

Create a Tableau Story

Data Set : Baseball Data

Author : Varun Kumar Sharma

Date : 30 November, 2018

Introduction :

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. We will create a visualization that shows differences among the performance of the baseball players.

Links :

Story1 : https://public.tableau.com/profile/varun4612#!/vizhome/Story1_Review/Story1

Story_Final : https://public.tableau.com/profile/varun4612#!/vizhome/Story_Final_6/Story1

Summary :

In this project we will create data visualization for Baseball Data set using Tableau and we will analyze various variables such as batting average, handedness, player's height, weight and homeruns, mutual relationship of these variables and how they determine player's performance.

Design :

Initial Story

Firstly I checked the columns and rows included in downloaded Baseball Data set. I was curious to check the handedness(R,L or B) distribution in the provided dataset. For this I create a Bar chart using Number of Records vs Handedness.

I create Bar chart to check effect of Height and Weight on Batting Average of baseball players.

I want to find the players names with top batting average listed along with their handedness and batting averages. For this I create a Bar chart by filtering the batting max. Batting Averages.

Now I want to check the effect of Height and Weight on Home Runs. For this I created Area chart with maximum Home Runs counts vs Height and Weight.

Finally I create scatter plots analysing the correlation between following variables in context of "Handedness" :

- Batting Average vs Height and Weight
- Batting Average and Home Runs

Feedback :

I shared a link to my project in my Study Group in Student Hub and got following feedbacks

Feedback from Ahlam A.

It's a pleasure, as we see the amount of effort spent here! Thank you. First, you need to write a title for your story explaining the contents of the story. People look a lot at these details. Second, you use the area chart to represent weight and height, which is not true. The area chart is used only when you want to represent data over a specified time period such as months or years. This situation requires you to use Histogram to represent weight and height! Add at least a filter to give your story more interaction. Good luck!

Feedback from Jayashree S

Please avoid using abbreviations like R, L, and B. Please include units for the labels in the x-axis and the y-axis.

Final Version :

Actions Taken -

- I wrote a relevant title for my story
“Baseball Data Anlysis - Differences among the performance of the baseball players”
- I changed chart type from Area chart to Bar chart for HomeRun vs Hieght and Weight.
- I added Filters on Dashboard to make my story more user interactive.
- I changed Handedness abreviations R,L and B to proper names to avoid confusion.
- I included units for labels in x-axis and y-axis by including units in column names Height and Weight.
- I fixed the issue of duplicate names.
- I resized the captions and made them more explanatory.
- I put multiple related charts on a single dashboard.
- I linked multiple charts to a single filter in conclusion slide.
- Design recontructed with updated plots & additional details.

Final Story

Firstly I checked the columns and rows included in downloaded Baseball Data set. To make basic analysis of Datset I used pandas. There are has 1157 rows and 6 Columns in this dataset.

I fixed duplicate names included in the dataset and also gave proper names to column and abreviations.

Once the dataset is clean I started to use make visualizations using Tableau.

I was curious to check the handedness(R,L or B) and other varaible distributions in the provided dataset. I wanted to check the which handedness did well. For this I create a histograms for each variable.I chose histogtams because a histogram is visually interesting combination of a vertical bar chart and a line chart. The continuous variable shown on the X-axis is broken into discrete intervals and the number of data we have in that discrete interval determines the height of the bar.Hence, Histograms are great for illustrating distributions of our data.

Height and weight are major factors in determining player's performance therefore I want to explore them in different ways. I created Bar charts to check the player's height and weight with maximum batting average and home runs. I chose Bar Chart because bar charts are best for comparing data that is grouped by discrete categories.

I created Line charts to find the relation between height and weight as the batting average and home run increases/decreases. I used trend line to check the variability. I used line chart because line chart is best to visualize a correlation or the lack thereof between these three variables.

I created scatter plot to check the range of height and weight where players performance is better. Another scatter plot was created to check relation of Batting Average and Home Runs in context of handedness. I used Scatter plot because it is useful plot for quickly understanding if there is a relationship between two variables and of what type (positive, negative or no correlation)

I want to find the players names with top batting average and HomeRun in context of their handedness. For this I create a horizontal Bar chart by filtering the batting max. Batting Averages and max. HomeRuns. Horizontal bar charts are typically used when the number of categories/names is large (greater than 10 or so)

So I started from exploring Handedness, then how Height and Weight effects on players performance and correlation between themselves in various contexts, checked relationship between batting average and home runs and finally checked the top performing players in context of their handedness. This sequence I followed because first I wanted to explore variables with their distributions and then proceeding towards bit complex correlations. Like this I was able to make effective story which draws the interest step by step. Using drag-n-drop functionalities of Tableau, the user can create a very interactive visual within minutes. The interface can handle endless variations

Conclusion :

Players are mostly right handed in given data for **Handedness**. Left handed players have more batting avg. and Home runs as compared to others.

It is seen that players with weight between 175 to 230 pounds and height between 69 to 75 inches performs better than others in terms of **Batting average** and **HomeRuns**

From the scatter plot there can be seen slight positive correlation between Home Runs and Batting averages which is as expected. For majority of players who have More Home Run, more will be their batting average. But there are some cases where player with high batting average doesn't have high count of Home runs.

There is not clear correlation between Batting average and Weight/Height in given dataset when zero HR and Batting Avg. are included. When excluded then slight positive correlation can be seen to certain limits and then negative. So excess of Height or Weight does effects Batting average/HomeRun in small ratio in provided data.

Interestingly, even if the numbers of right handed batsmen are more but still Left handed batemen are higher in numbers in top 20 highest batting averages and they have better performances then others.