HAL
open science

State of the Art of Sports Data Visualization

Charles Perin, Romain Vuillemot, Charles D Stolper, John T Stasko, Jo Wood, Sheelagh Carpendale

To cite this version:

Charles Perin, Romain Vuillemot, Charles D Stolper, John T Stasko, Jo Wood, et al.. State of the Art of Sports Data Visualization. Computer Graphics Forum, 2018, 37 (3), pp.1-24. 10.1111/cgf.13447 . hal-01806107

HAL Id: hal-01806107

https://hal.science/hal-01806107v1

Submitted on 1 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EUROVIS 2018
R. S. Laramee, G. E. Marai, and M. Sedlmair
(Guest Editors)

Volume 37 (2018), Number 3
STAR – State of The Art Report

State of the Art of Sports Data Visualization

C. Perin[1,2], R. Vuillemot[3], C. D. Stolper[4], J. T. Stasko[5], J. Wood[1], and S. Carpendale[2]

[1]Department of Computer Science, City, University of London, UK
[2]Department of Computer Science, University of Calgary, Canada
[3]École Centrale de Lyon, Université de Lyon, France
[4]Department of Mathematics and Computer Science, Southwestern University, USA
[5]School of Interactive Computing, Georgia Institute of Technology, USA

Abstract

In this report, we organize and reflect on recent advances and challenges in the field of sports data visualization. The exponentially-growing body of visualization research based on sports data is a prime indication of the importance and timeliness of this report. Sports data visualization research encompasses the breadth of visualization tasks and goals: exploring the design of new visualization techniques; adapting existing visualizations to a novel domain; and conducting design studies and evaluations in close collaboration with experts, including practitioners, enthusiasts, and journalists. Frequently this research has impact beyond sports in both academia and in industry because it is i) grounded in realistic, highly heterogeneous data, ii) applied to real-world problems, and iii) designed in close collaboration with domain experts. In this report, we analyze current research contributions through the lens of three categories of sports data: box score data (data containing statistical summaries of a sport event such as a game), tracking data (data about in-game actions and trajectories), and meta-data (data about the sport and its participants but not necessarily a given game). We conclude this report with a high-level discussion of sports visualization research informed by our analysis—identifying critical research gaps and valuable opportunities for the visualization community. More information is available at the STAR's website: `https://sportsdataviz.github.io/`.

1. Introduction

Sports have a long tradition of data collection and reporting. In the eighth century BCE, spectators recorded the results of the Ancient Olympic Games. In the 1870s, Henry Chadwick introduced the baseball scorecard for collecting statistical information about baseball games beyond simply team runs scored that is still in use (with some minor adjustments) today [Dic97].

There is ample evidence that successful analysis of sports data can result in better performances of a team [Lew04, MW12, Win12] and can have notable positive economic impact [OJK13]. Scouts use data to find talented players; trainers to improve player performances; analysts to find even the slightest edge over their rivals. Journalists report on sports using data meticulously maintained by warehousers such as the Elias Sports Bureau, STATS LLC, and Opta. Their reporting in the past decade has included the design of intricate and informative infographics of this data. Fans follow their favorite teams and players using statistics, once using newspaper box scores and now utilizing free sources such as the Sports Reference family of sites [Spo16a] and Fangraphs [Fan16]. Beyond the realm of professional sports, individuals actively participate in a recreational facet of sports data by collecting statistics on their own activities to track, maintain, and improve their performance.

Sports visualization offers new approaches to exploring, making sense of, and communicating sports data. As visualizations can be more accessible and more meaningful than traditional statistical analysis [Ber15, JSS*14], the number of visualizations of sports data has grown rapidly over the past decades. Print and online media companies, including The New York Times (e.g., [QW14, AQ16, AQ14, Tim07, DQ14, QR16a]), present sports data using infographics and interactive visualizations. Dedicated blogs such as FiveThirtyEight have become extremely popular for their analyses of sports events, players, and economics using visualizations. Television broadcast and betting websites increasingly use statistics and graphics to evaluate and predict sports events.

In visualization research, there has been a similarly growing interest in sports data over the past 20 years. The increasing numbers of publications related to the field at leading conferences such as IEEE VIS, EuroVis, and ACM CHI demonstrate this interest. The MIT Sloan Sports Analytics Conference has been gathering professionals and academics since 2007. In 2013, the first workshop on sports data visualization was organized at the IEEE VIS Conference [BCC*13]. In 2016, an IEEE Computer Graphics & Applications special issue was dedicated to sports visualization [BS16]. Figure 1 shows the distribution, by year, of published research articles on sports visualization we gathered.

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd. Published by John Wiley & Sons Ltd.

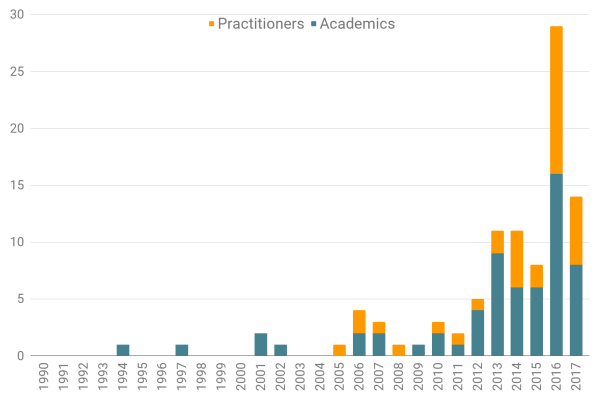

Figure 1: *The 98 sports data visualization articles from both academics and practitioners we collected, grouped by year. This chart emphasizes the recent growth in research surrounding sports data.*

While many academic papers, including other state of the art reports, include sports visualizations, sports as an application domain has not yet been the subject of a review from a visualization perspective. Visualization of team sports has recently been surveyed [GH17], however, this previous survey focuses solely on team sports and takes a fully spatio-temporal lens to looking at visualizations used in this context. Our paper differs in both aspects as we integrate both team sports and individual sports and discuss box-score data and meta-data in addition to (spatio-temporal) tracking data.

Existing sports visualization research stems from passionate, often distributed, communities working on sports data analysis and communication. With the increasing interest in sports visualization from both sports practitioners and visualization researchers, there is a need to document and analyze these efforts. In our review we identify three types of sports data that are used in sports visualizations: box score data (data containing statistical summaries of a sport event such as a game), tracking data (data about in-game actions and trajectories), and meta-data (data about the sport and its participants but not necessarily a given game). Using these three data types, we structure our collection of 98 works from both academics and practitioners in the space and lay out the opportunities and challenges for the future of this fascinating subfield.

We start by describing the methodology and scope of our review in section 2. Then, in section 3, we introduce box-score data, tracking data and meta-data, the three types of sports data we extracted from analyzing the literature in sports data visualization. Sections 4, 5, and 6 then organize the literature using these three types of sports data. Finally, we reflect on our review to discuss both core and under-explored topics in sports data visualization, unveiling challenges for the area of sports visualization in particular but also for information visualization in general.

2. Methodology and Scope

Before diving into the review of sports data visualization, in this section we describe our methodology for building our collection of related research papers and newspaper and blog articles. We then explain the main dimensions that emerged from our analysis of the literature and form the high-level structure of this report.

2.1. Building the Collection

To review the state of the art in sports data visualization, we started with personal collections of *seed papers* and *links to infographics*. Authors of this report had been building these collections over several years, in some cases from their previous work on sports data visualization or in others from their interest in a particular sport or player. The duplicates that resulted from merging these personal collections were important as these were cues of particular relevance of the contribution, sport, or event.

The second step was to **expand** the collection in order to remove biases on favorite sports and expand the scope of sports and events. We primarily used Google Scholar, the ACM DL Library, and IEEEXplore in this phase. We used keywords such as `sports visualization` and `sports data visualization` to search for recent papers that we did not know of and to reach other communities.

We further reviewed hundreds of infographics as they are a significant driver of advances in sports visualization. This review made it possible to contrast the work in the research community and the work from sports practitioners. For selecting these newspaper and blog articles among the thousands that exist, we decided to go for broadness and diversity and include any new visualization that seemed to offer a new perspective on sports data visualization. We also searched on the Web for non-research sports visualization that already had a citation entry. Google Images was also helpful for filtering out non-visual entries.

A third step was collection **completion**. As we identified authoritative sources of publications and newspapers (e.g. New York Times sports infographics), as well as the most prolific sports data sources and tracking technologies, we visited websites to browse complete collections of articles in conferences or tagged articles on newspapers websites (e.g. `data-visualization` on FiveFirthyEight and `olympics` on FlowingData).

The fourth and final step was to **reduce** the collection. In this step, we discarded fringe articles and infographics whose contribution was below average, such as by: not making a strong contribution to sports data analysis; not providing a novel lens on sports data (e.g., simply using standard charts and not providing any particular insight); or not describing the underlying process enough to understand the data or the context sufficiently. We also discarded academic papers without a visual contribution (e.g., data mining and algorithmic papers that are related to sports data but do not contribute to visualization research). We also discarded papers that are about a generic visualization technique and use a sport as an example scenario (e.g., [JSSS16]).

This process resulted in 281 initial papers and articles, out of which we selected 98 for review in this report.

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

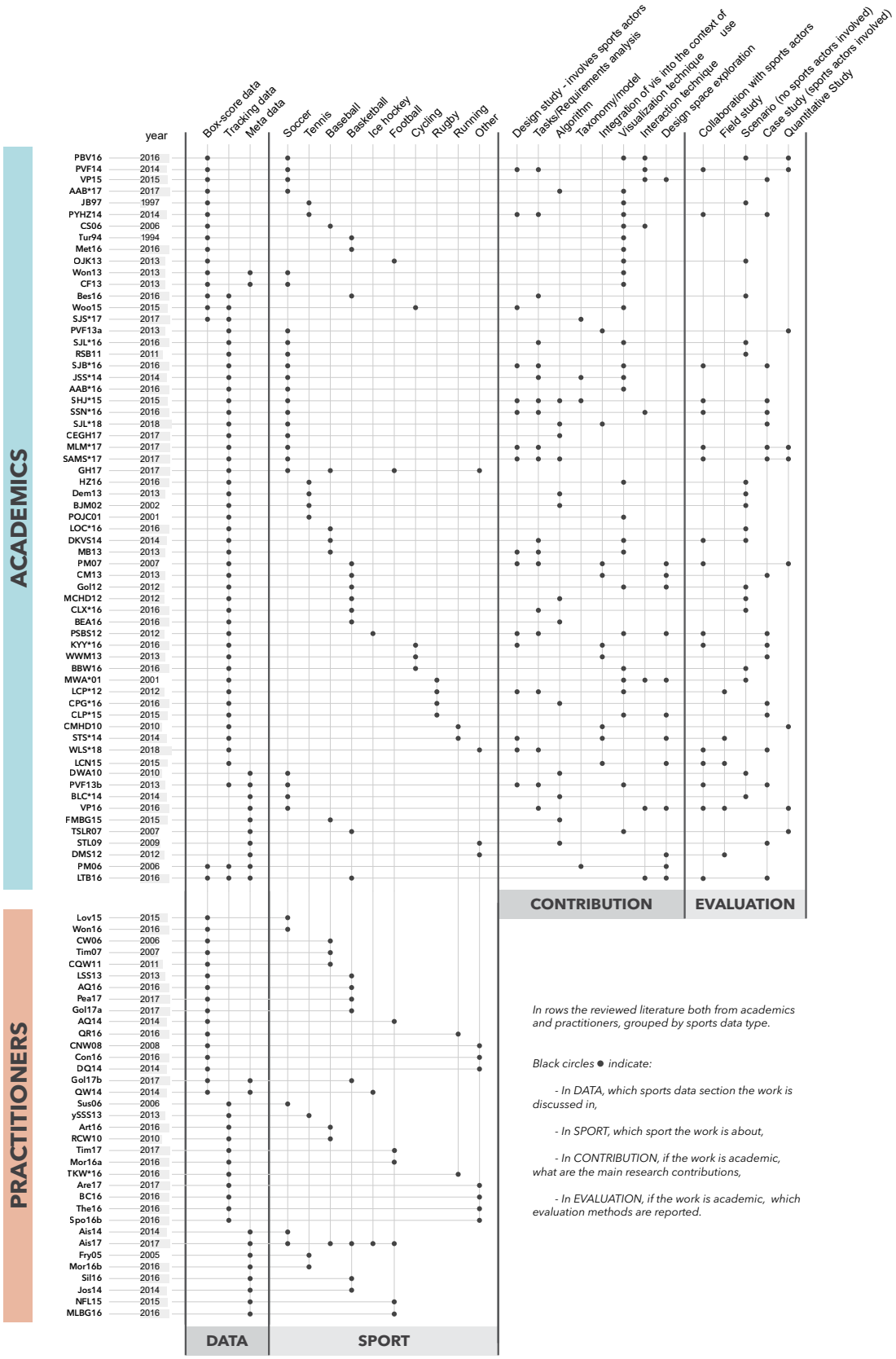

Figure 2: *The* 98 *articles we reviewed described in terms of sport, sports data type, contribution type and evaluation method.*

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

2.2. Analyzing the Collection

Two researchers independently open-coded a dozen elements of the collection with which they were the most familiar. After several iterations and refinements, we identified three overarching sports data types we review in this STAR and that we introduce in section 3:

Box score data corresponds to discrete, in-game event data and is detailed in section 4,
Tracking data corresponds to continuous spatio-temporal motion data and is detailed in section 5, and
Meta-data corresponds to discrete data beyond a specific game and is detailed in section 6.

In section 7 we further discuss trends that emerge from the research papers we reviewed in terms of **contribution type** and **evaluation method**. These descriptors make it possible to grasp the breadth and depth of contributions that emerge from sports data visualization, as well as extract the various strategies for conducting research in this area.

Figure 2 presents a summary of the 98 articles and research papers we analyzed in terms of the aforementioned descriptors.

3. Types of Sports Data

As explained in the previous section, we structure our survey based on the three types of sports data we derived from related papers analysis: *box-score data*, *tracking data* and *meta-data*.

One could apply almost any visualization to the sports domain, and many visualization types (e.g., node-link diagrams, time series line plots, and heatmaps) have been applied to sports. What makes sports a unique domain of interest is the variety and specificity of the data it generates. Thus, to discriminate the techniques used to visualize sports data, we look into the types of data that are visualized. In particular, while trajectory information and meta-data are often of a similar nature across sports, box-score data are highly sport-specific. The combination of these three types of sports data results in a rich and wide space of sports visualization.

We note ahead of time that these three data types are not perfectly orthogonal—the line between these data types is not always completely clear. For example, some professional sports leagues now use tracking data to compute box-score data and meta-data usually requires box-score and/or tracking data to be put in perspective. Manually collected event data with additional qualitative descriptors (e.g. location and speed of a shot on goal) could fit into either box-score data or the tracking data categories. The height of a baseball player or soccer goalie, which we consider to be meta-data, could also be considered tracking data as height can affect the size of the batters' box or the free space on goal. However, most data falls squarely into one of the three categories.

3.1. Box-Score Data

Sports have a long and constant tradition of recording critical observations. Spectators have been tracking the history of sports dating to very early competitions, such as the Ancient Olympic Games in 776 BCE. Since then, game results and details have been

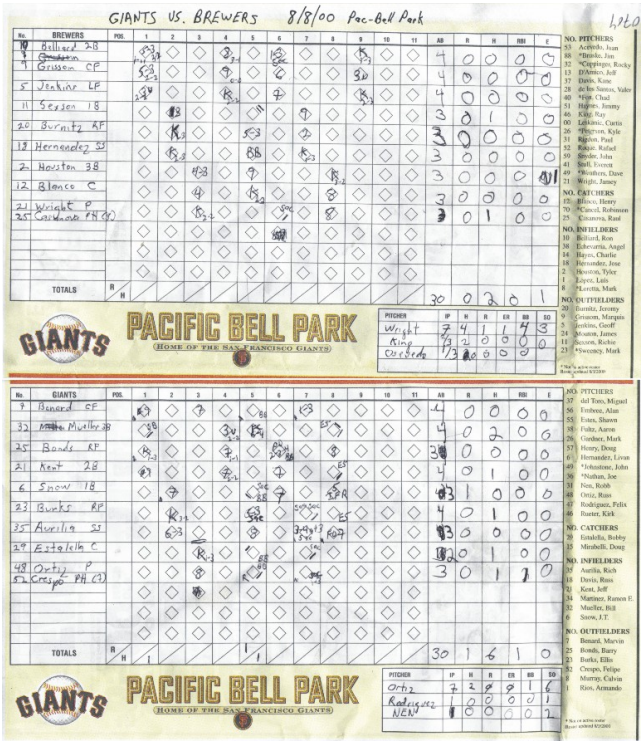

Figure 3: *Example of scorecard notation that captures events (e.g. errors, hits) in baseball. By Synaptidude at English Wikipedia [Public domain], via Wikimedia Commons.*

tracked by sport officials, journalists, and fans, using scorecards (see Figure 3). Such scorecards capture the structure of the game as discrete events such as substitutions, hits, and fouls by players over the course of the game.

The term *box-score* designates *the statistical summary of a game*. Although this term was originally used in baseball, it has been adopted by many sports including basketball and hockey. We use the term to describe any discrete data referencing in-game events. This can include both summary statistics (3 hits, 1 walk, 1 strikeout) as well as finer-grained event data (goal in the 83rd minute). We mentioned that the line between categories can be blurry. We tended to not consider any discrete data that had a spatio-temporal aspect to it (such as where on a court a tennis shot landed) as box-score data and instead categorized them as tracking data. Note that despite advances in technology, as of today most of these events and statistics are still generated by humans in sometimes quite tedious processes.

3.2. Tracking Data

Recent development in tracking and sensing technologies now enable collecting and analyzing new classes of data beyond the traditional box-score.

This tracking data contrasts with the box-score approach as the volume, variety, and precision grow exponentially and at cheaper costs. Most tracking data is collected by utilizing machine vision

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

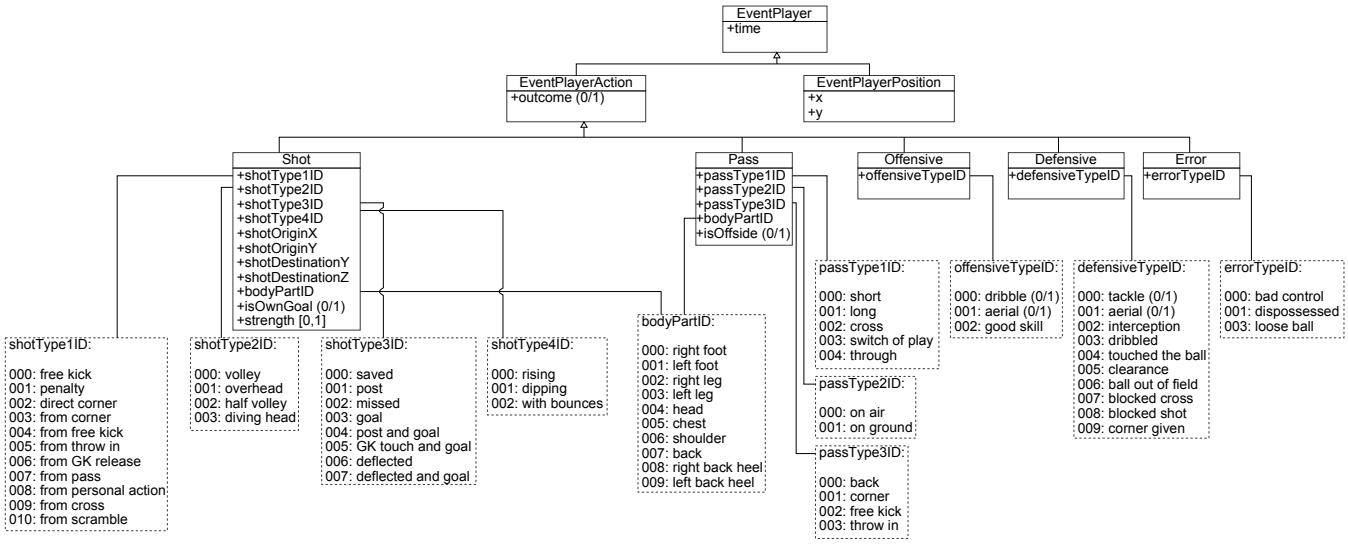

Figure 4: *Taxonomy of detailed soccer data used to collect and structure spatio-temporal events [PVF13a], based on Opta's data scheme. Image courtesy of the authors, used with permission.*

to gather precise spatio-temporal information about the players and equipment in real-time during the course of play. A variety of multi-camera tracking systems such as Hawk-Eye, K2 Panoramic Video system, Sportvision RACEf/x [Spo16b], PITCHfx/Statcast, SportVU, SportsCode, Catapult Sports' OptimEye S5 [Cat18], and Dartfish collect this form of data. (See [DKVS14] for a comprehensive list and comparison). A common underlying motivation for sports visualization research is to make use of the unprecedentedly detailed datasets generated by these technologies. Figure 4 demonstrates the complexity of the data with a taxonomy of soccer tracking data provided by one of the major sports data collection companies [PVF13a].

3.3. Meta-Data

Box-score and tracking data capture specific matches and events in those matches. Beyond this, there is additional data that surrounds sports that can add context to these events. Such data might concern rules, stadium capacities, physical characteristics of players, kit colors, team badges, and sponsor information to name a few. The data may also come from less sports-exclusive domains such as social media activity during a game.

Meta-data can also generate meta-data, for example correlations between box-score and tracking data or factors such as stadium formats (see Figure 5), players' diets, or weather conditions.

4. Visualization by type: Box-Score Data

We start our review with box-score data. 26 / 98 articles in our review focus solely on this type of sports data. The popularity of box-score data, in particular among practitioners, can be explained by its cheap and technology-independent means of acquisition. Although box-score data is often simple and small-scale, it is diverse and results in a wide variety of visualization approaches.

From an information visualization perspective, this type of data is different from for example, spatial data, as it is usually more abstract and strictly follows rules that define the types of events and occurrences recorded during a sports events. Sports rules are indeed the backbone of sports data, since tracking and meta-data (as defined in this paper) may be considered, in a sense, independent from the sport itself. Rules are, on the whole, stable over time but do occasionally change slightly—often driven by needs from media, TV broadcasting, and advertising—such as MLB's changes in 2015 [MLB15] to increase the pace of baseball games.

We separate the visualization of box-score data into two main categories: *time-evolving championship tables and rankings* relate to the explicit importance of time in sports and *scores, goals, and points* relate to the general sports rules.

4.1. Time-Evolving Championship Tables and Rankings

Sport competitions can span over long periods of time. Typically, team sports such as soccer, football, ice hockey, and basketball are played in championships, or leagues, every year. Teams play against each other two times or more over the course of several months. In some leagues this is either the entire competition (e.g., for European soccer championships) or the first phase of the competition (e.g., for the NBA and the NHL). The results of these games build standing tables, which determine who wins the competition, who is demoted to a minor league, and/or who progresses to a highly anticipated playoff stage.

Ranking (or standing) tables have been the standard way of representing and analyzing championships. They typically consist of vertical lists of teams' names and indicators of success such as wins and losses, points earned for victories, and/or the number of goals scored and conceded. The teams are ordered by *rank*, which is a derived index usually based on their number of points or ratio of wins to games played.

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

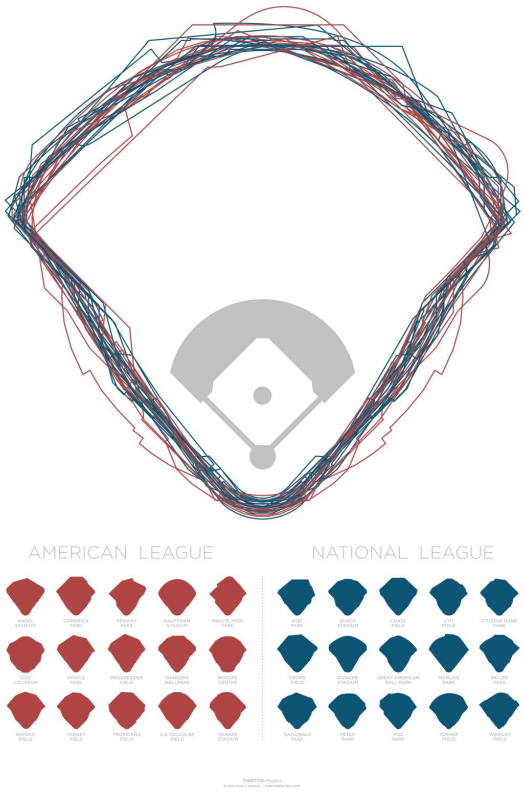

Figure 5: *Visual comparison of baseball stadium dimensions [Lou15] is an example of using meta-data to understand the influence of a baseball game's environment on teams' play. Image courtesy of the author, used with permission.*

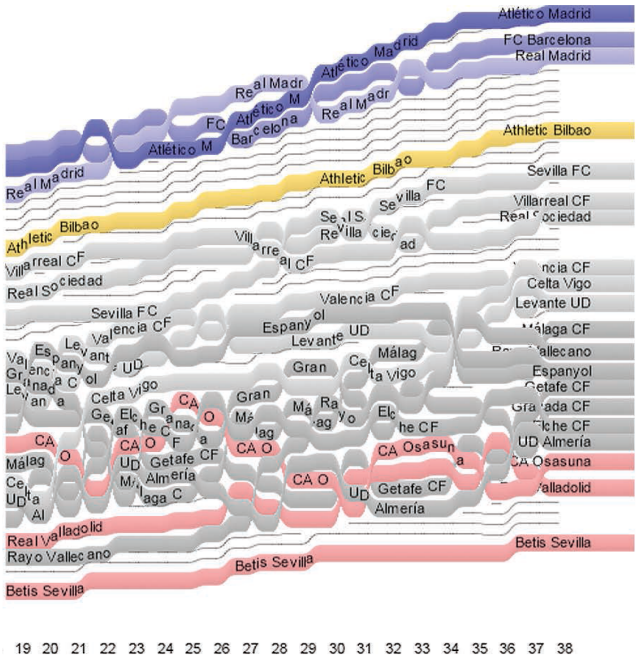

Figure 6: *The second half of the 2013–2014 Spanish Liga, visualized using GapChart [PBV16]. At each matchday, the lines representing the teams do not overlap. Teams that have the same number of points at a given time are shown with no white space between them. Thin gray lines indicate 3-point differences between two consecutively ranked teams. Image courtesy of the authors, used with permission.*

Allowing to navigate in the time dimension of tables. Because standard ranking tables do not convey the time-evolving aspect of championships [PVF14], existing work has looked into ways of combining the well-established and widely used ranking tables with visualizations that convey the changes over time.

In a series of works, Perin and Vuillemot [PVF14, VP15] explored new interaction techniques for navigating into the temporal dimension of ranking tables. Based on an analysis of tasks that soccer journalists and enthusiasts seek to perform, they proposed a way of interacting with the rank domain (such as the number of points of a team at a given time) to explore the time domain. They also proposed integrating a line chart visualization into the table and offered a way of transitioning between the two views [PVF14]. The authors further refined the technique to make it possible to navigate into ranking tables at different points in time by navigating directly along the trajectory of the rank of a team [VP15].

Showing relative gaps. Ranking tables have historically laid out all teams of a championship on a linear scale, i.e. with rows (teams) being simply ordered by rank. Showing rankings that evolve over time with a line graph then results in either a rank chart—where the y axis is an ordinal axis that conveys rank—or a score chart—where the y axis is a quantitative axis that conveys the result of the score function used to determine the rankings [PBV16].

While rank charts do not convey the magnitude of the gaps (or score differences) between teams, score charts convey these scores but result in overplotting and overlapping lines. A way of tackling this issue is to change the mapping on the y axis to convey not the rank or score of each team over time but the relative distances that separate the teams. This is particularly important when looking at championship data as the gaps between teams has more importance than the absolute score a team has. This technique facilitates: more accurate predicting of the final championship rankings; reflection on the full course of the competition; and comparison between the variability of different leagues. Both gap charts [PV13,PBV16] (see Figure 6) and Standings Tracer [Won16] (see Figure 7) take the approach of conveying the relative score differences between teams over time (note that their formulae for determining the mapping on the y axis differ slightly). Both techniques also feature a *focus* interaction to select a team and explore that team within the broader context of a season (see Figure 8).

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

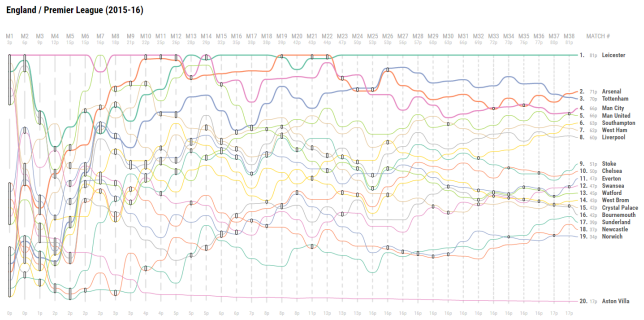

Figure 7: *Standings Tracer [Won16] showing the relative points of all teams in a championship over time. Image courtesy of the author, used with permission.*

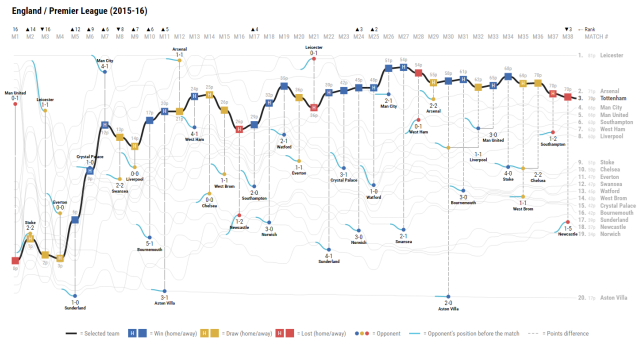

Figure 8: *Standings Tracer's [Won16] focus interaction. Selecting a team shows details about each game the team played during the championship. Color indicates the outcome of the game (win, loss, draw) and vertical lines connect the focus team's line graph to their opponent for each match day. This creates a fishbone of the team that shows the "gaps" between the focus team and each of its opponents. Image courtesy of the author, used with permission.*

4.2. Scores, Goals, and Points

Whether or not sports are framed within competitions, they obey specific rules. In many sports, teams or players participating in a game or competition need to score points to win each game (e.g., score goals in soccer, score points in tennis). Within a competition, points may be attributed according to their performance towards the competition (e.g., three points for a victory in most soccer championships and points given to each participant in a decathlon after each of ten events).

Contextualizing by overplotting. Many visualizations are dedicated to record-breaking points and scores. For example, Stephen Curry's 3-point record was put in context in one of The New York Times' infographics [AQ16]. This line chart shows one line for each of the 752 players who finished in the top-20 in 3-point attempts at the end of each season between 1980 and 2016. This visualization clearly shows how Stephen Curry is an outlier compared to the rest of the players, as he goes "off the charts". A similar and at least as famous infographic shows Barry Bonds's progress towards breaking the career home run record [CW06].

Many infographics from online newspapers create comparisons of players and teams by using multiple line charts. This is the case with The New York Times' Peyton Manning's touchdown passes record [AQ14]; Paths to the Top of the Home Run Charts [Tim07]; comparison of baseball players who reached 3000 hits in their career [CQW11] or of NBA players in terms of number of points scored during playoffs [Pea17]; and with the comparison of NBA players performances in terms of points, rebounds, and assists per game over the season [Gol17a]. One notable example of this technique is Five Thirty Eight's sumo project [Con16], showing the history of sumo from 1761 to 2017, using overplotting of line graphs with color coding and annotations. The density of information in those visualizations creates a baseline that usually gathers the majority of players from which a select few top performers visually stand out.

Other visualizations besides line charts make use of overplotting, such as scatterplots. For example, by overplotting all swimming world records over time [Are17] one can explore how swimming world records evolved across history, paces of change, and outliers. The Berlin 2016 marathon visualization [TKW*16] is another example that shows all the runners participating in the race on a single map. It is also possible to establish a goal and see where in the pack of runners one would need to be to achieve that goal.

Visualizing all statistics. In the visualization community, some complex exploratory tools support the analysis of scores, goals, and points in sports. SportsVis's baseline bar display [CS06] makes it possible to explore the performance of baseball teams or players over the course of one season. NBAVis [LSS13] consists of a scatterplot with multiple dimensions such as team, rebounds, assists, points, and blocks to explore NBA data. With Buckets [Bes16] and BKViz [LTB16], it is possible to explore the points a basketball team scored against each other team in the league, among many other facets of the data. Chung et al.'s knowledge-based ranking system features multiple coordinated views to explore the multiple facets of rugby games [CPG*16]. These approaches result in powerful tools in which many dimensions can be explored simultaneously—as opposed to more static infographics—but at the cost of being more complex to use.

Laying out in context. Often, sports visualizations that show scores, goals, or points show these values within the broader context of sport events (e. g., full games). For example, some sports like tennis have their scoring structured hierarchically. A tennis match consists of several sets, each set consists of several games, and each game consists of several points. The outcome of a game (who wins and who loses) follows this hierarchy, and the player who scored the most points is not guaranteed to win the match. To convey the bigger picture in which points are scored, treemaps have been used for visualizing this natural hierarchy [JB97].

As another example, Chris Love used Tableau to show all the shots on goal for all games of a Premier League's championship day [Lov15] (see Figure 10). The visualization shows a bar chart where the x axis represents time and each bar represents a 5 minute interval. The y axis—the length of each bar—encodes the shots on goal. This visualization conveys the dynamics of each game by showing important events (shots and goals) in the context of the timeline of the game.

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

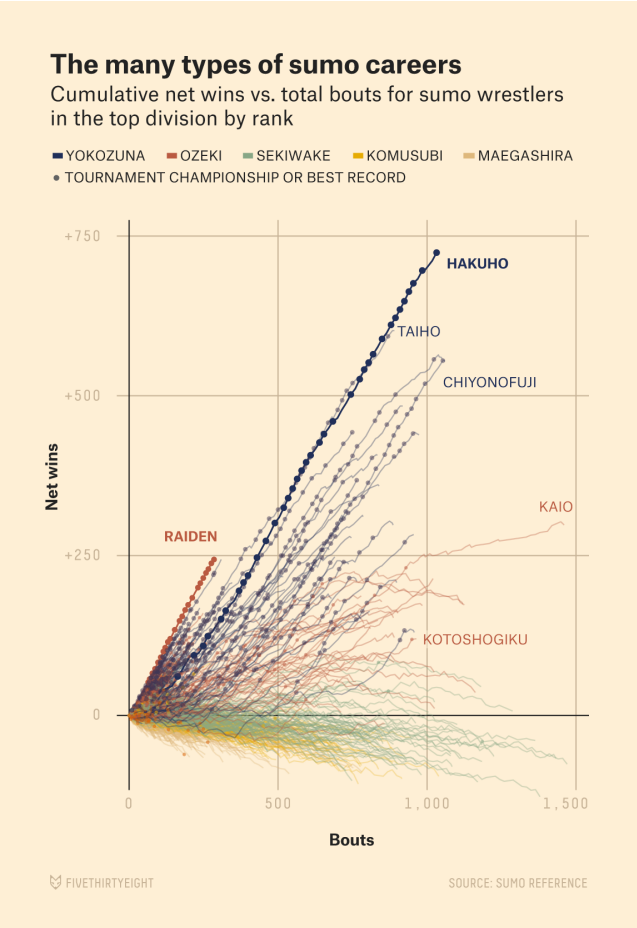

Figure 9: *A History of Sumo [Con16] is a stylized interactive visualization that presents the history of sumo using storytelling and exploration capabilities. Image courtesy of the author, used with permission.*

Showing relative differences rather than absolute values. In sports such as skiing, running, and racing, time is what matters most. The New York Times showed how close to winning participants were for each final of alpine skiing [Bou47], bobsled, luge, speed skating, and skeleton for the 2014 Winter Olympics [DQ14]. Instead of showing the absolute times of participants, they show the time difference relative to the winner. Relative times are particularly important in such sports where differences are often matters of milliseconds. For such short intervals, a sonification approach was used in the 2010 Winter Olympics to grasp the scale of infinitesimal differences between finish times [Ama10].

The New York Times also created an image showing all fastest men in the world by year, since 1896, with their position on the track at the time of the current world record [QR16a]. By representing each data point with a sprinter, the image looks like a photograph from the top of the track and the differences in performances are easily apparent to the reader.

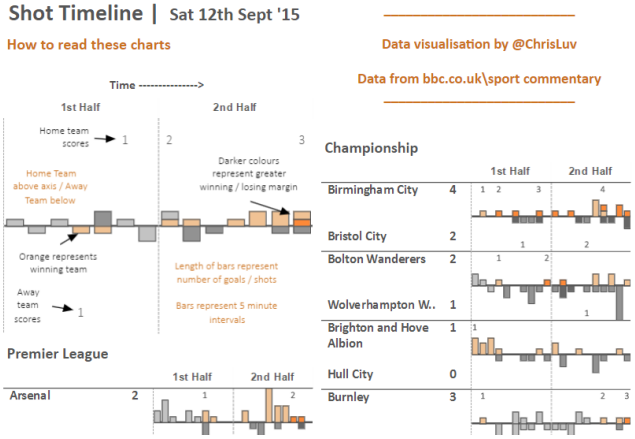

Figure 10: *Small multiples of football games putting the shots and goals in the context of soccer game timelines [Lov15] (cropped). Image courtesy of the author, used with permission.*

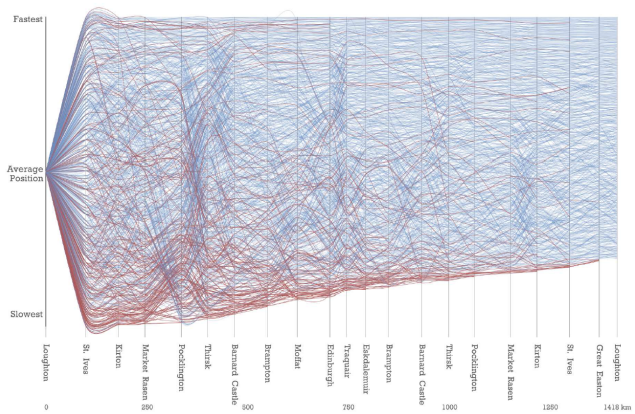

Figure 11: *Relative time differences between each cyclist participating in the race and the top performer at that moment [Woo15]. Red lines encode cyclists who did not finish the race. Image courtesy of the author, used with permission.*

In the context of golf, multiple Slope Graphs [Tuf01] have been used to show Tiger Woods's difference in performance compared to the rest of the world's top golfers in non-major and major tournaments [CNW08]. Once again, what matters is the relative difference between Woods and the other players rather than absolute numbers of strokes.

In the context of cyclist group racing, Wood created graphs (such as the one shown in Figure 11) showing the relative difference between each cyclist and the top performer at each point in time of the race [Woo15].

5. Visualization by type: Tracking Data

Sports *tracking* data is a natural evolution of box-score data that has become increasingly available due to recent advances in sensor and video motion detection technologies [GH17]. The majority

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

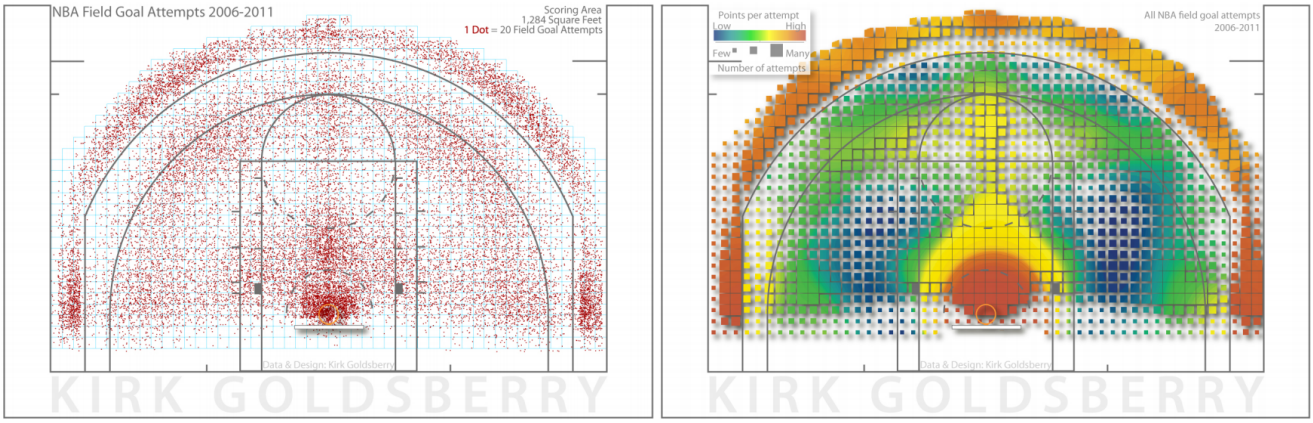

Figure 12: *Shot distribution and density in CourtVision [Gol12]. Sweet spots usually are below the hoop or behind the 3-point line. Image courtesy of the author, used with permission.*

of the articles we collected (55 / 98) make use of tracking data, for example to track player positions on a field, ball speed, and path after a shot. Tracking data usually consists of series of spatio-temporal events at a very fine grain of detail, with multiple dimensions such as player identities and type of action. The data scales with the number of games or players or by adding more sensors and cameras.

Because the data is spatio-temporal, it is not necessarily directly governed by the rules of the game. This is the primary difference as compared with box-score data which are summaries of events dictated by sports rules. Notably, new box-score data metrics—e. g., "Contested Points" and "Expected Goals"—can be derived from tracking data like shot attempts, player movements without the ball, or ball trajectories that did not result in a score change.

The richness and density of tracking data raises other challenges than the ones of box-score data, often involving some combination of algorithms, data mining, and visualization. In this section we separate general work in visualization of *events and trajectories* (often adopting a third-person perspective on the data) from work dedicated to representing the data from the *players' point of view* (often adopting a first-person perspective on the data).

5.1. Events and Trajectories

Event density map. The most straightforward way to plot raw tracking data is as events on a *dot map*, such as hit balls on a baseball pitch [DKVS14, MB13, LOC*16]. Although simple, dot maps provide enough details to characterize players' abilities and are a natural representations of ball trajectory origins or targets. Interestingly, they are commonly used as a visual signature characterizing a player or a team [Bes16] for complex play analysis [DKVS14].

Variations of density maps include using spatial binning [Gol12] to generate discretized heat maps using a preset color scale. The field is partitioned into regular areas which are each assigned a color according to the number of events that occurred in that area.

While these aggregations approximate otherwise accurate data (e.g. shots), they have been used to identify players' roles and teams' systems. For instance, "spray charts" [DKVS14] show where balls have been hit to—creating spatial groupings of hits that highlights areas of the field where events happen frequently.

Other variants of density maps include radial heatmaps to show ice hockey shots [PSBS12]. Density maps are often decorated with landmarks from the field of play—such as formalized line schemes—to provide scale and contextualize the data within the game's rules. In some cases, landmarks visually emerge on their own due to data distribution. In CourtVision [Gol12], the 3-point line appears due to the lack of shots just inside it (see Figure 12).

Events representation. Despite events having strong ties with their (x, y) position, they are often best represented with position in the display encoding task-dependent attributes. For example, baseball hit analysis has found that the combination of ball exit velocity and launch angle determines whether a hit will likely be a home run [Art16] (see Figure 13). A connected scatterplot by time allows the visualization's user to easily detect increases in average kick distance and field-goal percentage of kickers in American Football [Mor16a] (see Figure 14). GameFlow provides a grouping of events over time [CLX*16] to reveal patterns and similarities in sequences and sub-sequences. Events can also be locally grouped, such as using a K-means algorithm to group Rodger Federer's similar tennis shots [Dem13]. Passes can be represented as a network such as for the 2012 NFL championship [Gol13]. Cycling time trials have been represented using a matrix to visually compare time series [BBW16].

Trajectories. Trajectories result from the connection of discrete spatio-temporal events of the same entity. For instance, they might represent the spatio-temporal movement of players, balls, or other equipment during a game.

Trajectories are the second most frequent way of plotting raw tracking data and conveying the temporal and sequential structure of events [POJC01, MCHD12, SSN*16, JSS*14]. In their raw format, players' trajectories resemble random scribble. However,

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

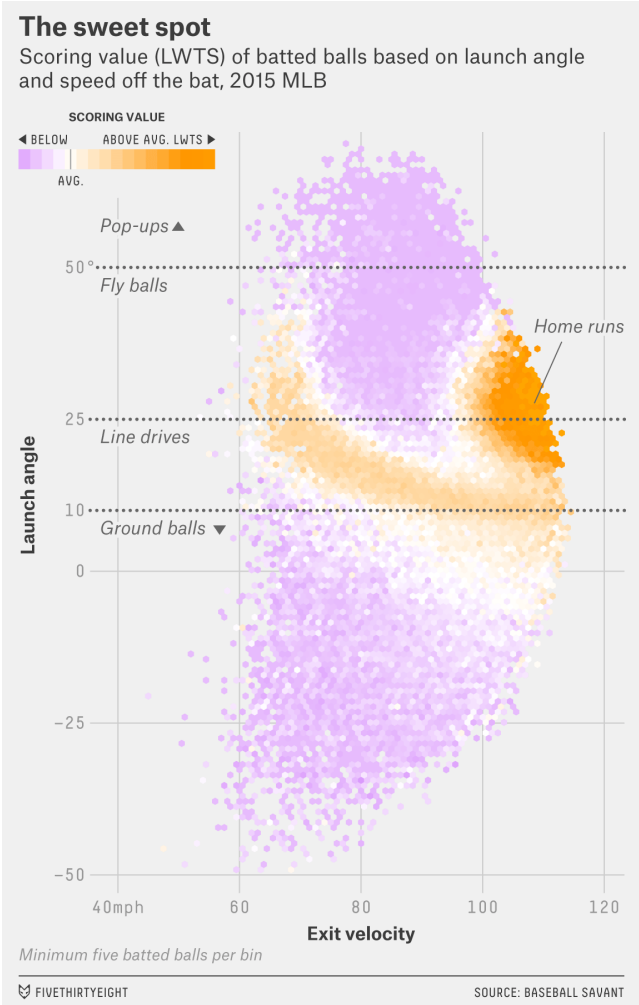

Figure 13: *A sweet spot appears when coloring ball hits by scoring value, which reveals the right trade off between ball velocity and angle. From Five Thirty Eight [Art16]. Image courtesy of the author, used with permission.*

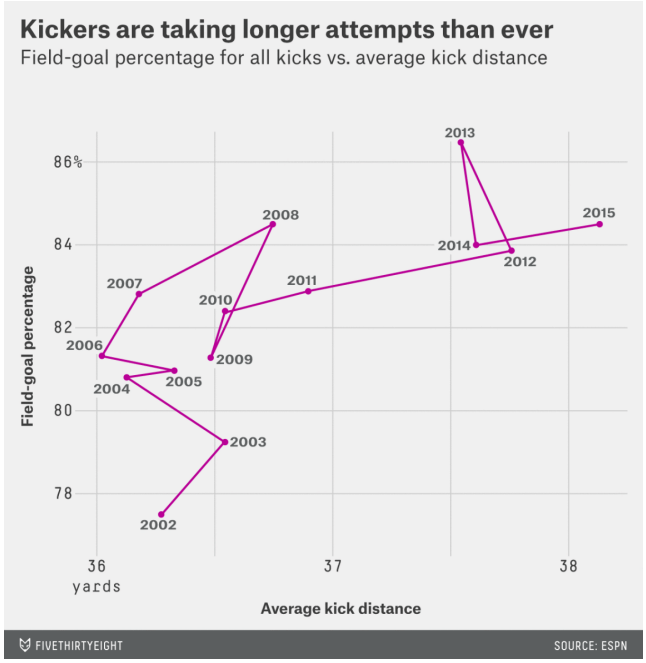

Figure 14: *A connected scatterplot shows trends over time. In this case, it emphasizes that kickers are taking longer attempts than ever. From Five Thirty Eight [Mor16a]. Image courtesy of the author, used with permission.*

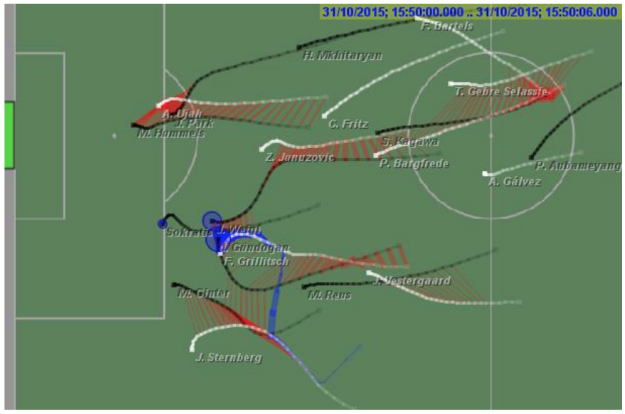

Figure 15: *Soccer pressure is illustrated by trails showing the pressure zone of the defense against the opponents [AAB*16]. Image courtesy of the authors, used with permission.*

some specific events are already recognizable, such as corner kicks in soccer [Sus06]. T-patterns [BJM02] aim at enhancing the presentation of spatial events by grouping them as zones. Interesting patterns are frequent connections of zones by the same player. Shorter trajectories (trails) also aim at reducing visual clutter, and are used in sports when entities are simultaneously moving such as during boat races [BC16], soccer games [AAB*16] (see Figure 15), or swimming races [CM16]. Trajectories are often used as visual signatures of phases of play [MLM*17, PVF13b]. These small multiples can represent segmented and abstracted sports events, such as football catches [Tim17], which then act as small visual summaries of particular game events.

Trails can convey movement in motion charts [MMZ17], and further be used as visual signatures of teams tactics [SJL*16]. Based on these trajectories, it is possible to compute and visualize new data such as players' "interaction spaces"—"duel areas" between players of two opposite teams [SJB*16] (see Figure 16) and pressure against adversaries [AAB*16, SJL*18, AAB*17]. It also makes it possible to classify the types of passes that soccer players make during a game [CEGH17].

A particular use of trajectories is for movement *deconstruction*, following a step-by-step analysis of complex sequences of actions such as those of a basketball rebound [MCHD12]. In a similar fashion, trails can emphasize a single movement. During the

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

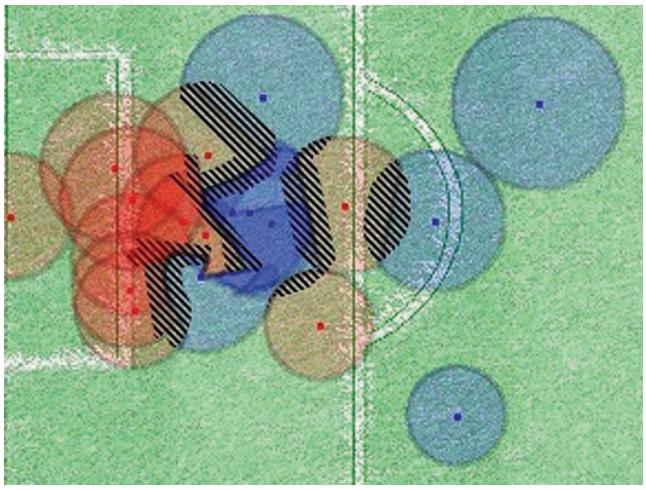

Figure 16: *Each circle represents a player. The radius of a circle represents the player's "interaction space", and the color of the circle encodes the team of the player. Hatching patterns show the potential "duel areas" of several players [SJB*16]. Image courtesy of the authors, used with permission.*

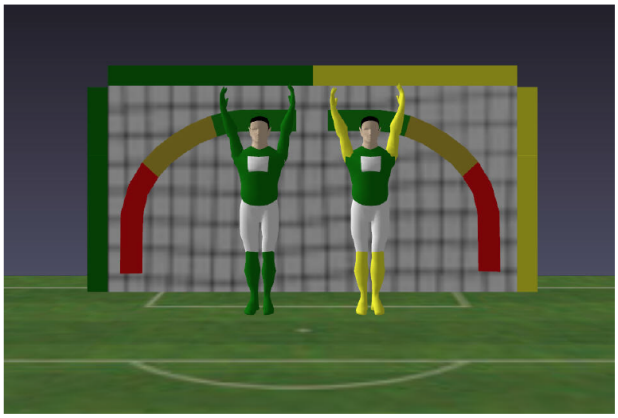

Figure 17: *A penalty shooter's point of view is reconstructed in 3D, showing her favorite target areas around the goalkeeper [RSB11]. Image courtesy of the authors, used with permission.*

2016 Olympics, Simone Biles's complex and fast-paced gymnastic movements [The16] were highlighted using a trails overlay atop the original video footage enabling visual persistence of multiple parts of her body during her routine.

Finally, one of the challenges with spatio-temporal trajectories within a bounded area such as a pitch is that representing an entire game or even just a subset of a game quickly results in overplotting, or 'spaghetti' representations. Tackling this issue is a hot topic in sports visualization, in particular, but also in visualizing spatio-temporal trajectories in general. For example, using different levels of spatio-temporal abstraction to simplify smoothing the trajectories [SAMS*17] (see Figure 18).

Detailed events with context. Besides space and time data, complex descriptors such as the ones listed in Figure 4 provide hundreds of ways a ball can be kicked or a player can behave with it. SoccerStories [PVF13b] (see Figure 19) represents complex multivariate records of each time the ball is kicked. While most visualizations apply the same representation to an entire dataset, their approach is to simplify such data using visual aggregation, where different visualizations are used piece-wise. Visualizations are organized on a spatial layout, and are centered where actions from the subsets spatially occur.

Real-time. As tracking technologies offer a real-time data feed, player position tracking can be broadcasted and analyzed live. Both Matchpad [LCP*12] and the exploration of the design space of glyphs in rubgy [CLP*15] summarize players' events in a compact display using glyphs to track and categorize plays. Both approaches provide a more effective sequence of events than the more standard textual play-by-play sequences of game events. In particular, it requires less attention when monitoring and provides compact game summaries. Major League Baseball Advanced Media's Gameday technology [MLB17] allows fans to see the trajectory and type of each pitch in real-time, and recent work augments the video stream of a soccer game with visualizations to convey the dynamics of the game [SJL*18].

5.2. Players' Point of View

First-person. While trajectories offer a bird's eye view on players, changing the point of view to the players' one allows viewers to understand the difficulty of a shot or a hit. A first-person point of view—such as a soccer goalkeeper's [RSB11], a baseball hitter [RCW10], or a baseball catcher [MLB17]—changes the camera and perspective on the data to orient the viewer to the way players experience the event. Events appear either as animated playback to re-create an actual play or as static heat maps with shot distributions such as tennis ball trajectories [ySS13]. First-person points of view usually include 3D objects such as a goalkeeper [RSB11] or a batter's silhouette and stadium features [MLB17] to suggest a more realistic environment.

Embedded tracking and visualization. Short feedback loops are important during any learning process. Several efforts have been made to provide those to players during training routines. Kaplan et al. display pedal pressure on training bikes [KYY*16] to optimize effort and prevent injuries. Treadmill screens can also provide real-time feedback to runners [CMHD10]. The use of small, portable devices such as wrist-bands to detect activity in basketball [BEA16] also provides individual feedback. Equipment can also be used as a ubiquitous mechanism to broadcast game-related data, such as heart rate data on a bicycle helmet [WWM13]. Most of these visualizations display data as numbers using no encoding. Besides performance indicators, traditional events such as scores on a display on a basketball hoop [CM13] or win/loss on jerseys [PM07] have been shown to provide better competition and engagement among players. GPS systems in racing cars [Spo16b] now collect speed, acceleration, time behind leader, amongst other

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

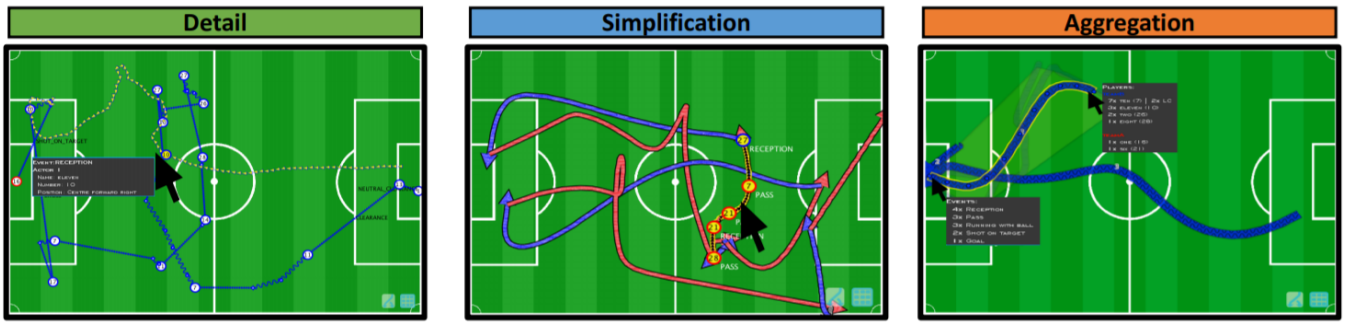

Figure 18: *Representing spatio-temporal trajectories of soccer players in detail, after simplification and aggregation [SAMS* 17]. Image courtesy of the authors, used with permission.*

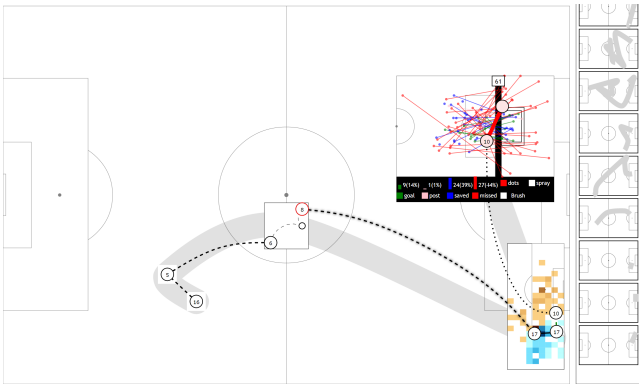

Figure 19: *SoccerStories [PVF13b] is the spatial connection of charts for the visual decomposition of a soccer phase. Image courtesy of the authors, used with permission.*

data that together provide monitoring data for mechanical assistants to anticipate repairs and prevent accidents.

6. Visualization by Type: Meta-Data

Sports *meta-data* provides content and context details for the two sports data types, box-score and tracking data. Meta-data can be *closely related* to the sport event. This includes the physical characteristics of players or the dimensions of the grounds. Meta-data can also be *loosely related* to the main sport activity. This includes Tweets or TV programs, which provide a proxy for sport activities. Sports meta-data does not have a particular data type, but has to be related to other data, which could include player identities, time, or location.

Meta-data goes beyond just providing context: it is often the object of research itself as it provides interesting data structures (e. g., trees, graphs) that require new visualizations and interactions. Meta-data can also be seen as an empty shell, similar to an empty box score paper, that is filled with observed or predicted data.

22 / 98 articles we collected visualize meta-data, making this type of sports data the least represented among of the three. However, because meta-data is also very diverse, visualizations of meta-data vary a lot. We structure our discussion of sports meta-data along three categories. First we discuss the visualization of *sports, players and games characteristics* such as physical landmarks on a pitch; then the visualization of *players and teams performance* such as team tactics; and then the visualization of *competition structure* and the thoroughly studied visualization of tournament brackets, in particular.

6.1. Sports, Players, and Game Characteristics

Meta-data plays an important role to convey a sport's identity and rules. Such information is needed to understand the importance of some events. For example, during a soccer game a foul in the penalty area has the risk of resulting in a penalty shot for the other team. Thus, such visual cues such as the boundaries of a field that players experience in reality are often replicated within the visualizations using the same codes.

Sport recognition. As we previously discussed, event data alone is sometimes enough to distinguish the sport itself. For example, this is the case for shots around a basketball hoop [Gol12]. The background pitch with recognizable landmarks is probably the most common way to tell which sport has been picked if this cannot simply be inferred. Additional landmarks can be added, such as those specific to sport events like a trophy showing for the UEFA Champion's League [Won13] or the FIFA World Cup [VP16].

Players and Teams Differentiation. Differentiating teams and players is crucial. In sports with two opponents, a differential color scheme can be used to differentiate entities. When multiple teams or players are involved, teams' semantic colors [Won13, PVF14] are used. These semantic colors originate from commonly associated colors (like the blues of Chelsea in soccer) or as representative colors from the badges that are teams' symbols [SM14]. Meta-data can also be used to critically discuss a sport in the context of more general topics. For example, showing the nationality of athletes for major sports league in North America and Europe over time shows how the sports have evolved throughout history [Ais17].

Game highlights. Sports events rarely are distributed uniformly

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

over time, and not all events are of equal interest, or value. An important body of research, especially in computer vision, uses video footage to detect potentially interesting sequences or errors. These algorithms are based on action detection [HBC09]—which is classic tracking data—but also using other individuals than players such as the referee of a game [CVS16]. Twitter can also be a proxy to quantify excitement during soccer games [Won13]. Peaks on social media usually reflect important events such as shots or missed shots. Using the volume of tweets over time has been used to plot sentiment changes in Tweets [HHEM*16] or to create physical data sculptures to convey the popular response to Olympic Games in The Emoto exhibit [DMS12].

6.2. Players' and Teams' Performance

Meta-data may augment sports data with details that are not captured by collected data and that can be used to create performance indicators.

Player performances. A player's physical abilities can be a predictor of performance for individual sports. However, team sports require additional skills such as synchronization, communication, and anticipation. The *plus-minus* indicator quantifies the positive or negative intervention on a field by visualizing the change of score when a player is in the game or not [Jos14]. Such indicator is helpful but requires alignment with other data sources such as game recaps to identify if the player is actually involved or not in the score change. A team's performance versus adversaries can be represented as temporal glyphs in a soccer matrix [CF13], which is a permutation of all possible games showing both results and progress of a championship.

It is sometimes necessary to add an additional entity that was not directly involved in a game as baseline for comparison. For example, using an ordinary, non-professional human [NFL15] or animal for speed scale. While this scale anchoring mechanism is recommended to grasp magnitudes variations and complex units [CVG13], especially by using physical objects from a similar domain, in sport it can be confusing to scale one sport using elements from another. For example, basketball landmarks have been used for scale to compare men's long jump medalists [QR16b]. This results in a confusing visualization as it makes it look like basketball shots are plotted and compared when this is not the case. Another type of baseline is referring to world-class sportsmen and sportswomen such as Stephen Curry [AQ16], Simone Biles [The16], Mariano Rivera [RCW10], or Tony Parker [Gol14] as ceilings or outliers for performance indicator values.

Team tactics and affinities. Each team has its own fashion of play. The most famous case is the Brazilian men's national soccer team with high-skilled players who, without playing a particular system, have a unique and enjoyable play signature. Many efforts have been devoted to identifying team styles [BLC*14] and spatial structures [FMBG15] using non-play events such as coach decisions, changes during games, and inter-individual patterns such as positioning or marking. Networks of passes [DWA10] also provide indicators on the team's structure while also showing success probability. Affinities between players is an important

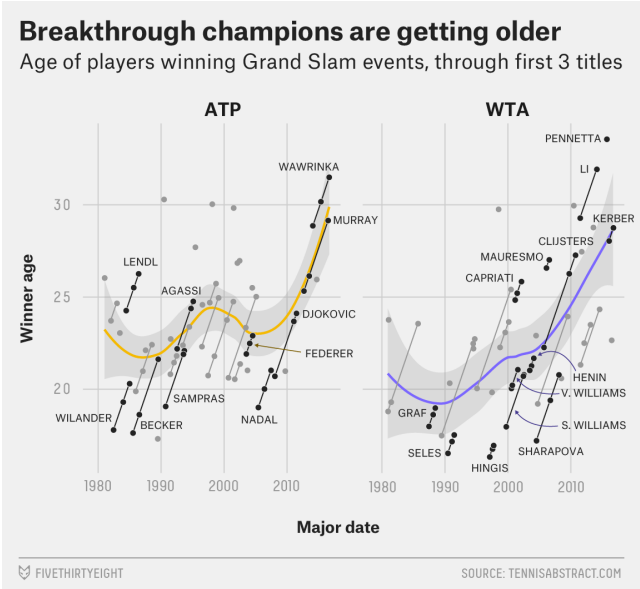

Figure 20: *Breakthrough champions are getting older. From Five Thirty Eight [Mor16b]. Image courtesy of the authors, used with permission.*

characteristic inside teams. However, these affinities may not only stem from the team they are evolving, for example when players evolve with their national team for international competitions but with their club for the majority of their playing time [Ais14]. Further connections between teams' popularity and spectators can be deduced from which games are televised in prime time in which locations [Gol15].

Expected performance. Meta-data include predictors of someone's ability to perform well at a given sport. Several visualizations show how NFL teams perform in terms of draft picking. Some compare the position in the draft (the earlier picked, the most successful the player is expected to be) and actual (subjective) career success estimation [MLBG16]. Others provide means of discussing successes and failures using storytelling strategies [Gol17b].

Players' age is a powerful predictor of within what range players will display their full abilities. Figure 20 shows the age of tennis tournament winners in relation with expected performances for a given age [Mor16b]. Salary is also an indicator of the magnitude of expectation for a given team at a given time, as one might expect the market value of a player or team to be proportional to their or its current or potential performance [Fry05]. Surprisingly, although betting is an important activity around sports, we did not find any sports data visualization making use of betting picks to visually track a player or a team outcome.

6.3. Competition Structure

Tournament trees describe competitions and are the second-most widely used competition structure behind rankings. A tournament

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

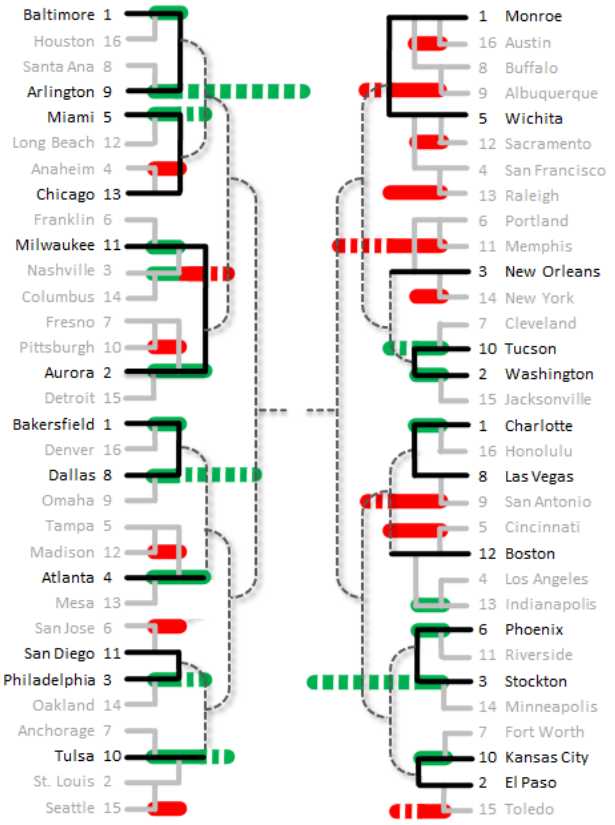

Figure 21: *AdaptiviTree [TSLR07] for visualizing prediction correctness on top of a tournament bracket using colored bars. Image courtesy of the authors, used with permission.*

(not a tournament tree) may start with a first phase, which is a mini-championship where small numbers of teams play against each other in a preliminary phase. In a tournament tree, or knock-out phase, two teams face each other at each stage of the tournament and only the winner progresses to the next stage.

Typical examples of tournament trees are tennis tournaments and NBA playoffs. Note that while in tennis tournaments, players play only one game to determine who progresses to the next stage, other formats exist. For example, in the NBA playoffs, teams face each other several times until one of the two teams wins a pre-determined number of games. In the European Champions' League, teams face twice—one home game and one away game for each team—and a set of rules determines which team progresses to the next stage.

Visualizing the tree structure of tournaments. Tournament trees have historically been represented graphically in their simplest form and printed on paper. While the structure is not space-efficient, it makes it possible to show both the individual games and the different stages of the tournament in one visualization. This enables tracking the path of a specific player or team.

Wongsuphasawat [Won13] designed a tree visualization of a soccer tournament based on Twitter data (see Figure 22). The New York Times infographics "Mapping Every Path to the NFL Playoffs" [QW14] uses trees to show all the possible paths of NFL teams that were in the race to the playoffs in 2014. This infographics was used to explain the many ways a team can reach the playoffs such as teams that get a week of rest and teams that host the first of their playoff game.

Supporting tournament predictions. Tournament trees have been used for supporting predictions. With iSee [STL09] and AdaptiviTree [TSLR07], researchers have investigated space-efficient tree visualizations for tournament-style brackets. Figure 21 shows an AdaptiviTree of teams participating in the NCAA men's basketball tournament. This visualization was designed to convey additional information on top of the tournament bracket. Specifically, in the context of online pick'em games, people predict the outcomes of games. AdaptiviTree shows the correctness of the predictions with colored bars on top of a tournament bracket.

Five Thirty Eight published an interactive visualization of tournament brackets and forecast [Sil16] that shows the round-by-round probabilities of teams to win the 2016 NCAA men's basketball tournament (see Figure 23).

Making tournament predictions has also been investigated from an input perspective. Vuillemot and Perin [VP16] explored the direct manipulation of tournament brackets to support people making their predictions using a visualization interface (see Figure 24). Instead of specifying with text or standard widgets which team progresses to the next stage of the tournament, users directly drag and drop teams to any stage of the tournament bracket. The visualization also shows all the possible paths for each team in the bracket.

7. Discussion and Perspectives

In addition to coding for the sports and data types, we identified the types of contribution and the evaluation methods of research papers. This provided a starting point for identifying emerging trends. Further, we discuss points that are raised by our review of the literature through the dual lens of sports data and research contributions and evaluations. We go on to emphasize challenges of sports data visualization, limitations in existing research, and opportunities for future work for sports data visualization in particular and for visualization in general.

7.1. Contributions and Evaluations

The research *contribution* columns of Figure 2 lists the types of research contributions we found in the research papers, based on the IEEE Infovis 2017 paper types [IEE17]. The *evaluation* columns of Figure 2 lists all the evaluation methods we found when reviewing the research literature. Note that the contribution and evaluation categories are not mutually exclusive, as a research paper in sports data visualization usually makes several contributions and can contain several evaluation methods.

Maybe the most valuable aspect of sports data for the research

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

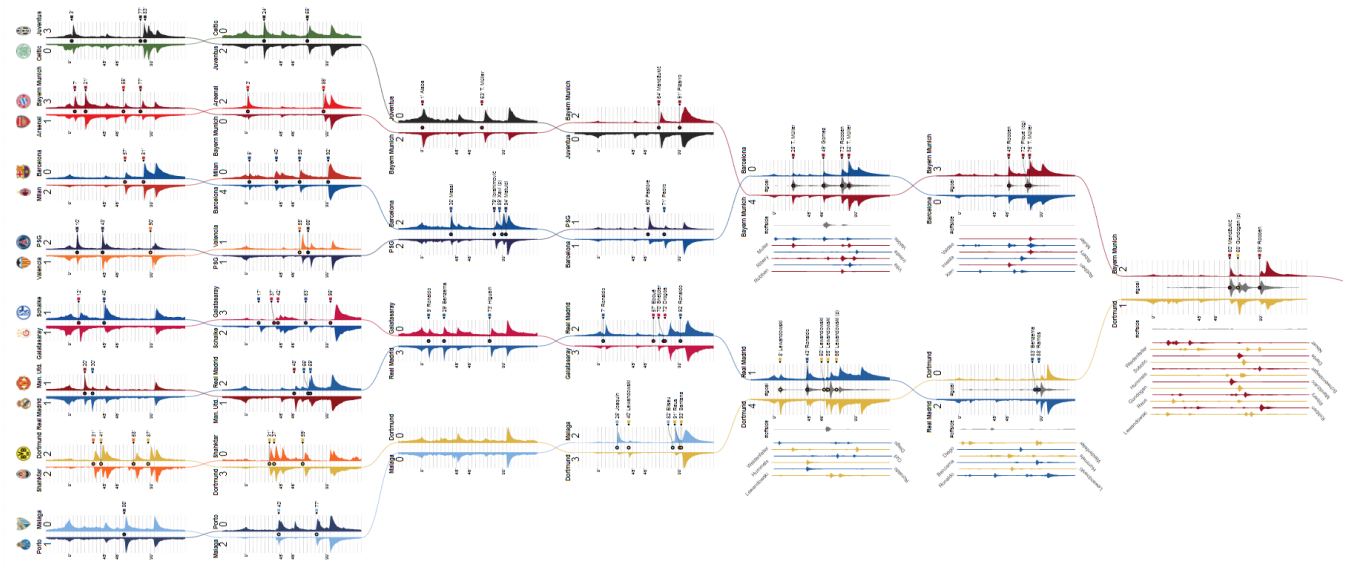

Figure 22: *All the games of the UEFA Champion's League visualized using the tree structure of the tournament. Individual games encode Twitter data to convey the excitement at each time [Won13]. Image courtesy of the author, used with permission.*

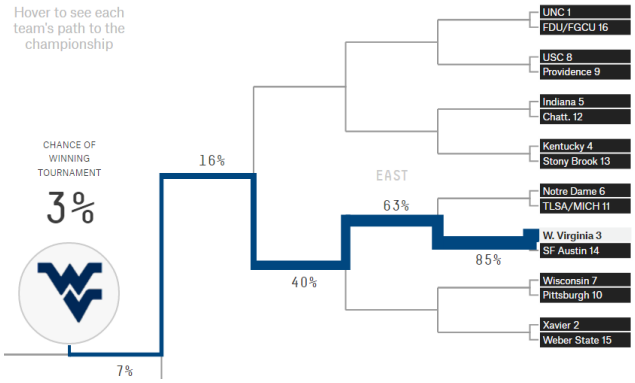

Figure 23: *2016 March Madness Predictions [Sil16] showing the probability of teams to win the 2016 NCAA men's basketball tournament. Probabilities are represented using the stroke width of the path that connects one team to the trophy. Image courtesy of the authors, used with permission.*

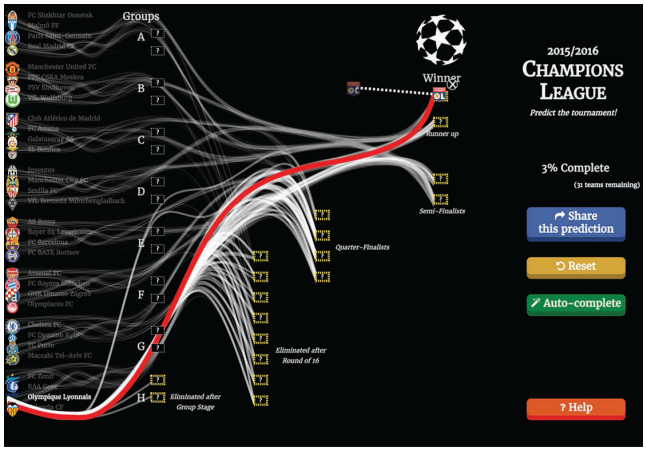

Figure 24: *Tournament brackets are meta-data that can be used to support an intuitive prediction-making process, by dragging and dropping teams according to their expected performance in the competition [VP16]. Image courtesy of the authors, used with permission.*

community is its wide variety. Data varies across sports, may be incomplete and uncertain, and can come in few values or consist of billions of data points. Tasks that sports experts seek to perform are not straightforward and often require tailored designs that do not generalize to other sports or even other contexts within a same sport. As a result, with few exceptions (e.g., [PBV16, TSLR07, PYHZ14]), most of the research papers we surveyed do not contribute a new generic technique but instead a tailored adaptation of an existing one. The most frequent visualization technique extensions we found were of node-link graphs (e.g., [PVF13b]), heatmaps (e.g., [PSBS12, Gol12, Bes16, AAB*16, MB13]), small multiples and glyphs (e.g., [HZ16, LCP*12, CPG*16, CF13, PYHZ14], treemaps (e.g., [JB97, CS06, Tur94], arc diagrams (e.g., [OJK13]), parallel coordinates (e.g., [OJK13, CPG*16], and line charts (e.g., [Woo15, PBV16, Won16]). For the same reasons, sports data visualization is also propitious to the design of new **interaction techniques**. There has, for example, been work on sketch and pattern recognition for querying spatio-temporal trajectories [SSN*16, MWA*01] and interaction

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

techniques were designed specifically for performing sports-specific tasks (e.g., [PBV16, CS06, PVF14, VP15, VP16]).

The process of designing visualization and interaction techniques for sports data requires a deep understanding of sports. However, sports has the useful property that *everyone can be an expert*, and academics conducting research in the area are usually passionate and knowledgeable about sports. This explains why we found several research papers that did not include any collaboration with sports experts. Instead, researchers used their own expertise to state required tasks or build scenarios to demonstrate the soundness and usefulness of their approaches (e.g., [Gol12, JB97, HZ16, Bes16, CLX*16]). Figure 2 highlights the variety of evaluation methods that can be employed in sports data visualization. These range from collaborations with sports actors to conduct case studies, to quantitative in-lab studies, to ad-hoc scenarios that illustrate the value of a technique or algorithm.

In our review, the most common method for evaluating visualization and interaction techniques is by conducting **case studies** that involve sports collaborators (e.g., [PVF13b, PYHZ14, PSBS12, PM07, SJB*16, WLS*18, SAMS*17]). This is usually made possible when the research project involved sports collaborators from the beginning. This type of process results in a **design study** [SMM12] where typically sports professionals and visualization researchers work together to understand a problem and propose a possible visualization solution. Exploring the problem space often involves a **tasks and requirements analysis** to better understand what the coaches, supporters, journalists, or analysts seek to achieve. Design studies and tasks and requirements analyses are an important part of sports visualization research. These contributions bring new knowledge to the field of visualization as they provide a deep understanding of a domain. This new knowledge then serves as a basis for the visualization community to explore further research directions.

Though sometimes directly applying an existing visualization technique or a slight variation of it is enough to address some needs, the solution space can be much bigger. In this case, **exploring the design space** of the particular problem or solution is a generative approach that expands the knowledge in the field (e.g., [CLP*15]). This is particularly true when it comes to less explored areas of visualization research. For example, when designing visualizations that **integrate into the context of use**, researchers have explored design spaces in the context of situated sports data collection and visualization (e.g., [PM07, CM13, KYY*16, WWM13]) and of the creation of physical artifacts representing sports data (e.g., [STS*14, LCN15]). Design space explorations can involve a wide variety of evaluation methods, including field studies [VP15, STS*14, LCN15, DMS12], scenarios [Gol12, MWA*01], case studies [PSBS12, VP15, CM13], and quantitative studies [PM07, VP16].

Visualization techniques are the most frequent contributions, yet there are other, equally important, facets to sports data visualization. These include **algorithms**, which can help with gathering new tracking data (e.g. [BEA16]), computing new meta-data (e.g., [FMBG15, DWA10]), and more generally compute new derived data by aggregating existing data (e.g., [AAB*16, BJM02, MCHD12]). Algorithms can also help model sports, usually by leveraging expert knowledge (e.g., [CPG*16, SHJ*15]). We found that the most common way of evaluating algorithms is through scenarios based on real data but not involving sports actors (e.g., [Dem13, BJM02, MCHD12, DWA10, BLC*14]). We also found a handful of papers that propose a **taxonomy or model** of a particular aspect of sports (e.g., [JSS*14, SHJ*15, PM06, SJS*17]). These papers are less tailored to a specific type of data or sport and tend to be more general.

7.2. The Role of Visualization in Sports

The two widely recognized uses of visualization are *analytical* (or exploratory) and *narrative* (or communicative). When visualization is used for analytical purposes, people typically seek to gain a better understanding of the data and its characteristics, learn about the phenomena modeled by the data, and gain insights about events captured through the data. Ultimately, knowledge gained from the data may be put to use for future decision-making and modifying performance. Narrative uses of visualization revolve around presenting data in a manner that is more easily understood and communicated. An ideal narrative visualization quickly allows the viewer to interpret the different cases and variables within a data set to see how they relate and to comprehend potential bigger picture messages. We find that visualizations of sports data are used widely for both these high-level purposes, though perhaps more commonly for communication and presentation.

Sports data visualizations used for analytical purposes focus on understanding an individual's or a team's performance. If the consumer of a visualization is a third-party analyst, then the investigator may be seeking to learn why an individual is successful in their sport or why a team has been losing so frequently. In some cases, the analyst may have specific hypotheses about performance and it may be possible to check them via statistical methods or data mining. However, the analyst may not harbor any initial suspicions about the data and may simply be seeking to explore it to discover unexpected insights. If the consumer of a visualization is an athlete or performer themself, then it is likely that the person is seeking to improve their performance and ultimately win more.

For example, consider visualizations of teams' shot patterns in basketball [Gol12]. Visualizations can demonstrate teams' tendencies and where the team is successful in shooting. Similarly, in American football, in any particular situation a team has hundreds of different plays to choose from and call. Very likely, teams follow patterns and have "favorite" plays to call in specific situations of field position and game score. Visualizations could communicate tendencies and patterns in these play selections. In tennis, what type of shot does a player typically hit given different styles of opponent shots? In soccer, what style of build-up and attack is more likely to lead to fruitful offensive possessions and, ultimately, goals? Where on the field do these attacks originate and flow and which players are involved [PVF13b, AAB*16]?

Visualizations used in these ways are potentially beneficial in scouting situations. A sports actor may be seeking to understand an opponent's tendencies and behaviors in order to develop their own strategies for competing against those tendencies. Alternately,

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

a performer may be self-scouting, that is, learning about their own behaviors and performance to optimize chances of success, move away from actions leading to failures, or to counter developing patterns that may be identified by opponents and used against them.

Sports data visualizations used for narrative purposes leverage the tremendous popularity of sports today. Billions of people worldwide likely have at least a passing interest in some sport, be it as a participant or spectator. Fans want to know why particular individuals and teams are successful, or, even more simply, they may seek to stay abreast of standings and performance. Statistics have played a huge role for this purpose in the past, but today's information consumer is more sophisticated. Effective visualizations can communicate nuances of performance data that cannot be conveyed easily by statistics.

One common narrative sports visualization is the presentation of standings. Depicting teams' positions in soccer leagues or tables [PBV16, VP16, Won16], bike racers positions at all the moments of a race [BW14, Woo15], or basketball teams' progress through a bracket [Sil16] all have been explored using different visualization techniques. Visualization seems particularly well-suited to communicating geographic-relevant aspects of sports data that are so common. Two- and three-dimensional diagrams correspond to the playing fields, pitches, courts, tracks, and courses that are pervasive in sports. Such visualizations are easy for consumers to interpret and understand, making the translation from data to representation especially meaningful.

7.3. Combining Data Types

In performing this first, detailed review of current work in sports data visualization, we arrived at a classification via the data type itself. This stresses the fact that most work considers only one type of data, with only a few exceptions including more than one type of data. This is true for academic work as much as for practitioner work.

One major difference that we observed between practitioner and academic work is that while practitioners make use of box-score, tracking, and meta-data in comparable proportions, academics focus much more on tracking data (see Figure 1).

A possible explanation is that research is driven by newly available sources of data. Another explanation could be that box-score data and meta-data are perceived as being more 'solved' issues than tracking data. There is, however, evidence from the works we reviewed that plenty of research on task analyses as well new visualization and interaction techniques has yet to conducted for both box-score and meta-data. Further, work combining all three data types is nearly non-existent, making this a promising avenue for the future of sports visualization.

Of particular note is the absence of box-score and tracking data being combined in a single visualization in our collection. There is, for instance, no work on time evolving detailed events (such as shots attempts in soccer) or how particular detailed events influenced the outcome of a game and even the winner of a tournament. Changes in rules, personal, and/or league-membership can also make it difficult to compare information at this intersection. The velocity of the data and the many possibilities result in keeping this area under-explored. Again, this is not specific to sports (e. g., in economics countries can split, merge, or disappear) but it does arise in this context.

7.4. Data Quality and Data Volume

A common theme in sports visualization research is that sports data visualization suffers from many of the same challenges as any other domain-focused visualization research. One of the ways that this manifests itself is in the quality of the data available. Box-score data is the gold standard as they officially track wins, losses, and records in both professional and recreational leagues, and therefore the data can be relied upon to be relatively accurate. Similarly, sensor-derived tracking data can be considered accurate within the limits the manufacturer specifies. In some cases this tracking data is now used to derive box-score data. There are still many challenges related to the accuracy of box-score data that still involves subjective interpretation and to the accuracy of tracking data that is collected from multiple, heterogeneous sources that is sometimes difficult to combine.

One challenge beyond accuracy is in the wealth of data beyond these two categories. We mentioned that tracking data produces more accurate and higher-volume data at a lower cost than box-score data. This is true for sensor-based tracking data (the majority of tracking data) in the long run; however, these sensor systems have a large upfront cost in purchasing and installing the systems. Therefore, high-quality tracking data usually only exists for the highest levels of professional sports. Recreational leagues and lower-level professional leagues can buy access to the data on the high-level leagues, but often do not have access to this same data for their own athletes [LM16]. For example, as Pileggi et al. [PSBS12] discuss, not all tracking data is sensor-derived. They dealt with a tracking dataset that was derived from human inspection, rather than from camera-based sensors. Thus, rather than the inches or centimeter accuracy provided by camera-based sensors, accuracy was at the foot granularity. Similarly, accurate meta-data, such as stadium capacities or player weights, is sometimes harder to acquire or will be at a coarser granularity than desired.

If and when tracking data is collected, there are still the challenges that come from visualizing large volumes of data. The work by the annual LDAV Symposium has worked to address the challenges of visualizing massive data sets. These challenges include accessing the data efficiently in a database, the additional costs of preprocessing that are ignored with smaller data sets, and handling noise in the sensor readings. Sports visualization can be a useful testbed for these techniques as they are researched and developed.

7.5. Missing Benchmark Datasets

A standard approach to demonstrating a research community has reached a certain level of structure and gains in maturity are reference data sets for comparison. For example, the Iris or Cars data sets are often included with novel, general-purpose visualization tools. These otherwise simple data sets have obvious

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

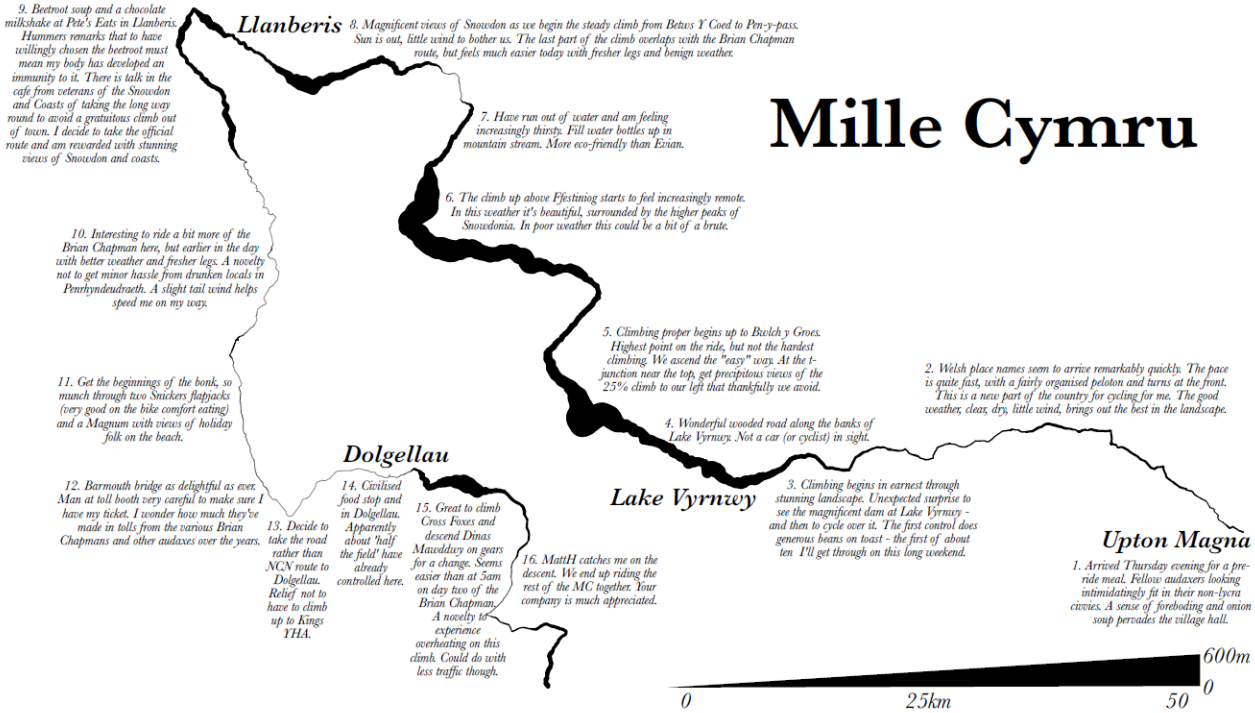

Figure 25: *Stylized and annotated cycling elevation map [Woo15]. Image courtesy of the author, used with permission.*

characteristics a novel visualization should be able to communicate and that experts know to look for.

With sports data, we have observed little use of similar datasets across projects, besides by the same authors. It seems there is a constant quest for most recent data. For instance, a ranking of 2012 soccer teams in the UK does not draw the same attention as a ranking of 2018 teams. By using the most recent data, viewers are more likely to contexualize the data with recent events and better understand the visualization, rather than depending on memory recall or additional resources. There is also a race for richer, more complex data, which make retro-comparison nearly impossible as both new attributes and higher detail usually cannot be acquired for older datasets.

One root cause of this is that "Sports" is a multi-billion dollar industry. While sports fans sometimes like to enjoy attending games as a hobby, the amount of money in play means that gaining a competitive advantage over rival teams or athletes can be significantly financially beneficial. Conversely, giving away a competitive advantage to a rival is financially detrimental. This has some notable benefits, such as a better "product on the pitch" for the fans and improved tracking technologies for teams. Unfortunately, this same competition discourages creation of open data sets or sharing of research findings, two cornerstones of academic research. We have seen this already with high costs for accessing relatively cheaply-obtained tracking data and limitations on sharing that data. One exciting possibility, however, is to explore new visualizations and algorithmic approaches to sports analytics using e-sports data [May]. E-sports also gather millions of enthusiasts and e-sports data shares the same characteristics as physical sports. The main difference is that e-sports data is more complete, more accurate, and easier to collect than sports data. This makes e-sports not only a domain with a bright future for visualization and visual analytics, but also a potential solution for bypassing the challenge of benchmark datasets in sports, at least temporarily.

7.6. Beyond Competition

An often forgotten facet of sports is their enjoyable nature. Sports are commonly practiced outside a competitive environment for physical and mental wellness.

In the context of cycling events, Wood [Woo15] created stylized and annotated elevation maps like the one shown in Figure 25. These maps not only display the data objectively, they also become artwork that reflect a personal experience. The annotations reinforce this connection between the cyclist's personal experience of the race and the raw tracking data. In a personal context, Wood highlights the need for both personal and synoptic visualization design. Sports records are beaten every year as technology and scientific advances progress; an individual performance is often either emphasized or de-emphasized by the teammates of the individual and the level of the opponents; geographic location plays a crucial role in major sports events (e.g., who is hosting the world cup will usually be among the favorites to win the title). Designing visualizations that are simultaneously relatable to individuals while still providing context around the individual—

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

typically other participants in a sport or competition—is an open challenge. In sports, it is particularly true that a performance does not mean anything without context.

The future is full of opportunities at the crossroads of sports data visualization, quantified-self, personal informatics, ubiquitous computing and personal data visualization. As people are more and more equipped with personal sensors, they record vast amounts of personal running, gym, and recreational sports activity data. There is an increasing amount of research building around personal fitness and activity tracking and visualization (e.g., [BPAC16, CMT*08, LML*06, STS*14, LCN15, KHM14, DMS12]). Challenges in personal sports data visualization are certainly related, but different, from those of sports visualization. Making visualizations personable and relatable artifacts in digital as well as physical forms that favor self-reflection are emerging topics in the visualization community (e.g., [LDF11, TBHC16, HTA*15, TPWC17]). Combining these topics of research offers many opportunities for future work.

7.7. Ethical Considerations of Sports Data

Having discussed personal/recreational data collection and visualization and the research opportunities it offers, we now end with the flip-side to this facet of sports data. Sports are one of the few private enterprises where so much information is public knowledge. Consider how most private businesses keep data on individual employee performance internal and even then usually restricted to certain individuals. An employee looking for a new job can usually expect that his or her potential employers cannot see the raw results of past performance reviews. Equivalently, consider the rigorous protections on your personally-identifiable health information mandated by most countries. Your doctor would never be allowed to publicize that you had an injury that might, in the future, affect your ability to do your job.

In sports, this is not the case. Individual performance is tracked and measured constantly. While enthusiasts approach box-score data as a way to mentally recreate games, athletes and teams use that same information to negotiate salaries and contract lengths. When using sports data, researchers must consider the risks and benefits to not only teams (representing management) but also to players (representing labor). Karkazis and Fishman [KF17] provide a thorough discussion of the ethics of tracking athletes.

8. Conclusions

We have provided a first, detailed review of current efforts in sports data visualization. Examining all this sports visualization work, from both academics and practitioners, has resulted in a classification of what is taking place in the current state of this domain, in particular the strong evidence that it all relies on three main data categories: box-score data, tracking data, and meta-data. These data types are distinct on several levels: the history from which they emerged, the reasons for which they have been collected, the purposes to which they are put, and people or audience who are interested in them. It is apparent that the work cleaves strongly along these lines—which both points to the strength of the classification and calls for an almost untouched exploration of the value of leveraging these data types in context of each other.

Furthermore, our classifications of sports visualizations by the type of research contribution they have made similarly provides actionable insights on how the research in information visualization can better be combined with sports data. For example, a clear call for more interaction research, space for taxonomies of sports visualizations, and for deploying field studies. The growing challenges and opportunities we have described for collecting, visualizing, and analyzing sports data, for both competitive and recreational purposes, are some of the many avenues for future work in sports visualization.

9. acknowledgements

We are grateful to all the authors who gave us permission to re-use their images in this publication. We also thank the organizers of the 1st Workshop on Sports Data Visualization [BCC*13]; the editors of the Sports Data Visualization special issue [BS16]; and all academics and practitioners who promote sports data visualization in their practice. This work has been partially supported by program "Investissements d'Avenir" from Université de Lyon program (ANR-11-IDEX-0007).

References

[AAB*16] ANDRIENKO G., ANDRIENKO N., BUDZIAK G., VON LANDESBERGER T., WEBER H.: Exploring Pressure in Football. 10, 15, 16

[AAB*17] ANDRIENKO G., ANDRIENKO N., BUDZIAK G., DYKES J., FUCHS G., VON LANDESBERGER T., WEBER H.: Visual analysis of pressure in football. *Data Mining and Knowledge Discovery 31*, 6 (2017), 1793–1839. 10

[Ais14] AISCH G.: The Clubs That Connect The World Cup. *The New York Times* (June 2014). 13

[Ais17] AISCH G.: Where Athletes in the Premier League, the N.B.A. and Other Sports Leagues Come From, in 15 Charts. *The New York Times* (Dec. 2017). 12

[Ama10] AMANDA COX: Fractions of a Second: An Olympic Musical - Interactive Graphic - NYTimes.com, 2010. 8

[AQ14] AISCH G., QUEALY K.: Why Peyton Manning's Record Will Be Hard to Beat. *The New York Times* (Oct. 2014). 1, 7

[AQ16] AISCH G., QUEALY K.: Stephen Curry's 3-Point Record in Context: Off the Charts. *The New York Times* (Apr. 2016). 1, 7, 13

[Are17] ARENAS I. D. L. T.: Swimming World Records throughout History, 2017. 7

[Art16] ARTHUR R.: The New Science Of Hitting, Apr. 2016. 9, 10

[BBW16] BECK F., BURCH M., WEISKOPF D.: A Matrix-Based Visual Comparison of Time Series Sports Data. 9

[BC16] BOSTOCK M., CARTER S.: The America's Cup Finale: Oracle's Path to Victory, 2016. 10

[BCC*13] BASOLE R., CLARKSON E., COX A., HEALEY C., STASKO J., STOLPER C.: *What's the score? The 1st Workshop on Sports Data Visualization*. Oct. 2013. http://workshop.sportvis.com/. 1, 19, 24

[BEA16] BAI L., EFSTRATIOU C., ANG C. S.: weSport: Utilising wrist-band sensing to detect player activities in basketball games. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)* (Mar. 2016), pp. 1–6. 11, 16

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

[Ber15] BERTIN M.: Why Soccer's Most Popular Advanced Stat Kind Of Sucks, 2015. 1

[Bes16] BESHAI P.: Buckets: Basketball Shot Visualization. 7, 9, 15, 16

[BJM02] BORRIE A., JONSSON G. K., MAGNUSSON M. S.: Temporal pattern analysis and its applicability in sport: an explanation and exemplar data. *Journal of Sports Sciences 20*, 10 (Jan. 2002), 845–852. 10, 16

[BLC*14] BIALKOWSKI A., LUCEY P., CARR P., YUE Y., SRIDHARAN S., MATTHEWS I.: Identifying Team Style in Soccer Using Formations Learned from Spatiotemporal Tracking Data. In *2014 IEEE International Conference on Data Mining Workshop* (Dec. 2014), pp. 9–14. 13, 16

[Bou47] BOUCHER, PIERRE: How to ski by the French method, 1947. 8

[BPAC16] BOTROS F., PERIN C., ASENIERO B. A., CARPENDALE S.: Go and Grow: Mapping Personal Data to a Living Plant. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2016), ACM, pp. 112–119. 19

[BS16] BASOLE R. C., SAUPE D.: Sports Data Visualization [Guest editors' introduction]. *IEEE Computer Graphics and Applications 36*, 5 (Sept. 2016), 24–26. 1, 19

[BW14] BEECHAM R., WOOD J.: Characterising group-cycling journeys using interactive graphics. *Transportation Research Part C: Emerging Technologies 47, Part 2* (2014), 194 – 206. Special Issue: Emerging Technologies Special Issue of {ICTIS}. 17, 24

[Cat18] CATAPULT SPORTS: World Leader in Sport Technology for Elite Sports, 2018. 5

[CEGH17] CHAWLA S., ESTEPHAN J., GUDMUNDSSON J., HORTON M.: Classification of Passes in Football Matches Using Spatiotemporal Data. *ACM Transactions on Spatial Algorithms and Systems (TSAS) 3*, 2 (2017), 6. 10

[CF13] CAVA R., FREITAS C. D. S.: Glyphs in matrix representation of graphs for displaying soccer games results. In *The 1st Workshop on Sports Data Visualization. IEEE* (2013). 13, 15

[CLP*15] CHUNG D. H., LEGG P. A., PARRY M. L., BOWN R., GRIFFITHS I. W., LARAMEE R. S., CHEN M.: Glyph sorting: Interactive visualization for multi-dimensional data. *Information Visualization 14*, 1 (2015), 76–90. 11, 16

[CLX*16] CHEN W., LAO T., XIA J., HUANG X., ZHU B., HU W., GUAN H.: GameFlow: Narrative Visualization of NBA Basketball Games. *IEEE Transactions on Multimedia 18*, 11 (Nov. 2016), 2247–2256. 9, 16

[CM13] CHATHAM A., MUELLER F. F.: Adding an Interactive Display to a Public Basketball Hoop Can Motivate Players and Foster Community. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (New York, NY, USA, 2013), UbiComp '13, ACM, pp. 667–676. 11, 16

[CM16] CROUSE K., MILLS D.: Katie Ledecky Smashes Her Own World Record in 400 Freestyle. *The New York Times* (Aug. 2016). 10

[CMHD10] CROWELL H. P., MILNER C. E., HAMILL J., DAVIS I. S.: Reducing impact loading during running with the use of real-time visual feedback. *journal of orthopaedic & sports physical therapy 40*, 4 (2010), 206–213. 11

[CMT*08] CONSOLVO S., MCDONALD D. W., TOSCOS T., CHEN M. Y., FROEHLICH J., HARRISON B., KLASNJA P., LAMARCA A., LEGRAND L., LIBBY R., OTHERS: Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008), ACM, pp. 1797–1806. 19

[CNW08] COX A., NGUYEN V., WARD J.: How Woods Wins a Major. *The New York Times* (May 2008). 8

[Con16] CONLEN M.: A History Of Sumo, May 2016. 7, 8

[CPG*16] CHUNG D. H. S., PARRY M. L., GRIFFITHS I. W., LARAMEE R. S., BOWN R., LEGG P. A., CHEN M.: Knowledge-Assisted Ranking: A Visual Analytic Application for Sports Event Data. *IEEE Computer Graphics and Applications 36*, 3 (May 2016), 72–82. 7, 15, 16

[CQW11] CARTER S., QUEALY K., WARD J.: Derek Jeter Joins an Exclusive Club, June 2011. 7

[CS06] COX A., STASKO J.: Sportsvis: Discovering meaning in sports statistics through information visualization. In *Compendium of Symposium on Information Visualization* (2006), Citeseer, pp. 114–115. 7, 15, 16, 24

[CVG13] CHEVALIER F., VUILLEMOT R., GALI G.: Using Concrete Scales: A Practical Framework for Effective Visual Depiction of Complex Measures. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2426–2435. 13

[CVS16] CARBOCH J., VEJVODOVA K., SUSS V.: Analysis of errors made by line umpires on ATP tournaments. *International Journal of Performance Analysis in Sport 16*, 1 (Apr. 2016), 264–275. 13

[CW06] COX A., WARD J.: Steroids or Not, the Pursuit is On. *The New York Times* (2006). 7

[Dem13] DEMAJ D.: Geovisualizing spatio-temporal patterns in tennis: An alternative approach to post-match analysis. In *Proceedings of the 26th International Cartographic Conference* (2013). 9, 16

[Dic97] DICKSON P.: *The Joy of Keeping Score: How Scoring the Game Has Influenced and Enhanced the History of Baseball*. Harcourt Brace, Feb. 1997. 1

[DKVS14] DIETRICH C., KOOP D., VO H. T., SILVA C. T.: Baseball4d: A tool for baseball game reconstruction amp; visualization. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct. 2014), pp. 23–32. 5, 9

[DMS12] DREW HEMMENT, MORITZ STEFANER, STUDIO NAND: Emoto - Installation * Studio NAND, 2012. 13, 16, 19

[DQ14] DANIEL JENNIFER S. C., QUEALY K.: How Close They Came to Winning. *The New York Times* (Feb. 2014). 1, 8

[DWA10] DUCH J., WAITZMAN J. S., AMARAL L. A. N.: Quantifying the Performance of Individual Players in a Team Activity. *PLOS ONE 5*, 6 (June 2010), e10937. 13, 16

[Fan16] FAN GRAPHS: Baseball Statistics and Analysis | FanGraphs Baseball, 2016. 1

[FMBG15] FRANKS A., MILLER A., BORNN L., GOLDSBERRY K.: Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics 9*, 1 (Mar. 2015), 94–121. arXiv: 1405.0231. 13, 16

[Fry05] FRY B.: Salary vs. Performance, 2005. 13

[GH17] GUDMUNDSSON J., HORTON M.: Spatio-temporal analysis of team sports. *ACM Computing Surveys (CSUR) 50*, 2 (2017), 22. 2, 8

[Gol12] GOLDSBERRY K.: Courtvision: New visual and spatial analytics for the nba. In *2012 MIT Sloan Sports Analytics Conference* (2012). 9, 12, 15, 16

[Gol13] GOLDSBERRY K.: Pass Atlas: A Map of Where NFL Quarterbacks Throw the Ball, Sept. 2013. 9

[Gol14] GOLDSBERRY K.: DataBall, 2014. 13

[Gol15] GOLDSBERRY K.: Which NFL Team Are You Stuck Watching Every Sunday?, Oct. 2015. 13

[Gol17a] GOLDENBERG R.: How Russell Westbrook can Average a Triple-Double, 2017. 7

[Gol17b] GOLDENBERG R.: Twenty Years of the NBA Redrafted, 2017. 13

[HBC09] HERVIEU A., BOUTHEMY P., CADRE J.-P. L.: Trajectory-based Handball Video Understanding. In *Proceedings of the ACM International Conference on Image and Video Retrieval* (New York, NY, USA, 2009), CIVR '09, ACM, pp. 43:1–43:8. 13

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

[HHEM*16] HOEBER O., HOEBER L., EL MESEERY M., ODOH K., GOPI R.: Visual Twitter Analytics (Vista): Temporally changing sentiment and the discovery of emergent themes within sport event tweets. *Online Information Review 40*, 1 (Feb. 2016), 25–41. 13

[HTA*15] HUANG D., TORY M., ASENIERO B. A., BARTRAM L., BATEMAN S., CARPENDALE S., TANG A., WOODBURY R.: Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics 21*, 3 (2015), 420–433. 19

[HZ16] HE X., ZHU Y.: TennisMatchViz: A Tennis Match Visualization System. *Electronic Imaging 2016*, 1 (Feb. 2016), 1–7. 15, 16

[IEE17] IEEE: IEEE InfoVis 2017 - Topics and Paper Types, 2017. 14

[JB97] JIN L., BANKS D. G.: TennisViewer: a browser for competition trees. *IEEE Computer Graphics and Applications 17*, 4 (July 1997), 63–65. 7, 15, 16

[Jos14] JOSE L.: Game on!, 2014. 13

[JSS*14] JANETZKO H., SACHA D., STEIN M., SCHRECK T., KEIM D. A., DEUSSEN O.: Feature-driven visual analytics of soccer data. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct. 2014), pp. 13–22. 1, 9, 16

[JSSS16] JANETZKO H., STEIN M., SACHA D., SCHRECK T.: Enhancing Parallel Coordinates: Statistical Visualizations for Analyzing Soccer Data. *Electronic Imaging 2016*, 1 (2016), 1–8. 2

[KF17] KARKAZIS K., FISHMAN J. R.: Tracking U.S. Professional Athletes: The Ethics of Biometric Technologies. *The American Journal of Bioethics 17*, 1 (Jan. 2017), 45–60. 19

[KHM14] KHOT R. A., HJORTH L., MUELLER F.: Understanding physical activity through 3d printed material artifacts. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (2014), ACM, pp. 3835–3844. 19

[KYY*16] KAPLAN O., YAMAMOTO G., YOSHITAKE Y., TAKETOMI T., SANDOR C., KATO H.: In-Situ Visualization of Pedaling Forces on Cycling Training Videos. 11, 16

[LCN15] LEE M.-H., CHA S., NAM T.-J.: Patina Engraver: Visualizing Activity Logs As Patina in Fashionable Trackers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY, USA, 2015), CHI '15, ACM, pp. 1173–1182. 16, 19

[LCP*12] LEGG P. A., CHUNG D. H. S., PARRY M. L., JONES M. W., LONG R., GRIFFITHS I. W., CHEN M.: MatchPad: Interactive Glyph-Based Visualization for Real-Time Sports Performance Analysis. *Computer Graphics Forum 31*, 3pt4 (June 2012), 1255–1264. 11, 15

[LDF11] LI I., DEY A. K., FORLIZZI J.: Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing* (2011), ACM, pp. 405–414. 19

[Lew04] LEWIS M.: *Moneyball: The art of winning an unfair game*. WW Norton & Company, 2004. 1

[LM16] LINDBERGH B., MILLER S.: *The Only Rule Is It Has to Work: Our Wild Experiment Building a New Kind of Baseball Team*, 1st edition ed. Henry Holt and Co., New York, May 2016. 17

[LML*06] LIN J. J., MAMYKINA L., LINDTNER S., DELAJOUX G., STRUB H. B.: Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *International Conference on Ubiquitous Computing* (2006), Springer, pp. 261–278. 19

[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. 9

[Lou15] LOU SPIRITO: The Differing Dimensions of Every Stadium in Baseball, Oct. 2015. 6

[Lov15] LOVE C.: Football Shots Timeline, 2015. 7, 8

[LSS13] LI F., STASKO J., STOLPER C. D.: NBAVis: Exploring 2012-13 NBA Season Data, 2013. 7

[LTB16] LOSADA A. G., THERON R., BENITO A.: BKViz: A Basketball Visual Analysis Tool. *IEEE Computer Graphics and Applications 36*, 6 (Nov. 2016), 58–68. 7

[May] MAYMIN P.: An Open-Sourced Optical Tracking and Advanced eSports Analytics Platform for League of Legends. In *2018 MIT Sloan Sports Analytics Conference*. 18

[MB13] MOON B., BRATH R.: Bloomberg sports visualization for pitch analysis. In *The 1st Workshop on Sports Data Visualization. IEEE* (2013). 9, 15

[MCHD12] MAHESWARAN R., CHANG Y.-H., HENEHAN A., DANESIS S.: Deconstructing the rebound with optical tracking data. In *Proc. 6th Annual MIT Sloan Sports Analytics Conference* (2012), pp. 1–7. 9, 10, 16

[MLB15] MLBPA/MLB NEWS RELEASE: MLBPA, MLB announce pace-of-game initiatives, replay modifications, 2015. 5

[MLB17] MLB: MLB.com, 2017. 11

[MLBG16] MEKO T., LU D., BERKOWITZ B., GAMIO L.: Legendary? Lousy? How your team has fared in the NFL draft, 2016. 13

[MLM*17] MACHADO V., LEITE R., MOURA F., CUNHA S., SADLO F., COMBA J. L.: Visual soccer match analysis using spatiotemporal positions of players. *Computers & Graphics 68* (2017), 84–95. 10

[MMZ17] METULINI R., MARISERA M., ZUCCOLOTTO P.: Space-Time Analysis of Movements in Basketball using Sensor Data. In *Statistics and Data Science: New Challenges, New Generations* (Apr. 2017). 10

[Mor16a] MORRIS B.: The Haters Are Losing The War On Kickers, Jan. 2016. 9, 10

[Mor16b] MORRIS B.: In Tennis, Even The Breakout Stars Are Veterans, Sept. 2016. 13

[MW12] MOSKOWITZ T., WERTHEIM L. J.: *Scorecasting: The hidden influences behind how sports are played and games are won*. Three Rivers Press, 2012. 1

[MWA*01] MOORE A., WHIGHAM P., ALDRIDGE C., HOLT A., HODGE K.: Rugby:(a) union of space and time. In *Proceedings of the 13th Annual Colloquium of the Spatial Information Research Centre* (2001), pp. 183–194. 15, 16

[NFL15] NFL: 'Run Rich Run': Rich Eisen vs. top prospects, 2015. 13

[OJK13] OWENS S. G., JANKUN-KELLY T. J.: Visualizations for exploration of american football season and play data. In *The 1st Workshop on Sports Data Visualization. IEEE* (2013). 1, 15

[PBV16] PERIN C., BOY J., VERNIER F.: Using Gap Charts to Visualize the Temporal Evolution of Ranks and Scores. *IEEE Computer Graphics and Applications 36*, 5 (Sept. 2016), 38–49. 6, 15, 16, 17, 24

[Pea17] PEARCE A.: LeBron James Scores 5,995th Playoff Point, Taking the Record From Michael Jordan. *The New York Times* (May 2017). 7

[PM06] PAGE M., MOERE A. V.: Towards Classifying Visualization in Team Sports. In *International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06)* (July 2006), pp. 24–29. 16

[PM07] PAGE M., MOERE A. V.: Evaluating a Wearable Display Jersey for Augmenting Team Sports Awareness. In *Pervasive Computing* (May 2007), LaMarca A., Langheinrich M., Truong K. N., (Eds.), Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 91–108. 11, 16

[POJC01] PINGALI G., OPALACH A., JEAN Y., CARLBOM I.: Visualization of Sports Using Motion Trajectories: Providing Insights into Performance, Style, and Strategy. In *Proceedings of the Conference on Visualization '01* (Washington, DC, USA, 2001), VIS '01, IEEE Computer Society, pp. 75–82. 9

[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. 9, 15, 16, 17, 24

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

[PV13] PERIN C., VERNIER F.: R2s2: A hybrid technique to visualize sport ranking evolution. In *What's the score? The 1st Workshop on Sports Data Visualization, IEEE VIS* (2013). 6, 24

[PVF13a] PERIN C., VUILLEMOT R., FEKETE J.-D.: Real-Time Crowdsourcing of Detailed Soccer Data. 5, 24

[PVF13b] 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. 10, 11, 12, 15, 16, 24

[PVF14] PERIN C., VUILLEMOT R., FEKETE J.-D.: A Table!: Improving Temporal Navigation in Soccer Ranking Tables. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY, USA, 2014), CHI '14, ACM, pp. 887–896. 6, 12, 16, 24

[PYHZ14] POLK T., YANG J., HU Y., ZHAO Y.: TenniVis: Visualization for Tennis Match Analysis. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 2339–2348. 15, 16

[QR16a] QUEALY KEVIN E. G., ROBERTS G.: Usain Bolt and the Fastest Men in the World Since 1896 - on the Same Track. *The New York Times* (Aug. 2016). 1, 8

[QR16b] QUEALY K., ROBERTS G.: Bob Beamon's Long Olympic Shadow, 2016. 13

[QW14] QUEALY KEVIN M. B., WARD J.: Mapping Every Path to the N.F.L. Playoffs. *The New York Times* (Dec. 2014). 1, 14

[RCW10] ROBERTS G., CARTER S., WARD J.: How Mariano Rivera Dominates Hitters, 2010. 11, 13

[RSB11] RUSU A., STOICA D., BURNS E.: Analyzing Soccer Goalkeeper Performance Using a Metaphor-Based Visualization. In *2011 15th International Conference on Information Visualisation* (July 2011), pp. 194–199. 11

[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. 11, 12, 16

[SHJ*15] STEIN M., HAUBLER J., JACKLE D., JANETZKO H., SCHRECK T., KEIM D. A.: Visual Soccer Analytics: Understanding the Characteristics of Collective Team Movement Based on Feature-Driven Analysis and Abstraction. *ISPRS International Journal of Geo-Information 4*, 4 (2015), 2159–2184. 16

[Sil16] SILVER REUBEN FISCHER-BAUM AND NATE J. B.: 2016 March Madness Predictions, Mar. 2016. 14, 15, 17

[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. 10, 11, 16

[SJL*16] STEIN M., JANETZKO H., LAMPRECHT A., SEEBACHER D., SCHRECK T., KEIM D. A., GROSSNIKLAUS M.: From Game Events to Team Tactics: Visual Analysis of Dangerous Situations in Multi-Match Data. In *International Conference on Technology and Innovation in Sports, Health and Wellbeing, Special Track "High level Sports in the XXI Century: contribution from Industry and University to the performance optimization"* (2016). 10

[SJL*18] STEIN M., JANETZKO H., LAMPRECHT A., BREITKREUTZ T., ZIMMERMANN P., GOLDLUCKE 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 24*, 1 (2018), 13–22. 10, 11

[SJS*17] STEIN M., JANETZKO H., SEEBACHER D., JAGER A., NAGEL M., HOLSCH J., KOSUB S., SCHRECK T., KEIM D. A., GROSSNIKLAUS M.: How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects. *Data 2*, 1 (Jan. 2017), 2. 16

[SM14] SETLUR V., MACKINLAY J. D.: Automatic Generation of Semantic Icon Encodings for Visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2014), CHI '14, ACM, pp. 541–550. 12

[SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (Dec. 2012), 2431–2440. 16

[Spo16a] SPORTS REFERENCE: Sports Reference | Sports Stats, fast, easy, and up-to-date, 2016. 1

[Spo16b] SPORTVISION: RACEF/X Technology, 2016. 5, 11

[SSN*16] SHAO L., SACHA D., NELDNER B., STEIN M., SCHRECK T.: Visual-Interactive Search for Soccer Trajectories to Identify Interesting Game Situations. *Electronic Imaging 2016*, 1 (Feb. 2016), 1–10. 9, 15

[STL09] SMITH G., TAN D., LEE B.: iSee: Interactive Scenario Explorer for Online Tournament Games. In *Proceedings of the International Conference on Advances in Computer Enterntainment Technology* (New York, NY, USA, 2009), ACE '09, ACM, pp. 190–197. 14

[STS*14] STUSAK S., TABARD A., SAUKA F., KHOT R. A., BUTZ A.: Activity Sculptures: Exploring the Impact of Physical Visualizations on Running Activity. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (Dec. 2014), 2201–2210. 16, 19

[Sus06] SUSKEN ROSENTHAL: Football Drawings Live, 2006. 10

[TBHC16] THUDT A., BAUR D., HURON S., CARPENDALE S.: Visual mementos: Reflecting memories with personal data. *IEEE transactions on visualization and computer graphics 22*, 1 (2016), 369–378. 19

[The16] THE NEW YORK TIMES: The Fine Line: What Makes Simone Biles the World's Best Gymnast. *The New York Times* (Aug. 2016). 11, 13

[Tim07] TIMES T. N. Y.: Paths to the Top of the Home Run Charts. *The New York Times* (July 2007). 1, 7

[Tim17] TIMES T. N. Y.: Odell Beckham's Hands Are Great, But His Feet Are Lethal. *The New York Times* (Jan. 2017). 10

[TKW*16] TROGER J., KLACK M., WENDLER D., PATZOLD A., MOLLER C., TIMCKE M.-L.: Berlin-Marathon 2016 - So schnell lauft Ihre Stadt, 2016. 7

[TPWC17] THUDT A., PERIN C., WILLETT W., CARPENDALE S.: Subjectivity in personal storytelling with visualization. *Information Design Journal 23*, 1 (July 2017), 48–64. 19

[TSLR07] TAN D., SMITH G., LEE B., ROBERTSON G.: AdaptiviTree: Adaptive Tree Visualization for Tournament-Style Brackets. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (Nov. 2007), 1113–1120. 14, 15

[Tuf01] TUFTE E. R.: *The Visual Display of Quantitative Information*, 2nd edition edition ed. Graphics Press USA, Cheshire, Conn, Jan. 2001. 8

[Tur94] TURO D.: Hierarchical Visualization with Treemaps: Making Sense of Pro Basketball Data. In *Conference Companion on Human Factors in Computing Systems* (New York, NY, USA, 1994), CHI '94, ACM, pp. 441–442. 15

[VP15] VUILLEMOT R., PERIN C.: Investigating the Direct Manipulation of Ranking Tables for Time Navigation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY, USA, 2015), CHI '15, ACM, pp. 2703–2706. 6, 16, 24

[VP16] VUILLEMOT R., PERIN C.: Sports Tournament Predictions Using Direct Manipulation. *IEEE Computer Graphics and Applications 36*, 5 (Sept. 2016), 62–71. 12, 14, 15, 16, 17, 24

[Win12] WINSTON W. L.: *Mathletics: How gamblers, managers, and sports enthusiasts use mathematics in baseball, basketball, and football*. Princeton University Press, 2012. 1

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

[WLS*18] WU Y., LAN J., SHU X., JI C., ZHAO K., WANG J., ZHANG H.: iTTVis: Interactive Visualization of Table Tennis Data. *IEEE transactions on visualization and computer graphics 24*, 1 (2018), 709–718. 16

[Won13] WONGSUPHASAWAT K.: A Narrative Display for Sports Tournament Recap. In *1st Workshop on Sports Data Visualization, IEEE VIS* (2013). 12, 13, 14, 15

[Won16] WONGSUPHASAWAT K.: Standings Tracer, 2016. 6, 7, 15, 17

[Woo15] WOOD J.: Visualizing Personal Progress in Participatory Sports Cycling Events. *IEEE Computer Graphics and Applications 35*, 4 (July 2015), 73–81. 8, 15, 17, 18, 24

[WWM13] WALMINK W., WILDE D., MUELLER F. F.: Displaying Heart Rate Data on a Bicycle Helmet to Support Social Exertion Experiences. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction* (New York, NY, USA, 2013), TEI '14, ACM, pp. 97–104. 11, 16

[ySS13] Y SANTACRUZ STUDIO E.: Corona Perspectives Tennis, 2013. 11

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.

Biographies of Authors

Charles Perin is a Lecturer (eq. Assistant Professor) in the department of Computer Science at City, University of London, United Kingdom and part of the giCentre. His research interests include the design and study of new interactions for information visualization, understanding how people use and interact with visualizations in their everyday lives, and human-centered design for specific applications such as sports and health. Perin has a PhD in computer science from Inria and CNRS-LIMSI at the University of Paris-Sud. Perin's strong personal interest in sports has led him to conduct cutting-edge research in sports visualization published at IEEE VIS, ACM CHI and IEEE CG&A. This includes publishing a design study in collaboration with soccer experts [PVF13b], improving the reading of ranks and scores in ranking charts [PBV16, PV13], and using sports as an application domain for designing new interaction techniques for sports enthusiasts [PVF14, VP15, VP16]. Perin participated in the 1st Workshop on Sports Data Visualization [BCC*13] at VIS 2013 and in the Data Technologies Workshop hosted by the U.S. Soccer Federation in 2015.

Romain Vuillemot is an Assistant Professor of computer science at École Centrale de Lyon, Université de Lyon, France. His research spans over the fields of information visualization, visual analytics, and human-computer interaction, with a focus on advanced interaction techniques to include humans in the loop of complex predictive analytics processes. Vuillemot has a PhD in computer science from INSA de Lyon and CNRS-LIRIS at the Université de Lyon. Vuillemot's connection to sport data lies in his interest in spatial and temporal data visualization, which naturally led him to design and implement novel techniques for sport data collection [PVF13a] and communication [PVF13b]. He recently focused on using direct manipulation for soccer ranking tables navigation [VP15], exploration [PVF14] and prediction [VP16].

Charles D. (Chad) Stolper is an Assistant Professor of Computer Science at Southwestern University in Georgetown, TX, USA. He received his PhD in Computer Science from the Georgia Institute of Technology in Atlanta, GA, USA. An avid sports fan his whole life, he has been an active member of the sports visualization community. At IEEE InfoVis 2012, he presented his work as part of the team that designed, developed, and evaluated the SnapShot ice hockey visualization system [PSBS12]. At IEEE VIS 2013, he co-organized the 1st Workshop on Sports Data Visualization [BCC*13] that brought together visualization researchers, visualization practitioners, and sports team executives to discuss the growing subfield. In 2015, he participated in the Data Technologies Workshop hosted by the U.S. Soccer Federation, which provided a forum for conversations between visualization researchers and soccer coaches, trainers, and executives. In addition to his work in sports visualization and visual analytics, he researches graph visualization and visualization-enabled storytelling.

John Stasko is a Professor in the School of Interactive Computing (IC) at the Georgia Institute of Technology, GA, USA. He is a widely published and internationally recognized researcher in the areas of information visualization and visual analytics, approaching each from a human-computer interaction perspective. He is a member of the ACM CHI Academy, and is an ACM Distinguished Scientist and an IEEE Fellow. Stasko is an avid sports fan and has co-authored prior sports data visualization papers [CS06, PSBS12]. Additionally, he was Co-Chair for the 1st IEEE Workshop on Sports Data Visualization [BCC*13] at IEEE VIS 2013 in Atlanta and was an invited speaker at the Data Technologies Workshop hosted by the U.S. Soccer Federation.

Jo Wood is a Professor of Visual Analytics and head of the department of Computer Science at City, University of London, United Kingdom. He is co-director of the giCentre, dedicated to visualization design and analytics. Starting in the fields of Geographic Information Systems and Cartography, he has 25 years of experience in visualization and geographic analysis with particular interests in visualization design for storytelling, personal visualization [Woo15] and geographic and movement analysis [BW14]. He is author of over 100 peer-reviewed publications including several IEEE best paper and honorable mention papers.

Sheelagh Carpendale is a Professor at the University of Calgary, Canada where she holds a Canada Research Chair in Information Visualization and the NSERC/AITF/SMART Industrial Research Chair in Interactive Technologies. She is the recipient of many awards including the NSERC Steacie and the British BAFTA. Her research draws upon her combined backgrounds in computer science and visual arts, benefiting from the interdisciplinary cross-fertilization to enable the design of innovative, people-centred information technologies. By studying how people interact with information both in work and social practices, she works towards designing more natural, accessible and understandable interactive visual representations of data. She combines information visualization, visual analytics and human-computer interaction with innovative interaction techniques to better support the everyday practices of people who are viewing, representing, and interacting with information.

© 2018 The Author(s)
Computer Graphics Forum © 2018 The Eurographics Association and John Wiley & Sons Ltd.