Data visualization of Baseball:

Links to the Project:
https://public.tableau.com/profile/alivelu#!/
https://public.tableau.com/profile/alivelu#!/vizhome/Aliveluinitial/Story1
https://public.tableau.com/profile/alivelu#!/vizhome/Alivelufinal/Story4

About:
Baseball game was one of the favorite games in the past. The important measures of player performance in baseball game are batting average and home runs hit.
The data is in CSV format. Open the CSV file in Tableau to understand more about variables. Belief is that more weight and taller the player, he can play best.
Let's see whether it's correct or not with this data.

Summary:

A total of 1157 rows are there. That means 1157 entries.
7 columns are present. That means 7 variables.

They are:
1. name
2. handedness
3. height
4. weight
5. average
6. HR
7. Number of records

Let's understand the variables:

Name: Name of the player
Handedness: Left-handed, Right-handed or both
Height: Height of the player
Weight: Weight of the player
Average: Batting Average
HR: Home runs
Number of records

Main focus:
With the help of data visualization, I want to understand the relationship between the performance of baseball players w.r.t to different variables mentioned above.

Design:

1. I have used Gantt bar chart for understanding relationship between a variable, player name and count of players.
Ex: How many persons are having the same height or weight and if possible their names

2. Used bar charts to visualize the distribution of different variables.
Ex: Players having more weight and their name

3. Packed bubble chart to know how many players belong to same weight category

4. A line chart to show the name of the person having the lowest and more weight and their names.

5. Pie charts to understand left handed, right handed, both handed and total number of home runs according to handedness.

My understanding from histograms:

2. As handedness is having 3 categories(left, right, both), I have used bar chart for the distribution of home runs based on handedness

3. Scatter plot used for the relationship between weight, height and home runs.

Feedback:

What do you notice in the visualization?

Ans:
I noticed different data graphs about base ball players like whether the players are left handed or right handed or both, how many home runs players had etc...

What questions do you have about the data?

Ans: What is the "average" graph in story3 is about?

My Ans: Or what changes I did:
I have added the following in the story section

These 7 variables are used here.
1. name
2. handedness
3. height
4. weight
5. average (Batting average) or avg
6. HR (home runs)
7. Number of records

What relationships do you notice?
Ans:
The right-handed batsman had more home runs than the left-handed batsman

What do you think is the main takeaway from this visualization?
Ans:
Easy representation of complex data in a quickly readable format

Is there something you don’t understand in the graphic?
Ans:
weight vs count of weight graph

I have added a note as following in the storyboard.
Note: Weight vs Count of weight graph is related to the total count of players with in a weight range.

My Findings:

Expected: Belief is that more weight and taller the player, he can play best.

Actual:

Reggie Jackson tops in home runs. More players are between batting average of 0.225 and 0.275. Very few players are having average home runs in between 300 and 563.
Rod Carew is having a batting average of 0.328. You can clearly see from the below graph (bottom graph) that 3 bins are having more number of players.
Dean Chance is having the lowest batting average of 0.066

From the top graph right-handed bats, men are more in between the batting average of 0.225 and 0.275.
From the bottom graph, we can see that Right handed bat's men are more in number for home runs.
Reggie Jackson had 544 home runs is left handed. Rod Carew had a batting average of 0.33 is also left-handed.

Sandy Alomar is having the lowest weight of 137 pounds with home runs 13.
Stefen Wever is having the weight of 243lbs and home runs 0. Between 220 lbs to 243 lbs, players have 0 runs.
Lesser the weight better the batting average.So I can say heavy weight means low performance related to batting average. Players of height in between 69 in and 75 in are having more number of records.
With the increase in height, batting average is decreasing but few exceptions are there. Freddie Patek is shorter in the team, i.e 63 in. 7 players are more taller than other players i.e 79 in.

My understanding from the first graph is home runs(hr) is independent of height and remaining variables. The number of records is increasing in between 63 to 73 inches.
From 73 to 79 in there is a downtrend in the number of records.

From second graph:

Weight is also independent of home runs. But from 130lbs to 175lbs number of records keeps an increase in trend. From 175 lbs to 182 lbs there is a downward trend related to number of records.
Again from 182lbs to 190lbs there is an upward trend in the number of records. From 190 to 243lbs there is a downward trend in the number of records.

From the third graph:
Players having batting average of 0 bin are having more total number of records. But players in between the 3 bins of 0.225 and 0.275 are having more number of records than other players.
I feel batting average and home run average are also not showing any relation.

Improvements:

Added a note for the shortcuts of variables for a better understanding

Resources:
https://www.youtube.com/watch?v=gWZtNdMko1k&list=PLWPirh4EWFpGXTBu8ldLZGJCUeTMBpJFK

https://www.youtube.com/watch?v=jj6-0cvcNEA

https://www.tutorialspoint.com/tableau/

Google for knowing and understanding baseball

Mentor Dylan Lennard from Udacity