Data Visualization in Tableau Project

Project name: Height - Weight Relation to Baseball Batting Average and HR

Important Links (Tableau links may require copy and paste):
Initial Draft: https://public.tableau.com/profile/f.hz#!/vizhome/Besbol_Fevz/Story1 (Tableau links)
Final Draft: https://public.tableau.com/profile/f.hz#!/vizhome/Height-WeightRelationtoBaseballBattingAverageandHR-FInal/Story1
Original Dataset: Baseball Data (original)
Final Dataset: Tableau\baseball_data_final.csv

Summary

Based on the baseball data, I performed an explanatory analysis on the historical baseball data obtained from the Udacity dataset link. My main research question is: Are height and weight related to baseball performance as measured by batting average (Avg) and home run (HR) career total? Despite not finding any solid evidence supporting this relationship, I found that data visualization can be used for a quick data quality check, identifying players with best (and worst) statistics, and how the two metrics stack up as baseball performance indicator.

Design

I focused on enriched scatter plots and histogram mainly due data limitation and their strengths to show multivariate relationship. There are not many variables to begin with in the dataset, but there are opportunities to transform the variables by having log transformations and categorization based on established BMI standards.

I know that Avg and HR are the desired outcomes. Naturally, I'd have the height, weight, and handedness as explanatory variables. When height and weight did not show interesting relationship, I combined both measures by creating a BMI categorization, a standard measurement to measure underweight, normal weight, and overweight. When I couldn't find observable relationship, I further categorized the three-category BMI scale into more categories, hoping to find interesting relationships.

In my beginning stage of exploring the data with visualization, I discovered that many of the listed played have 0.000 Avg. That did not make sense because professional players should have at least 10% of batting average in my opinion if they were truly major league baseball players. Based on research a small sample of these 0.000 Avg players, they all tend to be pitchers or spend minimum time in the major league. Pitchers tend to be taken out from batting because of their tendencies to be weak in batting. So, they tend to be substituted by pinch hitters during batting. Based on this information, I filtered out the 0.000 Avg players for the visualization. Besides this, as noted in my presentation, I found three players with the same name, which means that I have to exercise caution for analysis and reporting.

With regards to Avg and HR, I looked into their relationship. If they have close to perfect relationship – where linear or linear – I would focus on one of the two performance variables and extend the summary to the other variable. Obviously, a multiplier or its transformation is needed if I discovered a non-linear Avg-HR relationship. Another option with the variables of interest is to create an index-like score that takes into account Avg and HR. Unfortunately, my research was not fruitful into arriving into such score.

Since Avg and HR was not strongly related, analyses performed with one variable had to be replicated with the other. This is the reason for having similar scatter plots done separately on Avg and HR. In addition, the histogram was stratified into Avg and HR groups while taking into account handedness. I was hoping to *dim* (i.e. make the bars light colored) the histogram bars for subcategories with less than 10 players, but I couldn't find a way to do so in Tableau. The suggestion that I found regards to filtering these results out.

Response to Feedbacks

Feedback 3 contains the juiciest suggestions. With regards to his "**layout wasn't optimized**" comment, I went back and altered the story layout with either automatic or 1000 x 800 pixel layout. Since reviewer 3 commented "story tabs/headers and content seem a bit wordy", I went with a **simple story header**, similar to what he utilized in his Tableau story presentation. Instead of having one-sentence headers for the story, I made the header box smaller by limiting each header box to 3 words or less to focus on the dashboard's main ideas. I also include a **list of reference** in the presentation. From Feedback 4, I included some "**call outs**" to bring reader's attention to the specific section of interest in order to better convey the intended story or message.

All the four reviewers commented about the **color choice**. I went with Tableau 2.0 template. Since both handedness was the least of the three, I chose gray to be its color because it tends to be a neutral color. For right- and left-handedness, I chose contrasting colors, blue and orange. I also changed the default grey story background to olive color.

Feedback

Feedback 1

> Good work on the visualization. Is there any way the data could be aggregated into categories to make the data more accessible to the eyes? For instance it appears generally, that 'heavier' players tend to hit more home runs. Is this a valid perception based on your data? Does a player's age (or years in the pro leagues) play a role in BMI or (H & W) ratios and by extension, to HR or batting averages?

Feedback 2

> Good work, Fahzy. It will be easier for me to see the difference if you can change the color to clearly distinguish the left players from the right ones in your scatter charts.

Feedback 3

> Have you read Cole Nussbaumer Knaflic's, Storytelling with Data? This book helped me better understand the importance of visual design — layout, graphics, text, color, line weights, white space, etc. — to draw attention where you want it and tell a story. I picked up mine as part of a grad class on descriptive analytics / visualizations (and Tableau) and still refer to it frequently. I think you might enjoy it and find useful information throughout. Cole's book is at amazon (link) and she has a blog (link) that is insightful as well. It's a great reference for visual design of communications pieces.
>
> Good work on your visualization. Tableau can be a challenge, especially with the various layout options. What stood out to me at first was the layout wasn't optimized for a standard laptop monitor resolution or with a dynamic layout. This caused scrolling on some pages but not others. Sometimes that's necessary but helpful if it can be avoided. The conclusion seems to be on page one and in great detail rather than letting the audience come to the conclusion your data should drive them to. Sometimes that is necessary, but it didn't seem so in this case. It wasn't immediately clear what the topic was without a fair amount of reading. This may seem trivial, but a simple header with what the topic is would be helpful. I see it in the browser window name, but this is something that should pop out immediately to your audience, *imo*. The story tabs/headers and content seem a bit wordy. I get the sense that you found a story to tell, but are walking us through the exploratory analysis rather than focusing on the explanatory analysis (the key findings). It would be helpful if each slide/tab had say one or two main ideas and you focus us on those. If those charts are ideal for the type of info you are trying to convey, then the use of color to highlight the particular class/group in question and its position relative to the other data points might help highlight your takeaways for the slides. The color in the 2nd tab when it first appears doesn't have a legend so it's not clear what it means until the legend is seen on the 3rd tab. Would be helpful to have that legend any time, or at least the first time, it is used. The scatter plot marks could be smaller to help see more nuance as the data is heavily overlapping. Some log/nlog transformations might help spread out the ranges a bit as well. I couldn't tell if color was necessary in the context used, as it didn't seem like batting position (Left/Right/Both) was relevant to your objectives and conclusions. If you find it wasn't a distinguishing characteristic, it would help clean things up a bit to be rid of them.

The last thing I would note is that when dealing with data, sometimes it is helpful to cite your sources so people know how the data was obtained if they want to investigate further. A citation/reference at the bottom of the page or at the end may be helpful in that regard. I suppose a guiding principle would be that the data should be able to tell the story with as little narrative as possible.

For what it's worth, I have found that it's helpful to look at other people's visualizations for ideas on what aesthetics you may like and may like to try out. To that end, Tableau maintains a viz gallery, "viz of the day," where interesting work is often highlighted and published, located here. Also, so that this isn't a one-way road, here's the link to my previous work for your review / critique: https://public.tableau.com/profile/robert.fritts#!/vizhome/AGrowingEpidemic-SharpRiseinOpioidDeathsHitsEveryDemographic/Story

Hope that helps. Let me know if you have questions on any of the items noted above.

Keep up the good work. It's fun stuff to play with.

Feedback 4

My recommendations are more cosmetic:

Are you able to make the circles smaller on the first viz page? Might make the viz stand out

On viz page 3 I am just having a hard time understanding – maybe include some call outs? I forget if Tableau allows

Interesting story especially with baseball players where it seems like your weight doesn't necessarily mean you are a good or bad hitter!

Resources

A Growing Epidemic - Sharp Rise in Opioid Deaths Hits Every Demographic: https://public.tableau.com/profile/robert.fritts#!/vizhome/AGrowingEpidemic-SharpRiseinOpioidDeathsHitsEveryDemographic/Story

Download Lahman's Baseball Database: http://www.seanlahman.com/baseball-archive/statistics

Gary Lance: https://www.baseball-reference.com/players/l/lancega01.shtml

Jerry Ujdur: https://www.baseball-reference.com/players/u/ujdurje01.shtml

Maintaining a custom color palette on dimensions in Tableau https://tableauandbehold.com/2015/08/24/maintaining-a-custom-color-palette-on-dimensions-in-tableau/

Nardi Contreras: https://www.baseball-reference.com/players/c/contrna01.shtml

Pinch hitter: https://www.baseball-reference.com/bullpen/Pinch_hitter

Udactiy Beginner Data Sets: https://docs.google.com/document/d/1w7KhqotVi5eoKE3I_AZHbsxdr-NmcWsLTIiZrpxWx4w/pub?embedded=true