J-Stroke

A new way to think about NBA Players

Michael Krebs, Harsh Rana

J-stroke is a project undertaken to examine the current methods of basketball player evaluations and is a data visualization tool to complement currently used methods. The developers were very interested in building a tool that not only communicates the playing style and patters of a player, but also helps compare different players with different playing styles.

This document outlines the design and development process for this project, including important milestones and decisions.

Overview and Motivation

Basketball is a highly popular game, which involves players with many different types of playing styles working together to create a strong team. Thus, it is important to be able to analyze players, finding their strengths and weaknesses and playing patterns when building a team. The main motivation behind building J-stroke was to use the widely available basketball player data from the NBA to build one such tool for analyzing players. The main goals we wanted to reach were the following:

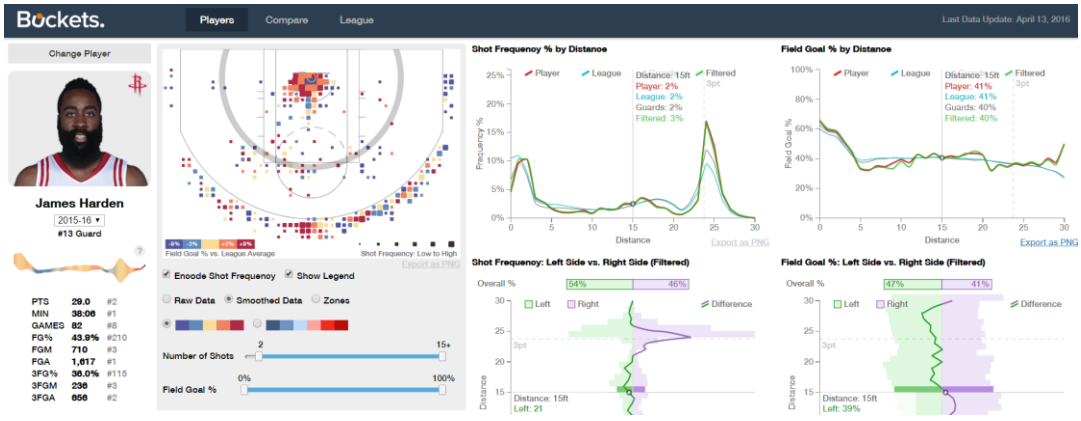
1. Visualize player shooting styles and patterns

2. Visualize strong and week shooting areas for a player

3. Incorporate multiply views to account for people with different preferences

Related works

One of the best projects we found related to this topic was Buckets by Peter Beshai. Peter build a data visualization dashboard to visualize players, analyze and compare shooting styles. His project was highly successful and was one of the inspirations for our own designs and interactions. The main view of the dashboard can be seen below-



As one can see from the dashboard above, buckets was a highly complex and granular tool, capable of much more than the scope of this project. We wanted to design J-stroke to have some of the same good parts of buckets, while adding additional ways to open up the field for new insights.

Questions

One of the main questions we wanted to address with this project was about using multiple views to account for user preferences in visualization. As one can see from the Buckets dashboard, Buckets does an awesome job of combing the needs to visually aligned people, with quantitatively aligned people by creating multiple views of the same data. We wanted to utilize a similar approach. Other questions which we wanted to address included efficiency of color encoding for showing successful/unsuccessful shots

Data

Our data was sources from stats.nba.com for this project. The website is the official NBA stats website and contains a very thorough API with a wide array of data. We were able to utilize the *shotchartdetail* endpoint to gather all the data about player shots and the *commonplayerinfo* endpoint for player information. To collect the data for different players, we also had to build a webscraper to scrape player ID's from the NBA stats website. Our data collection/scraping program was written in python and our code can be viewed in the repository in the python folder.

Some of the interesting fields that we discovered in the datasets were-

1. Player information like name, team, jersey number, height etc.

2. Shot type: jumpshot, layup etc.

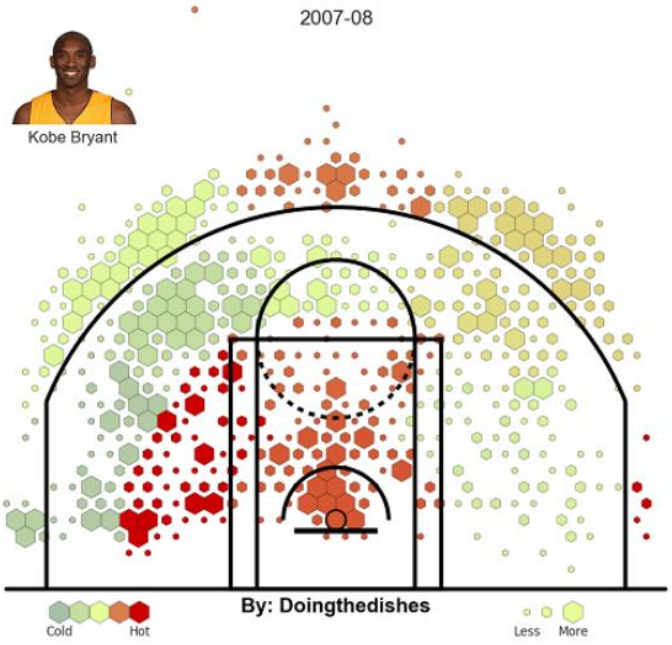
3. Shot location: x and y coordinates

4. shot result: made or missed

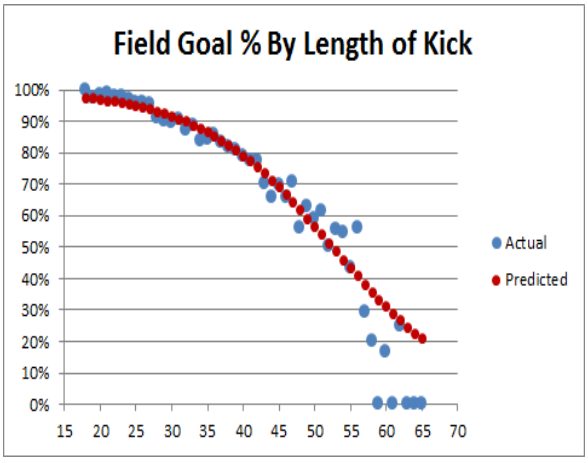
One interesting aspect about the dataset was that we had an incredible amount of data (upwards of 15000 data points per player) and to efficiently visualize this data, we utilized an aggregation-based approach which is described later in this document.

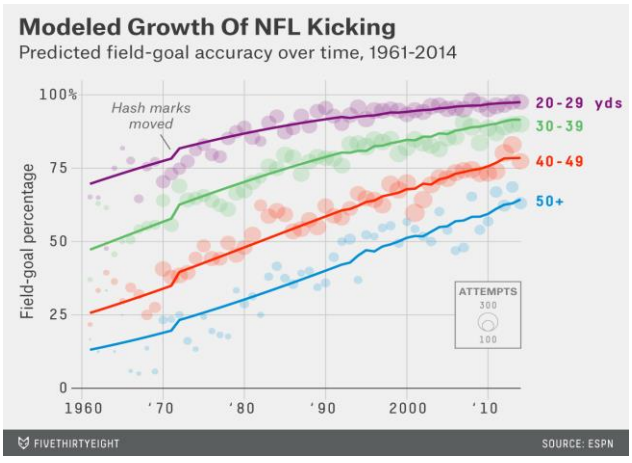
Exploratory Data Analysis

Once we'd collected and cleaned our data, we started out by creating some basic designs of the different visualizations we wanted to build. One visualization that we knew was highly effective at communicating this data is called a shot chart. A shot chart shows the players' shot attempts as points on a basketball court, often with visual encodings such as color and sized map to attributes like field goal percentage, shot type etc. This type of chart can be seen below-



Another visualization that we wanted to explore was a more quantitative version of the shot chart: one that mapped field goal percentage to distance. This would help eliminate some of the noise made by the x and y location and reduce it to only the distance attribute. This type of chart could be made using a few different types of charts as it would be a simple frequency graph. Some examples are shown below-

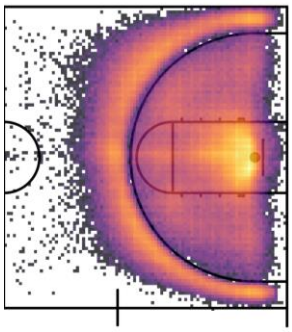


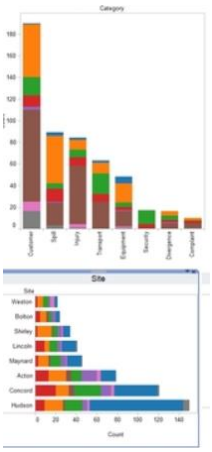


After trying out a few different types of charts, we decided to move forward with a line-based frequency graph as it would best communicate the fluidity of distance, similar to the game. Visualizing our data based on the above two charts showed us how easy it was to spot patterns and styles in players' shooting abilities and made us more confident moving forward with even better visualizations.

Design Evaluation

We'd already decided on two of the graphs we'd be using for the J-stroke dashboard, which started our journey of finding a third and potentially forth view. Our main ideology behind another view was to better combine the quantitative and visual aspects together into one such visualization. Some of the charts we explored included a heatmap which mapped distance to field goal percentage, heatmap of the basketball court, stacked bar chart showcasing different shot types inside bigger bars of shots attempted. Some of these chart examples are shown below.

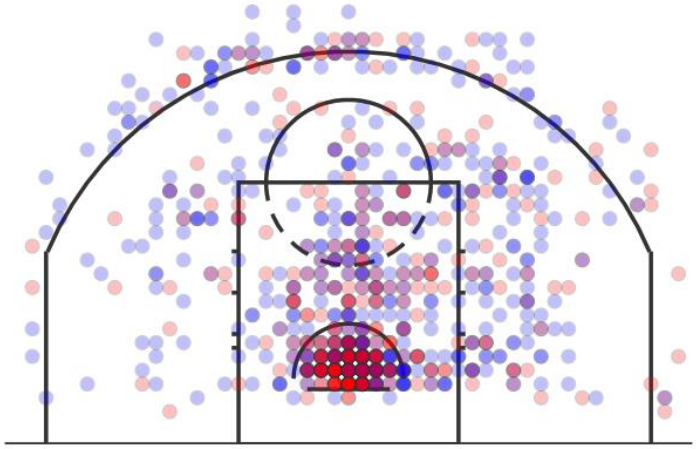




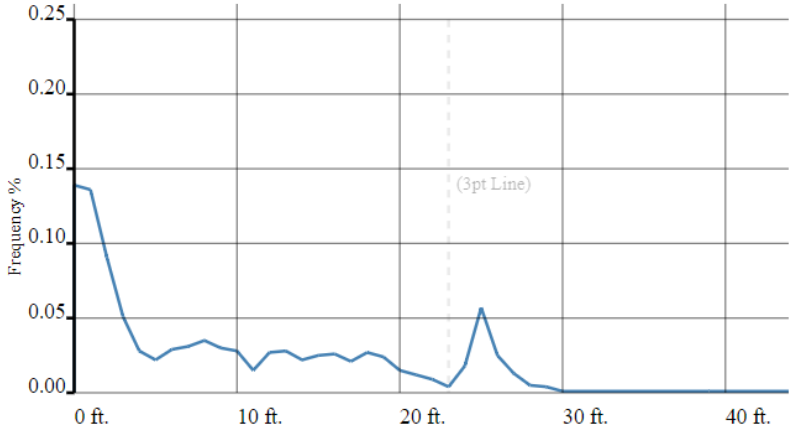
We finaly decided to go a different, and more complicated route. We decided to showcase the actual path a ball took to reach the basket. We got this idea from a Washington Post data visualization about basketball.

Implementation

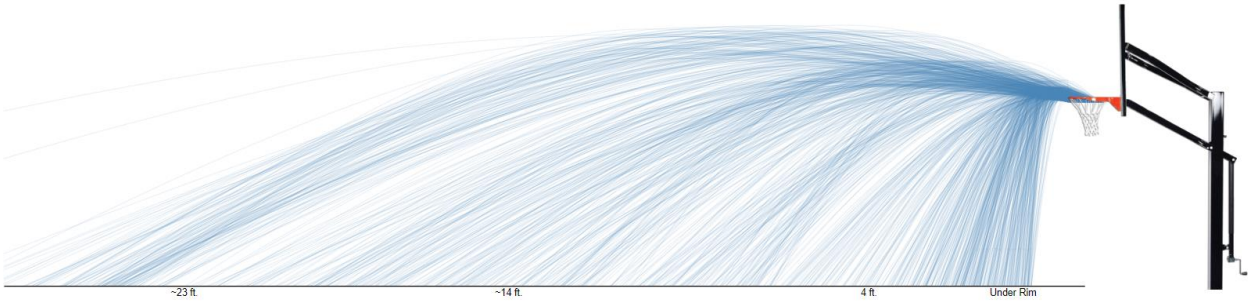
We started out by creating a single view dashboard containing the shot chart for a specific player. We utilized a single visual encoding channel, color, to showcase the field goal percent and used aggregation to combine similar shots together. Our shot chart can be see below-



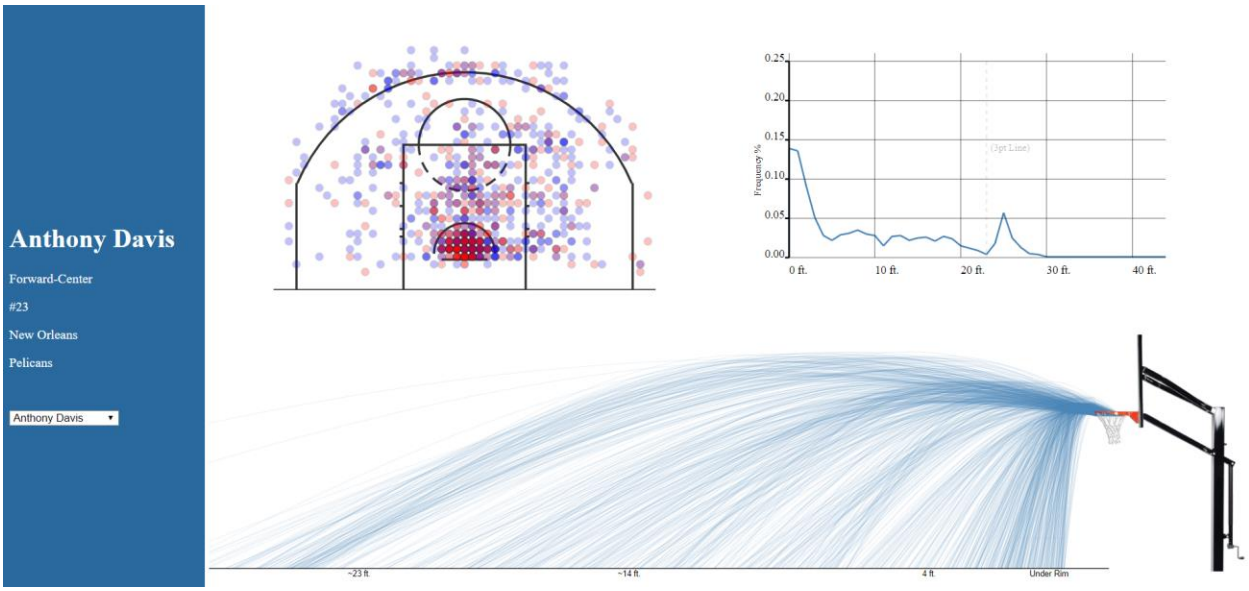
Next, we built the frequncy chart of field goal percentage and distance. This chart can be seen below-



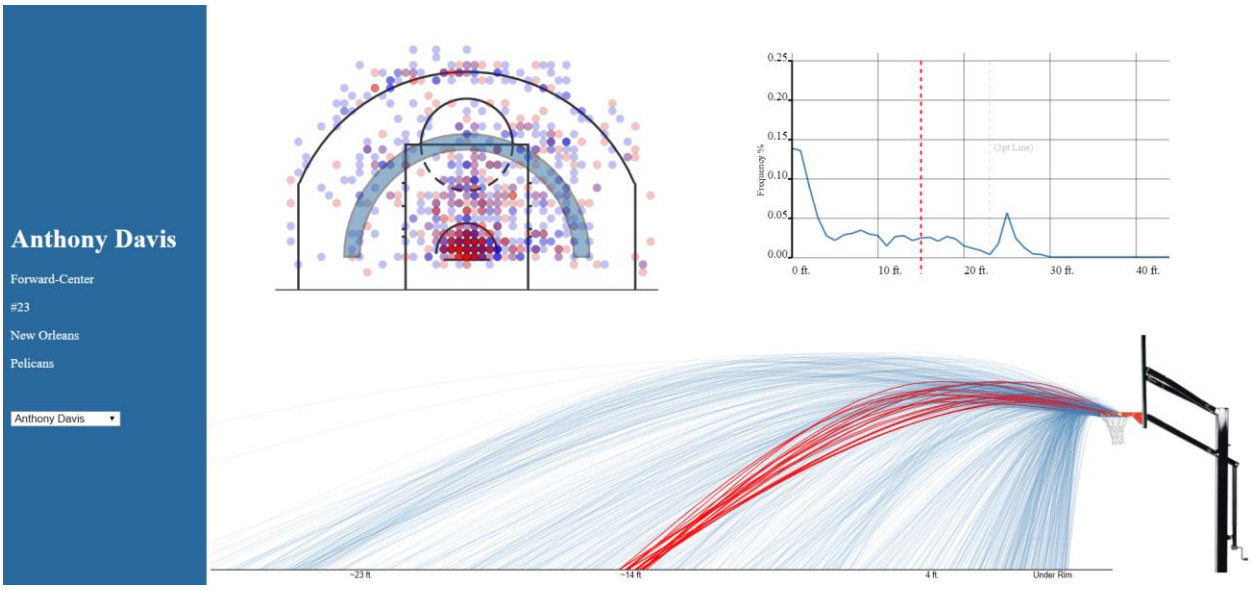
Finally, after exploring D3 techniques like line interpolations, seeded randomness, and arcs, we built the shot path/curve chart which can be seen below-



All these charts worked really well together with the shot chart being highly visual, the frequency chart being highly quantitative and the shit curve chart combining the two concepts. Our final dashboard can be seen below-

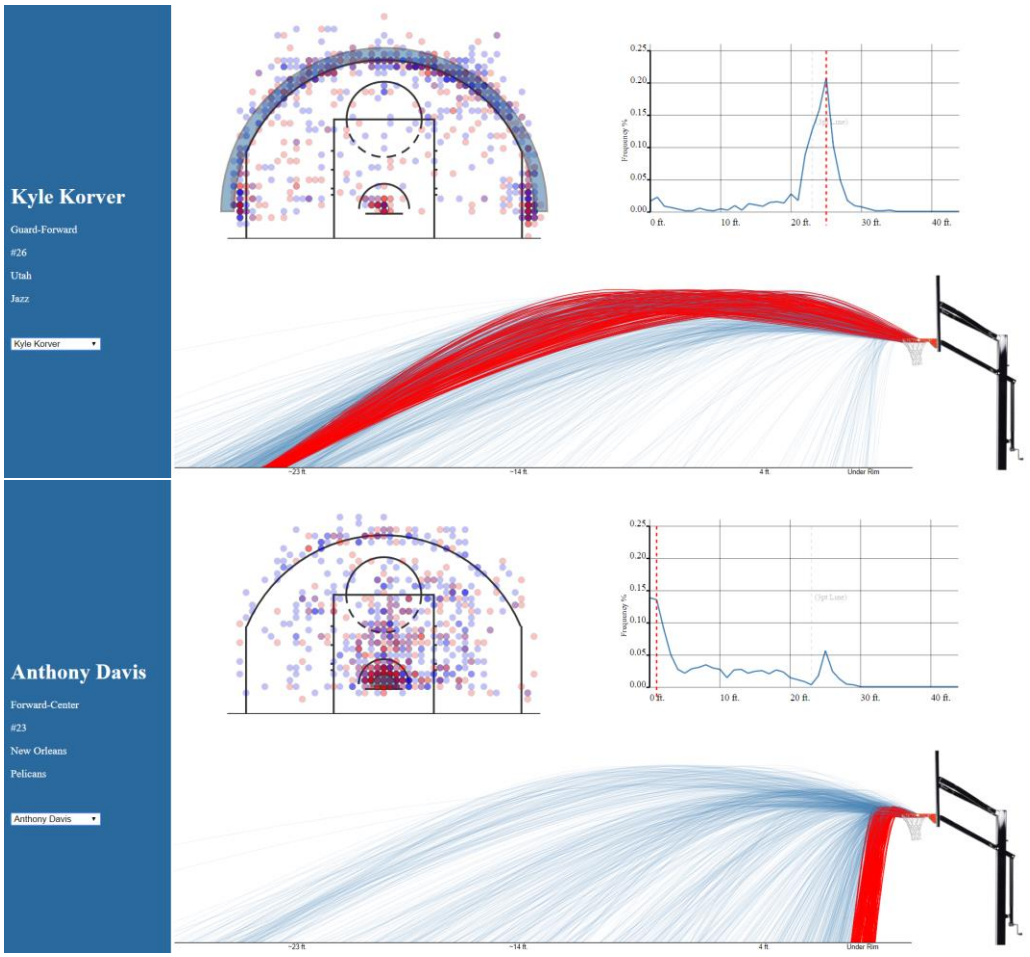


In true multiview dashboard style, we added a three-way binding between the charts and even visual feedback channels so a user could interact with the charts and find more value than static graphs. The interactions were hiver based and could be instantiated from any of the three graphs. This is showcased below-



Evaluation

We were very satisfied with our end product. Our visualization was robust, efficient and highly practical and user friendly. We used smart data wrangling techniques like aggregation and random sampling to make our dashboard extremely fast. We used visual feedback cues like on-hover size changes etc. to make the user experience more pleasant. We created a basic data set with different types of players so people could compare player. Such a comparison is shown below-



One can clearly see that Kyle Korver is a fierce three point shooter given his high field goal attempts and successful shots as showcased in the top graph, which Anthony Davis is a more involved and under-the-paint kind of shooter.

Such comparisons, pattern finding and evaluation was the reason we built this tool we're happy with how it came out. For the future we wish to improve the tool by adding filtering techniques, more data sets, a side-by-side comparison feature and other smaller improvements.