

A Story for the performance of Baseball Players

Author: Xueyun Zhou

1. Summary

This report is a Tableau story about the performance of baseball players measured by the number of home runs. The story finds that only a few players have large numbers of home runs; the players with median weight, of tall and very tall, with the overweight BMI, or of right-handedness are more likely to have large number of home runs; only the players with median and high rates of batting average are possible to have large number of home runs.

2. Design

The story visualizes a dataset^[2] about 1,157 baseball players.

The independent variables are grouped by values as following tables, in order to analyse which groups of factors are more likely to result in large numbers of home runs.

Table 1 Weight Groups

Group	Weight (lb)
Very light	[140, 161]
Light	(161, 182]
Median	(182, 203]
Heavy	(203, 224]
Very heavy	(224, 245]

Table 2 Height Groups

Group	Height (in)
Median	[64, 68]
Tall	(68, 72]
Very tall	(72, 76]
Giant	(76, 80]

Table 3 Rate of Batting Average Groups

Group	Rate of Batting Average
Very low	[0, 0.082]
Low	(0.082, 0.164]
Median	(0.164, 0.246]
High	(0.246, 0.328]

A new variable BMI is created by the formula^[1] as below to measure the health conditions.

$$BMI(kg/m^2) = \frac{Weight_{lb}}{Height_{in}^2} \times 703$$

The variable is also categorized by the following standard^[1].

Table 4 BMI Category

Category	BMI (kg/m ²)
Normal (healthy weight)	(18.5, 25]
Overweight	(25, 30]
Obese	>30

From the first scene of the story, it is noticeable from the histogram that the distribution of home runs is extremely right-skewed. There are a great number of players who have 0 HR, and a few players who have hundreds of home runs. Here we define that a good player with a large number of home runs are a player with over 100 home runs.

From the second and the third scenes, the boxplots tell that the players with median weight (182-203lb), of tall and very tall (69-76in), with the overweight BMI (25-30kg/m²), or of right-handedness are more likely to have large number of home runs.

From the fourth scene, the scatterplot illustrates that it is hardly possible that players with low and very low rates (0-0.164) of batting average have over 100 home runs. Only the players with median and high rates (0.164-0.328) of batting average are possible to have large number of home runs.

3. Feedback

3.1 First feedbacks from co-workers

The first version^[3] of the story is shared at

<https://public.tableau.com/profile/xueyun.zhou#!/vizhome/StoryforthePerformanceofBaseballPlayersv1/Story?publish=yes>.

3.1.1

Feedback: There are too many outliers for HR and Avg, especially zero.

Solution: Add a filter for the value of the variables in each dashboard, so the zero can be filtered out. By default zero is filtered out to decrease the fluctuation of the line.

3.1.2

Feedback: There is too much space for the two aggregation scatterplots on the right corner of the HR and Avg dashboards. As the values are not so large, the points are discrete. If the values were thousands, all the points would have been crowded and seemed to be the same value. This also results in distortion.

Solution: The large space is because all the points are far from the coordinate. If setting the coordinate to start from a value larger than zero, then it is against integrity. Because the points are discrete in these charts, this report temporarily does not update for this feedback.

3.1.3

Feedback: Some borders and grid lines are useless.

Solution: Remove the borders of boxplots and scatterplots. Remove the grid lines of boxplots. The borders for the bar chart *Avg & HR by Player* are remained because they can clearly divide the columns. The grid lines of scatterplots are remained because they can exhibit the values.

3.2 Second feedbacks from reviewers

The second version^[4] of the story is shared at

<https://public.tableau.com/profile/xueyun.zhou#!/vizhome/StoryforthePerformanceofBaseballPlayersv2/BaseballPlayerStory?publish=yes>.

3.2.1

Feedback: There are no specific conclusions in the Summary part.

Solution: Include a specific, clear finding from visualization in the Summary part.

3.2.2

Feedback: As the final visualization is only a story, the file is not tidy with too many supporting parts.

Solution: Hide the supporting worksheets and dashboards.

3.2.3

Feedback: There is a worksheet which is not used.

Solution: Delete the worksheets which are not used.

3.2.4

Feedback: The size of some dashboards and the story is set as fixed, so they are too small and not displayed properly on some devices.

Solution: Set the Size to be Automatic for all dashboards and the story.

3.2.5

Feedback: The story only shows the title and make a very simple summary of the visualization.

Solution: Illustrate the findings and conclusions from the story (in the title of the story and in separate text boxes). Make sure that from the written narrative, the audience can clearly understand each chart, the conclusions, the purpose and results of the analysis.

3.2.6

Feedback: The question, structure and the theme of the story are not clear. The story has no focus.

Solution: Once again, frame the story and structure, draw up a clear theme for analysis, and make visualization.

3.2.7

Feedback: The scatterplot and the line chart are overlapped.

Solution: Delete the line chart which is overlapped on the scatterplot.

3.2.8

Feedback: There is a table for each player, which is nonsense.

Solution: Delete the table for each player and add charts for summaries of the information.

3.2.9

Feedback: The titles are not standard, comprehensive and clear.

Solution: Update the titles so they can completely summarize the contents of the charts.

3.3 Third feedbacks from reviewers

The third version^[5] of the story is shared at

<https://public.tableau.com/profile/xueyun.zhou#!/vizhome/StoryforthePerformanceofBaseballPlayersv3/BaseballPlayerStory?publish=yes>.

3.3.1

Feedback: There are too many meaningless filters in the story.

Solution: Remove the filters that do not contribute to the conclusions.

3.3.2

Feedback: There are no color and shape legends.

Solution: Show color and shape legends for all charts which use color and shape as marks.

3.3.3

Feedback: The titles are lost in the first and fourth scenes of the story, because the worksheets are added instead of dashboards.

Solution: Insert the worksheets in dashboards, and add dashboards instead of worksheets into the story.

Resources

[1] En.wikipedia.org. (2019). *Body mass index*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Body_mass_index [Accessed 1 Sep. 2019].

[2] GitHub. (2019). *udacity/new-dand-advanced-china*. [online] Available at: <https://github.com/udacity/new-dand-advanced-china/blob/master/Tableau%E5%8F%AF%E8%A7%86%E5%8C%96/%E5%88%9B%E5%BB%BA%E4%B8%80%E4%B8%AA%20Tableau%E5%8F%AF%E8%A7%86%E5%8C%96%E6%95%85%E4%BA%8B.md> [Accessed 1 Sep. 2019].

[3] Public.tableau.com. (2019). *Tableau Public*. [online] Available at: <https://public.tableau.com/profile/xueyun.zhou#!/vizhome/StoryforthePerformanceofBaseballPlayersv1/Story?publish=yes> [Accessed 12 Sep. 2019].

[4] Public.tableau.com. (2019). *Tableau Public*. [online] Available at: <https://public.tableau.com/profile/xueyun.zhou#!/vizhome/StoryforthePerformanceofBaseballPlayersv2/BaseballPlayerStory?publish=yes> [Accessed 15 Sep. 2019].

[5] Public.tableau.com. (2019). *Tableau Public*. [online] Available at: <https://public.tableau.com/profile/xueyun.zhou#!/vizhome/StoryforthePerformanceofBaseballPlayersv3/BaseballPlayerStory?publish=yes> [Accessed 15 Sep. 2019].