Create a Tableau Story

Summary

I studied a dataset containing 1,157 different baseball players' physical attributes (handedness, height and weight) and some of their baseball performance statistics (batting average and total home runs). In my visualization, I looked at the impact of the combination of the physical attributes on a player's performance stats.

Original:
https://public.tableau.com/profile/elisa.marchione#!/vizhome/C939_16008729177380/Story1

Final:
https://public.tableau.com/profile/elisa.marchione#!/vizhome/C939-Final/Story1

Design

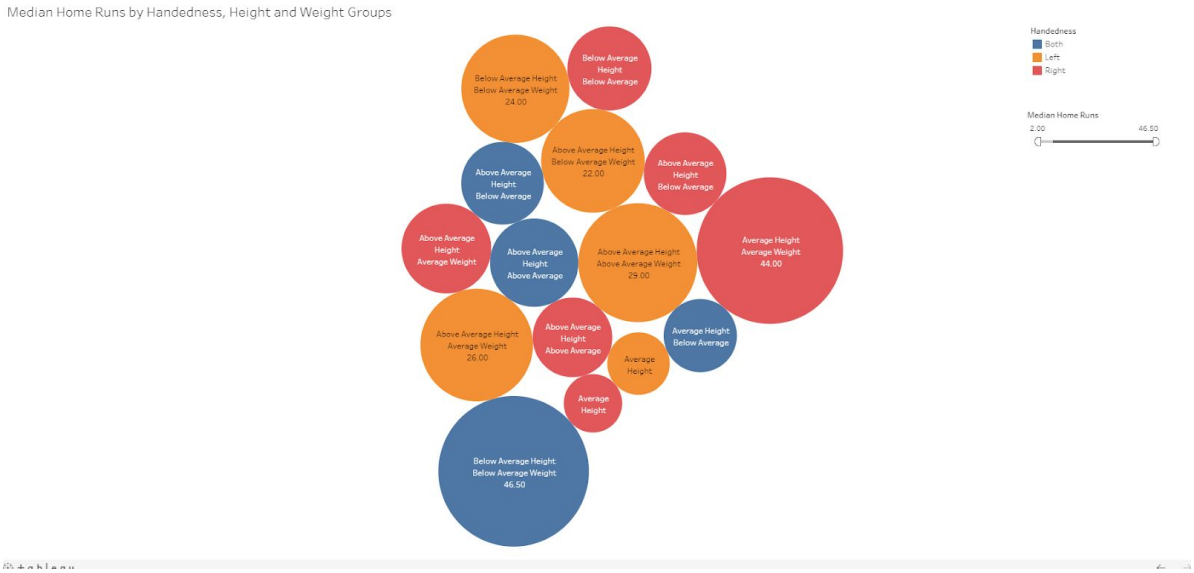
The design of my data changed quite a bit as my understanding of the data evolved. As I worked through the project more, I realized that I was still taking a more exploratory approach, as opposed to an explanatory one. I started out with way too many visualizations that showed the evolution of my thinking about the data. I started out with a bunch of histograms, as well as univariate, bivariate and multivariate charts. I honestly didn't have a concrete explanation that I was going for; that didn't end up coming until I'd had a few iterations of the project.

I started out with a lot of different charts, and the approach was too complicated. When I went to show my first iteration for feedback, I found that the reader was completely confused and did not get any sort of feel for the data that I was trying to explain. I ended up having to explain everything out loud, which was obviously an indication that my visualizations were not explanatory on their own.

My initial visualization ideas were to try and illustrate a relationship between height, weight and home runs. I took the approach of bucketing the height and weight values into their own categories ("Below Average", "Average" and "Above Average"). Then I set up some complicated visualizations that showed the combination of these categories in relation to the median number of Home Runs. The chart that I was the most happy with was one that showed all the different combinations of height and weight categories, and used size and labels to indicate the best values for Home Runs. I decided to name the labels with the full value of the category name ("Above Average Weight", for example) in order to remove any sort of confusion about what the category was.

I felt like there was a clear pattern that was indicated; however after I got feedback, I realized that it was much too complicated in scope.

As part of my first iteration, I also decided to do a bubble chart, which seemed very fun and energetic. I took the data from the above scatterplot and showed it in an interactive bubble chart.



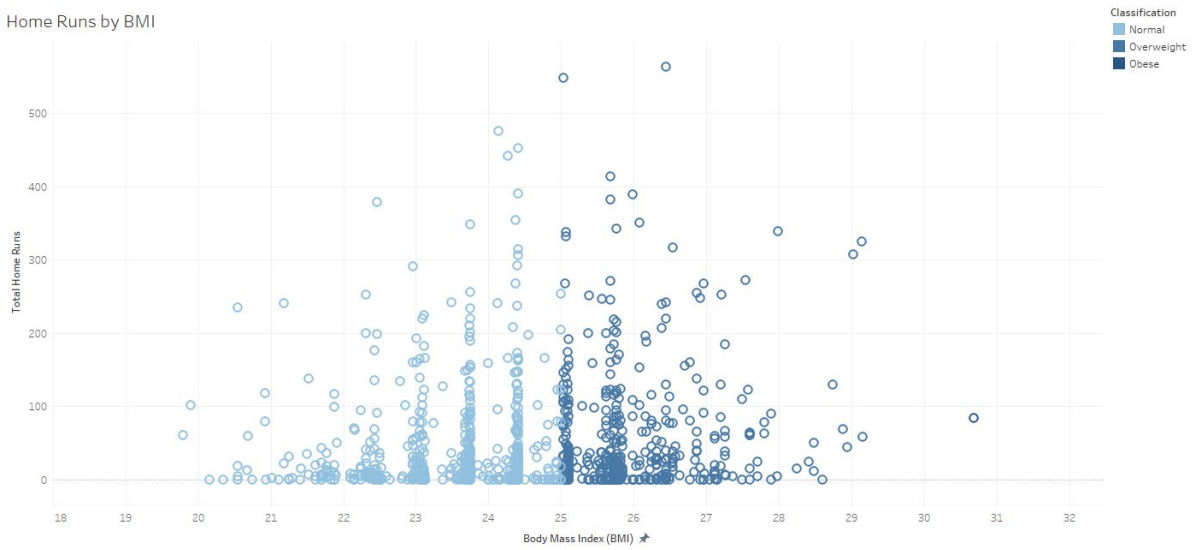
I liked the use of color for “Handedness” since it felt like a distinct way to show that categorization. I added an interactive filter for “Median Home Runs” so the user could narrow down which combinations were the best or worst.

The feedback I received made me realize that I should be even more clear and concise with the use of color. For example, in a different iteration with “Handedness”, I ended up using a color

scheme that made more intuitive sense. I tried using blue for “Right”, yellow for “Left” and then green (which is a combination of blue and yellow) for “Both”.

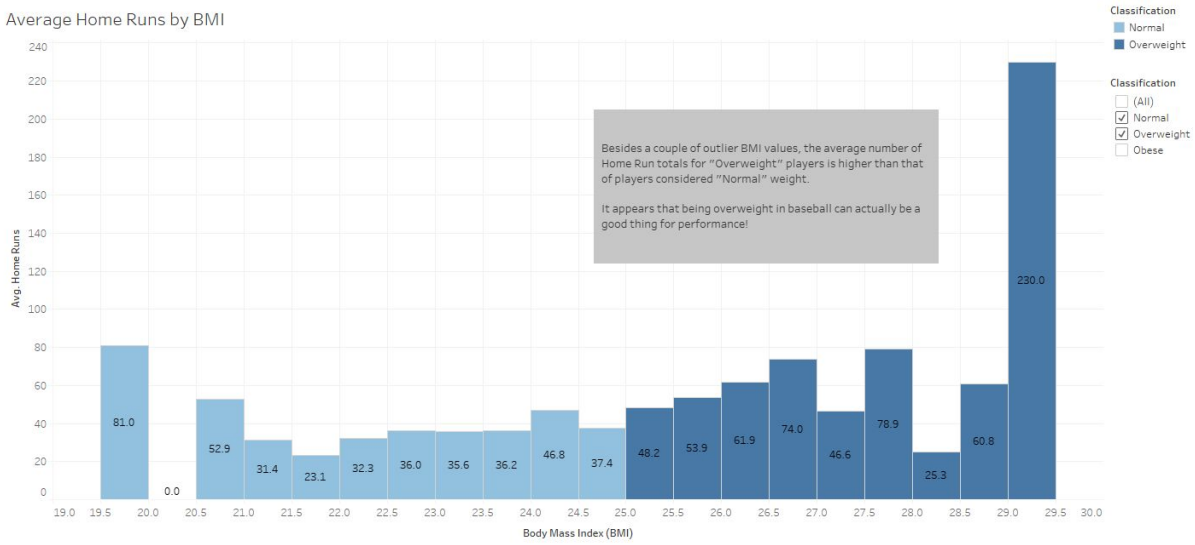
After receiving feedback, I realized that I really needed to simplify my approach. I ended up going back to the drawing board and almost starting over. I ended up finding a much better approach for the height and weight attributes, researching the calculation for Body Mass Index (BMI), and using that as my main physical attribute to study.

From here, I was able to create much simpler graphs as a basis. I created some histograms to take a look at the distributions of BMI. I also was able to create a nice scatterplot of BMI and Home Runs. With the use of color to indicate the BMI classification (“Normal” vs “Overweight”, for example), I was able to quickly illustrate that there were a lot of baseball players that were overweight and had high values of Home Runs.



I used the same color scheme throughout the entire story to have a unified understanding of the data. I also used feedback and set up the color scheme in a way that made sense. For the progression from “Normal” to “Overweight” to “Obese”, I used darker and darker blues.

My final graph is supposed to illustrate that baseball players can be overweight, and that it actually seems to contribute to a better performance when it comes to Home Runs. I used the unified color scheme to indicate BMI “Classification”. I also decided to label the number of “Average Home Runs” in order to fully describe the data effectively. I added interactivity with the “Classification” especially for the “Obese” data point because it was an outlier.



For my story, I used the feedback provided and simplified the approach. I made my chart titles much simpler, and found a different navigation method (using the “Numbers” instead of the “Caption Boxes”. Since I got rid of the caption boxes, I added “Annotations” to help tell a simple story on each page. I cleaned up the wording so it wasn’t as much of an exploratory story, but an explanatory one.

Feedback

- Move the storytelling portion down or to the side if possible. I keep missing it at the top of the page, and feel like I have to go back to read it later. Can the pages of the story be numbered or colored to show progression?
- Titles for the pages are way too long and not very engaging. Does it have to say histograms, for example?
- Explanation of data - missing things like units of measurements. What does batting average mean?
- Less descriptive labelling - less duplication. For histograms, could put something like Player Count or Players
- Don’t use the word “bins” or other terminology that’s not understandable to the layman
- Histograms side by side? They’re too long - makes it seem like I’m trying to compare the values
- Don’t talk about suspicions - talk more like a non-specific entity. Exploratory instead of explanatory
- Side-by-side for handedness is confusing. Looking for left-hand stats on the left side. It’s confusing without some sort of separation between the two graphs.
- For batting average bubbles - the text is so tiny. It’s not a great presentation because the circles are so close in size.
- Color of text is different on the bubbles - can you find a better color scheme?

- Change the sort order of the bubbles - left on left, right on right, both on bottom
- Maybe “Ambidextrous”, “Right hand”, “Left Handed” instead for aliases
- Remove the trendline - it’s confusing
- Speak more confidently and engaging - not just talking about decisions. Write in a more engaging, clear way - getting to the point faster. Instead of “There appears to be a correlation”, use something like “A correlation appears between …”
- Don’t use first person to talk about the data. Speak more professionally and confidently.
- Count of players in height and weight - use line histogram instead of bar chart histogram?
- Abbreviate Average to “Avg” - there’s too much going on, so using this might help alleviate some of the clutter.
- Swap order of heights so it’s in correct order.
- Dual Axis graph is confusing. There’s a lot going on with it. Hard to get a feel for what the story is that’s trying to be told. Nothing stands out.
- Last bubble chart - way to use symbols to indicate the type of height and weight?
- Get rid of decimals - the numbers are too long.

Resources

- https://en.wikipedia.org/wiki/Batting_average#:~:text=In%20baseball%2C%20the%20batting%20average,%22batting%20three%2Dhundred.%22
- https://www.healthline.com/health/average-height-for-men#u.s.-height
- https://www.healthline.com/health/mens-health/average-weight-for-men#:~:text=How%20much%20does%20the%20average,(about%2069.1%20inches)%20tall.&text=As%20time%20wears%20on%2C%20American,in%20both%20stature%20and%20weight.
- https://help.tableau.com/current/pro/desktop/en-us/datafields_typesandroles.htm
- https://www.cdc.gov/nccdphp/dnpao/growthcharts/training/bmiage/page5_2.html