Aaron Judge Visualizations

Group 2

By: Jackson Leech, Gavin Vazquez and Xavier James

Video Link:

https://drive.google.com/file/d/1mFp16Mu7K6hAbZSZNJs9aIjAg3A_KoW1/view?usp=sharing

Website:

https://xjames1313.github.io/CPSC4030_Aaron_Judge_Visualization/

GitHub Repo:

https://github.com/Xjames1313/CPSC4030_Aaron_Judge_Visualization/

Overview and Motivation:

Our idea for this visualization was inspired by Aaron Judge and his quest to break the American League record for home runs in a single season. Our group was invested in his greatness and we wanted to make a visual representation of just how legendary his season was. In doing so, we realized that his home run statistics were only half of the story. We wanted to show his entire career progression through our website. Furthermore, we wanted to show his failure through his quest in home runs by displaying strike outs. To be a great home run hitter, you must try to hit homeruns; trying to hit homeruns comes at a cost when you look at his number of strikeouts. In doing so we also wanted to analyze where he physically hit most of his home runs in the strike zone. We realized the combination of strikeouts, homeruns, strike zone heat map and hits that were not home runs would tell the full story of his season. We want a holistic view of his historic season, not just his homerun accomplishment.

Overall this project seemed simple at first, show each of his home runs in the MLB lifetime and we would accomplish our goals. Once we dove into that we found so much more worth showing about his journey to that peak. Our goal then became to make our website display his entire profile this season and give the viewer a more indepth look at what he did.

This project was difficult for us because we started with one goal and realized slightly late that we needed to do so much more for the viewer. It is not just an accomplishment that you write an article about. We wanted someone to see everything over multiple months of work in one minute of viewership. Our data expanded as our goals expanded and we found ourselves at the drawing board many times. Instead of starting big and working incrementally towards that goal, we started small and worked towards future goals. This involved a lot of failure we did not

anticipate because once we accomplished a smaller goal we hit a crossroads of making that visualization as good as it could be versus making another visualization to accomplish a different task.

Questions:

- How many homeruns did Aaron Judge hit in each of his seasons and where did they land geographically in the ballpark?
- How many strikeouts did Aaron Judge have in each of his seasons compared to how many home runs per season?
- Where in the strike zone did Aaron Judge make all of his contact in his career?
  his career?
- What is Aaron Judge's average exit velocity of hits in the strike zone?

Exploratory Data Analysis:

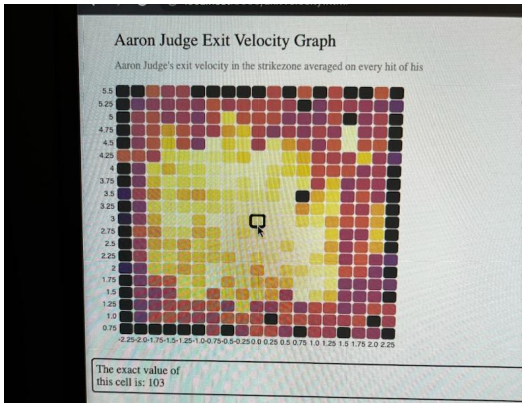
Overall, our data was pretty straight forward, we took homerun data from the baseball statistics website in csv format and began to dice through the data in excel in some basic graphs. Luckily, we had a lot of knowledge on Aaron Judge's season and career to the point where we are able to visualize how we wanted to showcase his different achievements. We gained a lot of insight into how historic his season was especially when we compared it to other players in the league which showed just how dominant he was. When looking at the dominance we realized we should highlight him compared to other greats in baseball to showcase his statistics. The overall numbers were somewhat easy to come by; we just had to look in the right places. We did have some data we had to derive: exit velocity, home run landing position, and total number of

contacts. It took a little time for us to turn the raw numbers we had into useful statistics but once we did it was easy to manipulate the data to show what we wanted the viewer to see.

Design Evolution:

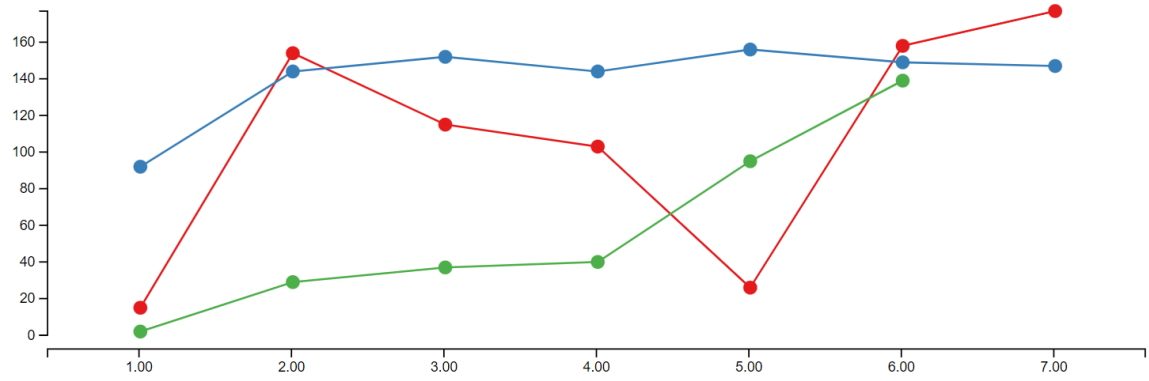
There were a variety of visualizations we looked at. The first thing that came to our head was to create a baseball field where we show the locations of all the homeruns hit. We tasked ourselves with creating a visualization of an actual baseball field and placing the points as an overlay on the field. This was challenging to accomplish with our knowledge of creating visualizations so we had to deviate into a graph that shows the same thing without the field. This still showed the distances and positioning to the user well. Hits in the strike zone was something we wanted to showcase next so that our viewers can see the hot spots for his contact with the ball. We stuck with our original idea of using a heat map to show the user how often different areas of the strike box were contact hits from Aaron Judge. For exit velocity, we went through a few different models that we wanted to display. The first one we looked at was trying to show different sized flaming baseballs to show the ranges of exit velocity. Although this would have been great for the theme we realized it would be much better to show how the location of the where the ball was hit affected velocity. When we completed this visualization you could see much easier how the locations allow for better full contact with the bat leading to higher exit velocity. Next we wanted there to be a way to show Aaron Judge's home runs and strikeouts throughout his career. We completed this chart and actually kept it for a long time as our visualization but had the idea to show this compared to Barry Bonds and Babe Ruth, two of baseball’s greats. This gives more depth to the information shown and we felt as if the user would appreciate the comparison across the decades more so. Creating this allowed for us to still

accomplish our original goal while showing even more information for the viewer. Overall I would say we deviated from the original proposal in a good way because we were able to show even more information in a visually appealing way compared to the proposal. As shown in our journal, this was one of our designs halfway through.

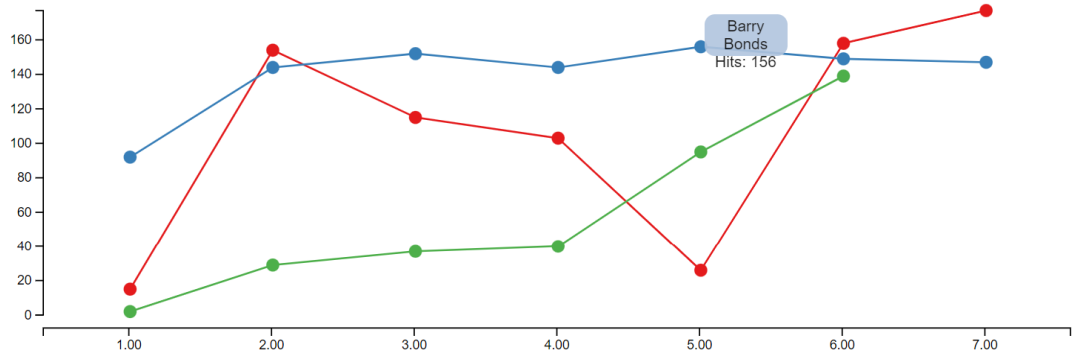


Implementation:

Player Comparison

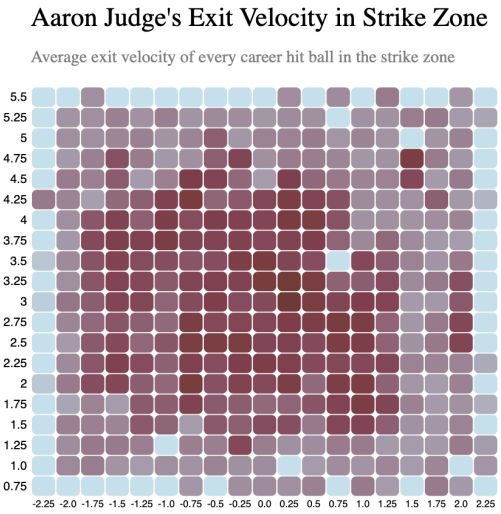


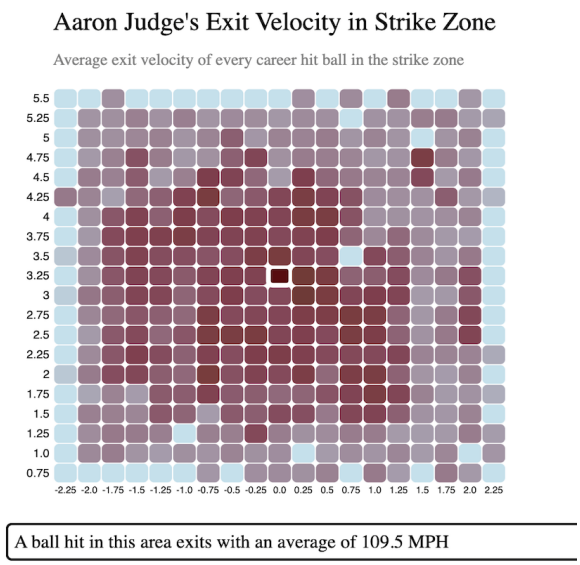
Player Comparison



- **Home run comparison with Barry Bonds, Babe Ruth, and Aaron Judge**

This visualization is used to represent the comparison of homeruns hit by three of the greatest home run hitters of all time. We used three different colors to represent the three different players. The tooltip part and interaction of this visualization shows the number of homeruns hit at each data point or year they were hit in their respective careers. It also displays clearly who hit the home runs the cursor is pointing to.

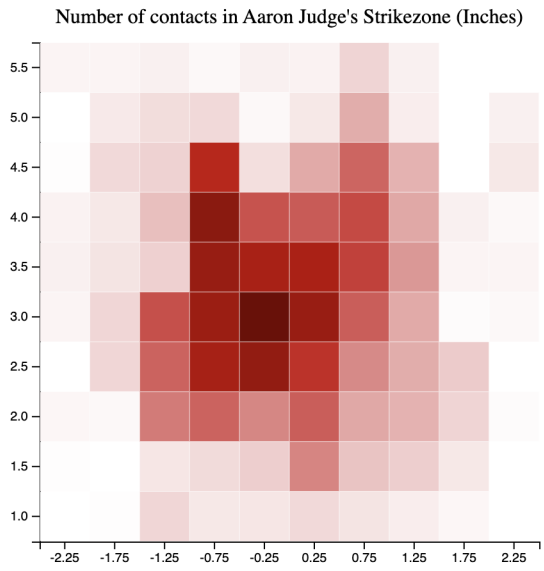


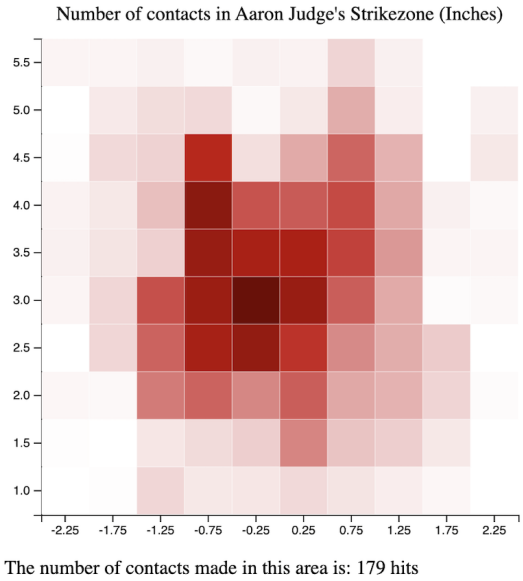


- **Exit Velocity:**

For this visualization we wanted to show a heatmap of where Aaron Judge hits the ball the hardest. Our visualization uses a blue to red color scheme to represent the “heat” of the ball off the bat. The darker the red, the faster the ball comes off the bat. This visualization also uses Toolkit to display the average velocity of the baseball's hit in the particular zone. As your mouse moves over each box, the mph of the hit is displayed at the bottom. We used a white outline so the user could see which box they were selecting, black seemed to make the visualization seem less innovative than white. The X and Y axis are in inches, meant to represent the average strike zone area.

- **Strike Zone Contact Heatmap:**

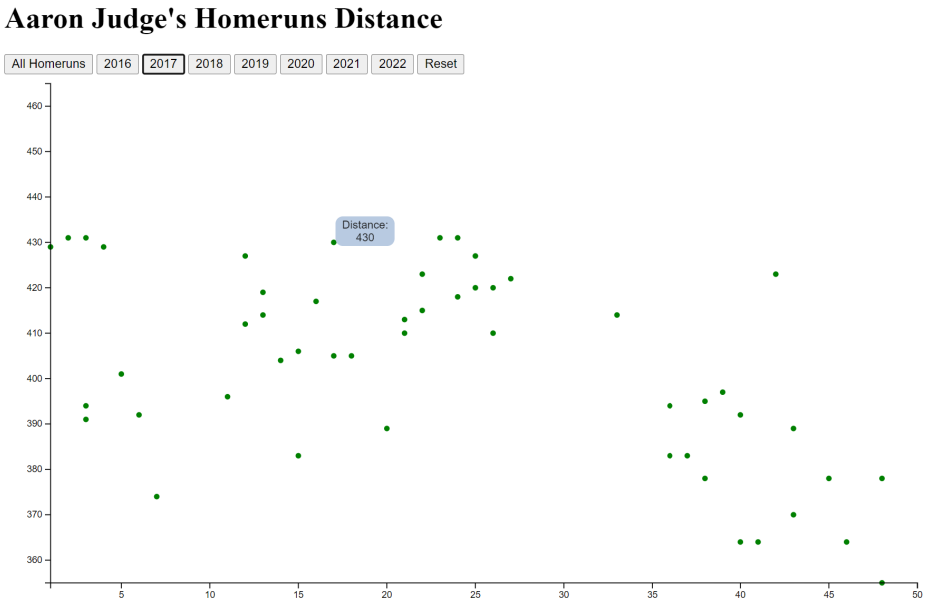


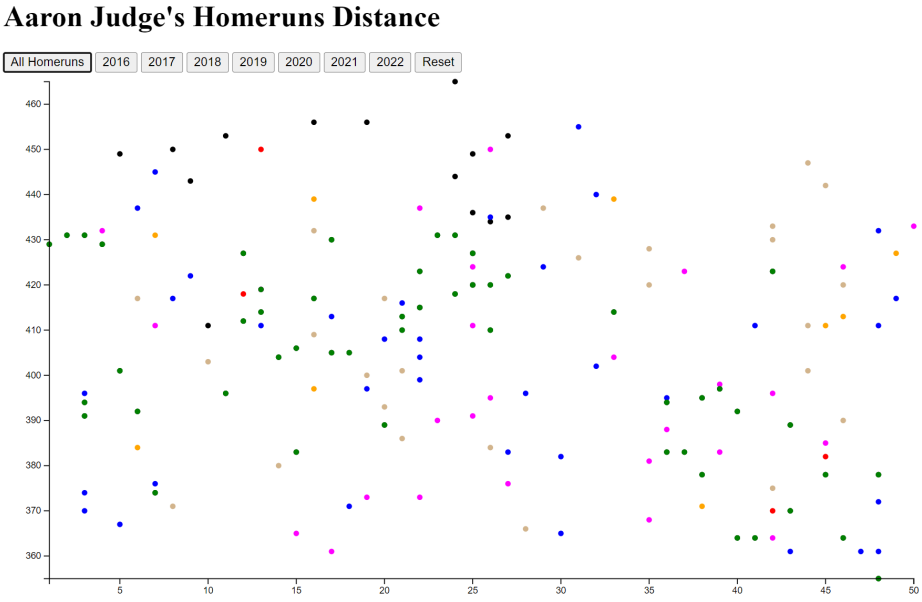


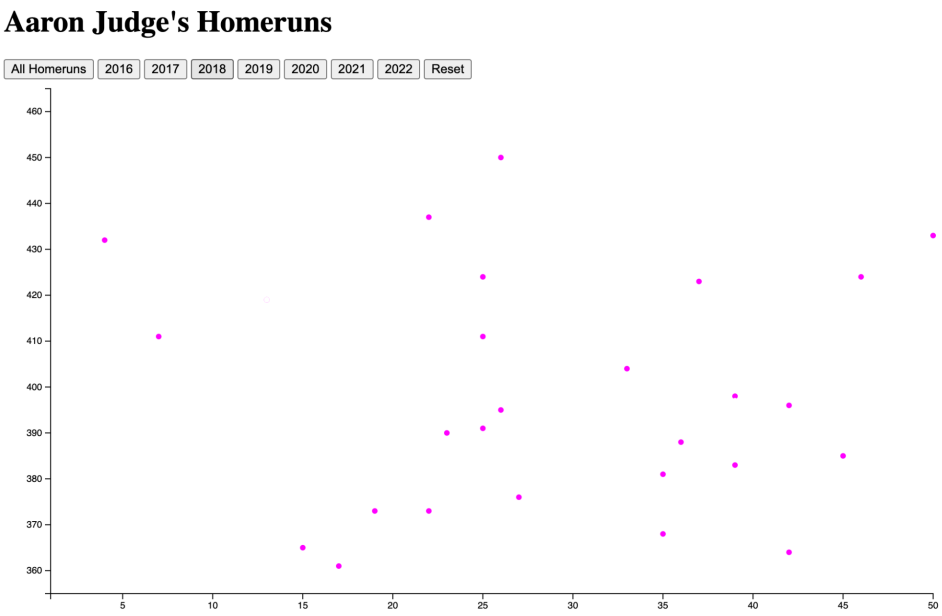
The number of contacts made in this area is: 179 hits

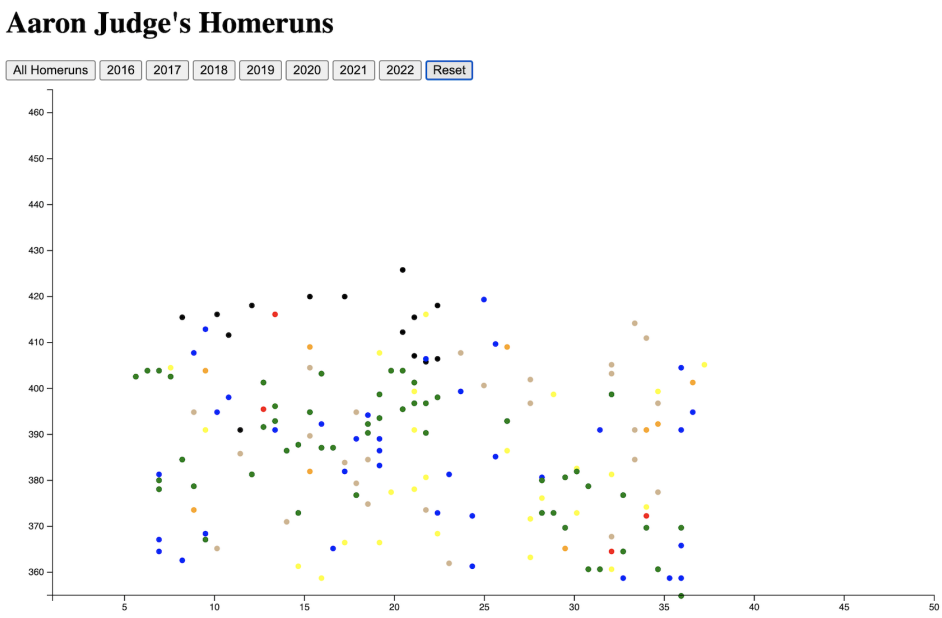
This visualization shows the number of contacts in the strike zone over Aaron Judge's career. We used white and red here to show the visual representation very clearly. The more hits you have the darker the red square. When you mouse over the box your cursor is on, it displays the number of hits made in that zone of the strike zone. We intended for the user to be able to look at this visualization and know what they were looking at relatively quickly. The X and Y axis are the same as the exit velocity visualization; the average strike zone of MLB players.

- **Home Run Landing position and distance**









This visualization was by far the most organic one we created. With almost zero examples of anything like it, we set out to make a visualization of all 206 homeruns of Aaron Judge's career. We had to gather data on all the positions and distances of all of his home runs. Once we had that data, we divided it into years, so we could display year by year and show the viewer his progression over time. The difficulty with that was the different data sets and how to incorporate

data from all the years of his career in one visualization. We also wanted the visualization to look like a baseball being hit off the bat so our animation has the balls “flying” from the center of the screen to their respective zones. The reset button was an absolutely crucial feature in this one. In order to properly show data and take it away we needed a way to make the data “go back” to the home plate. We also used a toolkit to display the data of distance and position once we moused over each data point.

Evaluation:

A lot of our data was very straight forward but when we visualized the exit velocity and hits in the strike zone we found very very interesting findings. The main finding being that the right side of the strike zone yielded the hardest hits. This seemed intuitive in hindsight considering the meat of the baseball bat would make contact with the ball but it was really interesting to see it visualized. Similarly this was true for the contact hits with the ball as well in the strike zone. For our questions, we answered all of them fairly easily. There were no huge issues in trying to answer the questions, it was more a matter of how we can show them in the most visually appealing way. We believe our visualizations accomplish what they are set out to but they could definitely be improved upon. There could be a lot more interaction and depth in our visualizations and with more time that would be what we would improve upon. There are a few visualizations that we could break down into further ones. This would create a lot more depth for our users and allow them to explore the data more. Another thing we could have done better is make them more baseball oriented as far as our visuals. We had a lot of trouble trying to incorporate baseball visuals into our project but found that it would be a significant improvement to grant that type of visual appeal to our users who come to see baseball visuals.

Journal and Date Log

9/10/2022:

- We gathered together to decide what our dataset topic and planned visualization will be.

10/15/2022

    - Finalized data collection and found what we need to make the project work

11/15/2022:

- Created the Home runs compared to Strikeouts visualization. Had trouble implementing the tooltip to hover next to the mouse when looking at a specific point on the line. Also kinda had trouble aligning both lines together.
- After looking at our original idea for the very first chart, we decided to add a little more effort to it. Basically by having a baseball design for it instead of it just being a regular bar graph.

11/16/2022:

- Having trouble with importing an image into the website as the background for the Home Run hits visualization.
- Fixed an issue with the Home Runs to Strikeouts visual, where the tooltip showed the number of strikeouts instead of home runs for the HR dots.
- Fixed an issue where the dots and lines didn't line up with the ticks on the x axis
    - ONLY fixed for certain sizes of the screen.
- Baseball Bar Chart is working, though it needs tooltips to have the bars mean anything.
    - Ball needs to be pulled down a bit more so the bar isn't cutoff

11/17/2022:

- Week has been rough, finally got the website working correctly
    - Exams, one groupmate possibly serious injury, the other went on a trip early.
    - Still managed to get a good bit done though.
    - We've found a guide to one of our baseball field designs but the last one with the 3 lanes may prove difficult.
        - One way to implement this may be to overlay a scatter plot on top of an edited field to show the 3 lanes (straightaway, left, and right). Or another way would be to calculate based on the Latitude whether it's in a certain lane or not.
    - Currently still have an issue with the tooltip in general with all of the visualizations. Might need to look into what version of d3 we're using.

11/23/2022

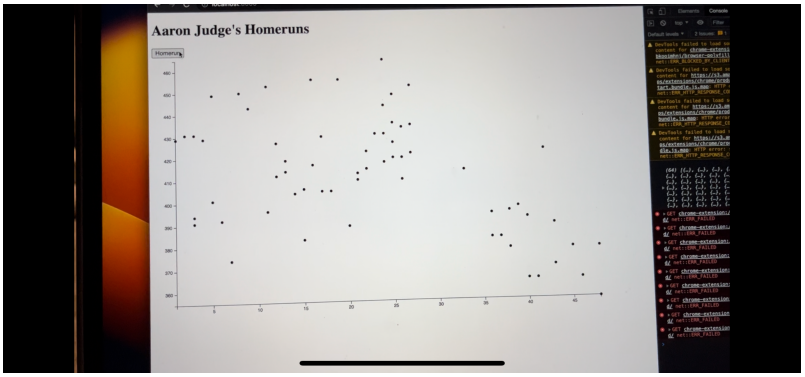
- We have created the visualization we want to have in the final form in a very basic way. By finding locations of where the baseballs landed we were able to graph all the homeruns accurately.
    - Essentially a scatter plot right now that we are going to add features to
    - Need to find a way to add more data and features so it is not so simple
    - Need to find a way to add the baseball field design so that it looks like we imagined

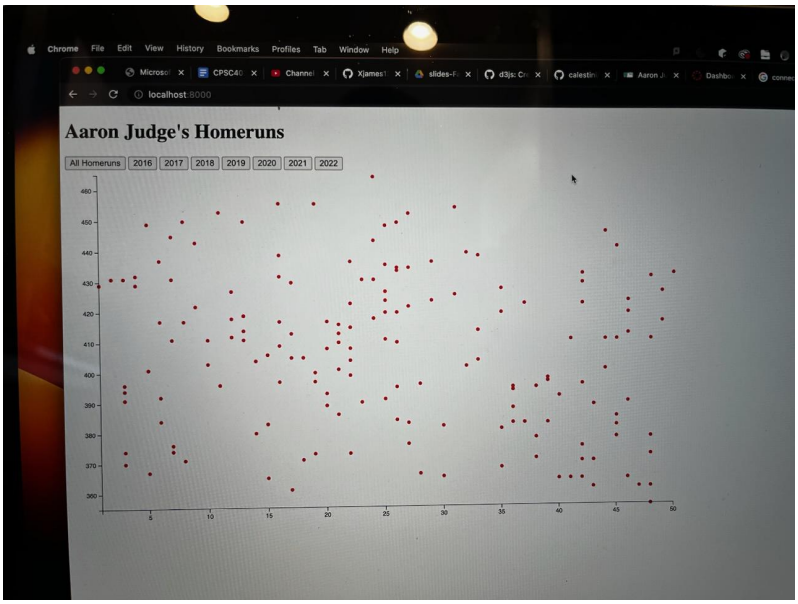
11/26/2022

- Group meeting to reconvene and establish new goals
    - Got on same page about final visualization so that we can finish soon

12/04/2022

- Finally got the visualization working like I want it to
    - Still a basic scatter plot but working on adding years to display the 206 homeruns
    - Took some serious work to find out how to display different data sets

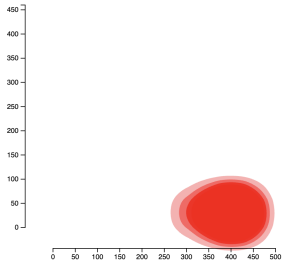


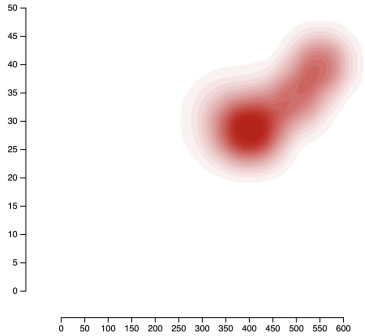
    - Later that night:
    - 

12/12/2022

- Started the grind to get an A on this project
    - Made a heatmap of the strike zone representing where Aaron judge hits all of his hits in his career

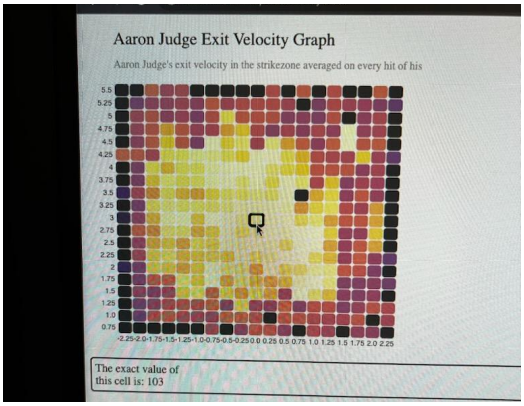
- Need to clean it up and add a tool kit and labels
- Tried to build d3 depth diagram - very difficult to do and not enough data
- Our attempt:





12/13/2022

- Completed the toolkit for the first two visualizations and
- then made the heatmap look better
  - Started on exit velocity graph and differentiated our data
  - Going to finish three visualizations tomorrow and finalize the last one
  - A lot of work, should have started earlier but I think we can still succeed
  - I need to change the color scheme so it makes more sense, darker in the middle and lighter on the outside for the less hard hit balls

- 

12/14/2022

- Made the last visualization and I am really proud of it.
    - Took a lot of learning but I feel very comfortable with d3 and json at this point
    - Need to finish up and make them all fit and look pretty on one slide
    - Going to record the video tomorrow