

Project-8-Tableau-Story-Baseball-Data

November 6, 2017

1 Effects of Handedness on Baseball Players Key Performance Metrics

- [Link](#) the to story.
- [Link](#) to the revised story with feedbacks.

1.1 Summary

This tableau story is based on Baseball Data from 1,157 baseball players. Dataset parameters include their handedness (right / left handed or both), height (in inches), weight (in pounds), batting average, and home runs.

This report will investigate creation of visualizations that show differences among the performance of the baseball players.

From data inspection I decided that the key performance metrics in dataset are **HR (Home Runs)** and **avg (Battling Average)**. Therefore, I tried to collect data and draw visualizations showing substantial differences for these metrics between different *handedness* groups.

1.2 Design

I started with exploring the distribution of each variable in the dataset, and created histograms for each. Each histogram also displays contribution from each handedness group with a different color.

I also created scatter plots to compare multiple variable at the same time. Height vs. Weight and HR vs. Avg use scatter plots. I also used bar chart for displaying average HR for each handedness group.

1.2.1 Dashboard 1

This dashboard shows distribution of *Battling Average* and *Home Runs*. Each of these variables are continuous. In order to display visualization for each value as histogram, I created a new variables by binning the original ones. I also added *handedness* variable as a color filter which created stacked columns.

All charts automatically filter *handedness* variable on hover.

Upon feedbacks I dropped the stacked bar. Instead, I added handedness into *Rows* area. This divided histograms into distinct groups. Distinct groups would display y-coordinates dependently; therefore, I changed axis settings to show range independently from others. This assured that we can distinguish differences between groups of different sample sizes. I also added a bubble chart to top right corner to show the amount of different groups. This bubble chart also functions

as filter view. However, adding this view, left bottom view wider. Therefore, I transpose the graph to efficiently allocate space.

Bin Sizes

- **Home Runs:** 26.8
- **Battling Average:** 0.0172

1.2.2 Dashboard 2

This dashboard shows distribution of *Height* and *Weight*. These variables are also continuous. In order to display visualization for each value as histogram, I created a new variables by binning the original ones. I also added *handedness* variable as a color filter which created stacked columns.

All charts automatically filter *handedness* variable on hover.

Upon feedbacks I changed this dashboard too. However, the changes are identical to the changes in **Dashboard 1**.

Bin Sizes

- **Height:** 1.92
- **Weight:** 7.6

1.2.3 Dashboard 3

This dashboard shows scatter plots for **Height vs Weight** and **HR vs Avg**. These variables are also continuous. In order to display scatter plot itself, I also added player names as a detail in *Marks* area. The points in scatter plots are sized in according to number of entries in each group. For all charts *handedness* variable is added as color Mark. There is also a bar chart showing average *Home Runs* value for each *handedness* group.

For **HR vs Avg** scatter plot I also used a scaled HR variable which is the square root of original values. This helps to distribute otherwise dense data points.

Upon feedbacks, I checked weight and height variables, to see if they are indeed categorical. However, I saw that they are continuous. The values are biased to multiples of 5. Therefore, values are dense around values like 170, 180 etc. As a results, I did no changes in this view.

All charts automatically filter *handedness* variable on hover.

2 Feedbacks

I shared my Workbook with my friend Burak Mert, upon completion of the first version. I asked for what he thinks and what could I have done better.

2.0.1 Interview with Burak Mert:

Q-1: What did you think about data first?

- I wondered if we could understand what could effect performances of players.

Q-2: Were you able to understand what you were looking at?

- Yes. I was guessing the visualization would come to the conclusion of comparing different stats for each handedness group. However,I was only able to be sure about that at the last dashboard. I think this is due to the stacked bars being a poor choice to show this difference. You may want to change that into something else, maybe group them accordingly.

Q-3: How was the overall structure of the visualizations?

- I felt like it was too short. But after checking number of features, I saw that there is actually very few variables. This could be the reason. In any case, I think you can enrich the existing visualizations a little bit more. I am not sure though.

Q-4: What do you think I could be changing for better?

- I don't have any idea right now about the distribution of handedness groups. I think you could add a pie chart or another chart showing the distribution. This distribution could also be used as filter for first two dashboards. I am also not sure if the scatter plot for weight vs height is a right choice. Variables somehow seem like categorical. You should check if they are indeed categorical.

3 Resources

- Udacity Data Visualization in Tableau Course
- Tableau Tutorials
- Tableau Community pages