Baseball Dataset Exploration

Summary:

In This write-up report I will summarize a data visualization process using Tableau Public software and I'll tell a story and highlight some patterns in baseball data set by creating an explanatory data visualization from this data set.

This data set contains 1,157 baseball players data including their handedness (right or left handed or both hands), height (in inches), weight (in pounds), batting average, and home runs.

1st Publish:

https://public.tableau.com/profile/mahmoudesmail#!/vizhome/BaseballDatasetExploration/Baseball

2nd Publish:

https://public.tableau.com/profile/mahmoudesmail#!/vizhome/BaseballDatasetExploration-Vol_2/Baseball

(3rd Publish)Link After Modifications Required by Udacity Reviewer:

https://public.tableau.com/profile/mahmoudesmail#!/vizhome/BaseballDatasetExploration-Vol_3/Baseball

Design:

Design Decisions:

- I used various visualizations methods in this exploration to find out interesting patterns and relationships between variables in this dataset
  e.g of variables in Dataset:
  handedness, home runs, height, weight and batting average
  These were the most interesting variables to find relations between them as they help to understand the nature of baseball sport and how well performing players got recognized for.

- Some of the used visualization methods:
  line plot, horizontal bar plot, area chart, scatterplot
  they're all used in some way to be able to figure out the most of hidden patterns in the data by the help of some analytical tools like line trend that was used with some line plots
  and also the hover interactions were used in all of visualizations to make the reader able to read any piece of information from the graphs.
  These types of plots and tools are used because they're the most appropriate things to easily read information from the graphs and also they're appropriate for the variables and relations that we have in the dataset.

- I took advantage of visual encodings and colors at the the horizontal bar plot of "Players With Most Home Runs with Their

Preferred Hand" and "Most average battings by name and preferred-hand" the colors here made me able to see the difference between the three different type of players Left-handed, Right-handed or both-handed
also it helps the reader to draw comparisons between players.
At the other plots I chosen to standardize the color of linear plots and bar plots to the blue color and pale blue color to make it readable and joyful for those who suffer from color-blindness

- Story points are sequenced in a logical way that makes the reader able to understand granular information discussed in the visualizations

Findings:

1st Publish:

- I chosen bar chart to draw a relation between Sum of no. of records according to handedness
  this type of plots is easy and readable for any reader.
  the plot states that, the most of players in the sum of records are right handed, in the second place comes left handed players and then both handed players.

- In the maximum Average of battings vs. height of player plot:
  I used the line chart to discover the direction of trend In the data by help of line trend analytical tool
  and the trend states that at low heights the average of batting is moderately low until we reach the height 72 inch where the

average of batting reaches 0.3280 which is the highest avg. after this transition when the height of player increases the average of battings decreases gradually until reaches the spike at 79 inch height and average batting increases a slight increase and then re-decrease again.

- By using of bar chart plots, I plotted Handedness of player vs. Home Runs
  and handedness of player vs. average batting
  And it clearly seen when we look at the visualizations that the right handed players are dominating the highest ratings in Home runs and average batting.

- Right handed players are dominating highest home runs and batting runs on the horizontal bar chart, But that doesn't mean that the highest individual player who has highest home runs is right handed, so I plotted a horizontal bar chart to show to highest home runs of players according to their preferred hand and as I expected, the highest home runs was scored by left-handed player. though, the dominance of highest home runs is by right-handed players

- From the area chart, it turns out that the Highest Average of batting is 0.328 at weight of 170 pound,
  While the least average is 0 at the weight 177, 217 and 223 pounds respectively, There's always many fluctuations on the graph which means the large weight is not always an inference for a good performance and there's always some outliers.

- From the area chart, The players with the best Home Runs are accompanied by certain weights.

The Highest Home Run is at weight 190 pound, the second best lies at 195 pound and the third is at 180 pound

- From the line chart it appears that the players with best Home Runs are also accompanied by certain specific heights. from 72 to 74 inch is the best region for players with high Home runs. Best Home Run is at 74 inch

- Some players with high home runs has the same performance of players with lower home runs in terms of average battings, I scatter plot is used in this condition because it's a good way to communicate the comparison and to describe the behavior of highest home runs against average of batting.

2nd publish:

- I added another horizontal plot to my story to test if the highest batting average individual player is right-handed or left-haned and as well as the sum of home runs. the highest average of batting goes for a left-handed player. even though the highest batting average accompanied by right handed players.

- I added another plot of min average of batting. vs height by using line plot to easily read info from the plot, the information found was The minimum average of batting decreases as the height of player increases which states that highest performances are accompanied by certain heights neither high height nor low heights.

Feedback:

Suggestions are taken from a colleague who has been graduated from the data analyst nanodegree a long time ago.

- Suggest:
  Adjust the sequence to be more narrative

  Act:
  I changed the sequence of 'Highest Rates of home runs' and added it before that story that starts with 'Highest average of batting..'
  to be more sequential and more logical to a reader

- Suggest:
  Add another layer of analysis according to name and handedness

  Act:
  as I made a visualization for sum of human run by name and handedness, I realized that I can make another visualization to sum of average batting by name and preferred hand

- Suggest: add some description of what you are trying to do to the title of your work

  Act: I added a summary of the dataset to the title of the story to make the reader get more interested and to avoid the reader being over-whelmed for not understanding what the nature of the dataset and visualizations trying to figure out.

- Suggest: Try to add another measure or type of count to your plots instead of sticking to (sum)

Act: as I made a max avg. plot against height I will add another type of measure against height which is (min avg.) cause I realized that if I inversed the measure I may figure out any other hidden pattern which may interfere or not interfere with my investigation.

Summary:

The Vast Majority of players are right-handed
and the highest Home Runs and batting averages are dominated by right-handed players whom their heights fall between 72 & 74 inches and weight fall between 170 and 180 pounds. but the highest individual home run was scored by left-handed player

Resources:

- https://public.tableau.com/en-us/s/resources
- https://www.theinformationlab.nl/2017/08/02/tableau-containers/