HoopInSight: Analyzing and Comparing Basketball Shooting Performance Through Visualization

Yu Fu and John Stasko

Abstract—Data visualization has the power to revolutionize sports. For example, the rise of shot maps has changed basketball strategy by visually illustrating where “good/bad” shots are taken from. As a result, professional basketball teams today take shots from very different positions on the court than they did 20 years ago. Although the shot map has transformed many facets of the game, there is still much room for improvement to support richer and more complex analytical tasks. More specifically, we believe that the lack of sufficient interactivity to support various analytical queries and the inability to visually compare differences across situations are significant limitations of current shot maps. To address these limitations and showcase new possibilities, we designed and developed HoopInSight, an interactive visualization system that centers around a novel spatial comparison visual technique, enhancing the capabilities of shot maps in basketball analytics. This article presents the system, with a focus on our proposed visual technique and its accompanying interactions, all designed to promote comparison of two different scenarios. Furthermore, we provide reflections on and a discussion of relevant issues, including considerations for designing spatial comparison techniques, the scalability and transferability of this approach, and the benefits and pitfalls of designing as domain experts.

Index Terms—sports data visualization, sports analytics, visual comparison, basketball

1 INTRODUCTION

With the increasing prevalence of sports data collection and analysis, sports visualization has become a powerful tool for accessing, analyzing, and communicating sports data and analytics [11, 38]. Only a handful of existing sports visualizations can be considered as having fundamentally altered the trajectory of a sport, however. In the realm of basketball, shot maps (or shot charts) [18, 19] are such an example—the locations where teams take shots have changed dramatically, driven in part by the insights available through shot maps [43]. Basketball teams have modified their game plans, aiming to create more opportunities for players to take shots from their “hot spots.” Such shifts in professional leagues (e.g., the NBA - National Basketball Association in the U.S.) have rippled out to millions of fans and players worldwide, affecting their appreciation of and engagement in the sport of basketball.

Although the use of data, visualization, and analytics in basketball has grown significantly over the past decade, one form of analysis that has not seen an accompanying focus is in-depth comparison of data and information. Comparison is one of the most important analytical activities in sports [11]. For example, how does one player compare to another, or how has a particular player’s performance improved or fallen over time? Media outlets rely on comparisons to assemble a wide range of compelling stories. Fans engage in casual or fierce arguments with friends and strangers about which player is better on different occasions. Analysts and coaches make comparisons to evaluate player and team performance, analyze various circumstances, and make appropriate adjustments in team make-up and strategy.

Our particular focus in this work is the comparison of shots taken in basketball. When a player “shoots” the ball toward the basket hoping that it goes through the hoop, this is called a “shot.” Obviously, whether the shot is made or missed is the most important metric, but many other accompanying variables exist for each shot. The second most important variable is likely the position on the basketball court where the shot was taken from. Additional variables include the type of shot (dunk, jumper, hook, etc.), the time of game/possession it was taken, and which other players were in the lineup at that time, among others.

In general, this basketball scenario can be abstracted as a spatio-temporal data analysis problem. The basketball court is our spatial grounding, which we divide up into a series of contiguous smaller regions for aggregation and convenience. Each region then has multiple events (shots) occur within it, each event having multiple quantitative and nominal variables. The primary variables that we focus on are frequency and efficiency. Frequency denotes how often shots are taken from a certain region and can be measured as the number of shots or as a proportion of total attempts. Efficiency denotes how accurate/effective a player/team is shooting the ball, i.e., the proportion of shots taken that are made. Our goal is to support our primary user group, sports analysts or other analytic-savvy stakeholders (e.g., fans, journalists), examining the shot behaviors of two references, such as different players or one player in different time periods, and to compare the performance and patterns of the two selected references, hopefully gaining insights from the comparison.

We have developed an interactive visualization technique and an accompanying system called HoopInSight to compare and analyze shooting patterns and performance in a way existing shot maps fail to provide. Our approach revolves around a novel visual technique with flexible interactions and a real-world data pipeline. More specifically, the data pipeline and interactivity enable users to swiftly construct a spectrum of comparative conditions/scenarios, e.g., Denver Nuggets’ shooting behaviors with or without their star player Nikola Jokic in the game. Meanwhile, our proposed visual technique allows users to rapidly compare multivariate spatio-temporal data across two chosen scenarios and derive insights, such as determining the locations where shots were taken more/less frequently and how much more/less efficient those shots were, and ultimately infer the contrasting patterns.

2 RELATED WORK

2.1 Sports Visualization

Perin et al. [38] conducted a comprehensive state-of-the-art review and identified two primary roles of visualization in sports: analytical and narrative. Visualizations used to analyze sports data focus on understanding the ever-increasing amount of data and extracting insights from it. These insights can lead to novel analytical storylines, help scouts understand opponents’ tendencies and behaviors, and even assist in strategy-making or team assembling. Narrative uses of sports visualization tend to be more prevalent, targeting a broader audience with more straightforward messages. Du and Yuan [11] focused on the use of visual analytics in competitive sports and proposed a more nuanced categorization of competitive sports data. They also listed feature...
compared as one of the three main tasks for competitive sports data visualizations. Our work falls into that category and seeks to advance the analytical use of visualization by incorporating multiple data types revolving around spatio-temporal information about shots.

While many sports visualization practitioners have primarily focused on using visualization for communication purposes, researchers also have sought to address challenging analytical problems leveraging both novel interactions and visual techniques (e.g., glyphs) [1, 10, 25, 37, 40, 44, 49]. In the domain of basketball, projects such as GameFlow [7] and BKViz [29] aim to demonstrate complex statistics about NBA games, such as game trends, play sequences, and team rotations. Another focus is the emerging notion of supporting sports journalists to cope with increasingly advanced basketball data and analytics, aiming to assist them in exploring such analytics and identifying novel insights to build narratives [13, 52]. Other researchers have endeavor to leverage tracking data and its derivatives to further advance the use of visualization in sports analytics [42, 48]. Recently, visualization researchers have sought to integrate sports visualization with natural language and VR/AR technology [8, 9, 27, 28], thus broadening the roles of sports visualization and benefiting more stakeholders. Reflecting on their projects and collaborations with different user groups, Lin et al. provide valuable insights into working with different stakeholder experts and future research opportunities [26].

2.2 Visualization for Basketball Spatial Analytics

Our work aims to push the boundaries of basketball spatial analytics through interactive visualization. The quest to understand basketball shooting behaviors through spatial analysis dates back to the 1940s. Research on this topic [21, 39], however, was scattered, until Goldsberry’s CourtVision technique paper [18] popularized the technology in both academia and industry, with his signature shot map paving the way for basketball spatial analytics. Ensuing research has emphasized developing and improving models [6, 22, 23, 32] to either reduce dimension or provide estimations/evaluations.

Although CourtVision was originally intended for spatial visual analytics, its primary use has been in sports journalism, partly due to Goldsberry’s long tenure as a journalist. To communicate sophisticated insights to broader audiences, he recast his invention into a set of variants, ranging from classic hexagonal bins to more recent contour lines to aesthetically appealing shot charts featuring 3D animation and embellished visual effects. Other practitioners have likewise explored different forms to engage audiences and deliver compelling stories with shot charts. For instance, Whitehead created a physical representation of Stephen Curry’s career three-pointers using toy trays [33, 45].

Despite the success in attracting audiences and driving narratives (e.g., “the midrange is dead” [3, 20, 50]), the analytic value of shot charts has seemingly been sacrificed in favor of storytelling and engagement. During an invited talk at the 2018 IEEE VIS conference, Goldsberry enumerated several reasons why NBA analytics departments have been slow to adopt visualization. Yet we have observed a trend that NBA teams and sports analysts have recently realized the importance of data visualization, although their employment of visualization has primarily focused on a mere showcase of their advanced metrics, with only a few exceptions (e.g., Buckets [4]) that employ interactivity for analysis. Based on our experience in the sports industry and visualization research, we argue that the insufficient realization of Goldsberry’s original vision is largely attributable to a lack of organic integration of analytics, visual representation, and interactivity. This motivated us to design and develop HoopInSight.

2.3 Visual Comparison for Information Visualization

Gleicher et al. [17] surveyed visualizations designed to support comparative tasks and proposed a general taxonomy of visual comparison. They classified visual comparison designs into three general categories: juxtaposition, which displays different objects separately; superposition, which overlays objects within the same space; and explicit encoding of the differences/relationships [16, 17]. Based on his framework [16], the targets we try to compare are the shooting performances of two entities (i.e., players/teams). The challenge lies in the complexity of individual items (multivariate spatial data) and their relationships. The strategy we adopt falls under the Summarize Somehow category — we compute the difference of each variable for each spatial unit. Our main visual technique for spatial comparison falls under the third category — explicit encoding of the relationships between two objects (i.e., shot maps). However, to combat decontextualization, we also include two reference objects in the system, making our system a combination of explicit encoding and juxtaposition.

3 UNDERSTANDING THE PROBLEM

3.1 Design As Domain Experts

As characterized by Sedlmair et al. [41], design studies in visualization research are projects that analyze “a specific real-world problem faced by domain experts” and “support solving this problem” with a visualization system. Collaboration between visualization researchers and domain experts is considered a “fundamental and mandatory part” of the framework. However, there are occasions when visualization designers/researchers are also well-versed in the domain for which they are designing visualization solutions. Our case fell into this territory where “the same person holds both roles” (visualization expert and domain expert), which Sedlmair et al. opted not to address in their design methodology article [41]. We characterize this distinctive design study method as Designing as Domain Experts (DaDE). In this project, after careful consideration, we ultimately decided to embrace and capitalize on DaDE for the following reasons:

(i) The first author is a domain expert in basketball journalism/analytics and a visualization researcher, and he has a comprehensive understanding of domain tasks and real-world data. The second author also has considerable experience in visualization research and is an enthusiastic and knowledgeable sports fan. (ii) The data we use are available through public APIs, thus, no necessity to collaborate with “data providers” exists. (iii) Potential stakeholders come from a variety of backgrounds, ranging from professional team management to sports analysts/journalists, as well as avid or casual fans who may or may not possess domain/data/visual literacy.

3.2 Basketball Shot Data and Shot Maps

The fundamental unit of data we are exploring is a shot taken by a player. Each shot is either made or missed, and we can aggregate these shots temporally (e.g., minute, game, season) and spatially (coordinates on the court). Within a particular spatial region and temporal window, two primary metrics will drive the analysis: frequency and efficiency. Our objective is to visually compare these and other metrics between two different situations or scenarios. For example, we may want to compare a player’s performance on shots from the “elbow” area [15] in the previous season to those same type of shots taken during the current season. More specifically, suppose he took 65 shots from that area in 2022 and made 27 (42%), but in 2023 he took 73 shots and made 28 (38%). The two important comparative metrics that we seek to communicate are that he took 8 more shots in the current year but his efficiency declined by 4%. Furthermore, we seek to do this for any players, teams, regions of the court, and time periods.

To understand current approaches to presenting this type of data, we collected and analyzed current shot map techniques and examples. Our collection comes from various sources, including research literature, renowned practitioners, and sports data websites.

Multiple methods for shot mapping exist, which differ in levels of aggregation and granularity, visual encoding (e.g., color scale, shapes), and filtering. Here we categorize these methods and explain how each aligns with a specific style of matching conventional geographic maps. Shot (Dot distribution maps): The Shot technique does not aggregate data; rather, it shows the actual location where each shot was taken, represented by a dot or other small glyph. Different colors or glyph shapes differentiate the binary results, indicating whether a shot was made or missed. A significant limitation of Shot is overplotting when a large number of shots is present. Therefore, Shot is often employed to show a small number of shots (e.g., shots taken in a single game). The
basic Shot technique is often straightforward and easy to understand for general audiences.

Cell (Proportional symbol maps): The Cell technique divides the court into small, predefined cells. It then aggregates the shots by cell and calculates different metrics on those aggregations. Cell typically uses some symbol (e.g., circle, rectangle, hexagon) to represent cell metrics, frequently employing symbol size to represent the shot frequency and color (usually diverging scale) to represent efficiency.

Zone (Choropleth maps): The Zone technique is commonly seen on data websites. It divides the court into larger, predefined zones, usually of varying sizes and shapes. While these divisions may vary across different websites, they typically follow domain conventions. Zone usually presents the aggregated efficiency for shots contained within each zone via the fill color of the region, but it does not represent the frequency. Consequently, Zone is easier for the general public to understand but holds less analytical value than Cell.

Others: Other less common shot map categories exist. For example, Contour line uses contour lines to depict the frequency of shots as a mountain landscape and Heatmap employs kernel density estimation to achieve the same, but with color intensity. Both Contour line and Heatmap usually present shot frequency but omit shot efficiency. They often support individual analysis of each scenario, as well as a shift in team strategy or playing style (DR1b at each location (scenarios (areas where the number of shots has increased or decreased across iterations from which shots are taken are often associated with players’ skill and can be aesthetically appealing. Such encodings are most useful in surfacing the most common places where shots were taken but may represent the data less precisely.

3.3 Task Analysis and Design Objectives

Since these different forms of shot charts have already become well-established industry standards, our goal is not to invent new chart variants but to focus on enhancing current capabilities and filling in the gaps of missing capabilities. In particular, we aim to: (1) enable more effective spatial comparisons, (2) support multiple analytical and comparative perspectives, and (3) communicate comparison within the context of original data.

Enable more effective spatial comparisons

Enable multivariate spatial comparison: The most notable analytical value of current shot maps lies in their ability to display two key metrics (Frequency and Efficiency) within a spatial context. We must similarly support exploration of these two key variables with respect to different spatial locations (i.e., positions on the court) for two different scenarios. We often support individual analysis of each scenario, as a comparison of differences and changes between the two scenarios. Ideally, viewers should be able to rapidly and clearly determine trends and patterns of the two variables across spatial locations.

Explicitly encode the differences: As reviewed earlier, Gleicher et al. [16, 17] categorize visual designs for comparison into three types: juxtaposition, superposition, and explicit encoding. People currently compare shot charts and similar sports data primarily by juxtaposing them as small multiples (e.g., [1,4]) beside each other and visually scanning back and forth. While this method is scalable, it can be cognitively challenging to compare multiple 2-D shot charts. The combination of many regions to examine and multiple variables to understand makes repetitive visual scanning to ascertain the differences quite difficult. We believe that explicitly encoding shot differences between two scenarios, that is, providing alternate views (DR2) that only represent the delta/difference, can further prepare and more direct analytical insights, which are difficult to uncover by simply scanning between individual shot maps. The above goals lead us to the first design requirement:

DR1 — Design explicit visual encodings that facilitate pre-attentive multivariate spatial comparison

This can be broken down into four sub-requirements. First, the locations from which shots are taken are often associated with players’ skill sets and teams’ game strategies. Therefore, we prioritize displaying areas where the number of shots has increased or decreased across scenarios (DR1a), as well as the specific increased/decreased quantities at each location (DR1b). Furthermore, the system should assist in detecting spatial patterns and trends in the data, which often signify a shift in team strategy or playing style (DR1c). Finally, although changes in efficiency may be inconsistent across neighboring locations, they can provide insights into a player/team’s shot-making ability, i.e., whether or not they improve their shooting performance from certain areas. Thus, the system also must visualize values and patterns of quantitative efficiency (DR1d).

Support multiple analytical and comparative perspectives

Current shot maps mainly provide two types of insights: where shots are taken and how efficient a player/team is at shooting. Together, they can disclose information such as whether a player/team took more shots in highly efficient areas. Comparison, however, can further lead to a wider range of analytical insights based on different references and scenarios. For example, by comparing a player before and after he was traded to another team, one can analyze the change in his roles. Comparing two players directly can yield insights into their different play styles. To support multiple analytical and comparative perspectives, the system must enable users to create different comparison scenarios, which leads to the second design requirement:

DR2 — Enable users to create different comparison scenarios

We can model the different types of comparisons via a comparison cross-table (entity × scenario). In this context, entity refers to players, teams, and opponents, while scenario involves lineups, shot types, time intervals, and other variables. This can be simply described as “comparing entities under different scenarios.” For instance, one could compare Michael Jordan’s (entity) shooting performance with and without Scottie Pippen on the court (lineup scenarios), or compare Jordan’s (entity) shooting pattern in his early career vs. later years (temporal scenarios). This design requirement can be divided into two sub-requirements: 1) obtain necessary data and structure the data to allow such filtering activities (DR2a); 2) design interactions to afford different entity selection and comparison scenario creation (DR2b).

Communicate comparison within the context of original data

An earlier goal emphasized the desire to explicitly encode comparison data. Only presenting an explicit comparison encoding (i.e., without the two source data views) can lead to decontextualization [17], however. Ideally, we seek to display comparison information as well as the original data from the two scenarios being compared. This consideration leads to the third design requirement:

DR3 — Present the two reference data sources (shot maps being compared) concurrently with comparison views to provide context and serve as devices for direct manipulation

We divide this requirement into three sub-requirements. First, the system should combine explicit encoding with juxtaposition and interactive connections to guide users’ attention and simplify analysis (DR3a). Second, the system should incorporate more visualizations to provide further context and additional metrics (DR3b). The insights extracted from these supplementary views can stand alone, but more importantly, they should assist users in identifying interesting subsets of the data or provide extensive insights. Finally, we aim to allow users to reconfigure comparison scenarios by directly manipulating supplementary views (DR3c).

4 HOOPInSIGHT

Our basketball shot visualization system is called HoopInSight and is driven by multiple raw datasets that we retrieved from the NBA API [36]. We merged and transformed the collected data into shot datasets grouped by players, teams, and seasons. Each data item (an individual shot event) contains 31 attributes, including the court coordinates (spatial), the result (numeric), the type (categorical) of each shot, team rotation data (textual), a video link connecting to the specific shot, and metadata. The data is updated on a daily basis. We also fetched supplemental data, such as player/team information. This data pipeline (retrieval, aggregation, and partition) helps us fulfill DR2a.

Within the system, users can select to examine a particular NBA player’s or team’s data and which season to display. We also support a special third focus entity (Opponents) that aggregates all opponents’ shooting data from games played against the selected team. This is a novel addition that enables users to analyze a team’s defensive
Fig. 1: HoopInSight user interface: It consists of three large columns: two Selection Views (S1 & S2) and the Comparison View (C). Each Selection View has a Shot Chart (SC1 & SC2) and multiple supplementary views (T, TR, ST). The Comparison View is divided into two sub-views (CV1 & CV2).

Fig. 2: The four visual encodings for a shot map (Stephen Curry 2022-23 season): (A) Shots, (B) Cell, (C) Zone, (D) Region

capabilities from the “shots allowed” perspective. Together, these capabilities partially fulfill DR2b (i.e., entity selection).

HoopInSight normalizes shot frequency by dividing the number of shots taken in a specific area by the total number of shots taken during the specified time period. We adopt the statistic, eFG% - Effective Field Goal Percentage [2, 34] to measure efficiency. eFG% adjusts for the fact that a 3-point field goal (shot made) is worth 50% more than a 2-point field goal. For example, if a player makes 4 out of 10 of his 3-point attempts, the traditional FG% (field goal percentage) is 40% while eFG% is 60%.

4.1 Interface and Visualizations

The user interface of HoopInSight (Figure 1) consists of three columns of views. Each column contains multiple visualizations. The two columns on the left and right of the interface are called a Selection View (Figure 1 S1 & S2). Each Selection View centers around a relatively standard Shot Chart (Figure 1 SC1 & SC2) that represents all the original shots taken by the selected entity under the specified scenarios. For example, Shot Chart 1 may show all of a player’s shots from the previous season, while Shot Chart 2 shows all the shots from this season. In between the Selection Views (Figure 1 C), in the center, is the Comparison View that employs two stacked sub-views to represent a comparison of the two Shot Charts, i.e., the difference/delta.

**Shot Chart (Figure 1 SC1 & SC2):** Fundamentally, this chart displays all shots taken by the selection, using one of four potential visual encodings that we implemented following the shot map typology categorized earlier in Section 3.2. The first, called the Shots encoding (Figure 2A), is a map of half the basketball court with a small point at each spot where a shot was taken. The color of the point indicates whether the shot was made (green) or missed (red). For the second encoding, Cell (Figure 2B), we divide the basketball half-court into 272 contiguous square cells (17 columns and 16 rows), each about 3ft (0.91m) wide and tall. We then aggregate all the shots within a cell and calculate the frequency and efficiency for that cell. We use circles to represent each cell, positioned at its center, with the size of the circle indicating the corresponding shot frequency and its color indicating the shot efficiency for that cell via a diverging color scale from orange (high efficiency) to blue (low efficiency). The third encoding, called Zone (Figure 2C), divides the half-court into meaningful zones known to basketball experts (e.g., left corner three, above the break 3, etc.). These zones are much larger than an individual cell of the Cell encoding and provide a chart roughly resembling a choropleth map. Again, all shots within a zone are aggregated. In this view, however, no circles are drawn. Instead, the fill color of the cell encodes the efficiency of shots
The central comparison encodings we explored did not choose. Designs C and D employ two sub-views with the top showing cells of frequency increase and the lower cells of frequency decrease.

via the same diverging color scale as in the Cell encoding. The Zone encoding does not represent frequency. Finally, the fourth encoding, Region (Figure 2D), is a hybrid of the Cell and Zone encoding. It contains the same cell circles as the Cell encoding, with circle size indicating frequency. Each circle’s color, however, is assigned based on the color of the zone its cell falls within. In other words, we use the zones as spatial filters to create smoother circle colors.

These four visual encodings provide insights at different levels of granularity: the Shots encoding allows users to see specific shots and their textual descriptions, as well as access a corresponding video clip. The Cell encoding offers visual representations of aggregated statistics in terms of efficiency and frequency. It also provides the basis for us to compare two selected targets. The Zone encoding allows viewers to focus only on efficiency — a simpler analytical task that fans tend to be more familiar with. The Region encoding provides a more balanced approach between Cell and Zone by still displaying frequency but with a smoother color encoding, as the efficiency tends to be more volatile across the cells. HoopInSight allows any of these four encodings to be used for the left and right Selection Views.

Comparison View (Figure 1C). The central Comparison View provides a visual comparison of the two displayed Shot Charts with a primary goal of communicating the differences (Δ) between the two selections, thus fulfilling design goals (DR1). As we sought to display the changes in both frequency and efficiency, we chose to compare the cells from the Cell shot encoding, and thus the Comparison View uses the same cell units. For each corresponding comparison cell, we use the data for S1 (D1) and the data for S2 (D2) to calculate the Δ (data for the comparison cell). The resulting Δ has three main attributes: its index on the court Δ(i,j), the difference in frequency ΔF, and the difference in efficiency ΔE. Δ(i,j) translates to the [x,y] position on the court, and the ΔF and ΔE are two quantitative variables we encode for each cell.

Our first challenge was to design a visual encoding that would be placed at each cell to represent the difference between the two corresponding original shot aggregations for that position. We experimented with using glyphs of one or more graphical shapes, and we conducted multiple rounds of “brainstorming” feedback sessions with our local visualization research group. During the first round, the authors introduced the motivations and analytical tasks and presented initial designs to elicit feedback. Other researchers provided suggestions on visual encodings or low-fidelity sketches. Multiple candidate visual encodings emerged, and we selected six and implemented each through a high-fidelity visualization prototype. We presented these candidates (Figure 3 & Figure 4b) again to our local group and gathered second-round feedback. One design (Figure 3A) used one consolidated view with two different shapes: circles denoted cells where frequency increased from Shot Chart 1 to Shot Chart 2, and diamonds denoted cells where the frequency decreased. The size of the glyph denoted the amount of frequency change and its color represented the change in efficiency between the two scenarios. A second design placed a four-quadrant axis on each cell, with a sector’s appearance indicating one of the four results (increasing/decreasing × ΔF/ΔE) and the radius indicating the quantity. With both of these designs, viewers felt that each glyph was comprehensible, but it was relatively difficult to discern trends and patterns across the entire view of glyphs showing all cells (DR1e).

A third design (Figure 3C) employed two sub-views, one (upper) with glyphs for cells with increasing frequency between the scenarios, and the other (lower) for cells with decreasing frequency. The design used circles for the glyphs with size and color encoding deltas in frequency and efficiency, respectively. Viewers felt that this design was too similar to the circles in the Shot Charts, however, which caused confusion. Another design (Figure 3D) substituted “slopes”, i.e., tilted lines, for the circles with angles of the line representing frequency differences. Although this design helped to show group patterns (DR1e), viewers felt it was too difficult to understand and decode the angles representing frequency differences DR1b. We also experimented with animations to denote the difference between two corresponding cells but found the animations unsuitable for analytical reasoning.

Ultimately, feedback helped us to narrow down to two promising visual encodings, which we now include in HoopInSight and allow users to select. We call the two designs that emerged Arrowhead and Needle (Figure 4). In each, a single glyph (a triangle or a sloped line, respectively) is drawn at each cell to encode the delta between the two selections. The first variable ΔF, the change in frequency, is encoded in the Arrowhead via the height of each triangle, with taller triangles indicating larger changes in frequency. The width of each triangle’s base represents the average of the two frequencies, thus showing wider triangles in regions where the player tends to take more shots. For example, in one cell, the number of shots taken in the two Shot Charts may have increased from 70 to 80, while in another, the shots increased from 10 to 20. Both of the corresponding triangles will have the same height because the delta between the two selections is +10 in both
cases. However, the triangle in the first case will be wider and larger overall, helping to indicate that this is a region where the player shoots frequently. The change in efficiency between the two regions, Δg, is denoted by the color of the triangle. We employ a diverging color scale, with a greener fill color indicating increased efficiency from selection 1 to 2 and a redder one signifying decreased efficiency.

In the Needle glyph technique, the length of the triangle denotes the change in shot frequency at the location and the thickness of the line denotes the average of the two frequencies, with thicker lines indicating regions where the player takes more shots. The color of each line encodes the difference in efficiency between the two cells.

Both designs also employ the idea of two sub-views, as was done in the earlier alternative designs shown in Figure 3C&D above. The top sub-view indicates all cells where the shot frequency increased from the first to the second selection, while the bottom sub-view indicates cells where the shot frequency decreased. Additionally, triangles in the upper Arrowhead sub-view (frequency increase from Selection View 1 to 2) are drawn pointing upwards, while those in the lower sub-view (frequency decrease) point downwards. Similarly, lines in the upper Needle sub-view slope to the upper-right, while those in the lower sub-view slope to the upper-left. We feel that this separated view makes the frequency and efficiency differences between the two periods more pre-attentively clear, enabling rapid determination of areas of shot increases (glyph locations on the court in the upper view) versus shot decreases (glyph locations in the bottom view), fulfilling one of our primary design goals (DR1a).

**Supplementary Views:** The two Selection Views also include multiple other visualizations that help to encode more variables about the shots in the two selections and serve as instruments for users to interact with and perform informed filtering, thus fulfilling DR3b&c.

On the top and right edges of the two Shot Charts are histograms showing aggregated statistics of the relevant row or column of regions on the half-court. The height of each bar represents the frequency, and color encodes the efficiency for the respective row or column.

To the left of each Shot Chart is a zoomable treemap, Shot Tree (Figure 1 (H1&T1)), that displays the frequency and efficiency statistics for various shot types (such as pull-up jump shots, floaters, and layups) and different shot zones (such as the left corner three-pointer, midrange, and restricted area). Below each Shot Chart is a traditional statistics table (Figure 1 (H1&T2)) that presents a higher-level statistical summary of the treemap. Below the table is Season Trend Chart (Figure 1 (ST1&S1&T2)), displaying aggregated game-based statistics over the course of the selected season with time moving from left to right. The upper line chart shows efficiency throughout the season, while the lower bar chart shows the number of shot attempts in each game.

### 4.2 Interaction

HoopInSight supports rich interactions that enable users to explore and create different comparison scenarios. Below we list some of the different styles of interaction [51] provided by the system.

**Filter:** The most prominent interaction in HoopInSight is filtering. Above the Shot Chart views are filter panels (Figure 5a) that enable users to filter shots (and thus make a selection) based on the opponent team, team lineups, and opponent lineups. For instance, users can apply filters to generate visualizations that represent one team’s shot map with or without a particular player, an opponent’s shooting patterns with or without a particular defensive player on the floor, or an individual player’s shooting pattern when he plays against a particular team.

The treemaps to the left of each Shot Chart also serve as categorical filters (Figure 5b) that support selection by shot actions and shot zones. For example, users can view the shot map of a particular player’s “step-back” jump shots. The Season Trend charts below the Shot Charts also serve as a temporal filter (Figure 5c) using brushing, allowing users to select any portion of a season that interests them.

Another important filtering operation provided by the system is freehand/lasso selection (Figure 8a). When using the Shots encoding in a Shot Chart, users can freehand draw a polygon to select an area of interest. HoopInSight aggregates the shots within the polygon and calculates corresponding statistics. The background of the selected area transitions into a color reflecting the shot efficiency within it. Hovering over the selected area brings out an on-demand view that includes a shot action treemap, a sunburst chart for an assist breakdown of the selected shots (i.e., which players assisted the made shots taken from the selected area), and a statistical summary of shot actions. When the focus is on team shooting performance, this view additionally displays the top 5 players who took the most shots in the selected area, along with their statistical summaries. The filtering interactions fulfill the design goals to facilitate scenario creation (DR2&DG3e).

**Select & Elaborate:** HoopInSight generally provides details on demand through hovering. For instance, hovering over each cell in the Comparison View displays a tooltip that shows corresponding statistics regarding frequency and efficiency.

**Connect:** HoopInSight provides a connect style interaction to link related cells and zones. Hovering over cells/zones in one view will simultaneously highlight the identical areas in other views (DR3a).

**Encode:** HoopInSight allows users to toggle between the four visual encodings for the Shot Charts and the two visual encodings for the Comparison Views. Also, hovering over legends in the Comparison Views can highlight the cells where efficiency increased/decreased.

In summary, the visual technique enables users to “see the differences” between two scenarios, whereas the role of interactivity lies in expediting the construction of various analytical scenarios and accessing contextual information and corresponding details on demand.

### 5 Case Studies

This section presents two case studies to illustrate the analytic capabilities of HoopInSight.

#### 5.1 Analyzing Nikola Jokic’s Offensive Impact

Nikola Jokic is the star player of the Denver Nuggets and has been named the NBA’s Most Valuable Player (MVP) twice. Recently, basketball analysts/journalists and thousands of fans have been engaged in a heated debate about whether he should win a third consecutive MVP award. While most model-derived advanced metrics point towards yet another historic performance from him, many traditional analysts/journalists have expressed skepticism toward such advanced metrics that often lack explainability and interpretability.
Perspective I: Individual shooting performance — Jokic is extremely efficient and consistent; he is also versatile on offense

To understand Jokic’s impact on offense, HoopInsight first supports analyzing his individual shooting performance. After selecting Jokic’s shooting profile from the current season, the Selection View displays the pertinent metrics (Figure 6). Illustrated by all the orange in the views, it becomes apparent that Jokic is an extremely efficient scorer. The Shot Map (Figure 6A) provides insights into Jokic’s most frequent shots — they are all over the box, with a moderate number of three-pointers from above the break. The Treemap (Figure 6B) reveals that as a center, Jokic rarely ends the offense with a dunk but instead takes many floaters and old-fashioned hook shots, implying that Jokic relies heavily on his excellent touch rather than athleticism. The Season Trend Chart (Figure 6C) shows that Jokic is consistently efficient throughout the season, with the exception of a few games.

Perspective II: Comparing Jokic with his counterparts — Different styles and efficiency

Another angle to analyze Jokic’s individual shooting performance is to directly compare him to his counterparts in the current MVP race — Giannis Antetokounmpo and Joel Embiid. Again, the Selection Views allow us to examine their overall efficiency throughout the season and their individual shooting performances. But it is the Comparison View that makes their distinctive shooting patterns extremely clear – Antetokounmpo is more dominant in the restricted area, where he shoots more frequently just under the rim and is even more efficient than Jokic, who shoots more frequently elsewhere and is significantly better away from the basket (Figure 7A). Jokic is also a better three-pointer shooter, especially from above the break. Compared to Embiid (Figure 7B), Jokic is not as dominant under the basket but is almost more dominant anywhere else in the box. While Embiid enjoys taking mid-range shots near the free-throw line, he surprisingly is not as accurate as Jokic. Two years ago and has thrived this season. The light green triangles (Figure 8c left) immediately reveal that Gordon has taken more shots in the restricted area with easy layups and dunks this year than he did two years ago with the Orlando Magic, indicating he has much better opportunities to score. One reason for this might be that Gordon has simply improved. To explore that notion, one can compare Gordon’s same-season performance and filter the data based on whether Jokic is on the court. The large green upward triangles (Figure 8c right) quickly reveal that Gordon shoots more near the rim while playing alongside Jokic — the frequency of his dunks has doubled, largely due to Jokic’s ability to create opportunities for him. This double-confirms Jokic’s positive impact on Aaron Gordon’s offensive metamorphosis.

5.2 Compare Big Men’s Defensive Influence

One area of analysis that existing shot maps rarely support is defense. Measuring the defensive impact of an individual player is notoriously difficult. Although advanced statistics and analytics can provide helpful aggregated information, they rarely allow for in-depth analysis, let alone insights into the intricate defensive schemes professional teams employ. Although HoopInsight does not provide a direct answer, it can offer insights from the “shots allowed” perspective and enable analysts and coaches to test their hypotheses or ask insightful questions. One usage scenario is to compare opponents’ shooting performance/patterns when a player is on the court versus when he is off the court. In this case (Figure 9), we compare the differences between teams’ opponents when a big man (center position) is on or off the court. Here we select four different scenarios.

Figure 9A represents the difference in shots taken against the Utah Jazz during the 2020-2021 season with and without their Defensive Player of the Year award winner, Rudy Gobert, on the court. This example, Selection 1 represents when Gobert was off the court and Selection 2 represents when he was on the court. Figure 9B-D follow the same procedure, but the focus teams are the Hawks, the Bucks, and the Mavericks from the current season, and the focus players are Clint

Fig. 6: Perspective I: Jokic’s individual shooting performance. Note: We add external annotations using white text and arrows on the figures in this section to aid explanation. These annotations are not part of the system.

Fig. 7: Perspective II: Compare Jokic with counterparts. A - Jokic (S2 on right) compared to Antetokounmpo (S1 on left). B - Jokic (S2) compared to Embiid (S1)
were "forced" by Utah's defense with Gobert anchoring it. The green
when Wood was on the court, opponents were seemingly encouraged to
with Jokic (S2 - right) vs. (b) Nuggets' performance
after their opponents' pick & roll actions [24].
abilities to get through screens and contest the mid-range jump shots
Utah Jazz gave up decreased (red triangles), suggesting those shots
nuanced insights. For example, the efficiency of the midrange shots
opponents' attempts in the midrange area, HoopInSight could reveal more
fewer shots from the two corners, suggesting that limiting penetration
attack the rim. Additionally, when Capela was present, opponents took
fewer mid-range shots and fewer and also less-efficient shots
opponents took more shots, while the bottom ones illustrate where their
is arguably the opposite. The top four sub-views show where their
players are well-known rim protectors, while the fourth player, Wood,
attack the rim. Moreover, when Capela was present, opponents took
more mid-range shots and fewer and also less-efficient shots
opponents took more 3-pointers and fewer mid-range shots.

6 Reflection and Discussion
6.1 Design Considerations For Spatial Comparison
    Consideration I: Different comparison granularities
    Drawing from Gleicher’s insight that "comparison is more than just
    finding differences" [16], we argue that in our case, spatial comparison
    is more than just finding individual spatial differences. Endert et al. [12] discussed three levels of visual aggregations (i.e., glyph level, multiple glyphs merge, and the aggregation of multiple glyphs into a field) on large displays and how different encodings would impact such aggregations. For HoopInSight’s comparison view, our goal was to facilitate recognition of group trends and patterns, while still maintaining the interpretability of the two variables associated with individual cells.

    This goal fundamentally determined how we chose glyphs. Superimposing glyphs on a map is a common way to visualize multivariate spatial data. [30, 46]. Some of our initial visual encoding designs, as well as later suggestions from colleagues, used more complex glyphs to encode more variables. For example, while the four-quadrant axis design we used (Figure 3B) can display individual differences and encode more variables in one cell, identifying group patterns can be problematic due to visual clutter. Similarly, using multiple color scales to encode different variables can also be visually confusing and impede our perception of group patterns.

    We recommend that designers consider the granularity of insights they want to convey and the priority of different levels. In our case, if the primary goal was to compare the multiple variables between individual cells, complex glyphs would likely be more effective. Conversely, if our primary goal was to communicate group patterns without worrying about individual interpretability, it would be reasonable to use a metric that aggregates individual cells and employ more perceptually dominant encodings (e.g., colors) to illustrate the patterns.

    Consideration II: Neighboring spatial data characteristics
    Design choices also depend on the strength of the connections between the data associated with individual cells and their adjacent cells. These connections may be based on factors such as whether the cells in a neighborhood exhibit roughly similar values or whether the delta follows a certain pattern. A strong connection indicates the existence of a "flow" — the data itself is inherently spatially dependent (such as wind maps or ocean current maps).

    Using U.S. election maps as an example [5, 14, 31], the voting behavior of a county tends to be strongly connected to the ones of its neighbors. The glyphs representing such data tend to merge together and form a group pattern. In such cases, occlusions are less likely to occur and less troubling. In fact, visualization designers can utilize this congruence to amplify patterns and even create affective influence, as demonstrated by The Washington Post’s political wind map [31]. However, when the data is less consistent between neighbors, using the same visual encoding may result in greater occlusion, making it challenging to reveal patterns. It might be more appropriate to apply less occlusion-prone encodings or adopt our approach — using two separated sub-views for increases and decreases, respectively.

6.2 Scalability and Transferability
By combining shot location data with further relevant information, HoopInSight can easily extend to provide more cutting-edge analytical insights. For instance, coupling data about team strategies, either of-
The right figure showcases when a spatial filter is applied. The sizes of
in-depth and comprehensive analysis. It also possesses great potential
active features and new visual comparison techniques can foster more
audience engagement. However, we believe that the addition of inter-
on developing novel metrics and the narrative approach that emphasizes
established solution. As previously noted, innovations in this field are
: In basketball, shot maps are a well-
costs between multiple roles.

collaborators, especially sports analytics companies, tend to care more
similar to what Oppermann and Munzner described in [35]. Second,
accessible data allows for early idealization and data sketches. This is
further share our reflections on adopting DaDE.

6.3 Reflection on Designing as Domain Experts

Earlier, we outlined the reasons for adopting the Designing as Domain
Experts (DaDE) method, such as adequate domain knowledge and
available and accessible data. We believe that this method can yield
significant benefits, but it does also incur dangerous pitfalls. Here, we
further share our reflections on adopting DaDE.

Benefits

More manageable process: While a productive collaboration can bring
many benefits, in reality, managing external interactions with collabor-
orators is rarely easy. Sedlmair et al. enumerate a dozen pitfalls that
may occur in the winnow and cast stage of a design study [41]. DaDE
can provide a faster design process by avoiding the significant time
external collaborations require. DaDE starkly contrasts our previous
Collaboration with two sports data companies. First, the available and
accessible data allows for early idealization and data sketches. This is
similar to what Oppermann and Munzner described in [35]. Second,
collaborators, especially sports analytics companies, tend to care more
about “just visualizing my data” (engineering) than “what visualization
can do differently” (research). Adopting DaDE allows us to manage
data acquisition and problem formation and reduce communication
costs between multiple roles.

Focusing on novel solutions: In basketball, shot maps are a well-
established solution. As previously noted, innovations in this field are
mainly focused on two directions: the modeling approach that focuses
on developing novel metrics and the narrative approach that emphasizes
audience engagement. However, we believe that the addition of inter-
active features and new visual comparison techniques can foster more
in-depth and comprehensive analysis. It also possesses great potential
to be integrated with other approaches. We observed that the afford-
dance and capabilities of interactive visualization have yet to be fully
grasped by stakeholders. They also tend to exhibit conventional tool
attachment [47]. As such, a traditional design study involving stake-
holders tends to result in projects that facilitate existing tasks rather
than challenging the common approaches. DaDE enables us to focus
particularly on analytical tasks that are beyond existing approaches.

Synergy of domain knowledge and visualization expertise: Expert
designers can use their knowledge in both fields to advance domain
tasks through visualization solutions. In our case, our domain knowl-
edge includes an understanding of high-level domain problems (what
tasks are valuable to support) and a familiarity with the characteristics
and availability of underlying data (what tasks can be supported). It
leaves the flexibility to enter a design iteration from either data ab-
straction or task abstraction and rapidly match data and tasks. Our
expertise in visualization serves as a catalyst for connecting tasks with
data, creating a synergistic effect. It helps break down high-level do-
main goals into low-level tasks that can be supported by interactive
visualization, as well as extract feasible tasks from multiple datasets
and transform the data into a visualizable form. We feel such synergy
is valuable, especially when the goal is to expand existing task space.

Pitfalls

Narrow domain perspectives: The term domain expert can be hetero-
geous in real life. Simply differentiating domain expert from casual
user or even visualization expert does not take adequate consideration
of the heterogeneity of domain experts. In our case, different types of
domain experts exist, such as sports journalists, sports analysts (data-
driven or traditional “eye-test”), coaches/scouts, etc. While they share
similar goals in analyzing shooting performance, their backgrounds
and interests can vary drastically, which may affect their perception
and adoption of such tools. Expert designers should be aware of their
own positional biases and avoid a narrow domain perspective.

Prioritizing experts over casual users: Although prioritizing domain
experts as end users may not necessarily be a pitfall, it is worth noting
that DaDE naturally caters to them if the aim is to solve complex
domain challenges. The resulting visualization interface will likely
be sophisticated. Furthermore, as evidenced by our case studies, the
insights provided by the visualization require strong domain knowledge
to perceive and interpret. For example, casual users who are less
familiar with the term “pick & roll” and different strategies to defend
against it are less likely to acquire insights from Figure 9.

Task/data explosion: Task/data explosion can occur when the domain
problem is unconstrained and the domain is data-rich. In such cases,
the iterative cycle of data and tasks may spin out of control. This means
that expert designers may keep pursuing additional tasks and data to
address the domain problem in a more exhaustive way. While this
may not necessarily be a pitfall, if the goal is to develop a functional
product to address the domain problem, researchers should be aware of
(visualization research) contribution saturation when new features are
more engineering.

Although DaDE is not a common method for design study, we
believe our situation is not a lone case. Domain experts may play
varying roles in the visualization design study process. We advocate
for research to explore these benefits and pitfalls more systematically
and provide methodological guidance for successful design endeavors.

7 Conclusion

This work aimed to enhance and expand the analytical capabilities of
the shot chart, a well-known visual technique for basketball spatial
analytics. To achieve this goal, we utilized our expertise in both visual-
ization and basketball analytics and developed a novel visual technique
for spatial comparison, along with an interactive visualization system
called HoopInSight. The system supports a wide range of comparative
analyses by allowing convenient selection of comparison entities and
rapid creation of various comparison scenarios. Ultimately, we shared
our reflections on and discussion about design considerations for visual-
spatial comparison, the scalability and transferability of our approach,
and adopting designing as domain experts (DaDE) methodology.
SUPPLEMENTAL MATERIALS

Submitted supplemental materials for this paper include a document with enlarged figures, a supplemental glossary explaining the domain terms, and two videos: a main system introduction video with one in-depth case study and a secondary video for more case studies. The system is available at: https://hoopinsight.netlittlapp/

REFERENCES


