

# SnapShot: Visualization to Propel Ice Hockey Analytics

Hannah Pileggi, Charles D. Stolper, J. Michael Boyle, and John T. Stasko, *Senior Member, IEEE*

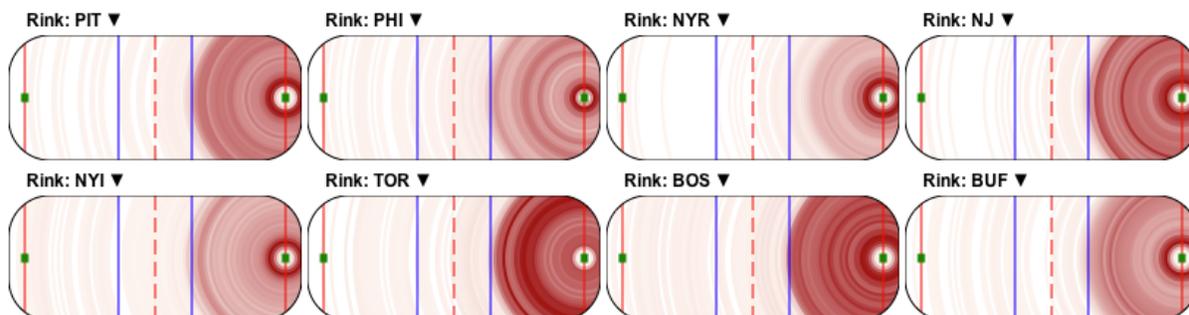


Fig. 1. Small multiples created by Analyst A of eight radial heat maps displaying patterns of shot lengths at each of the indicated rinks during the 2010-2011 NHL Regular Season.

**Abstract**—Sports analysts live in a world of dynamic games flattened into tables of numbers, divorced from the rinks, pitches, and courts where they were generated. Currently, these professional analysts use R, Stata, SAS, and other statistical software packages for uncovering insights from game data. Quantitative sports consultants seek a competitive advantage both for their clients and for themselves as analytics becomes increasingly valued by teams, clubs, and squads. In order for the information visualization community to support the members of this blossoming industry, it must recognize where and how visualization can enhance the existing analytical workflow. In this paper, we identify three primary stages of today’s sports analyst’s routine where visualization can be beneficially integrated: 1) exploring a dataspace; 2) sharing hypotheses with internal colleagues; and 3) communicating findings to stakeholders. Working closely with professional ice hockey analysts, we designed and built SnapShot, a system to integrate visualization into the hockey intelligence gathering process. SnapShot employs a variety of information visualization techniques to display shot data, yet given the importance of a specific hockey statistic, shot length, we introduce a technique, the *radial heat map*. Through a user study, we received encouraging feedback from several professional analysts, both independent consultants and professional team personnel.

**Index Terms**—Visual knowledge discovery, visual knowledge representation, hypothesis testing, visual evidence, human computer interaction.

## 1 INTRODUCTION

Sports are inherently competitive. Teams seek championships in their leagues and players strive to be the highest ranking athlete at their positions. In professional sports, the money flows to those who win. Winning dictates player salaries, team salary caps, and prioritizes television airtime.

In sports such as baseball, game data has both been collected and studied for over a century. An increasing number of professional sports teams are incorporating statistical analysis of game data into their organizations’ strategies. The use of analytics is cited as a major factor in the American League record 20-game win streak of the 2002 Oakland Athletics of Major League Baseball and the National Basketball Association Championship of the 2011 Dallas Mavericks [12, 9].

Although data analytics are becoming much more prevalent throughout sports, data visualization is still relatively unexplored. This surprises us because sports data lends itself well to visualization. A timeless example is Baseball Hall of Famer Ted Williams’s and John

Underwood’s 1969 strike-zone graphic accompanying the *Sports Illustrated* article “The Science of Batting” [28]. The iconic graphic places Williams himself next to a theorized map of pitch locations, colored according to his perceived success in hitting a pitch in that location.

A few teams in the National Hockey League, the highest league of professional ice hockey in North America, have begun to utilize analytics, but not visualization. There is evidence that visualization could be embraced by these teams. As Dan McKinnon, Director of Personnel for the Pittsburgh Penguins, shared with us, “We are always seeking out advantages to improve our strategy and thus our performance.”

In cooperation with a professional hockey analyst, we have begun to explore whether visualization could be advantageous for competitive hockey strategy. We utilized a user-centered methodology for designing a system, with our consulting analyst representing the “user”. As a group, we focused our efforts on three potential stages where visualization could integrate into the analyst’s workflow:

1. **Exploration** of a dataspace.
2. **Discussion** of hypotheses amongst analysts.
3. **Presentation** of findings to stakeholders

To support these three tasks, we have designed and built SnapShot, a system for exploring, discussing, and presenting hypotheses and findings using National Hockey League shot data. SnapShot displays the 72,926 regular-season, regulation-time shots that were taken during the 2010-2011 season. Our professional analyst identified the five most crucial variables of the available data to his investigations: the length of the shot, whether the shot was a goal or not, whether the

- Hannah Pileggi, Charles D. Stolper, and John T. Stasko are with the School of Interactive Computing and the Gvu Center at the Georgia Institute of Technology, e-mail: {hannah.pileggi,charles.stolper,stasko}@cc.gatech.edu.
- J. Michael Boyle is with the Sports Analytics Institute, LLC, e-mail: mike@sportsanalyticsinstitute.com.

Manuscript received 31 March 2012; accepted 1 August 2012; posted online 14 October 2012; mailed on 5 October 2012.

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

shooter was on the home team or the away team, the shooter’s team, and the location the shot was taken from on the ice. SnapShot’s visualizations allow an analyst to flexibly explore these variables on the entire data set or a subset of it.

We conducted an evaluation of SnapShot with three professional hockey analysts. Two were partners for independent consulting firms and the third was the Director of Team Personnel for a National Hockey League team. These evaluations captured valuable case studies of hockey analysts’ exploration processes. Feedback from each session also provided encouraging feedback that there is interest in capitalizing on visualization in this domain-specific dataspace.

The contributions of this paper are as follows:

- An analysis and description of a professional sports analyst’s workflow to inform sports visualization system design,
- The SnapShot system for visualizing hockey shots taken during the 2010-2011 National Hockey League season,
- Three case studies of professional ice hockey analysts incorporating visualization into their analysis.

## 2 RELATED WORK

The use of visualization to help fans, team officials, and analysts better understand sports data has grown in popularity recently. Many static infographics about sports data can be found on various websites and in sport publications. More comprehensive interactive visualization systems are less common, however. In the information visualization community, a few projects and systems have been developed including the use of treemaps to summarize tennis match results [11], sparkline-style visualizations to show baseball [25] team performance with different pitchers [5], and flexible representations of tournament brackets/ladders [22].

Major League Baseball (MLB) and the National Basketball Association (NBA) have pioneered the gathering, analysis, and visualization of game data acquired through computer vision technologies applied to video. Sportvision’s PITCHf/x [19] is a system for automatic statistical collection in baseball parks using two cameras mounted in the stadium to track the speed and location of a pitched baseball. Using PITCHf/x, statistics such as the pitcher with the fastest fastball or the pitcher with the sharpest-breaking curve can be analyzed.

A well-known example of baseball data visualization utilizing the PITCHf/x data is a video from the New York Times New Media Desk highlighting the pitching tendencies of Mariano Rivera, one of the best closing pitchers due to his “signature cutter” [4]. The NY Times video uses heat maps of Rivera’s pitch locations to explain why his pitches are so difficult to hit and why he is such a successful pitcher.

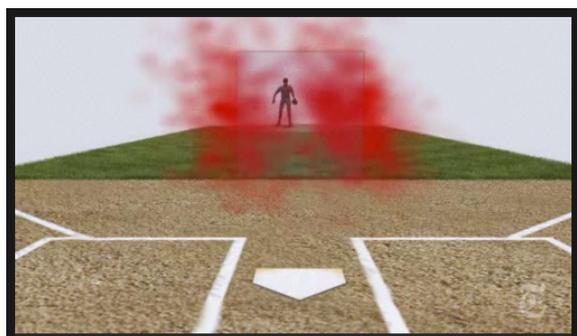


Fig. 2. Frame captured from “How Mariano Rivera Dominates Hitters” featuring a heat map of a his “signature cutter.” Image: *The New York Times*, June 29, 2010 © 2010 *The New York Times*. Used under license.

The company Stats LLC is using computer vision technology called SportVU [20] to capture movement and game action data from the National Basketball Association (NBA). Using this data, Goldsberry [7] created an application for visualizing several seasons worth of NBA data as a complex heat map called CourtVision (Figure 3). The heat

map regions are sized by shot attempts and colored by points per attempt, resulting in a series of court heat maps that help compare different teams and explore conventional wisdom about the teams.

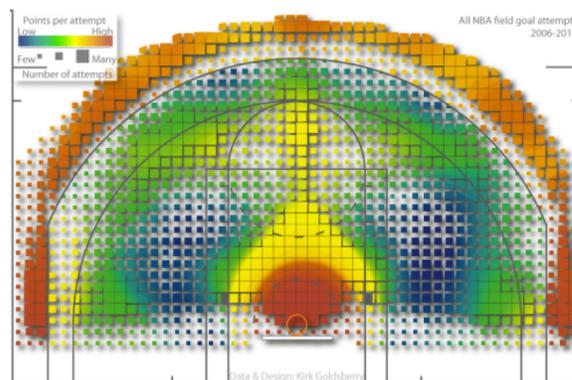


Fig. 3. Goldsberry’s heat map of NBA field goal attempts over a five year span [7]. Glyph size represents the number of shot attempts and color represents points per shot attempt. Image courtesy of Kirk Goldsberry.

Another basketball project [13], built with Pat Summitt’s mantra that “Rebounding wins championships” in mind, utilizes the SportVU optical tracking data in a multi-view system. Figure 4 shows two of the views, pairing scatterplot and heat map visualizations of the spatial distribution data for on-court rebounding.

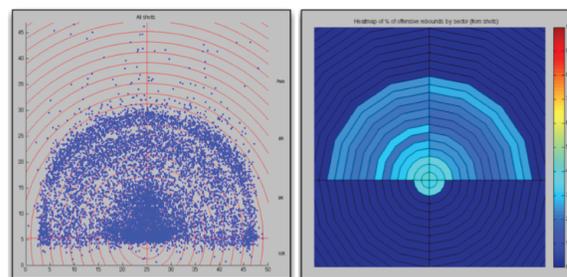


Fig. 4. Two sample views from Maheswaran et al.’s system for visualizing basketball rebound data [13]. The left window displays the spatial distributions of rebounds and the right window displays a heatmap of offensive rebound rates. Image courtesy of Rajiv Maheswaran.

Soccer (called Football by most countries other than the United States) is similar to ice hockey in terms of game structure and style. Games in both sports have a continuous flow with passing and player position structured by physical aspects of the pitch (or rink). Because soccer is so popular in Europe, it has garnered significant analytical exploration and a growing body of work has been published regarding performance improvement systems, many including visualization. One such example is a system called Soccer Scoop [15] [16] (Figure 5), which pairs a statistical tool with the ability to visualize goalkeepers through glyphs. Team managers can use the tool to visualize one goalie between games, or compare two separate goalies to find opportunities for improvement.

The Attribute Explorer [18], a well-known general purpose information visualization system, was adapted in 2008 to display soccer game events[1] (Figure 6). The tool was used for observing single matches and for collecting game events cumulatively for later analysis through coordinated views.

Most of the analytics work being done for ice hockey does not include visualization capabilities. A few hockey analysts have used systems such as Tableau [21] and Spotfire [23] for selected presentations of analysis data. Such visualizations obviously are general purpose, however, and do not show the data in the context of a rink.

Beyond in-house team analytics and hired consultants, news outlets and their web affiliates provide visual game data for consumption

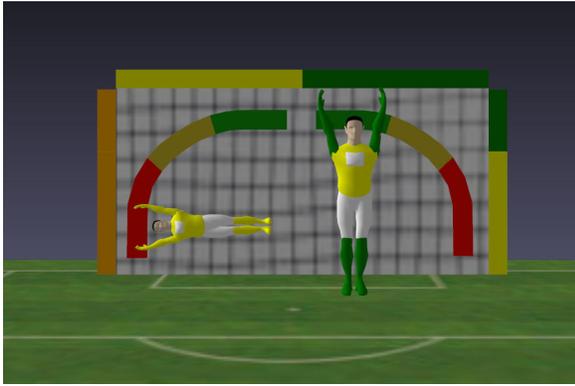


Fig. 5. Output image from Soccer Scoop comparing a goalkeeper's performance at home (left) versus away (right) games [15] [16]. Image courtesy of the Adrian Rusu.

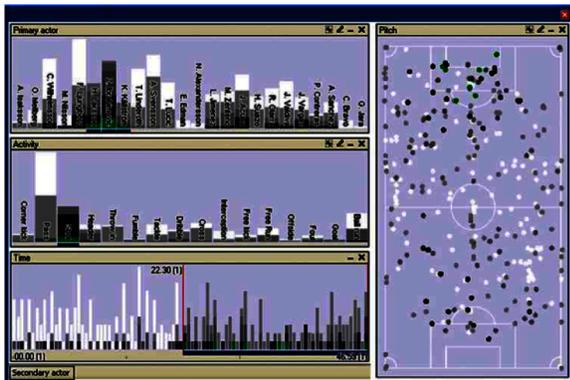


Fig. 6. Screenshot of soccer data in Attribute Explorer with bar charts on the left corresponding to (top to bottom) player, action, and time, with the pitch represented on the right with each instance plotted as a dot [1]. Image courtesy of Sage Journals.

by fans and viewers. The two most relevant examples of such web-based NHL data sharing are NHL.com's Ice Tracker (Figure 7) and ESPN.com's GameCast (Figure 8).

The National Hockey League created the Ice Tracker visualization system [14] for displaying events as they occur in real-time throughout a hockey game, displaying each event as a small glyph on a rink as well as on a timeline. If the event has a film clip associated with it, the glyph reflects this and allows the user to view the clip. The system provides filtering to display shots, goals, hits, penalties, and fights on the ice.

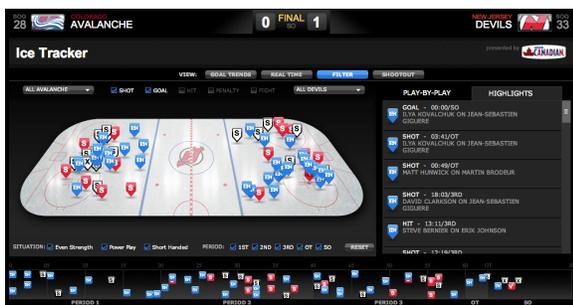


Fig. 7. The NHL.com's Ice Tracker system visualizes a recent game featuring the Colorado Avalanche at the New Jersey Devils. Image from <http://www.nhl.com/ice/icetracker.htm?id=20110206151>.

ESPN hosts a similar visualization system called GameCast [6]. GameCast differs from Ice Tracker in that it normalizes events on the ice so that the away team's events are on the left end and the home team's events are on the right end and the glyphs displayed are in the

relevant team's color scheme. Like Ice Tracker, GameCast can run in real-time during games or display a recap of a game.

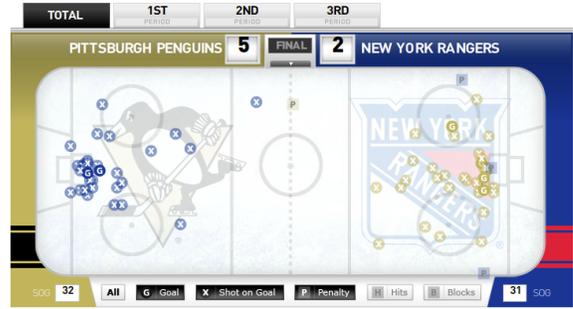


Fig. 8. The ESPN.com's GameCast system visualizes a recent game featuring the Pittsburgh Penguins at the New York Rangers. Image from <http://scores.espn.go.com/nhl/gamecast?gameId=400047896>.

The audience for both Ice Tracker and GameCast is not a team or organization, but rather the public at large. While the full visualization of a single game is available after it has concluded, neither system allows games to be combined for longer-term analysis, and thus neither are utilized by professional analysts in the course of their work.

### 3 UNDERSTANDING THE PROBLEM

Every strategical advantage increases a professional hockey team's chances of winning. In order for visualization to be one of these advantages, we set out to understand how analysts offer value to teams and where visualization can augment their workflow.

#### 3.1 Ice Hockey

Ice hockey is a sport played on an ice rink by two teams with two nets, a single puck and three or more referees. The objective of the game is for a team to score more goals than their opponent. Players are equipped with protective gear, skates and a hockey stick and score goals by shooting the puck into the opposing team's net using their sticks. A team is required to have six players on the playing surface at all times with the exception of when certain types of penalties are being served. At most times during the game a team has a one goaltender, two defensemen and three forwards on the ice. The forward positions subdivide into left wing, right wing and center. The rink itself (Figure 9) is divided into three zones: the offensive zone, neutral zone, and defensive zone. The zones are divided by *blue lines*, and there are also *red lines* at each goal and a center line at middle-ice in the center of the neutral zone.

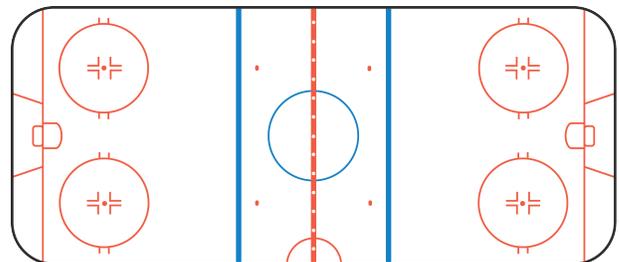


Fig. 9. Mandatory NHL rink markings.

Games consist of three 20-minute periods. When the score is tied at the end of regulation time, it is typical to have one or more sudden death overtime periods to decide the winner of the game. In some leagues/tournaments, there is a limit to the amount of overtime played before the game is decided in a shootout during which teams trade penalty shots until a winner is decided. Alternatively, some leagues/tournaments do not require an outright winner and leave games tied after regulation time and a set amount of overtime.

There are several elite hockey leagues in a handful of regions/countries. In North America, the most elite professional league

is the National Hockey League (NHL). In Eurasia, the most elite professional league is the Kontinental Hockey League (KHL). At a national level, there are several elite professional leagues including but not limited to the Elitserien (Sweden) and SM-liiga (Finland). Each of these top leagues has a network of minor professional and elite amateur leagues (for example, the CHL in Canada, and the NCAA in the United States).

### 3.2 NHL Hockey Data

The analyst with whom we developed SnapShot was specifically focused on the North American National Hockey League shot data. Hockey information has quirks. Some of these quirks are pervasive in all most competitive team sports: teams swapping ends of the rink each period and the importance of whether a team was home or away. Some of these quirks are more pronounced in hockey data. Hockey shots are taken on a rink with fixed dimensions and standardized reference lines which factor heavily into the game rules. Non-overtime rules, regular season overtime rules, and overtime playoff rules all differ from each other. The rules change during play based how many men are off the ice when the shot was taken due to penalties. Finally, while every team plays the same number of regular season games, sixteen of the thirty teams in the league also play in the postseason, each round of which can be up to seven games. (In 2011, the champion Boston Bruins and runners-up Vancouver Canucks each played in twenty-five more games than the playoff-missing Columbus Blue Jackets.)

The data set our analyst provided consists of 81,158 data points, representing each shot officially recorded by the NHL's stat keepers during the 2010-2011 season and distributed by the NHL to analysts. Shots are a specific instance of what the NHL refers to as *events*, which also includes but is not limited to hits, penalties, and faceoffs. Currently, events recorded in NHL games are not collected using computer vision (as is the case with Major League Baseball). Instead, small groups of 3-5 men and women are employed to manually record events of interest that occur during the game using a custom software interface. The recorder clicks on the location of the event on a display of the rink and then selects the player and the event. The result of this process is that for each shot we are provided with the season identifier (first year of the season), the official game number, whether the game was during the regular season or postseason, the date of the game, the home team, the away team, the period and the number of seconds into the period that the shot was taken, what advantage type the shooter's team had (even-handed, short-handed, power-play), the shooter's name, team, and position, what type of shot was taken (wrist-shot, backhand, tip-in, etc.), the zone in which the shot was taken (offensive, neutral, defensive), the (x,y) coordinate of the shot in feet relative to center ice, and a precomputed length of the shot.

Due to the manner in which the data are recorded, it is not as precise as in some other professional sports. For example, while Major League Baseball tracks pitches to within one inch using Sportvision's PITCHf/x, the official NHL records concerning shots are to the nearest foot. This includes both the x,y coordinate data of the shot as well as the length of the shot. There is no data about the path of the puck including where (if anywhere) on the net the shot ended or how fast the puck was travelling. Additionally, 1,292 of the 81,158 data points provided to us do not have x,y coordinate data associated with them. Each of these data points is the result of one of two cases:

*Null*– the NHL did not provide coordinate information for the events (0.20% of the total event data falls into this case)

*Multiple*– the NHL provided information on these events, but multiple events occurred at the same time so it cannot be determined which of the events belongs to each coordinate pair (1.66% of the total event data falls into this case).

### 3.3 The Sports Analysts' Workflow

The term "sports analysts" describes those economists, statisticians, computer scientists, and mathematicians who perform analysis on data derived from sports. These individuals use a variety of tools in the course of their work, including scripting languages, relational

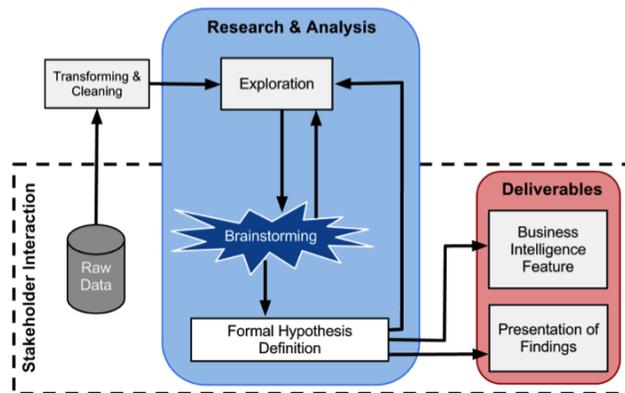


Fig. 10. Prototypical work flow of a sports analyst. The tasks on which we have focused the design efforts of SnapShot are: Exploration, Brainstorming, and Presentation of Findings.

databases, and statistical programs such as Excel, Stata, R, SPSS, and SAS. They draw upon statistics, econometrics, operations research, information systems, and applied mathematics during the course of their work. Analysts can be amateur fans, professional consultants, employees of a team, or employees a league. The analyst we worked with is employed by a small consulting firm. This allowed us to adopt an iterative, user-centered design process.

Our team member describes working as a hockey analyst in two modes: reactive and proactive. Reactive consultation occurs when a client, often a team, poses a question for which they desire an answer. Proactive consulting occurs when an analyst initiates and implements a new *business intelligence feature* to present to a client. Example business intelligence features include:

**Player Productivity Models** to inform lineup changes.

**Player Contribution to Winning Models** to inform contract renewals and player transactions.

**Team Performance Models** to track performance in key areas over various time intervals.

**Video Analysis Preprocessing** to prepare answers for questions that may arise during video analysis.

**Dataset Preparation** for stakeholders that request subsets of data to use for their own "in-house" analysis.

Our team member helped us generate a model of his typical workflow (Figure 10) to demonstrate and highlight the order in which these features may occur. Throughout the workflow he interacts with individuals he describes as *stakeholders*– those who are counting on analytics to inform their decisions. Stakeholders can be league officials or general managers, personnel officers, or coaches of teams. Usually these stakeholders will provide the raw data to our analyst, though the NHL or an authorized third party may also distribute data. Sometimes our analyst must extract raw data from multiple sources, such as official records and proprietary data.

Once he has the raw data, he must transform it into a form readable by his tools. He will also calculate derived attributes of the data at this stage. Once he has accomplished all of this, he is ready to begin his exploration (top of the blue box in Figure 10).

During initial exploration, he utilizes tools such as Excel, R, and SAS. He typically compiles tables of data subsets and creates static figures such as bar charts and scatterplot matrices. He uses these to quickly recognize correlated statistics or overwhelming outliers. Exploration will often include running simple macros and becoming familiarized with the data. He indicated that the five most crucial variables of the available data to him were the length of the shot, whether the shot was a goal or not, whether the shooter was on the home team

or the away team, the shooter’s team, and the location the shot was taken from on the ice.

Our analyst uses the foundational work created through such exploration as a starting point for brainstorming. He most often brainstorms independently or with his fellow consulting partners. This brainstorming process can include whiteboarding or video, telephone, and/or email communication amongst participants. Occasionally he brainstorms with outside analysts such as fellow attendees of the annual Sloan Sports Analytics Conference. The result of a brainstorming session is a set of hypotheses of varying levels of abstraction.

Once numerous hypotheses are generated, some are taken back to the exploratory stage. Our analyst creates summary tables or custom models using a quantitative tool. Despite the fact that this cycle (exploration, brainstorming, hypothesis generation) can iterate many times, our analyst can only pursue a finite number of hypotheses. Unconventional and complex hypotheses can drain precious time, money, and manpower because they are not easily proven or refuted. Further, these theories tend to return few beneficial findings to the stakeholder. Thus, he must ignore many potentially radical theories. Findings are those hypotheses that survive both brainstorming and statistical confirmation to become theories or models.

Finally, our analyst presents these findings to the project’s stakeholders. He accomplishes this through either a periodic report of a business intelligence feature or a PowerPoint presentation of recommendations.

We believe that visualization can enhance this prototypical sports analytics workflow as several of its components have been addressed by the information visualization research community. Tukey [26] and Inselberg [10] describe exploring data sets using visualization. Viegas and Wattenberg [27] and Heer et al. [8] outline how sharing of visualizations can encourage discussion. Segel and Heer [17] discuss using visualization to present findings.

## 4 SNAPSHOT SYSTEM

### 4.1 Design Rationale

Reflecting upon our discussions with the hockey analyst and our own experiences and research about the sport, we developed a set of design objectives for a visualization system. Below, we list important nuances of the problem space (in bold) along with how we address each in our system design.

#### **Location and distance from the net is crucial to shot data.**

We chose to use the ice hockey rink as a backdrop to our visualizations, allowing for an analyst to translate real-world context to the data in a system.

#### **Key facets of hockey data are categorical.**

We allow for mapping categorical variables to colors and dividing by categories into separate views.

#### **Open-ended exploration should allow for selection of both each individual shot or a defined subset of shots.**

We designed for analysis of subsets of data through universal transparent filtering. We allowed for the full range of data granularity: from one single shot to all 72,926 shots at once.

#### **Every analyst’s investigative lens is unique.**

We created an environment with multiple views to provide the flexibility that the variety in analytical approaches requires.

#### **The ability to record and revisit promising ideas encourages alternative exploration.**

With a single click, the system saves the current exploration state to a new browser tab or window. We termed this process *SnapShooting*.

#### **Collaboration strengthens hypotheses.**

Every view has a unique, sharable URL for enabling discussions amongst analysts to include visual evidence.

Table 1. Summary of data selection process.

81,158	Data points in 2010-2011 season
↔	5,551 Post-season data points
↔	75,607 Regular season data points
	↔ 2,192 Overtime data points
	↔ 73,415 Regulation data points
	↔ 463 No coordinates
	↔ 107 null condition
	↔ 356 multiple condition
	↔ 72,952 With coordinates
	↔ 26 Penalty shots
	↔ <b>72,926 Non-penalty shots</b>

### Comparisons drive insights.

We implemented small multiples views, *SnapShooting*, and concurrent, multi-technique displays to compare different partitions, selections, and representations of data.

### Ultimately, stakeholders expect a presentation of findings.

We enabled the system to export high-resolution images that can easily be incorporated into presentation materials.

These design choices led to the creation of an interactive visualization system, which we named SnapShot.

## 4.2 Data Pre-processing

SnapShot contains shots that occurred during regulation time during the regular season. Shot location is necessary for capitalizing on the contextual display of data and thus the system does not include data points lacking coordinate data due to either the null or multiple conditions. Discarding shots recorded more than once was a basic integrity issue. Removing null shots was a decision made at our discretion based on feedback from the analysts. When they run their statistical analyses they ignore these shots. Building a system attempting to conform to their current behavioral patterns and treatment of data, we chose to follow suit. This left our focus data set containing 72,926 regular-season, regulation-time shots. Table 1 summarizes this process.

NHL rules dictate that teams alternate which of the two ends of the rink each team attacks. While this both lowers any bias from any one rink end and allows fans at the arena to see both teams from two angles, it introduces unpredictability in the records concerning which end a given shot was taken against. SnapShot thus pre-processes the data to normalize all shots as if they were taken against the right end of the rink.

## 4.3 User Interface and Visualizations

SnapShot supports visualization through a consistent user interface. The canvas is in the upper lefthand corner of the window (Figure 12 (1)). For all non-small multiple visualizations this canvas size is fixed and features a hockey rink outline. Data can be filtered in the lower half of the system (Figure 11) through a series of interactions.



Fig. 11. Filtering widgets. The selected filters limit the data to only even-strength, away goals taken at the Minnesota rink, by left and right wingers.

SnapShot provides a rich set of filter controls: checkboxes and toggle-buttons for turning on and off shots, goals, shots by each position, and relative strength on the ice; drop-down selectors for teams

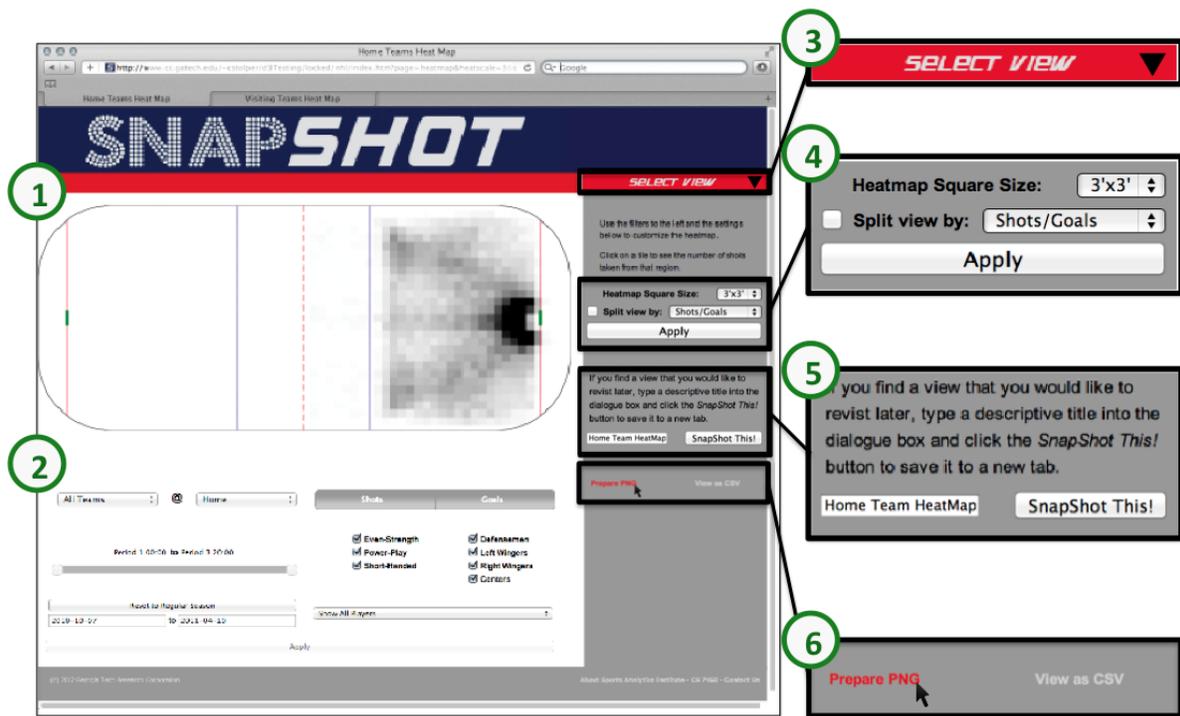


Fig. 12. Full view of the SnapShot system displaying a traditional heat map of shots and goals by home teams. The user interface includes: (1) Canvas where each visualization is drawn; (2) Data filtering widgets; (3) View selector drop-down menu; (4) View-specific settings; (5) SnapShot widget; (6) View and data export links.

and rinks (which also handle home and away) as well as players; a double-ended slider for time during the game; and a calendar for dates during the season. The filter controls are consistent across all views to highlight that the data behind different visualizations remains constant and provides the added benefit of minimizing context-switch dissonance between views.

The right side-bar of the system includes several important features. Navigation between visualization techniques is controlled by the upper right drop-down menu (Figure 12 (3)). Each active visualization displayed on the canvas has a corresponding group of settings (Figure 12 (4)). The text and filters in this area will change for each of the three view techniques in order to remain relevant to the current visualization. SnapShooting is conducted through the text box and button located on the right side. Below SnapShooting are the links for image (PNG) and data (CSV) exporting.

Earlier, we identified the five most important variables in the data to the analysts: shot length, shot/goal, shot location, shooter team, and home/away. SnapShot provides three visualization techniques (Shot Map, Traditional Heat Map, and Radial Heat Map) that highlight these facets, especially shot location and shot length. Each view is drawn using Data-Driven Documents (D3) javascript library [2] on an abstracted hockey rink including the border of the rink, the red, blue, and center lines, and the goals.

Here we describe each of the three techniques in detail.

#### 4.3.1 Shot Map

The shot map (Figure 13) primarily highlights shot location and secondarily the shot/goal, home/away, and shooter team distinctions. SnapShot represents every shot as a small SVG circle element with an opacity property of 0.1 drawn at the location where the shot originated. SnapShot colors each circle according to a categorical variable selected by the user from the available variable schemes of shot/goal, home/away, shooter's team, rink, period, shooter position, and advantage. Thus, the color of any given pixel is the alpha-blend of the colors of the last ten shots rendered at that position. SnapShot renders shots in the following order: shots then goals, each in reverse chronological order. Because the number of short-handed shots  $\ll$  number of power-play  $\ll$  even-strength shots, the rendering order when under

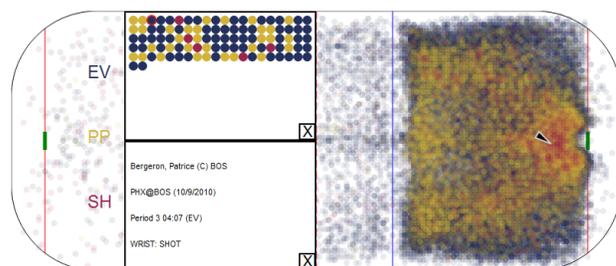


Fig. 13. Shot map of all 2010-2011 regular-season, regulation-time shots using the *advantage type* color scheme. The black arrow tooltip has selected a grid of the shot map and the open overlay windows are displaying every shot at that location (top) and the details of one particular shot selected at the point (bottom).

the advantage-type color scheme is: even-strength shots, even-strength goals, power-play shots, power-play goals, short-handed shots, short-handed goals.

The NHL records shot coordinates to the nearest foot. So for any given coordinate there are often multiple shots assigned to that location, even with filtered data. To still allow the user to see every shot, SnapShot allows the user to click on any coordinate and a black arrow tooltip will appear and pop up a details window displaying all shots taken from that coordinate as linearly-arrayed circles (Figure 13). The details pop-up displays circles arranged according to the standard rendering scheme (all shots followed by all goals), regardless of color scheme (in other words, for any color scheme with the same filters, a shot's circle will be in the same location in the tooltip for its coordinate). The user may click any of these circles to see the respective shot's full details. SnapShot also displays a legend of the currently selected color scheme on top of the defensive (left) end of the rink.

#### 4.3.2 Traditional Heat Map

The traditional heat map view (Figure 14) is designed primarily to highlight patterns of shot location density. SnapShot's heat map is a mono-chromatic grid, with each cell representing a fixed region on the ice. SnapShot draws each region as a black SVG rect element

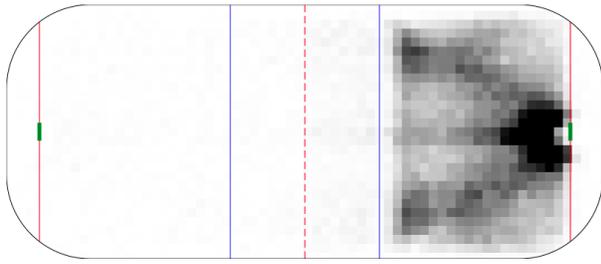


Fig. 14. In this traditional heat map, SnapShot displays all 2010-2011 regular-season, regulation-time shots using 3 foot by 3 foot squares.

with a set opacity. The opacity of the element is a function of the number of shots in the region, the natural log of the total number of shots corresponding to the selected filters, and the size of the region ( $opacity_{square} \leftarrow \frac{shots_{square}}{\ln shots_{filter} \cdot sidelength^3}$ ). This formula allows the analyst to identify patterns at scales from a single team at a single rink to every team at every rink. An analyst has the ability to adjust the size of cumulation regions to see the data at higher or lower granularities. At its finest granularity the traditional heat map can display grids of one-square foot of shots. Clicking on a cell opens a tooltip displaying the number of shots taken from within a respective region.

#### 4.3.3 Radial Heat Map

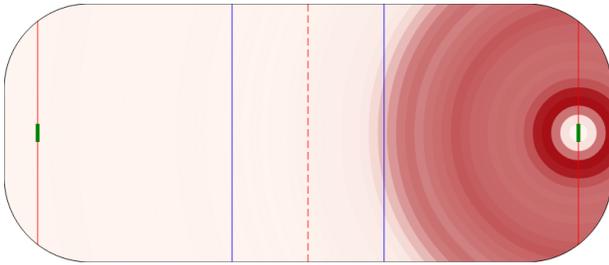


Fig. 15. Radial heat map of all 2010-2011 regular-season, regulation-time shots using 3-foot-wide rings.

SnapShot’s radial heat maps (Figure 15) convey information about shot length, the statistic identified as most interesting by the analysts we worked with. SnapShot visualizes the distance of shots from the net using a series of concentric rings surrounding the attacking (right-hand) goal representing cumulative bins of shots. Each ring represents a defined segment of the ice. SnapShot colors each ring based on the number of shots taken from its respective segment, normalized by the densest segment. The darkest red ring  $rgb(165, 15, 21)$  represents the densest segment. Using D3’s scale.linear interpolation function, SnapShot assigns colors to the remaining rings according to the normalized number of shots ( $[0..1] \rightarrow [rgb(255, 245, 240)..rgb(165, 15, 21)]$ ). Finally, SnapShot recolors empty segments’ rings white. The user can adjust the view’s granularity by modifying the width of the segments. Clicking on a ring displays a tooltip containing the number of shots taken from within its respective segment.

### 4.4 Comparison Capabilities

#### 4.4.1 Small Multiples

In order to highlight the shot/goal, home/away, and shooter team distinctions, SnapShot allows an analyst to view small multiples [24] of a given visualization (Figure 1), partitioned based on one of these variables. SnapShot supports displaying small multiples of each of the three visualization styles, partitioning the data based on not just shots/goals, home/away, and shooter’s team, but also rink, period, shooter position, and advantage type.

#### 4.4.2 Multi-Technique Display

The initial display of SnapShot consists of all three views tiled. The user may adjust the filters on the data and the results will be shown in all three views. This facilitates the analyst identifying patterns that appear throughout multiple visualization techniques and allows the user to compare between visualizations of the same partition, rather than between partitions of data.

#### 4.4.3 SnapShooting

As shown in Figure 12 (5), SnapShot supports comparing arbitrary subsets of data and arbitrary view techniques through *SnapShooting*, duplicating the current view to a new tab with an optional user-defined title. The user can toggle between these tabs as he or she wishes, and because all three views occupy the same space on a given tab, this toggling allows the user to compare two visualizations as if on top of each other, much like a flip chart. The analyst can thus visually identify the often subtle differences between two views.

### 4.5 Supporting Sharing

#### 4.5.1 Web Deployability and Bookmarking

In order to support teams of analysts, we have hosted SnapShot on a private web server. This allows team members to both access their views from any modern browser, as well as supporting the ability to share the visualization within the context of the tool used to build it. SnapShot’s views have uniquely identifiable URLs, allowing an analyst to bookmark a view in the browser or pass the view on to a colleague.

#### 4.5.2 Exporting

While bookmarking supports collaboration, it is less effective for supporting exploration or presentation. Analysts exploring the data using a visual tool will still want to confirm their findings in a statistical package. To facilitate this capability, SnapShot allows the user to export the currently displayed data which then can be imported into the analyst’s tool of choice (Figure 12 (6)). Analysts creating visualizations to enhance their communications with stakeholders rarely need the data itself. Instead, they prefer to have static images which can be placed in a slide deck, in a whitepaper, or on a webpage. SnapShot supports this use by allowing the user to export his or her current view as an image, including any tooltips currently displayed.

### 4.6 Implementation

SnapShot is written entirely in Javascript, making use of the jQuery toolkit<sup>1</sup>. The interactive visualizations are made up of SVG elements generated using the D3<sup>2</sup>, which also handles the data importing. The toggle buttons, calendar widget, and slider widget are all from the jQueryUI widget library<sup>3</sup> and the remaining widgets are native html. The exporting feature makes use of the canvg package<sup>4</sup>, which allows for converting SVG elements to PNG images. The GET-based bookmarking code is greatly simplified by Stephen Morley’s QueryData.js scripts<sup>5</sup>. The radial heat map colorscale and a selection of the categorical color schemes were chosen using ColorBrewer2 [3].

## 5 CASE STUDIES

To evaluate SnapShot, we worked with three professional hockey analysts. Two of the evaluators were consultant analysts and one was an internal analyst for an NHL team. We conducted one-hour sessions with each analyst remotely using screen-sharing software. For each session, we recorded both the screen activity and the participant’s comments. A SnapShot designer was present for each of the three study

<sup>1</sup><http://www.jquery.com>

<sup>2</sup><http://mbostock.github.com/d3/>

<sup>3</sup><http://www.jqueryui.com>

<sup>4</sup><http://code.google.com/p/canvg/>

<sup>5</sup><http://code.stephenmorley.org/javascript/parsing-query-strings-for-get-data/>

sessions but only as technical support. No design guidance or collaboration was present between the SnapShot creators and the case study analysts.

### 5.1 Analyst A: “Is This Rink Different Than Others?”

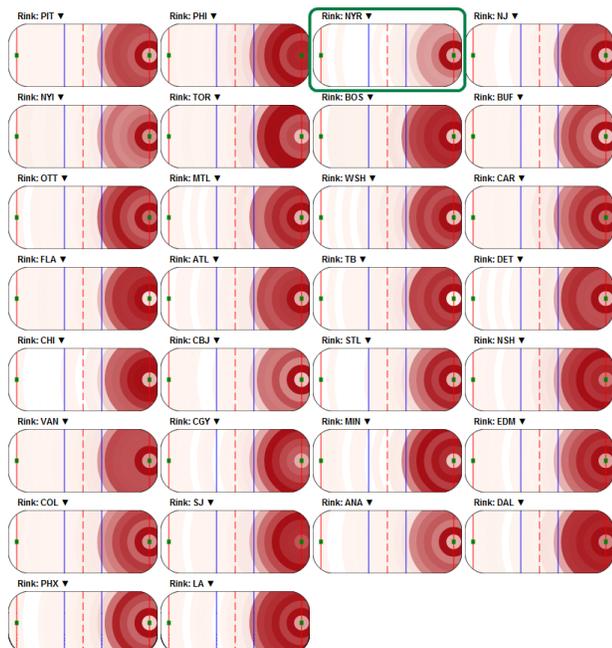


Fig. 16. Small multiples of radial heat maps for each of the thirty NHL teams at home created by the Analyst A. The analyst’s initial team of interest, the New York Rangers, is indicated by the green box.

Our first analyst (A) had been investigating shot lengths at a specific hockey arena, Madison Square Garden in New York. He had been doing so by studying raw data and game footage. He began his visual exploration with SnapShot by creating small multiples of each rink in the league (Figure 1) then increased the band-width to 10-feet (Figure 16). He remarked, “I will be able to see right away if this rink really is different than others.” As evidenced in the figure, he found that NYR (Madison Square Garden) does show a different pattern of shots than any of the other rinks.

He next filtered down to just shots by home teams, and he noted that patterns appeared in several other rinks that surprised him. Rinks such as Minnesota, Nashville, and Philadelphia were all recording high shot counts uniformly across the entire offensive side of the ice (from the right blue line to the net). This did not match his working knowledge of those teams, and he commented that Minnesota should look different from Nashville and Philadelphia due to their relatively weaker offense.

While the Rangers were his starting point, he put them on hold to explore the other three rinks. He SnapShotted six traditional heat maps: home shots only and away shots only for each of the three rinks. He told us that he was only concerned with even strength shots and filtered out power play shots and short handed shots. After studying the relative differences in shot patterns for the three teams, he changed the views to shot maps so that he could color shots by player position. He was interested in identifying which positions might be contributing to the atypical shot patterns.

Upon completion of the session, he mentioned that he was eager to investigate the three teams that SnapShot highlighted as similar (but should have been more dissimilar) and that he would use the image from Figure 16 to share his findings with interested parties.

### 5.2 Analyst B: “Animals Defend Their Homes”

Analyst B began the session by declaring, “Animals defend their homes”. He was interested in observing if hockey teams follow this ‘rule of nature’: do they play more defensively at home and offensively on the road? He hypothesized that “playing defensively” would

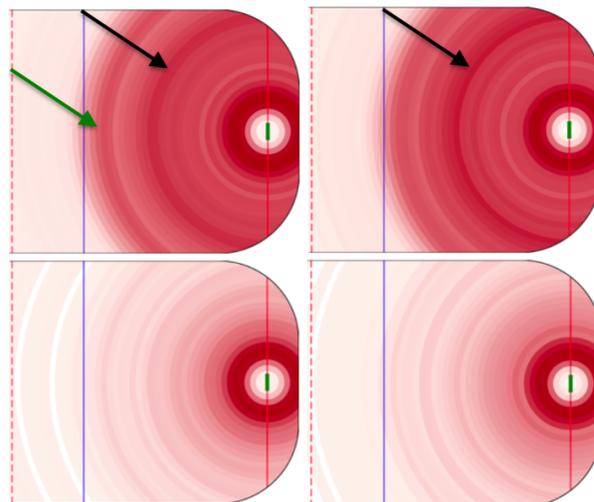


Fig. 17. Radial heat maps created by Analyst B. Top row showing shots and goals, bottom row showing only goals. Left heat maps are for home teams and right are for away teams. Black arrows indicate a shot distance density present in both the home (left) and away (right) radial heat map. The green arrow points to a region of longer distance shots that was of interest to the analyst.

imply longer even-strength shooting, because “the home team is not willing to give up any ground by over-attacking on offense”.

As his hypothesis was a high-level generalization, he chose to keep each filter at the highest level, and never drilled down to any specific locations, he chose to explore exclusively using radial heat maps. He SnapShotted two radial heat maps (Figure 17, top row): home shots and goals and away shots and goals, each at a two foot granularity. Because the views included all thirty teams, he mentioned that there may be several teams skewing the distribution of shot lengths, but that he still expected to find slight differences in shooting patterns.

By toggling back and forth between the two radial heat maps, the analyst noticed and pointed out two rings on the ice that interested him. While both heat maps had a red ring around the thirty foot shooting mark, the ring is comparatively darker on the visitor team map (each ring indicated with a arrow arrow in the top row of Figure 17). Additionally, the dark ring fifty feet from the net on the home map (green arrow in Figure 17) was not as pronounced on the away map. These two insights caused the analyst to theorize out loud, “More long shots from home teams and less of those thirty footers, I wonder which of these are actually goals?” This query led him to SnapShot two additional radial heat maps (Figure 17, bottom row): one of home goals and one of away goals. The distributions appeared to be nearly identical. This excited the analyst. He concluded that “They are shooting differently at home but they aren’t scoring differently!” As he ended his evaluation, the analyst asked if he could access the system in a non-evaluation capacity and use these visualizations in a client presentation about the “Animals defend their home” theory.

### 5.3 Analyst C: “Sweet Spots”

The third evaluator, Analyst C, chose to use SnapShot to investigate potential “sweet spot” areas on the ice, those areas with a high conversion rate of shots to goals. He commented, “Since most goalies have their stick in their right hand and glove in their left, I am wondering if they struggle to move left to right, and if shooters are capitalizing on this.” Essentially, he was hypothesizing that more goals result from shots on the goalie’s right side than left side.

To explore this possible phenomenon, the analyst created a shot map colored by shots and goals. He quickly turned off the blue (failed) shots and displayed only the red goals (Figure 18, top left). His initial impression was that there was little evidence that goals came from non-symmetrical locations on the ice so he created several shot maps based on his domain knowledge of hockey dynamics. One displayed

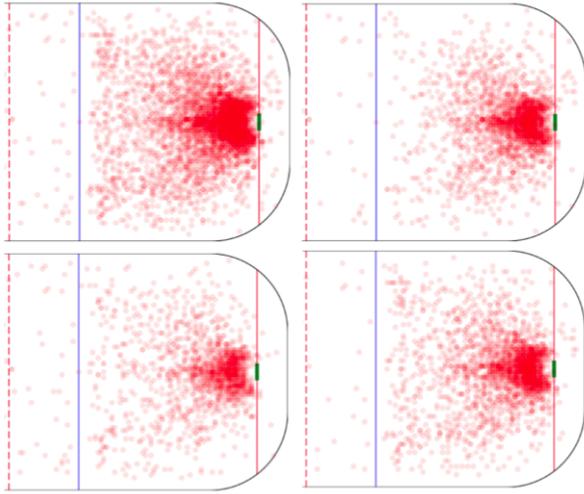


Fig. 18. Analyst C’s four shot maps of goals during the season. The maps depict, clockwise from top left: all goals, left and right winger only, goals made from January onwards, and goals during the 3rd period.

goals by just left and right wingers (Figure 18, top right); one displayed goals during the third period (Figure 18, bottom left); and one displayed only goals scored from January onward (Figure 18, bottom right). As he explored each image, he talked about theories common amongst NHL coaches such as “intuition says that a shot has a bigger advantage if the goalie isn’t set yet”.

None of the four shot maps that he created (and the corresponding traditional heat maps he created later) supported his hypothesis that players scored more goals shooting from the goalie’s right side. However, this dead end only intrigued the analyst more. In his words, “I like to feel that I’m being pushed to think more creatively. This was my first opportunity to explore these theories.” He mentioned that he would like to continue to search for “soft spots” using the system in the future, particularly areas outside of “the house”, the term he used to describe the high-trafficked pentagon defined by the net, faceoff circles, and the blue line.

## 6 DISCUSSION

In order to offer hockey analysts effective visualization capabilities, we designed the SnapShot system around nuances of their data and three stages of their existing workflow: exploration, collaboration, and presentation.

The case studies demonstrate SnapShot effectively allowing for visual exploration of data. As Analyst B explained, “I like that this allows me to go from abstract thoughts to specific images very quickly. Sometimes I don’t know what it is telling me initially, but I like the possibilities.”

During the evaluation of SnapShot, two participating analysts preferred radial heat maps because the view visualizes the variable they were most interested in, shot length. Analyst B told us in his closing comments, “I was suspicious at first, but I’ve fallen for it.” The Analyst C preferred shot maps because that view visualizes the variable he was most interested in, shot locations.

The analysts utilized small multiples representations, despite initial reservations. Upon his first encounter of them, Analyst A remarked, “I have to tell you, I am overwhelmed and very impressed all at once.” Yet because many of hockey’s statistics are categorical, the view proved its utility over the course of his investigation. By the end of his session, Analyst A was constantly referring back to it.

Analyst C shared promising evidence that SnapShot may provide the presentation capabilities for which it was designed, “I know many traditionalists in our sport who resist analytics, but maybe wouldn’t with more pictures involved.”

We were surprised that none of the evaluation participants expressed interest in sharing their visual findings with their peers, despite collaboration being a main design goal of SnapShot. However,

this does align with the initial exploration stage of the model workflow being an individual process, in contrast to the collaborative brainstorming exploration stage.

## 7 CONCLUSION AND FUTURE WORK

In our introduction, we explained that in hockey, winning on the ice starts with strong strategies and that hockey analysts contribute to these strategies through statistical analysis. In our related work, we described the current state of sports visualization. Through interactions with a professional hockey analyst, we developed and modelled a typical sports analyst workflow and we identified those areas where visualization can add value. We presented a hockey visualization system, SnapShot, designed to augment this workflow. Finally, we outlined and discussed three case studies of domain experts using SnapShot and the effectiveness of visualization for enhancing analysts’ exploration process and business intelligence features.

To the information visualization community, this paper contributes: a model workflow of a sports analyst, the SnapShot system for visualizing 2010-2011 National Hockey League shot data, and three case studies of hockey analysts exploring their data visually.

We propose several areas of future work for ourselves and the visualization community as a whole. We hope to expand the number of event types that SnapShot supports to include events such as hits, passes, and penalties. SnapShot only visualizes the data provided; we would like to add analytic capacity to the system in order to support visual analytics methods. Our second analyst expressed a desire to filter based on whether a shooter’s team is winning or losing, so we plan to accommodate this and other such analytic functions in later versions of SnapShot. We hope our documentation of a sports analysts’ workflow will inform future sports visualization system design. Our workflow model provides a foundation for designing systems to support sports analysts. There is opportunity to refine this model through investigation of the exploration stage collaboration amongst sports analysts. Finally, we intend for our case studies to highlight the potential for incorporating visualization into the growing ice hockey analytics community.

## ACKNOWLEDGMENTS

The authors wish to thank the study participants for giving their time to evaluate the system and the Information Interfaces Group, Yi Han, and the reviewers for their invaluable feedback on drafts. This work was supported in part by the National Science Foundation via Award CCF-0808863.

## REFERENCES

- [1] P.-A. Albinsson and D. Andersson. Extending the attribute explorer to support professional team-sport analysis. *Information Visualization*, 7(2):163–169, 2008.
- [2] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.
- [3] C. Brewer and M. Harrower. Colorbrewer 2.0. <http://www.colorbrewer2.org>.
- [4] S. Carter, G. Roberts, and J. Ward. How mariano rivera dominates hitters. *New York Times Magazine* <http://www.nytimes.com/interactive/2010/06/29/magazine/rivera-pitches.html>, June 2010.
- [5] A. Cox and J. Stasko. Sportsvis: Discovering meaning in sports statistics through information visualization. *Compendium of Symposium on Information Visualization*, pages 114–115, 2006.
- [6] ESPN. GameCast. <http://scores.espn.go.com/nhl/gamecast?gameId=400047896>, 2012.
- [7] K. Goldsberry. CourtVision: New Visual and Spatial Analytics for the NBA. *MIT Sloan Sports Analytics Conference*, 2012.
- [8] J. Heer, F. Viegas, and M. Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1029–1038, 2007.
- [9] J. Hollinger. Carlisle pushes all of the right buttons. *ESPN NBA*, June 2011.

- [10] A. Inselberg. Multidimensional detective. In *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pages 100–107, 1997.
- [11] L. Jin and D. C. Banks. Visualizing a tennis match. In *INFOVIS '96: Proceedings of the 1996 IEEE Symposium on Information Visualization (INFOVIS '96)*. IEEE Computer Society, Oct. 1996.
- [12] M. Lewis. *Moneyball: the art of winning an unfair game*. W.W. Norton, 2003.
- [13] R. Maheswaran, Y.-H. Chang, A. Henehan, and S. Danesis. Deconstructing the Rebound with Optical Tracking Data. *MIT Sloan Sports Analytics Conference*, pages 1–7, Feb. 2012.
- [14] NHL. Ice Tracker. <http://www.nhl.com/ice/icetracker.htm?id=20110206151>.
- [15] A. Rusu, D. Stoica, and E. Burns. Analyzing Soccer Goalkeeper Performance using a Metaphor-Based Visualization. In *15th International Conference Information Visualisation (IV)*, pages 194–199. IEEE, 2011.
- [16] A. Rusu, D. Stoica, E. Burns, B. Hample, K. McGarry, and R. Russell. Dynamic visualizations for soccer statistical analysis. *International Conference on Information Visualisation*, pages 207–212, 2010.
- [17] E. Segel and J. Heer. Narrative Visualization: Telling Stories with Data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010.
- [18] R. Spence and L. Tweedie. The attribute explorer: information synthesis via exploration. *Interacting with Computers*, 11:137–146, 1998.
- [19] SportsVision. PITCHfx. <http://www.sportvision.com/base-pitchfx.html>.
- [20] Stats LLC. SportVu Data Visualization Suite. <http://www.sportvu.com/>.
- [21] Tableau Software. Tableau. <http://www.tableausoftware.com>.
- [22] D. Tan, G. Smith, B. Lee, and G. Robertson. AdaptiviTree: Adaptive tree visualization for tournament-style brackets. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1113–1120, 2007.
- [23] Tibco. Spotfire. <http://www.spotfire.tibco.com>.
- [24] E. Tufte. *Envisioning information*. Graphics Press, 1990.
- [25] E. Tufte. *Beautiful evidence*. Graphics Press, 2006.
- [26] J. Tukey. *Exploratory Data Analysis*. Addison-Wesley Series in Behavioral Science. Addison-Wesley Pub. Co., 1977.
- [27] F. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1121–1128, 2007.
- [28] T. Williams and J. Underwood. Science of Batting. *Sports Illustrated*, pages 34–47, July 1968.