CS 6630 MAMBA MENTALITY
Jordan Hendriksen, Marko Miholjcic
November 15, 2020

The title Mamba Mentality comes from Kobe Bryant's autobiography, "The Mamba Mentality: How I Play."

Basic Information
Title: Mamba Mentality
Group Members:

- Jordan Hendriksen, UID: u1063288, Email: jordan.hendriksen@utah.edu
- Marko Miholjcic, UID: u0984549, Email: u0984549@utah.edu

Github Link: https://github.com/miholjcicmarko/DataVizProject

Overview and Motivation

Kobe Bryant is an inspirational figure in the world of sports. His tragic death has led many sports writers and analysts to analyze his legendary 20 year career. Numerous visualizations on talk shows and social media have attempted to convey his legendary career. Most of these visualizations fall short of capturing interesting details of his career or overwhelm the viewer with the large amount of data that Bryant's career presents. We feel that we are capable of developing an interactive visualization that will convey more details about Bryant's career while maintaining visual appeal. Additionally, both group members are passionate basketball fans. Therefore, we look forward to exploring and visualizing the data collected over the course of a legendary NBA career. Another reason for choosing this project is the opportunity to work with a large dataset. Communicating complex data and large datasets is an important skill to have because data has become important in industry. We feel that managing and resolving the problems that may arise with visualizing over 50,000 data points is good practice for our future careers as data scientists.

The dataset allows for different types of visualizations. We can look at trends and distributions of shots over the course of a season or over the course of the entire career and where they occur across the court. The dataset can inform the viewer about which areas of the court did Bryant shoot from the most and how did his shot tendencies change over time. Furthermore, the dataset allows Kobe to be compared to other players in the regular season and in the playoffs.

Another aspect that we wanted to capture in our visualization were the legendary game winners throughout Kobe Bryant's career. Bryant was defined as a clutch basketball player, and providing a central location for all of his famous shots will be appreciated by the average NBA basketball fan.

Related Work

Upon Bryant's tragic death, multiple data science competitions appeared on Kaggle. A large amount of the competitions revolved around analyzing and visualizing the data produced over the course of Bryant's career. The Los Angeles Times produced a visualization displaying the 30,000 shots that Bryant attempted throughout his career shown in Figure 1. The story-telling aspect of the Los Angeles Times visualizations highlighting important shots made in important games. In addition, the comparison of Bryant's different seasons was a nice visualization that we wanted to include in our visualization. Another inspiring visualization was discovered on Github. As shown in Figure 2, the visualization provided a shot chart for all the shots attempted during Bryant's career. This visualization provided a drop down menu where the user could choose which iconic game they wanted to observe.

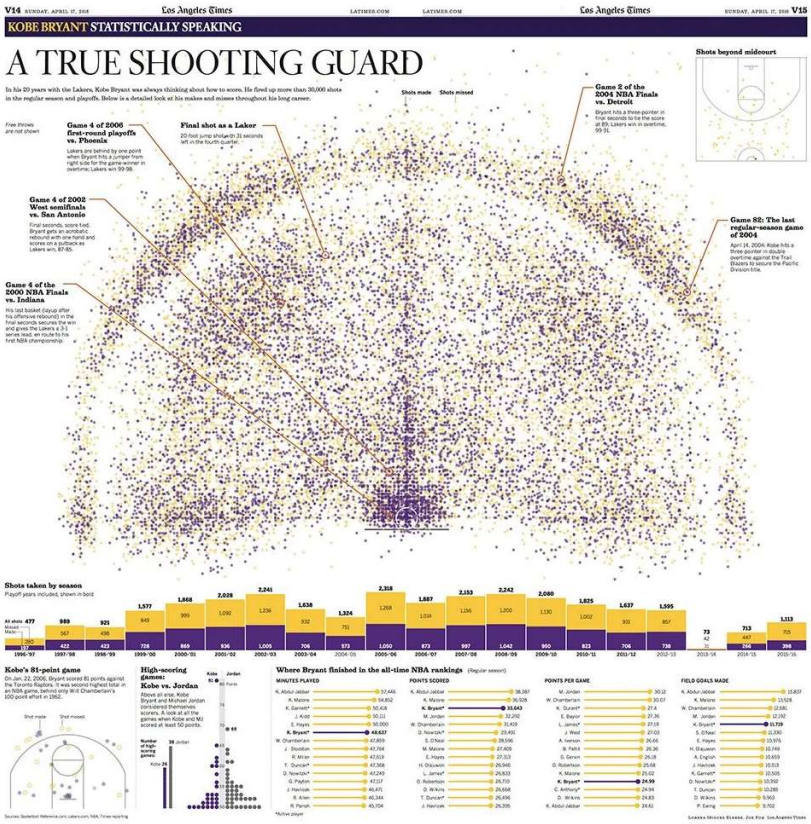

Figure 1: The Los Angeles Times visualization of Kobe Bryant's career. The link to the visualization (now broken) is: https://graphics.latimes.com/kobe-every-shot-ever/.

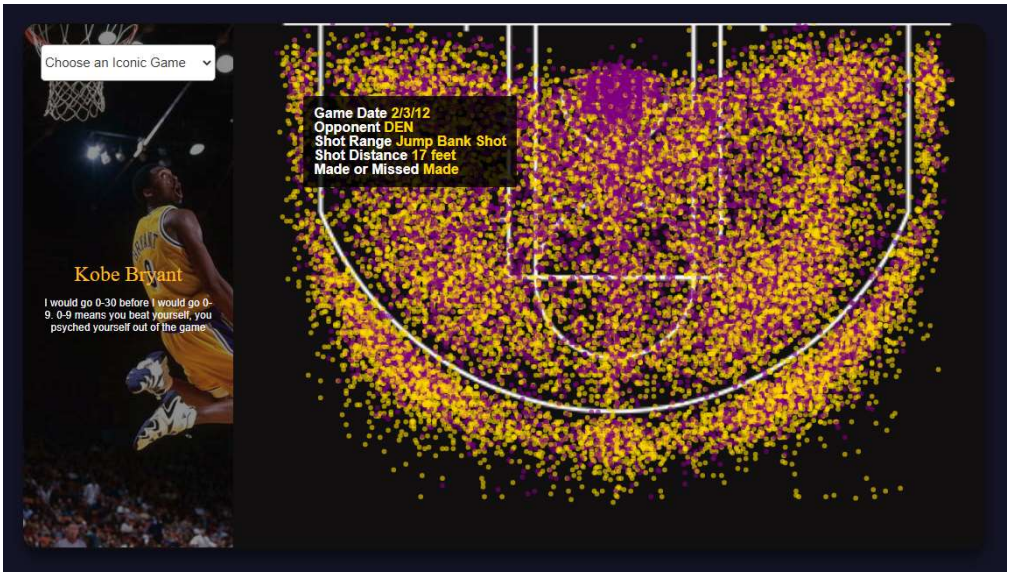

Figure 2: The visualization of Kobe Bryant's career found on Github. The link to the visualization is https://github.com/derekchoe/kobe-chart.

Questions

The primary questions that we are trying to answer with our visualization are:

- What are all of the significant shots that Bryant attempted and made over the course of his career?
- Which areas of the court did Bryant shoot from the most?
- Which parts of the court were the lowest percentage shot for Bryant?
- How did Kobe's shooting percentage and shot tendencies change over time?
- Where did buzzer beaters occur from?
- Did Kobe perform similarly in the playoffs and regular season?
- Were there particular seasons that he performed differently?
- How does Kobe compare to other players? Volume? Scoring?

Data Acquisition and Processing:

The data was collected from the nba_api.stats. We used the package, "shot chart detail." This package provided a shot chart for every basketball player in the NBA for every game they have played in during the regular season and playoffs. We utilized a loop to load in all of the data for Kobe Bryant for every year that he played in the NBA. The data has over 30,000 rows and 24 columns. Each row is a single shot attempt in Kobe Bryant's career. The columns provide details about the shot, such as shot type, the opposing team name, the location of the shot on the court, whether the shot was made, etc. The link to the api is https://pypi.org/project/nba-api/. The player ID and the seasons played were altered into order to access a different player's career.

Data clean up was required for this project. The importation of the data will be performed using python. The data was imported from the api and we checked for null values in any of the data points. We added columns to track which season corresponds to each shot and to track if the

shot occured in the regular season or the playoffs. We placed the data into a dataframe and converted it to a csv file for each player.

In addition, we derived the shot percentage for each hexagon in the heatmap for the career as well as cumulative shot percentage for each hexagon dependent on the year being filtered for. The percentage derivation was implemented using JavaScript by cycling through the shots taken and counting the number of shots made and then dividing by the total shots. This calculation was performed for each bin of the heatmap and upon drawing and was thus reperformed after any filtering or change of player took place.

For the brushing aspect of the visualization, the percentage of shots made within the brush area will be calculated, and the distributions needed for the additional visualizations will be created using D3. Necessary calculations, such as aggregate field goal percentage and aggregate field goal percentage for each year within the selected regions, are calculated by cycling first defining which bins are within the region. These bins are then cycled through and the needed data is pulled from each shot within the bin. This implementation should carry over to any region, number of bins, size of bins, or filters applied to the data.

Exploratory Data Analysis:

Our data was initially looked at using the scatter plot type visualizations that we had seen online. From these we were able to deduce whether a heatmap would be a useful visualization for this dataset.

Design Evolution:

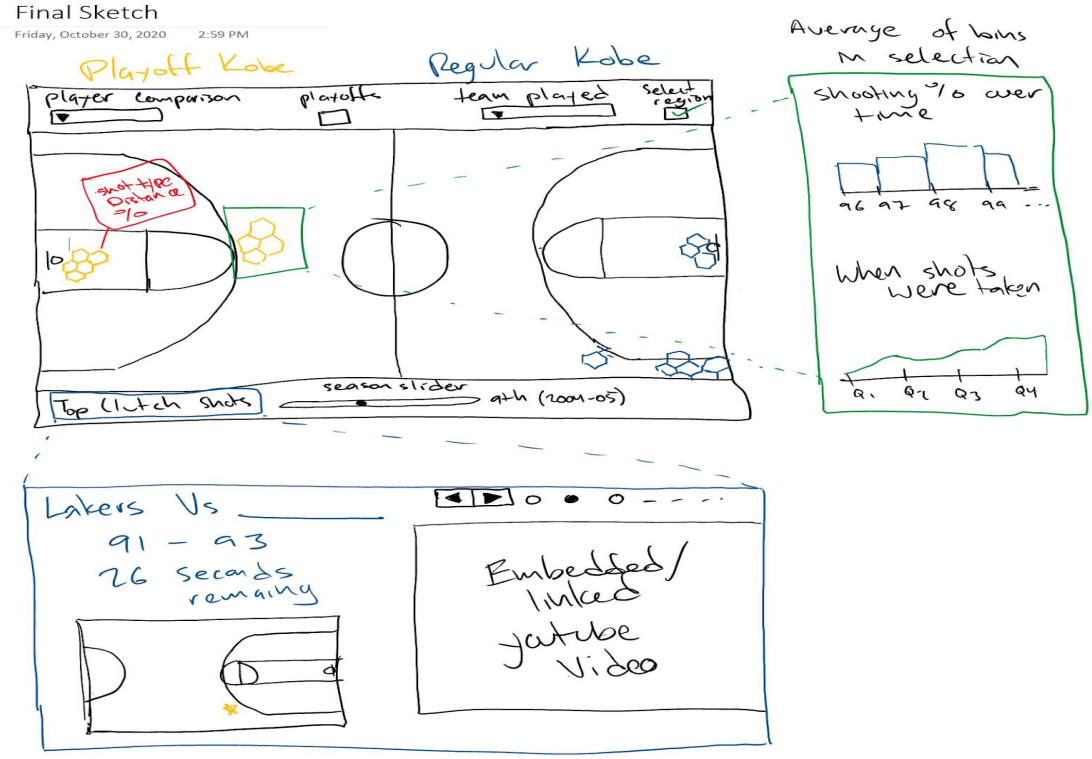

Figure 3: The final sketch we created as part of our design process prior to our project proposal. This is representative of our original aim as we began developing our visualization.

Incorporating additional labels for the season: When making the slider, we noticed that the NBA season includes two years. Therefore, we need to number the seasons. Additionally, the dataset did not differentiate between shots attempted during the regular season and playoffs. As a result, we need to add data labels for the two different categories.

Separate Slider for the Opposing Player: Initially, we planned on having a single slider in the visualization. However, we have noticed that the players had different career lengths. This posed an issue when filtering the data for the different players. Therefore, we have decided to create a second slider below the first one when the player comparison option is chosen.

Figure 4: Comparison between the original and the updated slider visualization. The second slider will match the color of the heatmap on the right in the final implementation.

Shot Attempt Encoding: In our proposal, we had no plan for representing the total number of shots taken from within a hex-tile. After implementing this we noticed that the decreasing opacity made it difficult to determine how the color/field goal percentage was changing across tiles (Figure 4).

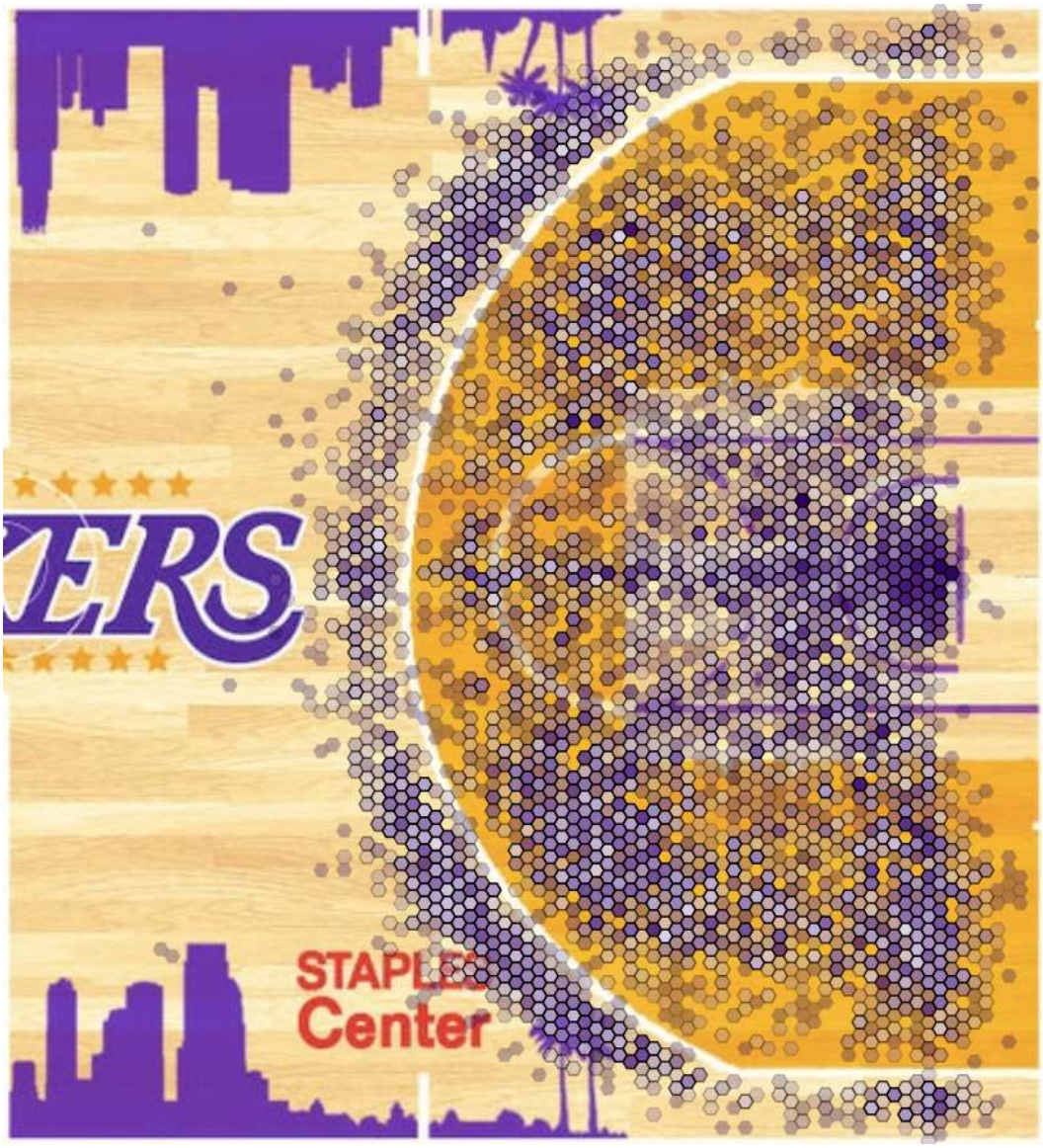

Figure 5: Original heatmap (shot data is over Kobe's entire career) implementation using opacity as the encoding method for number of shots taken. While the areas of high traffic are relatively clear, it is difficult to compare shot percentage between these high frequency areas and lower frequency areas.

This led us to try encoding the shot attempt number through another method - stroke color saturation. This encoding saturates with increasing shot attempts and is clamped at a set number of shots (12 for the career heatmaps) so that areas with a high rate of shooting are distinct, while still allowing for more direct comparison between the fill of tiles. Lower opacity is still applied to tiles with only a singular shot taken so that they are not as impactful as other regions of 100% accuracy.

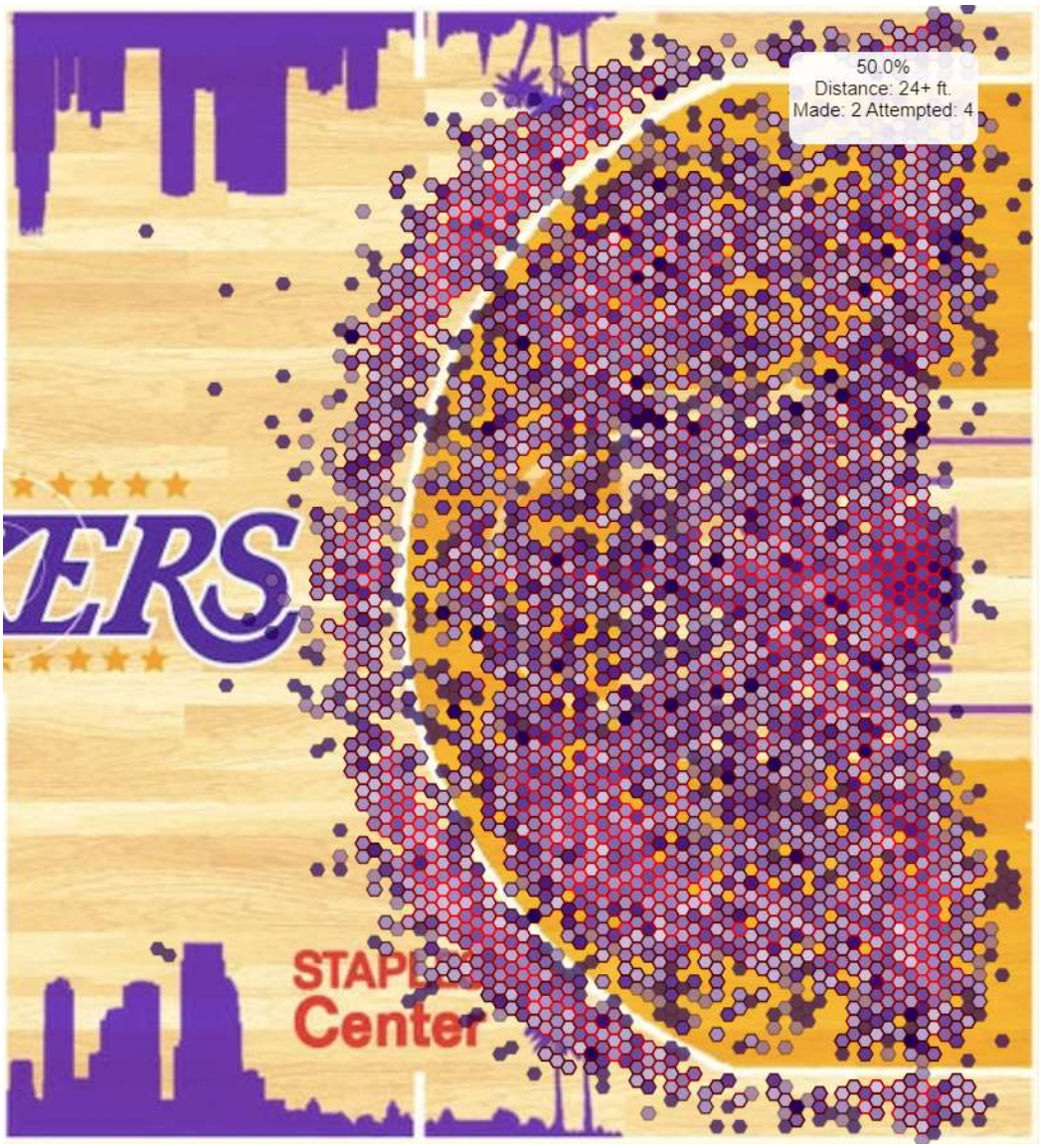

Figure 6: The current implementation of the heatmap seen in Figure 4. Rather than using opacity as the encoding method for shot frequency, saturation of the stroke color is used. This allows visualization of distinct areas that are high frequency (near the hoop, near three point line, center, etc.,) while also allowing a more direct comparison of fill between hexes.

Aggregate Statistics Vis Location: We originally planned on displaying the smaller visualizations generated upon brushing the heatmap adjacent to the court. However, to improve compatibility across display types and across visualization layouts, we decided to put these visualizations in a div similar to a tooltip that stays a certain distance from the selected region. This would allow us to increase the size of the court/heatmap and/or center it while there wouldn't have been room before. We may also make the div switch to the left side of the selection if it is within a certain distance of the end of the court.

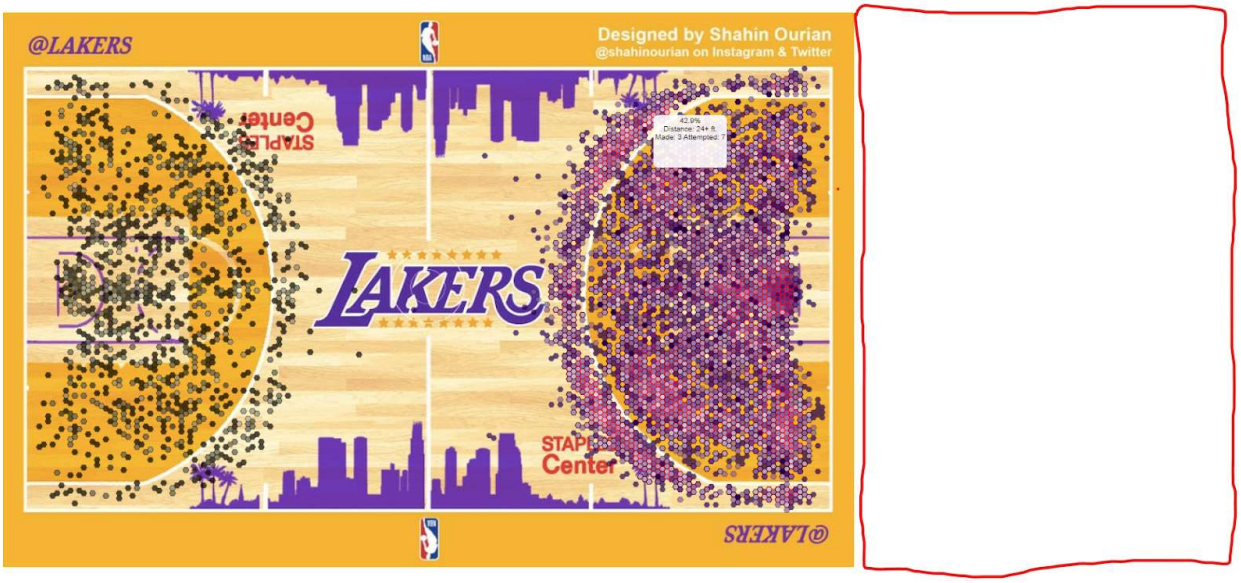

Figure 7: The planned location of the sub-visualizations in context of the full court visualization (boxed in red).

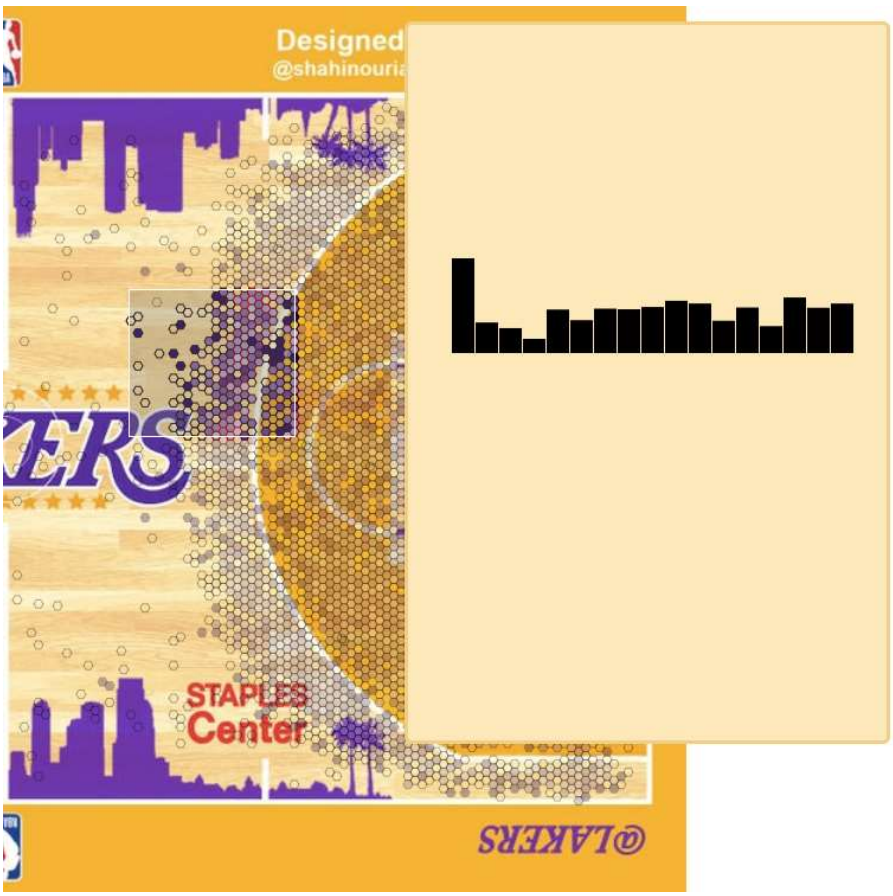

Figure 8: depiction of the current location for the sub-visualizations - this is set to maintain a certain distance from the right edge of the selection. The early implementation of a sub-visualization shown is the field goal percentage in the selected region across the years.

Exclusion of shots from beyond half court: Initially, we did not know that Kobe Bryant had attempted multiple shots beyond the half court mark. None of these ¾ court shots were made shot attempts. As a result, we decided to hide bins from beyond half court that did not include successful shots.

Implementation:

The heatmap is meant to display the distribution of shots throughout Kobe Bryant's career and his field goal percentages from different areas. This is accomplished through using a hexagonal heatmap that bins the shots by location and then uses color to encode the number of shots taken and the field goal percentage(Figure 6). The tool tip and brushing feature of the heatmap will display additional information about the shots Kobe took from within a hovered hexbin - this includes number of shots attempted, made, and the true field goal percentage (Figure 10). A slider will filter the data to show the shot distribution for specific years (Figure 11). The comparison drop down will allow current NBA players to compare with Kobe Bryant by drawing a heatmap of the same style on the other end of the court (Figure 9). The story mode

presentation will display the buzzer beaters that Kobe Bryant made over the course of his career as embedded videos. These can be toggled through using a set of Next/Back buttons and in the future will include the context of the shot (Figure 12).

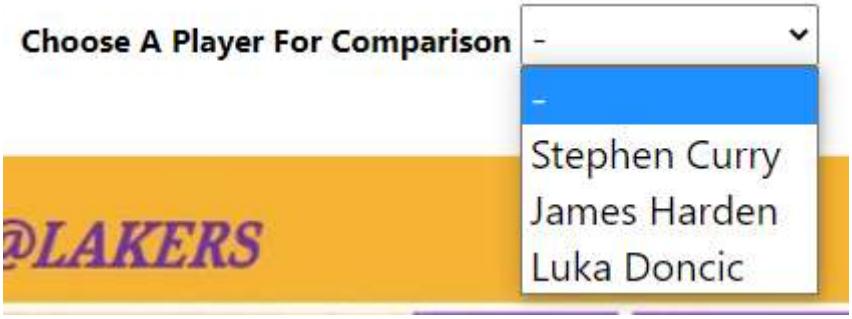

Figure 9: Current implementation of the player comparison dropdown above the left side heatmap.

Figure 10: Implementation of tooltip providing shot percentage, shot range, shots made, and attempted shots for the hovered hex (gold and white).

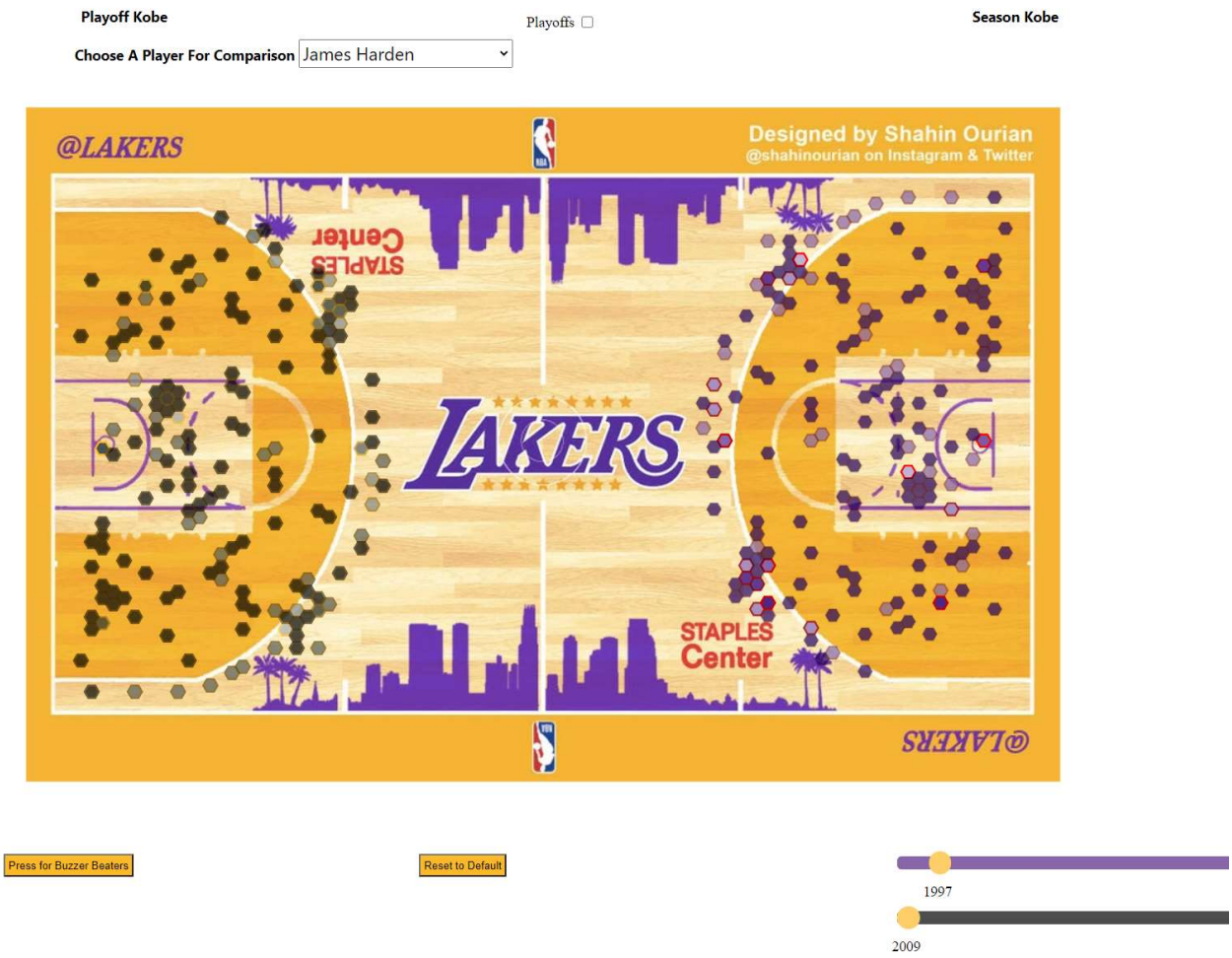

Figure 11: The current slider implementation for by year filtering and the resulting heatmaps for both players.

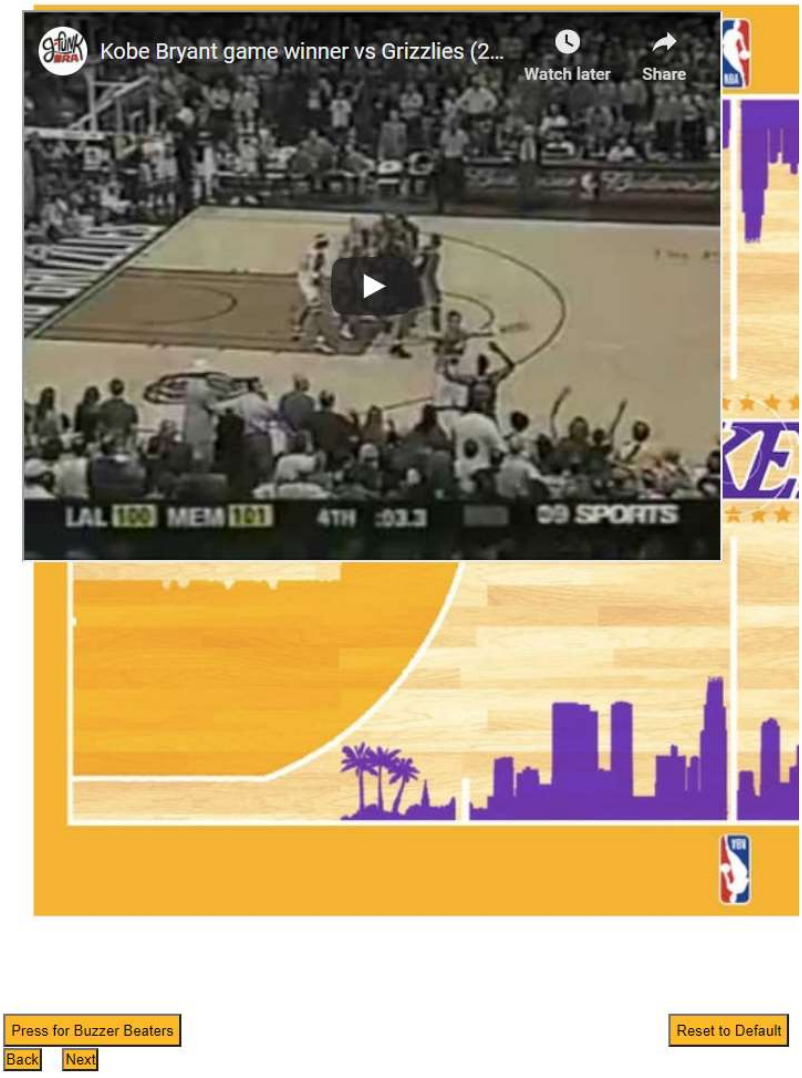

Figure 12: Current implementation of our story mode focused on buzzer beaters. Pushing the button back and next in the lower left will change the buzzer beater shown and reset to default should remove the overlay.

Evaluation:

The visualization gave insight into how Kobe Bryant has changed throughout his career. We noticed that early in his career, Bryant did not attempt many shot attempts from the beyond the 3 point line. As Bryant gets older, a larger distribution of jump shots are attempted from beyond the three point line. We noticed the large disparity in the number of 3 point attempts when comparing Kobe's earlier years with the shot distributions of players of 2020. The visualization shows how basketball has changed in the past twenty years. We were also able to see, when selecting different regions with high frequency, that in the first few years of his career Kobe's field goal percentage increased fairly linearly followed by a relative plateau across the

rest of his career (Figure 13). This pattern was also seen in some of the other players such as James Harden and in Stephen Curry's 3-point shot.

Figure 13: Demonstration showing how Kobe's and Harden's field goal percentage changed over the years in the paint. Interestingly both showed the same pattern of a linear increase for their first few years followed by a slower decline and plateau.