Eurographics Conference on Visualization (EuroVis) 2018
J. Heer, H. Leitte, and T. Ropinski
(Guest Editors)

Volume 37 (2018), Number 3

Baseball Timeline: Summarizing Baseball Plays Into a Static Visualization

Jorge P. Ono, Carlos Dietrich and Claudio T. Silva

New York University

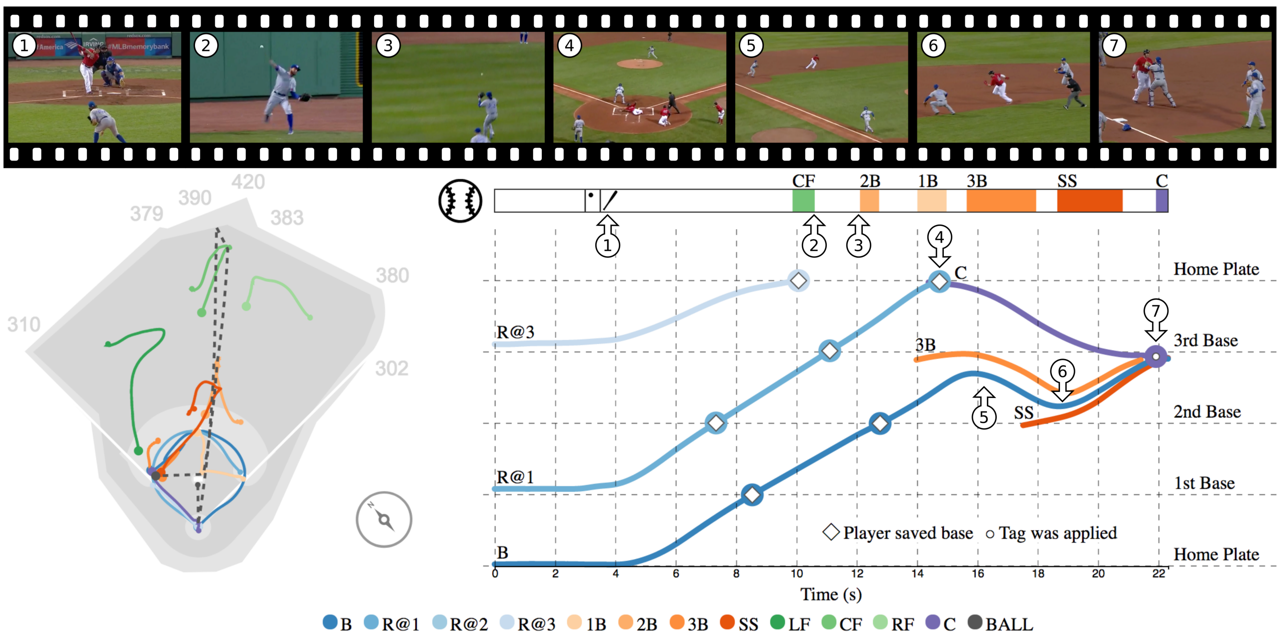


Figure 1: Visualization of a Toronto Blue Jays vs Boston Red Socks play. (1) Batter hits the ball. (2-3) Center fielder throws ball to second baseman. (4) Runner reaches home. (5) Rundown, batter stranded between two bases. (6) Batter changes direction. (7) Batter is tagged out.

Abstract

In sports, Play Diagrams are the standard way to represent and convey information. They are widely used by coaches, managers, journalists and fans in general. There are situations where diagrams may be hard to understand, for example, when several actions are packed in a certain region of the field or there are just too many actions to be transformed in a clear depiction of the play. The representation of how actions develop through time, in particular, may be hardly achieved on such diagrams. The time, and the relationship among the actions of the players through time, is critical on the depiction of complex plays. In this context, we present a study on how player actions may be clearly depicted on 2D diagrams. The study is focused on Baseball plays, a sport where diagrams are heavily used to summarize the actions of the players. We propose a new and simple approach to represent spatiotemporal information in the form of a timeline. We designed our visualization with a requirement driven approach, conducting interviews and fulfilling the needs of baseball experts and expert-fans. We validate our approach by presenting a detailed analysis of baseball plays and conducting interviews with four domain experts.

CCS Concepts

•***Human-centered computing*** → ***Information visualization;***

1. Introduction

The use of Play Diagrams is one of the easiest ways to explain rules, strategies and actions in sports. They are widely used in sports like basketball, baseball, football and soccer, as a clear way to represent and convey information. The diagram is basically a top-view mapping of the actions of the targets (the players and the ball) during a certain period of time. The language and the symbols used are usually well understood among managers, journalists and fans. Although efficient, these charts were created to convey just simple information. When several players move to a small region in the field, the visualization becomes cluttered and users cannot read the movements of the players clearly. The same happens when the diagram depicts a long play, where too many actions are packed into a single chart. In baseball, for example, all actions and movements usually happen around the bases, and many actions might happen in a single play. Figure 1 (left) shows an example of a diagram of a complex play, where several actions are cluttered around the third base.

The Major League Baseball (MLB) Statcast project [Med15], the new baseball tracking technology, frequently shares Play Diagrams with fans in their twitter account and in their At Bat application [Med17]. One example is shown in Figure 2 [†]. The diagram depicts the trajectory of the players and the ball during the play, but it cannot encode *when* each action took place. It is not possible to know if the runner reached third base, for example, or if he was tagged by the third baseman. What Figure 2 illustrates is that the time, and especially the relationship among the players through time, is critical for the understanding of complex situations. Since time is just implicitly represented on such diagrams, we have a limited canvas to express the information contained in the play.

The representation of complex plays usually makes use of additional techniques to overcome such limitations. Techniques like using color mappings for overlapping objects, decomposing the player actions, focusing (or zooming in) the complex parts of the play, are commonly seen on sport websites or other artistic depictions of plays. The use of animated diagrams, for example, is a technique that let the user follow the targets as they move through the play and offers a clear representation of time. An animated diagram, however, may result in adding complexity to the visualization. If the trajectories of the targets are kept in the diagram as the players move, the resulting diagram will still contain the characteristic overlapping of the static diagrams, and if they are not, they lack the elements that help the user to compare and analyze the dynamics in the play.

In this paper, we present Baseball Timeline, a new way to visualize baseball plays with a focus on the temporal aspect of the game. We used a requirement driven approach in order to design a visualization that is able to clearly and succinctly convey players actions and game events throughout time. We take advantage of the way the players move over the field in order to build a visual representation that encodes both position and time in a clear two-dimensional chart. Our encoding is inspired by Marey's Graph, a

[†] https://twitter.com/_dadler/status/893907355255156737

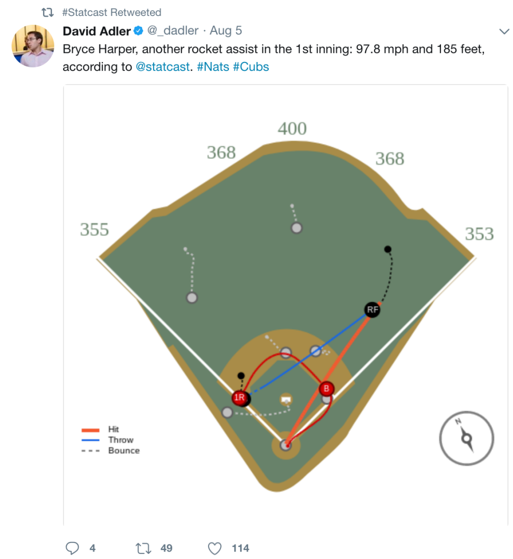


Figure 2: An example of the way the MLB Statcast [Med15] project makes use of Play Diagrams. Each play containing interesting features, like the speed of a throw, for example, is published as a diagram and a brief description of what set it apart.

successful visualization that was originally proposed for the presentation of train schedules. Our technique has the advantage of providing baseball analysts with an overview of the play while enabling them to identify interesting temporal events in the chart, for example, "The 1st baseman arrived at 1st base before the batter". Figure 1 (right) shows a baseball play using the timeline visualization, with the X axis representing time, and the Y axis, the position in respect to the bases. In our expert interviews, we show that the use of Baseball Timeline together with a Play Diagram improves the user understanding of the events in the game, even for complex and long plays.

Our work includes the following contributions:

- We present Baseball Timeline, a visualization driven approach to understanding the spatiotemporal evolution of baseball plays. We show how our design can be combined with the Play Diagram in order to create more meaningful play depictions.
- A study showing how Baseball Timeline can be used to analyze a set of six baseball plays. We highlight the insights that can be extracted from a timeline representation and how they improve the reader's understanding of the plays.
- An interview with four domain experts, who helped us validate and improve our design.

This paper is structured as follows: Section 2 reviews the most relevant works in the field of sports visualization and diagram creation. Sections 3 and 4 present an overview about baseball, and the requirements described by domain experts. Section 5 describes our original designs and failed attempts. Section 6 presents the Baseball Timeline and describes how to build the visualization in order to represent baseball plays. Section 7 presents use cases of our visualization applied to interesting baseball plays, and how it can be used to gain insight into the data. Section 8 presents the feedback we received from baseball experts on our visualization and how we improved our design based on their comments. Finally, Section 9 presents our conclusions, the limitations of our design and future work.

2. Related Work

This section is divided in two parts: first, we review studies related to sports diagrams. Then, we focus on spatiotemporal trajectory visualization techniques which inspired Baseball Timeline.

2.1. Sports Visualization

The use of diagrams is widespread in sports to represent strategy, formation or complex player actions. The diagrams can be roughly grouped in terms of the type of information that they portray: *event visualizations*, *single play trajectory*, *aggregated spatial data*, and *aggregated statistics data*.

Event visualizations present a high-level description of play or game highlights (or *events*), in order to contextualize and explain the main developments in the game. They are usually associated with a temporal representation, where events are sorted on time and provide a view of the action flow through the play. Event visualizations have been used in soccer [WG12] and basketball [CDBG14]. Wongsuphasawat et al. [WG12] used a Sankey diagram in order to portray the evolution of soccer events during a game season. The tool was created to assist users in the task of event sequence analysis. Cervone et al. [CDBG14] proposed a new metric to evaluate basketball players and their ball pass events. In the paper, the authors used small multiples of Play Diagrams in order to convey the temporal evolution of the game, and a time graph to show how the proposed metric changed over time. With the tool, coaches were able to identify the times at which it was advantageous for a player to pass the ball or attempt to score a point.

Single play trajectory visualization is a popular approach for the visualization of actions in sports. The movements of the targets (players and the ball) are depicted as Play Diagrams, i.e. top-view mappings of the field, where movements are represented as lines in the chart. We can find examples of these charts in basketball [TC10], baseball [DKVS14, LOC*16], soccer [PVF13, SJL*17] and tennis [POJC01]. Sport-specific information can be also encoded in the chart; Theron et al. added player speed information using colors and events, and information about team strategies, such as dribbling and defensive traps, by using icons. Dietrich et al. [DKVS14] added annotations to the tracking data, as a way to highlight baseball statistics like player top speed and ball velocity. Stein et al. [SJB*16] created an icon based representation for soccer plays, where movement and interaction among players are represented in an adapted Play Diagram. The authors also developed a system that shows the actual video footage as a diagram overlay, in order to give even more context to the visualization [SJL*17].

Aggregated spatial visualizations are used in sports in order to show trends in the position of players or ball over a large period of time. In baseball, for example, it can be used to identify trends in the ball landing position per pitch type [MB13,DKVS14,LOC*16]. In basketball, it can be used to show what are the regions of the court where players score more points [Gol12]. This data is traditionally visualized using heat maps or scatter plots, that are also contextualized with a top view of the field of play in the background. The chart may be customized, however, according to its intended purpose or audience. Pileggi et al. [PSBS12], for example, proposed a radial heat map to encode the proportion of shots taken by the distance to the goal in hockey games. Cross et al. [CS15] modeled players spatial batting ability and presented their models using heatmaps. According to the authors, this approach naturally highlights the areas in the strike zone where the batters had better performances. Losada et al. [LTB16] also used classic and radial heatmaps in order to analyze offense and defense in basketball games. The movements of the targets can be also aggregated in the visualization; Sacha et al. [SAMS*17] proposed a visualization that combines soccer player trajectories in order to summarize the overall movement during the play. The authors presented an approach that simplifies and cluster trajectories, thus reducing the visual cluttering and enabling the identification of soccer strategies.

Aggregated statistics visualizations are also a very widespread tool in sports, as they are able to present statistics about relevant sports metrics. Cox et al. [CS06] proposed Sportsvis, an interactive system that show statistics about baseball games. In Sportsvis, bar charts and treemaps are used to represent baseball statistics, such as the number of home runs and on-base percentage per game. The visualization of statistics is also studied in the context of soccer analytics. Albinsson et al. [AA08] proposed a system that used linked bar charts and scatter plots to enable the interactive querying and filtering of soccer statistical data.

All the aforementioned visualizations are effective at conveying the spatial information of the sport they represent. However, they are not focused on the *temporal* aspect of the data. In sports, the knowledge of how the players and the ball move during a play or drill is necessary to reveal the underlying strategy of the participants. The knowledge of *when* interactions among players took place, however, is critical for the understanding of complex situations. In baseball, for example, knowing that both the runner and the baseman arrived in a given base does not give enough information for inferring if the runner is safe or not. In the following sections, we present the related work on *spatiotemporal* diagrams, and how it can be extended in order to represent baseball plays.

2.2. Spatiotemporal Visualization

The visualization of spatiotemporal data is a challenging task. The data is frequently hard to analyze in full detail, and some type of simplification has to be performed in order to display it [AAB*13]. Many approaches have been used to represent both space and time in a graph, including (but not limited to) small multiples, space-time cube, clustering and time graphs. In this section, we briefly review each of these approaches.

Small multiples [Tuf01] is a popular way of displaying spatiotemporal information in charts [AAB*13]. In this approach, representative times (or *samples*) are chosen from the data. The samples are then shown as a set of juxtaposed charts, which allows an easy comparison between different samples. The downside of small multiples is that comparing situations where quantitative differences are not very prominent is a hard task [AAB*13]. Moreover, scalability with respect to time is also a concern [Mun14].

The space-time cube [Hag70] is a three-dimensional representation of spatiotemporal data that places the spatial data on two dimensions, and the time in the remaining dimension. This approach was explored to represent movement data as well [Kra03]. Although visually interesting, it might be not appropriate for identifying position of multiple elements over time, and often requires interaction, such as cube rotation, for users to make sense of the data [AAB*13]. In a recent survey of visual analytic techniques for movement, Andrienko et al. [AAB*13] suggest that the use of two displays, one mapping for spatial information, and another one for the time information, is more effective for the inspection of spatiotemporal data. Lukasczyk et al. [LMGH15] used this approach to visualize hotspot events over time: the spatial distribution was shown using kernel density estimates, while the temporal distribution was presented with an adaptation of Reeb graphs for time series.

Data clustering can be also used to simplify the spatiotemporal data before it is displayed. Andrienko et al. [AAS*12] developed an approach that groups similar spatial configurations together in order to guide the visual analysis. In their system, the trajectory data is presented using multiple views, i.e. map small multiples and time series, with clusters annotations encoded using color.

Marey's train schedule [Mar78] is perhaps the biggest inspiration for our work. The visualization portrayed the train schedule for Paris to Lyon in the end of the nineteenth century, by using a line chart with the horizontal axis representing time, and the vertical axis representing the train stations. The stations on the chart were separated proportionally to the actual distance between them, therefore it was possible to identify train velocity based on line angle and train crossings based on line crossings [Tuf01]. We adapt Marey's approach in order to create a spatiotemporal representation of baseball plays. The next section describes our approach.

3. Baseball Overview

Baseball is a bat-and-ball game that is played on a field shaped like a diamond. The field has a set of four bases placed at the corners of a ninety-foot square at the bottom of the diamond. The bases are labeled in counter-clockwise order starting at the bottom as home (or home plate), first, second, and third. The area right above the square is called *infield*, while the area above the dirt is called *outfield*. During the game, the teams alternate between the nine defensive and the four offensive roles. The defensive roles are the pitcher (P), the catcher (C), the basemen (1B, 2B and 3B), the short stop (SS) and the outfielders to the left (LF), center (CF) and right (RF). As the offensive roles, there are the batter (B) and zero to three runners on bases (R@1, R@2 and R@3). Figure 3 shows the field and the players average positions. The runners are not shown in the picture for conciseness, but their starting positions are next to the first, second and third bases.

Baseball is a very structured sport. It consists of nine innings, each of which are divided into two halves with teams taking turns on attack and defense. In general, a play starts when the pitcher

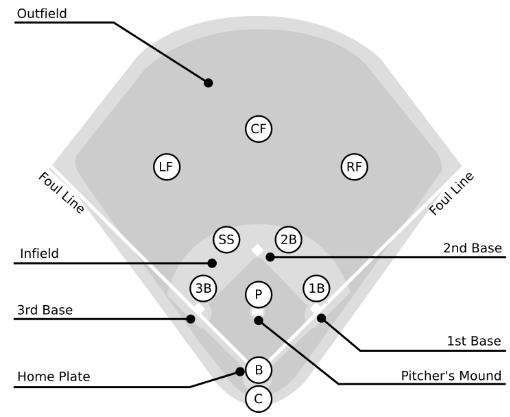


Figure 3: Baseball field of play and player positions: pitcher (P), batter (B), catcher (C), infielders (1B, 2B, 3B and SS) and outfielders (LF, CF and RF).

makes the first movement and finishes when the ball returns to the pitcher's glove or goes out of play. Every player has a fixed initial position, and the set of actions they perform is relatively limited. Players in the offensive role try to touch all four bases in anticlockwise order (1st, 2nd, 3rd and home plate). Meanwhile, players in the defensive role try to catch the ball and eliminate the attackers.

4. Data and Domain Requirements

In 2015, the MLB Advanced Media team unveiled the Statcast project [Med15], a system designed for the tracking of the players, the ball and a series of high-level game events with an unprecedented level of detail. The StatCast goal was to capture all actions performed on the playing field and process them to generate interesting content for the public. The project consists of both the hardware necessary for the tracking of targets as they move over the field (an optical and a doppler radar-based tracking system) and the software layer required to gather and process this information. The available data is extremely rich, enabling a deeper understanding of baseball games and making it possible for analysts to present interesting plays to fans in a high level of detail.

The baseball data is inherently visual, its availability was immediately followed by the research of interesting ways to visualize it. The visualization of the baseball data usually relies on mappings to convey the tracking information. Spray charts, or scatter plots that show ball landing position with respect to the ballpark, are frequently used to identify batting trends in players and teams. Meanwhile, Play Diagrams, or top-view mappings of the trajectory of the targets, are used to depict interesting situations or even entire plays. When analysts are concerned with the temporal information in the game, for example, identifying who was the first player to reach a base, they usually have to go back to the actual video footage of the play, or an animation of the tracking data. This approach to temporal analysis of games has three drawbacks: first, watching a play takes time, and the analyst might need to replay segments in order to identify the event of interest. Second, it relies on the user's memory to store the progression of the play, and they might need to rewind the video to review game events. Third, it requires the use of an electronic device in order to visualize the play, which is not ideal given that there is a lot of baseball information in static media being generated even today. Therefore, there is a clear need towards better visualization mechanisms to explore spatiotemporal information in baseball plays.

Based on these observations and on feedback we obtained from domain experts, we have compiled a list of requirements for a baseball play visualization that enables a detailed analysis of game events and player movements:

R1 Represent the entire play in a concise and clean manner: the visualization should present the information succinctly and without clutter. Moreover, it should not require much training to be understood.

R2 Identify players' actions, as well as relate those actions with the context in the play: the visualization should represent how players move in the baseball field throughout time, as well as how they interact with other players.

R3 Highlight relevant events in the play and enable the ordering of events by time: users should be able to easily see relevant events in the chart. For example, saving a base, catching the ball, or tagging a player must be easily identifiable actions.

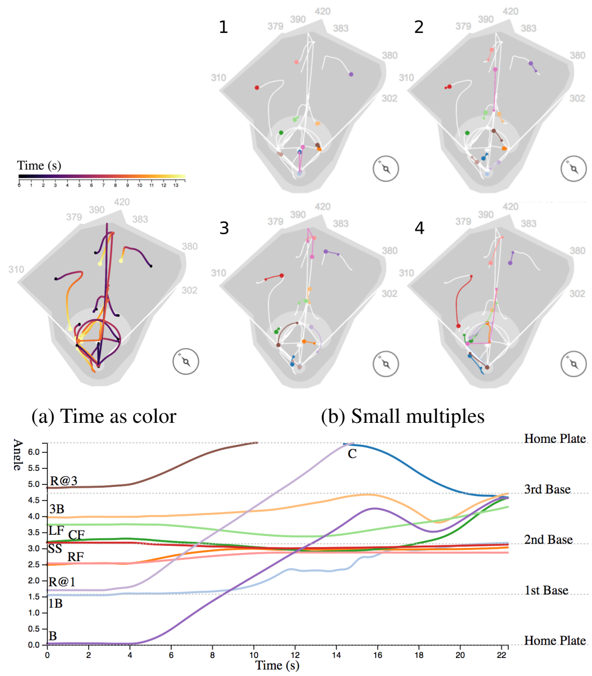
R4 Do not require interactive content: the visualization should not require interaction or animation in order to convey game information. This requirement makes the chart easily shareable both in social and static media (books, newspapers, magazines, and brochures).

5. Design considerations

The development of the Baseball Timeline was a iterative process, where we tested several well-known approaches for the spatiotemporal data visualization. This section presents some of these previous designs.

We started by focusing on small and incremental changes to the traditional Play Diagrams. The initial attempts were around time-based color mappings of the data and the automatic selection of interesting events for the creation of small multiples. Figure 4a shows one example of the time-as-color approach. The resulting chart is not effective at presenting the movement of players over time. As discussed by Munzner [Mun14], color is not an expressive visual channel to encode a continuous variable. Moreover, this visualization still had problems with cluttering, especially near the third base, and did not highlight game events. In summary, the time-color encoding does not meet the requirements R1, R2 and R3.

The second approach was focused on selecting interesting game events that would delimit the elements of a small multiple visualization. Figure 4b shows an example of this approach. The visual encoding had two problems: (1) as discussed by Andrienko et al. [AAB*13], it is hard to identify small movement changes in the chart, and (2) it is not possible to find the exact position of a player in time, given that player position is aggregated into each element. Small multiples solve the problem with clutter (R1), since it represents a small time frame in each chart. However, one could not read player position and game events from the chart, violating requirements R2 and R3.



(a) Time as color (b) Small multiples

(c) Bases as landmarks

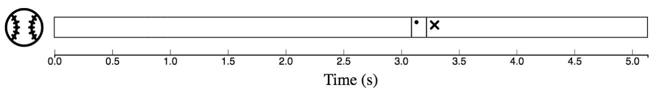
Figure 4: Design attempts that did not meet our requirements.

The initial attempts were not successful in conveying spatiotemporal information of baseball plays. We then moved to an approach where the play is taken as a time series, inspired by the work of Marey [Mar78]. Figure 4c shows the first result of the "Bases as landmarks" approach. Time is shown on the horizontal axis and spatial data (the position of the targets) on the vertical axis. Since most of the actions in baseball games occur close to the bases, we then took the bases as landmarks, and positioned players on the vertical axis according to how close they were to the bases. More specifically, player position was encoded as angle, with home plate being 0 radians and second base being π radians. Although this chart encodes the movements of the infielders, one could not read outfielder positions directly from it. As seen in Figure 4c, fielders LF, CF and RF are all cluttered on second base (around π radians), violating requirement R1. Events were not present on the chart either, violating R3.

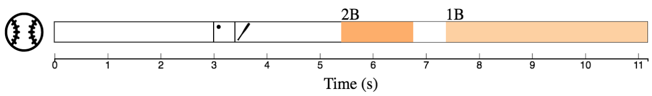
The "bases as landmarks" approach, although with limited success, is capable of representing the movements of the targets through time. The next session describes how we improved this approach in order to satisfy requirements R1-4.

6. Baseball Timeline

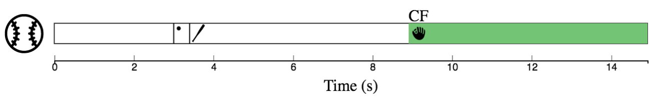
In order to fulfill the requirements given by the domain experts, we designed Baseball Timeline (BT), a visualization that enables the analysis of baseball plays by providing information regarding both temporal and spatial aspects of the tracking data. Figure 1 shows an instance of BT. The proposed visualization consists of three views, (1) the *Play Diagram*, (2) the *Ball Status*, and (3) the *Play Timeline*.



(a) 07/05/2015 - Los Angeles Angels @ Texas Rangers: Strike. The ball was pitched, but the batter did not hit it with the bat.

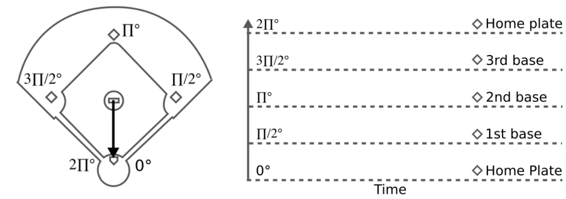


(b) 04/14/2017 - Chicago White Sox @ Minnesota Twins: Ball was batted, retrieved by the second baseman and thrown to the first baseman.



(c) 05/13/2016 - Oakland Athletics @ Tampa Bay Rays: Ball caught in the air by the center fielder.

Figure 5: Ball Status: this view shows the status of the ball throughout the play, from the moment it is pitched until the end of the play.



(a) Infield (b) Timeline view

Figure 6: Mapping of the player position to angle. (a) Infield with angle annotations and reference vector from Pitcher's Mound to Home Plate (b) Representation of player position in the Y axis.

Every player and the ball are associated with a color, which is the same on all three views. We have colored the batter and runners in shades of blue, basemen in shades of orange, fielders in green and catcher in purple.

The **Play Diagram** is well-known in the baseball community. It consists of a top-view mapping of the field, with the movements of the targets encoded as polylines. This view provides a spatial summary of the play to the user and was chosen in accordance to requirements R1 and R4. The starting position of every player and ball is encoded as a small circle, and the ending position of their trajectory is encoded with a big circle. The ball trajectory is shown with a dotted line in order to differentiate it from the players trajectory. Our dataset contains metadata about the game, including at which stadium the game took place. In case a custom image of the stadium is available, it is shown in the play diagram, together with a compass that shows the stadium orientation. Otherwise, a default image is displayed. Figure 1 shows an example of a play in the Fenway Park, a stadium in Boston, Massachusetts.

The **Ball Status** shows the temporal evolution of the ball possession during the play, from the moment the ball is pitched to the end of the play. It encodes all events related to the ball and makes clear, at any time, who has the possession of the ball. The horizontal axis represents the time, and the color of the bar, the player with the possession of the ball. Events are represented as icons on the timeline; the pitch is represented by a ball, the batting is represented by a bat, the moment the ball was caught (or gloved) as a mitt and a strike as an "x". This view fulfills requirements R2 and R4. Figure 5 shows three possible configurations for the Ball Status view: (1) 5a shows a strike, (2) 5b shows a batted ball and (3) 5c shows a ball caught in midair.

The **Play Timeline** shows the movements of the players in the vicinity of the bases. It consists in a two-dimensional line chart, with the horizontal axis encoding time, and the vertical axis, the mapping of the trajectories of the players. The chart is constrained to a region of interest around the area delimited by the bases, since this area encloses most of the complex interactions among baseball players. The position of the player is encoded as the angle with respect to the vector defined between the pitcher mound and the home plate. Figure 6 illustrates how this mapping is performed. In 6a, we show the infield and the four bases, with their respective angles. In 6b, we show how the angle is mapped in a line, with the bases annotated on the right. The home plate is represented by 0 radians, the 1st base, by $\pi/2$ radians, 2nd base by π radians and 3rd base by $3\pi/2$ radians. We encode the home plate again using 2π radians in order to present the full path of the runners through the bases without discontinuities. This encoding enables users to visualize the movement of the runners as they go over bases, as well as the strategy of the defense players, as they try to reach the runner or a base. Visual clutter is avoided by representing only players in the vicinity of the bases. However, even if all players were to be in this region, the chart would still be readable: according to Munzner [Mun14] (Chapter 12), superimposed line charts can be understood up to a few dozen lines shown simultaneously. The two most relevant events in baseball are shown using special marks: when a runner saves a base (reaches a base safely), we encode this action by a small diamond. When a runner is tagged, i.e. touched by a defensive player that is holding the ball, we encode this action by a small circle. These events are present in Figure 1.

Baseball is characterized by a set of interesting plays, and among them is the Grand Slam. A Grand Slam is a home run hit with all bases loaded (all bases occupied by runners), which grants four runs to the offensive team. Figure 7 shows a grand slam from the bottom half of the 7th inning of the Los Angeles Dodgers and Milwaukee Brewers game (June 3rd, 2017). In this play, batter Travis Shaw hits a grand slam to right center field and runners Eric Sogard, Domingo Santana and Jesus Aguilar score. Even by taking into consideration that this might be considered a simple play, the Play Timeline highlights some interesting aspects, like the reaction time of each runner, the way they move back and forth while waiting for the landing of the batted ball, and the speeds of the runners in relation to each other. This is the type of temporal data the Play Timeline was created to highlight.

Implementation Details

Our visualization was implemented as a client-server web application. The client is written in Javascript with the library D3.js, and

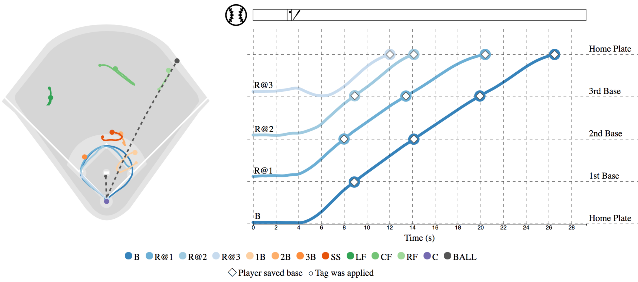


Figure 7: 06/03/2017 - Los Angeles Dodgers @ Milwaukee Brewers: Travis Shaw hits a Grand Slam

is responsible for rendering the Baseball Timeline chart. The server is written in Python, and is responsible for querying plays from the MLB server using the Statcast JSON API, preprocessing the data (tracking, events, as well as game metadata) and sending it to the client through a REST API. This infrastructure is ready to be deployed in order to present plays to fans and experts alike.

7. Analysis of Baseball Plays

In this section, we present six baseball plays and show how they are represented in the Baseball Timeline approach. The plays are shown in Figure 8, in the same way they were presented during the interviews with the baseball experts, which are detailed in the next session.

[Play 1] The first play we present is from the bottom of the first inning of the Toronto Blue Jays versus Boston Red Sox game in April 15, 2016. The play is shown in Figure 1 with annotations and image footage of the game, and without annotations on figure 8a. In this play, batter Travis Shaw hits a fly ball to center fielder Kevin Pillar and safely reaches 2nd base. Runners David Ortiz and Hanley Ramirez reach home plate and score. However, Travis Shaw is stranded between 2nd and 3rd bases, and at 22 seconds in the play, is tagged out by catcher Josh Thole. This is a complex play, and can be hardly understood directly from the Play Diagram. However, all the actions become clear in the timeline view. The batter path contains a sinuous line, as he moves back and forth 3rd base, while he is surrounded by the defensive players.

[Play 2] Top of the eighth inning of the St. Louis Cardinals versus Pittsburgh Pirates game (April 06, 2016). Batter Greg Garcia hits the ball in the direction of the right fielder Gregory Polanco. The fielder grabs the ball and throws to second baseman, who then assists the first baseman. Greg Garcia reaches second base safely, and the runner at first base, Jeremy Hazelbaker, reaches third base safely. Figure 8b shows how this information is conveyed by the timeline view. The actions around the 3rd base, especially, are clearly depicted through the time.

[Play 3] Top of the eighth inning of the Arizona Diamondbacks versus Kansas City Royals game, September 30, 2017. In this play, batter Rey Fuentes grounds out to first baseman Cheslor Cuthbert (Figure 8c). The batter is the third out, so the movement of the runner at 3rd, also depicted on the chart, does not contribute to the play. This is highlighted by the lack of the glyph at the end of the trajectory of the runner.

[Play 4] In the fourth play, on the other hand, the batter reaches first safe (top of the fourth inning of the Detroit Tigers and Chicago Cubs game, in August 18, 2015). The batter Tyler Collins hits a ground ball to second baseman Chris Coghlan, who throws it to first baseman Anthony Rizzo. The actions around the first base are clearly depicted in the timeline view, while they can be hardly understood in the Play Diagram. This play is particularly interesting, since the order of events completely changes its outcome. More specifically, the order at which players reach the base determines whether the offensive team scores or not.

[Play 5] The fifth example is a 6-4-3 (double play), when the defense team makes two outs during the same play (Figure 8e). This play is from the top of the fourth inning of the Tampa Bay Rays versus Toronto Blue Jays game, in April 15, 2015. The batter Rene Rivera hits a ground ball. Shortstop Jose Reyes takes the ball, pass it to second baseman Devon Travis, who passes it to first baseman Justin Smoak. Rivera is out at first and runner Kevin Kiermaier is out at second base. The timeline view shows the exact moment at which the runner and batter are out. With the timeline, we see that both the batter and the runner arrived at their bases approximately one second too late.

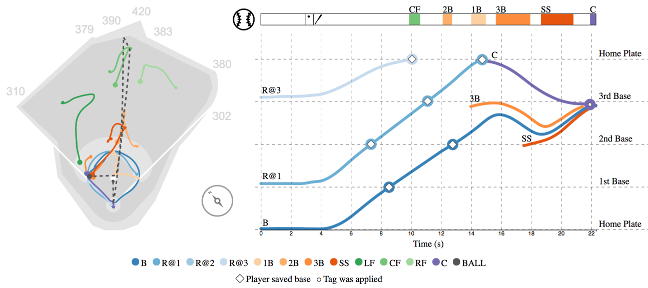
[Play 6] Finally, our last play is from the top first inning of the Washington Nationals versus Atlanta Braves game, in June 30, 2015. This play is relatively simple, with batter Danny Espinosa hitting the ball all the way to the outfield and making a double, and runner Denard Span scoring a run (Figure 8f). However, the timeline visualization can expand the reader's knowledge about the play even in simple cases: using the ball status view, we notice that Span had plenty of time to score a run and reach home plate. The third baseman only got the ball after the runner reached home, and by that time it was too late to tag him.

This section described six distinct baseball plays using Baseball Timeline. Our examples show that BT is a powerful tool to explain baseball plays without the need of animation. In the next section, we describe the feedback we received from four baseball experts on our visualization.

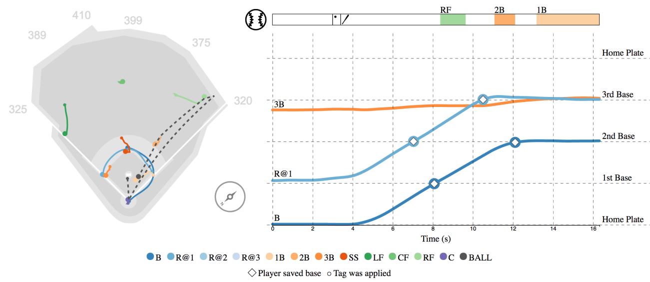
8. Domain Experts Interviews

We conducted interviews with four domain experts, two sabermetricians and two expert-fans, in order to evaluate and gain feedback on our visualization. The two sabermertricians, S1 and S2, are knowledgeable in data science and have published peer reviewed papers in the field of baseball analytics. The two expert-fans, F1 and F2, understand the sport deeply, and have been following it for more than 10 years. These four domain experts were chosen because we wanted to see how well our visualization conveys play information to professional experts and expert-fans alike.

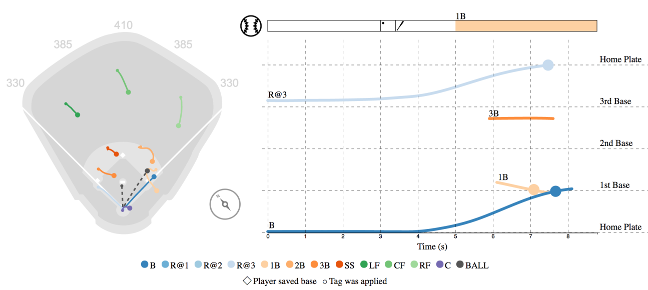
The interviews occurred as follows: firstly, we made sure the expert was familiar with the Play Diagram. Next, we presented our visualization and clarified any questions that they had. We showed them one play using the Play Diagram, followed by Baseball Timeline. Then, we began the qualitative evaluation of the tool. In order to perform the interview, we showed them the six baseball plays discussed in Section 7. For each play, we first showed the traditional Play Diagram and asked the expert to describe it. Next, we



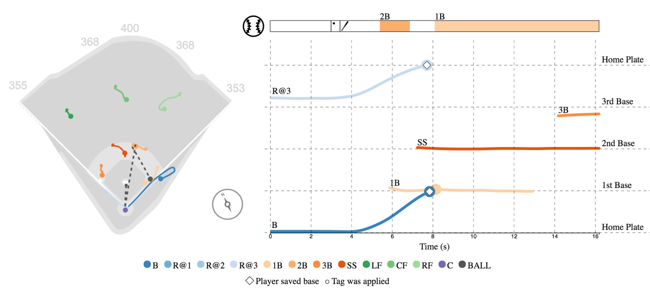
(a) 04/15/2016 - Toronto Blue Jays @ Boston Red Sox: runners David Ortiz and Hanley Ramirez score. Batter Travis Shaw is tagged out.



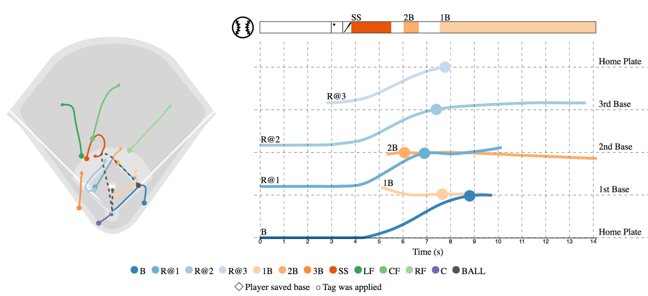
(b) 04/06/2016 - St. Louis Cardinals @ Pittsburgh Pirates: batter Greg Garcia doubles. Runner Jeremy Hazelbaker reaches third base.



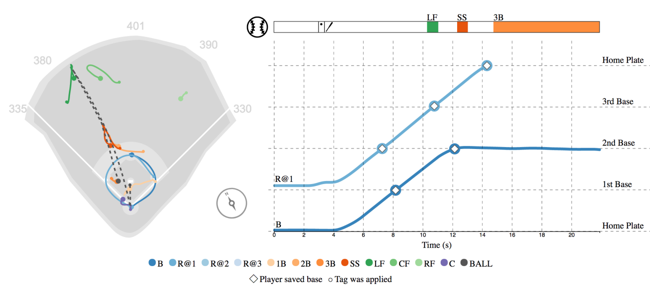
(c) 09/30/2017 - Arizona Diamondbacks @ Kansas City Royals: batter Rey Fuentes is out.



(d) 08/18/2015 - Detroit Tigers @ Chicago Cubs: batter Tyler Collins reaches first base just before first baseman with the ball.



(e) 04/15/2015 - Tampa Bay Rays @ Toronto Blue Jays: double play. Batter Rene Rivera and runner Kevin Kiermaier are out.



(f) 06/30/2015 - Washington Nationals @ Atlanta Braves: batter Danny Espinosa doubles. Runner Denard Span scores a run.

Figure 8: Six baseball plays represented by Baseball Timeline.

showed them Baseball Timeline, and asked them if they could identify any features that they did not notice before. We played the video footage of the play, and asked the expert if their conclusions were correct. After all six plays were analyzed, we asked the experts for feedback on our visualization. In order to steer the discussion, we asked them the following questions:

1. Did you have any difficulties learning to read the visualization? Why?
2. Did the timeline help you to better understand the play? Why?
3. Which parts of the chart do you think can be improved?
4. Can you think of any information that is not shown in the chart, but is relevant for play understanding?

8.1. Describing Plays

Our subjects described the six plays presented in Section 7 using the traditional Play Diagram followed by Baseball Timeline. Both experts and expert-fans made a few incorrect descriptions using the Play Diagram, but *described all plays correctly using Baseball Timeline*. Table 1 shows the correct and incorrect descriptions that our experts made using solely the Play Diagrams. We briefly describe the mistakes they made, and what caused them.

[Play 1] None of the participants were able to correctly describe Play 1 (Figure 8a) based on the Play Diagram. Due to clutter, S1 and F2 could not see how many runners there were on the bases. S2 and F1 could see the runners, but they thought that the play was a triple (batter reaching third base). After seeing the timeline and ball status view, all participants described the play correctly.

[Play 2] All participants described Play 2 correctly with the Play

	Play 1	Play 2	Play 3	Play 4	Play 5	Play 6
S1	✗	✓	✓	✗	✓	✓
S2	✗	✓	✓	✗	✗	✓
F1	✗	✓	✓	✗	✓	✓
F2	✗	✓	✗	✗	✓	✓

Table 1: Play descriptions *without* using Baseball Timeline (✓- correct description, ✗ - incorrect description). *With* Baseball Timeline, S1, S2, F1 and F2 described every play correctly.

Diagram. Play 2 (Figure 8b) had a very uncluttered Play Diagram. Moreover, there were no defense players protecting bases or tagging runners, therefore it was clear for the experts that both batter and runner at first were safe at their bases. The participants used Baseball Timeline to confirm their predictions.

[Play 3] S1, S2 and F1 described Play 3 (Figure 8c) correctly, stating that they knew the batter was out because all the players were moving out of the field when the play was over. F2, however, did not notice this fact, and could not describe the play, saying that he was not sure if the batter saved first base or not. After seeing the timeline and ball status views, F2 was able to correctly describe the plays.

[Play 4] None of the participants could describe Play 4 (Figure 8d) using the Play Diagram, particularly because they were not sure who arrived at base first: the batter or the first baseman. After seeing Baseball Timeline, all the experts described this play correctly.

[Play 5] S1, F1 and F2 described Play 5 (Figure 8e) correctly. S2, however, described Play 5 as being a ground out to second baseman after reading the Play Diagram. He realized that he was wrong after reading the timeline visualization, stating that the batter was actually safe at first base.

[Play 6] All participants described Play 6 (Figure 8f) correctly. Similarly to Play 2, the Play Diagram was uncluttered and it was visible that batter and runner at first were safe at second base and home plate respectively.

With this experiment, we noticed that the timeline representation is able to help readers to better understand and describe baseball plays, especially in cases where the Play Diagram was too cluttered or did not provide enough temporal information about the game events. All subjects, both experts and expert-fans, were able to correctly describe plays after seeing the Baseball Timeline representation.

8.2. Expert Feedback

Overall, the experts liked Baseball Timeline and did not have any difficulty understanding or reading it. All subjects stated that the timeline chart helped them to better understand the plays. S1 and F1 appreciated the fact that they could see interesting properties of the play with the timeline approach, such as player speed and reaction time based on the slope of the line. F1 also made positive comments regarding the ball status view, saying that determining the player in possession of the ball was hard using the Play Diagram, but this information was easy to see in the ball status chart. S1, S2 and F2 mentioned that the Play Diagram and the timeline complement each other, because while the first described spatial information in detail, the second showed baserunner and ball events which are hard to read on the traditional diagram.

We also received feedback on how to improve our visualization. Originally, we used the d3 color scheme "category20" to encode player positions. However, that resulted in similar player positions having vastly distinct colors. For instance, while the second baseman was represented by orange, the third baseman was represented by green. S2 and F1 suggested that we used similar colors for similar players, therefore we changed the color scheme to d3 "category20c", and that resulted in a more consistent color encoding: batter and runners are represented by shades of blue, basemen are encoded by shades of orange, and outfielders, by shades of green.

The second suggestion we received was regarding the choice of players to be shown in the timeline visualization. Originally, these players were chosen based on a threshold of the distance to the bases in the field. More specifically, we chose only players that moved within 10 feet of the square with corners at the 1st, 2nd, 3rd bases and the home plate. However, S1 and S2 suggested that even if the player moved close to the square, he might not do anything useful in the play. Therefore, in our prototype, we kept the distance threshold approach, but we also enabled the user to manually show or hide players as they see fit in order to create a better chart. This feature helped to reduce clutter in the visualization and made it more easily readable. Note that this interaction does not violate requirement R4: once the analyst is satisfied with his Play Timeline, the chart is static and therefore it can be printed and distributed in static media.

9. Conclusion

This paper presented Baseball Timeline, a visualization that represents how baseball plays evolve through time. Using BT, we were able to translate complex baseball plays into a two-dimensional graph that is easy to understand and conveys both the temporal and the spatial aspects of the game simultaneously. We showed six use cases of our technique and described how it helps users to understand the evolution of the play throughout time. We also performed interviews with both sabermetricians and baseball fans, who were able to read and describe the plays quickly with our visualization. While the Play Diagram is the industry standard for visualizing baseball plays, we realized that it is not the most effective way for doing so. More specifically, we have shown cases where experts were not able to describe plays using the Play Diagram, but were successful in doing so with Baseball Timeline.

Limitations and Future Work

Baseball Timeline was tailored to show the spatiotemporal evolution of plays near the bases. However, it cannot represent the movement of players in the outfield, as the mapping of angle to position is not invertible. More specifically, distinct positions in the outfield may correspond to the same position in the timeline view, which results in visual clutter. As future work, we would like to investigate how to address this issue, in order to convey the movement of infielders, outfielders and the ball throughout the play. Another limitation of Baseball Timeline is that it relies on Statcast data, which may contain missing information about the game. One example of missing information are errors in passes between players, for example, when the right fielder throws the ball to the first baseman, but he misses. Such information is not available in the Statcast description of the game; therefore we cannot display it in our visualization. Should more events be available in the future, we can display them using special marks in the Play Timeline, similarly to what is done with the events "player reached a base" and "player was tagged". Since the number of events in a baseball play is limited, and they are spread through time, we do not expect visual clutter to be an issue. The third limitation of our work is that it can only describe individual plays. A baseball game contains dozens of plays, therefore looking at individual BT charts is not a convenient way of understanding what happened in an entire game. We would like to address this issue in a future work, by creating summaries for entire baseball games that can be read and interpreted in a short amount of time. Moreover, given the vast amount of baseball tracking and statistics data available, we would like to enable users to perform advanced queries and find patterns in past games. This feature would allow analysts to identify interesting plays in the collection and rank players in order to find the season Most Valuable Player (MVP). We can build upon the work of Lage et al. [LOC*16], who proposed a methodology to enable the execution of complex queries in baseball data.

This paper was focused on the description of baseball plays with emphasis on the temporal events that happen during the game. We took advantage of patterns in the movement of the runners and used a visual encoding similar to Marey's Graph in order to convey spatiotemporal information about the play. As future work, we would like to explore similar approaches to other sports, such as soccer, basketball and hockey. We can also see our visualization applied to other domains with spatiotemporal data, especially those where common paths are found, for example, production lines and project management.

Acknowledgements

We would like to thank Dan Cervone, Justin Chiang, Nina Wiedemann, Fabio Sikansi, Lukas Younker and Neel Dey for their help with the paper. MLB Advanced Media provided partial funding for the project and access to data. This work was also supported in part by: the Moore-Sloan Data Science Environment at NYU; NSF awards CNS-1229185, CCF-1533564, CNS-1544753, CNS-1730396. J. P. Ono and C. T. Silva are partially supported by the DARPA D3M program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA.

References

[AA08] ALBINSSON P.-A., ANDERSSON D.: Extending the Attribute Explorer to Support Professional Team-Sport Analysis. *Information Visualization 7*, 2 (June 2008), 163–169. 3

[AAB*13] ANDRIENKO G., ANDRIENKO N., BAK P., KEIM D., WROBEL S.: *Visual Analytics of Movement*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. 3, 4, 5

[AAS*12] ANDRIENKO N., ANDRIENKO G., STANGE H., LIEBIG T., HECKER D.: Visual Analytics for Understanding Spatial Situations from Episodic Movement Data. *KI - Künstliche Intelligenz 26*, 3 (2012), 241–251. 4

[CDBG14] CERVONE D., D'AMOUR A., BORNN L., GOLDSBERRY K.: POINTWISE: Predicting points and valuing decisions in real time with NBA Optical Tracking Data. In *MIT Sloan Sports Analytics Conference* (2014), vol. 28. 3

[CS06] COX A., STASKO J.: Sportsvis: Discovering meaning in sports statistics through information visualization. In *Compendium of Symposium on Information Visualization* (2006), pp. 114–115. 3

[CS15] CROSS J., SYLVAN D.: Modeling spatial batting ability using a known covariance matrix. *Journal of Quantitative Analysis in Sports 11*, 3 (2015), 155–167. 3

[DKVS14] DIETRICH C., KOOP D., VO H. T., SILVA C. T.: Baseball4d: A tool for baseball game reconstruction amp; visualization. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2014), pp. 23–32. 3

[Gol12] GOLDSBERRY K.: Courtvision: New visual and spatial analytics for the nba. In *MIT Sloan Sports Analytics Conference* (2012). 3

[Hag70] HAGERSTRAAND T.: What About People in Regional Science? *Papers in Regional Science 24*, 1 (Jan. 1970), 7–24. 4

[Kra03] KRAAK M.: The space-time cube revisited from a geovisualization perspective. In *International Cartographic Conference* (2003), pp. 1988–1996. 4

[LMGH15] LUKASCZYK J., MACIEJEWSKI R., GARTH C., HAGEN H.: Understanding hotspots: A topological visual analytics approach. In *SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2015), ACM, p. 36. 4

[LOC*16] LAGE M., ONO J. P., CERVONE D., CHIANG J., DIETRICH C., SILVA C. T.: StatCast Dashboard: Exploration of Spatiotemporal Baseball Data. *IEEE Computer Graphics and Applications 36*, 5 (Sept. 2016), 28–37. 3, 10

[LTB16] LOSADA A. G., THERÓN R., BENITO A.: BKViz: A Basketball Visual Analysis Tool. *IEEE Computer Graphics and Applications 36*, 6 (Nov. 2016), 58–68. 3

[Mar78] MAREY E.-J.: *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine*. G. Masson, 1878. 4, 5

[MB13] MOON B., BRATH R.: Bloomberg sports visualization for pitch analysis. In *Workshop on Sports Data Visualization* (2013). 3

[Med15] MEDIA M. A.: Statcast Leaderboard. http://m.mlb.com/statcast/leaderboard, 2015. [Online; accessed 12-Dec-2017]. 2, 4

[Med17] MEDIA M. A.: At Bat. http://m.mlb.com/apps/atbat, 2017. [Online; accessed 12-Dec-2017]. 2

[Mun14] MUNZNER T.: *Visualization analysis and design*. CRC press, 2014. 4, 5, 6

[POJC01] PINGALI G., OPALACH A., JEAN Y., CARLBOM I.: Visualization of Sports Using Motion Trajectories: Providing Insights into Performance, Style, and Strategy. In *Conference on Visualization* (2001), VIS '01, IEEE Computer Society, pp. 75–82. 3

[PSBS12] PILEGGI H., STOLPER C. D., BOYLE J. M., STASKO J. T.: SnapShot: Visualization to Propel Ice Hockey Analytics. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (Dec. 2012), 2819–2828. 3

[PVF13] PERIN C., VUILLEMOT R., FEKETE J. D.: SoccerStories: A Kick-off for Visual Soccer Analysis. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2506–2515. 3

[SAMS*17] SACHA D., AL-MASOUDI F., STEIN M., SCHRECK T., KEIM D. A., ANDRIENKO G., JANETZKO H.: Dynamic visual abstraction of soccer movement. In *Computer Graphics Forum* (2017), vol. 36, Wiley Online Library, pp. 305–315. 3

[SJB*16] STEIN M., JANETZKO H., BREITKREUTZ T., SEEBACHER D., SCHRECK T., GROSSNIKLAUS M., COUZIN I. D., KEIM D. A.: Director's Cut: Analysis and Annotation of Soccer Matches. *IEEE Computer Graphics and Applications 36*, 5 (Sept. 2016), 50–60. 3

[SJL*17] STEIN M., JANETZKO H., LAMPRECHT A., BREITKREUTZ T., ZIMMERMANN P., GOLDLÜCKE B., SCHRECK T., ANDRIENKO G., GROSSNIKLAUS M., KEIM D. A.: Bring it to the Pitch: Combining Video and Movement Data to Enhance Team Sport Analysis. *IEEE Transactions on Visualization and Computer Graphics PP*, 99 (2017), 1–1. 3

[TC10] THERON R., CASARES L.: Visual Analysis of Time-Motion in Basketball Games. vol. 6133, pp. 196–207. 3

[Tuf01] TUFTE E. R.: *The visual display of quantitative information*. Graphics Press Cheshire, CT, 2001. 3, 4

[WG12] WONGSUPHASAWAT K., GOTZ D.: Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2659–2668. 3