

Sports Data Visualization

Rahul C. Basole ■ Georgia Institute of Technology

Dietmar Saupe ■ University of Konstanz

Few domains involve, generate, and celebrate data in the same manner as sports. For every major sport, analysts can and often do extract large amounts of data, which can be leveraged by media and fans (for entertainment), athletes (to improve individual and team performance), and organizations (to yield a deeper understanding of the sport itself). All major professional sporting events today—ranging from the Olympic Games and the Soccer World Cup to Formula One Racing, the NFL Super Bowl, and Wimbledon—generate multitudinous amounts of real-time data that, when curated and transformed, can provide valuable and potentially novel insight. These efforts are often done in collaboration with leading technology vendors, who also have recognized the tremendous value of sports analytics.

At the same time, we are seeing a surge in personal fitness and sports data generated by wearable technologies and sensors mounted on sports equipment. Individuals use this data to track and improve their performances as well as their personal health. Moreover, combining social media with data from training and competition has been shown to provide valuable incentives and motivation for sports and physical activity involvement.

Massive amounts of data are also generated and used for sports medicine applications, including preventative care and rehabilitation. Most major professional teams today embrace data-driven decisions and employ analytical staff that help prepare training plans, predict athlete injury risks, and prescribe personalized recovery strategies.

Although statistical analysis has primarily dominated much of the sports domain, such as the application of data analytics to personnel decisions

in baseball popularized in the book *Moneyball*,¹ there is a growing interest in applying existing and developing novel visual representations and interactive exploration tools to a range of sports contexts. Journalists, bloggers, and analysts from leading outlets such as ESPN, the *New York Times*, and FiveThirtyEight.com, for instance, use visualization to tell fascinating sports stories, including the groundbreaking three-point shooting performance by Stephen Curry during the 2015–2016 NBA season,² the most exciting goals during the EURO 2016 soccer tournament based on Twitter data (<https://interactive.twitter.com/euro2016/>), and the geography of basketball shots³ and football throws.⁴

The increasing popularity of sports data visualization is also reflected in greater scholarly interest. Prior research design and development of a diverse set of visualization techniques, systems, and applications to sports data can be found at leading conferences such as IEEE Information Visualization (InfoVis), ACM Conference on Human Factors in Computing Systems (CHI), International Association of Computer Science in Sport (IACSS), and MIT Sloan Sports Analytics. Some notable works include examinations of data in soccer,⁵ ice hockey,⁶ and tennis.⁷

Despite this growth in both academia and practice in recent years, many opportunities remain. The ubiquity, diversity, and relative accessibility of sports data makes it a particularly attractive domain for a range of visualization researchers. Sports data tends to be hypervariate, temporal, relational, hierarchical, or a combination thereof, which leads to some fascinating visualization challenges. Potential users of sports data visualizations

range from individuals (professionals and hobbyists) to groups and entire organizations, each with differing sets of tasks and requirements.

Motivated by the significant growth and popularity as well as overall potential of sports data visualization, we sought state-of-the-art research for this special issue. In particular, from a total of 17 submissions, we selected four representative articles for this issue, as well as a fifth that will appear in a future issue, that cover a broad range of sports data visualization topics. We thank all the reviewers for their valuable feedback and comments.

Special Issue Articles

In “StatCast Dashboard: Exploration of Spatiotemporal Baseball Data,” Marcos Lage, Jorge Piazzenti Ono, Daniel Cervone, Justin Chiang, Carlos Dietrich, and Claudio T. Silva present a visualization and analytics infrastructure to help query and facilitate analysis of StatCast, a novel Major League Baseball baseball data system provided that tracks player and ball locations as well as semantically meaningful game events. Rather than just looking at single-play statistics, their system facilitates an interactive visual examination of a collection of games and analysis of comprehensive trends. Through a simple query interface and a set of flexible interactive visualization tools, the authors enable analysts, players, and coaches to gain spatiotemporal insights into the game of baseball, including a precise quantification and assessment of a player’s defensive abilities. As more data and metrics are added to the StatCast engine, a myriad of additional analyses will become possible.

The article by Charles Perin, Jeremy Boy, and Frédéric Vernier, “Using Gap Charts to Visualize the Temporal Evolution of Ranks and Scores,” addresses a fundamental task many sports enthusiasts often have with understanding sports data. How do you effectively portray rankings over time? Traditional approaches generally include table presentations, but they have many limitations, including the inability to provide an understanding of the evolution of multiple metrics. Gap charts, a novel class of line charts, overcome many of these issues by encoding multiple performance metrics (such as rank and score magnitude), leveraging space-filling visualization techniques to create an overlap-free line chart. Using soccer and Tour de France data, the authors demonstrate the value and effectiveness of their approach.

In “Director’s Cut: Analysis and Annotation of Soccer Matches,” Manuel Stein, Halldór Janetzko, Thorsten Breitzkreutz, Daniel Seebacher, Tobias

Schreck, Michael Grossniklaus, Iain D. Couzin, and Daniel A. Keim focus their visual analytics system on soccer analysis based on video combined with motion and event data. The essential ingredients for match analysis are interaction spaces, free spaces, and pass options, all of which are intrinsically difficult for algorithms to estimate to the satisfaction of domain experts. The authors’ contribution provides visual-interactive and data-analysis support for annotating these important characteristics. Together with a rule-based annotation framework, experts can quickly identify key situations and features in matches. The goal is to greatly reduce the effort of traditional manually processed, annotated, and edited video sequences for match analysis and presentation.

Sports data tends to be hypervariate, temporal, relational, hierarchical, or a combination thereof, which leads to some fascinating visualization challenges.

Lastly, in “Sports Tournament Predictions by Direct Manipulation,” Romain Vuillemot and Charles Perin examine how to effectively provide visual analytic support to sports fans who desire to predict team performances in individual games or entire tournaments. The ability to support this effectively is challenging given the complexity of the mental process involved. The authors report on their multistage system design and development journey. Their results show that the design of a direct manipulation interface using a visual tool called Drag-and-Snap can significantly aid with prediction tasks.

Future Research Directions

The articles included in this special issue cover a broad range of topics. Still, they only scratch the surface of potential future research directions. We foresee that as sports become increasingly data driven, we will see many more studies in this field.

Real-time visualization and likewise sonification can be effectively used in feedback training to facilitate better motor learning and to optimize performance, both in sports as well as physical and rehabilitation medicine. Data gathering is becoming increasingly ubiquitous with cheaper sensor technology and easier ways to automatically upload measured data to cloud services that can instantly analyze, visualize, and compare it with

data from previous recordings or other participants.⁸ This has led to intriguing projects like the Quantified Self (self knowledge through numbers, <http://quantifiedself.com>) that also include physical and sports data. The massive amount of such heterogeneous data that is already collected poses great challenges for analysis and visualization, even when focused just on relevance for sports.

Another emerging domain of application may be competitive multiplayer video games, also known as eSports, a growing phenomenon that attracts as many as 100 million spectators per year.⁹ eSports are distinguished from traditional sports in that player actions are mediated through human-computer interfaces that offer direct access for real-time game analysis and visualization in parallel to the video game screen itself.

Advances such as these demonstrate that, while sports data visualization has many context-specific nuances and is clearly important to a diverse set of stakeholders within the domain, it also provides a fertile test ground to study novel visualization techniques that may eventually be applicable to problems in other fields.

References

1. M. Lewis, *Moneyball: The Art of Winning an Unfair Game*, WW Norton & Company, 2004.
2. G. Aisch and K. Quealy, "Stephen Curry's 3-Point Record in Context: Off the Charts," *New York Times*, 16 Apr. 2016; www.nytimes.com/interactive/2016/04/16/upshot/stephen-curry-golden-state-warriors-3-pointers.html.
3. K. Goldsberry, "CourtVision: New Visual and Spatial Analytics for the NBA," MIT Sloan Sports Analytics Conf., 2012; www.sloansportsconference.com/wp-content/uploads/2012/02/Goldsberry_Sloan_Submission.pdf.
4. K. Goldsberry, "Pass Atlas: A Map of Where NFL Quarterbacks Throw the Ball," *Grantland*, 6 Sept. 2013; <http://grantland.com/the-triangle/pass-atlas-a-map-of-where-nfl-quarterbacks-throw-the-ball/>.
5. C. Perin, R. Vuillemot, and J.D. Fekete, "SoccerStories: A kick-off for visual soccer analysis," *IEEE Trans. Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2506–2515.
6. H. Pileggi et al., "Snapshot: Visualization to propel ice hockey analytics," *IEEE Trans. Visualization and Computer Graphics*, vol. 18(12), 2012, pp. 2819–2828.
7. T. Polk et al., "Tennis: Visualization for Tennis Match Analysis," *IEEE Trans. Visualization and Computer Graphics*, vol. 20, no. 12, 2014, pp. 2339–2348.
8. M. Swan, "Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0," *J. Sensor and Actuator Networks*, vol. 1, no. 3, pp. 217–253.
9. "The Burgeoning Evolution of eSports: From the Fringes to Front and Center," PwC, Apr. 2016; https://www.pwc.com/us/en/industry/entertainment-media/assets/pwc_consumer-intelligence-series_esports_april-2016.pdf.

Rahul C. Basole is an associate professor in the School of Interactive Computing and the director of the Tennenbaum Institute at Georgia Tech. A former NCAA athlete and continuing sports enthusiast, his research focuses on computational enterprise science, information visualization, and strategic decision support. Basole has a PhD in industrial and systems engineering from the Georgia Institute of Technology. Contact him at basole@gatech.edu.

Dietmar Saupe is a professor of computer science at the Department of Computer and Information Science at the University of Konstanz, Germany. His research interests include multimedia signal processing for image and video quality assessment and sports informatics. Saupe has a doctorate degree in mathematics from the University of Bremen, Germany. Contact him at dietmar.saupe@uni-konstanz.de.