**Explanatory Analysis of Baseball Player Performances with Tableau**

**Luyuan Zhang, Aug. 2018**

**Tableau Story:**

Initial version: <https://public.tableau.com/profile/luyuan.zhang#!/vizhome/BaseballDataversion-0/DependenceofBaseballPlayerPerformancesonPhysicalTraits?publish=yes>

Final version: <https://public.tableau.com/profile/luyuan.zhang#!/vizhome/BaseballDataversion-1/DependenceofBaseballPlayerPerformancesonPhysicalTraits?publish=yes>

**Summary**:

* I investigated the baseball data set, that has 1157 baseball player’s name, handedness, height, weight, batting average and home runs.
* Following data transformations were performed for better visualization:
  + natural log transformation of home runs
  + binning of weight
  + binning of height
  + calculation and binning of BMI
* Scatter plots with size encoding number of players in the bin, were plotted to show correlations.
* Key observations are that smaller (short and light) players have better batting averages, and bigger (tall and heavy) players have better home runs.

**Design:**

I want to know whether baseball player performances are correlated with physical traits of height and weight. The first plot comes to mind is obviously scatter plots. However, when encoding details with player name, it is very difficult to see any correlations. To reduce noise level, I decided to group players into bins. When column values are weight or height bins, default plot is bar charts. Correlations are apparent in bar charts with outliers. Player counts in these outliers are only a few, compared to hundreds in the each other bins, and with the bars having same width, weights of outliers are enlarged, so I switched to scatter plot with size encoding player count. In this way, outliers become small dots, while majority of data are big dots, and audience attention will be focused on majority of the data instead of outliers.

After plotting correlations of batting average and home runs with height, weight and BMI, I found the two measures of performances have almost opposite dependences on the independent variables. However, the two performances show apparent positive correlations with each other. This is kind of a paradox, and is interesting. So I decided to show the correlation between the two performances first, then show their opposite dependences on height, weight and BMI, in hope that this layout will grasp audience interest.

**Feedback:**

I asked my friend, who is a data scientist, to take a look at my tableau story. Her feedback is “Try to make the axis title more readable”. So I renamed the axis titles so they can be easily understood by audience.

**Resources:**

<https://community.tableau.com>

<https://kb.tableau.com>

<https://en.wikipedia.org/wiki/Batting_average>