

p7-Baseball Tableau Story

August 12, 2018

0.1 Project 7- Create a Tableau Story - Baseball dataset

This is a project for Udacity Data Analyst Nanodegree Term2 Lesson 9.

0.2 Intro

This project investigates a data set containing 1,157 baseball players including their handedness (right or left handed), height (in inches), weight (in pounds), batting average, and home runs using Tableau.

The data used in this project can be found on this [link](#)

The goal of the project is to create a visualization that shows differences among the performance of the baseball players.

0.2.1 Questions

The following are some questions that I wish to answer through the visualization using Tableau:

Basics:

- What is the distribution of height and weight among players?
- What is the distribution of handedness among players in this dataset?
- What is the distribution of BMI among players in this dataset?
- What is the distribution of batting average among players in this dataset?
- What is the distribution of Home Runs among players in this dataset?

Effects of Handedness:

- What is the relationship between home runs and handedness?
- What is the relationship between batting average and handedness?

Relationship btw HRs and X:

- What is the relationship between home runs and batting average?
- What is the relationship between home runs and weight?

0.2.2 Final Visualizations

- The final tableau story adjusting from the feedback can be found on this [link](#).
- The first version can be found on this [link](#).

0.3 Summary

Here are some interesting insights I found in the data: - Left-handed players seem to be a great choice as higher batting average and lifetime home runs while most players are right-handed; both-handed doesn't go well. - Player with the highest lifetime home runs don't have the largest batting avg. Weight seems to have nothing to do with Lifetime Home Runs.

Here are something you should notice before dive in to the visualization: - This dataset contains players' height, weight, handedness, average batting rates, lifetime home runs. - Batting rate is counted as lifetime average and home runs are counted as lifetime accumulation.

0.4 Design

In this section, I will talk about the chart type, visual encodings I used for this Tableau story and why.

0.4.1 Scatter plot – Height-Weight, HomeRuns-Batting avg., Weight-HomeRuns

I chose scatter plot for these 3 plots since it's better for scatter plot to demonstrate the distribution between 2 variables.

Note that I placed "name" in color marks since compared to other marks like size, shape or detail, we can more easily tell the difference when the data points are close to each other.

0.4.2 Box plot – BMI, avg-batting, HomeRuns distribution

I chose box plot for BMIs, average batting rate and homeruns distribution since this gave me some basic infos for the data such as median, upper/lower hinge, upper/lower whisker.

Note that since average people don't have BMIs around 0-15, I actually zoomed in the data to only demonstrates datas that have BMIs between 18-32.

0.4.3 Bar chart – handedness, batting-handedness, HR-handedness

For handedness, batting-handedness and HR-handedness, I chose bar chart as visualization. This is due to the fact that one can easily compared the numbers of players within each categories – left-handed, right-handed and both-handed.

0.5 Feedback

In this section, I demonstrated my Tableau story to one of my best friend: Tony and iterate my Tableau story according to his feedback.

0.5.1 Here are the first and final version of my visualizations:

- The first version can be found on this [link](#).
- The final tableau story adjusting from the feedback can be found on this [link](#).

0.5.2 The following are the original feedback I collected from Tony:

What do you notice in the visualization?

The largest weight of the player is 111 kg. It seems that as the player has to run between bases there's limitation of their weight.

There might be some typos: batting-avg-Bar, HR-Bar -> Box.

The text box on the top is too small, you have to scroll in order to see all the text.

What questions do you have about the data?

What is the source of this dataset? there's no Derek Jeter or Chien-Ming Wang. Is it due to the dataset is collected in the old days or is there some data missing?

Home runs should be accumulated through out their lifetime, but there's no proper legends and the batting is counted as average. In the first glance, I thought home runs is also counted as average.

What is the relationship between weight and home runs? Does players with larger weight has larger amount of home runs?

What relationships do you notice?

Left-handed players has higher batting average and home runs this may also be a selection bias(Better player tend to practice becoming left-handed player as is better for base running)

What do you think is the main takeaway from this visualization?

If you want to gain better batting average, you should practice becomming lef-handed. Player with both handed on the contrary had worse total home runs.

Is there something you don't understand in the graphic?

What does the uppper, lower line in the box plot refers to?

0.6 Resources

Here are some resources that help we went through the whole project.

- In order to get to recap how Tableau works, I re-watch the video where Udacity has great [conversion with Trina](#).
- [Cannot save my workbook to Tableau Public](#)
- [Internal Error - An unexpected error occurred and the operation could not be completed.](#)