. I would suggest to use the same color coding for Handedness for all the charts.

Similar to bar charts add titles for you line charts also.

**Post-feedback Design -2**

* Same color coding is used for all charts
* Titles are added for line charts

##### **Interview #3**

Third person with whom I reviewed my graph with was a six sigma black belt of my team and here are the highlights from his comments: “This is an excellent visual for the given data set. However some minor improvements suggested are:

* Use a standard title for the plots
* Give a small description about the chart and it possible a link to the data set