

# Big Data in Sports Visualization

Josh Lipe-Melton  
Indiana University  
4400 E Sheffield Dr  
Bloomington, Indiana 47408  
jlipemel@umail.iu.edu

## ABSTRACT

This paper will focus on Big Data Analytics in "other sports" (not baseball) and how it is being used to improve and evaluate performance via spatial visualization

## KEYWORDS

sports, analytics, data visualization, spatial rendering

## 1 INTRODUCTION

Analytics and data visualization in sports other than baseball. General problem is that it is very difficult to quantify sports. Furthermore, it is difficult to make use of massive amounts of data for real sports analysis. Big data allows vast quantities of data points to be assessed in a small amount of time. Examples of use cases include SoccerStories, a data visualization tool for soccer which makes use of several different techniques to represent real world events in a series of diagrams. Snapshot is a similar example that uses complex formulas to turn huge quantities of data points into easily readable representations using graphs, symbols, and heat maps. Directorfis Cut is another program that enhances simple statistics with more useful context.

## 2 SHOT CHART VISUALIZATION

Team sport data analysts are met with the challenge of the dynamic, and even chaotic position of players and the ball or puck. Data visualization, however, allows decisions to be made based on large data sets that would otherwise be difficult to understand. Snap Shot is an example of a hockey data visualization technique that allows teams to identify the position, trajectory, and effectiveness of shots throughout the course of a game or season [3]. A specific use case was to analyze several theories about sweetspots on the ice, or positions that a shooter is more likely to score from. One such theory is that goalies tend to be right handed, and that makes it more difficult for them to block shots from their right side due to holding their stick in that hand [3]. This is an intuition held by many high level coaches that was proved to be false after numerous queries through SnapShot [3]. Similarly, NBA shooting was modeled by experts at MIT's Sloan Sports Analytics Conference [?]. Using color and size to represent different data points for shots taken layered onto different positions on the court, one can quickly see where players tend to shoot from and where they are most effective. This type of analysis could be done to show coaches and players where they should be shooting from or what types of plays to draw up for which players.

## 3 SOCCER

Similar to hockey or basketball, soccer is a difficult sport to quantify due to the seemingly subjective methods of evaluating different plays and the randomness of the locations at which plays start and end. One of the most basic and commonly employed data visualization techniques for soccer games is a timeline [2]. This method takes advantage of a predetermined statistic in the game: the length of each half. By using symbols to represent real world events, such as a ball representing a goal, and placing these symbols along the timeline, one is able to gain a limited understanding of the events in the game. This might give simple statistics such as shots, possession, or fouls, more context within the flow of the game and allow a user to gain more insight into what is happening.

### 3.1 Field Position Identification

Deeper analysis of a soccer game typically relates a game event such as a shot or pass with positions on the field. According to soccerstories, the soccer field is the primary object of observation and analysis in soccer. Analysts construct their mental model over the spatial arrangement of the team, and its motion, over time. [2] This type of data can be accrued through the use of wearable technology or through video [2]. One common method of this visualizing this data is a heatmap, through which player's most frequent positions is displayed by density [2]. A heat map allows for an intuitive and instantaneous evaluation of a player's positioning which might otherwise take the length of a match to evaluate. This technique could also be used to identify a player that does not run back to play defense or a player that gets pulled out of position easily. By identifying visual patterns to make insights such as these, teams can gain competitive advantages [2].

### 3.2 Set Piece Analysis

An important part of the scouting report for a soccer game is corner kicks and free kicks. According to researchgate, about 30

### 3.3 Flow Graph Uses

Another method of relating simple statistics to locations on the field is a flow graph, where the size of the nodes shows player's role in the game and the links show the connections between players [2]. A flow graph relates simple statistics about individual players to other individual players and complex insights to be made quickly, such as discerning which players like to pass to one another or which player has a greater impact on the game [2]. This could also be useful in evaluating a team's tendencies such as identifying an inability to attack down the right side or give up more shots on the left. These types of tendencies can be used to make decisions

such as what formation to play or what spaces to run into on counterattacks.

## 4 TEAM SHAPE ANALYSIS

There are several ways team based analysis can often be utilized in data visualization. Directorfis Cut creates an analysis of a teamfis fiback four,fi or defensive line, which can be portrayed by simply drawing a line connecting each of the defenders [1]. A teamfis coach can use this line to identify situations where the back line maintains good flat shape defensively in order to play an offside trap, or alternately situations where the defense gets stretched and could allow a player to get in behind [1].

### 4.1 Player to Player Spatial Relationships

Playerfis proximity to opposing players is another factor that could be useful in evaluating performance. Directorfis cut, for example, breaks down player proximity into three separate categories: no pressure, weak pressure, and strong pressure [1]. This is done by segmenting the soccer field into one meter by one meter squares [1]. Each square is then assigned several attributes regarding the closest player and their speed, direction, and proximity to the square [1]. fiPressurefi on each player can then be determined by viewing the attributes assigned to the square he or she occupies [1]. This can be extremely helpful in analysis of an individual playerfis ability to cope with situations that tend to force mistakes [1]. For example, a playerfis pass completion percentage, the number of passes they complete divided by the number of passes they attempt, can be a useful but slightly misleading statistic. A forward will tend to have lower pass completion percentage than a defender, for example, due to the closer proximity of the other teamfis defenders. By associating pass completion with no, weak, or strong pressure, however, individual players can be analyzed and compared to one another more easily.

## 5 CONCLUSIONS

In conclusion, data visualization is a widely used application of big data in sports. Technology such as player trackers and video analysis allows players, coaches, and managers to gather and communicate insights into sports performances. Many of these visualization techniques are combinations of simple statistics and context such as field position or time remaining. Basketball and hockey use data visualization for analysis of high percentage shots, while soccer focuses on tactics, formations, and player evaluations. There are numerous tools for each of these, all with various methods of carrying out these tasks.

## REFERENCES

- [1] Thorsten Breitkreutz Manuel Stein, Halldr Janetzko. [n. d.]. Director's Cut: Analysis and Annotation of Soccer Matches. ([n. d.]). Retrieved Oct 8, 2017 from <http://ieeexplore.ieee.org.proxyiub.uits.iu.edu/document/7579433/#full-text-section>
- [2] Charles Perin. [n. d.]. SoccerStories: A Kick-off for Visual Soccer Analysis. ([n. d.]). Retrieved Oct 8, 2017 from <http://ieeexplore.ieee.org.proxyiub.uits.iu.edu/document/6634087/#full-text-section>
- [3] Hannah Pileggi. [n. d.]. SnapShot: Visualization to Propel Ice Hockey Analytics. ([n. d.]). Retrieved Oct 7, 2017 from <http://ieeexplore.ieee.org.proxyiub.uits.iu.edu/document/6327288/>

## 6 BIBTEX ISSUES

Warning-I didn't find a database entry for "Sloan"

Warning-empty year in directorscut

Warning-empty year in SoccerStories

Warning-empty year in SnapShot

(There were 4 warnings)

## 7 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

### 7.1 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

### 7.2 Writing Errors

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

### 7.3 Citation Issues and Plagiarism

Overcitation

### 7.4 Character Errors

Erroneous use of quotation marks, i.e. use "quotes" , instead of " " "

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

### 7.5 Structural Issues

Acknowledgement section missing

The paper has less than 2 pages of text, i.e. excluding images, tables and figures